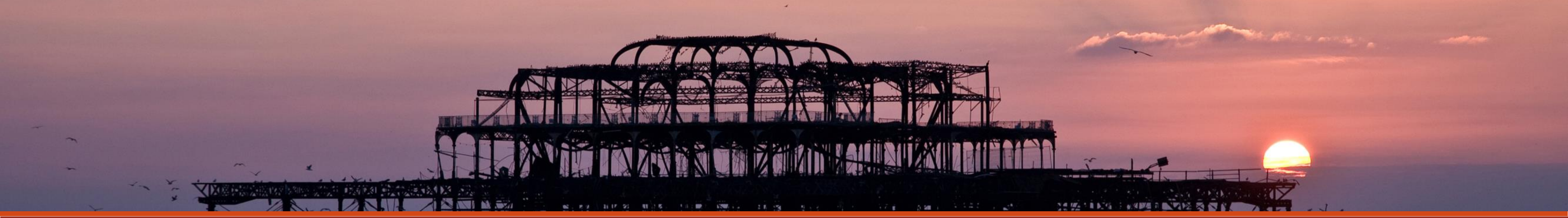
“Data-Driven background estimation techniques used in ATLAS searches for SUSY”

IOP HEPP/APP Brighton 2016

Calum Macdonald (University of Sheffield)



Data-driven techniques overview (1)

- **Why do we need them?**
  - Gives a backup to Monte-Carlo driven methods which are the primary choice for the majority of ATLAS SUSY searches

    ** Most searches which use Monte-Carlo are semi-data-driven: Monte-Carlo normalised in 'control regions'
  - The QCD (Multi-jet) background from jet mis-measurement take a lot of computing power to generate statistics in high E_T^{miss} regions → majority of SUSY searches

- **What are the advantages?**
  - Systematics uncertainties arising from detector measurements do not have to be evaluated
    - E.g. JET energy resolution, JET energy scale etc. etc.
  - JetSmearing/Template method – large increase in statistics vs MC based estimate
  - MC can be mis-modelling for many reasons

- **What are the disadvantages?**
  - Lower amount of statistics vs MC based estimate in the case of Z+jets from photon+jets
  - Need to evaluate theoretical uncertainties as many methods rely on MC in some way at either TRUTH or RECO (reconstruction) level for various corrections

Data-driven techniques overview (2)

Background	Method(s)	Run-II analyses usage
Many e.g. Top production, W+jets, Z+jets	Monte-Carlo based estimation, normalised in CRs (control regions)	Majority
Lost Lepton(s)	Lepton replacement method	Not yet used in Run-II
Tau(s)	Tau replacement method	Not yet used in Run-II
QCD/Multi-jet fluctuations	**JetSmearing**	**Sbottom, stop 0L, strong 0L, Monojet**
	Template / ABCD method	**Multi-jet (0-7 jets)**
Z+jets	**ZfromGamma**	**Sbottom, strong 0L, RJigSaw, 2 Lepton Z + MET**
	ZfromLightJets	**Sbottom**
Z+met	**JetSmearing**	**2 Lepton Z + MET**

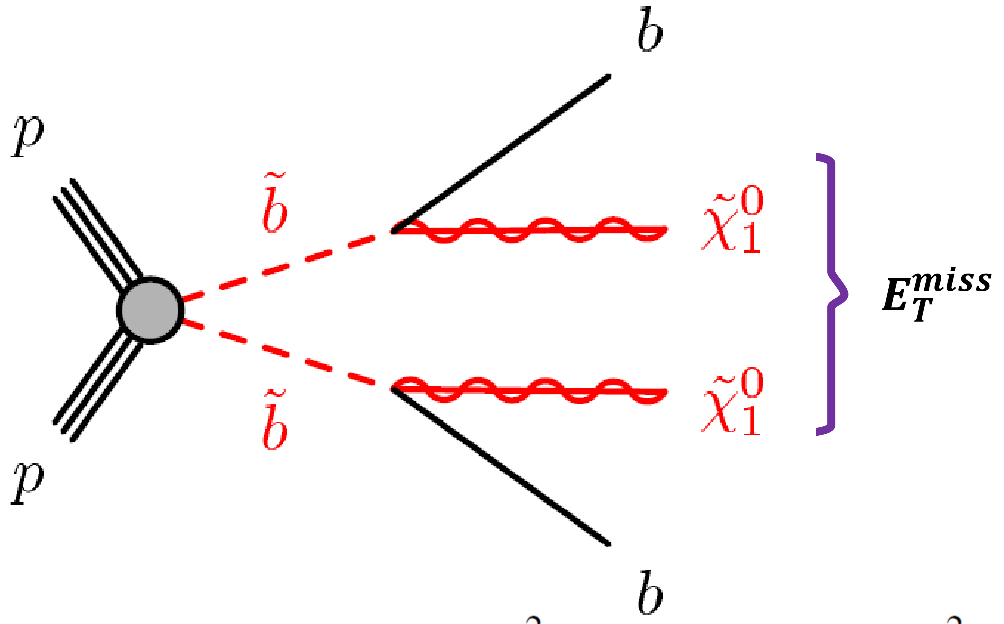
Outline of the Talk

Background	Method(s)	Run-II analyses usage
Many e.g. Top production, W+jets, Z+jets	Monte-Carlo based estimation, normalised in CRs (control regions)	Majority
Lost Lepton(s)	Lepton replacement method	Being used, but not public
Tau(s)	Tau replacement method	Not yet used in Run-II
QCD/Multi-jet fluctuations	**JetSmearing**	**Sbottom, stop 0L, strong 0L, Monojet**
	Template / ABCD method	**Multi-jet (0-7 jets)**
Z+jets	**ZfromGamma**	**Sbottom, strong 0L, RJigSaw, 2 Lepton Z + MET**
	ZfromLightJets	**Sbottom**
Z+met	**JetSmearing**	**2 Lepton Z + MET**

- Methods to estimate the Z+jets background
  - Sbottom analysis: overview of Z+b-jets estimation
- Template fit method for Multi-jet analysis
- JetSmearing method

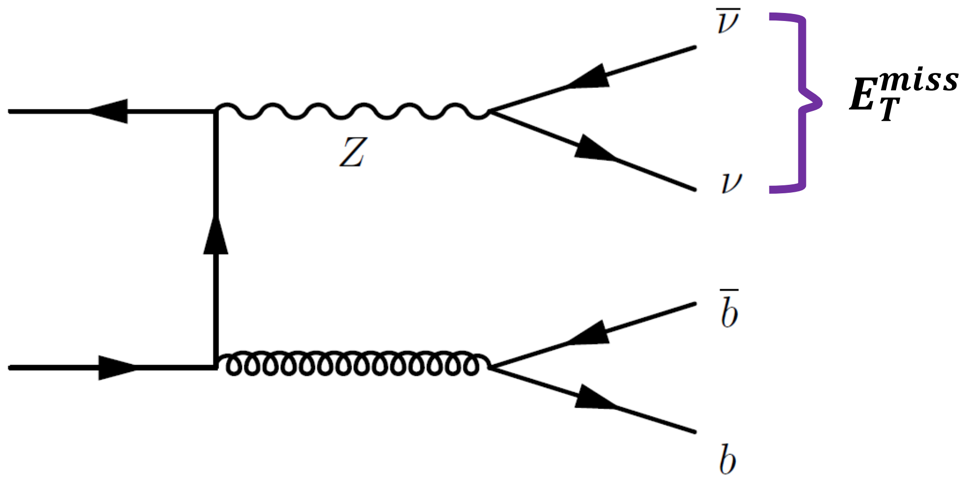
Generic search strategies e.g sbottom 0L (1)

- Public CONF note, December 2015:
  - https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2015-066/
- Cut and count signal regions (SRs)
  - Further details in IOP talk by John Anders (Liverpool)
- Sbottom pair production with 2 b-jets + large missing transverse energy
- m_{CT}(1st b-jet, 2nd b-jet) key discriminant



$$m_{CT}^2(v1,\ v2) = [E_T(v1) + E_T(v2)]^2 - [\mathbf{p}_T(v1) - \mathbf{p}_T(v2)]^2$$

