ATLAS NOTE

ATL-COM-PHYS-2016-XXX

14th June 2016

Draft version 0.1

Monte-Carlo Modelling of the Z^0 Boson in association with jets as a background to Standard Model Higgs Searches with Taus

Vincent Croft$^{\diamond}$ and Théo Jean Megy*

$^{\diamond}$*Radboud University Nijmegen/Nikhef*
**Albert-Ludwigs-Universitaet Freiburg*

Abstract

An assessment of the modelling of the Z→ $\tau\tau$ component of the background to searches for the Standard Model Higgs Boson decay into taus with simulated Monte-Carlo events and the associated errors using the parametrisation method. Performance is assessed using Run2 $Z \to \mu\mu$ data and propagated into several channels.

© 2016 CERN for the benefit of the ATLAS Collaboration.
Reproduction of this article or parts of it is allowed as specified in the CC-BY-3.0 license.

Contents

1 Introduction

The search of the Higgs boson is one of the primary goals of the ATLAS detector. Collisions at center of mass energies of $\sqrt{s}$=7 and 8TeV yielded evidence for several decay modes of the Higgs and a combination between the ATLAS and CMS experimental results claims evidence for a new particle compatible with the Standard Model Higgs Boson decaying into taus. The first collisions at $\sqrt{s}$ = 13TeV lead to the possibility of strengthening this evidence with a 'pure' result derived from each detector independently. The analysis laid out by the ATLAS group features a 'cut based' analysis whereby the final state topology of the Higgs is exploited to enhance any new signals against their backgrounds.

The primary irreducible background to Higgs Searches with Taus is the decay of the Z^0 boson into a tau anti-tau pair. The analysis targets kinematic variables such as the resonant mass of the tau pair to try to remove as many Z type events whilst retaining as many Higgs type events as possible. This means that the remaining Z→ $\tau\tau$ events are in a highly contrived region of phase-space where the theoretical prediction of the events is highly important to the sensitivity of the analysis.

In view of the complexity of the relevant event properties, in the past the ATLAS collaboration endeavoured to rely as little as possible on simulation. However Z→ $\tau\tau$ model cannot be obtained directly from the collision data due to background contributions, e.g. from events with other objects misidentified as tau decays. Events with two muons can be 'embedded' with simulated tau decays such that kinematic quantities can be preserved. However such a process requires a large Z→ $\mu\mu$ data set and extensive validation.

In practice, particle physics analyses use Monte-Carlo (MC) generators to compare predictions from theory to data. An extensive system of simulation and reconstruction mimics the effects of the detector such that theoretical models can be compared directly with physics objects in data. A full description of all relevant processes in simulated MC is considered by many to be the only way that a process in the ATLAS detector can be observed. Moreover, it allows us to produce new events for testing and refining our analysis regardless of the performance of the LHC and ATLAS detector. As such the production of high statistic, high prescision MC modeling of the Z→ $\tau\tau$ process is key to the search for Higgs Bosons in the first $\sqrt{s}$ = 13TeV data with the ATLAS detector.

The Z→ $\tau\tau$ process in MC is highly complex. Due to the properties of the Z^0 and the tau almost all observable quantities are correlated. To reduce Z and DY contribution to the higgs analysis the signal regions either have a high transverse mass (possibly with additional jets) or explicity have at least 2 additional jets. The production of these additional jets requires a very large number of additional QCD and EW production modes must be calculated.

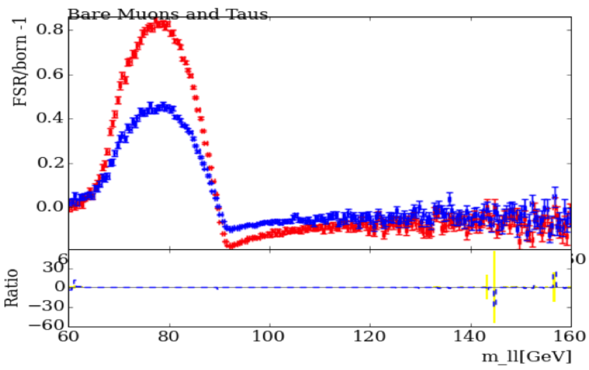
This note describes the generators used for producing simulated Z+jets events as used in the ATLAS $H \to \tau\tau$ analysis and demonstrates its performance using the first X data at $\sqrt{s}$ = 13TeV as collected by the ATLAS detector. 3 possible channels are considered here. In the case that both taus decay leptonically the Z→ $\tau\tau$ component is enhanced when one tau decays into an electron and one into a muon. In the case where both leptons measured have the same flavour the Z→ $\tau\tau$ component is surpressed and the Z→ $\ell\ell$ component can be used to cleanly asses the MC performance of the Jet activity associated with these events. The case where one tau decays hadronically and one leptonically is predicted to give the greatest quantity of Z→ $\tau\tau$ events and is therefore also considered.

