Fantômas4QCD:
pion PDFs with epistemic uncertainties

Aurore Courtoy

Instituto de Física
National Autonomous University of Mexico (UNAM)

QNP2024

Barcelona

09/07/24

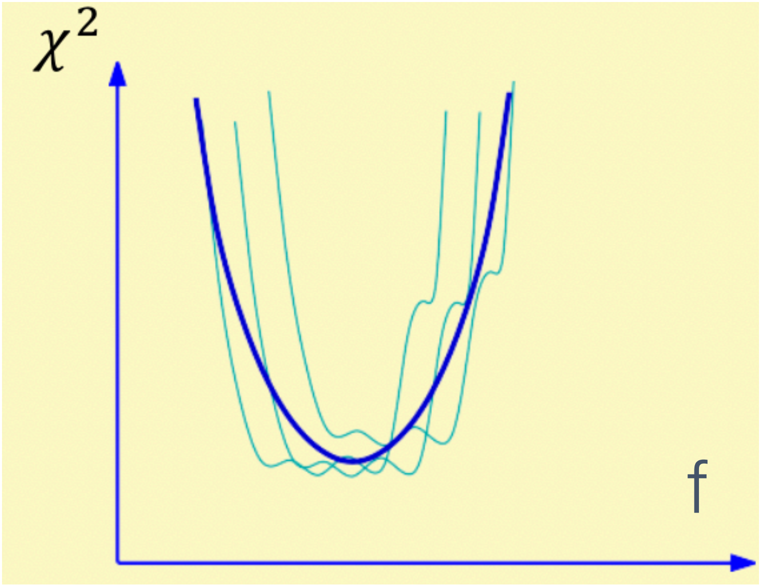
Fantômas4QCD

Main motivation of Fantômas4QCD:
to quantify the rôle of parametrization form in global analyses.

A new `c++` code automates series of fits using multiple functional forms, called `metamorph`.

It's based on Bézier curves —polynomial functional forms can approximate any arbitrary PDF shape.

Fantômas unlocks the concept of tolerance for cuts on the χ^2 shapes.



Kotz, AC, Nadolsky, Olness, Ponce-Chávez
[Phys.Rev.D109] and code release very soon!

Uncertainty quantification for PDFs

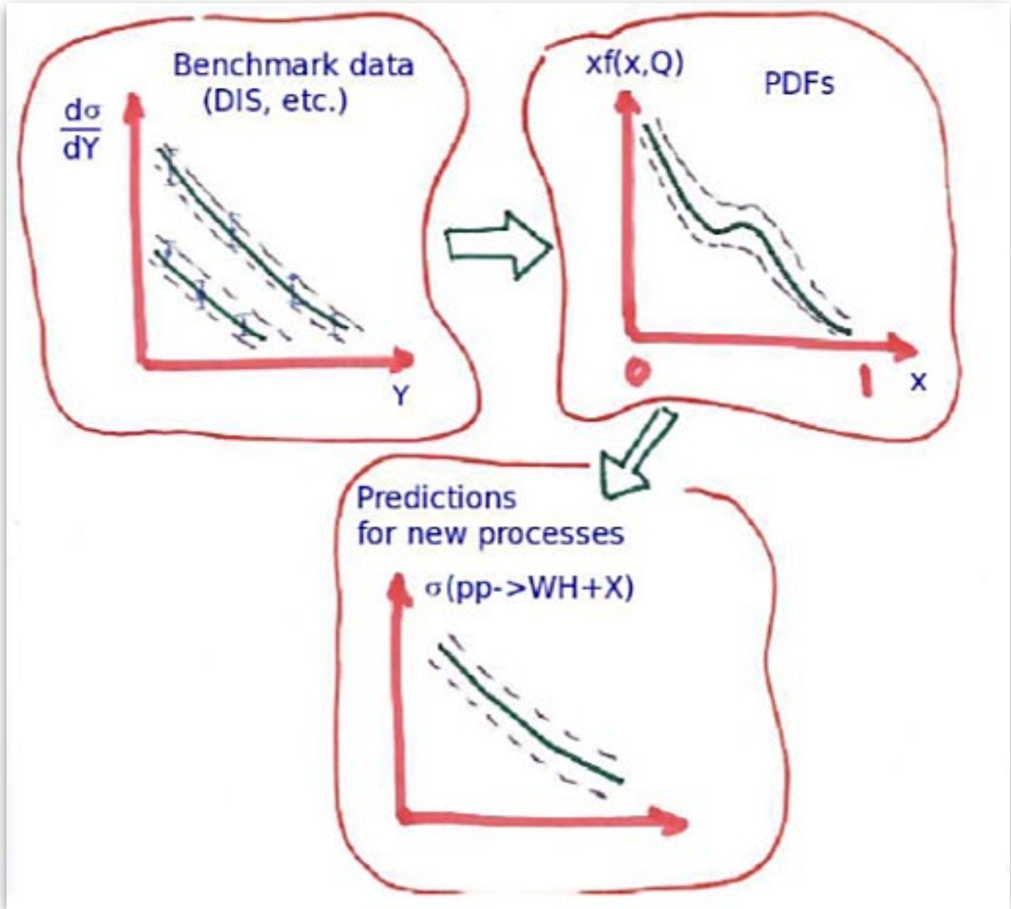
Low-energy QCD dynamics, encapsulated in PDFs, are learned from experimental data.

Uncertainty propagates from data and methodology to the PDF determination

I. assessment of uncertainty magnitude is key

II. advanced statistical problem

III. evolving topic in the era of AI/ML



Uncertainty quantification for PDFs

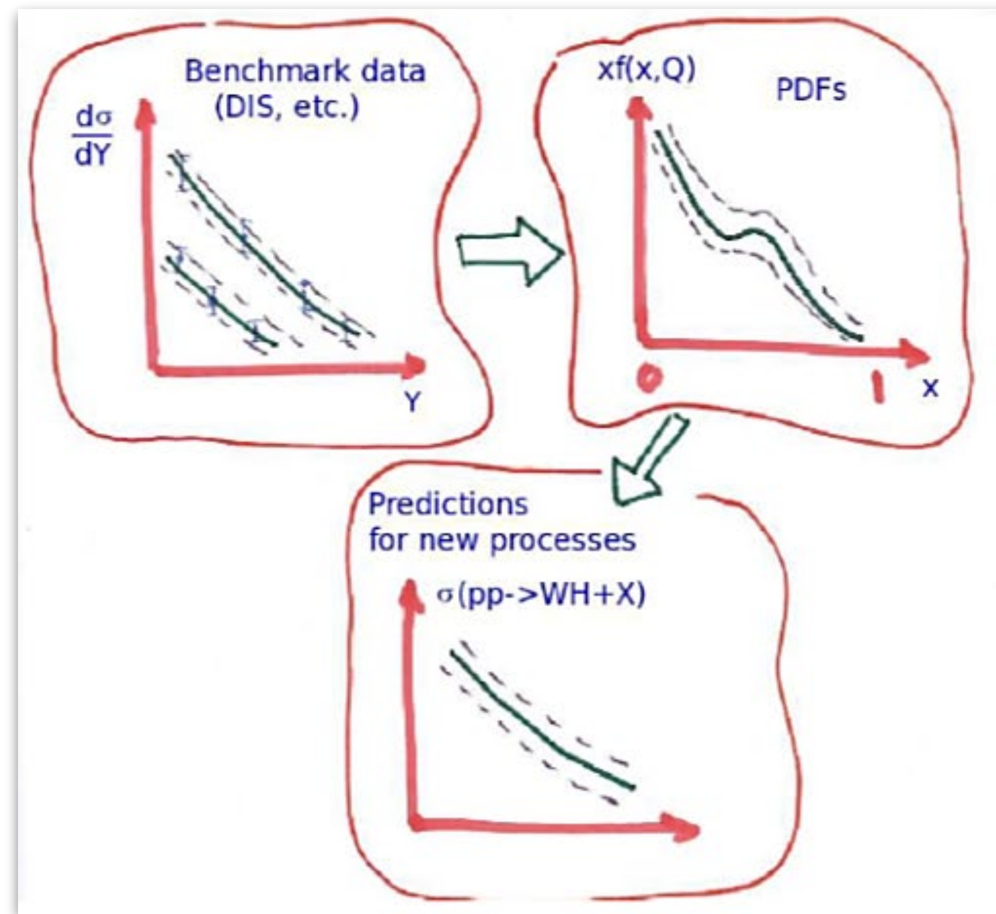
Low-energy QCD dynamics, encapsulated in PDFs, are learned from experimental data.

Uncertainty propagates from data and methodology to the PDF determination

I. assessment of uncertainty magnitude is key

II. advanced statistical problem

III. evolving topic in the era of AI/ML



Epistemic vs. aleatory uncertainties

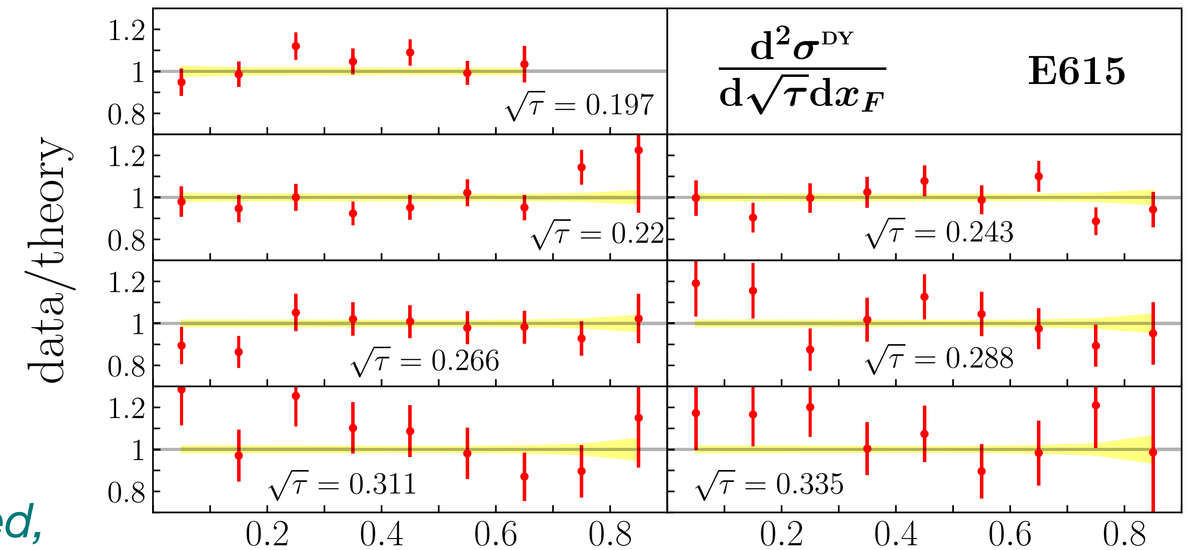
Uncertainty due to lack of knowledge —bias (may be reduced)

Statistical uncertainty propagated from experiments— irreducible

Data-based analyses

Global analyses involve searching for extrema of a (log-)likelihood function.

[JAM21]



<u>(Very) simplified:</u>

$$\chi^2 = \sum_i^{N_{\rm exp}} \frac{\left(D_i - \langle T(\{\mathbf{x},\mathbf{a}\})\rangle_i\right)^2}{\sigma_i^2} + \text{penalty terms}$$

discrete data point

theory prediction averaged,

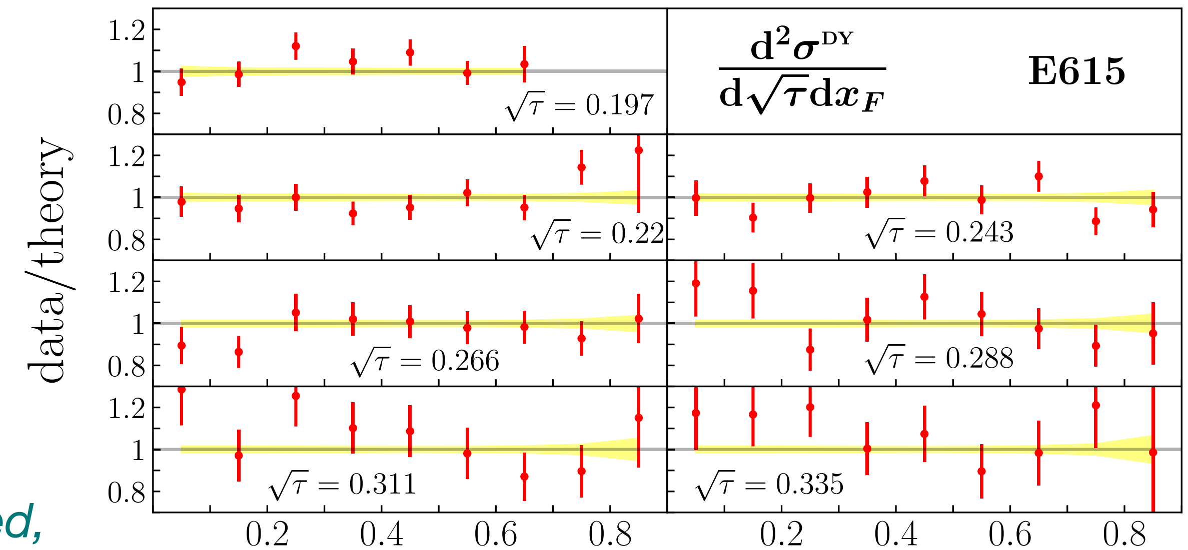
*as a function of the variables {**x**} and free parameters {**a**}*

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.

Data-based analyses

Global analyses involve searching for extrema of a (log-)likelihood function.

[JAM21]



<u>(Very) simplified:</u>

$$\chi^2 = \sum_i^{N_{\rm exp}} \frac{\left(D_i - \langle T(\{\mathbf{x},\mathbf{a}\})\rangle_i\right)^2}{\sigma_i^2} + \text{penalty terms}$$

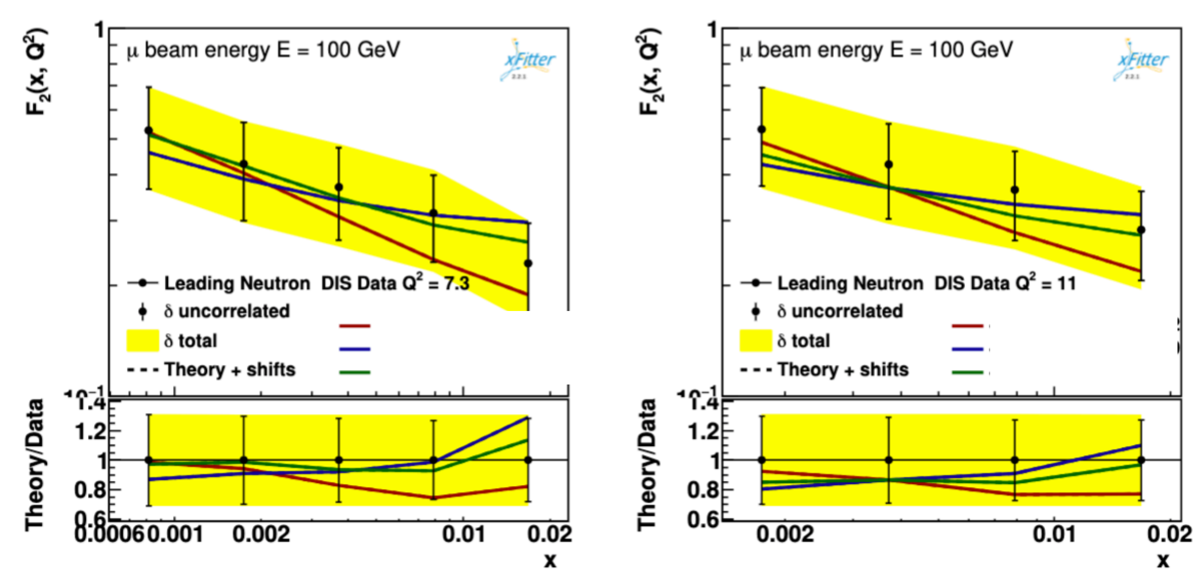
discrete data point

theory prediction averaged,

*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.

[Fantômas]



That's fine in the data region,
but the results may vary greatly outside
— extrapolation region.

Why not adopt more than one form?

Epistemic PDF uncertainty — a CT "signature" concept

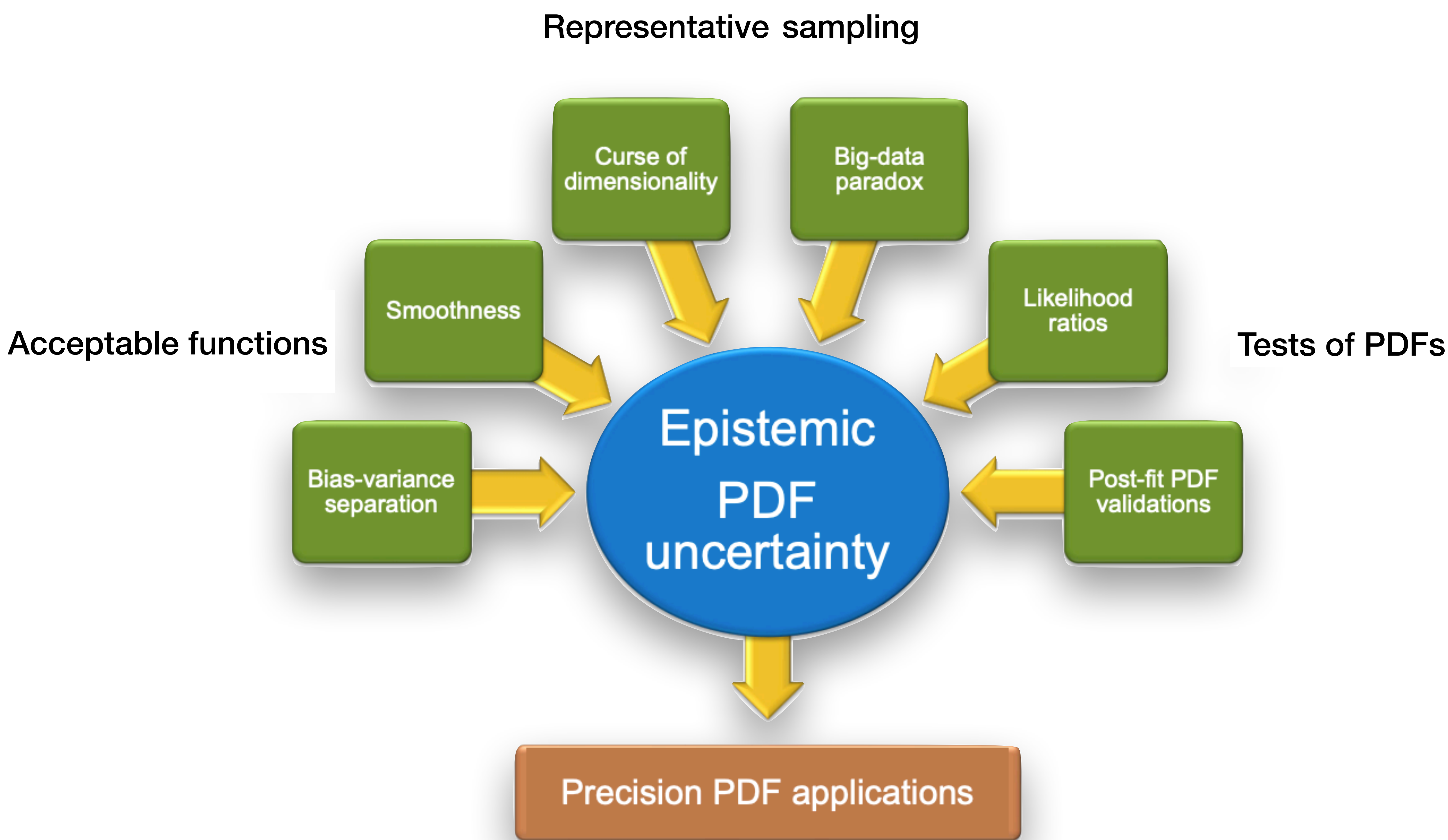
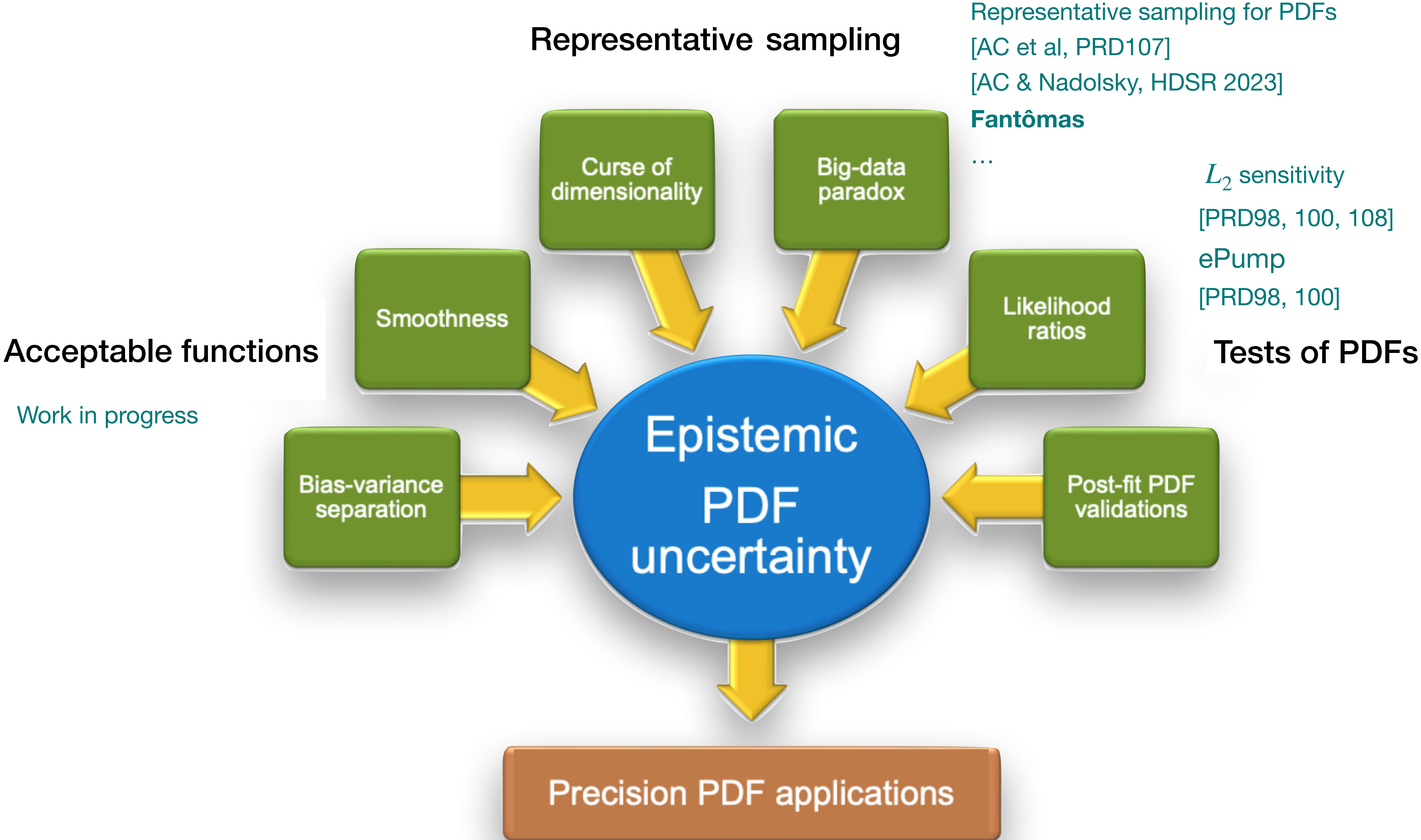


diagram by P. Nadolsky [DIS2023]

