QCD analyses for phenomenological PDF determination with comprehensive uncertainties

Bridging Theory and Experiment at the Electron-Ion Collider

BNL-INT Workshop @INT 06/02/25

Aurore Courtoy

Instituto de Física National Autonomous University of Mexico (UNAM)

In collaboration with CTEQ-TEA collab., Fantômas team, and students.





Dirección General de Asuntos del Personal Académico



CTEQ (Wu-Ki Tung Et Al.) group

From low to high energies

PDF determination is twofold:

- Iearned from data knowledge on low-energy dynamics and pQCD expansion
- used for precision physics at higher energies

The EIC may contribute to <u>learning</u> unpolarized PDFs – paradigm for this talk.



Kinematical coverage for collinear PDFs

A bit outdated...



[Prog.Part.Nucl.Phys.121]

so very small $x - x < 10^{-6}$.

Solve a large *x*: $x \ge 0.2$, includes the valence region.

Extrapolation regions for unpolarized PDFs

Both regions involve physics at the end-points -non-perturbative and more effects.



Parton Distribution Functions as an inverse problem

Parton Distribution Functions: are determined from data through an inverse problem.



Image from: del Debbio, SciPost Phys. Proc. 15, 028 (2024)

Parton Distribution Functions as an inverse problem

Parton Distribution Functions: are determined from data through an inverse problem.



Image from: del Debbio, SciPost Phys. Proc. 15, 028 (2024)

Reasons for complexity of inverse problems

- direct products
- convolutional problems
- Hausdorff moment problem

most common for PDF analyses

Ill-posed problem that accepts no unique solution: determining PDFs will involve <u>defining a set of solutions</u>.

A. Courtoy-IFUNAM_

UQ for pheno PDFs

CT18 unpolarized PDFs

CTEQ (Wu-Ki Tung Et Al.) group

CT18NNLO is the general-purpose PDF set published in 2019.

CT methodology is based on minimizing a χ^2 expressed in terms of *parametrizations* for the PDFs, finding the global minimum, and providing Hessian error PDFs to estimate the uncertainty.

CT25 in the making — stay tuned!



[[]Hou et al, Phys.Rev.D 103 (2021)]

Uncertainty Quantification in CT involves a two-tier analysis

Fixed tolerance criterion T^2 – beyond the $\Delta \chi^2 = 1$ prescription Experiment-based penalty

encompasses both the propagation of experimental and methodological uncertainties

CT18 unpolarized PDFs

CTEQ (Wu-Ki Tung Et Al.) group

CT18NNLO is the general-purpose PDF set published in 2019.

CT methodology is based on minimizing a χ^2 expressed in terms of *parametrizations* for the PDFs, finding the global minimum, and providing Hessian error PDFs to estimate the uncertainty.

CT25 in the making — stay tuned!



[[]Hou et al, Phys.Rev.D 103 (2021)]

Uncertainty Quantification in CT involves a two-tier analysis

Fixed tolerance criterion T^2 – beyond the $\Delta \chi^2 = 1$ prescription Experiment-based penalty

encompasses both the propagation of experimental and methodological uncertainties

UQ glossary	aleatoric/statistical uncertainties — fixed once for all				
	epistemic/systematic	uncertainties - e.g. from	methodology, can be reduced		
	A. Courtoy—IFUNAM	UQ for pheno PDFs	BNL-INT Workshop 2025		

CT18 unpolarized PDFs

CTEQ (Wu-Ki Tung Et Al.) group



Tools to explore and understand the spread of uncertainties in PDF analyses

The Hessian formalism allows for unique tools to visualize pulls from various experiments.

> [Wang et al, PRD98] [Hobbs et al., PRD100]

Comparative analysis for ATLASpdfs, CT18 and MSHT20 (+aN3LO)

[Jing et al., PRD108]





Parametrizations to tackle inverse problems

Two main approaches to parametrization in global analyses:



A. Courtoy—IFUNAM____

UQ for pheno PDFs

___BNL-INT Workshop 2025

Parametrizations to tackle inverse problems

Two main approaches to parametrization in global analyses:

I. use an explicit functional form — 20-50 parameters ← e.g., CT
 II. use neural networks — 20-50 hyperparameters ← e.g., NNPDF

The rôle of parametrization form in global analyses can be quantified

Fantômas4QCD

A new c++ code automates series of fits using <u>multiple</u> functional forms, called metamorph.

Just like *neural networks*, these polynomial functional forms can approximate any arbitrary PDF shape.

