

PDF CHALLENGES HL-LHC AND BEYOND

STEFANO FORTE
UNIVERSITÀ DI MILANO & INFN



UNIVERSITÀ DEGLI STUDI DI MILANO
DIPARTIMENTO DI FISICA



PUSHING THE BOUNDARIES

IPPP, DURHAM, JULY 5 2018

PDF CHALLENGES

AN EXAMPLE: HIGGS IN GLUON FUSION

YR4 vs. NOW

~~(APPR)~~ N³LO+N³LL QCD (EFT); NLO ~~PURE~~ EW; NLO EXACT HQ;
NNLO APPROX TOP; NNLO PDFs

$$\begin{aligned}\sigma(\text{LHC13}, m_H = 125 \text{ GeV}) &= 48.58 \text{ pb} \pm \cancel{2.2^{\text{TH}} (4.5\%)} \pm \cancel{1.6^{\text{PDF}+\alpha_s} (3.2\%)} \\ \sigma(\text{LHC13}, m_H = 125 \text{ GeV}) &= \text{pb} \pm 1.6^{\text{TH}} (3.3\%) \pm 1.4^{\text{PDF}+\alpha_s} (2.8\%)\end{aligned}$$

PDF+ α_s UNCERTAINTY

PDF: ~~$\pm 0.9 \text{ pb} (1.9\%)$~~ $\pm 0.5 \text{ pb} (1\%)$

α_s : $\pm 1.3 \text{ pb} (2.6\%)$

- BOTH PDF & THEORY **UNCERTAINTY RAPIDLY DECREASING**
- **CAN WE BELIEVE** IN 1% PDF UNCERTAINTIES?

PDFs TO 1% (OR BETTER)

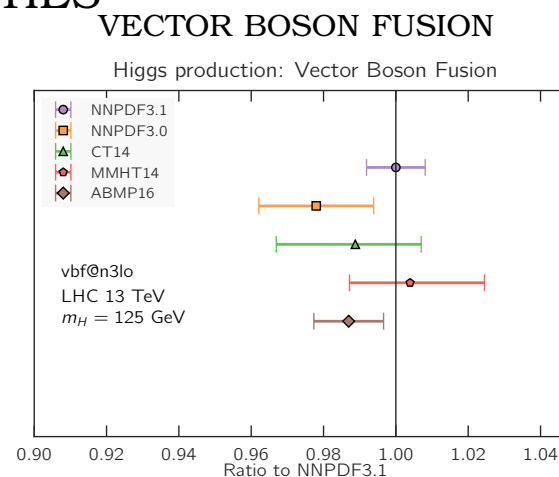
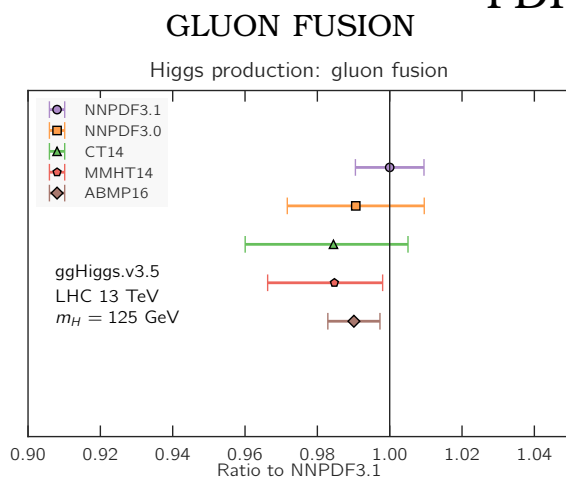
PDFs NOW

- CURRENTLY **PDF4LHC15 SET** RECOMMENDED FOR PRECISION PHENO
- **STATISTICAL COMBINATION** OF CT14, NNPDF3.0, NNPDF3.0 $\Rightarrow$ 2014 **DATA, MOSTLY PRE-LHC**
- **NNPDF3.1** AVAILABLE (2017) $\Rightarrow$ SIZABLE IMPACT OF **RUN I LHC DATA** $\Rightarrow$ **METHODOLOGICAL IMPROVEMENTS**

PROGRESS

HIGGS IN GLUON FUSION

PDF UNCERTAINTIES

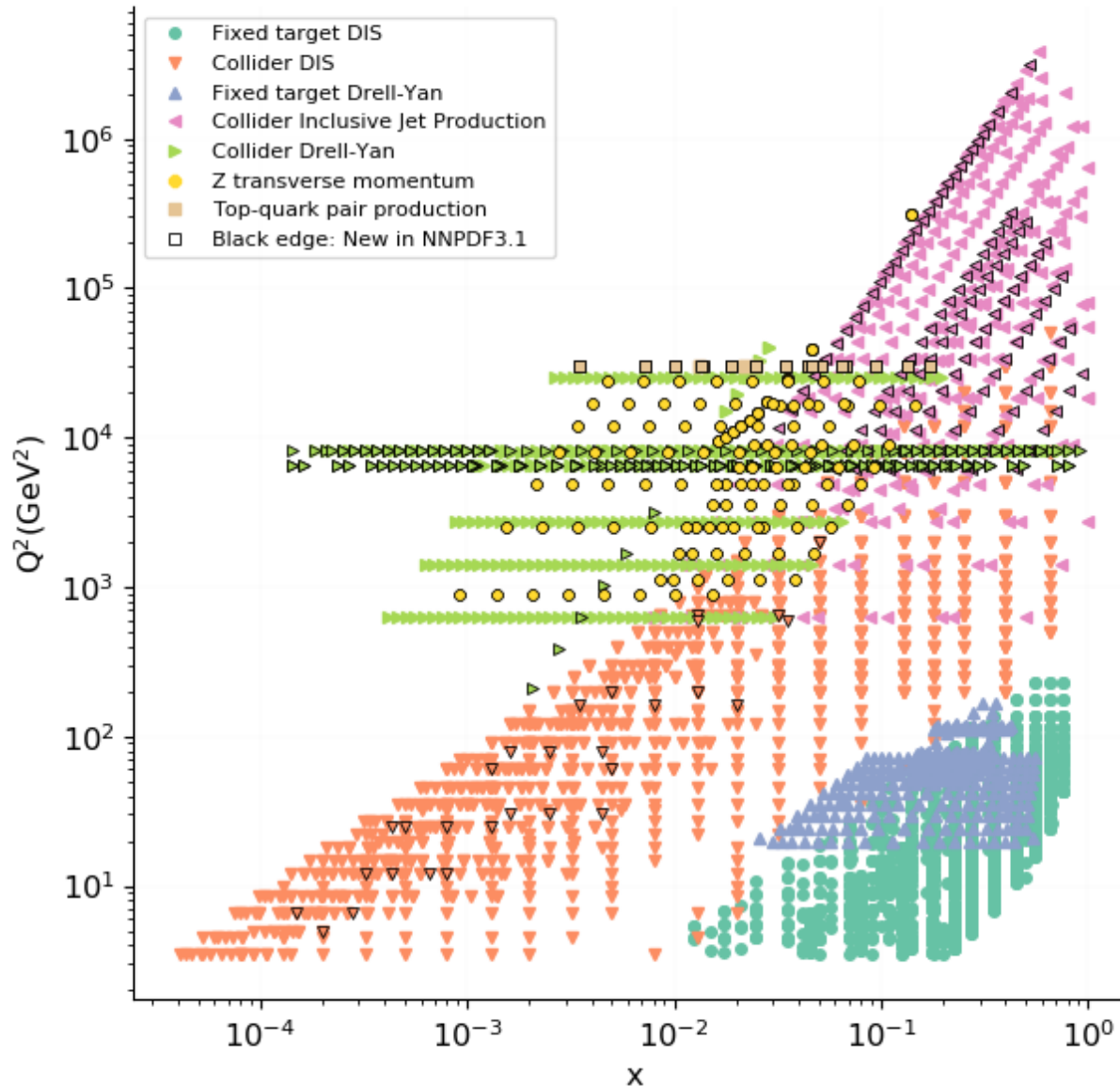


COMPARISON TO YR4

$$\Delta_{\text{ggF}}^{\text{PDF}} \sigma \approx 0.5 \text{ pb (1\%)} [\text{WAS } 0.9 \text{ pb (1.9\%)}]$$

WHY THE IMPROVEMENT? LHC DATA!: NNPDF3.0 vs NNPDF3.1

Kinematic coverage



NEW DATA: (BLACK EDGE)

