Fantômas For QCD: exploring parametrization uncertainties for pion and other PDFs

Precision QCD predictions for ep physics at the EIC (II)

CFNS

20/09/23

Aurore Courtoy

Instituto de Física

Universidad Nacional Autónoma de México (UNAM)

CONACYT
Consejo Nacional de Ciencia y Tecnología

dgapa
Dirección General de Asuntos del Personal Académico

Instituto de Física UNAM


U. P. LINKE

Towards quantifying epistemic uncertainties in global PDF analyses

Mainly based on

"Testing momentum dependence of the nonperturbative hadron structure in a global QCD analysis" [Phys.Rev.D 103]

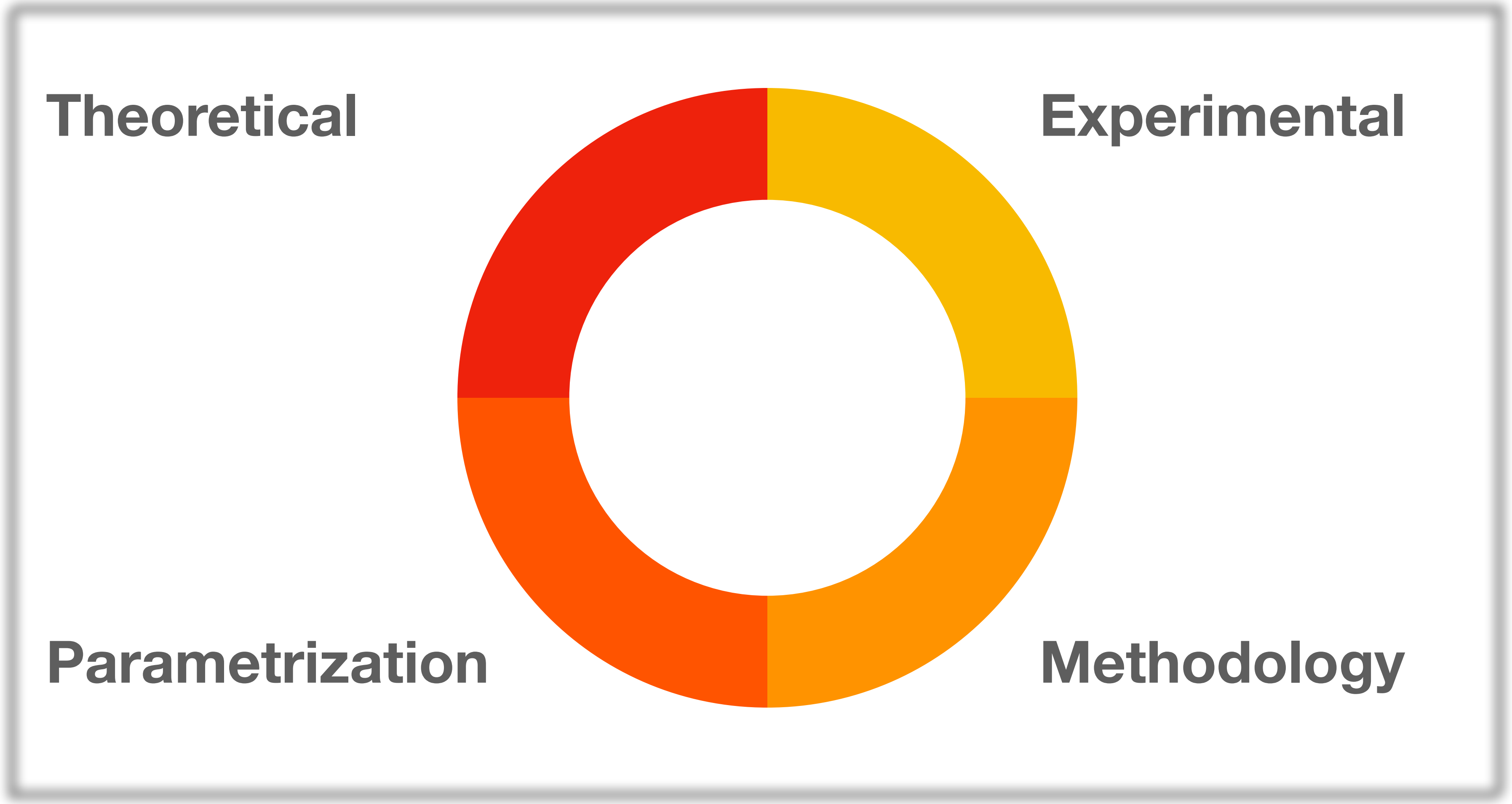
A.C. & Nadolsky

"Parton distributions need representative sampling" [Phys.Rev.D 107]

CTEQ-TEA collaboration

"An analysis of parton distributions in a pion with Bézier parametrizations " [upcoming]

L. Kotz, A. Courtoy, P. Nadolsky, F. Olness, D.M. Ponce-Chávez
DIS23 proceedings [2309.00152]



Fantômas4QCD

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

Fantômas4QCD: Our new `c++` code, Fantômas, automates series of fits using multiple functional forms.

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

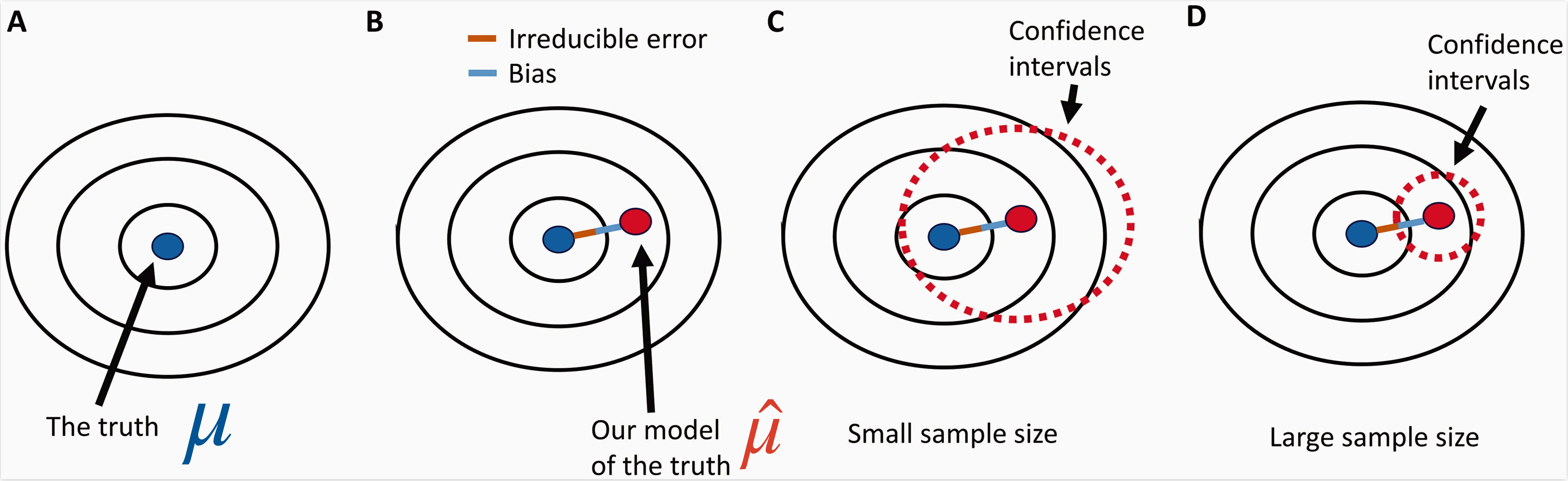
This code facilitates unbiased estimates of parametrization dependence.

CONACYT
Consejo Nacional de Ciencia y Tecnología

dgapa
Dirección General de Asuntos del Personal Académico

Inter-American Network of Networks of QCD challenges

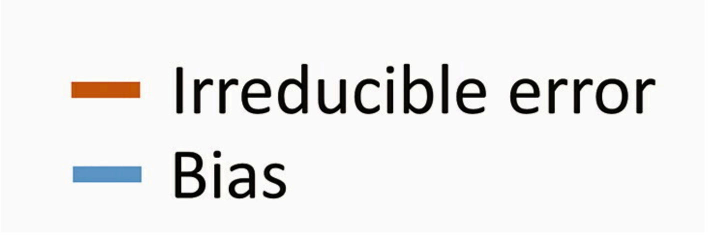
Sampling bias and big-data paradox



What uncertainties keep us from including *the truth*, μ?

Pavlos Msaouel (2022)
Cancer Investigation, 40:7, 567-576

The law of large numbers disregards the *quality of the sampling.*

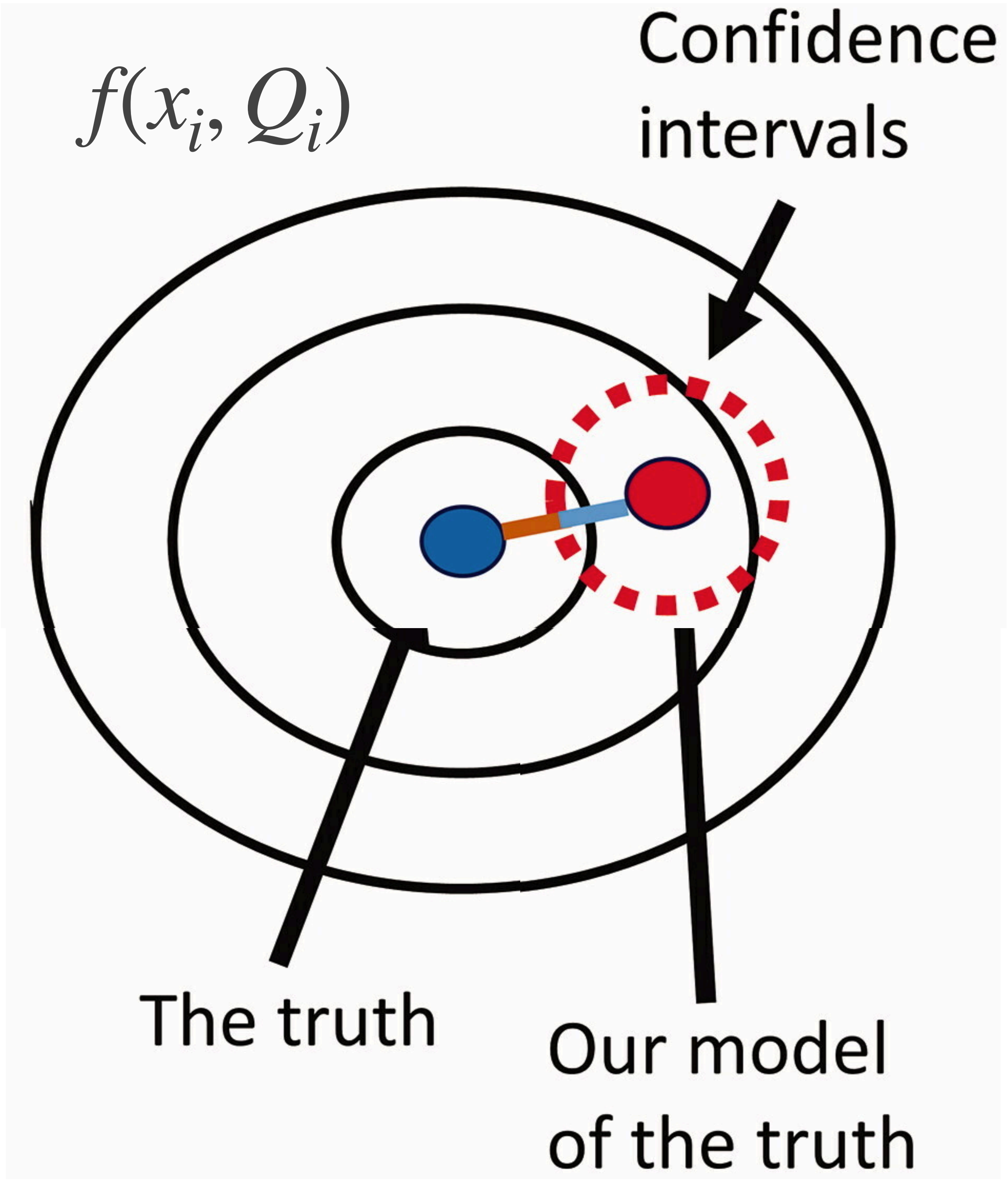


Xiao-Li Meng
The Annals of Applied Statistics
Vol. 12 (2018), p. 685

Physics phenomenology and accuracy

Is our determination from global analysis encompassing the true parton distribution function at given (x_i, Q_i)?

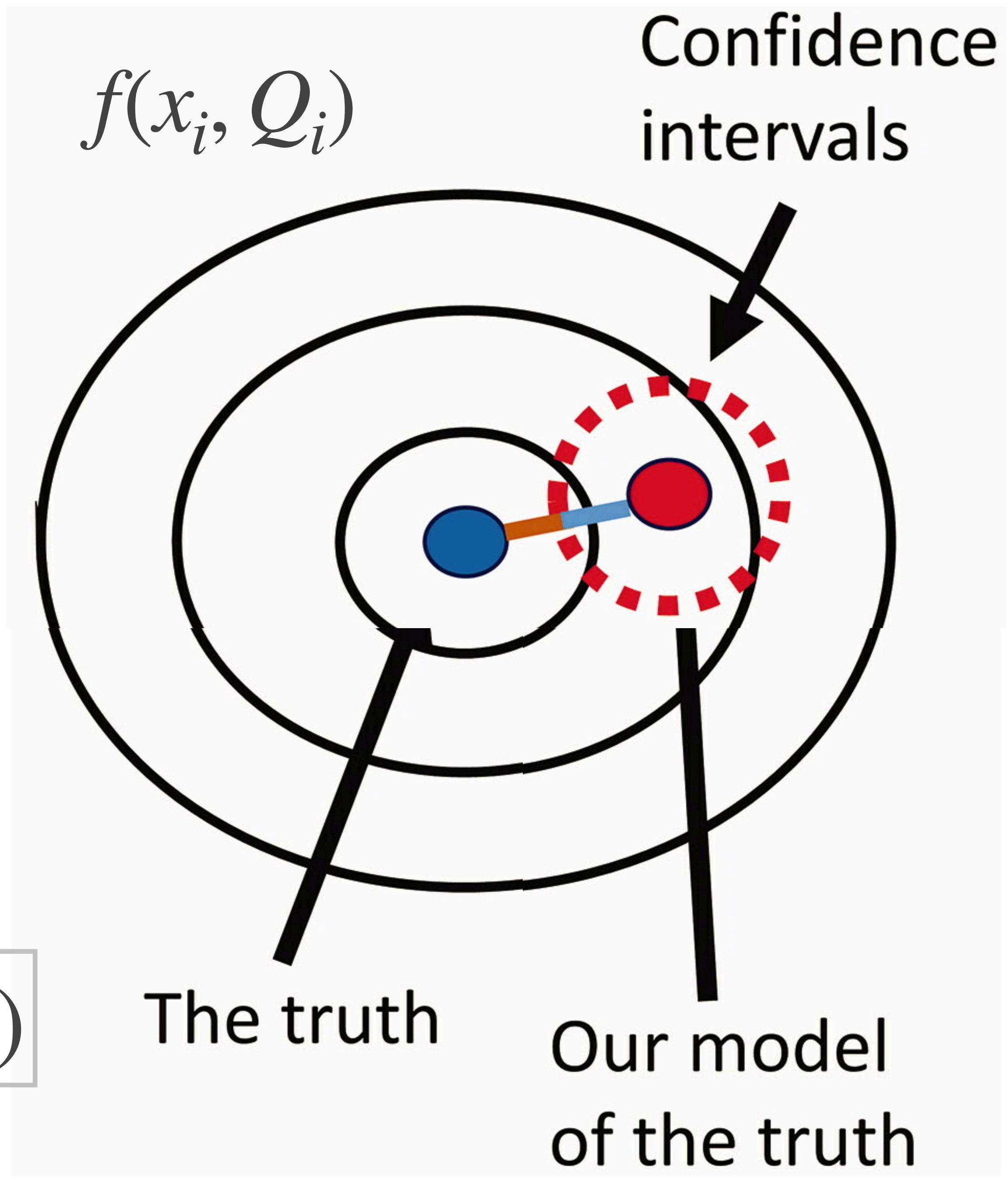
Large sample size



Physics phenomenology and accuracy

Is our determination from global analysis encompassing the true parton distribution function at given (x_i, Q_i)?

Large sample size



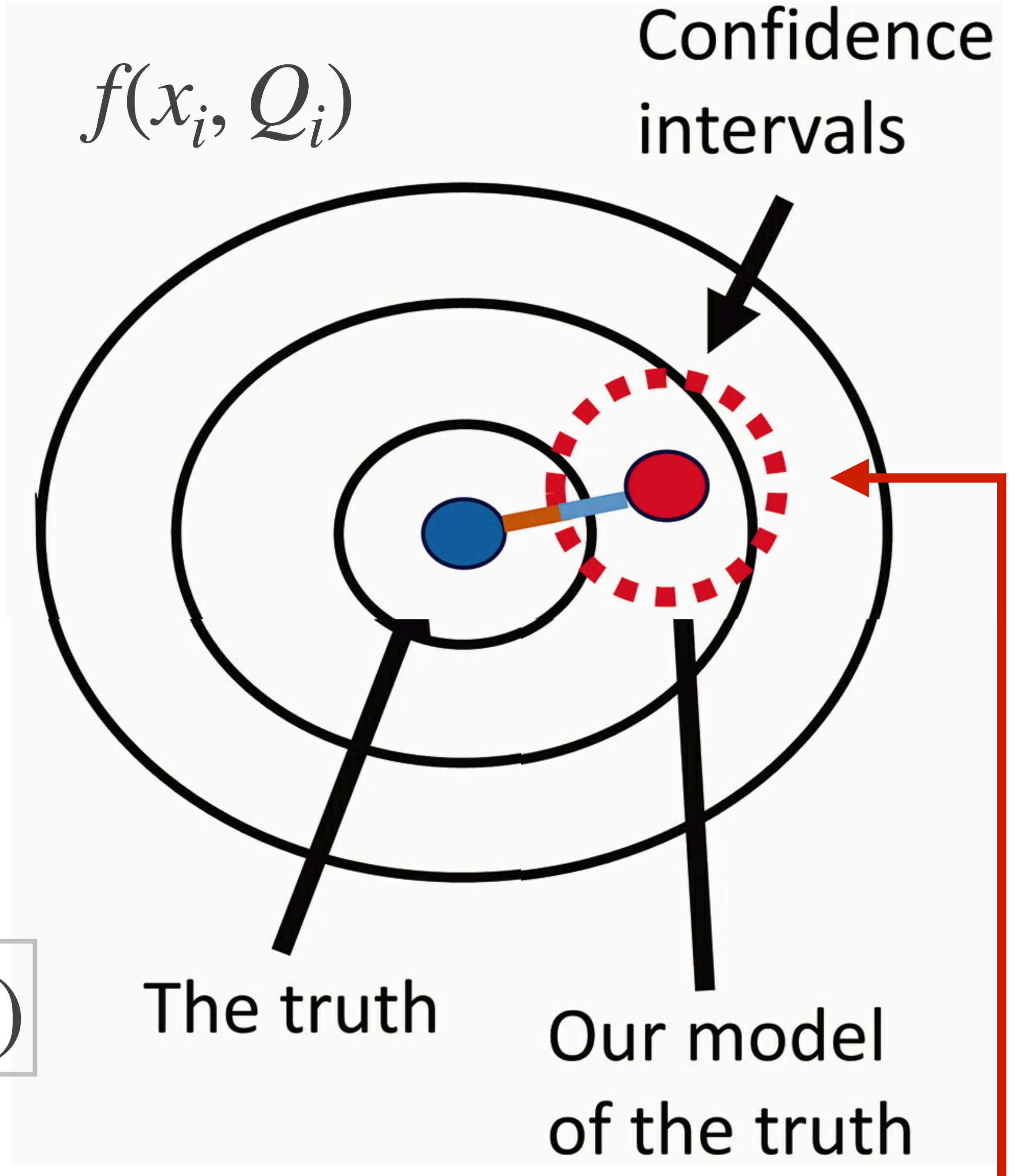
$$\mu - \hat{\mu} = \text{(data+sampling defect)} \times \text{(measure discrepancy)} \times \text{(inherent problem difficulty)}$$

depends on the sampling algorithm

- Irreducible error $\equiv$ statistical model, quality of data,…
- Bias

Physics phenomenology and accuracy

Is our determination from global analysis encompassing the true parton distribution function at given (x_i, Q_i)?



Large sample size

$$\mu - \hat{\mu} = \text{(data+sampling defect)} \times \text{(measure discrepancy)} \times \text{(inherent problem difficulty)}$$

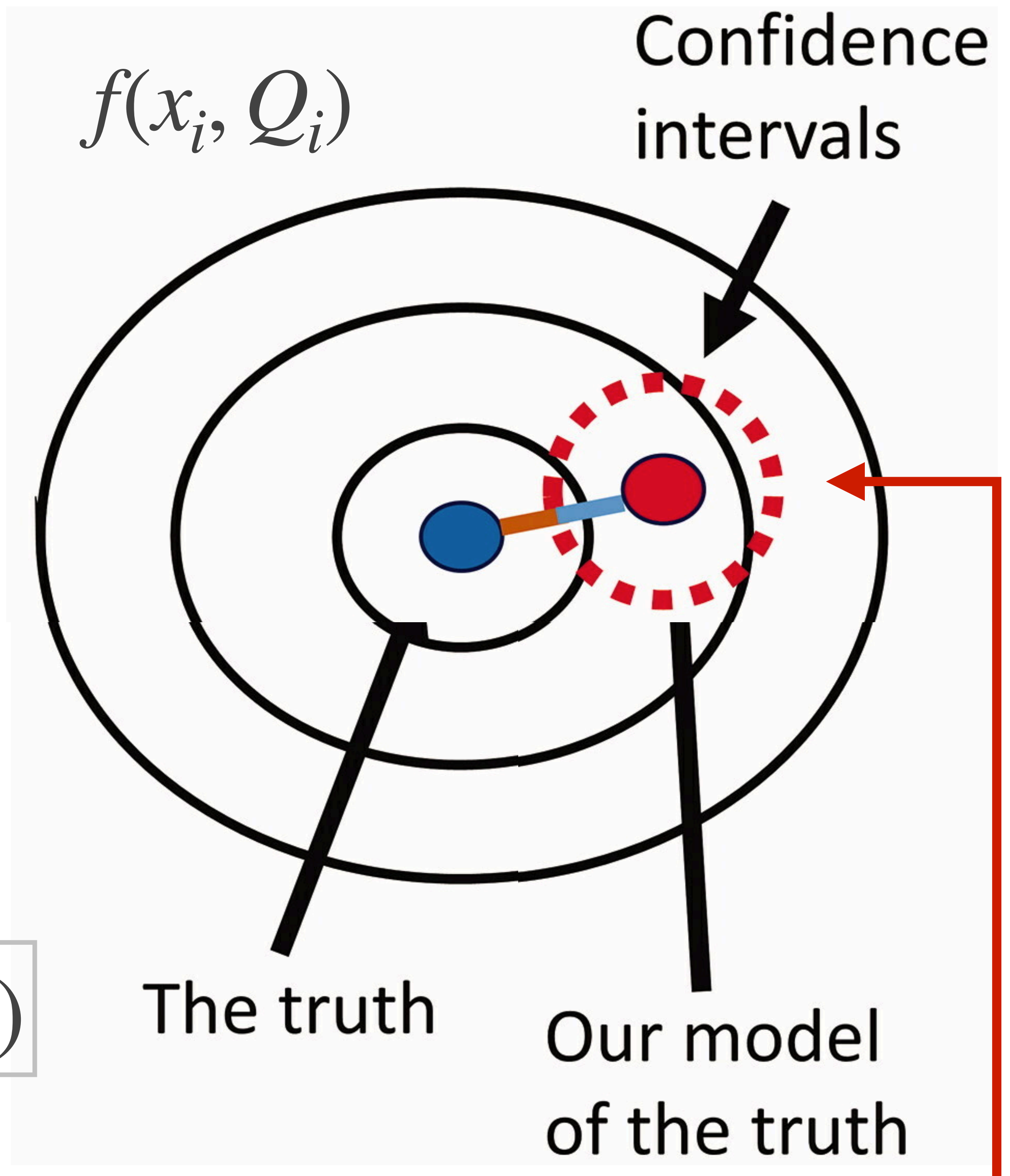
depends on the sampling algorithm

can tend to $\left(\sqrt{n}\right)^{-1}$ for random sampling

- Irreducible error $\equiv$ statistical model, quality of data,…
- Bias

Physics phenomenology and accuracy

Is our determination from global analysis encompassing the true parton distribution function at given (x_i, Q_i)?



Large sample size

$$\mu - \hat{\mu} = \text{(data+sampling defect)} \times \text{(measure discrepancy)} \times \text{(inherent problem difficulty)}$$

depends on the sampling algorithm

can tend to $\left(\sqrt{n}\right)^{-1}$ for random sampling

- Irreducible error $\equiv$ statistical model, quality of data,…
- Bias

To streamline the sampling over parametrization forms, we have designed `metamorph`.