Epistemic PDF uncertainty — a CT "signature" concept

Representative sampling for PDFs
[AC et al, PRD107]
[AC & Nadolsky, HDSR 2023]
Fantômas
…

L_2 sensitivity
[PRD98, 100, 108]
ePump
[PRD98, 100]



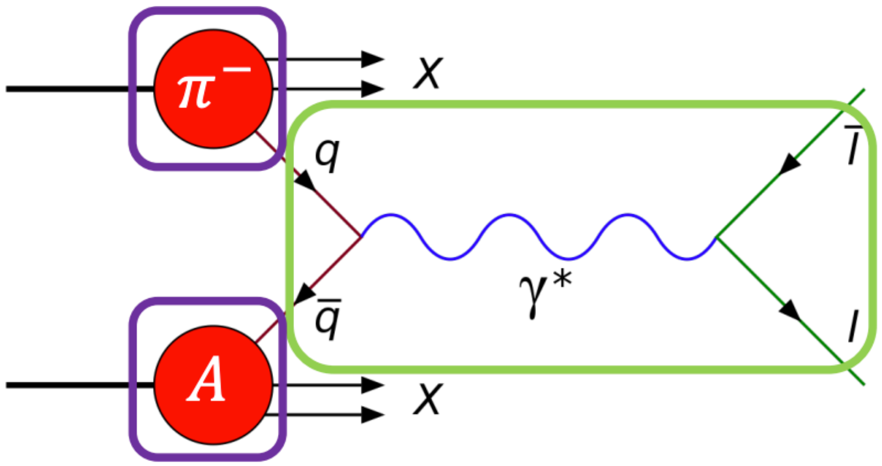
Work in progress

diagram by P. Nadolsky [DIS2023]

Pion PDF as sandbox

Framework to access pion PDFs available on xFitter.

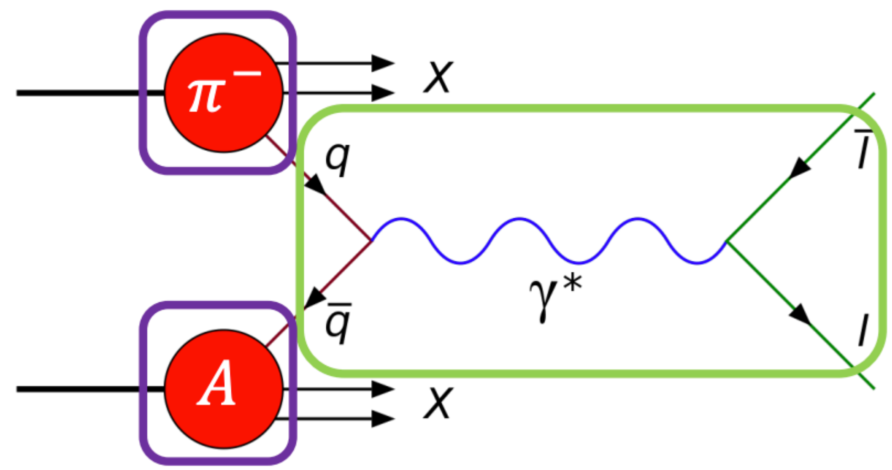
The first two global analyses of the pion data were performed by the SMRS and GRV groups in the early 1990s on Drell-Yan data.



Pion PDF as sandbox

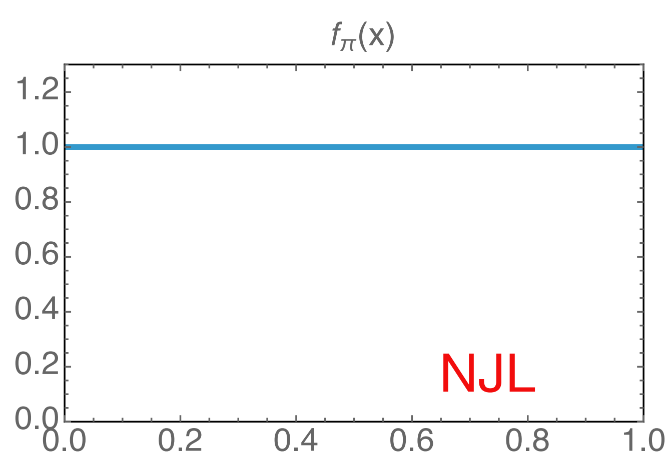
Framework to access pion PDFs available on xFitter.

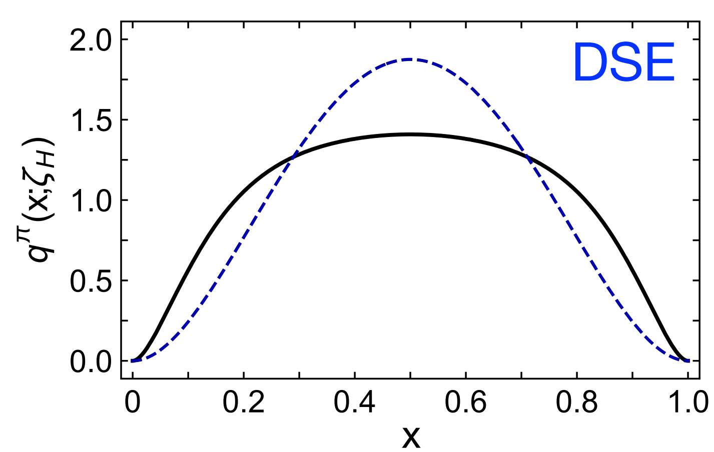
The first two global analyses of the pion data were performed by the SMRS and GRV groups in the early 1990s on Drell-Yan data.



QCD dynamics in the non-perturbative regime:

Physics interests are related to the *χ sym. breaking* predicts that the pion PDF is broader than what expected from a bound-state smearing.

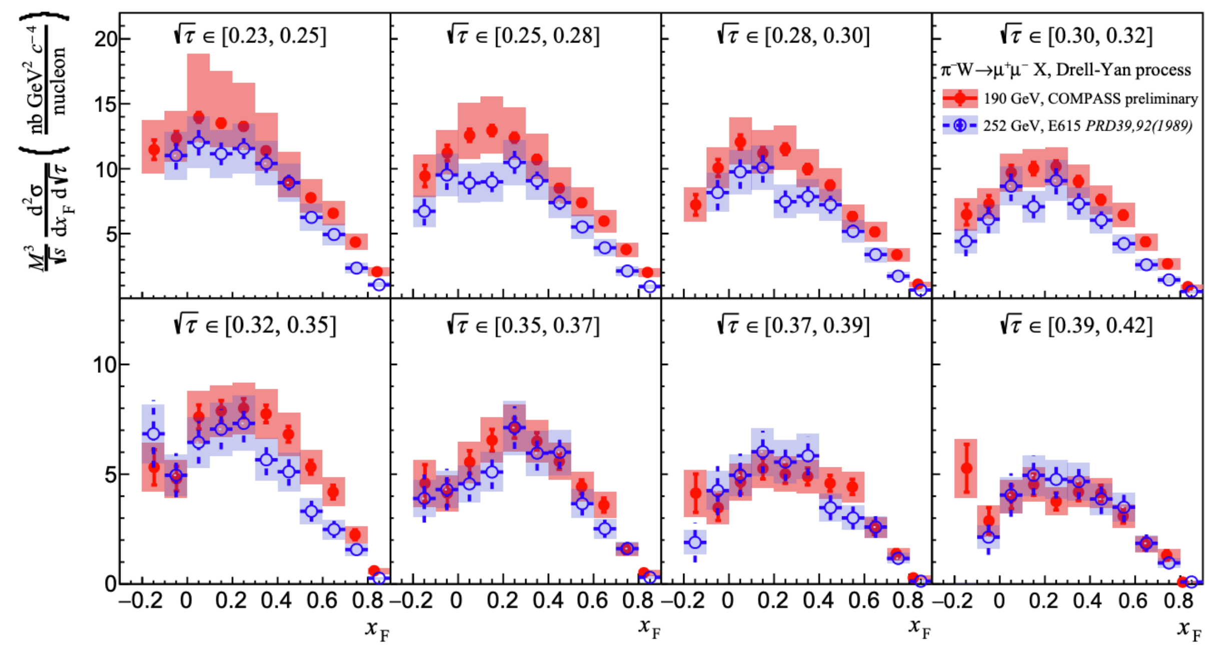




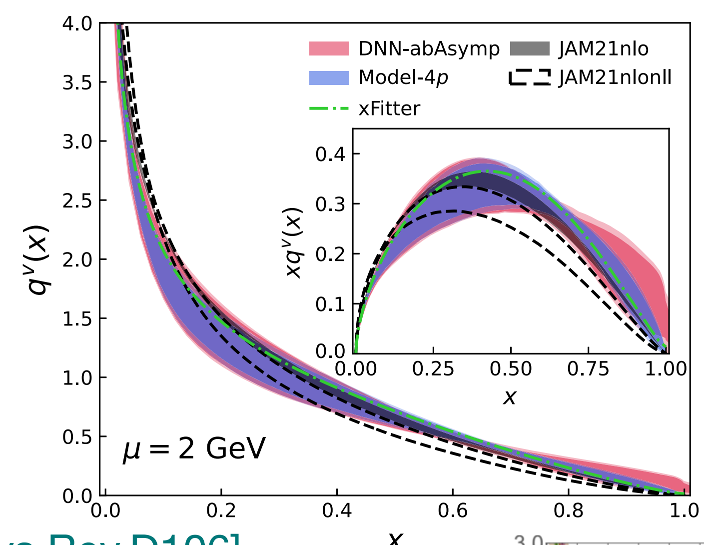
Model estimates for the pion valence PDF at a low hadronic scale

2020’s and the revival of pion structure studies

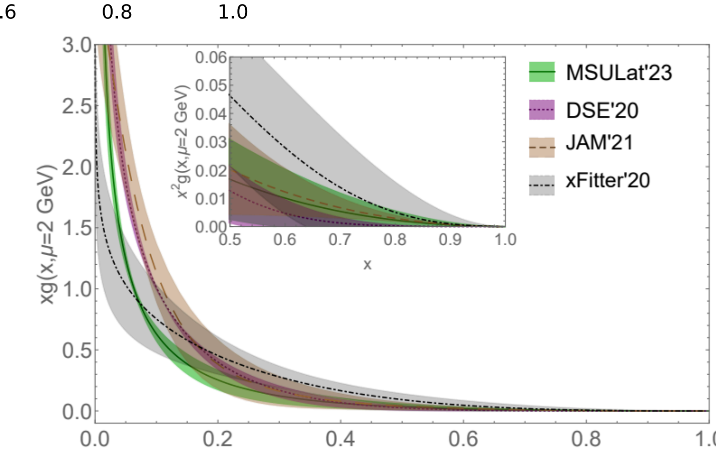
- **experimentally**:
  ⇒ pion beams or pion production in DIS experiments.
  ⇒ COMPASS++, future EIC



- **lattice QCD**:
  ⇒ reconstruction from moments of PDFs
  ⇒ quasi/pseudo-PDFs & related can be evaluated

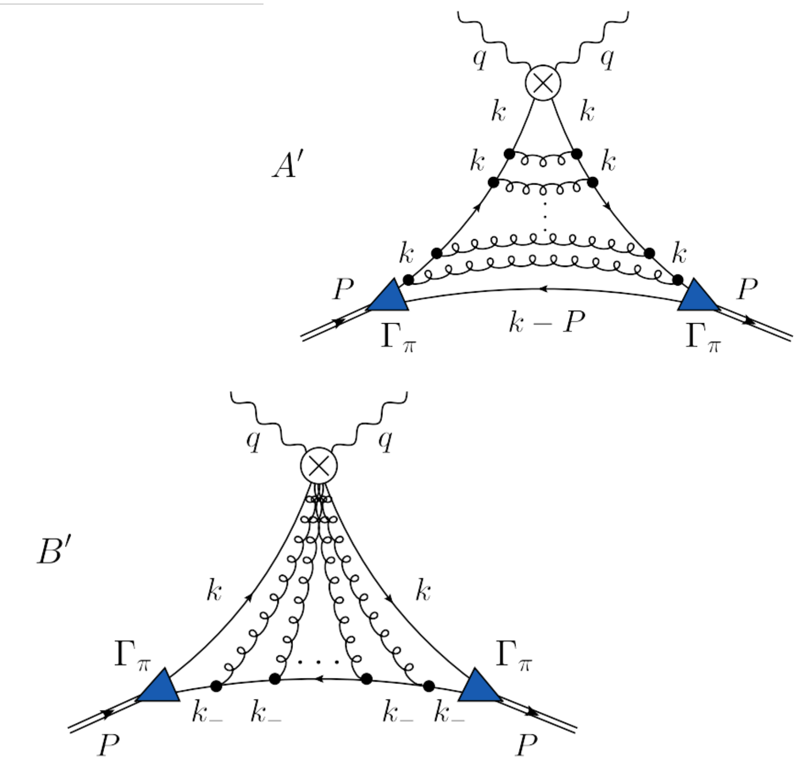


[X. Gao et al, Phys.Rev.D106]

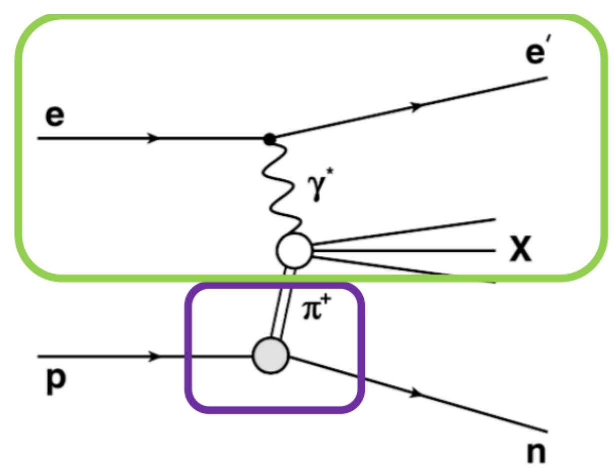


[MSULat, 2310.12034]

2020’s and the revival of pion structure studies



- **theoretically**:
  - ⇒ improvements in continuum approaches (DSE).
  - ⇒ more complexed objects have been studies, including ML/AI tools
  - ⇒ mixed models

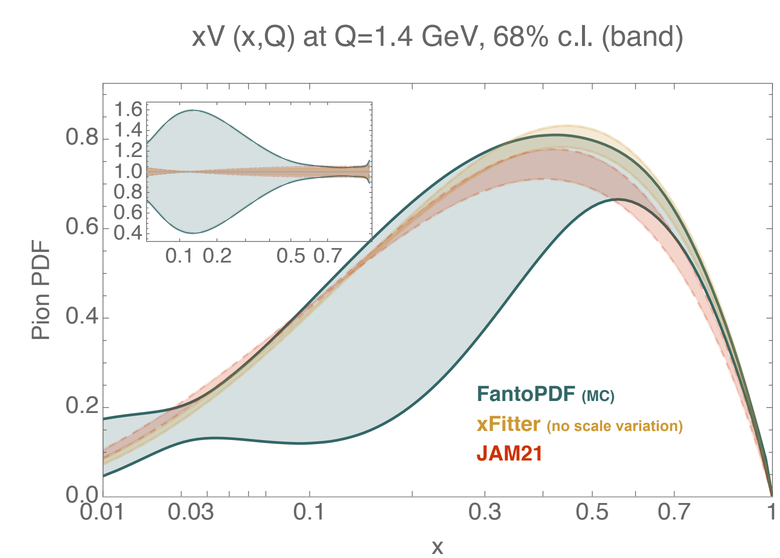


- **phenomenologically**
  - ⇒ 2018, first fit including DIS on the pion (leading-neutron detection)

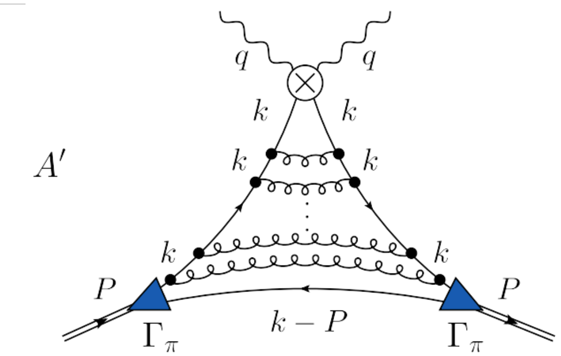
[JAM, Phys.Rev.L121]

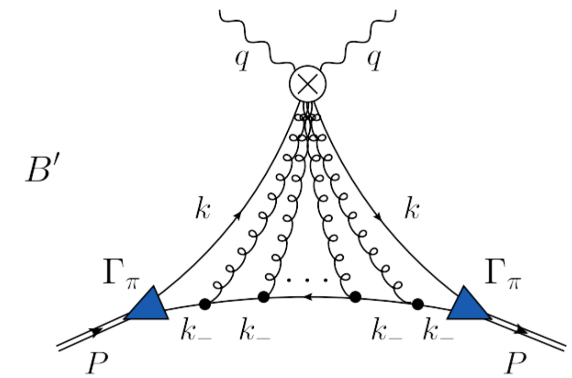
  - ⇒ 2023, first fit accounting for the *parametrization uncertainty*

[Fantômas, 2311.08447]

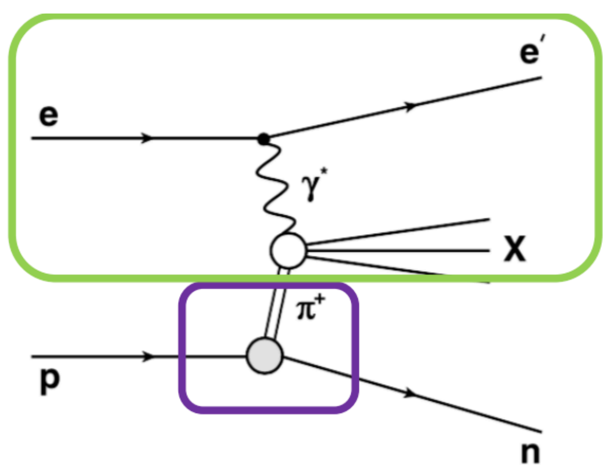


2020’s and the revival of pion structure studies





- **theoretically**:
  - ⇒ improvements in continuum approaches (DSE).
  - ⇒ more complexed objects have been studies, including ML/AI tools
  - ⇒ mixed models



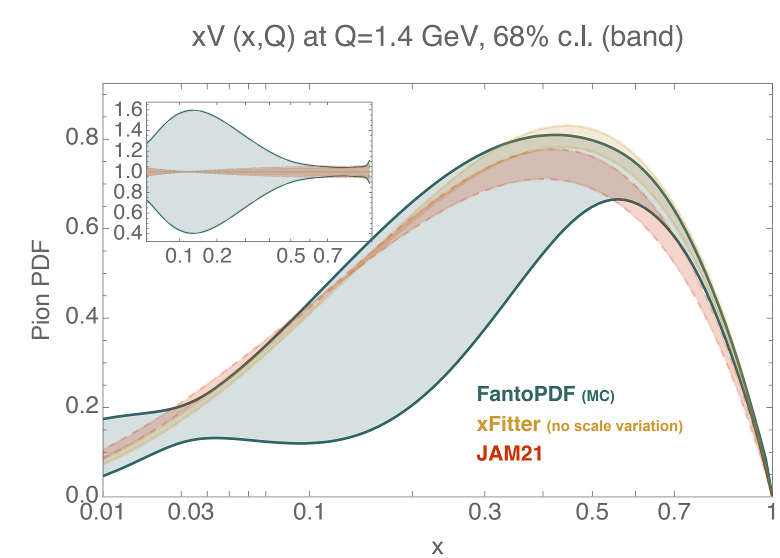
- **phenomenologically**
  - ⇒ 2018, first fit including DIS on the pion (leading-neutron detection)