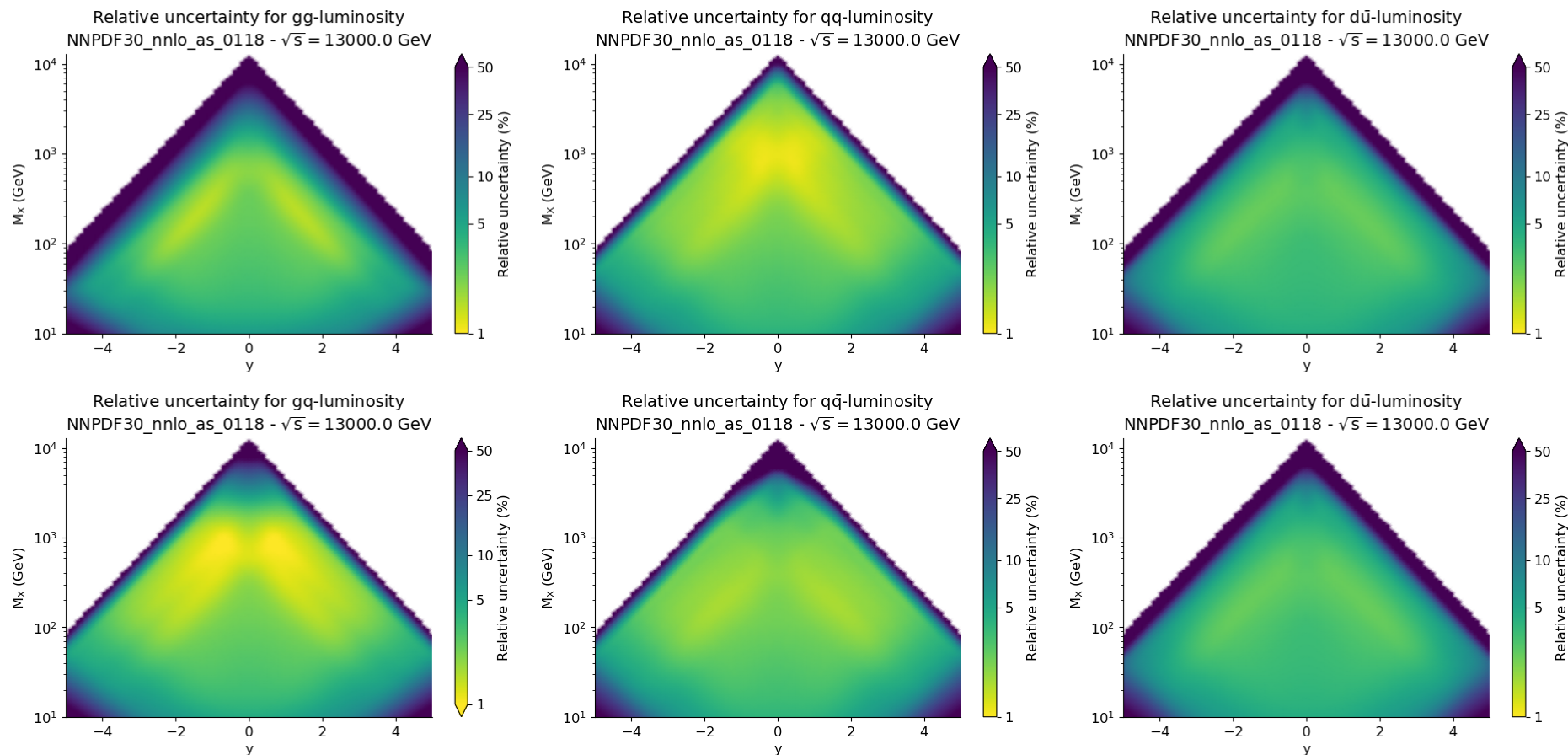
- HERA COMBINED F_2^b
- D0 W LEPTON ASYMMETRY
- ATLAS W, Z 2011, HIGH & LOW MASS DY 2011;
CMS $W^\pm$ RAPIDITY 8TeV
LHCb W, Z 7TeV & 8TeV
- ATLAS 7TeV JETS 2011, CMS 2.76TeV JETS
- ATLAS & CMS TOP DIFFERENTIAL RAPIDITY
- ATLAS $Z p_T$ DIFFERENTIAL RAPIDITY & INVARIANT MASS 8TeV,
CMS $Z p_T$ DIFFERENTIAL RAPIDITY 8TeV

THE IMPACT OF LHC DATA NNPDF3.0 (2014) PDF UNCERTAINTIES (NNLO)

GLUON

SINGLET

FLAVORS



- TYPICAL UNCERTAINTIES IN DATA REGION $\sim 3 - 5\%$

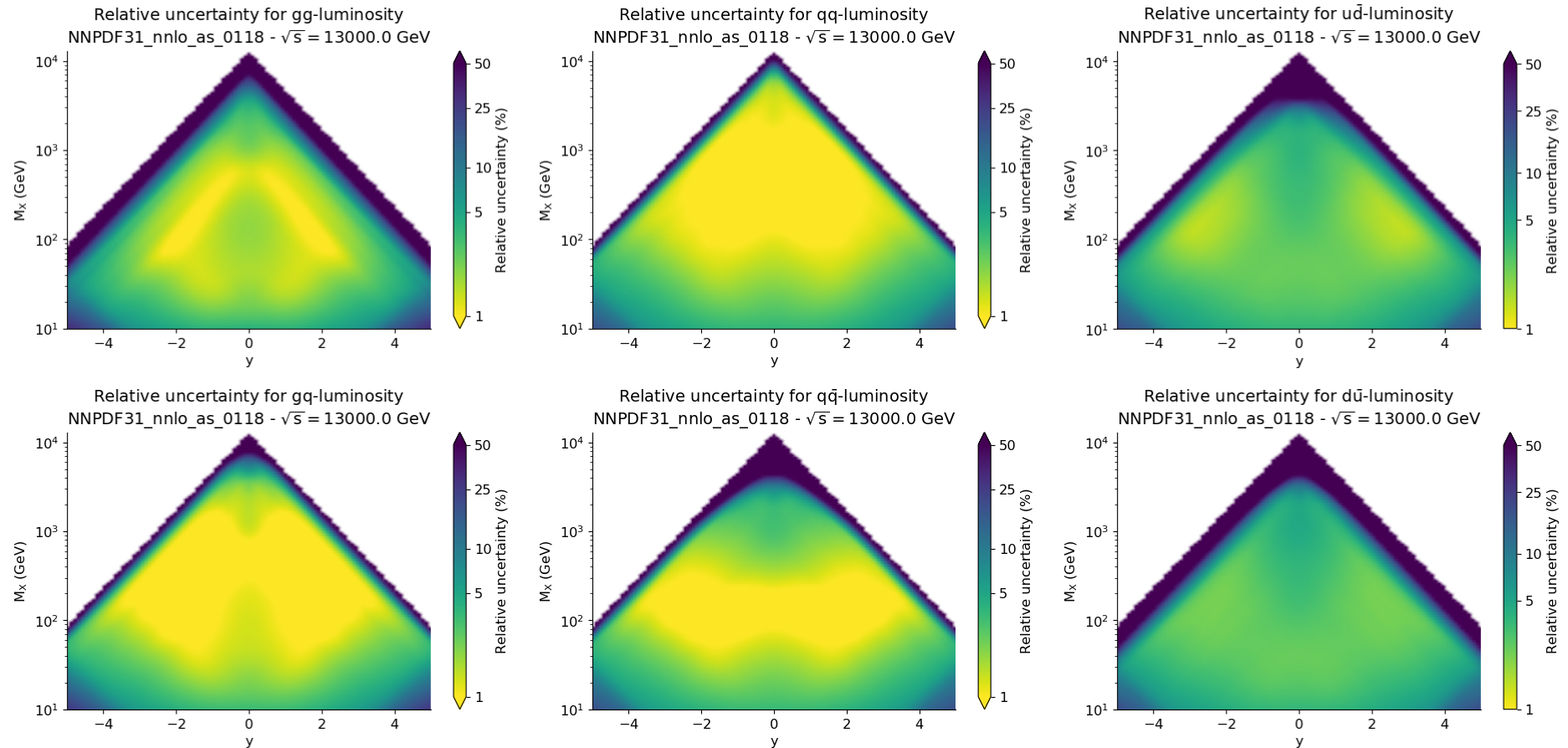
- .

THE IMPACT OF LHC DATA NNPDF3.1 (2017) PDF UNCERTAINTIES (NNLO)

GLUON

SINGLET

FLAVORS



- TYPICAL UNCERTAINTIES IN DATA REGION $\sim 1 - 3\%$
- NEW LHC DATA $\Rightarrow$ SIZABLE REDUCTION IN UNCERTAINTIES

“PDF” UNCERTAINTIES

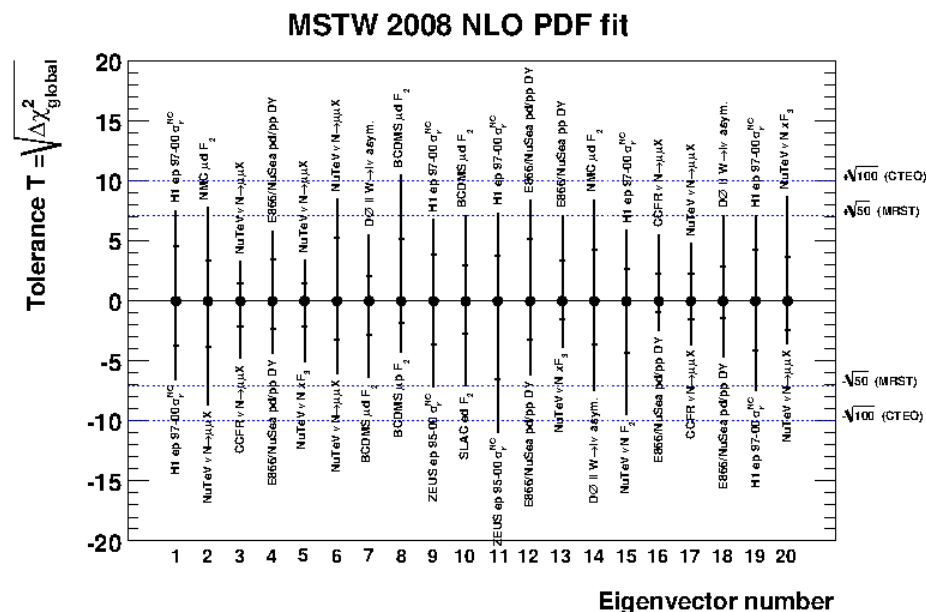
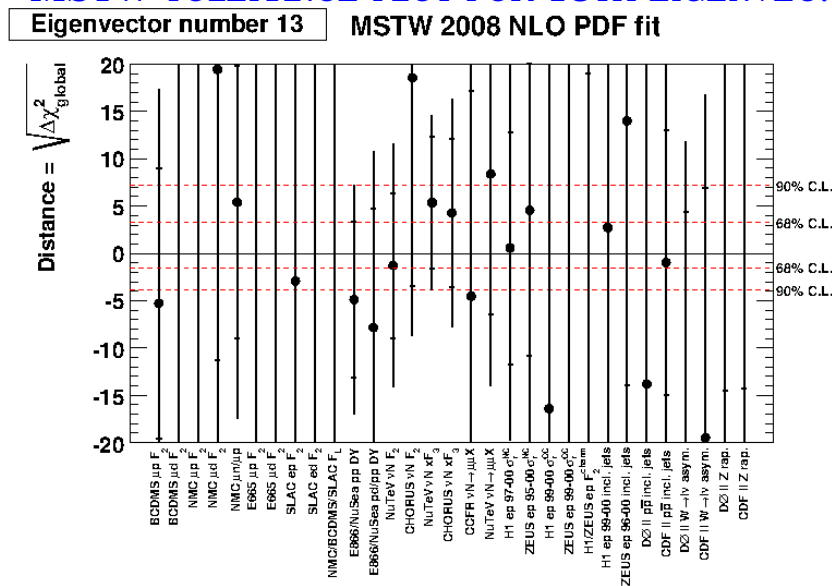
CAN WE TRUST PDF UNCERTAINTIES?

