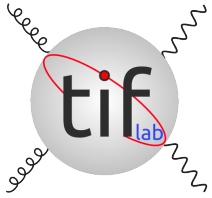
LESSONS LEARNT TOWARDS PDF4LHC20

STEFANO FORTE
UNIVERSITÀ DI MILANO & INFN

UNIVERSITÀ DEGLI STUDI DI MILANO
DIPARTIMENTO DI FISICA



PDF4LHC MEETING

CERN, DECEMBER 13, 2018

BENCHMARKING AND COMBINATION
SOME HISTORY

- 2011:
  - BENCHMARKING 1101.0536
  - PDF4LHC RECOMMENDATION (ENVELOPE) 1101.0538
- 2012-2014:
  - HXSWG BENCHMARKING: PDF CORRELATIONS 1201.3084
  - GLOBAL PDF SET BENCHMARKING: CODES, STATISTICAL METHODS & STANDARD CANDLES 1211.5142
  - HXSWG BENCHMARKING: PDF4LHC RECOMMENDATION 1307.1347
  - LES HOUCHES 2013 BENCHMARK: HQ SCHEME, COUNTING OF PERT. ORDERS, EW CORRECTIONS, CUTS, SCALE CHOICES, STATISTICAL TREATMENT, DATA 1405.1067
- 2015: PDF4LHC15
  - BENCHMARK & RECOMMENDATION 1507.03865
  - COMPARISONS TO LHC RUN I & PREDICTIONS FOR RUN II 1507.00556

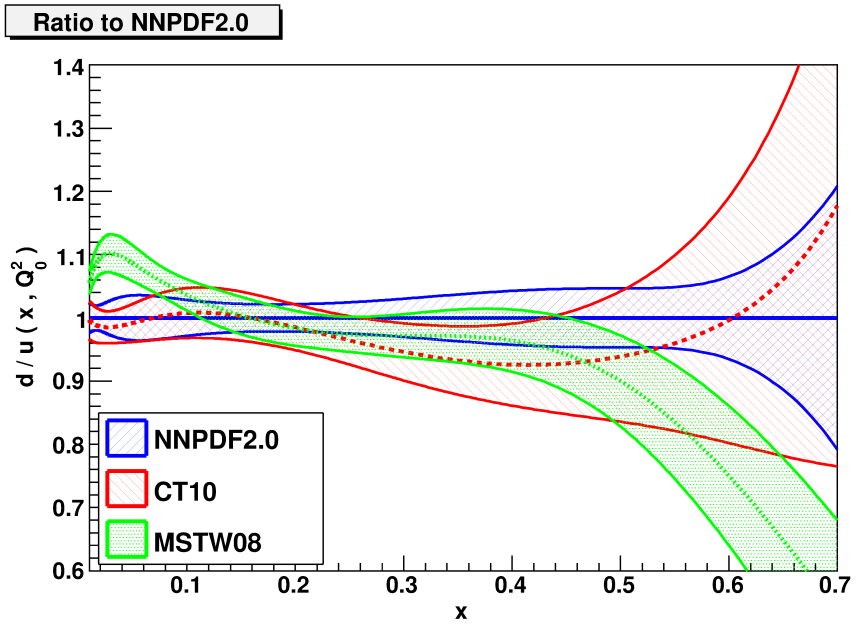
QUESTIONS

- SHOULD EVERYBODY USE THE SAME METHODOLOGY? XFITTER?
- SHOULD EVERYBODY USE THE SAME DATASET?
- CAN WE COMPUTE CORRELATIONS BETWEEN PDFS? DO WE NEED THEM FOR A COMBINATION?
- DO WE NEED THEORY UNCERTAINTIES ON PDFS?
- DO WE NEED A STUDY OF FUTURE DATA?