2 Simulated Samples

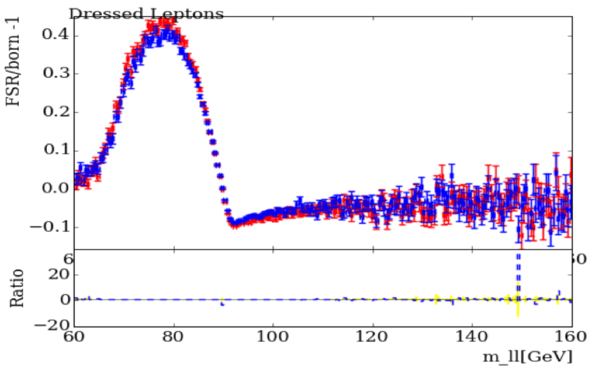
Monte Carlo (MC) simulations, normalised to the results of the highest order calculations available, are used in the following to compare data to Z + jets predictions and to estimate the contribution from background events. Signal events, containing a Z boson with associated jets, were simulated using the Sherpa v2.2.1 generator. Matrix elements were calculated for up to two partons at NLO and up to four additional partons at LO using the Comix and OpenLoops matrix element generators and merged with the Sherpa parton shower using the ME+PS@NLO prescription. The CT10 PDF set was used. Simulated samples of Z+jets production were also produced with the MadGraph5_aMC@NLO v2.2.2 generator using explicit matrix elements for up to four partons at leading order, interfaced to the Pythia v8.186 parton shower model. The A14 parton shower tune was used together with the NNPDF23LO PDF set. The EvtGen v1.2.0 program was used for properties of the bottom and charm hadron decays. The Powheg-Box v2 simulation program, interfaced with the Pythia v8.186 parton shower was also considered.

The Sherpa v2.2.1 and MadGraph5_aMC@NLO v2.2.2 generators are usually favoured over Powheg-Box as they are expected to better model the emission of additional partons. Usage of Sherpa v2.2.1 is strongly advised since the intrinsic parton showering allows for the calculation of up to 2 additional Jets at NLO in addition to the hard process. This formal accuracy is at the cost of additional computing time and as such MadGraph5_aMC@NLO is often preferred.

All generated events are then treated with a full simulation of the ATLAS detector and subsequently physics objects are reconstructed in the same manner as data is. Radiative emissions from muon and taus decays for example are handled by the The PHOTONS++ module within Sherpa. This holds routines to add QED radiation to tau-lepton decays. This has been achieved by an implementation of the YFS algorithm, structured in a way such that the formalism can be extended to scattering processes and to a systematic improvement to higher orders of perturbation theory. The application of PHOTONS++ therefore accounts for corrections that usually are added by the application of PHOTOS to the final state. Figure 1a shows that photon emissions are carefully considered within a MC generator. Figure 1b shows that if detector effects such as isolation and calorimeter calibration are taken into effect. This particular example is covered in detail elsewhere [embedding paper] as the differences between data and simulation need to be well understood especially when comparing the properties of Z→ $\tau\tau$ and Z→ $\mu\mu$ decays.



(a)



(b)

Figure 1: Simulation of Z decays into taus and muons as a ratio of the total cross-section. Plotted is the difference between the masses of the Z boson when radiative corrections are applied. Since the muon is lighter it radiates more energy as seen on the left. When this energy is clustered in a tight calorimeter cone around the lepton; as in the plot on the right, the differences between the two distributions are reconciled

3 Scale Factor Comparison Method

There are multiple montecarlo generators available for producing samples for use with ATLAS analyses. Since each generator has its own strengths and weaknesses the differences between them can be used to estimate the uncertainty associated with any mis-modelling produced by the generator.

As already mentioned the signal region in a higgs analysis is built from a series of cuts designed to reduce the contribution from processes such as $Z \rightarrow \tau\tau$ Typically this leads to large variations in the Z+jets differential distributions. The statistical variations can be controlled by including a Z enhanced control region into the likelihood fit. The data MC agreement can be then quantified by use of a normalisation factor.

