PARTON DISTRIBUTIONS

FOR PRECISION PHYSICS

STEFANO FORTE

UNIVERSITÀ DI MILANO & INFN

LHCP19

PUEBLA, MAY 21, 2019

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

SUMMARY

THE CHALLENGES OF PRECISION PHYSICS

- W MASS DETERMINATION AND PDFs
- "TENSION" VS. UNCERTAINTIES: TOP PRODUCTION
- DATA VS. METHODOLOGY

α_s: THE PITFALLS OF USING PDFs IN PRECISION ANALYSIS

- PARAMETER SPACE AND PDF SPACE
- CORRELATED REPLICAS

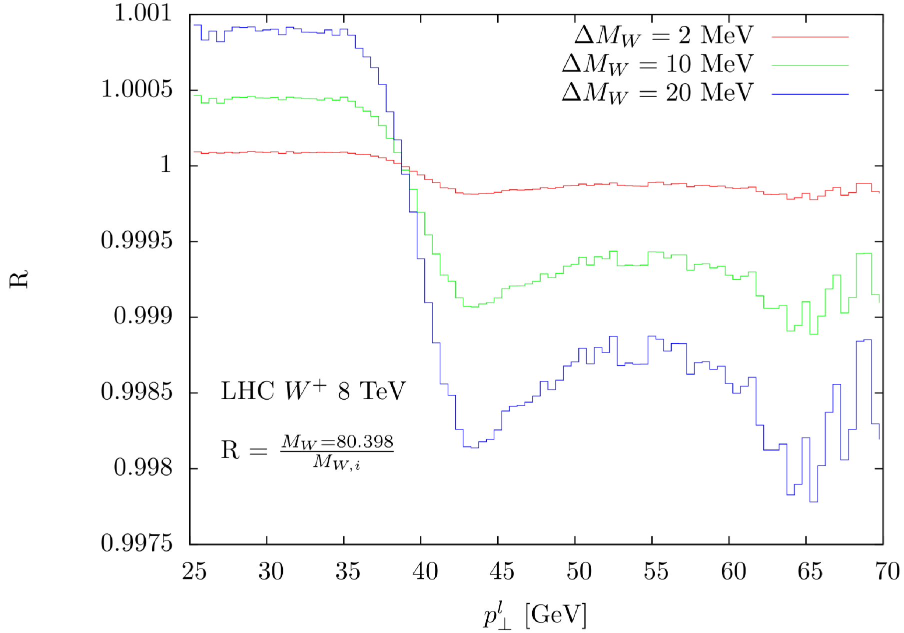
TOLERANCE: HOW ARE PDF UNCERTAINTIES DEFINED?

- $\Delta\chi^2$ AND FINITE-SIZE EFFECTS
- GAN REPLICA GENERATION

BETTER PDFs FOR PRECISION PHYSICS

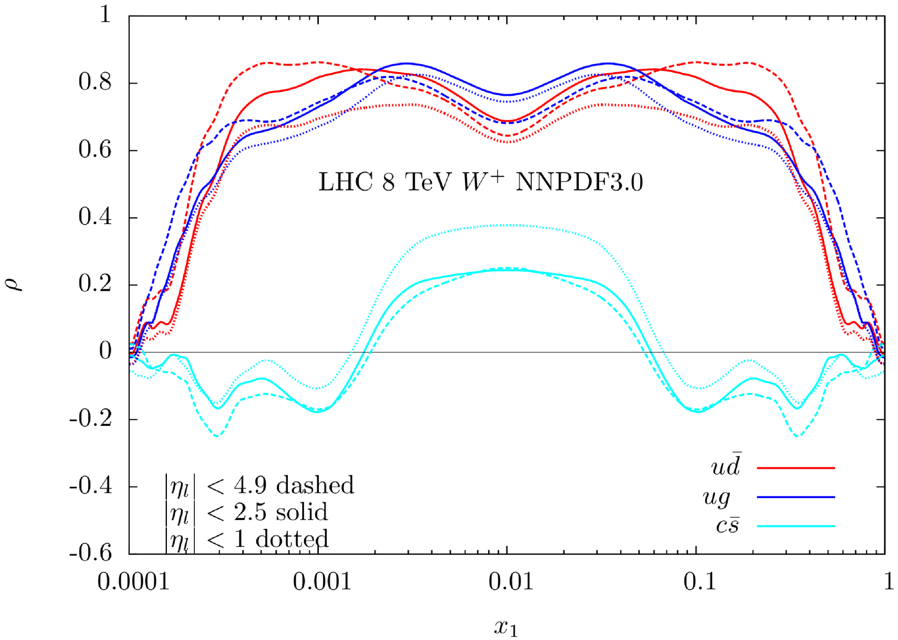
DETERMINING THE W MASS
THE TEMPLATE METHOD

p_T SHAPE VS. M_W



CORRELATION OF XSECT TO LUMI

($p_t^l = 40.5$ GEV, $p_T^W < 15$ GEV)

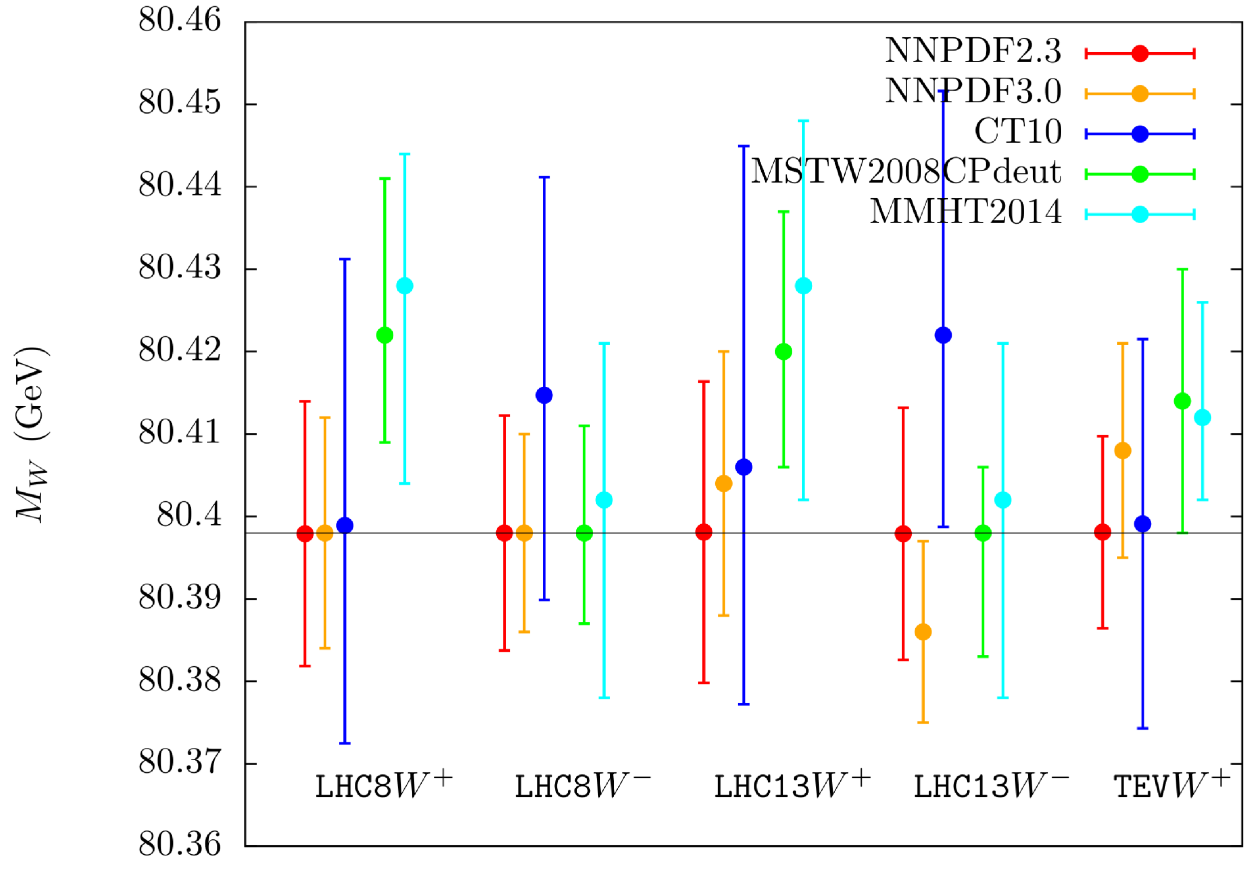


(Bozzi, Citelli, Vicini, 2015)

- TEMPLATE METHOD: W MASS EXTRACTED BY COMPARING OBSERVED SPECTRA TO THEORY: LEPTON PAIR TRANSVERSE MASS, LEPTON p_T^l
- SHAPE DEPENDS ON p_T^l: LARGER M_W, FASTER DROP AT HIGH p_T^l, LARGER XSECT AT SMALL p_T^l
- STRONG CORRELATION TO LEADING PARTON LUMIS ($u\bar{d}$ & $c\bar{s}$ FOR W^+) BUT ALSO TO NL LUMI (GLUON-INDUCED: ug)

DETERMINING THE W MASS

USING PDF4LHC15 GLOBAL SETS



(Bozzi, Citelli, Vicini, 2015), for CT confirmed by (Hussein, Isaacson, Huston, 2019)

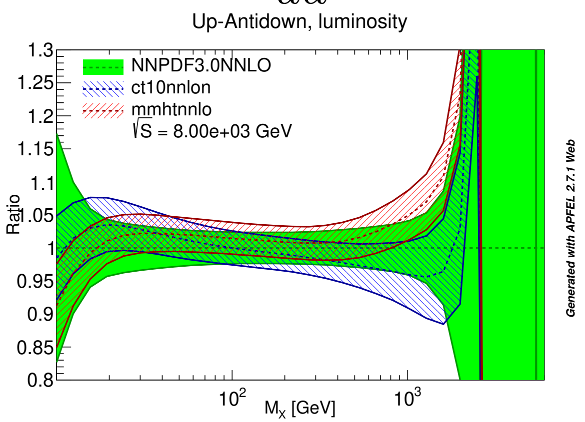
- STRONG DEPENDENCE ON PDF SET OF BOTH CENTRAL VALUE & UNCERTAINTY
- PECULIAR ASYMMETRY BETWEEN W^+ & W^-
- DIFFERENCE LARGE IN COMPARISON TO PDF UNCERTAINTY

WHAT'S GOING ON?

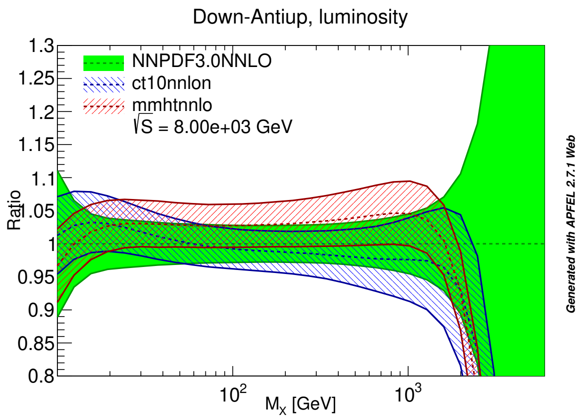
THE W MASS AND PDFs

LUMINOSITIES

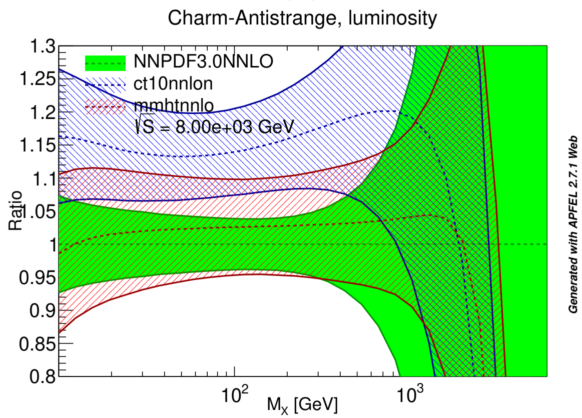
$u\bar{d}$



$d\bar{u}$



$c\bar{s}$



- FOR THE W p_T DISTRIBUTION THE HARD SCALE IS $M_X = \left(\sqrt{{p_T^W}^2 + m_W^2} + p_T\right)$
- DIFFERENT SETS HAVE SIGNIFICANTLY DIFFERENT M_X SLOPES $\Rightarrow$ DIFFERENT M_W
- UNCERTAINTIES IN $c\bar{s}$ LUMI SIGNIFICANTLY DIFFERENT

SMALL DIFFERENCES IN PDF AMPLIFIED TO LARGER DIFFERENCES IN SLOPE

- FLAVOR SEPARATION AFFECTED BY METHODOLOGY $\Rightarrow$ see plenary talk
- LHC DATA HELP $\Rightarrow$ see plenary talk

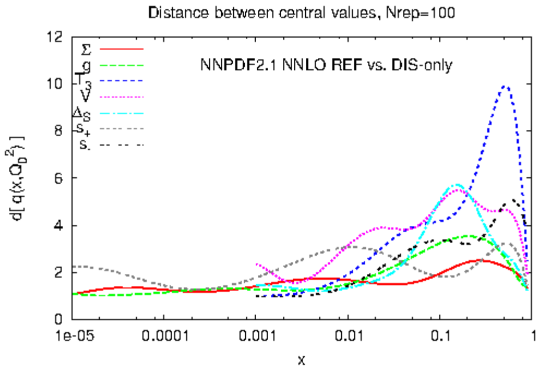
NEED BETTER PDFs!

