Contents

List of Figures

List of Tables

1 W boson pT spectrum

"Potentielle citation sans aucun rapport avec le sujet"

— Personne inconnue, *contexte à déterminer*

1.1 Unfolding

The measured W p_T spectrum is subject to various detector effects (finite resolution and acceptance, reconstruction efficiency, etc.) that distort the true underlying spectrum. Mathematically, the unfolding problem is an integral equation of the following form:

$$\int K(x,y) \cdot f(x) dx = g(y), \tag{1.1}$$

where we seek the function $f(x)$ assuming that $g(y)$ and the kernel $K(x,y)$ are known [**Schmitt:2016orm**]. The function $g(y)$ is convoluted (or folded) with the kernel hence the name of the problem. In experimental physics it is more common to use binned distributions instead of continuous functions:

$$\sum_i \mathbf{R}_{ij} \cdot \mathbf{T}_i + \mathbf{B}_j = \mathbf{D}_j, \tag{1.2}$$

with **D**, **T** and **B** being vectors that represent detector-level (measured), truth and background distributions respectively; each vector has n components that represent the bins in the corresponding distribution. The response matrix **R** represents bin-to-bin migrations caused by the detector effects. The response matrix is usually obtained through Monte-Carlo (MC) simulation, along with the corrections for the fiducial volume acceptance and efficiency. Each event is simulated on the truth and reconstructed levels, this means that element R_{il} of the migration matrix contains events that pass both reconstruction and truth cuts (R&T) and would go to bin i of the truth distribution and to bin j of the reconstructed distribution. Detector acceptance for bin i is defined as the ratio $A_i = \sum_j Rij/T_i$. Similarly the reconstruction efficiency for bin j is defined as the ratio $\epsilon_j = \sum_i Rij/D_j$. The underlying distribution estimate is presented in the following way:

$$\sum_i \mathbf{V}_{ij} \cdot (\mathbf{D}_i - \mathbf{B}_i) = \mathbf{U}_j, \tag{1.3}$$

where U vector provides the underlying distribution estimate and V is the unfolding transformation matrix.

There exist a diverse variety of methods to obtain the unfolding transformation. In the current analysis a Bayesian iterative method is used [**unfolding1**], [**unfolding2**]. The method allows to obtain the unfolding transformation provided that the response matrix, acceptance and efficiency corrections are known and the number of iterations is given. The number of iterations as well as the unfolded distribution binning are adjusted in order to minimize the unfolding bias and keep the uncertainty below the designated level.

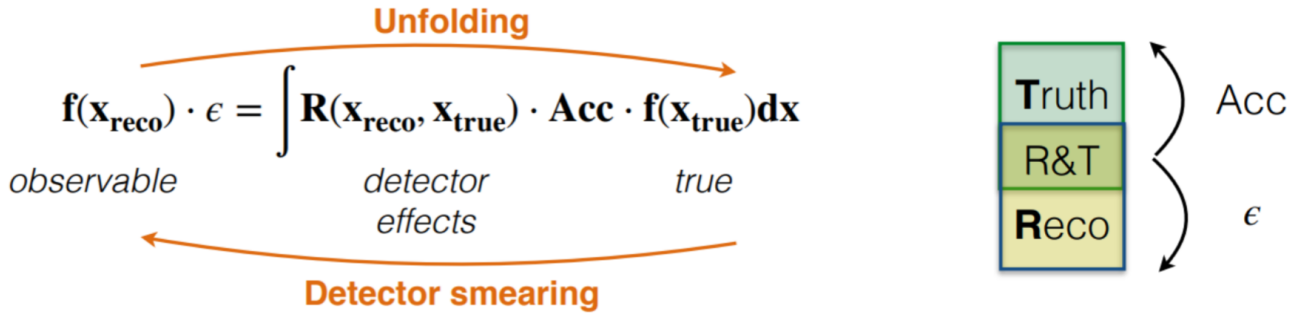

Figure 11: Schematic description of the unfolding procedure.

1.2 Uncertainties propagation

The detector-level uncertainties breakdown for the p_{T} distribution are presented here. Uncertainties breakdown for the rest of the observables are listed in Appendix A.
These uncertainties now have to be propagated to the unfolded level.

