Probing lepton flavour violation via neutrinoless $\tau \rightarrow 3\mu$ decays with the ATLAS detector

The ATLAS Collaboration

Abstract

This article presents the sensitivity of the ATLAS experiment to the lepton-flavour-violating decays of $\tau \rightarrow 3\mu$. A method utilising the production of $\tau$ leptons via $W \rightarrow \tau\nu$ decays is used. This method is applied to the sample of 20.3 fb$^{-1}$ of $pp$ collision data at a centre-of-mass energy of 8 TeV collected by the ATLAS experiment at the LHC in 2012. No event is observed passing the selection criteria, and the observed (expected) upper limit on the $\tau$ lepton branching fraction into three muons, $\text{Br}(\tau \rightarrow 3\mu)$, is $3.76 \times 10^{-7}$ ($3.94 \times 10^{-7}$) at 90% confidence level.
## Contents

1 Introduction .................................................. 3

2 The ATLAS detector ........................................ 4

3 Simulation and data samples ............................... 4

4 Trigger and reconstruction ............................... 5

5 Analysis procedure ...................................... 6
   5.1 Object selection ........................................ 7
   5.2 Loose event selection ................................. 8
   5.3 Multivariate analysis ................................. 9
   5.4 Tight event selection .............................. 10
   5.5 Background estimation ............................ 11
   5.6 Uncertainties and optimisation .................. 12

6 Results ...................................................... 13

7 Conclusions and outlook ............................... 14
1 Introduction

The observation of a lepton-flavour-violating (LFV) process involving charged leptons would be a major breakthrough in understanding the matter content of the universe and would support the hypothesis of leptogenesis [1]. In particular, LFV processes involving both a $\tau$ lepton and a muon are seen as most promising for such an observation, given the current measurements of neutrino oscillations [2]. In the Standard Model (SM), such processes have a vanishingly small branching fraction, e.g. $\text{Br}(\tau \to 3\mu) < 10^{-14}$ [3], while a number of models beyond the SM predict it to be of the order of $10^{-10} - 10^{-8}$ [4–6].

The current limits on branching fractions of neutrinoless $\tau$ lepton decays are of the order of few times $10^{-8}$ [7–10], for $Z$ boson LFV decays they are about $10^{-5}$ [2, 11, 12], and for the LFV decay of a Higgs boson to a $\tau$ lepton and a muon they are about 1% [13, 14]. The main experimental obstacles to improve the sensitivity with $\tau$ leptons are the small number of produced $\tau$ leptons world-wide.

In this article, a search for neutrinoless $\tau$ lepton decays to three muons is performed with 20.3 fb$^{-1}$ of $pp$ collision data collected with ATLAS detector in 2012 at 8 TeV centre-of-mass energy. The search is focused on a particular source of $\tau$ leptons, namely $W \to \tau \nu$ decays with subsequent $\tau \to 3\mu$ decay. In such events, $\tau$ leptons are produced with a transverse momentum ($p_T$) mostly in the range of $\sim 25 - 50$ GeV. Due to the relativistic boost of the $\tau$ lepton, the muons from the $\tau$ LFV decay are produced in close geometrical proximity to each other but isolated from other energetic particles in the event. The tau-neutrino from the $W$ boson decay appears as missing transverse momentum ($E_{\text{miss}}^T$) in the detector and together with the transverse momentum of the three muons ($p_{T}^{3\mu}$) gives a transverse mass, $m_T = \sqrt{2p_{T}^{3\mu}E_{\text{miss}}^T(1 - \cos \Delta \phi)}$, compatible with the $W$ boson decay, where $\Delta \phi$ is the angle between the directions of the $p_{T}^{3\mu}$ and the $E_{\text{miss}}^T$. The unique signature in the detector is three muons with invariant mass equal to the mass of the $\tau$ lepton and with a significant missing transverse momentum that is on average back-to-back with the three muons in the transverse plane. Since no energetic jet is expected in the majority of $W$ boson production events, very small hadronic activity is predicted beyond that from the soft underlying event or multiple simultaneous $pp$ collisions (pile-up). A large fraction of such $\tau$ leptons decay sufficiently far from the $W$ production vertex to give a fully reconstructable additional vertex. This allows the selection of three muons originating from a vertex which is displaced from the primary interaction vertex. The background events usually contain one or two muons originating from the decay of hadrons, including decays in flight, while the remaining tracks are hadrons mis-measured as muons, originating from e.g. a pile-up jet or a pion punching through the calorimeter. The dominant background is due to muons originating from decays of $b$- or $c$-hadrons (heavy flavour, HF). Although such decays are typically accompanied by jets of particles produced in the direction opposite to the HF jets, in a fraction of the events the associated jet is lost or mis-measured, mimicking the signal $E_{\text{miss}}^T$. A small light-flavour multi-jet contribution is also present while the contribution from leptonic decays of vector bosons is negligible.

The analysis strategy is as follows. Events with three muons associated with a common vertex are selected. A loose event selection is applied to collect a high-quality sample of candidate events satisfying $|m_{3\mu} - m_\tau| \lesssim 1$ GeV. The characteristics of the loose sample of events are then analysed with a boosted decision tree (BDT). The BDT input variables are chosen so that the BDT output and the three-muon mass are uncorrelated in the mass range used in the analysis. A tight selection, following an initial cut on the BDT output, is applied to separate the signal from the background. After the optimal cut on the BDT output is found, a search is performed for an excess of events at the $\tau$ lepton mass above the expected background level.
The branching fraction is calculated as

\[ \text{Br}(\tau \to 3\mu) = \frac{N_s}{(A_s \times \epsilon_s) N_{W\to\tau\nu}}, \]

where \( N_s \) is the number of observed events above the expected background level in a narrow region around the \( \tau \) lepton mass, \( A_s \times \epsilon_s \) is the detector acceptance times efficiency for the signal, and \( N_{W\to\tau\nu} \) is the number of \( \tau \) leptons produced via the \( W \rightarrow \tau\nu \) channel (additional contributions to the \( \tau \) lepton yield are estimated to be less than 3%).

2 The ATLAS detector

The ATLAS experiment [15] at the LHC is a multi-purpose particle detector with a forward-backward symmetric cylindrical geometry and a near 4\( \pi \) coverage in solid angle.\(^1\) It consists of an inner tracking detector surrounded by a thin superconducting solenoid providing a 2 T axial magnetic field, electromagnetic and hadronic calorimeters, and a muon spectrometer. The inner tracking detector (ID) covers the pseudorapidity range \(|\eta| < 2.5\). It consists of silicon pixel, silicon microstrip, and transition radiation tracking detectors. Lead/liquid-argon (LAr) sampling calorimeters provide electromagnetic (EM) energy measurements with high granularity. A hadronic (iron/scintillator-tile) calorimeter covers the central pseudorapidity range \(|\eta| < 1.7\). The endcap and forward regions are instrumented with LAr calorimeters for EM and hadronic energy measurements up to \(|\eta| = 4.9\).

The muon spectrometer (MS) comprises separate trigger and high-precision tracking chambers measuring the deflection of muons in a magnetic field generated by superconducting air-core toroids. The magnets’ bending power is in the range from 2.0 to 7.5 T m. The muon tracking chambers cover the region \(|\eta| < 2.7\) with three layers of monitored drift tubes, complemented by cathode-strip chambers in the forward region, where the background is highest. The muon trigger system covers the range \(|\eta| < 2.4\) with resistive-plate chambers in the barrel, and thin-gap chambers in the endcap regions.

A three-level trigger system is used to select events. The first-level trigger is implemented in hardware and uses a subset of the detector information to reduce the accepted rate to at most 75 kHz. This is followed by two software-based trigger levels that together reduce the accepted event rate to 400 Hz on average.

During the data-taking period, there were no dedicated triggers implemented for this analysis. A combination of seven muon triggers is used, where all triggers are constructed from at least two trigger objects. A detailed discussion of the trigger is given in Section 4.

3 Simulation and data samples

The results presented here are based on proton–proton collision data at a centre-of-mass energy of \( \sqrt{s} = 8\) TeV, collected by the ATLAS detector at the LHC during 2012. Data samples corresponding to an integ-

\(^1\) ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the \( z \)-axis along the beam pipe. The \( x \)-axis points from the IP to the centre of the LHC ring, and the \( y \)-axis points upwards. Cylindrical coordinates \((r, \phi)\) are used in the transverse plane, \( \phi \) being the azimuthal angle around the \( z \)-axis. The pseudorapidity is defined in terms of the polar angle \( \theta \) as \( \eta = -\ln(\tan(\theta/2)) \). Angular distance is measured in units of \( \Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \).
rated luminosity of 20.3 fb⁻¹ are used. Selected data events are required to have all relevant components of the ATLAS detector in good working condition.

