Intrinsic charm parton distribution functions from CTEQ-TEA global analysis

Sayipjamal Dulat,[1,2,*] Tie-Jiun Hou,[3,†] Jun Gao,[4,‡] Joey Huston,[2,§] Jon Pumplin,[2,¶] Carl Schmidt,[2,**]
Daniel Stump,[2,††] and C.-P. Yuan[2,‡‡]

[1]*School of Physics Science and Technology, Xinjiang University, Urumqi, Xinjiang 830046, China*
[2]*Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824, USA*
[3]*Institute of Physics, Academia Sinica, Taipei 115, Taiwan*
[4]*Department of Physics, Southern Methodist University, Dallas, Texas 75275-0181, USA*

We study the possibility of intrinsic (nonperturbative) charm in parton distribution functions (PDF) of the proton, within the context of the CT10 next-to-next-to-leading order global analysis. Three models for the intrinsic charm (IC) quark content are compared: (i) $\hat{c}(x) = 0$ (zero-IC model); (ii) $\hat{c}(x)$ is parametrized by a valence-like parton distribution (BHPS model); (iii) $\hat{c}(x)$ is parametrized by a sea-like parton distribution (SEA model). In these models, the intrinsic charm content, $\hat{c}(x)$, is included in the charm PDF at the matching scale $Q_c = m_c = 1.3$ GeV. The best fits to data are constructed and compared. Correlations between the value of m_c and the amount of IC are also considered.

I. CHARM QUARKS IN THE CTEQ-TEA GLOBAL ANALYSIS

The global analysis uses QCD theory to analyze a broad range of experimental data. In particular, the theoretical predictions for short-distance scattering processes allow the measurement, within some approximations, of universal parton distribution functions (PDFs) for the proton. These functions can then be used to predict hadronic cross sections in the QCD and electroweak theories, and in beyond-the-standard-model theories. With the new high-precision data becoming available from the LHC, the goal of QCD global analysis is to be able to make predictions that are accurate to be less than about one percent.

The most recent CT10NNLO PDFs (referred to as the CT10 PDFs in this paper) are based on next-to-next-to-leading order (NNLO) approximations for perturbative QCD [1]. That is, NNLO approximations are used for the running coupling $\alpha_S(Q)$, for the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolution equations, and for those hard matrix elements for which the NNLO approximation is available [2–4]. [Next-to-leading order (NLO) is used only for inclusive jet data.]

Another important approximation in the CT10 analysis concerns the treatment of charm quark effects. There are two issues: the dependence on the assumed charm quark mass and the possibility of a nonperturbative charm component, intrinsic charm (IC), in the proton. The first issue has been addressed in many papers [5–9] and was considered recently in the context of the CT10 analysis in Ref. [10]. In that work the general dependence on the charm quark mass was studied and a preferred value of $m_c(m_c) = 1.15^{+0.18}_{-0.12}$ GeV was obtained at 68% C.L., where the error is a sum in quadrature of the PDF and theoretical uncertainties. Here, $m_c(m_c)$ denotes the running mass of the charm quark, defined in the modified minimal-subtraction ($\overline{\text{MS}}$) scheme and evaluated at the scale of m_c. This value, constrained primarily by a combination of inclusive and charm production measurements in HERA deep-inelastic scattering, translates into $m_c^{\text{pole}} = 1.31^{+0.19}_{-0.13}$ GeV and $1.54^{+0.18}_{-0.12}$ GeV when using the conversion formula in Eq. (17) of Ref. [11] at the one-loop and two-loop order, respectively.[1] Either converted value is compatible with the value of $m_c^{\text{pole}} = 1.3$ GeV, which was assumed by CT10, and which we shall use as our standard charm mass value in this paper.

The second issue, the possibility of intrinsic charm, is our primary concern here. In the standard CT10 analysis, the charm and anti-charm quark PDFs were turned on at the scale $Q = Q_c = m_c = 1.3$ GeV with an initial $\mathcal{O}(\alpha_s^2)$ distribution, consistent with NNLO matching [13]. Thus, at higher Q, most of the charm PDF is generated from DGLAP evolution. However, one can also consider the possibility of including an additional contribution, $\hat{c}(x)$, to

*sdulat@msu.edu
†tjhou@msu.edu
‡jung@mail.smu.edu
§huston@pa.msu.edu
¶pumplin@pa.msu.edu
**schmidt@pa.msu.edu
††Stump@pa.msu.edu
‡‡yuan@pa.msu.edu

[1]The pole mass cannot be used to arbitrarily high accuracy because of nonperturbative infrared effects in QCD. The full quark propagator has no pole because the quarks are confined so that the pole mass cannot be defined outside of perturbation theory [12].

the initial charm and anticharm PDFs at the scale Q_c, beyond that required by matching. In principle, this intrinsic charm content would be suppressed by powers of (Λ_{QCD}/m_c), but since this ratio is not very small, it may be important.

Thus, in this paper we ask these questions: What are the nonperturbative c and $\bar{c}$ components of the proton? Is intrinsic charm significant? Accurate predictions of the c and $\bar{c}$ parton distributions will be relevant to some important LHC measurements. For example, production of $W^{\pm}$ and Z^0 involves $c\bar{d}$, $c\bar{s}$, $d\bar{c}$, $s\bar{c}$, and $c\bar{c}$ contributions. Another example is charm particle production at the LHC, which will depend quite directly on the c and $\bar{c}$ partons; some data for this process have already been published [14]. In addition, because of the momentum sum rule, an increase in the charm component of the proton must be compensated by a decrease in other components. So, in a model with IC, a compensating change in the gluon PDF could change the theoretical predictions of other processes, such as jet production at the Tevatron and Higgs boson production at the LHC, which are not directly related to charm.

The current paper updates previous work on the charm PDFs, which was based on the CTEQ6.5 global analysis [15]. However, there are some important advances with respect to the previous work. Most importantly, the PDFs in this paper will be based on the NNLO approximation of perturbative QCD; the previous work was NLO. Also, some more recent data is now available and is used here: the *combined* H1 and ZEUS data for both inclusive deep-inelastic scattering [16] and inclusive charm production [17] at HERA. Given these improvements to theory and experiment, we expect that this updated analysis will yield a better understanding of the charm PDF.

The outline of the paper is as follows. In Sec. II, we present three models of the intrinsic charm quark PDF and obtain fits and constraints on these models by performing a global analysis of the data, using a charm mass[2] of $m_c = 1.3$ GeV. In Sec. III we list the data used in the global analysis and present comparisons between theory and data for the data sets that are particularly sensitive to the charm quark PDF. In Sec. IV we use these PDFs (with the corresponding PDFs for other partons, of course) in calculations of processes at the LHC that would be particularly sensitive to the charm quark PDF. In Sec. V we consider the correlation of the amount of intrinsic charm with the value of the charm mass. In particular, we demonstrate this correlation by redoing the fit for one of the models with the larger value of the charm mass, $m_c = 1.67$ GeV. Sec. VI contains our conclusions. We also include an appendix, which contains the technical

definition of a novel variable that we use to assess the comparison of theory and data for the individual data sets.

II. FITS WITH AN INTRINSIC CHARM COMPONENT

Following our earlier study of c and $\bar{c}$ content of the proton [15], we will consider three models for the charm PDF $c(x, Q_c)$. We set the QCD coupling constant to $\alpha_s(M_Z) = 0.118$, which is close to the world average value, 0.1184 ± 0.0007, given by the Particle Data Group (PDG) [12]. In all three models, the charm PDF becomes active at the matching scale $Q_c = m_c = 1.3$ GeV. The first model is of the standard CT10 PDFs; they have an initial value of $c(x, Q_c) = \bar{c}(x, Q_c)$ that is fixed by NNLO matching[3] [13] and is $\mathcal{O}(\alpha_s^2)$. For $Q > Q_c$, additional charm content is generated by radiation, as required by the evolution equations of the renormalization group. For most values of Q (larger than a few GeV) this radiated charm content dominates the $\mathcal{O}(\alpha_s^2)$ initial contribution.

In the other two models, we assume some additional nonperturbative IC content, $\hat{c}(x)$, which is added to the $\mathcal{O}(\alpha_s^2)$ perturbative contribution at the scale Q_c. The second model, which we call the BHPS model, has an IC content that is parametrized by a valence-like nonperturbative function,

$$\hat{c}(x) = Ax^2[6x(1 + x)\ln x + (1 - x)(1 + 10x + x^2)]. \quad (1)$$

This model, which is representative of predictions from the light-cone picture of nucleon structure, is based on the original work of Brodsky, Hoyer, Peterson, and Sakai [18]; it was also considered in the NLO CTEQ6.5 study [15]. The third model, which we call the SEA model, has an IC content that is parametrized by a sea-like nonperturbative function,

$$\hat{c}(x) = A(\bar{d}(x, Q_0) + \bar{u}(x, Q_0)). \quad (2)$$

For all three models, we use the Fortran 95 package HOPPET [19] to include the IC with the NNLO matching and to evolve the PDFs at NNLO. We also use the NNLO S-ACOT-χ scheme [20], which is designed to approximately account for production threshold kinematics.

The other partons are parametrized at an initial scale $Q_0 = 1.295$ GeV with adjustable shape parameters. The parametrizations are essentially like CT10, but with some minor changes especially for the gluon PDF. The values of the shape parameters are varied to find the best fit to the global data set, which we shall describe in Sec. III. This best fit is obtained by minimizing a global χ^2 function with

---

[2]Throughout this paper, unless otherwise specified, the variable m_c will indicate the charm mass in the pole mass scheme.

[3]Note that, in the absence of intrinsic charm, the condition $c(x, Q_c) = \bar{c}(x, Q_c) = 0$ is only valid for $Q_c = m_c$ at NLO; at NNLO, the starting charm quark PDFs are $\mathcal{O}(\alpha_s^2)$ and nonzero, even for $Q_c = m_c$.

respect to the input shape parameters for different choices for IC. The fitting procedures include the treatment of systematic errors and other techniques that have been described in recent reports on CTEQ-TEA global analysis [1].

We note here that the scale parameters Q_0 and Q_c, as well as the mass scale introduced in the S-ACOT-χ [20,21] scheme could affect the determination of IC. In principle, any variation of the scale Q_0 can be absorbed into the parametrization of the initial PDFs, but in practice it can affect the overall χ^2 minimum. In the absence of IC, any variation of Q_c away from m_c would produce a change in the predictions at the next higher order in α_s. Similar dependence on the S-ACOT-χ rescaling variable should occur at higher order in α_s in the absence of IC. The effects induced from these latter two scales, Q_c and the rescaling variable χ, do not cancel in the presence of nonzero IC, but instead can be considered as part of the defining parametrization of IC. However, since we are concerned here with the *relative* change in the global χ^2 as IC is turned on, we can consider the dependence on these three auxiliary scales as part of the systematic theoretical uncertainty of the global analysis, which is no worse than their contribution to the uncertainty in global fits without IC. Thus, we keep them fixed, while varying the amount of IC. We keep fixed $Q_0 = 1.295$ GeV, $Q_c = m_c$, and we use the default definition of the ACOT-χ variable,

$$\chi = (1 + 4m_c^2/Q^2)x, \qquad (3)$$

for evaluating the coefficient functions in charm production; note that $\chi \sim x$ for $Q \gg m_c$. Different choices for these variables may affect the precise limits that we can place on IC, but we do not expect them to change our overall conclusions.

In contrast to the above three auxiliary scale parameters, we distinguish the charm quark mass parameter of QCD, m_c, which enters in the hard matrix elements through the coefficient functions. Since this is a fundamental parameter of the theory, it is possible that there is a correlation between the value of m_c and the amount of intrinsic charm that is physically significant. We shall postpone the discussion of this further until Sec. V. In this section and the following two sections we keep the charm mass fixed at the CTEQ standard value of $m_c = 1.3$ GeV.

To examine the dependence on the type and amount of intrinsic charm, we carry out a series of fits, varying the parameter A in (1) or (2). That is, we minimize χ^2 with respect to all variations of the input parameters, constrained by a fixed value of the intrinsic charm content, $\hat{c}(x)$, which we specify by its momentum fraction[4] at the scale Q_c,

---

[4]The notation $\langle x \rangle$ refers here to the momentum fraction or first moment of the given PDF; it does *not* signify the mean value of x, which is undefined, in general, if the 0th moment is undefined.

$$\langle x \rangle_{\rm IC} = \int_0^1 x[2\hat{c}(x)]dx, \qquad (4)$$

which in turn is determined by A. Here, we have multiplied by a factor of 2 in order to include both the c and $\bar{c}$ contributions. To satisfy the proton momentum sum rule at $Q \geq Q_c$, this momentum fraction has been subtracted from the total momentum fraction available to the light partons for $Q < Q_c$.

