UNIVERSITÉ Clermont Auvergne

ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA

technische universität dortmund

MAX PLANCK INSTITUTE FOR PHYSICS

NNLO QCD predictions for W boson production in association with jets and their use in PDF fits

Master's Thesis Defence

Alberto Saborido Patiño
ATLAS Computing Group, Max Planck Institute for Physics

September 29, 2025

University of Bologna — Dept. of Physics & Astronomy "A. Righi"

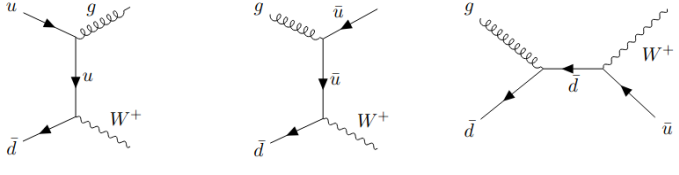
Introduction

The goal is to provide precise QCD predictions for W+jets processes and to use them together with experimental measurements to constrain the proton PDFs.

Introduction

The goal is to provide precise QCD predictions for W+jets processes and to use them together with experimental measurements to constrain the proton PDFs.

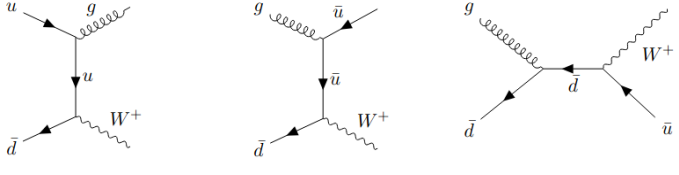
- LO



Introduction

The goal is to provide precise QCD predictions for W+jets processes and to use them together with experimental measurements to constrain the proton PDFs.

- LO

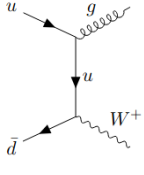


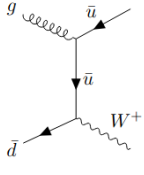
- NLO
  - Real emissions
  - Virtual corrections

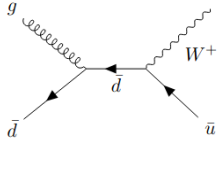
Introduction

The goal is to provide precise QCD predictions for W+jets processes and to use them together with experimental measurements to constrain the proton PDFs.

- LO

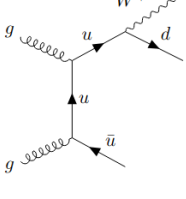


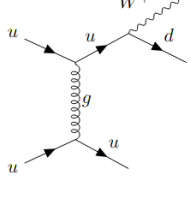


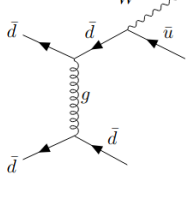


- NLO
  - Real emissions
  - Virtual corrections

  New initial state channels:







Introduction

- NNLO
  - Double real emissions
  - Double virtual corrections
  - Real-virtual corrections

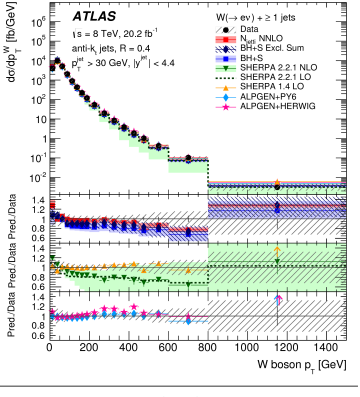
Introduction

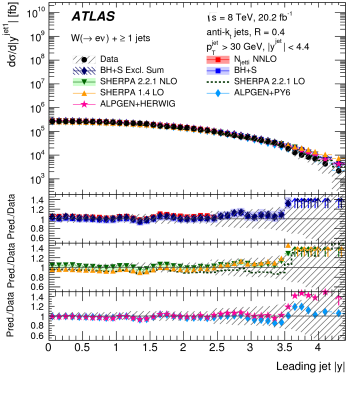
- NNLO
  - Double real emissions
  - Double virtual corrections
  - Real-virtual corrections

At **NLO** and **NNLO**, partonic cross sections contain infrared singularities arising from both real and virtual corrections. These singularities cancel each other but must be properly subtracted when applying numerical methods.

Introduction

We want to assess the impact in PDF fits that ATLAS data measured at $\sqrt{s} = 13\ \mathrm{TeV}$ has, but the corresponding measurement is still ongoing. Instead, the results presented here make use of published data at $\sqrt{s} = 8\ \mathrm{TeV}$ [1].





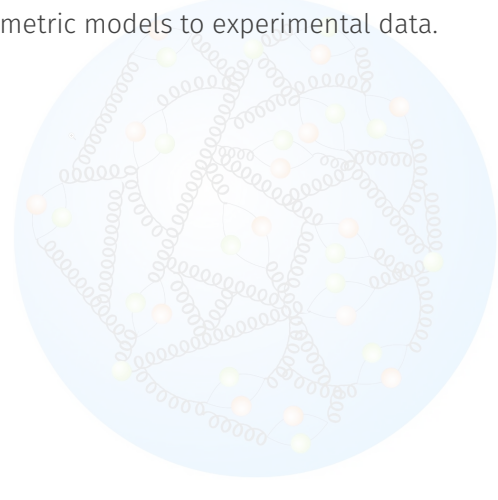
[1] Morad Aaboud et al. In: *JHEP* 05 (2018), p. 077. DOI: *10.1007/JHEP05(2018)077*. arXiv: *1711.03296 [hep-ex]*

Parton Distribution Functions

- Precise PDFs are essential for accurate predictions in proton–proton collisions.

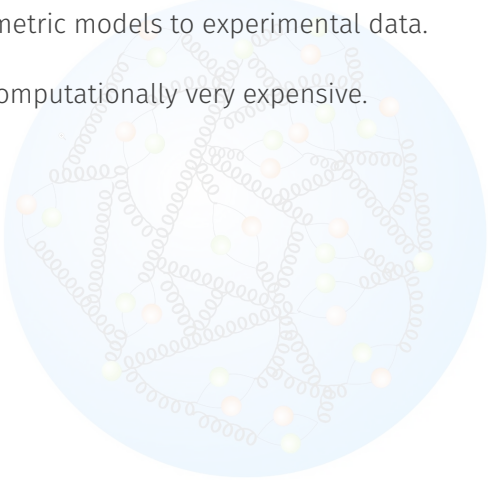
Parton Distribution Functions

- Precise PDFs are essential for accurate predictions in proton–proton collisions.
- PDFs are experimentally obtained by fitting parametric models to experimental data.



