

Epistemic and nuclear uncertainties for the parton distributions of the pion

Lucas Kotz¹, Aurore Courtoy^{2,*}, Pavel Nadolsky^{3,†} and Maximiliano Ponce-Chavez³

¹*Department of Physics, Southern Methodist University, Dallas, Texas 75275-0175, USA*

²*Instituto de Física, Universidad Nacional Autónoma de México,
Apartado Postal 20-364, 01000 Ciudad de México, Mexico*

³*Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824, USA*



(Received 25 May 2025; accepted 16 September 2025; published 7 October 2025)

We present `FantO10`, a Hessian ensemble of next-to-leading-order error parton distribution functions (PDFs) in a charged pion, constructed to enable precision comparisons with upcoming experimental data and *ab initio* QCD predictions. `FantO10` estimates several experimental, theoretical, and methodological uncertainties on pion PDFs. It follows a statistical procedure, grounded in de Finetti's theorem, to systematically determine the component of the epistemic uncertainty associated with the choice of PDF functional forms. This uncertainty is estimated by fitting the data with multiple functional forms generated in a versatile format based on polynomial universal approximators (Bézier curves). The most distinct fits are converted into a joint Monte Carlo replica ensemble and then combined into one Hessian ensemble through a major update of the METAPDF methodology, expanded in this article to combine PDFs for arbitrary hadrons and partonic flavor representations. This study also goes beyond the previous pion fits by estimating, for the first time, the uncertainty from `nCTEQ15` nuclear PDFs for the tungsten target in the E615 pion-nucleus Drell-Yan data—a third-party input in our analysis, also included in the expanded METAPDF combination. The upgraded METAPDF framework, thus, altogether provides a principled algorithm to propagate parametrization dependence and other types of the epistemic uncertainty into conveniently represented PDFs. Notably, `FantO10` modifies the key physics conclusions from the earlier, more constrained analyses of the pion's gluon and sea-quark content. It facilitates studies of PDF-induced correlations and comparisons to lattice QCD. We summarize the methodology, structure of the `FantO10` PDFs, and experimental implications.

DOI: [10.1103/h2vn-1wxp](https://doi.org/10.1103/h2vn-1wxp)

Introduction. The structure of pions is a subject of active inquiry thanks to the advent of new experiments that use either pion beams [1] or virtual pion targets [2–4], a challenging task owing to the pion's instability and small mass. Although pions seemingly possess a simpler constituent quark structure compared to nucleons, their internal dynamics remains intriguing due to imprint of chiral symmetry breaking on formation of pseudo-Goldstone bosons. Key open questions include the gluon content of pions at a given resolution scale and the role of quarks and gluons in pion's mass emergence through a combination of QCD-driven and electroweak Higgs mechanisms. Insights into both questions can be gleaned from exploring the flavor composition and shapes of parton distribution

functions (PDFs) of the pion in high-energy collisions. To interpret these results, one must understand and quantify uncertainties on pion PDFs.

Recent analyses of pion PDFs offer a valuable testing ground for new methods in solving inverse problems, taking upon statistics issues of significant depth. The JAM Collaboration has led many an effort [5–9], accounting for data inputs from leading-neutron deeply inelastic scattering (DIS) and even first constraints from lattice QCD, as well as threshold resummation. The `xFitter` Collaboration published an independent Hessian-based fit [10] of the pion PDFs that served as a progenitor for our study. On the other hand, progress has been made in the context of nucleon PDF fits [11–14] toward quantification of uncertainties coming from the model and methodology, referred to as *epistemic*, as in the sense of *systematic*, uncertainties, as opposed to the contribution coming from *aleatoric* or *statistical* uncertainties. Recent pion analyses [5–10] also explored some aspects of quantification of systematic uncertainties yet leaving room for further advancements.

It is with the aim of achieving PDF determination with comprehensive uncertainties that our team determined the

* Contact author: aurore@fisica.unam.mx

† Contact author: nadolsky@msu.edu

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

pion PDFs $f_a(x, Q_0)$ along a new methodology, Fantômas, introduced to systematically explore their functional dependence on the momentum fraction x at the initial scale Q_0 of QCD evolution. Our first publication [12] documenting this method described a versatile representation for the PDF, called a metamorph, based on a polynomial approximator chosen to be a Bézier curve. Metamorphs have direct physics interpretation and reduce computational needs for generation and refitting of multiple PDF parametrizations when estimating the associated uncertainty, which contributes a large part, sometimes half or more [15], of the total PDF uncertainty. Traditional polynomial forms of fitted PDFs often can be converted into the standard format of metamorphs, and then multiple fits based on descendant functional forms can be automated and parallelized.

