Using the M_{T2} and ϕ^* variables to investigate Monte Carlo methods of $t\bar{t}$ production

Caitlin Jones

September 8, 2015

Abstract

In this project the behaviour of Monte Carlo simulations for the event $t\bar{t} \to \ell^-\ell^+ b\bar{b}\nu\bar{\nu}$ in the ATLAS (A Toroidal LHC ApparatuS) detector were investigated. The performance of algorithms for resolving the pairing ambiguity for jets and leptons at the reconstructed level of Monte Carlo data was accessed. It was found that a combined selection based on the M_{T2} and M_{lb} variables gave the best performance with an efficiency to select the correct lepton-jet permutation of 0.516 and purity of 0.791. This was an improvement on a previous method which used only M_{lb}. Additionally the behaviour of the ϕ^* variable for 4 different Monte Carlo simulations at the parton, particle and reconstructed level was investigated. It was found that ϕ^* is sensitive to modelling the kinematics of $t\bar{t}$ events and is a good variable to tune Monte Carlo simulations.

1. Introduction

1.1. Top Quark Physics

This report focuses an analysis of Monte Carlo simulations of a $t\bar{t}$ decay in the ATLAS detector. The reason that such event are interesting is primarily due to the top quark's large mass. The mass is measured to be 173 GeV which is much greater than all other quark masses. The top quark's large mass mean its has very short lifetime and is the only quark which decays before hadronising. This means it is possible to determine properties of the top quark from its decay more easily as its quantum numbers are not diluted by hadronisation.

1.2. Monte Carlo Simulations

In order investigate the properties of the top it is necessary to use accurate Monte Carlo (MC) simulations. To generate each event, the Lorentz four vectors of the produced particles are found by evaluating an integral over the phase space of the momentum of the incoming and outgoing particles, with a matrix element in the integrand describing the interaction [1]. The approximate matrix element is found from perturbative next to leading order calculations from QCD. The integral is evaluated by Monte Carlo integration where for an integral :

$$I = \int f(x)dx \tag{1}$$

The estimate for its value is:

$$E = \frac{1}{N}\sum_{n=1}^{N} f(x_n). \tag{2}$$

When events are simulated, difference stages (levels), corresponding to the chronological progression of the event, are created. Three levels are considered in this project; parton, particle and reconstructed. The parton level, also called the truth level, corresponds to immediately after the initial hard scattering. The quarks are not hadronised and no leptons, such as τ, have decayed or radiated, and each particle is labeled so all information about the event is known. The particle level is the state after $10^{-10}s$, the hadrons have now formed particle jets, and the leptons may have decayed. It is not known which jets and leptons were the product of which decay, and there is missing momentum from neutrinos. The reconstruction level is the same as the particle level except the event is simulated to have passed through the detector, it is equivalent to the data recorded in real experiments.

1.3. The ATLAS Detector

The ATLAS detector is a cylindrical detector at the LHC (Large Hadron Collider) made up of an inner tracking detector surrounded by a 2T solenoidal magnet, calorimeters and a muon spectrometer [2]. The inner tracking detector is made up of a silicon pixel detector, a silicon micro strip detector and a transition radiation tracker. The inner tracking detector is inside the magnetic field so that momentum of charged particles can be determined. The calorimeters measure the energy of particles absorbed and have an electromagnetic and hadronic part. They are positioned around the barrel of the detector and as two end caps to cover both the high and low pseudorapidity regions. The muon spectrometer is made up of toroidal magnets providing a magnetic field for tracking and triggering chambers.

1.4. $t\bar{t}$ Decay

In this project the event topology considered was $t\bar{t} \to \ell^-\ell^+ b\bar{b}\nu\bar{\nu}$, see figure 1. This is the dilepton channel which is interesting as it has an easily suppressed background and a easily identifiable signal of two leptons. Jets are also present, with two possible jets originating from the b-quarks and potentially additional jets from hadronisation and pileup. The overall charge of the jets is unknown, therefore in order to analyze the properties

of the parent top quark the correct daughter lepton and jet must be identified for both the t and $\bar{t}$. The first section of this report discusses the use of a variables called M_{lb} and M_{T2}, [3], achieve the correct lepton-jet pairings.

