Prepared for submission to JHEP DESY-12-191, DO-TH 12/34, LPN12-118, SFB/CPP-12-79

Statistical issues in the parton distribution analysis of the Tevatron jet data

S.Alekhin,[a,b,1] **J.Blümlein,**[a] **S.-O.Moch**[a,c]

[a] *DESY, Platanenallee 6, D–15738 Zeuthen, Germany*

[b] *Institute for High Energy Physics, 142281 Protvino, Moscow region, Russia*

[c] *II. Institut für Theoretische Physik, Universität Hamburg, Luruper Chaussee 149, D-22761 Hamburg, Germany*

E-mail: sergey.alekhin@desy.de, johannes.bluemlein@desy.de, sven-olaf.moch@desy.de

Abstract: We analyse a tension between the D0 and CDF inclusive jet data and the perturbative QCD calculations, which are based on the ABKM09 and ABM11 parton distribution functions (PDFs) within the nuisance parameter framework. Particular attention is paid on the uncertainties in the nuisance parameters due to the data fluctuations and the PDF errors. We show that with account of these uncertainties the nuisance parameters do not demonstrate a statistically significant excess. A statistical bias of the estimator based on the nuisance parameters is also discussed.

[1] Corresponding author.

Contents

1 Introduction

Since the first observation of jet production at Tevatron this process is considered as a valuable source of information about the gluon distribution at large x. Indeed, the gluon distribution directly enters into the jet production cross section in contrast to the deep-inelastic-scattering (DIS) process, which provides only an indirect constraint on the gluon distribution, through the QCD evolution. The Tevatron jet production data [1, 2] are used in the global fits of parton distribution functions (PDFs) to improve accuracy of the gluon distribution, particularly at large x. At this end proper statistical treatment of the data is required since uncertainties in the data of Refs. [1, 2] are dominated by the correlated systematics and the simplest χ^2 estimator is inapplicable. In this case one should ideally use the χ^2 estimator including the covariance matrix, which encodes the error correlations. However, for the sake of implementation simplicity an alternative form of estimator is often employed [3]. This form is based on the so-called "nuisance" parameters, which describe a possible shift of the data due to systematic uncertainties. The nuisance parameters entering the estimator are fitted to the data simultaneously with other parameters describing the PDF shape. As a result, the number of fitted parameters dramatically grows. This difficulty is circumvented because the nuisance parameters enter into the estimator of Ref. [3] linearly therefore the χ^2 value can be minimized with respect to the nuisance parameters analytically. As an added feature, the approach based on the nuisance parameters allows for the visualization of any tension between the data and the fitted model since it shows how large a shift of the data provides the best agreement with the model. Moreover, in the same way the best values of the nuisance parameters can be estimated for any given data set, which is not included in the PDF fit, in order to check for potential problems with accommodation of the new data into the fit.

The ABKM09 PDFs [4] and their refined version, ABM11 PDFs [7], were extracted to next-to-next-to-leading-order (NNLO) in perturbative QCD from a combination of the world inclusive DIS data supplemented by the fixed-target data for the Drell-Yan process and dimuon production in the neutrino-nucleon collision. The Tevatron jet data were also included into a variant of the ABKM09 fit [4, 5] and good agreement with other data used in the fit has been achieved. The analysis of Ref. [5] is focused on the impact of the Tevatron data on the Higgs cross section estimate, cf. also [8], and statistical aspects in this analysis have not been detailed. In the present paper we fill this gap by giving a detailed calculation of the nuisance parameters for the ABKM09 and ABM11 PDFs with and without Tevatron jet data included. The paper is organized as follows. In Section 2 we give a brief outline of the formalism used in analysis of the correlated data. Section 3 contains a description of the systematic uncertainties in the Tevatron jet data and the corresponding nuisance parameters in comparison with ones obtained with other PDF sets. Particular attention is payed on the nuisance parameters for the normalization uncertainty and on the impact of this source of uncertainty on the fit results following suggestions of Ref. [6]. Section 4 contains a conclusion.

2 Basics of the correlated data analysis

In case measurements are subject to correlated systematic uncertainties the experimental data $\{y_i\}$ can be represented as follows,

$$y_i = f_i(\vec{\Theta}) + \mu_i \sigma_i + \sum_{k=1}^{N_{syst}} \lambda_k s_{k,i}^{rel} f_i(\vec{\Theta}), \tag{2.1}$$

where f_i is the mathematical expectation of the measurement i depending on the vector of model parameters $\vec{\Theta}$, σ_i is its uncorrelated uncertainty, $s_{k,i}^{rel}$ are the relative correlated uncertainties, which stem from N_{syst} independent sources, and the index i runs over all experimental data points. The independent random variables μ_i and λ_k describe the uncorrelated and correlated fluctuations in the data, respectively. By definition, the uncorrelated fluctuations are independent for each data point. In contrast, the correlated fluctuation due to each source k are common for all data points. Routinely they are related to systematic effects in the data normalization, calibration, corrections, etc. For cross section measurements these factors are applied to the data multiplicatively therefore the systematic errors are commonly multiplicative. With account of the data correlations the χ^2-estimator reads

$$\chi^2 = \sum_{ij} (y_i - f_i) E_{ij} (y_j - f_j). \tag{2.2}$$

The error matrix E_{ij} is the inverse of the positive definite covariance matrix C_{ij}. For the model of Eq. (2.1) it reads

$$C_{ij} = \sigma_i^2 \delta_{ij} + \sum_{k=1}^{N_{syst}} s_{k,i}^{rel} s_{k,j}^{rel} f_i f_j, \tag{2.3}$$

where δ_{ij} is the Kronecker symbol. Alternatively, the error correlations are often taken into account employing the following form of χ^2 [3]

$$\chi^2 = \sum_i \frac{[f_i - (1 - \sum_k \eta_k s_{k,i}^{rel}) y_i]^2}{\sigma_i^2} + \sum_{k=1}^{N_{syst}} \eta_k^2. \tag{2.4}$$

