Investigating the Sources of Mismeasured Missing Transverse Energy in Simulated Z + 1 jet dilepton decays

James Teoh

September 16, 2016

Abstract

We investigate the sources of mismeasured Missing Transverse Energy (MET) in ATLAS Run II Monte Carlo (MC) simulations, specifically looking at events in the tail of the MET distribution where the reconstructed MET is often significantly greater than the truth MET of the simulation. We find the inclusion of jets not associated with the hard physics vertex (pile-up jets) in the MET to be the main cause of mismeasured MET in the tails of Z + jets events and propose a cut using the Jet Tracker Vertex variable (JVT) to remove these jets.

1 Introduction

Missing Transverse Energy (MET) [1] is defined as the total momentum imbalance in the plane transverse to the beam axis. For head-on *pp* collisions, one may naively expect this to be zero since practically all parton momentum will be along the beam axis. However, the presence of undetectable particles will result in a genuine transverse momentum imbalance. These may be neutrinos or possibly new stable weakly interacting particles popular in SUSY models and dark matter searches [1]

When working with simulated data we have the luxury to define the true MET, $E_{\mathrm{T}}^{\mathrm{miss,True}}$ as the negative vector sum of the momenta of all detectable particles associated with the hard physics of the interaction in a given event. For a real particle detector such as ATLAS it is necessary to use the information of the calorimeters and trackers to reconstruct the MET and is usually calculated as follows:

$$\mathbf{E}_{\mathrm{x(y)}}^{\mathrm{miss,Reco}} = \mathbf{E}_{\mathrm{x(y)}}^{\mathrm{miss,e}} + \mathbf{E}_{\mathrm{x(y)}}^{\mathrm{miss,}\gamma} + \mathbf{E}_{\mathrm{x(y)}}^{\mathrm{miss,}\tau} + \mathbf{E}_{\mathrm{x(y)}}^{\mathrm{miss,jets}} + \mathbf{E}_{\mathrm{x(y)}}^{\mathrm{miss,softjets}} + \mathbf{E}_{\mathrm{x(y)}}^{\mathrm{miss,calo}\ \mu} + \mathbf{E}_{\mathrm{x(y)}}^{\mathrm{miss,CellOut}} + \mathbf{E}_{\mathrm{x(y)}}^{\mathrm{miss,}\mu} \quad (1)$$

Where $\mathbf{E}_{\mathrm{x(y)}}^{\mathrm{miss}}$ is the x(y) MET component related to $E_{\mathrm{T}}^{\mathrm{miss,Reco}}$ by the following

$$E_{\mathrm{T}}^{\mathrm{miss,Reco}} = \sqrt{\mathbf{E}_{\mathrm{x}}^{\mathrm{miss,Reco}\ 2} + \mathbf{E}_{\mathrm{y}}^{\mathrm{miss,Reco}\ 2}} \quad (2)$$

We focus on events in the well studied Zll decay in the MET tails (Events with $E_{\mathrm{T}}^{\mathrm{miss,Reco}} > 100$ GeV) accompanied by jets. We expect zero or small $E_{\mathrm{T}}^{\mathrm{miss,True}}$ as MET can only be generated from neutrinos produced when heavy quarks in the jets radiate W bosons. Fake $E_{\mathrm{T}}^{\mathrm{miss,Reco}}$ arises due the effects of limited inner tracker coverage, calorimeter resolution, reconstruction inefficiencies, pile-up etc. By minimizing and quantifying the fake $E_{\mathrm{T}}^{\mathrm{miss,Reco}}$ in Z + jet decays we hope to reduce the backgrounds for SUSY searches which often have high jet multiplicity and large MET in the final states (e.g. stop-1lepton).

We begin by quantifying the (reco) MET tails in MC Z + jet events before determining pileup as the main cause of mismeasured MET in Z + 1 Jet events. We then investigate new cuts on Jet vertex association and inner tracker variables, checking these do not bias or reduce MET reconstruction efficiencies in a $t\bar{t}$ background sample.

[1] also interchangeably known as Missing Transverse Momentum

2 MET comparisons for Z + jets Simulations

Preliminary studies comparing $E_{\mathrm{x(y)}}^{\mathrm{miss,Reco}}$ to $E_{\mathrm{x(y)}}^{\mathrm{miss,True}}$ were conducted using the Sherpa NNPDF3ONNLO Monte Carlo (MC) model to generate both $Z\mu\mu$ and Zee events with at least one jet.[2]Feynman diagrams illustrating how jets may be associated with the hard scatter (HS) of dilepton Z decays are shown in Fig. 1.

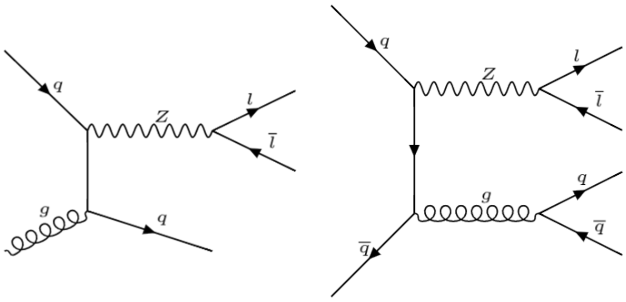

Figure 1: Feynman diagrams showing possible interactions producing a monojet and dijet in Zll decays. The quarks produced go on to produce jets through hadronization The presence of a virtual quark means the jets originate from the same vertex as the leptons

For all histograms, events were weighted by $\mathrm{Xs} * \kappa * \epsilon * 10^6$ and normalized by the number of events generated for each MC ID number. (e.g. Zee 500<Pt<700 BFilter, MC ID:363402 has a different weight to $Z\mu\mu$ 70 <Pt<140 CFilterBVeto, MC ID:363368) Xs, κ and ϵ represent the event cross section, kFactor and efficiency unique to each MC ID/event type. The 10^6 factor converts the weight into units of fb^{-1}.

