PARTON DISTRIBUTIONS
FOR LHC RUN II

STEFANO FORTE
UNIVERSITÀ DI MILANO & INFN

UNIVERSITÀ DEGLI STUDI
DI MILANO

INFN
Istituto Nazionale
di Fisica Nucleare

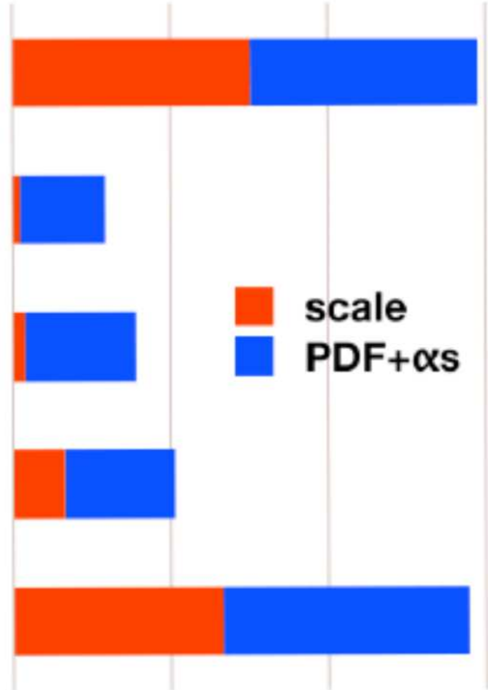
FIRST HIGGSTOOLS MEETING FREIBURG, APRIL 16, 2015

PDF RECAP SEQUENCE. . .

WHY WORRY ABOUT PDFs?

Higgs production

		σ (8 TeV)	uncertainty
NNLL QCD +NLO EW	gg→H	19.5 pb	14.7%
	VBF	1.56 pb	2.9%
NNLO QCD +NLO EW	WH	0.70 pb	3.9%
	ZH	0.39 pb	5.1%
NLO QCD	ttH	0.13 pb	14.4%

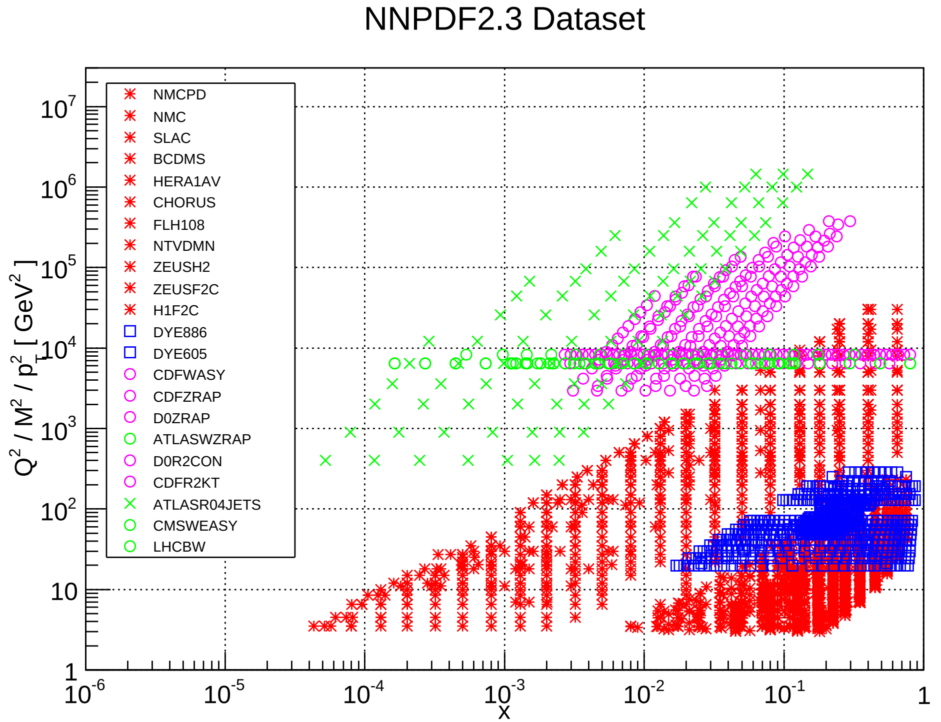


(J. Campbell, HCP2012)

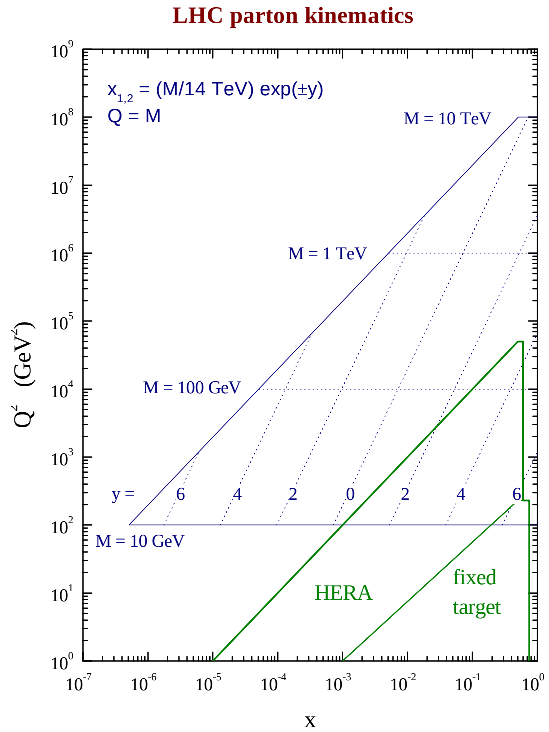
PDF uncertainty either dominant, or very large, or both
. . . and not only for the Higgs!
(W mass determination, new physics searches for heavy states,. . .)

PDFs for LHC Run I

$$\sigma_X(s, M_X^2) = \sum_{a,b} \int_{x_{\min}}^1 dx_1\, dx_2\, f_{a/h_1}(x_1) f_{b/h_2}(x_2) \hat{\sigma}_{q_a q_b \to X}\left(x_1 x_2 s, M_X^2\right)$$



LHC kinematics



	MSTW08	CT10	NNPDF2.3	HERAPDF1.5	ABM11/12	JR09
HERA DIS	✔	✔	✔	✔	✔	✔
Fixed-target DIS	✔	✔	✔	✗	✔	✔
Fixed-target DY	✔	✔	✔	✗	✔	✔
Tevatron W+Z+jets	✔	✔	✔	✗	✗	✗
LHC W+Z+jets	✗	✗	✔	✗	some, approx.	✗

RUN I PDF SETS: THE APPROACH

METHODOLOGY

- STATISTICAL TREATMENT: CTEQ, MSTW HESSIAN WITH DYNAMICAL TOLERANCE; HERAPDF, STANDARD HESSIAN+PARM. ERROR ANALYSIS; GJR, HESSIAN WITH FIXED TOLERANCE; ABKM STANDARD HESSIAN; NNPDF MONTE CARLO (ALSO STUDIED BY HERAPDF, MSTW)
- PARTON PARAMETRIZATION: CTEQ, MSTW, HERAPDF $x^\alpha(1-x)^\beta\times$ POLYNOMIALS; GJR: DITTO + VALENCELIKE ASSUMPTION; NNPDF NEURAL NETS; CHEBYSHEV POLYNOMIALS STUDIED BY HERAPDF, MSTW;
- COVARIANCE MATRIX, NORMALIZATION UNCERTAINTIES, OUTLIERS, THEORETICAL UNCERTAINTIES . . .

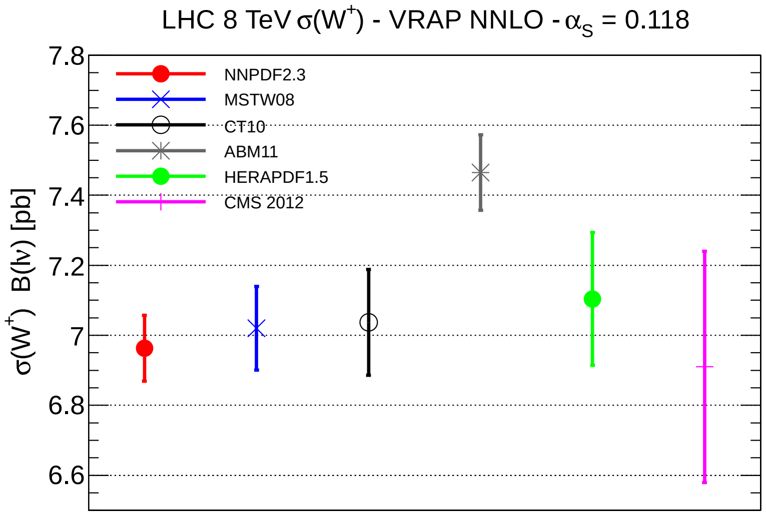
THEORY

- α_s VALUE: CTEQ: EXTERNAL PARAMETER, SEVERAL VALUES AVAILABLE; NNPDF: EXTERNAL PARAMETER, SEVERAL VALUES AVAILABLE, BEST-FIT DETERMINED; MSTW: FITTED, BUT ALSO VARIABLE AS EXT.PARAMETER; ABKM: FITTED, VARIABLE AS EXT.PARAMETER (ONLY CENTRAL VALUE); GJR: FITTED, NOT VARIABLE AS EXT. PARAMETER;
- HEAVY QUARKS: CTEQ: GM-VFN (SACOT-χ SCHEME); MSTW: GM-VFN (ACOT+TR' SCHEME); NNPDF: GM-VFN (FONLL SCHEME); ABKM: FFN ($N_f=3,4$ MATCHED WITH BMSN SCHEME); GJR: FFN ($N_f=3$)
- NUCLEAR CORRECTIONS, HIGHER TWISTS, KINEMATIC CUTS, "INITIAL SCALE", . . .

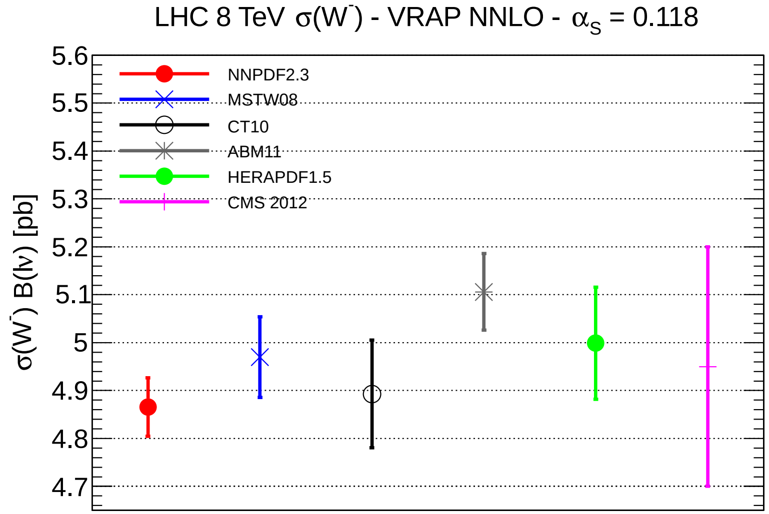
	MSTW08	CT10	NNPDF2.3	HERAPDF1.5	ABM11/12	JR09
NO. OF PDFS	7	6	7	5	6	5
STATISTICS	HESS.+DT	HESS.+DT	MC	HESS.+MODEL+PARM.	HESS.	HESS.+T
PDF PARMS.	20+8	25	259	14	24	12
HEAVY QUARKS	VFN TR	VFN ACOT	VFN FONLL	VFN TR	FFN	FFN

LHC EW standard candles

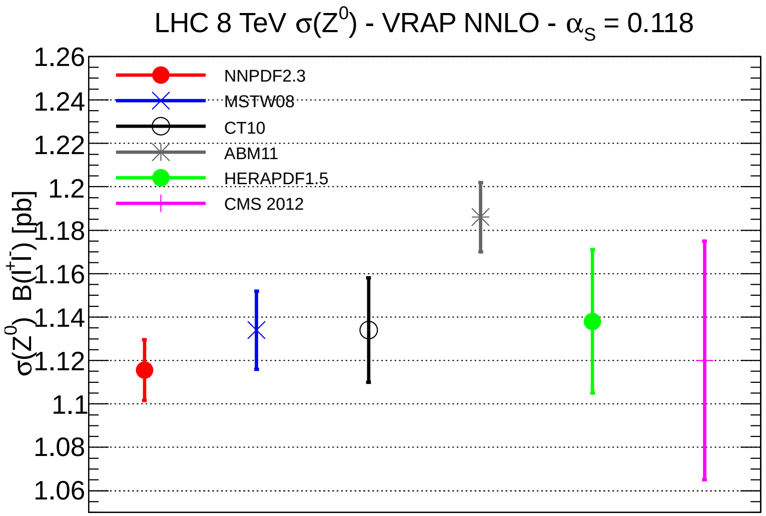
W^+



W^-



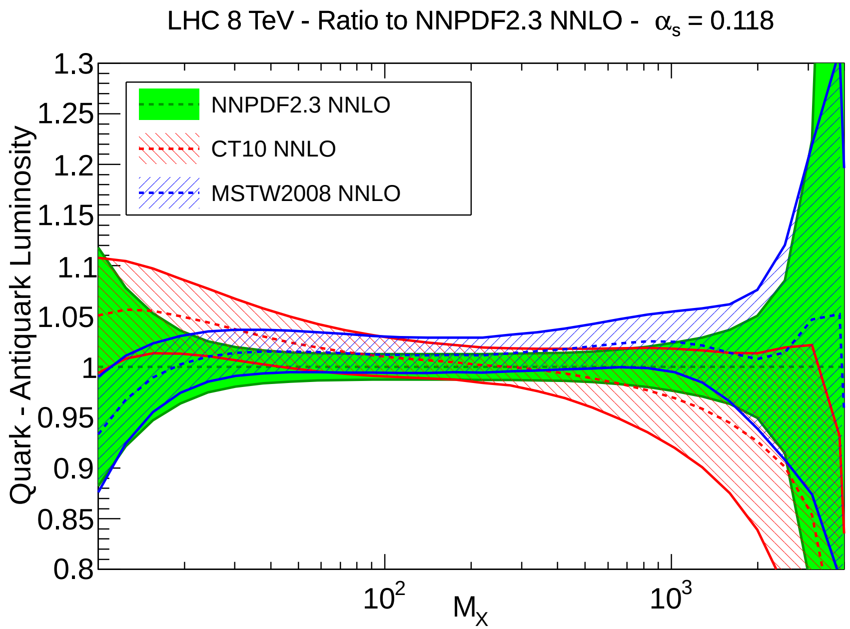
Z



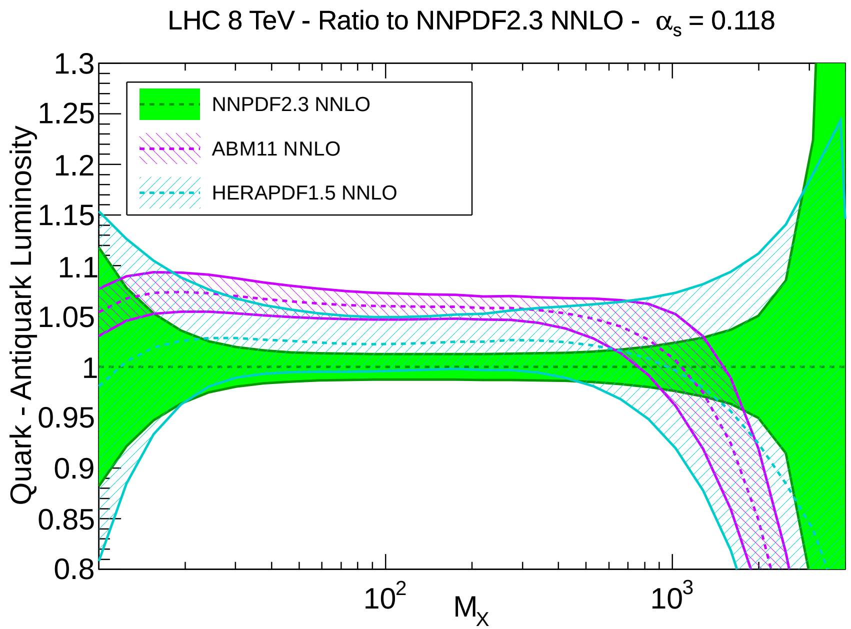
- Agreement/disagreement driven by data
- DIS-only fit (HERAPDF) safe, but large uncertainty
- Fits with smaller datasets prone to th. bias

PARTON LUMINOSITIES: QUARK SECTOR $(q\bar{q})$

GLOBAL PDF SETS
(ratio to NNPDF2.3)



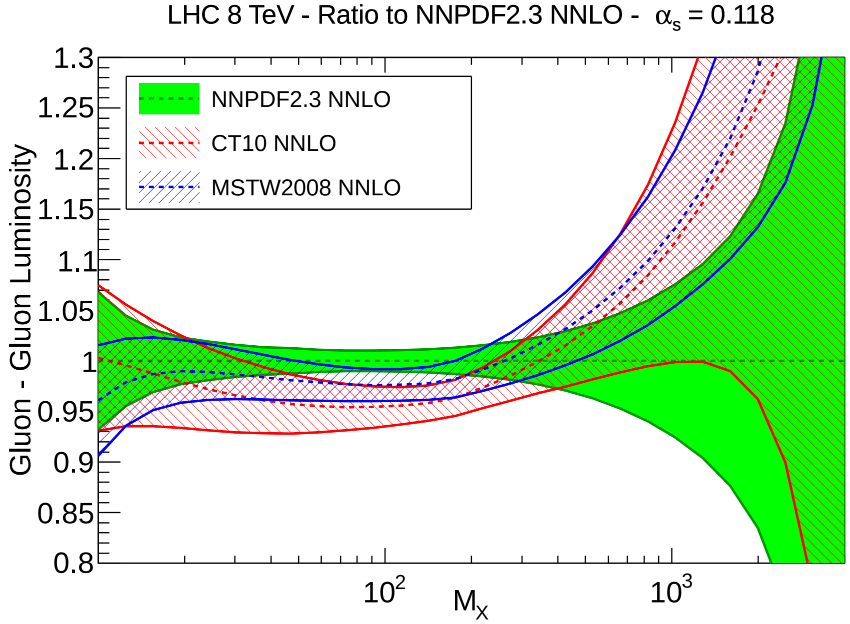
OTHER PDF SETS
(ratio to NNPDF2.3)