The Fantômas approach allows one to reduce the parametrization bias that arises when one fits only a small selection of functional forms [11]. Instead, we fit and combine multiple, equally acceptable PDF solutions [16,17]. We, the Fantômas team, have shown that consequently the uncertainty tends to increase with respect to other analyses for the pion PDFs based on restrictive parametrizations, in x intervals where there are little to no data. Such reduction of the parametrization bias, alternatively possible with neural-network PDF models [18,19], here is achieved with an interpretable polynomial representation. As a demonstration of possible physics implications, the metamorph-based fit [12] implemented in xFitter revealed a strong anticorrelation between the pion's sea and gluon distributions, weakening the earlier findings [5,8,10] of a large gluon content in the pion, as characterized by the gluon's momentum fraction.

In the follow-up to Ref. [12], several significant developments were made to publish a practical, minimally biased ensemble of pion PDF parametrizations as well as computing tools that produced them. We validated and released two versions of the C++ code implementing the Fantômas parametrization methodology, in the form of a stand-alone program, and as a module in the xFitter fitting framework [17,20].¹

To enable production of the current Hessian ensemble, we deeply revised the METAPDF toolkit [16] for the combination of PDF ensembles in order to handle PDFs for arbitrary flavor representations or provided as LHAPDF6 grids. Originally proposed in [16], the META analysis transforms all PDF solutions into a Monte Carlo (MC) replica ensemble and then converts it into a Hessian eigenvector ensemble. Until now, it supported only the proton case, which sufficed for its deployment for PDF4LHC PDF combinations [24,25].

¹Additional aspects of our pion fit, such as the exploration of constraints of individual datasets on pion PDFs obtained with the method of L_2 sensitivities [21,22], are documented in a Ph.D. thesis [23].

In addition to the bias-suppressed implementation of parametrizations via Bézier curves, we extended the METAPDF algorithm to include, for the first time, propagation of uncertainty from nuclear PDFs relevant to the Drell-Yan pion-tungsten scattering data. This enables systematic propagation of a broader class of third-party inputs for multiple unrelated uncertainties, e.g., from nuclear PDFs or fragmentation functions, into the final Hessian error PDF ensemble.

This article aims, first, to summarize a statistically justified combination methodology, grounded in de Finetti's theorem, for merging ensembles of PDFs originating from distinct parametrizations and external inputs. We then detail the steps of the METAPDF conversion from MC to Hessian replicas and explain our approach for including the nuclear PDF uncertainty. We conclude by reviewing physics implications of the new Fantô10 PDFs with the nuclear uncertainty.

The probability of models emerges from their sampling. As the epistemic probability on the space of functional forms of PDFs is unknown *a priori*, the philosophy of the Fantômas method is to learn it incrementally through representative sampling [11] of PDF models according to several guiding principles. First, each accepted PDF fit must describe well the fitted data, with less agreeing models disfavored in the final combination. Second, simpler models are preferred under otherwise equal conditions to facilitate generalization for future predictions according to information criteria [26] and their realizations in Bayesian model averaging for non-perturbative QCD quantities [27,28]. Third, models selected for the final combination must be representative of the available ensemble of good solutions. For example, when combining models with the same goodness of fit, it makes sense to keep diverse, relatively equidistant models according to their separation in physical PDF space, as clustering of the models in one corner of parameter space is not expected from their prior probability. Fourth, models must be conditionally independent and exchangeable when included in the combination. In this case, and in line with de Finetti's representation theorem for exchangeable sequences [29], PDF ensembles entering the final combination can be viewed as sampled from some epistemic probability distribution associated with an unknown latent parameter θ assigned to each sampled model. The final PDF error band then accounts for the aleatoric uncertainty, associated with random fluctuations of data as well as the emergent epistemic one that reflects sampling over the explored PDF models.

Model combination in the Fantô10 fit. We illustrate application of these general principles according to a simplified procedure that is sufficient at the current, still low, accuracy of pion data.² Reference [12] details settings

²Refinements in the combination procedure are planned for precise nucleon fits.

of our pion fit. It is performed at the next-to-leading order (NLO) in QCD and introduces three independent parametrizations for the valence, SU(3)-symmetric quark sea, and gluon distributions, denoted by $xV(x)$, $xS(x)$, and $xg(x)$, respectively, and given by a metamorph with up to four free parameters each. These PDFs are constrained by fixed-target measurements of Drell-Yan (DY) pair and direct photon production, covering $x \gtrsim 0.07$ and $Q \lesssim 15$ GeV, and by 29 points from leading-neutron DIS, covering $8.16 \times 10^{-4} < x_\pi < 0.0765$. Among these, the E615 and NA10 datasets on pion-tungsten DY process dominate the large- x constraints, and, hence, the pion PDFs are sensitive to the assumed nuclear ones, which we take from the nCTEQ15 NLO ensemble [30].