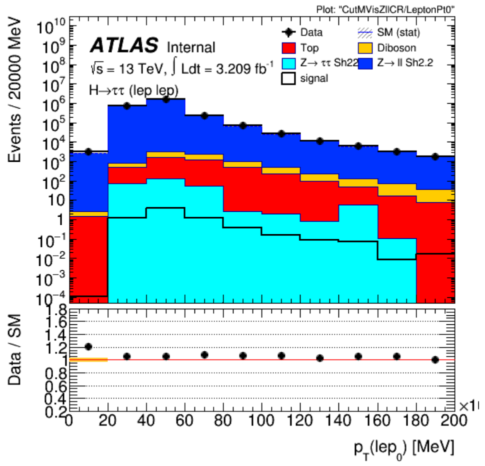
Figure [2] shows the distribution of $Z \rightarrow \ell\ell$ in an enhanced Z region. This is defined as being a window in the visible mass of 10GeV around the Z rest mass. Each cut gives slightly different agreement between data and mc section 6 goes into more detail on this. Even in this 'clean' Z region there is not a perfect agreement between data and MC. As such we can propagate this 'known' difference between data and MC from the control region to the signal region through the use of a Normalisation Factor (NF though also known as a Scale Factor SF) this can be defined as:

$$\mathrm{PN}_{\mathrm{Z}}^{\mathrm{SR}} = \frac{\mathrm{N}_{\mathrm{Z,MC}}^{\mathrm{SR}}}{\mathrm{N}_{\mathrm{Z,MC}}^{\mathrm{SR}}} \cdot (\mathrm{N}_{\mathrm{data}}^{\mathrm{CR}} - \mathrm{N}_{\mathrm{nonZ}}^{\mathrm{CR}}) = \mathrm{TF}_{\mathrm{Z}} \cdot (\mathrm{N}_{\mathrm{data}}^{\mathrm{CR}} - \mathrm{N}_{\mathrm{nonZ}}^{\mathrm{CR}}) \tag{1}$$

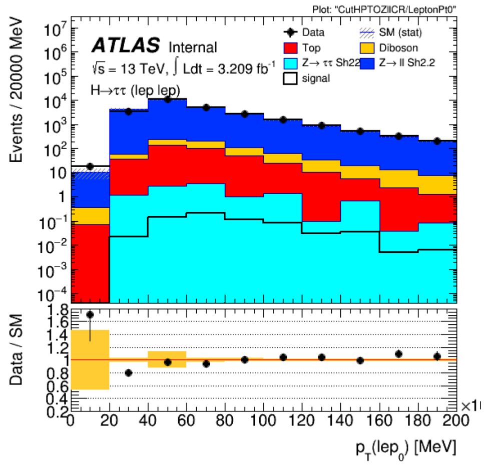
where $\mathrm{N}_{\mathrm{Z,MC}}^{\mathrm{CR}}$, $\mathrm{N}_{\mathrm{Z,MC}}^{\mathrm{CR}}$ etc are the expected number of Z/γ events in signal and control regions. The difference in the transfer factor as calculated for difference generators can be taken as the generator modelling error. This can be calculated for any definition of the Z control region. Table 1 shows the associated uncertainties taken from using Sherpa as the nominal generator for the $Z \rightarrow \ell\ell$ process.

	SR$^{\text{(Preselection)}}$	M_{Vis}	$\mathrm{E}_{\mathrm{T}}^{miss}$	X_1,X_2	$\Delta\phi_{\ell\ell}$
Sherpa vs PoPy	-58.27	-58.21	-43.78	-38.07	-28.10
Sherpa vs mg	-34.20	-29.61	-23.11	-16.99	-14.69
TOTAL	±67.57	±65.31	±49.50	±41.69	±31.71

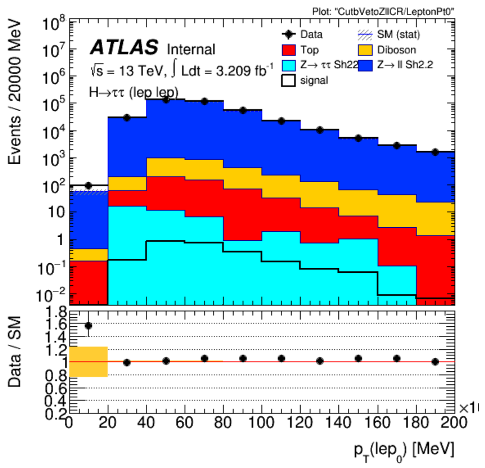
Table 1: Percentage uncertainties arising from the difference in yields as predicted by different MC generators. The SR column represents the 'raw' difference in generators. All others are calculated from the difference in normalisation factors



