# A Deep Neural Network for Missing Transverse Momentum Reconstruction in ATLAS

**Matthew Leigh**

University of Cape Town, South Africa

E-mail: matthew.leigh@cern.ch

**Abstract.** The missing transverse momentum ($\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$) of a proton-proton ($pp$) collision is an important observable as it serves as an experimental proxy for the total momentum carried away by undetected particles in the plane perpendicular to the beam line. Precisely measuring $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ is critical for the understanding many physical processes which take place at the Large Hadron Collider. The ATLAS experiment currently utilises several algorithms to reconstruct the $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$. These proceedings describe the development of a novel algorithm based on a deep neural network to improve $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction. The network reconstruction had resolution of 22.7 GeV when tested in simulated $t\bar{t}$, improving over than the next best performing algorithm which had a resolution of 27.3 GeV. The new estimate was more robust to an increase in pileup and outperformed all other methods across the full pileup range. The network was also demonstrated to generalise well in real $Z$ events, improving the $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution from 15.53 to 11.29 GeV.

## 1. Introduction

The ATLAS detector [1] is an example of a hermetic detector. It was thus designed to observe nearly all possible decay products produced in the high-energy $pp$ collisions provided by the Large Hadron Collider (LHC) at CERN. However, neutrinos and many theorised particles in beyond the Standard Model physics are weakly interacting and are able travel through the detector without leaving any measurable signal. Since the total linear momentum of the colliding protons in the transverse plane is negligible, the presence of these undetectable particles can be inferred by the resultant transverse momentum imbalance of all observable decay products emerging from the collision. The negative vectorial sum of the visible momenta in the transverse plane, $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$, therefore serves as an experimental proxy for the net transverse momentum of undetected particles. This is an essential part of many ongoing physics analyses at the LHC [2,3], and played a role in the discovery of the Higgs boson in 2012 [4]. Reconstructing $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ requires the output of all detector subsystems, as well as the most complete representation of the hard-scatter[1] which gets obscured by the presence of additional $pp$ interactions in the same bunch crossing, known as "*pileup*". Therefore, the performance of current $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction algorithms are expected to get worse as the luminosity of the LHC increases over the next few years. Monte-Carlo (MC) simulations are used in most analyses to provide a baseline from which one can analyse real data captured by ATLAS. To produce this simulated data, the underlying parton interaction is generated, soft radiation and hadronization modelling is applied, before the passage of the

[1] The hard-scatter is defined in ATLAS as the reconstructed vertex with the highest $\sum(p_{\mathrm{T}}^{\mathrm{track}})^2$.

resultant stable particles through a detector simulation, to estimate the signals they would produce [5]. The event reconstruction process applied to MC is identical to the one applied to real data. In this project we extracted and stored the true transverse momentum of all non-interacting particles in each MC event before the detector simulation, called $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss,\ true}}$. A deep neural network [6] was then trained to learn a mapping from a set of reconstructed observables to $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss,\ true}}$. This mapping was applied and evaluated on real and MC datasets, all of which are orthogonal from those used to train the network.

## 2. The ATLAS Detector

The ATLAS detector at the LHC is a general purpose particle detector with nearly $4\pi$ solid angle coverage and a nominal forward-backward symmetry. It consists of an inner tracking detector (ID) surrounded by a superconducting solenoid magnet which provides a 2 T magnetic field. Encompassing the ID are electromagnetic (EM) and hadronic calorimeters, and a muon-spectrometer (MS). The ID consists of a pixel detector, a semiconductor tracker and a transition radiation tracking detector. It provides tracking information in a pseudo rapidity range of $\eta < |2.5|$[2]. The EM calorimeter provides high granularity energy measurements over a range of $\eta < |3.2|$. Steel-scintillating hadronic calorimeters provide central coverage within $\eta < |1.7|$. End-cap regions of the detector contain additional calorimeters up to $\eta < |4.9|$. The MS is the outermost layer of the ATLAS detector and features three large air-core toroidal superconducting magnet systems with eight coils each. It contains precision tracking chambers covering $\eta < |2.7|$. ATLAS data-taking utilises a two-level trigger system. The Level-1 trigger is hardware-based and reduces the event rate to around 100 kHz. The Level-2 trigger is a software-based high level trigger, which further reduces the rate to approximately 1 kHz.

