PDFSense: A tool for exploring the sensitivity of hadronic experiments to nucleon structure

Bo Ting Wang,[1, *] Sean Doyle,[2, 3, †] Jun Gao,[2, 3] T. J. Hobbs,[1, ‡]
Tie-Jiun Hou,[4, §] Pavel Nadolsky,[3, ¶] and Fredrick I. Olness[3, **]

[1] *Department of Physics, Southern Methodist University,*
Dallas, TX 75275-0181, U.S.A.
[2] *School of Physics and Astronomy, INPAC,*
Shanghai Key Laboratory for Particle Physics and Cosmology,
Shanghai Jiao-Tong University, Shanghai 200240, China[††]
[3] *Department of Physics, Southern Methodist University,*
Dallas, TX 75275-0181, U.S.A.
[4] *School of Physics Science and Technology, Xinjiang University,*
Urumqi, Xinjiang 830046 China

We demonstrate a collection of tools and metrics for quantitatively studying the sensitivity of hadronic measurements to the underlying Parton Distribution Functions (PDFs).

Replace PACS with PhySH:`https://www.aps.org/publications/apsnews/201602/classification.cfm`

PACS numbers: 12.15.Ji, 12.38 Cy, 13.85.Qk
Keywords: parton distribution functions; large hadron collider; Higgs boson

Contents

I. INTRODUCTION

The determination of the collinear parton distribution functions (PDFs) of the nucleon has increasingly become a precision discipline in recent years with the advent of high statistics programs at both fixed-target experiments and colliders. Parton distribution functions (PDFs) are crucial for understanding the behavior of hadron collisions and then exploring the Standard Model (SM). PDFs describe the structure of hadrons, which affect the configurations of the final particles in the collisions. Therefore, the magnitudes of physical observables in hadron collisions strongly depend on PDFs. Currently, The Large Hadron Collider (LHC) produces a lot of experimental data. Owning to the fact that uncertainties in measurements constantly decrease, reducing the PDF uncertainties of physical observables and using the higher order PDFs will make it easier to find the inconsistency between SM and the data sets collected by the LHC and then discover new physics. Incorporating more (LHC) new data sets in the global fits of PDFs is a naive way to generate better PDF sets with small uncertainties.

However, incorporating more experimental data points will substantially increase the time for fitting PDF sets, especially when we fit higher order PDF sets. From here we know that how to select data sets in global fits will become extremely important in the near future. It is essential to know which data sets will effective constrain the higher order PDFs for the global fits in the limited time of computation. In addition, because physical predictions are sensitive to respective flavors and regions of $\{x, \mu\}$ in PDFs, we need to narrow down uncertainties of the specific regions of $\{x, \mu\}$ (in the PDFs). Where partonic x are momentum fractions and μ are QCD factorization scales. For example, if PDF values for the

[*] Electronic address: botingw@smu.edu
[†] Electronic address: seand@smu.edu
[‡] Electronic address: tjhobbs@smu.edu
[§] Electronic address: tiejiun.hou@foxmail.com
[¶] Electronic address: nadolsky@smu.edu
[**] Electronic address: olness@smu.edu
[††] Electronic address: jung49@sjtu.edu.cn

leading $\{x, \mu\}$ ranges and flavors that characterize kinematical quantities for Higgs production processes (e.g. at $\mu = 125$ GeV) are tightly constrained, the theoretical predictions for these processes are reliable (precise).

Using correlation between PDF uncertainties in two observables have been proposed to study constraints on PDFs and constraints on observables imposed by PDFs [1][2][3].

The approach can help us to find the $\{x, \mu\}$ ranges of PDFs affecting physical observables such as total cross section [3]. It is yet be established that how to know the ranges specifying PDFs constrained by experimental data sets.

Thus, establishing a better understanding of the relationships between the strength of constraints on PDF and experimental data sets will be a significant and beneficial contribution to particle physics.

We have developed and tested a systematic method to study the constraints on PDFs imposed by the experimental data sets. We will use established statistical observables to quantify the strength of these constraints.

After that, we introduce a new statistical package *PDFSense* to visualize the regions of partonic momentum fractions x and QCD factorization scales μ where the experiments impose strong constraints on the PDFs. Recent experimental data will be considered in the analysis in order to provide better constraints to various ranges of PDFs.

The remainder of the article proceeds as follows. Essential details regarding the PDFs and their standard determination via QCD global analyses are summarized in II. The basic statistical ingredients of our analysis are presented in Sec. III, along with an introduction to our analysis package *PDFSense.* To highlight the utility of the resulting framework, we perform an assessment of the potential impact of hypothetical LHeC pseudodata in Sec. IV; in the conclusion contained in Sec.V we emphasize a number of physics insights that may be gained with *PDFSense*, while a number of statistical details and supporting figures and tables are reserved for Apps. A and B, respectively.

II. PDF PRELIMINARIES

While various theoretical methods exist to compute nucleon PDFs in terms of models, their unambiguous evaluation in terms of QCD is not yet possible due to the fact that the PDFs can in general receive substantial nonperturbative contributions at infrared momenta. For this reason, precise PDF determination has proceeded mainly through the technique of QCD global analysis — a method enabled by the QCD factorization theorem.

In this approach, a highly flexible parametric form is ascribed for the various flavors in a given analysis at a relatively low scale Q_0^2. For example, in the commonly used Hessian Method one might take the input PDF for a given quark flavor f according to be an n-parameter form

$$f(x, Q_0^2) = a_{f0}\, x^{a_{f1}} (1-x)^{a_{f2}}\, F(a_{f3}, \dots, a_{fn}) \ , \tag{1}$$

in which $F(a_{f3}, \dots, a_{fn})$ is typically a suitable polynomial function, *e.g.*, a Chebyshev polynomial. In this circumstance, a best fit is then found for the parameters $a_{f0}, \dots, a_{fn}$ by minimizing a χ^2 function with respect to the world's data for which physical observables can be computed in terms of the PDFs by the factorization theorem; the resulting uncertainties are determined by requiring that the χ^2 function remain appropriately small under minor perturbations of the fit parameters $\{a_{fn}\}$ subject to the condition

$$\chi^2 \,<\, \chi^2_{min} + \chi^2_{torelance} \ . \tag{2}$$

From this information it is then possible to convert from a parametric basis $\{a_{fn}\}$ basis to a basis of eigenvectors $\{r_i\}$ in terms of which a set of PDF error replicas may be generated that encapsulate the constraint to the proton PDFs from the world's data.

Due to the presence of correlated systematic errors in many experimental data sets, it should be noted that the most appropriate χ^2 function is in general

$$\chi^2 = \sum_i r^2_{i,\,shift} + \sum_{k=1}^{K} \rho_k^2 \tag{3}$$

$$r_{i,\,shift} = \frac{1}{\sigma_i}\left(T_i - D_{i,\,shift}\right) \ , \tag{4}$$

where the sum over i accounts for the individual data points in the set of global data, and $r_{i,\,shift}$, T_i, and σ_i are the *residual*, theoretical prediction evaluated in terms of PDFs, and total uncorrelated uncertainty for the i^{th} data point, respectively. On the other hand, $D_{i,\,shift}$ represent the central values of the experimental data, shifted to incorporate the effect of systematic correlations as described in App. B of Ref. [4], with ρ_k in Eq. (3) the random shift parameters associated with K sources of correlated systematic error. We henceforth refer to the residuals corresponding to data shifted with respect to correlated errors simply as r_i for compactness.

In consequence of these defintions, both the residual r_i and χ^2 may be thought of as broadly characterizing the quality of a given fit to a large data set, with the scenario $\chi^2 \gg N_{data}$ implying a poor fit, whereas $\chi^2 \leq N_{data}$ is consistent with a strong description of the data. At the same time, fits that too sharply limit $\chi^2 \ll N_{data}$ correspond to an overfit of the data.

...select the experimental data sets and theoretical model in global fits ...write down parametrization functions of all flavors ...determine the best-fitted a_0, a_1, a_2...... of the parametrization functions by minimizing χ^2 determine uncertainties of the parametrization functions by requiring $\chi^2 < \chi^2_{min} + \chi^2_{torelance}$