- “PDF” UNCERTAINTIES REFLECT UNCERTAINTY FROM THE DATA & METHODOLOGY (NOT THEORY)
- UNCERTAINTIES ON GLOBAL FITS $\Rightarrow$ SIMILAR SIZE DESPITE DIFFERENT PROCEDURES
- DUE TO UNCERTAINTY TUNING

TOLERANCE (MMHT-CT)

GLOBAL MSTW TOLERANCE

MSTW TOLERANCE PLOT FOR 13TH EIGENVEC.

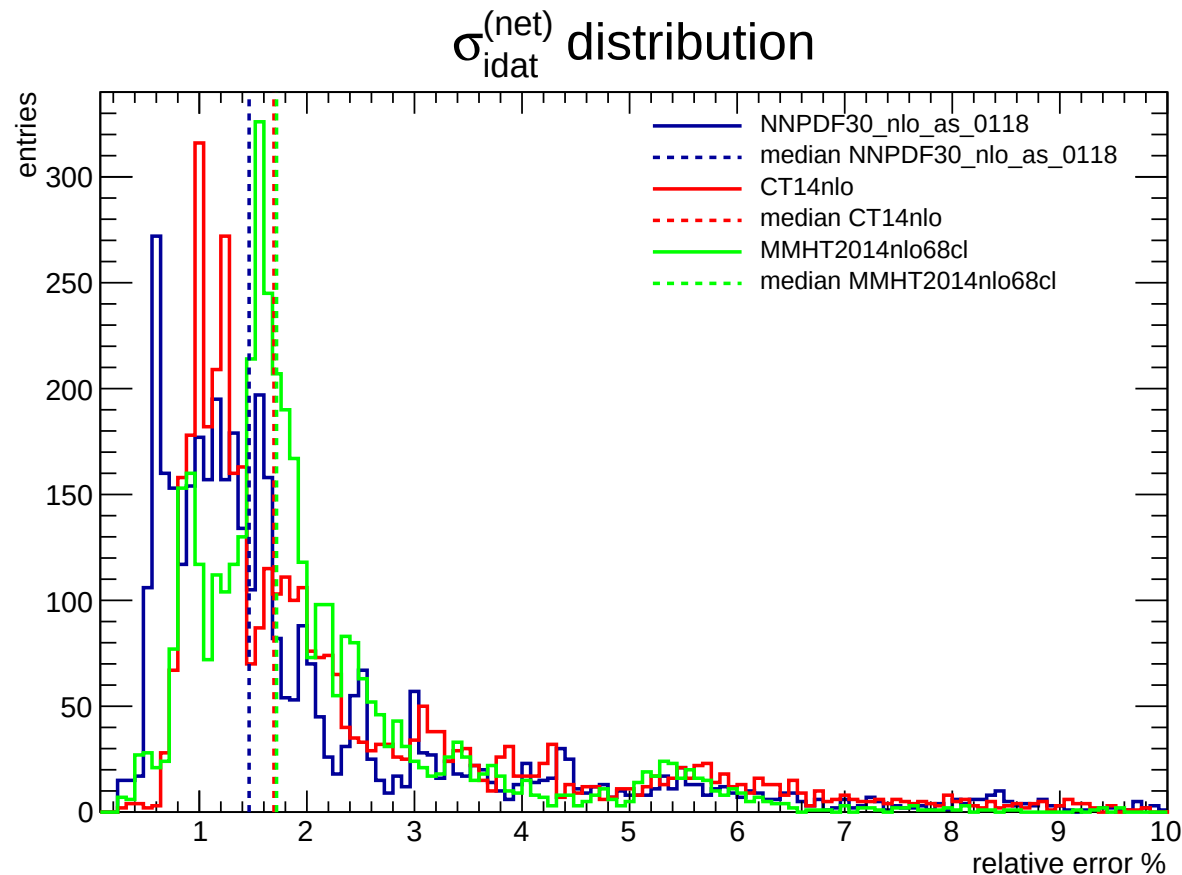


- (MSTW/MMHT) FOR EACH EIGENVECTOR IN PARAMETER SPACE DETERMINE CONFIDENCE LIMIT FOR THE DISTRIBUTION OF BEST-FITS OF EACH EXPERIMENT
- RESCALE $\Delta\chi^2 = T$ INTERVAL SUCH THAT CORRECT CONFIDENCE INTERVALS ARE REPRODUCED
- WHY DO WE NEED TOLERANCE?
- DO WE UNDERSTAND PDF UNCERTAINTIES?

PDF UNCERTAINTIES: HOW MUCH DO THEY VARY?

- COMPUTE **PERCENTAGE PDF UNCERTAINTY** ON ALL DATA INCLUDED IN GLOBAL FIT
- **COMPARE** GLOBAL FITS

PERCENTAGE PDF UNCERTAINTY ON PREDICTIONS



- **MEDIAN SIMILAR**
- **DISTRIBUTION VERY DIFFERENT!**
- **NNPDF: SMALLER MODE, BUT FAT TAIL $\Leftrightarrow$ GREATER FLEXIBILITY**

ARE PDF UNCERTAINTIES FAITHFUL?

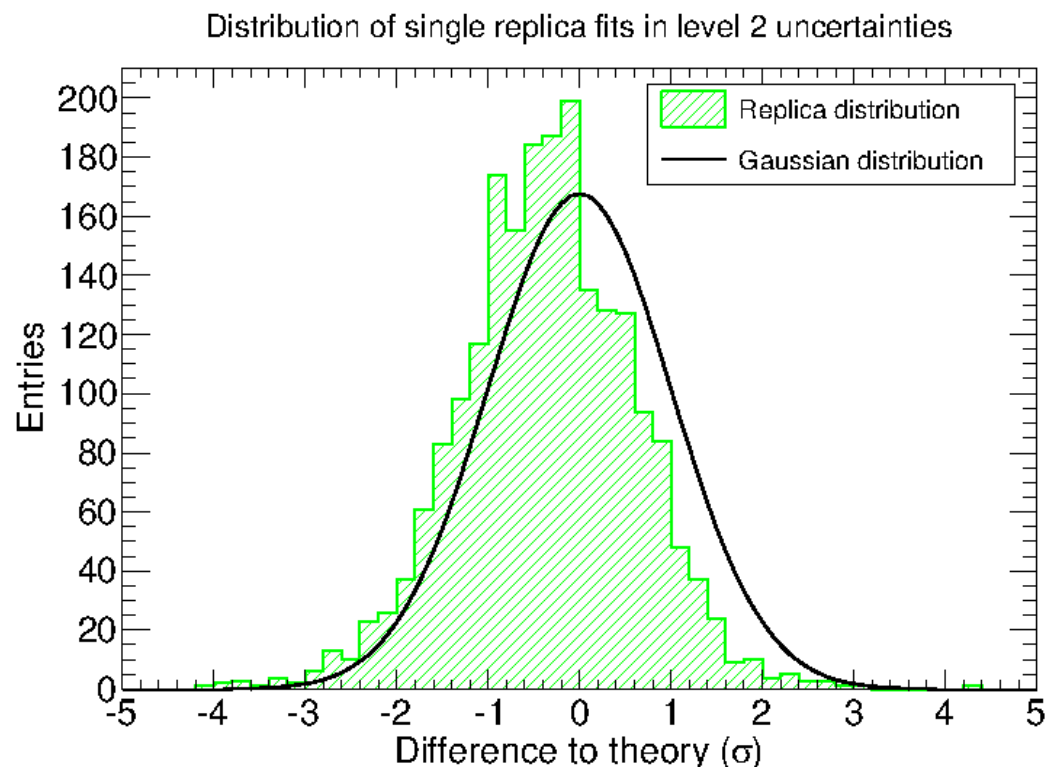
CLOSURE TESTING

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:
- TEST WHETHER PDF UNCERTAINTY FAITHFULLY REFLECTS DATA UNCERTAINTY:
E.G. TRUE VALUE WITHIN ONE-SIGMA 68% OF TIMES

CLOSURE-TESTING THE PDF UNCERTAINTIES RESULTS

UNCERTAINTIES: DISTRIBUTION OF DEVIATIONS BETWEEN FITTED AND “TRUE” PDFs, SAMPLED AT 20 POINTS BETWEEN 10^{-5} AND 1



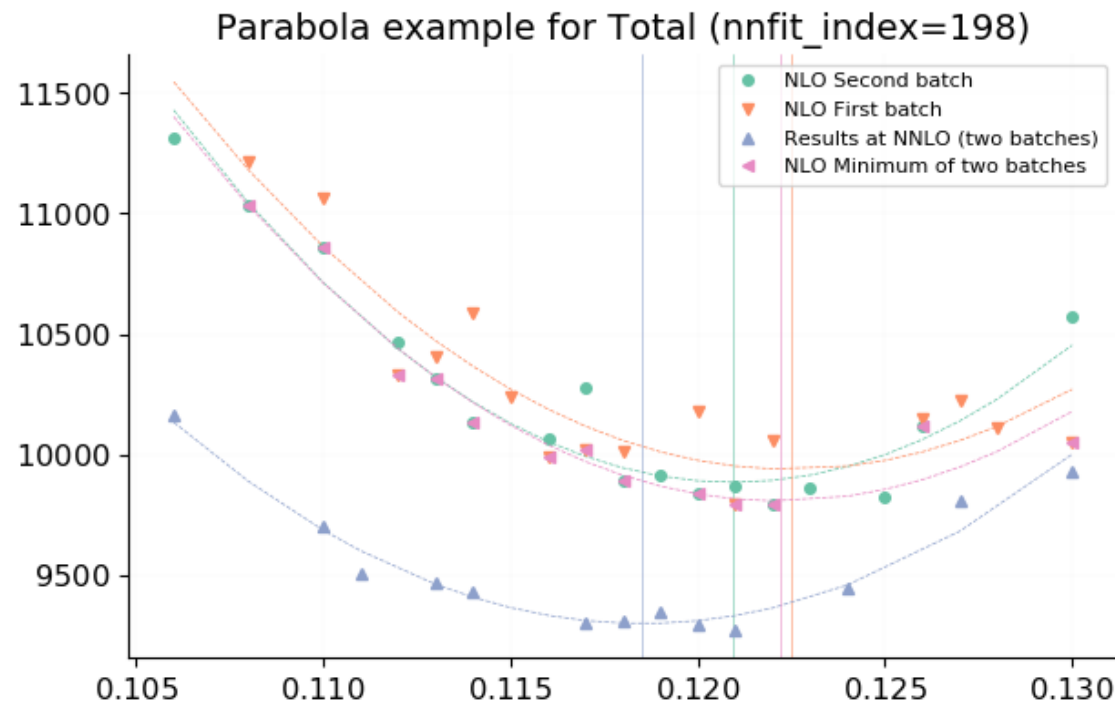
FIND 0.699% FOR ONE-SIGMA, 0.948% FOR TWO-SIGMA C.L.