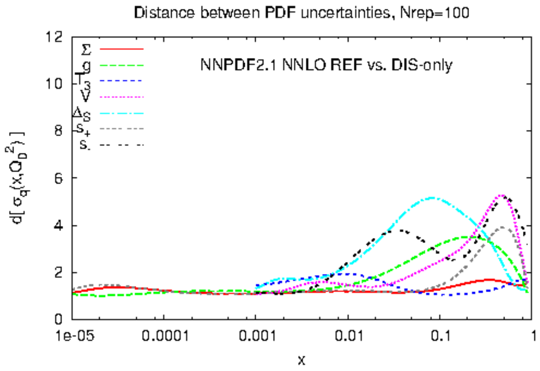
THE IMPACT OF LHC DATA

BEFORE LHC: PDFs MOSTLY DETERMINED BY DIS

NNPDF2.1 VS NNPDF2.1 DIS ONLY

DISTANCES (difference in units of st. dev.)

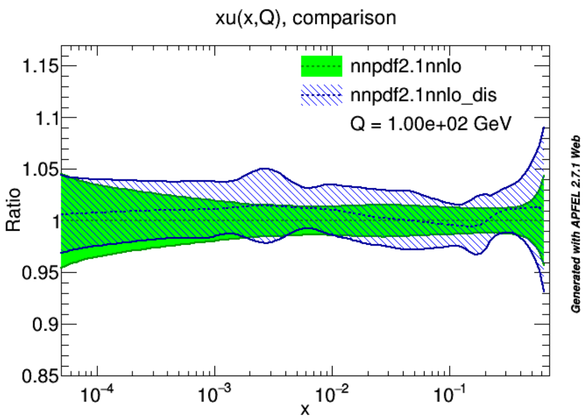




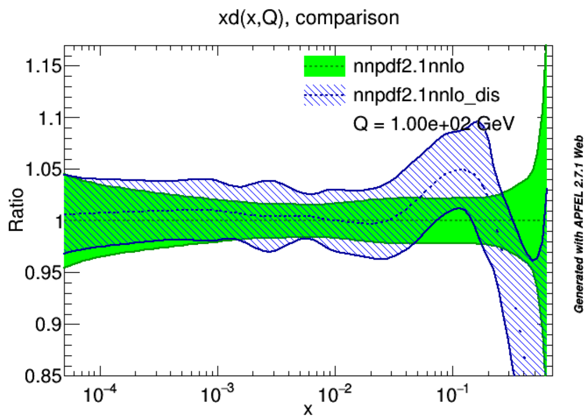
$d = 10 \Leftrightarrow$ one sigma difference

PDF COMPARISON

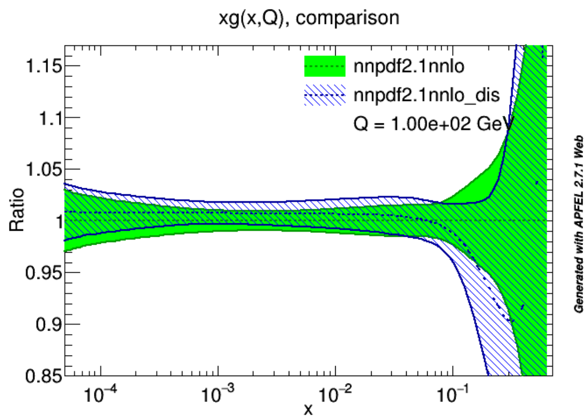
UP



DOWN



GLUON



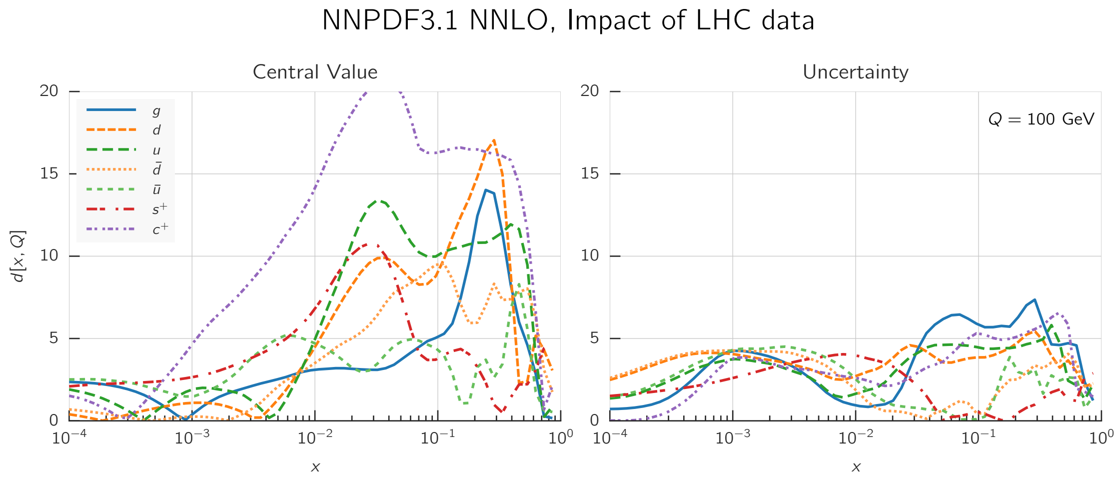
- ALL DIFFERENCES BELOW ONE SIGMA
- ONLY UP-DOWN SEPARATION SIGNIFICANTLY AFFECTED

THE IMPACT OF LHC DATA

NOW: PDFs LARGELY DETERMINED BY LHC DATA

NNPDF3.1 VS NNPDF3.1 NO LHC

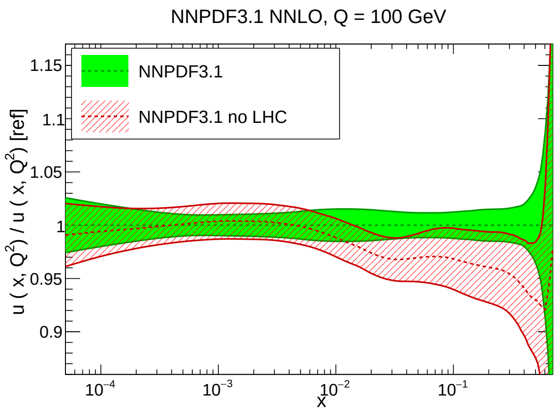
DISTANCES (difference in units of st. dev.)



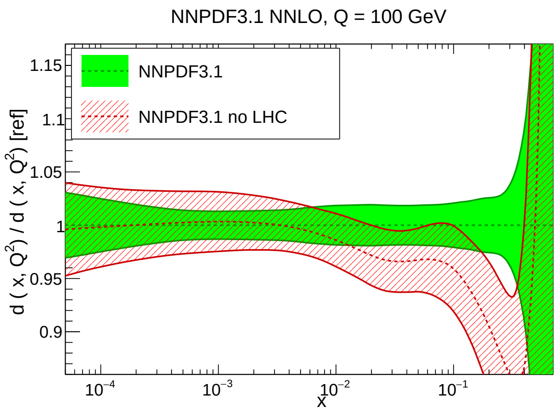
$d = 10 \Leftrightarrow$ one sigma difference

PDF COMPARISON

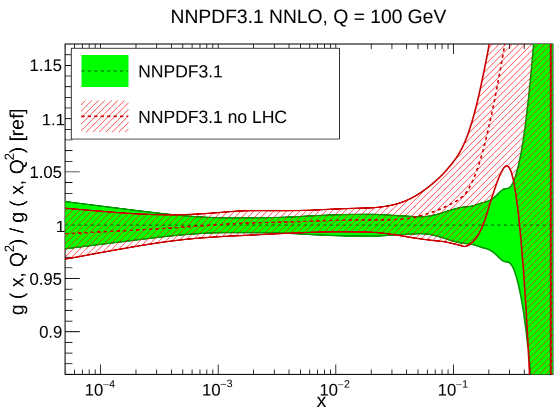
UP



DOWN



GLUON

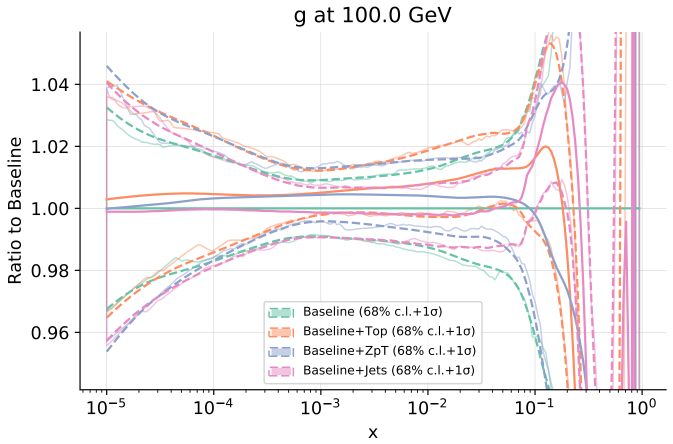


- MANY PDFs CHANGE BY MORE THAN ONE SIGMA
- BOTH FLAVOR SEPARATION & GLUON SIGNIFICANTLY AFFECTED

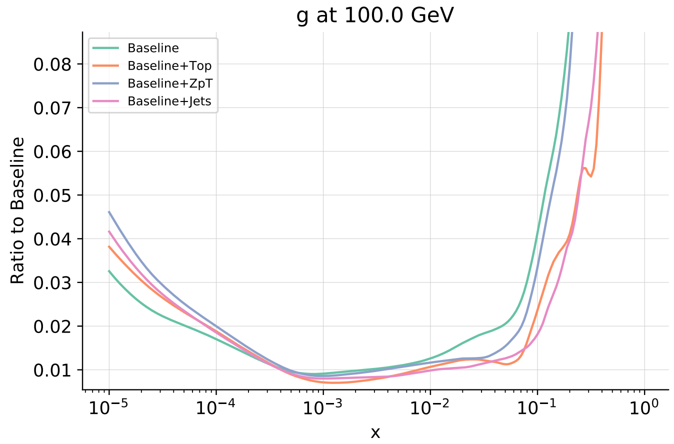
THE IMPACT OF LHC DATA
CONSISTENCY OF DIFFERENT OBSERVABLES
THE GLUON

- BEFORE LHC $\Rightarrow$ DIS SCALING VIOLATIONS, TEV JETS AT LARGE X
- AFTER LHC $\Rightarrow$ JETS; $Z\ p_t$, TOP

CENTRAL VALUE



UNCERTAINTY



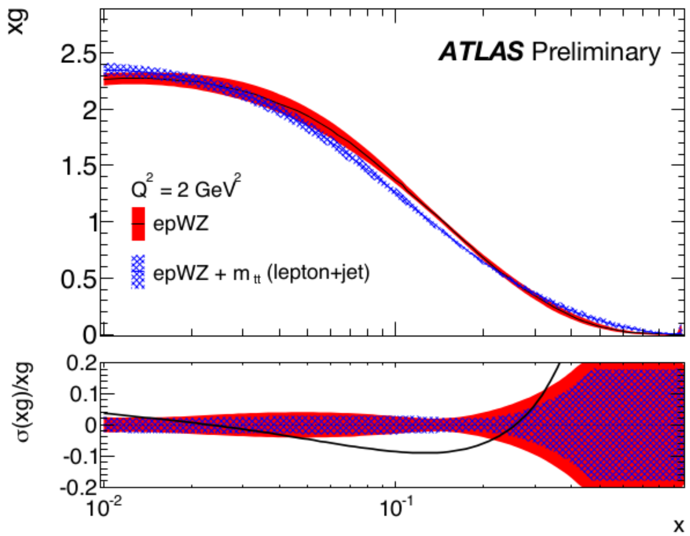
- TOP HAS LARGEST IMPACT, FOLLOWED BY JETS
- ALL LHC DATA PULL CENTRAL VALUE IN SAME DIRECTION!