III. PDFSENSE

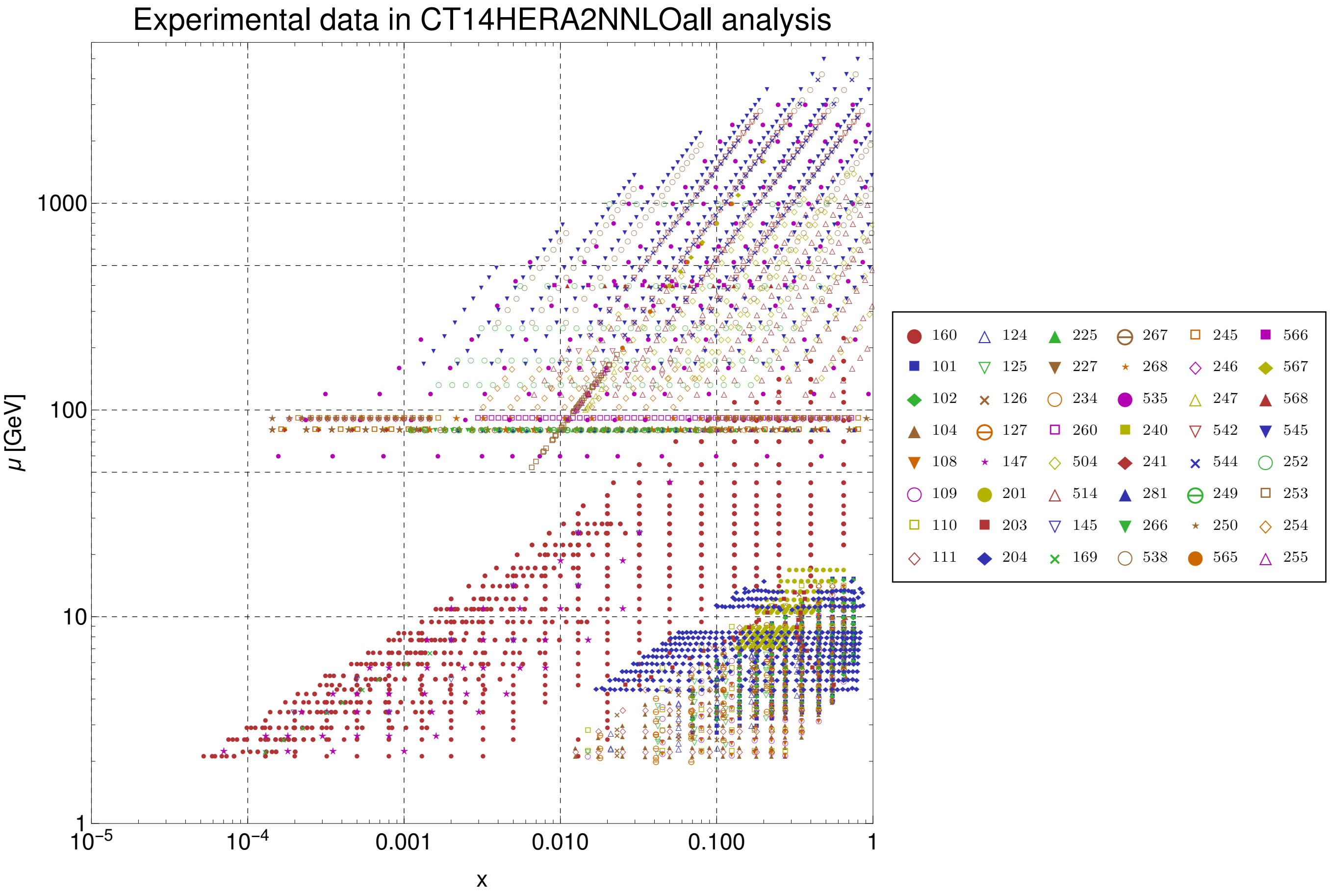


FIG. 1: A graphical representation in (x, μ) space of the full dataset treated in the present analysis, essentially corresponding to an expansion of the data fitted in the most recent CT14 analysis, which included fixed-target measurements from Run II of HERA as described in Ref. [5] (CT14HERA2); we take this data set to be the default in the present analysis and illustrate the potential of *PDFSense* to illustrate the differential impact of the separately labeled data sets represented here in PDF determinations, with the CT14HERA2 PDF set as an illustrative example. We note that the points are labeled according to experimental ID number; a detailed translation key to individual experiments is given in Tables I–III

We have developed and tested a systematic method to study the constraints on PDFs imposed by experimental data sets. We use established statistical observables to quantify the strength of these constraints. After that, We introduce a statistical technique to visualize the regions of partonic momentum fractions x and QCD factorization scales μ where the experiments impose strong constraints on the PDFs. Recent experimental data is considered in the analysis in order to provide better constraints to various ranges of PDFs.

To test the effectiveness of the proposed method, we study constraints on CT14HERA2 parton distributions [5] from various data sets. We include various types of experimental data sets in the analysis, including DIS processes, $Z \to l^+l^-, d\sigma/dy(l)$, $W \to l\nu$, and jet productions $(p_1 p_2 \to jjX)$.

For data sets of interest, we can demonstrate and identify values of correlation/sensitivity data by different colors on the $x - \mu$ plane $(2D - x - \mu$ figure), such as Figs. 6, which help us to rapidly estimate the distribution of the strength of constraints on the $x - \mu$ plane. We can also know the number of data points constraining PDFs by the histograms of the statistical quantities.

A. Statistical definitions

Among various quantities that characterize the impact of experimental data upon the PDFs, the correlations, which we define according to

$$c_f^i(x, \mu) \equiv Corr[f(x, \mu), r_i] \ , \tag{5}$$

encodes the quantitative relation between the PDFs $f(x, \mu)$ and residuals r_i, and can determine whether there exist predictive relationships between PDFs and goodness of fit to data points. The formal definition and origin of the correlation of Eq. (5) is given in detail in App. A.

Correlation illustrates the strength of the predictive relation between any two observables X and Y. We can use values of one observable to predict values of another observable very well when their correlation is close to ±1. correlations of Hessian uncertainties [1] have been used to see the simultaneous constraint on observables X and Y, and to get constraints on PDFs [1–3]. First, via measuring one physical observable, we are able to predict the value of another observable precisely. In addition, strong correlations are highly likely to show the signs of some physical relations, such as causation, between the two observables.

While the correlation $c^i_f(x,\mu)$ can provide insight into the relationship between the x, μ dependence of a set of PDFs and associated PDF uncertainty, it alone does not fully quantify the extent to which individual data points constrain fit uncertainties; similarly, the correlation alone does not encode by itself the potential impact of separate or new measurements in improving PDF determinations in terms of uncertainty reduction. For this purpose, we define the *sensitivity* of the i^{th} data point to the PDF of flavor f,

$$s^i_f \;\equiv\; \frac{\delta^{(\mathrm{PDF})} r_i}{\sqrt{\frac{1}{N}\sum_{i=1}^{N} r_i^2}}\, c^i_f(x,\mu)\ , \tag{6}$$

where the sum is up to N, which we take to be the total number of data points in a given experimental data set. In Eq. (6), the quantity $\delta^{(\mathrm{PDF})} r_i$ represents the variation of the residuals accross the set of Hessian error PDFs, which we normalize the quadrature-summed residual for a given experiment. This definition has the benefit of encoding not only the correlated relationship between the PDF at a given (x,μ) with the residual, but also the comparative size of the experimental uncertainty with respect to state-of-the-art PDF uncertainties. In consequence, for example, new experimental data with reported uncertainties much tighter than present PDF errors would register as high sensitivity points by the definition in Eq. (6). Fig. 5 and the associated discussion in App. A describe this point theoretically, and we illustrate this behavior with a number of practical examples below.

There are several ways to evaluate uncertainties on PDFs such as the Hessian method [1], the Monte Carlo method [6][7], and the Lagrange Multiplier [8]. Our PDF set input is CT14HERA2, which uses the Hessian method to estimate uncertainties information. This idea is based on the quadratic assumption. According to the quadratic assumption, we will get an elliptical shape of PDF parameter space around the best fit parameters $\vec{a_0}$ for a given tolerance parameter $\chi^2_{tolerance}$ satisfying $\chi^2(\vec{a}) < \chi^2(\vec{a_0}) + \chi^2_{torelance}$. If errors of an observable X along the ± directions of i-th dimension of the ellipse are X^+_i and X^-_i, the uncertainty of X based on the variation of parameter at i-th dimension could be approximated by $(X^+_i - X^-_i)/2$. According to the principle of error propagations, the X uncertainty via PDF parameter space is $\Delta X = \frac{1}{2}\sqrt{\sum\limits_i (X^+_i - X^-_i)^2}$.

Move these definitions to an appendix???

Our idea of studying PDF constraints from data sets uses the correlation between PDF Hessian sets and residual Hessian sets, where the Hessian correlation of two observables is defined as $cos\phi = \sum\limits_i (X^+_i - X^-_i)(Y^+_i - Y^-_i)/4\Delta X \Delta Y$. The correlation of any two observables X and Y could be used to see the simultaneous constraint of X and Y [1]. The ellipse of simultaneous constraint could be described by Lissajous figure

$$X = X_0 + \Delta X\, sin(\theta + \phi)$$

$$Y = Y_0 + \Delta Y\, sin(\theta + \phi)$$

where $0 < \theta < 2\pi$ traces the shape of the ellipse, and whether the shape is needle-like or circle-like is controlled by ϕ. If $|cos\phi| \simeq 1$, the shape is needle-like, which strongly constrains Y for a given δX (see Fig. **??**). Thus, the correlations $Corr(f_a(x,\mu), r_i)$ of PDFs $f_a(x,\mu)$ and the residuals r_i can determine the strength of constraints on PDFs imposed by r_i in experimental data points.

strength of constraints on PDFs imposed by r_i $(SOC(PDF))$

Although we can know the predictivities between PDFs and measurements through $Corr(f_a(x,\mu), r_i)$, $Corr(f_a(x,\mu), r_i)$ could not specify the strength of constraints on PDFs imposed by r_i $(SOC(PDF))$. For instance, the measurements with large uncertainties cannot effectively constrain $f_a(x,\mu)$ no matter how large $Corr(f_a(x,\mu), r_i)$ is, since r_i is not sensitive to the variation of $f_a(x,\mu)$. Therefore, we want to find a more representative statistical quantity for $SOC(PDF)$. To study $SOC(PDF)$ between PDFs and data sets, we study the variation of χ^2 and r_i associated with the fluctuation in $f_a(x,\mu)$. Fig. 6 shows r_i in data points depending on the variation in $f_a(x,\mu)$ error sets. We find that despite the fact that the correlation between r_i in two data points (red circles and blue circles) and $f_a(x,\mu)$ are the same, the fluctuation for $f_a(x,\mu)$ imposes different levels of impact to r_i. The r_i, represented by red circles, are more sensitive to $f_a(x,\mu)$, which indicates that when we constrain χ^2 for getting the new fitted $f_a(x,\mu)$ error sets, the data point represented by the red circles will more dramatically narrow down the range of the new $f_a(x,\mu)$ error sets so that r_i for error sets become smaller. Here we find the δr_i, which evaluates the fraction of theoretical and experimental uncertainties, indicating whether the theoretical uncertainties are apt to be constrained after the fitting. For the above reasons, we advise using $\delta r_i \times Corr(f_a(x,\mu), r_i)$ to quantify the sensitivity $(Sen(f_a(x,\mu), r_i))$ for r_i to $f_a(x,\mu)$, and using the sensitivity to estimate $SOC(PDF)$ for data point i.

B. Correlation and Sensitivity analysis

Having outlined the main statistical quantities of interest in the preceding subsection, we turn now to their implementation in the correlation and sensitivity analysis package *PDFSense*. As described in the previous sections, our aim is to quickly evaluate the impact of various hadronic data sets upon the present knowledge of the PDFs in a fashion that does not require a full QCD analysis of the type described in Sec. II. As a demonstration of how *PDFSense* achieves this, we treat the data set shown in Fig. 1 of the most recent NNLO CT14 analysis of the global data set augmented by the HERA Run II fitted in Ref. [5] (which we refer to as CT14HERA2), explicitly showing how various constituent data sets constrain PDFs in a full space of (x, μ). We have already noted the magnitude of this data set, which is decomposed into separate experiments in Fig. 1.