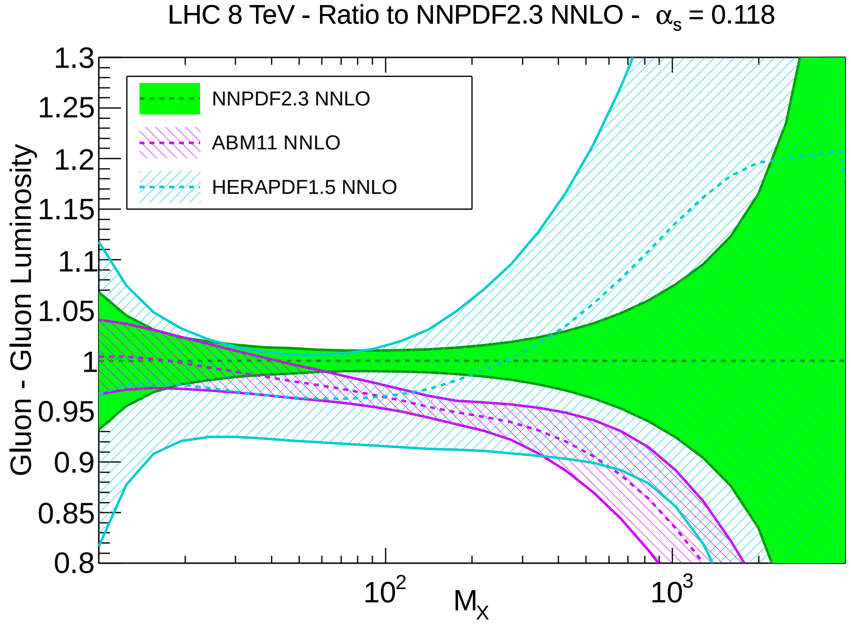
- CROSS-SECTIONS REFLECT UNDERLYING LUMINOSITIES
  FEWER DATA $\rightarrow$ LARGER UNCERTAINTIES (OR SYSTEMATIC BIAS)
- GLOBAL SETS: GOOD AGREEMENT IN THE REGION OF THE EW SCALE
- UNCERTAINTIES BLOW UP FOR LARGE-MASS FINAL STATES

PARTON LUMINOSITIES: GLUON SECTOR

GLOBAL PDF SETS
(ratio to NNPDF2.3)



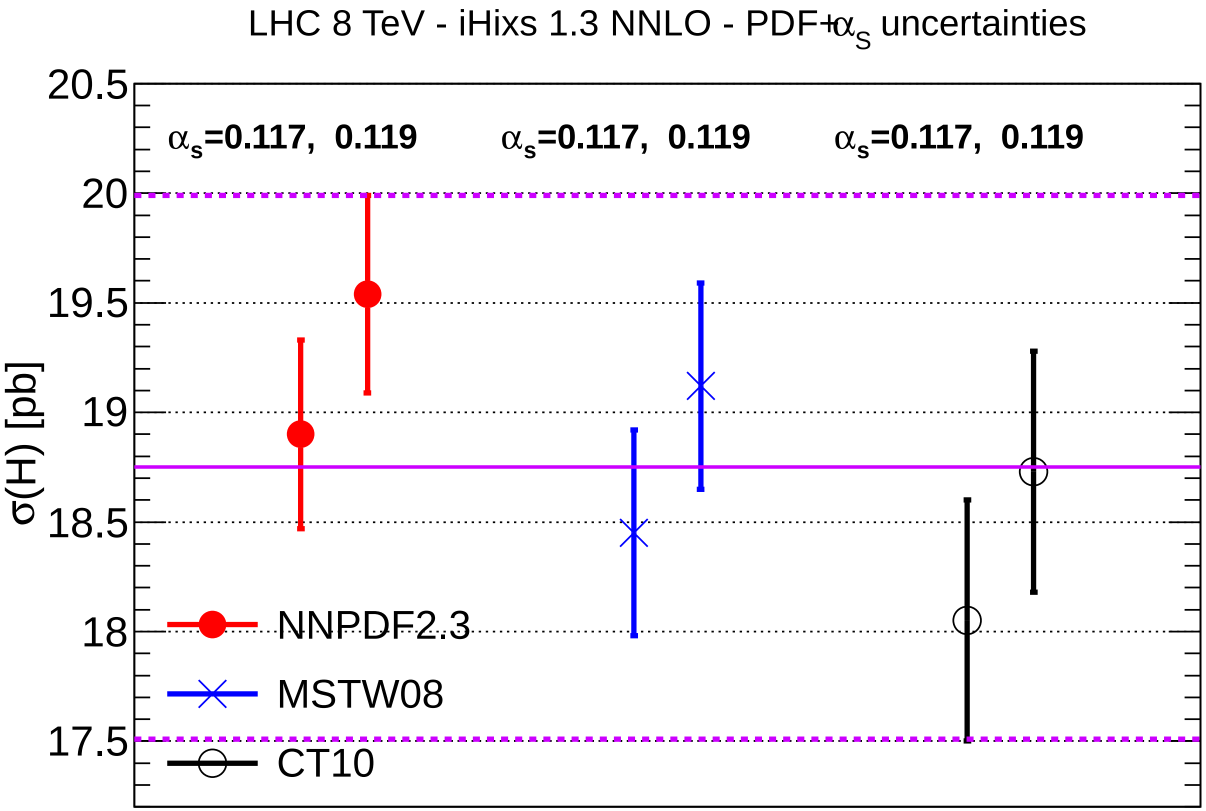
OTHER PDF SETS
(ratio to NNPDF2.3)



- FEWER DATA → LARGER UNCERTAINTIES (OR SYSTEMATIC BIAS)
- GLOBAL SETS: NOT SO GOOD AGREEMENT IN THE REGION OF THE EW SCALE
- UNCERTAINTIES BLOW UP FOR LARGE-MASS FINAL STATES

ISSUES

HIGGS IN GLUON FUSION

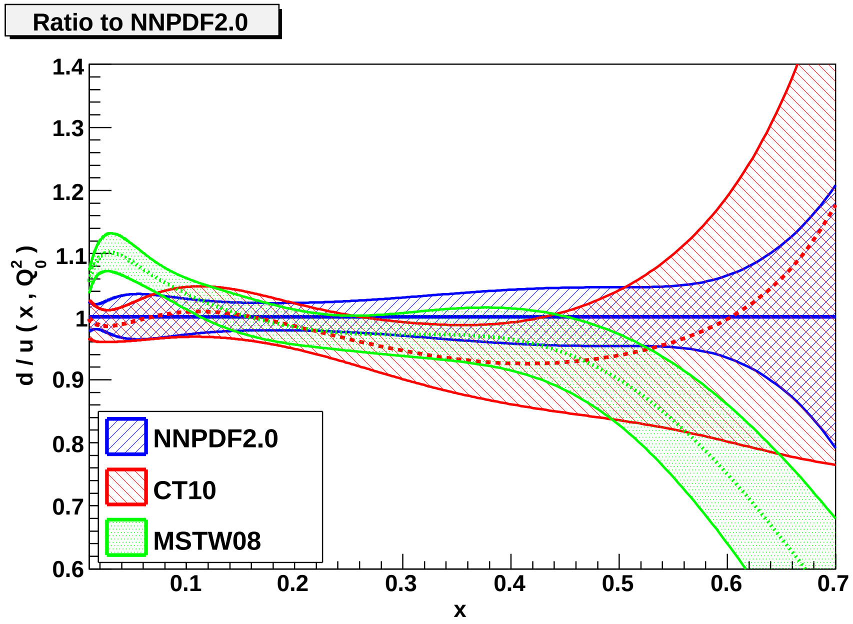


- DISCREPANCY NOT UNDERSTOOD DESPITE INTENSIVE BENCHMARKING
- CONSERVATIVE WAY OUT: TAKE THE ENVELOPE OF RESULTS (PDF4LHC PRESCRIPTION)

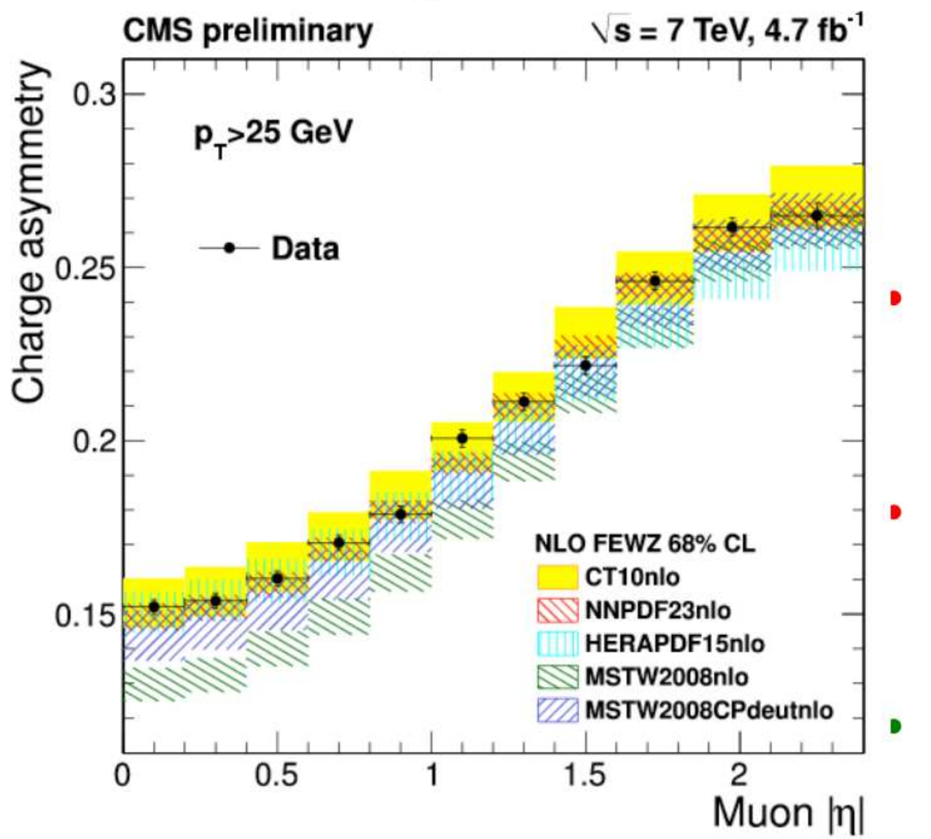
KNOWN ISSUES: METHODOLOGY

EXAMPLE: THE d/u RATIO

THE d/u RATIO



THE CMS W ASYMMETRY

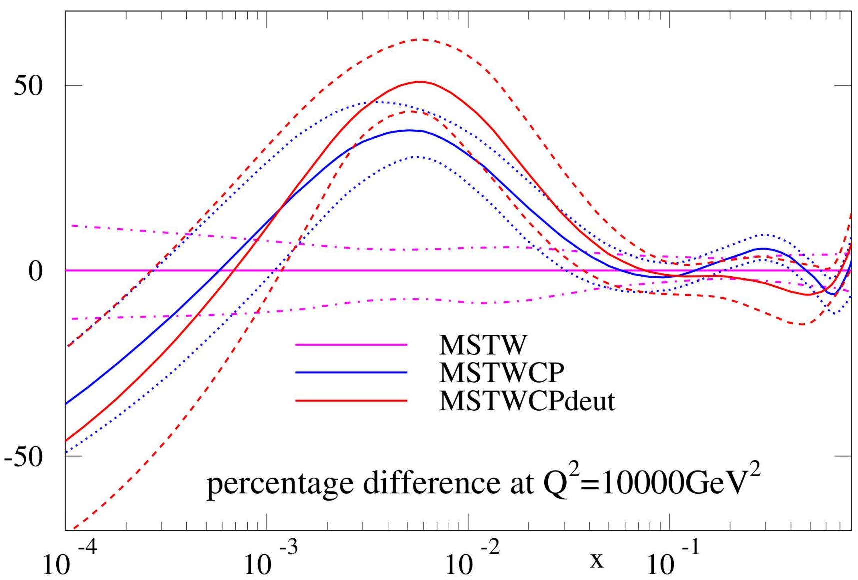


- LONG-STANDING DISCREPANCY IN THE d/u RATIO BETWEEN MSTW AND OTHER GLOBAL FITS
- RESOLVED BY W ASYMMETRY DATA
- EXPLAINED BY INSUFFICIENTLY FLEXIBLE PDF PARAMETRIZATION $\Rightarrow$ NEW MSTW08DEUT SET

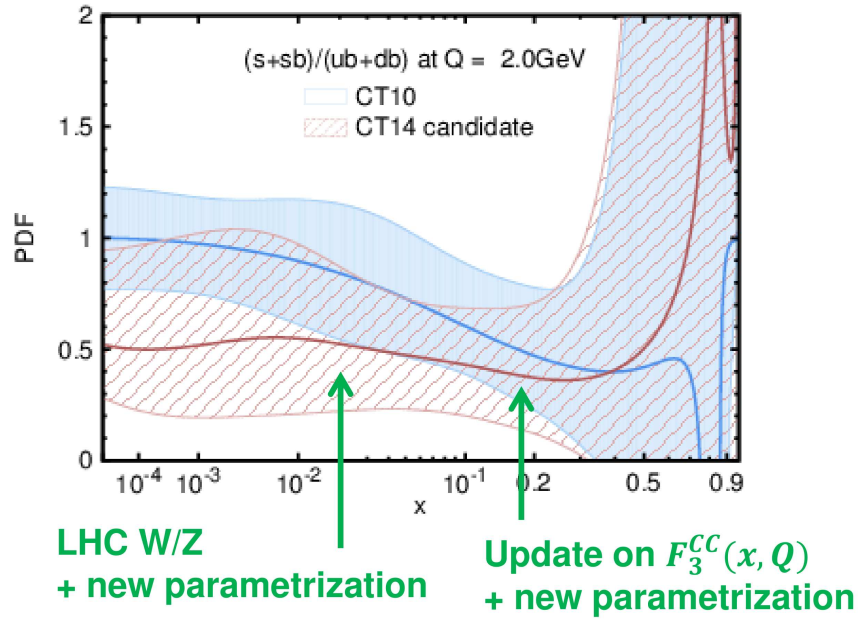
KNOWN ISSUES: METHODOLOGY

PARAMETRIZATION

MSTW08 DEUT VS DEFAULT



CTx PDF SET



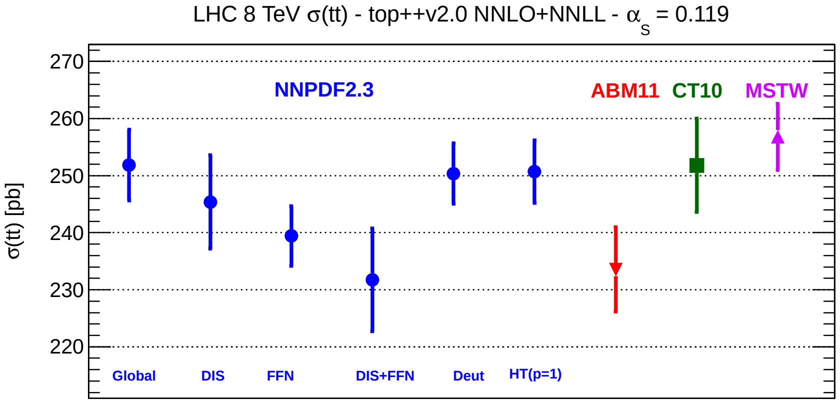
(MSTW, 2013)

(P.Nadolsky, PDF4LHC, Nov.3, 2014)

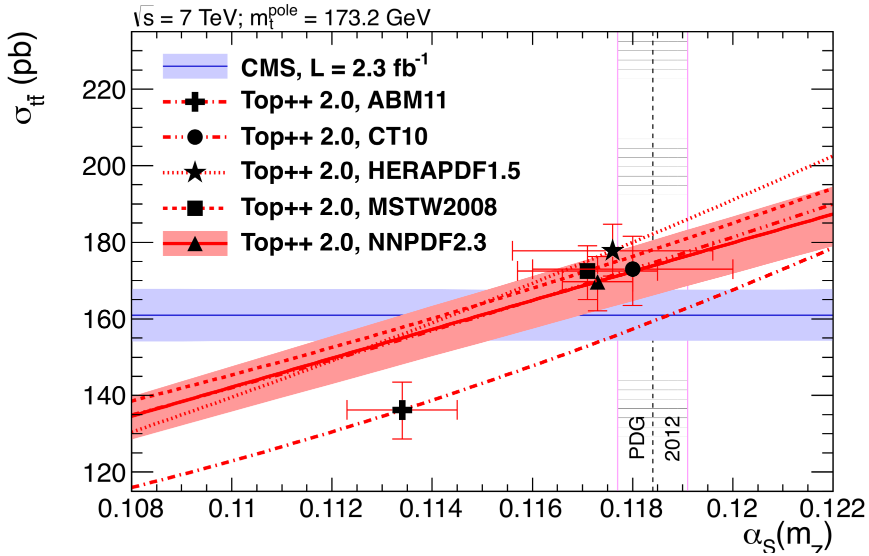
- IMPROVED PARAMETRIZATION (MORE PARAMETERS)
- $\Rightarrow$ MORE REALISTIC SHAPE
- MORE DATA $\Rightarrow$ BIGGER UNCERTAINTY

KNOWN ISSUES
THEORY AND DATA

FFN, DIS VS DEFAULT



TOP PRODUCTION



(CMS, 2014)

- FFN SCHEME NOT ADEQUATE FOR HIGH-ENERGY DATA
- DIS DATA ⇒ 'RUNAWAY' DIRECTION TOWARDS SMALL GLUON, SMALL α_s
- ⇒ CANNOT FIT TOP DATA

PARTON DISTRIBUTIONS FOR LHC RUN II

STEFANO FORTE
UNIVERSITÀ DI MILANO & INFN
FOR THE
NNPDF COLLABORATION: R. D. BALL, V. BERTONE, S. CARRAZZA,
C. D. DEANS, L. DEL DEBBIO, A. GUFFANTI, M. P. HARTLAND,
Z. KASSABOV, P. GROTH-MERRILD, J. I. LATORRE, J. ROJO AND M. UBIALI

UNIVERSITÀ DEGLI STUDI DI MILANO

INFN Istituto Nazionale di Fisica Nucleare

FIRST HIGGSTOOLS MEETING

FREIBURG, APRIL 16, 2015

SUMMARY

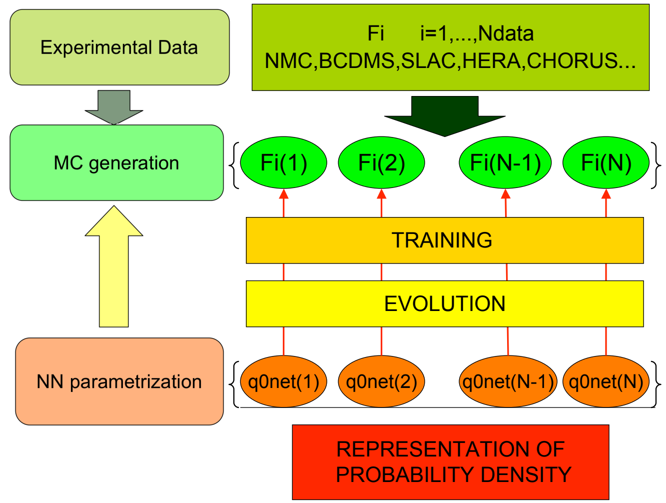
THE NNPDF METHODOLOGY ADDRESSES IN A QUANTITATIVE & SYSTEMATIC WAY ISSUES OF

- METHODOLOGY: CLOSURE TESTS
- THEORY CONSISTENCY: BENCHMARKS
  CASE STUDY I: JETS AND NNLO
- DATA: COMPATIBILITY & IMPACT ANALYSIS TOOLS
  CASE STUDY II: THE STRANGE PDF
  CASE STUDY III: CONSERVATIVE PARTONS
  CASE STUDY IV: ASSESSING DATA IMPACT

THE NNPDF METHODOLOGY
A BRIEF SUMMARY

MONTE CARLO APPROACH

- PSEUDODATA REPLICAS ARE GENERATED USING FULL INFO ON CORRELATED SYSTEMATICS
  $\Rightarrow$ MONTECARLO REPRESENTATION OF COVARIANCE MATRIX
- A PDF REPLICA SET IS FITTED TO EACH PSEUDO-DATA REPLICA
- THE SET OF PDF REPLICAS CAN BE USED TO COMPUTE ANY SET OF OBSERVABLES
  CENTRA VALUES (MEANS), UNCERTAINTIES (VARIANCES), CORRELATIONS (COVARIANCES), CONFIDENCE LEVELS. . .