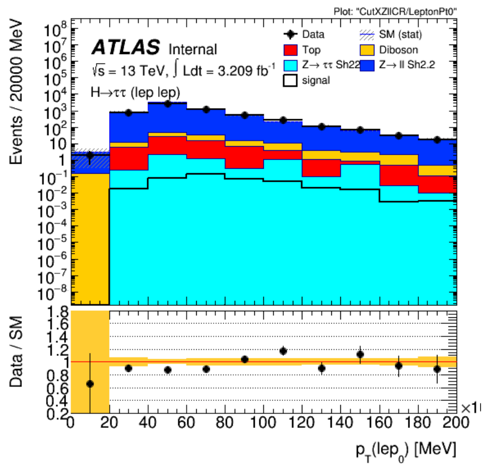
(a)



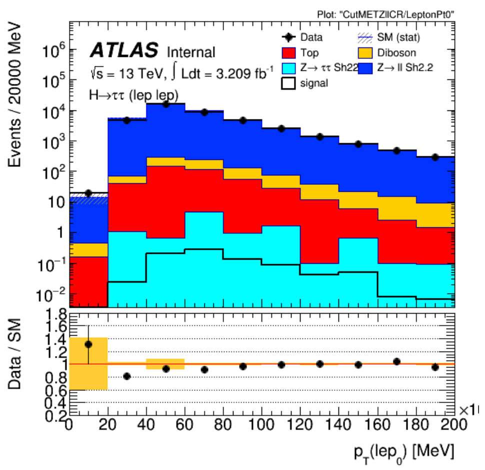
(b)



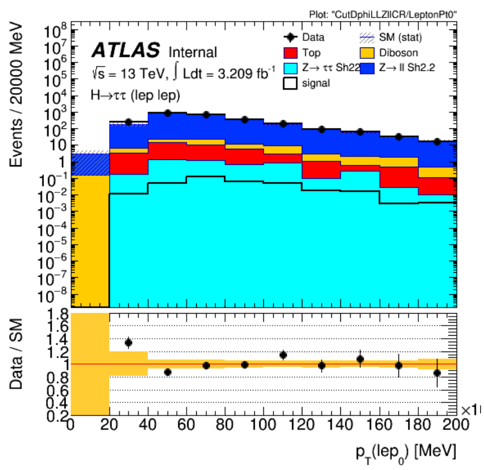
(c)



(d)