The second half of this report focuses on the variable ϕ^*. ϕ^* is an angular variable which has a sensitivity to the top quark p_T.

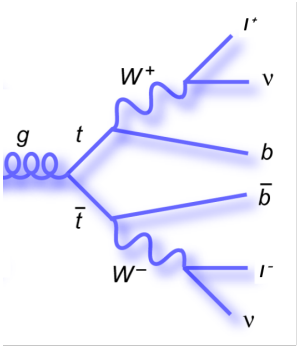

Figure 1: Feynman diagram of $t\bar{t} \to \ell^-\ell^+ b\bar{b}\nu\bar{\nu}$

2. Pairing with M_{T2} and M_{lb}

2.1. Pairing with only M_{lb}

Two methods to find lepton-jet pairings were considered, the first using one variable named M_{lb} and the second using two variables M_{lb} and M_{T2}. M_{lb} is defined as the combined invariant mass of the lepton and jet. The first pairing algorithm worked as follows:

- Randomly select two of the jets in the event and calculate the value of $M_{lb} = \frac{1}{2}(m_{l^+,b_i} + m_{l^-,b_j})$ where $m_{l^\pm,i}$ is the invariant mass of the Lorentz vector of the lepton-jet pair.
- Repeat this for every possible pair of jets and take the lepton and jet combination that gives the minimum value of M_{lb}
- If this value is less than 155 GeV then take the pair that produced that value as the b quark jets.

This method exploits the fact that the value of $m_{l^\pm,i}$ cannot be greater than the top mass and is normally much less as momentum is carried away by the neutrino. Therefore placing a cut on the value of M_{lb} removes incorrect events and biases the selection towards the correct pairing. The efficiency, number of events matched over number of events considered, and purity, number of correctly matched events over number of matched events, of the method are shown in table 1. It was checked if a pair was correctly matched by comparing to the parton level information.

Table 1: Efficiencies and Purity for the M_{lb} algorithm.

Number of b tagged	2	≥ 1	No Requirement
Efficiency	0.474	0.470	0.470
Purity	0.695	0.558	0.525

The table shows that requiring more b-tagged jets improves the purity but does not significantly improve the efficiency.

2.2. Defining the M_{T2} variable

The implementation of the second, pairing algorithm is based on that of [3]. The value of M_{T2} is found as follows:

- Create a range of partitions of the missing transverse momentum and assigning that momentum to two neutrinos as p^d_{iT}.
- For each possible combination of the jet and lepton for the top and anti-top quark calculate the value the transverse mass of the combination of :

$$M^2_{iT} = m^2_{i,vis} + 2(E^{vis}_{iT}E^d_{iT} - \overrightarrow{p^{vis}_{iT}}\overrightarrow{p^d_{iT}}), \quad (3)$$

  where:

$$E^{2,vis}_{iT} = m^2_{vis} + |\overrightarrow{p^{2,vis}_{iT}}|, \quad E_{iT} = |\overrightarrow{p^d_{iT}}|. \quad (4)$$

  Where $m_{i,vis}$, p^{vis}_{iT} is the mass and transverse momentum of the combined Lorentz vector of the lepton and jet. p^d_{iT} is the transverse momentum of the neutrino.

- The the value of M_{T2} is the minimum over different partitions of missing momentum of the maximum of the pair of transverse masses:

$$M_{T2} = \min_{p_T}\{\max\{M_{1T}, M_{2T}\}\} \quad (5)$$

Therefore the interpretation of M_{T2} is the reconstructed transverse mass of the parent particle that is minimised with respect to the partition of the missing transverse momentum. Therefore for any combination of leptons and jets to be physically possible its value of M_{T2} must be less than or equal to the mass of the parent particle, in this case the top quark.

2.3. Performance of the M_{T2} algorithm

To investigate the performance of the algorithm, for each event the value of M_{T2} and M_{lb} were calculated for all possible lepton-jet combinations. Two cuts at 176 GeV and 158 GeV for M_{T2} and M_{lb} respectively were performed (values chosen as the optimised values from [3]). Two types of selection were performed. The first was a tight selection required that only one combination passed both cuts.
The second selection was looser and required that either one or two combinations passed either the M_{T2} or the M_{lb} cut and took the combination with the minimum M_{T2} value.

