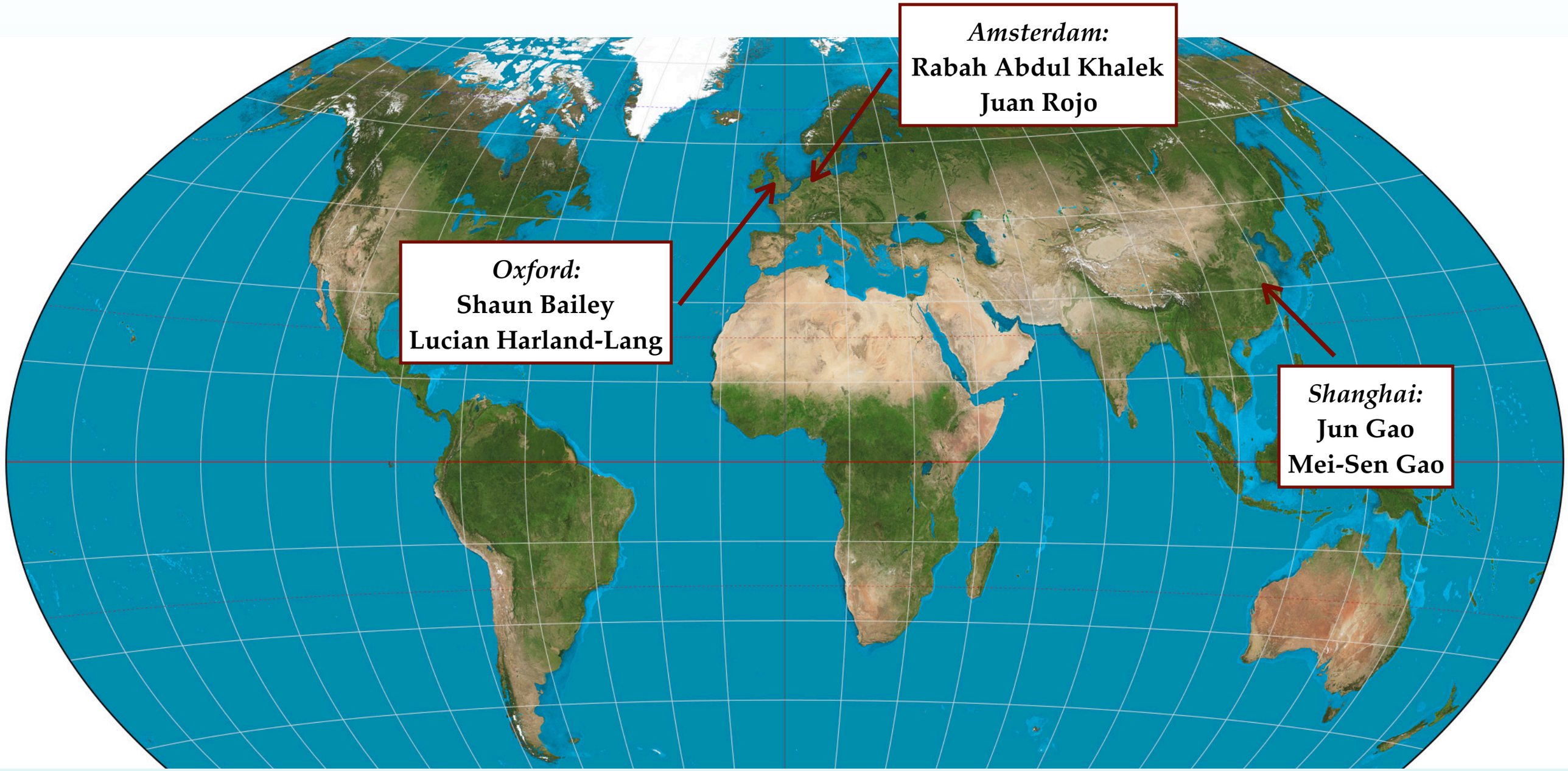
Quantifying the impact of HL-LHC (pseudo-)data on parton distributions

Juan Rojo
VU Amsterdam & Theory group, Nikhef

HL/HE-LHC Working Group 1 Meeting - QCD Physics

CERN, 02/05/2018

The team



Contactpersons within the LHC experiments:
Mario Campanelli (ATLAS)
Katerina Lipka (CMS)
Katharina Mueller (LHCb)

Towards ultimate parton distributions at the HL-LHC

- The HL-LHC will make possible a range of **measurements with direct impact on the PDFs**, for instance on the poorly known **large-*x* region**

- Together with the PDF contacts of ATLAS, CMS, and LHCb, our plan is the following:

  - Assess the **kinematical coverage** and **experimental uncertainties** that we can expect by the end of the HL-LHC on relevant PDF-sensitive measurements

