THEORY SYSTEMATICS IN PDFs

STEFANO FORTE
UNIVERSITÀ DI MILANO & INFN

CERN THEORY INSTITUTE

JUNE 16, 2017

THE PROBLEM:

THEORETICAL UNCERTAINTIES ON PDFs:

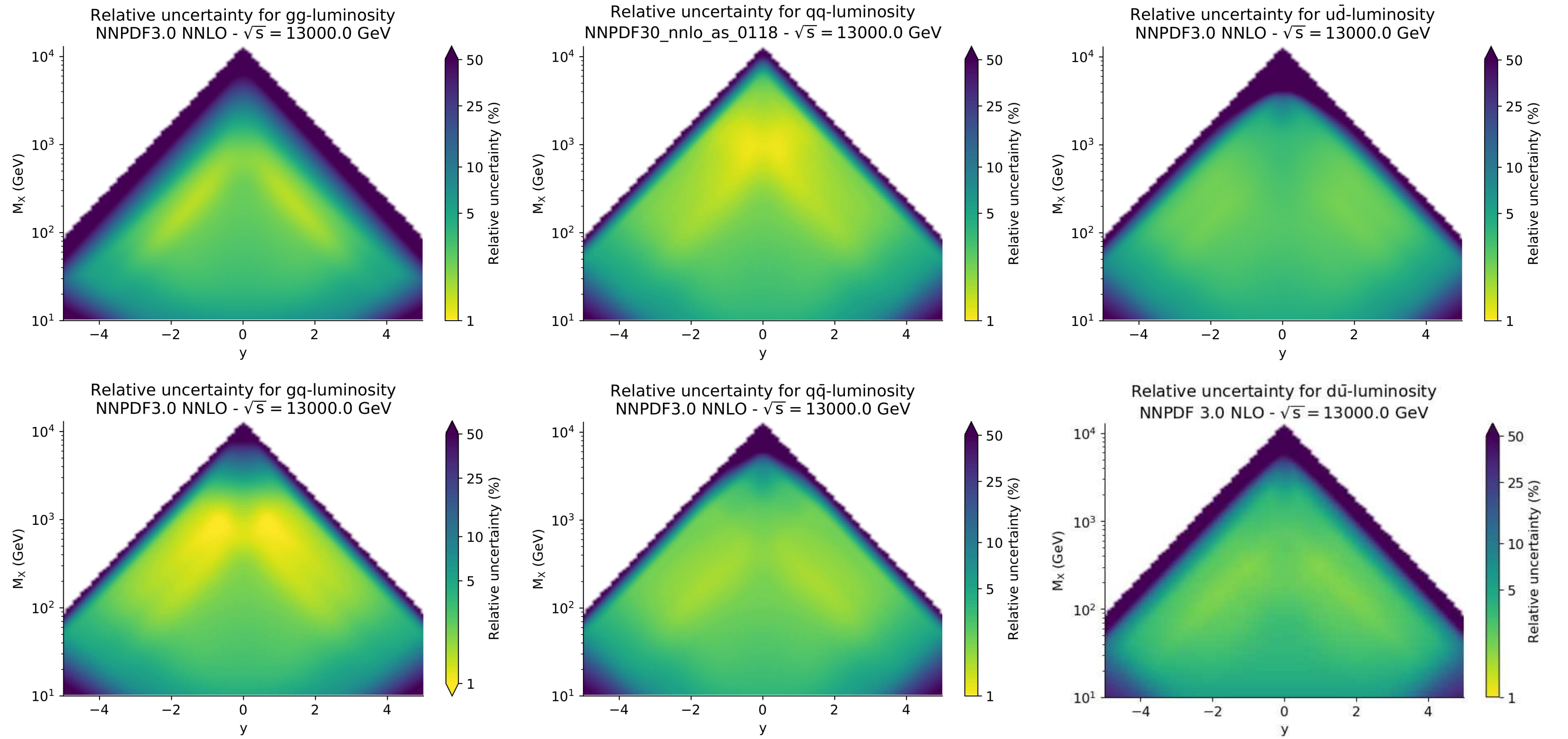
- PDFs ARE DETERMINED BY COMPARING TO DATA THEORY AT SOME FINITE ORDER
- AFFECTED BY THEORETICAL UNCERTAINTY JUST LIKE HARD CROSS-SECTIONS
- NOT INCLUDED IN CURRENT "PDF UNCERTAINTY"
  (ACCOUNTS ONLY DATA & METHODOLOGY)
- TERMINOLOGY:
  "PDF UNCERTAINTY" ⇔ PROPAGATED FROM DATA I.E. DATA+METHODOLOGY
  "THEORY UNCERTAINTY" ⇔ DUE TO MHO IN QCD EVOLUTION AND MATRIX ELEMENTS IN FIT
- NOTE PDF UNCERTAINTY TESTED TO BE FAITHFUL THROUGH CLOSURE TEST

THEORY UNCERTAINTIES: QUESTIONS

- HOW LARGE ARE THEY IN COMPARISON TO PDF UNCERTAINTIES
- HOW DO THEY RELATE TO THE UNDERLYING MHOU
- CAN WE ESTIMATE THEM

PDF UNCERTAINTIES: PAST $\Rightarrow$ PRESENT (NNPDF3.0 NNLO)