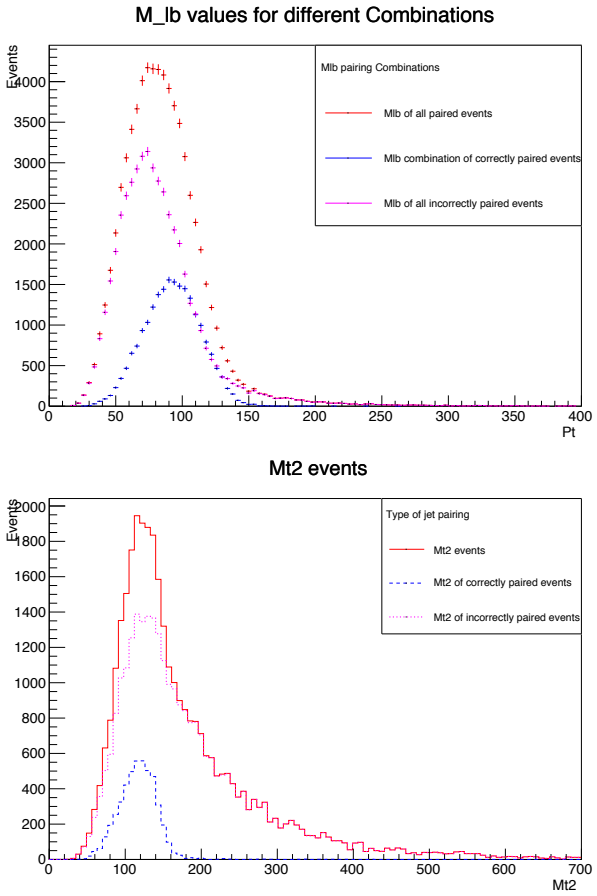

Figure 2: Plot of the value of $\frac{1}{2}(m_{l^+}b_i + m_{l^-}b_j)$ for correct and incorrect lepton-jet pairings (above). Plot of the value of M_{T2} for correct and incorrect combinations (below).

Figure 2 shows the values of M_{lb} and M_{T2} for correct and incorrect combinations of leptons and jets. A cutoff can be seen at approximately 176 GeV and 158 GeV for the correct combinations, suggesting they were appropriate values at which to cut. Additionally, it can be seen that for the M_{T2} distribution the fraction of incorrect combinations that are above the cut is higher than for M_{lb}, suggesting it is an better variable on which to cut.

The efficiency and purity of the algorithm for each type of cut is show in tables 2 and 3.

Table 2: Efficiencies and Purity for the M_{T2} algorithm with requirement only one jet to passes both cuts.

Number of b tagged	2	≥ 1	No Requirement
Efficiency	0.200	0.206	0.206
Purity	0.982	0.983	0.983

Table 3: Efficiencies and Purity for the M_{T2} algorithm with requirement that either one or two jets passes one or both of the cuts

Number of b tagged	2	≥ 1	No Requirement
Efficiency	0.516	0.511	0.511
Purity	0.791	0.790	0.788

For the tight selection the purity is very high although the efficiency is low, whereas for the loose selection the efficiency is much higher although the purity is lower. The loose selection is an improvement on both the purity and efficiency achieved from pairing using the M_{lb} variable alone. The efficiency and purity for both types of selection was only negligibly changed by the number of b-tagged jets required to consider the event. This suggests that M_{T2} could be used for b-tagging as it is able to find the jets originating from b quarks without requiring prior b-tagging. Plots of the values of M_{T2} for combinations which passed each selection are shown in figures 3 and 4. By comparing the combination to the parton level, values were separated into those which were the correct and incorrect combination. The differing purity for the two selections can be seen easily in the two figures.