1.2.1 Statistical uncertainty propagation using Bootstrap method

Bootstrap is a computer-based method of dataset parameters estimate and propagation using the analysis distribution resampling. In particular bootstraping used for the propagation of statistical uncertainties.
Both data and MC-simulated datasets have limited number of events, hence the statistical uncertainties due to fluctuations. In order to estimate the statistical uncertainty a number of pseudo-data sets is generated for both data and MC where each event is assigned a random weight w:

$$w = \mathcal{P}(n, 1), \tag{1.4}$$

where n is a random number generated with Poisson distribution with mean $\lambda = 1$, value $\mathcal{P}(n, 1)$ is a Poissonian probability of observing n events while expecting an average of 1 event.
The boostrapping defined in this way allows to take into account the correlated effect of statistical fluctuations across all observables and distributions in the analysis. For the determination of statistical uncertainty of the unfolded spectrum 400 bootstrap samples were generated. In both data and MC

cases the statistical uncertainty is estimated by composing the covariant matrix C_{kl}^{stat}:

$$C_{kl}^{stat} = \frac{1}{N_{bs}-1}\sum_{\alpha=1}^{N_{bs}}(U_k^\alpha - \langle U_k\rangle)(U_l^\alpha - \langle U_l\rangle), \tag{1.5}$$

where N_{bs} is the number of the Bootstrap toys used, vector U stands for the varied underlying distribution, $\langle U_k\rangle$ is the average underlying distribution. However, the variation is performed in a different way for Data and MC:

$$U_j^{\alpha,(MC)} = V_{ij}^\alpha \sum_i (D_i - B_i),$$

$$U_j^{\alpha,(Data)} = V_{ij} \sum_i (D_i^\alpha - B_i).$$

In the MC case it is the response matrix V^α to be varied (α index corresponds to the variation number), whereas in Data the toys are obtained by varying the measured distribution D_i^α. The statistical uncertainty for both cases is defined as:

$$\delta U_k = \sqrt{C_{kk}^{stat}}. \tag{1.6}$$

1.2.2 Systematic uncertainty propagation

Systematic uncertainties are broken down into a number of uncorrelated uncertainty sources, which include signal and background modelling uncertainties, calibration and efficiency uncertainties, physics modelling uncertainties. The systematic variations used for uncertainty estimate on the detector level are propagated to the level of underlying distribution in two different ways. For the background uncertainties:

$$U_j^a = V_{ij} \sum_i (D_i - B_i^a),$$

total background estimate B_i^a is varied in luminosity and cross-section of every back-ground (index a numbers the sources of uncertainty). For other sources of systematic uncertainty:

$$U_j^a = V_{ij}^a \sum_i (D_i - B_i),$$

response matrix variation is created. The corresponding covariance matrix is defined as:

$$C_{kj}^a = \delta U_k^a \delta U_l^a,$$

where the deltas are $\delta U_k^a = U_k^a - U_k^{Nom}$. The total covariance matrix is calculated as a sum:

$$C_{kl}^{tot} = C_{kl}^{stat,Data} + C_{kl}^{stat,MC} + \sum_a C_{kl}^a. \tag{1.7}$$

1.2.3 Unfolded uncertainty breakdown

Figures 12, 14, contain the systematic uncertainties breakdown for electron and muon channels for the reconstructed level distributions for 5 and 13 TeV. Similarly figures 13, 15 contain unfolded-level uncertainties.

At the detector level the designated level of uncertainty of below 1% is preserved up to 25 GeV for 5 TeV datasets and up to 50 GeV for 13 TeV samples in every channel. An increased role of background uncertainty is observed at 13 TeV due to the significantly higher cross-sections of diboson and top-antitop backgrounds. The scale and hierarchy of uncertainties are preserved at the unfolded level.