CONSISTENCY OF DIFFERENT OBSERVABLES

TOP PRODUCTION AND THE GLUON

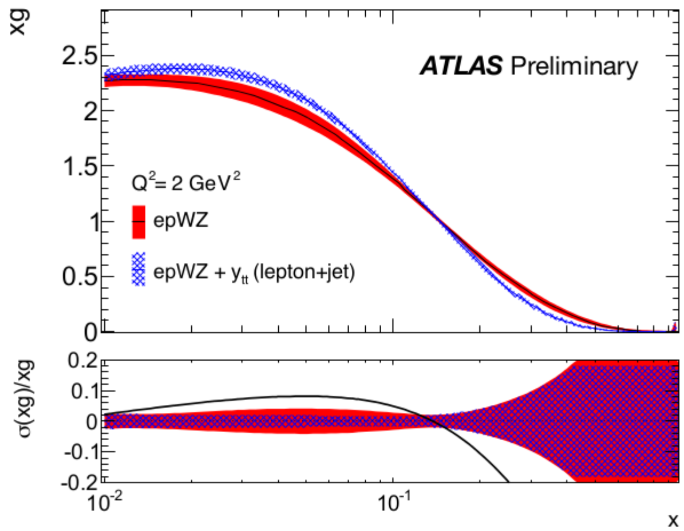
INCLUSION OF ATLAS TOP DATA IN HERA+TOP FIT (XFITTER)

HQ PAIR RAPIDITY DISTN



(a)

INVARIANT MASS DISTN.



(a)

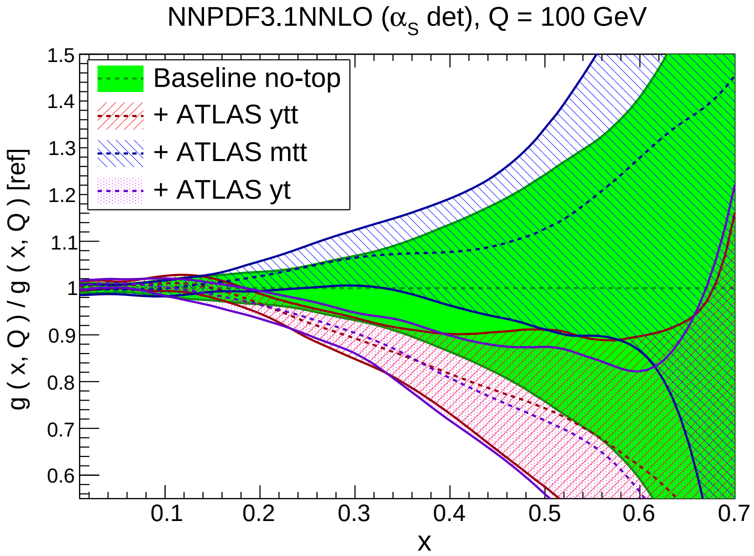
INCONSISTENCY?

CONSISTENCY OF DIFFERENT OBSERVABLES

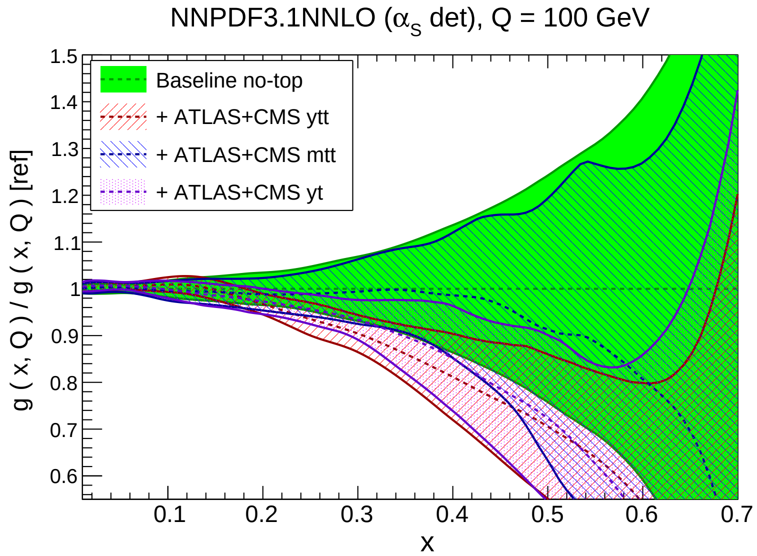
TOP PRODUCTION AND THE GLUON

INCLUSION OF ATLAS TOP DATA IN NNPDF3.1-LIKE FIT

ATLAS ONLY



ATLAS+CMS



- FOR ATLAS m_{tt} & y DISTRIBUTIONS PULL IN OPPOSITE DIRECTION
  $\Rightarrow$ COMPATIBLE WITHIN UNCERTAINTIES
- m_{tt} HAS MUCH LESS PULL
- FOR CMS, BOTH m_{tt} & y PULL IN THE SAME DIRECTION

CONSISTENCY!

LESSONS:

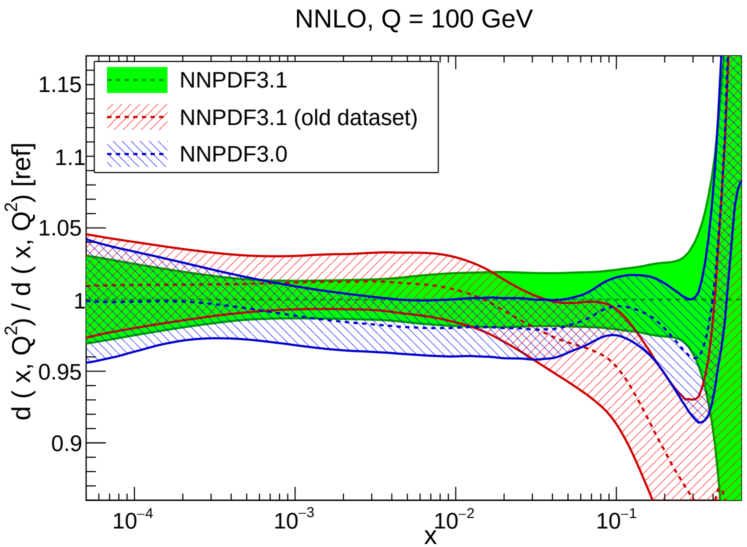
- BEWARE OF XFITTER HERA+X FITS
- IN A GLOBAL FIT, DIFFERENT DATA ALWAYS PULL IN DIFFERENT DIRECTIONS!

DATA VS. METHODOLOGY

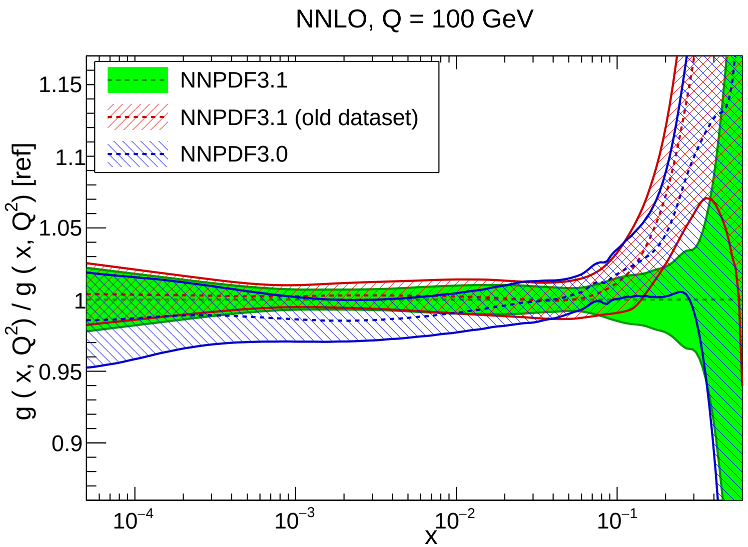
- EVEN WITH LHC DATA MAJOR METHODOLOGICAL CHOICES ⇒ SIGNIFICANT IMPACT
- EXAMPLE: HEAVY QUARKS INDEP. PARAMETRIZED ⇒ see plenary talk
- NNPDF3.1 VS NNPDF3.0: DATA AND METHODOLOGY HAVE SIMILAR IMPACT

NNPDF3.0 VS. NNPDF3.1 VS. NNPDF3.1 W/ NNPDF3.0 DATASET

DOWN



GLUON

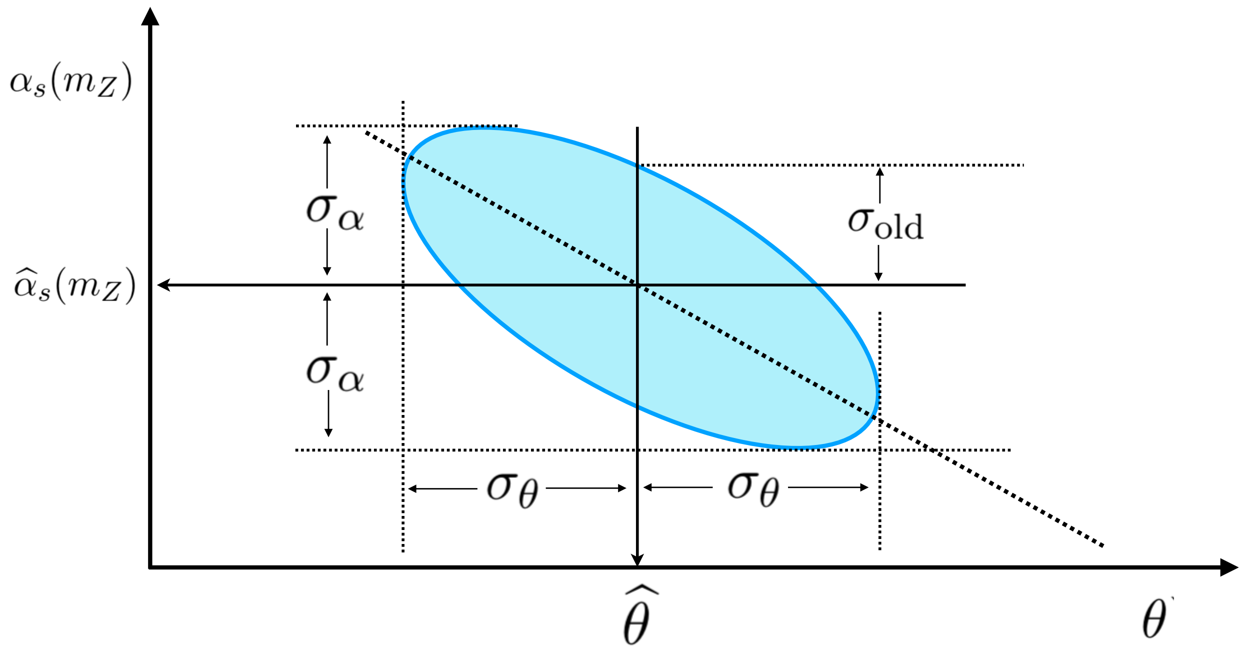


LHC DATA+ METHODOLOGICAL IMPROVEMENTS ⇒ BETTER PDFS

SM parameters from PDF-dependent observables

α_s DETERMINATION

WHAT'S THE PROBLEM?

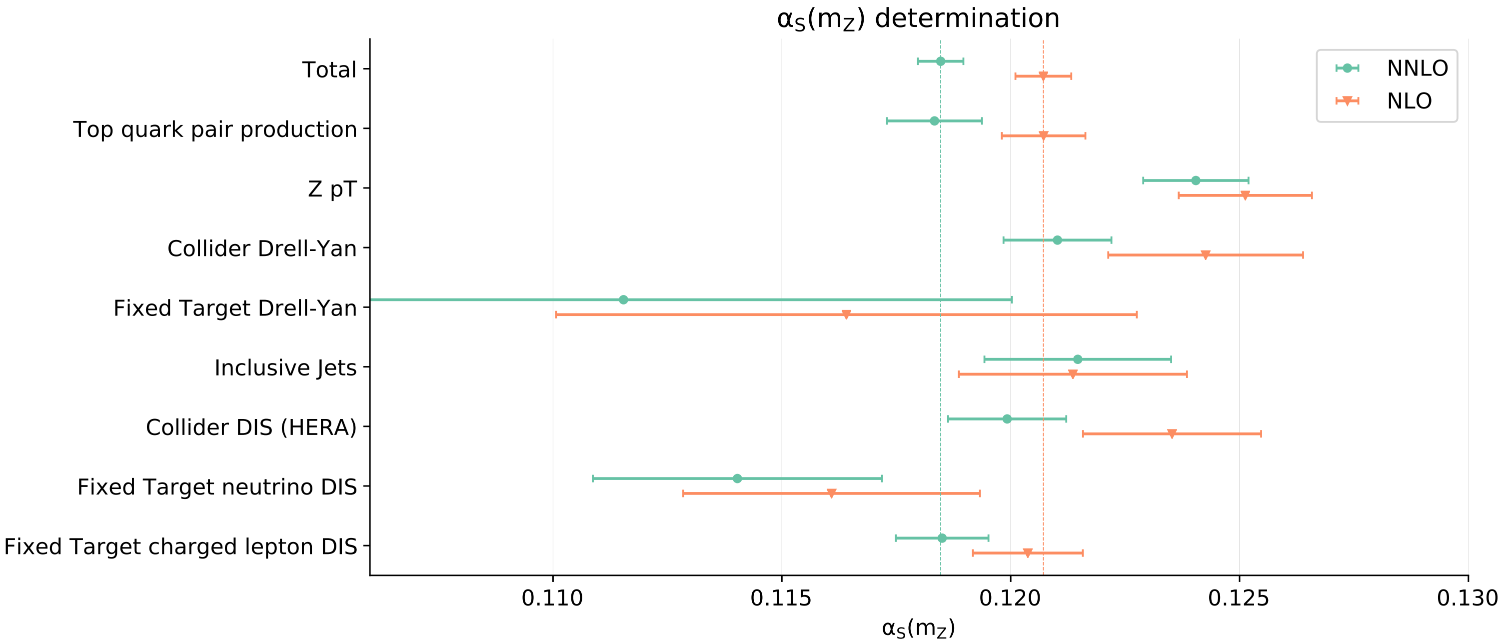


- MINIMUM DETERMINED ALONG THE "BEST PDF" LINE $\Rightarrow \sigma_{\rm old}$ FOR HIGHLY CORRELATED VARIABLES & UNEQUAL SEMIAXES, MAY UNDERESTIMATE ONE-σ ERROR $\Rightarrow \sigma_\alpha$

NEED SIMULTANEOUS MINIMIZATION IN (PDF, α_s) SPACE!

