Improved description of the di-tau final state in events with associated production of a W boson and jets in the ATLAS detector using the tau-promotion method

The ATLAS Collaboration

The production of W bosons and jets in the ATLAS detector is the main irreducible contribution in many analyses that select events with two reconstructed hadronically decaying tau leptons. The number of simulated events that are used to model this background is limited by the CPU resources available for the Monte Carlo simulation. In searches for new physics with two tau leptons in the final state, this limitation may have a deleterious effect on the estimation of the background and ultimately on the sensitivity of the analysis to new physics. This note presents a method to reduce the statistical uncertainties from Monte Carlo simulation of the associated production of a W boson together with jets for searches that contain two hadronically decaying tau leptons in the final state. The results of the method are presented in an example case, with a comparison to the nominal predictions from simulation.

1 Introduction

In searches for final states that contain two hadronically decaying tau leptons,[1] the majority of the selected $W(\tau\nu)$ + jets events that contribute to the background have one true tau lepton from the W-boson decay and at least one object which is misidentified as a tau lepton (fake tau lepton). As the misidentification rate (or fake rate for short) of a jet as a tau lepton is small, typically of the order of a few per mille to percent, most of the $W(\tau\nu)$ + jets events are discarded already when requiring the presence of two identified tau leptons. This fact helps to suppress the otherwise dominant $W(\tau\nu)$ + jets background, but also leads to large statistical fluctuations in the estimated event yields due to the small remaining number of simulated events that can be used for the background modeling. The idea of the tau-promotion method is to recover the discarded simulated events by artificially promoting an additional jet in each event to a fake tau lepton, and with an appropriate reweighting to reduce the statistical uncertainties on the $W(\tau\nu)$ + jets background estimation. Here, the $W(\tau\nu)$ + jets background refers to W-boson production, possibly in association with additional partons, where the W boson decays to a tau lepton and a tau neutrino.

The tau-promotion method is used in the ATLAS search for direct tau-slepton production in events with two hadronic tau leptons using the full 2015 – 2018 dataset of pp collisions at $\sqrt{s} = 13$ TeV [1] with an integrated luminosity of 139 fb^{-1}. This implementation serves as an example of the method in the following, but the method can be generally used in searches for final states that contain two hadronically decaying tau leptons. Section 2 briefly describes the simulated $W(\tau\nu)$ + jets samples, Section 3 the reconstruction and definition of physics objects, and Section 4 the signal-region selections from Ref. [1]. In Section 5, the idea and the implementation of the tau-promotion method itself are discussed. Section 6 presents the determination of the fake rates and the results from the tau-promotion method, comparing them to estimates obtained with the nominal simulated samples.

2 Simulated event samples

In this note, only Monte Carlo (MC) simulated samples of $W(\tau\nu)$ + jets events are used. The samples are the same as those used in the example case of the ATLAS search for direct tau-slepton production in events with two hadronic tau leptons. This section provides details on these samples for reference, but for the tau-promotion method itself the details are not of primary importance.

The samples were produced in exclusive ranges of the variable $\max[H_{\mathrm{T}}, p_{\mathrm{T}}(V)]$ at the parton level, where $p_{\mathrm{T}}(V)$ is the boson transverse momentum and H_{T} is the scalar p_{T} sum of all parton-level jets with $p_{\mathrm{T}} > 20$ GeV [2]. They are further split by a filter on the presence of light, c- or b-hadrons at the particle level. The total number of generated events in the samples covering the lowest range of $\max[H_{\mathrm{T}}, p_{\mathrm{T}}(V)]$, which has the largest cross section, corresponds to an effective luminosity of only around 6 fb^{-1}, which is much smaller than the integrated luminosity of the 2015 – 2018 dataset.

The samples were generated at next-to-leading order (NLO) in the strong coupling constant with SHERPA 2.2.1 [3, 4]. Matrix elements (ME) were calculated for up to two additional partons at NLO and four additional partons at leading order (LO), using the Comix [5] and OpenLoops [6, 7] generators and merged with the SHERPA parton shower (PS) [8] using the ME+PS@NLO prescription [4]. The NNPDF3.0NNLO [9] parton distribution function (PDF) set was used in conjunction with a dedicated PS tuning developed by the

---

[1] In this note, we are only concerned with hadronically decaying tau leptons unless otherwise stated, as electrons and muons from leptonic tau decays are not distinguished from prompt leptons in the reconstruction.

Sherpa authors. The expected yields were normalised using their next-to-next-to-leading order (NNLO) cross sections [10].

The generated events were passed through a detailed detector simulation [11] based on Geant4 [12]. They were overlaid with multiple pp collisions (pile-up) simulated with the soft strong interaction processes of Pythia 8.186 [13] using the A3 set of tuned underlying event and hadronization parameters [14] and the NNPDF2.3LO [15] PDF set. The distribution of the number of concurrent pp collisions in the simulation is similar to the Run-2 data-taking conditions.

3 Object reconstruction and definition

Events with at least one reconstructed primary vertex [16] are selected. A primary vertex must have at least two associated charged-particle tracks with transverse momentum $p_T > 500$ MeV and be consistent with the collision region. In events with multiple primary vertices, the one with the largest $\sum p_T^2$ of the associated tracks is chosen.

Jets are reconstructed from three-dimensional calorimeter energy clusters [17] using the anti-k_t algorithm [18, 19] with a radius parameter of 0.4 and further calibrated as described in Ref. [20]. Events are vetoed when containing jets induced by calorimeter noise or non-collision background, according to criteria similar to those described in Ref. [21]. Only jets with $p_T > 20$ GeV and $|\eta| < 2.8$ are further considered.[2] An additional requirement on the output of a jet-vertex tagger allows the exclusion of jets produced in pile-up processes [22]. Jets containing b-hadrons (b-jets) are identified using a working point of the MV2c10 algorithm [23] with 70 % efficiency.

Electrons are required to have $p_T > 17$ GeV, $|\eta| < 2.47$, and to satisfy the 'loose' working point according to a likelihood-based identification [24]. Muons are required to have $p_T > 17$ GeV and $|\eta| < 2.7$ and fulfil the 'medium' quality criteria of Ref. [25]. Events containing a muon candidate with a poorly measured charge-to-momentum ratio ($\sigma(q/p) / |q/p| > 0.2$) are rejected.