## 3. Missing Transverse Momentum Reconstruction at ATLAS

Over the past few years, different algorithms have been developed to reconstruct $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ at ATLAS [7]. ATLAS currently offers several variants for physics analyses with different requirements, several of which are used in this project [8]. Most definitions are object based and characterised by two contributions. The first one is from "*hard-event*" signals, which are signals associated with identified and calibrated physics objects, such as fully reconstructed jets, photons, electrons, and muons. Dedicated rejection procedures are carried out to ensure that all contributing objects were reconstructed from mutually exclusive detector signals. The second contribution to $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ is from "*soft-event*" signals. These are the leftover detector signals that were not used to reconstruct the objects mentioned above. Most algorithms use a Track Soft Term (TST), which is created purely from unused particle tracks in the ID which are associated with the hard-scatter. The tracks are required to have passed high-quality reconstruction requirements and have $\eta < |2.5|$ and $p_T > 0.4$ GeV. An example of this algorithm is shown in Equation 1, which show the negative vectorial summation of the terms:

$$\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}} = -\sum_{\substack{\text{selected}\\\text{electrons}}} \mathbf{p}_{\mathrm{T}}^{e} - \sum_{\substack{\text{selected}\\\text{muons}}} \mathbf{p}_{\mathrm{T}}^{\mu} - \sum_{\substack{\text{selected}\\\text{photons}}} \mathbf{p}_{\mathrm{T}}^{\gamma} - \sum_{\substack{\text{selected}\\\text{jets}}} \mathbf{p}_{\mathrm{T}}^{\mathrm{jet}} - \sum_{\substack{\text{unused}\\\text{tracks}}} \mathbf{p}_{\mathrm{T}}^{\mathrm{track}} \tag{1}$$

$$\underbrace{\underbrace{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss},e}}_{\text{selected electrons}} + \underbrace{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss},\mu}}_{\text{selected muons}} + \underbrace{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss},\gamma}}_{\text{selected photons}} + \underbrace{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss,\ jet}}}_{\text{selected jets}}}_{\text{hard terms}} + \underbrace{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss,\ soft}}}_{\text{soft term}}$$

[2] ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point in the center of the detector. The z-axis lies along the beam pipe, the x-axis points towards the center of the LHC ring, and the y-axis points vertically upwards. Cylindrical coordinates are used in the transverse plane where $\phi$ is the azimuthal angle around the z-axis. The pseudorapidity is defined in terms of the polar angle as $\eta \equiv -\log\tan\frac{\theta}{2}$, where $\theta$ is the polar angle off the z-axis.

The total transverse energy in the detector, $\Sigma E_{\mathrm{T}}$, is a scalar which quantifies the total event activity. This is a crucial variable for understanding the scale and resolution of $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$. It uses the same terms, and follows the same overlap removal procedures, but is created using the scalar sum of the various transverse momenta. Reconstruction of the magnitude of the missing transverse momentum ($|\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}|$) suffers from a positive observation bias towards non-vanishing values, due to it being strictly positive by definition. This is most noticeable in final states without genuine missing transverse momentum. This means that the recorded $|\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}|$ is often non-zero even when there are no undetected particles due to mismeasurements.

### *3.1. Hard Terms*

Muons are identified by matching an ID track with an MS track or segment, and are required to have $p_T > 10$ GeV to be included in the muon term for $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$. Electrons are reconstructed from clusters in the EM calorimeter associated with an ID track and must also have $p_T > 10$ GeV to contribute to $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ or $\Sigma E_{\mathrm{T}}$. Photons are identified from their distinctive electromagnetic showers in the calorimeters. In this project, photon candidates are required to satisfy a set of tight criteria to reduce backgrounds and must have $p_T > 25$ GeV. Jets are reconstructed from clusters of topologically connected calorimeter cells using the anti-kt algorithm [9] with a radius parameter of $R = 0.4$. All jets are first required to have $p_T > 20$ GeV after calibration and $\eta < |4.5|$. Jets are further decorated using a tagging algorithm to select jets likely from emerging from the hard-scatter, known as jet-vertex-tagging (JVT) [10]. The JVT value attributed to a jet ranges from 0 (likely from pile-up) to 1 (likely from the hard-scatter). All central jets ($\eta < |2.5|$) are required to have either a JVT value greater than 0.59 or $p_T > 60$ GeV.