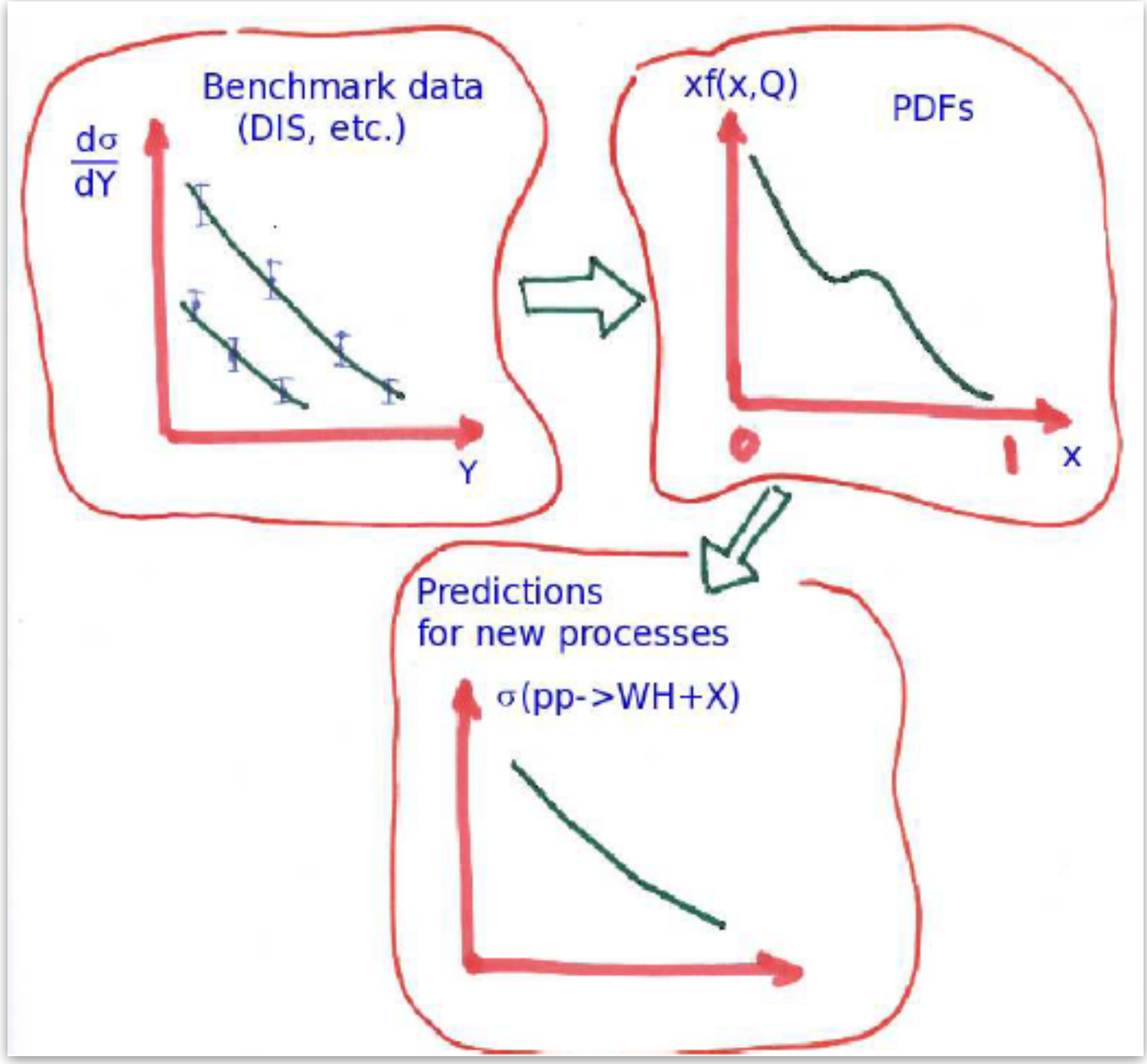
The shape of parton distributions

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

Shape in x extracted from data that are sensitive to specific PDF flavors, etc.

I. hints of behavior of partons at low scales

II. predictions for other (new) processes



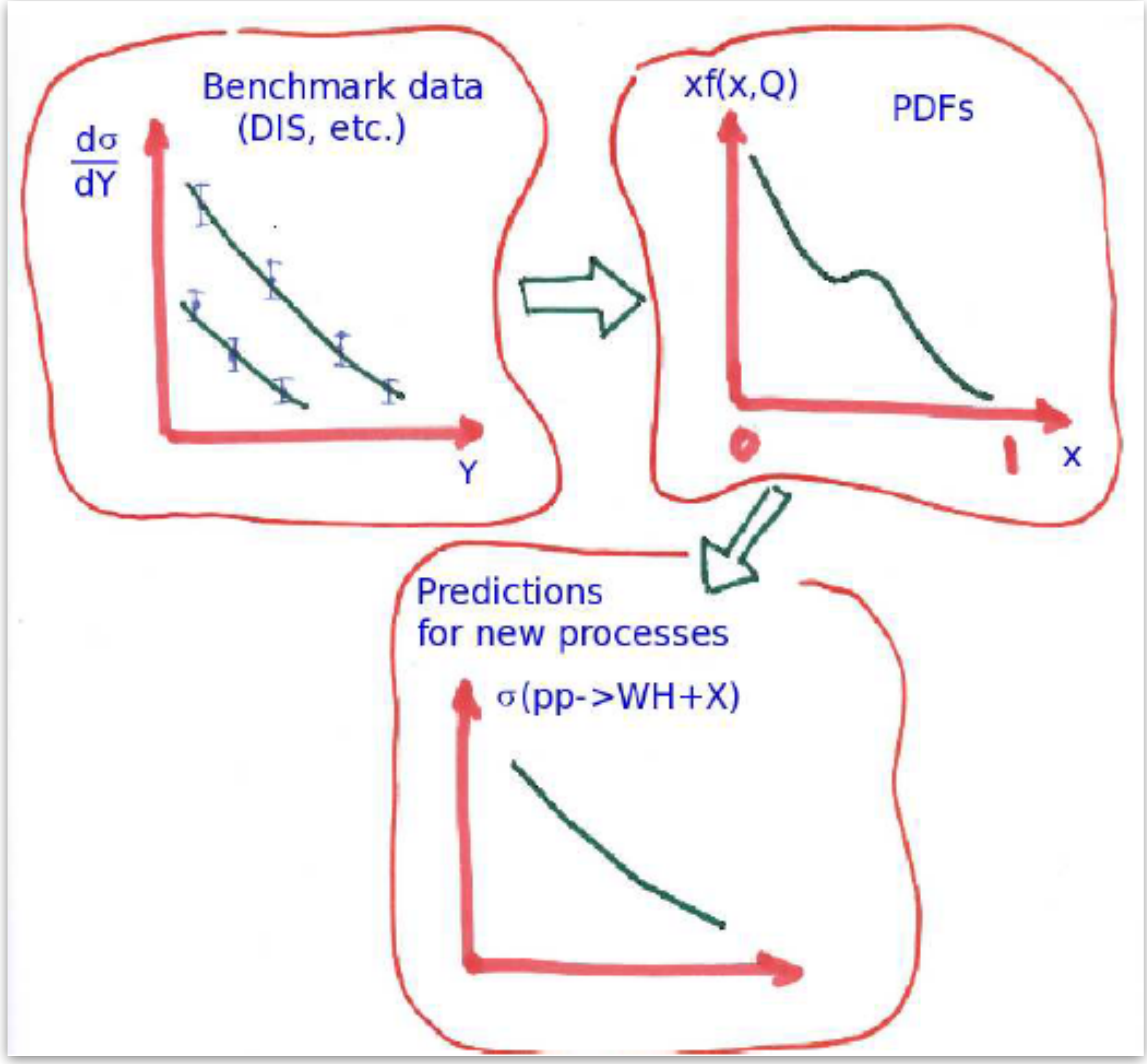
The shape of parton distributions

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

Shape in x extracted from data that are sensitive to specific PDF flavors, etc.

I. hints of behavior of partons at low scales

II. predictions for other (new) processes



Classes of *first principle* constraints for x-dependence

- positivity of cross sections
- support in $x \in [0,1]$
- end-point: $f(x=1)=0$
- sum rules: $< x >_n = \int_0^1 dx\, x^{n-1} f(x)$

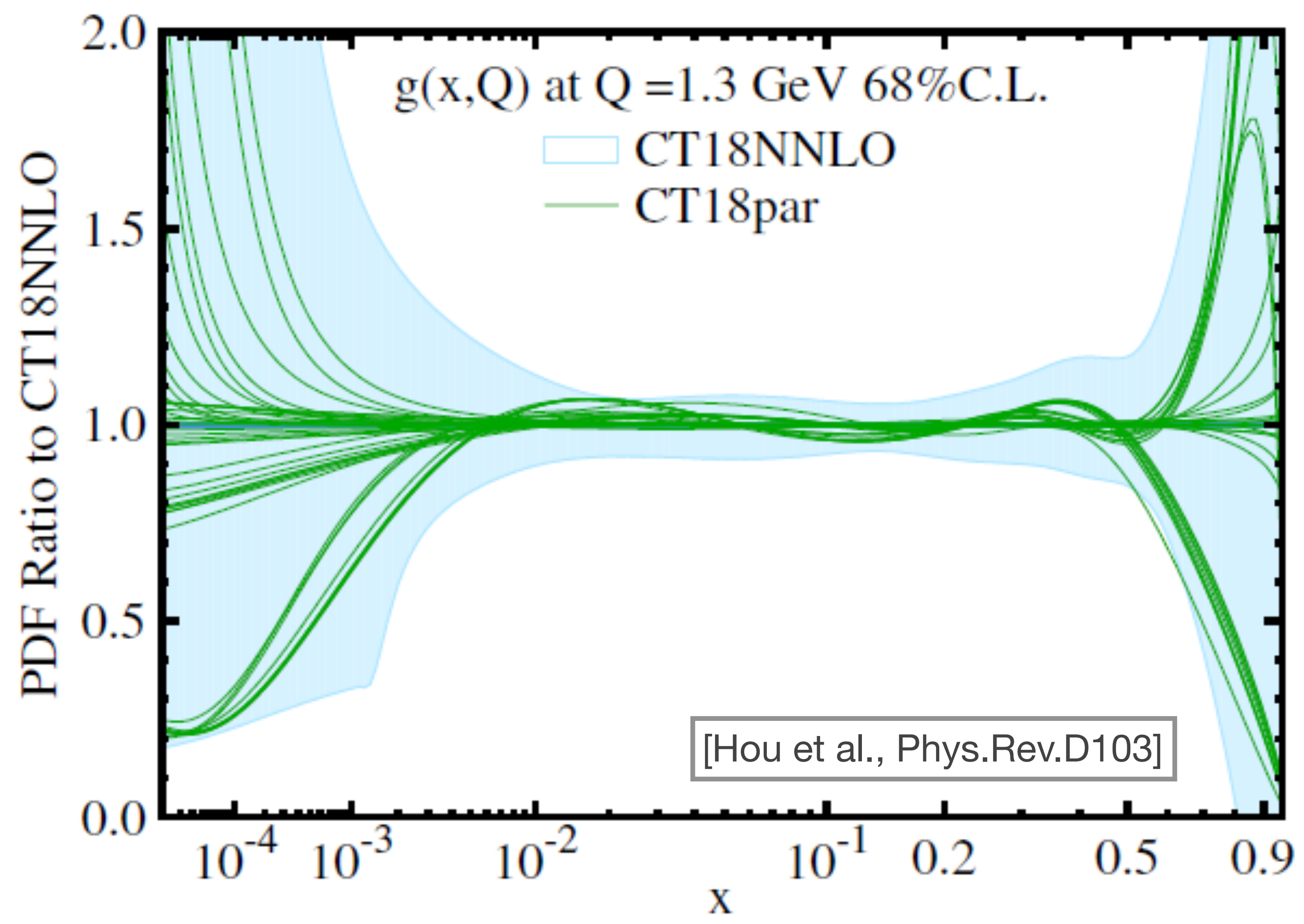
⇨ asymptotics usually ensured by a *carrier function*

⇨ sum rules imposed through normalization

Rôle of parametrization in previous analyses

CT18 PDF (unpolarized proton PDF)

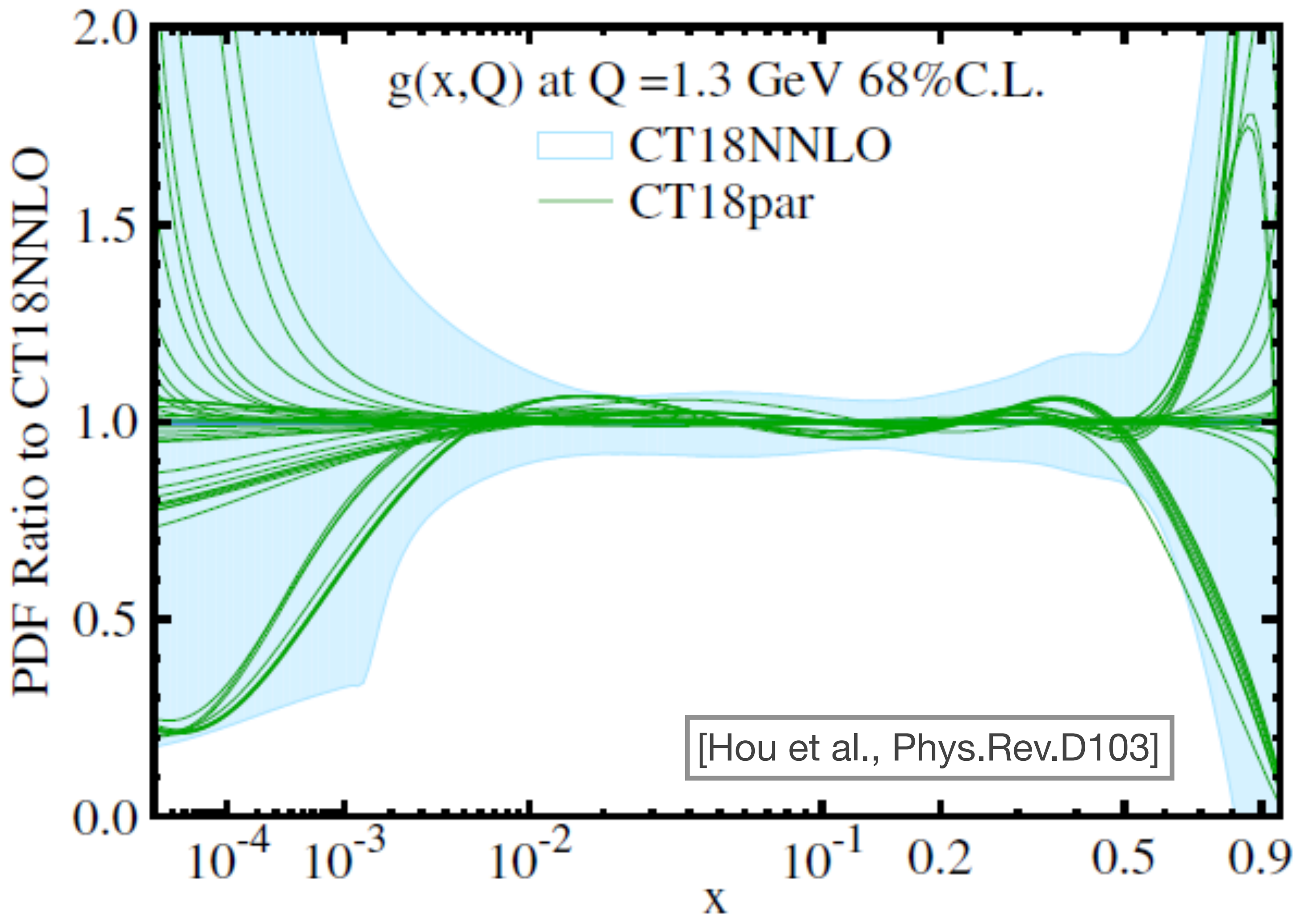
Hessian-based methodology
Inclusive of sampling bias/lack of knowledge
Tolerance criterion leads to cyan band



Rôle of parametrization in previous analyses

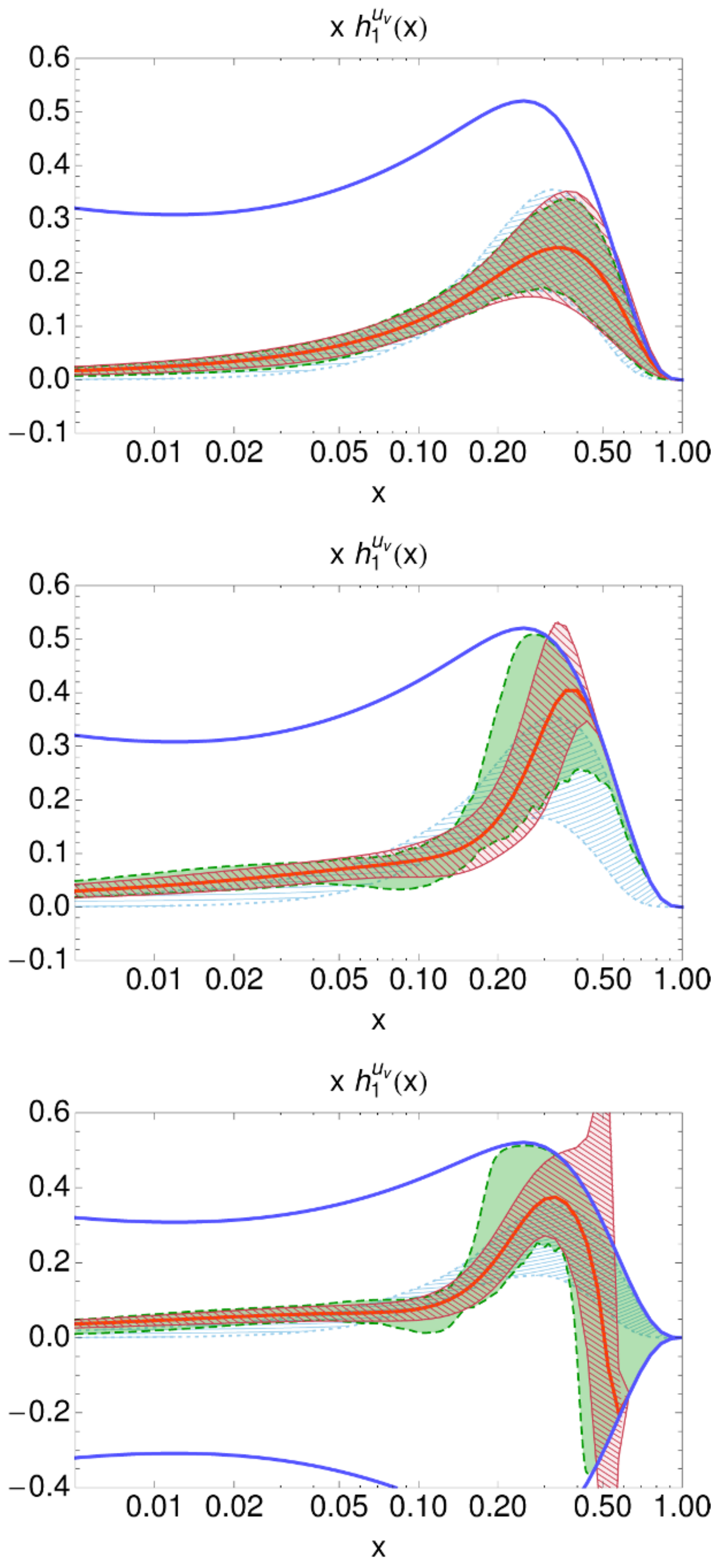
CT18 PDF (unpolarized proton PDF)

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).

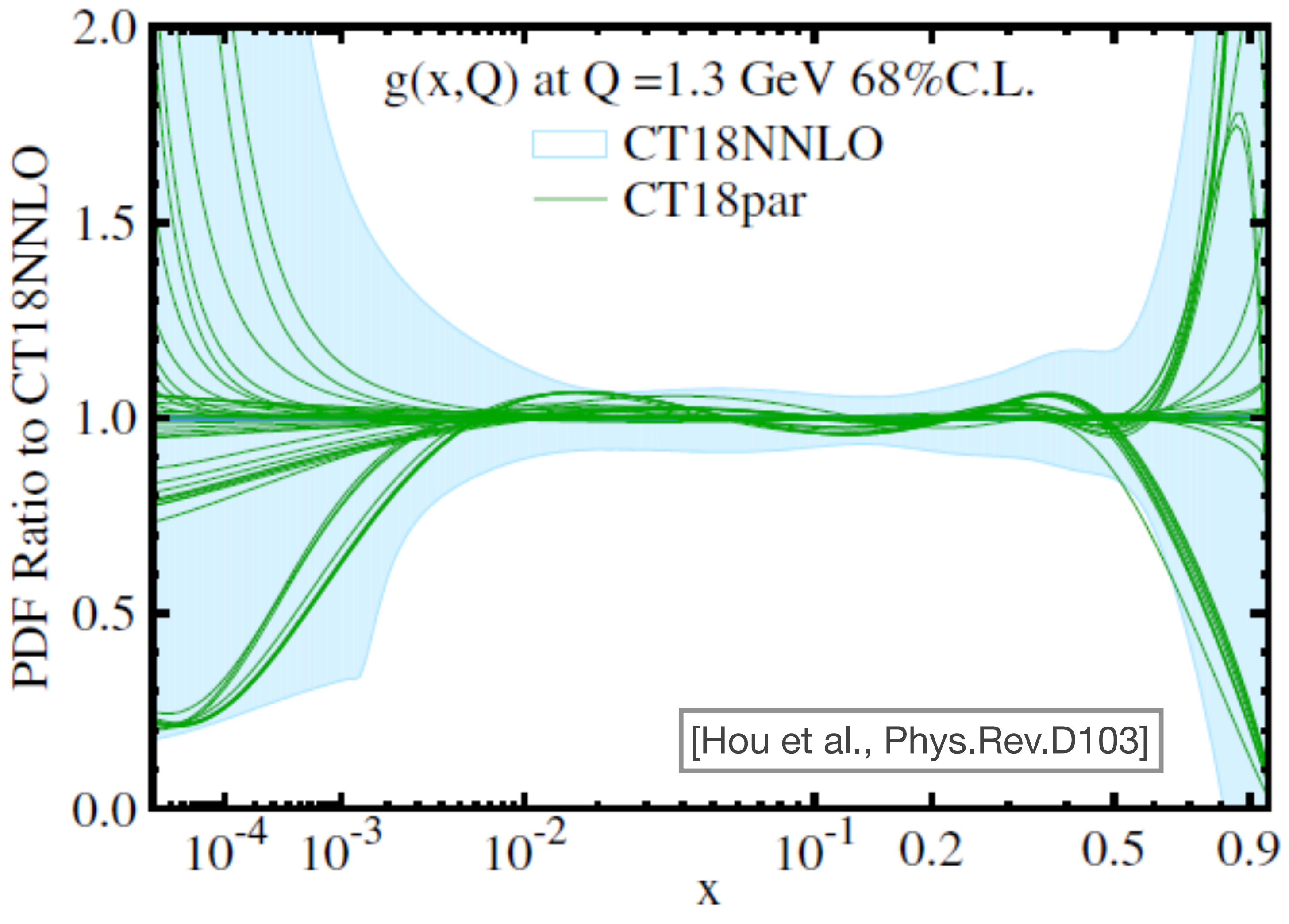


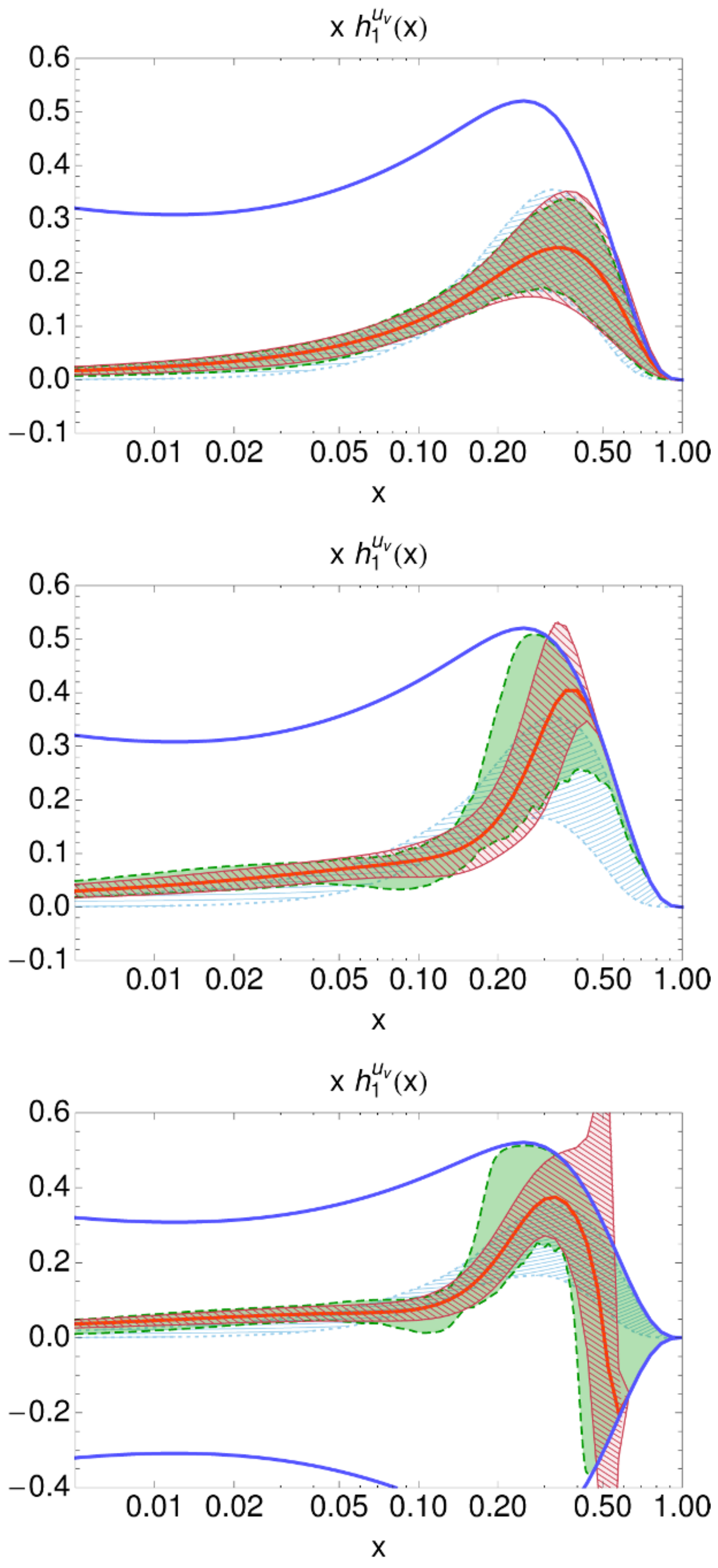
[Bacchetta, AC & Radici, JHEP03 (2013)]

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

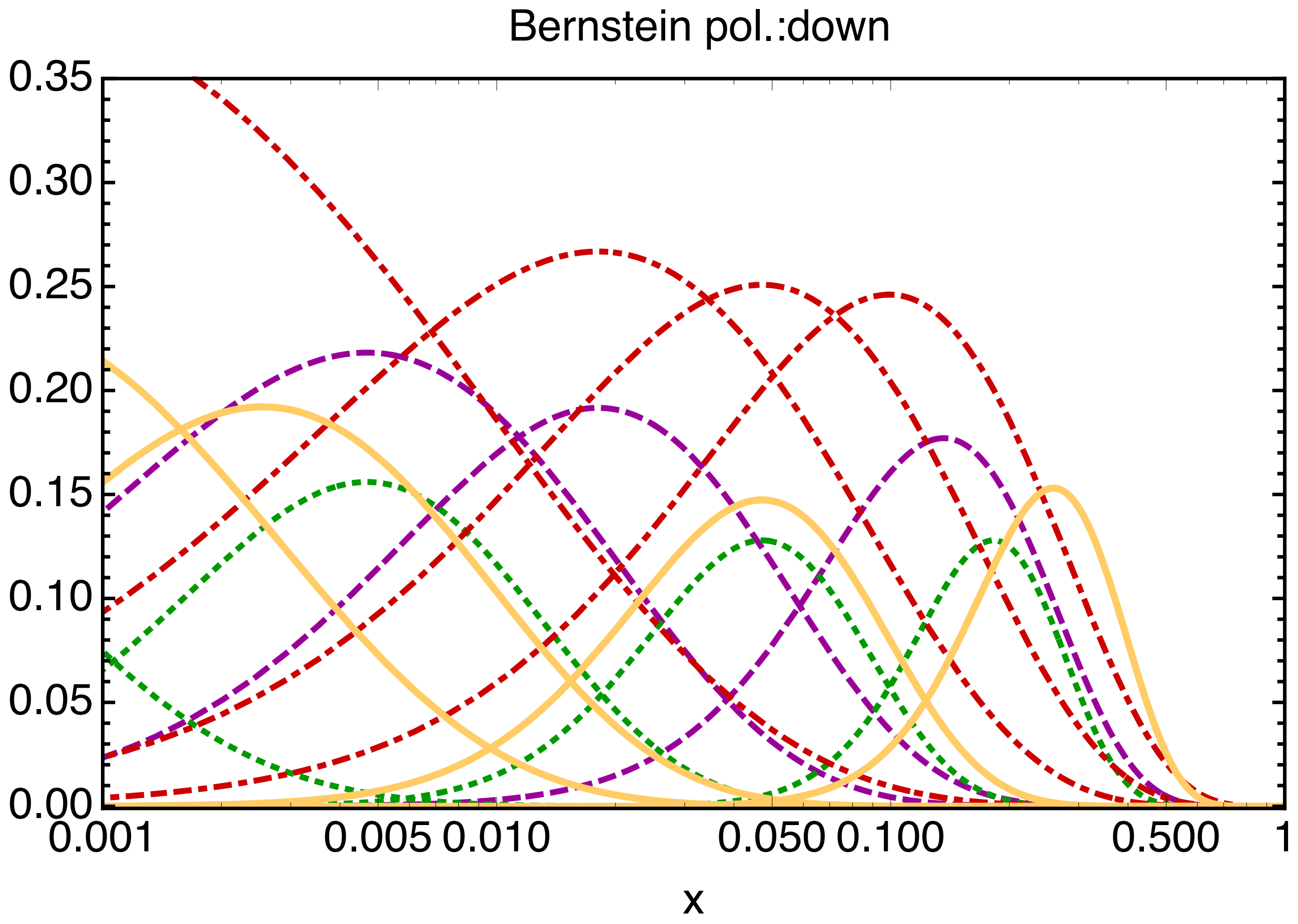




Pavia transversity PDF

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

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



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

Mexico transversity PDF

Variation of Bernstein polynomials to span the x range.