The Monte Carlo (MC) simulated $W \to \tau \nu \to (3 \mu)\nu$ signal sample is produced by the Pythia8 [16] event generator (version 8.175) using the AU2 [17] set of tuned parameters and the MSTW2008LO parton distribution function (PDF) set [18]. This signal sample is modelled using $W \to \tau \nu$ production where the $\tau$ lepton is forced to decay isotropically into three muons as in previous searches for this mode [7–10].

The detector response is modelled using GEANT4 [19, 20]. The number of $W\to\tau\nu$ events is 31% for the combination of all triggers used in the analysis. To evaluate the trigger performance in the region where the muons have a small angular separation, as is typical for the signal, a tag-and-probe study is performed using data events containing $\ell\nu$ candidates. For this study, the data are collected using a single-muon baseline trigger. The software-based trigger $E_T^{\text{miss}}$ trigger and one dimuon plus $E_T^{\text{miss}}$ trigger. The software-based trigger thresholds used for the muons range from 4 to 18 GeV in transverse momentum while the $E_T^{\text{miss}}$ threshold is 30 GeV. The trigger efficiency for simulated signal events within the muon-trigger acceptance (three generator-level muons with $p_T > 2.5$ GeV and $|\eta| < 2.4$) is $\sim 31\%$ for the combination of all triggers used in the analysis. To evaluate the trigger performance in the region where the muons have a small angular separation, as is typical for the signal, a tag-and-probe study is performed using data events containing high-momentum $J/\psi \rightarrow \mu\mu$ candidates. For this study, the data are collected using a single-muon baseline trigger with a $p_T$ threshold of 18 GeV. Single-muon efficiencies are measured separately for the different thresholds which define the six multi-muon triggers. Each multi-muon trigger efficiency is calculated as the product of the single-muon efficiencies. Correction factors are applied to account for the limited performance of the trigger system in identifying a pair of muons as two muon-trigger objects. At small angular separations ($\Delta R \leq 0.2$), where most of the signal is expected and where these limitations are most pronounced, these corrections must be taken into account. These factors are measured from the efficiency to identify two independent muon-trigger objects for different $\Delta R$ values between the tag- and the probe-muon. The total efficiency of the seven triggers is calculated considering correlations between any of the triggers. The trigger efficiency, measured from the data, is compared to the one measured in simulated $J/\psi$ events for the seven different triggers separately and jointly. Agreement between data and MC simulation was found to be within $11\%$ for all relevant values of $\Delta R$ and $p_T$, where the largest difference comes from events where the $\Delta R$ separation is smallest. The systematic uncertainty on $\mathcal{A}_k \times \epsilon_k$ due to the trigger is therefore taken to be $11\%$.

The approach for measuring the muon reconstruction efficiency is similar to that used to measure the trigger efficiency. While the trigger efficiency is measured with respect to muon reconstruction as the baseline, the reconstruction efficiency is measured with respect to ID tracking, which in turn is close to $100\%$ efficient [22]. Small deviations from the assumed value for ID tracking efficiency have a negligible impact on this measurement. The tag-and-probe procedure is performed using muons as tags and ID

4 Trigger and reconstruction

To maximise the signal acceptance times efficiency, events are required to pass at least one of seven triggers. These are six multi-muon triggers and one dimuon plus $E_T^{\text{miss}}$ trigger. The software-based trigger $E_T^{\text{miss}}$ is $30\%$ for the combination of all triggers used in the analysis. To evaluate the trigger performance in the region where the muons have a small angular separation, as is typical for the signal, a tag-and-probe study is performed using data events containing $\ell\nu$ candidates. For this study, the data are collected using a single-muon baseline trigger. The software-based trigger $E_T^{\text{miss}}$ trigger and one dimuon plus $E_T^{\text{miss}}$ trigger. The software-based trigger thresholds used for the muons range from 4 to 18 GeV in transverse momentum while the $E_T^{\text{miss}}$ threshold is 30 GeV. The trigger efficiency for simulated signal events within the muon-trigger acceptance (three generator-level muons with $p_T > 2.5$ GeV and $|\eta| < 2.4$) is $\sim 31\%$ for the combination of all triggers used in the analysis. To evaluate the trigger performance in the region where the muons have a small angular separation, as is typical for the signal, a tag-and-probe study is performed using data events containing high-momentum $J/\psi \rightarrow \mu\mu$ candidates. For this study, the data are collected using a single-muon baseline trigger with a $p_T$ threshold of 18 GeV. Single-muon efficiencies are measured separately for the different thresholds which define the six multi-muon triggers. Each multi-muon trigger efficiency is calculated as the product of the single-muon efficiencies. Correction factors are applied to account for the limited performance of the trigger system in identifying a pair of muons as two muon-trigger objects. At small angular separations ($\Delta R \leq 0.2$), where most of the signal is expected and where these limitations are most pronounced, these corrections must be taken into account. These factors are measured from the efficiency to identify two independent muon-trigger objects for different $\Delta R$ values between the tag- and the probe-muon. The total efficiency of the seven triggers is calculated considering correlations between any of the triggers. The trigger efficiency, measured from the data, is compared to the one measured in simulated $J/\psi$ events for the seven different triggers separately and jointly. Agreement between data and MC simulation was found to be within $11\%$ for all relevant values of $\Delta R$ and $p_T$, where the largest difference comes from events where the $\Delta R$ separation is smallest. The systematic uncertainty on $\mathcal{A}_k \times \epsilon_k$ due to the trigger is therefore taken to be $11\%$.

The approach for measuring the muon reconstruction efficiency is similar to that used to measure the trigger efficiency. While the trigger efficiency is measured with respect to muon reconstruction as the baseline, the reconstruction efficiency is measured with respect to ID tracking, which in turn is close to $100\%$ efficient [22]. Small deviations from the assumed value for ID tracking efficiency have a negligible impact on this measurement. The tag-and-probe procedure is performed using muons as tags and ID
tracks as probes. The baseline sample for the reconstruction efficiency measurement includes a large number of non-muon tracks, which must be subtracted. This is done in bins of probe-track $p_T (p_{trk})$ and bins of the angular separation between the tag-muon and the probe-track, $\Delta R_{\mu+trk}$. To describe the $J/\psi$ peak and the background, a small range in tag-muon plus probe-track invariant mass, $m_{\mu+trk} \in [2600, 3500]$ MeV, is fit to a double Gaussian function plus an exponential function and a second-order polynomial. In each $p_{trk}$ or $\Delta R_{\mu+trk}$ bin, the ratio of the $J/\psi$ peak component integral to the full shape ($J/\psi$ plus background) integral is used as a weight to correct the $p_{trk}$ or $\Delta R_{\mu+trk}$ shape itself. This is done separately for the probe-track distributions (denominators) and the muon-matched probe-track distributions (numerators). The ratio of the above two weighted distributions is defined as the reconstruction efficiency per $p_{trk}$ or $\Delta R_{\mu+trk}$ bin. The efficiency measured with this approach in data is compared with the one from simulation and the difference at small $\Delta R_{\mu+trk}$ results in an uncertainty of 13.1% per event.

5 Analysis procedure

The analysis procedure is divided into four steps. First, events containing three high-quality muon objects with a combined invariant mass of less than 2.5 GeV are selected. These muons are required to originate from a common vertex. Second, a loose selection is applied to this sample to obtain a background sample that can be used to train the BDT, which is constructed using the TMVA toolkit [23]. The loose selection cuts (using a number of vertex quantities as well as kinematic quantities) are chosen to obtain a large background sample for training, while rejecting background that is kinematically inconsistent with the signal. Before training the BDT, the data events are divided into three regions based on the three-muon mass value. These are the blinded region (which includes the signal region), a sideband region and a BDT training region as defined in Table 1. Third, a tight selection (tightening the loose selection with a few additional cuts) is applied while simultaneously placing an initial loose cut on the BDT score, denoted by $x > x_0$. The $x > x_0$ cut removes background-like events having a very low BDT score, while the tight selection further reduces the background in the blinded and sideband regions. Fourth, the background rejection as a function of the BDT cut is studied using data events in the sideband region passing the tight $x > x_0$ selection. This allows to optimise the final cut on the BDT score, denoted by $x > x_1$. The statistical analysis is performed for the tight $x > x_1$ selection.

<table>
<thead>
<tr>
<th>Region</th>
<th>Range in $m_{3\mu}$ [MeV]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal region</td>
<td>[1713, 1841]</td>
</tr>
<tr>
<td>Blinded region</td>
<td>[1690, 1870]</td>
</tr>
<tr>
<td>Sideband region</td>
<td>[1450, 1690] and [1870, 2110]</td>
</tr>
<tr>
<td>Training region</td>
<td>[750, 1450] and [2110, 2500]</td>
</tr>
</tbody>
</table>

Table 1: The different three-muon mass ranges used in the analysis.