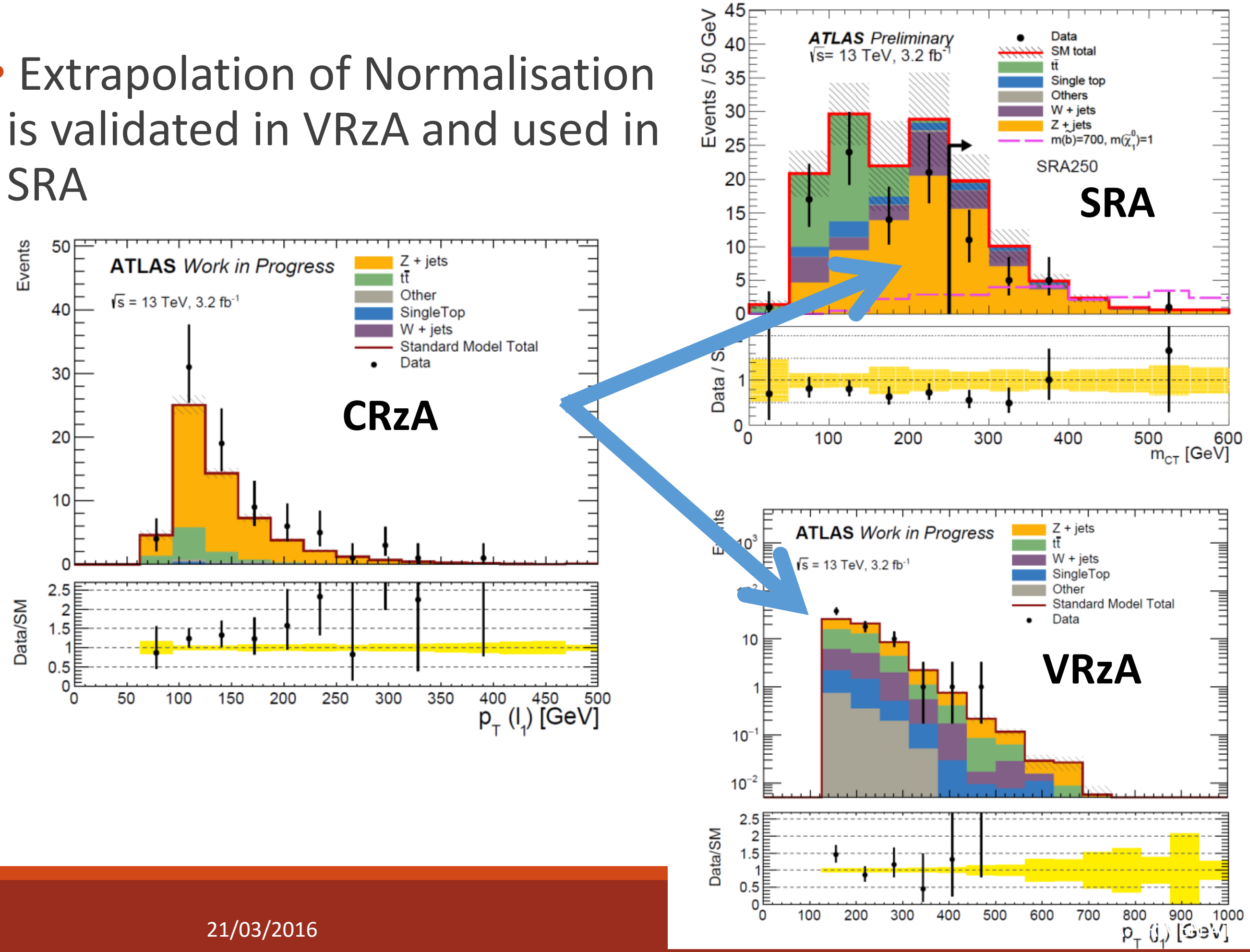
- Main background from Z+b-jets where the Z decays into neutrinos → which can give large missing transverse energy

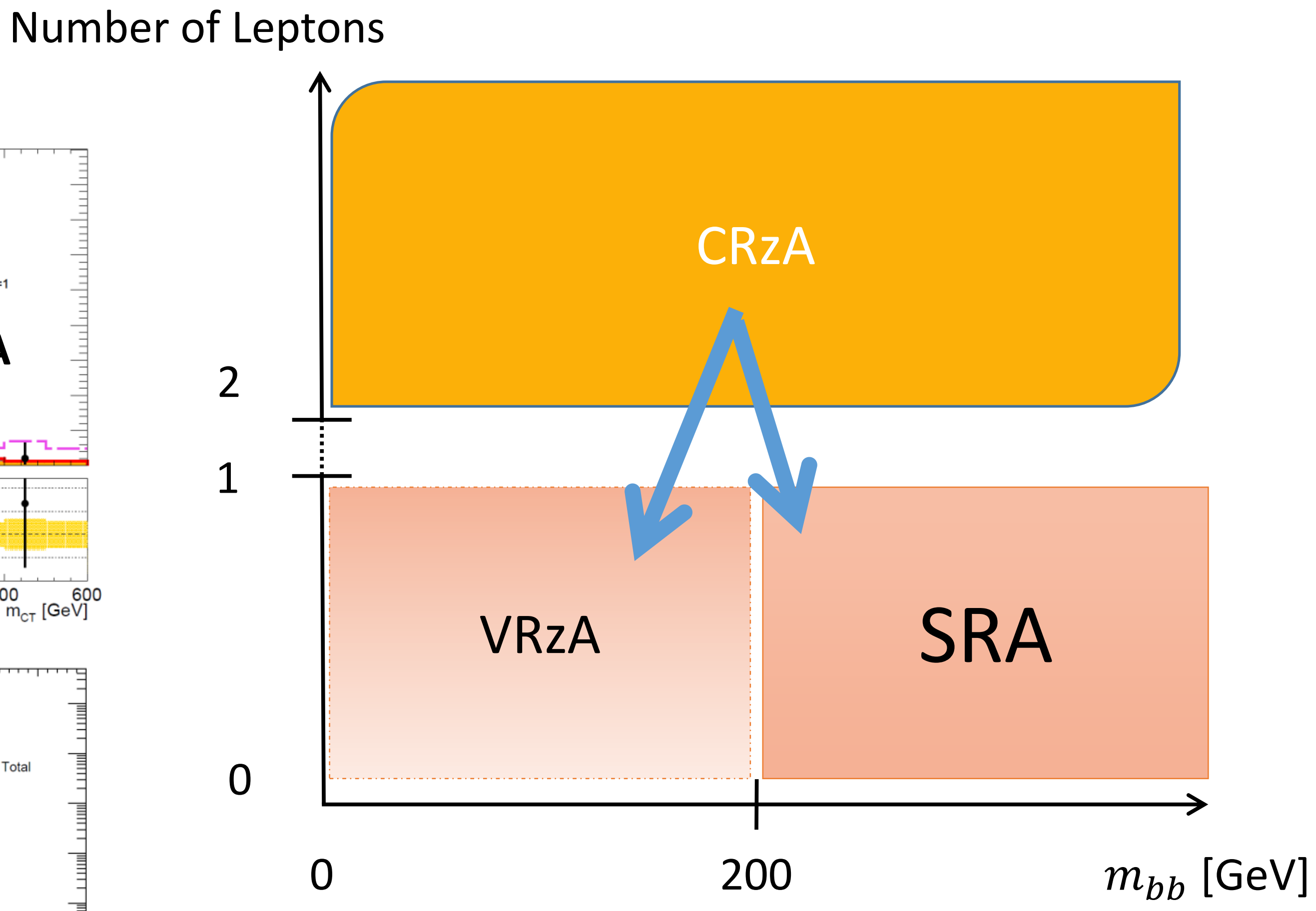


- General strategy is 'semi-data-driven'
  - **Monte carlo simulation is normalised in 'control regions' CRs and extrapolated to validation regions (VRs) and the SRs**
- **Alternatively use fully data-driven estimations**

Generic search strategies e.g sbottom 0L (2)

- CRzA
  - Control region for the Z+b-jets process
  - Float the normalisation of the process(es) to match the data
- Extrapolation of Normalisation is validated in VRzA and used in SRA





https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2015-066

Alternative estimations of the Z+b-jet background (1)

- Two basic data-driven methods that will be discussed exploiting replacing $Z \rightarrow \nu\nu$ in Z+b-jets

$\boldsymbol{Z \rightarrow \nu\nu}$ + b-jets from $\boldsymbol{Z \rightarrow ll}$ + light-jets

