New method for reducing parton distribution function uncertainties in the high-mass Drell-Yan spectrum

C. G. Willis, R. Brock,[*] D. Hayden, T.-J. Hou, J. Isaacson,[†] C. Schmidt, and C.-P. Yuan

Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48823, USA

Uncertainties in the parametrization of parton distribution functions (PDFs) are becoming a serious limiting systematic uncertainty in Large Hadron Collider (LHC) searches for beyond the standard model physics. This is especially true for measurements at high scales induced by quark and anti-quark collisions, where Drell-Yan continuum backgrounds are dominant. Tools are recently available which enable exploration of PDF fitting strategies and emulate the effects of new data in a future global fit. ePUMP is such a tool and it is shown that judicious selection of measurable kinematical quantities can reduce the assigned systematic PDF uncertainties by significant factors. This will be made possible by the huge statistical precision of future LHC standard model datasets.

I. INTRODUCTION

Beyond the standard model (BSM) physics at the Large Hadron Collider (LHC) would be found as deviations from standard model (SM) expectations, possibly in rates, but more typically in the kinematic distributions of final state objects or their combinations—of jets, leptons, and missing energy. Therefore the importance of accurately and precisely modeling SM physics cannot be overstated. While the electroweak properties of the SM are very precisely known, precision knowledge of parton distributions functions (PDFs) is becoming a limiting factor for many BSM searches. This limitation comes from the theoretical uncertainties becoming so large at high-mass that a clear deviation from the SM becomes hard to distinguish, and even upon discovery of new physics the characterization of this signal among various different theoretical models would be blurred.

As PDFs are not analytically calculable in the framework of perturbative quantum chromodynamics (QCD), their shapes must be modeled by globally fitting measured distributions from many combinations of varied experimental data. Most of these data come from legacy experiments, such as deep inelastic scattering (DIS) experiments, various fixed target hadron experiments, and the Fermilab

Tevatron. LHC experimental results are beginning to be used in global PDF fits, and in the coming decades new knowledge of PDFs will come from measurements at ATLAS [1], CMS [2], and LHCb [3]. We suggest that new strategies are worth exploring and we present one here.

Constraining PDFs and their uncertainties is now an intense research program. The systematic uncertainty in the PDF models arises from the (1) experimental uncertainties of the input data used in a global fit, (2) any theoretical assumptions made by the fitting groups, and/or (3) the chosen parametrizations characterizing the functional forms of the PDFs themselves. All of the global PDF fitting groups (CTEQ-TEA [4], MMHT [5], and NNPDF [6]) characterize their fits with Hessian error matrices or Monte Carlo replicas so that experiments can legitimately include PDF uncertainties as a component to any theoretical error for any measurement or limit.

In this paper we explore the PDF uncertainties as they apply to the BSM search for a resonant Z' gauge boson in the dilepton invariant mass spectrum. The dominant and irreducible background process to this search is the Drell-Yan (DY) process. Both ATLAS [7] and CMS [8] have recently completed their searches for new high-mass phenomena from the first $\sqrt{s} = 13$ TeV data-taking runs at the LHC. Both set comparable lower bounds on the mass of a hypothetical new vector boson and both publish extensive lists of their systematic uncertainties, including uncertainties attributed to our limited knowledge of PDF fitting.

To date, only 5% of the planned LHC data are in hand and yet these PDF uncertainties might already have limited future mass reaches for such searches. Not only are resonant Z' boson searches "at risk" but also W' boson searches and especially nonresonant (such as contact interactions) searches, which are very sensitive to sloped

---

[*]brock@pa.msu.edu
[†]Present Address: Theory Division, Fermi National Accelerator Laboratory, Batavia, Illinois 60510-5011, USA.

shape changes in the background. Furthermore, as we enter the new high integrated luminosity era of the LHC, experimental uncertainties will naturally be continually reduced, meaning that searches with even more complicated final states will eventually start to become limited predominantly by theoretical uncertainties. Therefore, it is critical that we improve our understanding of PDFs and their associated uncertainties.

A. Our strategy

Experiments utilize PDF fits which are global and agnostic respecting a basic principle of the parton model: PDF sets and uncertainties originate from all data and are applicable to all scattering. But knowledge of the PDFs is not uniform nor are all reactions similarly dependent on them. For example, DY production is less sensitive to knowledge of the gluon PDF than many BSM searches. Instead, precision predictions of DY processes depend significantly on knowledge of both the valence and sea quark densities which largely come from deep inelastic scattering and DY experiments. And to that end, hadron collider DY experimental inputs have been a part of PDF global fitting for years. For example, the CT14NNLO [4] fits utilized inputs from the W and Z boson charge asymmetry measurements from the Tevatron: [9–11] from CDF and [12,13] results from DØ.

And for the first time, in CT14NNLO the CTEQ-TEA group included LHC data from W/Z cross sections and the charged lepton asymmetry measurement from ATLAS [14], the charged lepton asymmetry in the electron [15] and muon decay channels [16] from CMS, and the W/Z lepton rapidity distributions and charged lepton asymmetry from LHCb [17]. But we will show that modern PDF global fits are not as potent for quark densities as are necessary for future precision measurements.

The only remedy to this problem is the addition of qualitatively new experimental inputs to global fitting, but the LHC is currently the only PDF "game in town." We propose a way to judiciously use LHC DY data itself as inputs to global fitting. The strategy would be to add Z boson peak and DY continuum data to global fitting from a well-measured, low-to-moderate invariant mass control region ($M < 1$ TeV). The resulting, "boutique" PDF sets could be used in an unbiased way to constrain the theoretical uncertainties in a kinematic search region relevant to modern BSM particle hunt, which is now in the $M > 5$ TeV region.

We further show that the DY kinematics can be exploited to enhance the impact on LHC DY data, namely emphasizing well-understood up-quark densities and deemphasize always limited sea-quark densities. This would require inputs which are differential in nature and not just asymmetry results near the Z boson peak.

The machinery of PDF global fitting groups is very complex and for physicists outside of the PDF groups,

testing new PDF analysis strategies can be cumbersome. This will change with the recent development of tools like EPUMP [18] (the Error PDF Updating Method Package, see Appendix A and [18] for details), which makes it possible to explore the effects of new kinematic inputs to a global fit without requiring a full global analysis. EPUMP is not a substitute for full global fitting, but can be used as a tool to probe the effects of new data.

In essence one can consider EPUMP to be a simulation of global fitting in an approximation described in Appendix A. Pseudodata can be added to an existing global fit in order to explore how that data might affect the central value and importantly, the uncertainties in the resulting candidate PDFs. All of the sum rules, QCD evolution, and uncertainties inherent in the "parent" global fit to which test data are added are preserved. While other PDF profiling tools exist such as xFitter [19], in this paper we choose to use EPUMP which has been thoroughly tested [18] against the CT14NNLO [4] global fits.

The work in this paper is the first published use of EPUMP. We demonstrate that new insight into kinematics of the DY process has emerged, and that considerable reduction in the quark and anti-quark PDF uncertainties is possible with new data inputs to PDF global fitting.

B. Our goals

Our goals in this paper are limited. We simply ask the optimistic questions: can qualitatively new data when combined with the current inputs of CT14HERA2 reduce future PDF uncertainties and if so, by how much? And would any reduction improve the overall precision of high mass DY backgrounds relevant to future Z' searches? We exploit the unprecedented statistical power of future LHC running and use DY kinematically motivated differential distributions to suggest that sensitivities to partons of special interest in DY production can be enhanced.

Our *ansatz* is to treat BSM DY searches as consisting of a *control region*—from which we envision mining DY data for global fitting—and a *signal region* to where those new global fits are extrapolated. Of course as in any control-signal region analysis, the assumption is that the control region contains only SM physics. We specifically explore the possibility that LHC DY data in a safe control region might be useful to further constrain PDFs appropriate to high-mass BSM searches for which the continuum DY is the dominant background. Having determined that this is worth consideration, our ultimate proposal is that the LHC experiments and the PDF fitting teams work together to explore inclusion of LHC DY data into global fitting when prepared in a particularly useful way.

We chose to do our work using the most recent CTEQ PDF global fit, namely CT14HERA2. This includes the most recent HERA1 and HERA2 data and utilizes an updated parametrization from the previous CT14NNLO sets. Since the recent experimental ATLAS publication [7] limits

were set using the CT14NNLO sets, we do make a brief comparison to show that the basic PDFs are very similar.

Our goals are limited to asking and answering our two questions above. To that basic end, what we do not do here are the following:

(i) An important part of the theoretical uncertainties include exploration of the parametrization assumed and potentially additional parametrization choices. While exploring functional choices would be an interesting exercise when attempting to extrapolate into a new kinematical regime, we do not do that here.

(ii) We do not attempt to optimize theoretical uncertainties associated with any other theoretical considerations like the strong coupling constant, electroweak couplings, or higher order electroweak and QCD effects.

(iii) We also make no effort to optimize or explore the full set of possible experimental uncertainties.

The paper is structured as follows. First, in Sec. II the current experimental results are briefly reviewed with an emphasis on the systematic uncertainties. Next, we review the kinematics of the DY process in Sec. III. Particular attention is paid to the role of the Collins-Soper (CS) angle (θ^*) [20], as this variable will be shown to possess hitherto unemphasized discrimination power between up- and down-type quark flavors which varies as a function of the invariant mass of the lepton pairs. Bearing this in mind, we then propose a new strategy for future PDF global fitting inspired by the use of EPUMP in Sec. IV. Then, in Sec. V, the results of such a strategy are assessed, first on the high-x behavior of the post-fit CT14HERA2 PDFs, and then on the expected event yields of the high-mass dilepton spectrum. Finally, concluding remarks are given in Sec. VI. Appendix A describes EPUMP in more detail while Appendix B provides some kinematical explanatory details.

II. CURRENT Z' BOSON SEARCH RESULTS

Both direct and indirect searches for Z' bosons have been conducted at several previous hadron collider experiments. Early results were obtained from the DØ [21] and CDF [22] experiments at the Tevatron, and more recently, the ATLAS [7,23,24] and CMS [8,25,26] experiments at the LHC. Table I shows the most recent results from the ATLAS [7] and CMS [8] collaborations.

As the highest energy particle collider, the LHC experiments' ability to set Z' limits is vastly improved compared to what was achievable at LEP and the Tevatron. The most stringent direct limits come from the ATLAS and CMS experiments where searches have been conducted at $\sqrt{s} = 7$, 8 and 13 TeV, with varying amounts of integrated luminosity. The most recent results for the combined electron and muon pair invariant mass are shown in Fig. 1 for these two experiments. These searches usually consider two types of Z' models. The first model considered is a simple U(1) gauge extension called the sequential standard model (SSM) [27], where the coupling of the new gauge boson to SM particles is the same as the Z boson. The second model considered is called the E_6 model [28], and gives rise to the additional gauge boson through the decomposition of the E_6 grand unified theory gauge group. This can lead to a variety of different Z' scenarios and coupling to SM particles, depending on the mixings of two states. Of these possible scenarios the Z'_χ has the widest width, and the Z'_ψ has the narrowest, leading to them often being used as two benchmarks to test both extremes of this model.

Of particular interest are the systematic errors due to PDF fitting uncertainties. The two LHC experiments treat these quite differently. Table II illustrates their assignments from CMS ([8]) and ATLAS ([7]). The PDF uncertainties for electron and muon pair backgrounds are shown for each experiment as are the total experimental uncertainties as quoted from each paper. The ATLAS experiment further assigns a "PDF Choice" uncertainty in accordance with the PDF4LHC forum [29] to account for differences among the PDF fitting groups' predictions as excursions from the nominal choice and its full error matrix.

