Dynamical jet vetoes for suppression of $\mathbf{t\bar{t}}$ background in inclusive $\mathbf{W^+W^-}$ production

Oscar Jakobsson
School of Physics and Astronomy, University of Manchester
The ATLAS Group, DESY Summer Student Programme 2018 (Hamburg)

Supervisors: Mateusz Dyndal and Valerie Lang
(Dated: September 5, 2018)

Precise measurements of the production of W boson pairs in association with jet production is an important test of theoretical predictions from pQCD, as well as providing a potential sector in which to search for new physics. In this report, we investigate the use of dynamical jet vetoes [1] – whereby jet vetoes are set on an event-by-event basis – for suppression of $t\bar{t}$ and Wt background in the decay chain $W^+W^- \to \ell^\pm\nu_\ell\ell^\mp\nu_\ell$ in final states with one or several associated jets. We find that, by setting the jet veto scale as a function of the transverse momentum (P_T) of the leading or subleading lepton, top background is able to be suppressed in both the high and low P_T(jet) regions. In addition, the effect of improved b-tagging efficiency is investigated through extrapolation of the current set of JetBTag working points.

1. INTRODUCTION

The lack of evidence at the large hadron collider (LHC) of new physics beyond the standard model (SM) at the TeV scale continues to reinforces the need for more precise SM measurements in the pursuit of subtle non-SM phenomena. The large data sets at $\sqrt{s} = 13$ TeV from the ATLAS and CMS detectors during run II of the LHC will offer unprecedented insight into this high energy regime of the SM [2]. In this regard, the study of scattering processes involving vector boson pair production is of particular interest since they provide an important probe into the electroweak sector of the SM. Precise measurements of di-boson production examines the non-Abelian structure of SM electroweak theory through gauge boson self-interactions involving WWZ and $WW\gamma$ vertices. In addition, the structure of the triple gauge couplings (TGCs) in these vertices, which arises from $SU(2)_L \times U(1)_Y$ gauge invariance and its associated spontaneous symmetry breaking, may be sensitive to the contributions from heavy new particles beyond the SM [3]. These so-called anomalous TGCs (aTGCs) have been searched for in electroweak interactions in previous analyses by the ATLAS and CMS experiments via the production of two oppositely charged W bosons decaying leptonically into electrons, muons and neutrinos [4, 5]. Recent analyses also include measurements of the W^+W^- production cross section at $\sqrt{s} = 13$ TeV by the ATLAS experiment [6]. In common for all these studies is the vetoing of WW events where any jets are present. This restriction considerably limits the phase space available for analysis, but is nevertheless necessary due to the excessive background of events involving top quarks. The sizeable contribution of top quarks to the background in $WW+$jets events is made apparent in figure.1. The reason for this dominant background from top quarks is that the same leptonic final-state configuration can be obtained from top-quark ($\bar{t}t$, Wt and $W\bar{t}$) production, where the top quarks decay into W bosons and bottom quarks, which in turn hadronise and give rise jets. This is evident if we look at the decay chains

$$W^+ + W^- + 2\text{jets} \qquad \text{and} \qquad t\bar{t} \to bW^+ + \bar{b}W^-$$
$$W^- \to \bar{\nu}_e + e^-, \quad W^+ \to \nu_\mu + \mu^+ \qquad\qquad W^- \to \bar{\nu}_e + e^-, \quad W^+ \to \nu_\mu + \mu^+$$

which both yield the same detector signature of leptons, two jets and missing momentum from neutrinos. Apart from providing a larger phase space in which to perform analyses, the inclusion of jets would also allow for the study of the decay of W bosons with high transverse momentum. This is because the production of W bosons in association with jets allows for the W bosons to recoil of the jets, resulting in P_T boosted W bosons, in contrast to the zero-jet events which produce near stationary W bosons. Investigations of this phenomena have been done in a previous analysis using data collected at $\sqrt{s} = 8$ TeV by the ATLAS detector [7]. However, the techniques used in the 8 TeV

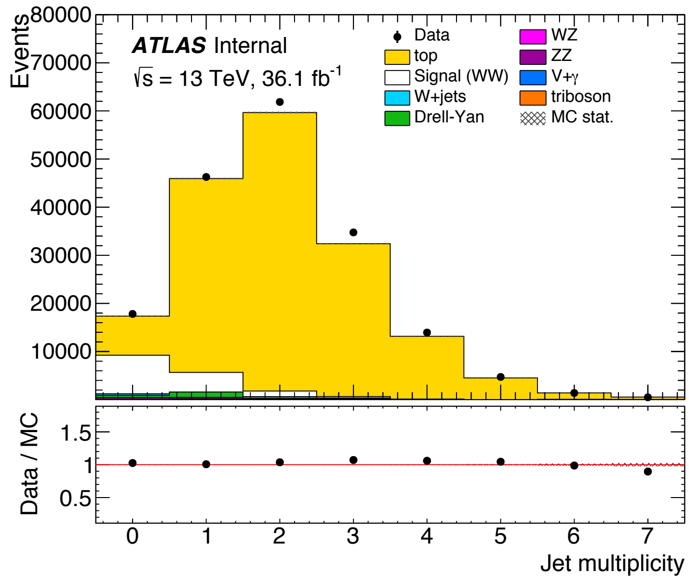

FIG. 1. Jet multiplicity distributions for WW signal and top background events before any jet-vetoes have been applied [8].

analysis may not be effective for the data sets at $\sqrt{s} = 13$ TeV because of the increase in top-quark production at this higher center-of-mass energy. Hence the top background events will form an ever more considerable background in the upcoming analyses of WW production, an issue which needs to be addressed if high precision measurements are to be done successfully. Consequently, it is the purpose of this report to investigate the effect of using dynamical jet vetoes and b-tagging to reduce top background in the $WW +$jets events.

