W+jets with ATLAS

Christina Nelson
she/they

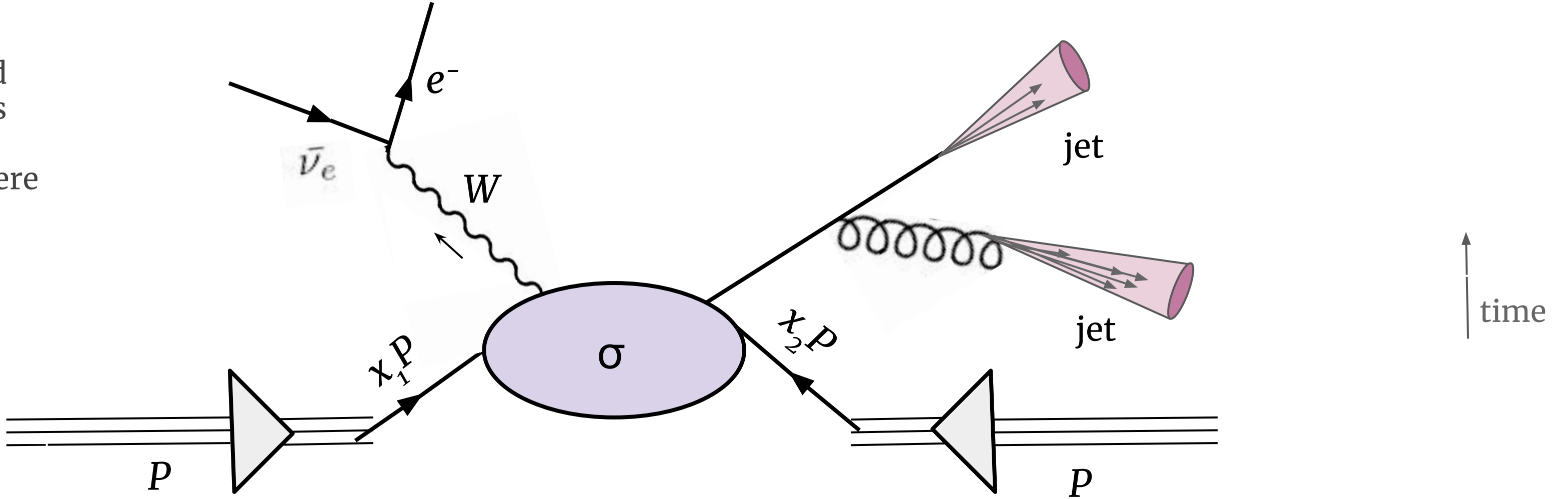
Supervisor: Prof Francois Corriveau

christina.nelson@mail.mcgill.ca

WNPPC2021

a Cross-section in pp Collisions at the LHC

Let's consider a hard parton-level process involving scattering into final state *X* where *Y* may denote any hadronic final state



The total inclusive cross section can be written as:

$$\sigma_{pp\to X+Y} = \int_0^1 dx_1 \int_0^1 dx_2 \sum_{1,2} f_1(x_1, \mu_F^2) f_2(x_2, \mu_F^2) \hat{\sigma}((x_1 P + x_2 P), \mu_R^2 \to X)$$

The total cross-section is factorized into a "hard part", and "normalization" from Parton distribution functions (PDFs)

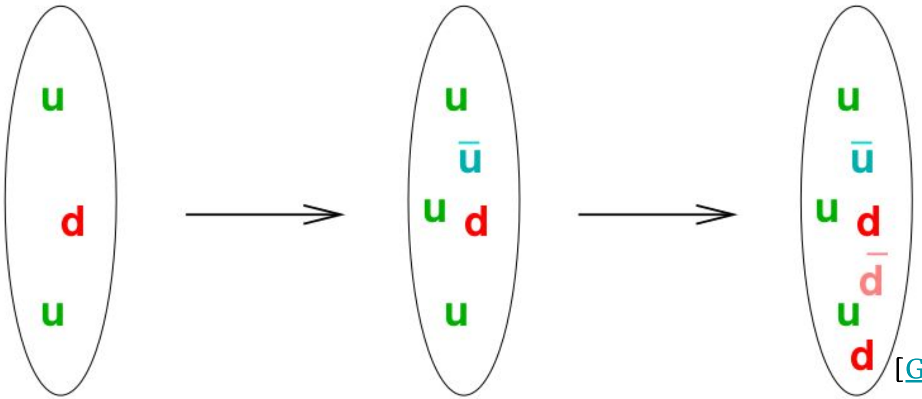
μ_R is the interaction scale

μ_F is the factorization scale

The sum runs over all species of quarks and antiquarks, weighted by probability

And integrated over all possible parton momenta, x_1 and x_2.

Charge Asymmetry of the W? How?



[G.Salam]

- The proton wavefunction fluctuates
- Virtual $u\bar{u}$ and $d\bar{d}$ pairs may be produced from gluons inside the proton (higher order processes)

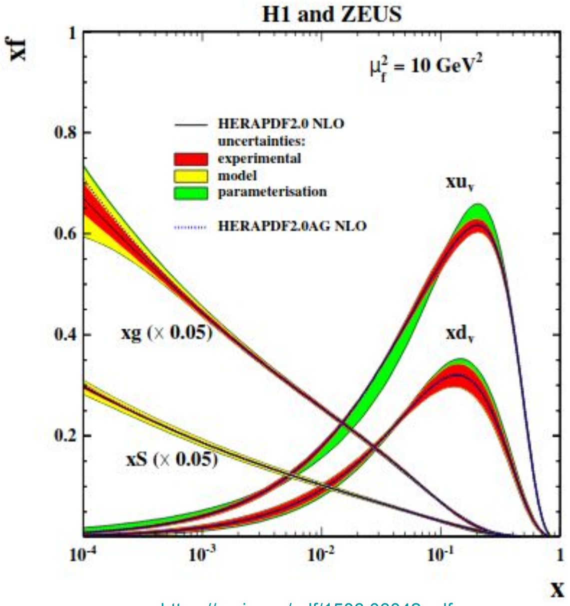
- The number of extra quark, anti-quark pairs can be infinite: $\int dx(u(x)+\bar{u}(x))=\infty$
- When we say that the proton is made up of *uud*, we mean: $\int dx(u(x)-\bar{u}(x))=2, \int dx(d(x)-\bar{d}(x))=1$
- $u-\bar{u}=u_v$ is called the valence distribution
- How can we measure the difference between u and ubar?
  - The photon interacts the same with both
  - The answer is the *W* boson! ;)