> [Kotz, AC, Nadolsky, Olness, Ponce-Chavez, PRD109] [Kotz, AC, Nadolsky, Ponce-Chavez, 2505.13594] [Kotz, AC, Hobbs, Nadolsky, Olness, Ponce-Chavez, Purohit, in progress]



Tolerance and parametrization studies



Uncertainty band reproduced, in part, by the spread of possible solutions

- → epistemic uncertainty
- → builds the $T^2 = \Delta \chi^2 > 1$ criterion

We (CT) are looking into information criteria to quantify the tolerance encompassing multiple sources of uncertainties.

Fantômas unlocks the concept of tolerance:

Solution with respective $\Delta \chi^2 = 1$ uncertainty can be bundled into a $\Delta \chi^2 > 1$ error band. Solution of constraints contributions

Regression for data-based analyses



as a function of the variables {x} and free parameters {a}

The theory input depends on the PDFs, whose parametrization is an input to the minimization procedure. The comparison to data for various parametrizations can lead to equally good χ^2 values.

Regression for data-based analyses



as a function of the variables {x} and free parameters {a}

The theory input depends on the PDFs, whose parametrization is an input to the minimization procedure. The comparison to data for various parametrizations can lead to equally good χ^2 values.

F₂(x, Q²) F₂(x, Q²) μ beam energy E = 100 GeV μ beam energy E = 100 GeV xFitte That's fine in the data region, but the results may vary greatly outside - extrapolation region. Leading Neutron DIS Data Q² = 11 - Leading Neutron DIS Data δ uncorrelated δ uncorrelated δ total δ total - - - Theory + shifts Theory + shifts Why not adopt more than one form? Theory/Data Theory/Data 1.2 1.2 0.6 0.002 0.01 0.02 0.002 0.01 0.02

A. Courtoy-IFUNAM_

UQ for pheno PDFs

BNL-INT Workshop 2025

Generation of parametrizations



Generation of parametrizations



Bézier curve characterized by **control points**, vector of $\mathscr{B} \rightarrow \underline{P}$:

$$\underline{P} = \underline{\underline{T}} \cdot \underline{\underline{M}} \cdot \underline{C}$$

matrix of x^l at $\{x_{CP}\}$

[AC & Nadolsky, PRD103] [Kotz, et al. PRD109]

Generation of parametrizations



Bézier-curve methodology - toy model



A. Courtoy—IFUNAM

UQ for pheno PDFs

BNL-INT Workshop 2025

Sandbox studies

Small(er) scale problems to develop and test the methodology, and learn more physics

 \odot pion PDFs: Fanto π PDFs — including model combination and UQ from nPDFs

- now released: Fantômas1.0 PDFs soon on LHAPDF
- developping: beta-testing the code for release
- developping: information-ish criteria for model selection
- twist-3 collinear PDF integral constraints and interpolation
 developping: information-ish criteria for model selection





Fantômas *π*PDFs

- \Rightarrow We generated $N \sim 100$ fits corresponding to N sets for $\{N_m, \underline{P}, \alpha_x\}$.
- ⇒ Well-behaved (convergence + fixed soft constraints) fits are kept.

$$\Rightarrow$$
 Fits within $\chi^2 + \delta \chi^2 = \chi^2 + \sqrt{2(N_{\text{pts}} - N_{\text{par}})}$ are kept.

The final bundle is generated from the 5 most diverse shapes at Q_0 .

 π^+ (MC) PDFs at Q=1.4 GeV, 68% c.l. (band)



Fantômas *m*PDFs

- \Rightarrow We generated $N \sim 100$ fits corresponding to N sets for $\{N_m, \underline{P}, \alpha_x\}$.
- ⇒ Well-behaved (convergence + fixed soft constraints) fits are kept.

$$\Rightarrow$$
 Fits within $\chi^2 + \delta \chi^2 = \chi^2 + \sqrt{2(N_{\text{pts}} - N_{\text{par}})}$ are kept.

The final bundle is generated from the 5 most diverse shapes at Q_0 .

 $\tilde{N} = 5$ exchangeable final independent solutions, from which we generate identically distributed MC distributions

- they can be interpreted as being drawn from a common distribution, characterized by a latent parameter θ .

 π^+ (MC) PDFs at Q=1.4 GeV, 68% c.l. (band)



Statistical justification of our model combination by de Finetti's theorem.