The extent of the MET tails was studied by first creating 2D histograms of $E_{\mathrm{T}}^{\mathrm{miss,True}}$ vs $E_{\mathrm{T}}^{\mathrm{miss,Reco}}$ and by plotting $E_{\mathrm{T}}^{\mathrm{miss,True}}$ for when $E_{\mathrm{T}}^{\mathrm{miss,Reco}}$ exceeded a given threshold. From both plots in Fig. 2 and Fig. 3 it was abundantly clear that the Reco MET was being catastrophically mismeasured in the MET tails.

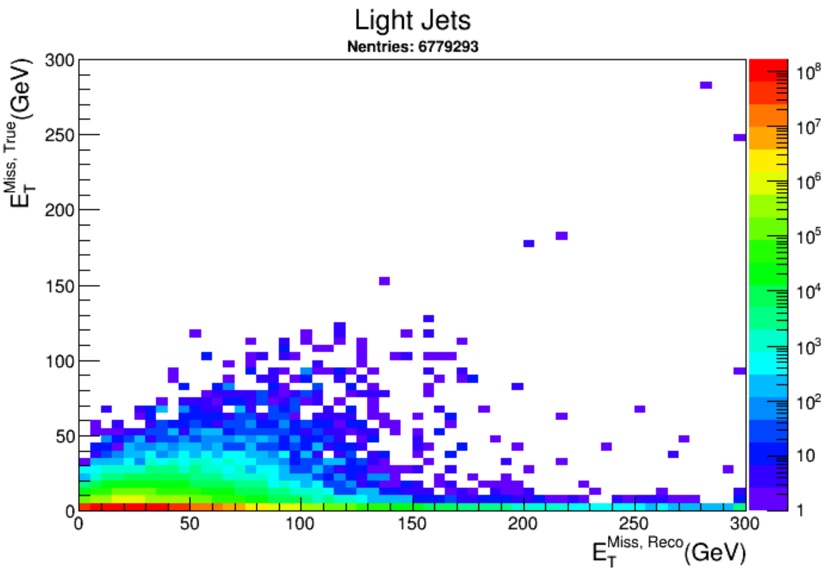

Figure 2: Weighted 2D histogram (5GeV binning) of Z + jets events with light flavor tagged jets only (parent quark u, d or s). The logarithmic scale clearly shows events in the Reco MET tails almost always have near zero truth MET. The same was seen for events with heavier flavor tags

[2]These were combined into one sample. Previous histograms of the MET distribution using PowHeg showed very little difference between $Z\mu\mu$ and Zee samples. MET is slightly more likely to be mismeasured in $Z\mu\mu$ due to muons emitting Bremsstrahlung in the outer layers of ATLAS, far away from the EM calorimeters. Requiring a dilepton invariant mass close to the Z mass also did not affect the MET distribution noticeably

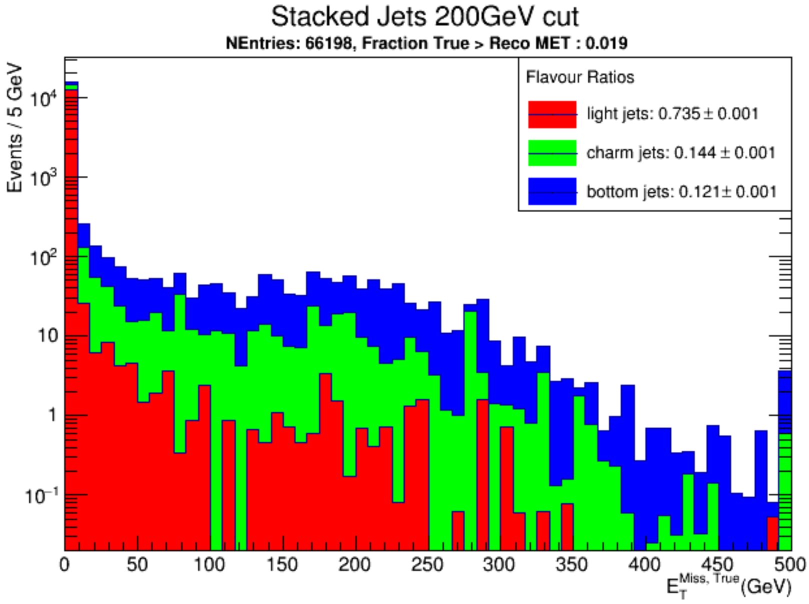

Figure 3: This stacked histogram plot shows the true MET for events with Reco MET > 200 GeV, the massive spike in the 0-5GeV bin for light jets again shows that these events should have a MET of 0 GeV. As well as showing the relative frequencies of jet flavors, this plot also shows events with light tagged jets only have the worse MET reconstruction. This trend was repeated for Reco MET > 100 - 300 GeV

3 Causes of mismeasured MET in Z + 1 Jet events

To isolate only the simplest events, we impose cuts requiring $E_{\mathrm{T}}^{\mathrm{miss,Reco}} > 100\mathrm{GeV}$, $E_{\mathrm{T}}^{\mathrm{miss,True}} < 10$ GeV and exactly one signal jet in the event. A signal jet is defined as a jet which has passed the JVT selection criteria imposed to reduce the number of low P_T[3] pileup (PU) jets in the central region where tracking information can be used to identify PU . The JVT selection is outlined as follows: include all jets with Jet P_T >60 GeV and $|\eta| < 2.7$[4] For central jets with $|\eta| < 2.4$ and 20< P_T <60GeV, these will only pass selection if the JVT (as described in **4**) associated with the primary interaction vertex exceeds 0.64.

It was suspected that mismeasured jets (including PU jets, removing the wrong jets via JVT selection, overlap removal etc.) were responsible for the large $E_{\mathrm{T}}^{\mathrm{miss,Reco}}$ in low $E_{\mathrm{T}}^{\mathrm{miss,True}}$ events. Swapping the Reco Jets for the truth jets in Fig. 4 confirmed this.