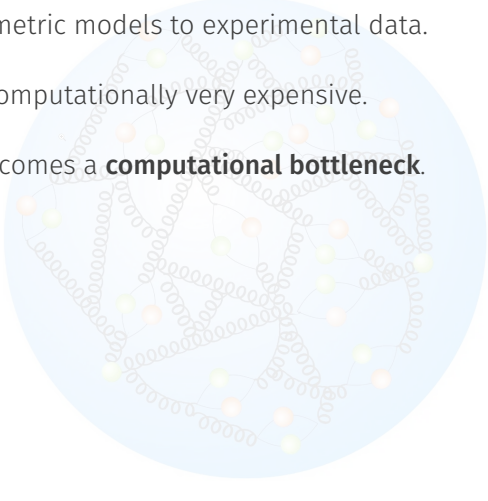
Parton Distribution Functions

- Precise PDFs are essential for accurate predictions in proton–proton collisions.
- PDFs are experimentally obtained by fitting parametric models to experimental data.
- High-order predictions in perturbative QCD are computationally very expensive.



Parton Distribution Functions

- Precise PDFs are essential for accurate predictions in proton–proton collisions.
- PDFs are experimentally obtained by fitting parametric models to experimental data.
- High-order predictions in perturbative QCD are computationally very expensive.
- Scanning different PDFs while performing a fit becomes a **computational bottleneck**.



Parton Distribution Functions

- Precise PDFs are essential for accurate predictions in proton–proton collisions.
- PDFs are experimentally obtained by fitting parametric models to experimental data.
- High-order predictions in perturbative QCD are computationally very expensive.
- Scanning different PDFs while performing a fit becomes a **computational bottleneck**.

SOLUTION: **To decouple the PDF dependence from the calculation of the partonic cross sections**

Parton Distribution Functions

- Precise PDFs are essential for accurate predictions in proton–proton collisions.
- PDFs are experimentally obtained by fitting parametric models to experimental data.
- High-order predictions in perturbative QCD are computationally very expensive.
- Scanning different PDFs while performing a fit becomes a **computational bottleneck**.

SOLUTION: **To decouple the PDF dependence from the calculation of the partonic cross sections**

$$\sigma = \sum_{a,b} \int dx_1\, dx_2\; f_a(x_1, \mu_F)\, f_b(x_2, \mu_F)\, d\hat{\sigma}_{ab}(x_1, x_2, \mu_R, \mu_F)$$

The PDFs, $f(x, \mu_F)$, are extended in a basis of interpolation kernels $E_i(x)$, and the resulting convolution with the partonic cross sections are pre-computed on a discrete grid.

Interpolation grids

$$\sigma = \sum_{a,b} \int dx_1\, dx_2\; f_a(x_1, \mu)\, f_b(x_2, \mu)\, d\hat{\sigma}_{ab}(x_1, x_2, \mu)$$

Interpolation grids

$$\sigma = \sum_{a,b} \int dx_1\, dx_2\; f_a(x_1, \mu)\, f_b(x_2, \mu)\, d\hat{\sigma}_{ab}(x_1, x_2, \mu)$$

The interpolation kernels satisfy the following properties:

$$E_i\big(x^{[j]}\big) = \delta_i^j \qquad\qquad \sum_i E_i(x) = 1$$

Interpolation grids

$$\sigma = \sum_{a,b} \int dx_1\, dx_2\; f_a(x_1, \mu)\, f_b(x_2, \mu)\, d\hat{\sigma}_{ab}(x_1, x_2, \mu)$$

The interpolation kernels satisfy the following properties:

$$E_i\big(x^{[j]}\big) = \delta_i^j \qquad\qquad \sum_i E_i(x) = 1$$

Thus, a continuous function $f(x, \mu)$ can be approximated as

$$f(x, \mu) \;\approx\; \sum_{i,k} f^{[i,k]}\, E_i(x)\, E_k(\mu), \qquad \text{with} \;\; f^{[i,k]} \equiv f\big(x^{[i]},\, \mu^{[k]}\big)$$

Interpolation grids

$$\sigma = \sum_{a,b} \int dx_1\, dx_2\; f_a(x_1, \mu)\, f_b(x_2, \mu)\, d\hat{\sigma}_{ab}(x_1, x_2, \mu)$$

The interpolation kernels satisfy the following properties:

$$E_i\big(x^{[j]}\big) = \delta_i^j \qquad\qquad \sum_i E_i(x) = 1$$

Thus, a continuous function $f(x, \mu)$ can be approximated as

$$f(x, \mu) \;\approx\; \sum_{i,k} f^{[i,k]}\, E_i(x)\, E_k(\mu), \qquad \text{with} \;\; f^{[i,k]} \equiv f\big(x^{[i]},\, \mu^{[k]}\big)$$

$\Rightarrow$ **The PDF–dependent term, $f^{[i,k]}$, is taken outside the integral. Its contribution can then be reconstructed efficiently by a fast sum over the grid nodes, weighted by $f^{[i,k]}$.**

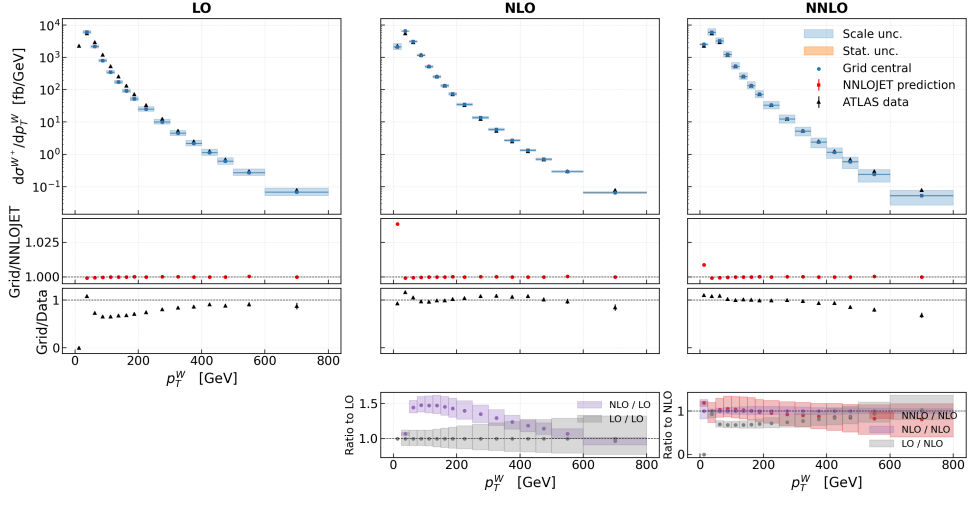
Technical details

GOAL: Produce interpolation grids for W+jets processes at NNLO accuracy in QCD and use them for PDF fitting

- **NNLOJET parton level event generator**
  - *Antenna subtraction scheme* for infrared divergences cancellation.
  - $\sim 300{,}000$ CPU hours to achieve an accuracy of $\sim 1\%$.
- **FastNLO libraries for grid production**
  - Fast re-evaluation of the grids for different PDF sets.
  - Fast re-evaluation for scale variations.