Based on information from tracking in the ID and three-dimensional clusters of energy in the electromagnetic and hadronic calorimeters, hadronically decaying tau leptons are reconstructed [26, 27]. The tau-lepton reconstruction algorithm is seeded by jets with $p_T > 10$ GeV and $|\eta| < 2.5$. The reconstructed energies of the hadronically decaying tau-lepton candidates are corrected to the tau energy scale [28]. Tau neutrinos from hadronic tau-lepton decays are not taken into account in the reconstruction and calibration of the tau energy and momentum. Hadronic tau-decay candidates are required to have $p_T > 20$ GeV and not considered if they are in the transition region between the barrel and end-cap calorimeters ($1.37 < |\eta| < 1.52$). They are required to have one or three associated charged-particle tracks (prongs) and the total electric charge of those tracks must be ± 1 times the electron charge. Tau-lepton candidates fulfilling these basic criteria will be referred to simply as *reconstructed tau leptons*. To better discriminate hadronically decaying tau leptons from jets, a multivariate algorithm (JetBDT) based on various track and cluster variables as input is used [28]. For 1-prong (3-prong) tau-lepton candidates, the identification efficiencies for tau leptons are 75 % (60 %) and 60 % (45 %) for the 'medium' and 'tight' working points, respectively. For electron discrimination, tau-lepton candidates associated with a single track that overlap with a good reconstructed

---

[2] 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, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the z-axis. The pseudorapidity is defined in terms of the polar angle θ as $\eta = -\ln\tan(\theta/2)$.

Table 1: Selection requirements of the two signal regions used in the ATLAS search for direct tau-slepton production [1].

SR-lowMass	SR-highMass
2 tight τs (OS)	2 medium τs (OS), ≥ 1 tight τ
asymmetric di-tau trigger	di-tau + E_T^{miss} trigger
$75 < E_T^{\mathrm{miss}} < 150$ GeV	$E_T^{\mathrm{miss}} > 150$ GeV

tau-lepton p_T and E_T^{miss} cuts (details are given in the text)
e/μ veto and 3rd medium τ veto
b-jet veto
Z/H veto ($m(\tau_1, \tau_2) > 120$ GeV)
$\Delta R(\tau_1, \tau_2) < 3.2$
$|\Delta\phi(\tau_1, \tau_2)| > 0.8$
$m_{\mathrm{T2}} > 70$ GeV

electron are removed. This requirement has about 95 % efficiency, and a rejection factor from 10 to 50 depending on the η range. Tau-lepton candidates fulfilling these criteria in addition will be referred to as (medium or tight) *signal tau leptons*. Signal tau leptons are a subset of the reconstructed tau leptons. Note that both reconstructed and signal tau leptons can either be true tau leptons (i. e. correctly identified tau leptons) or fake tau leptons (i. e. an object that is not a tau lepton, most often a jet, falsely identified as a tau lepton).

The missing transverse momentum (and its magnitude E_T^{miss}) is defined as the negative vector sum of the transverse momenta of all identified objects and an additional soft term. The soft term is constructed from all tracks associated with the primary vertex but not with any physics object. In this way, the E_T^{miss} is adjusted for the best calibration of the jets and the other identified physics objects listed above, while maintaining approximate pile-up independence in the soft term [29, 30]. Overlaps between objects in the E_T^{miss} calculation are resolved as described in Ref. [29].

The possible double counting of reconstructed objects is resolved in the following order. Tau candidates close to electron or muon candidates ($\Delta R < 0.2$) are removed, as are electrons that share a track with a muon. ΔR is a measure of the spatial separation of two objects in the detector and defined as $\Delta R = \sqrt{(\Delta y)^2 + (\Delta \phi)^2}$, where y is the pseudorapidity and ϕ the azimuthal angle. For electrons close to a jet ($\Delta R < 0.4$), the electron is removed, except when $\Delta R < 0.2$ and the jet is not b-tagged, in which case the jet is removed. For a muon close to a jet ($\Delta R < 0.4$), the muon is removed unless the jet has less than three tracks associated to it and is within $\Delta R = 0.2$. In the latter case, the jet is removed. Any remaining jet within $\Delta R = 0.2$ of a tau candidate is removed.

Differences between data and MC simulation are determined to be small in independent measurements and are corrected for using scale factors.

4 Event selection

Since the tau-promotion method was initially developed to improve the statistical uncertainty of $W(\tau\nu)$ + jets in the search for the direct production of a scalar tau-lepton pair, in this note an event selection is used

that is the same as that of the search for direct tau-slepton production in events with two hadronic tau leptons in $\sqrt{s}$ = 13 TeV pp collisions with the ATLAS detector [1]. Two signal regions (SRs) are defined in this search: SR-lowMass for the low-$E_\mathrm{T}^\mathrm{miss}$ regime ($75 < E_\mathrm{T}^\mathrm{miss} < 150$ GeV) and SR-highMass for the high-$E_\mathrm{T}^\mathrm{miss}$ regime ($E_\mathrm{T}^\mathrm{miss} > 150$ GeV). SR-lowMass uses an *asymmetric di-tau* trigger; SR-highMass uses a combined *di-tau* + E_T^{miss} trigger. Requirements on the leading tau-lepton candidate of $p_\mathrm{T}(\tau_1) > 95$ GeV ($50 - 75$ GeV) and the next-to-leading tau-lepton candidate $p_\mathrm{T}(\tau_2) > 65 - 75$ GeV (40 GeV) ensure that the asymmetric di-tau (di-tau + $E_\mathrm{T}^\mathrm{miss}$) trigger is fully efficient for the final event selection.

The reconstructed mass of the two leading tau-lepton candidates $m(\tau_1, \tau_2)$, the stransverse mass m_T2 [31,32], and angular variables are also used in the definition of the signal regions as listed in Table 1.