The signal region (SR) is defined as an interval around the $\tau$ lepton mass with a half-width corresponding to twice the resolution of the three-muon mass, $\sigma_3 = 32$ MeV, as obtained from the signal MC sample. The analysis was blinded in a slightly wider region to allow variation of the signal region definition. The signal MC sample is divided into two independent samples. One signal sample is used for the BDT training while the second signal sample is used for estimating the $\mathcal{R} \times \epsilon$. The background in the signal region is estimated from a fit to the three-muon mass distribution in the sidebands (SB) using the
tight+$x>x_0$ selection. This estimate is then scaled down to the final BDT score cut, $x_1$, using a fit to the BDT shape as explained below.

5.1 Object selection

Muons are selected to have a transverse momentum greater than 2.5 GeV and are required to pass stringent requirements on the track quality and the associated hits in both the ID and the MS. Only combined ID+MS measurements of track parameters are used. Several matching criteria [22] are imposed to reject non-muon tracks (e.g. tracks from hadron decays in flight). The performance of muon identification is validated in two dedicated dimuon control regions. One region is populated with muons from $J/\psi \rightarrow \mu\mu$ decay (in $2850 < m_{3\mu} < 3350$ MeV), while the second region has an enhanced fraction of non-muon tracks (in events with $m_{3\mu} < 750$ MeV).

Events with at least three selected muons are considered. All possible three-muon combinations are used as inputs to a vertex fit. The primary vertex (PV) is also refitted after removing the three tracks. Due to the $\tau$ lepton lifetime, the three-muon vertex is often separated from the PV. The characteristics of the separation between the three-muon vertex and the PV are therefore used to distinguish signal from background. Particularly, the two projections of the three-muon vertex displacement with respect to the PV in the transverse plane are used; $L_{xy}=L_T \cos \theta_{xy}$ and $d_{xy}^0=L_T \sin \theta_{xy}$ where $L_T$ is the transverse component of the vector connecting the PV and the three-muon vertex and $\cos \theta_{xy}=\vec{L}_T \cdot \vec{p}_{3\mu}/L_T p_{3\mu}$. The three-muon vertex fit probability, $p$-value, is also used (as calculated from the vertex fit $\chi^2$ and degrees of freedom). After fitting all possible vertices, exactly one three-muon candidate is allowed per event, satisfying $m_{3\mu} < 2500$ MeV and $|Q_{3\mu}| = 1$ where $Q_{3\mu}$ is the sum of the charges of the three-muon tracks.

Jets are used to separate the signal from the multi-jet backgrounds (predominantly HF), where more hadronic activity is expected. The jets are reconstructed from topological clusters formed in the calorimeter using the anti-$k_t$ algorithm [24] with a radius parameter $R = 0.4$. The jets are calibrated to the hadronic energy scale using energy- and $\eta$-dependent correction factors derived from simulation and with residual corrections from in situ measurements. A detailed description of the jet energy scale measurement and its systematic uncertainties can be found in Ref. [25]. Jets found within a cone of $\Delta R = 0.2$ around a selected three-muon candidate are removed. Jets are required to have $p_T > 30$ GeV and $|\eta| < 2.8$; only the leading jet satisfying these criteria is considered. There is no veto of events with more than one jet satisfying these criteria. The leading jet and the three-muon momenta are summed vectorially to define $\vec{\Sigma} = \vec{p}_{jet} + \vec{p}_{3\mu}$ with $\Sigma_T$ being the magnitude of its transverse component. For events where there are no jets satisfying these criteria (the majority of events for the signal), $\vec{\Sigma}$ is simply $\vec{p}_{3\mu}$.

The $E_T^{\text{miss}}$ is calculated as the negative vector sum of the transverse momenta of all high-$p_T$ objects reconstructed in the event, as well as a term for other activity in the calorimeter [26]. Clusters associated with electrons, hadronic $\tau$ lepton decays and jets are calibrated separately, with other clusters calibrated at the EM energy scale. This $E_T^{\text{miss}}$ is denoted hereafter by $E_T^{\text{miss,cal}}$. In addition, a track-based missing transverse momentum ($E_T^{\text{miss,trk}}$) is calculated as the negative vector sum of the transverse momenta of tracks with $|\eta| < 2.5$, $p_T > 500$ MeV and associated with the primary vertex. Both the calorimeter-based and track-based measurements of the $E_T^{\text{miss}}$ are used.
Several kinematic variables are defined from the reconstructed objects listed above. Two transverse masses are defined using the three-muon transverse momentum ($p_T^{3\mu}$) as $m_T = \sqrt{2p_T^{3\mu}E_T^{\text{miss}}(1 - \cos \Delta \phi_{3\mu})}$, where $\Delta \phi_{3\mu}$ is the angle between the $E_T^{\text{miss}}$ and $p_T^{3\mu}$ directions in the transverse plane. In these definitions, $E_T^{\text{miss}}$ can be either $E_{T,\text{cal}}^{\text{miss}}$ or $E_{T,\text{trk}}^{\text{miss}}$ to obtain $m_{T,\text{cal}}$ or $m_{T,\text{trk}}$ respectively. The $\Delta \phi_{3\mu}$ terms are $\Delta \phi_{3\mu}^{\text{cal}}$ or $\Delta \phi_{3\mu}^{\text{trk}}$ respectively. Similarly, the $\Delta \phi_{\Sigma^3}$ variable is the angle between the $E_T^{\text{miss}}$ and $\Sigma$ directions in the transverse plane. This adds two additional angles, $\Delta \phi_{\Sigma^3}^{\text{cal}}$ and $\Delta \phi_{\Sigma^3}^{\text{trk}}$ for $E_{T,\text{cal}}^{\text{miss}}$ and $E_{T,\text{trk}}^{\text{miss}}$ respectively. These $\Delta \phi_{\Sigma^3}$ variables provide good separation when a hard jet is found and thus $\Sigma$ deviates from $p_T^{3\mu}$ in magnitude and direction.

5.2 Loose event selection

After the three-muon candidates are formed from the selected muons, a loose event selection is performed, maintaining a signal efficiency of about 80% while rejecting about 95% of the background. This loose selection includes cuts on the displacement of the vertex from the PV, requirements on the three-muon kinematics and on the presence of other tracks (track isolation), and requirements on quantities involving $E_{T,\text{cal}}^{\text{miss}}$ and $E_{T,\text{trk}}^{\text{miss}}$. The loose selection comprises the following requirements:

- The $L_{xy}$ significance, $S(L_{xy})=L_{xy}/\sigma_{L_{xy}}$, must satisfy $-10<S(L_{xy})<50$, where $\sigma_{L_{xy}}$ is the uncertainty in the $L_{xy}$.
- The $a_{xy}^0$ significance, $S(a_{xy}^0)=a_{xy}^0/\sigma_{a_{xy}^0}$, must satisfy $S(a_{xy}^0)<25$, where $\sigma_{a_{xy}^0}$ is the uncertainty in $a_{xy}^0$.
- The three-muon track-fit probability product, $P_{\text{trks}} = p_1 \times p_2 \times p_3$ (where $p_i$ is the track fit $p$-value of track $i$), must satisfy $P_{\text{trks}} > 10^{-9}$.
- The three-muon transverse momentum must satisfy $p_T^{3\mu} > 10$ GeV.
- The missing transverse energies, $E_{T,\text{cal}}^{\text{miss}}$ and $E_{T,\text{trk}}^{\text{miss}}$, must both satisfy $10 < E_T^{\text{miss}} < 250$ GeV.
- The transverse masses, $m_{T,\text{cal}}$ and $m_{T,\text{trk}}$, must both satisfy $m_T > 20$ GeV.
- The three-muon track isolation is obtained from the sum of the $p_T$ of all tracks with $p_T^{\text{trk}} > 500$ MeV in a cone of $\Delta R_{\mu\text{max}}^{0.2}$ (and $\Delta R_{\mu\text{max}}^{0.30}$) around the three-muon momentum while excluding its constituent tracks; it must satisfy $\Sigma p_T^{\text{trk}}(\Delta R_{\mu\text{max}}^{0.2})/p_T^{3\mu} < 0.3$ (and $\Sigma p_T^{\text{trk}}(\Delta R_{\mu\text{max}}^{0.30})/p_T^{3\mu} < 1$). The largest separation, $\Delta R_{\mu\text{max}}^{3\mu}$, between any pair of the three-muon tracks is on average 0.07 for the signal.