NEURAL NETWORK PARAMETRIZATION

- EACH PDF IS PARAMETRIZED BY A NEURAL NETWORK
  37 PARAMETERS WITH DEFAULT $2-5-3-1$ ARCHITECTURE
- MINIMIZATION IS PERFORMED USING GENETIC ALGORITHMS
- BEST-FIT IS NOT MINIMUM OF χ^2 (WOULD BE FITTING NOISE) $\Rightarrow$ CROSS-VALIDATION
  DIVIDE DATA IN TWO SETS, FIT TRAINING SET, MONITOR VALIDATION SET, STOP WHEN VALIDATION χ^2 STARTS RISING

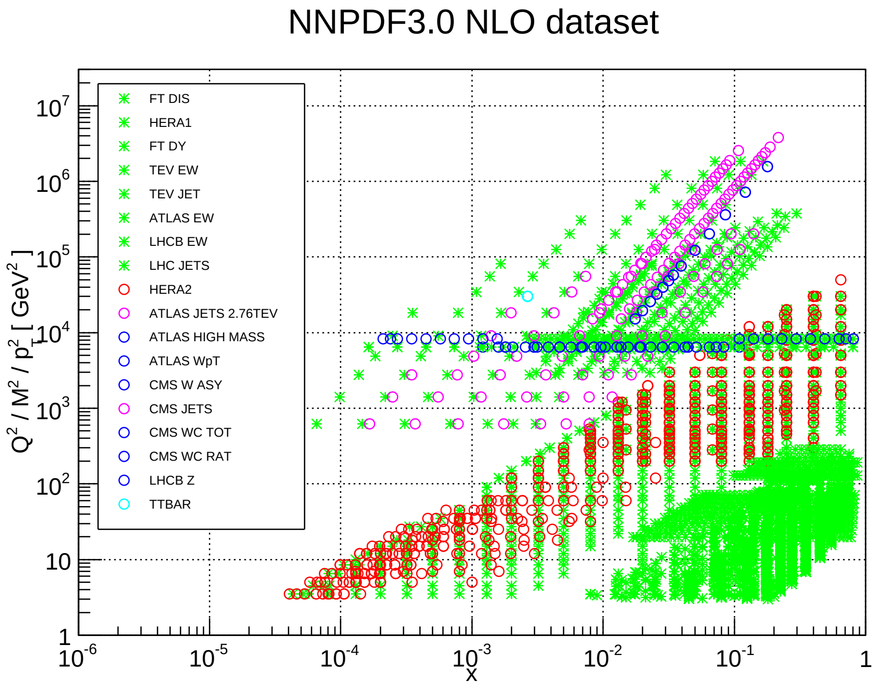
THE NNPDF3.0 DATASET

A BRIEF SUMMARY

NEW IN NNPDF3.0

- COMBINED HERA CHARM PRODUCTION (55 D.P)
- HERA II ZEUS+H1 STRUCTURE FUNCTIONS (778 D.P.)
- ATLAS 2.76TEV JETS (59 D.P.)
- ATLAS HIGH-MASS DRELL-YAN (5 D.P.)
- ATLAS W p_T (9 D.P.)
- CMS W MUON ASYMMETRY (11 D.P.)
- CMS DOUBLE-DIFFERENTIAL DRELL-YAN (110 D.P.)
- CMS JETS (133 D.P.)
- CMS $W+c$ (10 D.P.)
- LHCB Z RAPIDITY (9 D.P.)
- ATLAS & CMS TOP TOTAL XSECT (3+3 D.P.)

- TOTAL DATASET:
  4276/4078 (NLO/NNLO)



NNPDF3.0 THEORY AND METHODOLOGY

A BRIEF SUMMARY

NEW IN NNPDF3.0

- UNCERTAINTIES: ALL CAN BE ADDITIVE OR MULTIPLICATIVE
- EW CORRECTIONS INCLUDED
- FULL MIGRATION OF CODE TO C++
- PDF PARAMETRIZATION SCALE ($Q^2 = 1$ GeV) & BASIS (EVOLUTION EIGENSTATES)
- PREPROCESSING: BASIS INDEPENDENT, SELF-CONSISTENTLY DETERMINED
- GENETIC ALGORITHM: NODAL MUTATION
- IMPLEMENTATION OF CROSS-VALIDATION: LOOKBACK STOPPING

METHODOLOGY:
CLOSURE TESTING

CLOSURE TESTS:

THE BASIC IDEA

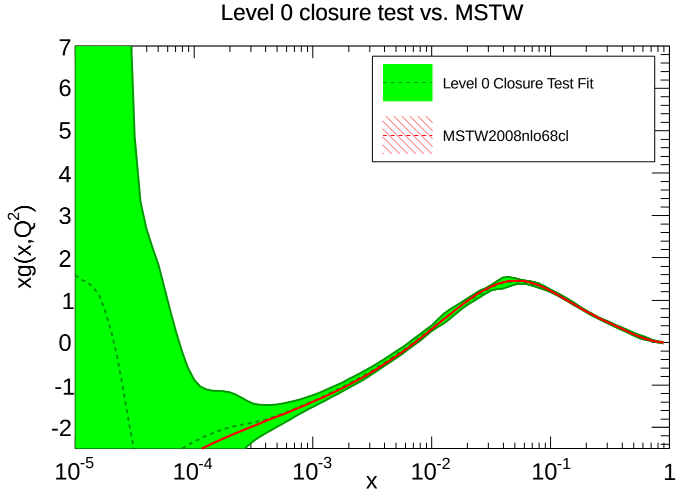
- ASSUME PDFs KNOWN: GENERATE FAKE EXPERIMENTAL DATA
- CAN DECIDE DATA UNCERTAINTY (ZERO, OR AS IN REAL DATA, OR . . .)
- FIT PDFs TO FAKE DATA
- CHECK WHETHER FIT REPRODUCES UNDERLYING "TRUTH":
  - CHECK WHETHER TRUE VALUE GAUSSIANLY DISTRIBUTED ABOUT FIT
  - CHECK WHETHER UNCERTAINTIES FAITHFUL
  - CHECK STABILITY

    (INDEP. OF METHODOLOGICAL DETAILS)

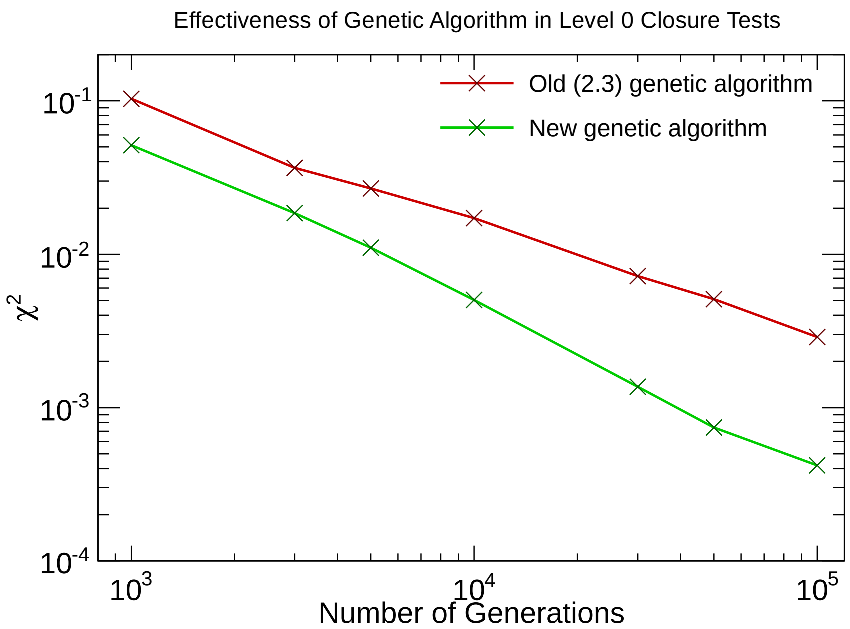
LEVEL-0 CLOSURE TESTS

- ASSUME VANISHING EXPERIMENTAL UNCERTAINTY
- MUST BE ABLE TO GET $\chi^2 = 0$
- UNCERTAINTY AT DATA POINTS TENDS TO ZERO (NOT NECESSARILY ON PDF!)
  DEFINE $\phi \equiv \sqrt{\langle \chi^2_{rep} \rangle - \chi^2}$, EQUALS FIT UNCERTAINTY/DATA UNCERTAINTY; CHECK $\phi \to 0$
- BEST FIT ON TOP OF "TRUTH" IN DATA REGION

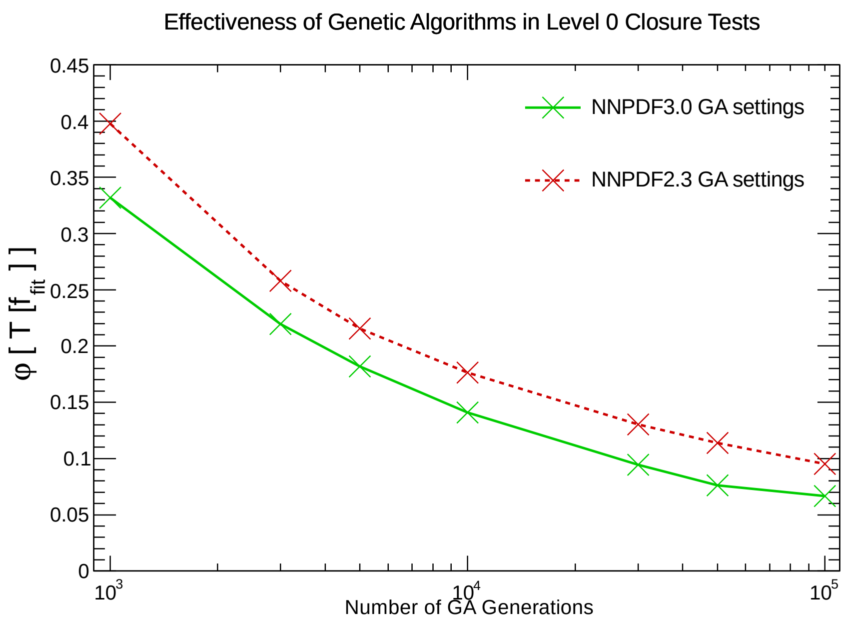
THE GLUON



χ^2 VS TRAINING LENGTH



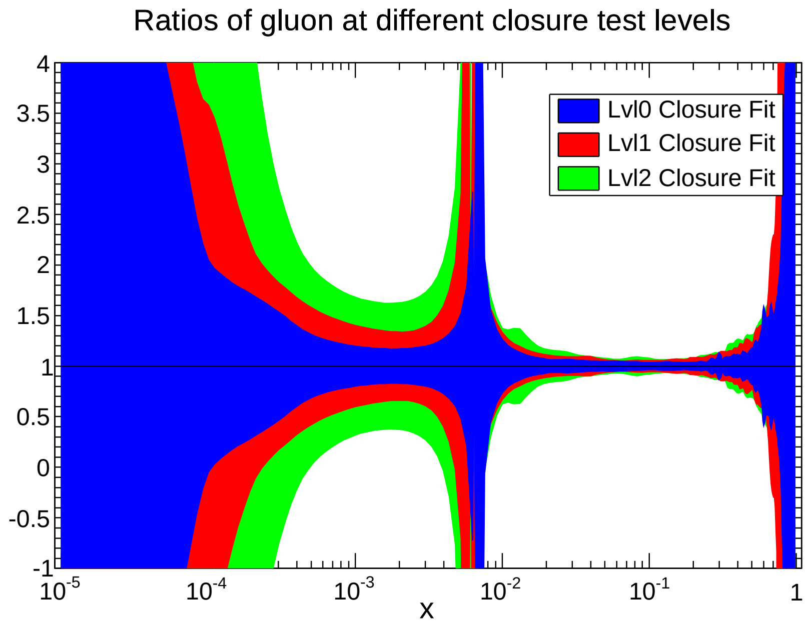
FRACTIONAL UNCERTAINTY VS TRAINING LENGTH

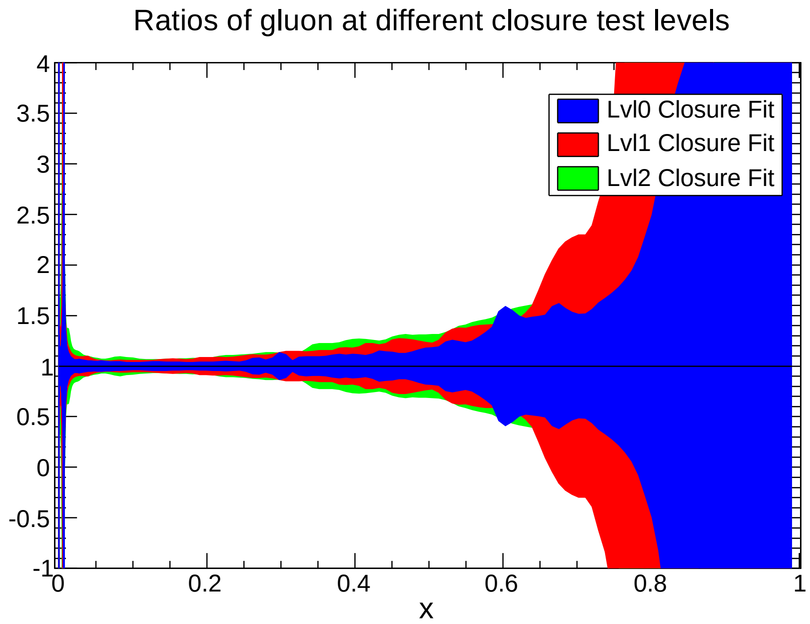


LEVEL-0, LEVEL-1 AND LEVEL-2

- LEVEL 0: FAKE DATA GENERATED WITH NO UNCERTAINTY
  $\rightarrow$ INTERPOLATION AND EXTRAPOLATION UNCERTAINTY
- LEVEL 1-2: FAKE DATA GENERATED WITH SAME UNCERTAINTY AS REAL DATA (INCLUDING CORRELATIONS)
- LEVEL 1: NO PSEUDODATA REPLICAS:
  $\Rightarrow$ REPLICAS FITTED TO SAME DATA OVER AND OVER AGAIN
  $\rightarrow$ FUNCTIONAL UNCERTAINTY DUE TO INFINITY OF EQUIVALENT MINIMA
- LEVEL 2: STANDARD NNPDF METHODOLOGY
  $\Rightarrow$ REPLICAS FITTED TO PSEUDODATA REPLICAS
  $\rightarrow$ DATA UNCERTAINTY
- THREE SOURCES OF UNCERTAINTY COMPARABLE IN DATA REGION

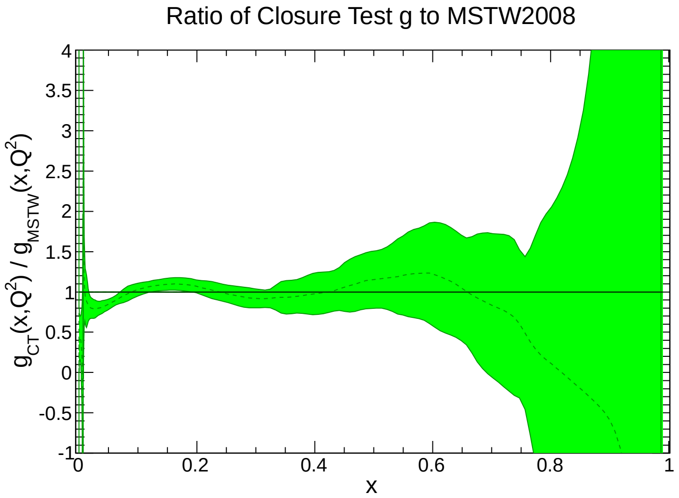
THE GLUON: LEVEL 0, LEVEL 1 AND LEVEL 2





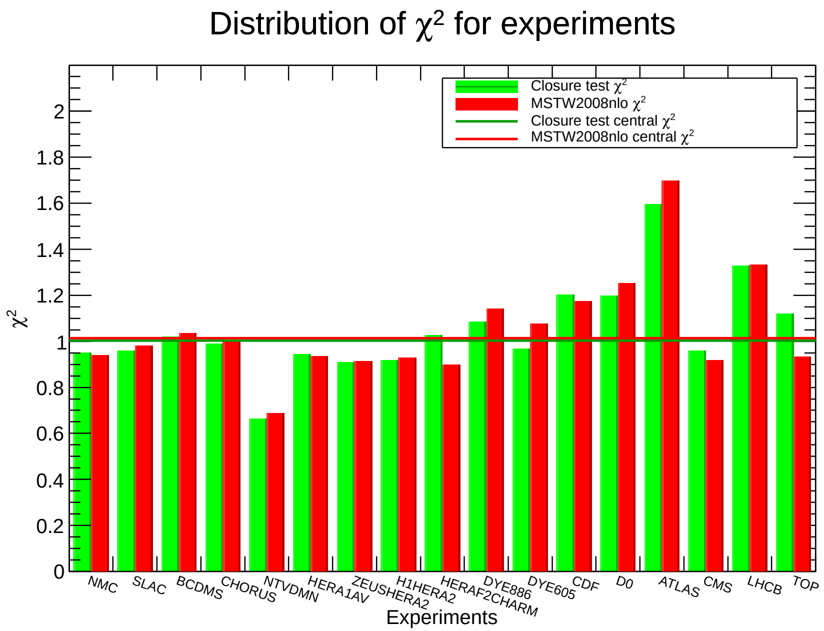
LEVEL-2: CENTRAL VALUES AND UNCERTAINTIES

THE GLUON: FITTED/"TRUE"

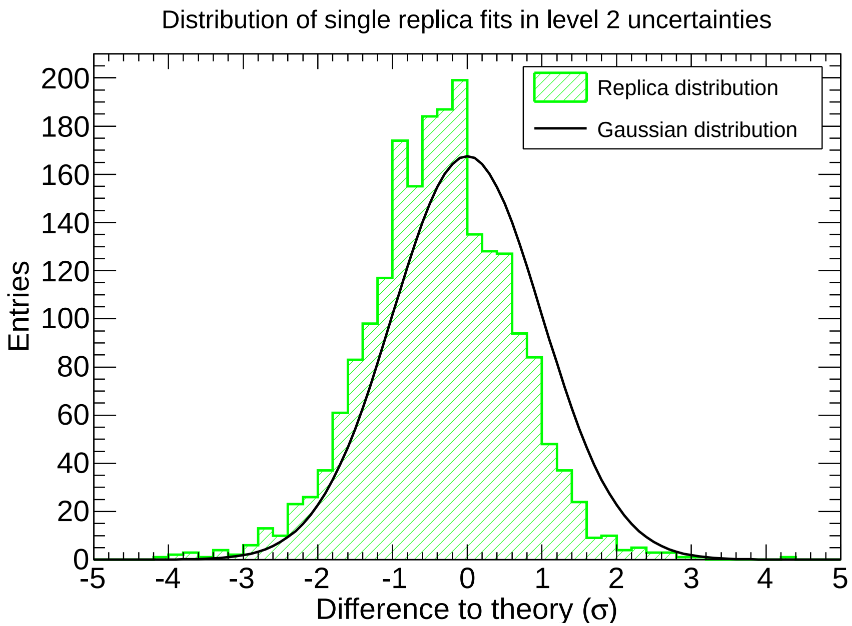


- CENTRAL VALUES: COMPARE FITTED VS. "TRUE" χ^2 BOTH FOR INDIVIDUAL EXPERIMENTS & TOTAL DATASET FOR TOTAL $\Delta\chi^2 = 0.001 \pm 0.003$
- UNCERTAINTIES: DISTRIBUTION OF DEVIATIONS BETWEEN FITTED AND "TRUE" PDFs SAMPLED AT 20 POINTS BETWEEN 10^{-5} AND 1 FIND 0.699% FOR ONE-SIGMA, 0.948% FOR TWO-SIGMA C.L.