We start by obtaining about 100 candidate fits to pion data with varied PDF parametrization forms (cf. Ref. [12]) and for the same central ensemble of nCTEQ15FullNuc_184_74 nuclear PDFs, referring to a tungsten-184 target, and nCTEQ15FullNuc_1_1 for the proton baseline in WA70 direct photon production. We then rank fits or models M according to an objective function defined as $\chi^2 = \chi_{D|M}^2 + \chi_M^2$, where $\chi_{D|M}^2 \equiv -2 \ln(P(D|M))$ is the log-likelihood for the fitted pion data D given M and $\chi_M^2 \equiv -2 \ln(P(M))$ represents the possible contribution from priors constraining M . In the context of the Fantômas fit, M corresponds to specific settings of the metamorphs, and χ_M^2 may be imposed to prevent runaway free parameters in some fits given the paucity of data, e.g., due to the insufficiently constrained asymptotic power B_S of the sea PDF $xS(x, Q_0)$, growing as x^{B_S} at $x \rightarrow 0$; see Sec. III in [12].

Among ~ 100 acceptable models, we find the best fit corresponding to the χ^2 minimum, $\chi_0^2 \equiv \min_{\{M\}} \chi_{D|M}^2$, and keep other solutions satisfying

$$\chi_{D|M}^2(M) \leq \chi_0^2 + \sqrt{2(N_{\text{pts}} - N_{\text{par}})}, \quad (1)$$

where $\sqrt{2(N_{\text{pts}} - N_{\text{par}})}$ is one standard deviation on $\chi_{D|M}^2$ with $N_{\text{pts}} - N_{\text{par}}$ degrees of freedom. In the current analysis, we selected $\tilde{N} = 5$ final models, where $\tilde{N}$ refers to the number of models, based on Eq. (1) and *ad hoc* metric estimate to select the most diverse shapes. The $\tilde{N}$ models were then combined using the METAPDF technique with the updated MCGEN [31] and MP4LHC packages [16] adapted for the pion quark contents. In a nutshell, Hessian uncertainties were generated for each model according to the $\Delta\chi^2 = 1$ criterion [32] and then converted into MC replicas, which were then bundled, in an unweighted way, to form the combined MC ensemble. The latter, in turn, was converted into a Hessian ensemble called Fantô10_15, with 15 eigenvector sets (one set per eigenvector direction). Figure 1 displays the resulting

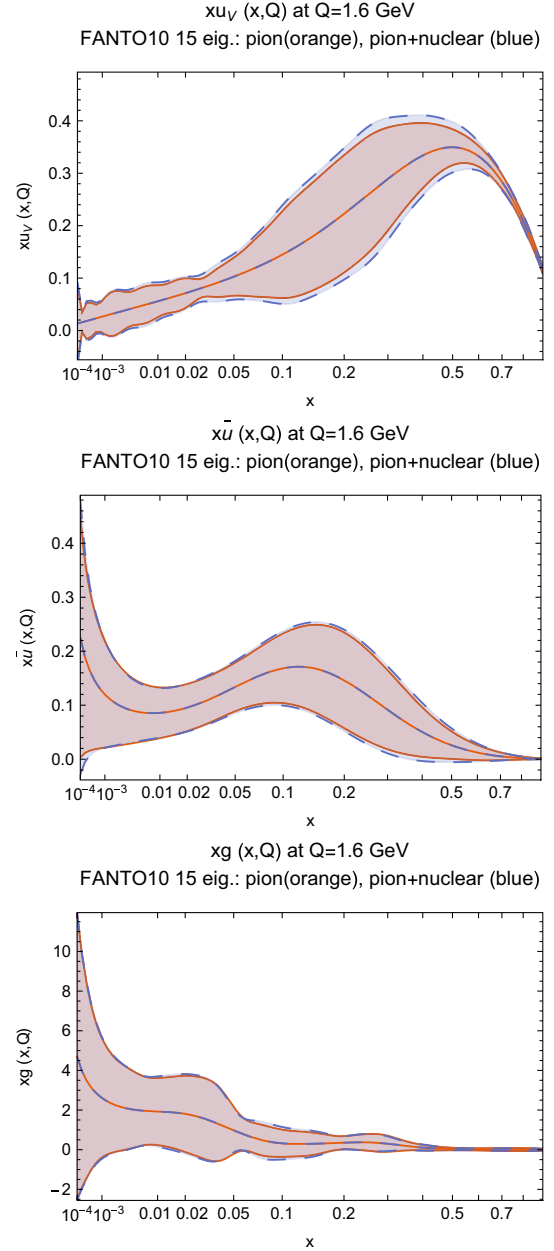


FIG. 1. The aleatoric + epistemic (orange) and aleatoric + epistemic + nuclear (blue) uncertainty bands for the Fantô10 set at $Q = 1.6$ GeV, with $\alpha_s(M_Z) = 0.118$.

