



# Approximate N<sup>3</sup>LO parton distribution functions with theoretical uncertainties: MSHT20aN<sup>3</sup>LO PDFs

J. McGowan<sup>1,a</sup> , T. Cridge<sup>1</sup>, L. A. Harland-Lang<sup>2</sup>, R. S. Thorne<sup>1</sup>

<sup>1</sup> Department of Physics and Astronomy, University College London, London WC1E 6BT, UK

<sup>2</sup> Rudolf Peierls Centre, Beecroft Building, Parks Road, Oxford OX1 3PU, UK

Received: 8 August 2022 / Accepted: 19 January 2023

© The Author(s) 2023

**Abstract** We present the first global analysis of parton distribution functions (PDFs) at approximate N<sup>3</sup>LO in the strong coupling constant  $\alpha_s$ , extending beyond the current highest NNLO achieved in PDF fits. To achieve this, we present a general formalism for the inclusion of theoretical uncertainties associated with the perturbative expansion in the strong coupling. We demonstrate how using the currently available knowledge surrounding the next highest order (N<sup>3</sup>LO) in  $\alpha_s$  can provide consistent, justifiable and explainable approximate N<sup>3</sup>LO (aN<sup>3</sup>LO) PDFs. This includes estimates for uncertainties due to the currently unknown N<sup>3</sup>LO ingredients, but also implicitly some missing higher order uncertainties (MHOUs) beyond these. Specifically, we approximate the splitting functions, transition matrix elements, coefficient functions and  $K$ -factors for multiple processes to N<sup>3</sup>LO. Crucially, these are constrained to be consistent with the wide range of already available information about N<sup>3</sup>LO to match the complete result at this order as accurately as possible. Using this approach we perform a fully consistent approximate N<sup>3</sup>LO global fit within the MSHT framework. This relies on an expansion of the Hessian procedure used in previous MSHT fits to allow for sources of theoretical uncertainties. These are included as nuisance parameters in a global fit, controlled by knowledge and intuition based prior distributions. We analyse the differences between our aN<sup>3</sup>LO PDFs and the standard NNLO PDF set, and study the impact of using aN<sup>3</sup>LO PDFs on the LHC production of a Higgs boson at this order. Finally, we provide guidelines on how these PDFs should be used in phenomenological investigations.

## Contents

### 1 Introduction

<sup>a</sup> e-mail: [j.mcgowan.18@ucl.ac.uk](mailto:j.mcgowan.18@ucl.ac.uk) (corresponding author)

2 Theoretical procedures	...
2.1 Hessian method with nuisance parameters	...
2.2 Multiple theory parameters	...
2.3 Decorrelated parameters	...
3 Structure functions at N <sup>3</sup> LO	...
$F_{2,q}$	...
$F_{2,H}$	...
4 N <sup>3</sup> LO splitting functions	...
4.1 Approximation framework: discrete moments	...
4.2 4-Loop approximations	...
$P_{qg}^{(3)}$	...
$P_{qg}^{\text{NS}, (3)}$	...
$P_{qg}^{\text{PS}, (3)}$	...
$P_{gq}^{(3)}$	...
$P_{gg}^{(3)}$	...
4.3 Predicted aN <sup>3</sup> LO splitting functions	...
4.4 Numerical results	...
5 N <sup>3</sup> LO transition matrix elements	...
5.1 3-Loop approximations	...
$A_{Hg}^{\text{PS}}$	...
$A_{Hg}^{\text{NS}}$	...
$A_{qg,H}^{\text{NS}}$	...
$A_{gg,H}$	...
$A_{gg,H}$	...
5.2 Predicted aN <sup>3</sup> LO transition matrix elements	...
5.3 Numerical results	...
6 N <sup>3</sup> LO heavy coefficient functions	...
6.1 Approximation framework: continuous information	...
6.2 Low- $Q^2$ N <sup>3</sup> LO heavy flavour coefficient functions	...
6.3 3-Loop approximations	...
$C_{H,g}$	...
NNLO	...
N <sup>3</sup> LO	...
$C_{H,g}$	...
NLO and NNLO	...

$N^3LO$ . . . . .	
$C_{q,q}^{\text{NS}}$ . . . . .	
$C_{q,q}^{\text{PS}}$ . . . . .	
$C_{q,g}$ . . . . .	
<b>7</b> $N^3LO$ <i>K</i> -factors . . . . .	
<b>7.1</b> Extension to $aN^3LO$ . . . . .	
<b>7.2</b> Numerical results . . . . .	
<b>8</b> MSHT20 approximate $N^3LO$ global analysis . . . . .	
<b>8.1</b> $\chi^2$ breakdown . . . . .	
<b>8.2</b> Correlation results . . . . .	
<b>8.3</b> Eigenvector results . . . . .	
<b>8.4</b> PDF results . . . . .	
<b>8.5</b> MSHT20a $N^3LO$ PDFs at $Q^2 = 2 \text{ GeV}^2$ . . . . .	
<b>8.6</b> Effect of a $x < 10^{-3}$ cut at $aN^3LO$ . . . . .	
<b>8.7</b> Posterior $N^3LO$ theory parameters . . . . .	
<b>8.8</b> $N^3LO$ contributions . . . . .	
<b>8.9</b> $\alpha_s$ variation . . . . .	
<b>8.10</b> Charm mass dependence . . . . .	
<b>9</b> $N^3LO$ predictions . . . . .	
<b>9.1</b> Higgs production – gluon fusion: $gg \rightarrow H$ . . . . .	
<b>9.2</b> Higgs production – vector boson fusion: $qq \rightarrow H$ . . . . .	
<b>10</b> Availability and recommended usage of MSHT20 $aN^3LO$ PDFs . . . . .	
<b>11</b> Conclusions . . . . .	
<b>Appendix A:</b> List of $N^3LO$ ingredients . . . . .	
<b>Appendix B:</b> $\chi^2$ results without HERA . . . . .	
<b>B.1</b> NNLO . . . . .	
<b>B.2</b> $aN^3LO$ . . . . .	
<b>Appendix C:</b> Dynamic tolerances . . . . .	
<b>C.1</b> Case 1: Decorrelated <i>K</i> -factor parameters . . . . .	
<b>C.2</b> Case 2: Correlated <i>K</i> -factor parameters . . . . .	
<b>Appendix D:</b> Higgs gluon fusion $\mu = m_H$ results . . . . .	
<b>References</b> . . . . .	

## 1 Introduction

In recent years, the level of precision achieved at the LHC has reached far beyond what was once thought possible. This has initiated a new era of high precision phenomenology that has pushed the need for a robust understanding of theoretical uncertainty to new levels. Due to the perturbative nature of calculations in Quantum Chromodynamics (QCD), with respect to the strong coupling constant  $\alpha_s$ , a leading theoretical uncertainty arises from the truncation of perturbative expansions [1,2]. The current state of the art for parton distribution functions (PDFs) is next-to-next-to leading order (NNLO) [3–10]. However, these PDF sets do not generally include theoretical uncertainties arising from the truncation of perturbative calculations that enter the fit. The consideration of these so-called Missing Higher Order Uncertainties

(MHOUs), and how to estimate them, is the topic of much discussion among groups involved in fitting PDFs [11–14].

More recently, a method of utilising a scale variation approach to estimating these uncertainties has been included in an NLO PDF fit [11]. This approach is based upon the fact that to all orders, a physical calculation must not depend on any unphysical scales introduced into calculations. Therefore varying the factorisation and renormalisation scales is, in principle, a first attempt at estimating the level of theory uncertainty from missing higher orders (MHOs). Motivated by the renormalisation group invariance of physical observables, this method is theoretically grounded to all orders. However, the method of scale variations has been shown to be less than ideal in practice [12,15]. An obvious difficulty is the arbitrary nature in the chosen range of the scale variation, as well as the choice of central scale. Expanding on this further, even if a universal treatment of scale variations was agreed upon, these variations are unable to predict the effect of various classes of logarithms (e.g. small- $x$ , mass threshold and leading large- $x$  contributions) present at higher orders. As an example, studies of fits including small- $x$  resummation have recently been done [16,17], showing significant PDF changes. Since it is these type of contributions that are often the most dominant at higher orders, this is an especially concerning pitfall in the use of scale variations to estimate MHOUs. Rather more subtle are the challenges encountered when considering and accounting for correlations between fit and predictions of PDFs [12,14]. An alternative method to the above is to parameterise the missing higher orders with a set of nuisance parameters, using the available (albeit incomplete) current knowledge [18,19].

In this paper we present the first study of an approximate  $N^3LO$  ( $aN^3LO$ ) PDF fit. In particular, we first consider approximations to the  $N^3LO$  structure functions and DGLAP evolution of the PDFs, including the relevant heavy flavour transition matrix elements. We make use of all available knowledge to constrain an approximate parameterisation of the  $N^3LO$  theory, including the calculated Mellin moments, low- $x$  logarithmic behaviour and the full results where they exist. Then for the case of hadronic observables (where less  $N^3LO$  information is available), we include approximate  $N^3LO$  *K*-factors which are guided by the size of known NLO and NNLO corrections. Based on the uncertainty in our knowledge of each  $N^3LO$  function, we obtain a theoretical confidence level (C.L.) constrained by a prior. The corresponding theoretical uncertainties are therefore regulated by our theoretical understanding or lack thereof. Applying the above procedure, we have performed a full global fit at approximate  $N^3LO$ , with a corresponding theoretical uncertainty included within a nuisance parameter framework. As we will show, adopting this procedure allows the correlations and sources of uncertainties to be easily controlled. The preferred form of the  $aN^3LO$  corrections is determined from the

fit quality to data, subject to theoretical constraints from the known information about higher orders.

We note that the source of the above uncertainty is due the (currently unknown) missing ingredients at  $N^3\text{LO}$ , and hence to be precise this corresponds to a ‘missing  $N^3\text{LO}$ ’ uncertainty. However under the common assumption that the dominant uncertainty from missing higher orders (MHOs) is due this uncertainty at the next not fully known  $N^3\text{LO}$  order, one can also expect this to provide a reasonable estimate of MHOs in the fit. Indeed, by allowing the unknown theory parameters to be determined by the fit to data, sensitivity to orders beyond  $N^3\text{LO}$  is explicitly introduced. As we will see, this is particularly transparent in the case of the hadronic  $K$ -factors, which are more directly interpreted as giving a full MHO uncertainty, although a similar sensitivity to higher orders (in particular at low  $x$ ) is observed in the DGLAP evolution of the PDFs. Therefore, while we assume that the majority of this uncertainty is due to the missing information at  $N^3\text{LO}$ , it is the case that some is associated with orders even beyond this, most obviously further effects due to small- $x$  logarithms. Nonetheless, there is in general a distinction between the missing  $N^3\text{LO}$  uncertainty we explicitly include and the uncertainty from MHOs beyond this and hence we will take care throughout this paper to distinguish the two where appropriate, even if the separation is not always clear cut. At present we assume that the majority of this uncertainty is due to the missing information at  $N^3\text{LO}$ , but that some is associated with orders even beyond this, most obviously further effects due to small- $x$  resummation. We will discover that, indeed, the results support this interpretation. As we will show, adopting this procedure allows the correlations and sources of uncertainties to be easily controlled. The preferred form of the  $aN^3\text{LO}$  corrections is determined from the fit quality to data, subject to theoretical constraints from the known information about higher orders.

The outline of this paper is as follows. In Sect. 2 we present the theoretical framework, describing the method and conventions used for the rest of the paper. Section 3 describes the structure functions and their role in QCD calculations. In Sects. 4, 5 and 6 we present our approximations for the  $N^3\text{LO}$  DIS theory functions, while in Sect. 7 we present the  $K$ -factors at  $aN^3\text{LO}$ . In Sect. 8 we present the MSHT  $aN^3\text{LO}$  PDFs with theoretical uncertainties and analyse the implications of the approximations in terms of a full MSHT global fit. Section 9 contains examples of using these  $aN^3\text{LO}$  PDFs in predictions up to  $N^3\text{LO}$ . Finally in Sects. 10 and 11 we present recommendations for how to best utilise these PDFs and summarise our results.

## 2 Theoretical procedures

In this section we describe the mathematical procedures used to implement  $N^3\text{LO}$  approximations into the MSHT PDF framework. These procedures are discussed in terms of the Hessian minimisation method employed by the MSHT fit and extended by theoretically grounded arguments to accommodate theoretical uncertainties.

The above will be achieved by adapting the underlying theory description of the data from NNLO to  $N^3\text{LO}$  (a formal description of how this will be done for the  $F_2$  structure function is discussed in Sect. 3). Not all the ingredients necessary for full  $N^3\text{LO}$  theory predictions are known, where there is missing information the  $N^3\text{LO}$  theory predictions will therefore include additional theoretical nuisance parameters, allowing their variation via an additional degree of freedom in specific theoretical pieces. These theoretical nuisance parameters will be constrained via an additional  $\chi^2$  penalty in the global fit and will accommodate a level of uncertainty for each added approximate  $N^3\text{LO}$  ingredient (more information on how these prior variations are decided is included in Sects. 4.1, 6.1 and 7.1). From this point, the fitting procedure remains similar to previous MSHT fits with a number of extra theory nuisance parameters which are treated in the same manner as experimental nuisance parameters inherent in PDF fits i.e. they can be fit to the data via an expanded Hessian matrix.

### 2.1 Hessian method with nuisance parameters

Following the notation and description from [14], in the Hessian prescription, the Bayesian probability can be written as

$$P(T|D) \propto \exp\left(-\frac{1}{2}(T - D)^T H_0(T - D)\right) \quad (2.1)$$

where  $H_0$  is the Hessian matrix and  $T = \{T_i\}$  is the set of theoretical predictions fit to  $N$  experimental data points  $D = \{D_i\}$  with  $i = 1, \dots, N$ . In this section we explicitly show the adaptation of this equation to accommodate extra theoretical parameters (with penalties) into the total  $\chi^2$  and Hessian matrices.

To adapt this equation to include a single extra theory parameter, we can make the transformation  $T \rightarrow T + tu = T'$ , where  $t$  is the chosen central value of the theory parameter considered and  $u$  is some non-zero vector such that  $uu^T$  is the theory covariance matrix for  $t$ . In defining this new theoretical prescription  $T'$ , we are making the general assumption that the underlying theory is now not necessarily identical to our initial NNLO theory<sup>1</sup>  $T$ .

<sup>1</sup> For the  $aN^3\text{LO}$  prescription defined in this paper this is indeed the case, although for any extra theory parameters that do not inherently change the theory from  $T$  (for example where there is no known  $N^3\text{LO}$

We now seek to include a nuisance parameter  $\theta$ , centered around  $t$ , to allow the fit to control this extra theory addition. We demand that when  $\theta = t$ ,  $T'$  remains unaffected with the theory addition unaltered from its central value  $t$ . This leads us to the expression,

$$T' + (\theta - t)u = T + tu + (\theta - t)u. \quad (2.2)$$

Redefining the nuisance parameter as the shift from its central value  $t$  ( $\theta' = \theta - t$ ) we define  $\theta'$  centered around 0. To constrain  $\theta'$  within the fitting procedure, we must also define a prior probability distribution  $P(\theta')$  centered around zero and characterised by some standard deviation  $\sigma_{\theta'}$ ,

$$P(\theta') = \frac{1}{\sqrt{2\pi}\sigma_{\theta'}} \exp(-\theta'^2/2\sigma_{\theta'}^2). \quad (2.3)$$

Throughout this paper, we refer to the chosen variation of theory predictions in the language of the standard deviation  $\sigma_{\theta'}$  presented here. A caveat to this however is that technically speaking, this standard deviation is chosen with a level of arbitrariness based on general assumptions and known information about the theory (we will show how this is done in more detail in Sects. 4.1, 6.1 and 7.1). Although this definition of  $\sigma_{\theta'}$  lacks the full extent of statistical meaning of a true standard deviation, the same is also true for scale variations as well as various experimental systematic uncertainties, which are often not strictly Gaussian. Furthermore, a more robust statistical meaning is recovered for the constraints on various theoretical parameters after a fit is performed, where we become less sensitive to a prior. Using this information and making the redefinition  $u \rightarrow u/\sigma_{\theta'}$  (in order to normalise the covariance matrix), we can update Eq. (2.1) to be

$$P(T|D\theta) \propto \exp\left(-\frac{1}{2}\left(T' + \frac{(\theta - t)}{\sigma_{\theta'}}u - D\right)^T\right. \\ \times H_0\left(T' + \frac{(\theta - t)}{\sigma_{\theta'}}u - D\right) \quad (2.4)$$

$$P(T'|D\theta') \propto \exp\left(-\frac{1}{2}\left(T' + \frac{\theta'}{\sigma_{\theta'}}u - D\right)^T\right. \\ \times H_0\left(T' + \frac{\theta'}{\sigma_{\theta'}}u - D\right). \quad (2.5)$$

From here, Bayes theorem tells us

$$P(T'|D\theta')P(\theta'|D) = P(\theta'|T'D)P(T'|D) \quad (2.6)$$

where our nuisance parameter  $\theta'$  is assumed to be independent of the data i.e.  $P(\theta'|D) = P(\theta')$ . Integrating over  $\theta'$

information to be included), this transformation still holds in the case that  $t = 0$ .

gives

$$P(T'|D) = \underbrace{\int d\theta' P(\theta'|T'D) P(T'|D)}_{=1} \\ = \int d\theta' P(T'|D\theta')P(\theta'). \quad (2.7)$$

Combining Eqs. (2.3), (2.5) and (2.7) it is possible to show that,

$$P(T'|D) \propto \int d\theta \exp\left(-\frac{1}{2}\left[\left(T' + \frac{\theta'}{\sigma_{\theta'}}u - D\right)^T\right.\right. \\ \times H_0\left(T' + \frac{\theta'}{\sigma_{\theta'}}u - D\right) + \theta'^2/\sigma_{\theta'}^2\left.\right]\right). \quad (2.8)$$

To make progress with this equation we consider the exponent and refactor terms in powers of  $\theta'$ ,

$$\left(u^T H_0 u + 1\right) \frac{\theta'^2}{\sigma_{\theta'}^2} + 2u^T H_0(T' - D) \frac{\theta'}{\sigma_{\theta'}} \\ + (T' - D)^T H_0(T' - D). \quad (2.9)$$

Defining  $M^{-1} = \frac{1}{\sigma_{\theta'}^2}(u^T H_0 u + 1)$  and completing the square gives,

$$M^{-1} \left[ \theta' + \frac{1}{\sigma_{\theta'}} M u^T H_0 (T' - D) \right]^2 \\ - \frac{1}{\sigma_{\theta'}^2} M \left( u^T H_0 (T' - D) \right)^2 + (T' - D)^T H_0 (T' - D). \quad (2.10)$$

In Eq. (2.10), we are able to simplify the first term by defining,

$$\bar{\theta}'(T, D) = \frac{1}{\sigma_{\theta'}} M u^T H_0 (D - T'). \quad (2.11)$$

Expanding the second term leaves us with,

$$\left(u^T H_0 (T' - D)\right)^2 = (T' - D)^T H_0 u u^T H_0 (T' - D). \quad (2.12)$$

The second and third term in Eq. (2.10) can then be combined to give,

$$(T' - D)^T \left( H_0 - \frac{1}{\sigma_{\theta'}^2} M H_0 u u^T H_0 \right) (T' - D). \quad (2.13)$$

Further to this we note that the following is true:

$$(H_0^{-1} + u u^T) \left( H_0 - \frac{1}{\sigma_{\theta'}^2} M H_0 u u^T H_0 \right) \\ = 1 + u u^T H_0 - \frac{1}{\sigma_{\theta'}^2} M u u^T H_0 \\ - \frac{1}{\sigma_{\theta'}^2} M u u^T H_0 u u^T H_0$$

$$= 1 + uu^T H_0 - \frac{1}{\sigma_{\theta'}^2} Muu^T H_0 - \frac{1}{\sigma_{\theta'}^2} Mu(\sigma_{\theta'}^2 M^{-1} - 1)u^T H_0 = 1. \quad (2.14)$$

Using Eq. (2.14) we are finally able to rewrite Eq. (2.8) as,

$$P(T'|D) \propto \int d\theta' \exp \left( -\frac{1}{2} M^{-1}(\theta' - \bar{\theta}')^2 - \frac{1}{2} (T' - D)^T (H_0^{-1} + uu^T)^{-1} (T' - D) \right). \quad (2.15)$$

At this point we can make a choice whether to redefine our Hessian matrix as  $H = (H_0^{-1} + uu^T)^{-1}$ , or keep the contributions completely separate. By redefining the Hessian we can include correlations between the standard set of MSHT parameters included in  $H_0$  and the new theoretical parameter  $\theta'$  contained within  $uu^T$ . However, by doing so we lose information about the specific contributions to the total uncertainty i.e. we cannot then decorrelate the theoretical and standard PDF uncertainties a posteriori. Whereas for the decorrelated choice, although we sacrifice knowledge related to the correlations between the separate sources of uncertainty, we are able to treat the sources completely separably. Interpreting Eq. (2.15) as in Eq. (2.1) we can write down the two  $\chi^2$  contributions,

$$\chi_1^2 = (T' - D)^T (H_0^{-1} + uu^T)^{-1} (T' - D) = (T' - D)^T H (T' - D), \quad (2.16)$$

$$\chi_2^2 = M^{-1}(\theta' - \bar{\theta}')^2, \quad (2.17)$$

where  $\chi_1^2$  is the contribution from the fitting procedure,  $\chi_2^2$  is the posterior penalty contribution applied when the theory addition strays too far from its fitted central value and  $M$  is the posterior error matrix for this contribution. This will be discussed further in following sections.

## 2.2 Multiple theory parameters

In the case of multiple  $N_{\theta'}$  theory parameters, Eq. (2.5) becomes

$$P(T'|D\theta') \propto \exp \left( -\frac{1}{2} \sum_{i,j}^{N_{\text{pts}}} \left( T'_i + \sum_{\alpha=1}^{N_{\theta'}} \frac{\theta'_{\alpha}}{\sigma_{\theta'_{\alpha}}} u_{\alpha,i} - D_i \right) \times H_{ij}^0 \left( T'_j + \sum_{\beta=1}^{N_{\theta'}} \frac{\theta'_{\beta}}{\sigma_{\theta'_{\beta}}} u_{\beta,j} - D_j \right) \right) \quad (2.18)$$

where we have explicitly included the sum over the number of data points  $N_{\text{pts}}$  in the matrix calculation for completeness.

The prior probability for all N<sup>3</sup>LO nuisance parameters also becomes

$$P(\theta') = \prod_{\alpha=1}^{N_{\theta'}} \frac{1}{\sqrt{2\pi} \sigma_{\theta'_{\alpha}}} \exp(-\theta'_{\alpha}^2 / 2\sigma_{\theta'_{\alpha}}^2). \quad (2.19)$$

Constructing  $P(T'|D)$  using Bayes theorem as before, results in the expression,

$$P(T'|D) \propto \int d^{N_{\theta'}} \theta' \exp \left( -\frac{1}{2} \left[ \sum_{i,j}^{N_{\text{pts}}} \left( T'_i + \sum_{\alpha=1}^{N_{\theta'}} \frac{\theta'_{\alpha}}{\sigma_{\theta'_{\alpha}}} u_{\alpha,i} - D_i \right) H_{ij}^0 \right. \right. \\ \left. \left. \times \left( T'_j + \sum_{\beta=1}^{N_{\theta'}} \frac{\theta'_{\beta}}{\sigma_{\theta'_{\beta}}} u_{\beta,j} - D_j \right) + \sum_{\alpha,\beta} \frac{\theta'_{\alpha}}{\sigma_{\theta'_{\alpha}}} \frac{\theta'_{\beta}}{\sigma_{\theta'_{\beta}}} \delta_{\alpha\beta} \right] \right). \quad (2.20)$$

Following the same procedure as laid out in the previous section, defining  $M_{\alpha\beta}^{-1} = (\delta_{\alpha\beta} + u_{\alpha,i} H_{ij}^0 u_{\beta,j}) / \sigma_{\theta'_{\alpha}} \sigma_{\theta'_{\beta}}$  and completing the square leaves us with,

$$(T'_i - D'_i) H_{ij}^0 (T'_j - D'_j) + \sum_{\alpha,\beta}^{N_{\theta'}} M_{\alpha\beta}^{-1} \left[ \left( \theta'_{\alpha} + \sum_{i,j}^{N_{\text{pts}}} \sum_{\delta=1}^{N_{\theta'}} \frac{1}{\sigma_{\theta'_{\alpha}}} M_{\alpha\delta} u_{\delta,i} H_{ij}^0 (T'_j - D_j) \right)^2 - \left( \sum_{i,j}^{N_{\text{pts}}} \sum_{\delta=1}^{N_{\theta'}} \frac{1}{\sigma_{\theta'_{\alpha}}} M_{\alpha\delta} u_{\delta,i} H_{ij}^0 (T'_j - D_j) \right)^2 \right], \quad (2.21)$$

where the summation over the  $\beta$  index in  $M_{\alpha\beta}^{-1}$  is implicit in the squared terms of the squared bracket expressions.

As in the previous section for a single parameter, we can define,

$$\bar{\theta}'_{\alpha}(T', D) = \sum_{i,j}^{N_{\text{pts}}} \sum_{\delta=1}^{N_{\theta'}} \frac{1}{\sigma_{\theta'_{\alpha}}} M_{\alpha\delta} u_{\delta,i} H_{ij}^0 (D_j - T'_j) \quad (2.22)$$

$$H_{ij} = \left( \left( H_{ij}^0 \right)^{-1} + \sum_{\alpha=1}^{N_{\theta'}} u_{\alpha,i} u_{\alpha,j} \right)^{-1} \quad (2.23)$$

which leads to the final expression for  $P(T|D)$ ,

$$P(T'|D) \propto \int d^{N_{\theta'}} \theta' \exp \left( -\frac{1}{2} \left[ \sum_{\alpha,\beta}^{N_{\theta'}} (\theta'_{\alpha} - \bar{\theta}'_{\alpha}) M_{\alpha\beta}^{-1} (\theta'_{\beta} - \bar{\theta}'_{\beta}) \right. \right. \\ \left. \left. + \sum_{i,j}^{N_{\text{pts}}} (T'_i - D_i) H_{ij} (T'_j - D_j) \right] \right). \quad (2.24)$$

which can be interpreted analogously to the single parameter case in (2.15).

### 2.3 Decorrelated parameters

In the treatment above we investigated the case of correlated parameters whereby the Hessian matrix was redefined in Eq. (2.23). In performing this redefinition we sacrifice the information contained within  $u_{\alpha,i}u_{\alpha,j}$  in order to gain information about the correlations between the original PDF parameters making up  $H_{ij}^0$  and any new N<sup>3</sup>LO nuisance parameters. As stated earlier, in this case, we can perform a fit to find  $H_{ij}$  but one is unable to separate this Hessian matrix into individual contributions.

As will be discussed in later sections, the  $K$ -factors we include in the N<sup>3</sup>LO additions are somewhat more separate from other N<sup>3</sup>LO parameters considered. The reason for this is that not only are they concerned with the cross section data directly, they are also included for processes separate from inclusive DIS.<sup>2</sup>

Hence, we have some justification to include the aN<sup>3</sup>LO  $K$ -factors' nuisance parameters as completely decorrelated from other PDF parameters (including other N<sup>3</sup>LO theory parameters). To do this we rewrite Eq. (2.23) as,

$$\begin{aligned} & \left( \left( H_{ij}^0 \right)^{-1} + \sum_{\alpha=1}^{N_{\theta'}} u_{\alpha,i}u_{\alpha,j} + \sum_{p=1}^{N_p} \sum_{\delta=1}^{N_{\theta_K}} u_{\delta,i}^p u_{\delta,j}^p \right)^{-1} \\ &= \left( H_{ij}^{-1} + \sum_{p=1}^{N_p} K_{ij,p}^{-1} \right)^{-1} = H'_{ij} \end{aligned} \quad (2.25)$$

where  $N_{\theta'} \rightarrow N_{\theta'} + N_{\theta_K}$ ,  $K_{ij,p}$  defines the extra decorrelated contributions from the N<sup>3</sup>LO  $K$ -factor's parameters, stemming from  $N_p$  processes;  $H_{ij}$  is the Hessian matrix including correlations with parameters associated with N<sup>3</sup>LO structure function theory; and  $H'_{ij}$  is the fully correlated Hessian matrix. It is therefore possible to construct these matrices separately and perform the normal Hessian eigenvector analysis (described in Sect. 8.3) on each matrix in turn. In doing this, we maintain a high level of flexibility in our description by assuming the sets of parameters (contained in  $H_{ij}^{-1}$  and  $K_{ij,p}$ ) to be suitably orthogonal.

## 3 Structure functions at N<sup>3</sup>LO

The general form of a structure function  $F(x, Q^2)$  is a convolution between the PDFs  $f_i(x, Q^2)$  and some defined process

<sup>2</sup> It is true that we may still expect some indirect correlation with the parameters controlling the N<sup>3</sup>LO splitting functions, which are universal across all processes. However, as we will show, these correlations are small and can be ignored.

dependent coefficient function  $C(x, \alpha_s(Q^2))$ ,

$$F(x, Q^2) = \sum_{i=q, \bar{q}, g} \left[ C_i(\alpha_s(Q^2)) \otimes f_i(Q^2) \right](x) \quad (3.1)$$

where we have the sum over all partons  $i$  and implicitly set the factorisation and renormalisation scales as  $\mu_f^2 = \mu_r^2 = Q^2$ , a choice that will be used throughout this paper for DIS scales. We also note that the relevant charge weightings are implicit in the definition of the coefficient function for each parton.

In Eq. (3.1), the perturbative and non-perturbative regimes are separated out into coefficient functions  $C_i$  and PDFs  $f_i$  respectively. Since these coefficient functions are perturbative quantities, they are an important aspect to consider when transitioning to N<sup>3</sup>LO.

The PDFs  $f_i(x, Q^2)$  in Eq. (3.1) are non-perturbative quantities. However, their evolution in  $Q^2$  is perturbatively calculable. In a PDF fit, the PDFs are parameterised at a chosen starting scale  $Q_0^2$ , which is in general different to the scale  $Q^2$  at which an observable (such as  $F(x, Q^2)$ ) is calculated. It is therefore important that we are able to accurately evolve the PDFs from  $Q_0^2$  to the required  $Q^2$  to ensure a fully consistent and physical calculation. To permit this evolution, we introduce the standard factorisation scale  $\mu_f$ .

The flavour singlet distribution is defined as,

$$\Sigma(x, \mu_f^2) = \sum_{i=1}^{n_f} \left[ q_i(x, \mu_f^2) + \bar{q}_i(x, \mu_f^2) \right], \quad (3.2)$$

where  $q_i(x, \mu_f^2)$  and  $\bar{q}_i(x, \mu_f^2)$  are the quark and anti-quark distributions respectively, as a function of Bjorken  $x$  and the factorisation scale  $\mu_f^2$ . The summation in Eq. (3.2) runs over all flavours of (anti-)quarks  $i$  up to the number of available flavours  $n_f$ .

This singlet distribution is inherently coupled to the gluon density. Because of this, we must consider the gluon carefully when describing the evolution of the flavour singlet distribution with the energy scale  $\mu_f$ . The Dokshitzer–Gribov–Lipatov–Altarelli–Parisi (DGLAP) [20] equations that govern this evolution are:

$$\begin{aligned} \frac{df}{d \ln \mu_f^2} &\equiv \frac{d}{d \ln \mu_f^2} \left( \begin{matrix} \Sigma \\ g \end{matrix} \right) \\ &= \left( \begin{matrix} P_{qq} & n_f P_{qg} \\ P_{gq} & P_{gg} \end{matrix} \right) \otimes \left( \begin{matrix} \Sigma \\ g \end{matrix} \right) \equiv \mathbf{P} \otimes \mathbf{f} \end{aligned} \quad (3.3)$$

where  $P_{ij} : i, j \in q, g$  are the splitting functions and the factorisation scale  $\mu_f$  is allowing the required evolution up to the physical scale  $Q^2$ . The matrix of splitting functions  $\mathbf{P}$  appropriately couples the singlet and gluon distribution by means of a convolution in the momentum fraction  $x$ . We note here that  $P_{qq} \equiv P_{q \rightarrow qg}$  is decomposed into non-singlet (NS) and a pure-singlet (PS) parts defined by,

$$P_{qq}(x) = P_{\text{NS}}^+(x) + P_{\text{PS}}(x), \quad (3.4)$$

where the  $P_{NS}^+$  is a non-singlet distribution splitting function which has been calculated approximately to four loops in [21].<sup>3</sup> The non-singlet part of  $P_{qq}$  dominates at large- $x$  but as  $x \rightarrow 0$ , this contribution is highly suppressed due to the relevant QCD sum rules. On the other hand, due to the involvement of the gluon in the pure-singlet splitting function (as described above), this contribution grows towards small- $x$  and therefore begins to dominate.

Turning to the splitting function matrix, each element can be expanded perturbatively as a function of  $\alpha_s$  up to  $N^3\text{LO}$  as,

$$\mathbf{P}(x, \alpha_s) = \alpha_s \mathbf{P}^{(0)}(x) + \alpha_s^2 \mathbf{P}^{(1)}(x) + \alpha_s^3 \mathbf{P}^{(2)}(x) + \alpha_s^4 \mathbf{P}^{(3)}(x) + \dots, \quad (3.5)$$

where we have omitted the scale argument of  $\alpha_s(\mu_r^2 = \mu_f^2) \equiv \alpha_s$  for brevity and  $\mathbf{P}^{(0)}, \mathbf{P}^{(1)}, \mathbf{P}^{(2)}$  are known [20, 22–27].  $\mathbf{P}^{(3)}$  are the four-loop quantities which we approximate in Sect. 4 using information from [21, 28–36].

Considering Eq. (3.1),  $\Sigma(Q^2)$  and  $g(Q^2)$  are the singlet and gluon PDFs respectively, evolved to the required  $Q^2$  energy of the process via Eq. (3.3). For more information on the relevant formulae used in this convolution, the reader is referred to [37].

Thus far, we have limited our discussion to only light quark flavours. However, as we move through the full range of  $Q^2$  values, the number of partons which are kinematically accessible increases. More specifically, as we pass over the charm and bottom mass thresholds (where  $Q^2 = m_{c,b}^2$ ) we must account for the heavy quark PDFs and their corresponding contributions.

To deal with the heavy quark contributions to the total structure function, whilst remaining consistent with the light quark picture described above, we consider

$$f_\alpha^{n_f+1}(x, Q^2) = \left[ A_{\alpha i}(Q^2/m_h^2) \otimes f_i^{n_f}(Q^2) \right] (x), \quad (3.6)$$

where we have an implied summation over partons  $i$  and  $A_{\alpha i}$  are the heavy flavour transition matrix elements [38, 39] which explicitly depend on the heavy flavour mass threshold  $m_h$ , where these contributions are activated.<sup>4</sup> We also denote the PDFs as  $f_i^{n_f}$  and  $f_i^{n_f+1}$  to indicate whether the PDF has been evolved with only light flavours ( $n_f$ ) or also with heavy flavours ( $n_f+1$ ). In this work we only consider contributions at heavy flavour threshold i.e. where  $Q^2 = m_h^2$ . We then define the PDFs:

$$f_q^{n_f+1}(x, Q^2) = \left[ A_{qq,H}(Q^2/m_h^2) \otimes f_q^{n_f}(Q^2) \right. \\$$

<sup>3</sup> In this discussion, we only consider the  $P_{NS}^+$  non-singlet distribution as this is the distribution which contributes to the singlet evolution. Other non-singlet distributions are briefly discussed in Sect. 4.

<sup>4</sup> The indices here run as  $\alpha \in \{H, q, g\}$  and  $i \in \{q, g\}$ , since  $n_f$  is the number of light flavours.

$$+ A_{qg,H}(Q^2/m_h^2) \otimes f_g^{n_f}(Q^2) \left] (x) \right. \quad (3.7a)$$

$$f_g^{n_f+1}(x, Q^2) = \left[ A_{gq,H}(Q^2/m_h^2) \otimes f_q^{n_f}(Q^2) \right. \\ + A_{gg,H}(Q^2/m_h^2) \otimes f_g^{n_f}(Q^2) \left] (x) \right. \quad (3.7b)$$

$$f_H^{n_f+1}(x, Q^2) = \left[ A_{Hq}(Q^2/m_h^2) \otimes f_q^{n_f}(Q^2) \right. \\ + A_{Hg}(Q^2/m_h^2) \otimes f_g^{n_f}(Q^2) \left] (x) \right. \quad (3.7c)$$

where we have an implicit summation over light flavours of  $q$  and a generalised theoretical description to involve heavy flavour contributions.<sup>5</sup> Equations (3.7a) and (3.7b) are the light flavour quark and gluon PDFs defined earlier, modified to include contributions mediated by heavy flavour loops. Whereas in Eq. (3.7c) we describe the heavy flavour PDF, perturbatively calculated from the light quark and gluon PDFs.

By considering the number of vertices (and hence orders of  $\alpha_s$ ) required for each of these transition matrix elements to contribute to their relevant ‘output’ partons, we are immediately able to show:

$$\begin{aligned} A_{qq,H} &= \delta(1-x) + \mathcal{O}(\alpha_s^2) & A_{gg,H} &= \delta(1-x) + \mathcal{O}(\alpha_s) \\ A_{qg,H} &= \mathcal{O}(\alpha_s^2) & A_{Hq} &= \mathcal{O}(\alpha_s^2) \\ A_{gq,H} &= \mathcal{O}(\alpha_s^2) & A_{Hg} &= \mathcal{O}(\alpha_s) \end{aligned} \quad (3.8)$$

where  $A_{qq,H}$  and  $A_{gg,H}$  include LO  $\delta$ -functions to ensure this description is consistent with the light quark picture discussed earlier. It is therefore the  $A_{Hg}$  transition matrix element which provides our lowest order contribution to the heavy flavour sector (i.e.  $g \rightarrow H\bar{H}$ ).

The insertion of scale independent contributions to  $A_{\alpha i}$  introduce unwanted discontinuities at NNLO into the PDF evolution. In order to ensure the required smoothness and validity of the structure functions across  $(x, Q^2)$ , these discontinuities must be accounted for elsewhere in the structure function picture. Equating the coefficient functions above the mass threshold  $m_h^2$  (describing the total number of flavours including heavy flavour quarks) and those below this threshold, discontinuities are able to be absorbed by a suitable redefinition of the coefficient functions. This procedure provides the foundation for the description of different flavour number schemes.

There are two number schemes which are preferred at different points in the  $Q^2$  range. Towards  $Q^2 \leq m_h^2$  we adopt the Fixed Flavour Number Scheme (FFNS). Towards  $\frac{Q^2}{m_h^2} \rightarrow \infty$ , the heavy contributions can be considered massless and therefore the Zero Mass Variable Flavour Number

<sup>5</sup> Note that the notation  $A_{\alpha i,H}$  is exactly equivalent to  $A_{\alpha i}$ . When  $H$  is not present in the final state of matrix element interactions, we opt for the  $A_{\alpha i,H}$  notation. This is to remind the reader that these elements are considering only those interactions involving a heavy quark.

Scheme (ZM-VFNS) is assumed. In order to join the FFNS and ZM-VFNS schemes seamlessly together, we ultimately wish to describe the General Mass Variable Number Scheme (GM-VFNS) [40] (which is valid across all  $Q^2$ ). This scheme can then account for discontinuities from transition matrix elements and re-establish a smooth description of the structure functions.

In [41] an ambiguity in the definition of the GM-VFNS scheme was pointed out (namely the freedom to swap  $\mathcal{O}(m_h^2/Q^2)$  terms without violating the definition of the GM-VFNS). We note here that since [42], MSHT PDFs have employed the TR scheme to define the distribution of  $\mathcal{O}(m_h^2/Q^2)$  terms, the specific details of which are found in [41,43,44]. The general method to relate the FFNS and GM-VFNS number schemes is to compare the prediction for a result e.g. the  $F_2$  structure function in the FFNS scheme:

$$\begin{aligned} F_2(x, Q^2) &= F_{2,q}(x, Q^2) + F_{2,H}(x, Q^2) \\ &= C_{q,i}^{\text{FF}, n_f} \otimes f_i^{n_f}(Q^2) + C_{H,i}^{\text{FF}, n_f} \otimes f_k^{n_f}(Q^2) \\ &= C_{q,q}^{\text{FF}, n_f} \otimes f_q^{n_f}(Q^2) + C_{q,g}^{\text{FF}, n_f} \otimes f_g^{n_f}(Q^2) \\ &\quad + C_{H,q}^{\text{FF}, n_f} \otimes f_q^{n_f}(Q^2) + C_{H,g}^{\text{FF}, n_f} \otimes f_g^{n_f}(Q^2) \end{aligned} \quad (3.9)$$

and the GM-VFNS scheme,

$$\begin{aligned} F_2(x, Q^2) &= \sum_{\alpha \in \{H, q, g\}} \left( C_{q,\alpha}^{\text{VF}, n_f+1} \otimes A_{\alpha i}(Q^2/m_h^2) \otimes f_i^{n_f}(Q^2) \right. \\ &\quad \left. + C_{H,\alpha}^{\text{VF}, n_f+1} \otimes A_{\alpha i}(Q^2/m_h^2) \otimes f_i^{n_f}(Q^2) \right), \end{aligned} \quad (3.10)$$

where  $F_{2,q}$  and  $F_{2,H}$  are the light and heavy flavour structure functions respectively.<sup>6</sup>  $C^{\text{FF}, n_f}$  and  $C^{\text{VF}, n_f+1}$  are the FFNS (known up to NLO [45,46] with some information at NNLO [47–49] including high- $Q^2$  transition matrix elements at  $\mathcal{O}(\alpha_s^3)$  [49–55]) and GM-VFNS coefficient functions respectively, and  $A_{\alpha i}(Q^2/m_h^2)$  are the transition matrix elements. We note that the above also applies to other structure functions and for clarity, in the following we consider the light and heavy structure functions separately.

$$F_{2,q}$$

Expanding the first term in Eq. (3.10) in terms of the transition matrix elements results in,

$$F_{2,q}(x, Q^2) = C_{q,H}^{\text{VF}, n_f+1} \otimes \left[ A_{Hq}(Q^2/m_h^2) \otimes f_q^{n_f}(Q^2) \right]$$

<sup>6</sup> The extra contribution from  $F_{2,H}$  allows for the possibility of final state heavy flavours.

$$\begin{aligned} &+ A_{Hg}(Q^2/m_h^2) \otimes f_g^{n_f}(Q^2) \Big] \\ &+ C_{q,q}^{\text{VF}, n_f+1} \otimes \left[ A_{qq,H}(Q^2/m_h^2) \otimes f_q^{n_f}(Q^2) \right. \\ &\quad \left. + A_{qg,H}(Q^2/m_h^2) \otimes f_g^{n_f}(Q^2) \right] \\ &+ C_{q,g}^{\text{VF}, n_f+1} \otimes \left[ A_{gq,H}(Q^2/m_h^2) \otimes f_q^{n_f}(Q^2) \right. \\ &\quad \left. + A_{gg,H}(Q^2/m_h^2) \otimes f_g^{n_f}(Q^2) \right], \end{aligned} \quad (3.11)$$

which is valid at all orders. The first term in Eq. (3.11) is the contribution to the light quark structure function from heavy quark PDFs (since the term contained within square brackets is exactly our definition in Eq. (3.7c)). Due to this, the coefficient function  $C_{q,H}$  describes the transition of a heavy quark to a light quark via a gluon and is therefore forbidden to exist below NNLO. The second and third terms here are the purely light quark and gluon contributions, with extra corrections from heavy quark at higher orders.

Using the definitions in Eq. (3.8) we can obtain an equation for  $F_{2,q}(x, Q^2)$  up to  $\mathcal{O}(\alpha_s^3)$  as,

$$\begin{aligned} F_{2,q}(x, Q^2) &= C_{q,q}^{\text{VF}, (0)} \otimes f_q(Q^2) \\ &+ \frac{\alpha_s}{4\pi} \left\{ C_{q,q, n_f+1}^{\text{VF}, (1)} \otimes f_q(Q^2) + C_{q,g, n_f+1}^{\text{VF}, (1)} \otimes f_g(Q^2) \right\} \\ &+ \left( \frac{\alpha_s}{4\pi} \right)^2 \left\{ \left[ C_{q,q, n_f+1}^{\text{VF}, (2)} + C_{q,q}^{\text{VF}, (0)} \otimes A_{qq,H}^{(2)} \right] \otimes f_q(Q^2) \right. \\ &\quad \left. + \left[ C_{q,g, n_f+1}^{\text{VF}, (2)} + C_{q,g, n_f+1}^{\text{VF}, (1)} \otimes A_{gg,H}^{(1)} \right. \right. \\ &\quad \left. \left. + C_{q,q}^{\text{VF}, (0)} \otimes A_{qg,H}^{(2)} \right] \otimes f_g(Q^2) \right\} \\ &+ \left( \frac{\alpha_s}{4\pi} \right)^3 \left\{ \left[ C_{q,q, n_f+1}^{\text{VF}, (3)} + C_{q,q, n_f+1}^{\text{VF}, (1)} \otimes A_{qq,H}^{(2)} \right. \right. \\ &\quad \left. + C_{q,g, n_f+1}^{\text{VF}, (1)} \otimes A_{qg,H}^{(2)} + C_{q,q}^{\text{VF}, (0)} \otimes A_{qq,H}^{(3)} \right] \otimes f_q(Q^2) \\ &\quad + \left[ C_{q,g, n_f+1}^{\text{VF}, (3)} + C_{q,g, n_f+1}^{\text{VF}, (1)} \otimes A_{gg,H}^{(2)} + C_{q,q, n_f+1}^{\text{VF}, (1)} \otimes A_{qg,H}^{(2)} \right. \\ &\quad \left. + C_{q,g, n_f+1}^{\text{VF}, (2)} \otimes A_{gg,H}^{(1)} + C_{q,q}^{\text{VF}, (0)} \otimes A_{qg,H}^{(3)} \right] \otimes f_g(Q^2) \\ &\quad \left. + C_{q,H}^{\text{VF}, (2)} \otimes A_{Hg}^{(1)} \otimes f_g(Q^2) \right\} + \mathcal{O}(\alpha_s^4) \end{aligned} \quad (3.12)$$

where  $C_{q,q}^{\text{VF}, (0)} = \delta(1-x)$  up to charge weighting. Eq. (3.12) defines the light quark structure function to  $N^3\text{LO}$  including heavy flavour corrections.<sup>7</sup>

<sup>7</sup> We also note that  $\alpha_s^{n_f+1} \neq \alpha_s^{n_f}$  and account for this, but omit in expressions such as Eq. (3.12) for simplicity.

$$F_{2,H}$$

Moving to the heavy quark structure function in Eq. (3.9), as above the second term in Eq. (3.10) can be expanded in terms of the transition matrix elements to obtain,

$$\begin{aligned} F_{2,H}(x, Q^2) = & C_{H,H}^{\text{VF}, n_f+1} \otimes \left[ A_{Hq}(Q^2/m_h^2) \otimes f_q^{n_f}(Q^2) \right. \\ & + A_{Hg}(Q^2/m_h^2) \otimes f_g^{n_f}(Q^2) \left. \right] \\ & + C_{H,q}^{\text{VF}, n_f+1} \otimes \left[ A_{qq,H}(Q^2/m_h^2) \otimes f_q^{n_f}(Q^2) \right. \\ & + A_{qg,H}(Q^2/m_h^2) \otimes f_g^{n_f}(Q^2) \left. \right] \\ & + C_{H,g}^{\text{VF}, n_f+1} \otimes \left[ A_{gq,H}(Q^2/m_h^2) \otimes f_q^{n_f}(Q^2) \right. \\ & + A_{gg,H}(Q^2/m_h^2) \otimes f_g^{n_f}(Q^2) \left. \right], \quad (3.13) \end{aligned}$$

which is valid at all orders. Similar to Eq. (3.11), we have a contribution from the heavy flavour quarks, the light quarks and the gluon respectively. However in this case, due to the required gluon intermediary, the coefficient functions associated with the light quark flavours and gluon are forbidden to exist below NNLO. Considering the  $C_{H,H}$  function, we are able to choose this to be identically the ZM-VFNS light quark coefficient function  $C_{q,q}$  up to kinematical suppression factors, since at  $Q^2 \rightarrow \infty$  these functions must be equivalent [40,44,56].

The full heavy flavour structure function then reads as,

$$\begin{aligned} F_{2,H}(x, Q^2) = & \frac{\alpha_s}{4\pi} \left[ C_{H,g}^{\text{VF}, (1)} + C_{H,H}^{\text{VF}, (0)} \otimes A_{Hg}^{(1)} \right] \otimes f_g(Q^2) \\ & + \left( \frac{\alpha_s}{4\pi} \right)^2 \left\{ \left[ C_{H,q}^{\text{VF}, (2)} + C_{H,H}^{\text{VF}, (0)} \otimes A_{Hq}^{(2)} \right] \otimes f_q(Q^2) \right. \\ & + \left[ C_{H,g}^{\text{VF}, (2)} + C_{H,g}^{\text{VF}, (1)} \otimes A_{gg,H}^{(1)} + C_{H,H}^{\text{VF}, (1)} \otimes A_{Hg}^{(1)} \right. \\ & \left. \left. + C_{H,H}^{\text{VF}, (0)} \otimes A_{Hg}^{(2)} \right] \otimes f_g(Q^2) \right\} \\ & + \left( \frac{\alpha_s}{4\pi} \right)^3 \left\{ \left[ C_{H,q}^{\text{VF}, (3)} + C_{H,g}^{\text{VF}, (1)} \otimes A_{gq,H}^{(2)} \right. \right. \\ & + C_{H,H}^{\text{VF}, (1)} \otimes A_{Hq}^{(2)} + C_{H,H}^{\text{VF}, (0)} \otimes A_{Hg}^{(3)} \left. \right] \otimes f_q(Q^2) \\ & + \left[ C_{H,g}^{\text{VF}, (3)} + C_{H,g}^{\text{VF}, (2)} \otimes A_{gg,H}^{(1)} + C_{H,g}^{\text{VF}, (1)} \otimes A_{gg,H}^{(2)} \right. \\ & + C_{H,H}^{\text{VF}, (2)} \otimes A_{Hg}^{(1)} \\ & \left. \left. + C_{H,H}^{\text{VF}, (1)} \otimes A_{Hg}^{(2)} + C_{H,H}^{\text{VF}, (0)} \otimes A_{Hg}^{(3)} \right] \otimes f_g(Q^2) \right\} \quad (3.14) \end{aligned}$$

where combining Eqs. (3.12) and (3.14), one can obtain the full structure function  $F_2(x, Q^2)$ . Equating the FFNS expansion from Eq. (3.9) to the above expressions in the GM-VFNS

setting, one can find relationships between the two pictures. In Sect. 6 we use this equivalence to enable the derivation of the GM-VFNS functions at  $N^3\text{LO}$ .

To summarise, we have identified the leading theoretical ingredients entering the structure functions and detailed how these affect the PDFs. As we will discuss further, when pushing these equations to  $N^3\text{LO}$ , there is already some knowledge available. For example, the  $N^3\text{LO}$  ZM-VFNS coefficient functions are known precisely for  $n_f = 3$  from [57], as are a handful of Mellin moments [21,35,36,50] and leading small and large- $x$  terms [28–34,49,51–54] associated with the splitting functions and transition matrix elements at  $N^3\text{LO}$ . Using this information, we approximate these functions to  $N^3\text{LO}$  and incorporate the results into the first approximate  $N^3\text{LO}$  global PDF fit.

## 4 $N^3\text{LO}$ splitting functions

Splitting functions at  $N^3\text{LO}$  allow us to more accurately describe the evolution of the PDFs. These functions are estimated here and the resulting approximations are included within the framework described in Sect. 2 and below in Sect. 4.1. In all singlet cases we set  $n_f = 4$  before constructing our approximations and ignore any corrections to this from any further change in the number of flavours.<sup>8</sup> In the non-singlet case, we calculate the approximate parts of  $P_{qg}^{NS(3)}$  with  $n_f = 4$  however, there is a relatively large amount of information about the  $n_f$ -dependence included from [21]. Therefore in the final result we choose to allow the full  $n_f$ -dependence to remain for the non-singlet splitting function.

### 4.1 Approximation framework: discrete moments

In order to estimate the missing  $N^3\text{LO}$  uncertainty in the splitting functions (also transition matrix elements considered in the following Sect. 5), and ultimately include these into the framework described in Sect. 2.2, one must acquire some approximation at  $N^3\text{LO}$ . Here we discuss using available sets of discrete Mellin moments for each function, along with any exact leading terms already calculated, to obtain  $N^3\text{LO}$  estimations. To perform the parameterisation of the unknown  $N^3\text{LO}$  quantities, we follow a similar estimation procedure as in [58,59] following the form,

$$F(x) = \sum_{i=1}^{N_m} A_i f_i(x) + f_e(x). \quad (4.1)$$

<sup>8</sup> An exception to this are the cases of  $P_{qg}$  and  $P_{qg}^{\text{PS}}$  where we have already defined  $P_{qg} \equiv n_f P_{qg}$  and  $P_{qg}^{\text{PS}} \equiv n_f P_{qg}^{\text{PS}}$ . Therefore the leading  $n_f$  dependence is already taken into account.

In Eq. (4.1),  $N_m$  is the number of available moments,  $A_i$  are calculable coefficients,  $f_i(x)$  are functions chosen based on our intuition and theoretical understanding of the full function, and  $f_e(x)$  encapsulates all the currently known leading exact contributions at either large or small- $x$ . To describe this, consider a toy situation where we are given four data points described by some unknown degree 9 polynomial. Along with this information, we are told the dominant term at small- $x$  is described by  $3x$ . In this case, one may wish to attempt to approximate this function by means of a set of 4 simultaneous equations formed from Eq. (4.1) equated to each of the four data points (or constraints). The result of this is then a unique solution for each chosen set of functions  $\{f_i(x)\}$ . However, a byproduct of this is that for each  $\{f_i(x)\}$ , one lacks any means to control the uncertainty in these approximate solutions. In order to allow a controllable level of uncertainty into this approximation, one must introduce an extra degree of freedom. This degree of freedom will be introduced through an unknown coefficient  $a \equiv A_{N_m+1}$ , which for convenience, will be absorbed into the definition of  $f_e(x) \rightarrow f_e(x, a)$ . In this toy example one is then able to choose to define the functions  $f_i(x)$  as,

$$\begin{aligned} f_1(x) &= x^3 \text{ or } x^4, \\ f_2(x) &= x^5, \text{ or } x^6 \\ f_3(x) &= x^7 \text{ or } x^8, \\ f_4(x) &= x^9, \\ f_e(x, a) &= 3x + ax^2, \end{aligned} \quad (4.2)$$

where we have prioritised approximating the small- $x$  behaviour more precisely than the large- $x$  behaviour. This could easily be adapted and even reversed depending on which region of  $x$  we are most sensitive to, however in this paper we will be more focused on small- $x$ . There is also an inherent functional uncertainty from the ambiguity in the choice of functions for  $f_{1,2,3}(x)$  in this toy example, in principle the number of functions in the functional variation can be larger than demonstrated here and indeed a larger choice of functions will be used for all  $f_i(x)$  when we apply this in practice in subsequent sections. Using these functions, one is then able to assemble a set of potential approximations to the overall polynomial, each uniquely defined by a set of functions and corresponding coefficients  $\{A_i, f_i\}$  for each value of  $a$ .

As mentioned, for the  $N^3LO$  additions considered in this framework we use the available calculated moments as constraints for the corresponding simultaneous equations. A summary of all the known and used ingredients for all  $N^3LO$  approximations is provided in Appendix A. The details of these known quantities will be discussed in detail in Sects. 4.2 and 5.1. We also mention here that towards the small- $x$  regime, the leading terms present in the splitting functions

and transition matrix elements exhibit the relations,

$$F_{gg}(x \rightarrow 0) \simeq \frac{C_A}{C_F} F_{gq}(x \rightarrow 0), \quad (4.3a)$$

$$F_{gq}(x \rightarrow 0) \simeq \frac{C_F}{C_A} F_{qg}(x \rightarrow 0), \quad (4.3b)$$

where  $F_{ij} \in \{P_{ij}, A_{ij,H}\}$  and  $C_A, C_F$  are the usual QCD constants. Although Eq. (4.3) are exact at leading order, it is known that as we expand to higher orders, these will break down due to the effect of large sub-leading logarithms. Due to this, we do not demand this relation as a constraint in our approximations. Instead we discuss the validity of Eq. (4.3) in comparison with the  $aN^3LO$  functions.

Following from [58, 59], we must choose a set of candidate functions for each  $f_i(x)$ . Our convention is to assign these functions such that at small- $x$ ,  $f_1(x)$  is dominant, while at large- $x$ ,  $f_{N_m}(x)$  is dominant. With  $f_i(x) \forall i \in \{2, \dots, N_m - 1\}$ , dominating in the region between. The sets of functions assigned to each  $f_i(x)$  are determined for each  $N^3LO$  function based on knowledge from lower orders and our intuition about what to expect at  $N^3LO$ .

Analogous to our toy polynomial example, we allow the inclusion of an unknown next-to-leading small- $x$  logarithm (NLL) term (NNLL in the  $P_{gg}$  case) into the  $f_e$  function of our parameterisation. The coefficient of this NLL (NNLL) term is then controlled by a variational parameter  $a$ . This parameter uniquely defines the solution to the sets of simultaneous equations considered i.e. for each set of functions  $f_i(x)$  there exists a unique solution for every possible choice of  $a$ . The final step to consider in this approximation is how to choose the prior allowed variation of  $a$  in a sensible way for each  $N^3LO$  approximation. To do this, we consider the criteria outlined below:

**Criterion 1:** At sufficiently small- $x$  ( $x < 10^{-5}$ ), for a fixed value of  $a$ , we require  $f_e(x, a)$  to be contained within the range of variation for  $F(x)$  predicted from the combinations of functions in (4.2). For example, after fixing  $a$ ,  $f_e(x, a)$  it should lie within the variation predicted for  $F(x)$  from the entire set of potential approximations defined in (4.2). In practice this means that we require the small- $x$  behaviour to not be in large tension with the large- $x$  description.

**Criterion 2:** At large- $x$  ( $x > 10^{-2}$ ) the  $N^3LO$  contribution should have relatively little effect. More specifically, we do not expect as large of a divergence as we do at small- $x$ . Due to this, we require that the trend of the  $N^3LO$  approximation follow the general trend of the NNLO function at large- $x$ .

The allowed variation in  $a$  gives us an uncertainty which, at its foundations, is chosen via a conservative estimate based on all the available prior knowledge about the function and lower orders being considered. We note that given we are including known information about the higher order, it is not guaranteed that a value of  $a = 0$  will satisfy either criterion 1 or 2. Indeed, typically the NLL coefficient in the splitting functions is the opposite sign to and larger than the LL contribution, for example in the NNLO splitting functions and the known NLL term in the N<sup>3</sup>LO splitting function  $P_{gg}$ . To determine a full predicted uncertainty for the function and allow for a computationally efficient fixed functional form, the variation of  $a$  can absorb the uncertainty from the ambiguity in the choice of functions  $f_i(x)$  (essentially expanding the allowed range of  $a$  – as will be shown in the following sections). Since the functions are approximations themselves, increasing the allowed variation of  $a$  to encapsulate the total uncertainty predicted by the initial treatment described above is a valid simplification.

A worked example following this procedure is provided for the  $P_{qg}^{(3)}$  and  $A_{Hg}^{(3)}$  functions in Sects. 4.2 and 5.1 respectively.

#### 4.2 4-Loop approximations

$$P_{qg}^{(3)}$$

We begin by considering the four-loop quark-gluon splitting function. Here we provide a more detailed explanation of the method described in Sect. 4.1 which will then be applied to the remaining splitting functions considered in this section. Four even-integer moments are known for  $P_{qg}^{(3)}$  ( $n_f = 4$ ) from [35, 36], along with the LL small- $x$  term from [28].

The functions made available for the  $P_{qg}$  analysis are,

$$\begin{aligned} f_1(x) &= \frac{1}{x} \quad \text{or} \quad \ln^4 x \quad \text{or} \quad \ln^3 x \quad \text{or} \quad \ln^2 x, \\ f_2(x) &= \ln x, \\ f_3(x) &= 1 \quad \text{or} \quad x \quad \text{or} \quad x^2, \\ f_4(x) &= \ln^4(1-x) \quad \text{or} \quad \ln^3(1-x) \quad \text{or} \\ &\quad \ln^2(1-x) \quad \text{or} \quad \ln(1-x), \\ f_e(x, \rho_{qg}) &= \frac{C_A^3}{3\pi^4} \left( \frac{82}{81} + 2\zeta_3 \right) \\ &\quad \times \frac{1}{2} \frac{\ln^2 1/x}{x} + \rho_{qg} \frac{\ln 1/x}{x}, \end{aligned} \quad (4.4)$$

where  $\rho_{qg}$  is the variational parameter. This is then varied between  $-2.5 < \rho_{qg} < -0.9$ , which has been chosen to satisfy the criteria described in Sect. 4.1. The set of functions in Eq. (4.4) is chosen from the analysis of lower orders. Specifically, following the pattern of functions from lower

orders, it can be shown that at this order we expect the most dominant large- $x$  term to be  $\ln^4(1-x)$  and  $\ln^4 x$  to be the highest power of  $\ln x$  at small- $x$ .

Figure 1 displays an example of the variation found from the different choices of functions that encapsulate the chosen range of  $\rho_{qg}$ . We also show the upper (A) and lower (B) bounds (at small- $x$ ) for the entire uncertainty (solid line) combining the variation in the functions and in the variation of  $\rho_{qg}$ . The upper ( $P_{qg}^{(3),A}$ ) and lower ( $P_{qg}^{(3),B}$ ) bounds are given by,

$$\begin{aligned} P_{qg}^{(3),A} &= 1.6699 \frac{1}{x} + 2.4167 \ln x \\ &\quad - 2.2011 x^2 + 0.0024228 \ln^4(1-x) \\ &\quad + \frac{C_A^3}{3\pi^4} \left( \frac{82}{81} + 2\zeta_3 \right) \frac{1}{2} \frac{\ln^2 1/x}{x} - 0.9 \frac{\ln 1/x}{x}, \end{aligned} \quad (4.5)$$

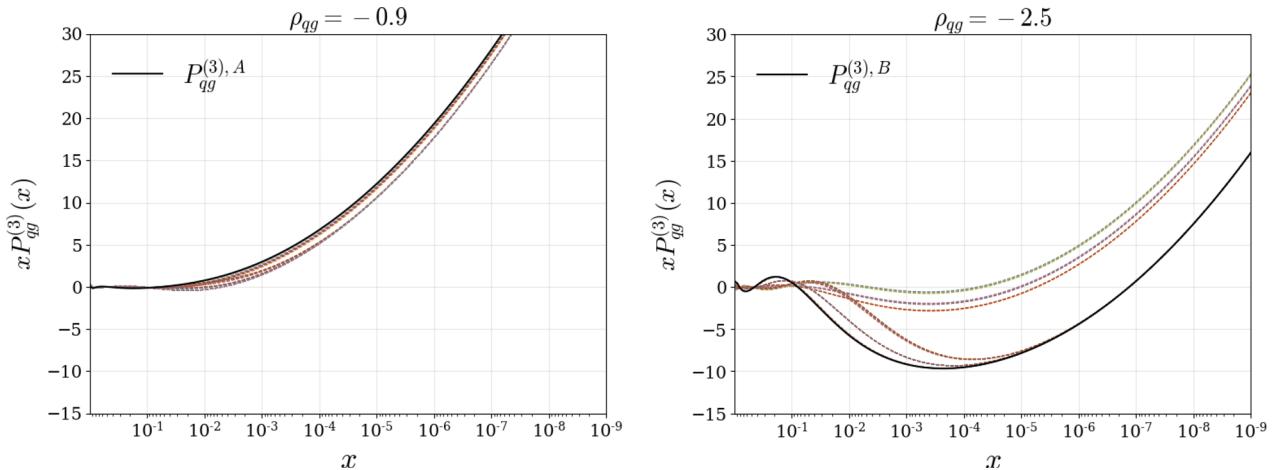
$$\begin{aligned} P_{qg}^{(3),B} &= 12.582 \ln^2 x + 5.3065 \ln x \\ &\quad + 1.7957 x^2 - 0.0041296 \ln^4(1-x) \\ &\quad + \frac{C_A^3}{3\pi^4} \left( \frac{82}{81} + 2\zeta_3 \right) \frac{1}{2} \frac{\ln^2 1/x}{x} - 2.5 \frac{\ln 1/x}{x}. \end{aligned} \quad (4.6)$$

Using this information, a fixed functional form is chosen to be,

$$\begin{aligned} P_{qg}^{(3)} &= A_1 \ln^2 x + A_2 \ln x + A_3 x^2 + A_4 \ln^4(1-x) \\ &\quad + \frac{C_A^3}{3\pi^4} \left( \frac{82}{81} + 2\zeta_3 \right) \frac{1}{2} \frac{\ln^2 1/x}{x} + \rho_{qg} \frac{\ln 1/x}{x} \end{aligned} \quad (4.7)$$

and  $\rho_{qg}$  is allowed to vary as  $-2.5 < \rho_{qg} < -0.8$ . This fixed functional form identically matches with the lower bound  $P_{qg}^{(3),B}$  and the expansion of the variation of  $\rho_{qg}$  enables (to within  $\sim 1\%$ ) the absorption of the small- $x$  upper bound uncertainty (predicted from  $P_{qg}^{(3),A}$ ) into the variation.<sup>9</sup> In other areas of  $x$  there are larger deviations from the upper bound ( $\sim 10\%$ ) when using this convenient fixed functional form. However, in these regions the function is already relatively small, therefore any larger percentage deviations are negligible. Also since the heuristic choice of variation found earlier is intended as a guide, we are not bound by any solid constraints to precisely reconstruct it with our subsequent choice of fixed functional form. Therefore it is entirely justified to be able to slightly adapt the shape of the variation in less dominant regions.

<sup>9</sup> Explicitly, the range is expanded from  $-2.5 < \rho_{qg} < -0.9$  to  $-2.5 < \rho_{qg} < -0.8$ , in order to absorb the functional variation lost by moving to a fixed functional form (for implementation purposes).



**Fig. 1** Combinations of functions with an added variational factor ( $\rho_{qg}$ ) controlling the NLL term. Combinations of functions at the upper (left) and lower (right) bounds of the variation are shown. The solid lines indicate the upper and

lower bounds for this function chosen from the relevant criteria

$$P_{qg}^{\text{NS}, (3)}$$

As discussed in Sect. 3, the quark-quark splitting function is comprised of a pure-singlet and non-singlet contribution. We approximate each part independently, although the final quark-quark singlet function will be almost completely dominated by the pure-singlet, except at very high- $x$ .

The four-loop non-singlet splitting function has been the subject of relatively extensive research and is known exactly for a number of regimes. For example in [21], some important exact contributions to the four-loop non-singlet splitting functions are presented, along with 8 even-integer moments for each of the + and – distributions [21]. In this discussion we are exclusively approximating the non-singlet + distribution, as this is the part that contributes to the full singlet quark-quark splitting function. The other relevant non-singlet distributions  $P_{\text{NS}}^{(3), -}$  and  $P_{\text{NS}}^{(3), \text{sea}}$  (described in [26]), are set to the central values predicted from [21] since any variation in these functions are negligible. All presently known information is used in this approximation, with results similar to that seen in [21] but with our own choice of functions.

$$\begin{aligned} f_1(x) &= \frac{1}{(1-x)_+}, & f_2(x) &= (1-x) \ln(1-x), \\ f_3(x) &= (1-x) \ln^2(1-x), \\ f_4(x) &= (1-x) \ln^3(1-x), & f_5(x) &= 1, \\ f_6(x) &= x, & f_7(x) &= x^2, & f_8(x) &= \ln^2 x, \\ f_e(x, \rho_{qg}^{\text{NS}}) &= C_F n_c^3 P_{L,0}^{(3)}(x) + C_F n_c^2 n_f P_{L,1}^{(3)}(x) \\ &\quad + P_{Ln_f}^{(3)+}(x) + \rho_{qg}^{\text{NS}} \ln^3 x \\ &\quad - 55.876 \ln^4 x - 2.8313 \ln^5 x - 0.14883 \ln^6 x \end{aligned}$$

$$\begin{aligned} &- 2601.7 - 2118.9 \ln(1-x) \\ &+ n_f (4.6584 \ln^4 x + 0.2798 \ln^5 x \\ &+ 312.16 + 337.93 \ln(1-x)) \end{aligned} \quad (4.8)$$

where the functions  $C_F n_c^3 P_{L,0}^{(3)}(x) + C_F n_c^2 n_f P_{L,1}^{(3)}(x)$  and  $P_{Ln_f}^{(3)+}(x)$  can be found in Equation (4.11) and Equation (4.14) respectively within [21], and  $\rho_{qg}^{\text{NS}}$  is our variational parameter. Note that the ansatz from Eq. (4.1) has been extended to include 8 pairs of functions and coefficients, to accommodate 8 known moments. Within the  $f_e(x, \rho_{qg}^{\text{NS}})$  part of Eq. (4.8), we have chosen to vary the coefficient of the most divergent unknown small- $x$  term ( $\ln^3 x$ ) with the variation across  $0 < \rho_{qg}^{\text{NS}} < 0.014$ . Due to the high level of information and larger number of functions allowed to be included, we ignore any functional uncertainty and explicitly define each function. Therefore the only variation needed to be considered as an uncertainty stems from the variation of  $\rho_{qg}^{\text{NS}}$ .

The resulting approximation is then,

$$\begin{aligned} P_{\text{NS}}^{(3), +} &= A_1 \frac{1}{(1-x)_+} + A_2 (1-x) \ln(1-x) \\ &\quad + A_3 (1-x) \ln^2(1-x) \\ &\quad + A_4 (1-x) \ln^3(1-x) + A_5 + A_6 x \\ &\quad + A_7 x^2 + A_8 \ln^2 x + f_e(x, \rho_{qg}^{\text{NS}}), \end{aligned} \quad (4.9)$$

where no alterations are made to the allowed range of  $0 < \rho_{qg}^{\text{NS}} < 0.014$ .

$$P_{qq}^{\text{PS}, (3)}$$

We now restrict our analysis to focus on approximating the pure-singlet part of  $P_{qq}^{(3)}$ , thereby providing a more accurate set of functions with a focus on the small- $x$  regime. To ensure the  $P_{qq}^{\text{PS}, (3)}$  function does not interfere with the large- $x$  regime (where the non-singlet description dominates) the ansatz from Eq. (4.1) is adapted to be:

$$P_{ij}^{(3)}(x) = \left\{ A_1 f_1(x) + A_2 f_2(x) + A_3 f_3(x) + A_4 f_4(x) \right\} (1-x) + f_e(x, \rho_{qq}^{\text{PS}}). \quad (4.10)$$

This modified parameterisation guarantees that any instabilities in the pure singlet approximation will not wash out the non-singlet behaviour at large- $x$ .