ARE PDF UNCERTAINTIES OPTIMAL? THE $\Delta\chi^2$ PROBLEM

- TOLERANCE MIGHT COMPENSATE FOR MISSING FUNCTIONAL UNCERTAINTY
- BUT WHAT IS $\Delta\chi^2$ FOR AN NNPDF FIT?
- CAN ANSWER USING HESSIAN CONVERSION! $\Delta\chi^2 = 16 \pm 15$
 - NON-PARABOLIC BEHAVIOUR NEAR MINIMUM ON SCALE OF UNCERTAINTIES?
 - INEFFICIENCY OF THE MINIMIZATION PROCEDURE?

MORE EFFICIENT MINIMIZATION?

- LOOK AT α_s DEPENDENCE (CORRELATED REPLICAS)
- **SIGNIFICANT FLUCTUATIONS** ABOUT PARABOLIC SHAPE
NOT DUE TO FINITE-SIZE MONTE CARLO SAMPLE



BATCH MINIMIZATION

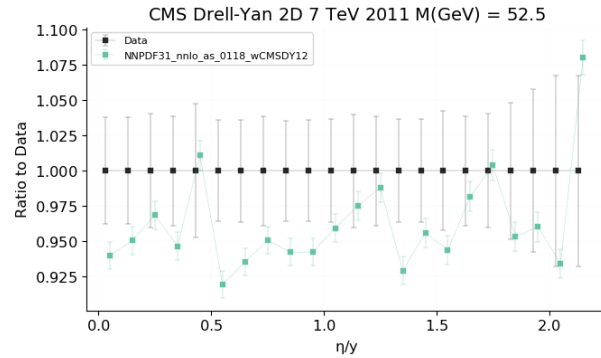
- MINIMIZE EACH REPLICA MORE THEN ONCE & KEEP BEST RESULTS
- SIGNIFICANT **STABILIZATION**

ARE UNDERLYING UNCERTAINTIES RELIABLE?

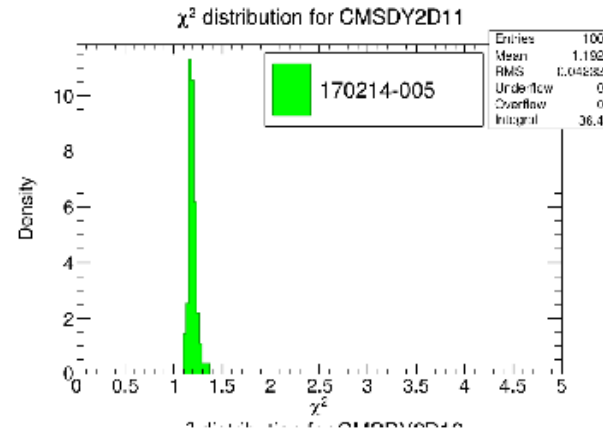
CORRELATIONS & THE COVARIANCE MATRIX

THE CMS DOUBLE-DIFFERENTIAL DRELL-YAN 2011

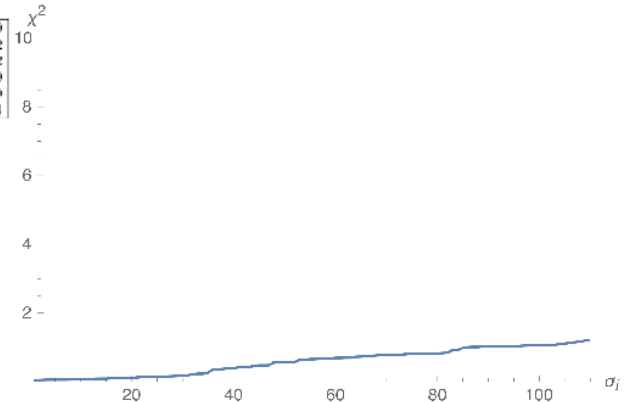
DATA/THEORY VS. DATA BIN



χ^2/dof HIST. OVER REPLICAS

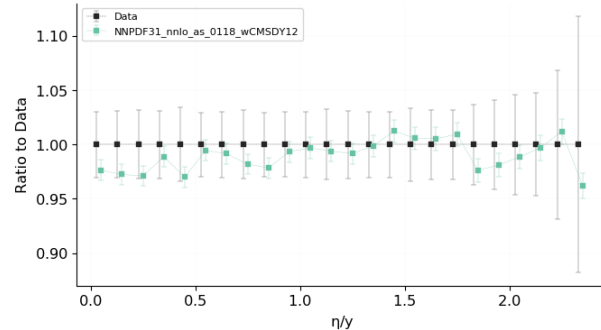


χ^2 AS COVMAT EIGVECS ADDED

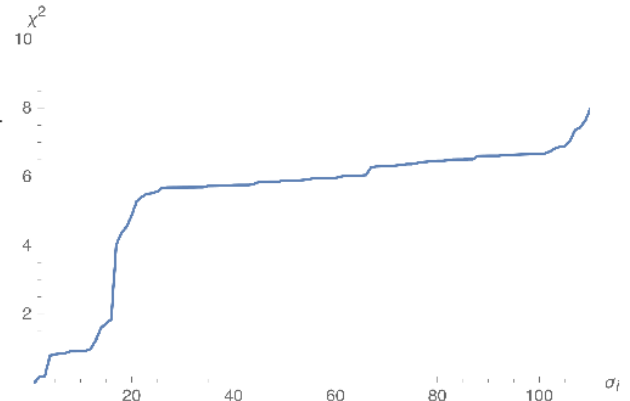
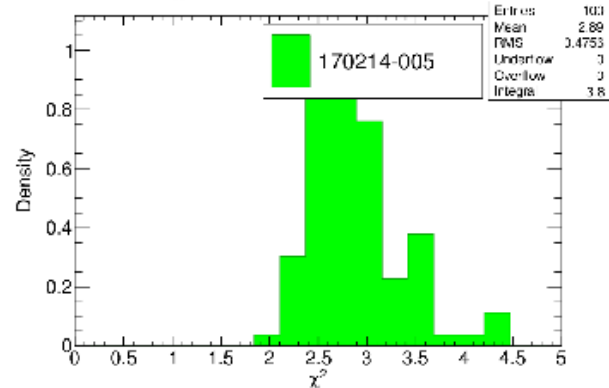


2012

CMS Drell-Yan 2D 8 TeV 2012 M(GeV) = 52.5



χ^2 distribution for CMSDY2D12



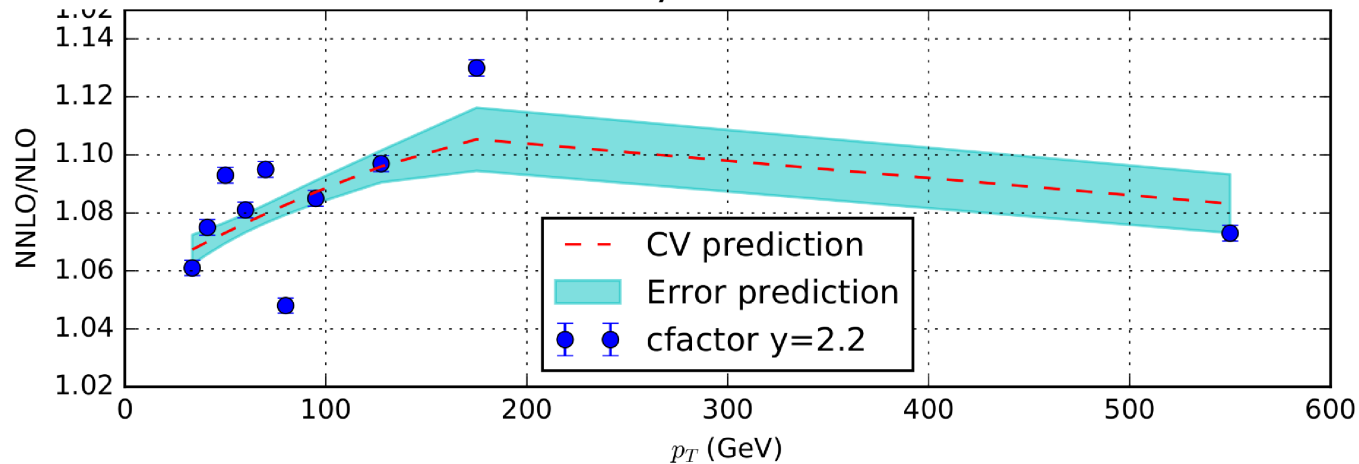
- FROM 2011 TO 2012, UNCORRELATED UNCERTAINTIES DOWN TO SUB-PERMILLE
- 2011: $\chi^2/dof \sim 1$; 2012: IMPOSSIBLE TO FIT BETTER THAN $\chi^2/dof \sim 3$
- PATHOLOGICAL BEHAVIOUR OF COVARIANCE MATRIX $\Rightarrow$ WHAT IS THE UNCERTAINTY ON IT?

THEORY UNCERTAINTIES

CAN WE TRUST NNLO CALCULATIONS?

