CTEQ-TEA projects and xFitter connections

CTEQ

Pavel Nadolsky

CTEQ-TEA (Tung et al.) working group

Shanghai Jiao Tong University: J. Gao

University of Manchester/Kennesaw State: M. Guzzi

Michigan State University: J. Huston, J. Pumplin, D. Stump, C. Schmidt, J. Winter, C.-P. Yuan

Southern Methodist University: T.-J. Hou, P. Nadolsky, B. T. Wang, K. P. Xie

Xinjiang University: S. Dulat

Some physics questions we examine

1. **CT17p (preliminary):** how to extract useful information about PDFs from the LHC data? Dozens of new experiments, some disagreements between data, massive (N)NLO computations
2. **CT14 HERA2:** is DGLAP factorization consistent with DIS and LHC data? Yes, within uncertainties *(arXiv:*1609.07968*)*
3. **CT14 QED** (*arXIv:1509.02905)* and **CT14 IC PDFs** *(T.-J. Hou et al., arXiv:1704.xxxxx)***:** allow for nonperturbative photon and charm distributions
4. **CT14 MC1 and MC2:** Monte Carlo replicas with asymmetric uncertainties and positivity
   (T.-J. Hou et al., arXiv:1607.06066v2, published in JHEP)
   - Public program **mcgen** to generate such replicas on **metapdf.hepforge.org/mcgen**

Theoretical calculations for the PDF analysis

Reliable PDF fits…

- …*must* control (N)NLO theoretical uncertainties and demonstrate agreement between fitted experiments
- …*need* fast computations for (N)NLO cross sections *and for* detailed studies of various uncertainties
- …*need* NNLO cross sections computed for all main *PDF* sets *and* a common PDF set (PDF4LHC'15?)

[NNLO/NLO K-factors may depend on the PDF at a percent level]

Data sets to be included in CT17

- Use as much relevant LHC data as possible using applgrid/fastNLO interfaces to data sets, with tabulated NNLO/NLO K-factors
- Include theoretical uncertainties when relevant
- Study tensions between experiments
- Parallelize the global PDF fitting to allow for faster turn-around time
- I will highlight some issues on the examples of the experiments in bold
- Some slides from the PDF4LHC meeting, more details at DIS2017

- **Combined HERA1+2 DIS (SACOT-chi)**
- LHCb W/Z rapidity at 7 and 8 TeV (applgrid)
- LHCb Z y at 13 TeV (applgrid)
- ATLAS W/Z rapidity at 7 TeV (applgrid)
- ATLAS 8 TeV DY (applgrid)
- ATLAS 7 TeV W/Z pT (applgrid)
- **ATLAS 7,8 Z p_T**, as a function of mass **(applgrid)**
- CMS W,Z p_T,as a function of y, at 8 TeV (applgrid)
- CMS W rapidity at 8 TeV (applgrid)
- CMS W/Z p_T at 8 TeV (applgrid)
- CMS inclusive jet cross section at 7 TeV with R=0.7 (fastNLO)
- **ATLAS inclusive jet cross section at 7 TeV with R=0.6 (applgrid)**
- ATLAS and CMS 7,8 TeV tT diff. distributions
- ATLAS low/high mass Drell-Yan at 7 TeV
- CMS low mass/high mass DY at 8 TeV

CT14 PDFs with HERA1+2 (=HERA2) combination

Phys.Rev. D95 (2017) 034003

- Make the following changes with regards to CT14
  - replace HERA-1 (N_{pt}=579) by HERA-1+2 (N_{pt} =1120)
  - remove NMC $F_2^p(x,Q)$ (N=201)
  - add 1 more parameter to strange quark PDF ⇒ slightly lower $s(x,Q)$

Separate the four HERA2 DIS processes;
(Q_{cut} = 2 GeV)

	N_{pts}	$\chi^2_{red.}$ / N_{pts}
NC e^+p	880	1.11
CC e^+p	39	1.10
NC e^-p	159	1.45
CC e^-p	42	1.52
totals		
[reduced χ^2] /N	1120	1.17
χ^2 / N	1120	1.25
R^2 / N	1120	0.08