Using four available even-integer moments for  $n_f = 4$  [35, 36] and the exact small- $x$  information [28], the chosen set of functions for this approximation is,

$$\begin{aligned} f_1(x) &= \frac{1}{x} \text{ or } \ln^4 x, \\ f_2(x) &= \ln^3 x \text{ or } \ln^2 x \text{ or } \ln x, \\ f_3(x) &= 1 \text{ or } x \text{ or } x^2, \\ f_4(x) &= \ln^4(1-x) \text{ or } \ln^3(1-x) \text{ or } \ln^2(1-x) \text{ or } \ln(1-x), \\ f_e(x, \rho_{qq}^{\text{PS}}) &= \frac{C_A^2 C_F}{3\pi^4} \left( \frac{82}{81} + 2\zeta_3 \right) \frac{1}{2} \frac{\ln^2 1/x}{x} \\ &\quad + \rho_{qq}^{\text{PS}} \frac{\ln 1/x}{x}, \end{aligned} \quad (4.11)$$

where  $\rho_{qq}^{\text{PS}}$  is varied as  $-0.7 < \rho_{qq}^{\text{PS}} < 0$ . For the variation produced from stable combinations of these functions, we coincidentally end up with the same functional form for both the upper  $P_{\text{PS}}^{(3), A}$  and lower  $P_{\text{PS}}^{(3), B}$  bounds. Therefore trivially, the fixed functional form is defined as:

$$\begin{aligned} P_{\text{PS}}^{(3)} = & \left\{ A_1 \frac{1}{x} + A_2 \ln^2 x + A_3 x^2 \right. \\ & \left. + A_4 \ln^2(1-x) \right\} (1-x) \\ & + \frac{C_A^2 C_F}{3\pi^4} \left( \frac{82}{81} + 2\zeta_3 \right) \frac{1}{2} \frac{\ln^2 1/x}{x} \\ & + \rho_{qq}^{\text{PS}} \frac{\ln 1/x}{x} (1-x) \end{aligned} \quad (4.12)$$

where the variation of  $\rho_{qq}^{\text{PS}}$  is unchanged and the entire predicted variation is encapsulated in this form.

$$P_{gq}^{(3)}$$

As with the previous singlet splitting functions, four even-integer moments for  $n_f = 4$  are known [35, 36] along with

the LL small- $x$  information [29–31]. The set of functions made available for the combinations in our approximation are stated as,

$$\begin{aligned} f_1(x) &= \frac{\ln 1/x}{x} \text{ or } \frac{1}{x}, \\ f_2(x) &= \ln^3 x, \\ f_3(x) &= x \text{ or } x^2, \\ f_4(x) &= \ln^4(1-x) \text{ or } \ln^3(1-x) \text{ or } \ln^2(1-x) \text{ or } \ln(1-x), \\ f_e(x, \rho_{gq}) &= \frac{C_A^3 C_F}{3\pi^4} \zeta_3 \frac{\ln^3 1/x}{x} \\ &\quad + \rho_{gq} \frac{\ln^2 1/x}{x}, \end{aligned} \quad (4.13)$$

where  $\rho_{gq}$  is set as  $\rho_{gq} = -1.8$ . In this case, the variation from the choice of functions is large enough to satisfy the criteria in Sect. 4.1 and encapsulate a sensible  $\pm 1\sigma$  variation without including any further variation in  $\rho_{gq}$ . Similarly to previous approximations, for stable variations we estimate this variation with the fixed functional form,

$$\begin{aligned} P_{gq}^{(3)} = & A_1 \frac{\ln 1/x}{x} + A_2 \ln^3 x + A_3 x \\ & + A_4 \ln(1-x) + \frac{C_A^3 C_F}{3\pi^4} \zeta_3 \frac{\ln^3 1/x}{x} \\ & + \rho_{gq} \frac{\ln^2 1/x}{x} \end{aligned} \quad (4.14)$$

where the allowed range of  $\rho_{gq}$  is expanded to  $-1.8 < \rho_{gq} < -1.5$  to approximate the variation from the choice of functions. As with the  $P_{qg}^{(3)}$  fixed functional form, this new range recovers a variation which is within  $\sim 1\%$  of the original, in the dominant areas of  $x$ .

$$P_{gg}^{(3)}$$

Finally we move to the approximation of the gluon-gluon splitting function, where four available even-integer moments for  $P_{gg}^{(3)}(n_f = 4)$  are known from [35, 36]. The list of functions (including the known small- $x$  LL and NLL terms from [29–33]) used for the approximation is,

$$\begin{aligned} f_1(x) &= \frac{1}{x} \text{ or } \ln^3 x \text{ or } \ln^2 x, \\ f_2(x) &= \ln x, \\ f_3(x) &= 1 \text{ or } x \text{ or } x^2, \\ f_4(x) &= \frac{1}{(1-x)_+} \text{ or } \ln^2(1-x) \text{ or } \ln(1-x), \\ f_e(x, \rho_{gg}) &= \frac{C_A^4}{3\pi^4} \zeta_3 \frac{\ln^3 1/x}{x} + \frac{1}{\pi^4} \left[ C_A^4 \left( -\frac{1205}{162} \right) \right. \\ &\quad \left. + \frac{C_A^2 C_F}{3\pi^4} \zeta_3 \frac{\ln^3 1/x}{x} \right] \end{aligned}$$

$$\begin{aligned}
& + \frac{67}{36} \zeta_2 + \frac{1}{4} \zeta_2^2 - \frac{11}{2} \zeta_3 \Big) \\
& + n_f C_A^3 \left( -\frac{233}{162} + \frac{13}{36} \zeta_2 - \frac{1}{3} \zeta_3 \right) \\
& + n_f C_A^2 C_F \left( \frac{617}{243} - \frac{13}{18} \zeta_2 + \frac{2}{3} \zeta_3 \right) \Big] \frac{1}{2} \frac{\ln^2 1/x}{x} \\
& + \rho_{gg} \frac{\ln 1/x}{x}, \tag{4.15}
\end{aligned}$$

where  $\rho_{gg}$  is varied as  $-5 < \rho_{gg} < 15$  and  $n_f = 4$ . The fixed functional form is then chosen to be,

$$\begin{aligned}
P_{gg}^{(3)} = & A_1 \ln^2 x + A_2 \ln x + A_3 x^2 \\
& + A_4 \ln^2(1-x) + \frac{C_A^4}{3\pi^4} \zeta_3 \frac{\ln^3 1/x}{x} \\
& + \frac{1}{\pi^4} \left[ C_A^4 \left( -\frac{1205}{162} + \frac{67}{36} \zeta_2 + \frac{1}{4} \zeta_2^2 - \frac{11}{2} \zeta_3 \right) \right. \\
& + n_f C_A^3 \left( -\frac{233}{162} + \frac{13}{36} \zeta_2 - \frac{1}{3} \zeta_3 \right) \\
& \left. + n_f C_A^2 C_F \left( \frac{617}{243} - \frac{13}{18} \zeta_2 + \frac{2}{3} \zeta_3 \right) \right] \frac{1}{2} \frac{\ln^2 1/x}{x} \\
& + \rho_{gg} \frac{\ln 1/x}{x}, \quad (n_f = 4) \tag{4.16}
\end{aligned}$$

where we maintain the variation of  $\rho_{gg}$  from above, as the fixed functional form manages to encapsulate the variation predicted, without any extra allowed  $\rho_{gg}$  variation.

#### 4.3 Predicted aN<sup>3</sup>LO splitting functions

Figures 2, 3 and 4 show the perturbative expansions for each splitting function up to approximate N<sup>3</sup>LO. Included with these expansions are the predicted variations ( $\pm 1\sigma$ ) from Sect. 4.2 (shown in green) and the aN<sup>3</sup>LO best fits (shown in blue – discussed further in Sect. 8). As a general feature, we observe that the singlet N<sup>3</sup>LO approximations are much more divergent than lower orders due to the presence of higher order logarithms at small- $x$ , further highlighting the need for an understanding of MHOUs beyond the default NNLO considered in current PDF sets in a way that is not reliant on the NNLO central value.

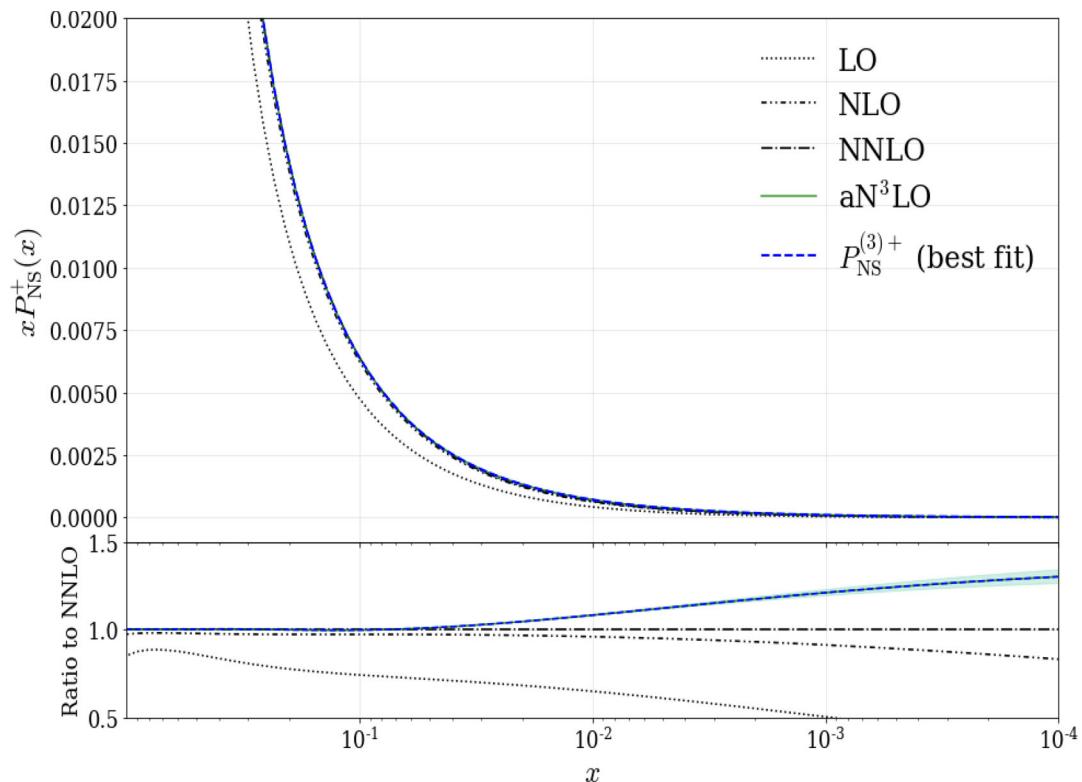
Considering the non-singlet case shown in Fig. 2, we see a very close agreement at large- $x$  between  $P_{qq}^{\text{NS}}$  expanded to NNLO and aN<sup>3</sup>LO. This is a general feature of the non-singlet distribution, since by design, this distribution is largely unaffected by small- $x$  contributions. The ratio plot in Fig. 2 provides clearer evidence for this, since it is only towards small- $x$  (where the non-singlet distribution tends towards 0) that any noticeable difference between NNLO and aN<sup>3</sup>LO can be seen.

The contributions to  $P_{qq}^{\text{PS}}$ ,  $P_{qg}$ ,  $P_{gq}$  and  $P_{gg}$  shown in Figs. 3 and 4 respectively, display a much richer description at

aN<sup>3</sup>LO. In all cases, the divergent terms (with  $x \rightarrow 0$ ) present in the approximations have a large effect from intermediate- $x$  ( $\sim 10^{-2}$ ) down to very small- $x$  values. The asymptotic relationships (red line) Eq. (4.3) defined using the best fit values of the aN<sup>3</sup>LO expansions (i.e. comparable to the blue dashed line) are also shown in Figs. 3 and 4. As discussed earlier, these relations are violated by large sub-leading small- $x$  terms and are therefore provided here as a qualitative comparison. Furthermore, we also observe a close resemblance to the N<sup>3</sup>LO asymptotic results in Fig. 4 of [34]. Specifically for quark evolution, we show that the data prefers a similar form ( $P_{qq}^{\text{PS}}$  and  $P_{qg}$ ) to the resummed splitting function results in [34] whereas for gluon evolution, this agreement is less prominent.

Superimposed onto these variations in Figs. 3 and 4 are the best fit values for the splitting functions, as predicted from a global fit of the full MSHT approximate N<sup>3</sup>LO PDFs. The full fit results will be discussed in more detail in Sect. 8, however we note here that the fit produces relatively good agreement with the prior allowed variations for each of the splitting functions. For all functions except for  $P_{gg}$ , the best fit results lie within their  $\pm 1\sigma$  variation range. This result implies that constraints from the data included in the global fit are in good agreement with the penalties describing quark evolution (i.e.  $P_{\text{PS}}$  and  $P_{qg}$  in Fig. 3). For the gluon evolution in Fig. 4 we observe a small level of tension with the data pushing towards a slightly harder small- $x$  gluon than preferred by the penalty constraints for  $P_{gg}$ . An important caveat to these best fit results is that the data included in the fit is sensitive to all orders in  $\alpha_s$ . Therefore by proxy, the best fit predictions are also sensitive to corrections at all orders. This will certainly be a driving factor for any violations away from the expected N<sup>3</sup>LO behaviour. However, since the ultimate goal of this investigation is to provide a theoretical uncertainty, the violation from higher orders is manifested into the defined penalties and therefore accounted for in the fit as a source of MHOU.

Finally, an important feature that can be seen across all these splitting function plots are points of zero aN<sup>3</sup>LO uncertainty in the high- $x$  regions. The regions where these points occur are where the moments are constraining the chosen fixed functional forms very tightly. In particular, for  $N_m$  moments (constraints) in Eq. (4.1), we are left with  $N_m - 1$  points of zero uncertainty predicted from our approximations. As stated, these points are dependent on the choice of our fixed functional form and are therefore regions where the uncertainty has been underestimated when compared to the functional uncertainty which the fixed form approximates. To provide a more complete estimate of the uncertainty in these areas, it would be necessary to smooth the uncertainty band out across these regions (or take into account several fixed functional forms). However, this shortcoming only occurs towards large- $x$ , where the uncertainty is naturally



**Fig. 2** Perturbative expansion up to  $aN^3LO$  for the non-singlet splitting function  $P_{qg}^{NS, +}$  including any corresponding allowed  $\pm 1\sigma$  variation (shaded green region). The best fit value (blue dashed line) displays the prediction for this function determined from a global PDF fit

smaller across these functions. Therefore if the uncertainty was smoothed, the effect would be negligible for the theoretical uncertainty this work aims to include in a PDF global fit. Further to this, these functions are ingredients in the DGLAP convolution where any smaller details are washed out by more dominant features inside convolutions with PDFs. For these reasons, we opt for computational efficiency and leave these points as shown.

#### 4.3.1 Moment analysis

Tracking back to the moments found for the splitting functions [35, 36] (shown in Table 1 and as a ratio in Fig. 5), we are able to identify the expected convergence in the perturbative expansions up to  $N^3LO$ . Figure 5 illustrates the relative size of the NNLO and  $N^3LO$  contributions to the low even-integer moments.

Until recently (at the time of writing), there were only 3 moments available for the functions  $P_{gq}$  and  $P_{gg}$  approximated here. However, in [36] an extra moment was published for these two gluon splitting functions. This extra information led to our predictions at small- $x$  being more in line with the resummation results in [34] mentioned earlier. This is an example of how extra information can be added as and when it is available to update any approximations and utilise our

full knowledge of the next highest order. By adopting this procedure, we immediately benefit from a slightly increased precision (with a relevant theoretical uncertainty) instead of having to delay the inclusion of higher order theory (for potentially decades) until a complete analytical calculation of the next order in  $\alpha_s$  is known.

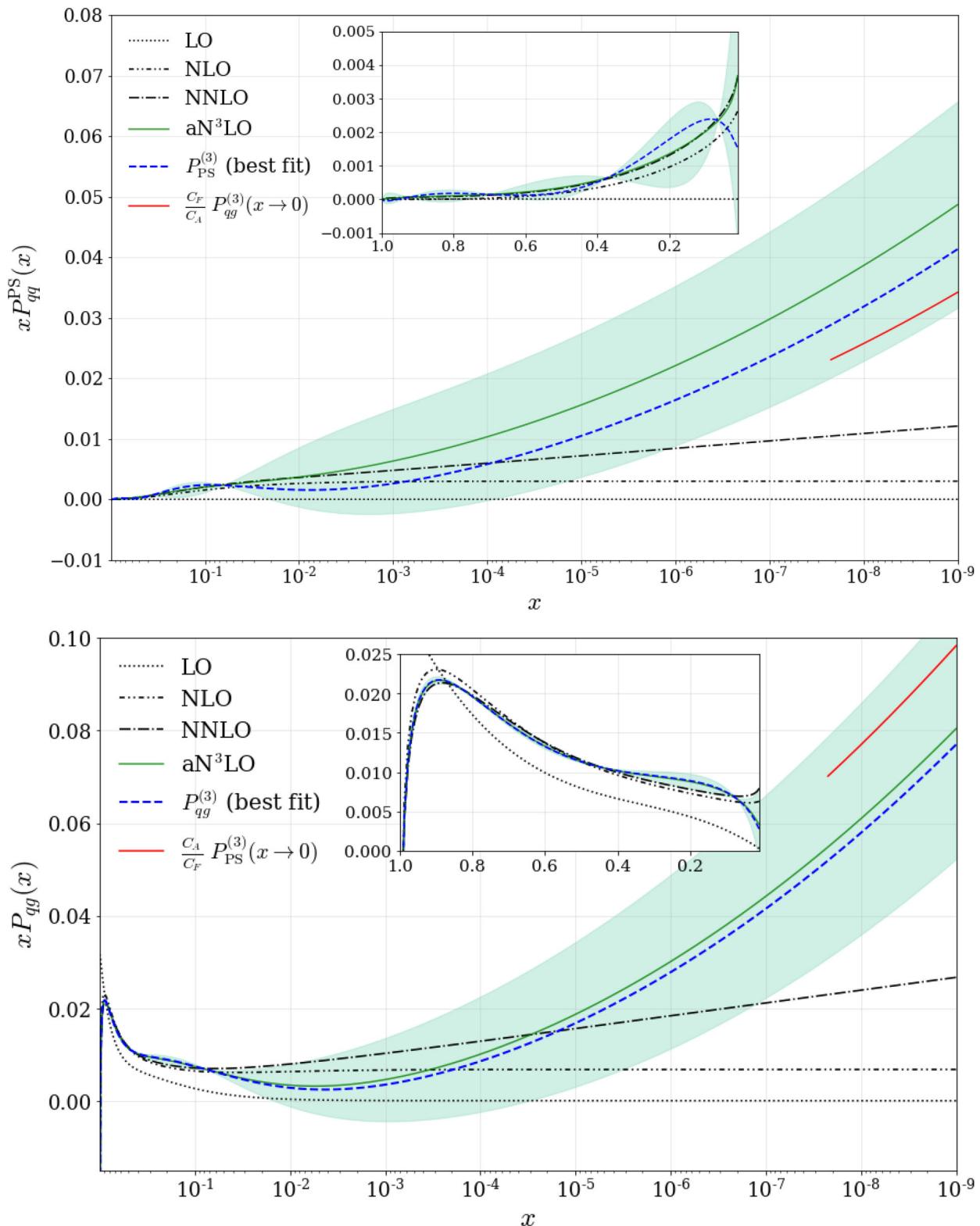
#### 4.4 Numerical results

We now consider the DGLAP evolution equations for the singlet and gluon shown in Eq. (3.3). We expand this equation to  $\alpha_s^4$  and investigate the effects of the variation in the  $N^3LO$  contributions.

For the purposes of this analysis, the approximate functions (4.17), taken from [27], are used as sample distributions at an energy scale of  $\mu_f^2 \simeq 30 \text{ GeV}^2$ , a scale chosen due to its relevance to DIS processes included in the MSHT global fit.

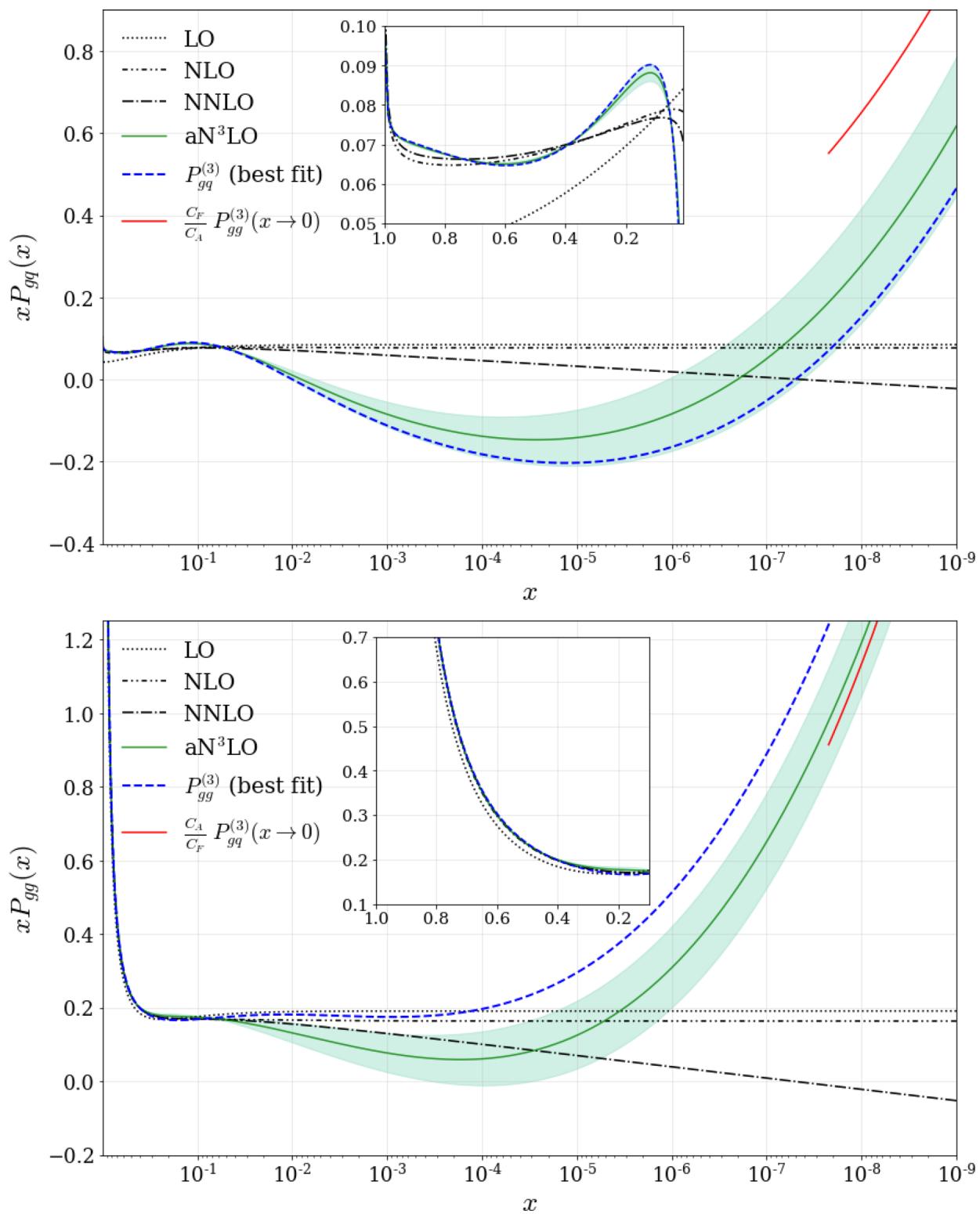
$$x\Sigma(x, \mu_f^2 = 30 \text{ GeV}^2) = 0.6 x^{-0.3} (1-x)^{3.5} (1+5x^{0.8}) \quad (4.17a)$$

$$xg(x, \mu_f^2 = 30 \text{ GeV}^2) = 1.6 x^{-0.3} (1-x)^{4.5} (1-0.6x^{0.3}) \quad (4.17b)$$



**Fig. 3** Perturbative expansions up to aN<sup>3</sup>LO for the quark singlet splitting functions  $P_{qg}^{PS}$  (top) and  $P_{qg}$  (bottom) including any corresponding allowed  $\pm 1\sigma$  variation (shaded green region). The best fit values (blue dashed line) display the

predictions for each function determined from a global PDF fit



**Fig. 4** Perturbative expansions for the gluon splitting functions  $P_{gq}$  (top) and  $P_{gg}$  (bottom) including any corresponding allowed  $\pm 1\sigma$  variation (shaded green region). The best fit value (blue dashed line) displays the prediction for this function determined from a global PDF fit

**Table 1** Numerical moments of singlet and gluon splitting function moments up to  $N^3\text{LO}$  for  $\alpha_s = 0.2$  and  $n_f = 4$ 

	Moment	LO	NLO	NNLO	$N^3\text{LO}$
$P_{qq}^{\text{PS}}$	$N = 2$	-0.056588	-0.06362642	-0.06395712	-0.06412109
	$N = 4$	-0.11104	-0.1261481	-0.12804822	-0.12835549
	$N = 6$	-0.14329	-0.16188618	-0.16433013	-0.16470246
	$N = 8$	-0.166448	-0.18751366	-0.19033329	-0.19074888
$P_{qg}$	$N = 2$	0.042442	0.05008496	0.04991043	0.04983007
	$N = 4$	0.023342	0.02203438	0.02110201	0.02112623
	$N = 6$	0.016674	0.01387744	0.01311037	0.01316929
	$N = 8$	0.013086	0.00979920	0.00919186	0.00927006
$P_{gq}$	$N = 2$	0.056588	0.06362642	0.06395712	0.06412109
	$N = 4$	0.015562	0.01903295	0.0195455	0.01965547
	$N = 6$	0.008892	0.0112073	0.01158133	0.0116615
	$N = 8$	0.006232	0.00801547	0.00831037	0.0083761
$P_{gg}$	$N = 2$	-0.042442	-0.05008496	-0.04991043	-0.04983007
	$N = 4$	-0.242978	-0.26161441	-0.26280015	-0.26326763
	$N = 6$	-0.32551	-0.35114066	-0.35335022	-0.35384552
	$N = 8$	-0.38091	-0.41151668	-0.41447721	-0.41495604

The expressions above are order independent and so provide a robust means to isolate the effects arising from higher orders in the splitting functions. For convenience we also assume

$$\alpha_s(\mu_r^2 = \mu_f^2 = 30 \text{ GeV}^2) \simeq 0.2. \quad (4.18)$$

where  $\mu_r$  and  $\mu_f$  are the renormalisation and factorisation scales respectively.

### *Singlet evolution*

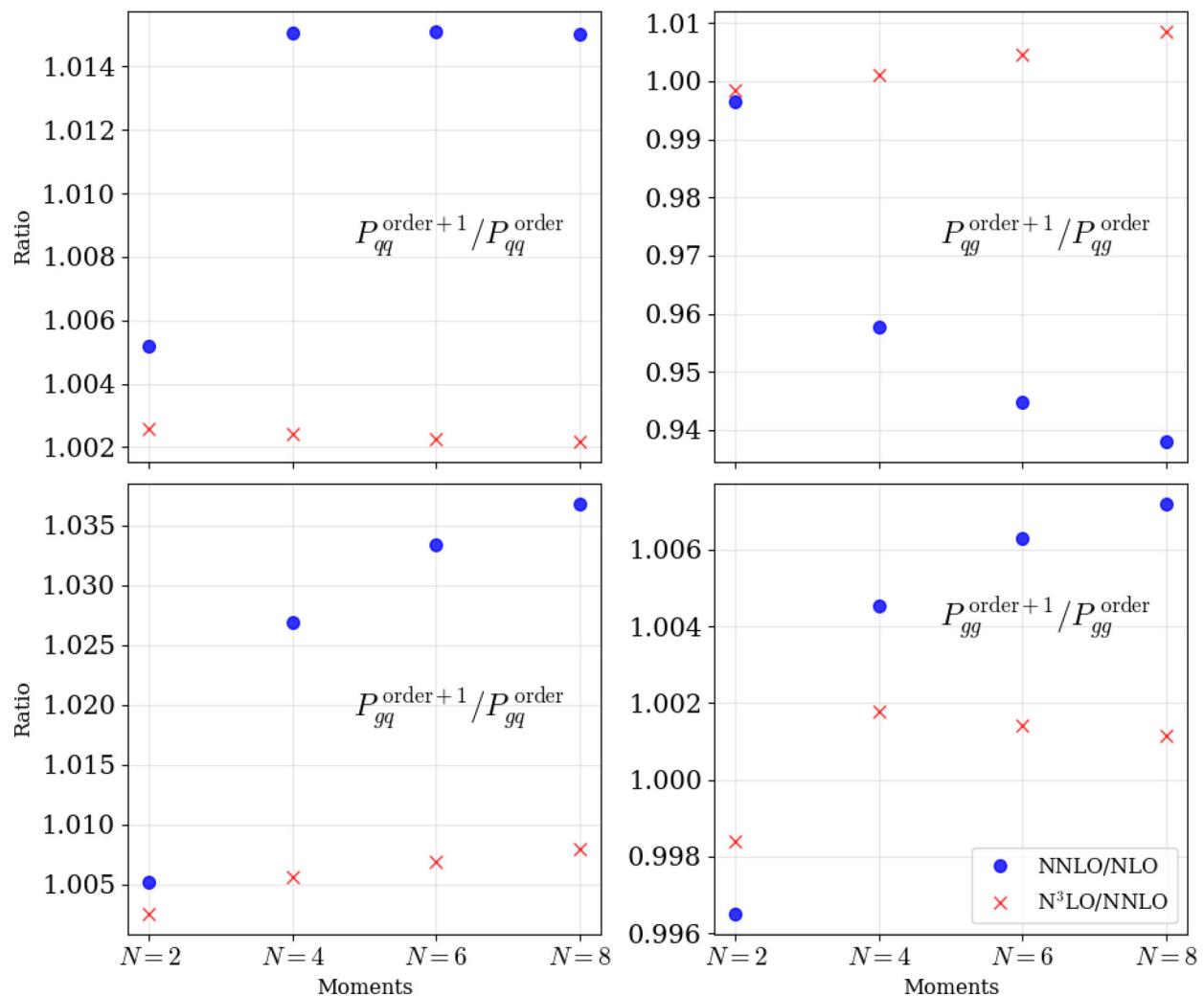
Figure 6 demonstrates the result of including the respective  $N^3\text{LO}$  expansions from Sect. 4.2 in an analysis of the evolution equation. Towards small- $x$  this variation increases due to the larger uncertainty in the  $P_{qq}^{\text{PS}}$  and  $P_{qg}$  splitting functions at a $N^3\text{LO}$ . On the right of Fig. 6, the difference plot displays the respective shifts from the previous order and demonstrates how this shift changes up to  $N^3\text{LO}$ . These results predict a reduction in the evolution of the singlet towards small- $x$  from NNLO. Inspecting Fig. 3, we can see that this reduction is stemming from the contribution of the gluon with the  $P_{qg}$  function at 4-loops, which is the dominant contribution to the evolution. Towards larger  $x$  values ( $10^{-2} < x < 10^{-1}$ ) we see a fractional increase in the quark evolution, also following the shape of the  $P_{qg}$  function. These results can therefore give some indication as to how we expect our gluon PDF to behave at  $N^3\text{LO}$ ; since the structure functions are directly related to the quarks (through LO), the singlet evolution should remain fairly constant. Therefore we can expect that the fit will prefer a slightly harder gluon at small- $x$  and a softer gluon between  $10^{-2} < x < 10^{-1}$  relative to NNLO.

Figure 6 displays a good level of agreement between the allowed  $N^3\text{LO}$  shift and the evolution at NLO and NNLO (within  $\pm 1\sigma$  variation bands from theoretical uncertainties). Also shown in Fig. 6 is the evolution prediction using the best fit results for  $P_{qq}^{(3)}$  and  $P_{qg}^{(3)}$  (red dashed). This prediction tends to follow slightly below the center of the  $1\sigma$  uncertainty band, where the data has balanced the two variations and is more in line with the NLO evolution than NNLO due to a negative contribution below  $10^{-2}$ . Considering the magnitude of shifts from each order, the predicted shift from NNLO to a $N^3\text{LO}$  is slightly larger than that from NLO to NNLO, contradicting what may be expected from perturbation theory. However, we remind the reader that these best fit results are, to some degree, sensitive to all orders in perturbation theory through the data constraint. Due to this, the resultant best fit can be thought of as an approximate asymptote to all orders. Interpreting the approximation in this way, restores our faith in perturbation theory and becomes an entirely plausible estimation of the missing higher orders.

Figure 6 also exhibits an example of how the points of zero uncertainty (discussed in Sect. 4.2) can affect the evolution predictions. We can see that at most the uncertainty is being underestimated by  $< 1\%$  and therefore, for the reasons discussed earlier, we do not consider these regions further here.

### *Gluon evolution*

Figure 7 displays the result of including the a $N^3\text{LO}$  splitting function contributions into the gluon evolution equation. As with the singlet evolution case, this extra contribution is currently inducing a notable variation at  $N^3\text{LO}$ . The general



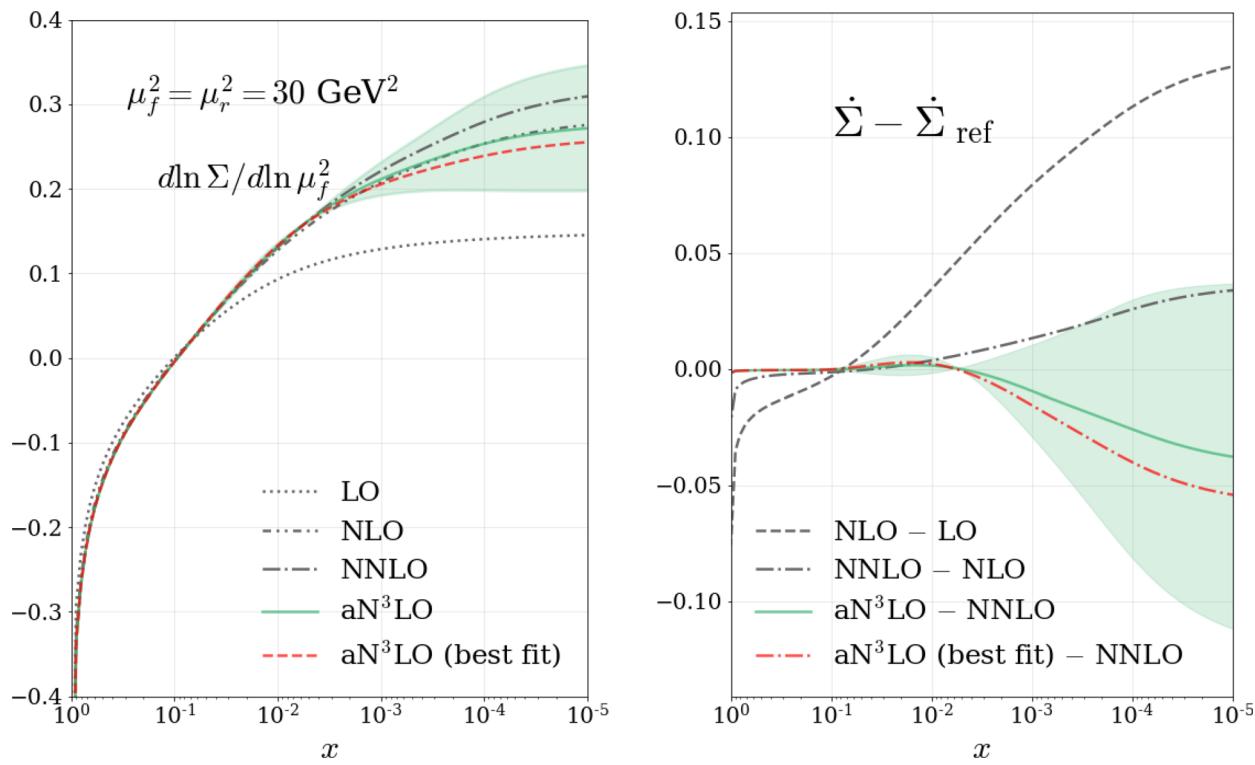
**Fig. 5** The low-integer numerical Mellin moments of relevant singlet splitting functions (excluding  $P_{q\bar{q}}^{\text{NS}, +}$ ) as a ratio between orders. In all cases the expected perturbative convergence is demonstrated

trend at small- $x$  is a reduction in the value of the evolution equation due to the N<sup>3</sup>LO prediction for  $P_{gg}$ . On the right hand side of Fig. 7 we observe the respective shifts from lower orders and how this shift changes up to N<sup>3</sup>LO.

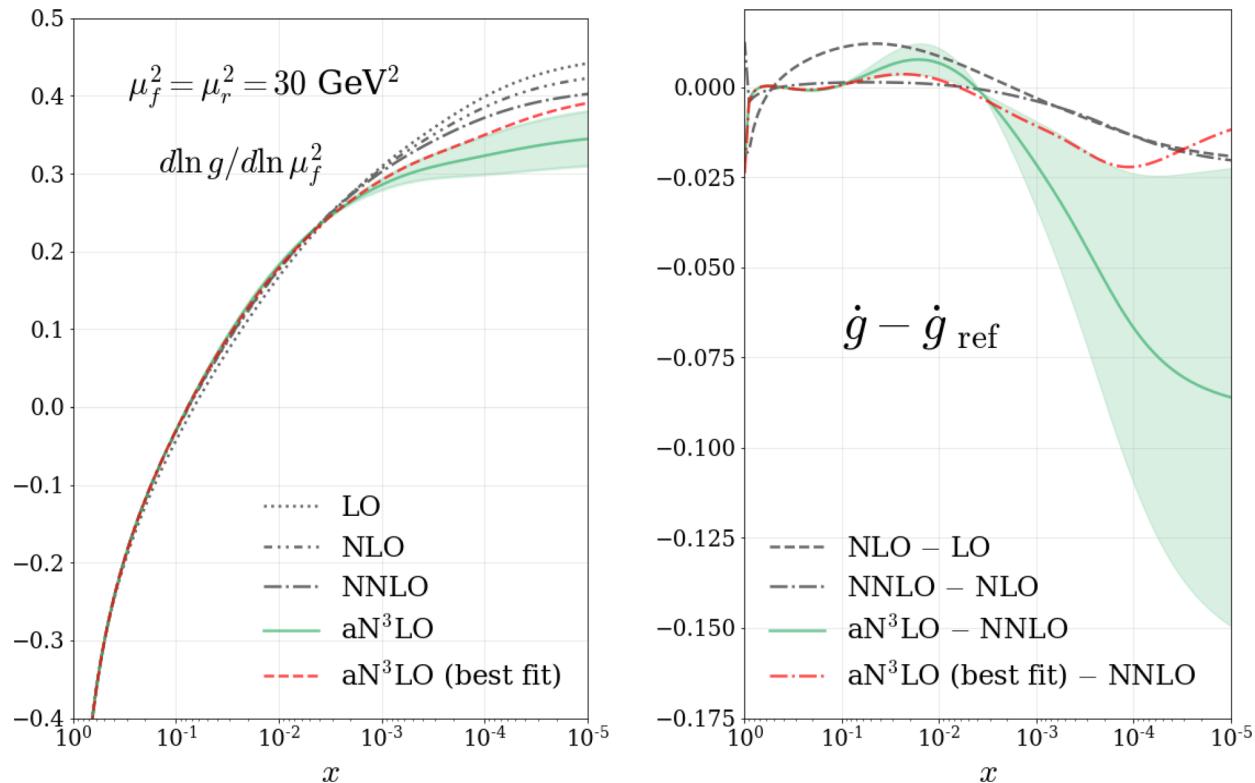
In the gluon evolution, there is a large variation coming from the uncertainty in the  $P_{gg}^{(3)}$  function. Therefore when  $P_{gg}^{(3)}$  is convoluted with the gluon PDF at small- $x$ , one could expect a potentially large shift from NNLO. The best fit gluon evolution prediction in Fig. 7 is produced by utilising the best fit results for  $P_{gq}^{(3)}$  and  $P_{gg}^{(3)}$  functions (red dashed). In this prediction we see that the fit prefers a reduction in the evolution from NNLO, which is contained within the  $\pm 1\sigma$  band until around  $x \lesssim 10^{-4}$ . Since at low- $Q^2$ , the quark and gluon are comparable at small- $x$ , this reduction is likely driven from the form of  $P_{gq}$  in Fig. 4. Combining this with the smaller gluon PDF at low- $Q^2$  therefore acts to slow the gluon evolution despite  $P_{gg}$  increasing. Furthermore, the best

fit is seemingly more in line with the perturbative expectation of the evolution than the chosen variation.<sup>10</sup> Since this variation is chosen from the known information about the perturbative expansions, this is a manifestation of how the framework we present here can capture the relevant sources of theoretical uncertainty (and account for these via a penalty in a PDF fit). This is encouraging, as even with the large amount of freedom for this gluon evolution, it seems that the data is constraining and balancing the two contributions from the splitting functions in a sensible fashion. As discussed in the singlet evolution case, the relative shift from NNLO to N<sup>3</sup>LO is slightly larger than one might hope for when dealing with a perturbative expansion. However, since this best fit is impacted to all orders from the experimental data (up to the

<sup>10</sup> Due to the presence of more divergent higher order logarithms at this level, it is not certain or by any means guaranteed that the shift at N<sup>3</sup>LO will follow the same trend outlined from lower orders.



**Fig. 6** The flavour singlet quark distribution evolution equation Eq. (3.3) shown for orders up to the approximate  $N^3LO$  (left). The relative shift between subsequent orders of the flavour singlet evolution (right) where  $\dot{\Sigma} = d \ln \Sigma / d \ln \mu_f^2$



**Fig. 7** The gluon distribution evolution equation Eq. (3.3) shown for orders up to the approximate  $N^3LO$  (left). The relative shift between subsequent orders of the gluon evolution (right) where  $\dot{g} = d \ln g / d \ln \mu_f^2$

leading logarithms at  $N^3\text{LO}$  i.e. even higher orders involve more divergent logarithms which are missed in this theoretical description), we can interpret this shift as an approximate all order shift and once again restore its validity in perturbation theory.

As with the singlet case above, negligible points of non-zero uncertainty are displayed in Fig. 7. For the reasons discussed in the singlet case and in Sect. 4.2, these are not an area of concern at the current level of desired uncertainty and are therefore not considered further.

## 5 $N^3\text{LO}$ transition matrix elements

Heavy flavour transition matrix elements,  $A_{ij}$ , as described in Sect. 3, are exact quantities that describe the transition of all PDFs with  $n_f$  active flavours into a scheme with  $n_f + 1$  active flavours. Due to discontinuous nature of  $A_{ij}$  at the heavy flavour mass thresholds, they are also present in the coefficient functions to ensure an exact cancellation of this discontinuity in physical quantities. This combination then preserves the smooth nature of the structure function, as demanded by the renormalisation group flows.

The general expansion of the heavy-quark transition matrix elements in powers of  $\alpha_s$  reads,

$$A_{ij} = \delta_{ij} + \sum_{\ell=1}^{\infty} \alpha_s^\ell A_{ij}^{(\ell)} = \delta_{ij} + \sum_{\ell=1}^{\infty} \alpha_s^\ell \sum_{k=0}^{\ell} L_\mu^k a_{ij}^{(\ell,k)}, \quad (5.1)$$

where at each order the terms proportional to powers of  $L_\mu = \ln(m_h^2/\mu^2)$  are determined by lower order transition matrix elements and splitting functions. Therefore the focus only needs to be on the  $a_{ij}^{(\ell,0)}$  expressions, as the rest are not only known [38, 39], but are guaranteed not to contribute at mass thresholds due to the presence of  $L_\mu$ . These  $\mu$ -independent terms can be decomposed in powers of  $n_f$  as

$$a_{ij}^{(3,0)} = a_{ij}^{(3,0), 0} + n_f a_{ij}^{(3,0), 1}, \quad (5.2)$$

where a number of the  $n_f$ -dependent and independent terms are known exactly. The  $n_f$  parts are however sub-leading and so as a first approximation, are set to zero in this work. In keeping with the framework set out in Sect. 4.1 for the  $N^3\text{LO}$  splitting functions, we will make use of the available known information (even-integer Mellin moments [50] and leading small and large- $x$  behaviour [49, 51–55]) about the heavy flavour transition matrix elements to approximate the  $\mu$ -independent contributions  $a_{ij}^{(3,0)}$ . As discussed above, we make the choice to completely ignore any terms that do not contribute at mass threshold since not only are these sub-leading but can also be ignored by explicitly setting  $\mu^2 = m_h^2$ .

### 5.1 3-Loop approximations

$$A_{Hg}$$

The  $A_{Hg}^{(3)}$  function is still under calculation at the time of writing. Currently the first five even-integer moments are known for the  $\overline{\text{MS}}$  scheme  $A_{Hg}^{(3)}$  [50], along with the leading small- $x$  terms [49].

The  $n_f$ -dependent contribution to the 3-loop unrenormalised  $A_{Hg}$  transition matrix element has also been approximated in [49], while all other contributions to  $A_{Hg}^{(3)}(n_f = 0)$  were already known. For this approximation we work in the  $\overline{\text{MS}}$  scheme using the framework set out in Sect. 4.1. We then approximate the function using the set of functions,

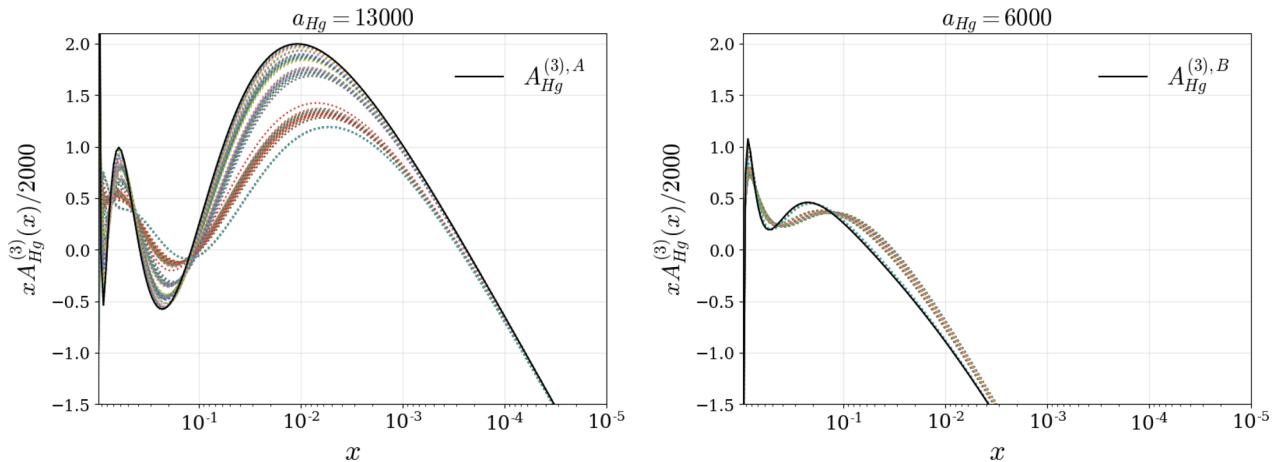
$$\begin{aligned} f_{1,2}(x) &= \ln^5(1-x) \quad \text{or} \quad \ln^4(1-x) \\ &\quad \text{or} \quad \ln^3(1-x) \\ &\quad \text{or} \quad \ln^2(1-x) \\ &\quad \text{or} \quad \ln(1-x), \\ f_{3,4}(x) &= 2-x \quad \text{or} \quad 1 \quad \text{or} \quad x \quad \text{or} \quad x^2, \\ f_5(x) &= \ln x \quad \text{or} \quad \ln^2 x, \\ f_e(x, a_{Hg}) &= \left( 224 \zeta_3 - \frac{41984}{27} - 160 \frac{\pi^2}{6} \right) \\ &\quad \times \frac{\ln 1/x}{x} + a_{Hg} \frac{1}{x} \end{aligned} \quad (5.3)$$

where  $a_{Hg}$  is varied as  $6000 < a_{Hg} < 13000$ . This variation is chosen from the criteria outlined in Sect. 4.1 and is comparable to that chosen in [49].

Figure 8 displays the approximation of the  $\overline{\text{MS}}$   $A_{Hg}^{(3)}$  with the variation from different combinations of functions in Eq. (5.3) at the chosen limits of  $a_{Hg}$ . Comparing with Fig. 3 in [49], we see a slightly larger range of allowed variation. A small proportion of this difference can be accounted for by the difference in renormalisation schemes, with the majority of this change being from the differences in the criteria from Sect. 4.1. The upper ( $A_{Hg}^{(3),A}$ ) and lower ( $A_{Hg}^{(3),B}$ ) bounds in the small- $x$  region (shown in Fig. 8) are given by,

$$\begin{aligned} A_{Hg}^{(3),A} &= 44.1703 \ln^5(1-x) + 268.024 \ln^4(1-x) \\ &\quad + 45271.0 x - 68401.4 x^2 \\ &\quad + 36029.8 \ln x + \left( 224 \zeta_3 - \frac{41984}{27} \right. \\ &\quad \left. - 160 \frac{\pi^2}{6} \right) \frac{\ln 1/x}{x} + 12000 \frac{1}{x} \end{aligned} \quad (5.4)$$

$$\begin{aligned} A_{Hg}^{(3),B} &= -18.9493 \ln^5(1-x) - 138.763 \ln^4(1-x) \\ &\quad - 31692.1 x + 33282.3 x^2 \\ &\quad - 3088.75 \ln^2 x + \left( 224 \zeta_3 - \frac{41984}{27} \right) \end{aligned}$$



**Fig. 8** Combinations of functions with an added variational factor ( $a_{Hg}$ ) controlling the NLL term. Combinations of functions at the upper (left) and lower (right) bounds of the variation are shown. The solid lines indicate the upper and

lower bounds for this function chosen from the relevant criteria

$$-160 \frac{\pi^2}{6} \left( \frac{\ln 1/x}{x} \right) + 6000 \frac{1}{x}. \quad (5.5)$$

Using this information, we then choose the fixed functional form,

$$\begin{aligned} A_{Hg}^{(3)} = & A_1 \ln^5(1-x) + A_2 \ln^4(1-x) \\ & + A_3 x + A_4 x^2 + A_5 \ln x \\ & + \left( 224 \zeta_3 - \frac{41984}{27} - 160 \frac{\pi^2}{6} \right) \frac{\ln 1/x}{x} + a_{Hg} \frac{1}{x} \end{aligned} \quad (5.6)$$

where the variation of  $a_{Hg}$  remains unchanged as it already encapsulates the predicted variation to within the  $\sim 1\%$  level.

$$A_{Hq}^{\text{PS}}$$

The  $A_{Hq}^{\text{PS}}$  transition matrix element has been calculated exactly in [53]. Here we attempt to qualitatively reproduce this result via an efficient parameterisation to an appropriate precision.

Using the expressions for the small and large- $x$  limits [53] and the known first six even-integer moments converted into  $\overline{\text{MS}}$  [50], we provide a user-friendly approximation as,

$$\begin{aligned} A_{Hq}^{\text{PS},(3)} = & (1-x)^2 \left\{ -152.523 \ln^3(1-x) \right. \\ & \left. - 107.241 \ln^2(1-x) \right\} - 4986.09 x \\ & + 582.421 x^2 - 1393.50 x \ln^2 x - 4609.79 x \ln x \\ & - 688.396 \frac{\ln 1/x}{x} + (1-x) 3812.90 \frac{1}{x} \\ & + 1.6 \ln^5 x - 20.3457 \ln^4 x \end{aligned}$$

$$\begin{aligned} & + 165.115 \ln^3 x - 604.636 \ln^2 x + 3525.00 \ln x \\ & + (1-x) \left\{ 0.246914 \ln^4(1-x) - 4.44444 \ln^3(1-x) \right. \\ & \left. - 2.28231 \ln^2(1-x) - 357.427 \ln(1-x) + 116.478 \right\} \end{aligned} \quad (5.7)$$

where the first two lines have been approximated and the last four lines are the exact leading small and large- $x$  terms. We note here that the approximated part of this parameterisation is in a much less important region of  $x$  than the exact parts, therefore any small differences in the approximated part from the exact function are unimportant.

$$A_{qq,H}^{\text{NS}}$$

Moving to the non-singlet  $A_{qq,H}^{\text{NS}}$  function, we attempt to parameterise the work from [51, 52]. Specifically, we make use of the known even integer moments up to  $N = 14$  [50], converted into the  $\overline{\text{MS}}$  scheme, with the even moments corresponding to the (+) non-singlet distribution.

As for  $A_{Hg}^{(3)}$ , the approximation is performed using the set of functions,

$$\begin{aligned} f_1(x) &= \ln x, \quad f_2(x) = \ln^2 x, \\ f_{3,4}(x) &= 1 \text{ or } x \text{ or } x^2 \text{ or } \ln(1-x), \\ f_5(x) &= 1/x, \quad f_6(x) = \ln^3(1-x), \\ f_7(x) &= \ln^2(1-x), \\ f_e(x, a_{qq,H}^{\text{NS}}) &= a_{qq,H}^{\text{NS}} \ln^3 x \end{aligned} \quad (5.8)$$

where  $a_{qg,H}^{\text{NS}}$  is varied as  $-90 < a_{qg,H}^{\text{NS}} < -37$ . To contain this variation in a fixed functional form we employ:

$$\begin{aligned} A_{qg,H}^{\text{NS}, (3)} &= A_1 \frac{1}{(1-x)_+} + A_2 \ln^3(1-x) \\ &+ A_3 \ln^2(1-x) + A_4 \ln(1-x) + A_5 \\ &+ A_6 x + A_7 \ln^2 x + a_{qg,H}^{\text{NS}} \ln^3 x \end{aligned} \quad (5.9)$$

where the variation of  $a_{qg,H}^{\text{NS}}$  is unchanged.

$$A_{gg,H}$$

The 3-loop  $A_{qg,H}$  function has been calculated exactly in [54]. As with the  $A_{Hg}^{\text{PS}}$  function above, we attempt to provide a simple and computationally efficient approximation to this exact form. To do this, we use the known even-integer moments (converted to the  $\overline{\text{MS}}$  scheme) and small and large- $x$  information from [50,54]. Gathering a fixed set of functions  $f_i(x)$  and omitting any variational parameter  $a_{qg,H}$ , due to the higher amount of information available, the resulting approximation to the  $\overline{\text{MS}}$   $A_{qg,H}^{(3)}$  is:

$$\begin{aligned} A_{qg,H}^{(3)} &= -237.172 \ln^3(1-x) - 201.497 \ln^2(1-x) \\ &+ 7247.70 \ln(1-x) + 39967.3 x^2 \\ &- 22017.7 - 28459.1 \ln x - 14511.5 \ln^2 x \\ &+ 341.543 \frac{\ln 1/x}{x} + 1814.73 \frac{1}{x} \\ &- \frac{580}{243} \ln^4(1-x) - \frac{17624}{729} \ln^3(1-x) \\ &- 135.699 \ln^2(1-x) \end{aligned} \quad (5.10)$$

where the first two lines have been approximated and the last two lines are the exact small and large- $x$  limits.

$$A_{gg,H}$$

Work is ongoing for the 3-loop contribution to  $A_{gg,H}$  [60,61]. Due to this, the entire approximation of  $A_{gg,H}^{(3)}$  presented here is based on the first 5 even-integer Mellin moments [50]. To reduce the wild behaviour of this approximation from only using the Mellin moment information (converted into the  $\overline{\text{MS}}$  scheme), we introduce a second mild constraint in the form of the relations in Eq. (4.3). These relations are closely followed by the gluon-gluon functions up to NNLO, but there is no guarantee that this behaviour will continue at  $\text{N}^3\text{LO}$ . This constraint is given as,

$$A_{gg,H}(x \rightarrow 0) \simeq \frac{C_A}{C_F} A_{qg,H}(x \rightarrow 0). \quad (5.11)$$

It can be expected that even though this relation may not be followed exactly, it should not stray too far from this general ‘rule of thumb’. Due to this a generous contingency of  $\pm 50\%$  is allowed when using this rule. Furthermore, to ensure this

relation is only used as a guide, we allow the variation to move beyond this rule as long as the criteria in Sect. 4.1 are still satisfied. As a result of this change in prescription and because the allowed variation is now on a much larger scale than that of any functional uncertainty, we choose a fixed functional form from the start and use the criteria described above to guide our choice of variation.

$$\begin{aligned} A_{gg,H}^{(3)} &= A_1 \ln^2(1-x) + A_2 \ln(1-x) \\ &+ A_3 x^2 + A_4 \ln x + A_5 x + a_{gg,H} \frac{\ln x}{x} \end{aligned} \quad (5.12)$$

where  $-2000 < a_{gg,H} < -700$ .

## 5.2 Predicted $\text{aN}^3\text{LO}$ transition matrix elements

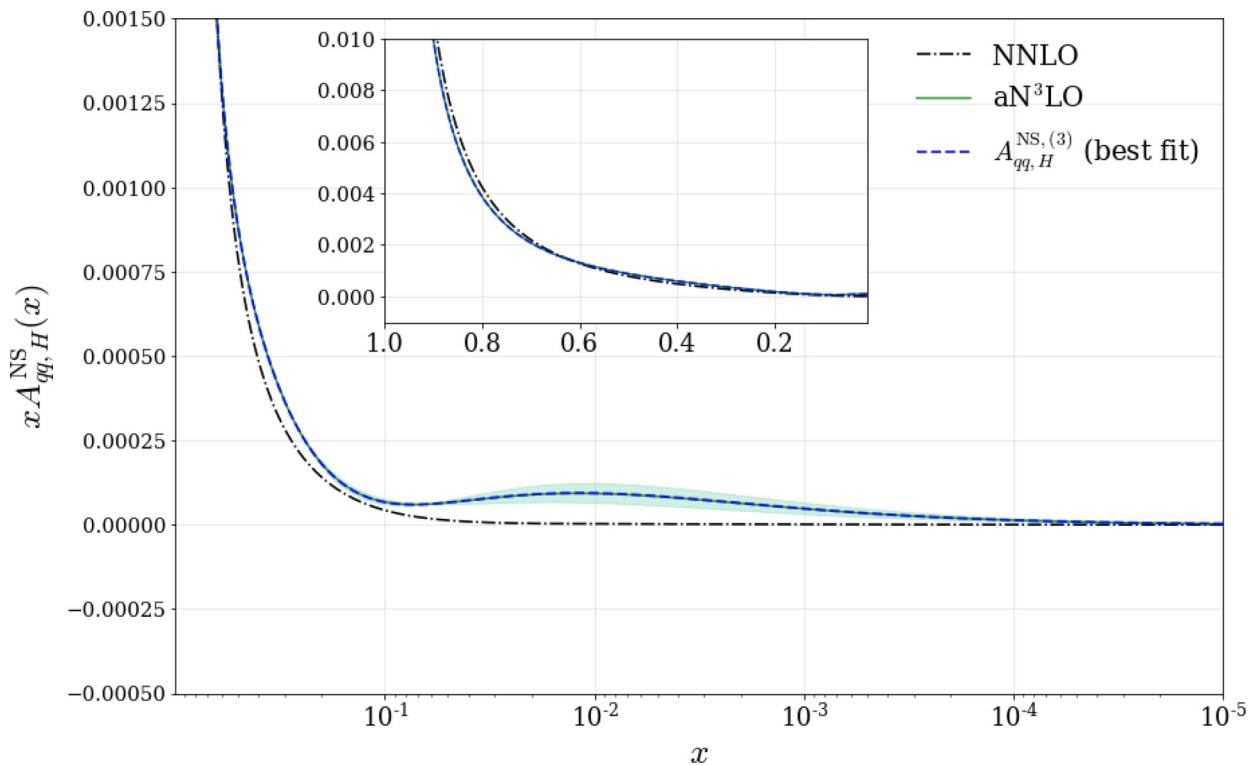
Figures 9, 10 and 11 show the perturbative expansions for each of the  $n_f$ -independent contributions to the transition matrix elements at the mass threshold value of  $\mu = m_h$ . Included with these expansions are the predicted variations ( $\pm 1\sigma$ ) from Sect. 5.1 (shown in green) and the approximate  $\text{N}^3\text{LO}$  best fits (shown in blue - discussed further in Sect. 8).

$A_{qg,H}^{\text{NS}}$  in Fig. 9 behaves as expected with little variation from NNLO until the magnitude of this function is very small. The approximations for the more dominant  $A_{Hg}^{\text{PS}}$  and  $A_{Hg}$  functions in Fig. 10 exhibit some slight sporadic behaviour towards large- $x$  due to the increased logarithmic influence. However, since this is in a region where the magnitude of these functions become small, any instabilities will have a minimal effect on the overall result. The major feature prevalent across both these functions is the large deviation away from the NNLO behaviour, especially at small- $x$  (and also mid- $x$  for  $A_{Hg}$ ).

Similarly for  $A_{gg,H}$  in Fig. 11 (upper), we see some irregular behaviour towards large- $x$ . As with  $A_{Hg}^{\text{PS}}$  and  $A_{Hg}$ , this behaviour is in a region where the magnitude of  $A_{gg,H}$  is small. As discussed in Sect. 5.1,  $A_{gg,H}^{(3)}$  is approximated without any variation due to the range of available information being large.<sup>11</sup> Due to this, and the fact that the region of potential instability (large- $x$ ) is highly suppressed, we can accept this function with negligible effect on any results. As more information becomes available about all these functions, it will be interesting to observe how the behaviour across  $x$  changes.

The  $A_{gg,H}$  function shown in Fig. 11 (lower) displays the  $\pm 50\%$  bounds of violation we allow for the relation Eq. (4.3). It follows that the allowed variation is conservative enough to include a generous violation of Eq. (4.3) at  $\text{N}^3\text{LO}$ , with the prediction that the function is positive at small- $x$ . This is an

<sup>11</sup> Although an exact expression has been calculated for  $A_{gg,H}^{(3)}$  [54], this function is not yet available in a computationally efficient format i.e. numerical grids.



**Fig. 9** Perturbative expansions for the transition matrix element  $A_{qg,H}^{NS}$  including any corresponding allowed  $\pm 1\sigma$  variation (shaded green region). This function is shown at the mass threshold value of

$\mu = m_h$ . The best fit value (blue dashed line) displays the prediction for this function determined from a global PDF fit

area where small- $x$  information would clearly be very beneficial. With this information currently in progress, it will be very interesting to compare how well this variation captures the true small- $x$   $A_{gg,H}$  behaviour.

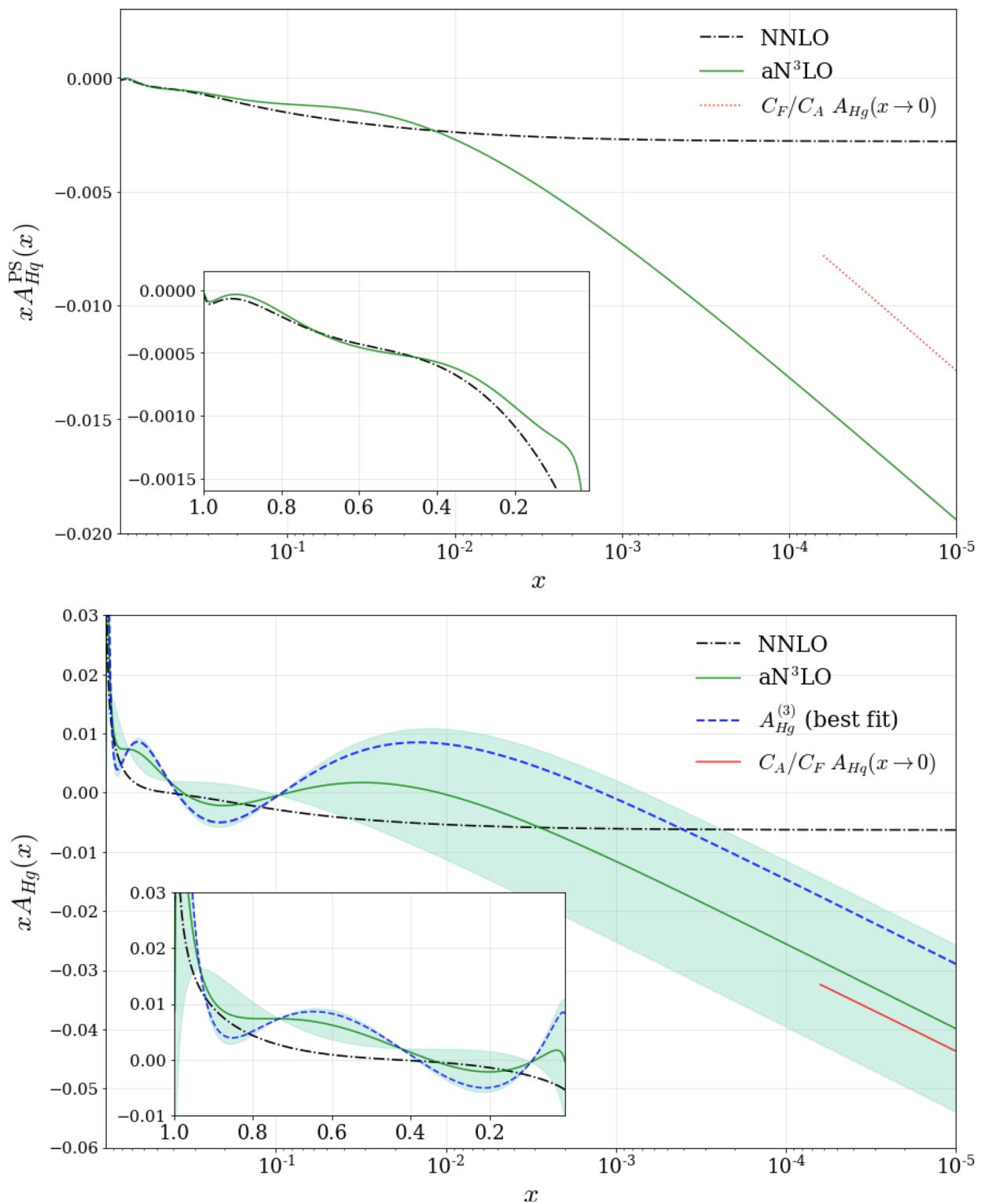
The final best fit values shown in Figs. 9, 10 and 11 are determined from a global PDF fit with various datasets seen to be constraining these functions within the  $\pm 1\sigma$  variations. As observed, we are able to show good agreement between the allowed variations and the best fit predictions. The perturbative expansion predicted for  $A_{gg,H}$  is the least well constrained while also violating its expected relation with  $A_{qg,H}$  more than one may originally expect. Since the small- $x$  region in all cases changes dramatically at  $N^3LO$ , one potential explanation is that this function is compensating for an inaccuracy in another area of the theory. However, when comparing with the relationship between  $A_{Hg}$  and  $A_{Hg}^{PS}$ , Eq. (4.3) also exhibits a significant violation at this order. This could suggest that for the  $N^3LO$  transition matrix elements, this relation may not be the best indicator of precision or consistency. Finally, we remember that the best fit in this case may be feeling a larger effect from higher orders, especially due to these functions only existing from NNLO. For example, in Sect. 4.3 we observed a high level of divergence introduced at 4-loops in the splitting functions.

The best fit results shown here may therefore be sensitive to a similar level of divergence further along in their corresponding perturbative expansions.

As previously discussed, this lack of knowledge is contained within our choice of the predicted variations of these functions. Therefore this treatment only seeks to add to the predicted level of theoretical uncertainty from missing  $N^3LO$  contributions, as one expects.