5 Description of the Tau-Promotion Method

In this section, the idea and the implementation of the tau-promotion method are described [33]. Most of the $W(\tau\nu)$ + jets events that contribute to the background in a selection of events with (at least) two signal tau leptons have one correctly identified true tau lepton from the W-boson decay and at least one fake tau lepton. As the fake rate is small, typically of the order of a few permille to percent, most of the simulated $W(\tau\nu)$ + jets background events are discarded already when requiring the presence of two signal tau leptons. The idea of the tau-promotion method is to recover the discarded events by picking one jet in every simulated event at random, artificially promoting it to a (fake) tau lepton, and then restoring the correct prediction of the $W(\tau\nu)$ + jets background yield by reweighting the event sample to the yield of the nominal sample. In this way, the statistical uncertainty of the $W(\tau\nu)$ + jets simulated samples can be reduced by a factor that is roughly given by $1/\sqrt{f}$, where $f \ll 1$ is the average fake rate.

A schematic overview of the procedure is shown in Figure 1. In this picture, every box symbolizes one simulated event, the size of the box indicating the event weight which is determined mainly by the number of simulated events, the Standard Model cross section of the $W(\tau\nu)$ + jets process and the assumed integrated luminosity. Events passing a two-tau-lepton selection are shaded in light green. The predicted event yield after reweighting, shown by the bottom-most shaded box, is estimated using a much larger number of simulated events than before. Each of the simulated events now has a smaller weight, thus yielding a much reduced statistical uncertainty.

In this implementation of the method, it is assumed that true tau leptons are very likely to be reconstructed and identified as signal tau leptons, so that the presence of only one additional fake tau lepton is enough to have at least two signal tau leptons in most events. In reality, the probability for true tau leptons to be reconstructed and identified as signal tau leptons is $45 - 75$ %, so the number of true tau leptons that are not identified as signal tau leptons is non-negligible. The method could be extended to also recover those events where the true tau lepton from the decay of the W boson is not identified as a signal tau lepton, but this is not done here. This means that true tau leptons must explicitly be excluded from the tau-promotion method because the probability of true tau leptons to be identified as signal tau leptons is by construction much larger than those of other objects such as jets, so that promoting a true tau lepton that has not been identified as signal tau lepton and applying the fake rates for jets would yield wrong results. Therefore, a truth matching based on the spatial separation between reconstructed tau leptons and true tau leptons is done (requiring $\Delta R < 0.2$ for a match), and only those reconstructed tau leptons that do not match a true tau lepton (non-truth-matched tau leptons) are considered for promotion. The tau lepton to be promoted is picked at random among all reconstructed tau leptons that do not pass the signal tau-lepton requirements and are not truth-matched to a true tau lepton.
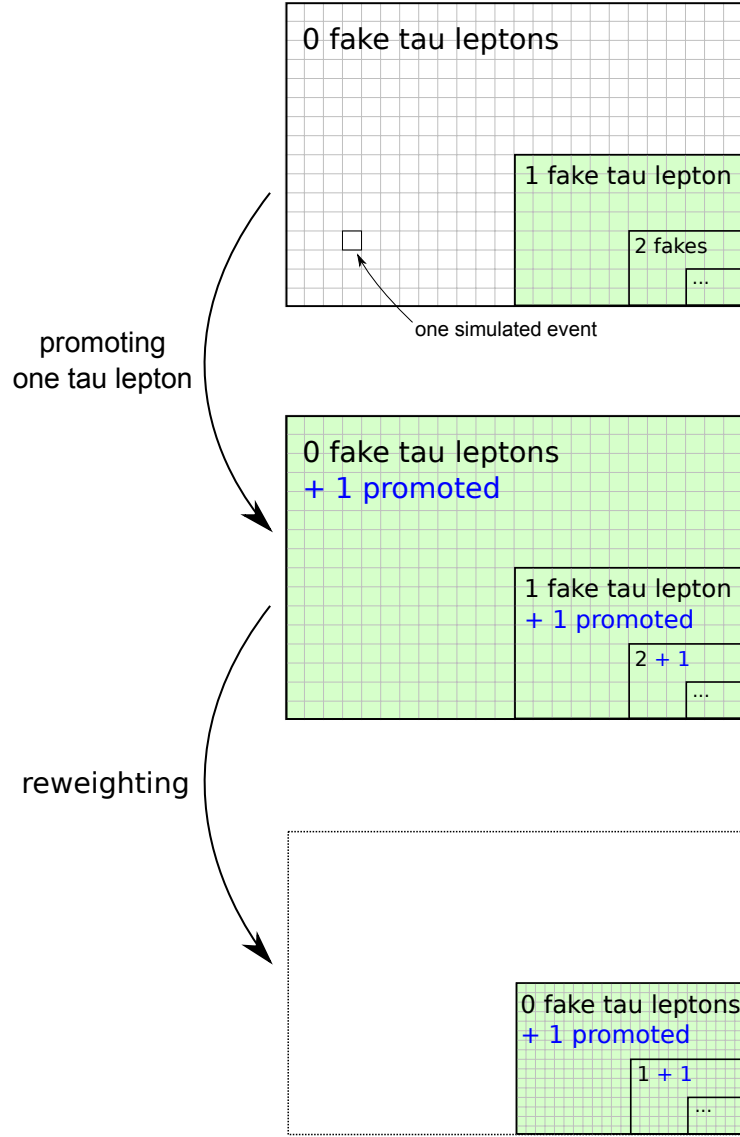
Figure 1: Schematic overview of the two steps of the tau-promotion method. Every box symbolizes one simulated event, and the size of the box indicates the event weight. Events passing a two-tau-lepton selection are shaded in light green. The predicted event yield after reweighting, shown by the bottom-most shaded box, has a much reduced statistical uncertainty (more boxes, smaller size) thanks to the promotion events with zero (or more) fake tau leptons.

The reweighting factor ω that is introduced to recover the normalization of the $W(\tau\nu)$ + jets event sample after promoting one reconstructed tau lepton can be computed as follows. Assuming there are n reconstructed tau leptons in an event, k of which pass the signal tau-lepton requirements, there will be $k + 1$ signal tau leptons after promotion.[3] The probability to observe an event with k signal tau leptons and $n - k$ non-signal, reconstructed tau leptons is given by

$$P(k) = \epsilon^k (1 - \epsilon)^{n-k} \binom{n}{k},$$
(1)

where ϵ is fake rate for the tau leptons.[4] After a reconstructed tau has been promoted, this binomial probability distribution becomes

$$P(k + 1) = \epsilon^{k+1} (1 - \epsilon)^{n-k-1} \binom{n}{k+1}.$$
(2)

The reweighting factor ω corrects the normalization after promotion back to the nominal value by weighting every event with the ratio of the probability for its occurence after to that before the promotion,

$$\omega = \frac{P(k+1)}{P(k)} = \frac{\epsilon}{1 - \epsilon} \cdot \frac{n - k}{k + 1},$$
(3)

where ϵ is the fake rate of the promoted tau lepton. The reweighting factor ω thus depends on the fake rate ϵ and includes a combinatorial factor C, which can be written as

$$C = \frac{\text{number of non-signal reconstructed tau leptons before promotion}}{\text{number of signal tau leptons after promotion}},$$
(4)

where the numerator is just the number of candidates for the promotion.