NUMERICAL INSTABILITIES

AN EXAMPLE: ATLAS 7 TeV p_T DISTRIBUTION
THE NNLO/NLO K -FACTOR



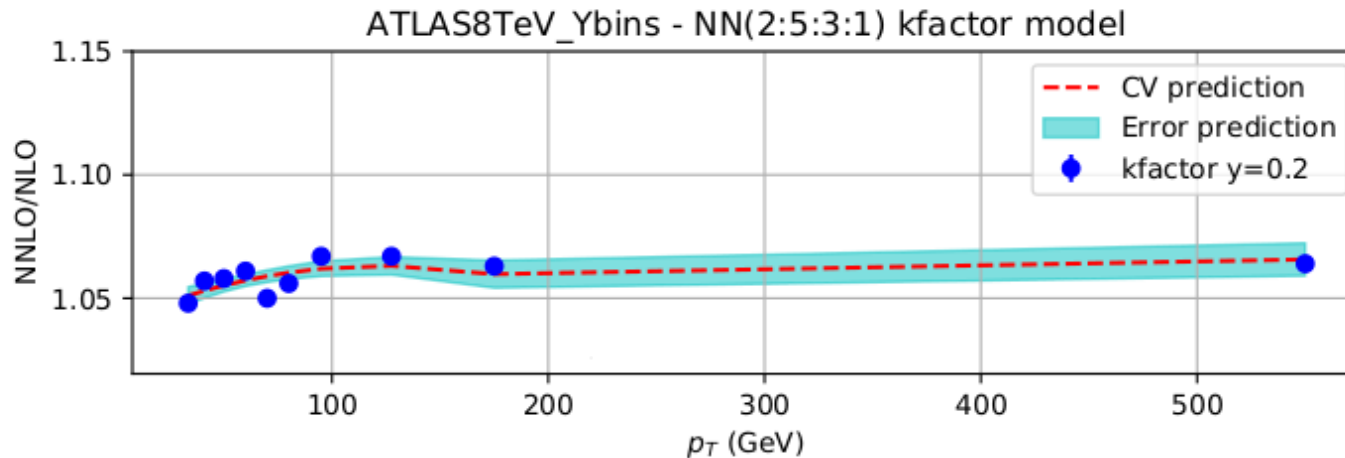
(Boughezal, Liu, Petriello, 2016)

- UNCORRELATED STATISTICAL UNCERTAINTIES AT PERMILLE LEVEL
- LARGE NNLO CORRECTIONS $\sim 10\%$
- NOMINAL K -FACTOR UNCERTAINTIES VERY SMALL: UNDERESTIMATED?
- FIT ONLY POSSIBLE WITH RELIABLE ESTIMATE OF UNCERTAINTY ON THEORY PREDICTION
- NNPDF3.1: EXTRA 1% THEORY UNCERTAINTY ESTIMATED BASED ON FLUCTUATIONS W.R. TO INTERPOLATION (SHADED IN PLOT)

CAN WE TRUST NNLO CALCULATIONS?

NUMERICAL INSTABILITIES

AN EXAMPLE: ATLAS 7 TeV p_T DISTRIBUTION
THE NNLO/NLO K -FACTOR



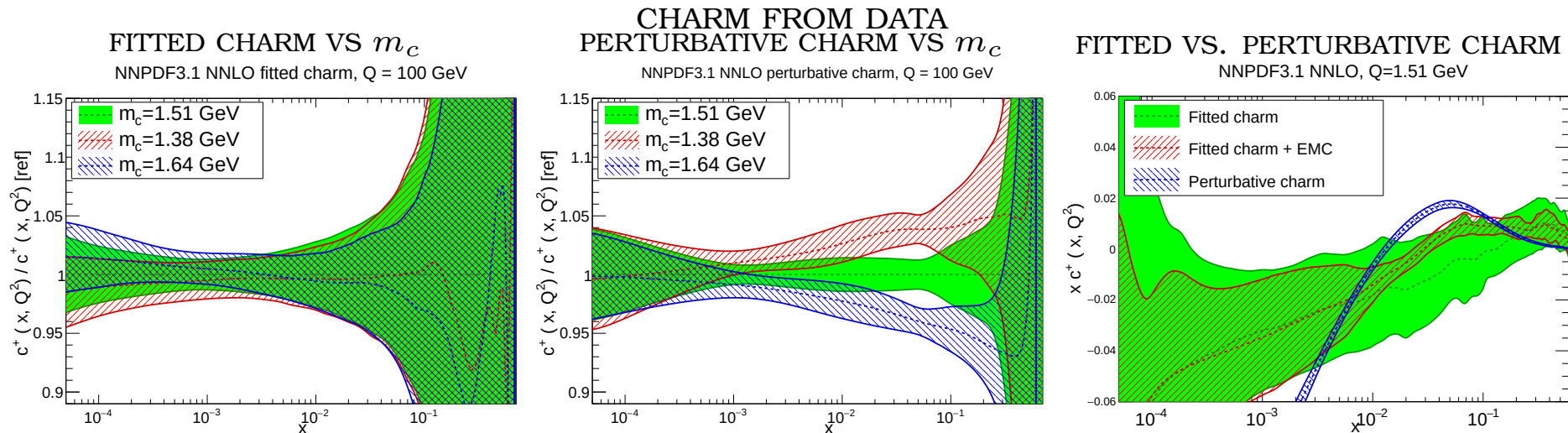
(Boughezal, Liu, Petriello, 2017)

- UNCORRELATED STATISTICAL UNCERTAINTIES AT PERMILLE LEVEL
- LARGE NNLO CORRECTIONS $\sim 10\%$
- NOMINAL K -FACTOR UNCERTAINTIES VERY SMALL: UNDERESTIMATED?
- FIT ONLY POSSIBLE WITH RELIABLE ESTIMATE OF UNCERTAINTY ON THEORY PREDICTION
- NNPDF3.1: EXTRA 1% THEORY UNCERTAINTY ESTIMATED BASED ON FLUCTUATIONS W.R. TO INTERPOLATION (SHADED IN PLOT)

CAN WE TRUST HEAVY QUARK PDFs?

- HEAVY QUARK PDFs DETERMINED BY (LOW-ORDER) MATCHING CONDITIONS + PERTURBATIVE EVOLUTION
- STRONG DEPENDENCE ON HQ (POLE) MASS

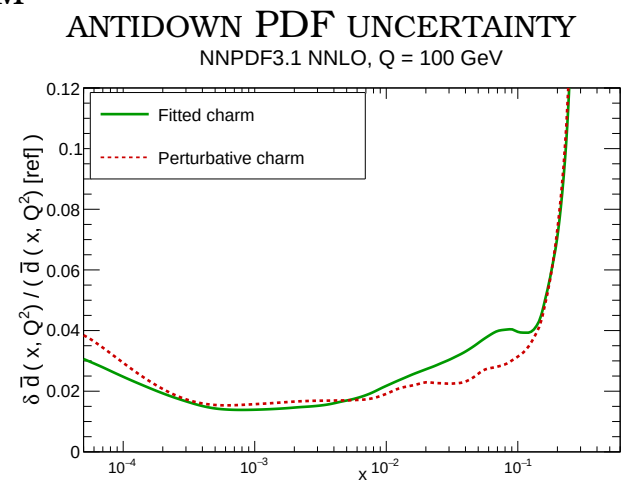
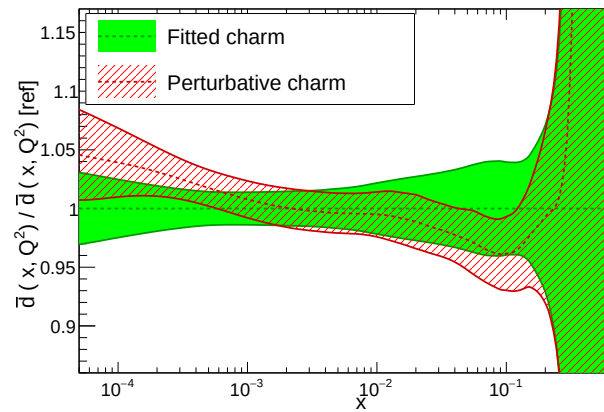
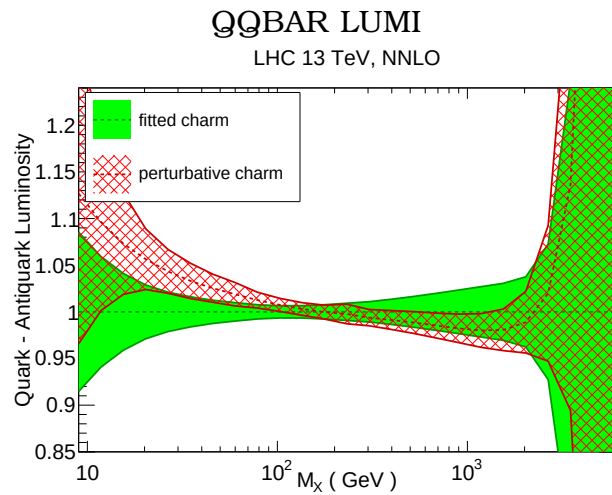
SOLUTION: HQ PDFs FROM DATA



- MASS DEPENDENCE GREATLY REDUCED
- SHAPE DIFFERS FROM NNLO MATCHING

CHARM FROM DATA IMPACT ON LIGHT QUARK PDFS

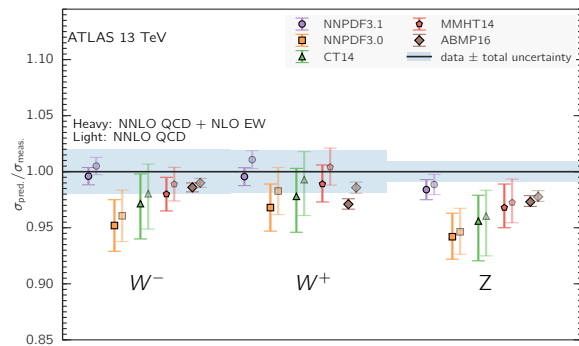
FITTED VS. PERTURBATIVE CHARM
ANTIDOWN PDF



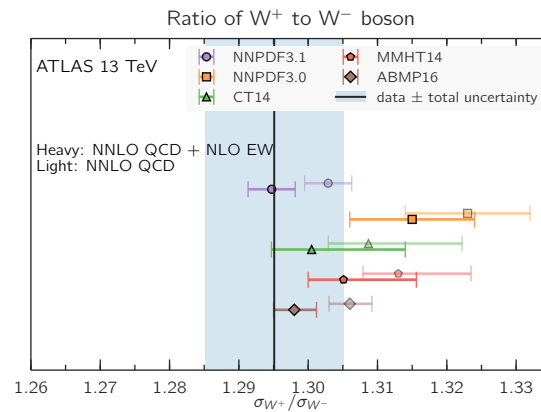
- FLAVOR DECOMPOSITION ALTERED
- UNCERTAINTIES ON LIGHT QUARKS NOT SIGNIFICANTLY INCREASED

