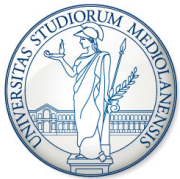


EXPLAINABLE PARTON DISTRIBUTIONS

STEFANO FORTE
UNIVERSITÀ DI MILANO & INFN



UNIVERSITÀ DEGLI STUDI DI MILANO
DIPARTIMENTO DI FISICA



Istituto Nazionale di Fisica Nucleare

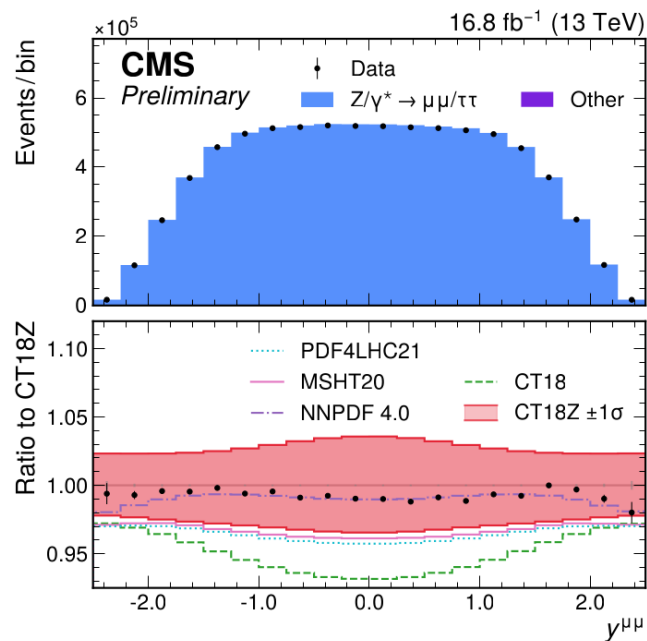
QCD predictions for
ep physics at the EIC

CFNS, Stony Brook,
21 September 2026

THE IMPACT OF PDFs

W MASS DETERMINATION

Parton Distribution Functions



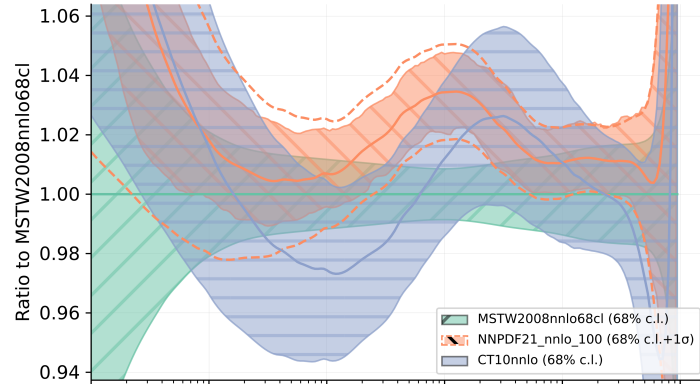
- **Good:** PDF sets are accompanied by uncertainty models with well defined correlations across phase space and between processes
- **Bad:** Different PDFs don't necessarily agree within their uncertainties
- Missing higher order uncertainties, resummation corrections in predictions usually not included
 - Partly mitigated by tolerance factors, etc

- **DOMINANT** SOURCE OF UNCERTAINTY
- DIFFERENT GROUPS **DISAGREE**
- CAN WE **EXPLAIN** THE DISAGREEMENT?

SETTING THE STAGE I PDFS IN TIME THE UP QUARK

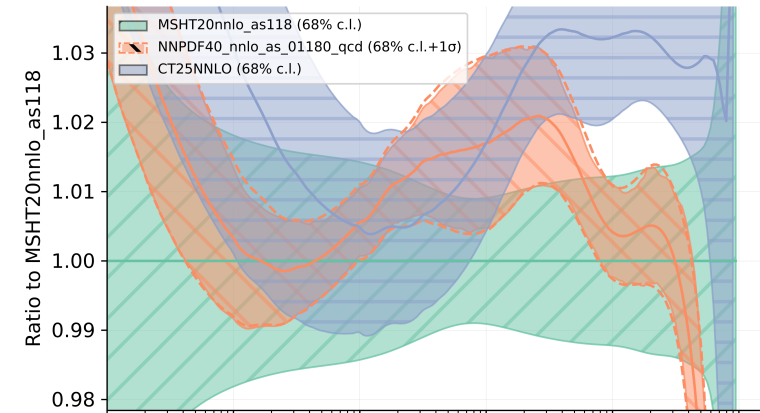
~2012

u at 100.0 GeV

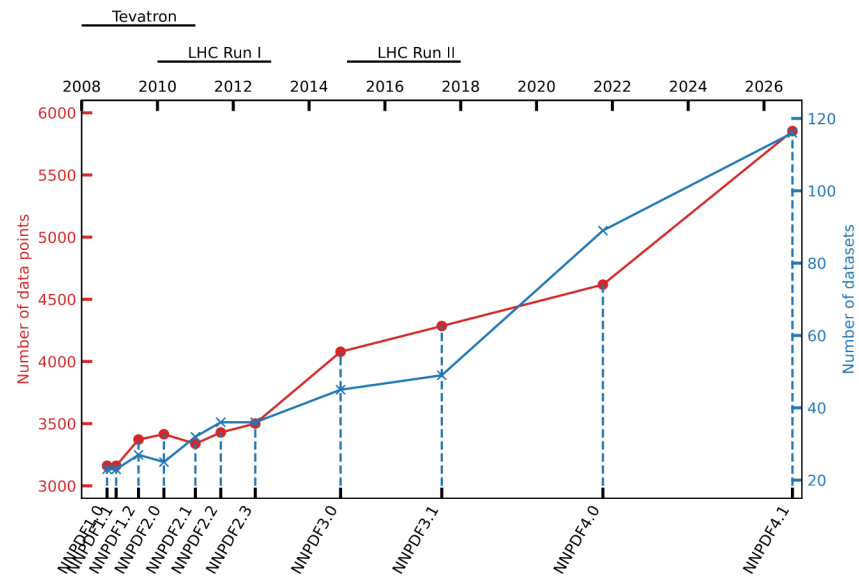


~2022

u at 100.0 GeV



THE DATA



- $\sim \times 2$ REDUCTION IN UNCERTAINTIES
- $\sim \times 1.5$ INCREASE IN # OF DATAPOINTS

SETTING THE STAGE II CURRENT AND FUTURE “GLOBAL” PDF SETS

- **CT**
 - CT18 (NNLO)
 - * release on LHAPDF August 2019; preprint December 2019
 - * paper published January 2021
 - CT25 (NNLO): partially available on CT website since few months
- **MSHT**
 - MSHT20 (NNLO)
 - * release on LHAPDF & preprint December 2020
 - * paper published April 2021
 - N^3 LO
 - * release on LHAPDF & preprint July 2022
 - * paper published March 2023
- **NNPDF**
 - NNPDF4.0 (NNLO)
 - * release on LHAPDF September 2021 & preprint September 2021
 - * paper published May 2022
 - N^3 LO
 - * release on LHAPDF & preprint July 2024
 - * paper published July 2024
 - NNPDF4.1 (N^3 LO): in preparation, due 2027

SOME MORE SPECIALIZED PDF SETS

- **ABMP**: simultaneous fit of QCD & EW parameters
- **ATLASPDF**: cross-correlation between ATLAS datasets
- **JAM**: large x , low scale JLAB data, higher twists

THE PROBLEM:

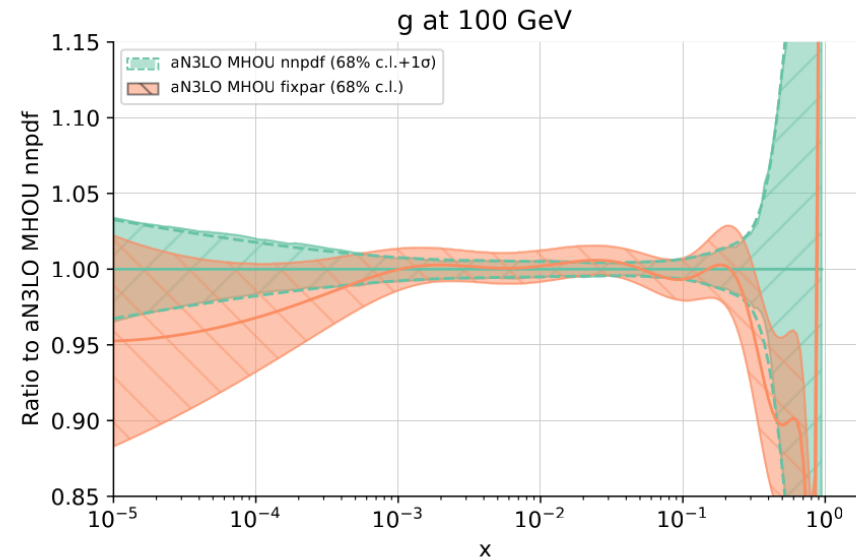
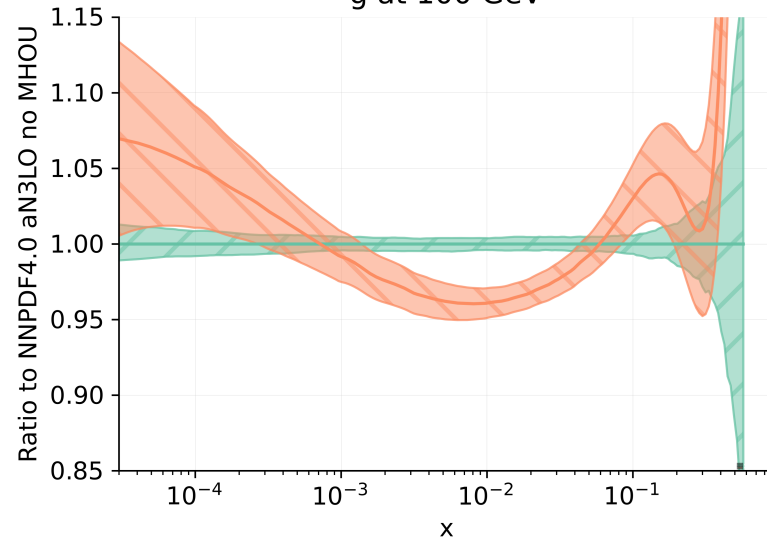
PDF UNCERTAINTIES & THEIR ORIGIN

- **DATA:** CORRELATIONS & CROSS-CORRELATIONS BETWEEN EXPERIMENTS
- **MODEL:** PDF IS ARE THE OUTPUT OF AN ILL-POSED REGRESSION PROBLEM
- **THEORY:** MISSING HIGHER ORDERS, TREATMENT OF HEAVY QUARKS (INCL. PARAMETRIZED CHARM), HIGHER TWISTS AND POWER CORRECTIONS, RESUMMATION, PHOTON PDF & EW CORRECTIONS

PROGRESS
UNCERTAINTY BREAKDOWN
 METHODOLOGY VS. DATA/THEORY
 N^3 LO GLUON **MSHT20** VS **NNPDF4.0**

DIFFERENT METHOD., SAME DATA/THEORY

NATIVE PDFS
 g at 100 GeV



(Cruz-Martinez, Giani, Harland-Lang, 2026)

- **NATIVE:** PUBLISHED SETS
- **FIX:**
 - **SAME (NNPDF) DATA/THEORY:** N^3 LO TH, FITTED CHARM, MHOUS, FONLL. . .
 - **DIFFERENT METHODOLOGY:** TOLERANCE, HESSIAN VS MC, POLY VS NN. . .

