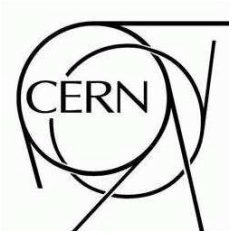
ATLAS INTERNAL NOTE

July 27, 2009



ATLAS Sensitivity to the Standard Model Higgs in the HW and HZ Channels at High Transverse Momenta

J.M. Butterworth[1], A.R. Davison[1], K. Jakobs[2],
V.E. Özcan[1], G. Piacquadio[2], C. Weiser[2]
[1] University College London [2] Albert-Ludwigs-Universität Freiburg

Abstract

Until recently it was thought that for Higgs boson searches at the Large Hadron Collider, WH and ZH production where the Higgs boson decays to $b\bar{b}$ would be poor search channels due to large backgrounds. Recent phenomenological studies have indicated that at high transverse momenta, employing state-of-the-art jet reconstruction and decomposition techniques, these processes can be recovered as promising search channels for the Standard Model Higgs boson around 120 GeV in mass. We investigate this claim using a realistic simulation of the ATLAS detector.

1 Introduction

A key aim of ATLAS is to elucidate the mechanism of mass generation, and hence electroweak symmetry breaking, either by discovering the Higgs boson or by proving that it does not exist. Current electroweak fits, together with the LEP and Tevatron exclusion limits, favour a light Higgs boson, i.e. one around 120 GeV in mass [1]. This mass region is particularly challenging for ATLAS, and any SM Higgs-boson discovery is expected to rely on a combination of several search channels, including $gg \to H \to \gamma\gamma$, $qqH \to qq\tau\tau$ ($qqH \to qqWW$) and the associated production with $t\bar{t}$ pairs [2].

Two significant channels that have generally been considered less promising are those of Higgs-boson production in association with a vector boson, $pp \to WH$ or ZH, followed by the dominant light Higgs boson decay, to two b-jets. Reconstructing W or Z associated $H \to b\bar{b}$ production would typically involve identifying a leptonically decaying vector boson, plus two jets tagged as containing b-hadrons. Identification is difficult due to large backgrounds and low signal acceptance.

A recent study [3] argued that focusing analysis on VH (where V is a W or Z) production in a boosted regime, in which both bosons have large transverse momenta and are back-to-back, has significant advantages over a more inclusive search. This region corresponds to only a small fraction of the total VH cross-section (about 5% for $p_{\mathrm{T}} > 200\,\mathrm{GeV}$), but kinematic acceptance is larger, while the backgrounds are reduced.

The use of jet-finding and b-tagging geared to identifying the characteristic structure of a fast-moving Higgs boson that decays to b and $\bar{b}$ in a common neighbourhood in angle is a key element of the analysis. While in Ref. [3] a reasonable attempt was made to estimate the capability of a detector to achieve this (based in part upon single-jet mass studies with the $k_\perp$ algorithm, see p.1779 of Ref. [2]), a more detailed study using a realistic simulation of ATLAS is clearly required. That is the main goal of this note.

The note is arranged as follows: First we outline the most novel part of the analysis, the Higgs selection (Section 2). Then we describe the simulated event samples used. In Section 4 we go through a cut-based analysis similar to the analysis of Ref. [3], but performed using a realistic detector simulation, and give results based on simple event counting for three sub-channels of the signal. The following sections examine some other aspects of the analysis in more detail (trigger, comparison of different jet algorithms and clustering methods, resolutions and efficiencies, cross-check of ATLFAST-II fast simulation against full simulation). Finally, we discuss how the results from the three studied sub-channels can be combined, before summarising our conclusions.

2 Higgs Candidate Identification

Following the discussion in Ref. [3], we briefly describe the overall strategy for clustering, and include some details of the actual ATLAS implementation in Athena.

When a fast-moving Higgs boson decays, it produces a single fat jet containing two b quarks. The identification strategy proposed in Ref. [3] uses the inclusive, longitudinally invariant Cambridge/Aachen (C/A) algorithm [4,5] to flexibly adapt to the fact that the $b\bar{b}$ angular separation varies significantly with the Higgs p_{T} and decay orientation. In this algorithm one calculates the angular distance $\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ between all pairs of objects (particles) i and j, recombines the closest pair, updates the set of distances and repeats the procedure until all objects are separated by a $\Delta R_{ij} > R$, where R is a parameter of the algorithm. It provides a hierarchical structure for the clustering, like the $k_\perp$ algorithm [6, 7], but in angles rather than in relative transverse momenta. It is implementated in FastJet 2.3 [8] and interfaced to Athena

via the JetRec package.

Two methods have been used to perform the clustering in Athena. In one method, jet clustering is performed in the ESD to AOD conversion step, using the FastJetKtFinder class and taking all available TopoClusters as input. FastJetKtFinder is steerable, and in our case the step is performed using the Cambridge-Aachen procedure described above. The full clustering history, above some minimum ΔR_{ij}, is stored in a JetClusteringHistory collection. The cell-by-cell part of the H1-style jet calibration [2,9] is applied to each of the stored jets before they are saved in the AOD. This means the detailed sub-jet analysis of Ref. [3] can be applied at the AOD analysis level. Alternatively, it is possible to start the clustering directly on the Topoclusters stored in the AOD. This has the advantage that the ESD level is not required. However, one is restricted to using the local hadron calibration as applied to topoclusters, which is less mature than the H1 calibration used in the previous method. The end results of both methods are discussed in Section 6.2.

Whichever approach is taken, all jets j corresponding to the history stage at $\Delta R_{ij} = 1.2$, that is the stage at which all jets are separated from each other by at least $\Delta R = 1.2$, are selected. Those jets with $p_{\mathrm{T}} > 200$ GeV and $|\eta| < 2.5$ are subjected to an iterative decomposition procedure involving two dimensionless parameters, μ and y_{cut}:

1. Break the jet j into two subjets by undoing its last stage of clustering. Label the two subjets j_1, j_2 such that $m_{j_1} > m_{j_2}$.

2. If there was a significant mass drop (MD), $m_{j_1} < \mu m_j$, and the splitting is not too asymmetric, $y = \frac{\min(p^2_{tj_1}, p^2_{tj_2})}{m^2_j} \Delta R^2_{j_1,j_2} > y_{\mathrm{cut}}$, then deem j to be the heavy-particle neighbourhood and exit the loop. Note that $y \simeq \min(p_{tj_1}, p_{tj_2}) / \max(p_{tj_1}, p_{tj_2})$.[1)]

3. Otherwise redefine j to be equal to j_1 and go back to step 1.

The two parameters μ and y_{cut} may be chosen independently of the Higgs mass and p_{T}. We use $\mu > 1/\sqrt{3}$, which ensures that if, in its rest frame, the Higgs decays to a Mercedes $b\bar{b}g$ configuration, then it will still trigger the mass drop condition. The cut on $y \simeq \min(z_{j_1}, z_{j_2}) / \max(z_{j_1}, z_{j_2})$ eliminates the asymmetric configurations that most commonly generate significant jet masses in non-b or single-b jets, due to the soft gluon divergence. We apply a cut at $y_{cut} = 0.1$.

The angular distance between j_1 and j_2, $R_{b\bar{b}}$, defines the distance between the two b-quarks. In order to obtain good b-tagging performance, a reliable separation and reconstruction of the two b-subjets is needed, so that the direction of the two b-subjets can be considered as a reasonable approximation for the direction of the outgoing b-partons after eventual QCD final state radiation. This is crucial in order to correctly associate to the two subjets their charged-particle tracks as reconstructed in the Inner Detector, avoiding to a large extent cross-talk between the two subjets. As a consequence the jet clustering procedure has to be optimized not only to provide a good invariant mass resolution for the Higgs candidate, but also to provide a good angular resolution on the direction of the two subjets and to try to select the $b\bar{b}$ pair out of a $b\bar{b}g$ configuration.

At this stage, the effective size of jet j will be just sufficient to contain the QCD radiation from the Higgs decay, which, because of angular ordering [12–14], will be almost entirely emitted in the two angular cones of size $R_{b\bar{b}}$ around the b quarks. Since this radius sets the angular scale (candidate-by-candidate) of the Higgs decay, it makes sense to recluster, or *filter* the candidate using this information. This involves rerunning the C/A algorithm on the jet constituents, using

[1)] Note also that this y_{cut} is related to, but not the same as, that used to calculate the splitting scale in Refs. [2, 10, 11], which takes the jet p_{T} as the reference scale rather than the jet mass.

a finer angular scale, $R_{\rm filt} < R_{b\bar{b}}$, and taking the three hardest objects (sub-jets) that appear — thus one captures the dominant $\mathcal{O}(\alpha_s)$ radiation from the Higgs decay, while eliminating much of the contamination from the underlying event. We follow Ref. [3] in using $R_{\rm filt} = \min(0.3, R_{b\bar{b}}/2)$. The jet j is accepted as a Higgs candidate if the two hardest subjets have b tags, while the third subjet, if present, is required not to be identified as a $b-$jet. The filtering procedure provides also an effective way to remove some of the contributions arising directly from the showering of the b-quark before hadronisation (i.e. no long-lifetime component) and thus improves the angular resolution of the two hardest subjets with respect to the two b-hadrons arising from the b-quark pair, which is a fundamental ingredient for b-tagging. For technical reasons, the $l\nu b\bar{b}$ analysis uses a variation on the procedure followed in Ref. [3], it requires that one of the two hardest jets is in j_1 and the other is in j_2. This is examined in more detail in Section 6.

Two b-tagging algorithms have been considered for this study: they are known in ATLAS under the names of Combined Tagger (COMB) and JetFitter-based Tagger (JetFitterCOMBNN). Both of them combine impact parameter information with the explicit determination of an inclusive secondary vertex, providing the highest b-tagging performance available within ATLAS, and are not intended to be used in the very early phase of data taking, because their optimization and calibration on data will require some time. More information about them can be found in Ref. [2, 15].

In addition, as described in Section 8, it turns out that one of the main backgrounds in the $l\nu b\bar{b}$ and $\bar{\nu}\nu b\bar{b}$ analyses is given by $t\bar{t}$ events, where the $b\bar{b}$ pair of the Higgs candidate is faked by one real b-quark from a top quark and by one c-quark coming out of the W boson produced by the same top quark. One way to reduce this background contribution is to optimize the b-tagging algorithms explicitly in order to improve the rejection of c-quark jets. The topological reconstruction of the primary vertex $\rightarrow$ b-hadron$\rightarrow$ c-hadron decay chain expected in b-jets and of the invariant masses at the b- and c-hadron decay vertices implemented in the JetFitter-based b-tagging algorithm can be useful to increase the separation power against c-jets, where only one vertex and a lower invariant mass are expected, and thus reject the $t\bar{t}$ background more efficiently.

Finally, the cut at $p_{\rm T} > 200\,{\rm GeV}$ is re-applied to the momentum of the filtered Higgs candidate. The filtered four-momentum, computed from the three highest $p_{\rm T}$ filtered subjets, is considered in all subsequent steps.

In the $l\nu b\bar{b}$ analysis, if more than one Higgs candidate is present in an event, the one with the first two filtered subjets having the highest b-jet probability (sum of b-weights) is chosen as Higgs candidate: this choice has the effect of selecting the correct Higgs candidate at an early analysis stage, when for example an additional hadronically decaying W boson can be present in the event, but is expected to have no effect on the final analysis, since the signal events with more than one Higgs candidate are removed by the remaining selection requirements. In the $ll b\bar{b}$ and $\nu\nu b\bar{b}$ analyses, all candidates are accepted. This discrepancy is not expected to be a significant effect in the signal region.

3 Event Samples and Detector Simulation

All the main event samples were produced with the Herwig Monte Carlo generator [16] using Jimmy [17] to simulate the underlying event, assuming a center of mass energy of 14 TeV. An exception is that the AcerMC [18] Monte Carlo generator was used to produce single top events, as well as additional $Wb\bar{b}$ samples used as a cross-check (see below). For the $Wb\bar{b}$ sample AcerMC was interfaced to Herwig for the parton showering and to Jimmy for the underlying event, while for the single-top sample Pythia [19] was used. For all processes, the

CTEQ6L1 (LO) PDFs were used [20]. In the case of the AcerMC processes, the factorization and renormalization scales Q^2 were set for Wt to $\sum_{i=[t,W]} \left({p_T^i}^2 + m_i^2 \right)/2$, while for $W + b\bar{b}$ they were set to $\sum_{i=partons} \left({p_T^i}^2 + m_W^2 \right)/2$.

For the $WH, W \to \mu\nu, e\nu$ analysis the dominant backgrounds are W+jet, $t\bar{t}$ and single top. For the other analyses, Z+jet dominates, but the other two are also considered.

The W+jet background, after b-tagging is applied, is dominated by the irreducible $q\bar{q} \to Wg \to Wb\bar{b}$ component. The kinematic region of our analysis, in which a $b\bar{b}$ pair is present at high p_{T}, is such that the leading-logarithmic parton-shower approximation as implemented in HERWIG (or PYTHIA), which we define as PS, can be expected to work well. In general, these approximations include the kinematically dominant terms where the ratio of scales involved (in this case loosely the p_{T} of the $b\bar{b}$ system and either its mass or the relative p_{T} of the b and $\bar{b}$) is large. However, since in the final selection the $b\bar{b}$ mass is relatively high (i.e. around the Higgs candidate mass), this ratio may for the lowest p_{T} events be around two, which is not very large. While there is good evidence (see for example Ref. [21, 22]) that parton showers do reproduce many event properties very well over a wide kinematic range, this does not directly extend to $b\bar{b}$ production, and we have made an important cross-check by comparing the rate and kinematics of $b\bar{b}$ production in HERWIG to a fixed order α_s calculation of $Wb\bar{b}$ production, as implemented in AcerMC (which we define as ME). AcerMC uses the full LO matrix element for $Wb\bar{b}$ and $Zb\bar{b}$ production, including the effect of the b-quark mass and a matching to the initial-state parton-shower where the $g \to b\bar{b}$ splitting can be significant. However, it does not match the matrix element on to the final-state parton-shower which can include the same splitting, and which is expected to be the dominant contribution in our kinematic region. This means that the AcerMC samples will be missing some fraction of the parton shower, since, due to fact that the $b\bar{b}$ pair is produced by the ME itself and due to angular ordering, the $b\bar{b}$ is in effect forced to be the widest angle splitting, whereas in the full parton shower in HERWIG contributions where a $g \to gg$ splitting is the widest in angle, but where a $b\bar{b}$ pair is neverthless still developed within the shower, will also be present.

In a previous ATLAS study [23], the PS and ME based approaches were compared in the low p_T region of the phase space, which is relevant for the inclusive WH analysis, with the following conclusions:

- The fraction of effective gluon splitting into heavy flavour is approximately a factor 1.4 higher in PS events than in ME events, due to higher order large logarithmic corrections to $g \to b\bar{b}$ splitting.[2)]
- The evolution of the partonic cascade is more realistically modeled in the PS Monte Carlo and leads to a softer b-quark spectrum than in the ME Monte Carlo.
- The PS approach misses contributions with high gluon virtuality.

In order to have a reference, first the AcerMC (ME) and HERWIG (PS) samples are compared, after the generator cuts which limit the p_T(W) to be higher than 150 GeV/c. Some kinematic distributions of the $b\bar{b}$ system as produced by HERWIG and AcerMC are shown in Fig. 1: $p_T(b\bar{b})$. It can be clearly seen that a low masses or low relative p_{T}, the PS prediction is clearly enhanced with respect to the ME, due to higher order $g \to b\bar{b}$ splitting contributions, while, at high masses or high relative p_{T} the PS approach breaks down, and only the ME approach can fill that region of phase space. This substantially confirms the conclusions already mentioned, in particular that at high gluon virtuality the PS result cannot be trusted.

[2)] It should be stressed that the parton shower used for this study was PYTHIA

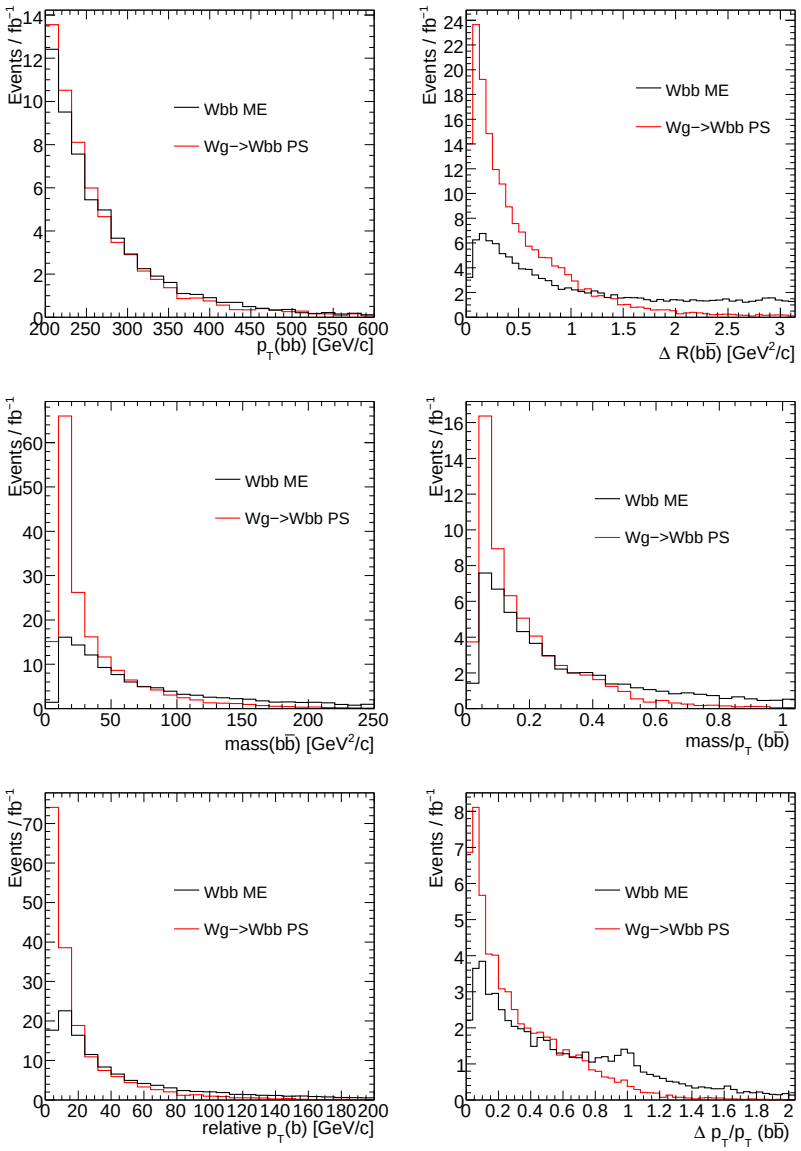


Figure 1: Distributions of the kinematics of the $b\bar{b}$ system in HERWIG and AcerMC, just after generator level cuts (p_T(W)$>$ 150 GeV/c): transverse momentum of the $b\bar{b}$ system (top left), $\Delta R(b\bar{b})$ (top right), invariant mass of the $b\bar{b}$ system (middle left), invariant mass of the $b\bar{b}$ system divided by $p_T(b\bar{b})$ (middle right), relative $p_T(b,\bar{b})$ (bottom left) and relative $p_T(b,\bar{b})$ divided by $p_T(b\bar{b})$ (bottom right). All distributions are normalised to their relative cross sections in fb.

However, in the specific region of phase space region which is relevant for our analysis, it turns out that, as shown in Fig. 2, the distributions from AcerMC (ME) and HERWIG (PS) are in good agreement with each other, especially around the mass region of most interest. This is not too surprising, since, after the requirement of having $p_T(b\bar{b})>$ 200 GeV/c, imposing $\Delta R > 0.3$ we cut out the region where the PS contribution is enhanced with respect to the ME, while imposing $\Delta R < 1.3$ the region of highest gluon virtuality (in particular where the two b-quarks tend to be more back to back) is removed as well. The first plot shows the p_{T} of the $b\bar{b}$ system, which is slightly harder for the AcerMC simulation, consistent with the extra gluon radiation present in HERWIG, where an extra gluon may carry off some of the p_{T}. The top right plot shows the ΔR distribution, which shows some fluctuations beyond statistical uncertainty, and with HERWIG having a tendancy to fall off faster at large ΔR as might be expected. The second and third rows of plots show the distributions of the relative p_{T} and the mass of the b and $\bar{b}$ respectively, as well as the same distributions divided by the p_{T} of the $b\bar{b}$ system. From this comparison we conclude that both the ME and PS approach are valid for the $b\bar{b}$ system in our kinematic region.

Given this, plus the fact that the lack of ME/PS matching in the final state for AcerMC will lead to a depletion of gluon radiation, which will have an impact on the reconstruction efficiency of the hadronic $b\bar{b}$ system, we have used the HERWIG samples for our main analysis. In this way, we have verified that both signal and backgrounds are simulated to the same (leading-logarithmic) degree of accuracy. In order to make also the AcerMC sample usable for the analysis, the W+jet sample generated with HERWIG is added to it, but a crude ME/PS matching for initial state radiation is performed, by removing from the inclusive W+jet sample produced with HERWIG the $qq \to Wg$ component where a b and a $\bar{b}$ are listed as daughter particles of the gluon (i.e. are produced in the parton shower), keeping the remaining $qq \to Wg$ component and all events produced by the $qg \to Wq$ Matrix Element; these will include also all events where no b-jet is present in the final state. Some such samples were used for some of the optimisation and verification studies.

All the above studies and samples are leading order and/or leading logarithmic. At LHC, a large enhancement in the $W + b\bar{b}$ cross section is expected at NLO, essentially because a new process ($qg \to Wb\bar{b}j$), which cannot just be described with additional initial state radiation, comes into play at NLO (see Ref. [24, 25]): however, in this case an additional hard jet is expected in the event and, due to the jet veto implemented in the analysis, this additional NLO contribution is expected to be highly reduced (see Ref. [26]). In addition, the leading-logarithmic contribution to this correction will again be included in HERWIG.

For the single top background, three different channels have to be considered: s and t-channel and Wt production. In the hadron level study performed in Ref. [3], it was noticed that the single top s- and t-channels produce a negligible level of background; while the s channel ($qq \to tb$) has a very low cross section (7.1pb at LO), the t channel $gq \oplus qg \to qt \oplus b$ has a large cross section (251pb at LO), but in both it is only in rare cases that they can fake the signal topology, where a heavy highly boosted object decaying to a b-quark pair is expected together with a high p_{T} W boson decaying into a lepton on the other side: in fact in these channels one of the two b-jets is either soft or it goes into a direction which is opposite to the one of the top, which the other b-jet needs to come from and which needs to be sufficiently boosted in order to produce a high p_{T} W boson. For this reason no single top events for the s and t-channels were produced in this study. However, the Wt channel needs to be considered, because its topology is extremely close to the signal, with one W on one side and a top on the other side, which fakes the Higgs candidate in case one of the three produced jets from the top is very soft and remains undetected, while the remaining c- and $b-$ quarks fake the $b\bar{b}$ pair. This background was not

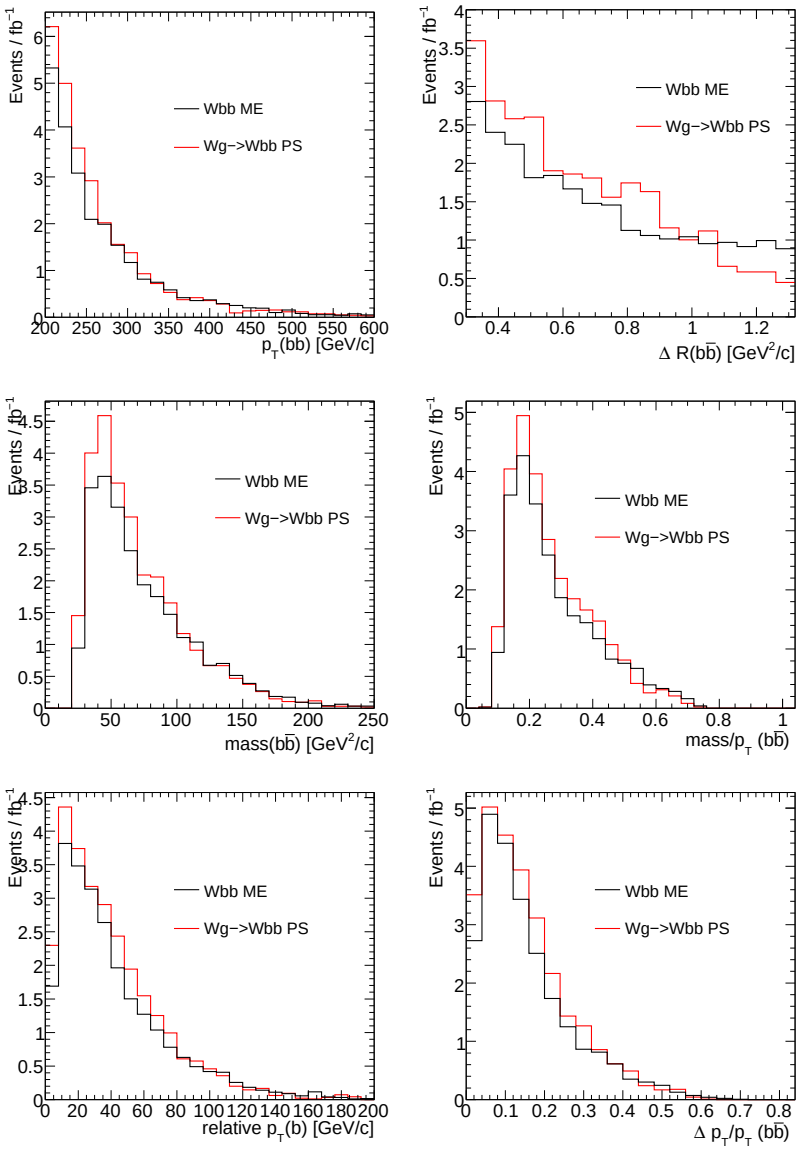


Figure 2: Distributions of the kinematics of the $b\bar{b}$ system in HERWIG and AcerMC. Top left shows the mass distribution for all $b\bar{b}$ pairs with $p_{\mathrm{T}} > 200$ GeV. The subsequent plots also have a requirement that $1.3 > \Delta R(b\bar{b}) > 0.3$; after the Higgs candidate identification, the $b\bar{b}$ pairs are constrained to be well within this range. All distributions are normalised to their relative cross sections in fb.

considered in Ref. [3], and is studied here for the first time.

A simple parton level study was also performed to make sure the $t\bar{t}Z$ background doesn't provide any significant contribution to this analysis. Infact, both the $l\nu b\bar{b}$ and $\nu\nu b\bar{b}$ channels are potentially affected by the $t\bar{t}Z$ background, in particulare in the channel where the Z boson decays into a pair of neutrinos: here the additional transverse missing energy increases the relative background selection efficiency and the fact that the $t\bar{t}$ system is not selected back-to-back makes the emission of two b-quarks or of a c and b-quark from the two top quarks inside a cone of radius $\Delta R = 1.2$ easier with respect to the $t\bar{t}$ background. However, even if the relative selection efficiency of the $t\bar{t}Z$ background is higher with respect to $t\bar{t}$, since the $t\bar{t}Z$ production cross section is only 0.86 pb at LO, after application of the following selection cuts at parton level:

- p_T(lepton)>30 GeV/c, $|\eta|$(lepton)<2.7, $E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV ($l\nu b\bar{b}$ channel only)
- p_T(W or Z boson candidate)>200 GeV/c
- no additional leptons with p_T >20 GeV/c and $|\eta|$ <2.5
- $cb-$ or $bb-$quark pair with $|\eta$(b/c)| <2.5, p_T(bb/cb)>200 GeV/c, $0.3 < \Delta R$(bb/cb)< 1.2
- no additional parton (light- or c-quark) with p_T >30 GeV/c
- $(100 < m_{bb/cc} < 140)$ GeV/c^2

and adding a conservative estimate of the b- and $c-$(mis)tagging efficiencies (respectively 70 % and 10 %), we find that, considering both the potential $bb-$ and $cb-$quark contributions, the contribution of the ttZ background in the $l\nu b\bar{b}$ analysis is < 0.6 events with $30fb^{-1}$ at 95 % confidence level, while its contribution in the $\nu\nu b\bar{b}$ analysis is < 0.4 events with $30fb^{-1}$ at 95 % confidence level. We don't consider the $t\bar{t}Z$ background with the Z boson decaying into a b-quark pair in more detail, since in this case there are two more jets on which the additional jet veto can apply, so that we can expect this contribution to be much smaller than $t\bar{t}Z$ with $Z \to \nu\nu$. We also don't need to consider the $t\bar{t}Z$ background for the $llb\bar{b}$ channel, since requiring m_{ll} to be compatible with the Z mass would again highly suppress this background (the small contribution with $Z \to ll$ is expected to play nearly no role, due to the additional jet and lepton veto which applies to the rest of the event). This confirms that no significant contribution is expected from the $t\bar{t}Z$ background to any of the channels considered in this study.

Finally, no background from QCD multi-jet events was considered. In order to reduce the impact of eventual jets faking leptons, an isolation requirement on muons and electrons is applied. There is nothing particular in this analysis which would enhance the QCD fake background compared to more normal W and Z studies, where such backgrounds are found to be very small [2], however the kinematical configuration and flavour composition selected in this analysis is sufficiently different that a more detailed study of such background would be of value; more detailed studies suffer from the extremely high statistics needed for generating such samples. In general, given the high p_{T} cuts adopted in the analysis both for the W and Higgs boson candidates, it should be more difficult for the QCD dijet background to pass through the analysis cuts than in more inclusive analyses.

More in detail, in the case of the $llb\bar{b}$ channel, the presence of QCD background is highly unlikely, since this would require the simultaneous presence of two correlated and isolated high p_T leptons going nearly into the same direction, with $p_T(ll) > 200$ GeV. In the case of the $l\nu b\bar{b}$ channel, the QCD background contribution could in principle arise either due to a fake or due to a real lepton (plus in both cases real or fake $E_{\mathrm{T}}^{\mathrm{miss}}$). While the fake lepton contribution

should be small, since it requires the correlated presence of a fake lepton and of a high amount of $E_{\mathrm{T}}^{\mathrm{miss}}$, such that $p_T(E_{\mathrm{T}}^{\mathrm{miss}} + l) > 200$ GeV, the contribution from a real lepton from the semi-leptonic decay of a b- or $c-$hadron for example in $gg \to b\bar{b}c$ or $gb \to b\bar{b}b$ could be potentially dangerous; in the case of these processes, the corner of phase space where the fragmentation of a b- or c-quark with $p_T > 200$ GeV into the corresponding leptonically decaying b- or c-hadrons doesn't produce additional particles stemming from the primary event is very small, so most of this background should fail the lepton isolation requirement. In the case of the $\nu\nu b\bar{b}$ channel the QCD background is probably more difficult to control, due to the absence of any lepton requirement: while the amount of QCD multi-jet events with real $E_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV will be negligible, the rejection of this background will depend on how well we will be able to control the extreme tail of the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution.

In order to produce the large number of events needed for this study, the generated events were fed into the AtlFast-II simulation, which corresponds to a full simulation of the ATLAS inner detector and muon system and a fast simulation of the calorimeter in its full granularity relying on the FastCaloSim package [27]. All samples were produced using release 13.0.40.5 of the ATLAS simulation and reconstruction software ATHENA, and rely on the ATLAS-CSC-02-00-00 version of the ATLAS Detector geometry: they do not include the simulation of Pile-Up and cavern background. The WH signal sample at a reference mass of 120 GeV was also passed through the full simulation of the ATLAS Detector, in order to cross-check how well the fast simulation of the calorimeter can reproduce the fine details needed to study the subjet clustering mechanism adopted in this study.