In principle, we can use the correlation & sensitivity mentioned above to quantify $SOC(PDF)$ for any points on the $x - \mu$ plane and data point i. Therefore, we can identify which regions in the $x - \mu$ plane have strong $SOC(PDF)$. Our objective is to characterize the strongly constrained ranges (Strong $SOC(PDF)$ Regions) imposed by the given data sets.

Given its centrality in many PDF analyses, we concentrate our demonstration in the present section upon the gluon distribution, $g(x, \mu)$. In fact, the gluon PDF remains dominated by substantial uncertainties at both $x \sim 0$ and in the elastic limit $x \to 1$, a fact which has driven an intense focus upon, *e.g.*, top quark and jet production processes at colliders, which themselves are typically measured at large center-of-mass energies $\sqrt{s}$ where the dependence upon the low x behavior of the gluon distribution is especially relevant.

We first consider plots of the correlations for the gluon distribution, both as a histogram and as fully represented in a space of (x, μ). A striking feature of the correlation plot is the large magnitude found for the inclusive jet production measurements (inverted triangles) for which large gluon-residual correlations $c_g(x, \mu) > 0.85$ are found, especially at the highest values of x and μ represented in Fig. 2.

We can also consider the analogous plot for the sensitivity of the data to the gluon distribution, which we plot in Fig. 3. Here, an evident feature is the comparitive reduction of the "temperature" of the plot due to the inclusion of $\delta^{(\mathrm{PDF})} r_i$ in the defintion of the sensitivity. In reasonable agreement with expectation, we also observe the most heightened sensitivities for the points at lowest x — especially for the combined HERA data set — in accord with the fact that PDF errors for the gluon distribution are in general least controlled at these x ranges.

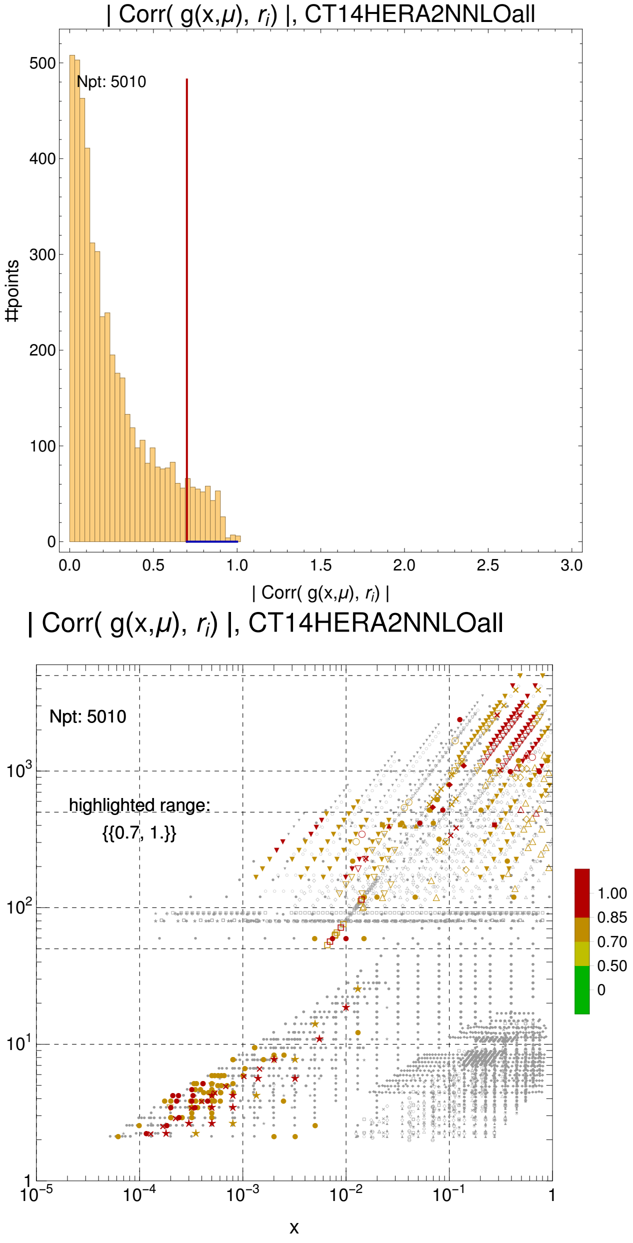


FIG. 2: Two representations of the correlation $c_g^i(x, \mu)$ of the gluon PDF $g(x, \mu)$ with the pointwise residual r_i of the CT14HERA2 analysis; in the upper panel we plot a histogram showing the distribution of correlations over 5010 points. In the lower panel we show the (x, μ) map of these correlations as produced by *PDFSense* for the full data set.

IV. CASE STUDY: LHEC

The need to disentagle, *e.g.*, the distribution of the gluon at low x and the detailed behavior of the light quark distributions at large x (which are inherently nonperturbative) has stimulated interest in a future high luminosity electron collider at the LHC, the Large Hadron-electron Collider (LHeC) [9], which could in principle afford access to new domains in x and Q^2 beyond that made available by recent fixed-target or collider experiments.

Such a machine would effectively represent an analogue to HERA, albeit at substantially higher energy and luminosity, a fact which would putatively enable it to probe various elusive phenomena, including the onset of

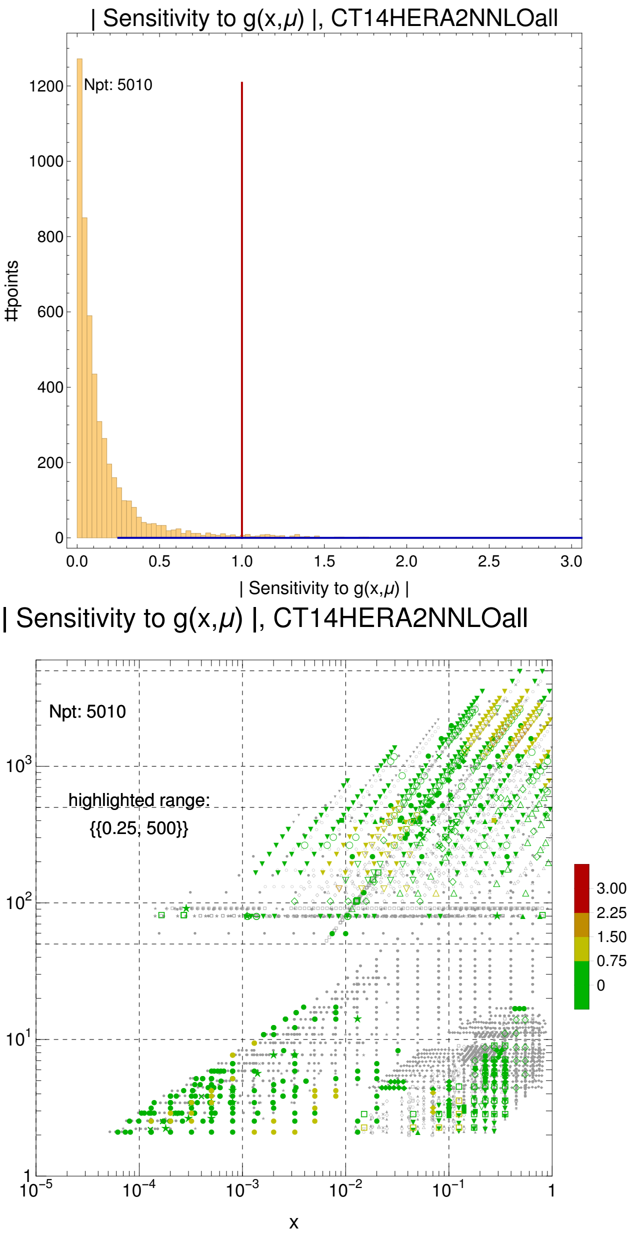


FIG. 3: Like Fig. 2, but for the gluon sensitivity $s_g^i(x,\mu)$ as defined in Eq. (6).

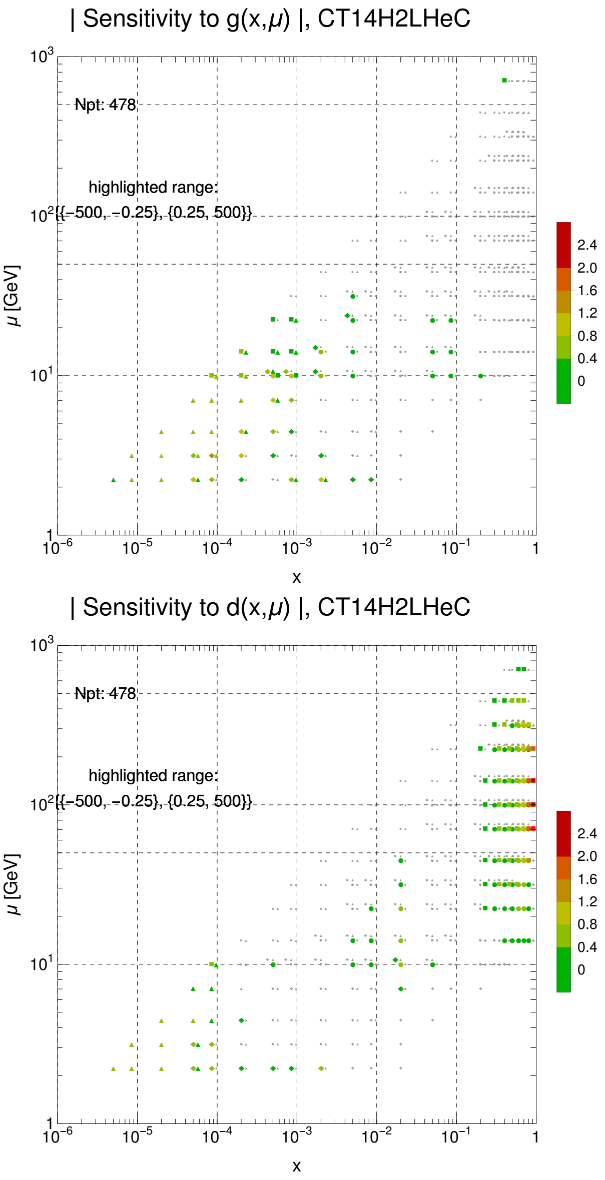


FIG. 4: Select sensitivity plots for the LHeC pseudodata, in this case showing $s_g^i(x,\mu)$ in the upper panel and $s_d^i(x,\mu)$ in the lower. The pseudodata are broken among four separate reaction channels — CC e^-p (circles) and e^+p (squares) as well as NC e^-p (diamonds) and e^+p (triangles).

gluonic saturation at low x, diffractive effects, and, potentially, nonperturbative phenomenology at high x and heavy quark production and hadronization. The reach of this machine would in principle be especially enhanced by the increased luminosity, which should afford a strong reduction in the size of experimental uncertainties.