  - To generate **APPLgrid/FastNLO tables** for these processes

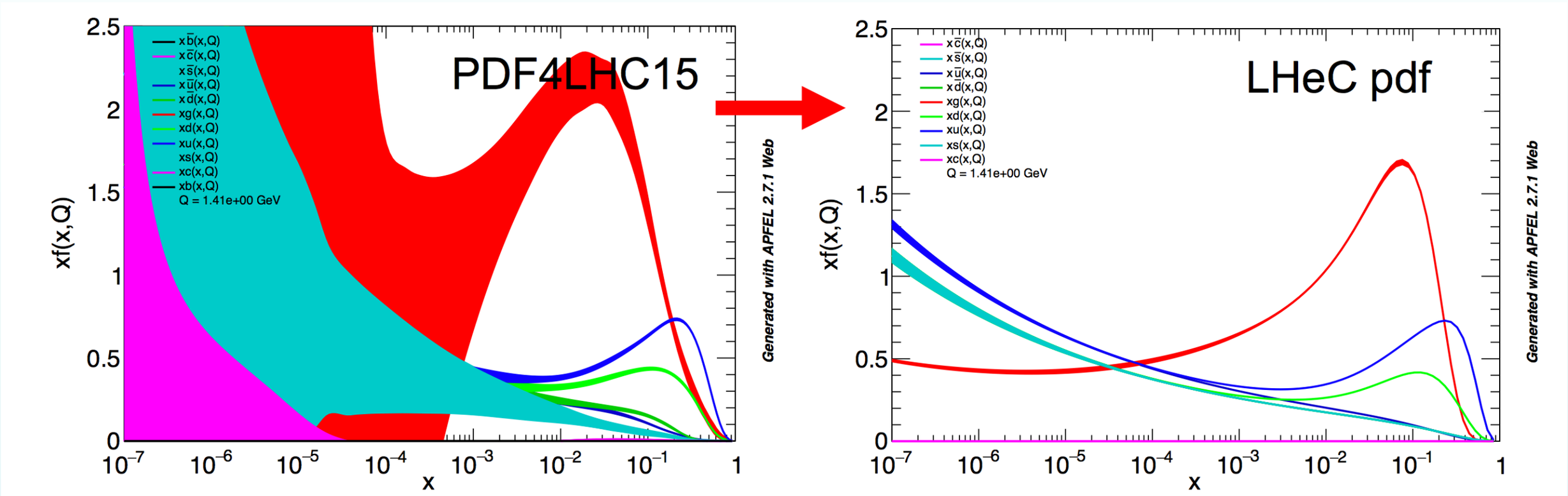
  - To include them in the PDF fit to gauge their **impact on the PDFs** and determine the ultimate precision in PDFs that can be expected in the HL-LHC era

  - To make the resulting PDF sets **public** in LHAPDF format so that they can be used for other HL/HE-LHC feasibility studies

*Similar studies have been carried out in the context of future **high-energy lepton-proton colliders**, but this is the first time that PDF prospects are quantified systematically for **future proton-proton colliders***

Towards ultimate parton distributions at the HL-LHC

The constraints on PDFs **from LHeC/FCC-eh measurements** have been studied in some detail

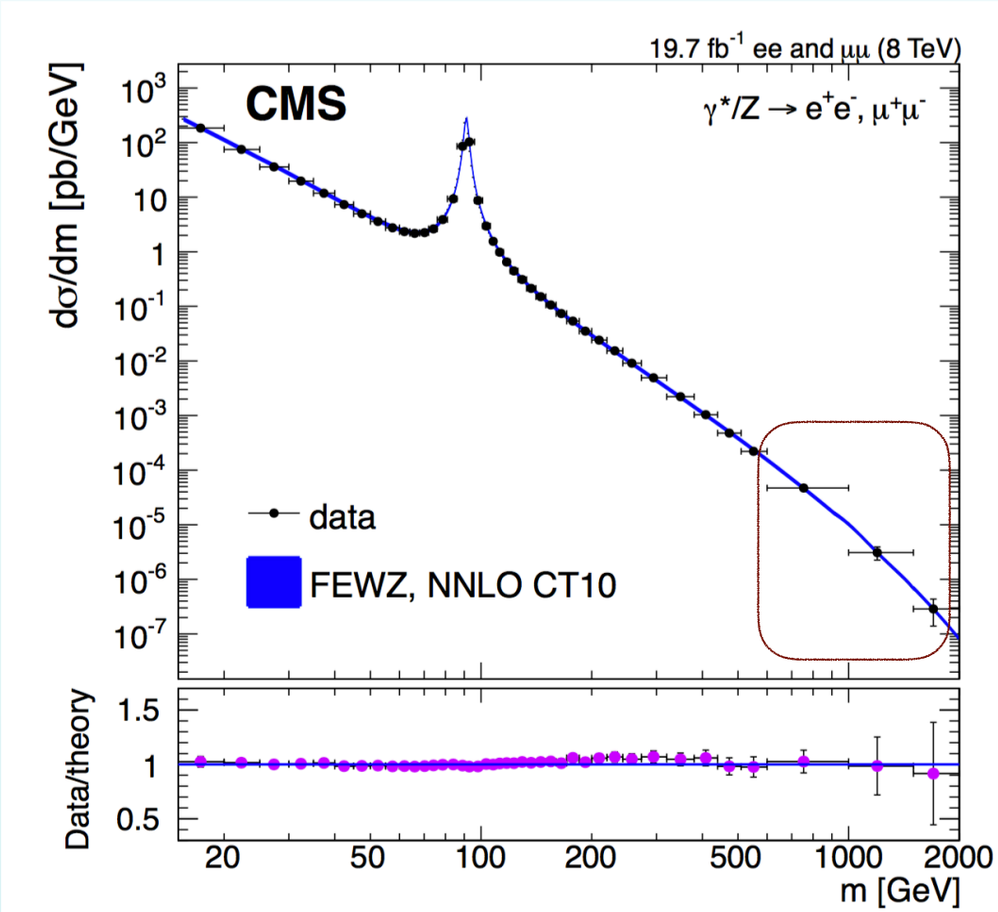


Our goal is to determine the corresponding picture in the case of the HL-LHC!

*Similar studies have been carried out in the context of future **high-energy lepton-proton colliders**, but this is the first time that PDF prospects are quantified systematically for **future proton-proton colliders***

Why better PDFs at the HL-LHC?

- BSM physics could manifest as **subtle deviations** wrt to the Standard Model predictions
- Even for high-mass resonances, **PDF uncertainties degrade or limit many BSM searches**
- The robustness of **global stress-tests of the SM** (electroweak fit, SM Effective Field Theory analysis) relies crucially in high-precision theoretical calculations



SMEFT expansion

$$\sigma(E) = \sigma_{SM}(E)\left(1 + \epsilon\frac{m_{SM}^2}{m_W^2} + \epsilon\frac{E^2}{m_W^2} + \ldots\right)$$

For $\mathbf{E} \simeq \mathbf{1}\ \mathbf{TeV}$, a measurement with $\delta\sigma/\sigma \simeq \mathbf{10}\%$ is sensitive to $\epsilon \simeq \mathcal{O}(\mathbf{0.1}\%)$!

