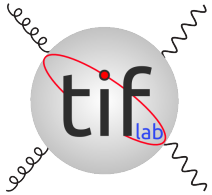
PARTON DISTRIBUTION UNCERTAINTIES AND THEIR CORRELATIONS

STEFANO FORTE
UNIVERSITÀ DI MILANO & INFN

UNIVERSITÀ DEGLI STUDI DI MILANO
DIPARTIMENTO DI FISICA



FEMTOCENTER SEMINAR

NOVEMBER 3, 2021

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 740006

SUMMARY

THE PROBLEM OF PDF UNCERTAINTIES

- UNCERTAINTY AND DISCOVERY
- TOLERANCE AND CONSISTENCY

UNCERTAINTIES AND DATA

- FUNCTIONAL UNCERTAINTIES
- CLOSURE TESTS

UNCERTAINTIES AND NO DATA

- EXTRAPOLATION
- FUTURE TESTS

PDF CORRELATIONS

- CORRELATION AND CROSS-CORRELATION
- PDF COMBINATION

DISCOVERY AT A HADRON COLLIDER AND PDF UNCERTAINTIES

THE DISCOVERY OF THE W (1984)

THEORETICAL PREDICTION

TABLE 2
Values (in nb) of the total cross sections for $W^{\pm}$ and Z^0 production

$\sqrt{s}$ (GeV)	$W^+ + W^-$ GHR	$W^+ + W^-$ DO1	$W^+ + W^-$ DO2	Z^0 GHR	Z^0 DO1	Z^0 DO2	$\frac{\sigma(W^+ + W^-)}{\sigma(Z^0)}$ GHR	$\frac{\sigma(W^+ + W^-)}{\sigma(Z^0)}$ DO1	$\frac{\sigma(W^+ + W^-)}{\sigma(Z^0)}$ DO2
540	4.2	4.3	4.1	1.3	1.3	1.2	3.1	3.4	3.5
700	6.2	6.3	6.1	2.0	1.9	1.8	3.1	3.3	3.4
1000	9.5	9.5	9.6	3.1	3.0	2.9	3.1	3.2	3.3
1300	12.5	12.5	12.9	4.0	3.9	3.9	3.1	3.2	3.3
1600	15.5	15.6	16.5	5.0	4.8	5.0	3.1	3.2	3.3

ALTARELLI, ELLIS, GRECO, MARTINELLI, 1984

EXPERIMENTAL DISCOVERY

EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH

CERN-EP/85-108
11 July 1985

W PRODUCTION PROPERTIES AT THE CERN SPS COLLIDER

UA1 Collaboration, CERN, Geneva, Switzerland

Aachen[1] – Amsterdam (NIKHEF)[2] – Annecy (LAPP)[3] – Birmingham[4] – CERN[5] – Harvard[6] – Helsinki[7] – Kiel[8] – London (Imperial College[9] and Queen Mary College[10]) – Padua[11] – Paris (Coll. de France)[12] – Riverside[13] – Rome[14] – Rutherford Appleton Lab.[15] – Saclay (CEN)[16] – Victoria[17] – Vienna[18] – Wisconsin[19] Collaboration

The corresponding experimental result for the 1984 data at $\sqrt{s} = 630$ GeV is

$$(\sigma \cdot B)_W = 0.63 \pm 0.05\,(\pm 0.09)\ \text{nb}.$$

This is in agreement with the theoretical expectation [14] of $0.47^{+0.14}_{-0.08}$ nb. We note that the 15%

- AGREEMENT AND UNCERTAINTIES AT 20% CONSIDERED TO BE SATISFACTORY
- RESULTS FROM DIFFERENT PDF SETS DIFFER BY AT LEAST 5%
- PDF UNCERTAINTIES??

DISCOVERY AT A HADRON COLLIDER AND PDF UNCERTAINTIES

THE DISCOVERY OF THE W (1984)

THEORETICAL PREDICTION

TABLE 2
Values (in nb) of the total cross sections for $W^{\pm}$ and Z^0 production

	$W^+ + W^-$	$W^+ + W^-$	$W^+ + W^-$	Z^0	Z^0	Z^0	$\frac{\sigma(W^+ + W^-)}{\sigma(Z^0)}$	$\frac{\sigma(W^+ + W^-)}{\sigma(Z^0)}$	$\frac{\sigma(W^+ + W^-)}{\sigma(Z^0)}$
$\sqrt{S}$ (GeV)	GHR	DO1	DO2	GHR	DO1	DO2	GHR	DO1	DO2
540	4.2	4.3	4.1	1.3	1.3	1.2	3.1	3.4	3.5
700	6.2	6.3	6.1	2.0	1.9	1.8	3.1	3.3	3.4
1000	9.5	9.5	9.6	3.1	3.0	2.9	3.1	3.2	3.3
1300	12.5	12.5	12.9	4.0	3.9	3.9	3.1	3.2	3.3
1600	15.5	15.6	16.5	5.0	4.8	5.0	3.1	3.2	3.3

ALTARELLI, ELLIS, GRECO, MARTINELLI, 1984

PDFS IN 1984

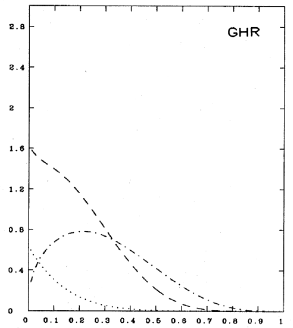


FIG. 25. Parton distributions of Glück, Hoffmann, and Reya (1982), at $Q^2=5$ GeV^2: valence quark distribution $x[u_v(x)+d_v(x)]$ (dotted-dashed line), $xG(x)$ (dashed line), and q_v (dotted line).

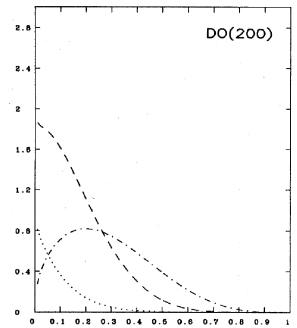


FIG. 27. "Soft-gluon" ($\Lambda=200$ MeV) parton distributions of Duke and Owens (1984) at $Q^2=5$ GeV^2: valence quark distribution $x[u_v(x)+d_v(x)]$ (dotted-dashed line), $xG(x)$ (dashed line), and $q_v(x)$ (dotted line).

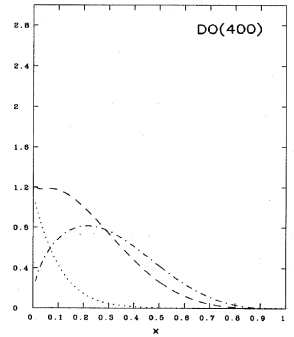


FIG. 26. "Hard-gluon" ($\Lambda=400$ MeV) parton distributions of Duke and Owens (1984) at $Q^2=5$ GeV^2: valence quark distribution $x[u_v(x)+d_v(x)]$ (dotted-dashed line), $xG(x)$ (dashed line), and $q_v(x)$ (dotted line).

Rev. Mod. Phys., Vol. 56, No. 4, October 1984

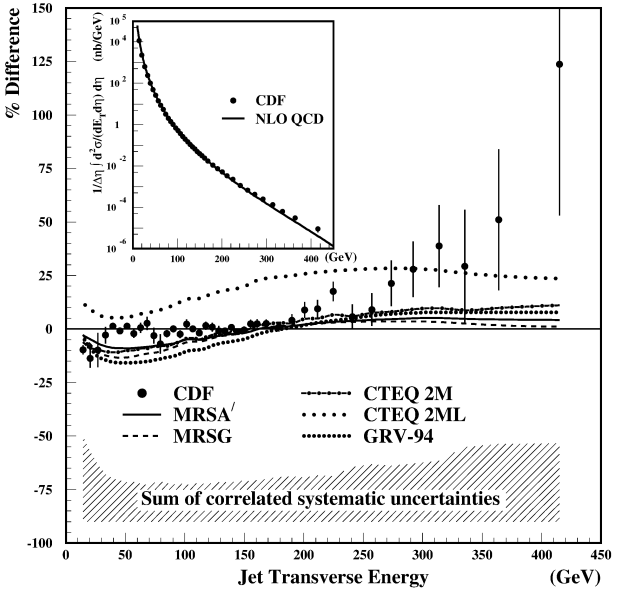
GHR VS DUKE-OWENS

- AGREEMENT AND UNCERTAINTIES AT 20% CONSIDERED TO BE SATISFACTORY
- RESULTS FROM DIFFERENT PDF SETS DIFFER BY AT LEAST 5%
- PDF UNCERTAINTIES $\Rightarrow$ SPREAD OF MODELS???
- NO BIG DEAL FOR DISCOVERY

DISCOVERY AT A HADRON COLLIDER AND PDF UNCERTAINTIES

THE DISCOVERY OF QUARK COMPOSITENESS (1995)

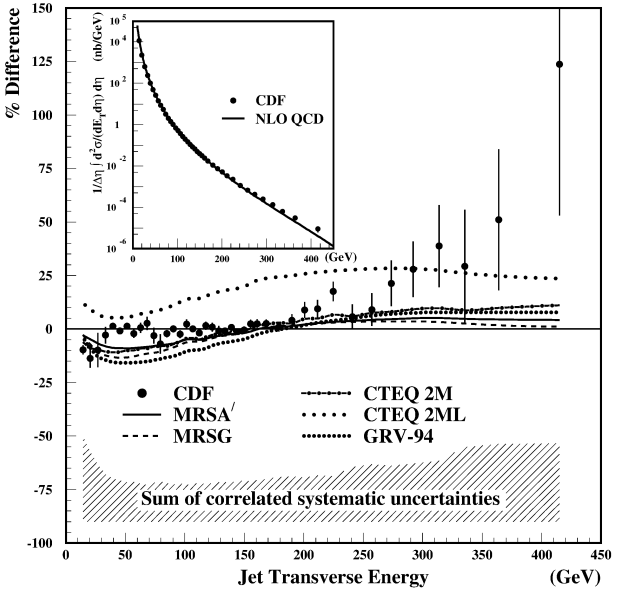
- DISCREPANCY BETWEEN QCD CALCULATION AND CDF JET DATA (1995)
- EVIDENCE FOR QUARK COMPOSITENESS
- .



DISCOVERY AT A HADRON COLLIDER AND PDF UNCERTAINTIES

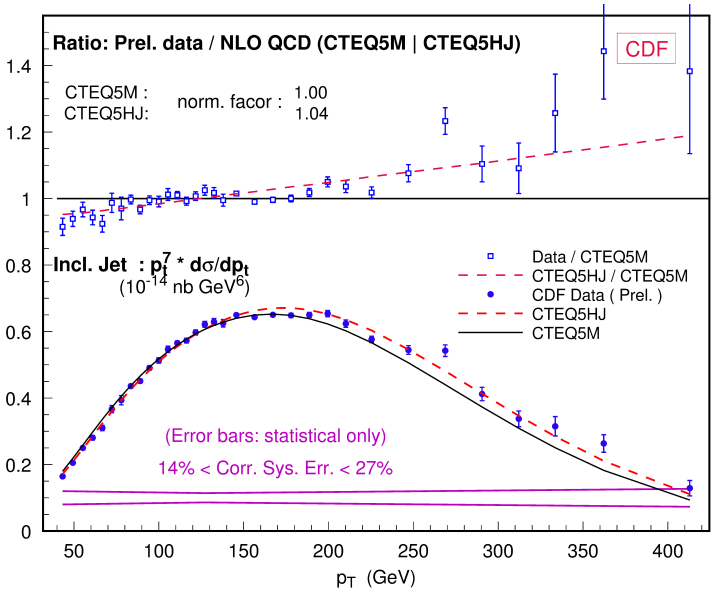
A BETTER DETERMINATION OF THE GLUON PDF (1995)

- DISCREPANCY BETWEEN QCD CALCULATION AND CDF JET DATA (1995)
- ~~EVIDENCE FOR QUARK COMPOSITENESS~~
- NO INFO ON PARTON UNCERTAINTY $\Rightarrow$ RESULT STRONGLY DEPENDS ON GLUON AT $x \gtrsim 0.1$

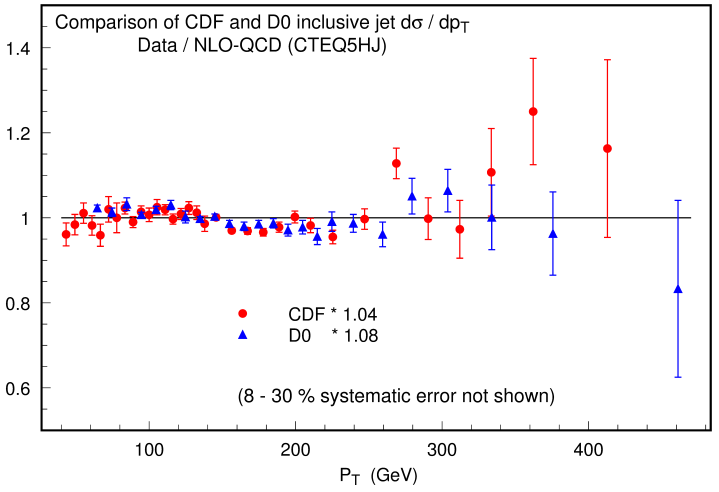


DISCREPANCY REMOVED IF JET DATA INCLUDED IN THE FIT

NEW CTEQ FIT (1996)



FINAL CTEQ FIT (1998)



WHAT'S THE PROBLEM $\sim$ 2000

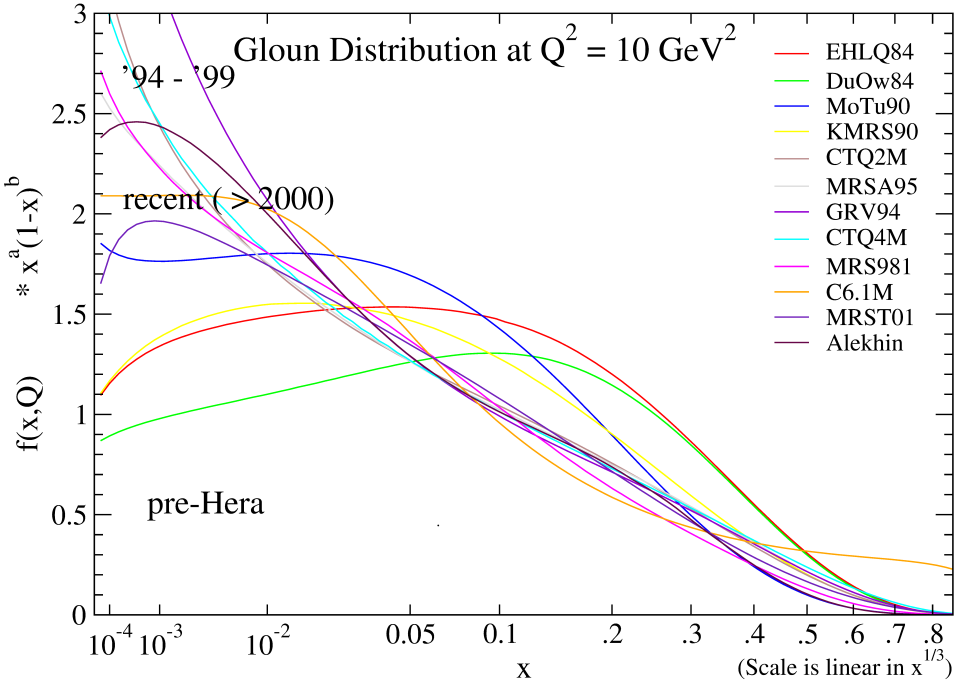
PDFs DETERMINED FITTING A MODEL-INSPIRED FUNCTIONAL FORM

gluon parametrization (MRST 2004)

$$xg(x,Q_0^2) = A_g(1-x)^{\eta_g}(1+\epsilon_g x^{0.5}+\gamma_g x)x^{\delta_g} - A_-(1-x)^{\eta_-}x^{-\delta_-}$$

- PROBLEM REDUCED TO FINITE-DIMENSIONAL
- WHO PICKS THE FUNCTIONAL FORM?

HISTORICAL COMPILATION OF GLUON PDFs

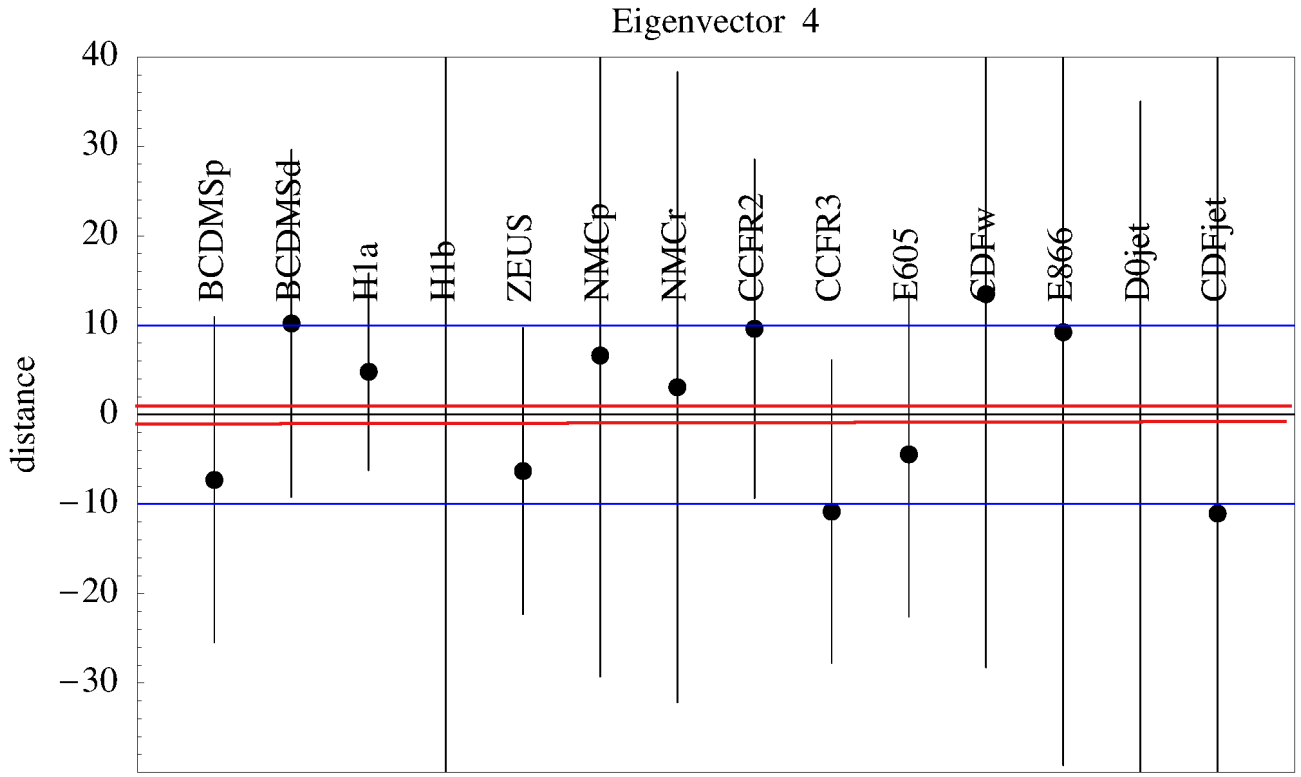


W.K.Tung, DIS 2004

FIRST PDFs WITH UNCERTAINTIES (2002)