Figure 1 shows the results of the two IC series. We plot here the global chi-square function, χ_F^2, versus the intrinsic charm content, $\langle x \rangle_{\rm IC}$, for the two models of intrinsic charm. The BHPS model is shown in blue; the SEA model is shown in red. The function χ_F^2, which is called $\chi_{\rm global}^2$ in Ref. [1], includes the treatment of correlated systematic errors. The parabolic curves are determined from fits with many values of $\langle x \rangle_{\rm IC}$. Two exemplary fits for each model are shown as dots. They have $\langle x \rangle_{\rm IC} = 0.57\%$ and 2% for the BHPS models; and 0.57% and 1.5% for the SEA models. These four examples will be used in the subsequent discussion.

Figure 1 provides a first step toward setting upper limits on the intrinsic charm. As χ_F^2 increases, the goodness of fit to the global data set decreases, and at some point we judge that the data has ruled out the theory; this point could define a maximum acceptable value of $\langle x \rangle_{\rm IC}$. However, we know from experience that relying only on the value of χ_F^2 is not always the best measure of the goodness of fit. For example, there may be PDFs with small *global* χ_F^2, but for which one or a few individual data sets are very poorly fit, balanced by especially good fits to other data sets. Furthermore, in the global χ_F^2, a particular experiment that is very sensitive to a fit parameter but has a small number of
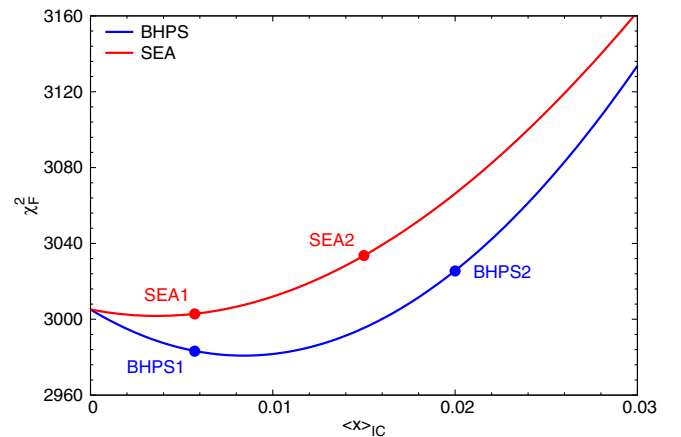

FIG. 1 (color online). The global chi-square function χ_F^2 versus the charm momentum fraction $\langle x \rangle_{\rm IC}$. The two approximately parabolic curves are determined from fits with many values of $\langle x \rangle_{\rm IC}$. Two exemplary fits for each IC model are shown as dots. Blue denotes the BHPS model; the dots have $\langle x \rangle_{\rm IC} = 0.57\%$ and 2%, which are denoted BHPS1 and BHPS2 below. Red denotes the SEA model; the dots have $\langle x \rangle_{\rm IC} = 0.57\%$ and 1.5%, which are denoted SEA1 and SEA2 below.
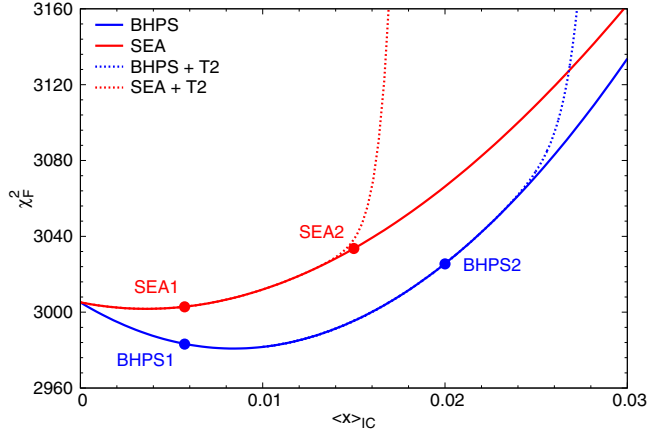
FIG. 2 (color online). Similar to Fig. 1, but the $\chi_F^2 + T_2$ versus charm momentum fraction, $\langle x \rangle_{IC}$, is also shown as the dotted curve for both the SEA and BPHS models.

data points, will be relatively underweighted compared to other data sets that are less sensitive but which have more data points. Thus, in order that the PDFs agree to a reasonable degree with all the individual data sets, we must also look at the χ^2 values from individual experiments, not just the overall sum of the χ^2 values.

To obtain a measure of the goodness of fit that includes the separate values of individual experiments, we introduce a "Tier-2 penalty," $T_2(i)$, for each experiment i. This measures the goodness of fit for that experiment; a large value means that the experiment i is not consistent with the theory. ($T_2(i)$ is not simply χ_i^2, but they are related. $T_2(i)$ is designed to increase more rapidly than χ_i^2 when χ_i^2 moves beyond the 90% C.L. Additional details are given in the appendix.)

The dotted curves in Fig. 2 show $\chi_F^2 + T_2$ versus $\langle x \rangle_{IC}$ for the two models of IC. Our usual choice for the "tolerance" of a PDF fit is

$$\Delta(\chi_F^2 + T_2) = \chi_F^2 + T_2 - (\chi_F^2 + T_2)_{min} < 100. \quad (5)$$

That is, a set of PDFs with $\Delta(\chi_F^2 + T_2) > 100$ is deemed to be such a poor fit to the data that it is ruled out (at the 90% C.L.). From Fig. 2 we see that, in the present study, the T_2 contribution turns on and rises very sharply with $\langle x \rangle_{IC}$, thus determining the upper limits on IC. The sharp increase in slope arises when one or more experiments become poorly fit as the charm momentum fraction increases. As we shall see and discuss in more detail in Sec. III, the dominant constraint on the SEA model is from the combined HERA charm production measurements; however, the main constraints on the BHPS model come from several measurements, but not the HERA charm production experiments. We conclude that the upper limits on $\langle x \rangle_{IC}$, at the 90% C.L., are

$$\langle x \rangle_{IC} \lesssim 0.025 \quad \text{for the BHPS model,}$$

$$\langle x \rangle_{IC} \lesssim 0.015 \quad \text{for the SEA model.}$$

Note that the four example PDFs are all within these limits. The global data does not rule out BHPS2 (which has $\langle x \rangle_{IC} = 2\%$) or SEA2 (which has $\langle x \rangle_{IC} = 1.5\%$). However, these are close to the upper limits for $\langle x \rangle_{IC}$, and can be taken as representative PDF models with amounts of IC near the largest acceptable value.

Figures 1 and 2 show that $\langle x \rangle_{IC} > 0.015$ is disfavored for the SEA model, while the BHPS model allows a larger charm content, up to about $\langle x \rangle_{IC} = 0.025$. Furthermore, the BHPS model provides a value of $\chi_F^2 + T_2$ that is *lower* than that of the standard CT10 PDFs for $\langle x \rangle_{IC} < 0.018$. The BHPS model with $\langle x \rangle_{IC} = 0.009$ gives the best fit to the global data set. However, the small decrease, $\Delta(\chi_F^2 + T_2) \simeq -20$, is really not significant enough to herald the discovery of intrinsic charm in a sense that the typical allowed χ^2 variation is $\lesssim 100$ in the CT10 gobal analysis. In the rest of the paper we shall focus on the four IC models marked as dots in Figs. 1 and 2, labeled as SEA1, SEA2, BHPS1, and BHPS2. The fraction of intrinsic charm component for each model is 0.57%, 1.5%, 0.57%, and 2%, respectively, cf. Table II.

The result in Fig. 1 is rather different from the corresponding result from our earlier NLO CTEQ6.5 study [15]. The order of magnitude of the dependence of χ_F^2 on $\langle x \rangle_{IC}$ is comparable to the earlier study. However, the behaviors of the BHPS model and the SEA model are reversed. In the current NNLO study, we find $\chi_{BHPS}^2 < \chi_{SEA}^2$, leading to a larger upper limit on $\langle x \rangle_{IC}$ for the BHPS model. In the older NLO study the inequality was the reverse. Because of the advances in both theory and data, listed above, we trust that the current results are more realistic regarding the issue of intrinsic charm. The most significant advance in the theoretical calculation lies in the fact that we have applied a better treatment of heavy parton mass in the current study.

More extensive comparisons between theory and experiment are given in Sec. III, while a discussion of the correlation of these fits with the value of the charm mass is deferred to Sec. V.

A. The charm quark PDF

Figure 3 shows the charm quark PDF for five cases. In each panel, the product of momentum fraction x and charm PDF $c(x, Q)$ is shown for the BHPS1, BHPS2, SEA1, SEA2, and CT10 PDFs. The four panels correspond to four values of Q: $Q = 2.0, 3.16, 8.0, 85$ GeV. CT10 has no IC; $\hat{c}(x) = 0$. But the charm distribution evolves rapidly from Q_c to 2 GeV. For $Q > 8$ GeV and $x < 0.01$, the CT10 charm PDF is comparable to the IC models.

The BHPS model has a valence-like charm distribution at Q_c. So for small Q, the charm density with $x > 0.1$ is significantly larger than our standard CT10 charm PDF.
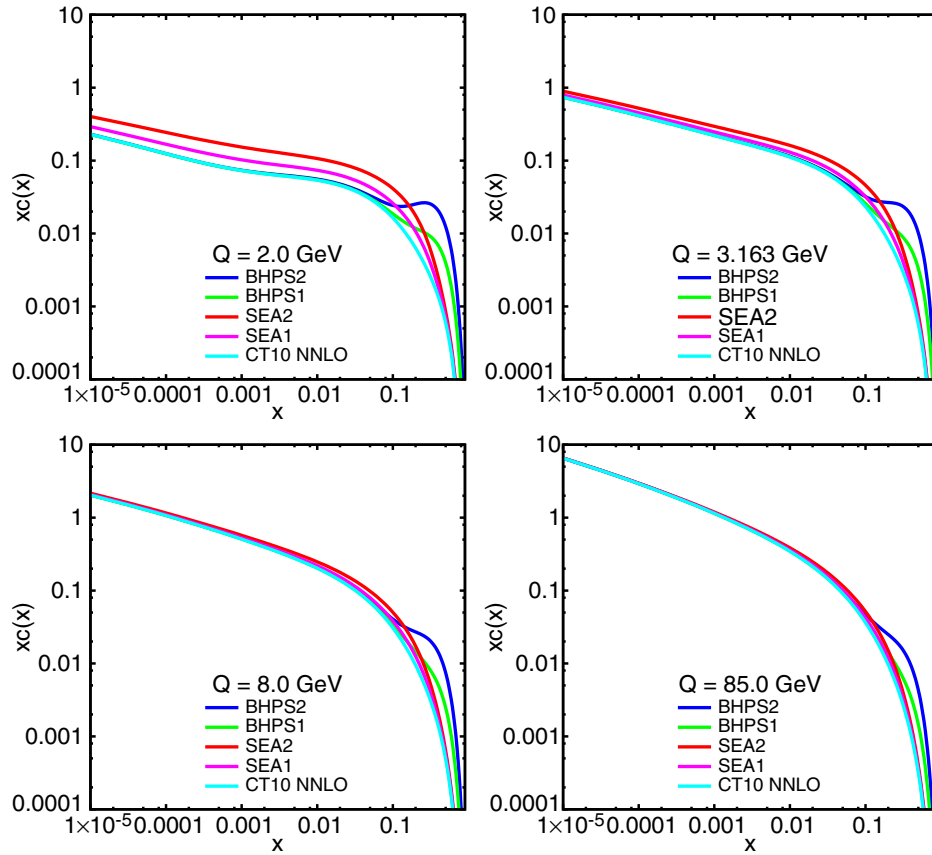
FIG. 3 (color online). Charm quark distribution $xc(x, Q)$ from the BHPS1 and BHPS2 PDFs (which have 0.57% and 2% $\langle x \rangle_{\mathrm{IC}}$); from SEA1 and SEA2 PDFs (which have 0.57% and 1.5% $\langle x \rangle_{\mathrm{IC}}$); and from CT10. The four graphs correspond to $Q = 2.0, 3.16, 8.0,$ and 85 GeV.

Even for Q as large as 85 GeV, $c(x, Q)$ is still notably larger for the BHPS model than for CT10, for large x. On the other hand, for small x, say $x < 0.05$, the BHPS model and CT10 charm densities are approximately equal for $Q > 8$ GeV. Because the IC in this model is concentrated at large x, the resulting PDFs can fit most data comparably well as CT10, as long as $\langle x \rangle_{\mathrm{IC}}$ is not too large; cf. Sec. III.

The SEA model has a sea-like charm distribution at Q_c. There is a large c density at Q_c for small and intermediate x ranges, say $x < 0.2$. The charm PDF evolves rapidly from Q_c to 2 GeV, but nevertheless $c(x, Q)$ continues to be notably larger than CT10 at small x, even for Q as large as 8 GeV. Because of this large charm density at low x, the SEA model has a more difficult time fitting the combined HERA charm data than CT10, or than the BPHS model for equal values of $\langle x \rangle_{\mathrm{IC}}$; cf. Sec. III.