CHARM FROM DATA IMPACT ON PHENOMENOLOGY

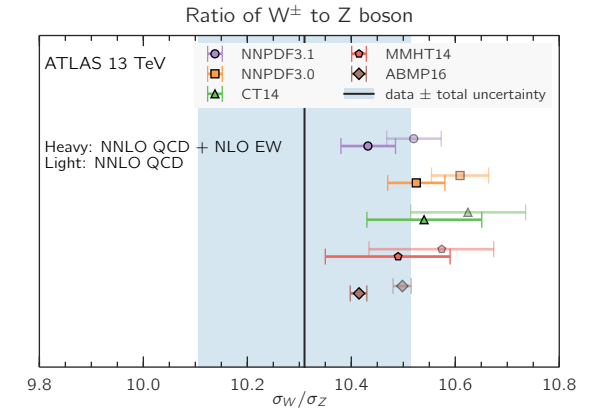
DRELL-YAN XSECTS



W^+/W^- XSECT RATIO



W/Z XSECT RATIO



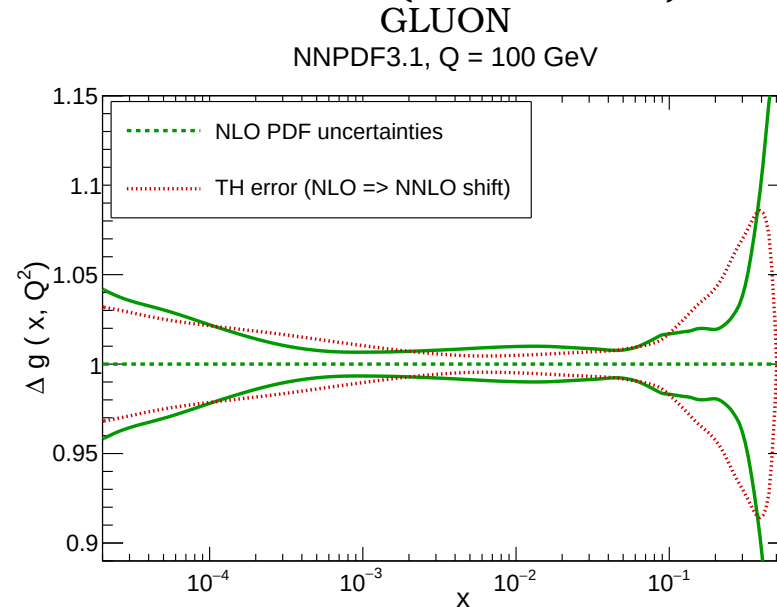
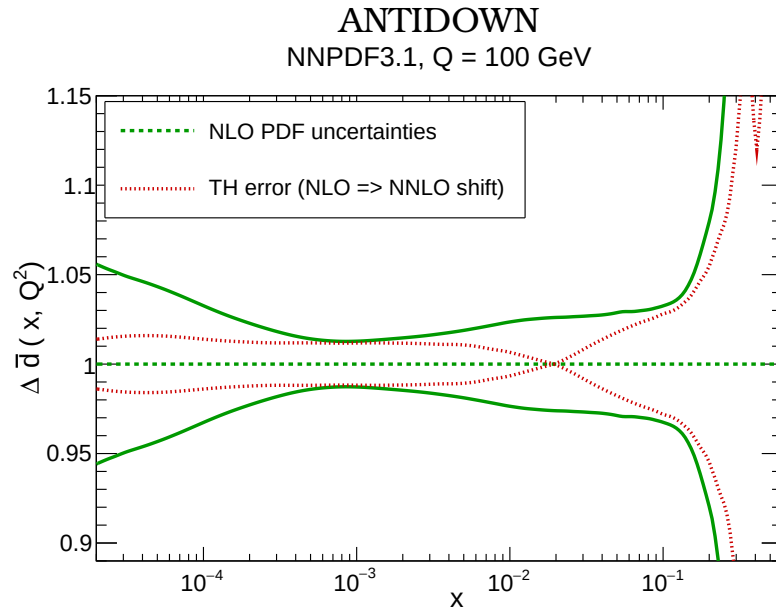
- W , Z CROSS-SECTIONS AT 13 TeV IN PERFECT AGREEMENT WITH DATA
THANKS TO FITTED CHARM!
- ELECTROWEAK CORRECTIONS IMPORTANT

WHAT ABOUT THE B PDF?

WHAT ABOUT MISSING HIGHER ORDERS?

- DOMINANT THEORY UNCERTAINTY ON QCD PREDICTIONS $\Rightarrow$ MHOU (SCALE)
- NOT INCLUDED IN PDF UNCERTAINTY
- HOW LARGE IS IT?
 $\Rightarrow$ AT NLO, CAN CHECK NLO-NNLO PDF SHIFT

NLO-NNLO SHIFT VS. NLO PDF UNCERTAINTY (NNPDF3.1)



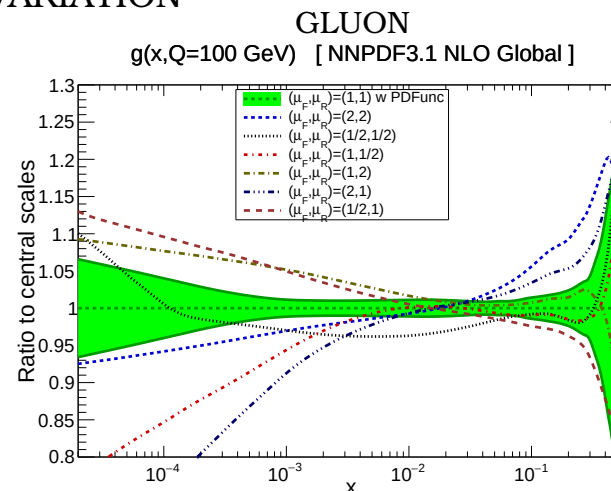
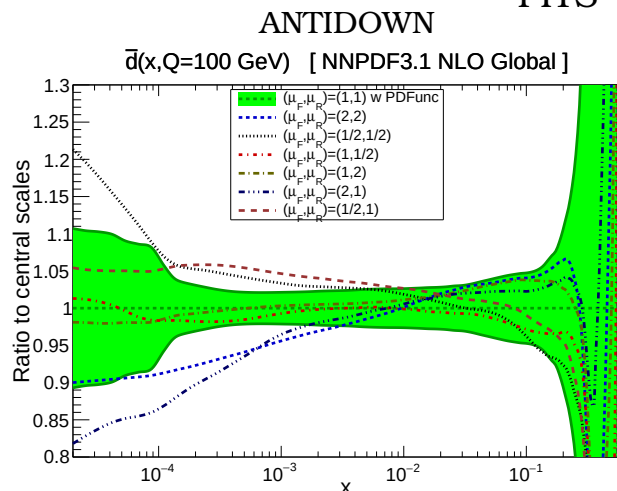
- TODAY: NLO PDF & MHOU UNCERTAINTIES COMPARABLE
- NEAR FUTURE: SHOULD WE WORRY ABOUT NNLO MHOU?

WHAT IS THE IMPACT OF SCALE VARIATION ON PDFs?

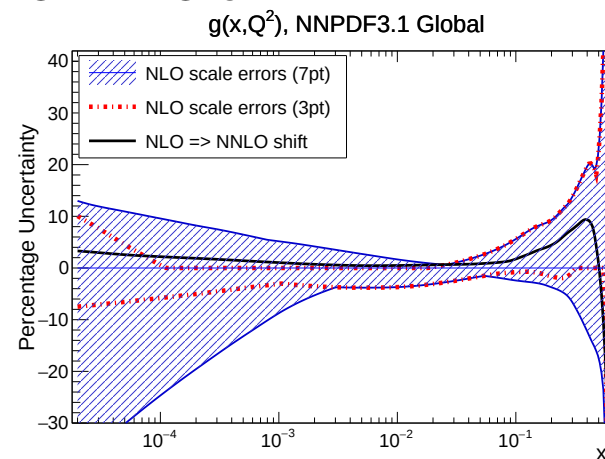
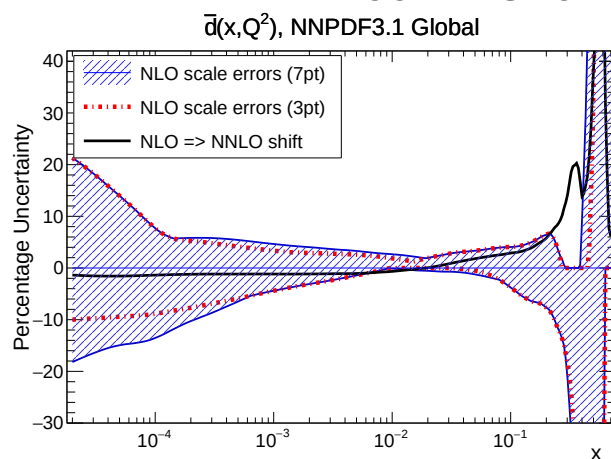
SIMPLEST IDEA FOR PDF MHOU ESTIMATE

- PERFORM FIT WITH VARIOUS SCALE CHOICES
- TAKE ENVELOPE OF RESULTS
- COMPARE TO NLO-NNLO SHIFT IN ORDER TO TUNE RANGE OF VARIATION & CORRELATION PATTERN: VARIATION FULLY CORRELATED ACROSS PROCESSES?