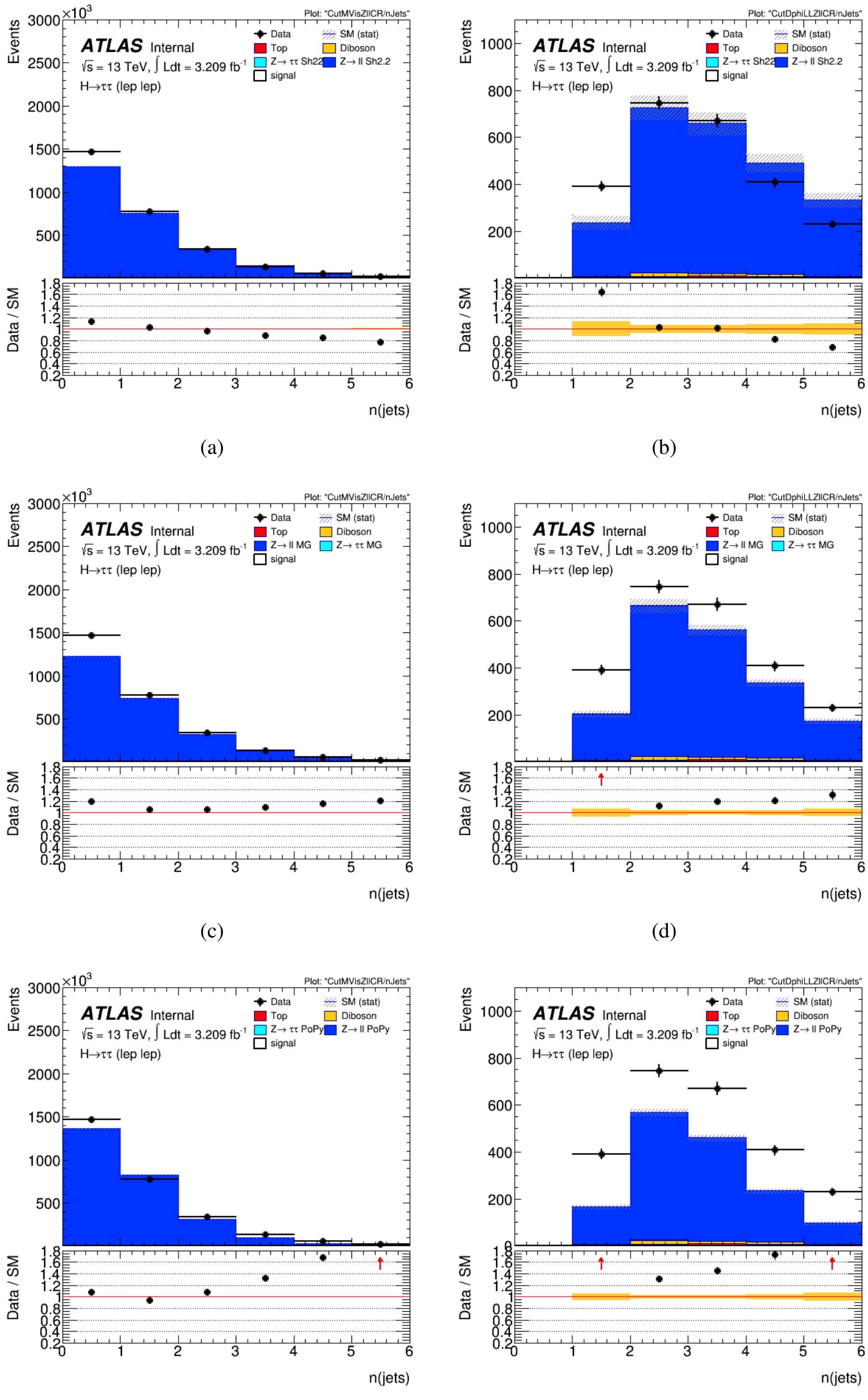


(e)



(f)

Figure 2: Distibutions in Nominal Z→ $\ell\ell$ Control region



(a) (b)

(c) (d)

(e) (f)

Figure 3: Jet Multiplicity in Base and Contrained Z→ $\ell\ell$ Control regions

4 Generator Systematics

Sherpa v2.2.1 and MadGraph5_aMC@NLO are able to produce scale variations to account for errors in modelling the V+jets process.

A global 5% uncertainty should be assigned on the total W/Z inclusive cross section. The prescription to estimate the uncertainties on the shapes requires the usage of alternative samples with the following variations:

- Renormalization scale variations: x 2 and x 1/2
- Factorization scale variations: x 2 and x 1/2
- Resummation scale variations: x 2 and x 1/2
- CKKW matching scale variations: nominal 20 GeV, variations setting it at 15 GeV and 30 GeV

A similar prescription holds for MadGraph5_aMC@NLO; a global 5% uncertainty should be assigned on the total W/Z inclusive cross section and the relevant parameters to be varied are:

- scalefact: value used for the variation of the factorization and renormalization scale: x 2 and 1/2
- kTdurham (in MG and Py8): nominal 30 GeV - variations could be 20 GeV and 50 GeV

The samples are produced for Sherpa v2.1.1 at EVGEN only, so truth codes must be used to estimate the uncertainties with respect to the truth-nominal. Each set of samples is normalized to the same cross section (to avoid double counting). The variations should be evaluated independently and added in quadrature.

Because of the large statistics of the samples (364M per lepton flavor for W, Z and Z to neutrinos), only 3 sets have been produced: Z->nunu+jets, W->enu+jets, Z->ee+jets. K-factors and cross-sections relevant for V+jets samples are collected centrally.

The recommendation for the uncertainty estimate for the V+jets samples is to take half the difference between the up and down variation (relative to the midway point). Individual contributions should be added up in quadrature for the various sources of scale uncertainty considered. The relative uncertainty can be directly applied to the Sherpa v2.1.1 nominal prediction as well as the Sherpa v2.2.1 nominal prediction (the formal accuracy being identical).

Because the Sherpa 2.1 nominal prediction needs a smoothing correction, the systematic variations cannot be compared to the nominal. Instead we recommend to evaluate the systematic uncertainties with respect to the midway point between the up and down variation, thereby symmetrizing the uncertainty.