α_s FROM A GLOBAL FIT

PULLS FROM DATA SUBSETS



PULLS DON'T ADD TO ZERO?!

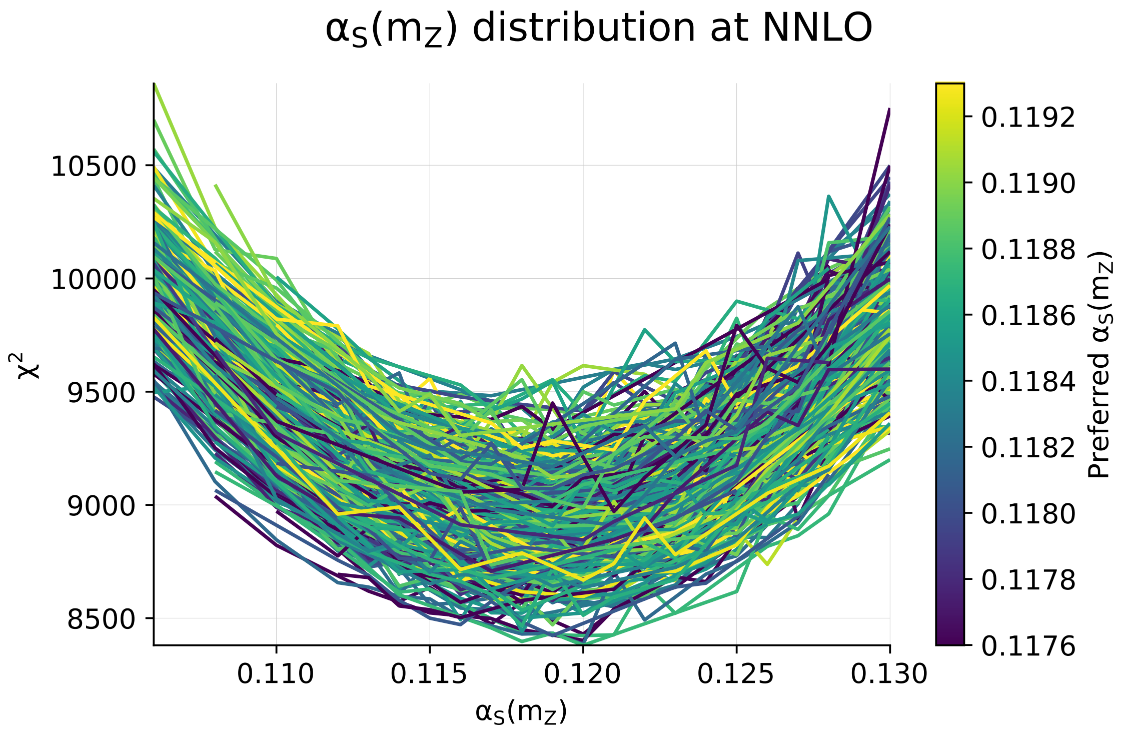
- PARTIAL VALUES ARE NOT PARTIAL BEST-FITS
- PDF SPACE HUGE $\Rightarrow$ MINIMUM AT DIFFERENT α_s VALUE WHEN INCLUDING NEW DATA, AGREEMENT WITH OTHER DATA ESSENTIALLY UNAFFECTED

$\Rightarrow$ CANNOT DETERMINE α_s WITHOUT ALSO DETERMINING THE PDF

THE CORRELATED REPLICA METHOD
NNPDF3.1 (2018)

- NNPDF METHOD $\Rightarrow$ EACH PDF REPLICA FITTED BY GA TO DATA REPLICA
- IDEALLY PERFORM GENETIC MINIMIZATION IN (PDF, α_s) SPACE
- PROBLEM THEORY PREDICTION $\Leftrightarrow$ PRECOMPUTED GRIDS DEPEND ON α_s $\Rightarrow$ DIFFICULT TO TREAT AS CONTINUOUS PARAMETER
- SOLUTION DETERMINE BEST-FIT PDF REPLICA TO EACH DATA REPLICA FOR SEVERAL (DISCRETE) α_s VALUES: C-REPLICA
  - EACH C-REPLICA $\Rightarrow$ χ^2 PROFILE $\Rightarrow$ α_s VALUE
  -

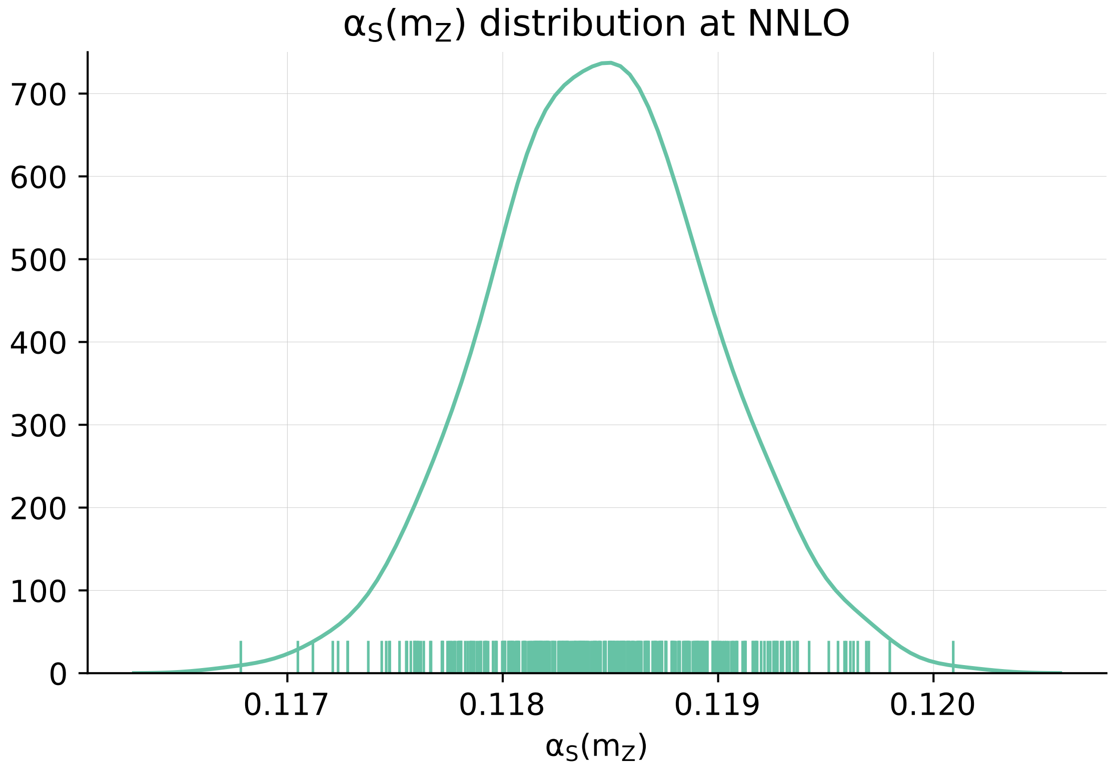
ENSEMBLE OF PARABOLAS



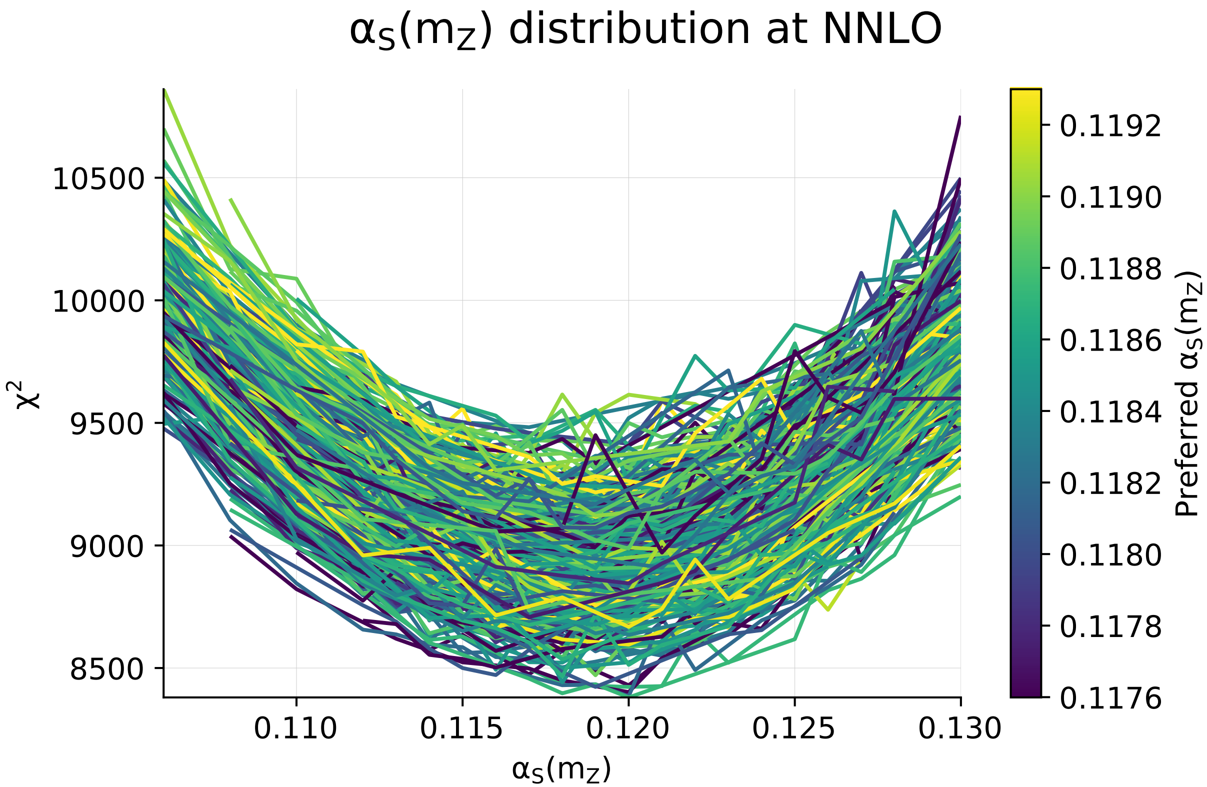
THE CORRELATED REPLICA METHOD
NNPDF3.1 (2018)

- NNPDF METHOD $\Rightarrow$ EACH PDF REPLICA FITTED BY GA TO DATA REPLICA
- IDEALLY PERFORM GENETIC MINIMIZATION IN (PDF, α_s) SPACE
- PROBLEM THEORY PREDICTION $\Leftrightarrow$ PRECOMPUTED GRIDS DEPEND ON α_s $\Rightarrow$ DIFFICULT TO TREAT AS CONTINUOUS PARAMETER
- SOLUTION DETERMINE BEST-FIT PDF REPLICA TO EACH DATA REPLICA FOR SEVERAL (DISCRETE) α_s VALUES:
  - EACH C-REPLICA $\Rightarrow$ χ^2 PROFILE $\Rightarrow$ α_s VALUE
  - EACH C-REPLICA $\Rightarrow$ BEST-FIT α_sREPLICA