The form of Eq. (2.4) allows for shifts of the data by the correlated uncertainty scaled with the values of the parameters η_k. The latter are fitted simultaneously with the theoretical model parameters $\vec{\Theta}$ and in this way describe the data shifts, which provide the best description of the fitted model. The form of Eq. (2.4) corresponds to the case, when the correlated uncertainties are additive, i.e. the statistical model of the data looks like

$$y_i^{add} = f_i(\vec{\Theta}) + \mu_i \sigma_i + \sum_{k=1}^{N_{syst}} \lambda_k s_{k,i}, \tag{2.5}$$

where $s_{k,i} = s_{k,i}^{rel} y_i$ and the covariance matrix, which should be used in Eq. (2.2) for this data model, reads

$$\left(C^{add}\right)_{ij} = \sigma_i^2 \delta_{ij} + \sum_{k=1}^{N_{syst}} s_{k,i} s_{k,j}. \tag{2.6}$$

The advantage of the estimator in Eq. (2.4) is essentially its technical simplicity since the vector of η_k, which provides the minimum of Eq. (2.4) can be found analytically as a product of two matrices

$$r_k = \sum_{k'=1}^{N_{syst}} {A^{-1}}_{kk'} B_{k'}, \tag{2.7}$$

where

$$A_{kk'} = \delta_{kk'} + \sum_i \frac{s_{k,i} s_{k',i}}{\sigma_i^2} \tag{2.8}$$

and

$$B_{k'} = \sum_i \frac{(y_i - f_i)}{\sigma_i^2} s_{k',i}. \tag{2.9}$$

The value of the estimator in Eq. (2.4) at $\eta_k = r_k$ reads

$$\chi^2_{min} = \sum_i \frac{(f_i - y_i)^2}{\sigma_i^2} - \sum_{k=1}^{N_{syst}} r_k B_k. \tag{2.10}$$

Since the inverse of the additive covariance matrix Eq. (2.6) is

$$\left(C^{add}\right)^{-1}_{i,j} = \frac{\delta_{ij}}{\sigma_i^2} - \frac{1}{\sigma_i^2 \sigma_j^2} \sum_{k,k'=1}^{N_{syst}} s_{k,i} s_{k',j} {A^{-1}}_{kk'} \tag{2.11}$$

the value of χ^2_{min} coincides with the one of Eq. (2.2) for the statistical model of data with additive systematic errors. The nuisance parameters r_k are random variables with average equal to zero and the variances, which read

$$V(r_k) = \sqrt{\sum_{ll'} {A^{-1}}_{kl} C^B_{ll'} {A^{-1}}_{l'k}} \tag{2.12}$$

where

$$C^B_{ll'} = \sum_{ij} s_{l,i} s_{l',j} \frac{C^{add}_{ij}}{\sigma_i^2 \sigma_j^2} \tag{2.13}$$

is the covariance matrix for the vectors $B_{l,l'}$ of Eq. (2.9)[1]. Through $f_i(\Theta)$ entering Eq. (2.9) the nuisance parameters depend on the fitted parameters $\vec{\Theta}$. For the data sets, which are not included into the fit, the nuisance parameters are generally bigger than the ones obtained from a fit, which includes those data sets, due to better a tuning of $\vec{\Theta}$ to the data in the latter case. In the following Section we analyze this trend for the different Tevatron jet data with respect to the ABKM09 [4] and ABM11 [7] fits considering two cases: before and after these data are included into the fit.

3 The Tevatron jet data in the ABKM09 and ABM11 fits

The Tevatron experiments CDF and D0 have accumulated big samples of events with hard jets in the final state and have performed elaborated analyses of these samples with different jet definition algorithms, cf. [9] for a recent review. For brevity we consider in the following only two Tevatron inclusive jet data sets [1, 2] obtained by the D0 and CDF collaborations, respectively, which nonetheless give a representative illustration of the issues discussed in the paper. Both data sets were collected in Run II and each corresponds to an integral luminosity of about 1fb^{-1}.

The D0 analysis of Ref. [1] is based on the midpoint cone algorithm for the jet definition. The D0 data cover the range of $-2.4 \div 2.4$ in the jet rapidity and $50 \div 600$ GeV in the transverse momentum of jet. The published correlated systematic uncertainties in the D0 data are due to the global normalization and 23 additional sources, including the jet energy calibration, resolution, etc. In the present analysis we consider all these sources taking the average in the case of asymmetric errors.[2] The distribution of the nuisance parameters r of Eq. (2.7), which correspond to these 23 sources of systematics, calculated for the NNLO ABKM09 PDFs are given in Fig. 1. The jet production cross sections are obtained with the FastNLO tool [10] and include the NLO corrections [11, 12] and the threshold resummation corrections of Ref. [13]. The D0 nuisance parameters spread in the range from -1.5 to 4.1 and in general their distribution is comparable to the normal Gaussian one. The maximal absolute value of r corresponds to the systematic uncertainty in the general normalization. This reflects the fact that the D0 data systematically overshoot the ABKM09 predictions, cf. Refs. [5, 7]. However with account of the errors in the nuisance parameters due to fluctuations in the data and due to the PDF uncertainties the statistical significance of the spread in the nuisance parameters reduces. To check in details the uncertainty in the D0 normalization nuisance parameter due to the data fluctuation we calculate it for 200 pseudo-data sets generated with Eq. (2.5) and the data errors of Ref. [1] taking a normal Gaussian distribution for the random variables μ and λ. The distribution of the normalization nuisance parameter obtained for these data sets is displayed in Fig. 2. It is comparable to the Gaussian distribution with the average of Eq. (2.7) and variance of Eq. (2.12), which are $r_{norm} = 4.1$ and $V(r_{norm}) = 0.85$, respectively. The error in

[1]Note that the variances of nuisance parameters differ from the square root of the diagonal elements of the inverse Hessian for Eq. (2.4) equal to A^{-1}_{kk}.