[JAM, Phys.Rev.L121]

Fantômas4QCD

  - ⇒ 2023, first fit accounting for the *parametrization uncertainty*

[Fantômas, 2311.08447]



Bézier curve

Bézier curves are convenient for interpolating discrete data

The interpolation through Bézier curves is unique if the polynomial degree= (# points-1), there's a closed-form solution to the problem,

$$\mathcal{B}^{(n)}(x) = \sum_{l=0}^{n} c_l \; B_{n,l}(x)$$

with the Bernstein pol.

$$B_{n,l}(x) \equiv \binom{l}{n} x^l (1-x)^{n-l}.$$

The Bézier curve can be expressed as a product of matrices:

- $\underline{T}$ is the vector of x^l
- $\underline{\underline{M}}$ is the matrix of binomial coefficients
- $\underline{C}$ is the vector of Bézier coefficient, c_l, to be determined

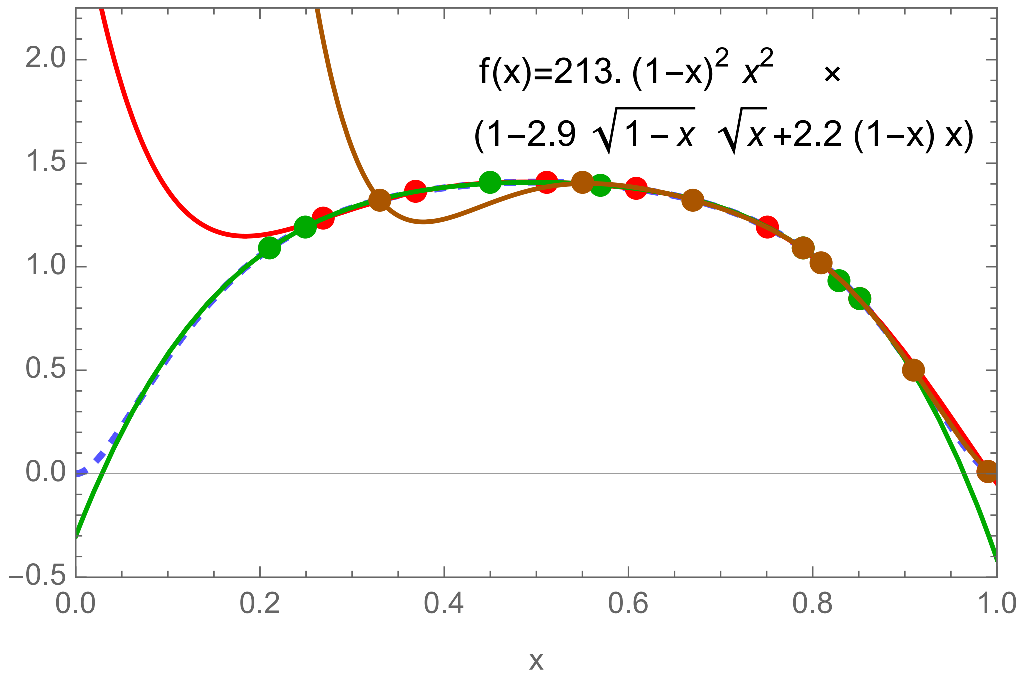
$$\mathcal{B} = \underline{T} \cdot \underline{\underline{M}} \cdot \underline{C}$$

We can evaluate the Bézier curve at chosen **control points**, to get a vector of $\mathscr{B} \rightarrow \underline{P}$

$\underline{\underline{T}}$ is now a matrix of x^l expressed at the control points.

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

Bézier-curve methodology for global analyses



The reconstructed function may depend on the position and number of control points.

Global analyses can exploit this property to generate many functional forms.

⇒ polynomial mimicry

Behaviour on top of asymptotics is embedded into a Bézier curve

⇨ asymptotics usually ensured by a *carrier function*

⇨ sum rules imposed through normalization

$$x\, q(x, Q_0^2) = \underline{A'_q}\, \underline{x^{B_q} (1-x)^{C_q}} \times \Big(1 + \mathcal{B}^{(N_m)}(x^{\alpha_x}, Q_0^2; \underline{v})\Big)$$

for q = PDF type
(flavor, combination or gluon)

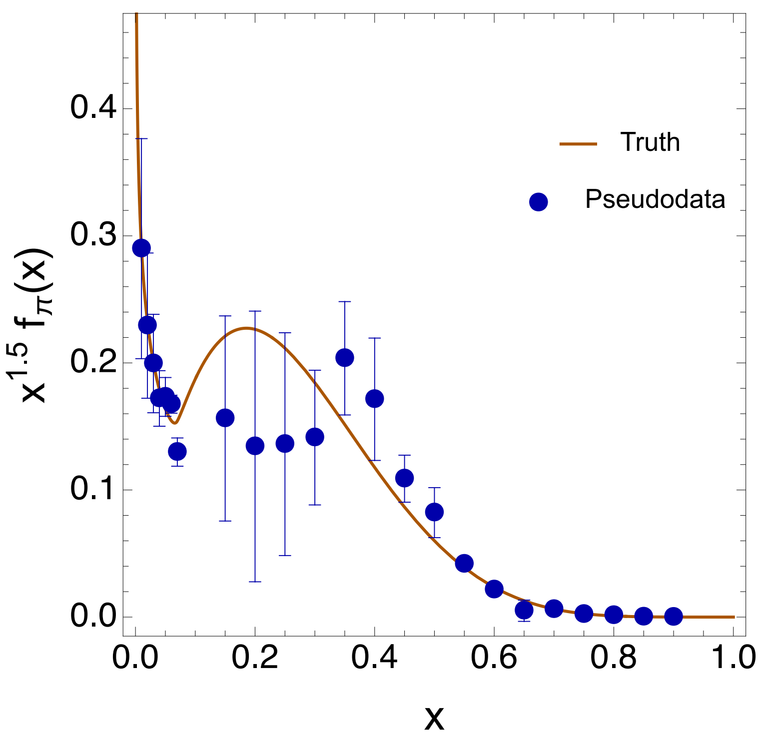
Fantômas4QCD program

⇨ $\mathscr{B}$ can modulate the PDFs in flexible ways at intermediate x using a set of free and fixed control points

Bézier-curve methodology for global analyses — toy model

Fantômas4QCD program

⇨ $\mathscr{B}$ can modulate the PDFs in flexible ways at intermediate x using a set of free and fixed control points through $\underline{P} = \underline{\underline{T}} \cdot \underline{\underline{M}} \cdot \underline{C}$



$$x\, q(x, Q_0^2) = A_q'\, x^{B_q} (1-x)^{C_q} \times \left(1 + \mathcal{B}^{(N_m)}(x^{\alpha_x}, Q_0^2; \underline{v})\right)$$

with $\underline{v} = \{\underline{C}, \underline{P}\}$

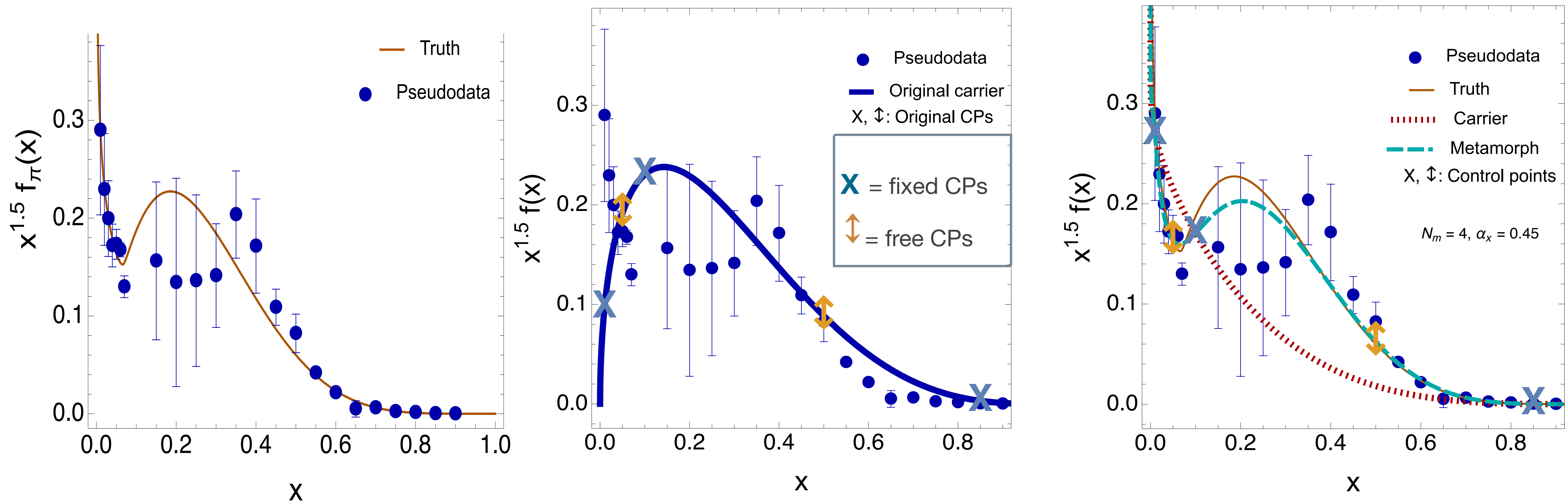
Classical fit: determines the vector $\underline{C}$

`metamorph` fit: determines the variations in vector $\underline{P}$

We parametrize the Bézier coefficients as the shifts of the position of the **control points:**

$$P_i = \mathcal{B}(x_i) \rightarrow P_i' = \mathcal{B}(x_i) + \delta\mathcal{B}(x_i)$$

Bézier-curve methodology for global analyses — toy model

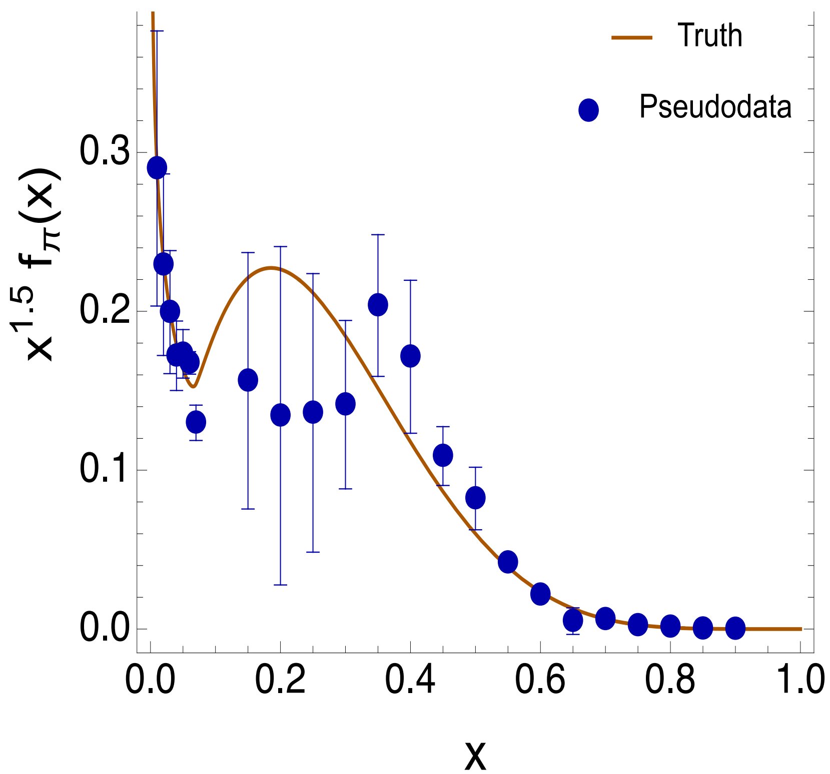


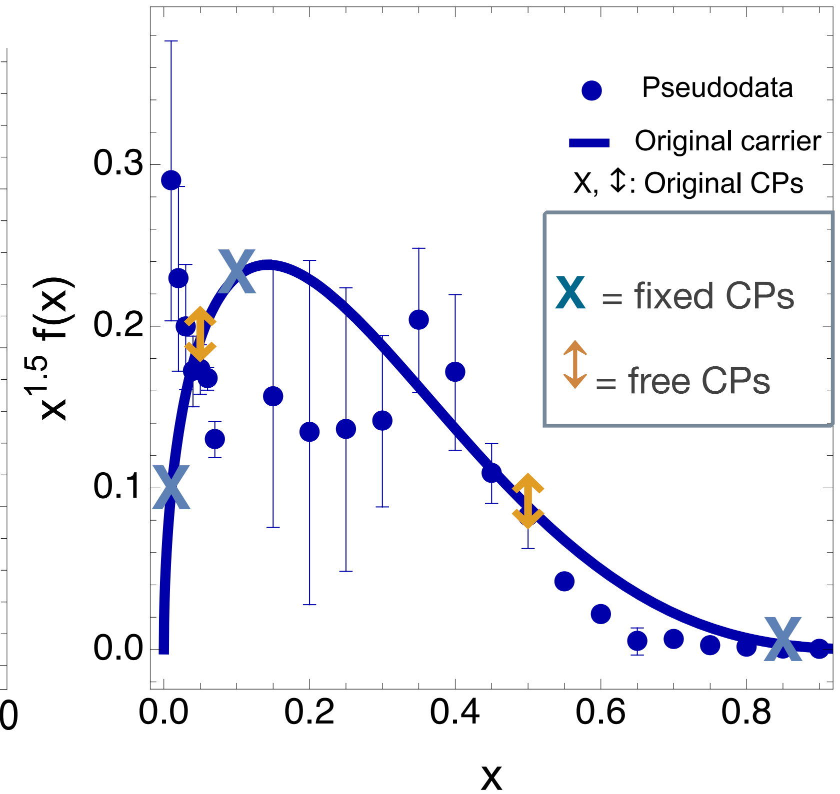
`metamorph` fit:

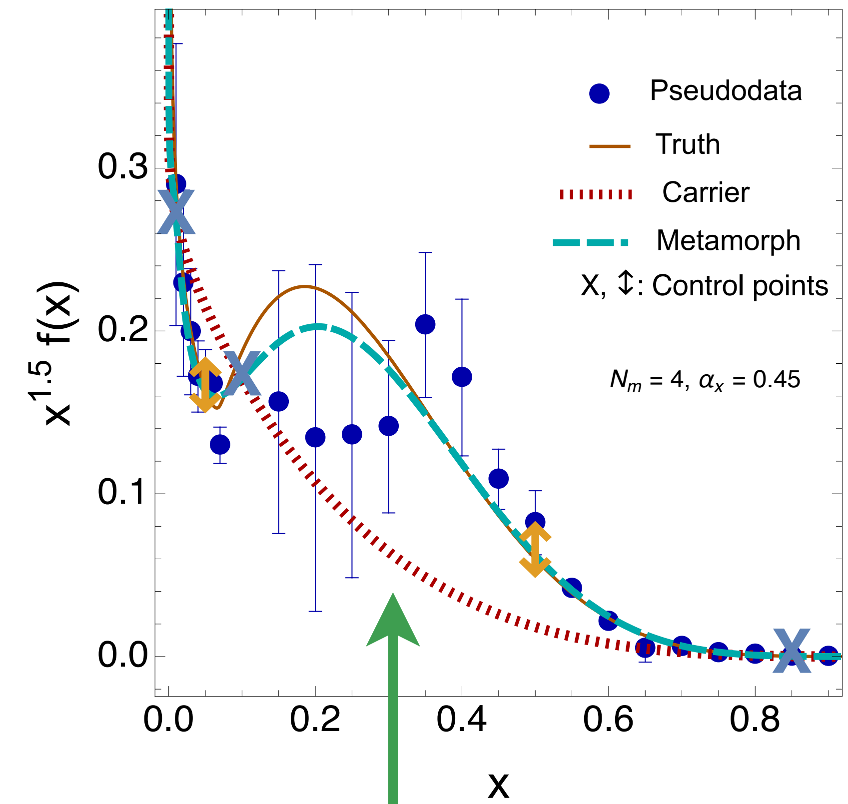
$$x\,q(x,Q_0^2) = A'_q\, x^{B_q}(1-x)^{C_q} \times \left(1+\mathcal{B}^{(N_m)}(x^{\alpha_x}, Q_0^2; \underline{v})\right)$$

with N_m= # CPs-1 for a square-matrices system.

Bézier-curve methodology for global analyses — toy model







`metamorph` fit:

$$x\, q(x, Q_0^2) = A'_q\, x^{B_q} (1-x)^{C_q} \times \left(1 + \mathcal{B}^{(N_m)}(x^{\alpha_x}, Q_0^2; \underline{v})\right)$$

with N_m= # CPs-1 for a square-matrices system.

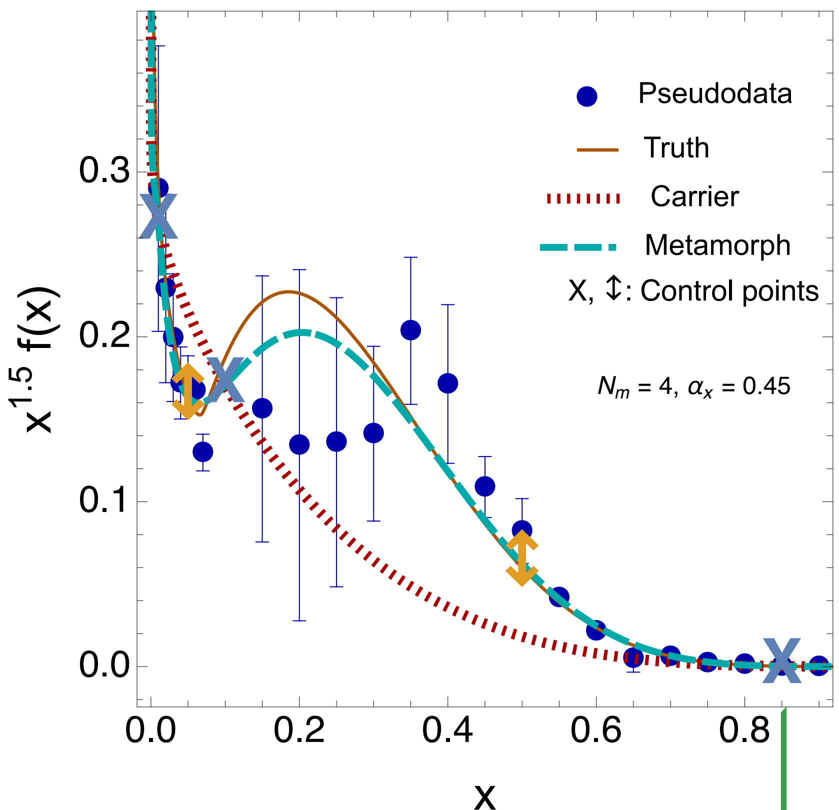
Shift of the control points $(\delta D_q, \ldots)$ replace free parameters

