Pion structure in global analyses:
exploring parametrization uncertainties with Fantômas4QCD

Parton Distribution Functions at a Crossroad

ECT*

19/09/23

CONACYT
Consejo Nacional de Ciencia y Tecnología

Instituto de Física
UNAM

Aurore Courtoy

Instituto de Física
Universidad Nacional Autónoma de México (UNAM)

dgapa
Dirección General de Asuntos del Personal Académico

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]*

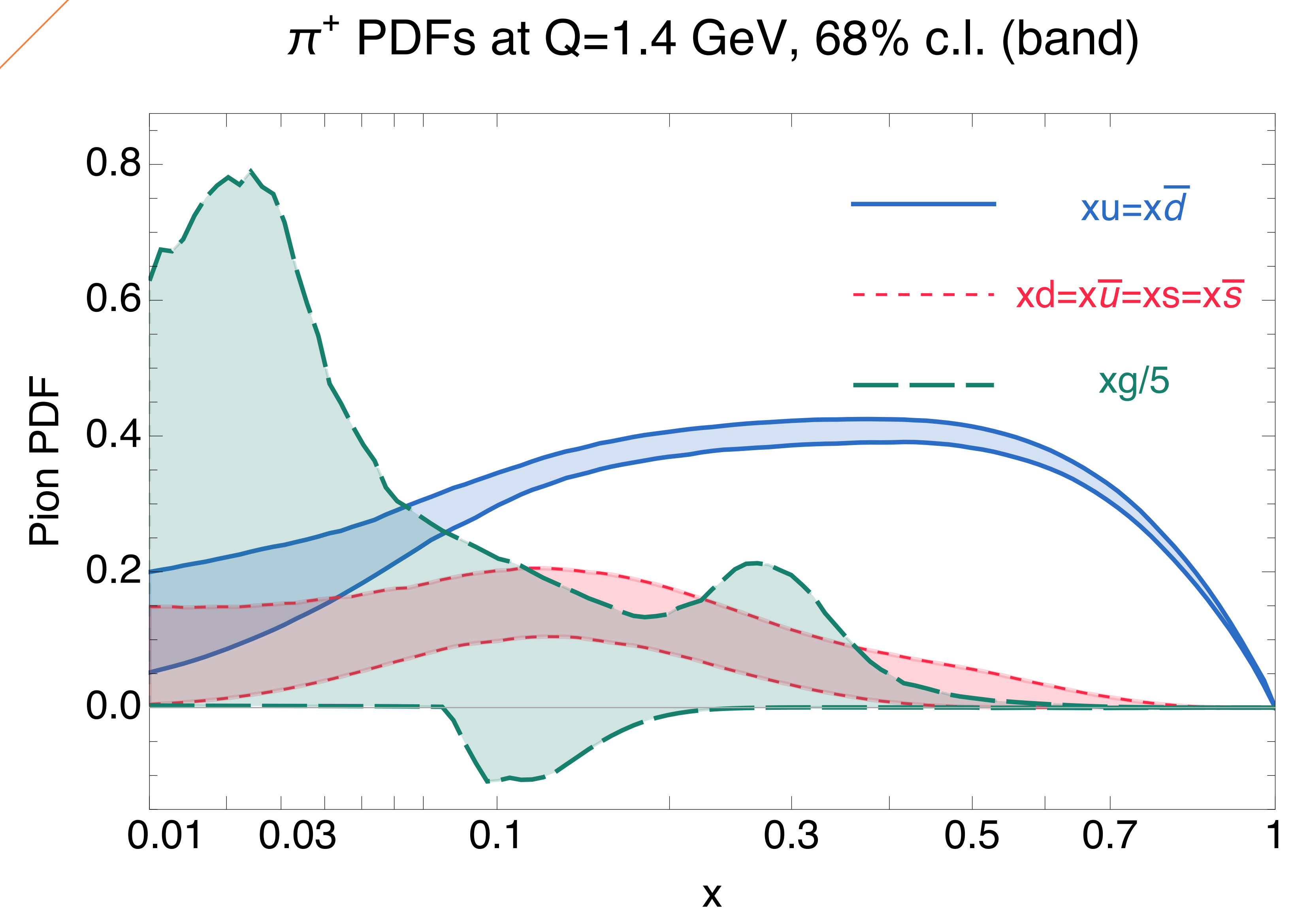
A.C. & Nadolsky

"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]



Why study the pion?

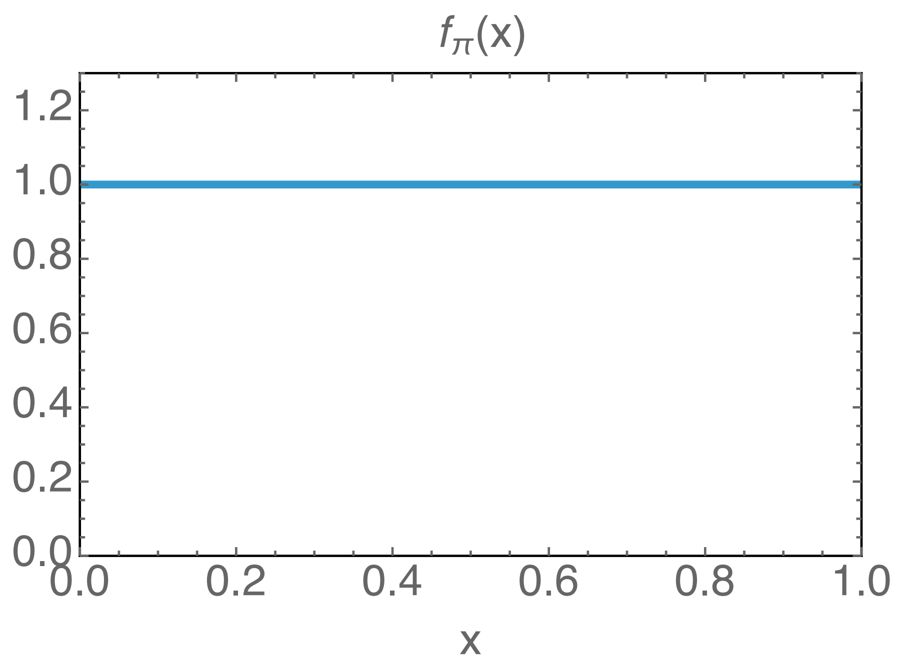
➡ xFitter's framework set up the pion PDF analysis— https://www.xfitter.org/xFitter/
➡ less data *wrt* proton, still at NLO accuracy
➡ recent "come back" thanks to increased fitting activity in the nuclear community —theory and experiment-wise

⇨ 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.

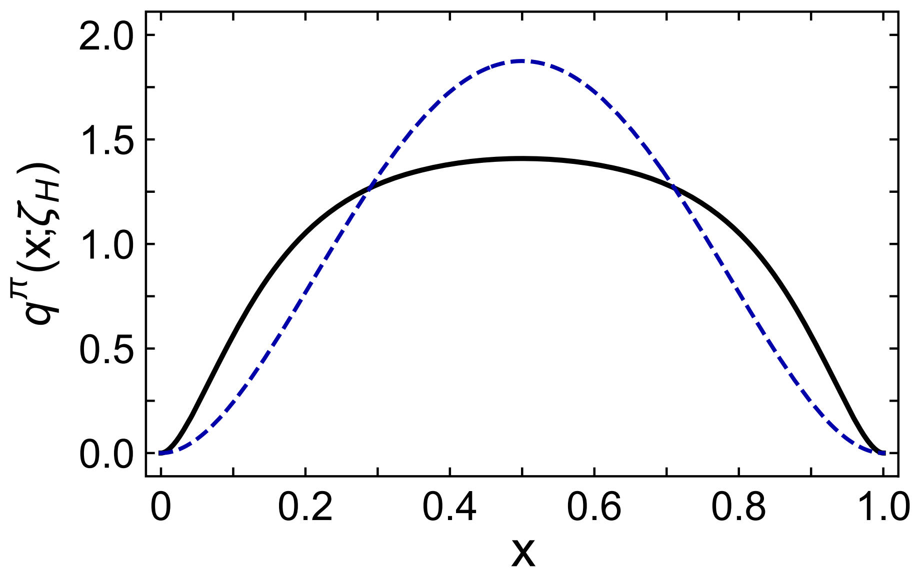
Why study the pion?

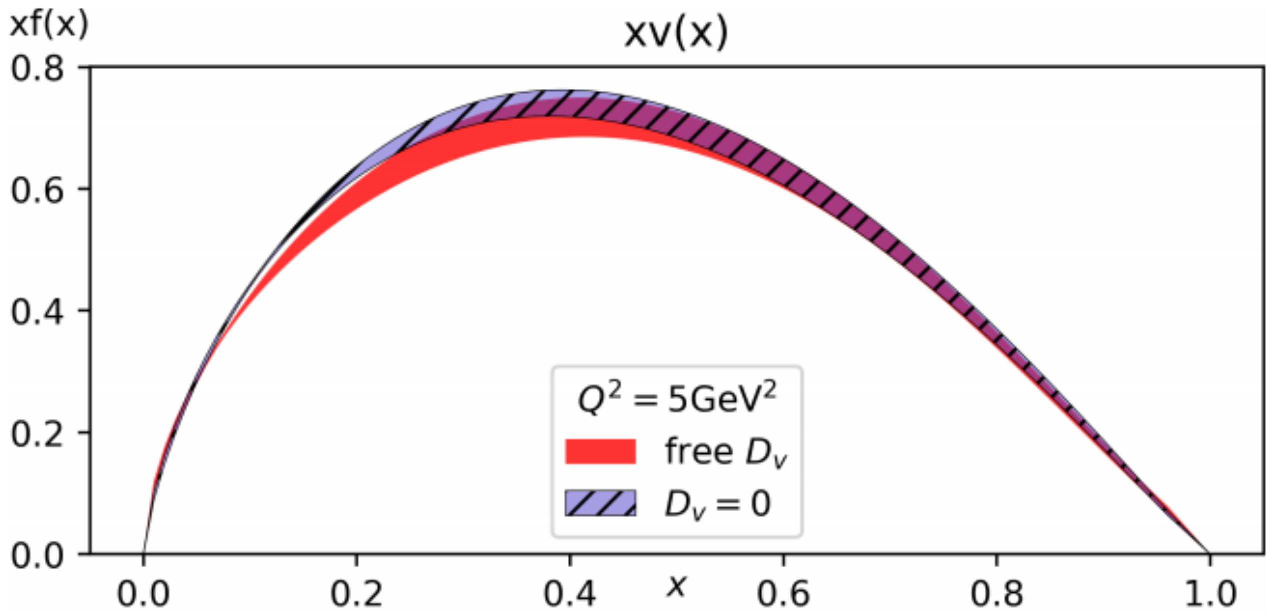
- ➡ xFitter's framework set up the pion PDF analysis— https://www.xfitter.org/xFitter/
- ➡ less data *wrt* proton, still at NLO accuracy
- ➡ recent "come back" thanks to increased fitting activity in the nuclear community —theory and experiment-wise

⇨ 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.

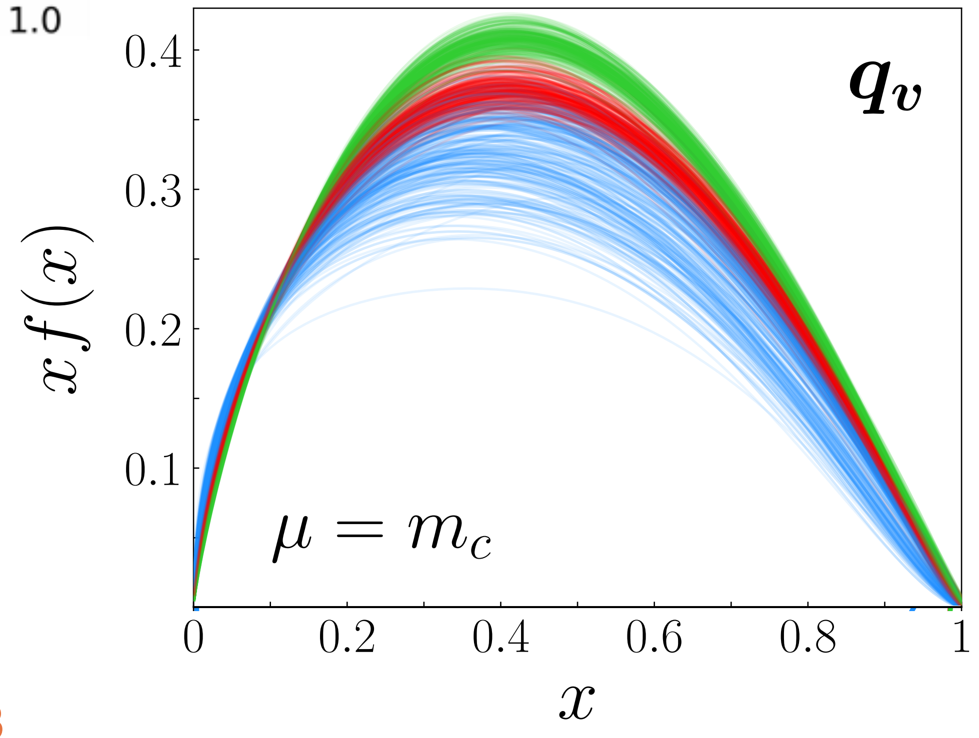


e.g. Nambu—Jona-Lasinio model, Schwinger-Dyson approaches, …





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



Fantômas4QCD

In any inference about primordial dynamics, dependence on the PDF functional form must be fully evaluated.

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.

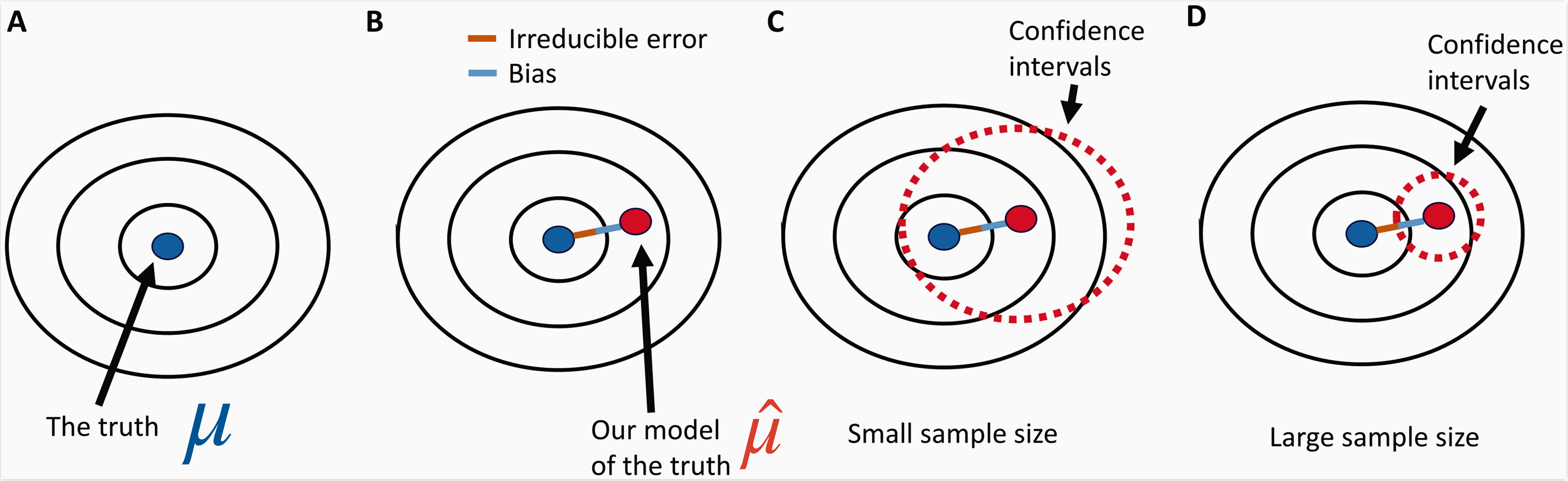
This code facilitates unbiased estimates of parametrization dependence.

CONACYT
Consejo Nacional de Ciencia y Tecnología

dgapa
Dirección General de Asuntos del Personal Académico

Inter-American Network of Networks of QCD challenges

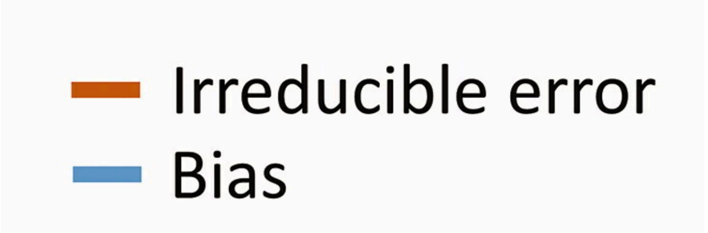
Sampling bias and big-data paradox



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.*

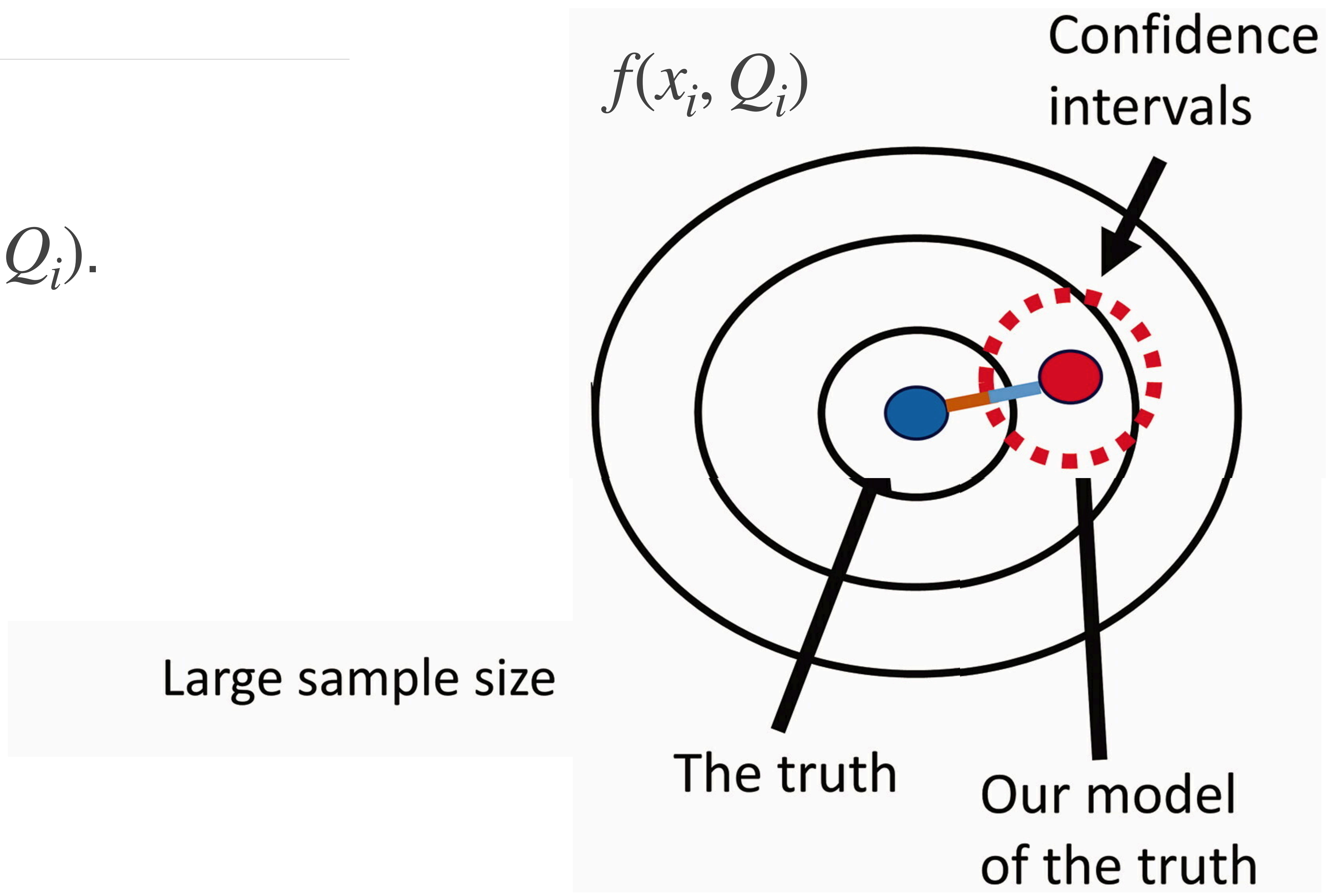


Xiao-Li Meng
The Annals of Applied Statistics
Vol. 12 (2018), p. 685

Physics phenomenology and accuracy

Suppose we know the true parton distribution function at given (x_i, Q_i).

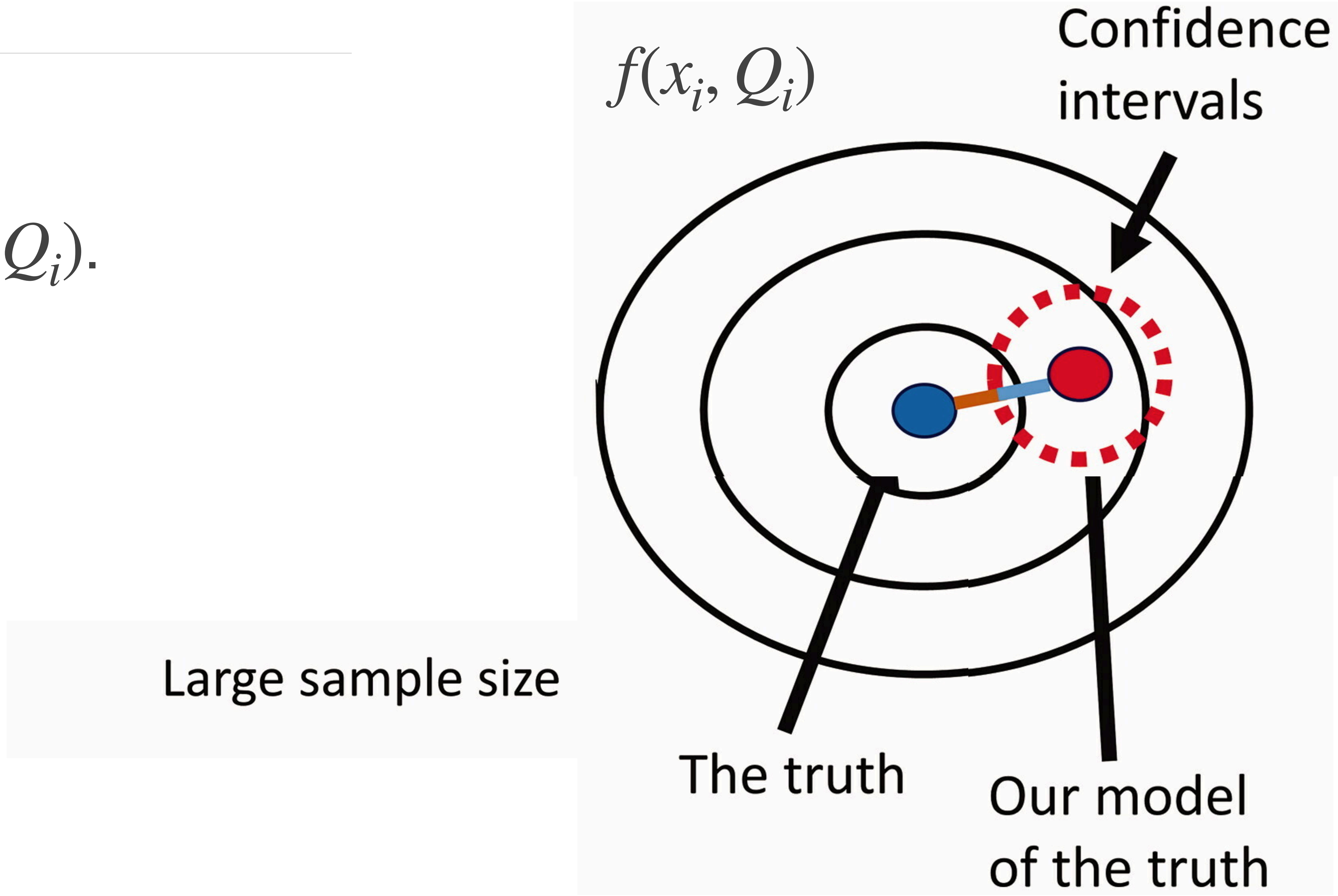
We want our determination from global analysis to encompass it.



Physics phenomenology and accuracy

Suppose we know the true parton distribution function at given (x_i, Q_i).

We want our determination from global analysis to encompass it.



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

depends on the sampling algorithm

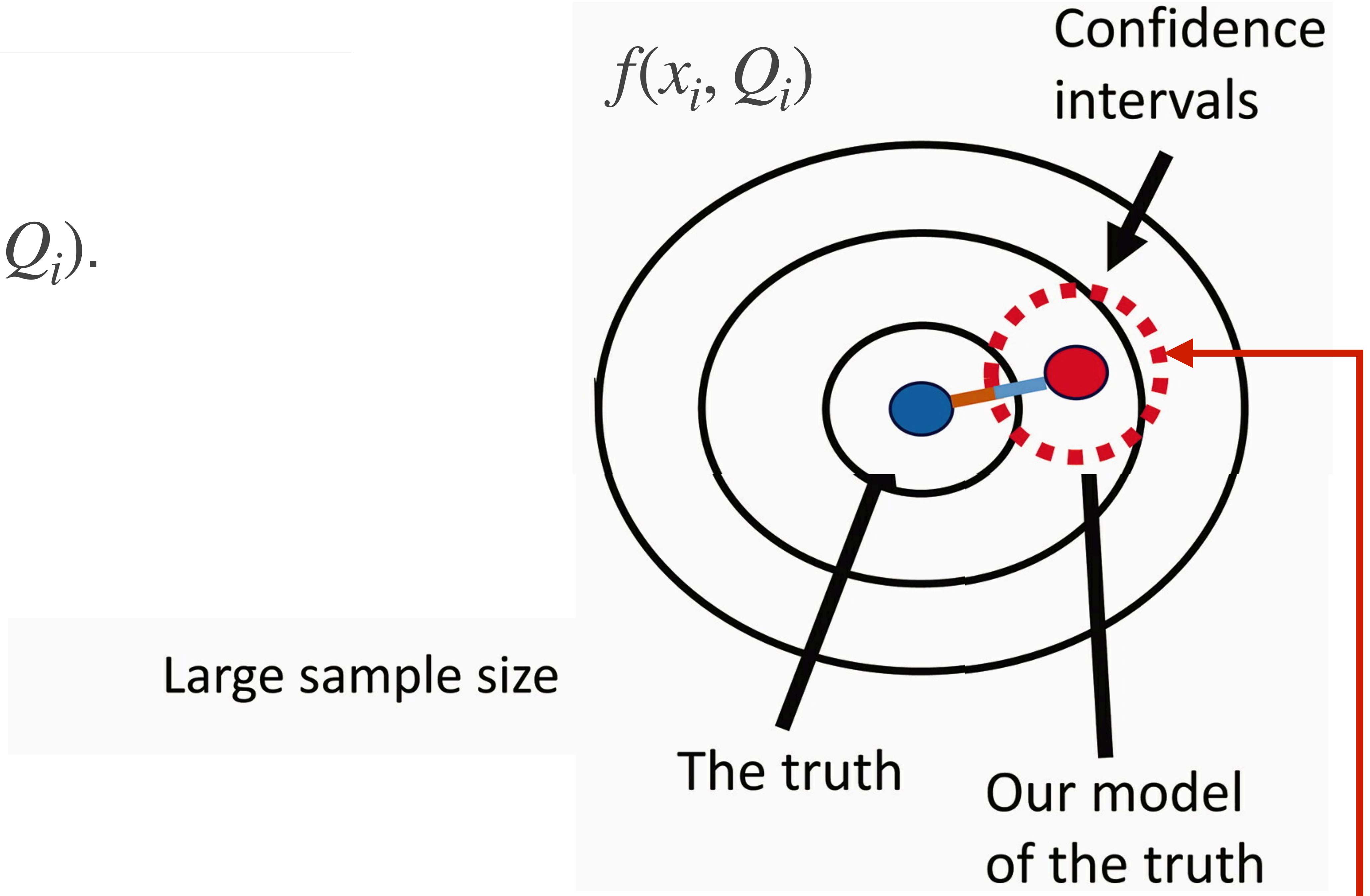
Irreducible error $\equiv$ statistical model, quality of data,…

Bias

Physics phenomenology and accuracy

Suppose we know the true parton distribution function at given (x_i, Q_i).