[2]The experimental data tables used in the analysis are available from `http://arxiv.org` as an attachment to the arXiv version of our paper.

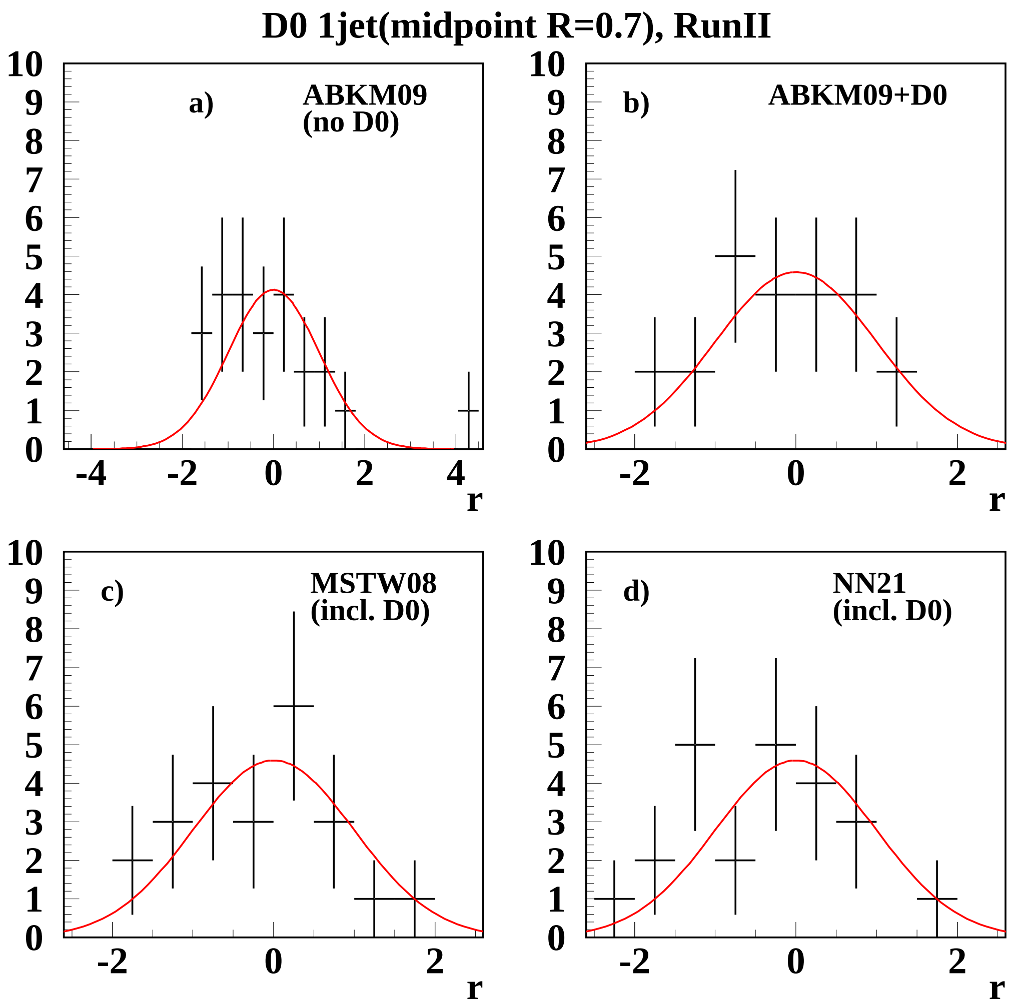

Figure 1. The distribution of nuisance parameters r for the D0 data [1] on the inclusive jet production calculated with the threshold NNLO corrections taken into account and different NNLO PDFs (a): ABKM09 [4]; b): variant of ABKM09 obtained from the fit with the D0 data included [5]; c): MSTW08 [15]; d): NN21 [16]). The curves superimposed display a normal Gaussian distribution normalized on the total number of the nuisance parameters.

the nuisance parameters due to PDFs is estimated in our analysis as a combination of their variation with the change in the PDFs between the central value and each of the 25 PDF sets describing the ABKM09 PDF uncertainties. For the D0 nuisance normalization parameter this gives an additional uncertainty of $\Delta^{PDF}(r_{norm}) = 0.95$. A combination of $V(r_{norm})$ and $\Delta^{PDF}(r_{norm})$ in quadrature gives the total uncertainty $\Delta^{tot}(r_{norm}) = 1.3$. With account of these uncertainties the D0 normalization nuisance parameter is consistent with zero within 3 standard deviations. Other D0 nuisance parameters are also consistent with zero within uncertainties, cf. Fig. 3, therefore the statistical significance of the excess in the normalization nuisance parameter is marginal. Indeed, in the variant of the ABKM09 fit with the D0 data included the nuisance parameters are in general much smaller due to