N_m= degree of polynomial can vary

δB_q & δC_q allow the carrier to vary

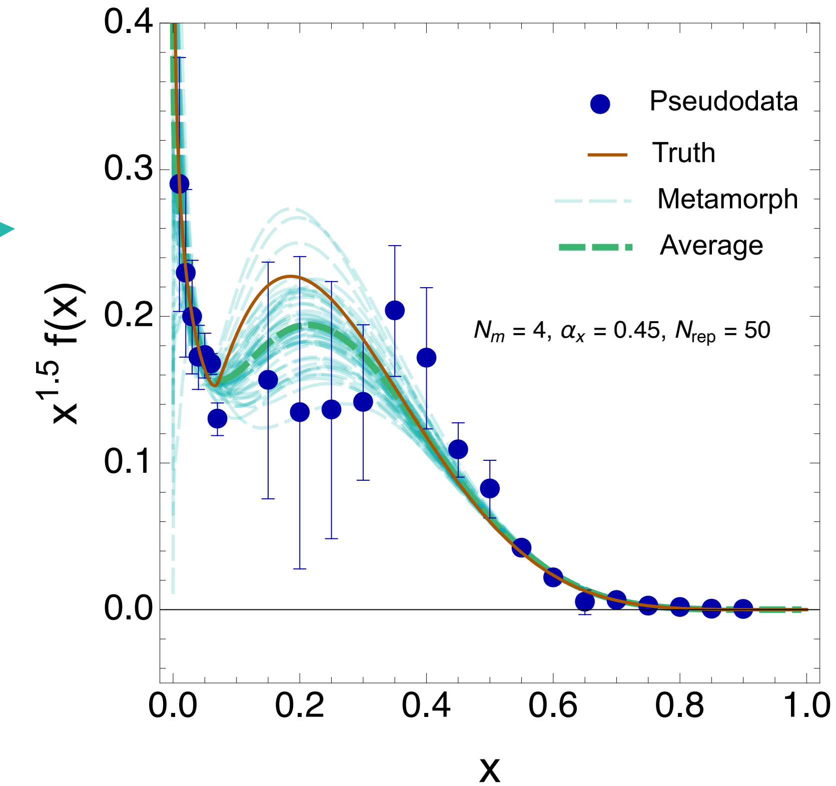
α_x can vary

Bézier-curve methodology for global analyses — toy model



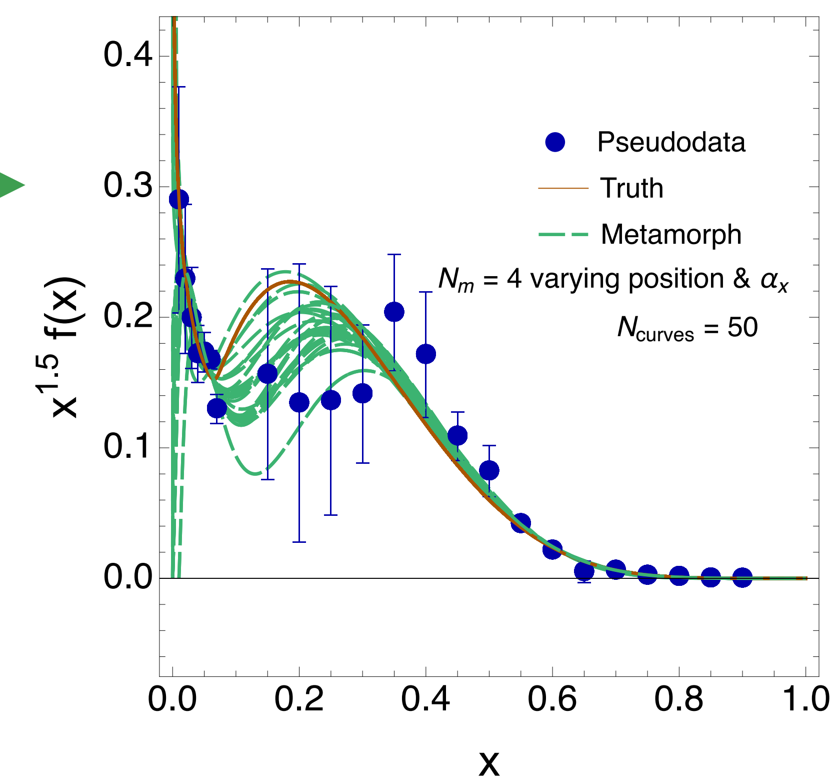
if bootstrapped

sampling on the distribution of data uncertainties



if sampled over `metamorph` settings

sampling over parametrizations



Both samplings can be done in the same analysis, they are not mutually exclusive.

metamorph routine in xFitter

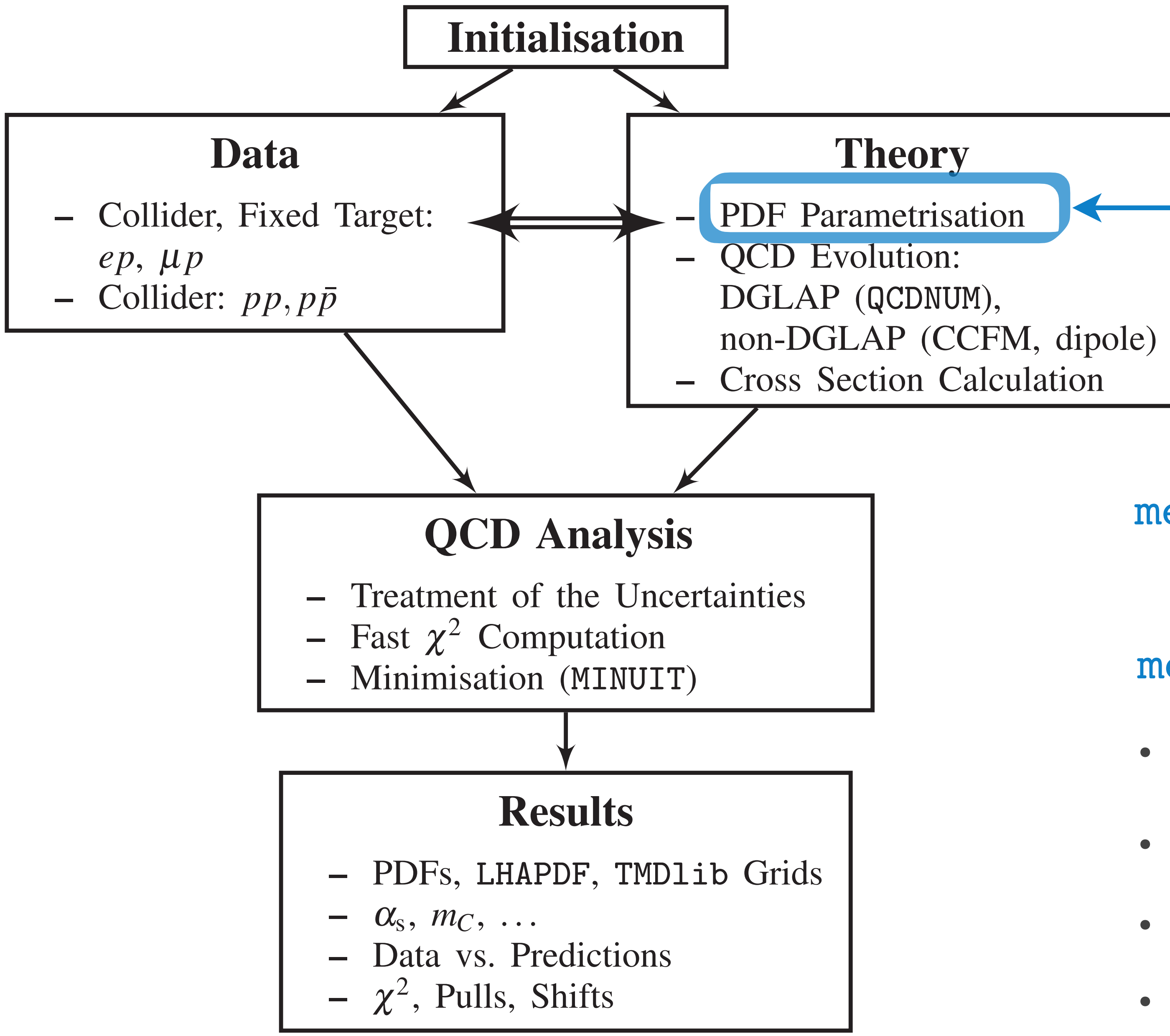


Figure 1: Schematic structure of the `xFitter` program.

`metamorph` routine — PhD thesis of L. Kotz (SMU)

`metamorph` requires inputs from the user:

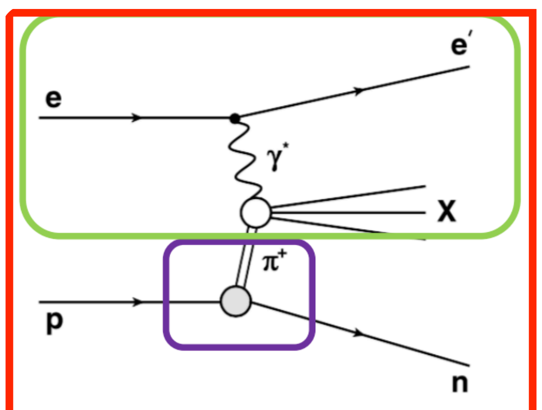
- N_m — degree of polynomial
- $\{x, f_{in}(x)\}$ of control points
- fixed or free control points
- stretching parameter

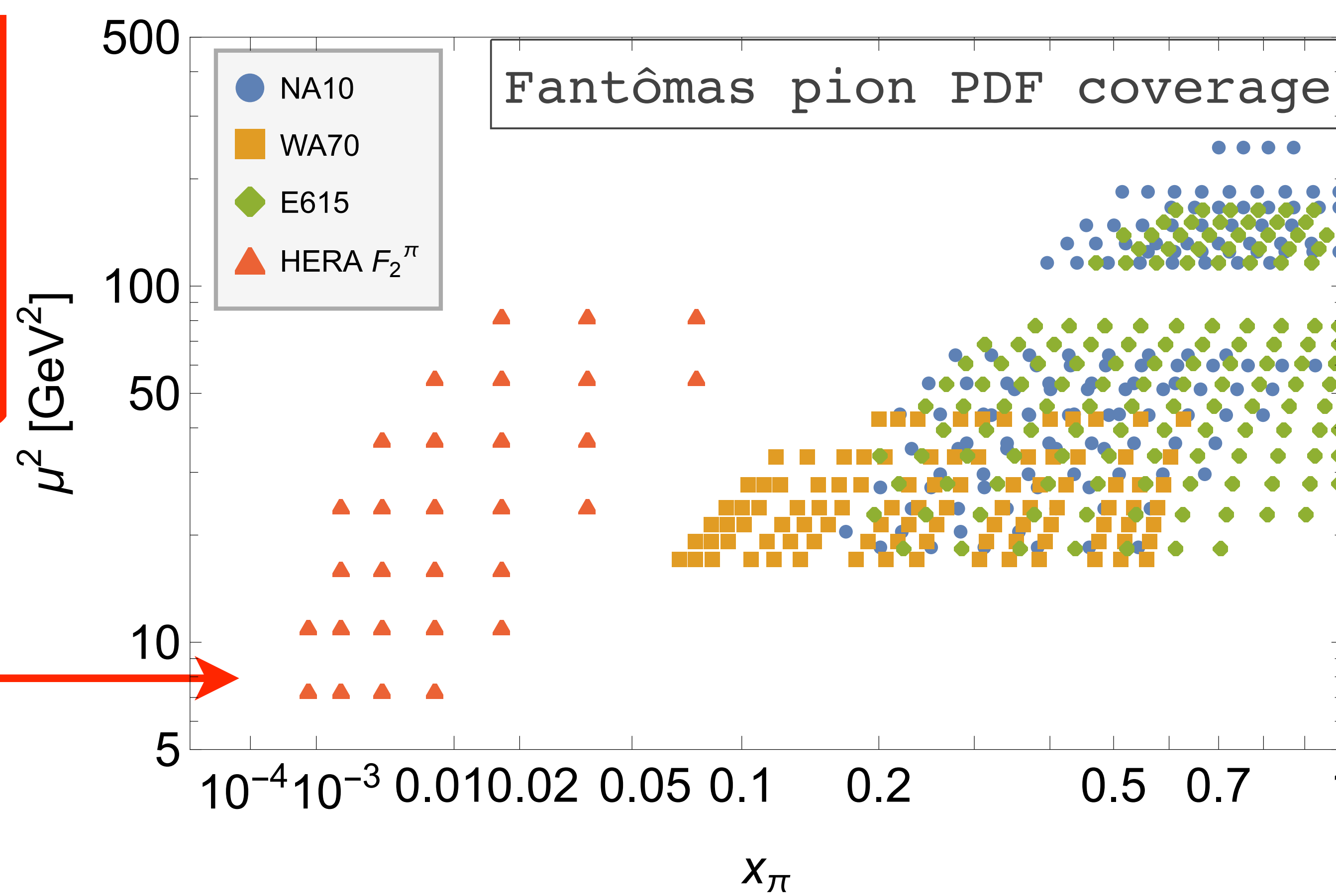
First application of Fantômas: pion PDF

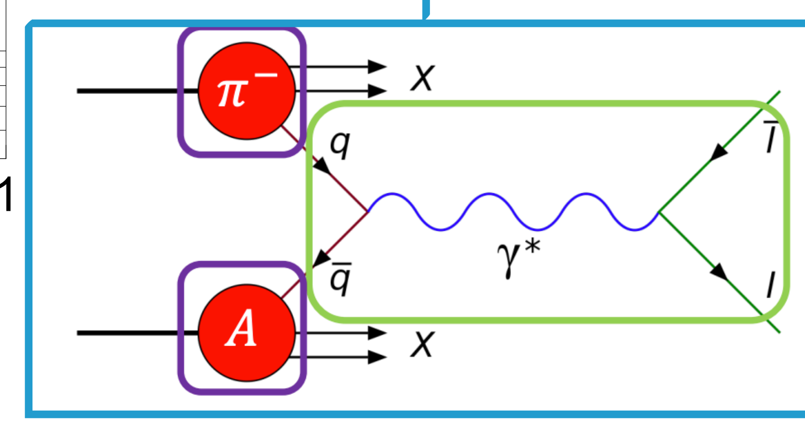
Previous (modern) pion analyses:

xFitter [PRD102]
JAM [PRL121, PRD103, PRL127]

We use the xFitter framework for pion PDFs.
We also extend the xFitter data by adding leading neutron (Sullivan process) data
— minimal small-x coverage [model-dependence in describing the pion as a target].







Diagrams from P. Barry

The Fantômas pion PDFs

First physics use of the Fantômas framework:

⇨ We generated $N \sim 100$ fits corresponding to N sets for $\{N_m, \underline{P}, \alpha_x\}$.

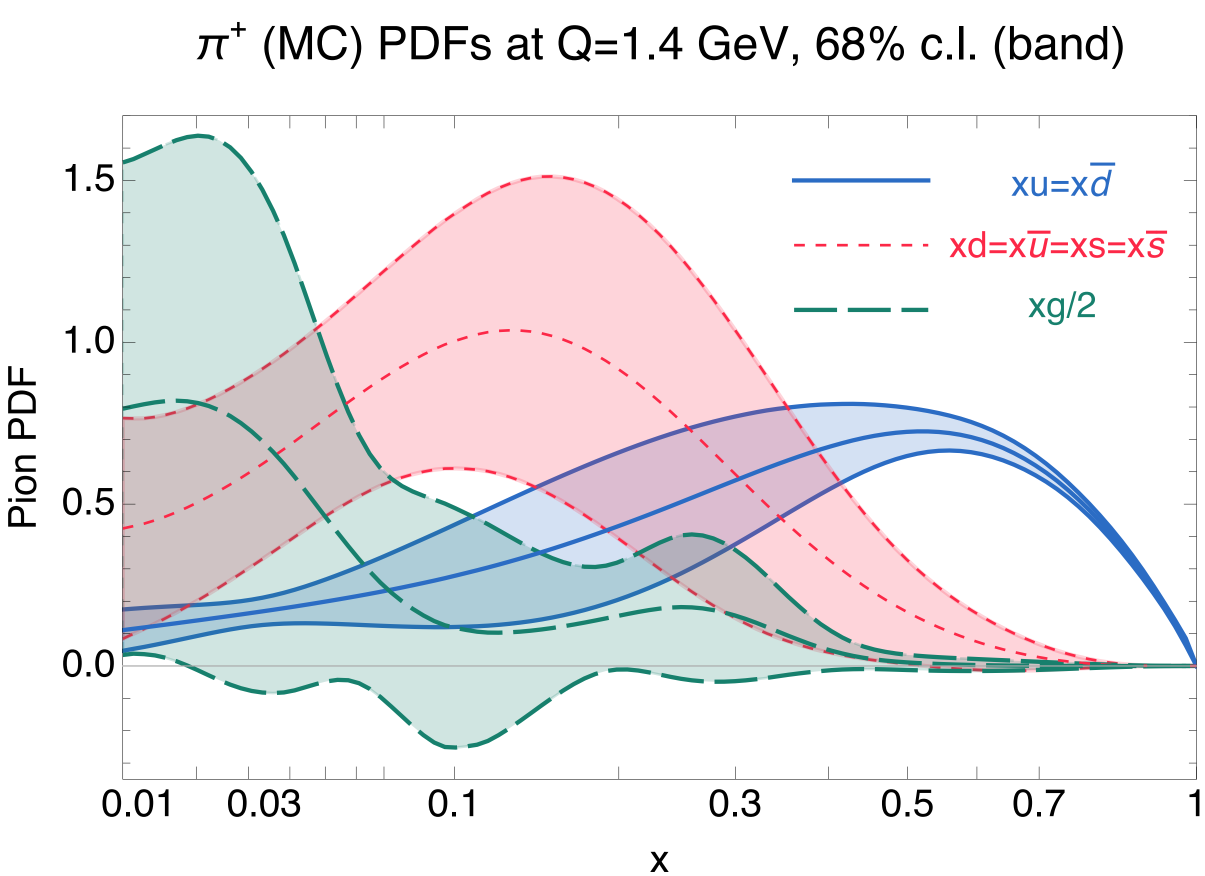
⇨ Well-behaved (convergence + soft constraints) fits are kept.

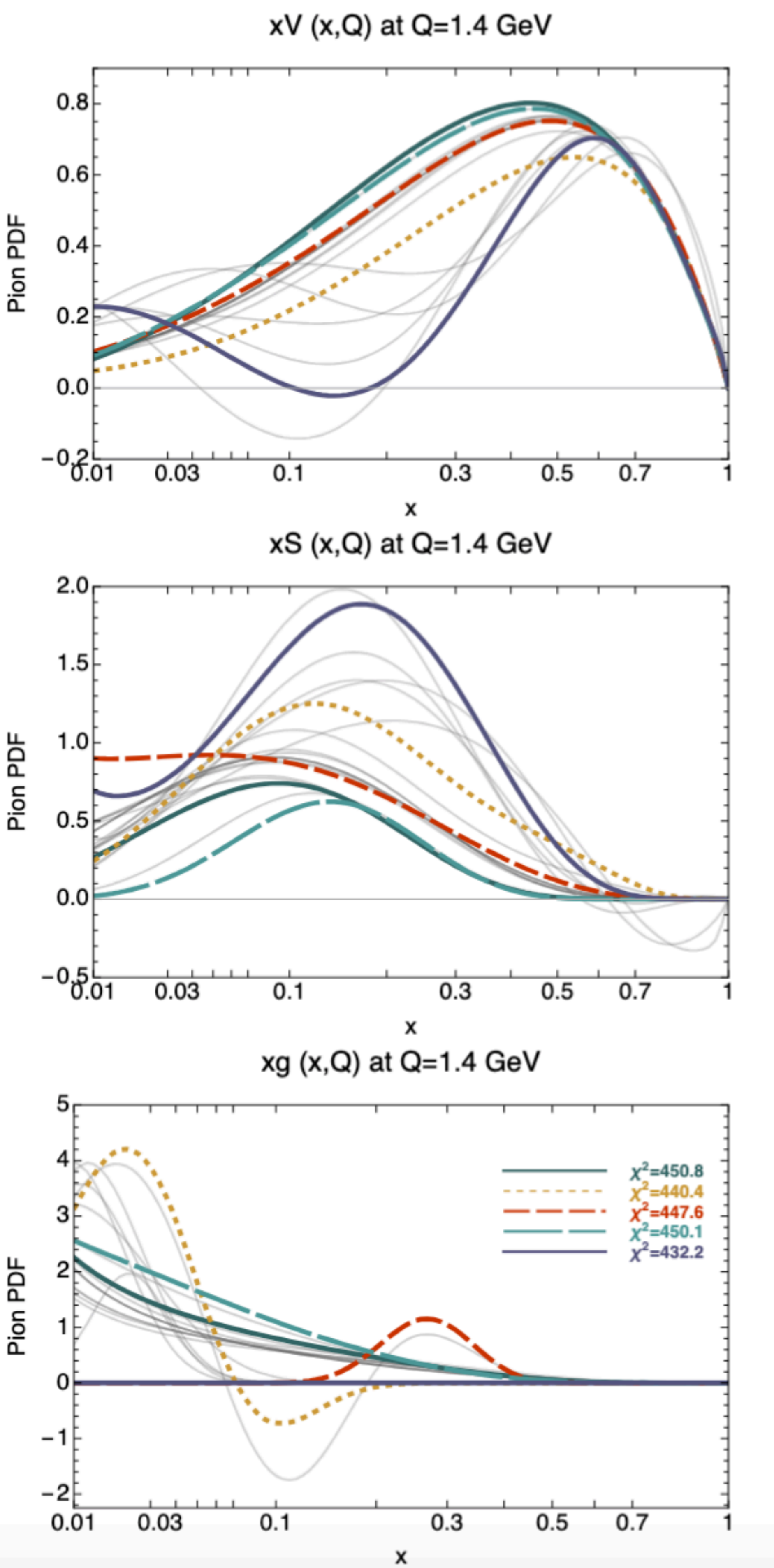
⇨ Fits within $\chi^2 + \delta\chi^2 = \chi^2 + \sqrt{2(N_{\rm pts} - N_{\rm par})}$ are kept.