To emphasize the differences between the models, relevant to LHC experiments, Fig. 4 compares $c(x, Q)$ for the four examples of the BHPS and SEA models at the large momentum scale $Q = 85$ GeV. For each case, the ratio $c(x, Q)_{\mathrm{IC}}/c(x, Q)_{\mathrm{CT10}}$ is plotted as a function of x. The shaded region is the uncertainty band for the charm distribution of CT10 at $Q = 85$ GeV. Note that the charm distributions with IC are allowed to be outside the uncertainty band of CT10, since CT10 was created

with the constraint that the charm PDF was radiatively generated.

Putting intrinsic charm into the c and $\bar{c}$ distributions will, of course, directly affect predictions involving charm
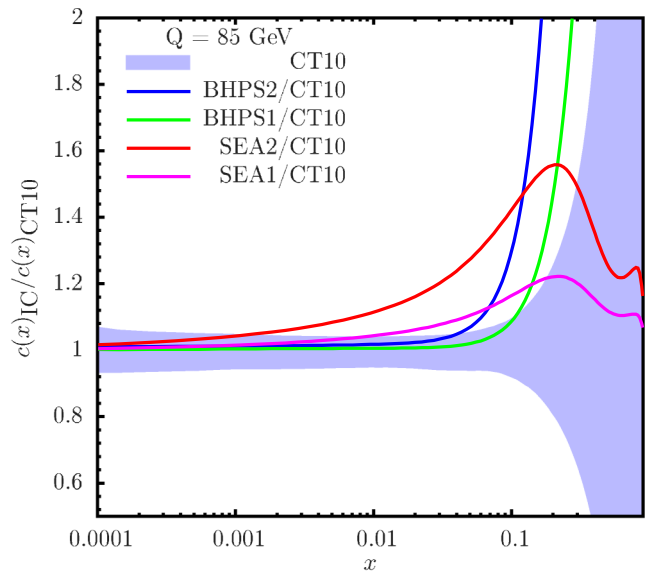

FIG. 4 (color online). Ratios of $c(x, Q)_{\mathrm{IC}}/c(x, Q)_{\mathrm{CT10}}$ for $Q = 85$ GeV.

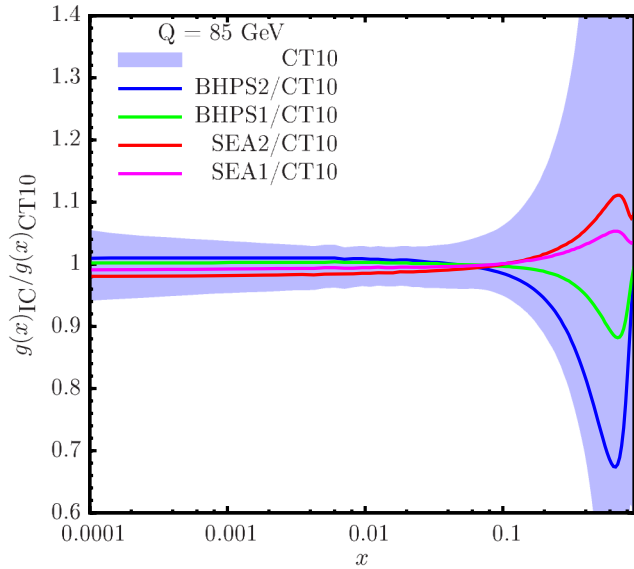FIG. 5 (color online). Ratios of $g(x)_{IC}/g(x)_{CT10}$ for $Q = 85$ GeV.

FIG. 7 (color online). Ratios of $u_v(x)_{IC}/u_v(x)_{CT10}$ for $Q = 85$ GeV.

quarks in the initial state. But there will also be indirect effects. Because of the momentum sum rule or constraints from the data, other parton distributions must change to balance the change in the charm distribution. Therefore, we should also look at other parton distributions in the PDF sets with intrinsic charm.

Figure 5 shows the gluon PDF for the same four examples of the BHPS and SEA models as in Fig. 4, again plotting the ratio to CT10 for the momentum scale $Q = 85$ GeV. Figure 6 shows the same for the $\bar{u} + \bar{d}$ quark PDF. The shaded regions show the uncertainty bands for CT10. We note that for the BHPS model, the presence of intrinsic
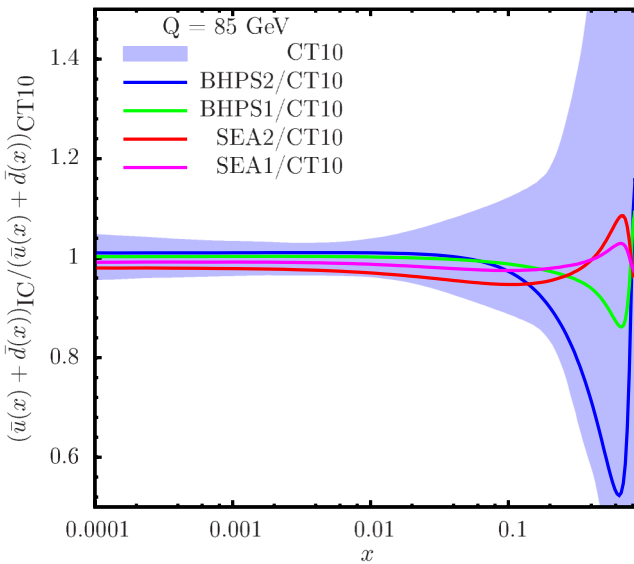
charm pulls momentum from the gluon and the $\bar{u} + \bar{d}$ quark at large x; whereas, for the SEA model, the presence of IC requires reduction of the $\bar{u} + \bar{d}$ quark PDF at small x to fit the inclusive deep inelastic scattering (DIS) data. Similarly, Figs. 7 and 8 are for changes of valence quarks. The u_v and d_v of the BHPS models are reduced in the valence region to balance the charm distribution there, and are increased in the small x region due to the valence number sum rules.

This will affect LHC predictions in interesting ways; cf. Sec. IV. Before concluding this section, we note that each IC fit is a central fit, with a different assumption on the intrinsic charm distribution at the scale Q_c. Hence, if we

FIG. 6 (color online). Ratios of $(\bar{u}(x) + \bar{d}(x))_{IC}/(\bar{u}(x) + \bar{d}(x))_{CT10}$ for $Q = 85$ GeV.
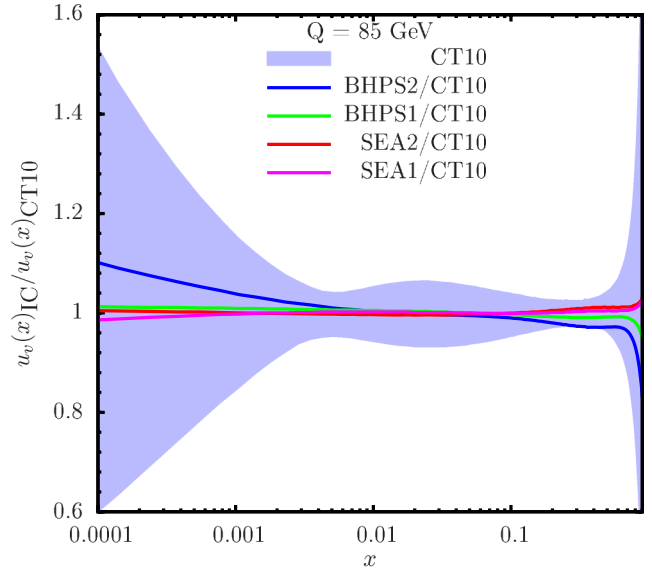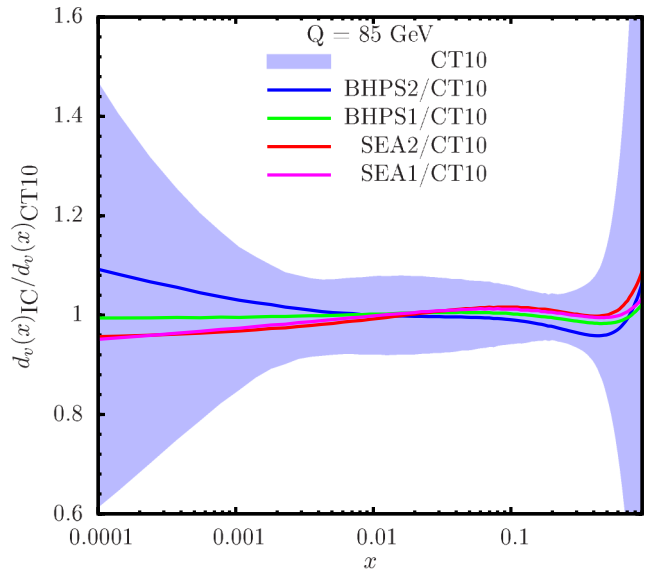

FIG. 8 (color online). Ratios of $d_v(x)_{IC}/d_v(x)_{CT10}$ for $Q = 85$ GeV.

computed eigenvector (Hessian) uncertainties for these IC fits, we would obtain an uncertainty range that expands on the CT10 uncertainty range for these gluon and quark distributions.

III. EXPERIMENTAL DATA SETS AND INTRINSIC CHARM

The experimental data sets included in the global analysis that led to the results of Sec. II are listed in Table I. With a few exceptions, they are the same as the data used in the standard CT10 fit [1]. We note that the CT10 PDFs were fitted to the individual F_2^c data sets from both H1 and ZEUS collaborations. As shown in Table I, in this study we have replaced those individual F_2^c data sets by the *combined* H1 and ZEUS reduced cross section data (called Data Set 147) for charm production at HERA [17].

We have not used the EMC data on charm production in muon scattering on iron [41] in part because we cannot be sure that nuclear effects will not be important there—although, of course, the same could be said for the NuTeV neutrino dimuon data, which we do include. An analysis of the EMC F_2^c data found $(0.86 \pm 0.60)\%$ for the IC probability [42]. However, subsequent studies [43–45] have shown that these EMC data are consistent with no IC, once the parton

distributions used for the analysis include an appropriate flexibility. A useful recent review of the IC data can be found in [46]. Including the EMC data is not likely to add any strong constraint to the upper limits of intrinsic charm obtained in our analysis, due to the uncertainty of nuclear corrections.

The impact of an intrinsic charm component on the fit to the different data sets can be visualized in several different ways. In Table I, we have listed the number of data points, N_{pt}, for each experiment, as well as the χ^2/N_{pt} for the CT10 fit, the BHPS2 fit, and the SEA2 fit, which have charm momentum fractions ($\langle x \rangle_{IC}$) of 2% and 1.5%, respectively. The change in χ^2/N_{pt} in the last two columns from the CT10 fit indicates which data sets are in conflict with a large intrinsic charm content.

One especially interesting data set for this paper is the *combined* H1 and ZEUS data set for charm production in deep-inelastic ep collisions at HERA [17]. This is Data Set 147 in Table I. Charm particles can be produced in two ways in an ep collision: the charm-excitation process $\gamma + c \to c + X$ and the charm-creation process $\gamma + g \to c + \bar{c}$. These are combined consistently in the calculations using the prescription introduced by ACOT [5]. However, in the presence of intrinsic charm, the former process becomes particularly important, since it is directly proportional to the

TABLE I. Experimental data sets employed in the CT10 analysis. N_{pt} = the number of points in the data set. The final three columns show χ^2/N_{pt} for each data set for CT10, the BHPS model with $\langle x \rangle_{IC} = 0.020$, and the SEA model with $\langle x \rangle_{IC} = 0.020$. The last row shows the global χ_F^2 for each PDF set for a total of 2625 data points.