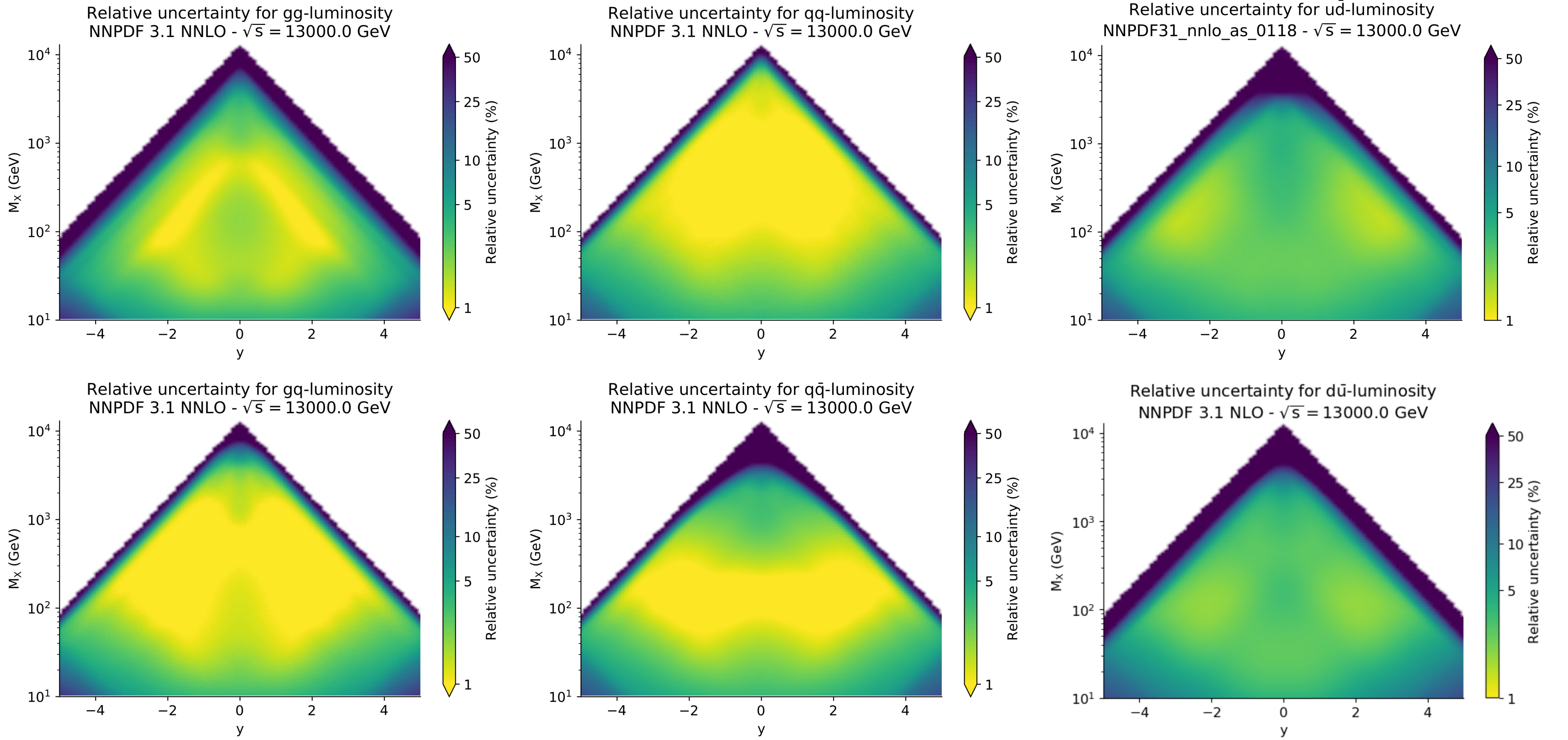
GLUON SINGLET FLAVORS



- GLUON BETTER KNOWN AT SMALL x, VALENCE QUARKS AT LARGE x, SEA QUARKS IN BETWEEN
- SWEET SPOT: VALENCE Q - G; UNCERTAINTIES DOWN TO 1%
- UP BETTER KNOWN THAN DOWN; FLAVOR SINGLET BETTER THAN INDIVIDUAL FLAVORS
- NO QUALITATIVE DIFFERENCE BETWEEN NLO AND NNLO

PDF UNCERTAINTIES: PRESENT ⇒ FUTURE (NNPDF3.1 NNLO)

GLUON SINGLET FLAVORS

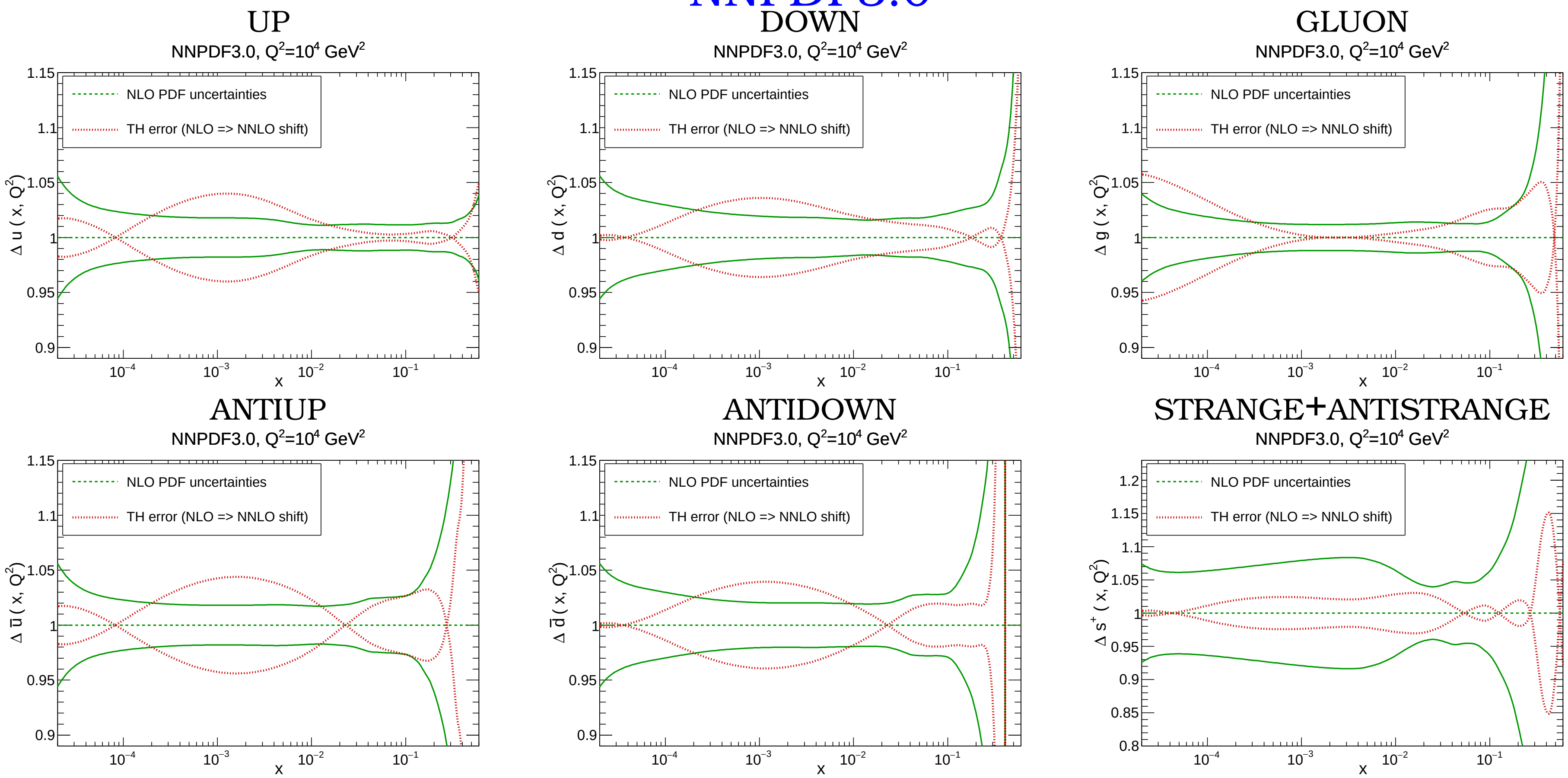


- GLUON BETTER KNOWN AT SMALL x, VALENCE QUARKS AT LARGE x, SEA QUARKS IN BETWEEN
- SWEET SPOT: VALENCE Q - G; UNCERTAINTIES DOWN TO 1%
- UP BETTER KNOWN THAN DOWN; FLAVOR SINGLET BETTER THAN INDIVIDUAL FLAVORS
- NO QUALITATIVE DIFFERENCE BETWEEN NLO AND NNLO
- NEW DATA ⇒ SIZABLE REDUCTION IN UNCERTAINTIES
- AT 1% LEVEL, CAN WE NEGLECT THEORY UNCERTAINTIES?

PDF UNCERTAINTIES vs. NLO THEORY UNCERTAINTIES

- AT NLO, THE THEORY UNCERTAINTY IS JUST THE NNLO-NLO SHIFT
- HOW DOES IT COMPARE TO THE PDF UNCERTAINTY?