"TOLERANCE"

one sigma & ten sigma intervals for typical
covariance matrix eigenvalue
vs best value and uncertainty from individual experiments

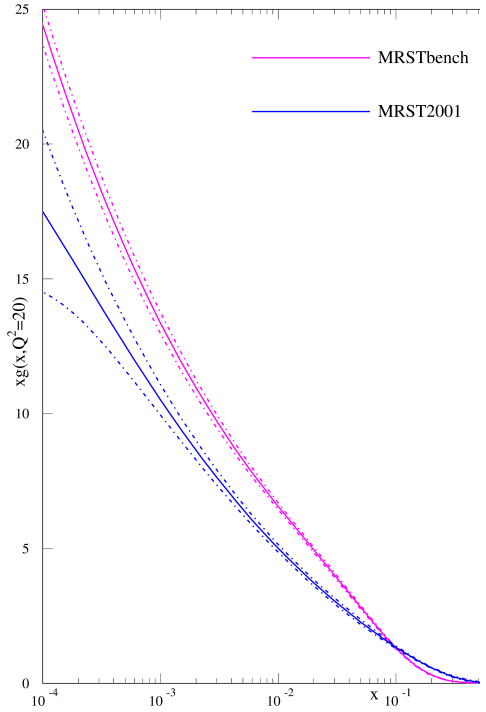


- SPREAD OF BEST-FIT FROM DIFFERENT DATA HUGE W.R. TO TEXTBOOK UNCERTAINTIES
- PDF UNCERTAINTIES RESCALED BY "TOLERANCE" $T \sim 10$

CAN WE TRUST PDF UNCERTAINTIES?
THE HERA-LHC BENCHMARK (2005)

- RESTRICTED AND VERY CONSISTENT DATASET USED
- RESULTS COMPARED TO THEN-BEST RESULT FROM FULL DATASET

BENCHMARK VS DEFAULT GLUON



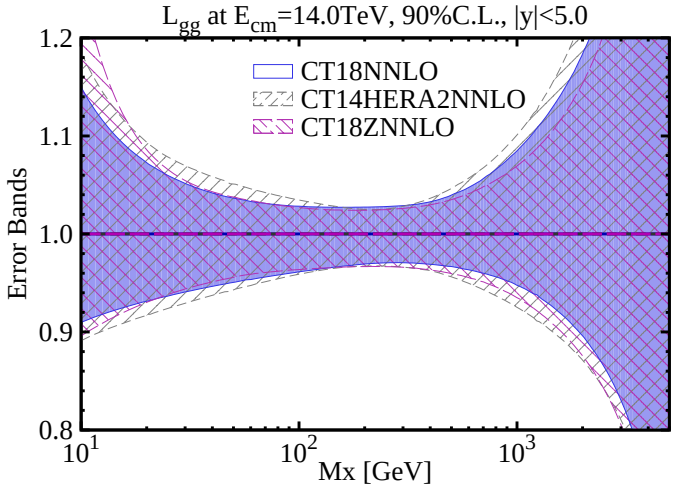
"...the partons extracted using a very limited data set are completely incompatible, even allowing for the uncertainties, with those obtained from a global fit with an identical treatment of errors...The comparison illustrates the problems in determining the true uncertainty on parton distributions." (R.Thorne, HERALHC, 2005)

CAN WE TRUST PDF UNCERTAINTIES?

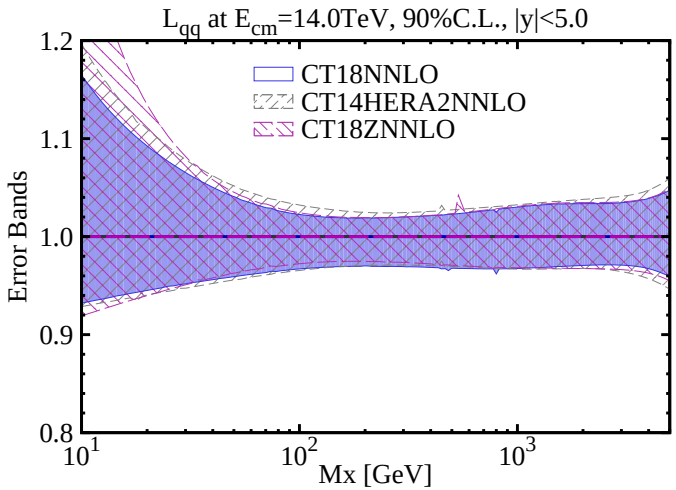
UNCERTAINTY REDUCTION?

CT18 VS. CT14: PARTON LUMINOSITY UNCERTAINTIES

GLUON-GLUON



QUARK-QUARK



MORE DATA $\Rightarrow$ BIGGER UNCERTAINTIES (?!)

PARTON PARAMETRIZATIONS

- CTEQ5 2002: $xg(x,Q_0^2) = A_0 x^{A_1}(1-x)^{A_2}(1+A_3 x^{A_4})$
- MRST-HERALHC 2005: $xg(x,Q_0^2) = A_g x^{\delta_g}(1-x)^{\eta_g}(1+\epsilon_g x^{0.5}+\gamma_g x) + A_{g'} x^{\delta_{g'}}(1-x)^{\eta_{g'}}$
- CT18: $g(x,Q=Q_0) = x^{a_1-1}(1-x)^{a_2}\left[a_3(1-y)^3 + a_4 3y(1-y)^2 + a_5 3y^2(1-y) + y^3\right]$; $y=\sqrt{x}$; $a_5 = (3+2a_1)/3$.

BIAS?

UNCERTAINTIES AND DATA

DATA $\Leftrightarrow$ PDFs

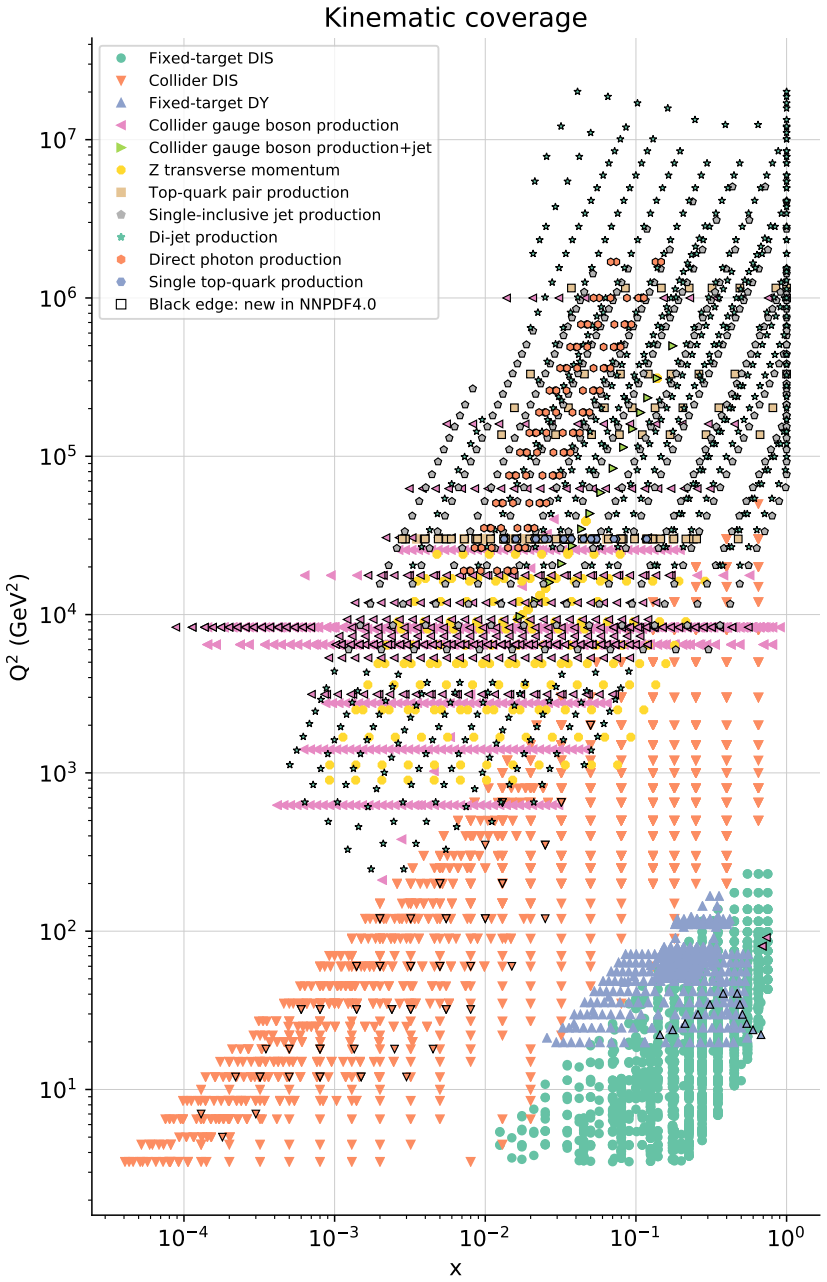
THE TASK

- LHC CROSS SECTION:
  - $\sigma = \sum_{ij} \hat{\sigma}_{ij} \otimes f_i^{(1)} \otimes f_j^{(2)}$
  - $\hat{\sigma}_{ij}$ PARTONIC CROSS SECTION WITH INCOMING PARTONS i, j
  - $f_i^{(j)}(x, Q^2)$ PDF FOR PARTON OF SPECIES i IN j-TH INCOMING PROTON
  - $\otimes$ CONVOLUTION OVER x
  - PDFs AT SCALE Q^2 DETERMINED FROM PDFs AT REFERENCE SCALE $f_i(x, Q_0^2)$ BY SOLVING EVOLUTION EQUATIONS
- PARTONIC CROSS SECTION COMPUTED PERTURBATIVELY
- PDFs DETERMINED COMPARING σ TO DATA

THE PROBLEM

- DETERMINE A MULTIVARIATE PROBABILITY DISTRIBUTION
- IN AN INFINTE DIMENSIONAL SPACE OF FUNCTIONS (DISTRIBUTIONS, REALLY) $\Rightarrow$ PROBABILITY OF "PROBABILITIES"
- USING INDIRECT & DISCRETE (THOUGH DENSE) DATA

the nnpdf4.0 dataset



PROBABILITIES IN A SPACE OF FUNCTIONS?

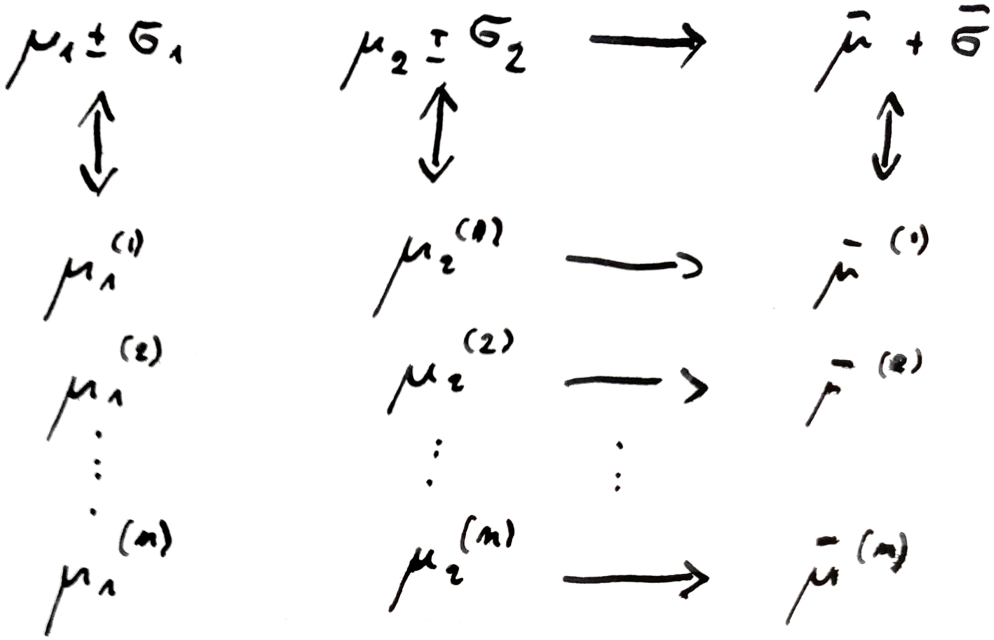
THE MONTE CARLO REPRESENTATION

A SIMPLE EXAMPLE

TWO MEASUREMENTS: $\mu_1 \pm \sigma_1; \qquad \mu_2 \pm \sigma_2$

MC COMBINATION: $\bar{\mu} \pm \bar{\sigma}; \quad \bar{\mu} = \dfrac{\frac{\mu_1}{\sigma_1^2} + \frac{\mu_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}; \; \bar{\sigma}^2 = \dfrac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}$

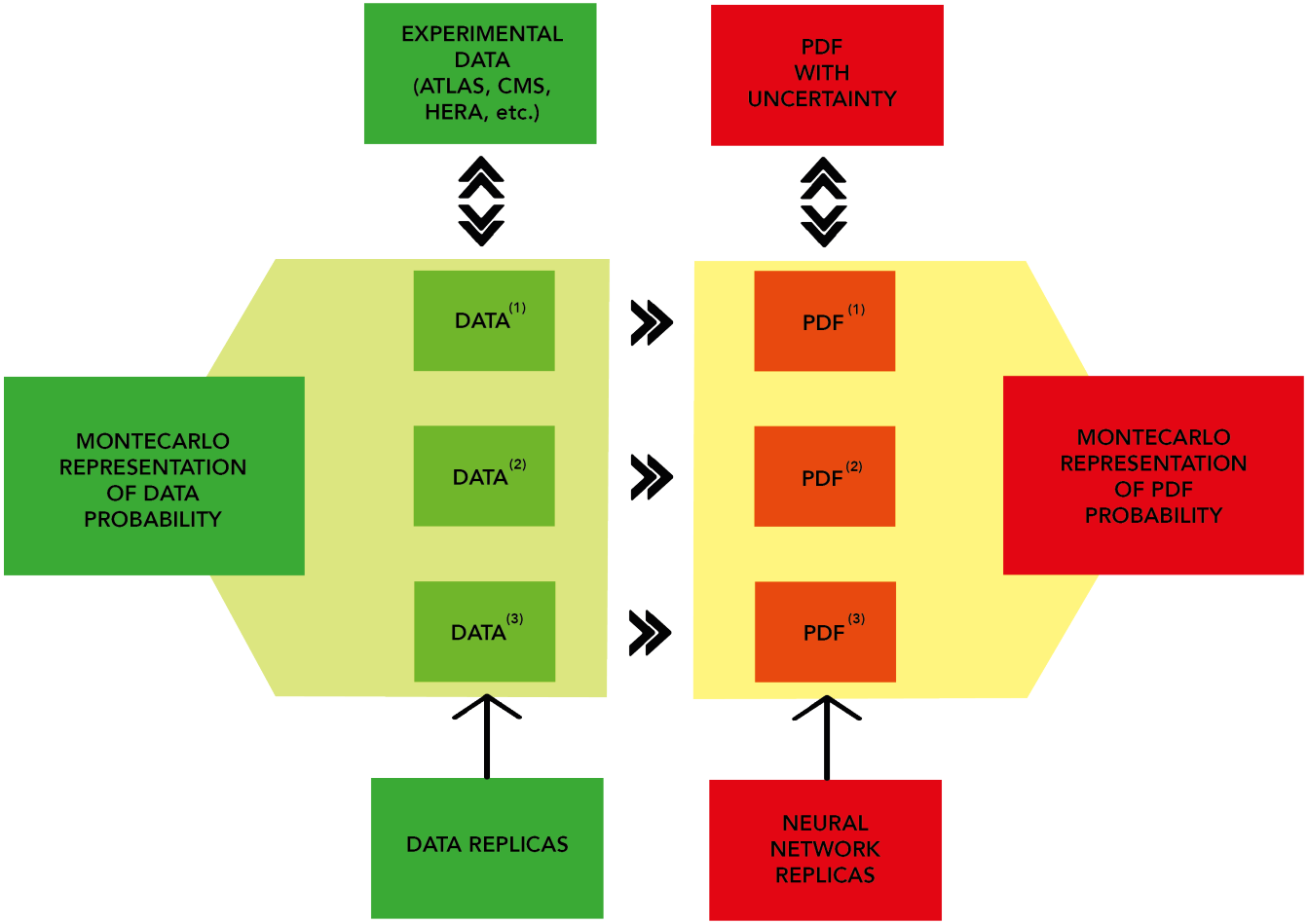
MONTE CARLO REPRESENTATION



$\mu^{(i)} \Leftrightarrow$ REPLICA SAMPLE $\Leftrightarrow$ REPRESENTATION OF PROBABILITY DISTRIBUTION

THE MONTE CARLO REPRESENTATION

REPLICA SAMPLE OF FUNCTIONS $\Leftrightarrow$ PROBABILITY DENSITY IN FUNCTION SPACE



FINAL PDF SET: $f_i^{(a)}(x,\mu)$; i =up, antiup, down, antidown, strange, antistrange, charm, gluon; $j = 1, 2, \ldots N_{\rm rep}$

UNDERSTANDING PDF UNCERTAINTIES

THE CLOSURE TEST

- ASSUME UNDERLYING "TRUTH" PDF (SAY A RANDOM NNPDF4.0 REPLICA)
- GENERATE DATA ACCORDING TO STATISTICAL AND CORRELATED SYSTEMATICS (SAY FOR NNPDF4.0 DATASET)
- PERFORM A FIT & COMPARED TO "TRUTH"

UNCERTAINTY LEVELS

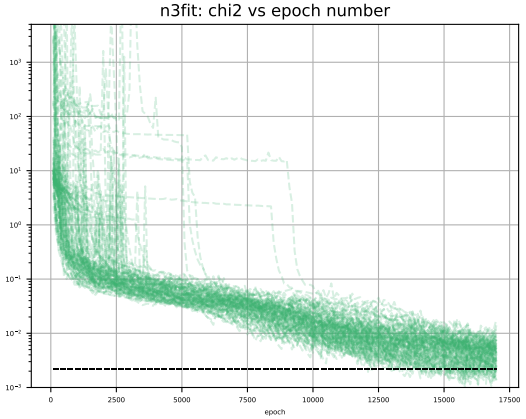
- LEVEL 0:
  - EACH DATAPOINT EQUAL TO THE "TRUTH VALUE"; ZERO UNCERTAINTY
  - FIT $\rightarrow$ MUST FIND $\chi^2 = 0$ (GET BACK "TRUTH")
  - $\chi^2 \approx 0$ BOTH REPLICA TO REPLICA AND AVERAGE TO TRUTH
  - INTERPOLATION/EXTRAPOLATION UNCERTAINTY
- LEVEL 1:
  - EACH PSEUDO- DATAPOINT IS OBTAINED AS A RANDOM FLUCTUATION WITH GIVEN COVARIANCE MATRIX ABOUT "TRUTH"
    $\Rightarrow$ "RUN OF THE UNIVERSE"
  - FIT DATA OVER AND OVER AGAIN
  - $\chi^2 \approx 1$ BOTH REPLICA TO REPLICA AND AVERAGE TO TRUTH
  - FUNCTIONAL UNCERTAINTY
- LEVEL 2:
  - DATA AS IN LEVEL 1
  - GENERATE DATA REPLICAS OF THESE "DATA"
  - FIT PDF REPLICAS TO DATA REPLICAS
  - $\chi^2 \approx 2$ REPLICA TO REPLICA; $\chi^2 \approx 1$ AVERAGE TO TRUTH
  - DATA UNCERTAINTY