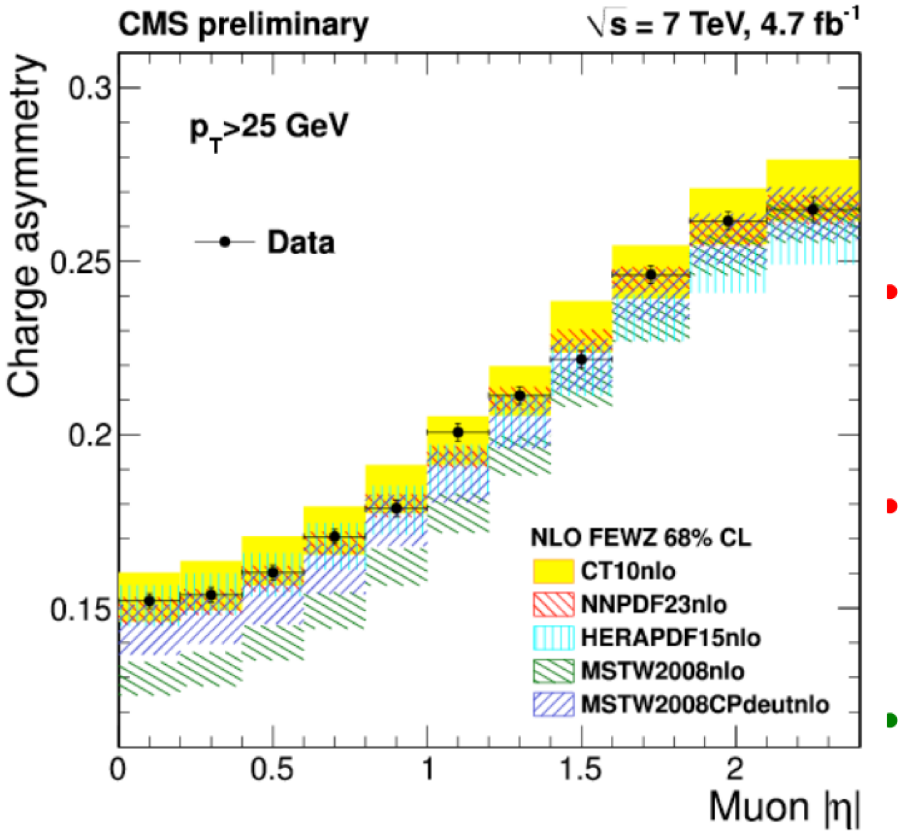
SHOULD EVERYBODY USE THE SAME METHODOLOGY?

HISTORY: WHY NOT

THE d/u RATIO IN 2011



THE CMS W ASYMMETRY IN 2012



- DISCREPANCY IN THE d/u RATIO BETWEEN MSTW AND OTHER GLOBAL FITS
- TRACED TO A PARAMETRIZATION ISSUE, RESOLVED IN MSTW08DEUT SET

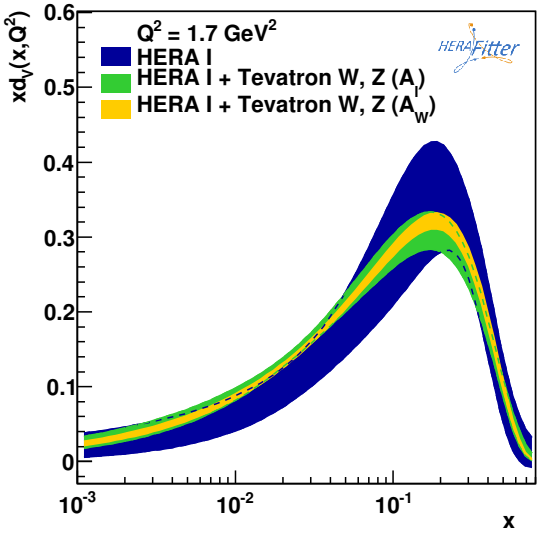
SIMILAR EXAMPLES WITH ANY PDF SET!

WHAT ABOUT XFITTER?