Whilst the JVT selection usually improves the MET by efficiently removing PU jets, for our Z + 1 jet sample the selection has an overall detrimental effect as shown in Fig. 5 Two mechanisms by which the JVT selection may adversely affect the MET are illustrated in Fig. 6. Firstly, low P_T jets from the hard scatter (HS) may fail the JVT selection and would be excluded from the MET. Secondly, dijets where one jet just passes the JVT selection (e.g. its P_T just exceeds 60GeV) and the other fails give rise to large fake $E_{\mathrm{T}}^{\mathrm{miss,Reco}}$ which would not be present if both jets were included in the MET calculation. These dijets are common occurring in approximately 30% of events where the single signal jet was identified as a PU jet. (Here a dijet was defined if the azimuthal angle between the jet passing and failing the JVT was between 2.8 and π)

[3] Jet's transverse momentum
[4] where η is jet pseudo rapidity

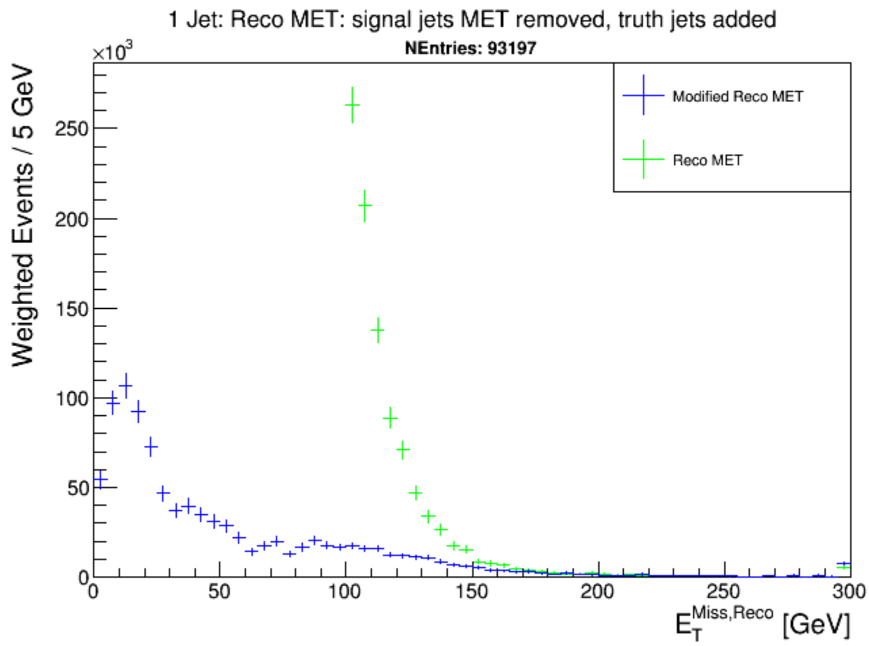

Figure 4: MET distribution for when all jets are swapped with truth jets i.e. the best possible improvement we can obtain by focusing only on the jets. This was done by first subtracting the MET of all the reconstructed jets from the total event MET, adding back the (more reliable) track Pt for that jet corresponding to the primary vertex (PV) and finally adding back all the truth jets to the MET calculation. This drastically reduces the MET tail in the 100-150 GeV region and is a strong indication that mis-reconstructed jets are the primary cause for MET tails. 83.5% of events now fall below the 100 GeV threshold

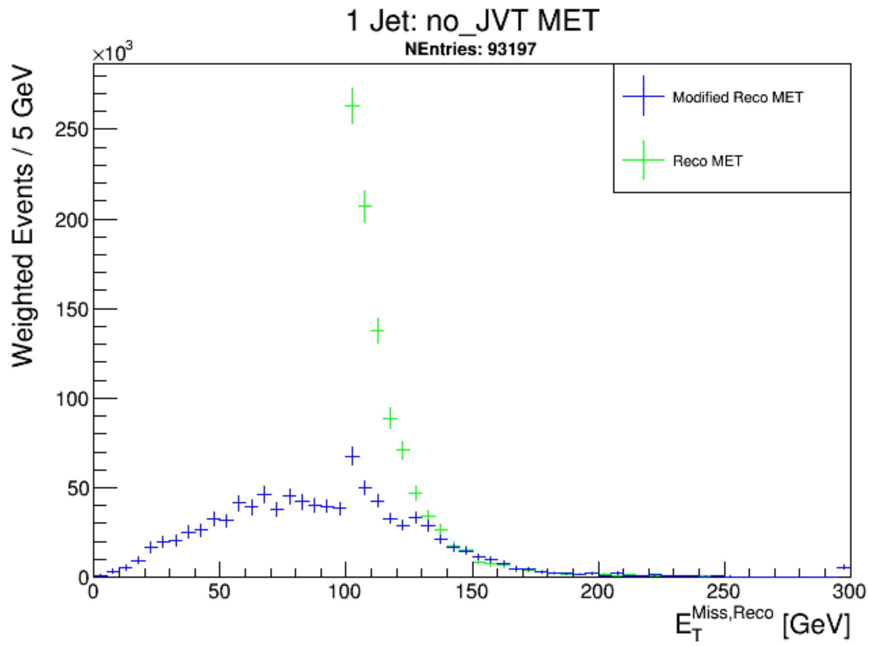

Figure 5: Surprisingly,removing the JVT selection improves the Reco MET significantly. Here, 58.2% of events now fall below the 100 GeV threshold