UNCERTAINTIES: TYPE AND SIZE

CLOSURE TEST RESULTS (NNPDF4.0)

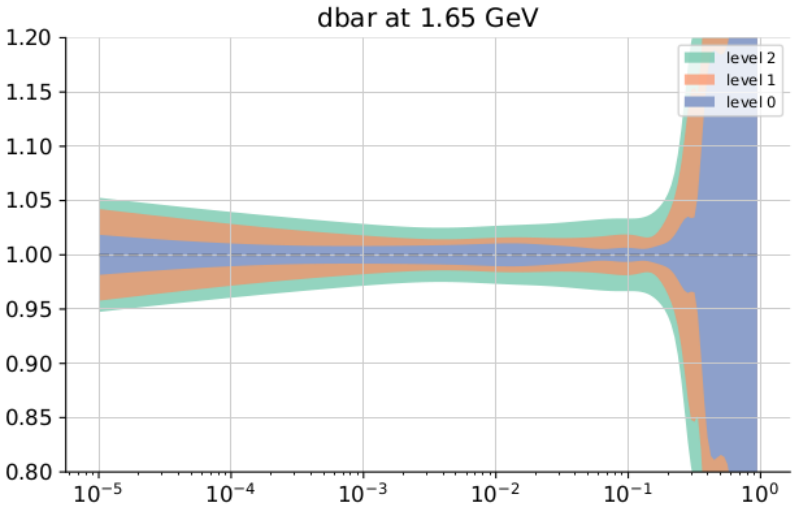
- LEVEL 0 $\chi^2 \approx 0$, YET UNCERTAINTY NONZERO
  $\Rightarrow$ NEURAL NETS $\Leftrightarrow$ MANY FUNCTIONAL FORMS
- LEVEL 1 REPLICAS ALL FITTED TO SAME DATA, YET UNCERTAINTY NONZERO
  $\Rightarrow$ DITTO
- LEVEL 0, 1 AND 2 UNCERTAINTIES COMPARABLE IN SIZE

LEVEL 0 χ^2 VS TRAINING

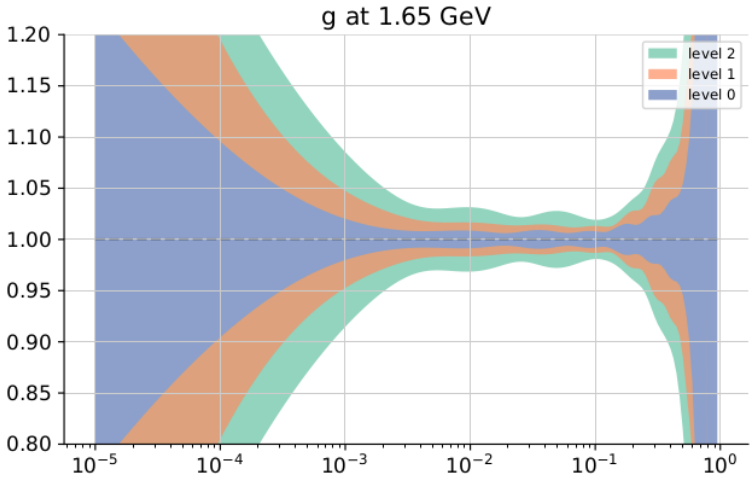


LEVEL 0/1/2 UNCERTAINTIES

ANTIDOWN

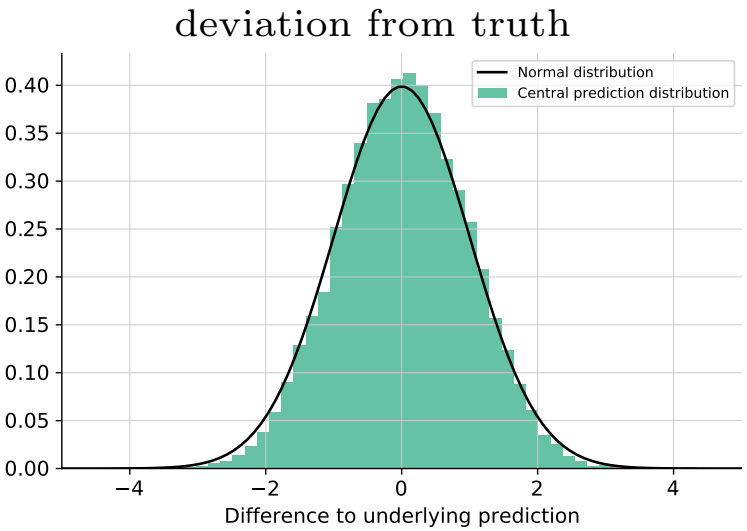


GLUON



ASSESSING UNCERTAINTIES
CLOSURE TEST RESULTS (NNPDF4.0)

- FIT TO DATA, COMPARE TO PREDICTIONS FOR DIFFERENT PROCESSES, SAME KINEMATIC REGION)
  fit to NNPDF3.1 dataset, compare to NNPDF4.0 data not in NNPDF3.1
- COMPARE PREDICTION TO TRUTH
- REPEAT FOR A NEW "RUN OF THE UNIVERSE"& ITERATE
- INDICATORS:
  - NORMALIZED HISTOGRAM OF DEVIATIONS FROM TRUTH
  - RMS DEVIATION (BIAS) / NOMINAL UNCERTAINTY (VARIANCE)
  - FRACTION OF TRUTH WITHIN ONE σ (SHOULD BE 68%)



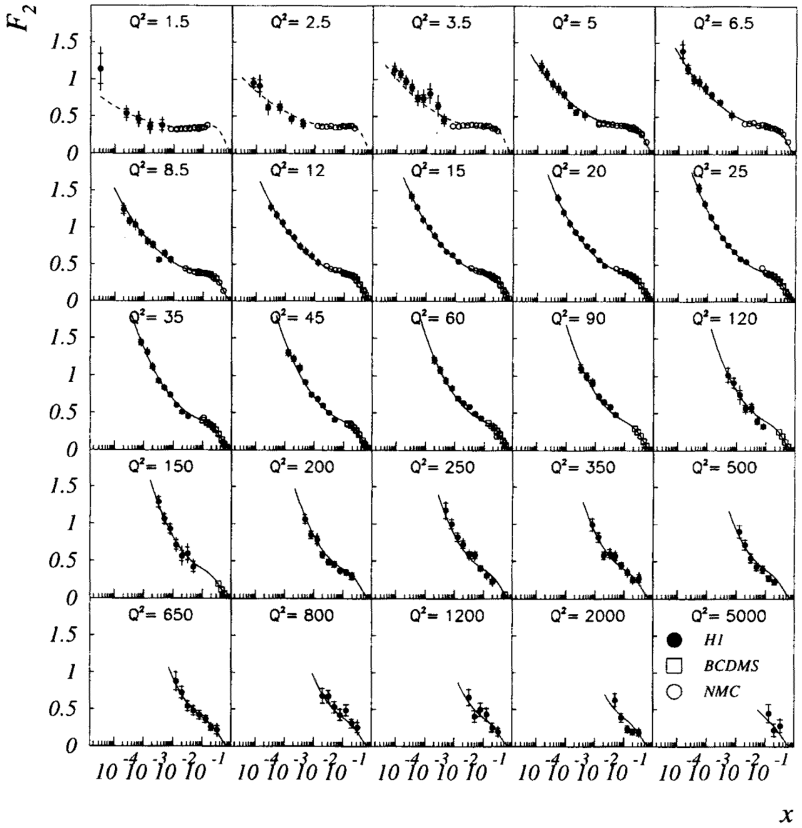
Dataset	$\sqrt{\text{bias/variance}}$	$\xi_{1\sigma}^{(\text{data})}$
DY	0.99 ± 0.08	0.69 ± 0.02
Top-pair	0.75 ± 0.06	0.75 ± 0.03
Jets	1.14 ± 0.05	0.63 ± 0.03
Dijets	0.99 ± 0.07	0.70 ± 0.03
Direct photon	0.71 ± 0.06	0.81 ± 0.03
Single top	0.87 ± 0.07	0.69 ± 0.04
Total	1.03 ± 0.05	0.68 ± 0.02

PDF EXTRAPOLATION

AN EXAMPLE: SMALL-x EXTRAPOLATION

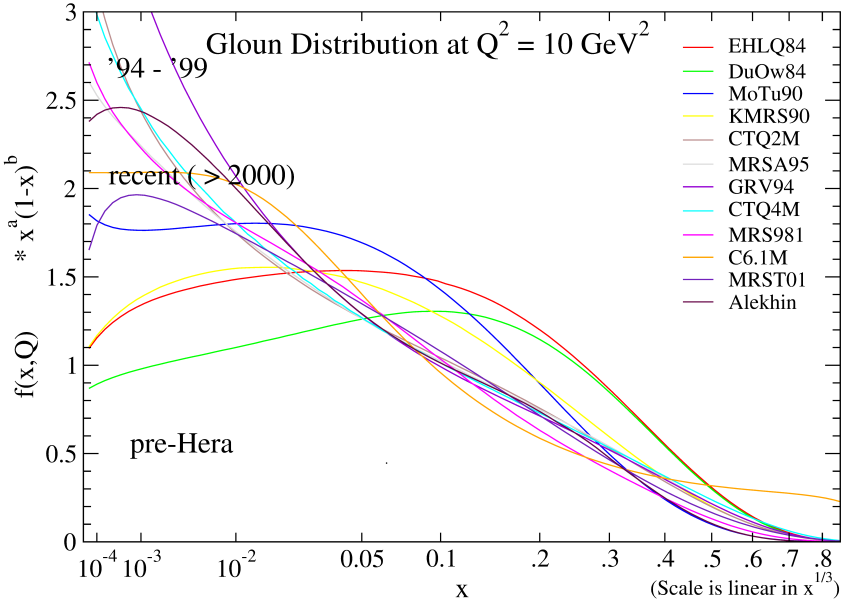
1995: THE RISE OF STRUCTURE FUNCTIONS AT HERA

FIRST HERA DATA VS OLDER DATA



A. de Roeck, Cracow epiphany conf. 1996

HISTORICAL COMPILATION OF GLUON PDFs



W.K.Tung, DIS 2004

- RISE OF F_2 AT HERA CAME $\Rightarrow$ SURPRIZE
- HINTED BY PRE-HERA DATA; VETOED BY THEORETICAL BIAS

EXTRAPOLATION

ARE UNCERTAINTIES OUTSIDE DATA REGION "REASONABLE"?

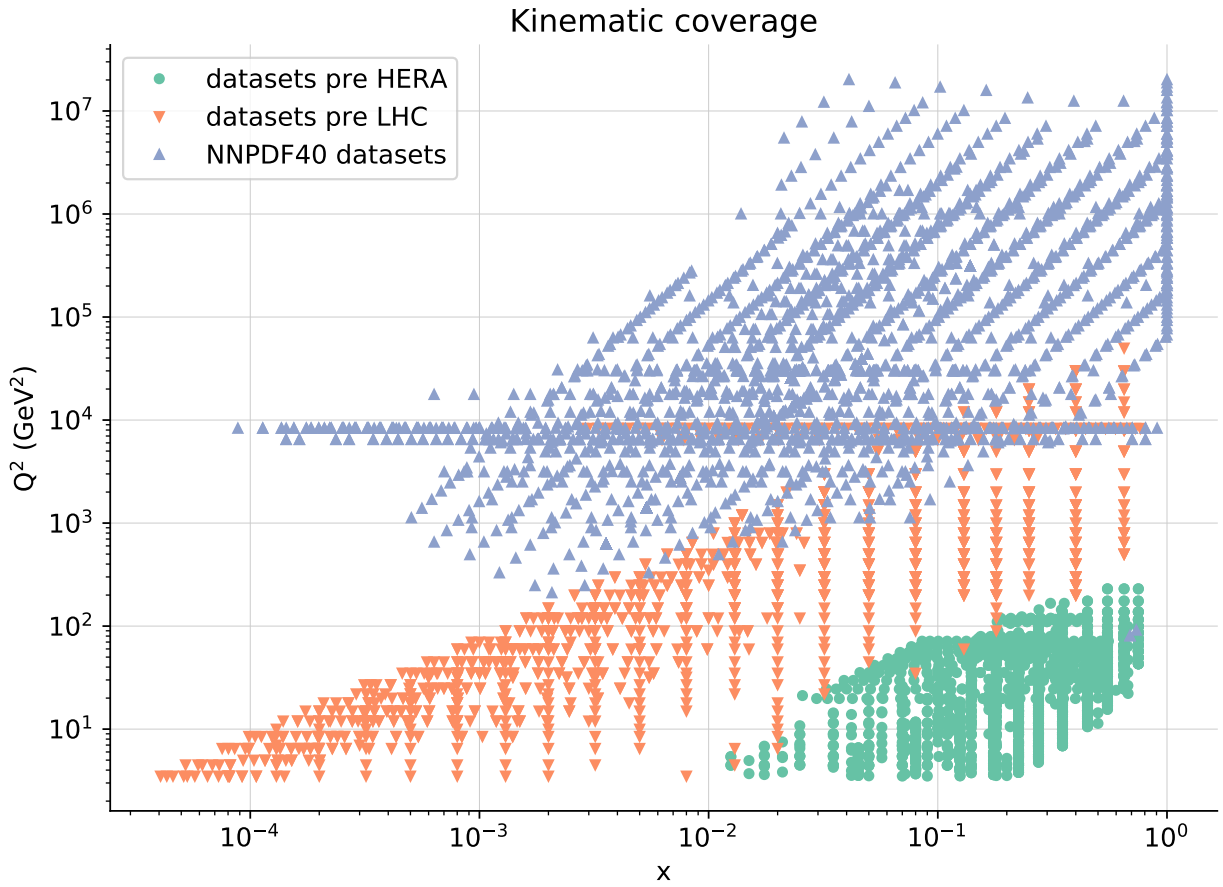
INFINITE UNCERTAINTIES IS NOT GOOD ENOUGH!

WHAT'S THE PROBLEM?

- DATA MAY INCORPORATE CONSTRAINTS THAT WE ARE UNAWARE OF
- DATA IN EXTRAPOLATION COULD VIOLATE "NATURAL" ASSUMPTIONS
- WHAT IS THE SPACE OF ACCEPTABLE SOLUTIONS?

FUTURE TESTS

IDEA: USE (REAL) HIERARCHICAL DATASETS



- DEFINE "PRE-HERA", " PRE-LHC" AND "CURRENT" DATASETS
  EACH LATER DATASET IS EXTRAPOLATION OF PREVIOUS
- DETERMINE PDFS & COMPARE TO "FUTURE" DATA
- COMPUTE χ^2 TO FUTURE DATA:
  - WITHOUT PDF UNCERTAINTIES $\Rightarrow$ IF $\gg 1$, MISSING INFORMATION
  - WITH PDF UNCERTAINTY $\Rightarrow$ IF ~ 1, TEST PASSED

MISSING INFO REPRODUCED BY UNCERTAINTY

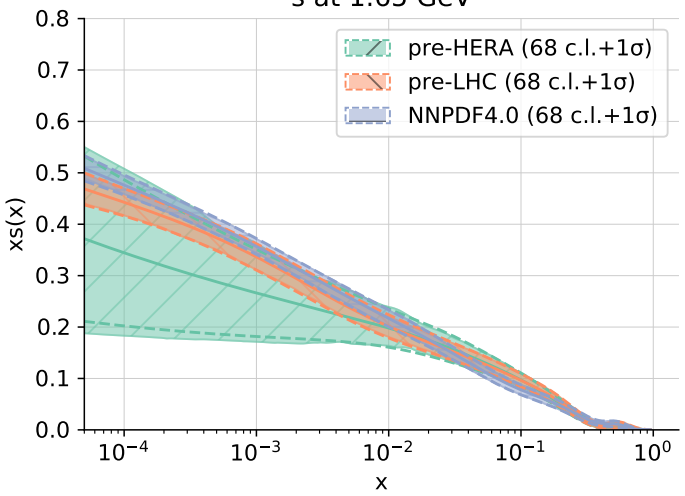
ASSESSING EXTRAPOLATION UNCERTAINTIES

FUTURE TEST RESULTS (NNPDF4.0)

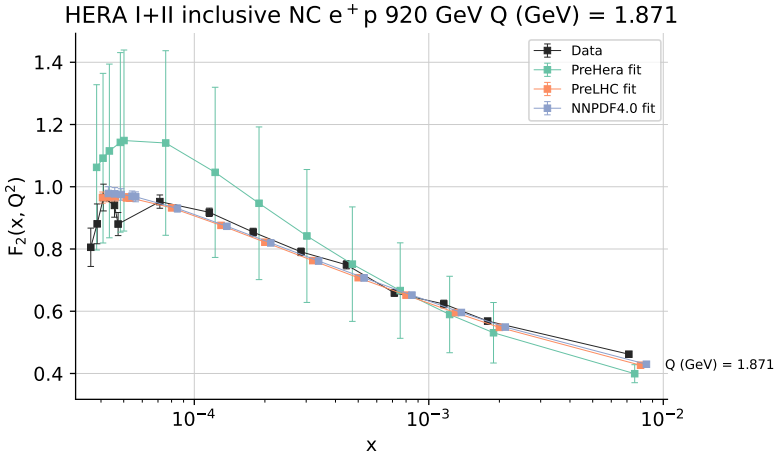
χ^2: FITTED VS EXTRAPOLATED: WITHOUT/WITH PDF UNC.

PROCESS	PRE-HERA	PRE-LHC	NNPDF4.0
FT DIS (NC)	1.05	1.18	1.23
FT DIS (CC)	0.80	0.85	0.87
FT DY	0.92	1.27	1.59
HERA	27.20/1.23	1.22	1.20
COLL. DY (TEV.)	5.52/1.02	0.99	1.11
COLL. DY (LHC)	18.91/1.31	2.63/1.58	1.53
TOP QUARK	20.01/1.06	1.30/0.87	1.01
JETS	2.69/0.98	2.12/1.10	1.26
TOTAL OUT OF SAMPLE	19.48/1.16	2.10/1.15	–

strange PDF



HERA F_2^p



PDFS ARE FUTURE-COMPATIBLE!