e^+p data are fitted fine

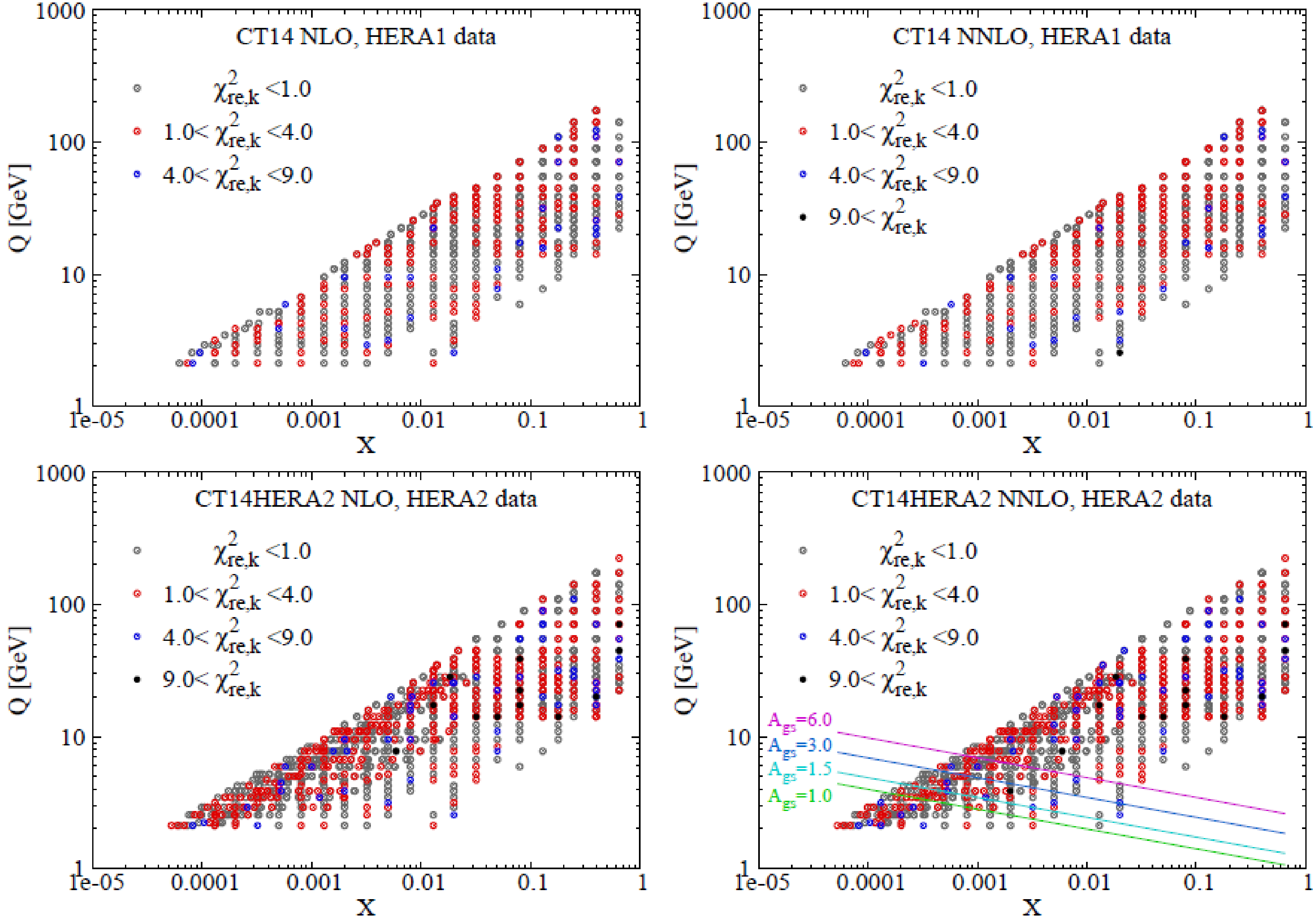
e^-p data are fitted poorly

reduced χ^2 values

χ^2 = [reduced χ^2] + R^2

The quadratic penalty for 162 systematic errors = 87.5

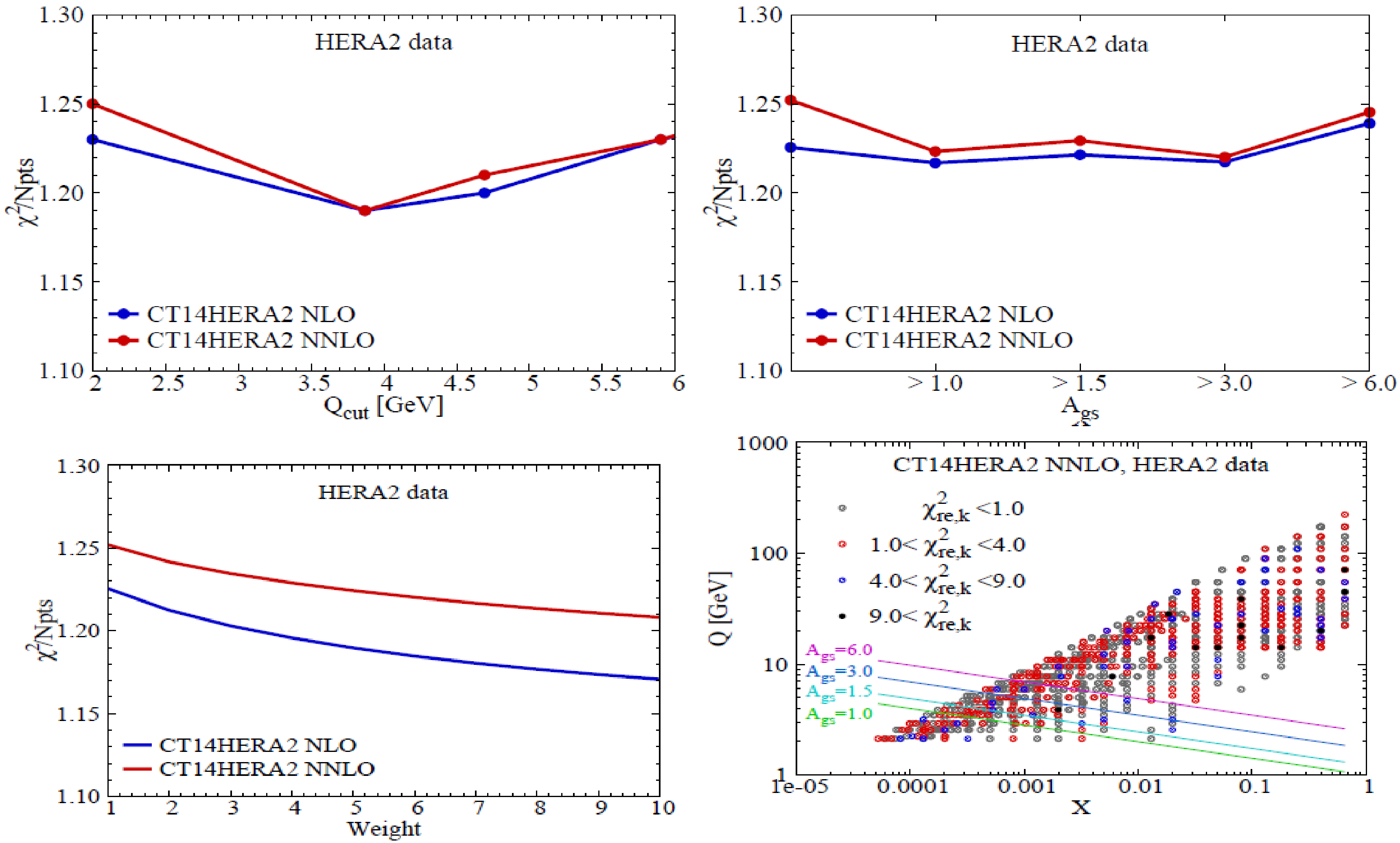
CT14 PDFs with HERA1+2 (=HERA2) data



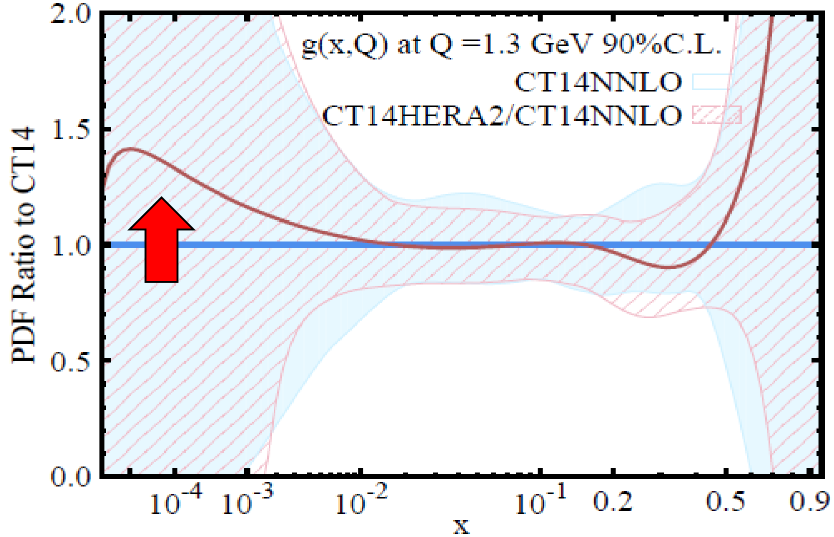
Points with excessive χ^2 are randomly scattered in the $\{x, Q\}$ plane

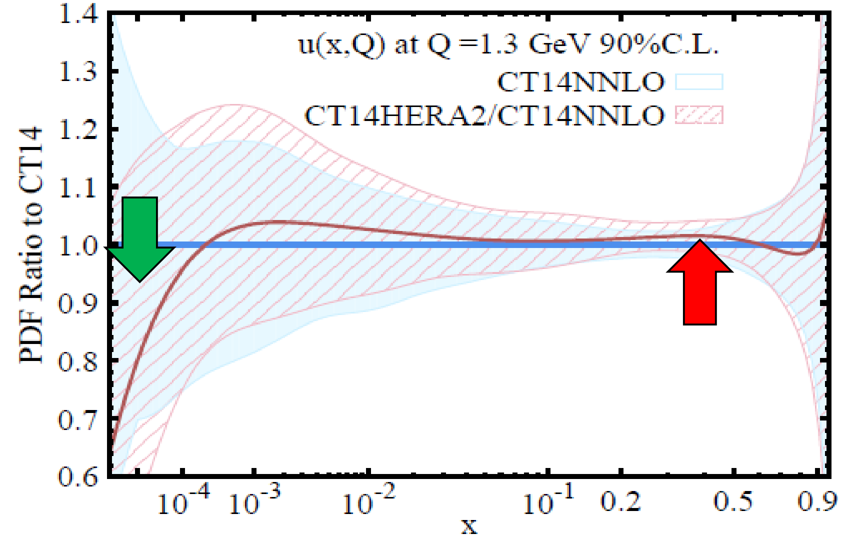
Stability of CT14 HERA2 fit

- The fit quality does not change much when we vary lower cuts on Q and $A_{gs} = x^{0.3}Q^2$, or increase statistical weight of the HERA2 data

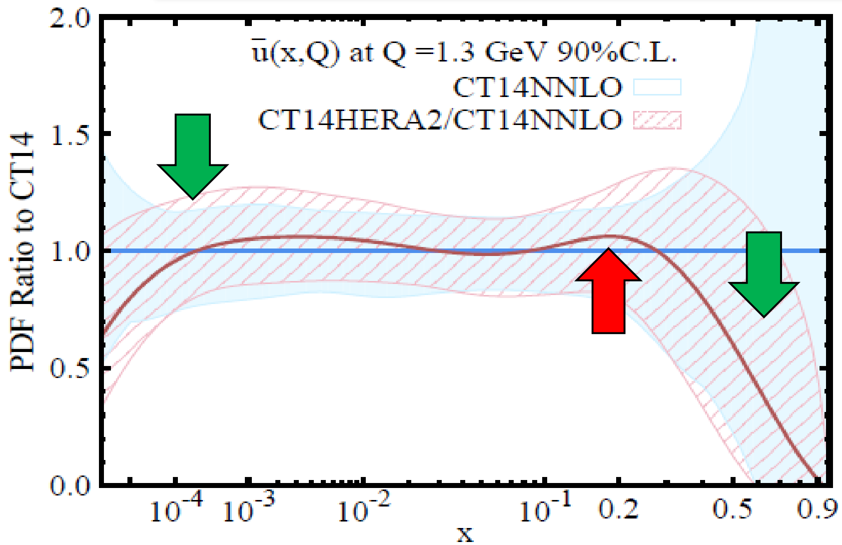


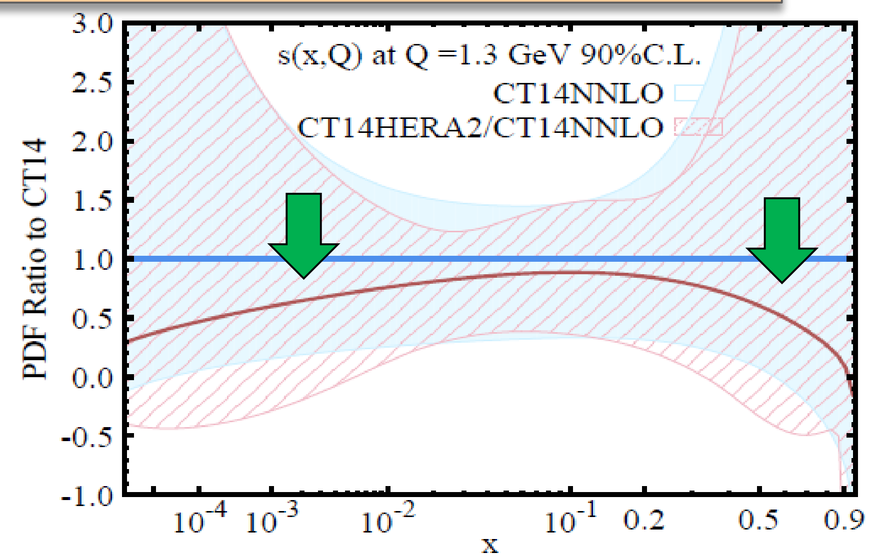
CT14HERA2 vs. CT14 PDFs: central sets



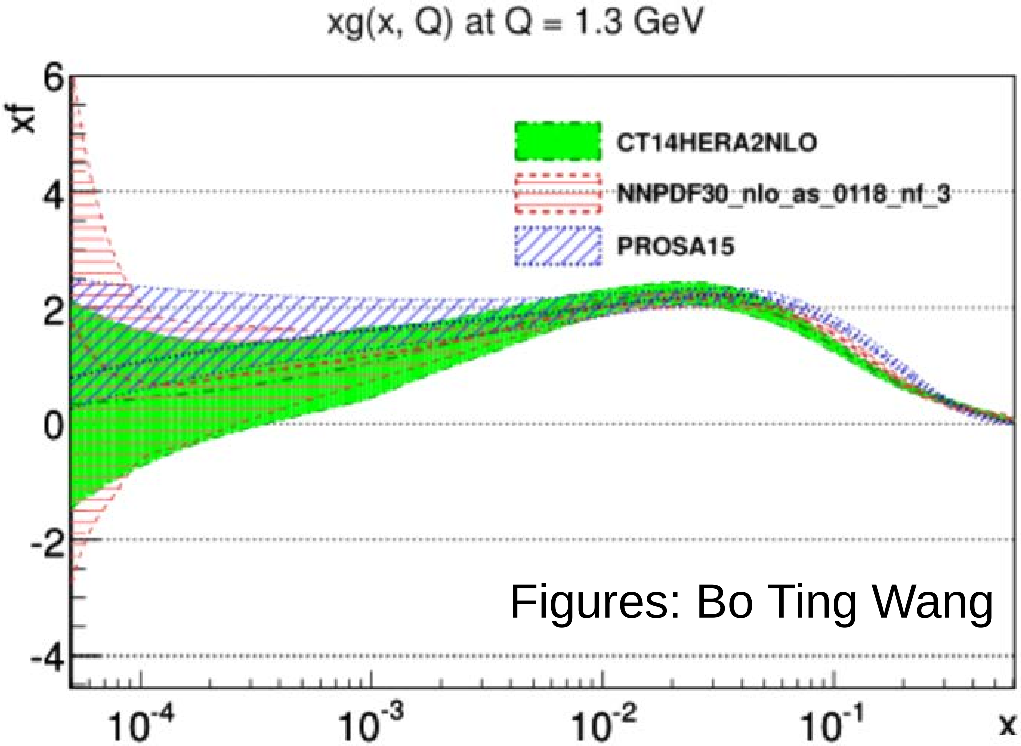


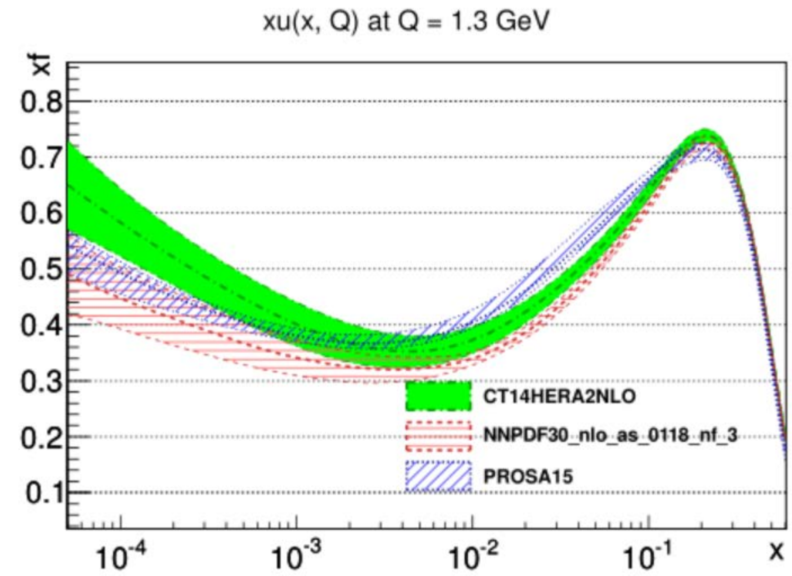
All changes are within CT14 uncertainties