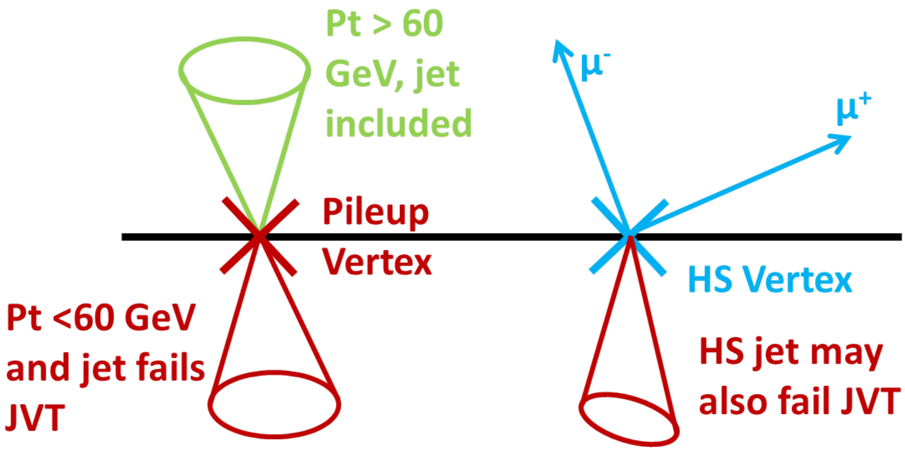

Figure 6: The illustration to the left describes a possible scenario where JVT selection may remove only one of the jets in a pileup dijet leading to fake $E_{\mathrm{T}}^{\mathrm{miss,Reco}}$. On the right, HS jets may also fail the JVT selection

Whilst the JVT selection is needed for other samples, we can separately quantify how much of the MET we can fix by correcting for these mechanisms. Fig. 7 and Fig. 8 confirm that the inclusion of pileup jets (which may be made worse by the JVT selection in this case) is the dominant cause of mis-measured $E_{\mathrm{T}}^{\mathrm{miss,jets,Reco}}$ and hence overall mis-measured MET in these events.

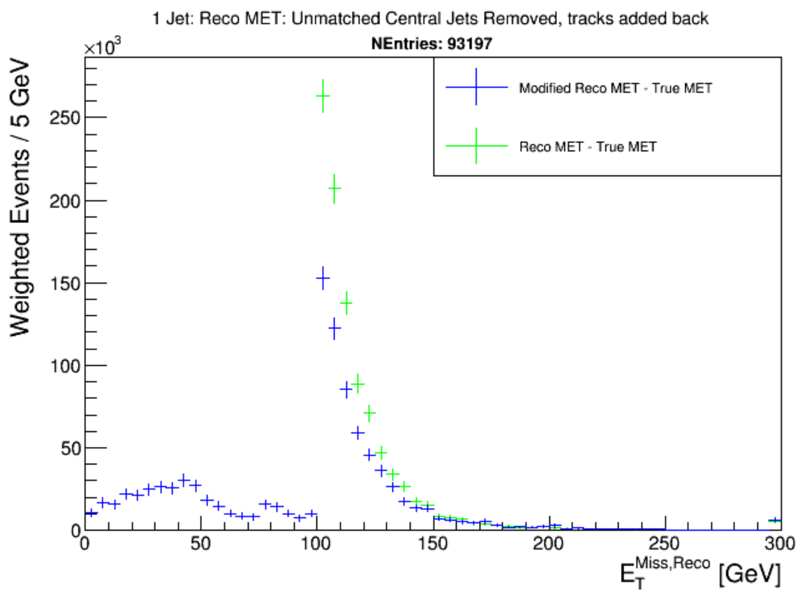

Figure 7: MET distribution for when all PU jets in the central region ($|\eta| < 2.4$) are removed from the MET calculation. For each jet, attempts were made to topologically match it to a truth jet by requiring $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.3$ (where ϕ is the azimuthal angle between the jet and the beam axis and $\Delta\phi$ is the difference in azimuthal angle between the reco and truth jet). Jets which could not be matched to any of the truth jets were designated as a pileup jet. The PU jet P_T was subtracted from the MET and the PV track P_T added back. With this change 34.9% of events fall below the 100GeV threshold

Fig. 9 studies the second scenario described in Fig. 6, the inclusion of HS jets failing the JVT selection but this does not have a significant effect on the MET tails. For the sake of completeness, Fig. 10 shows the effect of adding in truth jets which were not reconstructed at all. Whilst there will be many truth jets which simply are not reconstructed, these usually have negligible P_T and jet reconstruction efficiency for high P_T jets is good.

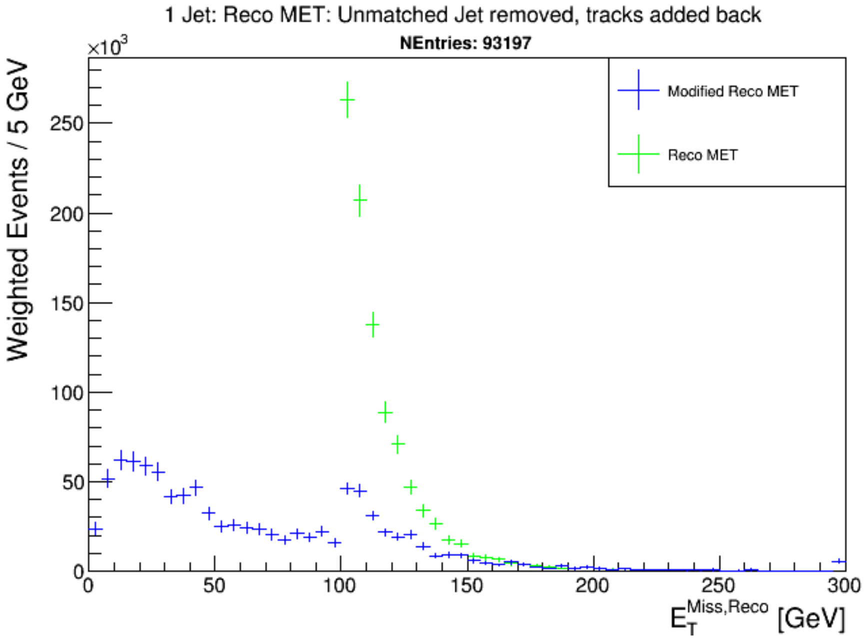

Figure 8: MET distribution for when all PU jets extending to $|\eta| < 4.5$ are removed from the MET calculation using the same method as in Fig. 8. This clearly illustrates the inclusion of pile-up is responsible for most of the mismeasured MET but we can only use the tracking info to fix jets with $|\eta| < 2.4$. 73.5% of events now fall below the 100GeV threshold.