The details of the samples are given in Table 1. For the signal samples, example Higgs boson masses of 115, 120 and 130 GeV were chosen. For technical reasons, the $t\bar{t}$ sample is divided into two, one where the two top quarks have a p_{T} >150 GeV (before eventual QCD radiation) and the complementary sample with the two top quarks having p_{T} <150 GeV.

It should be noted that the filter efficiency for the ZH signal samples is comparable to the WH signal samples, despite Z-boson's lower branching fraction to charged lepton final states. This is because of the fact that the filter is on Monte Carlo final-state objects and it is quite common to have non-Z electrons or muons in the event (for instance from the decays of the b-quarks).

4 Cut-based Analysis

As in Ref. [3], the analysis is performed in three subchannels based upon the decay of the vector boson. (a) Missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}} > 30\,\mathrm{GeV}$plus a lepton ($e$ or μ) with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$, consistent with a W of nominal mass with $p_{\mathrm{T}} > \hat{p}_T^{\mathrm{min}}$. (b) An e^+e^- or $\mu^+\mu^-$ pair with an invariant mass $80 < m < 100\,\mathrm{GeV}$and $p_{\mathrm{T}} > \hat{p}_T^{\mathrm{min}}$. (c) $E_{\mathrm{T}}^{\mathrm{miss}} > \hat{p}_T^{\mathrm{min}}$. The sample from selection (a) is dominated by $HW, W \to \mu\nu, e\nu$, sample (b) contains the $HZ, Z \to e^+e^-/\mu^+\mu^-$ signal, and sample (c) contains a mixture of $HZ, Z \to \nu\bar{\nu}$ and HW where a lepton from the W is outside the acceptance. The vector boson decays to τ leptons are not explicitly searched for but may contribute to some extent to samples (a) and (c).

Sample selection (b) is rather clean but has a low cross-section. Selections (a) and (c) have higher signal cross-sections but are more vulnerable to background from $t\bar{t}$ production. A value of $\hat{p}_T^{\mathrm{min}}$ of 200 GeV has been chosen.

4.1 $l\nu b\bar{b}$ channel

The procedure used to identify and separate the basic physics objects in the event is as follows:

Process	Generator cut	σ(pb)	Filter	Filter efficiency
$WH(115)$	none	1.157pb	$p_{\mathrm{T}}(H) > 150$GeV, $p_{\mathrm{T}}(W) > 100$GeV, $p_{\mathrm{T}}(e,\mu) > 15$GeV	$4.04 \pm 0.03\%$
$WH(120)$	none	0.953pb	$p_{\mathrm{T}}(H) > 150$GeV, $p_{\mathrm{T}}(W) > 100$GeV, $p_{\mathrm{T}}(e,\mu) > 15$GeV	$4.38 \pm 0.04\%$
$WH(130)$	none	0.602pb	$p_{\mathrm{T}}(H) > 150$GeV, $p_{\mathrm{T}}(W) > 100$GeV, $p_{\mathrm{T}}(e,\mu) > 15$GeV	$5.19 \pm 0.03\%$
$WH(120)$	none	0.953pb	$p_{\mathrm{T}}(H) > 150$GeV, $E_{\mathrm{T}}^{\mathrm{miss}} > 100$GeV	$4.39 \pm 0.04\%$
$ZH(115)$	none	0.660pb	$p_{\mathrm{T}}(H) > 150$GeV, $p_{\mathrm{T}}(Z) > 100$GeV, $p_{\mathrm{T}}(e,\mu) > 15$GeV	$3.21 \pm 0.02\%$
$ZH(120)$	none	0.545pb	$p_{\mathrm{T}}(H) > 150$GeV, $p_{\mathrm{T}}(Z) > 100$GeV, $p_{\mathrm{T}}(e,\mu) > 15$GeV	$3.51 \pm 0.02\%$
$ZH(130)$	none	0.347pb	$p_{\mathrm{T}}(H) > 150$GeV, $p_{\mathrm{T}}(Z) > 100$GeV, $p_{\mathrm{T}}(e,\mu) > 15$GeV	$4.18 \pm 0.03\%$
$ZH(120)$	none	0.545pb	$p_{\mathrm{T}}(H) > 150$GeV, $E_{\mathrm{T}}^{\mathrm{miss}} > 100$GeV	$2.34 \pm 0.03\%$
WW	$\hat{p}_T^{\mathrm{min}} = 150$GeV	2.059pb	$p_{\mathrm{T}}(e,\mu) > 15$GeV	$40.7 \pm 0.4\%$
WW	$\hat{p}_T^{\mathrm{min}} = 150$GeV	2.059pb	$E_{\mathrm{T}}^{\mathrm{miss}} > 100$GeV	$30.8 \pm 0.4\%$
ZZ	$\hat{p}_T^{\mathrm{min}} = 150$GeV	0.440pb	$p_{\mathrm{T}}(e,\mu) > 15$GeV	$61.2 \pm 0.2\%$
ZZ	$\hat{p}_T^{\mathrm{min}} = 150$GeV	0.440pb	$E_{\mathrm{T}}^{\mathrm{miss}} > 100$GeV	$47.9 \pm 0.5\%$
WZ	$\hat{p}_T^{\mathrm{min}} = 150$GeV	0.96pb	$p_{\mathrm{T}}(e,\mu) > 15$GeV	$33.6 \pm 0.2\%$
WZ	$\hat{p}_T^{\mathrm{min}} = 150$GeV	0.96pb	$E_{\mathrm{T}}^{\mathrm{miss}} > 100$GeV	$33.3 \pm 0.5\%$
$t\bar{t}$	$\hat{p}_T^{\mathrm{min}} = 150$GeV	112.7pb	$p_{\mathrm{T}}(e,\mu) > 20$GeV	$47.5 \pm 0.2\%$
$t\bar{t}$	$\hat{p}_T^{\mathrm{min}} = 150$GeV	112.7pb	$E_{\mathrm{T}}^{\mathrm{miss}} > 100$GeV	$21.5 \pm 0.2\%$
$t\bar{t}$	$\hat{p}_T^{\mathrm{max}} = 150$GeV	298.7pb	$p_{\mathrm{T}}(e,\mu) > 20$GeV	$39.8 \pm 0.5\%$
$Z + jet$	$\hat{p}_T^{\mathrm{min}} = 150$GeV	160.3pb	$p_{\mathrm{T}}(e,\mu) > 15$GeV	$13.2 \pm 0.2\%$
$Z + jet$	$\hat{p}_T^{\mathrm{min}} = 150$GeV	160.3pb	$E_{\mathrm{T}}^{\mathrm{miss}} > 100$GeV	$21.6 \pm 0.2\%$
$W + jet$	$\hat{p}_T^{\mathrm{min}} = 150$GeV	384.5pb	$p_{\mathrm{T}}(e,\mu) > 15$GeV	$21.1 \pm 0.1\%$
$W + jet$	$\hat{p}_T^{\mathrm{min}} = 150$GeV	384.5pb	$E_{\mathrm{T}}^{\mathrm{miss}} > 100$GeV	$19.5 \pm 0.2\%$
Wbb	none	89.96pb	$p_{\mathrm{T}}(W) > 150$GeV, $p_{\mathrm{T}}(e,\mu) > 15$GeV	$0.51 \pm 0.01\%$
Wt	none	57.896pb	$p_{\mathrm{T}}(W) > 150$GeV, $p_{\mathrm{T}}(top) > 100$GeV, $p_{\mathrm{T}}(e,\mu) > 15$GeV	$9.76 \pm 0.09\%$
Wt	none	57.896pb	$p_{\mathrm{T}}(W) > 150$GeV, $p_{\mathrm{T}}(top) > 100$GeV, $E_{\mathrm{T}}^{\mathrm{miss}} > 100$GeV	$9.72 \pm 0.08\%$

Table 1: Monte Carlo samples produced for the present study. All samples have been produced with the HERWIG Monte Carlo generator, except for the W$b\bar{b}$ and single top samples, which were produced with AcerMC. All cross sections are given at LO.

- The highest p_{T} muon or electron is identified.
- The Higgs boson candidate is then selected using the procedure described in Section 2.
- Additional jets in the event are identified to be used later for the additional jet veto, using a simple inclusive $k_\perp$ algorithm with $\Delta R = 0.4$.
- Additional electrons or muons in the event are found, to which a veto will later be applied.

For a muon to be accepted as the candidate lepton from the W boson decay, it must be reconstructed as a track segment both in the muon and in the inner detector system, and a good match is required between the two. In addition a loose isolation criterion is applied, so that that:

$$\frac{E_{T,cone}}{E_t(\mu)} < 25\%,$$

where $E_{T,cone}$ is the amount of transverse energy in the calorimeter in a cone of $\Delta R = 0.4$ around the track, as extrapolated to the calorimeter entrance.

For an electron to be accepted as the lepton from the W boson, the standard ATLAS medium electron-ID selection is required (isEM=medium) and the calorimeter isolation requirement is:

$$\frac{E_{T,cone}(\Delta R = 0.2)}{p_{\mathrm{T}}(e)} < 10\%.$$

For the leptons to be considered for the veto, they must not be matched to any of the two Higgs subjets (i.e. they must be further away than $\Delta R = 0.4$ from any of them and fail the isolation requirement); muons do not need to have combined inner detector - muon system tracks, while electrons only need to fullfil the loose electron-ID (isEM=loose) requirement. Jets overlapping with the highest p_{T} lepton in the event (ΔR(lepton,jet)< 0.2) are not considered for the jet veto.

After a basic definition of the physics objects, the event selection is applied. In this channel, two selections are defined. A "tight" set of cuts is used to allow a close comparison to the particle-level results of Ref. [3], where a simple event-counting estimate of the significance is given. This result is also used in the combination with the other channels studied in this note. In addition a set of "loose" cuts is also used, in order to provide input to a more sophisticated log-likelihood fit, described in Ref. [28].

The basic selection is shown in Table 2. The loose cut on $\Delta\phi$ is motivated by the fact that the W and the Higgs bosons are expected to be back to back in the transverse plane. In order for a $b-$ or light jet to be considered for the veto, they are required not to overlap with any of the two leading filtered subjets from the Higgs candidate (the ΔR cut between the Higgs subjets and the momenta of k_T jets has been optimized to a value of 0.32). The b-jet veto is applied at a b-tagging efficiency of $\approx 80\%$, so only $\approx 20\%$ of the b-jets are supposed to escape the veto. For convenience the veto on light jets in the table is divided into a veto on either the Higgs or W boson side, according to which of the two is nearer to the jet to veto in the transverse plane. Finally the values of Higgs mass, $p_{\mathrm{T}}(Higgs)$, p_{T}(additional jet) and $\Delta\eta$(W,H) are required to be within the range to be used for the maximum likelihood fit.

The dominant backgrounds are W+jets, $t\bar{t}$ and Wt. The W+jets background is strongly reduced by the requirement of two b-tagged subjets, so the remaining background is dominated by the irreducible $W+b\bar{b}$ contribution. The case of $t\bar{t}$ and Wt is more complex, since the presence of b and c quarks in the final state makes $b-$tagging less powerful in rejecting this background; however, in particular for $t\bar{t}$, vetoing additional $b-$jets and the remaining well identified light-jets in the event down to a certain p_{T} can suppress a good part of this contribution.

Since the Monte Carlo statistics available for the W+jet sample is limited, all the analysis optimization and variables distribution plots were obtained using the AcerMC $W + b\bar{b}$ sample, including the Herwig W+jet sample after removing the $Wg \to Wb\bar{b}$ component. However, since, as discussed in Section 3, the Herwig generator provides a better description of this process in the particular region of phase space selected by this analysis, the inclusive Herwig W+jet sample is used as a reference for the cut flow shown here and for the final mass distribution after all remaining selection cuts are applied, as well as for the final combination aimed to assess the discovery potential. The difference in terms of cut flow between the AcerMC $W + b\bar{b}$ sample and the Herwig sample after selection of the only $Wg \to Wb\bar{b}$ component is shown in Table 4, starting from the lepton cut, which removes the bias due to the fact that the W bosons were forced to decay leptonically in the AcerMC sample. As the table shows, while the number of expected events agrees within statistics after the final mass window cut due to the limited statistics of the Herwig sample, the jet veto and b-tagging cuts show some significant differences.

In order to explain such differences, the parton level study presented in Sec. 3 needs to be recalled here, where it was explained that in a significant fraction of the events in Herwig the widest angle splitting is $g \to gg(\to g b\bar{b})$, and not $g \to b\bar{b}$, so that either the two leading subjets happen to describe an underlying gb or $g\bar{b}$ pair, thus failing the b-tagging requirement, or yielding to the presence in addition to the $b\bar{b}$ pair of an additional hard jet in the event close to the $H \to b\bar{b}$ candidate, on which the veto on jets surrounding the $H \to b\bar{b}$ candidate does more easily apply.

More specifically, after the $\Delta\phi$ cut, the number of events predicted by the PS is higher than those predicted by the ME approach, compatible with our expectations. Focusing on this stage of the cut flow, the invariant mass distribution of the Higgs candidate is shown in Fig. 3 separately for the category where the two leading subjets are matched to a real $b\bar{b}$ pair and for the category where they are matched to a real gb or $g\bar{b}$ pair. In the first case, the distributions are very similar for the PS and ME approaches, although the region of higher masses is slightly lower in the PS approach. In the second case, the PS predicts a much higher rate than the ME, which is most probably due to the presence of secondary $g \to b\bar{b}$ splittings. However, after b-tagging and jet vetoes are applied, the gb or $g\bar{b}$ contribution is essentially completely removed; in addition to that the real $b\bar{b}$ contribution gets reduced more effectively in the PS than in the ME Monte Carlo, mostly due to the additional gluon radiation predicted in the PS approach, but also because in that case the b-jet efficiency turns out to be slightly lower. So while at the beginning the PS approach predicts a higher rate than the ME, after all analysis cuts the situation is the opposite. This difference can be explained by the different generator modelling of the jet shapes and substructure, for which we expect the PS based approach to provide a more reliable description.

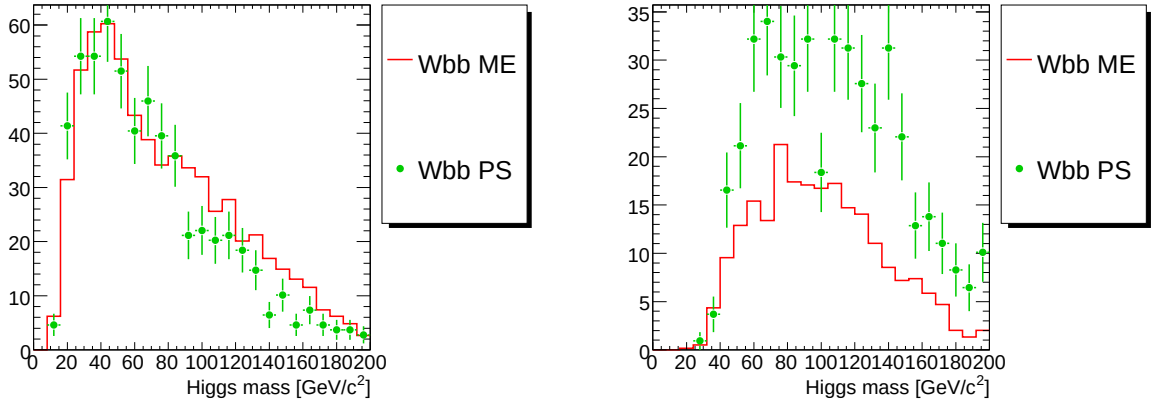


Figure 3: Comparison between ME and PS based approaches for the simulation of the W+$b\bar{b}$ background of the invariant mass of the Higgs candidate after the $\Delta\phi$ cut, for the category where the two leading subjets are matched to a real $b\bar{b}$ pair (left) and for the category where they are matched to a real gb or $g\bar{b}$ pair (right). The distributions are normalized to the number of events expected in 30 fb^{-1} of collected data.

The tight selection is applied on top of the loose one and is shown in Table 3. The difference is in a harder jet veto (any additional jets with $p_{\mathrm{T}} < 20$ GeV and up to $|\eta| < 5$ are vetoed) and in a tighter b-tagging requirement, explicitly optimized against charm quark-jets, imposed on the two leading Higgs subjets (more details on b-tagging in Section 8).

Since most of the Monte Carlo samples used were produced with a generator filter cut on the $2 \to 2$ Matrix Element process at $\hat{p}_T^{\mathrm{min}} = 150$ GeV (except for the $t\bar{t}$ background), it is reasonable to look at the shape of the variables used for the selection only after the bias introduced by this filter is removed. This is most effectively done requiring a single high p_{T} Higgs candidate to be reconstructed with $p_{\mathrm{T}} > 200$ GeV and on the other side a W boson with $p_{\mathrm{T}} > 200$ GeV (first four cuts in Table 2. After these first cuts are applied, the distributions for $p_{\mathrm{T}}(W)$ and for $E_{\mathrm{T}}^{\mathrm{miss}}$ are as shown in Fig. 4. In all distributions shown here, the WZ, ZZ and WW background components are summed up together in the diboson VV background, the two $t\bar{t}$ samples are also

	WH(115)	WH(120)	WH(130)	WZ	$t\bar{t}(p_T^{min})$	Wt	W+jets
After filter cuts	1402.3 ± 7.2	1252.8 ± 7.8	946.1 ± 4.9	9331	1609356	169519	2433885
1 Higgs candidate	646.2 ± 2.8	569.7 ± 3.0	429.7 ± 1.9	3509.7 ± 8.0	806175	69375	562030
filtered p_T > 200 GeV	581.7 ± 2.9	512.7 ± 3.2	387.6 ± 2.0	3108 ± 10	709271	60241	413406
Missing E_T > 30 GeV	413.7 ± 2.9	362.4 ± 3.2	273.6 ± 2.0	2183 ± 13	552284	46779	318400
p_T(W) > 200 GeV	194.3 ± 2.4	171.0 ± 2.6	128.0 ± 1.6	1216 ± 12	137946	18524	206331
p_T(e/μ)>30 GeV	166.2 ± 2.2	145.6 ± 2.4	108.1 ± 1.5	996 ± 11	115053	15724	178004
p_T(additional μ)<10 GeV	165.1 ± 2.2	144.6 ± 2.4	107.3 ± 1.5	942 ± 11	106836	14992	177542
p_T(additional e)<10 GeV	162.9 ± 2.2	142.9 ± 2.4	105.8 ± 1.5	885 ± 11	97305	13881	174941
$\Delta\phi$(W,H)< $\frac{2}{3}\pi$	161.9 ± 2.2	142.2 ± 2.4	104.9 ± 1.5	841 ± 11	84773	12999	167704
no additional b-jets p_T >15 GeV	150.7 ± 2.2	130.6 ± 2.3	95.1 ± 1.4	790 ± 10	30605	7805	160608
add. jets on W side p_T <60 GeV	133.2 ± 2.1	115.7 ± 2.2	83.6 ± 1.4	637.2 ± 9.5	19422	5870	121437
add. jets on H side p_T <60 GeV	119.8 ± 2.0	102.7 ± 2.1	75.2 ± 1.3	525.6 ± 8.8	13841	4370	94055
one subjet b-tagged	108.2 ± 1.9	91.4 ± 2.0	66.8 ± 1.2	126.1 ± 4.5	8638	2421	6964
both subjets b-tagged	54.3 ± 1.4	45.6 ± 1.4	32.85 ± 0.89	43.7 ± 2.7	576	161.4 ± 7.0	266
loose fit cuts	54.2 ± 1.4	45.4 ± 1.4	32.75 ± 0.89	43.0 ± 2.7	565	156.3 ± 6.9	257

	ZH(120)	WW	ZZ	$t\bar{t}(p_T^{max})$	Z+jets
After filter cuts	574.2 ± 3.3	25140	8079	3566400	649215
1 Higgs candidate	295.9 ± 1.3	8428.0 ± 5.0	3372.1 ± 8.1	160154	225597
filtered p_T > 200 GeV	267.8 ± 1.4	7355 ± 12	2993.4 ± 9.4	103170	175061
Missing E_T > 30 GeV	141.9 ± 1.3	6217 ± 15	1414 ± 10	75129	62883
p_T(W) > 200 GeV	33.86 ± 0.76	4166	457.5 ± 6.7	3320	13514
p_T(e/μ)>30 GeV	9.06 ± 0.41	3463	264.2 ± 5.3	2899	10135
p_T(additional μ)<10 GeV	5.69 ± 0.32	3441	177.9 ± 4.4	2508	7174
p_T(additional e)<10 GeV	3.56 ± 0.26	3361	121.7 ± 3.6	2037	4741
$\Delta\phi$(W,H)< $\frac{2}{3}\pi$	3.43 ± 0.25	3257	112.0 ± 3.5	1638	4081
no additional b-jets p_T >15 GeV	3.04 ± 0.24	3128	102.4 ± 3.3	494	3751
add. jets on W side p_T <60 GeV	2.35 ± 0.21	2525	82.2 ± 3.0	323	2746
add. jets on H side p_T <60 GeV	2.15 ± 0.20	2083 ± 14	68.9 ± 2.8	233	2146
one subjet b-tagged	1.94 ± 0.19	269.9 ± 5.8	16.8 ± 1.4	134 ± 14	207 ± 13
both subjets b-tagged	0.96 ± 0.13	5.43 ± 0.84	5.26 ± 0.77	7.6 ± 3.4	12.5 ± 3.1
loose fit cuts b-tagged	0.96 ± 0.13	5.43 ± 0.84	5.26 ± 0.77	7.6 ± 3.4	12.5 ± 3.1

Table 2: Number of events passing the selection at each stage for the signal processes of different Higgs boson masses and the background processes. Numbers are projected to 30 fb^{-1} of collected data, based on LO MC cross sections.

	WH(115)	WH(120)	WH(130)	WZ	$t\bar{t}(p_T^{min})$	Wt	W+jets
add. jets on W side p_T <20 GeV	98.9 ± 1.8	83.2 ± 1.9	62.5 ± 1.2	461.3 ± 8.3	7227	3343	86087
add. jets on H side p_T <20 GeV	67.0 ± 1.5	55.8 ± 1.6	41.17 ± 0.99	275.6 ± 6.6	1895	1142	48229
one subjet b-tagged	55.5 ± 1.4	46.4 ± 1.5	33.50 ± 0.90	49.8 ± 2.9	986	498 ± 12	1825
both subjets b-tagged	23.05 ± 0.91	19.51 ± 0.96	13.98 ± 0.59	16.5 ± 1.7	38.9 ± 4.9	18.2 ± 2.4	87.3 ± 9.0
112 GeV < mass(H) < 136 GeV	-	13.25 ± 0.79	-	1.18 ± 0.45	5.6 ± 1.9	4.2 ± 1.1	8.3 ± 2.8

	ZH(120)	WW	ZZ	$t\bar{t}(p_T^{max})$	Z+jets
add. jets on W side p_T <20 GeV	1.26 ± 0.15	1835 ± 14	46.3 ± 2.3	87 ± 12	1571
add. jets on H side p_T <20 GeV	0.93 ± 0.13	1134 ± 11	28.4 ± 1.8	18.3 ± 5.3	850
one subjet b-tagged	0.78 ± 0.12	68.9 ± 3.0	6.26 ± 0.84	7.6 ± 3.4	49.9 ± 6.2
both subjets b-tagged	0.333 ± 0.079	0.65 ± 0.29	1.68 ± 0.43	1.5 ± 1.5	3.9 ± 1.7
112 GeV < mass(H) < 136 GeV	0.278 ± 0.072	0.13 ± 0.13	0.11 ± 0.11	0.0 ± 1.5	0.78 ± 0.78

Table 3: Expected number of events with tighter selection cuts for an integrated luminosity of 30 fb^{-1}.

	$qq \to Wbb$	$qq \to Wg$
p_T(e/μ)>30 GeV	1176 ± 12	1359
p_T(additional μ)<10 GeV	1153 ± 12	1346
p_T(additional e)<10 GeV	1131 ± 12	1326
$\Delta\phi$(W,H)< $\frac{2}{3}\pi$	1087 ± 12	1316
no additional b-jets p_T >15 GeV	926 ± 11	1179
add. jets on W side p_T <20 GeV	663.0 ± 9.9	732
add. jets on H side p_T <20 GeV	446.8 ± 8.3	461
one subjet b-tagged	340.5 ± 7.3	323
both subjets b-tagged	99.6 ± 4.0	60.7 ± 7.5
loose fit cuts	97.4 ± 4.0	59.8 ± 7.4
112 GeV < mass(H) < 136 GeV	10.7 ± 1.3	6.4 ± 2.4

Table 4: Expected number of events for the tight selection cuts for the AcerMC $W + b\bar{b}$ sample compared to the HERWIG W+jet sample, where the only $qq \to Wg \to Wb\bar{b}$ component was selected. All numbers correspond to an integrated luminosity of 30 fb^{-1}.

added together, the W+jets and the Z+jets backgrounds are included in the V+jets background, and the ZH sample passing through the selection is added up to the WH signal. It should, however, be kept in mind that the background is dominated by W+jets, $t\bar{t}$, Wt and WZ at any stage of the analysis.

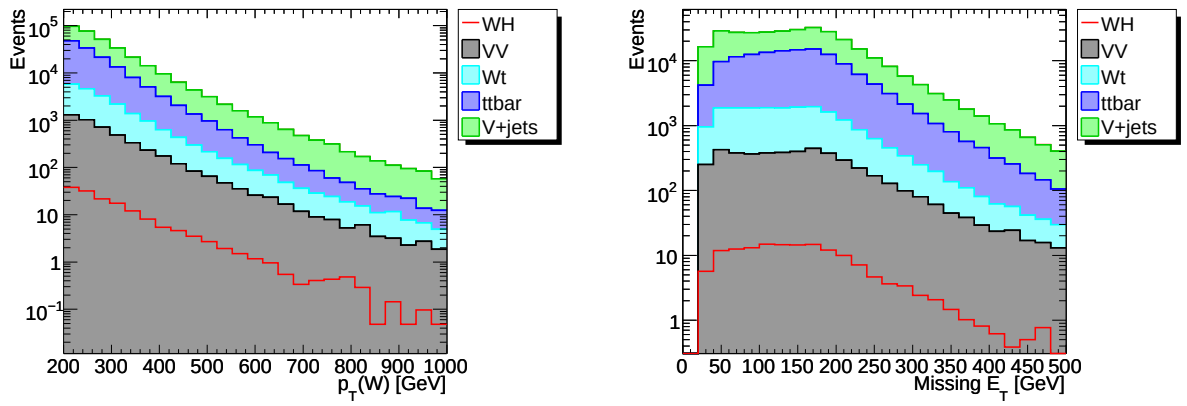


Figure 4: Distribution for $p_{\mathrm{T}}(W)$ (left) and $E_{\mathrm{T}}^{\mathrm{miss}}$ (right) after the first four cuts of Table 2. The backgrounds are added on top of each other, to provide the overall background shape. The signal is just added in foreground.

Fig. 5 shows the distribution of the $\Delta\phi$ between the Higgs and the W boson candidates, after the lepton veto cuts have been applied. As can be seen from the signal distribution, a cut at $\Delta\phi(W,H) > \frac{2}{3}\pi$ is extremely conservative, however this value has been chosen because the use of a NLO Monte Carlo generator to generate WH signal is expected to increase the rate at which the the W and Higgs bosons are not emitted exactly back to back due to some additional hard QCD radiation which is outside the kinematic region correctly described by the parton shower.

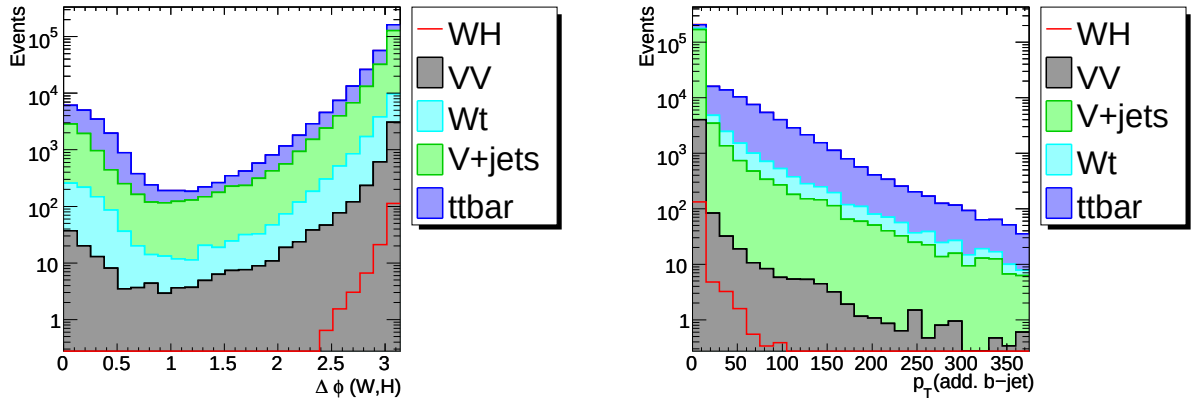


Figure 5: Distributions for $\Delta\phi$ between the Higgs boson and the W boson (left) and the distribution of transverse momenta of additional $b-$jets in the event, on which to apply the b-jet veto (right).

Fig. 5 (right) shows the distribution of additional $b-$jets in the event, after having applied the $\Delta\phi$ cut. Fig. 6 shows the $b-$weight distributions provided by the JetFitter algorithm, after having applied the jet veto cut of the *tight* selection (at 20GeV).

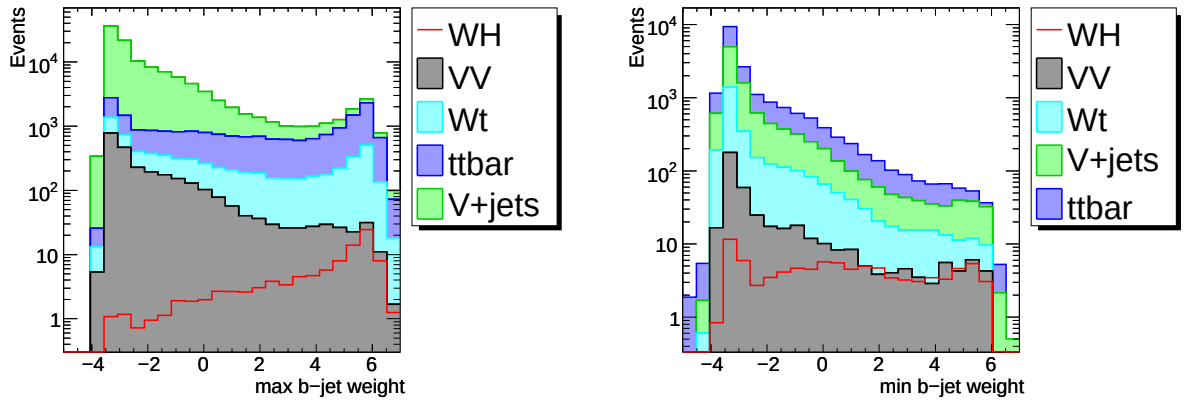


Figure 6: Distribution for the highest (left) and the lowest (right) of the $b-$weights of the two Higgs candidate subjets. The right distribution is shown after having applied the *tight* selection cut at 1.5 on the distribution on the left.

The mass distribution of the Higgs candidates is shown in Fig. 7, after application both of the loose and tight selections. In the latter case, both Higgs and Z boson peaks are well visible on top of the $t\bar{t}$, Wt and W+jets backgrounds, which are dominated by the irreducible $W + b\bar{b}$ contribution. The Wt and $t\bar{t}$ backgrounds are labeled as *top* background in the plots. Since the statistics for the reference HERWIG sample used for the W+jet background component is rather low, the same distributions are also shown adopting the AcerMC sample for the dominant $W + b\bar{b}$ component, in order to cross check the mass distribution.