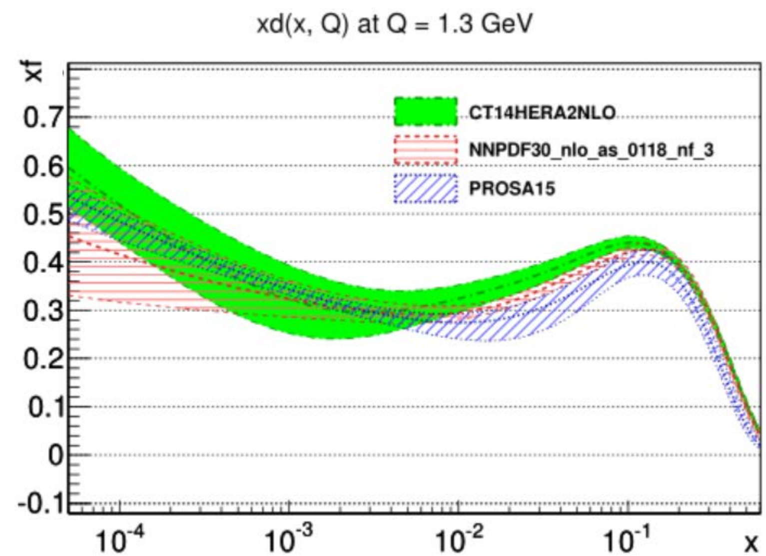




CT14HERA2 PDFs at small x



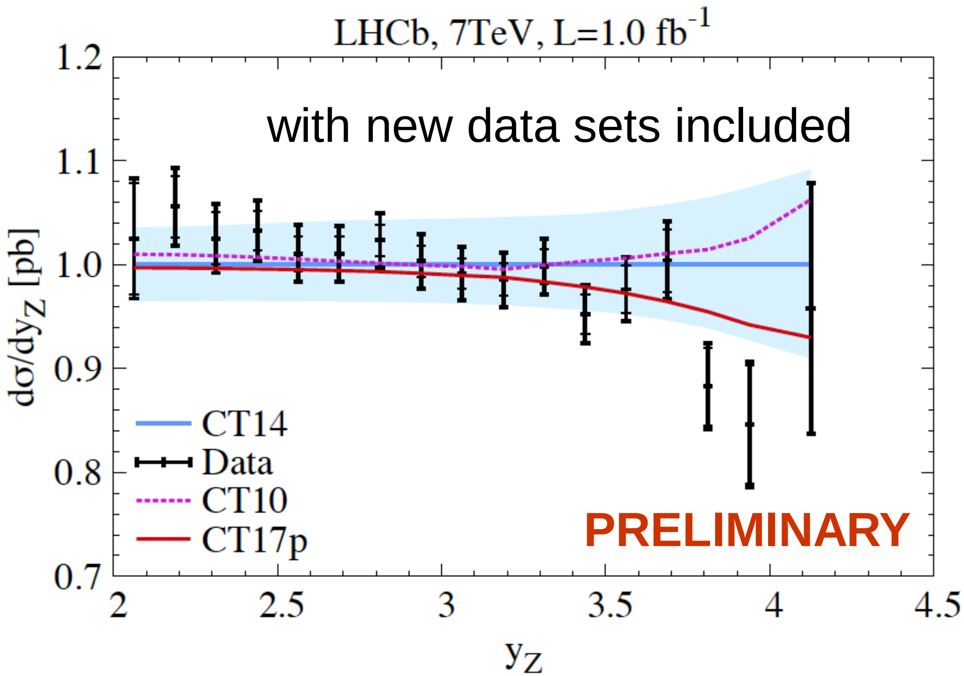


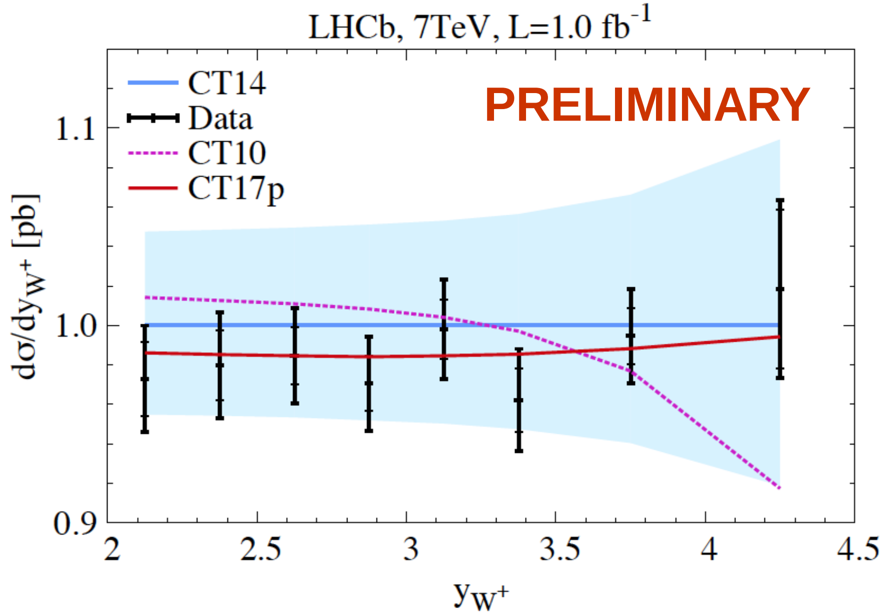


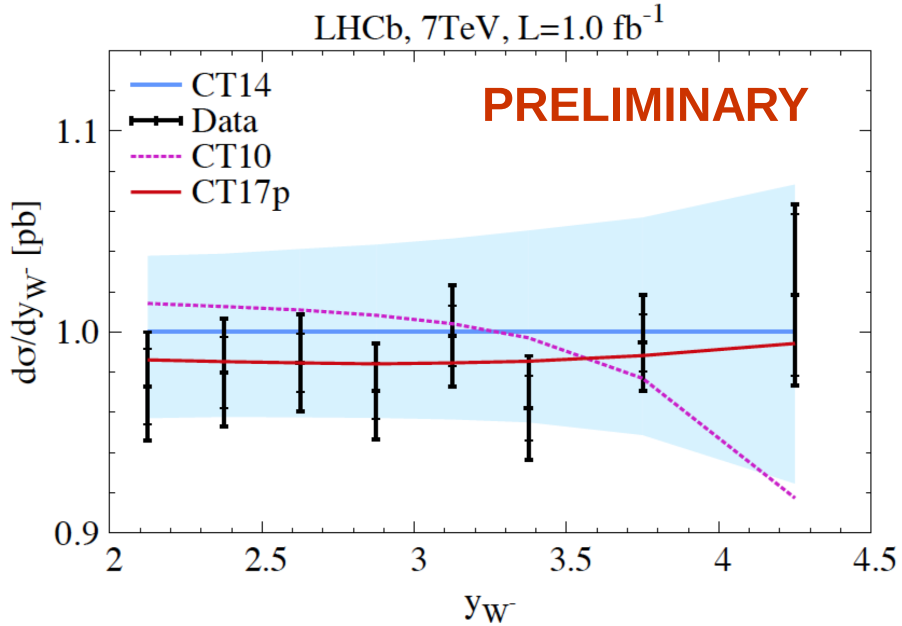
At $x < 10^{-2}$, the CT14HERA2 NLO $N_f = 3$ gluon is compatible with PROSA'15 PDFs fitted to 7 TeV LHCb heavy-flavor production data. Next rounds of LHCb measurements may help constrain the small-x gluon.

ATLAS, LHCb W,Z rapidity data

- Shown are LHCb 7 TeV: W,Z muon rapidity, LHCb 8 TeV Z(->ee) rapidity
- Unshifted data; outer error bars are with stat and syst errors added in quadrature; inner error bars are uncorrelated errors alone
- Major change is an increase in the strange quark distribution to that similar to CT10