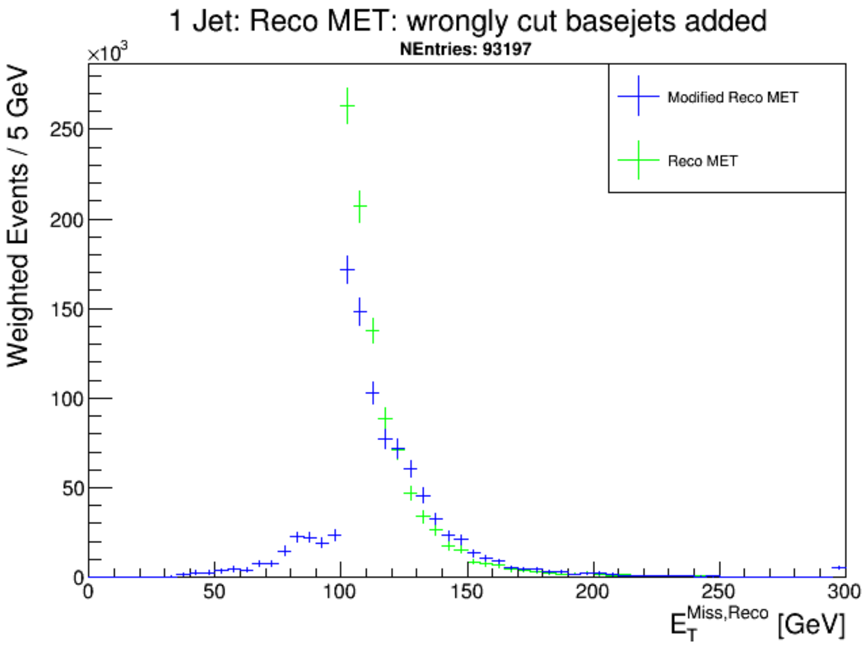

Figure 9: MET distribution for when missed HS jets are added back to the MET. This was done by ΔR matching jets failing the JVT to truth jets to identify them as missed HS jets. When adding these jets their track Pt was also subtracted (so the Jet MET was determined from only calorimeter info and not double counted with calorimeter and track info)

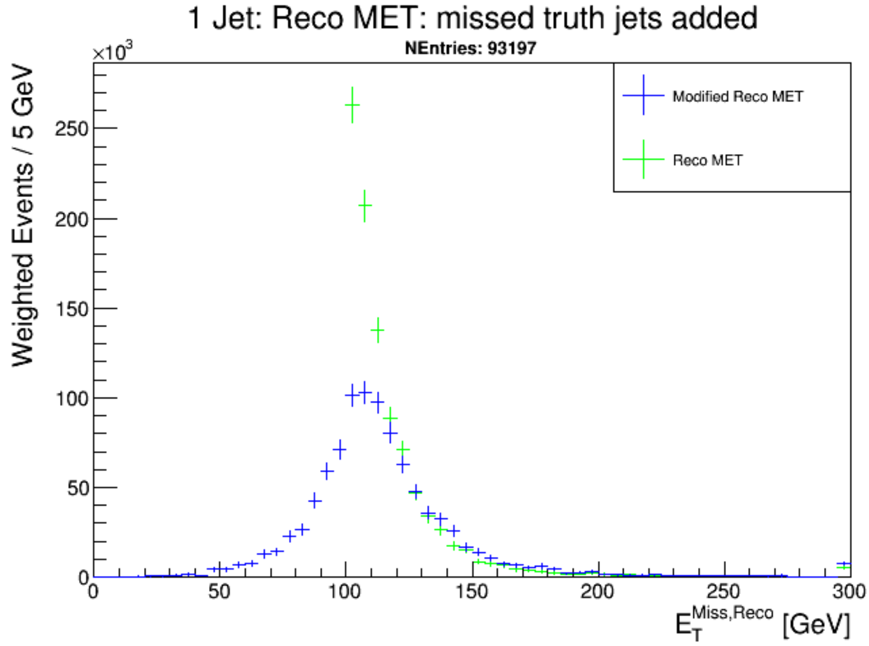

Figure 10: MET distribution for when missed truth HS jets are added back to the MET. This was achieved by using ΔR matching and adding the truth jets which did not match any reco jets. If anything, this appears to smear the MET tails slightly

Briefly the effect of electron overlap removal (OR) was investigated. This was a procedure by which a jet is removed if its topology overlaps with an electron the from HS vertex. OR occurred in 9% of events where the signal jet was a HS jet (as opposed to more than half of these events including pileup jets). From now on we only focus on reducing the number of pile-up jets included in the MET calculation.

4 Suppressing pile-up jets using jet vertex association variables and track info

By requiring additional cuts to each jet, it is possible to improve PU jet background rejection whilst maintaining a high signal efficiency for keeping HS jets. Three jet vertex association variables were investigated: JVF, JVT and R_{pT} which try to quantify how likely a jet is to be pileup (0 as strongest indication of pileup, 1 (or greater for R_{pT}) as a strong indication of HS). They are defined in [2]. . We focused on the most sophisticated variable, JVT which can be thought of as a modified JVF (the ratio of the transverse momentum of tracks associated with the PV to the transverse momentum of all tracks) and crucially does not depend on the total number of identified vertices in the interaction.

By plotting the JVT for PU and HS jets in Fig. 11 we can see that even by cutting on very low values of the JVT we can remove a significant fraction of the PU jets. Limiting ourselves to the central $|\eta| < 2.4$ region (for which there is adequate tracking information available to accurately calculate the JVT) we find we can reject 62.5% of pile-up jets whilst maintaining 98.6% of HS jets by removing jets with JVT< 0.05 (for the central region). Higher background rejections could be achieved (up to 70% by cutting at higher JVT values but these would likely be accompanied by reduced signal efficiency when used on other samples).