### *3.2. Alternate Definitions of* $\mathbf{E}_T^{miss}$

Five different algorithms are used and compared in this work. The first three are constructed as mentioned above and only differ in their treatment of forward jets. *Loose* $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ includes contributions from any forward jet with $p_T > 20$ GeV. *Tight* $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ removes all forward jets with $p_T < 30$ GeV. *FJVT* $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ once again uses the same terms as Loose, but all forward jets with $20 > pT > 50$ GeV are required to pass the "Loose" FJVT criteria [11]. *CST* $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ uses the same jet selection as the Tight, but replaces the TST with a more inclusive soft term created from unused calorimeter clusters [12], and although noise suppression is applied, no additional pileup suppression techniques are used. The fifth and final algorithm used in this work is somewhat distinct from the others. *Track* $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ is not object based and only employs ID tracks. It is therefore inherently more immune to pileup but ignores all neutral particles.

## 4. The Deep Neural Network

The main goal of this project was to produce a new definition, *Network* $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$, derived from a non-linear combination of all other algorithms. The form of this combination would be adaptable and unique for each event to maximise accuracy. The model chosen for this work was a dense feed-forward artificial neural network (ANN) [6]. The network did not receive any raw detector signals or even individual particles as inputs. This was due to the desired scope of the project, which was to provide a final level correction to the $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ of an event only, and not to redefine object selection or identification. 65 variables were chosen to be inputs for the neural network. These included the outputs of the 5 mentioned algorithms, their unique hard and soft terms, and additional variables which might indicate the accuracy of these definitions. These additional variables included the number of reconstructed vertices, which indicates the amount of pileup, the amount of forward jet activity and object based $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ significance [13].

*4.1. Network Training*

The network was trained on MC simulated datasets where the target variable on an event by event basis was $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss,\ true}}$. The dataset was made up of 8 million $t\bar{t}$, $Wt$, $W\bar{t}$ and diboson ($WW$, $WZ$, $ZZ$) processes. Underlying events were generated using Powheg-Box [14] and subsequent parton showers and hadronisation were handled by Pythia [15].

Over the course of this work, more than 3000 different networks architectures were tested and compared. Features that were varied included the network depth, width, hidden layer activation function, loss metric, optimising algorithm, learning rate, and regularisation technique. The final and most accurate network had five hidden layers, each with 1000 neurons featuring a Swish activation function [16]. After each non-linearity a dropout layer with $p = 0.2$ was applied. Gradient descent was performed by the ADAM optimiser [17] based on the Huber loss function [18].

## 5. Performance on $Z \to ee$ and $t\bar{t}$ events

*5.1. Event Selection*

Events involving $t\bar{t}$ production provide a good topology to measure $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss,\ true}}$ reconstruction performance in environments with large jet multiplicities. Studies in this region involved MC simulation only. The absence of background allowed for very loose selection criteria, so only standard ATLAS event quality checks were performed [19].

The $Z \to ee$ final state is ideal for the evaluation of $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ performance in real events. $Z$ boson production is relatively abundant at the LHC, has a well understood topology, and can be selected with high signal to background ratios. In this channel neutrinos are produced only through very rare heavy-flavour decays in the hadronic recoil. In standard ATLAS $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ evaluations [7, 8, 12], this process is associated with not having any genuine missing transverse momentum. Events were selected for this channel if they fired a single lepton trigger and contained exactly two oppositely charged electrons with $p_T > 25$ GeV, which have an invariant mass between 76 GeV and 106 GeV. Data used in this channel was recorded by the ATLAS experiment in 2017 at $\sqrt{s} = 13$ TeV, and had a total integrated luminosity of 44.3 fb$^{-1}$. All events had to pass standard detector quality assessment criteria. Monte-Carlo samples of $Z \to ee$ were generated using Sherpa [20]. Background contributions from $t\bar{t}$, single-$t$ and diboson events were generated using Powheg, with Pythia handling the parton shower.

*5.2. Results*

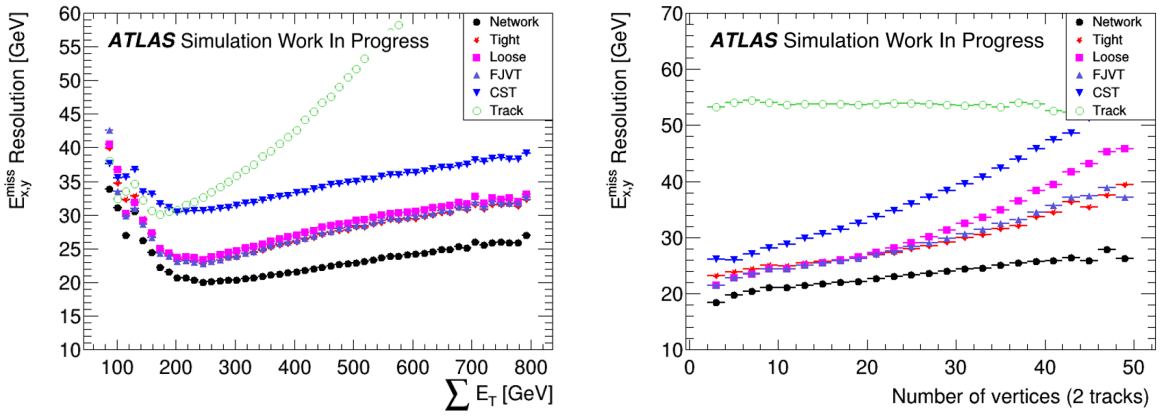
The resolution is determined by the width of the combined distribution of the differences between the measured $E_{x(y)}^{\mathrm{miss}}$ and the corresponding components of the true $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss,\ true}}$. The width is taken from the of the root mean square error (RMSE) of the distributions,