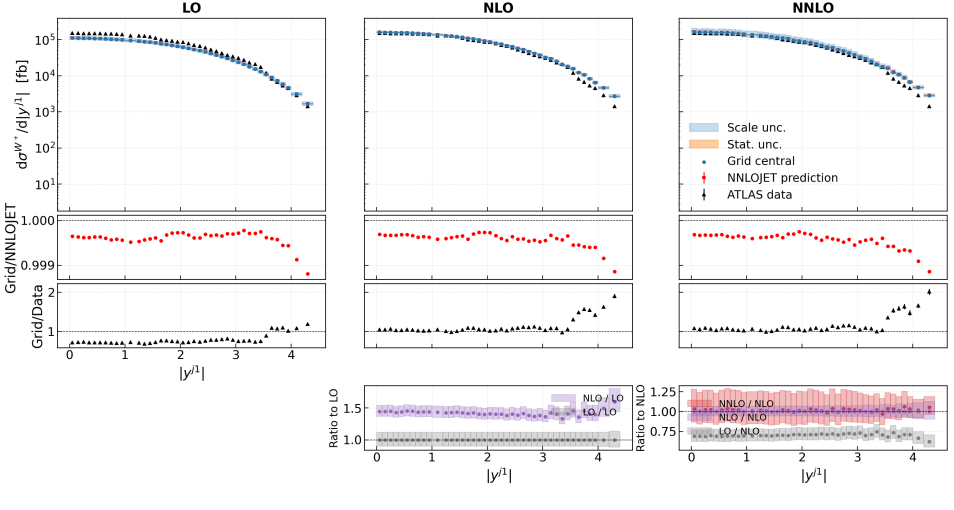
  Kernels: third-degree Lagrange polynomials.
  Number of nodes: 25 x nodes. 10 scale nodes ($\mu \equiv \mu_R = \mu_F$) for dynamic scales.
- ***xFitter* framework for QCD analysis**
  - PDF evolution via the DGLAP equations.
  - PDF fits by χ^2 minimisation with MINUIT.

Results at fixed scales $\mu = m_W$



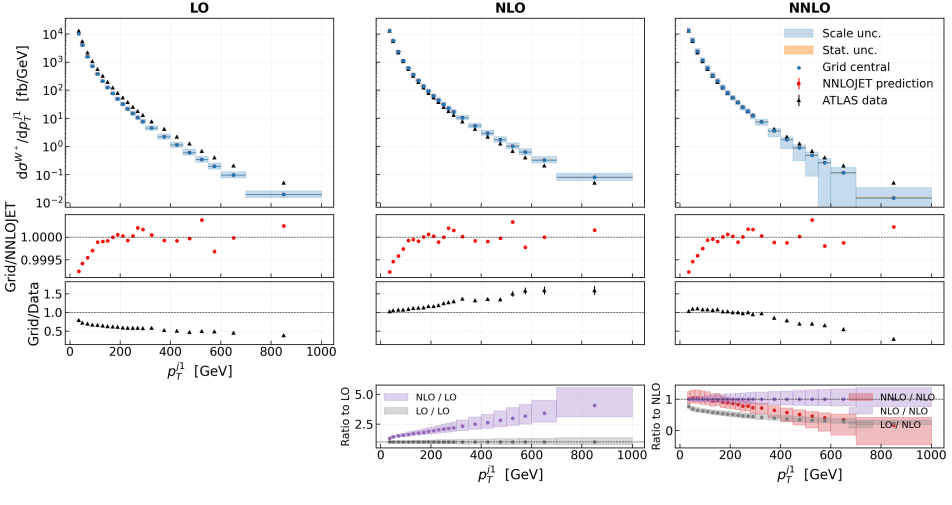
No LO contribution in the first bin due to kinematic cuts. Good agreement elsewhere.

Results at fixed scales $\mu = m_W$



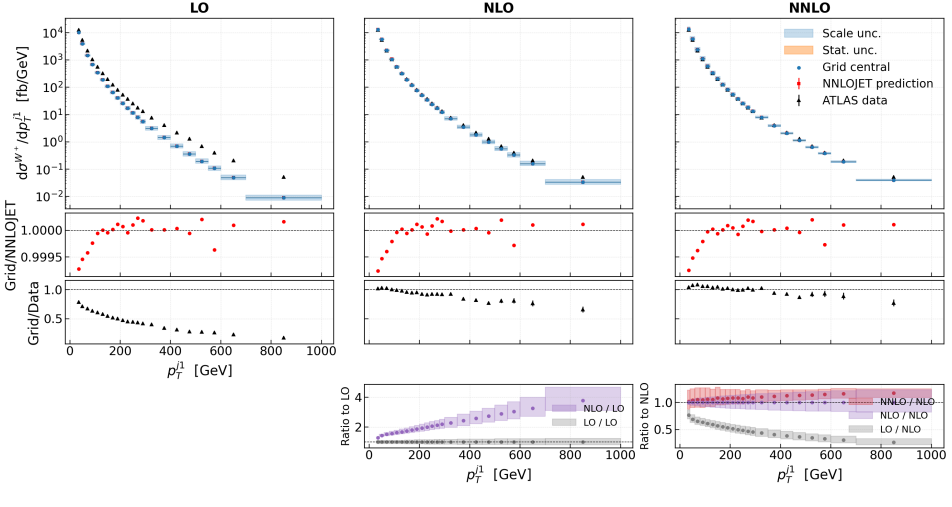
Small systematic shift in the closure test. Tension with data at high rapidity.

Results at fixed scales $\mu = m_W$



Successful closure test. NNLO contribution turns negative at high p_T for some scale variations.

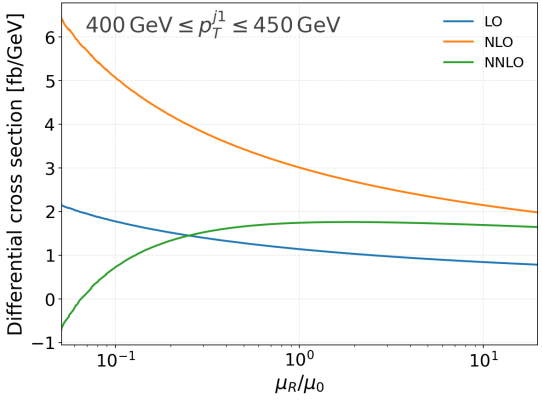
Results at $\mu = \sqrt{m_{l\nu}^2 + \sum_i p_T^{j_i 2}}$



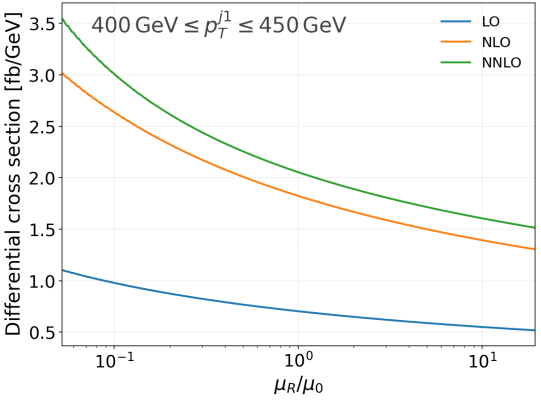
The closure test is still successful and the scale dependence is more stable.