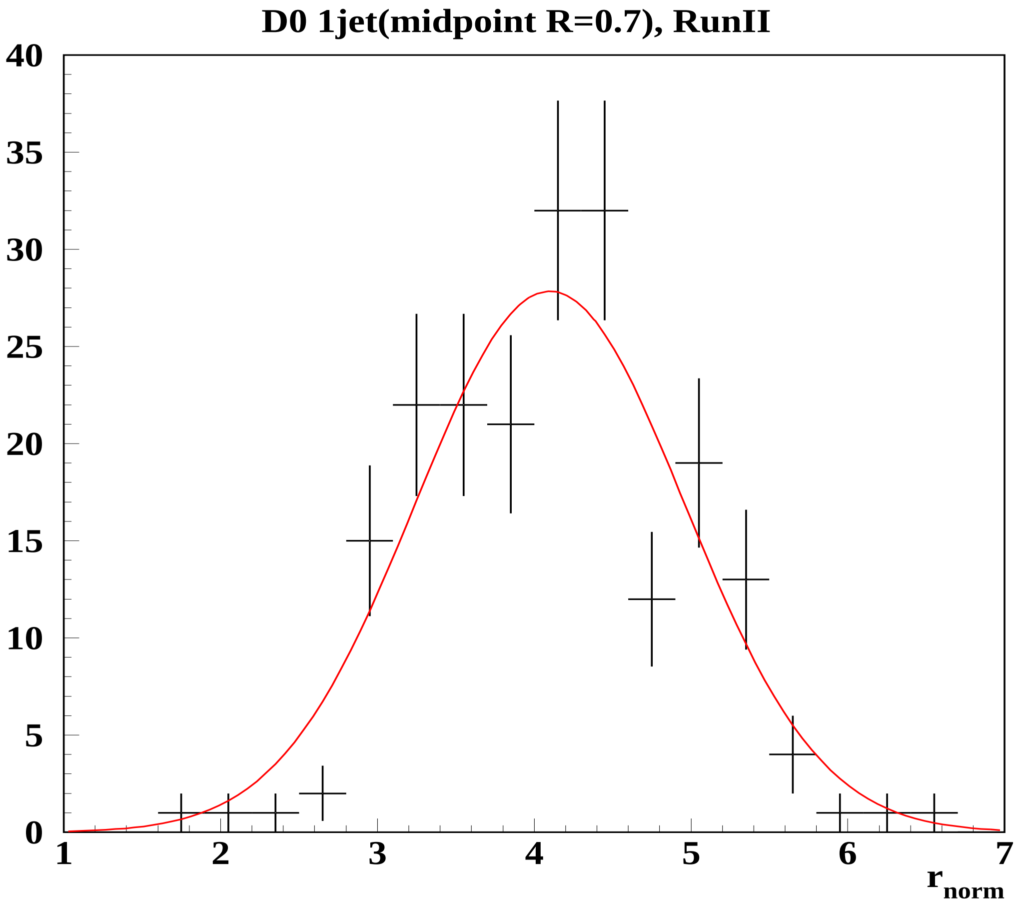

Figure 2. Distribution of the D0 normalization nuisance parameter obtained for 200 pseudo-data sets. The curve superimposed displays a Gaussian distribution with the average of Eq. (2.7) and the variance of Eq. (2.12).

better tuning of the PDFs to the data and the value of normalization nuisance parameter is 1.5 only that is consistent with zero within the errors. To make an explicit check of the impact of the D0 normalization uncertainty on the extracted PDFs we perform one more variant of the ABKM09 fit, with the normalization uncertainty in the D0 data dropped. It turns out that dropping this error does not lead to any essential deterioration of the D0 data description. For the variant of fit without the D0 normalization uncertainty taken into account the value of χ^2 grows by less than 1 for 110 data points. The change in the gluon distribution obtained from these two variants of the fit generally does not exceed its uncertainty, cf. Fig. 4, and for other PDFs it is even smaller. This shows that the normalization error does not play crucial role in the interpretation of the D0 inclusive jet data. This can be also understood qualitatively, since the normalization error in the data is 6.1% only, much smaller than other systematic uncertainties, therefore the latter easily overwhelm the impact of the normalization error.

The CDF data on the inclusive jet cross sections [2] were obtained with the k_T algorithm for the jet definition and cover the range of $-2.1 \div 2.1$ in the jet rapidity and $50 \div 600$ GeV

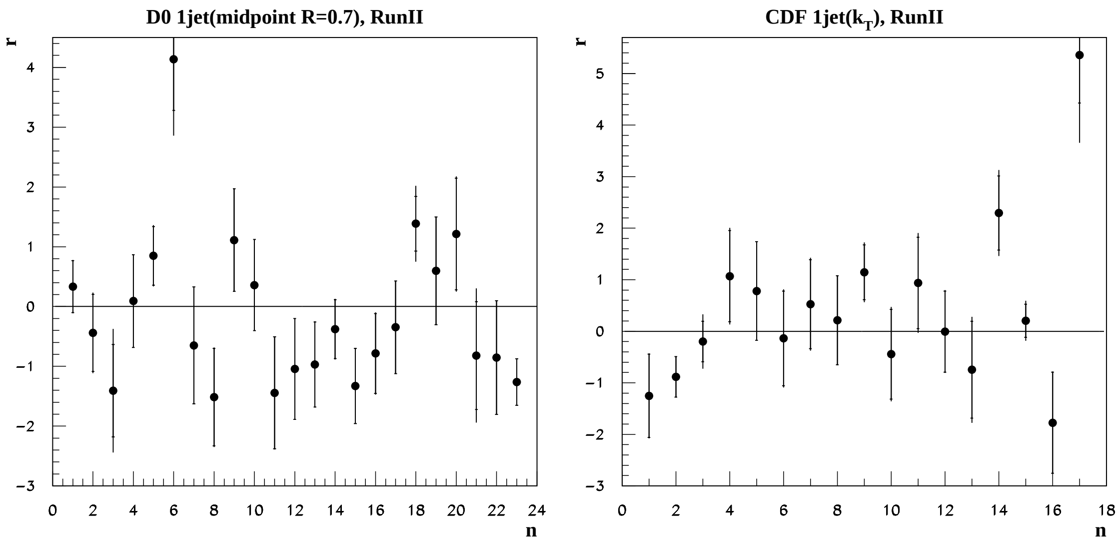

Figure 3. Values of the nuisance parameters r for the D0 data [1] (left) and the CDF ones [2] (right) with the uncertainties due to data fluctuation (inner bars) and the total uncertainties including the ones due to PDFs (outer bars) versus the nuisance parameter number n. The normalization nuisance parameters correspond to $n = 6$ and 17 for D0 and CDF, respectively.

