PARTON DISTRIBUTIONS

OR

THE ART OF SOLVING ILL-POSED PROBLEMS

Stefano Forte
Università di Milano & INFN

UNIVERSITÀ DEGLI STUDI DI MILANO
DIPARTIMENTO DI FISICA

Istituto Nazionale di Fisica Nucleare

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

PDFs NOW

- FROM PDF4LHC15 TO NEW GENERATION SETS
- THE IMPACT OF LHC DATA
- RESOLVING ISUES WITH DATA: STRANGENESS & CHARM

FITTING THE METHODOLOGY

- DATA VS. METHODOLOGY
- DATASET OPTIMIZATION
- METHODOLOGY OPTIMIZATION

THEORY UNCERTAINTIES

- PDF THEORY ERRORS
- SCALE VARIATION
- TESTING THEORY ERRORS

THE STATE OF THE ART

CONTEMPORARY PDF TIMELINE (ONLY PUBLISHED GLOBAL)

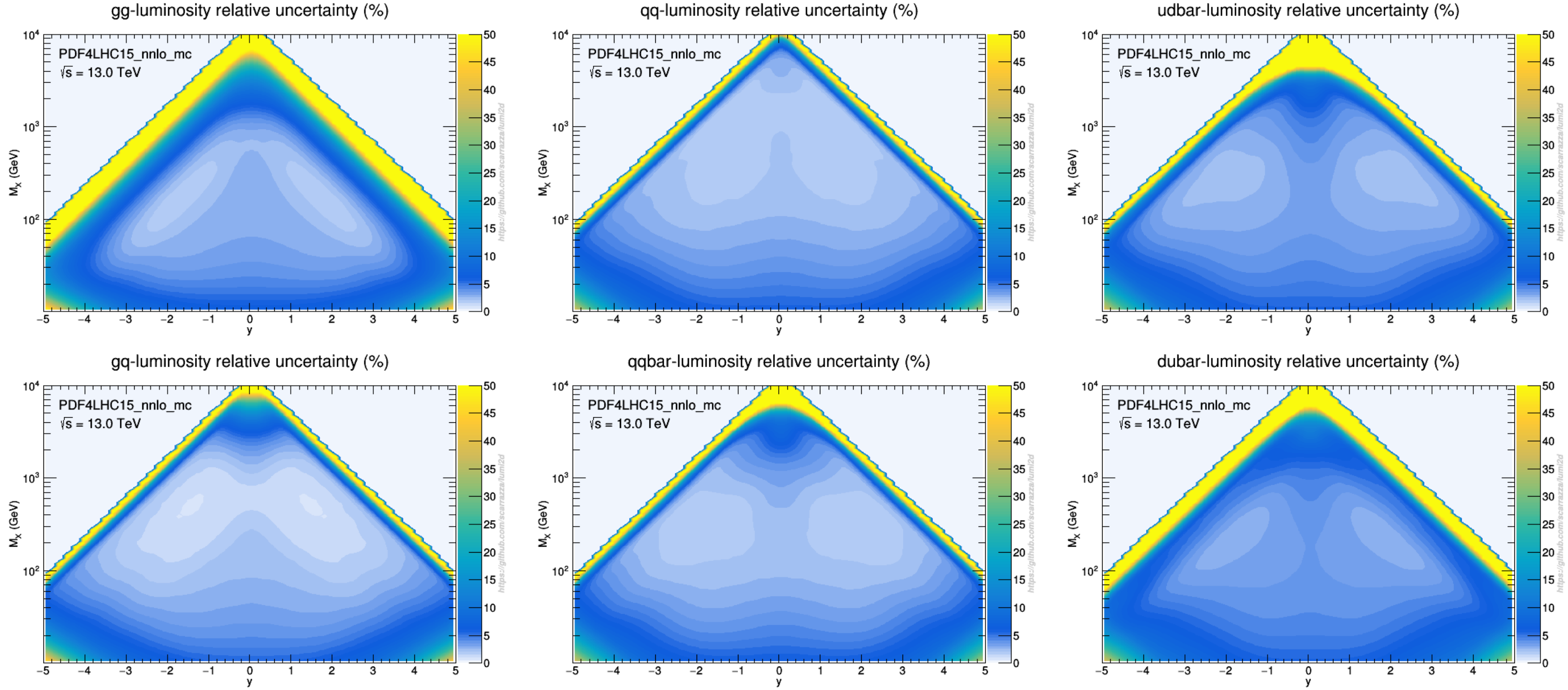
	2008		2009		2010		2011	2012		2013		2014		2015	2017	
SET / MONTH	CTEQ6.6 (02)	NNPDF1.0 (08)	MSTW (01)	ABKM09 (08)	NNPDF2.0 (02)	CT10 (NLO) (07)	NNPDF2.1 (NNLO) (07)	ABM11 (02)	NNPDF2.3 (07)	CT10 (NNLO) (02)	ABM12 (10)	NNPDF3.0 (10)	MMHT (12)	CT14 (06)	ABMP16 (01)	NNPDF3.1 (06)
F. T. DIS	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
ZEUS+H1-HI	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
COMB. HI	✗	✗	✗	✗	✔	✗	✔	✗	✔	✗	✔	✔	✗	✗	✔	✔
ZEUS+H1-HII	✗	✗	✗	✗	✗	✗	some	✗	✗	some	✗	✔	✗	✗	✔	✔
HERA JETS	✗	✗	✔	✗	✗	✗	✗	✗	✗	✗	✗	✗	✔	✗	✗	✗
F. T. DY	✔	✗	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
TEV W+Z	✔	✗	✔	✗	✔	✔	✔	✗	✔	✔	✗	✔	✔	✔	✗	✔
LHC W+Z	✗	✗	✗	✗	✗	✗	✗	✗	✔	✗	some	✔	✔	✔	some	✔
TEV JETS	✔	✗	✔	✗	✔	✔	✗	✔	✔	✔	✗	✔	✔	✔	✗	✔
LHC JETS	✗	✗	✗	✗	✗	✗	✗	✗	✔	✗	✗	✔	✔	✔	✗	✔
TOP TOTAL	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✔	✔	✗	✗	✔	✔
SINGLE TOP TOTAL	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✔	✗
TOP DIFFERENTIAL	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✔
W p_T	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✔	✗	✗	✗	✗
W+C	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✔	✗	✗	✗	✗
Z p_T	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✔

THEORY PROGRESS:

- MSTW, ABKM: all NNLO; NNPDF NNLO since 07/11 (2.1), CT since 02/13 (CT10);
  NNPDF THRESHOLD RESUMMATION (3.0RESUM, 07/15), SMALL x RESUMMATION (3.1SX, 10/17)
- MSTW, CT, NNPDF all GM-VFN; NNPDF since 01/11 (2.1);
  ABM FFN+ZM-VFN since 01/17 (ABMP16)
- NNPDF FITTED CHARM since 05/16 (NNPDF3IC)
- PHOTON PDF: (mrst2004qed), NNPDF2.3QED (08/13), NNPDF3.0QED (06/16), NNPDF3.1LUXQED (12/17)

PDF4LHC15: PDF UNCERTAINTIES (NNLO)

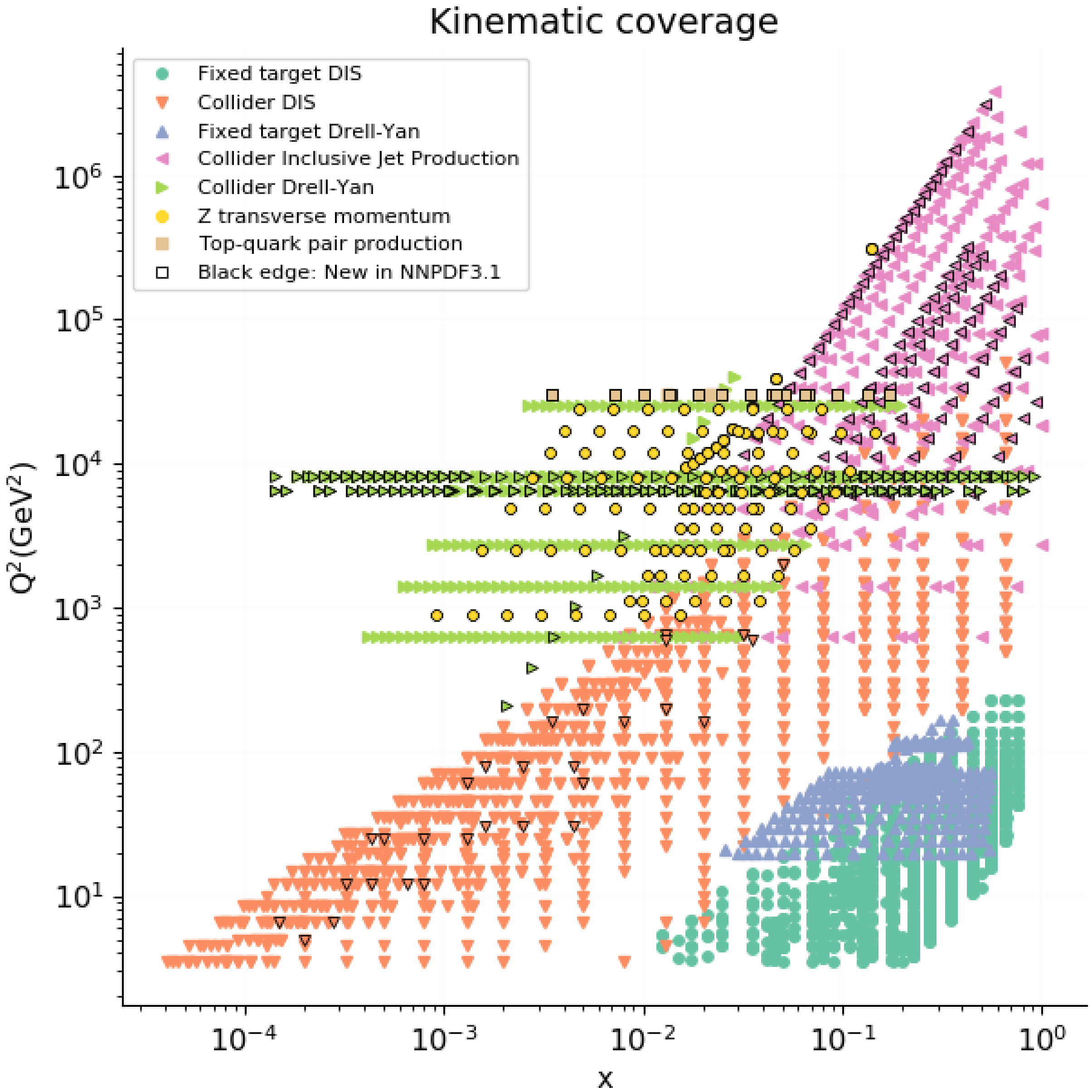
GLUON SINGLET FLAVORS



- GLUON BETTER KNOWN AT SMALL x, VALENCE QUARKS AT LARGE x, SEA QUARKS IN BETWEEN
- TYPICAL UNCERTAINTIES IN DATA REGION $\sim 3-5\%$
- SWEET SPOT: VALENCE Q - G; DOWN TO 1%
- UP BETTER KNOWN THAN DOWN; FLAVOR SINGLET BETTER THAN INDIVIDUAL FLAVORS
- NO QUALITATIVE DIFFERENCE BETWEEN NLO AND NNLO