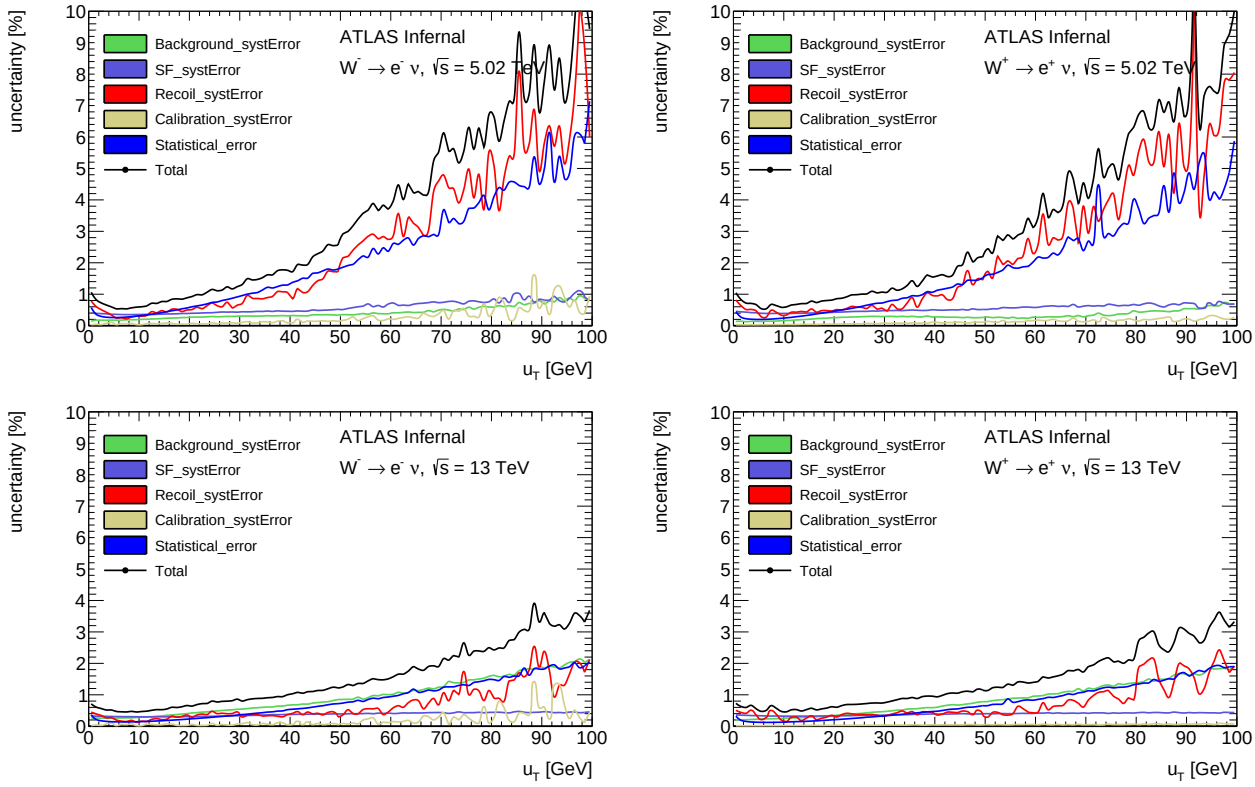

Figure 12: Breakdown of systematic uncertainties for 5 (a,b) and 13 (c,d) TeV in the electron channel at the reconstructed level

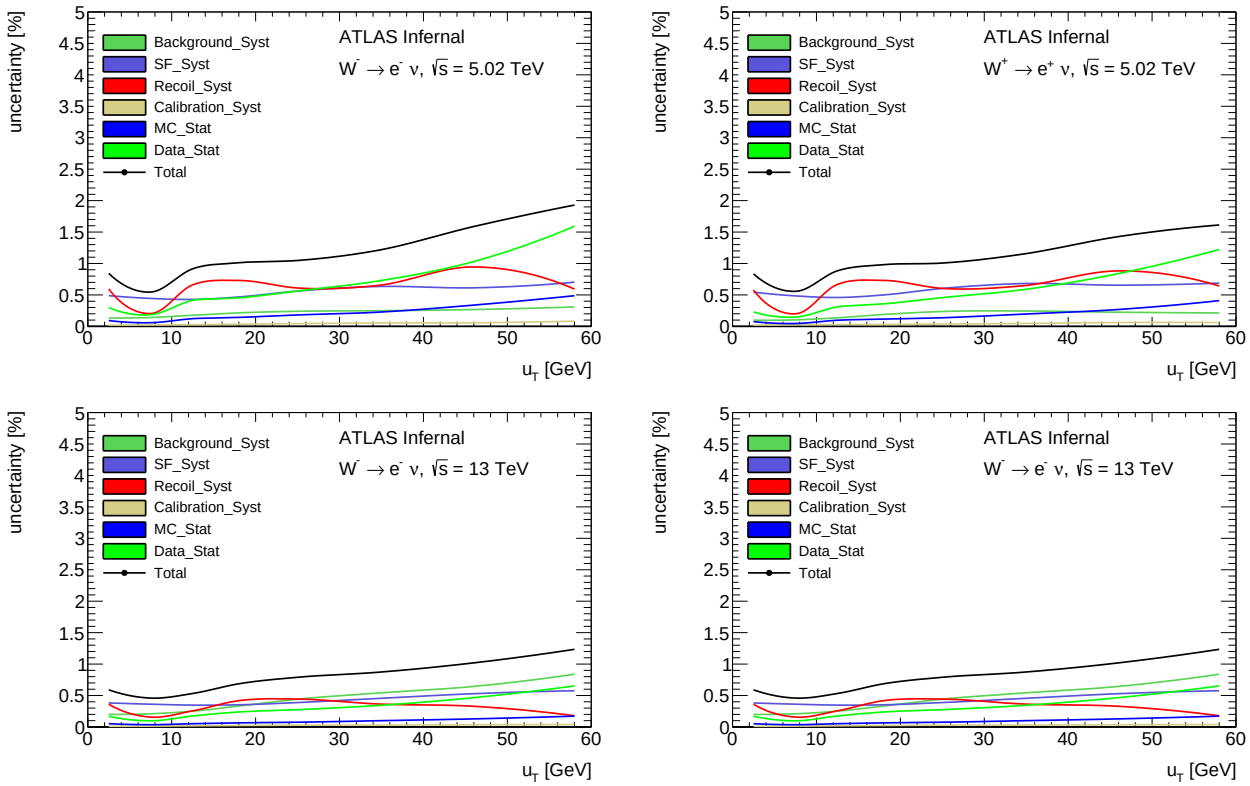

Figure 13: Breakdown of systematic uncertainties for 5 (a,b) and 13 TeV (c,d) in the electron channel at the unfolded level