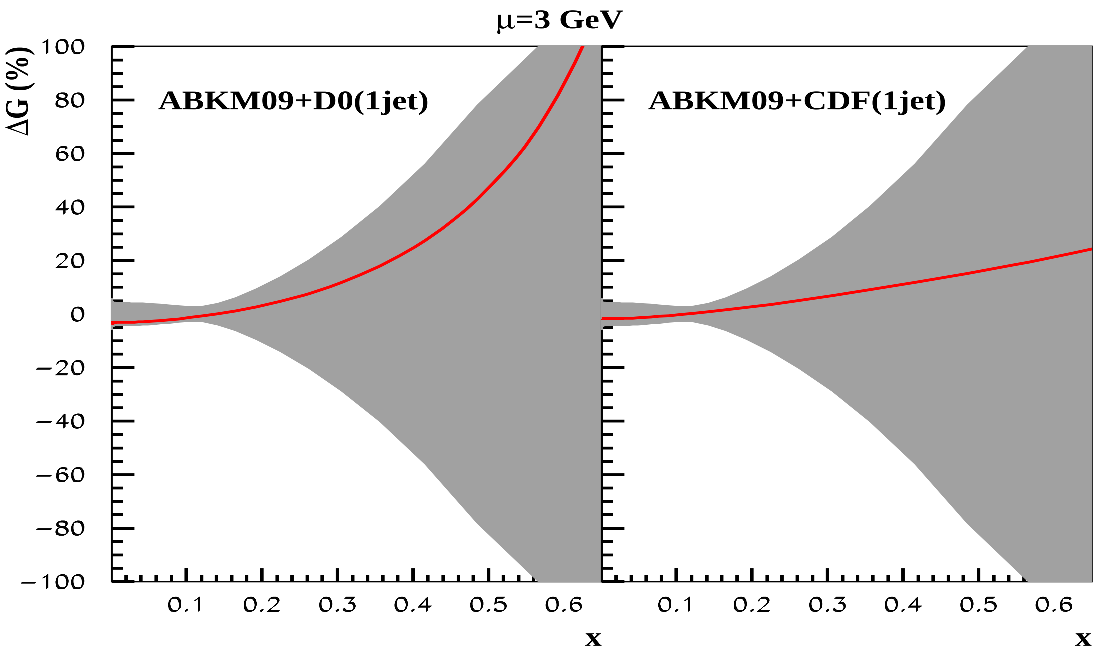

Figure 4. Relative variation of the gluon distribution due to dropping the normalization error in different Tevatron jet data sets (lines) compared to the uncertainties in the ABKM09 gluon distributions (shaded area) at the factorization scale $\mu = 3$ GeV versus x (left panel: D0; right panel: CDF).

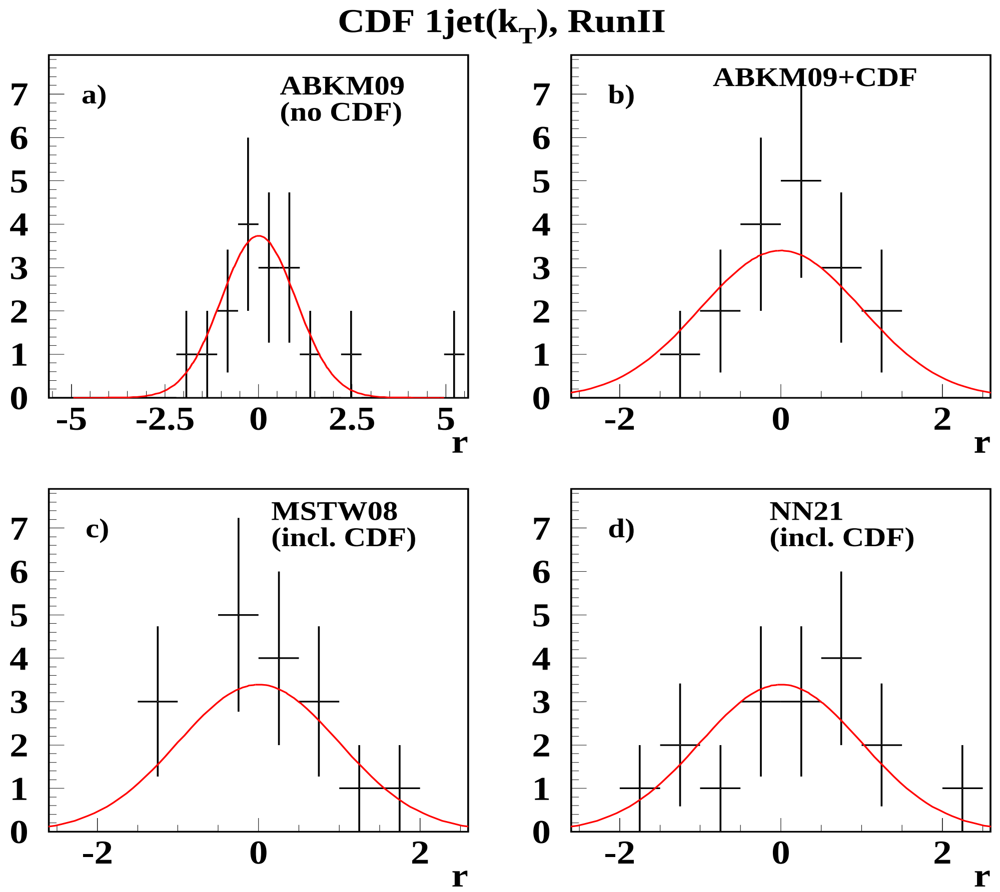

Figure 5. The same as Fig. 1 for the CDF data on inclusive jet production [2].

in the transverse momentum of jet. The correlated systematic uncertainties in the CDF jet data stem from 17 sources including the overall normalization. The distribution of the corresponding nuisance parameters calculated with the NNLO ABKM09 PDFs is displayed in Fig. 5. In general, it is in agreement with the normal Gaussian one with the only essential excess observed for the normalization nuisance parameter, which reaches the value of $r_{norm} = 5.4$. This is bigger than the D0 normalization nuisance parameter. However, due to bigger uncertainties in the CDF data the errors in this parameter are also bigger as compared to the D0 case. The variance of the CDF normalization nuisance parameter is $V(r_{norm}) = 0.93$ (to be compared to 0.85 for D0) and the uncertainty due to the PDFs is $\Delta^{PDF}(r_{norm}) = 1.43$ (to be compared to 0.95 for D0). The CDF error due PDFs is evidently enhanced due to the particular trend of the data with respect to the predictions based on the ABKM09 fit. In the D0 case the offset of data does not depend on the jet energy, while the CDF jet energy dependence is systematically tilted as compared to the predictions, cf. Figs. 1,2 in Ref. [5]. With account of these errors the CDF normalization nuisance parameter is consistent with 0 within 3 standard deviations. Other nuisance parameters for the CDF data are consistent with zero within uncertainties, cf. Fig. 3, therefore in total the