Another variable interest in the nTrk500PV variable which is the number of separate tracks with $P_T >$500 MeV associated with the jet and originating from the primary vertex. From Fig. 12 and ROC curve analysis we see we can also discriminate well between HS and PU jets with an optimal cut threshold requiring ≥ 3 tracks for the jet to pass. This has a Signal efficiency of 94.7% and background rejection of 91.1%, and removes almost all PU jets from the central region.

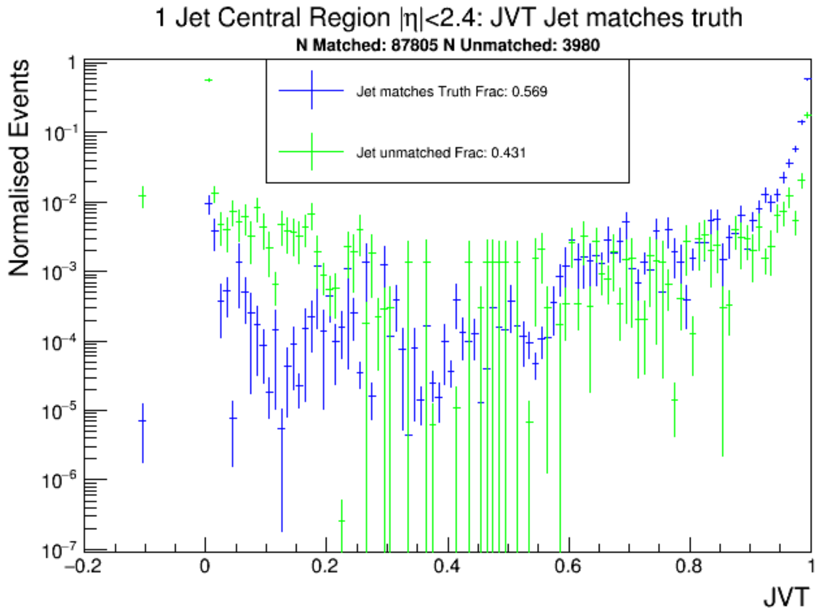

Figure 11: Histogram showing the JVT values (associated with the primary vertex) of HS jets and PU jets in the central $|\eta| < 2.4$ region. Unsurprisingly, most PU jets have a very low JVT we can cut on

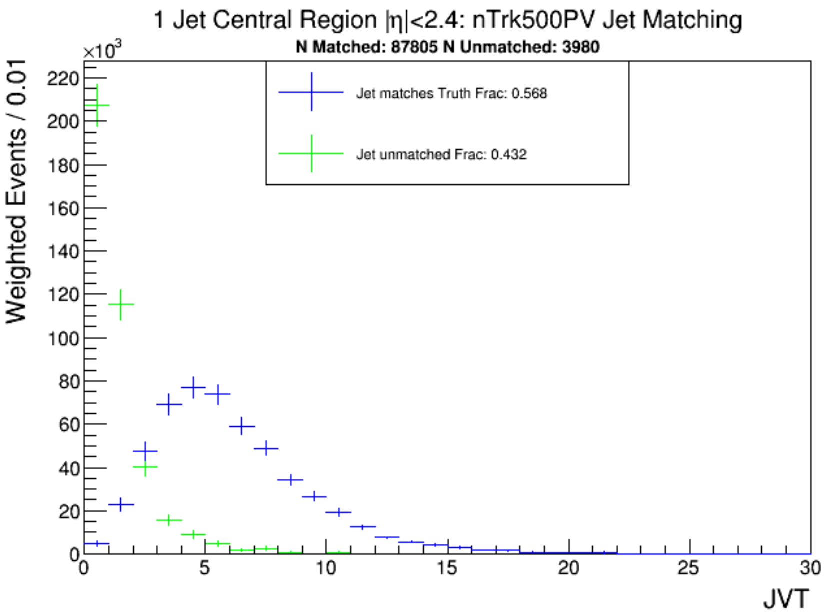

Figure 12: Histogram showing the number of tracks as described by the nTrk500PV variable for PU and HS jets

The next step was to naturally see how the MET distribution was affected by imposing these cuts. We focused on the central region as practically all jets in the forwards region would be removed by these cuts due to lack of tracking info. Fig. 13 and 14 demonstrate that the MET in the central region due to jets can be mostly corrected by these cuts with the nTrk500PV variable looking most promising. Fig. 15 further characterizes the performance of these cuts through a ratio plot with Fig. 7.

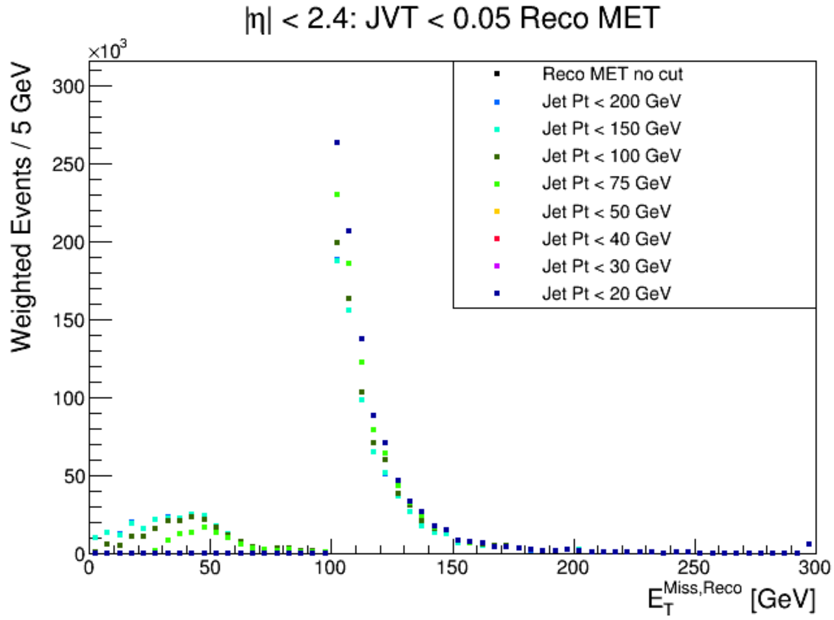

Figure 13: MET distributions for the JVT cut whilst varying the Jet P_T threshold required for the cut