A meaningful measurement for valence quark distributions in PDFs: Asymmetry (A)

$$A=\frac{\sigma_{W^+}-\sigma_{W^-}}{\sigma_{W^+}+\sigma_{W^-}}$$

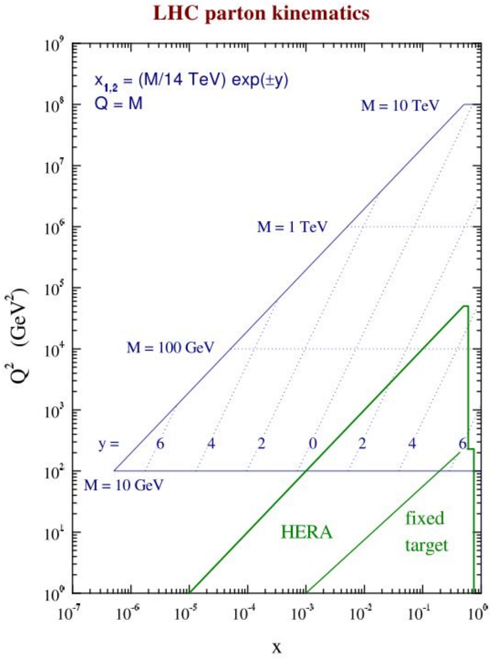
Parton Distribution Functions

- Historically, and still considered a standard, PDFs are obtained from DIS experiments like H1 and ZEUS
- This PDF plot shows the dynamics within the proton
  - Quarks may interact with each other
  - Gluons may produce virtual quark pairs



https://arxiv.org/pdf/1506.06042.pdf

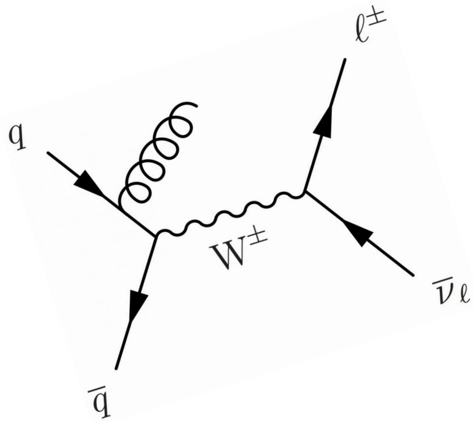
- LHC probes different regions of phase space compared to DIS experiments
- *W*+jets are sensitive to gluon PDF
- Charged *W*+jets measurements are sensitive to *u* and *d* valence PDFs (at low *x*)

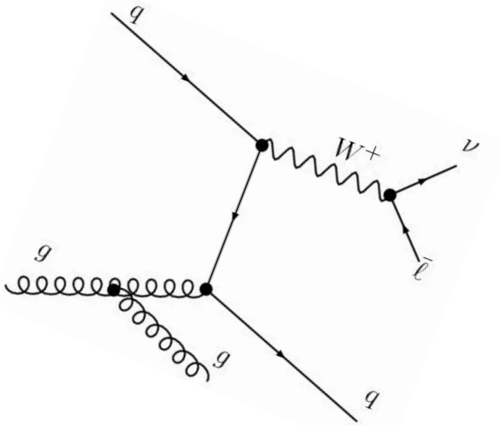


https://cds.cern.ch/record/1263062/files/parton_kin.png

x = the fraction of the proton's longitudinal momentum carried by the parton initiating the interaction

W+jets at the LHC





- *W*+jets at the LHC contains energy scales from the centre of mass of the colliding partons, to the *W* mass, to the scale of the strong interaction ΛQCD
- Provides valuable benchmarks for precision tests of the Standard Model
- Offers robust environment for probing perturbative QCD (pQCD) and providing inputs to PDFs in regions where pQCD is not viable.
- Contributes as a dominant background to Higgs measurements, and many Beyond the Standard Model searches, thus precision measurements are vital.

The W+jets Analysis

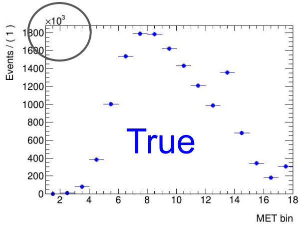
- We are doing the first measurement of W+jets in ATLAS using full run2 data (2015-2018) at 13 TeV
  - Canadian members of the group include: McGill (e channel and cross check μ channel) and University of Toronto (μ channel)

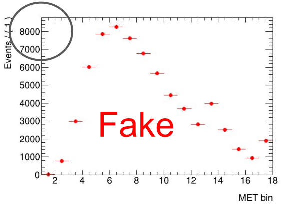
- Contributions currently in progress:
  - Multijet background estimation
  - Signal significance studies
  - Comparison & validation of Monte Carlo (MC) generators: Matrix Element (ME), matching /merging (MC tuning), Parton Showering (PS)
  - Beyond scope of this presentation: Parallelization of analysis code & using machine learning for unfolding measured observables to parton level information

- Study our W+jets measurement impact on Parton Distribution Functions (PDFs)