BSM physics might very well hiding itself in the tails of LHC distributions, but need to make sure first that PDF uncertainties are under control

The roadmap

- A non-exhaustive list of PDF-sensitive processes at the HL-LHC, with emphasis on **high-p_T region:**
  - ☑ High-mass Drell-Yan, both neutral and charged current (*quark flavour separation*)
  - ☑ The transverse momentum of the W and Z bosons (*gluon, antiquarks*)
  - ☑ Top-quark differential distributions (*gluon*)
  - ☑ Inclusive jets and dijets (*gluon, quarks*)
  - ☑ photon/W/Z+charm (*strange, charm*)
  - ☑ Direct (prompt) photon production (*gluon*)
- These processes can be computed with state-of-the-art theory: **NNLO QCD and NLO EW (incl. PI)**
- However since here we deal with **pseudo-data**, **NLO QCD** theory is sufficient. For same reason, no need to account for PI effects (do not affect PDF sensitivity either)
- In each case, generate pseudo-data with a sensible binning and different assumptions (conservative to optimistic) for the **systematic uncertainties** with input from the LHC PDF contacts
- Then determine **their impact in a global analysis**: this will represent the ultimate precision that one can expect on PDFs in the HL-LHC era

The roadmap

Generate **NLO APPLgrids** and the corresponding pseudo-data

Explore different options for **binning** and **systematic uncertainties**

↓

Quantify the effects of the individual processes on the

PDF4LHC15 set using Hessian profiling (*e.g.* with **ePump**)

↓

Combine all pseudo-data and perform **full-fledged fits** using

the CT, MMHT, and NNPDF frameworks

↓

Combine and reduce results using **meta-PDFs / MC2H compression**

and make the resulting sets **available via LHAPDF**

The roadmap

We are here!

Generate **NLO APPLgrids** and the corresponding pseudo-data

Explore different options for **binning** and **systematic uncertainties**

↓

Quantify the effects of the individual processes on the

PDF4LHC15 set using Hessian profiling (*e.g.* with **ePump**)