We want our determination from global analysis to encompass it.



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

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

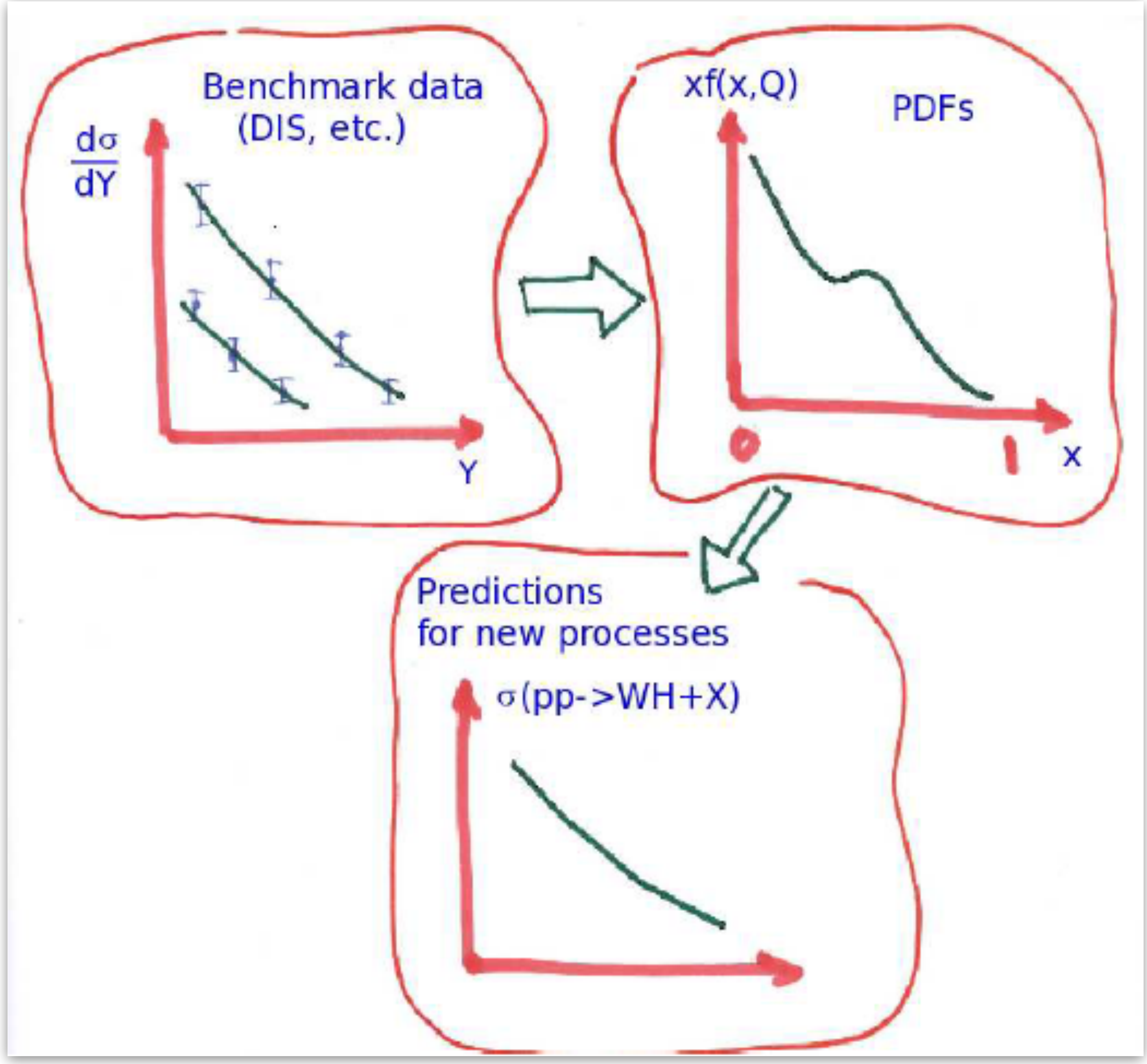
The shape of parton distributions

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



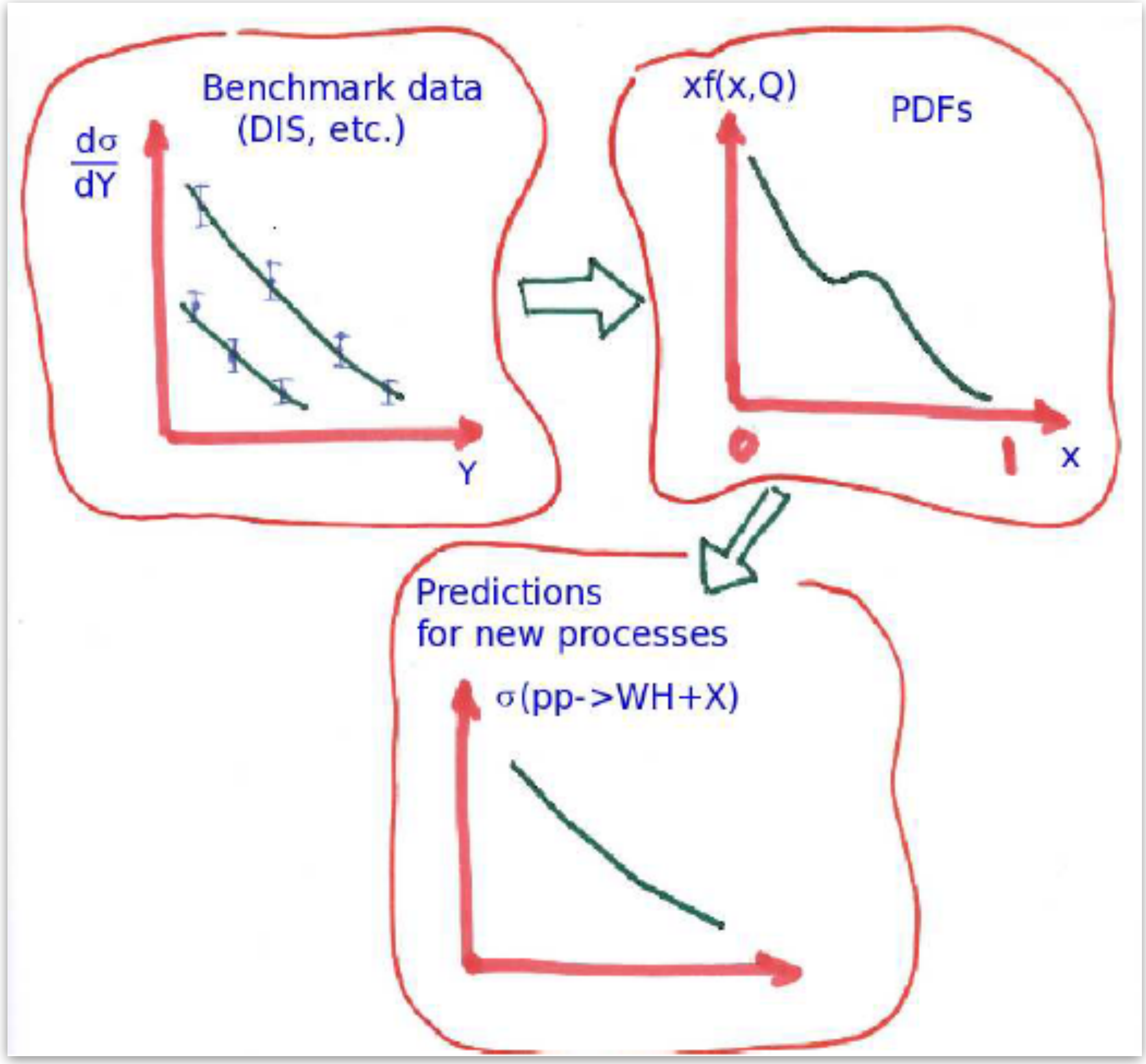
The shape of parton distributions

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

The shape of parton distributions

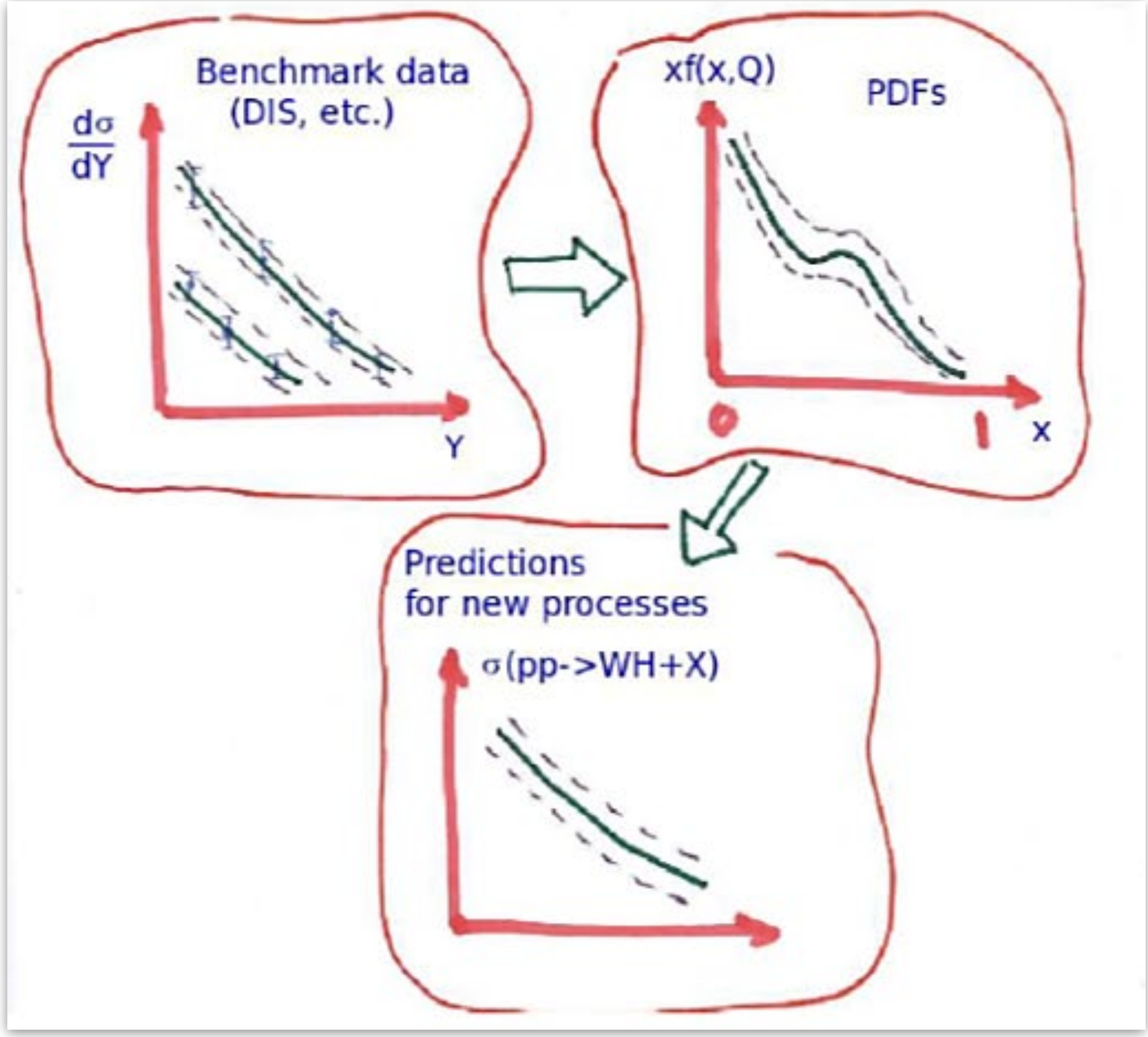
Low-energy QCD dynamics, encapsulated in PDFs, are learned from experimental data.

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

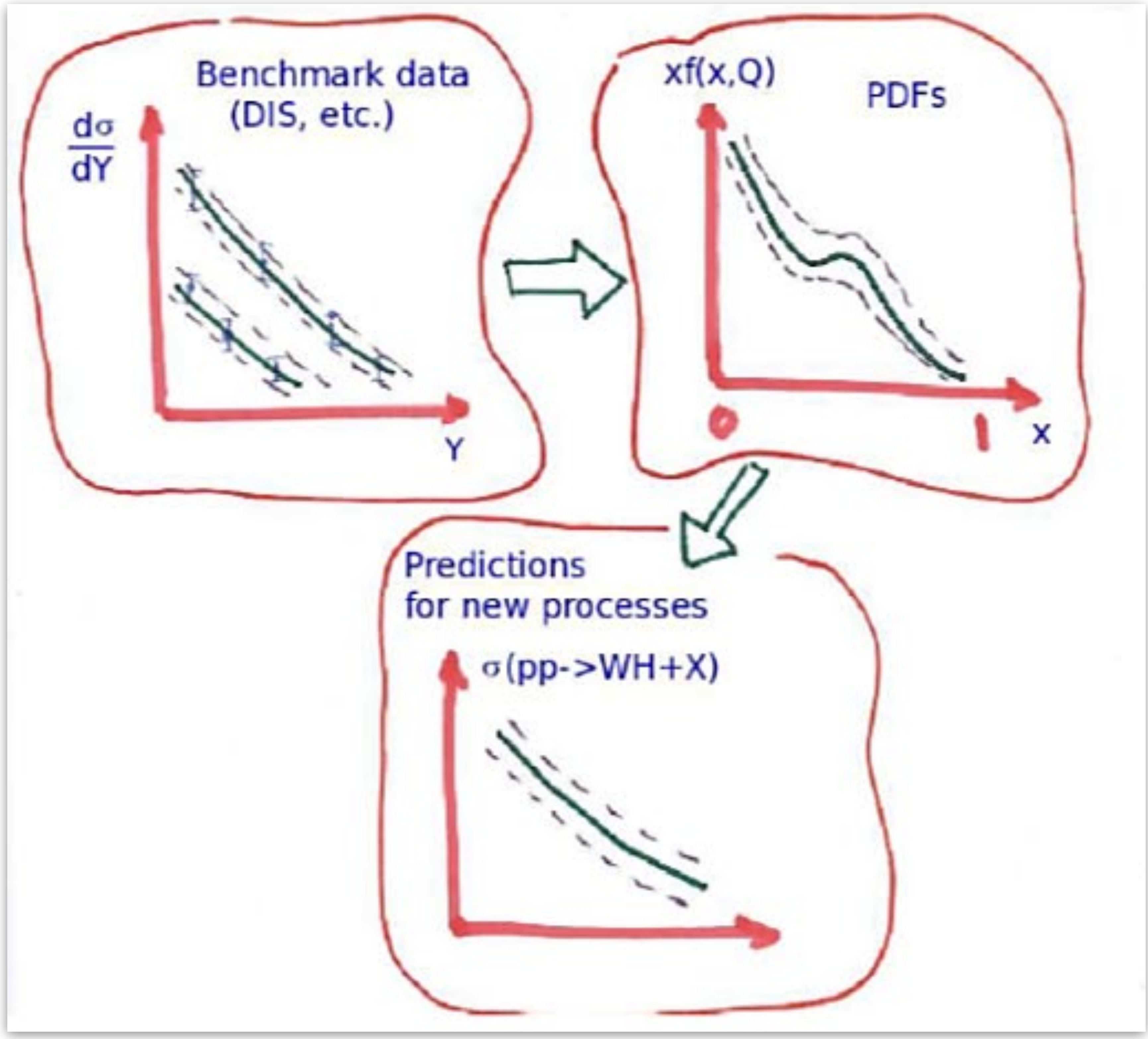
Low-energy QCD dynamics, encapsulated in PDFs, are learned from experimental data.

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

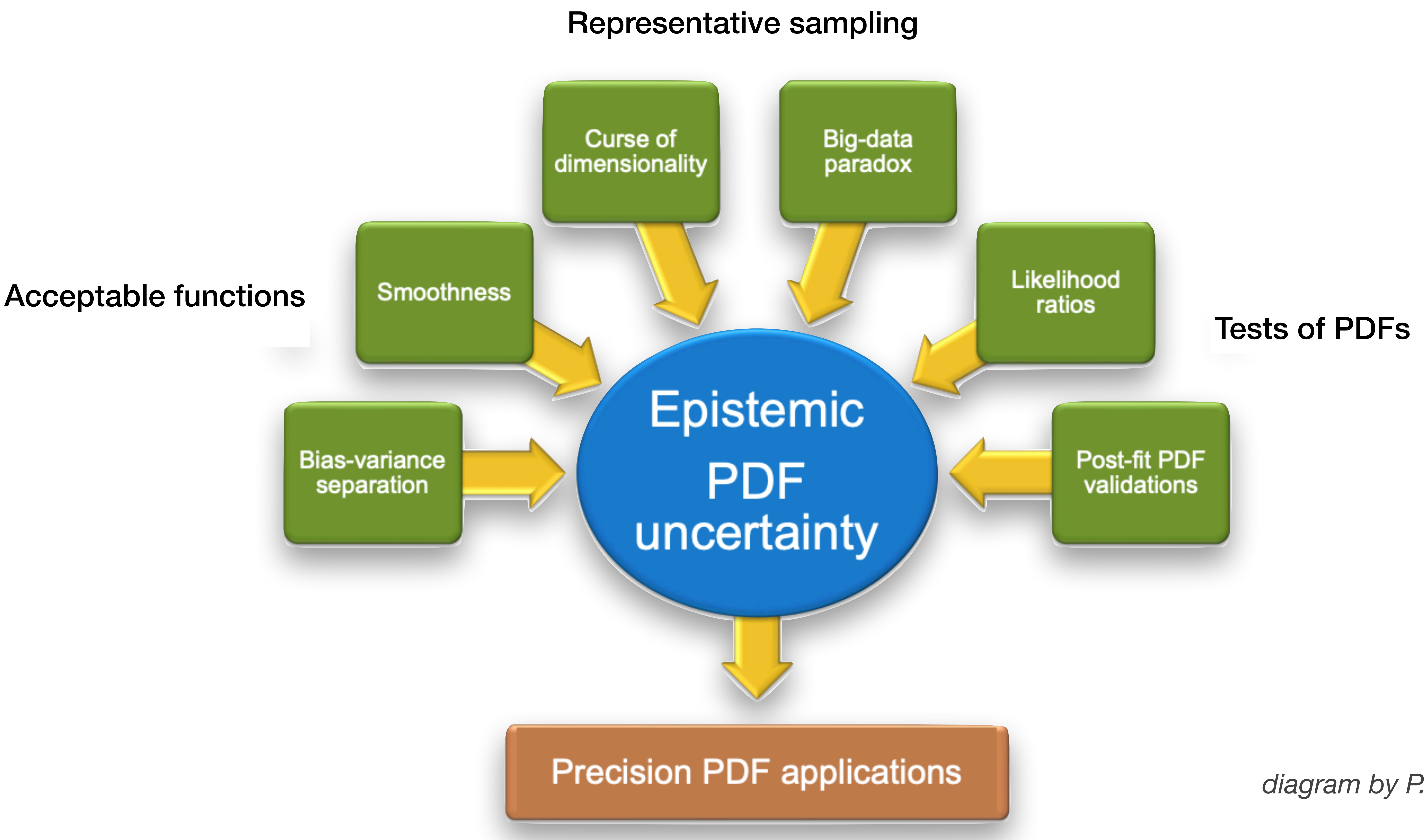
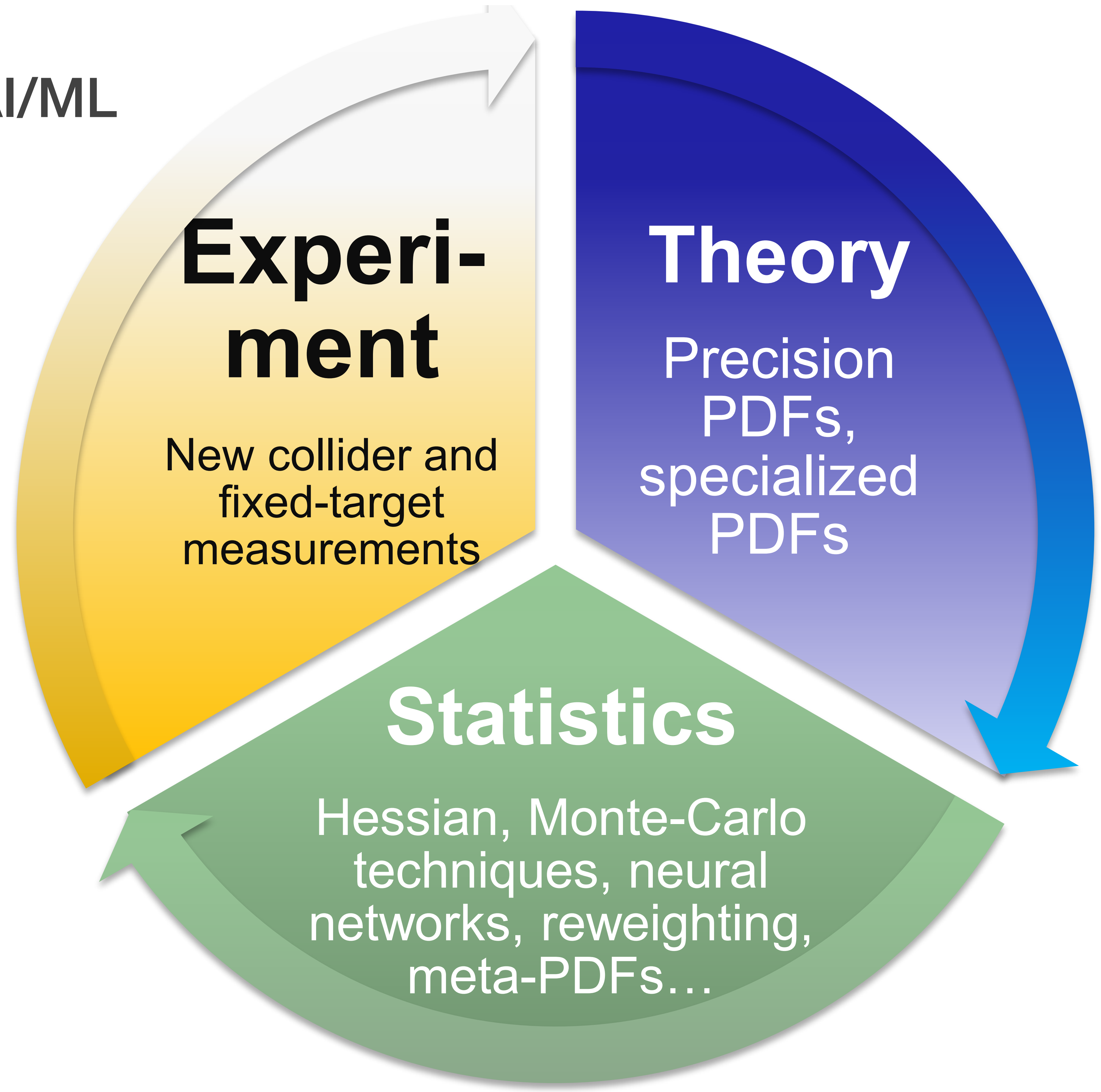


diagram by P. Nadolsky [DIS2023]

Epistemic uncertainties

Components of a global QCD fit — theorists working with data and statistics

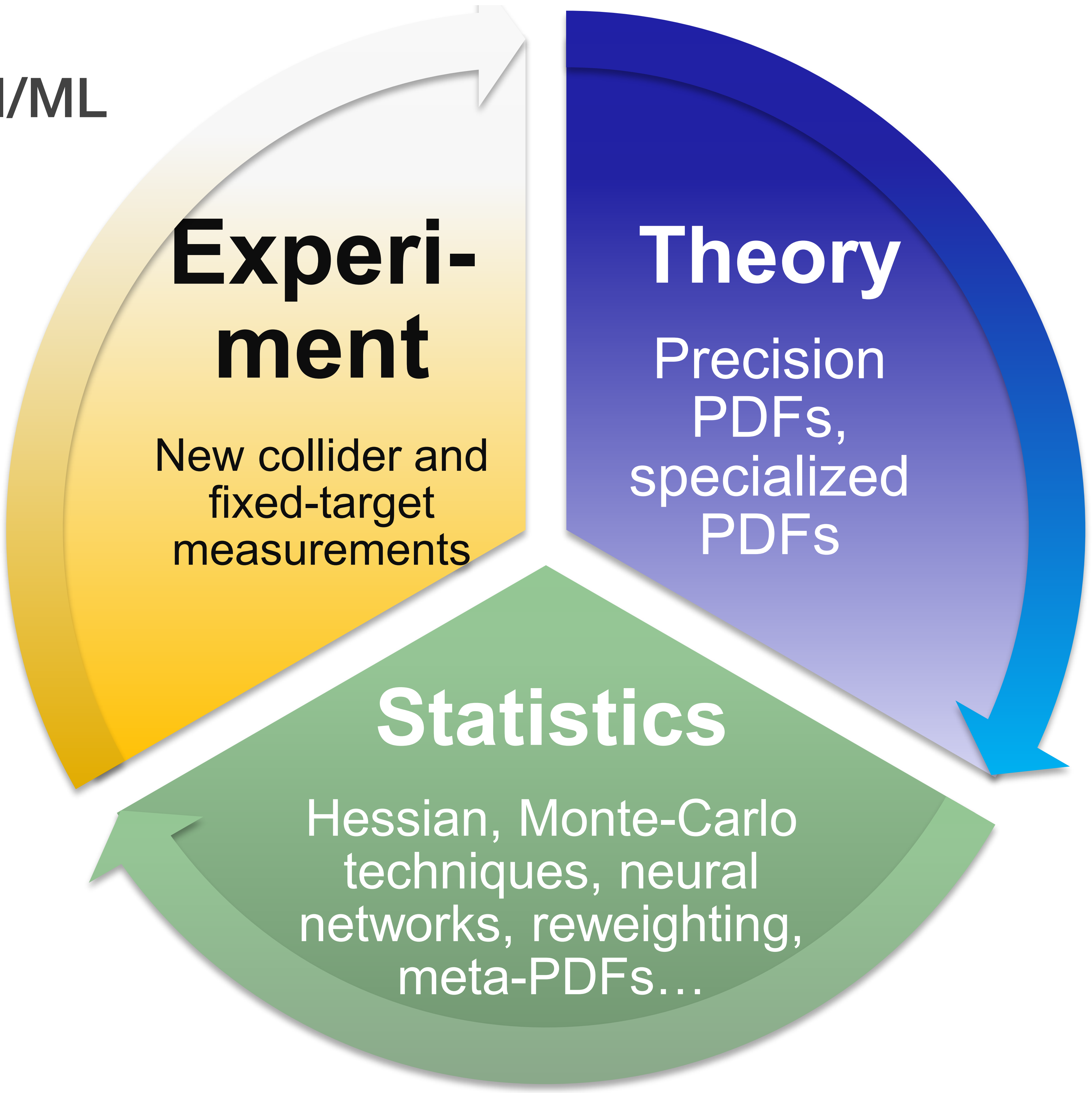
Global analyses with years of expertise — now evolving in the context of AI/ML



Epistemic uncertainties

Components of a global QCD fit — theorists working with data and statistics

Global analyses with years of expertise — now evolving in the context of AI/ML



Methodological choices are reflected in the epistemic uncertainty, including biases from sampling.

Priors, including choice of functional form or Bayesian *priors*, influence the sampling algorithm.