The two experiments report different assignments for PDF uncertainties. For example ATLAS assigns large uncertainties derived from a detailed treatment. For the di-electron channel the reported overall uncertainty is 26.3% which comes from: the combined PDF (variation plus choice) uncertainties of 20.8%, other non-PDF theory uncertainties of 10%, and total experimental uncertainties of 12.8%. Di-muon uncertainties are not as large, but for both measurements the PDF uncertainties compete unfavorably with the experimental uncertainties. CMS reports smaller PDF uncertainties and comparable experimental uncertainties.

TABLE I. Observed limits at 95% C.L. on the mass of a Z' boson from the most recent LHC experimental searches. The CMS analysis listed does not provide limits on the Z'_χ, which would otherwise be slightly higher than what was obtained for the Z'_ψ. The integrated luminosity for each analysis is rounded to the nearest whole number. The $\ell\ell$ channel refers searches that combine individual electron and muon channels.

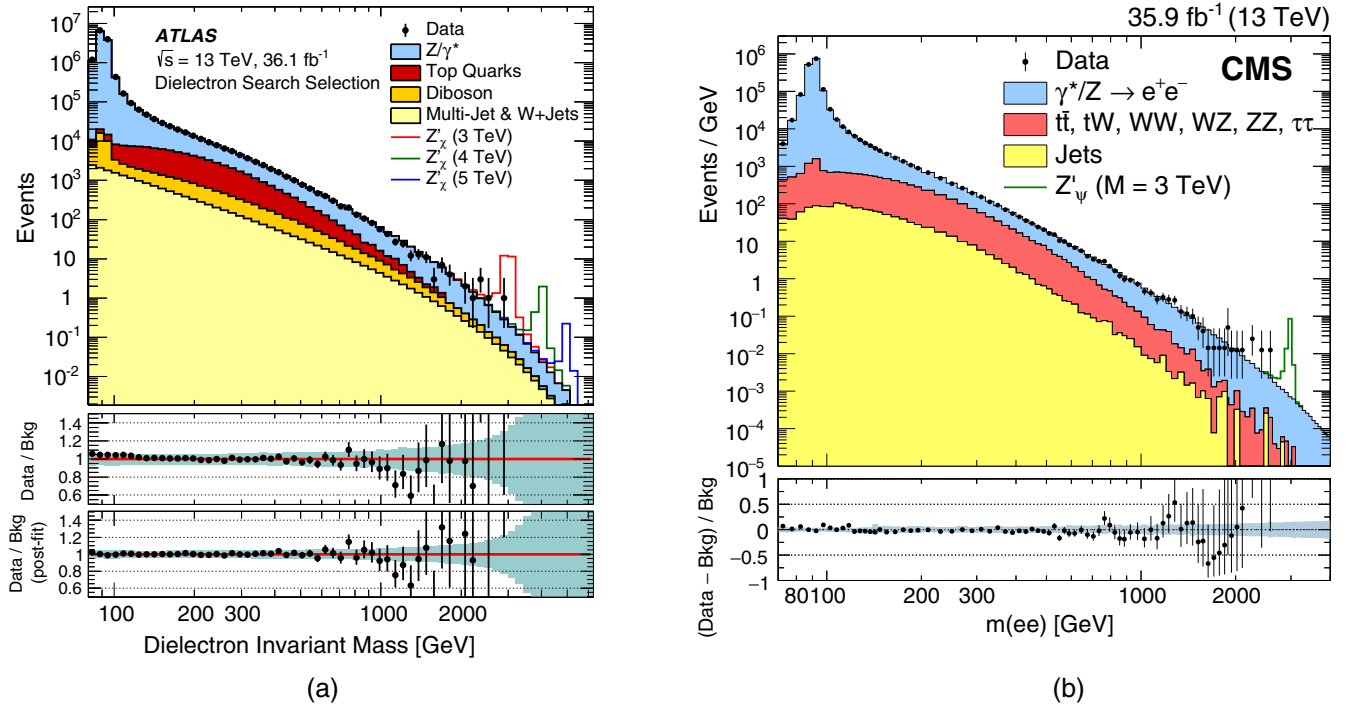| Collaboration | $\sqrt{s}$ [TeV] | $\mathcal{L}$ [fb^{-1}] | Channel | Lower limit on $M_{Z'}$ [TeV] | | | Reference |
				Z'_{SSM}	Z'_ψ	Z'_χ	
CMS	13	36	$\ell\ell$	4.5	3.9	$\cdots$	[8]
ATLAS	13	36	$\ell\ell$	4.5	3.8	4.1	[7]

FIG. 1.   Dielectron invariant mass search spectra in the (a) ATLAS [7] and CMS [8] dilepton analyses at the LHC.

Experimental systematic uncertainties will likely be reduced with more data, but the PDF uncertainties at this point are largely irreducible in the absence of new data of a qualitatively different sort (new DIS experiments?) or new ideas. We propose new ideas to address this using LHC data itself.

III. THE DRELL-YAN PROCESS

The general Drell-Yan process [30] of $pp \rightarrow \ell^+\ell^- + X$ at leading order originates from an s-channel exchange of an electroweak boson

$$q\bar{q} \rightarrow \gamma^*/Z \rightarrow \ell^+\ell^-. \tag{1}$$

Here, X denotes any additional final-state particles (radiated partons, the underlying event, multi-parton interactions, etc.). At next-to-leading order, the real corrections introduce three t-channel processes, listed in order of decreasing cross section at LHC energies,

$$qg \rightarrow \gamma^*/Z \rightarrow \ell^+\ell^- + q \tag{2}$$

$$\bar{q}g \rightarrow \gamma^*/Z \rightarrow \ell^+\ell^- + q \tag{3}$$

$$q\bar{q} \rightarrow \gamma^*/Z \rightarrow \ell^+\ell^- + g. \tag{4}$$

The leading order process is depicted in Fig. 2.

In each case, the vector boson decays into a pair of same-flavor, oppositely-charged leptons. For simplicity,

our discussion will center on the leading order process, but all of our results are based on next to leading order (NLO) plus next to leading log (NLL) calculations using the NLO-NLL RESBOS [32–34] package.

The DY triple-differential cross section can be represented as a function of the dilepton invariant mass $m_{\ell\ell}$, the dilepton rapidity $y_{\ell\ell}$, and the cosine of the lepton polar angle in the Collins-Soper rest frame $\cos\theta^*$. This was

TABLE II.   Published uncertainties due to the lack of PDF knowledge on the DY backgrounds for CMS and ATLAS for the 13 TeV LHC running. Each experiment determines PDF uncertainties from a common nominal PDF choice which is parenthetically indicated. The uncertainties have a scale dependence and can differ according to the dilepton channel. The results quoted here are evaluated at a mass of approximately 4 TeV, except the experimental results for CMS which are only quoted at 5 TeV in their paper. "PDF choice" for ATLAS results refers to the PDF4LHC forum recommendations [29]. "PDF variation" is the result from the full error matrix for the nominal PDF set.

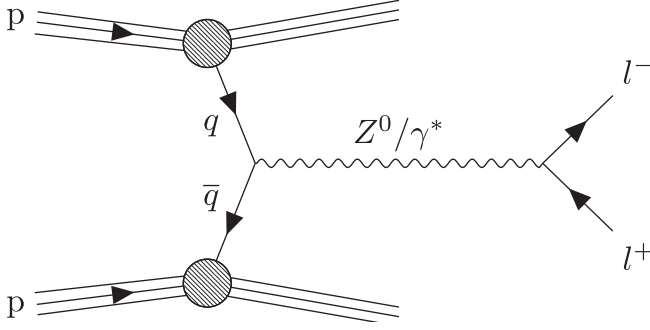| Systematic uncertainty	CMS (NNPDF2.3)		ATLAS (CT14NNLO)	
	ee [%]	$\mu\mu$ [%]	ee [%]	$\mu\mu$ [%]
PDF variation	$\cdots$	$\cdots$	19	13
PDF choice	$\cdots$	$\cdots$	8.4	1.9
PDF variation & choice	7	7	20.8	13.1
Combined experimental	12	15	12.8	18.9

FIG. 2.   A Feynman diagram [31] of the DY process initiated by a quark-antiquark pair as observed at the LHC.

measured by ATLAS [35] using data from the $\sqrt{s} = 8$ TeV LHC running for $46 < m_{\ell\ell} < 150$ GeV.

For the LO s-channel process, the DY triple-differential cross section can be written as

$$\frac{d^3\sigma}{dm_{\ell\ell}dy_{\ell\ell}d\cos\theta^*}$$
$$= \frac{\pi\alpha^2}{3m_{\ell\ell}s}\sum_q P_q[f_{q/P_1}(x_1, Q^2)f_{\bar{q}/P_2}(x_2, Q^2) + (q \leftrightarrow \bar{q})].$$

(5)

Here $\sqrt{s}$ is the center of mass energy of the LHC, and P_1 and P_2 are the 4-momenta of protons 1 and 2. In the standard fashion, x_1 and x_2 are the incoming parton momentum fractions such that $p_1 = x_1 P_1$ and $p_2 = x_2 P_2$. We take our notation from [35].

The functions $f_{q/P_1}(x_1, Q^2)$ and $f_{\bar{q}/P_2}(x_2, Q^2)$ are the PDFs for quark flavors q and $\bar{q}$, respectively. The term $(q \leftrightarrow \bar{q})$ accounts for the fact that either proton can carry a sea quark, as the LHC is a proton-proton collider.

Finally, the quantity P_q accounts for the parton-level dynamics in terms of important electroweak parameters, and exhibits dependencies on both dilepton mass and $\cos\theta^*$. Each factor in this formula matters in a high-mass extrapolation, and are discussed in detail in Appendix B 1.

The energy scale of the collision is set by the transferred four-momentum squared Q^2, which can be identified with the square of the dilepton invariant mass $m_{\ell\ell}^2$. Well-known kinematic definitions include

$$Q^2 = (p_1 + p_2)^2 = x_1 x_2 s,$$

(6)

and,

$$y_{\ell\ell} = \frac{1}{2}\ln\left(\frac{x_1}{x_2}\right),$$

(7)

which parametrizes the dilepton rapidity in terms of the x fractions of the initial-state partons at LO. From these, the variables are related, also at LO, by

$$x_1 = \frac{m_{\ell\ell}}{\sqrt{s}}e^{+y_{\ell\ell}}, \qquad x_2 = \frac{m_{\ell\ell}}{\sqrt{s}}e^{-y_{\ell\ell}}.$$

(8)