↓

Combine all pseudo-data and perform **full-fledged fits** using

the CT, MMHT, and NNPDF frameworks

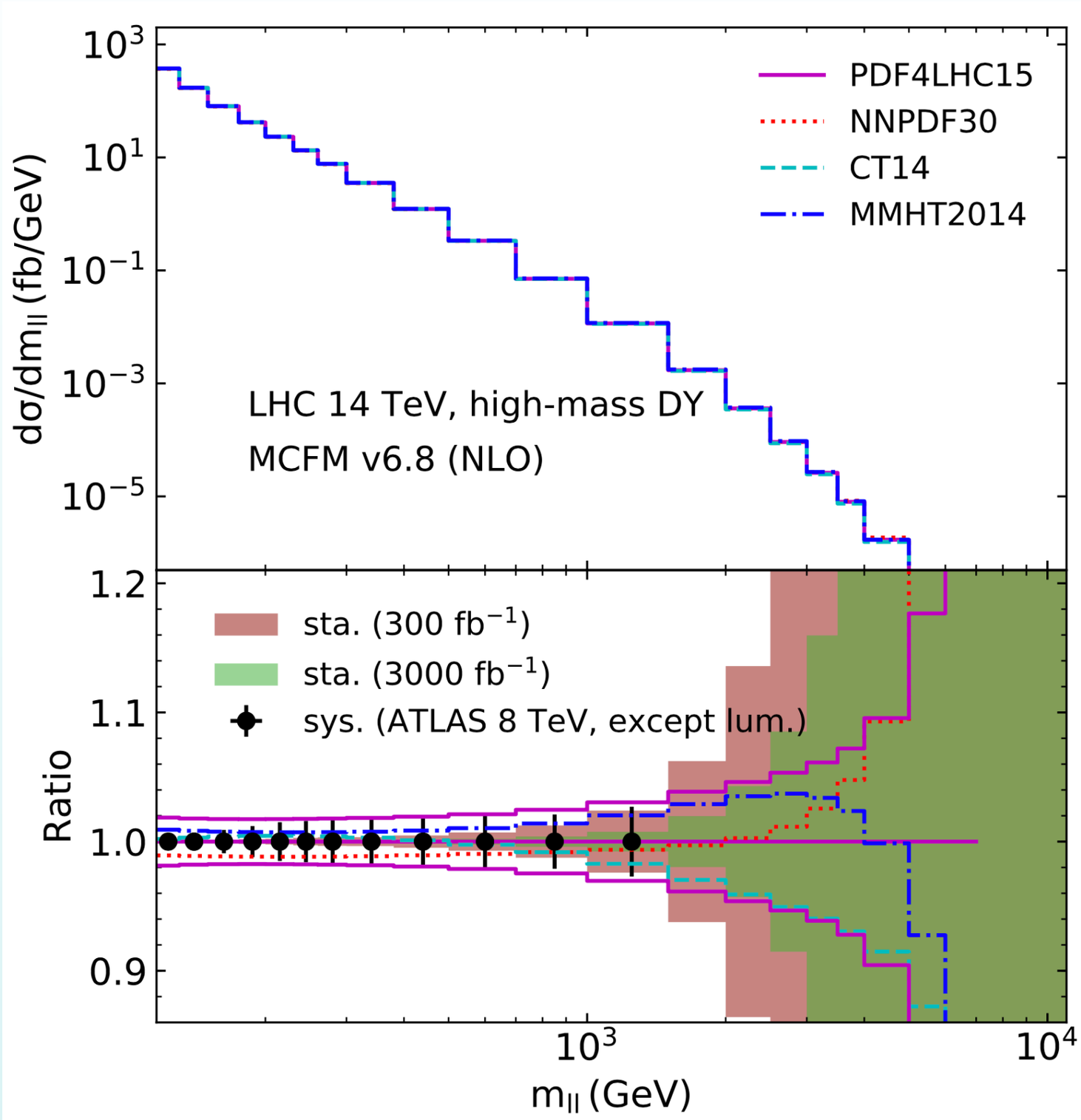
↓

Combine and reduce results using **meta-PDFs / MC2H compression**

and make the resulting sets **available via LHAPDF**

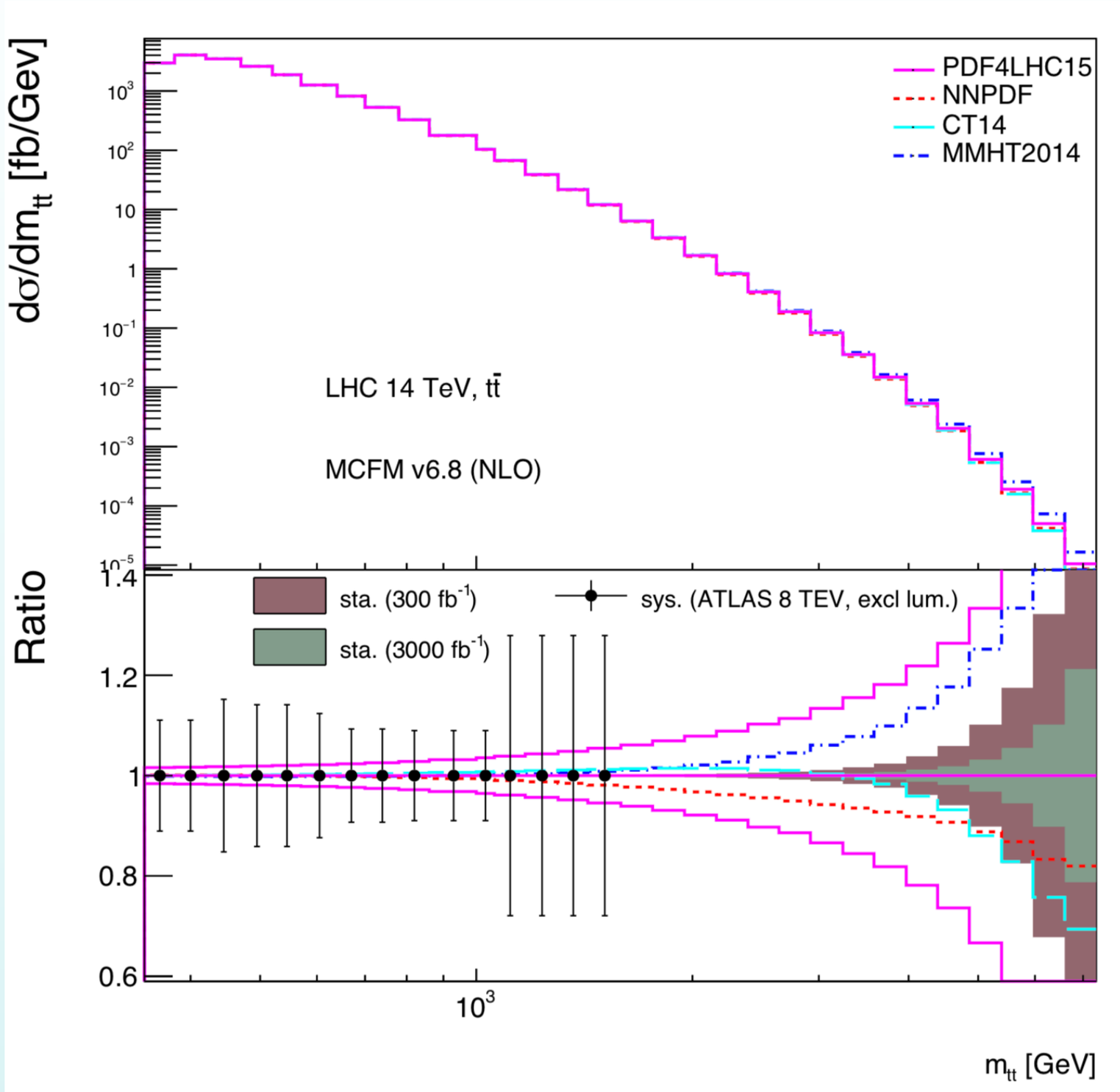
High-mass Drell-Yan

- Invariant mass distributions of dilepton pairs m_{ll} at 14 TeV
- Systematic errors from 8 TeV ATLAS measurement. Now discussing how to extend to HL-LHC case
- Binning still being optimised
- HL-LHC measurements **up to m_{ll}=3 TeV** should constrain the **poorly known large-*x* antiquarks**
- Need to account for the **effect of acceptances** in the statistical uncertainties