LEVEL-2 FITTED χ^2 VS "TRUE"



NORM. DISTRIBUTION OF DEVIATIONS



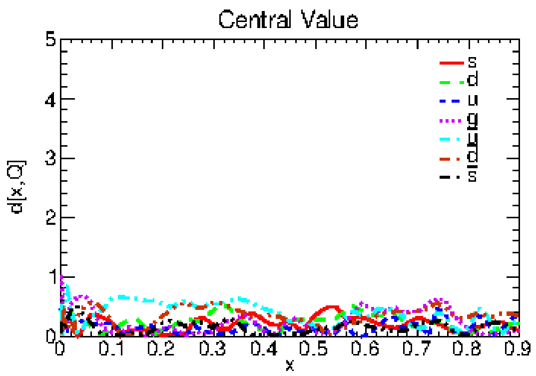
LEVEL-2 STABILITY TESTS

- CHANGE UNDERLYING PDF SET (CT10, NNPDF2.3)
- INCREASE MAXIMUM GA TRAINING LENGTH TO 80K
  TESTS EFFICIENCY OF CROSS-VALIDATION
- INCREASE NN ARCHITECTURE TO 2-20-15-1
  NUMBER OF FREE PARAMETRES INCREASE BY MORE THAT $10\times$
- CHANGE PDF PARAMETRIZATION BASIS
  OLD: ISOTRIPLET, $\bar{u}-\bar{d}$, $s+\bar{s}$, $s-\bar{s}$;
  NEW: ISOTRIPLET, SU(3)-OCTET, BOTH TOTAL $(q+\bar{q})$ & VALENCE $(q-\bar{q})$

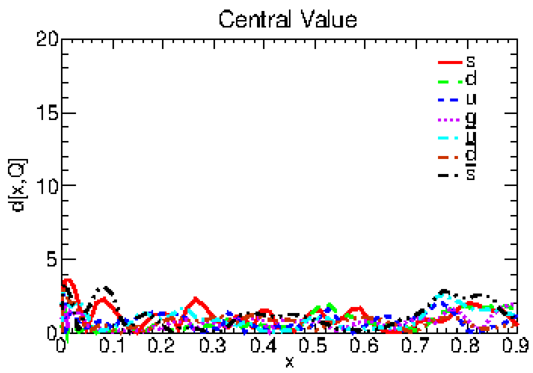
STATISTICAL EQUIVALENCE!

DISTANCES BETWEEN REF. AND NEW FIT:
difference in unites of standard deviation of the mean

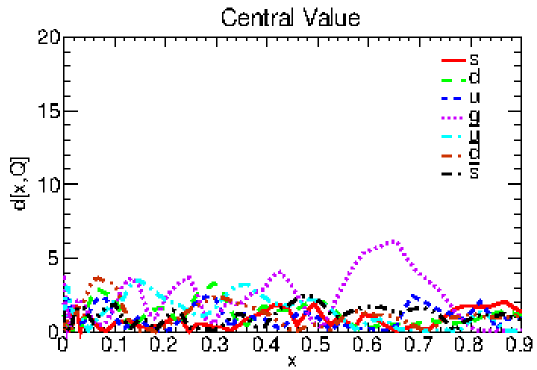
30K GENS VS 80K GENS



2.3 BASIS VS 3.0 BASIS



300 VS 37 PARMS



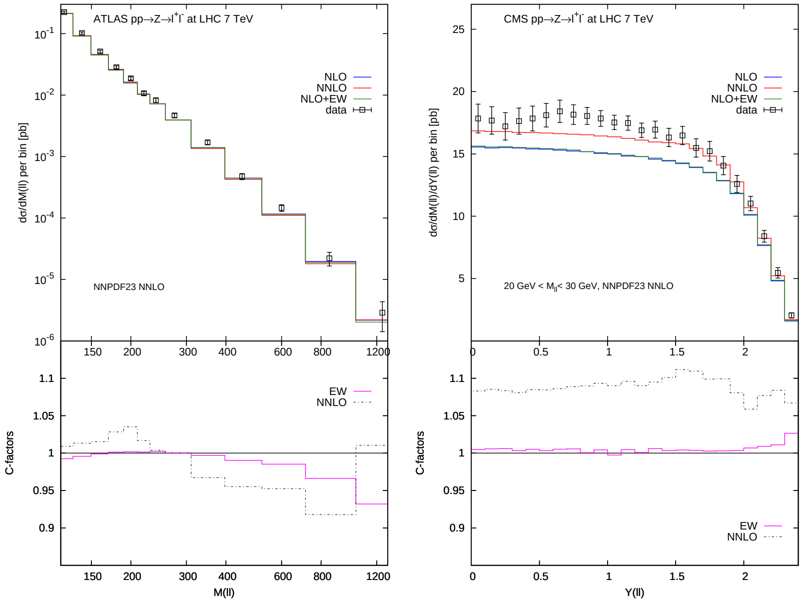
THEORY BENCHMARKING

CORRECTIONS AND CUTS

- PDFs provided at LO, NLO, NNLO in α_s
- NNLO QCD corrns benchmarked between codes: for DY MCFM vs FEWZ at NLO; FEWZ & DYNNLO for NNLO;
- size of NNLO monitored: in NLO fit, discard if $\gg$ than expt uncertainties lowest Q^2 bin of CMS 2dDY discarded
- virtual pure EW corrections included; FSR-corrected data used; QED corrections monitored highest Q^2 bin of CMS 2dDY discarded

NLO, NNLO & EW corrns
ATLAS and CMS

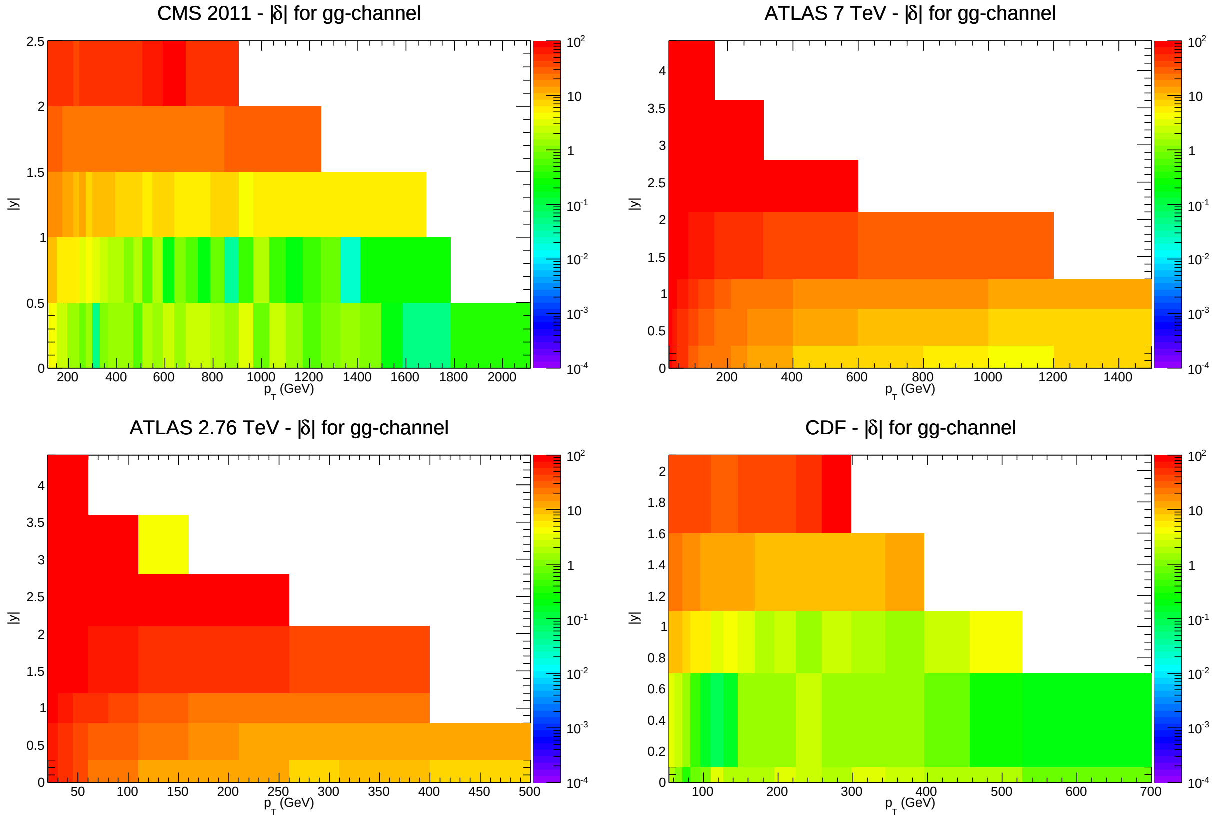
high mass DY



JET DATA AT NLO AND NNLO

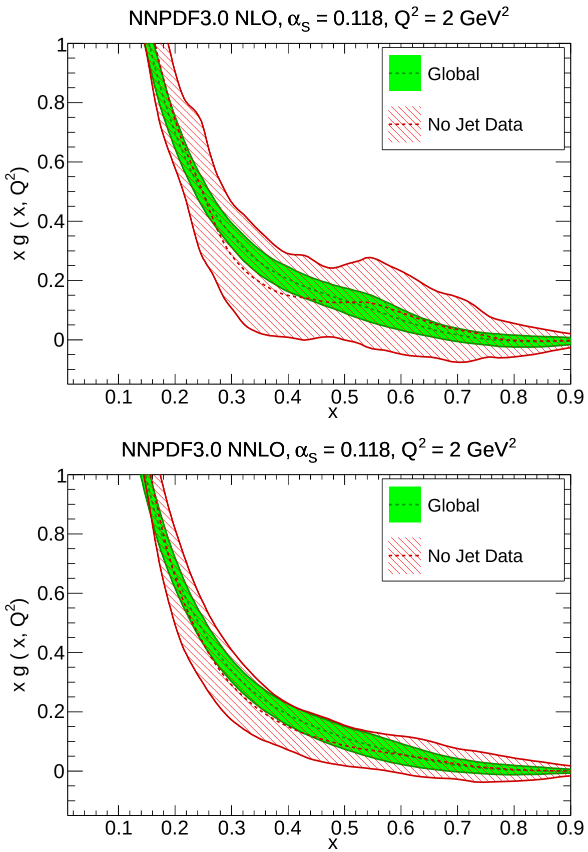
PERCENTAGE ACCURACY OF THRESHOLD APP.

IN GLUON CHANNEL



IMPACT OF THE JET DATA:

NLO&NNLO



- EXACT NNLO CORRECTIONS NOW KNOWN IN GLUON CHANNEL (GEHRMANN, GLOVER, DE RIDDER, PIRES, 2014)
- THRESHOLD APPROX AVAILABLE IN ALL CHANNELS (DE FLORIAN ET AL, 2014)
- ONLY ACCURATE AT HIGH p_T, CENTRAL RAPIDITY (CARRAZZA & PIRES, 2014)
- AT NNLO RETAIN ONLY BINS SUCH THAT ACCURACY BETTER THAN 10% CORRECTION THERE SMALLER THAN 15% $\Rightarrow$ MOSTLY CMS DATA SURVIVE
- IMPACT OF JET DATA STILL SIZABLE

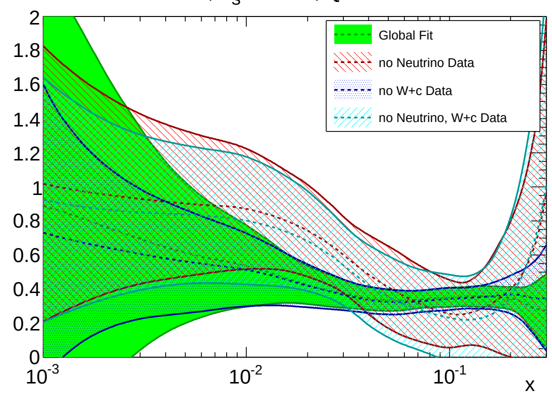
DATA CONSISTENCY & IMPACT ANALYSIS

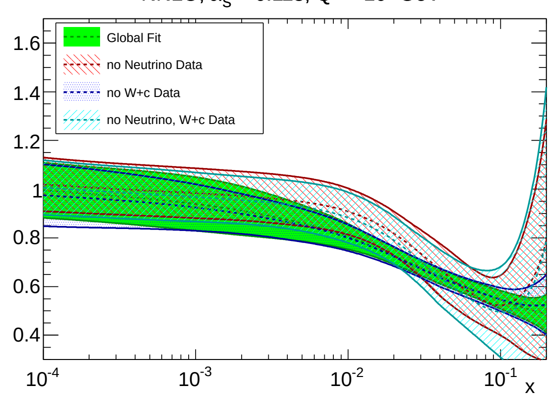
NUCLEON STRANGENESS

	$\chi^2_{\rm exp}$			
	GLOBAL	NO NEUTRINO	NO W+c	NO NEUTRINO/W+c
CHORUS	1.13	3.87	1.09	3.45
NUTEV	0.62	4.31	0.66	6.45
ATLAS W, Z 2010	1.21	1.05	1.24	1.08
CMS W+c 2011	0.86	0.50	0.90	0.61

STRANGENESS RATIO

AT LOW & HIGH SCALE





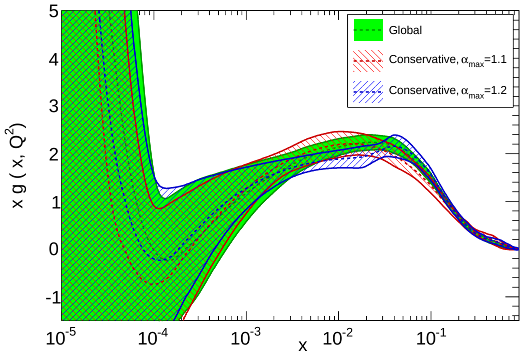
- IN GLOBAL FITS, STRANGENESS DETERMINED IN EQUAL PROPORTION BY NEUTRINO & W/Z PRODUCTION DATA
- NOW SPECIFIC INFORMATION FROM $W + c$
  BUT ONLY 10 CMS DATAPOINTS;
  ATLAS NOT YET INCLUDED BECAUSE DATA ARE HADRON-LEVEL
- DEFINE STRANGENESS RATIO $r(x, Q^2) = \frac{s+\bar{s}}{\bar{u}+\bar{d}}$
- ALL DETERMINATIONS CONSISTENT WITHIN UNCERTAINTIES
- AT PRESENT NEUTRINO DATA CONTROL BOTH CENTRAL VALUE & UNCERTAINTY
- W/Z PRODUCTION DATA FAVOR LARGER r BUT WITH MUCH LARGER UNCERTAINTY
- CMS $W + c$ DATA HAVE VERY LITTLE IMPACT, ON GLOBAL FIT OR ON FIT W/O NEUTRINO DATA (PERHAPS FAVOR HIGHER r)

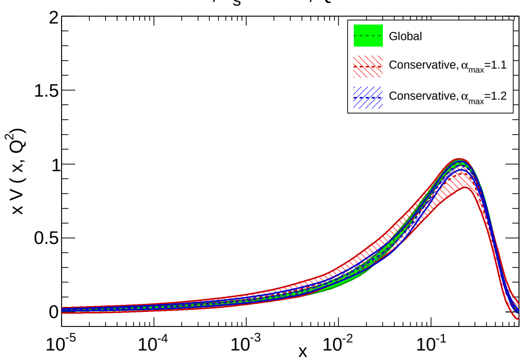
CONSERVATIVE PARTONS

- RESCALE ALL UNCERTAINTIES $\sigma \to \alpha\sigma$: $\chi^2 \to \chi^2/\alpha^2$ FOR A GIVEN EXPERIMENT
- DETERMINE PROBABILITY DISTRIBUTION $P(\alpha)$ (USING BAYES)
- DISCARD ALL EXPERIMENT FOR WHICH $P(\alpha)$ PEAKS WELL ABOVE ONE
  TWO OUT OF MEDIAN, MODE, MEAN, GREATER THAN $\alpha_{threshold}$
- $\chi^2 = 1.29$ FOR NNLO GOBAL, BECOMES $\chi^2 = 1.16$ FOR $\alpha_{threshold} = 1.3$, $\chi^2 = 1.10$ FOR $\alpha_{threshold} = 1.2$, $\chi^2 = 1.01$ FOR $\alpha_{threshold} = 1.1$, BUT CONSIDERABLE DETERIORATION OF UNCERTAINTIES

CONSERV. VS. DEFAULT

GLUON AND VALENCE





α PEAK FOR EXPERIMENTS DISCARDED IN CONS. FIT

WHEN INCLUDED OR EXCLUDED FROM FIT

	NNLO global fit			NNLO cons. fit $\alpha_{max} = 1.1$		
Experiment	mean	mode	median	mean	mode	median
NMC $\sigma^{NC,p}$	1.27	1.26	1.27	1.50	1.45	1.48
SLAC	1.13	1.09	1.12	1.61	1.37	1.48
BCDMS	1.20	1.19	1.20	2.02	1.86	1.92
CHORUS	1.10	1.09	1.09	2.55	1.69	2.32
ZEUS HERA-II	1.25	1.24	1.25	1.38	1.33	1.36
H1 HERA-II	1.35	1.34	1.34	1.51	1.47	1.49
HERA σ^c_{NC}	1.14	1.11	1.13	1.13	1.09	1.12
E886 p	1.15	1.14	1.15	2.18	1.62	2.03
CDF Z rapidity	1.39	1.32	1.36	1.56	1.40	1.50
CDF Run-II k_t jets	1.15	1.12	1.14	1.25	1.18	1.22
ATLAS W, Z 2010	1.17	1.12	1.15	1.38	1.25	1.32
ATLAS high-mass DY	1.00	1.34	1.63	1.63	1.19	1.45
CMS W muon asy	1.60	1.40	1.53	2.90	2.48	2.81
CMS W+c total	1.50	1.09	1.33	1.85	1.37	1.67
CMS W+c ratio	2.00	1.39	1.69	2.12	1.58	1.94
CMS 2D DY 2011	1.28	1.27	1.28	1.29	1.28	1.29
LHCb	1.20	1.12	1.17	1.58	1.22	1.48