NNPDF3.0

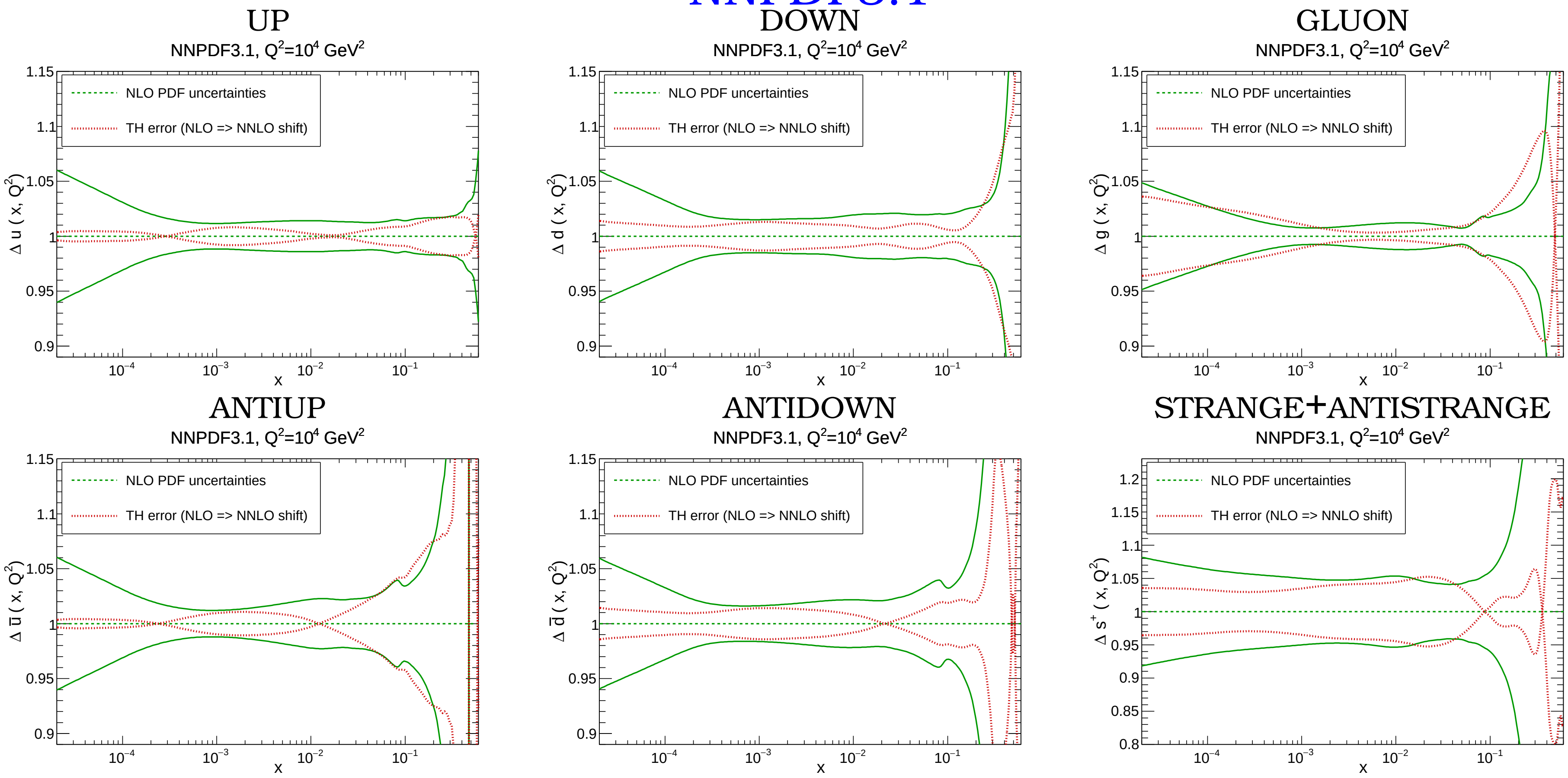


- IN CURRENT NLO SETS MISSING THEORY UNCERTAINTY COMPARABLE TO PDF UNCERTAINTY

PDF UNCERTAINTIES VS. NLO THEORY UNCERTAINTIES

- at NLO, the theory uncertainty is just the NNLO-NLO shift
- how does it compare to the PDF uncertainty?

NNPDF3.1



- new NLO sets: MISSING theory smaller but still comparable to PDF uncertainty
- can we estimate what happens at NNLO?

Cacciari-Houdeau for PDFs

(slides from 2011)

IDEA

- ASSUME COEFFICIENT OF PERTURBATIVE EXPANSION BOUNDED FROM ABOVE AND WITH SOME (UNIFORM?) DISTRIBUTION

- ESTIMATE SIZE OF NEXT ORDER BASED ON KNOWN COEFFICIENTS

series in α_s starting at α_s^0; uncertainty on k-th order:

$$\Delta_k = \begin{cases} \alpha_s^{k+1} \max\{|c_l|, \ldots, |c_k|\} \frac{n_c+1}{n_c} p & \text{if } p \leq \frac{n_c}{n_c+1} \\ \alpha_s^{k+1} \max\{|c_l|, \ldots, |c_k|\} \left[(n_c+1)(1-p)\right]^{-1/n_c} & \text{if } p > \frac{n_c}{n_c+1} \end{cases}$$

$n_c = k + 1$ number of known coefficients; P $\Rightarrow$ c.l. (one $\sigma \leftrightarrow P = 0.68$)

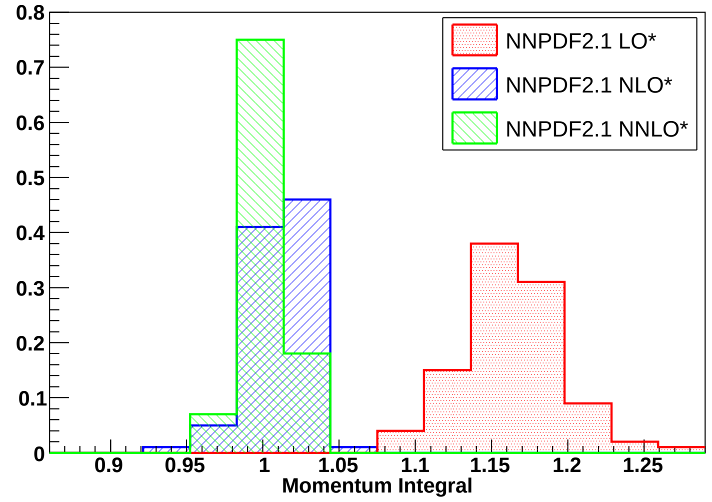
HOW WELL DOES IT WORK?

WHEN WE KNOW THE ANSWER

THE MOMENTUM SUM RULE

- PERFORM LO, NLO, NNLO PDF FITS WITHOUT MOMENTUM CONSTRAINT
- DETERMINE $[M] = \int_0^1 dx\Sigma(x) + g(x)$
  (AT ANY SCALE)
- VIEW LO, NLO, NNLO RESULTS AS SERIES IN α_s

- LO $[M] = 1.161 \pm 0.032^{\text{exp}}$
- NLO $[M] = 1.011 \pm 0.018^{\text{exp}}$; $\Delta^{\text{th,CH}} = 0.019$
- NNLO $[M] = 1.002 \pm 0.014^{\text{exp}}$; $\Delta^{\text{th,CH}} = 0.002$



WORKS QUITE WELL, ACCURACY IMPROVES WITH ORDER!