CORRELATIONS & COMBINATION

PDF CORRELATIONS

DEFINITION AND USE

example: up vs down PDFs

COVARIANCE: $\mathrm{Cov}[u,d](x,x') = \langle u(x,Q_0^2)d(x',Q_0^2)\rangle - \langle u(x,Q_0^2)\rangle\langle d(x',Q_0^2)\rangle$;

CORRELATION: $\rho[u,d](x,x') = \frac{\mathrm{Cov}[u,d](x,x')}{\sqrt{\mathrm{Var}[u](x)\mathrm{Var}[d](x')}}$

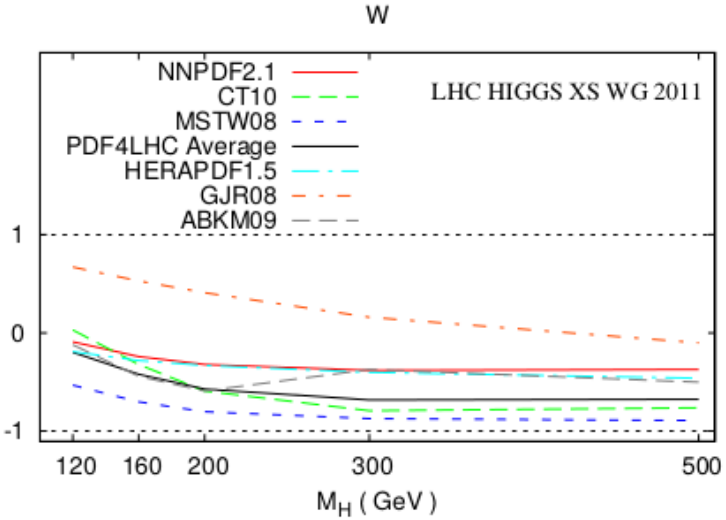
COMPUTATION IN MC APPROACH: $\langle u(x,Q_0^2)d(x',Q_0^2)\rangle = \frac{1}{N}\sum_{r=1}^{N} u^{(r)}(x,Q_0^2)d^{(r)}(x',Q_0^2)$;
$u^{(r)}(x,Q_0^2)$ REPLICAS

- CORRELATION INDUCED BY DATA, THEORY (E.G. SUM RULES), METHDOLOGY (E.G. ASSUMPTIONS ON EXTRAPOLATION)

- USED E.G. TO ASSESS CORRELATION BETWEEN SIGNAL AND BACKGROUND PROCESSES

PDF-INDUCED CORRELATIONS
BETWEEN HIGGS SIGNAL & BACKGROUND PROCESSES (HXSWG, YR2, 2011)

Higgs in gluon fusion vs. W production



UNDERSTANDING PDF CORRELATIONS

THE CROSS-CORRELATION

CORRELATE PDFS IN DIFFERENT SETS

example: up NNPDF4.0 vs down CT18

$$\mathrm{Cov}[u^N, d^C](x, x') = \langle u^{\mathrm{NNPDF}}(x, Q_0^2) d^{\mathrm{CT}}(x', Q_0^2)\rangle - \langle u^{\mathrm{NNPDF}}(x, Q_0^2)\rangle \langle d^{\mathrm{CT}}(x', Q_0^2)\rangle$$

S-CORRELATION VS F-CORRELATION

$\rho[u^N, u^C]$ DIFFERENT SETS, SAME PDF VS. $\rho[u^N, d^N]$ SAME SET, DIFFERENT PDFS

- SAME REPLICA MUST BE USED FOR NONZERO CORRELATION:
  IF REPLICAS UNCORRELATED $\langle u(x, Q_0^2) d(x, Q_0^2)\rangle \stackrel{?}{=} \frac{1}{N}\sum_{r=1}^{N} u^{(r)}(x, Q_0^2) d^{(r')}(x, Q_0^2) = \langle u\rangle\langle d\rangle$
  THEN CORRELATION VANISHES

- WHAT IS "THE SAME REPLICA" ACROSS PDF SETS?

- MAYBE A REPLICA FITTED TO THE SAME DATA REPLICA?

SUGGESTION

- FIT PDF REPLICAS $f_i^{(r,\,\mathrm{NNPDF})}(x, Q_0^2)$ & $f_i^{(r,\,\mathrm{CT})}(x, Q_0^2)$ for all x, i TO SAME DATA REPLICA

- COMPUTE COVARIANCE & CORRELATION USING

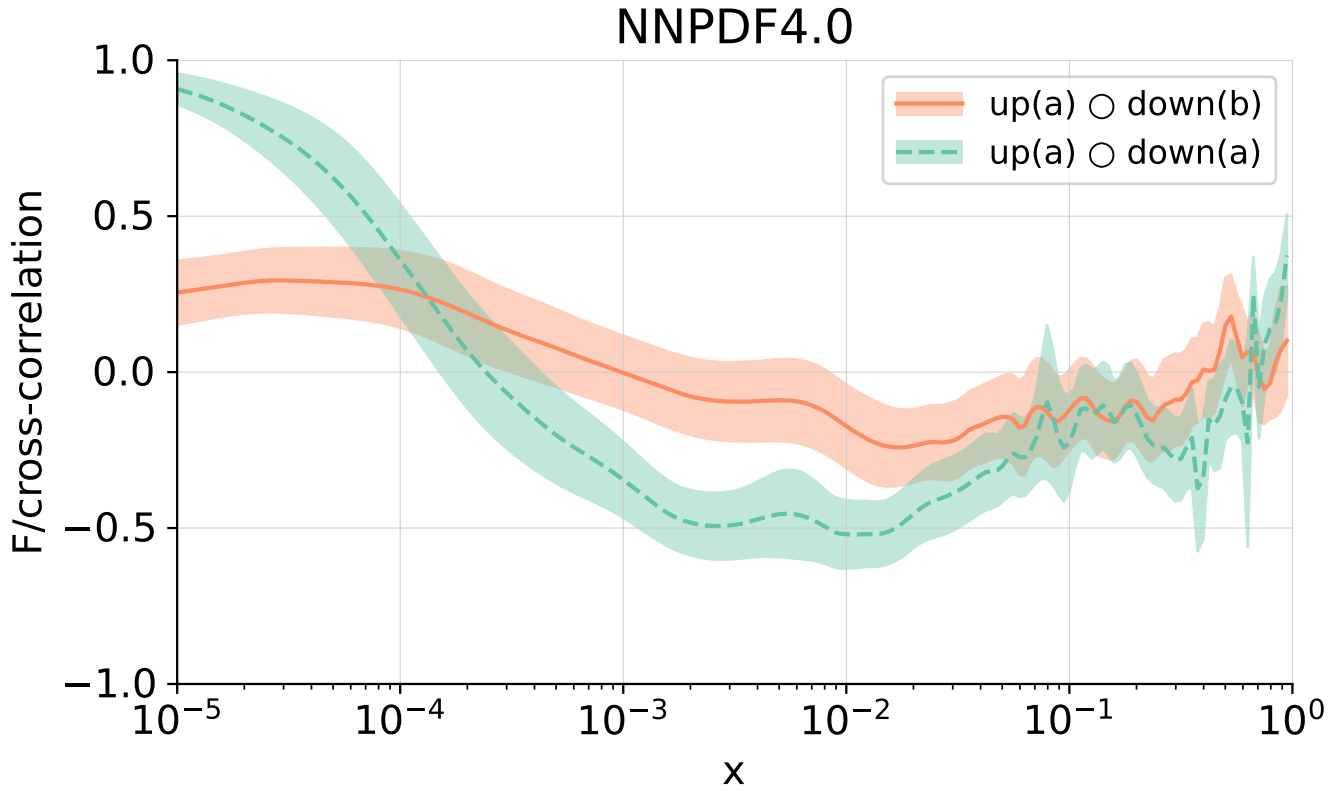
$$\langle u(x, Q_0^2) d(x, Q_0^2)\rangle \stackrel{?}{=} \frac{1}{N}\sum_{r=1}^{N} u^{(r,\,\mathrm{NNPDF})}(x, Q_0^2) d^{(r,\,\mathrm{CT})}(x, Q_0^2)$$

THE DATA-INDUCED CORRELATION PROBLEM

- THE DATA REPLICA DOES NOT DETERMINE UNIQUELY THE PDF REPLICA
  RECALL THE LEVEL-1 UNCERTAINTY
- IF $r \Leftrightarrow$ DATA REPLICA, THEN
  REPLICAS (UP QUARK) $u^{(r,r')}(x,Q_0^2)$; $r' \Leftrightarrow$ LEVEL-1 (METHDOLOLOGY) REPLICAS

$$\left|\frac{1}{N}\sum_{r=1}^{N} u^{(r,r')}(x,Q_0^2)d^{(r,r'')}(x,Q_0^2) - \langle u\rangle\langle d\rangle\right| \leq \left|\frac{1}{NM}\sum_{r=1}^{N}\sum_{r'=1}^{M} u^{(r,r')}(x,Q_0^2)d^{(r,r')}(x,Q_0^2) - \langle u\rangle\langle d\rangle\right|$$

- ONLY DATA-INDUCED CORRELATION INCLUDED!



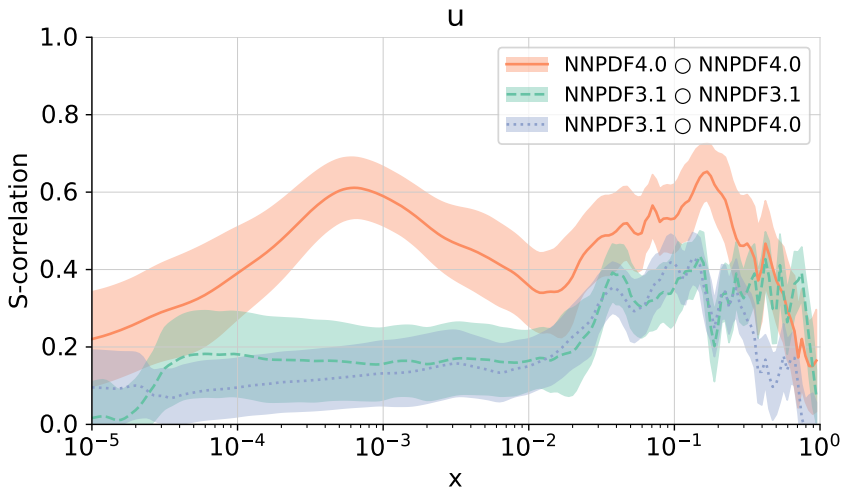
THE DATA-INDUCED SELF-CORRELATION

- COMPUTE THE S-CORRELATION OF A PDF SET TO ITSELF
  = THE F-CORRELATION OF A PDF TO ITSELF
- USE TWO DIFFERENT SETS OF PDF REPLICAS FITTED TO THE SAME DATA REPLICAS

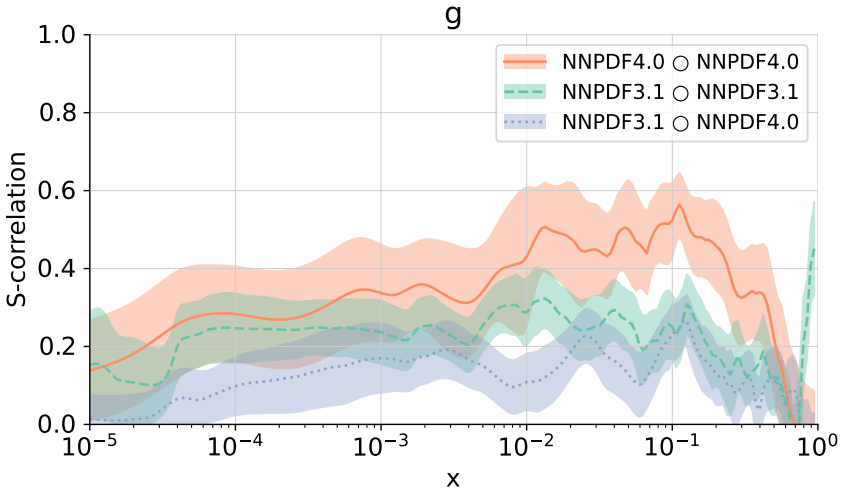
$$\langle u(x,Q_0^2)u(x,Q_0^2)\rangle = \frac{1}{N}\sum_{r=1}^{N} u^{(r,r')}(x,Q_0^2)u^{(r,r'')}(x,Q_0^2)$$

- DEVIATION OF CORRELATION FROM 100% MEASURES THE CORRELATION LOSS DUE TO UNCORRELATED METHODOLOGY
-
-
-

up quark



gluon



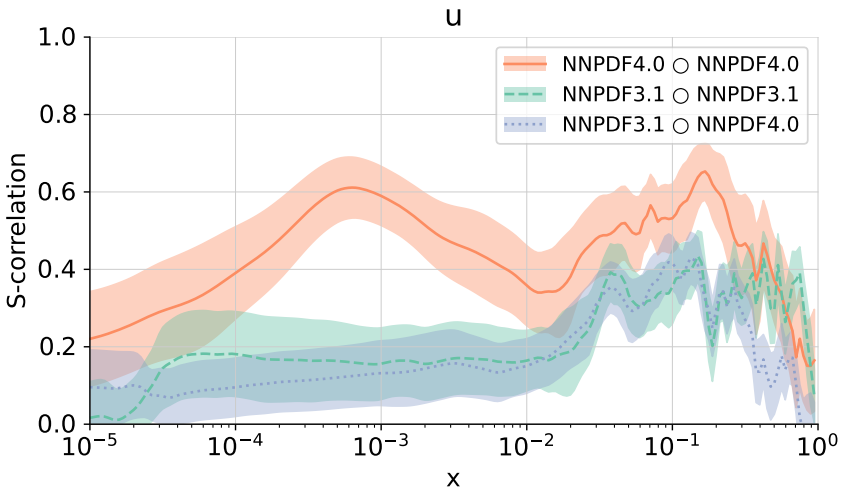
THE DATA-INDUCED SELF-CORRELATION

- COMPUTE THE S-CORRELATION OF A PDF SET TO ITSELF
  = THE F-CORRELATION OF A PDF TO ITSELF
- USE TWO DIFFERENT SETS OF PDF REPLICAS FITTED TO THE SAME DATA REPLICAS
- DEVIATION OF CORRELATION FROM 100% MEASURES THE CORRELATION LOSS DUE TO UNCORRELATED METHODOLOGY
- COMPARE TO RESULT BETWEEN DIFFERENT METHODOLOGIES (nnpdf3.1& nnpdf4.0)

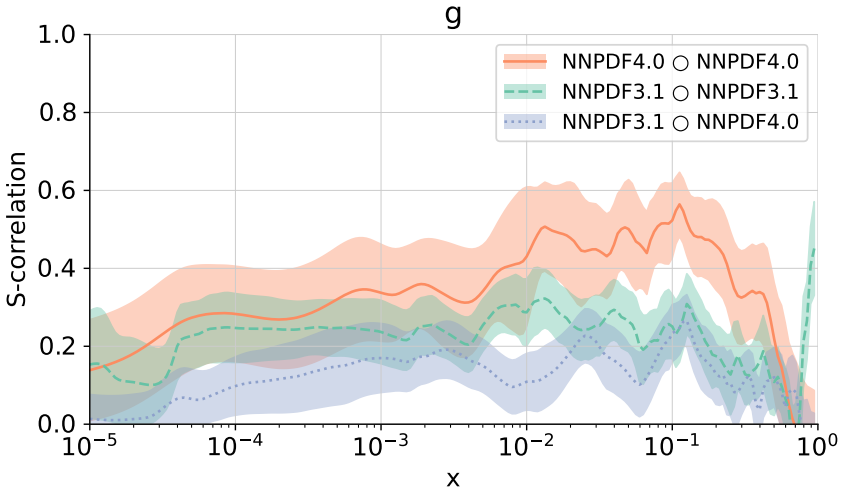
$$\langle u(x,Q_0^2)u(x,Q_0^2)\rangle = \frac{1}{N}\sum_{r=1}^{N} u^{(r,r',3.1)}(x,Q_0^2)u^{(r,r'',4.0)}(x,Q_0^2)$$

- HIGHER CORRELATION (nnpdf4.0) $\Rightarrow$ MORE EFFICIENT METHODOLOGY
- "WEAKEST LINK" (nnpdf3.1) : $3.1 \bigcirc 4.0 \approx 3.1 \bigcirc 3.1$

up quark



gluon



PDF COMBINATION

THE PDF4LHC SET

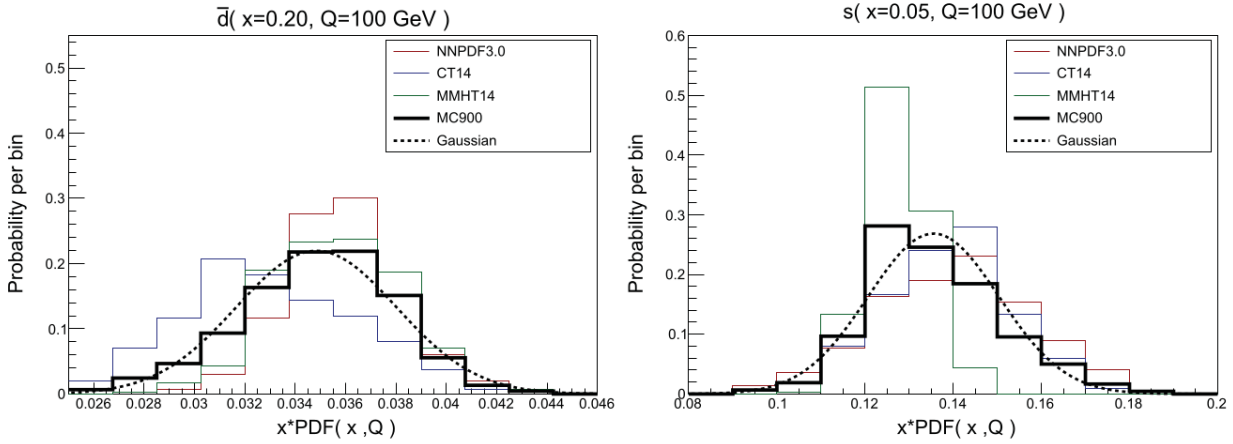
J. Phys. G: Nucl. Part. Phys. **43** (2016) 023001 (57pp) doi:10.1088/0954-3899/43/2/023001

Topical Review

PDF4LHC recommendations for LHC Run II

Jon Butterworth[1], Stefano Carrazza[2,4], Amanda Cooper-Sarkar[3], Albert De Roeck[4,5], Joël Feltesse[6], Stefano Forte[2], Jun Gao[7], Sasha Glazov[8], Joey Huston[9], Zahari Kassabov[2,10], Ronan McNulty[11], Andreas Morsch[4], Pavel Nadolsky[12], Voica Radescu[13], Juan Rojo[14] and Robert Thorne[1]

- SETS ASSUMD TO BE EQUALLY LIKELY
- BAYESIAN COMBINATION: COMBINED PDF SET
  $\Rightarrow$ 300 REPLICAS EACH FROM THREE UNDERLYING PDF SETS



COVARIANCE OF COMBINATION (TWO SETS, CT & NN):

$$\mathrm{Var}[u^{\mathrm{comb}}] = \frac{1}{2}(\mathrm{Var}[u^{\mathrm{CT}}] + \mathrm{Var}[u^{\mathrm{NN}}]) + \frac{1}{4}\langle u^{\mathrm{CT}}\rangle - \langle u^{\mathrm{NN}}\rangle)^2$$

- UNCERTAINTY ON COMBINATION BIGGER THAN AVERAGE UNCERTAINTY IF CENTRAL VALUES DISAGREE

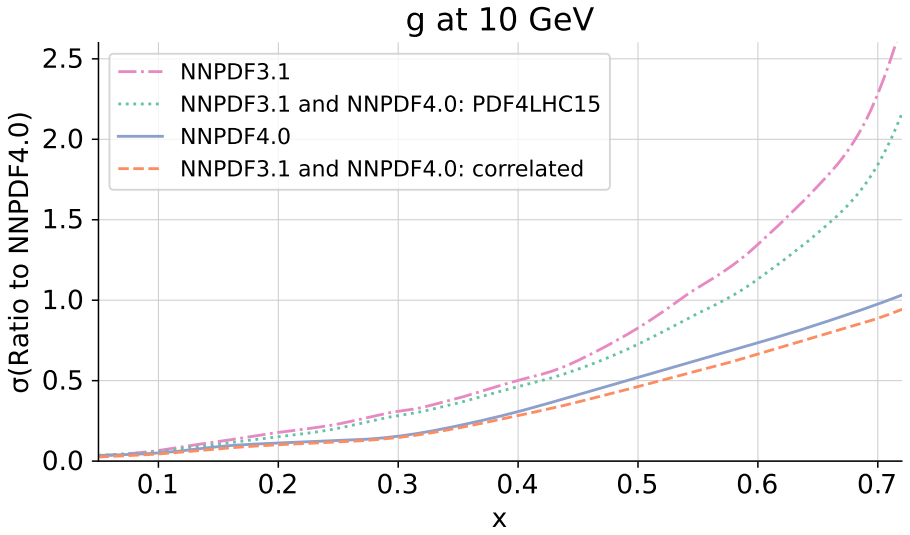
CORRELATED COMBINATION?