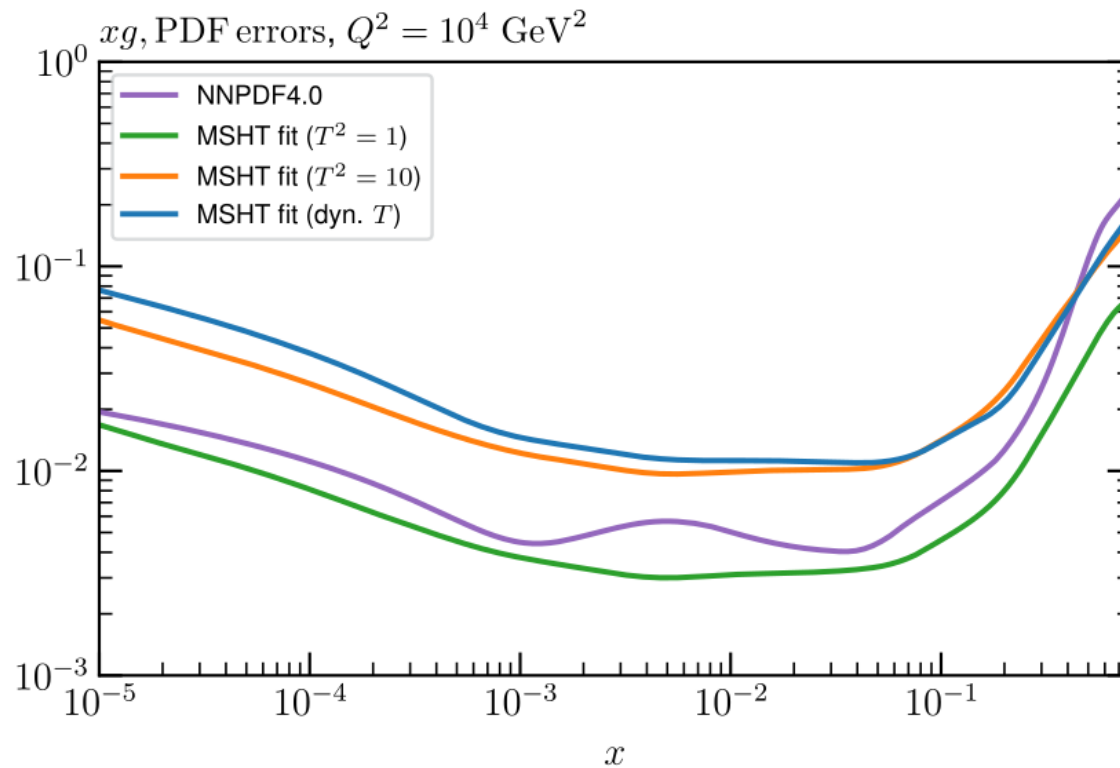
⇒ **GOOD AGREEMENT WITHIN UNCERTAINTY!**

METHODOLOGY & DATA

THE TOLERANCE PROBLEM

- HESSIAN PDF UNCERTAINTY (CT, MS^{HT}) $\Rightarrow$ TOLERANCE
- $\Delta\chi^2 = T^2$, ACCOMMODATE FLUCTUATIONS BETWEEN DATASETS
- DYNAMICAL: DIFFERENT T ALONG DIFFERENT HESSIAN EIGENVECTORS

MSHT TOLERANCE IMPACT VS. NNPDF



(Harland-Lang, Cridge, Thorne, 2025)

fixed NNPDF dataset, native NNPDF vs. MSHT methodology & theory

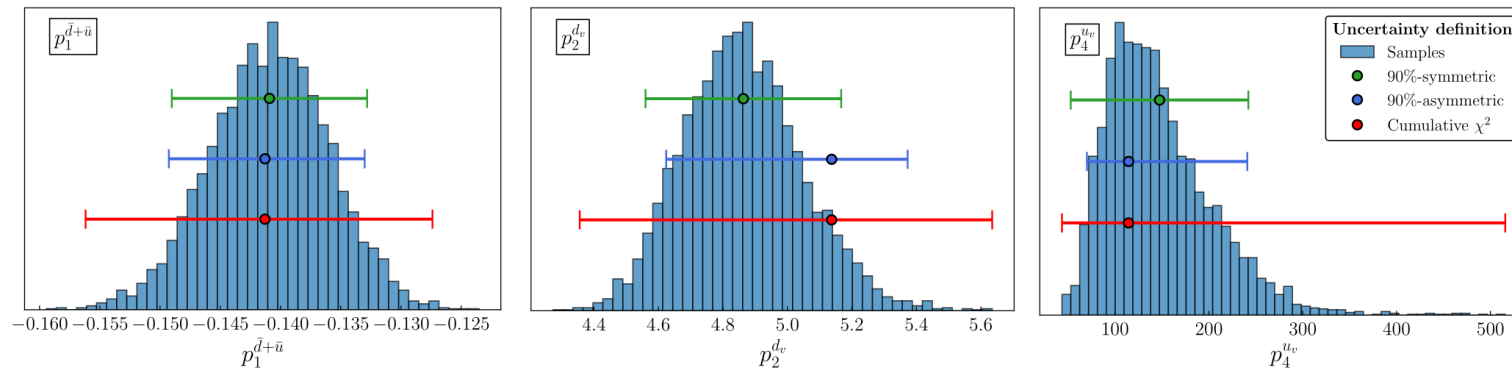
WHAT'S THE PROBABILITY DISTRIBUTION?

- PERFORM **BAYESIAN INFERENCE** FROM DATA USING SIMPLE PARM & PRIOR
- DETERMINE **POSTERIOR PROBABILITY** DISTRIBUTION (MCMC) (JAM; Kusina et al.; Ubiali et al.; Costantini; Colibri)

PARAMETRIZATION

$$x f_i(x, Q_0) = p_0^i x^{p_1^i} (1 - x)^{p_2^i} (1 + p_3^i \sqrt{x} + p_4^i x)$$

PROBABILITY DISTRIBUTIONS



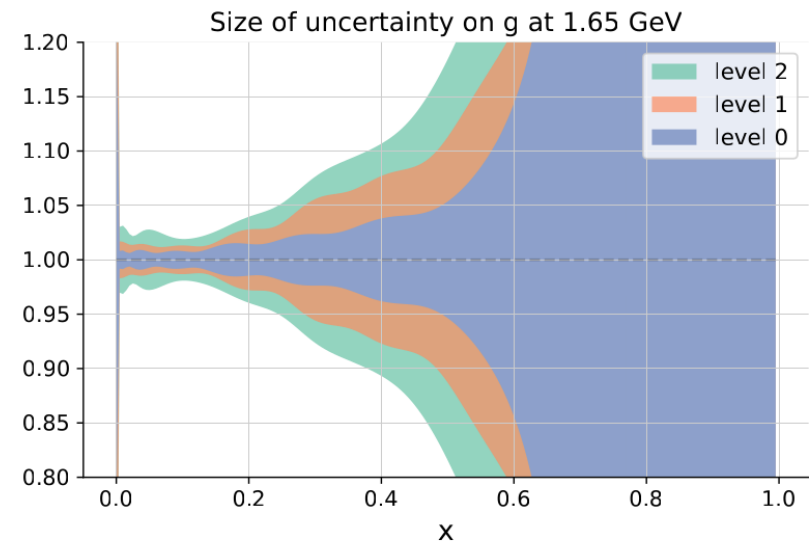
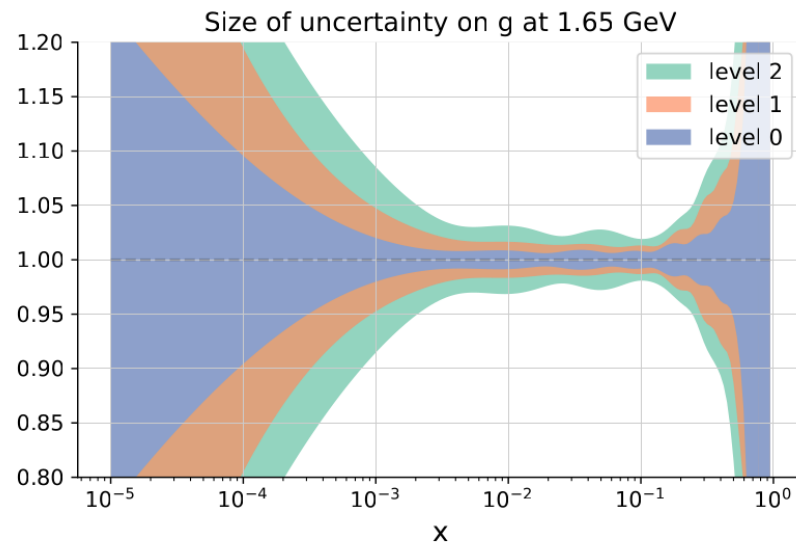
(Risse, Derkashanian, Ježo, Kovařík, Kusina, 2026)