Representative sampling can ultimately be used to optimize its contribution, e.g. through the study of largest effective dimensions.

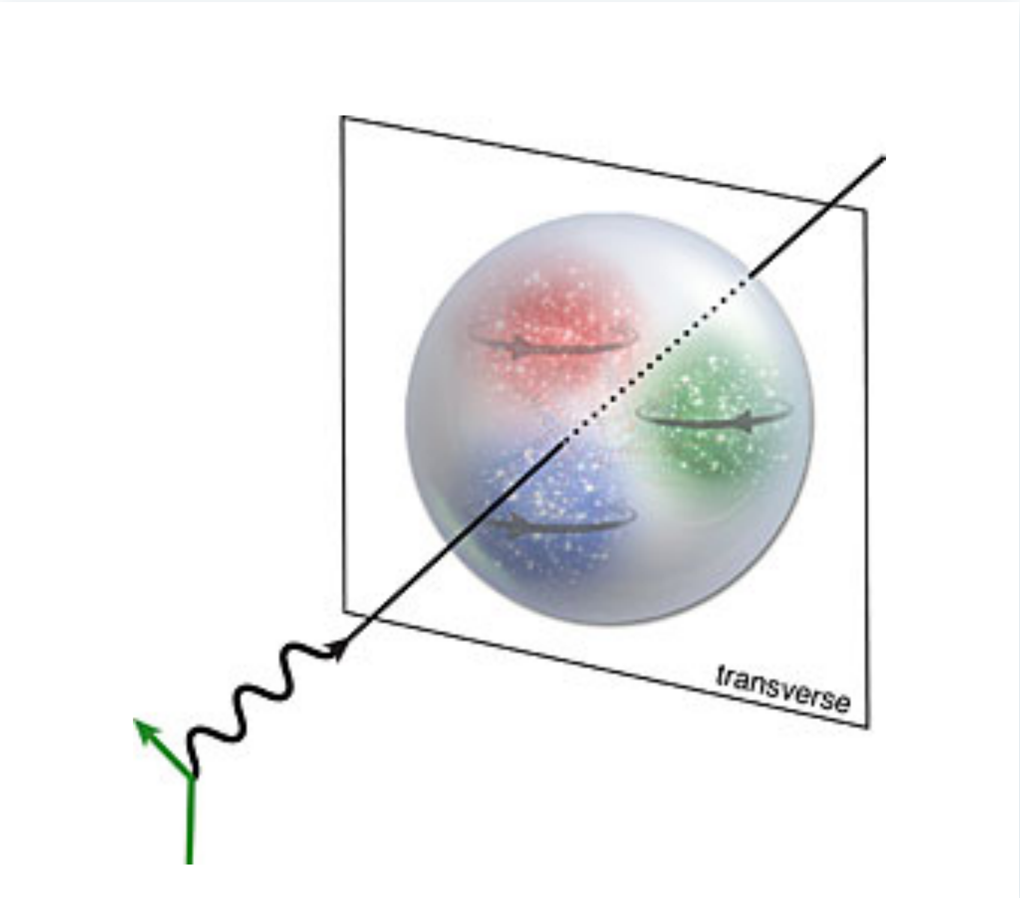
PDFs in QCD phenomenology

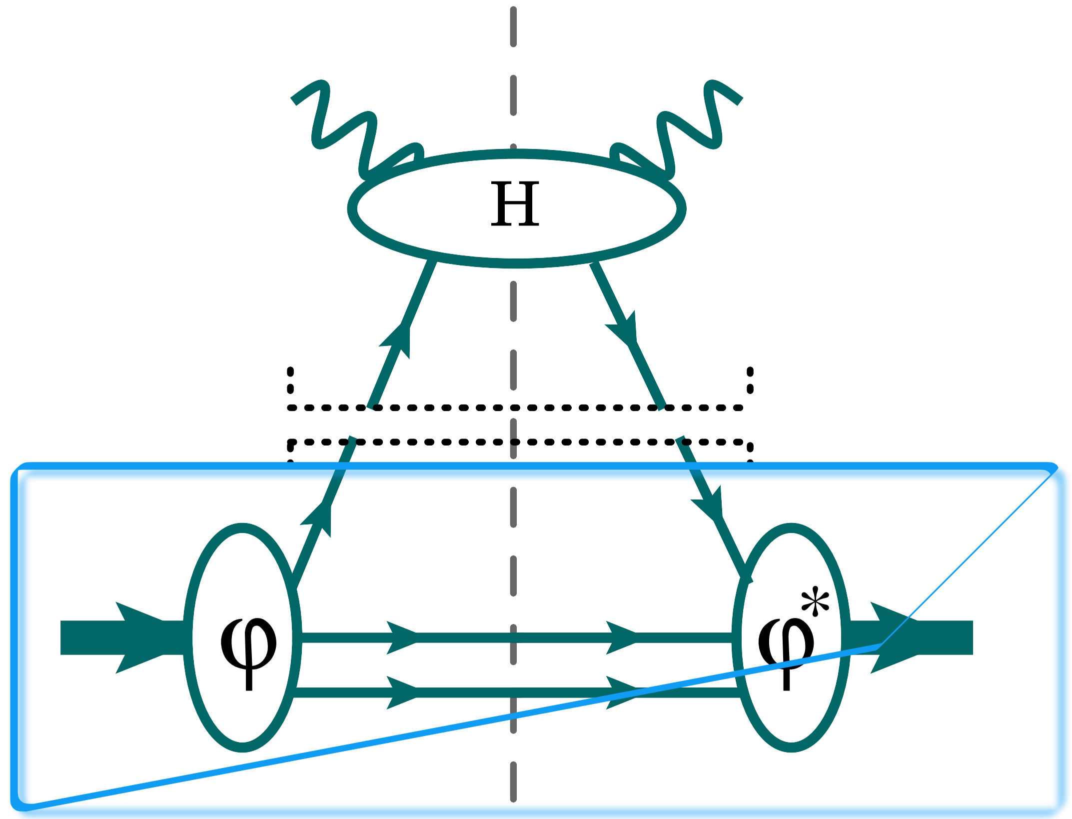
In the realm of QCD…

The two regimes of QCD demand that non-perturbative objects are accessed through *factorization theorems*,

i.e., theory is expressed through a convolution of hard and soft part to which corrections are added.

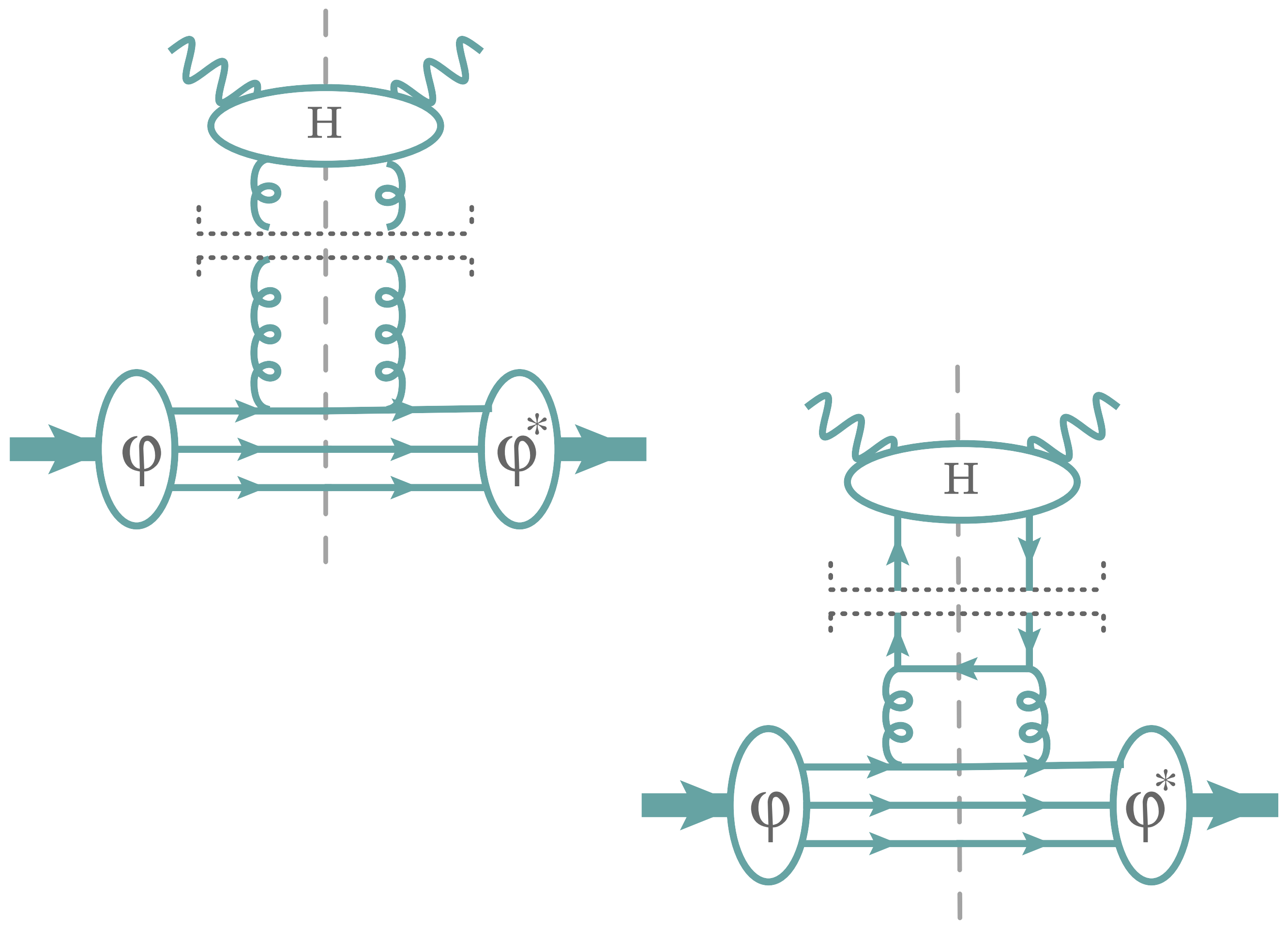
⇒ the data for observables is reproduced by the PDF times the hard contribution, and corrections.





$f_1(x)$ is the unpolarized PDF

Illustration in the $\overline{MS}$ scheme

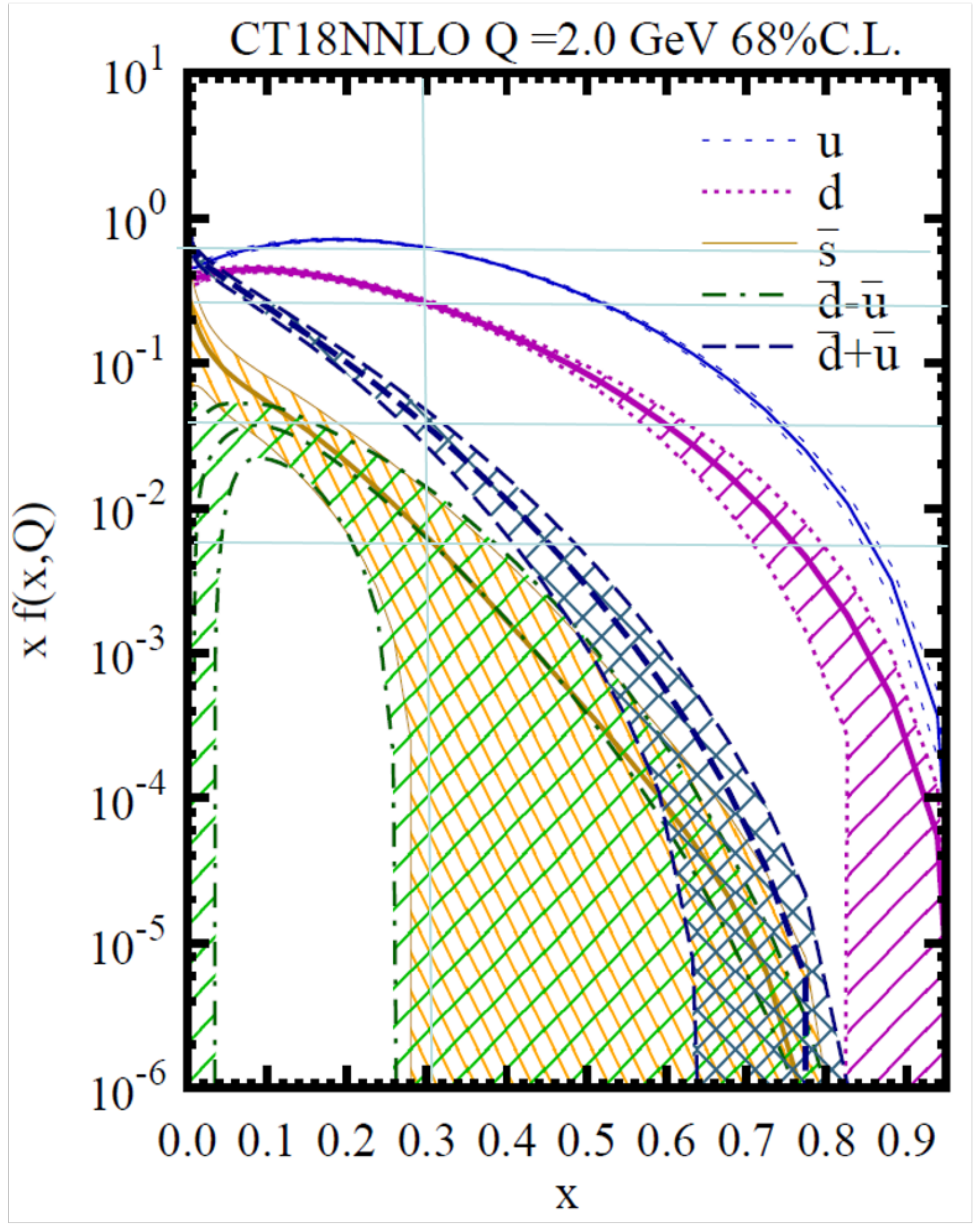


μ_0^2 $\mu^2 \sim Q^2$

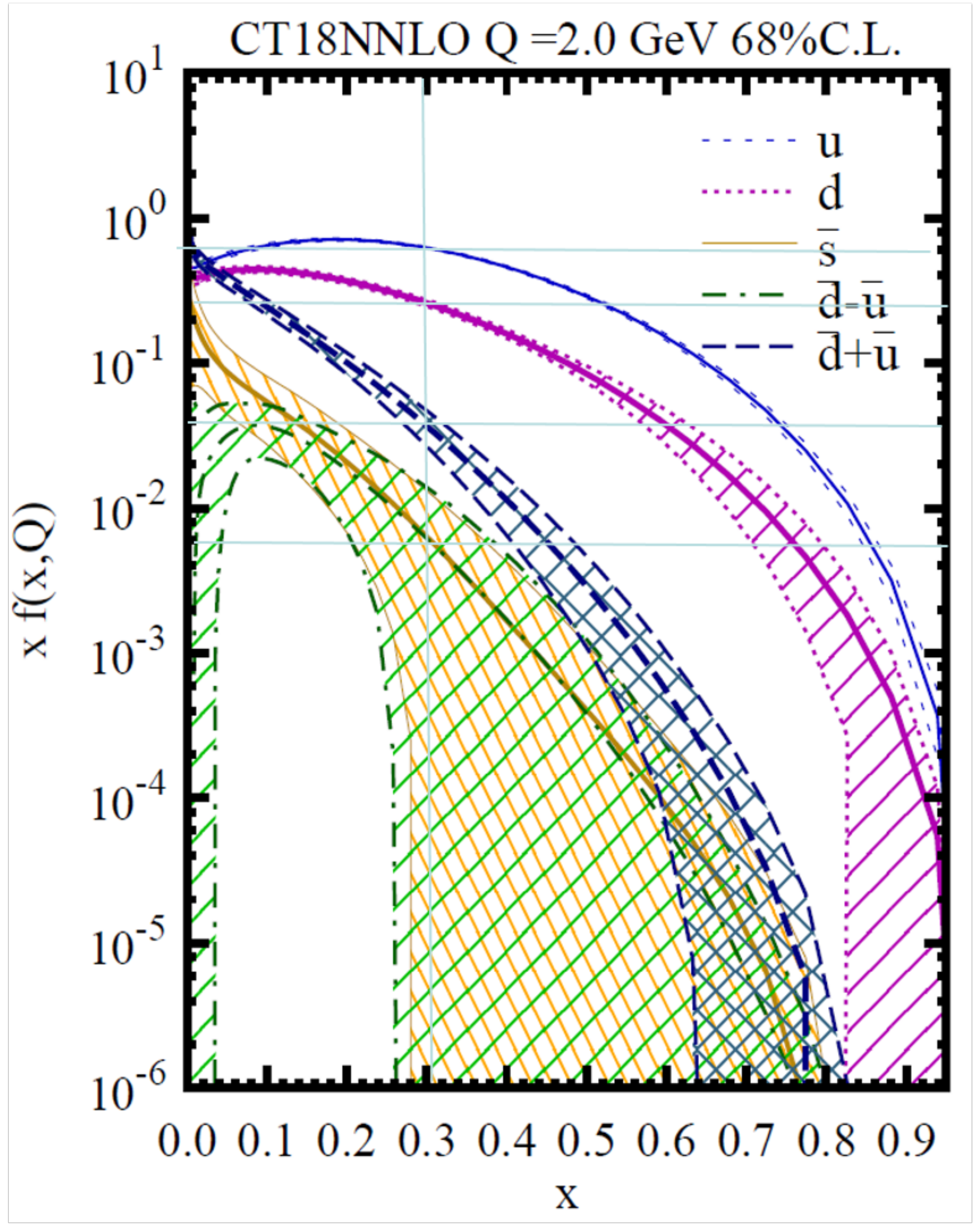
Proton PDF from the CT analysis

Uncertainties in CT18 and beyond:

two tier criteria to estimate uncertainties, inclusive of various sampling inaccuracy sources.



Proton PDF from the CT analysis



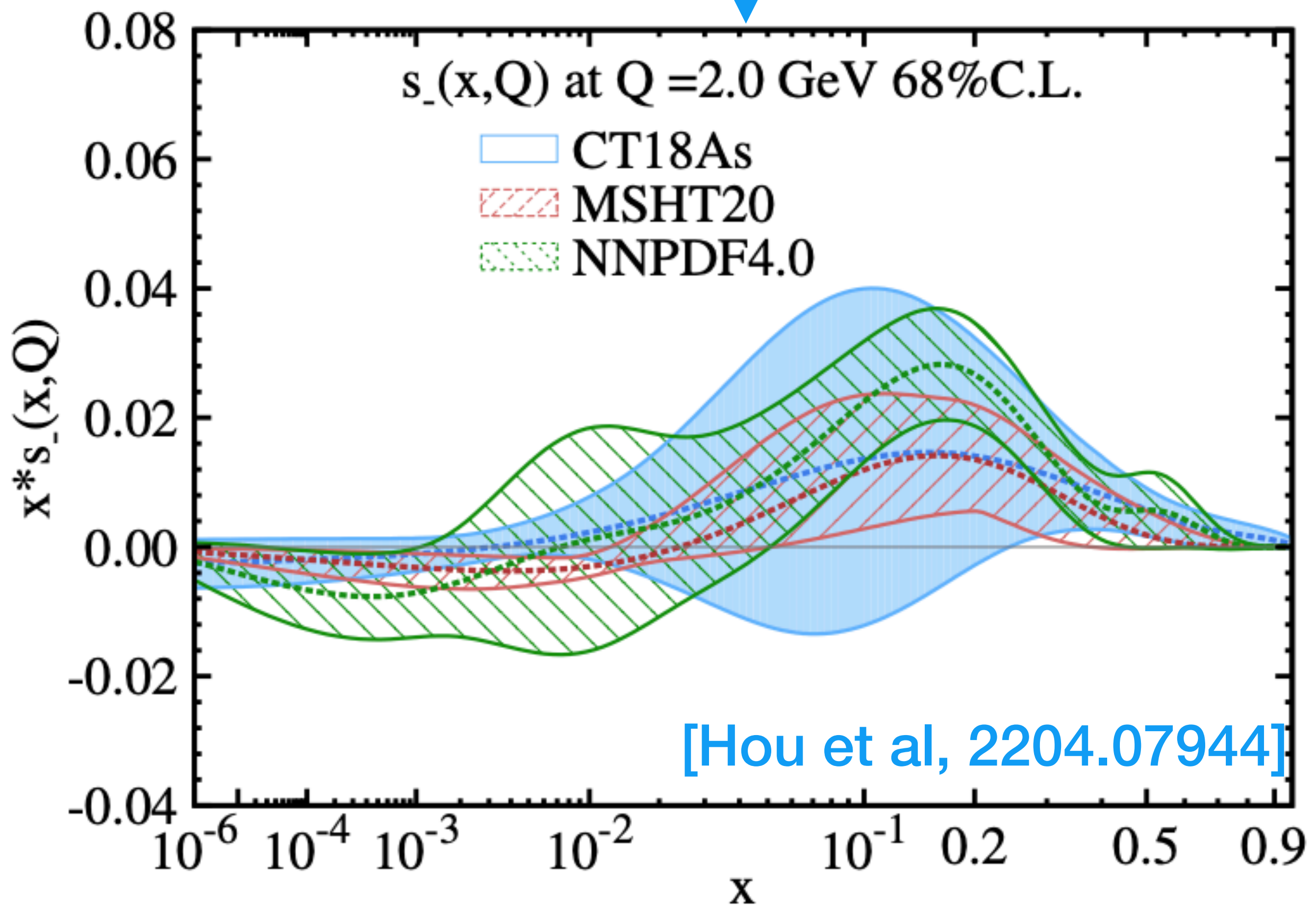
Uncertainties in CT18 and beyond:

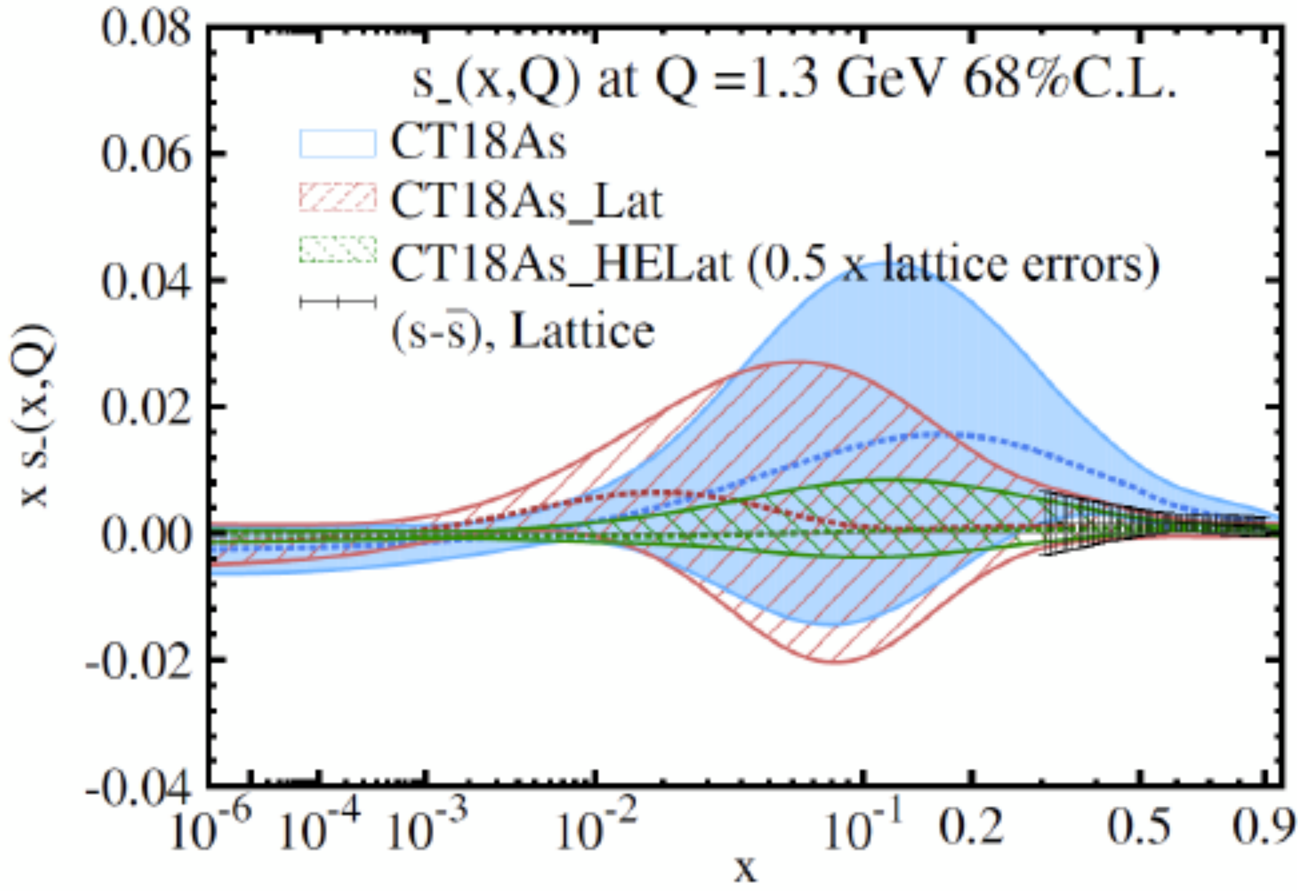
two tier criteria to estimate uncertainties, inclusive of various sampling inaccuracy sources.

Example, strange asymmetry

⇨ Initially zero in CT18 — $s = \bar{s}$,

⇨ Methodology allows to relax that constraint and, also, to incorporate lattice data in $s_- = s - \bar{s}$.





Data does not seem to constrain the strange asymmetry for now:

CT accounts for that thanks to the two tier uncertainty estimate.

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.

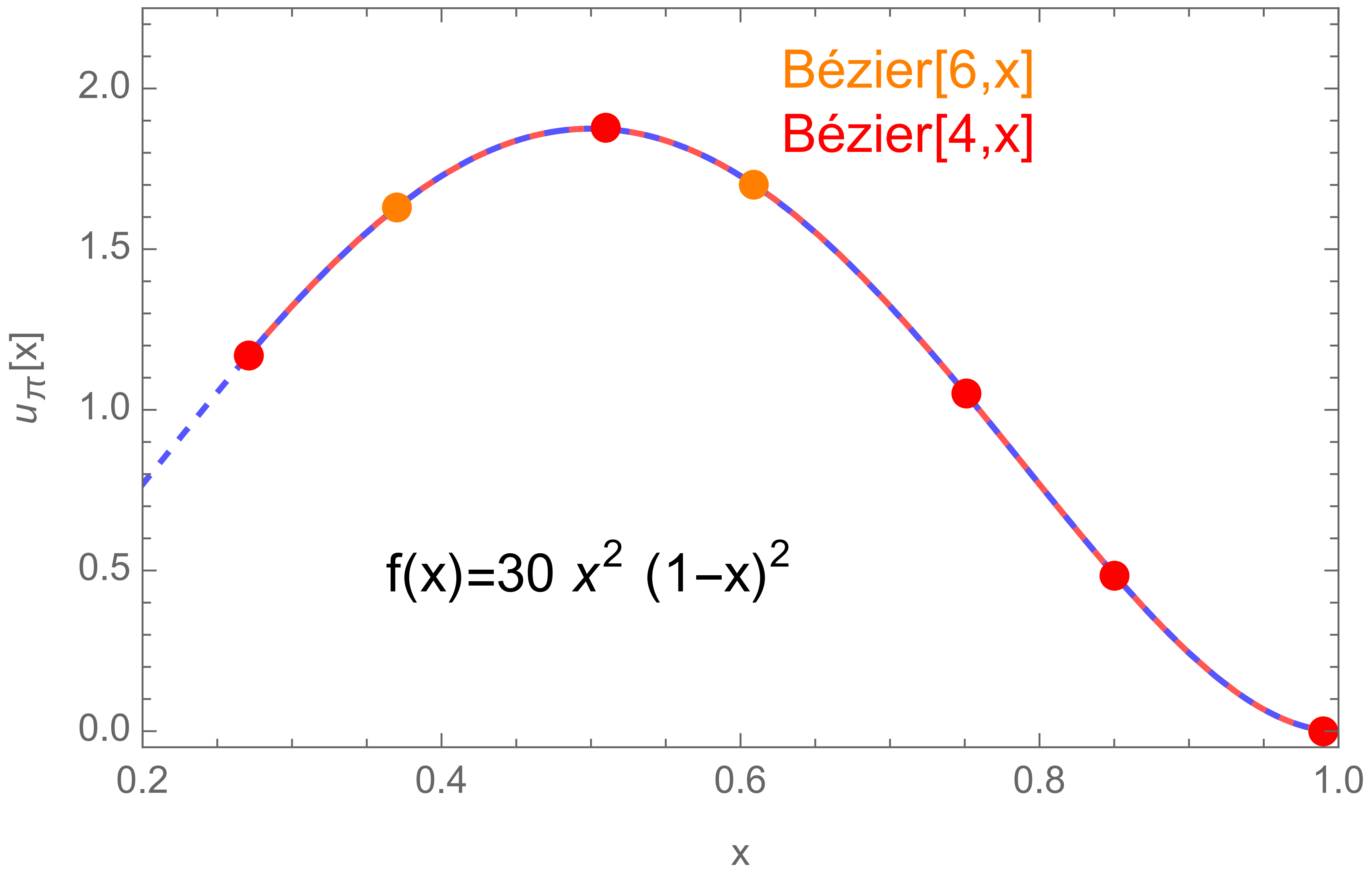
$$\underline{P} = \underline{\underline{T}} \cdot \underline{\underline{M}} \cdot \underline{C}$$

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.