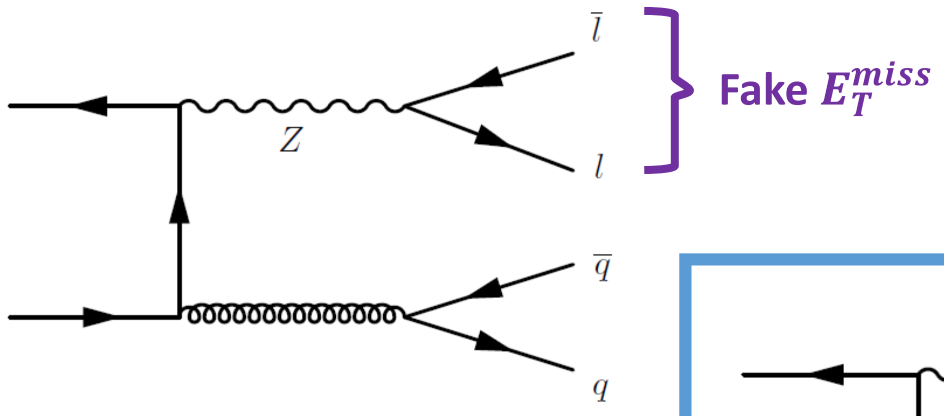
- Fake E_T^{miss} can be created from well measured 2 lepton events by vectorially adding the leptons to the real E_T^{miss}
- Relies on key discriminating variables e.g. m_{CT} to have similar shape
- Disadvantages:
  - Cross section lower for decays into two leptons
  - Compensate by using light-jets
    - Different Feynman diagrams
    - Kinematics of the jets therefore important
- Advantages:
  - Large amount of statistics
  - Dominant systematic uncertainty from stats

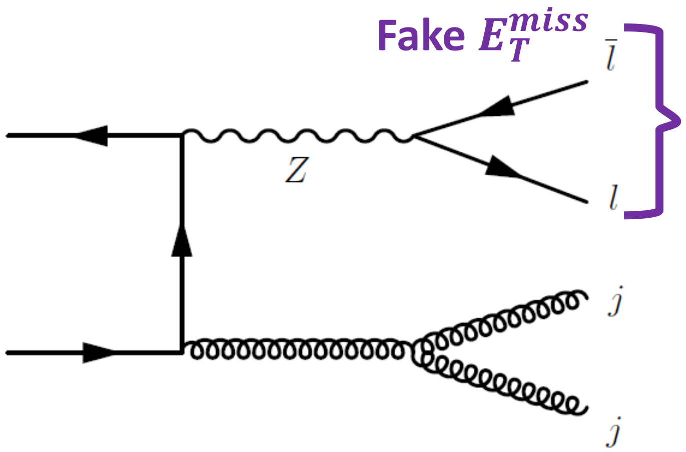
$\boldsymbol{Z \rightarrow \nu\nu}$ + b-jets from $\boldsymbol{\gamma}$ + b-jets

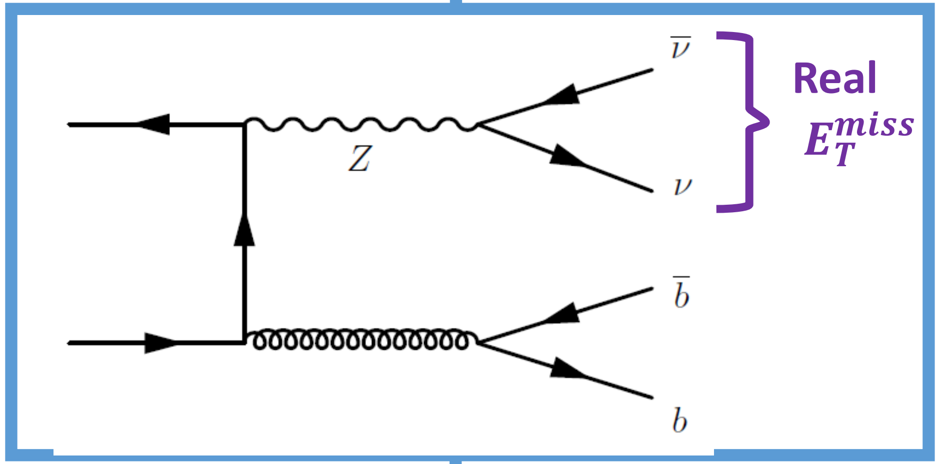
- Fake E_T^{miss} can be created from well measured photon events by vectorially adding the photon to the real E_T^{miss}
- Disadvantages:
  - Statistics are low
  - $\boldsymbol{\gamma}$ is massless, Z is massive → pT (boson) reweighting needed
- Advantages:
  - Cross section for $\boldsymbol{\gamma}$ production is higher than for Z$\boldsymbol{\nu\nu}$
  - No need to use light jets
    - Similar Feynman diagrams
    - Jet kinematics are not so important

Alternative estimations of the Z+b-jet background (2)

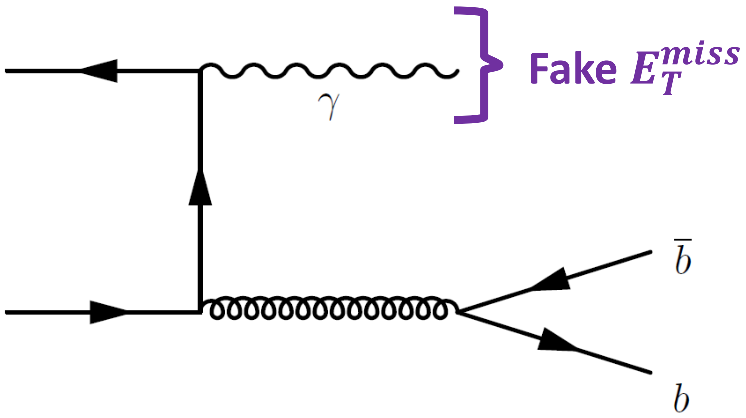
$Z \rightarrow \nu\nu$ + b-jets from $Z \rightarrow ll$ + light-jets







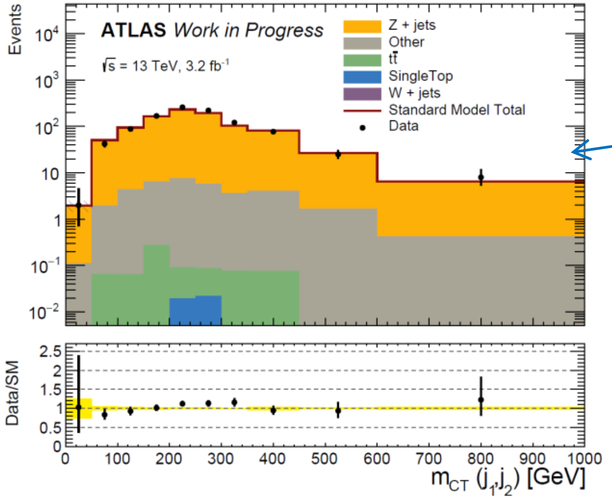
$Z \rightarrow \nu\nu$ + b-jets from γ + b-jets



"Cleaner" of the two methods: Light-jets method greatly affected with higher jet multiplicity because of gluon/quark jets

$Z \to \nu\nu$ + b-jets from $Z \to ll$ + light-jets (1)

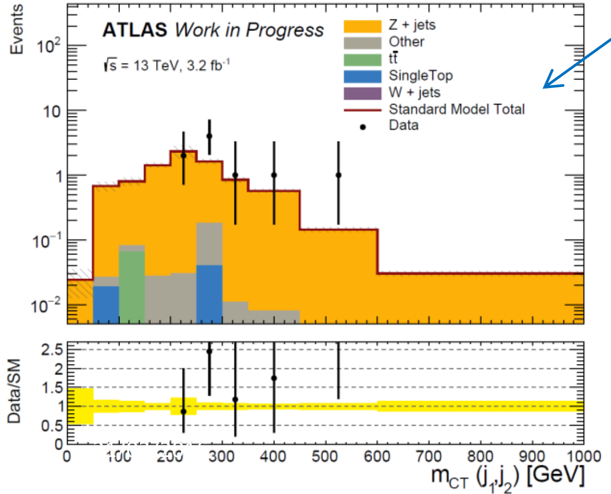
- You have started with 2 leptons + light-jets (0b2l) → want to estimate 0 leptons + b-jets (2b0l)
- 1) Emulate the SR in 0b2l events, take the data in this region – nonZ background 2) Correct shape per mct bin for 2L->0L 3) Normalise shape for 0B->2B with data