- SOME PARAMETERS **NONGAUSSIAN**
- **EFFECTIVE TOLERANCE** $T^2 \sim 10$

ILL-POSED REGRESSION: THE FUNCTIONAL UNCERTAINTY

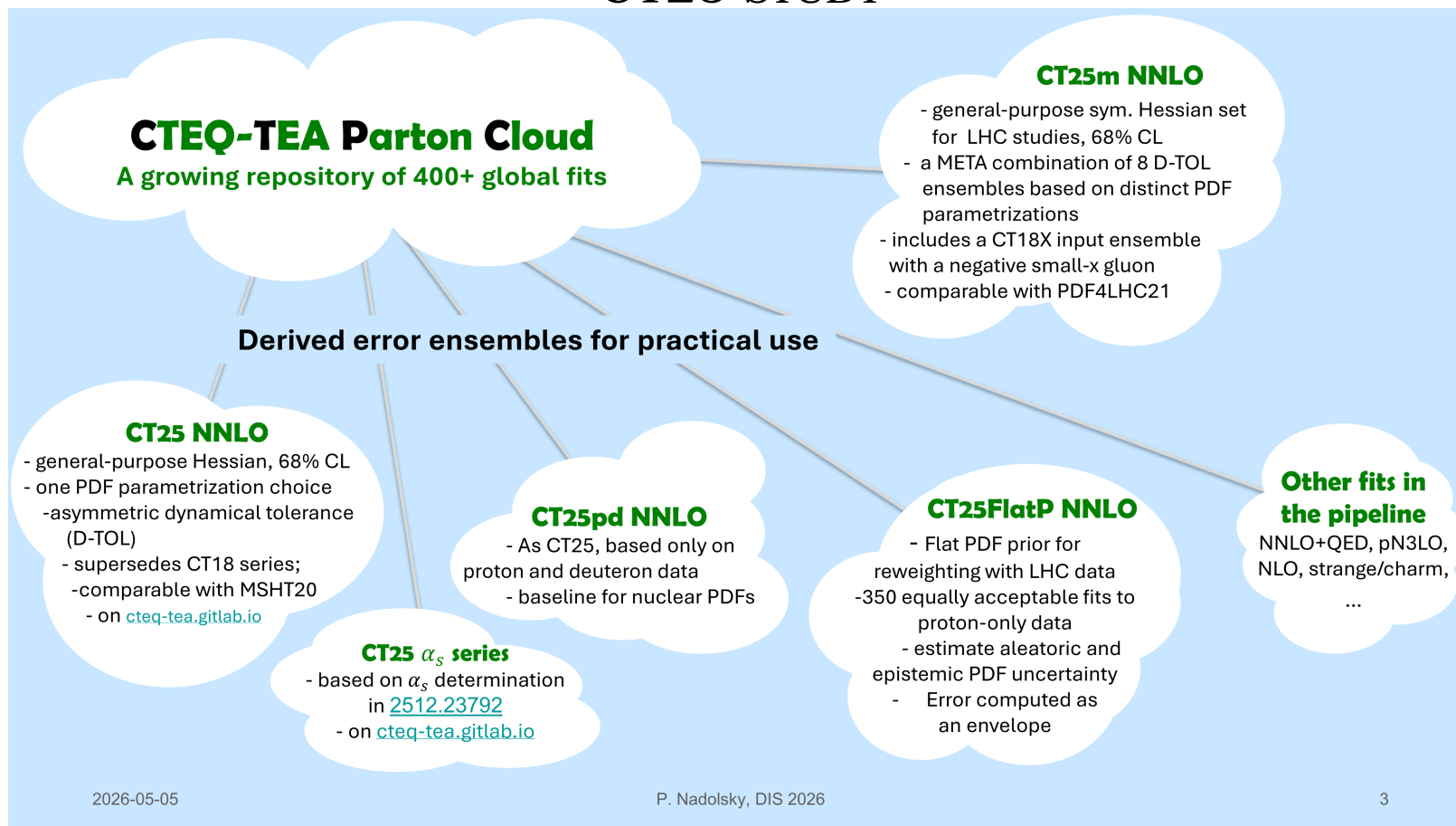
- INTERPOLATION/EXTRAPOLATION
- FUNCTIONAL: CHOICE OF FUNCTIONAL FORM
- DATA

NNPDF4.0 CLOSURE TEST BREAKDOWN



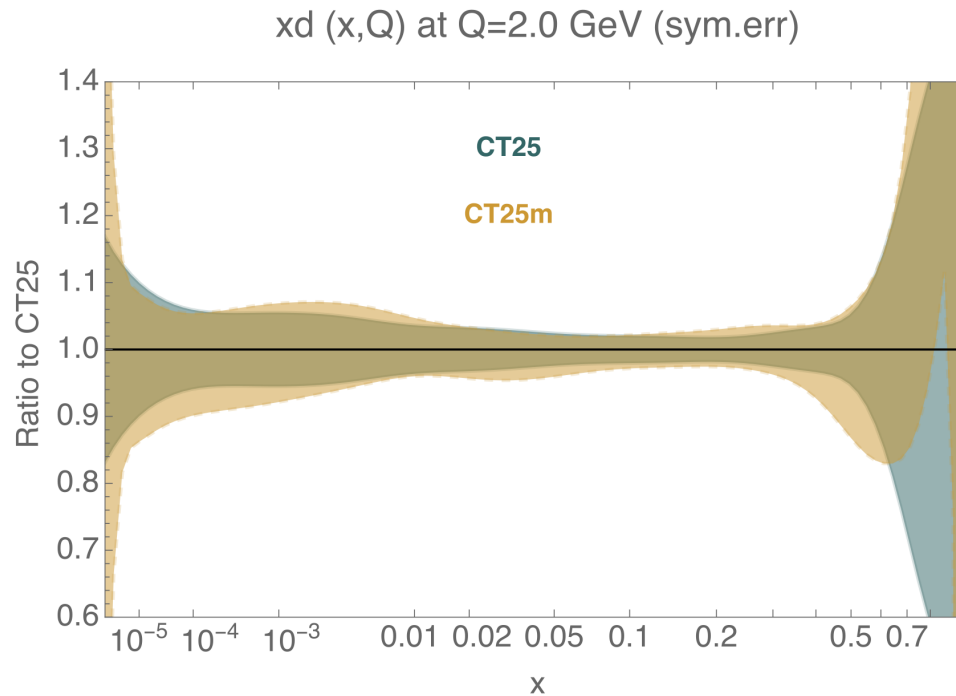
(Del Debbio, Giani, Wilson, 2022)

THE FUNCTIONAL UNCERTAINTY & TOLERANCE CT25 STUDY



(Nadolsky, DIS 2026)

ENSEMBLING VS. TOLERANCE



(Courtoy, DIS 2026)

TOLERANCE LARGELY ACCOUNTS FOR FUNCTIONAL UNCERTAINTY
QUESTIONS:

- CAN WE DO BETTER?
- CAN WE DISENTANGLE:
 - THE IMPACT OF THE DATA? $\Rightarrow$ DATA-FEATURE MAP
 - THE IMPACT OF THE METHDOLOGY? $\Rightarrow$ MODEL FEATURE MAP

TOWARDS EXPLAINABLE PDFs

THE SHAPLEY VALUE

A GAME-THEORETIC CONCEPT

INGREDIENTS

- A SET OF N **PLAYERS** f_j
- ALL 2^{N-1} **COALITIONS** OF $N - 1$ PLAYERS NOT INCLUDING THE i TH PLAYER
- A **VALUE FUNCTION** v

DEFINITION: MARGINAL CONTRIBUTION OF i -TH PLAYER

- VALUE FOR ALL SUBSETS OF S OUT OF $N - 1$ COALITIONS NOT INCLUDING i -TH PLAYER: $v(S)$
- CHANGE IN VALUE WHEN i TH PLAYER INCLUDED $[v(S \cup \{i\}) - v(S)]$
- WEIGH WITH NUMBER OF SUBSETS $w(S) = \left[N \binom{N-1}{S} \right]^{-1} = \frac{S! (N-S-1)!}{N!}$

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} w(S) [v(S \cup \{i\}) - v(S)]$$

XAI USAGE

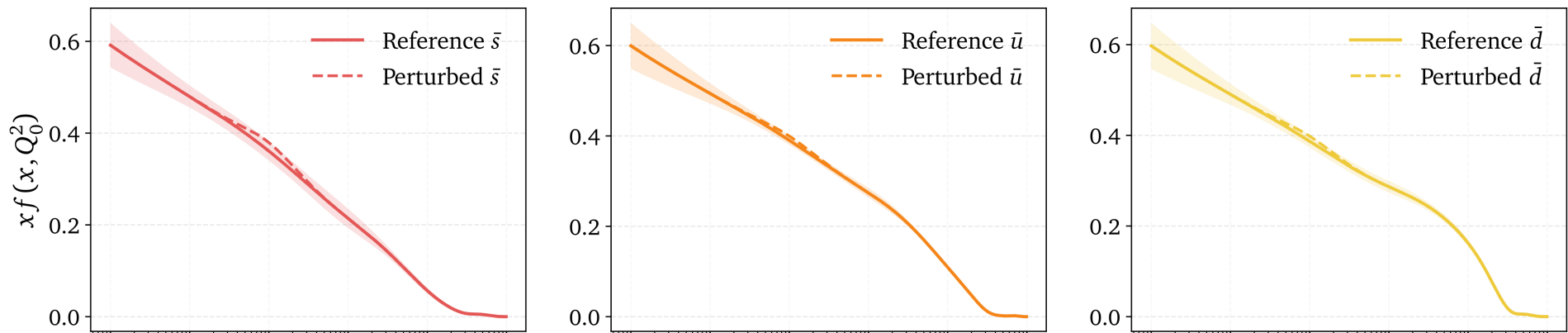
- **PLAYERS** $\Rightarrow$ **FEATURES**
- **GROWTH** 2^{N-1} COALITIONS WITH N FEATURES $\Rightarrow$ **ADDITIVE** (LINEARIZATION)

SHAPLEY FOR PDFs

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} w(S) [v(S \cup \{i\}) - v(S)]$$

- COALITIONS $\Rightarrow$ **PERTURBED** PDF: $f_j^{(S)} = \begin{cases} f_j^{\text{PERT}}, & j \in S, \\ f_j^{(0)}, & j \notin S, \end{cases}$

- FEATURES $\Rightarrow$ PDF **BUMP**: $\delta_j^\pm(x) = \pm \Delta_j^\pm(x_\mu); \exp \left[-\frac{(\log_{10}(x) - \log_{10}(x_\mu))^2}{2w^2} \right]$



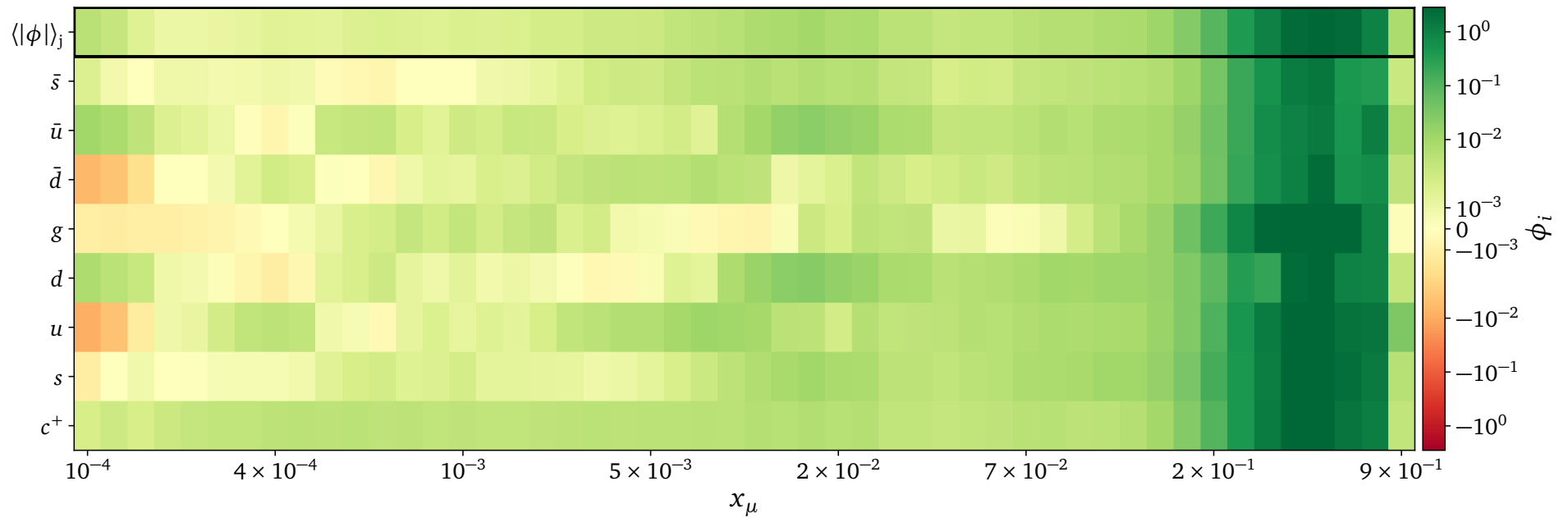
- VALUE $\Rightarrow$ **AGREEMENT** W. DATA $v(S) = \frac{1}{2} [\chi^2(f^{(S+)}; \mathcal{D}) + \chi^2(f^{(S-)}; \mathcal{D})]$
- $N = 8$ **FLAVORS** $\Rightarrow 2^7 = 128$ **COALITIONS** $\Rightarrow$ **EXACT**

WHAT IS THE PDF SHAPLEY INDEX GOOD FOR?