Study of the scale dependence

- $\mu_0 = m_W$



- $\mu_0 = \sqrt{m_{l\nu}^2 + \sum_i p_T^{j_i 2}}$



The dynamic scale improves convergence of the perturbative expansion in the p_T observables.

PDF fits setup

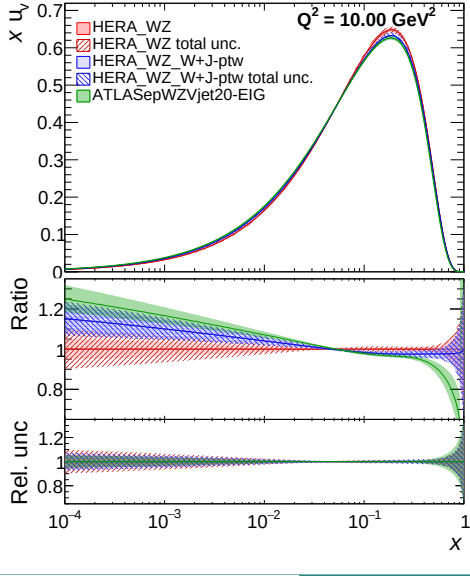
- **Included data**
  - Inclusive DIS HERA data[2]
  - Inclusive W and Z production (ATLAS)[3]
  - W+jets differential cross sections in p_T^W and $|y^{j1}|$ (ATLAS)[4]
- **Parameterisation of the PDFs**
  - *ATLASepWZVjet20* parameterisation[5]

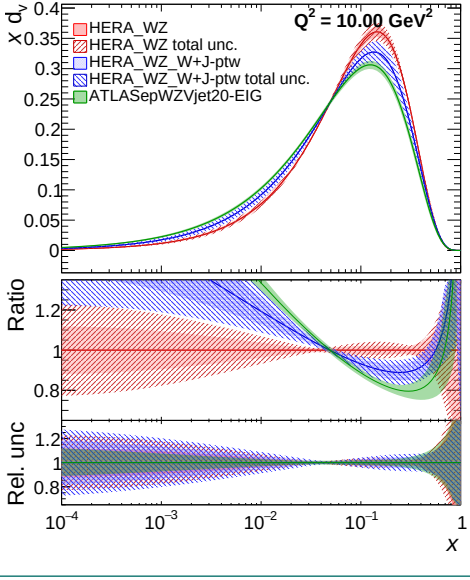
$$xq_i(x) = A_i x^{B_i}(1-x)^{C_i} P_i(x), \qquad P_i(x) = (1 + D_i x + E_i x^2) e^{F_i x}$$

$$xg(x) = A_g x^{B_g}(1-x)^{C_g} P_g(x) - A_g^{'} x^{B_g^{'}}(1-x)^{C_g^{'}}$$

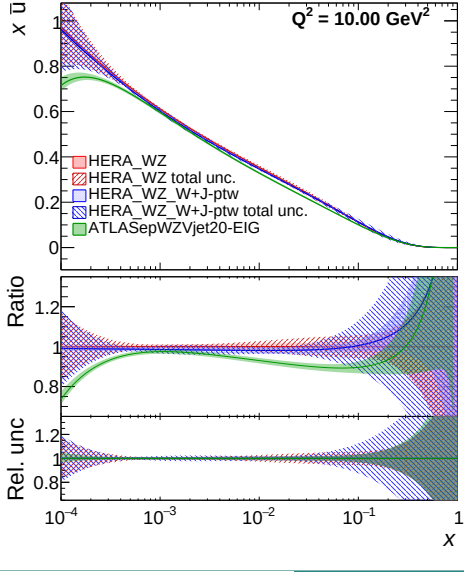
[2] H. Abramowicz et al. In: *Eur. Phys. J. C* 75.12 (2015), p. 580. DOI: *10.1140/epjc/s10052-015-3710-4*. arXiv: *1506.06042 [hep-ex]*
[3] Morad Aaboud et al. In: *Eur. Phys. J. C* 77.6 (2017), p. 367. DOI: *10.1140/epjc/s10052-017-4911-9*. arXiv: *1612.03016 [hep-ex]*
[4] Morad Aaboud et al. In: *JHEP* 05 (2018), p. 077. DOI: *10.1007/JHEP05(2018)077*. arXiv: *1711.03296 [hep-ex]*
[5] Georges Aad et al. In: *JHEP* 07 (2021), p. 223. DOI: *10.1007/JHEP07(2021)223*. arXiv: *2101.05095 [hep-ex]*