5 Reweighting Method

Monte Carlo simulations of physics processes are very heavily controlled in ATLAS. Each theoretical distribution must be validated, propagated through a detailed simulation of the ATLAS detector and unfolded through an digital approximation of instrumentation effects and reconstructed back in to physics objects in the same manner as the data. This leads to the production of MC samples being extremely costly in terms of computing resources.

Each sample produced is filtered by quark flavour and sliced in terms of the associated Boson Pt such the statistical accuracy of the samples can be enhanced. This means for a single physics process there is usually upwards of 20 samples required. If each generator variation would be simulated, more than 100 samples per process would be needed which is considered computationally impractical.

Truth level samples were used to produce a 2 dimensional parametrisation of the variations [ref SUSY tool note] This allows the assessment of generator effects on differential distributions whilst only requiring a single file containing the parametrised weights.

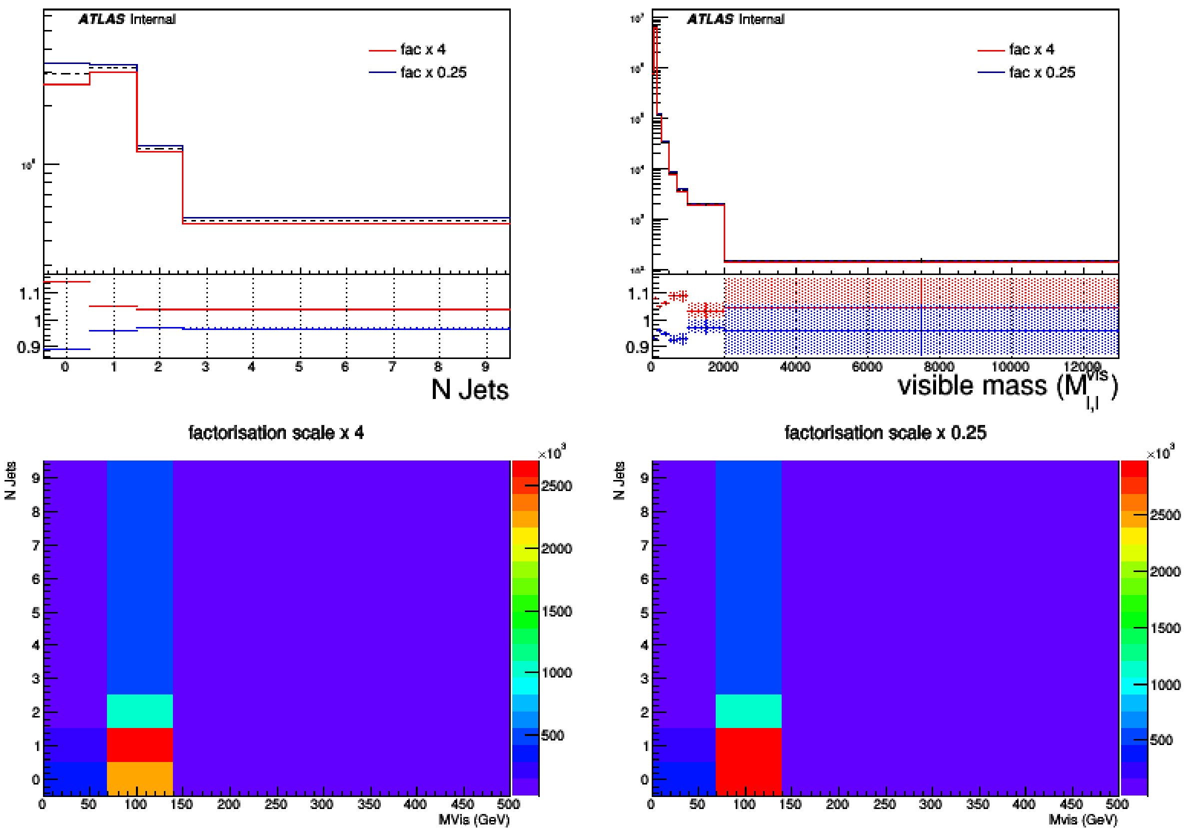
Preliminary investigations were performed by producing weights based on a 1-dimensional parameterisation based on either pT(Z) or njets. For the case where pT(Z) was used for the parameterisation process, poor modelling is found for the properties related to the jets (HT,pT(j1)etc),whilst using the jet multiplicity as the basis for the parameterisation leads to poor modelling ofthe ETmiss. Due to this a 2D parameterisation was performed using both pT(Z) and njets. [ref SUSY tool note] given the goal of using this tool to assess variations in the Z→ $\tau\tau$ process, instead of pT(Z) the Visible Mass (mvis) was considered. This is a value more commonly used in cases where tau neutrinos are liable to produce a truth contribution to Et miss and therefore offset the reconstructed mass of the parent boson.