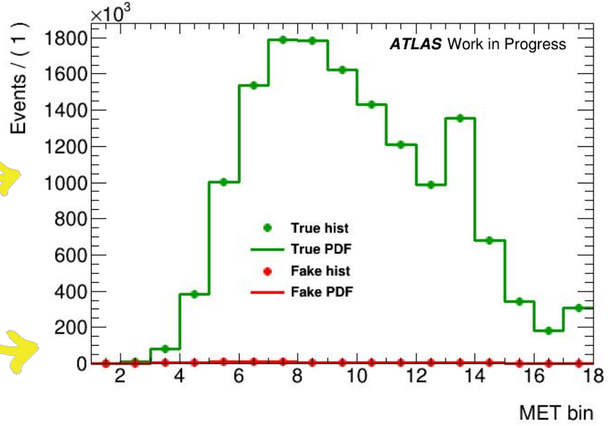
Multijet Background Estimation

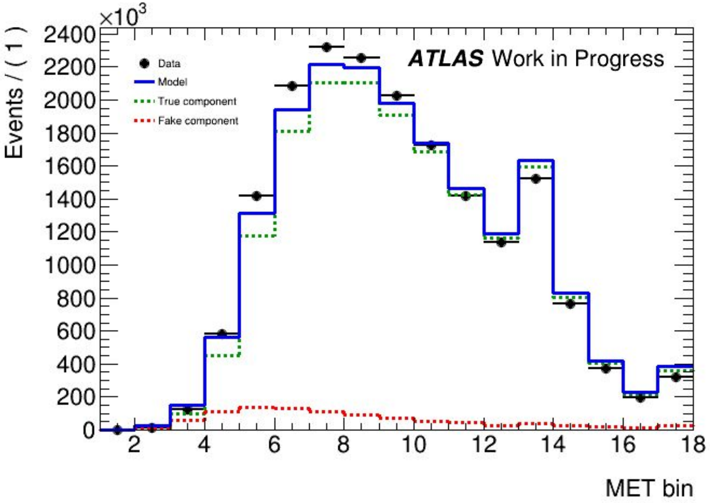
- A dominant background comes from jets that are mistaken for signal leptons, this is referred to as the multijet (MJ) background.
- We are using a template method with an extended maximum likelihood fit to the missing transverse energy (MET) distribution
- True template is made from MC signal + electroweak backgrounds in the SR w/o MET cut
- Fake MJ template is made from Data in the CR





Model: $n_{\mathrm{True}} \times \mathrm{True} + n_{\mathrm{Fake}} \times \mathrm{Fake}$





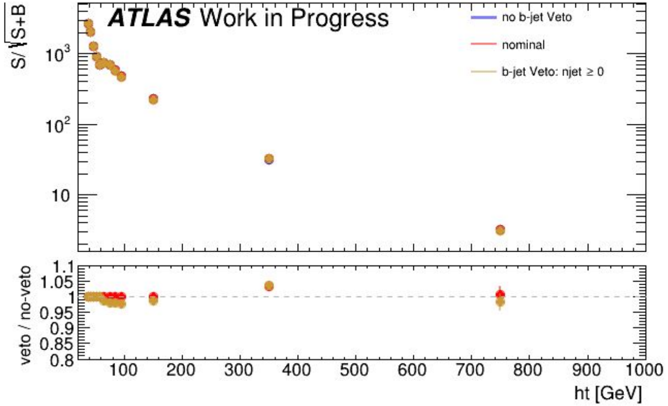
- Estimate MJ background by extracting the number of events from the fake template, integrated above the MET cut applied in the SR

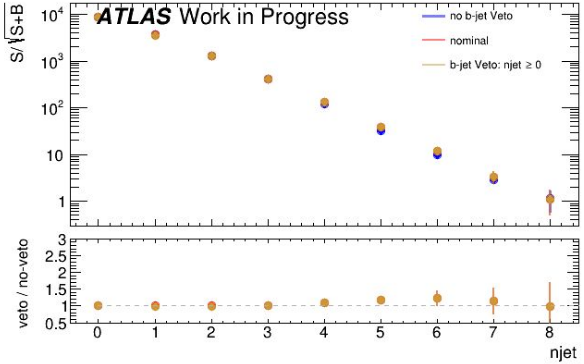
MET > 25 GeV

- SR = Signal Region, CR = Control Region, events selected such that MJ events are enhanced

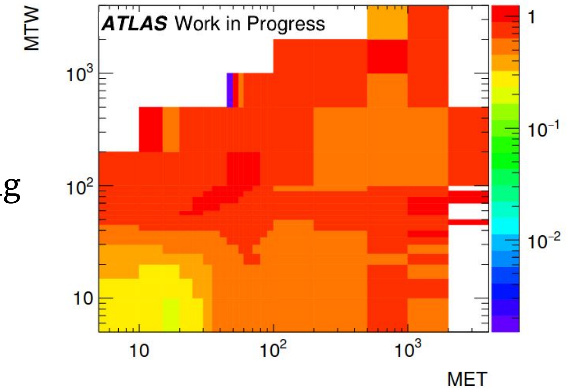
Significance Studies

- Applying a b-jet veto to selected jets aids in suppressing ttbar background.
- A study was done to assess if it is advantageous to apply the veto to all selected jets or nominally, when the number of jets is larger than 2.
- We found no significance gains when applying veto to all jets.





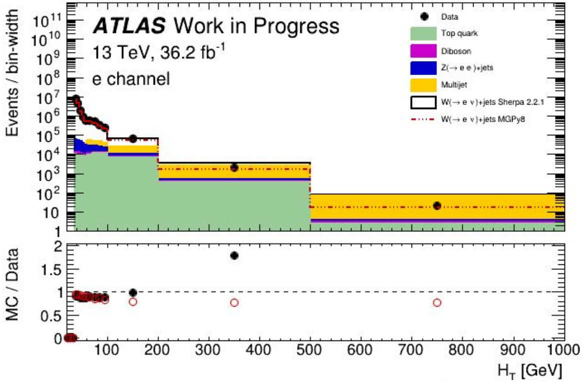
- Currently investigating using of a 2D cut: MET+MTW > 60 GeV



$$\mathrm{MTW} = \sqrt{2 \cdot p_T^l \cdot E_T^{miss} \cdot (1 - \cos(\Delta\phi(\mathrm{lep}, E_T^{miss})))}$$