The resulting significance, considering the mass range 112-136 GeV, in terms of $\frac{S}{\sqrt{B}}$ is 3.0 ± 0.3, where the quoted uncertainty comes from the limited available Monte Carlo statistics, and the signal to background ratio is $\frac{S}{B} \simeq \frac{2}{3}$. This number can be approximately compared to the particle-level result for this channel in Ref. [3] of 3.1. Note that in the particle-level study, high $E_{\mathrm{T}}^{\mathrm{miss}}$ events were in fact counted in the $E_{\mathrm{T}}^{\mathrm{miss}} b\bar{b}$ channel regardless of whether a lepton was identified, thus reducing the relative contribution to the significance from the $l\nu b\bar{b}$ channel compared to our result. The other two channels are considered next. Trigger efficiency and systematic uncertainties are not included. These are discussed in subsequent sections. Nevertheless our result is already enough to indicate that this channel is indeed an excellent prospect for low-mass Higgs searches with ATLAS.

4.2 $ll b\bar{b}$ channel

In this section extraction of the processes $ZH \to e^+e^- b\bar{b}$ and $ZH \to \mu^+\mu^- b\bar{b}$ is discussed. The requirement of leptonic Z decay leads to small branching ratios. However this is counteracted by the fact that it is hard for backgrounds such as $t\bar{t}$ to emulate this signature.

The analysis for this channel was performed at the AOD level. Hadronic reconstruction was performed using the TopoClusters stored in the AOD as produced by the reconstruction software in release 13.0.40. This procedure requires a correction to the jet energy scale as described in Section 6.

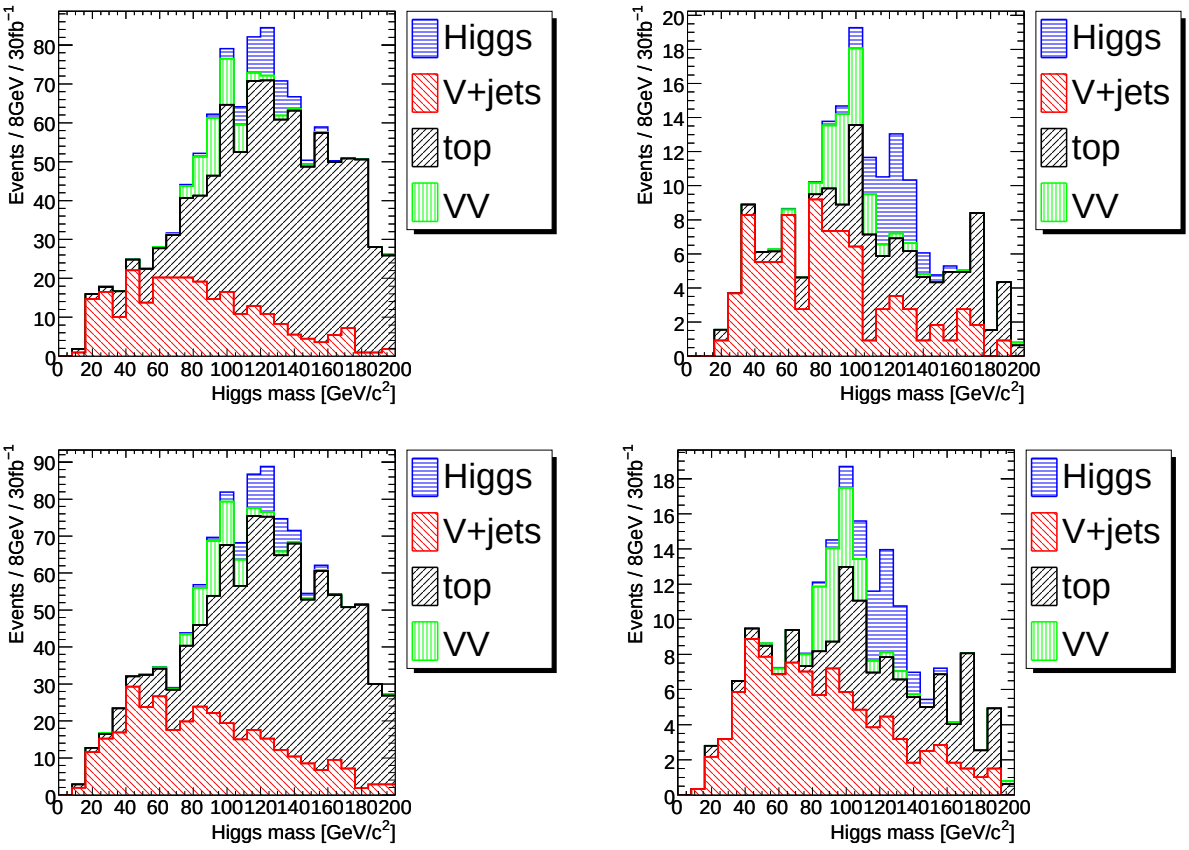


Figure 7: Distribution of the invariant mass of the Higgs candidate after all selection cuts for the *loose* (left) and *tight* (right) selection, using the reference HERWIG sample for the $W + jet$ background component (top) and using the higher statistics AcerMC sample for the $W + b\bar{b}$ background component, but without the minor contribution from W+light jets (bottom). The WH signal (for $m_H = 120$ GeV) is shown on top of the backgrounds. All distributions are normalized to an integrated luminosity of 30 fb^{-1}.

The selection consists of two parts, firstly a candidate for the hadronic $H \to b\bar{b}$ system is identified according to the procedure described in Section 2. A candidate for a leptonic Z is then defined as a pair of opposite sign, same flavour electrons or muons with an invariant mass such that $80 < M_{\ell^+\ell^-} < 100\,\text{GeV}$ where the highest p_T of the two leptons satisfies $p_\text{T} > 25\,\text{GeV}$ and the other satisfies $p_\text{T} > 20\,\text{GeV}$. The mass and p_T spectra of the Z candidates can be seen in Figure 8. The decay lepton p_T spectra can be found in Figure 9.

Considering this, the selection specific to this channel is:

- H candidate with $p_\text{T} > 200\,\text{GeV}$ and $|\eta| < 2.5$
- Z candidate with $p_\text{T} > 180\,\text{GeV}$
- Z and H candidates not within 1.2 radians of each other in ϕ
- The two leading subjets of the H candidate have b-weights > 1.0

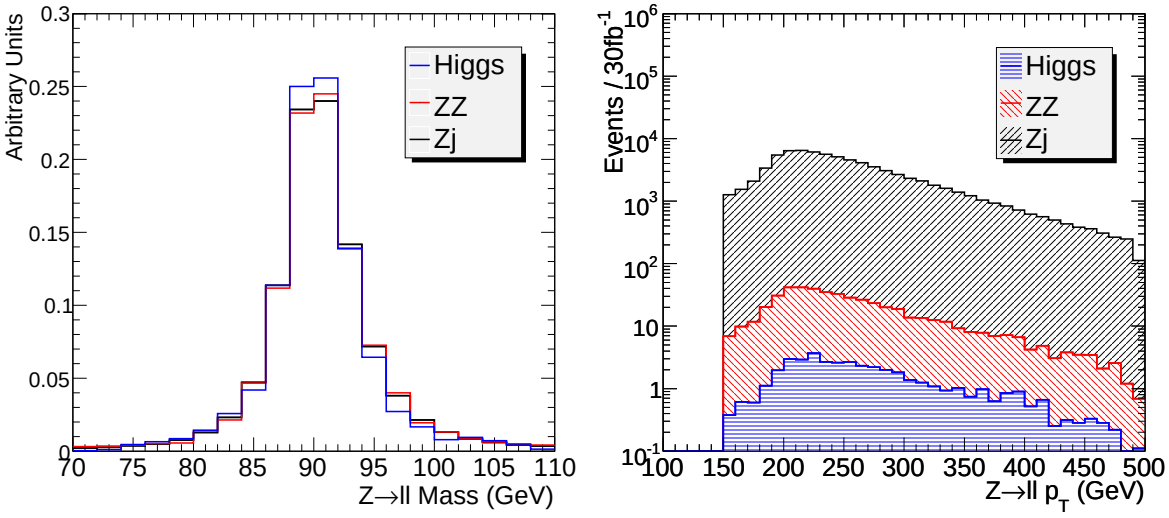


Figure 8: Mass and p_{T} spectra of Z candidates.

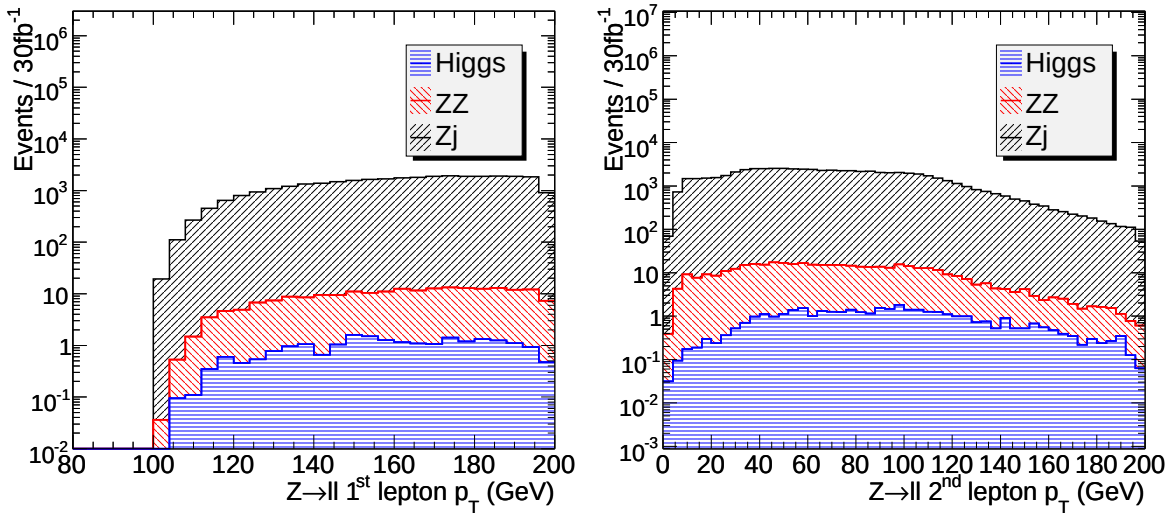


Figure 9: p_{T} spectra of Z candidate decay leptons.

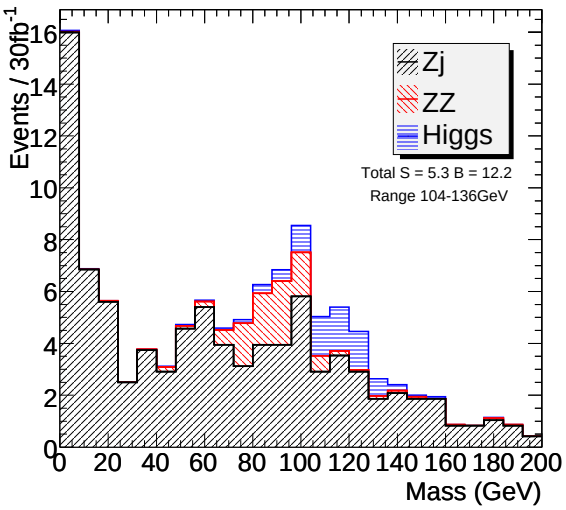


Figure 10: Masses of H candidates after ZH selection

Applying this selection we find that $t\bar{t}$ production appears to be a negligible background in this channel.

The cut flow for HZ, ZZ and $Z + jets$ samples can be found in Table 5.

The mass distribution of H candidates in all samples is shown in Figure 10. As in the previous section, both Higgs and Z boson peaks are visible on top of the backgrounds, which in this case is dominated by $Z+$ jets. The resulting significance in terms of $\frac{S}{\sqrt{B}}$ is 1.5 and the signal to background ratio is $\frac{S}{B} \simeq \frac{1}{3}$. Trigger efficiency and systematic uncertainties are not included. These are discussed in subsequent sections. This number can be compared to the particle-level result for this channel in [3] of 2.1.

4.3 $E_{\mathrm{T}}^{\mathrm{miss}} b\bar{b}$ channel

The aim of this channel is to extract events where a $H \to b\bar{b}$ system is produced in association with a large amount of missing E_{T}. These events come primarily from the process $ZH \to \nu\nu b\bar{b}$. There is also a contribution from $WH \to l\nu b\bar{b}$ where the charged lepton has not been correctly identified and therefore a W can not be reconstructed in the $l\nu b\bar{b}$ analysis. This channel offers greater signal cross-sections than the $llb\bar{b}$ channel but does not have the strong background rejection the $Z \to ll$ system provided.

As in the $llb\bar{b}$ analysis, the analysis for this channel was performed at the AOD level. Hadronic reconstruction was performed on the AOD TopoClusters with a jet energy scale correction applied as described in Section 6.

The $H \to b\bar{b}$ identification is as defined in Section 2. The requirement of no leptons ensures that the sample of events selected in this channel is independent of that extracted in the $l\nu b\bar{b}$ channel in Section 4.1.

The jets used for the additional jet veto are Cambridge-Aachen jets with $R = 0.4$. A range of possible values for the p_{T} cut used in the veto are scanned in Fig. 13. Although higher significances are achievable with lower values, a cut of 30 GeV has been chosen here as a value which should be robust against pile-up.

Considering these points, the selection for this channel is as follows:

	$ZH(120)$	ZZ	$Z + jets$
Generated	575 ± 3	3129 ± 11	647460 ± 367
$p_{Tjet} > 200GeV$	301 ± 2	1503 ± 7	325080 ± 260
$p_{TH} > 200GeV$	263 ± 2	1186 ± 7	232152 ± 220
$\|\eta_H\| < 2.5$	259 ± 2	1166 ± 6	228413 ± 218
$80GeV < m_Z < 100GeV$	33.6 ± 0.7	372 ± 4	63149 ± 115
$p_{TZ} > 200GeV$	29.5 ± 0.7	310 ± 3	51160 ± 103
$p_{TZ_l1} > 25GeV$	29.5 ± 0.7	310 ± 3	51160 ± 103
$p_{TZ_l2} > 20GeV$	28.8 ± 0.7	282 ± 3	46204 ± 98
$d\phi_{ZH} > 1.2$	23.0 ± 0.6	197 ± 3	33341 ± 83
b-tagged ($w > 1.0$)	8.0 ± 0.4	10.4 ± 0.6	91 ± 4
$104GeV < m_H < 136GeV$	5.3 ± 0.3	1.0 ± 0.2	11 ± 2

Table 5: Expected number of events in the $llb\bar{b}$ channel with 30fb^{-1}of data, after each selection criterion is applied.

- H candidate with $p_{\mathrm{T}} > 200\,\mathrm{GeV}$ and $|\eta| < 2.5$
- $E_{\mathrm{T}}^{\mathrm{miss}} > 200\,\mathrm{GeV}$
- No electron or muon with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$
- No veto jet with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$
- H candidate and $E_{\mathrm{T}}^{\mathrm{miss}}$ are not within 1.2 radians of each other in ϕ
- The two leading subjets of the H candidate have b-weights > 1.0

After applying this selection the final mass distribution of H candidates can be found in Fig. 11. The cut flow used to produce this plot is given in Table 6.

The $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions can be seen in Fig. 12 which also shows the p_{T} spectra of additional leptons, while the relevant distributions for veto jets are available as Fig. 13.

The final significance after applying this selection is expected to be 1.6σ for 30fb^{-1} of integrated luminosity. This number can be approximately compared to the particle-level result for this channel in [3] of 3.1, however, as already explained at the end of Sec. 4.1, in the hadron level study some of the $l\nu b\bar{b}$ events were considered as belonging to the $E_{\mathrm{T}}^{\mathrm{miss}}$ channel, increasing the relative contribution to the significance from the $E_{\mathrm{T}}^{\mathrm{miss}}$ channel compared to our result.

5 Trigger efficiency

Analyses for the $l\nu b\bar{b}$ and $llb\bar{b}$ channels make use of muons and electrons with $p_T > 25 \sim 30\,\mathrm{GeV}$. These leptons are of sufficiently high transverse momenta for triggering most events for instantaneous luminosities up to $2{\times}10^{33}$ cm^{-2} s^{-1}, given that the following single-lepton trigger items expected to stay unprescaled according to Ref. [2]:

- $e25i$
- $mu20i$
- $e105$

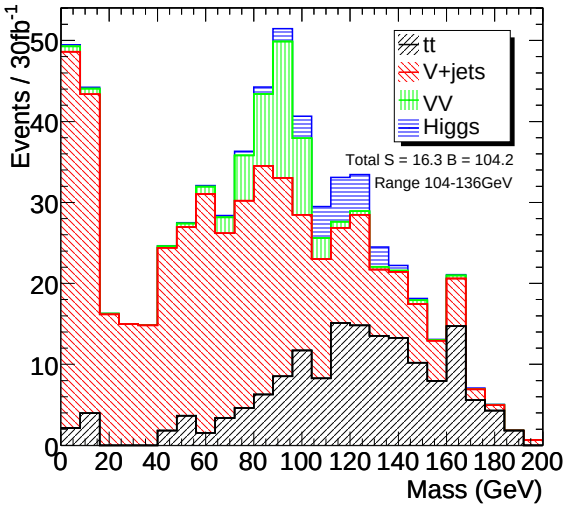


Figure 11: Masses of H candidates after $E_{\mathrm{T}}^{\mathrm{miss}} b\bar{b}$ selection

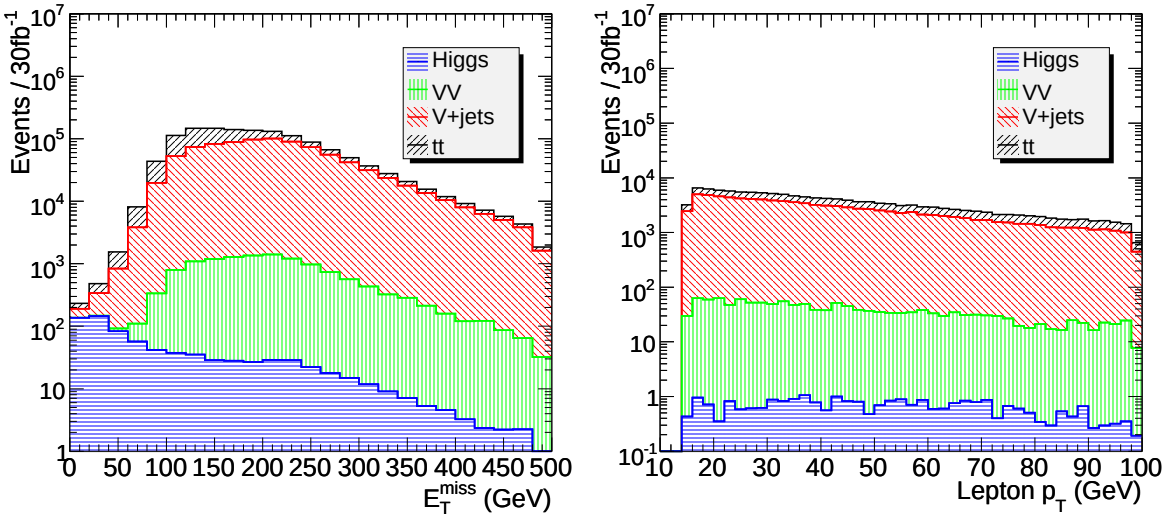


Figure 12: $E_{\mathrm{T}}^{\mathrm{miss}}$ and lepton p_{T} spectra, note that $E_{\mathrm{T}}^{\mathrm{miss}}$ spectra at low scales are significantly biased by generator level cuts

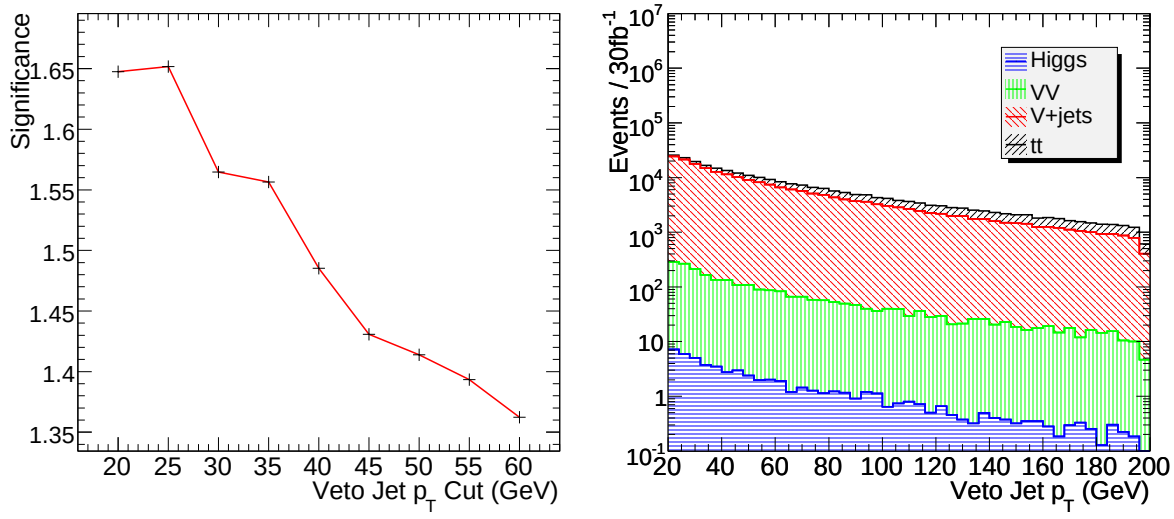


Figure 13: Left: Scan of possible p_{T} cut values for jets to be vetoed. Right: p_{T} spectrum of additional jets that are considered in the jet veto.

	$ZH(120)$	$WH(120)$	WW	WZ	ZZ
Generated	379 ± 2	1254 ± 8	18975 ± 86	9549 ± 61	4650 ± 15
$p_{\perp jet} > 200GeV$	166 ± 1	607 ± 5	9440 ± 61	4531 ± 42	2060 ± 10
$p_{\perp H} > 200GeV$	133 ± 1	513 ± 5	7321 ± 54	3560 ± 37	1573 ± 9
$\lvert\eta_H\rvert < 2.5$	130 ± 1	505 ± 5	7114 ± 53	3465 ± 37	1537 ± 9
$E_{\mathrm{T}}^{\mathrm{miss}} > 200GeV$	96.4 ± 0.9	71 ± 2	3403 ± 36	2047 ± 28	1132 ± 7
$p_{\perp lepton} < 30GeV$	94.8 ± 0.9	25 ± 1	2151 ± 29	1637 ± 25	1063 ± 7
$p_{\perp vetojet} < 30GeV$	62.7 ± 0.7	10.5 ± 0.7	1060 ± 20	914 ± 19	699 ± 6
$d\phi_{E_{\mathrm{T}}^{\mathrm{miss}}H} > 1.2$	62.7 ± 0.7	10.4 ± 0.7	1059 ± 20	914 ± 19	699 ± 6
b-tagged ($w > 1.0$)	20.9 ± 0.4	3.1 ± 0.4	1.2 ± 0.7	14 ± 2	36 ± 1
$104GeV < m_H < 136GeV$	14.4 ± 0.3	1.9 ± 0.3	0.0 ± 0.0	0.8 ± 0.6	3.3 ± 0.4

	$Z + jets$	$W + jets$	Wt	$t\bar{t}$
Generated	1031751 ± 833	2249139 ± 1445	168765 ± 233	729000 ± 816
$p_{\perp jet} > 200GeV$	438541 ± 543	1022862 ± 975	47314 ± 124	389393 ± 596
$p_{\perp H} > 200GeV$	266352 ± 423	630230 ± 765	38666 ± 112	331419 ± 550
$\lvert\eta_H\rvert < 2.5$	261090 ± 419	617078 ± 757	38418 ± 111	328448 ± 547
$E_{\mathrm{T}}^{\mathrm{miss}} > 200GeV$	188447 ± 356	291114 ± 520	11428 ± 61	82062 ± 274
$p_{\perp lepton} < 30GeV$	185851 ± 354	195853 ± 426	6696 ± 46	51338 ± 216
$p_{\perp vetojet} < 30GeV$	110517 ± 273	92467 ± 293	1344 ± 21	5375 ± 70
$d\phi_{E_{\mathrm{T}}^{\mathrm{miss}}H} > 1.2$	110516 ± 273	92426 ± 293	1343 ± 21	5373 ± 70
b-tagged ($w > 1.0$)	279 ± 14	127 ± 11	32 ± 3	129 ± 11
$104GeV < m_H < 136GeV$	32 ± 5	17 ± 4	11 ± 2	41 ± 6

Table 6: Expected number of events in 30fb^{-1}of data in the $E_{\mathrm{T}}^{\mathrm{miss}}b\bar{b}$ channel after each selection criterion is applied.

- $mu40$

The efficiencies for these trigger items have been studied in fully-simulated $l\nu b\bar{b}$ samples. Additionally, efficiencies for the non-isolated alternatives of the lower-threshold items have been measured as a reference, since it is well-known that the isolation criteria tend to be too strict for leptons with momenta significantly above the low-p_{T} thresholds.

Fig. 14 shows the efficiency for offline electrons that satisfy the $isEM = medium$ requirement. For electrons with true p_{T} above 25 GeV, while the $e25$ trigger item has an inclusive efficiency of $(98.6 \pm 0.2)\%$, the isolation requirement in $e25i$ starts to remove electrons at transverse momenta higher than 150 GeV, and the inclusive efficiency is reduced to $(94.4 \pm 0.4)\%$. This loss can be addressed by using the $e105$ trigger item, which provides an inclusive efficiency of $(92.7 \pm 0.6)\%$ for electrons with true p_{T} above 105 GeV.

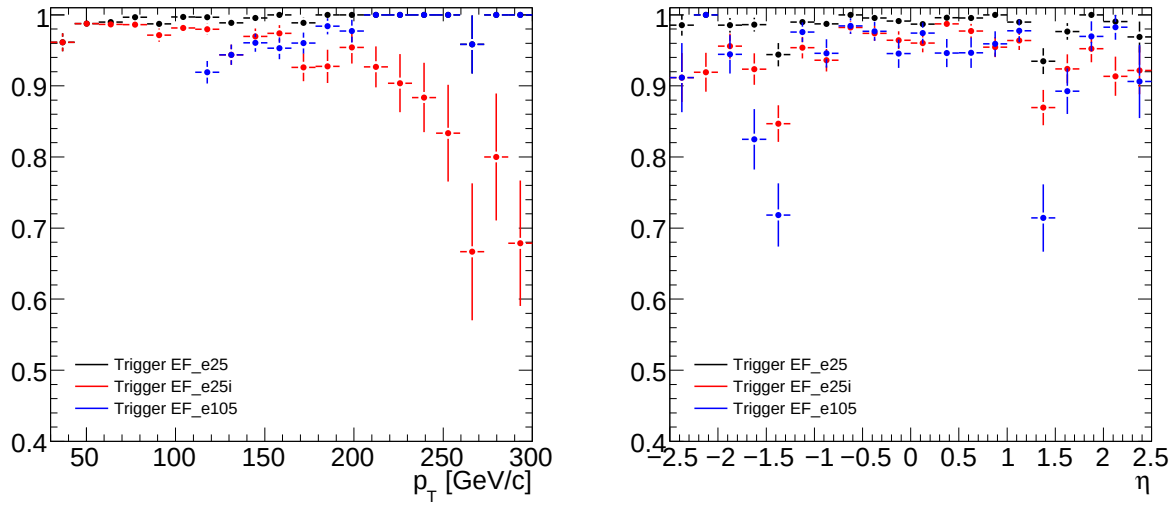


Figure 14: Single electron trigger efficiency for offline $isEM = medium$ electrons, as a function of p_T (left) and pseudorapidity (right). For the pseudorapidity distribution a cut on the transverse momentum of the Monte Carlo truth matched electron was done, corresponding to the trigger item threshold used (at 25 and 105 GeV).

The efficiencies of the trigger muon items are shown in Fig. 15 as a function of pseudorapidity and transverse momentum, measured for offline-reconstructed muons combined with inner detector tracks. For muons with true p_{T} above 20 GeV, the overall efficiency of the $mu20$ item is $(81.0 \pm 0.6)\%$, which is not high, but is well-understood (geometric acceptance of the L1 muon trigger). Further loss due to the additional isolation requirement brings the efficiency down to $(52.8 \pm 0.8)\%$ for $mu20i$. This dramatic loss is probably caused by the fact that the isolation requirement is based on an older version of the ATLAS detector geometry: The additional reduction in efficiency due to the isolation requirement on $mu20$-selected events is normally expected to be around 5%. This loss can be also recovered from quite effectively by using the $mu40$ trigger item in addition, which itself provides an efficiency of $(77.6 \pm 0.7)\%$ for muons of true p_{T} higher than 40 GeV.

The impact of the trigger requirements on the HV analyses can be investigated by applying the full *loose* offline selection on top of the trigger requirement. As an example, the effects of three different trigger strategies are shown for the WH signal in Table 7.

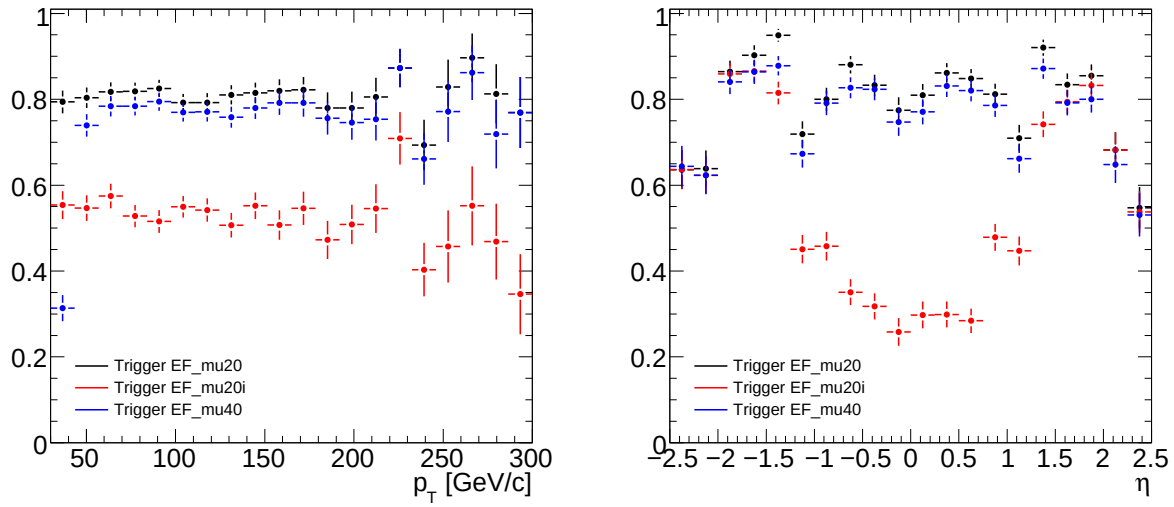


Figure 15: Trigger efficiencies for single-muon trigger items, measured for offline combined muons, as a function of p_{T} (left) and pseudorapidity (right). For the pseudorapidity plot, a cut on the transverse momentum of the Monte Carlo truth-matched electron was done, corresponding to the trigger item nominal threshold used (at 20 and 40 GeV).