Equation (8) provides the first hint to the source of the large PDF uncertainty in high-mass DY production. The $\sqrt{s} = 13$ TeV LHC is now probing extremely large values of $m_{\ell\ell}$, beyond a few TeV. As such, a central dilepton event with an invariant mass of $m_{\ell\ell} = 3$ TeV and rapidity of $y_{\ell\ell} = 0$ requires x fractions beyond $x \simeq 0.2$. This is beginning to probe regions of sea and even valence quark momentum fractions which are not well constrained by mostly DIS inputs. Figure 3 shows quark, antiquark and gluon momentum fractions from the CT14HERA2 PDF set evaluated at two scales Q^2.
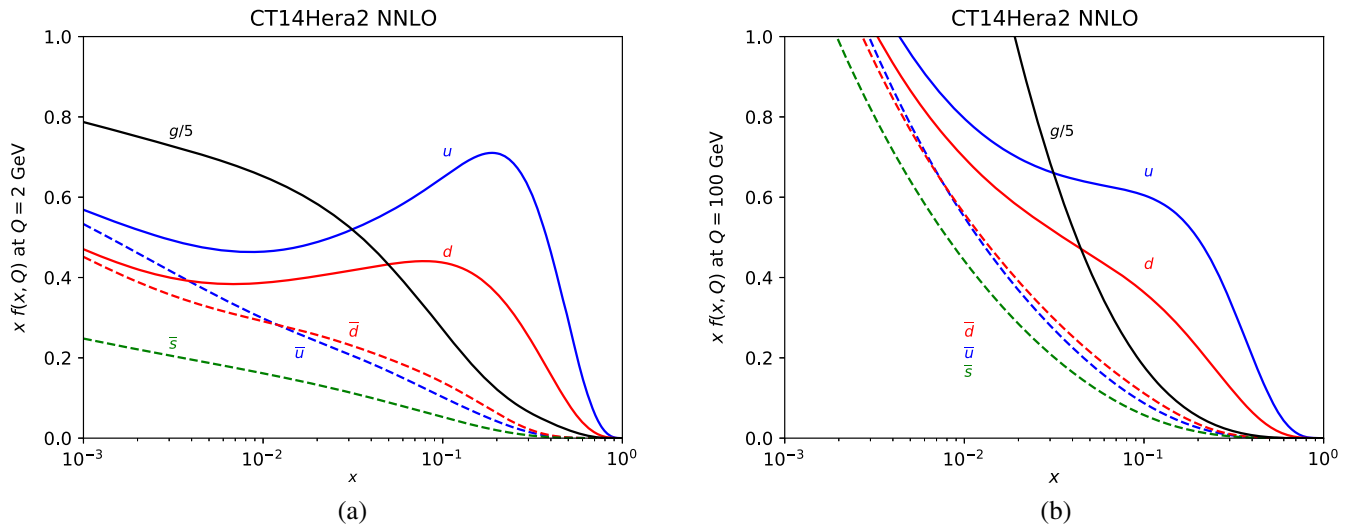

(a)

(b)

FIG. 3.   The CT14HERA2 PDFs of the CTEQ collaboration. Depicted are gluon, quark, and anti-quark PDFs as a function of x, evaluated at a scale of $Q = 2$ GeV (a) and $Q = 100$ GeV (b) [4].
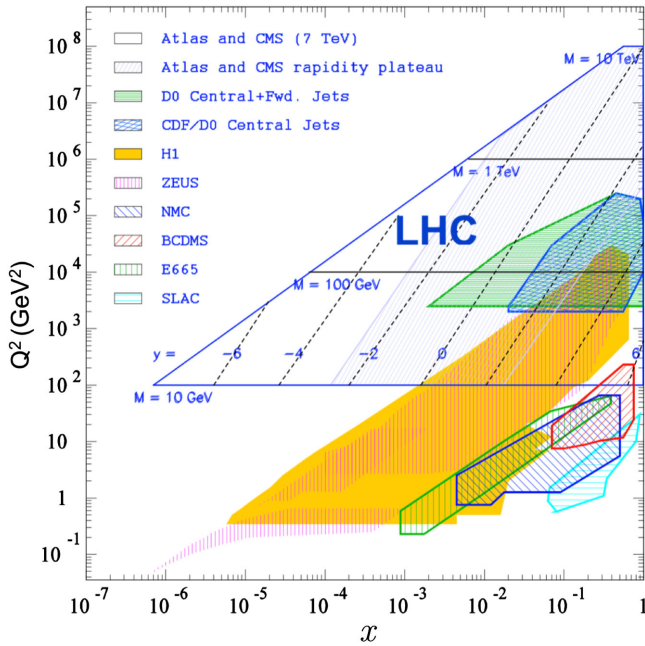
FIG. 4. The transferred momentum squared Q^2 vs the parton momentum fraction x at $\sqrt{s} = 7$ TeV. The regions probed by previous DIS, fixed-target, and collider-based experiments are labeled [36].

A. Behavior of PDFs at high parton x

The reason for this inherent highs-x uncertainty in the quark and anti-quark PDFs is due to the need to extrapolate experimental data—especially for quark and anti-quark fitting—as seen in Fig. 4. The only data which directly probe quark and antiquark PDFs for $x \gtrsim 0.2$ come from legacy deep-inelastic scattering experiments and HERA measurements. PDFs relevant for current and future LHC DY production scales of interest require an extrapolation of almost three orders of magnitude in mass and this proves difficult to do precisely with the current world data.

Figure 5 shows the PDF uncertainties for several individual parton flavors in the `CT14NNLO` and `CT14HERA2` PDF sets. It is not surprising that the $\bar{u}(x)$ and $\bar{d}(x)$ distributions are least precisely known at moderate-to-high x where input data are difficult to obtain and where their magnitudes have fallen to small fractions of their valence counterparts. But even the $u_V(x)$ and $d_V(x)$ distributions are poorly constrained in this region, although the up quark is much better determined than the down. Each of these distributions plays an important role in the initial-state quark-antiquark annihilation that results in the DY process. This significant lack of precision is the source of the large systematic uncertainties required in a high mass, dilepton Z' search. Figure 6 shows the iconic invariant mass distribution of dilepton pairs calculated using the RESBOS [32–34] Monte Carlo (MC) generator and the `CT14HERA2` PDF. The ratio band is the quoted `CT14HERA2` [37] PDF uncertainties of about 18% at $m_{\ell\ell} = 4$ TeV consistent with that quoted in the ATLAS result of 19% at $m_{ee} = 4$ TeV.

As DY data inputs are the only way to constrain high-x PDFs, a strategy is explored here that turns this lack of sensitivity into an opportunity. The DY continuum is well-measured and reliably SM physics. If PDF global fits were to include LHC DY data well below any search region, but high enough in invariant mass to better constrain the fits, this uncertainty could be reduced. Moreover, the amount of LHC data that will become available in the coming years will be staggering, so we have decided to explore DY kinematics further in hopes of finding/discovering sensitivities that would help to enhance the potential of high-x PDF fits.

We will show that there are DY observables, such as $\cos\theta^*$, that could in principle be incorporated in PDF global fitting, and the use of EPUMP tells us approximately how much reduction in PDF uncertainty is possible, as well as how much smaller the PDF systematic uncertainty might become in the DY differential mass spectrum. Due to the importance of $\cos\theta^*$, and the role it plays in our fitting strategy, a brief review is given in the next section.

B. The Collins-Soper polar angle

We have found particular power in $\cos\theta^*$ in Eq. (5). This angle is defined in the Collins-Soper (CS) [20] rest frame of the lepton-pair with the polar and azimuthal angles defined relative to the two proton directions. The z axis is defined in the Z boson rest frame so that it bisects the angle formed by the momentum of one of the incoming protons and the negative of the momentum of the other incoming proton. The y axis is constructed to be normal to the plane of the two proton momenta and the x axis which is chosen in order to create a right-handed Cartesian coordinate system.

The cosine of the polar angle θ^* defines the direction of the outgoing lepton ℓ^- relative to $\hat{z}$ in the CS frame and can be calculated directly from lab frame lepton quantities with

$$\cos\theta^* = \frac{P_z}{|P_z|} \frac{2(p_1^+ p_2^- - p_1^- p_2^+)}{M_{\ell\ell}\sqrt{M_{\ell\ell}^2 + P_T^2}}. \tag{9}$$

The sign of the z axis is defined on an event-by-event basis as the sign of the lepton pair momentum with respect to the z axis in the laboratory frame. Here, P_T and P_z are the transverse and longitudinal momentum of the dilepton system, respectively, and,

$$p_i^\pm = \frac{1}{\sqrt{2}}(E_i \pm p_{z,i}), \qquad i = 1, 2, \tag{10}$$

where the lepton (antilepton) energy and longitudinal momentum are E_1 and $p_{z,1}$ (E_2 and $p_{z,2}$), respectively. This definition requires the electric charge identification of each lepton. We define DY events as forward ($\cos\theta^* > 0$) or backward ($\cos\theta^* < 0$) according to the direction of the outgoing lepton in this frame of reference.
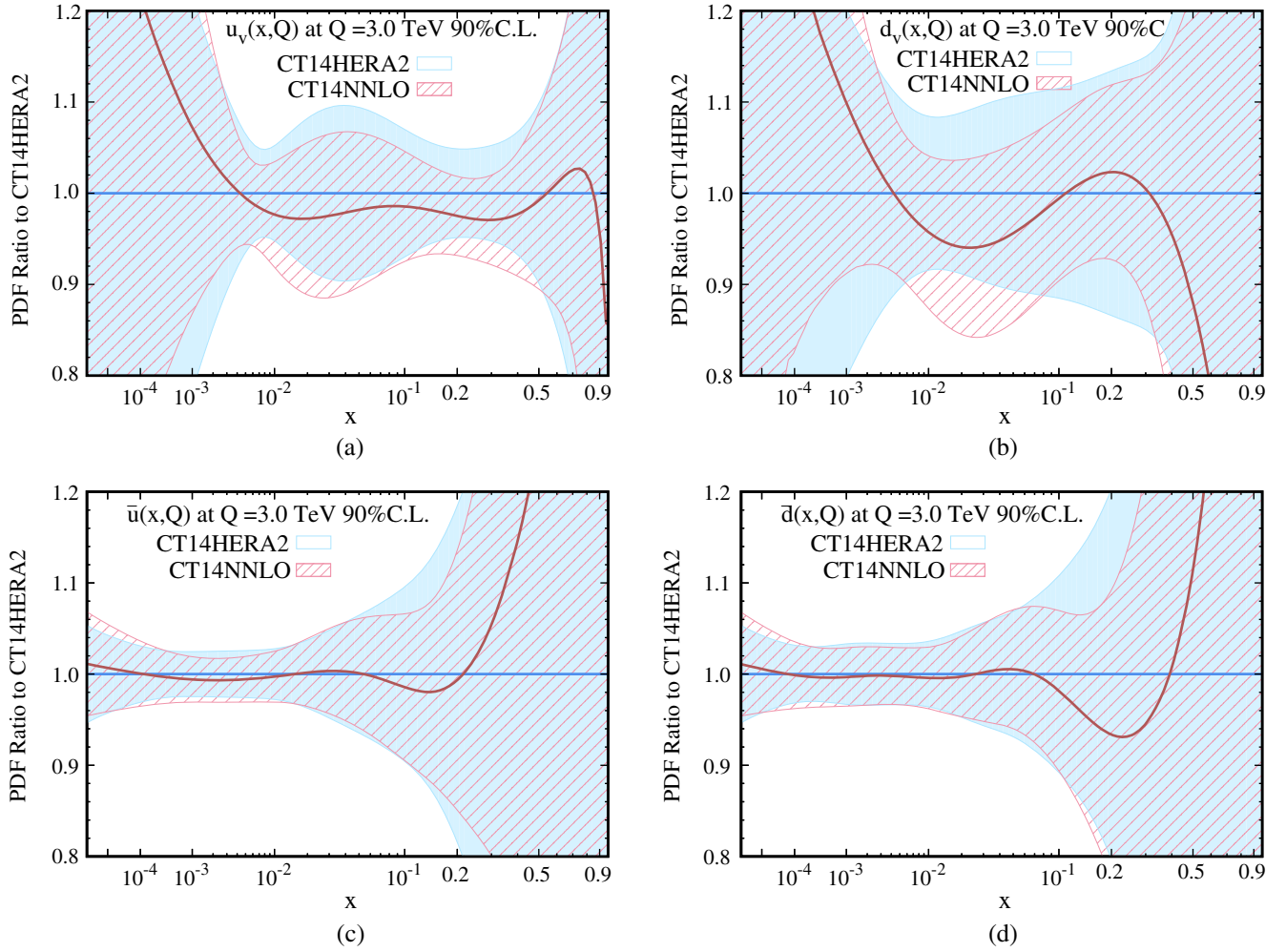
FIG. 5. PDF uncertainties associated with the (a) $u_v(x)$, (b) $d_v(x)$, (c) $\bar{u}(x)$, and (d) $\bar{d}(x)$ distributions in the CT14NNLO and CT14HERA2 PDF sets evaluated at a scale of $Q = 3$ TeV. The common denominator is the CT14HERA2 central set. At high values of x, such as $x \gtrsim 0.1$ relevant for high-mass DY production, the PDF uncertainties begin to diverge. As noted in the introduction, the differences between these two recent fits is minimal, justifying our choice of CT14HERA2 in this analysis.