ID	Experimental data set	N_{pt}	CT10	BHPS (0.020)	SEA (0.015)
101	BCDMS F_2^p [22]	339	1.158	1.087	1.220
102	BCDMS F_2^d [23]	251	1.157	1.119	1.187
103	NMC F_2^p [24]	201	1.656	1.668	1.582
104	NMC F_2^d/F_2^p [24]	123	1.210	1.311	1.207
108	CDHSW F_2^p [25]	85	0.832	0.833	0.781
109	CDHSW F_3^p [25]	96	0.809	0.867	0.810
110	CCFR F_2^p [26]	69	0.989	1.105	0.943
111	CCFR xF_3^p [27]	86	0.387	0.416	0.417
124	NuTeV ν di-μ SIDIS [28]	38	0.781	0.836	0.745
125	NuTeV $\bar{\nu}$ di-μ SIDIS [28]	33	0.852	0.905	0.864
126	CCFR ν di-μ SIDIS [29]	40	1.195	1.204	1.145
127	CCFR $\bar{\nu}$ di-μ SIDIS [29]	38	0.692	0.728	0.655
147	HERA charm production [17]	47	1.187	1.185	1.424
159	Combined HERA1 DIS [16]	579	1.068	1.086	1.059
201	E605 DY process $\sigma(pA)$ [30]	119	0.804	0.845	0.796
203	E866 DY process $\sigma(pd)/(2\sigma(pp))$ [31]	15	0.658	0.789	0.718
204	E866 DY process $\sigma(pp)$ [32]	184	1.271	1.286	1.277
225	CDF Run-1 W charge asymmetry [33]	11	1.292	1.191	1.285
227	CDF Run-2 W charge asymmetry [34]	11	0.978	0.995	0.978
231	V. M. Abazov *et al.* (D0 Run-2 W charge asymmetry [35]	12	1.928	2.006	1.972
234	V. M. Abazov *et al.* (D0 Run-2 W charge asymmetry [36]	9	1.501	1.709	1.371
260	V. M. Abazov *et al.* (D0 Run-2 Z rapidity dist. [37]	28	0.580	0.551	0.550
261	CDF Run-2 Z rapidity dist. [38]	29	1.586	1.466	1.535
504	CDF Run-2 inclusive jet [39]	72	1.398	1.431	1.311
514	V. M. Abazov *et al.* (D0 Run-2 inclusive jet [40]	110	1.044	0.950	1.012
	Totals	2625	3005	3034	3026

c and *c̄* component of the proton. Thus, one would expect data set 147 to be particularly sensitive to intrinsic charm. In addition, this set of combined H1 and ZEUS data has smaller systematic errors than the separate data sets, which were used in our previous analysis [15]. Therefore, we are interested to assess the influence of these newly available high-precision combined data.

Figure 9 shows a comparison of the H1 and ZEUS combined data on charm production with the theory predictions using IC parton distributions. The agreement

between data and theory for BHPS2 is satisfactory, and almost the same as for the CT10 PDFs. On the other hand, we see systematic differences between the data and SEA2, whose prediction is consistently higher than the data over the full range of Q^2 and x. Table II lists the values of χ^2 for CT10 and our four representative IC models with different amounts of IC. In general, the models with sea-like intrinsic charm do worse than the standard fit or the fits with the BHPS model with valence-like intrinsic charm. The sea-like models have more charm at low values of x, which is
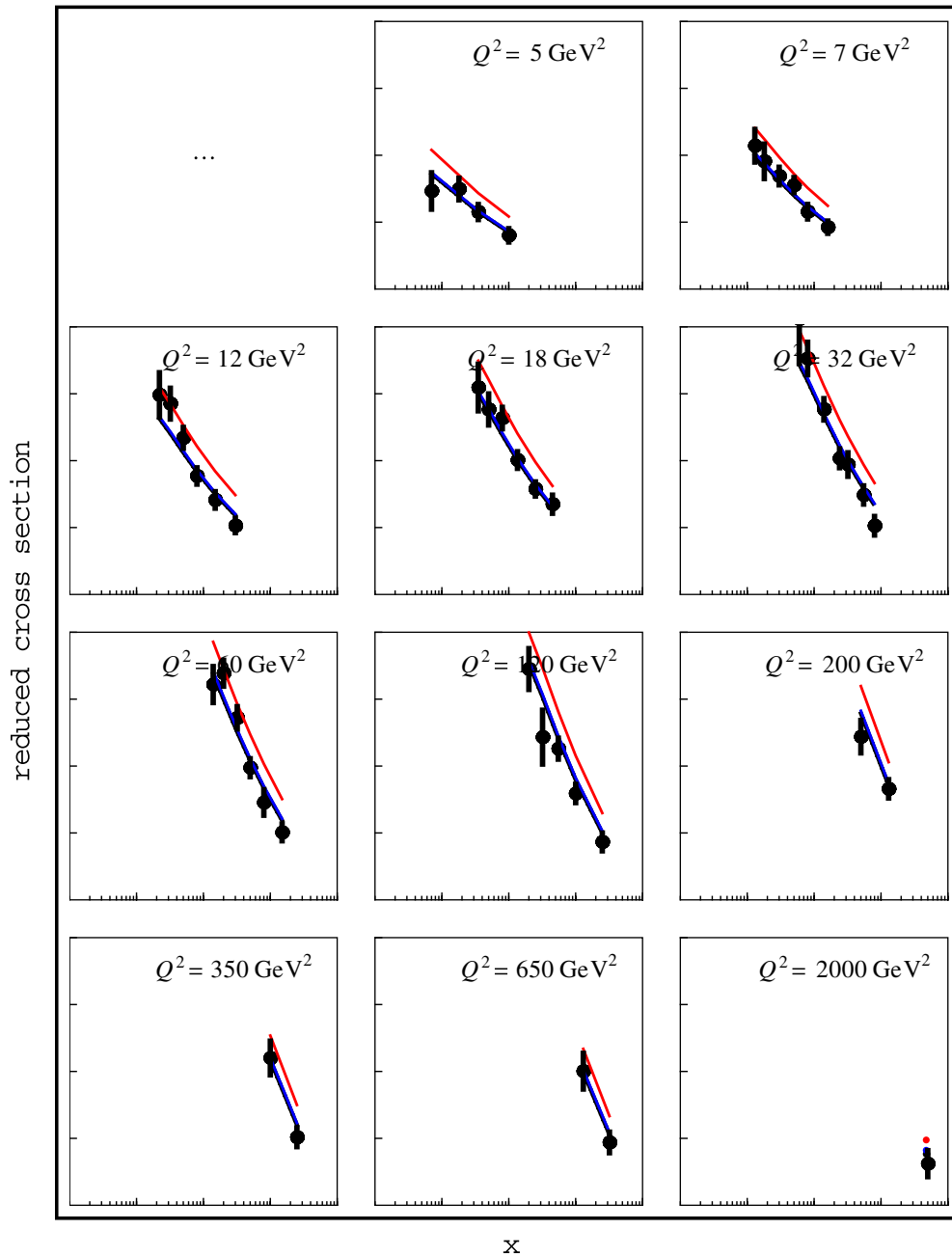

FIG. 9 (color online). Comparison of theory and data for the H1 and ZEUS combined data on charm production in ep collisions at HERA. Theory: The blue curves are the result for BHPS2; the red curves are the result for SEA2. The data is plotted as points; the error bars are the total errors. The CT10 prediction cannot be distinguished from that given by BHPS2.

TABLE II. χ^2 for different models of IC, for data set 147. This is the combined H1 and ZEUS data for inclusive charm production in $\bar{e}p$ or ep collisions at HERA. The number of data points is $N_{pt} = 47$.

Model	$\langle x \rangle_{IC}$	χ^2
CT10	0%	55.75
BHPS1	0.57%	55.78
BHPS2	2%	56.29
SEA1	0.57%	57.60
SEA2	1.5%	68.50

disfavored by this HERA data with small x. We note, however, that the increase in χ^2 for SEA2 is only about 13 units, to be compared to 47 data points.

The impact of intrinsic charm on the other data sets in Table I is not so obvious as for the charm production set 147. Some data sets favor the SEA models, while other data sets favor the BHPS models, while still others disfavor both. One other data set to note in particular is data set 101, which is the BCDMS F_2^p data. Although the decrease in χ^2/N_{pt} for BHPS2, and the increase in χ^2/N_{pt} for SEA2, are not obviously significant, the large number of data points $N_{pt} = 339$ ensures that these have an effect on the total χ^2. In this case the BHPS2 fit has χ^2 reduced by 24 units compared to CT10, while the SEA2 fit has χ^2 increased by 21 units. The net effect is that the full combination of all data sets to the total χ^2 disfavors the SEA model of intrinsic charm more than the BHPS model of intrinsic charm for a given intrinsic charm momentum fraction, as seen in Fig. 1.

A. An equivalent Gaussian variable

From the discussion in the last paragraph, we can see that the naive use of total χ^2 as the discriminating variable may overweight data sets with large numbers of points, even if the correlation with the fitting parameter is not very significant. It was for this reason that the Tier-2 penalty was added to χ^2 in the global fitting program. The Tier-2 penalty makes use of an equivalent Gaussian variable S_n, which gives a measure of the goodness of fit for each of the individual data sets. It is defined precisely in the appendix.

In other words, for a particular data set we map its $\{\chi^2, N\}$ value, assumed to obey a chi-square probability distribution, onto the variable S_n, which has the same probability but for a standard Gaussian distribution (cf. Appendix).

The values of S_n can then be interpreted in terms of probabilities in a normal distribution. Fits with S_n between -1 and 1 are accepted as reasonable, within the errors. Fits with $S_n > 3$ are considered poor fits. Fits with $S_n < -3$ actually fit the data much better than one would expect from normal statistical analysis; i.e., they have anomalously small residuals, presumably because the true experimental errors are smaller than the published values.
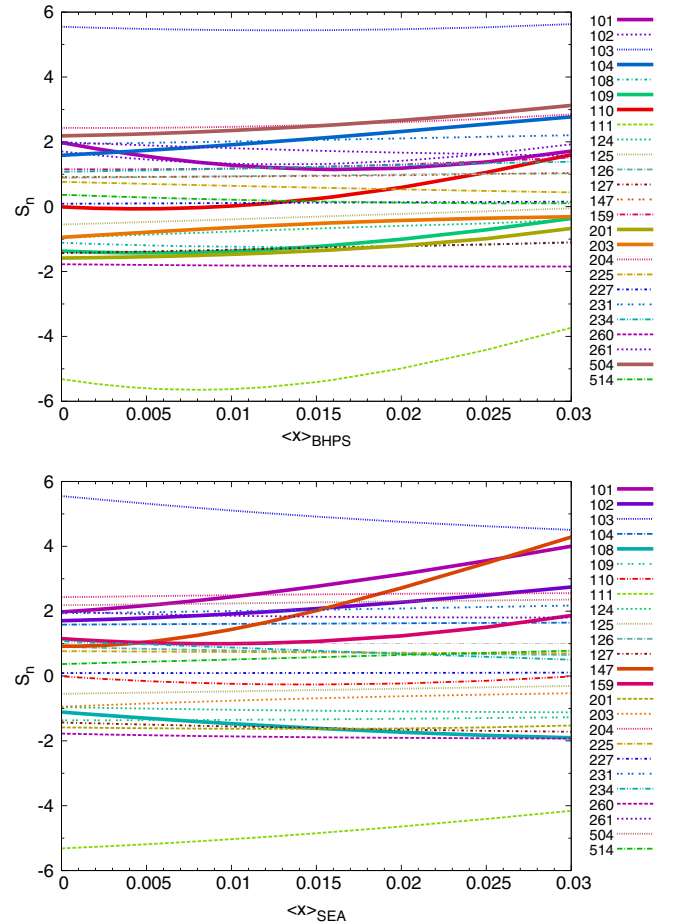

FIG. 10 (color online). Comparing S_n as a function of charm momentum fraction, $\langle x \rangle_{IC}$, for two models of intrinsic charm. Upper: BHPS model; lower: SEA model.

Figure 10 shows S_n for all 25 individual data sets, for the two models of IC, as a function of $\langle x \rangle_{IC}$. The BHPS model is shown above, and the SEA model is shown below. To focus on the data sets which are most affected by intrinsic charm, we select just those in Fig. 11.

For the BHPS model, we see that several of the data sets have a slow increase in S_n with momentum fraction $\langle x \rangle_{IC}$, while data set 101 (BCDMS F_2^p) has a significant decrease in S_n for small $\langle x \rangle_{IC}$, and then levels off for larger $\langle x \rangle_{IC}$. This is consistent with the results of Fig. 1 which showed that the BHPS models for $\langle x \rangle_{IC} \lesssim 0.015$ were actually slightly favored over zero intrinsic charm, but increasing $\langle x \rangle_{IC}$ further decreased the overall goodness of fit for the BHPS model. The increase in χ_F^2 or $\chi_F^2 + T_2$ seen in Figs. 1 or 2 at $\langle x \rangle_{IC} \lesssim 0.025$ for the BHPS model can be attributed to several experiments: ID 104 (NMC F_2^d/F_2^p), ID 110 (CCFR F_2^p), and ID 504 (CDF Run-2 inclusive jets); note that S_n increases for these experiments in Fig. 11. The abrupt increase of T_2 at $\langle x \rangle_{IC} \gtrsim 0.025$ comes from experiment ID 110.