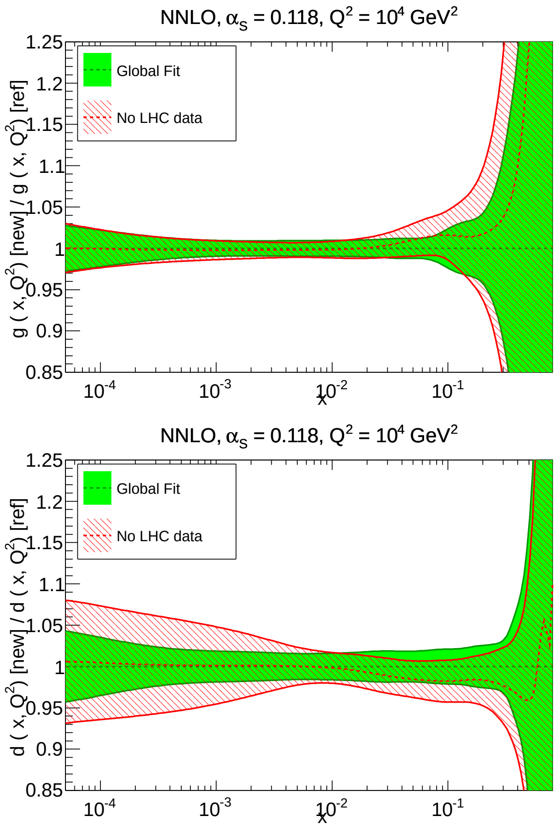
THE IMPACT OF LHC (AND HERA) DATA

- OVERALL MEASURE OF IMPACT:
  $\phi \Rightarrow$ FIT UNCERTAINTY/DATA UNCERTAINTY
- HERA-II IMPACT SIZABLE
- IMPACT OF LHC DATA MODERATE BUT VISIBLE
- IMPACT OF CMS OR ATLAS COMPARABLE TO (MODERATE) IMPACT OF NON-LHC, NON-HERA DATA

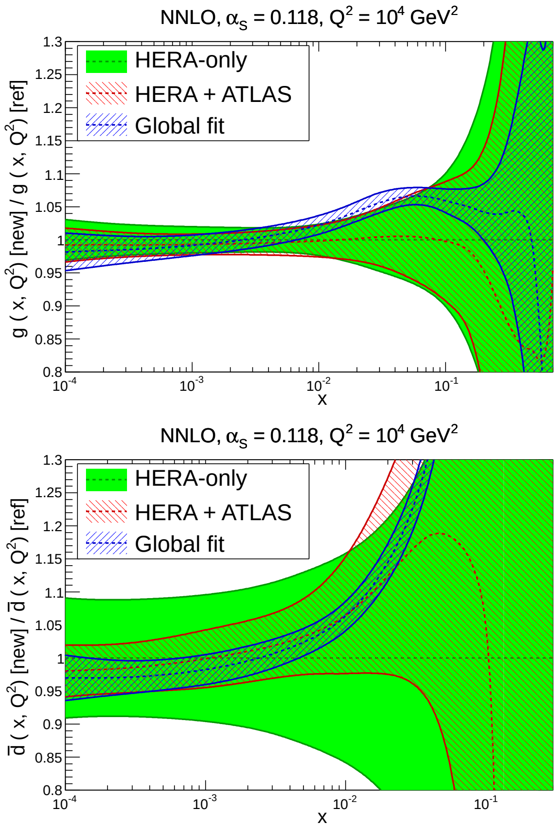
FRACTIONAL UNCERTAINTY

Dataset	φ_{χ^2} NLO	φ_{χ^2} NNLO
Global	0.291	0.302
HERA-I	0.453	0.439
HERA all	0.375	0.343
HERA+ATLAS	0.391	0.318
HERA+CMS	0.315	0.345
Conservative	0.422	0.478
no LHC	0.312	0.316

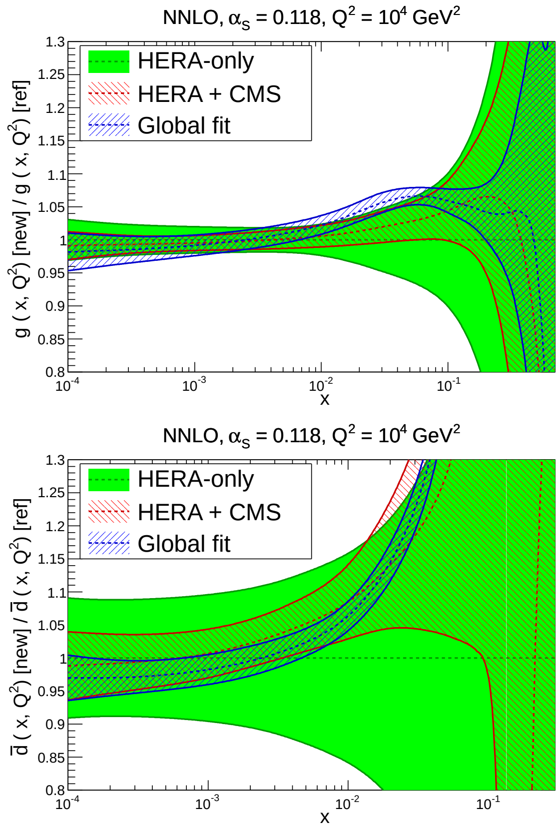
GLOBAL VS NO LHC: g& d



GLOBAL VS HERA+ATLAS: g& $\bar{d}$



GLOBAL VS HERA+CMS: g& $\bar{d}$



OUTLOOK

GLOBAL FITS: PROGRESS

- MMHT (SUCCESSOR OF MSTW08) PUBLISHED DECEMBER 2014
- CT14 PRESENTED IN PRELIMINARY FORM 2014-2015
- FOR ALL, PROGRESS IN METHODOLOGY & DATASET

METHODOLOGY

	NNPDF3.0	MMHT14	CT14
NO. OF FITTED PDFS	7	7	6
PARAMETRIZATION	NEURAL NETS	$x^a(1-x)^b\times$ CHEBYSCHEV	$x^a(1-x)^b\times$BERNSTEIN
FREE PARAMETERS	259	37	30-35
UNCERTAINTIES	REPLICAS	HESSIAN	HESSIAN
TOLERANCE	NONE	DYNAMICAL	DYNAMICAL
CLOSURE TEST	✔	✗	✗
REWEIGHTING	REPLICAS	EIGENVECTORS	EIGENVECTORS

- MMHT, CT10 LARGER # OF PARMS., ORTHOGONAL POLYNOMIALS
- NNPDF CLOSURE TEST

DATASET

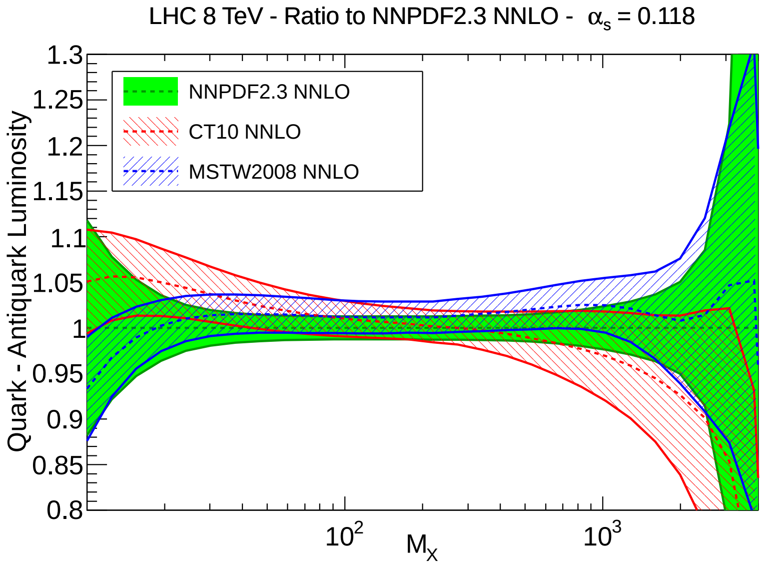
	NNPDF3.0	MMHT14	CT14(PREL)
SLAC P,D DIS	✔	✔	✗
BCDMS P,D DIS	✔	✔	✔
NMC P,D DIS	✔	✔	✔
E665 P,D DIS	✗	✔	✗
CDHSW NU-DIS	✗	✗	✔
CCFR NU-DIS	✗	✔	✔
CHORUS NU-DIS	✔	✔	✗
CCFR DIMUON	✗	✔	✔
NUTEV DIMUON	✔	✔	✔
HERA I NC,CC	✔	✔	✔
HERA I CHARM	✔	✔	✔
H1,ZEUS JETS	✗	✔	✗
H1 HERA II	✔	✗	✗
ZEUS HERA II	✔	✗	✗
E605 & E866 FT DY	✔	✔	✔
CDF & D0 W ASYM	✗	✔	✔
CDF & D0 Z RAP	✔	✔	✔
CDF RUN-II JETS	✔	✔	✔
D0 RUN-II JETS	✗	✔	✔
ATLAS HIGH-MASS DY	✔	✔	✔
CMS 2D DY	✔	✔	✗
ATLAS W,Z RAP	✔	✔	✔
ATLAS W PT	✔	✔	✗
CMS W ASY	✔	✔	✔
CMS W +C	✔	✗	✗
LHCB W,Z RAP	✔	✔	✔
ATLAS JETS	✔	✔	✔
CMS JETS	✔	✔	✔
TTBAR TOT XSEC	✔	✔	✗
TOTAL NLO	4276	2996	3248
TOTAL NNLO	4078	2663	3045

PROGRESS...

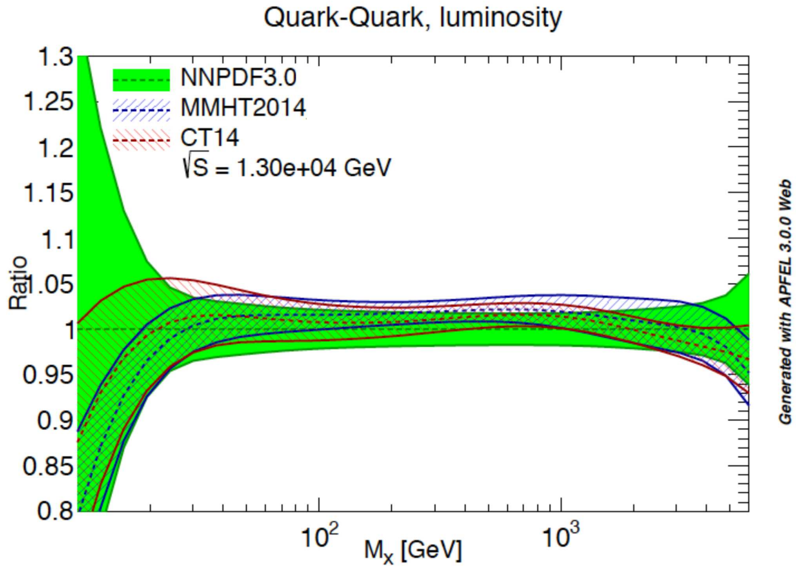
PARTON LUMINOSITIES: IMPROVED AGREEMENT

QUARK-QUARK

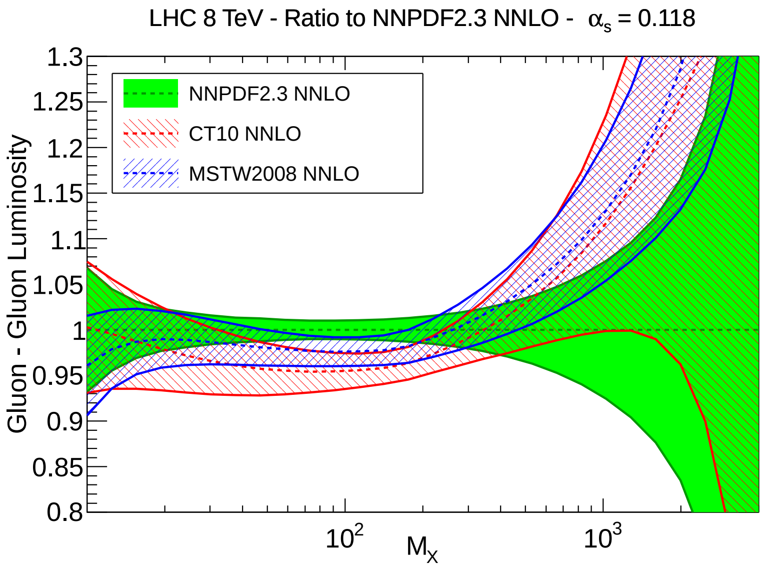
2012

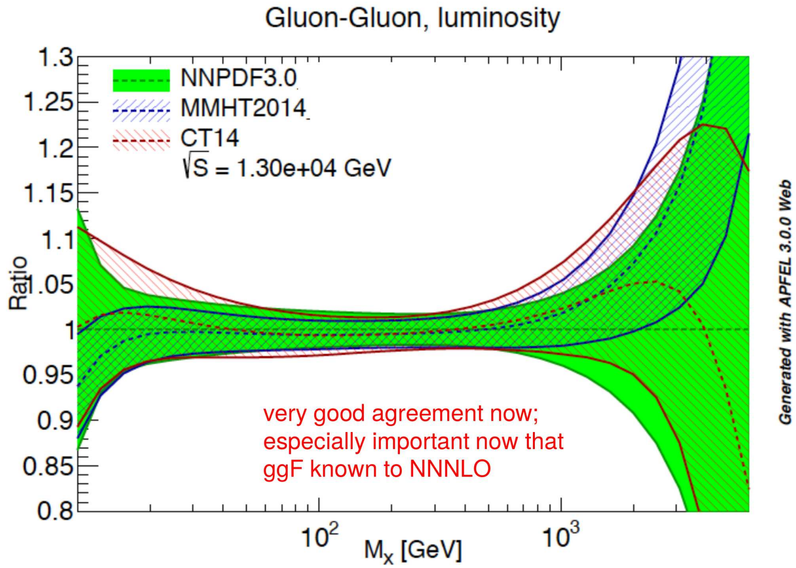


2015



GLUON-GLUON





“progress in convergence between the parton distribution functions will also be needed in order to reduce the theoretical uncertainties below the experimental measurement uncertainties.”

(J.Ellis, arXiv:1504.03654, April 15, 2015) "progress in convergence between the parton distribution functions will also be needed in order to reduce the theoretical uncertainties below the experimental measurement uncertainties."

(J.Ellis, arXiv:1504.03654, April 15, 2015)

PROGRESS!

HIGGS IN GLUON FUSION

A comparison of ggF at NNLO

	CT14	**MMHT2014**	**NNPDF3.0**
8 TeV	18.66 pb -2.2% +2.0%	18.65 pb -1.9% +1.4%	18.77 pb -1.8% +1.8%
13 TeV	42.68 pb -2.4% +2.0%	42.70 pb -1.8% +1.3%	42.97 pb -1.9% +1.9%

J.Huston, PDF4LHC, April 2015

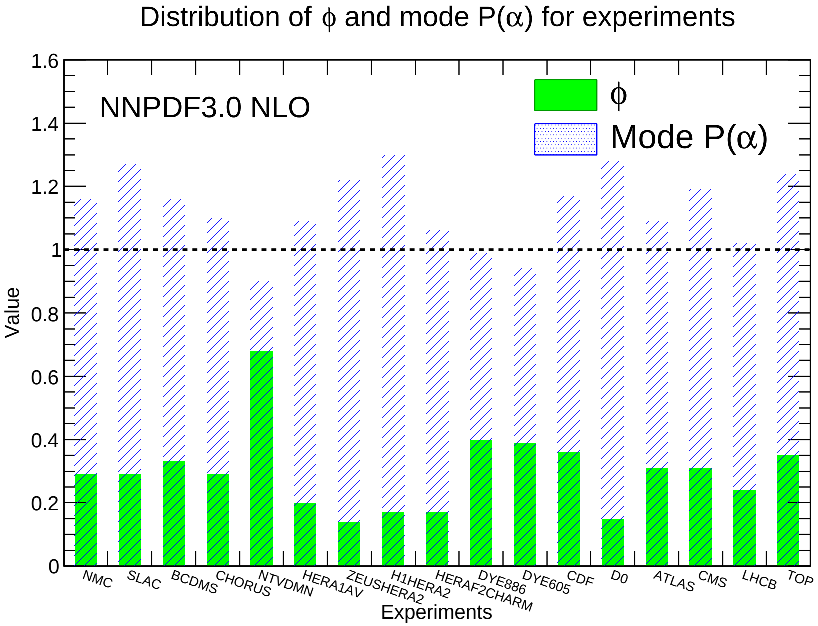
- almost perfect agreement between global PDF fits
- comes out of the box, thanks to methodological improvements

EXTRAS

DATA: WHERE LIES THE PROMISE

- CAN USE α & ϕ INDICATORS TO ASSESS OBJECTIVELY WHICH DATA COULD HAVE BIGGER IMPACT
- LARGE α MODE $\Rightarrow$ MUST IMPROVE COMPATIBILITY
- SMALL ϕ $\Rightarrow$ MUST REDUCE UNCERTAINTY

$P(\alpha)$ MODE & ϕ



	NNLO global fit		
Experiment	N_{dat}	$P(\alpha)$ mode	ϕ
ATLAS W, Z 2010	30	1.12	0.32
ATLAS 7 TeV jets 2010	90	0.92	0.25
ATLAS 2.76 TeV jets	59	0.57	0.34
ATLAS high-mass DY	5	1.34	0.11
CMS W electron asy	11	0.87	0.29
CMS W muon asy	11	1.40	0.51
CMS jets 2011	133	1.07	0.37
CMS $W+c$ total	5	1.09	0.24
CMS $W+c$ ratio	5	1.39	0.39
CMS 2D DY 2011	88	1.27	0.13
LHCb W, Z rapidity	19	1.12	0.24
$\sigma(t\bar{t})$	6	0.75	0.35

SOME PRELIMINARY INDICATIONS...

- 2.76TEV JET DATA SEEEM TO HAVE OVERESTIMATED UNCERTAINTIES, POTENTIALLY HIGH IMPACT
- TOP DATA: SUPRISINGLY HIGH IMPACT GIVEN SMALL DATA SAMPLE
- HIGH-MASS DY $\Rightarrow$ CONSISTENCY PROBLEM

AS SHOWN AT ATLAS-ITALY BOLOGNA MEETING

January 2014

LHC DATA

FOR PDF DETERMINATION

THE GOAL: A "COLLIDER ONLY" PDF DETERMINATION

- MEDIUM & LARGE x GLUON
  - TOP RAPIDITY DISTRIBUTIONS (PARTIAL NNLO)
  - INCLUSIVE W, Z p_T DISTRIBUTIONS (NNLO SOON)
  - JETS (PARTIAL NNLO)
  - PROMPT PHOTONS (NLO)
  - DIJETS? (NLO)
  - W POLARIZATION? (NLO)
- LIGHT FLAVOR SEPARATION
  - LOW–MASS & HIGH–MASS DRELL-YAN (NNLO)
  - DOUBLE-DIFFERENTIAL DY RAPIDITY DISTRIBUTIONS (NNLO)
  - Z RAPIDITY DISTRIBUTIONS (NNLO)
  - W ASYMMETRIES (NNLO)
- STRANGENES & HEAVY FLAVORS
  - STRANGENESS $\Rightarrow$ $W + c$ (NLO)
  - CHARM $\Rightarrow$ $Z + c$, $\gamma + c$ (NLO)
  - BOTTOM $Z + b$ (NLO)

CURRENT DESIDERATA
November 2014
LHC DATA
FOR PDF DETERMINATION
THE GOAL: A "COLLIDER ONLY" PDF DETERMINATION

- MEDIUM & LARGE x GLUON
  - TOP RAPIDITY DISTRIBUTIONS (NNLO) HIGH PRIORITY
  - INCLUSIVE W, Z p_T DISTRIBUTIONS (NNLO SOON)
  - JETS (PARTIAL NNLO) LARGE p_T, CENTRAL RAPIDITY
  - PROMPT PHOTONS (NLO)
  - DIJETS? (NLO)
  - W POLARIZATION? (NLO)
- LIGHT FLAVOR SEPARATION
  - LOW–MASS & HIGH–MASS DRELL-YAN (NNLO) NEED BETTER CONTROL ON SYSTEMATICS
  - DOUBLE-DIFFERENTIAL DY RAPIDITY DISTRIBUTIONS (NNLO) DITTO, PERHAPS FINER MASS BINNING
  - Z RAPIDITY DISTRIBUTIONS (NNLO)
  - W ASYMMETRIES (NNLO)
- STRANGENES & HEAVY FLAVORS
  - STRANGENESS $\Rightarrow$ $W + c$ (NLO)
  - CHARM $\Rightarrow$ $Z + c$, $\gamma + c$ (NLO)
  - BOTTOM $Z + b$ (NLO)

PRESENT...