	No trigger	Lepton triggers	Lepton triggers w/o isolation	Lepton + MET·jet triggers
After filter cuts	849.2 ± 7.6	849.2 ± 7.6	849.2 ± 7.6	849.2 ± 7.6
1 Higgs candidate	551.6 ± 3.6	232.5 ± 3.4	287.4 ± 3.6	313.5 ± 3.7
filtered $p_{\mathrm{T}} > 200$ GeV	497.7 ± 3.7	210.1 ± 3.3	260.1 ± 3.5	285.0 ± 3.6
$E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV	355.5 ± 3.7	183.2 ± 3.1	212.4 ± 3.3	254.3 ± 3.5
p_{T}(W)> 200 GeV	167.8 ± 3.0	130.2 ± 2.7	134.8 ± 2.8	166.0 ± 3.0
p_{T}(e/μ)> 30 GeV	140.6 ± 2.8	123.9 ± 2.7	125.9 ± 2.7	139.7 ± 2.8
p_{T}(additional μ)< 10 GeV	139.8 ± 2.8	123.2 ± 2.7	125.1 ± 2.7	138.8 ± 2.8
p_{T}(additional e)< 10 GeV	137.5 ± 2.8	121.1 ± 2.7	123.1 ± 2.7	136.5 ± 2.8
$\Delta\phi$(W,H)$< \frac{2}{3}\pi$	136.6 ± 2.8	120.5 ± 2.6	122.4 ± 2.7	135.6 ± 2.8
no additional b-jets $p_{\mathrm{T}} > 15$ GeV	123.8 ± 2.7	109.4 ± 2.5	111.3 ± 2.6	122.9 ± 2.7
add. jets on W side $p_{\mathrm{T}} < 60$ GeV	109.7 ± 2.5	97.2 ± 2.4	98.7 ± 2.4	109.0 ± 2.5
add. jets on H side $p_{\mathrm{T}} < 60$ GeV	98.9 ± 2.4	87.5 ± 2.3	89.0 ± 2.3	98.3 ± 2.4
one subjet b-tagged	88.8 ± 2.3	78.8 ± 2.2	80.0 ± 2.2	88.2 ± 2.3
both subjets b-tagged	44.6 ± 1.7	39.8 ± 1.6	40.4 ± 1.6	44.2 ± 1.7
loose fit cuts	44.4 ± 1.7	39.7 ± 1.6	40.3 ± 1.6	44.1 ± 1.7
$112 < m_H < 136$ GeV	29.5 ± 1.4	26.1 ± 1.3	26.5 ± 1.3	29.2 ± 1.4
trigger efficiency	-	(88.1 ± 0.6)%	(89.5 ± 0.6)%	(99.4 ± 0.2)%

Table 7: Expected number of events through the various analysis selection criteria for the WH signal, compared for different trigger selection strategies. Numbers are projected to 30 fb^{-1} of collected data.

The lepton trigger combination referred to in the table is an OR combination of the $mu20i + mu40 + e25i + e105$ trigger items. The efficiency for the same combination, but with the isolation requirements removed ($mu20$ and $e25$), is also shown in order to give an idea on the impact on the trigger efficiency: while the replacing $e25i$ with $e25$ does not does not make much of a difference in the overall efficiency, removing the muon isolation brings a small improvement

of about 1%. As mentioned earlier, this small loss is likely to disappear in the newer releases of ATLAS software, therefore we conclude that the overall efficiency expected from the lepton triggers is about 90%.

On the other hand, the remaining 10% is unlikely to be recovered with any combination of lepton triggers, since it is caused by the limited acceptance of the L1 trigger muons chambers (TGCs and RPCs). Therefore, we investigate trigger items from the jetTauEtmiss slice. While most jet triggers are expected to be heavily prescaled up to very high momenta, a jet AND $E_{\mathrm{T}}^{\mathrm{miss}}$ combination trigger with relatively low thresholds, $XE70 \cdot J62$, is expected to be among those that will be unprescaled at $2{\times}10^{33}$ cm^{-2} s^{-1} [2]. This signature is interesting not only when a genuine $E_{\mathrm{T}}^{\mathrm{miss}}$ signature is present, but also when we lose muons due to L1 acceptance, creating fake missing momentum. It is not yet clear whether recovering high-p_{T} muons this way is actually foreseen by the ATLAS L2 and EF trigger strategy, but, as can be seen in Fig. 16, it indeed appears to be quite effective. As such, when we combine the lepton triggers ($mu20i + mu40 + e25i + e105$) with the $E_{\mathrm{T}}^{\mathrm{miss}}$·jet trigger, an overall trigger efficiency of $(99.4 \pm 0.2)\%$ is obtained for our offline selected events in the $l\nu b\bar{b}$ channel.

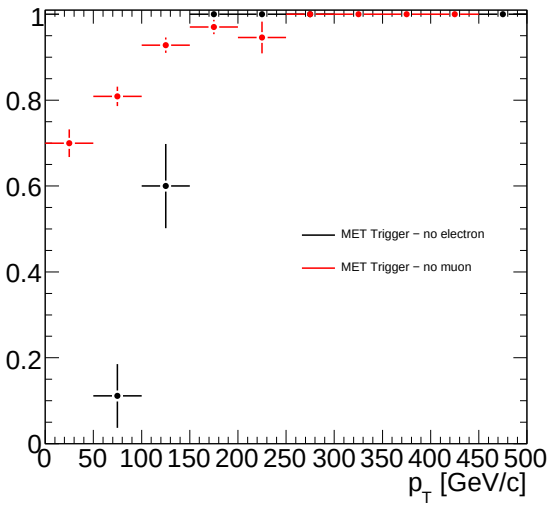


Figure 16: Efficiency of the $E_{\mathrm{T}}^{\mathrm{miss}}$·jet trigger as a function of the true p_{T} of the lepton from the W boson, for all events which didn't trigger any of the lepton triggers already considered for the WH channel, at the analysis stage after generator filter cuts.

As can be expected, among the channels studied, $llb\bar{b}$ provides the highest overall trigger efficiency due to the presence of two charged leptons. For the electron events, the efficiency of $e25i + e105$ is $(98.6 \pm 0.6)\%$, and for the muon events, the efficiency of $mu20i + mu40$ is $(94.5 \pm 1.1)\%$, in agreement with what one would expect from two charged leptons that can be considered mostly independent of each other for providing single-lepton trigger objects. Di-lepton trigger chains with lower p_{T} thresholds, namely $2e12i$ and $2mu10$ do not significantly increase the overall efficiency since the momentum of the leptons of interest are already higher than the single-lepton trigger thresholds. The efficiencies of these di-lepton triggers themselves are shown on the left-hand side of Fig. 17.

While the di-lepton chains are not of much use, the excellent performance of the $E_{\mathrm{T}}^{\mathrm{miss}}$·jet trigger can also be used to recover the $\mu\mu b\bar{b}$ events lost due to the L1 acceptance. The combined efficiency of $mu20i + mu40 + J80\_xe70$ is an excellent $(99.8 \pm 0.2)\%$.

Finally, the other ZH channel, $E_{\mathrm{T}}^{\mathrm{miss}}b\bar{b}$, does not benefit from the lepton triggers at all. However the large missing energy should be easy to trigger on. While it is not clear what the actual trigger threshold will be at high luminosity, it is likely to be around $100 - 125\,\mathrm{GeV}$- therefore events with $E_{\mathrm{T}}^{\mathrm{miss}} > 200\,\mathrm{GeV}$should have close to 100% efficiency. We have also considered the $E_{\mathrm{T}}^{\mathrm{miss}}$-jet triggers for this channel. The efficiency of such a trigger, $J80\_xe70$, is shown in Fig. 17 (right) as a function of the true p_{T} of the Z bosons for generated ZH events before any offline selection criteria are applied. The overall effiency is a remarkable $(97.2 \pm 0.4)\%$.

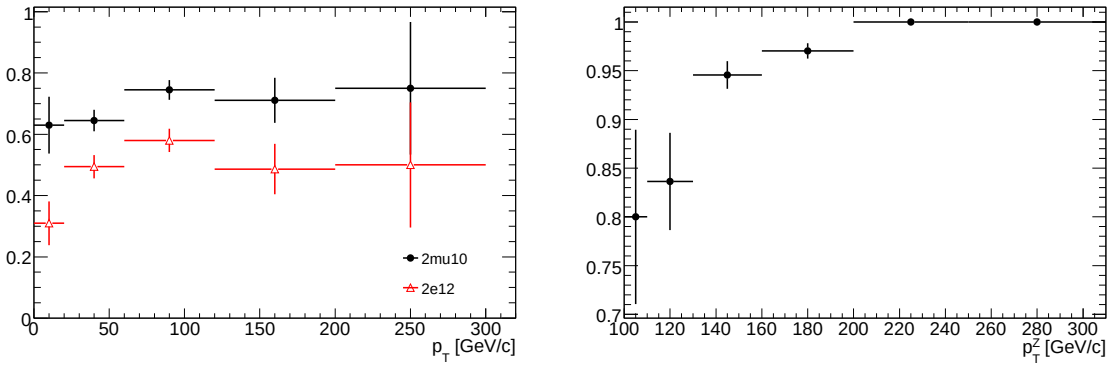


Figure 17: Left: Efficiencies of $2e12i$ and $2mu10$ trigger chains vs. the true p_{T} of the lower-p_{T} lepton from the Z_{ll} decay. Right: Efficiency of the $J80\_xe70$ trigger chain as a function of the true p_{T} of the Z boson, for the signal events generated in the $E_{\mathrm{T}}^{\mathrm{miss}}b\bar{b}$ channel (with true $E_{\mathrm{T}}^{\mathrm{miss}} > 100\,\mathrm{GeV}$).

In conclusion, due to the presence of high-p_{T} leptons and high $E_{\mathrm{T}}^{\mathrm{miss}}$, no significant loss is expected at the trigger for events from any of the studied channels.

6 Jet clustering: properties and optimization

6.1 Comparison to Traditional Jet Techniques

The present study is based on the reconstruction of the Higgs boson decaying into a pair of b-jets, by means of the jet clustering procedure described in Section 2. Since this is a new technique, we here compare it to a more conventional technique using dijets identified with the $k_{\perp}$ algorithm with $\Delta R = 0.4$. To do so, the cut flow needs to be adapted a bit, with the following differences with respect to the default selection:

- The Higgs mass is based on the invariant mass of two jets with the highest $b-$weight, instead of the invariant mass of the highest two or (if present) three subjets of the Higgs candidate
- A shift is applied to the value of the dijet invariant mass (+6GeV), to account for a small difference in the absolute jet energy scale calibration.

No change is applied to the remaining selection cuts.

Applying the new selection, a very large background contribution is seen, yielding a significance $\frac{S}{\sqrt{B}}$ of around 1.5, which can be explained by the fact that on conventional jets there is no limit on the ΔR of the dijet pair, while, since the subjet clustering method starts from C/A

jets with $\Delta R = 1.2$, the two subjets aperture is limited to this ΔR value. Most of the W+jets and $t\bar{t}$ background extends to very large values of ΔR, a region where only a negligible amount of signal can be found.

Therefore a further set of cuts is applied to the candidate pairs of $k_\perp$ jets, analogous to the two Higgs subjets have to be selected :

- Di-jet symmetry $\left(\frac{min(p_T(j_1),p_T(j_2))\Delta R(j_1,j_2)}{mass(j_1 j_2)}\right)^2 > 0.1$
- Mass drop $\frac{\max(mass(j_1),mass(j_2))}{mass(j_1 j_2)} < \frac{1}{\sqrt{3}}$
- $\Delta R(j_1, j_2) < 1.2$.

The result is shown in Table 8, for the signal and the main backgrounds. Since no specific charm-rejection was implemented for the $k_\perp$ jets, this has to be compared with Table 9, where the subjets analysis is implemented, but the last cuts differ from what was presented in Table 3, because no specific charm-rejection is implemented.

The signal efficiency is 11 % lower than in the mainstream analysis and at the same time the $t\bar{t}$ background increases by more than a factor of 2. The $b-$tagging efficiency, even with the same cut value on the $b-$weight as in the standard analysis, turns out to be higher. However tightening this cut does not significantly improve the overall analysis significance, which, considering also the remaining (minor) backgrounds, is ≈ 2.1, against ≈ 2.7 of the subjets based analysis without using explicit charm-jet rejection.

	WH(120)	WZ	$t\bar{t}(p_T^{min})$	Wt	W+jets
After filter cuts	858.4 ± 6.4	3935 ± 26	1229506	93155	582364
1 Higgs candidate	848.80 ± 0.68	3832.4 ± 4.1	1221084	92386 ± 15	555598
filtered $p_T > 200$ GeV	320.9 ± 3.1	1225 ± 12	132699	16053	160325
Missing $E_T > 30$ GeV	243.8 ± 2.9	884 ± 11	106361	12789	124545
p_T(W) > 200 GeV	148.1 ± 2.4	586.0 ± 9.2	36891	6957 ± 44	89777
p_T(e/μ)>30 GeV	126.5 ± 2.3	479.7 ± 8.4	30139	5935 ± 41	77987
p_T(additional μ)<10 GeV	124.8 ± 2.3	454.5 ± 8.2	27647	5554 ± 40	77740
p_T(additional e)<10 GeV	123.2 ± 2.3	429.2 ± 8.0	24975	5057 ± 38	76643
$\Delta\phi$(W,H)< $\frac{2}{3}\pi$	123.0 ± 2.3	416.4 ± 7.9	22798	4814 ± 37	75571
no additional b-jets p_T >15 GeV	121.5 ± 2.2	414.4 ± 7.9	21638	4717 ± 37	75539
jets on W side p_T <60 GeV	83.6 ± 1.9	298.4 ± 6.8	7235	2478 ± 27	55571
cutOnDeltaR2Jets	55.4 ± 1.6	145.7 ± 4.9	1342 ± 29	742 ± 15	12534
jets on H side p_T <60 GeV	39.9 ± 1.4	122.4 ± 4.5	487 ± 17	364 ± 10	10947
one subjet b-tagged	34.7 ± 1.3	30.7 ± 2.3	317 ± 14	196.3 ± 7.7	549 ± 22
both subjets b-tagged	17.06 ± 0.90	11.1 ± 1.4	54.6 ± 5.9	24.0 ± 2.7	46.0 ± 6.5
loose fit cuts	17.01 ± 0.90	11.1 ± 1.4	53.9 ± 5.8	22.5 ± 2.6	43.2 ± 6.3
112 GeV < mass(H) < 136 GeV	11.27 ± 0.73	1.01 ± 0.41	14.4 ± 3.0	3.6 ± 1.1	8.3 ± 2.8

Table 8: The table shows the expected events going through the selection at each stage for the signal and the main backgrounds, when using conventional k_T jets instead of the subjet clustering procedure. Numbers are projected to 30 fb^{-1} of collected data.

Fig. 19 shows the same comparison for the distance in pseudorapidity between the two jets. The k_T jets require a minimum distance between the jets of $\Delta R = 0.4$: this doesn't seem to reduce the signal contribution significantly, but it avoids the W+jet contribution to populate significantly the lower Higgs candidate invariant mass region, which is very helpfull in providing a mass sideband to normalize this background component on data.

	WH(120)	WZ	$t\bar{t}(p_T^{min})$	Wt	W+jets
add. jets on W side p_T <20 GeV	83.2 ± 1.9	461.3 ± 8.3	7227	3343 ± 31	86087
add, jets on H side p_T <20 GeV	55.8 ± 1.6	275.6 ± 6.6	1895 ± 34	1142 ± 18	48229
one subjet b-tagged	46.7 ± 1.5	53.5 ± 3.0	1039 ± 26	529 ± 13	2247 ± 45
both subjets b-tagged	20.82 ± 0.99	17.4 ± 1.7	53.9 ± 5.8	25.8 ± 2.8	104.8 ± 9.8
loose fit cuts	20.77 ± 0.99	17.0 ± 1.7	52.7 ± 5.7	24.3 ± 2.7	103.0 ± 9.7
112 GeV < mass(H) < 136 GeV	14.26 ± 0.82	1.01 ± 0.41	8.2 ± 2.3	6.7 ± 1.4	9.2 ± 2.9

Table 9: Cut flow for the subjets based analysis, but without using a specific charm-jet rejection. Numbers are projected to 30 fb^{-1} of collected data, according to LO cross sections.

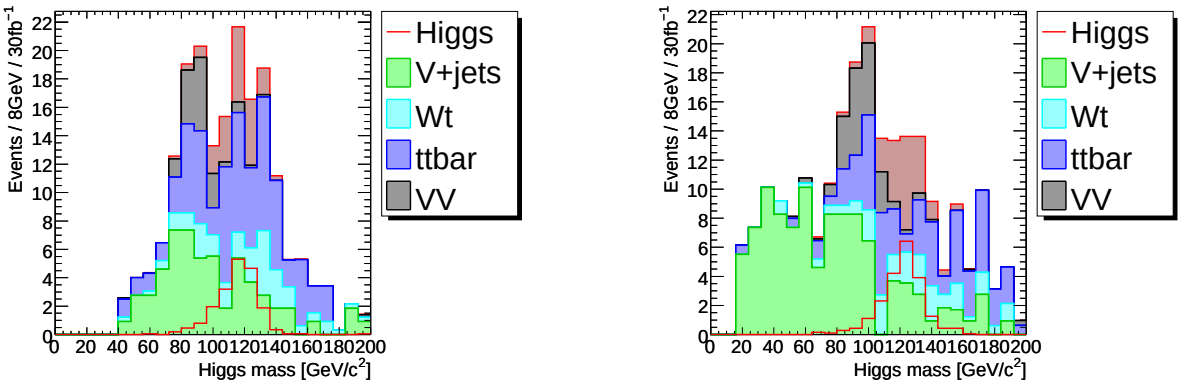


Figure 18: Invariant mass of the Higgs candidate for the conventional di-jet analysis (left) compared to the Higgs invariant mass obtained from the subjets based analysis (right), after full *tight* selection is applied. Numbers are projected to an integrated luminosity of 30 fb^{-1}.

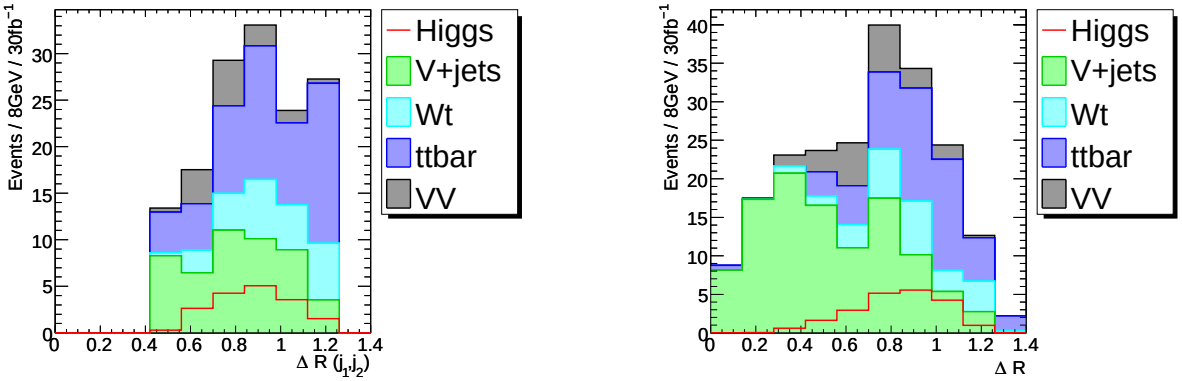


Figure 19: $\Delta R(j_1, j_2)$ between the two jets representing the Higgs candidate (left) compared to the same quantity referring to the distance in pseudorapidity between the two highest p_T subjets in the subjets based analysis (right), after full *tight* selection is applied. Numbers are projected to an integrated luminosity of 30 fb^{-1}.

The subjet clustering method seems to be able, for a fixed signal efficiency, to reject the background more efficiently. However this improvement, as shown in Fig. 20, gets visible only

after the application of b-tagging. Even looking only at the combinations which are matched at truth level with a combination of b- and c-jets, the subjets clustering method results in a higher amount of background before the b-tagging is applied. One possible explanation for this is that the three jet structure gets better decomposed by the subjet clustering method, so that, while one subjets turns out to be the b-jet, the other converges more frequently to one of the two single directions of the light or c-jet, moving in half of the cases away from the c-jet. However the algorithm which matches the jets with the flavour of the partons would still label also the light jet as a c-jet, since the c-jet would be in most of the cases still nearer than $\Delta R = 0.3$ to the c-quark. A confirmation of the fact that the subjet structure gets better decomposed is the fact that, in particular after b-tagging is applied, the mass tends to peak at higher values, towards the top mass, which corresponds to a higher amount of cases where all three subjets have been correctly included in the clustering.

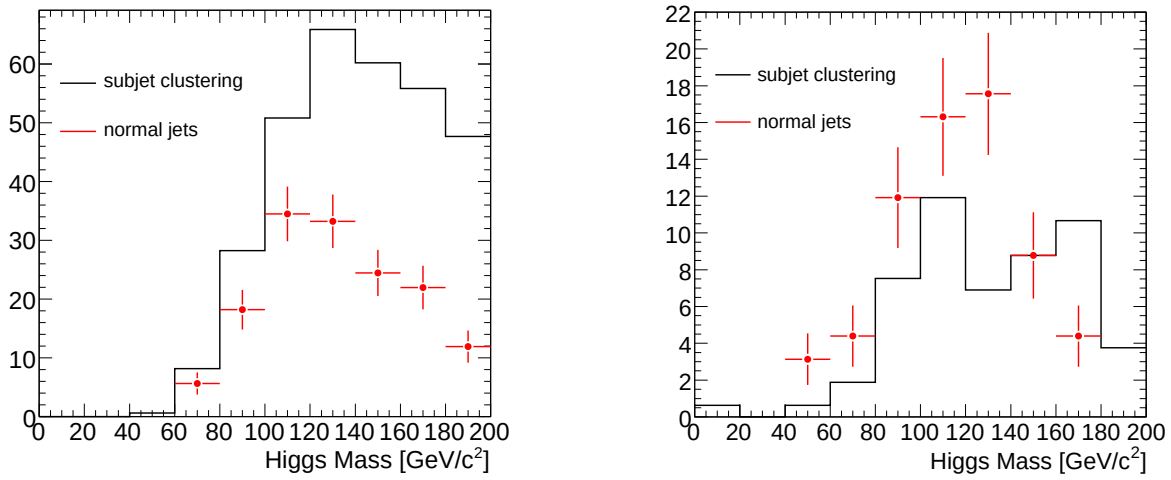


Figure 20: (Left) Invariant mass distribution of the Higgs candidate for the conventional di-jet analysis (red) compared to same quantity as obtained from the subjets based analysis (black), before applying b-tagging but selecting only the c-b jet combinations at true level. (Right) Same distribution, but after all analysis cuts. Both distributions are normalized to 30 fb^{-1} of integrated luminosity.

In addition the subjet clustering method is able to reconstruct subjets which are nearer in pseudorapidity than $\Delta R = 0.4$; this effect is however limited by the $b-$tagging algorithm, which provides an effective *turn on* curve as a function of ΔR. To clarify this effect, Fig. 21 shows the Higgs mass distribution for the subjet clustering method, normalized to the number of entries in the histogram, before and after applying $b-$tagging on both the subjets. It can be seen that the $b-$tagging efficiency drops for subjets nearer than $\Delta R = 0.2 - 0.3$. This effect is particularly important, because most of the higher p_{T} Higgs bosons, which are easier to separate from the $t\bar{t}$ background, accumulate at lower ΔR. This effect could be cured by developing a dedicated $b-$tagging algorithm, which simultaneously fits the two $PV \rightarrow b \rightarrow c$ decay chains expected in the two overlapping $b-$jets: this is however outside the scope of this note.

To recap, the parameters used for the subjet clustering algorithm in the present study are $\Delta R_{ij} = 1.2$, $y_{cut} = 0.1$, mass drop $= \frac{1}{\sqrt{3}}$, $p_{\mathrm{T}} > 200$ GeV and filtered $\Delta R = 0.3$.

A value of $\Delta R_{ij} = 1.2$ is nearly optimal, because it contains most of the two body decays of a Higgs boson with a $p_{\mathrm{T}} > 200$GeV. The mass drop value and the filtering ΔR were not changed

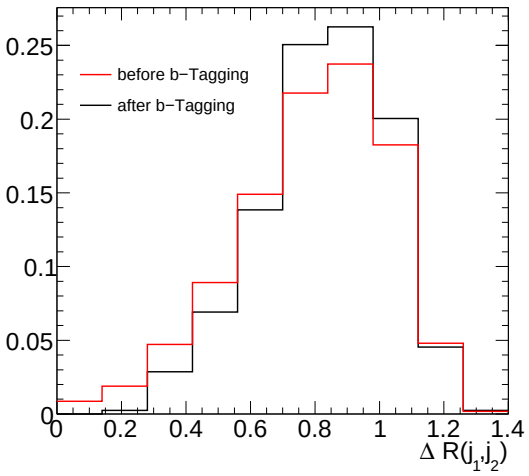


Figure 21: $\Delta R(j_1, j_2)$ between the two subjets representing the Higgs candidate before and after applying b-tagging on them. Both histograms are normalized to an area of 1.

during the present analysis, while different values were considered for y_{cut}, which provides an efficient way of removing very asymmetric subjets configurations (typical for example for QCD soft gluon emissions). This provides a way to reduce the W+jets background, but is not particularly well suited for rejecting the $t\bar{t}$ background, where the jets are emitted by the decay of top quark.

In Table 10 the signal and main backgrounds components are listed for different choices of the parameter y_{cut} (0.075,0.1 and 0.15), projected to an integrated luminosity of 30 fb^{-1}. Since the statistics is very low, the expected numbers of events without the final mass window cut are also listed. There is no dramatic change when varying the y_{cut} parameter, however the best significance is obtained for $y_{cut} \approx 0.1$. For this study a slightly tighter b-tagging requirement on the two leading subjets was applied with respect to the mainstream analysis.

Parameter y_{cut}	WH(120)	WZ	$t\bar{t}(p_T^{min})$	Wt	$W + b\bar{b}$
$y_{cut} = 0.075$	10.79 ± 0.72	1.01 ± 0.41	2.5 ± 1.3	2.12 ± 0.80	8.0 ± 1.2
$y_{cut} = 0.1$	10.79 ± 0.72	1.01 ± 0.41	2.5 ± 1.3	2.12 ± 0.80	7.9 ± 1.1
$y_{cut} = 0.15$	9.44 ± 0.67	1.01 ± 0.41	1.9 ± 1.1	2.12 ± 0.80	7.5 ± 1.1
Parameter y_{cut}	WH(120)	WZ	$t\bar{t}(p_T^{min})$	Wt	$W + b\bar{b}$
$y_{cut} = 0.075$	15.66 ± 0.86	12.7 ± 1.5	22.6 ± 3.8	8.5 ± 1.6	71.8 ± 3.4
$y_{cut} = 0.1$	15.42 ± 0.85	12.5 ± 1.4	21.3 ± 3.7	8.8 ± 1.6	69.6 ± 3.4
$y_{cut} = 0.15$	12.96 ± 0.78	12.3 ± 1.4	18.8 ± 3.4	8.2 ± 1.6	59.1 ± 3.1

Table 10: The table shows the expected number of events going through the full *tight* selection (top) and through the same selection expept for the mass windows cut (bottom), for the signal and the main backgrounds and for different choices of the parameter y_{cut}. Numbers are projected to 30 fb^{-1} of collected data.

Fig. 22 shows a comparison between the invariant mass distributions of the same three working points for y_{cut}. The effect of increasing the y_{cut} parameter on the WH signal is a decrease in efficiency and slight improvement in the mass resolution. This explains also why the expected number of events for the signal scales a bit differently when varying the y_{cut} parameter

before and after the mass window cut.

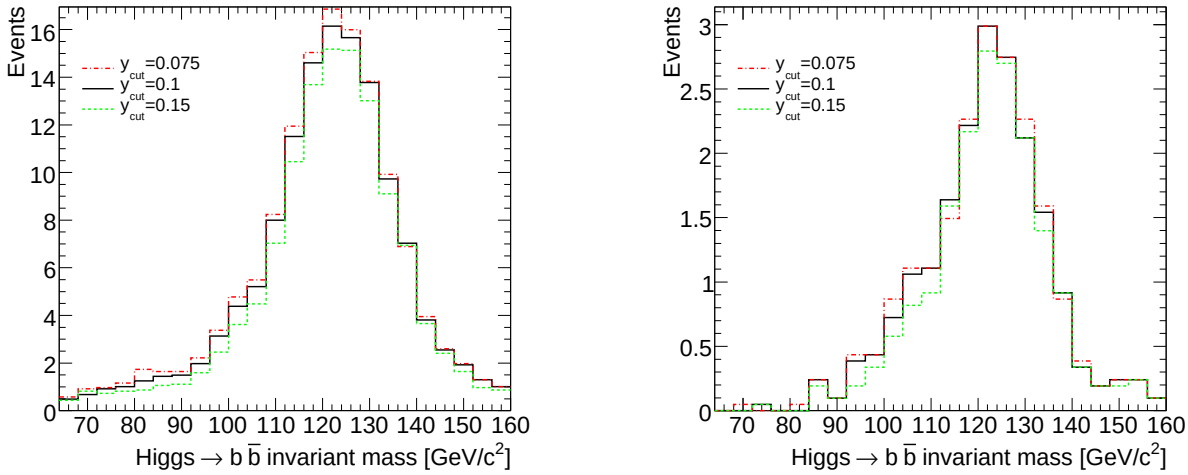


Figure 22: Invariant mass distribution of the Higgs candidate for various values of the y_{cut} clustering parameter, immediately after the additional lepton veto (left) and after the full selection (right). Numbers are projected to an integrated luminosity of 30 fb^{-1}.

6.2 Comparison of AOD and ESD based methods

As mentioned in Section 2, it is possible to perform the jet clustering steps either on AOD or ESD input. The limitation of the AOD based method is that only local calibration can be applied which is known to be incomplete. However the ESD method requires access to ESD datasets.

As a comparison of the two methods, a fit was applied to the signal peak shape (Fig. 23) to assess resolution effects, using the functional form

$$f(x; m, \sigma_{\pm}, \alpha_{\pm}) = exp\left[-\frac{(x-m)^2}{2\sigma_{\pm}^2 + \alpha_{\pm}(x-m)^2}\right] \quad (1)$$

The results of the fit can be found in Table 11.

Parameter	AOD-based	ESD-based
$mean$	117 ± 3 GeV	123 ± 1 GeV
σ_+	10 ± 2 GeV	9 ± 1 GeV
σ_-	14 ± 3 GeV	13 ± 1 GeV
α_+	0.1 ± 0.1 GeV	0.12 ± 0.02
α_-	0.15 ± 0.05 GeV	0.15 ± 0.02

Table 11: Fit parameters describing peak resolution for two different analysis methods. The AOD-based method is applied on the $WH \rightarrow l\nu b\bar{b}$ sample, while the ESD-based method is applied on the $ZH \rightarrow llb\bar{b}$ sample.

Since the absolute scale of the AOD-based jets was expected to be incorrect, a simple correction was applied of multiplcation by a scale factor of +6%. This factor was derived by

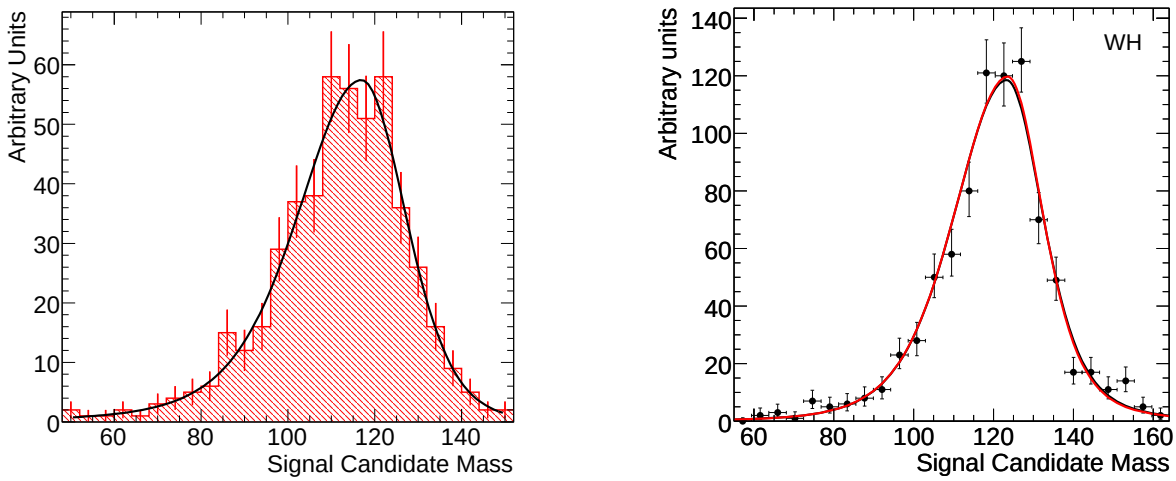


Figure 23: The peak observed in a $ZH \rightarrow llb\bar{b}$ sample after application of an AOD based analysis, compared to the peak in the $WH \rightarrow l\nu b\bar{b}$ sample after application of an ESD based analysis.

studying the energies of filtered Higgs candidate jets. Jets after the filtering procedure with a $p_{\perp} > 150GeV$ which match an equivalent truth jet with $dR < 0.3$ were selected and the fractional difference between the true and reconstructed energies plotted and fitted with a Gaussian as shown in Fig. 24. The mean of the Gaussian shown is -0.058 ± 0.002.