DATASET WIDENING

NNPDF3.0 vs NNPDF3.1

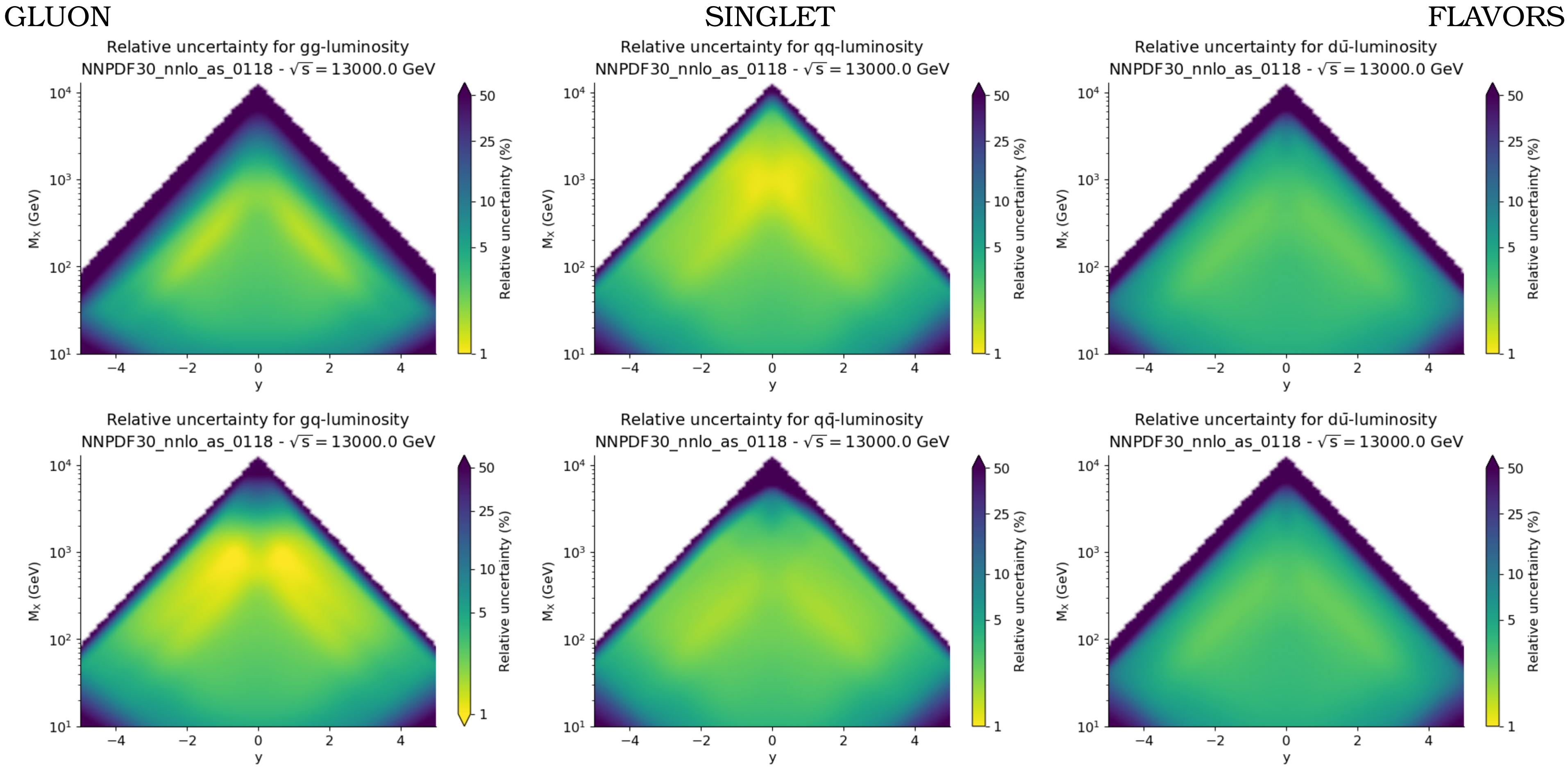


NEW DATA: (BLACK EDGE)

- 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

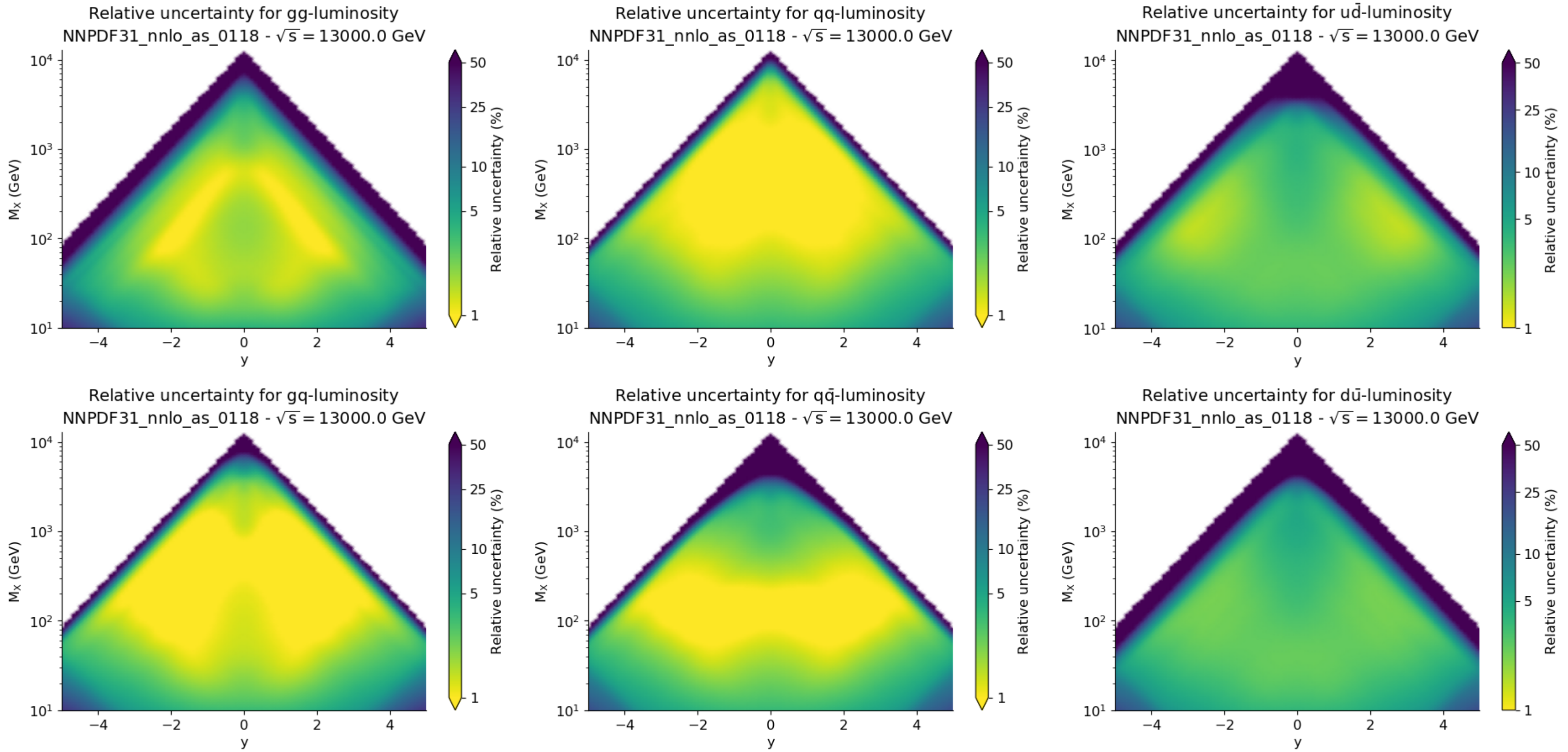
PDF UNCERTAINTIES IN DETAIL: NNPDF3.0 (NNLO)



- GLUON BETTER KNOWN AT SMALL x, VALENCE QUARKS AT LARGE x, SEA QUARKS IN BETWEEN
- TYPICAL UNCERTAINTIES IN DATA REGION $\sim 3-5\%$
- SWEET SPOT: VALENCE Q - G; DOWN TO 1%
- UP BETTER KNOWN THAN DOWN; FLAVOR SINGLET BETTER THAN INDIVIDUAL FLAVORS

THE IMPACT OF LHC DATA
PDF UNCERTAINTIES IN DETAIL: NNPDF3.1 (NNLO)

GLUON | SINGLET | FLAVORS

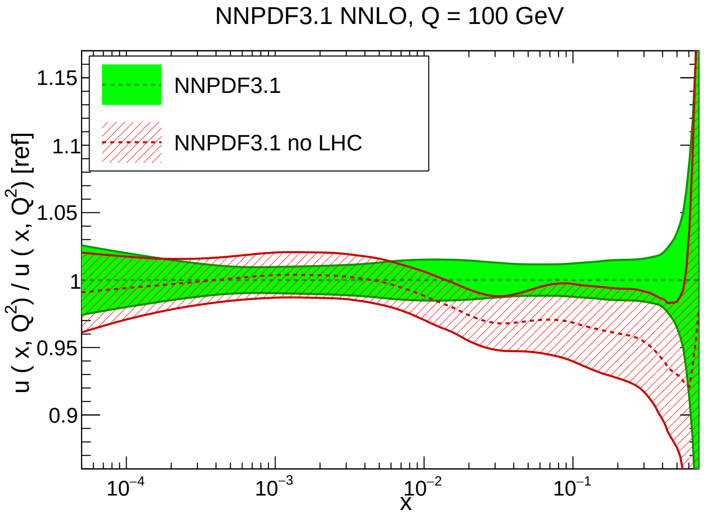


- GLUON BETTER KNOWN AT SMALL x, VALENCE QUARKS AT LARGE x, SEA QUARKS IN BETWEEN
- TYPICAL UNCERTAINTIES IN DATA REGION $\sim 1-3\%$
- SWEET SPOT: VALENCE Q - G; 1% OR BELOW
- UP BETTER KNOWN THAN DOWN; FLAVOR SINGLET BETTER THAN INDIVIDUAL FLAVORS
- NEW LHC DATA $\Rightarrow$ SIZABLE REDUCTION IN UNCERTAINTIES

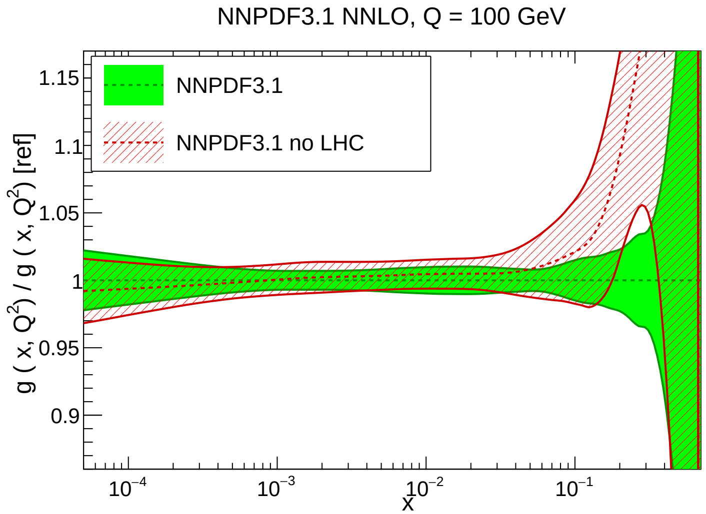
THE IMPACT OF LHC DATA

NEXT-GENERATION PDFS LARGELY DETERMINED BY LHC DATA: A FIRST!

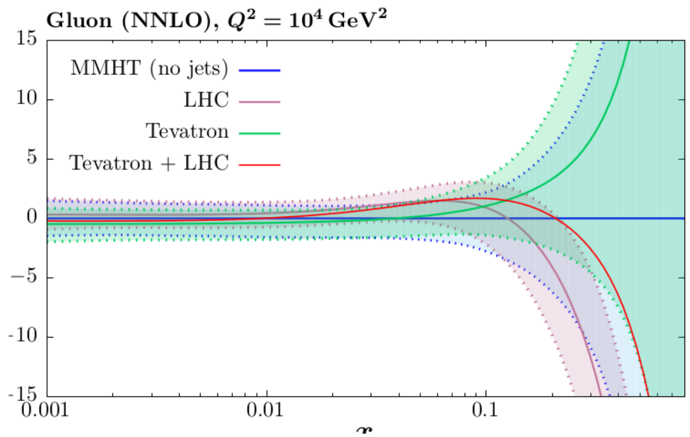
NNPDF3.1 up



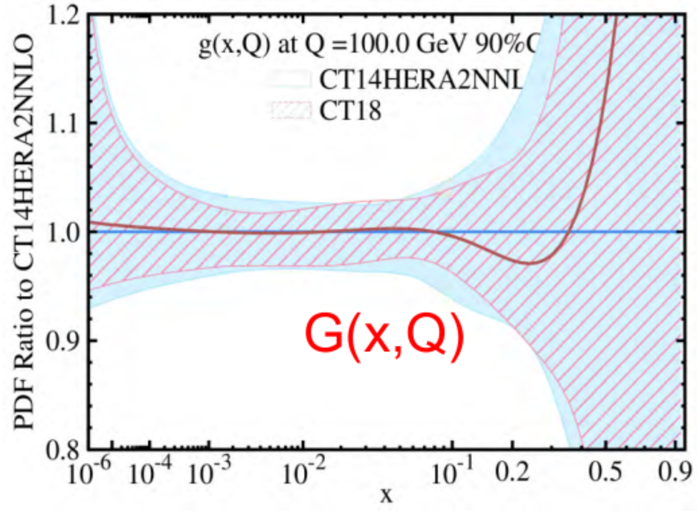
NNPDF3.1 glue



'MMHT' 19 glue (prelim., unpublished)



CT18 glue (preliminary, unpublished)