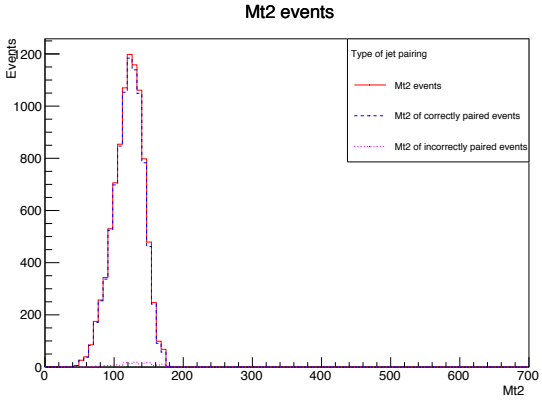

Figure 3: Plot of the value of M_{T2} for combinations that passed tight selection; blue dashed- correct combinations, magenta dotted- incorrect combinations, red full-all combinations. The plot shows that nearly all combinations which passed were correct i.e the purity is high.

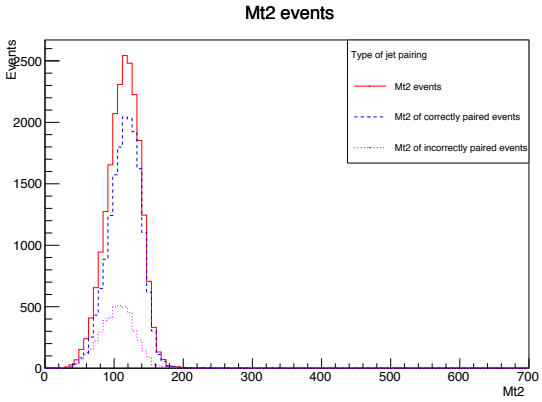

Figure 4: Plot of the value of M_{T2} of events that passed the looser selection; blue dashed- correct combinations, magenta dotted- incorrect combinations, red full-all combinations.

2.4. Discussion and further optimization

The M_{T2} and M_{lb} combined algorithm was an improvement on the M_{lb} pairing alone, in particular it was able to offer much higher purity. However in [3] the optimum efficiency and purity were found to be 0.574 and 0.985, respectively, with the equivalent of the tight selection. The discrepancy between the efficiencies could be due to a different method of b-tagging, or a lower average number of jets per event. Efficiencies could possibly be

improved by increasing the number of partitions of the missing momentum, or optimizing the partition in some other way.

3. Analysing the behaviour of ϕ^* for different Monte Carlo simulations.

In [4] the ϕ^* variable is introduced. It is constructed only from angular variables however it has a sensitivity transverse momentum p_T. Currently there is a tension between the observed data from the ATLAS and CMS experiments in the measured top quark p_T and the MC simulations, and it is unclear why is this the case. Improved tuning of MC simulations, adjusting the parameters of the simulation to match with data, could improve understanding of p_T behaviour. In order to be able to tune MCs the difference between MC predictions for a variable must be greater than the uncertainty of the measured variable. As ϕ^* is an angular variable it is measured much more precisely than energy-based variables; for example in the ATLAS detector the resolution of momentum is approximately 3 GeV ($\mathcal{O}(1\%)$) whereas for ϕ and η the uncertainty is $<< 1\%$. This implies that ϕ^* is a suitable variable to compare MC simulations with. In this project the value of ϕ^* for the decay $t\bar{t} \to \ell^-\ell^+ b\bar{b}\nu\bar{\nu}$ for 4 types of MC simulation was investigated.
ϕ^* is defined as:

$$\phi^* = \sin\theta^* \tan\frac{\phi_{acop}}{2}, \tag{6}$$

where ϕ_{acop} is $\pi - p_T^{(1)} + p_T^{(2)}$ where $p_T^{(1)}$ is the lepton transverse momentum and $p_T^{(2)}$ is that of the jet and

$$\cos\theta^* = \tanh\frac{\eta^{(2)} - \eta^{(1)}}{2} \tag{7}$$

where $\eta^{(1)}$ and $\eta^{(2)}$ are the pseudorapidity of the lepton and jet respectively. Figure 5 shows a schematic of how the variables are defined.