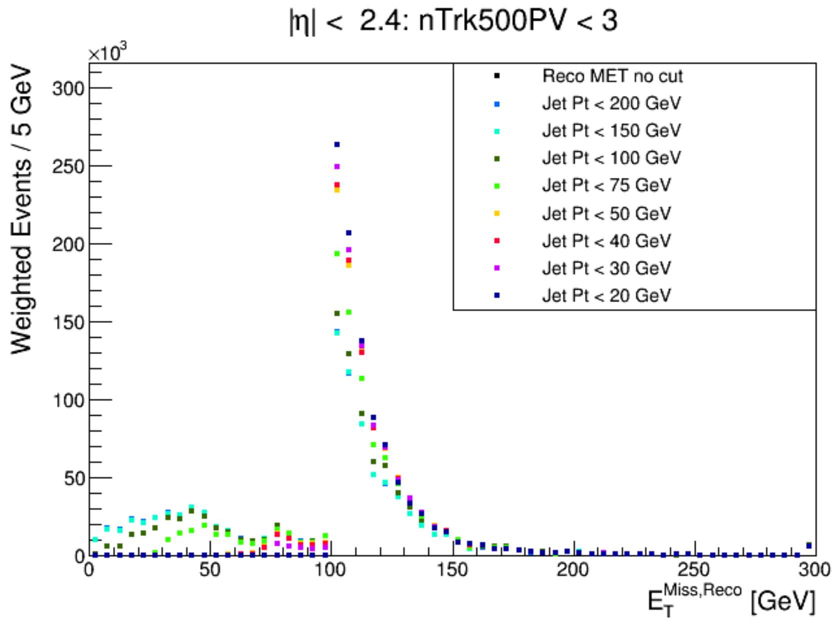

Figure 14: MET distributions for the JVT cut whilst varying the Jet P_T threshold required for the cut. This improves the MET almost as much as in Fig. 7

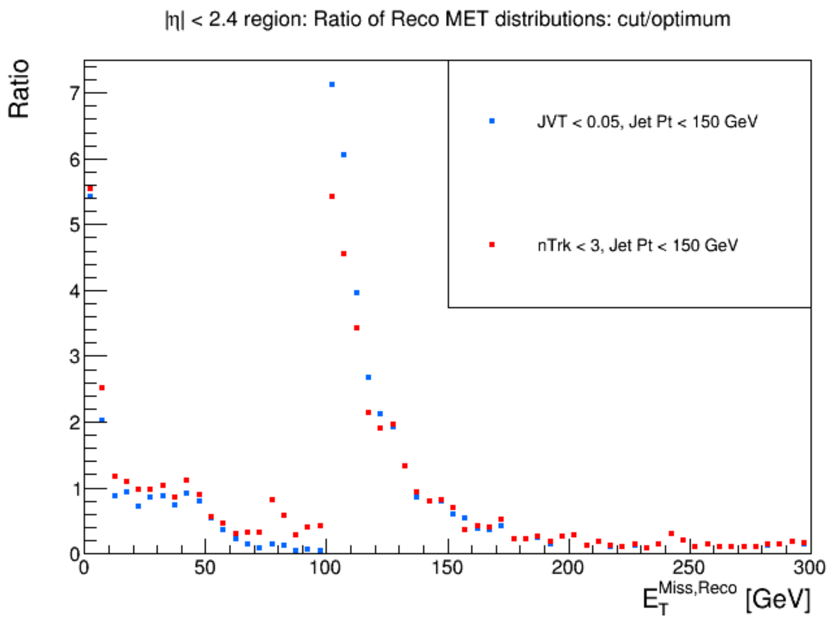

Figure 15: Ratio plots for the JVT and nTrk500PV cut (divided by the 'optimum' in Fig. 7) We can see nTrk500PV is a slightly better variable to cut on and that both cuts are close to optimum in the 0-50GeV range of the fixed MET distributions.

For these cuts to be useful they not only have to perform well on this specific Z + 1 jet sample but also on the samples and data used for real physics analysis. For instance, applying these cuts in the forwards region improves the MET distribution of the Z + 1 jet sample, mainly by removing most of the jets anyway due to a lack of tracking info, but this is not something we would want to do in general. Similarly, we only impose these cuts for jets falling below a P_T threshold . Physics analyses will often want to keep high P_T jets regardless. We found we could reduce this threshold to 150 GeV before cut performances were affected.

5 $t\bar{t}$ Validation

$t\bar{t}$ events are a major background in SUSY searches. The final state can often be indistinguishable from a SUSY decay with high jet multiplicity and large MET. We validate performance by studying the difference in absolute MET,$E_{\mathrm{T}}^{\mathrm{miss,Reco}} - E_{\mathrm{T}}^{\mathrm{miss,True}}$ which takes a near Gaussian distribution centered on zero. Our cuts should not broaden, shift or skew this distribution. From Fig. 16 and Fig. 17 we notice that the JVT cut has a minimal impact on the $t\bar{t}$ sample (MC ID: 410000) but realize the nTrk500PV cut broadens the distribution, making this a less than ideal cut. Cutting in the forwards region ($2.4 < |\eta|$) and partial forwards region ($2.4 < |\eta| < 2.7$) would significantly broaden and skew the distribution confirming that these cuts were not suitable without full tracker information.