To cohere with the prior studies and to coincide with the Pt(Z) slicing of the samples the mvis paraemetrisation is produced in bins of [0,70], [70-140], [280-500], [500-700], [700-100], [1000-2000], [2000-ECMS] GeV and the jet multiplicity is binned in terms of 0,1,2,>2 jets. This averages out large differences that may occur in extremely high jet multiplicity regions (that this analysis is blind to) and allows for low statistical error in every bin.

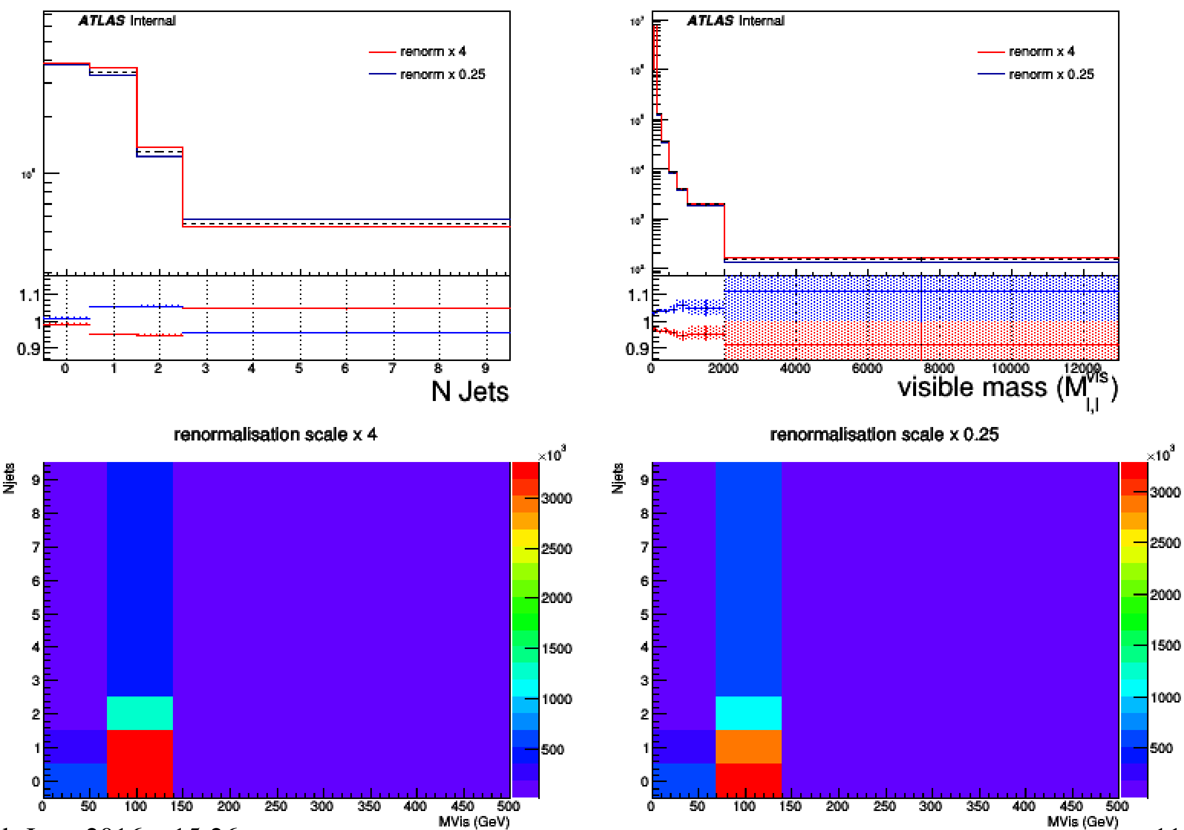
For a given mvis bin (i), and njets bin (j), the weights are calculated per sample (up and down variations are treated separately) using:

$$W_{i,j} = \frac{N_{i,j}^{Syst}}{N_{i,j}^{Nominal}} W_{i,j} = \frac{\sum_{m_{\mathrm{T},flavour}} \sigma^{Syst} \cdot k \cdot \epsilon \cdot N_{i,j}^{Syst,RawNo}}{\sum_{p_{\mathrm{T},flavour}} \sigma^{Nominal} \cdot k \cdot \epsilon \cdot N_{i,j}^{Nominal,RawNo}}$$