To explore this possibility, we implement the pseudodata [10] generated by Klein and Radescu into *PDFSense* to gauge the potential sensitivity of this information to the behavior and uncertainty of the proton PDFs as again represented by the most recent CT14HERA2 set. For the sake of illustration, we concentrate on two banner results — the sensitivity of the LHeC pseudodata to the gluon distribution (shown in the top panel of Fig. 4) as well as for the down quark distribution, which we plot in the lower panel, also of Fig. 4. Our plots explicitly delineate the separate experimental channels for the pseudodata (the unpolarized reduced cross sections $\sigma_r^{e^\pm}$ for both neutral current and charged current interactions, as accounted in the caption of Fig. 4).

V. CONCLUSIONS

The preceding article presented a new analysis tool for the rapid exploration of the impact of both existing and potential data on the PDF determinations.

VI. ACKNOWLEDGEMENTS

This work was supported in part by the U.S. Department of Energy under Grant No. DE-SC0010129 and by the National Natural Science Foundation of China under the Grant No. 11465018. The work of J.G. is sponsored by Shanghai Pujiang Program.

[1] J. Pumplin, D. Stump, R. Brock, D. Casey, J. Huston, J. Kalk, H. L. Lai, and W. K. Tung, Phys. Rev. **D65**, 014013 (2001), hep-ph/0101032. I, III A, III A, III A, III A, A, A, A

[2] P. M. Nadolsky and Z. Sullivan, eConf **C010630**, P510 (2001), hep-ph/0110378. I, A, ∗, A

[3] P. M. Nadolsky, H.-L. Lai, Q.-H. Cao, J. Huston, J. Pumplin, D. Stump, W.-K. Tung, and C.-P. Yuan, Phys. Rev. **D78**, 013004 (2008), 0802.0007. I, III A

[4] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, P. M. Nadolsky, and W. K. Tung, JHEP **07**, 012 (2002), hep-ph/0201195. II, A, A

[5] T.-J. Hou, S. Dulat, J. Gao, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, D. Stump, and C.-P. Yuan, Phys. Rev. **D95**, 034003 (2017), 1609.07968. 1, III, III B

[6] W. T. Giele and S. Keller, Phys. Rev. **D58**, 094023 (1998), hep-ph/9803393. III A

[7] W. T. Giele, S. A. Keller, and D. A. Kosower (2001), hep-ph/0104052. III A

[8] D. Stump, J. Pumplin, R. Brock, D. Casey, J. Huston, J. Kalk, H. L. Lai, and W. K. Tung, Phys. Rev. **D65**, 014012 (2001), hep-ph/0101051. III A

[9] J. L. Abelleira Fernandez et al. (LHeC Study Group), J. Phys. **G39**, 075001 (2012), 1206.2913. IV

[10] M. Klein, *Lhec data* (2017), URL `http://hep.ph.liv.ac.uk/~mklein/lhecdata/`. IV

[11] W.-K. Tung, H.-L. Lai, A. Belyaev, J. Pumplin, D. Stump, and C.-P. Yuan, JHEP **02**, 053 (2007), hep-ph/0611254. A, A

[12] D. Stump, J. Huston, J. Pumplin, W.-K. Tung, H. L. Lai, S. Kuhlmann, and J. F. Owens, JHEP **10**, 046 (2003), hep-ph/0303013. ∗

[13] A. C. Benvenuti et al. (BCDMS), Phys. Lett. **B223**, 485 (1989). B

[14] A. C. Benvenuti et al. (BCDMS), Phys. Lett. **B237**, 592 (1990). B

[15] M. Arneodo et al. (New Muon Collaboration), Nucl. Phys. **B483**, 3 (1997), hep-ph/9610231. B

[16] J. P. Berge et al., Z. Phys. **C49**, 187 (1991). B

[17] U.-K. Yang et al. (CCFR/NuTeV), Phys. Rev. Lett. **86**, 2742 (2001), hep-ex/0009041. B

[18] W. G. Seligman et al., Phys. Rev. Lett. **79**, 1213 (1997), hep-ex/9701017. B

[19] D. A. Mason, Ph.D. thesis, Oregon U. (2006), URL `http://lss.fnal.gov/archive/thesis/2000/fermilab-thesis-2006-01.pdf`. B

[20] M. Goncharov et al. (NuTeV), Phys. Rev. **D64**, 112006 (2001), hep-ex/0102049. B

[21] A. Aktas et al. (H1), Eur. Phys. J. **C40**, 349 (2005), hep-ex/0411046. B

[22] A. Aktas et al. (H1), Eur. Phys. J. **C45**, 23 (2006), hep-ex/0507081. B

[23] H. Abramowicz et al. (ZEUS, H1), Eur. Phys. J. **C73**, 2311 (2013), 1211.1182. B

[24] H. Abramowicz et al. (ZEUS, H1), Eur. Phys. J. **C75**, 580 (2015), 1506.06042. B

[25] F. D. Aaron et al. (H1), Eur. Phys. J. **C71**, 1579 (2011), 1012.4355. B

[26] G. Moreno et al., Phys. Rev. **D43**, 2815 (1991). B

[27] R. S. Towell et al. (NuSea), Phys. Rev. **D64**, 052002 (2001), hep-ex/0103030. B

[28] J. C. Webb et al. (NuSea) (2003), hep-ex/0302019. B

[29] F. Abe et al. (CDF), Phys. Rev. Lett. **77**, 2616 (1996). B

[30] D. Acosta et al. (CDF), Phys. Rev. **D71**, 051104 (2005), hep-ex/0501023. B

[31] V. M. Abazov et al. (D0), Phys. Rev. **D77**, 011106 (2008), 0709.4254. B

[32] R. Aaij et al. (LHCb), JHEP **06**, 058 (2012), 1204.1620. B

[33] V. M. Abazov et al. (D0), Phys. Lett. **B658**, 112 (2008), hep-ex/0608052. B

[34] T. A. Aaltonen et al. (CDF), Phys. Lett. **B692**, 232 (2010), 0908.3914. B

[35] S. Chatrchyan et al. (CMS), Phys. Rev. **D90**, 032004 (2014), 1312.6283. B

[36] S. Chatrchyan et al. (CMS), Phys. Rev. Lett. **109**, 111806 (2012), 1206.2598. B

[37] G. Aad et al. (ATLAS), Phys. Rev. **D85**, 072004 (2012), 1109.5141. B

[38] V. M. Abazov et al. (D0), Phys. Rev. **D91**, 032007 (2015), [Erratum: Phys. Rev.D91,no.7,079901(2015)], 1412.2862. B

[39] T. Aaltonen et al. (CDF), Phys. Rev. **D78**, 052006 (2008), [Erratum: Phys. Rev.D79,119902(2009)], 0807.2204. B

[40] V. M. Abazov et al. (D0), Phys. Rev. Lett. **101**, 062001 (2008), 0802.2400. B

[41] G. Aad et al. (ATLAS), Phys. Rev. **D86**, 014022 (2012), 1112.6297. B

[42] S. Chatrchyan et al. (CMS), Phys. Rev. **D87**, 112002 (2013), [Erratum: Phys. Rev.D87,no.11,119902(2013)], 1212.6660. B

[43] R. Aaij et al. (LHCb), JHEP **08**, 039 (2015), 1505.07024. B

[44] R. Aaij et al. (LHCb), JHEP **05**, 109 (2015), 1503.00963. B

[45] G. Aad et al. (ATLAS), JHEP **09**, 145 (2014), 1406.3660. B

[46] V. Khachatryan et al. (CMS), Eur. Phys. J. **C76**, 469 (2016), 1603.01803. B

[47] R. Aaij et al. (LHCb), JHEP **01**, 155 (2016), 1511.08039. B

[48] G. Aad et al. (ATLAS), JHEP **08**, 009 (2016), 1606.01736. B

[49] G. Aad et al. (ATLAS), Eur. Phys. J. **C76**, 291 (2016), 1512.02192. B

[50] V. Khachatryan et al. (CMS), Phys. Lett. **B749**, 187 (2015), 1504.03511. B

[51] V. Khachatryan et al. (CMS), JHEP **02**, 096 (2017), 1606.05864. B

[52] S. Chatrchyan et al. (CMS), Phys. Rev. **D90**, 072006 (2014), 1406.0324. B

[53] G. Aad et al. (ATLAS), JHEP **02**, 153 (2015), [Erratum: JHEP09,141(2015)], 1410.8857. B

[54] V. Khachatryan et al. (CMS), JHEP **03**, 156 (2017), 1609.05331. B

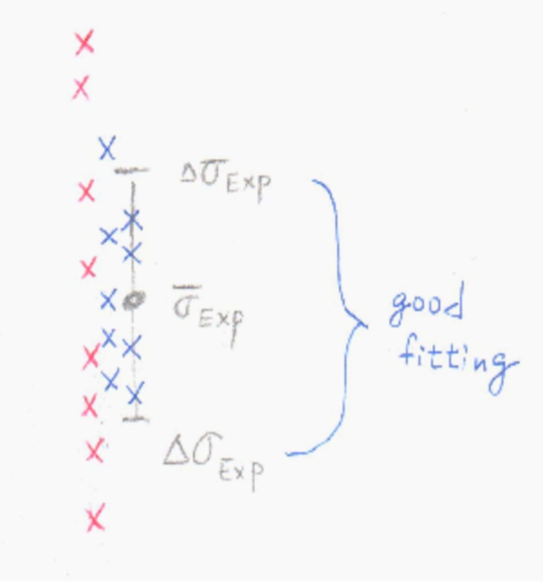


FIG. 5: Theoretical predictions and an experimental data point measurement with the error bar. Red crosses and blue crosses are two sets of theoretical prediction uncertainties.