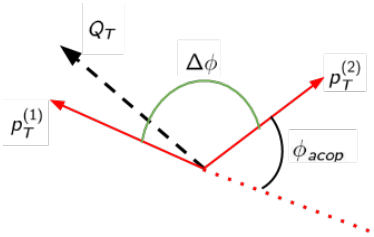

Figure 5: Sketch showing definition of $\Delta\phi$. Here Q_T is the lepton-jet transverse momentum, $p_T^{(1)}$ is the lepton transverse momentum and $p_T^{(2)}$ is that of the jet.

3.1. ϕ^* behaviour at parton level

The values of ϕ^* were calculated from the products of each t and $\bar{t}$ for each type of MC simulation. Figure 6 shows the absolute values ϕ^* showing that the majority of ϕ^* are close to zero. The lower section of the plot shows the normalised ratio of the MC to Powheg+Pythia6. All three of the the Powheg based simulations offer very similar predictions however aMC@NIO differs by a maximum of 5%.

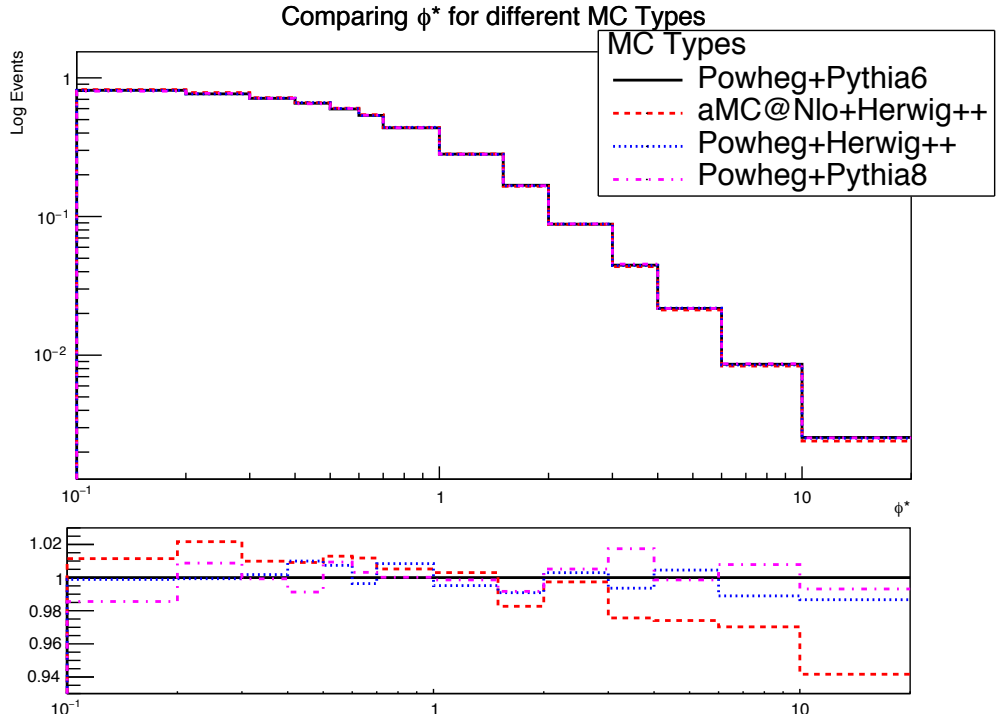

Figure 6: Normalised log log plot of ϕ^* for different MC simulations. The lower plot shows the ratio of each MC to Powheg+Pythia6.

In order to investigate this further the parton level events where spilt along the value of H_t^{VIS}. H_t^{VIS} is the scalar sum of all the visible products of the $t\bar{t}$ decay:

$$H_t^{VIS} = \sum_{\ell,jets} |p_T|. \tag{8}$$

The ratio plots of this splitting are shown in figure 7. The events were split such that the number of events in each plot was roughly equal. The values of ϕ^* for each simulation for low and high H_t^{VIS} differed at low and high ϕ^*. This could be used as a further way to determine if physical data agrees with one simulation over the others.

Although the differences at parton level suggest that ϕ^* is a good variable to distinguish MC simulations, comparisons at particle and reconstruction level are more important as parton level cannot be rigorously constructed from real data for different MC. Additionally, at the parton level the difference in behaviour of the three Powheg based simulations is so small that it would not be possible to say if one fitted the data better than the others.