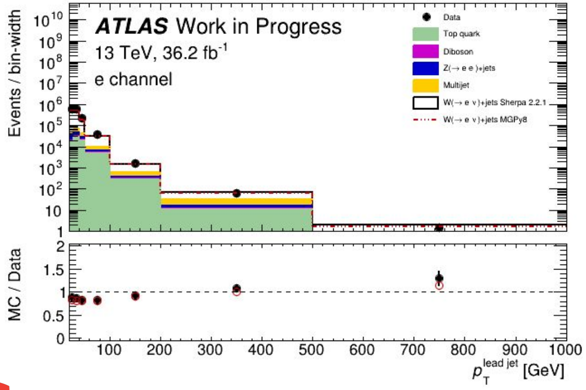
MC/Data Comparisons

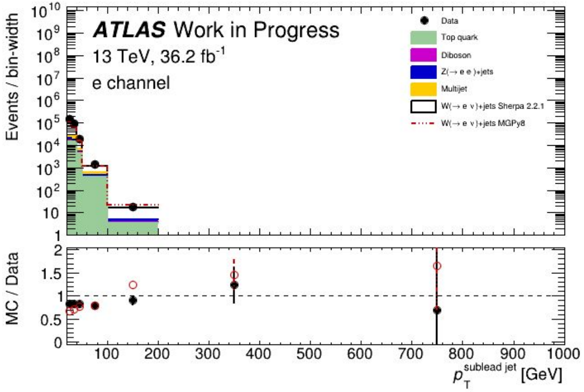
- Preliminary distributions with MJ estimation applied
- Sherpa221 and MGPy8 are compared
- Electron channel shown here



$$H_{\mathrm{T}} = \sum_j p_{\mathrm{T}}^j + \sum_l p_{\mathrm{T}}^l$$

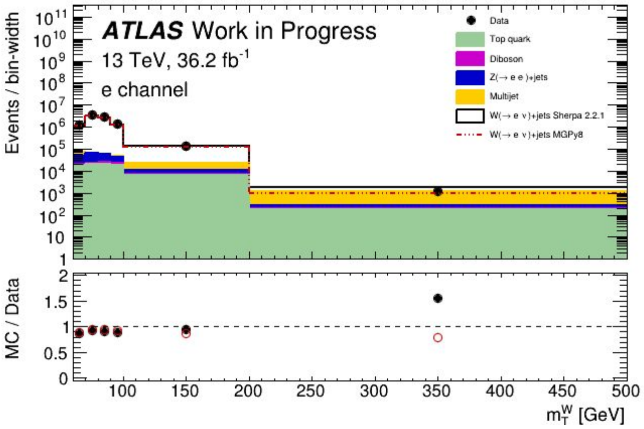
- Large disagreement in last couple of bins for H_T when using Sherpa221
- Leading jet p_T and subleading jet p_T have similar MC comparisons
- Need to check distribution for the lepton p_T



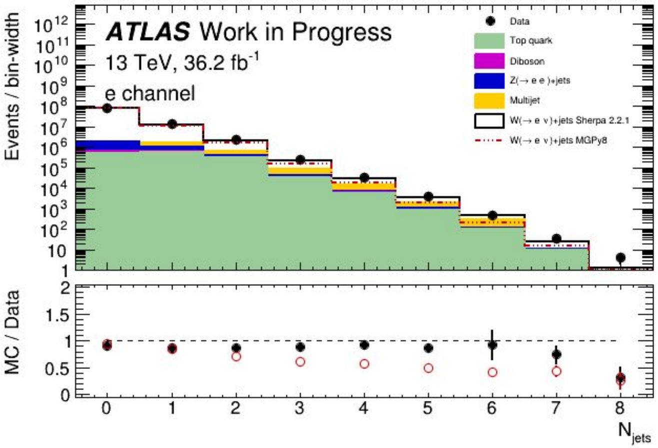


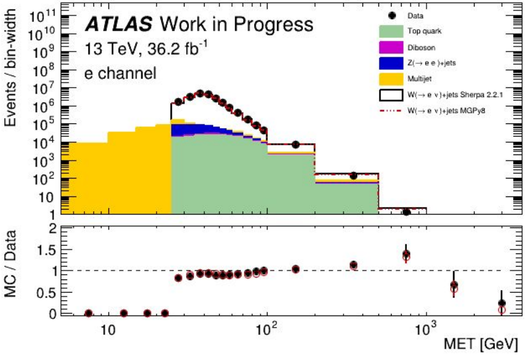
- NB: only statistical uncertainties shown here, and comparison done at reco leve (ie. unfolding not done yet)

MC/Data Comparisons



$$W\ m_{\mathrm{T}} = \sqrt{2 \cdot p_{\mathrm{T}}^{l} \cdot E_{\mathrm{T}}^{\mathrm{miss}} \cdot (1 - \cos(\Delta\phi(\mathrm{lep}, E_{\mathrm{T}}^{\mathrm{miss}})))}$$



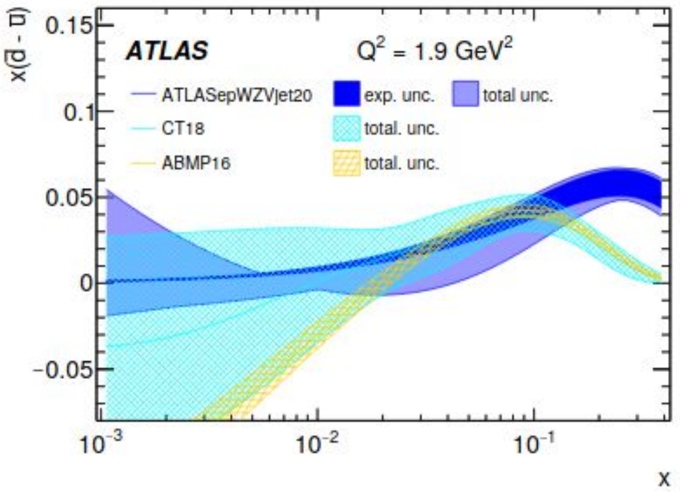


- Last p_T bin begins to diverge for Sherpa221 in MTW
- Slight downward trend in MC/Data comparison for Njets distribution
- Relatively good modelling of MET

$$E_{\mathrm{T}}^{\mathrm{miss}} = |\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}| = \sqrt{(E_x^{\mathrm{miss}})^2 + (E_y^{\mathrm{miss}})^2},$$
$$E_{x(y)}^{\mathrm{miss}} = -\sum_{i,\mathrm{hard}} p_{x(y),i} - \sum_{j,\mathrm{soft}} p_{x(y),j}$$

xFitter

- An open source QCD fit framework which extracts PDFs with data as inputs
- A recent result presenting PDF analysis using ATLAS 8 TeV data on W+jets and Z+jets combined with inclusive W and Z measurements at 7 TeV and HERA data