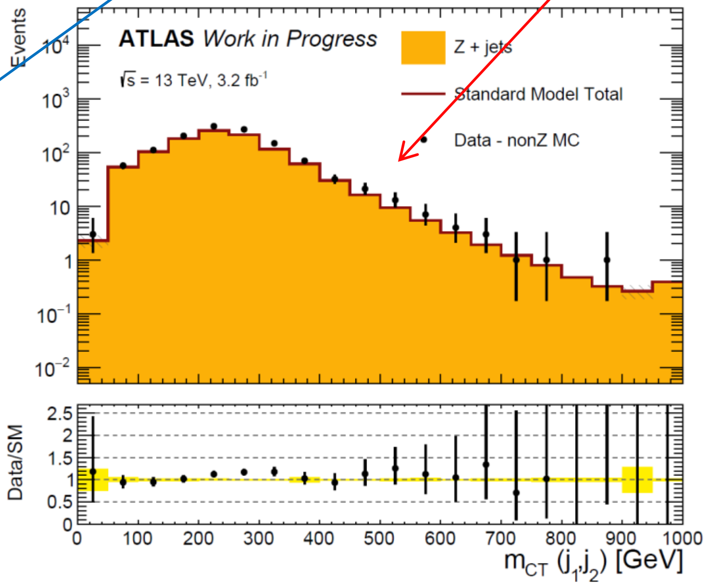


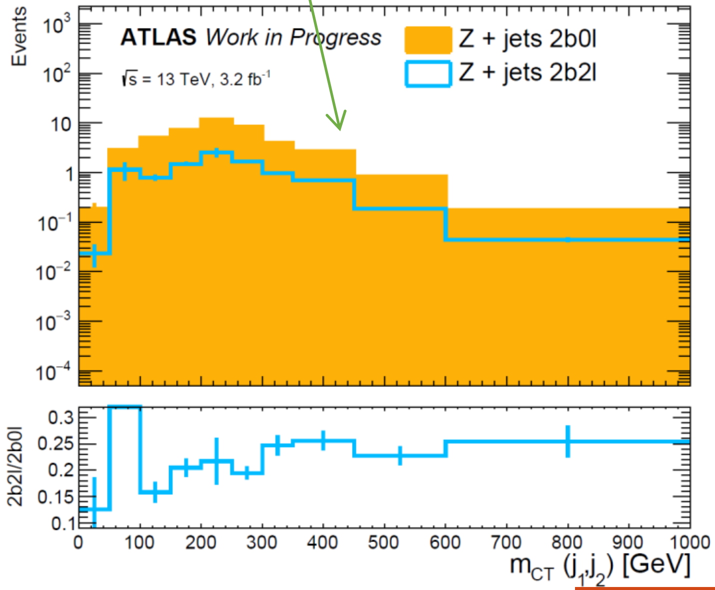
(sum)

$$N^{SR}_{Z\nu\nu} = \frac{\Gamma^{\rm data}_{Z\ell\ell;2b}}{\Gamma^{\rm data}_{Z\ell\ell;0b}} \times \int_X^\infty f^{\rm data}_{Z\ell\ell;0b}(m_{\rm CT}) \times \frac{f^{\rm MC}_{Z\nu\nu;2b}(m_{\rm CT})}{f^{\rm MC}_{Z\ell\ell;2b}(m_{\rm CT})} dm_{\rm CT}$$

(sum)





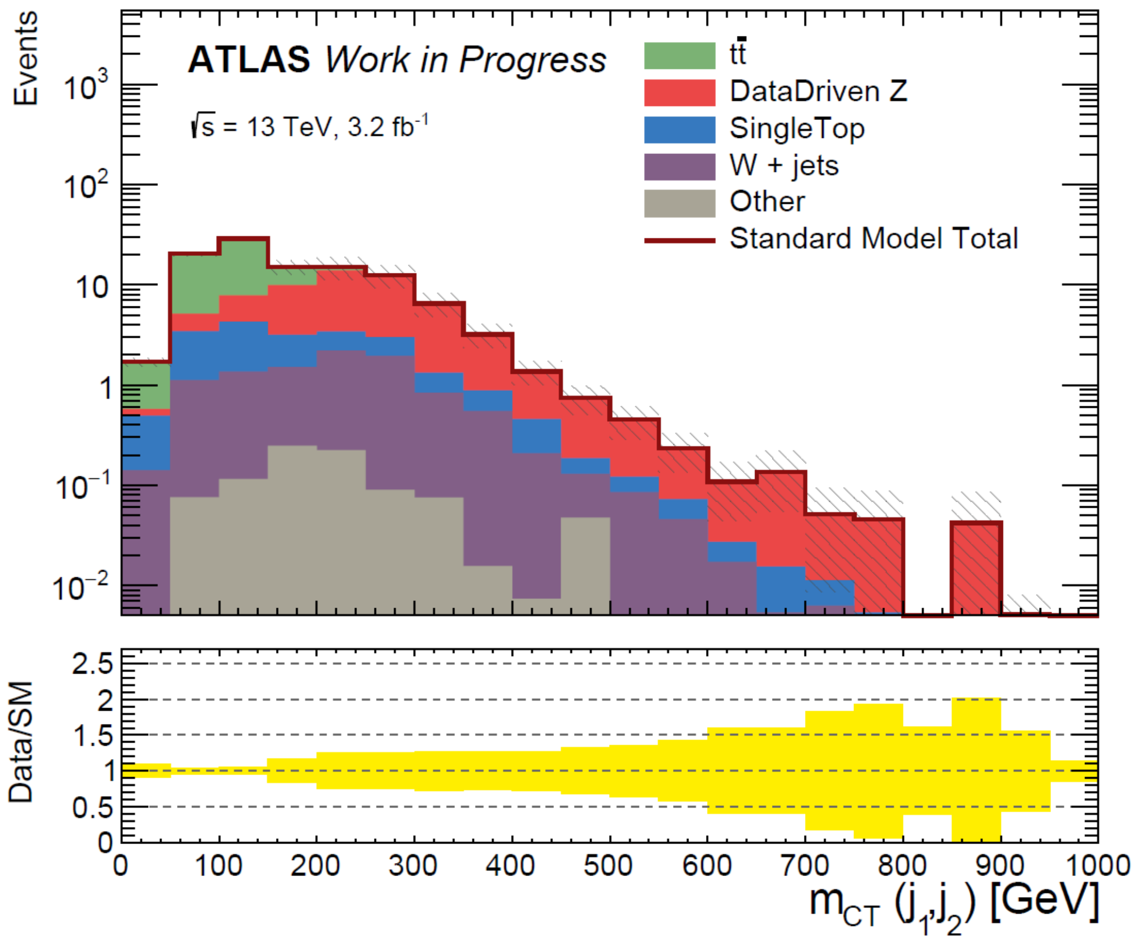


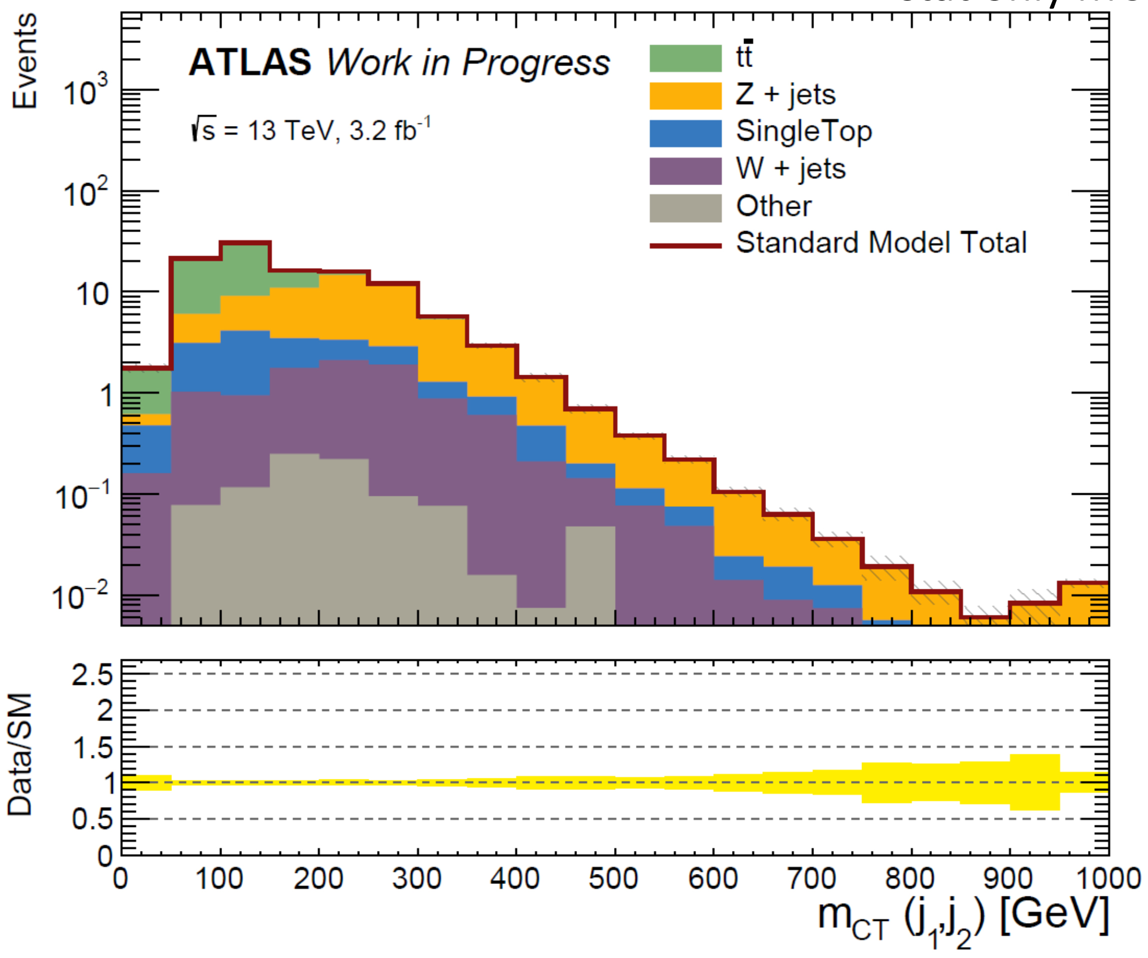
$$m^2_{\rm CT}(v1,\ v2) = [E_{\rm T}(v1) + E_{\rm T}(v2)]^2 - [\mathbf{p}_{\rm T}(v1) - \mathbf{p}_{\rm T}(v2)]^2$$