3.2. ϕ^* behaviour at particle and reconstructed level

In order to calculate ϕ^* for the particle and reconstructed level data, the M_{T2} algorithm with loose selection was used to find the lepton and jet pair for each top decay. The ratio plots of each level can be seen in figure 8.Only events where an electron and a muon were produced were considered.

The histogram binning is coarser for the reconstructed and particle level ratio plots as some events are lost to the inefficiency of the M_{T2} pairing algorithm and due to not recovering all events from parton level. However, it is still possible to see that difference between the MC simulations is around 5%. The shape of the particle and reconstructed level plot is different for large ϕ^* values than the parton level. This could be due to how showering was simulated. In particular the high ϕ^* of the aMC@NIO simulation differed between the particle and parton levels. Additionally, for parton and reconstruction levels there is good agreement between the aMC@NLO+Herwig++, Powheg+Herwig++, and Powheg+Pythia8 simulations but not between Powheg+Pythia6, which may be due to the former simulations being created more recently.

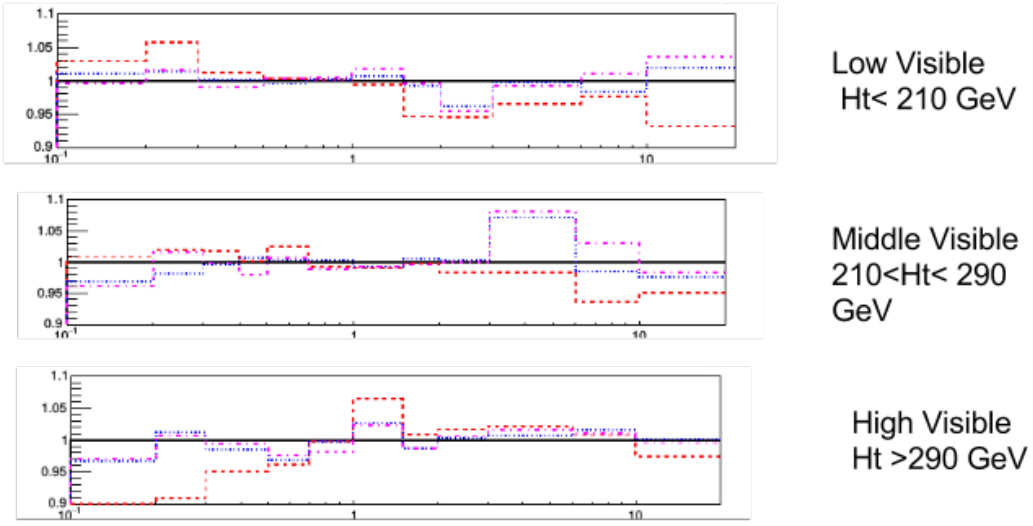

Figure 7: The ratio plot of parton level values of ϕ^* spilt along high, middle and low values of the visible H_t showing different behaviour for different simulations.