PDFs and uncertainties for three flavor combinations—valence $xu_V = xV/2$, $x\bar{u} = xS/6$, and gluon g —as solid orange bands.

As a resampling method, the METAPDF combination ensures that the MC PDFs are identically generated for all models' outputs. The $\tilde{N}$ models are exchangeable (the combination does not depend on the order in which the models are added), while the Fantômas methodology guarantees that all input Hessian sets are converted into error PDF bundles according to an identical procedure. PDF inference in the Fantômas framework, thus,

provides an ensemble of independent and identically distributed MC sets, and as a result the Fantômas methodology, followed by the combination via the METAPDF method, renders an estimate of epistemic uncertainty that improves as more functional forms are explored. The presented implementation sufficient for pion fits, which nevertheless goes beyond the previous results utilizing fixed parametrization forms, can be further augmented with features already deployed in precision fits, such as closure tests [33], cross validation, and expanded goodness-of-fit criteria [34].

Including nuclear uncertainty in pion PDFs. The Fantômas framework can handle other types of uncertainties, like the one coming from nuclear PDFs in the key Drell-Yan datasets from E615 and NA10. Given their size of the dataset and impact, studies of nuclear PDFs (nPDFs) [30,35–37] are currently far more phenomenologically mature and precise than those of pion PDFs. We find that the impact of pion data on the employed nPDF ensemble, here taken to be nCTEQ15 NLO [30], is a small perturbation. One can capture dominant constraints of the pion data on nPDFs by PDF reweighting. When estimating the total (pion + nuclear) uncertainty in Fantô10, we treat the nCTEQ15 uncertainty as its authors suggest, namely, that the nCTEQ15 Hessian eigenvector sets characterize an approximately normal probability distribution with respect to nuclear PDF parameters in the vicinity of the best fit given by the central nCTEQ15 set, at 1σ .

Thus, under our central assumption that the pion parametrization and nuclear uncertainties are independent, we combine the two uncertainties using the METAPDF technique [16] with some additional steps.

- (1) The central lines and orange-band results shown in Fig. 1 correspond to using the Drell-Yan cross section with the central PDF set (replica 0, $f^{(0)}$) of the nPDF ensemble. The central prediction is obtained by averaging the central predictions of $\tilde{N}$ PDF models, and the error bands are obtained as a bundle of $\tilde{N}$ respective error bands.
- (2) Next, to estimate the nPDF uncertainty, we generate $N_{\text{nMC}} = 100$ nuclear MC (nMC) PDF replicas from $2D + 1$ member sets of the nCTEQ15 Hessian ensemble [16,31]. A k th replica, $f^{(k)}$, is computed by drawing a random vector $\vec{R}^{(k)} = \{R_1^{(k)}, R_2^{(k)}, \dots, R_D^{(k)}\}$ from a D -dimensional standard normal distribution:

$$f^{(k)} = f^{(0)} + d^{(k)} - \Delta, \quad \text{where} \quad (2)$$

$$d^{(k)} \equiv \sum_{i=1}^D \left(\frac{f_{+,i}^{(k)} - f_{-,i}^{(k)}}{2} R_i^{(k)} + \frac{f_{+,i}^{(k)} + f_{-,i}^{(k)} - 2f^{(0)}}{2} \left(R_i^{(k)} \right)^2 \right), \quad (3)$$

and optional $\Delta \equiv \langle d \rangle$, with $\langle \cdot \rangle$ the average over all replicas [31].