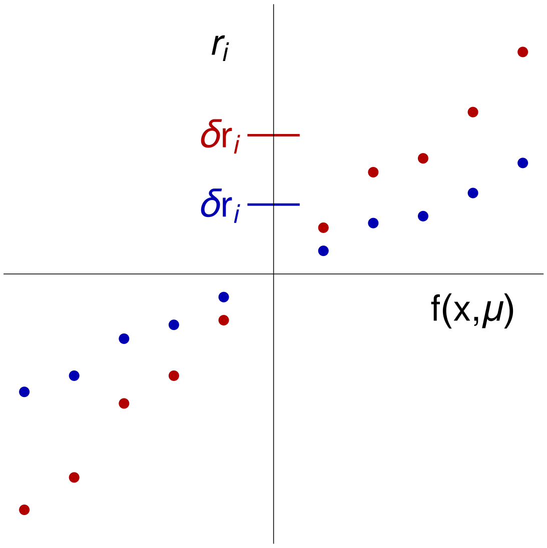


FIG. 6: The sensitivity of a data point to a PDF. Red circles and blue circles are residuals of two data points versus f(x,Q) values in PDF error sets

APPENDIX A: STATISTICAL FORMALISM

In this appendix we review a number of essential statistical issues that inform the calculations implemented in *PDFSense*. In many applications, it is instructive to study the correlations between the PDFs and the experimental observables. We review the relevant theoretical framework as presented in Ref.***. As an illustration of the relationship between residuals and experimental uncertainties discussed briefly in Sec. II we compare in Fig.5 two different fluctuations of theoretical values. Even though mean values of red crosses and blue crosses are the same, we can find the fluctuation of red crosses is easier to be detected because it's affection on residual values is larger than the fluctuation of blue crosses. Fig. 6 is the comparison of two different fluctuations of residuals depending on $f_a(x,\mu)$. Although both of red circles and blue circles are strongly correlated, red circles are more sensitive to $f_a(x,\mu)$ because the $f_a(x,\mu)$ fluctuation of red circles strongly affects values of residuals. Therefore, To understand the relationship between data sets and the constraints on PDFs imposed by these data sets, we should study whether χ^2 and r_i are sensitive to the variation of $f_a(x,\mu)$ values.

Let X be a variable that depends on the PDFs. We consider X as a function of the parameters $\{a_i\}$ that define the PDFs at the initial scale μ_0. Thus we have $X(\vec{a})$, where $\vec{a}$ forms a vector in an N-dimensional PDF parameter space, with N being the number of free parameters in the global analysis that determines these PDFs. In the Hessian formalism for the uncertainty analysis developed in [1] and used in all of our recent work, this parton parameter space is spanned by a set of orthonormal eigenvectors obtained by a self-consistent iterative procedure [4, 11].

If $\vec{a}_0$ represents the best fit obtained with a given set of theoretical and experimental inputs, the variation of $X(\vec{a})$ for parton parameters $\vec{a}$ in the neighborhood of $\vec{a}_0$ is given, within the Hessian approximation, by a linear formula

$$\Delta X(\vec{a}) = X(\vec{a}) - X(\vec{a}_0) = \vec{\nabla} X|_{\vec{a}_0} \cdot \Delta\vec{a}, \qquad \text{(A1)}$$

where $\vec{\nabla} X$ is the gradient of $X(\vec{a})$, and $\Delta\vec{a} = \vec{a} - \vec{a}_0$. As explained in detail in Refs. [1, 4, 11], the uncertainty range of the PDFs in our global analysis is characterized

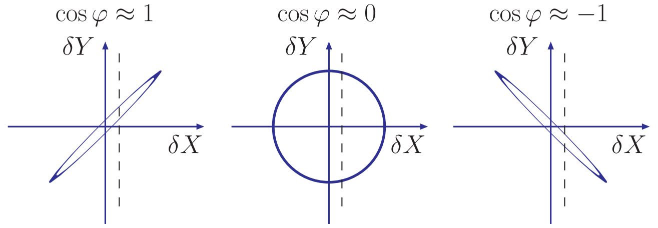


FIG. 7: Dependence on the correlation ellipse formed in the $\delta X - \delta Y$ plane on the value of $\cos\varphi$.

by a tolerance factor T, equal to the radius of a hypersphere spanned by maximal allowed displacements $\Delta\vec{a}$ in the orthonormal PDF parameter representation. T is determined by the criterion that all PDFs within this tolerance hypersphere should be consistent with the input experimental data sets within roughly 90% c.l. The detailed discussions and the specific iterative procedure used to construct the eigenvectors can be found in Refs. [1, 4, 11].

In practice, the results of our uncertainty analysis are characterized by $2N$ sets of published eigenvector PDF sets along with the central fit. We have 2 PDF sets for each of the N eigenvectors, along the ($\pm$) directions respectively, at the distance $|\Delta\vec{a}| = T$. The i-th component of the gradient vector $\vec{\nabla}X$ may be approximated by

$$\frac{\partial X}{\partial a_i} \equiv \partial_i X = \frac{1}{2}(X_i^{(+)} - X_i^{(-)}), \tag{A2}$$

where $X_i^{(+)}$ and $X_i^{(-)}$ are the values of X computed from the two sets of PDFs along the ($\pm$) direction of the i-th eigenvector. The uncertainty of the quantity X due to its dependence on the PDFs is then defined as

$$\Delta X = \left|\vec{\nabla}X\right| = \frac{1}{2}\sqrt{\sum_{i=1}^{N}\left(X_i^{(+)} - X_i^{(-)}\right)^2}, \tag{A3}$$

where for simplicity we assume that the positive and negative errors on X are the same.*

We may extend the uncertainty analysis to define a *correlation* between the uncertainties of two variables, say $X(\vec{a})$ and $Y(\vec{a})$. We consider the projection of the tolerance hypersphere onto a circle of radius 1 in the plane of the gradients $\vec{\nabla}X$ and $\vec{\nabla}Y$ in the parton parameter space [1, 2]. The circle maps onto an ellipse in the XY plane. This "tolerance ellipse" is described by Lissajous-style parametric equations,

$$\begin{aligned} X &= X_0 + \Delta X \cos\theta, && \text{(A4)} \\ Y &= Y_0 + \Delta Y \cos(\theta + \varphi), && \text{(A5)} \end{aligned}$$

where the parameter θ varies between 0 and 2π, $X_0 \equiv X(\vec{a}_0)$, and $Y_0 \equiv Y(\vec{a}_0)$. ΔX and ΔY are the maximal variations $\delta X \equiv X - X_0$ and $\delta Y \equiv Y - Y_0$ evaluated according to Eq. (A3), and φ is the angle between $\vec{\nabla}X$ and $\vec{\nabla}Y$ in the $\{a_i\}$ space, with

$$\begin{aligned} \cos\varphi &= \frac{\vec{\nabla}X \cdot \vec{\nabla}Y}{\Delta X \Delta Y} && \text{(A6)} \\ &= \frac{1}{4\Delta X\, \Delta Y}\sum_{i=1}^{N}\left(X_i^{(+)} - X_i^{(-)}\right)\left(Y_i^{(+)} - Y_i^{(-)}\right). \end{aligned}$$

The quantity $\cos\varphi$ characterizes whether the PDF degrees of freedom of X and Y are correlated ($\cos\varphi \approx 1$), anti-correlated ($\cos\varphi \approx -1$), or uncorrelated ($\cos\varphi \approx 0$). If units for X and Y are rescaled so that $\Delta X = \Delta Y$ (e.g., $\Delta X = \Delta Y = 1$), the semimajor axis of the tolerance ellipse is directed at an angle $\pi/4$ (or $3\pi/4$) with respect to the ΔX axis for $\cos\varphi > 0$ (or $\cos\varphi < 0$). In these units, the ellipse reduces to a line for $\cos\varphi = \pm 1$ and becomes a circle for $\cos\varphi = 0$, as illustrated by Fig. 7. These properties can be found by diagonalizing the equation for the correlation ellipse,

$$\left(\frac{\delta X}{\Delta X}\right)^2 + \left(\frac{\delta Y}{\Delta Y}\right)^2 - 2\left(\frac{\delta X}{\Delta X}\right)\left(\frac{\delta Y}{\Delta Y}\right)\cos\varphi = \sin^2\varphi. \tag{A7}$$

A magnitude of $|\cos\varphi|$ close to unity suggests that a precise measurement of X (constraining δX to be along the dashed line in Fig. 7) is likely to constrain tangibly the uncertainty δY in Y, as the value of Y shall lie within the needle-shaped error ellipse. Conversely, $\cos\varphi \approx 0$ implies that the measurement of X is not likely to constrain δY strongly.†

The parameters of the correlation ellipse are sufficient to deduce, under conventional approximations, a Gaussian probability distribution $P(X, Y|\text{CTEQ6.6})$ for finding certain values of X and Y based on the pre-LHC data sets included in the CTEQ6.6 analysis. If the LHC measures X and Y nearly independently of the PDF model, a new confidence region for X and Y satisfying both the CTEQ6.6 and LHC constraints can be determined by combining the prior probability $P(X, Y|\text{CTEQ6.6})$ with the new probability distribution $P(X, Y|\text{LHC})$ provided by the LHC measurement. For this purpose, it suffices to construct a probability distribution

$$\begin{aligned} &P(X, Y|\text{CTEQ6.6+LHC}) = \\ &= P(X, Y|\text{CTEQ6.6})P(X, Y|\text{LHC}) && \text{(A8)} \end{aligned}$$

which establishes the combined CTEQ6.6+LHC confidence region without repeating the global fit.