In contrast, the SEA model shows a strong increase in S_n as $\langle x \rangle_{IC}$ increases for several experiments, in particular
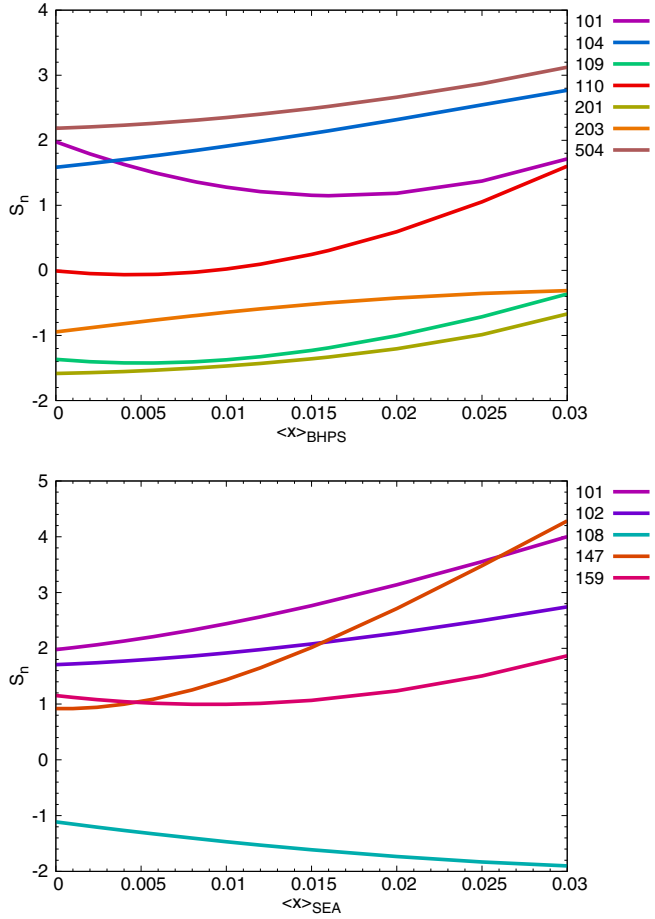
FIG. 11 (color online).   Comparing S_n as a function of charm momentum fraction, $\langle x \rangle_{IC}$, showing the data sets for which S_n changes the most significantly. Upper: BHPS model; lower: SEA model.

for the HERA combined charm experiment ID 147. (The CDHSW measurement of F_2^p, experiment ID 108, shows a strong decrease in S_n; but we note that that experiment is already anomalously well fit, even for no intrinsic charm.) Again, this is consistent with the rapid rise of χ^2 versus $\langle x \rangle_{IC}$ in Fig. 1 at $\langle x \rangle_{IC} \approx 0.015$; but Fig. 11 is more informative because it shows which experiments are responsible for the rapid rise in χ^2, i.e., which experiments conflict with the large IC. In fact, it is the HERA combined charm experiment 147 that gives the dominant contribution to the Tier-2 penalty for the SEA model, and determines the limit on charm momentum fraction from Fig. 2.

Overall, we can use the S_n plots to understand which data sets can or cannot be fit by a particular model of intrinsic charm. For instance, the BHPS model actually gives a *better* global fit for $\langle x \rangle_{IC} \sim 0.01$ because that significantly lowers S_n of the BCDMS F_2^p data set 101; at the same time it does not conflict with HERA inclusive charm data set 147. Evidently, intrinsic charm at large x at Q_c does not conflict with charm production at HERA (low x and large Q), while it can improve the agreement with the BCDMS data which are sensitive to quark distributions at high x values. On the other hand, any increase in the SEA model intrinsic charm (predominantly low x charm) worsens the fit to both data sets 101 and 147.

IV. PREDICTIONS FOR THE LHC

The inclusion of intrinsic charm changes the charm quark distributions at high Q values significantly, as well as affects the gluon and other quark PDFs through the PDF correlations in the global fit as shown in Sec. II.
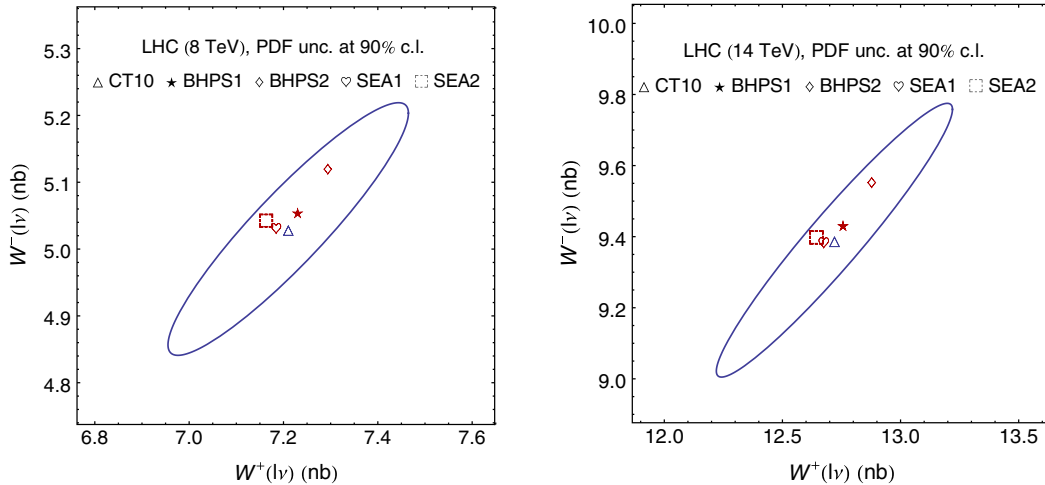

FIG. 12 (color online).   Correlation plot for the predictions of W^- and W^+ boson production cross sections at the LHC with $\sqrt{S} = 8$ TeV and 14 TeV. The tolerance ellipses are for 90% C.L. for the PDF uncertainties (unc.).
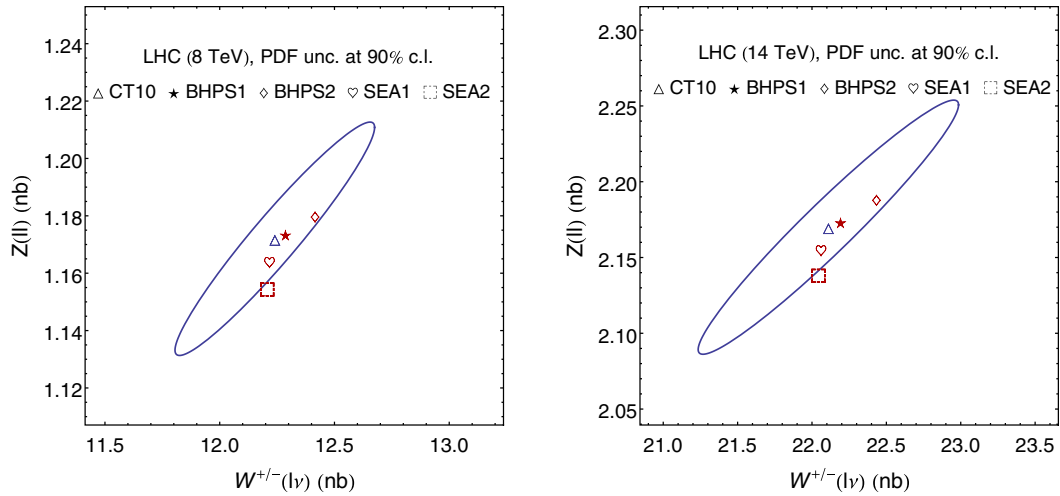
FIG. 13 (color online).   Correlation plot for the predictions of Z and $W^{\pm}$ boson production cross sections at the LHC with $\sqrt{S} = 8$ TeV and 14 TeV. The tolerance ellipses are for 90% C.L.

Thus, it may have impact on collider observables [47]. For example, in Figs. 12–15 we show the predictions of the NNLO total cross sections for W and Z boson production, Higgs boson production through gluon fusion, and top quark pair production at the LHC at $\sqrt{S} = 8$ and 14 TeV. The NNLO cross sections for W and Z boson production are computed with FEWZ2.1 [48,49]. The NNLO cross sections for Higgs boson and top quark pair production are obtained from iHix1.3 [50] and Top + +2.0 [51,52], with $m_h = 125$ GeV, $m_t = 173.3$ GeV, and the QCD scales set to the corresponding mass values. For each pair of total cross sections, we show the central predictions from CT10 with the 90% C.L. PDF tolerance ellipse as well as predictions from the four examples with intrinsic charm.

In Figs. 12–15, generally the predictions from the chosen IC models differ from the central predictions of CT10 by

less than 2%, which is smaller than the PDF uncertainties of CT10 but may not be negligible for precise predictions for the LHC. For the SEA models, the cross sections at the LHC are almost uniformly smaller, due to the reduction in momentum fraction remaining for the other noncharm partons. The only exception is W^- production through the process $\bar{c}s \to W^-$, which benefits from the increased $\bar{c}$ distribution. The production of Higgs bosons and top quark pairs, in particular, are reduced for nonzero IC SEA, due to the reduction in gluons and light sea quarks in the relevant regions of x and Q, as seen previously in Figs. 5 and 6.

For the BHPS models, the situation is more complicated, since the gluon and light sea quark PDFs are increased in the region $x \lesssim 0.1$ in the BHPS models. In this case the production of Z and $W^{\pm}$ (independent of sign) are increased and the production of top quark pairs is decreased
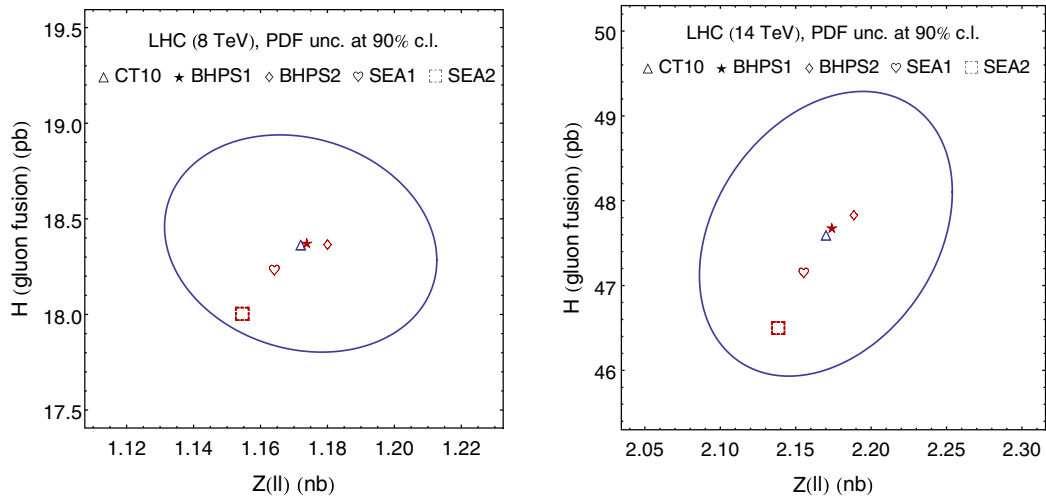

FIG. 14 (color online).   Correlation plot for the predictions of Higgs and Z boson production cross sections at the LHC with $\sqrt{S} = 8$ TeV and 14 TeV. The tolerance ellipses are for 90% C.L.
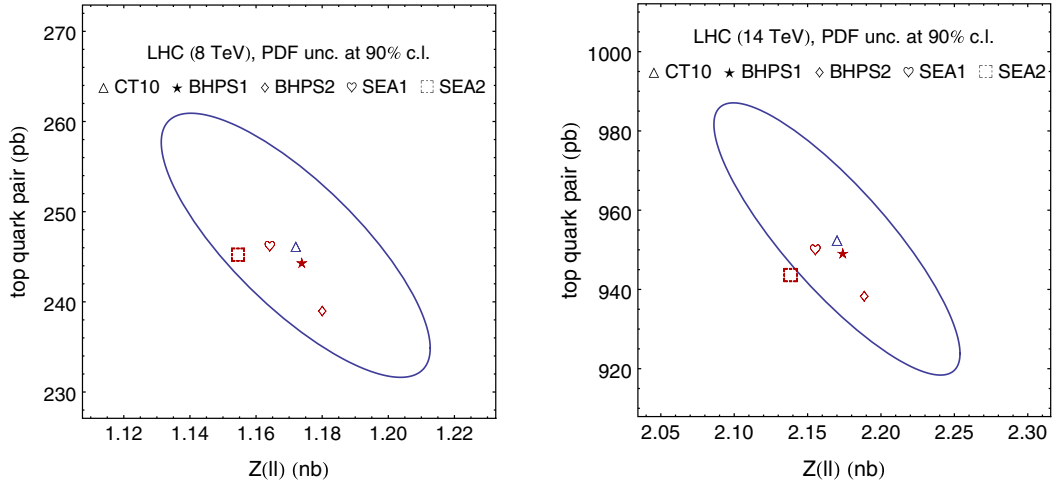
FIG. 15 (color online).   Correlation plot for the predictions of top quark pair and Z boson production cross sections at the LHC with $\sqrt{S} = 8$ TeV and 14 TeV. The tolerance ellipses are for 90% C.L.

by an increase of IC in the BHPS models. The production of Higgs bosons is fairly insensitive to the amount of IC for the BHPS model. Also, it is interesting to see from Figs. 12, 13, and 15 that the predictions from the BHPS models follow similar (anti) correlations as CT10 since they distribute along the diagonal direction of the ellipse, in distinction from the SEA models.