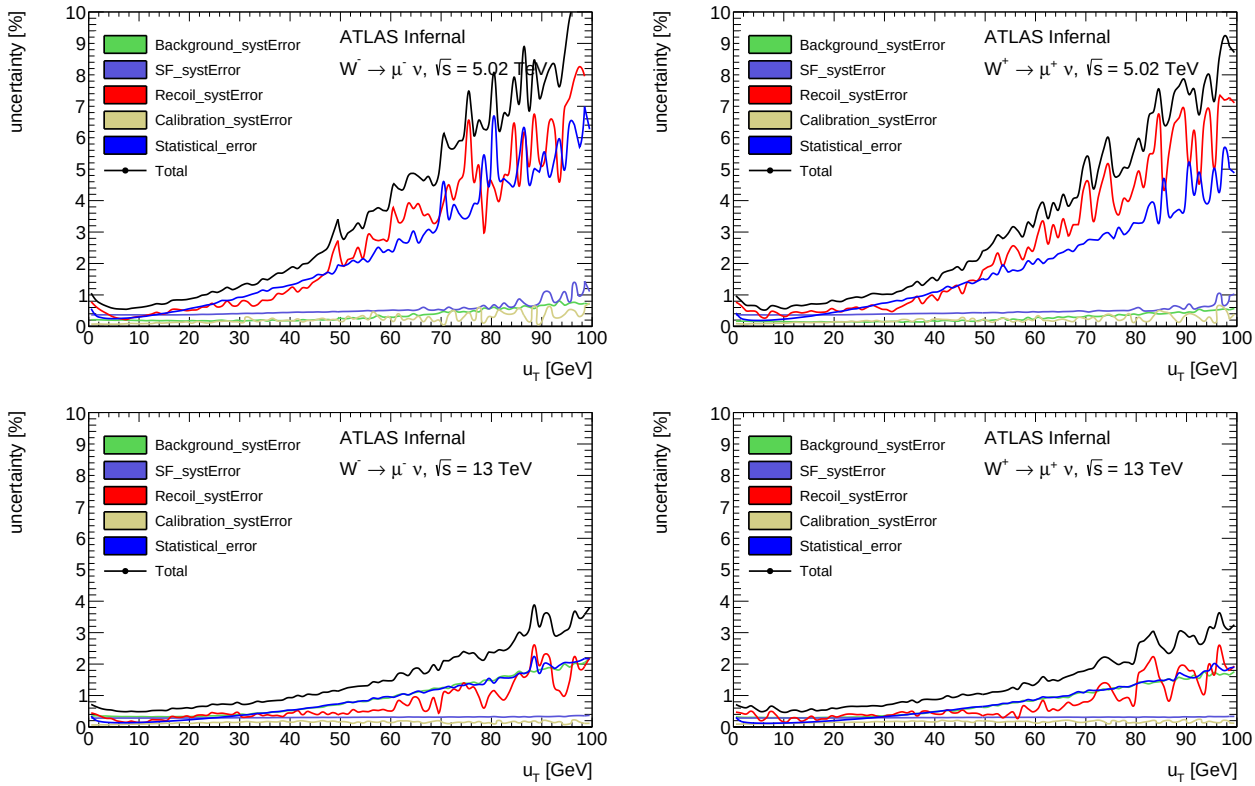

Figure 14: Breakdown of systematic uncertainties for 5 (a,b) and 13 (c,d) TeV in the muon channel at the reconstructed level