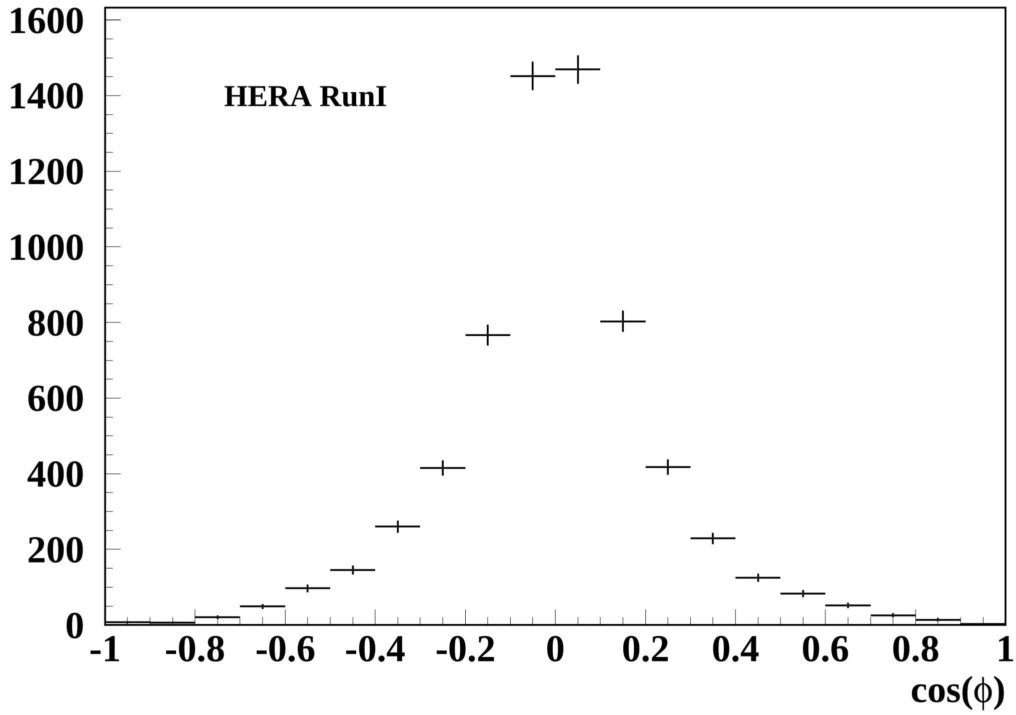

Figure 6. Distribution of the cosine of the angle between the systematic error vectors $\phi_{kk'}$, cf. Eq. (3.1), for the HERA data on the inclusive DIS structure functions [14]. Only the angles with $k > k'$ are histogrammed.

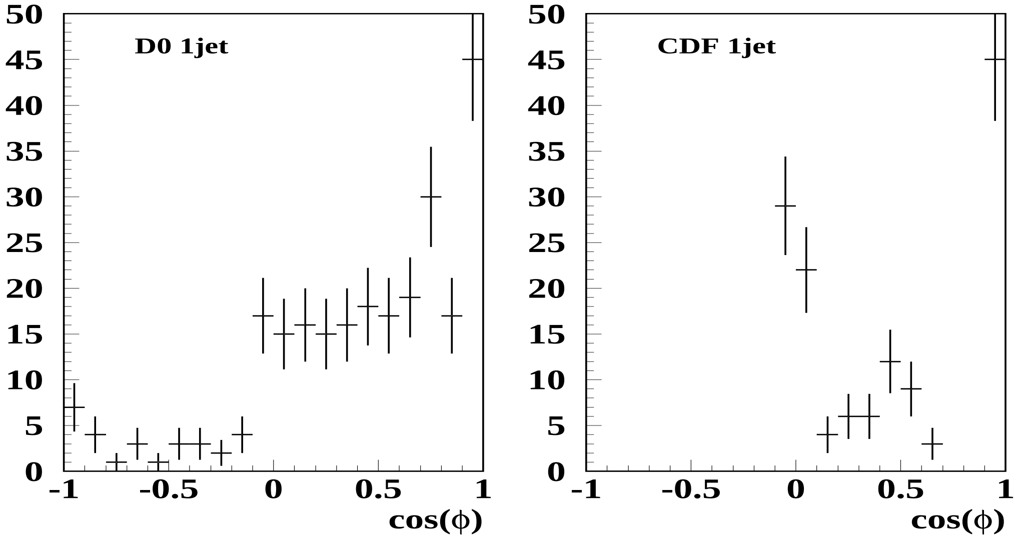

Figure 7. The same as in Fig. 6 for the D0 [1] (left) and CDF [2] (right) data on the inclusive jet production cross section.

statistical significance of the excess in the normalization nuisance parameter is marginal, as well as for the D0 jet data. In line with this observation the distribution of the CDF nuisance parameters in the variant of the ABKM09 fit, which includes the CDF data, is in agreement with the normal Gaussian one, cf. Fig. 5. Similarly to the D0 case the change in χ^2 due to dropping the CDF normalization uncertainty in the variant of the ABKM09 fit, which includes the CDF data, is marginal, i.e. less than 1 for 76 data points. The change in the PDFs due to dropping the normalization uncertainty is also marginal, cf. Fig. 4.