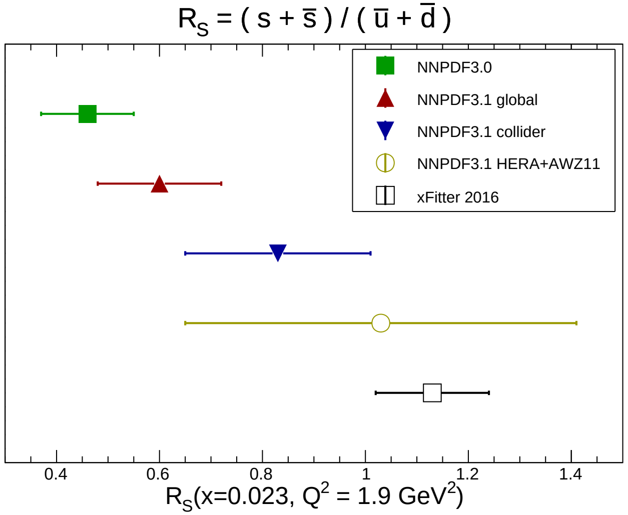
- SIGNIFICANT UNCERTAINTY REDUCTION
- MANY PDFS CHANGE BY MORE THAN ONE SIGMA
- BOTH FLAVOR SEPARATION & GLUON SIGNIFICANTLY AFFECTED

DATA VS. THEORY/METHODOLOGY
THE STRANGE PDF: DIS VS. W PRODUCTION

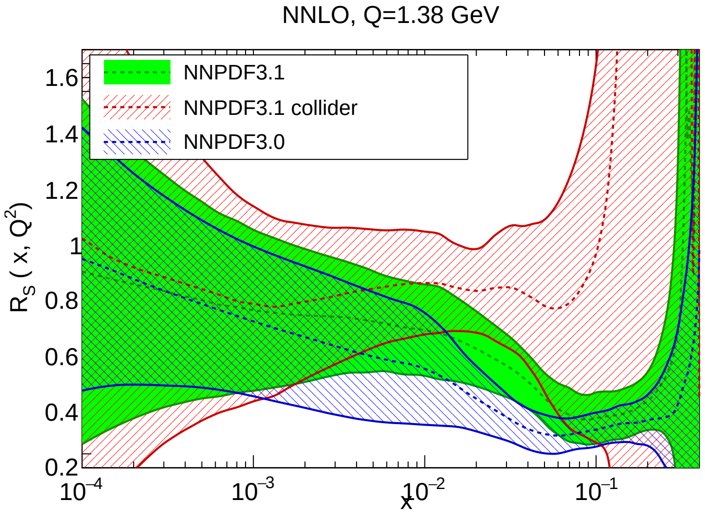
- STRANGE PDF CONTROLLED BY NEUTRINO DIS CHARM PRODUCTION + W PRODUCTION
- DIS DATA FAVOR "SUPPRESSED STRANGE" $\Rightarrow$ SMALL $R_s \equiv \frac{s+\bar{s}}{\bar{u}+\bar{d}}$
- ATLAS FAVORS ENHANCED STRANGENESS
- ATLAS IMPACT EXAGGERATED IN XFITTER ANALYSIS
- EVERYTHING CONSISTENT WITHIN UNCERTAINTIES IN GLOBAL FIT

THE STRANGENESS SUPPRESSION

XFITTER VS HERA+ATLAS VS. DIS ONLY VS ATLAS ONLY VS ALL



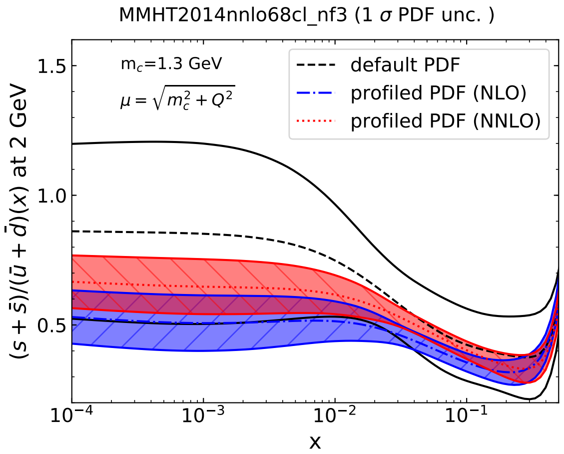
DIS ONLY VS ATLAS ONLY VS ALL



DATA VS. THEORY/METHODOLOGY
THE STRANGE PDF: DIS VS. W PRODUCTION

- MASSIVE CORRECTIONS TO CHARGED CURRENT DIS HITERTO INCLUDED TO NLO MASSLESS TO NNLO
- Gao, 2018 $\Rightarrow$ NNLO COMPUTED
- STRANGENESS ENHANCED BY NNLO CORRECTIONS

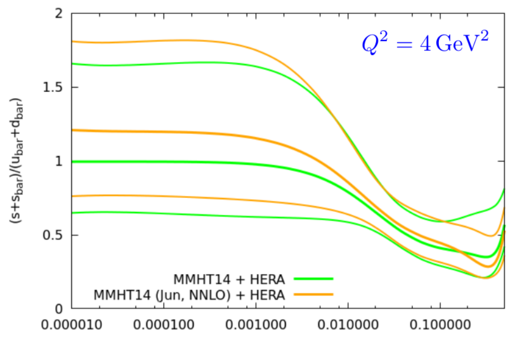
HERAPDF +NLO CC DIS VS NNLO CC DIS

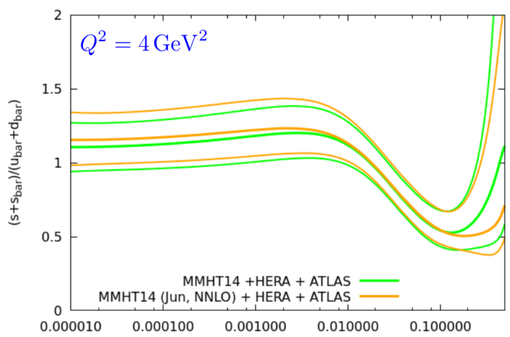


(Gao, 2108)

MMHT WITH NLO VS NNLO CC DIS

Preliminary





(Harland-Lang, Thorne, prelim.)

LESSONS:

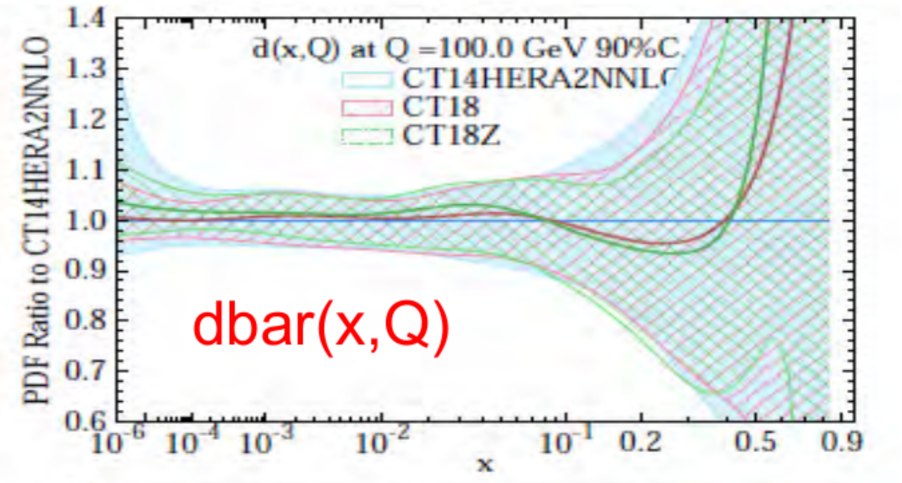
- BEWARE OF XFITTER HERA+X FITS
- IN A GLOBAL FIT DIFFERENT DATA ALWAYS PULL IN DIFFERENT DIRECTIONS!
- TENSION CAN BE RESOLVED BY BETTER THEORY

DATA VS. THEORY/METHODOLOGY
THE CHARM MASS AND TREATMENT
CT18 $\to$ CT18Z

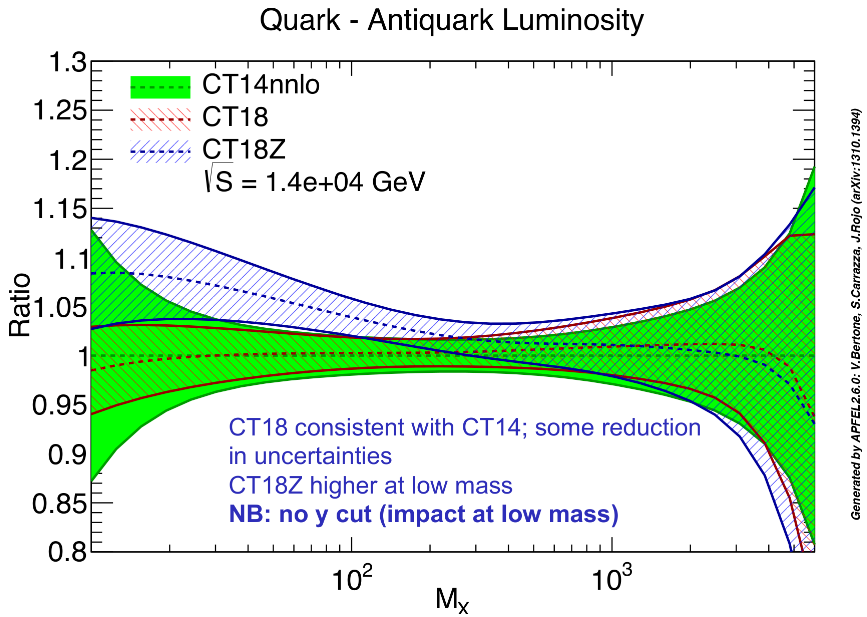
- ATLAS W AND Z 7TEV RAPIDITY INCLUDED
- CHARM MASS INCREASED
- x-DEPENDENT FACTORIZATION SCALE

CT18 VS. CT18Z (preliminary, unpublished)

DBAR PDF



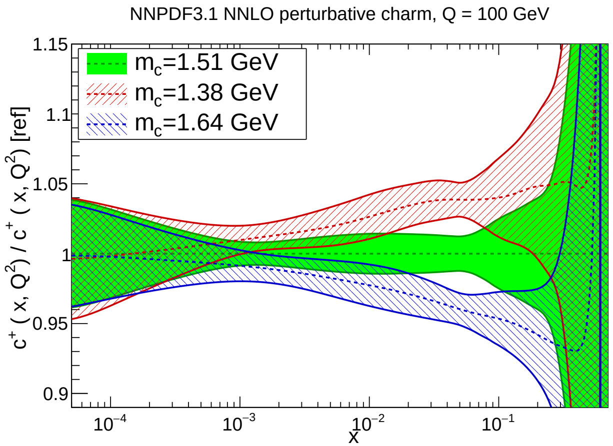
QQBAR LUMI



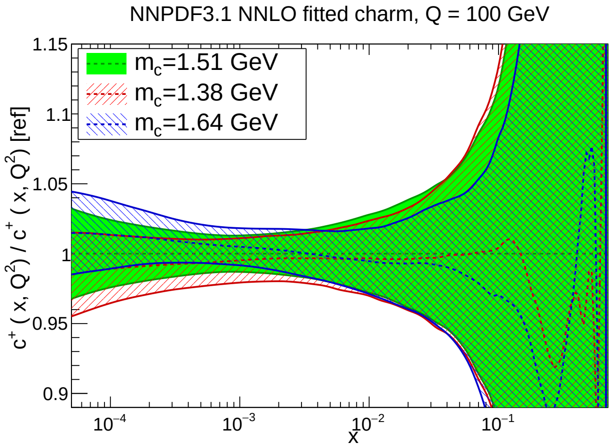
DATA VS. THEORY/METHODOLOGY
THE CHARM MASS AND TREATMENT
CHARM FROM DATA

- CHARM SHOULD NOT DEPEND STRONGLY ON CHARM MASS

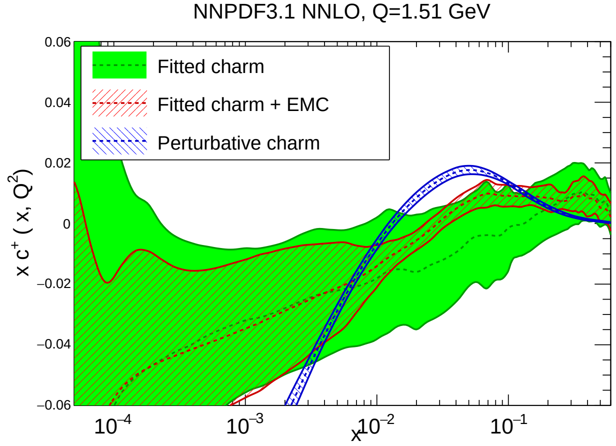
PERTURBATIVE CHARM VS m_c



FITTED CHARM VS m_c



FITTED VS. PERTURBATIVE CHARM



- ITS SHAPE SHOULD NOT BE DETERMINED BY FIRST-ORDER MATCHING (NO HIGHER NONTRIVIAL ORDERS KNOWN)
- MIGHT EVEN HAVE A NONPERTURBATIVE COMPONENT