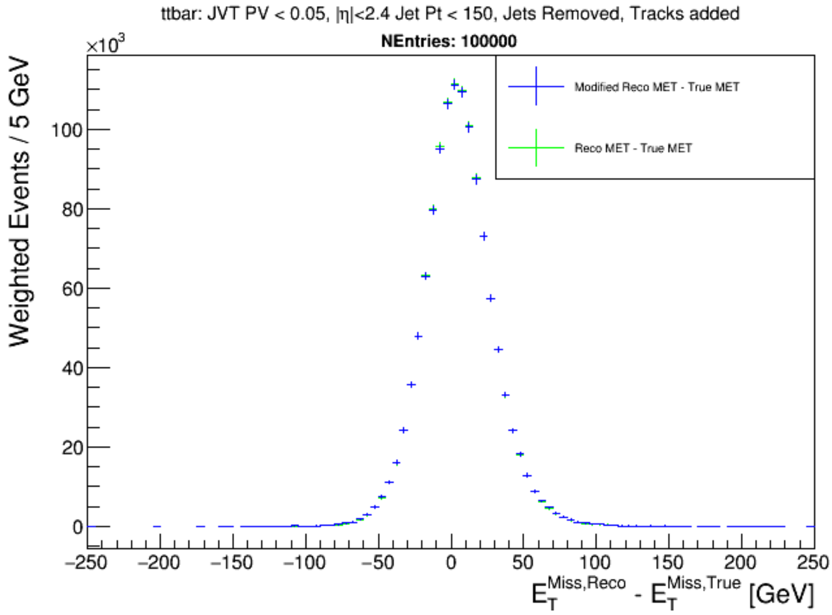

Figure 16: The difference in truth MET and reco MET before and after the JVT ¡ 0.05 cut is applied to the central region. There is practically no change to the MET, even out in the tails making the JVT cut look promising

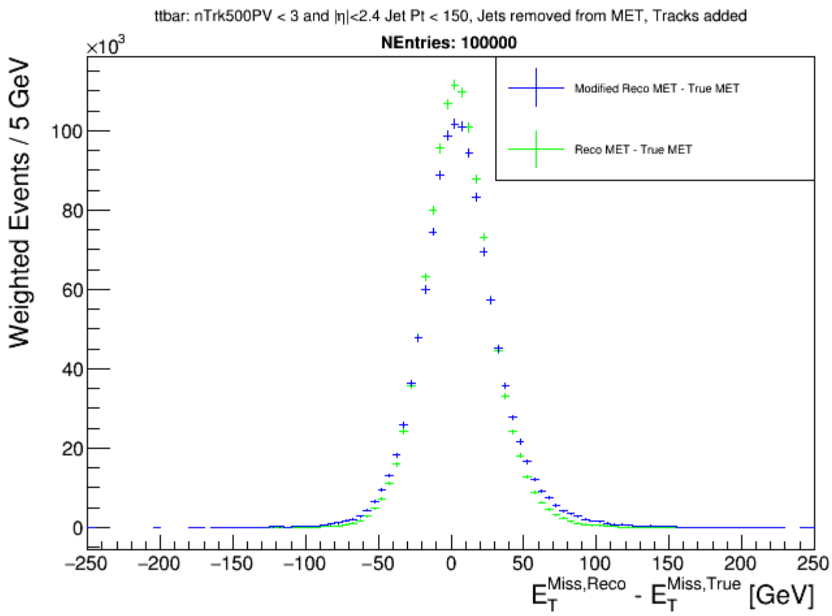

Figure 17: The difference in truth MET and reco MET before and after the nTrk500PV <3 cut is applied to the central region. There is a slight broadening of the MET difference and due to this, the JVT cut should be favored over the nTrk500PV cut (these cuts fix the MET by similar degrees)

6 Conclusion

We have shown for Z + 1jet events, the jet term is the dominant cause of mismeasured MET in the MET tails and this should be of concern for SUSY searches. The inclusion of Pile-up jets in the MET calculation is the primary reason for mis-measured MET whereas overlap removal, HS jets failing the JVT and jet reconstruction inefficiencies are relatively small effects. The inclusion of pile-up jets can be exacerbated by the JVT selection which often cuts only one of the jets in a pile-up dijet, generating fake $E_{\mathrm{T}}^{\mathrm{miss}}$. We have demonstrated that cuts in the central region ($|\eta| < 2.4$) where full tracker information is available can reduce the number of PU jets included and hence improve the MET distribution. The most promising cuts are JVT (associated with the primary vertex) < 0.05, Jet $P_T < 150$ GeV and nTrk500PV ¡3, Jet $P_T < 150$ GeV.

Although the nTrk500PV cut had a higher background rejection of PU jets, validation on a $t\bar{t}$ sample indicated the JVT had no adverse effects on this sample's MET whereas nTrk500PV reduced the MET resolution slightly but did not introduce bias. We recommend that the JVT cut can be implemented readily, perhaps after validation with MC simulations more closely resembling SUSY final states however the nTrk500PV cut should be used sparingly, possibly as an independent variable in conjunction with the JVT cut.

A significant proportion of the PU jets lie in the forwards region ($|\eta| > 2.4$) where tracker info is either absent or partial from large R jets. Attempts to address these using the cuts discussed in the forwards or partial forwards region would considerably reduce the MET resolution of the $t\bar{t}$ sample. Reducing the scope of physics analyses to the central region would alleviate this but this is not the ideal solution. It may also be possible to edit the JVT cut specifically to look for dijets and include them in the MET even when one fails the JVT.

References

[1] Performance of Missing Transverse Momentum Reconstruction in ATLAS with 2011 Proton-Proton Collisions at $sqrts = 7$ TeV. Technical Report ATLAS-CONF-2012-101, CERN, Geneva, Jul 2012.

[2] P Nef and A Schwartzman. Tagging and suppression of pileup jets. Technical Report ATLAS-COM-CONF-2014-025, CERN, Geneva, Apr 2014.