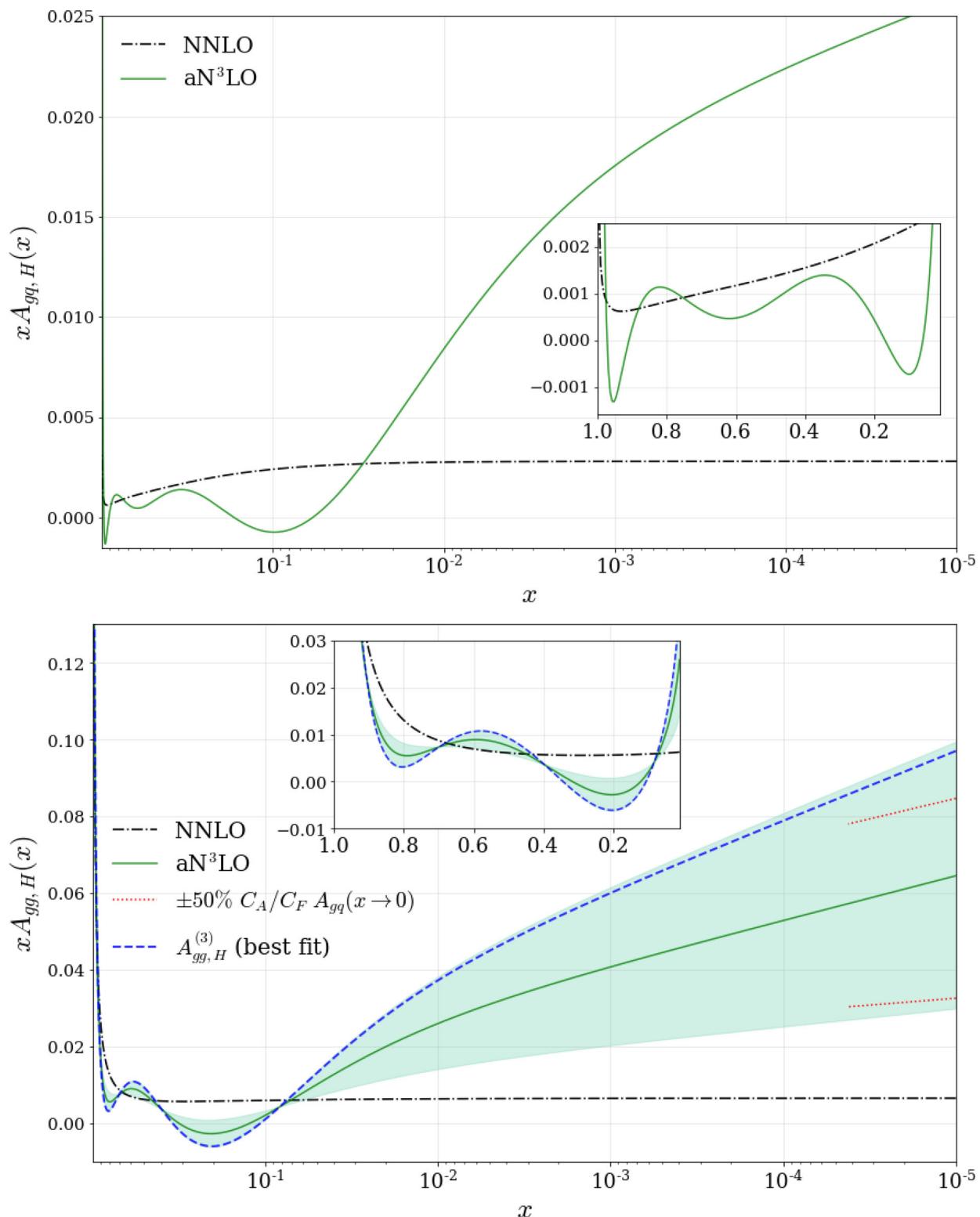
### 5.3 Numerical results

For these results, the same toy PDFs presented in Sect. 4.4 are employed which approximate the general order-independent PDF features at  $Q^2 \simeq 30 \text{ GeV}^2$ . Note that due to the higher  $Q^2$ , these results are more representative of the b-quark. The left plot in Fig. 12 shows the result of including the  $N^3LO$  transition matrix element approximations we have determined into Eq. (3.7c), which is describing the heavy quark distribution  $(H + \bar{H})(x, Q^2 = m_h^2)$ . The right plot in Fig. 12 is describing the heavy flavour contribution to the gluon at  $(x, Q^2 = m_h^2)$  in Eq. (3.7b) where the delta function describing the leading order contribution to  $A_{gg,H}$  has been subtracted out. The dominant contribution to the heavy quark (left plot) is stemming from the  $A_{Hg}$  function.



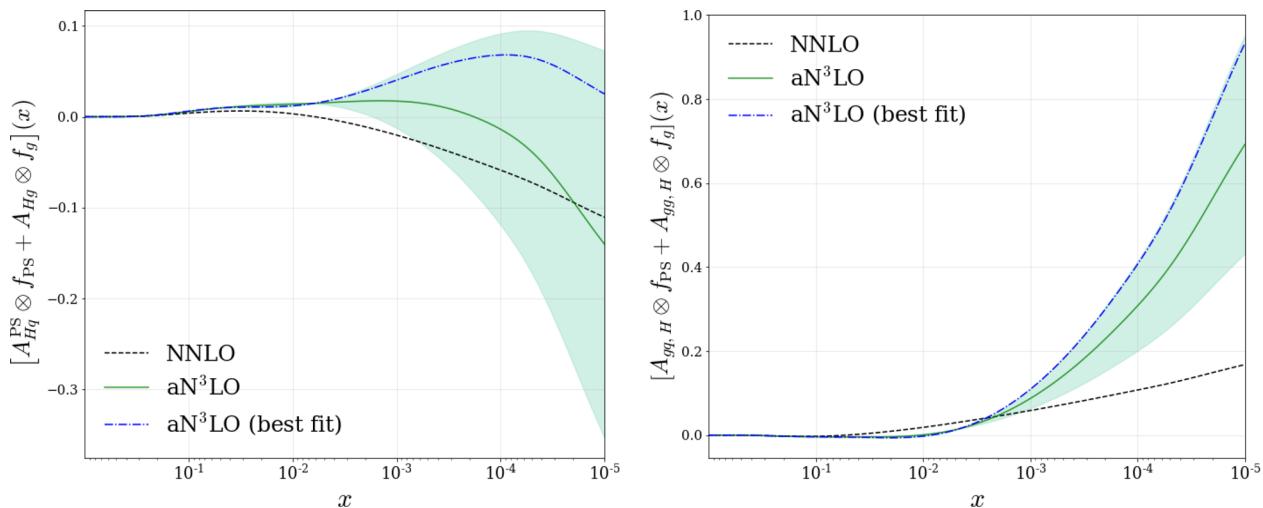
**Fig. 10** Perturbative expansions for the transition matrix elements  $A_{Hq}^{\text{PS}}$  and  $A_{Hg}$  including any corresponding allowed  $\pm 1\sigma$  variation (shaded green region). These functions are shown at the mass threshold

value of  $\mu = m_h$ . The best fit values (blue dashed line) display the predictions for these functions determined from a global PDF fit



**Fig. 11** Perturbative expansions for the transition matrix elements  $A_{gq,H}$  and  $A_{gg,H}$  including any corresponding allowed  $\pm 1\sigma$  variation (shaded green region). These functions are shown at the mass threshold

value of  $\mu = m_h$ . The best fit values (blue dashed line) display the predictions for these functions determined from a global PDF fit



**Fig. 12** Heavy flavour evolution contributions to the heavy quark ( $H + \bar{H}$ ) (left) and gluon (right) PDFs provided at  $\mu \simeq 30$  GeV $^2$ . These results include the  $\mu = m_h$  contributions from  $A_{Hq}^{\text{PS}}$ ,  $A_{Hg}$ ,  $A_{gq,H}$  and  $A_{gg,H}$  transition matrix elements up to aN $^3$ LO

Whereas the dominant contribution to the gluon (right plot) is from the  $A_{gg,H}$  function. As one might expect, the predictions at N $^3$ LO are more divergent at small- $x$ , however it is also true that the general trend from NNLO is being followed across most values of  $x$ .

The best fit functions predicted from a global fit show the preferred aN $^3$ LO contributions for both scenarios. The predicted behaviour from the global fit follows the results for the perturbative expansions in Sect. 5.2. For the ( $H + \bar{H}$ )( $x, Q^2 = m_h^2$ ) result (Fig. 12 left), the aN $^3$ LO result is positive across a much wider range of  $x$ . Since this is a perturbatively calculated PDF, this is an encouraging result that could potentially eliminate some of the more unphysical shortcomings at NNLO without demanding positivity of the PDF a priori.

## 6 N $^3$ LO heavy coefficient functions

The final set of functions considered are the Neutral Current (NC) DIS coefficient functions which, when combined with the PDFs, form the structure functions discussed in Sect. 3.<sup>12</sup>

<sup>12</sup> Charged current (CC) structure function data is limited to relatively high- $x$  values compared to NC data and is either comparatively low statistics, high- $Q^2$  proton target data from HERA or nuclear target data (again often quite low statistics) on heavy nuclear targets. In both cases the effect of N $^3$ LO corrections is small compared with uncertainties, especially when considering those involved with nuclear corrections. Also, heavy flavour contributions are less well known at high orders for CC structure functions. Hence, we do not include N $^3$ LO for these processes, except dimuon data, which is particularly important for the poorly constrained strange quark, but which is a semi-inclusive DIS process, and for which we parameterise N $^3$ LO corrections, as discussed in Sect. 7. An improvement would be necessary for more precise proton data, from the EIC for example.

We approximate the N $^3$ LO heavy quark coefficient functions which accompany the heavy flavour transition matrix elements from Sect. 5 and also the N $^3$ LO light quark coefficient functions. We note that our standard definition of the order of coefficient functions includes the longitudinal coefficient functions at order  $\alpha_s$  at LO, at order  $\alpha_s^2$  at NLO etc. This means we already include order  $\alpha_s^3$  coefficient functions for the longitudinal coefficient functions at NNLO, whereas many groups only consider order  $\alpha_s^2$  at NNLO. Since little is known about longitudinal coefficient functions at order  $\alpha_s^4$ , and the data constraints from  $F_L(x, Q^2)$  are very much less precise than from  $F_2(x, Q^2)$ , we simply remain at the precisely known order  $\alpha_s^3$  in this study.

### 6.1 Approximation framework: continuous information

In Sect. 4.1 we described the approximation framework employed for functions with discrete Mellin moment information, combined with any available exact information. For the N $^3$ LO coefficient function approximations, we have access to a somewhat richer vein of information than the discrete moments discussed for the framework used in approximating the N $^3$ LO splitting functions and transition matrix elements in Sects. 4 and 5. More specifically, approximations of the FFNS coefficient functions at  $\mathcal{O}(\alpha_s^3)$  are known for the heavy quark contributions to the heavy flavour structure function  $F_{2,H}(x, Q^2)$  at  $Q^2 < m_{c,b}^2$  [47–49]. These approximations include the exact LL and mass threshold contributions, with an approximated NLL term (the details of this are described in Sect. 6.2). Furthermore, the N $^3$ LO ZM-VFNS coefficient functions are known exactly [57]. Both of these contributions can then be combined with the transition matrix element approximations to define the GM-VFNS functions

in the  $Q^2 \leq m_c^2, m_b^2$  and  $Q^2 \rightarrow \infty$  regimes. Due to this, we base our approximations for the  $C_{H,\{q,g\}}^{(3)}$  functions on the known continuous information in the low and high- $Q^2$  regimes.

To achieve a reliable approximation for  $C_{H,\{q,g\}}^{(3)}$ , we first fit a regression model with a large number of functions in  $(x, Q^2)$  space made available to the model (in order to reduce the level of functional bias in the parameterisation). This produces an unstable result at the extremes of the parameterisation (large- $x$  and low- $Q^2$ ). However, it provides a basis for manually choosing a stable parameterisation to move between the two known regimes (low- $Q^2$  and high- $Q^2$ ).

Using the regression model predictions as a qualitative guide, we choose a stable and smooth interpolation between the two  $Q^2$  regimes (low- $Q^2$  and high- $Q^2$ ) as given in Eq. (6.1). This interpolation is observed to mirror the expected behaviour observed from lower orders, the regression model qualitative prediction having been calculated independently of lower orders and the best fit quality to data. By definition, we also ensure an exact cancellation between the coefficient functions and the transition matrix elements at the mass threshold energies as demanded by the theoretical description in Sect. 3.

For the contributions to the heavy flavour structure function  $F_{2,H}$  the final interpolations in the FFNS regime are defined as,

$$C_{H,\{q,g\}}^{(3)} = \begin{cases} C_{H,\{q,g\}, \text{low-}Q^2}^{(3)}(x, Q^2 = m_h^2) e^{0.3(1-Q^2/m_h^2)} \\ + C_{H,\{q,g\}}^{(3)}(x, Q^2 \rightarrow \infty)(1 - e^{0.3(1-Q^2/m_h^2)}), & \text{if } Q^2 \geq m_h^2, \\ C_{H,\{q,g\}, \text{low-}Q^2}^{(3)}(x, Q^2), & \text{if } Q^2 < m_h^2. \end{cases} \quad (6.1)$$

where  $C_{H,\{q,g\}, \text{low-}Q^2}^{(3)}$  are the already calculated approximate heavy flavour FFNS coefficient functions at  $Q^2 \leq m_h^2$ , and  $C_{H,\{q,g\}}^{(3)}(Q^2 \rightarrow \infty)$  is the limit at high- $Q^2$  found from the known ZM-VFNS coefficient functions and relevant subtraction terms, themselves found from Eq. (3.14). Both of these limits will be discussed in detail on a case-by-case basis in Sect. 6.

For the heavy flavour contributions to  $F_{2,q}$ , we have no information about the low- $Q^2$   $N^3\text{LO}$  FFNS coefficient functions. In this case, we use intuition from lower orders to provide a soft (lightly weighted) low- $Q^2$  target for our regression model in  $(x, Q^2)$ . However, since the overall contribution is very small from these functions, the exact form of these functions is not phenomenologically important at present. Further to this, our understanding from lower orders is that these functions have a weak dependence on  $Q^2$  and so the form of the low- $Q^2$  description is even less important. As with the  $C_{H,\{q,g\}}^{(3)}$  coefficient functions, the regression results provide an initial qualitative guide which exhibits instabilities in the

extremes of  $(x, Q^2)$ . We therefore employ a similar technique as before to ensure a smooth extrapolation across all  $(x, Q^2)$  into the unknown behaviour at low- $Q^2$ . For these functions, the ansatz used is given as,

$$C_{q,\{q,g\}}^{\text{FF, (3)}} = \begin{cases} C_{q,q}^{\text{FF, NS, (3)}}(x, Q^2 \rightarrow \infty) \\ (1 + e^{-0.5(Q^2/m_h^2)-3.5}), \\ C_{q,q}^{\text{FF, PS, (3)}}(x, Q^2 \rightarrow \infty) \\ (1 - e^{-0.25(Q^2/m_h^2)-0.3}), \\ C_{q,g}^{\text{FF, (3)}}(x, Q^2 \rightarrow \infty)(1 - e^{-0.05(Q^2/m_h^2)+0.35}), \end{cases} \quad (6.2)$$

where  $C_{q,\{q,g\}}^{\text{FF, (3)}}(x, Q^2 \rightarrow \infty)$  is the known limit at high- $Q^2$ .

## 6.2 Low- $Q^2$ $N^3\text{LO}$ heavy flavour coefficient functions

As previously mentioned in Sect. 3, the standard MSHT theoretical description of NNLO structure functions includes approximations to the low- $Q^2$  FFNS coefficient functions  $C_{H,\{q,g\}}^{(3),\text{FF}}$  from [47–49]. Within these functions are the precisely known LL small- $x$  terms and mass threshold information, along with an approximate NLL small- $x$  term added into the MSHT fit. In the NNLO fit these approximate NLL parameters play a very small role due to not only being sub-leading, but also only affecting the FFNS scheme below the mass thresholds. At NNLO they are therefore heuristically set to a value that is theoretically justified and suits the NNLO best fit. At  $N^3\text{LO}$  these functions begin to directly affect the form of the full GM-VFNS scheme across all  $(x, Q^2)$ . For this reason, these NLL parameters need to be considered as an independent source of theoretical uncertainty. In the  $aN^3\text{LO}$  fit, the NLL parameters are left free and included into the framework set out in Sect. 2.1.

The standard NNLO MSHT fit contains terms of the form,

$$C_{H,i}^{(3), \text{NLL}}(Q^2 \rightarrow 0) \propto -4 \frac{1}{x} + c_i^{\text{LL}} \frac{\ln 1/x}{x}, \\ \left( c_g^{\text{LL}} = \frac{C_F}{C_A} c_q^{\text{LL}} \right), \quad (6.3)$$

where  $i = q, g$  and  $c_i^{\text{LL}}$  is the precisely known leading small- $x$   $\log$  coefficient. In the  $aN^3\text{LO}$  fit, the NLL coefficient is allowed to vary by  $\pm 50\%$  ( $\pm 1\sigma$  variation). This conservative range is chosen to enable the release of tension with the variational parameters associated with the  $N^3\text{LO}$  transition matrix elements. Here we stress that this quantity is heuristically set even at NNLO, therefore our treatment is completely justified with the added benefit of now accounting for an uncertainty for this choice.

### 6.3 3-Loop approximations

$$C_{H,q}$$

In this section the  $C_{H,q}$  coefficient function is investigated. As discussed in Sect. 3,  $C_{H,q}$  contributes to the heavy flavour structure function  $F_{2,H}$ . We begin by isolating this function from Eq. (3.14) and relating the FFNS and GM-VFNS schemes at all orders from Eq. (3.9) and Eq. (3.13),

$$\begin{aligned} C_{H,q}^{\text{FF}} = & \left[ C_{H,H}^{\text{VF, NS}} + C_{H,H}^{\text{VF, PS}} \right] \otimes A_{Hq}^{\text{PS}} \\ & + C_{H,q}^{\text{VF}} \otimes \left[ A_{qq,H}^{\text{NS}} + A_{qq,H}^{\text{PS}} \right] + C_{H,g}^{\text{VF}} \otimes A_{gq,H}. \end{aligned} \quad (6.4)$$

Expanding this function we obtain:

$$\mathcal{O}(\alpha_s) : C_{H,q}^{\text{FF, (1)}} = 0 \quad (6.5)$$

$$\begin{aligned} \mathcal{O}(\alpha_s^2) : C_{H,q}^{\text{FF, (2)}} = & C_{H,H}^{\text{VF, (0)}} \otimes A_{Hq}^{\text{PS, (2)}} \\ & + C_{H,q}^{\text{VF, (2)}} \otimes A_{qq,H}^{\text{NS, (0)}} \end{aligned} \quad (6.6)$$

$$\begin{aligned} \mathcal{O}(\alpha_s^3) : C_{H,q}^{\text{FF, (3)}} = & C_{H,H}^{\text{VF, (1)}} \otimes A_{Hq}^{\text{PS, (2)}} \\ & + C_{H,H}^{\text{VF, (0)}} \otimes A_{Hq}^{\text{PS, (3)}} \\ & + C_{H,q}^{\text{VF, (3)}} \otimes A_{qq,H}^{\text{NS, (0)}} \\ & + C_{H,g}^{\text{VF, (1)}} \otimes A_{gq,H}^{(2)} \end{aligned} \quad (6.7)$$

where we recall that  $A_{qq,H}^{\text{NS, (0)}} = \delta(1-x)$ .

### NNLO

The first contribution from the heavy quarks appears at the  $\mathcal{O}(\alpha_s^2)$  level. Fortunately there is a complete picture of this order [43] which provides some experience with the behaviour of these functions before moving into unknown territory. Figure 13 shows the case for  $C_{H,q}^{\text{VF, (2)}}$  converging onto  $C_{H,q}^{\text{ZM, (2)}}$  at high- $Q^2$ , as required by the definition of the GM-VFNS scheme outlined in Sect. 3.

From Fig. 13, immediately some intuition can be built up surrounding the form of these functions. It can be observed that the GM-VFNS function at low- $Q^2$  is consistently more positive than at high- $Q^2$ . However, the values at low and high- $Q^2$  are of the same order of magnitude which provides evidence that the behaviour should not be substantially different across values of  $Q^2$  when estimating our N<sup>3</sup>LO quantities. Further to this, as  $x \rightarrow 0$  the overall magnitude of  $C_{H,q}^{(2)}$  becomes much larger, which is consistent with an inherently pure singlet quantity.

### N<sup>3</sup>LO

At  $\mathcal{O}(\alpha_s^3)$  the N<sup>3</sup>LO ZM-VFNS and low- $Q^2$  FFNS functions are known [47–49, 57] and parameterisations/approximations are available (up to the level of precision discussed in Sect. 6.2). Nevertheless, there is no direct information on how the full GM-VFNS function behaves at this order which is required for a full treatment of the heavy flavour coefficients. Using Eq. (6.7) to estimate the N<sup>3</sup>LO contribution, we have

$$\begin{aligned} C_{H,q}^{\text{VF, (3)}} = & C_{H,q}^{\text{FF, (3)}} - C_{H,H}^{\text{VF, (1)}} \otimes A_{Hq}^{\text{PS, (2)}} \\ & - C_{H,g}^{\text{VF, (1)}} \otimes A_{gq,H}^{(2)} - A_{Hq}^{\text{PS, (3)}}, \end{aligned} \quad (6.8)$$

where  $A_{Hq}^{\text{PS, (3)}}$  is the N<sup>3</sup>LO transition matrix element approximated in Sect. 5.1.

It must be the case that the discontinuities introduced into the heavy flavour PDF from the transition matrix elements (at the threshold value of  $Q^2 = m_h^2$ ) are cancelled exactly in the structure function. The cancellation of  $A_{Hq}^{\text{PS, (3)}}$  is therefore guaranteed by its inclusion into the GM-VFNS coefficient function in Eq. (6.8). Since in practice the transition matrix elements are convoluted with the PDFs separately to the coefficient functions, to ensure that this statement remains the case, the parameterisation will be performed in the FFNS number scheme. By doing this, we can explicitly switch to the GM-VFNS number scheme by including the subtraction term in Eq. (6.8). This procedure then ensures that  $A_{Hq}^{\text{PS, (3)}}$  is subtracted off exactly with no unphysical discontinuity.

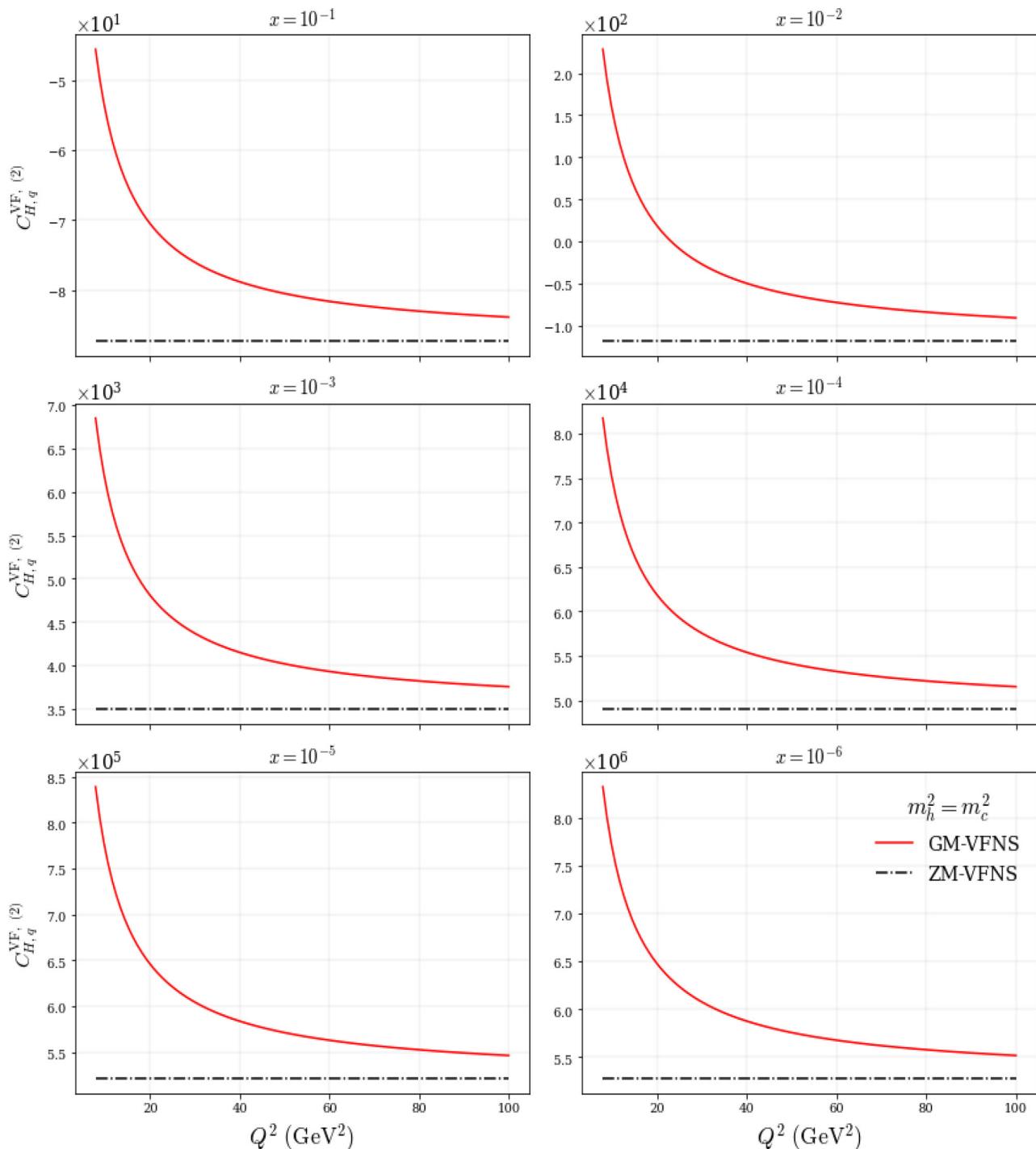
Following the methodology set out in Sect. 6.1, the two regimes we wish to interpolate between are the approximate  $C_{H,q}^{\text{FF, (3)}}(Q^2 \rightarrow 0)$  limit and

$$\begin{aligned} C_{H,q}^{\text{FF, (3)}}(Q^2 \rightarrow \infty) = & C_{H,q}^{\text{ZM, (3)}} + C_{H,H}^{\text{VF, (1)}} \otimes A_{Hq}^{\text{PS, (2)}} \\ & + C_{H,g}^{\text{VF, (1)}} \otimes A_{gq,H}^{(2)} + A_{Hq}^{\text{PS, (3)}}, \end{aligned} \quad (6.9)$$

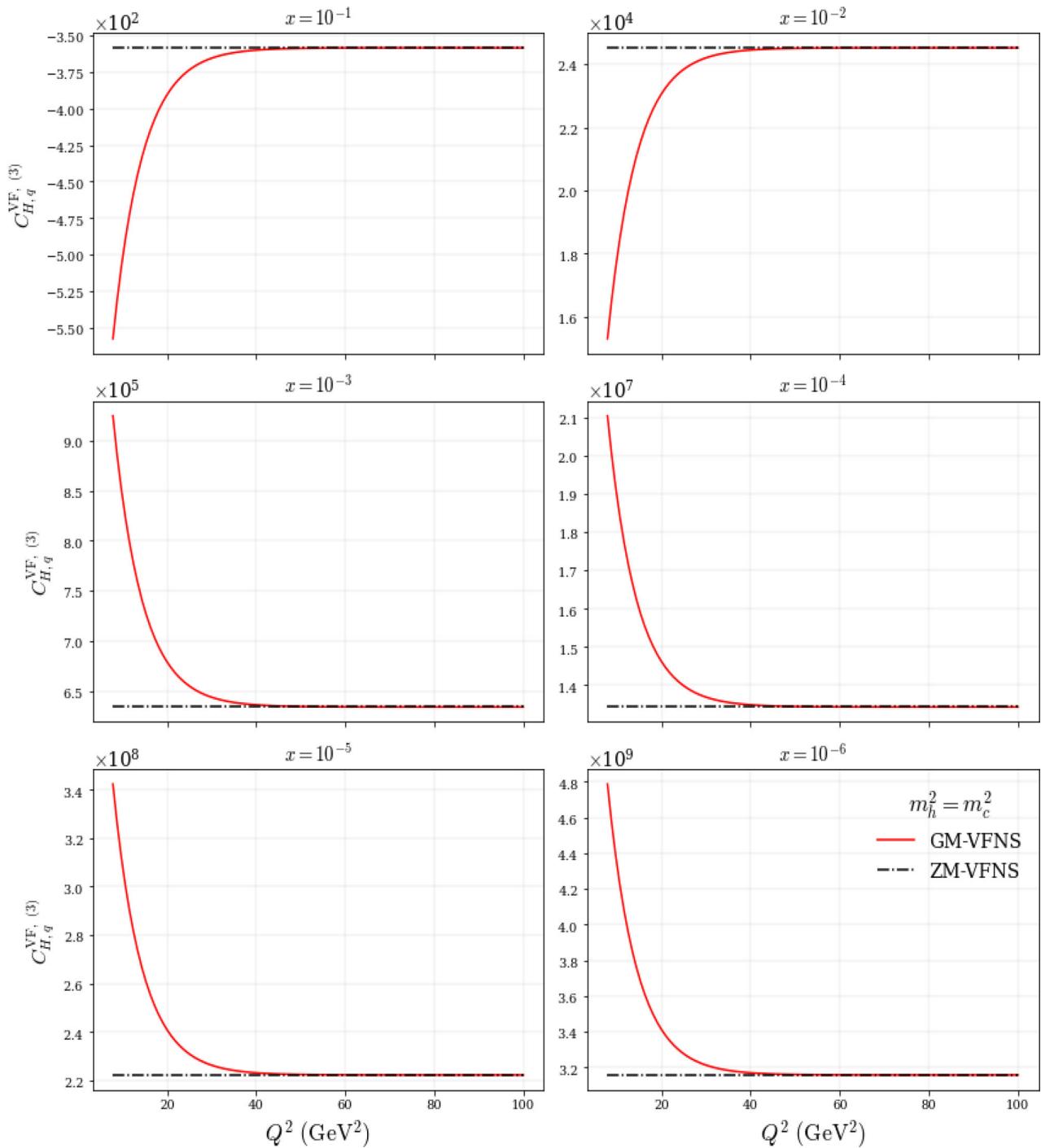
where  $C_{H,q}^{\text{VF, (3)}}$  is replaced with  $C_{H,q}^{\text{ZM, (3)}}$  in the high- $Q^2$  limit. Eq. (6.1) is then stable across all  $(x, Q^2)$ , exactly cancelling any discontinuity that would violate the RG flow, whilst also demanding that the known FFNS approximation (for  $Q^2 < m_h^2$ ) is followed.<sup>13</sup>

Figure 14 shows the result of estimating  $C_{H,q}^{\text{VF, (3)}}$  using the above approximation for  $C_{H,q}^{\text{FF, (3)}}$  and the relevant subtraction

<sup>13</sup> Since in practice the discontinuities from the transition matrix elements are added to PDFs regardless of what order coefficient function they are convoluted with, discontinuities of even higher order (e.g.  $\alpha_s^4$  and beyond) are also present in calculations. Because the order  $\alpha_s^3$  matrix elements are large these even higher order discontinuities are not insignificant. Therefore we add the same contributions to the unknown FFNS contributions below  $m_h^2$  to impose continuity on structure functions. Such corrections are extremely small, except right at the transition point where they eliminate minor unphysical discontinuities.



**Fig. 13** The NNLO GM-VFNS function  $C_{H,q}^{\text{VF}, (2)}$  compared with the NNLO ZM-VFNS function  $C_{H,q}^{\text{ZM}, (2)}$  across a variety of  $x$  and  $Q^2$  values. Mass threshold is set at the charm quark level ( $m_h^2 = m_c^2 = 1.4 \text{ GeV}^2$ )



**Fig. 14**  $N^3LO$  GM-VFNS function  $C_{H,q}^{VF,(3)}$  compared with the  $N^3LO$  ZM-VFNS function  $C_{H,q}^{ZM,(3)}$  across a variety of  $x$  and  $Q^2$  values (shown without the variation from the low- $Q^2$  NLL term discussed

in Sect. 6.2).  $C_{H,q}^{VF,(3)}$  is parameterised via Eqs. (6.8), (6.9) and (6.1). Mass threshold is set at the charm quark level ( $m_h^2 = m_c^2 = 1.4 \text{ GeV}^2$ )

term from Eq. (6.8). Note that this plot ignores any variation from the low- $Q^2$  NLL term discussed in Sect. 6.2, where this is fixed to its central value.

$C_{H,g}$

As with  $C_{H,q}$ , using Eq. (3.9) and Eq. (3.13) to isolate  $C_{H,g}$  and relate the FFNS and GM-VFNS schemes,

$$C_{H,g}^{\text{FF}} = C_{H,g}^{\text{VF}} \otimes A_{gg,H} + C_{H,g}^{\text{VF, PS}} \otimes A_{qg,H} + \left[ C_{H,H}^{\text{VF, NS}} + C_{H,H}^{\text{VF, PS}} \right] \otimes A_{Hg} \quad (6.10)$$

$$\mathcal{O}(\alpha_s) : C_{H,g}^{\text{FF, (1)}} = C_{H,g}^{\text{VF, (1)}} + C_{H,H}^{\text{VF, (0)}} \otimes A_{Hg}^{\text{(1)}} \quad (6.11)$$

$$\mathcal{O}(\alpha_s^2) : C_{H,g}^{\text{FF, (2)}} = C_{H,g}^{\text{VF, (2)}} + C_{H,g}^{\text{VF, (1)}} \otimes A_{gg,H} + C_{H,H}^{\text{VF, (0)}} \otimes A_{Hg}^{\text{(2)}} + C_{H,H}^{\text{VF, (1)}} \otimes A_{Hg}^{\text{(1)}} \quad (6.12)$$

$$\mathcal{O}(\alpha_s^3) : C_{H,g}^{\text{FF, (3)}} = C_{H,g}^{\text{VF, (3)}} + C_{H,g}^{\text{VF, (2)}} \otimes A_{gg,H}^{\text{(1)}} + C_{H,g}^{\text{VF, (1)}} \otimes A_{gg,H}^{\text{(2)}} + C_{H,H}^{\text{VF, NS+PS, (2)}} \otimes A_{Hg}^{\text{(1)}} + C_{H,H}^{\text{VF, (1)}} \otimes A_{Hg}^{\text{(2)}} + C_{H,H}^{\text{VF, (0)}} \otimes A_{Hg}^{\text{(3)}} \quad (6.13)$$

we uncover a NLO contribution to the heavy flavour structure function. This lower order contribution is a consequence of the gluon being able to directly probe the heavy flavour quarks, whereas a light quark must interact via a secondary interaction (hence the  $C_{H,q}$  coefficient function beginning at NNLO).

### NLO and NNLO

The NLO and NNLO contributions to  $C_{H,g}$  are known exactly [43]. To build some experience and check our understanding, we can observe how the lower order GM-VFNS functions converge onto their ZM-VFNS counterparts in Figs. 15 and 16.

At NLO and NNLO the magnitude of the functions is generally higher in the low- $Q^2$  limit than at high- $Q^2$ . In both cases, the function remains at the same order of magnitude across all  $Q^2$ . However, the relative change across  $Q^2$  is smaller at NLO, and similar to that seen for  $C_{H,q}^{(2)}$  at NNLO. Due to this, we can once again expect that although more of a scaling contribution at  $N^3\text{LO}$  may be present, it should not be too substantial across the range of  $Q^2$ .

### $N^3\text{LO}$

As with the  $C_{H,q}^{(3)}$  function at  $\mathcal{O}(\alpha_s^3)$ , the FFNS result at low- $Q^2$  is known (up to the level of precision discussed in

Sect. 6.2), as well as the exact ZM-VFNS function at high- $Q^2$  [47–49, 57]. Considering the form of  $C_{H,g}^{\text{VF, (3)}}$ , there is an extra complication coming from the transition matrix element  $A_{Hg}^{(3)}$ . As discussed in Sect. 5.1, the  $A_{Hg}^{(3)}$  function is not as well known as the  $A_{Hg}^{(3)}$  function considered earlier and is accompanied by the variational parameter  $a_{Hg}$ . Since it is a requirement for  $C_{H,g}^{(3)}$  to exactly cancel the PDF discontinuity introduced by  $A_{Hg}^{(3)}$ , this variation must be compensated for and included in the description,

$$C_{H,g}^{\text{VF, (3)}} = C_{H,g}^{\text{FF, (3)}} - C_{H,g}^{\text{VF, (2)}} \otimes A_{gg,H}^{\text{(1)}} - C_{H,g}^{\text{VF, (1)}} \otimes A_{gg,H}^{\text{(2)}} - C_{H,H}^{\text{VF, NS+PS, (2)}} \otimes A_{Hg}^{\text{(1)}} - C_{H,H}^{\text{VF, (1)}} \otimes A_{Hg}^{\text{(2)}} - A_{Hg}^{(3)} \quad (6.14)$$

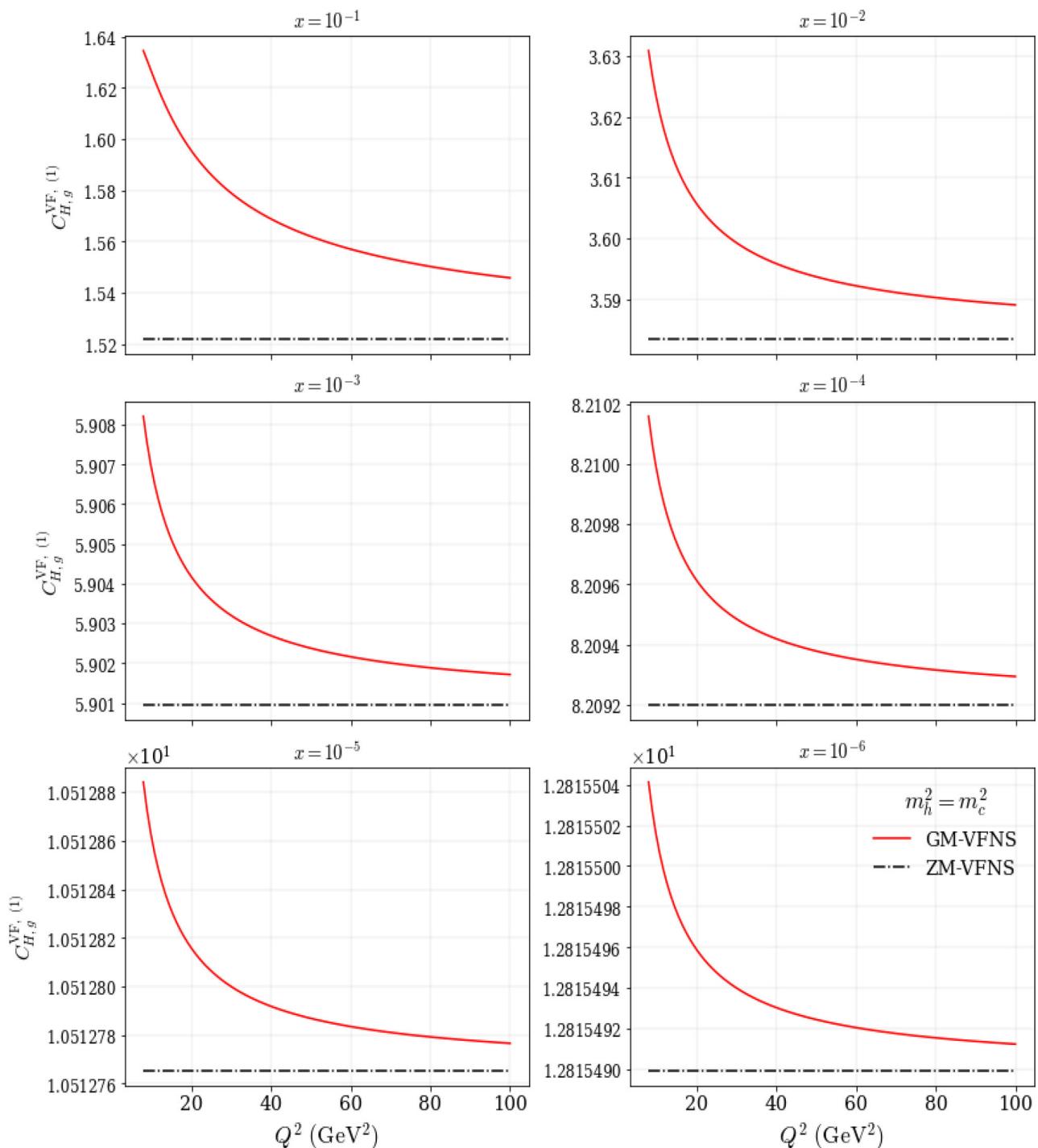
As in Sect. 6.3, transitioning to the FFNS number scheme ensures an exact cancellation via the subtraction term in Eq. (6.14). Using the exact information for  $C_{H,g}^{\text{FF, (3)}}(Q^2 \rightarrow 0)$  and the known high- $Q^2$  limit,

$$C_{H,g}^{\text{FF, (3)}}(Q^2 \rightarrow \infty) = C_{H,g}^{\text{ZM, (3)}} + C_{H,g}^{\text{VF, (2)}} \otimes A_{gg,H}^{\text{(1)}} + C_{H,g}^{\text{VF, (1)}} \otimes A_{gg,H}^{\text{(2)}} + C_{H,H}^{\text{VF, NS+PS, (2)}} \otimes A_{Hg}^{\text{(1)}} + C_{H,H}^{\text{VF, (1)}} \otimes A_{Hg}^{\text{(2)}} + A_{Hg}^{(3)} \quad (6.15)$$

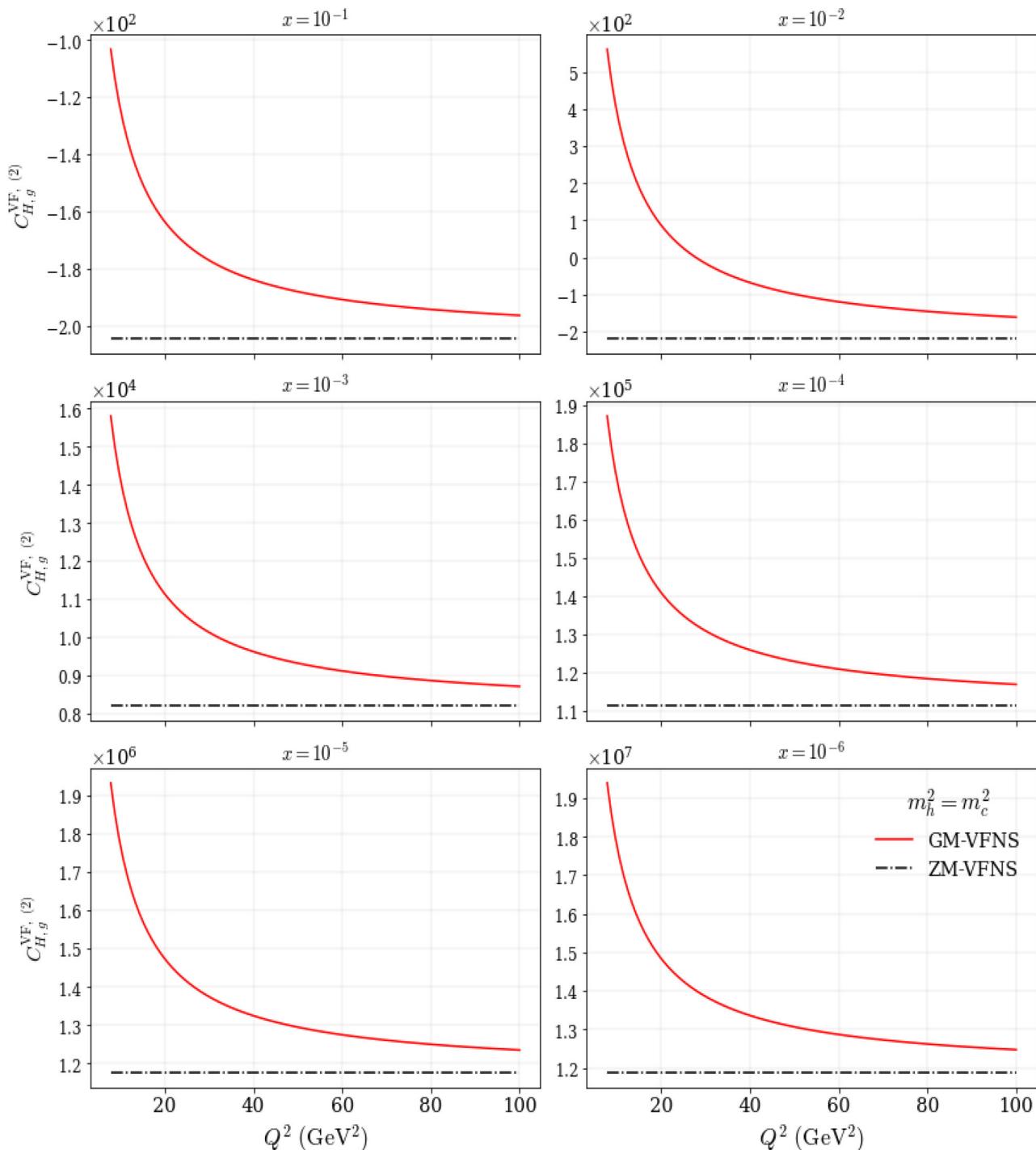
where  $C_{H,g}^{\text{VF, (3)}}$  is replaced with  $C_{H,g}^{\text{ZM, (3)}}$  in the high- $Q^2$  limit. Applying the framework set out in Eq. (6.1), the resulting parameterisation is stable across all  $(x, Q^2)$ . As  $A_{Hg}^{(3)}$  and its variation is explicitly included in Eq. (6.14) this ensures the continuity of the structure function with exact cancellations of discontinuities at mass thresholds.

Figure 17 displays our approximation for the GM-VFNS coefficient function across a range of  $(x, Q^2)$  via a parameterisation for  $C_{H,g}^{\text{FF, (3)}}$  and the relevant subtraction term in Eq. (6.14). Figure 17 also contains the uncertainty in this approximation stemming from  $A_{Hg}^{(3)}$  (see Sect. 5.1). Note that Fig. 17 ignores any variation from the low- $Q^2$  NLL term discussed in Sect. 6.2, where this is fixed to its central value. The uncertainty shown in Fig. 17 is suppressed as we move to high- $Q^2$  owing to the required convergence of the GM-VFNS onto the corresponding ZM-VFNS gluon coefficient function at  $N^3\text{LO}$ .

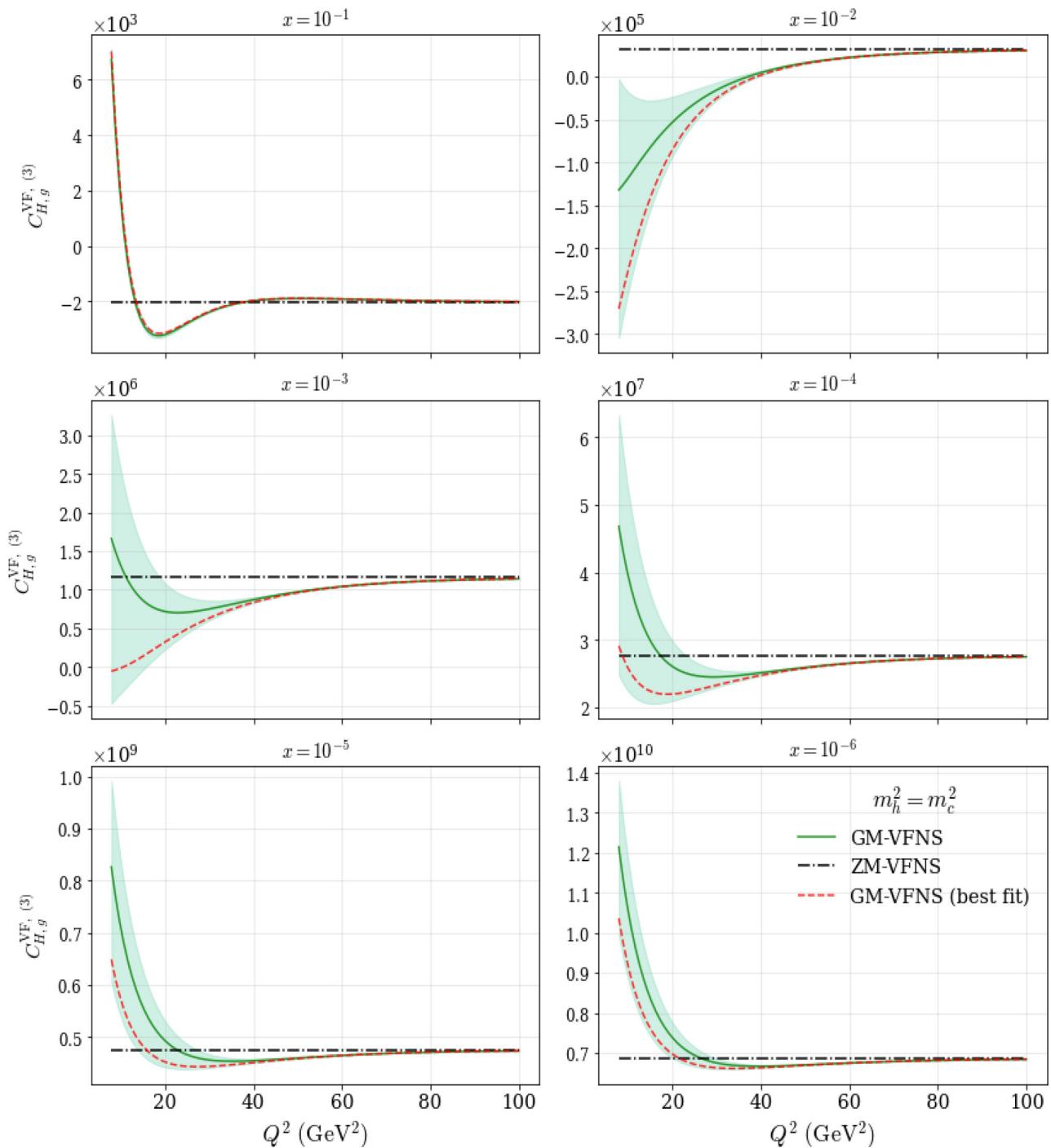
Included in Fig. 17 is the best fit prediction for  $C_{H,g}^{\text{VF, (3)}}$  (corresponding to the best fit of  $A_{Hg}^{(3)}$  approximated in Sect. 5). Overall we see the resultant shape of  $C_{H,g}^{(3)}$  is within our predicted range and follows a sensible shape that matches with the known high- $Q^2$  FFNS behaviour. Contrasting this with NNLO, the shape across the range of  $x$  values shown is less consistent. There is no guarantee that this should be the case, since we do not know how the perturbative nature of QCD will behave. However, we do maintain the relatively



**Fig. 15** The NLO GM-VFNS function  $C_{H,g}^{\text{VF},(1)}$  compared with the NLO ZM-VFNS function  $C_{H,g}^{\text{ZM},(1)}$  across a variety of  $x$  and  $Q^2$  values. Mass threshold is set at the charm quark level ( $m_h^2 = m_c^2 = 1.4 \text{ GeV}^2$ )



**Fig. 16** The NNLO GM-VFNS function  $C_{H,g}^{\text{VF}, (2)}$  compared with the NNLO ZM-VFNS function  $C_{H,g}^{\text{ZM}, (2)}$  across a variety of  $x$  and  $Q^2$  values. Mass threshold is set at the charm quark level ( $m_h^2 = m_c^2 = 1.4$  GeV $^2$ )



**Fig. 17** The N<sup>3</sup>LO GM-VFNS function  $C_{H,g}^{\text{VF}, (3)}$  compared with the N<sup>3</sup>LO ZM-VFNS function  $C_{H,g}^{\text{ZM}, (3)}$  across a variety of  $x$  and  $Q^2$  values (shown without the variation from the low- $Q^2$  NLL term discussed

in Sect. 6.2).  $C_{H,g}^{\text{VF}, (3)}$  is parameterised via Eqs. (6.14), (6.15) and (6.1). Mass threshold is set at the charm quark level ( $m_h^2 = m_c^2 = 1.4 \text{ GeV}^2$ )

consistent order of magnitude across the evolution in  $Q^2$ , therefore the exact form of the shape across  $Q^2$  will be less important in the resultant structure function picture.

$$C_{q,q}^{\text{NS}}$$

The light quark coefficient functions involve small heavy flavour contributions at higher orders from heavy quarks produced away from the photon vertex. As discussed in Sect. 6.1 the low- $Q^2$  FFNS function in this case is unknown. However, since the heavy flavour contributions to the light quark structure function  $F_{2,q}(x, Q^2)$  are very small, any choice of sensible variation in  $Q^2$  has a near negligible effect on the overall structure function. Further to this, as is apparent from lower order examples, it can be expected that the light quark coefficient functions remain relatively constant across  $Q^2$ .

Using Eqs. (3.9) and (3.11), the non-singlet coefficient function is stated as,

$$C_{q,q}^{\text{FF, NS}} = A_{qq,H}^{\text{NS}} \otimes C_{q,q}^{\text{VF, NS}}, \quad (6.16)$$

$$\mathcal{O}(\alpha_s^0) : C_{q,q, \text{NS}}^{\text{FF, (0)}} = C_{q,q, \text{NS}}^{\text{VF, (0)}} \quad (6.17a)$$

$$\mathcal{O}(\alpha_s^1) : C_{q,q, \text{NS}}^{\text{FF, (1)}} = C_{q,q, \text{NS}}^{\text{VF, (1)}} \quad (6.17b)$$

$$\mathcal{O}(\alpha_s^2) : C_{q,q, \text{NS}}^{\text{FF, (2)}} = C_{q,q, \text{NS}}^{\text{VF, (2)}} + A_{qq,H}^{\text{NS, (2)}} \quad (6.17c)$$

$$\begin{aligned} \mathcal{O}(\alpha_s^3) : C_{q,q, \text{NS}}^{\text{FF, (3)}} = & C_{q,q, \text{NS}}^{\text{VF, (3)}} + A_{qq,H}^{\text{NS, (3)}} \\ & + C_{q,q, \text{NS}}^{\text{VF, (1)}} \otimes A_{qq,H}^{\text{NS, (2)}}. \end{aligned} \quad (6.17d)$$

From Eq. (6.17) the FFNS contribution at LO and NLO is identical to the GM-VFNS and ZM-VFNS function at high- $Q^2$ . Physically for heavy quarks to affect light quarks, a larger number of vertices than are allowed at LO and NLO must be present to enable interactions involving heavy quarks. We therefore begin our discussion at NNLO.

### NNLO

At NNLO the functions included in Eq. (6.17c) are known exactly [39, 58]. Assembling these together, we provide an example of how the GM-VFNS function converges to the familiar ZM-VFNS function for the light quark. By performing this exercise, expectations as to how  $C_{q,q}^{\text{NS}}$  will behave at  $N^3\text{LO}$  can be constructed.

From Fig. 18  $C_{q,q, \text{NS}}^{\text{VF, (2)}}$  quickly converges onto the ZM-VFNS function with the difference between the low and high- $Q^2$  being within 10% at large- $x$  and within 0.01% at small- $x$ . This weak scaling with  $Q^2$  reinforces the statement that it is possible to approximate the  $N^3\text{LO}$  function relatively well without extensive low- $Q^2$  information.

### $N^3\text{LO}$

Equation (6.17d) involves a mixture of functions known exactly (ZM-VFNS high- $Q^2$  limit [57]) and functions that are completely unknown ( $C_{q,q, \text{NS}}^{\text{FF, (3)}}$ ). This presents an issue as it is no longer possible to rely on  $C_{q,q, \text{NS}}^{\text{FF, (3)}}$  to constrain the low- $Q^2$  limit. Nevertheless, by utilising the experience gained from NNLO, it is feasible to choose any sensible choice for the low- $Q^2$  limit. In practice, due to the observed weak scaling in  $Q^2$ , the exact form at low- $Q^2$  will not present any noticeable differences.

A naive choice for heuristically placing the  $C_{q,q, \text{NS}}^{\text{FF, (3)}}(Q^2 \rightarrow 0)$  function would be a constant value i.e. no scaling in  $Q^2$ . We propose to use the intuition from NNLO and the overall fit quality to give us potentially a more sensible and viable choice for the GM-VFNS approximation.<sup>14</sup> By inserting the high- $Q^2$  limit into the NS part of Eq. (6.2), the result is a crude approximation to  $C_{q,q, \text{NS}}^{\text{FF, (3)}}(Q^2 \rightarrow 0)$ . Combining this with Eq. (6.17d), we obtain a GM-VFNS parameterisation which is relatively constant across  $Q^2$  (similar to the NNLO behaviour) with any differences arising from the subtraction terms which are known.

Figure 19 shows the result of this approximation for the full  $C_{q,q, \text{NS}}^{\text{VF, (3)}}$  function. We notice that the behaviour is similar to that of NNLO across all  $(x, Q^2)$  and appropriately larger in magnitude to account for the extra contributions obtained at  $N^3\text{LO}$  compared to NNLO. By definition, the parameterisation converges well to the ZM-VFNS scheme with the magnitude at high- $Q^2$  (ZM-VFNS regime) remaining similar to that at low- $Q^2$  for each specific value of  $x$ . This final point gives assurances that even if this low- $Q^2$  guess is not entirely representative of the actual  $N^3\text{LO}$  function, the effects of including this approximation are virtually negligible in a PDF fit. Also shown in Fig. 19 is the variation in the  $C_{q,q, \text{NS}}^{\text{VF, (3)}}$  function stemming solely from the  $A_{qq,H}^{\text{NS, (3)}}$  function.

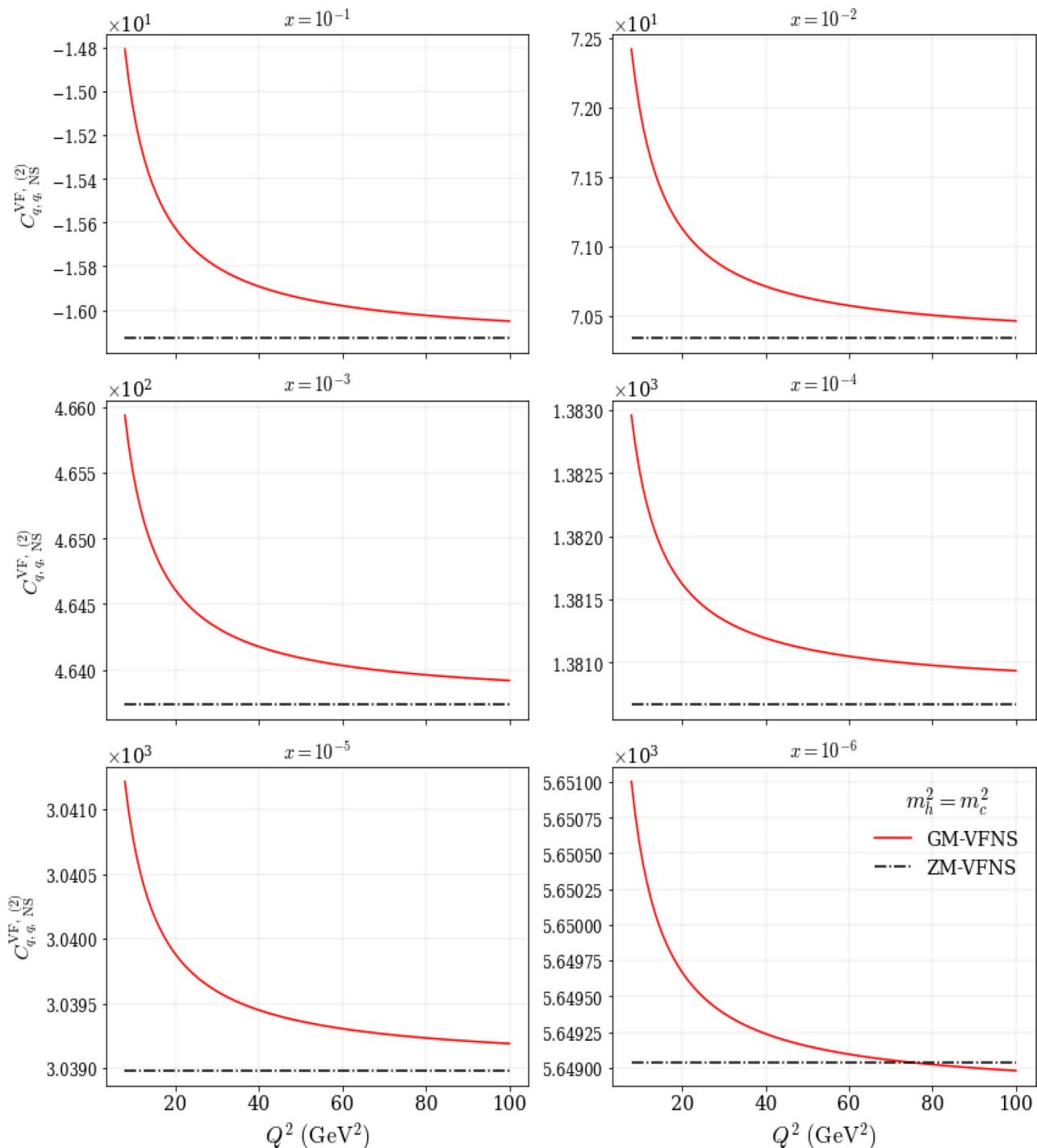
$$C_{q,q}^{\text{PS}}$$

To complete the light-quark GM-VFNS coefficient function picture the pure-singlet contribution from Eq. (3.9) and Eq. (3.11) is described by,

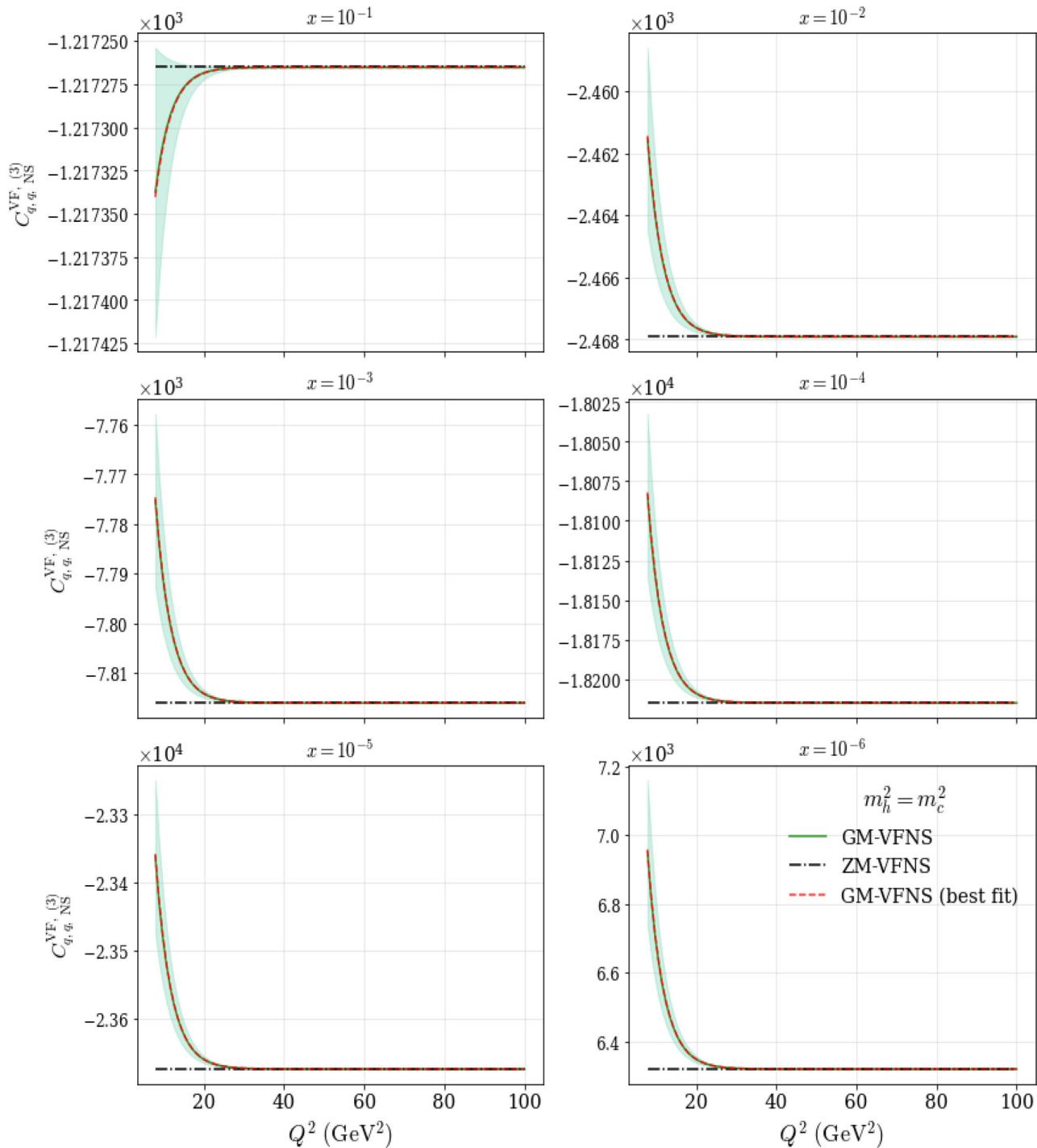
$$\begin{aligned} C_{q,q}^{\text{FF, PS}} = & C_{q,q}^{\text{VF, PS}} \otimes A_{qq,H}^{\text{PS}} + C_{q,g}^{\text{VF}} \otimes A_{gq,H} \\ & + C_{q,H}^{\text{VF, PS}} \otimes A_{Hq} \end{aligned} \quad (6.18)$$

$$\mathcal{O}(\alpha_s^0) : C_{q,q}^{\text{FF, PS, (0)}} = 0 \quad (6.19a)$$

<sup>14</sup> The differences in fit quality for sensible choices are < 0.05% compared to the overall  $\chi^2$  for the light quark NS coefficient function.



**Fig. 18** The NNLO GM-VFNS function  $C_{q,q,NS}^{VF,(2)}$  compared with the NNLO ZM-VFNS function  $C_{q,q,NS}^{ZM,(2)}$  across a variety of  $x$  and  $Q^2$  values. Mass threshold is set at the charm quark level ( $m_h^2 = m_c^2 = 1.4$  GeV $^2$ )



**Fig. 19** The  $N^3\text{LO}$  GM-VFNS function  $C_{q,q,\text{NS}}^{\text{VF},(3)}$  compared with the  $N^3\text{LO}$  ZM-VFNS function  $C_{q,q,\text{NS}}^{\text{ZM},(3)}$  across a variety of  $x$  and  $Q^2$  values.  $C_{q,q,\text{NS}}^{\text{VF},(3)}$  is parameterised via Eqs. (6.17d) and (6.2). Mass threshold is set at the charm quark level ( $m_h^2 = m_c^2 = 1.4 \text{ GeV}^2$ )

$$\mathcal{O}(\alpha_s^1) : C_{q,q}^{\text{FF, PS}, (1)} = 0 \quad (6.19\text{b})$$

$$\mathcal{O}(\alpha_s^2) : C_{q,q}^{\text{FF, PS}, (2)} = C_{q,q, \text{PS}}^{\text{VF}, (2)} \quad (6.19\text{c})$$

$$\mathcal{O}(\alpha_s^3) : C_{q,q}^{\text{FF, PS}, (3)} = C_{q,q, \text{PS}}^{\text{VF}, (3)} + C_{q,g}^{\text{VF}, (1)} \otimes A_{gq, H}^{(2)}. \quad (6.19\text{d})$$

As with the non-singlet analysis the heavy flavour contributions to the pure-singlet appear at higher orders to allow for the possibility of heavy quark contributions. In the pure-singlet case, the heavy flavour contributions are pushed one order higher than the non-singlet due to the requirement for an extra intermediary gluon.

$N^3LO$

In the pure-singlet case, the FFNS function is non-existent up until  $N^3LO$ . Because of this, we choose to parameterise the pure-singlet with a weak constraint suppressing the FFNS function  $C_{q,q}^{\text{FF, PS}, (3)}$  across all  $x$  for very low- $Q^2$ . The reason for this is that the coefficient functions acquire more contributions as they exist through higher orders. If  $C_{q,q}^{\text{FF, PS}, (3)}$  is beginning at this order, then one could expect the low- $Q^2$  form to be relatively small compared to the known ZM-VFNS function [57]. This is somewhat justified by the low- $Q^2$  kinematic restrictions for the singlet distribution which broadly manifest into a suppression at low- $Q^2$ . We reiterate here that the low- $Q^2$  form of this function is still essentially around the same magnitude across all  $Q^2$ . Therefore, as with  $C_{q,q}^{\text{FF, NS}, (3)}$ , it will be virtually negligible in the overall structure function.

After constructing the approximation for  $C_{q,q}^{\text{FF, PS}, (3)}$  with Eq. (6.2), Eq. (6.19d) is used to approximate the GM-VFNS function. The exact form of Eq. (6.19d) is chosen based on intuition and where the best fit quality can be achieved.<sup>15</sup>

It can be seen from Fig. 20 that the overall magnitude of  $C_{q,q, \text{PS}}^{\text{VF}, (3)}$  decreases substantially towards large- $x$  as one would expect from a pure-singlet function. Inspecting the predicted values of  $C_{q,q, \text{PS}}^{\text{VF}, (3)}$ , we can confirm that the non-singlet function from Fig. 19 begins to dominate at large- $x$ . Conversely towards small- $x$ ,  $C_{q,q, \text{PS}}^{\text{VF}, (3)}$  is much larger than  $C_{q,q, \text{NS}}^{\text{VF}, (3)}$ , thereby preserving the familiar interplay between quark distributions. The suppression of the FFNS parameterisation towards low- $Q^2$  is also seen to give sensible results in terms of the expected percentage change in magnitude through the range of  $Q^2$  values. Specifically we see  $< 10\%$  difference in magnitude between low and high- $Q^2$ . Since scale violating terms become more dominant at higher orders and we are essentially at leading order in terms of heavy flavour contributions, a high level of scaling with  $Q^2$  is not expected at this order.