The values of ΔX, ΔY, and $\cos\varphi$ are also sufficient to estimate the PDF uncertainty of any function $f(X, Y)$ of

* A more detailed equation for ΔX accounts for differences between the positive and negative errors [2, 12]. It is used for $t\bar{t}$ cross sections in Table and Fig.

† The allowed range of $\delta Y/\Delta Y$ for a given $\delta \equiv \delta X/\Delta X$ is $r_Y^{(-)} \le \delta Y/\Delta Y \le r_Y^{(+)}$, where $r_Y^{(\pm)} \equiv \delta\cos\varphi \pm \sqrt{1-\delta^2}\sin\varphi$.

ID#	**Experimental data set**	N_{pts}	In CT14HERA2?	$\mathcal{L}$
101	BCDMS F_2^p [13]	337	yes	No paper
102	BCDMS F_2^d [14]	250	yes	No paper
104	NMC F_2^d/F_2^p [15]	123	yes	No info
106	NMC σ_{red}^p [15]	201	no	No info
108	CDHSW F_2^p [16]	85	yes	No info
109	CDHSW F_3^p [16]	96	yes	No info
110	CCFR F_2^p [17]	69	yes	No info
111	CCFR xF_3^p [18]	86	yes	No info
124	NuTeV $\nu\mu\mu$ SIDIS [19]	38	yes	No info?
125	NuTeV $\bar{\nu}\mu\mu$ SIDIS [19]	33	yes	No info?
126	CCFR $\nu\mu\mu$ SIDIS [20]	40	yes	No info
127	CCFR $\bar{\nu}\mu\mu$ SIDIS [20]	38	yes	No info
145	H1 σ_r^b [21][22]	10	yes	57.4 pb^{-1}
147	Combined HERA charm production [23]	47	yes	1504 pb^{-1}
160	HERA1+2 Combined NC and CC DIS [24]	1120	yes	1 fb^{-1}
169	H1 F_L [25]	9	yes	121.6 pb^{-1}

TABLE I: Experimental data sets considered in this analysis: deep-inelastic scattering.

X and Y by relating the gradient of $f(X,Y)$ to $\partial_X f \equiv \partial f/\partial X$ and $\partial_Y f \equiv \partial f/\partial Y$ via the chain rule:

$$\begin{aligned}\Delta f = \left|\vec{\nabla} f\right| = \big[\,(\Delta X\ \partial_X f\ )^2 & \qquad \text{(A9)}\\ + 2\Delta X\ \Delta Y\ \cos\varphi\ \partial_X f\ \partial_Y f + (\Delta Y\ \partial_Y f)^2\,\big]^{1/2}.\end{aligned}$$

Of particular interest is the case of a rational function $f(X,Y) = X^m/Y^n$, pertinent to computations of various cross section ratios, cross section asymmetries, and statistical significance for finding signal events over background processes [2]. For rational functions Eq. (**??**) takes the form

$$\frac{\Delta f}{f_0} = \sqrt{\left(m\frac{\Delta X}{X_0}\right)^2 - 2mn\frac{\Delta X}{X_0}\ \frac{\Delta Y}{Y_0}\ \cos\varphi\ + \left(n\frac{\Delta Y}{Y_0}\right)^2}. \tag{A10}$$

For example, consider a simple ratio, $f = X/Y$. Then $\Delta f/f_0$ is suppressed ($\Delta f/f_0 \approx |\Delta X/X_0 - \Delta Y/Y_0|$) if X and Y are strongly correlated, and it is enhanced ($\Delta f/f_0 \approx \Delta X/X_0 + \Delta Y/Y_0$) if X and Y are strongly anticorrelated.

As would be true for any estimate provided by the Hessian method, the correlation angle is inherently approximate. Eq. (**??**) is derived under a number of simplifying assumptions, notably in the quadratic approximation for the χ^2 function within the tolerance hypersphere, and by using a symmetric finite-difference formula (A2) for $\{\partial_i X\}$ that may fail if X is not monotonic. With these limitations in mind, we find the correlation angle to be a convenient measure of interdependence between quantities of diverse nature, such as physical cross sections and parton distributions themselves. For collider applications, the correlations between measured cross sections for crucial SM and beyond SM processes will be of primary interest, as we shall illustrate in Sec. As a first example however, we shall present some representative results on correlations between the PDFs in the next section.

APPENDIX B: SUPPLEMENTARY INFORMATION

In this section we gather a number of additional results to supplement the main findings of this article; while the focus in the preceding sections was upon a presentation of the basic details of *PDFSense*, we collect a number of physics plots that highlight the breadth of phenomena that can be contained within the main output of *PDFSense* — the sensitivity plots in (x,μ) space. Before this, we also collate tables detailing the experimental information contained in the plots of the previous sections.

In Tables I–III we provide a detailed key for the individual experiments mapped in Fig. 1, including the physical process, number of points, and luminosities, where available. We group these tables broadly according to subprocess — Table I corresponds to DIS experiments, while Tables II and III collect various measurements for the hadroproduction of, *e.g.*, gauge boson, jet, and $t\bar{t}$ pairs — and thus provide a translation key for the experimental ID numbers given in Fig. 1.

In Tables IV and V we collect the flavor-specific (s_f) and overall ($\sum_f s_f$) sensitivities for the experimental datasets contained in this analysis. In Table IV we list the total and point-averaged sensitivities for each main flavor ($\bar{d}, \bar{u}, g, u, d, s$), while Table V gives the corresponding information for a number of quantities derived from these, as explained in the associated captions.

We also gather a number of additional (x,μ) sensitivity plots for quantities of special interest.

To explore the sensitivity to Higgs production, we demonstrate the $x-\mu$ plots of $|s_{H14}|$ for LHC data sets of jet

ID#	Experimental data set	N_{pts}	In CT14HERA2?	$\mathcal{L}$
201	E605 DY [26]	119	yes	No paper
203	E866 DY, $\sigma_{pd}/(2\sigma_{pp})$ [27]	15	yes	No info
204	E866 DY, $Q^3 d^2\sigma_{pp}/(dQdx_F)$ [28]	184	yes	No info
225	CDF Run-1 $A_e(\eta^e)$ [29]	11	yes	110 pb^{-1}
227	CDF Run-2 $A_e(\eta^e)$ [30]	11	yes	170 pb^{-1}
234	D$\emptyset$ Run-2 $A_\mu(\eta^\mu)$ [31]	9	yes	0.3 fb^{-1}
240	LHCb 7 TeV W/Z muon forward-η Xsec [32]	14	yes	35 pb^{-1}
241	LHCb 7 TeV W $A_\mu(\eta^\mu)$ [32]	5	yes	35 pb^{-1}
260	D$\emptyset$ Run-2 Z $d\sigma/dy_Z$ [33]	28	yes	0.4 fb^{-1}
261	CDF Run-2 Z $d\sigma/dy_Z$ [34]	29	yes	2.1 fb^{-1}
266	CMS 7 TeV $A_\mu(\eta)$ [35]	11	yes	4.7 fb^{-1}
267	CMS 7 TeV $A_e(\eta)$ [36]	11	yes	840 pb^{-1}
268	ATLAS 7 TeV W/Z Xsec, $A_\mu(\eta)$ [37]	41	yes	35 pb^{-1}
281	D$\emptyset$ Run-2 $A_e(\eta)$ [38]	13	yes	9.7 fb^{-1}
504	CDF Run-2 incl. jet $(d\sigma/dp_T^j dy_j)$ [39]	72	yes	1.13 fb^{-1}
514	D$\emptyset$ Run-2 incl. jet $(d\sigma/dp_T^j dy_j)$ (???) [40]	110	yes	0.7 fb^{-1}
535	ATLAS 7 TeV incl. jet $(d\sigma/dp_T^j dy_j)$ [41]	90	yes	35 pb^{-1}
538	CMS 7 TeV incl. jet $(d\sigma/dp_T^j dy_j)$ [42]	133	yes	5 fb^{-1}

TABLE II: Same as Table I, showing experimental data sets for production of vector bosons, single-inclusive jets, and $t\bar{t}$ pairs.

ID#	Experimental data set	N_{pts}	In CT14HERA2?	$\mathcal{L}$
245	LHCb 7 TeV Z/W muon forward-η Xsec [43]	33	no	1.0 fb^{-1}
246	LHCb 8 TeV Z electron forward-η $d\sigma/dy_Z$ [44]	17	no	2.0 fb^{-1}
247	ATLAS 7 TeV $d\sigma/dp_T^Z$ [45]	8	no	4.7 fb^{-1}
249	CMS 8 TeV W muon, Xsec, $A_\mu(\eta^\mu)$ [46]	33	no	18.8 fb^{-1}
250	LHCb 8 TeV W/Z muon, Xsec, $A_\mu(\eta^\mu)$ [47]	42	no	2.0 fb^{-1}
252	ATLAS 8 TeV Z $(d^2\sigma/d\|y\|_{ll}dm_{ll})$ [48]	48	no	20.3 fb^{-1}
253	ATLAS 8 TeV $(d^2\sigma/dp_T^Z dm_{ll})$ [49]	45	no	20.3 fb^{-1}
254	CMS 8 TeV $(d^2\sigma/dp_T^Z dy_Z)$ [50]	20	no	19.7 fb^{-1}
255	CMS 8 TeV $(d\sigma/dp_T^{W/Z})$ [51]	9	no	18.4 pb^{-1}
542	CMS 7 TeV incl. jet, R=0.7, $(d\sigma/dp_T^j dy_j)$ [52]	66	no	5 fb^{-1}
544	ATLAS 7TeV incl. jet, R=0.6, $(d\sigma/dp_T^j dy_j)$ [53]	60	no	4.5 fb^{-1}
545	CMS 8TeV incl. jet, R=0.7, $(d\sigma/dp_T^j dy_j)$ [54]	185	no	19.7 fb^{-1}
565	ATLAS 8 TeV $t\bar{t}$ $d\sigma/dp_T^t$, pseudodata	8	no	20.3 fb^{-1}
566	ATLAS 8 TeV $t\bar{t}$ $d\sigma/dy_{<t/\bar{t}>}$, pseudodata	5	no	20.3 fb^{-1}
567	ATLAS 8 TeV $t\bar{t}$ $d\sigma/dm_{t\bar{t}}$, pseudodata	7	no	20.3 fb^{-1}
568	ATLAS 8 TeV $t\bar{t}$ $d\sigma/dy_{t\bar{t}}$, pseudodata	5	no	20.3 fb^{-1}
191	LHeC CC e^-p			
192	LHeC CC e^+p			
193	LHeC NC e^-p			
194	LHeC NC e^+p			

TABLE III: Same as Table I, showing experimental data sets for production of vector bosons, single-inclusive jets, and $t\bar{t}$ pairs that are not incorporated in the CT14HERA2NNLO fit.

and $t\bar{t}$ processes and p_T^Z distribution of Z production in Fig. 11. We de-emphasize the points with $|s_{H14}| < 0.25$ via the gray color. The higher fraction of colored jet data indicates that the jet data in the present LHC measurement can constrain the theoretical prediction for Higgs cross section.