2. ANALYSIS

The current analysis by the ATLAS collaboration, at $\sqrt{s} = 13$ TeV using an integrated luminosity of 36.1fb^{-1}, includes the measurement of the W^+W^- production cross-section, both fiducial and differential, using the $W^+W^- \rightarrow e^{\pm}\nu_e\mu^{\mp}\nu_\mu$ zero-jet decay channel. The jet vetoes applied to obtain the zero-jet sample are two-fold: Firstly, backgrounds from top-quark production ($t\bar{t}$, Wt and $W\bar{t}$) are identified by their characteristic b-jet signature and subsequently rejected based on the events' jetBTag values – where each jet in the event gets a jetBTag $\in [-1, 1]$ which corresponds to the likelihood of the jet being a b-jet. The jetBTag cut-off is currently set at jetBTag $= 0.1758$ which corresponds to a b-jet detection efficiency of 84.95% [9], such that jets with jetBTag > 0.1758 are categorized as b-jets. Secondly, jet vetoes are applied such that events in which any jet with a transverse momentum exceeding 35 GeV are rejected, where the 35 GeV cut-off comes from optimizing the statics in the current zero-jet analysis [8]. Due to the associated uncertainties from these selections procedures – including detector inefficiencies, particle pileup and vertex reconstruction – scaling factor derived from comparison of MC and experimental data are applied to each event. Therefore, instead of the weight of an event being one it is instead given by `weight` $\times$ `BTagweight`.

This restricted phase space significantly excludes top events, resulting in a sample with a signal-to-background ratio of 2.4, in which the cross-sections are computed from 7690 prospective WW signal events and 3120 prospective top background events. It is however desirable to open up the zero-jet requirement to allow for measurements of phenomena associated with multi-jet events. To this end, the analysis in this report is performed on ntuples [1] from MC simulations at NLO using `Powheg+ Pythia8`. The WW signal is simulated from quark-quark and gluon-gluon initial states with W bosons decaying leptonically in multi-jet final states. The dominant background is simulated through the $t\bar{t}$, Wt and $W\bar{t}$ production modes with the same multi-jet leptonic final states.

In this analysis a p_T of > 25 GeV will be used as identifying criterion for the presence of a jet in the events. This differs from the current analysis' value of 35 GeV, which is optimized for the zero-jet analysis, and will thus not be used except for in the purpose of comparing to the current analysis. Furthermore, since the effect of varying the jetBTag cut-off value is investigated in this report, the BTagweight, which is only available for the official working

[1] See appendix section I for specific files used in the analysis.

weight cut	b-jet efficiency [%]
0.9977155	29.99
0.9769329	50.05
0.934906	60.03
0.8244273	69.97
0.645925	76.97
0.1758475	84.95

TABLE I. MV2c10 tagger: jetBTag cut-off values and associated b-tagging efficiencies. From "Expected flavour tagging performance in release 20.7" [9].

points, is omitted throughout this analysis.

By introducing jets in the analysis the previous statistical and systematic uncertainties are going to change. To fully judge the outcome of the analysis presented here one would have to re-compute these parameters, this is however beyond the scope of the work presented in his report. Instead, we use the signal-to-background ratio

$$S/B = \frac{\text{Number of WW events}}{\text{Number of Top events}} \ ,$$

as a measure of the relative success of our approach.

3. B-TAGGING

A straightforward approach to reduce the top quark background is to improve the identification of b-jets originating from single top quarks. Since each top quark decays almost exclusively via $t \to bW^+$ or $\bar{t} \to \bar{b}W^-$ we can categorize each detected b-jet as coming from a top event and reject the event accordingly. In anticipating of higher b-tagging efficiencies it is interesting to estimate the improvements we may expect to gain in top background suppression. This is done by taking the current working points together with the MC files, computing the b-tagging efficiency and then extrapolate into higher efficiencies to find what jetBTag values we might expect to become available.

The current working points are shown in table 1. To compute the b-tagging efficiency we begin by considering the efficiency by which we can identify if a single jet is a b-jet. To do this we use the Wt and $W\bar{t}$ MC data, such that we know that each jet we detect is most likely a b-jet from a top quark. Efficiencies are computed for jetBTag cut-off values in steps of 0.01 in the range $[-1, 1]$ by

$$\eta_b(\text{1-jet}) = \frac{\text{Number of WW+1 b-jet events}}{\text{Number of WW+1 jet events}} \ ,$$

where the number of b-jets events are defined as the number of jets with jetBTag above the given threshold and the number of jet events are defined as the number of jets above 25 Gev. Due to the simplicity and incomprehensiveness of this approach this efficiency will be lower than the ones corresponding to the current working points. However, by comparing the computed efficiencies to the current working points, a constant scaling factor is found and used to scale the computed efficiencies, the results of which is shown in figure 2. Extrapolation of the scaled efficiencies then provides estimates for the 90% and 95% working points at jetBTag$= -0.29$ and jetBTag$= -0.57$, respectively. We also note that the extrapolated 100% working point occurs at jetBTag$= -0.77$ and not at jetBTag$= -1$ as per the definition of the jetBTag variable. Although a value of -1 is unattainable in practice, the current difference for the 100% working point should be used to question the accuracy of the 95% working point.