6 Results

6.1 Computation of Fake Rates

From Eq. 3, the reweighting factor ω depends on both the fake rate ϵ and a combinatorial factor C. The fake rates are computed from the same set of simulated $W(\tau\nu)$ + jets events to which they are later applied. They are given by the ratio of the number of signal tau leptons over the total number of reconstructed tau leptons,

$$\epsilon = \frac{\text{number of signal tau leptons}}{\text{number of reconstructed tau leptons}},$$
(5)

where both numerator and denominator include only non-truth matched tau leptons (as explained above). Figure 2 shows the fake rate as a function of $p_{\mathrm{T}}(\tau)$, $E_{\mathrm{T}}^{\mathrm{miss}}$ and number of prongs. There is a strong dependence of the fake rates on these quantities. No strong dependence on the pseudorapidity η of the tau lepton has been found. The medium working point is used for signal tau leptons. The fake rates for the tight working point are smaller, as is demonstrated by the middle and the bottom row in the same figure, which show the fake rates for the same sample but for the medium and the tight working point, respectively.

---

[3] All of these counts exclude truth-matched tau leptons.

[4] As the fake rate depends on the properties of each tau lepton, it would be more precise to write a product of fake rates $\Pi_{i=1}^{k} \epsilon_i$ rather than a common fake rate ϵ^k etc., but later the fake rates for all of the non-promoted tau leptons cancel out in ω, and the essentials become clearer with this approximation. A more detailed derivation can be found in appendix A.
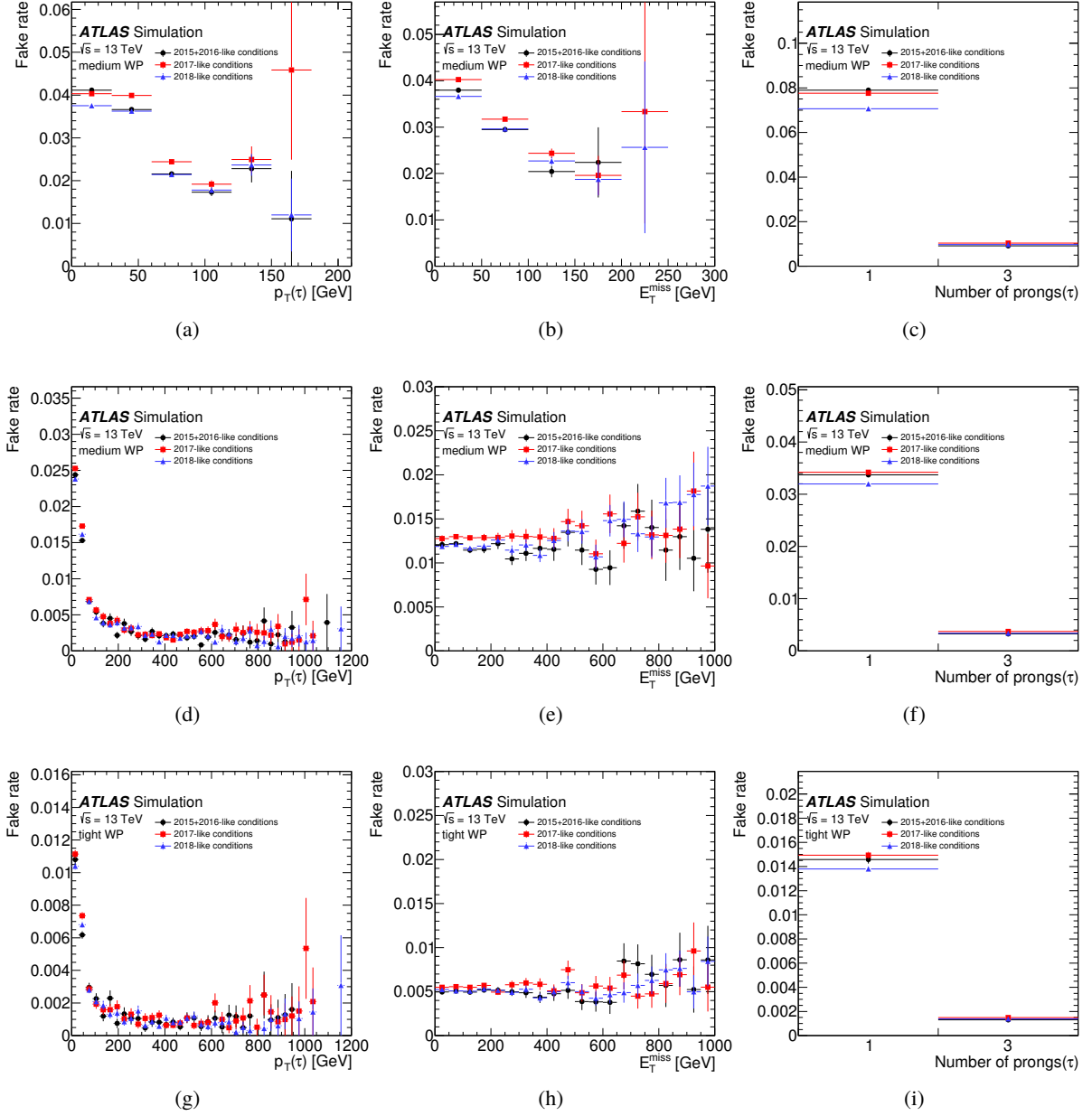
Figure 2: Fake rates as a function of $p_T(\tau)$ (left), E_T^{miss} (middle) and number of prongs (right) in $W(\tau\nu)+$ jets events with $\max[H_T, p_T(V)] < 70\,\text{GeV}$ and jets from light hadrons (top), and in $W(\tau\nu)+$ jets events with $\max[H_T, p_T(V)] > 1000\,\text{GeV}$ (middle and bottom), comparing the medium and tight work working points in the middle and bottom row, respectively. Small differences can be seen between the data-taking conditions in different years.