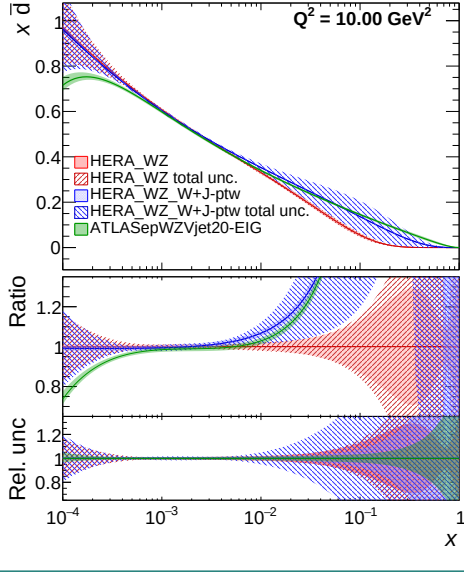
PDF fits: Including p_T^W spectrum



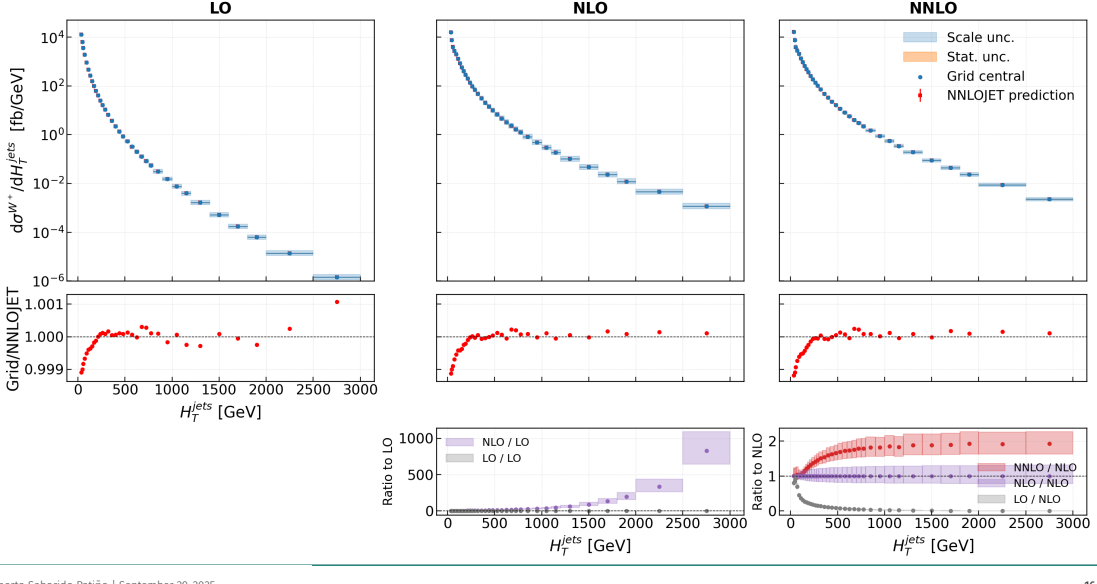


PDF fits: Including p_T^W spectrum





Predictions for the measurement at $\sqrt{s} = 13\,\mathrm{TeV}$



Summary

- Reliable NNLO QCD predictions for $W+$jets have been produced in grid format, making them usable for PDF fitting and representing the state of the art in accuracy.
- The behaviour of the predictions under scale variations has been systematically studied.
- PDF fits have been performed using ATLAS data at $\sqrt{s} = 8\ \mathrm{TeV}$. The impact of $W+$jets data in the fits is particularly relevant in the $x\bar{d}$ and $x\bar{s}$ distributions at high x.
- Grids for the $\sqrt{s} = 13\ \mathrm{TeV}$ scenario have been produced, enabling the inclusion of the corresponding measurements in future ATLAS PDF fits.

Backup

Interpolation grids

The interpolation kernels satisfy:

$$E_i\big(x^{[j]}\big) = \delta_i^j, \qquad \sum_{i=0}^{N} E_i(x) = 1, \qquad a \leq x \leq b.$$

Thus, a continuous function $f(x,\mu)$ can be approximated as

$$f(x,\mu) \;\approx\; \sum_{i,k} f^{[i,k]}\, E_i(x)\, E_k(\mu), \qquad \text{with} \;\; f^{[i,k]} \equiv f\big(x^{[i]},\, \mu^{[k]}\big)$$

and the following approximation holds:

$$\alpha_s(\mu)\, f_a(x_1,\mu)\, f_b(x_2,\mu) \approx \sum_{i,j,k} \alpha_s^{[k]}\, f_a^{[i,k]}\, f_b^{[j,k]}\, E_i(x_1)\, E_j(x_2)\, E_k(\mu)$$

An integral is then approximated by:

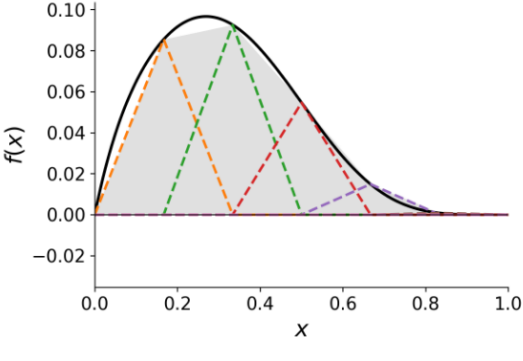
$$\int_a^b dx\, f(x,\mu)\, g(x,\mu) \;\approx\; \sum_{i,k} f^{[i,k]}\, g^{[i,k]}$$

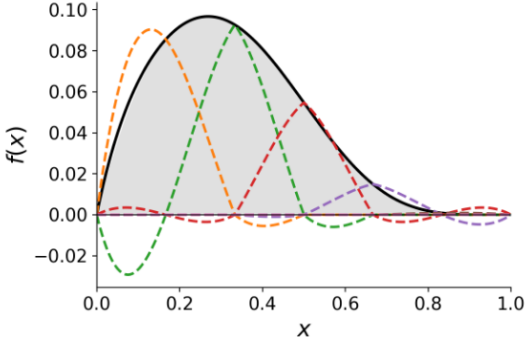
where

$$g^{[i,k]} \equiv \int_a^b dx\, E_i(x)\, E_k(\mu)\, g(x,\mu)$$

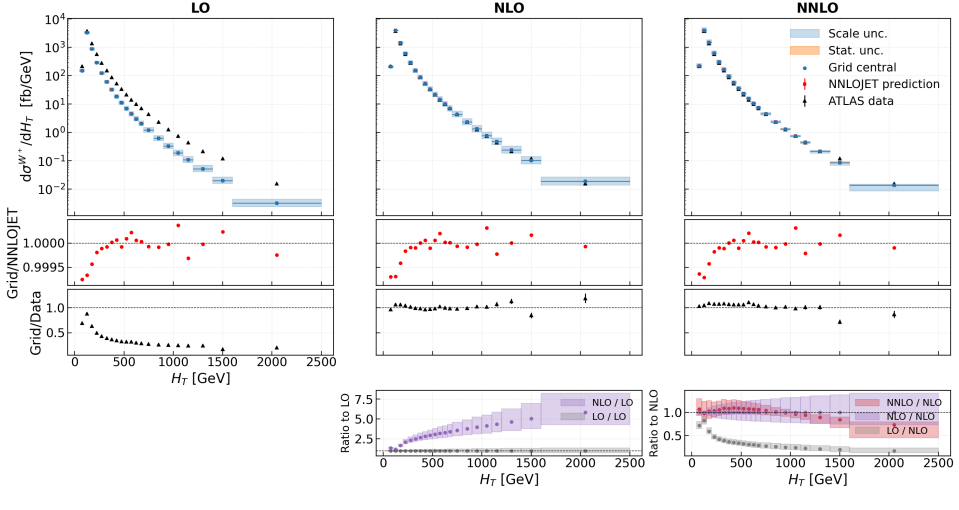
- This technique is applied with g being the partonic cross section.
- $g^{[i,k]}$ are the precomputed grid weights.

Interpolation examples



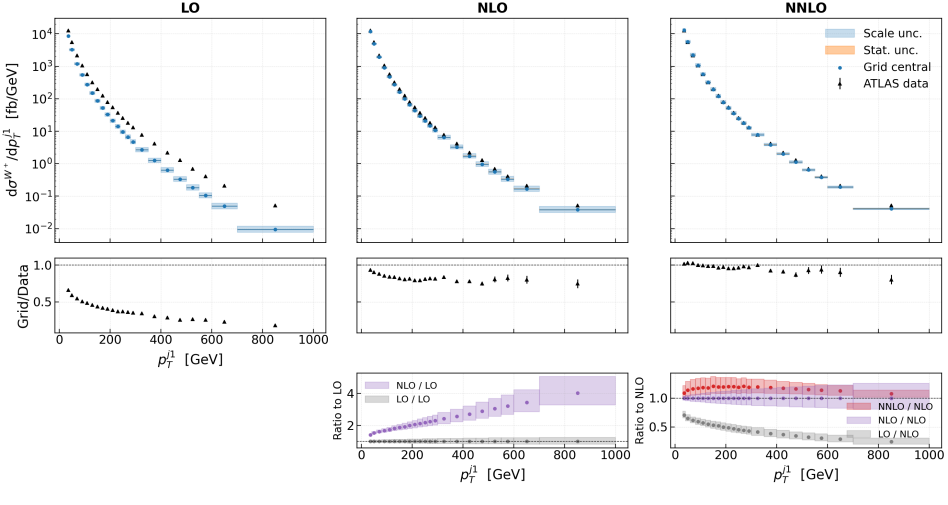


Results at fixed scales $\mu = m_W$



Successful closure test and good agreement with data for H_T.