- CORRELATED MEASUREMENTS $\Rightarrow$ WEIGHTED AVERAGE
- UNCERTAINTY ON AVERAGE SMALLER THAN EITHER OF COMBINED EQUAL IF 100% CORRELATED
- EQUAL UNCERTAINTIES $\mathrm{Var}[u^{\mathrm{CT}}] = \mathrm{Var}[u^{\mathrm{NN}}]$: $\mathrm{Var}[u^{\mathrm{comb}}] = \frac{1}{2}\left(1+\rho[u^{\mathrm{CT}},u^{\mathrm{NN}}]\right)\mathrm{Var}[u^{\mathrm{NN}}]$

PROBLEMS

- DOES NOT ACCOUNT FOR BIAS (DIFFERENT CENTRAL VALUES)
- HOW DO COMPUTE CORRELATIONS RELIABLY?
- IF DATA-INDUCED USED, CORRELATION UNDERESTIMATED $\Rightarrow$ UNCERTAINTY UNDERESTIMATED

UNCERTAINTY ON COMBINATION OF NNPDF3.1 & 4.0
PDF4LHC VS. CORRELATED



OUTLOOK

I DID NOR MENTION. . .

- THEORY UNCERTAINTIES AND MISSING HIGHER ORDERS
- DATA INCONSISTENCIES

AND WHAT IF. . .

- CLOSURE AND FUTURE TEST INDICATORS ARE ADOPTED AS FIGURES OF MERIT

GOING TO THE HEART OF GENERALIZATION IN MACHINE LEARNING

EXTRAS

LHC DATA

- CUTOFF DATE AROUND 06/2020
- DIJETS NOW INCLUDED ALONG WITH JETS

  CANNOT INCLUDE SIMULTANEOUSLY FROM SAME UNDERLYING DATASET

LHCB

Data set	NNPDF4.0	NNPDF3.1	ABMP16	CT18	MSHT20
LHCb Z 940 pb	✓	✓	✗	✗	✓
LHCb $Z \to ee$ 2 fb	✓	✓	✓	✓	✓
LHCb $W, Z \to \mu$ 7 TeV	✓	✓	✓	✓	✓
LHCb $W, Z \to \mu$ 8 TeV	✓	✓	✓	✓	✓
LHCb $Z \to \mu\mu, ee$ 13 TeV	✓	✗	✗	✗	✗

ATLAS

Data set	NNPDF4.0	NNPDF3.1	ABMP16	CT18	MSHT20
ATLAS W, Z 7 TeV (2010)	✓	✓	✓	✓	✓
ATLAS W, Z 7 TeV (2011)	✓	✓	✗	✓	✓
ATLAS low-mass DY 7 TeV	✓	✓	✗	✗	✗
ATLAS high-mass DY 7 TeV	✓	✓	✗	✗	✓
ATLAS W 8 TeV	✓	✗	✗	✗	✓
ATLAS DY 2D 8 TeV	✓	✗	✗	✗	✓
ATLAS high-mass DY 2D 8 TeV	✓	✗	✗	✗	✓
ATLAS $\sigma_{W,Z}$ 13 TeV	✓	✗	✓	✗	✗
ATLAS W^++jet 8 TeV	✓	✗	✗	✗	✓
ATLAS Z p_T 8 TeV	✓	✓	✗	✓	✓
ATLAS $\sigma_{t\bar{t}}^{\text{tot}}$ 7, 8 TeV	✓	✓	✓	✗	✗
ATLAS $\sigma_{t\bar{t}}^{\text{tot}}$ 13 TeV	✓	✓	✓	✗	✗
ATLAS $t\bar{t}$ lepton+jets 8 TeV	✓	✓	✗	✓	✓
ATLAS $t\bar{t}$ dilepton 8 TeV	✓	✗	✗	✗	✓
ATLAS single-inclusive jets 7 TeV, R=0.6	✗	✓	✗	✓	✓
ATLAS single-inclusive jets 8 TeV, R=0.6	✓	✗	✗	✗	✗
ATLAS dijets 7 TeV, R=0.6	✓	✗	✗	✗	✗
ATLAS direct photon production 13 TeV	✓	✗	✗	✗	✗
ATLAS single top R_t 7, 8, 13 TeV	✓	✗	✓	✗	✗
ATLAS single top diff. 7, 8 TeV	✓	✗	✗	✗	✗
ATLAS single top diff. 8 TeV	✓	✗	✗	✗	✗

CMS

Data set	NNPDF4.0	NNPDF3.1	ABMP16	CT18	MSHT20
CMS W electron asymmetry 7 TeV	✓	✓	✗	✓	✓
CMS W muon asymmetry 7 TeV	✓	✓	✓	✓	✗
CMS Drell-Yan 2D 7 TeV	✓	✓	✗	✗	✓
CMS W rapidity 8 TeV	✓	✓	✓	✓	✓
CMS Z p_T 8 TeV	✓	✓	✗	✓	✗
CMS $W + c$ 7 TeV	✓	✓	✗	✗	✓
CMS $W + c$ 13 TeV	✓	✗	✗	✗	✗
CMS single-inclusive jets 2.76 TeV	✗	✓	✗	✗	✓
CMS single-inclusive jets 7 TeV	✗	✓	✗	✓	✓
CMS dijets 7 TeV	✓	✗	✗	✗	✗
CMS single-inclusive jets 8 TeV	✓	✗	✗	✓	✓
CMS 3D dijets 8 TeV	✗	✗	✗	✗	✗
CMS $\sigma_{t\bar{t}}^{\text{tot}}$ 5 TeV	✓	✗	✓	✗	✗
CMS $\sigma_{t\bar{t}}^{\text{tot}}$ 7, 8 TeV	✓	✓	✓	✗	✓
CMS $\sigma_{t\bar{t}}^{\text{tot}}$ 13 TeV	✓	✓	✓	✗	✗
CMS $t\bar{t}$ lepton+jets 8 TeV	✓	✓	✗	✗	✓
CMS $t\bar{t}$ 2D dilepton 8 TeV	✓	✗	✗	✓	✓
CMS $t\bar{t}$ lepton+jet 13 TeV	✓	✗	✗	✗	✗
CMS $t\bar{t}$ dilepton 13 TeV	✓	✗	✗	✗	✗
CMS single top $\sigma_t + \sigma_{\bar{t}}$ 7 TeV	✓	✗	✓	✗	✗
CMS single top R_t 8, 13 TeV	✓	✗	✓	✗	✗

THEORY

ELECTROWEAK CORRECTIONS

- `PineAPPL` FAST INTERFACE TO `Madgraph5_aMC@NLO` AVAILABLE

  (Schwan, Carrazza, Nocera, Zaro 2020)

  $\Rightarrow$ FULL NLO EW+QCD POSSIBLE

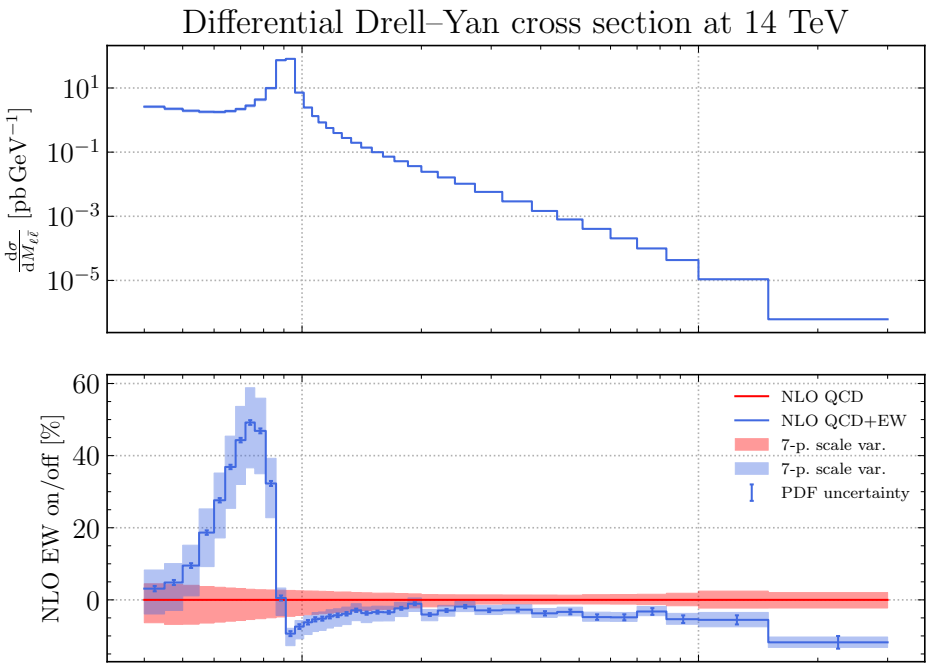
- DATA W/O FSR & PHOTON-INITIATED SUBTRACTION OFTEN NOT AVAILABLE

- CURRENTLY USED FOR DATASET SELECTION:

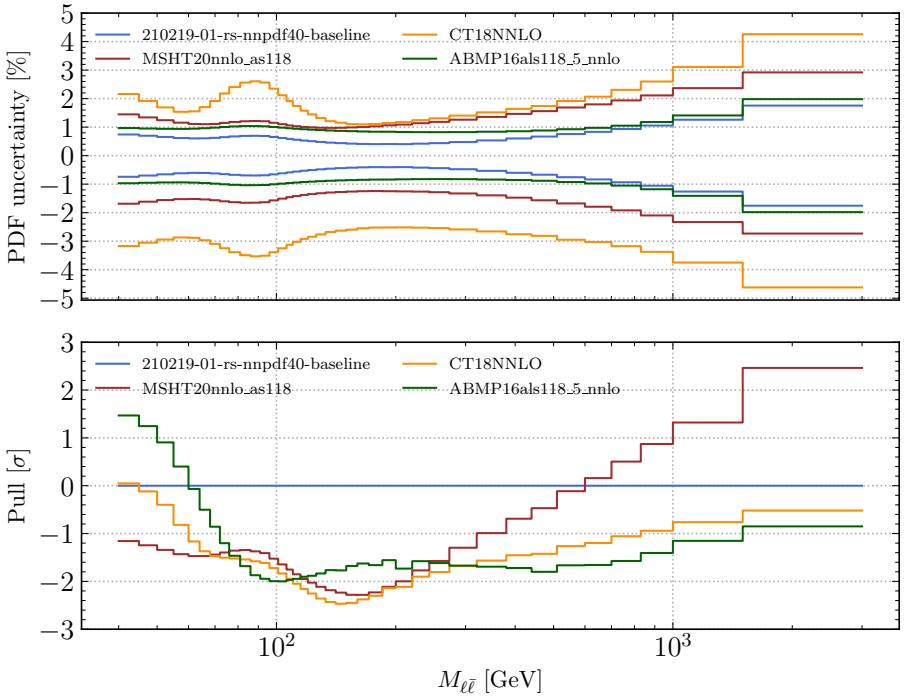
  $\Rightarrow$ DISCARDED IF EW CORRNS EXCEED THRESHOLD

EXAMPLE: DRELL-YAN AT 14 TEV

QCD+EW vs. QCD



PDF UNCERTAINTIES



NUCLEAR CORRECTIONS

- INCLUDED AS CONTRIBUTION TO COVARIANCE MATRIX (FULLY CORRELATED)

  (Ball, Nocera, Pearson, 2019)

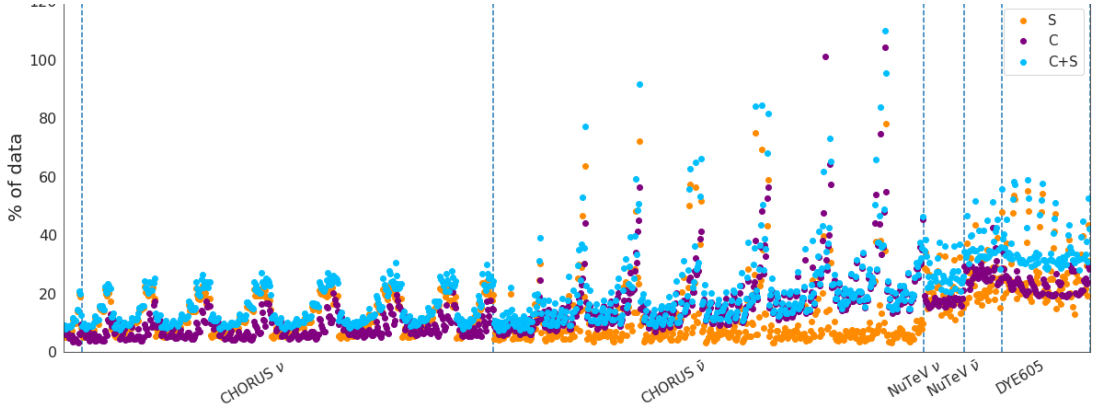
- COMPUTED AS SHIFT BETWEEN NUCLEAR & STANDARD PDF

- DEUTERIUM PDF DETERMINED FROM SELF-CONSISTENT NNPDF FIT

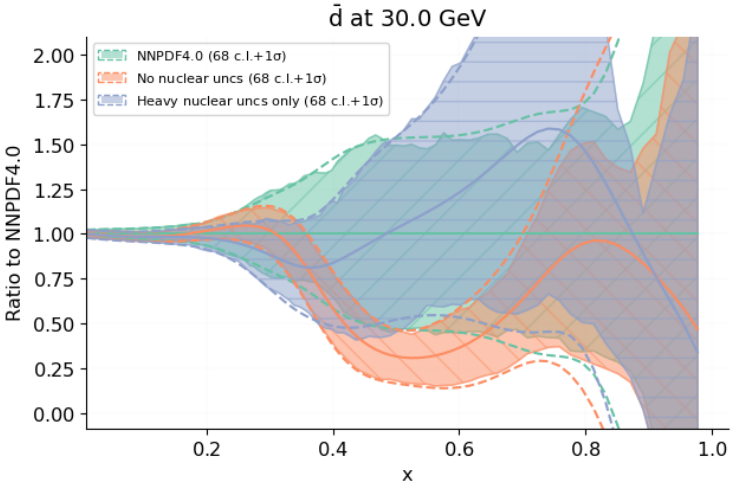
  (Ball, Nocera, Pearson, 2019)

- NUCLEAR PDFS FROM NNNPDF2.0 (Abdul Khalek, Ethier, Rojo, van Weelden, 2020)

DIAGONAL ERRORS



IMPACT ON PDFS



PDF POSITIVITY & INTEGRABILITY

- $\overline{\rm MS}$ PDFs ARE NON-NEGATIVE!(Candido, Hekhorn, Forte, 2020)
- PDF POSITIVITY IMPOSED (PREVIOUSLY: OBSERVABLE POSITIVITY)

  $\Rightarrow$ SMALLER LARGE x UNCERTAINTIES
- SEA NONSINGLET COMBINATIONS INTEGRABLE:

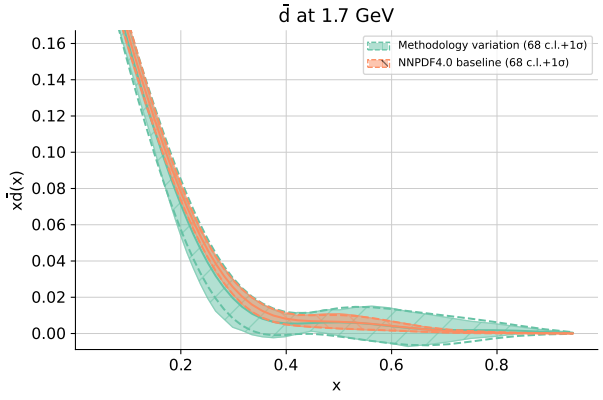
  GOTTFRIED $u + \bar{u} - (d + \bar{d})$

  STRANGENESS $u + \bar{u} + (d + \bar{d}) - 2(s + \bar{s})$

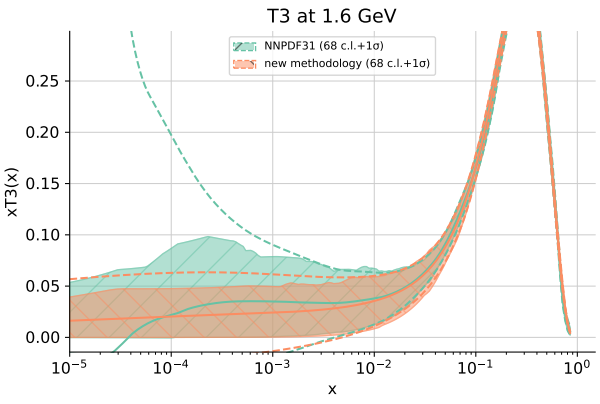
  $\Rightarrow$ SMALLER SMALL x UNCERTAINTIES