- INDEX **AGNOSTIC** TO PDF SET, DATA:
- EXISTING DATA & PDF $\Rightarrow$ **SENSITIVITY**
- NEW DATA & EXISTING PDF $\Rightarrow$ **IMPACT**

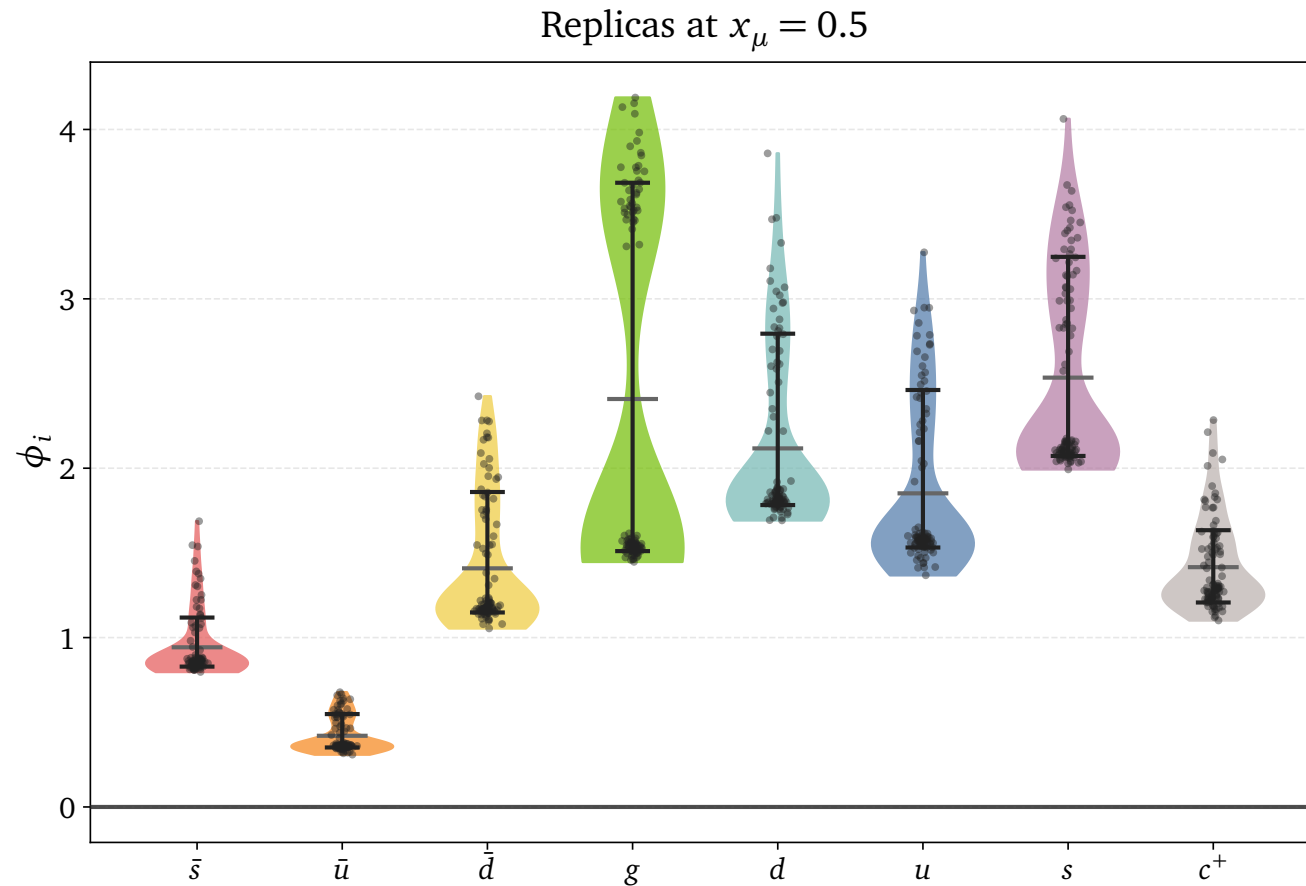
GLOBAL PDF IMPACT

NNPDF4.0 PDFs & DATASET



- VANISHES OUTSIDE DATA REGION
- GLUON $\rightarrow$ SMALLER x
- QUARKS MORE SENSITIVE THAN ANTIQUARKS

EXPLORING THE PROBABILITY DISTRIBUTION



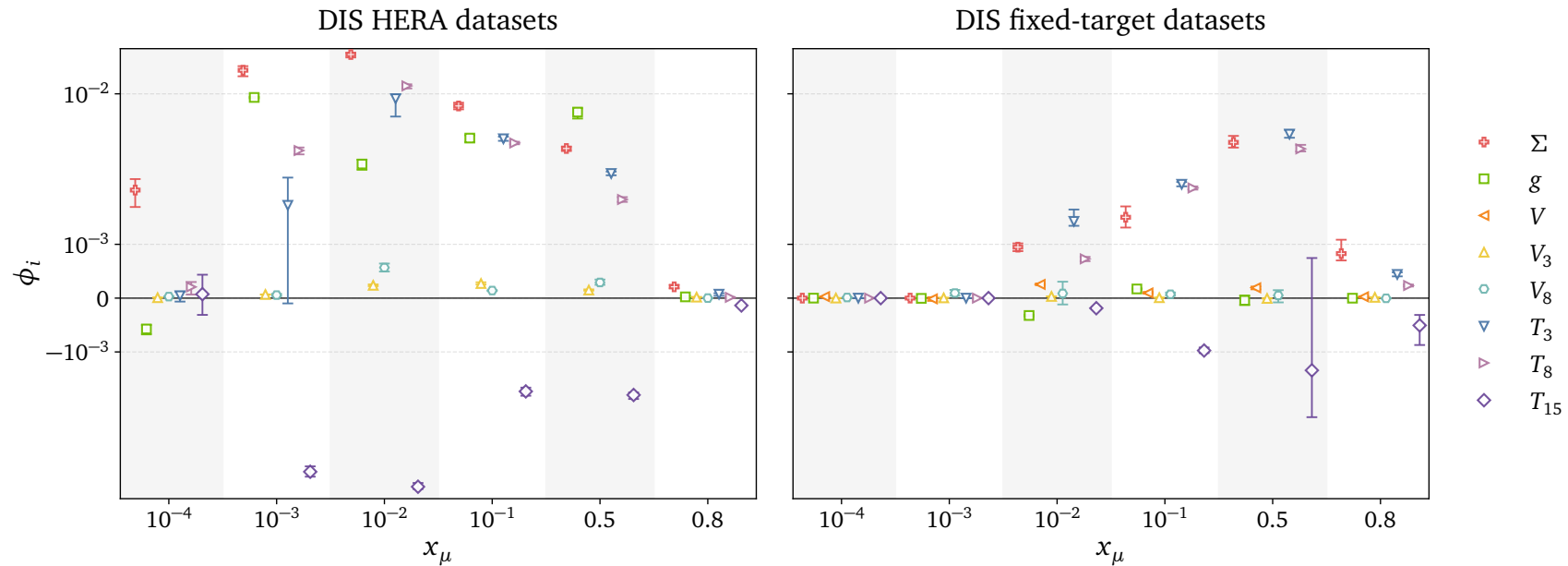
- INDIVIDUAL PDF REPLICAS $\Rightarrow$ DISTRIBUTION
- NONLOCAL CONSTRAINTS (SUM RULES) $\Rightarrow$ NONGAUSSIAN

DATASET IMPACT

- CHOICE OF DATASET
- CHOICE OF PDF BASIS

DEEP-INELASTIC SCATTERING

DEEP-INELASTIC SCATTERING, EVOLUTION EIGENSTATE BASIS



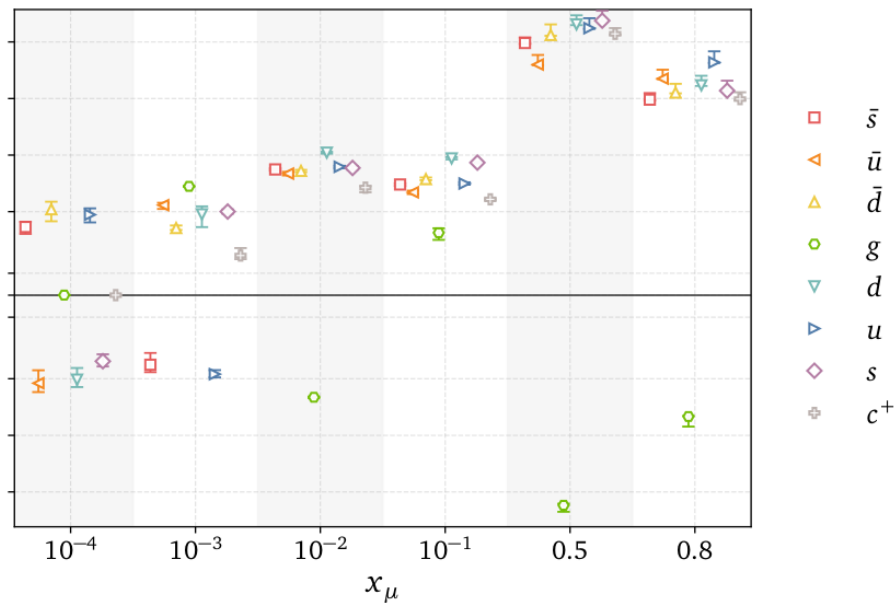
- HERA:
 - SMALLER x
 - MOSTLY GLUON AND SINGLET
 - LOW SENSITIVITY TO NONSINGLET VALENCE
- FIXED TARGET:
 - LARGER x
 - p, d NC $\Rightarrow$ TRIPLET T_3

DATASET IMPACT HADRONIC PROCESSES

- DRELL-YAN RAPIDITY:**
 - WIDE** x RANGE FOR QUARKS
 - SURPRISINGLY **LARGE GLUON** IMPACT AT LARGE x
- JETS:**
 - LARGEST **GLUON** IMPACT
 - MEDIUM-**LARGE** x