Our strategy was to explore the DY cross section with the goal of finding global PDF fitting inputs tailored specifically to DY physics. To that end we used the RESBOS MC (and the MADGRAPH generator [38] as a check), configured with the CT14HERA2 PDF set to study several kinematic distributions.

All simulation samples are produced in bins of true dilepton invariant mass in the range $m_{\ell\ell} = 40$ GeV to $m_{\ell\ell} = 1$ TeV at $\sqrt{s} = 13$ TeV. In order to roughly correspond to ATLAS [7] and CMS [8] acceptances, the lepton pseudorapidities were restricted. Central-central (CC) events are required to have both leptons with $|y_\ell| < 2.47$, and the dilepton rapidity, $|y_{\ell\ell}| < 2.47$. Central-forward (CF) events are required to have one lepton with $|y_\ell| < 2.47$, and the other with $2.47 < |y_\ell| < 3.6$, where the dilepton rapidity can extend out to $|y_{\ell\ell}| < 3.6$, which allows access to a wider range in x, cf. Eq. (11).

We found particular practical significance in focusing on the polar angle. Figure 7 shows several $\cos\theta^*$ distributions

of Eq. (9) in discrete slices of dilepton invariant mass. Each mass-slice is further decomposed into subprocesses that consist distinctly of up-type or down-type initial-state quarks. The up-type subprocesses include initial states of $u\bar{u}$, ug, and $\bar{u}g$, where u is the up quark or charm quark and g is the gluon. A similar definition applies to the d-type (down, strange, bottom) subprocesses, with u replaced by d. This is in accordance with the four DY reactions in Eqs. (1) and (4).

The distributions in Figs. 7(a)–7(c) are essentially the regions covered by an ATLAS measurement of the triple differential cross section during the 8 TeV running. [35] These are familiar as they show part of the source of the oft-measured forward-backward asymmetry in both $p - \bar{p}$ and pp on-resonance Z boson analyses [39].

Intriguingly, the relative up-type and down-type subprocesses are highly dependent on both mass and polar angle θ^*. This is especially true above the Z boson mass peak, in which the forward region ($\cos\theta^* > 0$) shows an
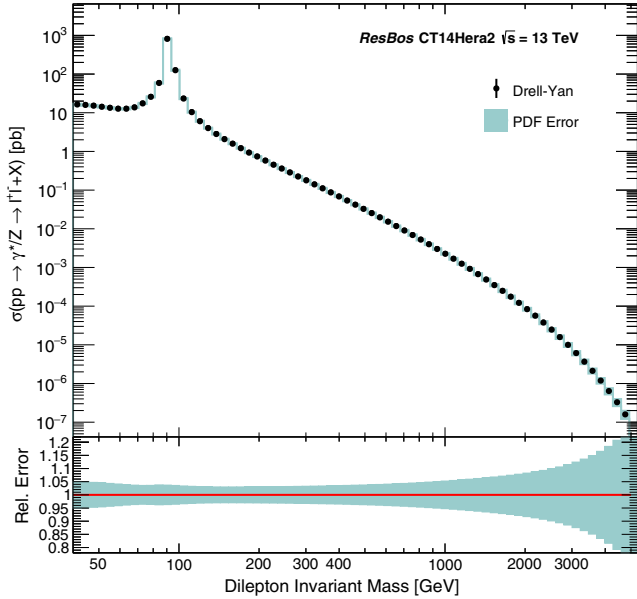
FIG. 6.   The dilepton invariant mass spectrum, generated with the RESBOS MC generator and the CT14HERA2 PDF set.

increasing degree of separation between the rates associated with the up-type and down-type DY sub-processes. Indeed, in this region the contribution to the total cross section is due almost entirely to the up-type subprocess by itself: almost by a factor of four. *At high mass and high polar angle, the LHC DY process proceeds almost entirely through the $u\bar{u}$ subprocess, effectively making the LHC a $u\bar{u}$ collider.*

Why is this the case? Appendix B explains this conclusion as a fortuitous conspiracy of electroweak couplings and parton luminosities which collectively favor up quarks and antiquarks over their down-like counterparts.

Notice that we have not really learned anything new since DY kinematics is an old subject. But high-mass behavior in regions only statistically available at the LHC is revealing and the question is whether $\cos\theta^*$ behavior as a function of mass should be an important discrimination as an input to global PDF fitting. This is where EPUMP comes in.

IV. A PROPOSED STRATEGY TO PDF ERROR REDUCTION FOR DY

We attempt to shed light on two questions:
(1) If $\cos\theta^*$ data were incorporated in global fitting, how significant might the reduction in PDF uncertainties be?
(2) Would those decreased errors be a significant reduction in the overall theoretical uncertainties in future BSM, high-mass DY searches?

In order to answer Question 1, EPUMP was used, which can update an existing PDF set with new experimental data (or pseudodata) in order to produce an improved best-fit

and Hessian error PDFs. The EPUMP workflow can be seen in Fig. 8.

For this analysis, "pseudodata" are used to mimic a possible future LHC dataset for PDF fitting. As any dataset has finite statistics, the resulting uncertainties in the new PDFs will reflect whatever statistical precision is modeled in the pseudodata. The effects of new PDFs and uncertainties can then be used to reevaluate the PDF systematic uncertainty on the high-mass dilepton event yield.

Furthermore, we imagine a signal region (SR) as $m_{\ell\ell} > 1$ TeV and a control region (CR) to be for $0.04 < m_{\ell\ell} < 1$ TeV. Since new physics should lie above the current limits of approximately $m_{\ell\ell} \sim 3$ TeV (as in Sec. II), it would be "fair" to use low-mass DY data to constrain the high-mass DY spectrum.

A. PDF update strategy

EPUMP requires standard inputs to emulate the global fit—the templates in Fig. 8. We describe our strategy here. The analysis was performed at "truth level," such that the acceptance and efficiency effects associated with the reconstruction and identification of prompt, high-p_T leptons in an LHC detector are neglected. However, leptons are well measured at the LHC, so this is an acceptable first look at this technique. Additional dilepton backgrounds were neglected, but are well understood by the LHC experiments as can be seen in Fig. 1. These backgrounds include $t\bar{t}$ production, Wt single top production, WW, WZ, and ZZ diboson production, and $W + $ jets and multijet production in the electron channel.

B. EPUMP template construction

Naively, one might imagine only using $m_{\ell\ell}$ in the CR to predict the improvement in the SR, but our awareness of the significant differential quark sensitivities to $\cos\theta^*$ (and moderate sensitivity to $y_{\ell\ell}$) plus the knowledge that future LHC running will provide enormous continuum DY datasets led us to explore dividing pseudo-data into many bins of dilepton mass $m_{\ell\ell}$, as well as $y_{\ell\ell}$ and $\cos\theta^*$.

The fiducial region considered for our analysis is designed explicitly to probe the PDFs at high x, and is defined by

$$40 \text{ GeV} < m_{\ell\ell} < 1000 \text{ GeV},$$
$$|y_{\ell\ell}| < 3.6, \qquad -1 < \cos\theta^* < 1. \qquad (11)$$

DY samples were generated using the RESBOS MC generator with the CT14HERA2 PDF set for the $\sqrt{s} = 13$ TeV LHC. Events were further required to pass a loose event selection in order to construct the finalized Data templates. Dilepton events with an invariant mass of $m_{\ell\ell} > 80$ GeV were required to satisfy $p_T > 30$ GeV, while low-mass events in the interval of $40 < m_{\ell\ell} < 80$ must satisfy
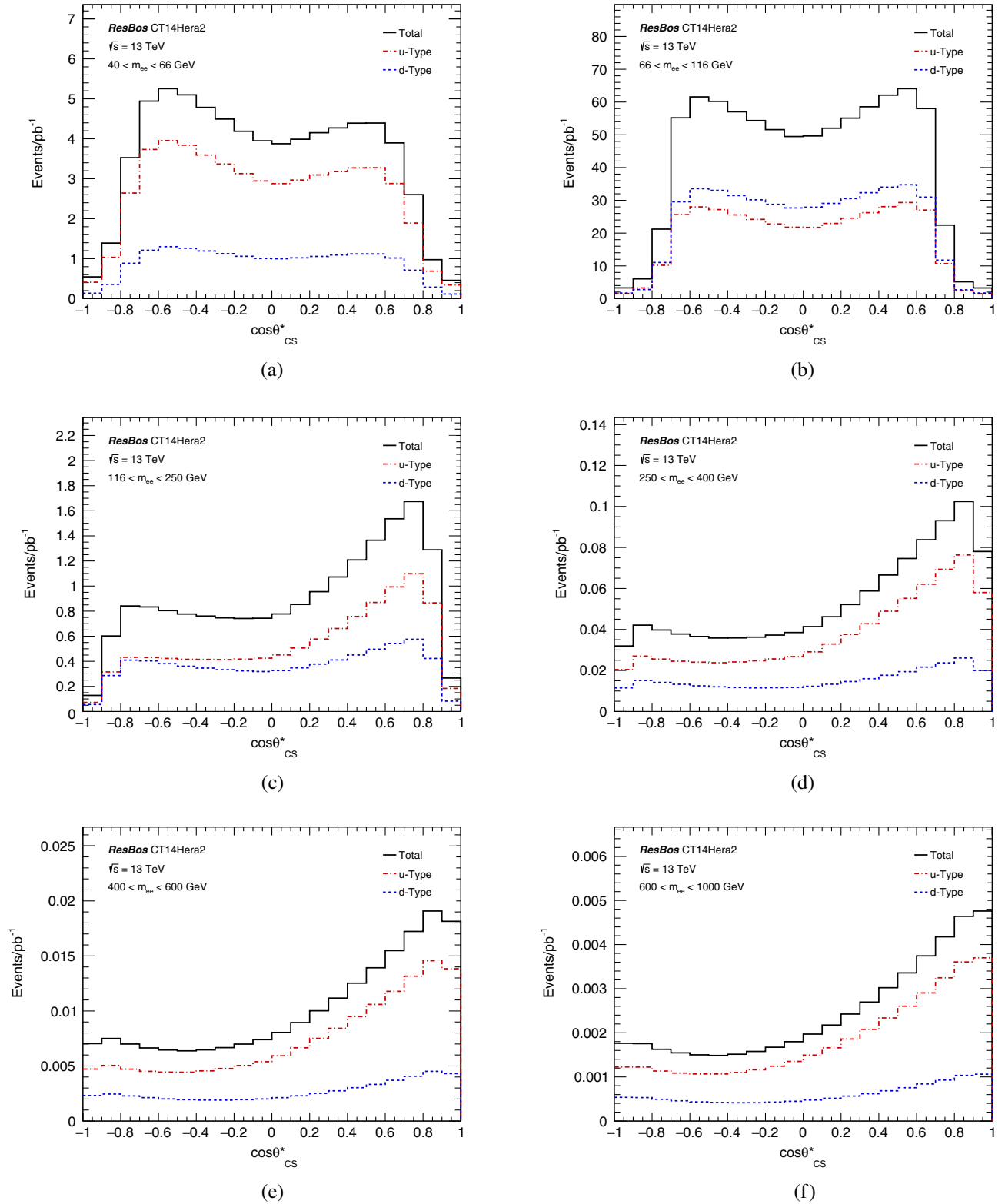
FIG. 7. The lepton angular distribution $\cos \theta^*$ in slices of dilepton invariant mass, $m_{\ell\ell}$, ranging from 40 GeV to 1 TeV (a)-(f). The up-type and down-type DY subprocesses are shown as well, which exhibit a strong angular dependence, especially at high mass. The CT14HERA2 PDF set is used.
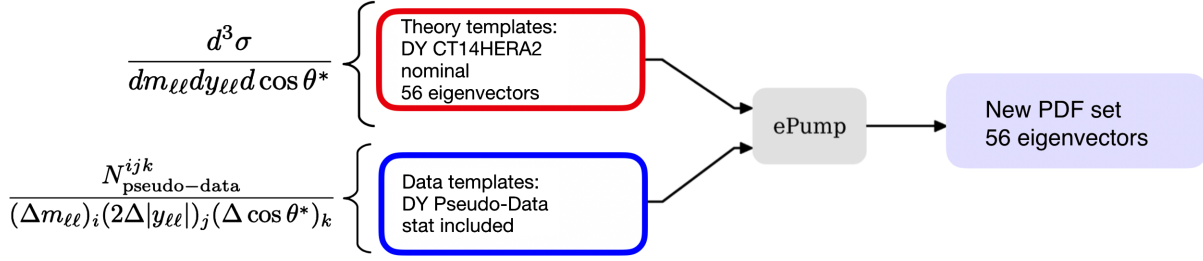
FIG. 8.　The EPUMP package requires two inputs to generate an updated PDF set: an existing Theory template of a PDF set (parameters + uncertainties) and binned Data template of (pseudo-) data, including statistical uncertainties from integrated luminosity assumptions.