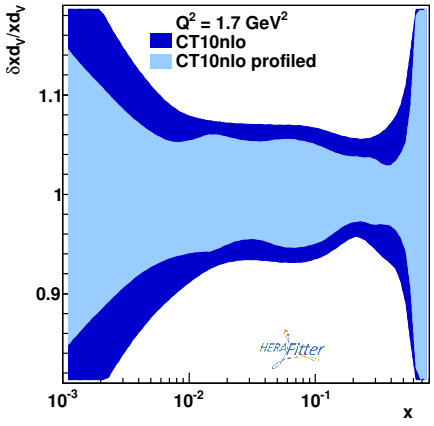
- OFTEN USED TO ASSESS IMPACT OF X IN "HERA+X" FITS

IMPACT OF THE TEVATRON W ASYMMETRY

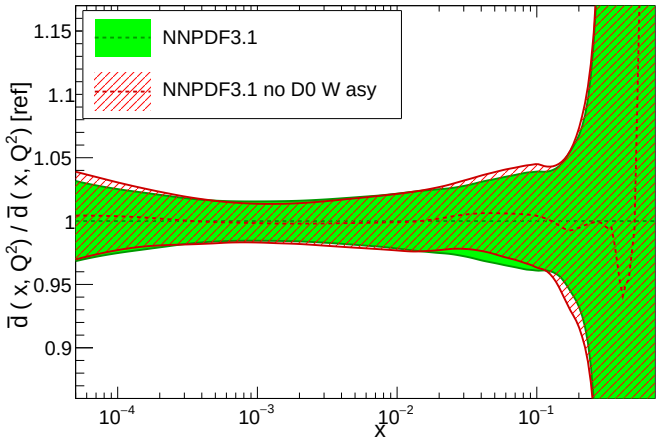
XFITTER: IMPACT ON HERA



XFITTER: IMPACT ON CT10



NNPDF3.1: IMPACT ON GLOBAL FIT



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

LESSONS LEARNT

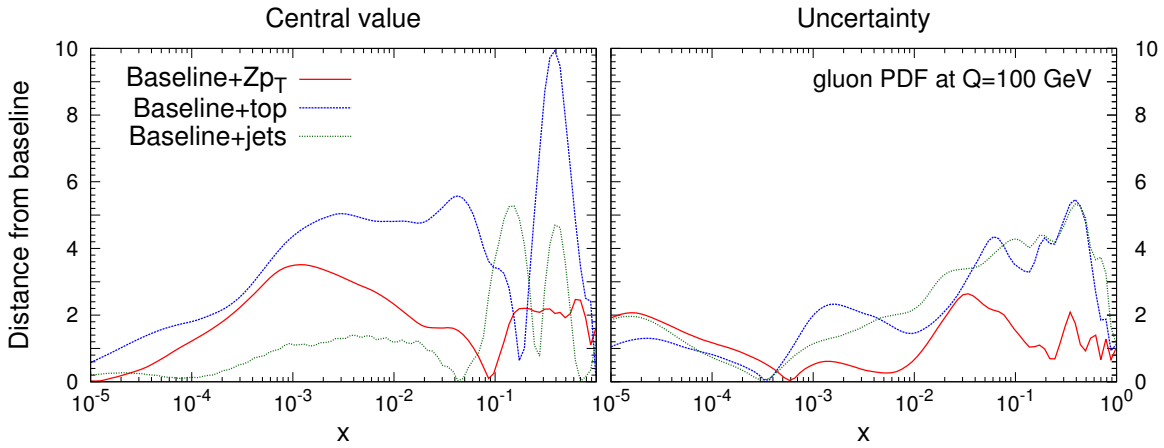
- DIFFERENCES IN PDF PREDICTIONS DRIVEN BY METHODOLOGY $\Leftrightarrow$ PDF UNCERTAINTIES "INFINITE" (FINITE DATA, INFINITE INFORMATION)
- SAME METHODOLOGY $\Rightarrow$ BIAS+UNDERESTIMATED UNCERTAINTIES
- XFITTER ONLY METHODOLOGY COULD BE RESTRICTIVE
- HERA+X FITS COULD BE POTENTIALLY MISLEADING

SHOULD EVERYBODY USE THE SAME DATASET?