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

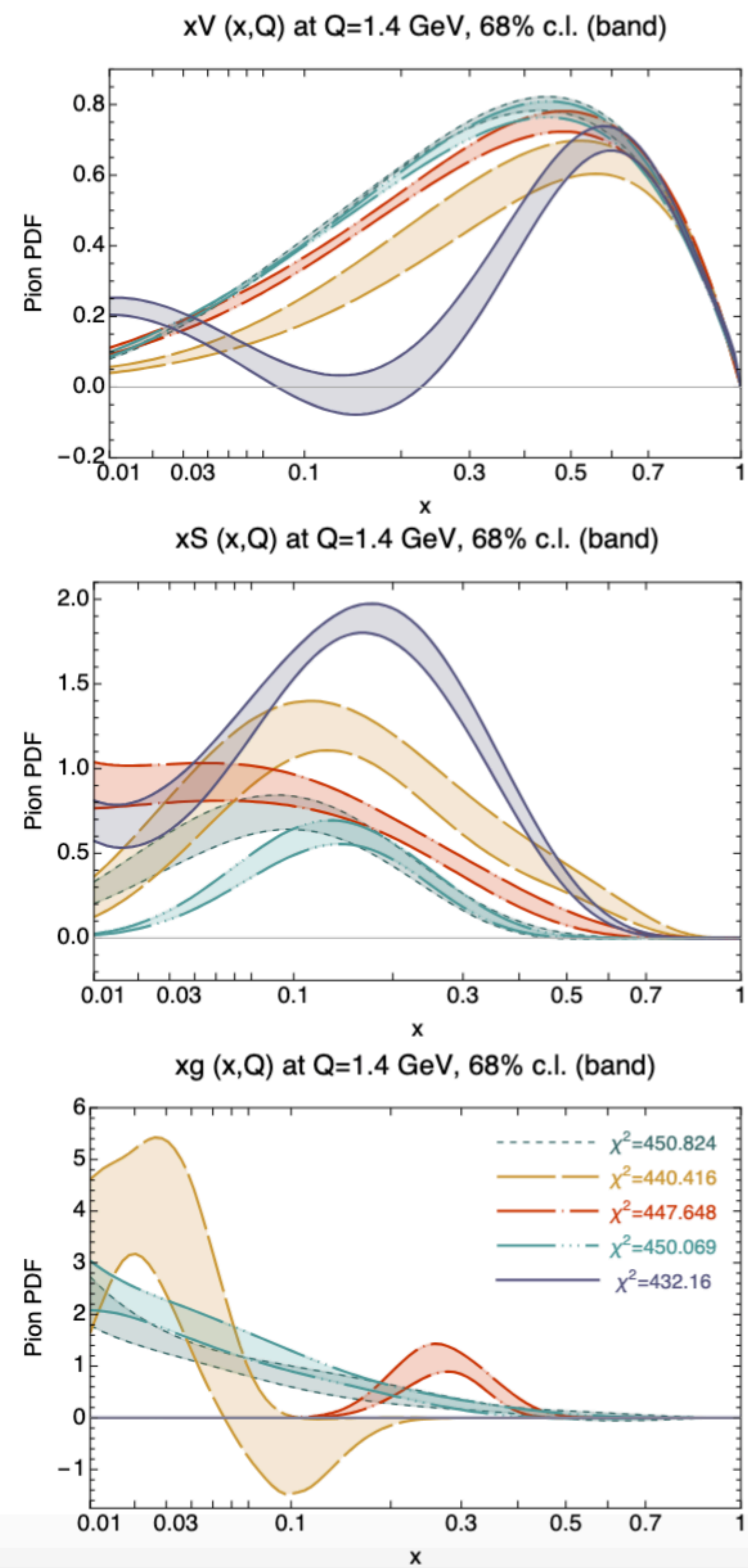
⇨ Bundled uncertainty using METAPDF (`mcgen`)

[Gao & Nadolsky, JHEP07]





Hessian-error bands evaluated with the default $\Delta\chi^2 = 1$ criterion



For π^+ PDFs,

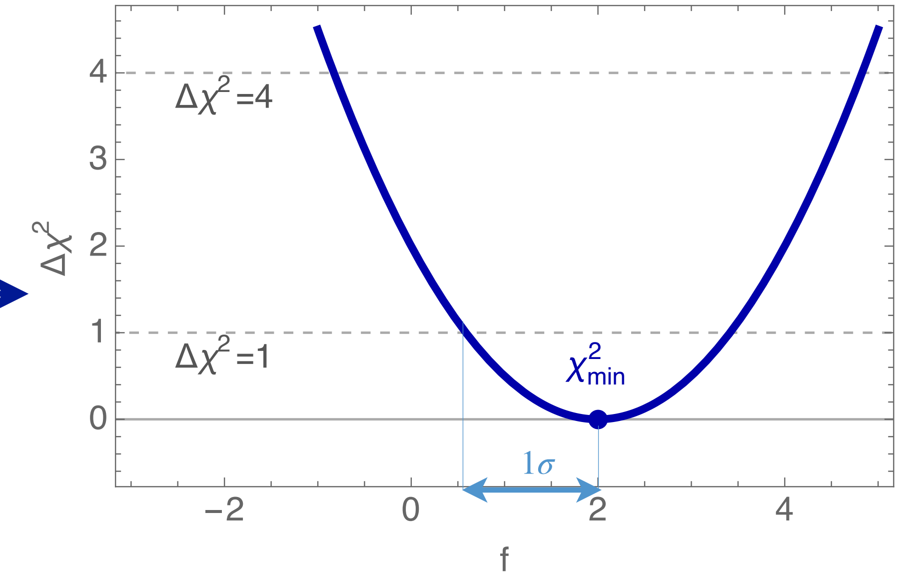
$q = V = 2(u - \bar{u})$,

$q = S = 6\bar{s}$

$q = g$

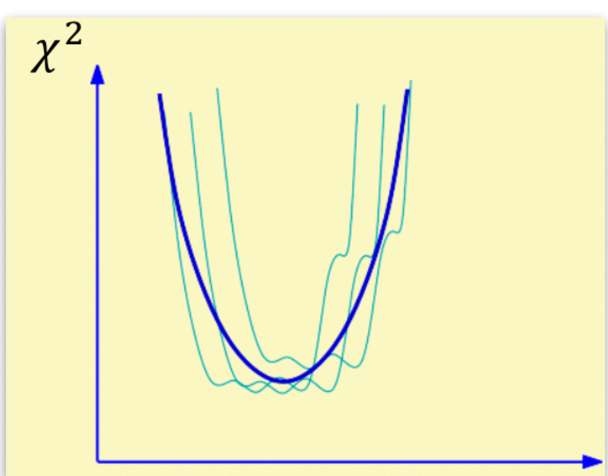
Uncertainties in global analyses

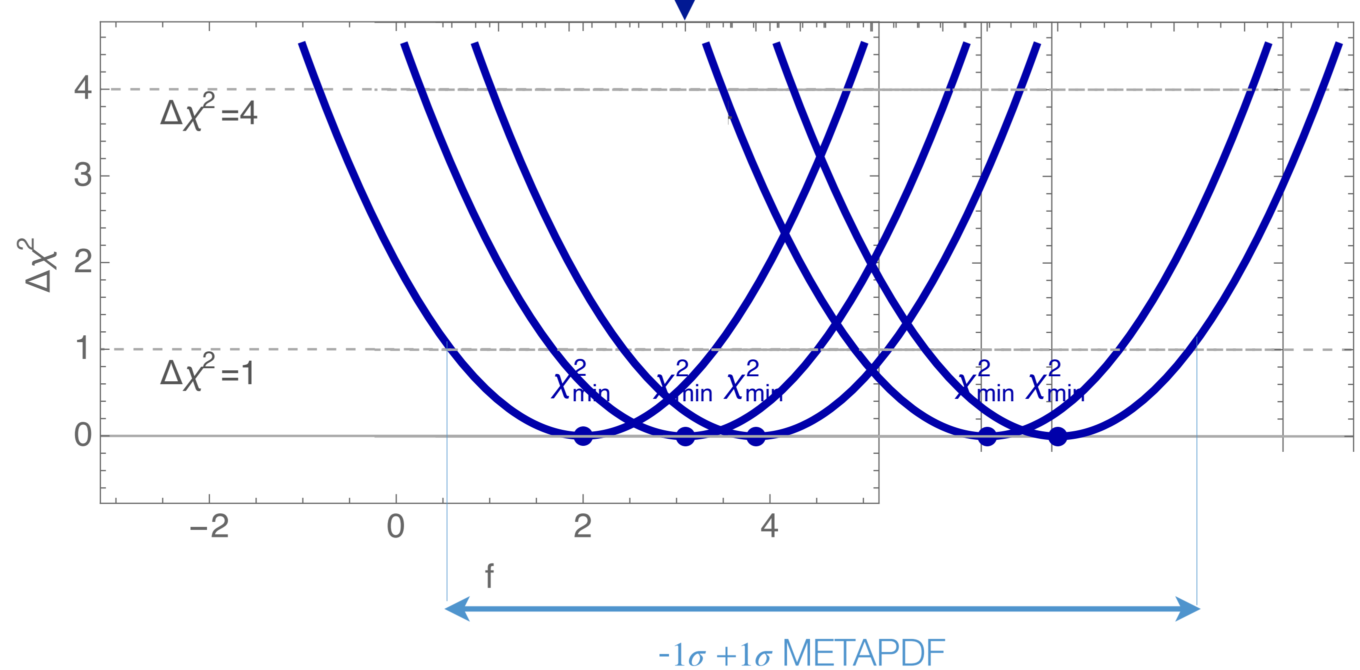
The χ^2 is a paraboloid in N_{par} dimensions.
We can project each dimension as



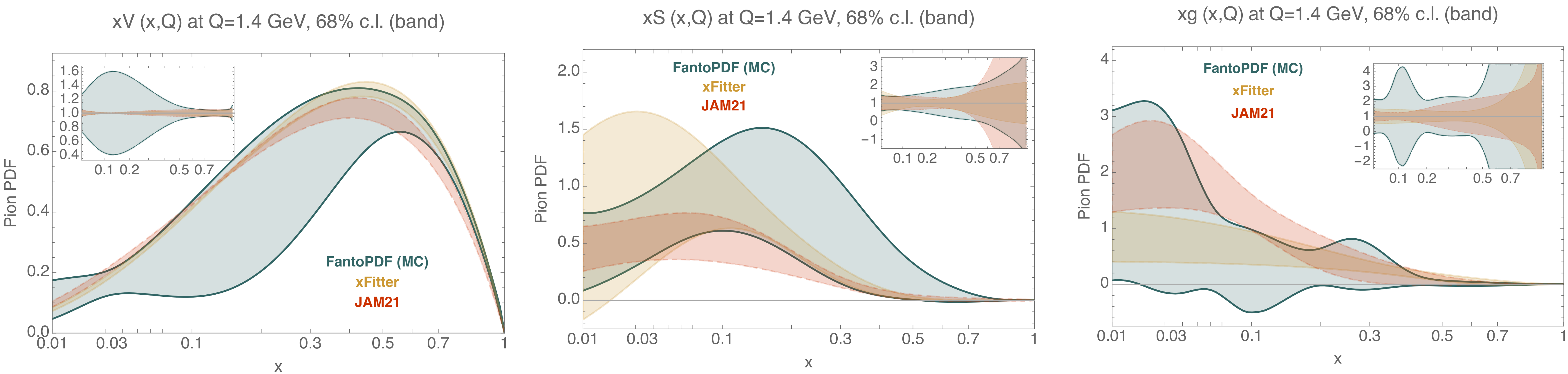
The $\Delta\chi^2 = 1$ criterion accounts for the 68% experimental uncertainty for the fixed settings of the fit. Additionally, we account for the uncertainty due to the PDF functional form using the METAPDF method.

Origin of $y-$axis varies according to $\chi^2 + \delta\chi^2$.





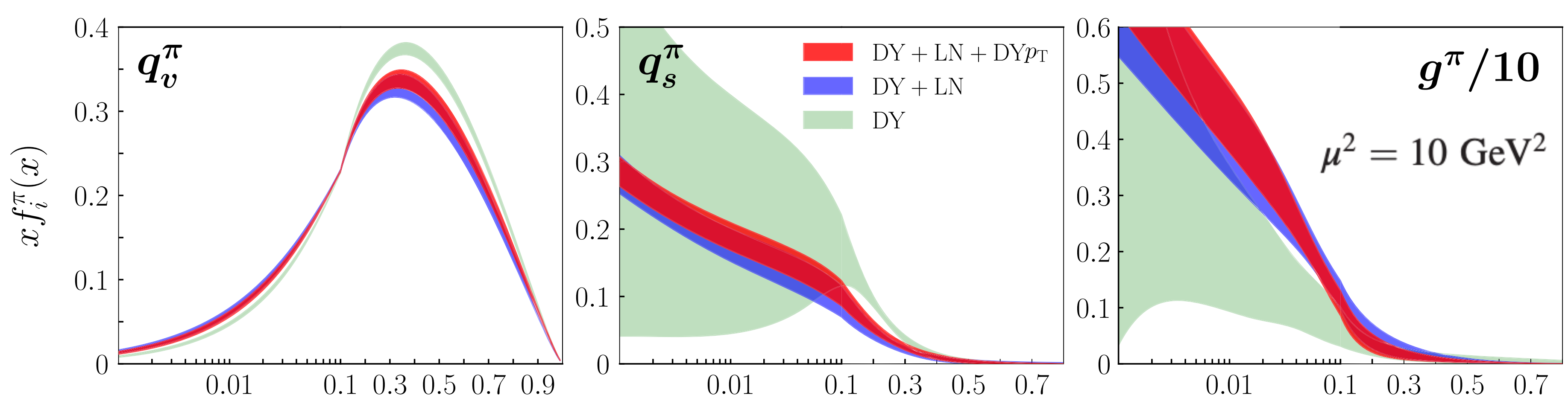
The Fantômas pion PDFs



<u>Comparison of methodologies:</u>

bootstrap+ IMC vs. `metamorph` parametrization in xFitter

Addition of leading-neutron data does not reduce the uncertainties for Fantômas



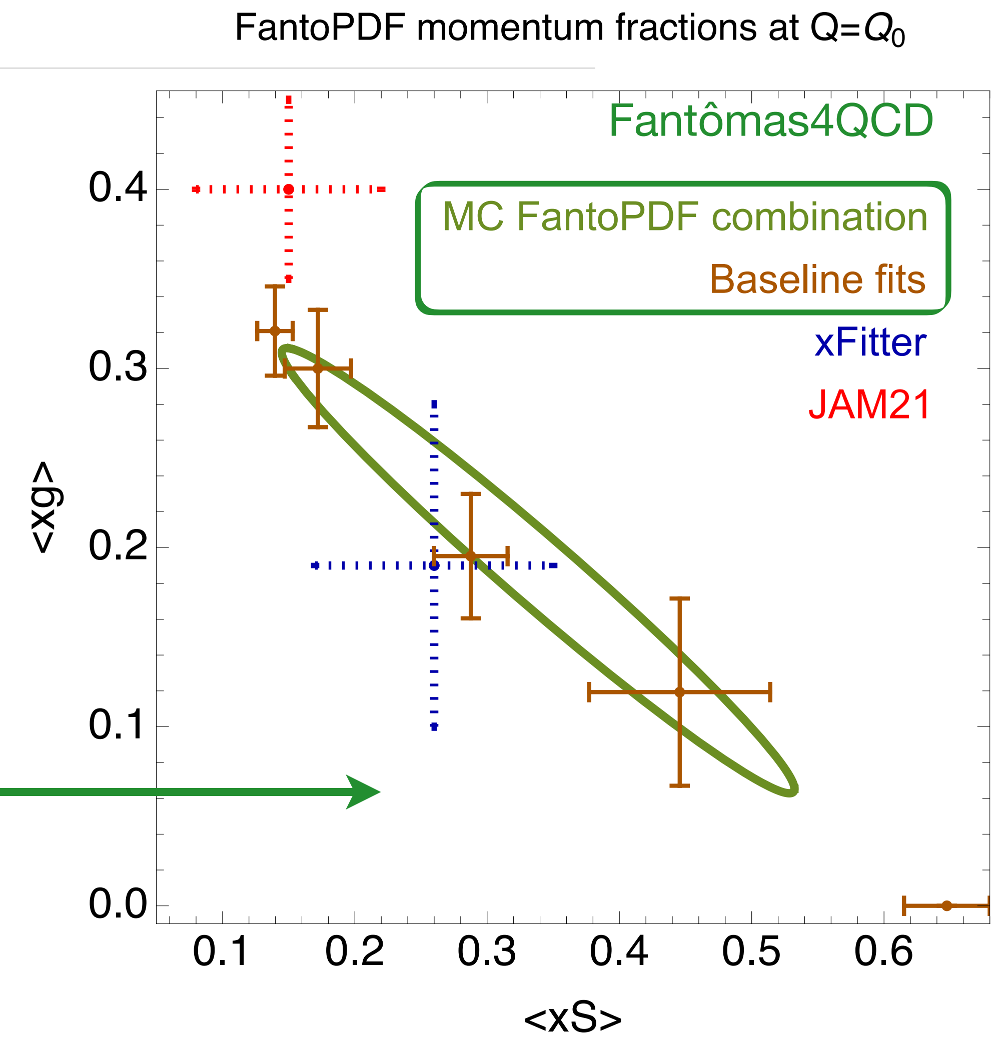
Distribution of the pion momentum

Momentum fraction x weighted by the PDF for $q = V, S, g$

$$\langle xq(Q^2)\rangle = \int_0^1 dx\, x\, f^q_{1,\pi}(x, Q^2)$$

Highlight on the separation of sea and gluon distributions.

The addition of leading-neutron data does not dramatically change the momentum fractions once the uncertainty appropriately include representative sampling.



Distribution of the pion momentum

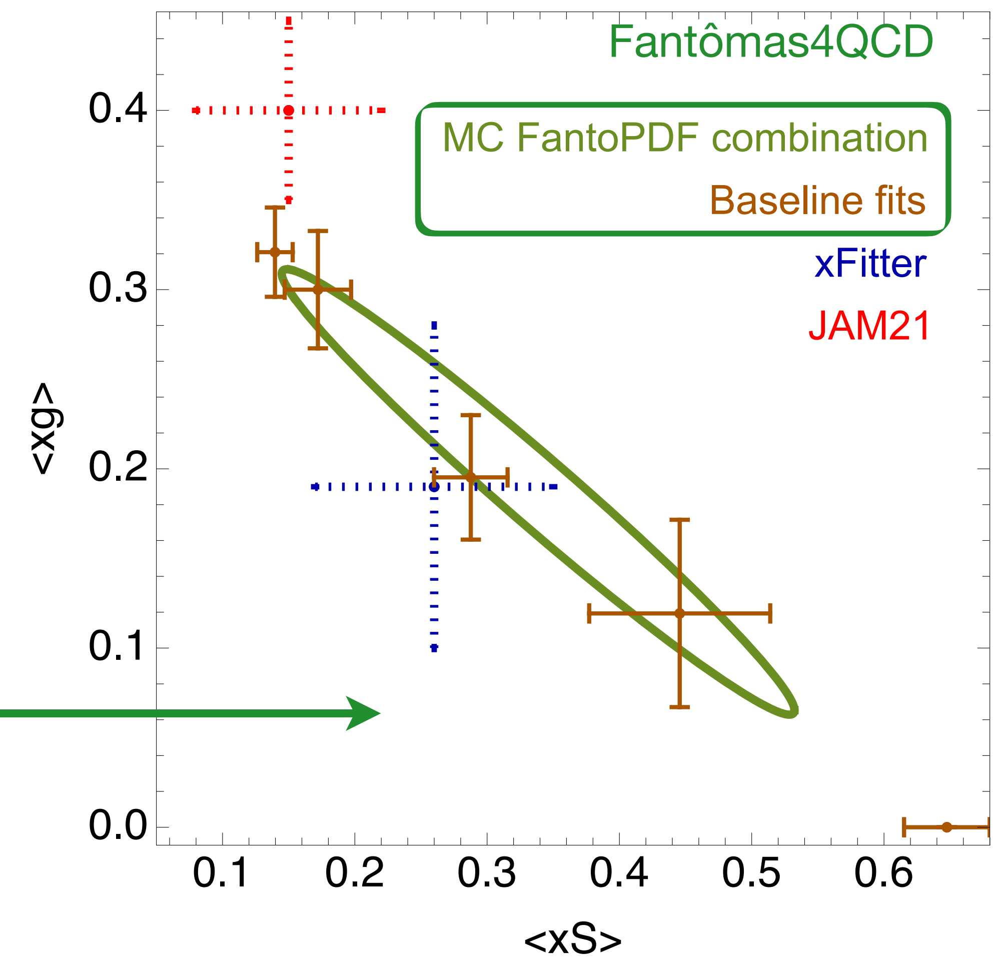
Momentum fraction x weighted by the PDF for $q = V, S, g$

$$\langle xq(Q^2)\rangle = \int_0^1 dx\, x f^q_{1,\pi}(x, Q^2)$$

Highlight on the separation of sea and gluon distributions.

The addition of leading-neutron data does not dramatically change the momentum fractions once the uncertainty appropriately include representative sampling.