$$\mathrm{RMSE} = \begin{cases} RMS(E_{x(y)}^{\mathrm{miss}} - E_{x(y)}^{\mathrm{miss,\ true}}) & t\bar{t} \text{ sample } (E_T^{\mathrm{miss,\ true}} > 0) \\ RMS(E_{x(y)}^{\mathrm{miss}}) & Z \to ee \text{ sample } (E_T^{\mathrm{miss,\ true}} \approx 0) \end{cases} \tag{2}$$

The performance of the different algorithms was compared to each other in simulated $t\bar{t}$ events. In Figure 1a, the resolution is plotted with respect to the $\Sigma E_{\mathrm{T}}$, which can be taken as a measurement of the hardness of the interaction, providing a useful scale for the evaluation of the resolution. The network outperforms all other estimates and shows less degradation with increased event activity. Furthermore, from the plots in Figure 1b, the network shows an excellent resilience to pileup, based on the consistently greater resolution with respect to the number of reconstructed vertices.

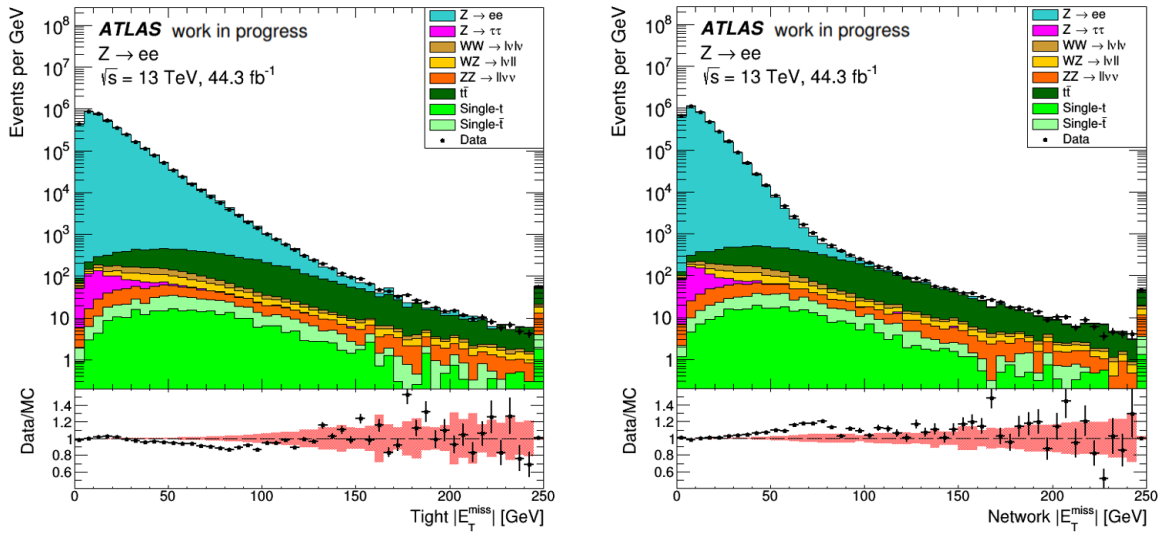
Data vs MC simulation comparisons for the Tight and the Network $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ algorithms in the $Z \to ee$ channel are shown in Figure 2. The error on the ratio points show the uncertainty of the MC samples, accounting for uncertainty in luminosity, cross section and statistics. Systematic

uncertainties have yet to be added to this work. Figure 2b shows good agreement between data and MC in the bulk and the distributions display same overall shape. This shows that the network, which was trained exclusively on MC datasets, is still able to generalise well to real data. Over 99% of events in the MC $Z \to ee$ contribution, which significantly dominates all other processes, have $|\mathbf{E}_{\mathrm{T}}^{\mathrm{miss,\ true}}| < 3$ GeV. Both the Tight and Network $|\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}|$ distributions show a tail in $Z \to ee$ events extending beyond this value, indicating that they both suffer from the aforementioned observation bias. However, the width of this tail is greatly reduced when using the network, leading to a greater distinction between the signal and background processes. The average resolution using all algorithms applied to both datasets are shown in Table 1.