The loose cuts on the significances, $S(L_{xy})$ and $S(a_{xy}^0)$, are applied to allow the three-muon vertex to be separated from the PV, while still being compatible with the $\tau$ lepton lifetime. The requirement on $P_{\text{trks}}$ imposes a goodness-of-fit criterion on the three-muon candidate. This value is based on an examination of signal-like events found in the sideband region in the data. As this is not the only quality requirement imposed on the individual muon objects, it is kept loose in this part of the selection. The efficiency for this cut to select signal events is $\sim 98\%$, while it is rejecting $\sim 13\%$ of the background events. The kinematic and the isolation variables are very effective in separating the $W$ boson properties of the signal from the HF and the light-flavour multi-jet background, which tend to be non-isolated and with low values of $p_T$, $E_T^{\text{miss}}$ and $m_T$. The associated cuts remain very loose in this part of the selection to ensure that the sample sizes are large enough for the BDT training.
5.3 Multivariate analysis

The events passing the loose selection described above are used as input to the BDT training. There are 6649 events passing the loose selection in the signal MC sample (out of $10^5$), where 6000 of these events are used for the BDT training and the rest are used for testing the BDT output. Similarly, the number of data events passing the loose selection in the training region is 4672, where 4000 of these events are used for the BDT training. The BDT input variables include kinematic distributions, track and vertex quality discriminants, vertex geometry parameters, and isolation. The following variables (sorted by their importance ranking) are used as inputs to the BDT:

1. The calorimeter-based transverse mass, $m_{T\text{cal}}$.
2. The track-based missing transverse momentum, $E_{T,\text{trk}}^{\text{miss}}$.
3. The isolation variable, $\Sigma p_T^{\text{trk}}(\Delta R_{\text{max}}^{3\mu} + 0.20)/p_T^{3\mu}$.
4. The transverse component of the vector sum of the three-muon and leading jet momenta, $\Sigma_{T}$.
5. The track-based transverse mass, $m_{T\text{trk}}$.
6. The difference between the $E_{T,\text{cal}}^{\text{miss}}$ and $E_{T,\text{trk}}^{\text{miss}}$ directions, $\Delta\phi_{\text{trk}}^{\text{cal}}$.
7. The calorimeter-based missing transverse momentum, $E_{T,\text{cal}}^{\text{miss}}$.
8. The track-based missing transverse momentum balance $p_T^{3\mu}/E_{T,\text{trk}}^{\text{miss}} - 1$.
9. The difference between the three-muon and $E_{T,\text{cal}}^{\text{miss}}$ directions, $\Delta\phi_{3\mu}^{\text{cal}}$.
10. Three-muon vertex fit probability, $p$-value.
11. The three-muon vertex fit $a_{xy}^{0}$ significance, $S(a_{xy}^{0})$.
12. The track fit probability product, $P_{\text{trks}}$.
13. The three-muon transverse momentum, $p_T^{3\mu}$.
14. The number of tracks associated with the PV (after refitting the PV while excluding the three-muon tracks), $N_{\text{trk}}^{\text{PV}}$.
15. The three-muon vertex fit $L_{xy}$ significance, $S(L_{xy})$.
16. The calorimeter-based missing transverse momentum balance, $p_T^{3\mu}/E_{T,\text{cal}}^{\text{miss}} - 1$.

This configuration was found to give the optimal balance between background rejection and signal efficiency.

The $\Sigma_T$ variable is introduced to avoid vetoing events with at least one jet fulfilling the requirements listed in section 5.1. Although the majority of signal events do not have jets, it is found that keeping such events increases the $\mathcal{A}_s \times \epsilon_s$ by $\sim 15\%$ and also ultimately leads to better rejection power, owing to the significantly larger training and sideband samples. The variables $\Delta\phi_{\text{trk}}^{\text{cal}}$, $\Delta\phi_{3\mu}^{\text{cal}}$, $p_T^{3\mu}/E_{T,\text{trk}}^{\text{miss}} - 1$, $p_T^{3\mu}/E_{T,\text{cal}}^{\text{miss}} - 1$ and $N_{\text{trk}}^{\text{PV}}$ are complementary to the $E_{T,\text{cal}}^{\text{miss}}$-related variables used in the loose selection as well as here. These variables are also very effective in distinguishing the $W \rightarrow \tau\nu$ production of the signal from the HF and light-flavour multi-jet background. The vertex $p$-value is a variable complementary to
the $S(L_{xy})$ and $S(a_{xy}^0)$ variables used in the loose selection as well as here. The HF and light-flavour multi-jet backgrounds have mostly random combinations of selected muon objects which do not originate from the same vertex. This variable peaks at very low values for the background while for the signal it is distributed uniformly and thus provides excellent separation.

After training the BDT with data events from the training region and signal MC events from the first signal MC sample, the BDT response is calculated for the data events in the sidebands and for events in the second signal sample. The BDT score, $x$, ranges between -1 and +1. Events with a very low BDT score, within $-1 \leq x \leq -0.9$ are removed from further consideration, defining $x_0 \equiv -0.9$.

In order to assess potential modelling problems in the signal MC sample, the BDT input distributions and the BDT response are validated against single-muon data. These data contain mainly $W \rightarrow \mu\nu$ events with a small fraction (< 10%) of background. The single-muon selection is formulated to be as close as possible to the main analysis selection where the differences are mostly driven by the different triggers used (one single-muon trigger with no isolation requirement and with a threshold of 24 GeV is used in the validation) and the exclusion of variables which do not have equivalents in the $W \rightarrow \mu\nu$ case, e.g. the three-muon vertex variables. The training samples used for this validation study, for both data and signal, are the same samples as used in the main analysis, constructed with the same loose selection as described in the previous section. All input variables are used for the training, excluding the $p$-value, $S(L_{xy})$, $S(a_{xy}^0)$ and $P_{trks}$, which cannot be calculated in a single-muon ($W \rightarrow \mu\nu$) selection. The resulting BDT setup is hereafter referred to as “partial BDT”. After training the partial BDT, the response is tested on the second signal sample and on the single-muon data, using the single-muon selection and where the three muon objects in the signal sample are treated as one object (muon). The $N_{\text{trk}}^{\text{PV}}$ distribution of the signal sample is also modified by subtracting two tracks to reflect the difference with respect to a single-muon selection. The responses in data and simulation are compared and are found to agree within 10% throughout most of the phase-space for all variables. The ratio of the partial BDT responses for the single-muon data and signal MC events is used as an event weight while applying the full selection and calculating the weighted $A_s \times \epsilon_s$ as described in the next sections. The difference between the weighted and unweighted $A_s \times \epsilon_s$ is found to be 4% and is taken as a modelling uncertainty.

Any variable which may bias the BDT response by only selecting events very close to the $\tau$ lepton mass is not included in the BDT input list. The distribution of the three-muon mass has been examined in several bins of $x$ above $x_0$ using both the loose and the tight samples, where no hint of potential peaking background around the $\tau$ lepton mass has been found. In addition, the shape of the three-muon mass distribution has been found to be insensitive to the BDT cut, as expected given the small correlation coefficient between $x$ and $m_{3\mu}$, which is found to be about -0.05.

### 5.4 Tight event selection

Additional tight cuts are applied after the BDT training and the application of the $x>x_0$ cut on the BDT score. The following requirements are tightened or added:

- A number of the loose requirements are tightened, namely $P_{\text{trks}} > 8 \times 10^{-9}$, $m_T^{\text{cal}} > 45$ GeV, $m_{\text{trk}}^{\text{cal}} > 45$ GeV and $1 < S(L_{xy}) < 50$.
- Three-muon vertex fit probability must have $p$--value > 0.2.
- The angle between the $\Sigma_T$ and $E_{T,\text{cal}}^{\text{miss}}$ ($E_{T,\text{trk}}^{\text{miss}}$) directions is required to be $\Delta \phi_{2T}^{\text{cal}} > 2$ ($\Delta \phi_{2T}^{\text{trk}} > 2$).
- The same-charge two-muon mass, \( m_{SS} \), and opposite-charge two-muon mass, \( m_{OS1} \) or \( m_{OS2} \), satisfy \( m_{SS} > 300 \text{ MeV} \), \( m_{OS1} > 300 \text{ MeV} \) and \( m_{OS2} > 300 \text{ MeV} \), where \( m_{OS1} \) (\( m_{OS2} \)) is the mass of the two opposite-charge muon pairs with the highest (second highest) summed scalar \( p_T \) among the three muons.

- The event is rejected if \( |m_{OS} - m_{\omega}| < 50 \text{ MeV} \) or \( |m_{OS} - m_{\phi}| < 50 \text{ MeV} \) if either of the \( p_T^{3\mu} \), the \( E_{T,\text{cal}}^{\text{miss}} \) or the \( E_{T,\text{trk}}^{\text{miss}} \) is lower than 35 GeV.

- The event is rejected if \( |m_{OS} - m_{\phi}| < 50 \text{ MeV} \) if \( |m_{3\mu} - m_{D_s}| < 100 \text{ MeV} \).

In the above notation, \( m_{OS} \) is \( m_{OS1} \) or \( m_{OS2} \) and \( m_{\omega}, m_{\phi} \) and \( m_{D_s} \) are the masses of the \( \omega, \phi \) and \( D_s \) mesons respectively, taken from Ref. [2].

The requirement on the three-muon vertex fit probability is applied in order to ensure a high-quality fit. The cuts on \( \Delta \phi^{\text{cal}}_{2\gamma} (\Delta \phi^{\text{trk}}_{2\gamma}) \) are applied in order to further suppress the HF and multi-jet background where the three-muon candidate is typically produced within or near a jet.