- NNPDF2.3 LO AND NLO ARE THE DEFAULT SETS IN MADGRAPH5_AMC@NLO
- NNPDF2.3QED LO IS THE DEFAULT SET IN PYTHIA8, AND USED IN CURRENT REFERENCE TUNE (MONASH 2013)
- NNPDF3.0 IS THE DEFAULT INTERNAL PDF SET IN SHERPA2.2

...AND FUTURE

NNPDF3.1 WILL INCLUDE:

- RUNNING HEAVY QUARK MASSES & INDEPENDENT CHARM PDFS
- AND NEW DATA:
  - HERA-II COMBINED DATA (IF AVAILABLE)
  - ATLAS LOW-MASS DY `arXiv:1404.1212`
  - ATLAS PROMPT PHOTON `arXiv:1311.1440`
  - ATLAS $W + c$ `arXiv:1402.6263`
  - ATLAS Z p_t `arXiv:1107.2381`
  - ATLAS INCLUSIVE JETS 7 TEV 5 FB^{-1} `arXiv:1410.8857`
  - CMS W p_T 8 TEV (IF AVAILABLE)
  - LHCB $\to \mu\nu$ RAP. DISTN. `arxiv:1408.4354`
  - ATLAS+CMS TOP RAPIDITY `arXiv:1207.5644, 1407.0371, 1211.2220`
  - POSSIBLY, DIJETS+ THREE JETS

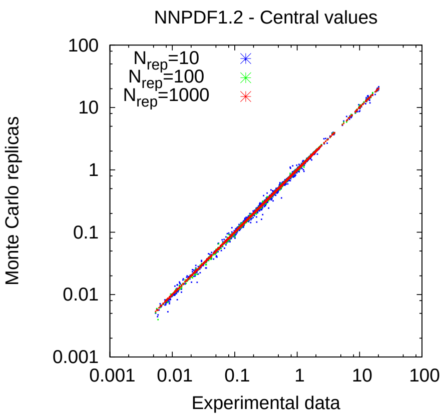
THE NNPDF APPROACH:
THE NEURAL MONTE CARLO

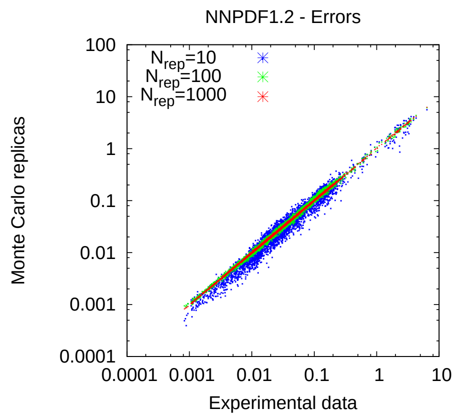
BASIC IDEA: MONTE CARLO SAMPLING OF THE PROBABILITY MEASURE IN THE (FUNCTION) SPACE OF PDFs

- START FROM MONTE CARLO SAMPLING OF DATA SPACE
- SPACE OF FUNCTIONS HUGE
  5 BINS FOR 10 PTS$\times$ 7 FCTNS $\to 5^{70} \sim 10^{49}$ BINS
- IMPORTANCE SAMPLING: DATA TELL US WHICH BINS ARE POPULATED

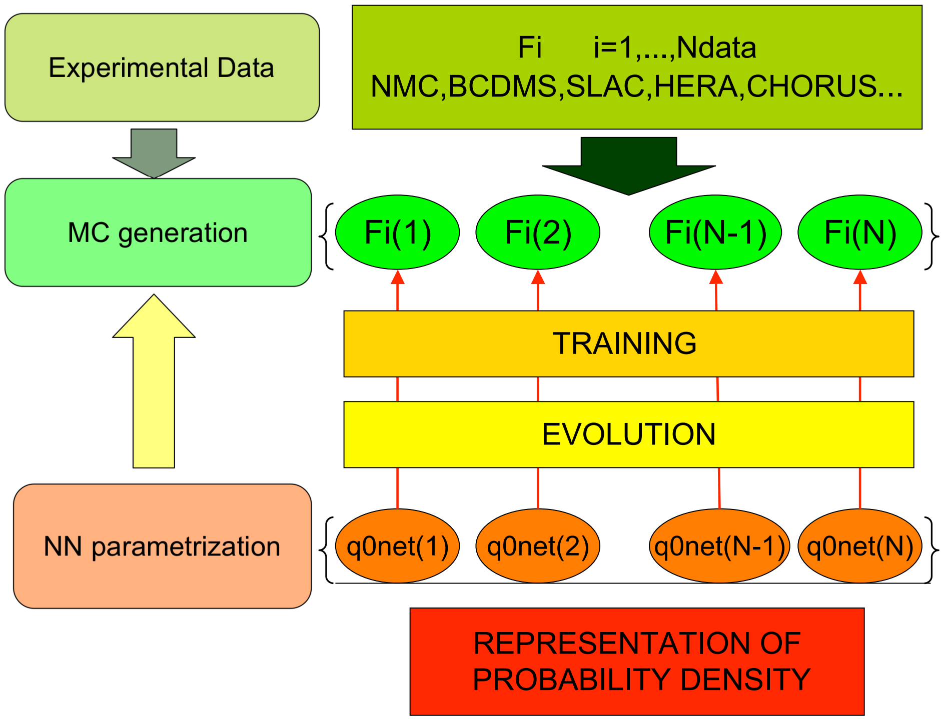
replica averages vs. central values

replica standard dev. vs. uncertainties



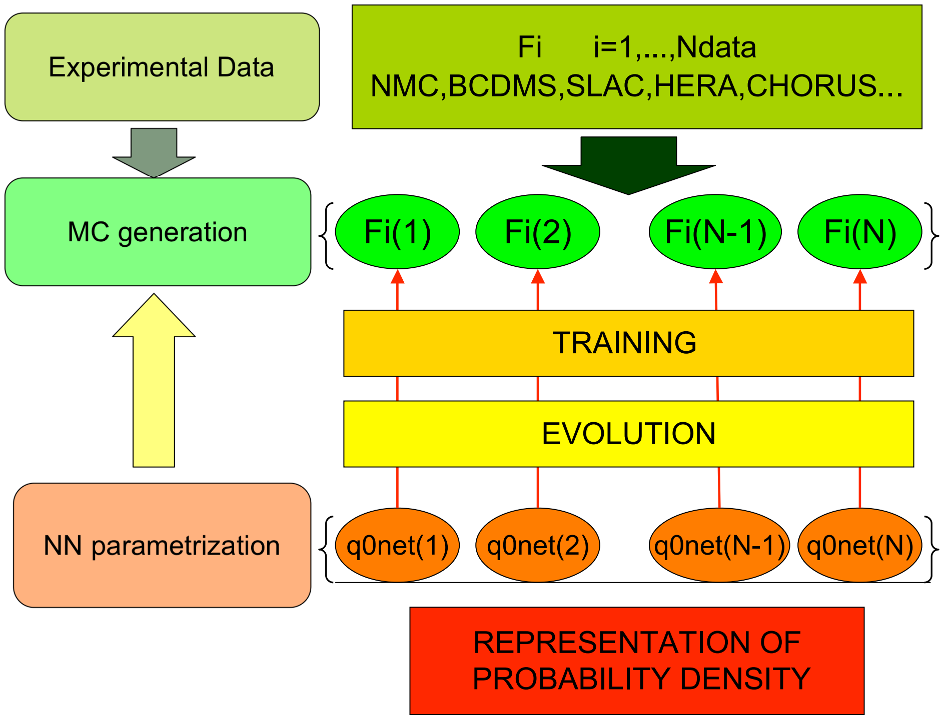


10 REPLICAS ENOUGH FOR CENTRAL VALS, 100 FOR UNCERTAINTIES, 1000 FOR CORRELNS



DATA MONTE CARLO ⇒ PDF MONTE CARLO
NEURAL NETWORK PARM+ CROSS-VALIDATION METHOD

- EACH PDF↔ NEURAL NETWORK PARAMETRIZED BY 37 PARAMETERS
- NNPDF1.2: $37 \otimes 7 = 259$ PARMS (RECALL MSTW, CTEQ → 20 FREE PARAMETERS)
  "INFINITE" NUMBER OF PARAMETERS⇒ CAN REPRESENT ANY FUNCTION
- COMPLEX SHAPES (LARGE NO.OF PARAMETERS) REQUIRE LONGER FITTING
- FIT STOPS WHEN QUALITY OF FIT TO RANDOMLY SELECTED "VALIDATION" DATA (NOT FITTED) STOPS IMPROVING
- CAN OBTAIN A FIT WITH χ^2 LOWER THAN BEST FIT ("OVERLEARNING")

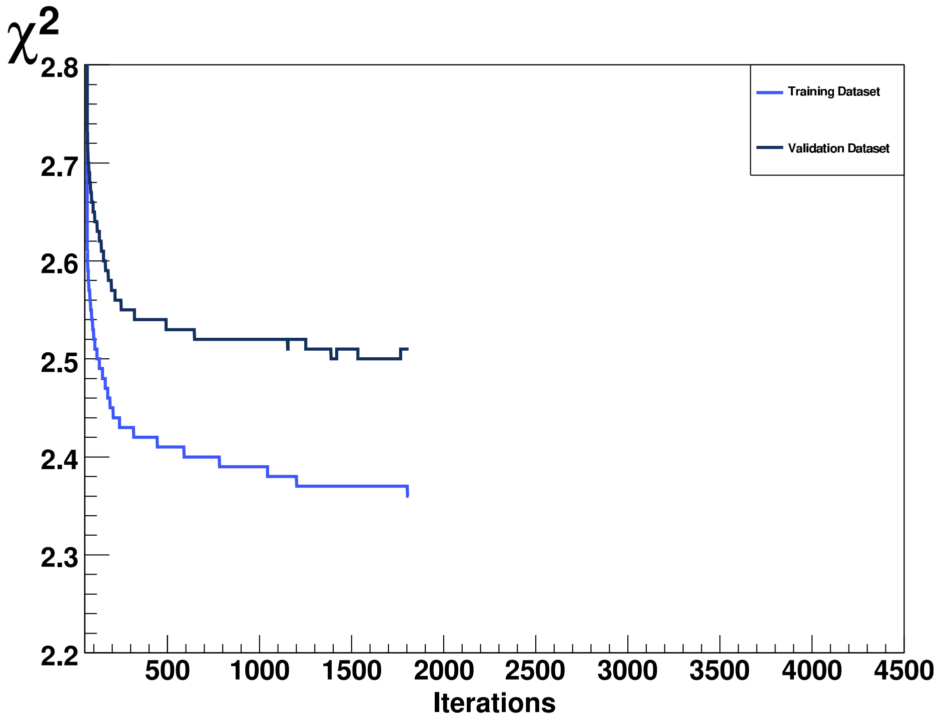


DATA MONTE CARLO ⇒ PDF MONTE CARLO
CROSS-VALIDATION

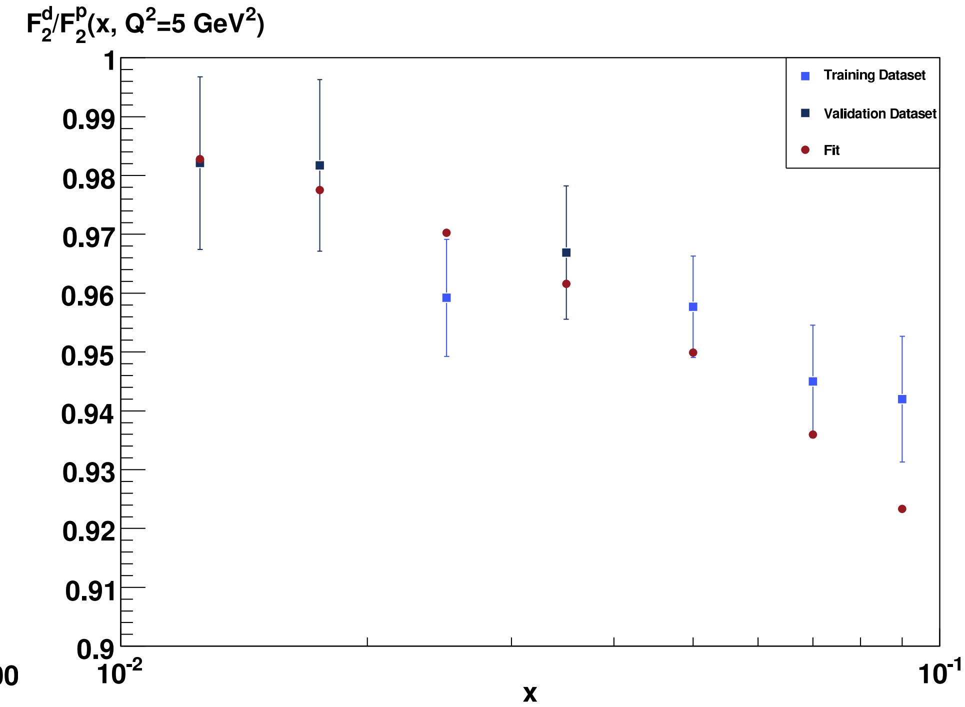
- REPLICAS ARE FITTED TO A DATA SUBSET
- A DIFFERENT SUBSET OF DATA USE FOR EACH REPLICA
- OPTIMAL FIT WHEN FIT TO VALIDATION (CONTROL) DATA STOPS IMPROVING
-

OPTIMAL FITTING

χ^2



FIT TO DATA

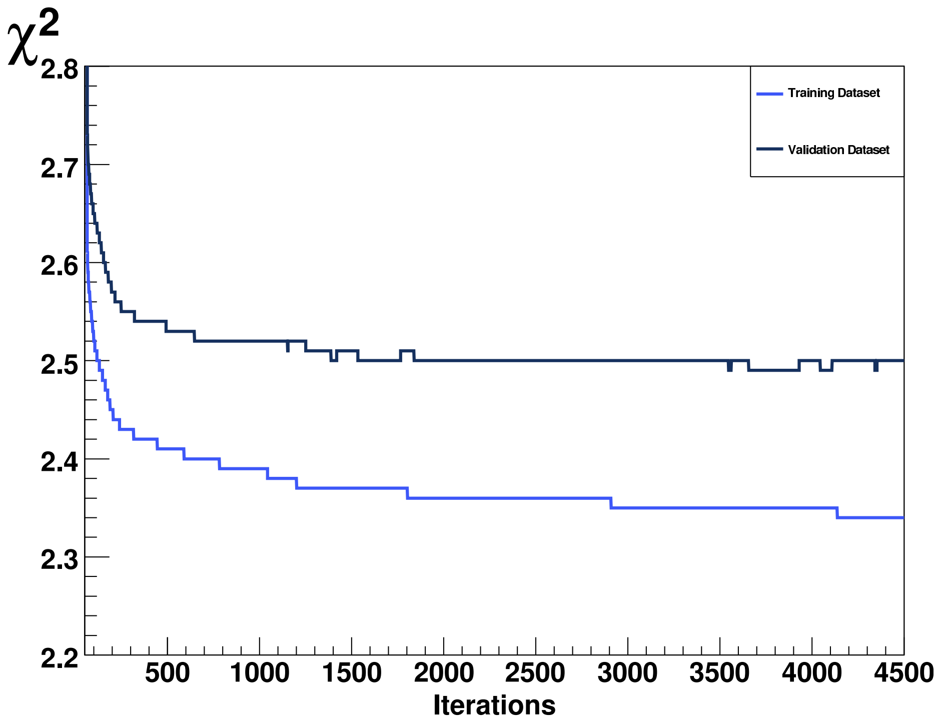


DATA MONTE CARLO ⇒ PDF MONTE CARLO
CROSS-VALIDATION

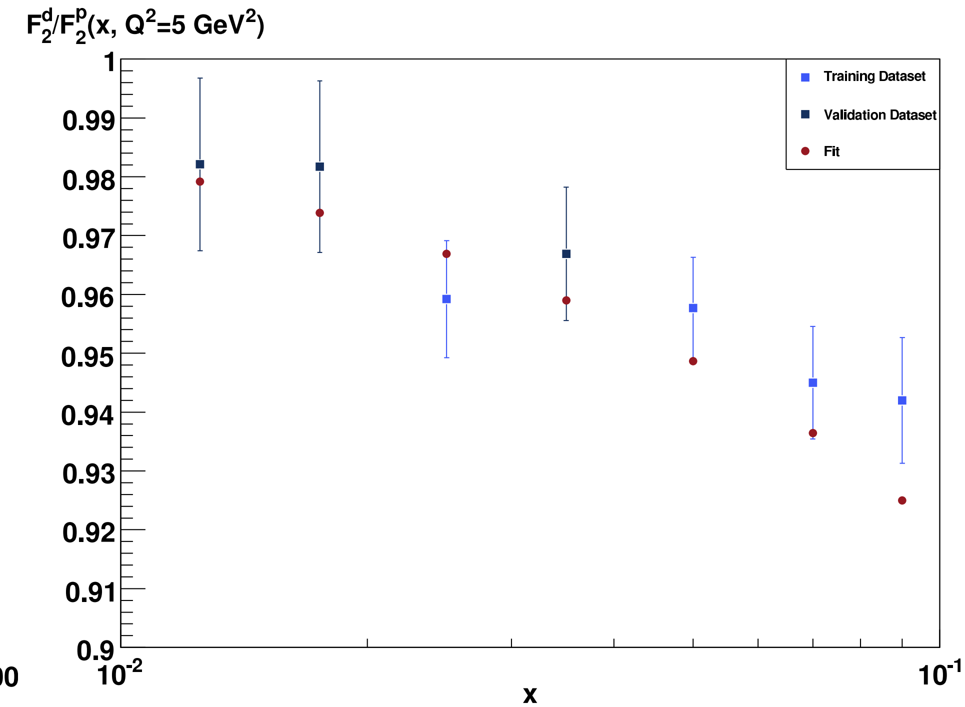
- REPLICAS ARE FITTED TO A DATA SUBSET
- A DIFFERENT SUBSET OF DATA USE FOR EACH REPLICA
- OPTIMAL FIT WHEN FIT TO VALIDATION (CONTROL) DATA STOPS IMPROVING
- THE BEST FIT IS NOT AT THE MINIMUM OF THE χ^2

OVERFITTING

χ^2



FIT TO DATA

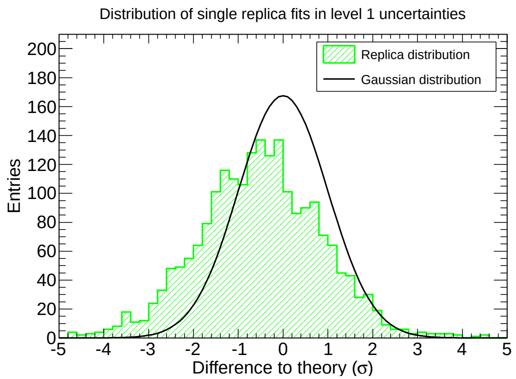


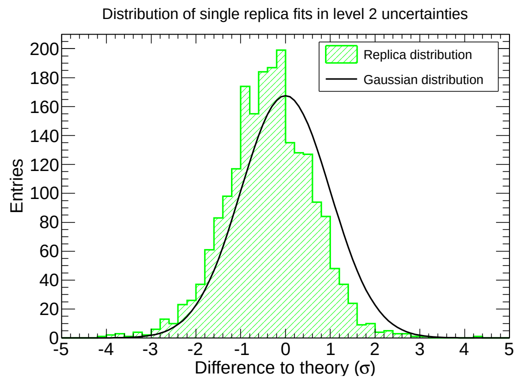
LOOKBACK CROSS-VALIDATION

- IN PREVIOUS NNPDF RELEASES, "INCREASING" AND "DECREASING" TRAINING AND VALIDATION χ^2 DEFINED IN TERMS OF THRESHOLD VALUES δ_{tr} AND δ_{val}:
  INCREASE: $\chi^2_{val}(N_{gen} + \Delta) > \chi^2(N_{gen} + \Delta) + \delta_{val}$
- IN NNPDF3.0 USE LOOKBACK:
  FIT IS RUN UP TO SOME LARGE # OF GA GENERATIONS
  THEN ONE "LOOKS BACK" FOR ABSOLUTE MIN. OF VALIDATION χ^2
- CHECK THAT RESULTS ARE INDEPENDENT OF THE LARGE # OF GA GENS
- CHECK THAT RESULTS ARE INDEPENDENT OF FLUCTUATIONS IN VALUE OF ABSOLUTE MINIMUM
  DIFFERENT STOPPING POINTS, BUT INDISTINGUISHABLE PDFS

LEVEL 1 VS. LEVEL 2
DO WE NEED THE PSEUDODATA REPLICAS?