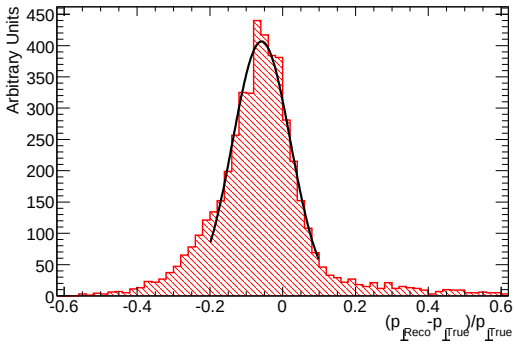


Figure 24: The fractional $p_{\perp}$ difference between hadron level and reconstructed jets in an HZ sample analysed at the AOD level.

As previously mentioned in Section 2. There is a small technical difference in the clustering procedures applied to AOD and ESD samples. In the AOD method, at the filtering stage, the 3 filtered subjets are the 3 highest $p_{\perp}$ objects in the jet (global). However in the ESD method, the first 2 filtered subjets are required to come from opposite halves of the jet (separate), as determined by the $R_b b$ splitting. A comparison was performed by applying both techniques to the same $HZ \rightarrow llb\bar{b}$ sample. The final distributions produced can be seen in Fig. 25. As expected, the differences are minimal. A similar comparison was performed using a Z+jets sample consisting of $700,000$ events and no difference was observed in the final candidate mass

spectrum.

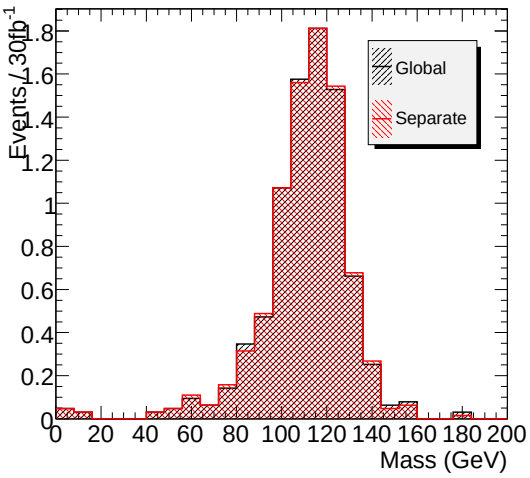


Figure 25: The same $HZ \to llb\bar{b}$ sample clustered using two separate techniques. The difference is not found to be significant.

6.3 Choice of μ parameter

As described in Section 2, one of the free parameters of the clustering procedure is μ which determines how large a drop in mass is considered a significant splitting.

For technical reasons, although the analysis on which this work is based [3] used a parameter of $\mu = 0.67$, this analysis uses $\mu = 1/\sqrt{3}$. Some testing was performed to ensure that this did not have a significant impact on the final analysis. A comparison between samples processed with the two different values of μ can be seen in Fig. 26 demonstrating that there is indeed no significant impact due to this change.

7 Mass resolution

The filtered four momentum of the Higgs candidate, obtained as described in Section 2, can be further refined. In fact, even in the case of using H1-style calibration, only the cell-by-cell part of the jet calibration is used. Several different types of jet-level correction can also be derived, and indeed for the more commonly used cone or $k_\perp$ jets, such and p_{T} and η dependent corrections are indeed used in ATLAS. In principle the same kind of additional corrections should be derived for Cambridge/Aachen algorithm and for the the special jet clustering method used in the present analysis. In this section we study some methods of improving the mass resolution of the Higgs candidate.

One such correction comes from the fact that the $b-$quarks produce $b-$hadrons which decay semi-leptonically into a muon or electron plus neutrino with a branching ratio of around 20 %. While correcting for the missing neutrino is more difficult and can be done on a statistical basis only, the muon, when present, can be just added to the subjet four momentum[3)]. Fig. 27(top) shows the Higgs candidate mass distribution (after having applied the basic H1-style cell based

[3)] The energy deposition of a muon in the calorimeter is expected to be negligible, around 2 GeV

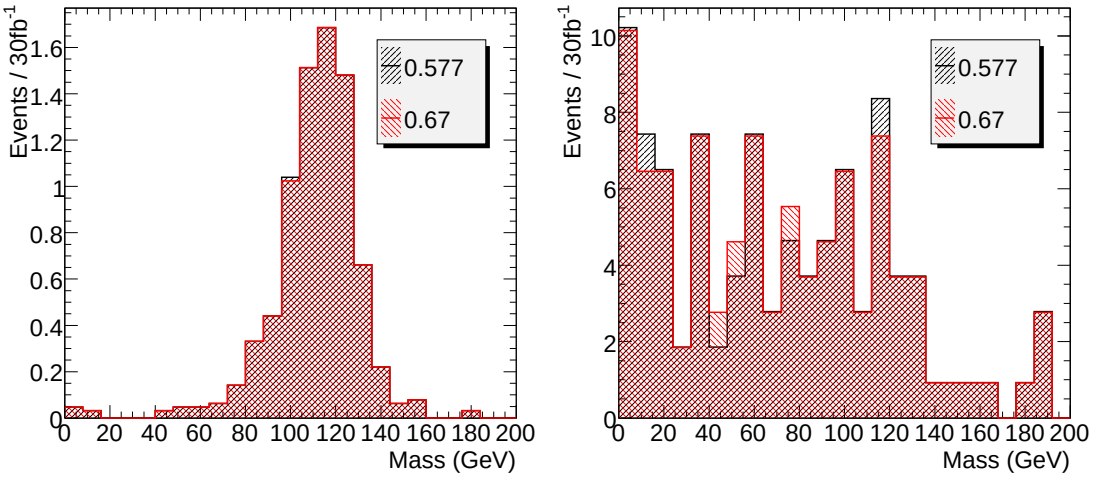


Figure 26: Comparison of signal and background mass distributions for two different μ values.

calibration and after all analysis selection cuts), while in Fig. 27(bottom) the Higgs candidate is corrected for the muon energy. In this way the low mass left tail is reduced and an improvement in the RMS of around ≈ 0.6 can be obtained. For computing mean values and RMS, only normalized mass residuals in a $(-30, 30)\%$ window are considered, to reduce the impact of outliers.

A further small improvement is expected correcting for the missing energy of the neutrinos from the semileptonic $b-$decays, but this is not studied here.

A final p_{T} and η dependent jet scale correction is not straightforward to perform using the standard calibration tools available in ATLAS, because the two subjets in which the Higgs candidate decays (or a similar boosted heavy objects) are strongly correlated: when the first subjet misses energy, most probably it was taken by the second subjet. The correlation in energy between the two subjets is illustrated in Fig. 28, which displays the average shift in p_{T} for a subjet with respect to the true p_{T} of the underlying b-quark as a function of the subjet p_{T}, separating the subjets into two categories on an event basis: given the two subjets of a certain Higgs candidate, the one nearest in ΔR to the originating parton goes into one category, the remaining one in the other. The former tends to lose part of its energy, the latter acquires energy from the other jet. This illustrates the difficulty of deriving a calibration for the individual subjets; this is a more acute instance of the fact that calibrating to an initial (coloured) parton is in principle wrong, since QCD radiation implies that any such correction will be highly model dependent and measurements based on it will not in fact be physically well-defined. However, this effect does not strongly feed into the Higgs mass resolution, since the hardest gluon radiation from either of the b-quarks is included in the Higgs candidate (see Section 2). And since the Higgs itself is a colour singlet, it undergoes no QCD radiation itself, and a MC-based calibration to the Higgs mass itself will be less model-dependent than an attempt to calibrate to leading-order quarks or gluons.

Following the above arguments, a complete calibration strategy for a boosted heavy object reconstructed with this particular jet clustering method could consist of using high statistics Monte Carlo sample with different Higgs masses to minimize the deviation of the reconstructed mass from the MC truth mass, applying corrections which depend on the p_{T} and η of either the subjets or of the whole candidate system. In order to perform the same calibration on data,

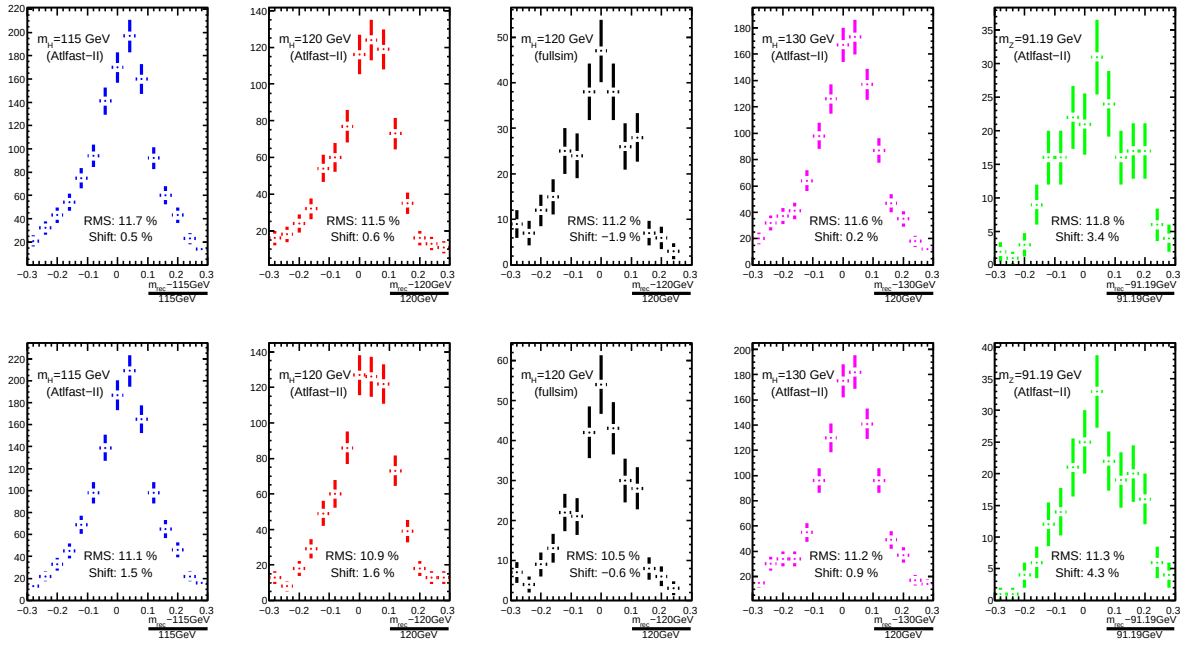


Figure 27: (Top) Higgs candidate mass distribution as delivered by jet clustering, after having applied the H1-style cell based calibration, for the $WH \to l\nu b\bar{b}$ process and different masses and for the $WZ \to l\nu b\bar{b}$ process. In the case of a 120 GeV Higgs mass the mass distribution on a sample produced using the complete full simulation of the ATLAS Detector is also shown for comparison. (Bottom) Higgs candidate mass distribution as shown on the left, but accounting also for the four momentum of a muon from a semi-leptonic $b-$decay, when present.

the decay products of a boosted W bosons for example in $t\bar{t}$ events could be used (the $b-$scale correction being obtained elsewhere and applied on top of it). Eventually, the Z peak seen in ZZ events in the dilepton channel (Section 4.2) will provide a good calibration, but this will require very large luminosities.

Only a very basic jet scale calibration will be performed here, where a relative invariant mass shift correction is applied as a function of the transverse momentum of the Higgs candidate. This relative shift is shown in Fig.29 as a function of the Higgs candidate p_{T}; to increase the available statistics all the Monte Carlo samples produced with the AtlFast-II simulation for Higgs masses of 115, 120 and 130 GeV are summed up together.

We correct for this effect by fitting a simple function of the form:

$$f(p_{\mathrm{T}}) = c_0 - \frac{c_1}{p_{\mathrm{T}}^6}$$

to describe this dependence. The fit result is shown in Fig.30, first with the full selection applied (left) and then, to crosscheck the validity of the functional choice, at an earlier analysis stage (just after the additional b-jet veto cut), with more statistics. It is no wonder that the $b-$tagging requirement affects the invariant mass distribution, since it requires the two subjets to be real $b-$subjets and thus removes much of the the combinatorial background present in the signal. The resulting coefficient and fit quality values are listed in Tab. 12.

After applying the p_{T} dependent calibration function to the signal Monte Carlo Samples, one

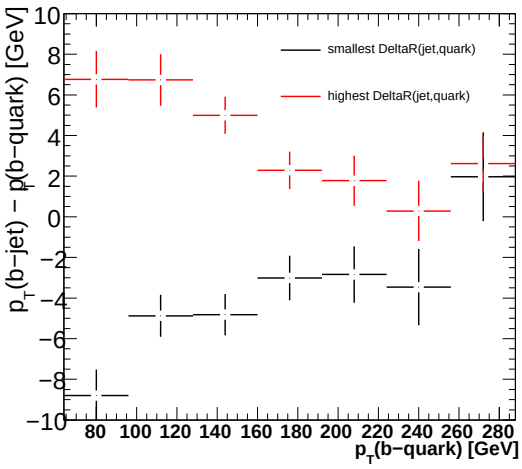


Figure 28: The figure shows the difference in p_{T} of the reconstructed subjet with respect to the b-quark p_{T} (before *FSR*), as a function of subjet p_{T}. For each Higgs candidate the two subjets are subdivided into two categories, in black the one nearest to the originating b-quark, in red the other.

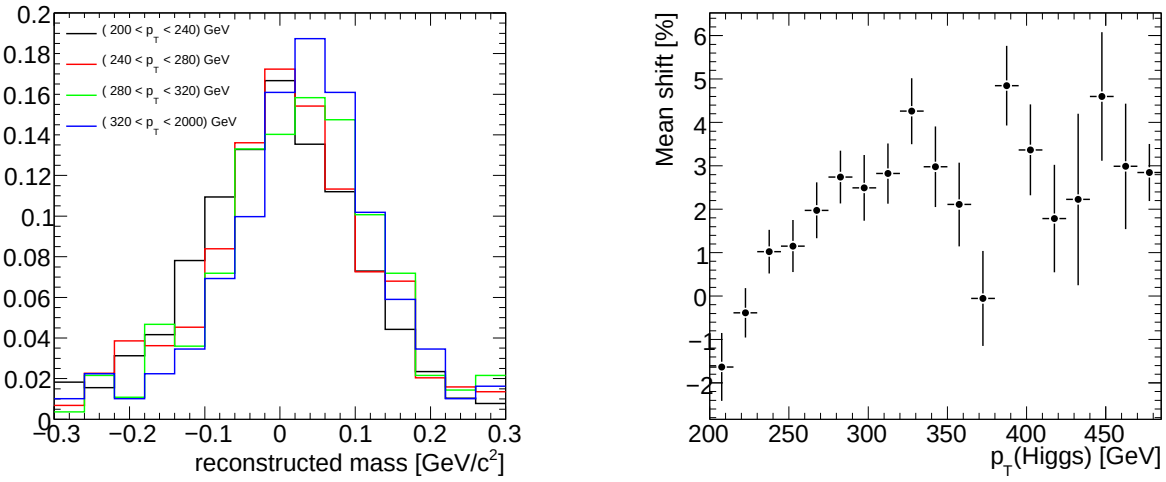


Figure 29: Pull of the reconstructed Higgs mass in various bins of $p_T(Higgs)$ (left) and average shift of reconstructed Higgs mass as a function of $p_T(Higgs)$. Full analysis preselection cuts are applied.

obtains the mass distributions shown in Fig. 31. The resolution improves another bit, the Higgs mass scale tends to be corrected, but some residual dependence on the Higgs candidate invariant mass remains. This indicates that a more complex calibration procedure is needed to account for the non-linearity in the reconstructed mass (at lower mass a bias is visible to smaller values, at higher mass to higher values). Given the statistics available for the study and the scope of this note, we will consider this calibration as acceptable. For the cut based analysis of the $l\nu b\bar{b}$ channel presented in Sec. 4.1, only the muon energy correction was taken into account, whereas

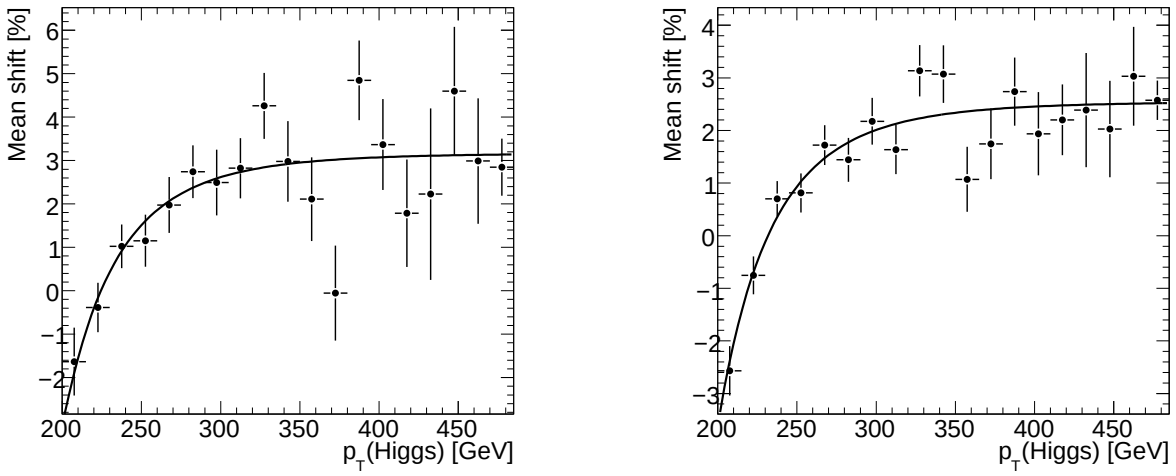


Figure 30: Fit of the dependence of the mass shift as a function of Higgs candidate p_T, after all selection cuts atr applied (left) and just after having applied the additional b-jets veto (right). A χ^2 based minimization procedure is used.

Parameter	value
c_0	3.17 ± 0.25
c_1	$(4.05 \pm 0.53) \times 10^{14}$ MeV
χ^2/ndf	11.31/17
Prob(χ^2)	72%

Table 12: Fit results to fix the jet scale as a function of Higgs candidate p_T.

the absolute energy scale and p_T(Higgs) dependent calibration was only used for the fit based approach presented in Ref. [28], where it is particularly important that the Higgs invariant mass distribution doesn't depend on p_T(Higgs).

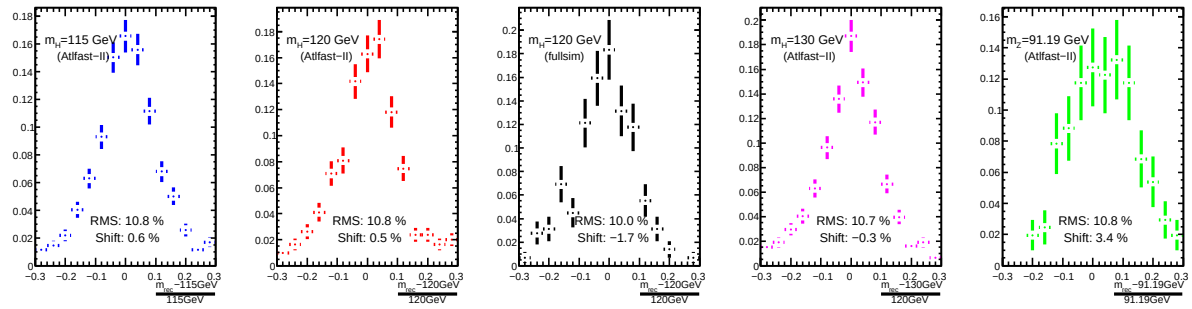


Figure 31: Mass distribution after application of simple p_T dependent calibration. The full simulation is shown just for comparison purposes, as well as the $Z \to b\bar{b}$ peak from the WZ sample.

8 B-tagging performance

The identification of the two b-quarks originating from the Higgs boson is crucial for separating the signal from the large backgrounds, most of which are dominated by u, d and s quark jets.

A key point of this study is to demonstrate, using a realistic Monte Carlo simulation, that achieving a good b-tagging performance on the two $b-$subjets into which the Higgs boson decays is possible. All generated Monte Carlo events used here were passed through the full simulation of the ATLAS Inner Detector: since b-tagging relies primarily on the charged particle tracks originating from the fragmentation of a b-jet as reconstructed in the tracking devices, the b-tagging results are expected to be as realistic as the full simulation of the ATLAS Detector.

In order to apply b-tagging on them, tracks have to be assigned to a certain subjet. The simplest way to perform this task is to take all tracks whose momenta at the point of closest approach to the interaction point on the transverse plane are within a certain ΔR from the subjet direction as determined by the jet clustering algorithm. For this subjet direction three different strategies have been studied;

- Using the directions of the two unbranched subjets before the filtering procedure
- Using the direction of two or three subjets as determinined after filtering the unbranched subjets, applying b-tagging on all of them and choosing the two with highest b-jet probability in the end
- Using the direction of subjets as determinined after filtering the unbranched subjets, choosing the two highest p_{T} ones in the end.

The subjet direction after filtering yields a better estimate of the direction of the outgoing b-hadron, since it filters out contributions from QCD radiation, and thus improves b-tagging performance. At the same time, it turns out that applying b-tagging to all the subjets after filtering, and then choosing the two with highest b-weight, increases considerably the amount of $t\bar{t}$ background. So the last of the three listed strategies is adopted in the following.

For the distance in pseudorapidity ΔR between the subjet direction and the track momenta a fixed value of 0.4 is used, without any dependence on $\Delta R(subjet1, subjet2)$, p_{T} or η of the two subjets: such a dependence might be used in future to improve further the b-tagging performance. Every track can be assigned to one only subjet, so in case of overlaps a track is assigned to the nearest in ΔR of the two subjets.

The direction of the subjet momentum has been corrected to take the displacement of the primary vertex in z with respect to the nominal position of the interaction point into account. In fact the direction associated with each calorimeter cell is only correct in the hypothesis of a particle originating in (0,0,0). While the transverse displacement of the primary vertex with respect to the nominal position of the beam spot is negligible, the longitudinal displacement is not and can be accounted for with the correction:

$$\theta_{corrected} = arccotan\left(cotan\left(\theta_{subjet}\right) - \frac{z_{PV}}{r_{cal}}\right)$$

where r_{cal} is the average distance in the transverse plane from $(0,0)$ of the energy deposits in the calorimeter. Even though for r_{cal} a jet based value could be used, we use an average value of 1.750 m, which has been chosen by minimizing the subjet direction mean squared residual. This correction provides a small improvement in the subjet direction resolution with respect to the b-quark direction.

No attempt has been made in this study to combine the information coming from the *spatial* b-tagging algorithms (based on the b-hadron lifetime) with the b-tagging algorithms based on the reconstruction of a lepton from a semileptonic b-hadron decay.

After the tracks are matched to the subjets, b-tagging is applied on them, running all standard Athena b-tagging algorithms[4]. The output is then retrieved for the combined and "JetFitter" based algorithms, using the strings "COMB" and "JetFitterCOMBNN".

The b-tagging performance has been analyzed by considering the two highest p_{T} subjets from all Higgs candidates present in the three following samples: WH(120), $t\bar{t}$ and W+jets. All selection cuts have been applied, apart from the additional jet veto and the b-tagging cuts, and the mass window cut has been loosened from 100 to 140 GeV, to increase the available statistics, still remaining in the kinematic region relevant for the present analysis. According the the conventional definition widely adopted in ATLAS, a subjet is labelled as a jet of a certain quark flavour (b, c or light) by exploiting Monte Carlo truth information: it is labelled as a b-jet if a b-quark with $p_{\mathrm{T}} > 5$ GeV is present nearer than $\Delta R = 0.3$ from the sub jet axis; if not, if a c-quark with $p_{\mathrm{T}} > 5$ GeV is nearer than $\Delta R = 0.3$ from the sub jet axis, it is considered as a c-jet, otherwise it is considered as a light-jet.

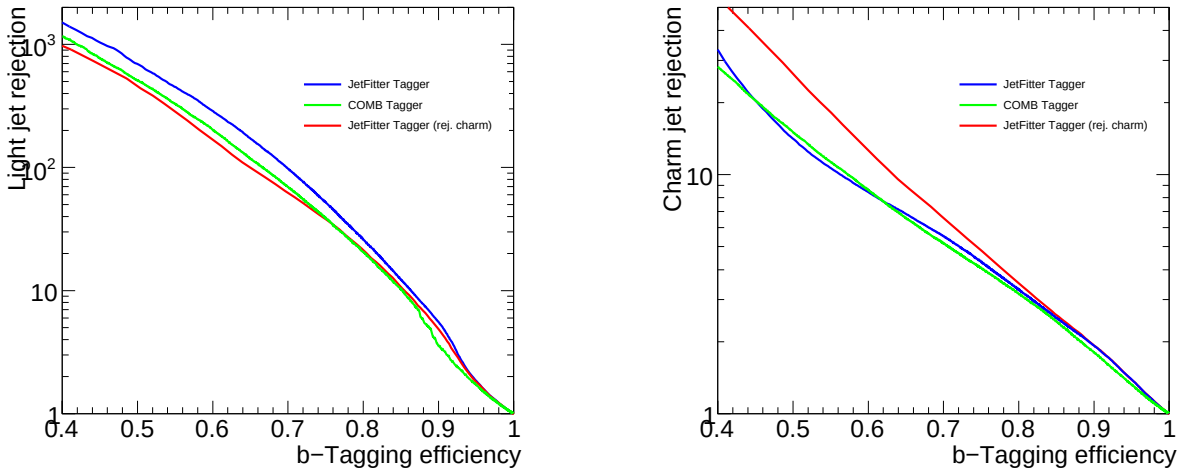


Figure 32: Rejection against light-quark jets (left) and against charm-quark jets (right) as a function of the B-tagging efficiency on the single subjet.

The rejection achievable against light- and c-quark jets is shown in Fig. 32 as a function of the b-tagging efficiency on the b-subjet[5]. Both the *JetFitter* and *COMB* Taggers are based on a weight of the following type:

$$w = log\left(\frac{P(b)}{P(b) + P(l)}\right), \tag{2}$$

where P(b) is the posterior probability to be a b-jet, P(u) the posterior probability to be a light-jet, and are trained to get the best possible separation of $b-$jets from light-quark jets. Looking at the performance of the two algorithms, it turns out that both perform very well, with JetFitter

[4] b-tagging algorithms are run using Athena rel. 14.0.0, with no relevant change applied, and release 13 calibration files for "conventional" jets are used for the *PDFs*.

[5] The rejection is defined as the inverse of the misidentification efficiency.

performing better in rejecting light-quark jets, in particular at b-tagging efficiencies below 80 %.

In Table 13 the rejection against light-quark jets for few fixed b-tagging efficiency points is listed explicitly. The region of b-tagging efficiencies which is relevant for the present analysis is between 60 % and 75 % (which would yield a signal efficiency on the Higgs candidate, under the hypothesis of uncorrelated b-weights, between 36 and 56 %).

b-tagging efficiency	COMB	JetFitter	JetFitter (rej. charm)
40 %	1162 ± 143	1450 ± 75	906 ± 37
50 %	510 ± 16	691 ± 25	441 ± 13
60 %	203 ± 4	287 ± 7	160 ± 3
70 %	69 ± 1	98 ± 1	60.6 ± 1
80 %	20.6 ± 0.1	26.1 ± 0.2	21.1 ± 0.1

Table 13: Rejection against light-quark jets at different subjet b-tagging efficiencies. The various b-tagging algorithms are described in the text.

At 70 % b-tagging efficiency (corresponding to $\approx 50\%$ signal efficiency), a rejection of light quark jets around 100 is expected to be achievable using *JetFitter*, corresponding to 1% light-jet misidentification efficiency. This is exactly the value considered in the generator level study [3] in the most optimistic scenario.

Such an efficient rejection of the background is however only valid in the hypothesis that the background is dominated by light-jets, so that charm-jets do not play an important role. We will however see (e.g. Fig.35) that, at least in the WH analysis, apart from the large irreducible $W + b\bar{b}$ component, the background is dominated by combinations of a light- and a charm-jet faking the two subjets representing the decay products of the Higgs boson.

b-tagging efficiency	COMB	JetFitter	JetFitter (rej. charm)
40 %	28 ± 1	31 ± 1	51 ± 1
50 %	15.0 ± 0.2	14.1 ± 0.2	25.6 ± 0.5
60 %	8.6 ± 0.1	8.4 ± 0.1	12.2 ± 0.2
70 %	5.15 ± 0.04	5.52 ± 0.04	6.47 ± 0.06
80 %	3.18 ± 0.02	3.30 ± 0.02	3.48 ± 0.02

Table 14: Rejection against c-quark jets at different subjet b-tagging efficiencies. The b-tagging algorithms used are described in the text.

Since charm-quarks fragment into c-hadrons which possess a significant lifetime and have similar decay multiplicities as b-hadrons, separating b-jets from c-jets is much harder than separating b-jets from light jets, as the right plot in Fig. 32 shows. The rejection values for some fixed efficiency points are also listed in Table 14. As a result, the rejection of charm-jets is, given a certain b-tagging efficiency, more than an order of magnitude worse with respect to rejecting light-jets.

In order to improve the rejection against charm-jets, the b-tagging algorithm based on *JetFitter* provides additional information[6)], which make a better separation against charm-jets

[6)] in particular it tries to identify the $PV \to b \to c$ decay chain topology, which is not present in a c-jet

possible by the use of a discriminator of the kind:

$$w = log\left(\frac{P(b)}{P(b) + c_l P(l) + (1 - c_l)\, P(c)}\right), \tag{3}$$

where P(c) is the posterior probability for a certain subjet to be a c-jet and c_l is a number between 0 and 1 providing the a-priori probability for a background subjet to be a c- or a light-jet.

A coefficient $c_l = 1$ makes the new discriminator in Eq.3 equivalent to the old in Eq. 2. Conversely, a coefficient $c_l = 0$ provides the highest rejection against c-jets, at the cost of a reduced light-jet rejection. The performance corresponding to $c_l = 0$ is shown for both rejecting light- and charm-jets in Fig.32, Table 13 and Table 14, with the label "Jet Fitter (rej. charm)". In general the optimal choice for the coefficient c_l depends on the flavour content of the background in term of fraction of c- and light-jets.