The first two-muon mass requirement is applied to suppress candidates originating from one prompt muon object and two muons from a converted photon. The second requirement on the two-muon masses is applied to prevent the low-mass mesons, \( \rho/\omega \) and \( \phi \), from entering into the region close to the \( \tau \) lepton mass when combined with an additional track. In the selected three-muon event sample, these resonances appear as two clear peaks in the mass distribution of oppositely charged muon pairs in data. Since the resonances lie in the middle of the signal distribution, the low-\( p_T \) and \( E_{T,\text{miss}}^{\text{cal}} \) requirement ensures that these can still be distinguished from the signal, and thus it removes the resonances while still maintaining a high enough signal efficiency. Finally, the last requirement is applied to remove a potential \( D_s \rightarrow \pi + \phi(\mu\mu) \) contamination from the high-mass sideband. The cuts listed above comprise the tight selection where the tight+\( x>x_0 \) selection is used to estimate the background for any cut on \( x \) above \( x_0 \).

Figure 1 shows the three-muon mass distribution and the BDT response distribution. Figures 2 and 3 show the distributions of the BDT inputs sorted by the separation rank as reported by TMVA during the BDT training. Figure 4 shows the distributions of the complementary variables which are used in the loose or tight selection but not in the BDT.

### 5.5 Background estimation

The events passing the tight+\( x>x_0 \) selection are used to estimate the expected number of background events in the signal region for higher cuts on \( x \) as described below.

The signal MC and sideband data BDT responses are shown in Fig. 5 after the tight+\( x>x_0 \) selection. The distinct shapes illustrate the power of the method in separating the background from the signal. The analytical function also shown in Fig. 5 is a result of a fit to the sideband data, excluding the blinded region, using an unbinned maximum-likelihood estimator. The fit function used is \( a_0 + a_1(x + 1)^{a_2} + a_3(x + 1)^{a_4} \), where \( a_i \) are the free fit parameters. The parameter \( a_2 \) is required to be negative while the other are required to be non-negative. This function can exhibit rising behaviour at both ends of the \( x \) distribution \( (x \rightarrow \pm 1) \) and it is used to scale the quantities measured in \( x>x_0 \) to the corresponding quantities in \( x>x_1 \) as explained below.

The three-muon mass distribution of the tight+\( x>x_0 \) data is fit simultaneously in the two sidebands to a second-order polynomial in \( m_{3\mu} \) while excluding the blinded region. This is also done with an unbinned maximum-likelihood estimator. The integral of the resulting fit function in the signal region gives the
expected number of background events, $N_b(x_0)$ in the signal region before applying the final $x_1$ cut. The statistical uncertainty of $N_b(x_0)$ is calculated by scaling the statistical error in the number of events in the sidebands, according to the ratio of analytical integrals in the signal region and sidebands. Figure 6 shows the three-muon mass distribution in the sidebands for the $\text{tight} + x > x_0$ selection as black points together with the fit result. The signal is also shown for reference, scaled up arbitrarily to match the scale of the data.

For any $x_1$ cut value above $x \sim 0.6$, where most of the signal is expected, the estimated $N_b(x_0)$ in the signal region can be then scaled down according to the ratio of the integrals of the BDT analytical function above and below this cut. This ratio is denoted hereafter by $R(x_1)$. The extrapolation procedure can be written as $N_b(x_1) = R(x_1) \times N_b(x_0)$ where $N_b(x_1)$ is estimated in the signal region for $x > x_1$.

### 5.6 Uncertainties and optimisation

The sources of systematic uncertainty associated with the extrapolation procedure in the background estimation are the BDT and sideband fit function choice, the definition of the sideband ranges and the definition of $x_0$. To estimate this uncertainty, each of these definitions and choices is varied individually while calculating $N_b(x_0)$ and $R(x_1)$. For each fit function (BDT and sideband), different parameterisations are considered. In addition, to construct the variation of the $\text{tight} + x > x_0$ sample with which the two fits are performed, nine different sideband range variations and ten different $x_0$ variations are used. The fits, and consequently also the extrapolation procedure, are found to be stable against these variations. The dominant uncertainty component is the impact on $R(x_1)$ of varying the sideband ranges definition. The differences from the nominal values of $R(x_1)$ and $N_b(x_0)$ are summed in quadrature and are translated to uncertainties in $N_b(x_1)$. The systematic uncertainty associated with the extrapolation procedure used to obtain $N_b(x_1)$ increases with $x_1$ from $\sim 45\%$ at $x_1 = 0.6$ to $\sim 80\%$ at $x_1 \approx 1$. The statistical uncertainty of $N_b(x_1)$ is $\sim 19\%$, independent of $x_1$. 

![Figure 1: The three-muon mass distribution in (a) and the BDT score in (b). The BDT score distribution of the data is shown for the sideband region. The loose data are shown as hollow circles, while the loose signal MC events are shown as light solid grey area. The tight+ x > x_0 data are shown as the solid black circles, while the tight+ x > x_0 signal MC events are shown as the dark solid grey area. The area of the signal MC shapes is normalised to the area of the loose data shapes and the relative normalisation difference between the loose and the tight+ x > x_0 MC signal distributions prior to the normalisation is maintained. For illustration, the signal is not constrained to the SR.](image-url)
The systematic uncertainty in the signal acceptance times efficiency has contributions from reconstruction (13.1%), trigger (11%) and MC modelling (4%) as discussed in the previous sections. In addition, there is a small (2.1%) contribution due to jet and $E_{T}^{miss}$ calibration. The number of $\tau$ leptons produced via the $W \rightarrow \tau\nu$ channel and its uncertainty (3.9%) are estimated as described in Section 3. These uncertainties are independent of $x_1$ in the range of interest.

The BDT cut is optimised by minimising the expected upper limit on the branching fraction given in Eq. (1), where $N_s$ becomes the upper limit on the number of observed events above the expected background level in a narrow region around the $\tau$ lepton mass. The procedure is performed by varying $x_1$ between 0.6 and 1.0 in steps of 0.001 while extracting $N_b(x_1)$ and its associated errors as explained above. To obtain the upper limit on $N_s$ for each $x_1$ cut, a single-bin counting experiment is performed using the HistFitter [27] statistical framework, supplied with $N_b(x_1)$ and its uncertainties. For compatibility with previous searches, the limit on $N_s$ and on $\text{Br}(\tau \rightarrow 3\mu)$ is given at 90% confidence level (CL). In each iteration, $A_s \times \epsilon_s$ is calculated for the specific $x_1$ cut using a signal sample that is different from the one used for the BDT training.

During the iterative optimisation process, the extrapolation of the number of events in the sideband region to high $x_1$ cuts using the BDT shape is tested against a cut-and-count procedure. The two procedures are found to agree very well within the uncertainties, and the extrapolation procedure gives a more conservative result throughout the examined $x_1$ range. The resulting optimal cut is at $x_1 = 0.933$.

6 Results

Figure 6 shows the three-muon mass distributions in the full mass range, including the blinded region, for the $\text{tight} + x > x_1$ selection in red squares. Only one event with a three-muon mass of 1860 MeV survives the selection in the full mass range (sideband and blinded regions). This event is found in the range between the signal region and the right sideband region and it does not affect the background estimation or the observation in the signal region.

The event counts entering the different regions at the different steps of the analysis for signal and data are given in table 2.

<table>
<thead>
<tr>
<th>Phase</th>
<th>Data SB</th>
<th>Data SR</th>
<th>Signal MC SR [out of $2 \times 10^5$]</th>
</tr>
</thead>
<tbody>
<tr>
<td>loose</td>
<td>2248</td>
<td>580</td>
<td>12672</td>
</tr>
<tr>
<td>loose+$x &gt; x_0$</td>
<td>736</td>
<td>203</td>
<td>12557</td>
</tr>
<tr>
<td>tight</td>
<td>42</td>
<td>9</td>
<td>5503</td>
</tr>
<tr>
<td>tight+$x &gt; x_0$</td>
<td>28</td>
<td>7</td>
<td>5501</td>
</tr>
<tr>
<td>tight+$x &gt; x_1$</td>
<td>0</td>
<td>0</td>
<td>4616</td>
</tr>
</tbody>
</table>

Table 2: The event count for the different steps of the analysis in the sideband and signal regions. The signal sample used to evaluate the $A_s \times \epsilon_s$ has $2 \times 10^5$ events.

The signal acceptance times efficiency is calculated from the second signal MC sample after applying the full $\text{tight} + x > x_1$ selection. This selection corresponds to $A_s \times \epsilon_s = 0.0231 \pm 0.0005_{\text{jet}}/E_{T}^{\text{miss}} \pm 0.0009_{\text{modelling}} \pm 0.0025_{\text{trig}} \pm 0.0030_{\text{reco}}$. With this selection, the expected background yield is $N_b(x_1) = 0.193 \pm 0.131_{\text{syst}} \pm$.
0.037_{\text{stat}}. The systematic uncertainty on $A_s \times \epsilon_s$ is dominated by the uncertainties in the reconstruction and trigger efficiency measurements. The systematic uncertainty on $N_b(x_1)$ is dominated by the uncertainty in the extrapolation of the background from the $\text{tight}+x>x_0$ selection to the $\text{tight}+x>x_1$ selection.