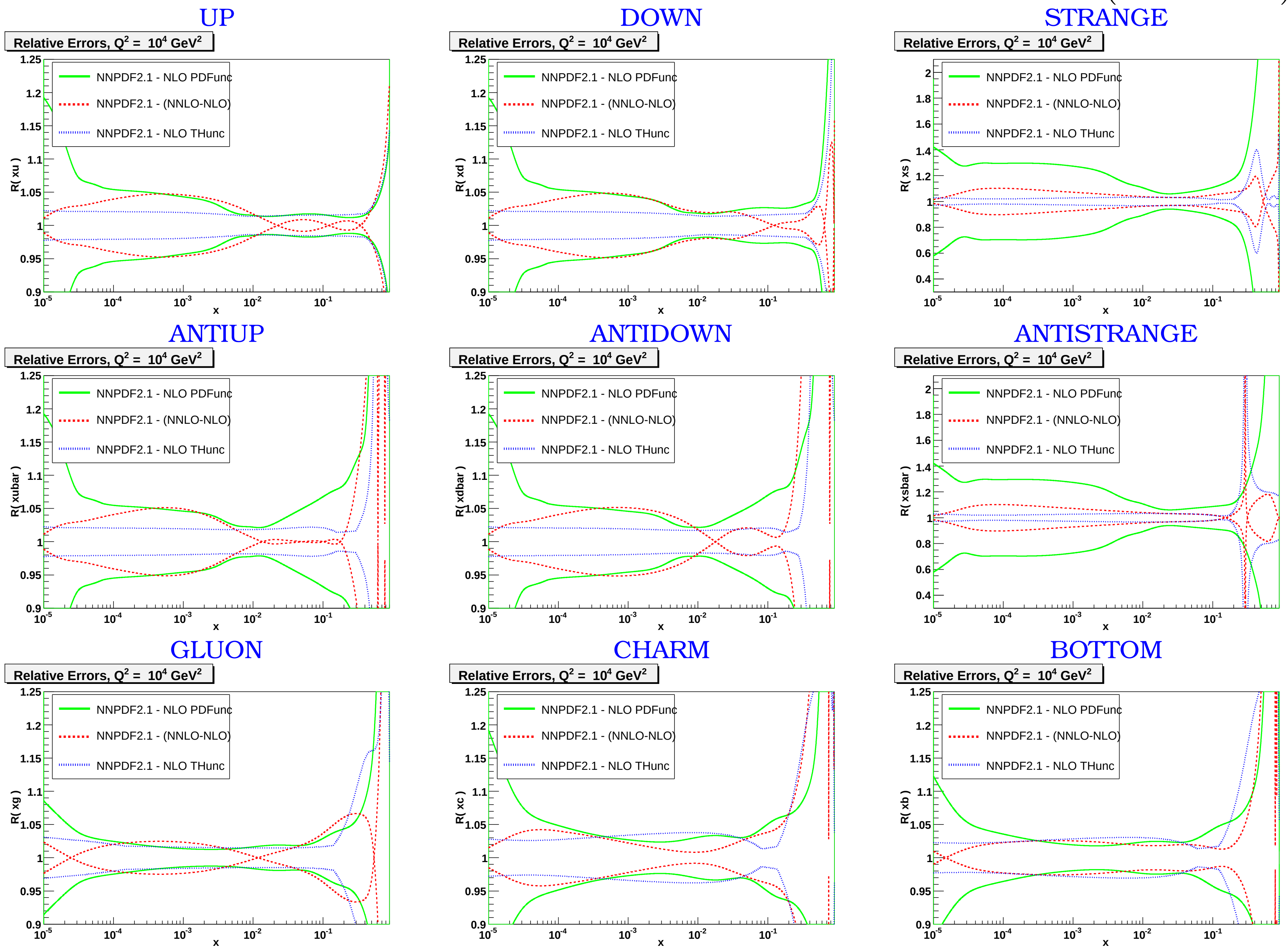
The Cacciari-Houdeau method

application to PDFs

- consider PDFs for given x, Q^2 as a series in α_s
- at NLO use formula with $k = 1$ etc. (henceforth, $\alpha_s = 0.119$)

THEORETICAL UNCERTAINTIES

NLO PDF unc. vs NLO-NNLO shift vs NLO Cacciari-Houdeau (NNPDF2.1)

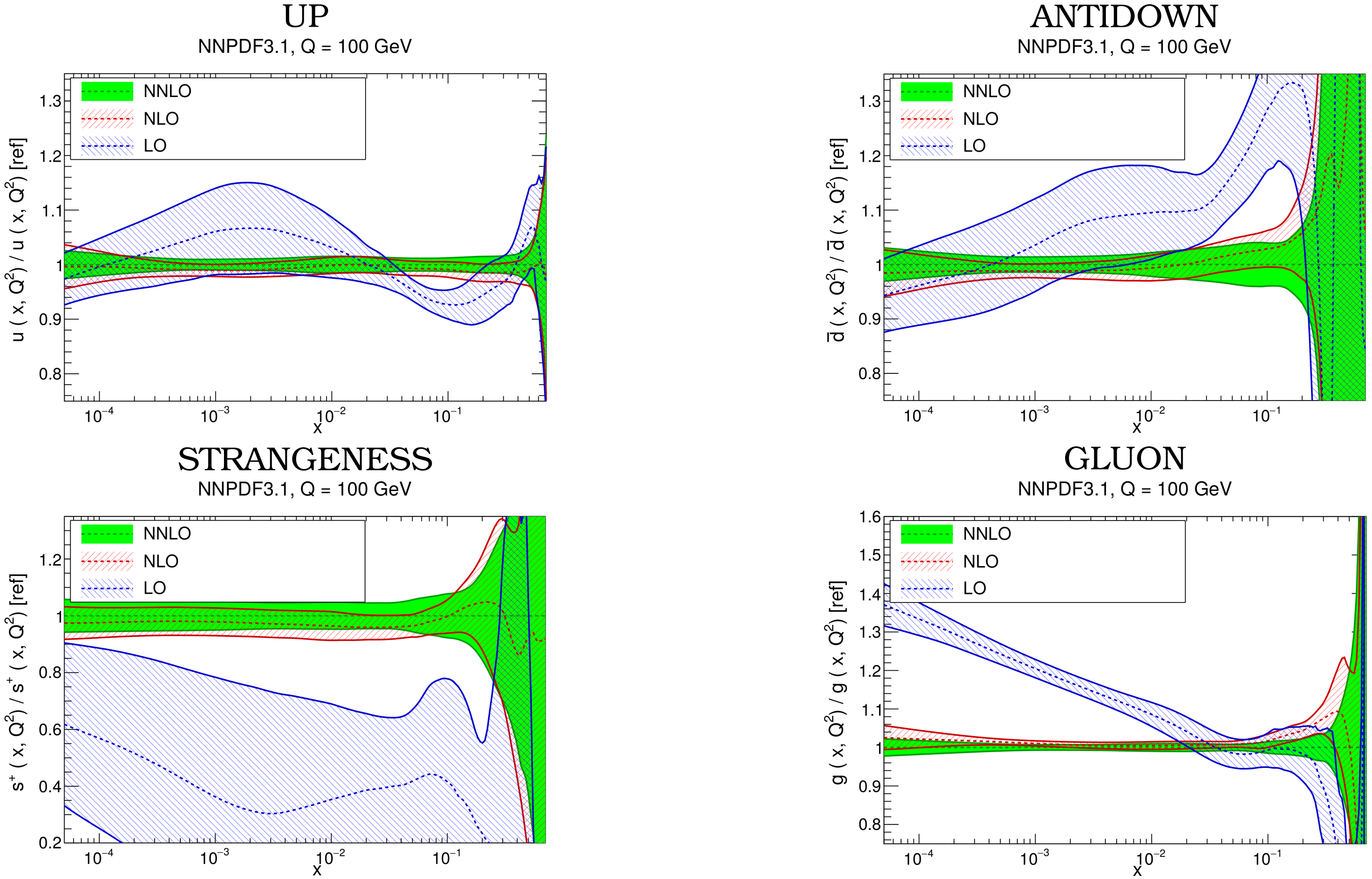


THE PERTURBATIVE EXPANSION FOR PDFs

OBJECTIONS:

- CAN WE REALLY TAKE PDFs AS SERIES IN α_s?
- IS IT MEANINGFUL TO EXTRAPOLATE BASED ON ONE SINGLE ORDER?
- LO PDFs ARE SPECIAL: POSITIVITY, GLUON, α_s

LO VS. NLO VS. NNLO PDFS



CACCIARI HOUDEAU FOR PDFS
PROS AND CONS

- CORRECTLY INCLUDES UNCERTAINTY DUE TO IMPERFECT FIT QUALITY LEADING TO SUBOPTIMAL CONVERGENCE
- POTENTIALLY UNSTABLE AT LOW PERTURBATIVE ORDERS DUE TO SPECIAL NATURE OF LO FITS