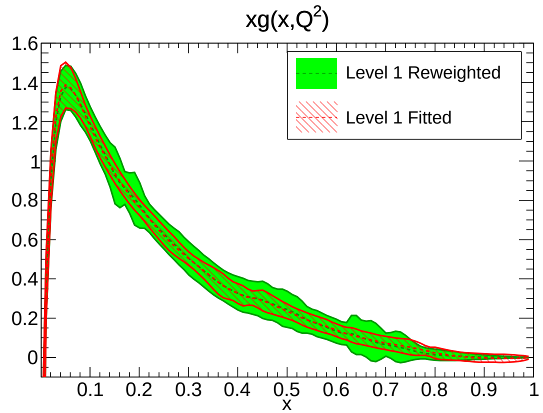
TRUTH-FIT: LEVEL 1 VS. LEVEL 2

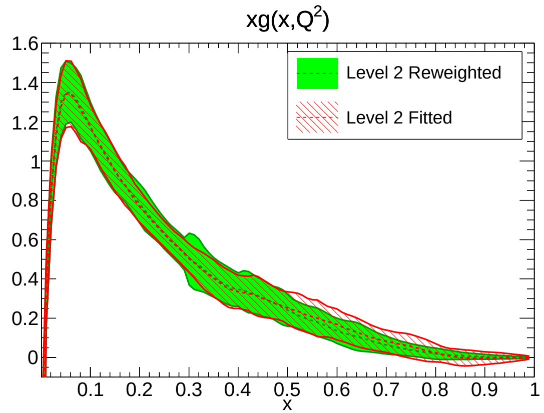




REWEIGHT VS REFIT:

LEVEL 1 VS. LEVEL 2

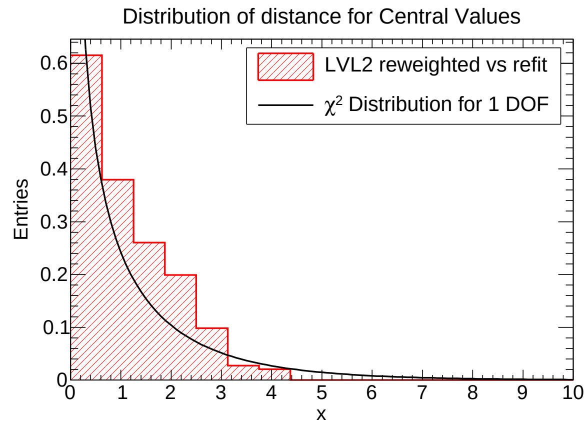




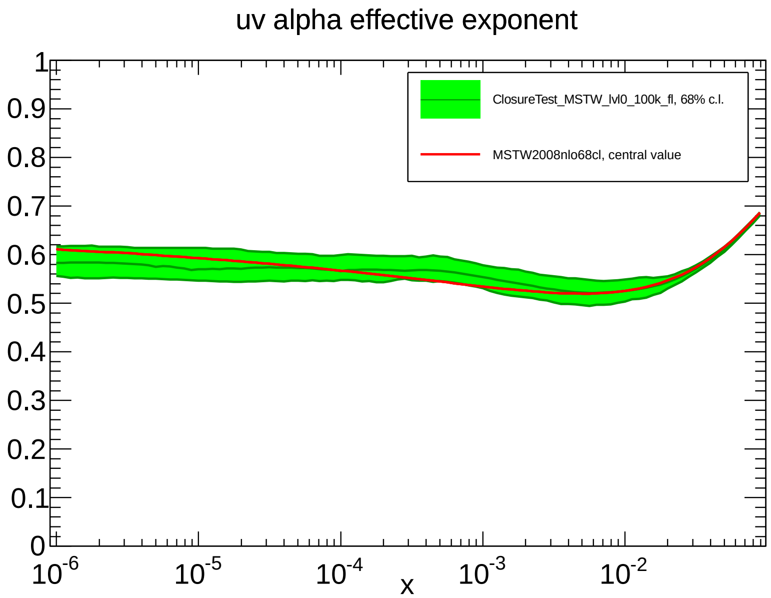
- LEVEL 1 UNDERESTIMATES UNCERTAINTIES
- LEVEL 0 TRUTH-FIT DISTN. WIDER THAN PREDICTED
- REWEIGHT VS. REFIT: X STAT. EQUIVALENCE FOR LEVEL 2
- REWEIGHT VS. REFIT: REFITTED SIGMA TOO SMALL FOR LEVEL 1

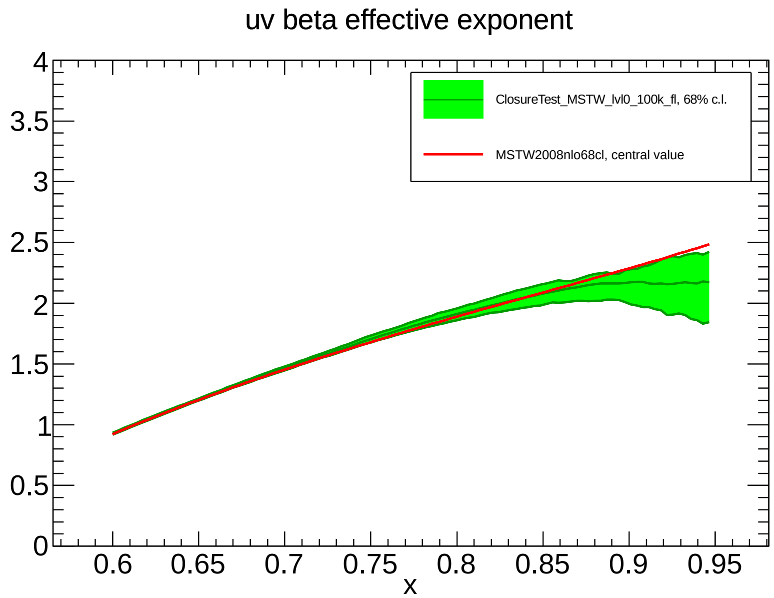
REWEIGHT-REFIT

(LEVEL 2)



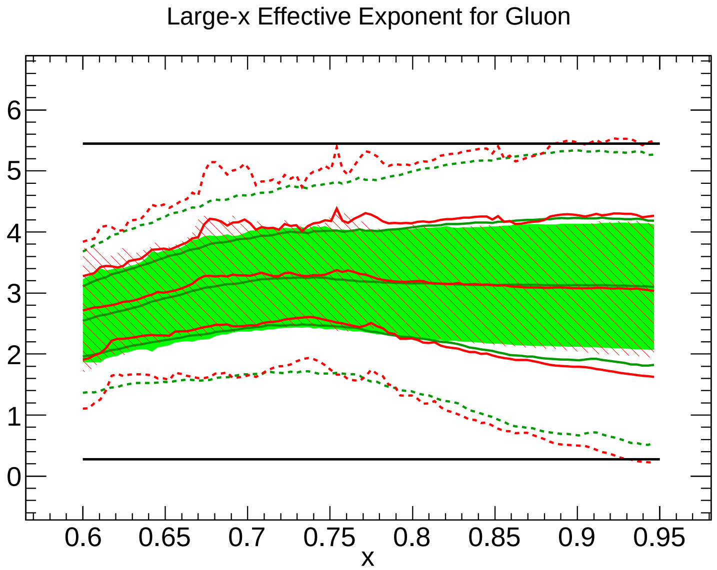
ASYMPTOTIC EXPONENTS
LEVEL 0 TEST

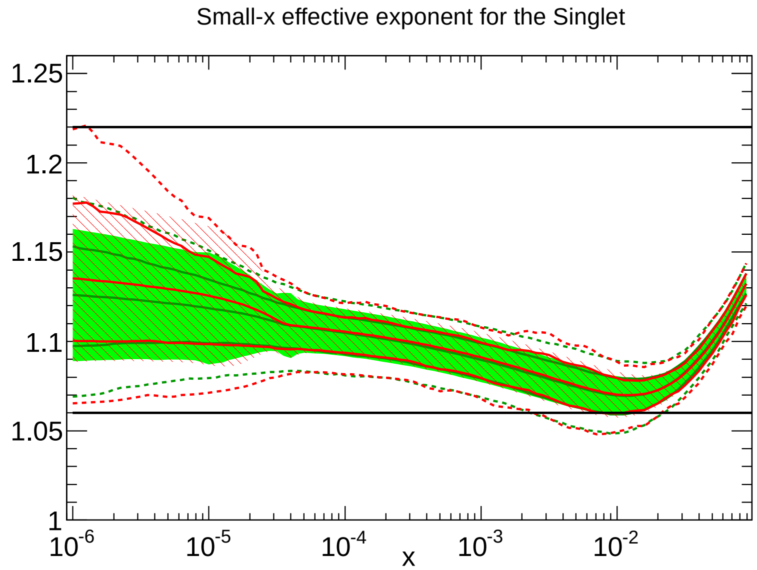




- AS $x \to 0$, PDF GOES AS x^α, AS $x \to 1$, PDF GOES AS $(1-x)^\beta$
- DEFINE EFFECTIVE EXPONENTS $\alpha_{\text{eff},i}(x) \equiv \frac{\ln f_i(x)}{\ln 1/x}$, $\beta_{\text{eff},i}(x) \equiv \frac{\ln f_i(x)}{\ln(1-x)}$
- COMPARE VALUES FROM FIT WITH VALUES OBTAINED FROM UNDERLYING "TRUE" (MSTW08) FORM

PREPROCESSING

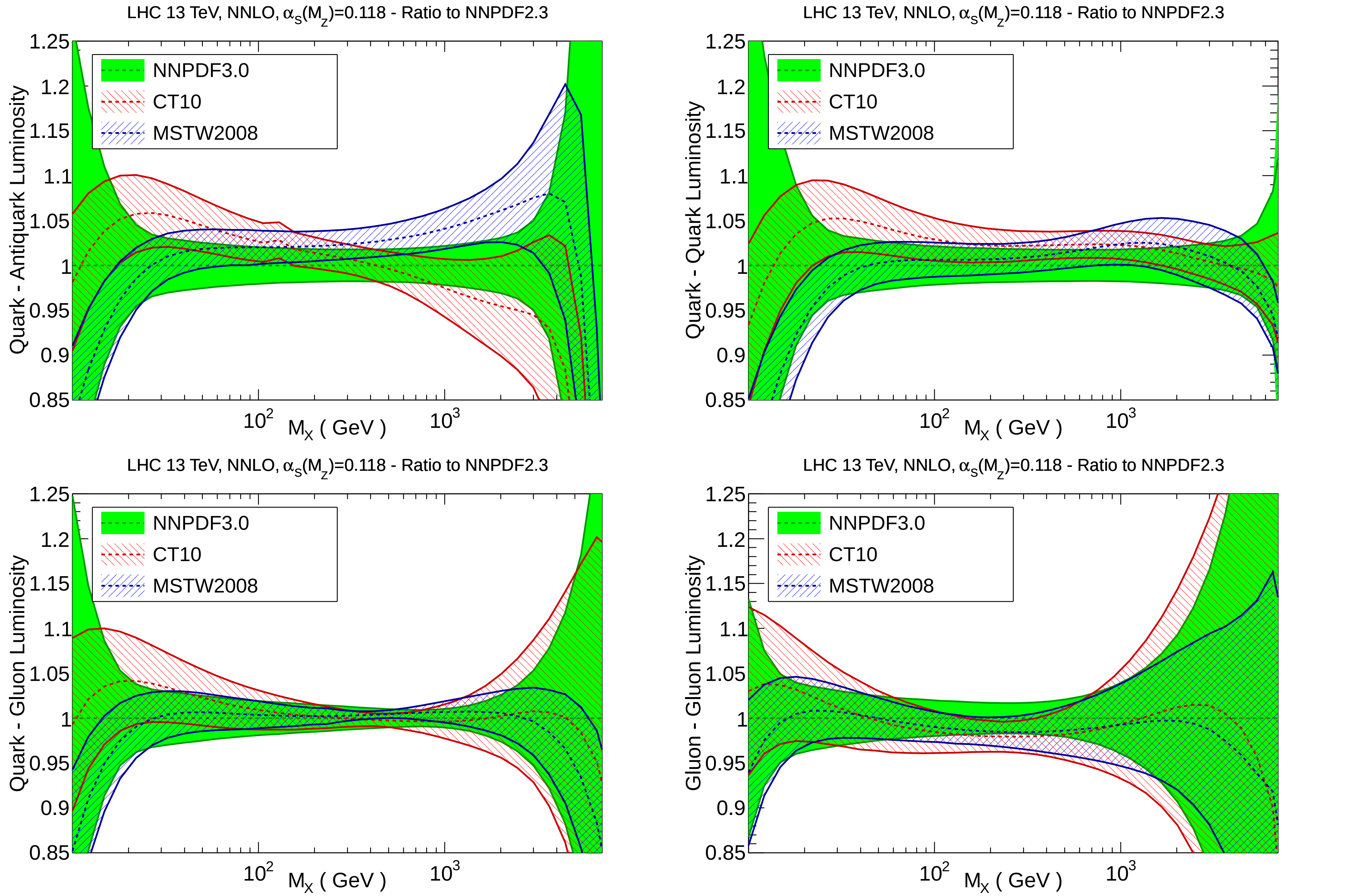




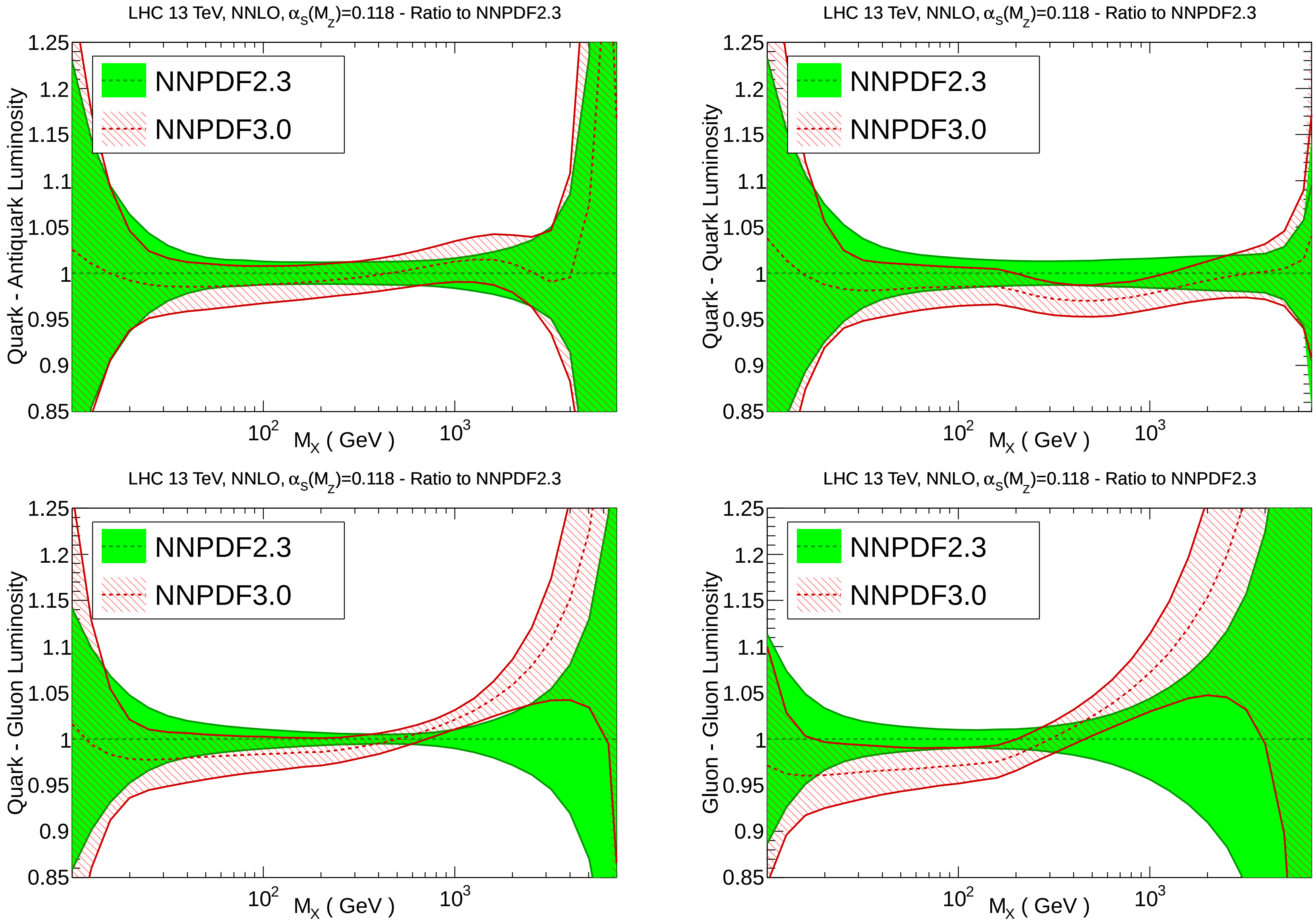
- PDFs ARE PARAMETRIZED WITH NEURAL NETWORKS TIMES PREPROCESSING FUNCTION: $f_i(x) = x^{\alpha_i}(1-x)^{\beta_i} NN(x)$
- GOAL IS TO SPEED UP TRAINING WITHOUT BIASING RESULT
- α_i, β_i RANDOM REPLICA BY REPLICA WITH UNIFORM DISTRIBUTION IN RANGE
- CHECK THAT THE RANGE IS WIDER THAN THE 90% C.L. OF EFFECTIVE EXPONENTS