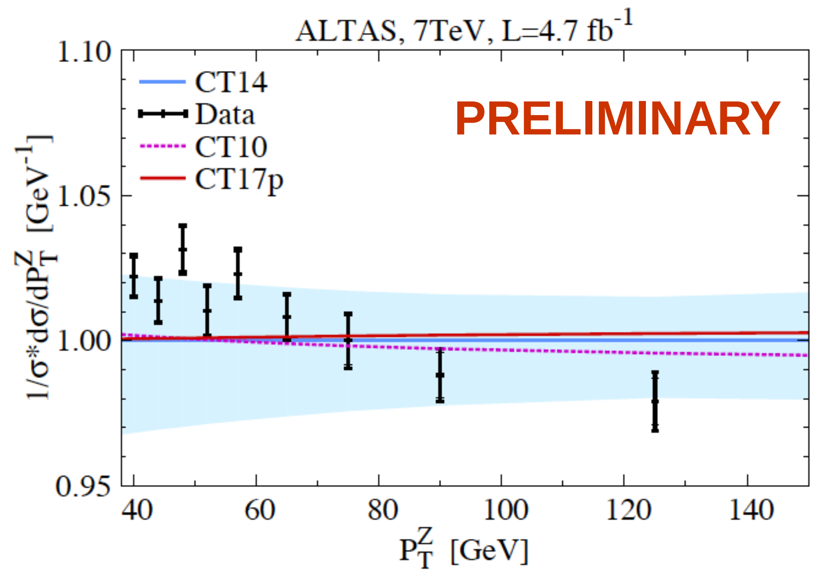


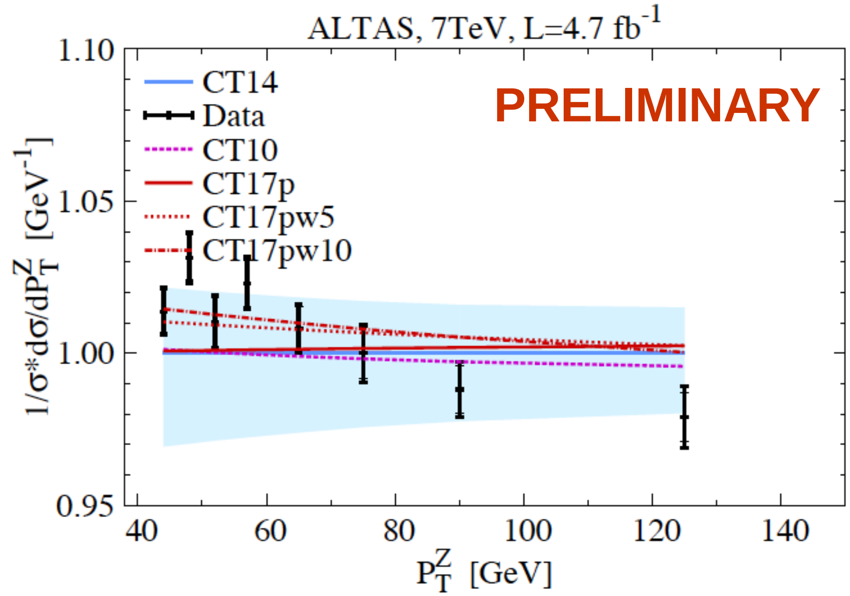




ATLAS Z p_T

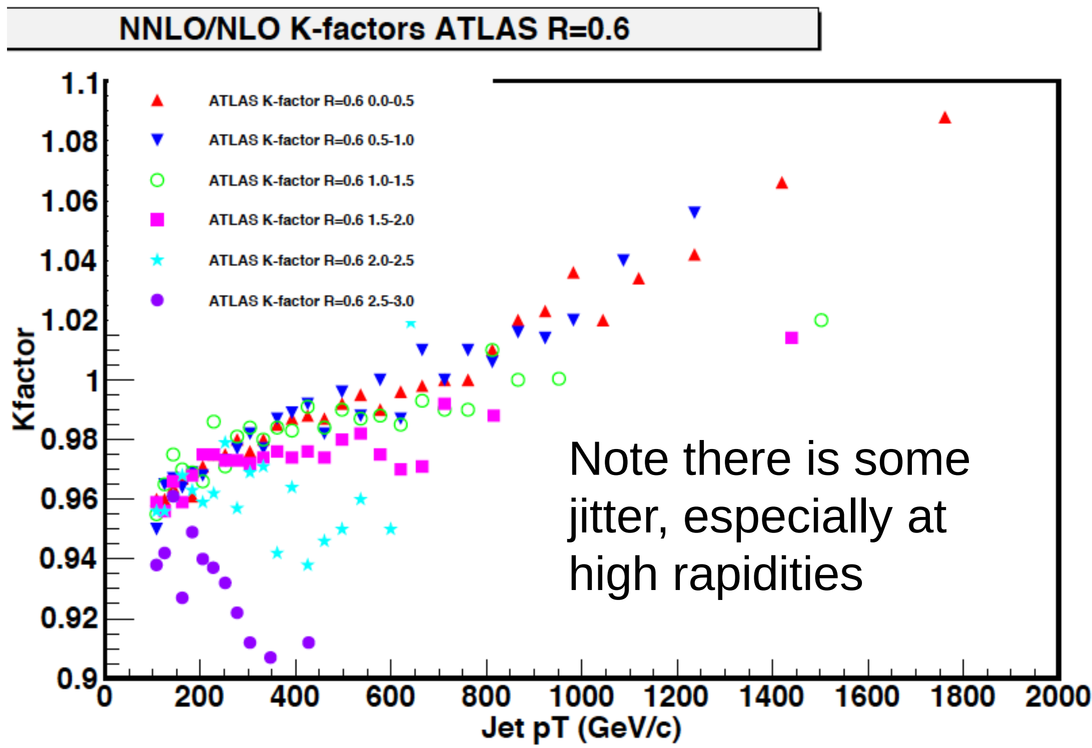
- Start with using normalized Z p_T data in range from 40 to 150 GeV/c at NLO, with $\mu_R=\mu_F=m_T$
  - prefers softer gluon at high x
  - adding additional weight =5 (w5) or 10 (w10) improves agreement
- Also trying absolute cross section
- Will add in NNLO/NLO K-factors
  - NNLO/NLO K-factors for this process have jitter, need an additional uncorrelated error





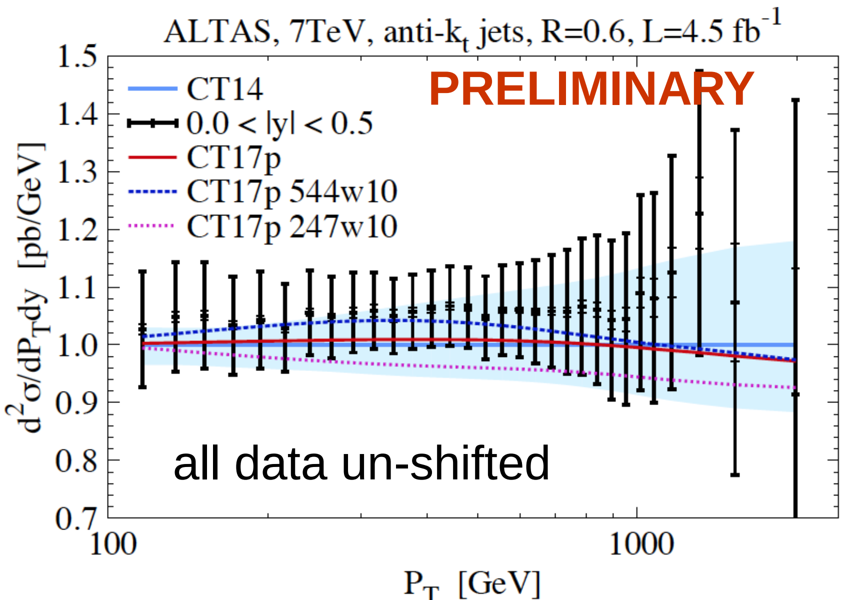
Some jet issues

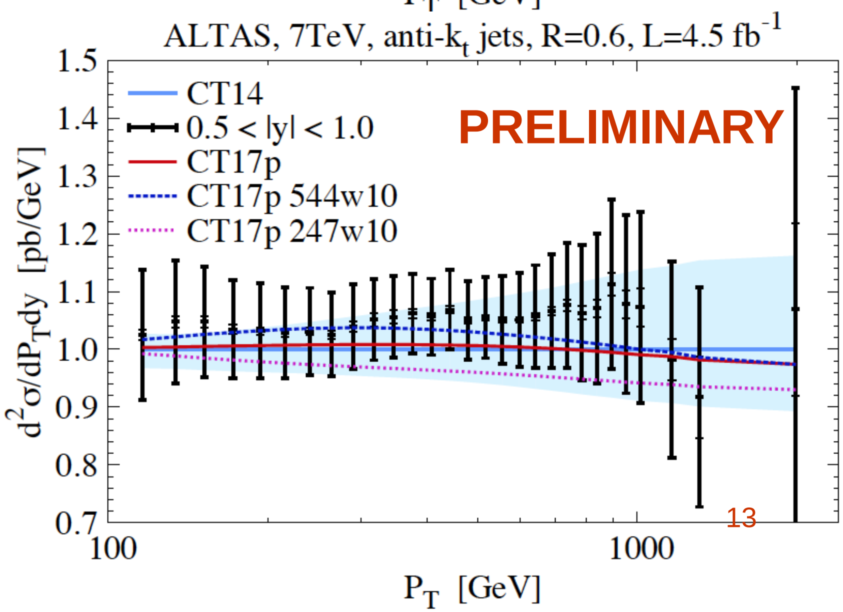
- For LHC jet cross sections, use largest jet sizes available (R=0.6 ATLAS, R=0.7 CMS) and a scale choice of p_T^{jet}
  - traditional choice for CT PDF fitting
- With such choices, K-factors (NNLO/NLO) nearly constant and close to unity
- Known tension within ATLAS jet rapidity data sets
- Not quite fair to choose just one rapidity interval (say 0.0-0.5) if another rapidity interval (say 0.5-1.0) gives a different PDF result
- Currently under investigation; what is the envelope of PDFs resulting from fits to individual rapidity bins, i.e. how unanimous are the rapidity bin data
- Decision whether to use ATLAS data set without resolution of rapidity correlations is pending