- (3) We repeat the pion *central* fit for N_{nMC} nPDF replicas and $\tilde{N}$ pion parametrization models that were selected when fitting with nCTEQ15 replica 0. From the resulting $N_{\text{nMC}} \times \tilde{N} = 500$ replicas, we keep only 396 that satisfy $\chi^2 < \chi_0^2 + \sqrt{2(N_{\text{dat}} - N_{\text{par}})}$, similarly to the criteria used by reweighting techniques [38,39]. These central fits quantify fluctuations in the nuclear PDFs. Since the fitting procedure, here in xFitter, imposes physical constraints such as sum rules, we check that integrals of every generated PDF over the covered interval $x \in [10^{-3}, 1]$ comply with the momentum sum rule at the level of about 1% and valence sum rule at the level better than 10%.
- (4) Next, we add aleatoric and nuclear uncertainties in each of $\tilde{N}$ baseline pion models M_i . The former is taken to be the same as in the baseline fit with the central nCTEQ15 replica (found using the $\Delta\chi^2 = 1$ prescription). The latter is computed from variations of central fits for M_i when the nCTEQ15 replica 0 is replaced by an alternative nMC replica generated in step 3. In practice, when generating the METAPDF MC replicas from M_i , we replace Eq. (2) by

$$f^{(k)} = f_{\text{nMC}|M_i}^{(0,k)} + d^{(k)} - \Delta, \quad (4)$$

with $f_{\text{nMC}|M_i}^{(0,k)}$ being a randomly drawn nMC central replica (different for each k) from the ones generated for model M_i . With the rest of MC generation staying the same, fluctuations of $f^{(k)}$ in M_i are now enhanced by random contributions from the nMC uncertainty transmitted via $f_{\text{nMC}|M_i}^{(0,k)}$.

- (5) The new MC sets for the $\tilde{N}$ models can be now combined and converted into a Hessian set with the METAPDF algorithm, following [16].

The resulting Hessian PDFs, Fantô10_n15, now include the nuclear uncertainty through the shifts of the input MC replicas in the previous steps. For the representative PDFs in Fig. 1, the Fantô10_n15 error bands are shown in blue color. The uncertainty coming from the convolution with nPDFs is moderate but not entirely negligible compared to the intrinsic fitting uncertainty shown in orange, especially for the valence $xV(x, Q)$ that is better constrained. For xV , we can better visualize the uncertainties without and with the nuclear component by plotting them as ratios to the central PDF set in Fig. 2. As the sea and gluon PDFs, xS and xg , remain poorly known, their nuclear uncertainty can be neglected for the time being. Repeating a central pion fit by replacing the nCTEQ15 tungsten PDFs with EPPS21 [36] ones resulted in a negligible difference compared to the current full error bands.

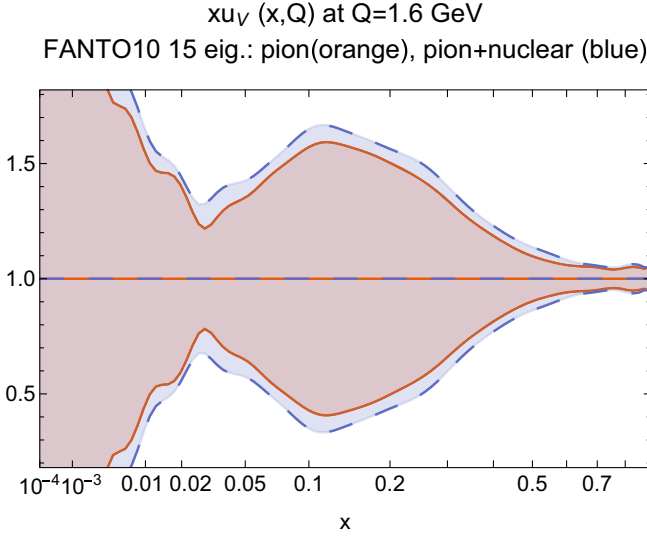


FIG. 2. Aleatoric + epistemic (orange) and aleatoric + epistemic + nuclear (blue) uncertainty bands for the valence up quark PDF at $Q = 1.6$ GeV, normalized to the respective central PDF.

Phenomenology implications. The compact Hessian representation of the Fanto10 facilitates phenomenological studies, such as computing PDF-induced correlations [40–42] both without and with the nuclear uncertainty. Among these, we confirm a strong anticorrelation between the gluon and sea PDFs and their momentum fractions, as depicted in the upper panel in Fig. 3. Since the dominant available data stem from πN Drell-Yan pair production, which is primarily sensitive to a mix of quark and antiquark contributions, the gluon and sea-quark components remain entangled, with no clear sensitivity pattern currently allowing for their separation [23]. With Fanto10_15 Hessian PDFs, we can trace how the anticorrelation originates from specific intervals of x . The lower panel in Fig. 3 depicts a heat map of the correlation cosine [42] between $xS(x_S, Q)$ and $xg(x_g, Q)$ in the plane of respective momentum fractions x_S and x_g . Reference [43] collects plots of correlation cosines for other PDF pairs. While the gluon at $x_g > 0.2$ is moderately correlated with xS at all x_S , we see a high anticorrelation between xg and xS at $x_g < 0.2$, producing the observed anticorrelation between momentum fractions $\langle xS \rangle$ and $\langle xg \rangle$ in the upper panel in Fig. 3.