IMPACT ON PDFS

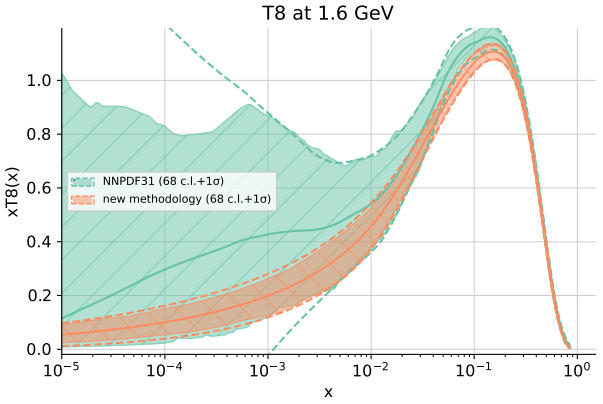
POS.: ANTIDOWN



INTEGR.: GOTTFRIED SR



INTEGR.: STRANGE SEA SR



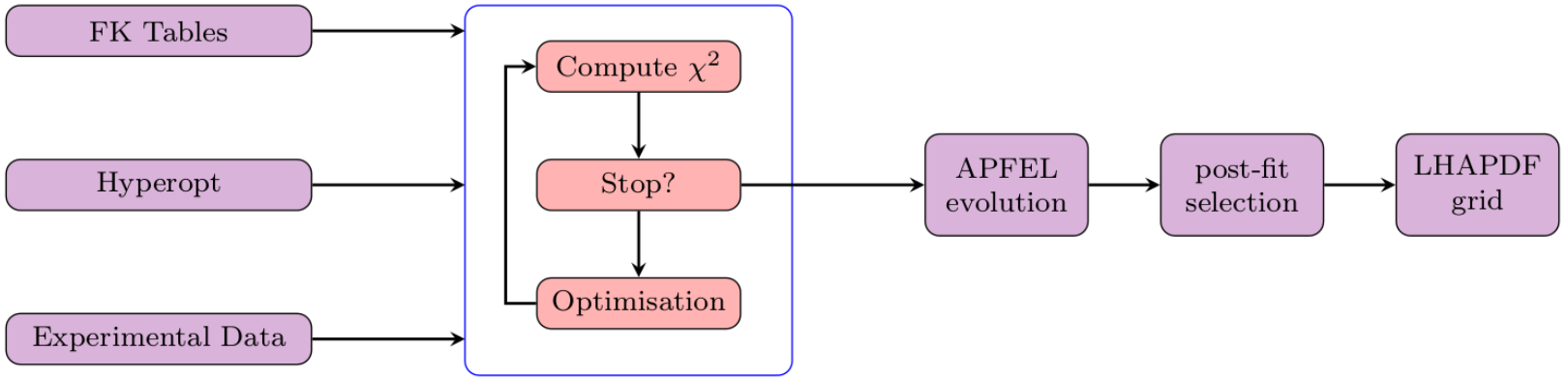
METHODOLOGY

THE NNPDF CODE STRUCTURE

- MODULAR PYTHON-BASED CODE
- HIGH DEGREE PARALLELIZATION & HARDWARE ACCELERATION

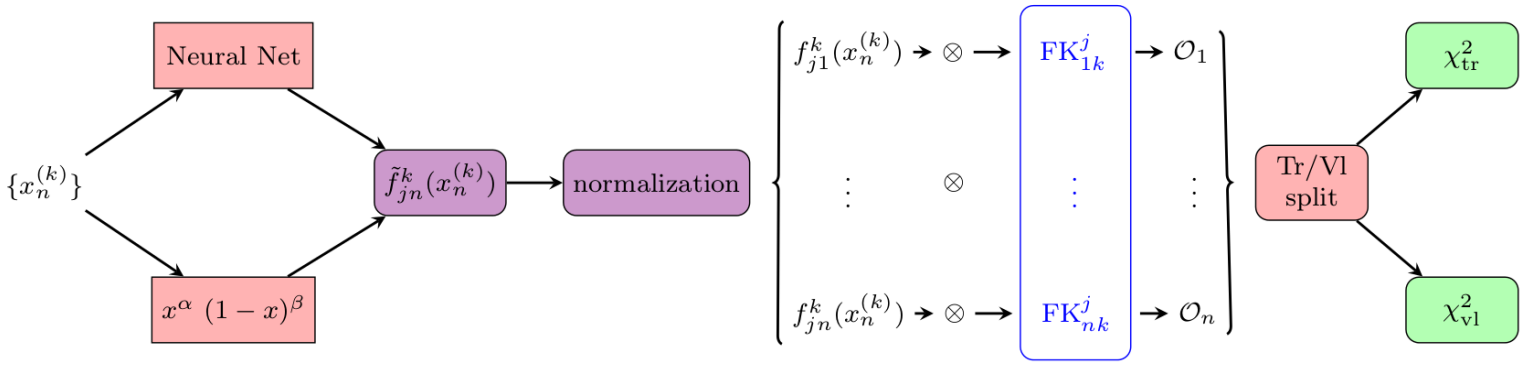
AVERAGE FITTING TIME PER REPLICA AND USE OF RESOURCES
SAME DATASET FOR OLD AND NEW METHODOLOGIES IN CPU AND GPU
CPU: INTEL(R) CORE(TM) I7-4770 AT 3.40GHZ; GPU: NVIDIA TITAN V

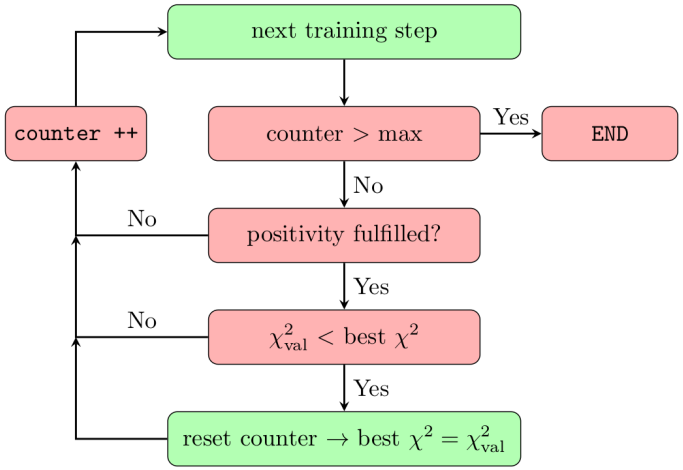
	NNPDF31 CODEBASE	NNPDF40 CODEBASE IN CPU	NNPDF40 CODEBASE IN GPU
TIME	15.2 H.	38 $\pm$ 5 MIN.	6.6 MIN.
RAM USE	1.5 GB	6.1 GB	NA



MINIMIZATION AND CROSS-VALIDATION

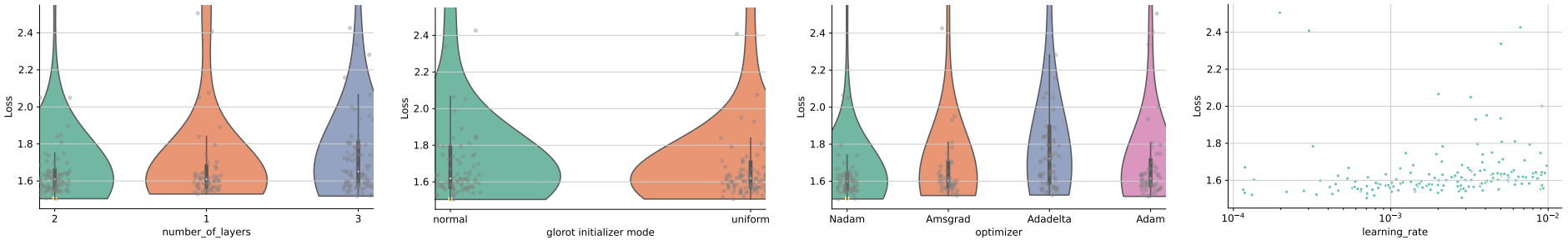
- DATA REPLICAS $\Rightarrow$ PDF REPLICAS
- EACH PDF REPLICA: PREPROCESSED NEURAL NET
- NEURAL NET $\Rightarrow$ OBSERVABLES
- RANDOM TRAINING-VALIDATION SPLIT, χ^2 TO TRAINING DATA REPLICAS MINIMIZED
- TRAINING STOPS IF VALIDATION χ^2 GROWS FOR A WHILE (PATIENCE)
- LOWEST VALIDATION χ^2 $\Rightarrow$ OPTIMAL FIT



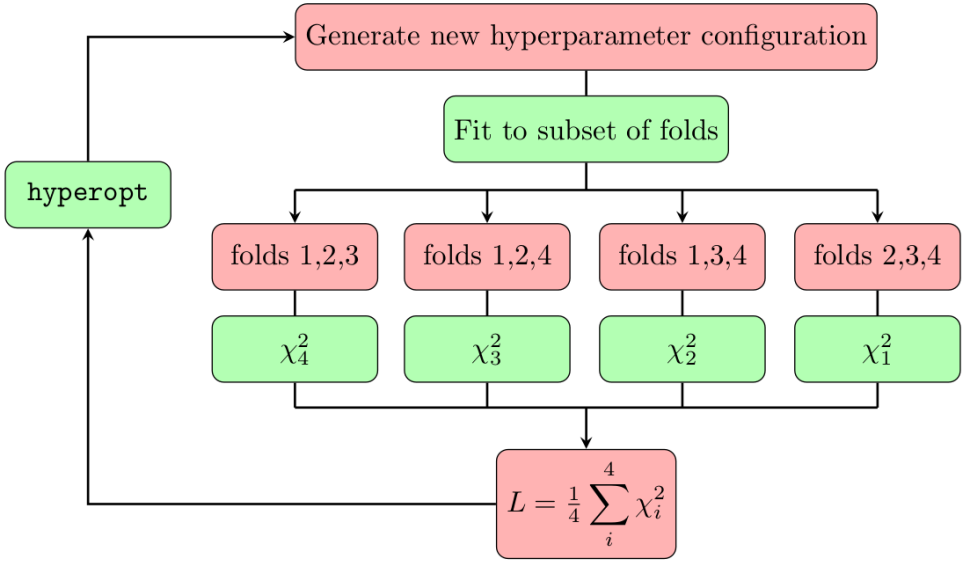

next training step
counter ++
counter > max
Yes
END
No
No
positivity fulfilled?
Yes
No
$\chi^2_{\text{val}} < \text{best } \chi^2$
Yes
reset counter → best $\chi^2 = \chi^2_{\text{val}}$

HYPEROPTIMIZATION

- PARAMETRIZATION AND MINIMIZATION PARAMETERS VARIED
- SCAN OF PARAMETER SPACE
- BAYESIAN UPDATING LEADS TO BEST METHODOLOGY



K-FOLDING



Fold 1		
CHORUS σ^{ν}_{CC}	HERA I+II inc NC e^+p 920 GeV	BCDMS p
LHCb Z 940 pb	ATLAS W,Z 7 TeV 2010	CMS Z p_T 8 TeV (p_T^{ll}, y_{ll})
DY E605 $\sigma^{p}_{\rm DY}$	CMS Drell-Yan 2D 7 TeV 2011	CMS 3D dijets 8 TeV
ATLAS single-$\bar{t}$ y (normalised)	ATLAS single top R_t 7 TeV	CMS $t\bar{t}$ rapidity $y_{t\bar{t}}$
CMS single top R_t 8 TeV		

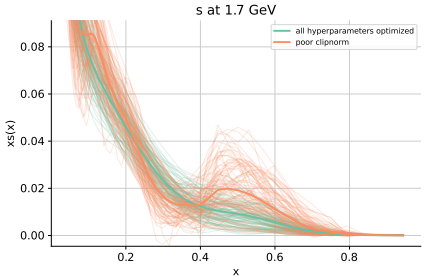
Fold 2		
HERA I+II inc CC e^-p	HERA I+II inc NC e^+p 460 GeV	HERA comb. $\sigma^{\rm red}_{bb}$
NMC p	NuTeV σ^{p}_{c}	LHCb $Z \to ee$ 2 fb
CMS W asymmetry 840 pb	ATLAS Z p_T 8 TeV (p_T^{ll}, M_{ll})	D0 $W \to \mu\nu$ asymmetry
DY E886 $\sigma^{p}_{\rm DY}$	ATLAS direct photon 13 TeV	ATLAS dijets 7 TeV, R=0.6
ATLAS single antitop y (normalised)	CMS $\sigma^{\rm tot}_{tt}$	CMS single top $\sigma_t + \sigma_{\bar{t}}$ 7 TeV

Fold 3		
HERA I+II inc CC e^+p	HERA I+II inc NC e^+p 575 GeV	NMC d/p
NuTeV σ^{ν}_{c}	LHCb $W,Z \to \mu$ 7 TeV	LHCb $Z \to ee$
ATLAS W,Z 7 TeV 2011 Central selection	ATLAS W^++jet 8 TeV	ATLAS HM DY 7 TeV
CMS W asymmetry 4.7 fb	DYE 866 $\sigma^{d}_{\rm DY}/\sigma^{p}_{\rm DY}$	CDF Z rapidity (new)
ATLAS $\sigma^{\rm tot}_{tt}$	ATLAS single top y_t (normalised)	CMS $\sigma^{\rm tot}_{tt}$ 5 TeV
CMS $t\bar{t}$ double diff. $(m_{t\bar{t}}, y_t)$		

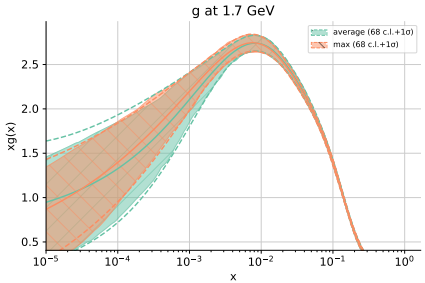
Fold 4		
CHORUS σ^{p}_{CC}	HERA I+II inc NC e^+p 820 GeV	LHCb $W,Z \to \mu$ 8 TeV
LHCb $Z \to \mu\mu$	ATLAS W,Z 7 TeV 2011 Fwd	ATLAS W^-+jet 8 TeV
ATLAS low-mass DY 2011	ATLAS Z p_T 8 TeV (p_T^{ll}, y_{ll})	CMS W rapidity 8 TeV
D0 Z rapidity	CMS dijets 7 TeV	ATLAS single top y_t (normalised)
ATLAS single top R_t 13 TeV	CMS single top R_t 13 TeV	

- HYPEROPTIMIZATION $\Rightarrow$ OVERFITTING (χ^2 TOO GOOD)
- CHECK GENERALIZATION POWER: K-FOLDING
  - DIVIDE DATA IN FOLDS
  - EXCLUDE ONE FOLD IN TURN FROM FIT
  - OPTIMIZE ON THE χ^2 OF THE EXCLUDED FOLDS
  - BEST AVERAGE OR BEST WORST

NO K-FOLDING



K-FOLDING VARIATION

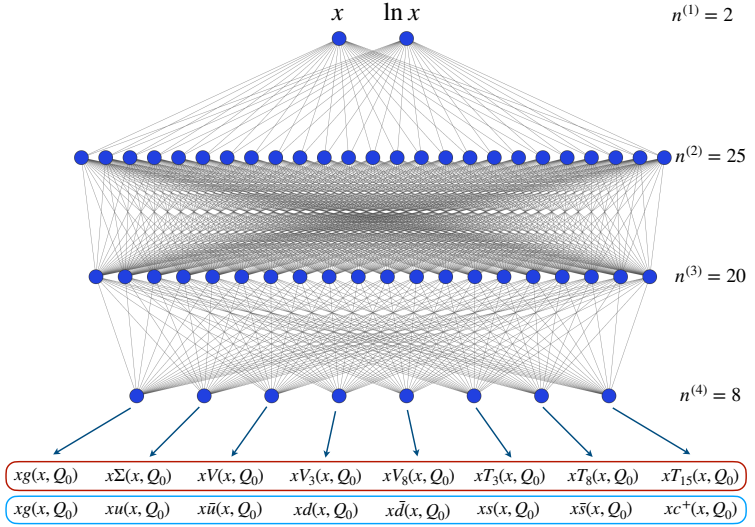


THE ML METHODOLOGY

HYPEROPTIMIZED PARAMETERS

Parameter	NNPDF4.0	L as in Eq. (3.21)	Flavour basis Eq. (3.2)
Architecture	25-20-8	70-50-8	7-26-27-8
Activation function	hyperbolic tangent	hyperbolic tangent	sigmoid
Initializer	`glorot_normal`	`glorot_uniform`	`glorot_normal`
Optimizer	`Nadam`	`Adadelta`	`Nadam`
Clipnorm	6.0×10^{-6}	5.2×10^{-2}	2.3×10^{-5}
Learning rate	2.6×10^{-3}	2.5×10^{-1}	2.6×10^{-3}
Maximum # epochs	17×10^{3}	45×10^{3}	45×10^{3}
Stopping patience	10% of max epochs	12% of max epochs	16% of max epochs
Initial positivity $\Lambda^{(\rm pos)}$	185	106	2
Initial integrability $\Lambda^{(\rm int)}$	10	10	10

NN ARCHITECTURE



- HYPEROPT ADAPTS TO EXTERNAL CHOICES (E.G. PARAMETRIZATION BASIS)
- SIMILAR RESULTS CAN BE OBTAINED WITH RATHER DIFFERENT SETTINGS

DATASET SELECTION

DATA "TENSION"

PROBLEMATIC DATA

$$\frac{\chi^2 - 1}{\sigma_{\chi^2}} \gg 1 \Leftrightarrow \text{POOR FIT QUALITY}$$

- MISSING HIGHER-ORDER CORRECTIONS
- NO RESUMMATION WHERE NEEDED
- ILL-CONDITIONED COVARIANCE MATRIX
- EXPERIMENTAL ISSUES

THE WEIGHTED FIT METHOD

- FLAG PROBLEMATIC DATASETS:
  - LARGE χ^2
  - LARGE FROBENIUS NUMBER OF COVMAT (EIGENVALUES TOO SMALL)
- REPEAT GLOBAL FIT WITH LARGE WEIGHT GIVEN TO EACH PROBLEMATIC DATASET IN TURN

- χ^2 OF DATASET
  - UNCHANGED $\Rightarrow$ INTERNAL INCONSISTENCY

  - DECREASES $\Rightarrow$ TENSION

- GLOBAL χ^2
  - UNCHANGED $\Rightarrow$ CONSISTENT, KEEP
  - INCREASES $\Rightarrow$ INCONSISTENT, DISCARD

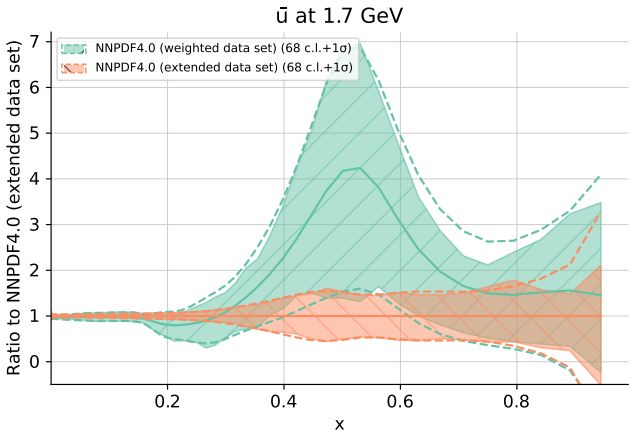
DATA "TENSION"

PROBLEMATIC DATA

$$\frac{\chi^2 - 1}{\sigma_{\chi^2}} \gg 1 \Leftrightarrow \text{POOR FIT QUALITY}$$

INCONSISTENT!

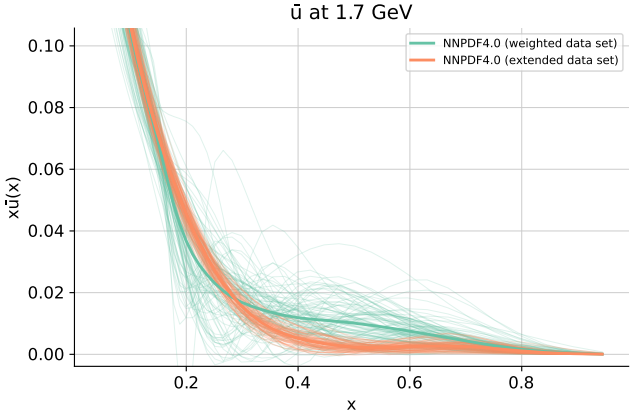
- MISSING HIGHER-ORDER CORRECTIONS
- NO RESUMMATION WHERE NEEDED
- ILL-CONDITIONED COVARIANCE MATRIX
- EXPERIMENTAL ISSUES



THE WEIGHTED FIT METHOD

INCONSISTENT!