Likelihood-ratio test

Independent contributions to uncertainty:

the parametrization contributes to the (log)-likelihood but constraints on the parameters, ..., contribute to the prior.

$$\chi^2_{\rm tot} = \chi^2 + \chi^2_{\rm prior}$$

$$P(a|D) \propto P(D|a) P(a)$$

$$\Leftrightarrow \quad \exp(-\chi_{\text{tot}}^2) \propto \exp(-\chi^2) \exp(-\chi_{\text{prior}}^2)$$

On which basis are PDFs accepted or rejected?

Likelihood ratios:

two replicas can be ordered according to their relative likelihood or relative prior.



Model combination



[Gao & Nadolsky, JHEP07]

A. Courtoy-IFUNAM_

UQ for pheno PDFs

BNL-INT Workshop 2025

Criterion for shape selection

So far, we've used an ad hoc criterion for the selection of the most diverse shapes.

In progress: automatize the selection based on shapes and use of information criteria — likelihood-ratio test and quantititative criteria



Third-party uncertainty propagation

For processes with more than one unknown function:

SIDIS — fragmentation and PDF Drell-Yan — PDF and nuclear PDF

•••

Prescription to propagate the uncertainty for Hessian-based input. It uses mcgen, model combination and selection criteria.

Here:

nCTEQ15 nuclear uncertainties on DY is propagated. Small effect.









Pion momentum fractions

[Kotz, et al., 2505.13594]



"Our selection of the" data supports a zero-gluon solution, while it has a hard time finding solutions leading to higher gluon momentum fractions.

Data sets: Drell-Yan (~65%) + prompt-photon (~25%) and leading-neutron DIS (~10%) — entanglement/ anticorrelation between sea and gluon distributions. Differs from JAM's pool.

UQ for pheno PDFs

Moments from the lattice

	Name	$Q[{ m GeV}]$	$\langle x(u+ar{u})_{\pi^+} angle$	$\langle xg angle$
	FantoPDF	2	0.331(25)	0.24(10)
	HadStruct [19]	2	0.2541(33)	_
[Gao et al., PRD102]		3.2	0.216(19)(8)	_
	ETM [46]	2	0.261(3)(6)	-
	ETM [91]	2	$0.601(28) _{u+d}$	0.52(11)
ا [Meyer et al., PRD77]		2	_	0.37(8)(12)
[Shanahan et al., PRD99]		2	-	0.61(9)
[MSU, 2310.12034]		2	_	0.364(38)(36)
	ZeRo Coll. [95]	2	0.245(15)	_
[Martinelli et al., PLB196]		7	0.02	_

Lattice can access either quarks or gluons

- only the recent ETM coll. results have both.

Variations in ensemble settings among lattice results.

Name	$Q [{ m GeV}]$	$\langle xV angle$	$\langle xS angle$	$\langle xg angle$
FantoPDF (DY+ γ +LN)	$\sqrt{1.9}$	0.49(8)	0.34(19)	0.18(12)
xFitter [9] (DY+ γ)	$\sqrt{1.9}$	0.55(6)	0.26(15)	0.19(16)
xFitter w/o scale variation	$\sqrt{1.9}$	0.55(2)	0.26(9)	0.19(9)
JAM'18 [8] (DY)	1.27	0.60(1)	0.30(5)	0.10(5)
JAM'18 [8] (DY+LN)	1.27	0.54(1)	0.16(2)	0.30(2)
JAM'21 [11] (DY+LN)	1.27	0.53(2)	0.14(4)	0.34(6)
CT18 NLO (proton)	$\sqrt{1.9}$	0.443(6)	0.160(10)	0.396(10)

[Kotz, et al., PRD109]

Moments from the lattice

	Name	$Q[{ m GeV}]$	$\langle x(u+ar{u})_{\pi^+} angle$	$\langle xg angle$
	FantoPDF	2	0.331(25)	0.24(10)
	HadStruct [19]	2	0.2541(33)	_
[Ga	ao et al., PRD102]	3.2	0.216(19)(8)	_
	ETM [46]	2	0.261(3)(6)	-
	ETM [91]	2	$0.601(28) _{u+d}$	0.52(11)
[Me	ver et al., PRD77]	2	_	0.37(8)(12)
[Shanahan et al., PRD99]		2	_	0.61(9)
[M	SU, 2310.12034]	2	_	0.364(38)(36)
	ZeRo Coll. [95]	2	0.245(15)	-
[Martine	elli et al., PLB196]	7	0.02	_

Lattice can access either quarks or gluons — only the recent ETM coll. results have both.