$Z \to \nu\nu$ + b-jets from $Z \to ll$ + light-jets (2)

- Full method in the SRs/VRs vs Z+b-jets from pure MC: m_{CT}

*stat only MC

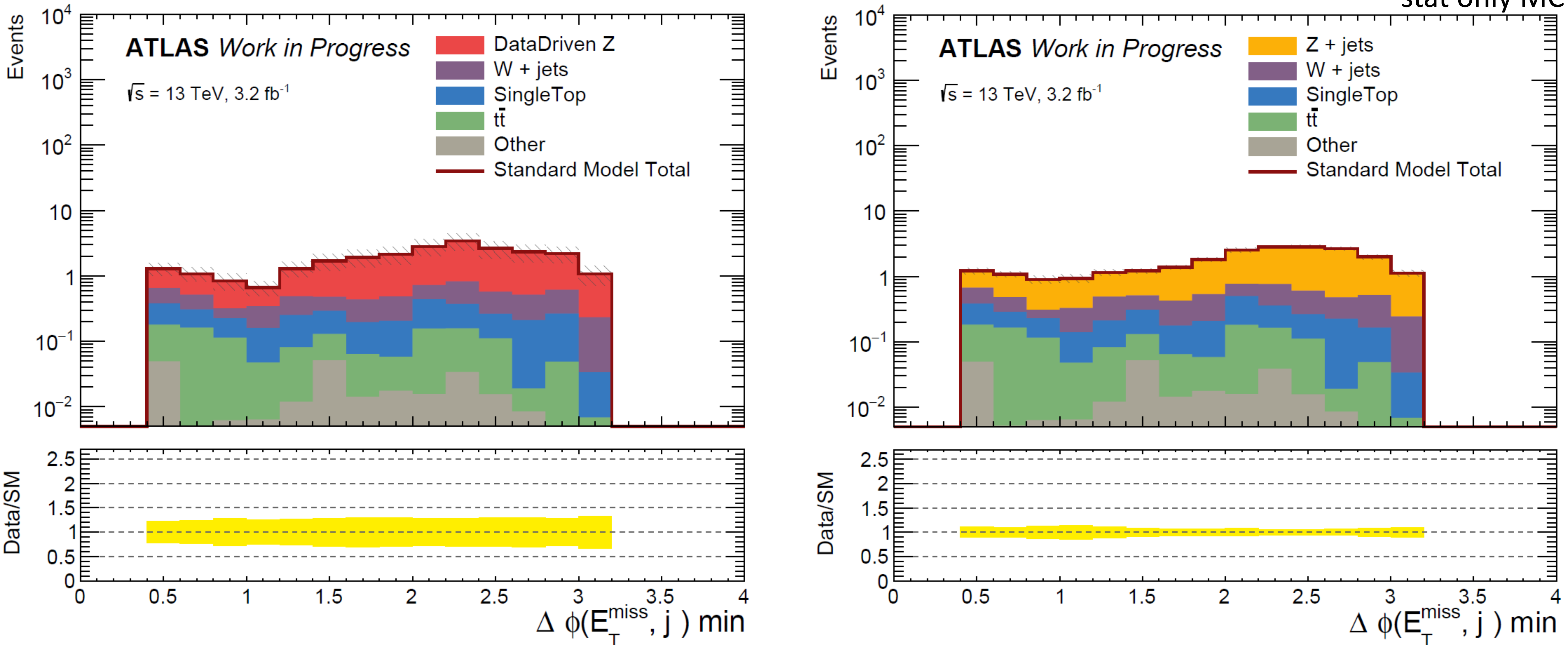




$Z \rightarrow \nu\nu$ + b-jets from $Z \rightarrow ll$ + light-jets (4)

- Full method in SRA250 vs Z+b-jets from pure MC: $\min \Delta\varphi(E_T^{miss}, jets)$

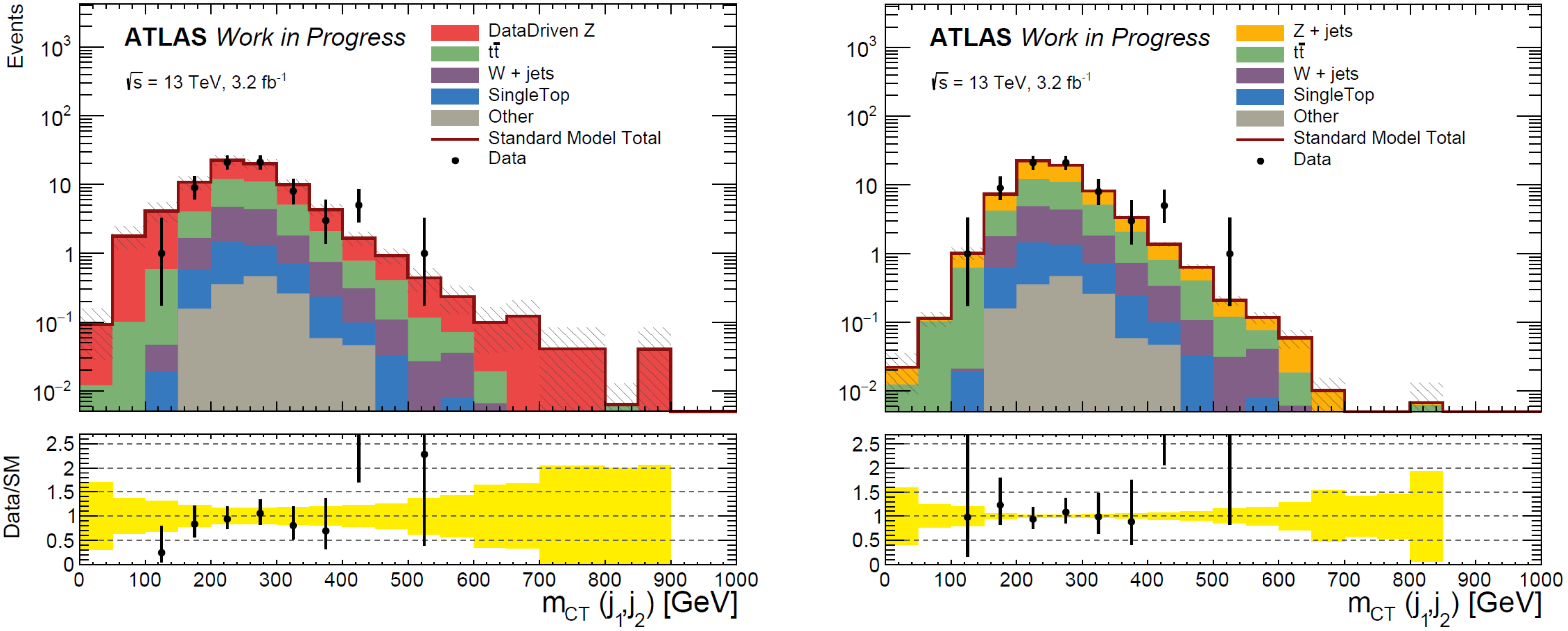
*stat only MC



$Z \to \nu\nu$ + b-jets from $Z \to ll$ + light-jets (5)