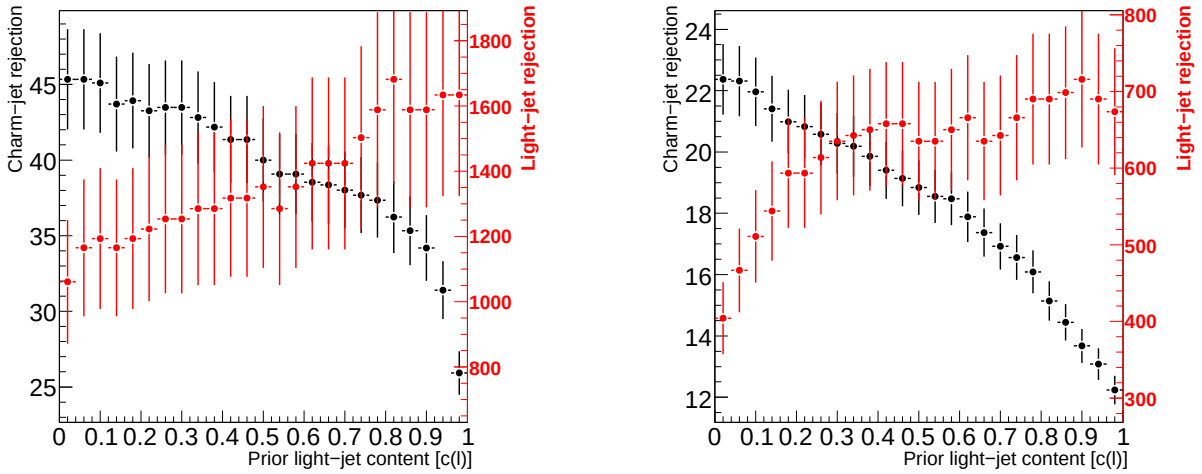


Figure 33: Rejection of c-jets (black) and light-jets (red) as a function of the coefficient c_l for $b-$subjet efficiencies of 40 (left) and 50 (right) %.

The performance in term of light- and c-jet rejection achievable at fixed b-tagging efficiencies of 40, 50, 60 and 70 % as a function of the coefficient c_l is shown in Fig.33 and Fig.34. At all efficiencies shown except for 40 %, it seems that the light-jet rejection saturates at a value of c_l around 50 % (0.5). This is probably due to the fact that the calibration of the algorithm was performed on generic Monte Carlo samples and on "conventional" jets in the present of quite a different event and jet topology. As a result, the c_l no longer has a clear statistical interpretation and has to be considered as an effective parameter: at each possible $b-$tagging efficiency, given a certain background composition, there is one only value for c_l which minimizes the amount of background.

As was shown in Table 3, where the tighter selection cuts were applied, the most important backgrounds are $t\bar{t}$ and W+jets. In order to understand the impact of b-tagging on the analysis, it is worth to analyse the flavour composition of the Higgs candidates selected in these two main backgrounds, and the breakdown of the different subjets flavour combinations as a function of the signal efficiency corresponding to a certain $b-$tagging cut on the two subjets. This is shown in Fig.35 for $t\bar{t}$ and in Fig.36 for W+jets. All remaining selection cuts, in the *tight* version, including the final mass window cut, are already applied, except for b-tagging.

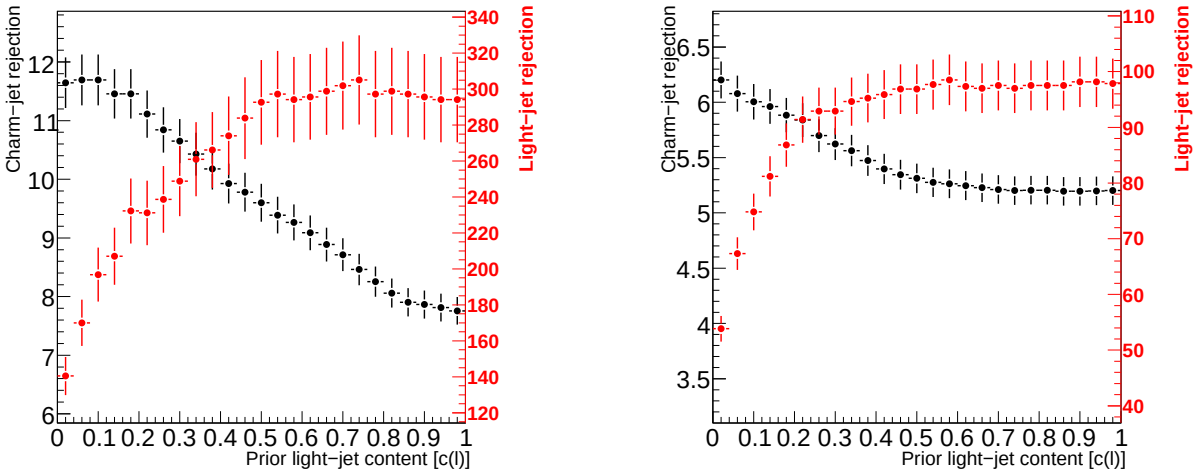


Figure 34: Rejection of c-jets (black) and light-jets (red) as a function of the coefficient c_l for $b-$subjet efficiencies of 60 (left) and 70 (right) %.

As a reference, the number of signal events is also shown in the plots. In the $t\bar{t}$ background the dominant contribution is given by subjet combinations with $b-l$ and $b-c$ subjets faking a Higgs candidate.

In general, in order to pass the signal selection, the two top quarks need to be highly boosted. One of the top quarks produces in a very asymmetric way a highly boosted W boson, which is required in the signal selection, together with a low p_{T} b-jet, which passes through the $b-$ and light jet veto. The opposite high p_{T} top quark produces a b-jet and a W boson which decays hadronically: out of these three jets, the combinations are selected where one jet is lost or has low p_{T} and the other two fake the Higgs subjets. However the combinations with both subjets originating from the W boson are highly suppressed, since their invariant mass peaks at the W boson mass, so the main remaining contribution comes from the $b-$jet combined with one of the two jets from the W boson ($b-l$ or $b-c$). The cases where a $b-b$ pair is selected as Higgs candidate in $t\bar{t}$ is much more rare, since in this case one $b-$quark needs to come from one top and the other b from the other top; so the top quarks need to have a lower p_{T}, which makes the simultaneous production of a highly boosted W boson very difficult.

While $b-$tagging can easily reduce the number of $b-l$ subjet combinations to an acceptable level, it is much harder to reduce the $b-c$ component. It can be however seen (in the bottom plot of Fig. 35) that the explicit rejection of charm-jets by JetFitter can help in rejecting $b-c$ subjet combinations, at the cost of letting through more $b-l$ combinations.

In $W+$jets the dominant contribution is given by the $l-l$ and $l-c$ subjet combinations, as expected from pure QCD production. The most dangerous contribution comes however from $b\bar{b}$ pairs (e.g. from gluon splitting), which cannot be reduced by applying b-tagging, and up to a certain point also from the $l-c$ subjet combinations.

In order to determine the optimal $b-$tagging strategy for the present analysis, the significance, defined as $\frac{S}{\sqrt{B}}$ has been analyzed as a function of the signal efficiency given for a certain b-tagging requirement. This is shown in Fig. 37, which corresponds to applying the *tight* selection cut flow. For comparison, also the result including the higher statistics $W+b\bar{b}$ sample is shown.

Including a veto on the presence of an eventual third $b-$subjet improves the statistical

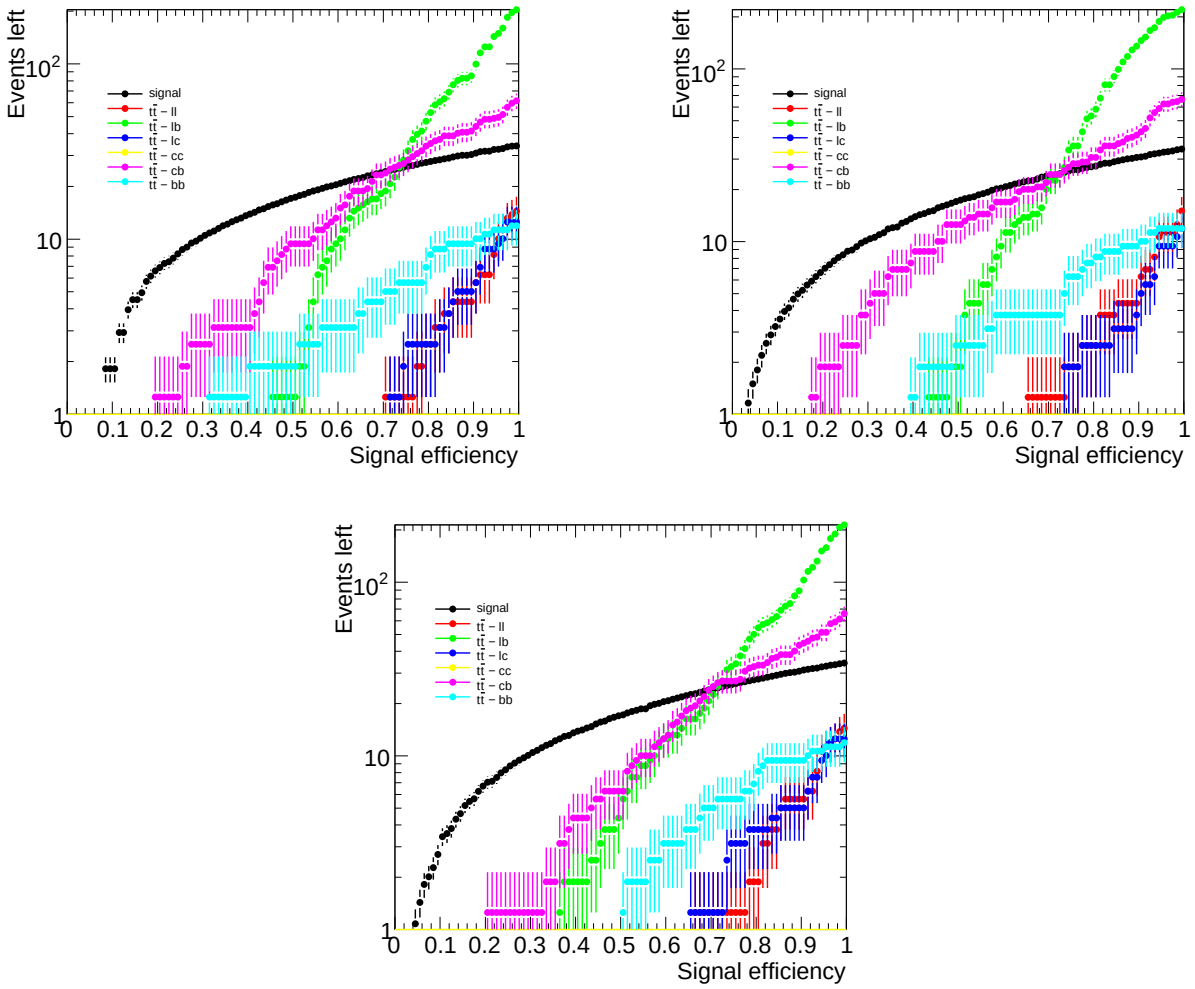


Figure 35: Number of events left as a function of the signal efficiency corresponding to applying a certain b-tagging cut for all various subjets flavour combinations present in the $t\bar{t}$ background. Different $b-$tagging algorithms are used: JetFitter (top left), COMB (top right), JetFitter rejecting charm-jets (bottom).

significance only marginally, therefore it was not included in the final analysis. The results for two different b-tagging strategies are shown: COMB and JetFitter with a value for c_l of 0.2.

The COMB Tagger yields a maximum significance of around 2.6 ± 0.2 at $\approx 45\%$ signal efficiency, while JetFitter with $c_l = 0.2$ provides doesn't provide a clear maximum, but is relatively flat at a significance of $3.0 - 3.1 \pm 0.3$ between $\approx 35 - 45\%$ signal efficiency. In addition, with JetFitter a significantly better signal over background ratio can be obtained, in particular at lower signal efficiencies.

The JetFitter algorithm and a working point at 40% signal efficiency[7)] has been chosen for the cut flow in the *tight* version described in Section 4 for the $l\nu b\bar{b}$ channel.

[7)] the quoted efficiency corresponds to around 35% $b\bar{b}$-tagging efficiency before mass window cut

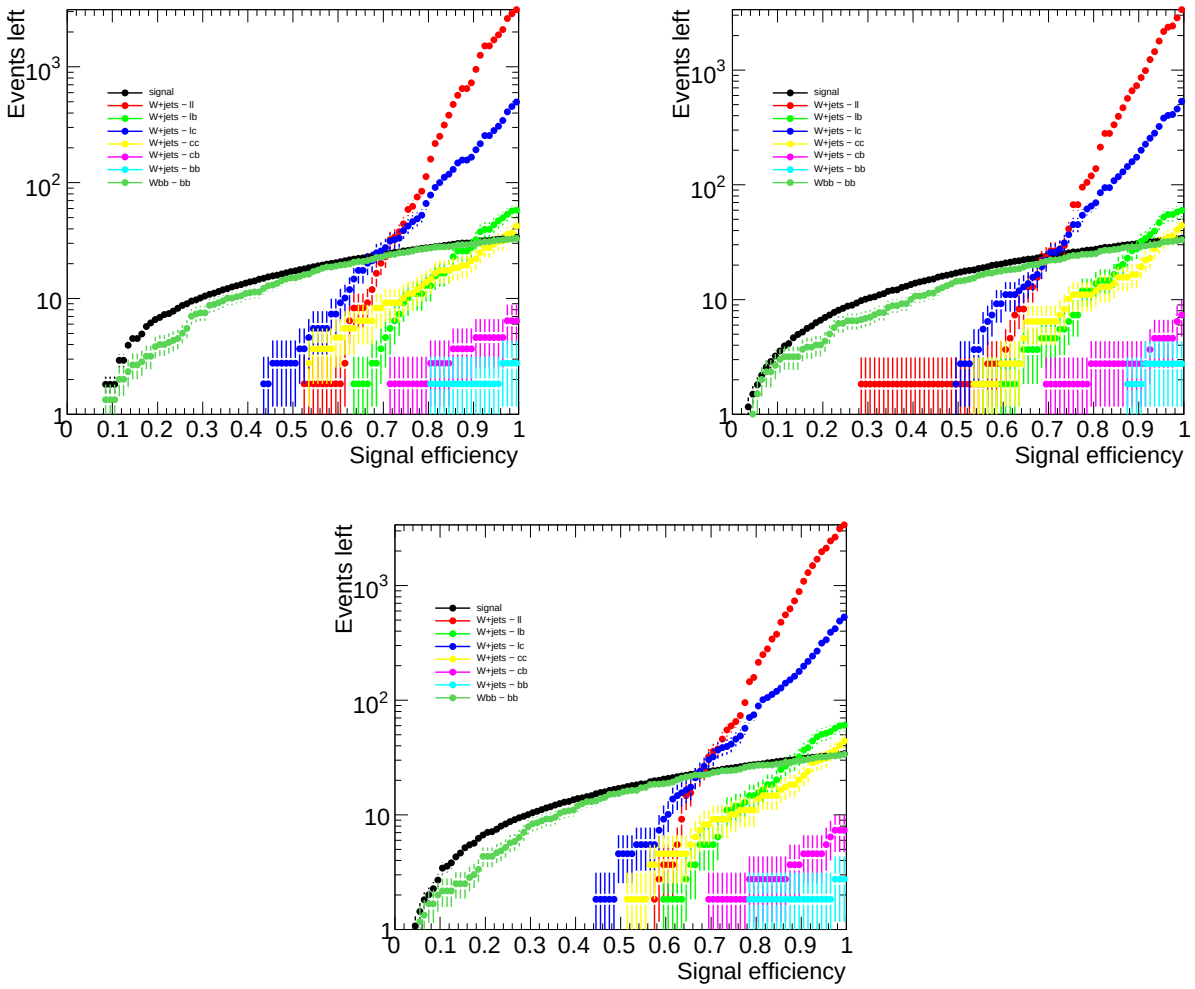


Figure 36: Number of events left as a function of the signal efficiency corresponding to applying a certain b-tagging cut for all various subjets flavour combinations present in the W+jets background. Different $b-$tagging algorithms are used: JetFitter (top left), COMB (top right), JetFitter rejecting charm-jets (bottom).

9 Lepton Reconstruction

The analysis of the $b\bar{b}\nu l$ relies on the identification of a high p_{T} lepton in the event, and that for $b\bar{b}ll$ relies upon finding two such leptons. The electron and muon efficiency as a function of p_{T} and η for the Higgs signal samples based on the Atlfast-II simulation is shown in Fig. 38, for all high p_{T} leptons from the decay of an associated W boson. In order to increase the available statistics no selection cuts are applied. The analysis relyies on requiring a *medium isEM* electron, while the veto is applied to *loose isEM* electron.

The same plots for the reconstructed *muons*, again from the W boson, are shown in Fig. 39. The analysis requires the presence of a *combined* muon, while for the veto a *standalone* muon is required.

The inefficiency around $\eta \approx 0$ is particularly important for this analysis, since an eventual high p_{T} muon landing in this region will escape the veto, and give a very big contribution to

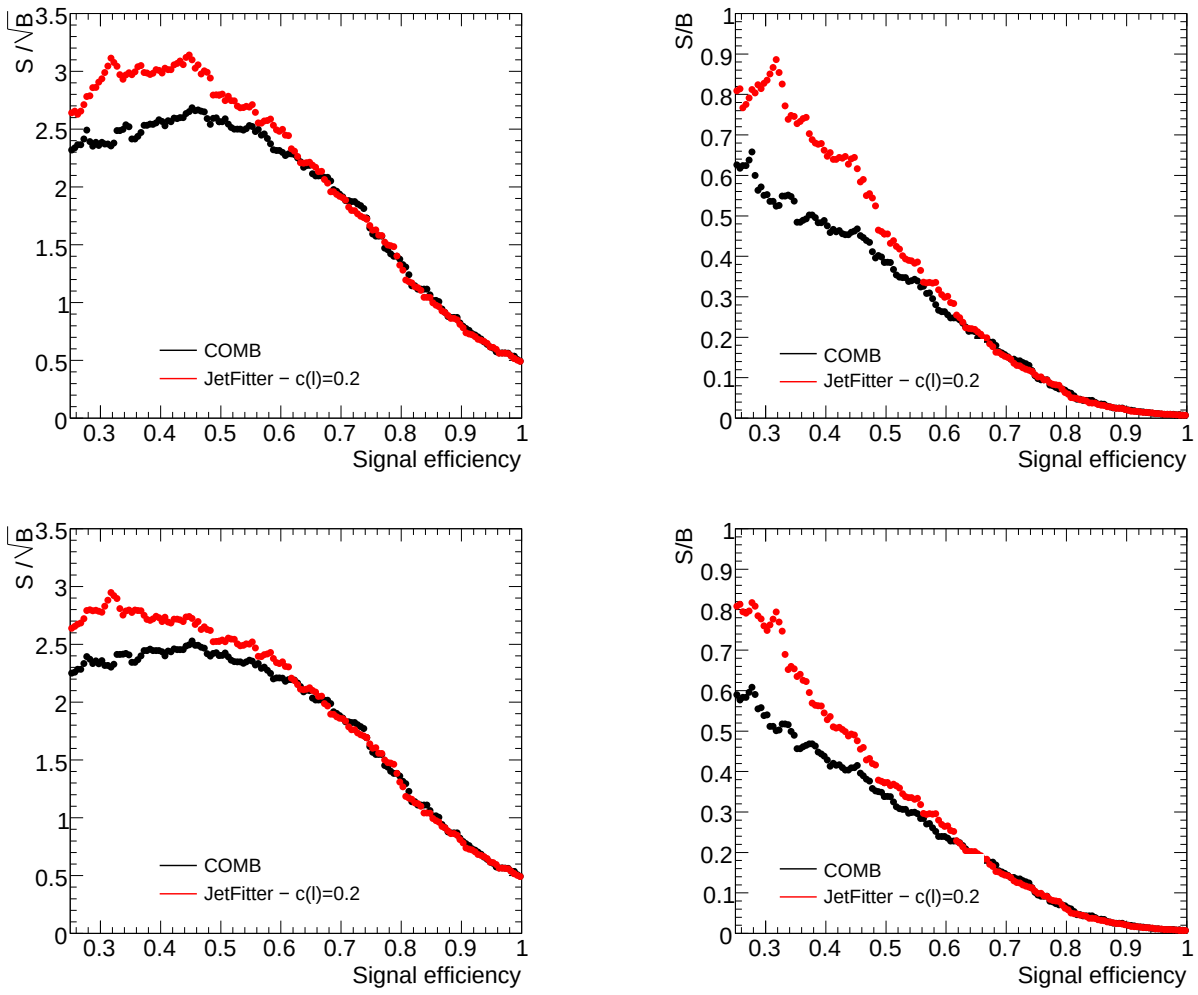


Figure 37: Statistical significance (left) and signal over background ratio (right) as a function of the signal efficiency corresponding to a certain b-tagging cut, for the reference Monte Carlo samples (top) and for the samples containing the higher statistics $Wb\bar{b}$ background component generated with AcerMC (bottom).

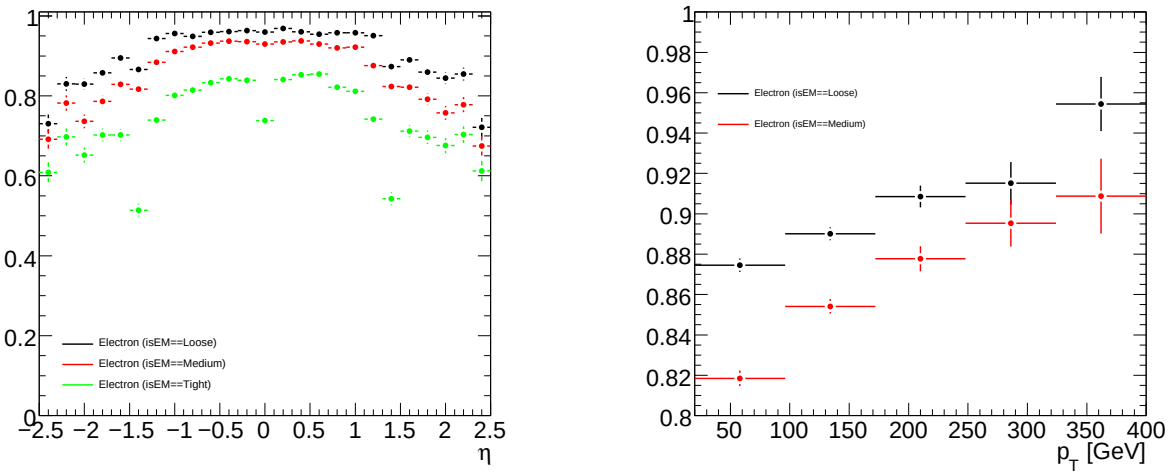


Figure 38: Electron efficiency as a function of electron pseudorapidiy η (left) and of electron p_{T} (right). Efficiencies are given for different object-ID definitions: isEM=Loose, isEM=Medium and isEM=Tight.

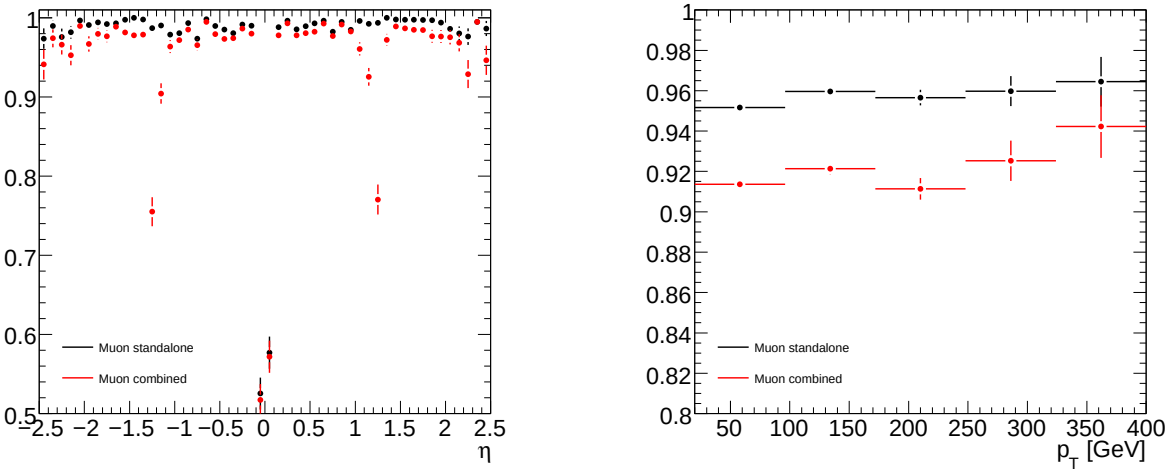


Figure 39: Muon efficiency as a function of muon pseudorapidiy *eta* (left) and of muon p_{T} (right). Efficiencies are given for muon as reconstructed in the Muon System and for muons combined with tracks reconstructed in the Inner Detector.

the $E_{\mathrm{T}}^{\mathrm{miss}}$. In this way a Z boson can be reconstructed as a W (in addition cases where the second lepton is outside acceptance, $|\eta > |2.5$ for the electron and $|\eta| > 2.7$ for the muon, also contribute to this).

Signal and background are dominated by cases where the selected high p_{T} lepton comes from a W boson (or in few cases from a Z boson): it turns out that these are isolated leptons, so if AtlFast-II can reproduce their efficiency correctly on the signal sample, it should be able to correctly reproduce their efficiency on the background too. Fig. 40 shows the ratio between the number of electrons in AtlFast-II and in full simulation, as a function of p_{T} and η.

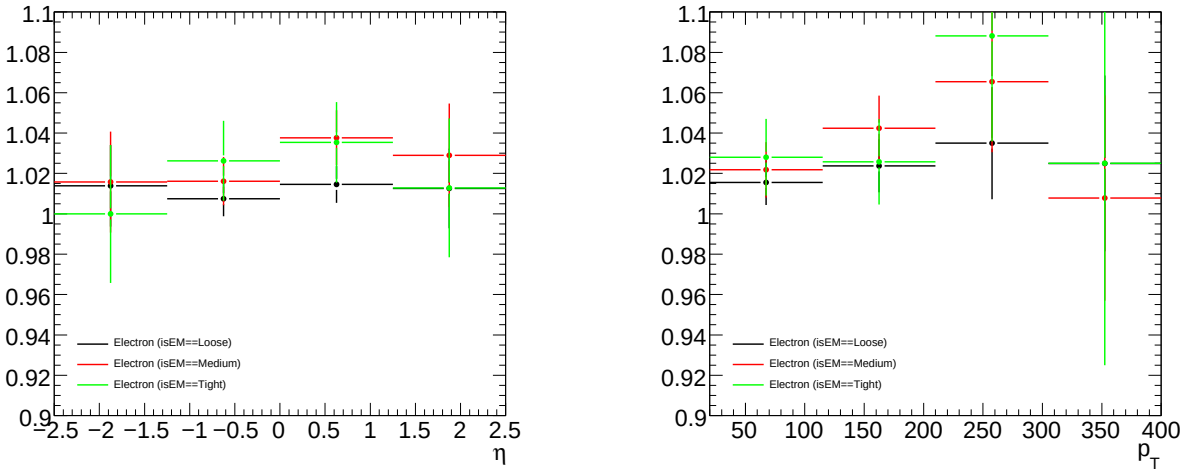


Figure 40: Ratio between electron efficiency in AtlFast-II and full simulation as a function η (left) and of p_{T} (right).

The same ratio is shown for muons in Fig. 41.

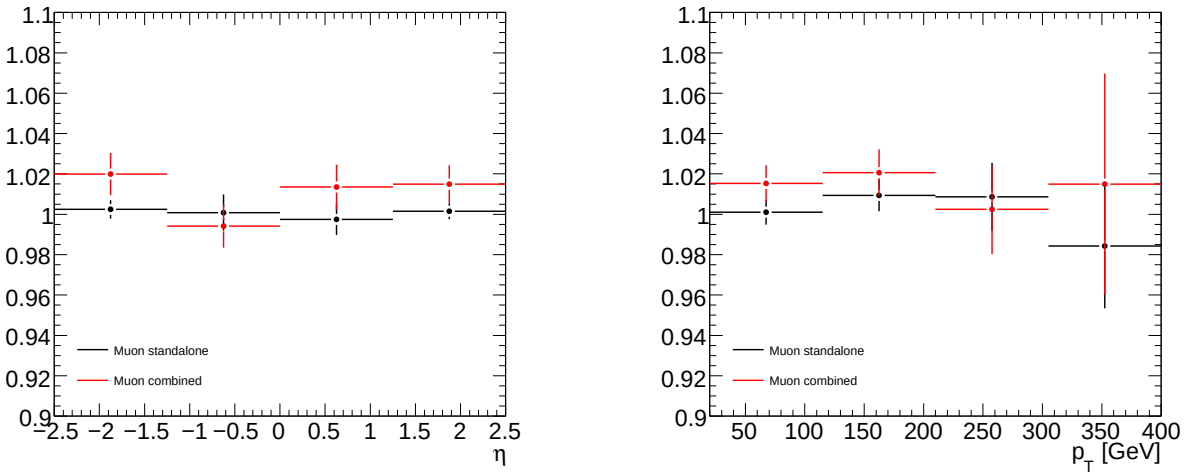


Figure 41: Ratio between muon efficiency in AtlFast-II and full simulation as a function η (left) and of p_{T} (right).

For the muons, AtlFast-II and the complete full simulation of the ATLAS Detector are expected to be the same, since all muons in the event pass through the full simulation of the ATLAS Detector.

The overall shifts (AtlFast-II versus complete full simulation) are shown in Tab. 15.

As can be seen in the table, AtlFast-II reproduces the muon efficiencies well (as expected), while a small shift to higher efficiency values is seen for the electrons. This at most a $2-4\%$ effect, but needs to be accounted for in the estimation of the signal efficiency for the Monte Carlo Toys.

Object	Definition	Shift AtlFast-II / FullSim (%)
Muon	standalone	-0.1 ± 0.4
Muon	combined	0.8 ± 0.5
Electron	isEM==loose	2.0 ± 0.8
Electron	isEM==medium	3.2 ± 1.0
Electron	isEM==tight	3.1 ± 1.4

Table 15: Overall shift of efficiencies for AtlFast-II with respect to full simulation, for both electrons and muons.

Since in the analysis a cut on p_{T} is applied, it is worth checking whether AtlFast-II reproduces the p_{T} resolution of a muon or electron correctly. The resolution in p_{T} for both muons and electrons is shown in Fig. 42.

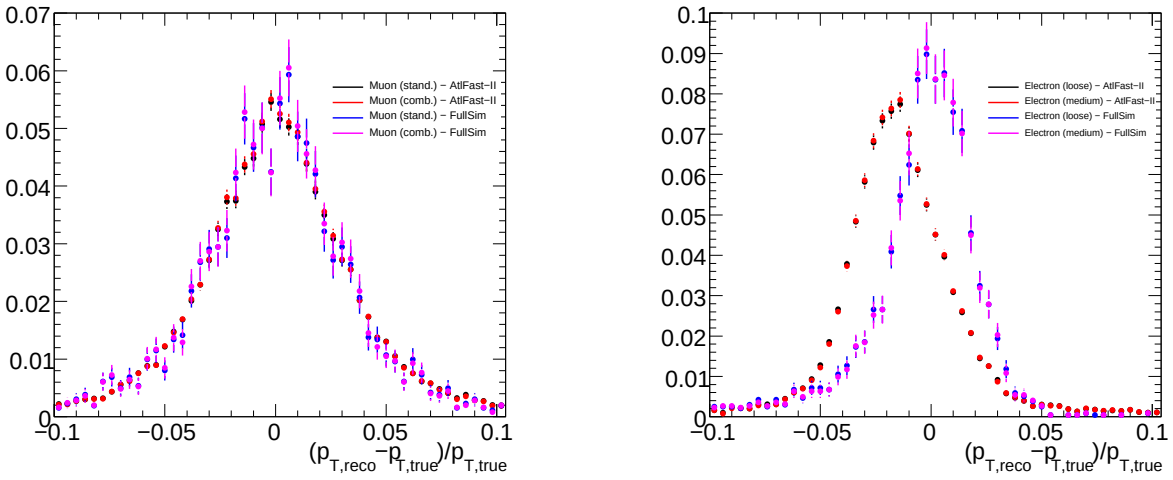


Figure 42: Lepton p_{T} resolution in AtlFast-II and in full simulation for muons (left) and electrons (right).