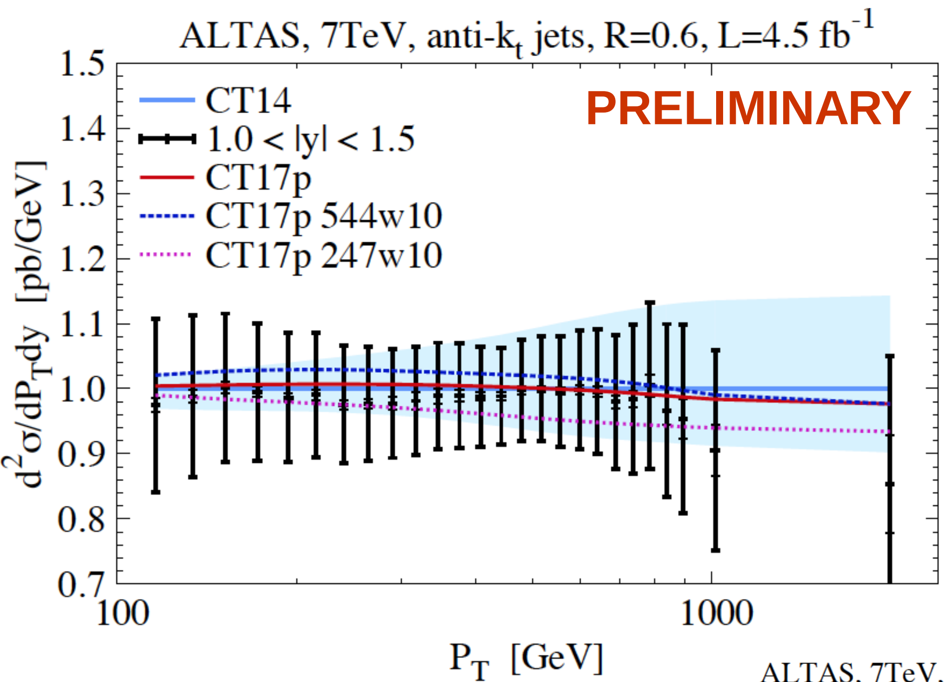
ATLAS 7 TeV jet data

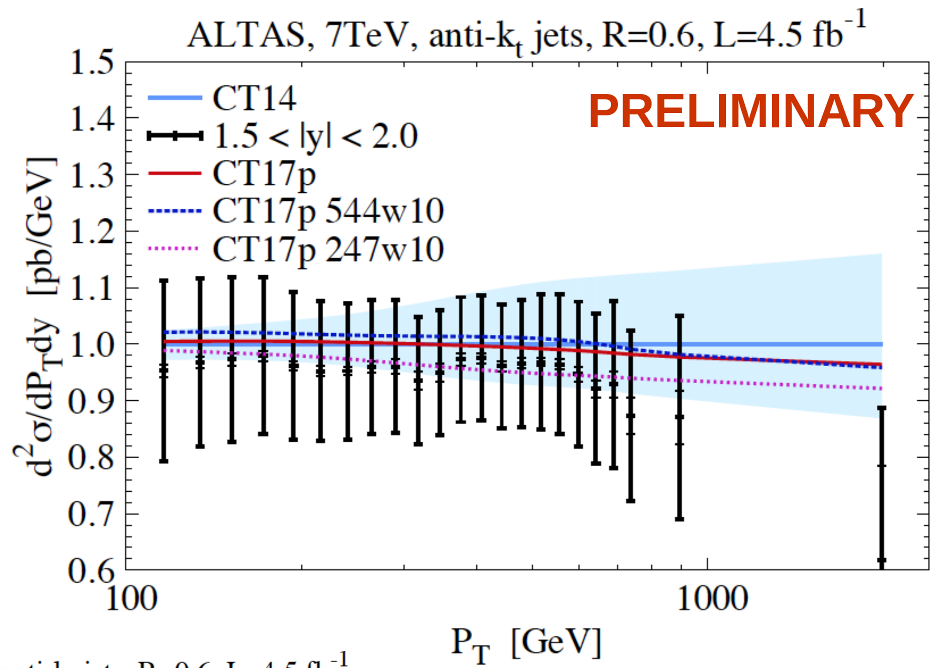
- Cannot get a good fit simultaneously to all rapidity bins: χ^2/DOF~2, even with weight 10 for combined jet data
  - outer error bars are with stat and syst errors added in quadrature; inner error bars are uncorrelated errors alone
- Individual rapidity bins can be fit well
- In general, ATLAS jet data favors a larger gluon PDF around x=0.2
- Some tension with ATLAS Z p_T data regarding high x gluon
  - jet data with weight=10 **(544w10)** likes a harder gluon (x~0.2)
  - Z p_T data with weight=10 (normalized, **247w10**) likes a softer gluon

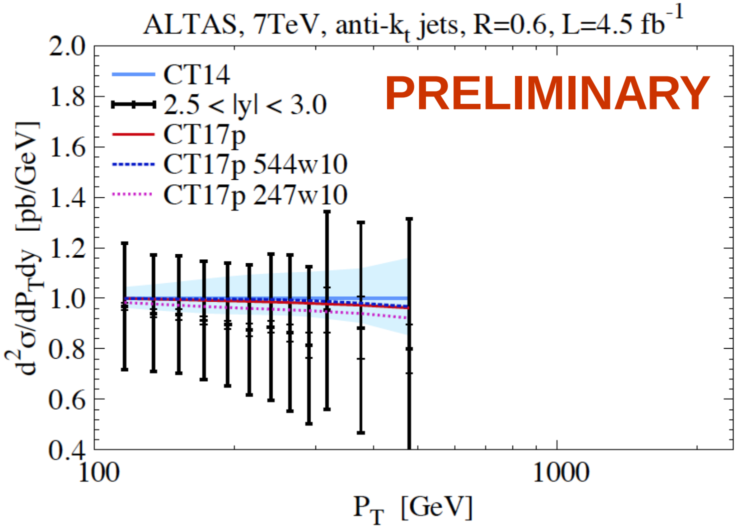




ATLAS 7 TeV jet data







Top distributions

- There are several distributions measured by ATLAS and CMS that have information on the high x gluon
  - m_{tT}, $y_{tT,}$ $p_T^{t,T}$ directly
  - $y_{t,T}$, p_T^{tT} indirectly
- Only one distribution should be used, unless a correlation model can be developed
  - which one?
  - do they give the same answer? if not, do we understand why?
  - how do the constraints/trends from each distribution compare?
  - similar to what we were talking about with ATLAS jet data in different y bins
- We are currently doing exploratory studies at NLO using MCFM and DiffTop and at approx. NNLO using DiffTop
  - starting with the p_T and rapidity of the top quark

- ATLAS and CMS have different trends; in this case, ATLAS favors harder gluon (than NNPDF3.0) at high x, CMS weaker gluon
- In general, the ATLAS and CMS top results are in tension internally, and with each other (the latter more so in the case of normalized distributions where the experimental errors are smaller)
- This is similar to the tension that exists between the ATLAS and CMS jet data, although there the tension is in the opposite direction
- If tension, then gluon PDF uncertainty may not decrease and may even increase
- Study in progress

Software for PDF analysis developed at SMU

1. **SACOT-chi at NNLO + IM scheme at NNNLO** in **xFitter**

(T.-J. Hou, M. Guzzi, Bowen Wang)

- Ready to be used in comparisons to other schemes, studies of $m_c(m_c)$, etc.

2. Local generation of **APPLgrids** using MCFM and aMCFast on US-based high-performance computing facilities

- Can coordinate APPLgrid generation with other groups

3. **Standalone public codes** – can be used/distributed with xFitter:

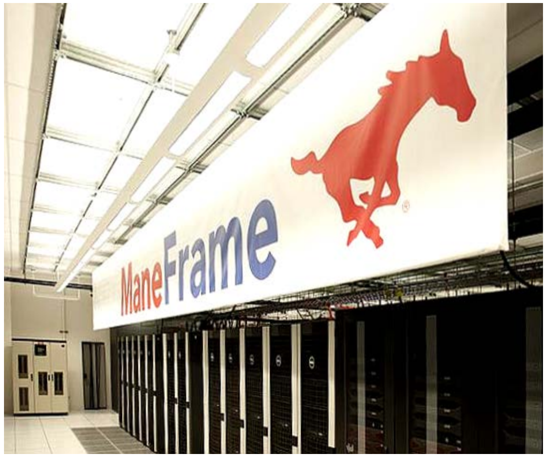
- **MCGEN** (metapdf.hepforge.org/mcgen): a C++ code to convert Hessian eigenvector PDF sets in the LHAPDF format into MC replica sets according to several methods

- **Mathematica packages**:
  a. Combination of PDF sets using meta-parametrizations
  b. N-dimensional matrix tools: reduction of correlation matrices, reconstruction of n-dimensional ellipsoids from discrete data,….

High-performance computing facilities

ManeFrame HPC center, SMU

- Currently 1100 nodes (8 cores/node, 2.8 GHz), **upgrade to** 316 nodes with with 11584 cpu cores and 132608 accelerator cores, 256/768/1536 GB RAM per node, 36 Xeon Phi processors (64 cores/385 GB RAM)



HPC center, Michigan State University

- 2016 nodes with two 2.4Ghz 14-core Intel Xeon E5-2680v4 Broadwell CPU's, 90 NVIDIA GPU's, 28 Xeon Phi nodes

National Energy Reseach Scientific Computing Center (www.nersc.gov)

- **Cori:** Cray XC40, 700,000+ cores
- **Edison:** Cray XC30, 67000+ 2-socket cores, 64GB RAM/node

Example: generation of NLO ApplGrids

-without kinematic cuts; for PDF comparisons

-with kinematic cuts and a variety of QCD scales; for the PDF analysis http://metapdf.hepforge.org/2016_pdf4lhc/

Predictions for LHC observables based on PDF4LHC15 PDFs

SMU gallery for basic processes at 7, 8, 13 TeV

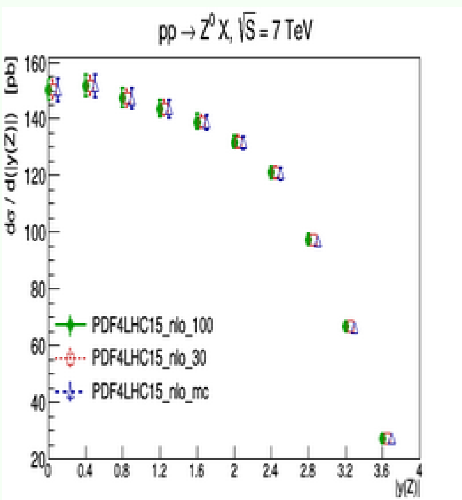
Developed by Bo Ting Wang and Keping Xie

arXiv: 1605.04692

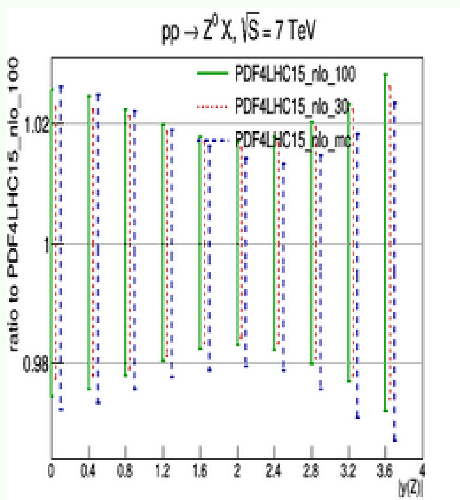
1. $p + p \rightarrow Z + X$
   - PDF4LHC15 PDFs (NLO)
   - PDF4LHC15 PDFs (NNLO)
   - PDF4LHC15_nnlo_100, HERA, ABM PDFs
   - PDF4LHC15_nnlo_100, CT14, MMTH14, NNPDF3.0 PDFs
2. $p + p \rightarrow W^+ + X$
   - PDF4LHC15 PDFs (NLO)
   - PDF4LHC15 PDFs (NNLO)
   - PDF4LHC15_nnlo_100, HERA, ABM PDFs
   - PDF4LHC15_nnlo_100, CT14, MMTH14, NNPDF3.0 PDFs