This degeneracy reflects the weakness of experimental constraints on the gluon and sea, with even zero gluon contents still allowed by the data at the low scale of the order of $Q = 1\text{--}2$ GeV. (See the detailed discussion in [12].) The degeneracy could be lifted with additional data at $x < 0.1$ and from a broader range of processes—see, e.g., [2,4,5]—thereby helping to answer the long-standing question of the pion’s gluon content. Looking forward, COMPASS++’s range in x will extend that on the old NA10

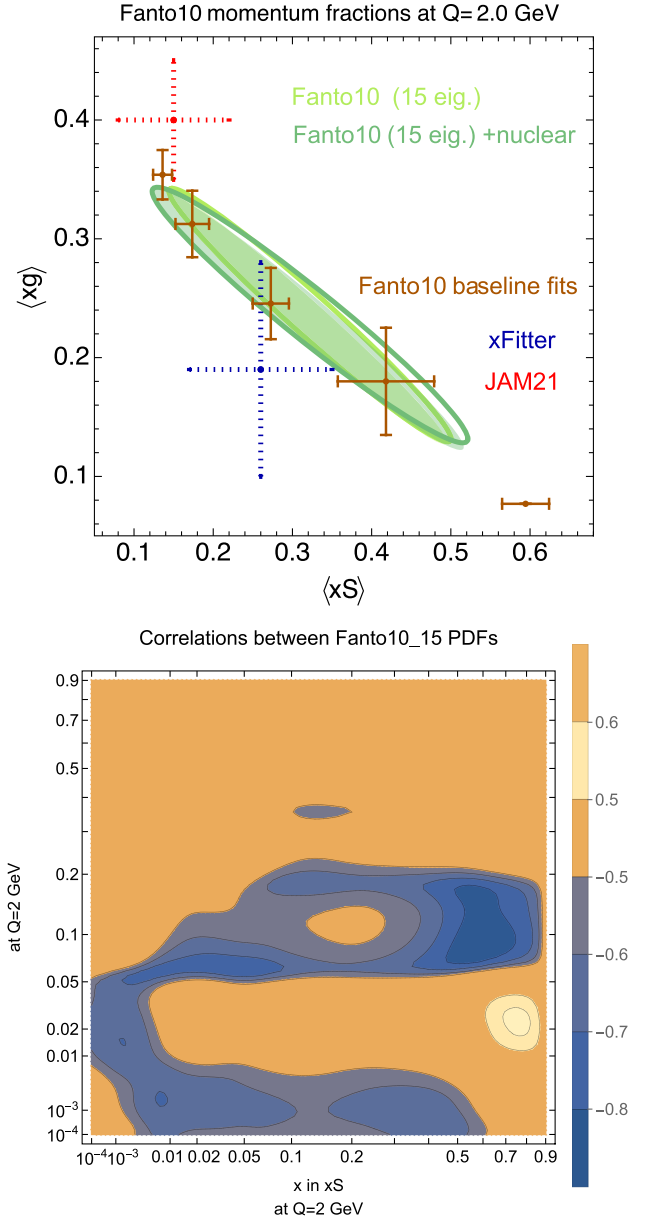


FIG. 3. Top: pion sea and gluon momentum fractions obtained with five Fantomas baseline fits (brown error bars) and combined Fanto10_15 and Fanto10_n15 ensembles at 1σ level. We also show respective results from the JAM21 [7] (in red) and xFitter (in blue) studies [10]. See discussion in [12]. Bottom: the correlation cosine [42] between $xS(x_S, Q)$ and $xg(x_g, Q)$ for Fanto10_15 PDFs. We use $Q = 2$ GeV in both panels.

and E615 sets, possibly covering the mid- x region [1,44]. The Electron-Ion Collider may provide valuable insights via leading-neutron DIS [3], although such measurements will require reliable theoretical input for the pion flux in the proton. Jefferson Lab, operating at 22 GeV, could also contribute to this effort, albeit with different nuclear uncertainties to be considered [4]. Less conventional

approaches include studies involving J/ψ production, using historical data from experiments such as NA3 or E0705 [45]. While Drell-Yan data prefer a small gluon content in the pion for NLO global analyses, elevating the pion fits to the $\mathcal{O}(\alpha_s^2)$ (NNLO) accuracy may change this trend, as the pQCD corrections of this order to fixed-target DY process are known to be large. Some lattice evaluations report $\langle xg \rangle_\pi > 30\%$ (e.g., [46]), just like claimed by the JAM global analysis [7]. Only at NNLO may the detailed interplay between the pion and nuclear uncertainties become relevant.