The shape of parton distributions

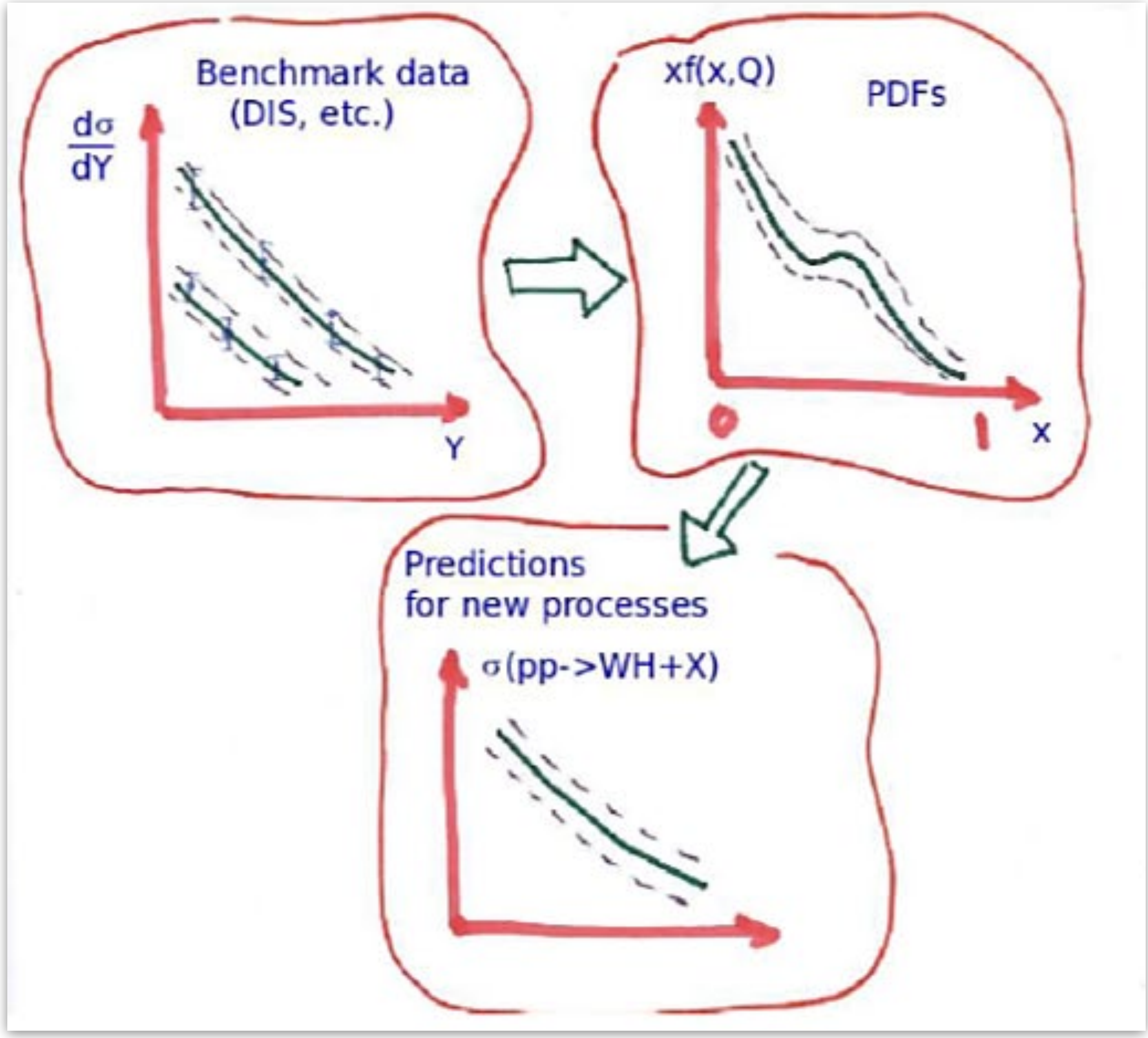
Low-energy QCD dynamics, encapsulated in PDFs, are learned from experimental data.

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



The shape of parton distributions

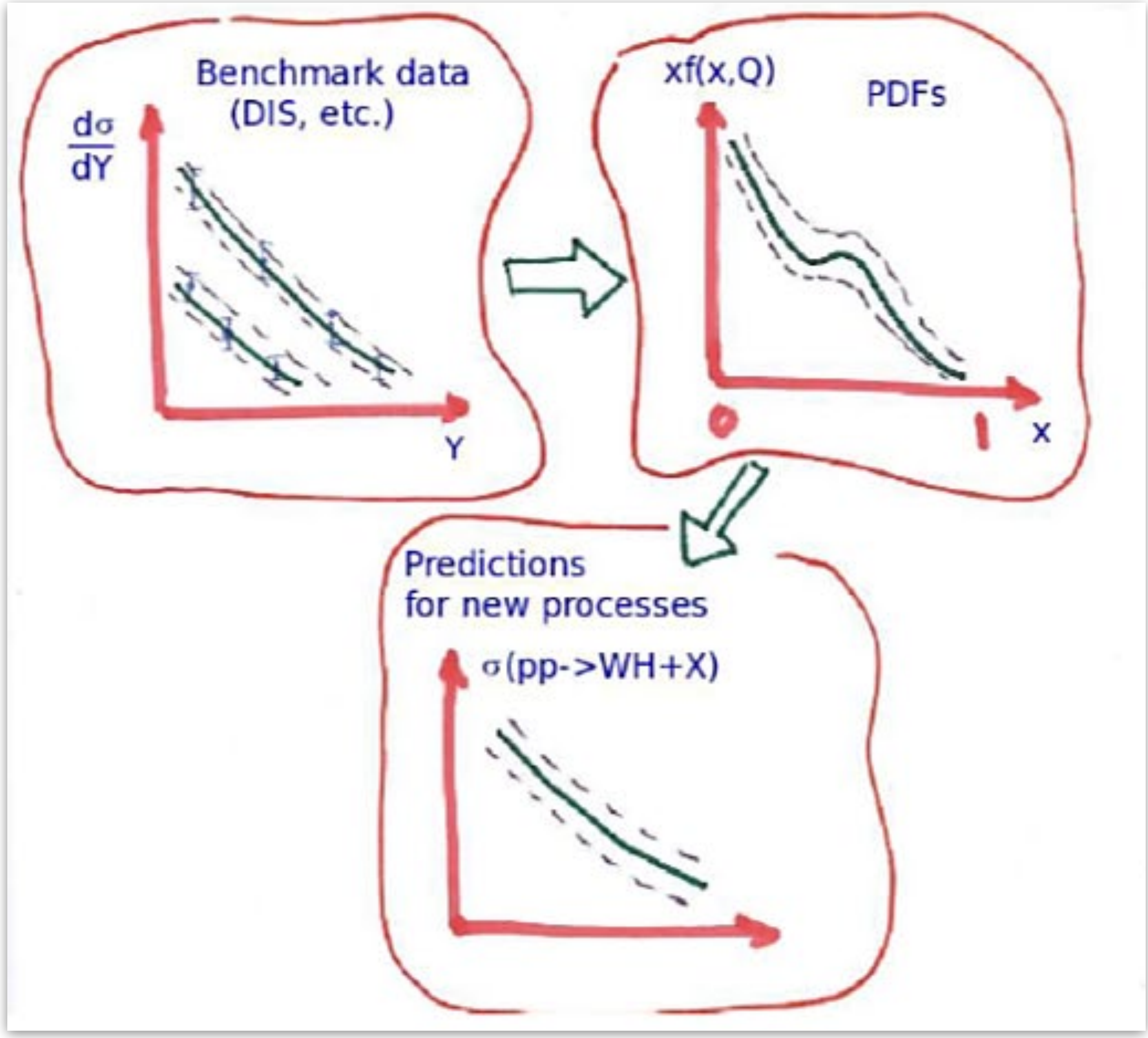
Low-energy QCD dynamics, encapsulated in PDFs, are learned from experimental data.

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

Hypothesis testing and parton distributions

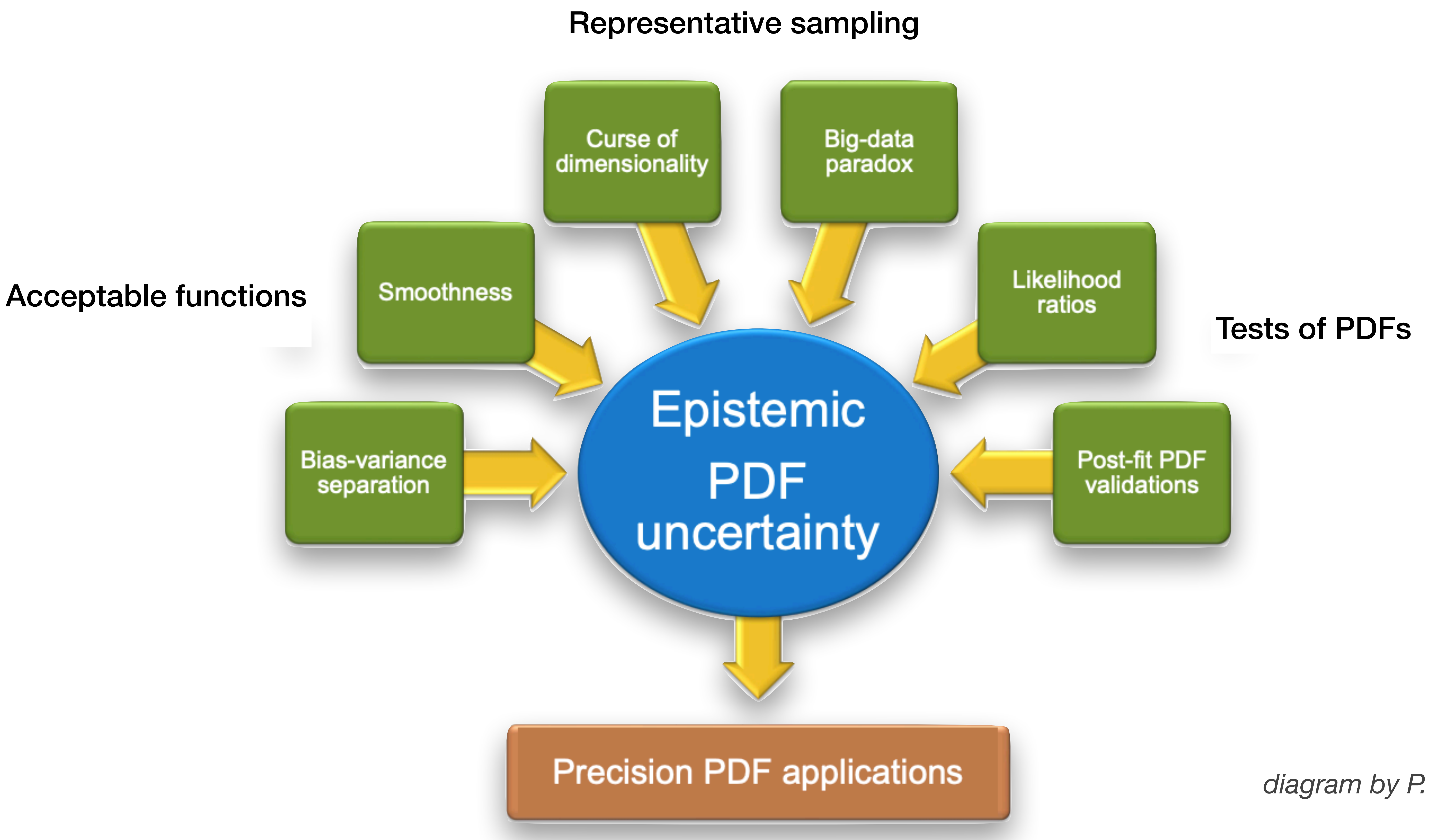


diagram by P. Nadolsky [DIS2023]

Epistemic uncertainties

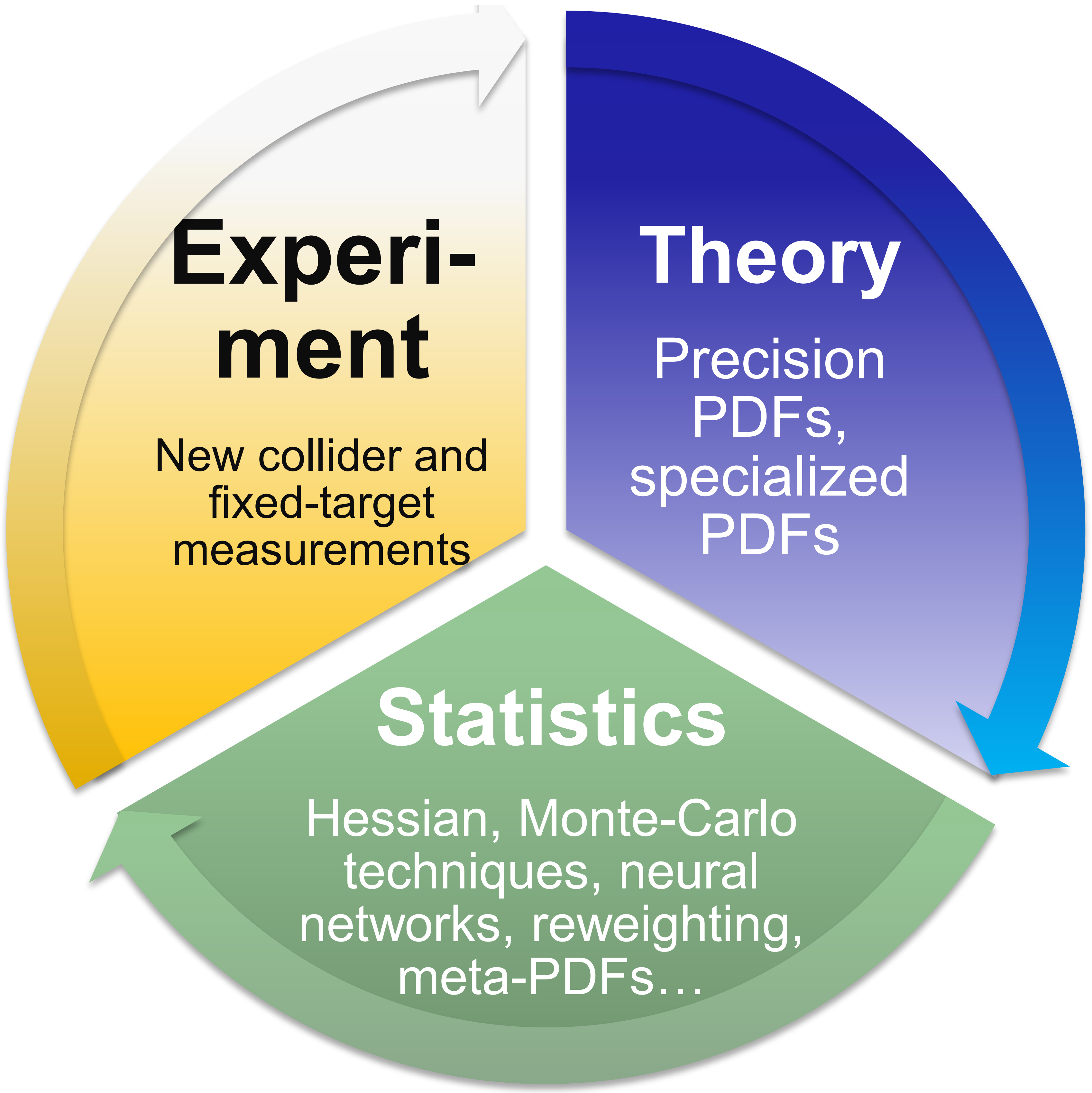
How do we estimate the epistemic uncertainty of our analysis?

Global analyses in both Hessian and MC/ML approaches estimate experimental, theoretical, and epistemic uncertainties

The latter is due to methodological choices that

- can be estimated by sampling over analysis workflows, parametrization forms, analysis settings
- are associated with the prior probability

While challenging in general, such estimation is facilitated by several representative sampling techniques.



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

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

Bézier curve

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

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

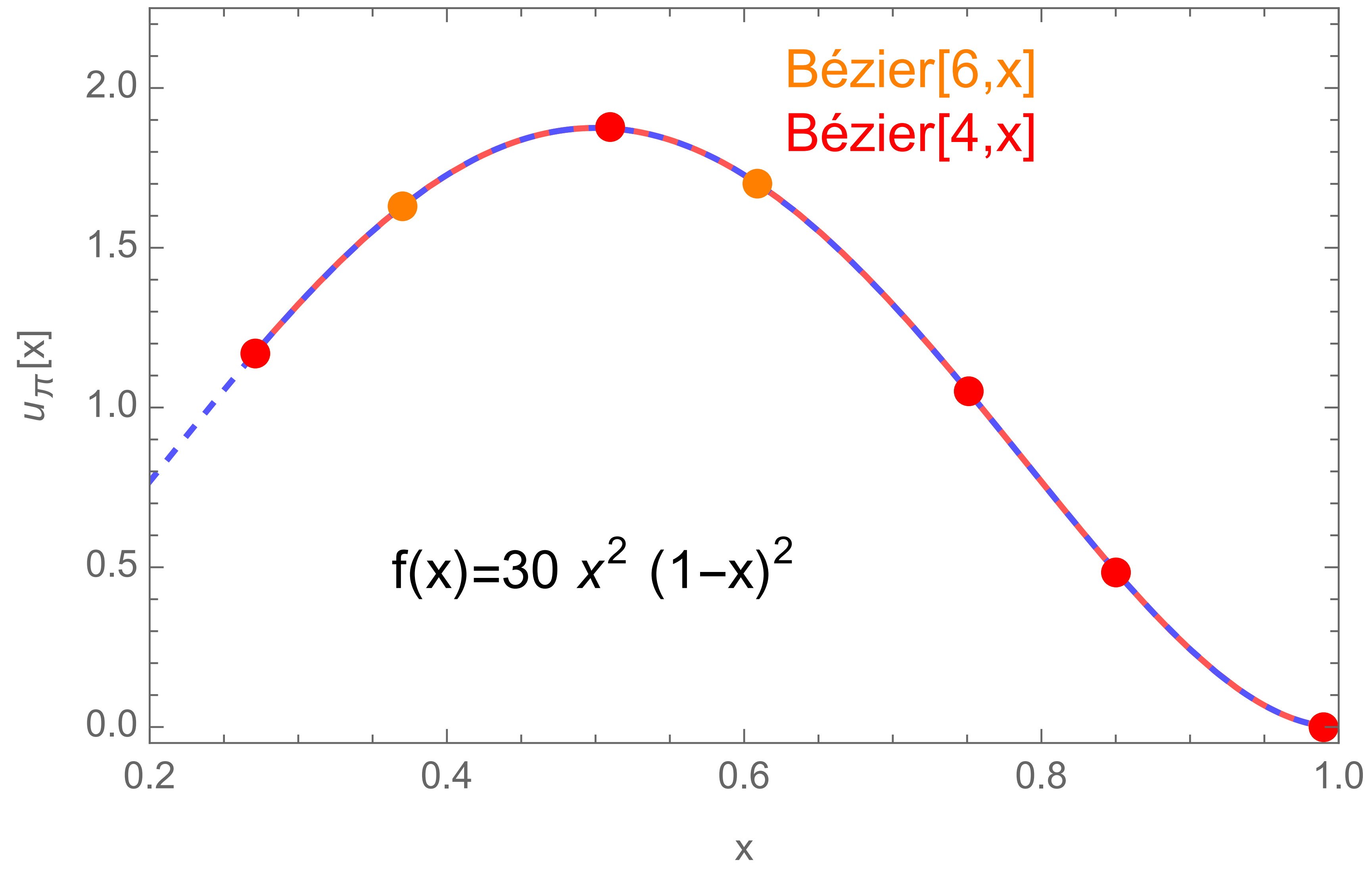
$$\underline{P} = \underline{\underline{T}} \cdot \underline{\underline{M}} \cdot \underline{C}$$

Such that the coefficients can be expressed in terms of known matrices

$$\underline{C} = \underline{\underline{M}}^{-1} \cdot \underline{\underline{T}}^{-1} \cdot \underline{P}$$

The orange/red points represent the control points, the number of which is related to the degree of the polynomial.

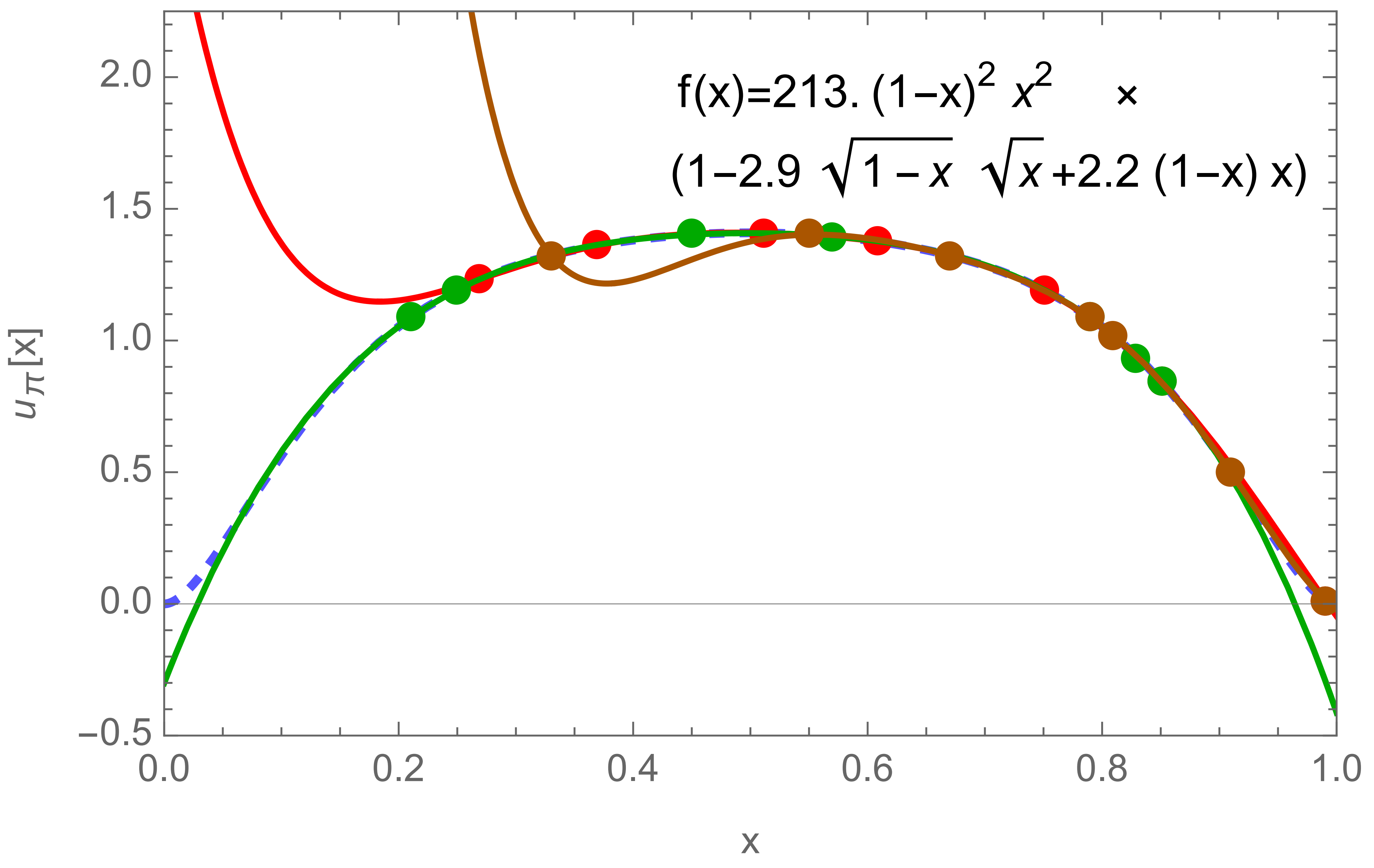
For simple functions, the interpolation is unique for any set of control points.