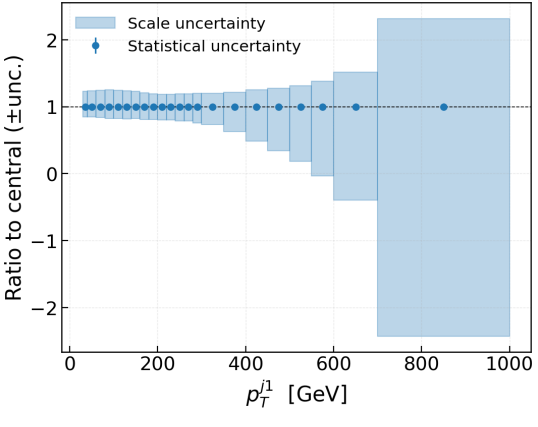
Test at $\mu = 8m_W$



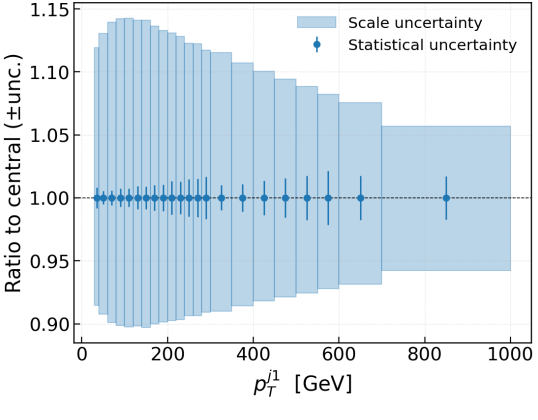
More reliable predictions and better agreement with data at high p_T.

Relative uncertainties for p_T^{j1}

- $\mu = m_W$



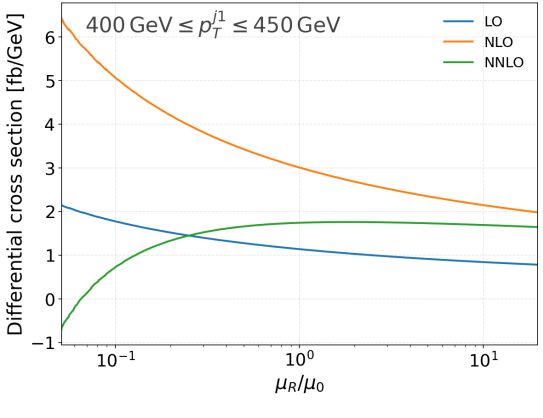
- $\mu = 8m_W$



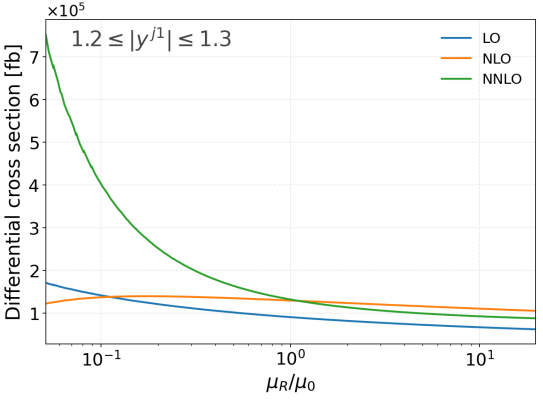
Scale uncertainties are significantly reduced.

Analysis of the scale dependence when $\mu_0 = m_W$

- p_T^{j1}

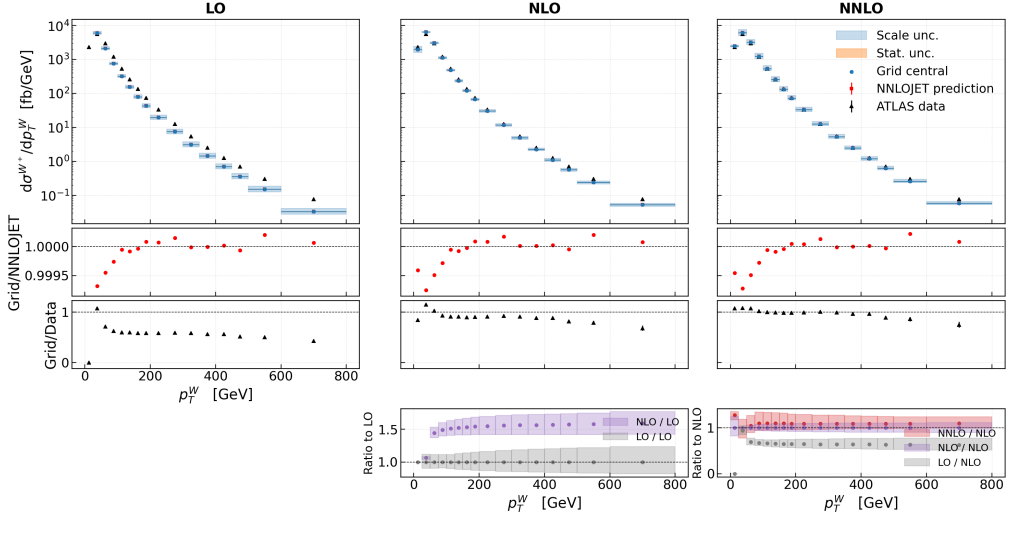


- $|y^{j1}|$

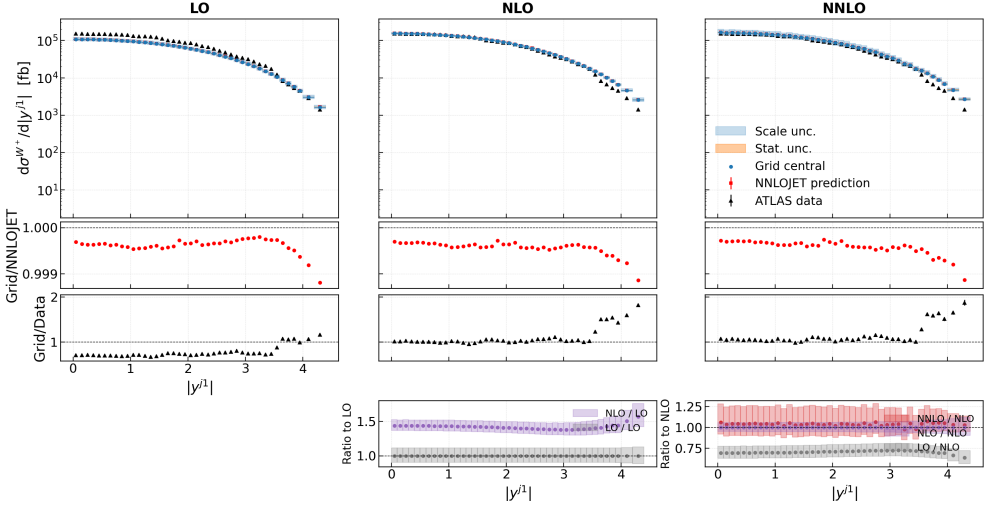


The factorisation scale is fixed.