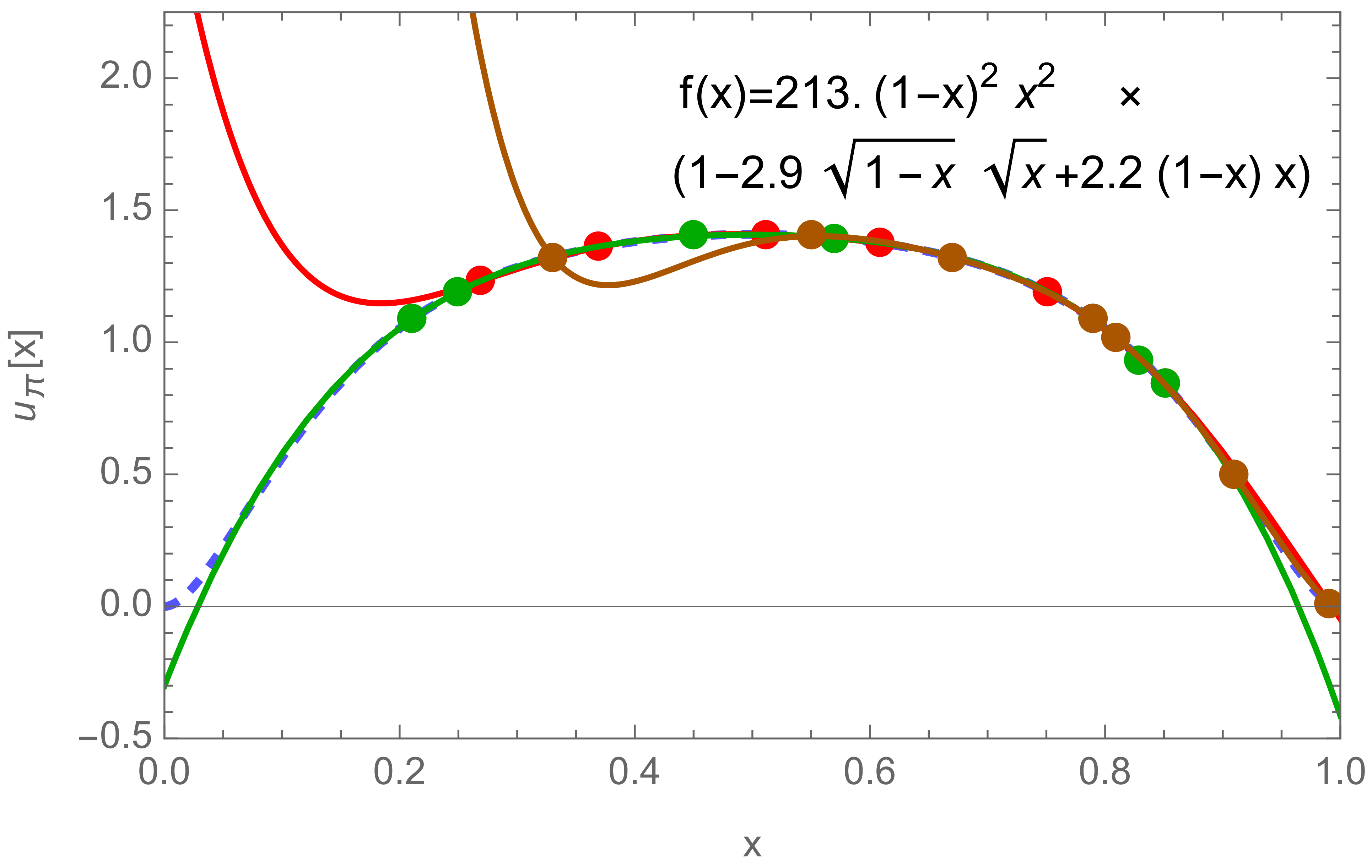


Bézier-curve methodology for global analyses

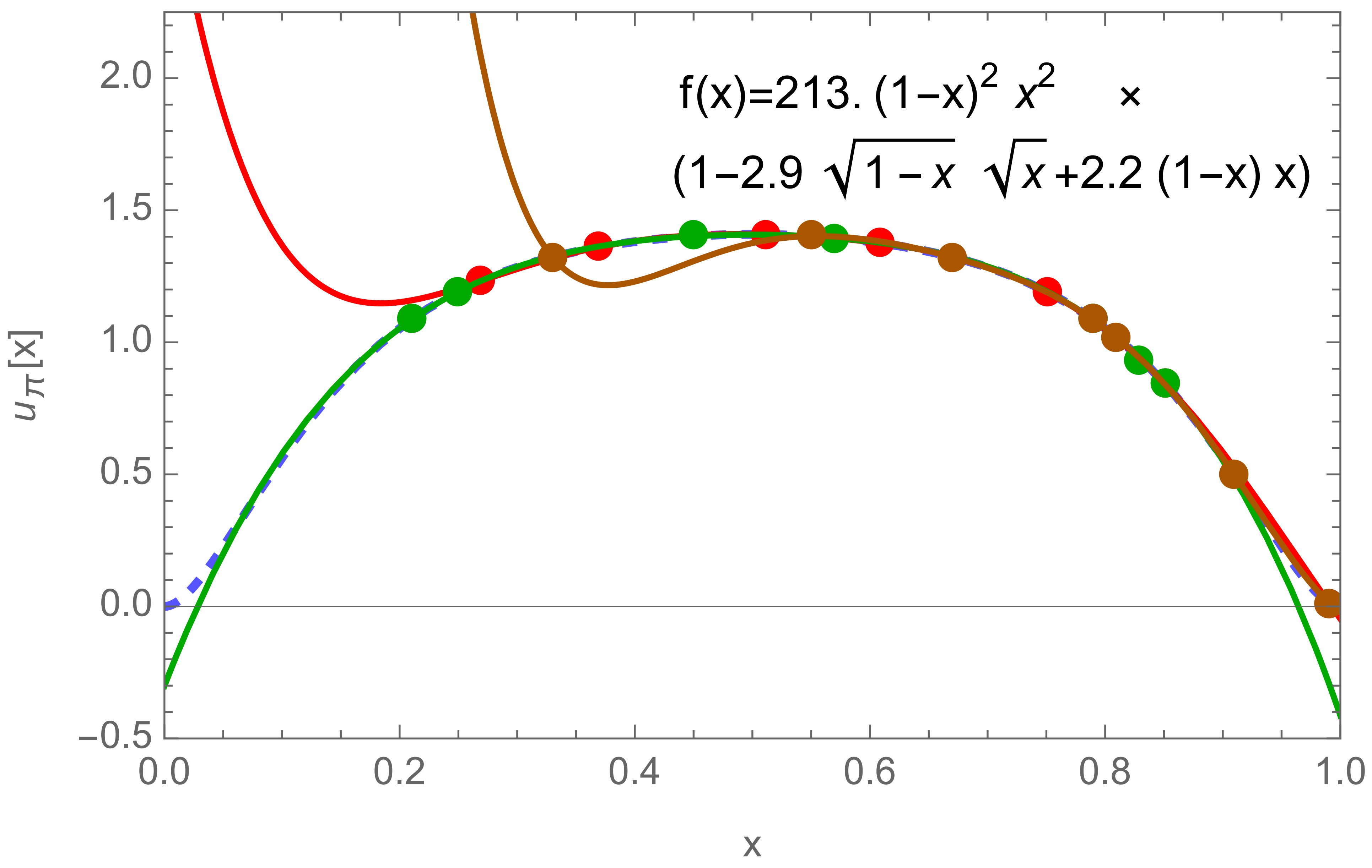
<u>Reconstruction of a more complex parametrization</u>

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

⇨ lowest powers of the expansion cannot be meaningfully reconstructed.



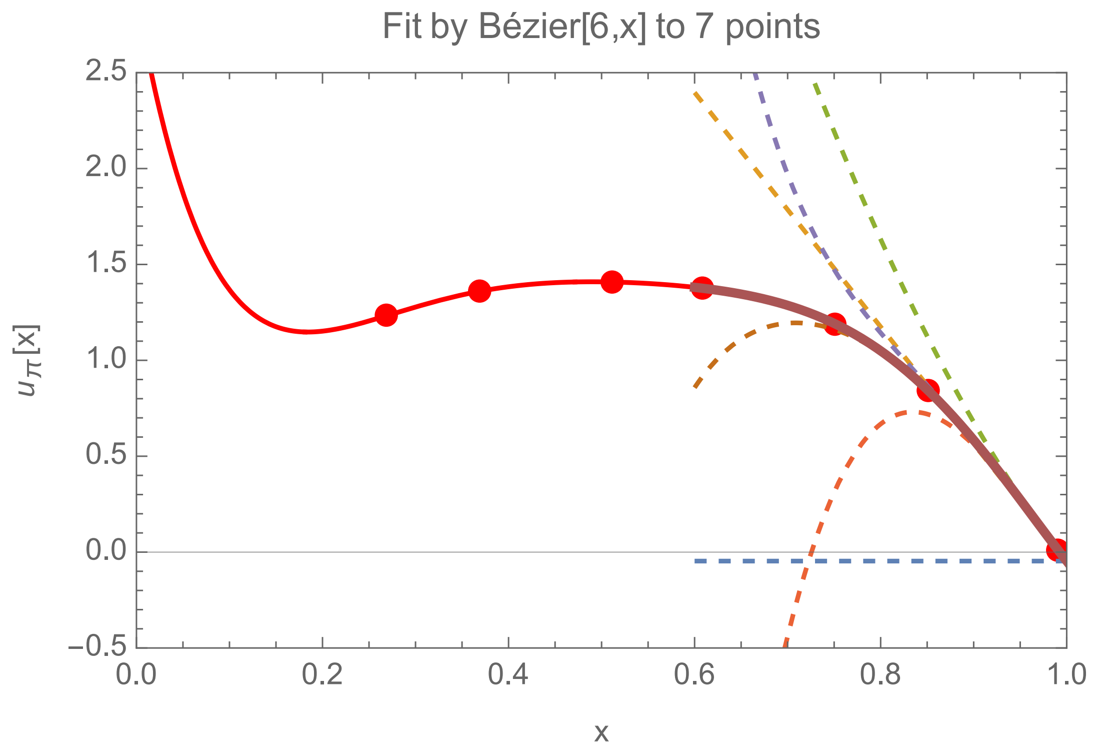
Bézier-curve methodology for global analyses

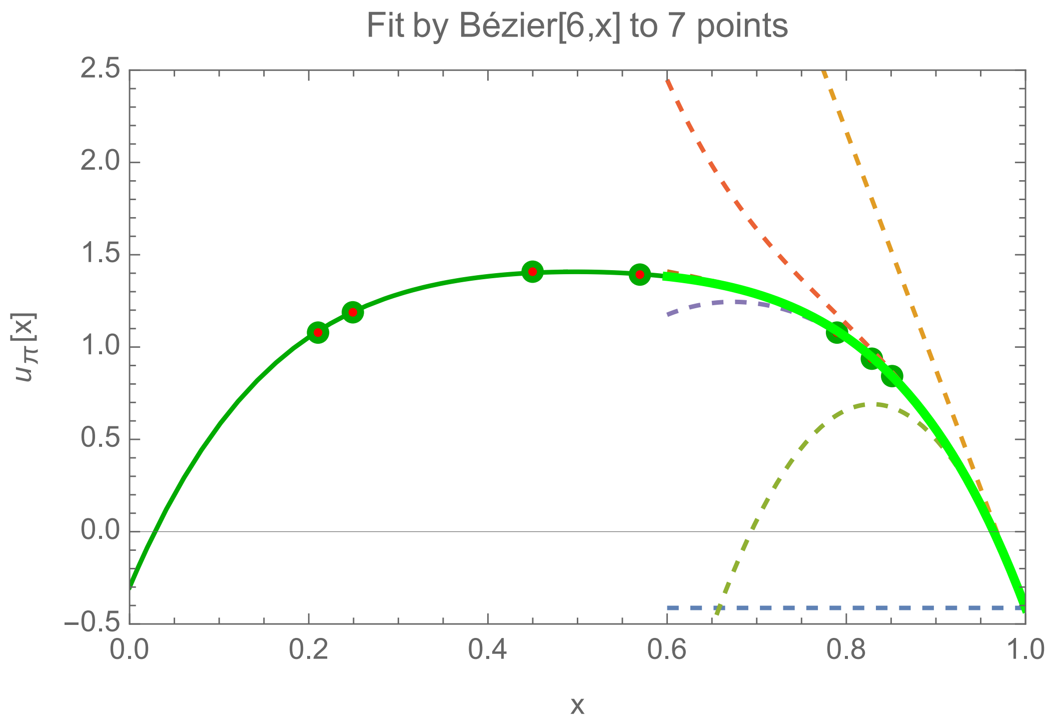


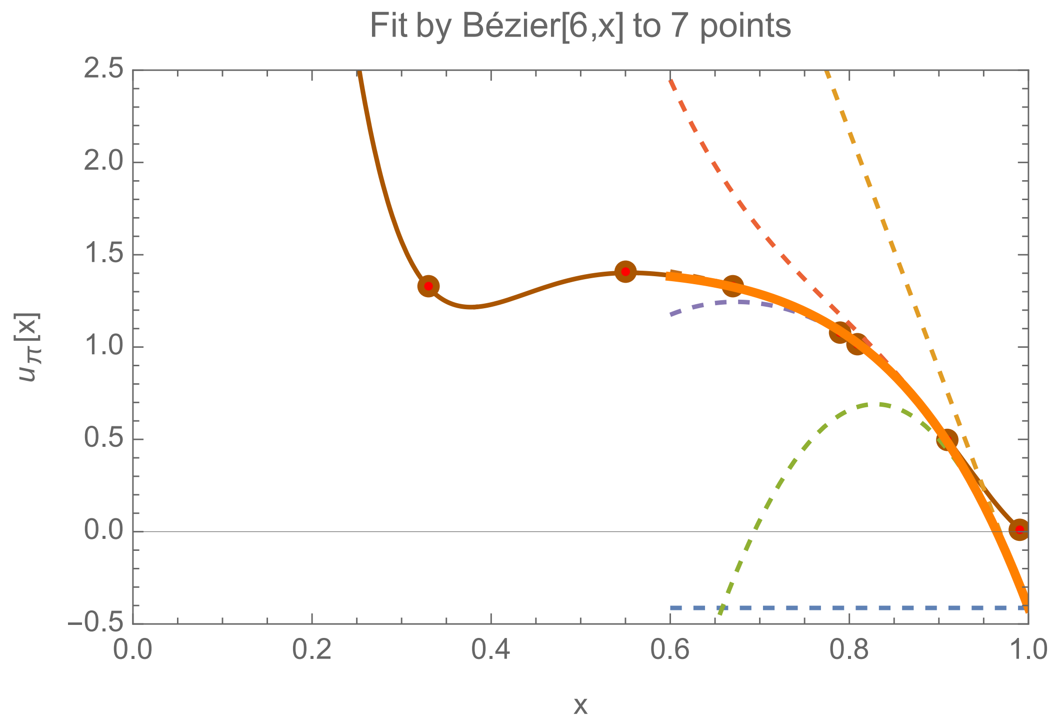
Reconstruction of a more complex parametrization

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

⇨ lowest powers of the expansion cannot be meaningfully reconstructed.





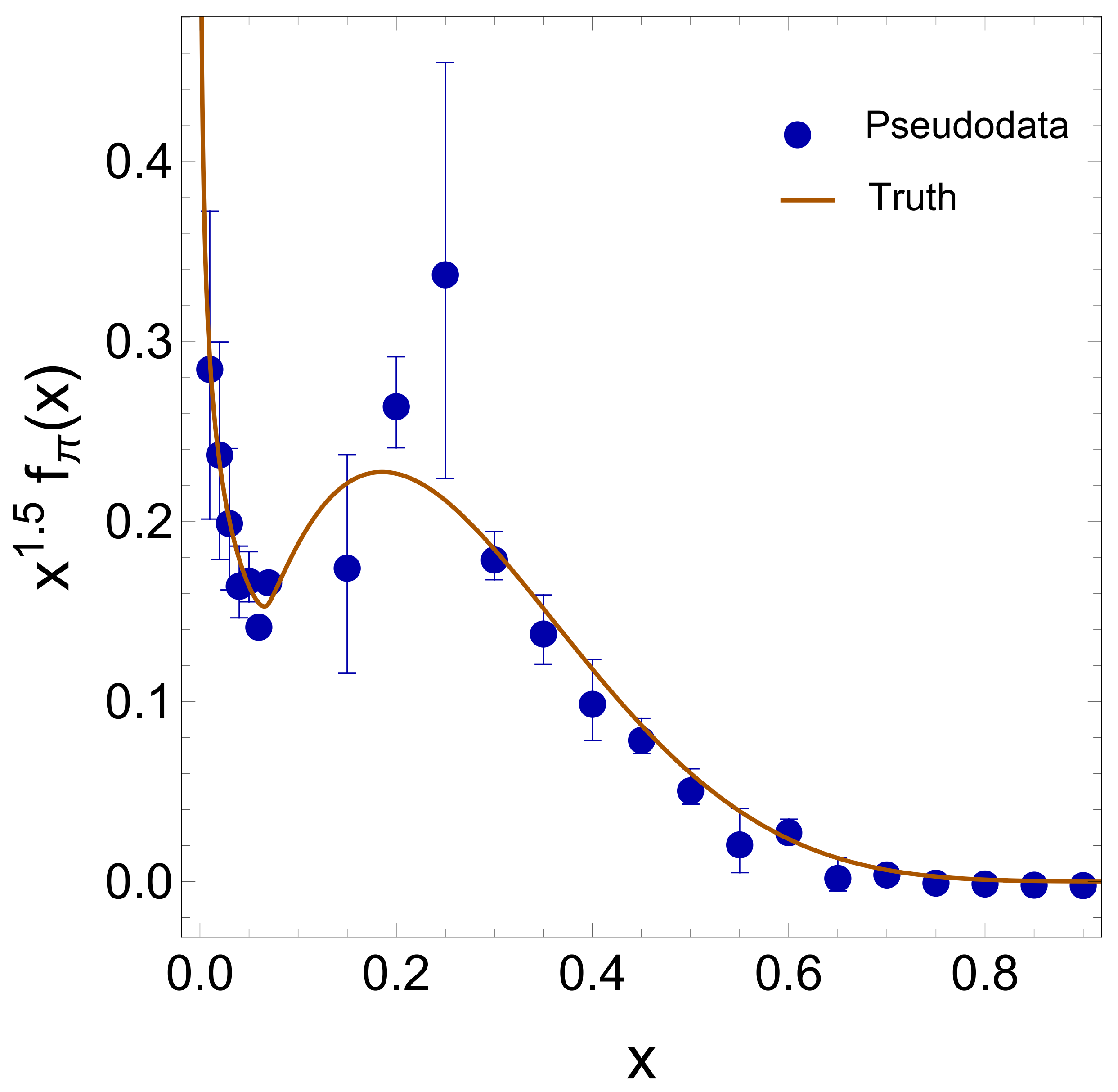


This property can be exploited in favor of global analyses:
by varying settings of the Bézier curves, we generate a variety of curves, beyond reconstruction.

Bézier-curve methodology for global analyses — toy model

Fantômas4QCD program

- ⇨ **From interpolation to minimization over parameters through $\mathscr{B}$**
- ⇨ Exploit polynomial mimicry to systematically improve and flexibilize parametrization of PDFs.



Classical fit:

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

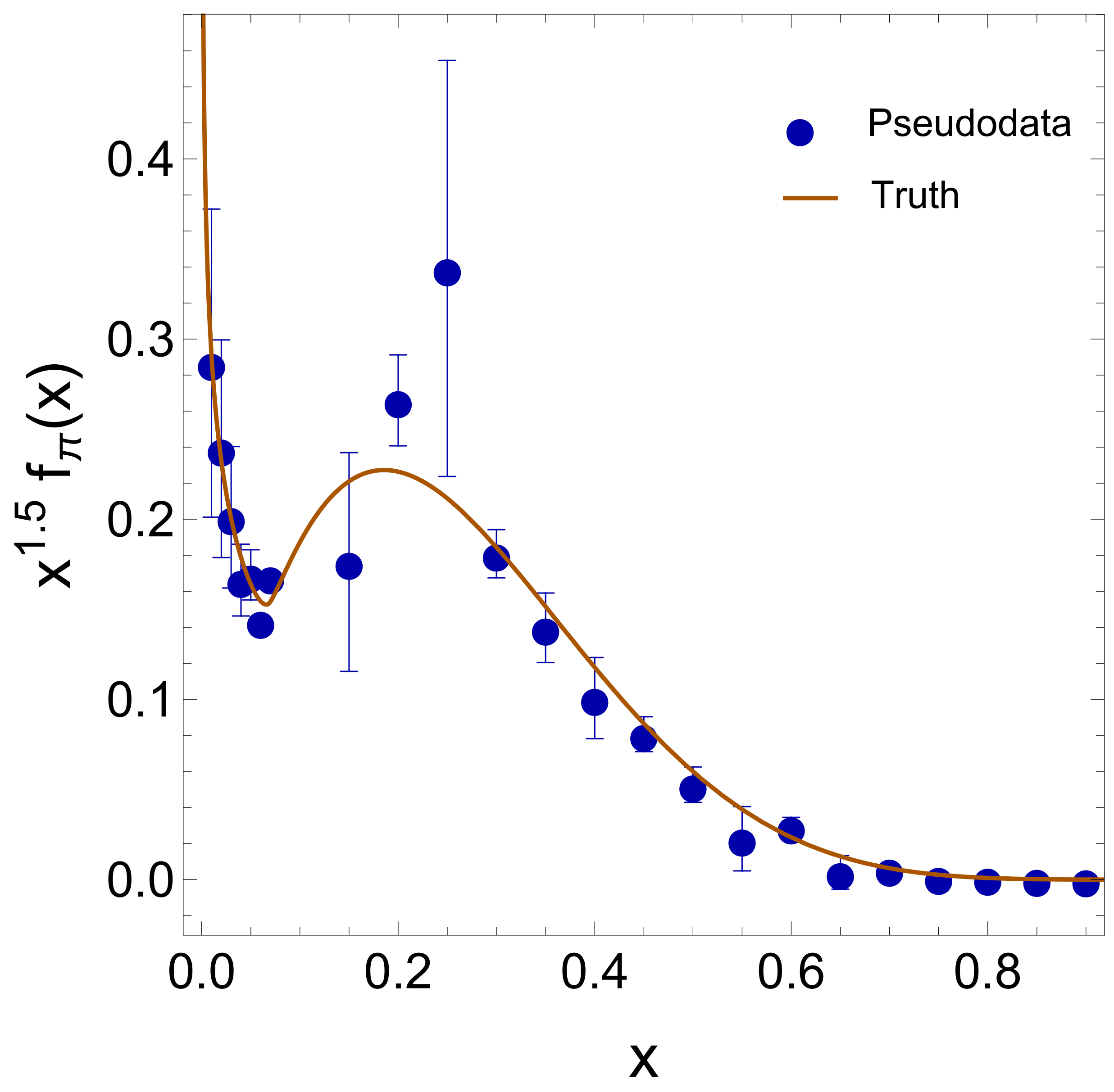
metamorph fit:

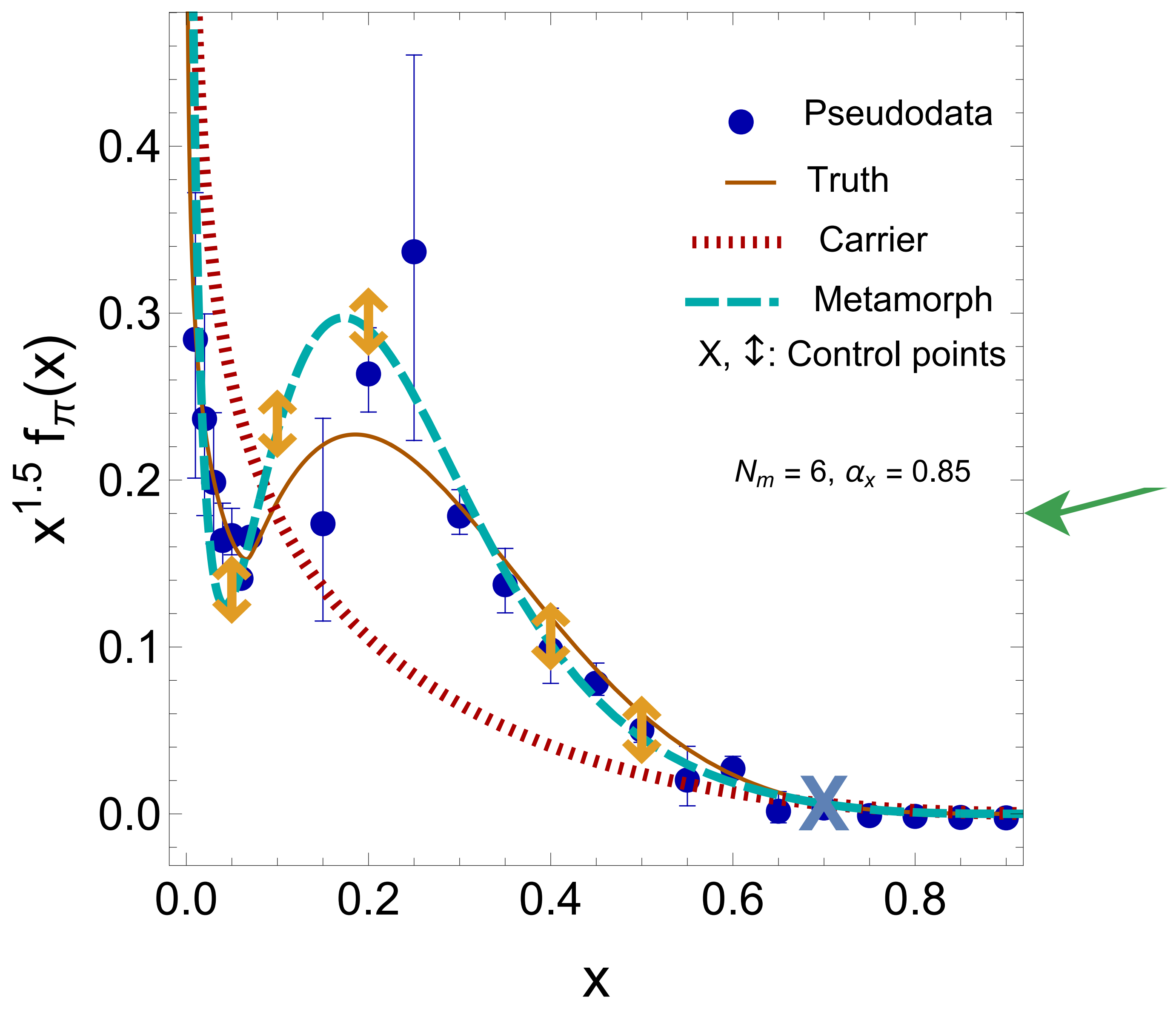
$$x\,q^{\rm Fanto}(x,Q_0^2) = a_q\,x^{B_q+\delta B_q}(1-x)^{C_q+\delta C_q} \times \left(1+\mathcal{B}^{(N_m)}(x,Q_0^2;\{\delta D_q,\delta E_q,\cdots\})\right) \quad \blacktriangleright \ \underline{P}$$