- FLAG PROBLEMATIC DATASETS:
  - LARGE χ^2
  - LARGE FROBENIUS NUMBER OF COVMAT (EIGENVALUES TOO SMALL)
- REPEAT GLOBAL FIT WITH LARGE WEIGHT GIVEN TO EACH PROBLEMATIC DATASET IN TURN



- χ^2 OF DATASET
  - UNCHANGED $\Rightarrow$ INTERNAL INCONSISTENCY

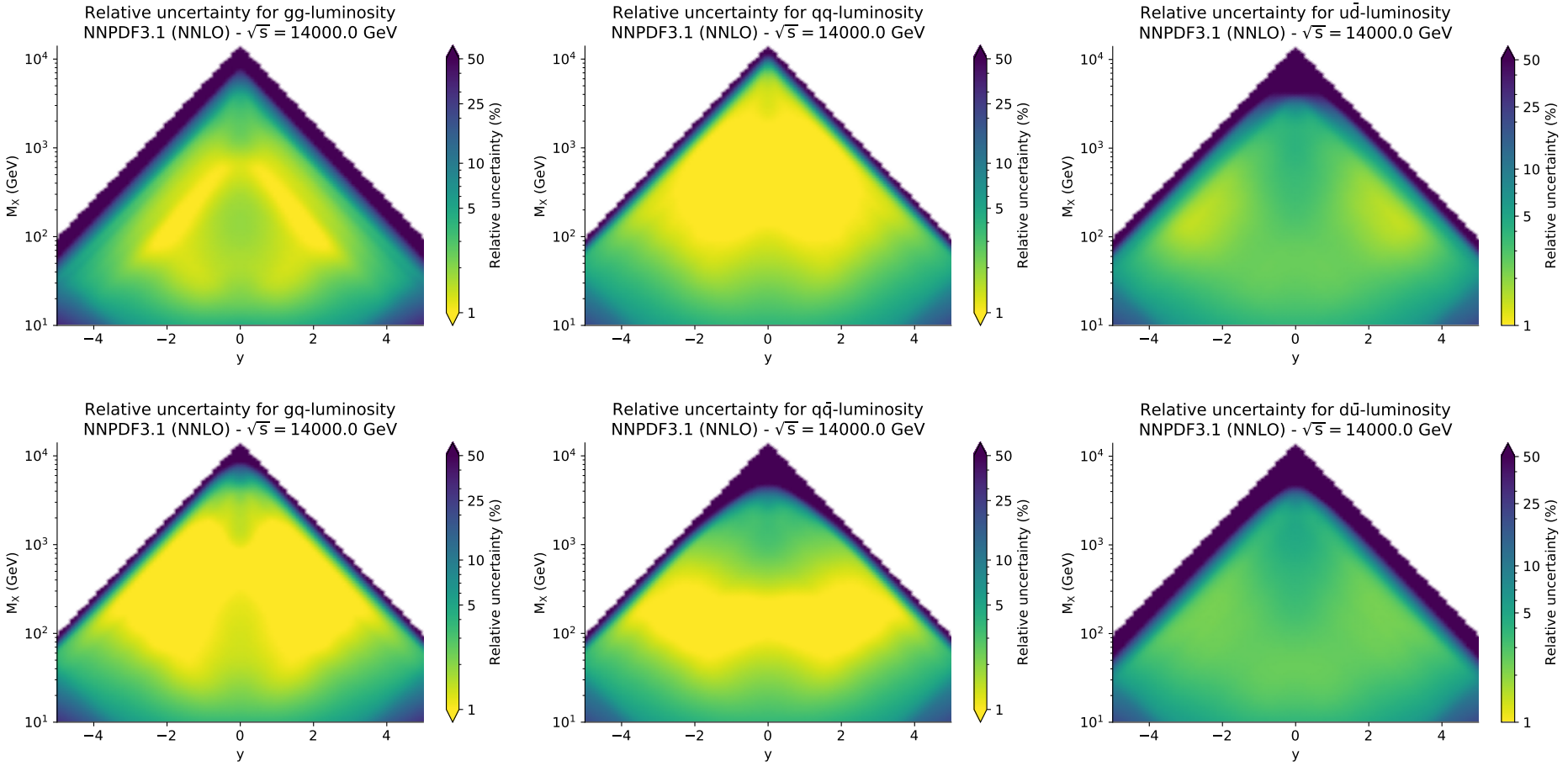
  - DECREASES $\Rightarrow$ TENSION

- GLOBAL χ^2
  - UNCHANGED $\Rightarrow$ CONSISTENT, KEEP
  - INCREASES $\Rightarrow$ INSONSISTENT, DISCARD

UNCERTAINTIES

UNCERTAINTIES: FROM NNPDF3.1...

GLUON SINGLET FLAVORS

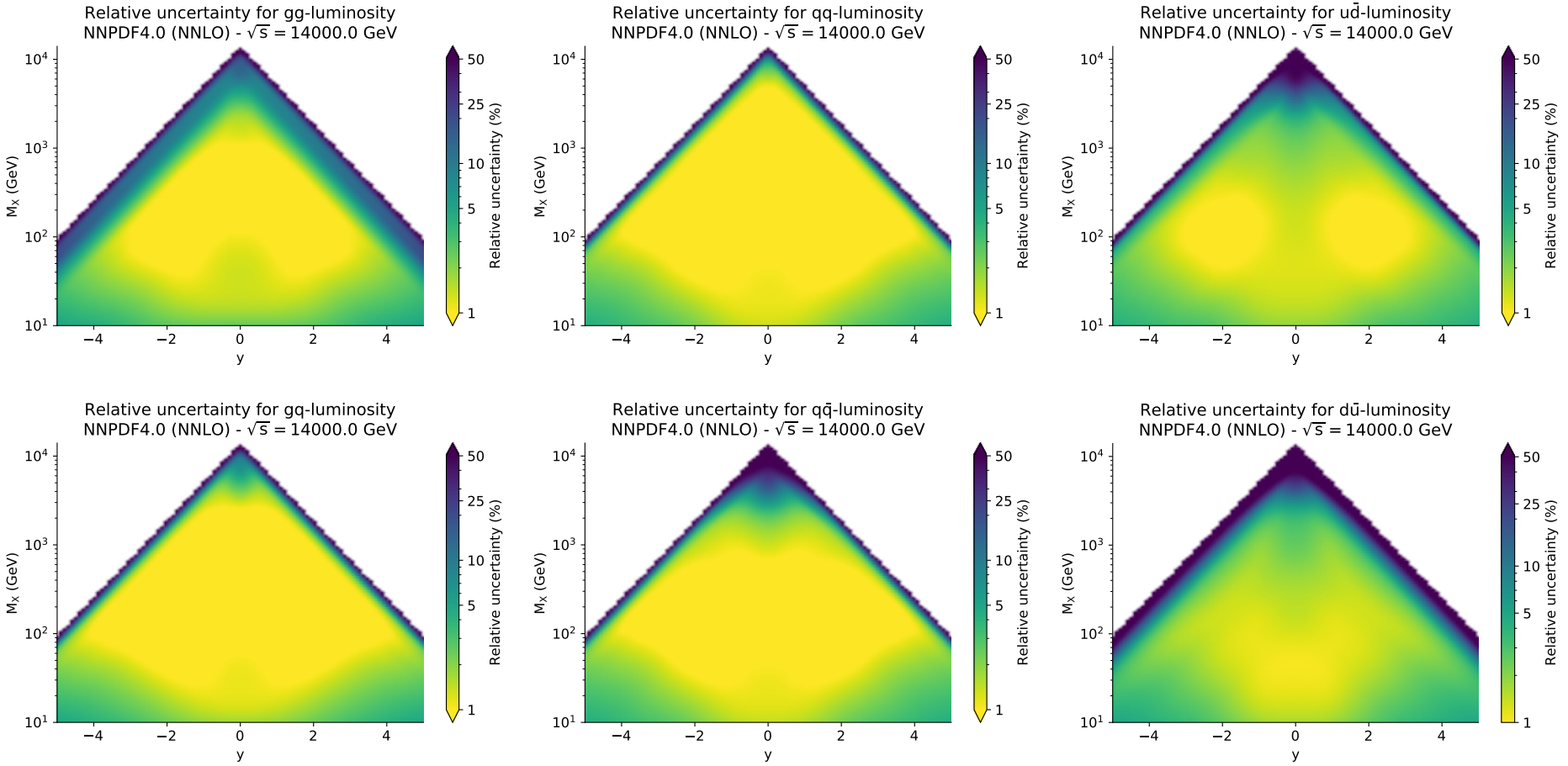


- TYPICAL UNCERTAINTIES IN DATA REGION: SINGLET $\sim 3\%$, NONSINGLET $\sim 5\%$

- DATA REGION: $10^2 \lesssim M_X \lesssim 10^3$ TeV, $-2 \lesssim y \lesssim 2$

UNCERTAINTIES: ...TO NNPDF4.0

GLUON SINGLET FLAVORS

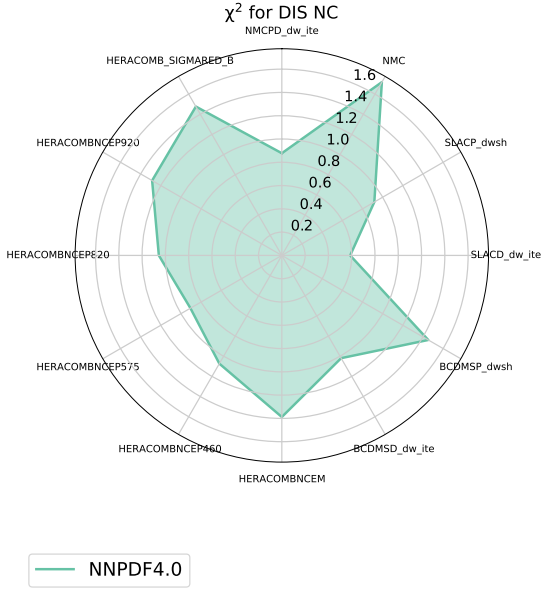


- TYPICAL UNCERTAINTIES IN DATA REGION: SINGLET $\sim 1\%$, NONSINGLET $\sim 2-3\%$

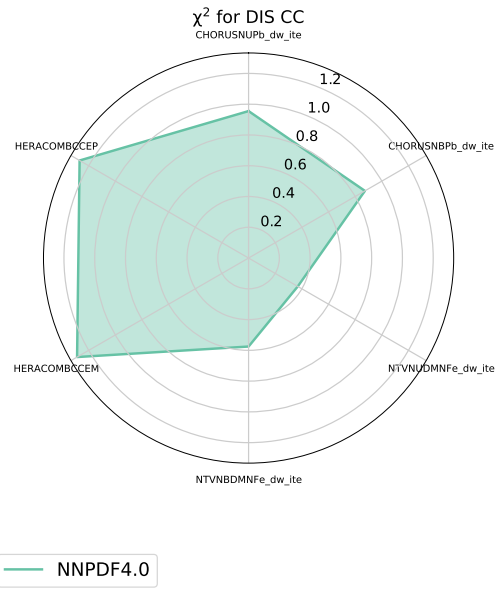
- DATA REGION: $10 \lesssim M_X \lesssim 3 \cdot 10^3$ TEV, $-4 \lesssim y \lesssim 4$

NNPDF4.0 FIT QUALITY

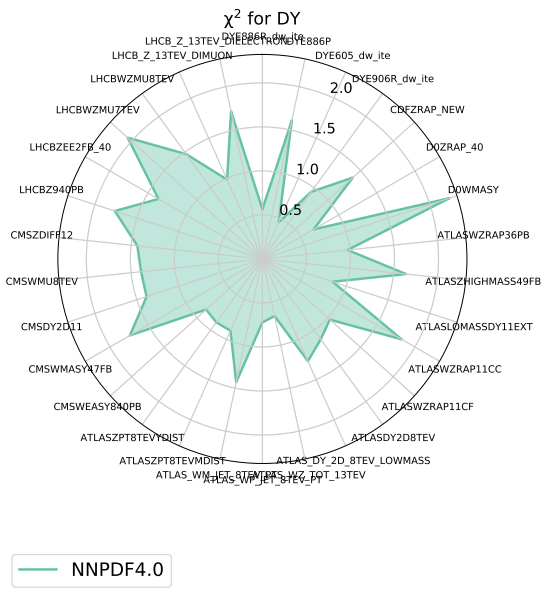
DIS NC



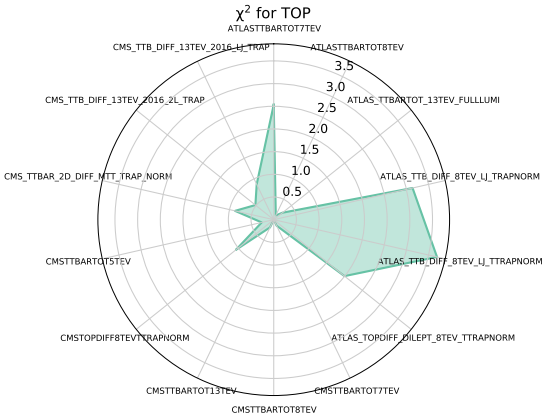
DIS CC



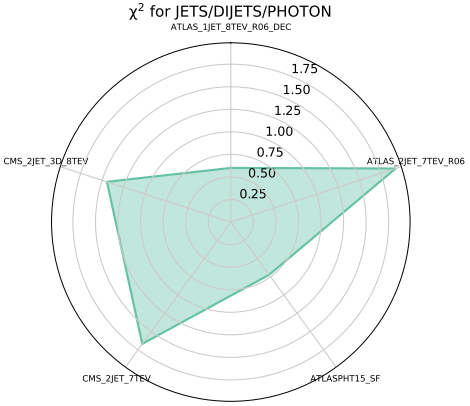
DY



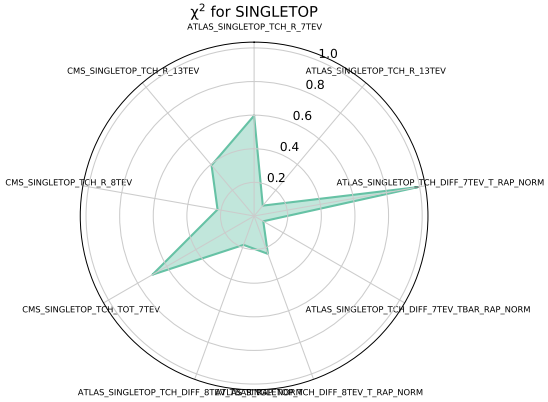
TOP PAIR



JETS



SINGLE TOP



- OUTLIERS $\Rightarrow$ FLAGGED DATASETS

- LARGE DATASETS (DIS) WELL FITTED

STABILITY

PARAMETRIZATION BASIS

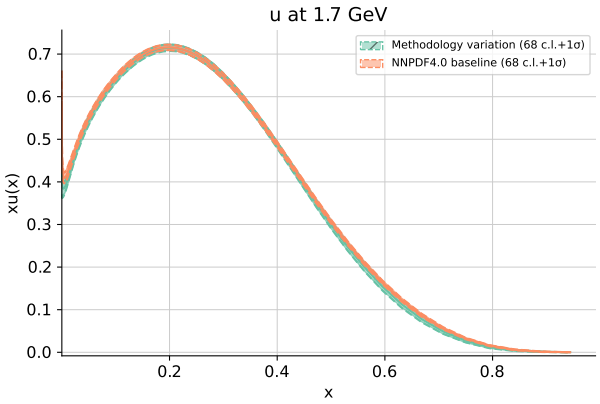
- PDFs BY DEFAULT PARAMETRIZED IN "EVOLUTION BASIS":

  SINGLET $\Sigma = \sum_i q_i + \bar{q}_i$, VALENCE $V = \sum_i q_i - \bar{q}_i$, TRIPLET $T_3 = u + \bar{u} - (d + \bar{d})$, . . .

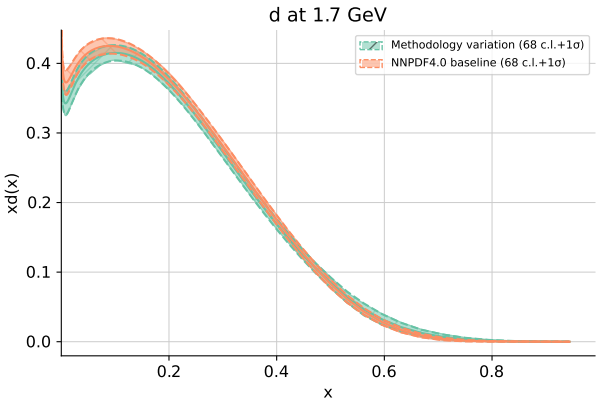
- WHAT IF ONE CHOOSES THE "FLAVOR BASIS": u, $\bar{u}$, d, $\bar{d}$?

- COMPLETE STABILITY OF RESULTS!