Bézier-curve methodology for global analyses

Reconstruction of a more complex parametrization

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

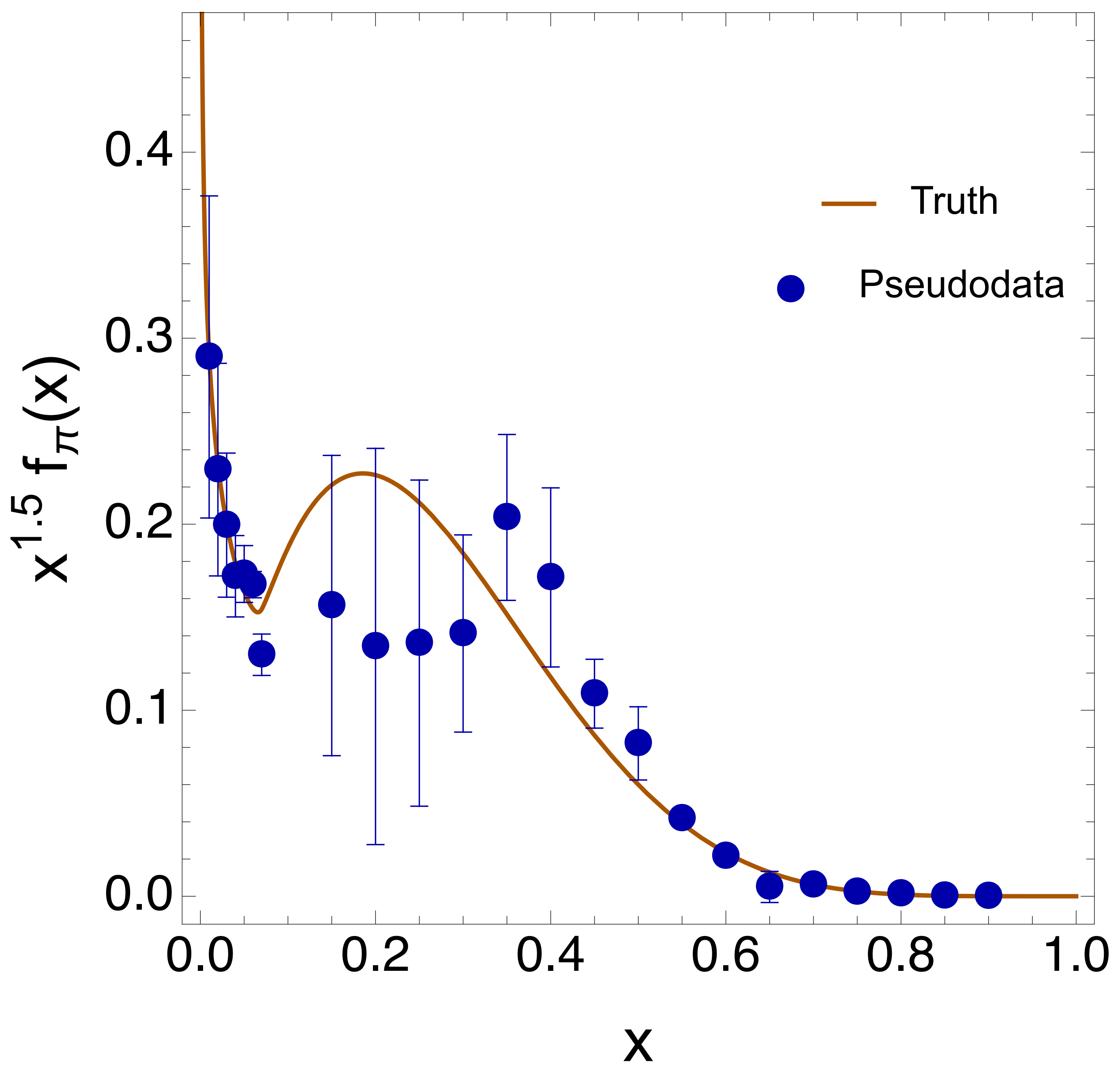


Global analyses can exploit this property to generate many functional forms.

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



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

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

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

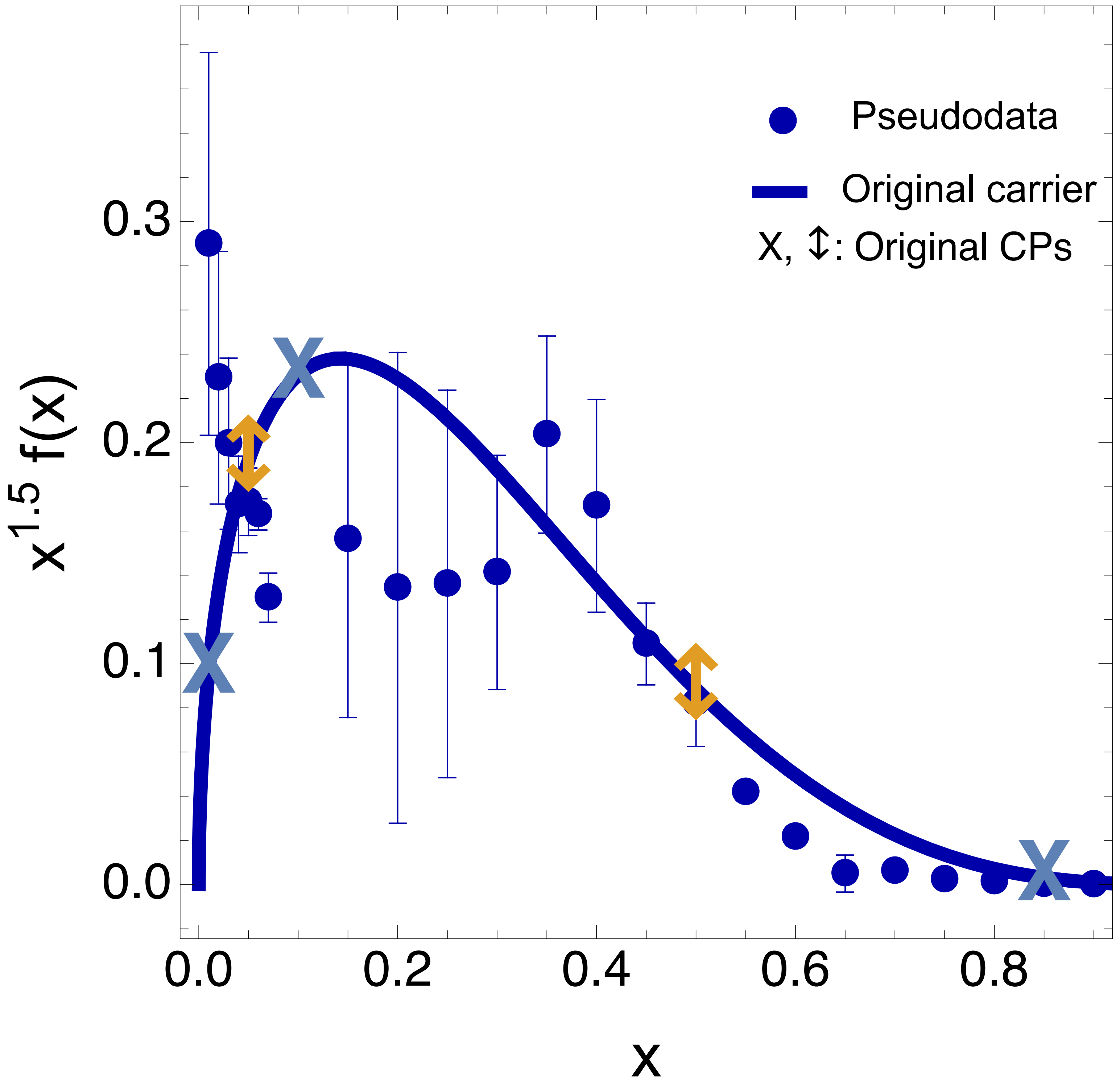
Classical fit: determines the vector $\underline{C}$

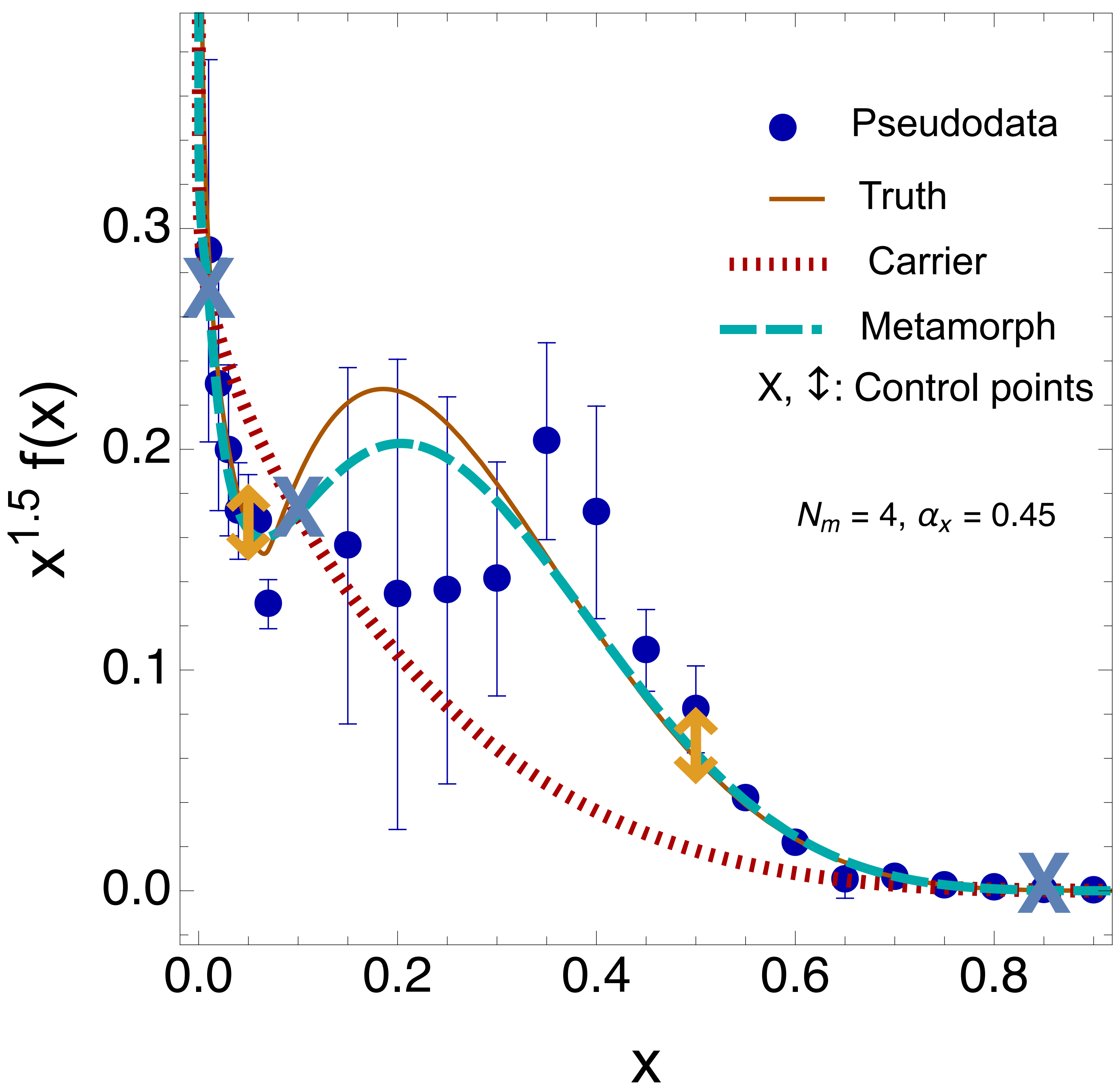
`metamorph` fit: determines the vector $\underline{P}$

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

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

Bézier-curve methodology for global analyses — toy model





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

N_m= degree of polynomial can vary

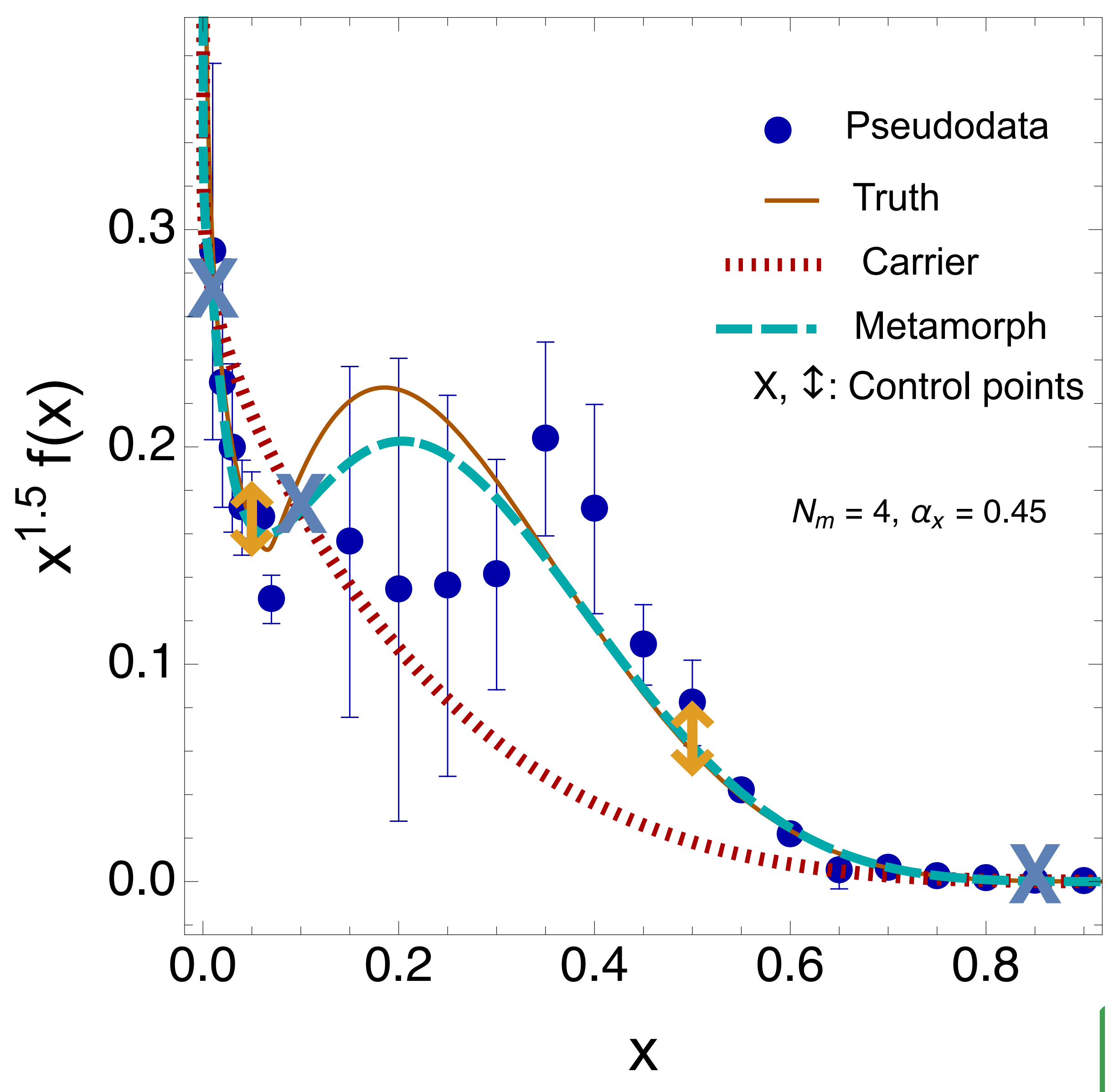
$\delta B_q \,\&\, \delta C_q$ allow the carrier to vary

α_x can vary

`metamorph` fit:

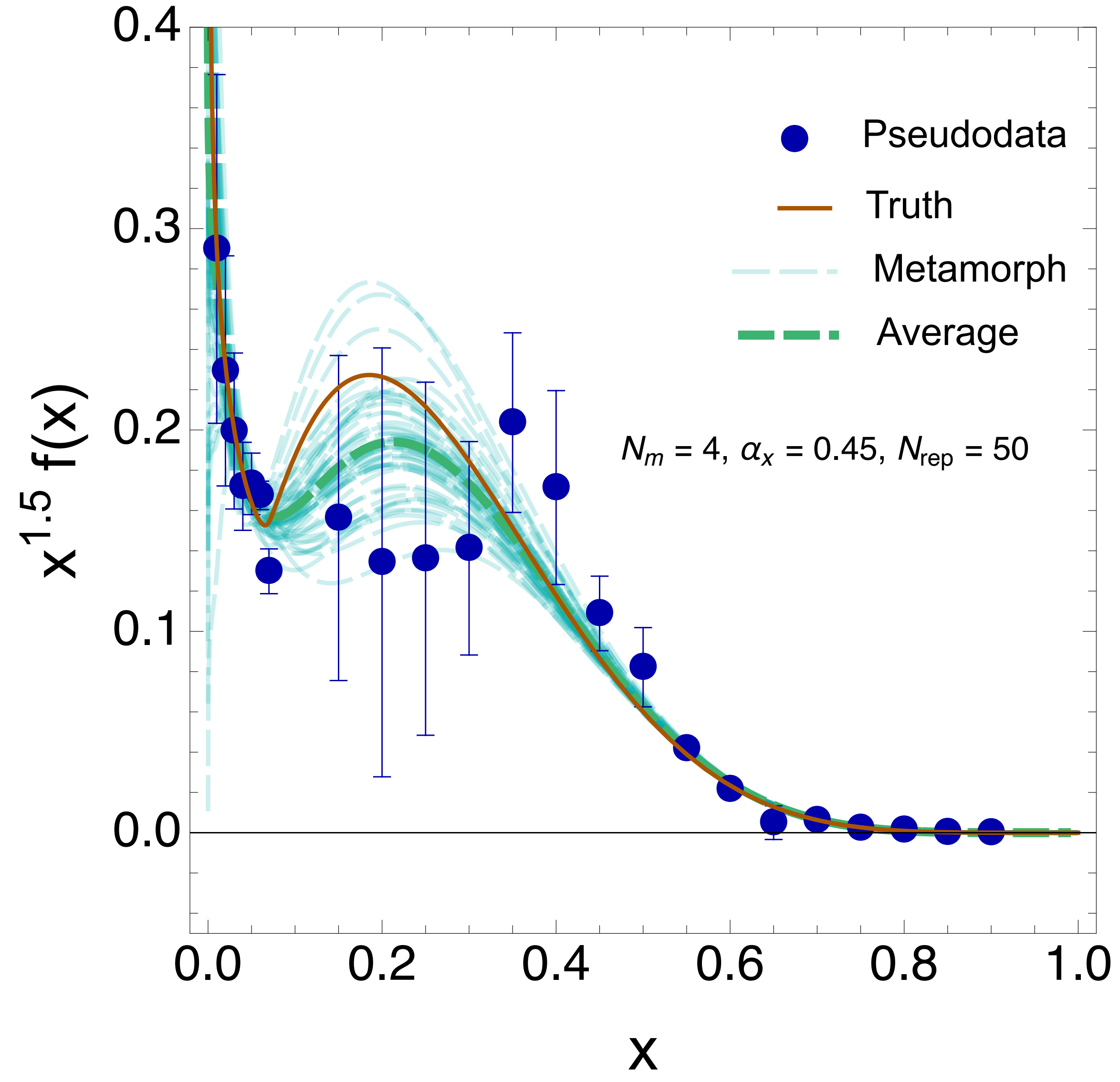
$$x\, q(x, Q_0^2) = A'_q\, x^{B_q} (1-x)^{C_q} \times \left(1 + \mathcal{B}^{(N_m)}(x^{\alpha_x}, Q_0^2; \underline{v})\right)$$

Bézier-curve methodology for global analyses — toy model



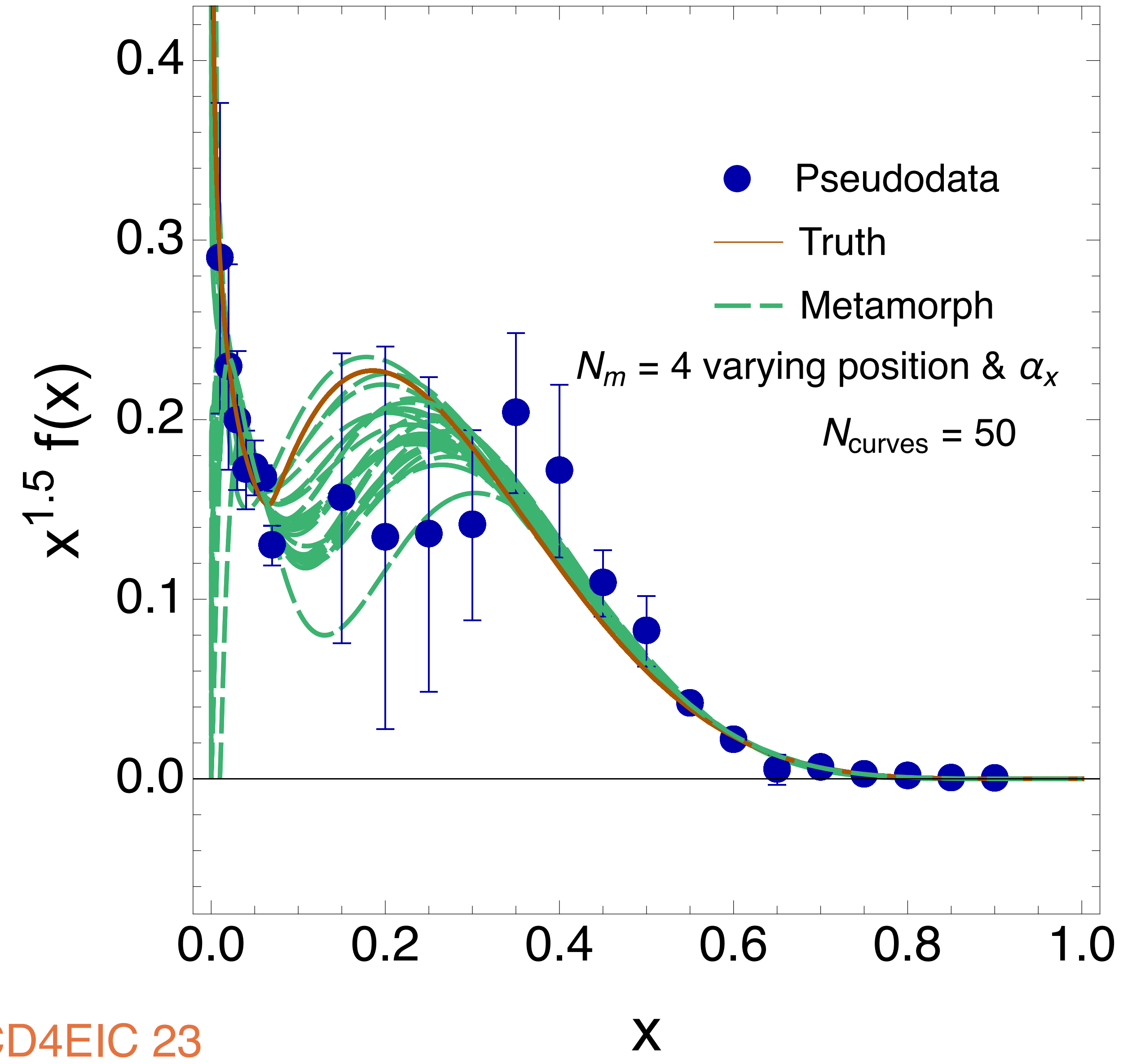
if bootstrapped

sampling on the distribution of data uncertainties



if sampled over functional forms

sampling over parametrizations



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

`metamorph` routine in xFitter

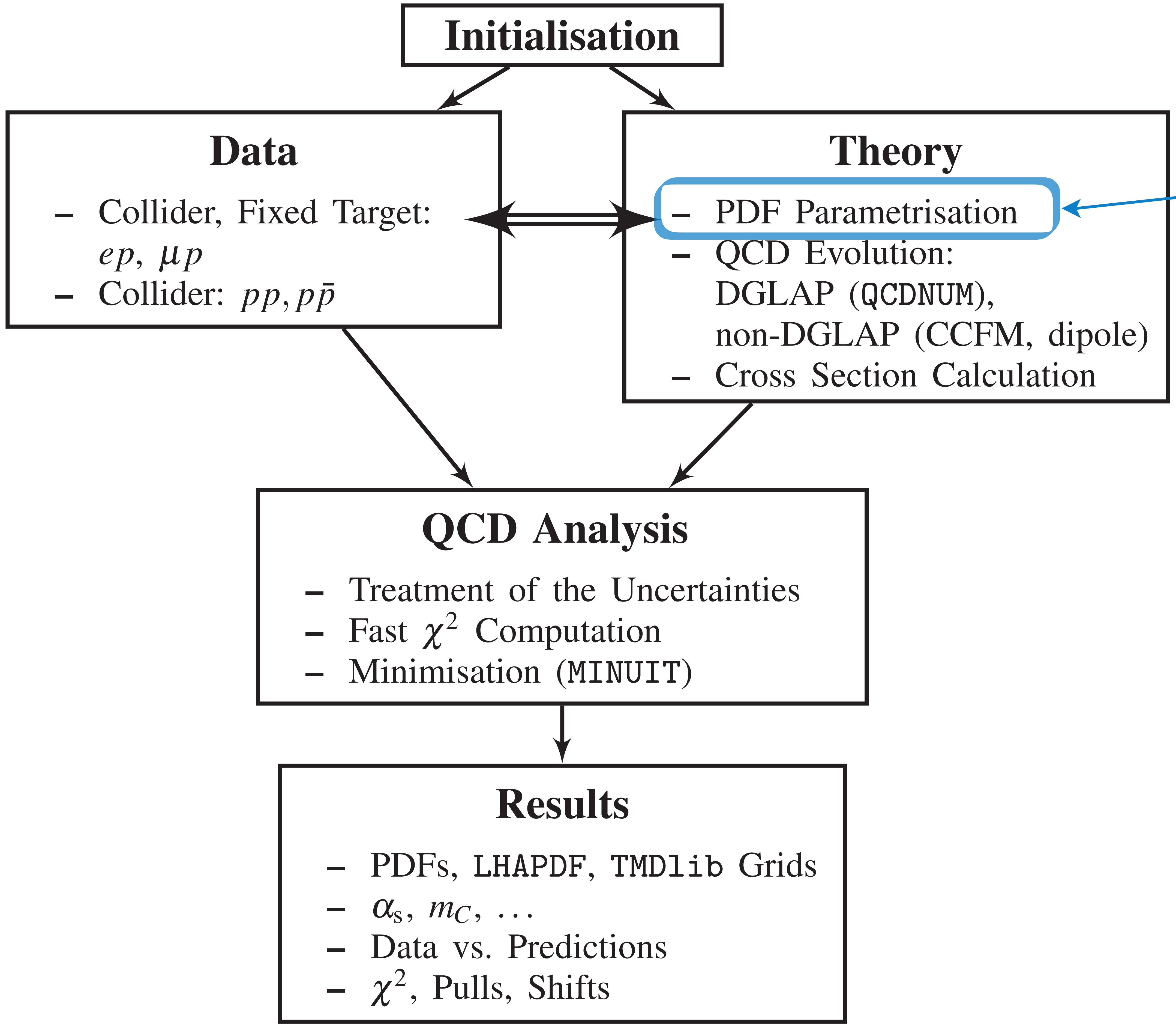


Figure 1: Schematic structure of the `xFitter` program.