We can also check the effects of the IC on the rapidity distribution of the vector boson production. Charm quark contributions there have different shapes compared to the light quarks depending on the momentum profile of the charm quark of the IC models. Figs. 16–18 give the NNLO predictions of the rapidity distributions of $W^{\pm}$ and Z boson

productions from CT10 and the four IC models. They are calculated with the program Vrap0.9 [53], and both the renormalization and factorization scales are chosen to be the mass of the vector boson. The differences are small among four IC models, and their predictions are all within the CT10 PDF uncertainties. In Fig. 19 we further plot the ratios of the rapidity distribution, $d\sigma_{W^{\pm}}/d\sigma_{Z}$, which are presumably more sensitive to the charm quark contributions due to the cancellations of the uncertainties from light quark contributions. This is evident from comparing the size of the error band induced by the PDF uncertainty in Figs. 16–18 to that in Fig. 19. We can see that in the small or intermediate rapidity region the predictions of SEA2 lie
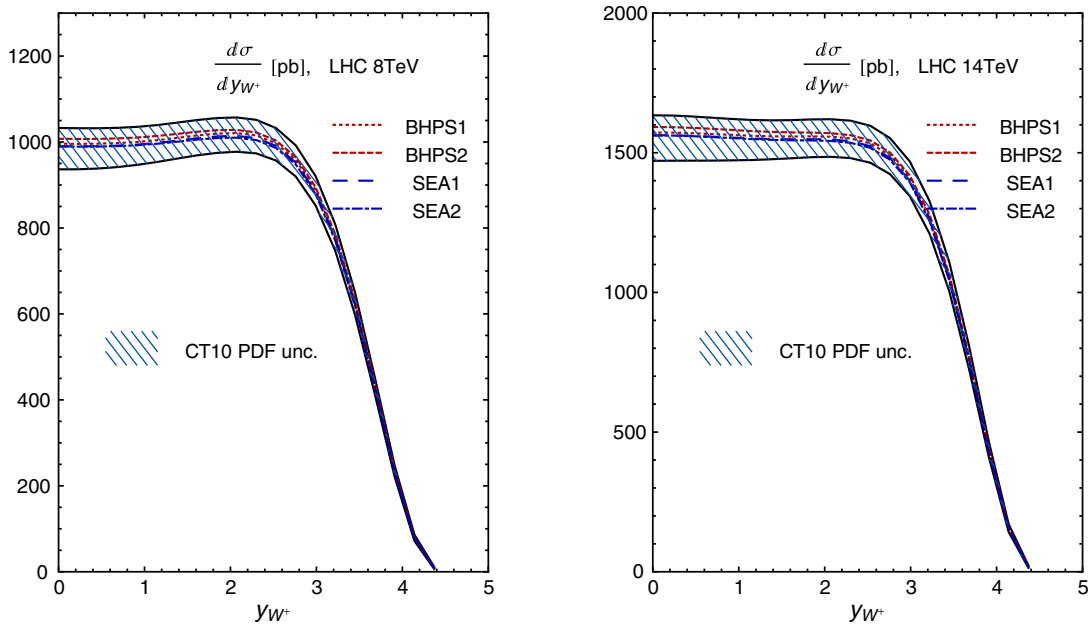

FIG. 16 (color online).   Rapidity distribution of the W^{+} boson at the LHC with $\sqrt{S} = 8$ TeV and 14 TeV. The PDF uncertainties are for 90% C.L.
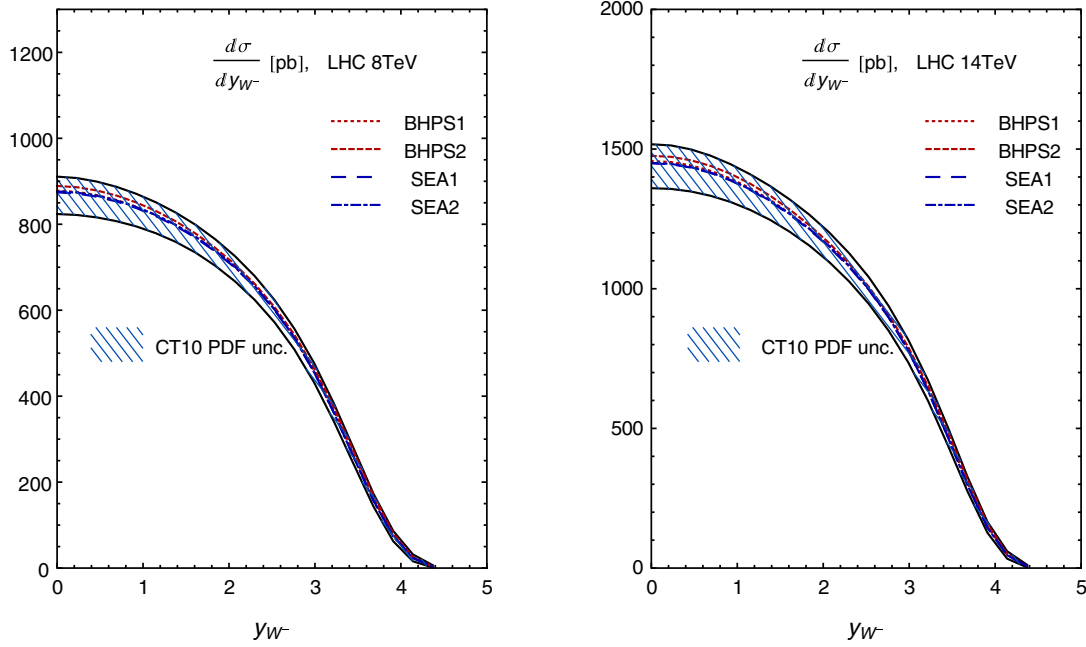
FIG. 17 (color online). Rapidity distribution of the W^- boson at the LHC with $\sqrt{S} = 8$ TeV and 14 TeV. The PDF uncertainties are for 90% C.L.

outdide the CT10 PDF uncertainties, similar to that for BHPS2 in the large rapidity region.

As already pointed out in our previous IC study [15], and recently in Refs. [54,55], the partonic process $g + c \to \gamma/Z + c$ is directly sensitive to the initial state charm distribution, although precision measurements could be experimentally challenging at the LHC. Figure 20 shows

predictions of the differential cross sections of an on-shell Z boson production in association with a charm quark at the LHC from CT10 and the IC models. The matrix element calculations of the process are only available up to NLO in QCD [56] and are implemented in program MCFM [57]. We simply convolute them with our NNLO PDFs in order to show the relative
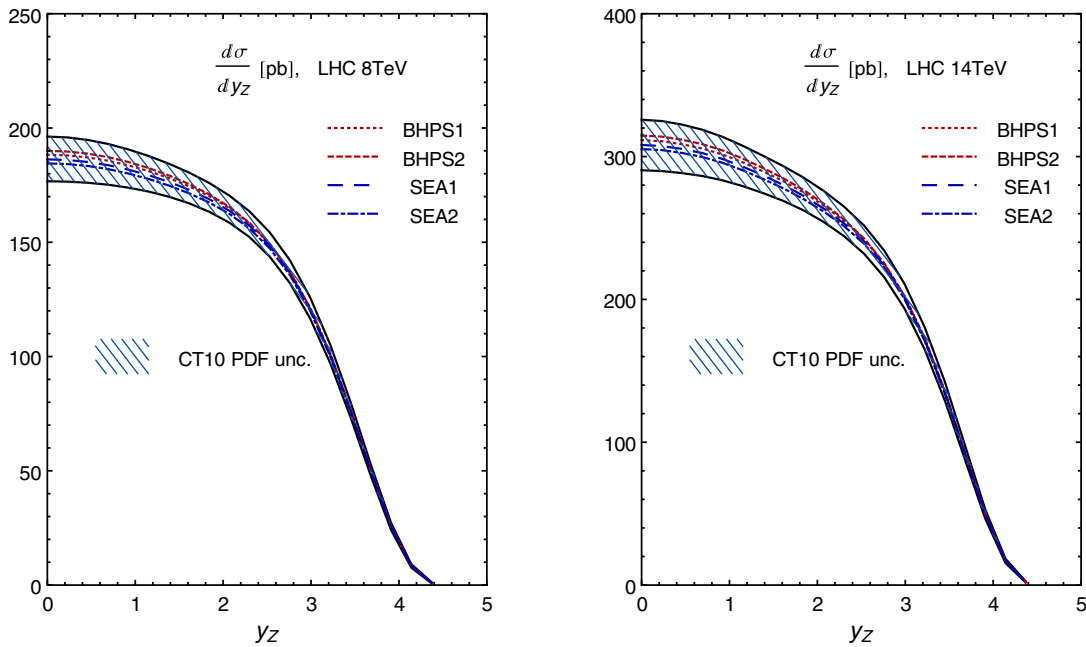

FIG. 18 (color online). Rapidity distribution of the Z boson at the LHC with $\sqrt{S} = 8$ TeV and 14 TeV. The PDF uncertainties are for 90% C.L.
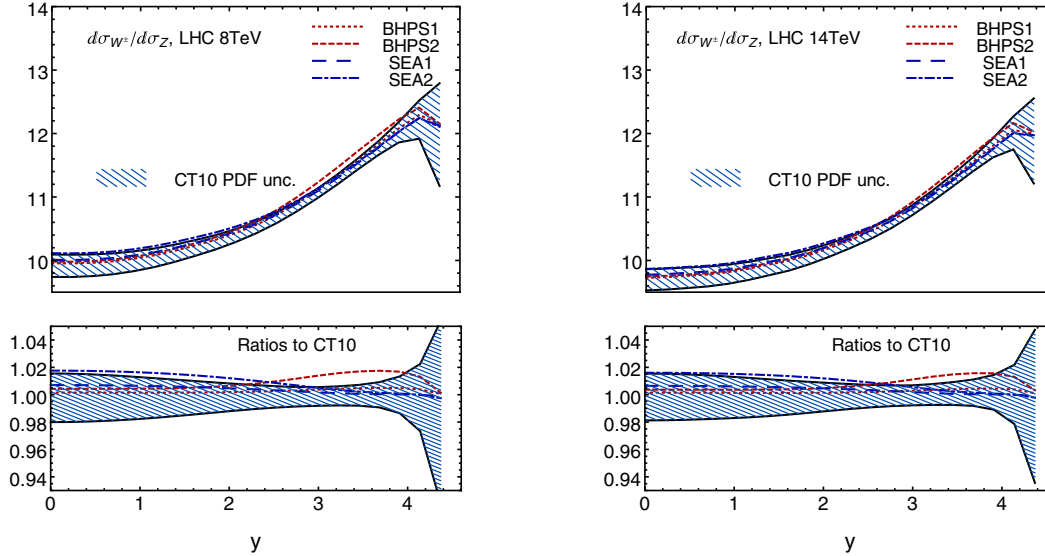
FIG. 19 (color online).   The ratio $\frac{d\sigma_{W^++W^-}}{d\sigma_Z}$ at the LHC with $\sqrt{S} = 8$ TeV and 14 TeV for four IC models at NNLO. Upper panel: differential distributions; lower panel: ratios normalized to the CT10 central prediction. The PDF uncertainties are for 90% C.L.
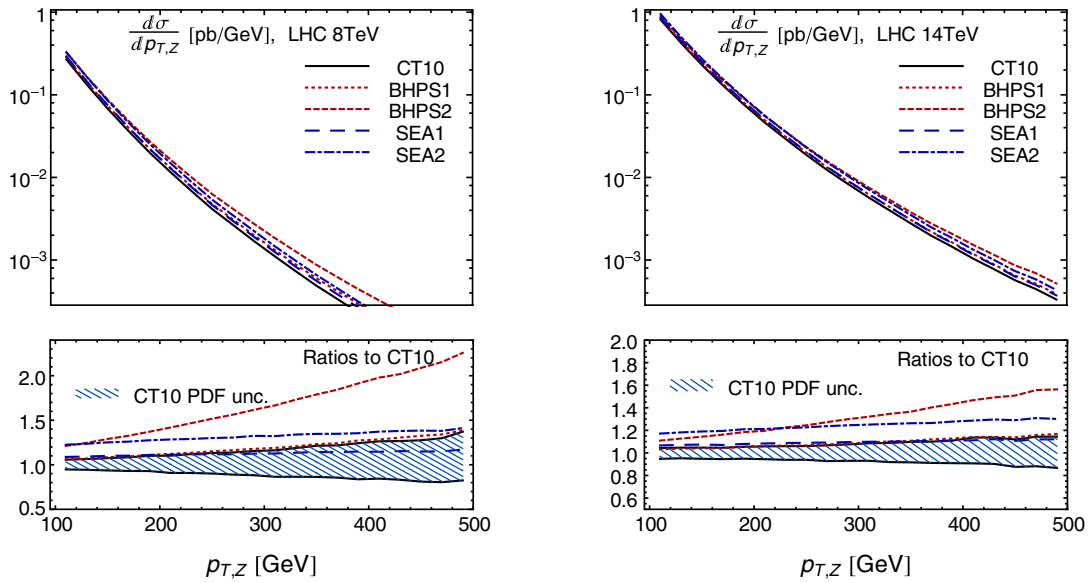

FIG. 20 (color online).   Transverse momentum distribution of the Z boson in the production of $pp \rightarrow Zc$ at the LHC with $\sqrt{S} = 8$ TeV and 14 TeV. Upper panel: differential cross sections; lower panel: ratios normalized to the CT10 central prediction. The PDF uncertainties are for 90% C.L.

changes of the cross sections in the presence of IC. Figure 20 gives the transverse momentum distribution of the Z boson with kinematic cuts of $p_T > 50$ GeV, and $|\eta| < 2.1$ applied on the charm quark jet using the anti-k_T algorithm with the radius parameter $R = 0.5$. In this calculation, both the renormalization and factorization scales are chosen to be the scale sum of the transverse momenta of final state particles. We have checked that in the large p_T region, the theoretical uncertainty induced by varying these scales simultaneously by a factor of 2

is much smaller than the error induced by the CT10 error PDFs.