While the muon p_{T} variable is perfectly reproduced (as expected), the electron p_{T} shows a shift of around 2% to lower p_{T} values. This shift can be neglected in the present analysis, as will be shown in the evaluation of systematic uncertainties. The resolution is slightly worse in AtlFast-II with respect to full simulation (the calibration of the EM scale was obviously performed on full simulation). This, again, is expected to have no effect on the present analysis, but it will be taken into account.

In Table 16 and 17 the cut flow is shown, separately for the WH signatures for a muon or a lepton from the W boson, starting from the $E_{\mathrm{T}}^{\mathrm{miss}}$ cut.

10 Comparison with fast simulation

In the present study we rely on the AtlFast-II fast simulation of the ATLAS Detector. As discussed in Section 8, this is in fact the same as full simulation as far as tracking and specifically b-tagging are concerned, and in Sec 9 it was shown that the lepton efficiency agrees reasonably well between the two simulations. Here we will concentrate on how well AtlFast-II describes the

	WH(120)		WZ		$t\bar{t}(p_T^{min})$		Wt		W+$b\bar{b}$	
	e	μ	e	μ	e	μ	e	μ	e	μ
$E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV	71	74	470	526	59263	55783	7916	7809	579	596
p_T(W) > 200 GeV	71	74	470	526	59263	55783	7916	7809	579	596
p_T(e/μ)>30 GeV	71	74	470	526	59263	55783	7916	7809	579	596
p_T(additional μ)<10 GeV	71	74	462	480	54637	52194	7534	7458	567	586
p_T(additional e)<10 GeV	70	73	423	462	49582	47720	6997	6884	560	571
$\Delta\phi$(W,H)< $\frac{2}{3}\pi$	70	72	394	448	38568	46204	6282	6716	530	558
no additional b-jets p_T >15 GeV	65	66	371	418	15029	15577	3858	3946	450	475
add. jets on W side p_T <60 GeV	57	59	301	336	9642	9779	2924	2947	408	431
add. jets on H side p_T <60 GeV	51	52	249	277	6858	6983	2170	2200	367	394
one subjet b-tagged	45	46	60	66	4344	4294	1217	1204	317	336
both subjets b-tagged	23	22	21	23	295	281	78	83	111	113
loose fit cuts	23	22	21	22	290	275	76	81	108	111
112 GeV < mass(H) < 136 GeV	15.3	14.5	1.4	2.2	60	69	20	18	16	13

Table 16: The table shows the expected events going through the selection at each stage for the signal and the main backgrounds, starting from the $E_{\mathrm{T}}^{\mathrm{miss}}$ cut, separately for electrons and muons from the W boson. Numbers are projected to 30 fb^{-1} of collected data.

	ZH(120)		WW		ZZ		$t\bar{t}(p_T^{max})$		Z+jets		W+jets	
	e	μ	e	μ	e	μ	e	μ	e	μ	e	μ
$E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV	3.6	5.4	1693	1771	105	159	1545	1354	4407	5728	87155	89491
p_T(W) > 200 GeV	3.6	5.4	1693	1771	105	159	1545	1354	4407	5728	87155	89491
p_T(e/μ)>30 GeV	3.6	5.4	1693	1771	105	159	1545	1354	4407	5728	87155	89491
p_T(additional μ)<10 GeV	3.3	2.4	1682	1759	98	80	1351	1157	4209	2965	86961	89235
p_T(additional e)<10 GeV	1.5	2.1	1642	1719	49	72	1101	936	1991	2750	85648	87967
$\Delta\phi$(W,H)< $\frac{2}{3}\pi$	1.4	2.0	1571	1687	43	69	746	892	1411	2671	79967	86421
no additional b-jets p_T >15 GeV	1.2	1.8	1507	1622	38	64	249	246	1283	2468	76707	82722
add. jets on W side p_T <60 GeV	0.8	1.5	1220	1305	30	52	168	156	921	1825	58063	62430
add. jets on H side p_T <60 GeV	0.8	1.4	1004	1079	25	44	120	113	712	1434	45024	48258
one subjet b-tagged	0.7	1.3	132	138	5.5	11	73	61	67	140	3103	3252
both subjets b-tagged	0.35	0.6	2.2	3.2	1.5	3.8	5	3	6	6	67	52
112 GeV < mass(H) < 136 GeV	0.19	0.46	0	0.4	0	0.4	3	0	0	0.8	9	6

Table 17: The table shows the expected events going through the selection at each stage for the reference signal and the remaining backgrounds, starting from the $E_{\mathrm{T}}^{\mathrm{miss}}$ cut, separately for electrons and muons from the W boson. Numbers are projected to 30 fb^{-1} of collected data.

Higgs candidate efficiency, using the the WH signal. Table 18 shows the projected number of events in 30 fb^{-1} passing the single selection cuts.

The same cut flow is shown in Table 19 in the form of relative decrease of events in percent with respect to the previous cut.

In Table 20 the cut flow is shown again separately for selecting either an electron or a muon from the W boson candidate.

The overall signal efficiency in AtlFast-II is $(11 \pm 7)\%$ higher than in full simulation, after the full *loose* selection, except for the mass window cut, is applied. Starting from the events which passed the generator filter selection, the special jet clustering procedure outlined in Section 2 is applied, where exactly one physics object with $p_{\mathrm{T}} > 200$GeV splitting in a sufficiently symmetric way in two nearby lying jets is required to exist in the event. This yields the filtered mass distribution of Fig. 43. No lepton in the event is required yet.

Since no constraint is applied at generator level to the signal to decay the W boson in a lepton and a neutrino, a second mass peak is visible on the left of the Higgs boson mass peak, corresponding to a W boson decaying hadronically.

In AtlFast-II $5 \pm 1\%$ more mono-jets representing Higgs candidates are found. In particular AtlFast-II shows a slightly incresed right shoulder in the Higgs mass. In order to understand to what kinematic region for the Higgs candidate this corresponds to, Fig. 44 shows the filtered p_{T} distribution in case the Higgs mass is between 100 and 150 GeV, while Fig. 45 shows the distribution for the ΔR between the two highest p_{T} subjets and the asymmetry y parameter which triggers the splitting condition during jet clustering, in the same Higgs mass window.

	WH(120) AtlFast-II	WH(120) FullSim	Ratio
After filter cuts	1253 ± 8	1253 ± 12	1.00 +/- 0.01
1 Higgs candidate	569.7 ± 3.0	542.8 ± 4.6	1.05 +/- 0.01
filtered $p_{\mathrm{T}} > 200$ GeV	512.7 ± 3.2	490.2 ± 4.8	1.05 +/- 0.01
$E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV	362.4 ± 3.2	353.5 ± 4.8	1.03 +/- 0.02
p_{T}(W) > 200 GeV	171.0 ± 2.6	164.4 ± 3.8	1.04 +/- 0.03
p_{T}(e/μ)>30 GeV	145.6 ± 2.4	137.3 ± 3.6	1.06 +/- 0.03
p_{T}(additional μ)<10 GeV	144.6 ± 2.4	136.6 ± 3.6	1.06 +/- 0.03
p_{T}(additional e)<10 GeV	142.9 ± 2.4	134.1 ± 3.5	1.07 +/- 0.03
$\Delta\phi$(W,H)< $\frac{2}{3}\pi$	142.2 ± 2.4	133.1 ± 3.5	1.07 +/- 0.03
no additional b-jets p_{T} >15 GeV	130.6 ± 2.3	119.3 ± 3.4	1.09 +/- 0.04
jets on W side p_{T} <60 GeV	115.7 ± 2.2	107.0 ± 3.2	1.08 +/- 0.04
jets on H side p_{T} <60 GeV	102.7 ± 2.1	95.1 ± 3.0	1.08 +/- 0.04
one subjet b-tagged	91.4 ± 2.0	85.9 ± 2.9	1.06 +/- 0.04
both subjets b-tagged	45.6 ± 1.4	41.2 ± 2.1	1.11 +/- 0.07
loose fit cuts	45.4 ± 1.4	41.0 ± 2.1	1.11 +/- 0.07
112 GeV < mass(H) < 136 GeV	29.8 ± 1.2	27.0 ± 1.7	1.10 +/- 0.08

Table 18: The table shows the expected events going through the selection at each stage for the signal produced through AtlFast-II compared to full simulation. Numbers are projected to 30 fb^{-1} of collected data.

	WH(120) AtlFast-II	WH(120) FullSim
1 Higgs candidate	45 ± 1	43 ± 2
filtered $p_{\mathrm{T}} > 200$ GeV	90 ± 1	90 ± 1
$E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV	71 ± 2	72 ± 2
p_{T}(W) > 200 GeV	47 ± 3	47 ± 3
p_{T}(e/μ)>30 GeV	85 ± 3	84 ± 3
p_{T}(additional μ)<10 GeV	99.3 ± 0.7	99.5 ± 0.6
p_{T}(additional e)<10 GeV	98.8 ± 0.9	98 ± 1
$\Delta\phi$(W,H)< $\frac{2}{3}\pi$	99.5 ± 0.6	99.3 ± 0.7
no additional b-jets p_{T} >15 GeV	92 ± 2	90 ± 3
jets on W side p_{T} <60 GeV	89 ± 3	90 ± 3
jets on H side p_{T} <60 GeV	89 ± 3	89 ± 3
one subjet b-tagged	89 ± 3	90 ± 3
both subjets b-tagged	50 ± 5	48 ± 5
loose fit cuts	99.7 ± 0.8	99 ± 1
112 GeV < mass(H) < 136 GeV	66 ± 7	66 ± 7

Table 19: The table shows the expected number of events going through the selection at each stage for the signal produced through AtlFast-II compared to full simulation.

Again, no requirement on the W boson side is applied yet. The difference is mainly in candidates near to the 200 GeV p_{T} threshold, corresponding to a larger aperture between the two Higgs subjets ΔR and a more symmetric configuration of the two subjets, are reconstructed.

The $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution, normalized to the number of entries in the histogram, is shown in Fig. 46, after applying the Higgs candidate selection: for AtlFast-II it has more entries at lower $E_{\mathrm{T}}^{\mathrm{miss}}$ values, most probably because a lower p_{T} Higgs boson corresponds to a lower p_{T} back to back W boson.

Fig. 47 shows the p_{T} distribution for the p_{T} of the hardest lepton, for electrons (on the left) and for muons (on the right), again straight after the Higgs selection. Since these distributions are again normalized to the number of their respective entries, they hide the effect of the different

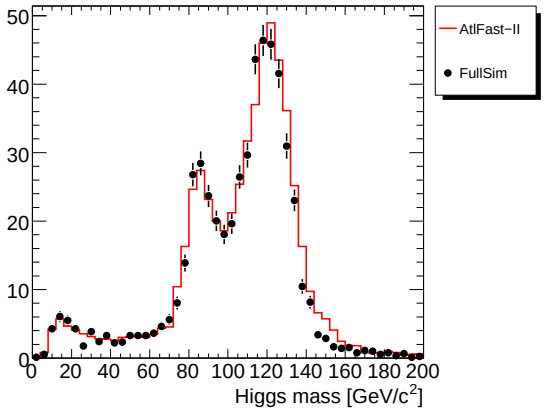


Figure 43: Comparison of filtered mass of Higgs candidate for the WH signal produced with AtlFast-II and full simulation. Distributions are normalized to the process cross sections.

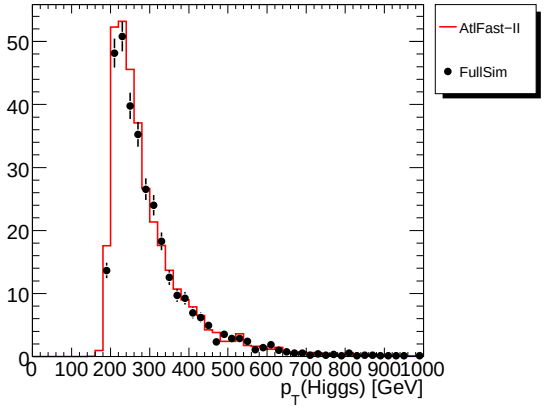


Figure 44: Comparison of filtered p_{T} of Higgs candidate for the WH signal produced with AtlFast-II and full simulation. Distributions are normalized to the process cross sections.

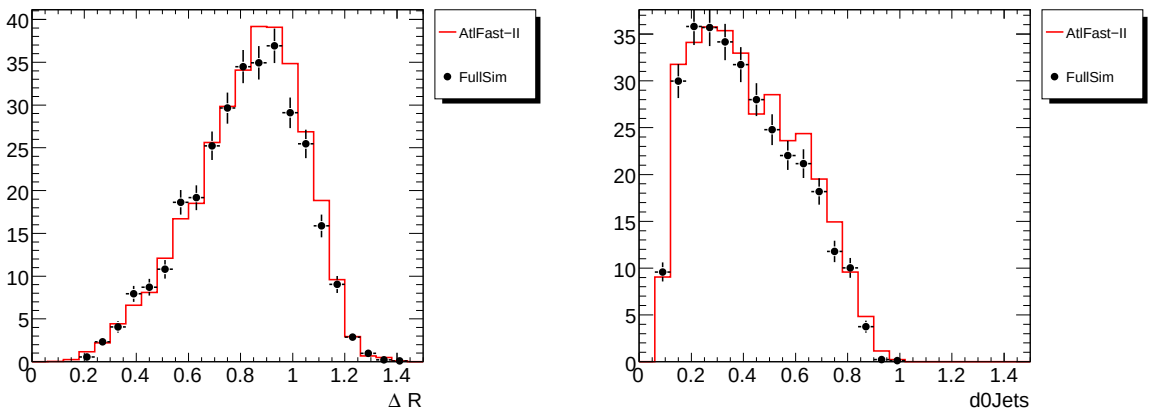


Figure 45: Distributions for ΔR (left) and d_0 (y_{clust}) between the two highest p_{T} subjets of the Higgs candidate. Distributions are normalized to the process cross sections.

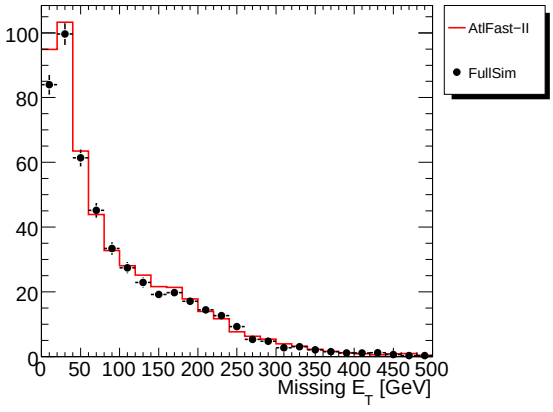


Figure 46: Comparison of $E_{\mathrm{T}}^{\mathrm{miss}}$ in the event for the WH signal sample produced with AtlFast-II and the corresponding one produced with full simulation. Distributions are normalized to the number of entries.

	WH(120) AtlFast-II		WH(120) FullSim	
	e	μ	e	μ
1 Higgs candidate	71	74	67	70
filtered $p_{\mathrm{T}} > 200$ GeV	71	74	67	70
$E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV	71	74	67	70
p_{T}(W) > 200 GeV	71	74	67	70
p_{T}(e/μ)>30 GeV	71	74	67	70
p_{T}(additional μ)<10 GeV	71	74	66	70
p_{T}(additional e)<10 GeV	70	73	65	69
$\Delta\phi$(W,H)< $\frac{2}{3}\pi$	70	72	64	69
no additional b-jets p_{T} >15 GeV	65	66	57	62
add. jets on W side p_{T} <60 GeV	57	59	51	56
add. jets on H side p_{T} <60 GeV	51	52	46	49
one subjet b-tagged	45	46	42	44
both subjets b-tagged	23	22	21	21
loose fit cuts	23	22	20	20
112 GeV < mass(H) < 136 GeV	15.3	14.5	14	13

Table 20: The table shows the expected events going through the selection at each stage for the signal and the main backgrounds, separately for selecting a muon or an electron from the W boson candidate. Numbers are projected to 30 fb^{-1} of collected data.

efficiency for the two different lepton families: at the point where the cut on the p_{T} of the leptons is applied, in AtlFast-II $(49.2 \pm 0.9)\%$ of the events contains an electrons, the remaining ones a muon, while in FullSim $(48.7 \pm 1.4)\%$ of the events contains an electron, the remaining ones a muon. So from applying first the Higgs candidate selection, the statistics is not enough to be able to compare the electron and muon efficiencies, this is why a more general study on electron and muon efficiencies for the WH signal samples was reported in Section 9: according to it, we expect the muon efficiency to be comparable in AtlFast-II and in full simulation, while the electron efficiency (medium electron ID requirement) is expected to be higher in AtlFast-II by a factor $3.2 \pm 1.0\%$.

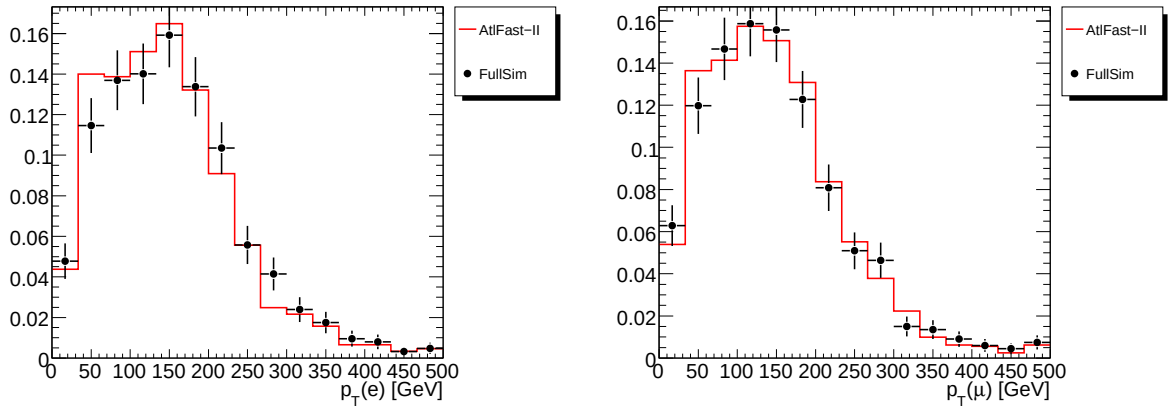


Figure 47: Distributions for p_{T} of the hardest lepton in the events, for electrons (left) and for muons (right), after the Higgs selection cuts have already been applied. Distributions are normalized to the number of entries.

Fig. 48 shows the $\Delta\phi$(Higgs,W) variable, after both Higgs and W boson selections are ap-

plied. The two distributions agree well in the limit of the available statistics. Fig.49 shows the comparison for the p_{T} of additional b-jets in the event.

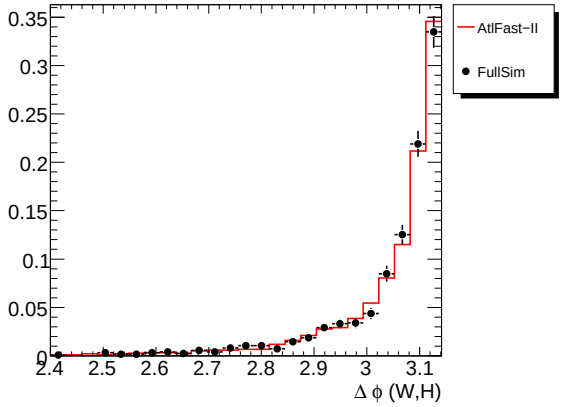


Figure 48: Comparison of $\Delta\phi$(Higgs,W) in AtlFast-II and full simulation, after the Higgs and W boson selection cuts have been applied. Distributions are normalized to the number of entries.

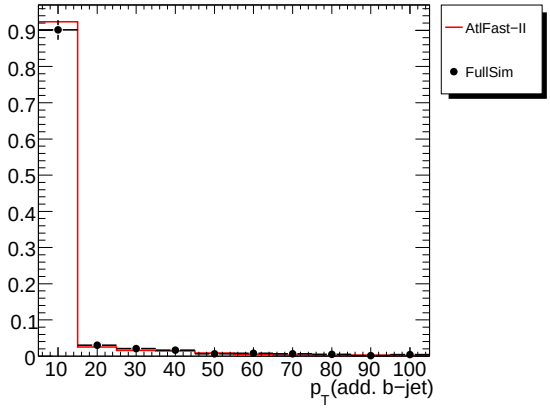


Figure 49: Comparison of p_{T}(add b-jet) variable in AtlFast-II and full simulation, after all previous selection cuts have been applied. Distributions are normalized to the number of entries.

Fig.50 shows the comparison for the p_{T}(additional jet) distribution, where the additional jet is on the side of the W boson (left) or on the side of the H boson (right). All selection cuts before the veto on the additional jet on the W boson side are applied for both distributions. While on the W boson side everything seems to be very compatible, on the Higgs boson side the p_{T} spectrum of additional jets in the events looks a slightly softer, but only at $p_{\mathrm{T}} < 60$GeV (so it does not affect our *loose* selection).

Fig. 51 shows the comparison for the b-weight distribution, after initial Higgs and W boson selection and after the $\Delta\phi > \frac{2}{3}$ requirement. The jet vetoes are not applied in order to increase the available statistics. The distributions agree very well, as is expected from the use of the full simulation of the Inner Detector.

Fig. 52 shows the $b-$weight distribution of the third subjet, after initial Higgs and W boson

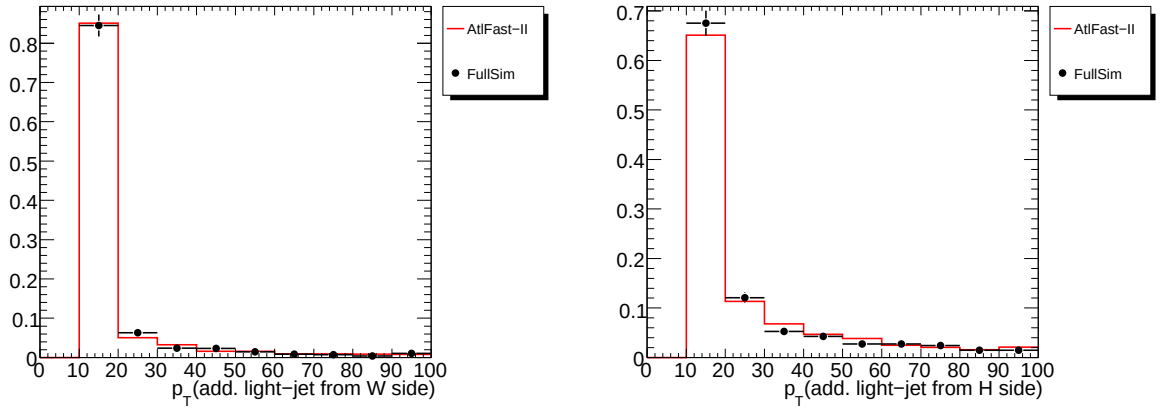


Figure 50: Distributions for p_{T} of the hardest additional jet on the W boson side (left) and on the Higgs boson side (right) for AtlFast-II and full simulation. Distributions are normalized to the number of entries.

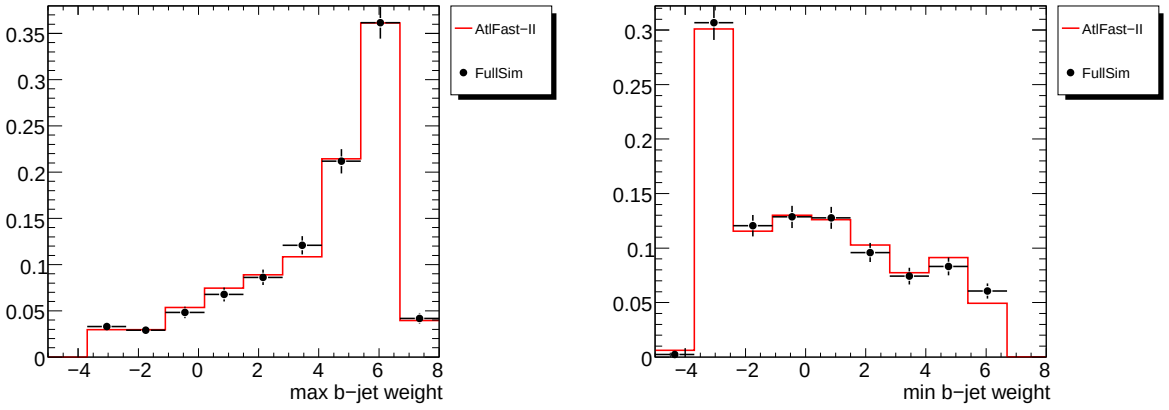


Figure 51: Distributions for maximum and minimum $b-$weight of the two subjets, for both AtlFast-II and full simulation. Distributions are normalized to the number of entries.

selection, after the $\Delta\phi > \frac{2}{3}$ requirement and after the b-tag requirement on the first two subjets.

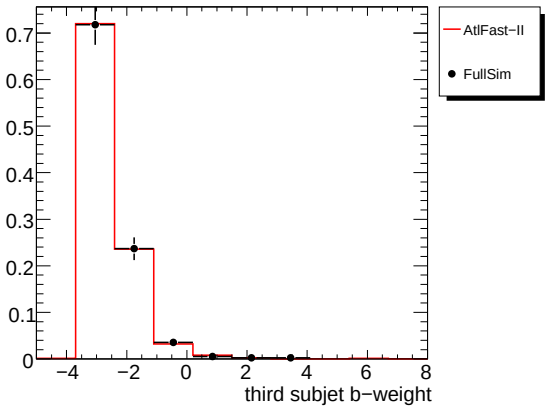


Figure 52: Comparison of the $b-$weight of the third subjet in AtlFast-II and in full simulation. Distributions are normalized to the number of entries.

Finally Fig. 53 shows the filtered mass distribution just after the initial W and H boson selection and the $\Delta\phi$(W,H) cut (left) and after all selection cuts of the *loose* selection. The expected number of events in 30 fb^{-1} is shown.

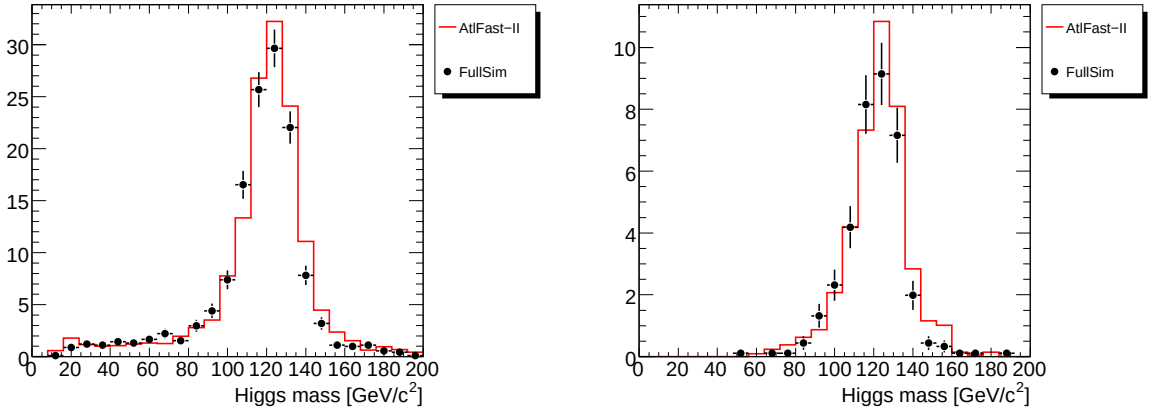


Figure 53: Distributions for the filtered mass before b-Tagging and jet veto cuts (left) and afterwards (rights). Distributions are normalized to their respective cross sections and to 30 fb^{-1} of collected data.

In order to compare the shape of the invariant mass distribution in AtlFast-II and full simulation, the final mass distribution is also shown normalized to the number of entries in Fig. 54. In addition a Gauss fit is performed iteratively, until the $\pm 1.5\sigma$ region around the fitted mean corresponds to the interval chosen for the fit (for the mean μ(AtlFast-II)$= 123.0 \pm 0.4$, μ(FullSim)$= 121.5 \pm 0.7$). Apart from the small shift to higher values, the shape in AtlFast-II is reproduced very well and the resolution is comparable.

In the limit of the available statistics of the samples used for the present comparison, the overall analysis efficiency for signal events produced with AtlFast-II agrees reasonably well with

full simulation. The small discrepancy between AtlFast-II and full simulation seen in the earlier steps of the cut flow can be explained by a slightly reduced mono-jet reconstruction efficiency (-5%), plus a small decrease in the electron efficiency by -3%.

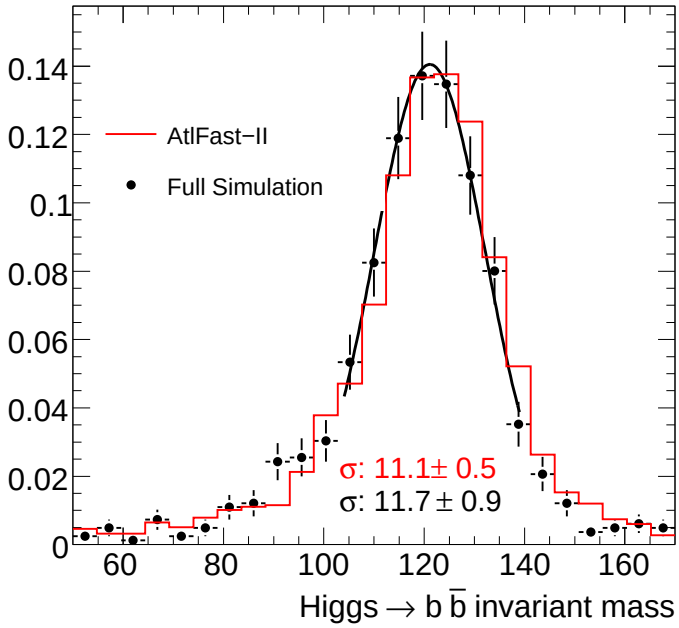


Figure 54: Comparison between final mass distribution for signal as produced with AtlFast-II compared to full simulation.

11 Combination of channels

The results of the analyses must be combined in order to evaluate an overall significance. In this chapter a profile likelihood method is applied. The procedure used here is heavily derived from that used in the Higgs chapter of Ref. [2].

Assuming simple event counting experiments, the likelihood function for a given experiment i where n events are observed is Poisson.

$$L_i(\mu) = \frac{(\mu s_i + b_i)^{n_i}}{n_i!} e^{-(\mu s_i + b_i)} \tag{4}$$

Where b and s represent expected background and signal levels respectively. The symbol μ parameterises the level of signal present, $\mu = 1$ representing the standard model case.

Since we are combining several analyses, the form of a combined likelihood function is required. The likelihood function for N experiments of the type described above can be expressed as:

$$L(\mu) = \prod_{i=1}^{N} L_i(\mu) \tag{5}$$

We also need to include the effects of the systematic uncertainty introduced by our imperfect understanding of the backgrounds. The backgrounds of the separate channels are neither fully

correllated nor completely independent. To reflect this, we break the backgrounds into 3 groups, which are independent of each other but completely correllated between the separate channels. In a full analysis, the background would be estimated via a combination of several separate techniques, which is beyond the scope of this note. The 3 groups are:.

- $t\bar{t}$, WW, WZ and ZZ
- $W + jets$ and Wt
- $Z + jets$

We then assume a Gaussian uncertainty on the level of each background.

In a complete treatment there would be many partially correlated uncertainties on the background level due to factors such as luminosity, cross-section and detector efficiency. Splitting the background into 3 completely uncorrelated uncertainties should provide enough degrees of freedom to give a rough but somewhat realistic estimate of how this uncertainty affects the final significance.