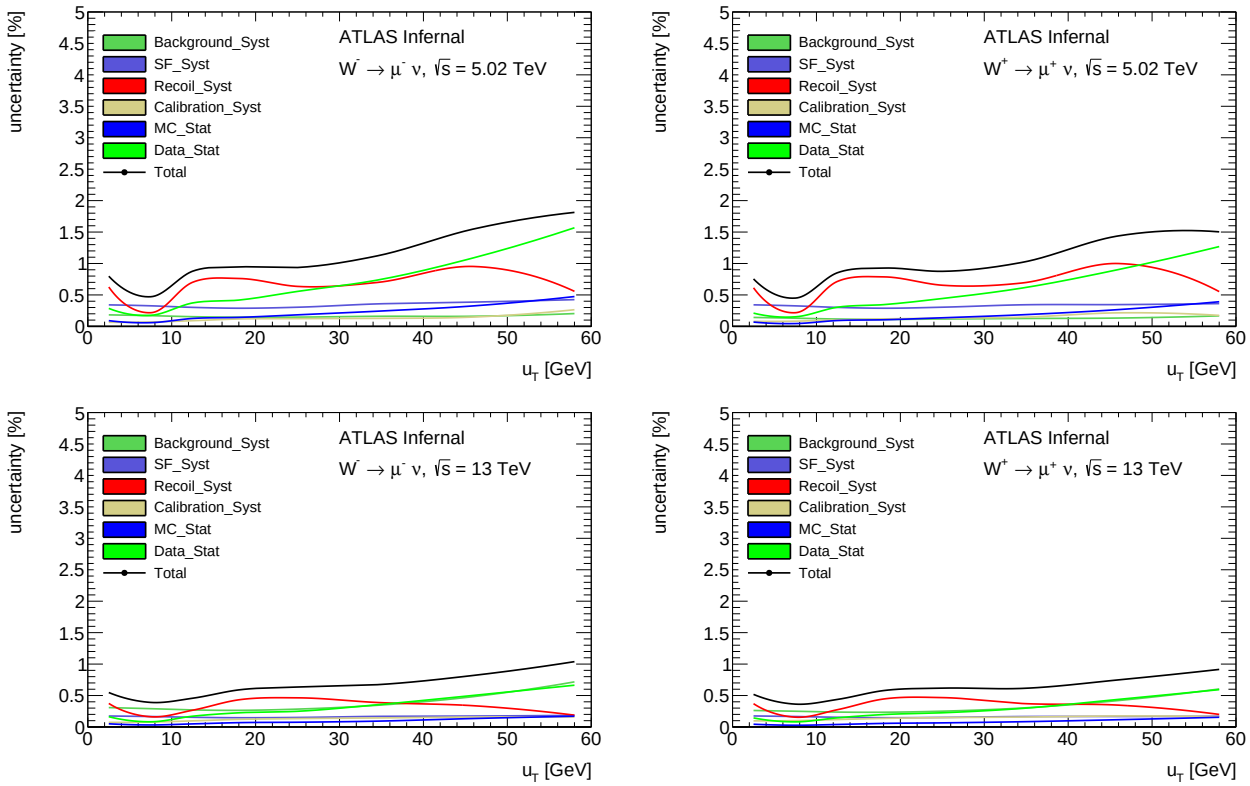

Figure 15: Breakdown of systematic uncertainties for 5 (a,b) and 13 TeV (c,d) in the muon channel at the unfolded level

1.3 Unfolding bias

One of the uncertainties associated with unfolding usage is called unfolding bias and may arise because the procedure relies on the MC simulation of the distribution, which is used as a prior hypothesis for the Bayesian algorithm. Possible discrepancies between the modelled and true distribution lead to erroneous bin-to-bin migrations and can lead to distortions of the spectrum.
In order to estimate the bias induced by the unfolding procedure it is necessary to quantify how much the unfolded result is impacted by the assumed MC distribution. A set of samples with a different distribution at the truth level though compatible at the detector level is generated.

The truth distribution is reweighted until a good agreement between the data and MC is reached at the reconstruction level. The agreement is estimated in the kinematic region of $u_T < 100$GeV using the χ^2 criterion. The truth reweighting procedure is applied to MC samples with a different distribution: Pythia8, Sherpa and DYRES were used. Fig. demonstrates the initial difference in the distributions.

The results are presented on fig 16 for 5 GeV bins and 3 unfolding iterations. The obtained bias is close to the precision goal of the measurement ($\sim$ 1%) for the 5 TeVdataset. The 13 TeV dataset shows a larger bias, which can be explained by a larger discrepancy between data and Monte-Carlo. Worse resolution in 13 TeV suggests a neccesity to try a broader binning comparing to 5 GeV.