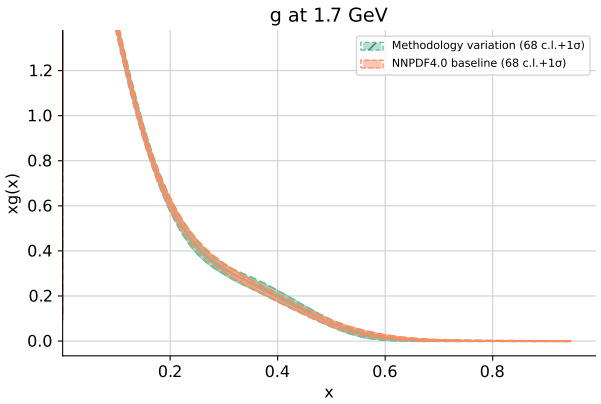
UP



DOWN



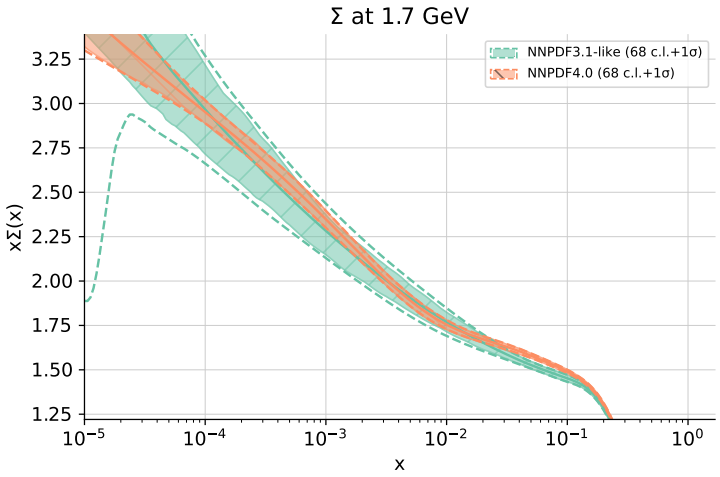
GLUON



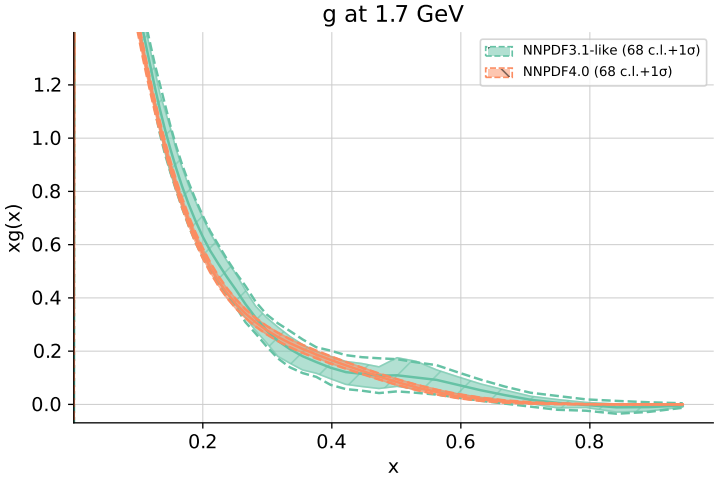
NNPDF4.0 vs. NNPDF3.1

- FULL BACKWARD COMPATIBILITY
- SUBSTANTIAL REDUCTION IN UNCERTAINTY

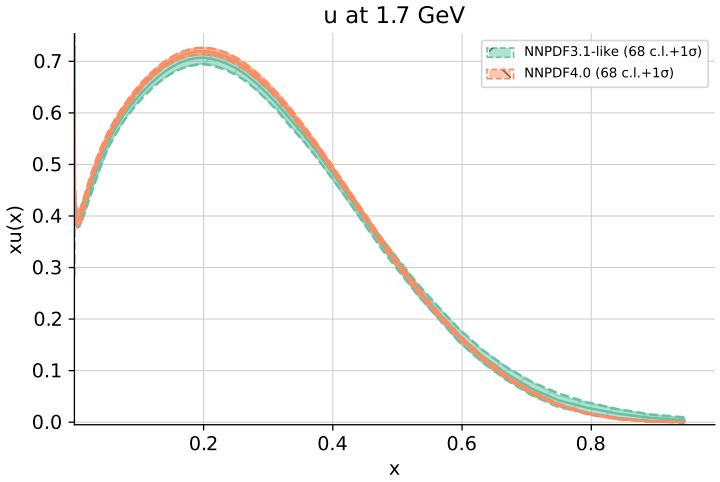
SINGLET



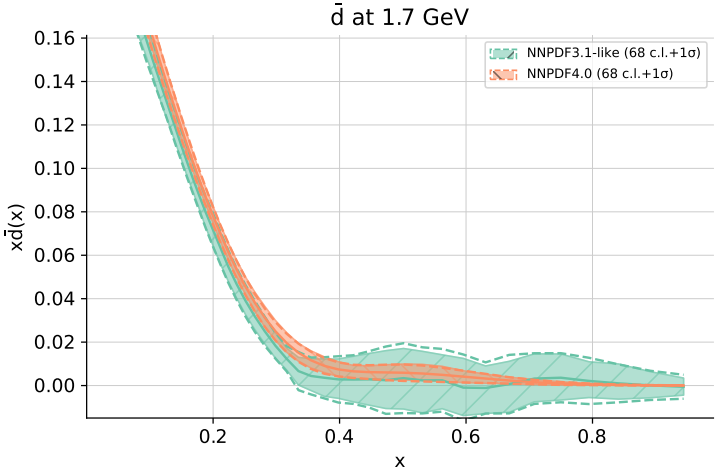
GLUON



UP



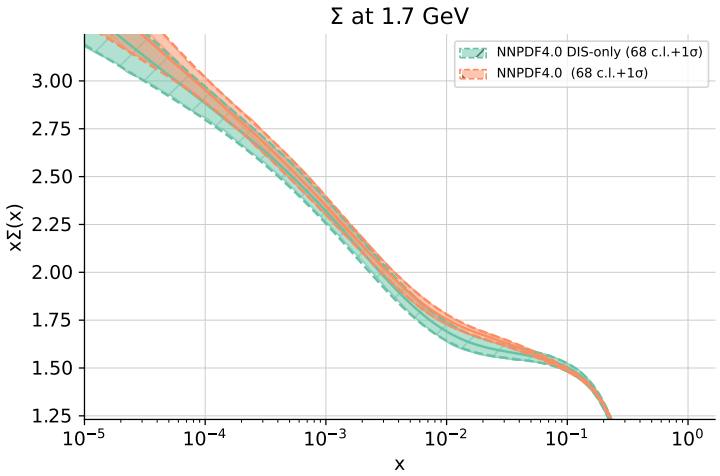
ANTIDOWN



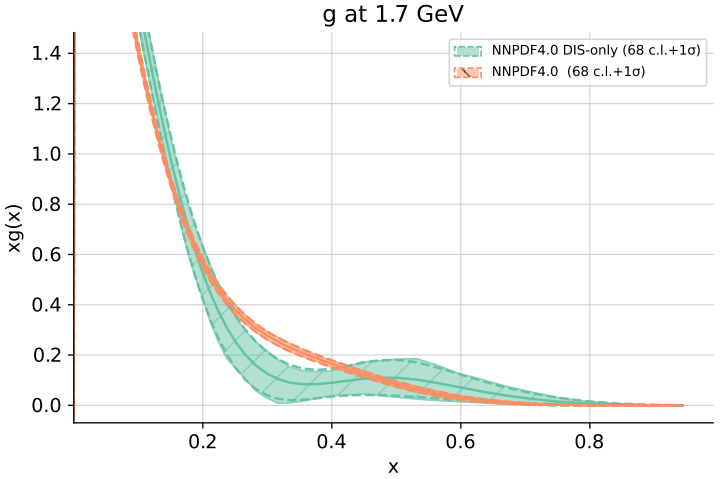
NNPDF4.0 vs DIS-only

- DIS-only fit NO LONGER COMPETITIVE
- HADRONIC DATA NEEDED FOR PRECISION

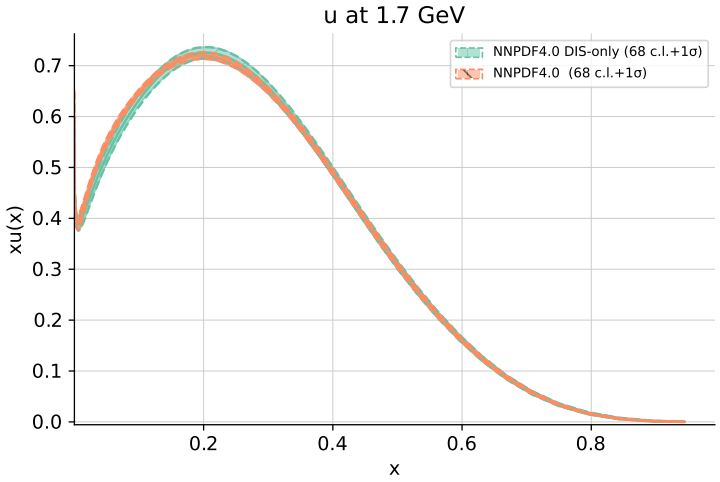
SINGLET



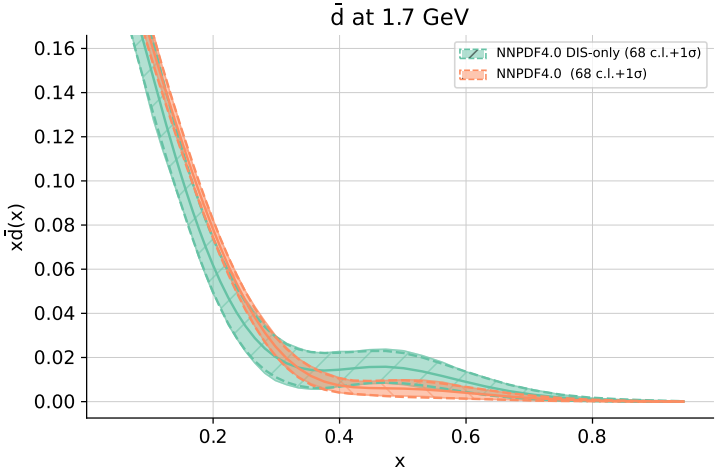
GLUON



UP



ANTIDOWN

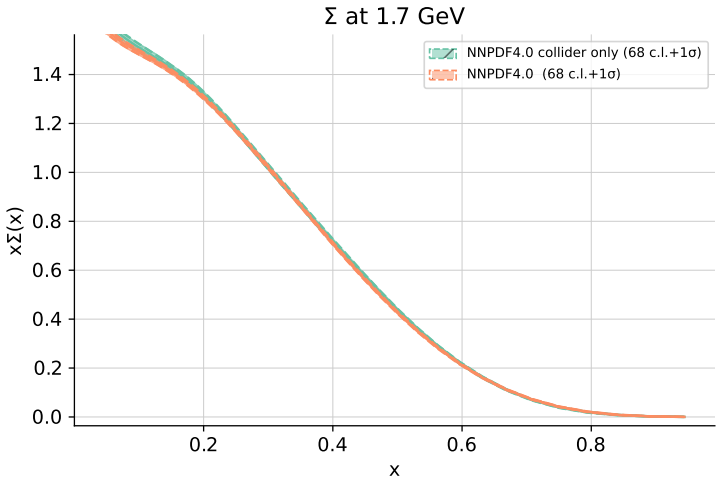


NNPDF4.0 VS COLLIDER ONLY

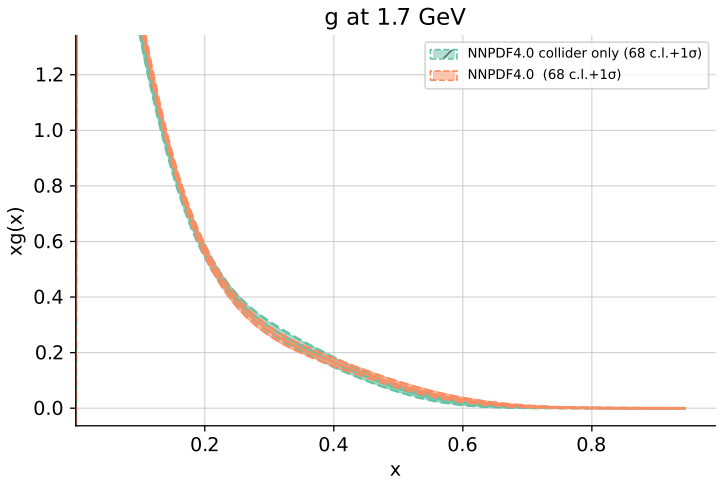
- COLLIDER ONLY COMPETITIVE!

- ONLY DEUTERIUM FIXED-TARGET DATA STILL RELEVANT

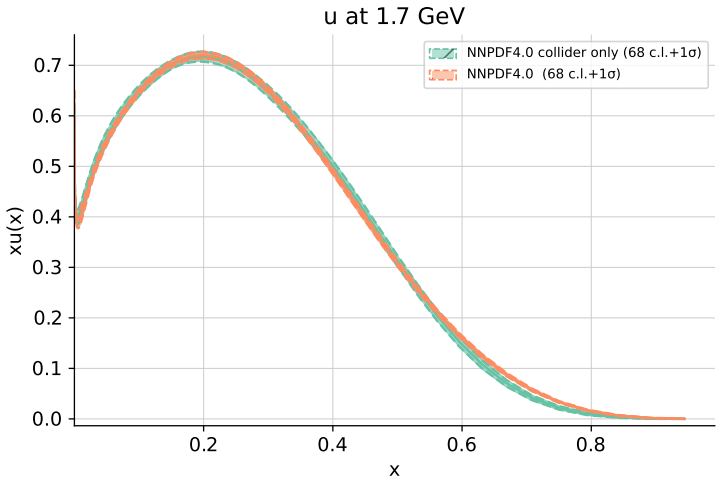
SINGLET



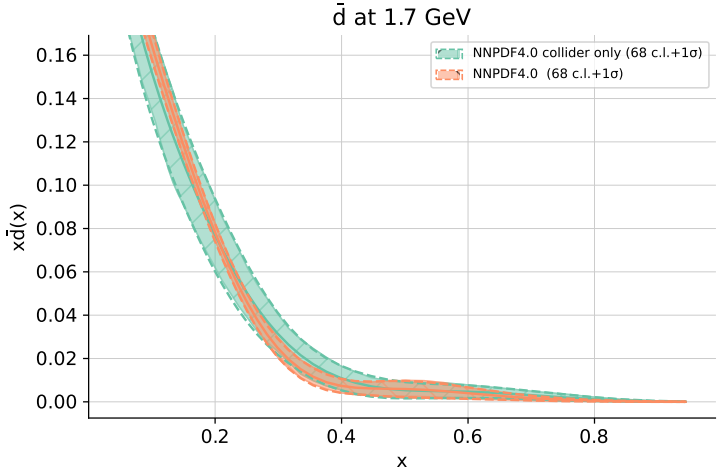
GLUON



UP



ANTIDOWN

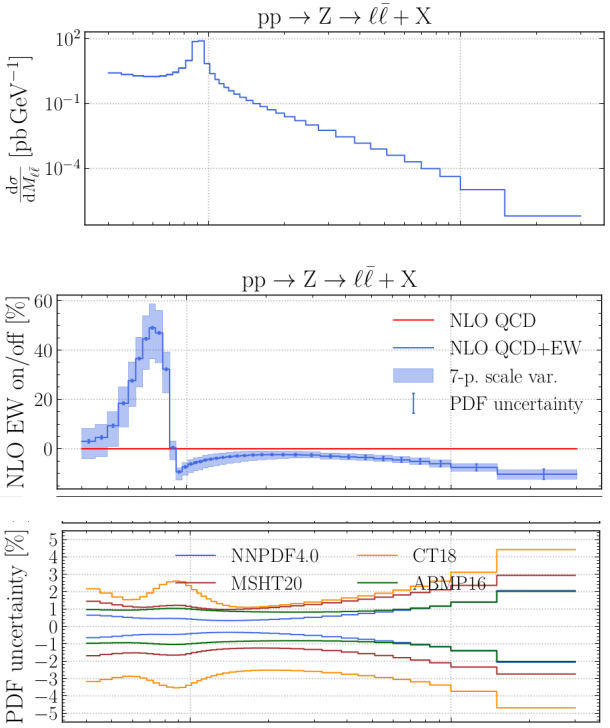


PHENOMENOLOGY

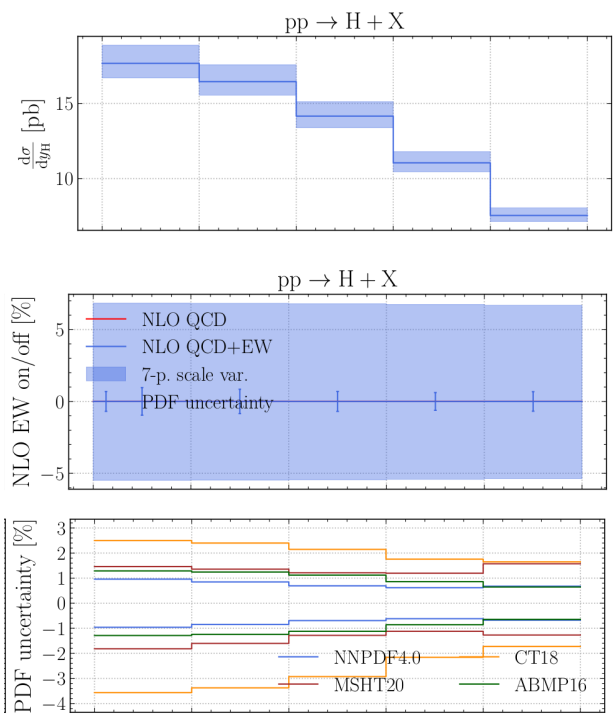
REPRESENTATIVE PROCESSES

- EW CORRECTIONS UNDER CONTROL
- SMALL UNCERTAINTIES

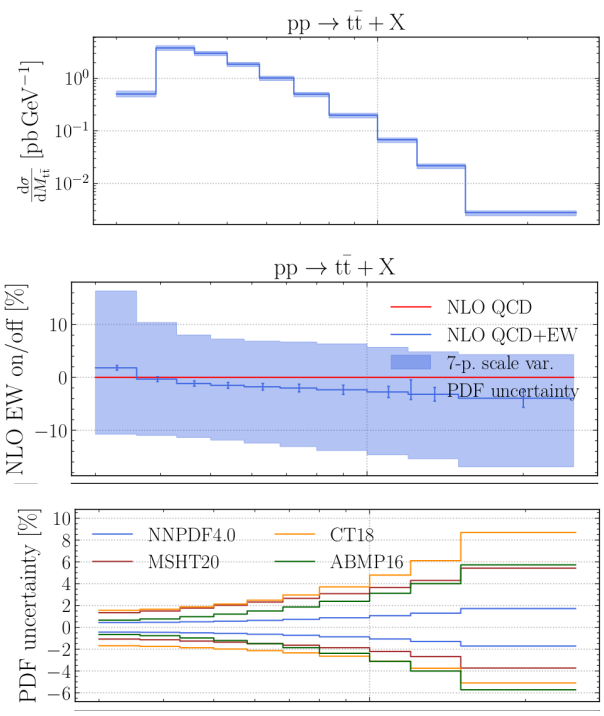
Z INVARIANT MASS



HIGGS RAPIDITY



TOP PAIR INV. MASS



DELIVERY

AN OPEN SOURCE CODE!

- THE FULL NNPDF CODE WILL BE MADE PUBLIC!
- INCLUDING HYPEROPTIMIZATION, EVOLUTION, THEORY, FITTING, VISUALIZATION
- FULLY DOCUMENTED CODE

An open-source machine learning framework for global analyses of parton distributions

The NNPDF Collaboration: Richard D. Ball · Stefano Carrazza · Juan Cruz-Martinez · Luigi Del Debbio · Stefano Forte · Tommaso Giani · Shayan Iranipour · Zahari Kassabov · Jose I. Latorre · Emanuele R. Nocera · Rosalyn L. Pearson · Juan Rojo · Roy Stegeman · Christopher Schwan · Maria Ubiali · Cameron Voisey · Michael Wilson

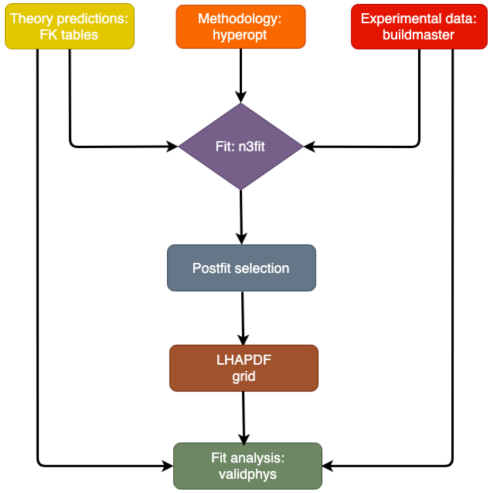


Fig. 2.1. Workflow for an NNPDF fit