ENSEMBLE OF α_s VALUES



α_s FROM CORRELATED REPLICAS



- NNPDF3.1 DATASET (ONLY NNLO JET DATA) $\Rightarrow$ 3979 DATAPOINTS
- 400 C-REPLICAS FOR 21 α_s VALUES: $\alpha_s(M_z) = 0.106, 0.108, 0.102, 0.112, 0.113, 0.114, 0.115, 0.116, 0.117, 0.118, 0.119, 0.120, 0.121, 0.122, 0.123, 0.124, 0.125, 0.126, 0.127, 0.128, 0.130$
- EXPERIMENTAL UNCERTAINTY $\Leftrightarrow$ STANDARD DEVIATION OVER REPLICA SAMPLE

THE RESULT

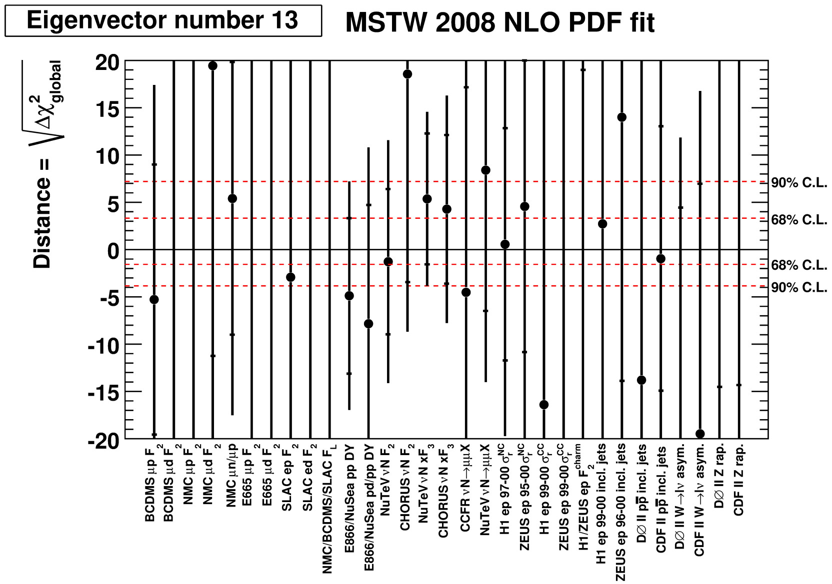
$$\alpha_s^{\mathrm{NNLO}}(M_Z) = 0.11845 \pm 0.00052^{\mathrm{exp}} \ (0.4\%)$$

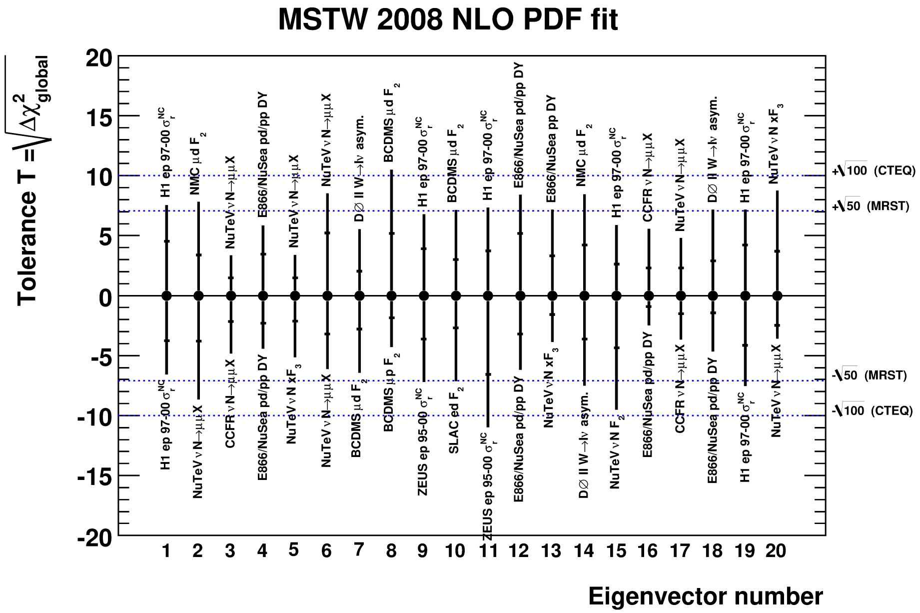
THE MEANING OF PDF UNCERTAINTIES

PDF UNCERTAINTIES: TOLERANCE (MMHT-CT)

GLOBAL MSTW TOLERANCE

MSTW TOLERANCE PLOT FOR 13TH EIGENVEC.





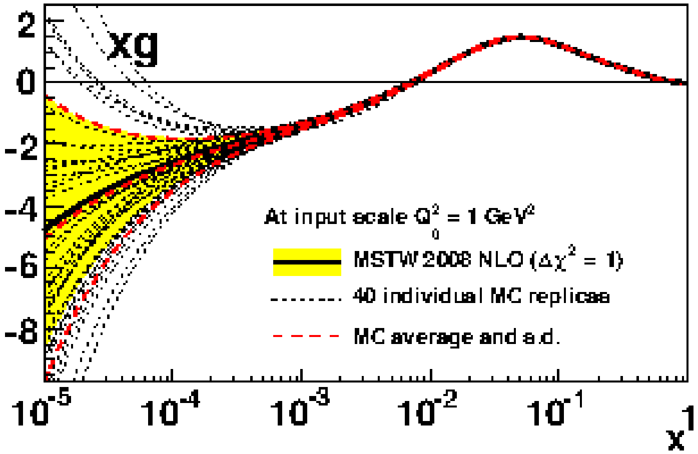
- (MSTW/MMHT) FOR EACH EIGENVECTOR IN PARAMETER SPACE DETERMINE CONFIDENCE LIMIT FOR THE DISTRIBUTION OF BEST-FITS OF EACH EXPERIMENT
- RESCALE $\Delta\chi^2 = T$ INTERVAL SUCH THAT CORRECT CONFIDENCE INTERVALS ARE REPRODUCED
- SIMILAR PROCEDURE ADOPTED BY CTEQ

WHAT ABOUT NNPDF?

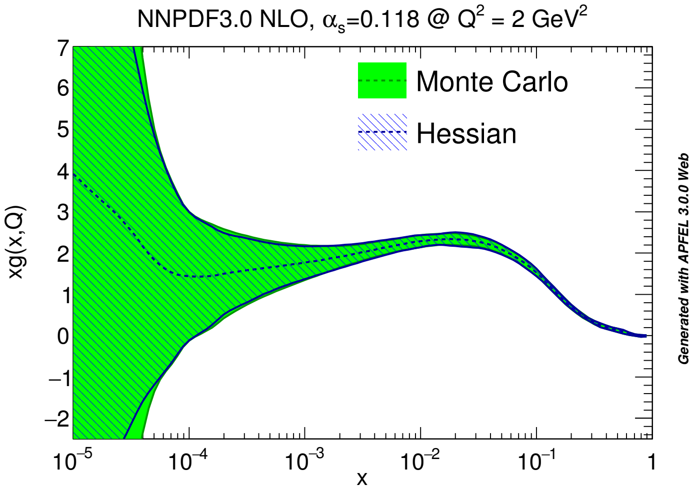
MC ⇔ HESSIAN

TWO DIFFERENT REPRESENTATIONS OF PDF UNCERTAINTIES

- TO CONVERT HESSIAN INTO MONTECARLO GENERATE MULTIGAUSSIAN REPLICAS IN PARAMETER SPACE
- ACCURATE WHEN NUMBER OF REPLICAS SIMILAR TO THAT WHICH REPRODUCES DATA



(Thorne, Watt, 2012)



(Carrazza, SF, Kassabov, Rojo, 2015)

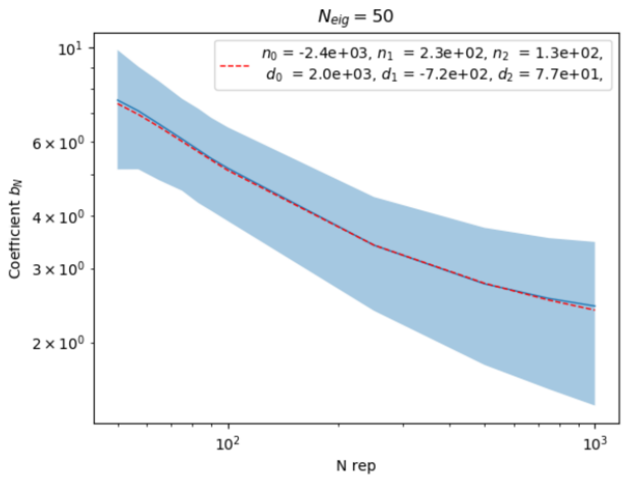
- TO CONVERT MONTE CARLO INTO HESSIAN, SAMPLE THE REPLICAS $f_i(x)$ AT A DISCRETE SET OF POINTS & CONSTRUCT THE ENSUING COVARIANCE MATRIX
- EIGENVECTORS OF THE COVARIANCE MATRIX AS A BASIS IN THE VECTOR SPACE SPANNED BY THE REPLICAS BY SINGULAR-VALUE DECOMPOSITION
- NUMBER OF DOMINANT EIGENVECTORS SIMILAR TO NUMBER OF REPLICAS ⇒ ACCURATE REPRESENTATION

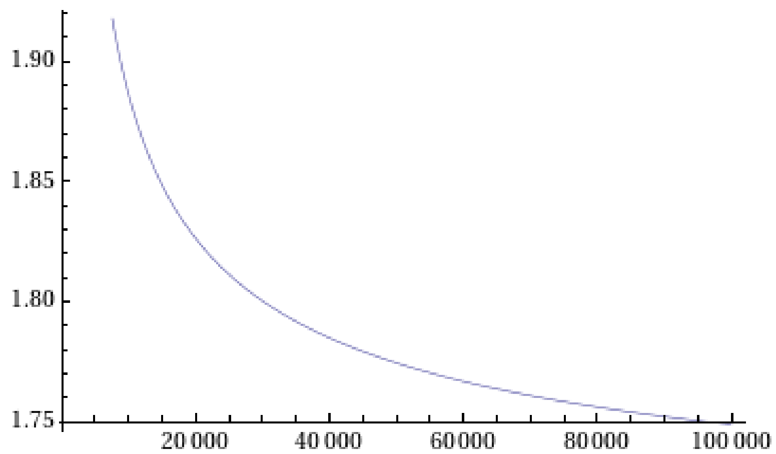
WHAT IS THE NNPDF "TOLERANCE"?

- PERFORM HESSIAN CONVERSION OF NNLO NNPDF3.1 PDFS
  50 OR 100 EIGENVECTORS
- DETERMINE χ^2 ALONG EACH EIGENVECTOR DIRECTION
- FIT A QUARTIC POLYNOMIAL
- STUDY DEPENDENCE ON
  NONGAUSSIANITY, NUMBER OF REPLICAS, NUMBER OF EIGENVECTORS,...