<sup>15</sup> The differences in fit quality for sensible choices are  $< 0.1\%$  compared to the overall  $\chi^2$  for the light quark PS coefficient function.

$$C_{q,g}$$

Finally the gluon-light quark coefficient function is constructed from Eqs. (3.9) and (3.11) to be,

$$C_{q,g}^{\text{FF}} = C_{q,q}^{\text{VF}} \otimes A_{qg, H} + C_{q,g}^{\text{VF}} \otimes A_{gg, H} + C_{q,H}^{\text{VF, PS}} \otimes A_{Hg} \quad (6.20)$$

$$\mathcal{O}(\alpha_s^0) : C_{q,g}^{\text{FF}, (0)} = 0 \quad (6.21\text{a})$$

$$\mathcal{O}(\alpha_s^1) : C_{q,g}^{\text{FF}, (1)} = C_{q,g}^{\text{VF}, (1)} \quad (6.21\text{b})$$

$$\mathcal{O}(\alpha_s^2) : C_{q,g}^{\text{FF}, (2)} = C_{q,g}^{\text{VF}, (2)} + C_{q,g}^{\text{VF}, (1)} \otimes A_{gg, H}^{(1)} + A_{qg, H}^{(2)} \quad (6.21\text{c})$$

$$\mathcal{O}(\alpha_s^3) : C_{q,g}^{\text{FF}, (3)} = C_{q,g}^{\text{VF}, (3)} + A_{qg, H}^{(3)} + C_{q,q}^{\text{VF}, (1)} \otimes A_{qg, H}^{(2)} + C_{q,g}^{\text{VF}, (2)} \otimes A_{gg, H}^{(1)} + C_{q,g}^{\text{VF}, (1)} \otimes A_{gg, H}^{(2)} + C_{q,H}^{\text{VF, (2)}} \otimes A_{Hg}^{(1)}. \quad (6.21\text{d})$$

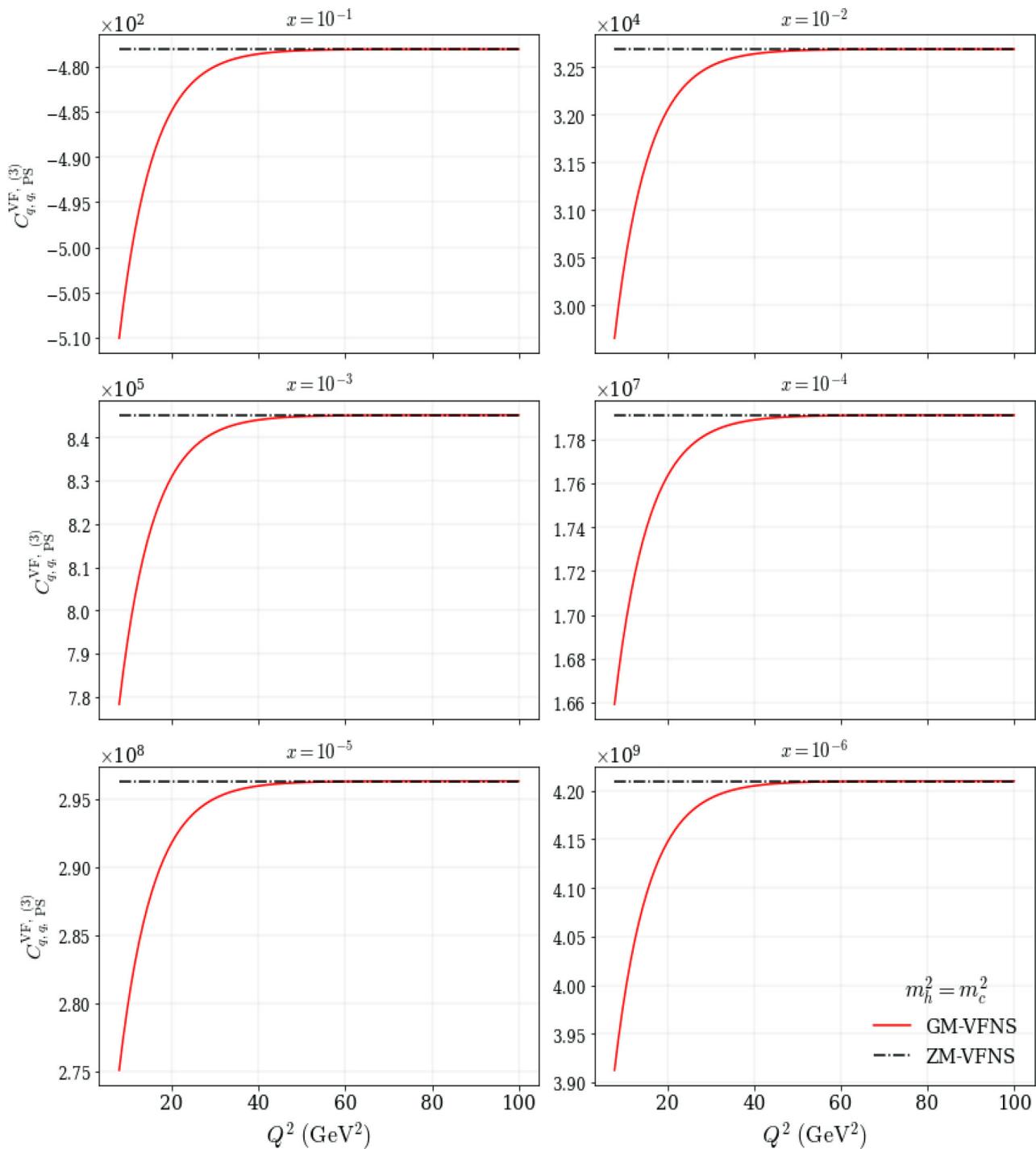
For  $C_{q,g}$ , the FFNS function is non-existent up to NNLO, similar to  $C_{q,q, \text{NS}}^{\text{FF}, (3)}$ . However, the  $A_{qg, H}$  contribution at NNLO is sub-leading in  $n_f$  [39] and is therefore not considered here.

$N^3LO$

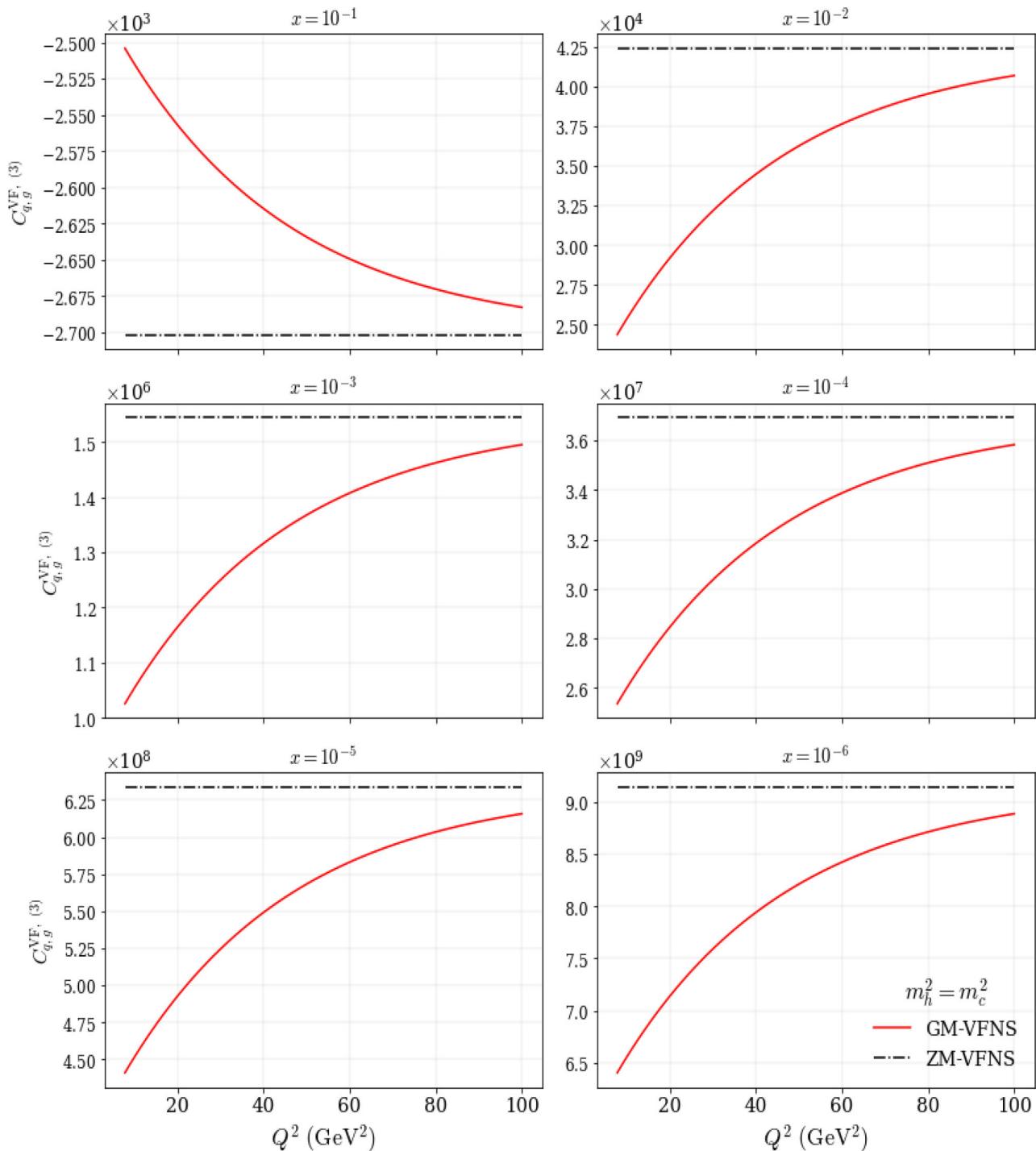
At  $N^3LO$  in Eq. (6.21d), no information is available for the  $C_{q,g}^{\text{FF}, (3)}$  at low- $Q^2$ . Whereas at high- $Q^2$  the ZM-VFNS function is known [57]. To construct the parameterisation, we apply the same method described for  $C_{q,q, \text{PS}}^{\text{FF}, (3)}$ . Specifically, by applying a suppression to the FFNS parameterisation in the low- $Q^2$  limit. After constructing the parameterisation for  $C_{q,g}^{\text{FF}, (3)}$  with Eq. (6.2), Eq. (6.21d) is used to approximate the GM-VFNS function. Since there is no information in the low- $Q^2$  limit, the parameterisation in Eq. (6.2) is chosen roughly based on how the fit prefers the evolution in  $Q^2$  to behave.

Figure 21 illustrates the GM-VFNS function in Eq. (6.21d) with Eq. (6.2) as  $C_{q,g}^{\text{FF}, (3)}$  across a range of  $x$  and  $Q^2$ .  $C_{q,g}^{\text{VF}, (3)}$  increases in magnitude when moving to smaller  $x$  and by definition converges onto the ZM-VFNS function. The convergence in this case is chosen to be less steep than for the light quark convergences due to some minor tensions in the fit.<sup>16</sup> The magnitude of  $C_{q,g}^{\text{VF}, (3)}$  across the entire range of  $Q^2$  is still relatively constant, although less flat than the behaviour predicted for  $C_{q,q, \text{PS/NS}}^{\text{VF}, (3)}$ . However, considering Eq. (6.21d), some justification for this behaviour can be offered. When

<sup>16</sup> The differences in fit quality for sensible choices of Eq. (6.2) are  $< 0.5\%$  compared to the overall  $\chi^2$  for the light quark gluon coefficient function.



**Fig. 20** The N<sup>3</sup>LO GM-VFNS function  $C_{q,q,PS}^{VF,(3)}$  compared with the N<sup>3</sup>LO ZM-VFNS function  $C_{q,q,PS}^{ZM,(3)}$  across a variety of  $x$  and  $Q^2$  values.  $C_{q,q,PS}^{VF,(3)}$  is parameterised via Eqs. (6.19d) and (6.2). Mass threshold is set at the charm quark level ( $m_h^2 = m_c^2 = 1.4 \text{ GeV}^2$ )



**Fig. 21** The N<sup>3</sup>LO GM-VFNS function  $C_{q,g}^{\text{VF}, (3)}$  compared with the N<sup>3</sup>LO ZM-VFNS function  $C_{q,g}^{\text{ZM}, (3)}$  across a variety of  $x$  and  $Q^2$  values.  $C_{q,g}^{\text{VF}, (3)}$  is parameterised via Eqs. (6.21d) and (6.2). Mass threshold is set at the charm quark level ( $m_h^2 = m_c^2 = 1.4 \text{ GeV}^2$ )

comparing the contributions to the FFNS functions in the NS, PS and gluon cases (Eqs. (6.17d), (6.19d) and (6.21d) respectively), the  $A_{Hg}$  and  $A_{gg,H}$  contributions involved in  $C_{q,g}^{\text{FF}}$  are much larger than the contributions from  $A_{gq,H}$ ,  $A_{Hq}$  and  $A_{qq,H}^{\text{NS}}$ . Therefore we can expect a larger difference across  $Q^2$  for the  $C_{q,g}^{\text{VF},(3)}$  function. With this being said, the specific form at low- $Q^2$  is not very important in current PDF fits, only that the form is continuous and valid.

## 7 $\text{N}^3\text{LO } K\text{-factors}$

Thus far the primary concern has been the  $\text{N}^3\text{LO}$  additions to the theoretical form of the DIS cross section. However, to complement these changes it is necessary to extend other cross section data to the same order. With these ingredients it is possible to maintain a consistent approximate  $\text{N}^3\text{LO}$  treatment across all datasets. At the time of writing,  $K$ -factors which provide exact transformations for each dataset up to NNLO are available.<sup>17</sup> Although there has been progress in  $\text{N}^3\text{LO}$  calculations for various processes including Drell–Yan (DY), top production and Higgs processes [63–76], there is still missing information on how these  $K$ -factors behave above NNLO. In this section we investigate the effects of the  $K$ -factors for each dataset when extended to  $\text{N}^3\text{LO}$ . Five process categories are considered separately: Drell–Yan, Jets,  $p_T$  jets,  $t\bar{t}$  production and Dimuon data. Inside each of these process categories we assume a perfect positive correlation between the behaviour of datasets i.e. all Drell–Yan  $K$ -factor shifts from NNLO are positively correlated. Clearly this treatment is a simplification, based on the expectation of a high degree of correlation between datasets concerned with the same processes. In practice, the uncertainty introduced from including these  $K$ -factors is already relatively small compared to the other sources of MHOUs already discussed, therefore any correction to this is guaranteed to be small (this will be shown more clearly in Sect. 8).

### 7.1 Extension to $\text{aN}^3\text{LO}$

The extension to  $\text{aN}^3\text{LO}$  is parameterised with a mixture of the NLO and NNLO  $K$ -factors. This allows control of the magnitude and shape of the transformation from NNLO to  $\text{aN}^3\text{LO}$ , using the known shifts from lower orders.

The basic idea is presented as,

$$K^{\text{N}^3\text{LO}/\text{LO}} = a_{\text{NNLO}} K^{\text{NNLO}/\text{LO}} + a_{\text{NLO}} K^{\text{NLO}/\text{LO}}, \quad (7.1)$$

where  $K^{\text{N}^3\text{LO}/\text{LO}}$ ,  $K^{\text{NNLO}/\text{LO}}$  and  $K^{\text{NLO}/\text{LO}}$  are the relevant  $K$ -factors with respect to the LO cross section, and  $a_{\text{N}(\text{N})\text{LO}}$

<sup>17</sup> An exception to this is the CMS 7 TeV  $W + c$  [62] dataset where  $K$ -factors are available only up to NLO.

are variational parameters controlling the mixture of NNLO and NLO  $K$ -factors included in the  $\text{N}^3\text{LO } K$ -factor approximation. Hence we have 2 parameters for each of the five processes included in the fit, and now 20 theory nuisance parameters in total – 10 controlling  $\text{aN}^3\text{LO } K$ -factors, 5 controlling  $\text{aN}^3\text{LO}$  splitting functions and 5 controlling heavy flavour  $\text{aN}^3\text{LO}$  contributions.

To describe this formalism in terms of physical observables we consider the cross section,

$$\sigma = \sigma_0 + \sigma_1 + \sigma_2 + \dots \equiv \sigma_{\text{NNLO}} + \dots, \quad (7.2)$$

where there is an implicit order of  $\alpha_s^{p+i}$  absorbed into the definition of  $\sigma_i$  beginning at the relevant LO for each process, i.e.  $p = 0$  for DY.

$K^{\text{NLO}/\text{LO}}$  is then the relative shift from  $\sigma_{\text{LO}}$  to  $\sigma_{\text{NLO}}$ ,

$$K^{\text{NLO}/\text{LO}} = \frac{\sigma_0 + \sigma_1}{\sigma_0} = 1 + \frac{\sigma_1}{\sigma_0}. \quad (7.3)$$

Similarly for NNLO we have,

$$K^{\text{NNLO}/\text{LO}} = \frac{\sigma_0 + \sigma_1 + \sigma_2}{\sigma_0} = 1 + \frac{\sigma_1}{\sigma_0} + \frac{\sigma_2}{\sigma_0}. \quad (7.4)$$

Moving to  $\text{N}^3\text{LO}$ , we write

$$\sigma = \sigma_0 + \sigma_1 + \sigma_2 + \sigma_3 + \dots \equiv \sigma_{\text{N}^3\text{LO}} + \dots, \quad (7.5)$$

where  $\sigma_3 = a_1\sigma_1 + a_2\sigma_2$  is approximated as some superposition of the two lower orders, with  $(a_1, a_2) = (0, 0)$  reproducing the NNLO case.

Pushing forward with this approximation and using the definitions for  $\sigma_{1,2}$  in terms of  $K$ -factors (Eqs. (7.3) and (7.4)) we have,

$$\sigma_{\text{N}^3\text{LO}} = \sigma_{\text{NNLO}} + a_1\sigma_1 + a_2\sigma_2 \quad (7.6)$$

$$= \sigma_{\text{NNLO}} + a_1\sigma_0(K^{\text{NLO}/\text{LO}} - 1) + a_2\sigma_0(K^{\text{NNLO}/\text{LO}} - K^{\text{NLO}/\text{LO}}) \quad (7.7)$$

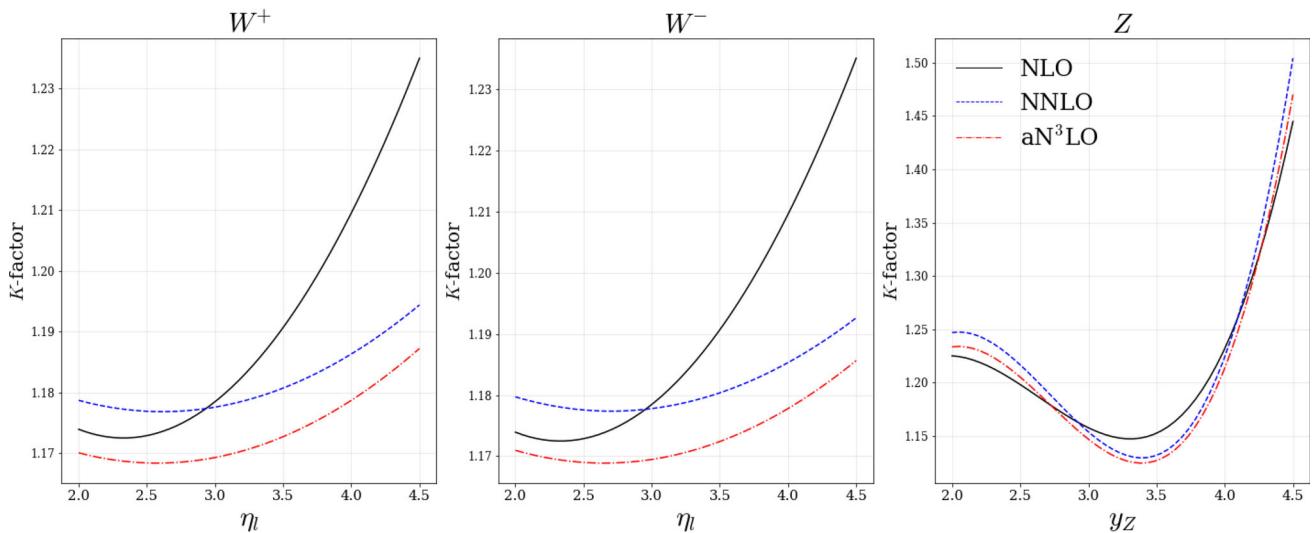
since,

$$\sigma_1 = \sigma_0(K^{\text{NLO}/\text{LO}} - 1) \quad (7.8)$$

$$\sigma_2 = \sigma_0(K^{\text{NNLO}/\text{LO}} - \sigma_1 - \sigma_0) = \sigma_0(K^{\text{NNLO}/\text{LO}} - K^{\text{NLO}/\text{LO}}). \quad (7.9)$$

From here one can obtain,

$$\begin{aligned} K^{\text{NNLO}/\text{LO}} - K^{\text{NLO}/\text{LO}} &= \frac{\sigma_2}{\sigma_0} = \frac{\sigma_2 + \sigma_0}{\sigma_0} - 1 \\ &\approx \frac{\sigma_2 + \sigma_1 + \sigma_0}{\sigma_1 + \sigma_0} - 1 \\ &= K^{\text{NNLO}/\text{NLO}} - 1, \end{aligned} \quad (7.10)$$



**Fig. 22**  $K$ -factor expansion up to  $aN^3LO$  shown for the LHCb 2015  $W$ ,  $Z$  dataset [77, 78]. The  $K$ -factors shown here are absolute i.e. all with respect to LO ( $K^{N^mLO/LO} \forall m \in \{1, 2, 3\}$ )

assuming  $\sigma_1 \ll \sigma_0$ , which is in general true for a valid perturbative expansion. Using (7.10)  $\sigma_{N^3LO}$  can be expressed by,

$$\sigma_{N^3LO} \simeq \sigma_{NNLO} \left( 1 + a_1(K^{NLO/LO} - 1) + a_2(K^{NNLO/NLO} - 1) \right), \quad (7.11)$$

where  $\sigma_2 \ll \sigma_1 \ll \sigma_0$ .

This defines the proposed approximated  $N^3LO$  cross section. It is given in terms of extra contributions from lower order shifts, which are controlled by variational parameters  $a_1$  and  $a_2$ . It is also true that the contributions to  $N^3LO$  are expected to be suppressed by  $\alpha_s/\pi$  in the NNLO case and  $(\alpha_s/\pi)^2$  in the NLO case to account for the strengths of each contribution. Currently this is taken into account within the variational parameters  $a_1, a_2$ . However for the purpose of this description, it is more appropriate to explicitly redefine  $a_1, a_2 = a_s^2 \hat{a}_1, a_s \hat{a}_2$  where  $a_s = \mathcal{N} \alpha_s$  and  $\mathcal{N}$  is some normalisation factor. This then results in,

$$K^{N^3LO/LO} = K^{NNLO/LO} \left( 1 + \hat{a}_1 \mathcal{N}^2 \alpha_s^2 (K^{NLO/LO} - 1) + \hat{a}_2 \mathcal{N} \alpha_s (K^{NNLO/NLO} - 1) \right), \quad (7.12)$$

where the LO cross section  $\sigma_0$  is cancelled and Eq. (7.12) is written in terms of the  $K$ -factor shifts only. Equation (7.12) also implicitly includes the correct order  $\mathcal{O}(\alpha_s^3)$  in the parameterisation through (7.3) and (7.4). We can then choose  $\mathcal{N}$  in order to set the approximate magnitude of our variational parameters  $\hat{a}_1, \hat{a}_2$ . Given  $\alpha_s \sim 0.1$  for the processes considered, if we neglect  $\mathcal{N}$  (i.e. choose  $\mathcal{N} \sim 1$ ), then our order by order reduction in the magnitude of the  $K$ -factors would be  $\sim 10\%$  for  $\mathcal{O}(1)$  for variational parameters, however from previous orders we see that typically  $K$ -factors tend to be

30–40% of the previous order, therefore we instead choose  $\mathcal{N} = 3$ . This then ensures the natural scale of variation allowed is also of this order with  $\mathcal{O}(1)$  variational parameters describing the admixture of NLO and NNLO  $K$ -factors, with conservative penalties applied accordingly.

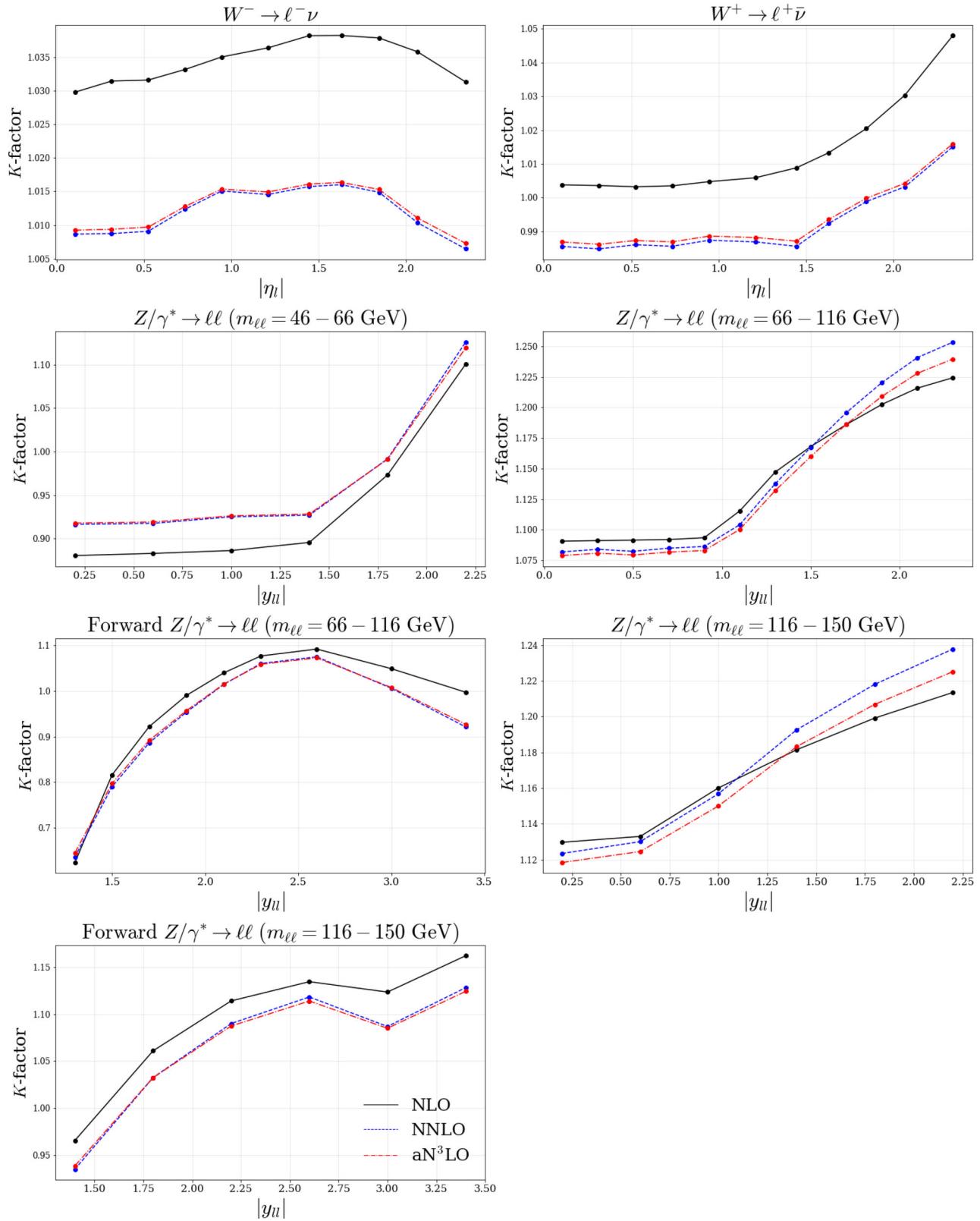
Reflecting on this, it is worth noting that these fitted  $K$ -factors will be sensitive to all orders, not just  $N^3LO$ . Considering these  $K$ -factors as approximating asymptotic behaviour to all orders in perturbation theory when assessing the stability of predictions, we can be less concerned with any somewhat large shifts from NNLO to  $aN^3LO$ , as we will specifically see in the case of Figs. 25 and 26. Finally, we remind the reader that at higher orders, new terms with more divergent leading logarithms appear which are missed by the current theoretical description. Due to this, the all-orders asymptotic description will still remain approximate up to the inclusion of more divergent leading logarithms in  $(x, Q^2)$  limits at even higher orders.

## 7.2 Numerical results

Using this formalism for the  $aN^3LO$   $K$ -factors, we present the global fit results for each of the five process categories considered.

### Drell–Yan processes

For the Drell–Yan processes (all calculated at  $\mu_{r,f} = m_{ll}/2$ ), a reduction of  $\sim 1\text{--}2\%$  in the  $K$ -factor shift is predicted across most of the corresponding datasets at  $aN^3LO$ . This is in agreement with recent work [64]. An example of this reduction is shown in Fig. 22.



**Fig. 23**  $K$ -factor expansion up to  $aN^3LO$  shown for the ATLAS 7 TeV high precision  $W$ ,  $Z$  dataset [79]. The  $K$ -factors shown here are absolute i.e. all with respect to LO ( $K^{N^mLO}/LO \forall m \in \{1, 2, 3\}$ )

**Table 2** Table showing the relevant DY datasets and how the individual  $\chi^2$  changes from NNLO by including the  $N^3LO$  treatment of  $K$ -factors, and theoretical  $N^3LO$  additions discussed earlier. The result with purely NNLO  $K$ -factors included for all data in the fit is also given

DY Dataset	$\chi^2$	$\Delta\chi^2$ from NNLO	$\Delta\chi^2$ from NNLO (NNLO $K$ -factors)
E866/NuSea $pp$ DY [80]	215.4/184	-9.7	+2.7
E866/NuSea $pd/pp$ DY [81]	8.4/15	-2.0	-1.1
DØ II $Z$ rap. [82]	17.4/28	+1.0	+0.7
CDF II $Z$ rap. [83]	40.6/28	+3.4	+1.4
DØ II $W \rightarrow \nu\mu$ asym. [84]	16.5/10	-0.8	-1.1
CDF II $W$ asym. [85]	18.2/13	-0.8	-0.8
DØ II $W \rightarrow \nu e$ asym. [86]	30.7/12	-3.2	-3.2
ATLAS $W^+, W^-, Z$ [87]	30.0/30	+0.1	+0.3
CMS $W$ asym. $p_T > 35$ GeV [88]	6.7/11	-1.1	-1.1
CMS $W$ asym. $p_T > 25, 30$ GeV [89]	7.9/24	+0.5	+0.3
LHCb $Z \rightarrow e^+e^-$ [90]	23.2/9	+0.5	+0.7
LHCb $W$ asym. $p_T > 20$ GeV [91]	12.6/10	+0.1	+0.3
CMS $Z \rightarrow e^+e^-$ [92]	17.6/35	-0.4	-0.3
ATLAS High-mass Drell–Yan [93]	18.4/13	-0.5	-0.9
CMS double diff. Drell–Yan [94]	133.2/132	-11.3	+9.7
LHCb 2015 $W, Z$ [77, 78]	103.2/67	+3.8	-0.1
LHCb 8 TeV $Z \rightarrow ee$ [95]	30.3/17	+4.1	+1.7
CMS 8 TeV $W$ [96]	11.5/22	-1.2	-0.2
ATLAS 7 TeV high prec. $W, Z$ [79]	119.3/61	+2.7	-8.3
DØ $W$ asym. [97]	12.2/14	+0.1	+1.3
ATLAS 8 TeV High-mass DY [98]	62.8/48	+5.7	+1.5
ATLAS 8 TeV $W$ [99]	53.0/22	-4.4	-0.1
ATLAS 8 TeV double diff. $Z$ [100]	80.4/59	+4.5	+6.7
Total	1069.4/864	-8.8	+10.4

Conversely, Fig. 23 displays an example where the  $K$ -factor shift has much less of a contribution at  $N^3LO$ . This is a feature of the ATLAS datasets included in the fit due to the impact of chosen  $p_T$  cuts which reduce the sensitivity to higher orders.

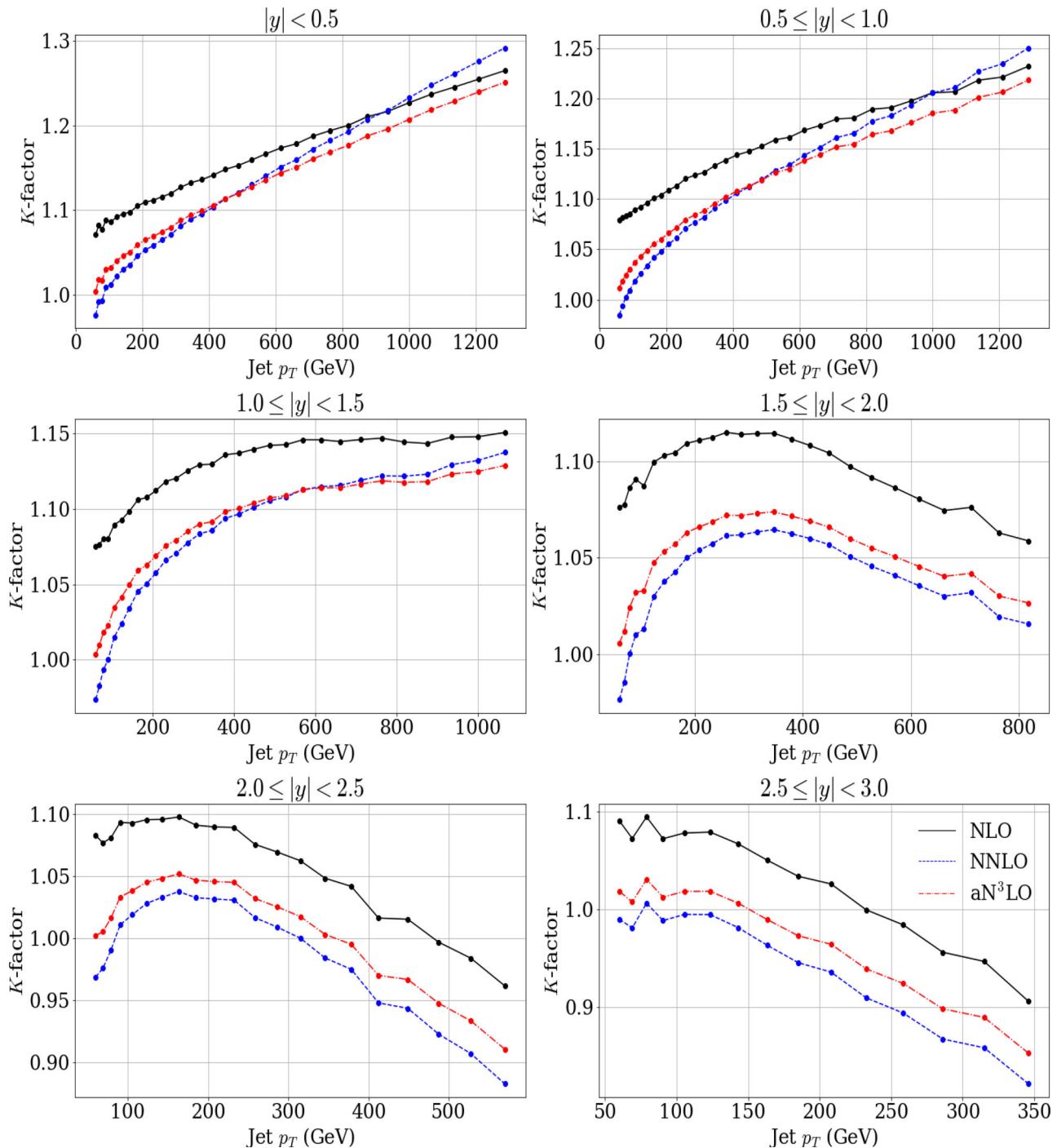
Table 2 demonstrates that in most cases, the new fitted DY  $aN^3LO$   $K$ -factors are producing a slightly better fit with a moderate cumulative effect. We remind the reader that we have included a total of 20 extra parameters into the fit. These extra 20 parameters are fit across all datasets and multiple processes, whereas the decrease here is for a subset of datasets corresponding to the DY processes included in a global fit.

Across these datasets, the  $K$ -factors act to extend the description of these processes to approximate  $N^3LO$ . The result of including this procedure is a better fit in the DY regime while also relaxing tensions with other processes included in the fit. Comparing the  $\Delta\chi^2$  results with and without  $aN^3LO$   $K$ -factors, we can see the extent to which the  $K$ -factors and all other  $N^3LO$  additions are reducing the overall  $\chi^2$ .

In some individual cases, the dataset  $\chi^2$  becomes slightly worse relative to NNLO (e.g. ATLAS 8 TeV High-mass DY [98]), whilst in a few others the  $\chi^2$  improvement upon addition of the  $aN^3LO$  splitting functions, transition matrix elements and coefficient function pieces is seen to deteriorate upon addition of the  $aN^3LO$   $K$ -factors, e.g. such as for the ATLAS 7 TeV high prec.  $W, Z$  [79], which is observed to prefer  $N^3LO$  theory with NNLO  $K$ -factors. The addition of the  $aN^3LO$   $K$ -factors do nonetheless result in a net reduction in  $\chi^2$  and for a large number of cases the  $aN^3LO$   $K$ -factors allow for a slight reduction in the individual  $\chi^2$ . The CMS double diff. Drell–Yan [94] shows a particularly large reduction when these are added on top of the  $aN^3LO$  theory, this is a dataset which shows some tension with the DIS  $N^3LO$  additions which is then eased by the addition of the  $aN^3LO$   $K$ -factors.

#### Jet production processes

The jets processes (all calculated for  $\mu_{r,f} = p_T^{jet}$ ) show a general increase in the  $K$ -factor shifts from NNLO as seen in Fig. 24, which displays the  $K$ -factor expansion up to  $aN^3LO$



**Fig. 24**  $K$ -factor expansion up to  $aN^3LO$  shown for the CMS 7 TeV jets dataset ( $R = 0.7$ ) [101]. The  $K$ -factors shown here are absolute i.e. all with respect to LO ( $K^{N^mLO}/LO \forall m \in \{1, 2, 3\}$ )

for the CMS 7 TeV jets dataset [101]. It is apparent that there is a mild shift to  $N^3LO$  from the NNLO  $K$ -factor. This behaviour follows what one might expect for a perturbative expansion considering the forms of the NLO and NNLO functions.

A  $\chi^2$  summary of the Jets datasets is provided in Table 3. By combining the  $N^3LO$  structure function and DGLAP additions (Sects. 4–6) with NNLO  $K$ -factors, the fit exhibits a substantial increase in the  $\chi^2$  from Jets data. Including  $aN^3LO$   $K$ -factors acts to reduce some of this tension with around half the initial overall  $\chi^2$  increase still remaining. We

**Table 3** Table showing the relevant jet datasets and how the individual  $\chi^2$  changes from NNLO by including the N<sup>3</sup>LO treatment of  $K$ -factors. The result with purely NNLO  $K$ -factors included for all data in the fit is also given

Jets dataset	$\chi^2$	$\Delta\chi^2$ from NNLO	$\Delta\chi^2$ from NNLO (NNLO $K$ -factors)
CDF II $p\bar{p}$ incl. jets [102]	66.5/76	+6.1	+1.7
D $\emptyset$ II $p\bar{p}$ incl. jets [103]	113.3/110	-6.9	-3.0
ATLAS 7 TeV jets [104]	215.9/140	-5.6	+3.7
CMS 7 TeV jets [101]	186.8/158	+11.0	+9.3
CMS 8 TeV jets [105]	271.3/174	+10.0	+21.5
CMS 2.76 TeV jet [106]	109.8/81	+6.9	+9.0
Total	963.6/739	+21.5	+42.2

note that in the case of the ATLAS 7 TeV jets [104], it is well known that there are issues in achieving a good fit quality across all rapidity bins (see [107] for a detailed study as well as [108] where the 8 TeV data are presented and the same issues observed). In [107, 108] the possibility of decorrelating some of the systematic error sources where the degree of correlation is less well established, was considered and indeed in our study we follow such a procedure, as described in [3]. Alternatively, however, it might be that the issues in fit quality could at least part be due to deficiencies in theoretical predictions, such as MHOs. To assess this, we revert to the default ATLAS correlation scenario and repeat the global fit. We find that the  $\chi^2$  deteriorates by +40.7 points to 256.6, which is very close to the result found in a pure NNLO fit [3]. In other words, in our framework the impact of MHOUs does not resolve this issue.

The  $\chi^2$  results for datasets in Table 3 show evidence for some tensions with the N<sup>3</sup>LO form of the high- $x$  gluon. It is also apparent that the CMS data is in more tension than ATLAS datasets with N<sup>3</sup>LO structure function and DGLAP theory. Therefore it will be interesting to see how this behaviour changes when considering this data as dijets in the global fit [109]. We do not consider the dijet data here, though this will be addressed in a future publication.

#### $Z p_T$ and vector boson + jets processes

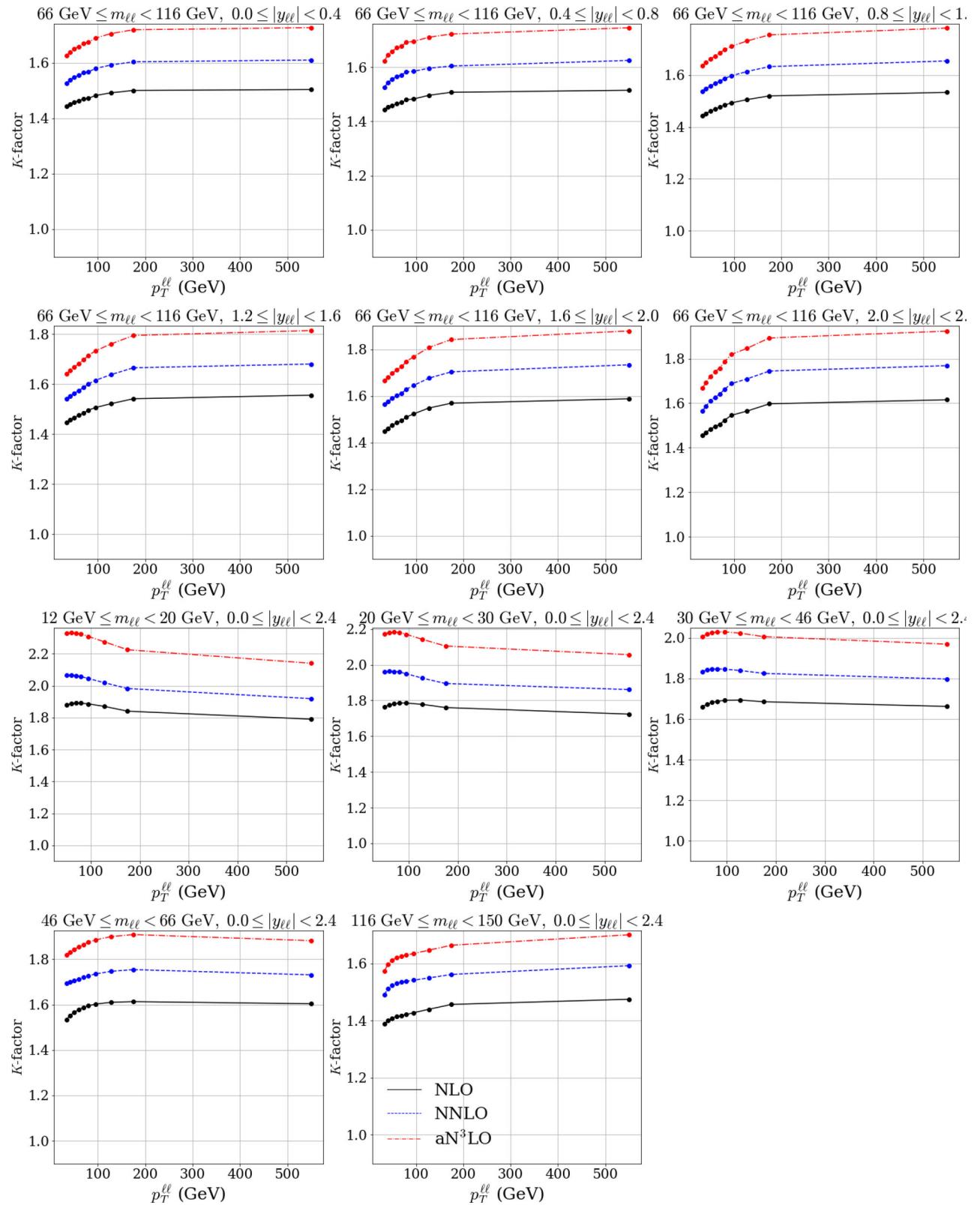
In the case of  $Z p_T$  & vector boson + jet processes (all calculated at  $\mu_{r,f} = \sqrt{p_{T,ll}^2 + m_{ll}^2}$ ), the  $K$ -factor shift is almost completely dominated by the ATLAS 8 TeV  $Z p_T$  dataset [110] (due to the larger number of data points included in this dataset) shown in Fig. 25. The gluon is less directly constrained than the quarks in a global fit. Therefore it can be expected that the significant modifications at small- $x$  will indirectly affect the high- $x$  gluon, where these processes are most sensitive. Considering the jet production processes in Table 3, when performing separate PDF fits not including ATLAS 8 TeV  $Z p_T$  data [110], we find a reduction of  $\Delta\chi^2 = -7.0$  in CMS 8 TeV jets data [105] eliminating most of the tension for this dataset (similar to MSHT20 NNLO results in Table 16 of [3]). Further to this, when not including HERA

and ATLAS 8 TeV  $Z p_T$  data we find a reduction of  $\Delta\chi^2 = -26.4$  in CMS 8 TeV jet data [105] and  $\Delta\chi^2 = -12.7$  in CMS 2.76 TeV jet data [106].

Although the overall magnitude of the  $K$ -factor in Fig. 25 may seem large, this new shift is contained within a 15% increase from NNLO (due to the NLO and NNLO  $K$ -factors also being significant). Moreover, not only does the size of this shift have some dependence on the central scale, but this shift may be more correctly interpreted as the preferred all-orders cross section rather than simply the pure N<sup>3</sup>LO result.

The extent of the  $\chi^2$  reduction in the  $Z p_T$  datasets is shown in Table 4. Note that around  $\sim 68\%$  of the improvement to the ATLAS 8 TeV  $Z p_T$  [110]  $\chi^2$  is due to the extra N<sup>3</sup>LO theory included in the DGLAP and DIS descriptions. It is also known the ATLAS 8 TeV  $Z p_T$  data [110] previously exhibited a significant level of tension with many datasets (including HERA data) at NNLO [3]. This was investigated by performing a global PDF fit with and without HERA data and comparing the individual  $\chi^2$ 's from each dataset. At NNLO it was found that the ATLAS 8 TeV  $Z p_T$  dataset [110] reduced by  $\Delta\chi^2 = -39.2$  when fitting to all non-HERA data (see Table 19). At aN<sup>3</sup>LO we observe that the ATLAS 8 TeV  $Z p_T$  dataset [110] actually increased by  $\Delta\chi^2 = +12.8$  when fitting to all non-HERA data (see Table 20). The aN<sup>3</sup>LO additions therefore eliminate this tension previously observed at NNLO, suggesting that this issue at NNLO in fitting the ATLAS 8 TeV  $Z p_T$  dataset [110] was a sign of MHOs. This is in contrast with the result observed for ATLAS 7 TeV jets [104] where the issues with fit quality were not alleviated by the inclusion of known higher order N3LO information and approximations for the remaining missing pieces.

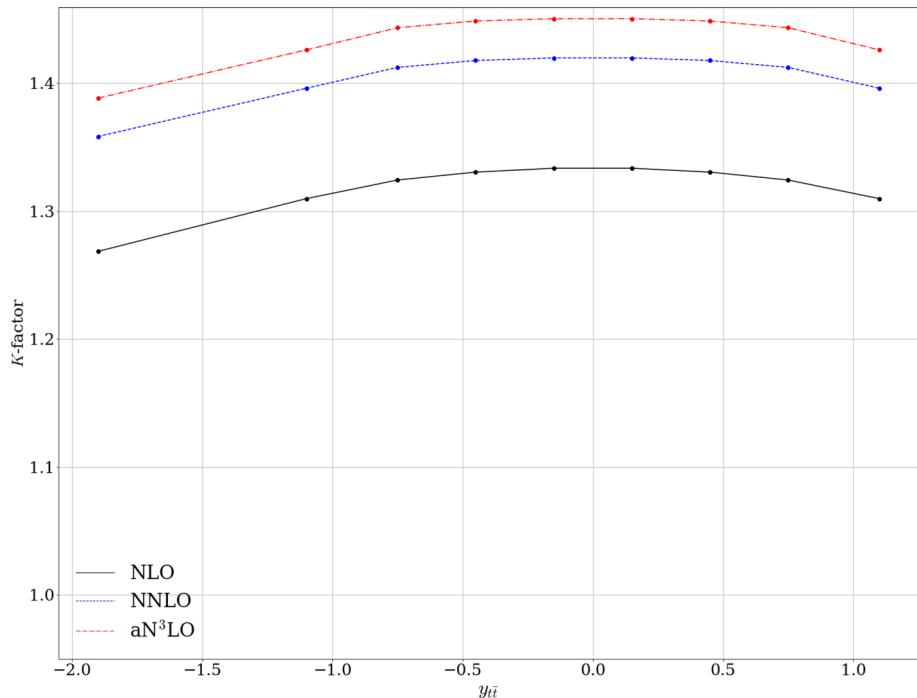
Finally we remind the reader that the CMS 7 TeV  $W + c$  dataset [62] does not include a  $K$ -factor at NNLO. To overcome this, we tie the overall N<sup>3</sup>LO  $K$ -factor shift to the NLO value ( $K^{\text{NNLO}/\text{NLO}} = 1$  in Eq. (7.12)), therefore contributing as an overall normalisation effect.



**Fig. 25**  $K$ -factor expansion up to  $aN^3LO$  shown for the ATLAS 8 TeV  $Z$   $p_T$  dataset [110]. The  $K$ -factors shown here are absolute i.e. all with respect to LO ( $K^{N^mLO/LO} \forall m \in \{1, 2, 3\}$ )

**Table 4** Table showing the relevant  $Z$   $p_T$  and Vector Boson jet datasets and how the individual  $\chi^2$  changes from NNLO by including the  $N^3LO$  treatment of  $K$ -factors, and theoretical  $N^3LO$  additions discussed earlier

$p_T$ Jets dataset	$\chi^2$	$\Delta\chi^2$ from NNLO	$\Delta\chi^2$ from NNLO (NNLO $K$ -factors)
CMS 7 TeV $W + c$ [62]	10.8/10	+2.2	+0.2
ATLAS 8 TeV $Z$ $p_T$ [110]	108.4/104	-80.0	-54.5
ATLAS 8 TeV $W$ + jets [111]	18.8/30	+0.7	-0.3
Total	138.0/144	-77.2	-54.7

**Fig. 26**  $K$ -factor expansion up to  $aN^3LO$  shown for the CMS 8 TeV single diff.  $t\bar{t}$  dataset [112]. The  $K$ -factors shown here are absolute i.e. all with respect to LO ( $K^{N^mLO/LO} \forall m \in \{1, 2, 3\}$ )**Table 5** Table showing the relevant Top Quark datasets and how the individual  $\chi^2$  changes from NNLO by including the  $N^3LO$  treatment of  $K$ -factors, and theoretical  $N^3LO$  additions discussed earlier. The result with purely NNLO  $K$ -factors included for all data in the fit is also given

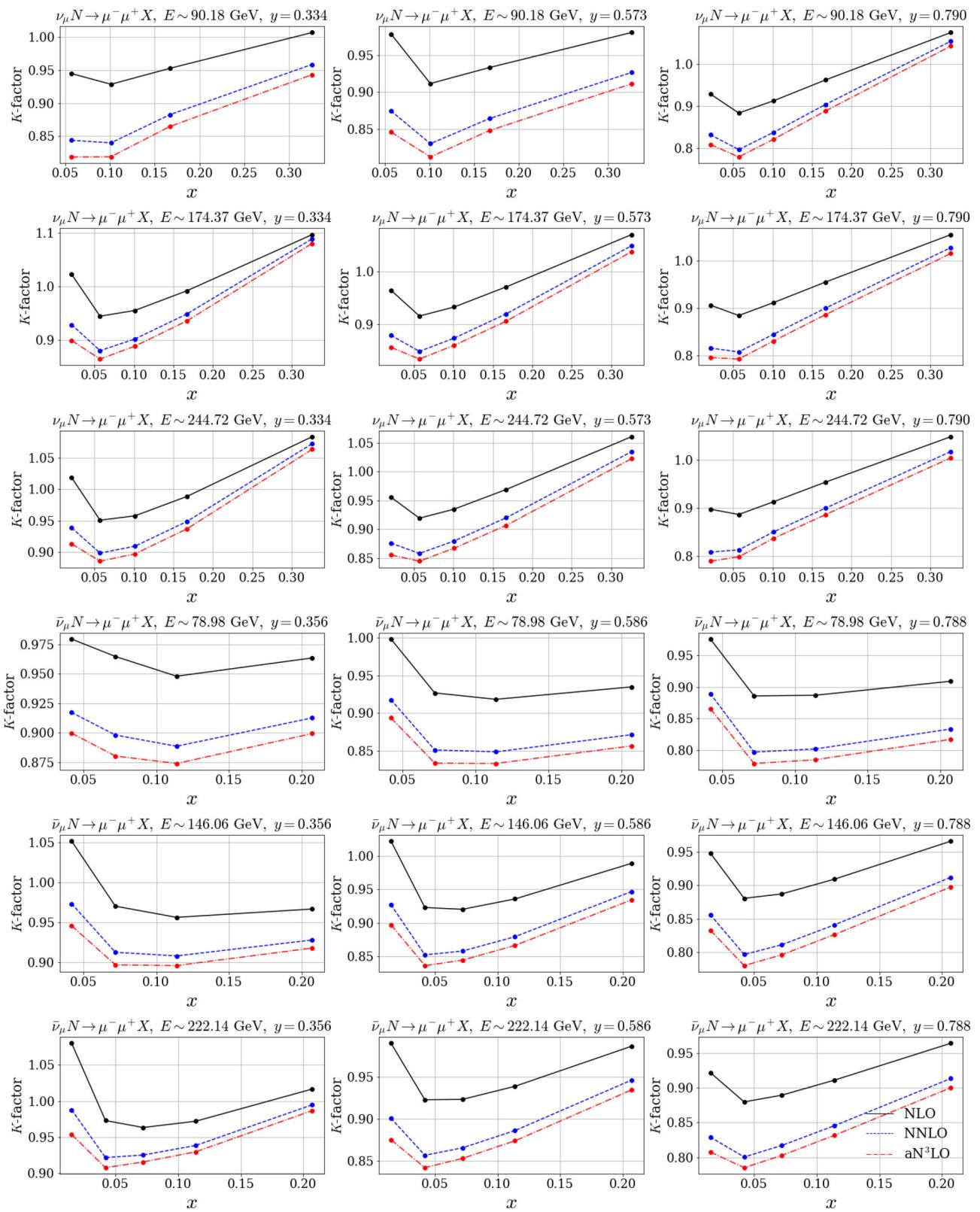
Top dataset	$\chi^2$	$\Delta\chi^2$ from NNLO	$\Delta\chi^2$ from NNLO (NNLO $K$ -factors)
Tevatron, ATLAS, CMS $\sigma_{t\bar{t}}$ [117–129]	14.3/17	-0.2	-1.0
ATLAS 8 TeV single diff. $t\bar{t}$ [130]	24.2/25	-1.4	-0.8
ATLAS 8 TeV single diff. $t\bar{t}$ dilep. [131]	2.7/5	-0.7	-0.7
CMS 8 TeV double diff. $t\bar{t}$ [132]	23.6/15	+1.0	+4.0
CMS 8 TeV single diff. $t\bar{t}$ [112]	10.3/9	-2.9	-4.0
Total	75.1/71	-4.2	-2.5

### Top quark processes

Moving to top quark processes, for the single differential datasets the scale choice for  $\mu_{r,f}$  is  $H_T/4$  with the exception of data differential in the average transverse momentum of the top or anti-top,  $p_T^t$ ,  $p_T^{\bar{t}}$ , for which  $m_T/2$  is used. For the double diff. dataset the scale choice is  $H_T/4$  and for the inclusive top  $\sigma_{t\bar{t}}$  a scale of  $m_t$  is chosen. Figure 26 displays the  $K$ -factor shifts up to  $N^3LO$  for the CMS 8 TeV single diff.  $t\bar{t}$  dataset [112], which shows the greatest reduction in

its  $\chi^2$ . A familiar perturbative pattern can be seen for this process's  $K$ -factors, with the shift at  $aN^3LO$  increasing by around 3–4% from NNLO. This is in agreement with a recent  $\sim 3.5\%$  predicted increase in the  $N^3LO$   $t\bar{t}$  production  $K$ -factor at 8 TeV in [66], whereby an approximate  $N^3LO$  cross section for  $t\bar{t}$  production in proton-proton collisions has been calculated employing a resummation formalism [113–116].

The  $\chi^2$  results in Table 5 display a mildly better fit for top processes, with most datasets not feeling a large overall effect from the  $N^3LO$  additions. Comparing with and without



**Fig. 27**  $K$ -factor expansion up to  $aN^3LO$  shown for the NuTeV  $\nu N \rightarrow \mu\mu X$  dataset [133]. The  $K$ -factors shown here are absolute i.e. all with respect to LO ( $K^{N^mLO}/LO \forall m \in \{1, 2, 3\}$ )

aN<sup>3</sup>LO  $K$ -factors, we see a slightly better fit overall, with most of the reduction in overall  $\chi^2$  stemming from CMS 8 TeV double diff.  $t\bar{t}$  data [132].

### Semi-inclusive DIS dimuon processes

The final set of results to consider in this section are the aN<sup>3</sup>LO  $K$ -factors associated with semi-inclusive DIS dimuon cross sections (with  $\mu_{r,f}^2 = Q^2$ ). Although the dimuon cross section is associated with the DIS process described from our approximate N<sup>3</sup>LO structure function picture, it is a semi-inclusive DIS process. Therefore it is sensible to treat this process as entirely separate from DIS. The NNLO cross-sections used in this case are a general-mass variable flavour number scheme extension of the results in [134], as described in more detail in [3]. The  $K$ -factors shown in Fig. 27 (for the NuTeV  $\nu N \rightarrow \mu\mu X$  data [133]) are somewhat similar to NNLO. The reason for this is mostly due to these datasets also including a branching ratio ( $\text{BR}(D \rightarrow \mu)$ ) which absorbs any overall normalisation shifts. This behaviour is not a concern since in practice these two work in tandem and when combined together it makes no difference where the normalisation factors are absorbed into.

Investigating the change in the BR's with the addition of N<sup>3</sup>LO contributions in Table 6, the BR at N<sup>3</sup>LO decreases substantially from NNLO, with little difference from the addition of aN<sup>3</sup>LO  $K$ -factors. The predicted dimuon BR at aN<sup>3</sup>LO is inside the allowed  $\pm 1\sigma$  range of  $0.092 \pm 0.010$ . When performing a fit with the BR fixed at its central value ( $\text{BR} = 0.092$ ), one is able to observe the effect of manually forcing the normalisation into the  $K$ -factor variation alone. The result of this is a worse global fit quality  $\Delta\chi^2 = +11.2$ , where +3.9 units arise from an increased penalty for the Dimuon  $K$ -factor description and +2.3 units from a slightly worse fit to the Dimuon datasets listed in Table 7. The rest of the observed increase in  $\chi^2_{\text{global}}$  is dominated by a +4.1 increase in the ATLAS 7 TeV high prec.  $W, Z$  [79] data due to a smaller strange quark PDF (compensating the higher BR in dimuon datasets). Returning to consider the case of the  $K$ -factors and BR together, the predicted effect on dimuon datasets is very similar. However due to the errors accounting for a larger allowed shift in the BR relative to the  $K$ -factors, the fit favours moving the BR by a larger amount to reduce the penalty  $\chi^2$  contribution from  $K$ -factors which explains the results shown in Table 6.

Table 7 further confirms the expectation that the Dimuon datasets are not too sensitive to N<sup>3</sup>LO additions. The results with and without a full treatment of aN<sup>3</sup>LO  $K$ -factors are also similar in magnitude. It is therefore clear that the dimuon BR's are compensating for any indirect normalisation effects from the form of the PDFs in the full aN<sup>3</sup>LO fit, as opposed to the aN<sup>3</sup>LO  $K$ -factors.

**Table 6** Table displaying dimuon branching ratios (BRs) at NLO, NNLO, aN<sup>3</sup>LO and aN<sup>3</sup>LO with NNLO  $K$ -factors

	NLO	NNLO	aN <sup>3</sup> LO	aN <sup>3</sup> LO (NNLO $K$ -factors)
$\text{BR}(D \rightarrow \mu)$	0.095	0.088	0.082	0.081

## 8 MSHT20 approximate N<sup>3</sup>LO global analysis

With the inclusion of all N<sup>3</sup>LO approximations discussed in earlier sections resulting in 20 extra free parameters from the NNLO MSHT20 fit, we now present the results for the first approximate N<sup>3</sup>LO global PDF fit with theoretical uncertainties from missing N<sup>3</sup>LO contributions and implicitly some MHOs beyond this. This includes the results for the best fit for the nuisance parameters describing the theoretical uncertainty. We remind the reader that these are parameterised specifically to represent the missing uncertainty at N<sup>3</sup>LO, which is currently the dominant source of uncertainty due to missing higher orders. However, the fit will also be influenced, to a limited extent, by effects at even higher orders. Later in the section we discuss this in more detail.

### 8.1 $\chi^2$ breakdown

Table 8 shows the global  $\chi^2$  results for an aN<sup>3</sup>LO best fit, inclusive of penalties associated with the new theory variational parameters (from Eq. (2.17)). The theory parameters are labelled as:  $A_{Hg}(a_{Hg})$ ,  $A_{gg,H}(a_{gg,H})$ ,  $A_{qq,H}^{\text{NS}}(a_{qq,H}^{\text{NS}})$  for the transition matrix elements;  $P_{qq}^{\text{NS}}(\rho_{qq}^{\text{NS}})$ ,  $P_{qq}^{\text{PS}}(\rho_{qq}^{\text{PS}})$ ,  $P_{qg}(\rho_{qg})$ ,  $P_{gq}(\rho_{gq})$  and  $P_{gg}(\rho_{gg})$  for the splitting functions; and  $c_q^{\text{NLL}}$  and  $c_g^{\text{NLL}}$  correspond to the NLL parameters discussed in Sect. 6.2. These are supplemented by the 10 additional nuisance parameters for the NLO and NNLO  $K$ -factors for the five process categories. These 20 additional parameters and their associated penalties are also shown in Table 8.

The extra N<sup>3</sup>LO theory and level of freedom introduced has allowed the fit to achieve a total  $\Delta\chi^2 = -150.4$  compared to MSHT20 NNLO total  $\chi^2$  (Table 7 from [3]). Comparing with lower order PDF fits, we find a smooth convergence in the fit quality which follows what one may expect from an increase in the accuracy of a perturbative expansion ( $\chi^2/N_{\text{pts}} = \text{LO}: 2.57, \text{NLO}: 1.33, \text{NNLO}: 1.17, \text{N}^3\text{LO}: 1.14$ ). In part, this is due to the extra freedom in the  $K$ -factors, which will almost always act to reduce this  $\chi^2$  due to the minimisation procedure. However, even with this freedom, in most cases the N<sup>3</sup>LO theory (non  $K$ -factor) contributions include large divergences from NNLO. With this in mind, we must conclude that the fit is preferring a description different from the current NNLO standard.

At NNLO (Table 19), the tension between HERA and non-HERA datasets accounted for  $\Delta\chi^2 = -61.6$  reduction in the overall fit quality when the former was removed,

**Table 7** Table showing the relevant Dimuon datasets and how the individual  $\chi^2$  changes from NNLO by including the N<sup>3</sup>LO treatment of  $K$ -factors, and theoretical N<sup>3</sup>LO additions discussed earlier. The result with purely NNLO  $K$ -factors included for all data in the fit is also given

Dimuon dataset	$\chi^2$	$\Delta\chi^2$ from NNLO	$\Delta\chi^2$ from NNLO (NNLO $K$ -factors)
CCFR $\nu N \rightarrow \mu\mu X$ [133]	68.3/86	+0.6	+1.1
NuTeV $\nu N \rightarrow \mu\mu X$ [133]	56.7/84	-1.8	-0.6
Total	125.0/170	-1.2	+0.5

**Table 8** Full breakdown of  $\chi^2$  results for the aN<sup>3</sup>LO PDF fit. The global fit includes the N<sup>3</sup>LO treatment for transition matrix elements, coefficient functions, splitting functions and  $K$ -factor additions with their variational parameters determined by the fit

Dataset	$N_{\text{pts}}$	$\chi^2$	$\Delta\chi^2$ from NNLO
BCDMS $\mu p$ $F_2$ [135]	163	174.4	-5.8
BCDMS $\mu d$ $F_2$ [135]	151	144.3	-1.7
NMC $\mu p$ $F_2$ [136]	123	121.5	-2.6
NMC $\mu d$ $F_2$ [136]	123	104.2	-8.4
SLAC $ep$ $F_2$ [137, 138]	37	31.6	-0.4
SLAC $ed$ $F_2$ [137, 138]	38	22.8	-0.2
E665 $\mu d$ $F_2$ [139]	53	63.9	+4.2
E665 $\mu p$ $F_2$ [139]	53	67.5	+2.9
NuTeV $\nu N$ $F_2$ [140]	53	35.7	-2.6
NuTeV $\nu N$ $x F_3$ [140]	42	34.8	+4.1
NMC $\mu n/\mu p$ [141]	148	131.6	+0.8
E866/NuSea $pp$ DY [80]	184	215.4	-9.7
E866/NuSea $pd/pp$ DY [81]	15	8.4	-2.0
HERA $ep$ $F_2^{\text{charm}}$ [142]	79	143.7	+11.4
NMC/BCDMS/SLAC/HERA $F_L$ [135, 136, 138, 143–145]	57	45.6	-22.9
CCFR $\nu N \rightarrow \mu\mu X$ [133]	86	68.3	+0.6
NuTeV $\nu N \rightarrow \mu\mu X$ [133]	84	56.7	-1.8
CHORUS $\nu N$ $F_2$ [146]	42	29.2	-1.0
CHORUS $\nu N$ $x F_3$ [146]	28	18.1	-0.3
HERA $e^+ p$ CC [147]	39	49.7	-2.3
HERA $e^- p$ CC [147]	42	64.9	-5.3
HERA $e^+ p$ NC 820 GeV [147]	75	84.3	-5.6
HERA $e^- p$ NC 460 GeV [147]	209	247.7	-0.6
HERA $e^+ p$ NC 920 GeV [147]	402	474.0	-38.7
HERA $e^- p$ NC 575 GeV [147]	259	248.5	-14.5
HERA $e^- p$ NC 920 GeV [147]	159	243.0	-1.4
CDF II $p\bar{p}$ incl. jets [102]	76	66.5	+6.1
DØ II $Z$ rap. [82]	28	17.4	+1.0
CDF II $Z$ rap. [83]	28	40.6	+3.4
DØ II $W \rightarrow \nu\mu$ asym. [84]	10	16.5	-0.8
CDF II $W$ asym. [85]	13	18.2	-0.8
DØ II $W \rightarrow ve$ asym. [86]	12	30.7	-3.2
DØ II $p\bar{p}$ incl. jets [103]	110	113.3	-6.9

**Table 8** continued

Dataset	$N_{\text{pts}}$	$\chi^2$	$\Delta\chi^2$ from NNLO
ATLAS $W^+, W^-, Z$ [87]	30	30.0	+0.1
CMS W asym. $p_T > 35$ GeV [88]	11	6.7	-1.1
CMS W asym. $p_T > 25, 30$ GeV [89]	24	7.9	+0.5
LHCb $Z \rightarrow e^+e^-$ [90]	9	23.2	+0.5
LHCb W asym. $p_T > 20$ GeV [91]	10	12.6	+0.1
CMS $Z \rightarrow e^+e^-$ [92]	35	17.6	-0.4
ATLAS high-mass Drell–Yan [93]	13	18.4	-0.5
Tevatron, ATLAS, CMS $\sigma_{t\bar{t}}$ [117–129]	17	14.3	-0.2
CMS double diff. Drell–Yan [94]	132	133.2	-11.3
LHCb 2015 $W, Z$ [77, 78]	67	103.2	+3.8
LHCb 8 TeV $Z \rightarrow ee$ [95]	17	30.3	+4.1
CMS 8 TeV $W$ [96]	22	11.5	-1.2
ATLAS 7 TeV jets [104]	140	215.9	-5.6
CMS 7 TeV $W + c$ [62]	10	10.8	+2.2
ATLAS 7 TeV high prec. $W, Z$ [79]	61	119.3	+2.7
CMS 7 TeV jets [101]	158	186.8	+11.0
DØ $W$ asym. [97]	14	12.2	+0.1
ATLAS 8 TeV $Z$ $p_T$ [110]	104	108.4	-80.0
CMS 8 TeV jets [105]	174	271.3	+10.0
ATLAS 8 TeV sing. diff. $t\bar{t}$ [130]	25	24.2	-1.4
ATLAS 8 TeV sing. diff. $t\bar{t}$ dilep. [131]	5	2.7	-0.7
ATLAS 8 TeV high-mass DY [98]	48	62.8	+5.7
ATLAS 8 TeV $W +$ jets [111]	30	18.8	+0.7
CMS 8 TeV double diff. $t\bar{t}$ [132]	15	23.6	+1.0
ATLAS 8 TeV $W$ [99]	22	53.0	-4.4
CMS 2.76 TeV jet [106]	81	109.8	+6.9
CMS 8 TeV sing. diff. $t\bar{t}$ [112]	9	10.3	-2.9
ATLAS 8 TeV double diff. $Z$ [100]	59	80.4	+4.5
Low- $Q^2$ coefficient			
$c_q^{\text{NLL}} = -3.868$	0.004	$c_g^{\text{NLL}} = -5.837$	0.844
Transition matrix elements			
$a_{Hg} = 12214.000$	0.601	$a_{qg,H}^{\text{NS}} = -64.411$	0.001
$a_{gg,H} = -1951.600$	0.857		
Splitting functions			
$\rho_{qg}^{\text{NS}} = 0.007$	0.000	$\rho_{gg} = -1.784$	0.802
$\rho_{qg}^{\text{PS}} = -0.501$	0.186	$\rho_{gg} = 19.245$	3.419
$\rho_{qg} = -1.754$	0.015		
K-factors			
$\text{DY}_{\text{NLO}} = -0.307$	0.094	$\text{DY}_{\text{NNLO}} = -0.230$	0.053
$\text{Top}_{\text{NLO}} = 0.041$	0.002	$\text{Top}_{\text{NNLO}} = 0.651$	0.424
$\text{Jet}_{\text{NLO}} = -0.300$	0.090	$\text{Jet}_{\text{NNLO}} = -0.691$	0.478
$p_T \text{Jets}_{\text{NLO}} = 0.583$	0.339	$p_T \text{Jets}_{\text{NNLO}} = -0.080$	0.006
$\text{Dimuon}_{\text{NLO}} = -0.444$	0.197	$\text{Dimuon}_{\text{NNLO}} = 0.922$	0.850
$\text{N}^3\text{LO}$ penalty total	9.262/20	Average penalty	0.463
		Total	4961.2/4363
		$\Delta\chi^2$ from NNLO	-150.4

with the majority of this tension between HERA and ATLAS 8 TeV  $Z$   $p_T$  [110] data. Whereas comparing fit results with and without HERA data at  $N^3LO$ , we find  $\Delta\chi^2 = -48.0$ . Although the overall difference is not too substantial we do report a substantial shift in the leading tensions, where most of the tension with HERA data is now residing with NMC  $F_2$  [136] and CMS 8 TeV jets [105] data. Tensions with NMC  $F_2$  [136] data are also seen to some extent at NNLO where we show a  $\Delta\chi^2 = -20.6$  in a fit omitting HERA data (combining the NMC  $F_2$  datasets shown in Table 19). However at  $N^3LO$ , Table 20 shows a  $\Delta\chi^2 = -23.4$  reduction from NMC  $F_2$  data in a fit omitting HERA data. Therefore whilst the  $N^3LO$  additions remove tensions with  $Z$   $p_T$  data, it remains that the HERA data is preferring the high- $x$  quarks to be lower than favoured by NMC data. This is suggestive of higher twist effects for NMC data at low- $Q^2$  (as we observe a worse fit to low- $Q^2$  data). We also emphasise that when conducting a fit at NNLO with  $Z$   $p_T$  data removed, an improvement of  $\Delta\chi^2 = -41.3$  is observed in the rest of the data, whereas at  $N^3LO$  an improvement of  $\Delta\chi^2 = -69.3$  is observed in all other datasets without removing  $Z$   $p_T$ , therefore these results are not purely an effect of removing any  $Z$   $p_T$  tension. Considering tensions with CMS 8 TeV jets [105] data, as discussed in Sect. 7, in general the jets datasets show tensions with the  $N^3LO$  description (especially for CMS 8 TeV jets [105]), therefore it will be interesting to observe how this picture evolves when considering this data in the form of dijets.