We can clearly see that in the large $p_{T,Z}$ region, predictions from the fits with IC models deviate significantly from CT10 as a direct result of the charm distribution changes shown in Fig. 20. Thus, it shows a potential for discriminating the possible IC models. In particular, at large values of $p_{T,Z}$ it is very sensitive to the presence of IC in the BHPS models, since they have a large enhancement at large x. However, the cross sections are small, and

there are further suppressions from the decay branching ratios of the Z boson, as well as the charm quark jet-tagging efficiency. A detailed study of the feasibility of this process at the LHC is needed, but it is beyond the scope of the current work.

V. CORRELATIONS BETWEEN THE CHARM MASS AND INTRINSIC CHARM

The curves in Fig. 3 show that the charm distribution in the SEA model is very similar in shape to the charm distribution of CT10 with no IC. It mostly differs in overall normalization. This then begs the question whether it can be distinguished from some other physics that may produce more $c\bar{c}$ radiation, such as a change in the charm quark mass, m_c. In this section we investigate the correlation between m_c and IC in the SEA model.

A lighter charm quark mass could mimic the effects of intrinsic charm because, with a smaller mass, the charm PDF turns on sooner, resulting in more charm at a given value of Q. Conversely, a larger value of the charm mass would result in a smaller charm content at a given value of Q, which could then be made up with some intrinsic charm. This possibility is intriguing because the standard value of the charm mass $m_c^{\text{pole}} = 1.3$ GeV, which was used in the CT10 analysis [1] and was shown to be favored by the global analysis data [10], is smaller than the world average value given by the PDG [12], which was obtained using a mostly orthogonal set of data. The PDG gives a world average value of $m_c(m_c) = 1.275$ GeV, which translates into pole mass values of $m_c^{\text{pole}} = 1.46$ GeV and $m_c^{\text{pole}} = 1.67$ GeV when using the conversion formula in Eq. (17) of Ref. [11] at the one-loop and two-loop order, respectively. Thus, one may wonder if the lower value of m_c used in the global analysis is actually hiding some evidence for intrinsic charm.

Rather than performing a combined fit of m_c and IC, we shall just demonstrate the effect by rerunning the analysis of Sec. II, using the larger value of $m_c = 1.67$ GeV. We
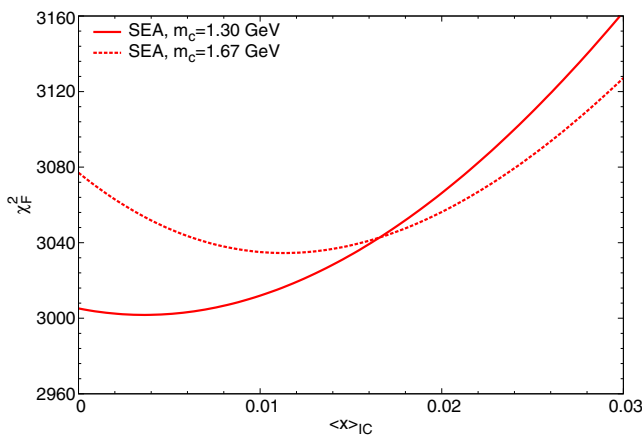
emphasize that we are not advocating this large value of the charm mass, but are just using it to observe the correlation between fits of the charm mass and intrinsic charm. In Fig. 21 we plot the results of the global fitting for χ_F^2 versus the intrinsic charm content, $\langle x \rangle_{\text{IC}}$, for the SEA model, both for our standard charm mass $m_c = 1.3$ GeV and for the larger value of $m_c = 1.67$ GeV. We keep the same initial scale for the light partons, $Q_0 = 1.295$ GeV, for both choices of m_c. As seen previously, the dependence of χ_F^2 on $\langle x \rangle_{\text{IC}}$ for $m_c = 1.3$ GeV is quite flat for small charm content and then begins to rise. The curve looks like a quadratic function with a minimum close to $\langle x \rangle_{\text{IC}} = 0$. This is consistent with the fact the $m_c = 1.3$ GeV is near the best fit of the charm mass with zero IC to the global analysis data. On the other hand, for $m_c = 1.67$ GeV, the χ_F^2 begins at a higher value for zero IC and then noticeably decreases to a minimum at around $\langle x \rangle_{\text{IC}} = 0.01$. Thus, if a larger value of the charm mass could be decisively shown to be required, then the global analysis would prefer nonzero IC, even for the SEA model, although this preference is not currently statistically significant.

We note, however, that the overall value of χ_F^2 is still worse for $m_c = 1.67$ GeV and $\langle x \rangle_{\text{IC}} = 0.01$ than for $m_c = 1.3$ GeV with zero intrinsic charm. We investigate this further in Fig. 22, where we compare S_n as a function of $\langle x \rangle_{\text{IC}}$ for the same set of experiments that were sensitive to the SEA model in Fig. 11 for $m_c = 1.3$ GeV. The main thing to note here is that the S_n for data set 147, the combined charm production at HERA, is significantly worse for $m_c = 1.67$ GeV than for $m_c = 1.3$ GeV. For $m_c = 1.67$ GeV and zero IC, S_n of data set 147 is more than 5, which indicates a very poor fit to the data. Increasing the IC content makes the fit better for this charm mass, but the S_n is still a good bit higher than that with $m_c = 1.3$ GeV and no IC. In addition, the fit to the combined HERA1 DIS data set 159 worsens as the charm
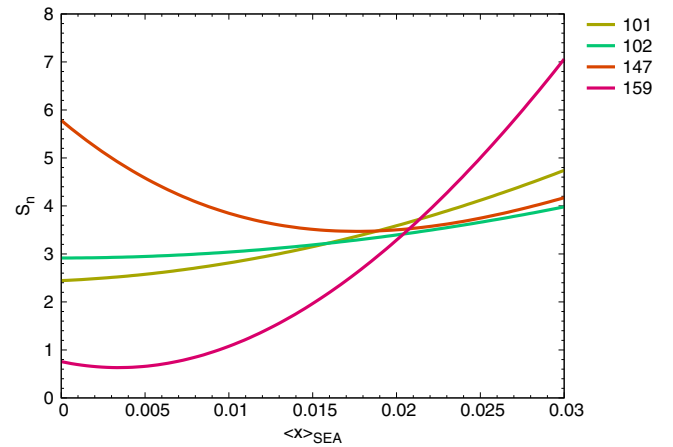

FIG. 21 (color online). The global chi-square function χ_F^2 versus the charm momentum fraction $\langle x \rangle_{\text{IC}}$ for the SEA model with $m_c = 1.3$ GeV and $m_c = 1.67$ GeV.

FIG. 22 (color online). Comparisons of S_n as a function of $\langle x \rangle_{\text{IC}}$, showing the data sets for which S_n changes the most significantly for the SEA model with $m_c = 1.67$ GeV.

momentum fraction increases for the SEA model with $m_c = 1.67$ GeV.

VI. CONCLUSIONS

The hypothesis of IC is a long-standing question in high-energy physics, combining QCD theory (both perturbative and nonperturbative) and phenomenology [18]. The purpose of this paper is to reassess the theory based on an up-to-date global analysis of QCD at the NNLO approximation.

The final conclusion of this work is that the wide range of short-distance processes that are commonly used in global analysis can be described accurately without an IC component of the proton; however, they do not rule out a small IC component. Quantitatively, we can construct PDFs that are acceptable fits to the global data (acceptable within the 90% C.L.) with $\langle x \rangle_{\text{IC}} \leq 1.5\%$ for a sea-like IC in the SEA model at $Q_c = 1.3$ GeV; or with $\langle x \rangle_{\text{IC}} \leq 2.5\%$ for a valence-like IC (i.e., concentrated at large x) in the BHPS model, cf. Fig. 2.

The HERA combined data on inclusive charm production in *ep* deep-inelastic scattering is particularly important. It constrains the SEA model more strongly than other data. On the other hand, it neither constrains nor favors the large-*x* IC of the BHPS model.

We also investigated the correlation between IC content and the value of the charm quark mass in the global fits. We found that a larger value of m_c was better fit with larger amounts of intrinsic charm, even for the SEA model. (The best fit for $m_c = 1.67$ GeV required about $\langle x \rangle_{\text{IC}} \approx 1\%$ for the SEA model.) However, the HERA combined data on inclusive charm production has a significant tension with larger values of the m_c, which is only somewhat alleviated by including nonzero IC. Thus, the overall best fit to the data with the SEA model still appears to be with the smaller charm quark mass $m_c = 1.3$ GeV and small or zero IC.

Predictions for LHC measurements, based on the IC PDFs, are interesting. The figures in Sec. IV show that the IC models with the largest allowed IC are right up against the uncertainty limits for the CT10 PDFs. In other words, the PDF *uncertainties* are just as large as the IC *effects*. Global analysis cannot say more about IC until the PDFs are more accurately determined. (Special interactions, especially sensitive to the charm quark but not used in global analysis, might be able to say more.) But more accurate measurements at the LHC will reduce the PDF uncertainties. The results of Sec. IV show that new LHC results will impact the limit on intrinsic charm. Or, if 2% intrinsic charm is *real*, then the results of Sec. IV indicate that it will show up as a discrepancy between theory and data in some common LHC measurements. For example, Fig. 14 shows that the BHPS models can strongly modify the $p_{T,Z}$ distribution in associated production of Z boson and charm jet.

Some recent research regarding intrinsic charm has focused on inclusive charm production at the LHC or Tevatron [14,58–60], or on associated production of charm with a prompt photon [61–65]. The IC PDFs constructed here can be used to make up-to-date predictions for those processes. This set of CT10 IC PDFs will be made available via an Internet web site.

APPENDIX: AN EQUIVALENT GAUSSIAN VARIABLE

We introduce the variable S_n to simplify the comparison of data and theory for the many data sets included in the global analysis. S_n is the equivalent *Gaussian* variable that matches the likelihood of the appropriate *chi-square* measure.

We could, of course, just use the individual values of χ^2/N_{pt} as measures of the goodness of fit for the different data sets, e.g., as in Table I. However, that variable has different meanings for different values of N_{pt}. For example, $\chi^2/N_{\text{pt}} = 11.0/10$ has a much different meaning than $\chi^2/N_{\text{pt}} = 1100.0/1000$, although they have the same value. The chi-square probability for $\chi^2/N_{\text{pt}} \geq 11.0/10$ is 0.358, while the chi-square probability for $\chi^2/N_{\text{pt}} \geq 1100.0/1000$ is only 0.015. Different data sets in the global analysis have very different numbers of data N_{pt}, so the ratio χ^2/N_{pt} alone is not sufficient to characterize the quality of the fit.

The variable S_n is designed to clarify the goodness of fit by transforming the cumulative probability from the chi-square distribution (which can be misleading because it depends on N_{pt}) to the normal distribution (which is more familiar).

The variable S_n is a function of χ^2 and N_{pt}. Let $P(\chi^2, N_{\text{pt}})$ denote the χ^2-probability distribution function for N_{pt} variables. Its cumulative distribution function (CDF) is

$$C(\chi^2, N_{\text{pt}}) = \int_0^{\chi^2} P(\xi, N_{\text{pt}}) d\xi. \tag{A1}$$

The definition of S_n is

$$C(\chi^2, N_{\text{pt}}) = \int_{-\infty}^{S_n} \frac{e^{-x^2/2} dx}{\sqrt{2\pi}}. \tag{A2}$$

We find that this variable is very helpful in judging the goodness of fit for the individual experiments in the global analysis; cf. Sec. III. The pure definition, as given above, is not very convenient for computation. Therefore, we use an accurate approximation for S_n [66],

$$S_n \approx L(\chi^2, N_{pt}) \qquad (A3)$$

$$L = \frac{(18N_{pt})^{3/2}}{18N_{pt} + 1} \left\{ \frac{6}{6 - \ln(\chi^2/N_{pt})} - \frac{9N_{pt}}{9N_{pt} - 1} \right\}. \qquad (A4)$$

Ideally, the variable S_n has an approximately Gaussian distribution with mean 0 and standard deviation 1. For a good fit to data, the value of S_n should be roughly between -1 and 1. A fit with $S_n > 3$ should be considered a poor fit. For a fit with $S_n < -3$, equivalent to $\chi^2 \ll N_{pt}$, we might assume that the actual systematic errors are smaller than the values used in the calculation of χ^2.