FINITE-SIZE EFFECTS

$\Delta\chi^2 = T^2$ VS NUMBER OF REPLICAS



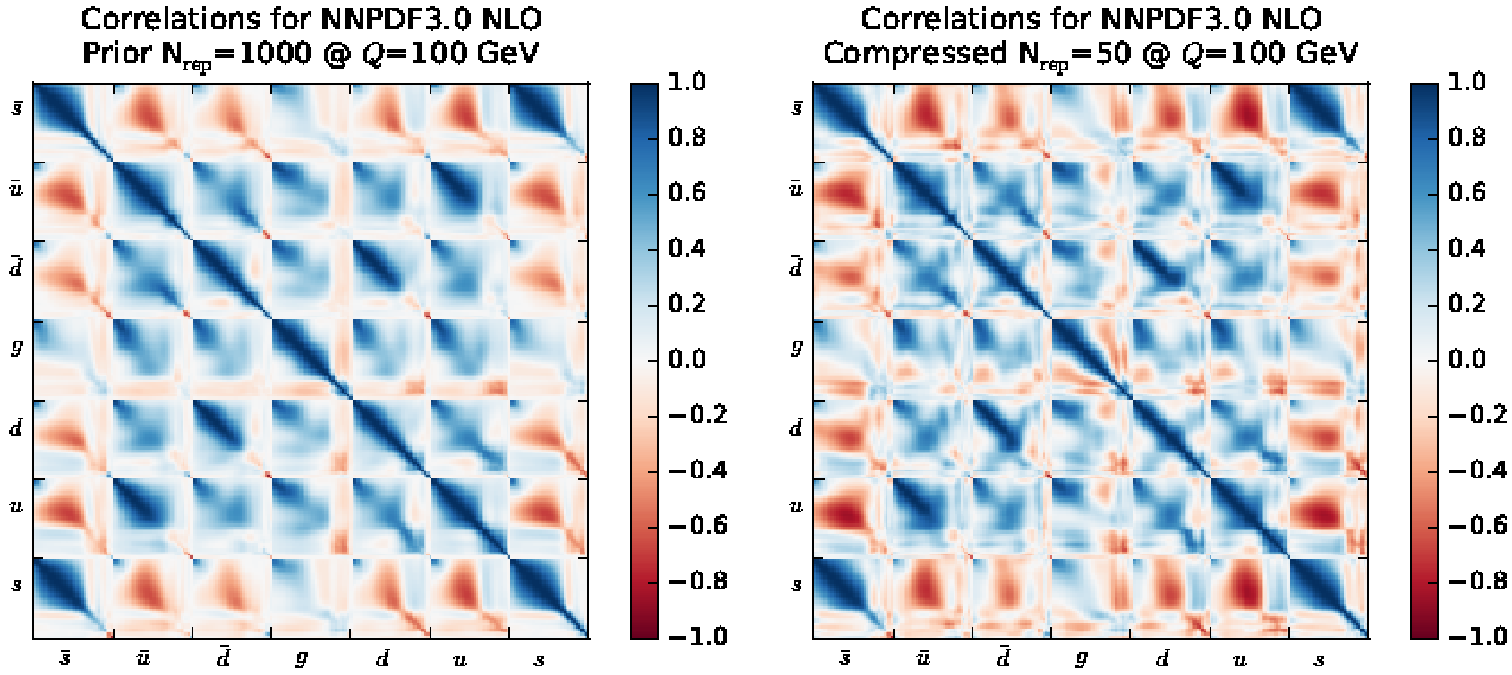


(Talon, MS thesis, 2019)

- NO SIGNIFICANT NONGAUSSIANITIY, DEVIATION FROM PARABOLIC,...
- SIGNIFICANT DEPENDENCE ON NUMBER OF REPLICAS
- ASYMPTOTIC TOLERANCE $T = 1.3 \pm 0.3$; $\Delta\chi^2 = 1.7 \pm 0.7$
- FOR $N_{\rm rep} = 100$, $T = 2.3$, EVEN FOR $N_{\rm rep} = 1000$, $T = 1.6$

DO WE HAVE TO FIT 10000 REPLICAS? DO WE HAVE TO USE 10000 REPLICAS?

SOLVING THE PROBLEM....
MONTECARLO COMPRESSION



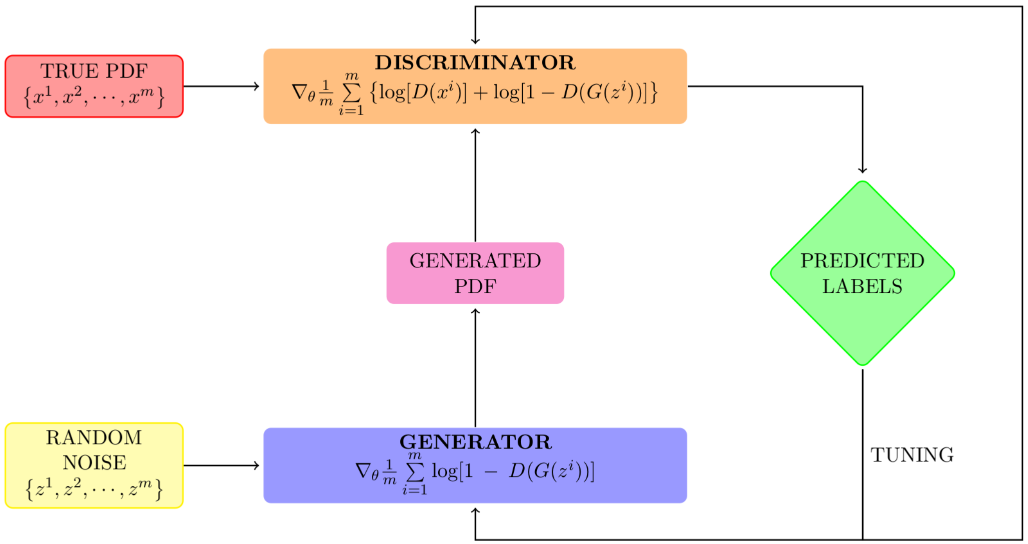
(Carrazza, Latorre, Kassabov, Rojo, 2015)

- START WITH LARGE REPLICA SAMPLE
- SELECT (BY GENETIC ALGORITHM) SUBSET OF REPLICAS $\Rightarrow$ STATISTICAL FEATURES OPTIMIZED TO PRIOR
- FOR ALL PDFs ON A GRID OF POINTS
  MINIMIZE DIFFERENCE OF FIRST FOUR MOMENTS, CORRELATIONS; OUTPUT OF KOLMOGOROV-SMIRNOV TEST (NUMBER OF REPLICAS BETWEEN MEAN AND σ, 2σ, INFINITY)
- 50 COMPRESSED REPLICA REPRODUCE 1000 REPLICA SET TO PRECENT ACCURACY

SOLVING THE PROBLEM....
GAN REPLICA GENERATION

- CAN WE REDUCE THE NUMBER OF COMPRESSED REPLICAS WITHOUT LOSS OF INFORMATION? SOLUTION FOR USER
- CAN WE INCREASE THE NUMBER OF REPLICAS WITHOUT REFITTING? SOLUTION FOR PDF FITTER

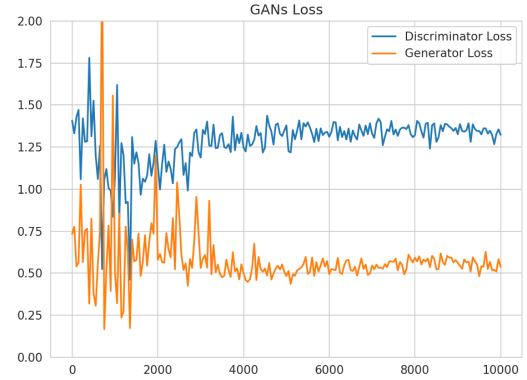
GENERATIVE ADVERSARIAL NETWORKS



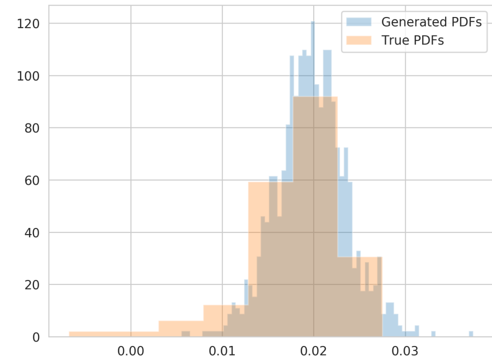
- TRAIN A NETWORK TO SIMULATE THE TRUE DISTRIBUTION (GENERATOR)
- TRAIN A NETWORK TO DISCRIMINATE TRUTH FROM SIMULATION (DISCRIMINATOR)
- TRAIN THE GENERATOR TO TRICK THE DISCRIMINATOR

SOLVING THE PROBLEM....
GAN REPLICA GENERATION

GAN TRAINING



UP VALENCE AT FIXED x

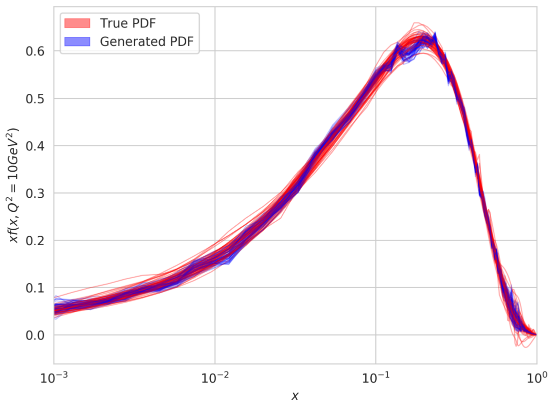


(Carrazza, Rabemananjara, preliminary)

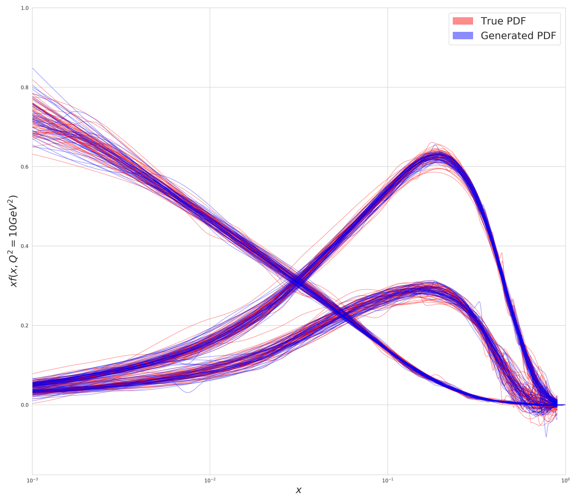
- 1D GAN: REPRODUCE THE INFORMATION IN THE UNDERLYING REPLICA SET, BUT NO GAIN (WIGGLY REPLICAS)
  $\Rightarrow$ REDUCE THE NUMBER OF COMPRESSED REPLICA WITH FIXED NUMBER OF FITTED REPLICAS W/O INFORMATION LOSS
- 2D GAN: COMBINE CORRELATED INFORMATION FROM UNDERLYING REPLICA SET INFERRING THE TRUE UNDERLYING DISTTRIBUTION
  $\Rightarrow$ REDUCE THE NUMBER OF INPUT REPLICAS W/O INFORMATION LOSS

N3PDF
Machine Learning • PDFs • QCD

ONE-DIMENSIONAL



TWO-DIMENSIONAL



OUTLOOK

SUMMARY

USE OF PDFs FOR PRECISION PHYSICS

DOES NOT ALLOW SHORTCUTS

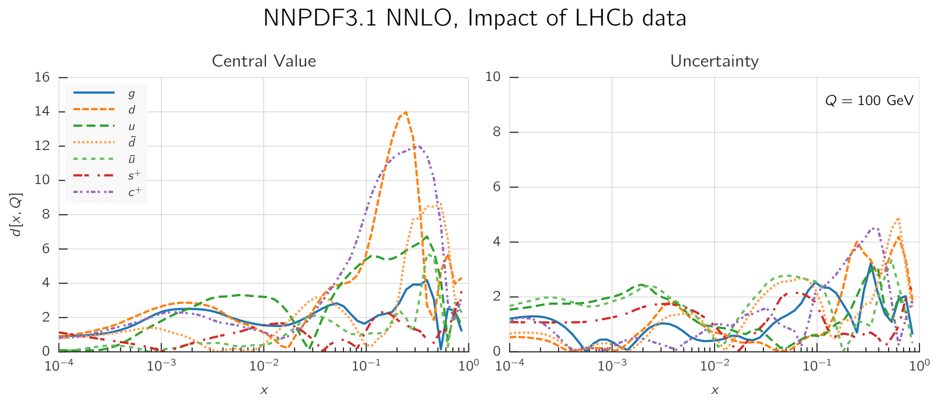
- CANNOT PICK THE DATASET
- MUST OPTIMIZE STATISTICS
- REMEMBER PDFS LIVE IN A SPACE OF FUNCTIONS

EXTRAS

THE IMPACT OF LHC DATA
FLAVOR SEPARATION

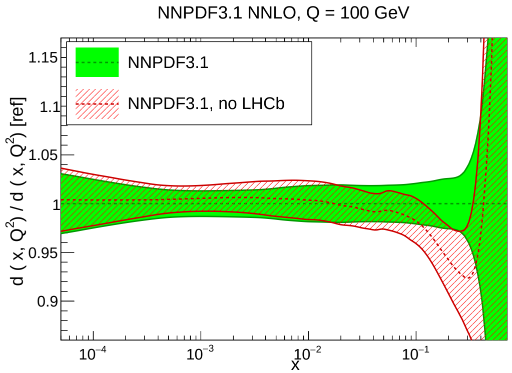
- BEFORE LHC $\Rightarrow$ CC DIS, TEV FIXED-TARGET DY, W ASYM.
- AFTER LHC $\Rightarrow$ WIDE RANGE OF W, Z PRODUCTION DATA

IMPACT OF LHCB
DISTANCES (difference in units of st. dev.)