$p_T > 15$ GeV. In addition, events must consist of leptons which are distributed as central-central or central-forward.

Events passing these selections were binned in EPUMP template histograms, which parametrize the triple-differential cross section of Eq. (5), according to

$$\mathcal{L}_{\text{int}}\left(\frac{d^3\sigma}{dm_{\ell\ell}d|y_{\ell\ell}|d\cos\theta^*}\right)_{ijk}$$
$$= \frac{N^{ijk}_{\text{pseudo-data}}}{(\Delta m_{\ell\ell})_i(2\Delta|y_{\ell\ell}|)_j(\Delta\cos\theta^*)_k}, \qquad (12)$$

where i, j, and k correspond to the bin indices of each distribution of interest. Note that in a realistic measurement, the numerator of Eq. (12) would be replaced by $N^{ijk}_{\text{data}} - N^{ijk}_{\text{bkg}}$, where the background component arises from the standard dilepton background processes.

The total number of pseudo-data events are given by $N^{ijk}_{\text{pseudo-data}}$, the integrated luminosity of the pseudo-dataset is $\mathcal{L}_{\text{int}}$, and $(\Delta m_{\ell\ell})_i$, $(2\Delta|y_{\ell\ell}|)_j$, and $(\Delta\cos\theta^*)_k$ are the

corresponding bin widths. The factor of two in the denominator accounts for the modulus in the rapidity bin width. The bins used to parametrize Eq. (12) are

 (i) $40 < m_{\ell\ell} < 1000$: {40, 66, 80, 91, 102, 116, 145, 200, 275, 381, 525, 725, 1000} GeV
 (ii) $0 < |y_{\ell\ell}| < 2.4$: {0.0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0, 2.2, 2.4}
 (iii) $2.4 < |y_{\ell\ell}| < 3.6$: {2.4, 2.6, 2.8, 3.0, 3.2, 3.4, 3.6}
 (iv) $-1 < \cos\theta^* < 1$: {−1.0, −0.7, −0.4, 0.0, 0.4, 0.7, 1.0}.

where two $|y_{\ell\ell}|$ regions explicitly call out the CC and CF selections. The total number of measurement bins is $N_{\text{bins}} = 12 \times 18 \times 6 = 1296$ for the fiducial region considered and they define the N_{new} data points that supplement Eq. (A1).

Events were generated as if they came from a future integrated luminosity and so uncertainties in the EPUMP results are scattered according to the statistics of such a hypothetical LHC input dataset. For each bin the DY cross section estimate $\sigma^{ijk}_{\text{Drell-Yan}}$ was scaled by a characteristic integrated luminosity $\mathcal{L}_{\text{int}}$ to arrive at a definite DY event
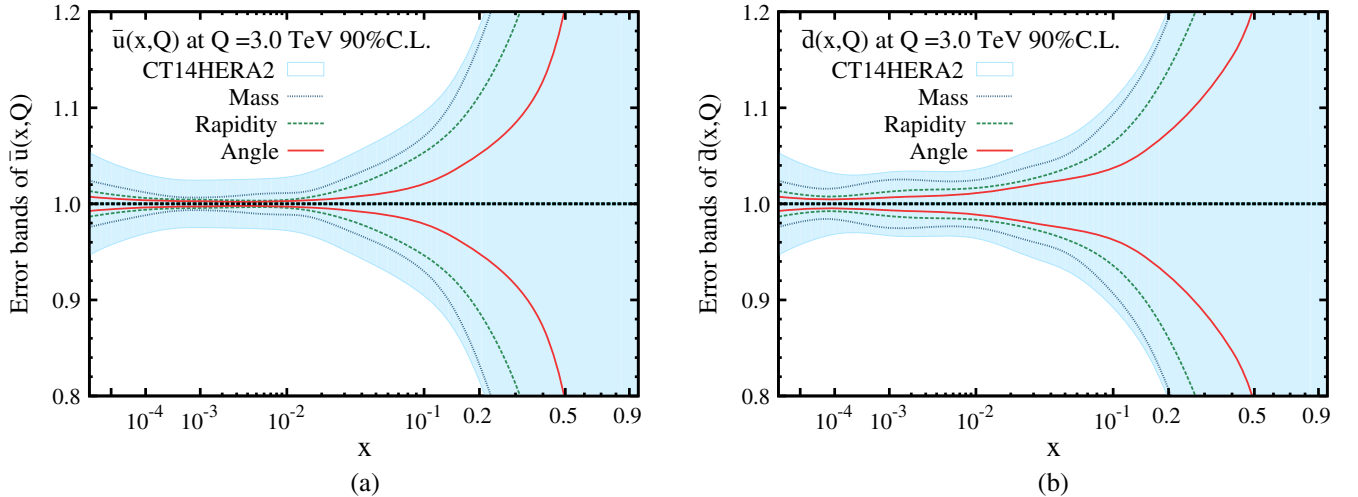

FIG. 9.　Impact of the 3000 fb^{-1} update on the (a) CT14HERA2 $\bar{u}(x)$ and (b) $\bar{d}(x)$. The shaded background shows the uncertainties resulting from the current CT14HERA2 uncertainties. The dotted curve labeled "mass" corresponds to the error reduction by sending only binned $(\Delta m_{\ell\ell})$ to EPUMP; i.e., integrated over the $y_{\ell\ell}$ and $\cos\theta^*$ dimensions. The dashed curve labeled "rapidity" adds the cumulative effect of binned $(\Delta|y_{\ell\ell}|)$ and $(\Delta m_{\ell\ell})$ to EPUMP. Finally, the solid curve labeled "angle" adds the cumulative effect of binned $(\Delta\cos\theta^*)$, $(\Delta|y_{\ell\ell}|)$, and $(\Delta m_{\ell\ell})$ to EPUMP.
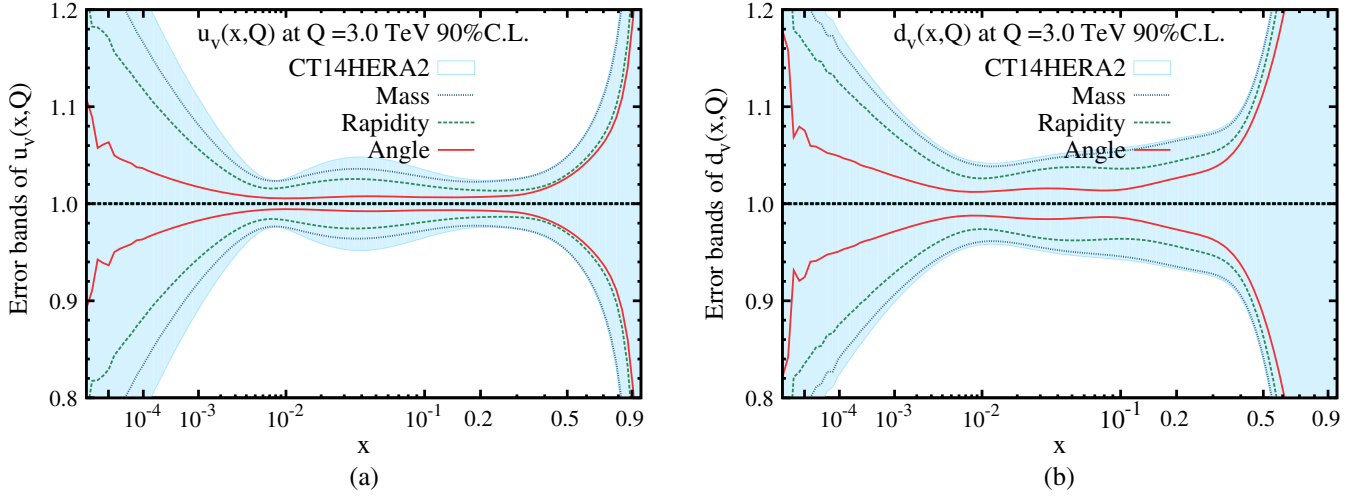
FIG. 10.   Impact of the 3000 fb^{-1} update on the (a) CT14HERA2 $u_v(x)$ and (b) $d_v(x)$. The shaded background shows the uncertainties resulting from the current CT14HERA2 uncertainties. The dotted curve labeled "mass" corresponds to the error reduction by sending only binned $(\Delta m_{\ell\ell})$ to EPUMP; i.e., integrated over the $y_{\ell\ell}$ and $\cos\theta^*$ dimensions. The dashed curve labeled "rapidity" adds the cumulative effect of binned $(\Delta|y_{\ell\ell}|)$ and $(\Delta m_{\ell\ell})$ to EPUMP. Finally, the solid curve labeled "angle" adds the cumulative effect of binned $(\Delta\cos\theta^*)$, $(\Delta|y_{\ell\ell}|)$, and $(\Delta m_{\ell\ell})$ to EPUMP.

yield $N^{ijk}_{\text{Drell-Yan}}$. The resulting yield was assumed to be the mean of a Poisson distribution, which was then used to throw a random number according to Poisson statistics, thereby populating the bin with $N^{ijk}_{\text{pseudo-data}}$ pseudodata events. Note that the pseudodata were treated as those of one "experiment," but in practice ATLAS, CMS, and LHCb would all be sources of fitting input data. For illustration we chose two future LHC scenarios for integrated luminosities: $\mathcal{L}_{\text{int}} = 300$ fb^{-1} approximating the data set for one experiment following Run-3 of the LHC, and $\mathcal{L}_{\text{int}} = 3000$ fb^{-1}, approximating that of the final dataset for one experiment of the high luminosity (HL) LHC.

V. PDF UPDATE RESULTS

We can answer Question 1 by reevaluating the effect of the 3000 fb^{-1} DY pseudodataset on the CT14HERA2 PDFs, as well as Question 2 by assessing the reduction of the PDF systematic uncertainty in the high-mass dilepton spectrum.