$p + p \rightarrow Z + X$

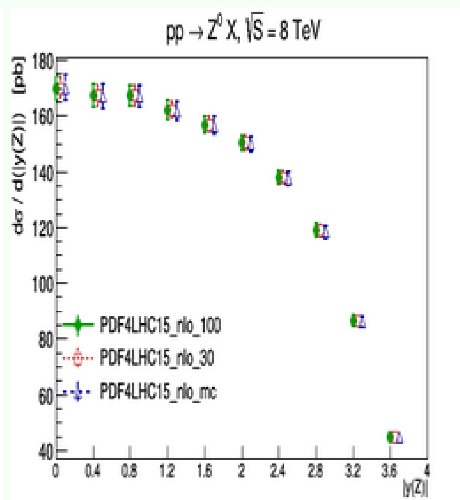
PDF4LHC15 PDFs (NLO)



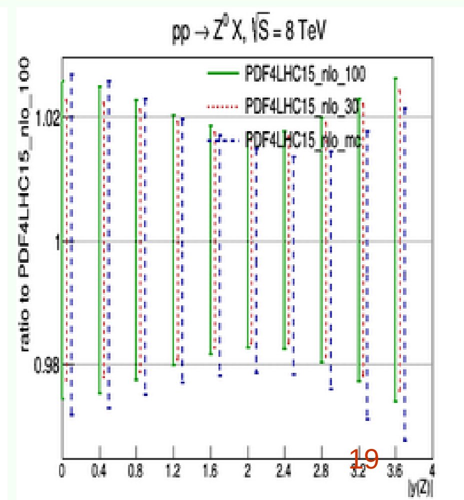
[pdf]



[pdf]



[pdf]



[pdf]

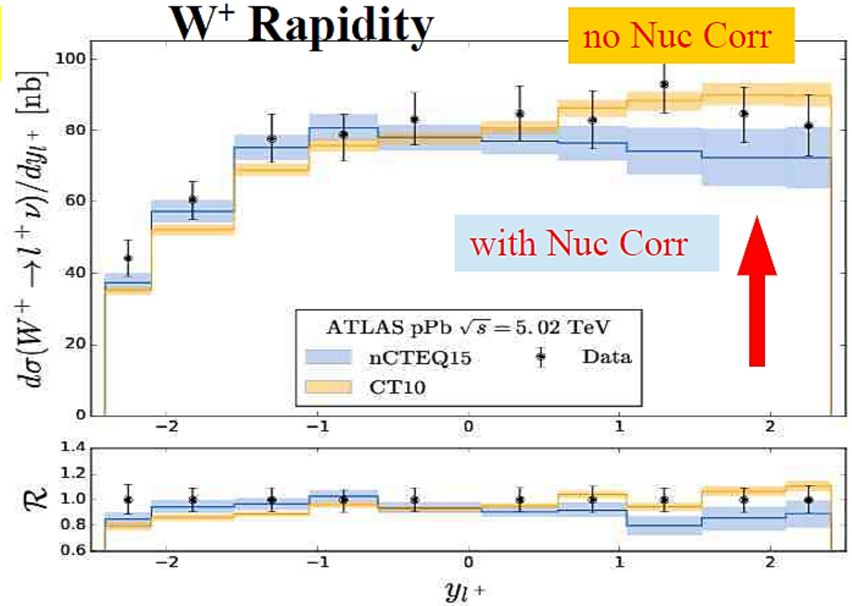
Gallery of phenomenological comparisons for LHC

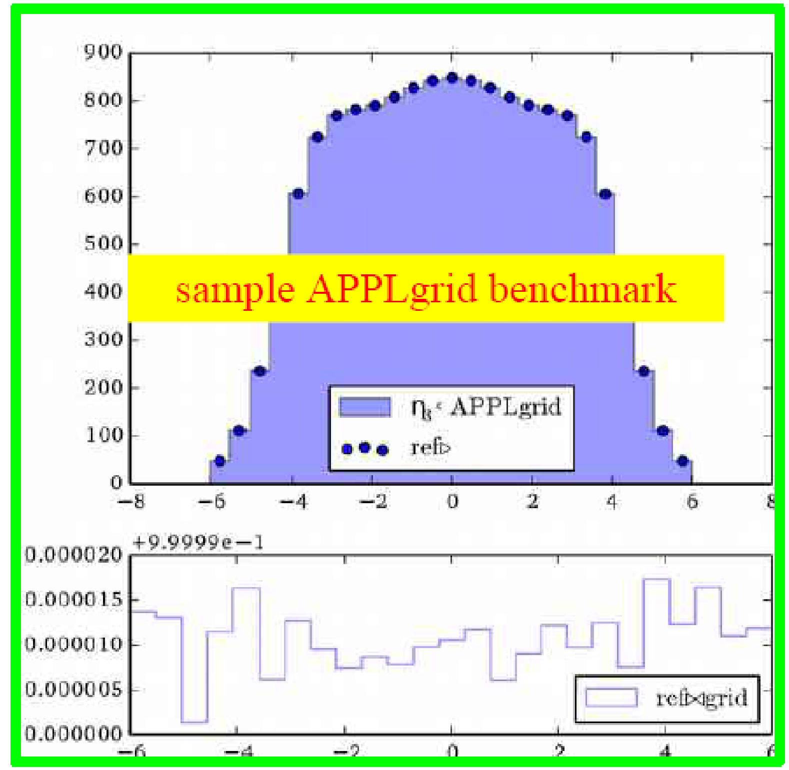
Process	Order	Type of calculation
•p + p $\rightarrow$ Z + X	NLO	aMCFast/APPLgrid
•p + p $\rightarrow$ W^+ + X	NLO	aMCFast/APPLgrid
•p + p $\rightarrow$ W^- + X	NLO	aMCFast/APPLgrid
•p + p $\rightarrow$ $t\bar{t}$ + X	NLO	aMCFast/APPLgrid
•p + p $\rightarrow$ $t\bar{t}$ + X	NLO	aMCFast/APPLgrid
•p + p $\rightarrow t\bar{t}\gamma\gamma$+X	NLO	aMCFast/APPLgrid
•ATLAS inclusive jets	NLO	NLOJET++/APPLgrid
•ATLAS inclusive dijets	NLO	NLOJET++/APPLgrid
•P + p $\rightarrow$ W^+ c + X	NLO	aMCFast/APPLgrid
•P + p $\rightarrow$ W^- c + X	NLO	aMCFast/APPLgrid
•P + p $\rightarrow$ H + X	LO,NLO	MCFM
•P + p $\rightarrow$ H+ jet + X	LO, NLO	MCFM

Compared PDFs: PDF4LHC15_100, _30, _MC, ABM'12, CT14, HERA2.0, MMHT14, NN3.0

Both full (MCFM) and fast (ApplGrid) calculations. AppGrlids are generated with minimal cuts and can be downloaded.

ApplGrid calculations for nCTEQ'15 $\left(p + Pb \to (W,Z)X\right)$





lead kinematics

larger x ⟷ **smaller x**

larger x	smaller x
"OK" nuclear correction	too much suppression
previous data constraints	minimal data constraints

APPLgrids are being generated for nuclear scattering processes

See talk by Olek Kusina on Tuesday afternoon

Example: **mcgen**, core functionalities

1. Monte-Carlo sampling of Hessian PDFs in the LHAPDF6 format according to various probability distributions $P(f(x, Q_0; a))$: normal, log-normal, Watt-Thorne, with/without the bias correction
   – captures mild non-Gaussian effects that are important (e.g., positivity of PDFs), when the PDF parametrization forms are unknown or difficult to use
2. Inline algebraic operations with LHAPDF6 grids (addition, multiplication, averaging of LHAPDF .dat files)

```
Usage examples
 > mcgen.x generate mcgen.card
 > mcgen.x average average.dat input1.dat input2.dat ...
 > mcgen.x add sum.dat input1.dat input2.dat weight1 weight2
 > mcgen.x multiply prod.dat input1.dat input2.dat power1
power2
```

Why advanced Monte-Carlo sampling methods?

The Watt-Thorne sampling method *(arXiv:1205.4024)* is adequate when nonlinearities are small (e.g., for symmetrically distributed PDFs at moderate x)

It has problems with reproducing asymmetric bands (which may be biased), or with reproducing positive-definite PDFs at large x

This is traced to interesting aspects of asymmetric probability distributions, as discussed in version 2 of 1607.06066.

Log-normal sampling with a bias correction in **mcgen** resolves these issues. It produces positive-definite MC replicas, with the MC mean PDF equal to the Hessian central PDF, and with the MC and Hessian uncertainties in close agreement.

CT14 Monte-Carlo replica ensembles (MC1 and MC2) with asymmetric errors and positivity

arXiv:1607.06066

N_{rep} Monte-Carlo replicas are constructed from predictions $X_{\pm i}$ for Hessian eigenvector sets as

NEW

$$X^{(k)} = X(\{0\}) + \delta X^{(k)} - \Delta$$

$$\delta X^{(k)} \equiv \sum_{i=1}^{D} \frac{X_{+i} - X_{-i}}{2} R_i^{(k)} + \frac{1}{2} \sum_{i,j=1}^{D} (X_{+i} + X_{-i} - 2X_0) \left(R_i^{(k)}\right)^2$$

Random real values $R_i^{(k)}$ are sampled from the standard normal distribution

CT14 MC1 replicas are constructed from $X = f_a(x, Q_0)$, can be negative
CT14 MC2 replicas are constructed from $X = log\,[f_a/f_{a0}]$, where $f_{a0}(x, Q_0)$ is the central Hessian PDF; **each replica PDF is non-negative**

The uncertainties are given by **asymmetric** standard deviations,

$$\delta_{68}^{MC,>} X = \sqrt{\langle (X - \langle X \rangle)^2 \rangle_{X > \langle X \rangle}}$$

$$\delta_{68}^{MC,<} X = \sqrt{\langle (X - \langle X \rangle)^2 \rangle_{X < \langle X \rangle}}$$