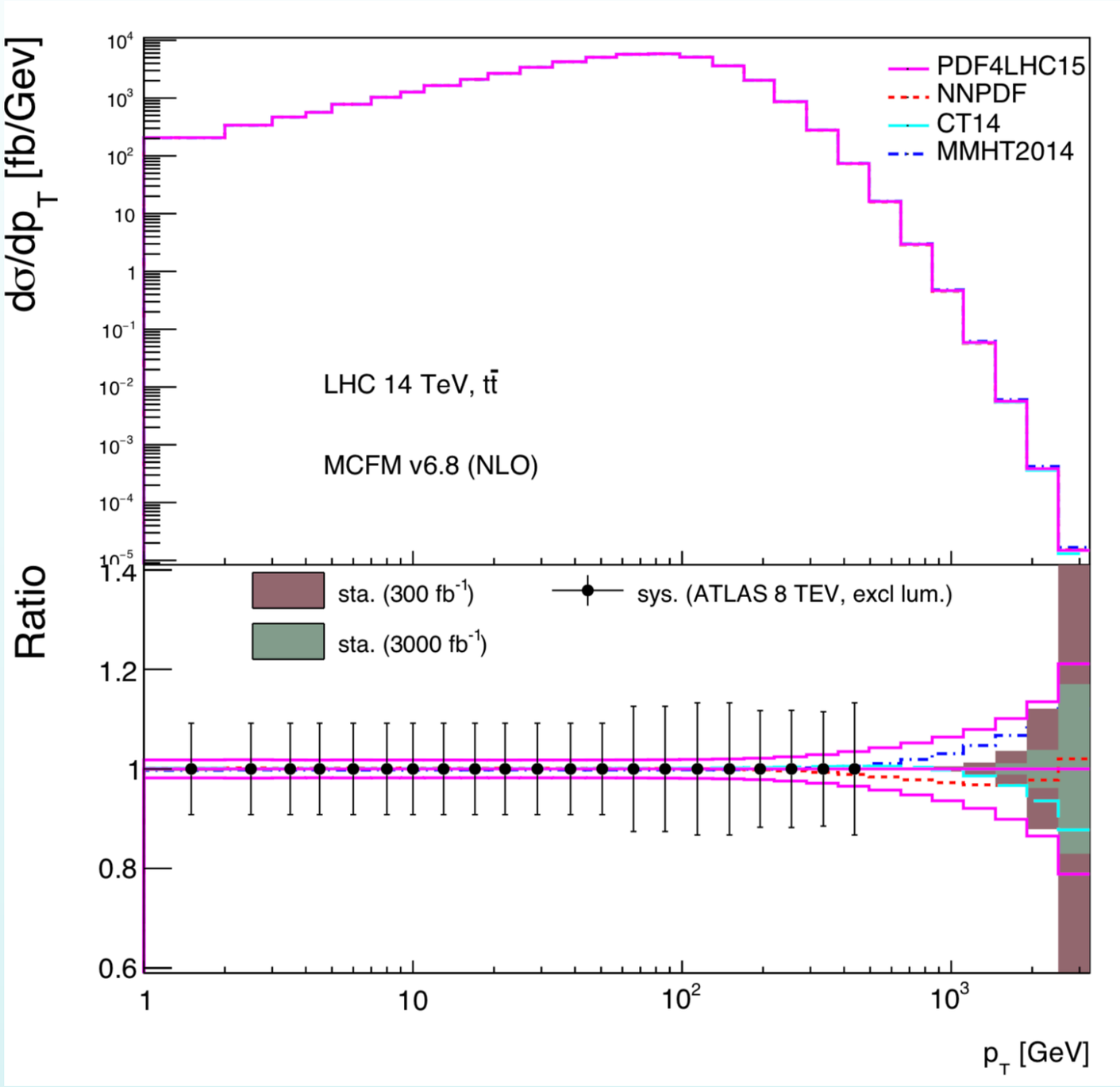
Top-quark pair production

- Differential distributions in top-quark pair production at 14 TeV. Optimisation of the binning still in progress.
- Systematic errors from 8 TeV ATLAS measurement. Now discussing how to extend to HL-LHC case
- Here showing the pair invariant mass distribution $\mathbf{m_{tt}}$
- HL-LHC measurements **up to $\mathbf{m_{tt}}$=4 TeV** should constrain the **poorly known large-*x* gluon,** but need improved control over systematics
- Still need to account here for the **effect of acceptances** in the statistical uncertainties