`metamorph` requires inputs from the user:

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

Why study the pion?

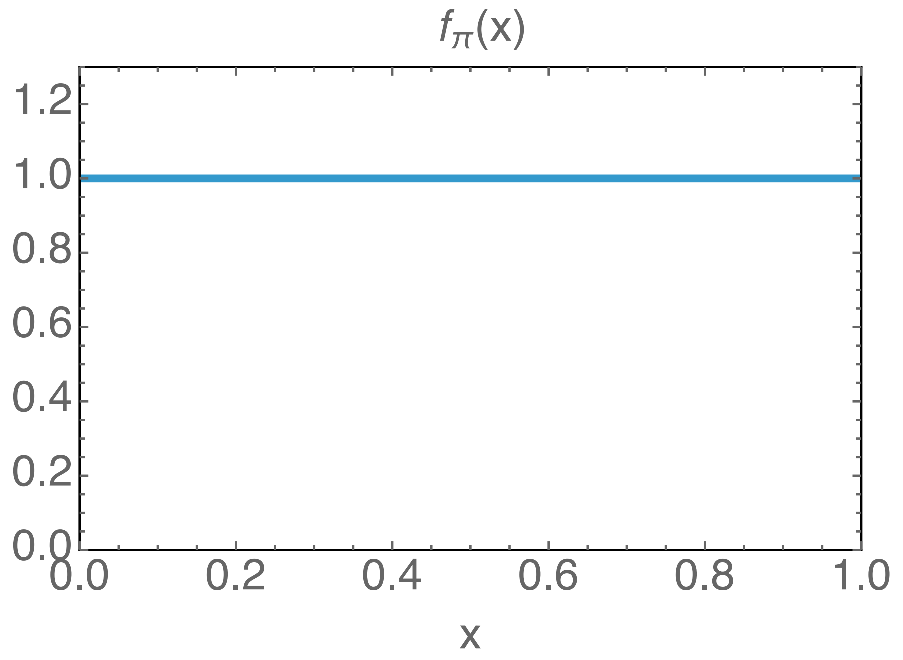
- ➡ xFitter's framework set up the pion PDF analysis— https://www.xfitter.org/xFitter/
- ➡ less data *wrt* proton, still at NLO accuracy
- ➡ recent "come back" thanks to increased fitting activity in the nuclear community —theory and experiment-wise

⇨ Pion PDFs are closely related to the dynamics of QCD in non-perturbative regime— trickier interpretation due to its pseudo-Goldstone nature and ansatze for exclusive-to-inclusive relations.

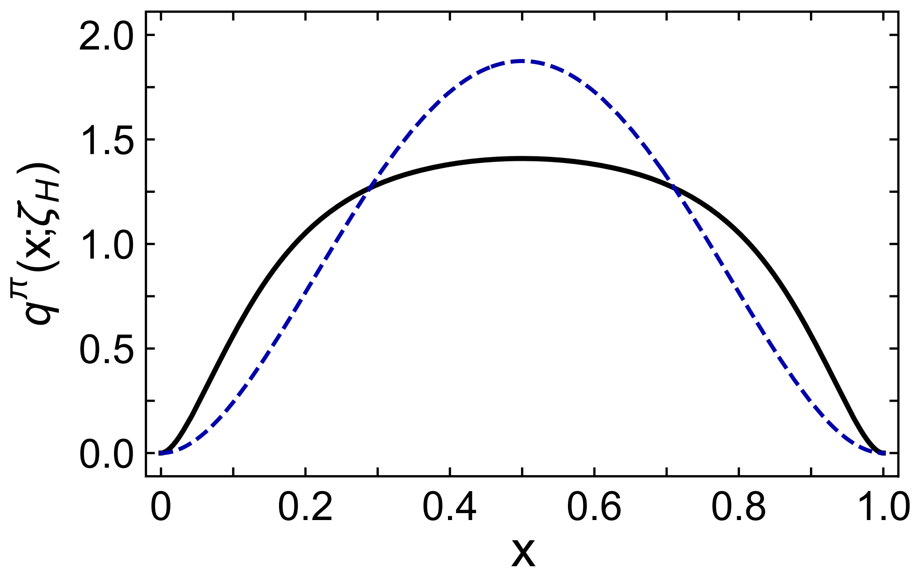
Why study the pion?

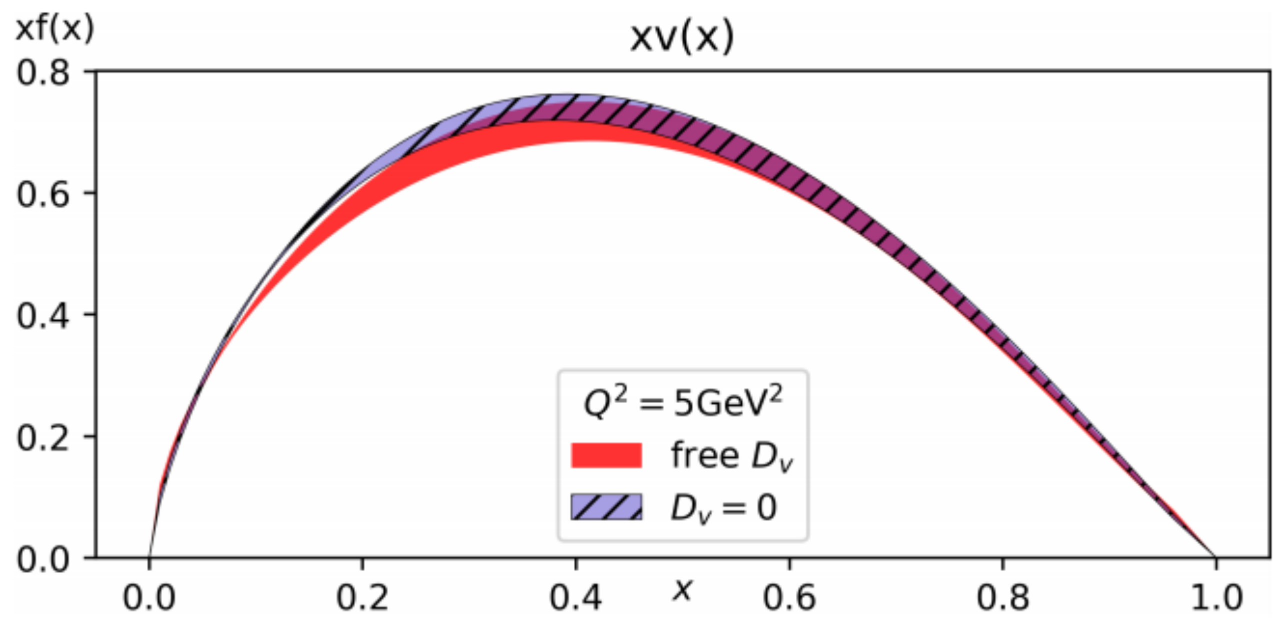
- ➡ xFitter's framework set up the pion PDF analysis— https://www.xfitter.org/xFitter/
- ➡ less data *wrt* proton, still at NLO accuracy
- ➡ recent "come back" thanks to increased fitting activity in the nuclear community —theory and experiment-wise

⇨ Pion PDFs are closely related to the dynamics of QCD in non-perturbative regime— trickier interpretation due to its pseudo-Goldstone nature and ansatze for exclusive-to-inclusive relations.

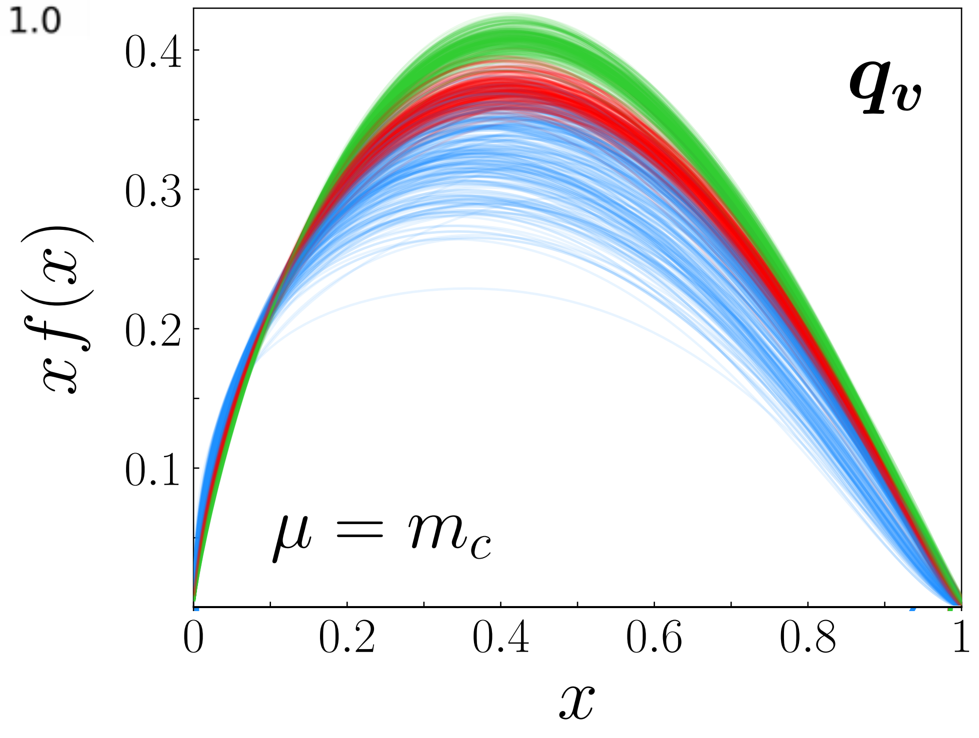


e.g. Nambu—Jona-Lasinio model, Schwinger-Dyson approaches, …



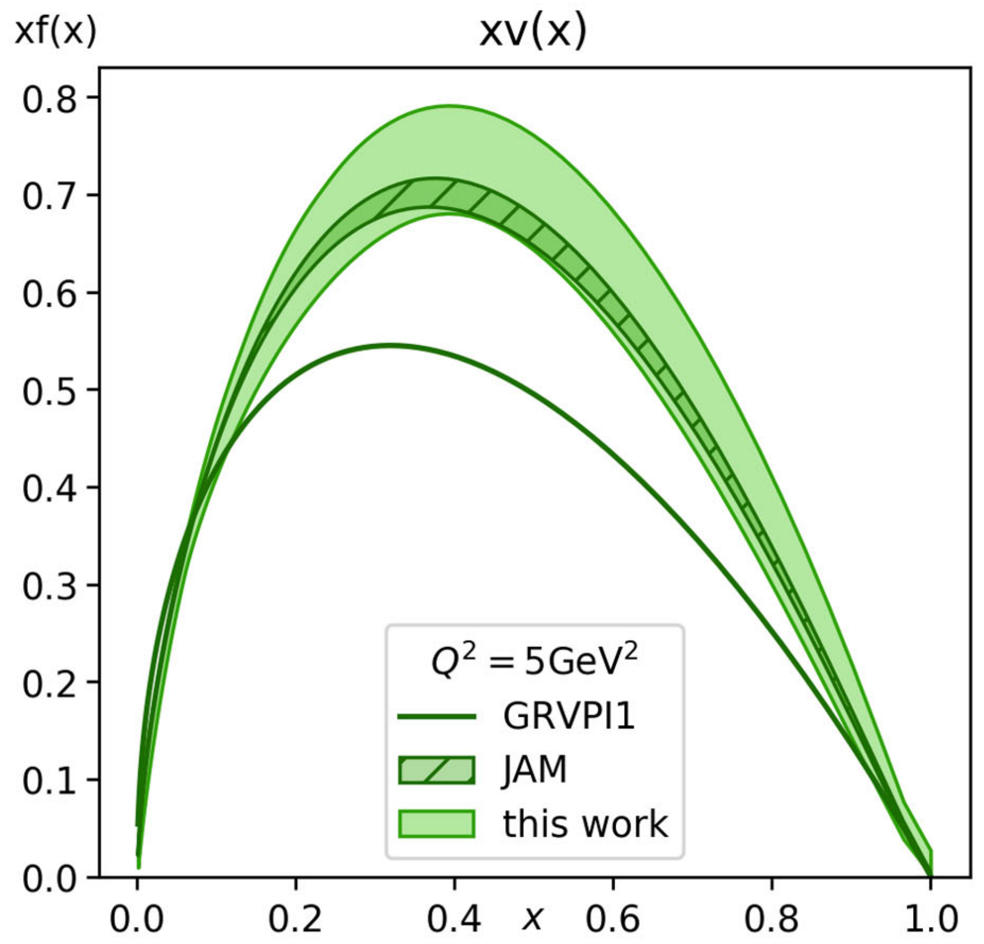


Global analysis groups:
xFitter [*PRD* 102 (2020)]
JAM [PRL 121 (2018), *PRD* 103, *PRL* 127 (2021)]
…



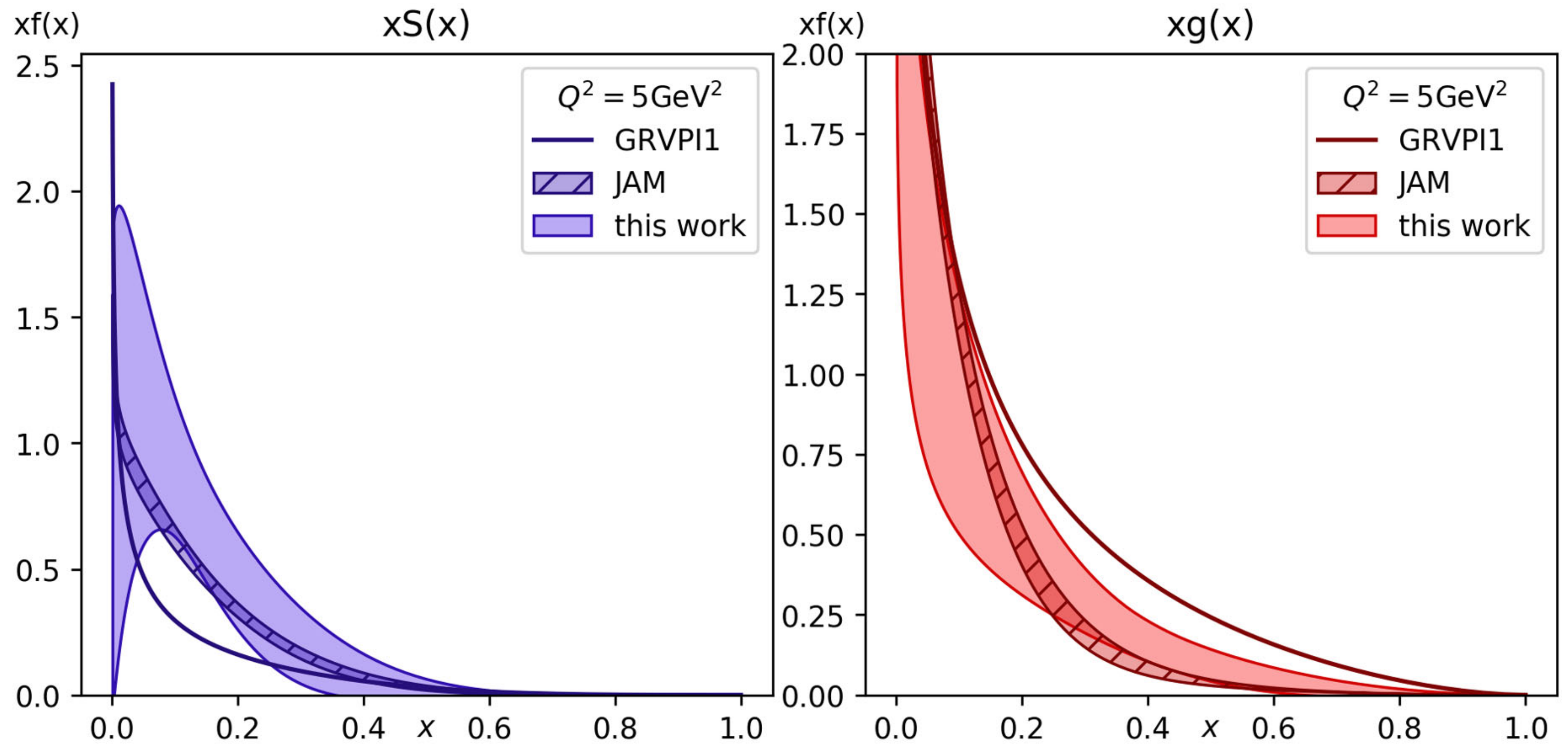
State-of-the-art of pion PDF in global analyses

Pioneer pion-induced Drell-Yan analyses (GRV, SMRS….) replaced by modern analyses



by **xFitter** [from which I took the plot]

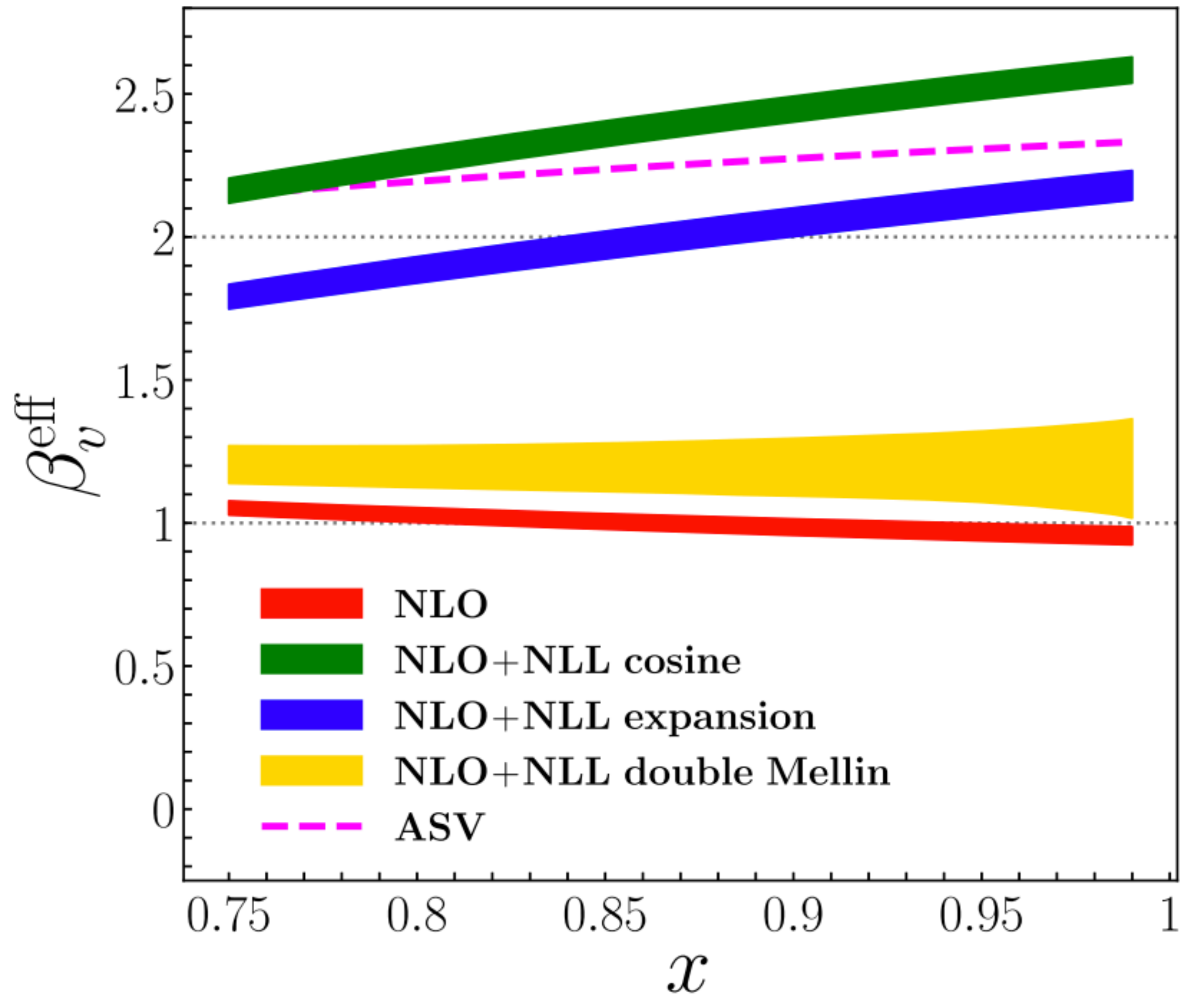
complemented by [model-dependent] leading-neutron data [**JAM**]



include large-x resummation

➡ ASV [Aicher etal., PRL105]

➡ **JAM21** [Barry et al., PRL127]

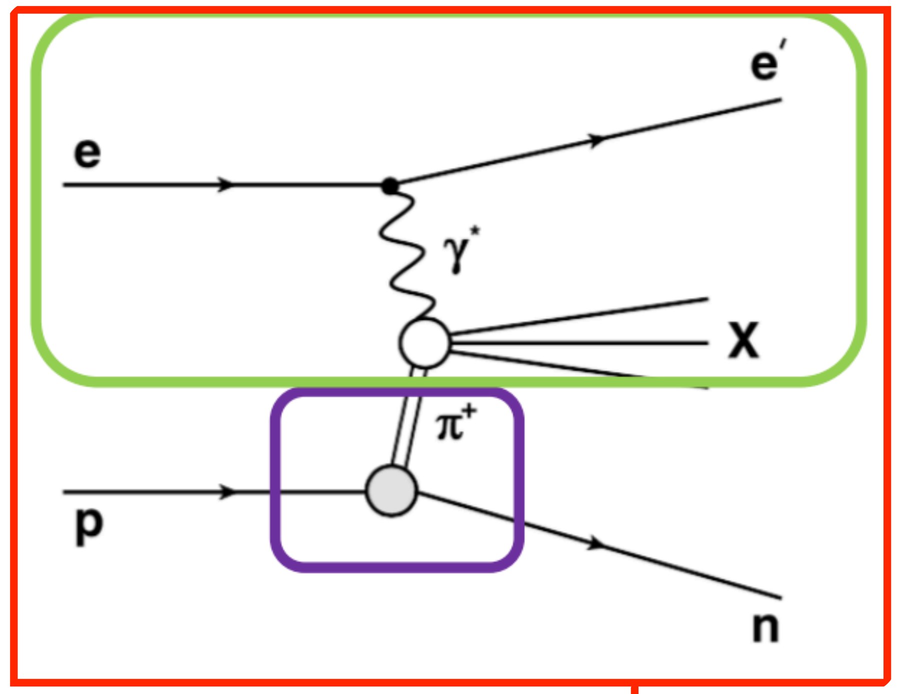


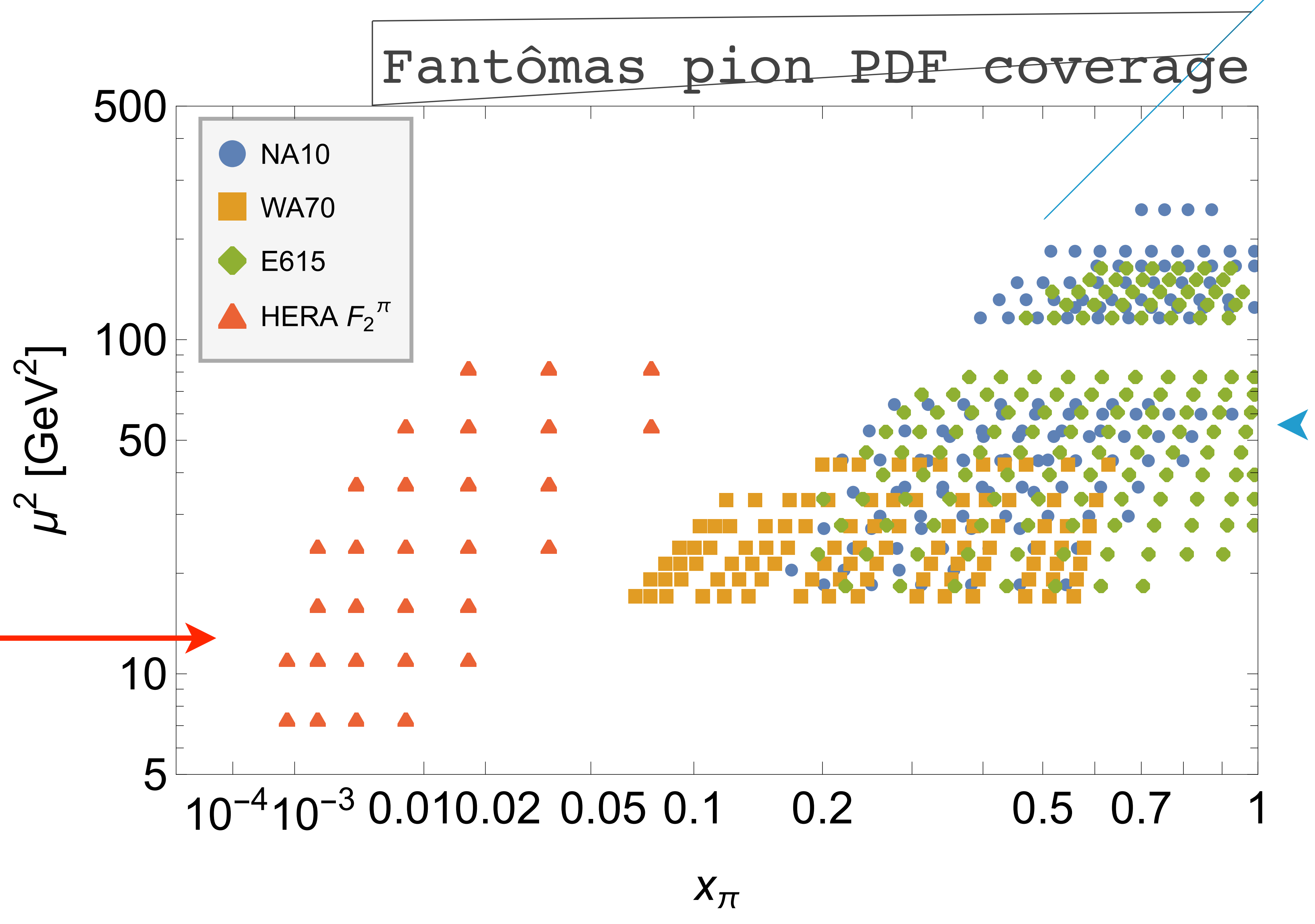
Data for pion PDF

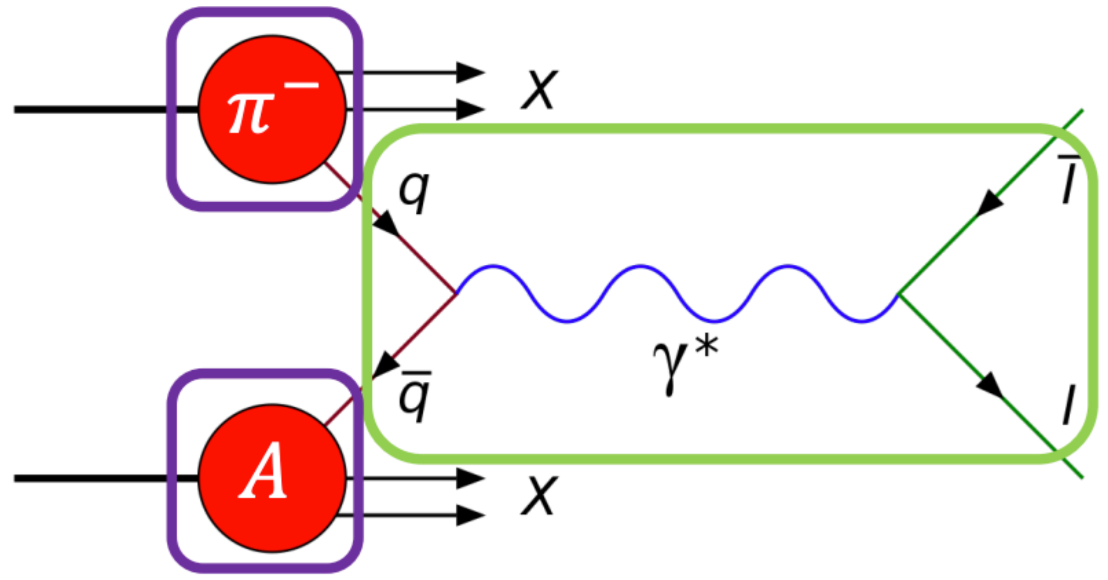
We use the xFitter framework, in which `metamorph` was implemented as an independent parametrization.