FITS WITH SCALE VARIATION



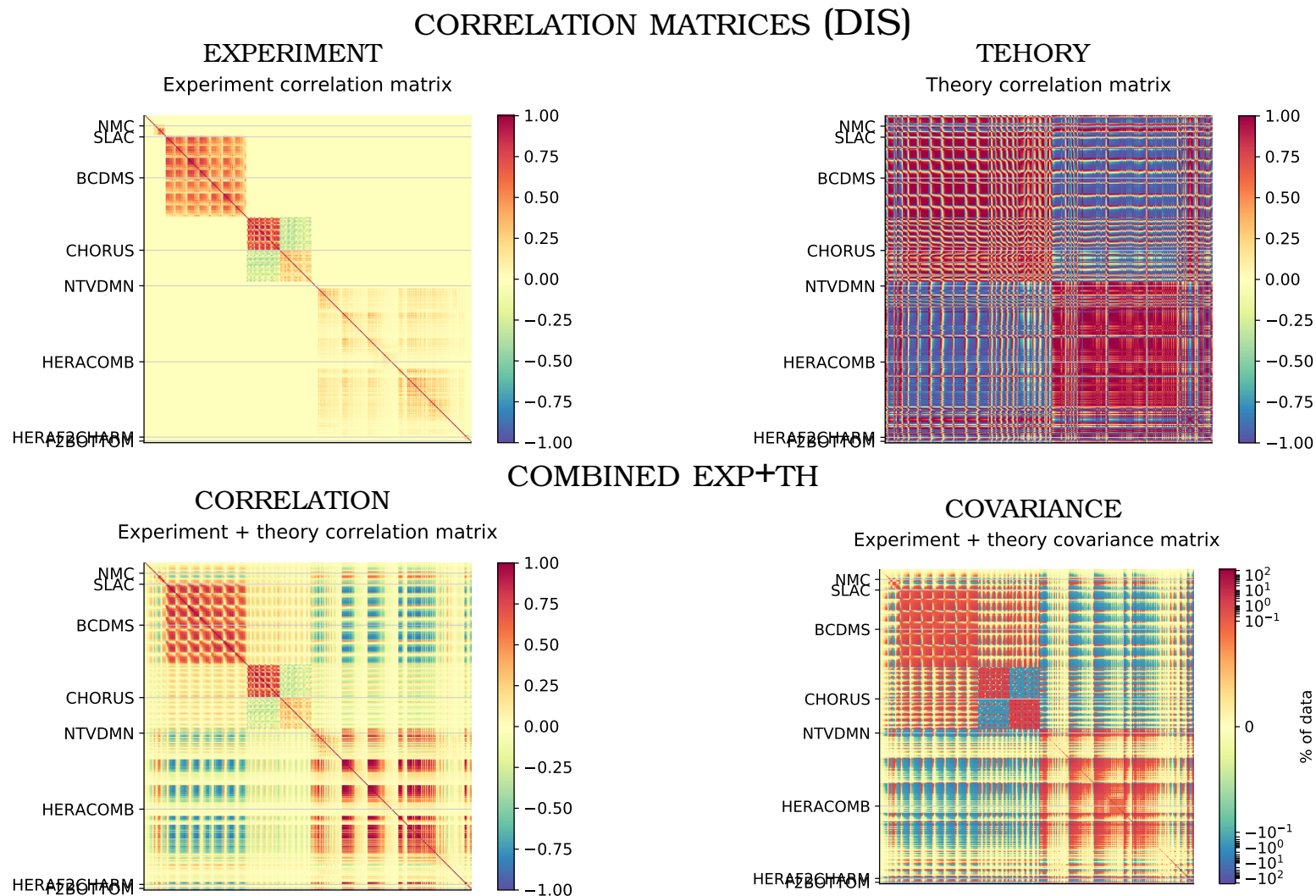
SCALE UNCERTAINTY VS NLO-NNLO SHIFT



HOW CAN WE INCLUDE MHOu IN PDF ERRORS?

BETTER IDEA FOR PDF MHOu ESTIMATE

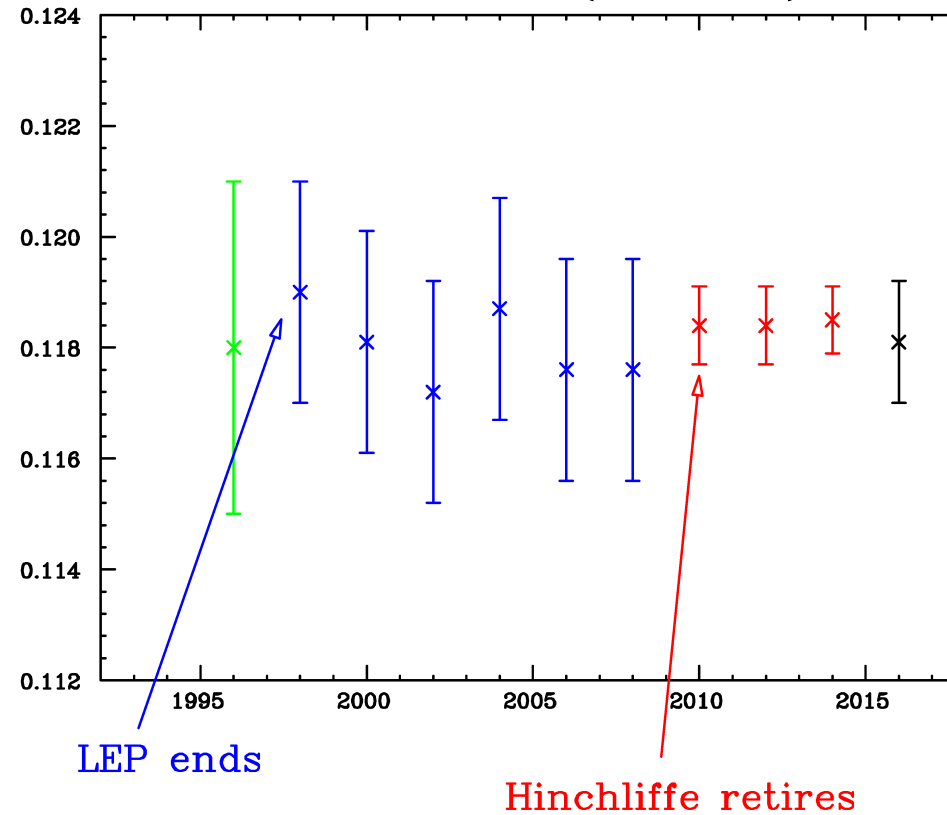
- PERFORM CALCULATION WITH VARIOUS SCALE CHOICES
- USE DISTRIBUTION OF RESULTS TO COMPUTE THEORY COVARIANCE MATRIX
- PERFORM FIT WITH EXTRA CONTRIBUTION TO COVARIANCE MATRIX



$$\alpha_s$$

THE PROBLEM OF α_s

THE PDG VALUE (VS.TIME)



THE PROBLEM

EXAMPLE: HIGGS IN GLUON FUSION:

IF NO CHANGE IN α_s UNCERTAINTY,

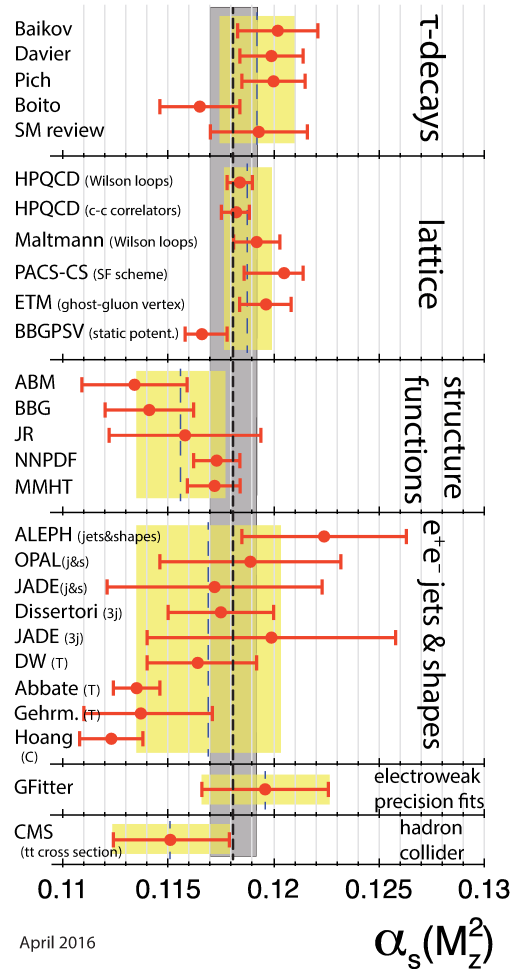
$$\Delta_{\text{ggF}}^{\text{PDF}} \sigma \approx 0.5 \text{ pb (1\%)} [\text{vs YR4 } 0.9 \text{ pb (1.9\%)}]$$

LEADS TO

$$\Delta_{\text{ggF}}^{\text{PDF}+\alpha_s} \sigma \approx 1.4 \text{ pb (2.8\%)} [\text{vs YR4 } 1.6 \text{ pb (3.2\%)}]$$

WHERE DOES THE PDF AVERAGE COME FROM?

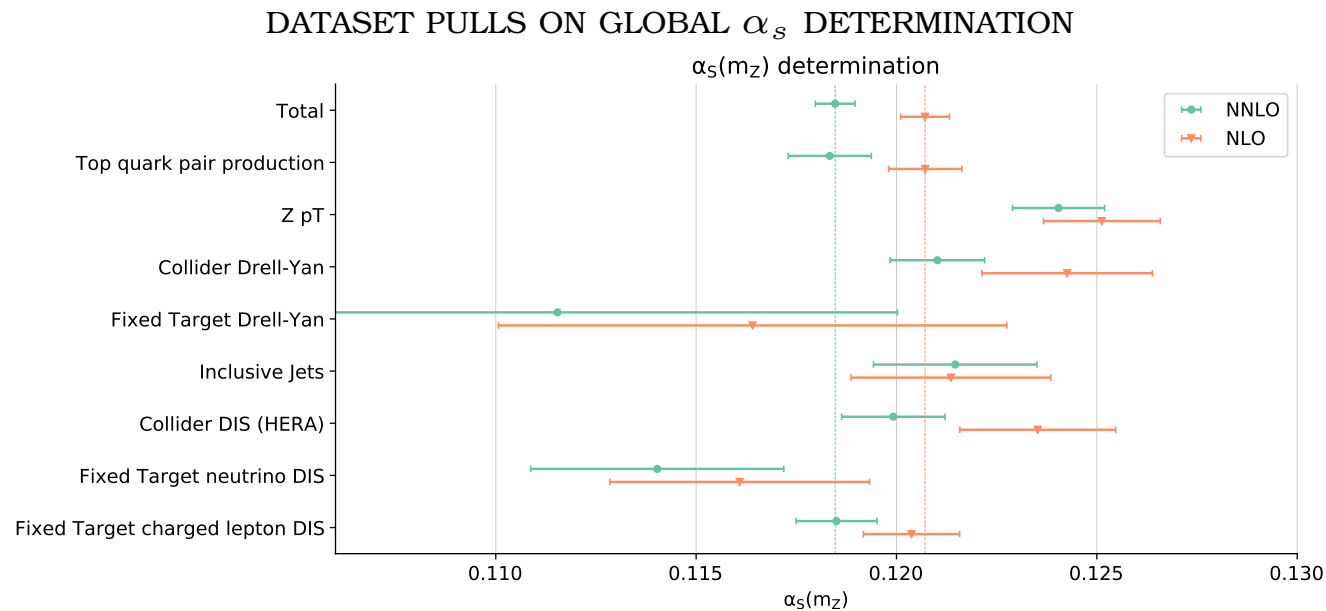
PDG PRE-AVERAGES



- PDG AVERAGE: χ^2 -AVERAGING OF SIX PRE-AVERAGES, CHOSEN TO BE MAXIMALLY UNCORRELATED
- χ^2 AVERAGING $\Rightarrow$ UNCERTAINTIES INFLATED OR CORRELATED UNTIL $\chi^2/dof = 1$
- EACH PRE-AVERAGE IS THE SIMPLE (UNWEIGHTED) AVERAGE OF ITS COMPONENTS, UNCERTAINTY $\leftrightarrow$ AVERAGE OF UNCERTAINTIES (OR STANDARD DEVIATION OF VALUES IF LARGER)

WHAT'S THE PROBLEM?