The same analysis is also carried out on the $t\bar{t}$ MC data samples with two final state jets, where the efficiency

$$\eta_b(\text{2-jet}) = \frac{\text{Number of WW+1 b-jet events} + 2\,(\text{Number of WW+2 b-jet events})}{\text{Number of WW+2 jet events}} \ ,$$

is found by selecting all two-jet events and calculating the number of jets with jetBTag above the threshold. From the resulting scaled efficiencies, estimates for the 90% and 95% working points are found to be jetBTag$= -0.28$ and jetBTag$= -0.55$, respectively, with an extrapolated 100% working point occurring at jetBTag$= -0.72$, all of which are close to the values found for the one-jet Wt and $W\bar{t}$ cases.

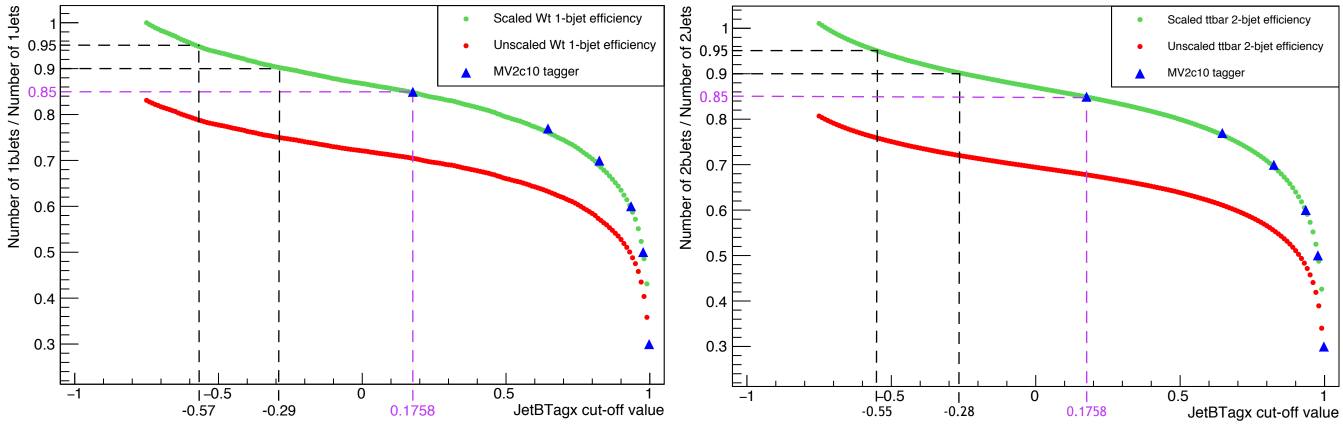

FIG. 2. Extrapolated b-tagging efficiencies for (left) Wt events and (right) $t\bar{t}$ events. The current working-point benchmarks for the MV2c10 tagger [9] are shown as well. Scaling factors are found to be 1.203 ± 0.023 and 1.252 ± 0.009 for Wt and $t\bar{t}$ respectively.

3.1 Signal-to-background improvements with b-tagging efficiencies

In this section we present the expected improvements in signal-to-background ratio as higher b-tagging efficiencies become available. For reference, the improved S/B for the current analysis' zero-jet sample is shown in figure 3, from which we see that a $S/B = 2.6 - 2.7$ is to be expected at the 90% working point for jetBTag=$-0.2 - 0.3$. Figure 3 further shows that this S/B improvement would only come at the cost of reducing the WW signal events from 7690 to 7600. In addition, in figure 4 we show the resulting S/B if one decides to simply remove the jet P_T veto of 35 GeV, which shows that at the extrapolated 90% working point the $S/B = 1.1$.

However, the removal of the jet veto would first and foremost enable the study of the jet multiplicities in isolation. It is therefore instructive to look at the S/B in the $WW + 1 -$jet and $WW + 2 -$jet production modes. The results of which are shown in figure 5 and 6, where a threshold of $P_T = 25$ GeV has been used for jet-identification.

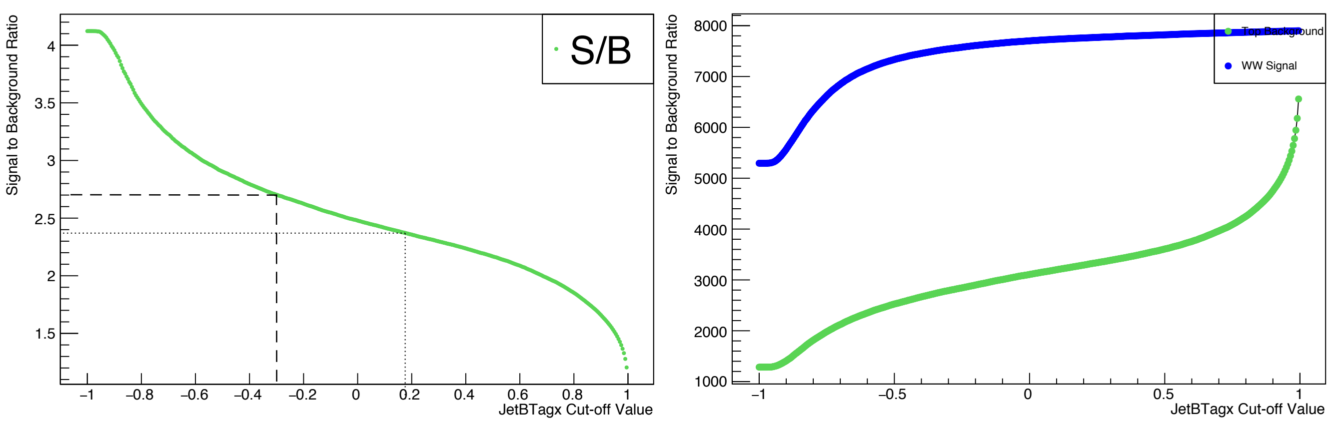

FIG. 3. For reference – current analysis – no jets: Signal-to-background ratio and absolute event counts as a function of jetBTag value for strict jet veto at 35GeV. The dotted lines show the values at the current working point as well as the extrapolated 90% working point. Value at jetBTag=0.1758 differs from S/B ratio in the internal documentation due BTagweight not being used.