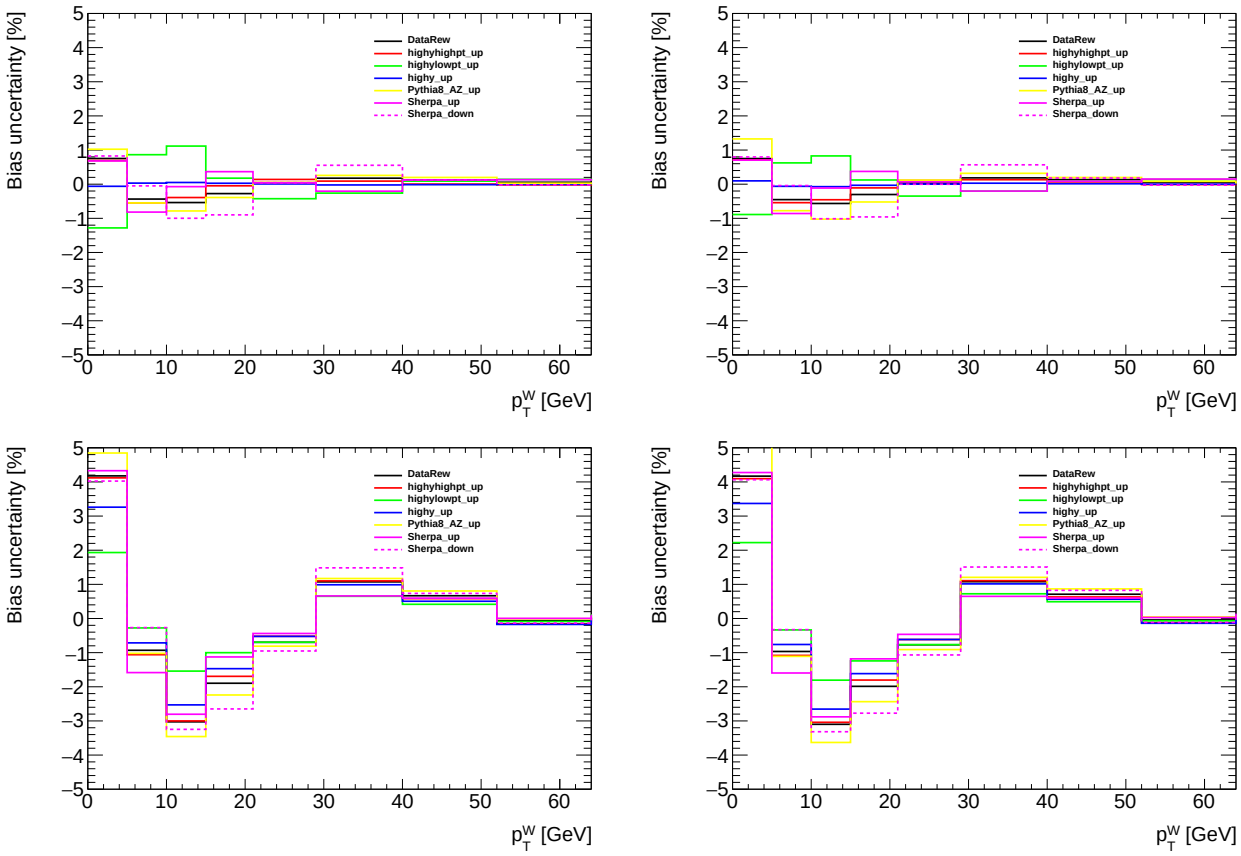

Figure 16: Unfolding bias on p_T^W in the electron channel after 3 iterations, for W^- (left) and W^+ (right), at 5 TeV (top) and 13 TeV (bottom).

1.4 Results

The comparison of unfolded spectrum to different theoretical predictions is presented at Figure 17 for electron channel and at 18 for the muon channel. The estimated experimental uncertainties raise from 1% at low p_{T}^{W} to about 5% (2%) at p_{T}^{W} =100 GeV, at 5 TeV (13 TeV).

The predicions are generated using Powheg AZNLO, Pythia AZ, Sherpa and DYRES. Powheg and Pythia agree with the data to a similar extent. A softer spectrum is predicted by Sherpa, while DYRES is on the opposite side compared to the data. The observed behaviour holds for both energies, both charges and both decay channels.