NNPDF3.0 vs CT10 & MSTW08

PARTON LUMINOSITIES



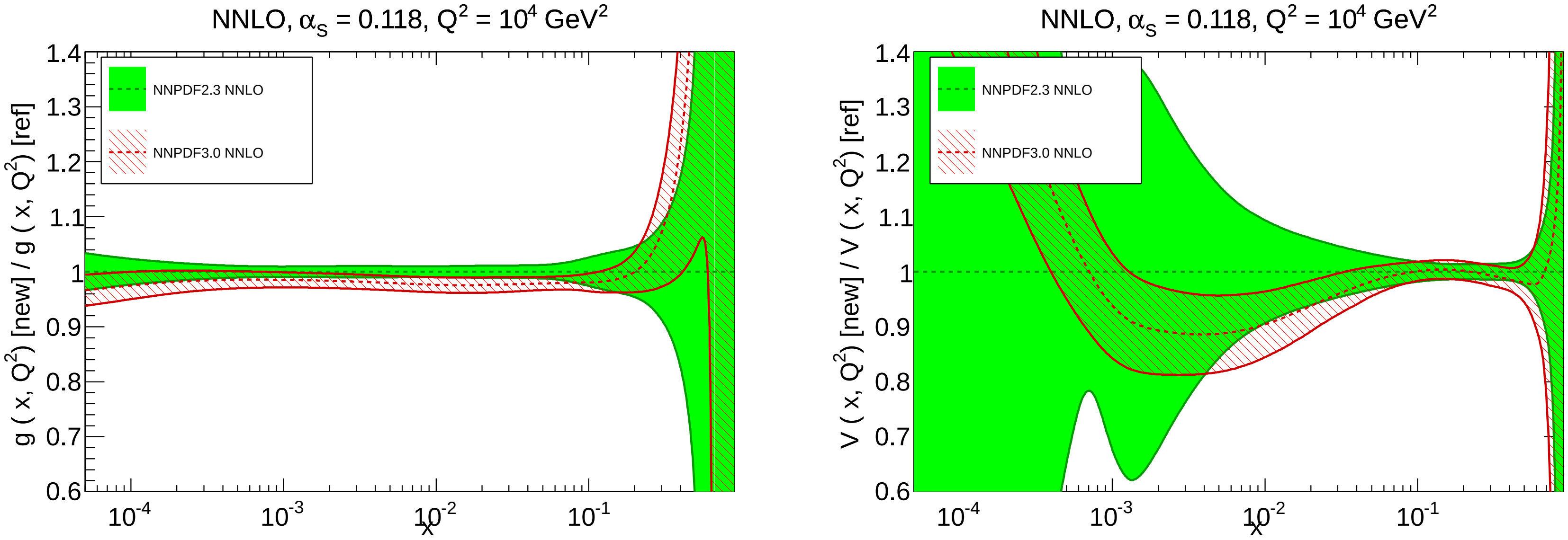
NNPDF2.3 vs NNPDF3.0
PARTON LUMINOSITIES



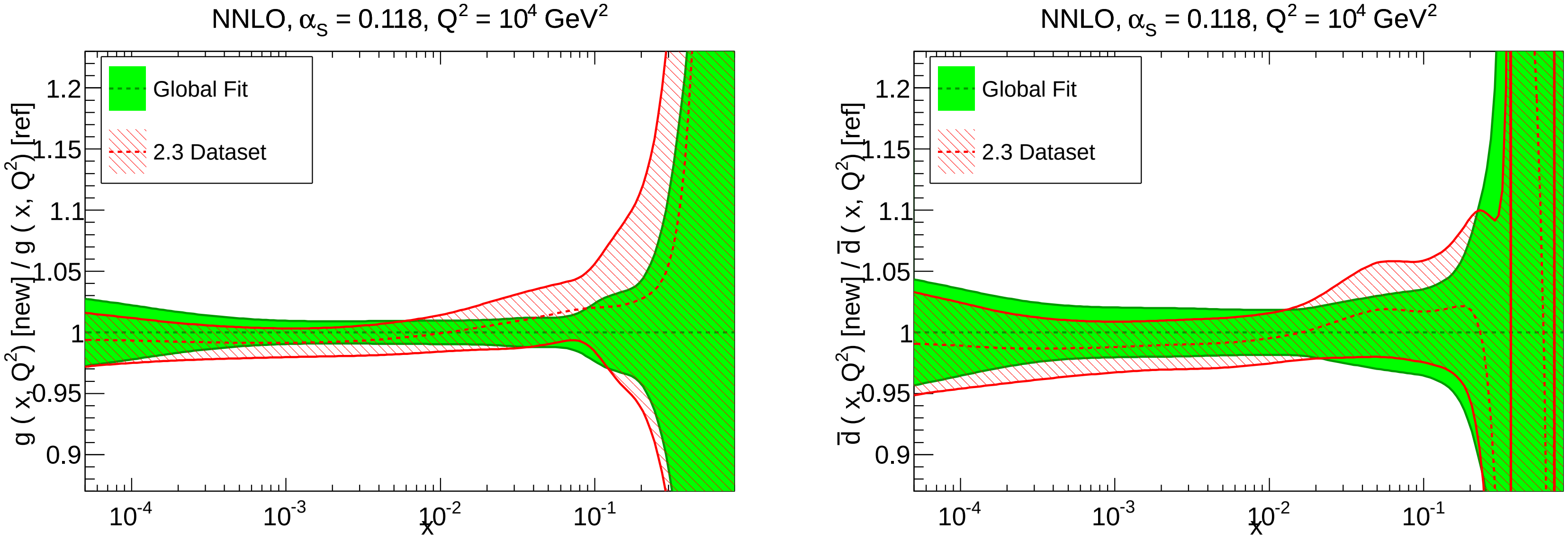
NNPDF2.3 vs NNPDF3.0

Methodology vs. Data

NNLO 2.3 vs 3.0: gluon & valence

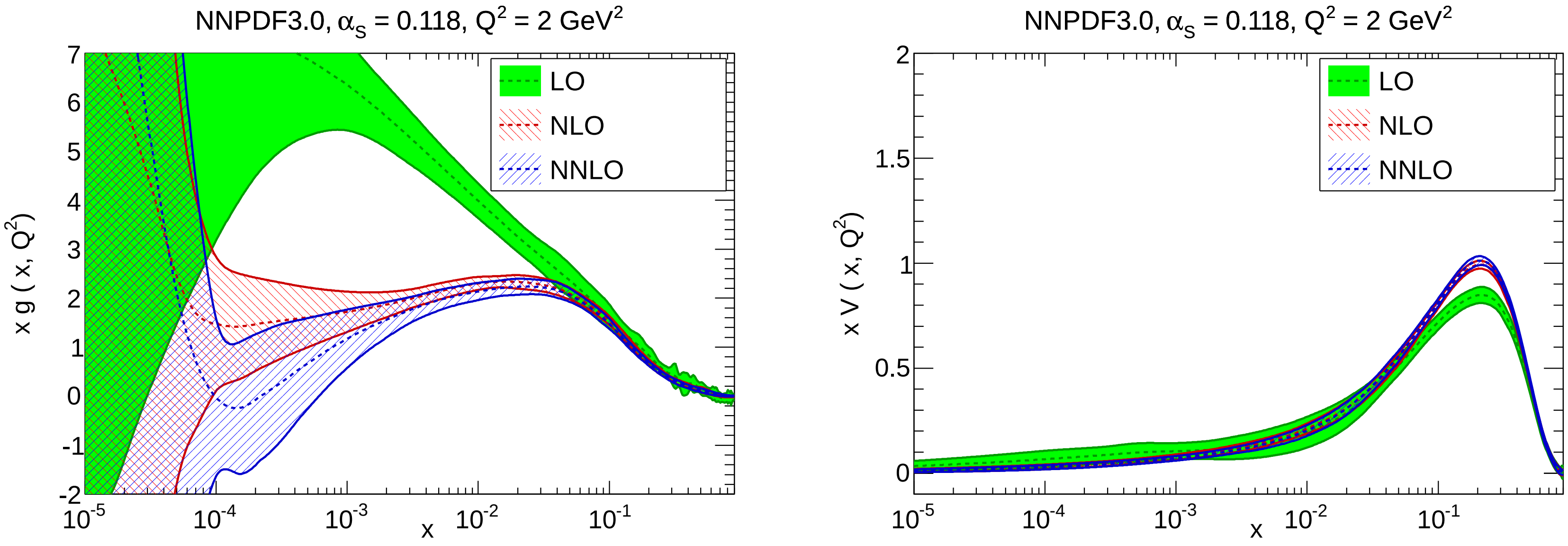


3.0 default vs 2.3-like dataset: gluon & antidown

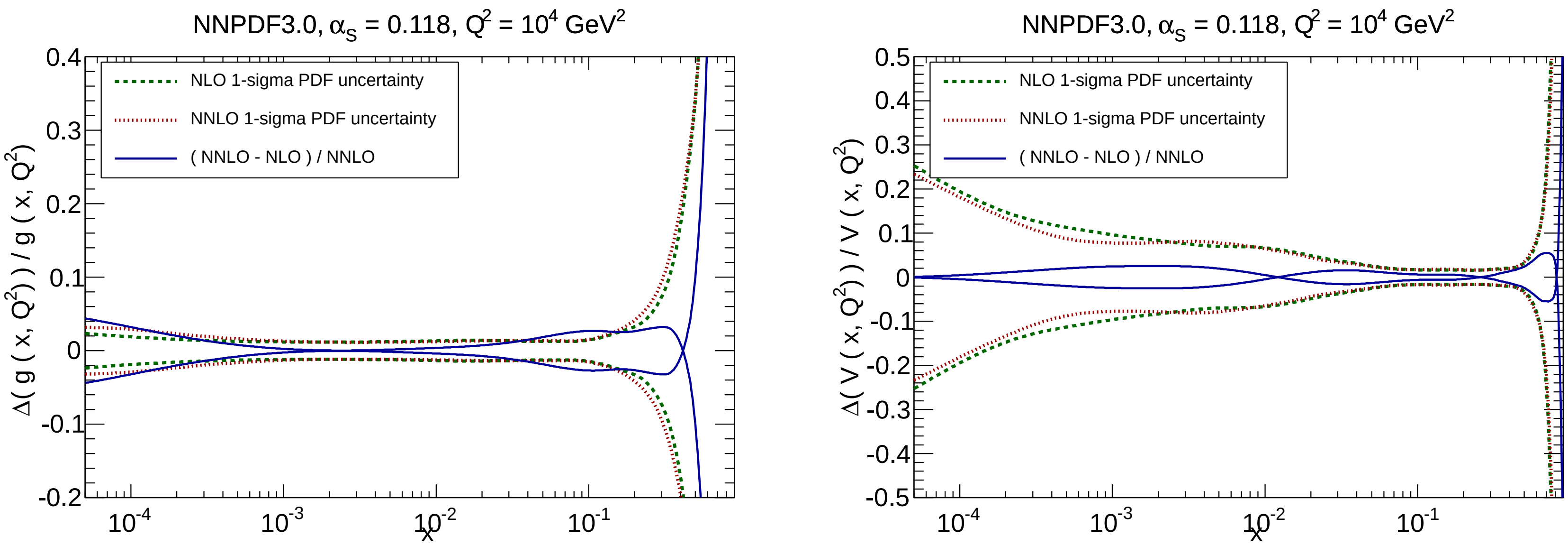


PERTURBATIVE STABILITY

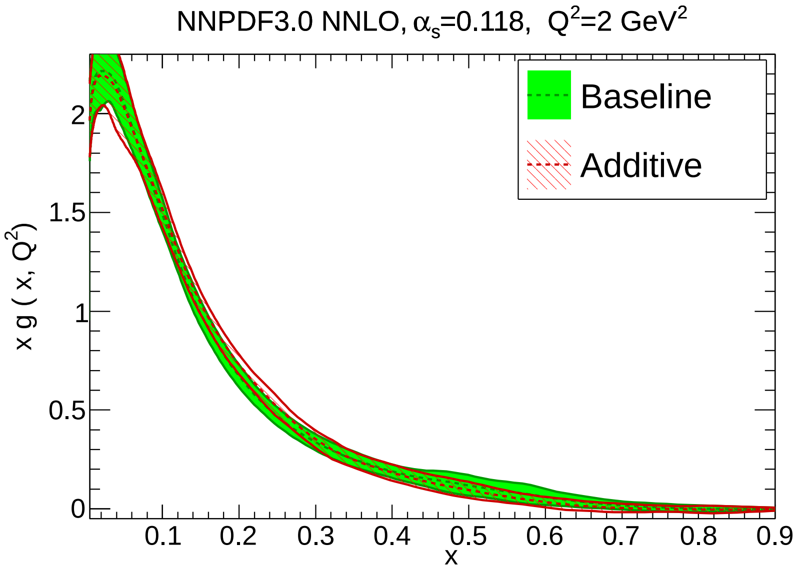
LO, NLO, NNLO: GLUON & VALENCE

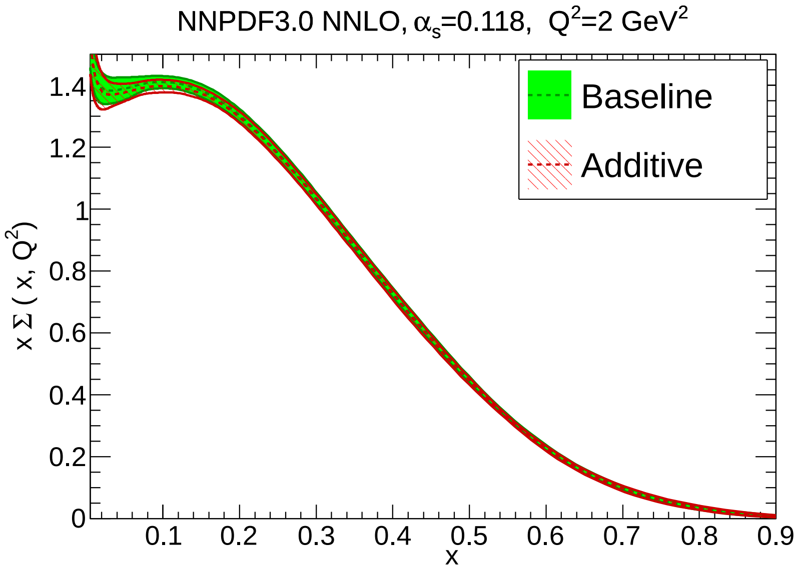


NLO, NNLO PDF UNC. & NLO TH. UNC.



MULTIPLICATIVE VS. ADDITIVE UNCERTAINTIES





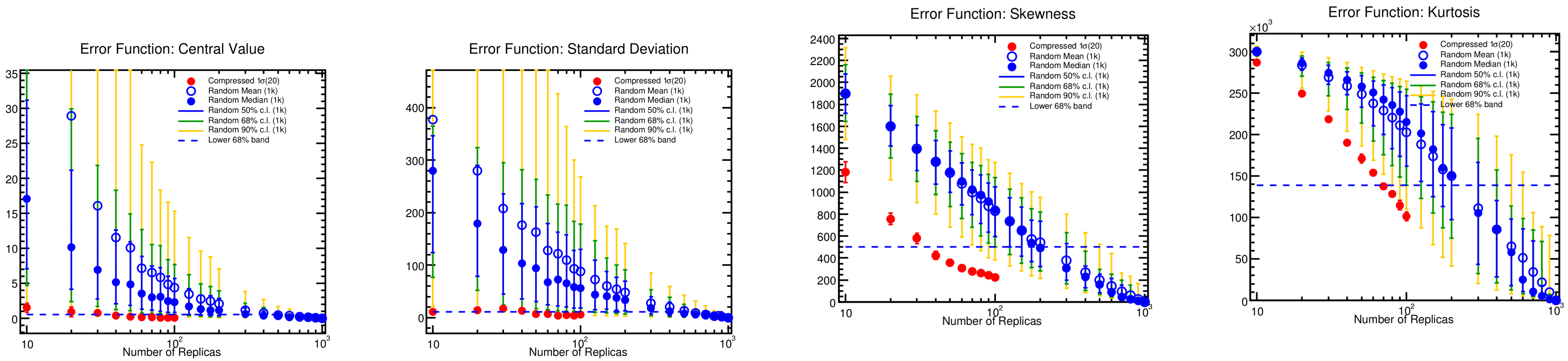
- NORMALIZATION UNCERTAINTY PROPORTIONAL TO MEASURED VALUE ⇒ MULTIPLICATIVE
- WHAT ABOUT OTHER UNCERTAINTIES?
  IN COLLIDER MOSTLY MULTIPLICATIVE (A. Glazov)
- IN NNPDF3.0 DEFAULT ALL UNCERTAINTIES MULTPLICATIVE FOR HERA, TEVATRON, LHC
- REPEAT FIT WITH ALL UNCERTAINTIES ADDITIVE (AS IN NNPDF2.3)
- t_0 METHOD USED FOR MULTIPLICATIVE UNCERTAINTIES
- NEGLIGIBLE CHANGE SEEN

COMBINED PDFs & META-PDFS

- FORTHCOMING PDF4LHC PRESCRIPTION BASED ON COMBINED SET
- ALL SETS CAN BE TURNED INTO MC (Thorne and Watt, 2013)
  $\Rightarrow$ COMBINED (POSSIBLY WEIGHTED) SET
- CAN ONE REDUCE THE SIZE OF THE COMBINED SET?
  – META-PDFS: FIT A FUNCTIONAL FORM TO THE COMBINATION
    FOR EACH PROCESS, PICK THE DOMINANT EIGENVECTORS (Gao and Nadolsky)
  – COMPRESSION: CONSTRUCT A SMALLER REPLICA SET WHICH
    PRESERVES LOWER MOMENTS OF PROBABILITY DISTRIBUTION (Carrazza and Latorre)

MOMENTS OF COMPRESSED SAMPLE

DISTANCE TO 1000 REPLICA SAMPLE



BEYOND NNPDF3.1

- MORE CLOSURE TESTS:
  - INCONSISTENT DATA
  - QUANTITATIVE MODELS OF FUNCTIONAL UNCERTAINTY
  - TOLERANCE & $\Delta\chi^2$ RULE
- THEORY UNCERTAINTY ON PDFS
  - SCALE VARIATION
  - PERTURBATIVE SERIES: CACCIARI-HOUDEAU
- RESUMMED PDFS
  - SMALL x AND EVOLUTION
  - THRESHOLD RESUMMATION AND NEW PHYSICS SEARCHES
- QED PDF SET
  - MIXED QCD-QED EFFECTS
  - DETERMINATION OF PHONEON FROM WW PRODUCTION
- MONTE CARLO PDFS