FITTED VS. PERTURBATIVE:
SUPPRESSED AT MEDIUM-SMALL x,
ENHANCED AT VERY SMALL, VERY LARGE x

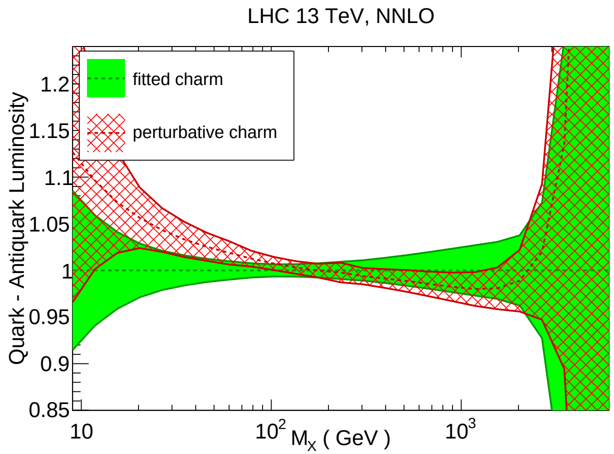
DATA VS. THEORY/METHODOLOGY

THE CHARM MASS AND TREATMENT

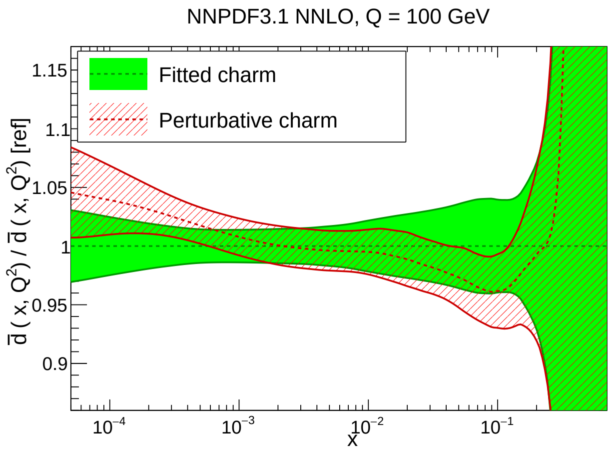
CHARM FROM DATA IMPACT ON LIGHT QUARK PDFS

FITTED VS. PERTURBATIVE CHARM

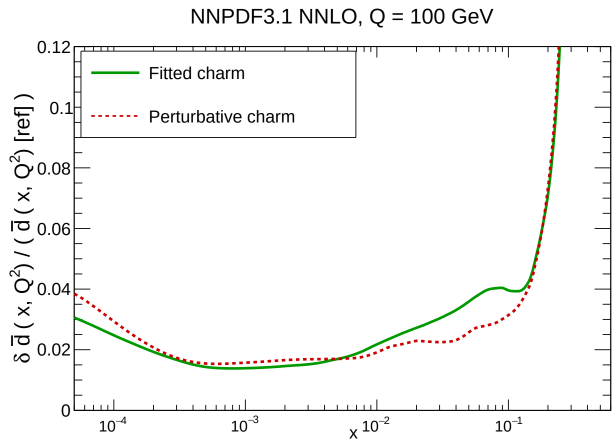
QQBAR LUMI



ANTIDOWN PDF



ANTIDOWN PDF UNCERTAINTY



- QUARK LUMI AFFECTED BECAUSE OF CHARM SUPPRESSION AT MEDIUM-x
- FLAVOR DECOMPOSITION ALTERED
- UNCERTAINTIES ON LIGHT QUARKS NOT SIGNIFICANTLY INCREASED
- AGREEMENT OF 13TeV W,Z PREDICTED CROSS-SECTIONS IMPROVES!

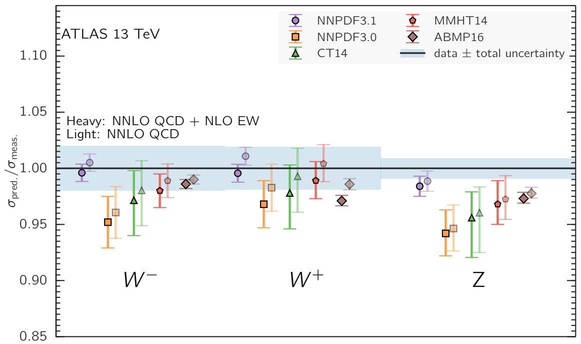
DATA VS. THEORY/METHODOLOGY

THE CHARM MASS AND TREATMENT

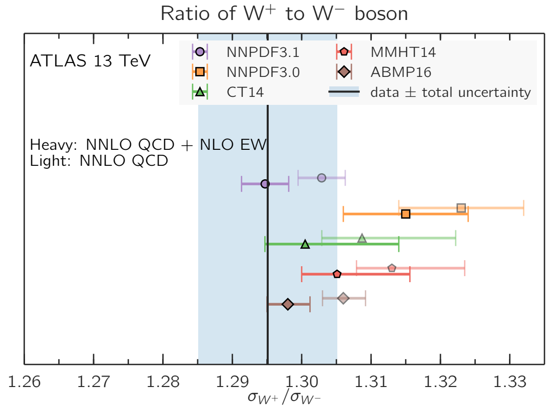
CHARM FROM DATA

IMPACT ON PHENOMENOLOGY

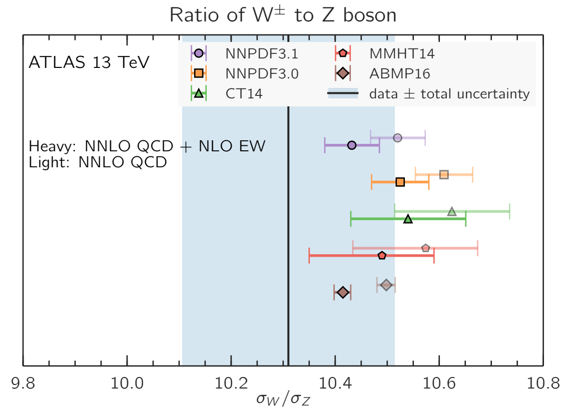
DRELL-YAN XSECTS



W^+/W^- XSECT RATIO



W/Z XSECT RATIO



- W, Z CROSS-SECTIONS AT 13 TEV IN PERFECT AGREEMENT WITH DATA THANKS TO FITTED CHARM!

LESSONS:

- TENSIONS CAN REVEAL METHODOLOGICAL ISSUES
- MORE LIKELY AS DATASET INCREASES, EXPERIMENTAL UNCERTAINTIES DECREASE
- RESOLVED BY MORE COMPLEX METHODOLOGY

OPTIMIZING THE METHODOLOGY

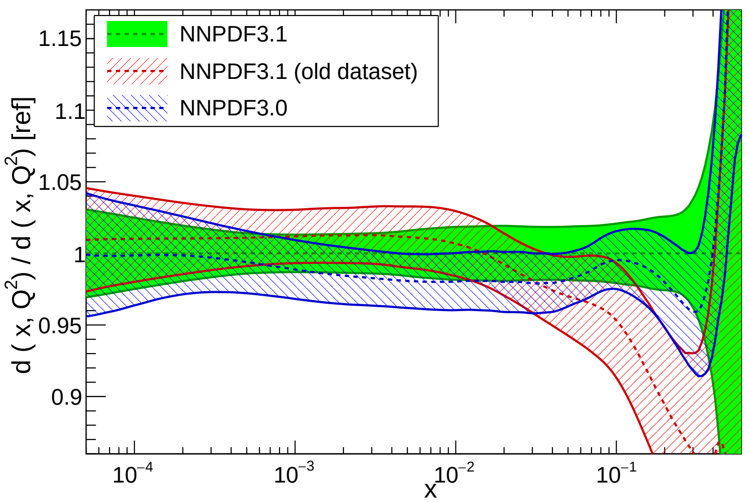
DATA VS. METHODOLOGY

- NEW DATA ⇒ MAJOR METHODOLOGICAL CHOICES ⇒ SIGNIFICANT IMPACT
- NNPDF3.1 VS NNPDF3.0: DATA AND METHODOLOGY HAVE SIMILAR IMPACT

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

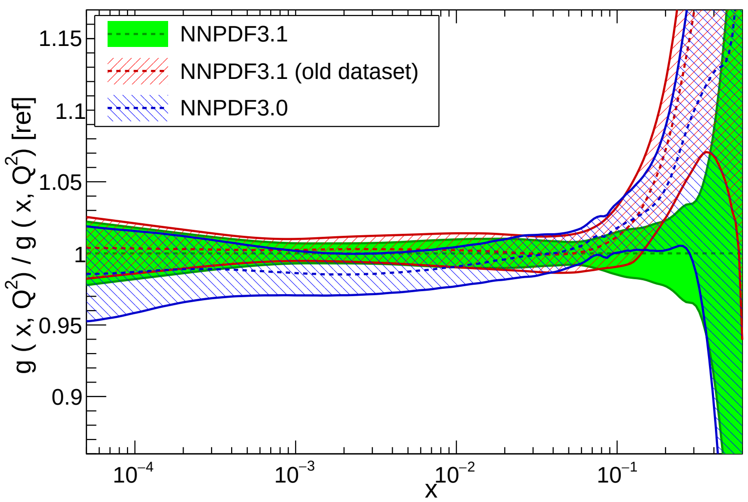
DOWN

NNLO, Q = 100 GeV



GLUON

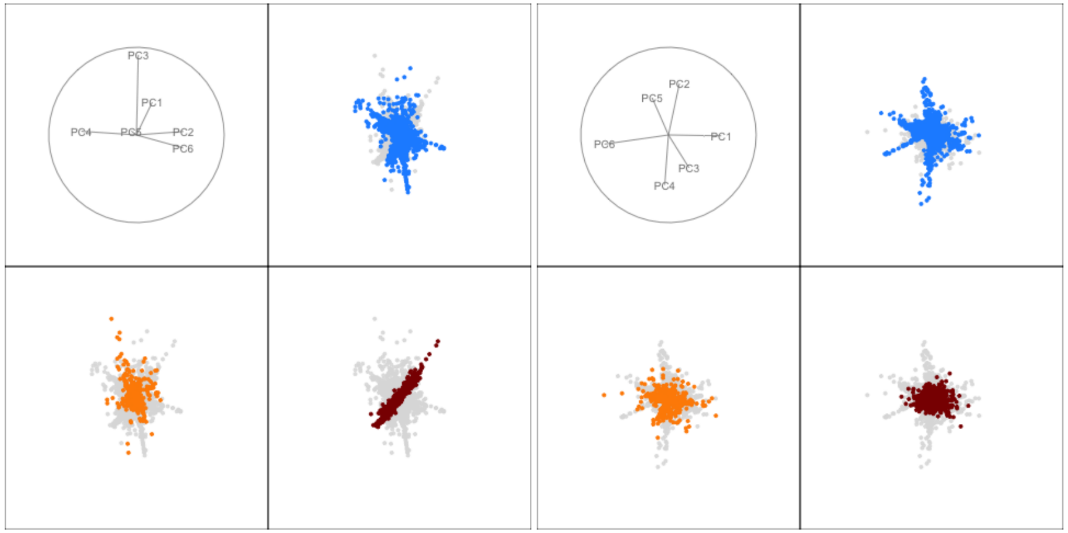
NNLO, Q = 100 GeV



DATA ⇒ THEORY ISSUES
VISUALIZATION TOOLS
TRACING THE IMPACT OF DATA

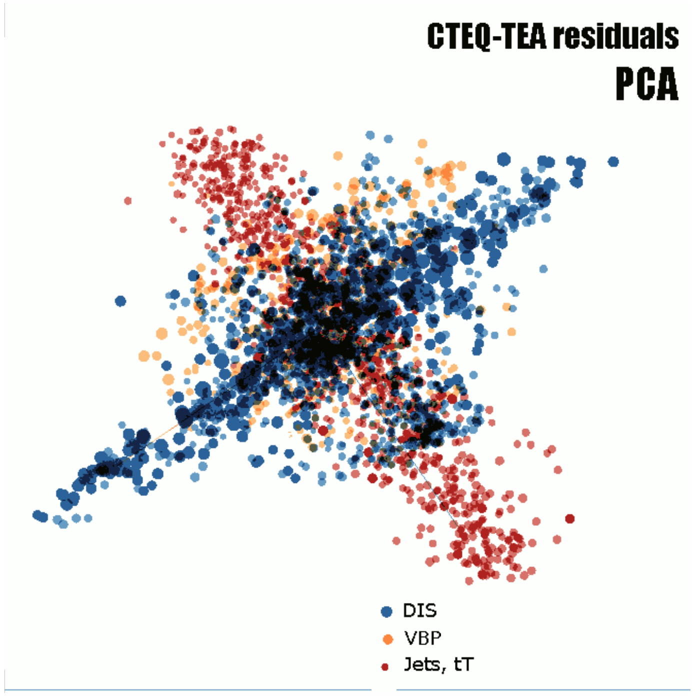
- DETERMINE THE SHIFT IN DATAPOINTS INDUCED BY CHANGES IN PDF
- CT14: 28 HESSIAN PARAMETER VARIATIONS
- VISUALIZE SHIFT VECTORS FOR 4000 DATAPOINTS
  - BY PCA
  - BY TOUR