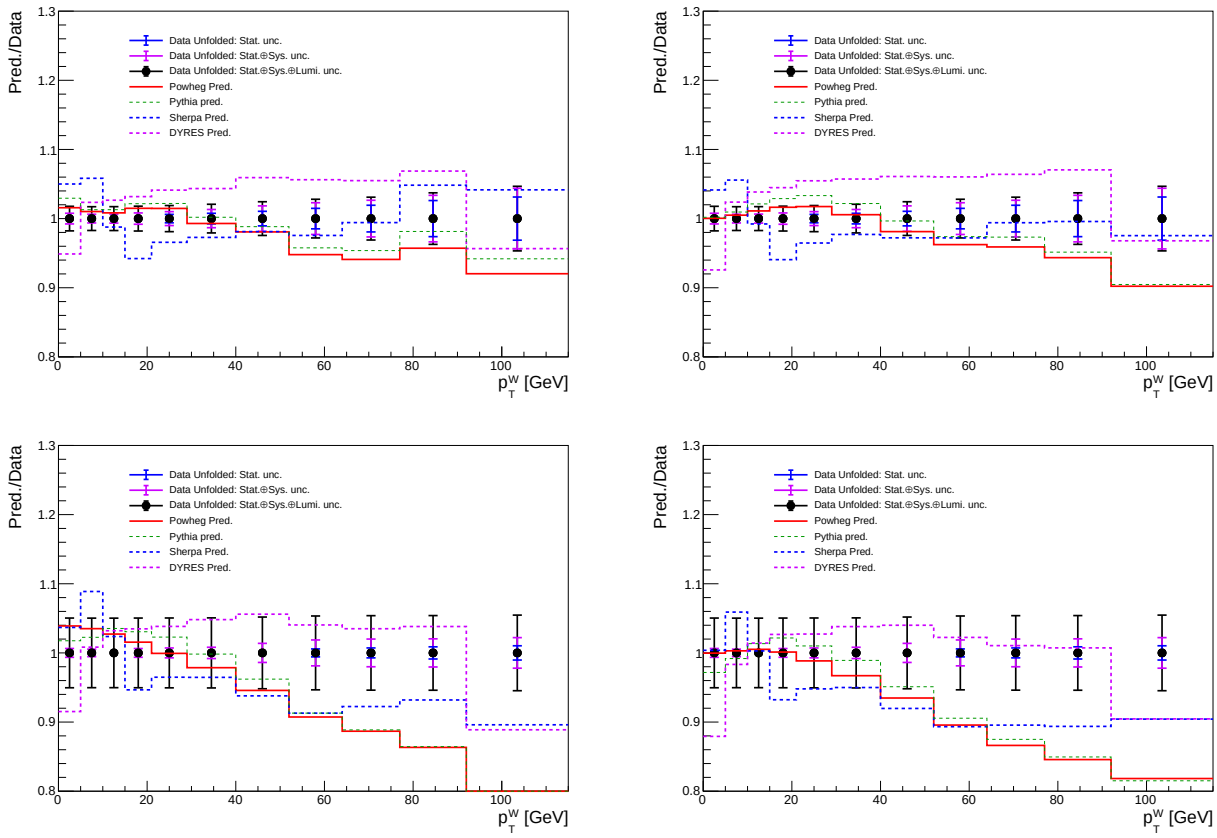

Figure 17: Unfolded measurement results in the W^- (left) and W^+ (right) electron channels, at 5 TeV (top) and 13 TeV (bottom).

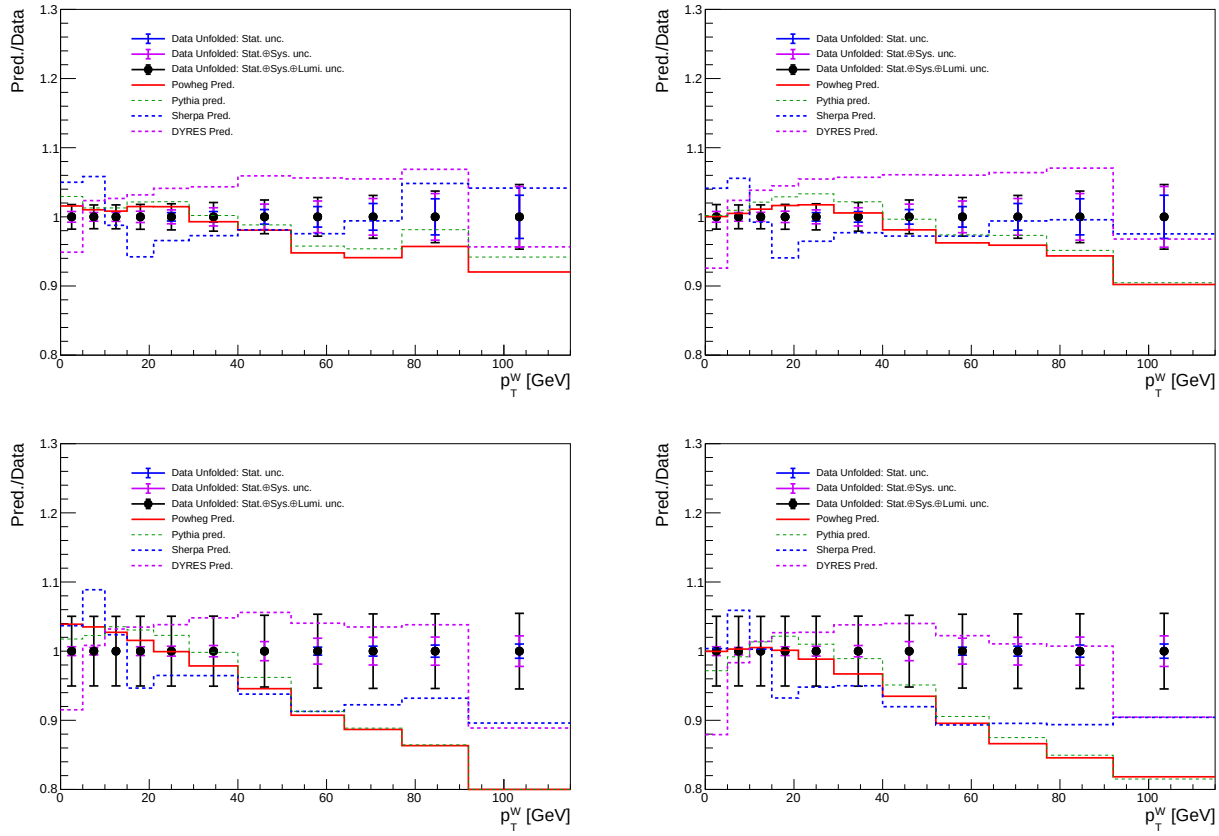

Figure 18: Unfolded measurement results in the W^- (left) and W^+ (right) muon channels, at 5 TeV (top) and 13 TeV (bottom).