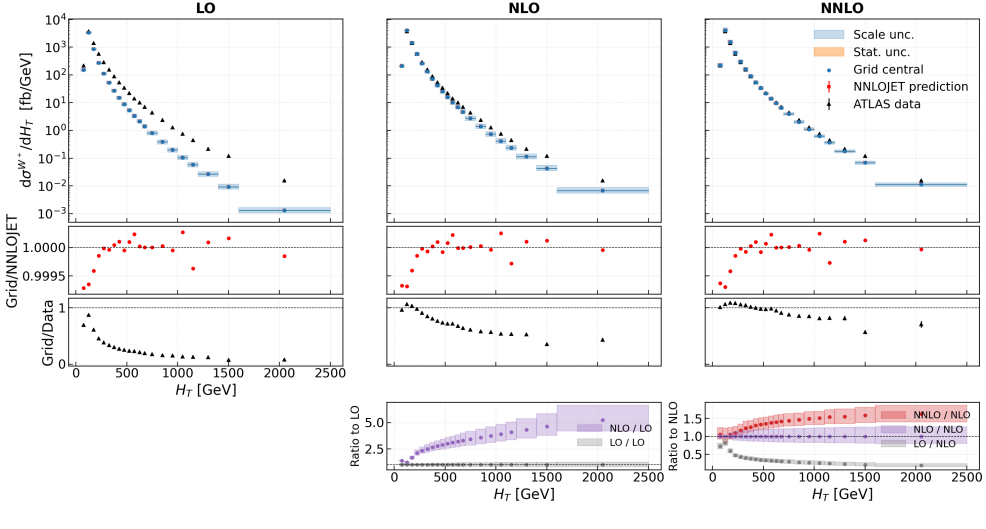
Results at $\mu = \sqrt{m_{l\nu}^2 + \sum_i p_T^{j_i 2}}$



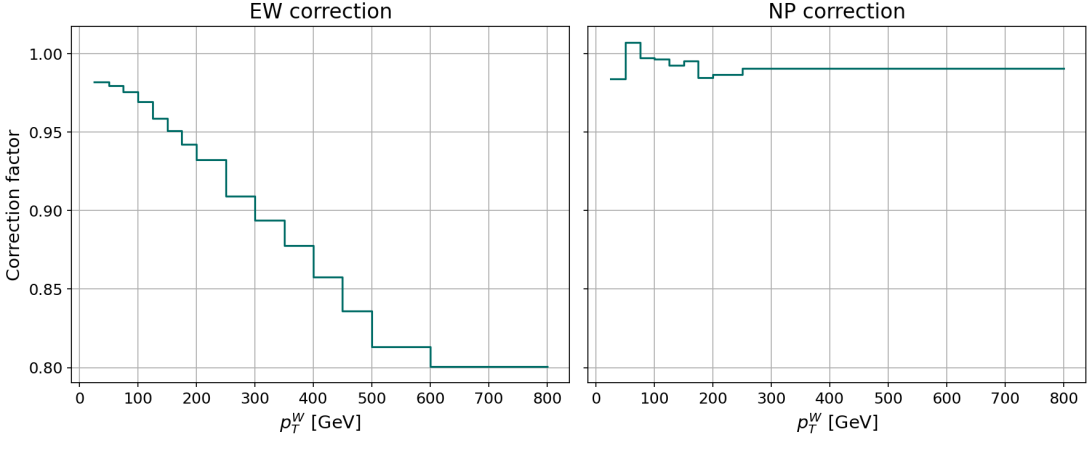
Results at $\mu = \sqrt{m_{l\nu}^2 + \sum_i p_T^{j_i 2}}$



Results at $\mu = \sqrt{m_{l\nu}^2 + \sum_i p_T^{j_i 2}}$



NP and EW corrections



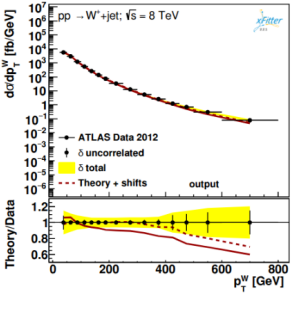
χ^2 definition

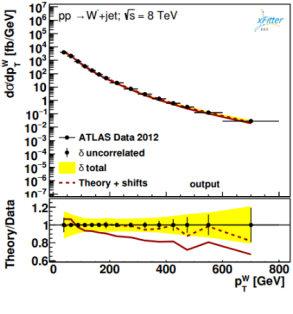
$$\chi^2 = \sum_i \frac{\left[D_i - T_i\left(1 - \sum_j \gamma_{ij} b_j\right)\right]^2}{\delta^2_{i,\text{uncor}} T_i^2 + \delta^2_{i,\text{stat}} D_i T_i} + \sum_j b_j^2 + \sum_i \log\left(\frac{\delta^2_{i,\text{uncor}} T_i^2 + \delta^2_{i,\text{stat}} D_i T_i}{\delta^2_{i,\text{uncor}} D_i^2 + \delta^2_{i,\text{stat}} D_i^2}\right)$$

- D_i and T_i: measured and predicted values in bin i
- $\delta_{i,\text{stat}}$, $\delta_{i,\text{uncor}}$: relative statistical and uncorrelated systematic uncertainties
- b_j: nuisance parameters introduced to handle correlated systematics
- γ_{ij}: relative effect of systematic source j on data point i
- Nuisance parameters are penalized via the regularisation term $\sum_j b_j^2$
- The logarithmic penalty term arises from transitioning the likelihood to χ^2 and avoids biases

Fit accuracy, p_T^W spectrum

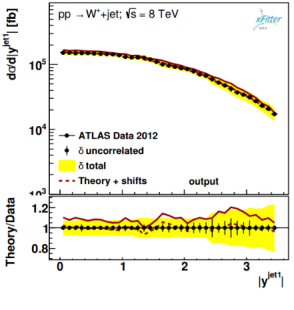
Dataset	output
HERA1+2 NCep 820	52 / 59
HERA1+2 NCep 460	200 / 177
HERA1+2 CCep	44 / 39
ATLAS high mass CF Z rapidity 2011	4.3 / 6
ATLAS high mass CC Z rapidity 2011	6.4 / 6
ATLAS W+ lepton rapidity 2011	12 / 11
HERA1+2 NCem	225 / 159
HERA1+2 CCem	62 / 42
ATLAS 8 TeV W^-+jets W PT	8.9 / 15
HERA1+2 NCep 575	189 / 221
ATLAS W- lepton rapidity 2011	12 / 11
ATLAS 8 TeV W^++jets W PT	13 / 15
HERA1+2 NCep 920	345 / 310
ATLAS peak CF Z rapidity 2011	6.2 / 9
ATLAS peak CC Z rapidity 2011	15 / 12
Correlated χ^2	117
Log penalty χ^2	-7.68
Total χ^2 / dof	1306 / 1078