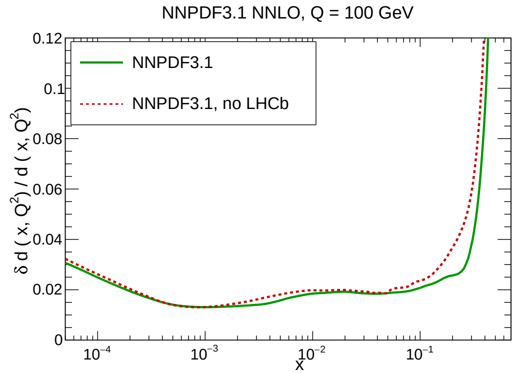


PDF COMPARISON: DOWN

CENTRAL VALUE



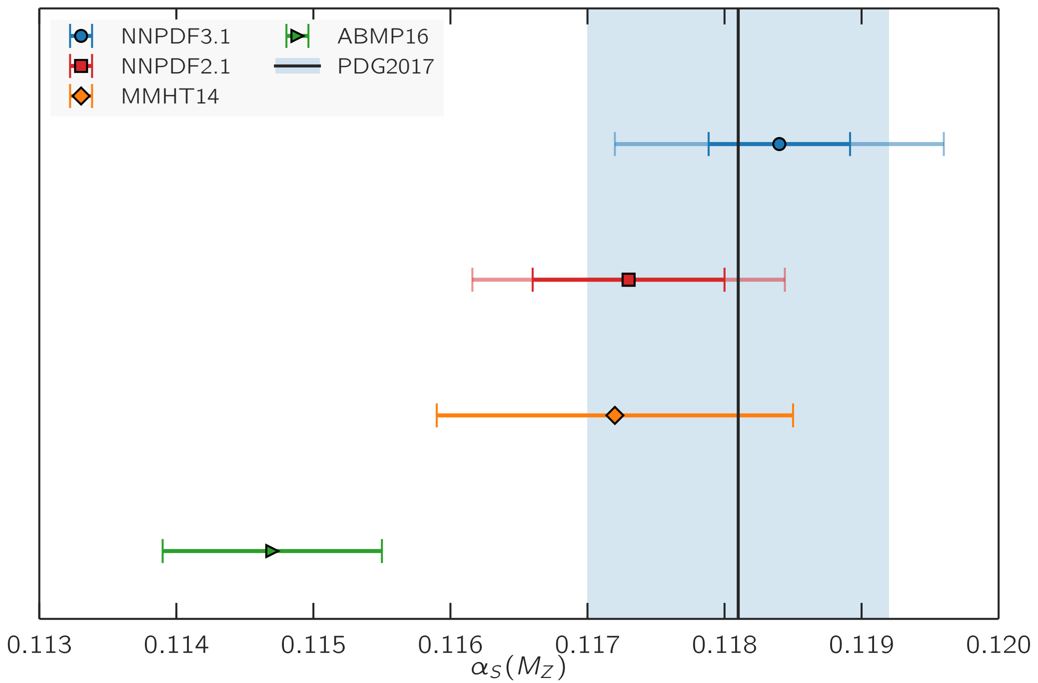
UNCERTAINTY



- SIZABLE SHIFT OF CENTRAL VALUE BY ALMOST ONE SIGMA
- LARGE x UNCERTAINTY DOWN BY LARGE FACTOR!

α_s FINAL RESULT & COMPARISON

$$\alpha_s^{\mathrm{NNLO}}(M_Z) = 0.1185 \pm 0.0005^{\mathrm{exp}} \pm 0.0001^{\mathrm{meth}} \pm 0.0011^{\mathrm{th}} = 0.1185 \pm 0.0012\ (1\%)$$

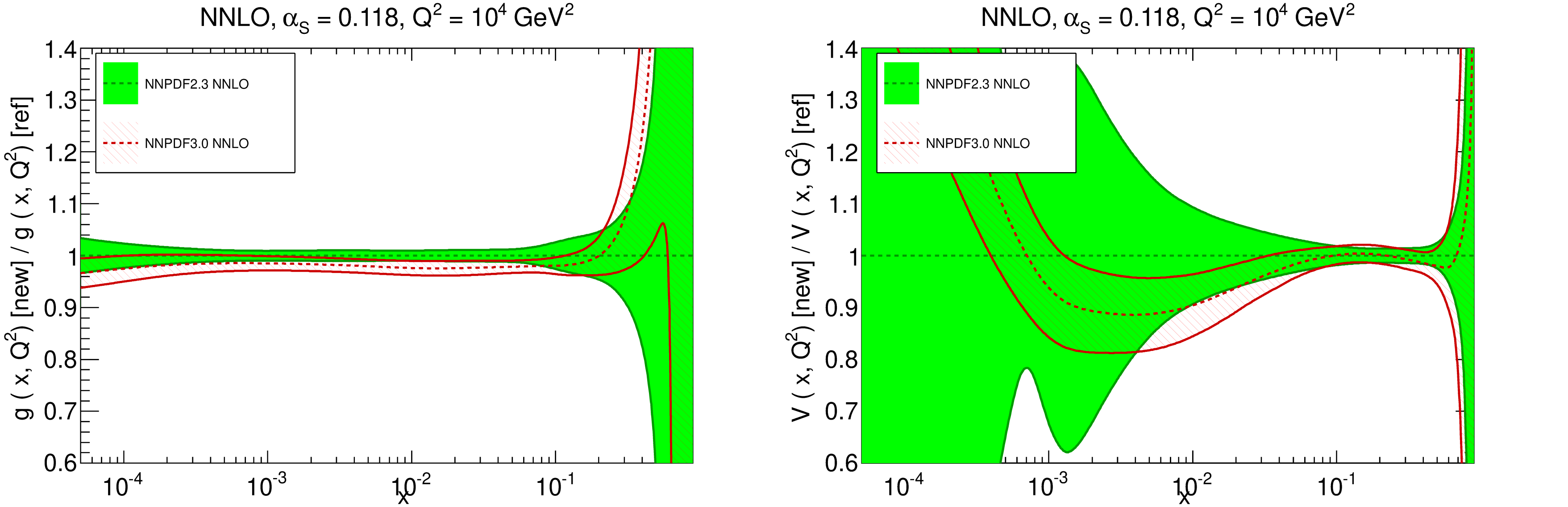


- SIGNIFICANTLY SMALLER EXP. UNCERTAINTY IN COMPARISON TO PREVIOUS NNPDF2.1 DETERMINATION (DESPITE MORE CONSERVATIVE ESTIMATE)
- SOMEWHAT LARGER CENTRAL VALUE THAN MMHT

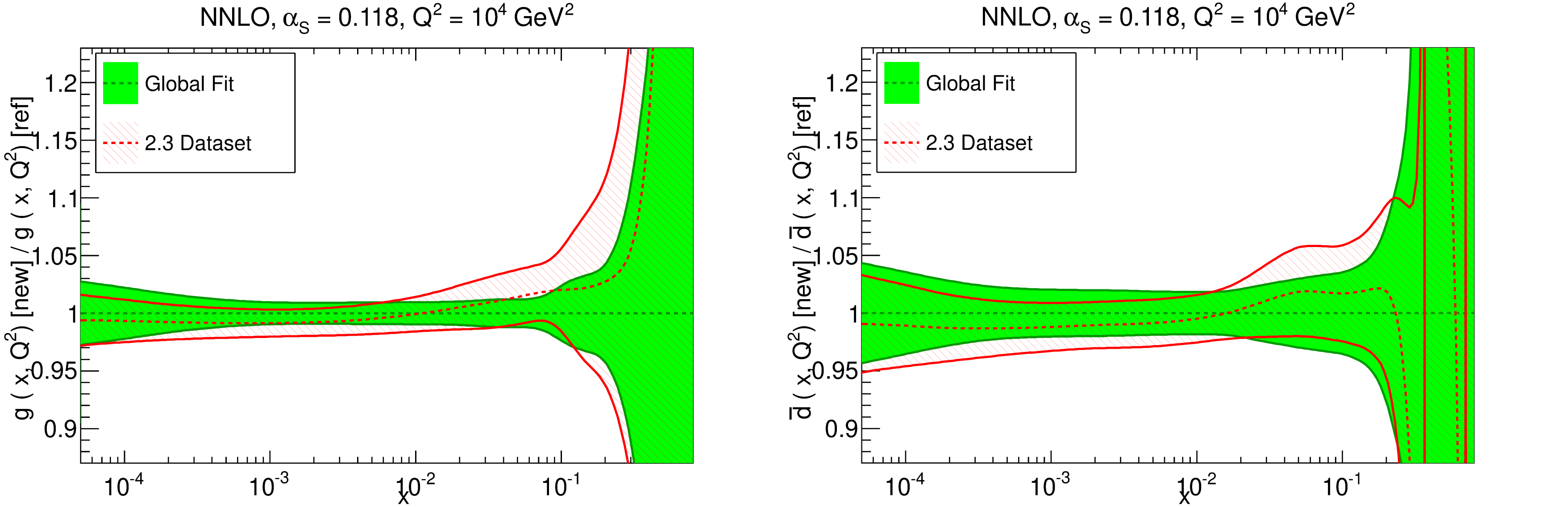
CONSISTENCY VS INFORMATION LOSS

- PDF SETS MUST BE BACKWARD CONSISTENT (THEY ARE)
- PDF UNCERTAINTY MIGHT IMPROVE EVEN WITH UNCHANGED DATASET (THEY DO)

NNPDF 2.3 VS 3.0: GLUON & VALENCE



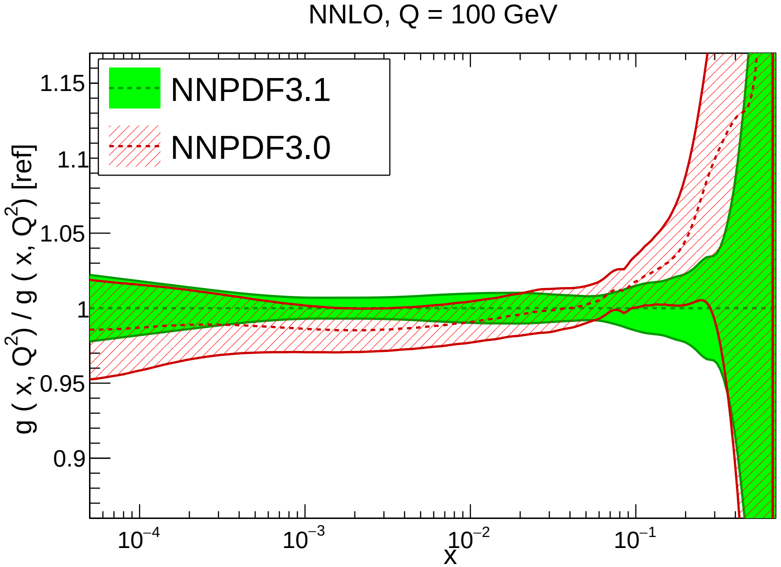
NNPDF 3.0 DEFAULT VS 2.3-LIKE DATASET: GLUON & ANTIDOWN

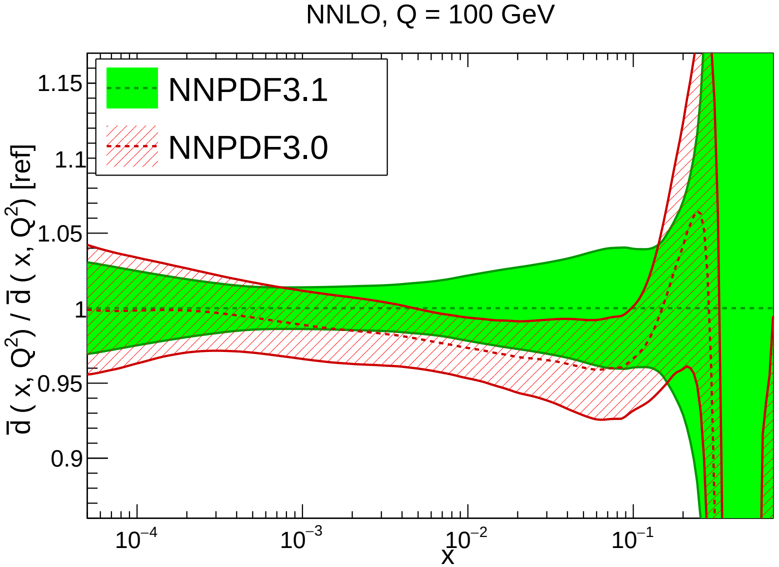


CONSISTENCY VS INFORMATION LOSS

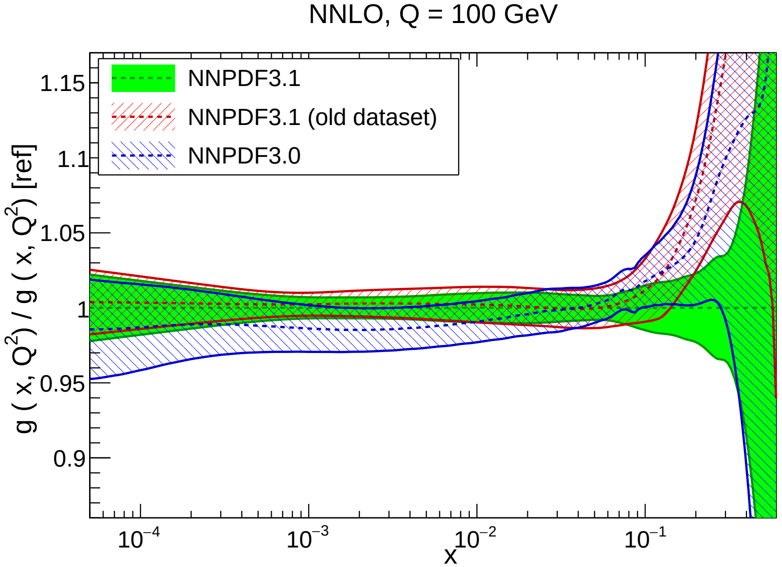
- PDF SETS MUST BE BACKWARD CONSISTENT (THEY ARE)
- PDF UNCERTAINTY MIGHT IMPROVE EVEN WITH UNCHANGED DATASET (THEY DO)