To take into account the above dependencies, the fake rates are measured in bins of $p_T(\tau)$, E_T^{miss} and number of prongs. A dedicated binning has been derived for each $\max[H_T, p_T(V)]$ range of the $W(\tau\nu)+$ jets sample. It is chosen such that the fake rates are as accurate as possible while keeping the relative statistical uncertainty of the fake rates typically below 10 % in each bin.

8

In addition, a possible pile-up dependence of the fake rates is taken into account by measuring the fake rates separately for each of the three Monte Carlo simulation campaigns which correspond to data-taking conditions in the years 2015 + 2016, 2017 and 2018. Figure 3 shows the fake rates in the resulting binning. Again, the medium working point is used for signal tau leptons.

In general, the fake rates are expected to differ between gluon- and quark-initiated jets. This difference is being integrated over in both the measurement of the fake rates and their application in the current implementation of the tau-promotion method. The fake rates are further expected to depend on the flavor of the quark from which a jet emerges. Again, this difference is not accounted for here, but the fraction of $W(\tau\nu)$ + jets events with jets arising from light quark flavors (up, down or strange) is around 90 % for $W(\tau\nu)$ + jets events with $\max[H_\mathrm{T}, p_\mathrm{T}(V)] < 70\,\mathrm{GeV}$, decreasing to around 60 % for $\max[H_\mathrm{T}, p_\mathrm{T}(V)] > 1000\,\mathrm{GeV}$, and thus light-flavor jets are dominant by far. Therefore, this dependence is not accounted for here, also because the contribution from heavy-flavour fakes is further reduced by the b-jet veto of the analysis selection.

6.1.1 Binning

Note that the current, very fine binning has many bins at large $p_\mathrm{T}(\tau)$ and $E_\mathrm{T}^\mathrm{miss}$ where the fake rates are either exactly 0 or 1. These values for the fake rates are, of course, not physical but the result of the lack of events to compute the true average fake rate. Further optimization of the binning in order to improve the prediction of the tau-promotion method is discussed in Ref. [33]. This approach takes into account the available statistics and uses much coarser bins at high values of the kinematic variables, where the number of events is too low to allow for a determination of the fake rates.
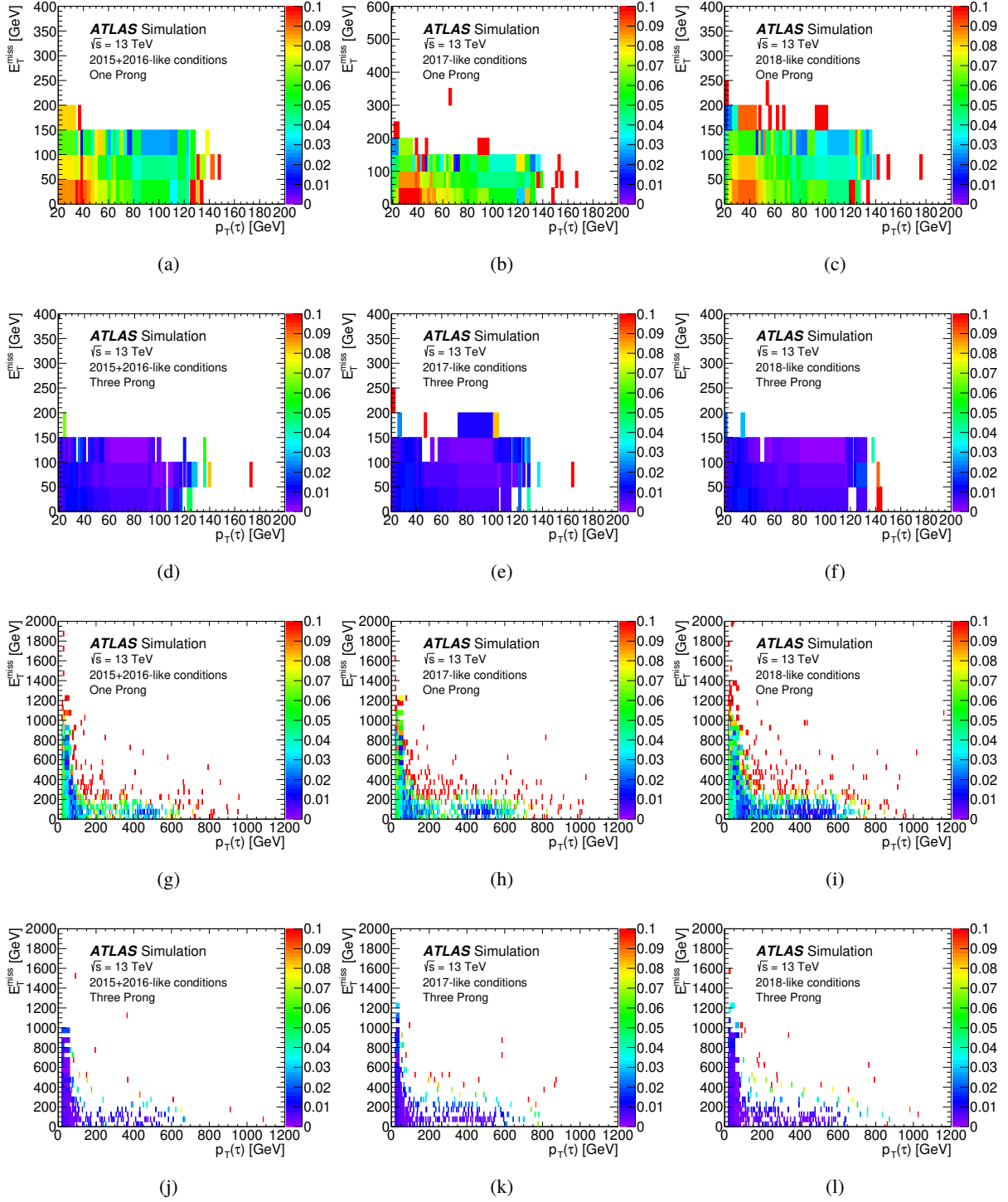
Figure 3: Color-coded fake rates for the medium working point as a function of $p_\mathrm{T}(\tau)$ and $E_\mathrm{T}^\mathrm{miss}$ for one prong (first and third row) and three prong fake tau leptons (second and fourth row). The upper two rows are for $W(\tau\nu)$ + jets events with $\max[H_\mathrm{T}, p_\mathrm{T}(V)] < 70\,\mathrm{GeV}$ and jets from light hadrons, and the bottom two rows for $W(\tau\nu)$ + jets events with $\max[H_\mathrm{T}, p_\mathrm{T}(V)] > 1000\,\mathrm{GeV}$.