NEW

f_a/f_{a0} values for individual replicas

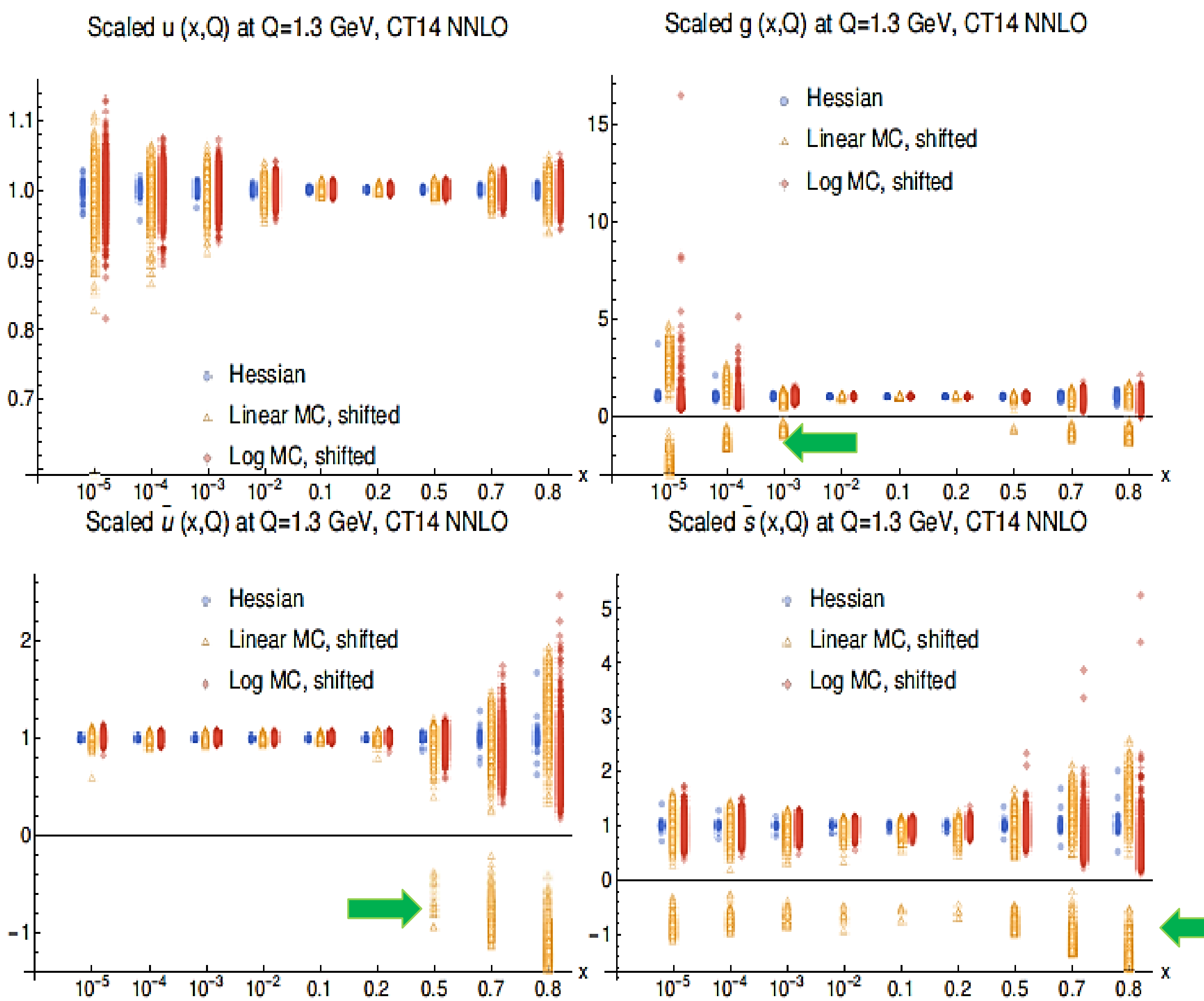


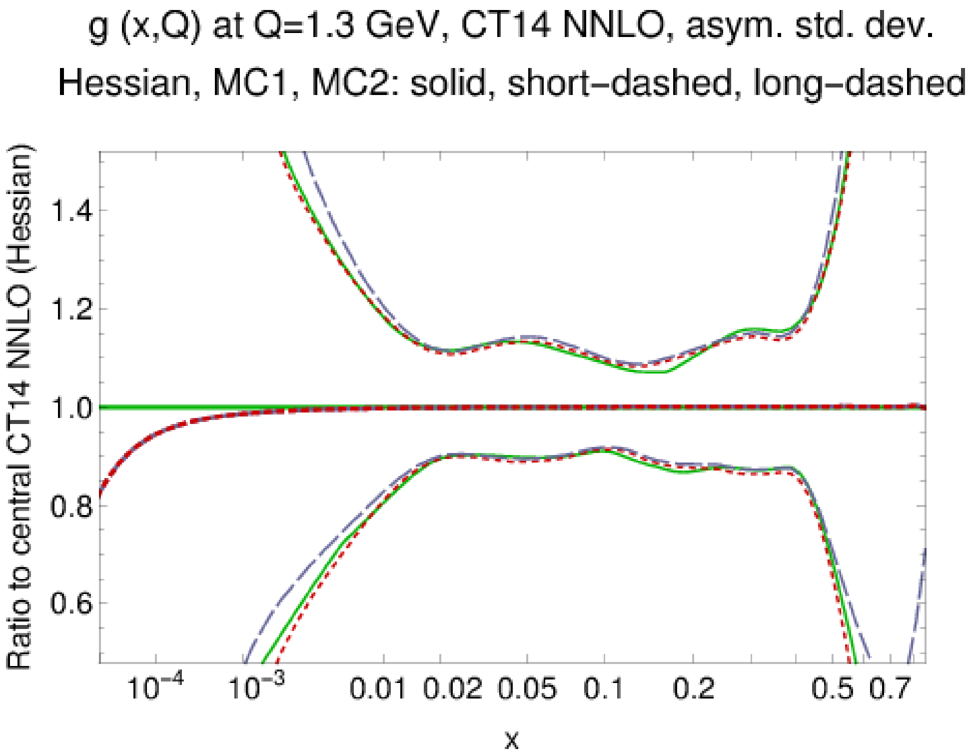
FIG. 2: Distributions of individual replicas for MC1 (linear MC, shifted) and MC2 (log MC, shifted) ensembles.

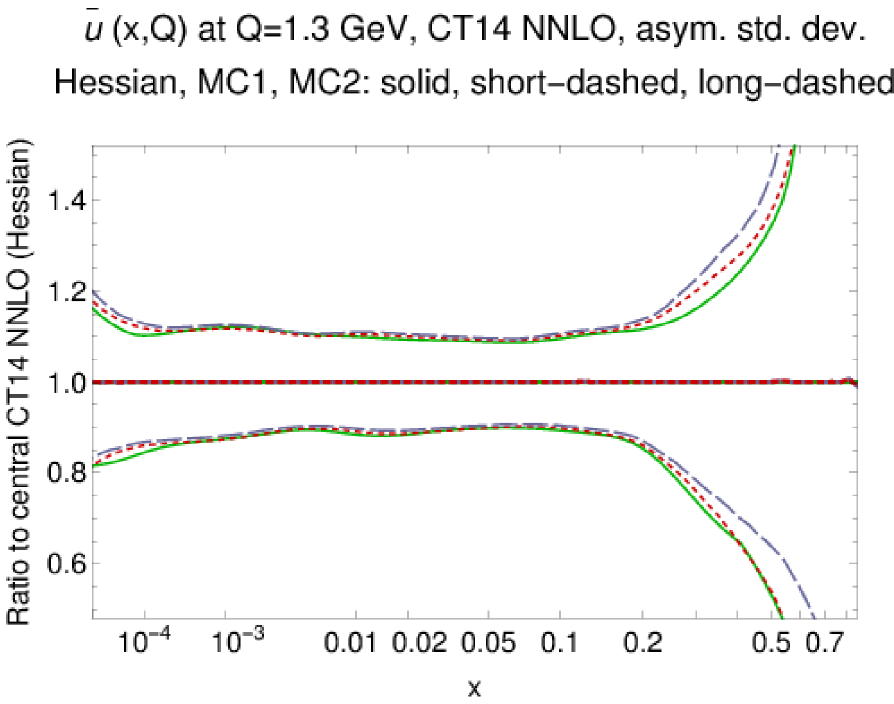
The vertical axes have scale $\left|\frac{f}{f_{a0}}\right|^{0.2} sign(f)$ to visualize relative variations of $\pm$ signs in an extended magnitude range

A fraction of 1000 MC1 replica PDFs can be negative (**green arrows**).

But, all Hessian and MC2 PDFs are positive

Asymmetric standard deviations for PDFs





Excellent agreement between the Hessian, MC1, MC2 bands at intermediate x

More pronounced differences at small and large x; there are ambiguities in reconstructing MC replicas from the Hessian PDFs when PDF uncertainties are large

Comparisons for other flavors at
http://hep.pa.msu.edu/cteq/public/ct14/MC/

Software for PDF analysis developed at SMU

1. **SACOT-chi at NNLO + IM scheme at NNNLO** in **xFitter**

(T.-J. Hou, M. Guzzi, Bowen Wang)

- Ready to be used in comparisons to other schemes, studies of $m_c(m_c)$, etc.

2. SMU's local expertise in generating **APPLgrid** using MCFM and aMCFast on US-based high-performance computing facilities
   - Can coordinate APPLgrid generation with other groups

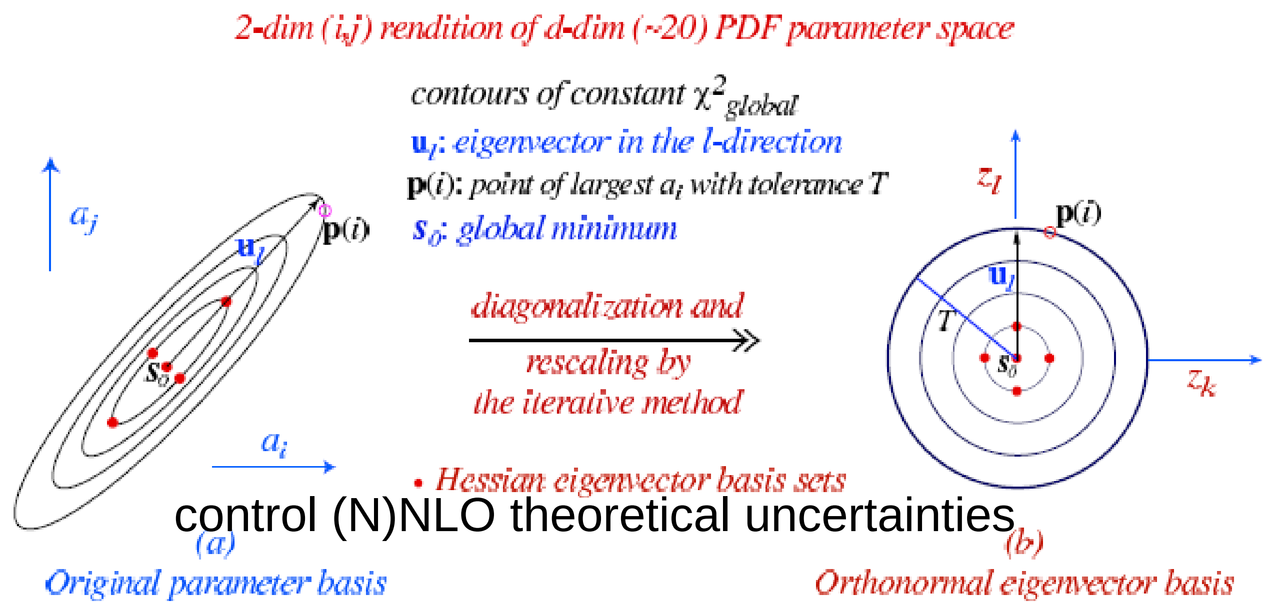
3. **Standalone public codes** – can be used/distributed with xFitter:

- **MCGEN** (metapdf.hepforge.org/mcgen): a C++ code to convert Hessian eigenvector PDF sets in the LHAPDF format into MC replica sets according to several methods

- **Mathematica packages**:
   a. Combination of PDF sets using meta-parametrizations
   b. N-dimensional matrix tools: reduction of correlation matrices, reconstruction of n-dimensional ellipsoids from discrete data,….