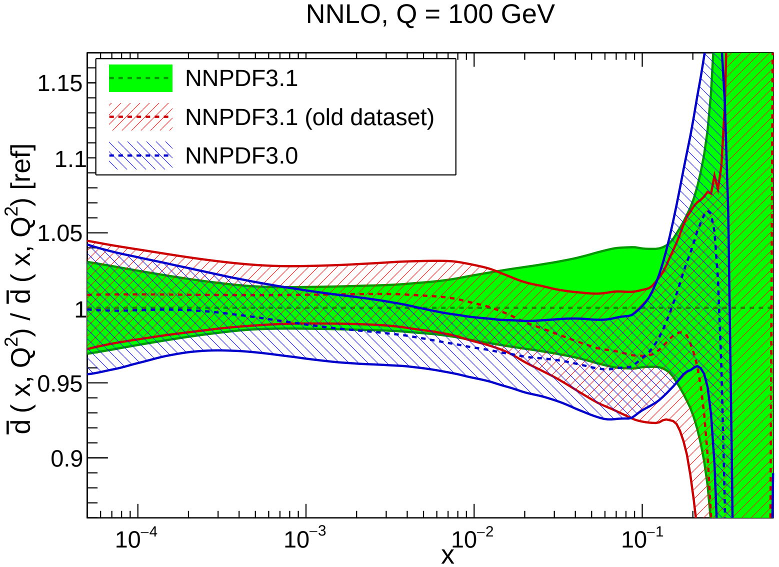
NNPDF 3.1 VS 3.0: GLUON & ANTIDOWN





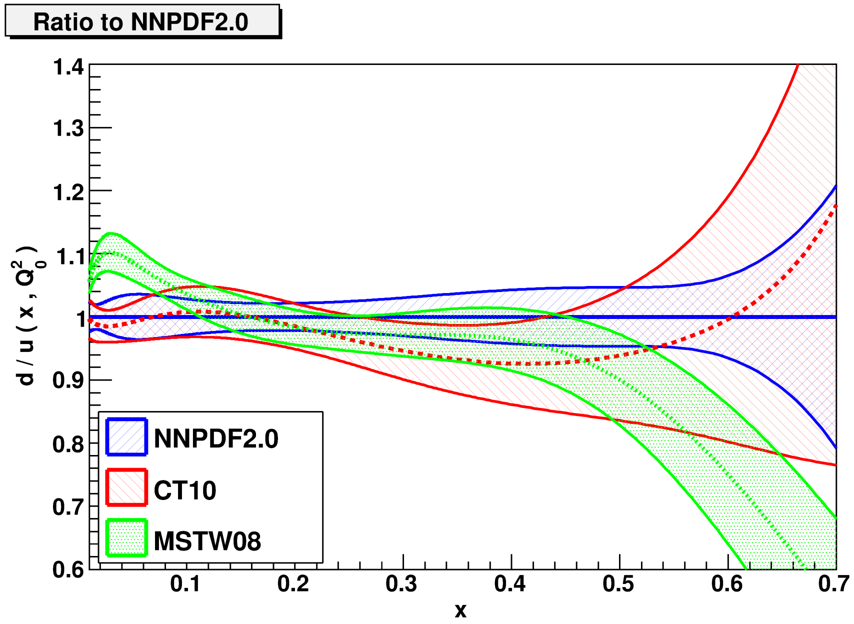
NNPDF 3.1 DEFAULT VS 3.0-LIKE DATASET



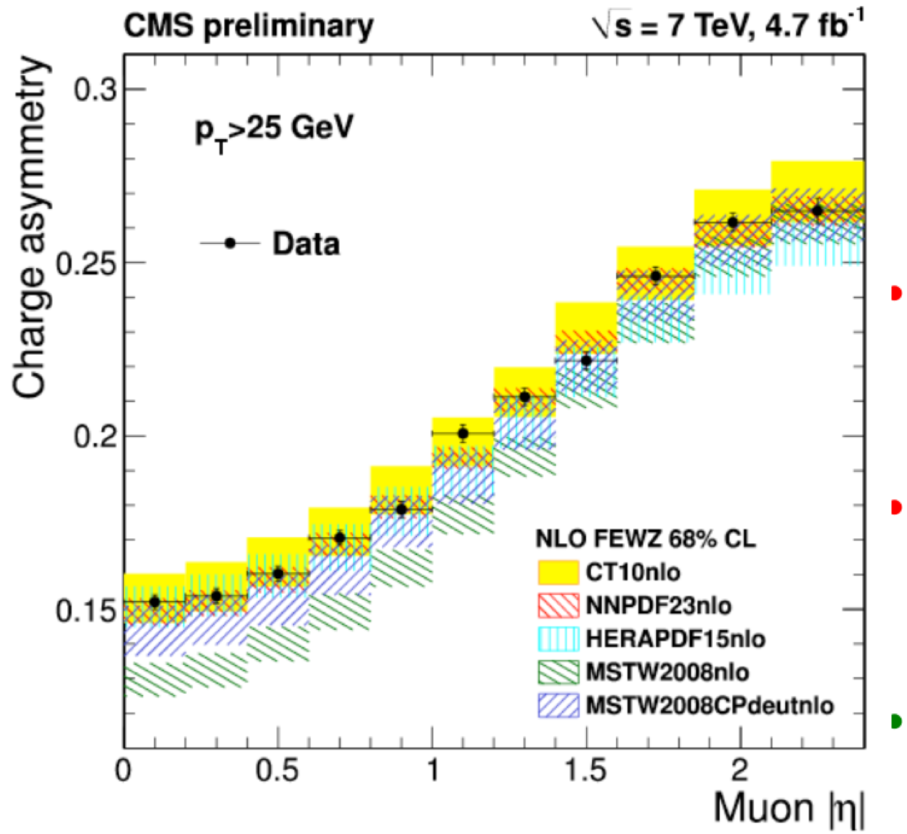


EXAMPLE OF DATA-DRIVEN PROGRESS
MSTW/MMHT: THE d/u RATIO

THE d/u RATIO



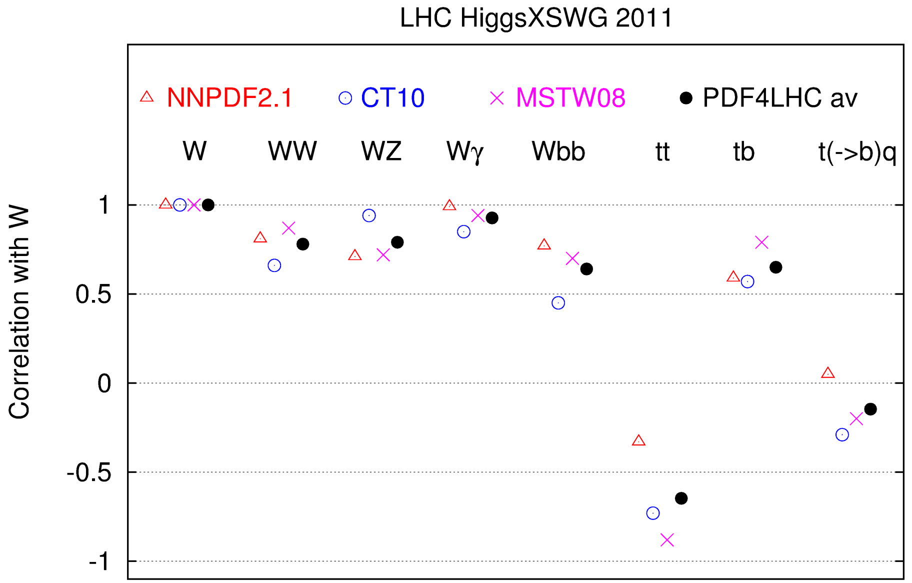
THE CMS W ASYMMETRY



- 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 $\Rightarrow$ FIXED IN MSTW08DEUT/MMHT

CORRELATING PDFS

CORRELATION BETWEEN HIGGS SIGNAL AND BACKGROUND (HXSWG, YR2)



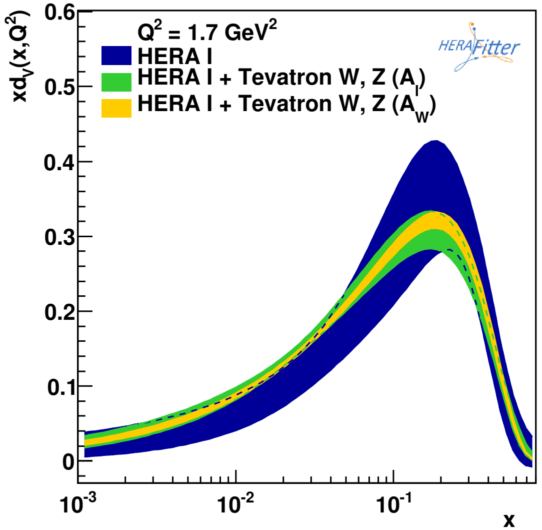
- CORRELATION BETWEEN PROCESSES AND PDFS, PROCESSES AND PROCESSES, PDF AND PDFS TRIVIAL TO COMPUTE $\Rightarrow$ NO NEED TO RUN DEDICATED FITS
- PREVIOUS EXERCISES SUGGEST VERY LARGE CORRELATION (SHOULD BE 100% FOR SAME DATA)
- IN PDF4LHC15 CORRELATION ASSUMED TO BE 100%: SIMPLE AVERAGE WEIGHTED AVERAGE DUBIOUS AND DANGEROUS
  - PDFS W/ SMALLER UNCERTANITY GET LARGER WEIGHT UNCERTAINTY DOMINATED BY METHODOLOGY $\Rightarrow$ SMALLER UNCERTAINTY COULD JUST BE BIAS!
  - UNCERTAINTY REDUCED IF CORRELATION LESS THAN 100% CAN WE BELIEVE IT IN THE ABSENCE OF NEW INFORMATION?

WHAT ABOUT XFITTER?

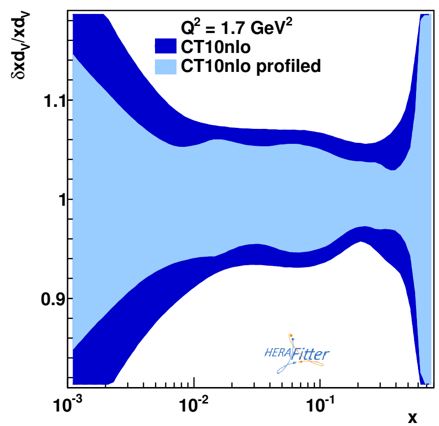
- OFTEN USED TO ASSESS IMPACT OF X IN "HERA+X" FITS

IMPACT OF THE TEVATRON W ASYMMETRY

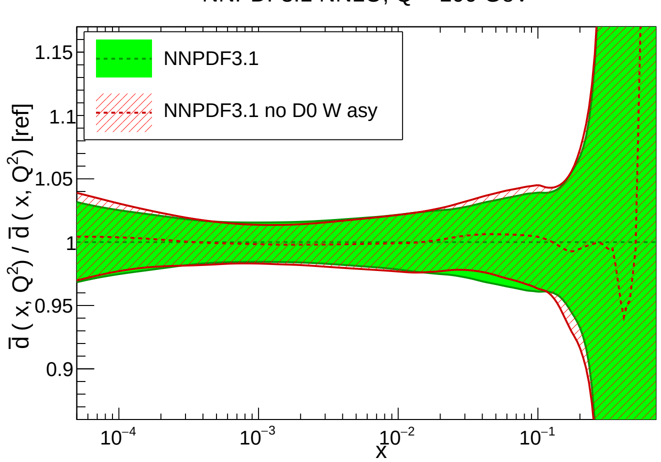
XFITTER: IMPACT ON HERA



XFITTER: IMPACT ON CT10



NNPDF3.1: IMPACT ON GLOBAL FIT



- IMPACT EXAGGERATED BY
  - COMPARISON TO SMALL DATASET
  - SOMEWHAT RESTRICTIVE PARAMETRIZATION