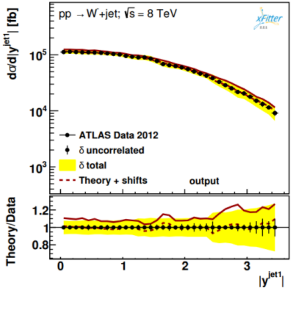




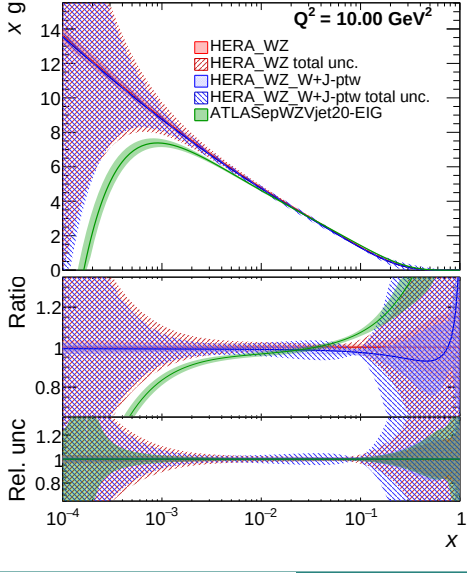
Fit accuracy, $|y^{j1}|$ spectrum

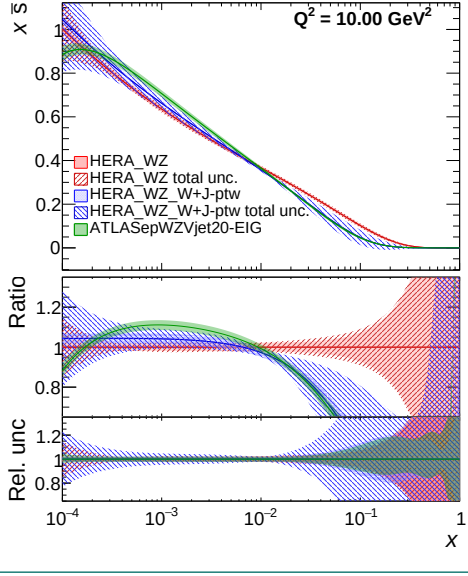
Dataset	output
HERA1+2 NCep 820	53 / 59
HERA1+2 NCep 460	199 / 177
HERA1+2 CCep	44 / 39
ATLAS high mass CF Z rapidity 2011	4.1 / 6
ATLAS high mass CC Z rapidity 2011	6.7 / 6
ATLAS W+ lepton rapidity 2011	13 / 11
ATLAS 8 TeV W^+ +jets jet 1 rap	21 / 35
HERA1+2 NCem	221 / 159
HERA1+2 CCem	59 / 42
HERA1+2 NCep 575	189 / 221
ATLAS 8 TeV W^- +jets jet 1 rap	26 / 35
ATLAS W- lepton rapidity 2011	11 / 11
HERA1+2 NCep 920	344 / 310
ATLAS peak CF Z rapidity 2011	6.8 / 9
ATLAS peak CC Z rapidity 2011	13 / 12
Correlated χ^2	102
Log penalty χ^2	+11
Total χ^2 / dof	1322 / 1118



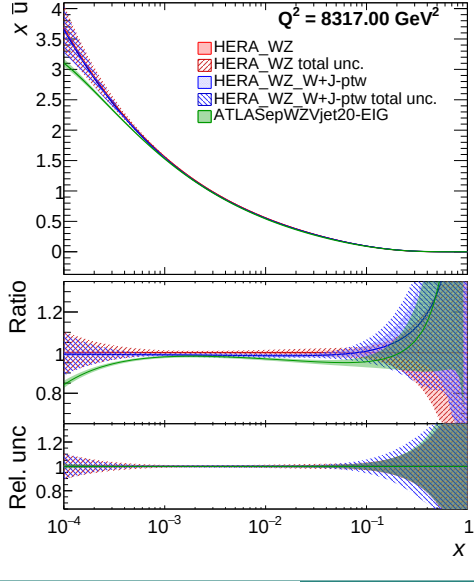


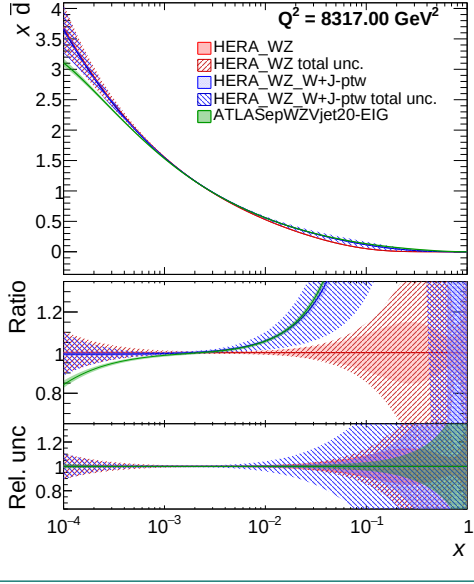
More PDF fits



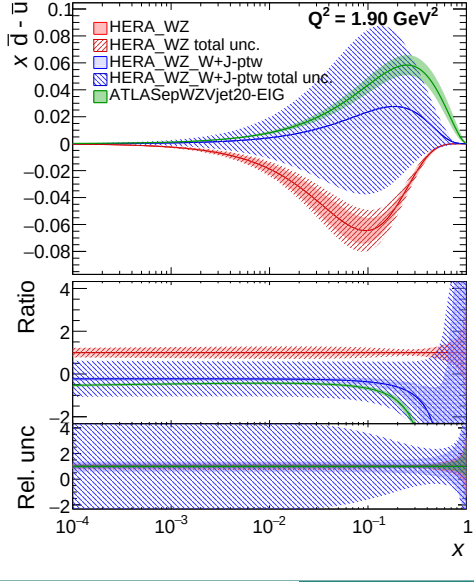


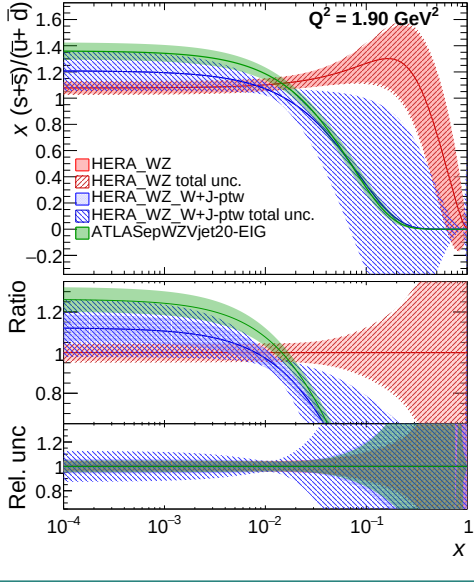
More PDF fits





More PDF fits





More PDF fits: Including $|y^{j1}|$ spectrum