TOUR OF CT14



(Cook, Laa, Valencia, 2018)

PCA PROJECTION OF CT14:
PDFSENSE

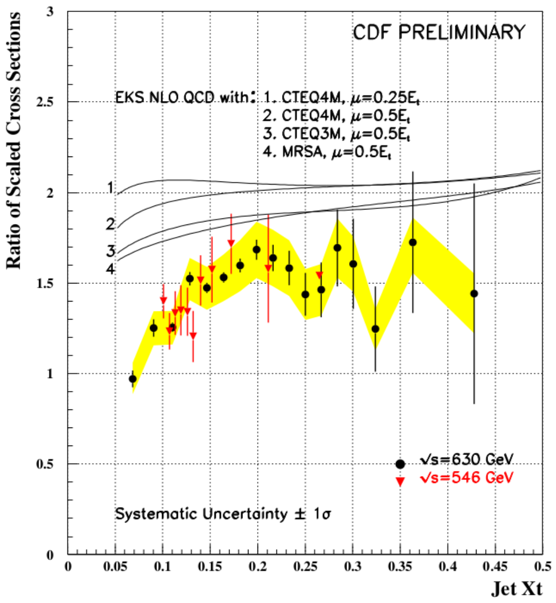


(Wang et al., 2018)

LESSONS FROM THE PAST: TWENTY YEARS BACK....

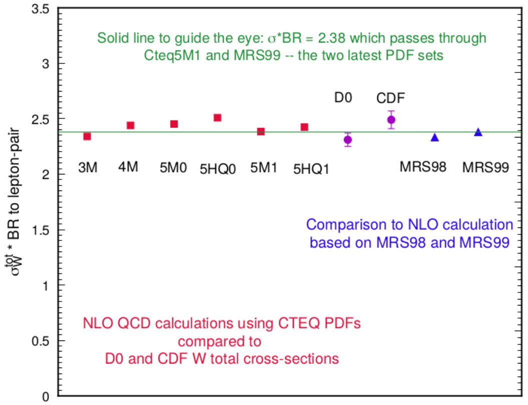
- NO PDF UNCERTAINTIES
- UNCERTAINTY ⇔ PDF SET VARIATION
- METHDOLOGY ↔ AGREEMENT WITH DATA

CDF JETS



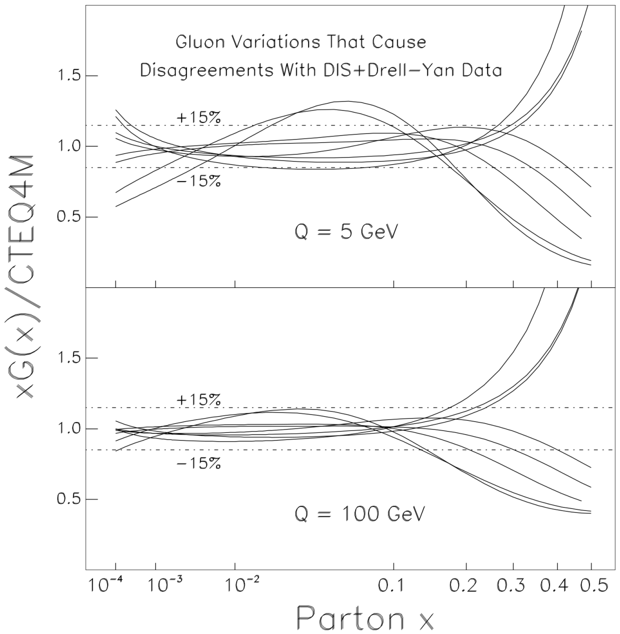
(Mangano, 1999)

W PRODUCTION



(Huston, Tung et al, 2000)

GLUON UNCERTAINTY



(Huston et al., 1998)

IS IT POSSIBLE TO DO BETTER?

FITTING THE METHODOLOGY
HYPEROPTIMIZATION SCANS
NNPDF3.1 DIS ONLY SCAN (preliminary)

(Carrazza, Cruz-Martinez, preliminary)

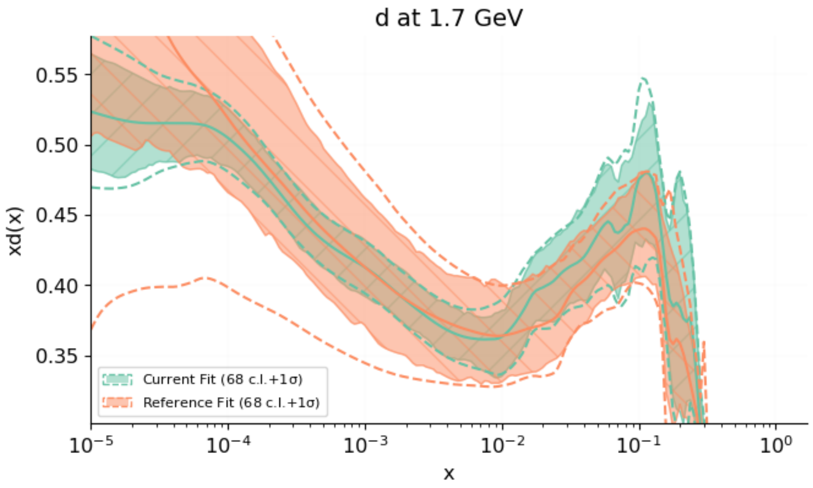
- EXPLORE METHODOLOGICAL ALTERNATIVES (ARCHITECTURE, MINIMIZATION, POSITIVITY,. . .)
- SCAN AUTOMATICALLY & ZOOM IN
- COULD BE DONE BY REINFORCEMENT LEARNING?

FITTING THE METHODOLOGY
WHAT IS "PROPER LEARNING"?

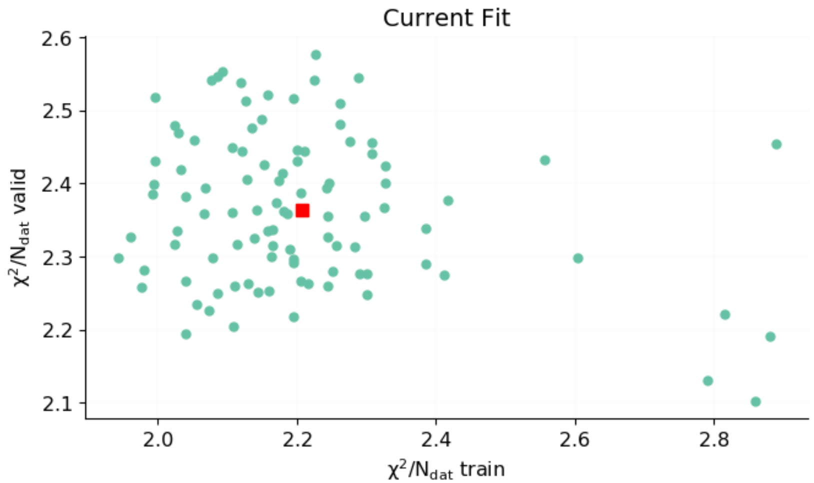
- CROSS-VALIDATION $\Rightarrow$ RANDOMLY DIVIDE DATA INTO TRAINING AND VALIDATION SETS
- MINIMIZE ON TRAINING; OPTIMIZE ON VALIDATION
- CORRELATIONS BETWEEN TRAINING AND VALIDATION
- HYPEROPTIMIZE ON VALIDATION $\Rightarrow$ OVERLEARNING

NNPDF3.1 DIS ONLY HYPEROPT ON VALIDATION
4 LAYERS (7 FOR GLOBAL); $\chi^2 = 1.01$

d QUARK



TRAINING VS VALIDATION χ^2



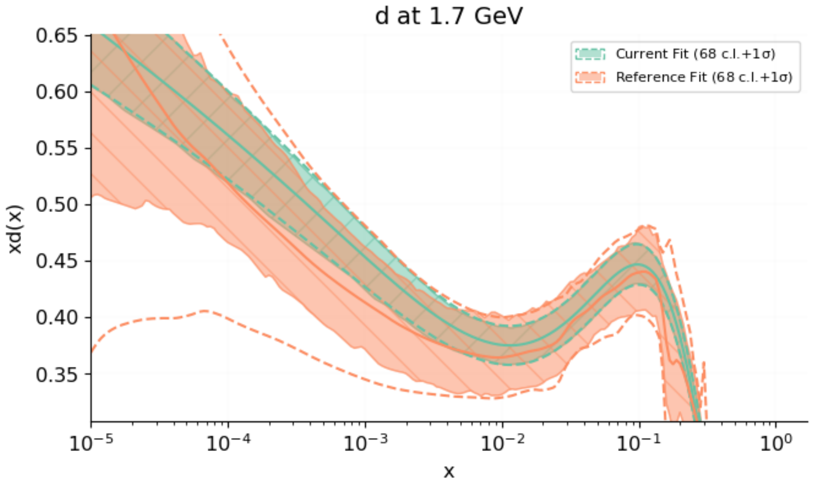
N3PDF
Machine Learning • PDFs • QCD

FITTING THE METHODOLOGY
WHAT IS "PROPER LEARNING"?

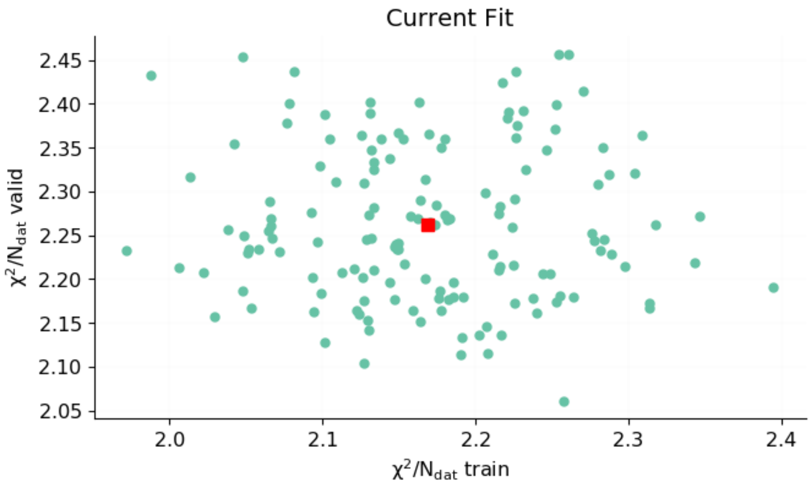
- NEED A COMPLETELY UNCORRELATED "TEST" SET
- OPTIMIZE ON WEIGHTED AVERAGE OF VALIDATION AND TEST
  $\Rightarrow$ NO OVERLEARNING
- BETTER SOLUTIONS? WHAT IS "OPTIMAL"?

NNPDF3.1 DIS ONLY HYPEROPT ON VALIDATION+TEST
2 LAYERS (SAME FOR GLOBAL); $\chi^2 = 1.14$

d QUARK



TRAINING VS VALIDATION χ^2



N3PDF
Machine Learning • PDFs • QCD

FITTING THE METHODOLOGY
WHAT IS “PROPER LEARNING”?

- BETTER SOLUTIONS? WHAT IS “OPTIMAL”?
- ARE STATISTICAL PROPERTIES OF THE ENSEMBLE RELEVANT?
- SHOULD THEORY PREJUDICE PLAY A ROLE?