No.	Exp. ID	N_{pts}	$\sum_f s_f$	$\sum_f \overline{s}_f/N_f$	$R[s_{\bar{u}}]$	$R[\overline{s}_{\bar{u}}]$	$R[s_{\bar{d}}]$	$R[\overline{s}_{\bar{d}}]$	$R[s_g]$	$R[\overline{s}_g]$	$R[s_u]$	$R[\overline{s}_u]$	$R[s_d]$	$R[\overline{s}_d]$	$R[s_s]$	$R[\overline{s}_s]$
1	160	1120	620.	0.0922	B		A		A		A		B		C	
2	545	288	397.	0.234	B		B		A	2	C		C		B	
3	542	132	233.	0.295	C	1	C	1	B	2					C	
4	201	238	225.	0.158	B	1	B	1			C					
5	111	86	218.	0.423	C	2	C	2			B	2	C	1		
6	204	368	206.	0.0942	C		B		C		C					
7	101	337	184.	0.0909			C		C		B		C			
8	104	123	169.	0.229	C	1					C	1	B	1		
9	102	250	141.	0.0938	C				C		C		C			
10	109	96	115.	0.199	C	1	C	1			C	1	C			
11	538	222	109.	0.0834					C							
12	110	69	89.3	0.216					C	1				1		
13	250	84	82.9	0.165	C								C	1		
14	108	85	82.4	0.161									C			
15	268	82	79.3	0.161												
16	249	66	78.3	0.198		1								1		
17	252	94	68.5	0.121												
18	203	30	66.6	0.37	C	2	C	2						1		
19	245	66	60.3	0.152												
20	124	38	58.9	0.258											C	2
21	266	22	58.8	0.445		2		1		1		1		2		
22	514	176	56.8	0.0549					C							
23	535	150	54.8	0.0617					C							
24	127	38	49.4	0.217											C	2
25	126	40	48.	0.2											C	2
26	125	33	36.7	0.185												1
27	504	118	36.1	0.0518												
28	544	118	34.5	0.0487												
29	234	18	30.	0.278								1		1		1
30	267	22	28.6	0.216		1								1		
31	281	26	28.	0.179										1		
32	260	56	23.2	0.0691												
33	225	22	17.7	0.134										1		
34	253	45	17.2	0.0638												
35	147	47	15.1	0.0537												
36	240	28	14.6	0.0867												
37	254	40	14.3	0.0596												
38	246	34	14.2	0.0696												
39	241	10	12.2	0.204		1								1		
40	227	22	7.39	0.056												
41	568	10	6.8	0.113						1						
42	566	10	6.38	0.106						1						
43	565	8	6.15	0.128						1						
44	247	8	5.84	0.122												
45	169	9	3.99	0.0739						1						
46	567	7	3.9	0.0928						1						
47	255	9	2.48	0.0459												
48	145	10	1.14	0.0191												

TABLE IV: We have defined the flavor-specific sensitivity s_f and its point-averaged counterpart $\overline{s}_f = s_f/N_{pts}$. Using these quantities we tabulate the total overall (*i.e.*, flavor summed) sensitivity and a flavor dependent sensitivity for the various experiments in our data set, ordering the table in descending magnitude for the total overall sensitivity; thus, row 1 for the combined HERA Run I + Run 2 dataset has the greatest overall sensitivity while row 48 for the H1 σ_r^b reduced cross section has the least overall sensitivity according to that metric. For each flavor, we award particularly sensitive experiments a rank $R[s_f] = \mathrm{A}, \mathrm{B}, \mathrm{C}$ or $R[\overline{s}_f] = 1, 2, 3$ based on their total and point-averaged sensitivities, respectively. These ranks are decided using the criteria $R[s_f] = C$ for $s_f \in [20, 50]$, $R[s_f] = B$ for $s_f \in [50, 100]$, and $R[s_f] = A$ for $s_f > 100$ according to the total sensitivities for each flavor and, analogously, $R[\overline{s}_f] = 3$ for $\overline{s}_f \in [0.25, 0.5]$, $R[\overline{s}_f] = 2$ for $\overline{s}_f \in [0.5, 1]$, and $R[\overline{s}_f] = 1$ for $\overline{s}_f > 1$ according to the point-averaged sensitivities. Experiments with mean sensitivities falling below the lowest ranks (that is, with $s_f < 20$ or $\overline{s}_f < 0.25$) are not awarded a rank for that category/flavor. Note that we take $N_f = 6$ for the light quark + gluon flavors to compute $\sum_f \overline{s}_f/N_f$ within this table.

No.	Exp. ID	$R[s_{u_v}]$	$R[\bar{s}_{u_v}]$	$R[s_{d_v}]$	$R[\bar{s}_{d_v}]$	$R[s_{\bar{d}/\bar{u}}]$	$R[\bar{s}_{\bar{d}/\bar{u}}]$	$R[s_{d/u}]$	$R[\bar{s}_{d/u}]$	$R[s_{H7}]$	$R[\bar{s}_{H7}]$	$R[s_{H8}]$	$R[\bar{s}_{H8}]$	$R[s_{H14}]$	$R[\bar{s}_{H14}]$
1	160	B		C		C		B		B		B		B	
2	545	C				C		C		A	1	A	2	A	2
3	542									B	2	B	2	B	2
4	201	C		C		C									
5	111	B	2	B	2			C	1						
6	204	B		C		C		C				C		C	
7	101	B		C		C		C		C					
8	104	C	1	C		C	1	B	2						
9	102	C		C				C		C		C			
10	109	C	1	C	1										
11	538									C		C		C	
12	110														
13	250					C	1	C	1						
14	108														
15	268														
16	249					C	1		1						
17	252														
18	203		1		1	B	3		1						
19	245						1		1						
20	124														
21	266		1		1		2	C	2						
22	514														
23	535														
24	127														
25	126														
26	125														
27	504														
28	544														
29	234		1		1		1		1						
30	267						1		1						
31	281				1				1						
32	260														
33	225				1				1						
34	253														
35	147														
36	240														
37	254														
38	246														
39	241						1		1						
40	227														
41	568										1		1		
42	566										1		1		
43	565										1		1		1
44	247														
45	169														
46	567														
47	255														
48	145														

TABLE V: A horizontal continuation of the information in Table IV, specifically, containing the flavor-dependent total and mean sensitivities of a number of derived quantities, as opposed to the individual flavors given in Table IV. Specifically, going across, the total and mean sensitivities are tabulated for valence distributions of the u and d quarks, the partonic flavor ratios $\bar{d}/\bar{u}$ and d/u, and the Higgs production cross section $\sigma_{pp\to H^0X}$ at 7, 8, and 14 TeV, respectively. The ranking criteria and ordering are again as described in Table IV.

