ATLAS NOTE

ATL-COM-PHYS-2016-XXX

14th July 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[◇], Théo Jean Megy[★], Elias Coniavitis[★], Julian Glatzer[♣], Daniele Zanzi[◇]

[◇]*Radboud University Nijmegen/Nikhef*
[★]*Albert-Ludwigs-Universitaet Freiburg*
[♣]*CERN*
[◇]*University of Melbourne*

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. Performance assessed via a comparison of different generators and verified using using the parametrisation method. Performance is assessed using Run2 $Z \rightarrow \mu\mu$ data and propagated into several channels.

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 8 TeV 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 $\sqrt{s} = 13$ TeV 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 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$\rightarrow \tau\tau$ 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$\rightarrow \tau\tau$ 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$\rightarrow \mu\mu$ 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$\rightarrow \tau\tau$ process is key to the search for Higgs Bosons in the first $\sqrt{s} = 13$TeV data with the ATLAS detector.

The Z$\rightarrow \tau\tau$ process in MC is highly complex. Due to the properties of the Z and the tau almost all observable quantities are correlated. To reduce Z and Drell-Yan (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 the calculation of a very large number of additional QCD and EW production modes.

This section describes the generators used for producing simulated Z+jets events as used in the ATLAS $H \rightarrow \tau\tau$ analysis and demonstrates its performance using the first X data at $\sqrt{s} = 13$TeV as collected by the ATLAS detector. 3 possible channels are considered: Fully leptonic, one leptonic and one hadronic and fully hadronic tau decay di-tau final states. In final states containing hadronic tau decays QCD-jets that 'fake' are always present meaning that a pure Z$\rightarrow \tau\tau$ final state is impossible to construct. In fully leptonic final states where the two leptons have the same flavour (ee,$\mu\mu$) the Z$\rightarrow \tau\tau$ component is suppressed in favour of a Z$\rightarrow \ell\ell$ background component. It can be assumed that the modelling of the Z boson in MC is correct and agnostic to the reconstructed final state within the statistical error of the samples produced. This allows the modelling of the Z$\rightarrow \tau\tau$ background to be assessed via and in conjunction with the Z$\rightarrow \ell\ell$ background.

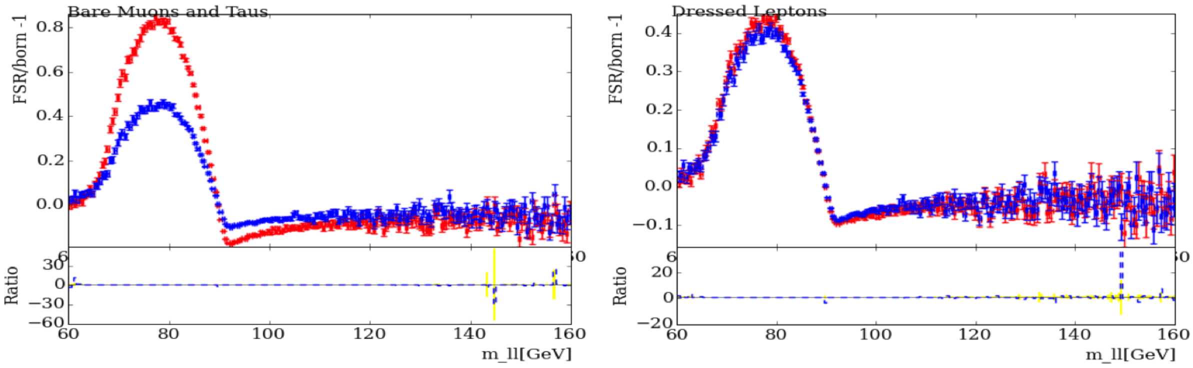
Firstly in subsection [2] the samples considered are described. Studies into the Z lineshape between channels follows in subsection [3] and elucidates the assumptions into the calculation of the systematic error attributed to this background process. The Z control region is described and prefit distributions using $5fb^{-1}$ of data at $\sqrt{s}$ = 13 TeV are shown in section [4]. The normalization and systematic contribution of this background into the analysis fit model and associated Nuisance Parameters (NPs) are described in section [5]. An alternative approach using samples corresponding to theoretical scale variations is documented in section [6]. Finally the contribution of Electroweak (t-channel, VBF-like) Zjj processes and the effects of the Z and Tau polarisations are neglected in this analysis due to the investigations seen in section [7].