EXAMPLE: TOP PRODUCTION AND THE GLUON

INCLUSION IN THE NNPDF3.1 SET: COMPARISON OF IMPACT VS. JETS, Z p_t

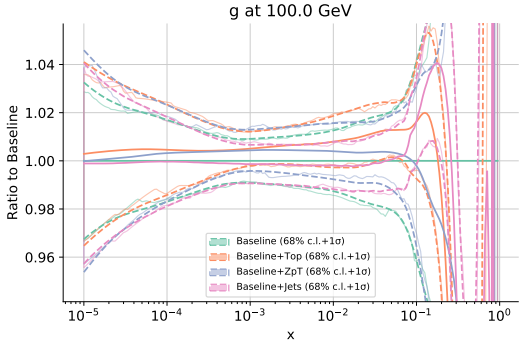
DISTANCES (difference in units of st. dev.)



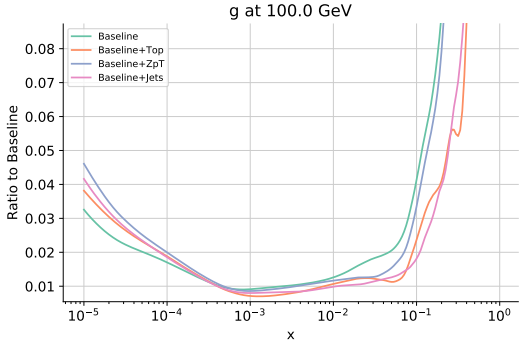
(Nocera, Ubiali, 2017)

PDF COMPARISON

CENTRAL VALUE



UNCERTAINTY



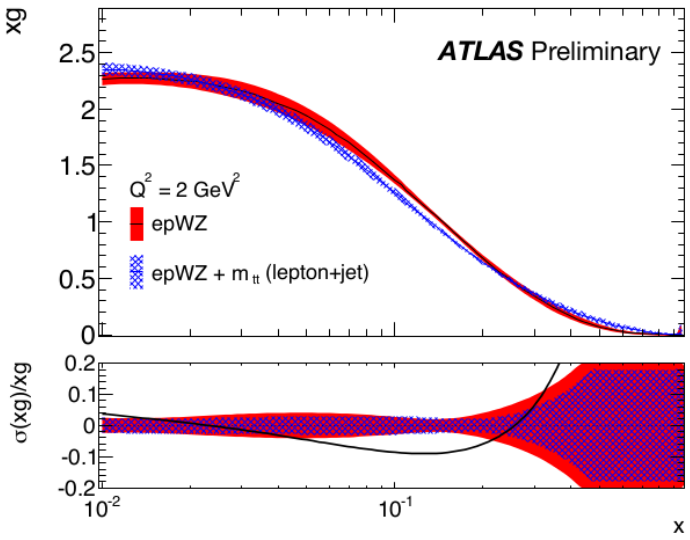
- TOP HAS LARGEST IMPACT, FOLLOWED BY JETS
- ALL LHC DATA PULL CENTRAL VALUE IN SAME DIRECTION!