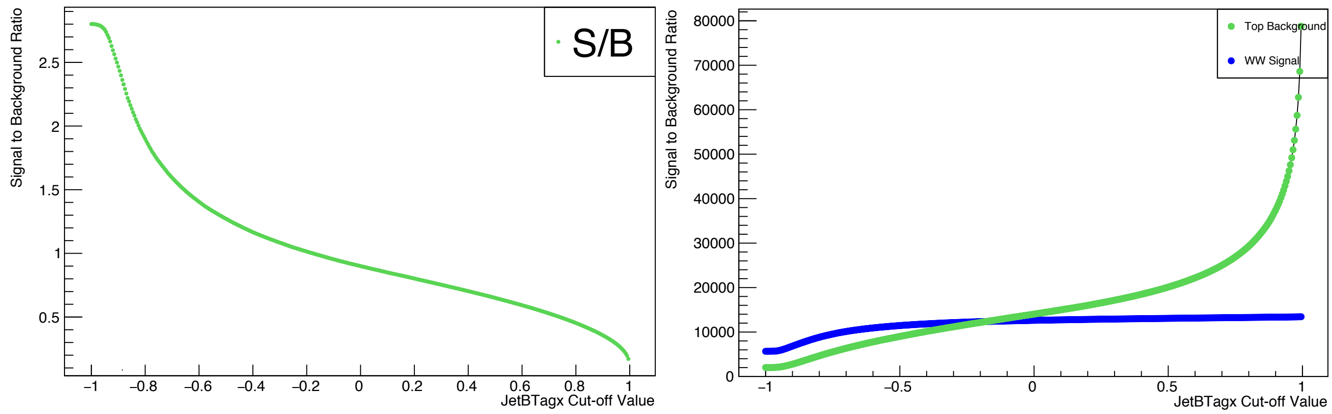

FIG. 4. For reference – no bjets: Signal-to-background ratio and absolute event counts as a function of jetBTag without jet P_T veto.

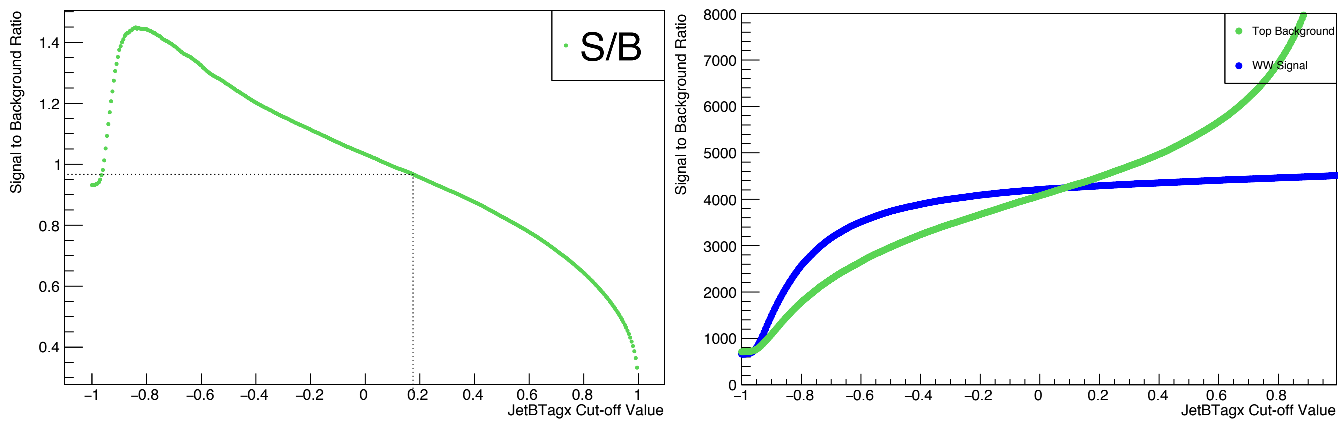

FIG. 5. $WW + 1$ jet: Signal-to-background ratio and absolute event counts as a function of jetBTag value for strict jet veto at 25GeV.