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. Processes such as 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 all 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 the correct implementation of photon emissions produced within a MC generator can lead to drastically different physical distributions. Figure 1b shows that if detector effects such as isolation and calorimeter calibration are taken into effect even large theoretical effects can be brought into good agreement with data. This particular example was covered in detail during run1 [embedding paper] and is stated now only to highlight that differences between data and simulation need to be well understood especially when comparing the properties of Z$\rightarrow \tau\tau$ and Z$\rightarrow \mu\mu$ decays.



(a) (b)

Figure 1: Simulation of Z decays into taus and muons as a ratio of the total cross-section. Figure (a) is the difference between the masses of the Z boson when radiative corrections are applied. Since the muon is lighter it radiates more energy in the form of photons and therefore gives a smaller visible mass. Figure (b) shows that this energy is clustered in a tight calorimeter cone around the lepton. When the calorimeter deposits as clustered in a cone of dR = 0.2 around the seed muon are also considered the differences between the two distributions are reconciled

3 Z lineshape and Truth distributions

Distribution of Ptz in different regions in ZCR, ll, lh, hh VBF and Boosted regions. As shown by Julian here

Ratio of ZpT Reco/Truth correlation As shown by Elias here

4 $Z \to \ell\ell$ Modelling

4.1 Z Region Definition

Definition here

4.2 Data MC Comparison Between Generators

Plots here

5 Z Fit Model

Description of Fit Model, including NFs and NPs.

6 Scale Variations from Theory

6.1 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. Unlike in the purely experimental approach taken in comparing generators, the variations produced at Generator level correspond to the major theoretical assumptions underlying the MC generation of events.

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.

6.2 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$\rightarrow\tau\tau$ 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.

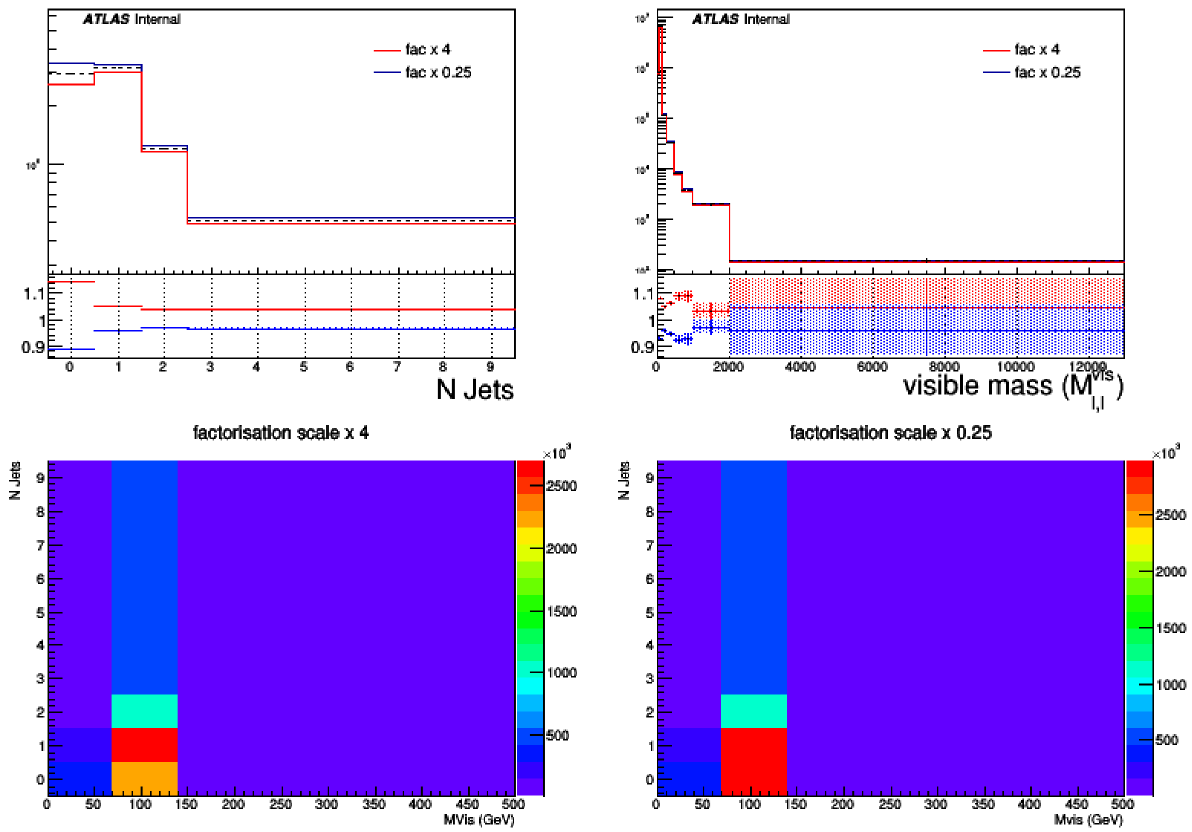
For a given mvis bin (i), and njets bin (j), the weights are calculated per sample (up and down variations are treated separately) using:

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

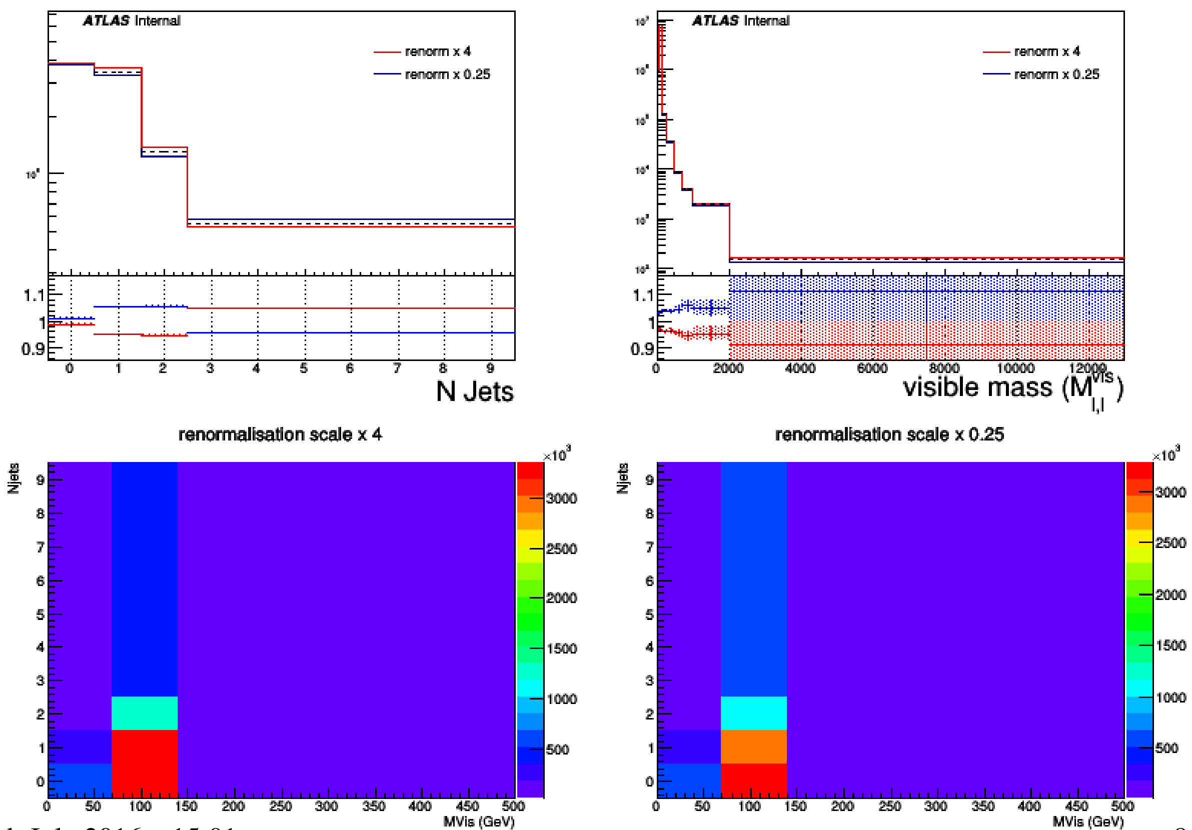
The largest variations expected are due to factorisation and renormalisation scales these can be seen in figure [2] and the corresponding weights to be applied to the reconstructed nominal samples at truth level can be seen in figure [3]

6.3 Distributions of ZCR with Scale Variations

plots of ZCR with scale variations here.