We also extend the xFitter data:

- pion-induced Drell-Yan → constraints valence PDF at large x
- prompt photons → may constrain gluon PDF at largish x
- leading neutron (Sullivan process) → only constraints on sea and gluon at $x \lesssim 0.1$ [Fantômas uses the H1 prescription]







Diagrams from P. Barry

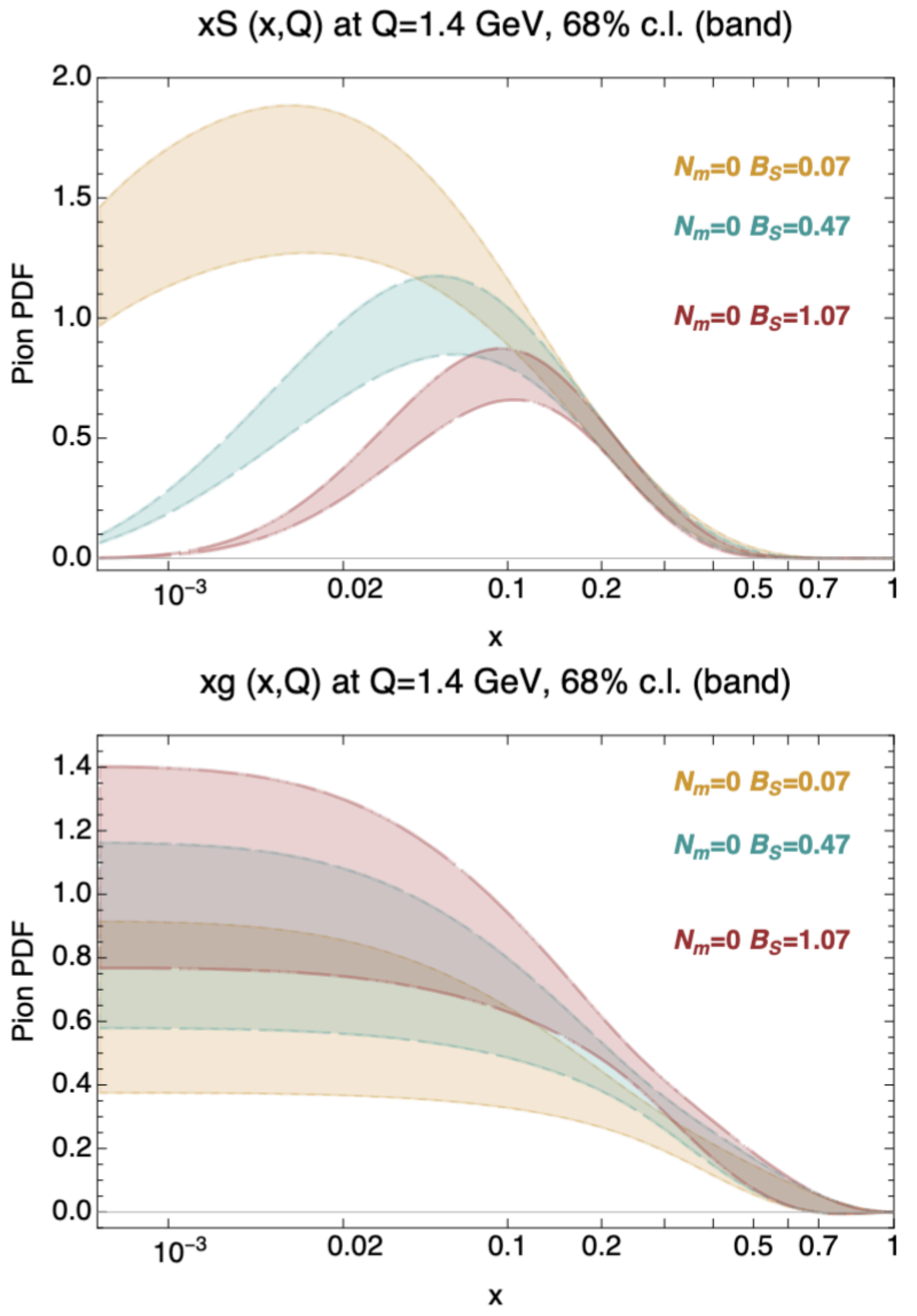
Drell-Yan only analysis

Previous analyses used a fairly basic parametrization

$$xf_{q/\pi}(x, Q_0) = Nx^{\alpha}(1-x)^{\beta} \times \left(1 + \gamma\sqrt{x} + \cdots\right)$$

With a rigid parametrization, in Drell-Yan only analysis, the sea and gluon pion distributions are not well determined.

We can achieve equally good or better fits by varying the small-x behaviour of the sea PDF [B_S] within xFitter uncertainty.



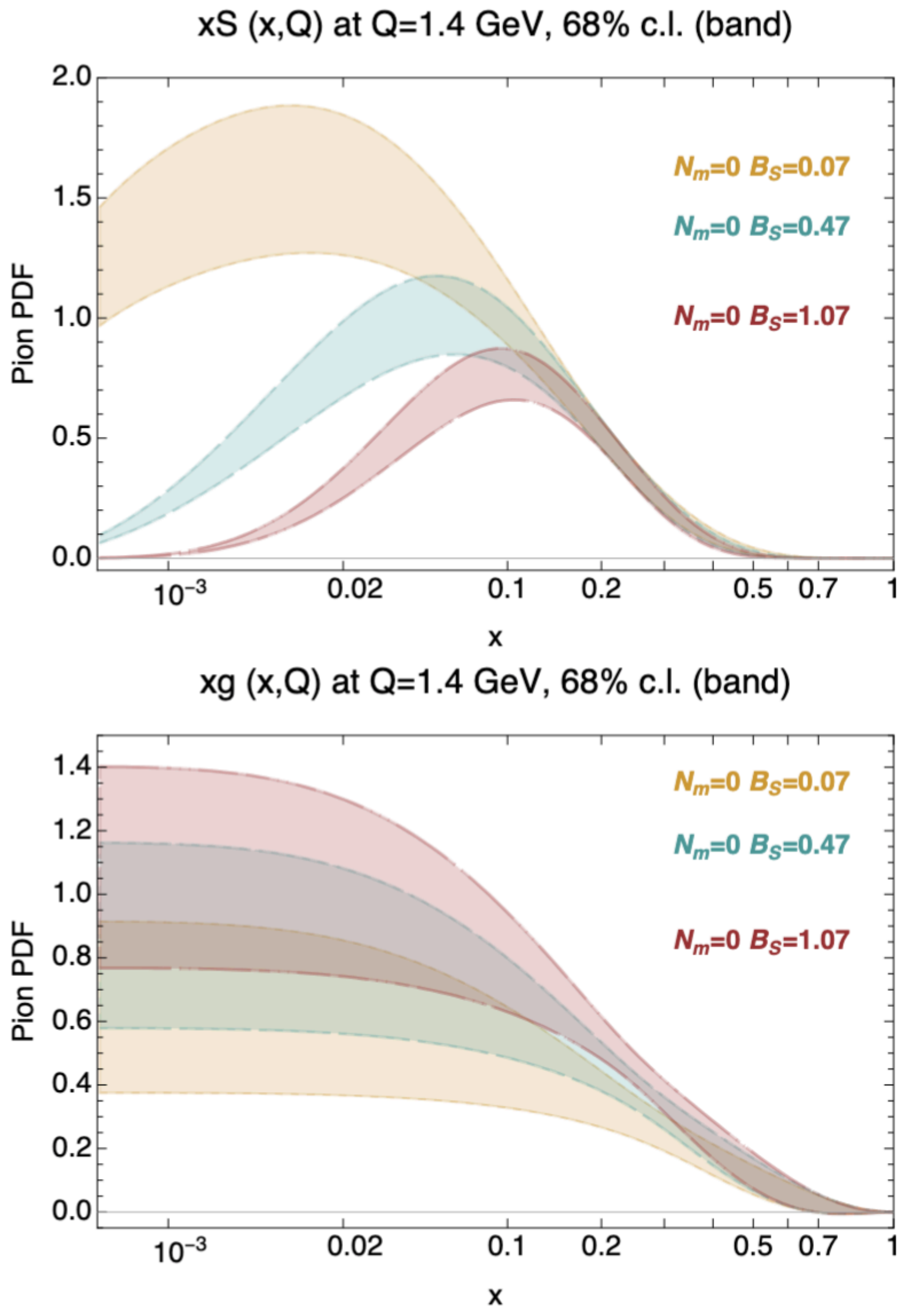
Drell-Yan only analysis

Previous analyses used a fairly basic parametrization

$$xf_{q/\pi}(x, Q_0) = Nx^{\alpha}(1-x)^{\beta} \times \left(1 + \gamma\sqrt{x} + \cdots\right)$$

With a rigid parametrization, in Drell-Yan only analysis, the sea and gluon pion distributions are not well determined.

We can achieve equally good or better fits by varying the small-x behaviour of the sea PDF [B_S] within xFitter uncertainty.



Need for complementary processes— universality and flavor separation

⇨ JAM (and HERA before them) proposed to use leading-neutron data

⇨ future experiments at EIC and JLab22(?)

The Fantômas pion PDFs

[Kotz, Ponce-Chávez, AC, Nadolsky & Olness]
Proceedings in 2309.00152.

First physics use of the Fantômas framework:

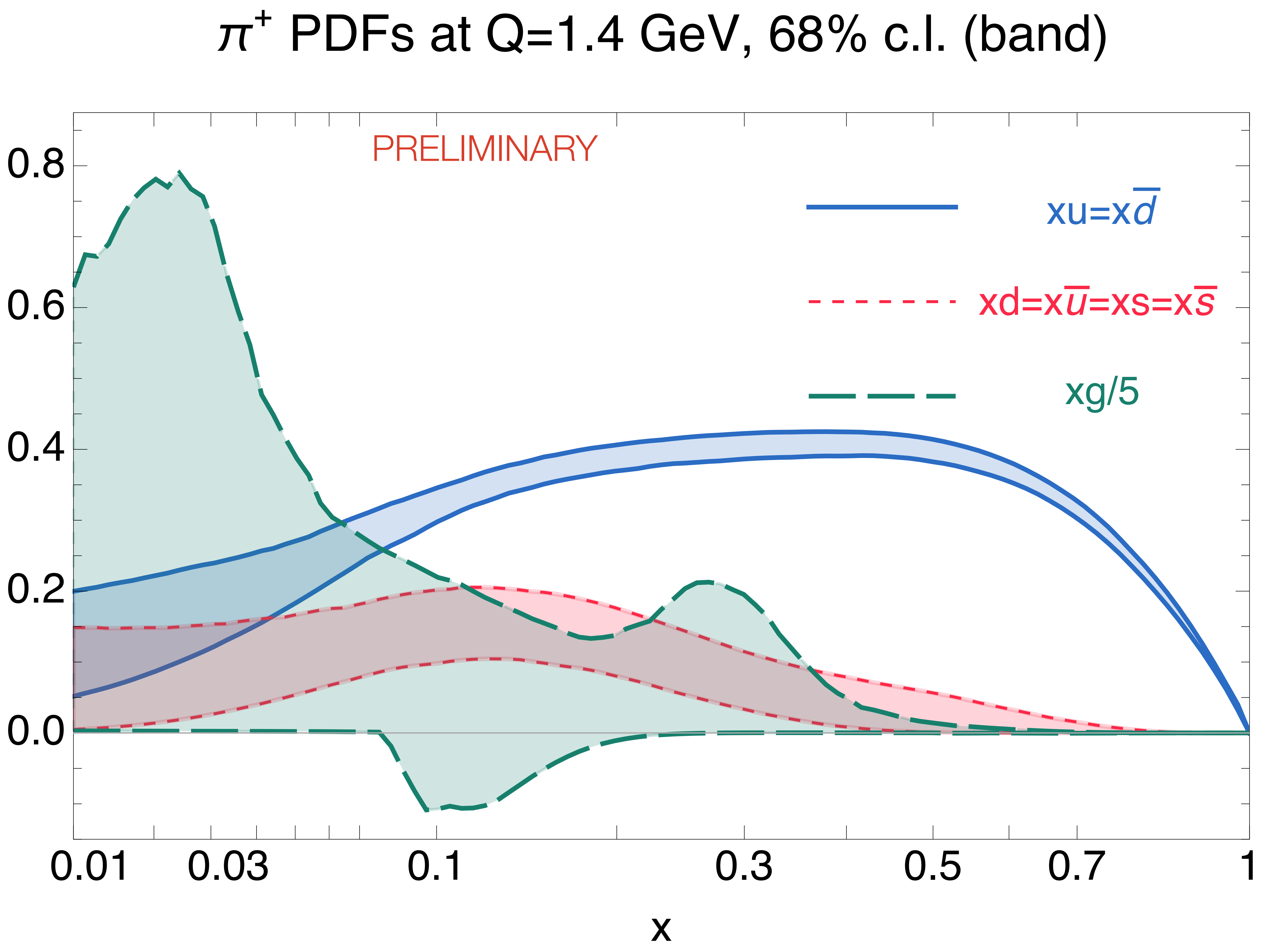
⇨ We generated $N \sim 75$ fits corresponding to N sets for $\{N_m, \underline{P}, \alpha_x\}$.

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

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

⇨ The final bundle is generated from the 4 most diverse shapes at Q_0.

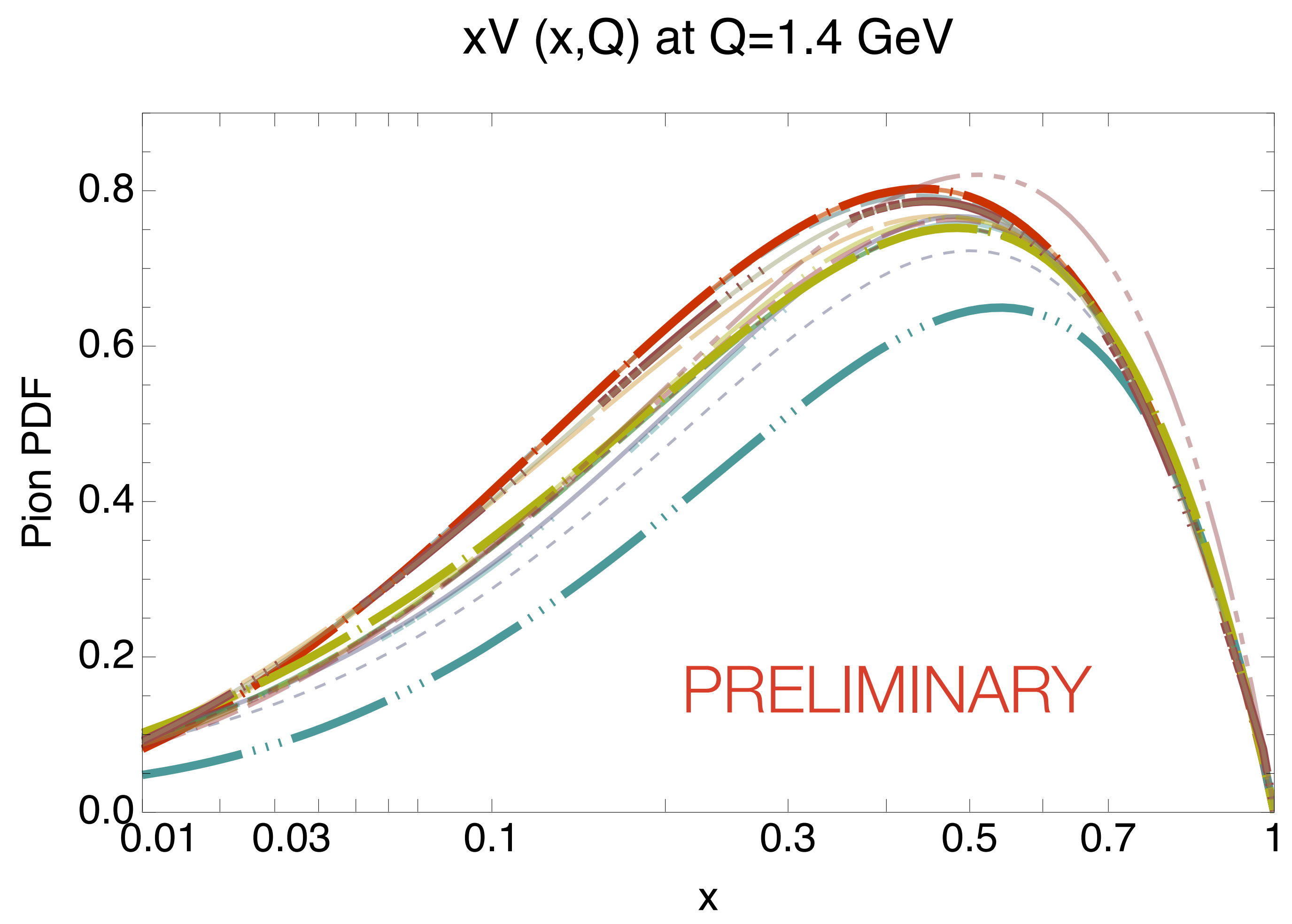
⇨ Bundled uncertainty with `mcgen` [Gao & Nadolsky, JHEP07]

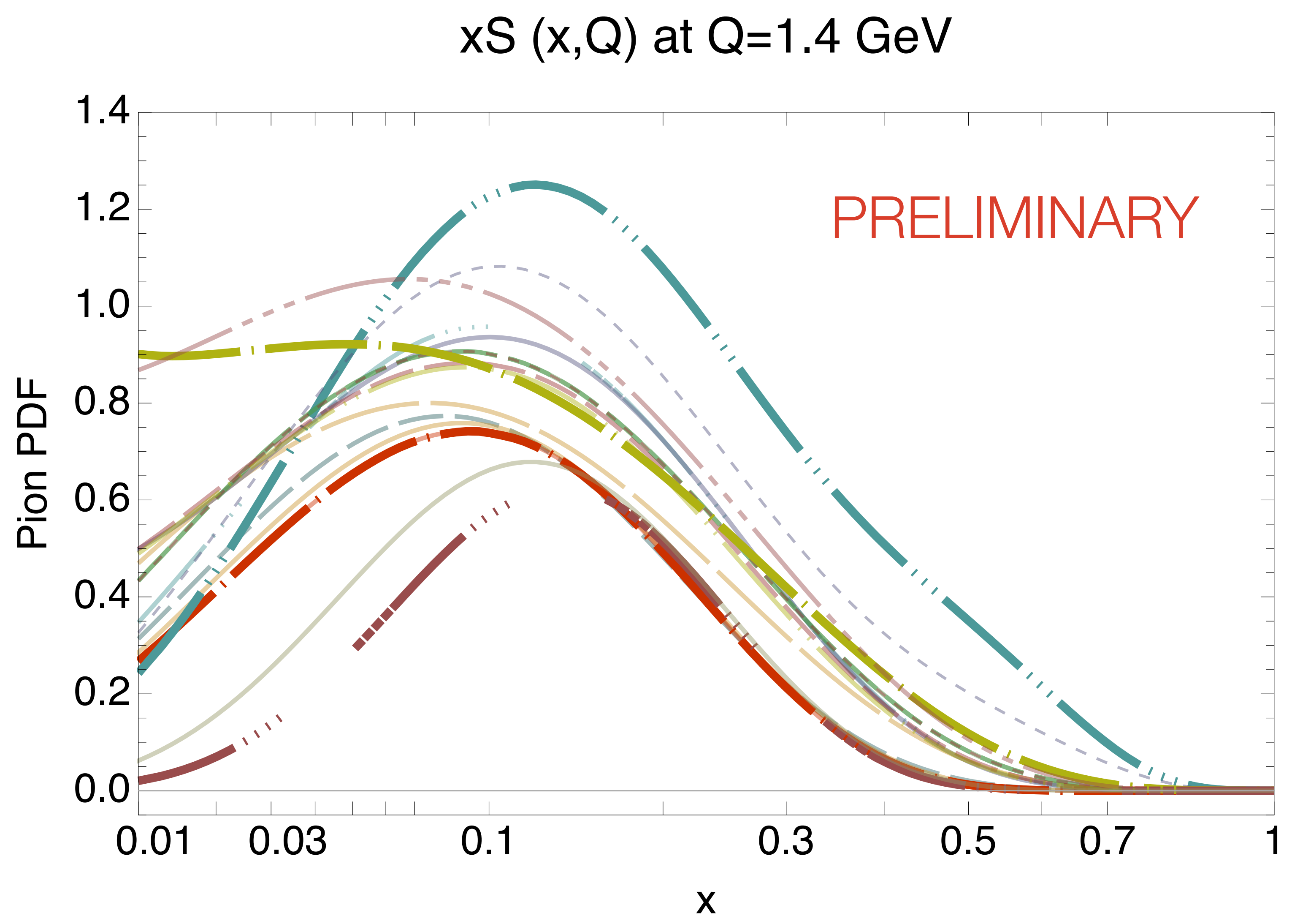


Fantômas parametrizations for the pion PDF

Fantômas analysis uses varying sets of

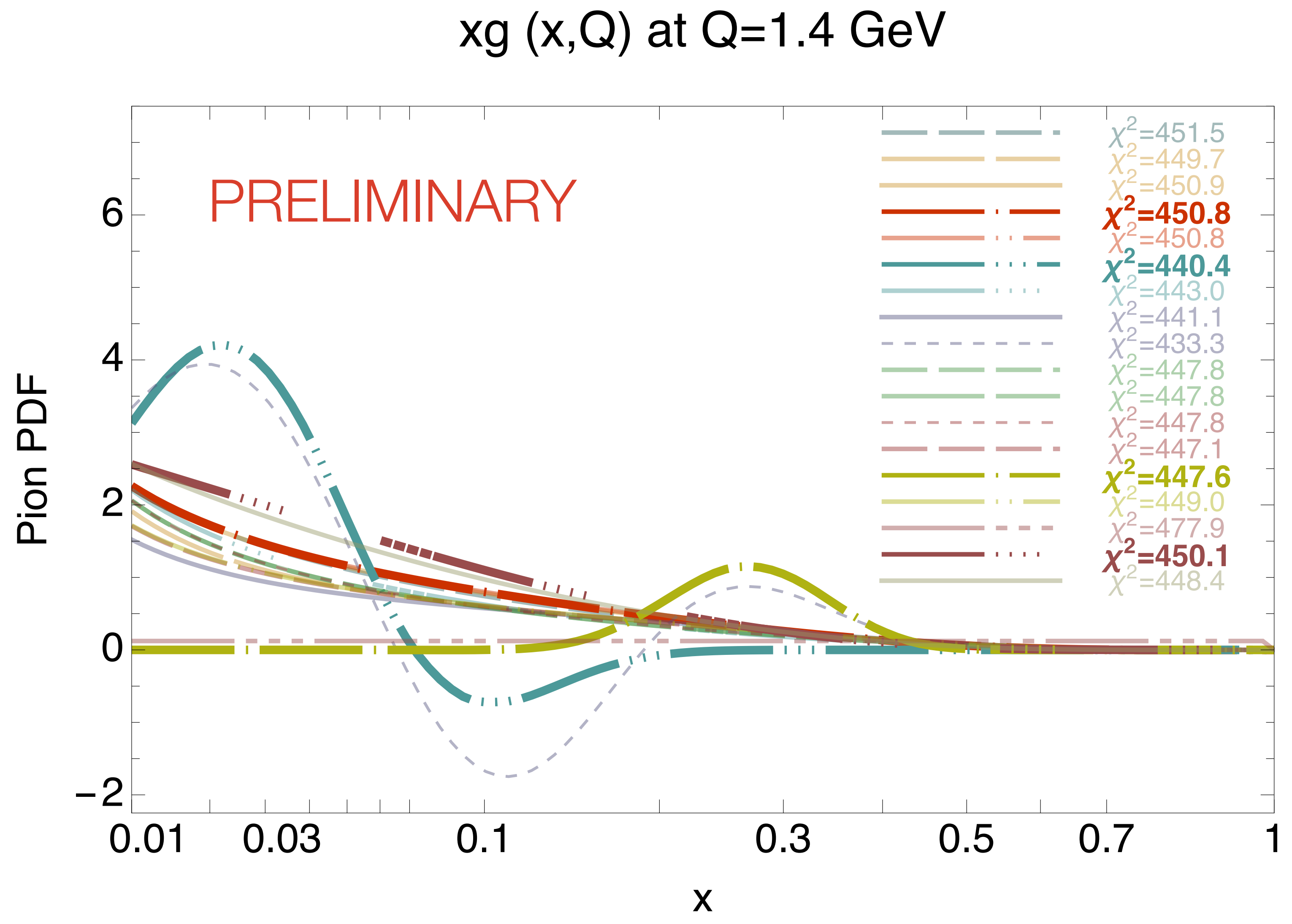
- degree of polynomial (0,1,2),
- position of fixed/free control points,
- stretching parameter of the argument





Extrapolation region for pion PDF is around $x = 0.1$ at Q_0.

Negative gluon are found to be possible at such a low scale [confirming JAM's findings].