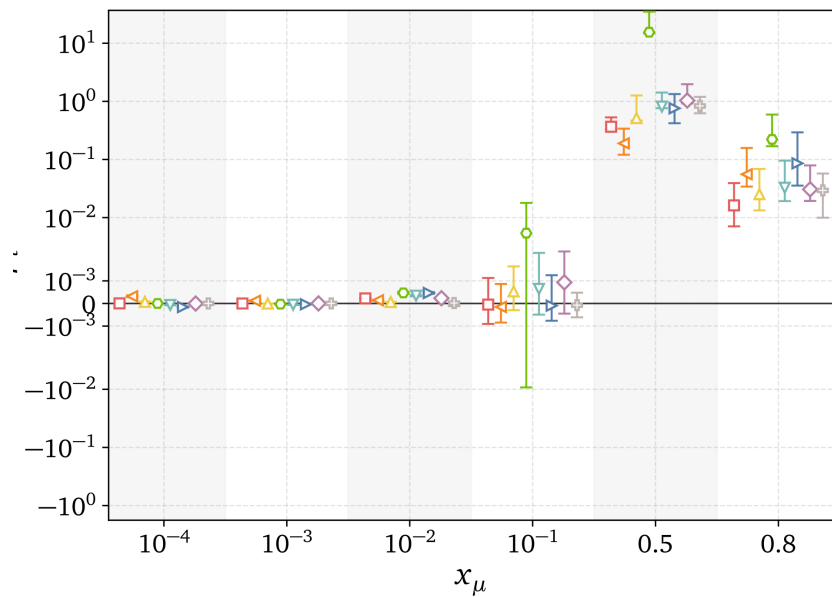
DRELL-YAN RAPIDITY

DY y datasets



JETS AND DIJETS

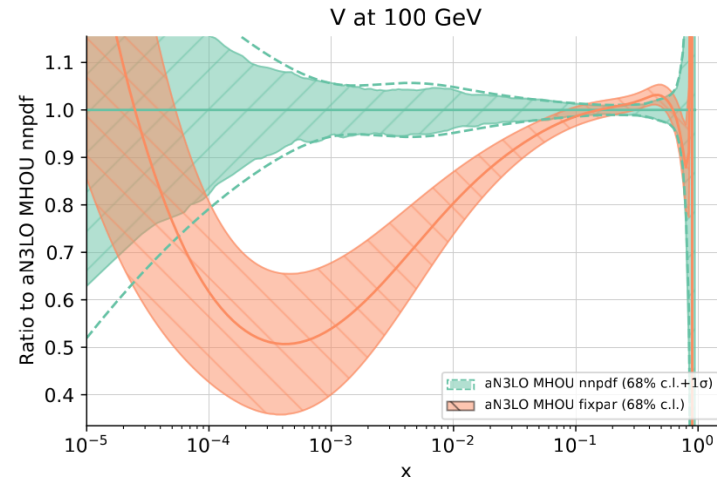
Jet datasets



HIDDEN ASSUMPTIONS

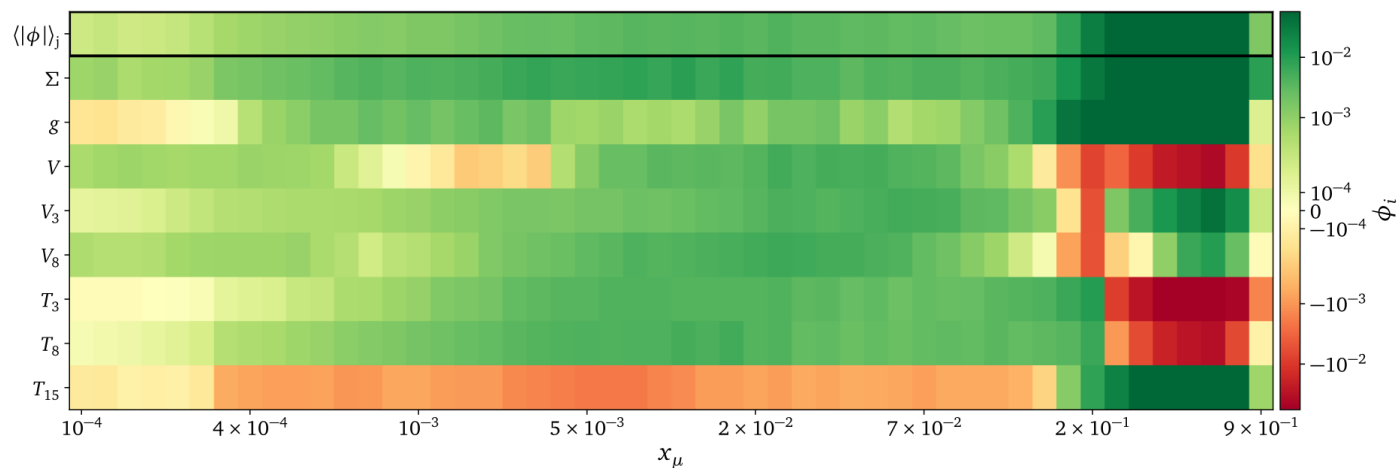
DIFFERENT METHODOLOGY, SAME DATA/THEORY AGAIN

THE TOTAL VALENCE (N³LO)



(Cruz-Martinez, Giani, Harland-Lang, 2026)

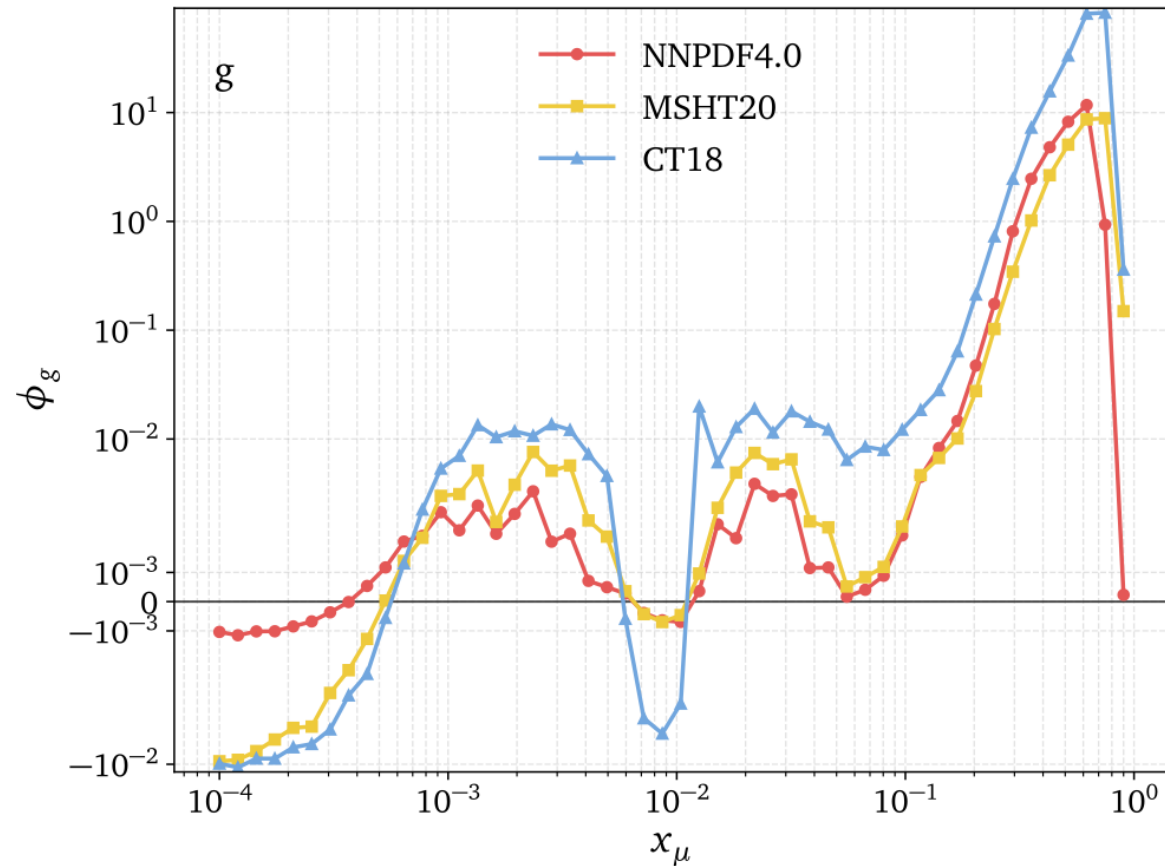
THE SHAPLEY ANALYSIS



- VALENCE **LOSS OF SENSITIVITY**
- DIFFERENT METHODOLOGIES **LOCK ONTO DIFFERENT MININIMA**

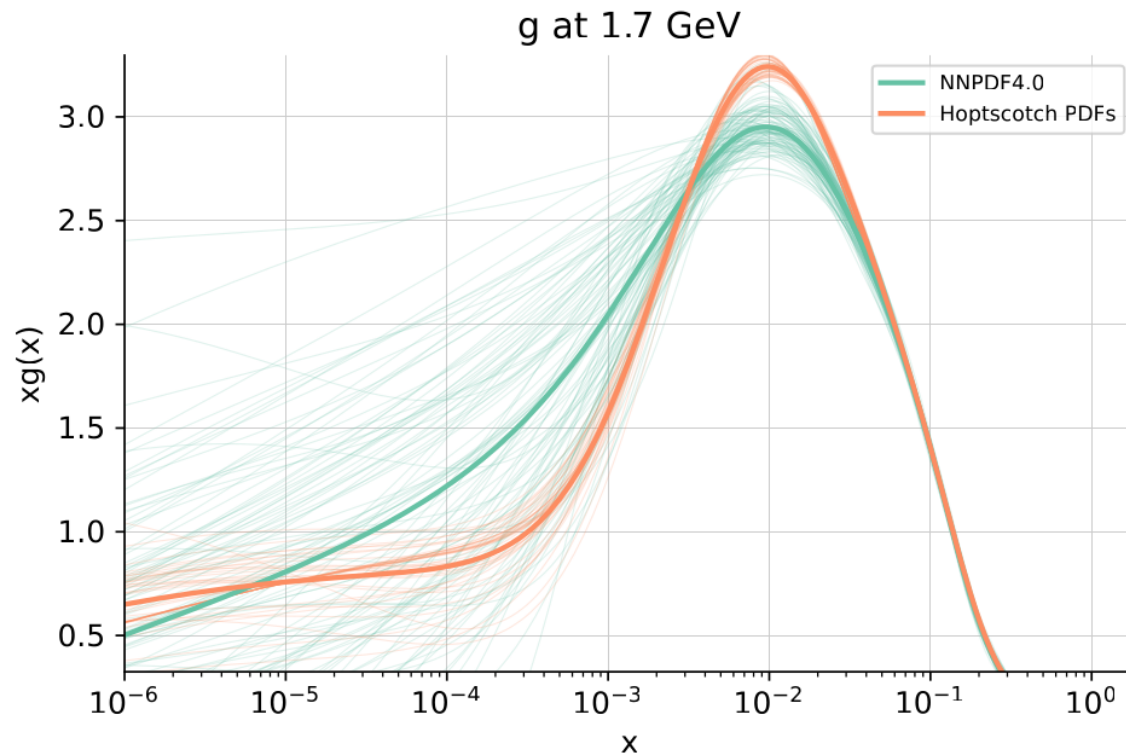
THE IMPACT OF ASSUMPTIONS

GLUON (NNPDF4.0 DATASET)



(Bonnet-guerrini, Carrazza, SF, Groenendijk, Piuri, Winterhalder , DIS 2026)