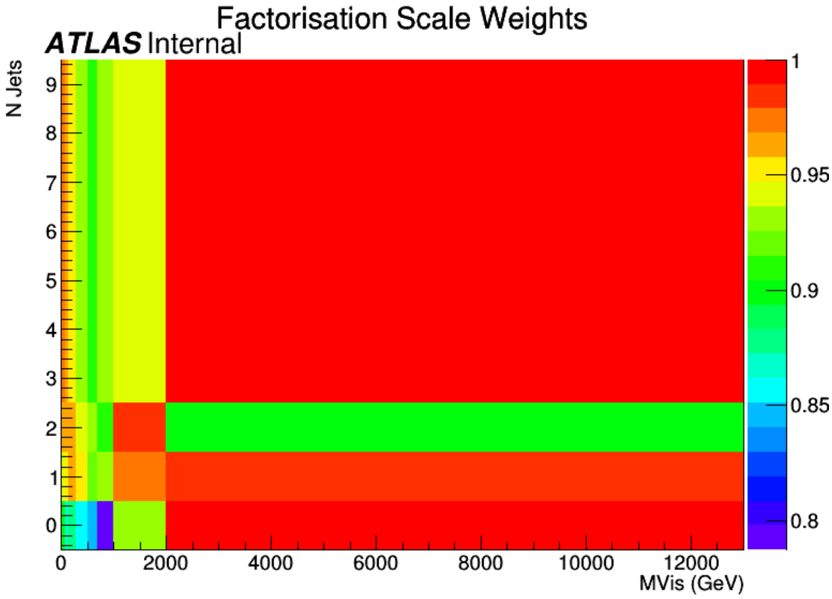


(a)

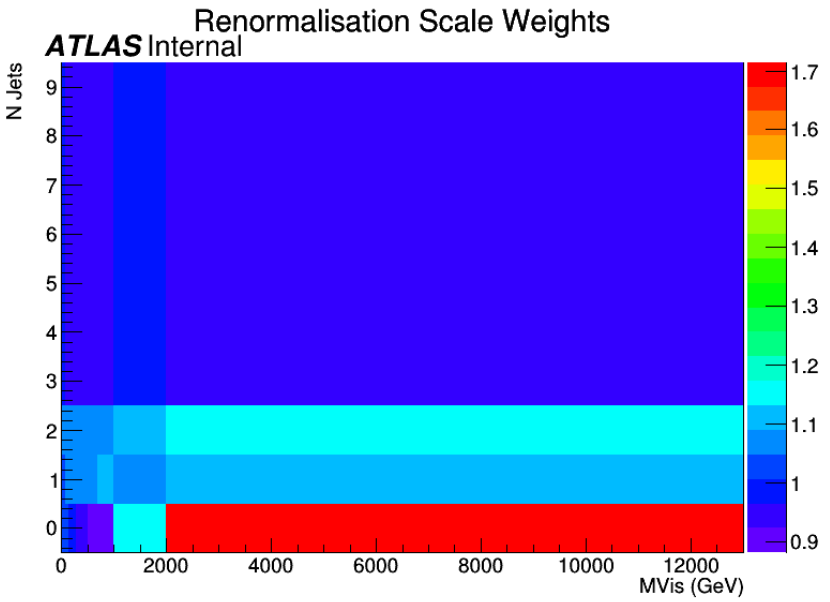


(b)

Figure 2: 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 3: Weights produced to emulate changes in (a)facorisation and (b) renomalisation scales. The weight should be positive for down variations and positive for up