FantoPDF momentum fractions at Q=Q_0



Name	Q[GeV]	$\langle x(u+\bar{u})_{\pi^+}\rangle$	$\langle xg\rangle$
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)
[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	–

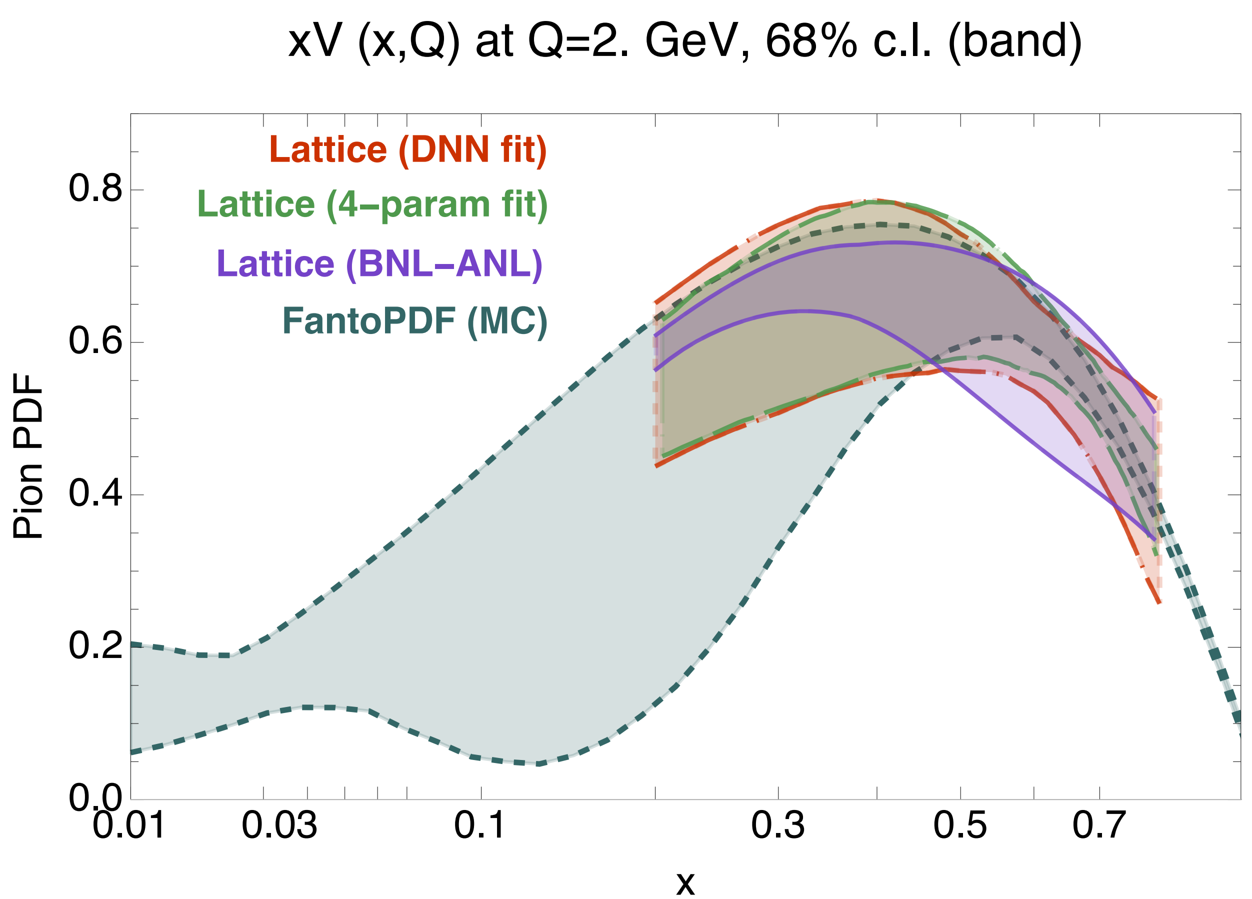
Lattice provides complementary access to momentum fractions— only the recent ETM coll. results have both.

All lattice results are work with different ensemble settings.

Pion PDF compared with lattice QCD results

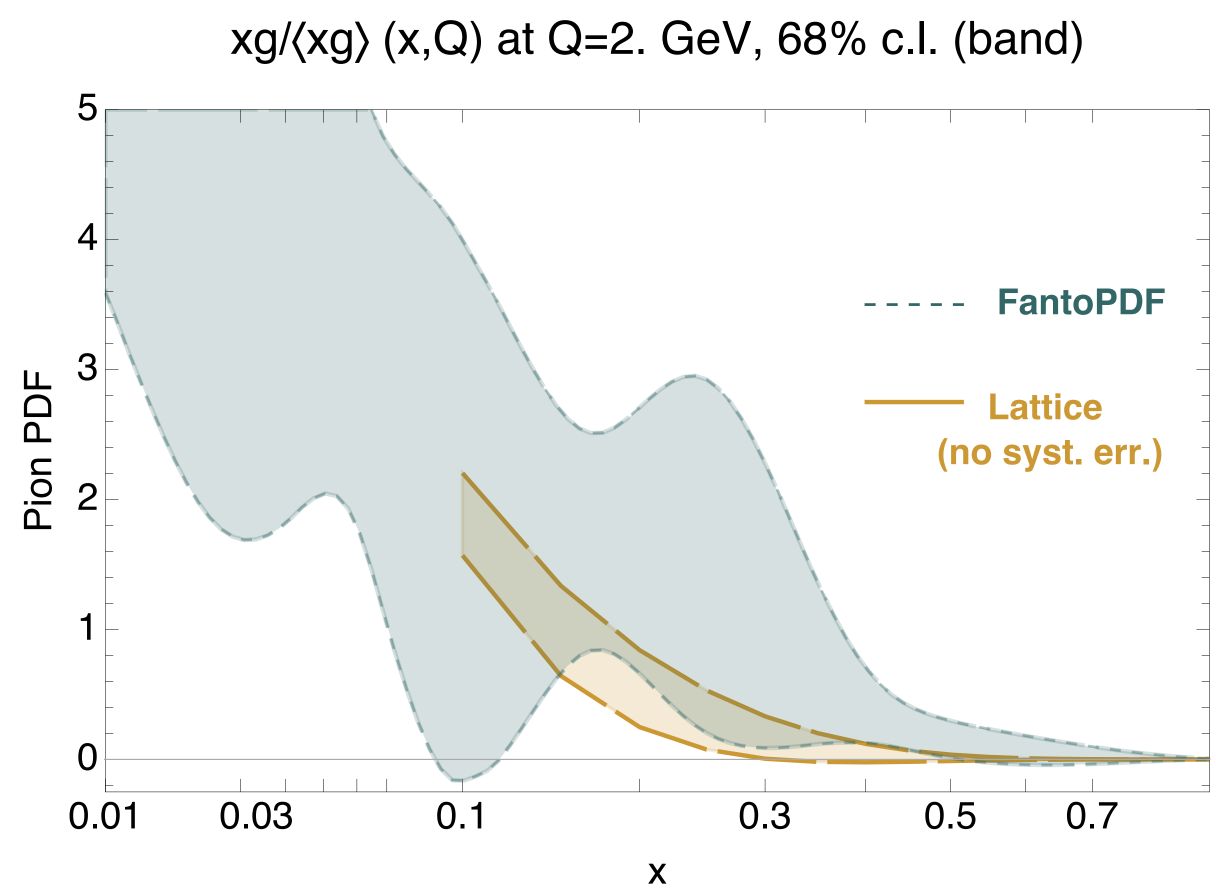
Valence pion PDF compared to lattice results from

[X. Gao, PRL128 & PRD106]



Gluon shape averaged to momentum fraction given by

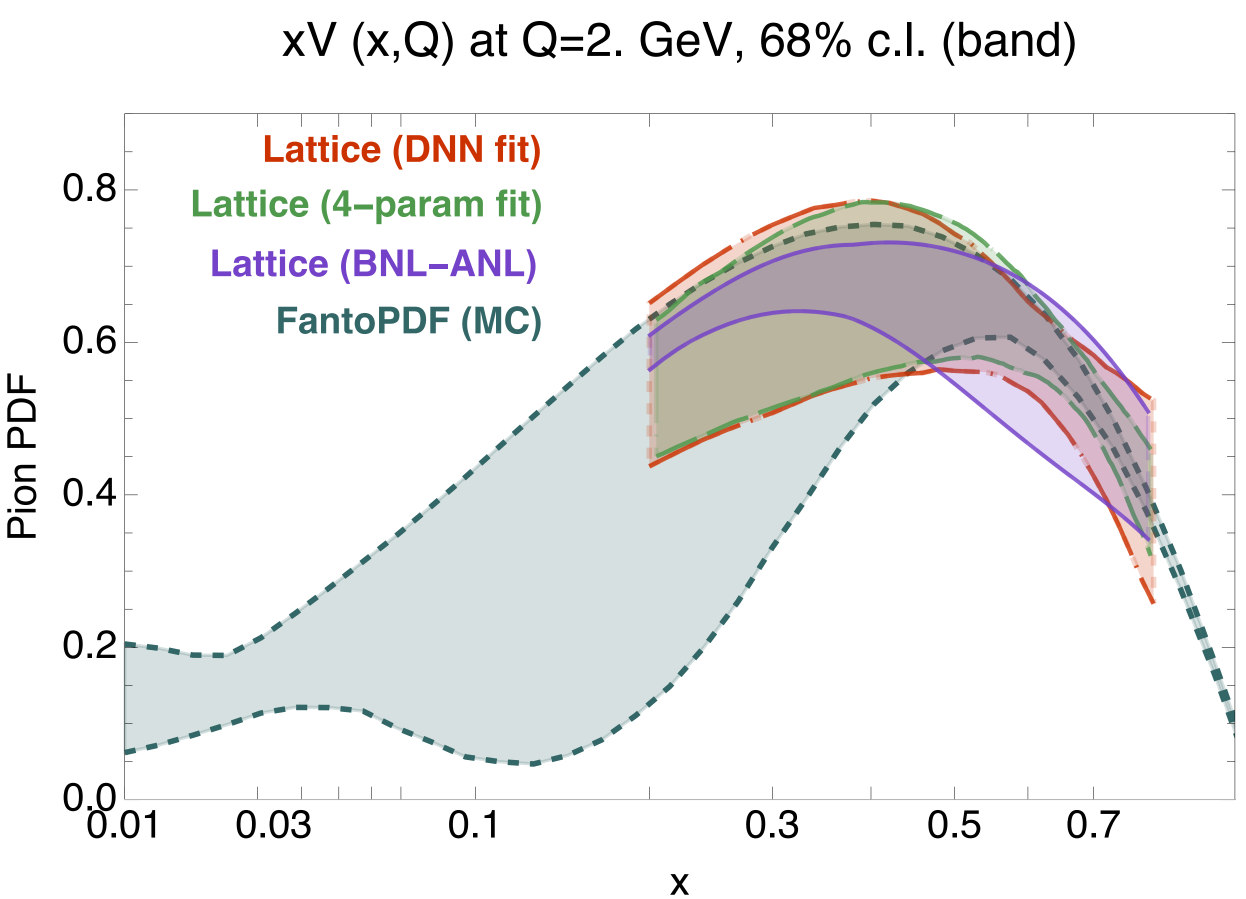
[Fan & Lin, PLB 823 (2021)]



Lattice PDFs come from solving inverse problems, too

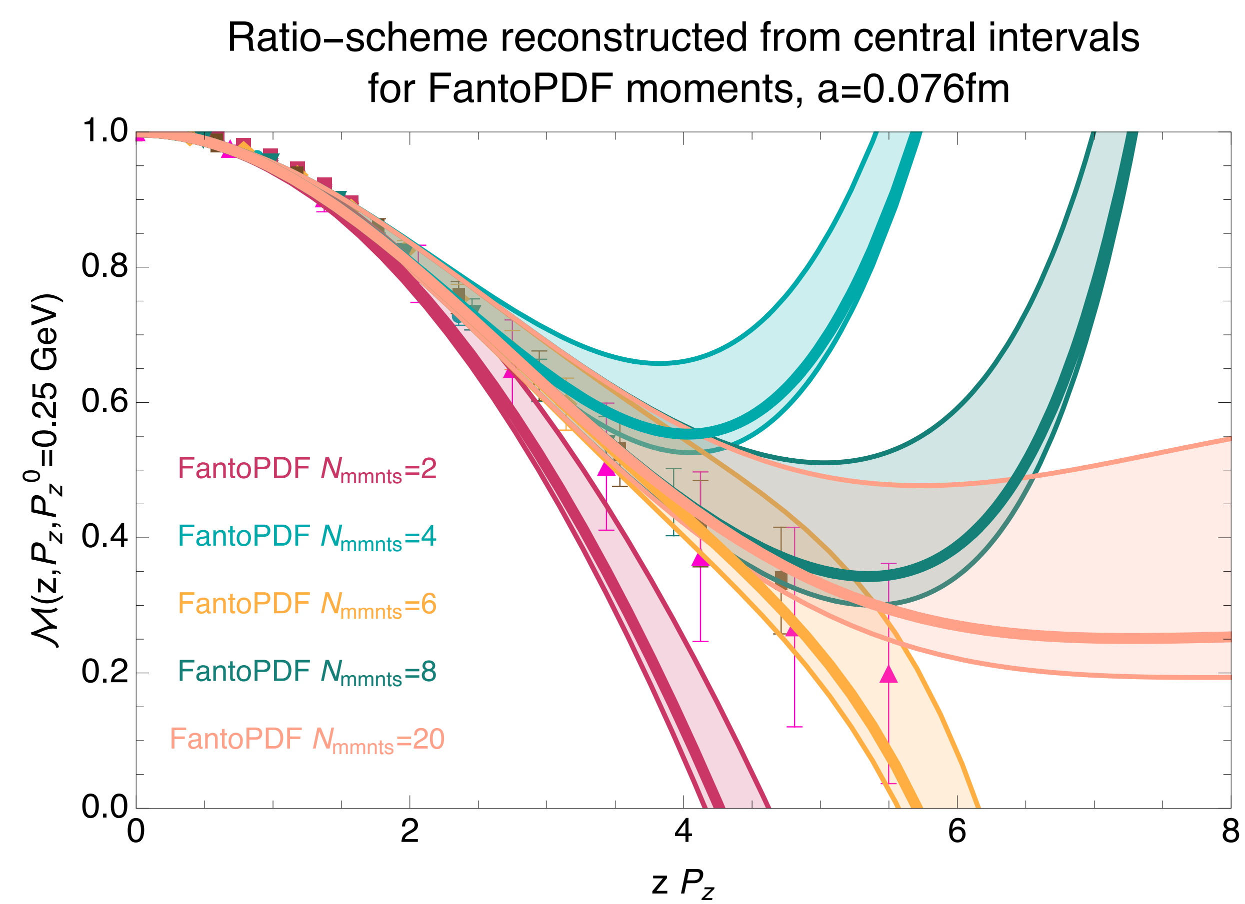
Valence pion PDF compared to lattice results from

[X. Gao, PRL128 & PRD106]

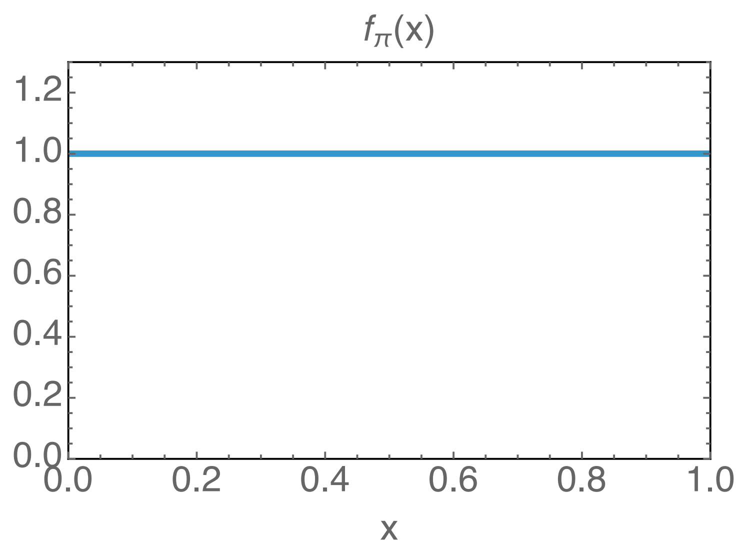


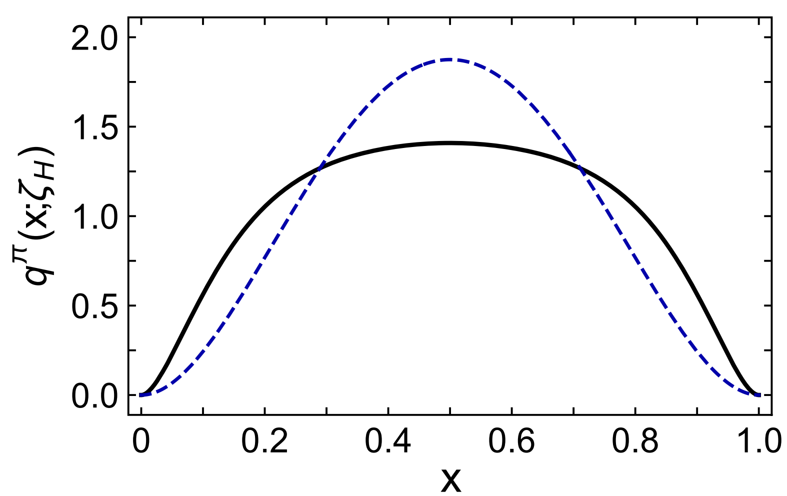
Matrix element through OPE in Ioffe-time from Fantômas

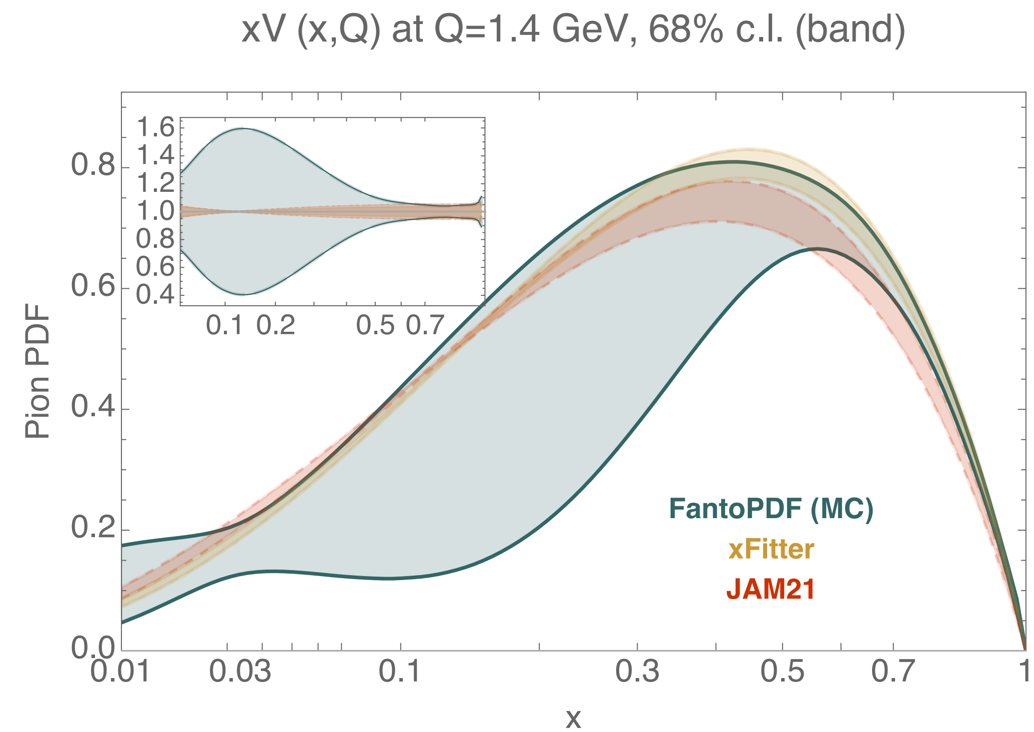
[in progress]



Evolution of nonperturbative manifestations

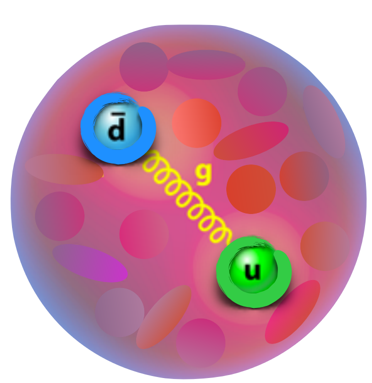


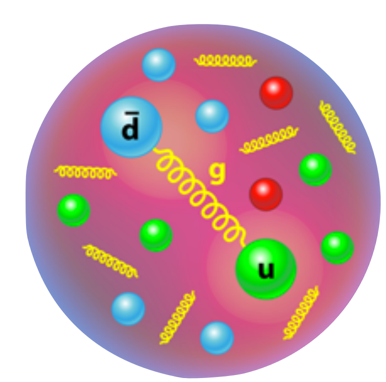




Pheno PDF given at Q_0^2

but obtained from data at higher Q^2





Hints of the mechanism that drives the pion structure

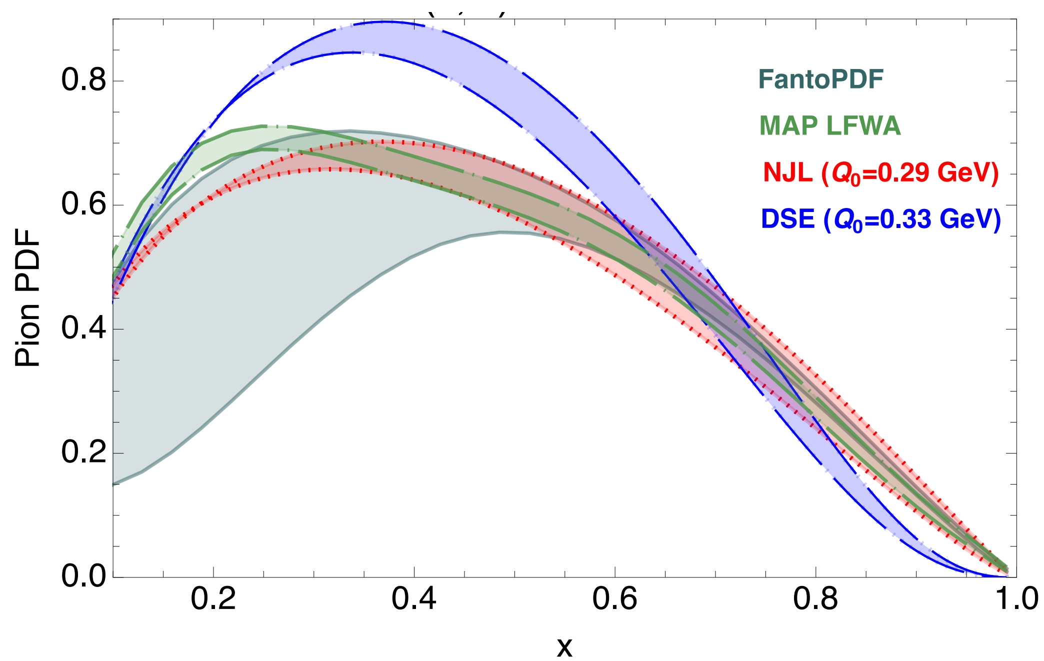
When testing polynomial shapes predicted from models, polynomial mimicry affects any interpretation.
No *if and only* conditions are possible given the state-of-the-art. [A.C. & Nadolsky, PRD103]

Contact-like kernel (NJL) and momentum-dependent kernel @ all order (DSE) calculations prescribe different initial conditions (Q_0^2 & *shape*), that evolve to different predictions at the scale of the data.