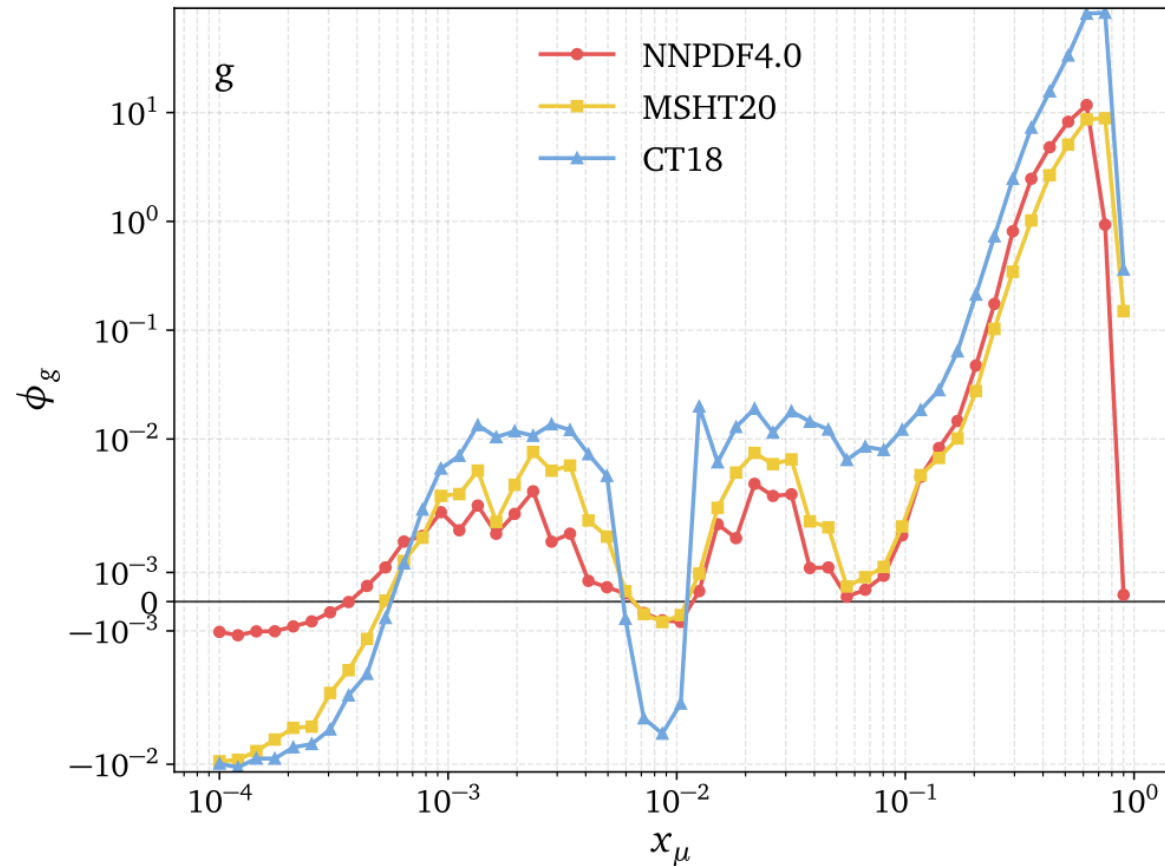
- **LOSS** OF SENSITIVITY FOR $x \sim 10^{-2}$
- **SEEN** IN “HOPSCOTCH” PDFs (Courtoy et al., 2022)



(NNPDF, 2022)

- NNPDF $\Rightarrow$ BUMP **DISFAVORED** BY ML MODEL (OVERLEARNING)
- CT/MSHT $\Rightarrow$ BUMP **FORBIDDEN** BY FUNCTIONAL FORM

THE IMPACT OF THEORY CONSTRAINTS

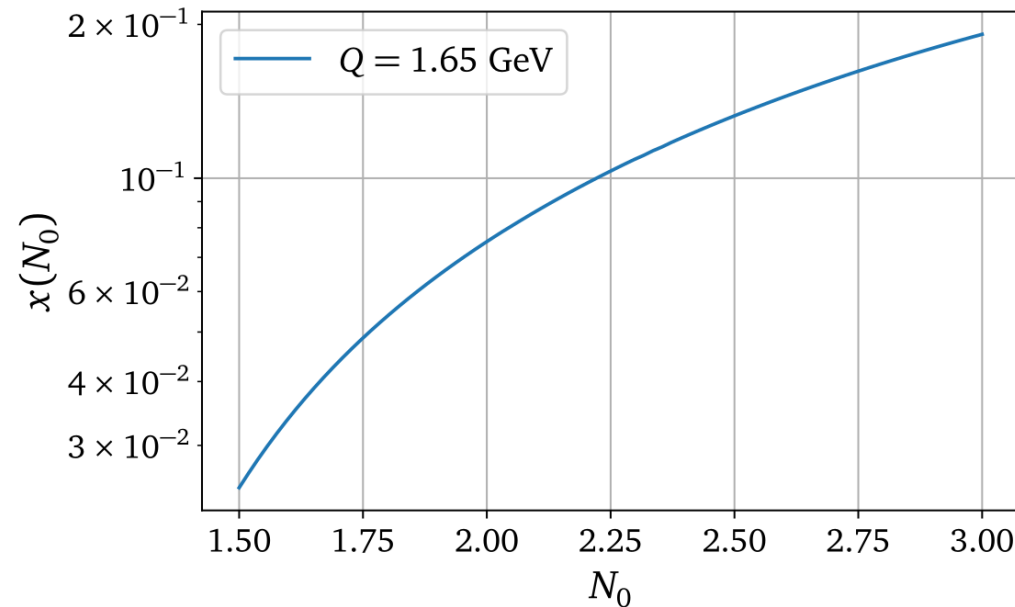


(Bonnet-guerrini, Carrazza, SF, Groenendijk, Piuri, Winterhalder , DIS 2026)

- **LOSS** OF SENSITIVITY FOR $x \sim 0.07$
- **EXPLAINED** BY MOMENTUM SR

x VALUE CORRESPONDING TO N BY SADDLE

Saddle point of gluon PDF



- ANOMALOUS DIMENSION MATRIX $\gamma_{ij}(2) \Rightarrow$ ZERO EIGENVALUE
- FLAT DIRECTION (GLUON)
- $N = 2$ CORRESPONDS TO $x = 0.07$ BY SADDLE MELLIN INVERSION

LESSON LEARNT

- A BSM BUMP COULD BE HIDDEN
- DIFFERENT DATA NEEDED

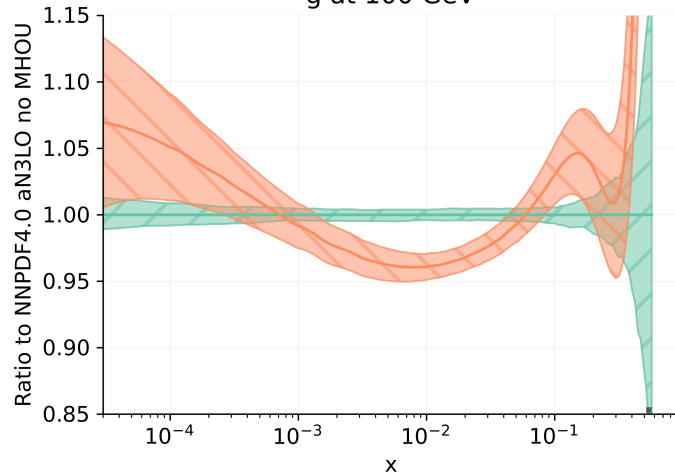
EXPLAINING THEORY UNCERTAINTIES

THE IMPACT OF $N^3\text{LO}$

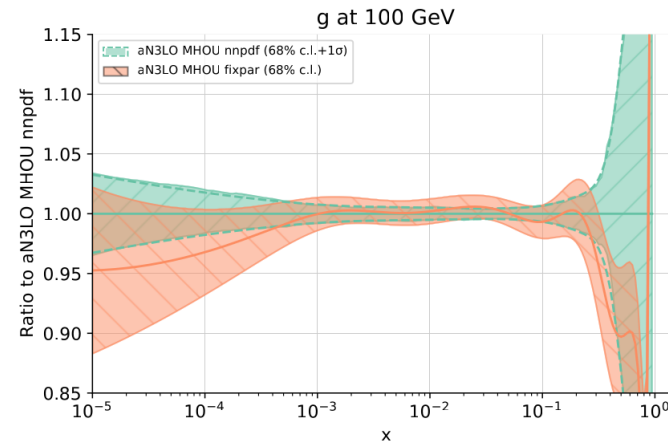
MSHT20 VS NNPDF4.0 GLUON, METHODOLOGY VS. DATA/THEORY AGAIN

$N^3\text{LO}$

NATIVE PDFS
g at 100 GeV

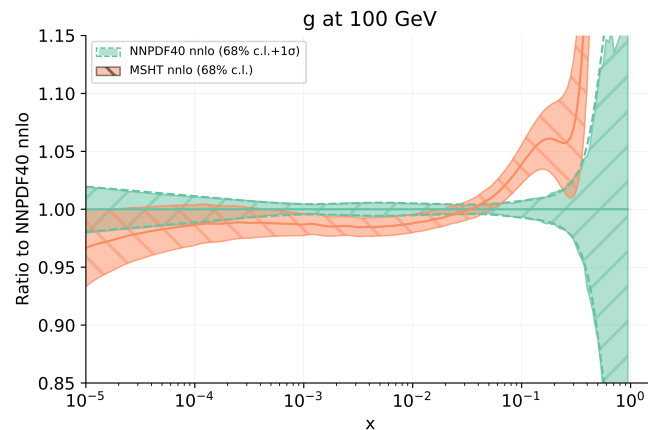


DIFFERENT METHOD., SAME DATA/THEORY

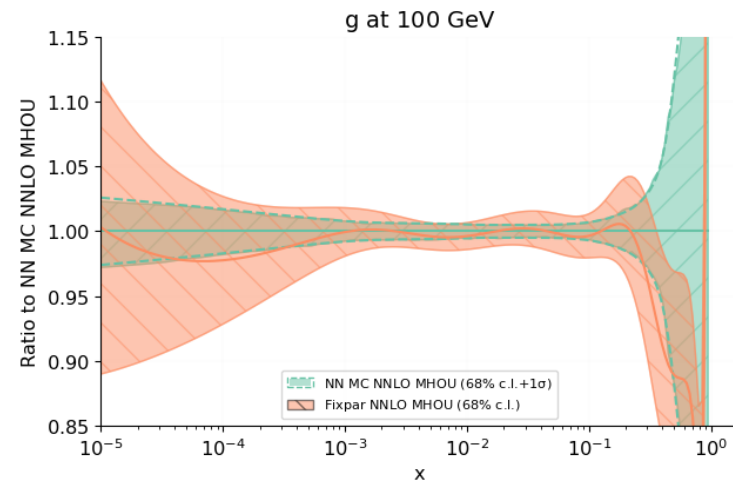


NNLO

(NO MHOU)



(WITH MHOU)