SCALE VARIATION IN PDF FITS

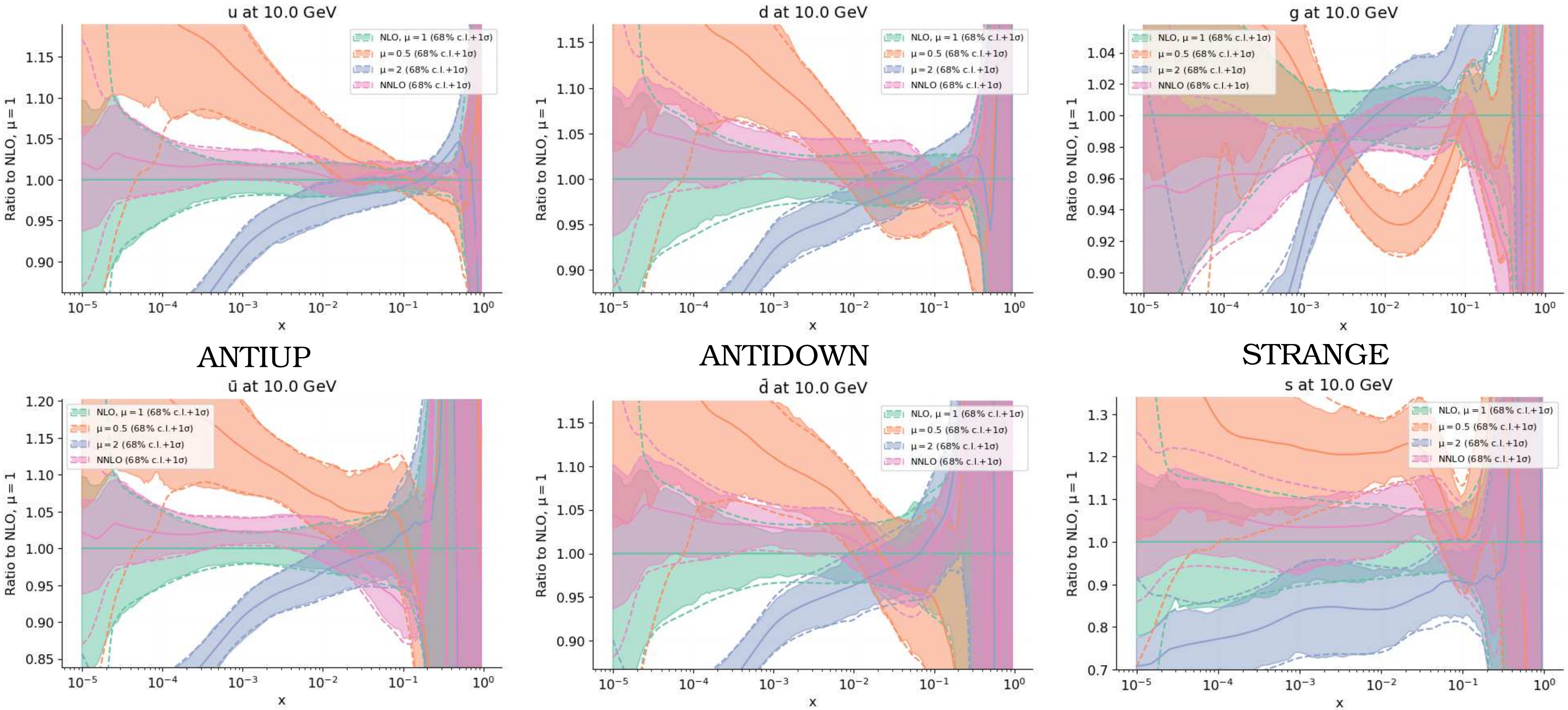
- REPEAT PDF DETERMINATION WITH DIFFERENT CHOICES OF RENORMALIZATION AND FACTORIZATION SCALE
- HOW SHOULD THE SCALE VARIATION CORRELATE BETWEEN DATAPOINTS?

NNPDF3.1: CENTRAL VS $2\times$ μ_R & μ_F

PRELIMINARY (INCONSISTENT DATASET)

UP — DOWN — GLUON

ANTIUP — ANTIDOWN — STRANGE



- AT MEDIUM & LARGE x, APPEARS TO BE IN BROAD AGREEMENT WITH ACTUAL NLO-NNLO SHIFT BUT SOMEWHAT OVERESTIMATES IT
- DISREGARD SMALL x REGION (INCONSISTENT DATASET)

NEW IDEA
PDF THEORY ERROR AS A FIT UNCERTAINTY

(Del Debbio, Ubiali, unpublished)

- PDFs ARE DETERMINED BY MAXIMIZING THE LIKELIHOOD

$$P = N \exp - \left(\frac{d-t}{2\sigma^2_{exp}} \right)$$

d, t ARE REALLY VECTORS AND $1/\sigma^2$ THE INVERSE COVARIANCE MATRIX

- CAN VIEW THIS AS THE PROBABILITY OF THE THEORY t BEING CORRECT GIVEN DATA d, WHICH BY BAYES IS

$$P(t|d) \propto P(d|t)P(t)$$

- IF THEORY WAS KNOWN EXACTLY, THEN $P(t) = \delta(t - t^{\mathrm{exact}})$
- IN ACTUAL FACT ONLY SOME PERTURBATIVE RESULT t_p IS EXACTLY KNOWN SO $t^{\mathrm{exact}} = t_p + \Delta_p$, WHERE Δ_p INCLUDES MHO
- ASSUMING Δ TO BE GAUSSIANLY DISTRIBUTED, WITH UNCERTAINTY σ_{th} AND INTEGRATING OUT

$$P = N \exp \left[\frac{d - t_p}{2\left(\sigma^2_{exp} + \sigma^2_{th}\right)} \right]$$

- THEORETICAL UNCERTAINTY ADDED IN QUADRATURE, PROPAGATES INTO PDF UNCERTAINTY UPON MINIMZATION
- SCALE VARIATION FOR EACH DATA POINT $\Rightarrow$ EIGENVECTOR OF COVARIANCE MATRIX (NUISANCE PARM.)

SUMMARY

- AT 1% LEVEL MUST INCLUDE THEORY UNCERTAINTIES ON PDFS
- NLO-NNLO SHIFT COMPARABLE TO NLO PDF UNCERTAINTIES
- CACCIARI-HOUDEAU & SCALE VARIATION PROMISING BUT PROBLEMATIC
- THEORY UNCERTAINTIES CAN BE INCLUDED IN OVERALL PDF UNCERTAINTY THROUGH BAYESIAN ARGUMENT