These observations do not support the conclusion of Ref. [6] about the crucial significance of the normalization uncertainty in the accommodation of Tevatron jet into the ABKM09 fit. As an explanation of this disagreement we point out that in the analysis of Ref. [6] the errors in nuisance parameters due to the PDF uncertainties and the experimental errors in the data are not considered. This leads to an overestimation of the statistical significance in the nuisance parameter excesses in the analysis of Ref. [6]. Another concern about the conclusion of Ref. [6] is related to the relevance of a rigorous statistical treatment of the systematic uncertainties in the Tevatron jet data. Commonly, the different sources of systematics are assumed to be independent, cf. Eqs. (2.1,2.5). This also was assumed in the present study and in Ref. [6]. We have checked this hypothesis for the Tevatron jet data plotting the cosine of angles between the systematic uncertainty vectors $s_{k,i}$, which are defined as

$$\cos(\phi_{kk'}) = \frac{\sum_i s_{k,i} s_{k',i}}{\sqrt{\sum_i s_{k,i}^2 \sum_i s_{k',i}^2}}. \tag{3.1}$$

Naively, the distribution of $\cos(\phi_{kk'})$ should peak at $\cos(\phi) = 0$ and be symmetric with respect to this peak for the case of independent sources of the systematic uncertainties. In particular, such a picture is observed for the HERA data on the inclusive deep-inelastic-scattering (DIS) structure functions, cf. Fig. 6. However, this is not the case for the D0 and CDF data, cf. Fig. 7. For both CDF and D0 data the distributions peak at $\cos(\phi) = 1$ and are quite asymmetric, particularly in the case of CDF. This indicates a strong collinearity of many systematic uncertainty vectors. In case these systematic errors really stem from one of a few sources only, the PDF fits based on the Tevatron jet data should be revisited. Note that the vectors $s_{k,i}$ corresponding to the normalization uncertainty are collinear to many other systematic error vectors for these data. Evidently, this also explains the big error in the normalization parameter since the corresponding nuisance parameters are mixed due to this collinearity.

The distributions of the D0 and CDF nuisance parameters for the variants of NNLO ABM11 fit [7], which include the Tevatron jet data in a similar way to Ref. [5], are in agreement with ones for the ABKM09 fit, cf. Fig. 8. In turn, both ABKM09 and ABM11 nuisance parameter distributions are similar to the ones obtained with the MSTW08 [15] and NN21 [16] PDFs, which are also tuned to the Tevatron jet data, cf. Figs. 1 and 5. The remaining differences can be explained by the specific data selection in the fits and the fitted model peculiarities, like e.g. heavy-quark treatment, high-twist contributions, and others, cf. Ref. [7]. It can also appear due to different statistical estimators used in the PDF fit. In particular, the ABKM09 and ABM11 fits are based on the covariance

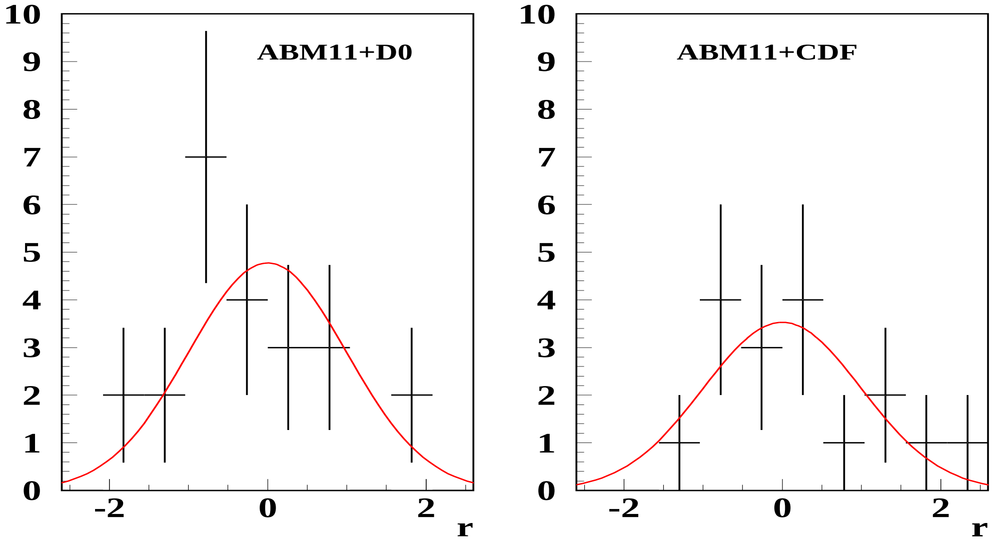

Figure 8. The same as Fig. 1 for the D0 data [1] (left) and the CDF data [2] (right) and the variants of NNLO ABM11 fit [7] including the D0 and CDF jet data, respectively.

matrix estimator of Eq. (2.2), while in the MSTW08 fit the one of Eq. (2.4) is employed. As we have pointed out in Section 2, in the first case the systematic errors are considered as multiplicative and in the second case as additive. Note, that an additive treatment of the errors in the cross sections leads to a statistical bias in the fitted parameters (cf. Refs. [17–19] and references therein for a discussion). Therefore it may have an impact on the nuisance parameter values which depend on the fitted PDF parameters as well. In the NNPDF fit [16, 20] the normalization errors are treated in a special way, which allows to minimize the bias. However, the covariance matrix of Eq. (2.6) is still used to take into account other correlated systematic errors (cf. Eq. (1) in Ref. [20]). Since for the Tevatron jet data the latter dominate, the bias appears also in the NNPDF fit.