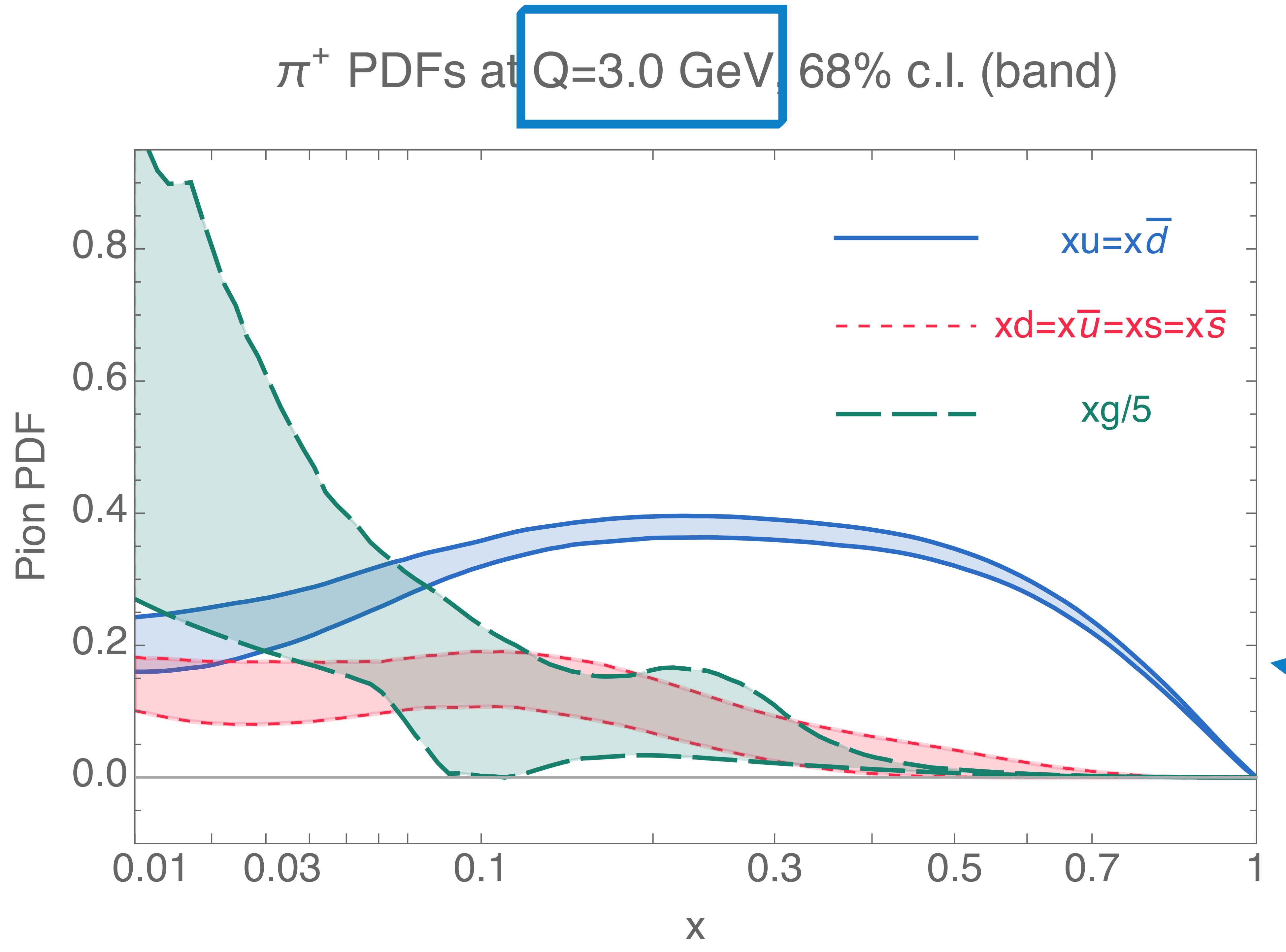
Bold curves correspond to our selection for the final Fantômas set.

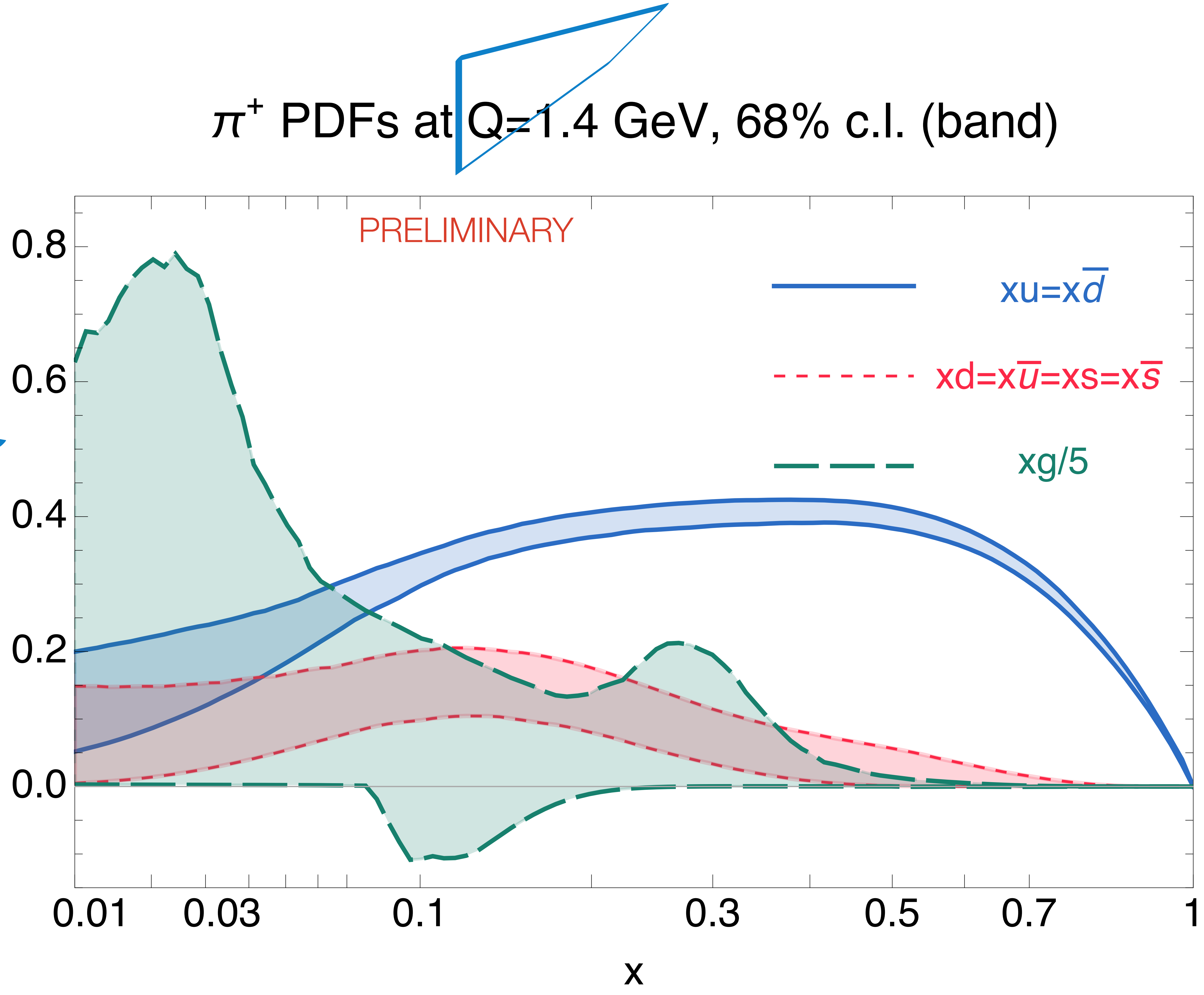
Representative curves within χ^2 range: $\chi^2 + \delta\chi^2 = \chi^2 + \sqrt{2(N_{\text{pts}} - N_{\text{par}})} \simeq 440 + 30$

for 408 points and 7-13 parameters.

The Fantômas pion PDFs

[Kotz, Ponce-Chávez, AC, Nadolsky & Olness]
Proceedings in 2309.00152.



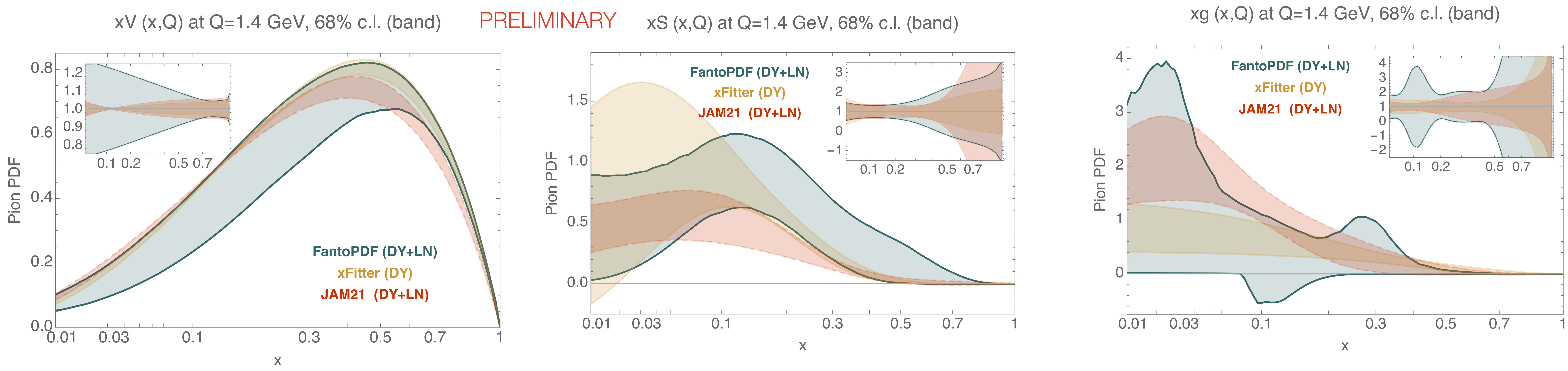


Scale recommendations for pion PDFs?

pQCD and non-perturbative arguments?

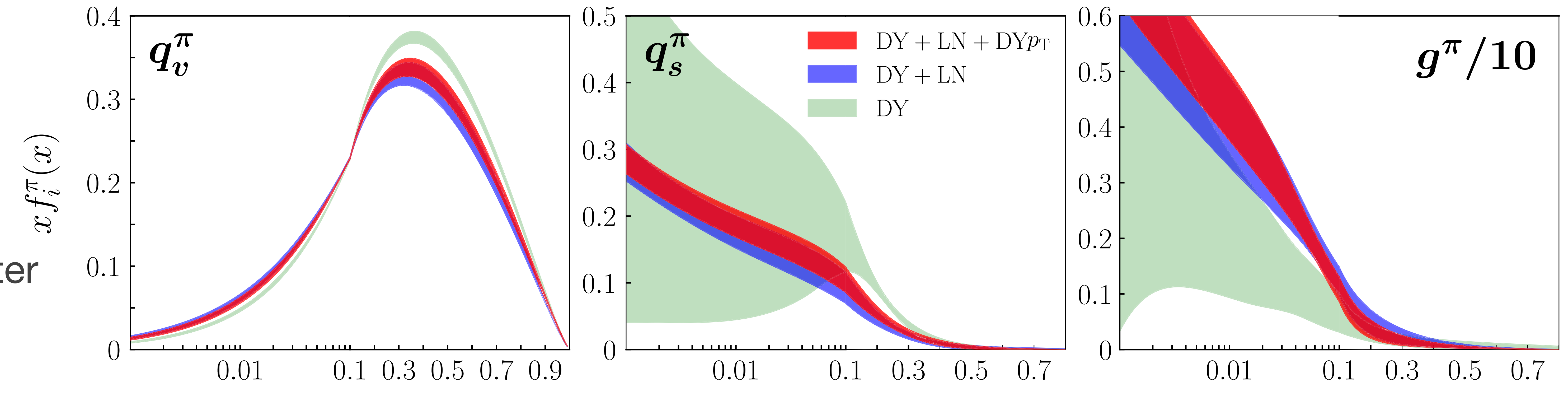
The Fantômas pion PDFs

*[Kotz, Ponce-Chávez, **AC,** Nadolsky & Olness]*
Proceedings in 2309.00152.



Comparison of methodologies:

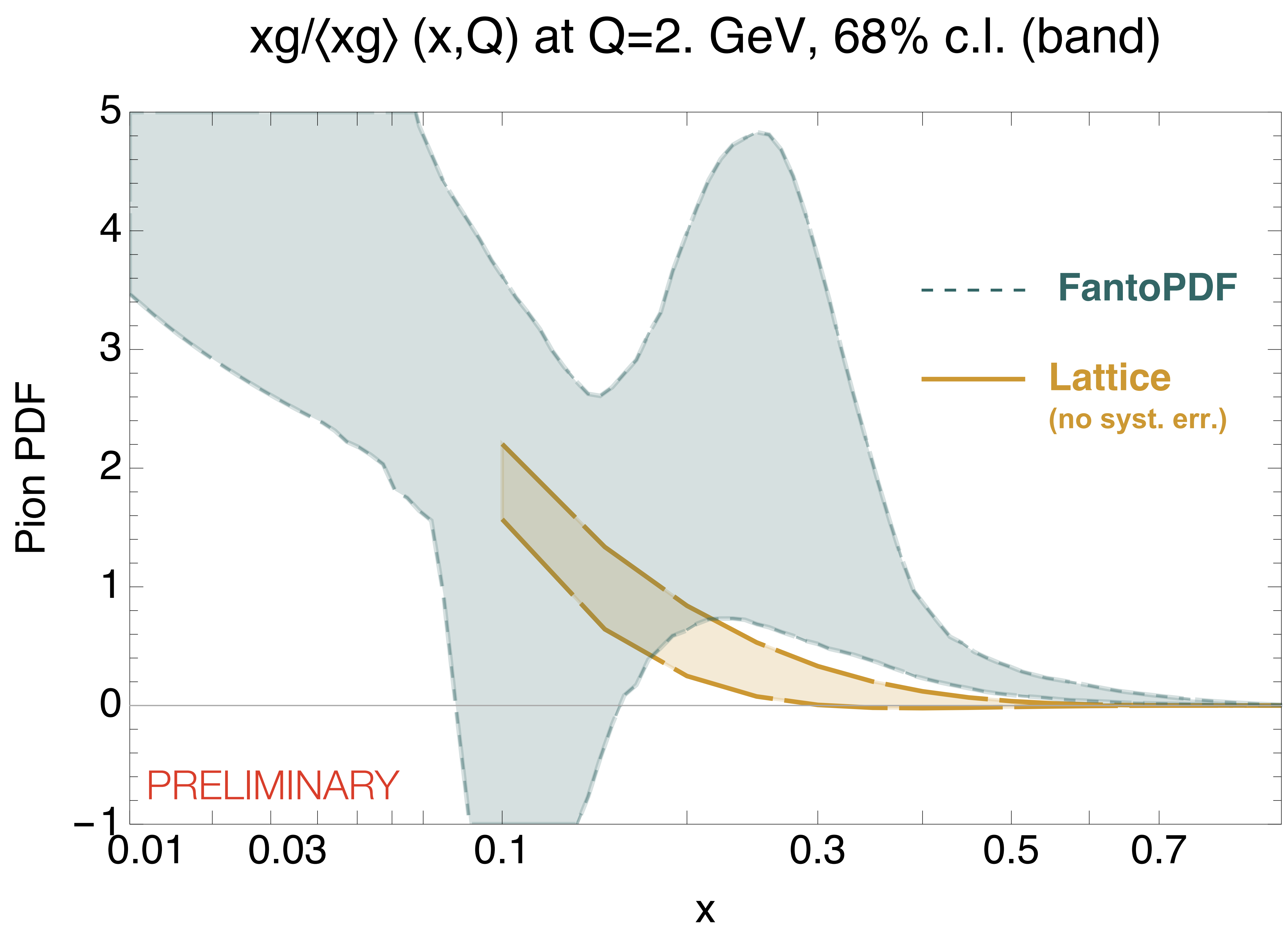
bootstrap+ IMC vs. `metamorph` parametrization in xFitter



JAM [Phys.Rev.D103]

Gluon PDF compared with lattice QCD results

Gluon shape averaged to momentum fraction given by [Fan & Lin, PLB 823 (2021)]



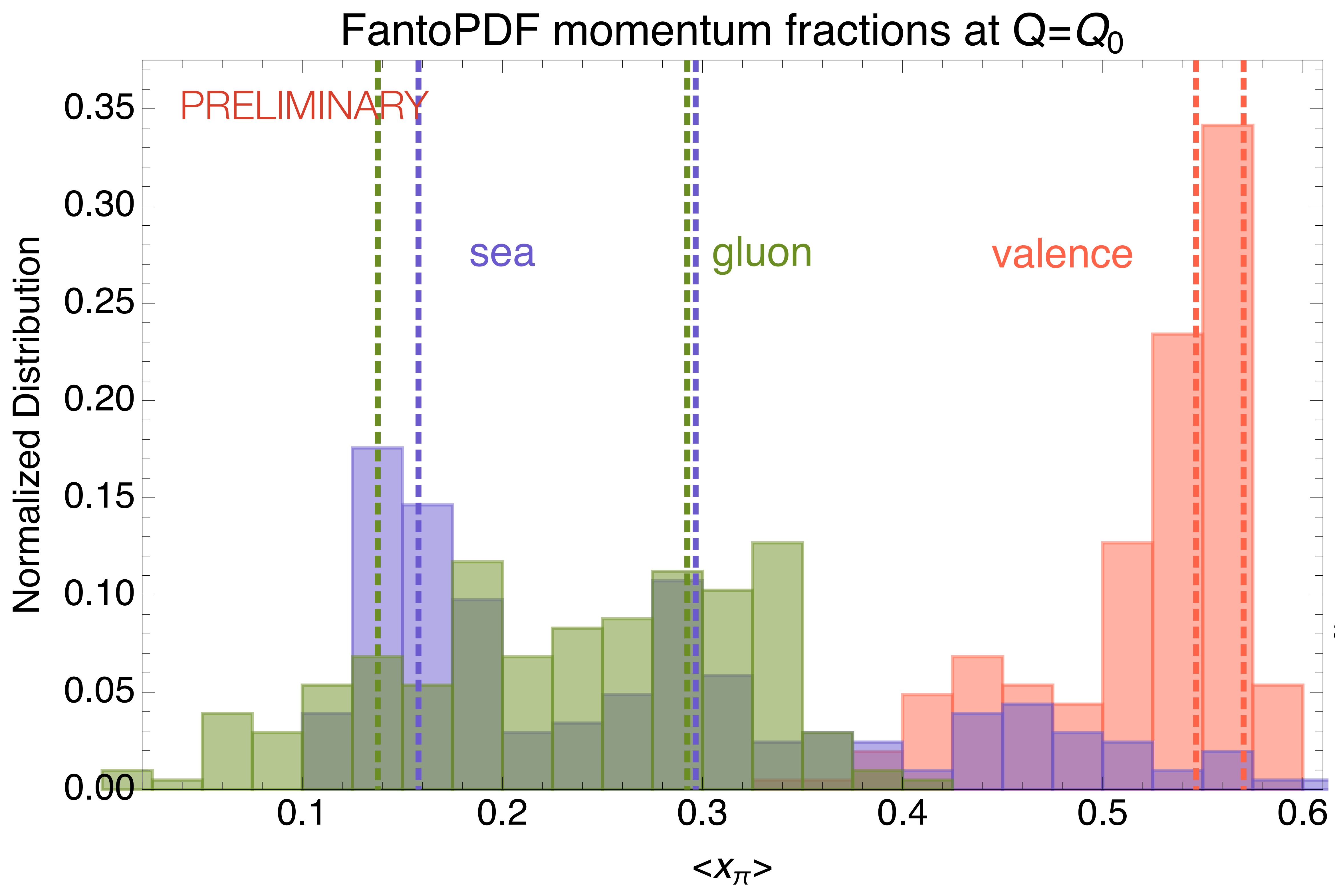
Momentum fractions

As it turns out, the valence sector was not as exciting as expected — sea and gluon separation got most of our attention!

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

Increased uncertainty on all three $\langle xf_q \rangle$.

Valence fraction $\langle xf_v \rangle (Q = 2\,\mathrm{GeV}) = 0.48 \pm 0.05$ compatible with lattice results.

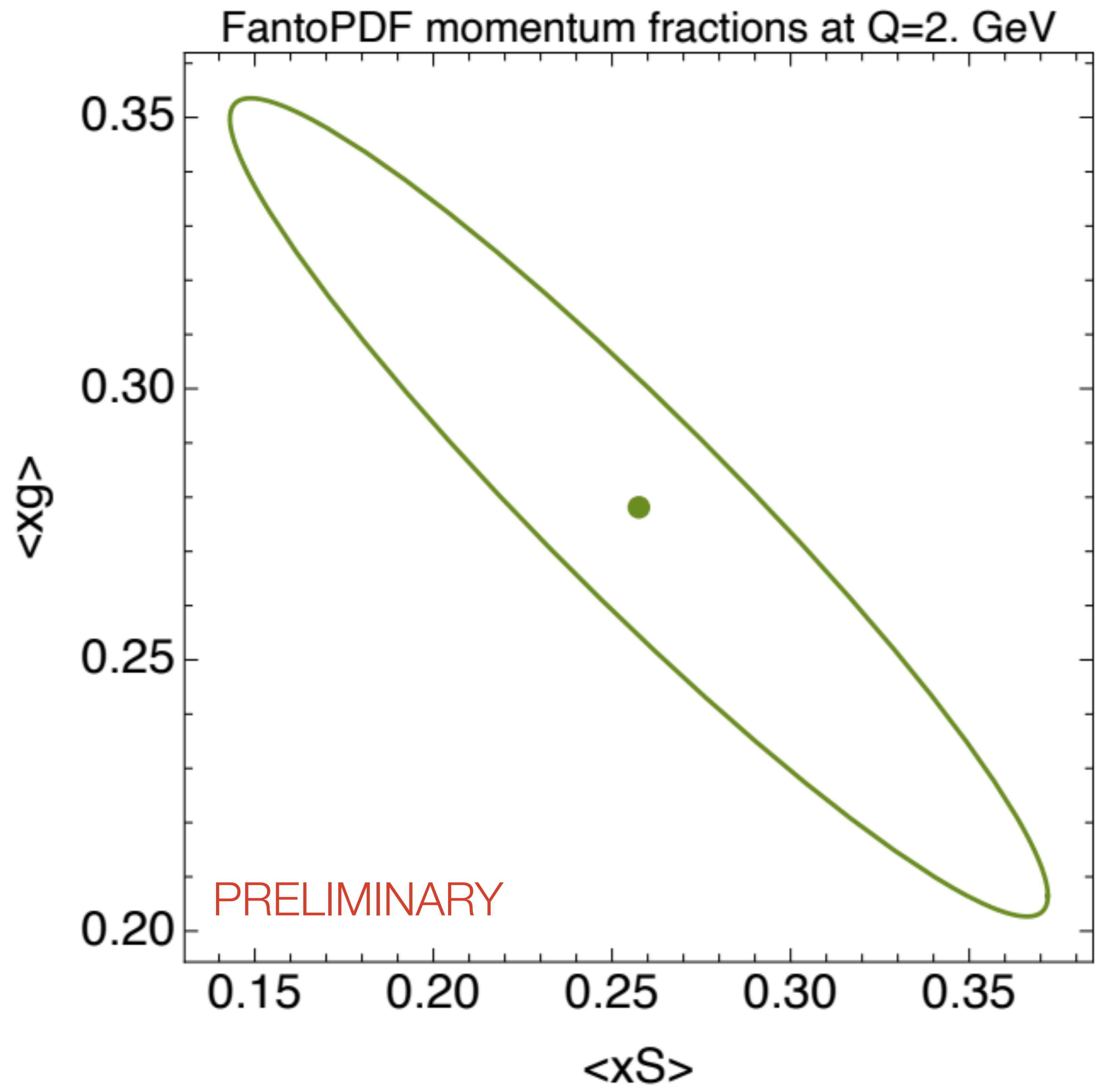


Momentum fractions

Momentum-fraction distributions for gluon and sea are largely (anti)-correlated.

We obtain $\langle x f_g \rangle (Q = 2\,\text{GeV}) = 0.28 \pm 0.08.$

Funny fact: some lattice results for gluon momentum fraction suggest a very large fraction of the momentum is carried by the gluon, in an incompatible proportion *wrt* the valence.



Hypothesis testing from PDFs

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

In any inference about primordial dynamics, unbiased determination of the PDF functional form must be fully evaluated to consider an *iif* validation of polynomial shapes.

Hypothesis testing from PDFs

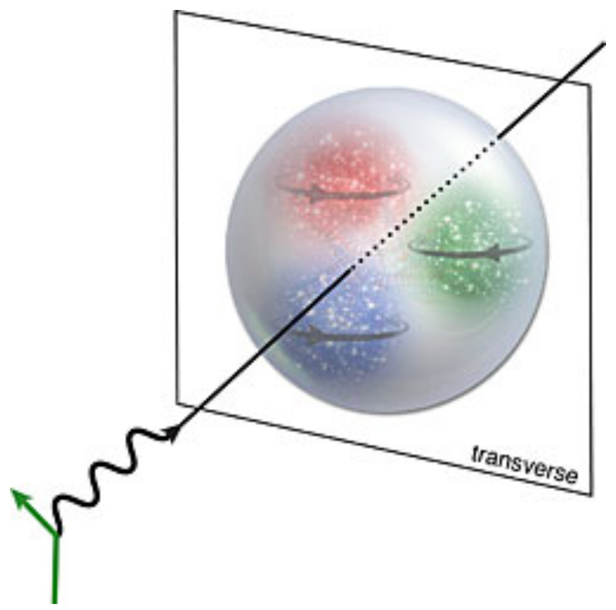
⇨ <u>Hypothesis testing for functional behavior constraints</u> — *do PDFs fall off like* $(1-x)^{\beta}$*?*

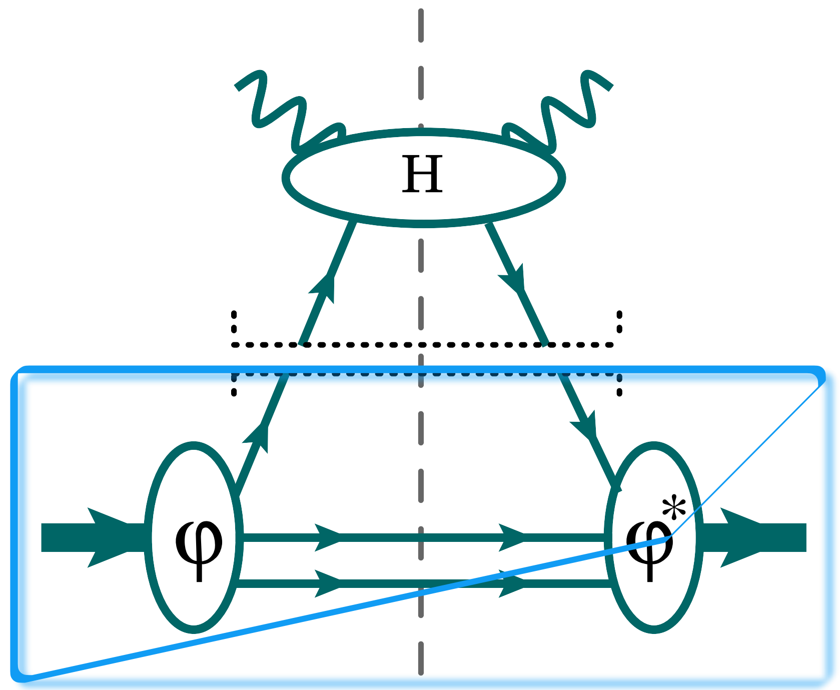
In any inference about primordial dynamics, unbiased determination of the PDF functional form must be fully evaluated to consider an *iif* validation of polynomial shapes.