A. Impact on CT14HERA2 PDFs

Question 1 asked whether explicit inclusion of $\cos\theta^\star$ data might have a useful effect in reducing the uncertainties in the parton fits. The answer can be seen in the following four plots in Figs. 9 and 10. In order to see the effect of each of the quantities in the EPUMP-simulated refitting, there are four sets of results in each plot. Figure 9 shows the impact of the EPUMP update with the 3000 fb^{-1} scenario on the $\bar{u}(x)$ and $\bar{d}(x)$ sea distributions and Fig. 10, the impact on the $u_v(x)$ and $d_v(x)$ valence distributions.

The sea distributions show a considerable reduction in uncertainty at high x. For example, in both the $\bar{u}(x)$ and $\bar{d}(x)$ distributions, the PDF uncertainty is reduced from its pre-update value of approximately 70% to 20% at $x = 0.5$. The improvement in the valence distributions at $x \gtrsim 0.5$ is less dramatic, but substantial improvement is observed in the ranges of $x \lesssim 0.5$. The post-update $u_v(x)$ distribution remains better constrained than $d_v(x)$ at high x, where the uncertainty measures 2.6% as compared to 11% at $x = 0.5$, respectively. Table III lists the pre- and post-update

TABLE III.   Impact of 3000 fb^{-1} update on the CT14HERA2 $u_v(x)$ and $d_v(x)$ valence and $\bar{u}(x)$ and $\bar{d}(x)$ sea distributions for several values of x using the standard triple-differential templates at $Q = 3$ TeV. To be compared with the "Angle" curves of Figs. 9 and 10.

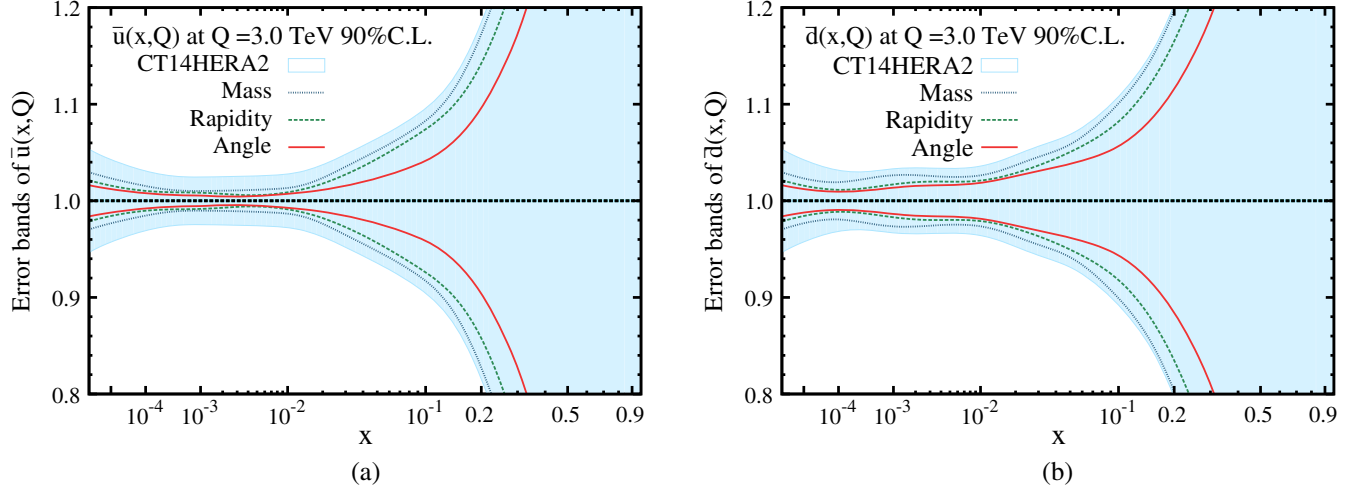|	$u_v(x)$		$d_v(x)$		$\bar{u}(x)$		$\bar{d}(x)$	
x	δ_{pre} [%]	δ_{post} [%]	δ_{pre} [%]	δ_{post} [%]	δ_{pre} [%]	δ_{post} [%]	δ_{pre} [%]	δ_{post} [%]
0.1	3.4	0.7	5.8	1.5	9.8	2.2	11	3.8
0.3	2.6	0.9	7.5	3.6	30	8.3	32	11
0.5	4.8	2.6	16	11	71	20	69	20
0.7	12	7.0	45	30	280	77	250	67

FIG. 11.   Impact of the 300 fb^{-1} update on the (a) CT14HERA2 $\bar{u}(x)$ and (b) $\bar{d}(x)$. The shaded background shows the uncertainties resulting from the current CT14HERA2 uncertainties. The dotted curve labeled "mass" corresponds to the error reduction by sending only binned $(\Delta m_{\ell\ell})$ to EPUMP; i.e., integrated over the $y_{\ell\ell}$ and $\cos\theta^*$ dimensions. The dashed curve labeled "rapidity" adds the cumulative effect of binned $(\Delta|y_{\ell\ell}|)$ and $(\Delta m_{\ell\ell})$ to EPUMP. Finally, the solid curve labeled "angle" adds the cumulative effect of binned $(\Delta\cos\theta^*)$, $(\Delta|y_{\ell\ell}|)$, and $(\Delta m_{\ell\ell})$ to EPUMP.

uncertainties for several parton flavors and values of x explicitly.

Figures 11 and 12 show the reduction in uncertainties for the 300 fb^{-1} scenario and Table IV is the corresponding comparison for the 300 fb^{-1} scenario.

The answer to Question 1 is that a global PDF fit which includes DY LHC data below 1 TeV in mass, and binned in rapidity and $\cos\theta^*$, would dramatically improve the precision in our knowledge of the up and down PDFs. During the LHC era DY measurements of this kind are likely the only way to reduce uncertainties on the PDFs at high x; no other input data are capable of achieving this improvement.

B. Impact on the high-mass Drell-Yan spectrum

With an updated set of PDFs, we can answer Question 2: the effect of new PDFs on the systematic uncertainty on high-mass DY cross section. Rather than the enormous extrapolation required of current-day PDFs, the extrapolation from our control region to our signal region is modest and impactful. In order to make contact with primarily the ATLAS dilepton analysis [7], the invariant mass distribution assessed here utilizes leptons that originate in the central-central final state only.

The results are presented in Fig. 13, which shows the impact of the 3000 fb^{-1} pseudo-dataset on the high-mass
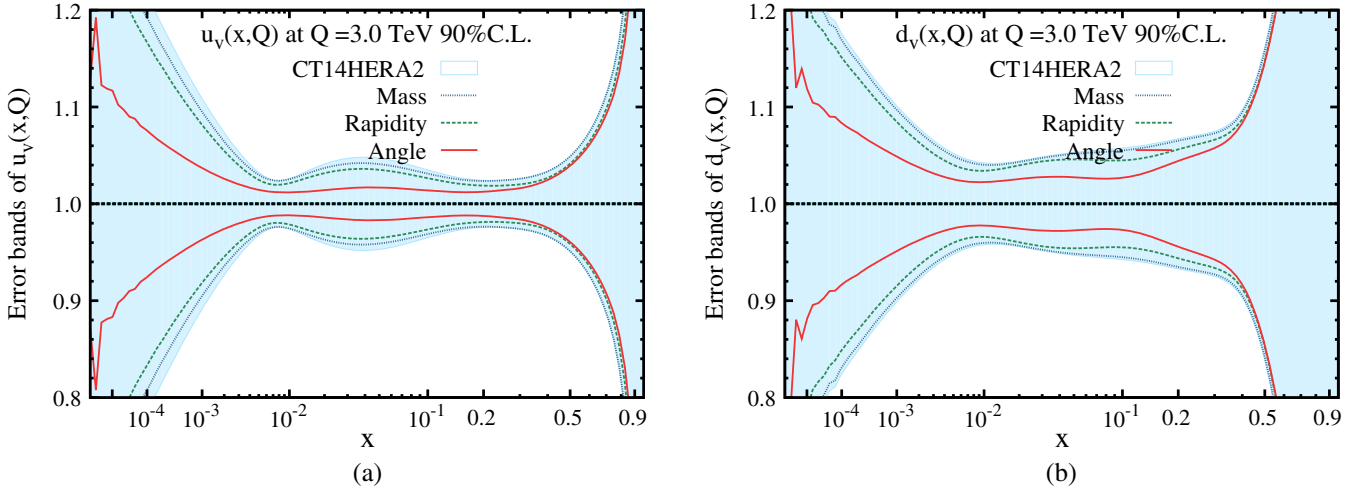

FIG. 12.   Impact of the 300 fb^{-1} update on the (a) CT14HERA2 $u_v(x)$ and (b) $d_v(x)$. The shaded background shows the uncertainties resulting from the current CT14HERA2 uncertainties. The dotted curve labeled "mass" corresponds to the error reduction by sending only binned $(\Delta m_{\ell\ell})$ to EPUMP; i.e., integrated over the $y_{\ell\ell}$ and $\cos\theta^*$ dimensions. The dashed curve labeled "rapidity" adds the cumulative effect of binned $(\Delta|y_{\ell\ell}|)$ and $(\Delta m_{\ell\ell})$ to EPUMP. Finally, the solid curve labeled "angle" adds the cumulative effect of binned $(\Delta\cos\theta^*)$, $(\Delta|y_{\ell\ell}|)$, and $(\Delta m_{\ell\ell})$ to EPUMP.

TABLE IV.   Impact of 300 fb^{-1} update on the CT14HERA2 $u_v(x)$ and $d_v(x)$ valence and $\bar{u}(x)$ and $\bar{d}(x)$ sea distributions for several values of x using the standard triple-differential templates at $Q = 3$ TeV. To be compared with the "Angle" curves of Figs. 11 and 12.

x	$u_v(x)$		$d_v(x)$		$\bar{u}(x)$		$\bar{d}(x)$	
	δ_{pre} [%]	δ_{post} [%]	δ_{pre} [%]	δ_{post} [%]	δ_{pre} [%]	δ_{post} [%]	δ_{pre} [%]	δ_{post} [%]
0.1	3.4	1.4	5.8	2.7	9.8	4.3	11	6.0
0.3	2.6	1.6	7.5	5.7	30	17	32	19
0.5	4.8	3.9	16	14	71	43	69	41
0.7	12	9.7	45	41	280	180	250	160

PDF systematic uncertainty. The PDF uncertainty is evaluated at several characteristic values of dilepton mass, which are listed in Table V. At $m_{\ell\ell} = 5$ TeV, the PDF systematic uncertainty is reduced from 31% to 8.9%, a reduction of roughly a factor of 3.5. Similarly, at $m_{\ell\ell} = 3$ TeV, the uncertainty is reduced from 15% to 3.7%, roughly a factor of 4. In each case, a substantial improvement is obtained compared to the current state-of-the-art predictions (as depicted in Fig. 7). The PDF uncertainty assessed in the ATLAS dilepton analysis is, for example, 13% and 29% at $m_{\ell\ell} = 3$ and $m_{\ell\ell} = 5$ TeV, respectively.

It is worth remembering that many differential cross section analyses of DY data around the Z peak have been performed over the years, including a triple differential cross section measurement by ATLAS [35]. We found that because of the extremely high rate, including using the triple differential cross sections to global PDF fitting from the low mass region should indeed be important. However, because of the surprising sensitivities to the parton density flavors, and the enormous rates from the 3000 fb^{-1} running, about half of the above uncertainty improvement came from the high statistics, low mass region and about

half came from the high mass continuum, but low cross section region. Therefore we advocate using the entire dilepton invariant mass spectrum—from below the Z peak to approximately 1 TeV—as the control region for inputs to future PDF global fitting. The only assumption this carries is that no new physics lurks in the continuum below that boundary.