Since a naturally richer description of the small- $x$  regime is being included at  $N^3LO$ , which has a direct effect on the HERA datasets, the reduction of important tensions from NNLO is even further justification for the inclusion of the  $N^3LO$  theory. The extra  $N^3LO$  additions are allowing the large- $x$  behaviour of the PDFs to be less dominated by data at small- $x$ , while also producing a better fit quality at small- $x$  (i.e. for HERA data). Some of the above observations are also made in [16, 17] where studies of including small- $x$  resummation results into a PDF fit have been reported.

Reflecting on the chosen prior distributions for each of the sources of missing  $N^3LO$  uncertainty, Table 8 confirms that no especially large penalties are being incurred in this new description. These results therefore demonstrate that the fit is succeeding in leveraging contributions (such as  $P_{qq}^{(3)}$  and  $P_{qg}^{(3)}$  in the quark evolution part of Eq. (3.3)) to produce a better overall fit.

### DIS processes

To complement the discussions in Sect. 7, we isolate the  $\chi^2$  results from DIS data in Table 9. This data is directly affected by the  $N^3LO$  structure functions constructed approximately in Sects. 3–6. A substantial decrease in the total  $\chi^2$

from NNLO is observed across DIS datasets. Considering the results in Table 9 in the context of Tables 2, 3, 4, 5, 6 and 7, a better fit quality is observed for all DIS and non-DIS datasets than at NNLO with the inclusion of  $N^3LO$  contributions. As the DIS data makes up over half of the total data included in a global fit, it is the dominant force in deciding the overall form of the PDFs, especially at small- $x$  (discussed further in Sect. 8.4). Table 9 further reinforces the point that the  $N^3LO$  description is flexible enough to fit to HERA and non-HERA data, without being largely constrained by tensions between the small- $x$  (HERA dominated) and large- $x$  (non-HERA dominated) regions.

### 8.2 Correlation results

The correlation matrix shown in Fig. 28 illustrates the correlations between extra  $N^3LO$  theory parameters and the subset of the MSHT20 parameters which are included in the construction of Hessian eigenvectors (see Sect. 8.3 and [3] for details). It is apparent that the correlations between  $K$ -factor parameters for each process (shown in green) and other PDF and theory parameters are usually small, with some exceptions e.g. for the TopNLO parameter. Due to this there is an argument that each process'  $K$ -factor parameters could be treated separately from all other parameters in the Hessian prescription (see Sect. 4.1) which allows for a more flexible PDF set that can be decorrelated from a process. By using the uncorrelated Hessian results for a process NNLO hard cross sections can be transformed to a  $N^3LO$  and therefore provide more reliable predictions (more details in Sect. 10). This is a fairly intuitive result, since most correlations are showing a natural separation between the process dependent and process independent physics in the DIS picture.<sup>18</sup> Mathematically, the  $K$ -factors are directly associated with the hard cross section, whereas other  $N^3LO$  theory parameters ( $\rho_{ij}$  and  $a_{ij}$ ) are having a direct effect on the PDFs. Figure 28 therefore begins to motivate the inclusion of the ‘pure’ theory (splitting functions and transition matrix elements) parameters within the standard MSHT eigenvector analysis [3], with the decorrelation of the  $K$ -factor parameters, as discussed in Sect. 2.3. We investigate and compare both treatments (complete correlation and  $K$ -factor decorrelation) throughout the rest of this section. We show in Sect. 8.4 that while the decorrelation of  $K$ -factors is not complete, both treatments result in similar uncertainty bands, therefore confirming that the effect of making the assumption of full decorrelation is minimal in practice. Note that although the  $c_i^{\text{NLL}}$  parameters also show minimal correlation with other parameters, we include these within the ‘pure’ theory group of parameters (i.e. cor-

<sup>18</sup> The same pattern can be seen for  $c_{\{q,g\}}^{\text{NLL}}$  parameters which are involved in the DIS hard cross section.

**Table 9** Table showing the relevant DIS datasets and how the individual  $\chi^2$  changes from NNLO by including the  $N^3LO$  contributions to the structure function  $F_2(x, Q^2)$ . The result within purely NNLO  $K$ -factors included for all data in the fit is also given

DIS dataset	$\chi^2$	$\Delta\chi^2$ from NNLO	$\Delta\chi^2$ from NNLO (NNLO $K$ -factors)
BCDMS $\mu p$ $F_2$ [135]	174.4/163	-5.8	-5.6
BCDMS $\mu d$ $F_2$ [135]	144.3/151	-1.7	+0.6
NMC $\mu p$ $F_2$ [136]	121.5/123	-2.6	-3.8
NMC $\mu d$ $F_2$ [136]	104.2/123	-8.4	-11.5
SLAC $ep$ $F_2$ [137, 138]	31.6/37	-0.4	-0.0
SLAC $ed$ $F_2$ [137, 138]	22.8/38	-0.2	-0.8
E665 $\mu p$ $F_2$ [139]	63.9/53	+4.2	+4.5
E665 $\mu d$ $F_2$ [139]	67.5/53	+2.9	+2.5
NuTeV $\nu N$ $F_2$ [140]	35.7/53	-2.6	-1.3
NuTeV $\nu N$ $x F_3$ [140]	34.8/42	+4.1	+2.1
NMC $\mu n/\mu p$ [141]	131.6/148	+0.8	+2.2
HERA $ep$ $F_2^{\text{charm}}$ [142]	143.7/79	+11.4	+13.8
NMC/BCDMS/SLAC/HERA $F_L$ [135, 136, 138, 143–145]	45.6/57	-22.9	-23.2
CHORUS $\nu N$ $F_2$ [146]	29.2/42	-1.0	-0.8
CHORUS $\nu N$ $x F_3$ [146]	18.1/28	-0.3	-0.5
HERA $e^+ p$ CC [147]	49.7/39	-2.3	-1.0
HERA $e^- p$ CC [147]	64.9/42	-5.3	-4.9
HERA $e^+ p$ NC 820 GeV [147]	84.3/75	-5.6	-5.1
HERA $e^- p$ NC 460 GeV [147]	247.7/209	-0.6	-0.7
HERA $e^+ p$ NC 920 GeV [147]	474.0/402	-38.7	-36.4
HERA $e^- p$ NC 575 GeV [147]	248.5/259	-14.5	-14.1
HERA $e^- p$ NC 920 GeV [147]	243.0/159	-1.4	-2.1
Total	2580.9/2375	-90.8	-86.2

related with  $\rho_{ij}$  and  $a_{ij}$ ) as they are essential ingredients in the underlying DIS theory.

### 8.3 Eigenvector results

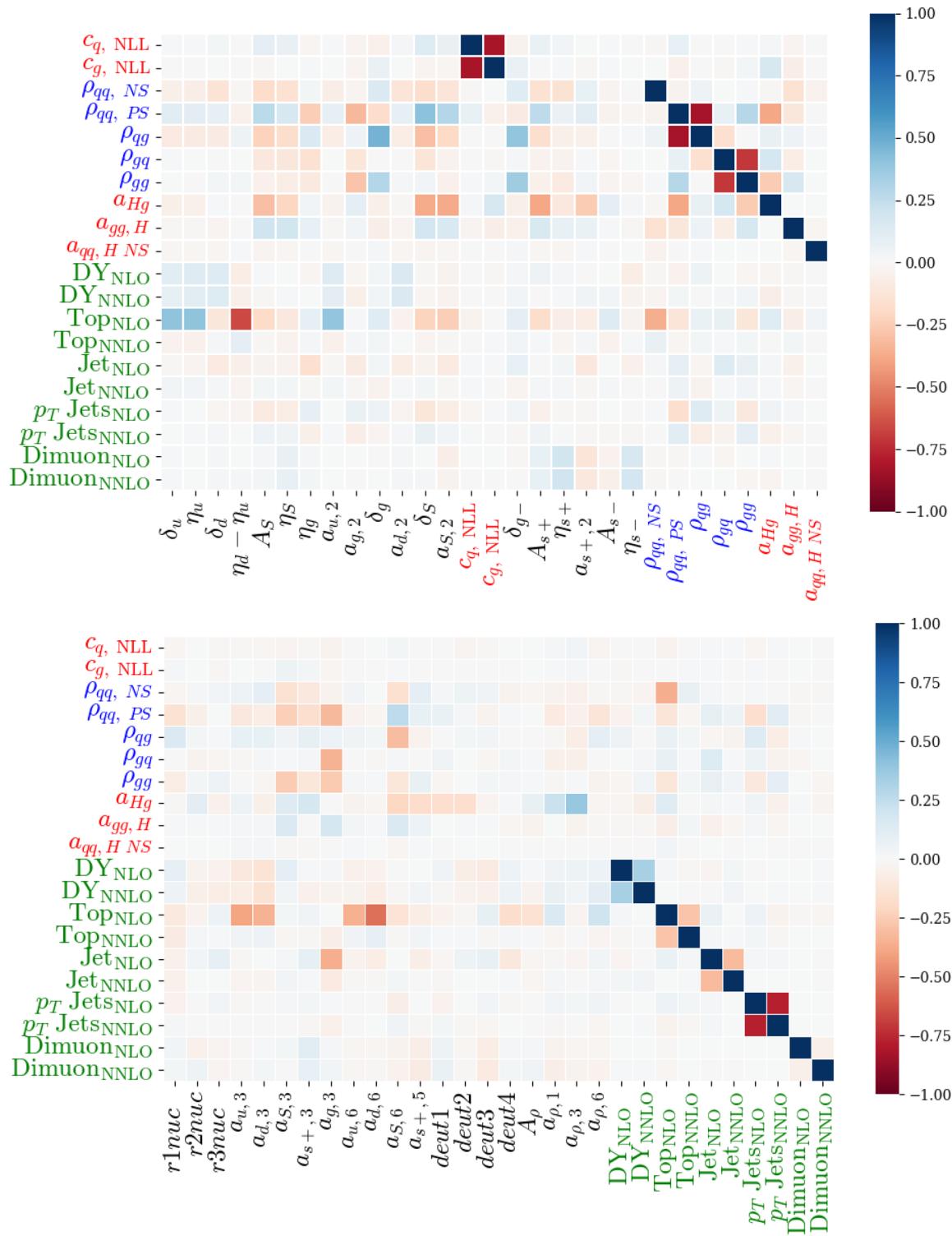
In the MSHT fitting procedure (described in [3]) the eigenvectors of a Hessian matrix are found, which encapsulate the sources of uncertainties and corresponding correlations. Combining these with the central PDFs, forms the entire PDF set with uncertainties. In this eigenvector analysis a dynamical rescaling of each eigenvector  $e_i$  is performed via a tolerance factor  $t$  to encapsulate the 68% confidence limit (C.L.).

$$a_i = a_i^0 \pm t e_i, \quad (8.1)$$

where  $a_i^0$  is the best fit parameter.  $t$  is then adjusted to give the desired tolerance  $T$  for the required confidence interval defined as  $T = \sqrt{\Delta\chi^2_{\text{global}}}$  (for 68% C.L.). In a quadratic approximation, for suitably well-behaved eigenvectors,  $t = T$  is true. Although for eigenvectors with larger eigenvalues, it is possible to observe significant deviations from  $t = T$ . The standard MSHT fitting procedure involves allowing all relevant parameters from [3] to vary when find-

ing the best fit, now including all  $N^3LO$  theory parameters ( $\rho_{ij}$ ,  $a_{ij}$ ,  $c_i^{\text{NLL}}$ ,  $K_{\text{NLO/NNLO}}$ ) discussed in this work. After accounting for high degrees of correlation between parameters (described in [42]), the result is a Hessian matrix which in general, depends on a subset of the parameters that were allowed to vary in a best fit and provides a set of suitably well-behaved eigenvectors. The standard MSHT NNLO PDF eigenvectors are based on a set of 32 parameters, reduced from the 52 parameters allowed to vary in the full fit. In the following analysis we are therefore concerned with a smaller number of parameters, specifically the 32 parameters from the standard MSHT fitting procedure plus an extra 20  $N^3LO$  parameters (shown in Fig. 28).

A standard choice of tolerance  $T$  is  $T = \sqrt{\Delta\chi^2_{\text{global}}} = 1$  for a 68% C.L. limit. However, this assumes all datasets are consistent with Gaussian errors. In practice, due to incomplete theory, tensions between datasets and parameterisation inflexibility, this is known not to be the case in a global PDF fit. To overcome this, a 68% C.L. region for each dataset is defined. Then for each eigenvector, the value of  $\sqrt{\Delta\chi^2_{\text{global}}}$  for each chosen  $t$  is recorded (ideally showing a quadratic behaviour). Finally, a value of  $T$  is chosen to ensure that all



**Fig. 28** Correlation matrix for all  $N^3LO$  theory parameters included in the fit against the subset of the MSHT20 parameters (shown in black) used in constructing the Hessian eigenvectors. This is shown for the case where the  $K$ -factors correlations with the first 42 parameters are

included.  $N^3LO$  theory parameters associated with the splitting functions are coloured blue, the parameters affecting the transition matrix elements and coefficient functions are in red and the  $K$ -factor parameters are in green

datasets are described within their 68% CL in each eigenvector direction. For a fuller mathematical description of the dynamical tolerance procedure used in MSHT PDF fits, the reader is referred to [42]. In this section we present a demonstration of how well the resultant eigenvectors follow the quadratic assumption based on  $t = T$ , including the specific choices of dynamical tolerances and which dataset/penalty constrains this tolerance in each eigenvector direction.

#### PDF + $N^3\text{LO}$ DIS theory + $N^3\text{LO}$ $K$ -factor (decorrelated) parameters

As discussed in Sect. 8.2, when determining the eigenvectors and therefore PDF uncertainties, we can choose to either include the correlations between the 10  $K$ -factor parameters added with the other 42 parameters (encompassing the standard 32 MSHT eigenvector parameters and the 10 new theory parameters from the splitting functions, transition matrix elements and coefficient functions) or to decorrelate the 10  $K$ -factor parameters.

In this section we address the scenario where we decorrelate the  $K$ -factors as

$$H_{ij}^{-1} + \sum_{p=1}^{N_p} K_{ij,p}^{-1} \quad (8.2)$$

and consider each term individually.

Figure 29 shows the map of eigenvectors produced from  $H_{ij}$  in Eq. (8.2), where we have included the new  $N^3\text{LO}$  DIS theory parameters (splitting functions in blue and coefficient functions/transition matrix elements in red) correlated with the PDF parameters. Eigenvectors 35 and 36 are prime examples of where the eigenvectors have specifically encompassed the correlation/anti-correlation between the two NLL FFNS coefficient function parameters  $c_i^{\text{NLL}}$  ( $i \in \{q, g\}$ ). Whereas the splitting functions naturally give rise to a much more complicated mixing with other PDF parameters as these directly affect the evolution of the PDFs. Due to the direct impact of  $\rho_{ij}$ 's on the PDFs (via DGLAP evolution), combined with the large contributions to the evolution shown at  $N^3\text{LO}$ , this result is as expected.

Another somewhat pleasing aspect is the recovery of a natural separation between eigenvectors associated with the  $N^3\text{LO}$  coefficient function/transition matrix elements and our original PDF parameters (incl.  $N^3\text{LO}$  splitting functions). This separation is reminiscent of our DIS picture, whereby the splitting functions are much more intertwined with the raw PDFs and the transition matrix elements have a symbiotic relationship with the coefficient functions (see GM-VFNS description in Sect. 3). Due to this, the form of these eigenvectors has not only some level of physical interpretation inherited from our underlying theory, but also offers a

useful way to access the different sources of  $N^3\text{LO}$  additions within the PDF set.

In Fig. 30 the eigenvectors resulting from the  $\sum_{p=1}^{N_p} K_{ij,p}^{-1}$  terms in Eq. (8.2) are shown. These eigenvectors are constructed in pairs, describing the correlation and anti-correlation of the two  $K$ -factor parameters (controlling the NLO and NNLO contributions to  $N^3\text{LO}$ ) for each process  $p$  contained within the corresponding  $K_{ij,p}$  correlation matrix.

Table 10 shows further information regarding the  $K$ -factor parameter limits from each eigenvector. In most cases the parameter limits are well within the allowed variation ( $-1 < a < 1$ ), which is an indication that the data included in the fit is constraining these parameters rather than the individual penalties for each parameter.<sup>19</sup>

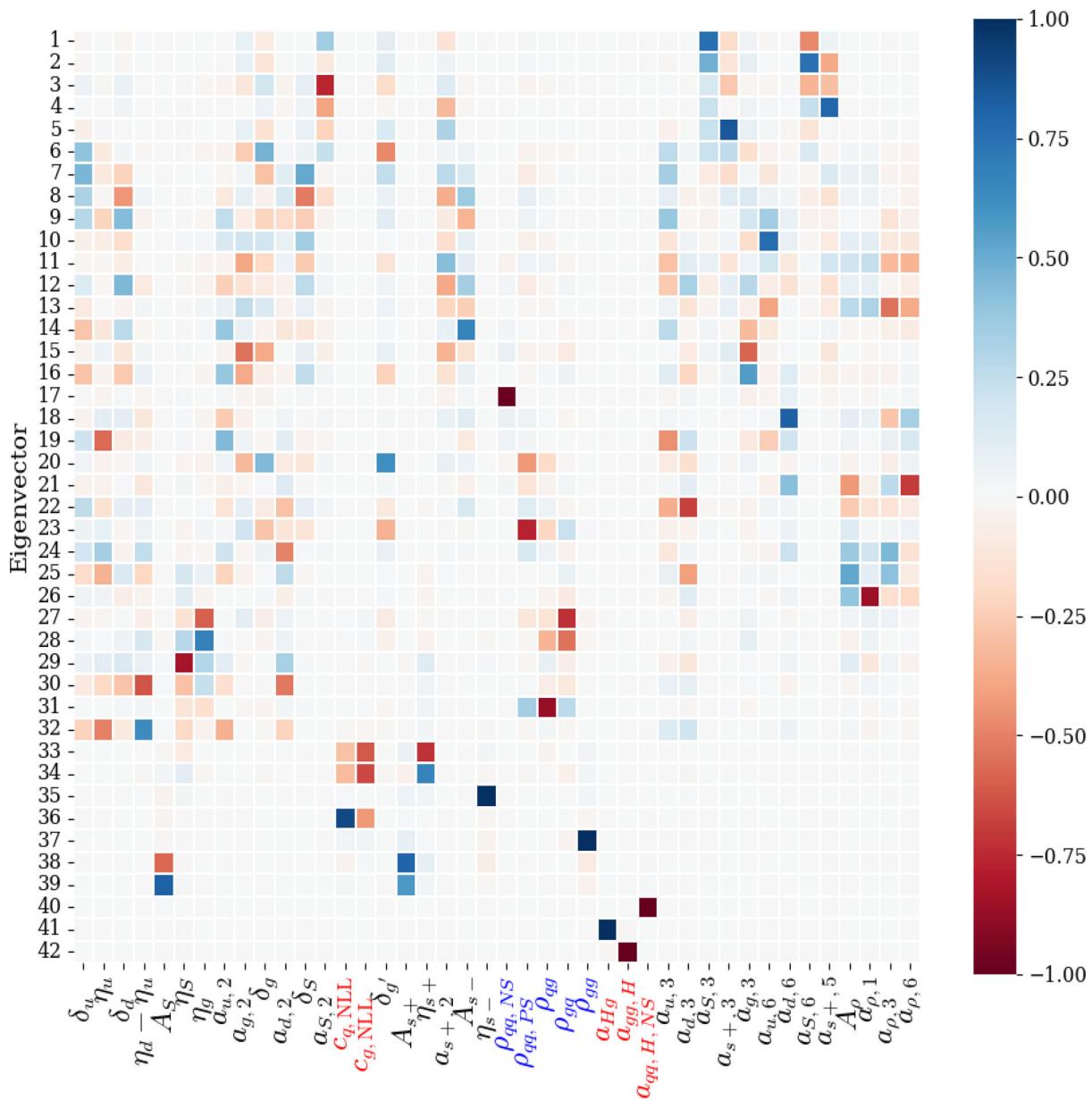
To assess whether the eigenvectors are violating the quadratic treatment, four examples displaying this behaviour are shown in Fig. 31, with a full analysis provided in Appendix 1. Additionally, Table 11 provides a summary of all tolerances found within the eigenvector scans.

There is relatively consistent agreement between  $t$  and  $T$  across all eigenvectors with later eigenvectors (i.e. higher #) generally becoming less quadratic (a feature which is built into the fit). Eigenvectors 31, 41 and 42 displayed in Fig. 31 are shown in Table 11 to be either dominated or limited by at least one new  $N^3\text{LO}$  parameter. Conversely, eigenvector 26 is much more dominated by the original PDF parameters from MSHT20 NNLO. Comparing these cases, the eigenvectors associated more strongly with the  $N^3\text{LO}$  parameters exhibit a similar level of agreement (and occasionally better) with the desired quadratic behaviour as eigenvectors more closely associated with the original PDF parameters.

The last 5 sets of eigenvectors (i.e. the last 10 where a set contains 2 eigenvectors for a particular process) we see in Table 11 are the decorrelated  $K$ -factor eigenvectors, where there are correlated/anti-correlated eigenvectors for each process. For all  $K$ -factor cases, Table 11 provides sensible results with either the dominant datasets or parameter penalties constraining each eigenvector direction. One interesting feature one can observe here is a sign of tension between the ATLAS 8 TeV  $Z$   $p_T$  [110] and ATLAS 8 TeV  $W + \text{jets}$  [111] datasets where the limiting factors in Table 11 for eigenvectors 49 and 50 show that these datasets are preferring a slightly different  $K$ -factor.

To provide some extra level of comparison between the eigenvectors shown here and the eigenvectors found in the NNLO case, the average tolerance  $T$  for a  $N^3\text{LO}$  (decorrelated  $K$ -factors) set is 3.34, compared to the NNLO average  $T$  of 3.37.

<sup>19</sup> We remind the reader that the Dimuon datasets also include a branching ratio factor which is providing some compensation with these  $K$ -factor parameters (as discussed in Sect. 7).



**Fig. 29** Correlation matrix of the first 42 (total 52) eigenvectors found with the  $N^3LO$  parameters added into the analysis in the case where the  $K$ -factors are decorrelated from these first 42 parameters. Parameters

associated with the splitting functions are coloured blue, those affecting the transition matrix elements and coefficient functions are in red

#### PDF + $N^3LO$ DIS theory + $N^3LO$ $K$ -factor (correlated) parameters

In this section we address the scenario,

$$H'_{ij} = \left( H_{ij}^{-1} + \sum_{p=1}^{N_p} K_{ij,p}^{-1} \right)^{-1}. \quad (8.3)$$

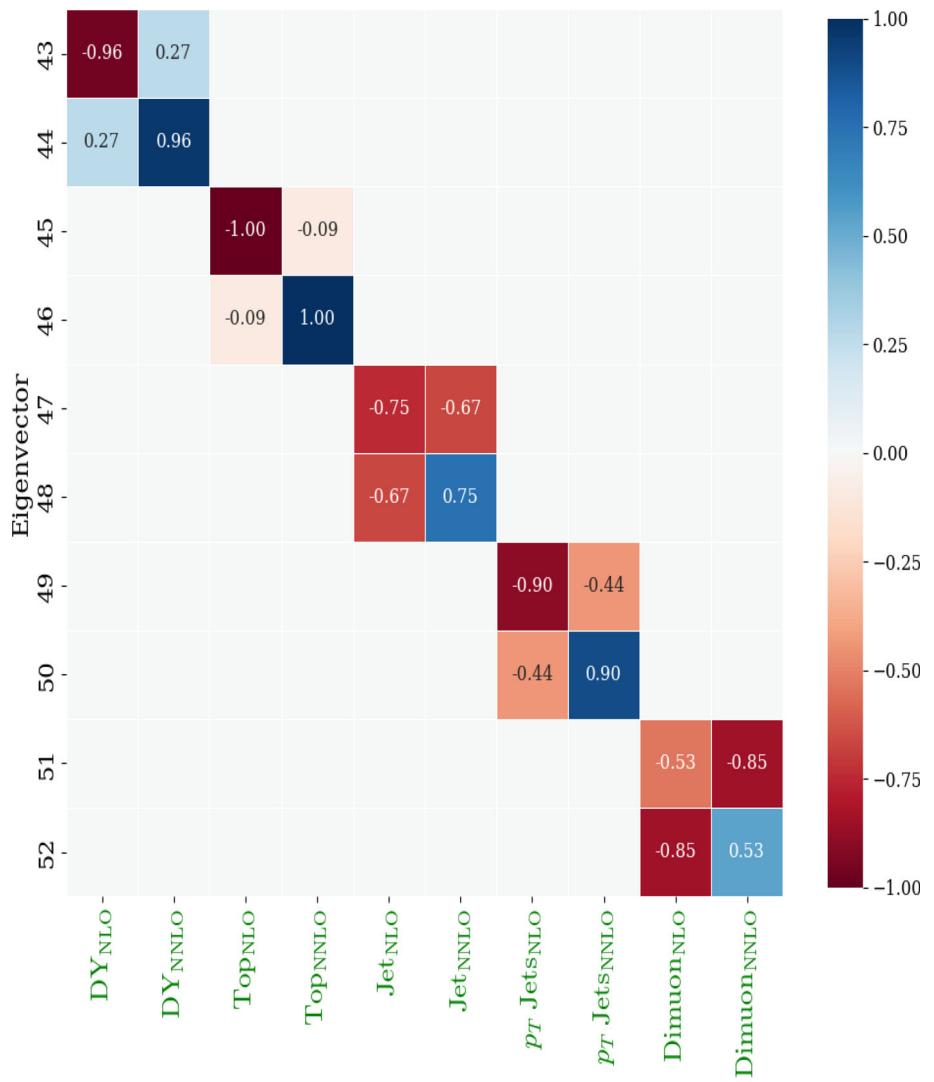
Moving to an analysis including a  $N^3LO$   $K$ -factors as correlated parameters with PDF and other  $N^3LO$  theory param-

eters. This provides a comparison to the case of decorrelated  $K$ -factors and justification for treating the cross section behaviour separately to the PDF theory behaviour.

Figure 32 shows a map of eigenvectors with the extra 10  $N^3LO$   $K$ -factor parameters (shown in green) included into the correlations considered. As expected, the result of including the correlations between PDF parameters and a  $N^3LO$   $K$ -factors results in a slightly more intertwined set of eigenvectors (although a high level of decorrelation remains). Specifically, due to the much higher number of DY datasets included

**Fig. 30** Map of the 10  $K$ -factor eigenvectors found with the  $N^3\text{LO}$  parameters added into the analysis in the case where the  $K$ -factors are decorrelated from these first 42 parameters..

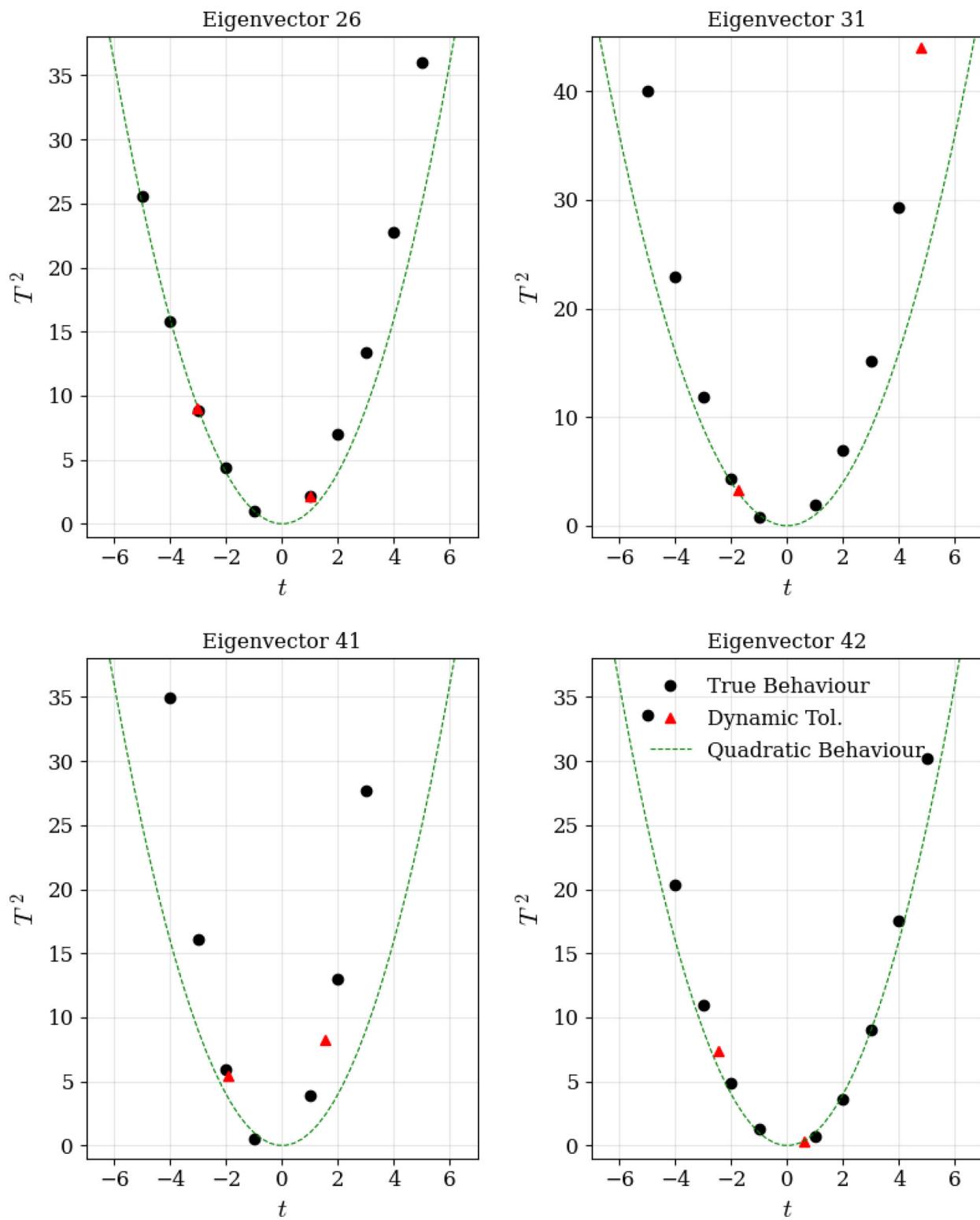
Combined with the 42 eigenvectors shown in Fig. 29, these form the total 52 eigenvectors in the decorrelated case. Parameters associated with the  $K$ -factor parameters are in green



**Table 10** Limiting values for specific  $K$ -factor parameters for each of the processes considered in the decorrelated case. Parameter values are shown in the positive and negative limits for each eigenvector. The scale choices for top quark processes are described in Sect. 7.2 to be  $H_T/4$

for the single differential datasets with the exception of data differential in the average transverse momentum of the top or antitop,  $p_T^t, p_T^{\bar{t}}$ , for which  $m_T/2$  is used. For the double diff. dataset the scale choice is  $H_T/4$  and for the inclusive top  $\sigma_{t\bar{t}}$  a scale of  $m_t$  is chosen

Matrix	Central values		Eigenvector	+ Limit		- Limit		Scale
	$a_{\text{NLO}}$	$a_{\text{NNLO}}$		$a_{\text{NLO}}$	$a_{\text{NNLO}}$	$a_{\text{NLO}}$	$a_{\text{NNLO}}$	
$K_{ij}^{\text{DY}}$	-0.307	-0.230	43	-0.449	-0.155	-0.213	-0.280	$m/2$
			44	-0.253	0.126	-0.377	-0.691	
$K_{ij}^{\text{Top}}$	0.041	0.651	45	-0.564	0.455	0.692	0.862	Section 7.2
			46	0.026	1.210	0.070	-0.456	
$K_{ij}^{\text{Jets}}$	-0.300	-0.691	47	-0.515	-0.957	0.105	-0.189	$p_T^{\text{jet}}$
			48	-0.725	-0.033	0.036	-1.212	
$K_{ij}^{p_T \text{ Jets}}$	0.583	-0.080	49	0.388	-0.406	0.812	0.301	$p_T$
			50	0.480	0.624	0.680	-0.742	
$K_{ij}^{\text{Dimuon}}$	-0.444	0.922	51	-1.109	-0.208	-0.103	1.502	$Q^2$
			52	-1.091	1.359	0.981	-0.039	



**Fig. 31** Dynamic tolerance behaviour for 4 selected eigenvectors in the case of decorrelated  $K$ -factor parameters. The black dots show the fixed tolerance relations found for integer values of  $t$ , whereas the red

triangles show the final chosen dynamical tolerances for each eigenvector direction. For an exhaustive analysis of all eigenvectors see Fig. 50

**Table 11** Tolerances resulting from eigenvector scans with decorrelated  $K$ -factors for each process. The average tolerance for this set of eigenvectors is  $T = 3.34$ 

#	$t+$	$T+$	Limiting factor (+)	$t-$	$T-$	Limiting factor (-)	Primary parameter
1	3.88	3.86	ATLAS 7 TeV high prec. $W, Z$ [79]	2.94	3.13	ATLAS 8 TeV double diff. $Z$ [100]	$a_{S,3}$
2	5.35	5.56	HERA $e^+ p$ NC 920 GeV [147]	3.10	2.98	NMC $\mu d$ $F_2$ [136]	$a_{S,6}$
3	4.20	4.47	NuTeV $\nu N \rightarrow \mu\mu X$ [133]	2.30	2.17	NMC $\mu d$ $F_2$ [136]	$a_{S,2}$
4	4.87	4.85	ATLAS 8 TeV double diff. $Z$ [100]	1.81	1.80	NuTeV $\nu N \rightarrow \mu\mu X$ [133]	$a_{s+,5}$
5	3.30	3.63	ATLAS 7 TeV high prec. $W, Z$ [79]	2.95	2.77	NuTeV $\nu N \rightarrow \mu\mu X$ [133]	$a_{s+,3}$
6	4.89	5.38	ATLAS 7 TeV high prec. $W, Z$ [79]	5.47	5.27	NMC $\mu d$ $F_2$ [136]	$\delta_{g'}$
7	5.30	5.47	DØ II $W \rightarrow ve$ asym. [86]	3.44	3.41	DØ $W$ asym. [97]	$\delta_S$
8	4.12	4.05	DØ II $W \rightarrow ve$ asym. [86]	2.19	2.64	DØ $W$ asym. [97]	$\delta_S$
9	3.88	3.93	BCDMS $\mu p$ $F_2$ [135]	6.78	7.11	CMS $W$ asym. $p_T > 25, 30$ GeV [89]	$\delta_d$
10	3.24	3.44	BCDMS $\mu p$ $F_2$ [135]	5.64	5.83	DØ II $W \rightarrow \nu\mu$ asym. [84]	$a_{u,6}$
11	2.22	4.41	E866/NuSea $pp$ DY [80]	5.31	4.79	ATLAS 7 TeV high prec. $W, Z$ [79]	$a_{s+,2}$
12	2.62	2.81	DØ $W$ asym. [97]	5.49	5.71	ATLAS 8 TeV $W$ [99]	$\delta_d$
13	1.12	3.37	E866/NuSea $pp$ DY [80]	3.87	3.19	DØ $W$ asym. [97]	$a_{\rho,3}$
14	2.01	2.34	NuTeV $\nu N \rightarrow \mu\mu X$ [133]	3.54	3.60	NuTeV $\nu N \times F_3$ [140]	$A_{s-}$
15	3.69	3.75	ATLAS 8 TeV $Z$ $p_T$ [110]	3.65	4.05	NMC $\mu d$ $F_2$ [136]	$a_{g,3}$
16	4.42	5.08	ATLAS 8 TeV $W$ [99]	4.96	4.84	ATLAS 8 TeV $Z$ $p_T$ [110]	$a_{g,3}$
17	1.23	1.02	$\rho_{qq}^{NS}$	1.24	1.05	$\rho_{qq}^{NS}$	$\rho_{qq}^{NS}$
18	3.09	3.50	DØ $W$ asym. [97]	2.93	3.06	DØ $W$ asym. [97]	$a_{d,6}$
19	4.69	4.57	BCDMS $\mu p$ $F_2$ [135]	3.85	5.06	CMS 7 TeV jets [101]	$\eta_u$
20	5.40	5.37	NuTeV $\nu N \rightarrow \mu\mu X$ [133]	5.04	5.60	HERA $e^+ p$ NC 920 GeV [147]	$\delta_{g'}$
21	2.10	2.36	E866/NuSea $pd/pp$ DY [81]	1.26	1.69	E866/NuSea $pd/pp$ DY [81]	$a_{\rho,6}$
22	3.04	3.27	NuTeV $\nu N \rightarrow \mu\mu X$ [133]	1.90	2.27	DØ $W$ asym. [97]	$a_{d,3}$
23	5.61	6.20	HERA $e^+ p$ NC 920 GeV [147]	5.43	6.13	HERA $e^+ p$ NC 920 GeV [147]	$\rho_{qq}^{PS}$
24	3.53	3.70	E866/NuSea $pd/pp$ DY [81]	1.47	1.82	DØ $W$ asym. [97]	$a_{d,2}$
25	1.60	2.03	E866/NuSea $pd/pp$ DY [81]	4.37	4.88	CMS $W$ asym. $p_T > 35$ GeV [88]	$A_\rho$
26	1.00	1.46	E866/NuSea $pd/pp$ DY [81]	3.02	3.00	DØ $W$ asym. [97]	$a_{\rho,1}$
27	1.60	2.09	$\rho_{gg}$	4.16	5.53	ATLAS 7 TeV high prec. $W, Z$ [79]	$\rho_{gg}$
28	1.51	2.16	$\rho_{gg}$	3.20	4.07	ATLAS 8 TeV sing. diff. $t\bar{t}$ dilep. [131]	$\eta_g$
29	2.99	3.21	CMS 8 TeV $W$ [96]	2.04	2.52	NuTeV $\nu N \times F_3$ [140]	$\eta_S$
30	0.97	1.30	DØ $W$ asym. [97]	3.56	4.90	ATLAS 8 TeV $W$ [99]	$\eta_d - \eta_u$
31	4.78	6.64	HERA $e^+ p$ NC 920 GeV [147]	1.77	1.81	$\rho_{gg}$	$\rho_{qg}$
32	2.51	7.32	BCDMS $\mu p$ $F_2$ [135]	1.80	3.96	DØ $W$ asym. [97]	$\eta_d - \eta_u$
33	2.71	3.35	CMS 7 TeV $W + c$ [62]	2.92	3.37	NuTeV $\nu N \rightarrow \mu\mu X$ [133]	$\eta_{s+}$
34	3.53	3.91	HERA $ep$ $F_2^{\text{charm}}$ [142]	3.87	4.82	CMS 7 TeV $W + c$ [62]	$\eta_{s+}$
35	3.67	3.50	NuTeV $\nu N \rightarrow \mu\mu X$ [133]	2.72	3.62	NuTeV $\nu N \rightarrow \mu\mu X$ [133]	$\eta_{s-}$
36	1.25	1.26	$c_q^{\text{NLL}}$	1.41	1.57	$c_q^{\text{NLL}}$	$c_q^{\text{NLL}}$
37	0.71	0.77	$\rho_{gg}$	4.25	5.33	NuTeV $\nu N \rightarrow \mu\mu X$ [133]	$\rho_{gg}$
38	2.72	2.71	CMS 8 TeV $W$ [96]	3.60	4.21	$\rho_{gg}$	$A_{s+}$
39	1.56	5.03	ATLAS 7 TeV high prec. $W, Z$ [79]	1.60	5.82	ATLAS 7 TeV high prec. $W, Z$ [79]	$A_S$
40	0.97	1.00	$a_{qq,H}^{\text{NS}}$	1.04	1.01	$a_{qq,H}^{\text{NS}}$	$a_{qq,H}^{\text{NS}}$
41	1.56	2.87	HERA $ep$ $F_2^{\text{charm}}$ [142]	1.93	2.34	$\rho_{gg}$	$a_{Hg}$
42	0.63	0.53	$a_{gg,H}$	2.46	2.72	$a_{gg,H}$	$a_{gg,H}$

**Table 11** continued

#	$t+$	$T+$	Limiting factor (+)	$t-$	$T-$	Limiting factor (-)	Primary parameter
43	2.28	2.83	ATLAS 7 TeV high prec. $W, Z$ [79]	1.51	1.14	CMS double diff. Drell–Yan [94]	$DY_{NLO}$
44	3.06	3.09	E866/NuSea $pp$ DY [80]	3.96	4.52	ATLAS 7 TeV high prec. $W, Z$ [79]	$DY_{NNLO}$
45	5.05	1.46	ATLAS 8 TeV sing. diff. $t\bar{t}$ dilep. [131]	5.44	1.46	Tevatron, ATLAS, CMS $\sigma_{t\bar{t}}$ [117–129]	$Top_{NLO}$
46	1.34	1.31	$Top_{NLO}$	2.65	2.68	ATLAS 8 TeV sing. diff. $t\bar{t}$ [130]	$Top_{NNLO}$
47	1.47	1.57	CDF II $p\bar{p}$ incl. jets [102]	2.78	2.96	$a_{gg, H}$	$Jet_{NLO}$
48	3.26	3.38	CMS 2.76 TeV jet [106]	2.58	3.01	CMS 7 TeV jets [101]	$Jet_{NNLO}$
49	2.20	2.39	ATLAS 8 TeV $Z$ $p_T$ [110]	2.58	2.56	ATLAS 8 TeV $W$ + jets [111]	$p_T$ $Jets_{NLO}$
50	2.36	2.43	ATLAS 8 TeV $Z$ $p_T$ [110]	2.22	2.23	ATLAS 8 TeV $W$ + jets [111]	$p_T$ $Jets_{NNLO}$
51	1.38	1.55	Dimuon $_{NLO}$	0.71	0.81	Dimuon $_{NNLO}$	Dimuon $_{NNLO}$
52	0.85	0.85	Dimuon $_{NLO}$	1.86	1.87	Dimuon $_{NLO}$	Dimuon $_{NLO}$

in the global fit, these  $N^3LO$   $K$ -factor parameters tend to be included across more of a spread of eigenvectors. On the other hand, the Dimuon  $K$ -factors are almost entirely isolated within two eigenvectors, similar to the decorrelated case.

Once again, to investigate deviations from the quadratic behaviour, Fig. 33 illustrates examples of the tolerance behaviours of selected eigenvectors, with a full analysis provided in Appendix 1. Further to this, Table 12 displays the tolerances and limiting datasets/parameters for the 52 correlated eigenvectors. It is difficult to compare and contrast these results with the decorrelated case, since the eigenvectors are inherently different. However in both cases, the eigenvectors are similarly well behaved, exhibit relatively good consistency between  $t$  and  $T$  and are therefore providing valid descriptions for a PDF fit.

For most of the 12 eigenvectors with  $N^3LO$   $K$ -factors as primary parameters, there is expected behaviour, with the eigenvectors constrained either by their own penalties or by dominant datasets for the associated process. However, due to the extra correlations considered, there are a small number of eigenvector directions which are not as trivial to explain (e.g. eigenvector 31). We therefore recover the lack of correlation between  $K$ -factor parameters seen within Fig. 28 in the set of correlated PDF eigenvectors presented here. Further to this, comparing the  $t$  and  $T$  values found for eigenvectors associated with  $N^3LO$   $K$ -factors in Tables 11 and 12, one can observe clear similarities between eigenvectors. This suggests that even when correlating the  $K$ -factor parameters, the fit succeeds in decorrelating the individual processes, thereby motivating our original assumption that the correlations with  $K$ -factors can be ignored. Another similarity one can observe between Tables 11 and 12 is the suggestion of some tension between ATLAS 8 TeV  $Z$   $p_T$  [110] and ATLAS 8 TeV  $W$  + jets [111] datasets seen in the limiting factors of eigenvector 39 in the correlated case.

Eigenvectors 27, 29 and 52 displayed in Fig. 33 can be seen from Table 12 to be associated with the new  $N^3LO$  the-

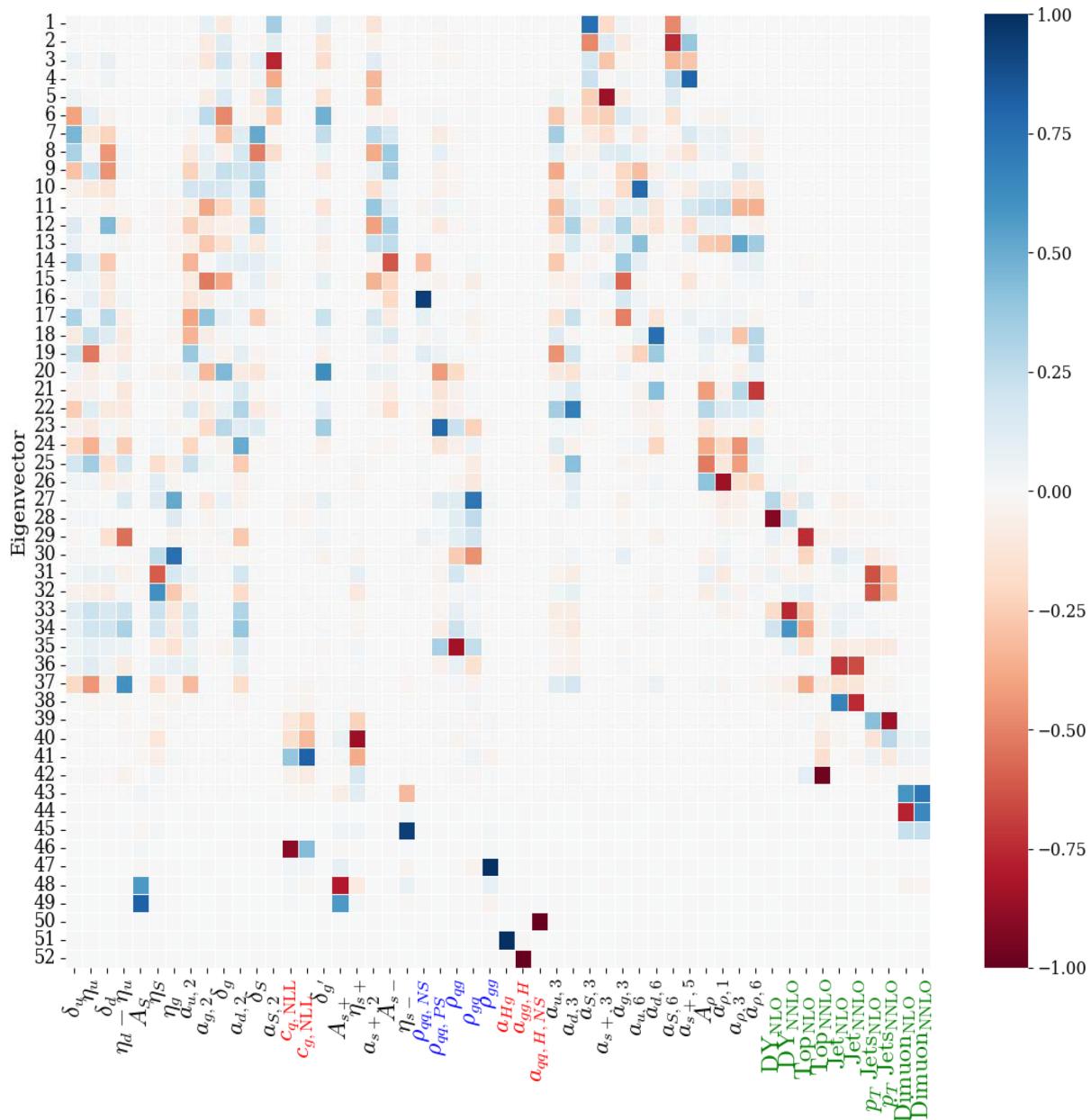
ory parameters. Whereas eigenvector 37 is primarily focused on an original PDF parameter. One can observe a similar level of quadratic behaviour across all four of these eigenvector tolerances. Comparing all eigenvectors in the decorrelated/correlated cases, the behaviours are similarly well behaved. The average tolerance  $T$  for the  $aN^3LO$  (with correlated  $K$ -factors) case is 3.57, slightly higher than the  $NNLO$  average of 3.37 and the  $aN^3LO$  (with decorrelated  $K$ -factors) average of 3.34.

#### 8.4 PDF results

Figure 34 displays the overall shape of the PDFs including the  $N^3LO$  additions compared to the standard  $NNLO$  set. We provide this comparison to accompany the results described in earlier sections. At small- $x$  and low- $Q^2$  the gluon exhibits a marked enhancement due to the large small- $x$  logarithms inserted at  $N^3LO$ . The changes induced from specific  $N^3LO$  contributions are investigated in Sect. 8.8.

Shown in Figs. 35 and 36 are the ratios for each flavour of  $aN^3LO$  PDF compared to the  $NNLO$  set with their 68% confidence intervals at low and high- $Q^2$  respectively. The shaded  $aN^3LO$  regions indicate the PDF uncertainty produced with the decorrelated  $((H_{ij}^{-1} + \sum_{p=1}^{N_p} K_{ij, p}^{-1})^{-1}) aN^3LO$   $K$ -factors for each process. As a comparison to these shaded regions, the bounds of uncertainty for the fully correlated  $(H'_{ij}) N^3LO$   $K$ -factor parameters are also provided (red dashed line).

Considering Fig. 35 we present the  $aN^3LO$  PDF set at  $Q^2 = 10 \text{ GeV}^2$  with the bottom quark PDF at  $Q^2 = 25 \text{ GeV}^2$ . These PDF ratios better display the substantial increase in the gluon at small- $x$ , reminiscent of the gluon PDF presented in [16, 17]. The predicted harder small- $x$  gluon is then accommodated for by reductions in the PDFs at large and small- $x$  (particularly the gluon near  $x = 10^{-2}$ ) from  $NNLO$ . Another prominent feature is the enhanced charm and bottom quark at  $N^3LO$ . Since the heavy flavour quarks are perturbatively calculated in the MSHT framework, this amplification

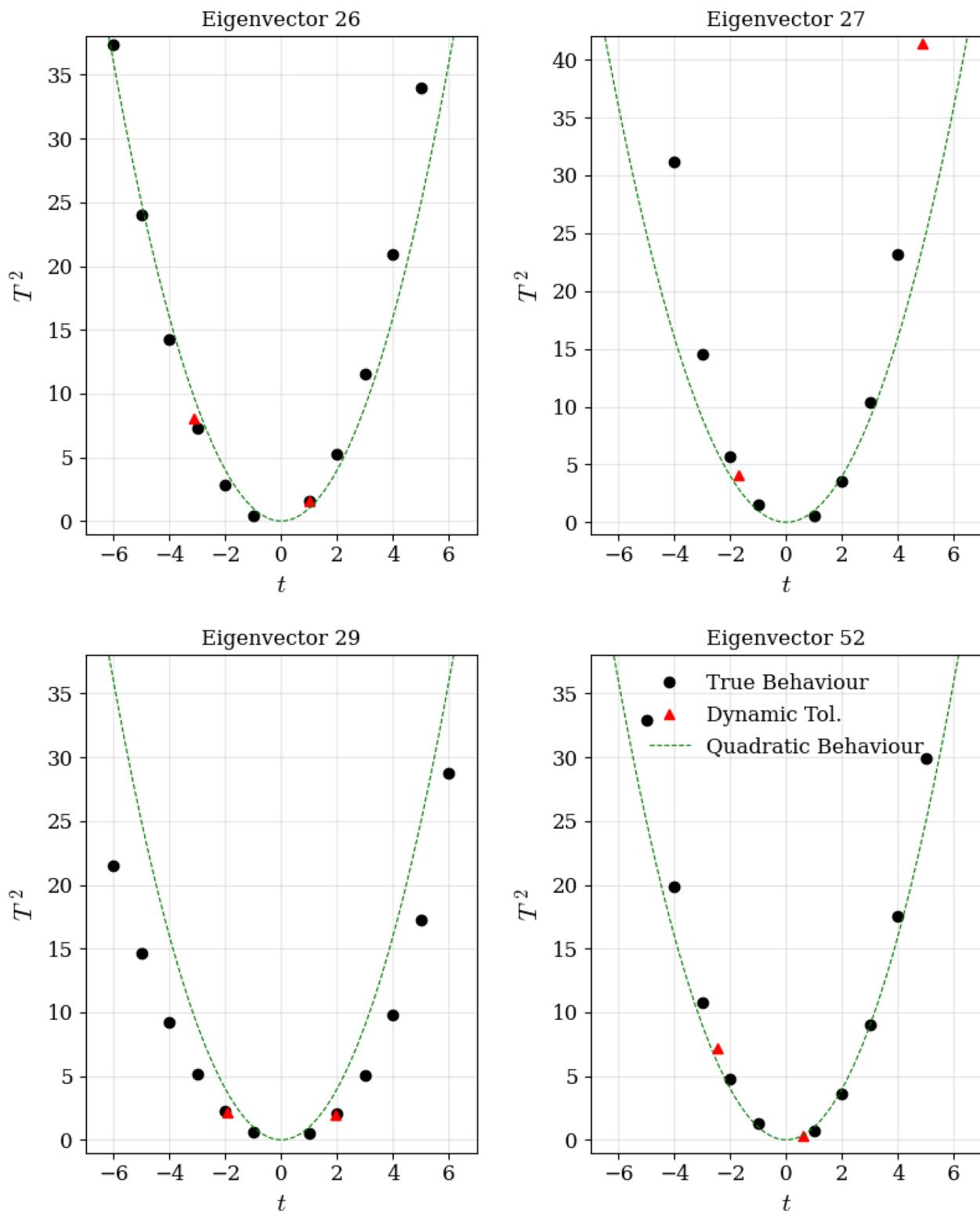


**Fig. 32** Map of eigenvectors found with the  $N^3LO$  theory and  $K$ -factor parameters added into the analysis. Parameters associated with the splitting functions are coloured blue, those affecting the transition matrix elements and coeffi-

cient functions are in red and the  $K$ -factor parameters are in green

is a feature of the transition matrix element  $A_{Hg}^{(3)}$  at high- $x$ , combined with the increase in the gluon PDF at small- $x$  (as these two ingredients are convoluted together). Comparing with Fig. 95 in [3], we observe that the approximate  $N^3LO$  charm quark now follows a much closer trend to the CT18 PDF and is therefore even more significantly different from the NNPDF NNLO fitted charm at large- $x$  than MSHT20 at NNLO. In the high- $Q^2$  setting shown in Fig. 36 we observe similar albeit less drastic effects to those described above.

Also contained in Figs. 35 and 36 are the relative forms of NNLO PDFs when fit to all non-HERA data (full  $\chi^2$  results are provided in Appendix B). Comparing the *non-HERA* NNLO PDFs with a $N^3LO$  PDFs, there are some similarities in the shapes and magnitudes of a handful of PDFs in the intermediate to large- $x$  regime, most noticeably the light quarks. At small- $x$  the HERA data heavily constrains the PDF fit and therefore these similarities rapidly break down. However, this analysis displays further evidence that including



**Fig. 33** Dynamic tolerance behaviour for 4 selected eigenvectors in the case of correlated  $K$ -factor parameters. The black dots show the fixed tolerance relations found for integer values of  $t$ , whereas the red triangles show the final chosen dynamical tolerances for each eigenvector direction. For an exhaustive analysis of all eigenvectors see Fig. 51

gles show the final chosen dynamical tolerances for each eigenvector direction. For an exhaustive analysis of all eigenvectors see Fig. 51

**Table 12** Tolerances resulting from eigenvector scans with correlated  $K$ -factors for each process. The average tolerance for this set of eigenvectors is  $T = 3.57$ 

#	$t+$	$T+$	Limiting factor (+)	$t-$	$T-$	Limiting factor (-)	Primary parameter
1	3.76	3.67	ATLAS 7 TeV high prec. $W, Z$ [79]	2.95	3.06	ATLAS 8 TeV double diff. $Z$ [100]	$a_{S,3}$
2	3.67	3.52	NMC $\mu d$ $F_2$ [136]	5.37	5.54	HERA $e^+ p$ NC 920 GeV [147]	$a_{S,6}$
3	4.15	4.33	NuTeV $\nu N \rightarrow \mu\mu X$ [133]	2.58	2.39	NMC $\mu d$ $F_2$ [136]	$a_{S,2}$
4	4.90	4.82	ATLAS 8 TeV double diff. $Z$ [100]	2.21	2.16	NuTeV $\nu N \rightarrow \mu\mu X$ [133]	$a_{s+,5}$
5	2.89	2.61	NuTeV $\nu N \rightarrow \mu\mu X$ [133]	3.34	3.55	CCFR $\nu N \rightarrow \mu\mu X$ [133]	$a_{s+,3}$
6	5.58	5.23	NMC $\mu d$ $F_2$ [136]	5.49	5.84	ATLAS 7 TeV high prec. $W, Z$ [79]	$\delta_{g'}$
7	5.34	5.44	DØ II $W \rightarrow \nu e$ asym. [86]	3.46	3.37	DØ $W$ asym. [97]	$\delta_S$
8	4.11	3.82	DØ II $W \rightarrow \nu e$ asym. [86]	2.18	2.48	DØ $W$ asym. [97]	$\delta_S$
9	6.39	6.59	CMS $W$ asym. $p_T > 25, 30$ GeV [89]	4.20	4.07	BCDMS $\mu p$ $F_2$ [135]	$\delta_d$
10	3.14	3.11	BCDMS $\mu p$ $F_2$ [135]	5.66	5.66	DØ II $W \rightarrow \nu\mu$ asym. [84]	$a_{u,6}$
11	2.07	3.99	E866/NuSea $pp$ DY [80]	5.88	5.00	ATLAS 7 TeV high prec. $W, Z$ [79]	$a_{g,2}$
12	2.76	2.76	DØ $W$ asym. [97]	5.60	5.67	CMS 8 TeV $W$ [96]	$\delta_d$
13	3.73	3.01	DØ $W$ asym. [97]	1.26	3.96	E866/NuSea $pp$ DY [80]	$a_{\rho,3}$
14	3.85	3.55	NuTeV $\nu N x F_3$ [140]	2.24	2.30	NuTeV $\nu N \rightarrow \mu\mu X$ [133]	$A_{s-}$
15	3.71	3.51	ATLAS 8 TeV $Z$ $p_T$ [110]	3.44	3.72	NuTeV $\nu N \rightarrow \mu\mu X$ [133]	$a_{g,3}$
16	1.78	1.18	$\rho_{qq}^{NS}$	1.78	1.24	$\rho_{qq}^{NS}$	$\rho_{qq}^{NS}$
17	4.76	4.49	DØ $W$ asym. [97]	4.37	4.81	ATLAS 8 TeV $W$ [99]	$a_{g,3}$
18	3.44	3.81	DØ $W$ asym. [97]	3.18	2.94	DØ $W$ asym. [97]	$a_{d,6}$
19	5.33	5.02	DØ $W$ asym. [97]	3.77	4.42	CMS 7 TeV jets [101]	$\eta_u$
20	6.04	5.82	HERA $e^+ p$ NC 920 GeV [147]	5.08	5.63	HERA $e^+ p$ NC 920 GeV [147]	$\delta_{g'}$
21	2.14	2.02	E866/NuSea $pd/pp$ DY [81]	1.29	1.39	E866/NuSea $pd/pp$ DY [81]	$a_{\rho,6}$
22	1.95	2.03	DØ $W$ asym. [97]	2.88	2.83	NuTeV $\nu N \rightarrow \mu\mu X$ [133]	$a_{d,3}$
23	5.33	5.87	HERA $e^+ p$ NC 920 GeV [147]	5.66	6.18	HERA $e^+ p$ NC 920 GeV [147]	$\rho_{qq}^{PS}$
24	1.47	1.52	DØ $W$ asym. [97]	3.51	3.48	E866/NuSea $pd/pp$ DY [81]	$a_{d,2}$
25	4.34	4.71	CMS $W$ asym. $p_T > 35$ GeV [88]	1.64	1.59	E866/NuSea $pd/pp$ DY [81]	$A_\rho$
26	1.00	1.24	E866/NuSea $pd/pp$ DY [81]	3.12	2.84	DØ $W$ asym. [97]	$a_{\rho,1}$
27	4.86	6.43	CMS double diff. Drell–Yan [94]	1.68	2.01	$\rho_{gq}$	$\rho_{gq}$
28	2.36	2.66	ATLAS 7 TeV high prec. $W, Z$ [79]	1.64	1.36	CMS double diff. Drell–Yan [94]	$DY_{NLO}$
29	1.95	1.41	$\rho_{qq}^{NS}$	1.94	1.47	$\rho_{qq}^{NS}$	$Top_{NLO}$
30	1.84	2.22	$\rho_{gq}$	2.70	2.56	ATLAS 8 TeV sing. diff. $t\bar{t}$ dilep. [131]	$\eta_g$
31	3.53	3.52	ATLAS 8 TeV $Z$ $p_T$ [110]	2.60	2.73	NuTeV $\nu N x F_3$ [140]	$p_T \text{ Jets}_{NLO}$
32	2.85	3.15	NuTeV $\nu N x F_3$ [140]	3.84	3.57	ATLAS 8 TeV $W$ + jets [111]	$p_T \text{ Jets}_{NLO}$
33	5.40	6.16	ATLAS 7 TeV high prec. $W, Z$ [79]	1.63	1.68	DØ $W$ asym. [97]	$DY_{NNLO}$
34	3.94	4.19	ATLAS 8 TeV $W$ [99]	1.56	1.86	DØ $W$ asym. [97]	$DY_{NNLO}$
35	5.06	6.73	HERA $e^+ p$ NC 920 GeV [147]	2.00	1.84	$\rho_{gq}$	$\rho_{gq}$
36	2.31	2.48	CDF II $p\bar{p}$ incl. jets [102]	4.73	4.88	ATLAS 7 TeV jets [104]	$Jet_{NLO}$
37	2.72	7.05	BCDMS $\mu p$ $F_2$ [135]	1.67	3.04	DØ $W$ asym. [97]	$\eta_d - \eta_u$
38	2.63	2.90	$Jet_{NLO}$	3.24	3.35	CMS 2.76 TeV jet [106]	$Jet_{NNLO}$
39	2.41	2.34	ATLAS 8 TeV $W$ + jets [111]	2.88	2.99	ATLAS 8 TeV $Z$ $p_T$ [110]	$p_T \text{ Jets}_{NNLO}$
40	2.98	3.44	CMS 7 TeV $W + c$ [62]	2.43	2.36	NuTeV $\nu N \rightarrow \mu\mu X$ [133]	$\eta_{s+}$
41	4.59	5.42	HERA $ep$ $F_2^{\text{charm}}$ [142]	2.93	3.20	HERA $ep$ $F_2^{\text{charm}}$ [142]	$c_g^{\text{NLL}}$
42	2.74	2.83	CMS 8 TeV double diff. $t\bar{t}$ [132]	1.36	1.28	$Top_{NLO}$	$Top_{NNLO}$

**Table 12** continued

#	$t+$	$T+$	Limiting factor (+)	$t-$	$T-$	Limiting factor (-)	Primary parameter
43	0.83	0.71	DimuonNNLO	1.36	1.46	$p_T$ JetsNNLO	DimuonNNLO
44	0.81	0.80	DimuonNNLO	2.07	2.05	$p_T$ JetsNNLO	DimuonNLO
45	1.45	1.46	DimuonNLO	2.09	2.38	$p_T$ JetsNNLO	$\eta_{s-}$
46	1.41	1.57	$c_q^{\text{NLL}}$	1.25	1.26	$c_q^{\text{NLL}}$	$c_q^{\text{NLL}}$
47	0.71	0.77	$\rho_{gg}$	4.27	5.21	NuTeV $\nu N \rightarrow \mu\mu X$ [133]	$\rho_{gg}$
48	3.62	4.07	$\rho_{gg}$	2.72	2.63	CMS 8 TeV $W$ [96]	$A_{s+}$
49	1.53	4.81	ATLAS 7 TeV high prec. $W, Z$ [79]	1.56	5.54	ATLAS 7 TeV high prec. $W, Z$ [79]	$A_S$
50	0.97	1.00	$a_{qq,H}^{\text{NS}}$	1.04	1.00	$a_{qq,H}^{\text{NS}}$	$a_{qq,H}^{\text{NS}}$
51	1.56	2.69	HERA $ep$ $F_2^{\text{charm}}$ [142]	1.93	1.98	$\rho_{gg}$	$a_{Hg}$
52	0.63	0.54	$a_{gg,H}$	2.46	2.68	$a_{gg,H}$	$a_{gg,H}$

$N^3\text{LO}$  contributions, even though approximate, reduces tensions between the HERA and non-HERA data (when considering the reduction in tension seen in Table 19). The  $aN^3\text{LO}$  PDFs are seemingly able to fit to HERA and non-HERA datasets with superior flexibility than at NNLO.

While in principle the negativity of quarks is possible in the  $\overline{MS}$  scheme, it is unlikely to be correct at very high scales and the behaviour can lead to issues concerning negative cross section predictions [148, 149]. In the case of the  $\bar{d}$ , the form of this PDF has a negative central value above  $x \sim 0.5$  with a minimum of  $\sim -0.001$  at  $x \sim 0.6$ . It is also noted that although the  $\bar{d}$  central value becomes negative in this region, it is still positive within PDF uncertainties. These features are not uncommon in PDF analyses and are discussed in detail in [8]. The proposed smoothing of parameterisations employed in [8] ensures the definite positive nature of PDFs in the high- $x$  region. Comparing the negativity of the approximate  $N^3\text{LO}$   $\bar{d}$  PDF with that in [8], the  $\bar{d}$  PDF presented here is much less negative and positive within PDF uncertainties. Due to this and the fact that this effect is only apparent in the  $\bar{d}$ , we present these PDFs as they are. We also note that in the current MSHT20 fit, recent results surrounding the  $\bar{d}/\bar{u}$  from the SeaQuest collaboration [150] are not included at the time of writing. It is therefore only the E866 / NuSea  $pd/p$  DY dataset [81] that is constraining this ratio, which is not as precise as the more recent results. However, SeaQuest results suggest a preference for a higher  $\bar{d}$  at large- $x$ , therefore including this data may in fact help constrain the high- $x$   $\bar{d}$  behaviour seen here.

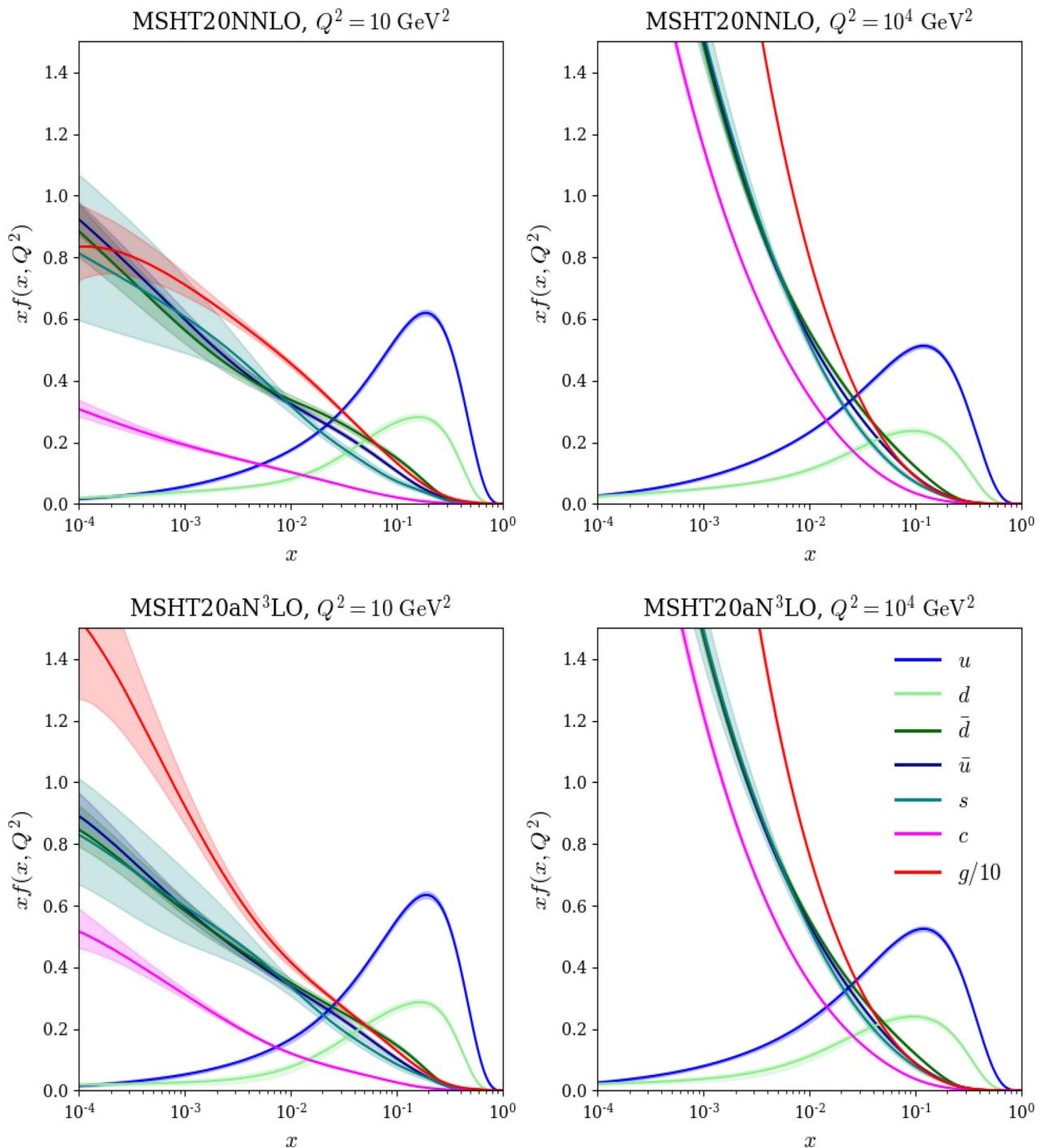
Figures 37 and 38 express the  $aN^3\text{LO}$  PDFs with decorrelated (green shaded region) and correlated (red dashed lines)  $aN^3\text{LO}$   $K$ -factors at low and high- $Q^2$  respectively (again with the bottom quark provided at  $Q^2 = 25 \text{ GeV}^2$  at low- $Q^2$ ) as a ratio to the  $N^3\text{LO}$  central value. For comparison we also include the level of uncertainty predicted with all  $N^3\text{LO}$  theory fixed (blue shaded region) i.e. only considering the variation *without*  $N^3\text{LO}$  theoretical uncertainty.