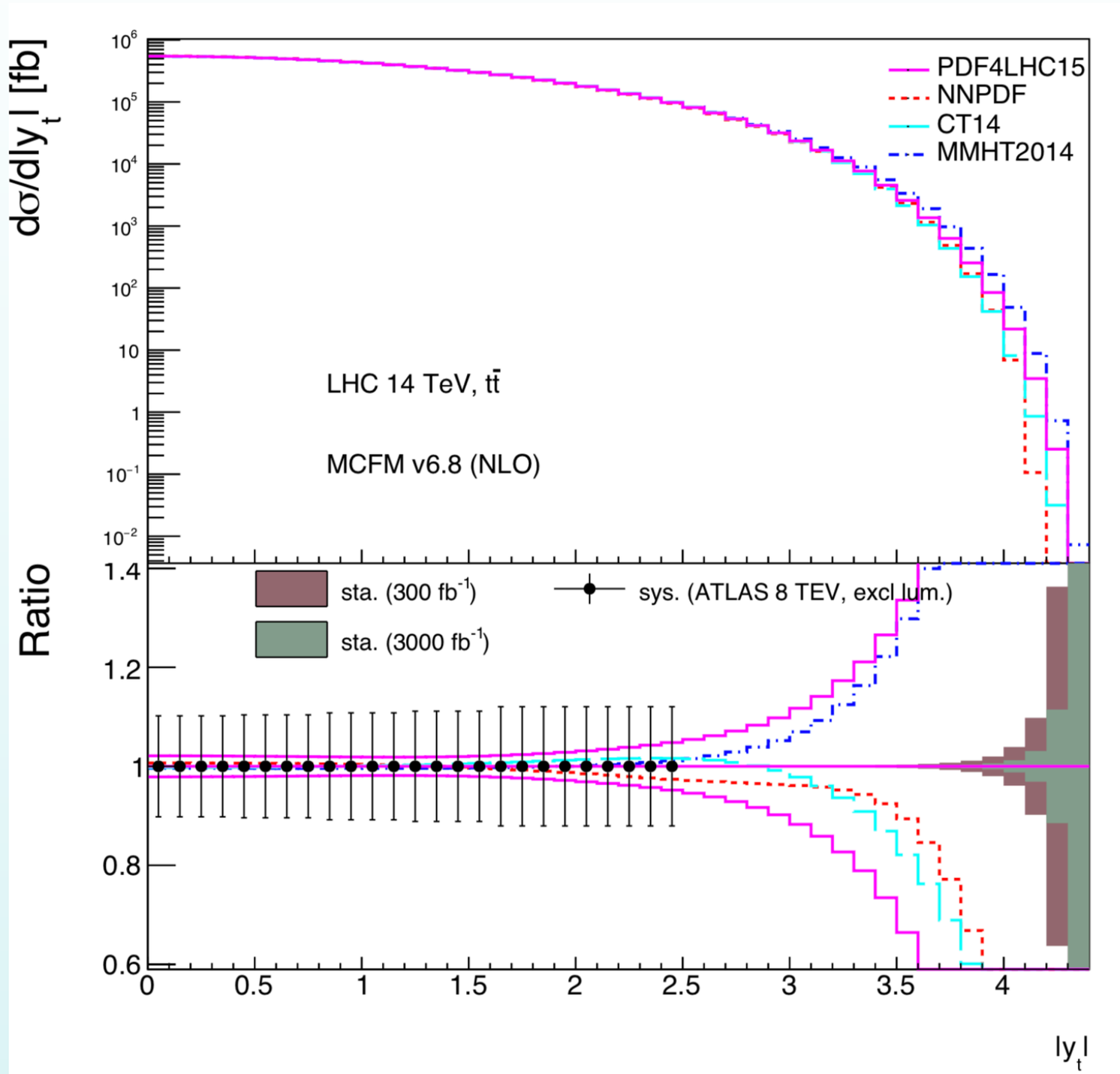
Top-quark pair production

- Differential distributions in top-quark pair production at 14 TeV. Optimisation of the binning still in progress.
- Systematic errors from 8 TeV ATLAS measurement. Now discussing how to extend to HL-LHC case
- Here showing the top transverse momentum distribution
- HL-LHC measurements **in the TeV** region should constrain the **poorly known large-*x* gluon,** but need improved control over systematics
- As before, still need to account here for the **effect of acceptances** in the statistical uncertainties



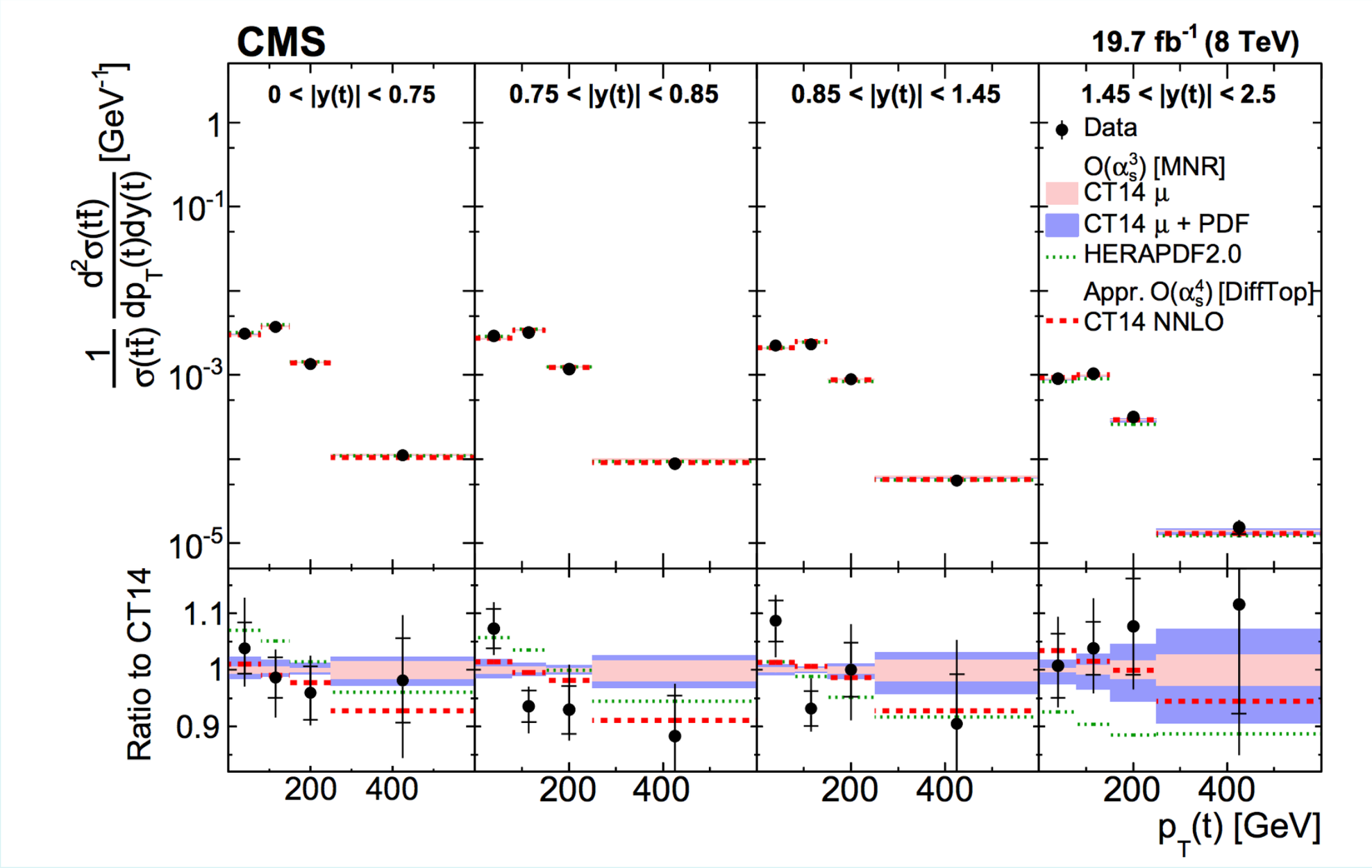
Top-quark pair production

- Differential distributions in top-quark pair production at 14 TeV. Optimisation of the binning still in progress.
- Systematic errors from 8 TeV ATLAS measurement. Now discussing how to extend to HL-LHC case
- Here showing the top rapidity distribution
- HL-LHC measurements **in the TeV** region should constrain the **poorly known large-*x* gluon,** but need improved control over systematics
- As before, still need to account here for the **effect of acceptances** in the statistical uncertainties