The systematic uncertainties in $N_b$ are taken into account when calculating the limit on the number of signal events, $N_s$, via one nuisance parameter. The systematic uncertainties in the product $(A_s \times \epsilon_s) \cdot N_W \to \tau\nu$ are summed in quadrature and taken into account as the uncertainty in the signal via one nuisance parameter when calculating the limit. The expected (median) limit on the branching fraction for $N_o = N_b(x_1)$ is $3.94 \times 10^{-7}$ at 90% CL. No events are observed in the signal region and the observed limit on the branching fraction is therefore $3.76 \times 10^{-7}$ at 90% CL.

7 Conclusions and outlook

This article presents a search with the ATLAS detector for neutrinoless $\tau \to 3\mu$ decays using 20.3 $\text{fb}^{-1}$ of 2012 LHC $pp$ collision data, utilising $\tau$ leptons produced in $W \to \tau\nu$ decays. No events are observed in the signal region for the final selection while $0.193 \pm 0.131_{\text{syst}} \pm 0.037_{\text{stat}}$ background events are expected. This results in an observed (expected) upper limit of $3.76 \times 10^{-7}$ ($3.94 \times 10^{-7}$) on $\text{Br}(\tau \to 3\mu)$ at 90% CL. Although this limit is not yet competitive with searches performed at $B$-factories [7, 8] and at LHCb [9], it demonstrates the potential of LHC data collected by ATLAS as a probe of lepton flavour violation in $\tau$ lepton decays. This analysis utilises single $\tau$ lepton production in an environment very different from $B$-factories, which rely on $\tau$ lepton pair production in $e^+e^-$ collisions. The method and sample presented here were used to improve the ATLAS muon trigger and reconstruction of low-$p_T$, collimated muons relevant to the $\tau \to 3\mu$ search. The analysis is limited by the number of $W \to \tau\nu$ decays and by the systematic uncertainty, which depends on the size of the data sample. With the much larger data sets anticipated at Run 2 of the LHC, the sensitivity of ATLAS to lepton-flavour-violating decays will be improved significantly.

Acknowledgements

We thank CERN for the very successful operation of the LHC, as well as the support staff from our institutions without whom ATLAS could not be operated efficiently.

We acknowledge the support of ANPCyT, Argentina; YerPhI, Armenia; ARC, Australia; BMWFW and FWF, Austria; ANAS, Azerbaijan; SSTC, Belarus; CNPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN; CONICYT, Chile; CAS, MOST and NSFC, China; COLCIENCIAS, Colombia; MSMT CR, MPO CR and VSC CR, Czech Republic; DNRF and DNSRC, Denmark; IN2P3-CNRS, CEA-DSM/IRFU, France; GNSF, Georgia; BMBF, HGF, and MPG, Germany; GSRT, Greece; RGC, Hong Kong SAR, China; ISF, I-CORE and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; FOM and NWO, Netherlands; RCN, Norway; MNiSW and NCN, Poland; FCT, Portugal; MNE/IFA, Romania; MES of Russia and NRC KI, Russian Federation; JINR; MESTD, Serbia; MSSR, Slovakia; ARRS and MIZŠ, Slovenia; DST/NRF, South Africa; MINECO, Spain; SRC and Wallenberg Foundation, Sweden; SERI, SNSF and Cantons of Bern and Geneva, Switzerland; MOST, Taiwan; TAEK, Turkey; STFC, United Kingdom; DOE and NSF, United States of America. In addition, individual groups and members have received support from BCKDF, the Canada Council, CANARIE,
CRC, Compute Canada, FQRNT, and the Ontario Innovation Trust, Canada; EPLANET, ERC, FP7, Horizion 2020 and Marie Skłodowska-Curie Actions, European Union; Investissements d’Avenir Labex and Idex, ANR, Région Auvergne and Fondation Partager le Savoir, France; DFG and AvH Foundation, Germany; Herakleitos, Thales and Aristeia programmes co-financed by EU-ESF and the Greek NSRF; BSF, GIF and Minerva, Israel; BRF, Norway; the Royal Society and Leverhulme Trust, United Kingdom.

The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN and the ATLAS Tier-1 facilities at TRIUMF (Canada), NDGF (Denmark, Norway, Sweden), CC-IN2P3 (France), KIT/GridKA (Germany), INFN-CNAF (Italy), NL-T1 (Netherlands), PIC (Spain), ASGC (Taiwan), RAL (UK) and BNL (USA) and in the Tier-2 facilities worldwide.
For illustration, the signal is not constrained to the SR. The area of the signal MC shapes is normalised to the area of the tight shown as the solid black circles, while the loose signal MC events are shown as light solid grey area. The reference between the signal MC events are shown as the dark solid grey area. The \( m_T^1 \) data in the sidebands are shown as hollow circles, while the loose signal MC events are shown as light solid grey area. The tight+\( x\times x \) data in the sidebands are shown as the solid black circles, while the tight+\( x\times x \) signal MC events are shown as the dark solid grey area. The area of the signal MC shapes is normalised to the area of the loose data shapes and the relative normalisation difference between the loose and the tight+\( x\times x \) MC signal distributions prior to the normalisation is maintained. For illustration, the signal is not constrained to the SR.
Figure 3: The BDT inputs ranked 9–16. $\Delta \phi^{\text{3d}}$ in (a), $p$-value in (b), $S \,(a_{0}^{\text{3d}})$ in (c) $p_{\text{trk}}^{3\mu}$ in (d), $E_{\text{T}}^{\text{miss}} \,(100)$ in (e), $N^{\text{trk}} \,(100)$ in (f), $S(L_{3\mu})$ in (g) and $p_{\text{trk}}^{3\mu} / E_{\text{T}}^{\text{miss}} - 1$ in (h). The loose data in the sidebands are shown as hollow circles, while the loose signal MC events are shown as light solid grey area. The tight data in the sidebands are shown as the solid black circles, while the tight signal MC events are shown as the dark solid grey area. The area of the signal MC shapes is normalised to the area of the loose data shapes and the relative normalisation difference between the loose and the tight signal MC signal distributions prior to the normalisation is maintained. For illustration, the signal is not constrained to the SR.
Figure 4: The complementary variables used in the loose or tight selection but not as inputs for the BDT. \( \Delta \phi_{\text{cal}}^{\text{SS}} \) in (a), \( \Delta \phi_{\text{trk}}^{\text{SS}} \) in (b), \( \Sigma p_{T}^{\text{trk}}(\Delta R_{\text{max}}^{\text{SS}} + 0.30)/p_{T}^{\text{SS}} \) in (c), \( m_{\text{SS}} \) in (d), \( m_{\text{OS1}} \) in (e) and \( m_{\text{OS2}} \) in (f). The loose data in the sidebands are shown as hollow circles, while the loose signal MC events are shown as light solid grey area. The tight+\( x_{0} \) data in the sidebands are shown as the solid black circles, while the tight+\( x_{0} \) signal MC events are shown as the dark solid grey area. The area of the signal MC shapes is normalised to the area of the loose data shapes and the relative normalisation difference between the loose and the tight+\( x_{0} \) MC signal distributions prior to the normalisation is maintained. For illustration, the signal is not constrained to the SR.
Figure 5: The distribution of the BDT score of the data in the sideband region (SB) for the $\text{tight}+x>x_0$ selection. The line shows the result of a fit to the BDT score distribution, while the hatched area shows the uncertainty in the fit due to the SB range definition, the $x_0$ cut location and the fit function choice. The solid grey area shows the signal shape (obtained from MC simulation), normalised to the area of the data for the $\text{tight}+x>x_0$ selection. For illustration, the signal is not constrained to the SR.
Figure 6: The three-muon mass distribution in the range [1450, 2110] MeV shown for the $\text{tight}+x>x_0$ selection by solid black circles and for the $\text{tight}+x>x_1$ selection by the solid red square. The sideband and signal regions are indicated by the arrows. The $\text{tight}+x>x_0$ data are fit in the two sidebands simultaneously, excluding the events in the blinded region. The hatched area shows the uncertainty in the fit due to the SB range definition, the $x_0$ cut location and the fit function choice. The solid grey area shows the signal shape (obtained from MC simulation), normalised to the area of the data for the $\text{tight}+x>x_0$ selection.
References