$x(\bar{d}-\bar{u})$ distribution as a function of Bjorken x

ATLASepWZVjet20

Comparison to global PDFs show differences at higher values of *x*

https://arxiv.org/pdf/2101.05095.pdf

- Fit done at NNLO in pQCD using recent theoretical developments for W/Z production with one jet
- Evaluated at Q^2 = 1.9 GeV^2

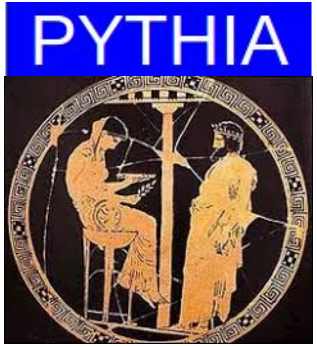
Summary

- W+jets is a benchmark measurement
- It is an important validation of SM precision, including electroweak and hard QCD effects
- W+jets is also a dominant background to many Higgs analyses and Beyond the Standard Model searches
- Background estimation from multijets has been developed
- Signal significance studies underway
- Comparing the latest MC generators and modelling
- Improving the W+jets measurement by using full run2 data
- And plan to use xFitter to perform a QCD analysis to extract PDFs from our results using ATLAS data at 13 TeV centre of mass energy

Backup

MC Generator Comparisons

- MC Signal Samples
  - Sherpa v2.2.1, with generator tune NNPDF3.0 NNLO (Sherpa221)
  - Alternative Signal Samples
    - Powheg v1_r2856 + Pythia8 v8.186 + EvtGen v1.2.0 + Photospp, generator tune AZNLO CTEQ6L1 (PPy)
    - MadGraph5_aMC@NLO v2.3.2.p1 + Pythia v8.210 + EvtGen v1.2.0, generator tune A14 NNPDF23LO (MGPy8)
- Background Samples
  - Single Top: Powheg + EvtGen v1.6.0, generator tune A14 NNPDF23LO
  - ttbar: Powheg+Pythia8 v8.230 + EvtGen v1.6.0, A14NNPDF23LO
  - Diboson, $Z\to ee$, $Z\to\mu\mu$, $Z\to\tau\tau$, $W\to\tau\nu$ Sherpa221



Event Selection I

Electron Selection (exactly 1)	
Pseudorapidity	$\|\eta\|<2.4$ (excluding $1.37<\|\eta\|<1.52$)
Electron ID	Tight : uses E_{T} and η requirements
Electron p_{T}	> 35 GeV
Electron Isolation	Tight:$E_{\mathrm{T}}^{\mathrm{cone20}}/p_{\mathrm{T}} < 0.06$, $p_{\mathrm{T}}^{\mathrm{varcone20}}/p_{\mathrm{T}} < 0.06$
Impact Parameters	$\|d_0/\sigma_{d0}\| < 5$ and $\|z_0 \cdot \sin(\theta) < 0.5\|$

$W \rightarrow e\nu$ Selection	
Triggers	unprescaled single lepton triggers
Electron-jet distance	$\Delta R(e,j) > 0.4$
MET	$E_{\mathrm{T}}^{\mathrm{miss}} > 25$ GeV
Transverse mass	$m_{\mathrm{T}}^{W} > 60$ GeV
B-jet veto	No b-jet if nSelectedJet ≥ 3

ATLAS EXPERIMENT

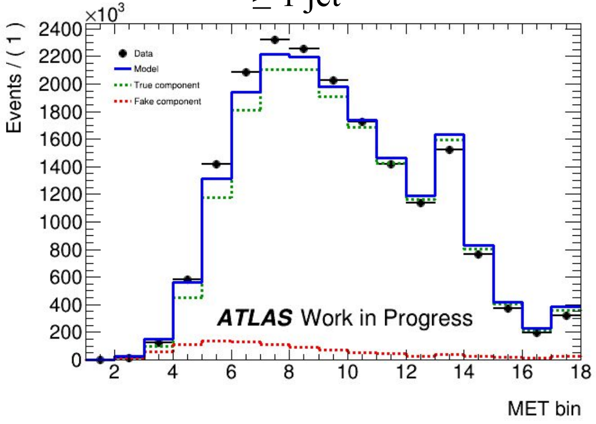
Event Selection II

Jet Selection	
Jet p_T	> 25 GeV
Rapidity	$\|y\| < 2.5$
Jet Vertex Tagger	track to vertex association : 'Loose'
cleaning	'LooseBad': high efficiency working point

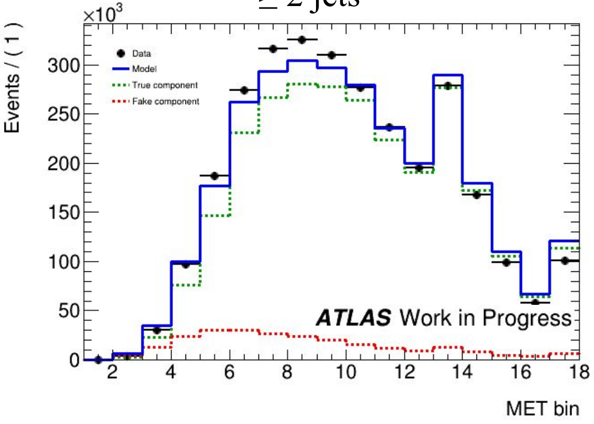
- The signal region (SR) selection is outlined in the above tables.
- The control region (CR) selection is the same except:
  - Prescaled single lepton triggers are used
  - Electron ID is 'Loose but not Tight'
  - Electron Isolation is 'Inverted Tight'