6.2 Comparisons of Predicted Event Yields

The $W(\tau\nu)$ + jets yields predicted using the tau-promotion method in combination with the fake-rate measurements from above are compared with the nominal $W(\tau\nu)$ + jets yields in Table 2 after a selection requiring the given number of the signal tau leptons. The numbers are raw event numbers without any event weights (except for the reweighting factor ω of the tau-promotion method). The reweighted yield shows a residual discrepancy of roughly 10 % with respect to the nominal sample. The residual discrepancy is taken into account in the following by introducing a set of scale factors $\text{SF}^{\text{offline}} = \frac{\text{Nominal}}{\text{Promoted}}$ of 0.9188 ± 0.0022 (0.8942 ± 0.0034) for the medium (tight) working points of the tau-lepton identification, where the stated uncertainty is the statistical uncertainty.

Since the tau-promotion method does not take into account any correlations between the online (trigger) and offline identification of tau leptons, requiring a positive trigger decision would lead to another discrepancy between the promoted and the nominal yields. To account for this, another set of corrections is computed using the nominal $W(\tau\nu)$ + jets samples after applying relevant requirements that affect the trigger efficiency from the selections described in Section 4. For the di-tau + $E_{\text{T}}^{\text{miss}}$ trigger, these are the requirements of exactly two medium tau leptons, the b-jet veto, the veto on electrons and muons, $E_{\text{T}}^{\text{miss}} > 150$ GeV and $p_{\text{T}}(\tau_{1,2})$. For the asymmetric di-tau trigger, these are the requirements of exactly two tight tau leptons, the b-jet veto, the veto on electrons and muons, $75 < E_{\text{T}}^{\text{miss}} < 150$ GeV and $p_{\text{T}}(\tau_{1,2})$. The efficiencies $\epsilon^{\text{trigger}}$ for the trigger selection are computed on the nominal sample from the ratio of event yields before and after applying the trigger selection (both including the relevant offline requirements) to be $\epsilon^{\text{trigger}} = 0.531 \pm 0.033$ (0.52 ± 0.07) for the di-tau + $E_{\text{T}}^{\text{miss}}$ (asymmetric di-tau) trigger, where the stated uncertainty is the statistical uncertainty. For the promoted sample, this efficiency is applied instead of the requirement of a positive trigger decision. Both $\text{SF}^{\text{offline}}$ and $\epsilon^{\text{trigger}}$ are calculated for the full set of simulated events corresponding to the whole 2015 – 2018 dataset.

The event yields for the $W(\tau\nu)$ + jets sample when applying the selection criteria of the two signal regions defined in Section 4 one by one are given in Tables 3 and 4. The yields for the promoted and nominal samples are shown separately, together with the relative and absolute statistical uncertainties and the ratio of the promoted to the nominal yields. The promoted yields include $\text{SF}^{\text{offline}}$ and, starting from the line "Trigger", also $\epsilon^{\text{trigger}}$. The statistical uncertainties given in the tables do not account for the fact that $\epsilon^{\text{trigger}}$ has a statistical uncertainty, too. The statistical uncertainty on $\epsilon^{\text{trigger}}$ is larger for the asymmetric di-tau than for the di-tau + $E_{\text{T}}^{\text{miss}}$ trigger and amounts to 13 % for the former trigger. Adding it in quadrature to the statistical uncertainty in the event yields after the selection would give a relative uncertainty of 22 % instead of 17 % on the promoted event yield after the final selection step in Table 4. The sizeable statistical uncertainty of $\epsilon^{\text{trigger}}$ arises from the fact that it must be determined on the nominal sample that suffers from the low statistics that motivated the development of the method. It may be possible to reduce

Table 2: The predicted event yields and their absolute and relative statistical uncertainties after basic selections based on tau-lepton multiplicities that are used as part of the SR-lowMass (top row) and SR-highMass (bottom row, still without any requirement on the number of tight tau leptons) definitions. Events with a third medium tau lepton are excluded from both selections.

| | Nominal | | Promoted | | |
	Yields	Rel. unc.	Yields	Rel. unc.	$\frac{\text{Promoted}}{\text{Nominal}}$
2 tight τs	$72\,290 \pm 270$	0.37 %	$80\,840 \pm\ \ 50$	0.062 %	1.118
2 medium τs	$212\,400 \pm 500$	0.22 %	$231\,160 \pm 130$	0.054 %	1.088

Table 3: Remaining event yields for $W(\tau\nu)$ + jets when applying the selection criteria of SR-lowMass (defined in Section 4) one by one. Only statistical uncertainties are shown in the table. Events with a third medium tau lepton are excluded already in the first step. The trigger in SR-lowMass is the asymmetric di-tau trigger.

SR-lowMass	Nominal			Promoted			$\dfrac{\text{Promoted}}{\text{Nominal}}$	
	Yields		Rel. unc.	Yields		Rel. unc.		
2 tight τs	52 300	± 1400	2.7 %	55 400	± 500	0.81 %	1.060 ±	0.030
b-jet veto	50 400	± 1400	2.7 %	52 400	± 400	0.85 %	1.040 ±	0.030
e/μ veto	50 200	± 1400	2.8 %	52 300	± 400	0.85 %	1.041 ±	0.030
Trigger	87	± 17	20 %	106.1	± 3.1	3.0 %	1.22 ±	0.24
2 τs with OS	79	± 17	21 %	73.2	± 2.9	4.0 %	0.92 ±	0.20
Z/H veto	76	± 17	22 %	71.7	± 2.9	4.1 %	0.94 ±	0.21
$|\Delta\phi(\tau_1, \tau_2)| > 0.8$	76	± 17	22 %	70.8	± 2.9	4.1 %	0.93 ±	0.21
$\Delta R(\tau_1, \tau_2) < 3.2$	51	± 15	29 %	54.0	± 2.8	5.1 %	1.05 ±	0.31
$m_{\text{T2}} > 70$ GeV	−0.05 ±	0.29	570 %	1.73 ±	0.29	17 %	−30	± 190

Table 4: Remaining event yields for $W(\tau\nu)$ + jets when applying the selection criteria of SR-highMass (defined in Section 4) one by one. Only statistical uncertainties are shown in the table. Events with a third medium tau lepton are excluded already in the first step. The trigger in SR-highMass is the di-tau + $E_{\text{T}}^{\text{miss}}$ trigger.