Division of Physics, TOBB University of Economics and Technology, Ankara, Turkey
5 LAPP, CNRS/IN2P3 and Université Savoie Mont Blanc, Annecy-le-Vieux, France
6 High Energy Physics Division, Argonne National Laboratory, Argonne IL, United States of America
7 Department of Physics, University of Arizona, Tucson AZ, United States of America
8 Department of Physics, The University of Texas at Arlington, Arlington TX, United States of America
9 Physics Department, University of Athens, Athens, Greece
10 Physics Department, National Technical University of Athens, Zografou, Greece
11 Institute of Physics, Azerbaijan Academy of Sciences, Baku, Azerbaijan
12 Institut de Física d’Altes Energies (IAFE), The Barcelona Institute of Science and Technology, Barcelona, Spain, Spain
13 Institute of Physics, University of Belgrade, Belgrade, Serbia
14 Department for Physics and Technology, University of Bergen, Bergen, Norway
15 Physics Division, Lawrence Berkeley National Laboratory and University of California, Berkeley CA, United States of America
16 Department of Physics, Humboldt University, Berlin, Germany
17 Albert Einstein Center for Fundamental Physics and Laboratory for High Energy Physics, University of Bern, Bern, Switzerland
18 School of Physics and Astronomy, University of Birmingham, Birmingham, United Kingdom
19 (a) Department of Physics, Bogazici University, Istanbul; (b) Department of Physics Engineering, Gaziantep University, Gaziantep; (c) Department of Physics, Dogus University, Istanbul, Turkey
20 Centro de Investigaciones, Universidad Antonio Narino, Bogota, Colombia
21 (a) INFN Sezione di Bologna; (b) Dipartimento di Fisica e Astronomia, Università di Bologna, Bologna, Italy
22 Physikalisches Institut, University of Bonn, Bonn, Germany
23 Department of Physics, Boston University, Boston MA, United States of America
24 Department of Physics, Brandeis University, Waltham MA, United States of America
25 (a) Universidade Federal do Rio De Janeiro COPPE/EE/IF, Rio de Janeiro; (b) Electrical Circuits Department, Federal University of Juiz de Fora (UFJF), Juiz de Fora; (c) Federal University of Sao Joao del Rei (UFSJ), Sao Joao del Rei; (d) Instituto de Física, Universidade de Sao Paulo, Sao Paulo, Brazil
26 Physics Department, Brookhaven National Laboratory, Upton NY, United States of America
27 (a) Transilvania University of Brasov, Brasov, Romania; (b) National Institute of Physics and Nuclear Engineering, Bucharest; (c) National Institute for Research and Development of Isotopic and Molecular Technologies, Physics Department, Cluj Napoca; (d) University Politehnica Bucharest, Bucharest; (e) West University in Timisoara, Timisoara, Romania
28 Departamento de Física, Universidad de Buenos Aires, Buenos Aires, Argentina
29 Cavendish Laboratory, University of Cambridge, Cambridge, United Kingdom
30 Department of Physics, Carleton University, Ottawa ON, Canada
31 CERN, Geneva, Switzerland
32 Enrico Fermi Institute, University of Chicago, Chicago IL, United States of America
33 (a) Departamento de Física, Pontificia Universidad Católica de Chile, Santiago; (b) Departamento de Física, Universidad Técnica Federico Santa María, Valparaíso, Chile
34 (a) Institute of High Energy Physics, Chinese Academy of Sciences, Beijing; (b) Department of Modern Physics, University of Science and Technology of China, Anhui; (c) Department of Physics, Nanjing University, Jiangsu; (d) School of Physics, Shandong University, Shandong; (e) Department of Physics and Astronomy, Shanghai Key Laboratory for Particle Physics and Cosmology, Shanghai Jiao Tong University, Shanghai; (also affiliated with PKU-CHEP); (f) Physics Department, Tsinghua University, Beijing 100084, China
35 Laboratoire de Physique Corpusculaire, Clermont Université and Université Blaise Pascal and CNRS/IN2P3, Clermont-Ferrand, France
36 Nevis Laboratory, Columbia University, Irvington NY, United States of America
37 Niels Bohr Institute, University of Copenhagen, Kobenhavn, Denmark
38 (a) INFN Gruppo Collegato di Cosenza, Laboratori Nazionali di Frascati; (b) Dipartimento di Fisica, Università della Calabria, Rende, Italy
39 (a) AGH University of Science and Technology, Faculty of Physics and Applied Computer Science, Krakow; (b) Marian Smoluchowski Institute of Physics, Jagiellonian University, Krakow, Poland
40 Institute of Nuclear Physics, Polish Academy of Sciences, Krakow, Poland
41 Physics Department, Southern Methodist University, Dallas TX, United States of America
42 Physics Department, University of Texas at Dallas, Richardson TX, United States of America
43 DESY, Hamburg and Zeuthen, Germany
44 Institut für Experimentelle Physik IV, Technische Universität Dortmund, Dortmund, Germany
45 Institut für Kern- und Teilchenphysik, Technische Universität Dresden, Dresden, Germany
46 Department of Physics, Duke University, Durham NC, United States of America
47 SUPA - School of Physics and Astronomy, University of Edinburgh, Edinburgh, United Kingdom
48 INFN Laboratori Nazionali di Frascati, Frascati, Italy
49 Fakultät für Mathematik und Physik, Albert-Ludwigs-Universität, Freiburg, Germany
50 Section de Physique, Université de Genève, Geneva, Switzerland
51 (a) INFN Sezione di Genova; (b) Dipartimento di Fisica, Università di Genova, Genova, Italy
52 (a) E. Andronikashvili Institute of Physics, Iv. Javakhishvili Tbilisi State University, Tbilisi; (b) High Energy Physics Institute, Tbilisi, Georgia
53 II Physikalisches Institut, Justus-Liebig-Universität Giessen, Giessen, Germany
54 SUPA - School of Physics and Astronomy, University of Glasgow, Glasgow, United Kingdom
55 II Physikalisches Institut, Georg-August-Universität, Göttingen, Germany
56 Laboratoire de Physique Subatomique et de Cosmologie, Université Grenoble-Alpes, CNRS/IN2P3, Grenoble, France
57 Department of Physics, Hampton University, Hampton VA, United States of America
58 Laboratory for Particle Physics and Cosmology, Harvard University, Cambridge MA, United States of America
59 (a) Kirchhoff-Institut für Physik, Ruprecht-Karls-Universität Heidelberg, Heidelberg; (b) Physikalisches Institut, Ruprecht-Karls-Universität Heidelberg, Heidelberg; (c) ZITI Institut für technische Informatik, Ruprecht-Karls-Universität Heidelberg, Mannheim, Germany
60 Faculty of Applied Information Science, Hiroshima Institute of Technology, Hiroshima, Japan
61 (a) Department of Physics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong; (b) Department of Physics, The University of Hong Kong, Hong Kong; (c) Department of Physics, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China
62 Department of Physics, Indiana University, Bloomington IN, United States of America
63 Institut für Astro- und Teilchenphysik, Leopold-Franzens-Universität, Innsbruck, Austria
64 University of Iowa, Iowa City IA, United States of America
65 Department of Physics and Astronomy, Iowa State University, Ames IA, United States of America
66 Joint Institute for Nuclear Research, JINR Dubna, Dubna, Russia
67 KEK, High Energy Accelerator Research Organization, Tsukuba, Japan
68 Graduate School of Science, Kobe University, Kobe, Japan
69 Faculty of Science, Kyoto University, Kyoto, Japan
70 Kyoto University of Education, Kyoto, Japan
71 Department of Physics, Kyushu University, Fukuoka, Japan
Instituto de Física La Plata, Universidad Nacional de La Plata and CONICET, La Plata, Argentina

Physics Department, Lancaster University, Lancaster, United Kingdom

(a) INFN Sezione di Leccce; (b) Dipartimento di Matematica e Fisica, Università del Salento, Lecce, Italy

Oliver Lodge Laboratory, University of Liverpool, Liverpool, United Kingdom

Department of Physics, Jožef Stefan Institute and University of Ljubljana, Ljubljana, Slovenia

School of Physics and Astronomy, Queen Mary University of London, London, United Kingdom

Department of Physics, Royal Holloway University of London, Surrey, United Kingdom

Department of Physics and Astronomy, University College London, London, United Kingdom

Louisiana Tech University, Ruston LA, United States of America

Laboratoire de Physique Nucléaire et de Hautes Energies, UPMC and Université Paris-Diderot and CNRS/IN2P3, Paris, France

Department of Physics, Jožef Stefan Institute and University of Ljubljana, Ljubljana, Slovenia

School of Physics and Astronomy, Queen Mary University of London, London, United Kingdom

Department of Physics, Royal Holloway University of London, Surrey, United Kingdom

Department of Physics and Astronomy, University College London, London, United Kingdom

Louisiana Tech University, Ruston LA, United States of America

Laboratoire de Physique Nucléaire et de Hautes Energies, UPMC and Université Paris-Diderot and CNRS/IN2P3, Paris, France

Fysiska institutionen, Lunds universitet, Lund, Sweden

Departamento de Física Teorica C-15, Universidad Autonoma de Madrid, Madrid, Spain

Institut für Physik, Universität Mainz, Mainz, Germany

School of Physics and Astronomy, University of Manchester, Manchester, United Kingdom

CPPM, Aix-Marseille Université and CNRS/IN2P3, Marseille, France

Department of Physics, University of Massachusetts, Amherst MA, United States of America