Comparing the two different  $aN^3\text{LO}$  sets in Figs. 37 and 38, in general there is good agreement between the total uncertainties considering the cases with correlated (red dash) and decorrelated (green shaded)  $aN^3\text{LO}$   $K$ -factors. The differences that are apparent between the two  $aN^3\text{LO}$  cases, are relatively small across all PDFs, with slightly larger effects only where the PDF itself tends towards zero i.e. valence quarks at small- $x$ .

A larger distinction is observed when comparing the sets *with* and *without* theoretical uncertainty (where  $N^3\text{LO}$  theory is fixed at the best fit value). In general there is an expected substantial increase in the PDF uncertainties when taking into account the missing  $N^3\text{LO}$  uncertainty for the gluon (and therefore the heavy quarks). In particular, the form of the  $N^3\text{LO}$  bottom quark uncertainty is reminiscent of the  $(H + \bar{H})$  prediction from Fig. 12. One can therefore directly observe the effect of the  $A_{Hg}$  theoretical uncertainty on the bottom quark directly above its mass threshold. In other areas, the *without* theoretical uncertainty PDF set exhibits a comparable uncertainty to  $aN^3\text{LO}$  and is even shown to increase the overall 68% confidence intervals in certain regions of  $(x, Q^2)$  due to  $N^3\text{LO}$  parameters being fixed (i.e.  $u_v$  and  $d_v$  PDFs in Figs. 37 and 38). As the fit now resides in a different  $\chi^2$  landscape where a best fit has been achieved through fitting the  $N^3\text{LO}$  theory, fixing the  $aN^3\text{LO}$  theory parameters is likely to have a substantial effect across all PDFs.

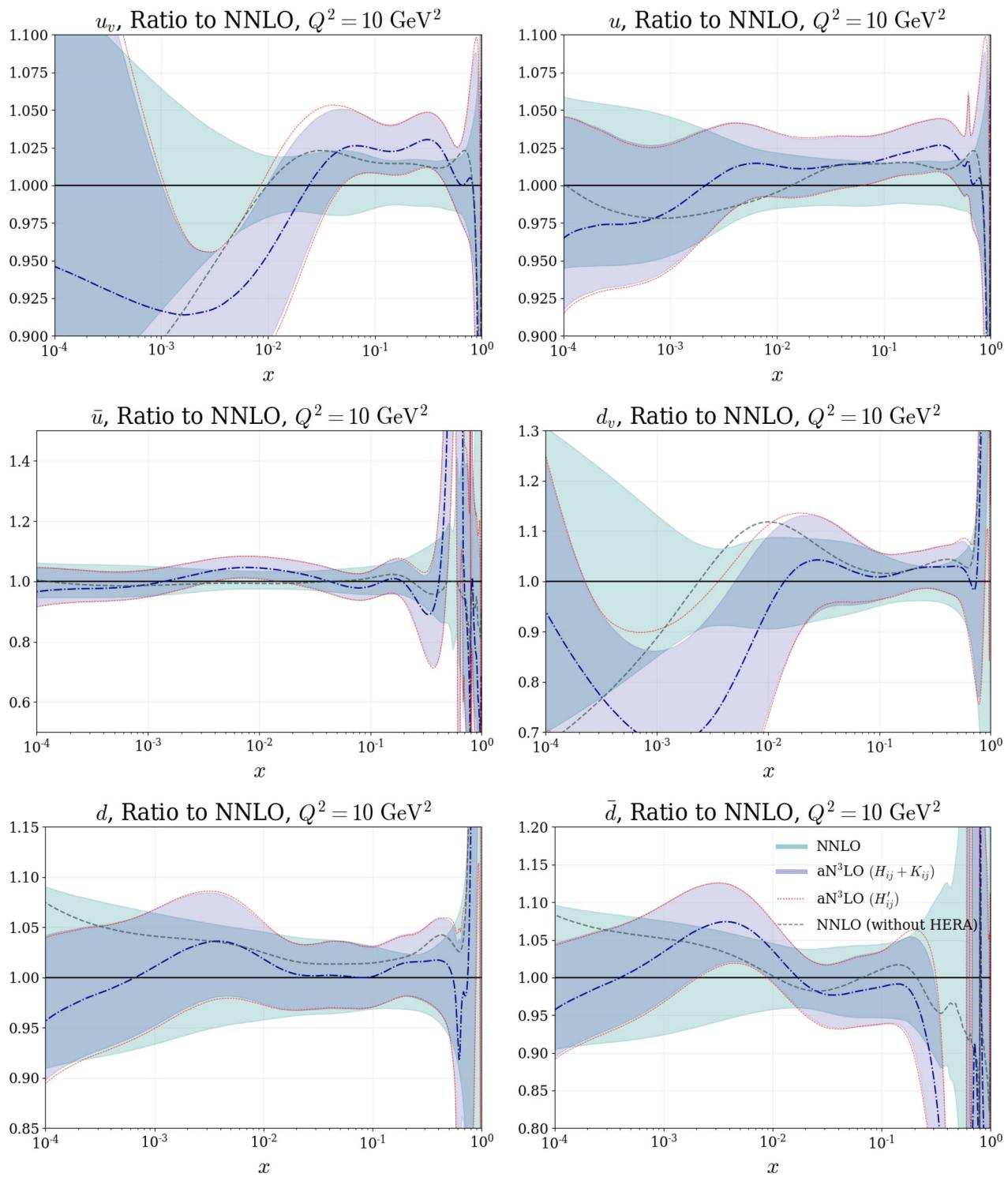
An important point made by Figs. 37 and 38 is that the difference between the decorrelated and correlated cases is much smaller than the difference of not including theoretical uncertainties at all (blue shaded region). This analysis therefore provides evidence to support the original assumption of being able to decorrelate the cross section ( $aN^3\text{LO}$   $K$ -factors) and PDF theory (including other  $N^3\text{LO}$  theory).

Along with the separate cases of uncertainty illustrated in Figs. 37 and 38, we also display the central values of an  $aN^3\text{LO}$  fit to all non-HERA data and an  $aN^3\text{LO}$  fit with NNLO  $K$ -factors. Examining the form of the no HERA



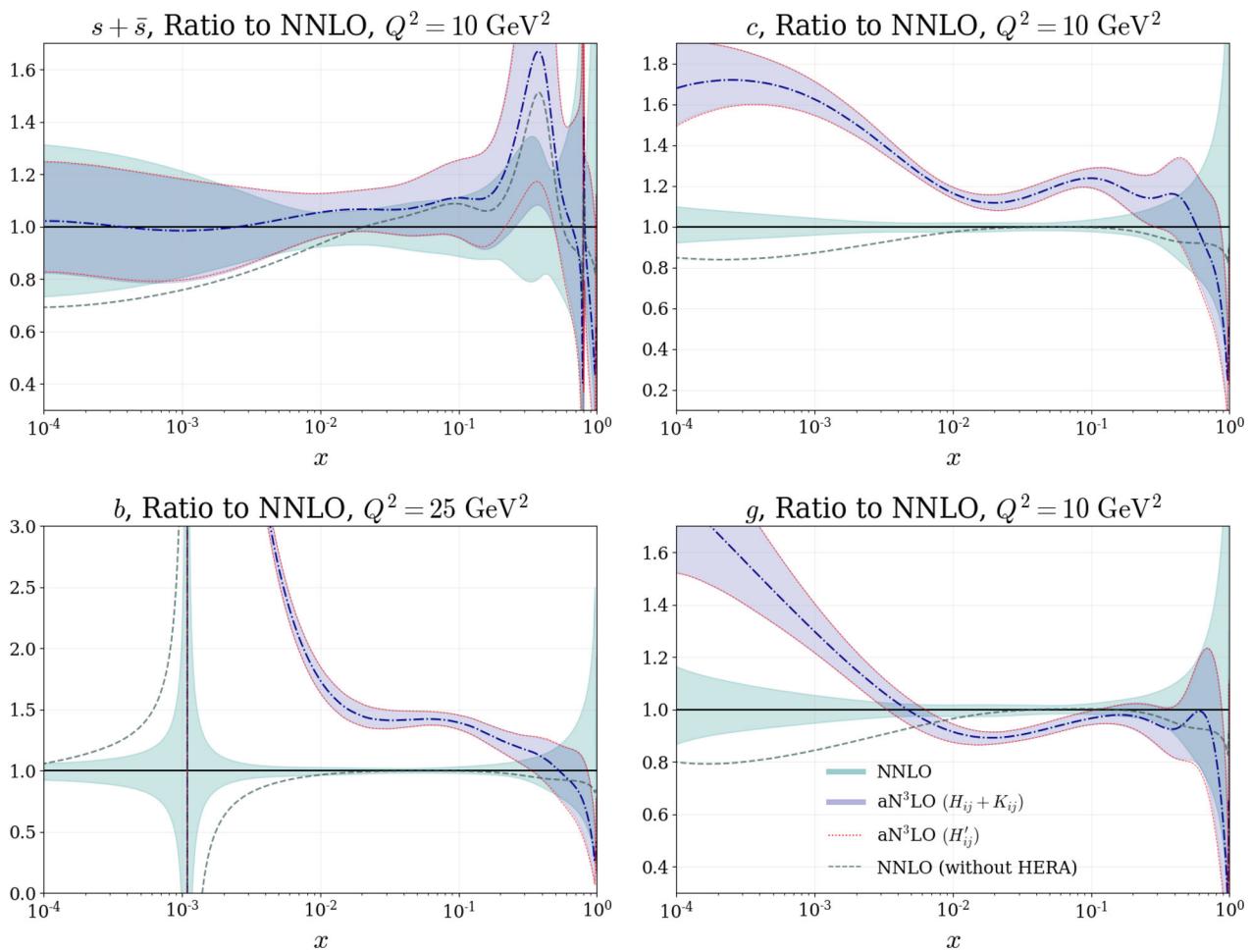
**Fig. 34** General forms of NNLO (top) and  $aN^3LO$  (bottom) PDFs at low (left) and high (right)  $Q^2$ . Several main features can be compared and contrasted such as the marked increase in the gluon and charm at small- $x$  (note the dif-

ference in y-axis scale between NNLO (top) and  $aN^3LO$  (bottom))



**Fig. 35** Low- $Q^2$  ratio plots showing the aN<sup>3</sup>LO 68% confidence intervals with decorrelated ( $H_{ij} + K_{ij}$ ) and correlated ( $H'_{ij}$ )  $K$ -factor parameters, compared to NNLO 68% confidence intervals. Also shown are the central values at NNLO when fit to all non-HERA datasets which show

similarities with N<sup>3</sup>LO in the large- $x$  region of selected PDF flavours. All plots are shown for  $Q^2 = 10 \text{ GeV}^2$  with the exception of the bottom quark shown for  $Q^2 = 25 \text{ GeV}^2$

**Fig. 35** continued

$aN^3LO$  PDFs for  $x > 10^{-2}$ , we show some agreement with the standard  $N^3LO$  central value across most PDFs (more so at high- $Q^2$  than low- $Q^2$ ). Whereas the form at small- $x$  gives some insight into the importance of HERA data in constraining PDFs in this region. In slightly better agreement across all  $x$  are the  $aN^3LO$  PDFs with NNLO  $K$ -factors, which compliment the  $\chi^2$  results in Sect. 7 and Sect. 8.1 arguing that the form (and fit results) of  $aN^3LO$  PDFs is mostly determined from the extra PDF + DIS coefficient function  $N^3LO$  additions i.e. not  $aN^3LO$   $K$ -factors which prefer a softer high- $x$  gluon (similar to the  $N^3LO$  no HERA case – also shown in Figs. 37 and 38).

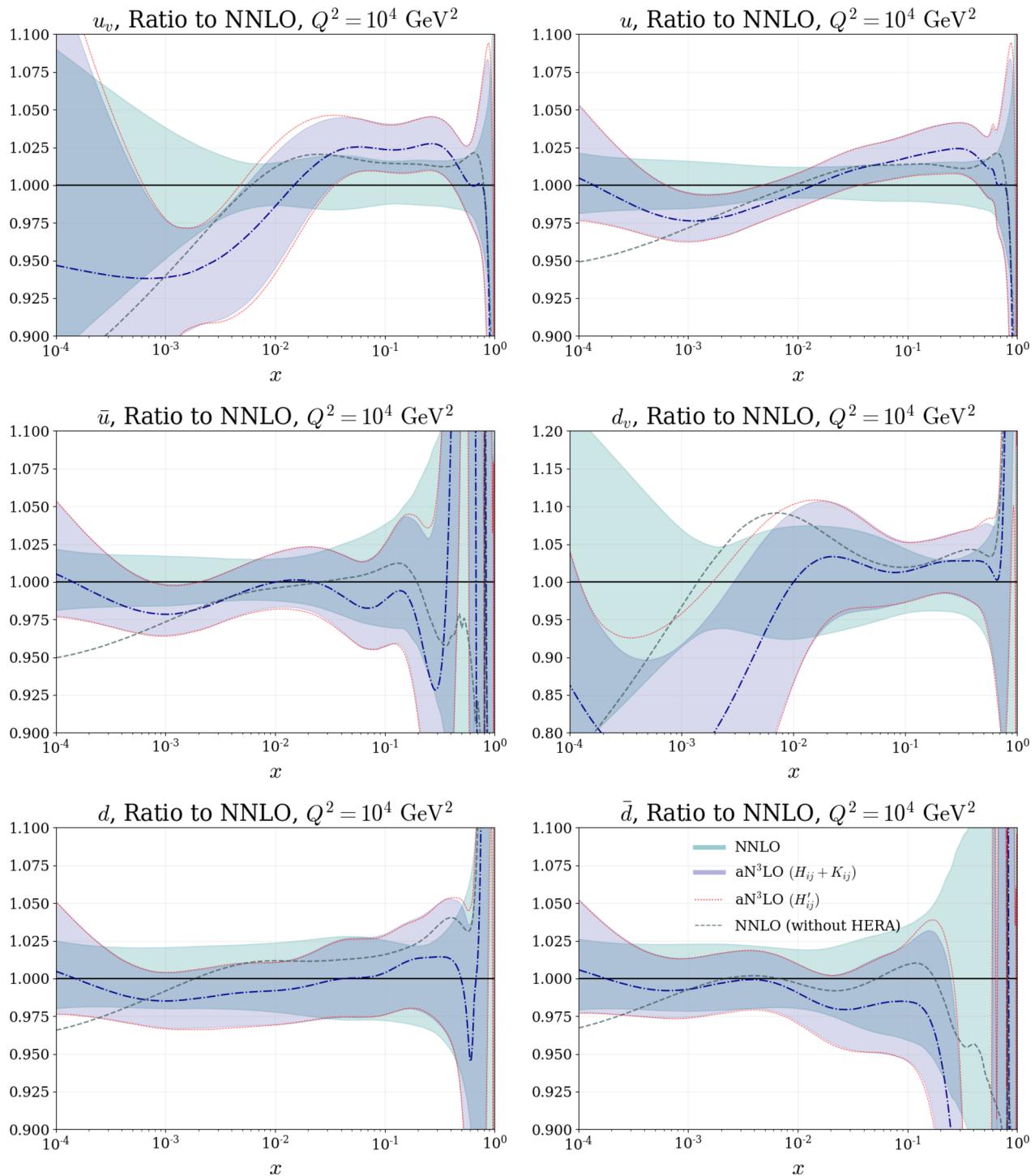
### 8.5 MSHT20aN<sup>3</sup>LO PDFs at $Q^2 = 2$ GeV<sup>2</sup>

Figure 39 compares the MSHT NNLO and  $aN^3LO$  PDF sets at  $Q^2 = 2$  GeV<sup>2</sup>. In this very low- $Q^2$  regime, some major differences are evident between NNLO and  $aN^3LO$  sets at  $Q^2 = 2$  GeV<sup>2</sup>, especially towards small- $x$ . For example, the gluon PDF is predicted to be much harder across this region, such that it is now positive across all  $x$  values considered

here. The effect of this can be immediately seen in the sea and heavy quarks.

Since the charm quark is directly coupled to the gluon PDF (through a convolution with  $A_{Hg}$ ), the charm PDF receives a notable enhancement at small- $x$  and also remains positive across all  $x$  values considered.<sup>20</sup> Another interesting feature is the reduction in uncertainty of the strange quark at small- $x$ . It may seem counter intuitive to have an uncertainty reduction by adding sources of theoretical uncertainty, however we should recall that the underlying theory has also been altered. Although one can expect an uncertainty increase in PDFs across  $(x, Q^2)$ , there are exceptions to this e.g. where tensions are relieved by introducing the  $N^3LO$  theory. The shift in the  $\chi^2$  landscape then has the potential to result in more precise regions of  $(x, Q^2)$  (in this case manifesting in an uncertainty reduction for the strange quark towards small- $x$ ).

<sup>20</sup> Since this is a convolution, it is the higher small- $x$  gluon, combined with the high- $x$  enhancement of  $A_{Hg}$  at  $N^3LO$  which gives rise to this increase in the charm PDF.



**Fig. 36** High- $Q^2$  ratio plots showing the aN<sup>3</sup>LO 68% confidence intervals with decorrelated ( $H_{ij} + K_{ij}$ ) and correlated ( $H'_{ij}$ )  $K$ -factor parameters, compared to NNLO 68% confidence intervals. Also shown are the

central values at NNLO when fit to all non-HERA datasets which show similarities with N<sup>3</sup>LO in the large- $x$  region of selected PDF flavours. All plots are shown for  $Q^2 = 10^4 \text{ GeV}^2$

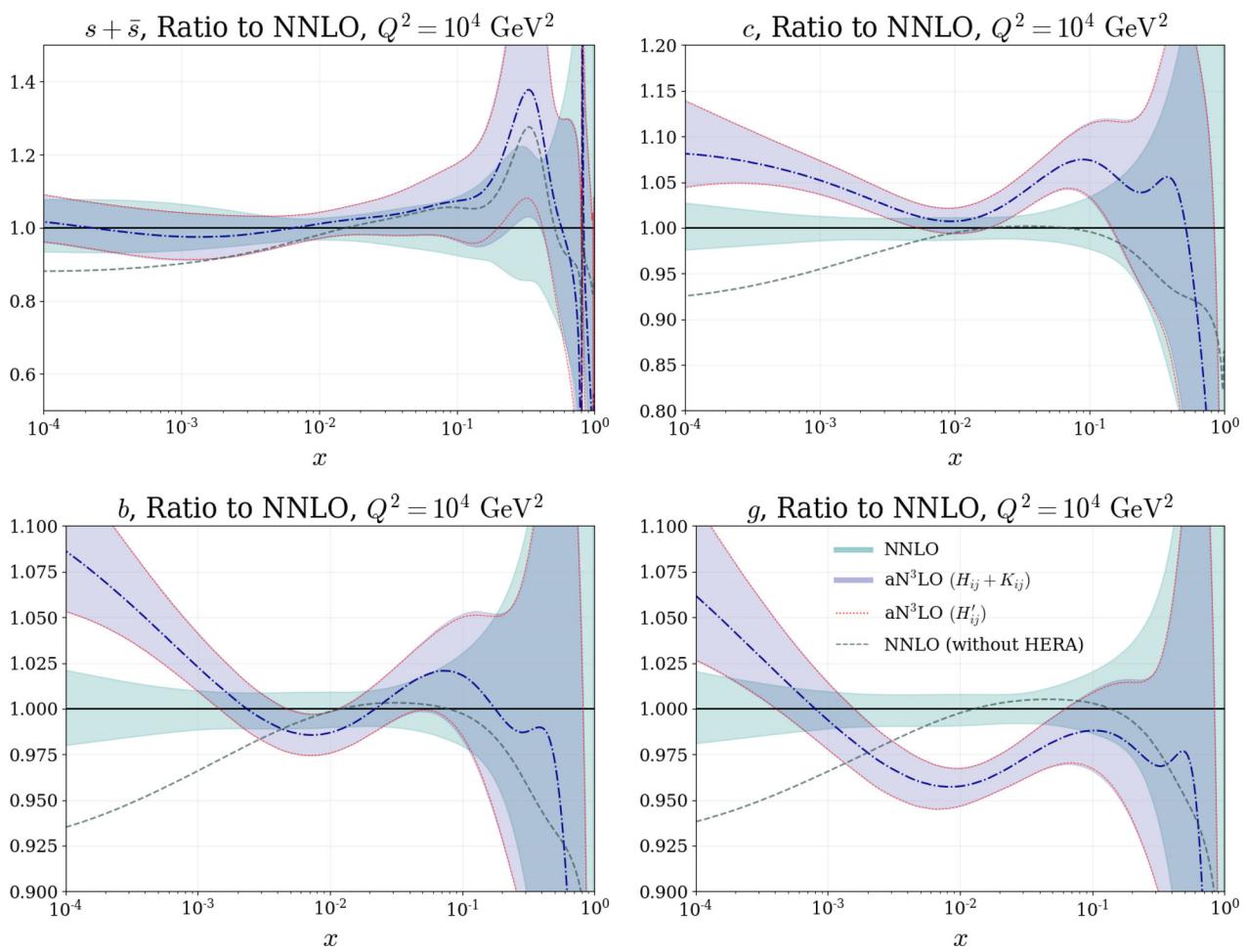
**Fig. 36** continued

Figure 40 displays the ratios of the  $aN^3LO$  MSHT PDFs to their NNLO counterparts at  $Q^2 = 2 \text{ GeV}^2$ . Here the specific shifts of each PDF are displayed more clearly. We note that there are many similar features shown here to those discussed for Figs. 35 and 36. Even in this very low- $Q^2$  regime, the uncertainty difference between correlated and decorrelated  $aN^3LO$   $K$ -factor PDF sets is minimal in all relevant regions of  $x$ .

#### 8.6 Effect of a $x < 10^{-3}$ cut at $aN^3LO$

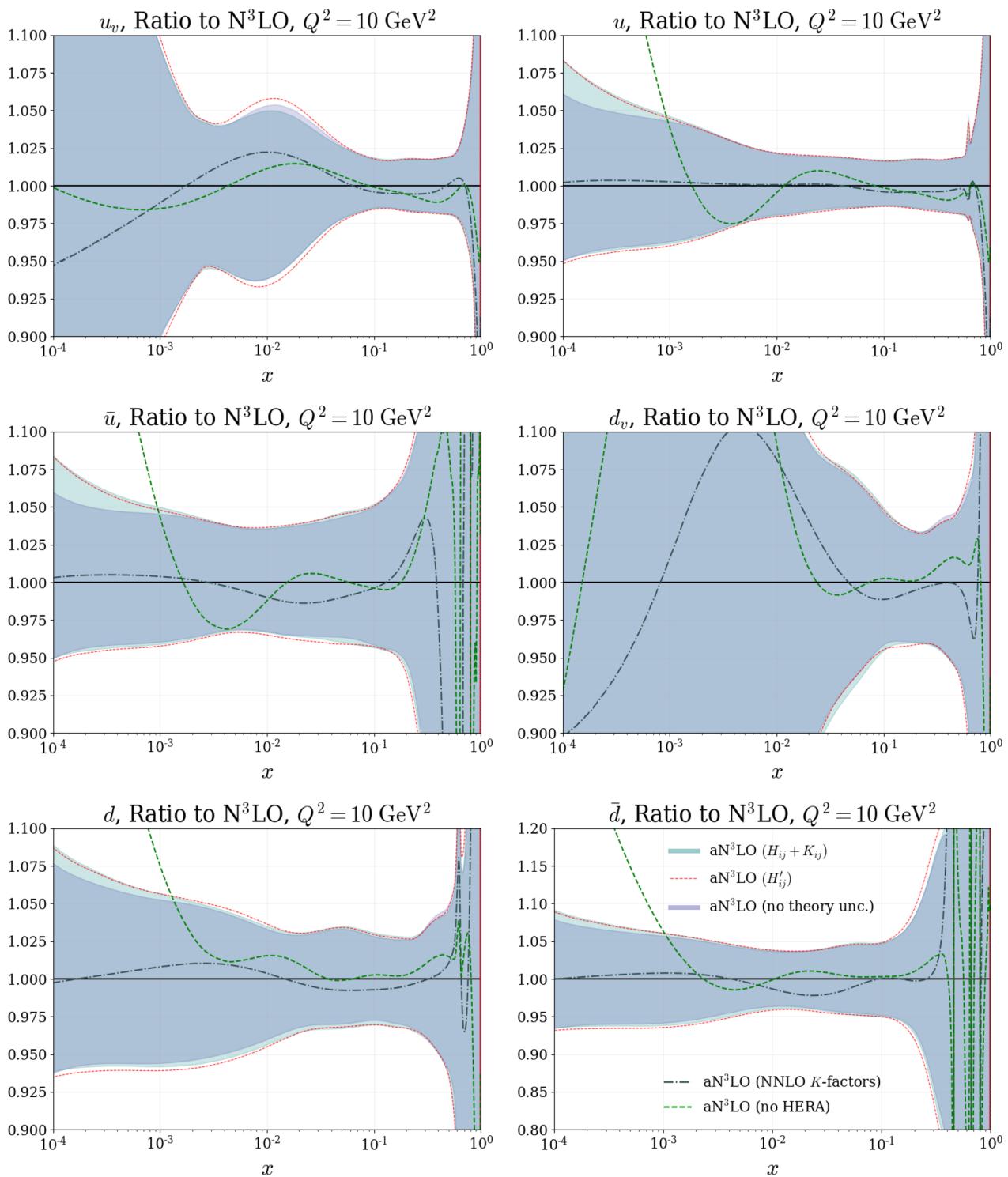
In this section we include results from a global PDF fit with small- $x$  ( $x < 10^{-3}$ ) data omitted. This analysis is provided to shed some light on the tensions between regions of  $x$  at  $aN^3LO$  while also providing some context with regards to the form of the PDFs in different regions of  $x$ .

Immediately, in Fig. 41, one can observe that omitting all small- $x$  data results in a set of less constrained PDFs for  $x < 10^{-3}$  (most notably in the gluon sector). However, also in Fig. 41 it can be observed that overall, the large- $x$  behaviour of these PDFs is very similar across both fits, indicating that the full fit is able to sufficiently fit both large and small- $x$  regions simultaneously. We provide this analysis as a cross-check to further support the reliability of our procedure, showing that the small- $x$  behaviour is not overwhelmingly attempting to fit to any all order, or specifically small- $x$  resummation, result at the expense of the large- $x$  description. We also note that while there is some definite change in the central values of the PDFs at small  $x$ , in most cases this is very well within uncertainties, and at most at the level of about one standard deviation, particularly for the gluon for  $x$  just below  $10^{-3}$ . This distinct, but limited shift in the best fit PDFs suggests that effects beyond  $N^3LO$  are certainly not insignificant at very low  $x$ , they are also not dominating the pull on the fit.

Following from the previous section, it is also interesting to examine where the  $aN^3LO$  theory contributions and their uncertainties reside after a global PDF fit.

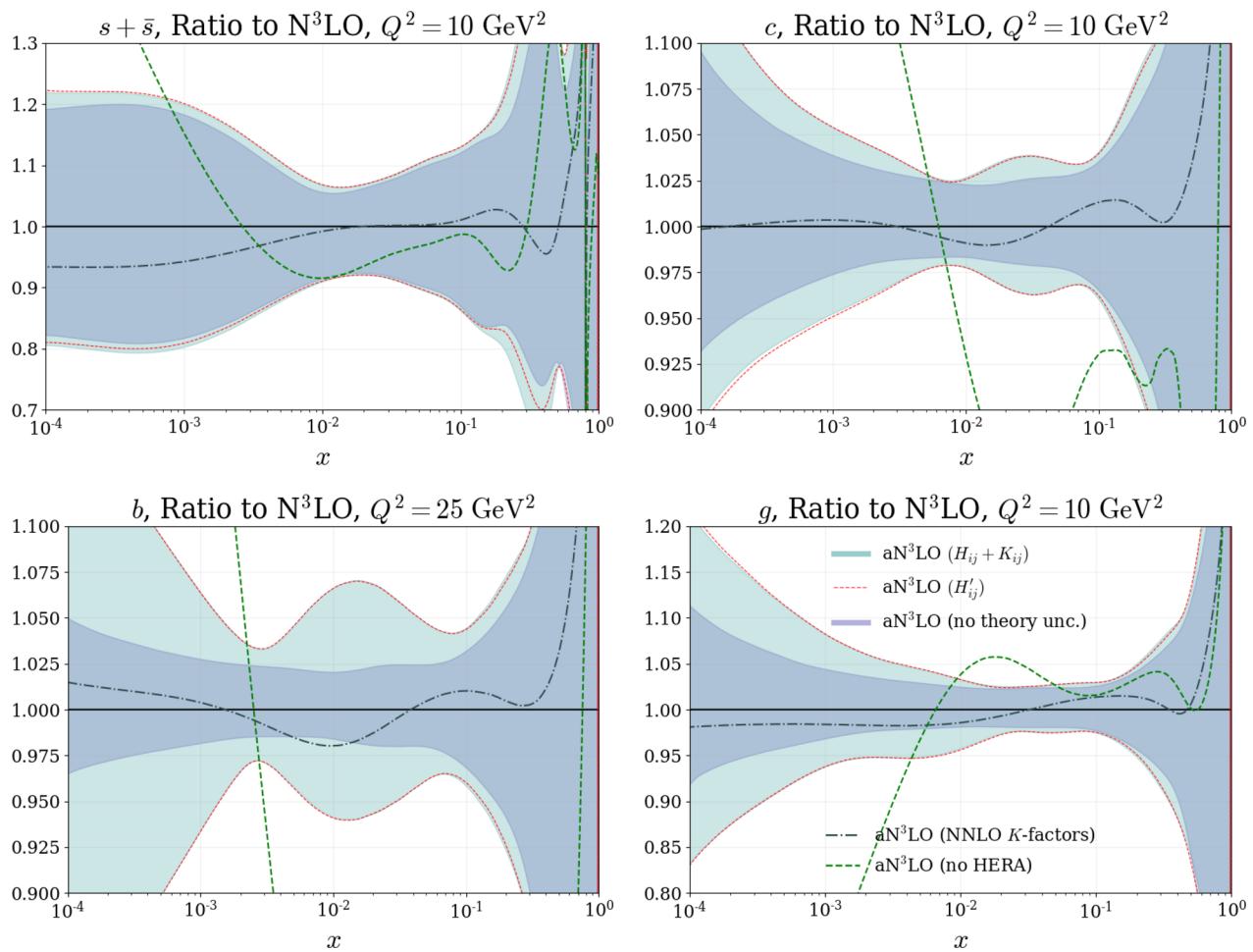
#### 8.7 Posterior $N^3LO$ theory parameters

Immediately, in Fig. 41, one can observe that omitting all small- $x$  data results in a set of less constrained PDFs for  $x < 10^{-3}$  (most notably in the gluon sector). However, also in Fig. 41 it can be observed that overall, the large- $x$  behaviour of these PDFs is very similar across both fits, indicating that the full fit is able to sufficiently fit both large and small- $x$  regions simultaneously. We provide this analysis as a cross-check to further support the reliability of our procedure, showing that the small- $x$  behaviour is not overwhelmingly attempting to fit to any all order, or specifically small- $x$  resummation, result at the expense of the large- $x$  description. We also note that while there is some definite change in the central values of the PDFs at small  $x$ , in most cases this is very well within uncertainties, and at most at the level of about one standard deviation, particularly for the gluon for  $x$  just below  $10^{-3}$ . This distinct, but limited shift in the best fit PDFs suggests that effects beyond  $N^3LO$  are certainly not insignificant at very low  $x$ , they are also not dominating the pull on the fit.



**Fig. 37** Low- $Q^2$  ratio plots showing the aN<sup>3</sup>LO 68% confidence intervals with decorrelated and correlated  $K$ -factor parameters, compared to the aN<sup>3</sup>LO central value. Also shown are the central values at aN<sup>3</sup>LO

when fit to all non-HERA datasets and the central values with all  $K$ -factors set at NNLO. All plots are shown for  $Q^2 = 10 \text{ GeV}^2$  with the exception of the bottom quark shown for  $Q^2 = 25 \text{ GeV}^2$

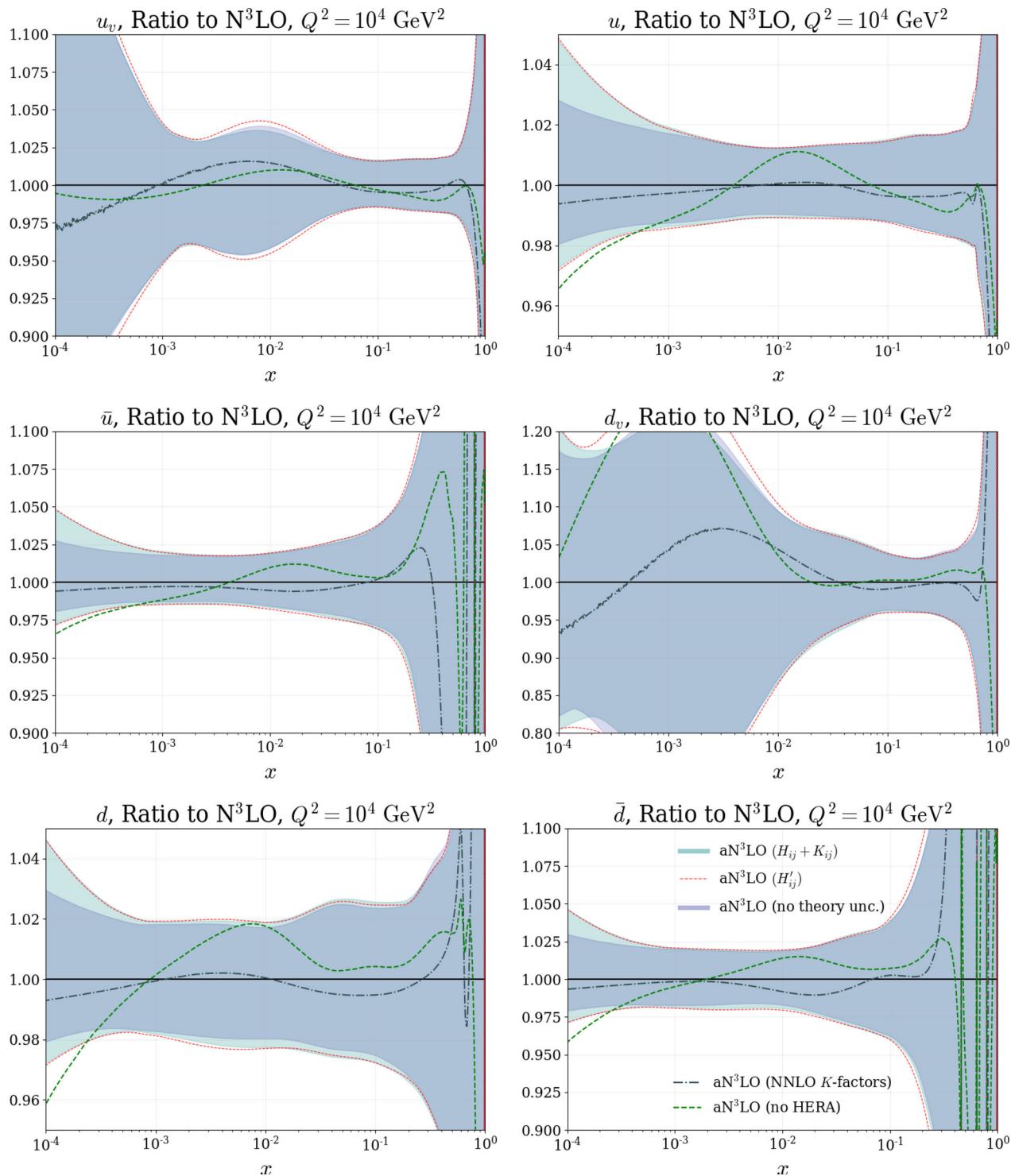
**Fig. 37** continued

Displayed in Table 13 are the predicted posterior limits on each (non  $K$ -factor) aN<sup>3</sup>LO theory parameter. Here one can directly compare these variations with the prior variations decided in earlier sections. Also provided is a comparison of these posterior limits across a fit with and without small- $x$  ( $x < 10^{-3}$ ) data included. This comparison complements the previous section by showing a similar trend in the central values predicted in both cases (i.e. with an overlap of uncertainties). As with the PDFs, there is some significant evidence of tensions, but these are not severe, and the central values of many parameters are extremely stable. Furthermore this is evidence that the small- $x$  behaviour is influencing, but not likely to be dominating the behaviour of aN<sup>3</sup>LO parameters in a manner which is significantly adversarial to the preference of data at  $x > 10^{-3}$ .

Figure 42 displays a comparison of the prior and posterior variations predicted for the perturbative expansions of the rel-

event splitting functions and transition matrix elements discussed in Sects. 4 and 5. We exclude the non-singlet quantities from this comparison as the variations predicted for these quantities are very similar to their priors (as can be seen in Table 13) and have a small overall effect on the PDFs. It is true that once a fit is performed, the variation of the aN<sup>3</sup>LO theoretical nuisance parameters becomes less sensitive to the prior variation, suggesting that the initial uncertainty estimate was conservative. Nevertheless in Fig. 42, one can observe that all posterior variations overlap with their corresponding priors, in most cases quite considerably. We also note that the most drastic differences between prior and posterior variations are as expected relating to the gluon PDF.

Figure 43 contains a comparison between the aN<sup>3</sup>LO functions posterior variations with and without small- $x$  ( $x < 10^{-3}$ ) data included in a global fit. These results accompany those presented in Table 13 and further display the reasonable



**Fig. 38** High- $Q^2$  ratio plots showing the aN<sup>3</sup>LO 68% confidence intervals with decorrelated and correlated  $K$ -factor parameters, compared to the aN<sup>3</sup>LO central value. Also shown are the central values at aN<sup>3</sup>LO

when fit to all non-HERA datasets and the central values with all  $K$ -factors set at NNLO. All plots are shown for  $Q^2 = 10^4$  GeV<sup>2</sup>

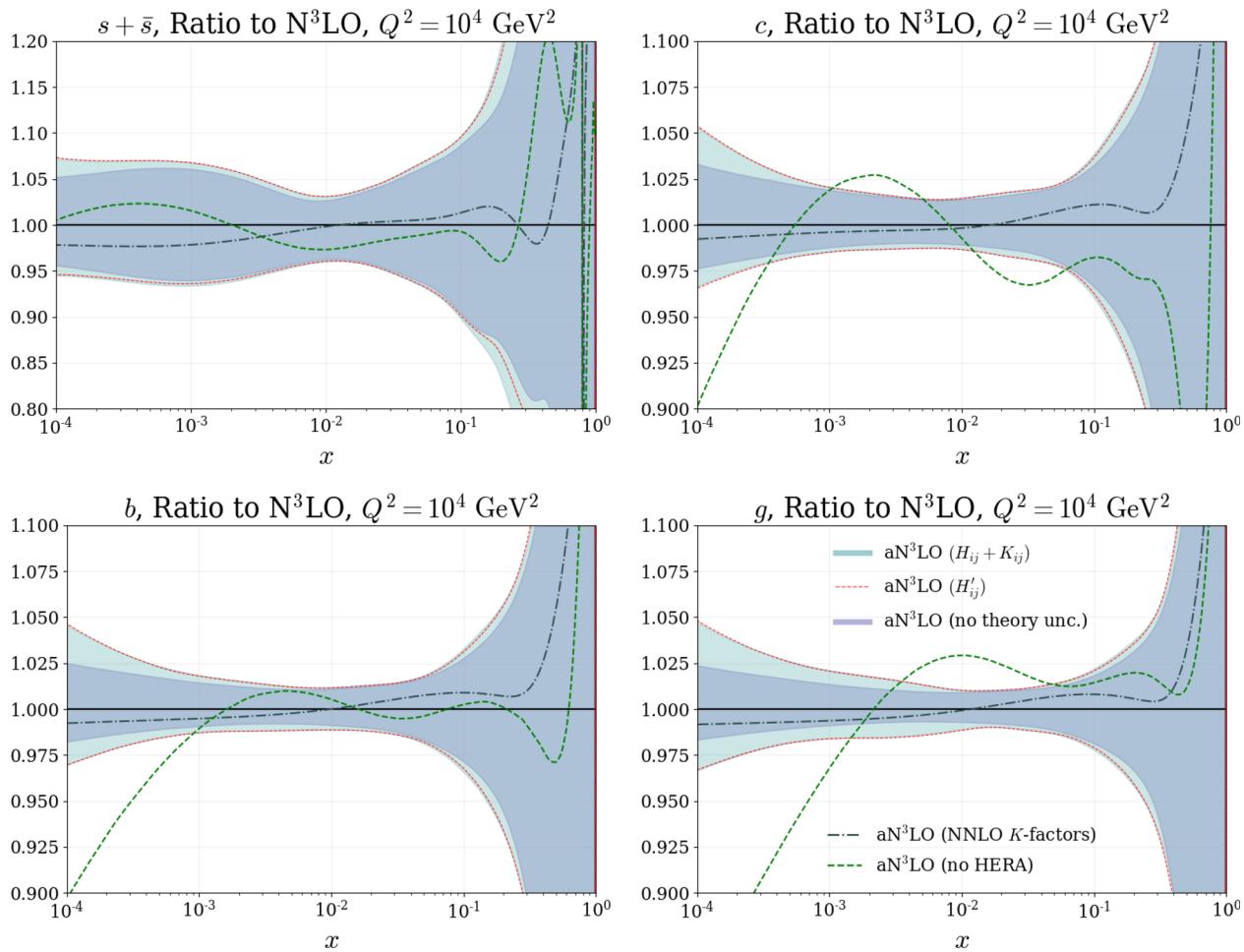
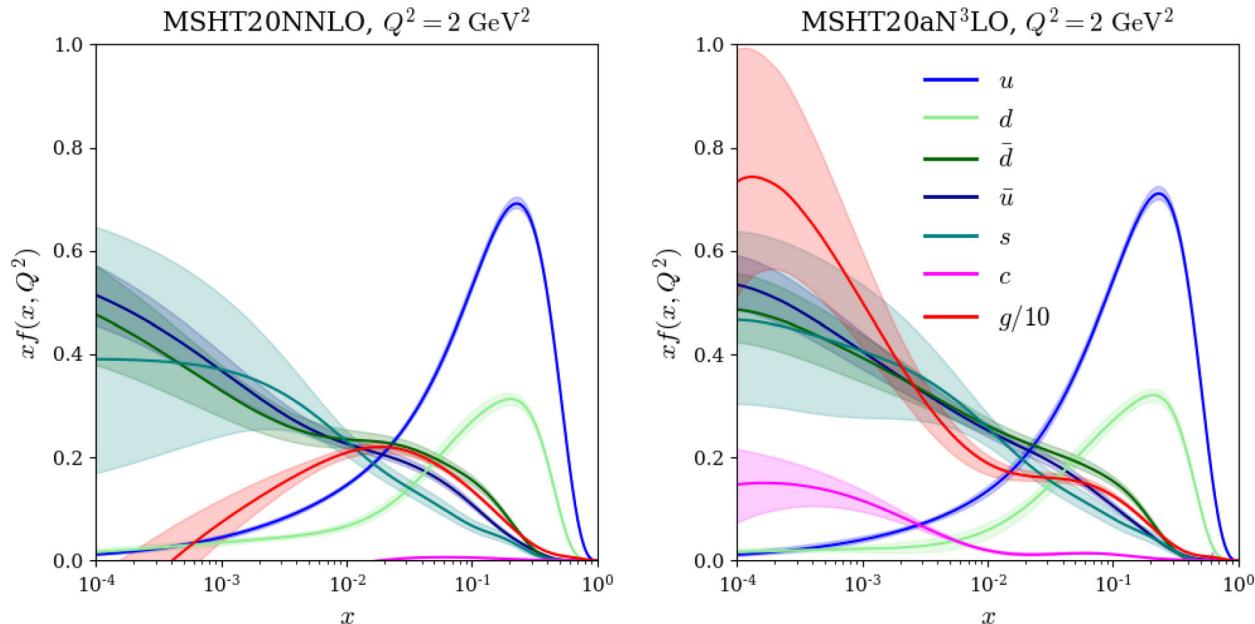
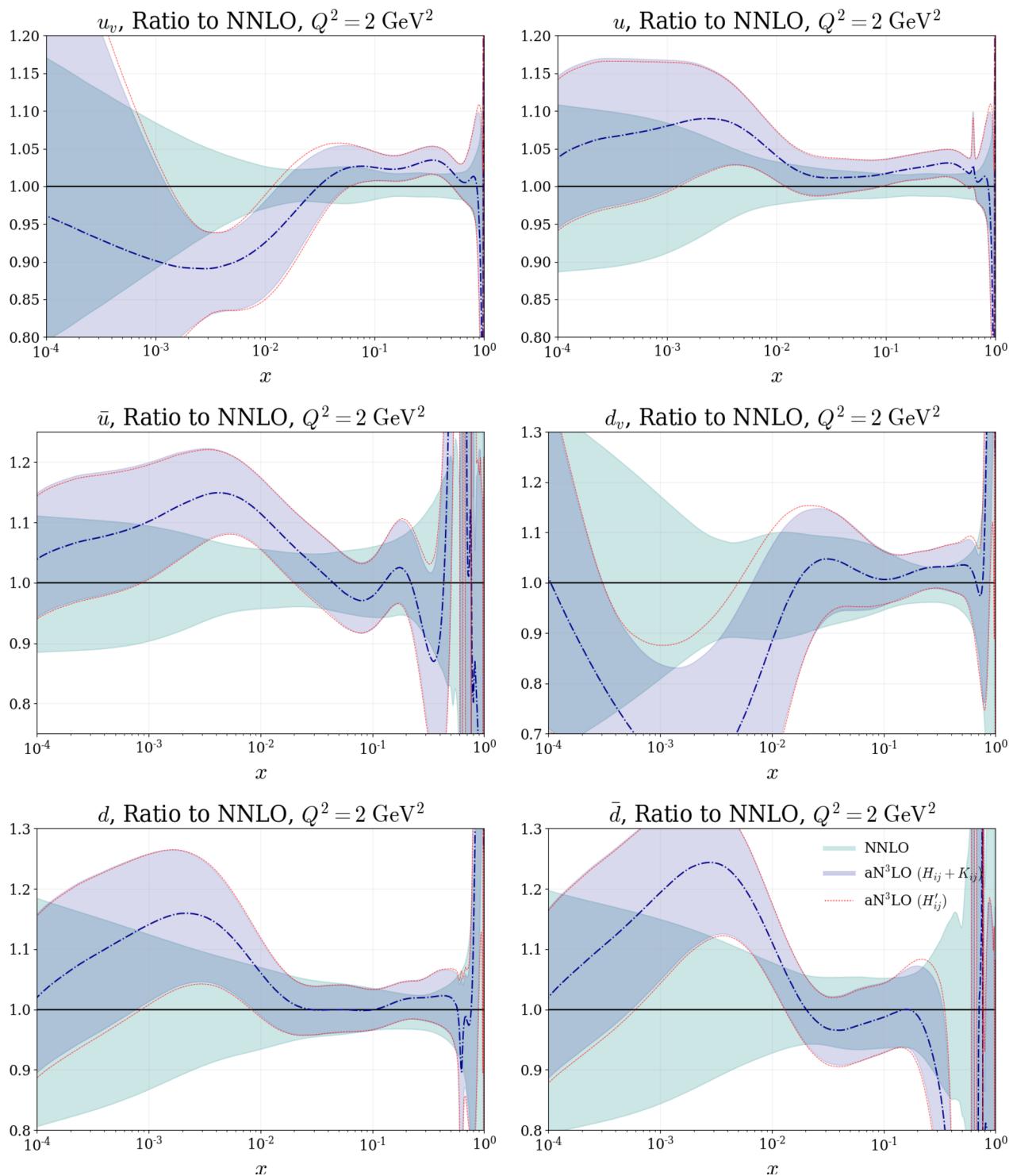
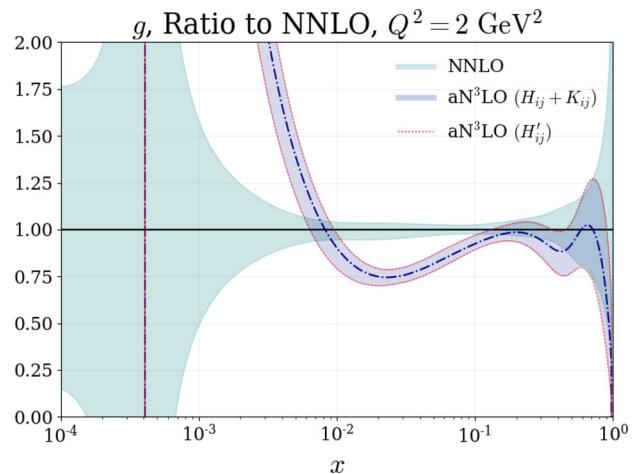
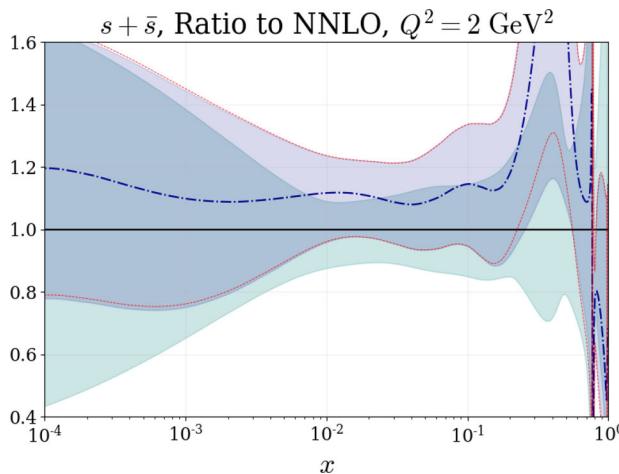


Fig. 38 continued

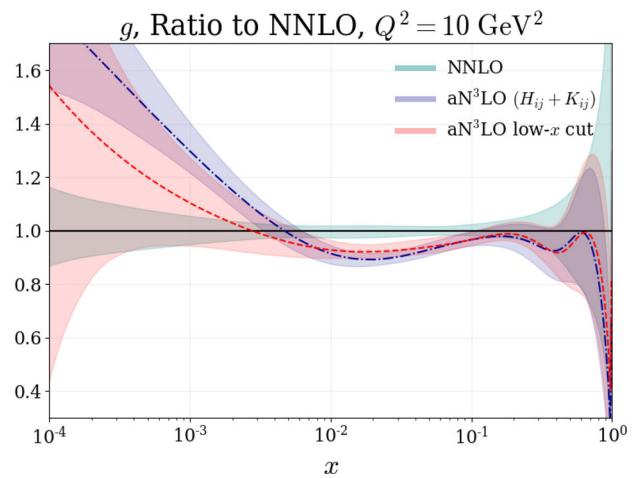
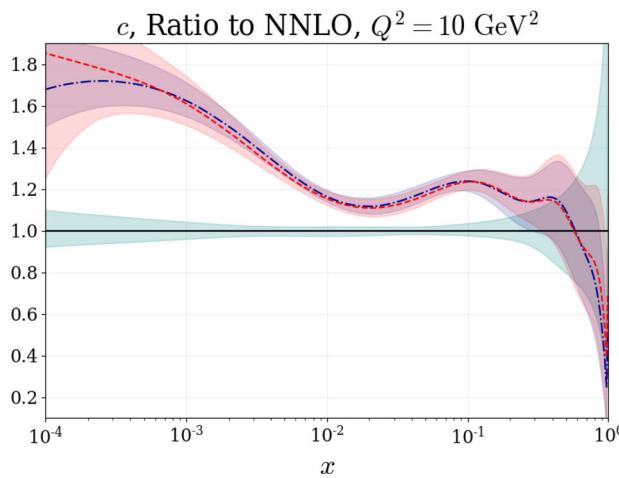
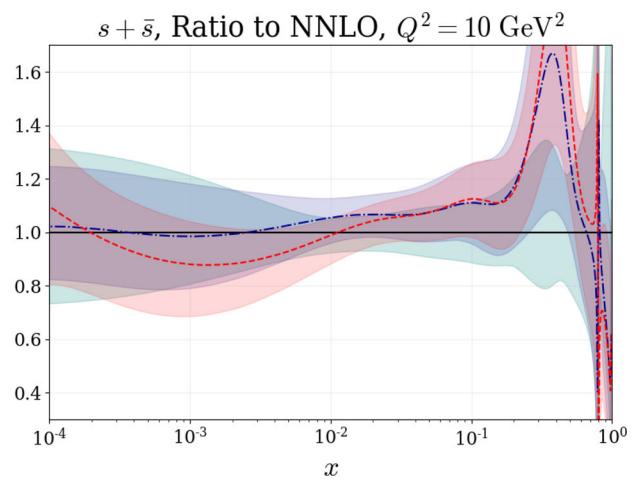
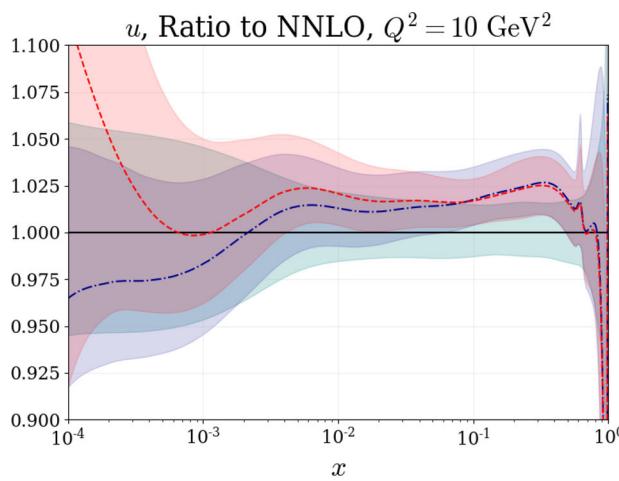
Fig. 39 General forms of NNLO (left) and aN<sup>3</sup>LO (right) PDFs at  $Q^2 = 2 \text{ GeV}^2$ . Axis are set to the same scale to highlight the main differences between NNLO and aN<sup>3</sup>LO. Specifically in the gluon and heavy flavour sectors



**Fig. 40** Very low- $Q^2$  ratio plots showing the aN<sup>3</sup>LO 68% confidence intervals with decorrelated and correlated  $K$ -factor parameters, compared to NNLO 68% confidence intervals. All plots are shown for  $Q^2 = 2 \text{ GeV}^2$



**Fig. 40** continued



**Fig. 41** Low- $Q^2$  PDF ratios showing  $\text{aN}^3\text{LO}$  PDFs fitted with and without small- $x$  ( $< 10^{-3}$ ) data included in a global fit. All plots are shown for  $Q^2 = 10 \text{ GeV}^2$  with the exception of the bottom quark shown for  $Q^2 = 25 \text{ GeV}^2$

**Table 13** Posterior predicted  $\pm 1\sigma$  limits on aN<sup>3</sup>LO theoretical nuisance parameters for splitting functions, transition matrix elements and coefficient functions

Parameter	Default fit			Small- $x$ cut fit		
	Central	+ Limit	- Limit	Central	+ Limit	- Limit
Low- $Q^2$ coefficient						
$c_q^{\text{NLL}}$	-3.868	-1.891	-6.132	-5.822	-4.333	-10.373
$c_g^{\text{NLL}}$	-5.837	-4.444	-7.429	-6.995	-3.701	-7.991
Transition matrix elements						
$a_{qg,H}^{\text{NS}}$	-64.411	-38.778	-91.850	-65.103	-40.073	-93.225
$a_{Hg}$	12214.000	12966.856	11279.376	12524.000	13831.976	11286.843
$a_{gg,H}$	-1951.600	-1577.155	-3418.568	-1392.600	-512.817	-2190.354
Splitting functions						
$\rho_{qg}^{\text{NS}}$	0.007	0.015	-0.002	0.006	0.020	-0.005
$\rho_{qg}^{\text{PS}}$	-0.501	-0.254	-0.644	-0.505	-0.285	-0.692
$\rho_{qg}$	-1.754	-1.157	-1.897	-1.309	-0.620	-1.881
$\rho_{gg}$	-1.784	-1.548	-2.212	-1.622	-1.367	-1.877
$\rho_{gg}$	19.245	21.505	9.025	12.997	16.142	6.611

agreement between the two fits, but there is some degree of tension occurring mainly in the cases of  $P_{gg}$ ,  $A_{Hg}$  and  $A_{gg,H}$ . In all cases the predicted variations overlap, with most central values being stable (i.e. contained well within the uncertainty predictions). However, for  $P_{gg}$ ,  $A_{Hg}$  and  $A_{gg,H}$  the fit with the small  $x$  cut does result in posterior functions which are more consistent with the prior functions, again suggesting that for these functions the posterior values are influenced, to a significant, but not overwhelming extent by terms beyond N<sup>3</sup>LO, most likely those associated with small- $x$  resummation. Hence, as with the PDFs this provides evidence that the aN<sup>3</sup>LO predictions are reasonably consistent across all values of  $x$  but are influenced to a limited extent by the small- $x$  region. This supports our view that while we are explicitly determining the missing N<sup>3</sup>LO corrections, which are indeed overall the dominant part of the missing higher order corrections, the fit is also probing some even higher order corrections, particularly at small  $x$ .

### 8.8 N<sup>3</sup>LO contributions

In this section all but one N<sup>3</sup>LO contribution will be switched off, in particular only splitting functions, or only heavy or light flavour coefficient functions with their relevant transition matrix elements. In all cases the aN<sup>3</sup>LO  $K$ -factors are left free to allow the fit some freedom in manipulating the cross sections of other datasets. In practice however, fixing these  $K$ -factors at the NNLO values has a minimal effect on the shape of the PDFs in all cases (as demonstrated in Figs. 37 and 38).

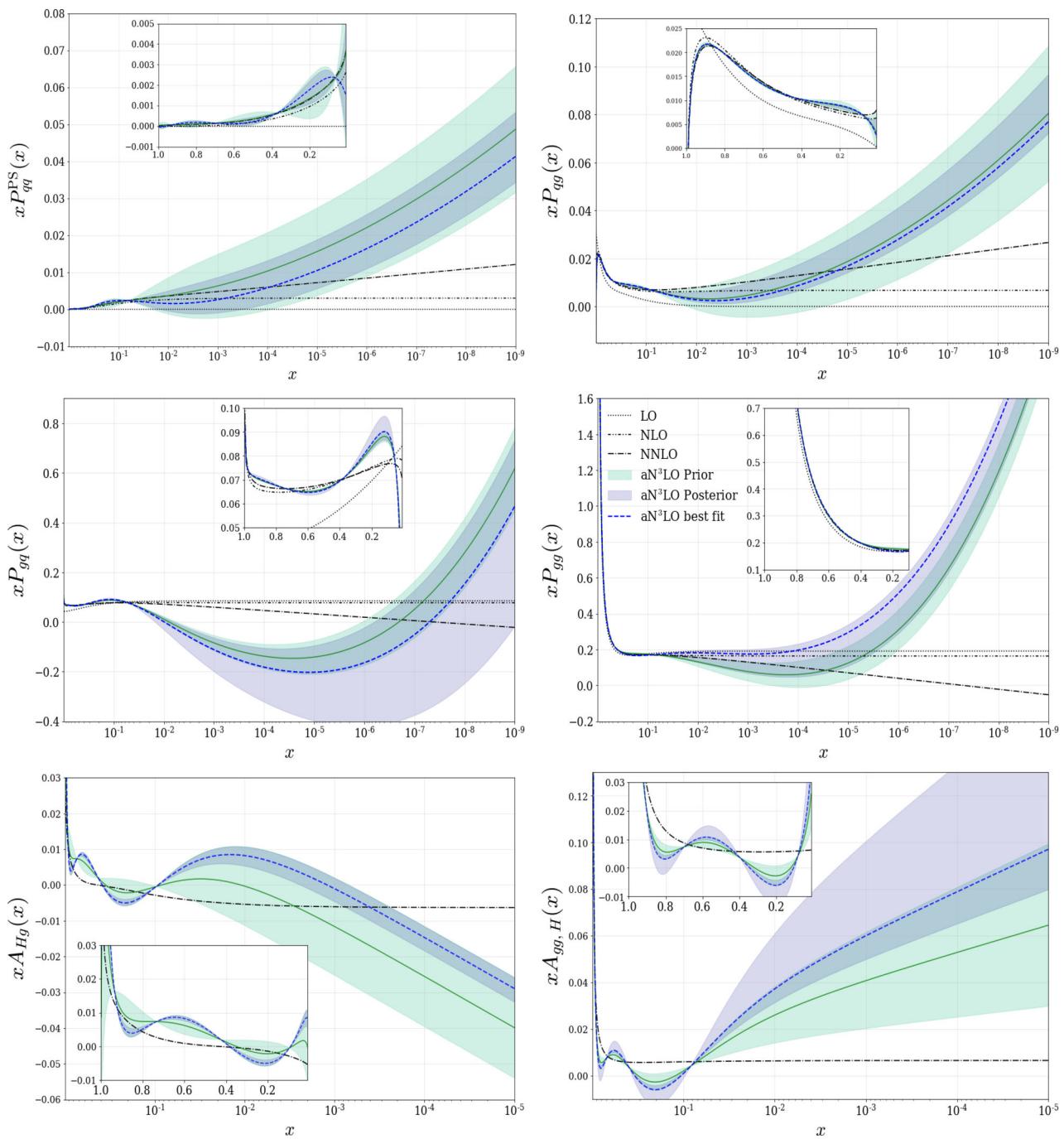
The deconstructed aN<sup>3</sup>LO PDFs as a ratio to the NNLO MSHT PDFs for various flavours at  $Q^2 = 10 \text{ GeV}^2$  (with the

bottom quark given at  $Q^2 = 25 \text{ GeV}^2$ ) are shown in Fig. 44. Across the more tightly constrained light quark PDFs, all contributions lie very close to the aN<sup>3</sup>LO  $\pm 1\sigma$  uncertainty bands (blue shaded region and solid line). The additive and compensating nature of these contributions is also clear in a handful of the ratios from Fig. 44. In other areas the full description is biased towards a single contribution, for example the charm and bottom quarks follow the contribution from heavy flavours as one may expect. Conversely, to some extent the gluon follows the splitting functions much more closely as these contributions indirectly couple the gluon to the more constraining data.<sup>21</sup>

### 8.9 $\alpha_s$ variation

As in the standard MSHT20 NNLO PDF fit, we present the best fit aN<sup>3</sup>LO PDFs with  $\alpha_s(m_Z) = 0.118$ , the common value chosen in the PDF4LHC combination [8]. However, investigating the true minima in  $\alpha_s(m_Z)$ , the  $\chi^2$  profiles in Fig. 45 prefer a value of around  $\alpha_s(m_Z) = 0.1170$ . This result follows the trend from lower orders whereby the best fit values are  $\alpha_s(m_Z) = 0.1174 \pm 0.0013$  at NNLO and  $\alpha_s(m_Z) = 0.1203 \pm 0.0015$  at NLO [151]. Following from NNLO, the aN<sup>3</sup>LO  $\alpha_s(m_Z)$  prediction is also slightly lower than the NNLO world average central value at around  $\alpha_s(m_Z) = 0.1179 \pm 0.0010$  [152]. In any case, the preferred aN<sup>3</sup>LO  $\alpha_s(m_Z)$  value stated here is in agreement with the MSHT20 NNLO result and the world average within uncertainties. A full analysis is left for a future publication.

<sup>21</sup> An exception to this can be seen around  $x \sim 10^{-2}$  where the contributions act cumulatively. We make this point as this region of  $x$  is of interest for Higgs calculations such as those discussed in Sect. 9.



**Fig. 42** Posterior variations of the aN<sup>3</sup>LO splitting functions and transition matrix elements predicted from a full global fit (blue shaded band) compared to the prior variations in each case (green shaded band)

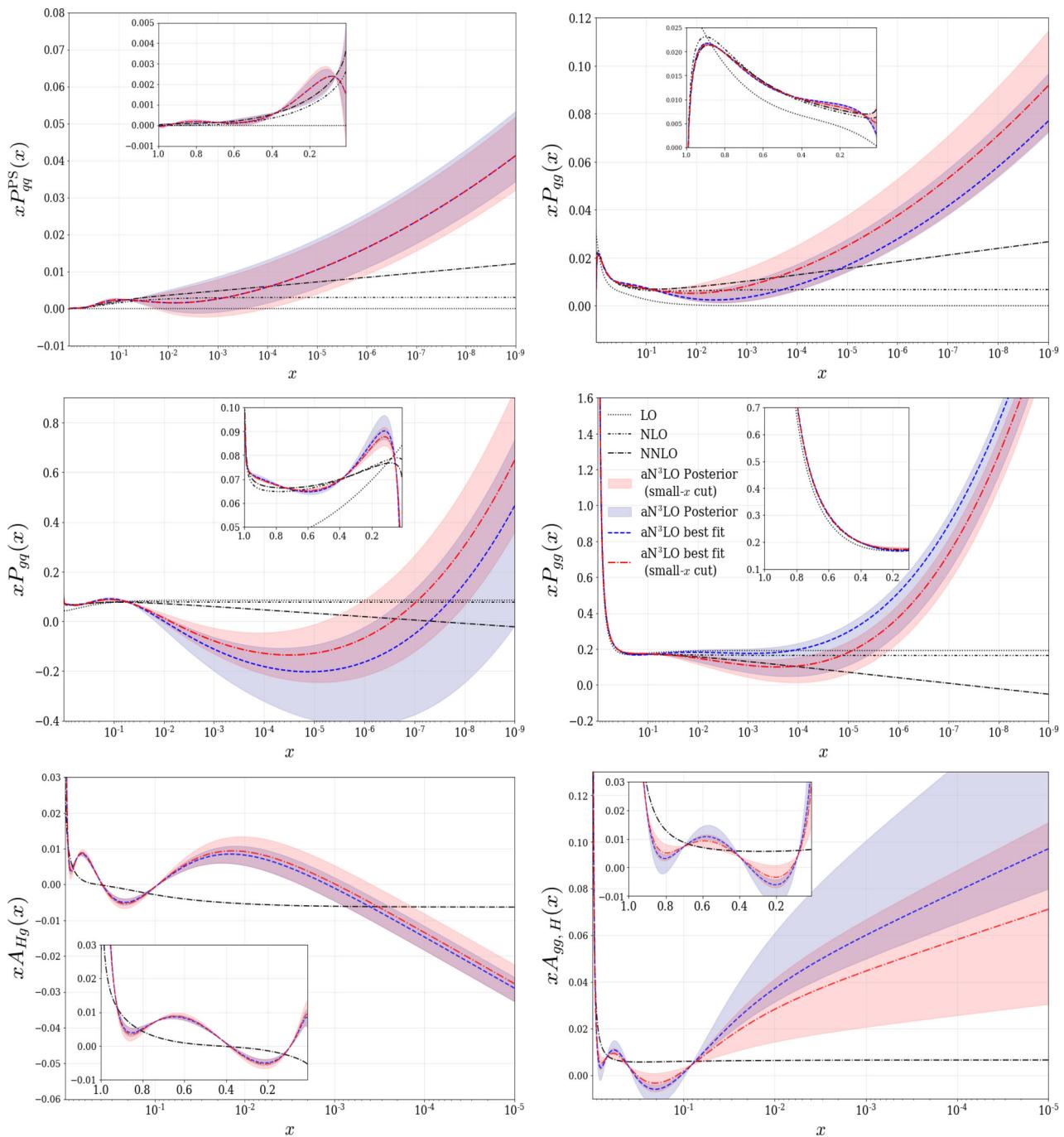
## 8.10 Charm mass dependence

In a standard MSHT fit [3], aN<sup>3</sup>LO PDFs are produced with the charm pole mass  $m_c = 1.40$  GeV. Figure 46 displays the  $\chi^2$  results when varying this charm mass. The predicted minimum at NNLO (for MSHT20 PDFs) is in the range  $m_c = 1.35 - 1.40$  GeV [151], whereas at aN<sup>3</sup>LO we show a minimum in the region of  $m_c = 1.42 - 1.47$  GeV. This

aN<sup>3</sup>LO result therefore shows a slightly better agreement with the world average [152]<sup>22</sup> of  $m_c = 1.5 \pm 0.2$  GeV.

Considering Fig. 47, one is then able to analyse the effect of this slightly higher charm mass on the form of the PDFs.

<sup>22</sup> There is some ambiguity in this value since the transformation from  $\overline{MS}$  to the pole mass definition is not well-defined (see [151] for more details).