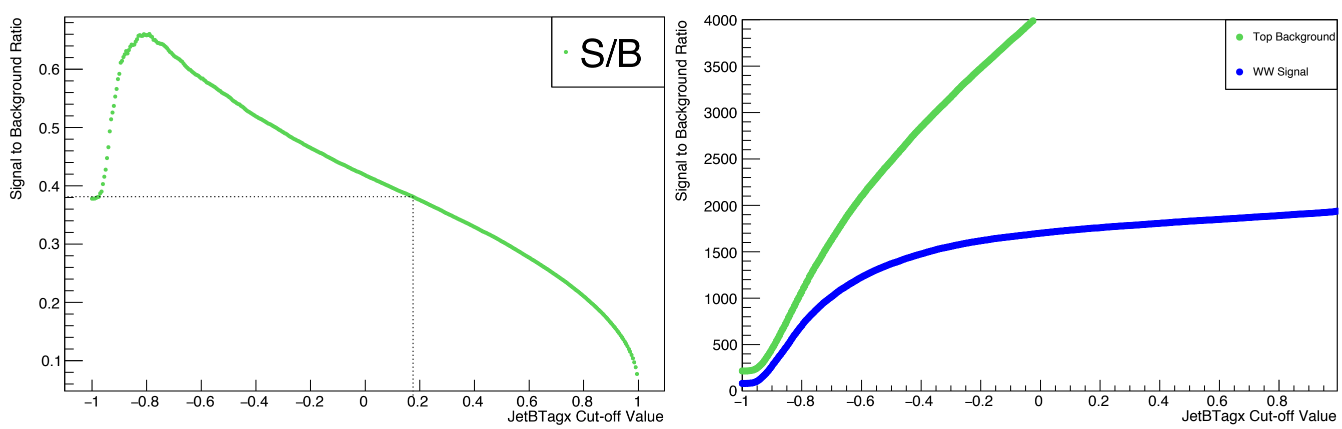

FIG. 6. $WW + 2$ jet: Signal-to-background ratio and absolute event counts as a function of jetBTag value for strict jet veto at 25GeV.

4. DYNAMICAL JET VETOES

As can be seen from figures 5 and 6, the one- and two-jet WW production modes suffer from signal-to-background ratios below 1 and 0.4, respectively, for the case where only a b-jet veto is applied. The purpose of this section is to investigate the possibility of selectively vetoing regions of phase space where the top background is high as a means of increasing the S/B ratio. The approach will be to set the jet veto scale ($P_T^{\text{jet-veto}}$) on an event-by-event basis, a so-called dynamical jet veto, as a function of one or several of the available kinematic variables. We show that setting $P_T^{\text{jet-veto}}$ as a function of the transverse momentum of the leading or subleading lepton yields significant improvement in the S/B for $WW+1-$jet events.

The use of dynamical jet vetoes has been explored in the context of di-boson production before [10]. The initial inspiration for the analysis presented in this report stems from theoretical work by Pascoli *et al* [1], in which setting the upper acceptance threshold for jets is against the P_T of the leading lepton, $P_T^{\text{jet-veto}} > P_T^{LeadLep}$ is shown to improve the suppression of top backgrounds for the production of a single W boson decaying into a hypothetical heavy neutrino. In this section we endeavour to investigate the merits of this approach for the analysis of inclusive WW production.

4.1 Non-linear dynamical jet veotes

There appears to be no *a priori* reason why the dynamic veto should be strictly linear. In fact, as figure 7 shows, if we generate a S/B density plot in the phase spaces of the leading and subleading leptons, and subsequently remove bins below a specified lowest S/B threshold, we see that the dominant top background distribution follows a shape that is assuredly non-linear. Since these removed regions contain an excessive amount of top background in relation to the WW signal, they are ultimately the regions that would need to be vetoed if we are to achieve higher S/B ratios. Consequently, we match a function to the regions and employ it as a dynamical jet veto in the analysis such that for each event we check whether any of the associated jets falls within the region enclosed by the function and discard the event if that is the case.

For both regions we make use of simplified Landau functions of the form

$$f(x) = \frac{ax^2}{\left(x^2+b^2\right)^2}\ , \quad x > c \tag{1}$$

where $x = P_T^{\text{Lep}}$ and $f(x) = P_T^{\text{Jet}}$. The constants $a, b, c \in \mathbb{R}$ are chosen to yield the highest fractional overlap with the removed region with respect to the overlap of the regions outside the removed area. For the purposes of the analysis carried out in this section, this approach is deemed sufficient since it adequately displays the effects of non-linear vetoes. In a more comprehensive analysis, however, the choice of S/B thresholds, histogram binning and fitting of functions would need to be investigated in detail to avoid the inadvertent introduction of biases.

The result of applying these non-linear vetoes in the $WW+1$ jet mode is shown in figure 8, where the S/B is computed with respect to the jetBTag cut-off value following the vetoing of events that fall within a given lepton region. The case where no veto other than the jetBTag cut-off has been applied is also included for reference. From this we observe that the leading and subleading non-linear vetoes consistently yield improvement in S/B ratio on the order 18% and 25%, respectively. In addition, the vetoing of events that fall within either the leading or subleading regions is included, from which we note that this combination out performs the individual leptonic vetoes with an average improvement of 38%. The downside of this top background suppression is displayed in figure 9, where it can be seen that the remaining WW signal after the vetoes constitutes $50-70\%$ of the initial $WW+1$ jet sample. However, we also note that the variation in WW signal among the vetoes is proportional to their S/B increase; hence it should be possible to specify a lowest number of acceptable WW events and optimize the veto functions with respect to the S/B ratio accordingly.

It is of further interest to compute the correlation between these regions. Doing so shows that on average 80% of the vetoed WW events fall events fall within both regions, while 72% of the top background events do the same. Had there been a strong correlation between the top events in these two regions, together with a low correlation of WW events, we could have used this as a veto for top events by specifying that only events within both regions are to be rejected. As it stands however, we are not able to use this to our advantage, although a search for this would have been conducted if it was feasible within the time frame of the project.