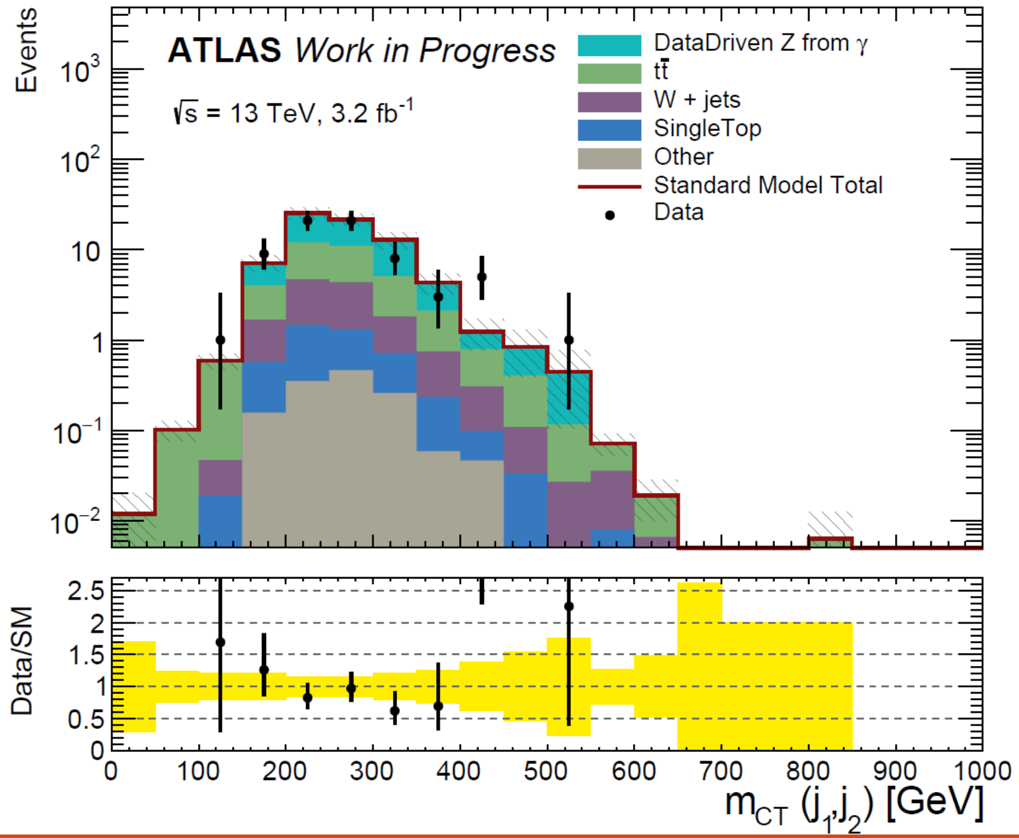
- Full method in VRA vs Z+b-jets from pure MC: $\min \Delta\varphi(E_T^{miss}, jets)$

*stat only MC

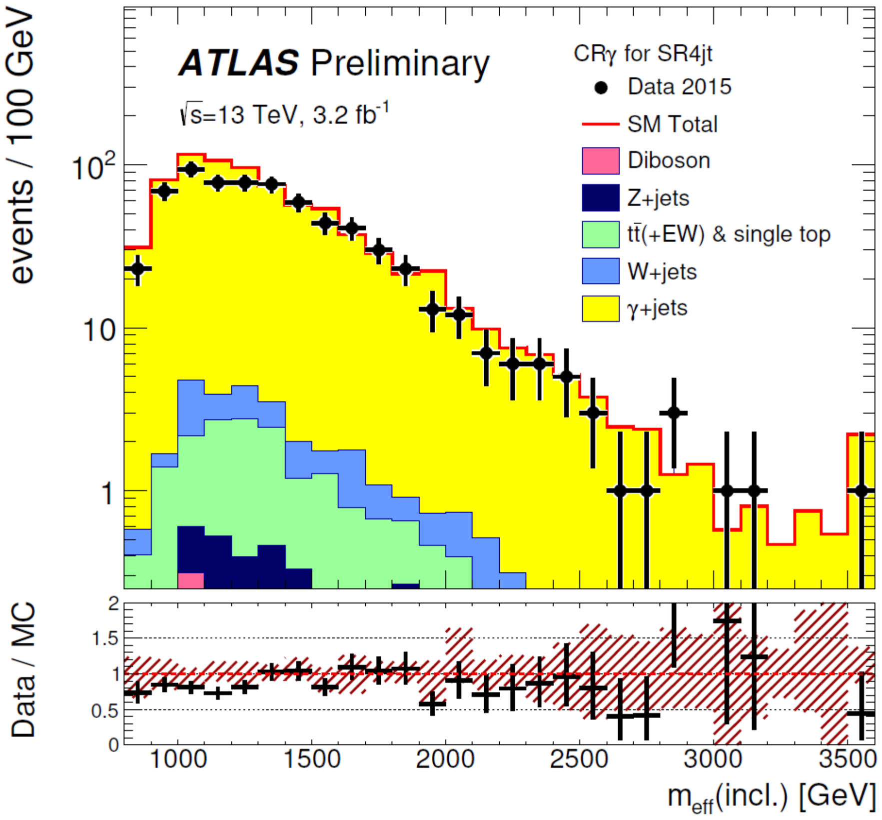


$Z \rightarrow \nu\nu$ + jets from $Z \rightarrow \gamma$ + jets (1)

- Used by sbottom 0L – specifically requiring b-jets
- Also used by the Strong 0L, RJigSaw, Z+met analyses for estimation of Z+jets background from γ + jets



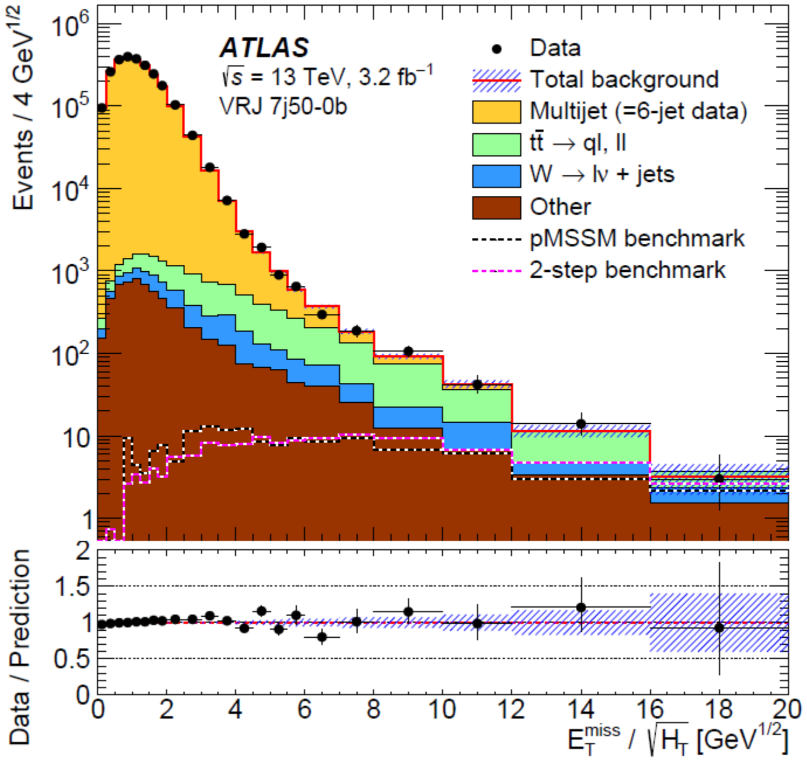
- Strong 0L analysis, EOYE:
  - ATLAS-CONF-2015-062 https://cds.cern.ch/record/2114828/files/ATLAS-CONF-2015-062.pdf



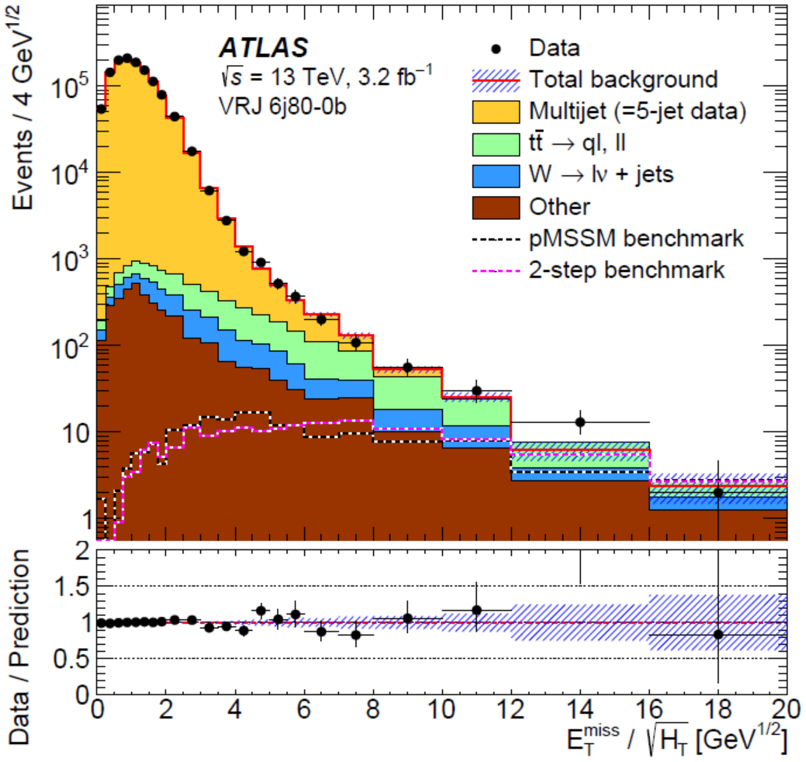
Multi-jet (QCD) Estimation from Template Fit