Variations in ensemble settings among lattice results.

-Gluon momentum fraction varies greatly!

Name	$Q \; [{ m GeV}]$	$\langle xV angle$	$\langle xS angle$	$\langle xg angle$
FantoPDF (DY+ γ +LN)	$\sqrt{1.9}$	0.49(8)	0.34(19)	0.18(12)
xFitter [9] (DY+ γ)	$\sqrt{1.9}$	0.55(6)	0.26(15)	0.19(16)
xFitter w/o scale variation	$\sqrt{1.9}$	0.55(2)	0.26(9)	0.19(9)
JAM'18 [8] (DY)	1.27	0.60(1)	0.30(5)	0.10(5)
JAM'18 [8] (DY+LN)	1.27	0.54(1)	0.16(2)	0.30(2)
JAM'21 [11] (DY+LN)	1.27	0.53(2)	0.14(4)	0.34(6)
CT18 NLO (proton)	$\sqrt{1.9}$	0.443(6)	0.160(10)	0.396(10)

[Kotz, et al., PRD109]

EIC: what is the gluon content of the pion?

	Science Question	Key Measurement	Key Requirements		
Collow Doport			 Need to uniquely determine 		
	What are the quark and gluon energy	Pion structure function data over a range	$e + p \rightarrow e' + X + n \text{ (low } -t)$		
	contributions to the pion mass?	of x and Q^2 .	 CM energy range ∼10-100 GeV 		
			 Charged and neutral currents desirable 		
	Is the pion full or empty of gluons as viewed at	Bion structure function data at large O^2	• CM energy ~100 GeV		
	large Q^2 ?	Pion structure function data at large Q^{-} .	 Inclusive and open-charm detection 		
		Kaon structure function data over a range	 Need to uniquely determine 		
	what are the quark and gluon energy contributions to the kaon mass?	Raon structure function data over a range of x and Ω^2	$e + p \rightarrow e' + X + \Lambda / \Sigma^0 \ (\text{low} - t)$		
	contributions to the kaon mass:	of x and Q^{-1} .	• CM energy range ~10-100 GeV		
	Are there more or less gluons in kaons than in V_{aon} attraction data at large Q^2		• CM energy ~100 GeV		
	pions as viewed at large Q ² ?	Sons as viewed at large Q^2 ? Kaon structure function data at large Q^2 .			
			 Need to uniquely determine exclusive process 		
	Can we get quantitative guidance on the	Pion form factor data	$e + p \rightarrow e' + \pi^+ + n \ (\text{low} - t)$		
	emergent pion mass mechanism?	for $Q^2 = 10-40 \; (\text{GeV}/c)^2$.	• $e + p$ and $e + D$ at similar energies		
			• CM energy ~10-75 GeV		



While COMPASS++ is confirming NA10's results,

we need the EIC: leading-neutron DIS



- pulls from future data in specific kinematics?
- correlations of PDFs with observables?
- disentaglement of sea and gluon PDF?
- model dependence?

UQ for pheno PDFs

Interpolation with constraints — scalar PDF

Set-up:

Decomposition of the scalar PDF

CLAS12 data on beam-spin asymmetry

Point-by-point extraction through DiFF at LO [AC et al., PRD106]

A. Courtoy-IFUNAM_

Interpolation with constraints — scalar PDF

Set-up:

Decomposition of the scalar PDF





CLAS12 data on beam-spin asymmetry

Point-by-point extraction through DiFF at LO [AC et al., PRD106]

Problem that would allow a unique solution, were it not for integral constraints.

The Bézier curve system can be solved exactly for specific dimensions in interpolation mode.

 \Rightarrow developing methodologies to impose moments.

Caveats: the method leads to a large number of models!

⇒ developing metrics to be used in penalties in loss function.



[E. López Rosa, Master thesis, UNAM] [E. López-Rosa. et al, in progress]

___UQ for pheno PDFs_

BNL-INT Workshop 2025

EIC: studies of multiparton dynamics

Yellow Report



Thanks to the sandbox studies, we can move on to our main goal:

sampling bias and model combination for the unpolarized PDF.

Unpolarized PDFs, EIC and precision physics

Uncertainties differ among various global analyses collaborations



Near future:

study of the building of the tolerance criterion for CT25, enhanced parametrization study,...

Not-so-near future:

complementarity of data, especially at large-x values and for the sea sector, will matter \Rightarrow EIC?