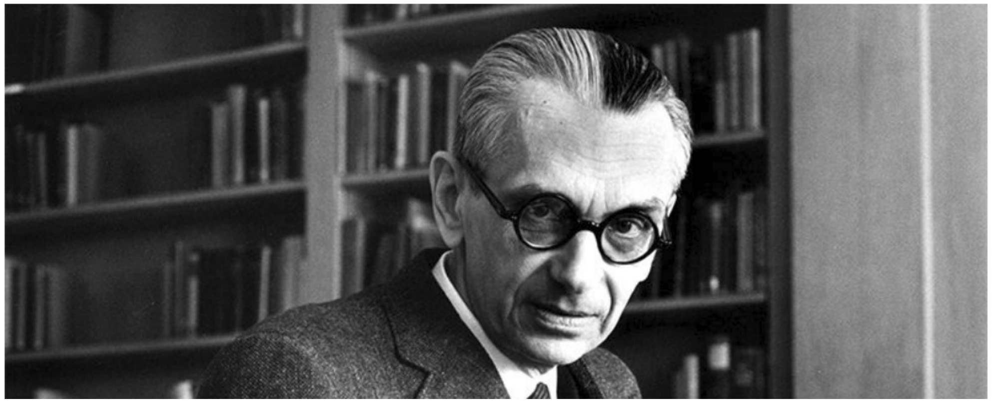
MENU nature
International journal of science

NEWS · 08 JANUARY 2019

Machine learning leads mathematicians to unsolvable problem

Simple artificial-intelligence problem puts researchers up against a logical parc
by famed mathematician Kurt Gödel.

Davide Castelvecchi



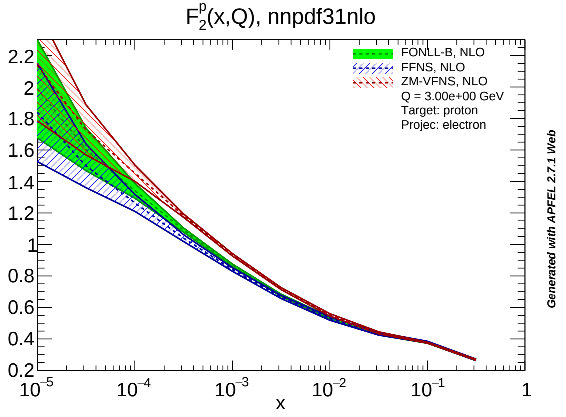
THEORY UNCERTAINTIES

THE MISSING HIGHER ORDER UNCERTAINTY

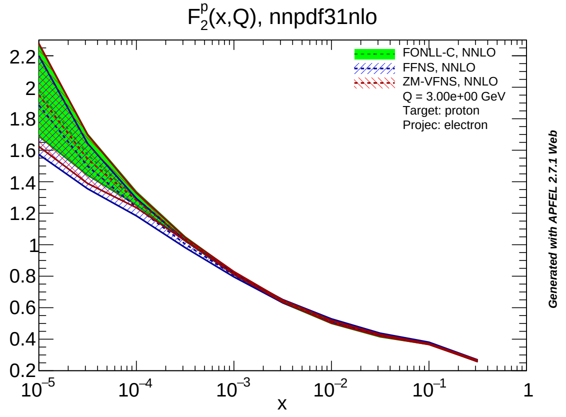
- DOMINANT THEORY UNCERTAINTY ON QCD PREDICTIONS $\Rightarrow$ MHOU (SCALE)
- NOT INCLUDED IN PDF UNCERTAINTY
- WHAT IS IT?

F_2^p PROTON DIS STRUCTURE FUNCTION

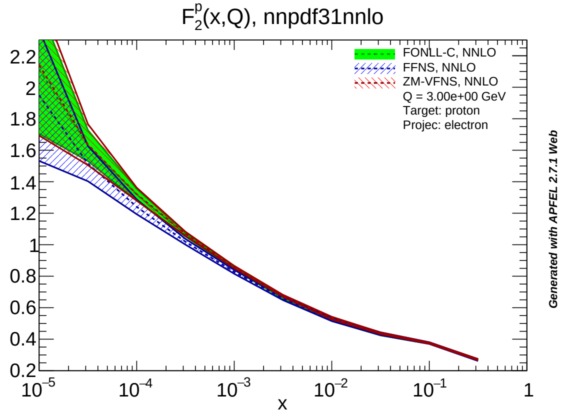
NLO THEORY NLO PDFS



NNLO THEORY NLO PDFS



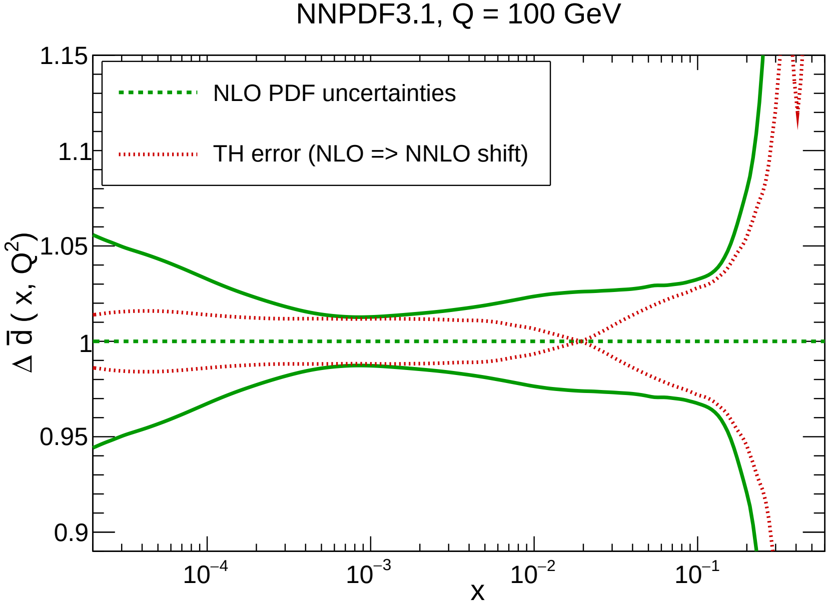
NNLO THEORY NNLO PDFS



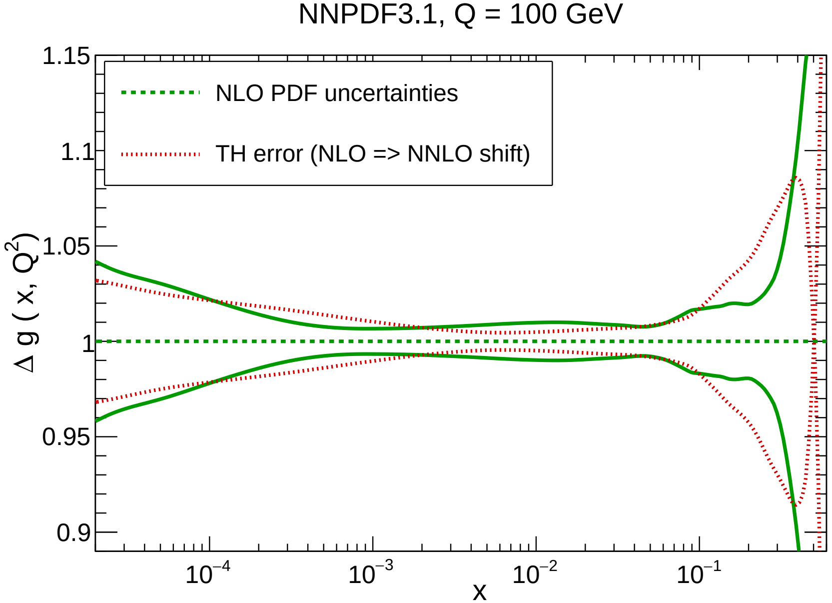
THE MISSING HIGHER ORDER UNCERTAINTY ON PDFs
HOW BIG IS IT?

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

ANTIDOWN



GLUON



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

THE MISSING HIGHER ORDER UNCERTAINTY ON PDFs
CAN WE ESTIMATE IT?

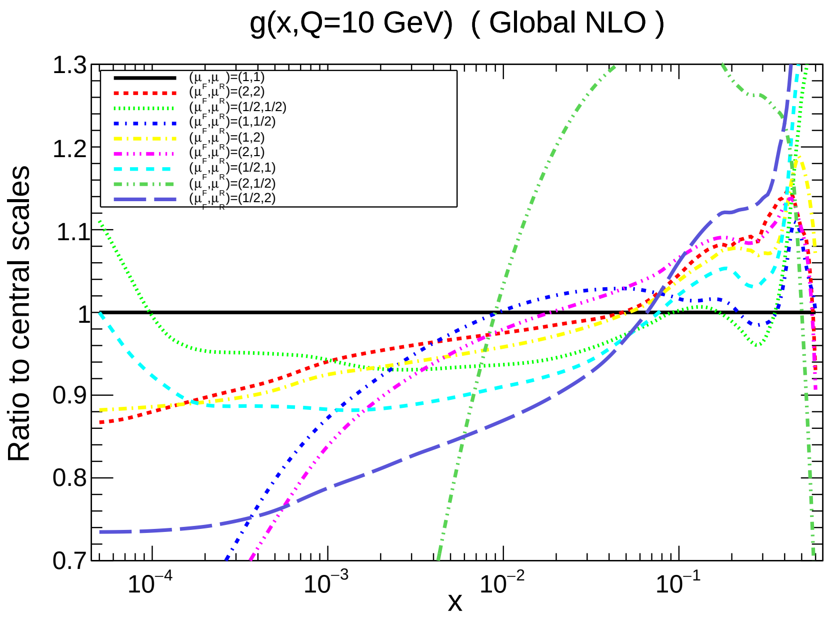
SCALE VARIATION IN PDF FITTING

NAIVE IDEA FOR PDF MHOU ESTIMATE

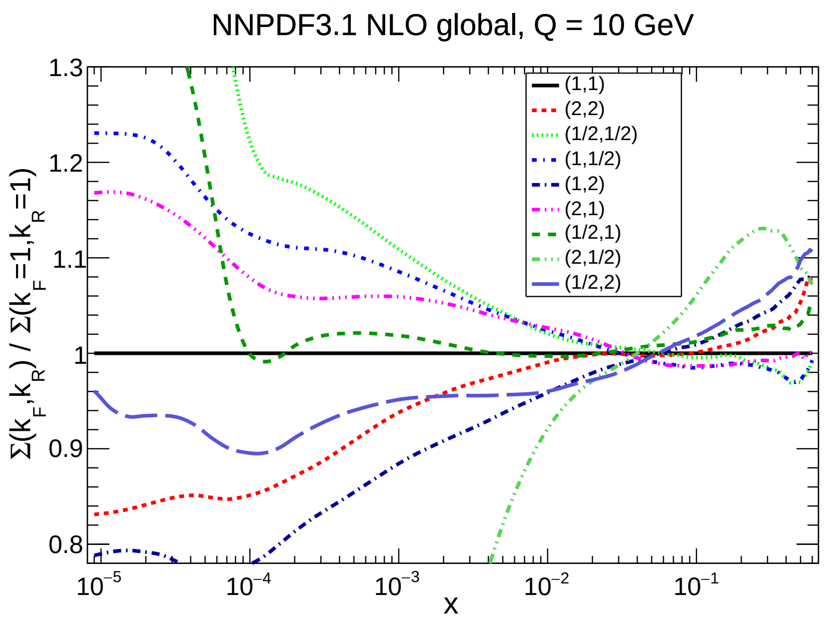
- PERFORM FIT WITH VARIOUS SCALE CHOICES
- TAKE ENVELOPE OF RESULTS
- 7-POINT $\Rightarrow$ OK!; 9-POINT $\Rightarrow$ UNSTABLE!
- RESULTS DEPEND STRONGLY ON THE CHOICE OF ENVELOPE

FITS WITH SCALE VARIATION

GLUON



QUARK SINGLET



THE THEORY COVARIANCE MATRIX

(NNPDF, 2019)

- ASSOCIATE MHOU TO NUISANCE PARAMETER $\Rightarrow$ THEORY COVARIANCE MATRIX S_{ij}
- $S_{ij} = \frac{1}{N}\sum_k \left(T_i^{(k)} - T_i^{(0)}\right)\left(T_j^{(k)} - T_j^{(0)}\right)$
  $\left(T_i^{(k)} - T_i^{(0)}\right)$: k-TH SHIFT OF i-TH DATAPOINT ABOUT CENTRAL PREDICTION $T_i^{(0)}$.
- SHIFT: GUESS FOR POSSIBLE MHO TERMS $\Rightarrow$ SCALE VARIATION

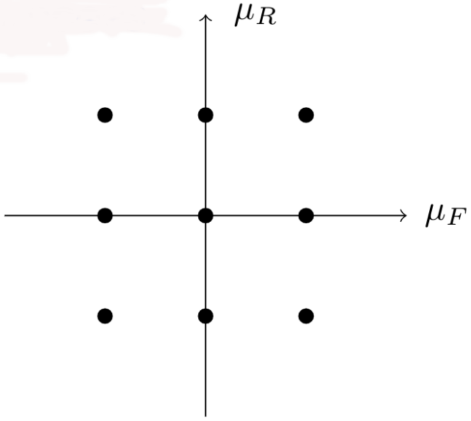
SCALE VARIATION

NINE-POINT SCALE VAR.