- http://arxiv.org/abs/1602.06194
- More analysis details in IOP talk by Will Fawcett (Oxford)
- Relies on $E_T^{miss}/\sqrt{H_T}$ being invariant with jet multiplicity (7 jets vs 6 jets)→
- Shape of the $E_T^{miss}/\sqrt{H_T}$ distribution is measured in CRs with lower jet multiplicity than the SRs
- Then normalised in a second CR with the same jet multiplicity as the SR but with $E_T^{miss}/\sqrt{H_T}$ < 1.5

7 jets, Multijet from 6 jets



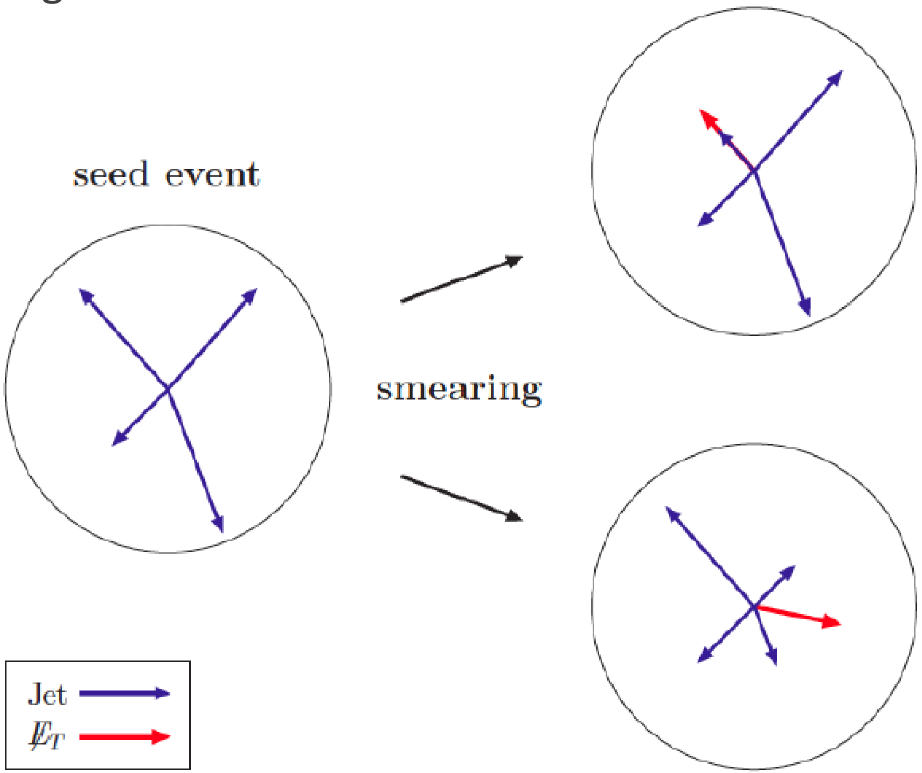
6 jets, Multijet from 5 jets



- Gives estimate in SR with $E_T^{miss}/\sqrt{H_T}$ >4

JetSmearing – QCD background estimation (1)

- 'JetSmearing' technique used to estimate QCD background from Multi-jet fluctuations
- Used by various groups for QCD estimate
  - Also used by Z+met analysis for Z+jets background

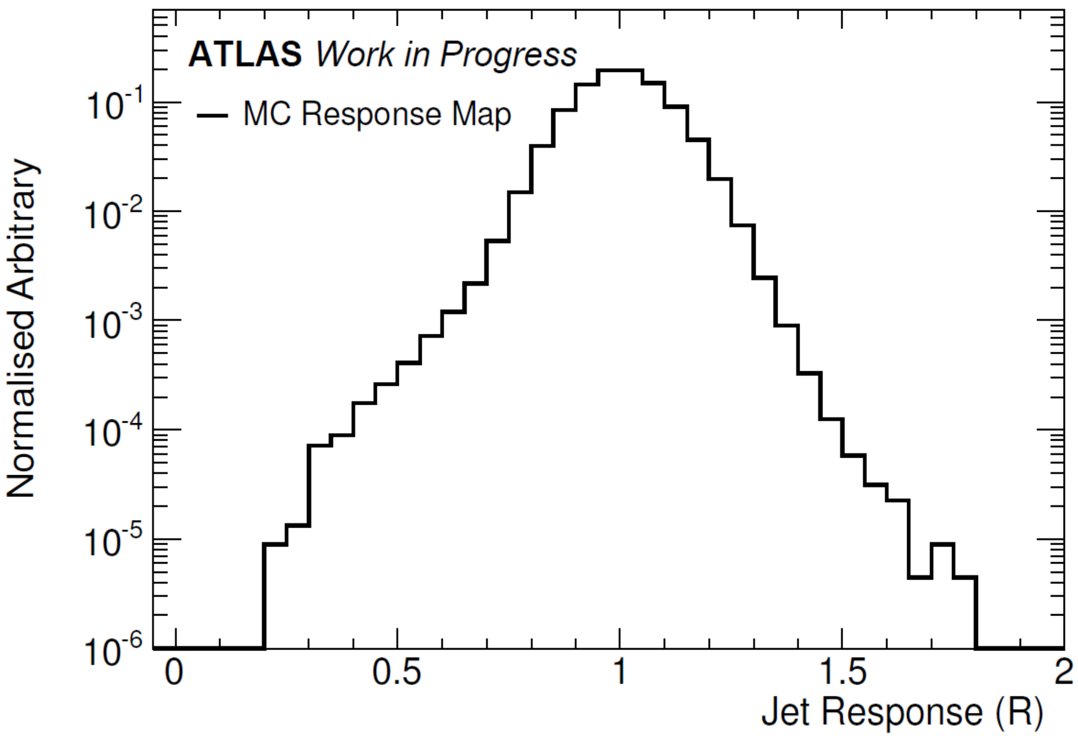


- Jets can be mis-measured for many different reasons (jet response)
- This leads to up and down fluctuations of the jet pT giving rise to potentially large missing energy
- Response (R) is measured in di-jet MC and corrected in data (di-jet balance analysis)
- Idea of jet smearing:
  - take well measured jet events and "smear" the jets by altering their LorentzVectors based on response maps (R)

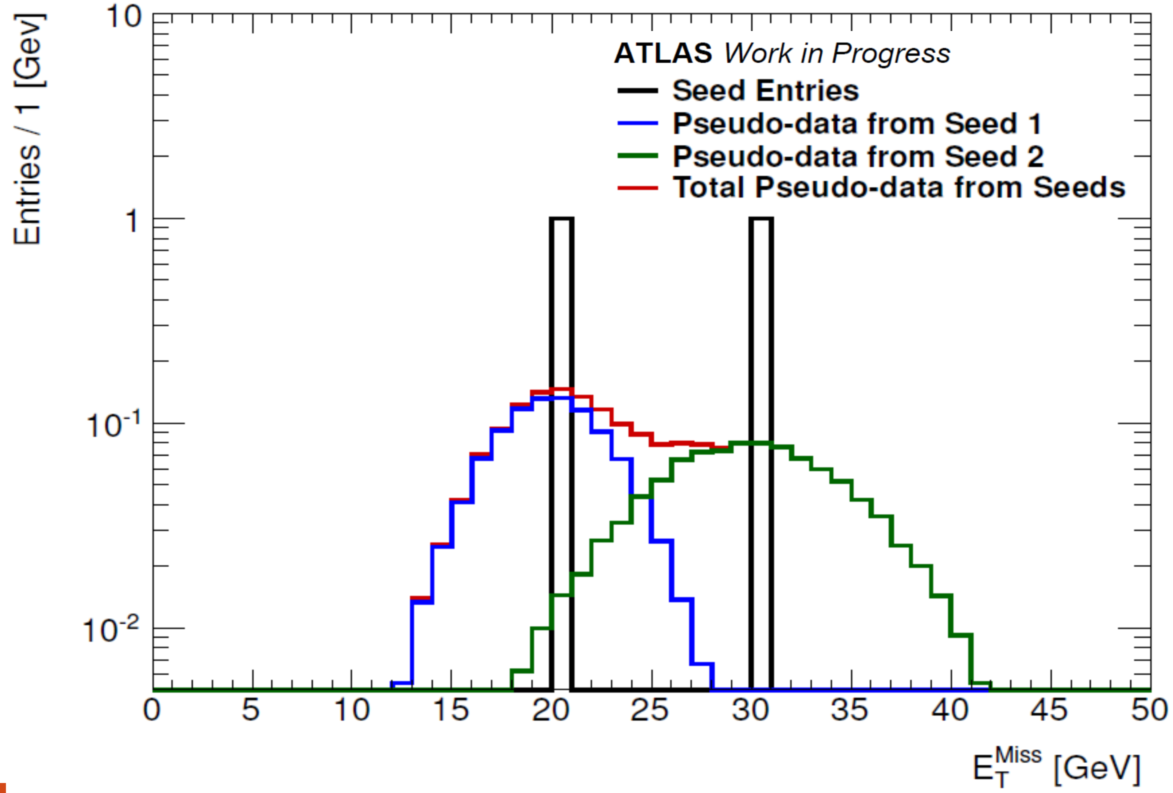
JetSmearing – QCD background estimation (2)