In reality, the variable S_n is unexpectedly large for some data sets, which are never fit very well in global analysis. Therefore, we do not judge the absolute values of S_n, but rather the relative values—relative to the best global fit— when comparing different sets of PDFs. For example, in Figs. 10 and 11 we assume it is not the absolute value of S_n for a chosen experiment that is relevant, but the change in the value as a function of $\langle x \rangle_{IC}$, which informs about the sensitivity of the experiment to the amount of intrinsic charm.

1. The Tier-2 penalty

When comparing the quality of agreement between theory and data, especially when comparing alternative PDFs to a central fit, we impose an additional *penalty*, the Tier-2 penalty, for large increases of S_n.

To compute the Tier-2 penalty for a given PDF and a given experiment, we first compute χ^2 for that experiment using the PDF under consideration. If $\chi^2_{BestFit} > N_{pt}$, we rescale this χ^2 value by a factor of $N_{pt}/\chi^2_{BestFit}$, where $\chi^2_{BestFit}$ is the χ^2 value for using the best-fit PDF. The value of S_n is then

defined as the point in a Gaussian distribution that has the same cumulative probability as the chi-squared distribution for the given (rescaled) χ^2 and N_{pt}. It is obvious that after rescaling the χ^2, $S_n = 0$ for the best fit with $\chi^2_{BestFit} > N_{pt}$. On the other hand, for any experiment for which the above ratio is larger than 1, i.e., for $\chi^2_{BestFit} < N_{pt}$, we do not rescale its χ^2 for calculating S_n. Hence, $S_n < 0$ for the best fit with $\chi^2_{BestFit} < N_{pt}$. Technically, we define the Tier-2 penalty for experiment i by $(S_n(i))^p \times \Theta(S_n)$, where the exponent p is taken to be 16 in this analysis, and the step function $\Theta(S_n)$ indicates that the penalty only applies to the case that $S_n > 0$.

There are two possible outcomes of the comparison for a given experiment. The first case is that after the rescaling, $S_n = 0$ for the best fit, so that the Tier-2 penalty is zero for the best fit, but rises immediately when the PDFs are moved in a direction that makes S_n increase. The second case is that $S_n < 0$ for the best fit, and the Tier-2 penalty remains to be zero for PDFs close to the best fit; but the penalty turns on once S_n becomes positive. By using $\chi^2 + T_2$ to judge the goodness of fit, we can exclude test PDFs that strongly violate some particular data set at the 90% C.L., and also exclude test PDFs for which two or more data sets come too close to the 90% C.L. limit.

In summary, the Tier-2 penalty is not used in finding the central fit, but it is used in constructing the error PDFs in the Hessian error analysis. It helps to avoid error PDFs that conflict (outside the 90% C.L.) with individual data sets. We also use the Tier-2 penalty when estimating uncertainties of alternative PDFs in the manner of Fig. 2. The simple requirement that the global $\Delta\chi^2$ must be less than some chosen "tolerance" is not adequate; we must also ensure that no individual experiment would definitively rule out the alternative PDFs by including the Tier-2 penalty in the global analysis. Furthermore, the rescaling procedure described above, for calculating the Tier-2 penalty, is particularly sensible in cases where a given experiment cannot be fit well by any choice of PDFs.

[1] J. Gao, M. Guzzi, J. Huston, H.-L. Lai, Z. Li, P. Nadolsky, J. Pumplin, D. Stump, and C.-P. Yuan, Phys. Rev. D **89,** 033009 (2014).

[2] S. Moch, J. A. M. Vermaseren, and A. Vogt, Nucl. Phys. **B688,** 101 (2004); A. Vogt, S. Moch, and J. A. M. Vermaseren, *ibid.* **B691,** 129 (2004).

[3] J. Sanchez Guillen, J. Miramontes, M. Miramontes, G. Parente, and O. A. Sampayo, Nucl. Phys. **B353,** 337 (1991); W. L. van Neerven and E. B. Zijlstra, Phys. Lett. B **272,** 127 (1991); E. B. Zijlstra and W. L. van Neerven, *ibid.* **273,** 476 (1991); Nucl. Phys. **B383,** 525 (1992); E. Laenen, S. Riemersma, J. Smith, and W. L. van Neerven, *ibid.* **B392,**

162 (1993); S. Riemersma, J. Smith, and W. L. van Neerven, Phys. Lett. B **347,** 143 (1995); M. Buza, Y. Matiounine, J. Smith, R. Migneron, and W. L. van Neerven, Nucl. Phys. **B472,** 611 (1996).

[4] C. Anastasiou, L. J. Dixon, K. Melnikov, and F. Petriello, Phys. Rev. Lett. **91,** 182002 (2003); C. Anastasiou, L. J. Dixon, K. Melnikov, and F. Petriello, Phys. Rev. D **69,** 094008 (2004).

[5] M. Aivazis, J. C. Collins, F. I. Olness, and W.-K. Tung, Phys. Rev. D **50,** 3102 (1994).

[6] M. Buza, Y. Matiounine, J. Smith, and W. van Neerven, Eur. Phys. J. C **1,** 301 (1998).

[7] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, Eur. Phys. J. C **70**, 51 (2010).

[8] R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, J. Rojo, and M. Ubiali, Nucl. Phys. **B849**, 296 (2011).

[9] S. Alekhin, J. Blmlein, K. Daum, K. Lipka, and S. Moch, Phys. Lett. B **720**, 172 (2013).

[10] J. Gao, M. Guzzi, and P. N. Nadolsky, arXiv:1304.3494v1.

[11] K. G. Chetyrkin, J. H. Kuhn, and M. Steinhauser, Comput. Phys. Commun. **133**, 43 (2000).

[12] J. Beringer *et al.* (Particle Data Group), Phys. Rev. D **86**, 010001 (2012).

[13] M. Buza, Y. Matiounine, J. Smith, R. Migneron, and W. L. van Neerven, Nucl. Phys. **B472**, 611 (1996); M. Buza, Y. Matiounine, J. Smith, and W. L. van Neerven, Eur. Phys. J. C **1**, 301 (1998).

[14] R. Aaij *et al.* (LHCb Collaboration), Nucl. Phys. **B871**, 1 (2013).

[15] J. Pumplin, H.-L. Lai, and W.-K. Tung, Phys. Rev. D **75**, 054029 (2007).

[16] F. Aaron *et al.* (H1 and ZEUS Collaboration), J. High Energy Phys. 01 (2010) 109.

[17] H1 and ZEUS Collaborations, Eur. Phys. J. C **73**, 2311 (2013).

[18] S. J. Brodsky, P. Hoyer, C. Peterson, and N. Sakai, Phys. Lett. **93B**, 451 (1980).

[19] G. P. Salam and J. Rojo, Comput. Phys. Commun. **180**, 120 (2009).

[20] M. Guzzi, P. M. Nadolsky, H.-L. Lai, and C.-P. Yuan, Phys. Rev. D **86**, 053005 (2012).

[21] W.-K. Tung, S. Kretzer, and C. Schmidt, J. Phys. G **28**, 983 (2002).

[22] A. C. Benvenuti *et al.* (BCDMS Collaboration), Phys. Lett. B **223**, 485 (1989).

[23] A. C. Benvenuti *et al.* (BCDMS Collaboration), Phys. Lett. B **237**, 592 (1990).

[24] M. Arneodo *et al.* (New Muon Collaboration), Nucl. Phys. **B483**, 3 (1997).

[25] J. P. Berge *et al.* (CDHSW Collaboration), Z. Phys. C **49**, 187 (1991).

[26] U.-K. Yang *et al.* (CCFR/NuTeV Collaboration), Phys. Rev. Lett. **86**, 2742 (2001).

[27] W. G. Seligman *et al.*, Phys. Rev. Lett. **79**, 1213 (1997).

[28] D. A. Mason, Ph.D. thesis, Oregon University [Institution Report No. FERMILAB-THESIS-2006-01,UMI-32-11223].

[29] M. Goncharov *et al.* (NuTeV Collaboration), Phys. Rev. D **64**, 112006 (2001).

[30] G. Moreno *et al.* (E605 Collaboration), Phys. Rev. D **43**, 2815 (1991).

[31] R. Towell *et al.* (FNAL E866/NuSea Collaboration), Phys. Rev. D **64**, 052002 (2001).

[32] J. Webb *et al.* (NuSea Collaboration), arXiv:hep-ex/0302019

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

[34] D. Acosta *et al.* (CDF Collaboration), Phys. Rev. D **71**, 051104 (2005).

[35] V. Abazov *et al.* (D0 Collaboration), Phys. Rev. Lett. **101**, 211801 (2008).

[36] V. Abazov *et al.* (D0 Collaboration), Phys. Rev. D **77**, 011106 (2008).

[37] V. Abazov *et al.* (D0 Collaboration), Phys. Lett. B **658**, 112 (2008).

[38] T. A. Aaltonen *et al.* (CDF Collaboration), Phys. Lett. B **692**, 232 (2010).

[39] T. Aaltonen *et al.* (CDF Collaboration), Phys. Rev. D **78**, 052006 (2008).

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

[41] J. J. Aubert *et al.* (European Muon Collaboration), Nucl. Phys. **B213**, 31 (1983).

[42] B. W. Harris, J. Smith, and R. Vogt, Nucl. Phys. **B461**, 181 (1996).

[43] F. M. Steffens, W. Melnitchouk, and A. W. Thomas, Eur. Phys. J. C **11**, 673 (1999).

[44] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne, Eur. Phys. J. C **4**, 463 (1998).

[45] M. Gluck, E. Reya, and A. Vogt, Eur. Phys. J. C **5**, 461 (1998).

[46] T. J. Hobbs, J. T. Londergan, and W. Melnitchouk, arXiv:1311.1578v1.

[47] P. M. Nadolsky, H.-L. Lai, Q.-H. Cao, J. Huston, J. Pumplin, D. Stump, W.-K. Tung, and C.-P. Yuan, Phys. Rev. D **78**, 013004 (2008).

[48] R. Gavin, Y. Li, F. Petriello, and S. Quackenbush, Comput. Phys. Commun. **182**, 2388 (2011).

[49] R. Gavin, Y. Li, F. Petriello, and S. Quackenbush, Comput. Phys. Commun. **184**, 208 (2013).

[50] C. Anastasiou, S. Buehler, F. Herzog, and A. Lazopoulos, J. High Energy Phys. 12 (2011) 058.

[51] P. Baernreuther, M. Czakon, and A. Mitov, Phys. Rev. Lett. **109**, 132001 (2012).

[52] M. Czakon, P. Fiedler, and A. Mitov, Phys. Rev. Lett. **110**, 252004 (2013).

[53] C. Anastasiou, L. J. Dixon, K. Melnikov, and F. Petriello, Phys. Rev. D **69**, 094008 (2004).

[54] W. J. Stirling and E. Vryonidou, Phys. Rev. Lett. **109**, 082002 (2012).

[55] V. A. Bednyakov, M. A. Demichev, G. I. Lykasov, T. Stavreva, and M. Stockton, arXiv:1305.3548.

[56] J. M. Campbell, R. K. Ellis, F. Maltoni, and S. Willenbrock, Phys. Rev. D **69**, 074021 (2004).

[57] J. M. Campbell and R. K. Ellis, Nucl. Phys. B, Proc. Suppl. **205–206**, 10 (2010).

[58] B. A. Kniehl, G. Kramer, I. Schienbein, and H. Spiesberger, Phys. Rev. D **79**, 094009 (2009); Eur. Phys. J. C **72**, 2082 (2012).

[59] E. V. Bugaev and P. A. Klimai, J. Phys. USSR **37**, 055004 (2010).

[60] G. I. Lykasov, V. A. Bednyakov, A. F. Pikelner, and N. I. Zimin, Europhys. Lett. **99**, 21002 (2012).

[61] V. M. Abazov *et al.* (D0 Collaboration), Phys. Rev. Lett. **102**, 192002 (2009); (Phys.Lett. B **719**, 354 (2013).

[62] T. P. Stavreva and J. F. Owens, Phys. Rev. D **79**, 054017 (2009).

[63] W. J. Stirling and E. Vryonidou, Phys. Rev. Lett. **109**, 082002 (2012).

[64] A. V. Lipatov, M. A. Malyshev, and N. P. Zotov, J. High Energy Phys. 05 (2012) 104.

[65] V. A. Bednyakov, M. A. Demichev, G. I. Lykasov, T. Stavreva, and M. Stockton, EPJ Web Conference, **60**, 20047 (2013).

[66] Toby Lewis, Australian Journal of statistics **30A**, 160 (1988).