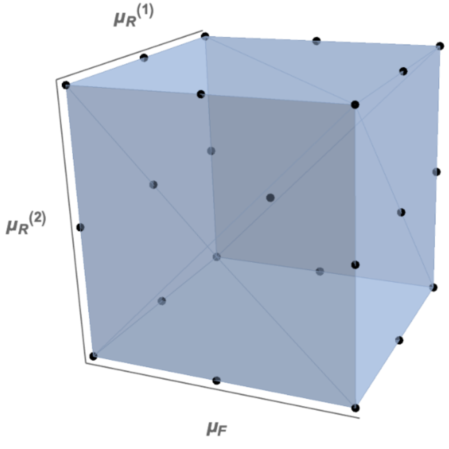
EXPERIMENTS AND PROCESSES

Process Type	Datasets
DIS NC	NMC, SLAC, BCDMS, HERA NC
DIS CC	NuTeV, CHORUS, HERA CC
DY	CDF, D0, ATLAS, CMS, LHCb (y, p_T, M_{ll})
JET	ATLAS, CMS inclusive jets
TOP	ATLAS, CMS total+differential cross-sections

SAME PROCESS



DIFFERENT PROCESSES



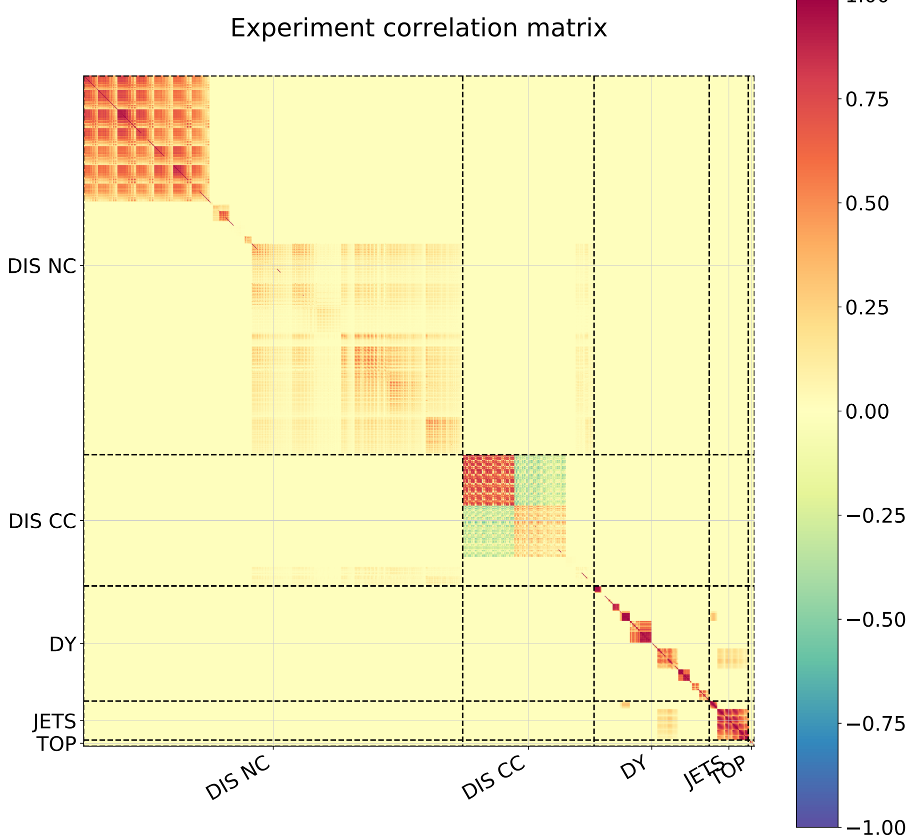
- CLASSIFIY DATA INTO PROCESSES
- PICK A SET OF SCALE VARIATIONS
- DECIDE HOW TO CORRELATE SCALE VARIATION BETWEEN DIFFERENT PROCESSES
- RENORMALIZATION $\Rightarrow$ MATRIX ELEMENT; FACTORIZATION $\Rightarrow$ EVOLUTION

THE THEORY COVARIANCE MATRIX: CORRELATIONS

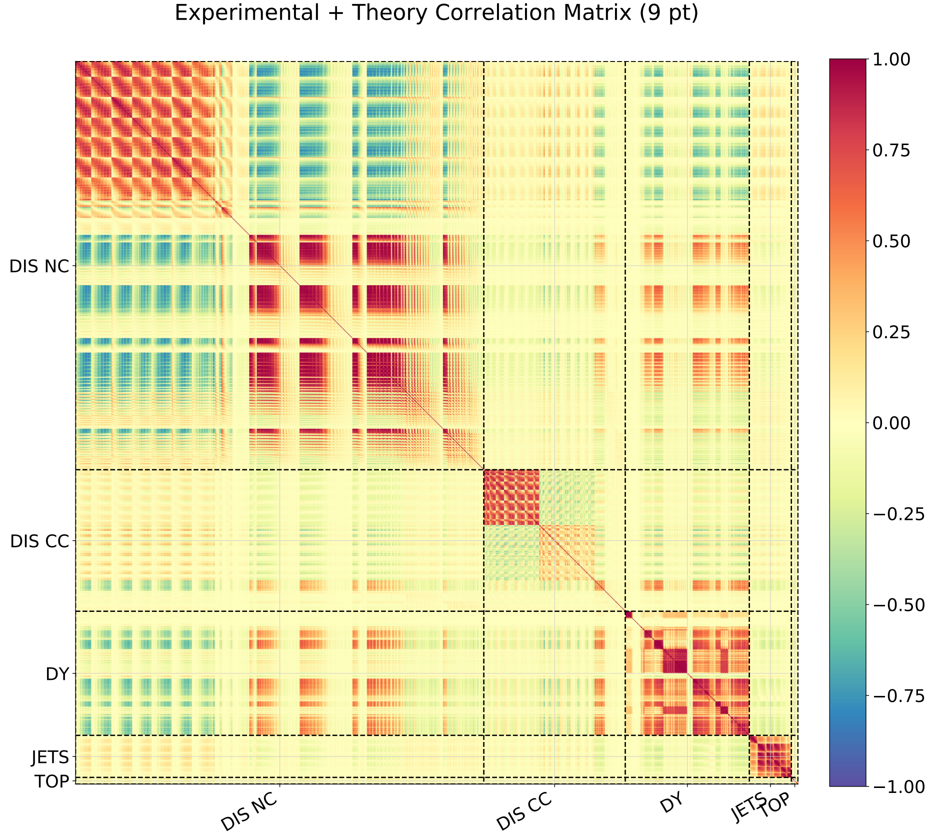
- INDEPENDENT NUISANCE PARAMETERS $\Rightarrow$ TH. AND EXP. ERRORS COMBINE IN QUADRATURE
  $\chi^2 = \sum_{i,j=1}^{N_{\rm dat}} \left(D_i - T_i^{(0)}\right) [S+C]_{ij}^{-1} \left(D_i - T_i^{(0)}\right)$
- REN. SCALE $\Rightarrow$ CORRELATIONS INDUCED BETWEEN EXPERIMENTALLY UNRELATED MEASUREMENTS OF SAME PROCESS
- FACT. SCALE $\Rightarrow$ CORRELATIONS INDUCED BETWEEN DIFFERENT PROCESSES

THE CORRELATION MATRIX

EXPERIMENT



EXPERIMENT+THEORY



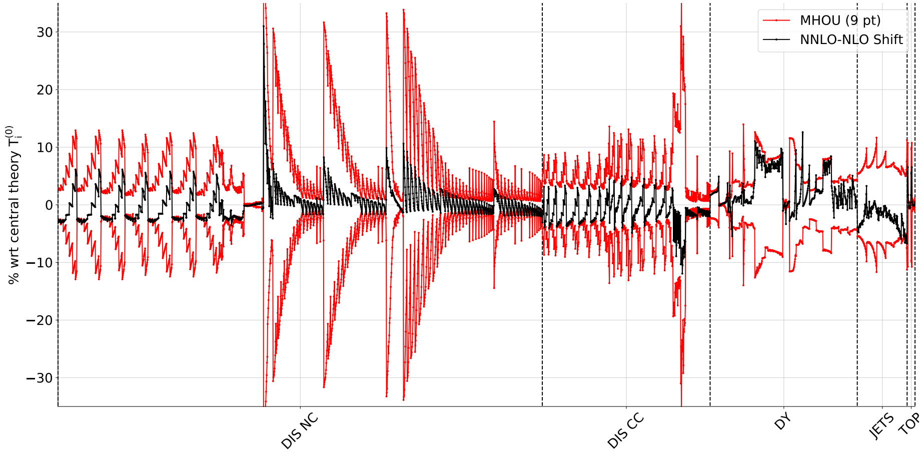
NNPDF

THE THEORY COVARIANCE MATRIX: VALIDATION

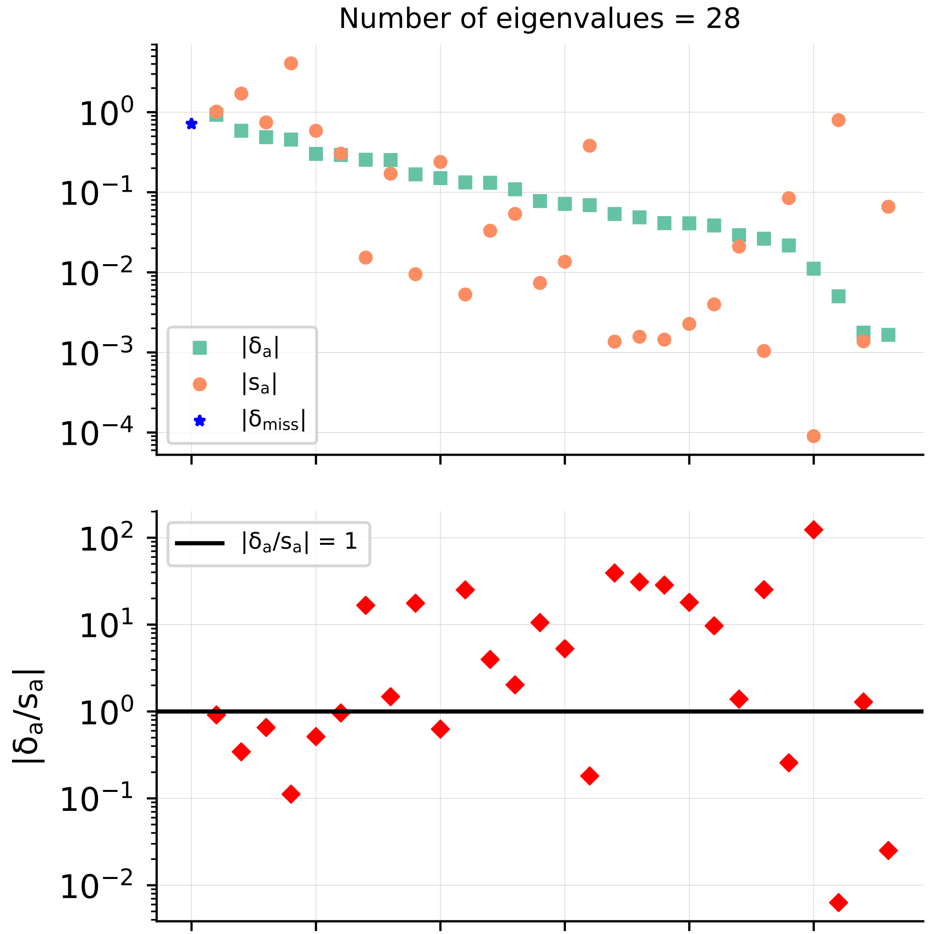
- COMPARE NLO THEORY COVMAT TO OBSERVED NLO-NNLO SHIFTS
- DETERMINE EIGENVECTORS e_i OF COVMAT $\Rightarrow$ 28 EVECS FOR 9PT, FIVE PROCESSES
- DETERMINE VECTOR OF SHIFTS δ
- DETERMINE PROJECTION OF δ IN SUBSPACE SPANNED BY e_i: IS IT CONTAINED IN IT?
- DETERMINE SIZE δ_i OF PROJECTIONS OF δ ALONG e_i: ARE THEY OF COMPARABLE SIZE?