4 Conclusion

We have analyzed a tension between the D0 and CDF inclusive jet data and the perturbative QCD calculations, which are based on the NNLO ABKM09 and ABM11 PDFs with account of the NLO and NNLO threshold resummation corrections to the parton cross sections. The nuisance parameters employed to quantify the tension are calculated for each source of systematic uncertainty in the data minimizing the χ^2-estimator, which allows for shifts of the data by the value of systematic error scaled with the corresponding nuisance parameter. For some sources, in particular for the normalization uncertainty, the nuisance parameter values are relatively big. However, the analysis of their uncertainties due to the data fluctuations and the PDF errors shows that the nuisance parameter errors are as well substantial. In particular, this happens due to many systematic uncertainty vectors including the normalization ones being collinear and, as a result, the corresponding nuisance parameters are mixed. In view of those big uncertainties the statistical significance of the excesses in the normalization nuisance parameters is marginal. Furthermore, this conclusion is explicitly checked by considering the variants of ABKM09 fit, which include the Tevatron jet data without any normalization uncertainty taken into account. The results of these fits are quite similar to the ones including the normalization uncertainties. These observations do not support the conclusion about the crucial role played by the normalization uncertainty in the accommodation of the Tevatron jet data into the ABKM09 fit mentioned in [6] disregarding the nuisance parameter errors. Besides, the statistical analysis of Ref. [6] lacks rigor since the nuisance parameters are derived for the statistically biased estimator, while the ABKM09 fit is based on the estimator, which is asymptotically unbiased [18]. At the same time, despite a serious statistical issue does not appear in the variants of the ABKM09 fit including the Tevatron jet data [5], the latter are finally not yet used in the ABM11 fit [7] in view of yet lacking complete NNLO corrections, which may have an impact both on determination of the strong coupling constant and on the parton distribution functions.

Acknowledgments

This work has been supported in part by DFG Sonderforschungsbereich Transregio 9, Computergestützte Theoretische Teilchenphysik, and EU Network LHCPHENOnet PITN-GA-2010-264564.

References

[1] V. M. Abazov *et al.* [D0 Collaboration], "Measurement of the inclusive jet cross-section in $p\bar{p}$ collisions at $s^{(1/2)}$ =1.96-TeV," Phys. Rev. Lett. **101** (2008) 062001 [arXiv:0802.2400 [hep-ex]].

[2] A. Abulencia *et al.* [CDF Collaboration], "Measurement of the Inclusive Jet Cross Section using the $\boldsymbol{k_{\mathrm{T}}}$ algorithm in $\boldsymbol{p\overline{p}}$ Collisions at$\sqrt{\boldsymbol{s}}$ = 1.96 TeV with the CDF II Detector," Phys. Rev. D **75** (2007) 092006 [Erratum-ibid. D **75** (2007) 119901] [hep-ex/0701051].

[3] D. Stump, J. Pumplin, R. Brock, D. Casey, J. Huston, J. Kalk, H. L. Lai and W. K. Tung, "Uncertainties of predictions from parton distribution functions. 1. The Lagrange multiplier method," Phys. Rev. D **65** (2001) 014012 [hep-ph/0101051].

[4] S. Alekhin, J. Blümlein, S. Klein and S. Moch, "The 3, 4, and 5-flavor NNLO Parton from Deep-Inelastic-Scattering Data and at Hadron Colliders," Phys. Rev. D **81** (2010) 014032 [arXiv:0908.2766 [hep-ph]].

[5] S. Alekhin, J. Blümlein and S. -O. Moch, "Parton Distributions and Tevatron Jet Data," arXiv:1105.5349 [hep-ph].

[6] R. S. Thorne and G. Watt, "PDF dependence of Higgs cross sections at the Tevatron and LHC: Response to recent criticism," JHEP **1108** (2011) 100 [arXiv:1106.5789 [hep-ph]].

[7] S. Alekhin, J. Blümlein and S. Moch, "Parton Distribution Functions and Benchmark Cross Sections at NNLO," Phys. Rev. D **86** (2012) 054009 [arXiv:1202.2281 [hep-ph]].

[8] S. Alekhin, J. Blümlein, P. Jimenez-Delgado, S. Moch and E. Reya, "NNLO Benchmarks for Gauge and Higgs Boson Production at TeV Hadron Colliders," Phys. Lett. B **697** (2011) 127 [arXiv:1011.6259 [hep-ph]].

[9] M. Wobisch, "Recent QCD results from the Tevatron," Nucl. Phys. Proc. Suppl. **222-224** (2012) 204 [arXiv:1202.0205 [hep-ex]].

[10] T. Kluge, K. Rabbertz and M. Wobisch, "FastNLO: Fast pQCD calculations for PDF fits," hep-ph/0609285.

[11] Z. Nagy, "Three jet cross-sections in hadron hadron collisions at next-to-leading order," Phys. Rev. Lett. **88** (2002) 122003 [hep-ph/0110315].

[12] Z. Nagy, "Next-to-leading order calculation of three jet observables in hadron hadron collision," Phys. Rev. D **68** (2003) 094002 [hep-ph/0307268].

[13] N. Kidonakis and J. F. Owens, "Effects of higher order threshold corrections in high E(T) jet production," Phys. Rev. D **63** (2001) 054019 [hep-ph/0007268].

[14] F. D. Aaron *et al.* [H1 and ZEUS Collaboration], "Combined Measurement and QCD Analysis of the Inclusive $e^{\pm}$ p Scattering Cross Sections at HERA," JHEP **1001** (2010) 109 [arXiv:0911.0884 [hep-ex]].

[15] R. S. Thorne, A. D. Martin, W. J. Stirling and G. Watt, "Parton Distributions and QCD at LHCb," arXiv:0808.1847 [hep-ph].

[16] R. D. Ball *et al.* [NNPDF Collaboration], "Unbiased global determination of parton distributions and their uncertainties at NNLO and at LO," Nucl. Phys. B **855** (2012) 153 [arXiv:1107.2652 [hep-ph]].

[17] G. D'Agostini, "On the use of the covariance matrix to fit correlated data," Nucl. Instrum. Meth. A **346** (1994) 306.

[18] S. I. Alekhin, "Statistical properties of the estimator using covariance matrix," hep-ex/0005042.

[19] P. Jimenez-Delgado, "The role of the input scale in parton distribution analyses," Phys. Lett. B **714** (2012) 301 [arXiv:1206.4262 [hep-ph]].

[20] R. D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, J. Rojo and M. Ubiali, Nucl. Phys. B **838** (2010) 136 [arXiv:1002.4407 [hep-ph]].