(a) (b)

**Figure 1.** Comparisons of the $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution using different algorithms with respect to the $\Sigma E_{\mathrm{T}}$ (a) of the event and the number of reconstructed vertices (b).




(a) (b)

**Figure 2.** Data to Monte-Carlo comparisons of the magnitude of the $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction using the Tight (a) and the Network (b) definitions in the $Z \to ee$ channel.

| | $t\bar{t}$ | $Z \to ee$ |
|---|---|---|
| Tight | 27.30 | 15.53 |
| Loose | 28.20 | 18.06 |
| FJVT | 27.48 | 16.55 |
| CST | 34.31 | 24.24 |
| Track | 53.96 | 18.48 |
| **Network** | **22.70** | **11.29** |

**Table 1.** The $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution, measured in GeV, of the different algorithms measured on the the simulated $t\bar{t}$ and real $Z \to ee$ datasets.

## 6. Conclusions

The measurement of $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ is an important contribution to many different physics analyses. The high luminosity at the LHC, however, means that there are large amounts of pileup interactions which degrade its reconstruction quality. Therefore there are a number of existing algorithms to reconstruct and estimate $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ depending on the topology of the event. A neural network was trained to combine these different definitions and produced a reconstruction method which lead to a greater $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution in both data and MC, as well as an excellent stability with pileup. While more quantitative analysis still needs to be performed on the resolution improvements, this project reflects the ongoing work to improve reconstruction performance at ATLAS.

## 7. References


[1] The ATLAS Collaboration 2008 *Journal of Instrumentation* **3** S08003
[2] The ATLAS Collaboration 2019 *Phys. Rev. D* **100**(1) 012006
[3] The ATLAS Collaboration 2018 *Phys. Rev. D* **98**(9) 092002
[4] The ATLAS Collaboration 2012 *Physics Letters B* **716** 1 – 29 ISSN 0370-2693
[5] The ATLAS Collaboration 2010 *The European Physical Journal C* **70** 823–874 ISSN 1434-6052
[6] Goodfellow I and Bengio Y 2016 *Deep Learning* (MIT Press) `http://www.deeplearningbook.org`
[7] The ATLAS Collaboration 2018 *The European Physical Journal C* **78** 903 ISSN 1434-6052
[8] The ATLAS Collaboration 2018 $E_{\mathrm{T}}^{\mathrm{miss}}$ performance in the ATLAS detector using 2015-2016 LHC p-p collisions Tech. Rep. ATLAS-CONF-2018-023 CERN Geneva
[9] Cacciari M, Salam G P and Soyez G 2008 *Journal of High Energy Physics* **2008** 063–063
[10] The ATLAS Collaboration 2016 *The European Physical Journal C* **76** 581 ISSN 1434-6052
[11] The ATLAS Collaboration 2017 *The European Physical Journal C* **77** 580 ISSN 1434-6052
[12] The ATLAS Collaboration 2017 *The European Physical Journal C* **77** 241 ISSN 1434-6052
[13] The ATLAS Collaboration 2018 Object-based missing transverse momentum significance in the ATLAS detector Tech. Rep. ATLAS-CONF-2018-038 CERN
[14] Nason P 2004 *Journal of High Energy Physics* **2004** 040–040
[15] Sjstrand T, Mrenna S and Skands P 2006 *Journal of High Energy Physics* **2006** 026–026
[16] Ramachandran P *et al.* 2017 Searching for activation functions (*Preprint* `1710.05941`)
[17] Kingma D P and Ba J 2014 *arXiv preprint arXiv:1412.6980*
[18] Huber P J 1964 *Ann. Math. Statist.* **35** 73–101
[19] The ATLAS Collaboration 2010 Data-Quality Requirements and Event Cleaning for Jets and Missing Transverse Energy Reconstruction with the ATLAS Detector in Proton-Proton Collisions at a Center-of-Mass Energy of $\sqrt{s} = 7$ TeV Tech. Rep. ATLAS-CONF-2010-038 CERN
[20] Gleisberg T *et al.* 2009 *Journal of High Energy Physics* **2009** 007–007