Light-front quark model with data-inferred parameters finds a similar large-x behavior.

[Ruiz-Arriola; Ding et al, PRD101]
[Pasquini et al, PRD107]

$2x(u-\bar{u})$ at $Q^2 = 10\,\mathrm{GeV}^2$



Comparing shapes, by evolving models from dangerously small scales.

Hints of the mechanism that drives the pion structure

When testing polynomial shapes predicted from models, polynomial mimicry affects any interpretation.

No *if and only* conditions are possible given the state-of-the-art. [A.C. & Nadolsky, PRD103]

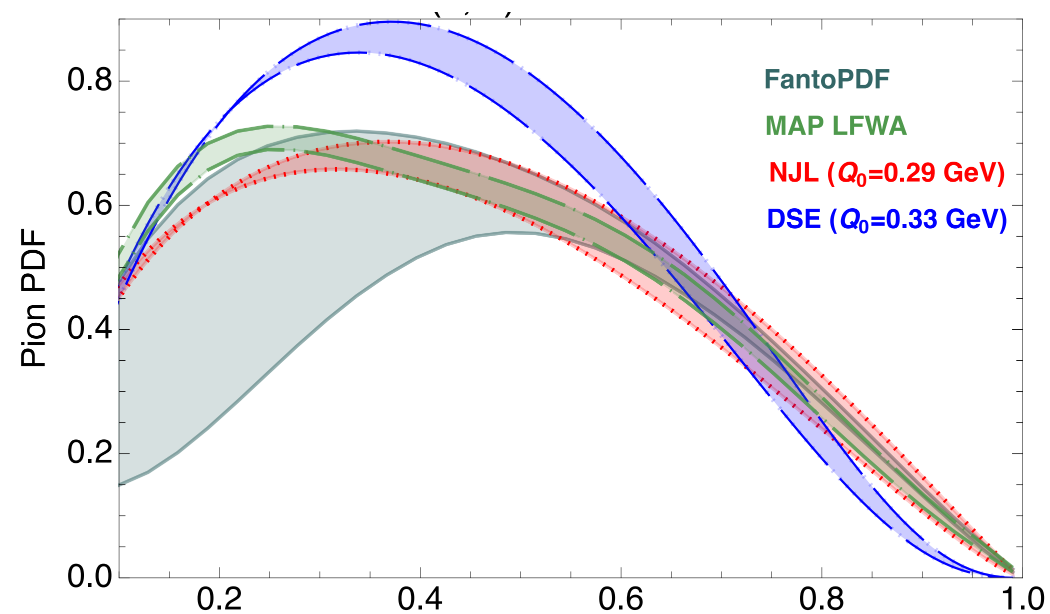
Contact-like kernel (NJL) and momentum-dependent kernel @ all order (DSE) calculations prescribe different initial conditions (Q_0^2 & *shape*), that evolve to different predictions at the scale of the data.

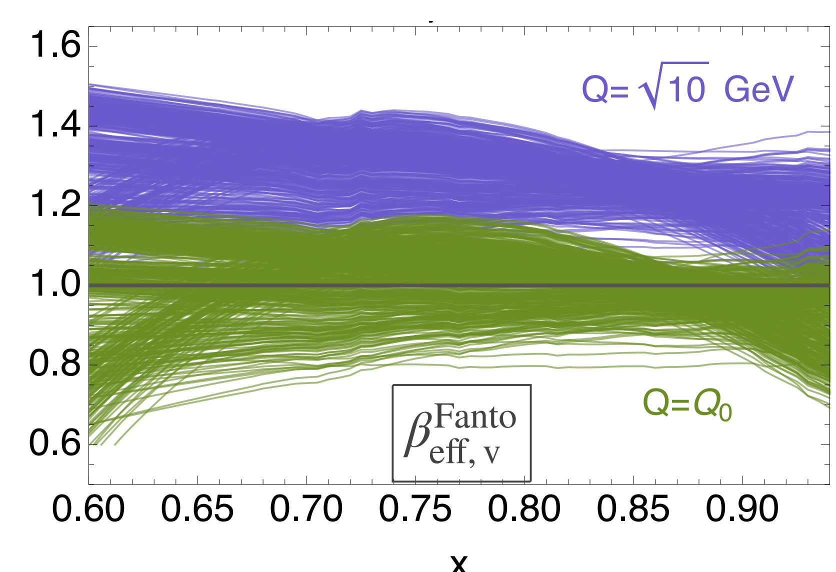
Light-front quark model with data-inferred parameters finds a similar large-x behavior.

[Ruiz-Arriola; Ding et al, PRD101]

[Pasquini et al, PRD107]

$2x(u-\bar{u})$ at $Q^2 = 10\,\mathrm{GeV}^2$





Quark-counting rules: $f_{q_v/\pi}(x) \xrightarrow[x\to 1]{} (1-x)^2$

All pheno analyses find $f_{q_v/\pi}(x, Q_0^2) \xrightarrow[x\to 1]{} (1-x)^{\beta_{\mathrm{eff}}=1}$

Conclusions

Towards epistemic PDF uncertainties with Fantômas4QCD.

Towards augmenting the aleatory $\Delta\chi^2 = 1$ uncertainties with the uncertainty due to parametrization.

The pion structure is currently being studied (again) on various (improved) fronts — lattice, experiments, theory.

⇨ Rôle of the parametrization in the sampling accuracy: we make use of Bézier-curve methodology

Fantômas4QCD framework
`metamorph` can be used to study many functions

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

⇨ Uncertainties come from various sources in global analyses.
Extension to sampling accuracy, here sampling occurs over parametrization forms.

⇨ Sea-gluon separation requires more data — a very interesting sector!
⇨ End-point behavior appears to play an important rôle.

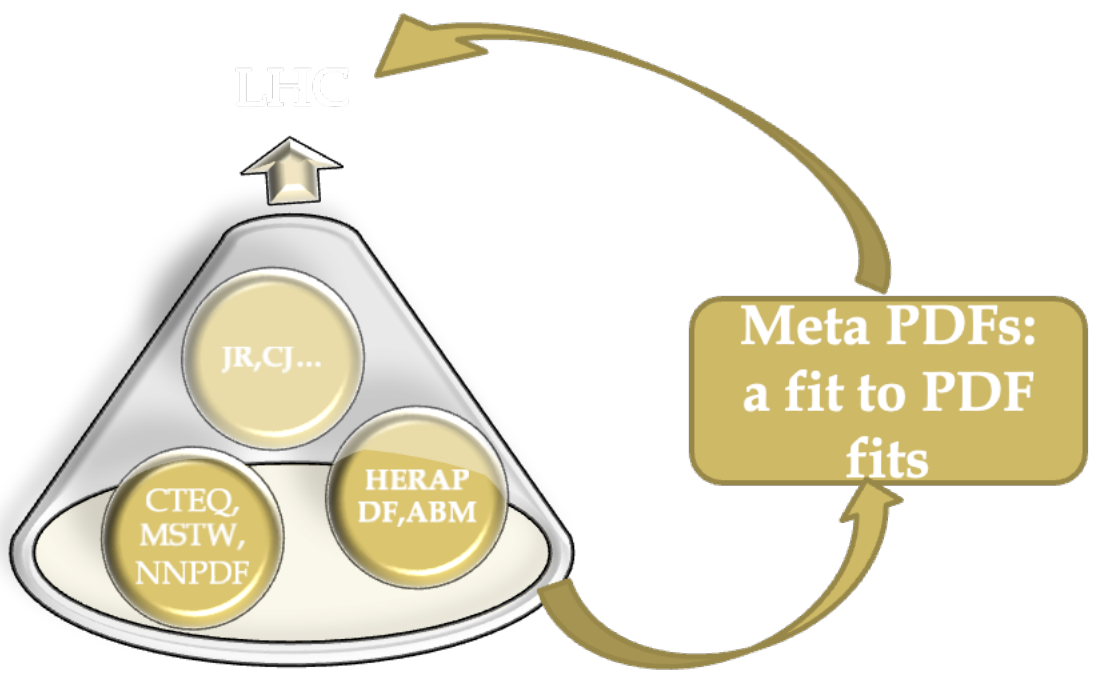

FANTOMAS
LA AMENAZA ELEGANTE
Presenta:
'LA INTELIGENCIA EN LLAMAS'
Un episodio excepcional. . . arde la cultura del mundo. . ¡Vea a Fantomas en apuros, entrevistándose con los más grandes escritores contemporáneos!
MEJORES REVISTAS
$2.00
M.N.
Con TABLA
Bingo
PEPSI

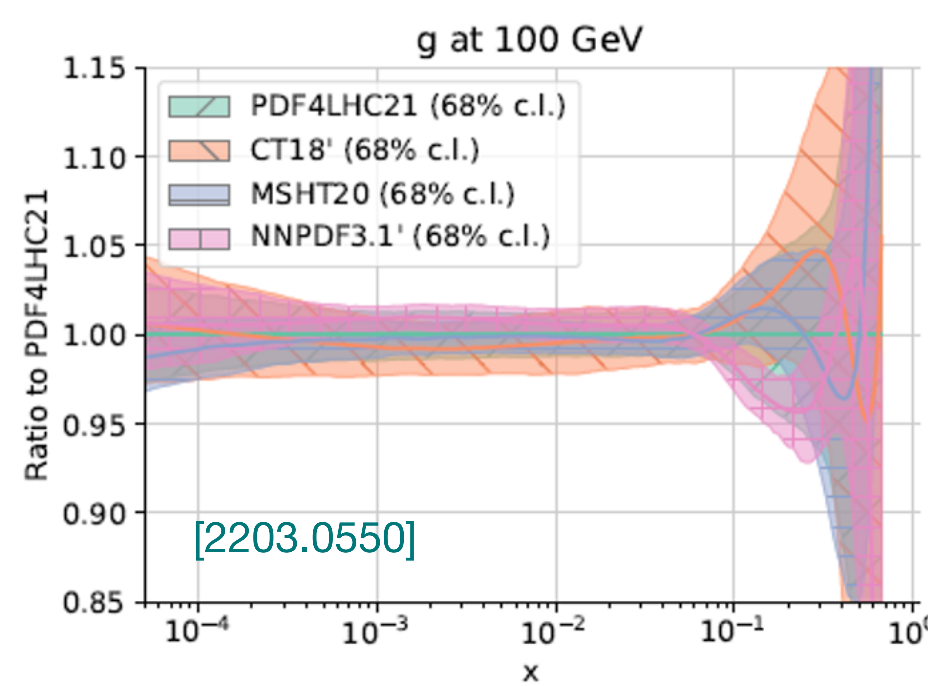
Back up

A META combination of parton distributions

[Gao & Nadolsky, JHEP07]

- A technique to compare and combine PDF ensembles from various groups
- Relies on the Hessian→MC→Hessian conversion using multi-dimensional sampling and PCA from 150 to 30-40 PDF parameters
- Combines CT, MSHT, NNPDF sets in the PDF4LHC 2015 and 2021 combinations



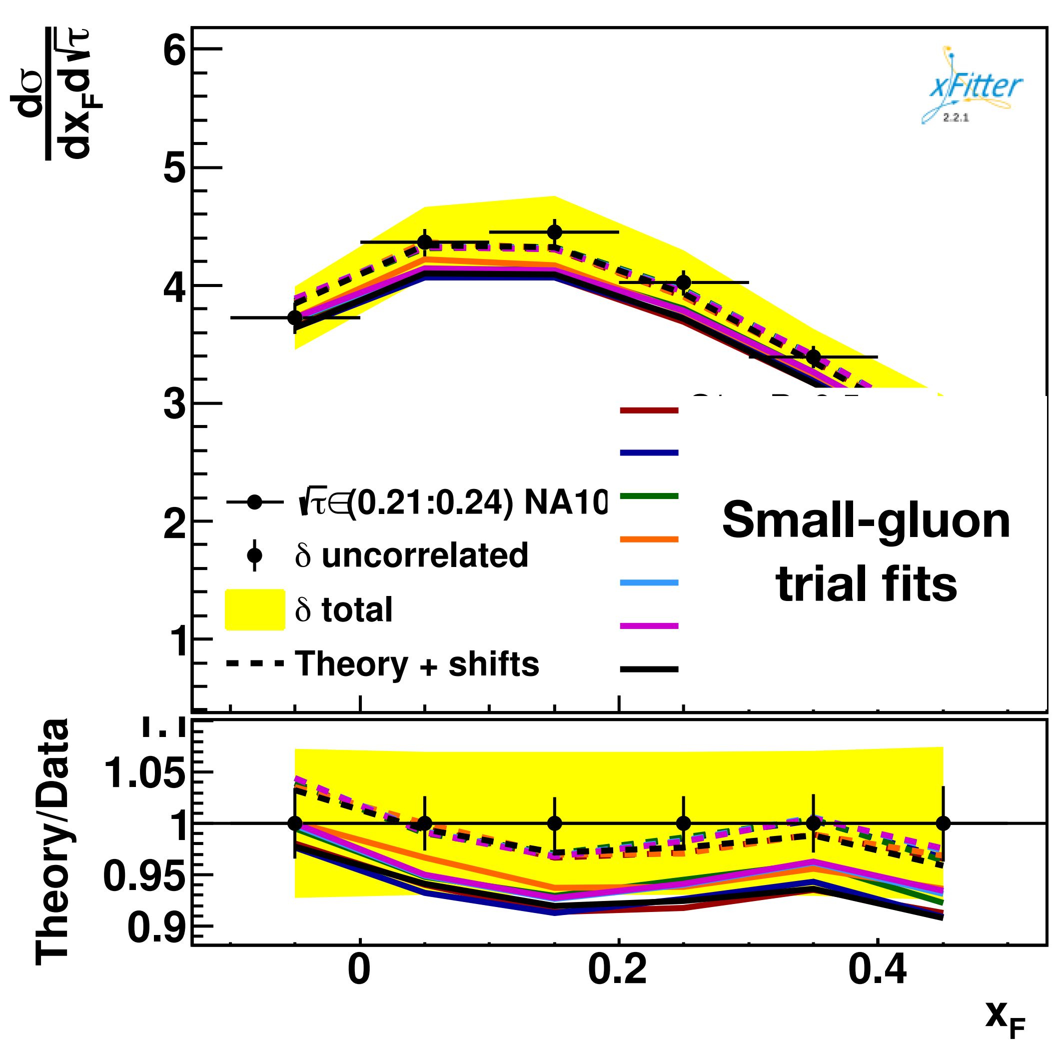


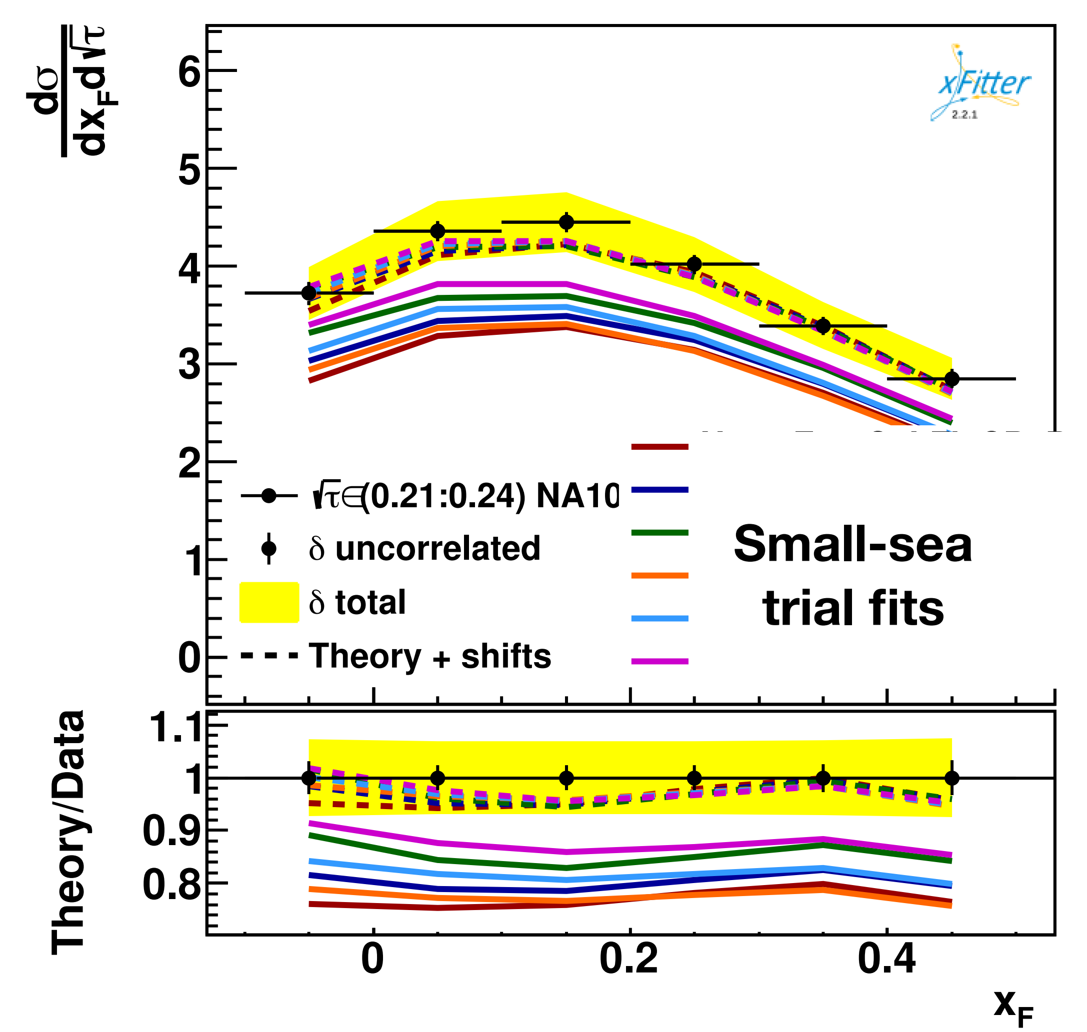
Update on the `mp4lhc` and `mcgen` codes in the context of Fantômas [Kotz et al, in progress]

Sea and gluon behavior

Data sets vary between JAM and Fantômas: higher number of NA10 data points for us.

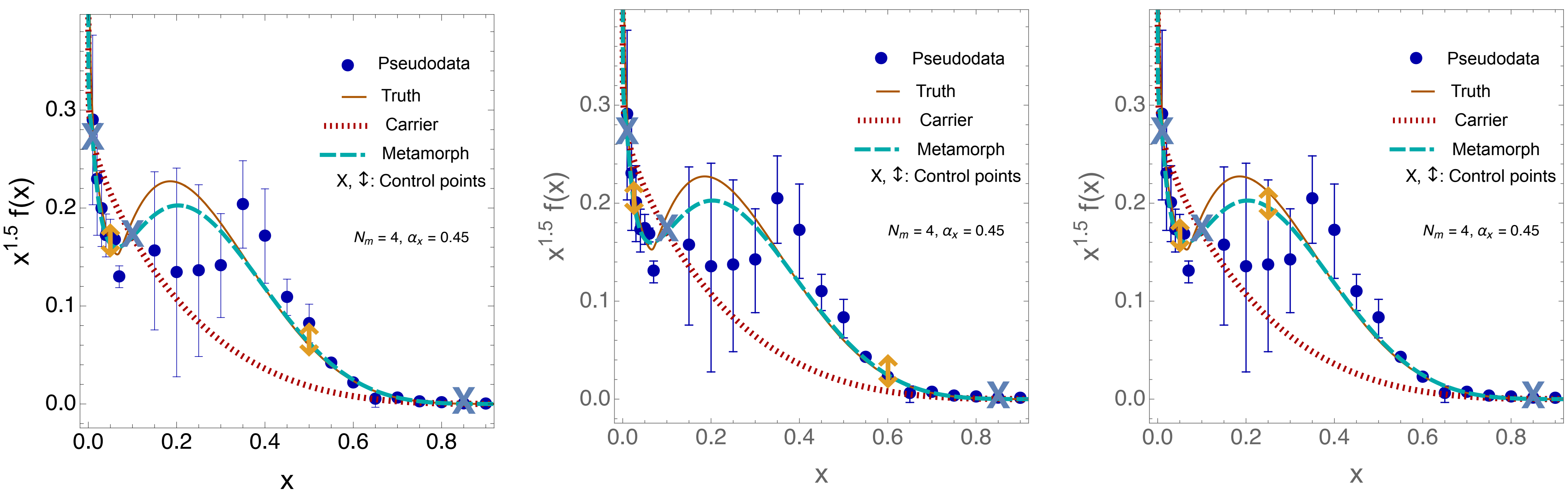
We explored small gluon and small sea scenarios:
zero-gluon solutions are allowed; zero-sea ones are unfavored.