SR-highMass	Nominal			Promoted			$\dfrac{\text{Promoted}}{\text{Nominal}}$
	Yields		Rel. unc.	Yields		Rel. unc.	
2 medium τs	146 700	± 2300	1.6 %	168 700	± 800	0.50 %	1.150 ± 0.019
b-jet veto	141 400	± 2300	1.6 %	159 700	± 800	0.52 %	1.130 ± 0.019
e/μ veto	140 700	± 2300	1.6 %	159 200	± 800	0.52 %	1.132 ± 0.019
Trigger	267	± 19	7.0 %	292	± 4	1.3 %	1.09 ± 0.08
2 τs with OS	202	± 16	8.0 %	180.3 ±	2.9	1.6 %	0.89 ± 0.07
Z/H veto	142	± 14	9.8 %	143.0 ±	2.5	1.7 %	1.01 ± 0.10
≥ 1 tight τ	128	± 14	11 %	115.9 ±	2.1	1.9 %	0.91 ± 0.10
$|\Delta\phi(\tau_1, \tau_2)| > 0.8$	120	± 14	11 %	111.2 ±	2.1	1.9 %	0.93 ± 0.11
$\Delta R(\tau_1, \tau_2) < 3.2$	85	± 12	14 %	80.5 ±	1.9	2.3 %	0.95 ± 0.13
$m_{\text{T2}} > 70$ GeV	1.8 ±	1.7	97 %	2.9 ±	0.5	17 %	1.6 ± 1.6

this uncertainty by determining the individual trigger efficiencies for each trigger leg, parametrized in the respective relevant offline quantities (tau-lepton p_{T} and $E_{\text{T}}^{\text{miss}}$). Instead of computing a global trigger efficency from the event yields after the offline trigger requirements as is done in the current implementation, the trigger decision can then be emulated under the assumption that the total trigger efficiency can be factorized into the trigger efficiencies of the trigger legs.

Reasonable agreement between the nominal and promoted samples is observed within statistical uncertainties, except for the yields after the m_{T2} requirement, where the statistical uncertainty of the nominal sample is very large. Another problem that can occur when only a small number of simulated events pass a selection can be seen in the last line of Table 3, where events with negative weights dominate and lead to an unphysical prediction of a negative expected yield. Also this problem can be solved by increasing the number of events passing the selection after applying the tau-promotion method.
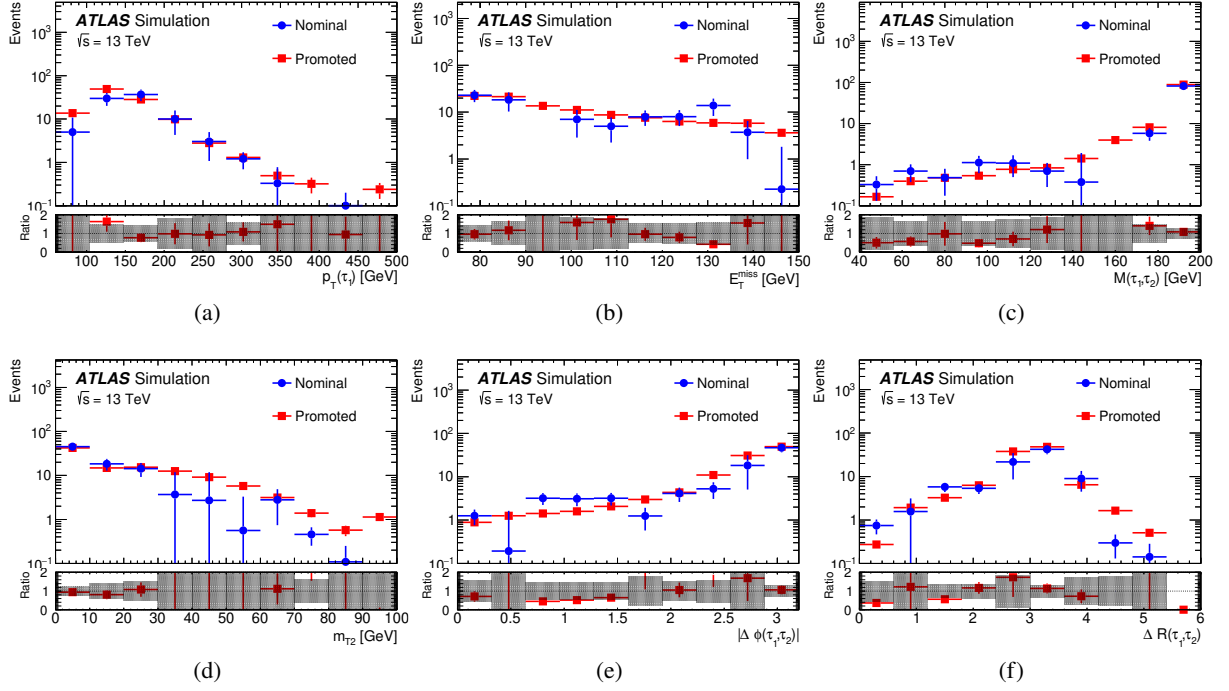
Figure 4: Kinematic distributions for the nominal and promoted samples after the asymmetric di-tau trigger selection that is part of the SR-lowMass definition. The error bars for the nominal sample show the statistical uncertainties, the error bars for the promoted sample comprise the statistical uncertainties and a flat additional 25 % uncertainty. The rightmost bin includes the overflow.