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

$$P_i = \mathcal{B}(x_i) \to P'_i = \mathcal{B}(x_i) + \delta\mathcal{B}(x_i)$$
$$\to \underline{P'} = (\mathcal{B}_0(x_1)+\delta D, \mathcal{B}_0(x_2)+\delta E, \cdots)$$

Bézier-curve methodology for global analyses — toy model





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

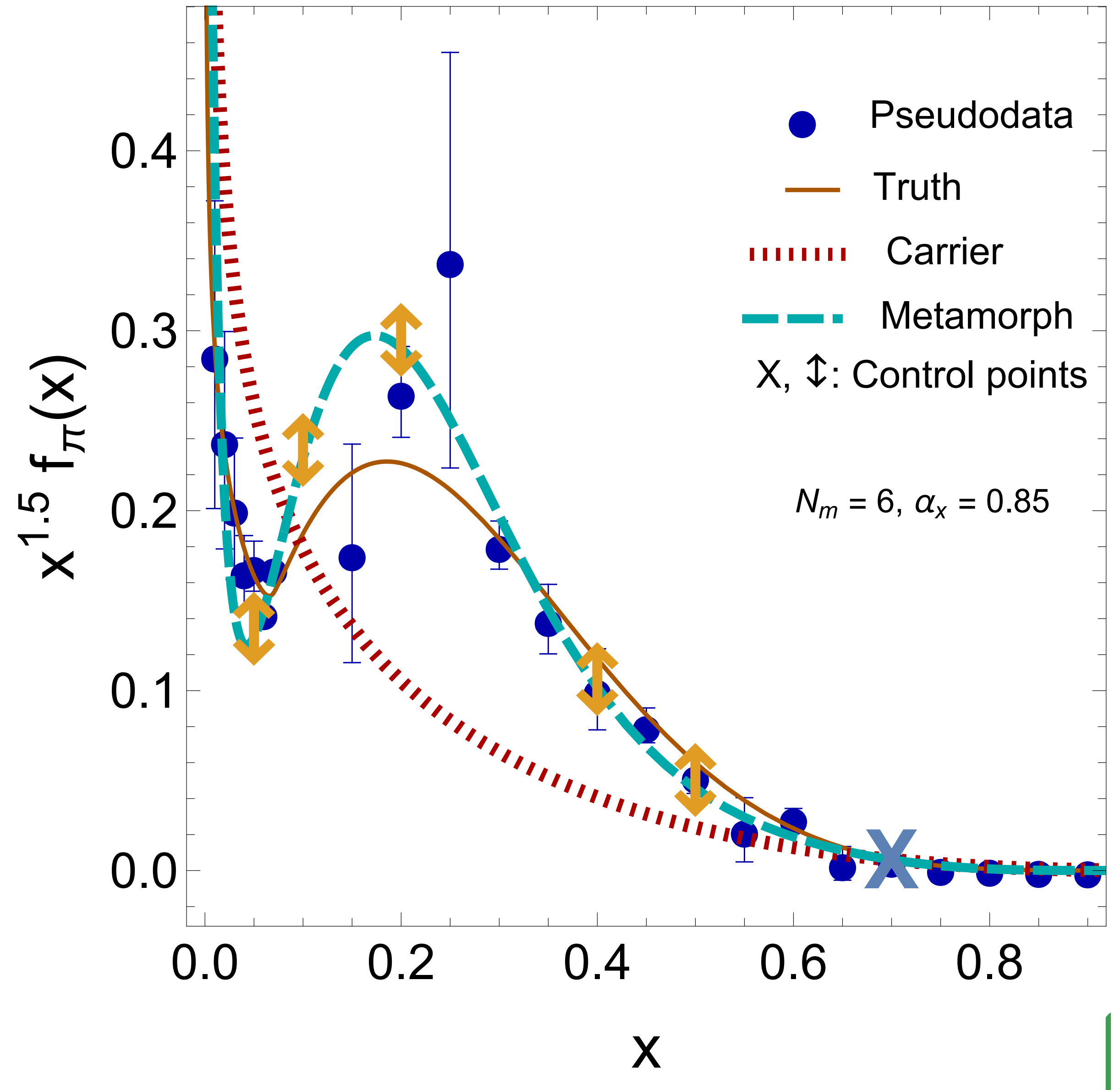
N_m= degree of polynomial can vary

δB_q & δC_q allow the carrier to vary

Classical fit: $$x\, q(x, Q_0^2) = A'_q\, x^{B_q} (1-x)^{C_q} \times \left(1 + \mathcal{B}^{(N_m)}(x, Q_0^2)\right)$$

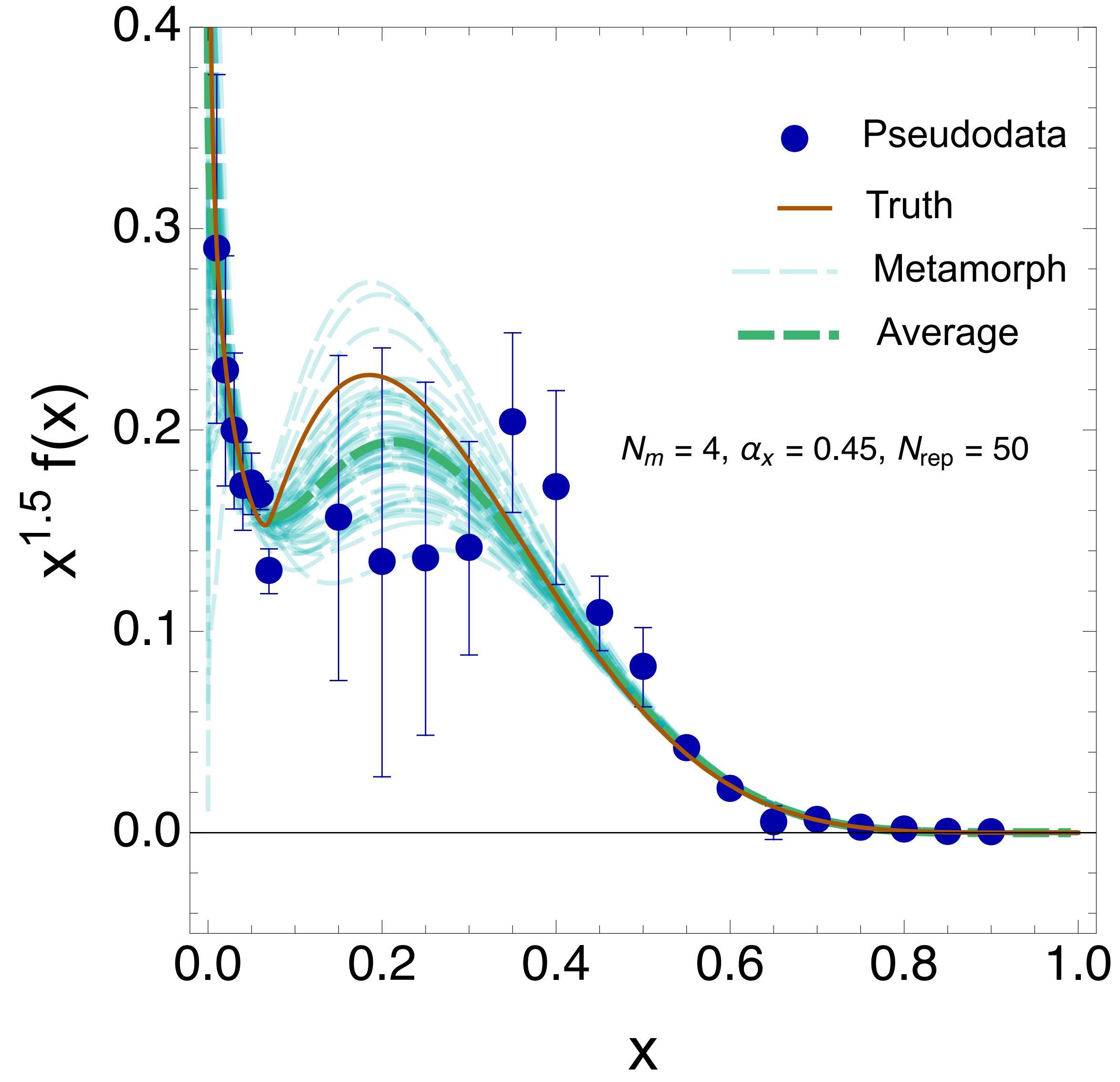
`metamorph` fit: $$x\, q^{\text{Fanto}}(x, Q_0^2) = a_q\, x^{B_q+\delta B_q} (1-x)^{C_q+\delta C_q} \times \left(1 + \mathcal{B}^{(N_m)}(x, Q_0^2; \{\delta D_q, \delta E_q, \cdots\})\right)$$

Bézier-curve methodology for global analyses — toy model



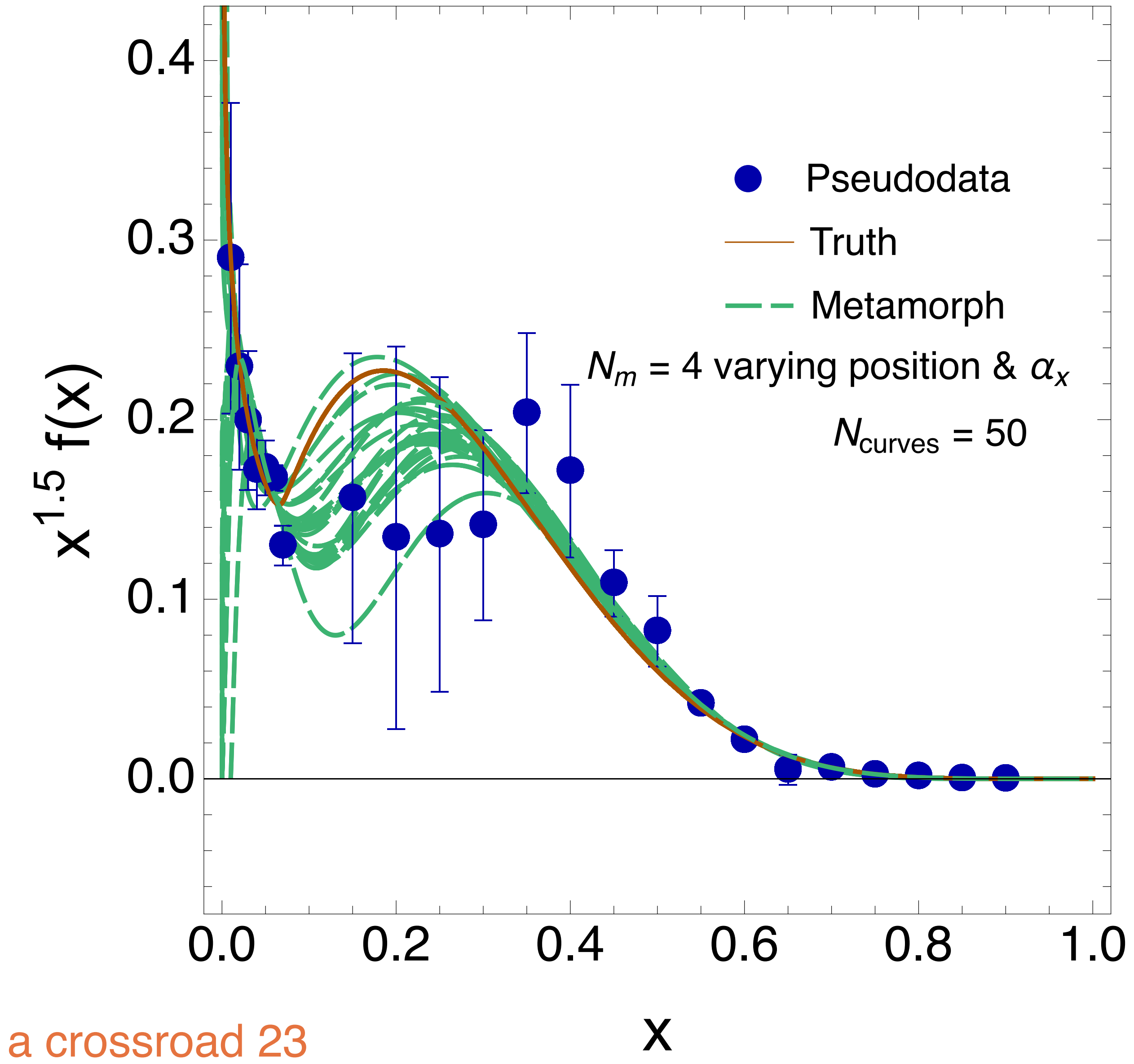
if bootstrapped

sampling on the distribution of data uncertainties



if sampled over functional forms

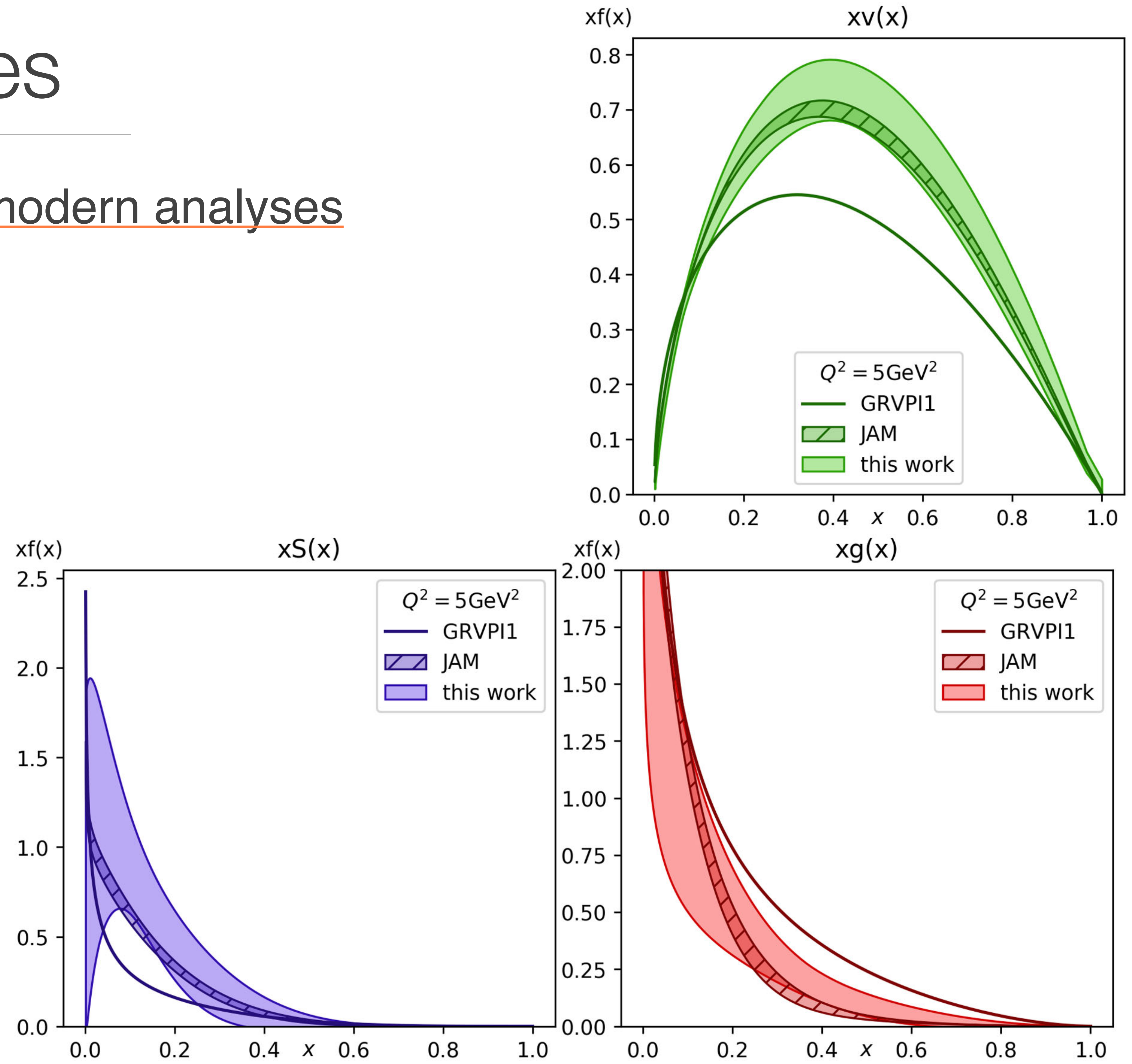
sampling over parametrizations



Both sampling can be done simultaneously, they are not mutually exclusive.

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]

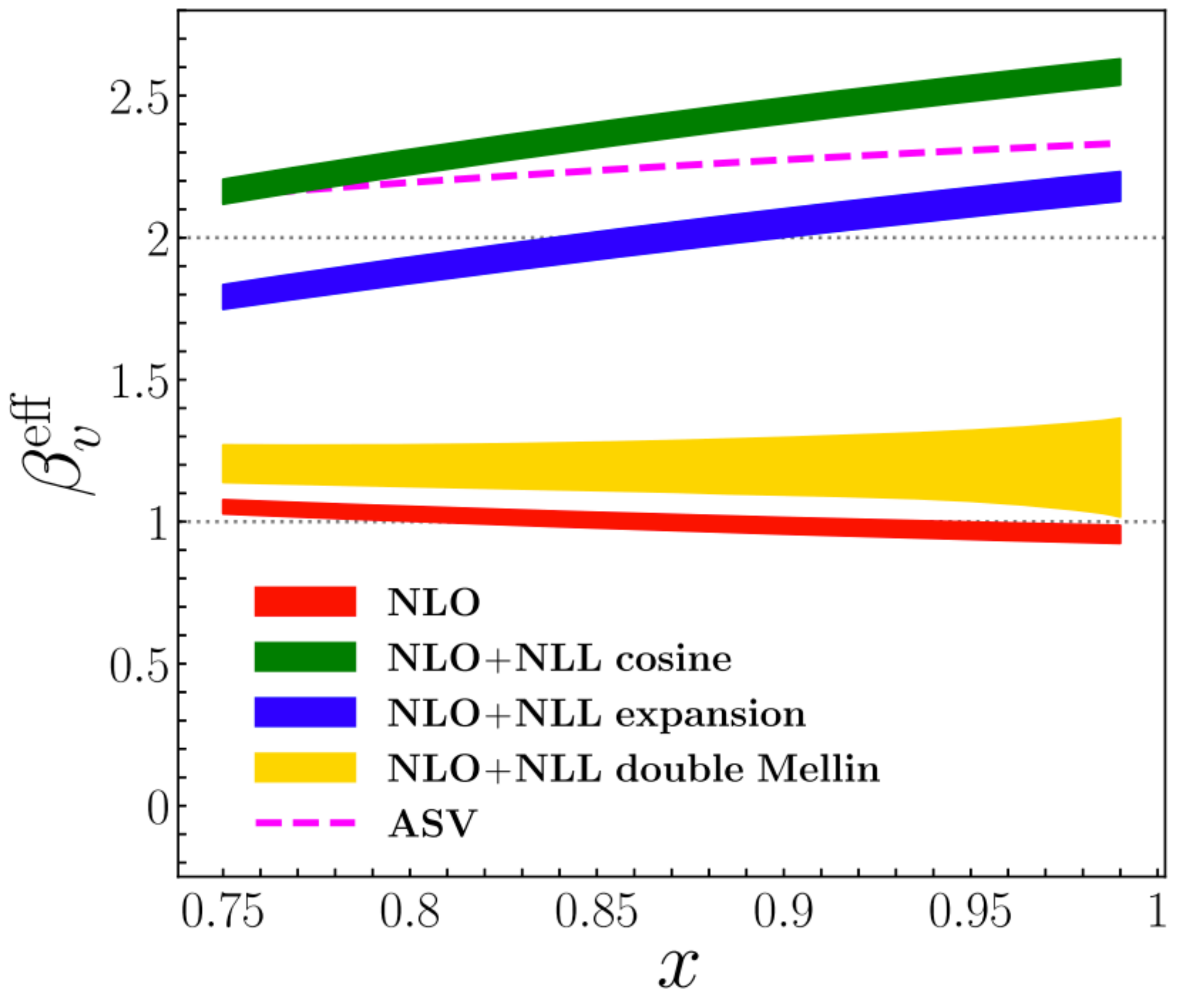
complemented by [model-dependent] leading-neutron data [**JAM**]

include large-x resummation

➡ ASV (2010, early LxR)

➡ JAM21 (2021, updated LxR)

** LxR experts favor the double Mellin resummation.

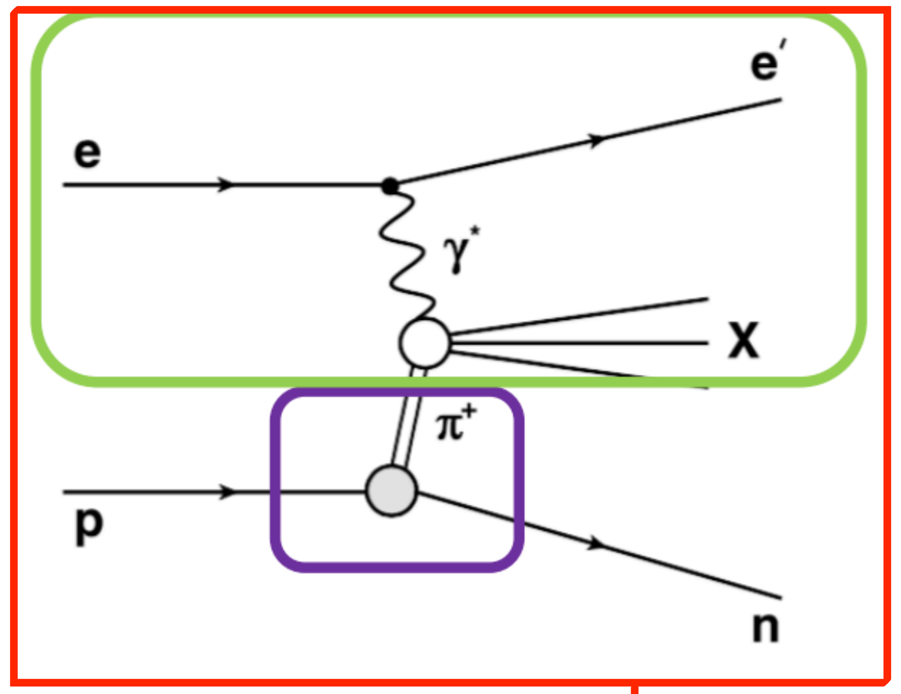


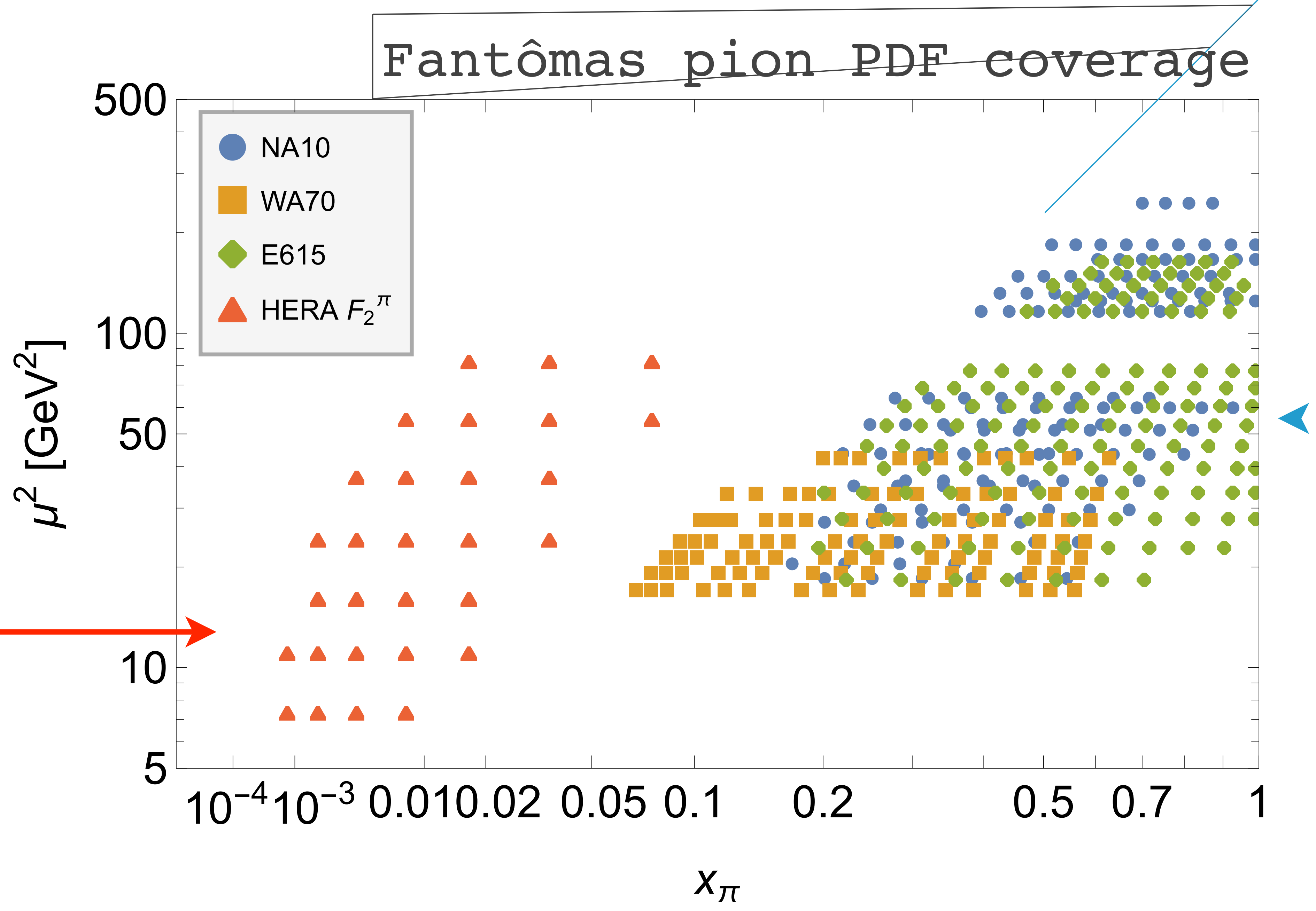
Data for pion PDF

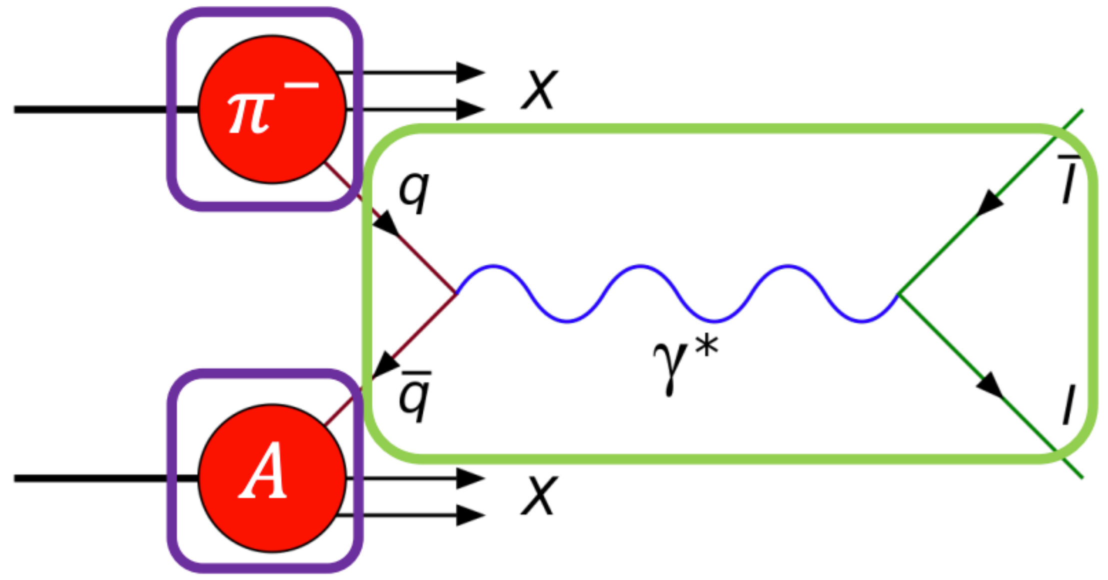
We use the xFitter framework, in which `metamorph` was implemented as an independent parametrization.