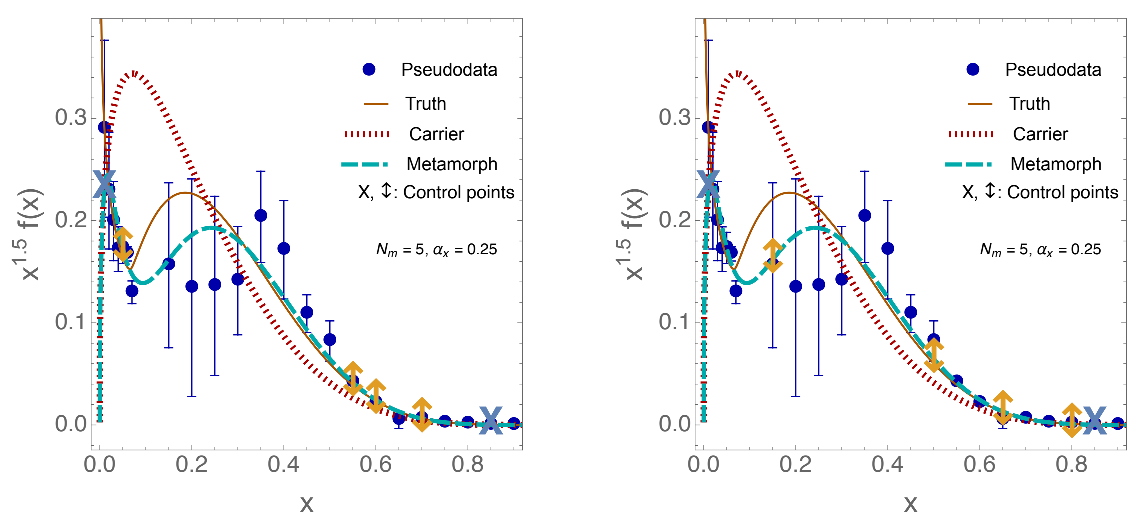


Rôle of the control points

X = fixed CPs
↕ = free CPs



The fixed CPs are the intersection points between the carrier and the final `metamorph`.
The fixed CPs set the shape of the curve ; the free CPs act through the minimization procedure.



. . .

find the balance between fit and interpolation

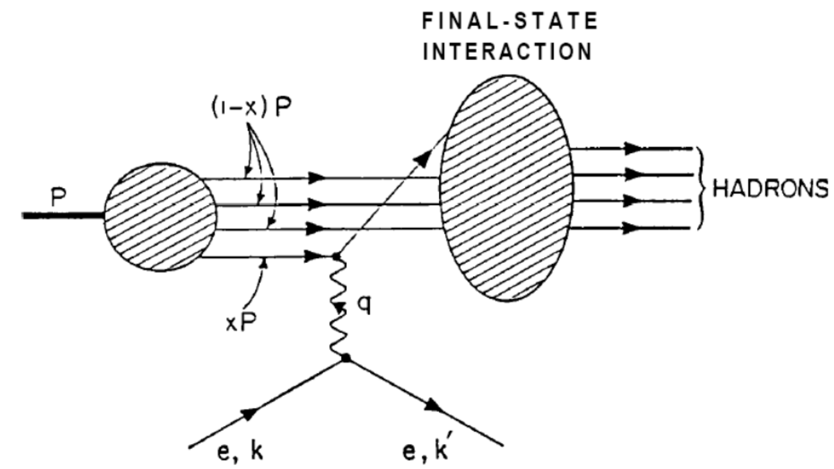
Testing quark counting rules

⇨ <u>Hypothesis testing for functional behavior constraints</u> — *do PDFs fall off like* $(1-x)^\beta$*?*

Quark-counting rules:

Early-QCD predicted behavior for structure functions when one quark carries almost all the momentum fraction

$$f_{q_v/P}(x) \xrightarrow[x\to 1]{} (1-x)^3, \qquad f_{q_v/\pi}(x) \xrightarrow[x\to 1]{} (1-x)^2$$



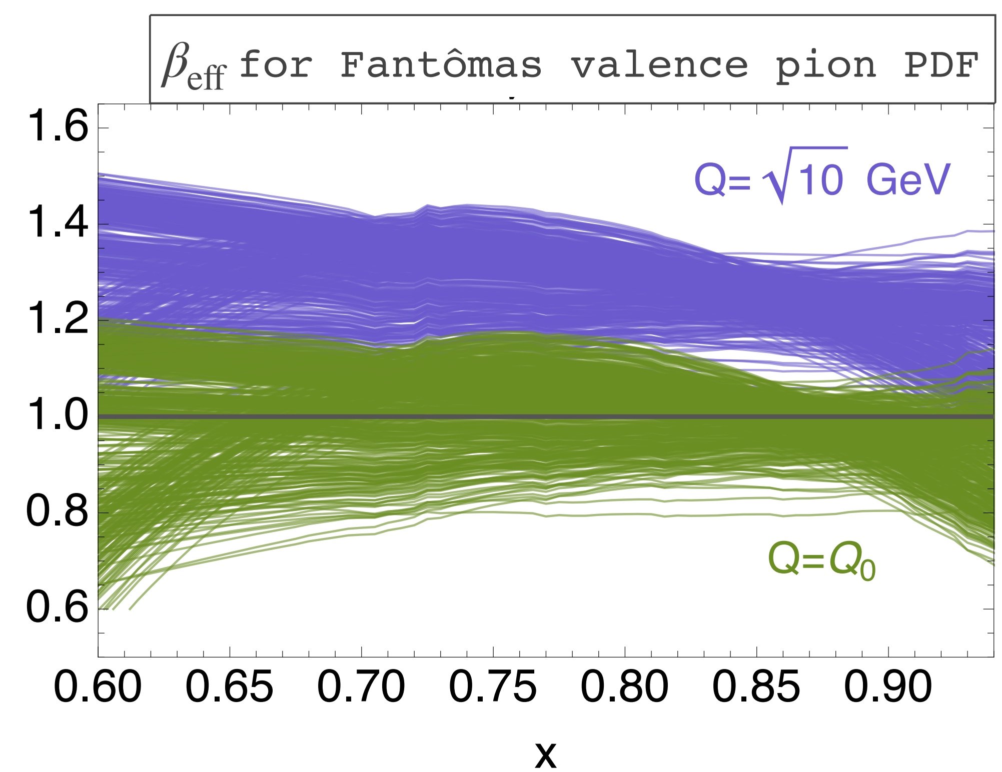
At NLO (MSbar), the valence PDF is well determined at large x

⇨ doesn't fall very much like $(1-x)^2$

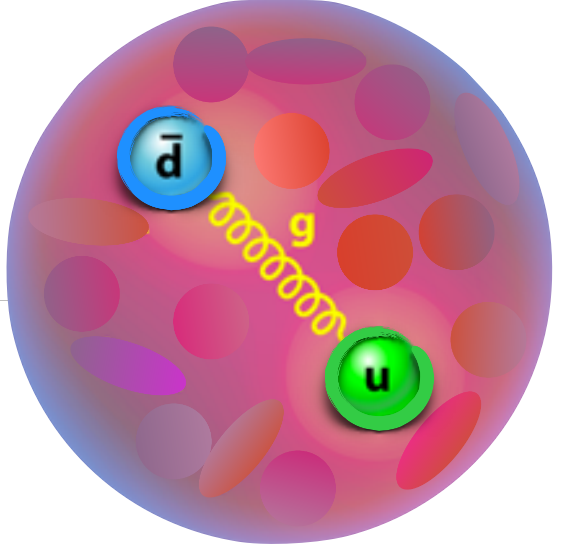
⇨ very similar to JAM and xFitter at large x

<u>This result can be understood through non-perturbative QCD corrections as well as polynomial mimicry.</u>

[A.C. & Nadolsky, PRD103]



Pion structure — amazing pioneer work



Pion's quark and gluon structure can be, and has been, studied in connection with its fundamental properties:

- **phenomenologically** — early QCD predictions on the shape of form factors and structure functions

  70'-80' [Brodsky, Lepage, Farrar, Soper,…]

- **theoretically** — quark models, EFTs with pion or quark degrees of freedom, Dyson-Schwinger eqs.… that, ideally, incorporate the physics of χ sym. breaking and pion as a bound-state,

  e.g., Nambu—Jona-Lasinio model, chiral-quark soliton model, continuum…

  — systematics and representations of non-perturbative functions.

  90'-00' [Ruiz-Arriola, Polyakov, …]

taking care of *many* relevant analytical aspects of the pion structure.

- **experimentally**:
  - exclusive processes — that give access to form factors

    70'-20' [DESY, SPS, Belle, BaBar, JLab…]

  - inclusive processes — the deeply inelastic ones

    70'-30' [SPS, Fermilab, HERA (DESY), future: JLab+EIC]

Shape of pion PDFs

Quantities that characterize the distribution of quarks inside hadrons
⇨ **Parton Distribution Functions (PDFs)**

What is the fraction of longitudinal momentum carried by a struck parton from the hadron?
⇒ variable x, with $x \in [0,1]$

How does that momentum fraction change with the increase of the virtuality of the probe, Q^2?
⇒ $f_1(x) \to f_1(x, Q^2)$ where the Q^2 dependence is known precisely in perturbative QCD

Shape of pion PDFs

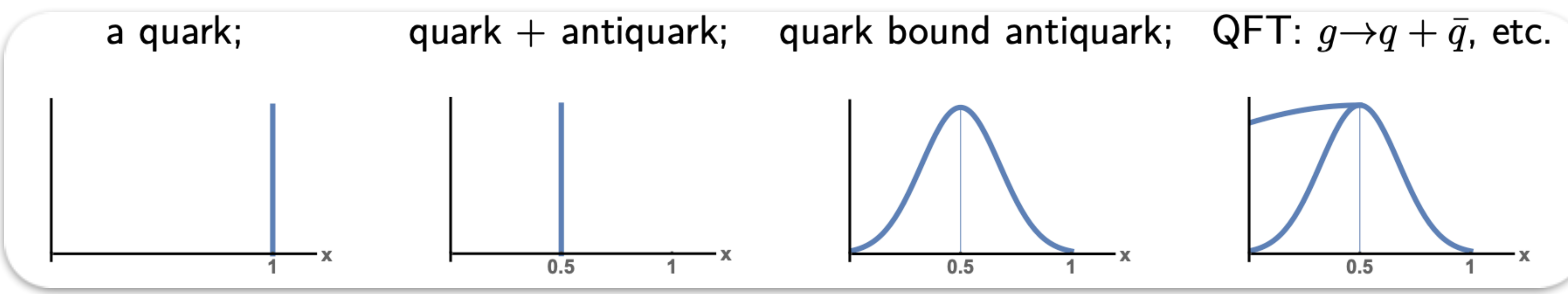
Quantities that characterize the distribution of quarks inside hadrons
⇨ **Parton Distribution Functions (PDFs)**

What is the fraction of longitudinal momentum carried by a struck parton from the hadron?
⇒ variable x, with $x \in [0,1]$

How does that momentum fraction change with the increase of the virtuality of the probe, Q^2?
⇒ $f_1(x) \to f_1(x, Q^2)$ where the Q^2 dependence is known precisely in perturbative QCD

Expectations from

from Minghui Ding

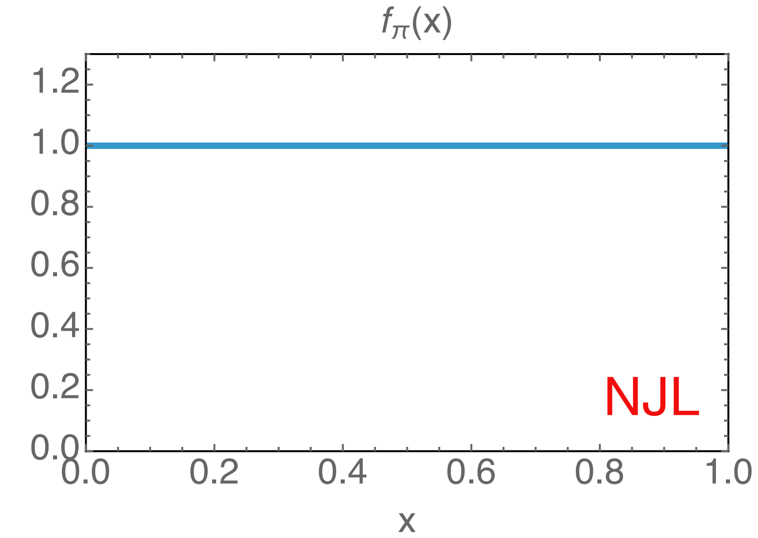


elastic limit — quark model $|\pi\rangle = |q\bar{q}\rangle$ — bound quarks — + QCD radiations

⇨ **Is one or more of those pictures correct in Nature? If yes, at which resolution scale?**

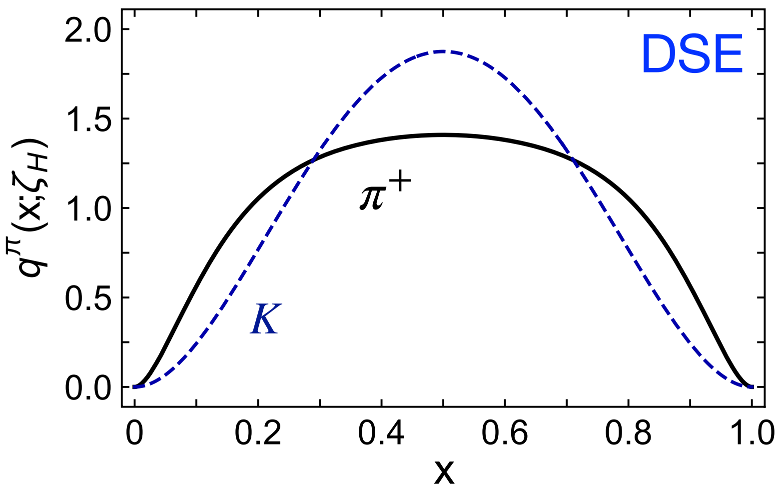
Hints about the low-energy QCD mechanisms

From the many theoretical studies on the pion, we expect manifestations of non-perturbative dynamics to be encapsulated in PDFs and related objects.



χ sym. breaking predicts that the pion PDF is broader than what expected from a bound-state smearing.

- in contact-like interaction models, it is exactly $f_1^\pi(x, Q_0^2) = 1$
- in mmt-dependent kernel models, it is broad, and 0 at the end-points
- *no approach that incorporates χ sym. finds a peak at $x = 0.5$*



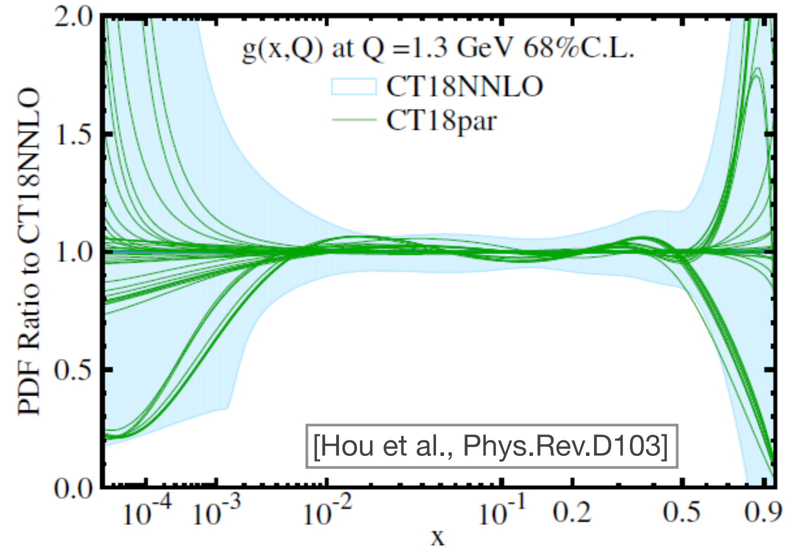
⇨ The shape of the valence pion PDF is intrisically related to the emergence of hadronic mass.

Proton and pion have very different dynamical contributions to their mass, the origin of this difference is also manifest through the respective shape of their PDFs.

Rôle of parametrization in previous analyses

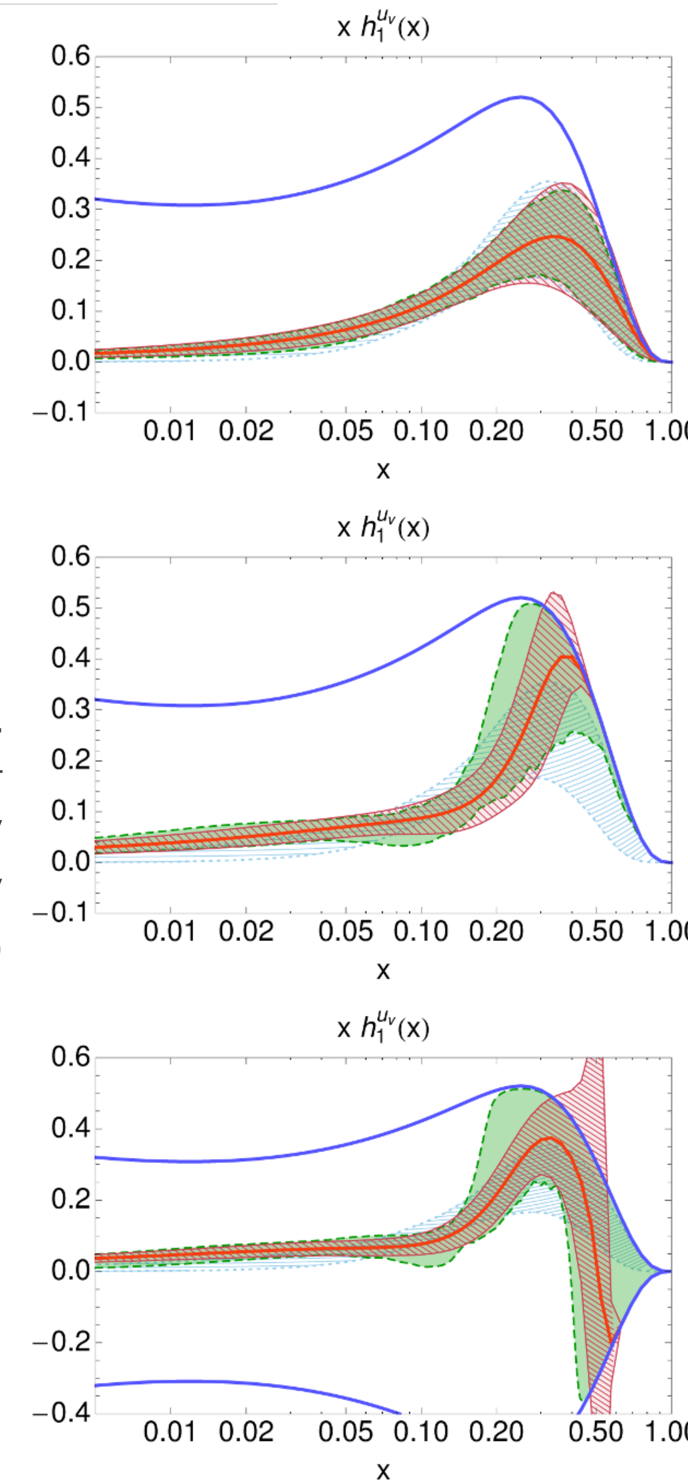
CT18 PDF (unpolarized proton PDF)

Hessian-based methodology
Inclusive of sampling bias/lack of knowledge
Tolerance criterion leads to cyan band



Pavia transversity PDF

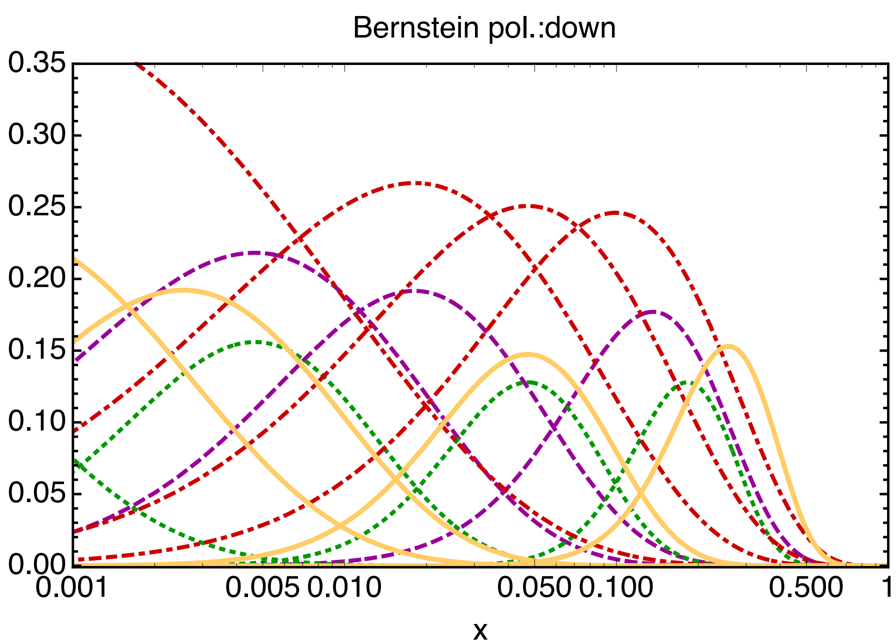
Hessian-based (with bootstrap) methodology
Variation on functional form (in early analyses)



[Bacchetta, AC & Radici, JHEP03 (2013)]

Mexico transversity PDF

Variation of Bernstein polynomials
to span the x range



[Benel, AC & Ferro, EPJC 80 (2020)]