This approach requires separation of the background component of the likelihood function into three separate contributions and the addition of some terms.

$$\begin{aligned} L(\mu, C_t, C_w, C_z) &= \prod_{i=1}^{N} \frac{(\mu s_i + C_t t_i + C_w w_i + C_z z_i)^{n_i}}{n_i!} e^{-(\mu s_i + C_t t_i + C_w w_i + C_z z_i)} \times \\ & \quad Gaus(C_t) \times Gaus(C_w) \times Gaus(C_z) \end{aligned}$$

$$Gaus(C_x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-\frac{(C_x - 1)^2}{2\sigma_x^2}}$$

The new constants C_x allow the different groups of background to vary in a manner regulated by the Gaussian terms. We can now proceed to calculate a significance. To do this we first define the likelihood ratio and for convenience an associated test statistic.

$$\lambda(\mu) = \frac{L(\mu, \hat{C}_t, \hat{C}_v)}{L(\hat{\mu}, \hat{\hat{C}}_t, \hat{\hat{C}}_v)}$$

$$q(\mu) = -2 \ln(\lambda(\mu))$$

The variables $\hat{C}_t$ and $\hat{C}_v$ maximise L for the given μ, whereas $\hat{\mu}$, $\hat{\hat{C}}_t$ and $\hat{\hat{C}}_v$ maximise L overall.

In order to calculate a discovery significance, we attempt to reject the background-only case, where the true $\mu = 0$. Given this we also do not allow $\hat{\mu} < 0$; in such cases, we fix $\hat{\mu} = 0$. This is because although experiments where the observation is less than the expected background-only count may not look consistent with the background prediction, they certainly do not imply the presence of a Higgs signal. First we define a function which represents the distribution of possible outcomes of experiments where only background is present.

$$f(x|\mu_{true}) = P(q(0) = x|\mu_{true} = 0) \tag{6}$$

The p-value associated with the outcome of an actual experiment can then be defined as the fraction of experiments at least as unlikely as this one:

$$p = \int_{q(0)}^{\infty} f(x|0)dx \tag{7}$$

Which can be associated to a significance using the Gaussian integral.

$$Z = \Phi^{-1}(1-p) \tag{8}$$

The biggest computational challenge is determining the form of f with sufficient accuracy. The function could be determined via Monte Carlo but given that a significance of 5σ corresponds to a p-value $O(10^{-7})$, the number of pseudo-experiments required is $> O(10^8)$. As described in Ref. [2], the form of f when $\mu > 0$ can be approximated by a chi-square distribution with one degree of freedom. However, for experiments where $\mu < 0$ we are setting $\mu = 0$, so these will always give $q(0) = 0$. Therefore the form of f can be described as follows, the only unknown parameter being w, the fraction of experiments where $\mu > 0$:

$$f(x|0) = w f_{\chi^2_1}(x) + (1-w)\delta(x) \tag{9}$$

There are many possible ways to determine the value of w. A simple one used here is to perform a smaller number of Monte Carlo pseudo-experiments ($O(10^6)$). Each pseudo-experiment consists of generating the constants C_x according to their Gaussian uncertainty. These constants are used to modify the background sample expectation values, which are then in turn used to generate Poisson random event counts. In the single fits, the nominal values for the constants C_x are used, so that their difference with respect to the expectation values used for generation correctly reflect their expected systematic uncertainty. The test statistic $q(0)$ is then calculated for each pseudo-experiment and used to fill a histogram.

If the lowest bin in this histogram starts at $x = 0$ and ends at a small number $x = a$, the contents of this bin are the integral:

$$\int_0^a f(x|0)dx = N_{exp} \times \left(w \int_0^a f_{\chi^2_1}(x)dx + (1-w) \right) \tag{10}$$

This equation is trivially soluble for w since the integral of the chi-square distribution is calculable. Typically the statistical uncertainty on this value of w is small since approximately half of all Monte Carlo experiments have values of $\mu < 0$. This value of w can then be used to test the validity of the chi-square approximation of f for experiments where $\mu > 0$. In this analysis, the agreement appears to be very good, as seen in Fig. 55.

With the form of f determined sufficiently well, many experiments containing signal ($\mu_{true} = 1$) can be generated and their significance calculated. Out of this distribution of possible outcomes, the median is taken as a reasonable expectation of performance. We are using the combination for our three channels: $ll b\bar{b}$, $l\nu b\bar{b}$ and $\nu\nu b\bar{b}$. Signal and background expectations are taken from the analyses performed, the values for which are given in Table 21. Initially, a perfect understanding of the backgrounds is assumed and the results are as seen in Fig. 56. Here a median significance of 3.7σ is observed, which is consistent with what would be expected from adding the $S/\sqrt{(B)}$ significances in quadrature. A variety of possible background uncertainties are then tested, the results of which can be found in Table 22.

The exact background uncertainties for each channel depend on the mixture of different background samples. A 10% uncertainty in each of the three background samples corresponds to a 9%, 7% and 6% uncertainty on the total background level in the $ll b\bar{b}$, $l\nu b\bar{b}$ and $\nu\nu b\bar{b}$ channels respectively. These numbers scale linearly with sample uncertainty.

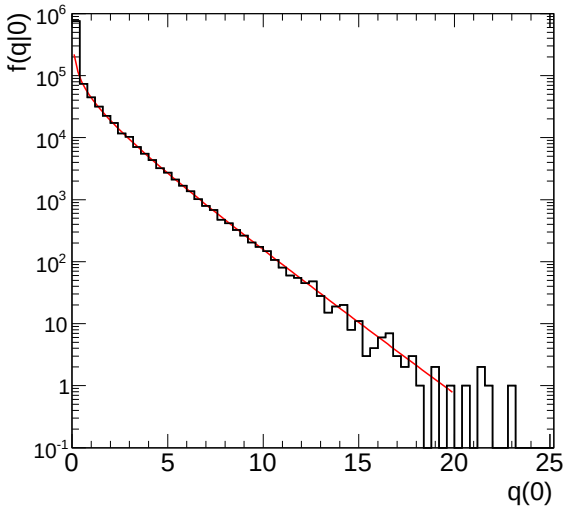


Figure 55: $q(0)$ for background-only experiments, the histogram shows Monte Carlo in excellent agreement with the solid line representing a one degree of freedom chi-squared distribution

Channel	s_i	t_i	w_i	z_i	$S/\sqrt{B}$
$ll b\bar{b}$	5.34	0.98	0.0	11.2	1.5
$l\nu b\bar{b}$	13.5	7.02	12.5	0.78	3.0
$\nu\nu b\bar{b}$	16.3	45.2	27.4	31.6	1.6
Combined					3.7

Table 21: Experiment expectations used for combination.

One method of determining the background event rates experimentally is with a fit. This approach is explored in Ref. [28], which implies that using this method alone, at 30fb^{-1}an uncertainty of 15% per background sample should be achievable. Given that after 30fb^{-1}of data ATLAS should also have strong understanding of the background from other sources we estimate that an uncertainty of 10% or better is realistic.

The significance of 3.7σ in the perfect case is found to be reduced to 3.2σ in the case of a 10% uncertainty and 3.0σ in the case of a 15% uncertainty on the expected level of each background sample. It is useful to note that the combination of the three separate channels with differing background compositions protects against the effects of this systematic uncertainty.

12 Summary and Outlook

We have presented a first study of the ATLAS sensitivity to the HZ and HW associated production channels at high-p_{T} for a low-mass Standard Model Higgs boson using a realistic detector simulation. The analysis closely follows that of Ref. [3], but makes use of the full GEANT-based ATLAS simulation for the tracking, vertexing and b-tagging, uses the ATLFAST-II simulation of the calorimeter, and the full ATLAS reconstruction framework throughout. The sensitivity to the differences between ATLFAST-II and the full simulation is evaluated and the impact on the signal efficiency in the $l\nu b\bar{b}$ channel found to be within 7%. The trigger efficiencies

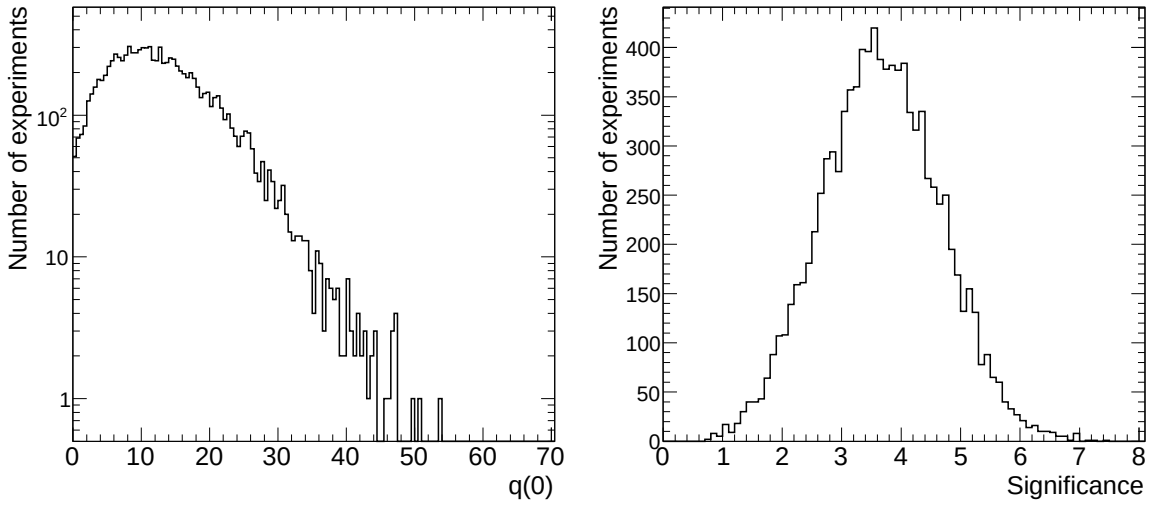


Figure 56: $q(0)$ and significance for a range of possible signal experiments.

σ_t	σ_w	σ_z	Significance
Perfect	Perfect	Perfect	3.7
5%	5%	5%	3.5
10%	10%	10%	3.2
15%	15%	15%	3.0
20%	20%	20%	2.8
30%	30%	30%	2.5
50%	50%	50%	2.2
75%	75%	50%	2.0
50%	10%	10%	2.8

Table 22: Significances for different scenarios with differing background uncertainties.

are found to be very high for all channels considered. All the expected significant backgrounds are studied, including that from Wt, which was not considered in Ref. [3]. Sensitivities are given as a function of the eventual systematic uncertainty in the background.

For the cut-based analysis presented here, the combined sensitivity of these channels after an integrated luminosity $L = 30$ fb^{-1} of data, considering only statistical errors, is 3.7σ for a Higgs-boson mass of 120 GeV. If the major backgrounds have a systematic uncertainty of around 15%, this drops to 3.0σ, and if the systematic uncertainty is as high as 50% this sensitivity drops to around 2.2σ. For $L = 10$ fb^{-1} we might expect sensitivity of up to 2.1σ. This is comparable to the ATLAS sensitivity in any other single channel [2] in this region. In addition, this channel would give the best information on the $H \to b\bar{b}$ coupling and will therefore be critical for determining the parameters of the Higgs sector [29].

All numbers are based on signal and background processes generated at LO and normalized to their respective LO cross sections. While a first evaluation of the impact of the NLO estimate of their cross sections was done in Ref. [3] at parton level and the impact on the significances found to be small, the use of dedicated NLO generators, whenever possible, is foreseen for a future update of the analysis.

Further improvements can be expected in this analysis. The b-tagging might be calibrated and further optimized for this specific kinematic region, aiming at a higher b-tagging efficiency while preserving a similar light and charm-jet rejection. The jet calibrations could be redone for these specific jets, hopefully improving the mass resolution. Perhaps most importantly, the background can be extracted directly from the data, and studies in this direction are presented in Ref. [28]. Finally, the use of sophisticated multivariate techniques similar to those currently being applied at the Tevatron should enhance the sensitivity compared to the relatively simple cut-based analysis presented here.

The simulation of pile-up and of cavern background was not considered in the present study: the potential impact of pile-up on the dijet mass resolution, on the jet veto efficiency and on the b-tagging efficiency will be evaluated in the near future.

We note that many Supersymmetric models and other extensions of the Standard Model may lead to significantly enhanced cross sections for $H \to b\bar{b}$, increasing the importance of studying this channel in early data.

In conclusion, we have confirmed with a realistic detector simulation that by studying the high-p_{T} regions and employing state-of-the-art subjet techniques, the HZ and HW channels can be reinstated as one of the promising search and measurement channels for the low-mass Standard Model Higgs at ATLAS.

13 Acknowledgments

Thanks to Glen Cowan, Michael Duehrssen, Borut Kersevan and HSG5 for useful discussions.

A Sample Production

This section describes the production and subsequent validation of event samples used in this note. In general, the samples used in this analysis were initially produced privately, but were validated against equivalent centrally produced datasets, and are therefore treated as extensions of these centrally-produced datasets. The list of such samples, with dataset names is provided in Table 23 and Table 24.

The only samples which have not been validated in this way are the AcerMC $Wb\bar{b}$ events. These events however have not been used in any results submitted for further publication. All

events submitted for further publication beyond this note have been validated and are therefore included in the above pair of tables.

A.1 Validation of Event Generation

Validation of the event generation step was performed by comparing EVGEN files from the centrally and privately produced datasets. Various variables relating to different classes of objects were extracted and their distributions compared both visually and statistically [30]. These included p_{T} and rapidity of jets, electrons, muons and b-quarks. The only significant differences found are in the Wt samples, where it was found that due to different versions of Pythia the centrally produced samples contain somewhat more QCD radiation. However, the differences will not have a large effect on the overall significance, and are understandable in terms of the Pythia version change and are reproducible. In addition the privately produced samples used in the analysis are the more conservative of the two, since the extra radiation in the centrally-produced samples would lead to more background events failing the extra-jet veto. Based on this comparison it was concluded that there are no significant issues requiring regeneration of any of the samples.

The job options for these samples are included in the ATLAS software distribution from release 15.3.0 onwards.

A.2 Validation of AtlFast-II Simulation

The premise here was not to validate AtlFast-II itself (since dedicated studies, also using release 13 have already been performed). The aim was to validate the specific use of AtlFast-II for this production, checking that it had been used exactly as it would have been in central production.

Based on discussion with both the validation team [31] and the AtlFast-II development team [32] it was agreed to validate the private production against the last official validation samples produced with release 13.0.40.5. This was the last validated version of AtlFast-II under release 13 and these samples represent the last validated simulation, digitization and reconstruction under this release also.

In order to perform the validation a validation sample was chosen (in this case dataset 005200). The EVGEN files for this validation sample were taken and simulated privately using exactly the same configuration as the samples used in the analysis. The resulting 50,000 events were then compared against the centrally produced validation sample [33]. The two samples were found to be perfectly compatible, showing that the AtlFast-II production had indeed been done using the same settings as it would have been under central production.

References

[1] T. Gadfort and S. Griso, presented at Moriond QCD and High Energy Interactions March 14th - March 21st 2009.

[2] The ATLAS Collaboration, G. Aad *et al.*, arXiv:0901.0512.

[3] J. M. Butterworth, A. R. Davison, M. Rubin and G. P. Salam, Phys. Rev. Lett. **100**, 242001 (2008), [0802.2470].

[4] Y. L. Dokshitzer, G. D. Leder, S. Moretti and B. R. Webber, JHEP **08**, 001 (1997), [hep-ph/9707323].

[5] M. Wobisch and T. Wengler, hep-ph/9907280.

[6] S. Catani, Y. L. Dokshitzer, M. H. Seymour and B. R. Webber, Nucl. Phys. **B406**, 187 (1993).

[7] S. D. Ellis and D. E. Soper, Phys. Rev. **D48**, 3160 (1993), [hep-ph/9305266].

[8] M. Cacciari and G. P. Salam, Phys. Lett. **B641**, 57 (2006), [hep-ph/0512210].

[9] H1 Collaboration, I. Abt *et al.*, Nucl. Instrum. Meth. **A386**, 348 (1997).

[10] J. M. Butterworth, B. E. Cox and J. R. Forshaw, Phys. Rev. D **65** (2002), [hep-ph/0201098].

[11] J. M. Butterworth, A. Davison, E. Ozcan and P. Sherwood, ATLAS-PHYS-INT-2007-015.

[12] A. H. Mueller, Phys. Lett. **B104**, 161 (1981).

[13] B. I. Ermolaev and V. S. Fadin, JETP Lett. **33**, 269 (1981).

[14] A. Bassetto, M. Ciafaloni and G. Marchesini, Phys. Rept. **100**, 201 (1983).

[15] G. Piacquadio and C. Weiser, J. Phys. Conf. Ser. **119**, 032032 (2008).

[16] G. Corcella *et al.*, hep-ph/0210213.
G. Corcella *et al.*, JHEP **01**, 010 (2001), [hep-ph/0011363].

[17] J. M. Butterworth, J. R. Forshaw and M. H. Seymour, Z. Phys. **C72**, 637 (1996), [hep-ph/9601371].

[18] B. P. Kersevan and E. Richter-Was, hep-ph/0405247.

[19] T. Sjostrand, S. Mrenna and P. Skands, JHEP **05**, 026 (2006), [hep-ph/0603175].

[20] J. Pumplin *et al.*, JHEP **07**, 012 (2002), [hep-ph/0201195].

[21] CDF, F. Abe *et al.*, Phys. Rev. **D54**, 4221 (1996), [hep-ex/9605004].

[22] ZEUS, J. Breitweg *et al.*, Phys. Lett. **B443**, 394 (1998), [hep-ex/9810046].

[23] R.-W. Elzbieta, ATL-PHYS-2003-018.

[24] R. K. Ellis and S. Veseli, Phys. Rev. **D60**, 011501 (1999), [hep-ph/9810489].

[25] F. Febres Cordero, L. Reina and D. Wackeroth, Phys. Rev. **D78**, 074014 (2008), [0806.0808].

[26] J. Campbell, R. K. Ellis and D. L. Rainwater, Phys. Rev. **D68**, 094021 (2003), [hep-ph/0308195].

[27] M. Duerhssen, ATLAS-PHYS-INT-2008-043.

[28] G. Piacquadio, ATLAS Note ATL-COM-PHYS-2009-308.

[29] R. Lafaye, T. Plehn, M. Rauch, D. Zerwas and M. Duehrssen, arXiv:0904.3866 [hep-ph].

[30] Website, EVGEN Validation Plots `http://www.hep.ucl.ac.uk/~adamd/validation`.

[31] I. Vivarelli, private communication.

[32] M. Duehrssen, private communication.

[33] Website, AtlFast-II Validation Plots http://wwwhep.physik.uni-freiburg.de/~giacinto/plotsAtlfastIIValidation.

	WH with lepton filter
Central EVGEN	mc08.209140.WH120lnbb_Herwig.evgen.EVNT.e419_tid067460
Private EVGEN	user08.NicolaGiacintoPiacquadio.ganga.datafiles.WHhighPtStudy.130405.009140.CSC.009140.WHlnbb_Herwig.evgen.EVNT.v13000405
Central AOD	mc08.209140.WH120lnbb_Herwig.recon.AOD.e419_a84/
Private AOD	user08.NicolaGiacintoPiacquadio.ganga.datafiles.WH120highPtStudy.13.0.40.5.AtlFastII.009140.CSC.009140.WHlnbb_Herwig.recon.AOD.v13000405
	ZH with lepton filter
Central EVGEN	mc08.209138.ZH120llbb_Herwig.evgen.EVNT.e419_tid067458
Private EVGEN	user08.NicolaGiacintoPiacquadio.ganga.datafiles.WHhighPtStudy.130405.009138.CSC.009138.ZHllbb_Herwig.sub2.evgen.EVNT.v13000405
Central AOD	mc08.209138.ZH120llbb_Herwig.recon.AOD.e419_a84/
Private AOD	user08.NicolaGiacintoPiacquadio.ganga.datafiles.WHhighPtStudy.AtlfastII.130405.009138.CSC.009138.ZHllbb_Herwig.recon.AOD.v13000405
	W+jets with lepton filter
Central EVGEN	mc08.209718.Jimmy_Wj_1Lepton20_ptMin150.evgen.EVNT.e419_tid067448
Private EVGEN	users.NicolaGiacintoPiacquadio.ganga.datafiles.WHhighPtStudy.009995.Wj_PtMin150_LeptonFilter15_Herwig.0.evgen.EVNT.v13000405
Central AOD	mc08.209718.Jimmy_Wj_1Lepton20_ptMin150.recon.AOD.e419_a84/
Private AOD	users.NicolaGiacintoPiacquadio.ganga.datafiles.WH120highPtStudy.009995.Wj_PtMin150_LeptonFilter15_Herwig.0.recon.AOD.v13000405
	Z+jets with lepton filter
Central EVGEN	mc08.209197.Jimmy_Zj_1Lepton15_ptMin150.evgen.EVNT.e419_tid067465
Private EVGEN	users.NicolaGiacintoPiacquadio.ganga.datafiles.WHhighPtStudy.009996.Zj_PtMin150_LeptonFilter15_Herwig.1.evgen.EVNT.v13000405 users.NicolaGiacintoPiacquadio.ganga.datafiles.WHhighPtStudy.009996.Zj_PtMin150_LeptonFilter15_Herwig.1.sub2.evgen.EVNT.v13000405
Central AOD	mc08.209197.Jimmy_Zj_1Lepton15_ptMin150.recon.AOD.e419_a84/
Private AOD	users.NicolaGiacintoPiacquadio.ganga.datafiles.WH120highPtStudy.009996.Zj_PtMin150_LeptonFilter15_Herwig.1.recon.AOD.v13000405
	WW with lepton filter
Central EVGEN	mc08.209193.Jimmy_WW_1Lepton15_ptMin150.evgen.EVNT.e419_tid067461
Private EVGEN	users.NicolaGiacintoPiacquadio.ganga.datafiles.WHhighPtStudy.009994.WW_PtMin150_LeptonFilter15_Herwig.evgen.EVNT.v13000405
Central AOD	mc08.209193.Jimmy_WW_1Lepton15_ptMin150.recon.AOD.e419_a84/
Private AOD	users.NicolaGiacintoPiacquadio.ganga.datafiles.WH120highPtStudy.009994.WHhighPtStudy.009994.WW_PtMin150_LeptonFilter15_Herwig.recon.AOD.v13000405
	WZ with lepton filter
Central EVGEN	mc08.209195.Jimmy_WZ_1Lepton15_ptMin150.evgen.EVNT.e419_tid067463
Private EVGEN	users.NicolaGiacintoPiacquadio.ganga.datafiles.WHhighPtStudy.009997.WZ_PtMin150_LeptonFilter15_Herwig.1.evgen.EVNT.v13000405
Central AOD	mc08.209195.Jimmy_WZ_1Lepton15_ptMin150.recon.AOD.e419_a84/
Private AOD	users.NicolaGiacintoPiacquadio.ganga.datafiles.WH120highPtStudy.2.009997.WZ_PtMin150_LeptonFilter15_Herwig.3.recon.AOD.v13000405
	ZZ with lepton filter
Central EVGEN	mc08.209194.Jimmy_ZZ_1Lepton15_ptMin150.evgen.EVNT.e419_tid067462
Private EVGEN	users.NicolaGiacintoPiacquadio.ganga.datafiles.WHhighPtStudy.009990.realZZ_PtMin150_LeptonFilter15_Herwig.evgen.EVNT.v13000405
Central AOD	mc08.209194.Jimmy_ZZ_1Lepton15_ptMin150.recon.AOD.e419_a84/
Private AOD	users.NicolaGiacintoPiacquadio.ganga.datafiles.WH120highPtStudy.009990.ZZ_PtMin150_LeptonFilter15_Herwig.recon.AOD.v13000405
	ttbar with lepton filter (ptmin=150GeV)
Central EVGEN	mc08.209198.Jimmy_ttbar_1Lepton20_ptMin150.evgen.EVNT.e419_tid067466
Private EVGEN	users.NicolaGiacintoPiacquadio.ganga.datafiles.WH120highPtStudy.009995.ttbarLO_PtMin150_LeptonFilter20_Herwig.1.recon.ESD.v13000405
Central AOD	mc08.209198.Jimmy_ttbar_1Lepton20_ptMin150.recon.AOD.e419_a84/
Private AOD	users.NicolaGiacintoPiacquadio.ganga.datafiles.WH120highPtStudy.009995.ttbarLO_PtMin150_LeptonFilter20_Herwig.1.recon.AOD.v13000405
	ttbar with lepton filter (ptmax=150GeV)
Central EVGEN	mc08.209721.Jimmy_ttbar_1Lepton20_ptMax150.evgen.EVNT.e419_tid067451
Private EVGEN	users.NicolaGiacintoPiacquadio.ganga.datafiles.WH120highPtStudy.009992.ttbarLO_PtMax150_LeptonFilter20_Herwig.recon.ESD.v13000405
Central AOD	mc08.209721.Jimmy_ttbar_1Lepton20_ptMax150.recon.AOD.e419_a84/
Private AOD	users.NicolaGiacintoPiacquadio.ganga.datafiles.WH120highPtStudy.009992.ttbarLO_PtMax150_LeptonFilter20_Herwig.recon.AOD.v13000405
	Wt with lepton filter
Central EVGEN	mc08.209717.AcerMC_Wt_1Lepton20.evgen.EVNT.e419_tid067447
Private EVGEN	user09.NicolaGiacintoPiacquadio.evgen.mc13.CSC.009977.singleTop_Wt_AcerMC.EVNT
Central AOD	mc08.209717.AcerMC_Wt_1Lepton20.recon.AOD.e419_a84/
Private AOD	user08.NicolaGiacintoPiacquadio.ganga.datafiles.WHhighPtStudy.WH.130405.009977.mc13.009977.singleTop_Wt_AcerMC.AtlFastII.simul.AOD.v13000405.v13000405

Table 23: Produced datasets relating to lepton filtered samples

	ZH with $E_{\mathrm{T}}^{\mathrm{miss}}$filter
Central EVGEN	mc08.209139.ZH120nnbb_Herwig.evgen.EVNT.e419_tid067459
Private EVGEN	user08.NicolaGiacintoPiacquadio.ganga.datafiles.WHhighPtStudy.bbPlusMET.130405.009139.CSC.009139.ZHnnbb_Herwig.evgen.EVNT.v13000405
Central AOD	mc08.209139.ZH120nnbb_Herwig.recon.AOD.e419_a84/
Private AOD	user08.NicolaGiacintoPiacquadio.ganga.datafiles.WHhighPtStudy.bbPlusMET.130405.009139.CSC.009139.ZHnnbb_Herwig.recon.AOD.v13000405
	Z+jets with $E_{\mathrm{T}}^{\mathrm{miss}}$filter
Central EVGEN	mc08.209714.Jimmy_Zj_MET100_ptMin150.evgen.EVNT.e419_tid067444
Private EVGEN	user08.NicolaGiacintoPiacquadio.ganga.datafiles.WHhighPtStudy.bbPlusMET.130405.009986.CSC.009986.ZplusJets_Herwig.evgen.EVNT.v13000405
Central AOD	mc08.209714.Jimmy_Zj_MET100_ptMin150.recon.AOD.e419_a84/
Private AOD	user08.NicolaGiacintoPiacquadio.ganga.datafiles.WHhighPtStudy.bbPlusMET.130405.009986.CSC.009986.ZplusJets_Herwig.recon.AOD.v13000405
	W+jets with $E_{\mathrm{T}}^{\mathrm{miss}}$filter
Central EVGEN	mc08.209713.Jimmy_Wj_MET100_ptMin150.evgen.EVNT.e419_tid067443
Private EVGEN	user08.NicolaGiacintoPiacquadio.ganga.datafiles.WHhighPtStudy.bbPlusMET.130405.009985.CSC.009985.WplusJets_Herwig.evgen.EVNT.v13000405
Central AOD	mc08.209713.Jimmy_Wj_MET100_ptMin150.recon.AOD.e419_a84/
Private AOD	user08.NicolaGiacintoPiacquadio.ganga.datafiles.WHhighPtStudy.bbPlusMET.130405.009985.CSC.009985.WplusJets_Herwig.sub2.recon.AOD.v13000405
	WW with $E_{\mathrm{T}}^{\mathrm{miss}}$filter
Central EVGEN	mc08.209710.Jimmy_WW_MET100_ptMin150.evgen.EVNT.e419_tid067440
Private EVGEN	user08.NicolaGiacintoPiacquadio.ganga.datafiles.WHhighPtStudy.bbPlusMET.130405.009984.CSC.009984.WW_Herwig.evgen.EVNT.v13000405
Central AOD	mc08.209710.Jimmy_WW_MET100_ptMin150.recon.AOD.e419_a84/
Private AOD	user08.NicolaGiacintoPiacquadio.ganga.datafiles.WHhighPtStudy.bbPlusMET.130405.009984.CSC.009984.WW_Herwig.recon.AOD.v13000405
	WZ with $E_{\mathrm{T}}^{\mathrm{miss}}$filter
Central EVGEN	mc08.209712.Jimmy_WZ_MET100_ptMin150.evgen.EVNT.e419_tid067442
Private EVGEN	user08.NicolaGiacintoPiacquadio.ganga.datafiles.WHhighPtStudy.bbPlusMET.130405.009987.CSC.009987.WZ_Herwig.evgen.EVNT.v13000405
Central AOD	mc08.209712.Jimmy_WZ_MET100_ptMin150.recon.AOD.e419_a84/
Private AOD	user08.NicolaGiacintoPiacquadio.ganga.datafiles.WHhighPtStudy.bbPlusMET.130405.009987.CSC.009987.WZ_Herwig.recon.AOD.v13000405
	ZZ with $E_{\mathrm{T}}^{\mathrm{miss}}$filter
Central EVGEN	mc08.209711.Jimmy_ZZ_MET100_ptMin150.evgen.EVNT.e419_tid067441
Private EVGEN	user08.NicolaGiacintoPiacquadio.ganga.datafiles.WHhighPtStudy.bbPlusMET.130405.009980.CSC.009980.ZZ_Herwig.evgen.EVNT.v13000405
Central AOD	mc08.209711.Jimmy_ZZ_MET100_ptMin150.recon.AOD.e419_a84/
Private AOD	user08.NicolaGiacintoPiacquadio.ganga.datafiles.WHhighPtStudy.bbPlusMET.130405.009980.CSC.009980.ZZ_Herwig.recon.AOD.v13000405
	ttbar with $E_{\mathrm{T}}^{\mathrm{miss}}$filter
Central EVGEN	mc08.209715.Jimmy_ttbar_MET100_ptMin150.evgen.EVNT.e419_tid067445
Private EVGEN	user08.NicolaGiacintoPiacquadio.ganga.datafiles.WHhighPtStudy.bbPlusMET.130405.009983.CSC.009983.ttbar_Herwig.evgen.EVNT.v13000405
Central AOD	mc08.209715.Jimmy_ttbar_MET100_ptMin150.recon.AOD.e419_a84/
Private AOD	user08.NicolaGiacintoPiacquadio.ganga.datafiles.WHhighPtStudy.bbPlusMET.130405.009983.CSC.009983.ttbar_Herwig.recon.AOD.v13000405
	Wt with $E_{\mathrm{T}}^{\mathrm{miss}}$filter
Central EVGEN	mc08.209716.AcerMC_Wt_MET100.evgen.EVNT.e419/
Private EVGEN	user09.NicolaGiacintoPiacquadio.evgen.mc13.CSC.009967.singleTop_Wt_AcerMC.EVNT
Central AOD	mc08.209716.AcerMC_Wt_MET100.recon.AOD.e419_a84/
Private AOD	dq2_user08.veysierkcanozcan.ganga.datafiles.singleTop_Wt_AcerMC.009967.AtlFastII_batch0.simul.AOD.v13004005.v13004005

Table 24: Produced datasets relating to $E_{\mathrm{T}}^{\mathrm{miss}}$filtered samples