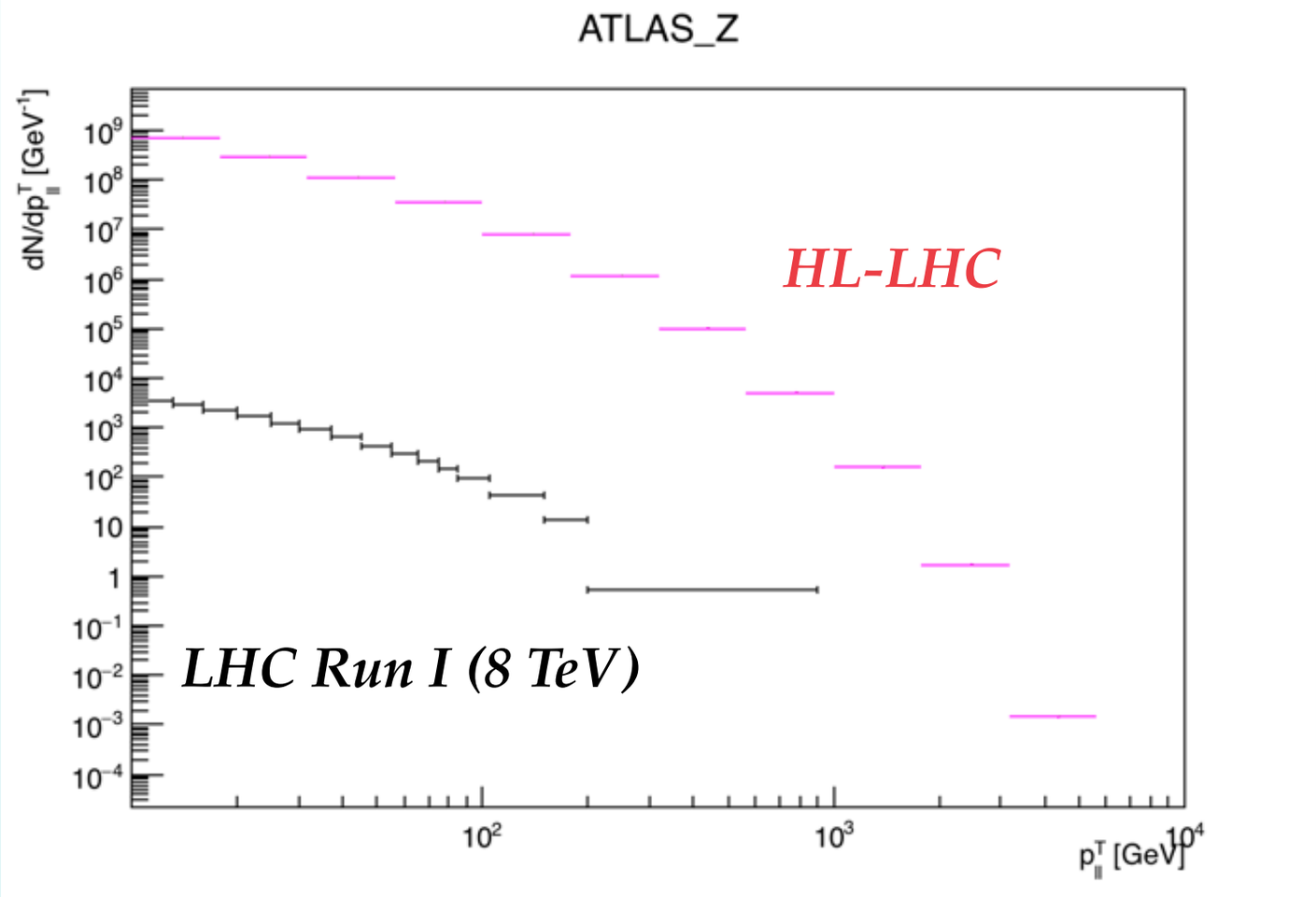
Top-quark pair production

- Generate also pseudo-data for **double-differential distributions**, along lines of CMS measurement
- Assume that **correlations between different distributions are know**, so that one can include in the fit multiple distributions without double counting



Transverse momentum of Z bosons

- Differential distributions in inclusive *Z* boson production
- Here showing the transverse momentum of **Z** bosons
- Still working on binning and systematics, but is clear the **marked improvement as compared to the 8 TeV case**
- Need to account for the **effect of acceptances** in the statistical uncertainties



Quantifying impact on PDFs

- Although the ultimate goal is to produce a full-fledged fit with the HL-LHC pseudo-data, for individual processes their impact of PDFs will be quantified by applying **Hessian profiling** to PDF4LHC15

Hessian Updating

- PDF parametrization $f(x,Q;\mathbf{z})$: (parameters $\mathbf{z}$)

  best-fit: $f^0 = f(x,Q;\mathbf{0})$, error PDFs: $f^{\pm i} = f(x,Q;\pm\mathbf{e}^i)$

- Updated Chi-square function :

$$\Delta\chi^2(\mathbf{z}) = \Delta\chi^2_{\text{old}}(\mathbf{z}) + \left(X^E_\alpha - X_\alpha(\mathbf{z})\right) C^{-1}_{\alpha\beta}\left(X^E_\beta - X_\beta(\mathbf{z})\right)$$

- Hessian approximation :

$$\Delta\chi^2_{\text{old}}(\mathbf{z}) = T^2\mathbf{z}^2 \quad (T = \text{tolerance parameter})$$

$$X_\alpha(\mathbf{z}) = X_\alpha(\mathbf{0}) + \Delta X_\alpha \cdot \mathbf{z} \quad \text{with} \quad \Delta X^i_\alpha = \tfrac{1}{2}\left(X_\alpha(+\mathbf{e}^i) - X_\alpha(-\mathbf{e}^i)\right)$$

- Minimize to find new best fit:

$$\mathbf{z}^0_{\text{new}} = (\mathbf{1}+\mathbf{M})^{-1}\mathbf{A} \quad \text{with} \quad A^i = \tfrac{1}{T^2}\left(X^E_\alpha - X_\alpha(\mathbf{0})\right) C^{-1}_{\alpha\beta}\Delta X^i_\beta, \quad M^{ij} = \tfrac{1}{T^2}\Delta X^i_\alpha C^{-1}_{\alpha\beta}\Delta X^j_\beta$$