- Response map define as:
  - $R = E_T^{RECO} \ / \ E_T^{TRUTH}$
  - Separately for b-veto and b-tagged jets

Truth jet p_T range of 100 – 120 GeV

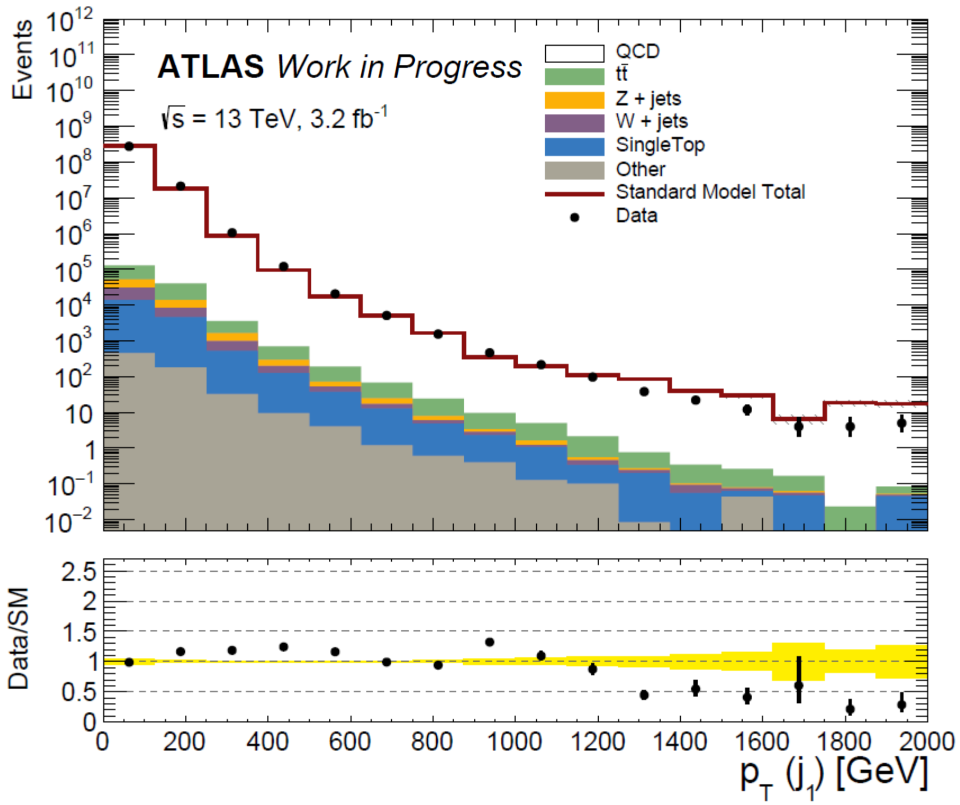


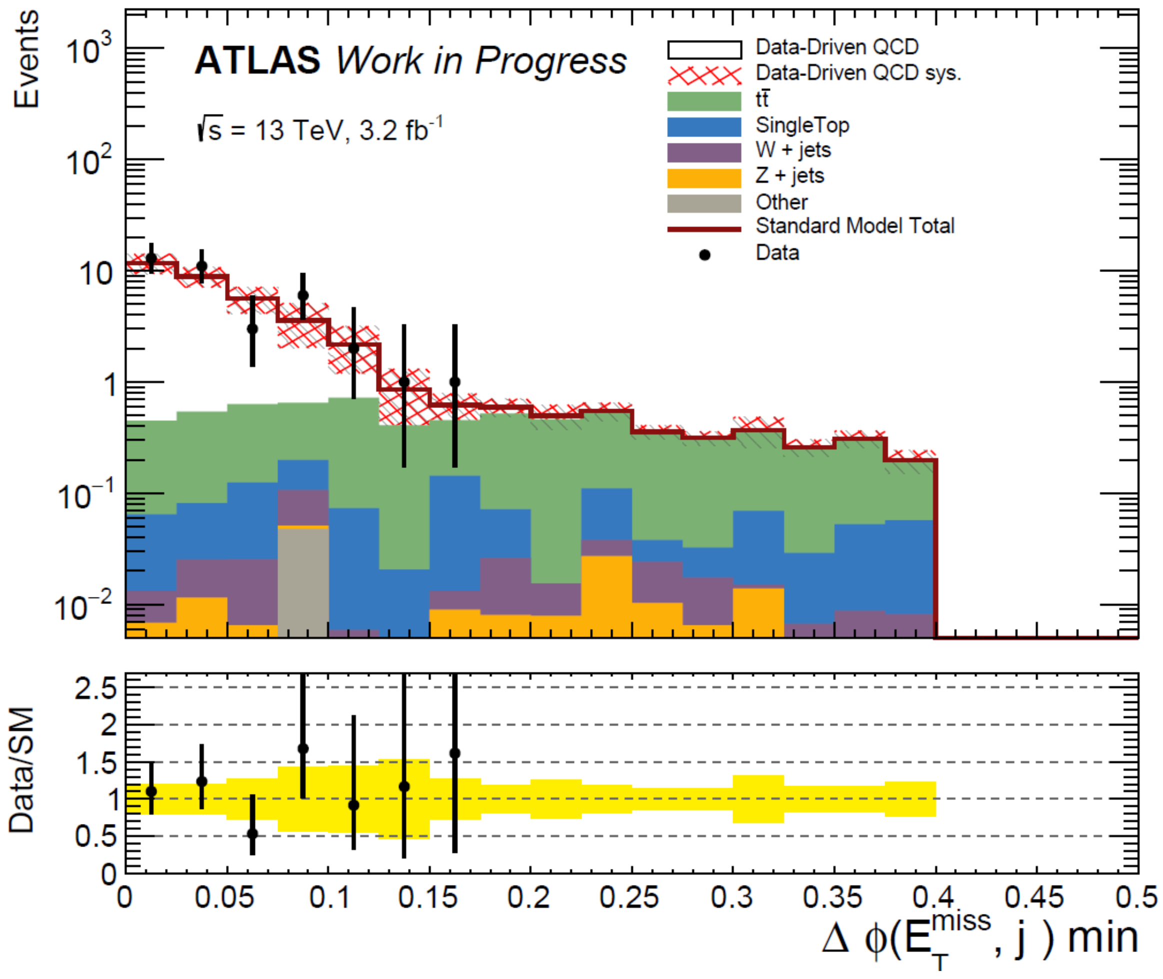
- Seed data events should, by definition of the MET significance cut ($E_T^{miss}/\sqrt{E_T} < 0.7$), be well measured:
  - R~1 , low MET



JetSmearing – QCD background estimation (3)

- After pseudo-data from smearing well measured jet events is created (left), this can be normalised in a high MET Control region (right) to give an estimation in the SR





Summary

- Many data-driven techniques for estimating SM backgrounds are used in ATLAS searches for SUSY
- This talk has focused on ways of estimating Z+jet backgrounds from an extrapolation over the b-jet multiplicity (ZfromLight) and using photon+jet data (ZfromGamma)
- The template fit method used by the Multi-jet analysis has been discussed
- A short overview was given of the JetSmearing technique which is used to estimate large MET backgrounds arising from the mis-measurement of multiple jets

Background	Method(s)	Run-II analyses usage
Many e.g. Top production, W+jets, Z+jets	Monte-Carlo based estimation, normalised in CRs (control regions)	Majority
Lost Lepton(s)	Lepton replacement method	Not yet used in Run-II
Tau(s)	Tau replacement method	Not yet used in Run-II
QCD/Multi-jet fluctuations	**JetSmearing**	**Sbottom, stop 0L, strong 0L, Monojet**
	Template / ABCD method	**Multi-jet (0-7 jets)**
Z+jets	**ZfromGamma**	**Sbottom, strong 0L, RJigSaw, 2 Lepton Z + MET**
	ZfromLightJets	**Sbottom**
Z+met	**JetSmearing**	**2 Lepton Z + MET**