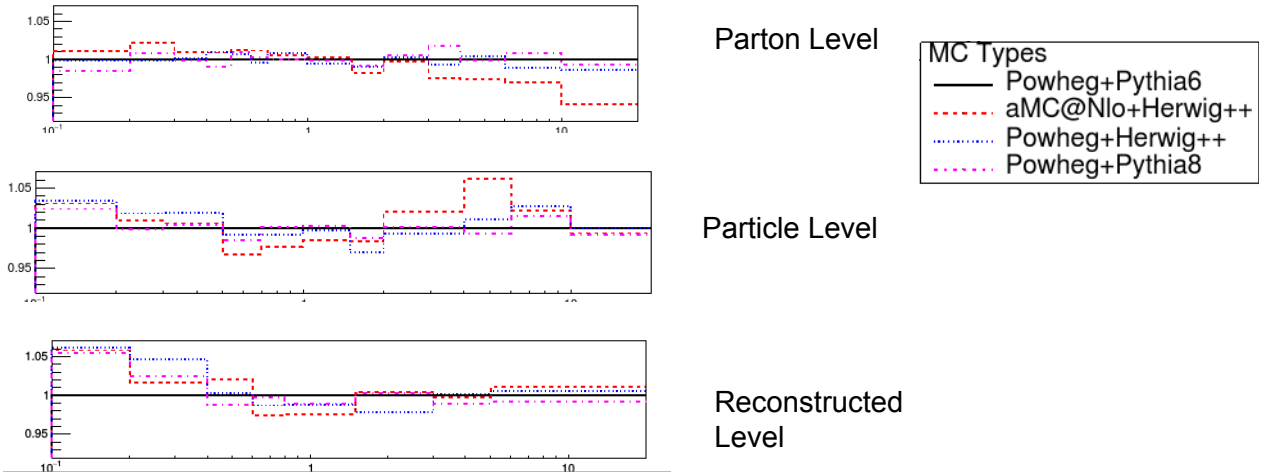

Figure 8: Comparison of ϕ^* behaviour for parton, particle and reconstructed level. Particle and recon level have much larger differences between MCs.

The similarity between the particle and reconstructed levels is due to the good resolution of angular variables in the detector so the value of ϕ^* is not significantly altered by detector effects. Figure 9 is a migration matrix for ϕ^* for the particle against reconstructed level. The figure confirms this good agreement between the two levels as there is is very little smearing. For comparison figure 10 shows a migration matrix of parton against reconstructed level showing less agreement between the two levels. Both figures 9 and 10 used from the Powheg+Pythia6 simulation and the other simulations have very similar behaviour. The varying bin sizes in each histogram are to maximise the resolution of each plot.

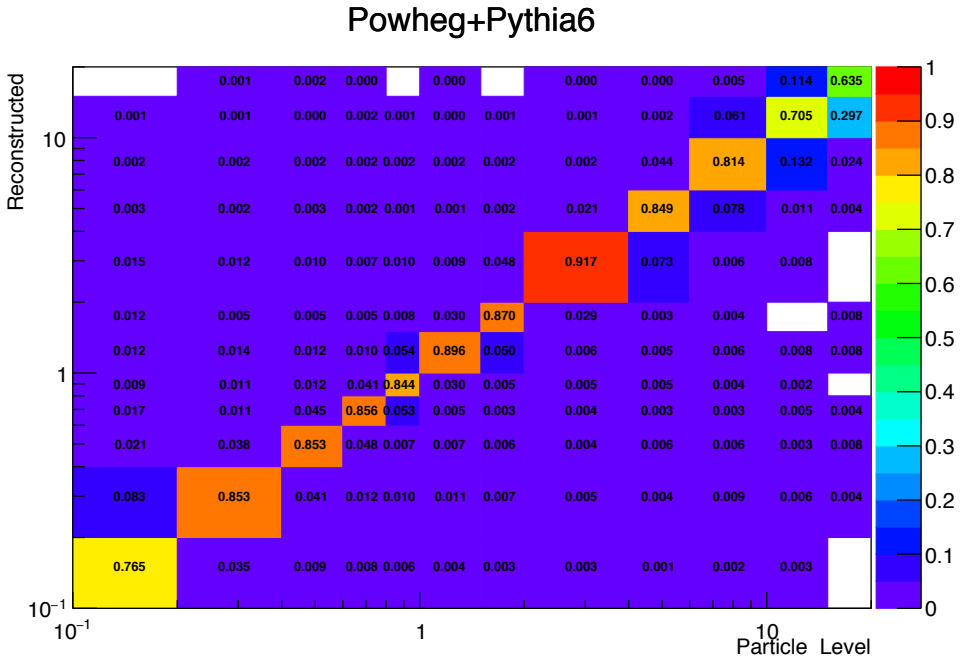

Figure 9: A log-log normalised migration matrix of ϕ^* for the particle versus reconstructed level. The data is an good agreement particularly at low ϕ^* where most of the statistics are.