VI. OUTLOOK

The impact of a future DY cross section measurement on the CT14HERA2 PDF uncertainty was assessed using the EPUMP package at the $\sqrt{s} = 13$ TeV LHC with 300 fb^{-1} and 3000 fb^{-1} of DY pseudodata. The fiducial region considered for the PDF update was based on three variables: the dilepton mass ($m_{\ell\ell}$), the dilepton rapidity ($y_{\ell\ell}$), and the cosine of the polar angle in the CS-frame ($\cos\theta^*$). These regions were divided into 1296 histogram bins and used to construct EPUMP pseudodata and signal templates, which were designed to probe the PDFs in the extreme kinematic regions of (x, Q^2) only accessible at the LHC.
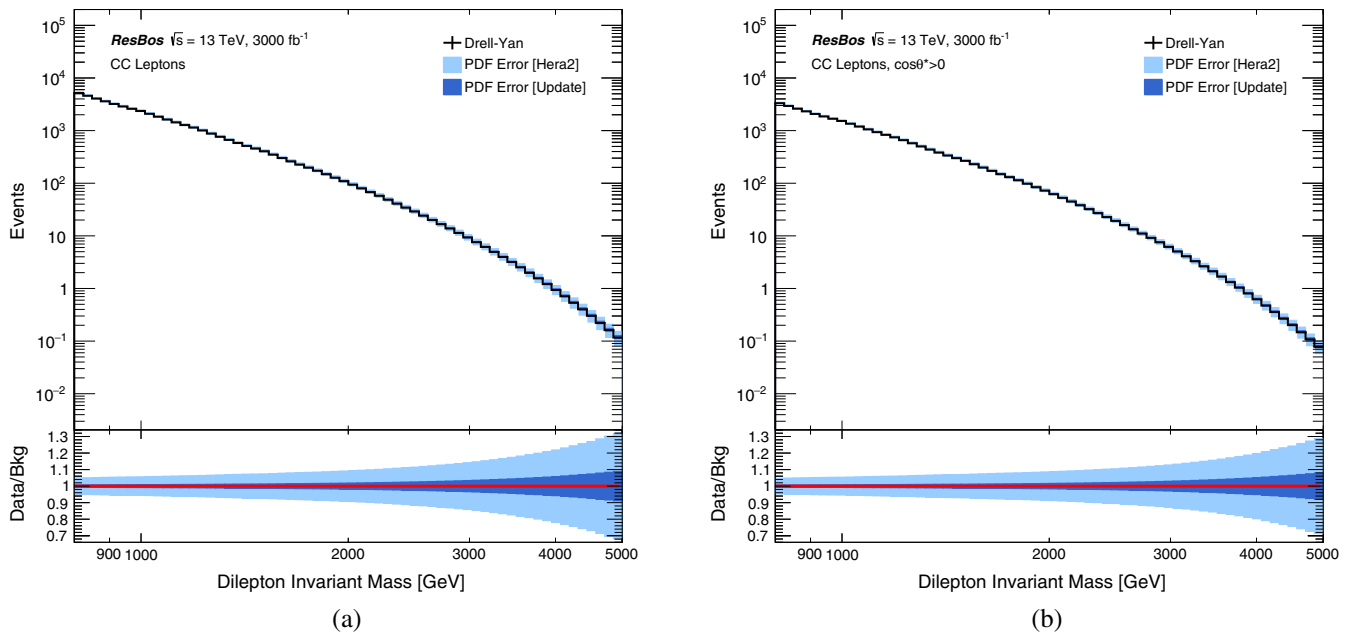

FIG. 13.   The dilepton invariant mass distribution for (a) central-central dilepton events, and (b) with an additional $\cos\theta^* > 0$ requirement added to the selection. The ratio subplot depicts the CT14HERA2 PDF uncertainty before and after the 3000 fb^{-1} update.

TABLE V. The estimated PDF uncertainty in several invariant mass bins for the distributions shown in Fig. 13. Two selections are tested: firstly the central-central selection, and secondly for the central-central selection with an additional $\cos\theta^* > 0$ requirement. For each selection, the current CT14HERA2 uncertainty estimates are shown in the first column, and the result of the 3000 fb^{-1} update is shown in the second. The preupdate values for the central-central selection are consistent with those assessed in the ATLAS dilepton analysis [7].

	CC selection		$CC + \cos\theta^*$ selection	
$m_{\ell\ell}$ [TeV]	$\delta_{\rm pre}^{\rm PDF}$ [%]	$\delta_{\rm post}^{\rm PDF}$ [%]	$\delta_{\rm pre}^{\rm PDF}$ [%]	$\delta_{\rm post}^{\rm PDF}$ [%]
1	5.9	1.0	5.6	0.9
2	9.6	2.0	8.9	1.7
3	15	3.7	13	3.2
4	22	6.0	20	5.3
5	31	8.9	28	8.0

The CT14HERA2 PDF set was used for the update, but similar effects would be observed in other PDF sets. The results showed a significant reduction in the uncertainties associated with all parton flavors, especially $\bar{u}(x)$ and $\bar{d}(x)$ sea distribution at high x. Likewise, these reduced PDF uncertainties, when propagated to the dilepton invariant mass spectrum, lead to a significantly improved description at high mass.

These proof-of-concept results indicate a great deal of improvement can still be obtained from precision PDF measurements at LHC. The use of $\cos\theta^*$ as an additional dimension in future PDF global fits is absolutely crucial, as it supplements the more standard double-differential measurements in invariant mass and rapidity; when used in conjunction, as was done here, the reduction in uncertainty can be dramatic.

For these reasons, DY cross section measurements could be vital to the success of future searches and measurements at the LHC. Not only will the PDF uncertainty that affects the high-mass dilepton analysis be reduced, improving the discovery potential of many non-resonant new physics models, but also the inclusion of new and robust data into the modern PDF global fits might even bring the uncertainty estimates of the various global fitting groups into better agreement.

Such an opportunity might result in a reduction of the "PDF choice" uncertainty when all PDF groups include

triply differential DY data as discussed here. Obviously the goal would be to reach a stage in which the largest uncertainty would cease to be due to the PDFs. Table VI compares these uncertainties explicitly, where the uncertainty on the QCD background estimate is not included in calculating the post-update PDF uncertainty which will be reduced well below the current experimental uncertainty.

Therefore, for the reasons outlined in this paper, experiments at the LHC and global fitting groups should seriously consider the inclusion of precision measurements of the DY triple-differential cross section over a large invariant mass region in order to further constrain the PDF uncertainties in future PDF global fits.

APPENDIX A: THE EPUMP PACKAGE

In a standard PDF global fit, the PDFs are determined by minimizing the function,

$$\chi_{\rm global}^2 = \sum_{n=1}^{N_{\rm exp}} \chi_n^2, \qquad (A1)$$

which consists of contributions from $N_{\rm exp}$ fitted experiments, χ_n^2. In the simplest case with no correlations between data points, the contribution from an experiment can be written

$$\chi_n^2 = \sum_{i=1}^{N_n} \frac{(T_{ni}(\mathbf{z}) - D_{ni})^2}{\sigma_{ni}^2}, \qquad (A2)$$

where D_{ni} is the experimental data value, σ_{ni} is the experimental error (combined systematic and statistical), and $T_{ni}(\mathbf{z})$ is the theory prediction, which depends on the

TABLE VI. The post-update PDF uncertainty as compared to the current experimental and dominant theoretical uncertainties in the electron channel of the dilepton analysis. As the PDF uncertainty will be reduced well below the current experimental uncertainty, attention will be shifted to the reduction of others, such as the "PDFChoice" uncertainty, improving the discovery potential of future iterations of the dilepton analysis.

	CC selection	ATLAS dilepton analysis				
$m_{\ell\ell}$ [TeV]	$\delta_{\rm post}^{\rm PDF}$ [%]	$\delta^{\rm PDF}$ [%]	$\delta^{\rm Choice}$ [%]	$\delta^{\rm Theory}$ [%]	$\delta^{\rm Exp}$ [%]	$\delta^{\rm Total}$ [%]
2	2.0	8.7	<1.0	9.8	11.0	14.7
4	6.0	19.0	8.4	23.0	12.8	26.3

PDFs, which in turn are described by a finite number of parameters, **z**. In practice, the χ_n^2 for modern experiments will include correlated errors among data points, and there may be additional terms added to impose constraints on the theoretical parameters, but the general procedure is unchanged. The central or best-fit PDFs are obtained by minimizing χ_{global}^2 with respect to **z**. In addition, χ_{global}^2 is, to a good approximation, a quadratic function of the parameters around the minimum. This is the basis for the Hessian approximation for PDF errors, which utilize PDF eigenvector sets, two for each PDF parameter, to evaluate the uncertainty due to the PDFs for any physical observable. Each PDF eigenvector set corresponds to a movement in the parameter space along the eigenvector directions of the Hessian error matrix around the global minimum of χ_{global}^2 at a defined confidence level (C.L.). For CT14HERA2 PDFs there are 56 eigenvector sets defined at the 90% C.L.

If the contribution from a new experiment, $\chi_{N_{\text{exp}}+1}^2$, is added to the global analysis, the exact solution of the problem would require finding the new minimum of Eq. (A1), as well as diagonalizing the new Hessian matrix. Since this requires the full data sets from all experiments in the global analysis, as well as the theory calculations for every data point evaluated at many parameter values, it is an onerous and time-consuming task even for the global analysis teams that specialize in this endeavor.

This is where a tool such as EPUMP is advantageous. EPUMP works by using the fact that the original χ_{global}^2 is well-approximated by the known quadratic function and the fact that the theory predictions for the new observables, $T_{N_{\text{exp}}+1,i}(\mathbf{z})$, can be approximated using the original Hessian error PDFs. Under these approximations, the minimization and Hessian diagonalization can be performed algebraically [18,40], with the numerical computations taking seconds, rather than hours or days.

Figure 8 illustrates the use of EPUMP. In order to perform the PDF update, EPUMP requires two sets of inputs: data templates and theory templates. The data templates consist of the new experimental data values and their statistical and systematic uncertainties, including correlations, exactly as would be included in a standard global analysis. In the case of our present study these are the event counts of the new pseudodata, along with their associated statistical uncertainties. The theory templates consist of the corresponding theory predictions for the same observables, evaluated using the central PDF and each of the Hessian eigenvector PDFs. Note that any number of new data sets can be included in the update by EPUMP, with any number of data points per new data set.

The output of EPUMP is an updated central and Hessian eigenvector PDFs, which approximate the result that would be obtained from a full global reanalysis that includes the new data. As an additional benefit, EPUMP can also directly output the updated predictions and uncertainties for any other observables of interest (such as the cross section in the signal

region), without the necessity to recalculate using the updated PDFs. For more details about the use of EPUMP, see Ref. [18]. The code for EPUMP and more specific details of its usage can be obtained at the website http://hep.pa.msu.edu/epump/.

In the present study, we have used EPUMP to assess the reduction of PDF uncertainties from various kinematic selection choices on the Drell-Yan data. It should be noted that if the included new data deviate more from the prediction (based on CT14HERA2), the result of the EPUMP analysis will be less reliable. This is due to the nature of the EPUMP method which assumes a quadratic dependence of χ^2 and a linear dependence of observables when the PDFs vary. Since the pseudodata (generated by MMHT14) and the theory predictions (from CT14HERA2) do not differ much, we expect the results of the EPUMP analysis in our study will hold to a very good approximation. However, should the future data deviate significantly from the theory predictions (from CT14HERA2), a full global analysis, probably with an extended nonperturbative parametrization form, must be carried out.