MJ Template Fits for njet multiplicities

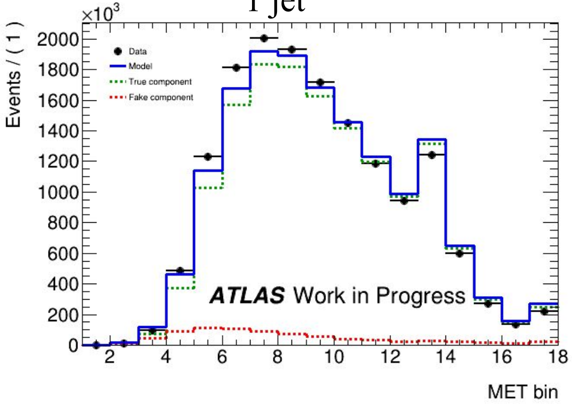
≥ 1 jet



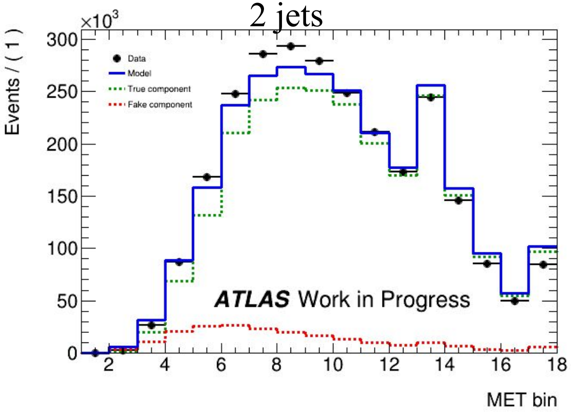
≥ 2 jets



1 jet

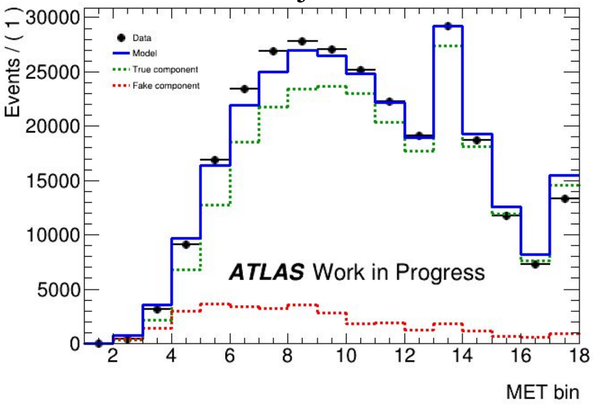


2 jets

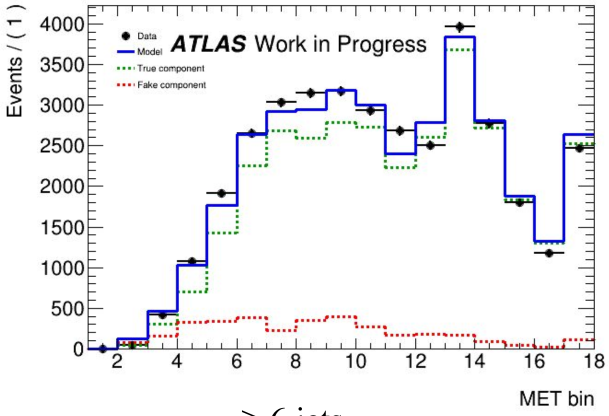


MJ Template Fits for njet multiplicities

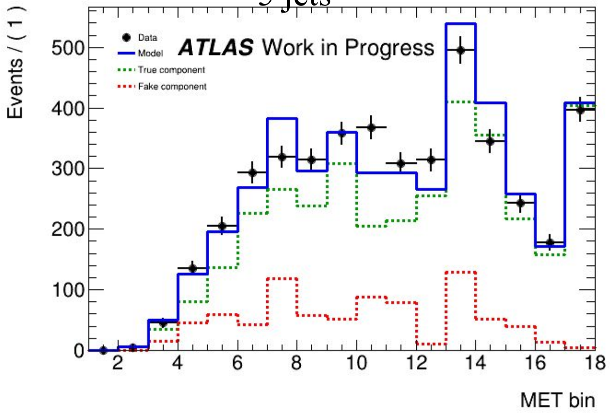
3 jets



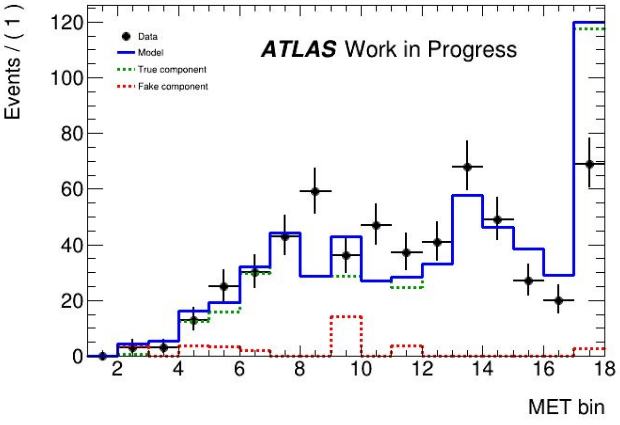
4 jets



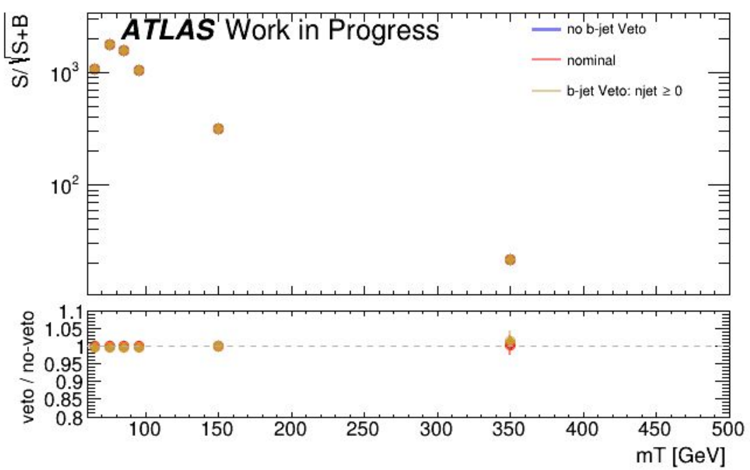
5 jets

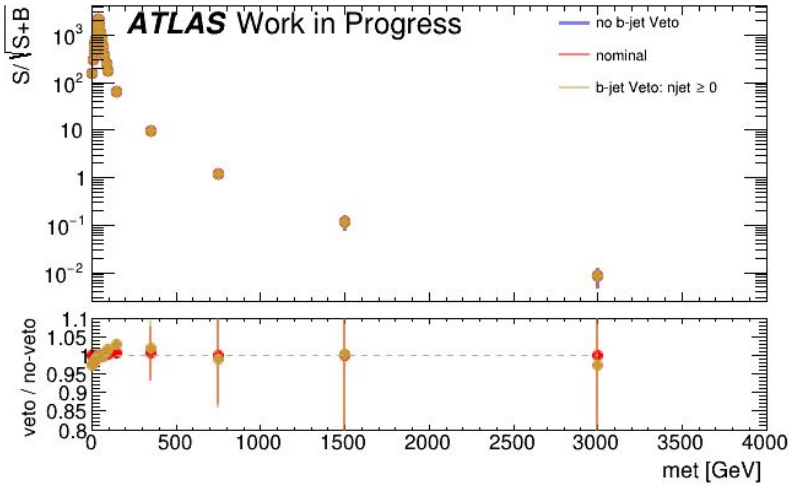


≥ 6 jets

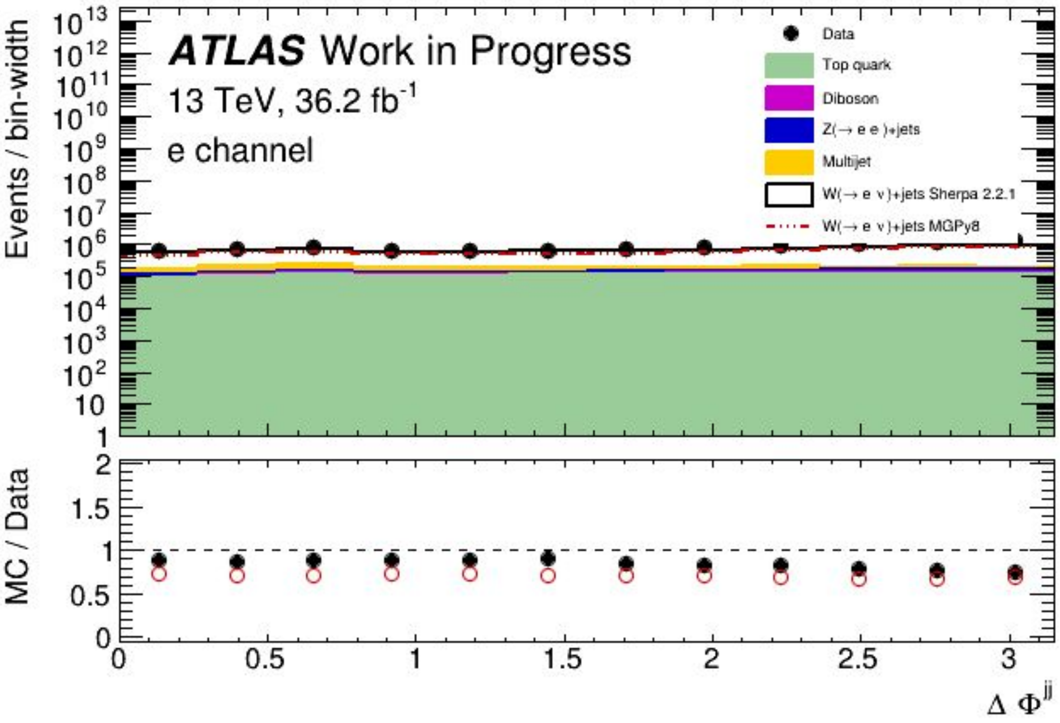


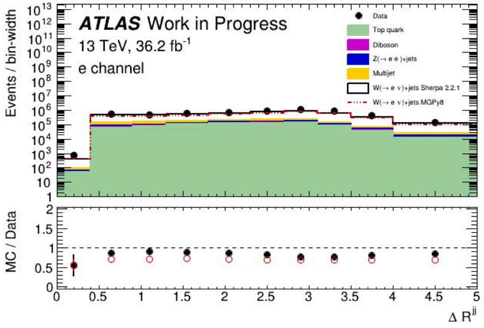
MET and MTW Significance Studies

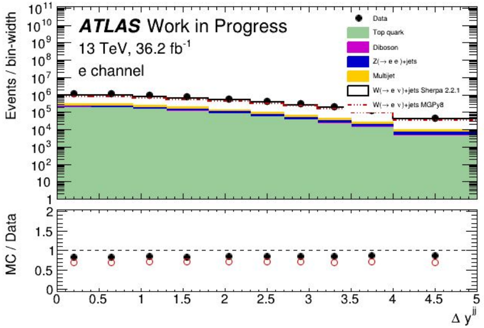