Given the unconventional nature of the dynamical vetoes applied so far it is crucial to investigate their effect on the kinematic variables at hand. Specifically, due to the vetoed phase space regions, the P_T distributions of both the jets and the leptons warrant particular attention. Figure 10 shows the P_T distribution for the jets in the $WW+1$ jet events before and after the combined veto has been applied. We note the substantial indentation in the distribution

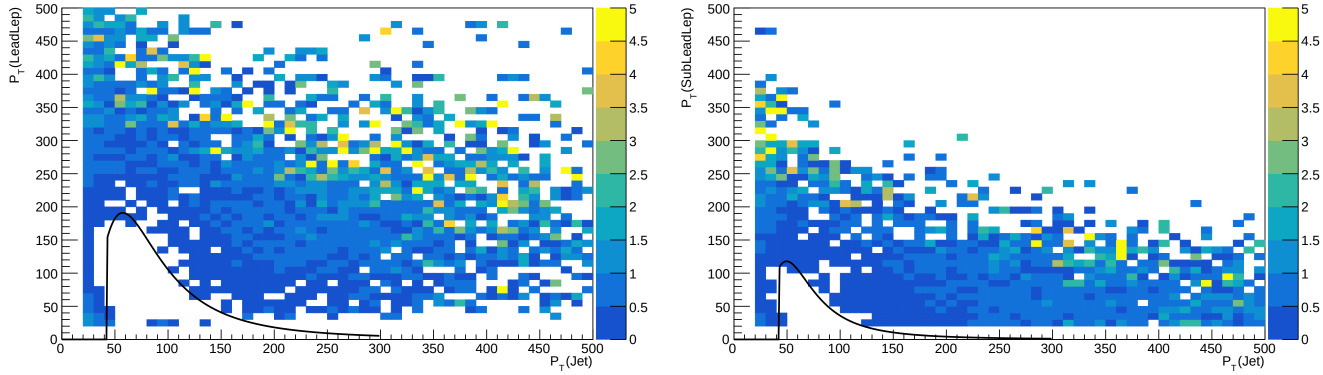

FIG. 7. S/B density distributions in the (a) P_T^{Jet} vs. P_T^{LeadLep} phase space and (b) P_T^{Jet} vs. $P_T^{\mathrm{SubleadLep}}$ phase space. The S/B acceptance threshold is set at 30% and any bin below this threshold is removed. The matched functions, defined by equation 1, used to set the dynamical jet vetoes are also shown.

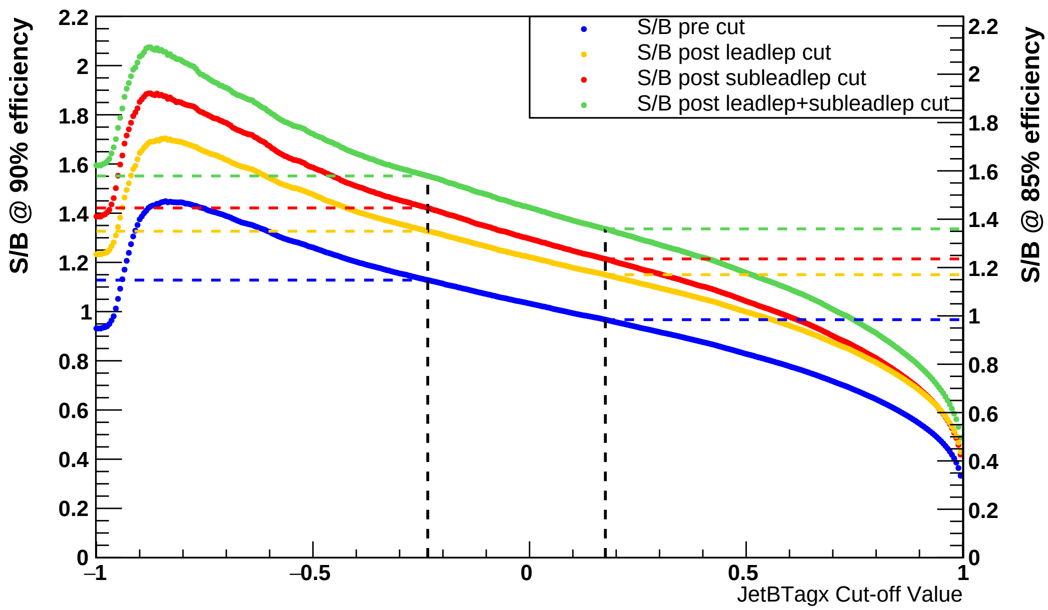

FIG. 8. Signal-to-background ratios for the $WW + 1$ jet production modes.

caused by the veto regions location in the lower part of the P_T(jet) spectrum in figure 7. Furthermore, since we are considering events with only one associated jet, the P_T(jet) spectrum remains unaltered past the extent of the regions covered by the dynamical vetoes.

The effect of the modification of the P_T(jet) spectrum introduced by the vetoes will have to be carefully considered depending on what analysis is to be carried out. For the purposes of this work it is relevant to also look at the P_T spectrum of the leptons since the presence of energetic W bosons recoiling of jets is ultimately not seen through the presence of high P_T jets, but rather in the detection of energetic leptons or large missing P_T from neutrinos. Hence $P_T^{\ell\ell}$, the combined transverse momentum of the leptons, and P_T^{miss} are shown in figure 11, again before and after dynamic vetoes have been applied. From this we note that both P_T distributions suffer only minor modifications in the low P_T region, in spite of the large modification of the jet P_T spectrum seen in figure 10.