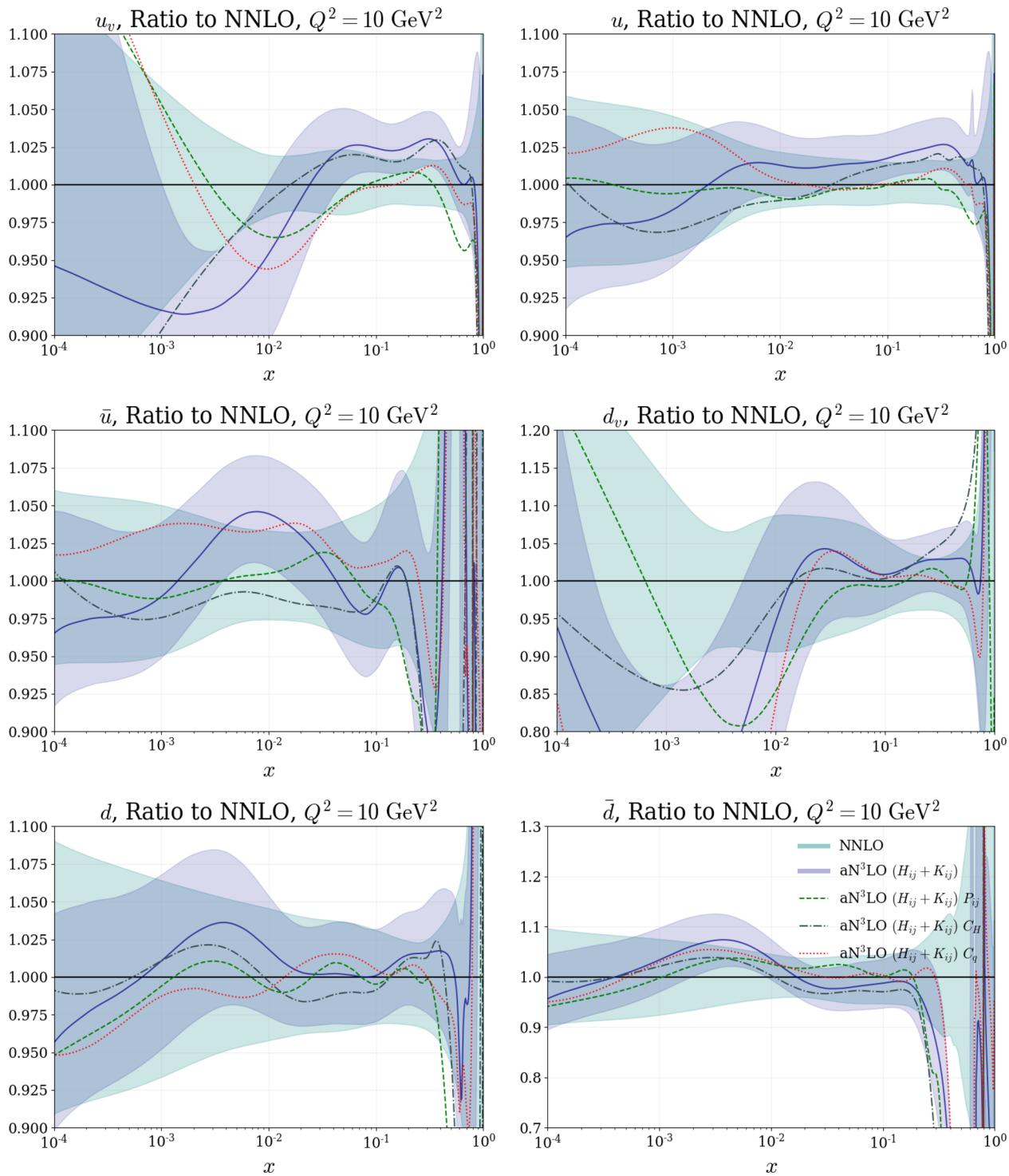


**Fig. 43** Posterior variations of the  $aN^3LO$  splitting functions and transition matrix elements predicted from a full global fit (blue shaded band) compared to a fit with small- $x$  ( $x < 10^{-3}$ ) data removed from a fit (red shaded band)

As one can expect, the charm PDF is subject to the largest difference and is suppressed by a higher  $m_c$ . The extra suppression from a higher charm mass allows the fit to suppress the  $c + \bar{c}$  sea contribution. This is then compensated by an increase in the  $\bar{u}$  and  $\bar{d}$  distributions which stabilises the overall sea contribution.

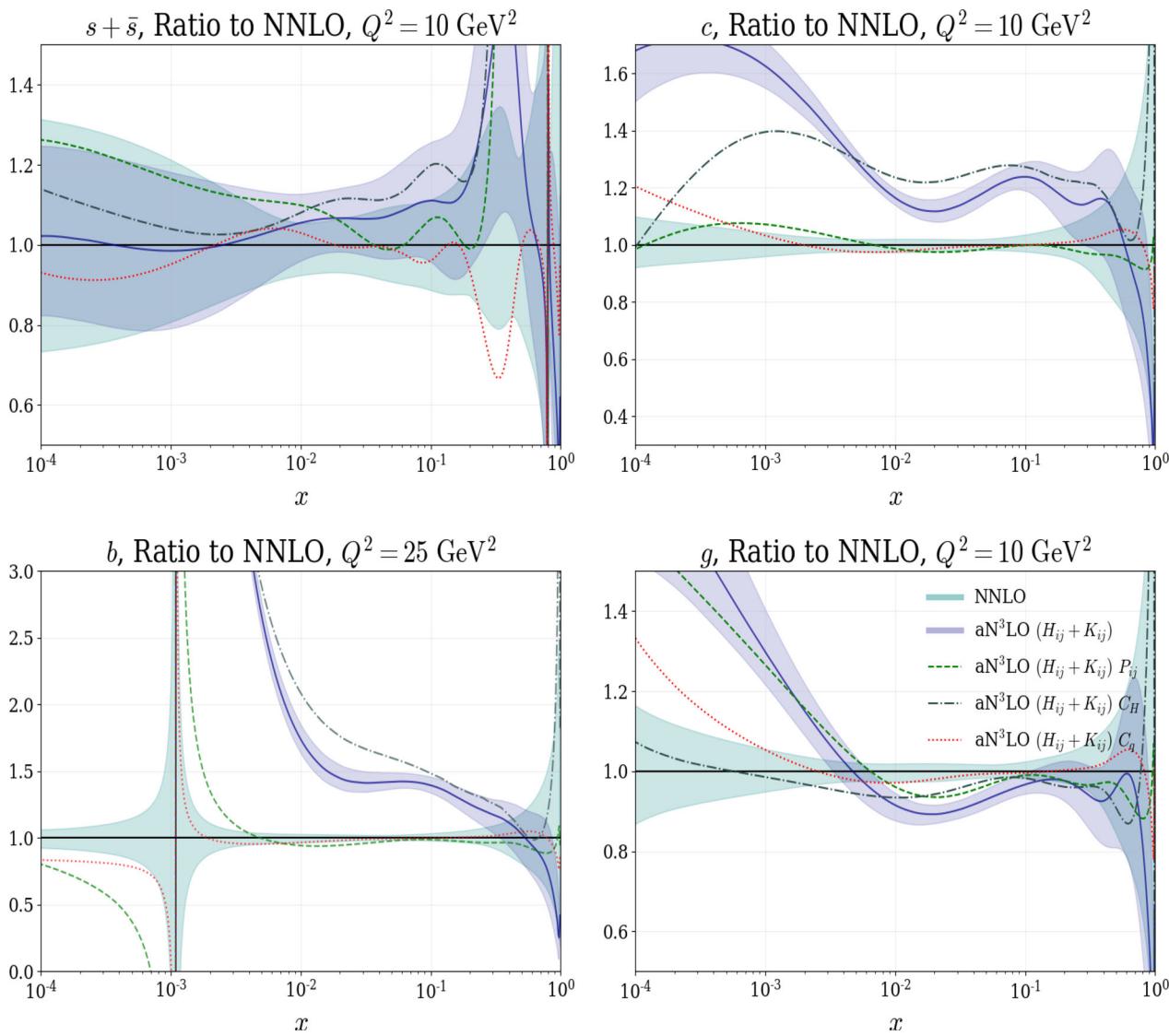
## 9 $N^3LO$ predictions

With the increasing number of hard cross section calculations at  $N^3LO$ , there is a growing demand for  $N^3LO$  accuracy in PDFs. In this section we investigate the effect of the MSHT approximate  $N^3LO$  PDFs on Higgs production via gluon fusion and vector boson fusion (VBF). The hard cross sections for these processes have been calculated to  $N^3LO$



**Fig. 44** Low- $Q^2$  PDF ratios showing  $aN^3LO$  (with decorrelated  $K$ -factors) 68% confidence intervals compared to NNLO 68% confidence intervals with varying theory contributions. All plots are shown for  $Q^2 = 10 \text{ GeV}^2$  with the exception of the bottom quark shown for  $Q^2 = 25 \text{ GeV}^2$ . The PDFs included are: NNLO (green shaded),

All  $N^3LO$  contributions (blue shaded), only splitting functions (green dashed), only heavy flavour coefficient functions and transition matrix elements (dark grey dash-dot) and only light flavour coefficient functions and transition matrix elements (red dotted)



**Fig. 44** continued

accuracy [67–76, 153, 154]. We present a full  $N^3LO$  computation for each prediction with our approximate  $N^3LO$  PDFs, including theoretical uncertainties. In future work, the intention will be to expand this analysis to include results for  $N^3LO$  DY [63] and approximate  $N^3LO$  top production [66] cross sections.

Note that in this section we follow the notation used previously and denote the  $aN^3LO$  results with decorrelated  $K$ -factors as  $(H_{ij} + K_{ij})^{-1}$  and those with correlated  $K$ -factors with  $H_{ij}'^{-1}$ . In all cases, scale variations are found via the 9-point prescription [11] for results with NNLO PDFs. Whereas for  $aN^3LO$  PDFs, although the extra information introduced is at  $N^3LO$ , the data (and therefore all relevant theory nuisance parameters) which are included in the global fit are sensitive to all orders. In particular, we include theoretical uncertainties into our  $aN^3LO$  fit which incorporate

MHO effects on the PDFs. Therefore we argue (and in these cases demonstrate) that the factorisation scale variation is contained within the PDF uncertainties. Due to this, it is only the renormalisation scale which requires variation in predictions involving  $aN^3LO$  PDFs.<sup>23</sup>

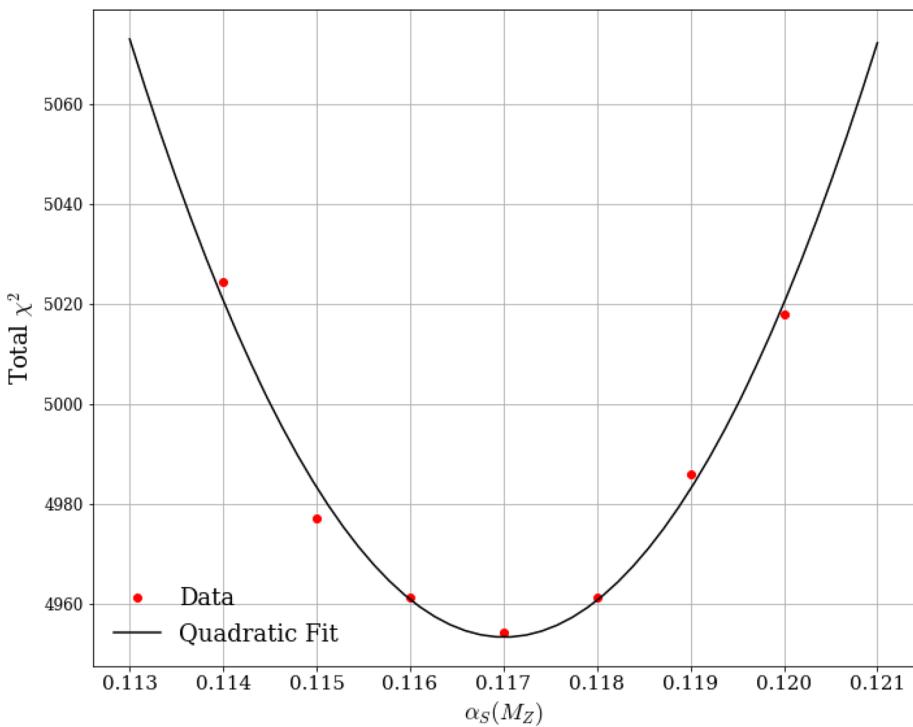
### 9.1 Higgs production – gluon fusion: $gg \rightarrow H$

Table 14 and Fig. 48 (left) show predictions at a central scale of  $\mu = \mu_f = \mu_r = m_H/2$  for the Higgs production cross section via gluon fusion<sup>24</sup> at the LHC for  $\sqrt{s} = 13$  TeV, where  $m_H = 125$  GeV is the Higgs mass and no fiducial

<sup>23</sup> This is to quantify the theoretical MHOU in the hard cross section, whereas the  $aN^3LO$  PDFs now come with an estimated MHOU.

<sup>24</sup> Results are obtained with the code `ggHiggs` [67–76, 155].

**Fig. 45** Quadratic fit to the total  $\chi^2$  results from various  $\alpha_s(m_Z)$  starting scales. The minimum of the quadratic fit provides a rough estimate of  $\alpha_s(m_Z) = 0.1170$  at aN<sup>3</sup>LO

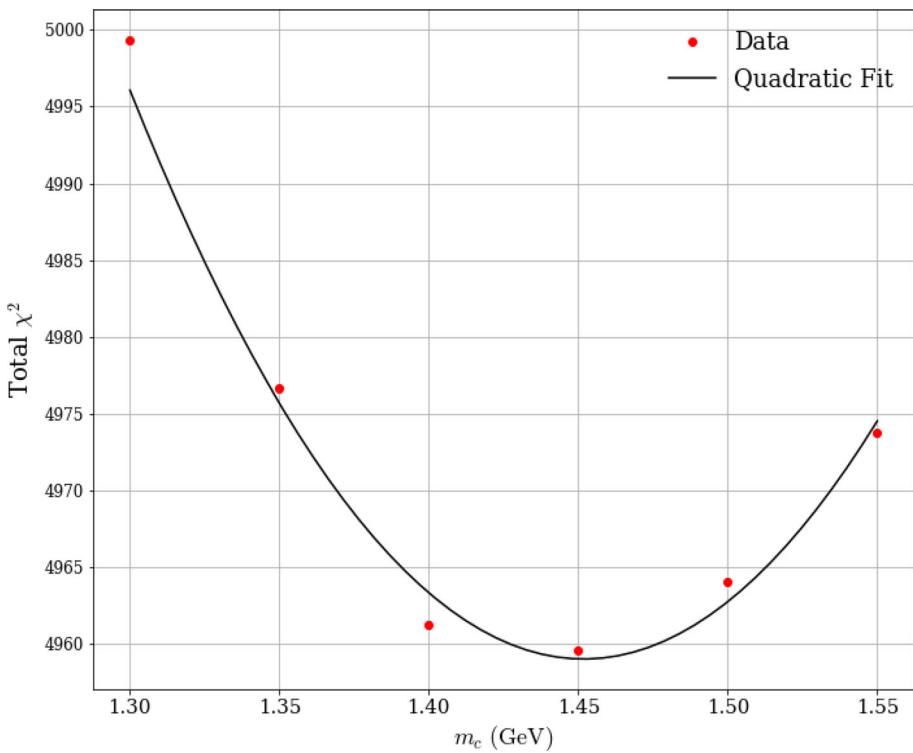


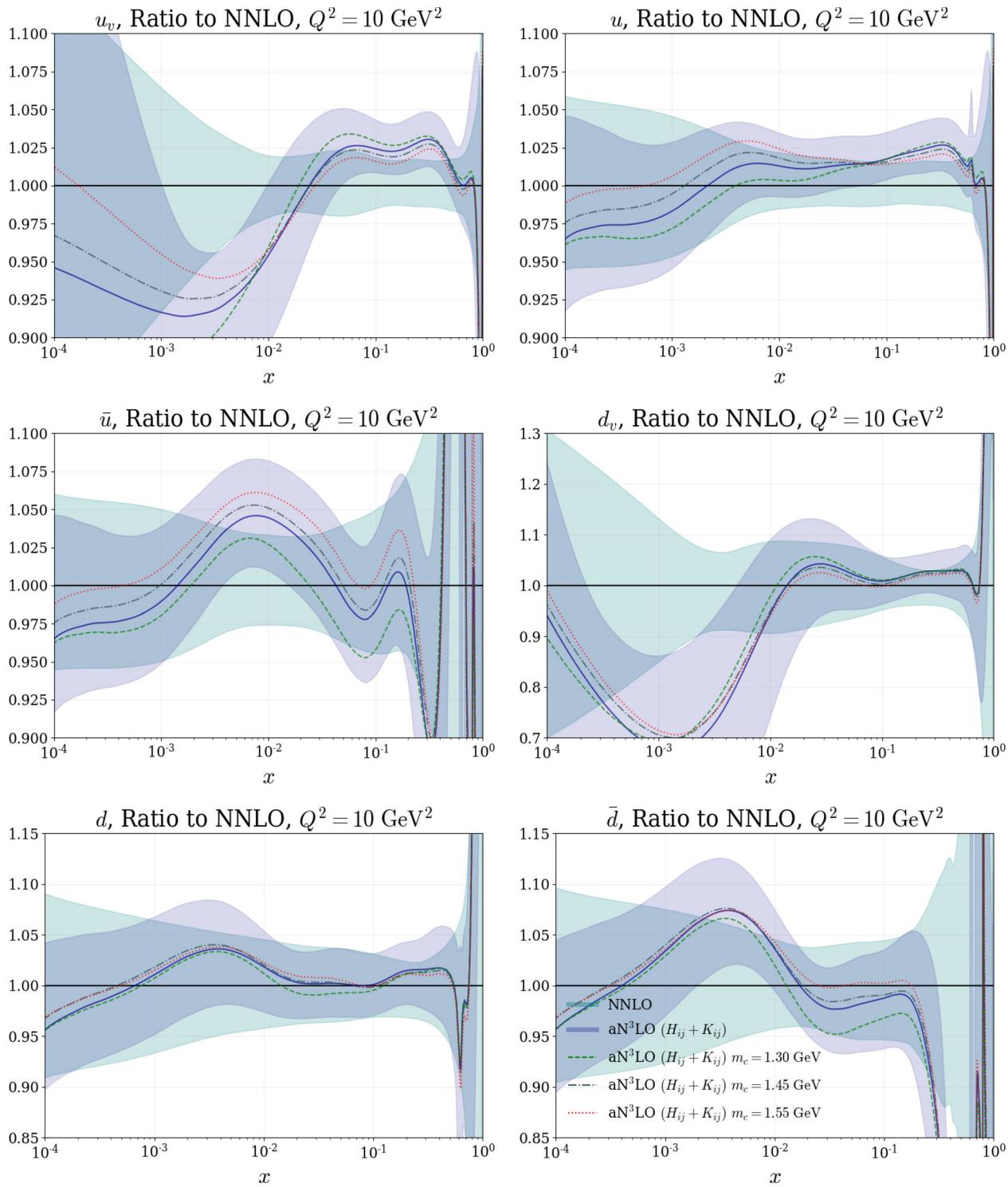
cuts are applied. Figure 48 (right) displays the same analysis for the gluon fusion cross section with  $\mu = \mu_f = \mu_r = m_H$  (numerical results provided in Table 21).

Considering the  $\mu = m_H/2$  and  $\mu = m_H$  central value results displayed in Table 14 and Fig. 48, it can be observed that aN<sup>3</sup>LO PDFs predict a lower central value than NNLO PDFs across all hard cross section orders. One can also notice

an overlap in all cases between predictions from NNLO and aN<sup>3</sup>LO PDFs. However for  $\mu = m_H/2$ , whilst the error bands for predictions with N<sup>3</sup>LO hard cross section and NNLO and N<sup>3</sup>LO PDFs overlap, their central values are outside each other's respective error bands. Since estimating MHOUs via scale variations is a somewhat ambiguous procedure (and is therefore estimated conservatively to reflect this),

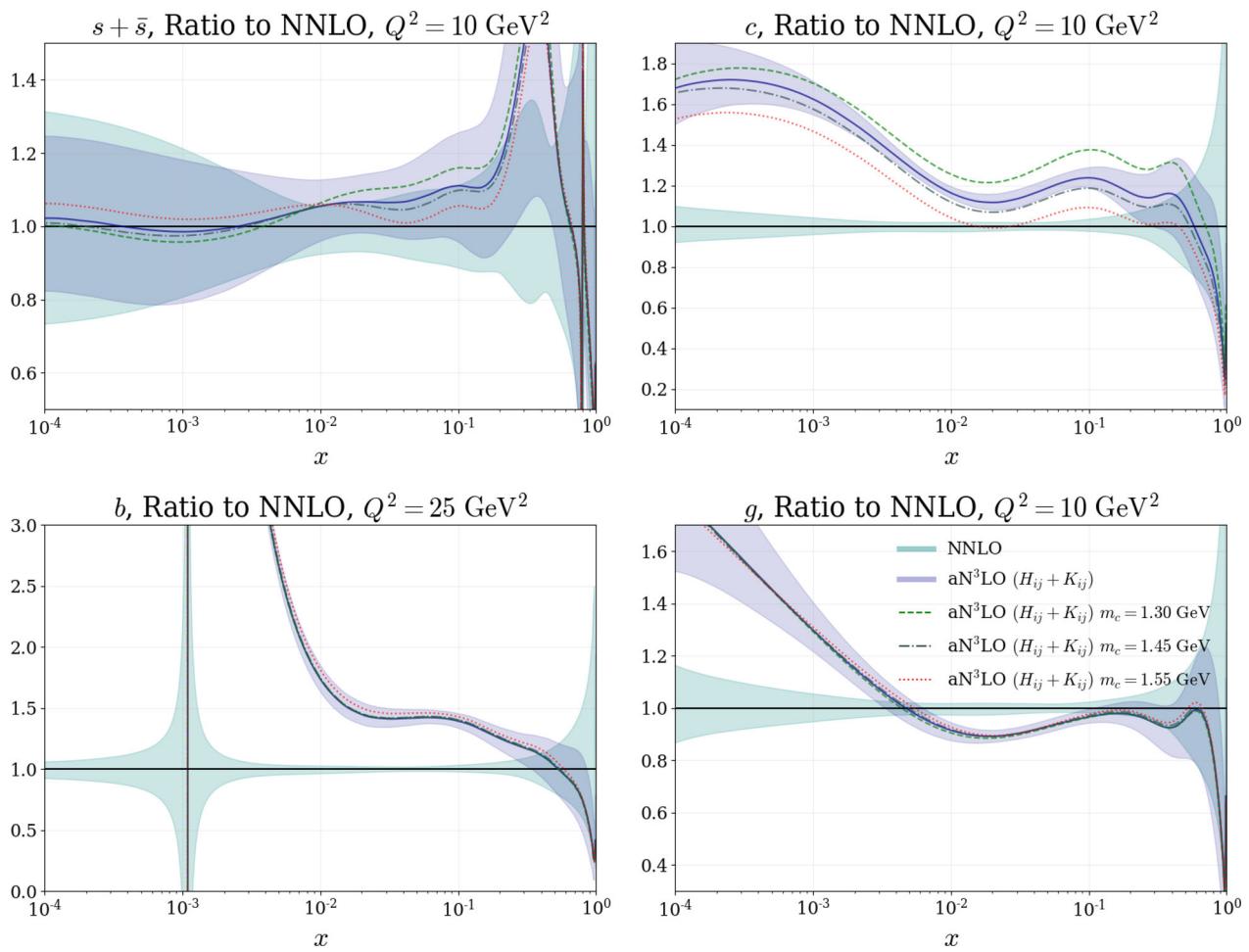
**Fig. 46** Quadratic fit to the total  $\chi^2$  results from various charm masses ( $m_c$ ). The minimum of the quadratic fit provides a rough estimate of  $m_c = 1.45$  GeV at aN<sup>3</sup>LO





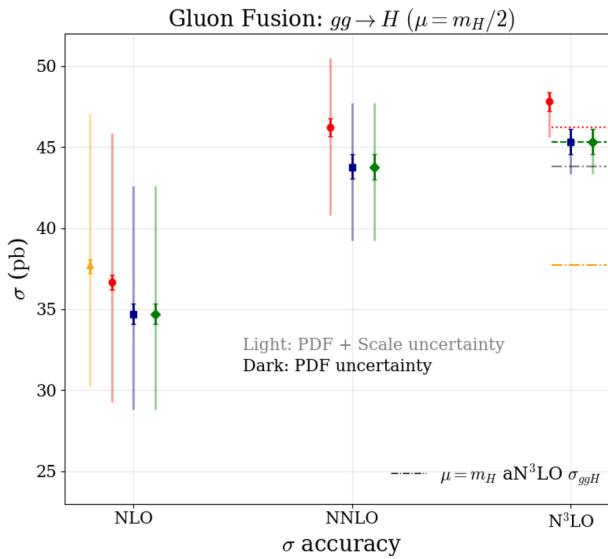
**Fig. 47** Low- $Q^2$  PDF ratios showing aN<sup>3</sup>LO (with decorrelated  $K$ -factors) 68% confidence intervals compared to NNLO 68% confidence intervals with varying fixed values for the charm mass. All plots are shown for  $Q^2 = 10 \text{ GeV}^2$  with the exception of the bottom quark shown

for  $Q^2 = 25 \text{ GeV}^2$ . The PDFs included are:  $m_c = 1.40 \text{ GeV}$  (standard MSHT20 choice) (blue solid),  $m_c = 1.30 \text{ GeV}$  (green dashed),  $m_c = 1.45 \text{ GeV}$  (grey dotted dashed),  $m_c = 1.50 \text{ GeV}$  (red dotted)

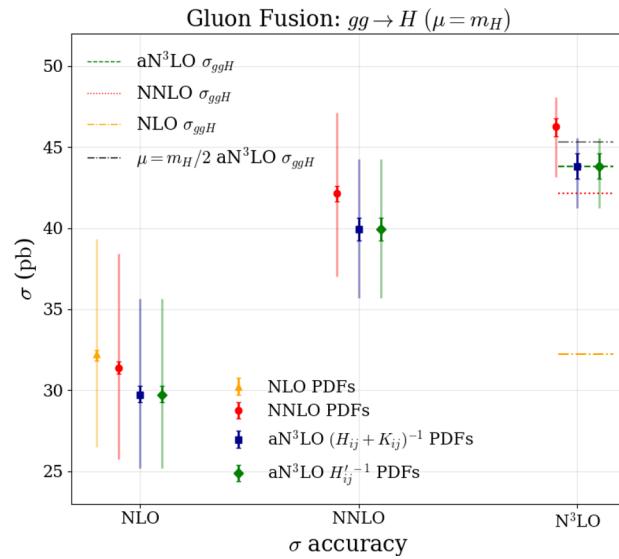
**Fig. 47** continued**Table 14** Higgs production cross section results via gluon fusion (with  $\sqrt{s} = 13$  TeV) using  $N^3LO$  and  $NNLO$  hard cross sections combined with  $NNLO$  and  $aN^3LO$  PDFs. All PDFs are at the standard choice

$\alpha_s(m_Z) = 0.118$ . These results are found with  $\mu = m_H/2$  unless stated otherwise, with the values for  $\mu = m_H$  supplied in Table 21

$\sigma$ order	PDF order	$\sigma + \Delta\sigma_+ - \Delta\sigma_-$ (pb)	$\sigma$ (pb) + $\Delta\sigma_+ - \Delta\sigma_-$ (%)
PDF uncertainties			
$N^3LO$	$aN^3LO$ (no theory unc.)	$45.296 + 0.723 - 0.545$	$45.296 + 1.60\% - 1.22\%$
	$aN^3LO (H_{ij} + K_{ij})$	$45.296 + 0.832 - 0.755$	$45.296 + 1.84\% - 1.67\%$
	$aN^3LO (H'_{ij})$	$45.296 + 0.821 - 0.761$	$45.296 + 1.81\% - 1.68\%$
$NNLO$	$NNLO$	$47.817 + 0.558 - 0.581$	$47.817 + 1.17\% - 1.22\%$
		$46.206 + 0.541 - 0.564$	$46.206 + 1.17\% - 1.22\%$
PDF + Scale uncertainties			
$N^3LO$	$aN^3LO$ (no theory unc.)	$45.296 + 0.723 - 1.851$	$45.296 + 1.60\% - 4.09\%$
	$aN^3LO (H_{ij} + K_{ij})$	$45.296 + 0.832 - 1.923$	$45.296 + 1.84\% - 4.25\%$
	$aN^3LO (H'_{ij})$	$45.296 + 0.821 - 1.926$	$45.296 + 1.81\% - 4.25\%$
$NNLO$	$NNLO$	$47.817 + 0.577 - 2.210$	$47.817 + 1.21\% - 4.62\%$
		$46.206 + 4.284 - 5.414$	$46.206 + 9.27\% - 11.72\%$



**Fig. 48** Higgs production cross section results via gluon fusion (with  $\sqrt{s} = 13$  TeV) at two central scales:  $\mu = m_H/2$  (left) and  $\mu = m_H$  (right). Displayed are the results for  $aN^3LO$  PDFs with decorrelated  $K$ -factors  $((H_{ij} + K_{ij})^{-1})$ , correlated  $K$ -factors  $(H'_{ij}^{-1} = (H_{ij} + K_{ij})^{-1})$



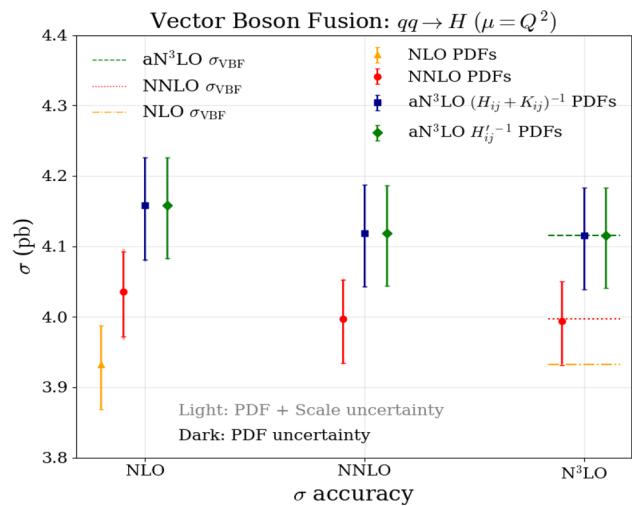
each with a scale variation band from varying  $\mu_r$  by a factor of 2. In the NNLO and NLO PDF cases, both scales  $\mu_f$  and  $\mu_r$  are varied by a factor of 2 following the 9-point convention [11]

these results highlight the benefit of being able to exploit a higher level of control over MHOUs i.e. via nuisance parameters. By predicting a different central value we include a more accurate estimation for higher order predictions which may not be contained within scale variations, especially at unmatched orders in perturbation theory.

Examining the predicted central values further, Fig. 48 suggests that the increase in the cross section theory at  $N^3LO$  is compensated by the PDF theory at  $N^3LO$ , suggesting a cancellation between terms in the PDF and cross section theory at  $N^3LO$ . This point is important to consider when combining unmatched orders in physical calculations, since we must be open to the possibility that unmatched cancellations in physical calculations can lead to inaccurate predictions, as our results suggest here.

Further to this, the change in the gluon PDF is largely driven by the predicted form of  $P_{qg}$  at  $aN^3LO$  and DIS data. Therefore the relevant changes in the gluon at  $aN^3LO$  are most likely due to indirect effects i.e. not directly related to gluon fusion predictions. Due to this, there is no reason to believe that the observed level of convergence should happen at  $aN^3LO$  for both choices of  $\mu$ . However, owing to the inclusion of known information at higher orders, one can be confident that the prediction is more accurate than NNLO, whichever way it moves.

Comparing PDF uncertainty values calculated using NNLO and  $aN^3LO$  PDFs, another prominent feature one can notice in Table 14 is an increase in PDF uncertainties. We find that the PDF uncertainty without  $N^3LO$  theory uncertainties included (i.e. using only the eigenvector description from



**Fig. 49** Higgs production cross section results via vector boson fusion (with  $\sqrt{s} = 13$  TeV) at a central scale set to the vector boson momentum. Displayed are the results for  $aN^3LO$  PDFs with decorrelated  $K$ -factors  $((H_{ij} + K_{ij})^{-1})$ , correlated  $K$ -factors  $(H'_{ij}^{-1} = (H_{ij} + K_{ij})^{-1})$  each with a scale variation band from varying  $\mu_r$  by a factor of 2. In the NNLO and NLO PDF cases, both scales  $\mu_f$  and  $\mu_r$  are varied by a factor of 2 following the 9-point convention [11]

the first 32 eigenvectors and with  $N^3LO$  parameters fixed at the best fit) also includes a marginal increase in the positive direction compared to NNLO. Mathematically, the reason for this comes back to the fact that the best fit is inherently different from the NNLO theory, residing in a completely novel  $\chi^2$  landscape. In turn, this means it is not guaranteed that the PDF uncertainty will remain consistent across the distinct

**Table 15** Higgs production cross section results via the vector boson fusion process (with  $\sqrt{s} = 13$  TeV) using  $N^3LO$  and  $NNLO$  hard cross sections combined with  $NNLO$  and  $aN^3LO$  PDFs. All PDFs are at the

standard choice  $\alpha_s(m_Z) = 0.118$ . These results are found with  $\mu = Q^2$  where  $Q^2$  is the vector boson momentum

$\sigma$ order	PDF order	$\sigma + \Delta\sigma_+ - \Delta\sigma_-$ (pb)	$\sigma$ (pb) + $\Delta\sigma_+ - \Delta\sigma_-$ (%)
PDF uncertainties			
$N^3LO$	$aN^3LO$ (no theory unc.)	$4.1150 + 0.0638 - 0.0724$	$4.1150 + 1.55\% - 1.76\%$
	$aN^3LO (H_{ij} + K_{ij})$	$4.1150 + 0.0682 - 0.0755$	$4.1150 + 1.66\% - 1.83\%$
	$aN^3LO (H'_{ij})$	$4.1150 + 0.0678 - 0.0742$	$4.1150 + 1.65\% - 1.80\%$
$NNLO$	$NNLO$	$3.9941 + 0.0558 - 0.0631$	$3.9941 + 1.40\% - 1.58\%$
	$NNLO$	$3.9974 + 0.0557 - 0.0633$	$3.9974 + 1.39\% - 1.58\%$
PDF + Scale uncertainties			
$N^3LO$	$aN^3LO$ (no theory unc.)	$4.1150 + 0.0638 - 0.0724$	$4.1150 + 1.55\% - 1.76\%$
	$aN^3LO (H_{ij} + K_{ij})$	$4.1150 + 0.0683 - 0.0755$	$4.1150 + 1.66\% - 1.83\%$
	$aN^3LO (H'_{ij})$	$4.1150 + 0.0678 - 0.0742$	$4.1150 + 1.65\% - 1.80\%$
$NNLO$	$NNLO$	$3.9941 + 0.0560 - 0.0631$	$3.9941 + 1.40\% - 1.58\%$
	$NNLO$	$3.9974 + 0.0576 - 0.0642$	$3.9974 + 1.44\% - 1.61\%$

PDF sets.<sup>25</sup> In the case of gluon fusion, the leading contribution to the positive uncertainty direction is an eigenvector primarily dominated by PDF parameters, while in the negative direction a  $N^3LO$  splitting function parameter dominates (eigenvector 9 and 31 in the  $(H_{ij} + K_{ij})^{-1} N^3LO$  case – see Table 11). As discussed in Sect. 8.8, the gluon predominantly follows the splitting function contributions, therefore it is not surprising that this eigenvector is having a noticeable effect. Phenomenologically, the increase in predicted uncertainties from the inclusion of the theoretical uncertainties is a reflection of the estimated PDF MHOUs in this particular cross section, and acts to replace factorisation scale variation. As a consistency check, we find that when performing a 9-point scale variation procedure with  $aN^3LO$  PDFs, the values calculated (for both choices of  $\mu$ ) are within the predicted PDF uncertainties. This is therefore a further verification of our MHOUs and that the  $\mu_f$  variation is intrinsic in the PDF uncertainties.

Finally Fig. 48 also demonstrates the increased stability of predictions when considering the two different central scales  $\mu$  at  $N^3LO$ . As predicted from perturbation theory, the scale dependence is reduced and central values become more in agreement when increasing the order of either the PDFs or hard cross section. Furthermore, the  $aN^3LO$   $\sigma$  central predictions for both choices of  $\mu$  are contained within the uncertainty bands of each other. This is true by definition for the  $NNLO$  PDFs since the factorisation scale  $\mu_f$  variation includes both choices of  $\mu$ , whereas for  $aN^3LO$  PDFs this result is not guaranteed and is therefore intrinsic in the PDF (and renormalisation scale  $\mu_r$  variation) uncertainty.

<sup>25</sup> As we can see from Sect. 8.4, the theory uncertainty is also not guaranteed to add to the total uncertainty (and in fact acts to reduce the uncertainty in some areas of  $(x, Q^2)$ ).

## 9.2 Higgs production – vector boson fusion: $qq \rightarrow H$

Table 15 and Fig. 49 show the predictions at various orders in  $\alpha_s$  for Higgs production cross sections via vector boson fusion<sup>26</sup> at the LHC for  $\sqrt{s} = 13$  TeV up to  $N^3LO$  [153, 154], again no fiducial cuts are applied in this comparison. The predictions shown are calculated with  $\mu_f^2 = \mu_r^2 = Q^2$  as the central scale where  $Q^2$  is the vector boson squared momentum.

For this process one can follow the increase in the cross section as higher order PDFs are used. Contrasting with the case of gluon fusion, Fig. 49 displays little cancellation between the terms added in the  $aN^3LO$  PDF description and the  $N^3LO$  cross section. However, the cross section for VBF produces around a  $\sim 3 - 4\%$  change order by order and is therefore fairly constant. Considering this relatively small difference between orders, this lack of cancellation is not a major concern. Further to this, the vector boson fusion process is much more reliant on the quark sector which, compared to the gluon, is relatively constant order by order (see Sect. 8.4). The reason for this stems from the more direct data constraints on the shape of quark PDFs.

Comparing the  $aN^3LO$  VBF cross section (with MHO theoretical uncertainties) with the  $NNLO$  cross section result (with  $NNLO$  PDFs) including MHOUs via scale variations, we see that the scale variation MHOUs are negligible against the PDF uncertainties at  $aN^3LO$ . This result is in part due to the fact that the scale variation for  $aN^3LO$  is only being included for the renormalisation scale. However at  $NNLO$ , the extra MHOU predicted was still only a small contribu-

<sup>26</sup> Results are obtained with the inclusive part of the code `proVBFH` [153, 154, 156].

**Table 16** Higgs production cross section results via the vector boson fusion process (with  $\sqrt{s} = 13$  TeV) using N<sup>3</sup>LO and NNLO hard cross sections combined with NNLO and decorrelated aN<sup>3</sup>LO PDFs whilst

$\sigma$ order	PDF order	$\sigma + \Delta\sigma_+ - \Delta\sigma_-$ (pb)	$\sigma$ (pb) + $\Delta\sigma_+ - \Delta\sigma_-$ (%)
N <sup>3</sup> LO	aN <sup>3</sup> LO $n_f = 5$	$4.1150 + 0.0683 - 0.0755$	$4.1150 + 1.66\% - 1.83\%$
	aN <sup>3</sup> LO $n_f = 4$	$4.0270 + 0.0685 - 0.0765$	$4.0270 + 1.70\% - 1.90\%$
	aN <sup>3</sup> LO $n_f = 3$	$2.7248 + 0.0653 - 0.0673$	$2.7248 + 2.40\% - 2.47\%$
NNLO	NNLO $n_f = 5$	$3.9974 + 0.0557 - 0.0633$	$3.9974 + 1.39\% - 1.58\%$
	NNLO $n_f = 4$	$3.9118 + 0.0561 - 0.0634$	$3.9118 + 1.44\% - 1.62\%$
	NNLO $n_f = 3$	$2.6845 + 0.0539 - 0.0641$	$2.6845 + 2.01\% - 2.39\%$

tion. Therefore considering these results further, the effects of higher orders in both cases are expected to be small, which provides some agreement with the argument that there is little scope for cancellation between orders for VBF. As for the gluon fusion prediction in Sect. 9.1, we confirm that any further factorisation scale variation (i.e. using the 9-point prescription) is contained within the predicted PDF uncertainties; hence further motivating our previous argument that factorisation scale variation is not necessary with aN<sup>3</sup>LO PDFs.

Another feature of the VBF results is that the level of uncertainty at full aN<sup>3</sup>LO is only increased slightly from the calculation involving NNLO PDFs. Comparing this to the gluon fusion results, where the uncertainty was more noticeably increased in both directions, it is evident that these approximate N<sup>3</sup>LO additions are having a smaller effect on the VBF calculation. Once again, the origin of this is due to the nature of the process. VBF involves mostly the quark sector and is therefore much less affected by the extra N<sup>3</sup>LO theory we have introduced (due to direct constraints from data). As we have presented in previous sections, most of the uncertainty in the N<sup>3</sup>LO theory resides in the small- $x$  regime which is more directly probed by the gluon sector than in the quark sector.

Lastly we briefly discuss the  $n_f$  dependence of the VBF cross section. In VBF the scaling of contributions follows as  $n_f^2$  due to the presence of two input quark flavours in the process. In Table 16 we observe that the VBF cross section receives a large contribution when including the charm quark ( $n_f = 3 \rightarrow 4$ ) due to this scaling. We also show that at aN<sup>3</sup>LO, this is where most of the difference in the central value and uncertainty from NNLO is accounted for. This is a consequence of the predicted enhancement of the charm PDF at aN<sup>3</sup>LO, discussed in Sect. 8.4. Beyond  $n_f = 4$  the bottom contribution to VBF in the  $W^\pm$  channel (the dominant channel) is heavily suppressed, since due to the CKM elements  $b$  must transition to  $t$  most of the time. Therefore the VBF cross section only receives a small contribution moving from  $n_f = 4$  to  $n_f = 5$ .

varying the number of active flavours  $n_f$ . All PDFs are at the standard choice  $\alpha_s(m_Z) = 0.118$ . These results are found with  $\mu = Q^2$  where  $Q^2$  is the vector boson momentum

## 10 Availability and recommended usage of MSHT20 aN<sup>3</sup>LO PDFs

We provide the MSHT20 aN<sup>3</sup>LO PDFs in LHAPDF format [157]:

<http://lhapdf.hepforge.org/>

as well as on the repository:

<http://www.hep.ucl.ac.uk/msht/>

The approximate N<sup>3</sup>LO functions (for  $P_{ij}(x)$  and  $A_{ij}(x)$ ) are provided as lightweight FORTRAN functions or as part of a Python framework in the repository:

[https://github.com/MSHTPDF/N3LO\\_additions](https://github.com/MSHTPDF/N3LO_additions)

We present the aN<sup>3</sup>LO eigenvector sets with and without correlated  $K$ -factors as discussed in Sect. 8, with the default set being provided with decorrelated  $K$ -factors.<sup>27</sup>

[MSHT20an3lo\\_as118](#)

[MSHT20an3lo\\_as118\\_Kcorr](#)

Both these PDF sets contain a central PDF accompanied by 104 eigenvector directions (describing 52 eigenvectors) and can be used in exactly the same way as previous MSHT PDF sets i.e. the MSHT20 NNLO PDFs with 64 eigenvector directions.

As presented in this work, the aN<sup>3</sup>LO PDFs include an estimation for missing N<sup>3</sup>LO contributions (the leading theoretical uncertainty) and implicitly some MHOU beyond this within their PDF uncertainties. Due to this, we argue and motivate in Sect. 9 that factorisation scale variations are no longer necessary in calculations involving aN<sup>3</sup>LO PDFs. However the renormalisation scale should continue to be varied to provide estimates of MHOU in the hard cross section piece of physical calculations.

In the case that the hard cross section for a process is available up to N<sup>3</sup>LO the recommendation is to use the aN<sup>3</sup>LO PDFs, since unmatched ingredients in cross section calculations can ignore important cancellations (between the PDFs and hard cross section).

<sup>27</sup> These grids are updated from a previous version of the MSHT20 aN<sup>3</sup>LO PDFs and should be used in favour of any sets downloaded before the latest upload date of this submission.

If a process is included within the global fit and the hard cross section is known only up to NNLO (i.e. those discussed in Sect. 7), we recommend the use of the decorrelated version of the aN<sup>3</sup>LO PDF set. Using these PDFs and the details provided in Table 10, the hard cross section can be transformed from NNLO to approximate N<sup>3</sup>LO. From here the two approximate N<sup>3</sup>LO ingredients can be used together to give a full approximate N<sup>3</sup>LO result.

If a process is not included in the global PDF fit and the hard cross section is known only up to NNLO, the standard NNLO PDF set remains the default choice. However, we recommend the use of these aN<sup>3</sup>LO PDFs as an estimate of potential MHOUs. In this case the aN<sup>3</sup>LO PDF set + NNLO hard cross section prediction should be reflected in any MHOU estimates for the full NNLO prediction. For example, when the hard cross section is known only up to NNLO Equation (3.13) from [75] can be adapted to be,

$$\delta(\text{PDF} - \text{TH}) = \frac{1}{2} \left| \frac{\sigma_{\text{aN}^3\text{LO}}^{(2)} - \sigma_{\text{NNLO}}^{(2)}}{\sigma_{\text{aN}^3\text{LO}}^{(2)}} \right| \quad (10.1)$$

where  $\delta(\text{PDF} - \text{TH})$  is the predicted PDF theory uncertainty on the  $\sigma$  prediction,  $\sigma_{\text{aN}^3\text{LO}}^{(2)}$  is the NNLO hard cross section with aN<sup>3</sup>LO PDFs and  $\sigma_{\text{NNLO}}^{(2)}$  is the full NNLO result. A caveat to this treatment is that the theory uncertainty is sensitive to unmatched cancellations and should therefore be used with care (and caution), therefore the NNLO set remains the default in evaluating PDF uncertainties.

## 11 Conclusions

In this paper we have presented the first approximate N<sup>3</sup>LO global PDF fit. This follows the MSHT20 framework [3], where the aN<sup>3</sup>LO PDF set also incorporates estimates for theoretical uncertainties from missing N<sup>3</sup>LO contributions and implicitly some MHOUs beyond this. In addition, the framework presented for obtaining these PDFs provides a means of utilising higher order information as and when it is available. In contrast, previously, complete information of the next order was required for theoretical calculations in PDF fits. This provides a significant advantage moving forward in precision phenomenology, since as we move to higher orders, this information takes increasingly longer to calculate. We have analysed the resulting set of PDFs, denoted MSHT20aN<sup>3</sup>LO, and made two sets available as described in Sect. 10. The aN<sup>3</sup>LO PDF fits have been performed to the same set of global hard scattering data and PDF parameterisations included for the MSHT20 NNLO PDF fits.

The NNLO theoretical framework for MSHT20 PDFs has been extended in Sect. 2 to include the addition of general N<sup>3</sup>LO theory parameters into the fit. Subsequently, we have outlined how these N<sup>3</sup>LO theory parameters can be included

into the Hessian procedure as controllable nuisance parameters where they are not yet known. Two methods of handling subsets of the N<sup>3</sup>LO theory parameters in the Hessian matrix have then been discussed; i.e. including or ignoring correlations with aN<sup>3</sup>LO  $K$ -factors across distinct processes.

In Sects. 4–7 we have presented the N<sup>3</sup>LO additions to the relevant splitting functions, transition matrix elements, heavy coefficient functions and  $K$ -factors. We present usable and computationally efficient approximations to N<sup>3</sup>LO based on known information in the small and large- $x$  regimes and the available Mellin moments (and make these available as described in Sect. 10). In all cases the best fit prediction for each N<sup>3</sup>LO function is in good agreement with the prior expected behaviour. Also in Sect. 7, we find good agreement with recent progress towards N<sup>3</sup>LO DY and top production  $K$ -factors [64, 66]. As more information becomes available surrounding each of these functions, the framework we present here can be easily adapted, aiding in the reduction in sources of MHOUs from N<sup>3</sup>LO. As we have stressed, we interpret our theoretical uncertainty as being mainly due to the remaining uncertainty at N<sup>3</sup>LO, but with some small, but significant contribution from even higher orders, particularly at small- $x$ . Our results seem consistent with this interpretation. However, in the future we expect the N<sup>3</sup>LO description to become more exact. Hence, at some point the remaining N<sup>3</sup>LO uncertainty will become comparable to, or smaller than effects beyond N<sup>3</sup>LO. We would then have to modify our procedure. However, we expect that once the N<sup>3</sup>LO theory becomes very largely known, there will at this point also be more information known about even higher orders (i.e. N<sup>4</sup>LO), which could then be incorporated in a similar manner to maintain an estimate for MHOUs. Alternatively, in the event that the available information is not suitable to provide approximations (or indeed to complement these approximations), a treatment similar in principle, but more sophisticated in practice, to that of the  $K$ -factors may be adopted for DIS quantities. On this note, we acknowledge that the method of constructing aN<sup>3</sup>LO  $K$ -factors for non-(inclusive) DIS processes presented here is a first step towards a more robust and flexible procedure, which is left for future work.

Combining together all N<sup>3</sup>LO information, in Sect. 8 the results of an approximate N<sup>3</sup>LO global PDF fit are presented. The new MSHT20 approximate N<sup>3</sup>LO PDFs show a significant reduction in  $\chi^2$  from the MSHT20 NNLO PDF set, with the leading NNLO tensions between HERA and non-HERA datasets heavily reduced at aN<sup>3</sup>LO (most notably with the ATLAS 8 TeV  $Z$   $p_T$  dataset [110]). With this being said, the aN<sup>3</sup>LO set does fit selected Jets datasets worse in an aN<sup>3</sup>LO global fit than at NNLO, although these are an exception to the behaviour seen for the other datasets. In performing a fit not including ATLAS 8 TeV  $Z$   $p_T$  data we provide evidence that similar tensions seen at NNLO (see [3]) remain between this dataset and jet production data at aN<sup>3</sup>LO. Fur-

ther to this, we show that since HERA and ATLAS 8 TeV  $Z$   $p_T$  data are more in agreement in the form of the high- $x$  gluon at  $aN^3LO$ , one can observe that the tension with the jet production data is shared between HERA and ATLAS 8 TeV  $Z$   $p_T$  data. Finally, as discussed, we highlight that in future work it will be interesting to observe if this increased tension may be alleviated when considering these jet datasets instead as dijet cross sections.

Investigating the correlations present within an  $aN^3LO$  PDF fit, a natural separation between process independent and process dependent parameters can be observed. With this motivation, a PDF set with decorrelated  $aN^3LO$   $K$ -factor eigenvectors is constructed. The validity of this is then also verified by comparison with a second PDF set which includes correlations between all parameters. Each of these sets exhibits similarly well behaved eigenvectors and levels of dynamical tolerance.

Considering the form of the individual PDFs, the  $aN^3LO$  PDFs include a much harder gluon at small- $x$  due to contributions from the splitting functions as discussed in Sect. 8.8. This enhancement then translates into an increase in the charm and bottom PDFs due to the gluon input into the heavy flavour sector via the transition matrix elements. At very low- $Q^2$  the result of the  $N^3LO$  additions is a non-negative charm and gluon PDF at small- $x$ . As a consistency check, the fit dependence on  $\alpha_s$  and  $m_c$  has been investigated. In both of these cases we show a preference for values which suppress the heavy flavour contributions (slightly lower  $\alpha_s$  and slightly higher  $m_c$  than NNLO). Considering the predicted  $aN^3LO$   $\alpha_s$ , we observe a slightly lower than  $1\sigma$  effect when comparing with the NNLO world average. While an extensive analysis of the  $aN^3LO$   $\alpha_s$  value is left for further study, since the world average is determined by NNLO results, one could expect a small systematic effect from moving to  $N^3LO$ .

Taking this analysis further and using the approximate  $N^3LO$  PDFs as input to  $N^3LO$  cross section calculations, we consider the cases of gluon and vector boson fusion in Higgs production. We present the first  $aN^3LO$  calculation for these cross sections and show how the  $aN^3LO$  prediction differs from the case with NNLO PDFs including scale variations, highlighting the importance of matching orders in calculations. In VBF we provide an example where cancellation is not realised between orders. However in this case the quark sector is much more constrained and due to the smaller variation between orders, there is naturally less scope for cancellation.

In summary, we have presented a set of approximate  $N^3LO$  PDFs that are able to more accurately predict physical quantities involving PDFs (given that all ingredients in these calculations are included at  $N^3LO$  or  $aN^3LO$ ). In producing these PDFs, we have provided a more controllable method for estimating theoretical uncertainties from MHOs in a PDF fit than scale variations. While some ambiguity remains in this

method in how the prior variations are chosen, we argue that the current knowledge and intuition surrounding each source of uncertainty can be utilised as and when available. This is therefore much more in line with what one can expect a theoretical uncertainty to encompass. Another potential shortcoming is the possibility of fitting to sources of uncertainty other than higher orders (or higher order corrections elsewhere in theory calculations included in a PDF fit). Although this is a possibility, the position of the considered sources of uncertainty in the underlying theory combined with the prior variations and penalties should act to minimise this effect. In any case, if a separate source of uncertainty is significantly affecting the fit, this will present itself as a source of tension with the  $N^3LO$  penalties and the  $\chi^2$  (and PDF uncertainty) will be adapted accordingly.

In future work it will be interesting to investigate the effects in the high- $x$  gluon, which is a region of phenomenological importance and where the interpretation of LHC constraints is not always straightforward. We also note that there are  $N^3LO$  results available from di-lepton rapidity in DY processes [64]. Considering the results in Sect. 7 which display an agreement with these recent results, we hope that these approximate  $N^3LO$  PDFs may be of interest in this analysis. Similarly for recent results considering top production [66]. Furthermore, any approximate information from these results could be included in the  $N^3LO$   $K$ -factor priors, which was not done for this iteration of the  $aN^3LO$  PDFs. Finally, in order to continually improve the description of  $aN^3LO$  PDFs, the inclusion of more sub-leading sources of MHOUs could be addressed. With the upcoming wealth of experimental data from future colliders such as the HL-LHC and the EIC, it will be of interest to gain a better understanding of the transition matrix elements and also describe better the charged current and longitudinal structure functions, where currently theoretical uncertainties are much smaller than the experimental uncertainties.

**Acknowledgements** J.M. thanks the Science and Technology Facilities Council (STFC) part of U.K. Research and Innovation for support via Ph.D. funding. T.C. and R.S.T. thank STFC for support via grant awards ST/P000274/1 and ST/T000856/1. L.H.L. thanks STFC for support via grant awards ST/L000377/1 and ST/T000864/1. We would like to thank Xuan Chen, Thomas Gehrmann, Nigel Glover and Alex Huss for providing details of  $N^3LO$  calculations. We would like to thank members of the PDF4LHC working group for numerous discussions on PDFs and theoretical uncertainties. We would also like to thank Alan Martin for long collaboration on the MSHT series of PDFs.

**Data Availability Statement** This manuscript has no associated data or the data will not be deposited. [Authors' comment: This manuscript has associated data in the form of grids for parton distributions functions which can be found at <https://www.hep.ucl.ac.uk/msht/index.shtml> and <https://lhapdf.hepforge.org/pdfsets>].

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you

give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Funded by SCOAP<sup>3</sup>. SCOAP<sup>3</sup> supports the goals of the International Year of Basic Sciences for Sustainable Development.

## Appendix A: List of N<sup>3</sup>LO ingredients

**Table 17** List of all the N<sup>3</sup>LO ingredients used to construct the approximate N<sup>3</sup>LO splitting functions and transition matrix elements. Where only a citation is provided, extensive knowledge i.e. beyond NLL is used. This table is a non-exhaustive list of the current knowledge about these functions, however information beyond that which is provided here is not currently in a usable format for phenomenological studies

N <sup>3</sup> LO function	No. of moments	Moments (even only)	Small- $x$	Large- $x$
$P_{qq}^{\text{NS}}$	8	$N = 2-16$ [21]	[21]	[21]
$P_{qq}^{\text{PS}}$	4	$N = 2-8$ [35, 36]	LL [28]	N/A
$P_{qg}$	4	$N = 2-8$ [35, 36]	LL [28]	N/A
$P_{gq}$	4	$N = 2-8$ [35, 36]	LL [29-31]	N/A
$P_{gg}$	4	$N = 2-8$ [35, 36]	LL and NLL [29-33]	N/A
$A_{qq,H}^{\text{NS}}$	7	$N = 2-14$ [50]	N/A	N/A
$A_{Hq}^{\text{PS}}$	6	$N = 2-12$ [50]	[53]	[53]
$A_{Hg}$	5	$N = 2-10$ [50]	LL [49]	N/A
$A_{gq,H}$	7	$N = 2-14$ [50]	[54]	[54]
$A_{gg,H}$	5	$N = 2-10$ [50]	N/A	N/A

**Table 18** List of all N<sup>3</sup>LO ingredients used to construct the approximate N<sup>3</sup>LO GM-VFNS coefficient functions. Note that lower order components that contribute to these functions are also known and are cited in the text. This table only considers contributing 3-loop functions

GM-VFNS N <sup>3</sup> LO function	Known N <sup>3</sup> LO components
$C_{H,q}$	$C_{H,q}^{(3), \text{FF}} (Q^2 \leq m_h^2)$ LL [47-49], $C_{H,q}^{\text{VF}, (3)}$ [57]
$C_{H,g}$	$C_{H,g}^{(3), \text{FF}} (Q^2 \leq m_h^2)$ LL [47-49], $C_{H,g}^{\text{ZM}, (3)}$ [57]
$C_{q,q}^{\text{NS}}$	$C_{q,q, \text{NS}}^{\text{ZM}, (3)}$ [57]
$C_{q,q}^{\text{PS}}$	$C_{q,q, \text{PS}}^{\text{ZM}, (3)}$ [57]
$C_{q,g}$	$C_{q,g}^{\text{ZM}, (3)}$ [57]

Tables 17 and 18 summarise the available (at the time of writing) and used information regarding the N<sup>3</sup>LO splitting functions and coefficient functions respectively. The formalism presented in Sect. 2 currently makes use of all this information and is able to be adapted as and when more information becomes available.

## Appendix B: $\chi^2$ results without HERA

### B.1 NNLO

Table 19 shows the differences in  $\chi^2$  found when omitting HERA data from a PDF fit using the MSHT NNLO PDFs. This table is copied here from [3] for the ease of the reader. We see similarities between these results and the  $\Delta\chi^2$ 's seen in the case of N<sup>3</sup>LO PDFs. Specifically the ATLAS 8 TeV  $Z$   $p_T$  displaying a substantial reduction from the global NNLO fit. This therefore provides evidence that the inclusion of the N<sup>3</sup>LO contributions is aiding in reducing tensions between the HERA and non-HERA datasets.

**Table 19** The change in  $\chi^2$  for a NNLO fit (with negative indicating an improvement in the fit quality) when the combined HERA data sets including  $F_L$  and heavy flavour data are removed, illustrating the tensions of these data sets with several of the other data sets in the global fit.  $\Delta\chi^2$  represents the change from a full global fit at the same order in  $\alpha_s$

Dataset	$N_{\text{pts}}$	$\chi^2$	$\Delta\chi^2$
BCDMS $\mu p$ $F_2$ [135]	163	174.7	-5.5
BCDMS $\mu d$ $F_2$ [135]	151	143.9	-2.1
NMC $\mu p$ $F_2$ [136]	123	119.6	-4.5
NMC $\mu d$ $F_2$ [136]	123	96.6	-16.1
SLAC $ep$ $F_2$ [137, 138]	37	33.0	+0.9
SLAC $ed$ $F_2$ [137, 138]	38	24.1	+1.1
E665 $\mu d$ $F_2$ [139]	53	63.5	+3.9
E665 $\mu p$ $F_2$ [139]	53	68.9	+4.3
NuTeV $\nu N$ $F_2$ [140]	53	38.0	-0.3
NuTeV $\nu N$ $x F_3$ [140]	42	27.5	-3.2
NMC $\mu n/\mu p$ [141]	148	132.7	+1.9
E866/NuSea $pp$ DY [80]	184	228.0	+2.9
E866/NuSea $pd/pp$ DY [81]	15	9.1	-1.3
CCFR $\nu N \rightarrow \mu\mu X$ [133]	86	66.2	-1.5
NuTeV $\nu N \rightarrow \mu\mu X$ [133]	84	49.0	-9.5
CHORUS $\nu N$ $F_2$ [146]	42	29.6	-0.6
CHORUS $\nu N$ $x F_3$ [146]	28	18.2	-0.3
CDF II $p\bar{p}$ incl. jets [102]	76	60.9	+0.5

**Table 19** continued

Dataset	$N_{\text{pts}}$	$\chi^2$	$\Delta\chi^2$
DØ II $Z$ rap. [82]	28	16.6	+0.3
CDF II $Z$ rap. [83]	28	38.7	+1.5
DØ II $W \rightarrow \nu\mu$ asym. [84]	10	17.4	+0.1
CDF II $W$ asym. [85]	13	19.0	+0.0
DØ II $W \rightarrow \nu e$ asym. [86]	12	30.0	-3.9
DØ II $p\bar{p}$ incl. jets [103]	110	119.3	-0.9
ATLAS $W^+, W^-, Z$ [87]	30	29.5	-0.4
CMS $W$ asym. $p_T > 35$ GeV [88]	11	6.6	-1.2
CMS $W$ asym. $p_T > 25, 30$ GeV [89]	24	7.5	+0.1
LHCb $Z \rightarrow e^+e^-$ [90]	9	24.2	+1.5
LHCb $W$ asym. $p_T > 20$ GeV [91]	10	12.1	-0.3
CMS $Z \rightarrow e^+e^-$ [92]	35	17.3	-0.6
ATLAS High-mass Drell–Yan [93]	13	16.9	-2.0
Tevatron, ATLAS, CMS $\sigma_{t\bar{t}}$ [117–129]	17	14.2	-0.4
CMS double diff. Drell–Yan [94]	132	134.2	-10.3
LHCb 2015 $W, Z$ [77,78]	67	97.4	-1.9
LHCb 8 TeV $Z \rightarrow ee$ [95]	17	24.4	-1.8
CMS 8 TeV $W$ [96]	22	13.7	+0.9
ATLAS 7 TeV jets [104]	140	228.0	+6.5
CMS 7 TeV $W + c$ [62]	10	9.2	+0.6
ATLAS 7 TeV high prec. $W, Z$ [79]	61	116.8	+0.2
CMS 7 TeV jets [101]	158	179.5	+3.8
DØ $W$ asym. [97]	14	11.3	-0.8
ATLAS 8 TeV $Z$ $p_T$ [110]	104	149.3	-39.2
CMS 8 TeV jets [105]	174	259.5	-1.8
ATLAS 8 TeV sing. diff. $t\bar{t}$ [130]	25	24.5	-1.1
ATLAS 8 TeV sing. diff. $t\bar{t}$ dilep. [131]	5	2.3	-1.1
ATLAS 8 TeV High-mass DY [98]	48	60.9	+3.7
ATLAS 8 TeV $W +$ jets [111]	30	16.4	-1.7
CMS 8 TeV double diff. $t\bar{t}$ [132]	15	23.3	+0.8
ATLAS 8 TeV $W$ [99]	22	54.4	-3.0
CMS 2.76 TeV jet [106]	81	102.9	+0.0
CMS 8 TeV sing. diff. $t\bar{t}$ [112]	9	10.6	-2.6
ATLAS 8 TeV double diff. $Z$ [100]	59	108.3	+22.7
Total	3042	3379.6	-61.6

## B.2 aN<sup>3</sup>LO

Table 20 shows the differences in  $\chi^2$  found when omitting HERA data from a PDF fit using the MSHT aN<sup>3</sup>LO PDFs. These results show that at aN<sup>3</sup>LO the fit no longer experiences large tensions between HERA and ATLAS 8 TeV  $Z$   $p_T$  [110] datasets. The main tensions at N<sup>3</sup>LO are now concerning the Jets data with HERA (and most likely some non-HERA datasets). This result is not unexpected due to the known issues surrounding jets especially as we move to higher precision [109].

**Table 20** The change in  $\chi^2$  for an N<sup>3</sup>LO fit (with negative indicating an improvement in the fit quality) when the combined HERA data sets including  $F_L$  and heavy flavour data are removed, illustrating the tensions of these data sets with several of the other data sets in the global fit.  $\Delta\chi^2$  represents the change from a full global fit at the same order in  $\alpha_s$ 

Dataset	$N_{\text{pts}}$	$\chi^2$	$\Delta\chi^2$
BCDMS $\mu p$ $F_2$ [135]	163	175.8	+1.4
BCDMS $\mu d$ $F_2$ [135]	151	144.2	-0.0
NMC $\mu p$ $F_2$ [136]	123	113.6	-7.8
NMC $\mu d$ $F_2$ [136]	123	87.6	-16.6
SLAC $e p$ $F_2$ [137,138]	37	30.7	-0.9
SLAC $e d$ $F_2$ [137,138]	38	23.2	+0.4
E665 $\mu d$ $F_2$ [139]	53	65.2	+1.3
E665 $\mu p$ $F_2$ [139]	53	69.0	+1.5
NuTeV $\nu N$ $F_2$ [140]	53	35.4	-0.4
NuTeV $\nu N$ $x F_3$ [140]	42	29.2	-5.6
NMC $\mu n/\mu p$ [141]	148	131.1	-0.5
E866/NuSea $pp$ DY [80]	184	217.6	+2.3
E866/NuSea $pd/pp$ DY [81]	15	8.2	-0.2
CCFR $\nu N \rightarrow \mu\mu X$ [133]	86	67.0	-1.3
NuTeV $\nu N \rightarrow \mu\mu X$ [133]	84	47.6	-9.1
CHORUS $\nu N$ $F_2$ [146]	42	29.0	-0.2
CHORUS $\nu N$ $x F_3$ [146]	28	18.5	+0.4
CDF II $p\bar{p}$ incl. jets [102]	76	65.9	-0.6
DØ II $Z$ rap. [82]	28	17.7	+0.3
CDF II $Z$ rap. [83]	28	42.1	+1.5
DØ II $W \rightarrow \nu\mu$ asym. [84]	10	18.9	+2.4
CDF II $W$ asym. [85]	13	19.2	+0.9
DØ II $W \rightarrow \nu e$ asym. [86]	12	31.0	+0.2
DØ II $p\bar{p}$ incl. jets [103]	110	114.2	+0.9
ATLAS $W^+, W^-, Z$ [87]	30	29.5	-0.5
CMS $W$ asym. $p_T > 35$ GeV [88]	11	7.0	+0.2
CMS $W$ asym. $p_T > 25, 30$ GeV [89]	24	7.8	-0.2
LHCb $Z \rightarrow e^+e^-$ [90]	9	22.5	-0.7
LHCb $W$ asym. $p_T > 20$ GeV [91]	10	12.6	+0.0
CMS $Z \rightarrow e^+e^-$ [92]	35	17.1	-0.5
ATLAS High-mass Drell–Yan [93]	13	17.5	-0.9
Tevatron, ATLAS, CMS $\sigma_{t\bar{t}}$ [117–129]	17	14.4	+0.1
CMS double diff. Drell–Yan [94]	132	129.5	-3.7

**Table 20** continued

Dataset	$N_{\text{pts}}$	$\chi^2$	$\Delta\chi^2$
LHCb 2015 $W, Z$ [77, 78]	67	96.7	-6.5
LHCb 8 TeV $Z \rightarrow ee$ [95]	17	27.9	-2.4
CMS 8 TeV $W$ [96]	22	11.6	+0.1
ATLAS 7 TeV jets [104]	140	217.7	+1.8
CMS 7 TeV $W + c$ [62]	10	10.8	+0.0
ATLAS 7 TeV high prec. $W, Z$ [79]	61	118.0	-1.3
CMS 7 TeV jets [101]	158	187.8	+1.0
DØ $W$ asym. [97]	14	10.1	-2.0
ATLAS 8 TeV $Z$ $p_T$ [110]	104	121.2	+12.8
CMS 8 TeV jets [105]	174	259.8	-11.5
ATLAS 8 TeV sing. diff. $t\bar{t}$ [130]	25	24.1	-0.2
ATLAS 8 TeV sing. diff. $t\bar{t}$ dilep. [131]	5	3.0	+0.3
ATLAS 8 TeV High-mass DY [98]	48	65.2	+2.4
ATLAS 8 TeV $W +$ jets [111]	30	18.0	-0.8
CMS 8 TeV double diff. $t\bar{t}$ [132]	15	22.8	-0.8
ATLAS 8 TeV $W$ [99]	22	48.0	-5.0
CMS 2.76 TeV jet [106]	81	103.0	-6.8
CMS 8 TeV sing. diff. $t\bar{t}$ [112]	9	12.3	+2.0
ATLAS 8 TeV double diff. $Z$ [100]	59	86.1	+5.7
Low- $Q^2$ coefficient			
$c_q^{\text{NLL}} = -3.844$	0.006	$c_g^{\text{NLL}} = -3.875$	0.004
Transition matrix elements			
$a_{Hg} = 17788.000$	5.607	$a_{qg,H}^{\text{NS}} = -63.950$	0.000
$a_{gg,H} = -1334.500$	0.001		
Splitting functions			
$\rho_{qg}^{\text{NS}} = 0.007$	0.000	$\rho_{gg} = -1.647$	0.001
$\rho_{qg}^{\text{PS}} = -0.579$	0.429	$\rho_{gg} = 9.237$	0.023
$\rho_{qg} = -1.343$	0.131		
K-factors			
$\text{DY}_{\text{NLO}} = -0.291$	0.085	$\text{DY}_{\text{NNLO}} = -0.228$	0.052
$\text{Top}_{\text{NLO}} = -0.204$	0.042	$\text{Top}_{\text{NNLO}} = 0.412$	0.170
$\text{Jet}_{\text{NLO}} = -0.254$	0.065	$\text{Jet}_{\text{NNLO}} = -0.861$	0.741
$p_T \text{Jets}_{\text{NLO}} = 0.461$	0.213	$p_T \text{Jets}_{\text{NNLO}} = 0.016$	0.000
$\text{Dimuon}_{\text{NLO}} = -0.329$	0.109	$\text{Dimuon}_{\text{NNLO}} = 0.587$	0.345
	Total	3311.8/3042	
	$\Delta\chi^2$ from $\text{N}^3\text{LO}$	-48.0	

**Appendix C: Dynamic tolerances**

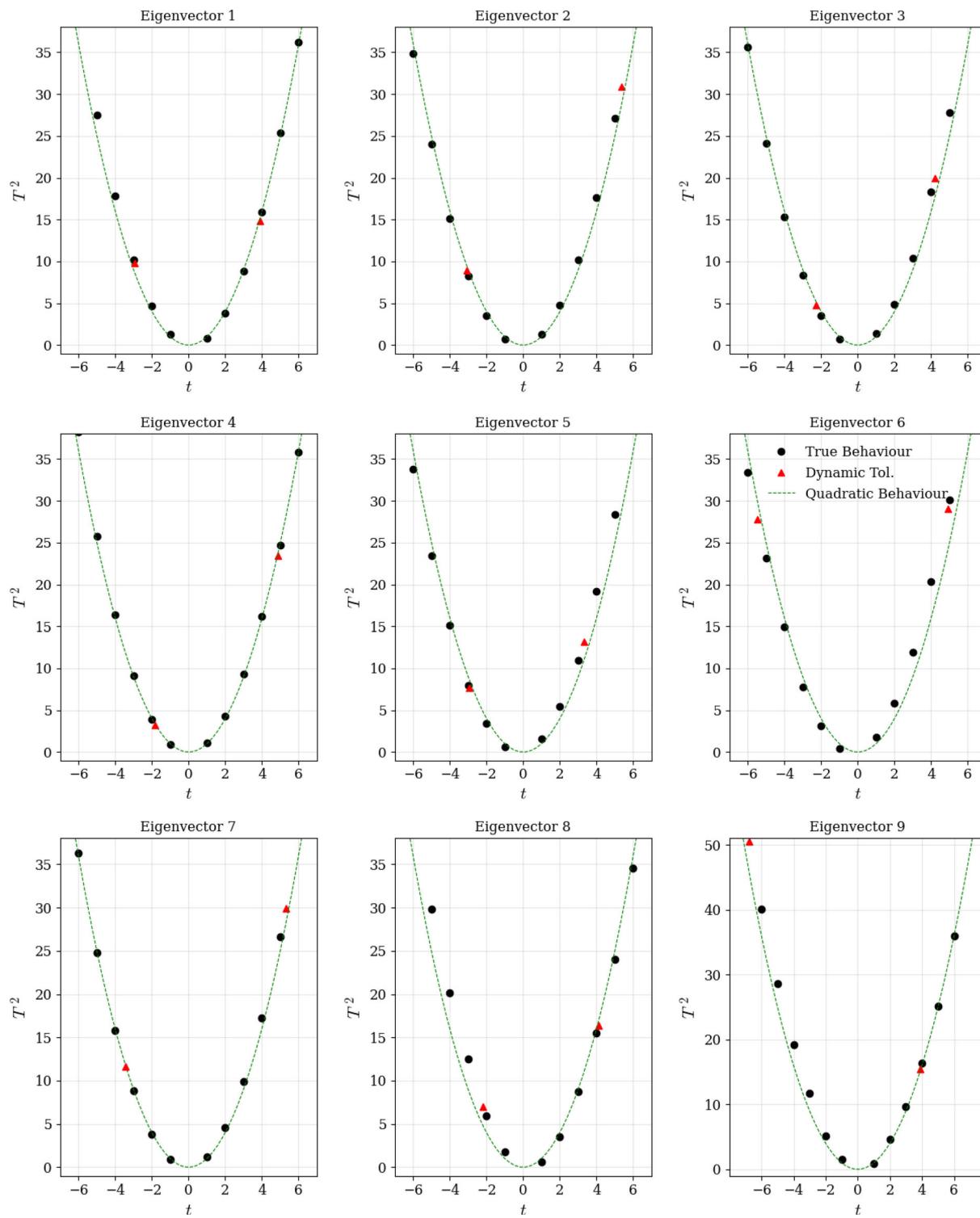
In this section we provide an exhaustive breakdown of the  $\Delta\chi^2_{\text{global}}$  behaviour for all eigenvectors found where  $\text{N}^3\text{LO}$   $K$ -factor parameters are considered completely decorrelated ( $H_{ij} + K_{ij}$ ) or correlated ( $H'_{ij}$ ) with all other parameters.