TOP PRODUCTION AND THE GLUON

CONSISTENCY OF DIFFERENT OBSERVABLES

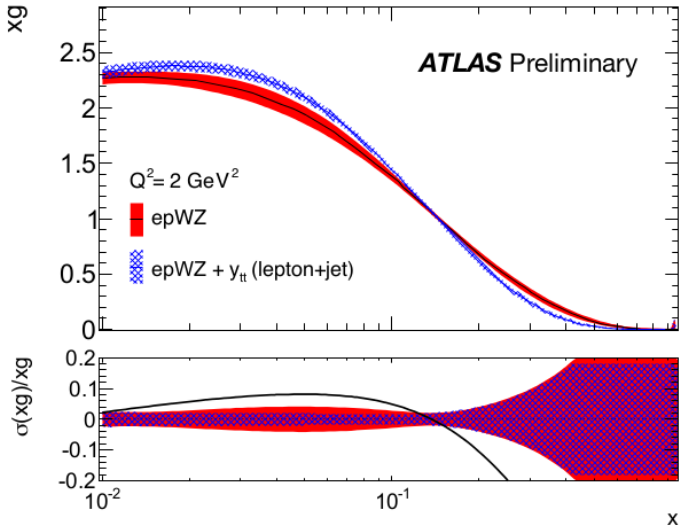
INCLUSION OF ATLAS TOP DATA IN HERA+TOP FIT (XFITTER)

HQ PAIR RAPIDITY DISTN



(a)

INVARIANT MASS DISTN.



(a)

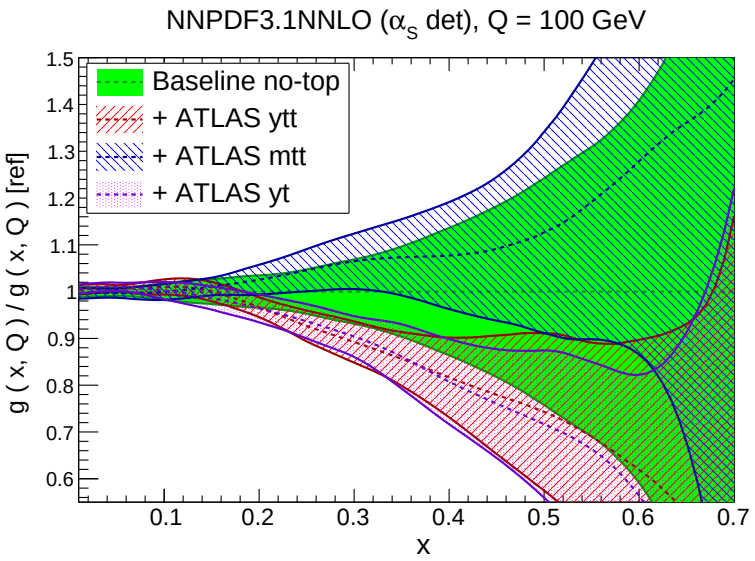
INCONSISTENCY?

TOP PRODUCTION AND THE GLUON

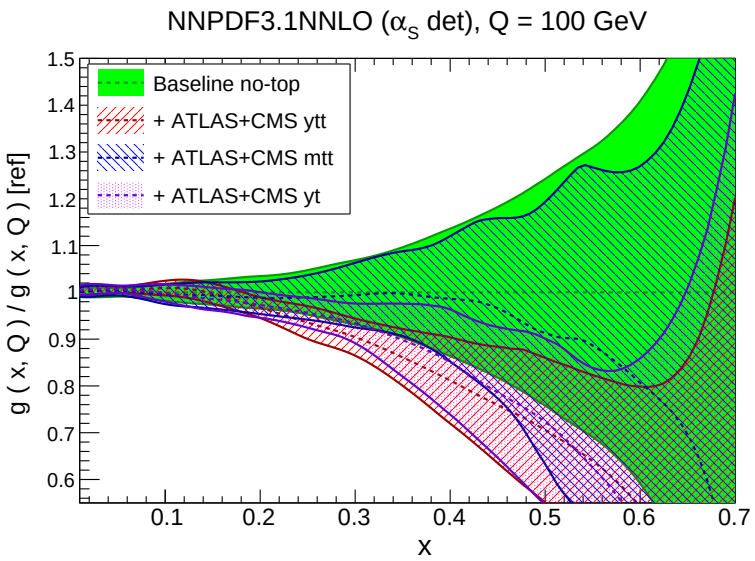
CONSISTENCY OF DIFFERENT OBSERVABLES

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

ATLAS ONLY



ATLAS+CMS



CONSISTENCY!