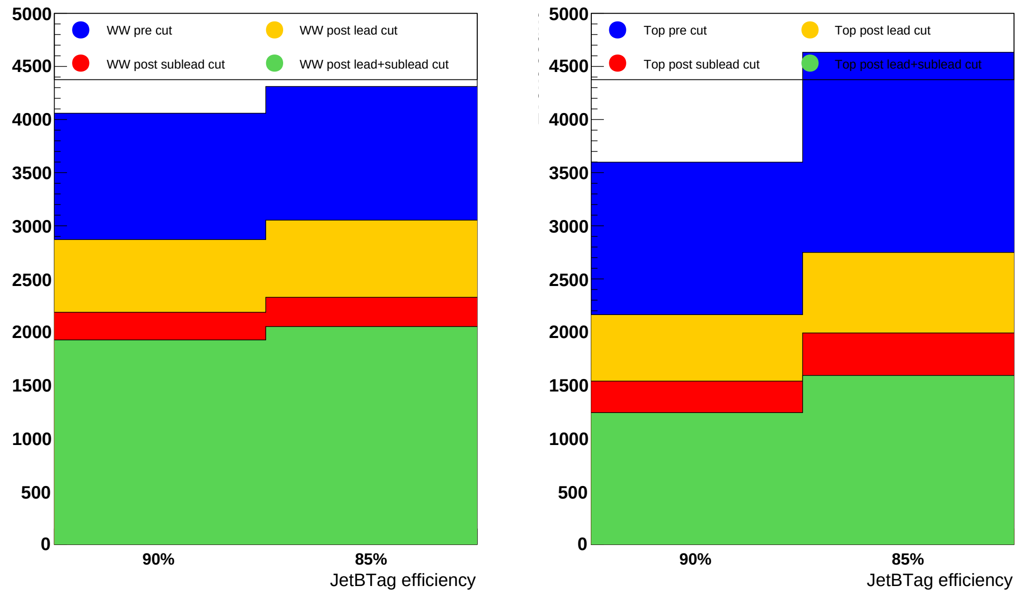

FIG. 9. Absolute counts for the S/B ratios computed in figure 8, for the $WW + 1$ jet production mode, following non-linear dynamical vetoes on the leading and subleading leptons.

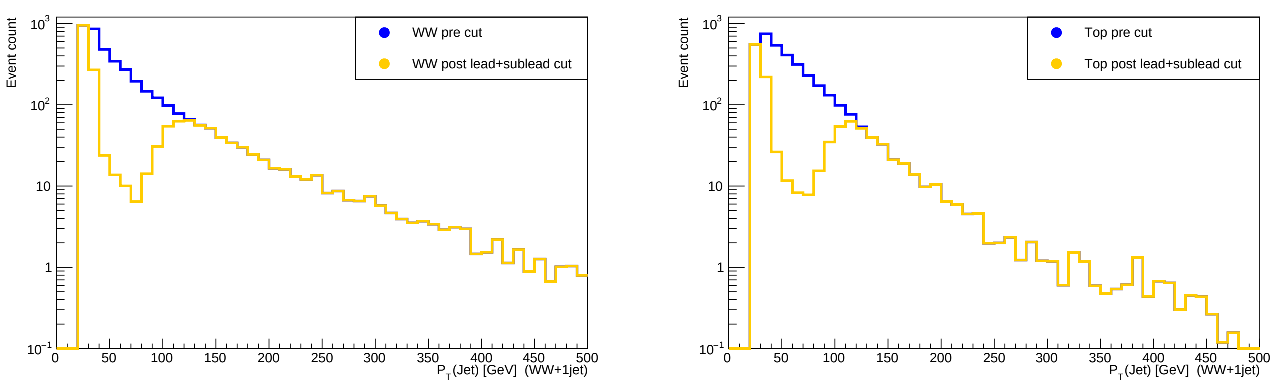

FIG. 10. Jet P_T distributions for the $WW + 1$ jet events, for the WW signal and top background events, before and after the combined veto has been applied.