Conclusions. In this paper, we have presented the final Hessian ensembles of NLO pion PDFs, `Fanto10_15` and `Fanto10_n15`, determined using the `Fantômas` methodology. Our framework introduces a statistically motivated model combination to account for the uncertainty arising from the inference of the PDF parametrization [12,47] and other epistemic uncertainties. We have further extended the analysis by incorporating an uncorrelated, third-party source of uncertainty propagated from the `nCTEQ15` ensemble of nuclear PDFs describing fitted cross sections from the pion-tungsten Drell-Yan process.

The described methodology is broadly applicable to inverse problems involving multiple nonperturbative inputs. Examples include semi-inclusive deep inelastic scattering, where PDFs are convoluted with fragmentation functions, or leading-neutron DIS, where uncertainties in the pion flux must be taken into account. Our approach may

also prove useful in the determination of nuclear PDFs themselves, where the baseline proton PDF typically lacks an associated uncertainty [35].

Interpolation grids for the `Fanto10_15` and `Fanto10_n15` ensembles, without and with the nuclear uncertainty, respectively, can be downloaded from the CTEQ-TEA website [43] and LHAPDF library [48]. We use $\alpha_s(M_Z) = 0.118$, and the grids cover the range of $1.3 \leq Q \leq 1000$ GeV and $0.0001 \leq x \leq 1$, for the maximal active flavor number $N_f = 5$. The supplemental material about `Fanto10` ensembles includes an extended collection [43] of plots of PDF-PDF correlations (similar to the lower Fig. 3) and an examination of L_2 sensitivities [22] of fitted experiments to `Fanto10_15` PDFs in Chapter 4 of [23].

Acknowledgments. We thank T. Hobbs, F. Olness, and CTEQ-TEA colleagues for various discussions. This study has been financially supported by a National Science Foundation AccelNet project, by the UNAM Grants No. DGAPA-PAPIIT IN111222 and No. IN102225, and by the U.S. Department of Energy under Grant No. DE-SC0010129. P. N. is grateful for support from Wu-Ki Tung Endowed Chair in particle physics.

Data availability. The data that support the findings of this article are openly available [43].

-
- [1] V. Andrieux, Pion-induced Drell-Yan cross-section measurements at COMPASS, dIS2024 workshop.
 - [2] R. Abdul Khalek *et al.*, *Nucl. Phys. A* **1026**, 122447 (2022).
 - [3] J. Arrington *et al.*, *J. Phys. G* **48**, 075106 (2021).
 - [4] A. Accardi *et al.*, *Eur. Phys. J. A* **60**, 173 (2024).
 - [5] P. C. Barry, N. Sato, W. Melnitchouk, and C.-R. Ji, *Phys. Rev. Lett.* **121**, 152001 (2018).
 - [6] N. Y. Cao, P. C. Barry, N. Sato, and W. Melnitchouk (Jefferson Lab Angular Momentum (JAM) Collaboration), *Phys. Rev. D* **103**, 114014 (2021).
 - [7] P. C. Barry, C.-R. Ji, N. Sato, and W. Melnitchouk (Jefferson Lab Angular Momentum (JAM) Collaboration), *Phys. Rev. Lett.* **127**, 232001 (2021).
 - [8] P. C. Barry *et al.* (Jefferson Lab Angular Momentum (JAM) Collaboration and HadStruc Collaboration), *Phys. Rev. D* **105**, 114051 (2022).
 - [9] W. Good, P. C. Barry, H.-W. Lin, W. Melnitchouk, A. NieMiera, and N. Sato, *arXiv:2507.22730*.
 - [10] I. Novikov *et al.*, *Phys. Rev. D* **102**, 014040 (2020).
 - [11] A. Courtoy, J. Huston, P. Nadolsky, K. Xie, M. Yan, and C. P. Yuan, *Phys. Rev. D* **107**, 034008 (2023).
 - [12] L. Kotz, A. Courtoy, P. Nadolsky, F. Olness, and M. Ponce-Chavez, *Phys. Rev. D* **109**, 074027 (2024).
 - [13] B. Kriesten and T. J. Hobbs, *Phys. Rev. D* **111**, 014028 (2025).
 - [14] M. N. Costantini, M. Madigan, L. Mantani, and J. M. Moore, *J. High Energy Phys.* **12** (2024) 064.
 - [15] T.-J. Hou *et al.*, *Phys. Rev. D* **103**, 014013 (2021).
 - [16] J. Gao and P. Nadolsky, *J. High Energy Phys.* **07** (2014) 035.
 - [17] L. Kotz, A. Courtoy, T. J. Hobbs, P. Nadolsky, F. Olness, M. Ponce-Chavez, and V. Purohit, *arXiv:2507.22969*.
 - [18] R. D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, J. Rojo, and M. Ubiali (NNPDF Collaboration), *Nucl. Phys. B* **838**, 136 (2010).
 - [19] R. D. Ball *et al.* (NNPDF Collaboration), *J. High Energy Phys.* **04** (2015) 040.
 - [20] `FantômasQCD` online repository, <https://gitlab.com/cteq-tea/public/fantomas4qcd>.
 - [21] T. J. Hobbs, B.-T. Wang, P. M. Nadolsky, and F. I. Olness, *Phys. Rev. D* **100**, 094040 (2019).
 - [22] X. Jing *et al.*, *Phys. Rev. D* **108**, 034029 (2023).
 - [23] L. Kotz, A novel approach to model the pion structure using advanced polynomial functions, Ph.D. thesis, Southern Methodist University (SMU), Dallas, TX, 2025.
 - [24] J. Butterworth *et al.*, *J. Phys. G* **43**, 023001 (2016).