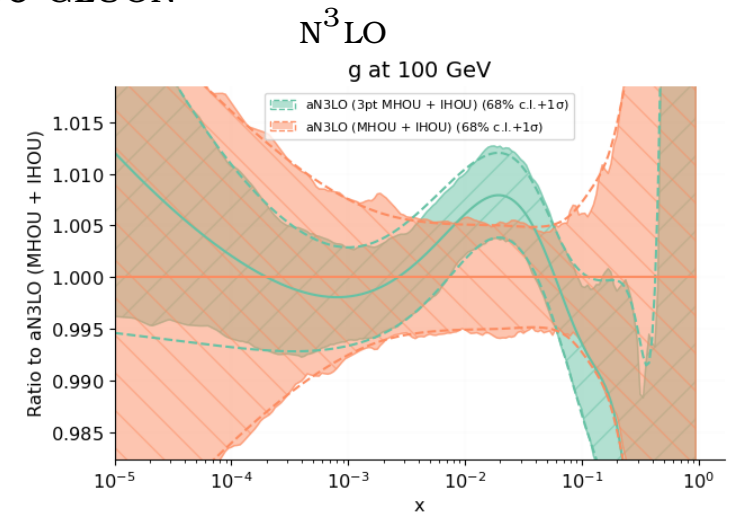
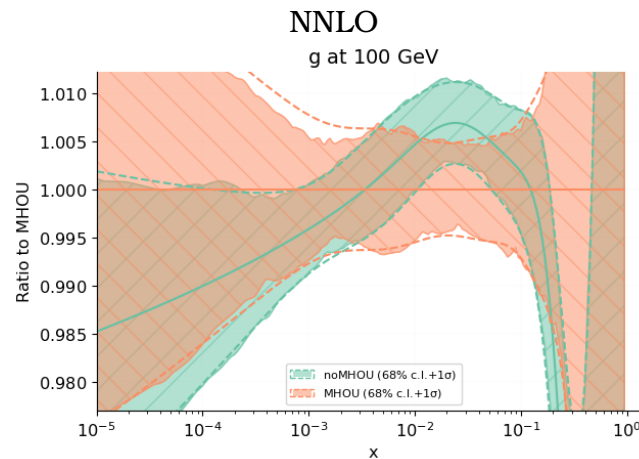
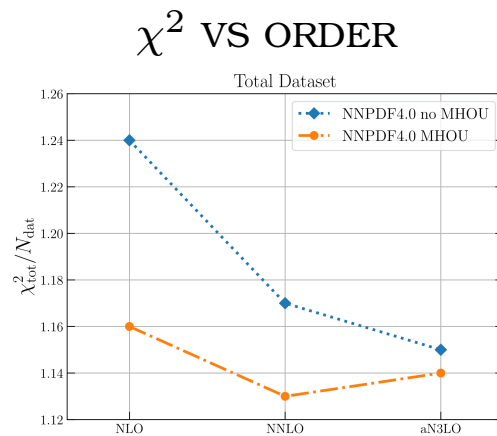


- $N^3\text{LO}$ DISCREPANCY LARGER THAN NNLO $\Rightarrow$ 2022 APPROXIMATION
- CURRENT $N^3\text{LO}$ THEORY EXACT: DIS, EVOLUTION, MATCHING;
HADRONIC PROCESSES MISSING $\Rightarrow$ MHOU

THE IMPACT OF MHOUS

- MSHT
 - INCLUDED ONLY AT N³LO THROUGH NUISANCE PARAMETERS
 - UNCERTAINTIES ESTIMATED BASED ON FORM OF MISSING TERMS
- NNPDF
 - INCLUDED AT ALL PERTURBATIVE ORDERS (2025) THROUGH COVMAT
 - MHOU COVMAT ESTIMATED THROUGH SCALE VARIATION
 - N³LO IHOU COVMAT FROM APPROX. SPREAD
- OTHER PDF SETS: NOT INCLUDED

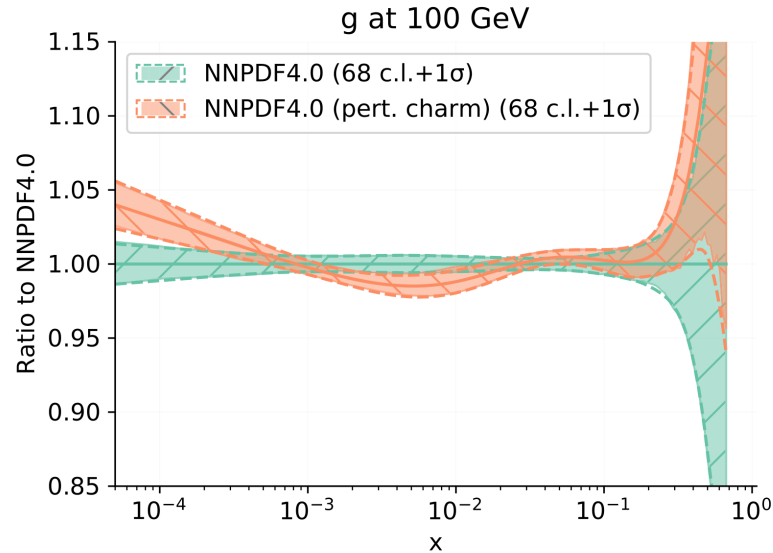
NO MHOU RATIO TO MHOU NNPDF4.0 GLUON



- IMPACT ON CENTRAL VALUE
- MHOU: N³LO SMALLER THAN NNLO
- N³LO: IHOU SMALLER THAN MHOU
- N³LO REQUIRED FOR HADRONIC PROCESS CONSISTENCY

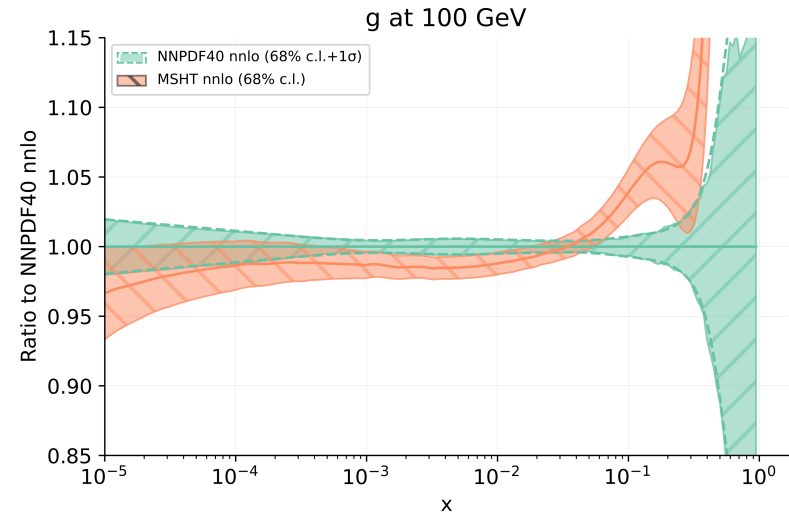
THE IMPACT OF FITTING CHARM

NNPDF4.0 FITTED VS PERT. CHARM



RECALL:

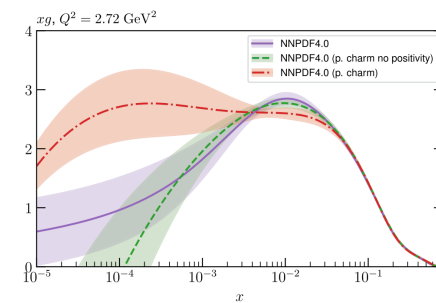
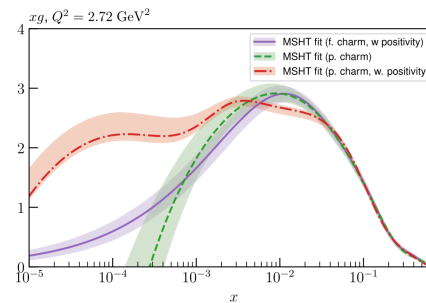
NNPDF4.0 VS MSHT20 NNLO NO MHOU



- **FITTED** CHARM $\Leftrightarrow$ **MISSING** HO **MATCHING**
- **EXPLAINS** RESIDUAL MSHT/NNPDF **DIFFERENCE** IN $x \gtrsim 0.003$ REGION

GLUON: **PERTURBATIVE** VS **FITTED** CHARM
MSHT NNPDF

- **MSHT VS NNPDF** FITTED CHARM **SAME IMPACT** ON GLUON
- **SAME** FITTED CHARM PDF, MSHT **SOMEWHAT LARGER UNC.**

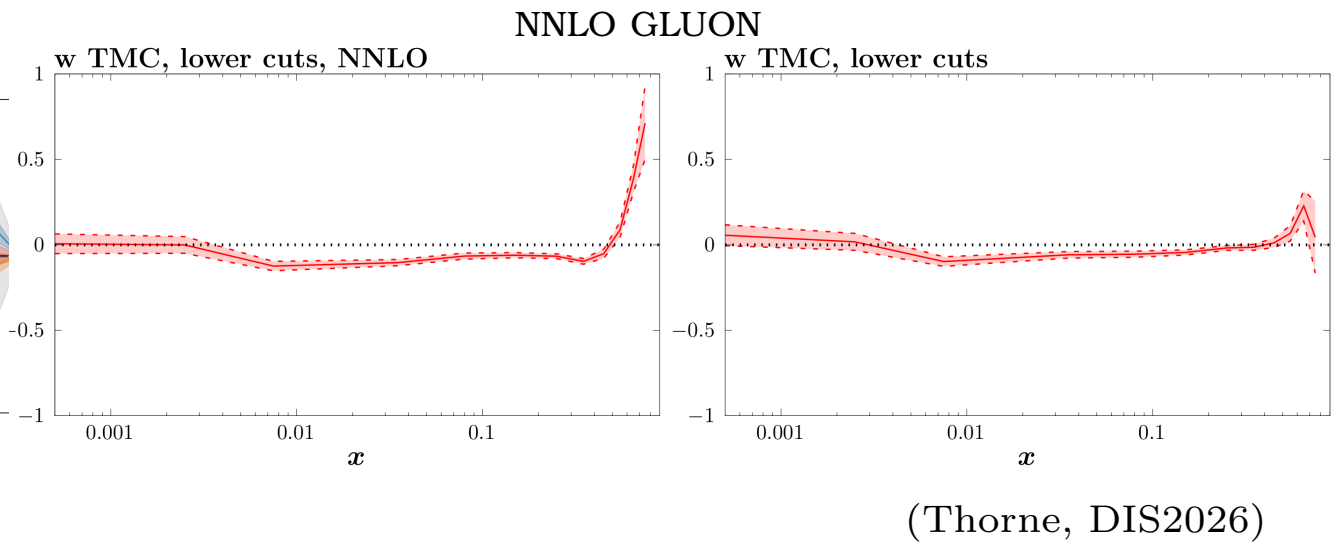
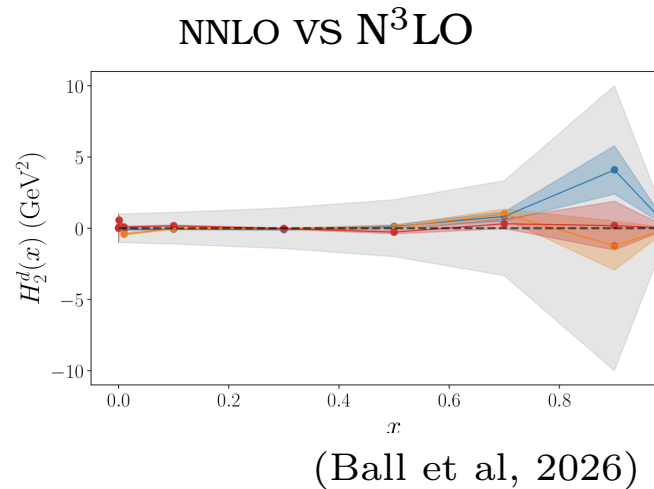


(Harland-Lang, Cridge, Thorne, 2025)