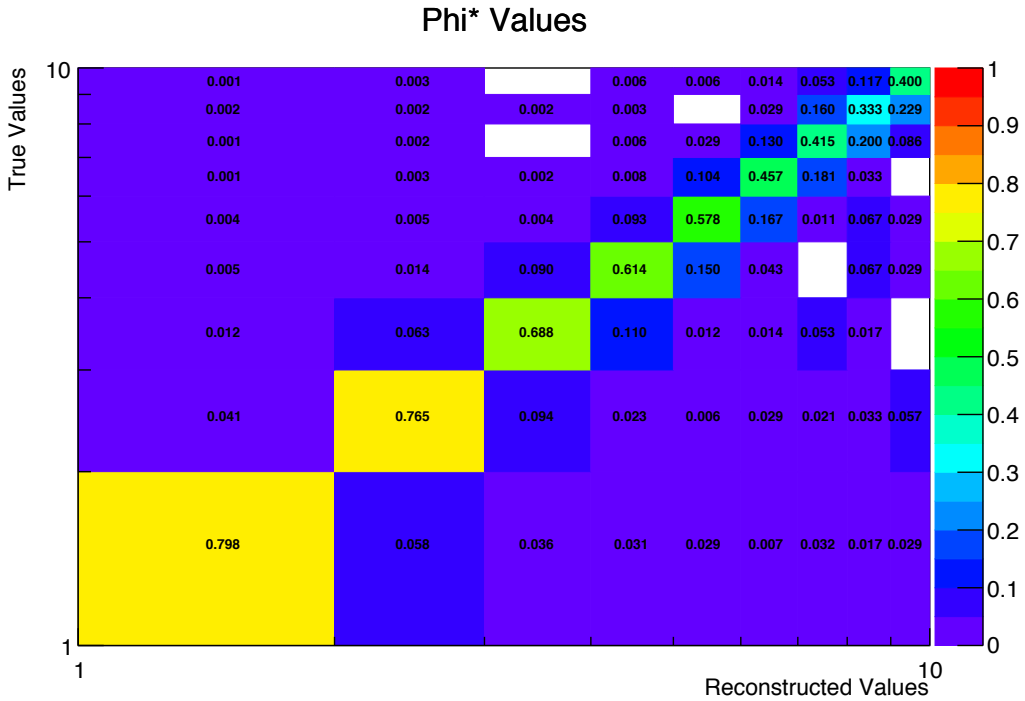

Figure 10: A log-log normalised migration matrix of ϕ^* for the parton versus reconstructed level. There is some symmetric smearing.

4. Conclusion and Extensions

In this project two variables M_{lb} and M_{T2} were used to resolve pairing ambiguities at the particle and reconstructed level in Monte Carlo data for the event $t\bar{t} \to \ell^-\ell^+b\bar{b}\nu\bar{\nu}$. It was found that pairing with both variables was an improvement on pairing with M_{lb} alone and the optimum selection gave a efficiency of 0.561 and a purity of 0.791. Further work could optimise the position of the cuts for MC used here instead of taking them from [3].

The variable ϕ^* was considered for differing MC simulations at parton, particle and reconstructed level. At parton level it was found that the maximum difference between MC simulations was 5% between Powheg+Pythia6 and aMC@NIO+Herwig++. Variation in differences was found when the parton level was spilt along H_t^{vis}, confirming at ϕ^* is sensitive to p_T. At particle and reconstructed level the difference between MCs was also at a maximum of 5%. There was also good agreement between the particle and reconstructed level as show in figure 9. These results suggest that ϕ^* is a good variable to tune Monte Carlos to as it is sensitive to p_T and generates differences in Monte Carlo data that are large enough that the measured data could be sensitive to the differences. An extension of this work could be to repeat this work with at much large number of MC statistics and with optimised binning to improve the resolution of the ratio plots to better extract information on the behaviour of the MCs.

References

[1] Weinzierl, Stefan.
"Introduction to monte carlo methods."
arXiv preprint hep-ph/0006269
2000.

[2] Aad, Georges, et al.
"The ATLAS experiment at the CERN large hadron collider."
Journal of Instrumentation
2008.

[3] Baringer, Philip, et al.
"Revisiting combinatorial ambiguities at hadron colliders with M_{T2}"
Journal of High Energy Physics
2011.

[4] Vesterinen, M., and T. R. Wyatt.
"A novel technique for studying the Z boson transverse momentum distribution at hadron colliders."
Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment
2009.