**C.1 Case 1: Decorrelated  $K$ -factor parameters**

Figure 50 displays the tolerance landscape for each eigenvector found from the decorrelated ( $H_{ij} + K_{ij}$ ) Hessian described in Sect. 2. Across all 52 eigenvectors (42 PDF +  $\text{N}^3\text{LO}$  DIS theory and 10  $\text{N}^3\text{LO}$   $K$ -factor) we show an overall general agreement with the quadratic assumption similar to that found at NNLO.

**C.2 Case 2: Correlated  $K$ -factor parameters**

Figure 51 displays the tolerance landscape for each eigenvector found from the correlated ( $H'_{ij}$ ) Hessian described in Sect. 2. Across all 52 eigenvectors we show an overall general agreement with the quadratic assumption similar to that found at NNLO.



**Fig. 50** Dynamic tolerances for each eigenvector direction in the case of complete decorrelation between the theory and PDF parameters, and the  $K$ -factor parameters included in the PDF fit

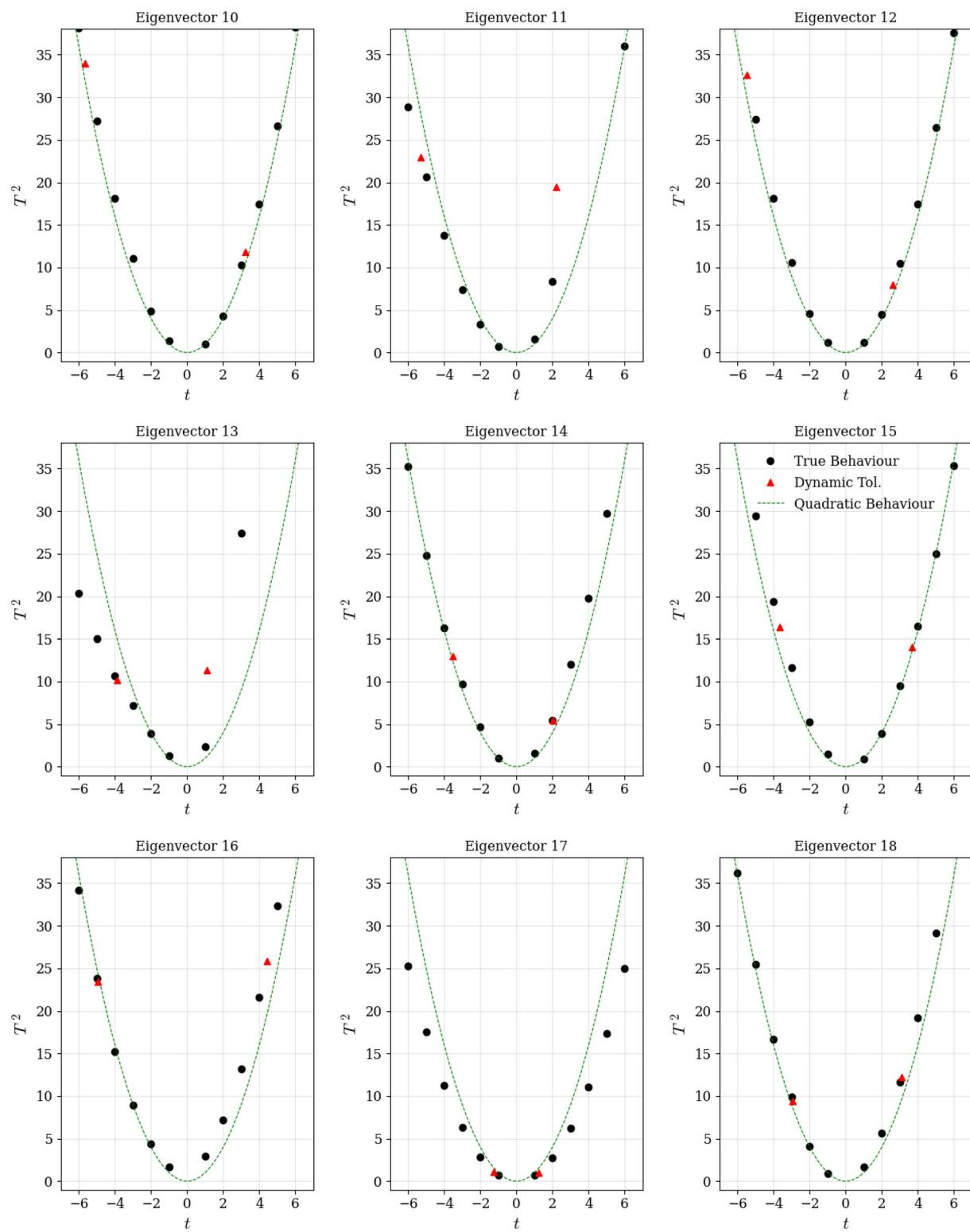
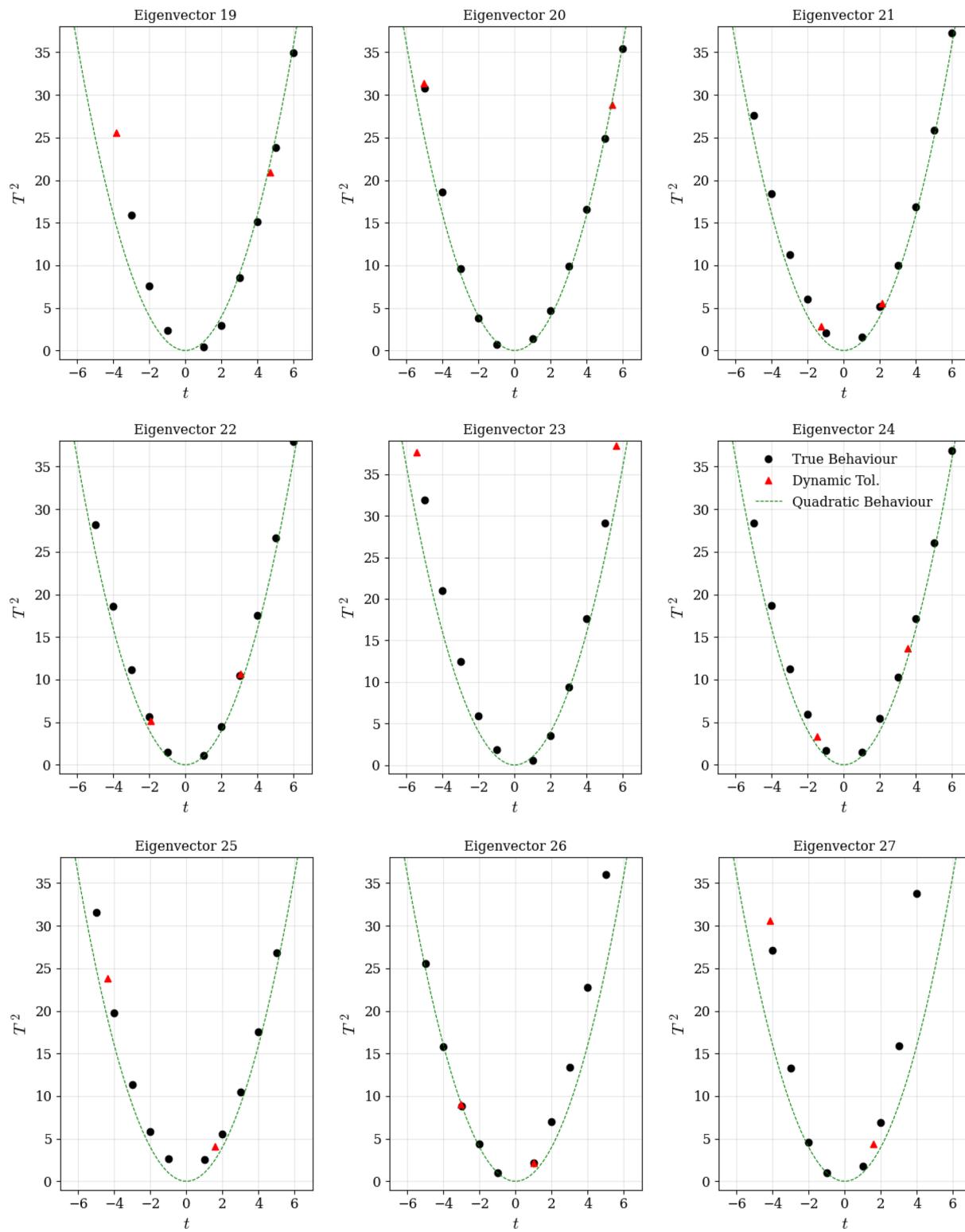


Fig. 50 continued

**Fig. 50** continued

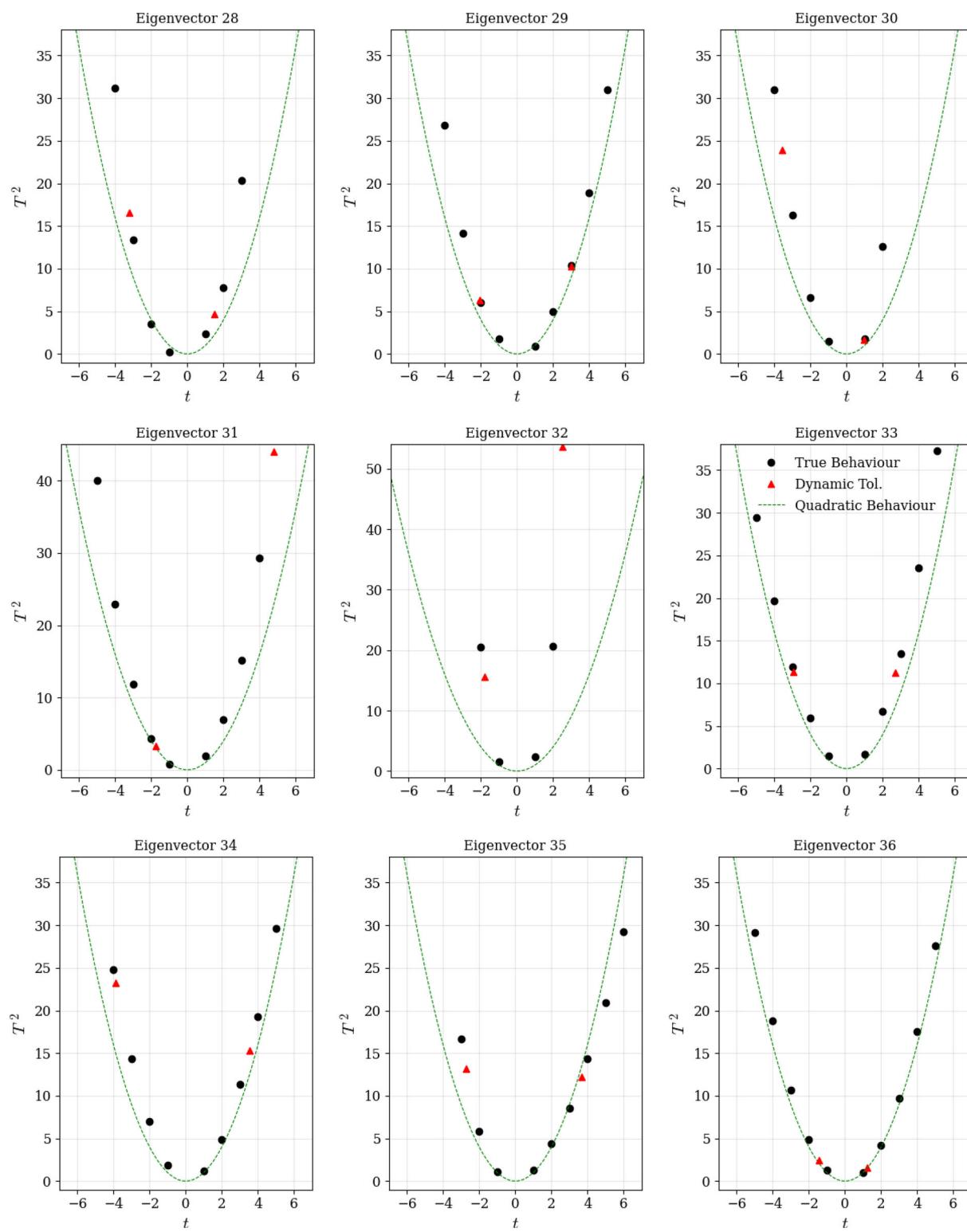


Fig. 50 continued

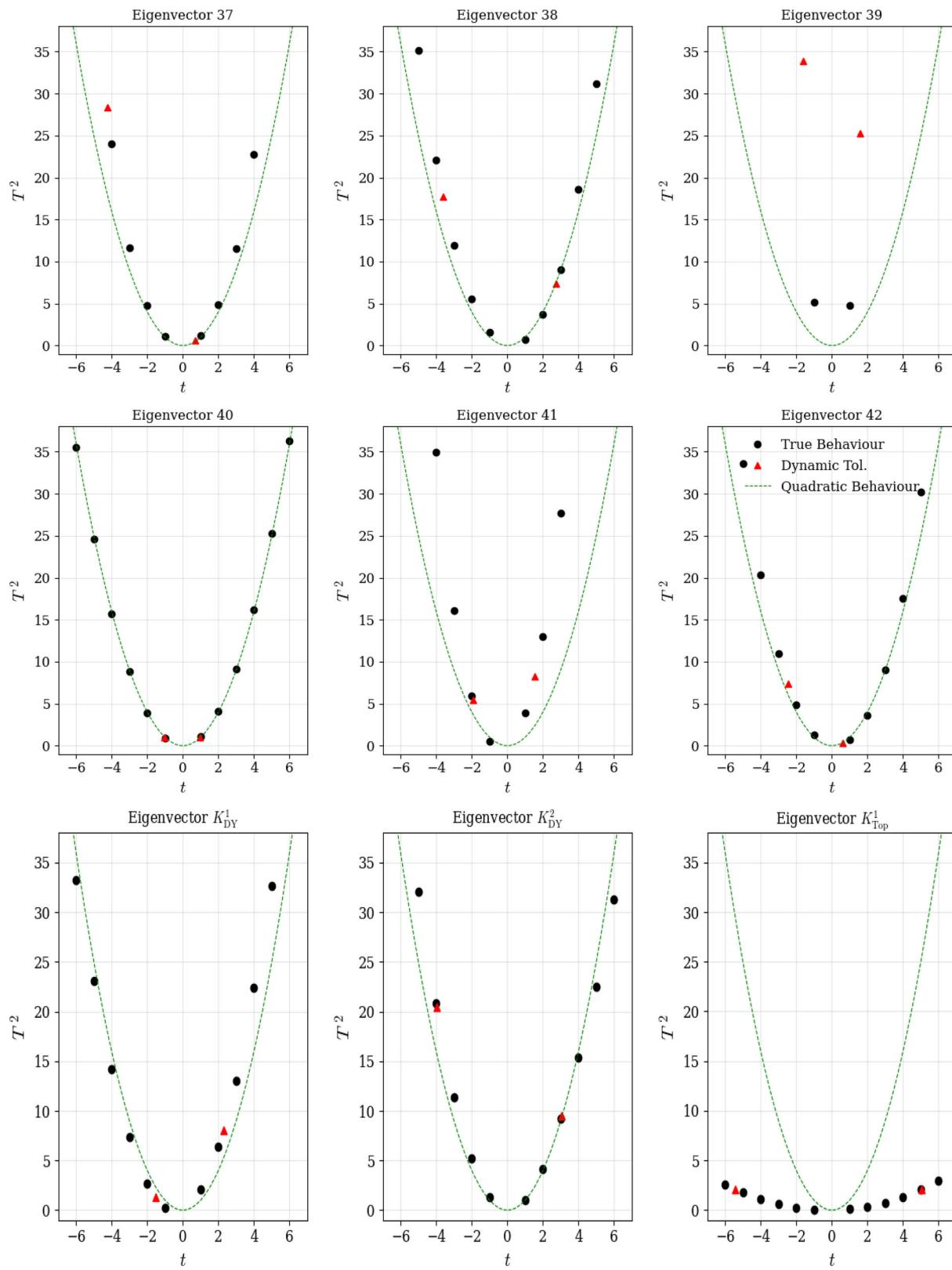


Fig. 50 continued

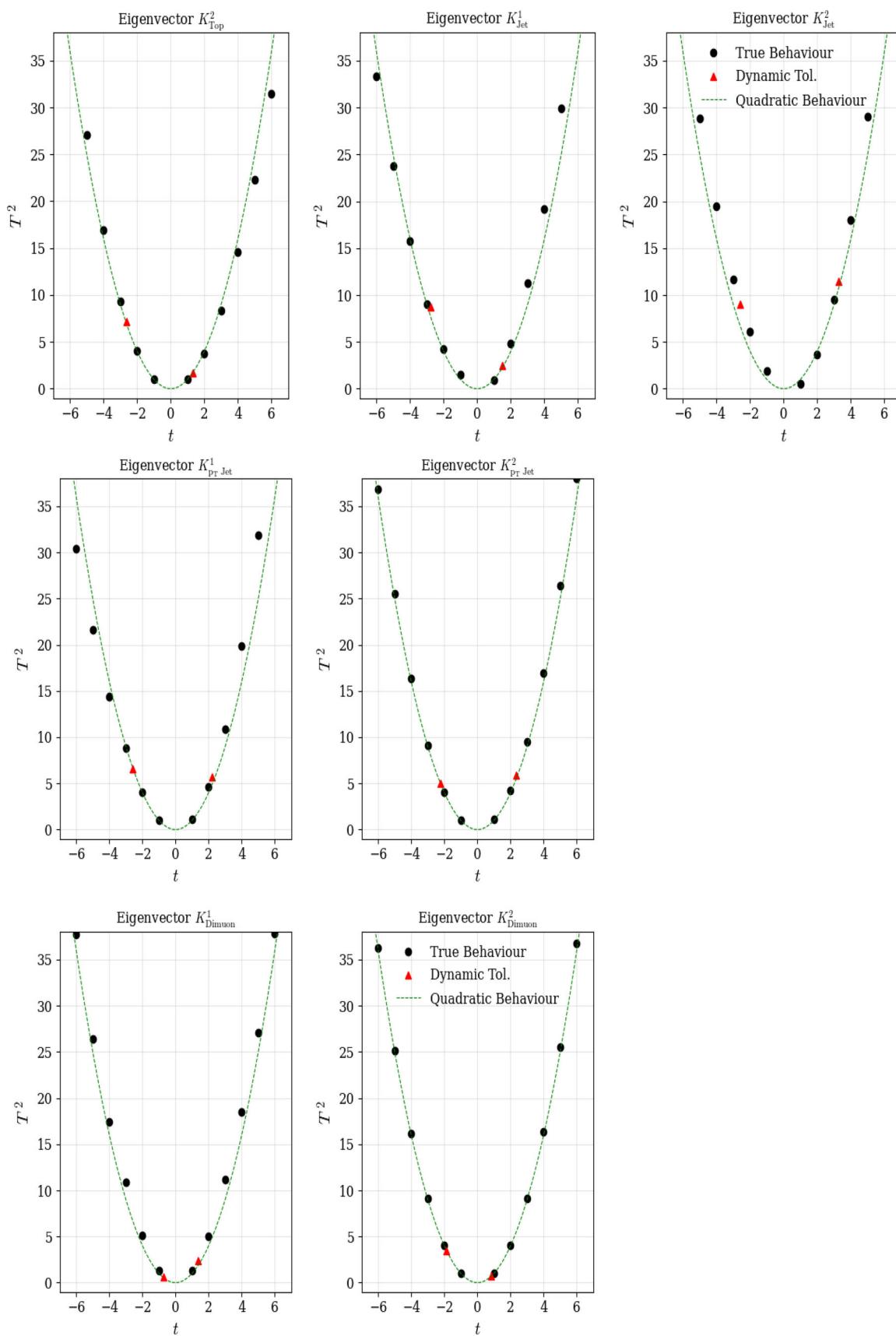
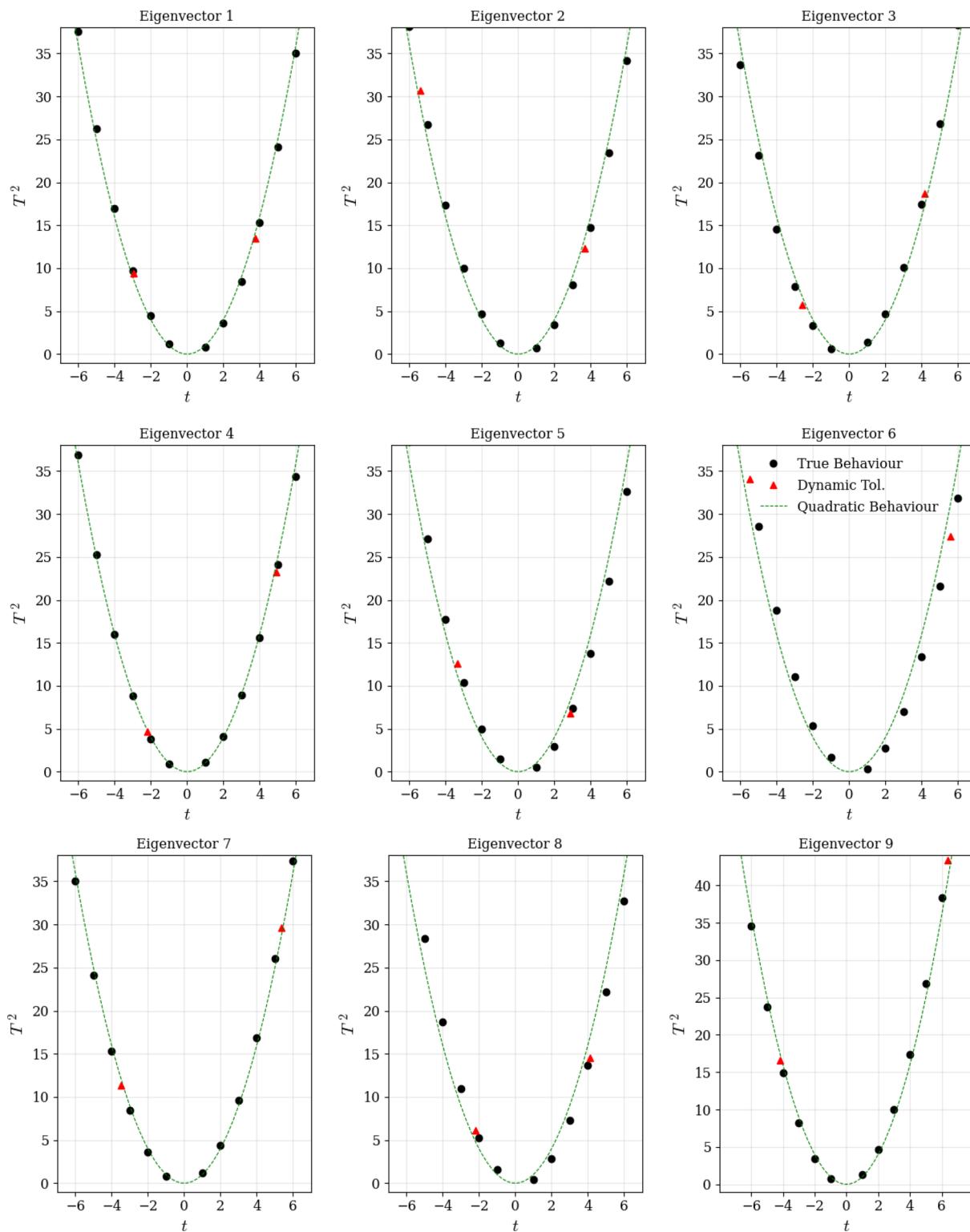


Fig. 50 continued



**Fig. 51** Dynamic tolerances for each eigenvector direction in the case of complete correlation between all theory, PDF and  $K$ -factor parameters included in the PDF fit

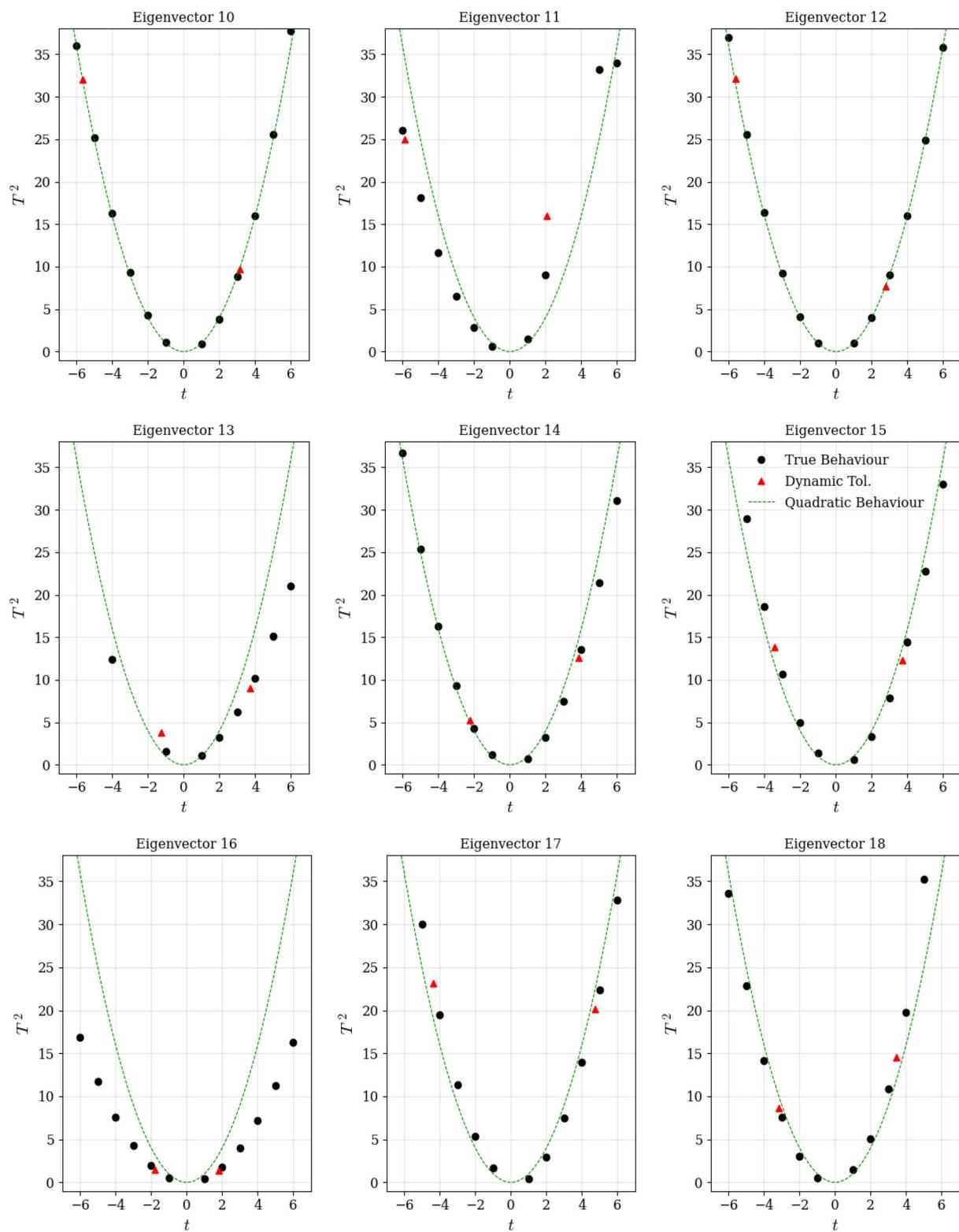
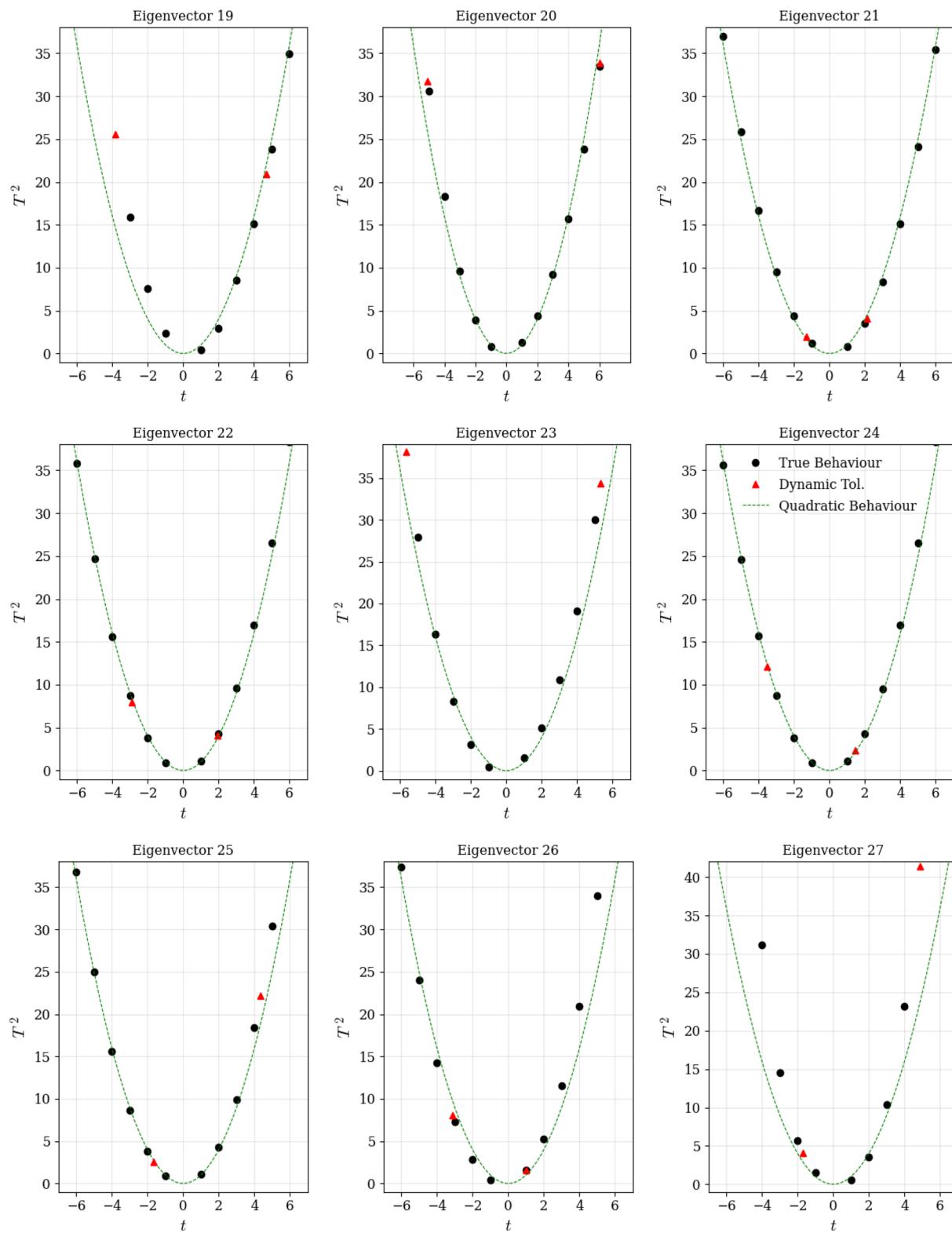


Fig. 51 continued



**Fig. 51** continued

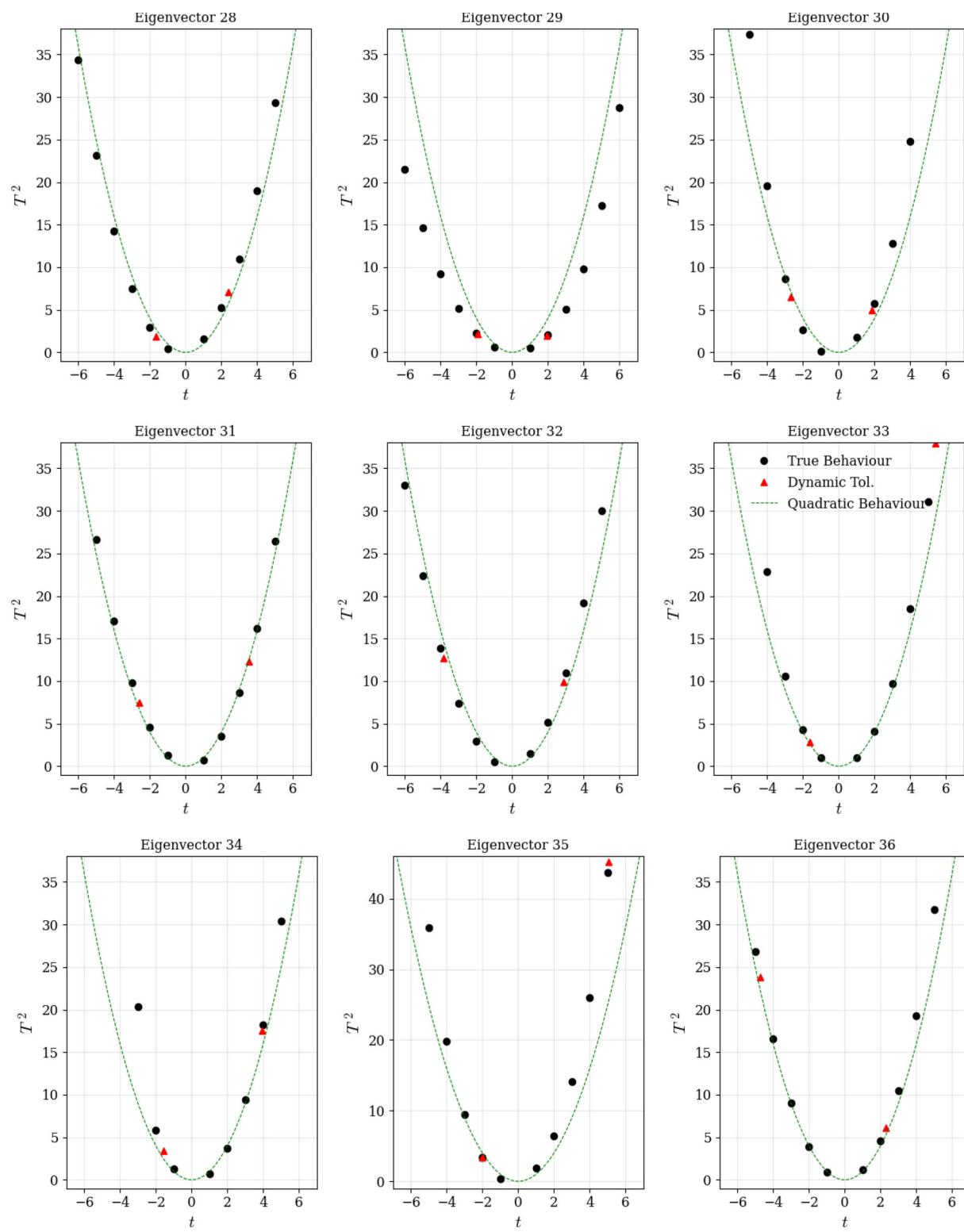
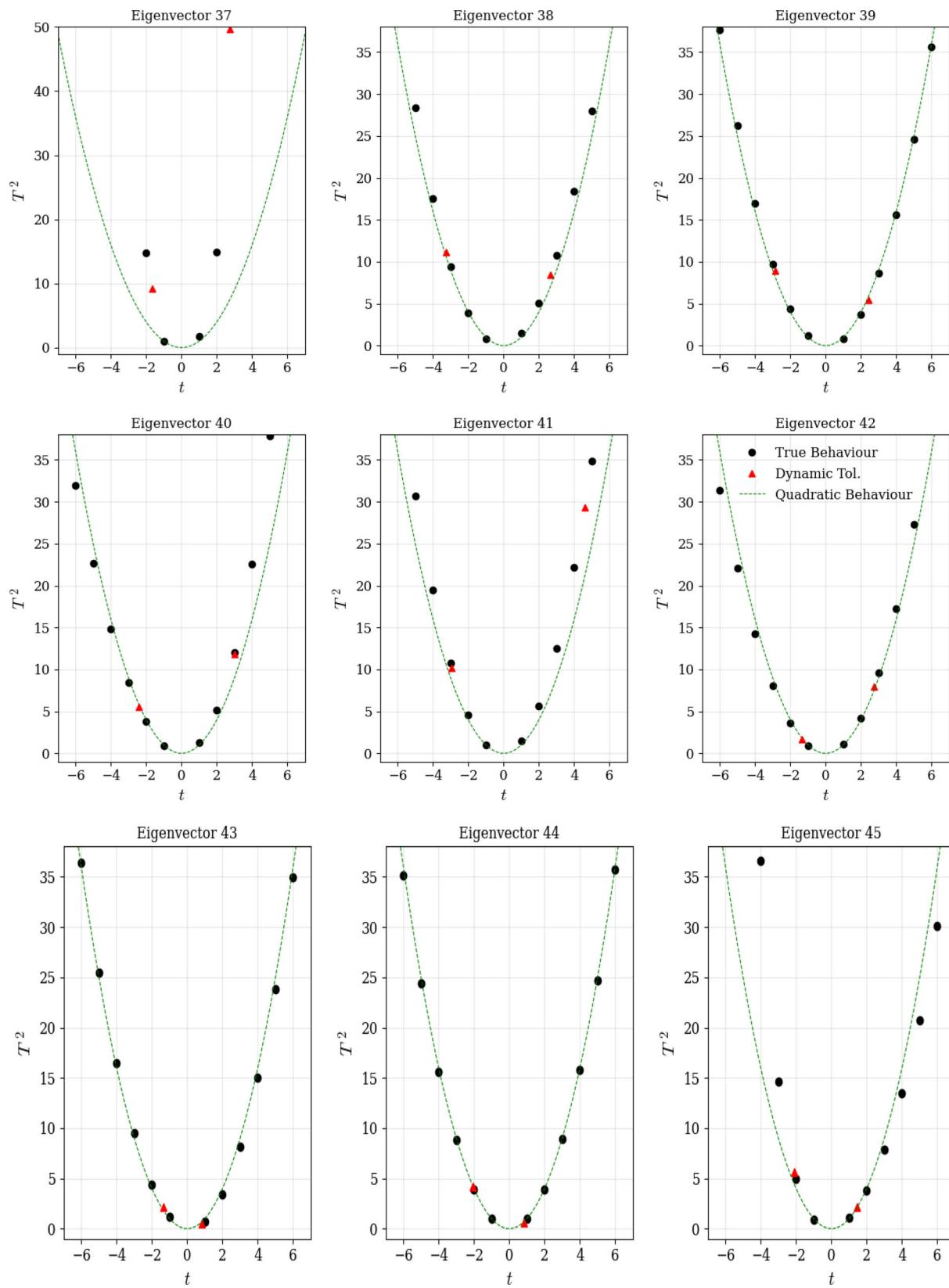
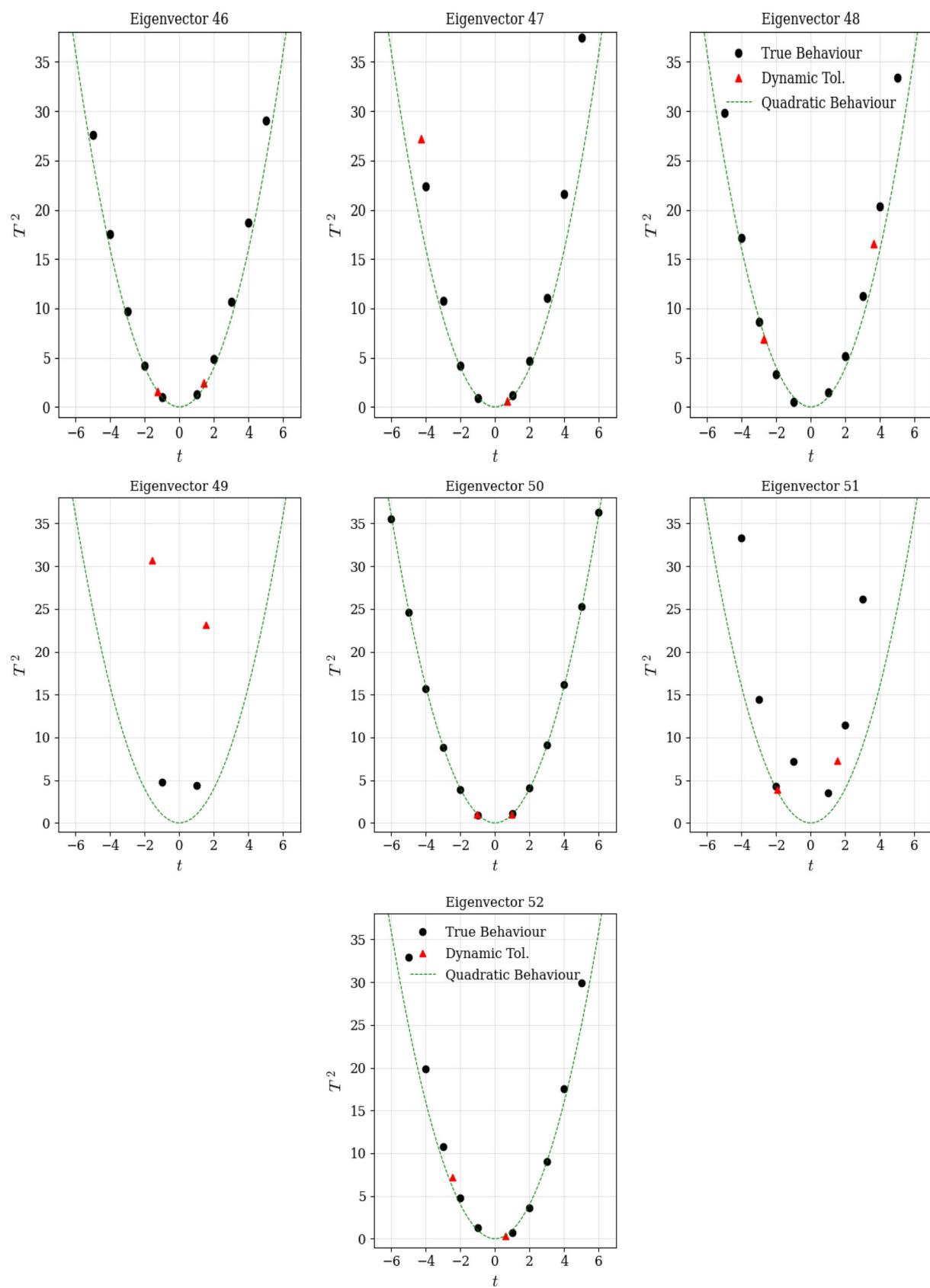


Fig. 51 continued

**Fig. 51** continued

**Fig. 51** continued

**Table 21** Higgs production cross section results via gluon fusion using  $N^3LO$  and  $NNLO$  hard cross sections combined with  $NNLO$  and  $aN^3LO$  PDFs. All PDFs are at the standard choice  $\alpha_s = 0.118$ . These results

$\sigma$ order	PDF order	$\sigma + \Delta\sigma_+ - \Delta\sigma_-$ (pb)	$\sigma$ (pb) + $\Delta\sigma_+ - \Delta\sigma_-$ (%)
PDF uncertainties			
$N^3LO$	$aN^3LO$ (no theory unc.)	$43.803 + 0.685 - 0.526$	$42.709 + 2.81\% - 3.14\%$
	$aN^3LO (H_{ij} + K_{ij})$	$43.803 + 0.795 - 0.732$	$42.709 + 3.30\% - 3.17\%$
	$aN^3LO (H'_{ij})$	$43.803 + 0.785 - 0.737$	$42.709 + 3.39\% - 3.08\%$
$NNLO$	$NNLO$	$46.243 + 0.524 - 0.563$	$46.243 + 1.13\% - 1.22\%$
	$NNLO$	$42.129 + 0.472 - 0.510$	$42.129 + 1.12\% - 1.21\%$
PDF + Scale uncertainties			
$N^3LO$	$aN^3LO$ (no theory unc.)	$43.803 + 1.723 - 2.519$	$42.709 + 4.68\% - 6.44\%$
	$aN^3LO (H_{ij} + K_{ij})$	$43.803 + 1.770 - 2.570$	$42.709 + 4.89\% - 6.45\%$
	$aN^3LO (H'_{ij})$	$43.803 + 1.766 - 2.571$	$42.709 + 4.95\% - 6.41\%$
$NNLO$	$NNLO$	$46.243 + 1.845 - 3.078$	$46.243 + 3.99\% - 6.66\%$
	$NNLO$	$42.129 + 4.989 - 5.106$	$42.129 + 11.84\% - 12.12\%$

## Appendix D: Higgs gluon fusion $\mu = m_H$ results

Provided in Table 21 are the results analogous to those in Table 14 but with the central scale set to  $\mu = \mu_f = \mu_r = m_H$ . These results show a higher level of stability for  $aN^3LO$  PDFs with the chosen central scale. By the renormalisation group arguments, this scale dependence should disappear at all orders in perturbation theory. Therefore the results here suggest that the  $aN^3LO$  PDFs are following this trend.

## References

1. M. Mojaza, S.J. Brodsky, X.-G. Wu, Phys. Rev. Lett. **110**, 192001 (2013). [arXiv:1212.0049](#)
2. M. Czakon, D. Heymes, A. Mitov, JHEP **04**, 071 (2017). [arXiv:1606.03350](#)
3. S. Bailey, T. Cridge, L.A. Harland-Lang, A.D. Martin, R.S. Thorne, Eur. Phys. J. C **81**, 341 (2021). [arXiv:2012.04684](#)
4. T.-J. Hou et al., Phys. Rev. D **103**, 014013 (2021). [arXiv:1912.10053](#)
5. NNPDF, R.D. Ball et al., Eur. Phys. J. C **77**, 663 (2017). [arXiv:1706.00428](#)
6. NNPDF, R.D. Ball et al., Eur. Phys. J. C **82**, 428 (2022). [arXiv:2109.02653](#)
7. NNPDF, R.D. Ball et al., Eur. Phys. J. C **81**, 958 (2021). [arXiv:2109.02671](#)
8. PDF4LHC Working Group, R.D. Ball et al., J. Phys. G **49**, 080501 (2022). [arXiv:2203.05506](#)
9. S. Alekhin, J. Blümlein, S. Moch, R. Placakyte, Phys. Rev. D **96**, 014011 (2017). [arXiv:1701.05838](#)
10. ATLAS, G. Aad et al., Eur. Phys. J. C **82**, 438 (2022). [arXiv:2112.11266](#)
11. NNPDF, R. Abdul Khalek et al., Eur. Phys. J. C **79**, 931 (2019). [arXiv:1906.10698](#)
12. L.A. Harland-Lang, R.S. Thorne, Eur. Phys. J. C **79**, 225 (2019). [arXiv:1811.08434](#)
13. J. Gao, L. Harland-Lang, J. Rojo, Phys. Rep. **742**, 1 (2018). [arXiv:1709.04922](#)
14. R.D. Ball, R.L. Pearson, Eur. Phys. J. C **81**, 830 (2021). [arXiv:2105.05114](#)
15. M. Bonvini, Eur. Phys. J. C **80**, 989 (2020). [arXiv:2006.16293](#)
16. R.D. Ball et al., Eur. Phys. J. C **78**, 321 (2018). [arXiv:1710.05935](#)
17. xFitter Developers' Team, H. Abdolmaleki et al., Eur. Phys. J. C **78**, 621 (2018). [arXiv:1802.00064](#)
18. F.J. Tackmann, Theory Uncertainties from Nuisance Parameters, SCET 2019 Workshop, (2019)
19. F.J. Tackmann, Beyond Scale Variations: Perturbative Theory Uncertainties from Nuisance Parameters, DESY-19-021, (2023)
20. G. Altarelli, G. Parisi, Nucl. Phys. B **126**, 298 (1977)
21. S. Moch, B. Ruijl, T. Ueda, J.A.M. Vermaseren, A. Vogt, JHEP **10**, 041 (2017). [arXiv:1707.08315](#)
22. E.G. Floratos, D.A. Ross, C.T. Sachrajda, Nucl. Phys. B **152**, 493 (1979)
23. A. Gonzalez-Arroyo, C. Lopez, Nucl. Phys. B **166**, 429 (1980)
24. W. Furmanski, R. Petronzio, Phys. Lett. B **97**, 437 (1980)
25. E.G. Floratos, C. Kounnas, R. Lacaze, Nucl. Phys. B **192**, 417 (1981)
26. S. Moch, J.A.M. Vermaseren, A. Vogt, Nucl. Phys. B **688**, 101 (2004). [arXiv:hep-ph/0403192](#)
27. A. Vogt, S. Moch, J.A.M. Vermaseren, Nucl. Phys. B **691**, 129 (2004). [arXiv:hep-ph/0404111](#)
28. S. Catani, F. Hautmann, Nucl. Phys. B **427**, 475 (1994). [arXiv:hep-ph/9405388](#)
29. L.N. Lipatov, Sov. J. Nucl. Phys. **23**, 338 (1976)
30. E.A. Kuraev, L.N. Lipatov, V.S. Fadin, Sov. Phys. JETP **45**, 199 (1977)
31. I.I. Balitsky, L.N. Lipatov, Sov. J. Nucl. Phys. **28**, 822 (1978)
32. V.S. Fadin, L.N. Lipatov, Phys. Lett. B **429**, 127 (1998). [arXiv:hep-ph/9802290](#)
33. M. Ciafaloni, G. Camici, Phys. Lett. B **430**, 349 (1998). [arXiv:hep-ph/9803389](#)
34. M. Bonvini, S. Marzani, JHEP **06**, 145 (2018). [arXiv:1805.06460](#)
35. A. Vogt et al., PoS **LL2018**, 050 (2018). [arXiv:1808.08981](#)
36. S. Moch, B. Ruijl, T. Ueda, J.A.M. Vermaseren, A. Vogt, Phys. Lett. B **825**, 136853 (2022). [arXiv:2111.15561](#)
37. R.K. Ellis, W.J. Stirling, B.R. Webber, *QCD and Collider Physics* (Cambridge University Press, Cambridge, 2011)

are found with  $\mu = m_H$  unless stated otherwise, with the values for  $\mu = m_H/2$  supplied in Table 14

38. M. Buza, Y. Matiounine, J. Smith, W.L. van Neerven, Eur. Phys. J. C **1**, 301 (1998). [arXiv:hep-ph/9612398](#)

39. M. Buza, Y. Matiounine, J. Smith, W.L. van Neerven, Nucl. Phys. B **485**, 420 (1997). [arXiv:hep-ph/9608342](#)

40. M.A.G. Aivazis, J.C. Collins, F.I. Olness, W.-K. Tung, Phys. Rev. D **50**, 3102 (1994). [arXiv:hep-ph/9312319](#)

41. R.S. Thorne, R.G. Roberts, Phys. Rev. D **57**, 6871 (1998). [arXiv:hep-ph/9709442](#)

42. A.D. Martin, W.J. Stirling, R.S. Thorne, G. Watt, Eur. Phys. J. C **63**, 189 (2009). [arXiv:0901.0002](#)

43. R.S. Thorne, Phys. Rev. D **73**, 054019 (2006). [arXiv:hep-ph/0601245](#)

44. R.S. Thorne, Phys. Rev. D **86**, 074017 (2012). [arXiv:1201.6180](#)

45. E. Laenen, S. Riemersma, J. Smith, W.L. van Neerven, Nucl. Phys. B **392**, 229 (1993)

46. B.W. Harris, J. Smith, Nucl. Phys. B **452**, 109 (1995). [arXiv:hep-ph/9503484](#)

47. S. Catani, M. Ciafaloni, F. Hautmann, Nucl. Phys. B **366**, 135 (1991)

48. E. Laenen, S.-O. Moch, Phys. Rev. D **59**, 034027 (1999). [arXiv:hep-ph/9809550](#)

49. H. Kawamura, N.A. Lo Presti, S. Moch, A. Vogt, Nucl. Phys. B **864**, 399 (2012). [arXiv:1205.5727](#)

50. I. Bierenbaum, J. Blümlein, S. Klein, Nucl. Phys. B **820**, 417 (2009). [arXiv:0904.3563](#)

51. J. Ablinger et al., Nucl. Phys. B **886**, 733 (2014). [arXiv:1406.4654](#)

52. J. Blümlein, P. Marquard, C. Schneider, K. Schönwald, Nucl. Phys. B **971**, 115542 (2021). [arXiv:2107.06267](#)

53. J. Ablinger et al., Nucl. Phys. B **890**, 48 (2014). [arXiv:1409.1135](#)

54. J. Ablinger et al., Nucl. Phys. B **882**, 263 (2014). [arXiv:1402.0359](#)

55. J. Blümlein et al., PoS QCDEV2017, 031 (2017). [arXiv:1711.07957](#)

56. M.A.G. Aivazis, F.I. Olness, W.-K. Tung, Phys. Rev. D **50**, 3085 (1994). [arXiv:hep-ph/9312318](#)

57. J.A.M. Vermaseren, A. Vogt, S. Moch, Nucl. Phys. B **724**, 3 (2005). [arXiv:hep-ph/0504242](#)

58. W.L. van Neerven, A. Vogt, Nucl. Phys. B **568**, 263 (2000). [arXiv:hep-ph/9907472](#)

59. W.L. van Neerven, A. Vogt, Nucl. Phys. B **588**, 345 (2000). [arXiv:hep-ph/0006154](#)

60. J. Ablinger et al., Nucl. Phys. B **885**, 280 (2014). [arXiv:1405.4259](#)

61. J. Ablinger et al., Nucl. Part. Phys. Proc. **258–259**, 37 (2015). [arXiv:1409.1435](#)

62. CMS, S. Chatrchyan et al., JHEP **02**, 013 (2014). [arXiv:1310.1138](#)

63. C. Duhr, F. Dulat, B. Mistlberger, Phys. Rev. Lett. **125**, 172001 (2020). [arXiv:2001.07717](#)

64. X. Chen et al., Phys. Rev. Lett. **128**, 052001 (2022). [arXiv:2107.09085](#)

65. C. Duhr, B. Mistlberger, JHEP **03**, 116 (2022). [arXiv:2111.10379](#)

66. N. Kidonakis, Snowmass Summer Study (2022). [arXiv:2203.03698](#)

67. R.D. Ball, M. Bonvini, S. Forte, S. Marzani, G. Ridolfi, Nucl. Phys. B **874**, 746 (2013). [arXiv:1303.3590](#)

68. M. Bonvini, R.D. Ball, S. Forte, S. Marzani, G. Ridolfi, J. Phys. G **41**, 095002 (2014). [arXiv:1404.3204](#)

69. M. Bonvini, S. Marzani, C. Muselli, L. Rottoli, JHEP **08**, 105 (2016). [arXiv:1603.08000](#)

70. T. Ahmed, M.C. Kumar, P. Mathews, N. Rana, V. Ravindran, Eur. Phys. J. C **76**, 355 (2016). [arXiv:1510.02235](#)

71. M. Bonvini, S. Marzani, Phys. Rev. Lett. **120**, 202003 (2018). [arXiv:1802.07758](#)

72. M. Bonvini, Eur. Phys. J. C **78**, 834 (2018). [arXiv:1805.08785](#)

73. M. Bonvini, EPJ Web Conf. **60**, 12008 (2013). [arXiv:1306.6633](#)

74. C. Anastasiou et al., JHEP **03**, 091 (2015). [arXiv:1411.3584](#)

75. C. Anastasiou et al., JHEP **05**, 058 (2016). [arXiv:1602.00695](#)

76. B. Mistlberger, JHEP **05**, 028 (2018). [arXiv:1802.00833](#)

77. LHCb, R. Aaij et al., JHEP **08**, 039 (2015). [arXiv:1505.07024](#)

78. LHCb, R. Aaij et al., JHEP **01**, 155 (2016). [arXiv:1511.08039](#)

79. ATLAS, M. Aaboud et al., Eur. Phys. J. C **77**, 367 (2017). [arXiv:1612.03016](#)

80. J.C. Webb, Measurement of continuum dimuon production in 800-GeV/C proton nucleon collisions. PhD thesis, New Mexico State U. (2003). [arXiv:hep-ex/0301031](#)

81. NuSea, R.S. Towell et al., Phys. Rev. D **64**, 052002 (2001). [arXiv:hep-ex/0103030](#)

82. D0, V.M. Abazov et al., Phys. Rev. D **76**, 012003 (2007). [arXiv:hep-ex/0702025](#)

83. CDF, T.A. Altonen et al., Phys. Lett. B **692**, 232 (2010). [arXiv:0908.3914](#)

84. D0, V.M. Abazov et al., Phys. Rev. D **88**, 091102 (2013). [arXiv:1309.2591](#)

85. CDF, T. Altonen et al., Phys. Rev. Lett. **102**, 181801 (2009). [arXiv:0901.2169](#)

86. D0, V.M. Abazov et al., Phys. Rev. Lett. **101**, 211801 (2008). [arXiv:0807.3367](#)

87. ATLAS, G. Aad et al., Phys. Rev. D **85**, 072004 (2012). [arXiv:1109.5141](#)

88. CMS, S. Chatrchyan et al., Phys. Rev. Lett. **109**, 111806 (2012). [arXiv:1206.2598](#)

89. CMS, S. Chatrchyan et al., JHEP **04**, 050 (2011). [arXiv:1103.3470](#)

90. LHCb, R. Aaij et al., JHEP **02**, 106 (2013). [arXiv:1212.4620](#)

91. LHCb, R. Aaij et al., JHEP **06**, 058 (2012). [arXiv:1204.1620](#)

92. CMS, S. Chatrchyan et al., Phys. Rev. D **85**, 032002 (2012). [arXiv:1110.4973](#)

93. ATLAS, G. Aad et al., Phys. Lett. B **725**, 223 (2013). [arXiv:1305.4192](#)

94. CMS, S. Chatrchyan et al., JHEP **12**, 030 (2013). [arXiv:1310.7291](#)

95. LHCb, R. Aaij et al., JHEP **05**, 109 (2015). [arXiv:1503.00963](#)

96. CMS, V. Khachatryan et al., Eur. Phys. J. C **76**, 469 (2016). [arXiv:1603.01803](#)

97. D0, V.M. Abazov et al., Phys. Rev. Lett. **112**, 151803 (2014). [arXiv:1312.2895](#). [Erratum: Phys. Rev. Lett. **114**, 049901 (2015)]

98. ATLAS, G. Aad et al., JHEP **08**, 009 (2016). [arXiv:1606.01736](#)

99. ATLAS, G. Aad et al., Eur. Phys. J. C **79**, 760 (2019). [arXiv:1904.05631](#)

100. ATLAS, M. Aaboud et al., JHEP **12**, 059 (2017). [arXiv:1710.05167](#)

101. CMS, S. Chatrchyan et al., Phys. Rev. D **90**, 072006 (2014). [arXiv:1406.0324](#)

102. CDF, A. Abulencia et al., Phys. Rev. D **75**, 092006 (2007). [arXiv:hep-ex/0701051](#). [Erratum: Phys. Rev. D **75**, 119901 (2007)]

103. D0, V.M. Abazov et al., Phys. Rev. D **85**, 052006 (2012). [arXiv:1110.3771](#)

104. ATLAS, G. Aad et al., JHEP **02**, 153 (2015). [arXiv:1410.8857](#). [Erratum: JHEP **09**, 141 (2015)]

105. CMS, V. Khachatryan et al., JHEP **03**, 156 (2017). [arXiv:1609.05331](#)

106. CMS, V. Khachatryan et al., Eur. Phys. J. C **76**, 265 (2016). [arXiv:1512.06212](#)

107. L.A. Harland-Lang, A.D. Martin, R.S. Thorne, Eur. Phys. J. C **78**, 248 (2018). [arXiv:1711.05757](#)

108. ATLAS, M. Aaboud et al., JHEP **09**, 020 (2017). [arXiv:1706.03192](#)

109. R. Abdul Khalek et al., Eur. Phys. J. C **80**, 797 (2020). [arXiv:2005.11327](#)

110. ATLAS, G. Aad et al., Eur. Phys. J. C **76**, 291 (2016). [arXiv:1512.02192](#)

111. ATLAS, M. Aaboud et al., JHEP **05**, 077 (2018). [arXiv:1711.03296](#)

112. CMS, V. Khachatryan et al., Eur. Phys. J. C **75**, 542 (2015). [arXiv:1505.04480](#)

113. N. Kidonakis, G.F. Sterman, Nucl. Phys. B **505**, 321 (1997). [arXiv:hep-ph/9705234](https://arxiv.org/abs/hep-ph/9705234)
114. N. Kidonakis, Phys. Rev. Lett. **102**, 232003 (2009). [arXiv:0903.2561](https://arxiv.org/abs/0903.2561)
115. N. Kidonakis, Phys. Rev. D **82**, 114030 (2010). [arXiv:1009.4935](https://arxiv.org/abs/1009.4935)
116. N. Kidonakis, 15th International Symposium on Radiative Corrections, (2021). [arXiv: 2109.14102](https://arxiv.org/abs/2109.14102)
117. CDF, D0, T.A. Aaltonen et al., Phys. Rev. D **89**, 072001 (2014). [arXiv:1309.7570](https://arxiv.org/abs/1309.7570)
118. ATLAS, G. Aad et al., Eur. Phys. J. C **71**, 1577 (2011). [arXiv:1012.1792](https://arxiv.org/abs/1012.1792)
119. ATLAS, G. Aad et al., Phys. Lett. B **707**, 459 (2012). [arXiv:1108.3699](https://arxiv.org/abs/1108.3699)
120. ATLAS, G. Aad et al., Phys. Lett. B **711**, 244 (2012). [arXiv:1201.1889](https://arxiv.org/abs/1201.1889)
121. ATLAS, G. Aad et al., JHEP **05**, 059 (2012). [arXiv:1202.4892](https://arxiv.org/abs/1202.4892)
122. ATLAS, G. Aad et al., Phys. Lett. B **717**, 89 (2012). [arXiv:1205.2067](https://arxiv.org/abs/1205.2067)
123. ATLAS, G. Aad et al., Eur. Phys. J. C **73**, 2328 (2013). [arXiv:1211.7205](https://arxiv.org/abs/1211.7205)
124. CMS, S. Chatrchyan et al., Phys. Rev. D **85**, 112007 (2012). [arXiv:1203.6810](https://arxiv.org/abs/1203.6810)
125. CMS, S. Chatrchyan et al., JHEP **11**, 067 (2012). [arXiv:1208.2671](https://arxiv.org/abs/1208.2671)
126. CMS, S. Chatrchyan et al., Phys. Lett. B **720**, 83 (2013). [arXiv:1212.6682](https://arxiv.org/abs/1212.6682)
127. CMS, S. Chatrchyan et al., Eur. Phys. J. C **73**, 2386 (2013). [arXiv:1301.5755](https://arxiv.org/abs/1301.5755)
128. CMS, S. Chatrchyan et al., JHEP **05**, 065 (2013). [arXiv:1302.0508](https://arxiv.org/abs/1302.0508)
129. CMS, S. Chatrchyan et al., JHEP **02**, 024 (2014). [arXiv:1312.7582](https://arxiv.org/abs/1312.7582). [Erratum: JHEP **02**, 102 (2014)]
130. ATLAS, G. Aad et al., Eur. Phys. J. C **76**, 538 (2016). [arXiv:1511.04716](https://arxiv.org/abs/1511.04716)
131. ATLAS, M. Aaboud et al., Phys. Rev. D **94**, 092003 (2016). [arXiv:1607.07281](https://arxiv.org/abs/1607.07281)
132. CMS, A.M. Sirunyan et al., Eur. Phys. J. C **77**, 459 (2017). [arXiv:1703.01630](https://arxiv.org/abs/1703.01630)
133. NuTeV, M. Goncharov et al., Phys. Rev. D **64**, 112006 (2001). [arXiv:hep-ex/0102049](https://arxiv.org/abs/hep-ex/0102049)
134. E.L. Berger, J. Gao, C.S. Li, Z.L. Liu, H.X. Zhu, Phys. Rev. Lett. **116**, 212002 (2016). [arXiv:1601.05430](https://arxiv.org/abs/1601.05430)
135. BCAMS, A.C. Benvenuti et al., Phys. Lett. B **223**, 485 (1989)
136. NMC, M. Arneodo et al., Nucl. Phys. B **483**, 3 (1997). [arXiv:hep-ph/9610231](https://arxiv.org/abs/hep-ph/9610231)
137. L.W. Whitlow, E.M. Riordan, S. Dasu, S. Rock, A. Bodek, Phys. Lett. B **282**, 475 (1992)
138. L.W. Whitlow, S. Rock, A. Bodek, E.M. Riordan, S. Dasu, Phys. Lett. B **250**, 193 (1990)
139. E665, M.R. Adams et al., Phys. Rev. D **54**, 3006 (1996)
140. NuTeV, M. Tzanov et al., Phys. Rev. D **74**, 012008 (2006). [arXiv:hep-ex/0509010](https://arxiv.org/abs/hep-ex/0509010)
141. NMC, M. Arneodo et al., Nucl. Phys. B **487**, 3 (1997). [arXiv:hep-ex/9611022](https://arxiv.org/abs/hep-ex/9611022)
142. H1, ZEUS, H. Abramowicz et al., Eur. Phys. J. C **73**, 2311 (2013). [arXiv:1211.1182](https://arxiv.org/abs/1211.1182)
143. H1, F.D. Aaron et al., Phys. Lett. B **665**, 139 (2008). [arXiv:0805.2809](https://arxiv.org/abs/0805.2809)
144. H1, F.D. Aaron et al., Eur. Phys. J. C **71**, 1579 (2011). [arXiv:1012.4355](https://arxiv.org/abs/1012.4355)
145. ZEUS, S. Chekanov et al., Phys. Lett. B **682**, 8 (2009). [arXiv:0904.1092](https://arxiv.org/abs/0904.1092)
146. CHORUS, G. Onengut et al., Phys. Lett. B **632**, 65 (2006)
147. H1, ZEUS, F. Aaron et al., JHEP **01**, 109 (2010). [arXiv:0911.0884](https://arxiv.org/abs/0911.0884)
148. A. Candido, S. Forte, F. Hekhorn, JHEP **11**, 129 (2020). [arXiv:2006.07377](https://arxiv.org/abs/2006.07377)
149. J. Collins, T.C. Rogers, N. Sato, Phys. Rev. D **105**, 076010 (2022). [arXiv:2111.01170](https://arxiv.org/abs/2111.01170)
150. SeaQuest, J. Dove et al., Nature **590**, 561 (2021). [arXiv:2103.04024](https://arxiv.org/abs/2103.04024). [Erratum: Nature **604**, E26 (2022)]
151. T. Cridge, L.A. Harland-Lang, A.D. Martin, R.S. Thorne, Eur. Phys. J. C **81**, 744 (2021). [arXiv:2106.10289](https://arxiv.org/abs/2106.10289)
152. Particle Data Group, P.A. Zyla et al., PTEP **2020**, 083C01 (2020)
153. M. Cacciari, F.A. Dreyer, A. Karlberg, G.P. Salam, G. Zanderighi, Phys. Rev. Lett. **115**, 082002 (2015). [arXiv:1506.02660](https://arxiv.org/abs/1506.02660). [Erratum: Phys. Rev. Lett. **120**, 139901 (2018)]
154. F.A. Dreyer, A. Karlberg, Phys. Rev. Lett. **117**, 072001 (2016). [arXiv:1606.00840](https://arxiv.org/abs/1606.00840)
155. M. Bonvini, <https://www.ge.infn.it/~bonvini/higgs/>, version 4.1 (2020)
156. M. Cacciari, F. Dreyer, A. Karlberg, G. Salam, and G. Zanderighi, <https://provbfl.hepforge.org/>, version 2.0.2 (2018)
157. A. Buckley et al., Eur. Phys. J. C **75**, 132 (2015). [arXiv:1412.7420](https://arxiv.org/abs/1412.7420)