Α.	Courtoy-	IF	U	N	AI	١

UQ for pheno PDFs

Unpolarized PDFs, EIC and precision physics

- Most likely look for "new interactions"
- Small deviations from SM : PRECISION
- EFT description / BSM model



Credit: Daniel de Florian (DIS2023)

Large invariant mass in parton luminosity related to large-x values. Precision on LHC observables may rely on our knowledge of large-x PDFs.

Uncertainties on large-x PDFs hamper the searches for new physics.

Looking for New Physics in the tails of Drell-Yan like processes involves PDFs at large *x* and $Q : (x_1x_2 \ge 0.3, Q = 8 \text{ TeV}).$

[Fu et al., PhysRevD109, Ball et al., EPJC82]



Conclusions

Global QCD analyses are important to access hadron-structure information. Inverse problems, which sets of solutions are evolving.

Our take today: moving towards epistemic PDF uncertainties with polynomial approximators — Bézier curves.

Rôle of the parametrization in the <u>sampling accuracy</u>: we make use of **Bézier-curve methodology**

- ⇒ the EIC will contribute to learning PDFs in complementary kinematical ranges
- \Rightarrow Uncertainty quantification requires going beyond the $\Delta \chi^2 = 1$ prescription
- → Posterior means likelihood + priors use of information-like criteria to select solutions
- \Rightarrow The model combination also stems from a distribution de Finetti
- ⇒ Sandbox studies

<u>Pion PDFs: Fantomas1.0 available</u> metamorph can be used to study many functions

Reliable uncertainty on the PDF analysis (to NLO) re: larger where no data constrains $q^{\pi}(x, Q^2)$

Collinear twist-3 PDF

Playground for my group

Integration of constraints from moments Development of metrics for shape selection

Fantômas standalone code

Fantômas Unconfined: global QCD fits with Bézier parameterizations

Lucas Kotz^{a,1}, Aurore Courtoy^{b,2,*}, Tim Hobbs^{c,3}, Pavel Nadolsky^{d,4,**}, Fredrick Olness^{a,5}, Maximiliano Ponce-Chavez^d, Varada Purohit^a

xFitter embedded in "PDF Parametrization"



Figure 1: Schematic structure of the **xFitter** program.

Companion standalone to be plugged wherever the user decides to

- C++ module
- MetamorphCollection class object
- reads steering cards
- check sum rules
- shuffles

Fantomas1.0: <u>https://cteq-tea.gitlab.io/project/00pdfs/#mesonPDFs</u>

Latin American contribution to the EIC

LASF4RI



Latin American Strategy Forum for Research Infrastructure

Developing a strategy to strengthen Latin American Scientific Collaborations and their impact.

Brazilian Journal of Physics (2025) 55:145 https://doi.org/10.1007/s13538-025-01778-x



Latin American Strategy Forum for Research Infrastructure (III LASF4RI Contribution)

The Glue That Binds Us All—Latin America and the Electron-Ion Collider

A. C. Aguilar¹ · A. Bashir² · J. J. Cobos-Martínez³ · A. Courtoy⁴ · B. El-Bennich⁵ · D. de Florian⁶ · T. Frederico⁷ · V. P. Gonçalves⁸ · M. Hentschinski⁹ · R. J. Hernández-Pinto¹⁰ · G. Krein¹¹ · M. V. T. Machado¹² · J. P. B. C. de Melo¹³ · W. de Paula⁷ · R. Sassot¹⁴ · F. E. Serna¹⁵ · Supporting authors outside Latin America: L. Albino¹⁶ · I. Borsa¹⁷ · L. Cieri¹⁸ · I. M. Higuera-Angulo¹⁹ · J. Mazzitelli²⁰ · Á. Miramontes¹⁸ · K. Raya²¹ · F. Salazar²² · G. Sborlini²³ · P. Zurita²⁴

A. Courtoy aurore@fisica.unam.mx

⊠ B. El-Bennich bennich@unifesp.br

Received: 1 October 2024 / Accepted: 20 April 2025 © The Author(s) under exclusive licence to Sociedade Brasileira de Física 2025

UQ for pheno PDFs_

Presenta:

'LA INTELIGENCIA EN LLAMAS'

NAZA BLEGANTE

Un episodio excepcional... arde la cultura del mundo. ¡Vea a **Fantoman** apuros, entrevistándos con los más grandes escritores contemporáneos!

RES REVIS

M.N

Con TABLA

Bindo 💽