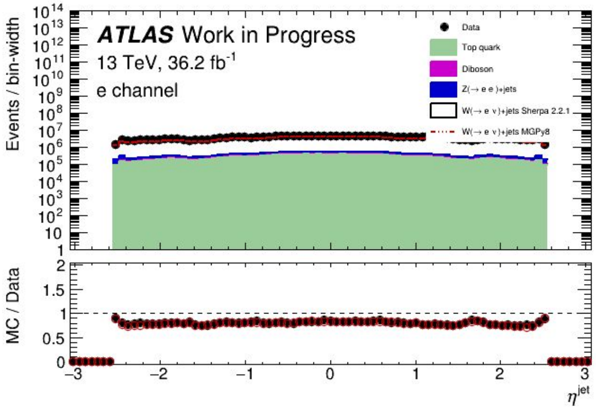
MC/Data Kinematic Distributions

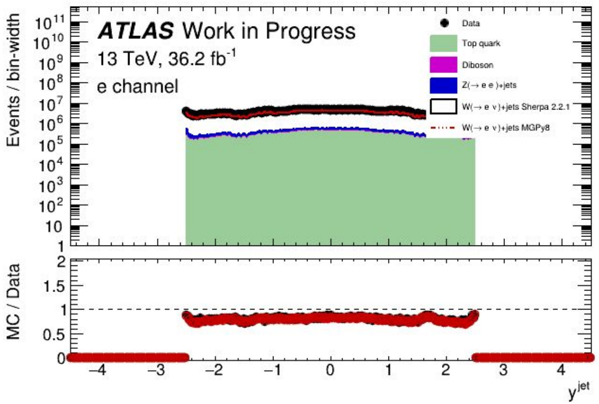


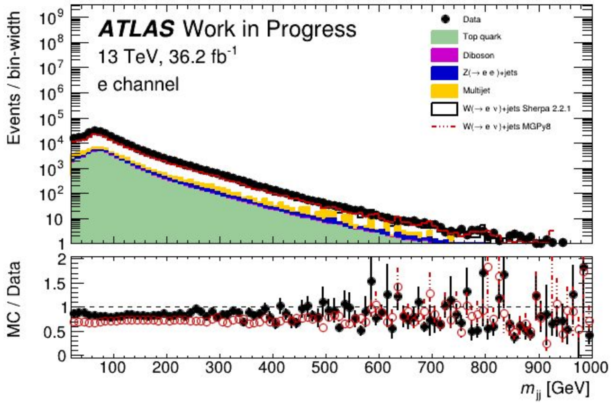




MC/Data Kinematic Distributions

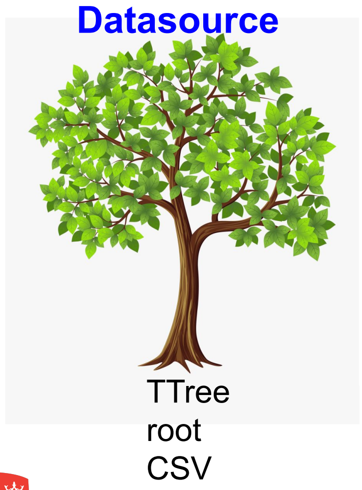


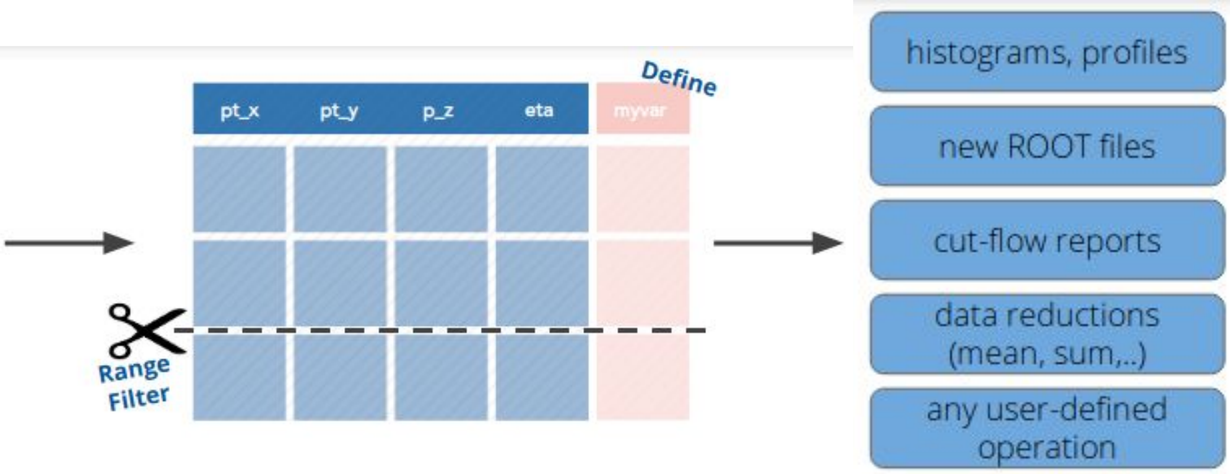


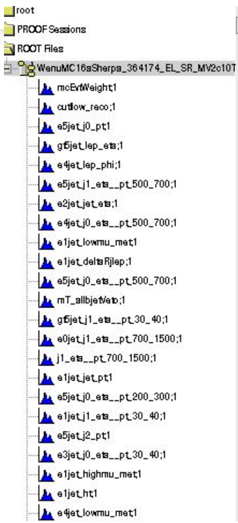


RDataFrames

- Low-level optimisation using multi-threading (MT) parallelisation and caching allows efficient exploitation of resources
- Inspired by other dataframe APIs like pandas DataFrames, RDataFrames is made for High Energy Physics Users.







- Cuts are applied as filters to columns as row operations

https://root.cern.ch/doc/v614/classROOT_1_1RDataFrame.html

ML for Unfolding

- Starting on implementation of using machine learning (ML) for unfolding in W+jets
- Unfolding is the process of getting particle-level (aka "truth") information. In the most simple description unfolding corrects for detector effects. Truth level distributions are independent of the experimental setup, thus and may be directly compared to the latest advances in theory, and compared to other experiments.
- A new method named "OmniFold" shows impressive results when applying ML to unfolding.
  - Unbinned, works for arbitrarily high-dimensional data, incorporates information from full phase space
  - Arxiv: https://arxiv.org/pdf/1911.09107.pdf