- GLOBAL EW FIT THEORETICALLY SAFE, BUT NOT VERY PRECISE
- τ QUITE PRECISE, BUT LOW SCALE $\Rightarrow$ PERTURBATIVE ACCURACY?
- LATTICE: VERY PRECISE, BUT WHAT IS THE ACCURACY OF THE TRUNCATION?
- ANY PROCESS WITH HADRONS IN THE FINAL STATE: DEPENDENCE ON THE PDF



(NNPDF, 2018)

PULLS DO NOT ADD TO ZERO $\Leftrightarrow$ BEST FIT PDF FOR DATASET IS NOT GLOBAL BEST FIT
 α_s USING EXTERNAL PDF SET POTENTIALLY BIASED DUE TO α_s -PDF CORRELATION

OUTLOOK

WHAT I TALKED ABOUT AND WHEN

- “PDF” UNCERTAINTIES
 - RELIABLE EXPERIMENTAL COVARIANCE MATRICES $\Rightarrow$ LHC RUN II
 - OPTIMIZED MINIMIZATION $\Rightarrow$ LHC RUN II
- THEORY UNCERTAINTIES
 - THEORY PREDICTIONS WITH SUB-PERCENT NUMERICAL ACCURACY $\Rightarrow$ NOW – LHC RUN II
 - FITTED HEAVY QUARK PDFs $\Rightarrow$ NNPDF3.1 (2018), PDF4LHC20
 - PDFs WITH THEORY UNCERTAINTY $\Rightarrow$ NNPDF4.0 (2019)
- CONSENSUS α_s WITH SUB-PERCENT ACCURACY $\Rightarrow$ HL-LHC

WHAT I DID NOT TALK ABOUT AND WHY

- HIGHER TWISTS, NUCLEAR CORRECTIONS $\Rightarrow$ HISTORY!
 - FIXED-TARGET DATA OBSOLETE
 - COLLIDER-ONLY PDFs
- RESUMMED PDFs, N³LO PDFs, EW CORRECTIONS $\Rightarrow$ LABORIOUS BUT TRIVIAL
 - AUTOMATIZATION OF HO CORRECTIONS
 - PUBLIC CODES
- PS-PDFs, MC-PDFs $\Rightarrow$ SOME WAY TO GO

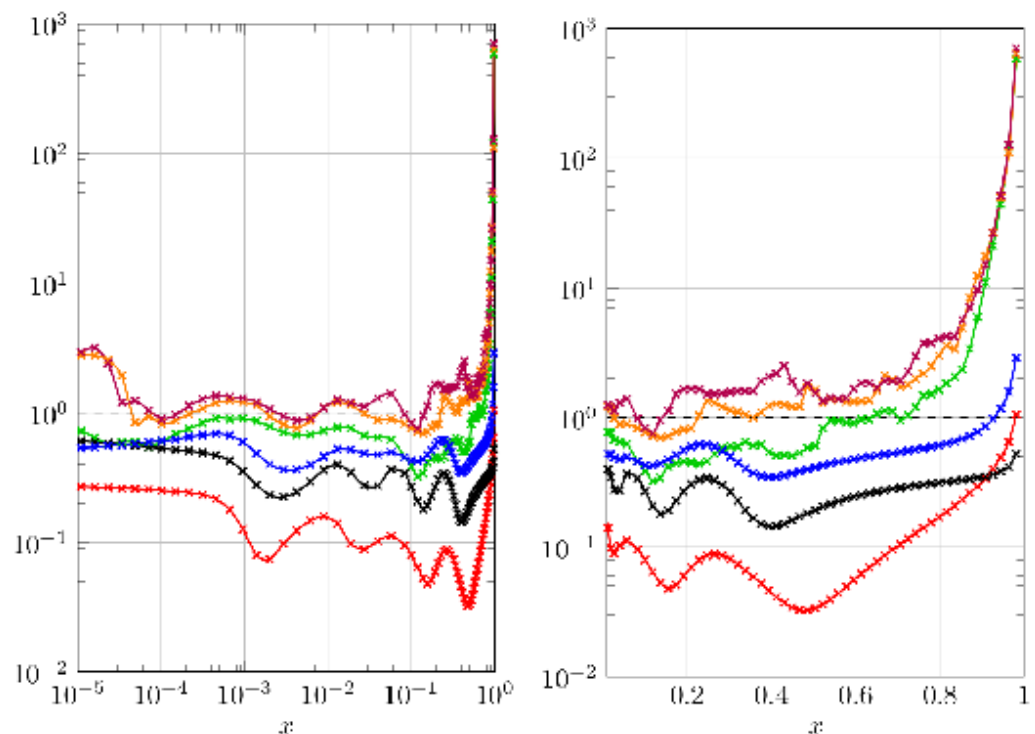
EXTRAS

CLOSURE TESTING THE SOURCES OF UNCERTAINTY

- LEVEL 0: ZERO UNCERTAINTY
 - CHECK WHETHER MINIMIZATION EFFICIENT
 - CHECK FOR INTERPOLATION UNCERTAINTY
- LEVEL 1: DATA UNCERTAINTY, BUT NO REPLICAS
 - CHECK FOR UNIQUENESS OF BEST FIT $\Rightarrow$ “FUNCTIONAL” UNCERTAINTY
- LEVEL 2: AS IN STANDARD PROCEDURE
 - CHECK WHETHER TRUE VALUE GAUSSIANLY DISTRIBUTED ABOUT FIT
 - CHECK WHETHER UNCERTAINTIES FAITHFUL

CLOSURE-TESTING: THE PARAMETRIZATION DEPENDENCE

GLUON PDF UNCERTAINTY NORMALIZED TO MSTW08



(C. Mascaretti, 2016)

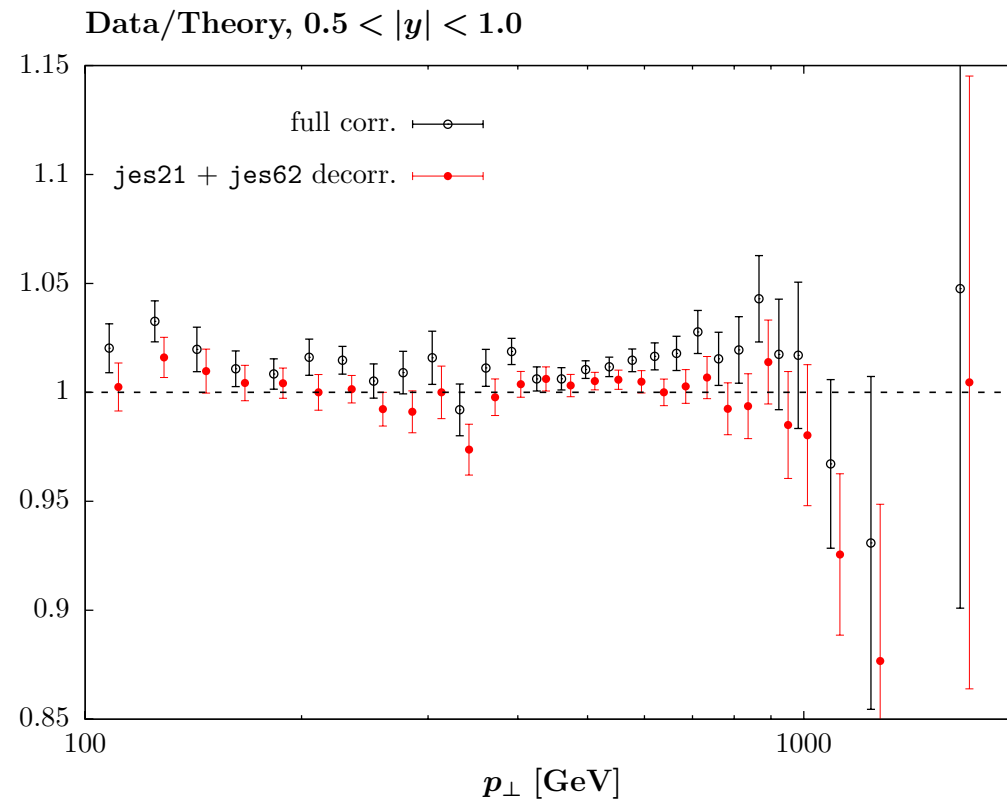
- CLOSURE TEST PERFORMED WITH DATA GENERATED BASED ON MST08 FUNCTIONAL FORM
- REFITTED EITHER WITH NNPDF OR MSTW-CT FUNCTIONAL FORM
- LEVEL 0: VANISHING DATA UNCERTAINTY
 - MSTW-CT: FIT HAS ZERO UNCERTAINTY
 - NNPDF: ABOUT HALF OF TOTAL UNCERTAINTY
- LEVEL 1: NOMINAL DATA UNCERTAINTY, BUT REPLICAS FITTED W/O PSEUDODATA
 - MSTW-CT: FIT HAS SMALL UNCERTAINTY
 - NNPDF: ABOUT 2/3 OF FINAL UNCERTAINTY
- LEVEL 2
 - NNPDF UNCERTAINTY LARGER THAN MSTW-CT
 - NNPDF UNCERTAINTY SIMILAR TO MSTW WITH TOLERANCE

“STANDARD” PARAMETRIZATION
MISSES INTERPOLATION &
FUNCTIONAL UNCERTAINTY?

CORRELATIONS & THE COVARIANCE MATRIX

THE ATLAS 7TeV JETS

- EACH RAPIDITY BIN CAN BE FITTED WITH $\chi^2/dof \sim 1$
- EACH LEADS TO INDISTINGUISHABLE BEST-FIT PDFS
- IF ALL BINS FITTED SIMULTANEOUSLY, $\chi^2/dof \sim 3$



(Harland-Lang, Martin, Thorne, 1016)

- MISESTIMATED CORRELATIONS?
- CAN SINGLE OUT WHICH CORRELATION OUGHT TO BE REMOVED