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

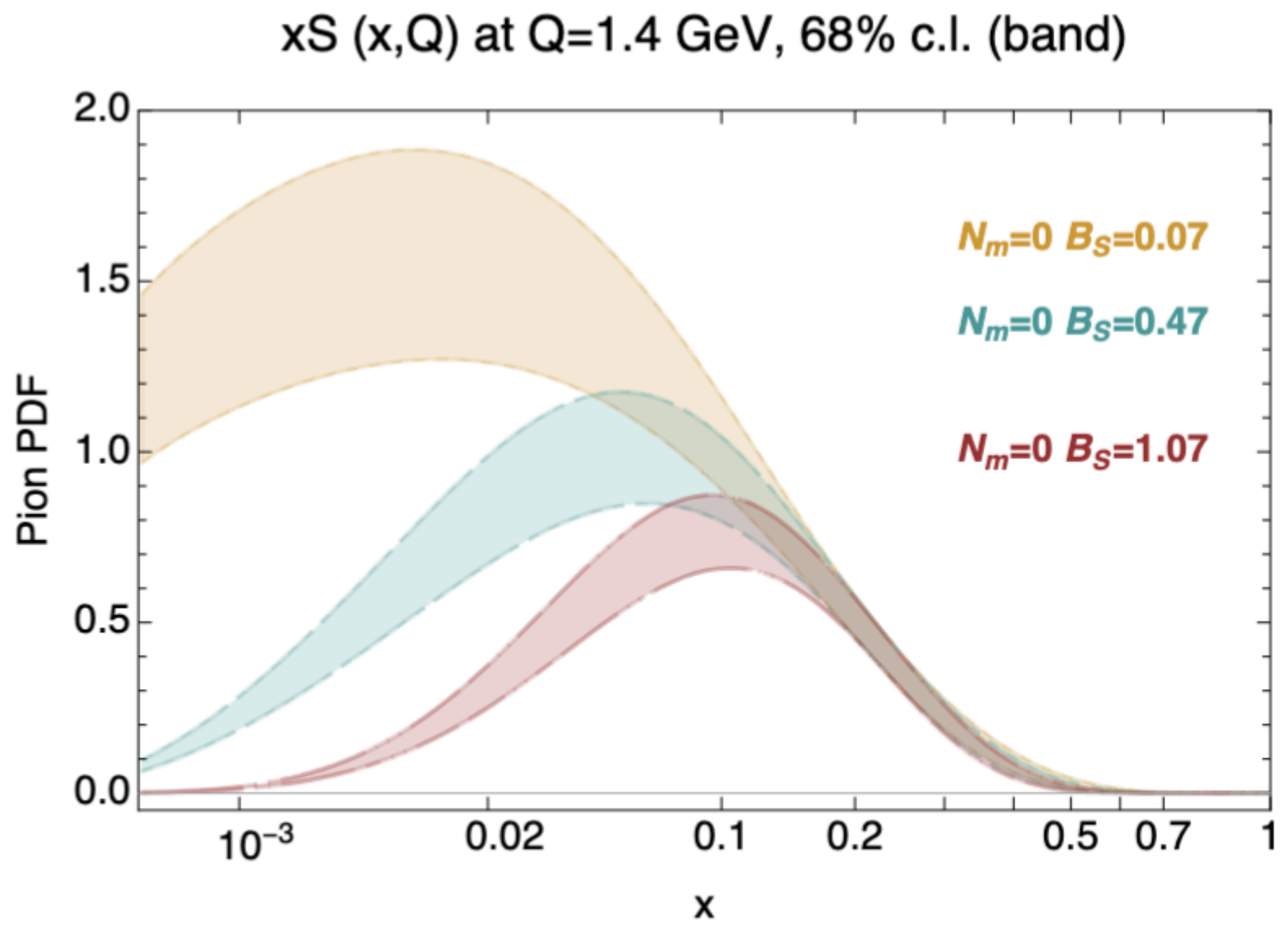
Drell-Yan only analysis

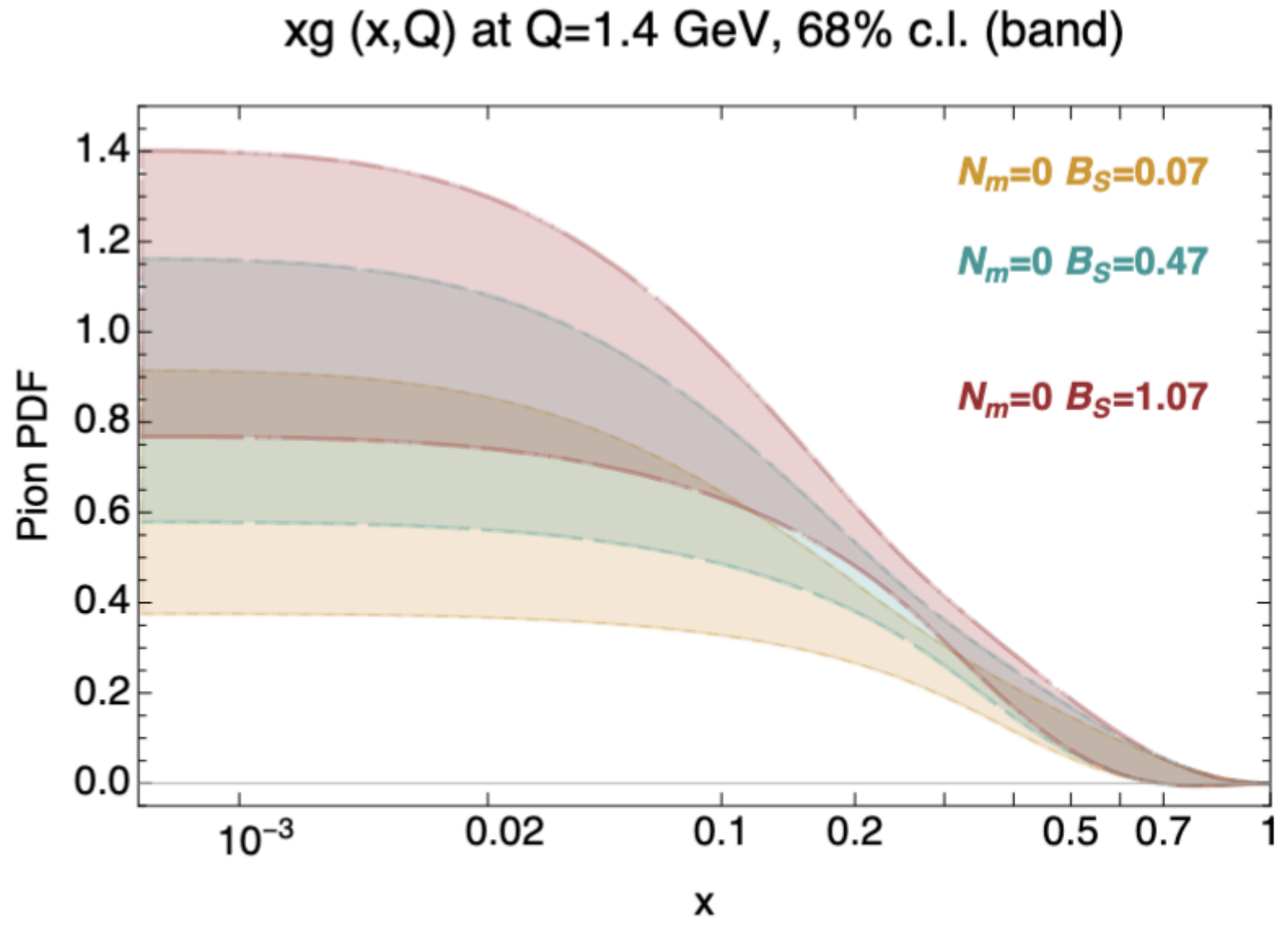
Previous analyses used a fairly basic parametrization

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

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.





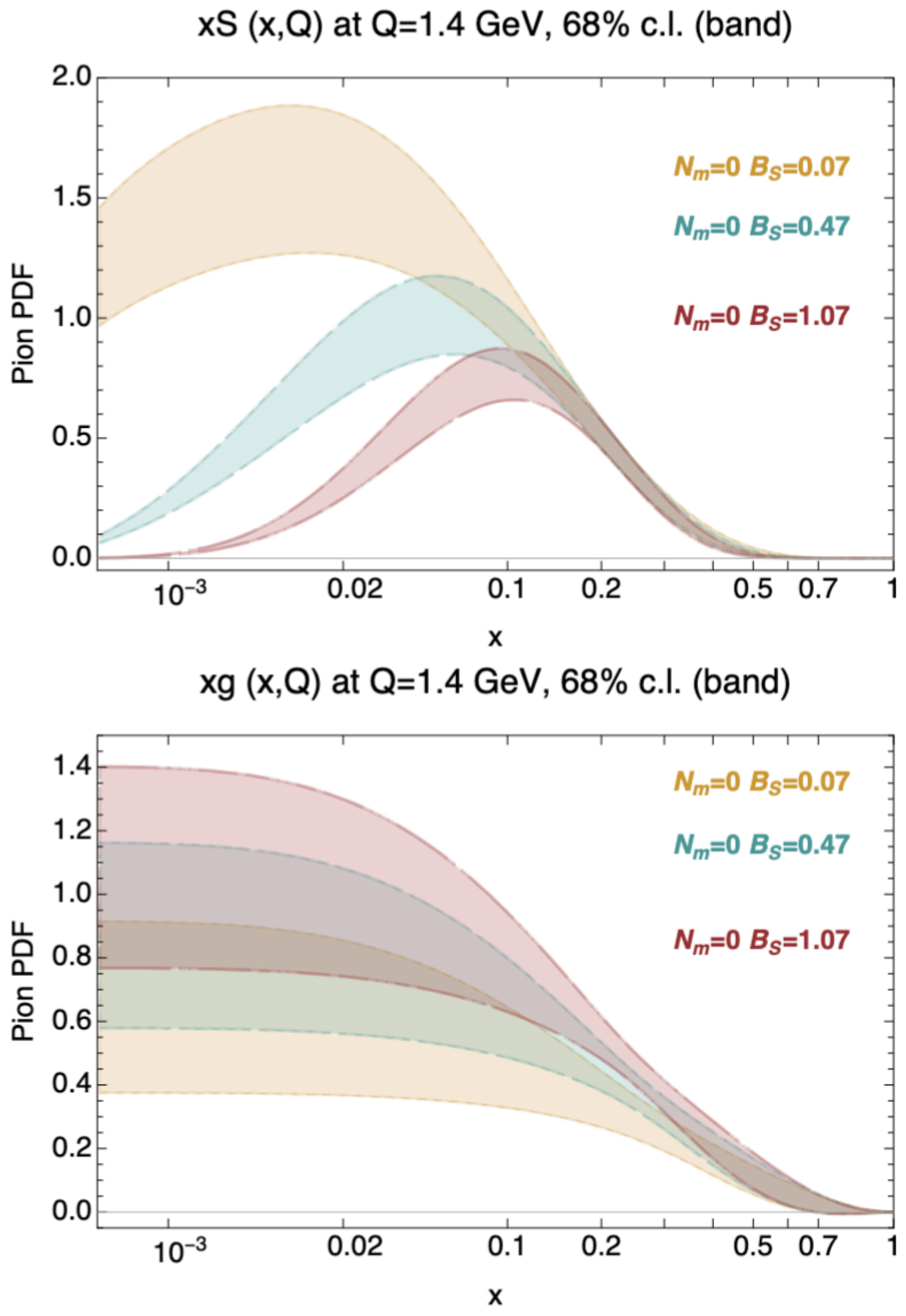
Drell-Yan only analysis

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

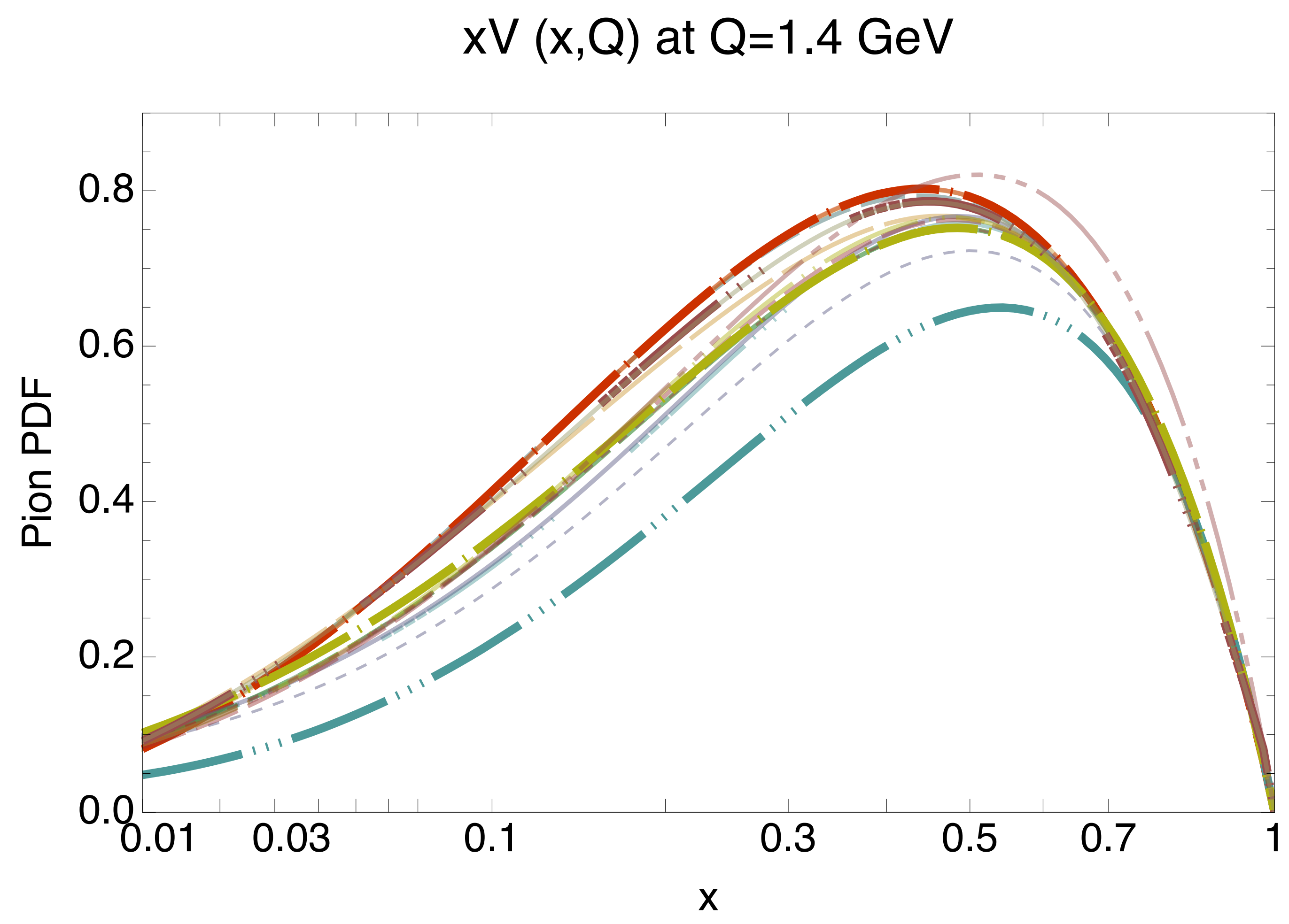
⇨ JAM (and HERA before them) proposed to use leading-neutron data

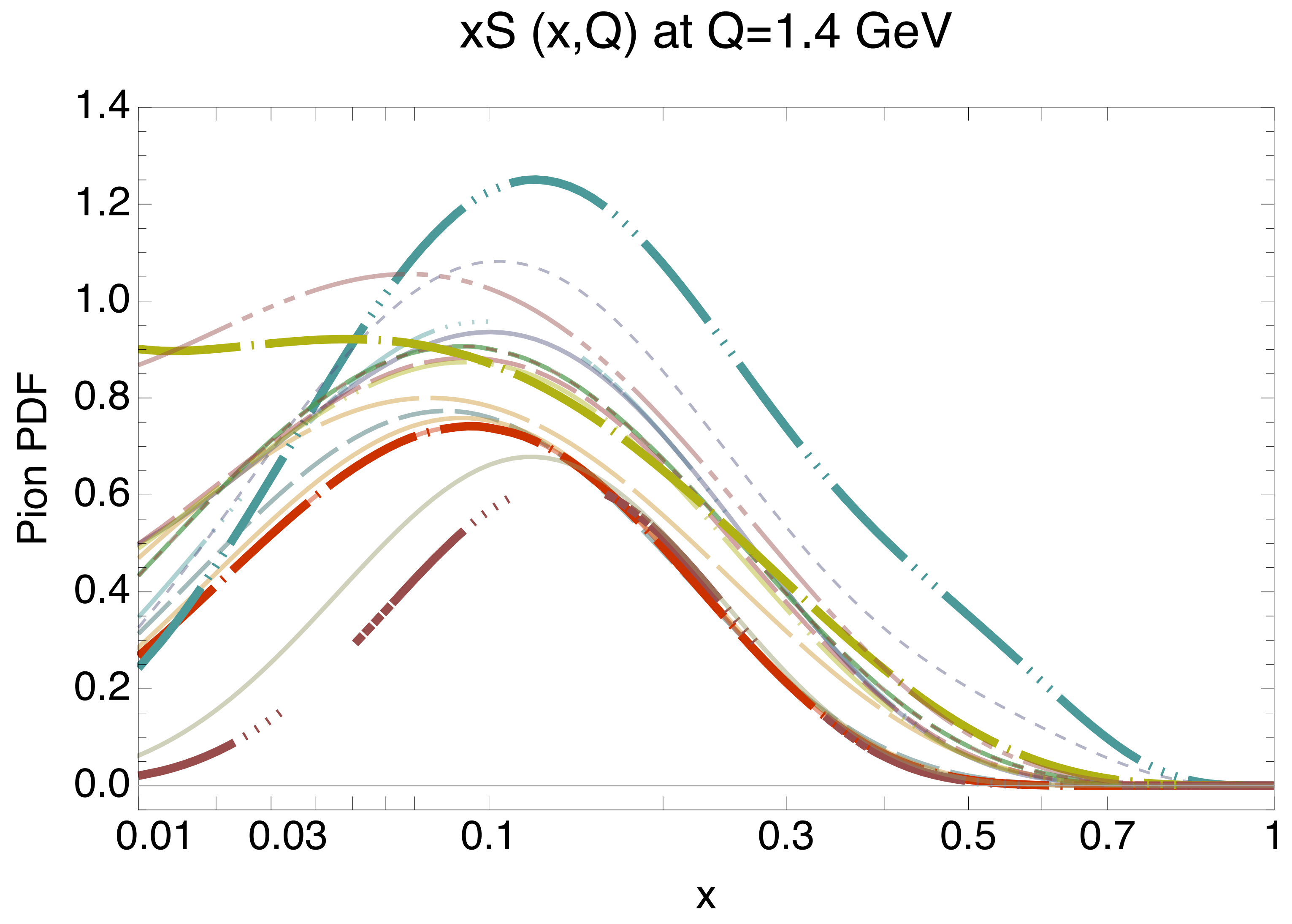
⇨ future experiments at EIC and JLab22(?)

Fantômas parametrizations for the pion PDF

Fantômas analysis uses varying sets of

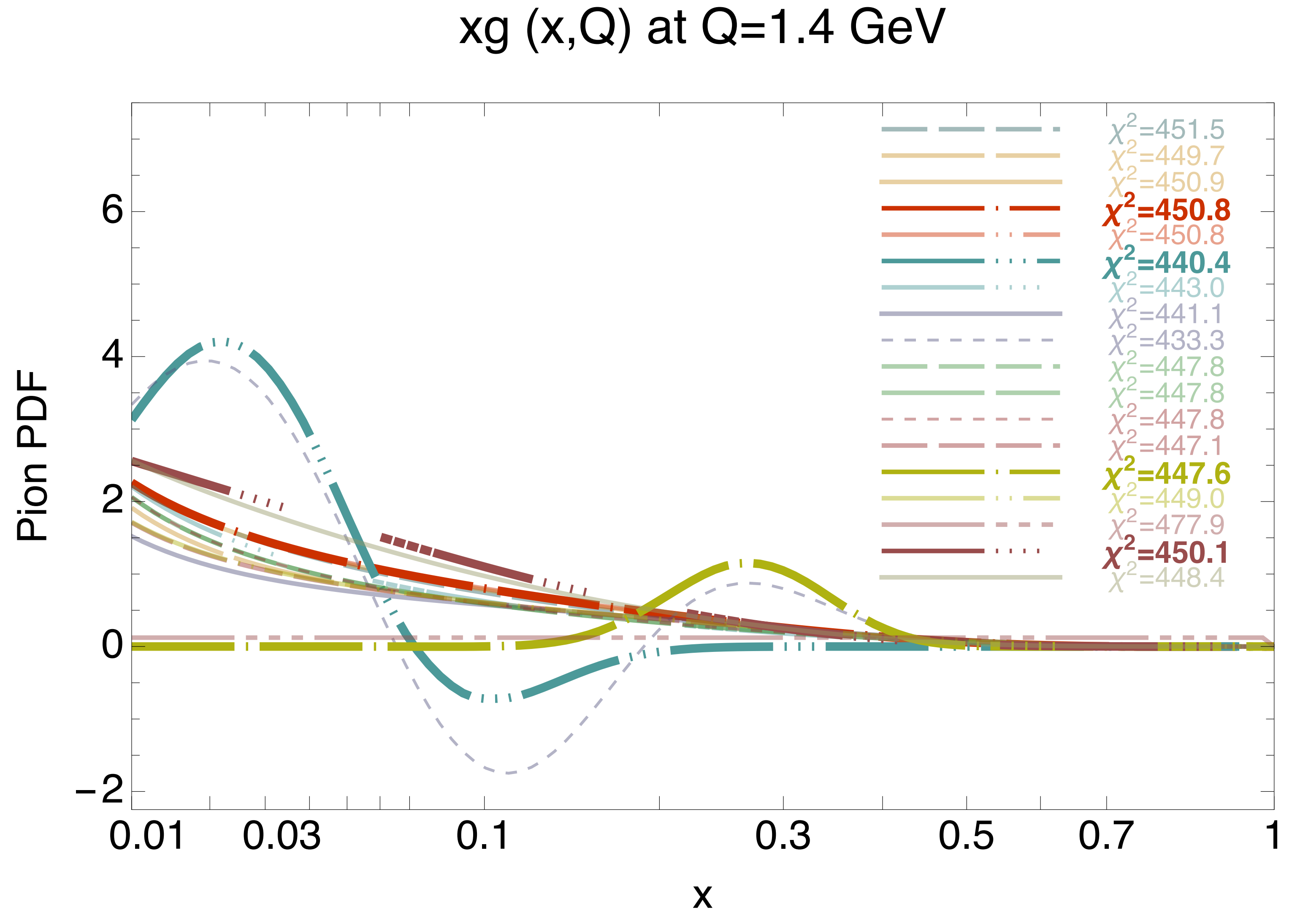
- degree of polynomial (0,1,2),
- position of fixed/free control points,
- stretching parameter of the argument





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

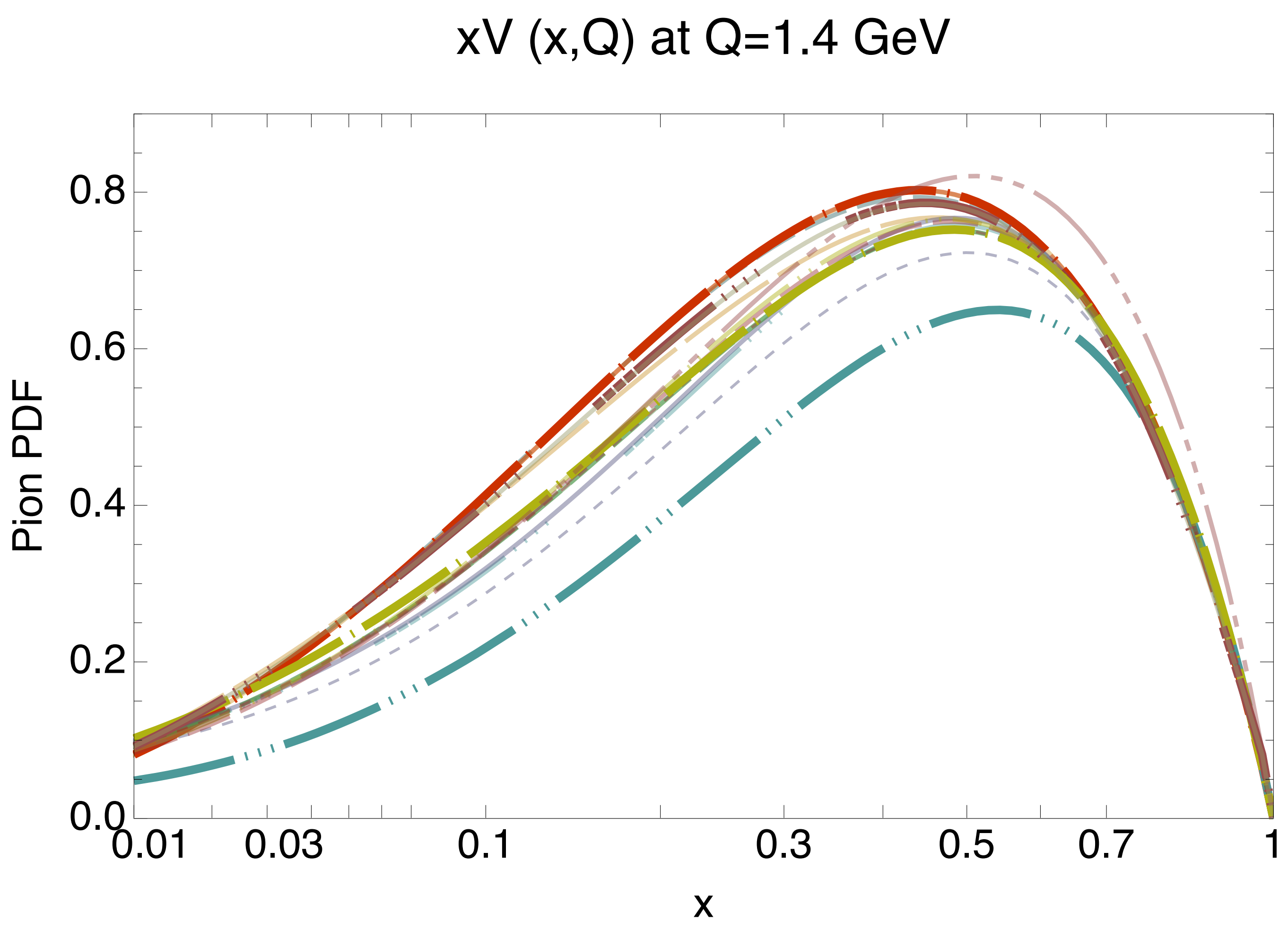
Negative gluon are found to be possible at such a low scale [confirming JAM's findings].