7 Contribution of EW Zjj

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

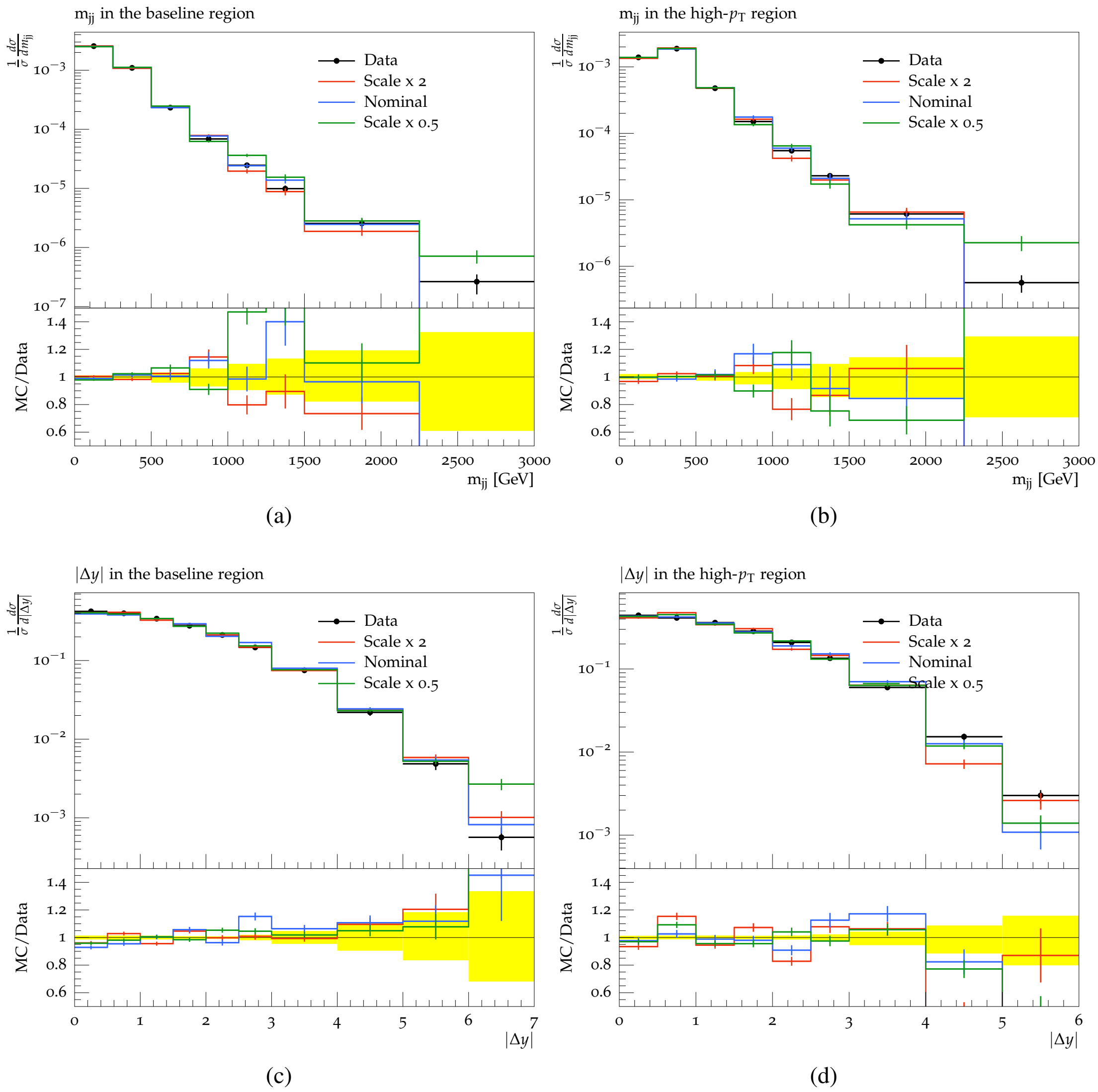


Figure 4: Kinematic distributions of the mass of the dijet system (above) and the difference in rapidity (below) for events with two same flavour leptons in the Z kinematic region. As used for validation of Z MC and for searches for EW production of Zjj at ATLAS

This is run for 2M MC events and normalised to crossection. This is MG5aMC@NLO+pythia8 at 8TeV. This is the process that includes the EW diagrams such that interference is calculated. I'm currently re-running a sample with Strong production and EW production to show effect of interference. Eta 15/07/2016.

If required I can run at 13TeV comparing distributions in our preselection, VBF and Boosted regions with and without EW and with and without interference but no data. Eta 21/07/2016

In progress.

8 Effect of Z and Tau polarisations on Z lineshape

Daniele had vatious suggestions for plots and descriptions here.