Paukkunen and Zurita, JHEP 1412 (2014) 100

Explore dependence of results with respect to choice of tolerance T

Huston PDF4LHC March 2018

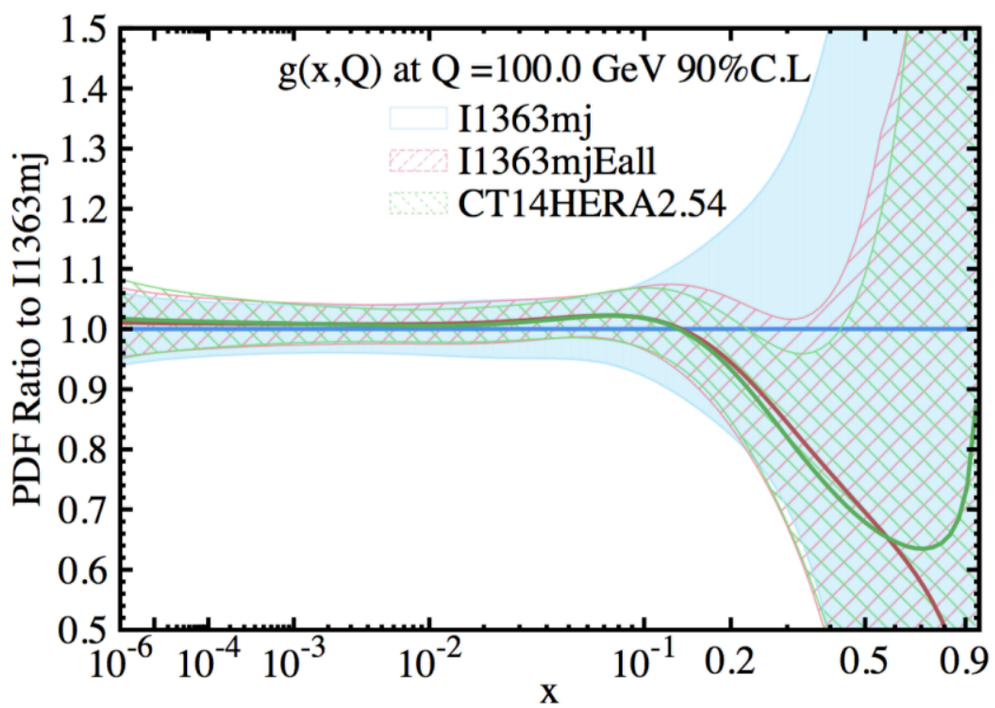
Quantifying impact on PDFs

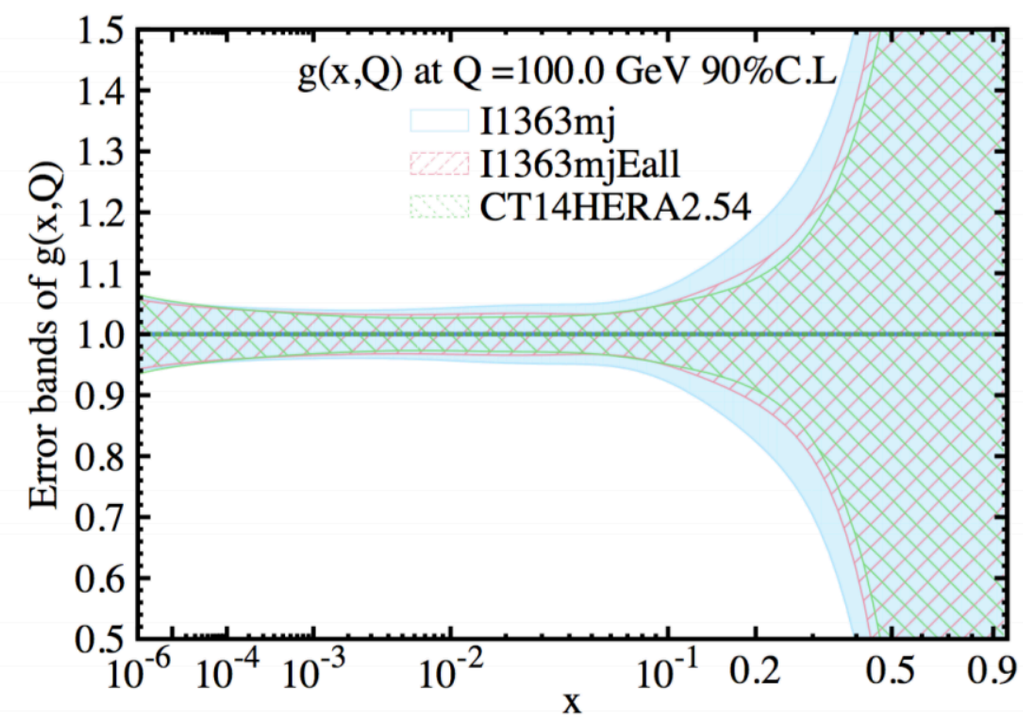
- Although the ultimate goal is to produce a full-fledged fit with the HL-LHC pseudo-data, for individual processes their impact of PDFs will be quantified by applying **Hessian profiling** to PDF4LHC15

Test 1: CT14HERA2 minus Jets

- Remove all CDF, D0, Atlas7, CMS7 jet data from CT14HERA2 and refit.
- Add back the 4 data sets and compare with CT14HERA2.

Hessian profiling (ePump) equivalent to full refit





Final goal: produce full fit, to be combined a la PDF4LHC and then distributed to the HL/HE-LHC community via LHAPDF

Next steps

- Finalise the **production of APPLgrids** for these and the rest of processes to be considered
- Use these grids to generate **HL-LHC pseudo-data** for various scenarios re systematic uncertainties
- Quantify **impact on PDF4LHC15 using Hessian profiling** as implemented in ePump
- Perform full refits and combine into a **"HL-LHC" PDF set,** available for the whole community
- Document this effort in a stand-alone publication, of which part of it will go into the **HL/HE-LHC Yellow Report write-up**