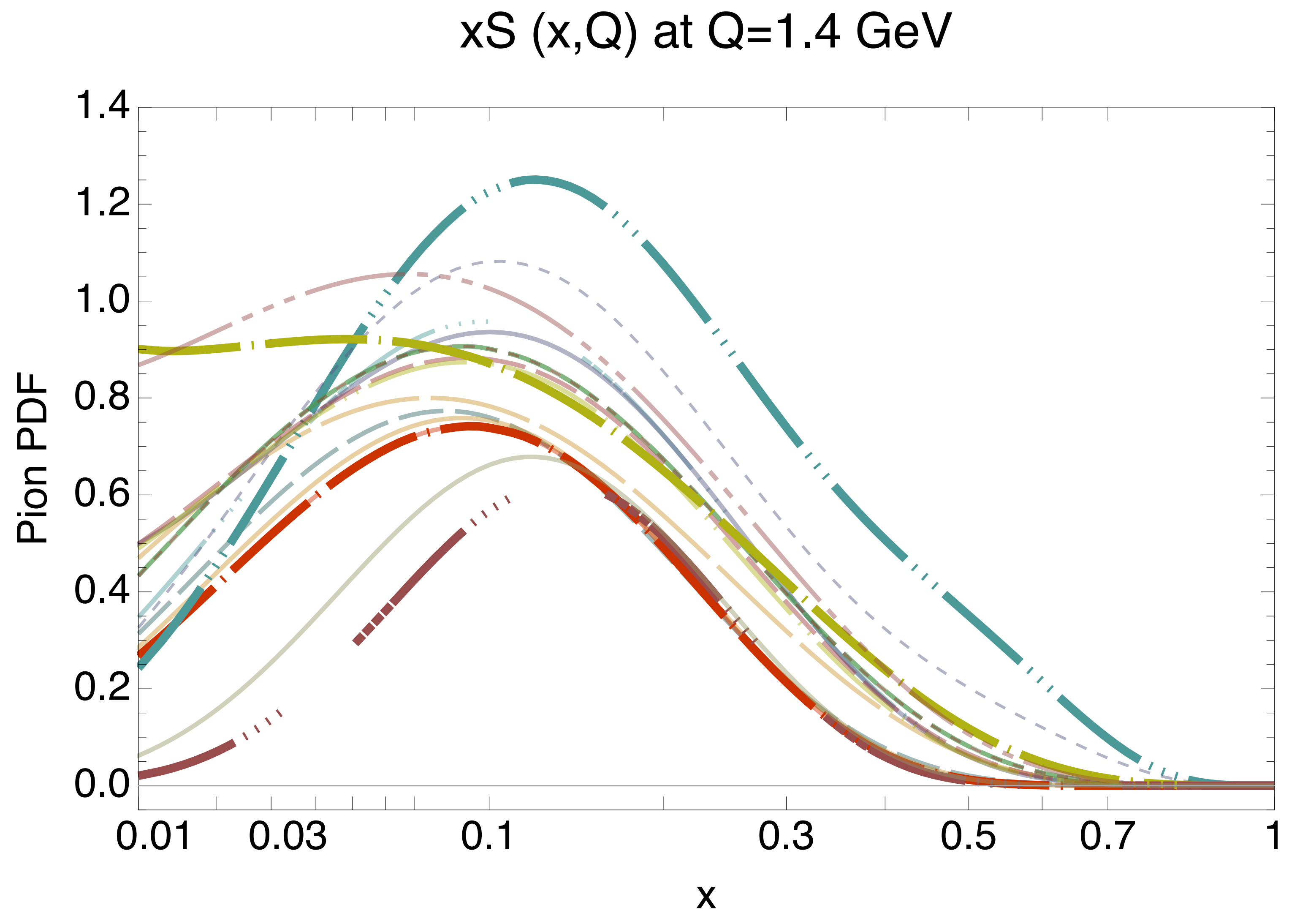


Fantômas parametrizations for the pion PDF

Fantômas analysis uses varying sets of

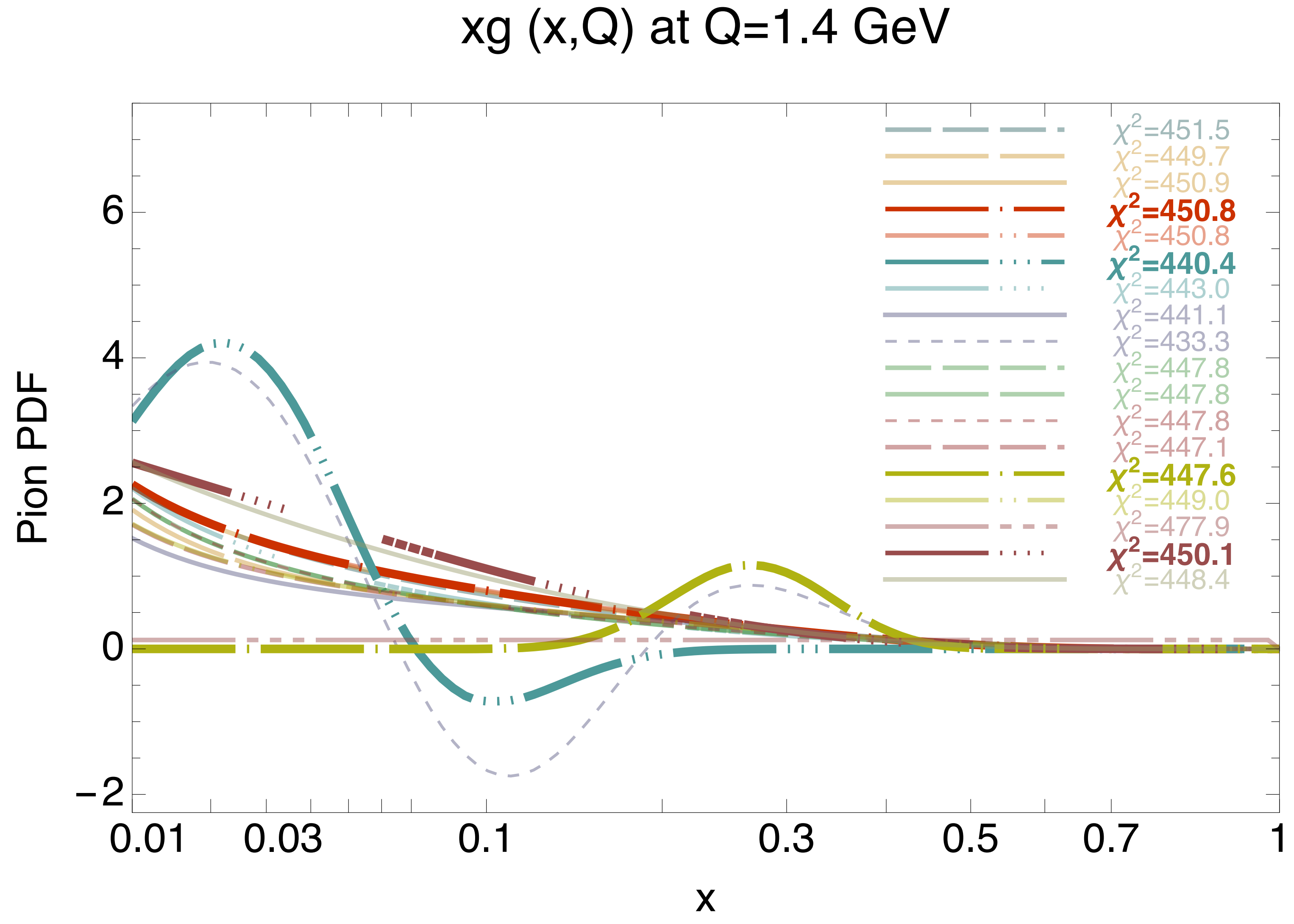
- degree of polynomial (0,1,2),
- position of fixed/free control points,
- stretching parameter of the argument





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].



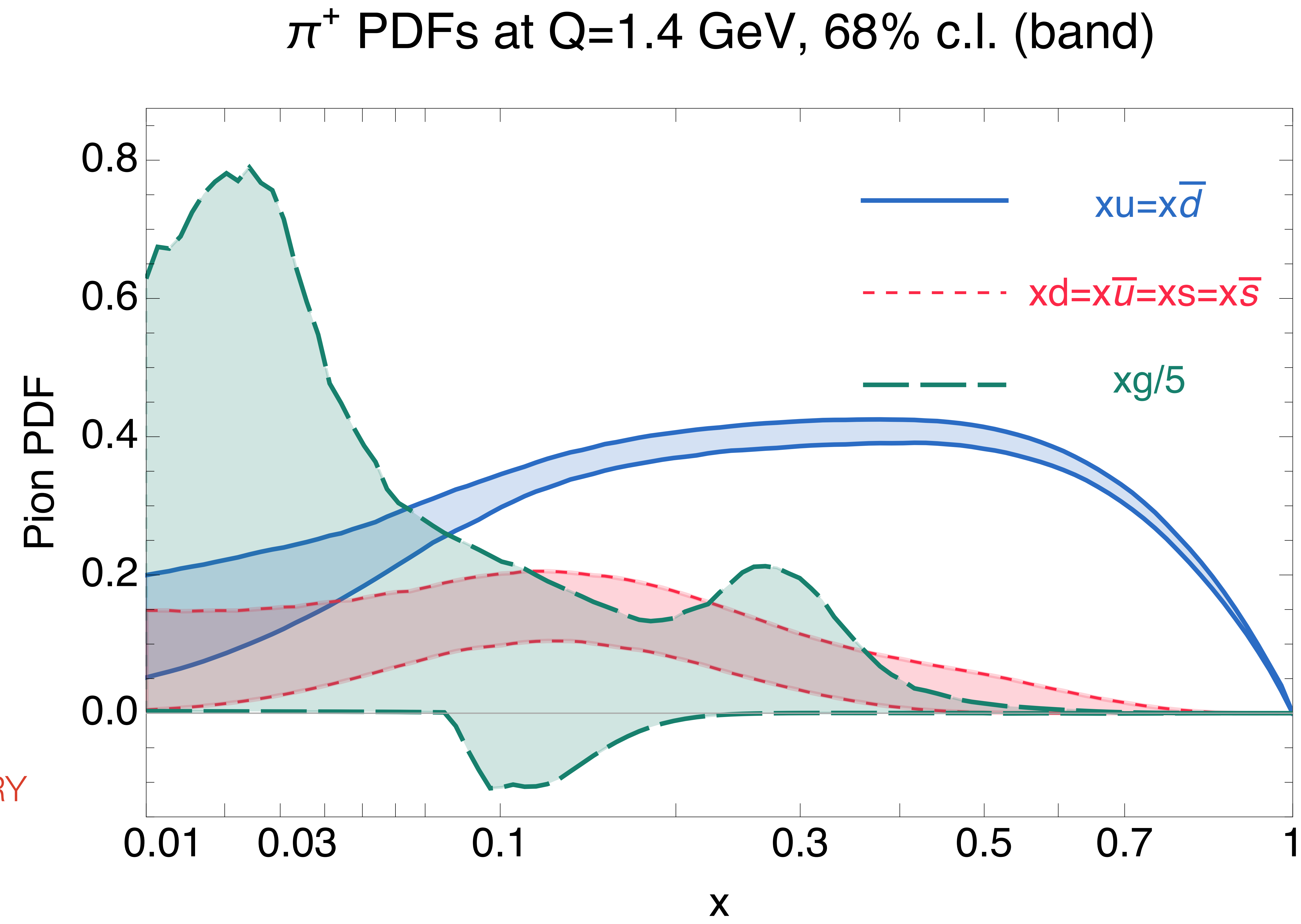
Bold curves correspond to our selection for the final Fantômas set.

Criterion of 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 NLO pion PDFs



PRELIMINARY

For a selection of $\{N_m, CP\}$ sets.

Bundled uncertainty with `mcgen`

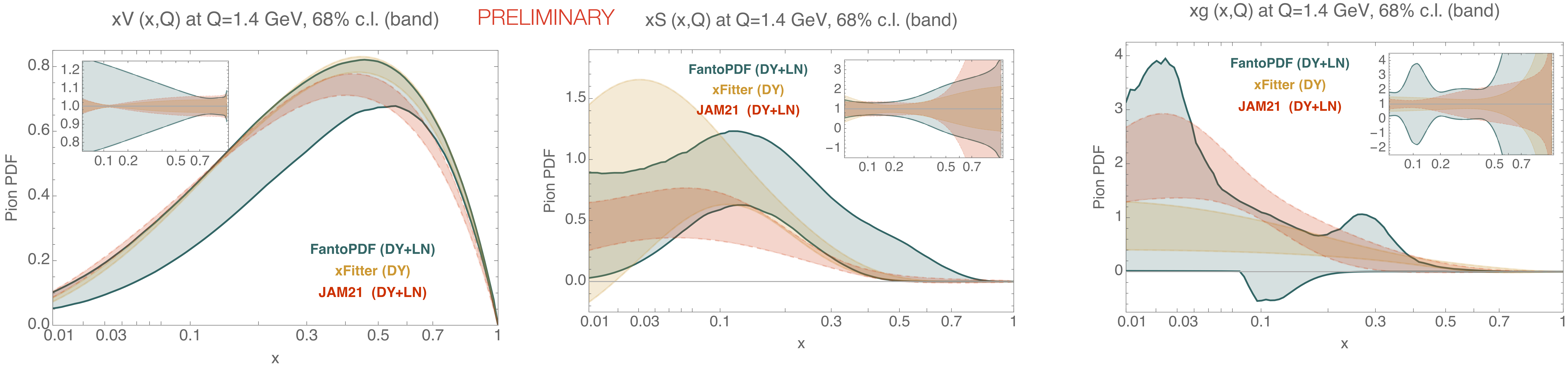
[Gao & Nadolsky, JHEP07]

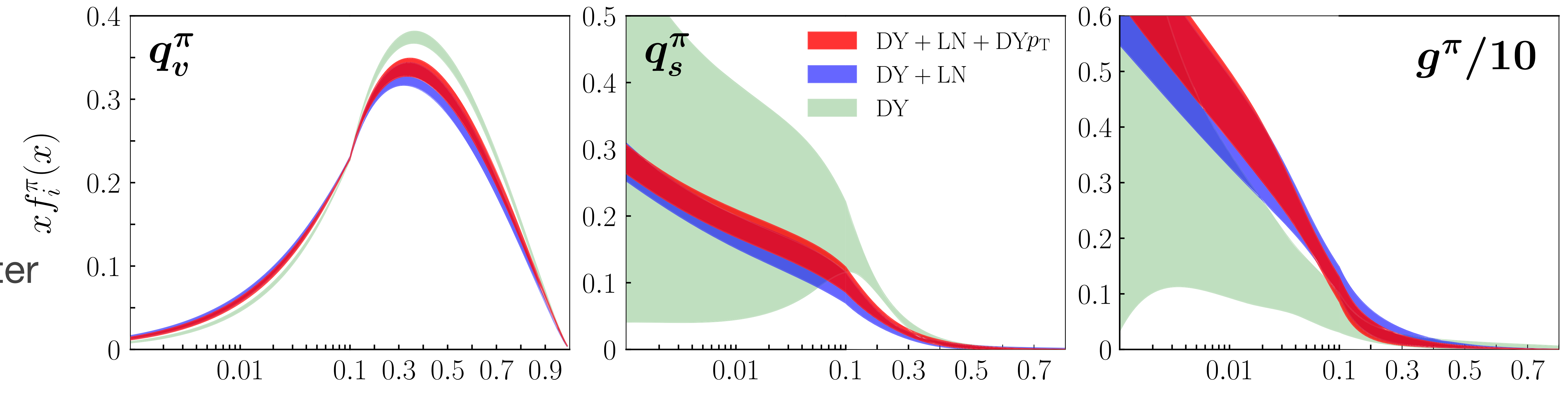
*[Kotz, Ponce-Chávez, **AC,** Nadolsky & Olness]*

Proceedings in 2309.00152.

The Fantômas pion PDFs

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



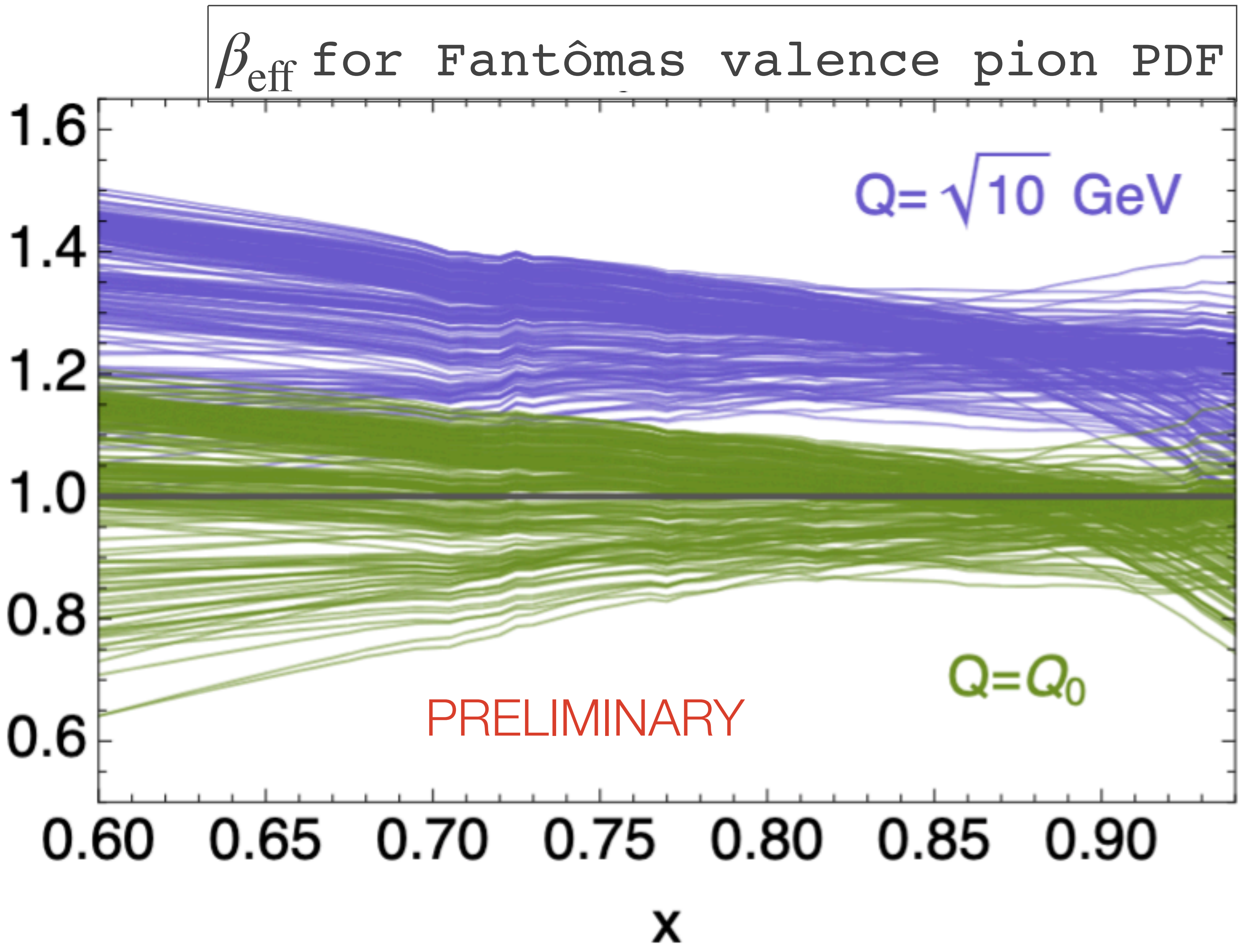


JAM [Phys.Rev.D103]

<u>Comparison of methodologies:</u>

bootstrap+ IMC vs. `metamorph` parametrization in xFitter

Large-x behavior of the valence pion PDF



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

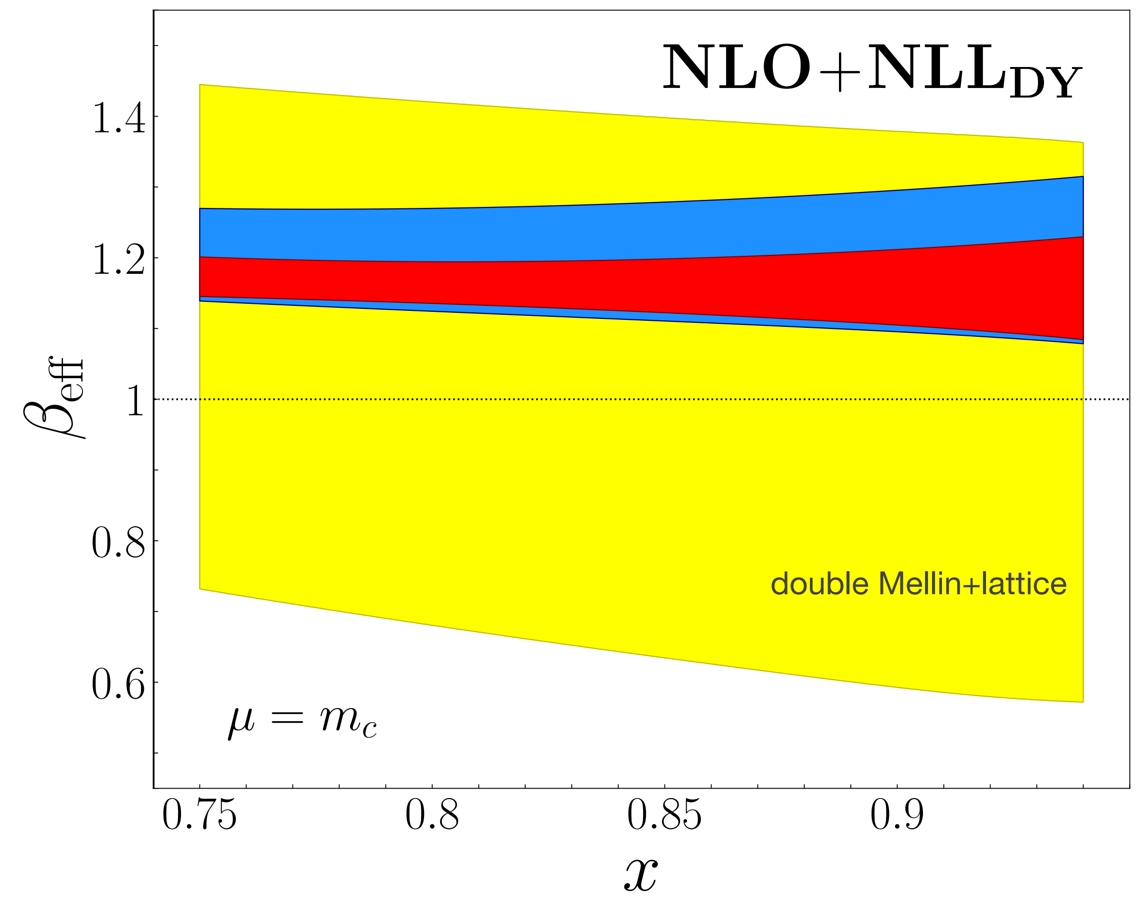
⇨ doesn't fall very much like $(1-x)^2$

⇨ very similar to JAM and xFitter at large x

Corrective terms might need to be taken into account [large-x resummation].
JAM did and found an exponent between 1 to ~2.5, depending on the prescription [JAM, PRL127].

Lattice studies contribute to the information on hadron structure. Mindful analysis of the determination of the effective exponent of the PDF fall-off on the lattice [Gao et al., PRD102].

⇨ inverse problem



[JAM+lattice, PRD105]

State-of-the-art of the pion at large x

Polynomial mimicry prevents functional behaviors from being validated as *if and only if* conditions.

Mathematical equivalence of polynomials of different orders can be illustrated with Bézier curves.

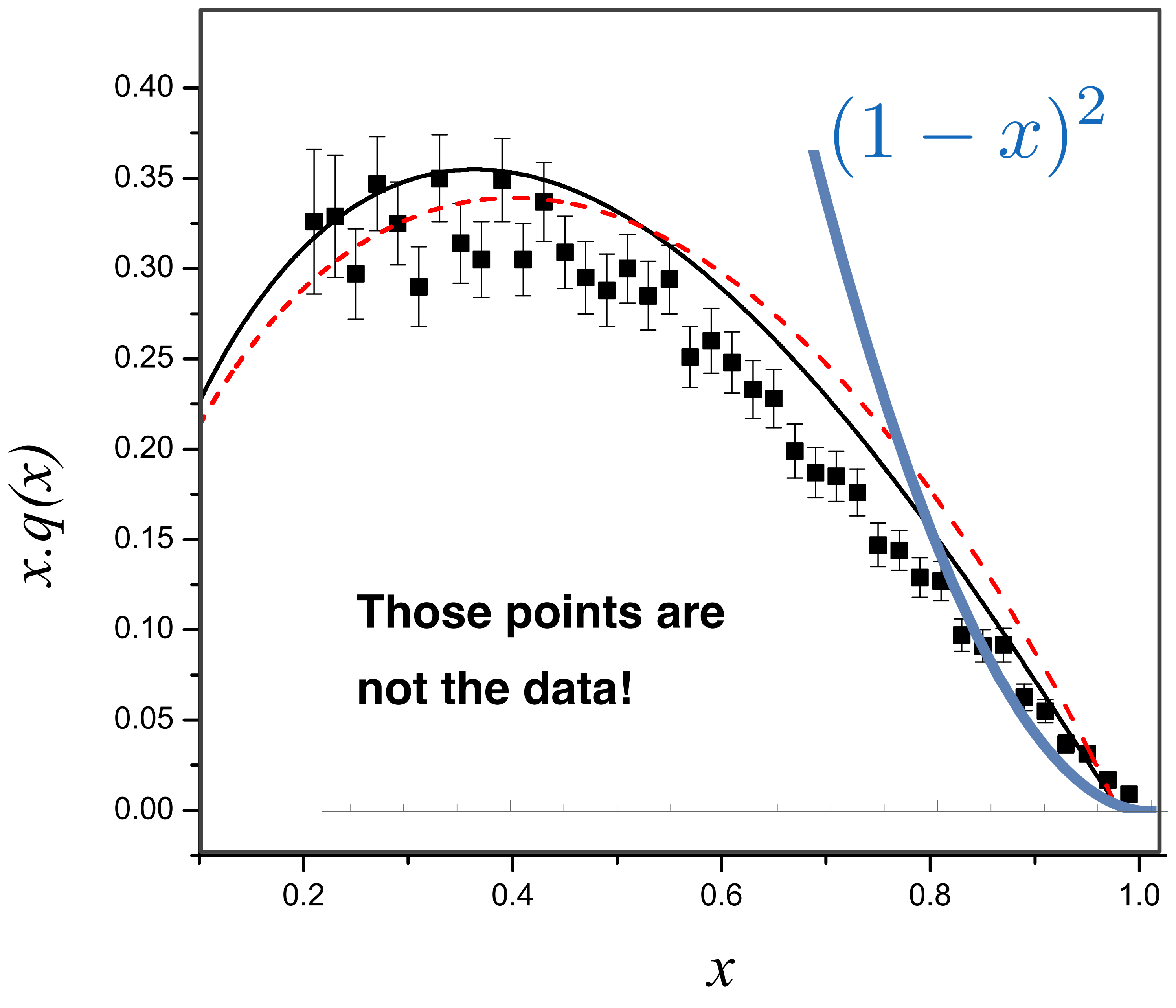
QCD corrections, at low and large Q^2, also inhibit the $(1-x)^\beta$ power to be tested.