The largest variations expected are due to factorisation and renormalisation scales these can be seen in figure [4] and the corresponding weights to be applied to the reconstructed nominal samples at truth level can be seen in figure [5]

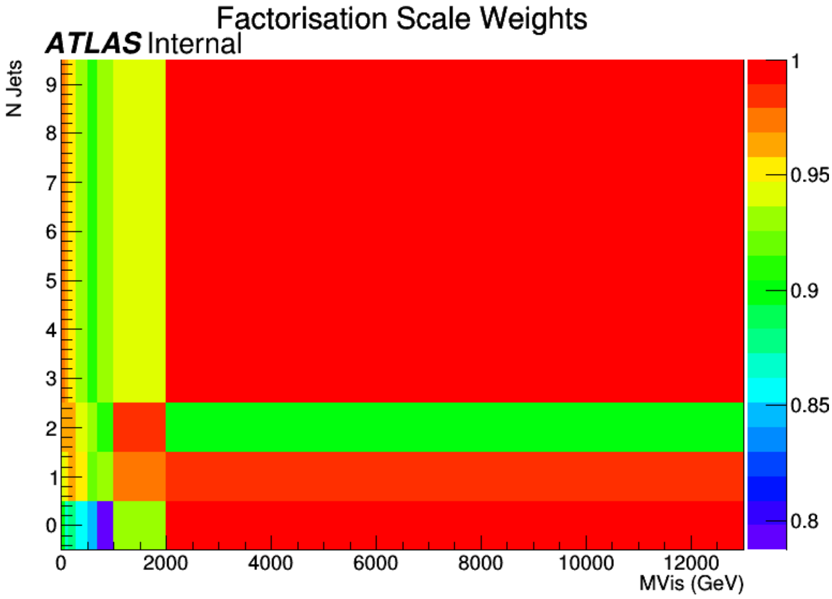


(a)

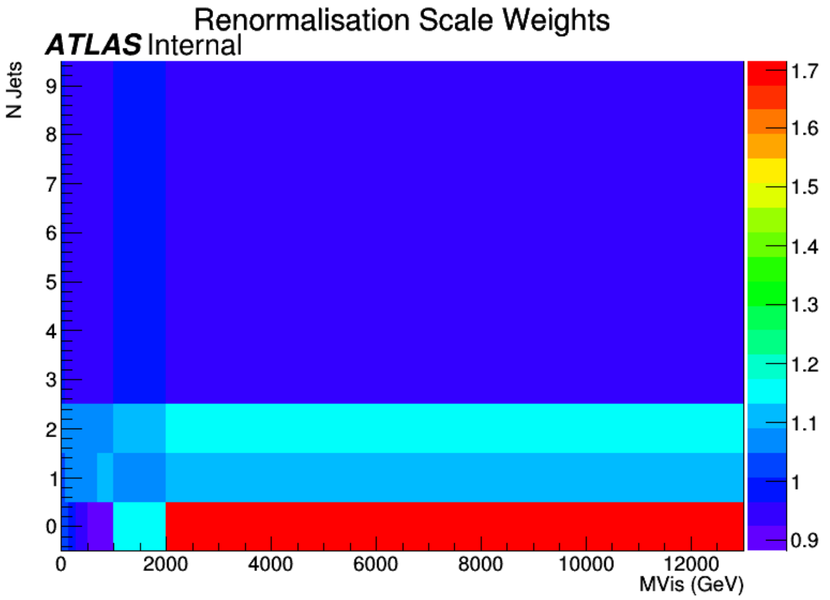


(b)

Figure 4: Two dimensional parametrisation of variations due to changes in (a) factorisation and (b) renomalisation scale. The error can be taken as half the difference between up and down variations



(a)



(b)

Figure 5: Weights produced to emulate changes in (a)facorisation and (b) renomalisation scales. The weight should be positive for down variations and positive for up

6 $Z \to \ell\ell$ Modelling

6.1 Z Region Definition

Definition here

6.2 Data MC Comparison Between Generators

Plots here

7 Modelling events with Hadronic taus

Section describing ztautau performance Using lephad ZCR

7.1 Z$\to$ $\mu\mu$ comparison with $\to$ e^+e^-

7.2 Propagation to Z$\to$ $\tau\tau$

8 Validation of MG5aMC@NLO+pythia8 variations

Validation of generator variations produced for Madgraph5 aMC@NLO+pythia8 using Zjj data produced at $\sqrt{s} = 8$TeV.

In progress.