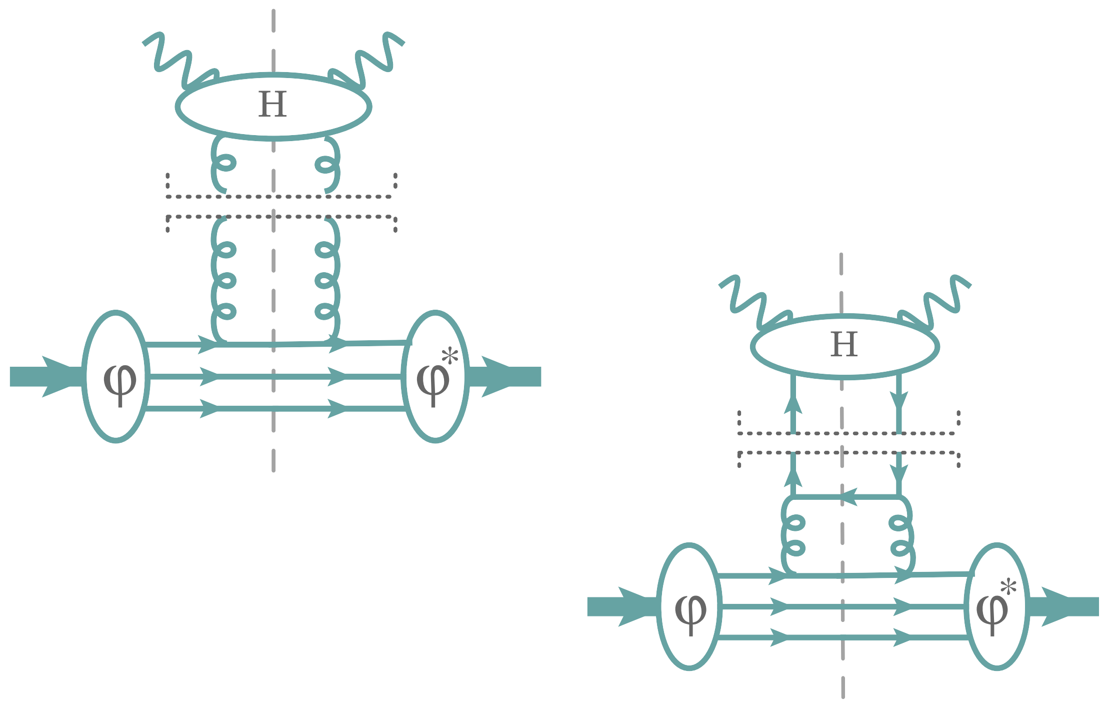
Correlators evaluated in non-perturbative approaches from open-vertices diagrams

PDFs extracted from data thanks to factorization theorems

$$\frac{1}{2}\int\frac{d\lambda}{2\pi}\,e^{i\lambda x}\,\langle P|\bar{\psi}(0)\gamma^{\mu}\psi(\lambda n)|P\rangle = f_1(x)p^{\mu} + \mathcal{O}(M^2)$$







μ_0^2 $\mu^2 \sim Q^2$

Hypothesis testing from PDFs

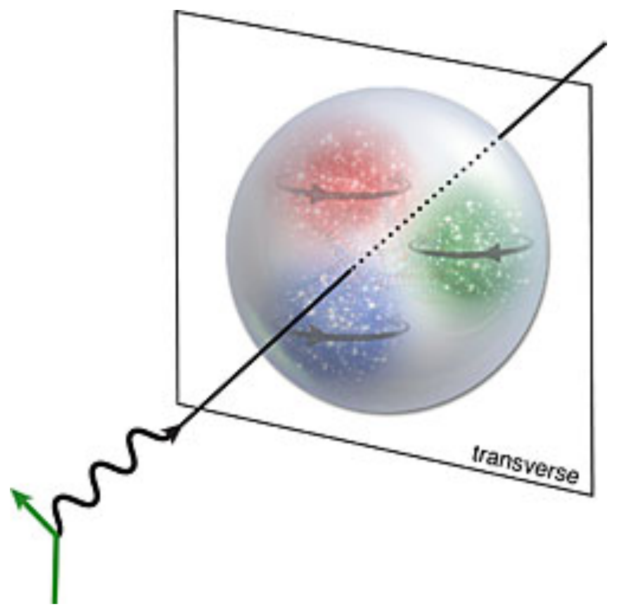
⇨ Hypothesis testing for functional behavior constraints — *do PDFs fall off like* $(1-x)^{\beta}$*?*

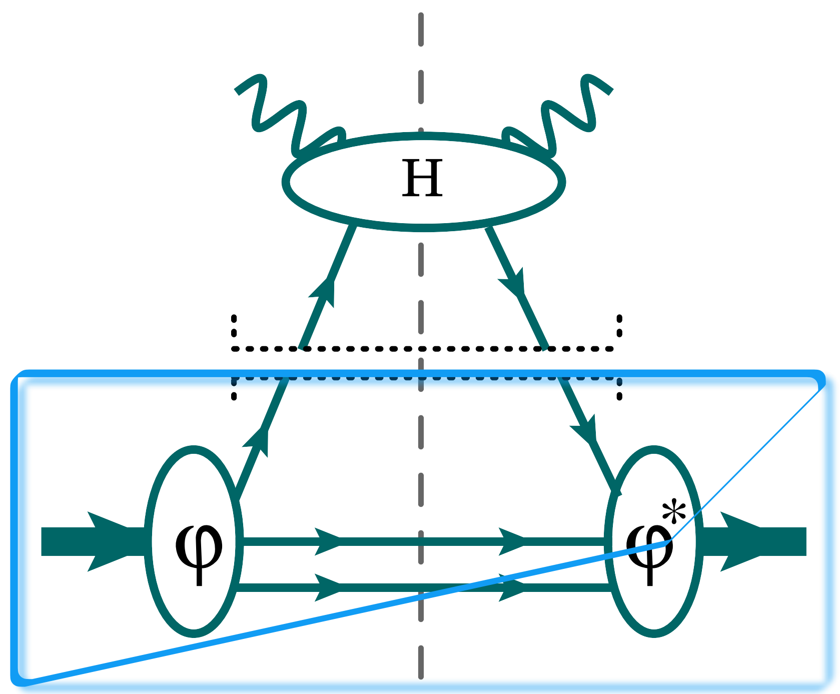
In any inference about primordial dynamics, unbiased determination of the PDF functional form must be fully evaluated for an *iif* validation of polynomial shapes.

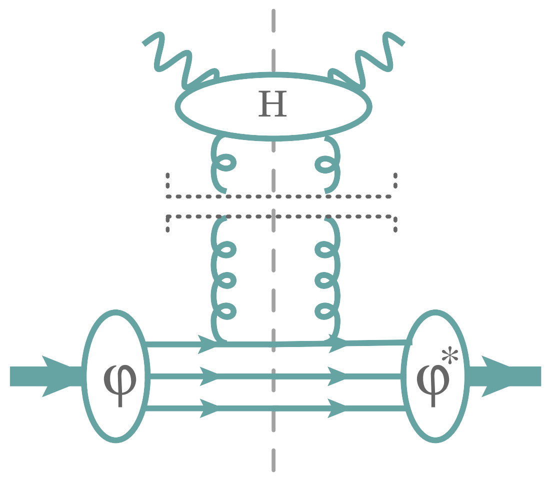
Quark-counting rules:

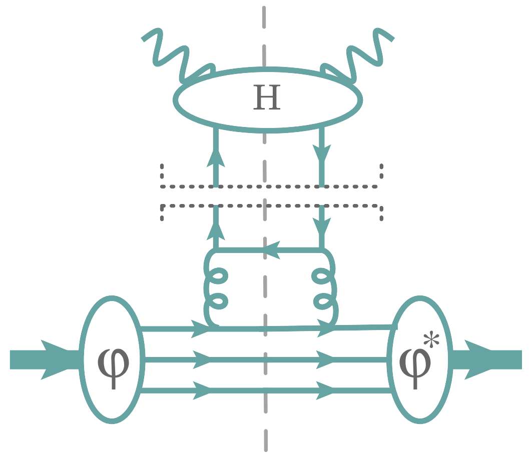
Early-QCD predicted behavior for structure functions when one quark carries almost all the momentum fraction

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



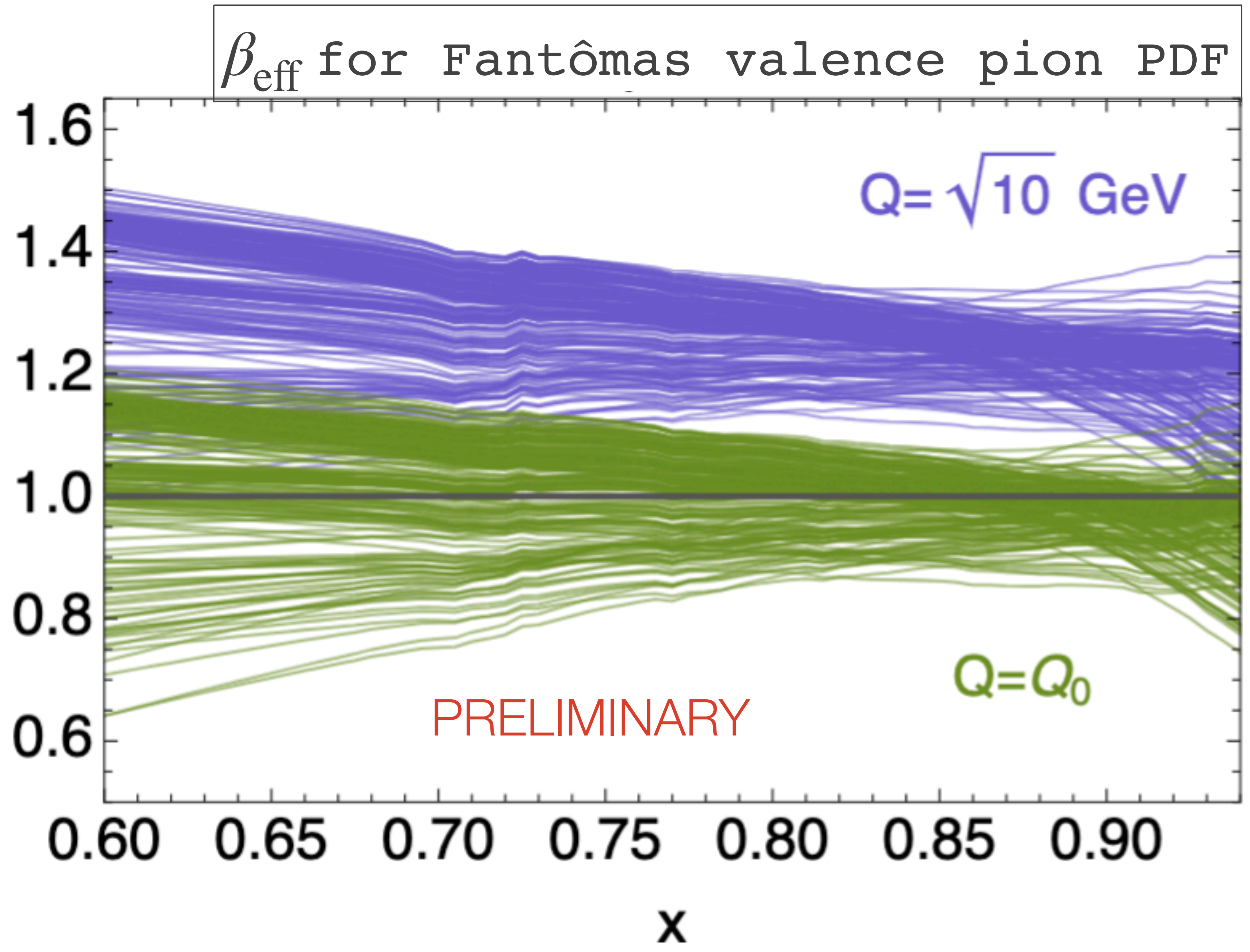






μ_0^2 $\quad$ $\mu^2 \sim Q^2$

Large-x behavior of the valence pion PDF

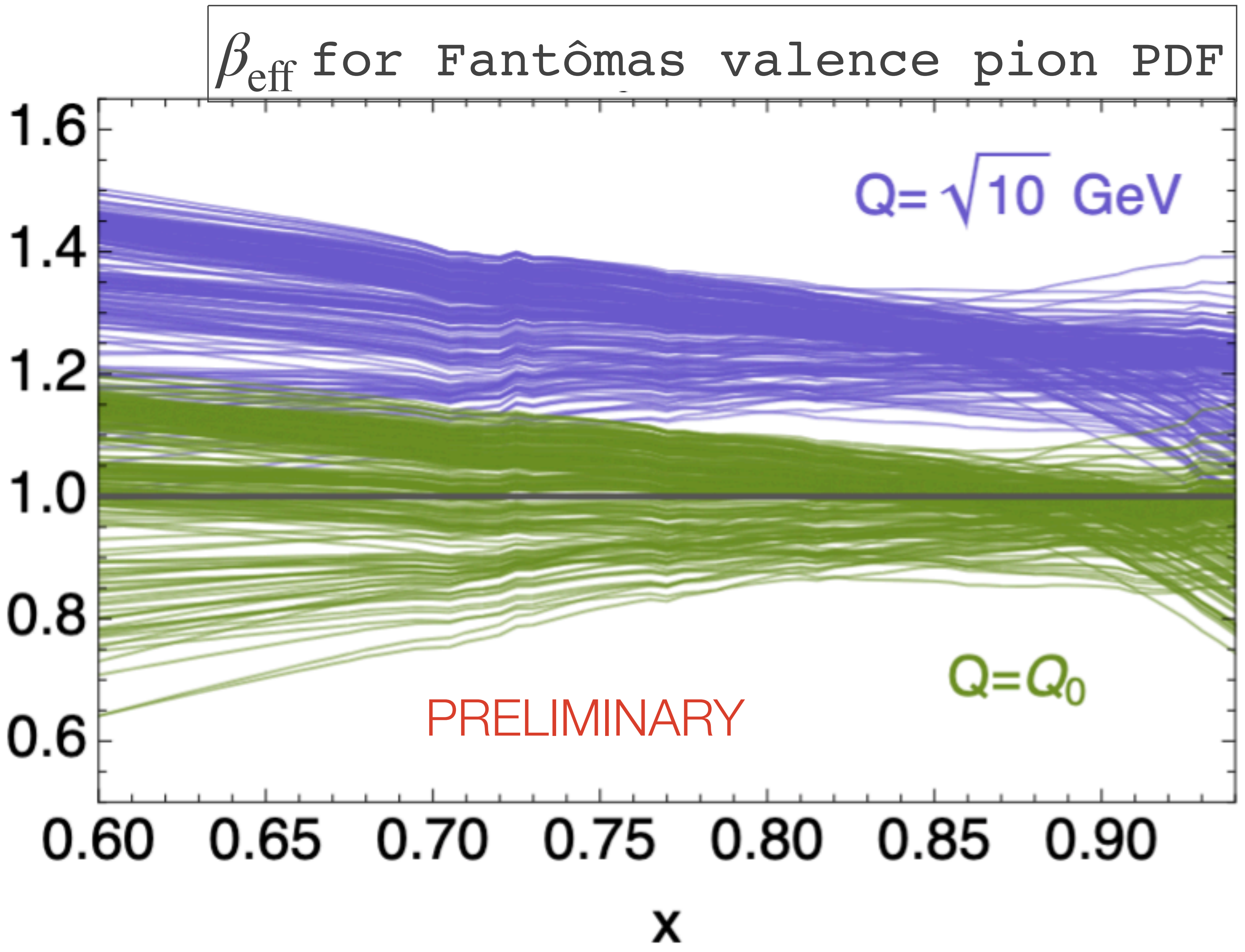


At NLO (MSbar), the valence PDF is well determined at large x

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

⇨ very similar to JAM and xFitter at large x

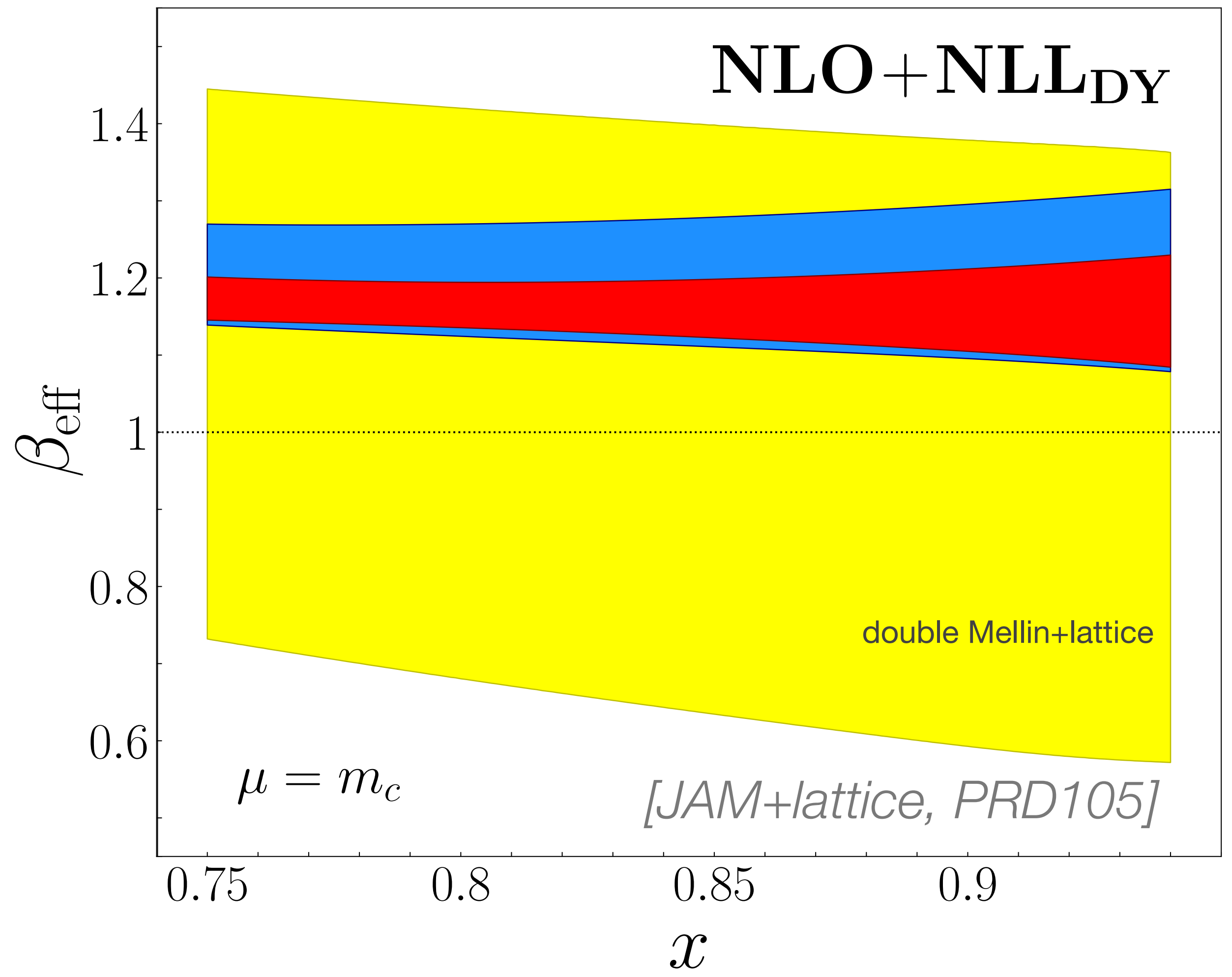
Large-x behavior of the valence pion PDF



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



Corrective terms might need to be taken into account [large-x resummation].
ASV found compatibility with $(1-x)^2$.
JAM did it and found an exponent between 1 to ~2.5, depending on the prescription.

Lattice studies: mindful analyses of the determination of the effective exponent of the PDF fall-off on the lattice [Gao et al., PRD102] ⇨ inverse problem

The Fantômas pion PDF

First analysis within the Fantômas framework!

Towards epistemic uncertainty: sampling over parameter space more representative

Pion PDFs with representative sampling over the space of solutions — here, parametrization is extended.

Not included (for now): uncertainties from scale dependence, nuclear PDF set, threshold resummation.

The Fantômas pion PDF

First analysis within the Fantômas framework!

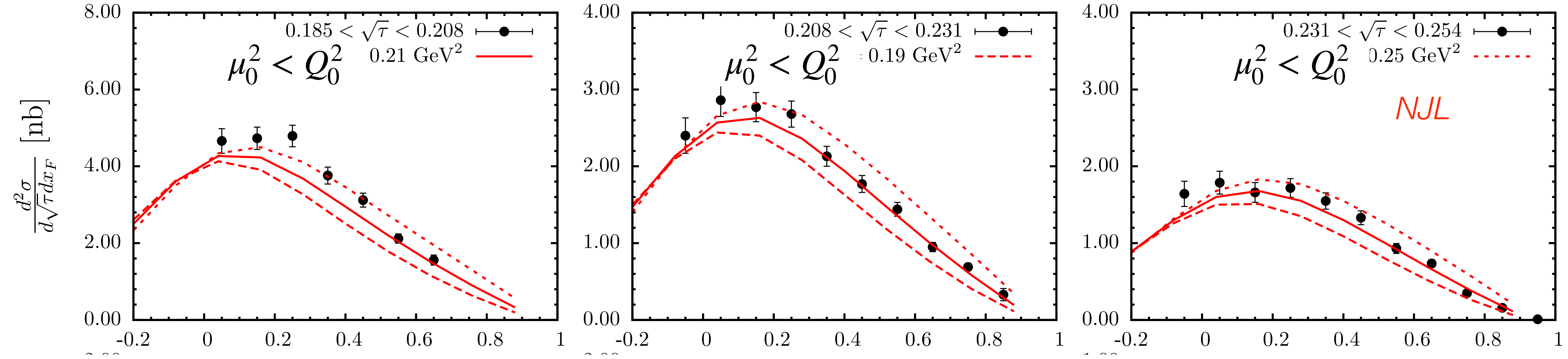
Towards epistemic uncertainty: sampling over parameter space more representative

Pion PDFs with representative sampling over the space of solutions — here, parametrization is extended.

Not included (for now): uncertainties from scale dependence, nuclear PDF set, threshold resummation.

What's next?

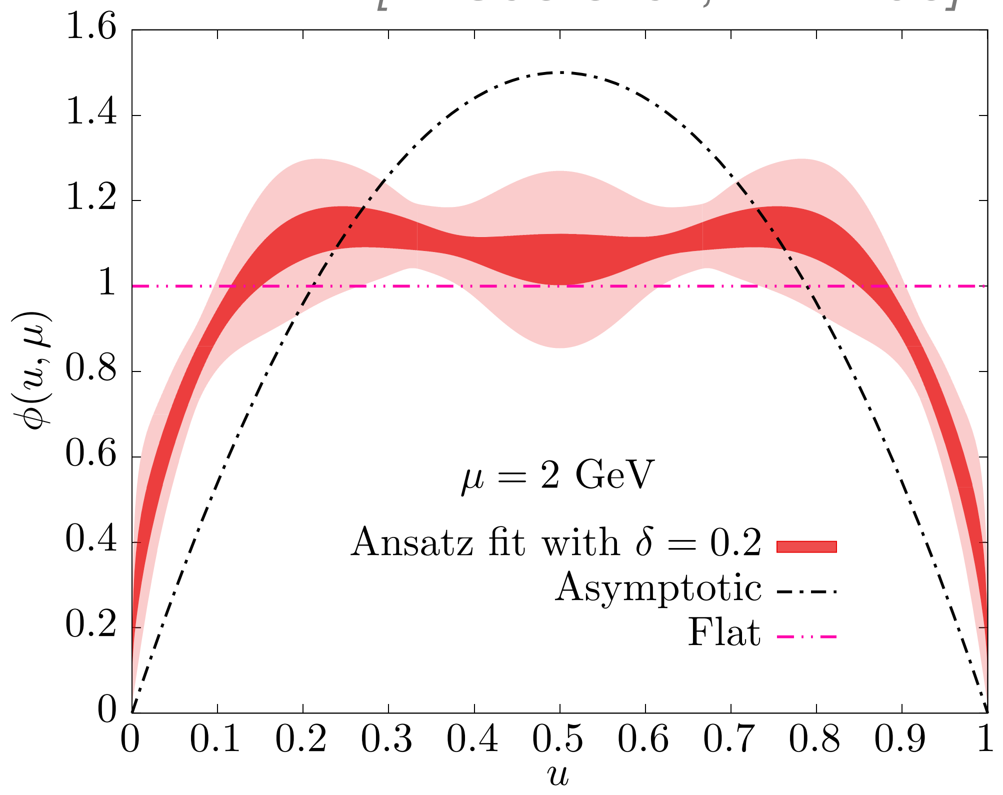
Q_0 plays an important rôle in PDF analyses. Can we improve our understading of the pion from data by varying the phenomenological starting scale? How to update the parameter-fixing of non-perturbative models?



[Ceccopieri et al, EPJC78]

From PDFs to DAs

[X. Gao et al., PRD106]



Wiggly and flat DA at $Q = 2$ GeV.

Second moment at $Q = 2$ GeV similar to NJL at Q_0 !

The Fantômas project beyond the pion

The metamorph functional form can be used for various correlation functions

- unpolarized proton PDF —extending CT's use of Bernstein basis
- twist-3 proton PDF — IFUNAM in charge
- nuclear PDF?
- helicity? transversity? — possibility to impose positivity constraints
- looking for more suggestions…

Everyone is welcome to use it and/or collaborate on it.

Fantômas will be included in the original xFitter framework

Conclusions

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

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

  Fantômas4QCD framework *[to appear very soon]*
  `metamorph` can be used to study many functions

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

  - Sea-gluon separation requires more data — a very interesting sector!
  - End-point behavior of valence pion distributions seems to follow a $(1-x)$ fall-off.

- Fantômas code can be used in inverse problems for other correlation functions — transversity, nuclear PDFs,…
- positivity constraints can be implemented, too


FANTOMAS
LA AMENAZA ELEGANTE