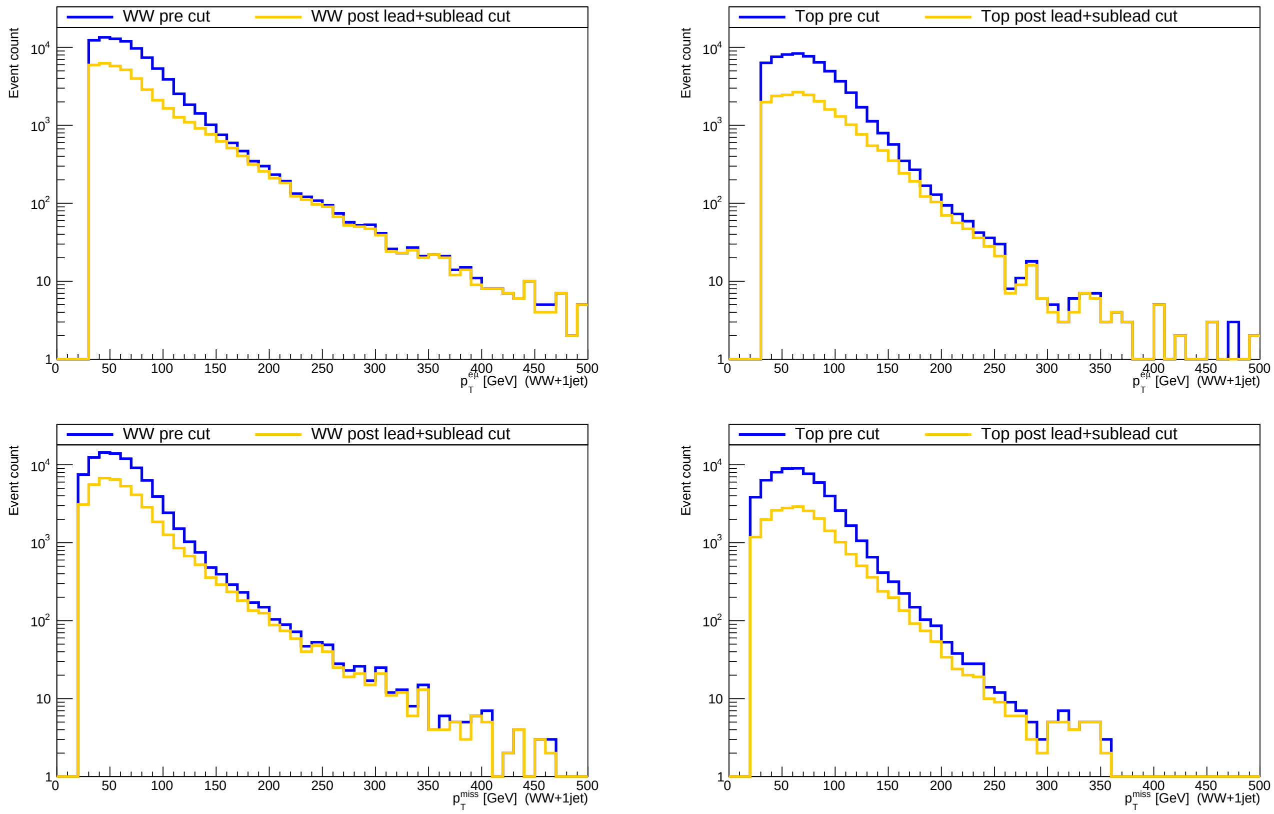

FIG. 11. Lepton P_T distributions: $P_T^{\ell\ell}$ distribution (top) and $P_T^{\rm miss}$ distribution (bottom) for the $WW+1$ jet events, for the WW signal and top background events, before and after the combined veto has been applied.

5. CONCLUSION

In this report the first preliminary investigations into the use of non-linear dynamical jet vetoes for top background suppression in inclusive WW production have been presented. We show that improvements in signal-to-background ratios of up to 40% can be achieved by selectively vetoing regions in the P_T phase space of jets and leptons with a 50% loss of WW signal events. It is further showed that this veto, although it significantly alters the jet P_T spectrum, only slightly modifies the P_T spectrum of the leptons and missing neutrinos, which appears promising for the analysis of the decay of high P_T W bosons.

ACKNOWLEDGEMENTS

The author expresses gratitude to Mateusz Dyndal and Valerie Lang for providing excellent supervision throughout this project. A special thanks goes to the ATLAS summer students and the ATLAS group in general for the numerous social events and enjoyable times spent together.

I. APPENDIX A1. FILES

The files [2] used in this analysis, containing the simulated MC top background at NLO using `Powheg+Pythia8` are:
`mc_410503__nominal.root -->`
`mc_410015__nominal.root -->`
`mc_410016__nominal.root -->`
The files containing the simulated MC WW signal at NLO using `Powheg+Pythia8` are:
`mc_361600__nominal.root --> qq initial state`
`mc_361077__nominal.root --> gg initial state`

References

[1] Silvia Pascoli, Richard Ruiz, and Cedric Weiland, "Safe Jet Vetoes," , 1–8 (2018), arXiv:1805.09335.
[2] F. Febres Cordero, P. Hofmann, and H. Ita, "Ww+ 3-jet production at the large hadron collider in next-to-leading-order qcd," Physical Review D **95**, 034006 (2017).
[3] Zhengkang Zhang, "Time to Go beyond Triple-Gauge-Boson-Coupling Interpretation of W Pair Production," Physical Review Letters **118**, 1–5 (2017).
[4] The ATLAS collaboration, "," Journal of High Energy Physics **2016** (2016), 10.1007/JHEP09(2016)029.
[5] CMS collaboration, "Measurement of the W+W-cross section in pp collisions at s= 8 TeVand limits on anomalous gauge couplings," European Physical Journal C **76** (2016), 10.1140/epjc/s10052-016-4219-1.
[6] The ATLAS collaboration, "," Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics **773**, 354–374 (2017).
[7] The ATLAS collaboration, "," Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics **763**, 114–133 (2016).
[8] Philippe Calfayan, Mateusz Dyndal, Yifei Han, Beate Heinemann, Valerie Lang, Rebecca Linck, Kristin Lohwasser, Karolos Potamianos, Christophe Roland, and Elias Rüttinger, "ATLAS Internal - Measurement of fiducial and differential W W production cross sections at s = 13 TeV with the ATLAS detector," , 1–196 (2018).
[9] ATLAS Flavour Tagging Working Group, "Expected flavour tagging performance in release 20.7," `https://twiki.cern.ch/twiki/bin/view/AtlasProtected/BTaggingBenchmarksRelease20`, accessed: 2018-09-05.
[10] Christopher Frye, Marat Freytsis, Jakub Scholtz, and Matthew J. Strassler, "Precision diboson observables for the LHC, page 22," Journal of High Energy Physics **2016**, 22 (2016), arXiv:1510.08451.

This document was typset in LaTeX.

[2] https://cernbox.cern.ch/index.php/s/75LtL6Y749cD9Gz