9PT PRESCRIPTION

DIAGONAL UNCERTAINTIES VS. SHIFTS



EIGENVALUES VS. SHIFT PROJECTIONS



ANGLE BETWEEN SHIFT AND PROJECTED SHIFT

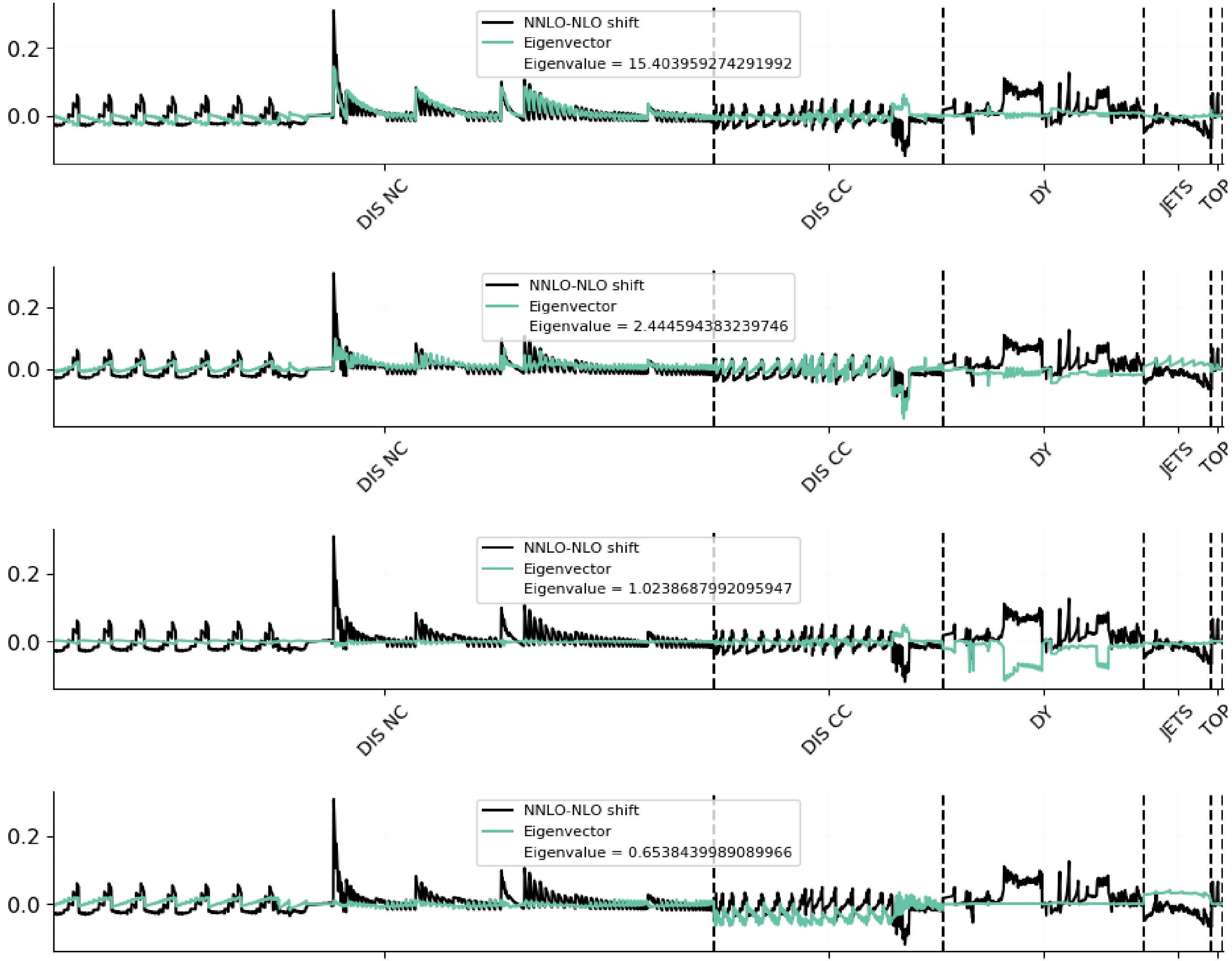
- DIS NC: 32°
- DIS CC: 16°
- DY: 21°
- JETS: 15°
- TOP: 3°
- GLOBAL: 27°

THE THEORY COVARIANCE MATRIX: VISUALIZING PROCESS IMPACT

- PROJECT THE SHIFT VECTOR δ ON EACH EIGENVECTOR
- LOOK AT THE INDIVIDUAL ~ 3000 COMPONENTS
- GROUP POINTS BY PROCESS

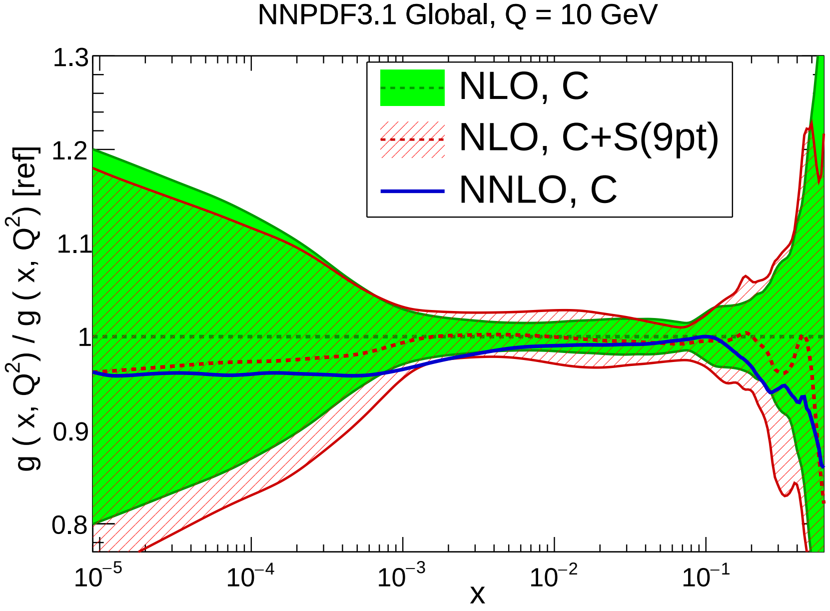
- RELATION BETWEEN SCALE VARIATION EIGENVECTORS & PROCESSES

projection of the shift vector along the four dominant eigenvectors

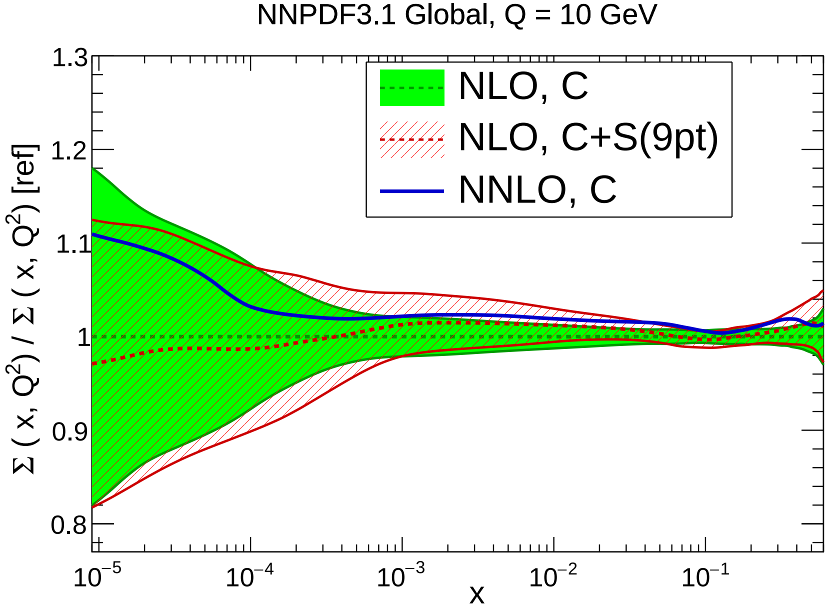


PDFs WITH THEORY UNCERTAINTIES 9PT

GLUON



SINGLET



	C	$C+S^{(3\mathrm{pt})}$	$C+S^{(9\mathrm{pt})}$
χ^2	1.139	1.139	1.109
ϕ	0.314	0.310	0.315

- FIT QUALITY χ^2 IMPROVES
- RELATIVE ERROR ϕ ON PREDICTION DOES NOT CHANGE
- PDF UNCERTAINTY ALMOST UNCHANGED
- CENTRAL VALUE MOVES TOWARDS KNOWN NNLO

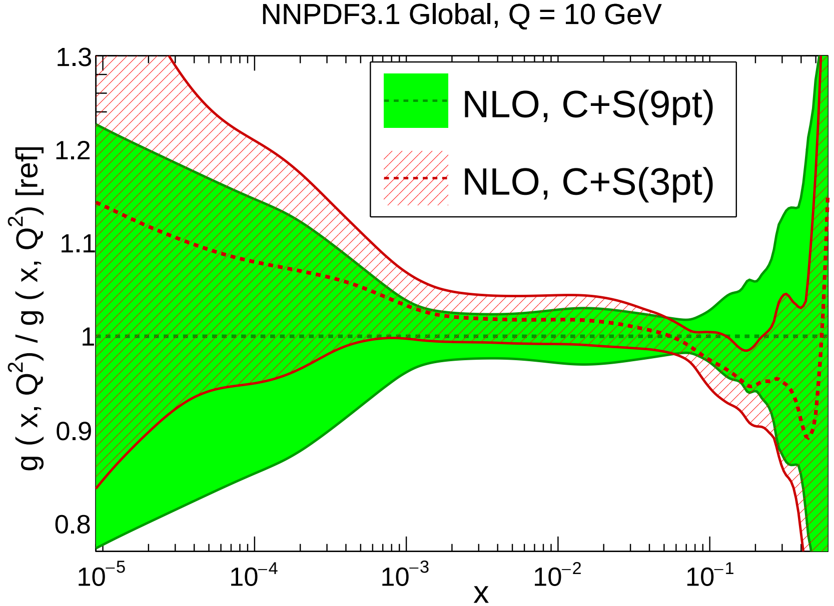
EQUALLY PRECISE BUT MORE ACCURATE RESULT!

PDFs with theory uncertainties

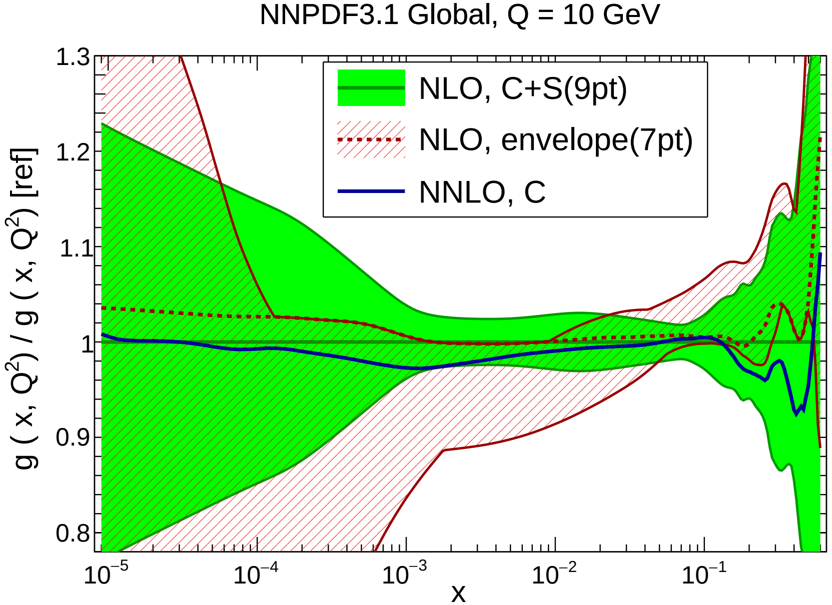
- results mildly dep. on prescription ⇒ 3pt closer to result w/o theory uncertainty
- "unstable" scale variations ⇒ no impact on fit
- 7pt envelope rather more conservative
  envelope does not include exp. uncertainty

the gluon PDF

3pt vs 9pt prescription



9pt prescription vs envelope



NNPDF

OUTLOOK

SUMMARY

- BETTER DATA $\Leftrightarrow$ BETTER THEORY AND BETTER METHODOLOGY
- CAN WE REDUCE ARBITRARINESS IN PDF METHPDOLOGY?
- CAN WE ASSESS ACCURATELY OUR THEORETICAL IGNORANCE?