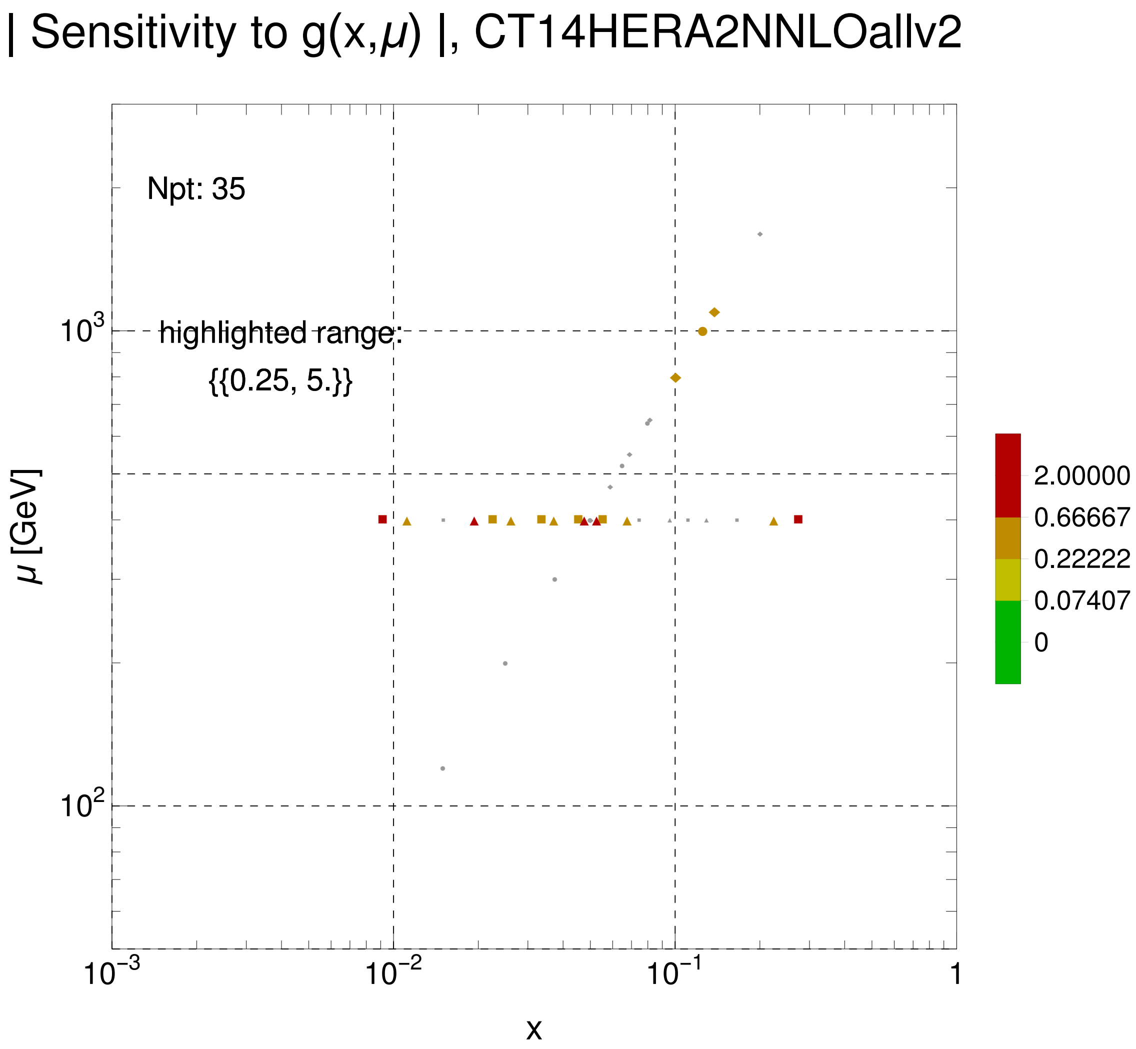


FIG. 8: Sensitivity plot for the ttbar production.

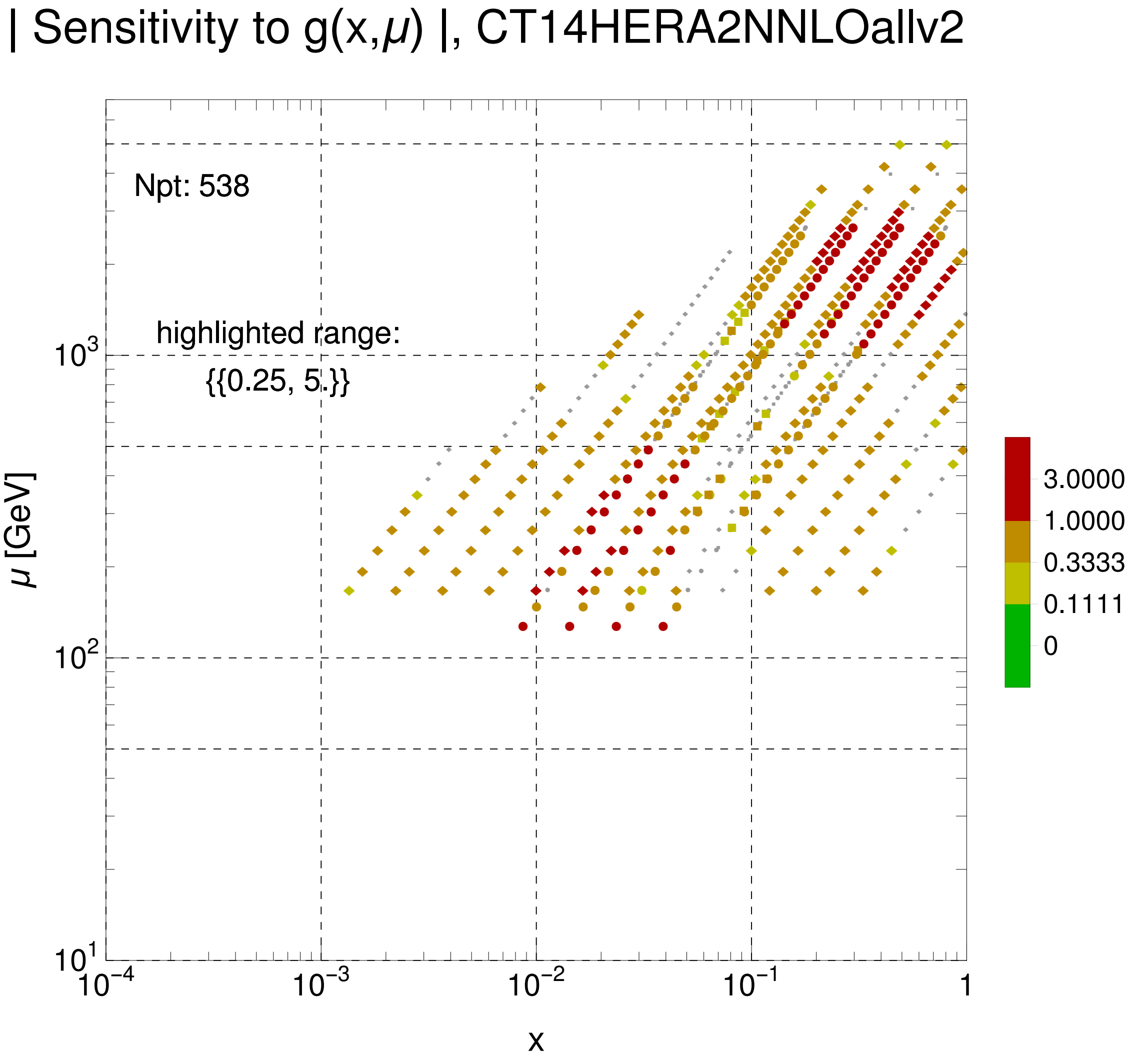
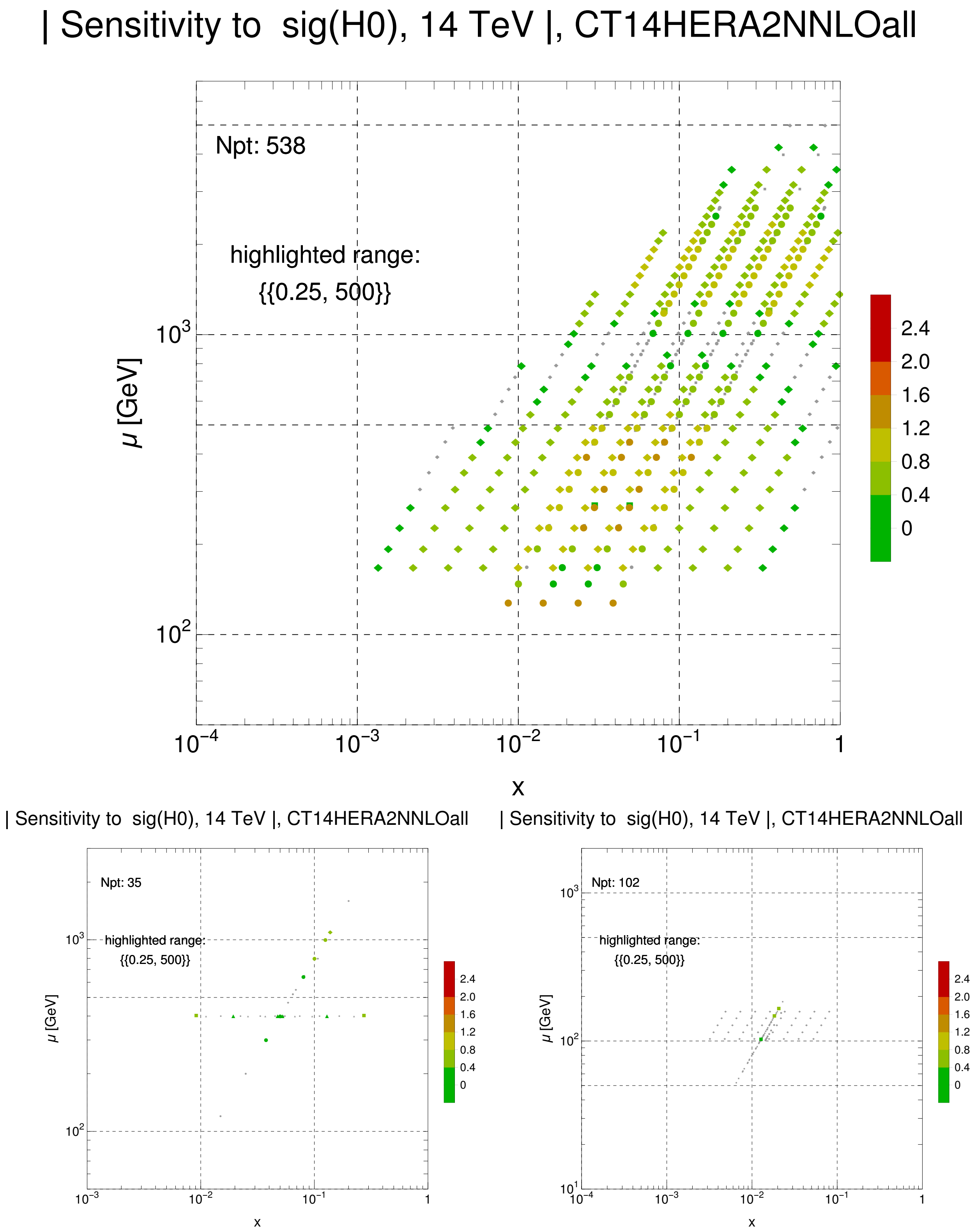


FIG. 9: sensitivity for the jet production

FIG. 10: sensitivity to the Zpt production



| Sensitivity to sig(H0), 14 TeV |, CT14HERA2NNLOall

| Sensitivity to sig(H0), 14 TeV |, CT14HERA2NNLOall

| Sensitivity to sig(H0), 14 TeV |, CT14HERA2NNLOall

FIG. 11: Sensitivity plot for the Higgs production. The left, middle, and right figures display the sensitivity to the pseudodata of Higgs cross section at 14 TeV for the jet processes (ID = 542, 544, 545 in III), $t\bar{t}$ processes (ID = 565˜568 in III), and $d\sigma/dp_T^Z$ of Z processes (ID = 247, 253, 254, 255 in III).