NO EFFECT THAT REQUIRES MORE THAN 10% ACCURACY IN MEASUREMENT IS WORTH INVESTIGATING

Walther Nernst

~~NO EFFECT THAT REQUIRES MORE THAN 10% ACCURACY IN MEASUREMENT IS WORTH INVESTIGATING~~

Walther Nernst

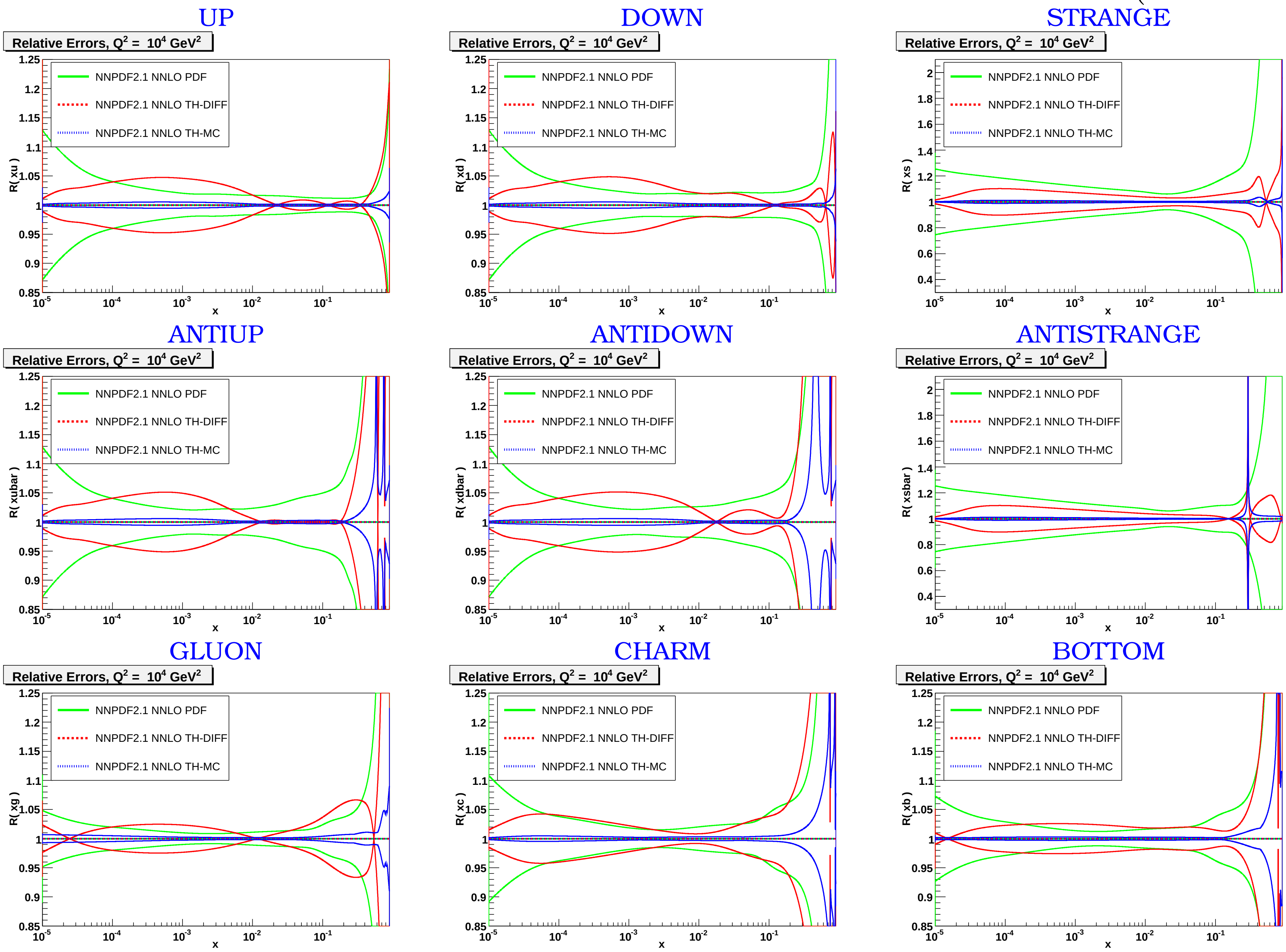
ACCURACY OF OBSERVATION IS THE EQUIVALENT OF ACCURACY OF THINKING

Wallace Stevens

EXTRAS

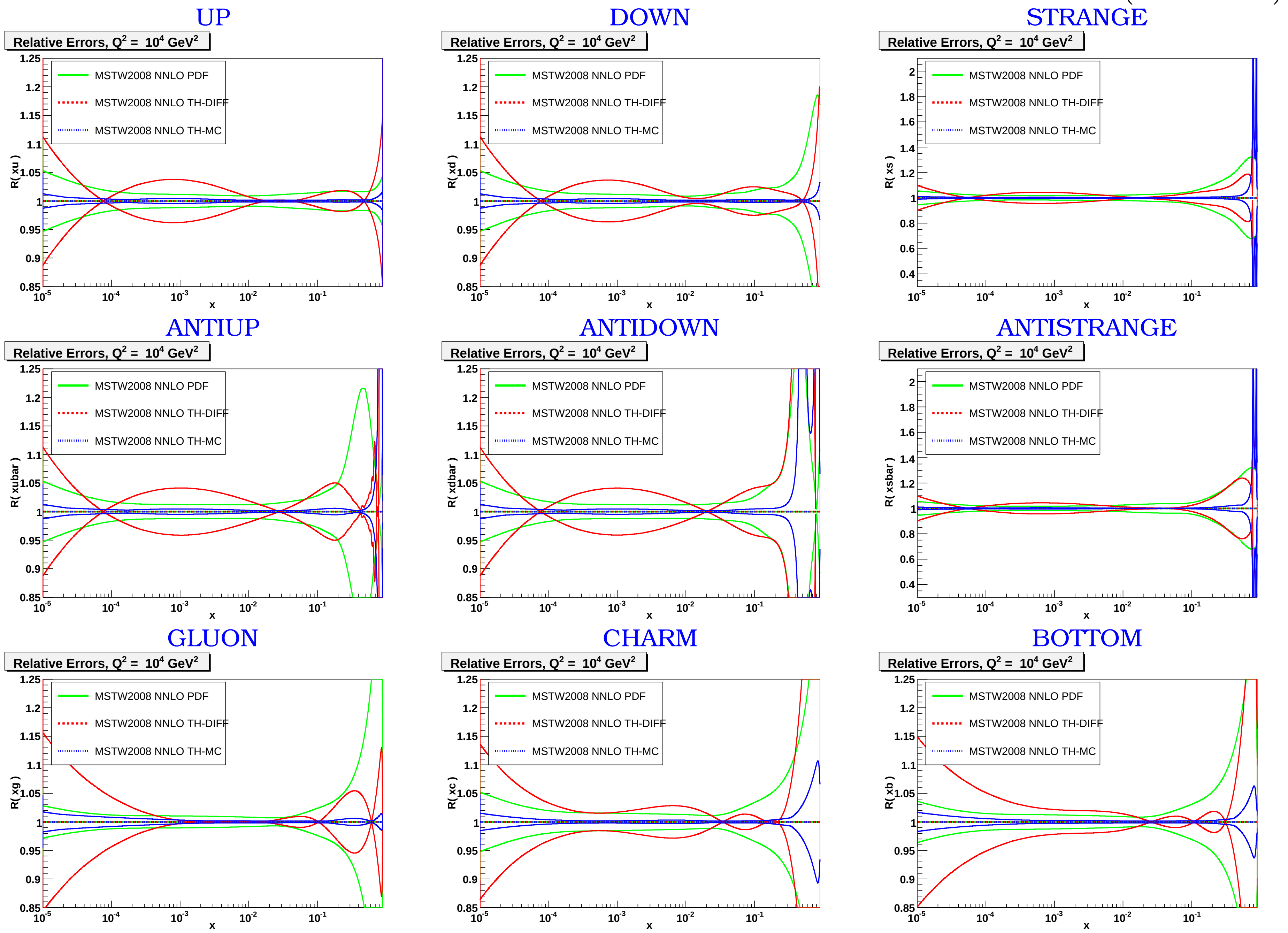
THEORETICAL UNCERTAINTIES

NNLO PDF unc. vs NLO-NNLO shift vs NNLO Cacciari-Houdeau (NNPDF2.1)



THEORETICAL UNCERTAINTIES

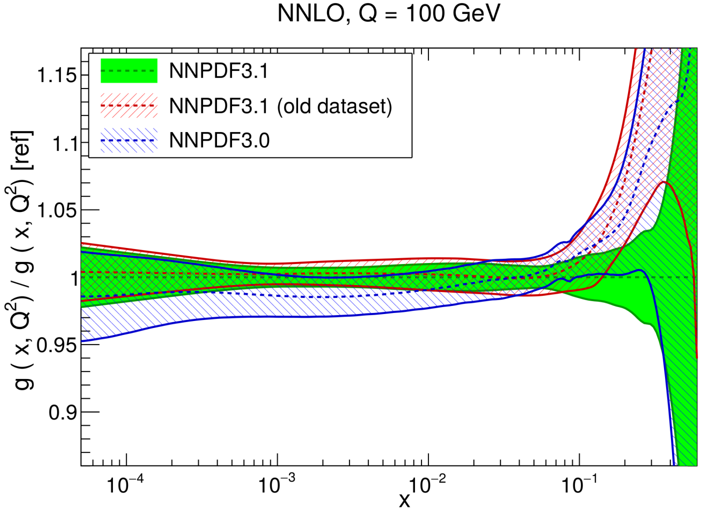
NNLO PDF unc. vs NLO-NNLO shift vs NNLO Cacciari-Houdeau (MSTW08)



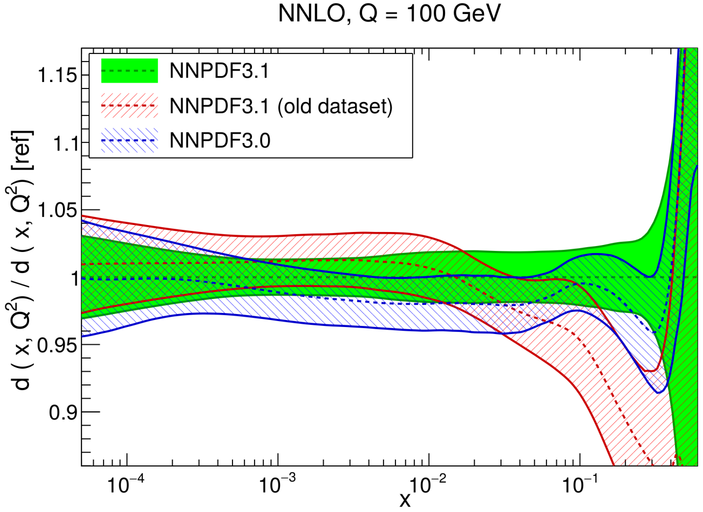
NNPDF3.1

- METHODOLOGY: INDEPENDENTLY PARAMETRIZED CHARM PDF
- LHC DATA : TOP RAPIDIDITY, Z p_T, LHCB & ATLAS W & Z PRODUCTION ADDED TO NNPDF3.0 DATASET

GLUON



DOWN



- FITTING CHARM ENHANCES GLUON IN GGF RANGE BY SEVERAL PERCENT
- NEW DATA (ESPECIALLY LHCB & ATLAS w z 2011) ENHANCE QUARKS IN VALENCE REGION

CLOSURE TESTS:

THE BASIC IDEA

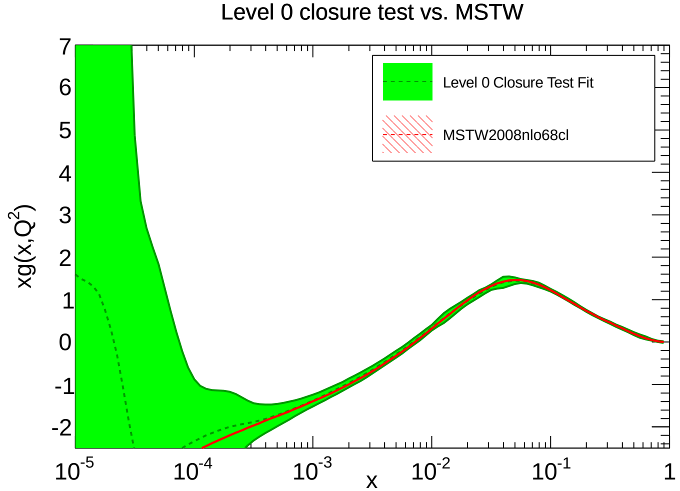
- ASSUME PDFS KNOWN: GENERATE FAKE EXPERIMENTAL DATA
- CAN DECIDE DATA UNCERTAINTY (ZERO, OR AS IN REAL DATA, OR . . .)
- FIT PDFS TO FAKE DATA
- CHECK WHETHER FIT REPRODUCES UNDERLYING "TRUTH":
  - CHECK WHETHER TRUE VALUE GAUSSIANLY DISTRIBUTED ABOUT FIT
  - CHECK WHETHER UNCERTAINTIES FAITHFUL
  -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

    (INDEP. OF METHODOLOGICAL DETAILS)

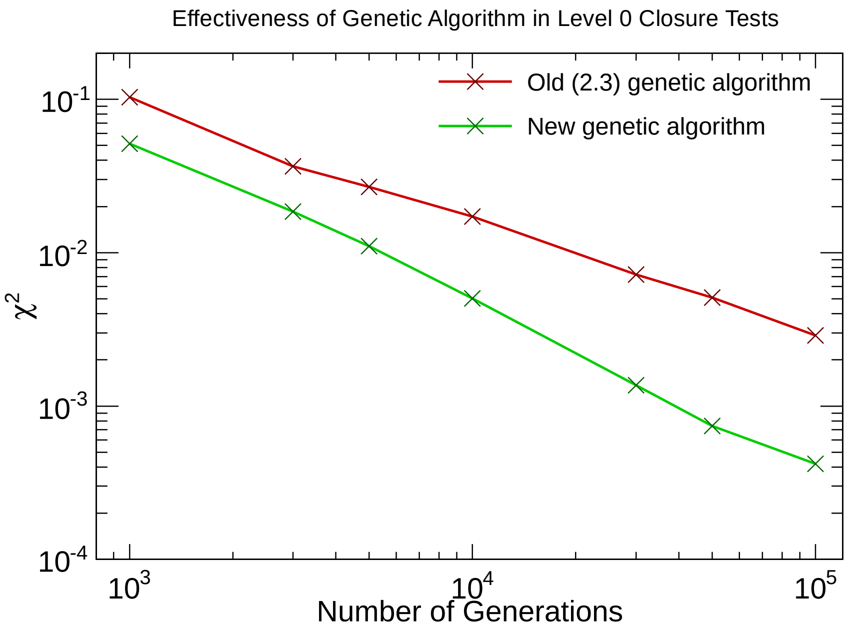
LEVEL-0 CLOSURE TESTS

- 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

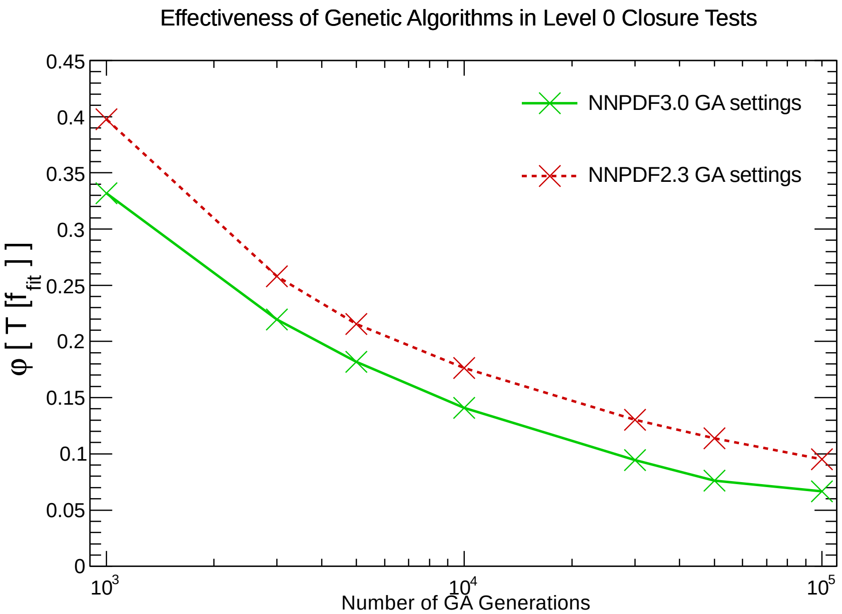
THE GLUON



χ^2 VS TRAINING LENGTH



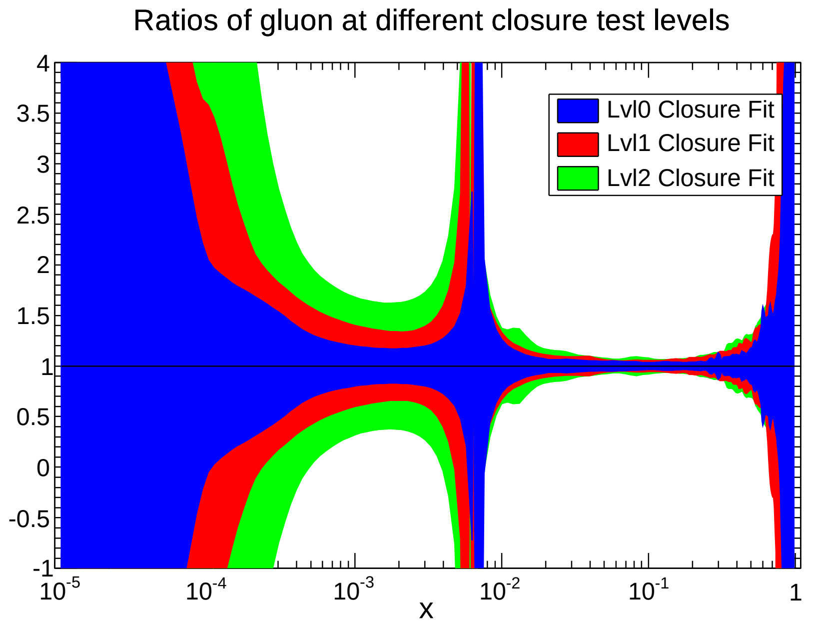
FRACTIONAL UNCERTAINTY VS TRAINING LENGTH