Backup slides

Hessian PDFs



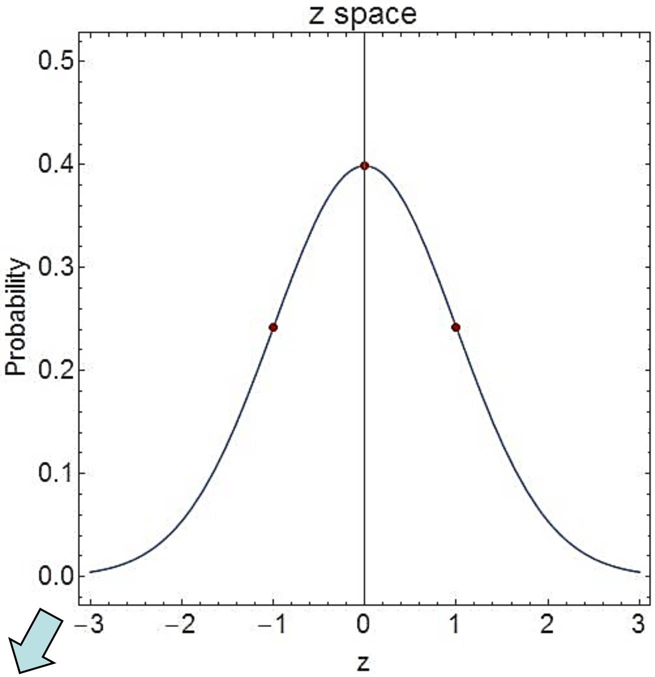
control (N)NLO theoretical uncertainties

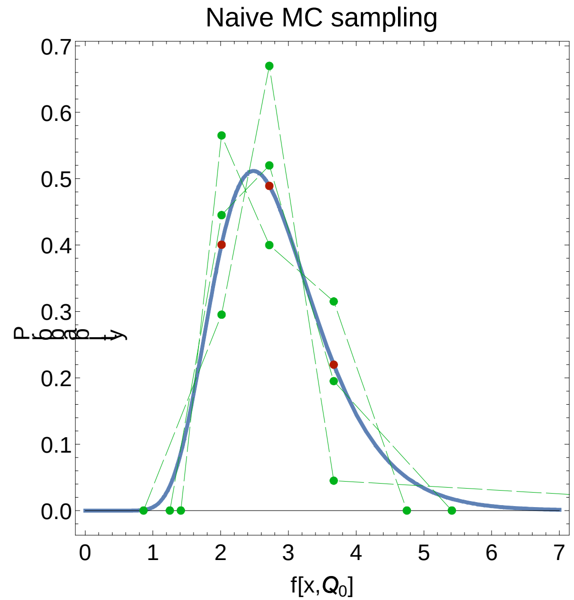
Along any direction z_i:

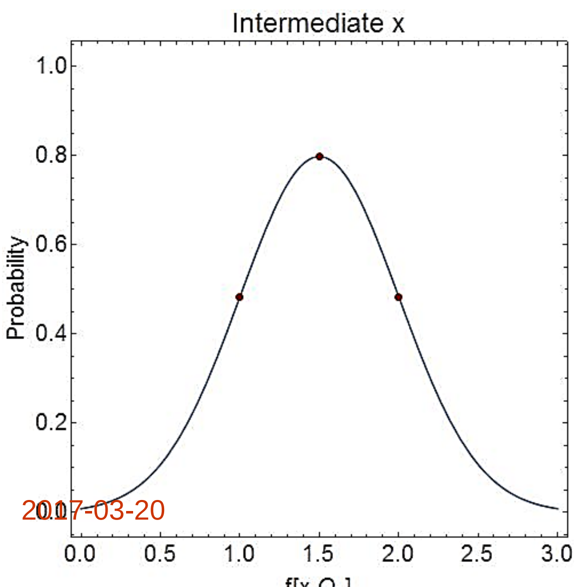
- the best-fit set corresponds to cumulative probability $\mathrm{P(z)} = 0.5$
- two 68% extreme sets $z_i^{\pm}$ correspond to $P(z) = 0.16$ and 0.84

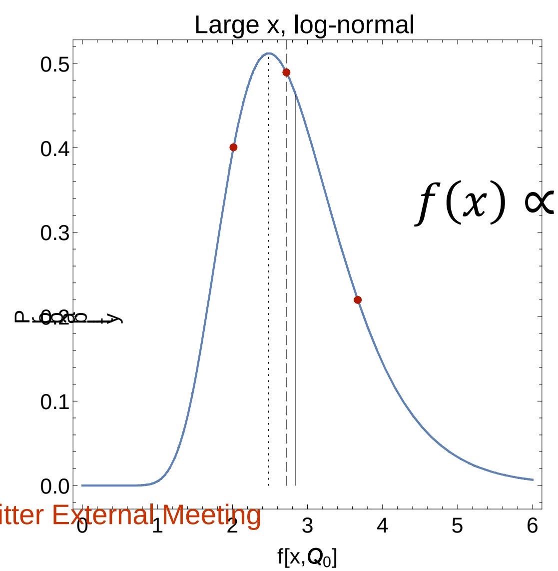
Cumulative probabilities $P(f(z))$ for the corresponding PDFs do not need to be 0.5, 0.16, and 0.84!

Beyond the linear approximation,



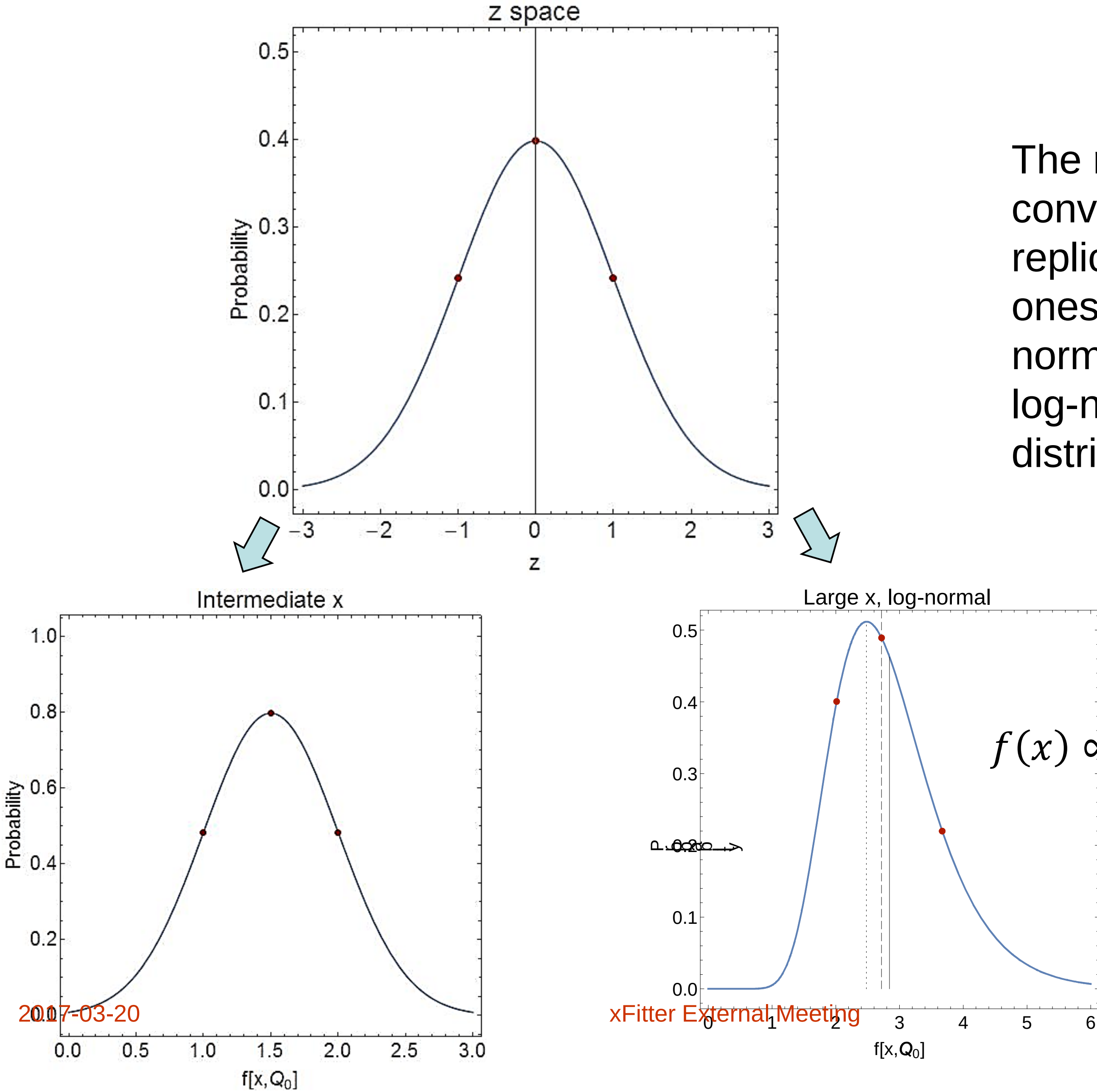






$$f(x) \propto (1-x)^{a_2}$$

The x-space probability can be non-Gaussian



The **mcgen** code converts Hessian replicas into the MC ones according to the normal (CT14 MC1) or log-normal (CT14 MC2) distributions

$$f(x) \propto (1-x)^{a_2}$$