POWER CORRECTIONS

HIGHER TWISTS IN DIS

- EMPHASIZED IN PDF FITS INCLUDING LOW Q^2 DIS DATA (JAM, ABMP)
- TREATED WITH CUTS IN CURRENT GLOBAL FITS
- INCLUDED IN RECENT NNPDF, MSHT STUDIES: $\Rightarrow$ LARGELY REABSORBED AT N³LO
- INCLUDED TOGETHER WITH LINEAR POWER CORRNS WITH LOWER CUT

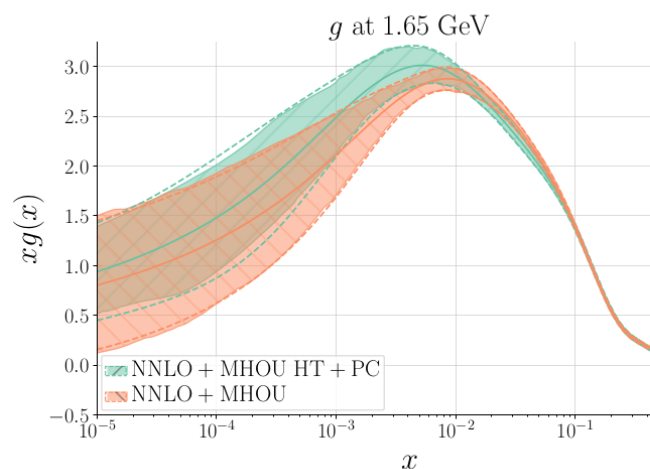


POWER CORRECTIONS

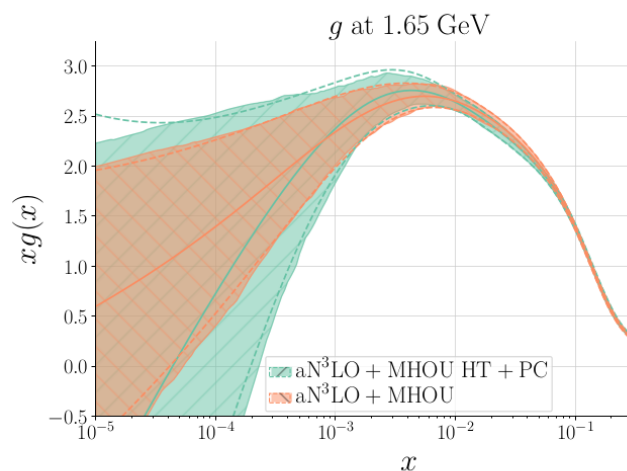
LINEAR PCs FOR JETS

- INCLUDED IN NNPDF STUDY
- MODERATE BUT **VISIBLE IMPACT** ON GLUON AT $10^{-3}x \lesssim x \lesssim 10^{-2}$

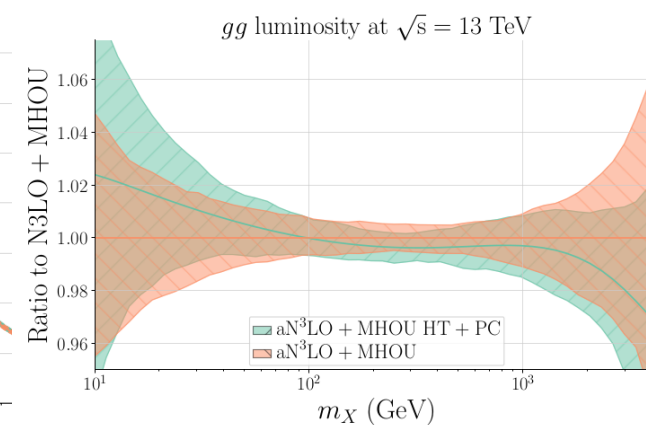
NNLO GLUON



N³LO GLUON



N³LO GG LUMI

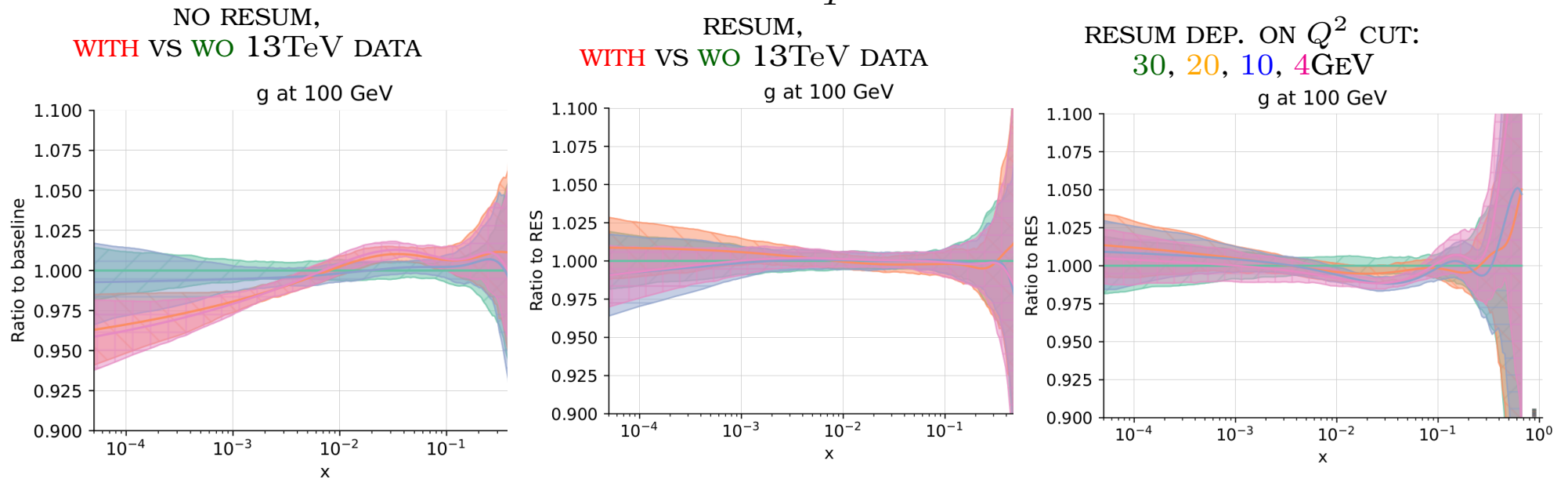


(Ball, Chiefa, Stegeman, 2026)

TRANSVERSE MOMENTUM RESUMMATION

- **LITTLE IMPACT** WITH 8TeV Z p_T DATA & $p_t > 30$ GeV
- **NEEDED** IN ORDER TO ACCOUNT FOR 13TeV **ATLAS DATA**
- GLUON **STABLE** UPON LOWERING p_t **CUT DOWN TO** $p_t > 4$ GeV

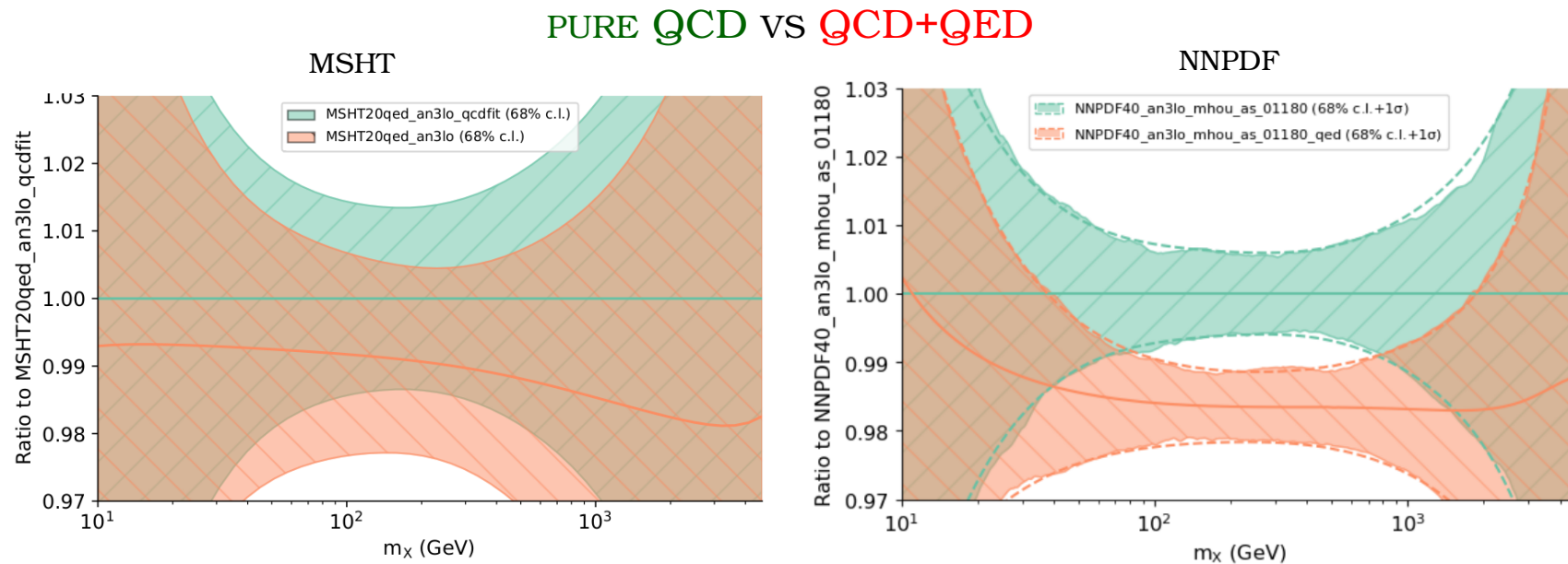
IMPACT OF Z P_T DATA



(Cruz-Martinez, Nocera, Rottoli, Torrielli, 2026)

QED & THE PHOTON PDF

- **COMPUTABLE** USING THE LUX FORMALISM (Manohar, Nason, Salam, Zanderighi 2017)
- **REDUCES GLUON MOMENTUM** FRACTION AT THE PERCENT LEVEL
- **PERCENT LEVEL IMPACT** ON GLUON LUMI
- RECENT LH STUDY $\Rightarrow$ MSHT-NNPDF **DIFFERENCE REDUCED** USING SMALLER (NNPDF3.1) DATASET



(Cridge, Cruz-Martinez, Huston, 2026)

THEORY UNCERTAINTIES

EXECUTIVE SUMMARY

WELL-UNDERSTOOD THEORY IMPACT:

- $N^3\text{LO}$, MHOU, FITTED CHARM $\Rightarrow$ **SIZABLE** IMPACT
- PHOTON PDF $\Rightarrow$ **VISIBLE IMPACT** ON GLUON
- LINEAR POWER CORRECTIONS, p_t RESUMMATION $\Rightarrow$ **MODERATE, LOCALIZED** IMPACT

NO REASON TO USE INFERIOR THEORY AT **CURRENT LEVEL OF PRECISION!**

SUMMARY

- ILL-DEFINED INVERSE PROBLEM $\Rightarrow$ DATA + ASSUMPTIONS
- UNCERTAINTIES $\Rightarrow$ SEPARATE
 - DATA UNCERTAINTIES $\Rightarrow$ STATISTICAL
 - MODEL UNCERTAINTIES $\Rightarrow$ BAYESIAN
- THIS TALK $\Rightarrow$ EXPLAINING DATA IMPACT
- FUTURE WORK $\Rightarrow$ EXPLAINING MODEL ASSUMPTIONS