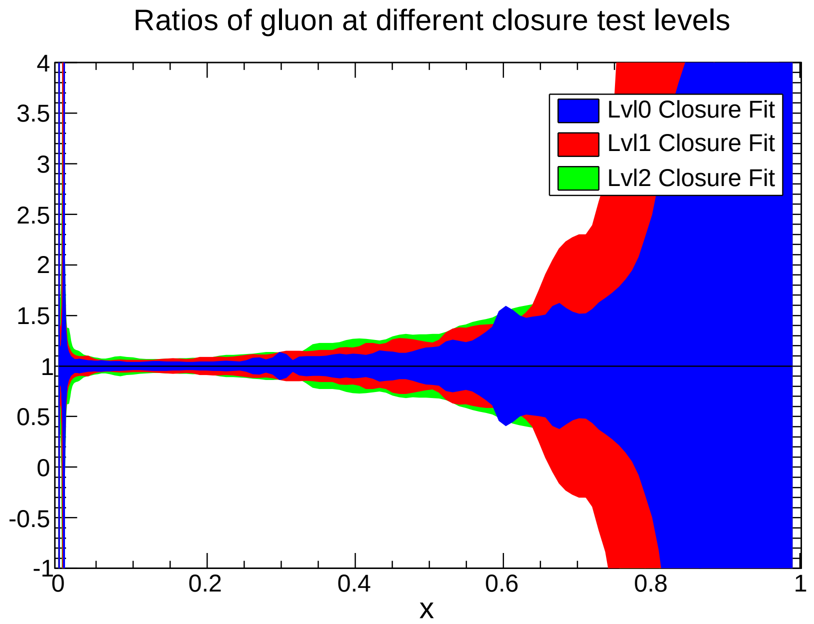


LEVEL-0, LEVEL-1 AND LEVEL-2

- 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

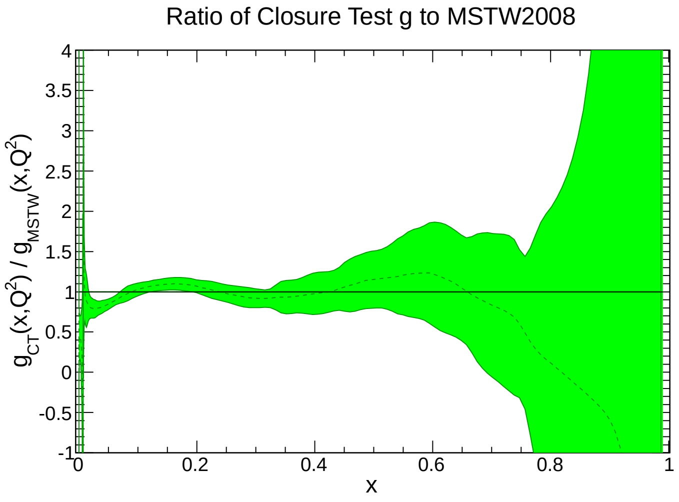
THE GLUON: LEVEL 0, LEVEL 1 AND LEVEL 2





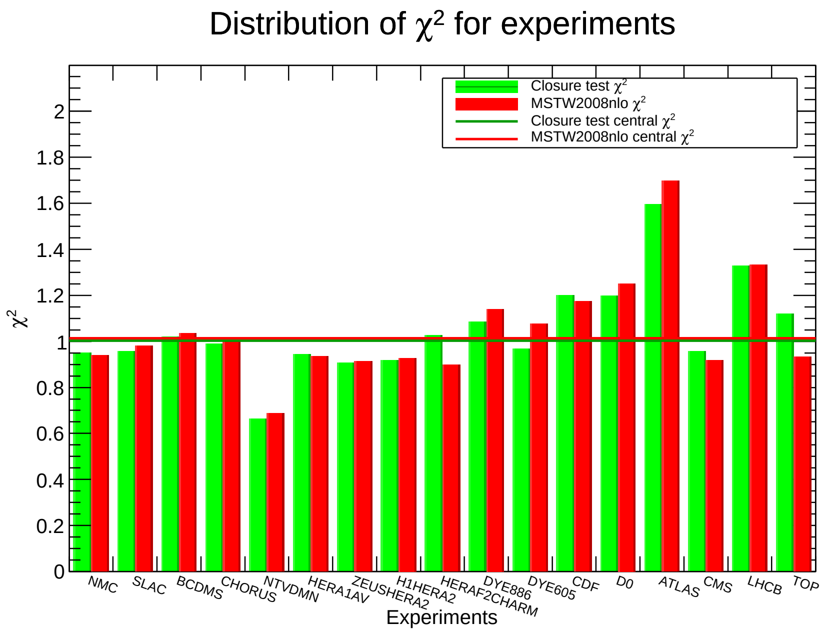
LEVEL-2: CENTRAL VALUES AND UNCERTAINTIES

THE GLUON: FITTED/"TRUE"

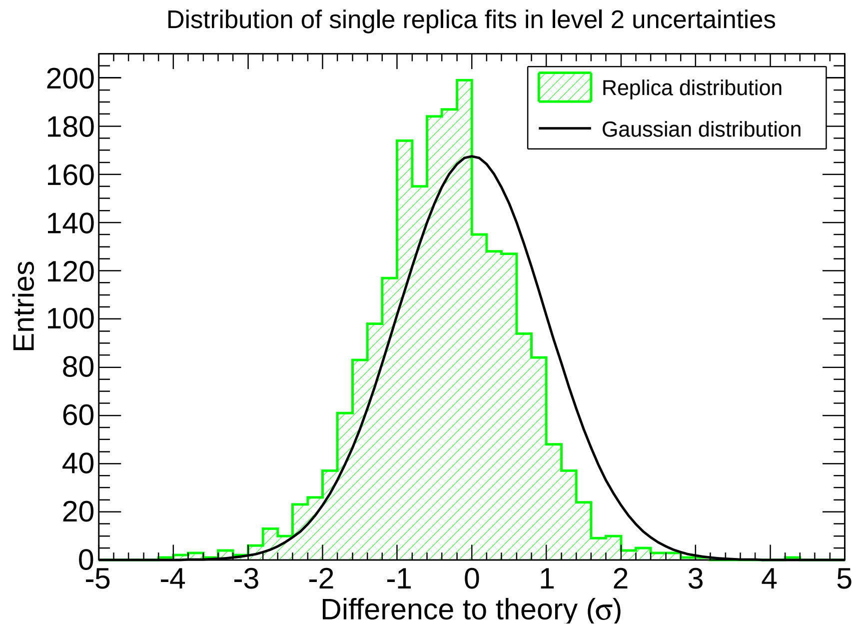


- CENTRAL VALUES: COMPARE FITTED VS. "TRUE" χ^2 BOTH FOR INDIVIDUAL EXPERIMENTS & TOTAL DATASET
  FOR TOTAL $\Delta\chi^2 = 0.001 \pm 0.003$
- UNCERTAINTIES: DISTRIBUTION OF DEVIATIONS BETWEEN FITTED AND "TRUE" PDFs
  SAMPLED AT 20 POINTS BETWEEN 10^{-5} AND 1
  FIND 0.699% FOR ONE-SIGMA, 0.948% FOR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

LEVEL-2 FITTED χ^2 VS "TRUE"



NORM. DISTRIBUTION OF DEVIATIONS



LEVEL-2 STABILITY TESTS

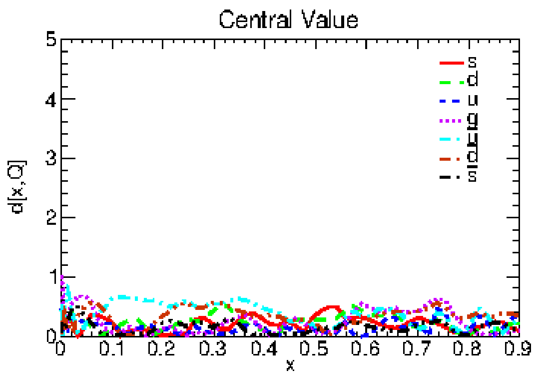
- CHANGE UNDERLYING PDF SET (CT10, NNPDF2.3)
- INCREASE MAXIMUM GA TRAINING LENGTH TO 80K
  TESTS EFFICIENCY OF CROSS-VALIDATION
- INCREASE NN ARCHITECTURE TO 2-20-15-1
  NUMBER OF FREE PARAMETRES INCREASE BY MORE THAT $10\times$
- CHANGE PDF PARAMETRIZATION BASIS
  OLD: ISOTRIPLET, $\bar{u}-\bar{d}$,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

STATISTICAL EQUIVALENCE!

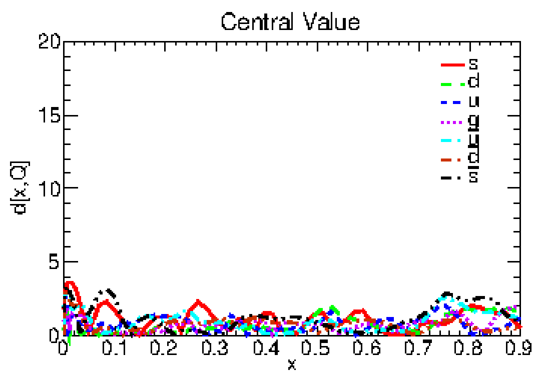
DISTANCES BETWEEN REF. AND NEW FIT:

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