- [25] R. D. Ball *et al.* (PDF4LHC Working Group), *J. Phys. G* **49**, 080501 (2022).
- [26] H. Akaike, *IEEE Trans. Autom. Control* **19**, 716 (1974).
- [27] E. T. Neil and J. W. Sitison, *Phys. Rev. D* **109**, 014510 (2024).
- [28] M. Yan, T.-J. Hou, Z. Li, K. Mohan, and C. P. Yuan, *Phys. Rev. D* **112**, 034008 (2025).
- [29] B. De Finetti, *Theory of Probability* (Wiley, New York, 1974), Vol. 1–2.
- [30] K. Kovarik *et al.*, *Phys. Rev. D* **93**, 085037 (2016).
- [31] T.-J. Hou *et al.*, *J. High Energy Phys.* 03 (2017) 099.
- [32] J. C. Collins and J. Pumplin, [arXiv:hep-ph/0105207](#).
- [33] L. A. Harland-Lang, T. Cridge, and R. S. Thorne, *Eur. Phys. J. C* **85**, 316 (2025).
- [34] K. Kovarik, P. M. Nadolsky, and D. E. Soper, *Rev. Mod. Phys.* **92**, 045003 (2020).
- [35] M. Klasen and H. Paukkunen, *Annu. Rev. Nucl. Part. Sci.* **74**, 49 (2024).
- [36] K. J. Eskola, P. Paakkinen, H. Paukkunen, and C. A. Salgado, *Eur. Phys. J. C* **82**, 413 (2022).
- [37] R. Abdul Khalek, R. Gauld, T. Giani, E. R. Nocera, T. R. Rabemananjara, and J. Rojo, *Eur. Phys. J. C* **82**, 507 (2022).
- [38] R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, J. Rojo, and M. Ubiali (NNPDF Collaboration), *Nucl. Phys.* **B849**, 112 (2011); **B854**, 926 (E) (2012); **B855**, 927(E) (2012).
- [39] R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, J. I. Latorre, J. Rojo, and M. Ubiali, *Nucl. Phys.* **B855**, 608 (2012).
- [40] J. Pumplin, D. Stump, R. Brock, D. Casey, J. Huston, J. Kalk, H. L. Lai, and W. K. Tung, *Phys. Rev. D* **65**, 014013 (2001).
- [41] P. M. Nadolsky and Z. Sullivan, eConf **C010630**, P510 (2001), [arXiv:hep-ph/0110378](#).
- [42] 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).
- [43] CTwebsite, CTEQ-TEA parton distribution functions, <https://cteq-tea.gitlab.io/project/00pdfs/#mesonPDFs>.
- [44] C.-Y. Hsieh, Measurement of pion-induced Drell-Yan cross section from COMPASS data, 12th Workshop on Hadron Physics and Opportunities Worldwide (2024).
- [45] W.-C. Chang, J.-C. Peng, S. Platchkov, and T. Sawada, *Phys. Rev. D* **102**, 054024 (2020).
- [46] C. Alexandrou *et al.* (Extended Twisted Mass Collaboration), *Phys. Rev. Lett.* **134**, 131902 (2025).
- [47] A. Courtoy, *Proc. Sci. QNP2024* (**2025**) 050 [[arXiv:2408.03318](#)].
- [48] LHAPDF library, <https://lhapdf.hepforge.org/>.