APPENDIX B: HIGH MASS FAVORING OF $u\bar{u}$ IN pp COLLISIONS

That the $u\bar{u}$ contribution is nearly a factor of four more than the down quark contribution was not expected. In this Appendix we show how this comes about.

1. Relearning Drell-Yan kinematics

We will exploit a novel feature of the DY subprocess cross section from Eq. (5) and the definition of $\cos\theta^*$ of Eq. (9), to form the basis of the PDF update with EPUMP in Sec. IV.

The function P_q of Eq. (5) encodes the parton-level dynamics with

$$P_q = C_q^0(1 + \cos^2\theta^*) + C_q^1 \cos\theta^*, \qquad \text{(B1)}$$

which is a weighted sum of an even function $(1 + \cos^2\theta^*)$ and an odd function $\cos\theta^*$. In the calculation of the total inclusive cross section, the odd term integrates to zero, but is responsible for inducing the well-known γ^*/Z forward-backward asymmetry A_{FB}.

The asymmetry coefficients C_q^0 and C_q^1 of Eq. (B1) include the electroweak couplings of the initial-state quarks and final-state leptons, and describe the $m_{\ell\ell}$ spectrum as

TABLE VII. Vector and axial-vector couplings of the SM fermions. Rows specify couplings within each respective fermion generation.

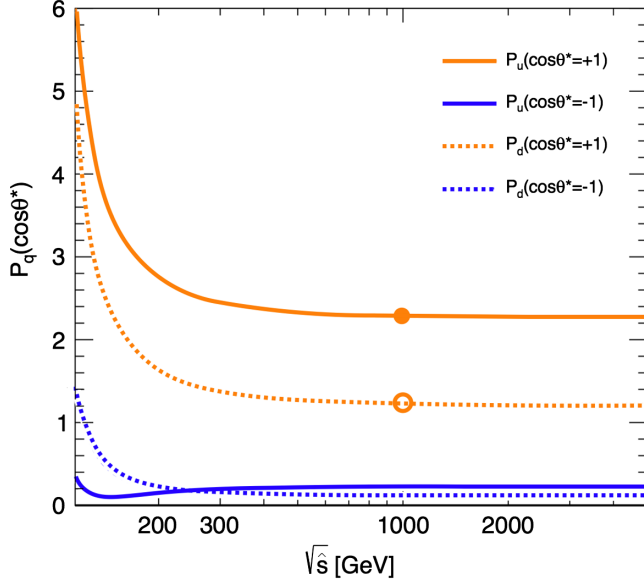|	v_f	a_f
u	$+\frac{1}{4} - \frac{2}{3}x_W$	$-\frac{1}{4}$
d	$-\frac{1}{4} + \frac{1}{3}x_W$	$+\frac{1}{4}$
ν_ℓ	$+\frac{1}{4}$	$-\frac{1}{4}$
ℓ	$-\frac{1}{4} + x_W$	$+\frac{1}{4}$

FIG. 14. The DY parton-level kinematics as described Eq. (B1). The function P_q is evaluated at values of $\cos \theta^* = \pm 1$ for both up (solid)- and down (dashed)-type quarks. The dots indicate the values the P_u and P_d functions at $\cos \theta^* = 1.0$ and $\sqrt{\hat{s}} = 1$ TeV.

$$C_q^0(m_{\ell\ell}) = Q_\ell^2 Q_q^2 + 2Q_\ell Q_q v_\ell v_q \chi_1(m_{\ell\ell})$$
$$+ (a_\ell^2 + v_\ell^2)(a_q^2 + v_q^2)\chi_2(m_{\ell\ell})$$
$$C_q^1(m_{\ell\ell}) = 4Q_\ell Q_q a_\ell a_q \chi_1(m_{\ell\ell}) + 8a_\ell v_\ell a_q v_q \chi_2(m_{\ell\ell}).$$
(B2)

Where

$$\chi_1(m_{\ell\ell}) = \frac{1}{\sin \theta_W \cos \theta_W} \frac{m_{\ell\ell}^2(m_{\ell\ell}^2 - m_Z^2)}{(m_{\ell\ell}^2 - m_Z^2)^2 + \Gamma_Z^2 m_Z^2},$$

$$\chi_2(m_{\ell\ell}) = \frac{1}{\sin^2 \theta_W \cos^2 \theta_W} \frac{m_{\ell\ell}^4}{(m_{\ell\ell}^2 - m_Z^2)^2 + \Gamma_Z^2 m_Z^2}. \quad (B3)$$

Here m_Z and Γ_Z are the mass and decay width of the SM Z boson, and Q_f, v_f, and a_f are the electric charge, and the electroweak vector and axial-vector couplings of each
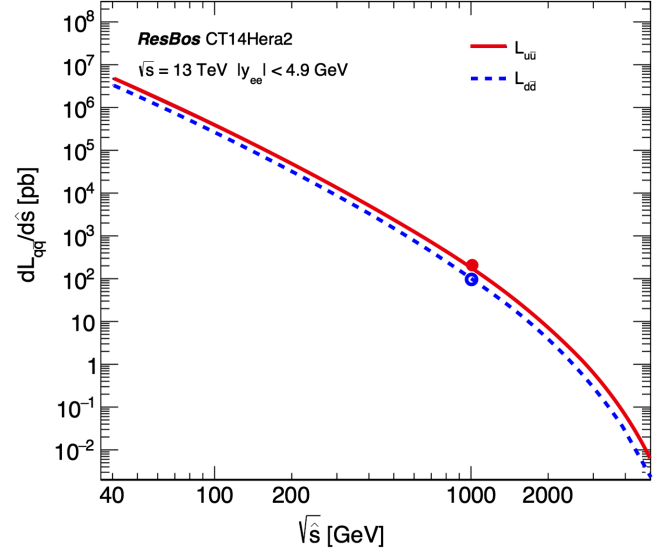

FIG. 15. Parton luminosities for $u\bar{u}$ and $d\bar{d}$ DY sub-processes. The dots indicate the values of the luminosity functions at $\sqrt{\hat{s}} = 1$ TeV.

fermion, whose values are shown in Table VII. The function χ_1 results from γ^*/Z interference, while χ_2 arises from pure Z boson exchange.

Figure 7(f) shows that as $\cos \theta^*$ nears $+1$, the up quark dominates DY production by almost a factor of four over that of the down quark. Taking apart Eqs. (5), (B1), (B2), and (B3) explains this observation. Figure 14 shows the quantities P_u and P_d evaluated at $\cos \theta^* = \pm 1$. The closed (open) circles tag P_u (P_d) at a $\sqrt{\hat{s}} = 1$ TeV for the $\cos \theta^* = 1.0$ curves. The ratio of P_u/P_d is about 2. Figure 15 shows the separate parton luminosity functions $\mathcal{L}_{u\bar{u}}$ and $\mathcal{L}_{d\bar{d}}$ for the leading order $u\bar{u}$ and $d\bar{d}$ sub-processes in accordance with the CT14HERA2 PDF set. Here too, the closed and open circles refer to the up-quark and down-quark parton luminosity functions and again, the ratio of $\mathcal{L}_{u\bar{u}}/\mathcal{L}_{d\bar{d}}$ is approximately 1.5. The product of these contributions (i.e., $P_u/P_d \times (\mathcal{L}_{u\bar{u}}/\mathcal{L}_{d\bar{d}})$) to the rates confirms the near factor of 4 ratio observed in Fig. 7 at $\cos \theta^*$ near 1.

[1] G. Aad *et al.* (ATLAS Collaboration), J. Instrum. **3**, P07007 (2008).

[2] S. Chatrchyan *et al.* (CMS Collaboration), J. Instrum. **3**, S08004 (2008).

[3] A. A. Alves, Jr. *et al.* (LHCb Collaboration), J. Instrum. **3**, S08005 (2008).

[4] S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, D. Stump, and C.-P. Yuan, Phys. Rev. D **93**, 033006 (2016).

[5] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, Eur. Phys. J. C **63**, 189 (2009).

[6] R. D. Ball, V. Bertone, S. Carrazza, C. S. Deans, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, J. I. Latorre, J. Rojo, and M. Ubiali, J. High Energy Phys. 04 (2015) 040.

[7] M. Aaboud *et al.*, J. High Energy Phys. 10 (2017) 182.

[8] A. M. Sirunyan *et al.* (CMS Collaboration), J. High Energy Phys. 06 (2018) 120.

[9] F. Abe *et al.* (CDF Collaboration), Phys. Rev. Lett. **74,** 850 (1995).

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

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

[12] V. M. Abazov *et al.* (D0 Collaboration), Phys. Rev. D **91,** 032007 (2015); **91,** 079901 (2015).

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

[14] G. Aad *et al.* (ATLAS Collaboration), Phys. Rev. D **85,** 072004 (2012).

[15] S. Chatrchyan *et al.* (CMS Collaboration), Phys. Rev. Lett. **109,** 111806 (2012).

[16] S. Chatrchyan *et al.* (CMS Collaboration), Phys. Rev. D **90,** 032004 (2014).

[17] R. Aaij *et al.* (LHCb Collaboration), J. High Energy Phys. 06 (2012) 058.

[18] C. P. Y. C. Schmidt, J. Pumplin, and P. Yuan, Phys. Rev. D **98,** 094005 (2018).

[19] V. Bertone *et al.* (xFitter Developers' Team), Proc. Sci. DIS2017 (**2018**) 203.

[20] J. C. Collins and D. E. Soper, Phys. Rev. D **16,** 2219 (1977).

[21] V. M. Abazovai *et al.* (DØ Collaboration), Phys. Lett. B **695,** 88 (2011).

[22] T. Aaltonen *et al.* (CDF Collaboration), Phys. Rev. Lett. **106,** 121801 (2011).

[23] G. Aad *et al.* (ATLAS Collaboration), Phys. Rev. D **90,** 052005 (2014).

[24] M. Aaboud *et al.* (ATLAS Collaboration), Phys. Lett. B **761,** 372 (2016).

[25] V. Khachatryan *et al.* (CMS Collaboration), J. High Energy Phys. 04 (2015) 025.

[26] V. Khachatryan *et al.* (CMS Collaboration), Phys. Lett. B **768,** 57 (2017).

[27] P. Langacker, Rev. Mod. Phys. **81,** 1199 (2009).

[28] D. London and J. L. Rosner, Phys. Rev. D **34,** 1530 (1986).

[29] J. Butterworth *et al.*, J. Phys. G **43,** 023001 (2016).

[30] S. D. Drell and T.-M. Yan, Phys. Rev. Lett. **25,** 316 (1970).

[31] J. Ellis, Comput. Phys. Commun. **210,** 103 (2017).

[32] G. A. Ladinsky and C.-P. Yuan, Phys. Rev. D **50,** R4239 (1994).

[33] C. Balázs and C.-P. Yuan, Phys. Rev. D **56,** 5558 (1997).

[34] F. Landry, R. Brock, P. M. Nadolsky, and C.-P. Yuan, Phys. Rev. D **67,** 073016 (2003).

[35] M. Aaboud *et al.* (ATLAS Collaboration), J. High Energy Phys. 12 (2017) 059.

[36] M. Boonekamp, F. Chevallier, C. Royon, and L. Schoeffel, Acta Phys. Pol. B **40,** 2239 (2009).

[37] 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. D **95,** 034003 (2017).

[38] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, J. High Energy Phys. 07 (2014) 079.

[39] V. Shalaev, I. Gorbunov, and S. Shmatov, EPJ Web Conf. **177,** 04010 (2018).

[40] H. Paukkunen and P. Zurita, J. High Energy Phys. 12 (2014) 100.