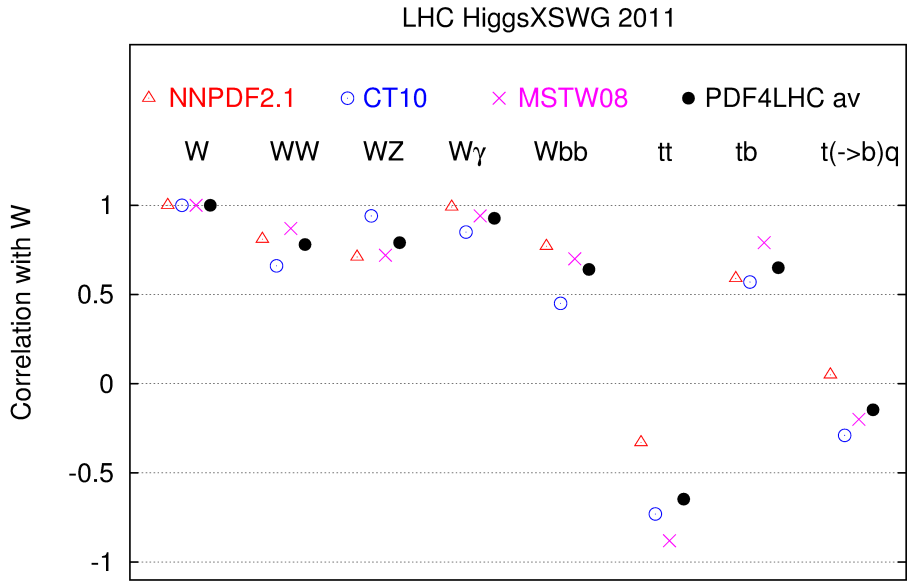
ATLAS INVARIANT MASS HAS VERY LITTLE PULL
$\Rightarrow$ RESULTS CONSISTENT WITHIN UNCERTAINTIES

LESSONS LEARNT

- WIDEST DATASET IN PRINCIPLE BEST, BUT
- NOT ALL METHODOLOGIES MAY ACCOMMODATE ALL DATA
- DATA-METHODOLOGY INTERPLAY $\Rightarrow$ CAREFUL BENCHMARKING

CORRELATING PDFS

CORRELATION BETWEEN HIGGS SIGNAL AND BACKGROUND (HXSWG, YR2)



- CORRELATION BETWEEN PROCESSES AND PDFS, PROCESSES AND PROCESSES, PDF AND PDFS TRIVIAL TO COMPUTE ⇒ 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 ⇒ SMALLER UNCERTAINTY COULD JUST BE BIAS!
  - UNCERTAINTY REDUCED IF CORRELATION LESS THAN 100% CAN WE BELIEVE IT IN THE ABSENCE OF NEW INFORMATION?

LESSONS LEARNT

- DATA-DATA, DATA-PDF, PDF-PDF CAN BE COMPUTED WITHOUT ANY NEW FIT
- DIFFERENT PDF SETS BASED ON SAME DATA HIGHLY CORRELATED
- MORE PRECISE PDFS NOT NECESSARILY MORE ACCURATE $\Rightarrow$ WEIGHTED AVERAGE NOT ADVISABLE
- NON-100% CORRELATIONS LARGELY DRIVEN BY METHODOLOGY $\Rightarrow$ CORRELATED AVERAGE NOT ADVISABLE

MORE QUESTIONS & ANSWERS

- Q: DO WE NEED THEORY UNCERTAINTIES
- A: YES, SEE Harland-Lang AND Rojo's TALKS $\Rightarrow$ ALMOST DONE
- Q: DO WE NEED STUDIES OF FUTURE DATA?
- A: YES, SEE Rojo's TALK $\Rightarrow$ ALREADY DONE

FINAL LESSONS

- GET READY FOR A NEW COMBINATION
- PRELIMINARY BENCHMARKING NECESSARY $\Rightarrow$ INVOLVE EXPERIMENTS & EWWG
- MORE PRECISE DATA REQUIRE MORE ACCURATE COMBINATION