Department of Physics, McGill University, Montreal QC, Canada

School of Physics, University of Melbourne, Victoria, Australia

Department of Physics, The University of Michigan, Ann Arbor MI, United States of America

Department of Physics and Astronomy, Michigan State University, East Lansing MI, United States of America

(a) INFN Sezione di Milano; (b) Dipartimento di Fisica, Università di Milano, Milano, Italy

B.I. Stepanov Institute of Physics, National Academy of Sciences of Belarus, Minsk, Republic of Belarus

National Scientific and Educational Centre for Particle and High Energy Physics, Minsk, Republic of Belarus

Department of Physics, Massachusetts Institute of Technology, Cambridge MA, United States of America

Group of Particle Physics, University of Montreal, Montreal QC, Canada

P.N. Lebedev Physical Institute of the Russian Academy of Sciences, Moscow, Russia

Institute for Theoretical and Experimental Physics (ITEP), Moscow, Russia

National Research Nuclear University MEPhI, Moscow, Russia

D.V. Skobeltsyn Institute of Nuclear Physics, M.V. Lomonosov Moscow State University, Moscow, Russia

Fakultät für Physik, Ludwig-Maximilians-Universität München, München, Germany

Max-Planck-Institut für Physik (Werner-Heisenberg-Institut), München, Germany

Nagasaki Institute of Applied Science, Nagasaki, Japan

Graduate School of Science and Kobayashi-Maskawa Institute, Nagoya University, Nagoya, Japan

(a) INFN Sezione di Napoli; (b) Dipartimento di Fisica, Università di Napoli, Napoli, Italy

Department of Physics and Astronomy, University of New Mexico, Albuquerque NM, United States of America

Institute for Mathematics, Astrophysics and Particle Physics, Radboud University Nijmegen/Nikhef, Nijmegen, Netherlands

Nikhef National Institute for Subatomic Physics and University of Amsterdam, Amsterdam, Netherlands
Department of Physics and Astronomy, University of Sheffield, Sheffield, United Kingdom
Department of Physics, Shinshu University, Nagano, Japan
Fachbereich Physik, Universität Siegen, Siegen, Germany
Department of Physics, Simon Fraser University, Burnaby BC, Canada
SLAC National Accelerator Laboratory, Stanford CA, United States of America

(a) Faculty of Mathematics, Physics & Informatics, Comenius University, Bratislava; (b) Department of Subnuclear Physics, Institute of Experimental Physics of the Slovak Academy of Sciences, Kosice, Slovak Republic

(a) Department of Physics, University of Cape Town, Cape Town; (b) Department of Physics, University of Johannesburg, Johannesburg; (c) School of Physics, University of the Witwatersrand, Johannesburg, South Africa

Department of Physics, Stockholm University; (b) The Oskar Klein Centre, Stockholm, Sweden
Physics Department, Royal Institute of Technology, Stockholm, Sweden
Departments of Physics & Astronomy and Chemistry, Stony Brook University, Stony Brook NY, United States of America

Department of Physics and Astronomy, University of Sussex, Brighton, United Kingdom
School of Physics, University of Sydney, Sydney, Australia
Institute of Physics, Academia Sinica, Taipei, Taiwan
Department of Physics, Technion: Israel Institute of Technology, Haifa, Israel
Raymond and Beverly Sackler School of Physics and Astronomy, Tel Aviv University, Tel Aviv, Israel

Department of Physics, Aristotle University of Thessaloniki, Thessaloniki, Greece
International Center for Elementary Particle Physics and Department of Physics, The University of Tokyo, Tokyo, Japan
Graduate School of Science and Technology, Tokyo Metropolitan University, Tokyo, Japan
Department of Physics, Tokyo Institute of Technology, Tokyo, Japan
Department of Physics, University of Toronto, Toronto ON, Canada

(a) TRIUMF, Vancouver BC; (b) Department of Physics and Astronomy, York University, Toronto ON, Canada
Faculty of Pure and Applied Sciences, and Center for Integrated Research in Fundamental Science and Engineering, University of Tsukuba, Tsukuba, Japan
Department of Physics and Astronomy, Tufts University, Medford MA, United States of America
Department of Physics and Astronomy, University of California Irvine, Irvine CA, United States of America

(a) INFN Gruppo Collegato di Udine, Sezione di Trieste, Udine; (b) ICTP, Trieste; (c) Dipartimento di Chimica, Fisica e Ambiente, Università di Udine, Udine, Italy
Department of Physics and Astronomy, University of Uppsala, Uppsala, Sweden
Department of Physics, University of Illinois, Urbana IL, United States of America
Instituto de Física Corpuscular (IFIC) and Departamento de Física Atómica, Molecular y Nuclear and Departamento de Ingeniería Electrónica and Instituto de Microelectrónica de Barcelona (IMB–CNM), University of Valencia and CSIC, Valencia, Spain
Department of Physics, University of British Columbia, Vancouver BC, Canada
Department of Physics and Astronomy, University of Victoria, Victoria BC, Canada
Department of Physics, University of Warwick, Coventry, United Kingdom
Waseda University, Tokyo, Japan
Department of Particle Physics, The Weizmann Institute of Science, Rehovot, Israel
Department of Physics, University of Wisconsin, Madison WI, United States of America
Fakultät für Physik und Astronomie, Julius-Maximilians-Universität, Würzburg, Germany
Fakultät für Mathematik und Naturwissenschaften, Fachgruppe Physik, Bergische Universität Wuppertal, Wuppertal, Germany
Department of Physics, Yale University, New Haven CT, United States of America
Yerevan Physics Institute, Yerevan, Armenia
Centre de Calcul de l’Institut National de Physique Nucléaire et de Physique des Particules (IN2P3), Villeurbanne, France
a Also at Department of Physics, King’s College London, London, United Kingdom
b Also at Institute of Physics, Azerbaijan Academy of Sciences, Baku, Azerbaijan
c Also at Novosibirsk State University, Novosibirsk, Russia
d Also at TRIUMF, Vancouver BC, Canada
e Also at Department of Physics, California State University, Fresno CA, United States of America
f Also at Department of Physics, University of Fribourg, Fribourg, Switzerland
g Also at Departament de Física de la Universitat Autònoma de Barcelona, Barcelona, Spain
h Also at Departamento de Física e Astronomia, Faculdade de Ciencias, Universidade do Porto, Portugal
i Also at Tomsk State University, Tomsk, Russia
j Also at CPPM, Aix-Marseille Université and CNRS/IN2P3, Marseille, France
k Also at Università di Napoli Parthenope, Napoli, Italy
l Also at Institute of Particle Physics (IPP), Canada
Also at Particle Physics Department, Rutherford Appleton Laboratory, Didcot, United Kingdom
Also at Department of Physics, St. Petersburg State Polytechnical University, St. Petersburg, Russia
Also at Louisiana Tech University, Ruston LA, United States of America
Also at Institutio Catalana de Recerca i Estudis Avancats, ICREA, Barcelona, Spain
Also at Department of Physics, The University of Michigan, Ann Arbor MI, United States of America
Also at Graduate School of Science, Osaka University, Osaka, Japan
Also at Department of Physics, National Tsing Hua University, Taiwan
Also at Department of Physics, The University of Texas at Austin, Austin TX, United States of America
Also at Institute of Theoretical Physics, Ilia State University, Tbilisi, Georgia
Also at CERN, Geneva, Switzerland
Also at Georgian Technical University (GTU), Tbilisi, Georgia
Also at Ochadai Academic Production, Ochanomizu University, Tokyo, Japan
Also at Manhattan College, New York NY, United States of America
Also at Hellenic Open University, Patras, Greece
Also at Institute of Physics, Academia Sinica, Taipei, Taiwan
Also at LAL, Univ. Paris-Sud, CNRS/IN2P3, Université Paris-Saclay, Orsay, France
Also at Academia Sinica Grid Computing, Institute of Physics, Academia Sinica, Taipei, Taiwan
Also at School of Physics, Shandong University, Shandong, China
Also at Moscow Institute of Physics and Technology State University, Dolgoprudny, Russia
Also at Section de Physique, Université de Genève, Geneva, Switzerland
Also at International School for Advanced Studies (SISSA), Trieste, Italy
Also at Department of Physics and Astronomy, University of South Carolina, Columbia SC, United States of America
Also at School of Physics and Engineering, Sun Yat-sen University, Guangzhou, China
Also at Institute for Nuclear Research and Nuclear Energy (INRNE) of the Bulgarian Academy of Sciences, Sofia, Bulgaria
Also at Faculty of Physics, M.V.Lomonosov Moscow State University, Moscow, Russia
Also at National Research Nuclear University MEPhI, Moscow, Russia
am Also at Department of Physics, Stanford University, Stanford CA, United States of America
an Also at Institute for Particle and Nuclear Physics, Wigner Research Centre for Physics, Budapest, Hungary
ao Also at Flensburg University of Applied Sciences, Flensburg, Germany
ap Also at University of Malaya, Department of Physics, Kuala Lumpur, Malaysia
* Deceased