Emergence of Hadronic Mass: broadens the PDF at μ_0^2 [μ_0 the hadronic scale]

Quark counting rules: $(1-x)^2$ tail at mid-Q^2 values

⇨ concurring effects that will not be distinguishable at a scale $Q^2 > \mu_0^2$.

PDF broadening in NJL vs. QCRs



Many interesting talks/results/discussions during this and the CFNS [QCD4EIC] workshops!

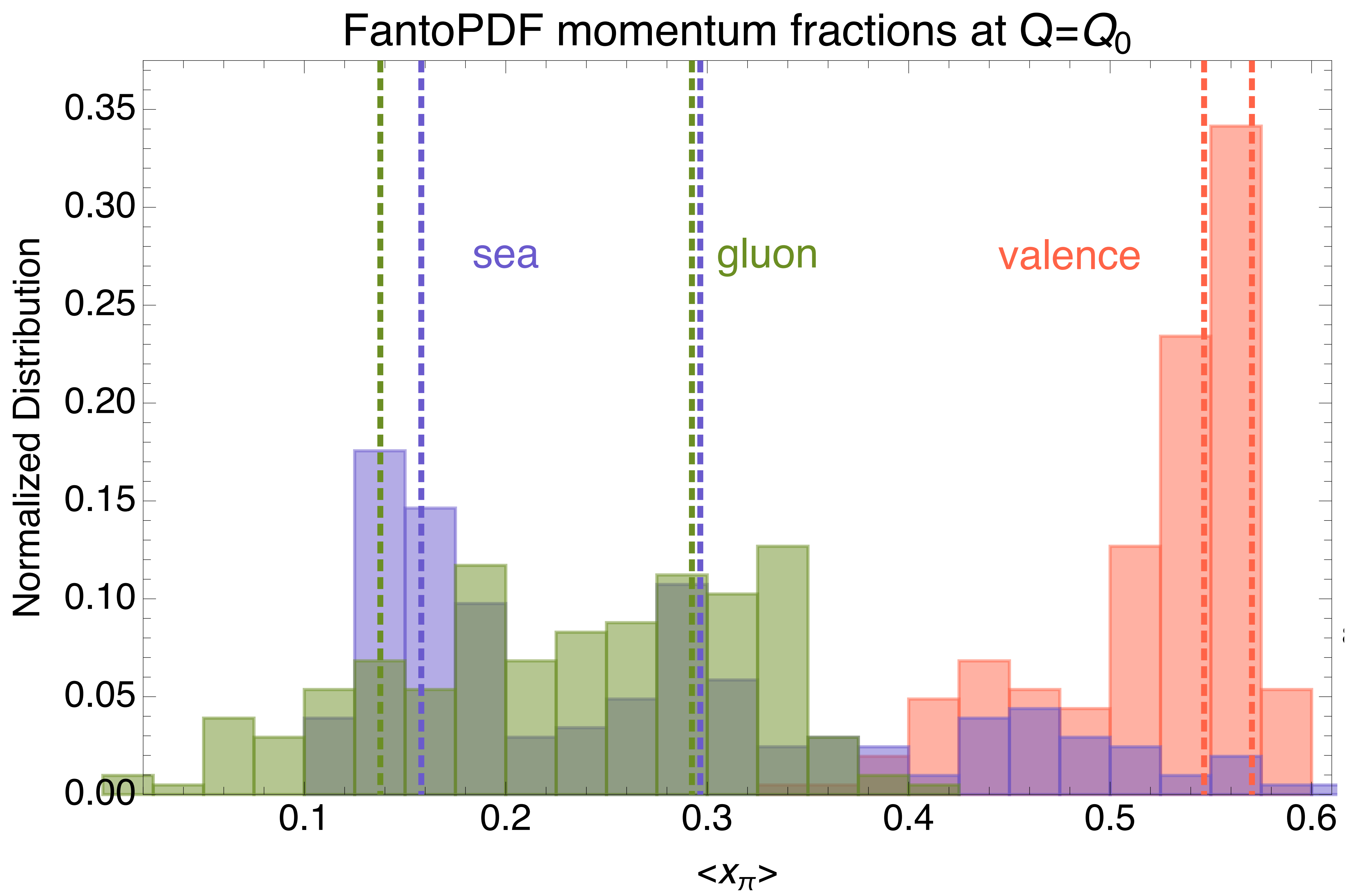
Momentum fraction

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 $< xf_q >$.

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



Momentum fraction

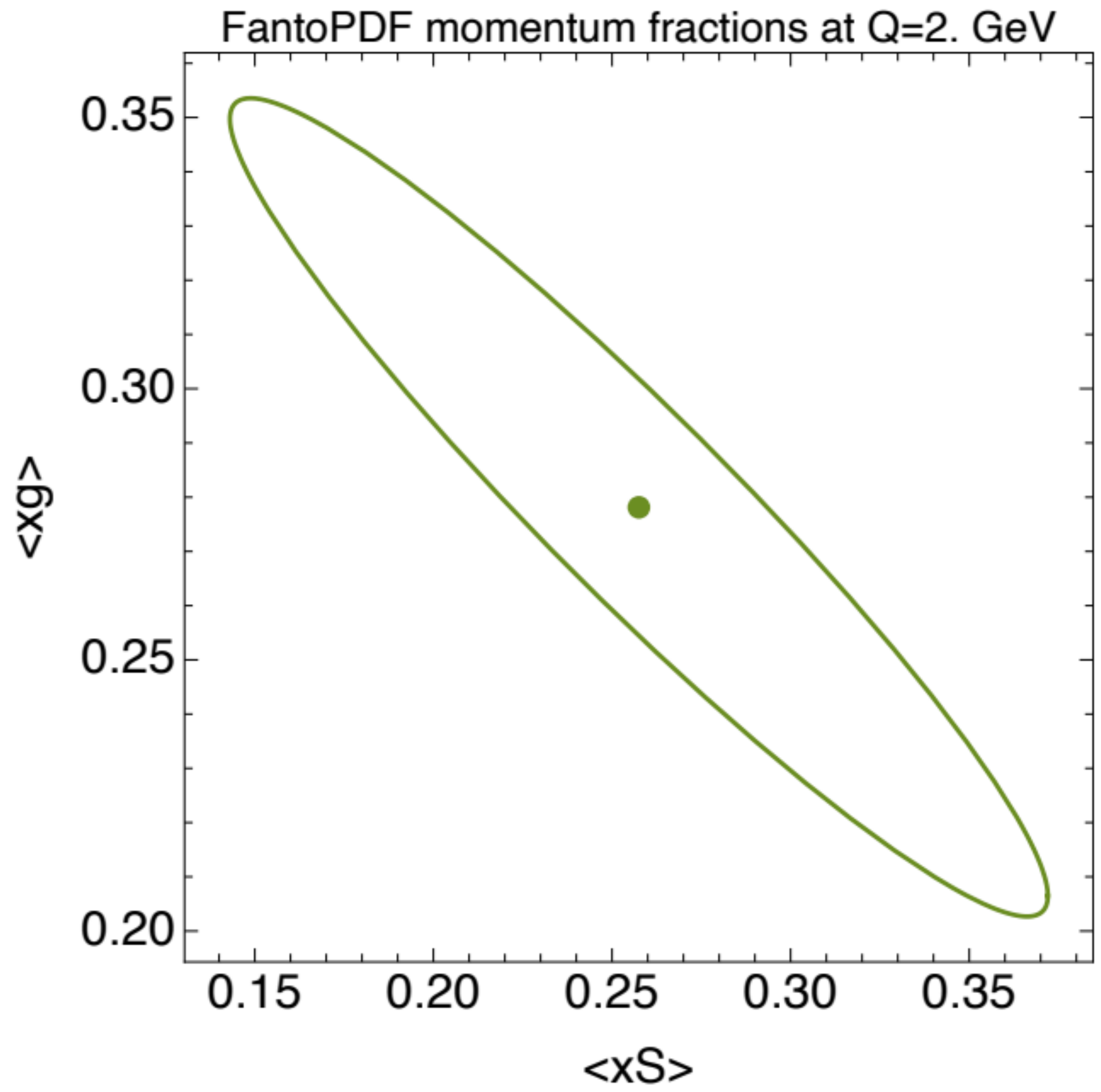
As it turns out, the valence sector was not as exciting as expected — sea and gluon separation got most of our attention!

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

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

We obtain $< xf_g > (Q = 2\,\text{GeV}) = 0.28 \pm 0.08.$

<u>Funny fact:</u> 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.



The Fantômas pion PDF

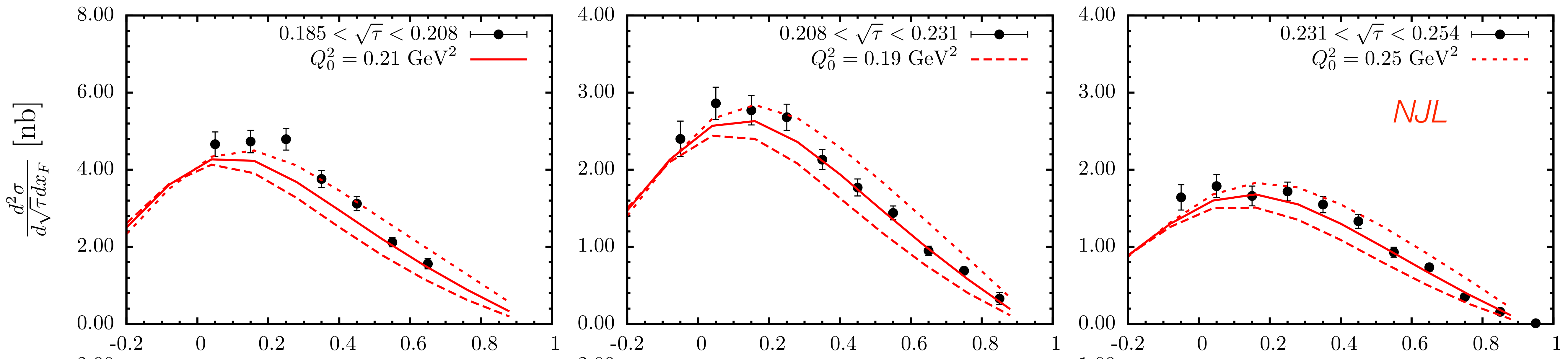
Towards epistemic uncertainty: sampling over parameter space more representative

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?

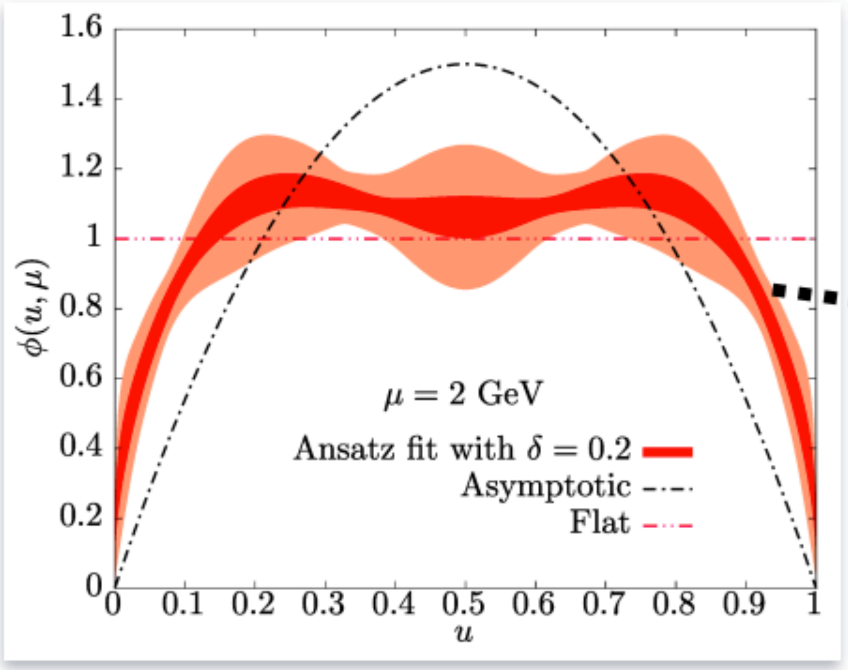
Q_0 plays an important rôle in PDF analyses. Can we improve our understading of the pion from data by varying the phenomenological starting scale? How to update the parameter-fixing of non-perturbative models?



[Ceccopieri et al, EPJC78]

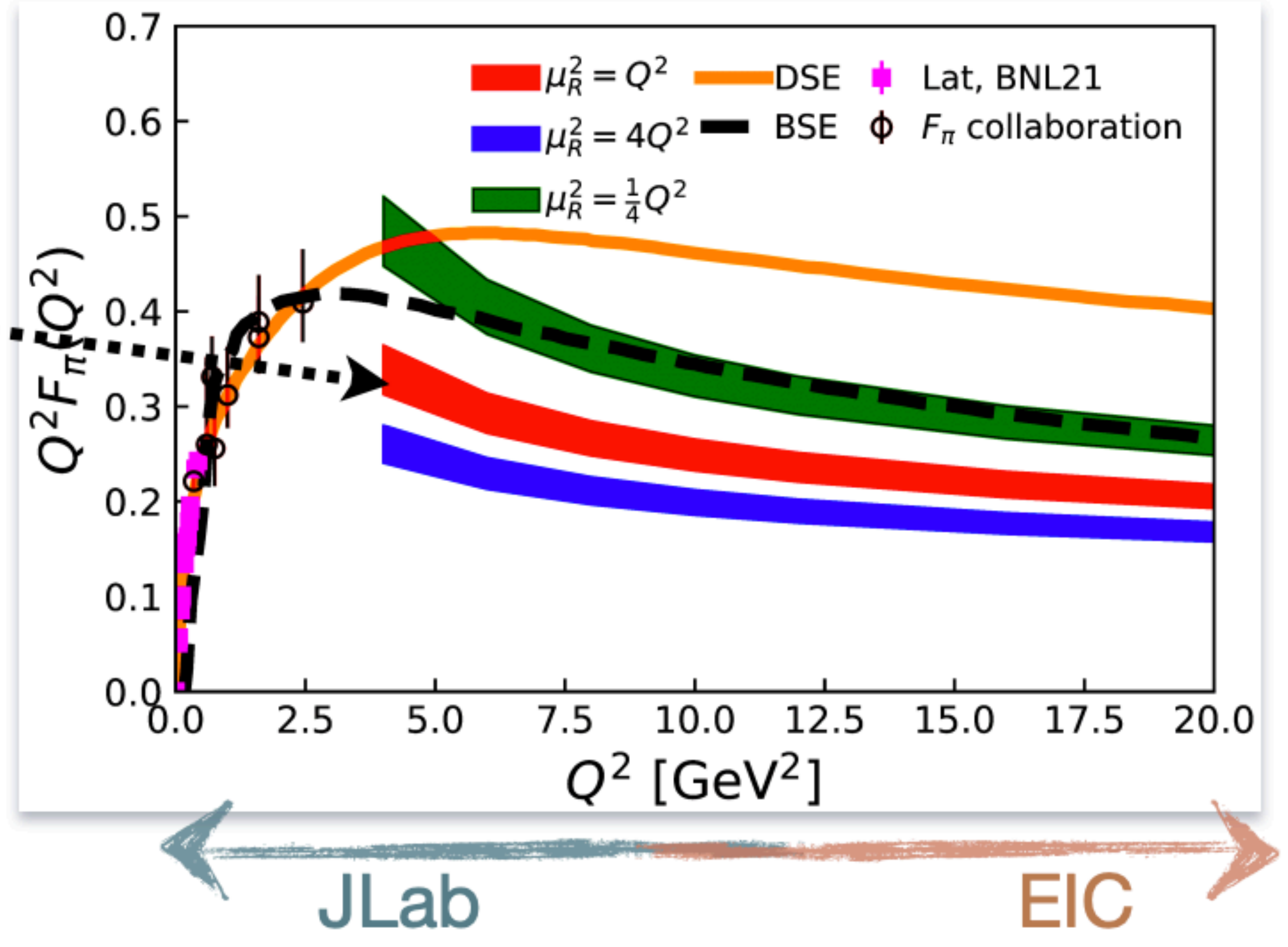
Pion objects and the inverse problem — (one of the many) Fantômas extension(s)

Chiral symmetry seems to control the pion DA over the Q spectrum.

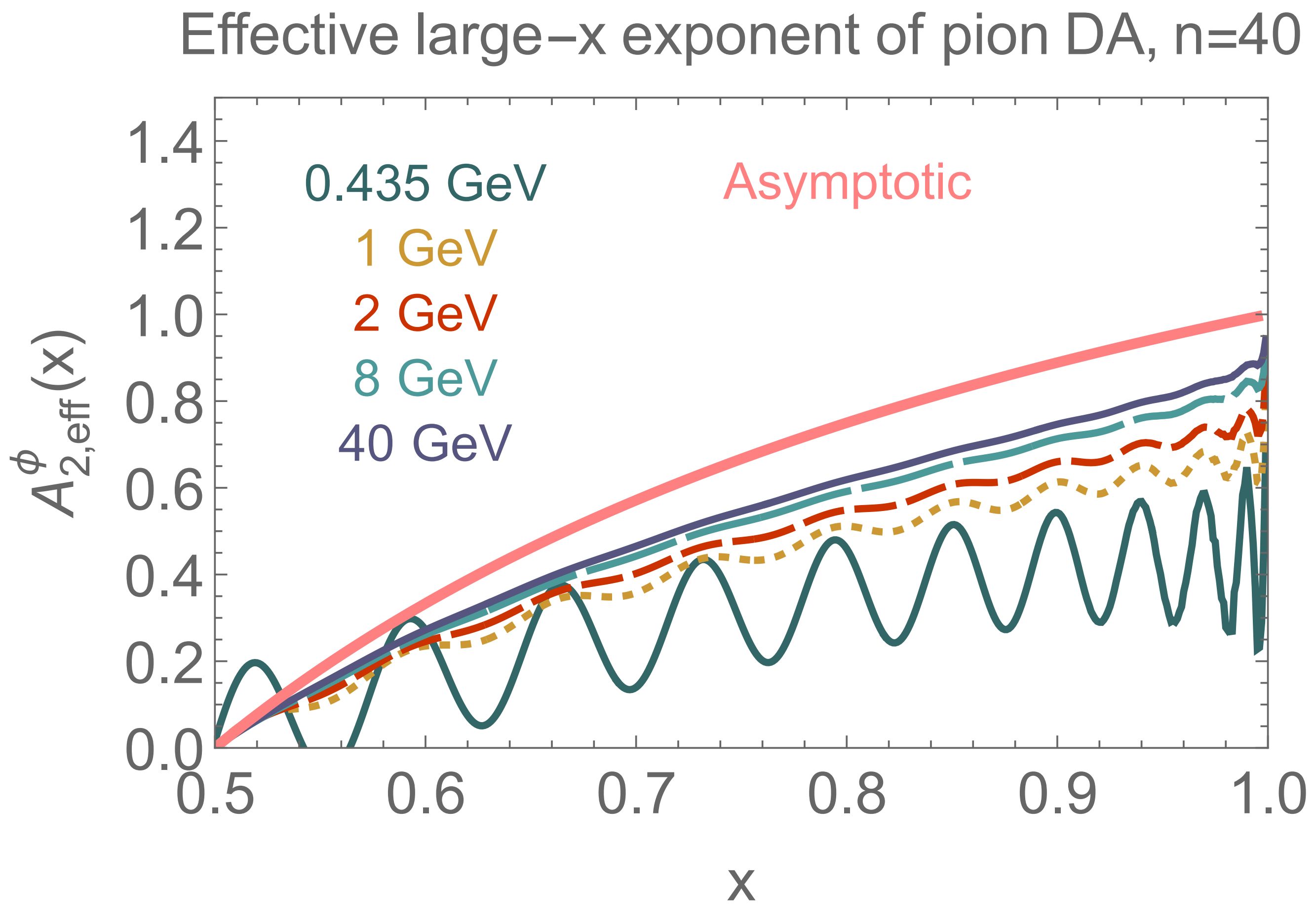


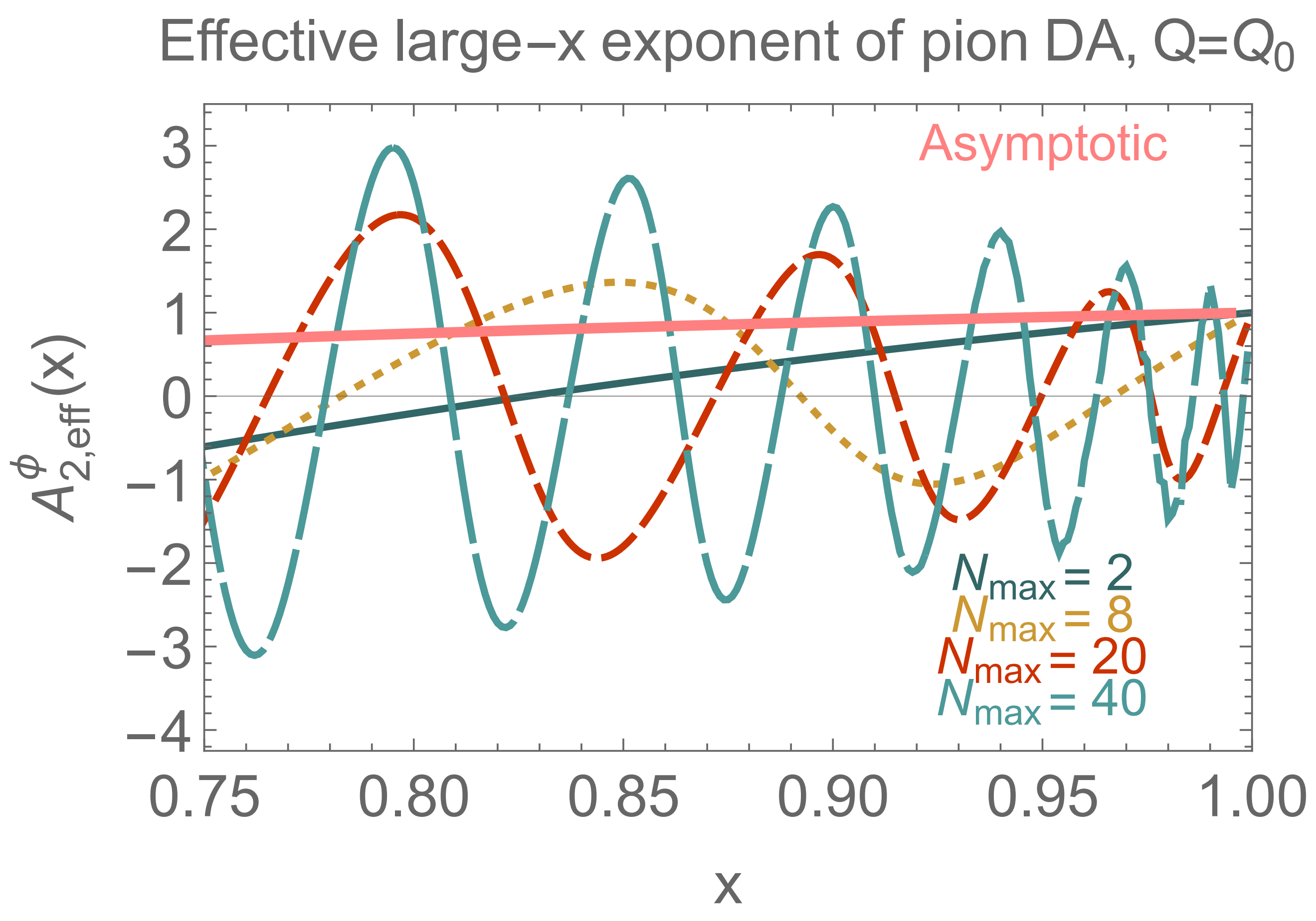
[X. Gao et al., PRD106]

& Swagato Mukherjee at DIS23.



Large-x convergence of the evolved pion DA seems to be key to the problem





[***AC*** *et al, in progress*]

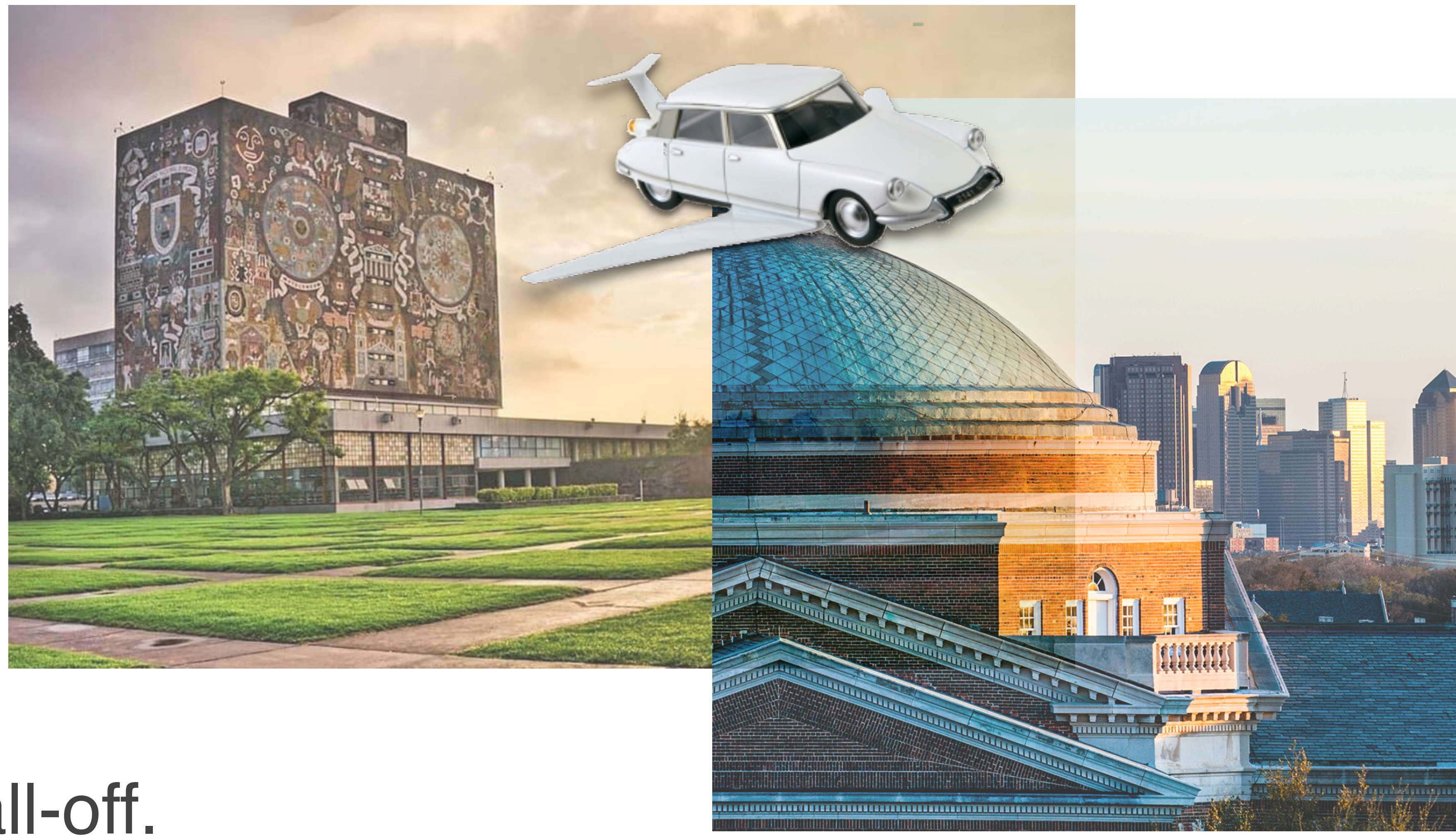
Conclusions

- 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 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.

  Uncertainty quantification in non-perturbative calculations?
  At what Q will pQCD (constraints) take over?


U. P. LINKE


FANTOMAS