Bibliography

[1] Jan Kretzschmar. *Samples and Physics modelling for low pile-up runs taken in 2017 and 2018*. Tech. rep. ATL-COM-PHYS-2019-075. Geneva: CERN, Feb. 2019. URL: https://cds.cern.ch/record/2657141.

[2] Paolo Nason. "A New method for combining NLO QCD with shower Monte Carlo algorithms". In: *JHEP* 11 (2004), p. 040. DOI: 10.1088/1126-6708/2004/11/040. arXiv: hep-ph/0409146.

[3] Stefano Frixione, Paolo Nason, and Carlo Oleari. "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method". In: *JHEP* 11 (2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph].

[4] Simone Alioli et al. "NLO vector-boson production matched with shower in POWHEG". In: *JHEP* 0807 (2008), p. 060. DOI: 10.1088/1126-6708/2008/07/060. arXiv: 0805.4802 [hep-ph].

[5] Simone Alioli et al. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX". In: *JHEP* 06 (2010), p. 043. DOI: 10.1007/JHEP06(2010)043. arXiv: 1002.2581 [hep-ph].

[6] T. Sjöstrand, S. Mrenna, and P. Skands. "Brief Introduction to PYTHIA 8.1". In: *Comput. Phys. Comm.* 178 (2008), p. 85. DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820v1 [hep-ph].

[7] ATLAS Collaboration. "Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector". In: *JHEP* 09 (2014), p. 145. DOI: 10.1007/JHEP09(2014)145. arXiv: 1406.3660 [hep-ex].

[8] Piotr Golonka and Zbigniew Was. "PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays". In: *Eur. Phys. J.* C45 (2006), pp. 97–107. DOI: 10.1140/epjc/s2005-02396-4. arXiv: hep-ph/0506026.

[9] Stefan Höche et al. "NLO matrix elements and truncated showers". In: *JHEP* 1108 (2011), p. 123. DOI: 10.1007/JHEP08(2011)123. arXiv: 1009.1127 [hep-ph].

[10] Richard D. Ball et al. "Parton distributions with LHC data". In: *Nucl. Phys. B* 867 (2013), p. 244. DOI: 10.1016/j.nuclphysb.2012.10.003. arXiv: 1207.1303 [hep-ph].

[11] ATLAS Collaboration. *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model*. ATL-PHYS-PUB-2016-017. 2016. URL: https://cds.cern.ch/record/2206965.

[12] S. Catani and M. Grazzini. "An NNLO subtraction formalism in hadron collisions and its application to Higgs boson production at the LHC". In: *Phys. Rev. Lett.* 98 (2007), p. 222002. DOI: 10.1103/PhysRevLett.98.222002. arXiv: hep-ph/0703012 [hep-ph].

[13] S. Catani et al. "Vector boson production at hadron colliders: A Fully exclusive QCD calculation at NNLO". In: *Phys. Rev. Lett.* 103 (2009), p. 082001. DOI: 10.1103/PhysRevLett.103.082001. arXiv: 0903.2120 [hep-ph].

[14] L.A. Harland-Lang, A. D. Martin, P. Motylinski, R. S. Thorne. "Parton distributions in the LHC era: MMHT 2014 PDFs". In: *Eur. Phys. J. C* 75.5 (2015), p. 204. DOI: 10.1140/epjc/s10052-015-3397-6. arXiv: 1412.3989 [hep-ph].

[15] ATLAS Collaboration. "Measurement of $W^{\pm}$ and Z-boson production cross sections in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Phys. Lett. B* 759 (2016), p. 601. DOI: 10.1016/j.physletb.2016.06.023. arXiv: 1603.09222 [hep-ex].

[16] ATLAS Collaboration. "Measurements of W and Z boson production in pp collisions at $\sqrt{s} = 5.02$ TeV with the ATLAS detector". In: *Eur. Phys. J. C* 79 (2019), p. 128. DOI: 10.1140/epjc/s10052-019-6622-x. arXiv: 1810.08424 [hep-ex].

[17] CMS Colaboration. "Measurement of the inclusive $t\bar{t}$ cross section in pp collisions at $\sqrt{s} = 5.02$ TeV using final states with at least one charged lepton". In: *JHEP* 03 (2018), p. 115. DOI: 10.1007/JHEP03(2018)115. arXiv: 1711.03143 [hep-ex].