6.3 Comparisons of Kinematic Distributions

Figures 4 and 5 show the kinematic distributions for the nominal and promoted samples after applying the trigger selection. A conservative systematic uncertainty of 25 % is applied for the promoted samples. This roughly reflects the level of agreement of the nominal and promoted samples for those selections in Tables 3 and 4 where the statistical uncertainty of the nominal sample is not exceedingly large. The distributions for $p_T(\tau)$ and E_T^{miss} are expected to agree within statistical uncertainties by construction, as these variables are used in the binning of the fake rates, except for a potential influence of the trigger selection, for which a flat scale factor is used. Additional distributions are included in the figures, and show reasonable agreement between the promoted and the nominal samples, with the remaining discrepancies being covered by the uncertainties in all cases. It proves difficult to obtain a high-purity selection of $W(\tau\nu)$ + jets events in the di-tau final state that would allow comparing the predictions of the tau-promotion method (or the nominal simulation) to data. This is due to the overwhelming background of multi-jet events (with two fake tau leptons) and the background from Z-boson and diboson events (with two true tau leptons). However, the above comparisons show that the results from the tau-promotion method and the predictions of the nominal simulation are consistent.
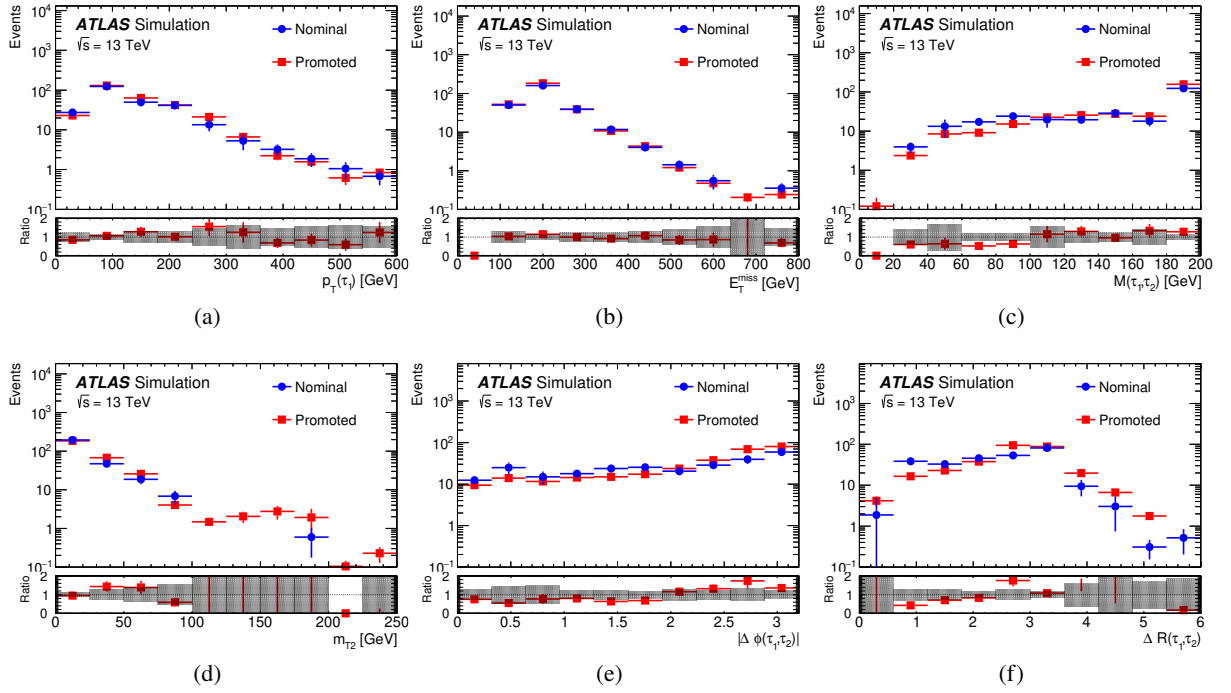
13

Figure 5: Kinematic distributions for the nominal and promoted samples after the di-tau + $E_{\mathrm{T}}^{\mathrm{miss}}$ trigger selection that is part of the SR-highMass definition. The error bars for the nominal sample show the statistical uncertainties, the error bars for the promoted sample comprise the statistical uncertainties and a flat additional 25 % uncertainty. The rightmost bin includes the overflow.

7 Conclusion

In this note, the methodology of the tau-promotion method and the results of an example implementation have been presented. This method reduces significantly the statistical uncertainties of the simulated $W(\tau\nu)$ + jets background for an event selection that requires two tau leptons to be reconstructed and pass the signal-identification requirements. The application of the tau-promotion method requires the fake rates for tau leptons as input, which have been measured in simulated events. To account for the trigger efficiencies being different in the nominal samples and in the result of the tau-promotion method, scale factors have been introduced. Using the signal-region selections from the ATLAS search for direct production of tau sleptons in Run 2 data as an example for the application of the method, reasonable agreement between the nominal simulated samples and the results from the tau-promotion method is found with respect to the event yields and the kinematic distributions. The relative statistical uncertainties of the predicted event yields are reduced by a factor between 5 and 15, depending on the tightness of the selection.

A Derivation of Reweighting Factors

This appendix gives an alternative derivation of the reweighting factor ω that is used to recover the normalization of the $W(\tau\nu)$ + jets event sample after promoting one reconstructed tau lepton without the simplification of a common fake-rate value for all tau leptons.

Assuming there are n reconstructed tau leptons in an event, k of which pass the signal tau-lepton requirements, there will be $k + 1$ signal tau leptons after promotion.[5] In the following, Ω_A denotes a set holding the indices of those tau leptons that pass the signal tau-lepton requirements, where the tau leptons in an event are indexed in an arbitrary but fixed order. The probability to observe the event specified by Ω_A is

$$P(\Omega_A) = \left(\prod_{i \in \Omega_A} \epsilon_i \right)\left(\prod_{i \notin \Omega_A} (1 - \epsilon_i) \right),\tag{6}$$

where ϵ_i is the (binned) fake rate for the ith reconstructed tau lepton.

Ω_A is an element of the subset Ω_k of the power set $\mathbb{P}$ of $\{1, \ldots, n\}$ holding all sets with k elements,

$$\Omega_k = \left\{ \Omega \in \mathbb{P}(\{1, \ldots, n\}) : |\Omega| = k \right\} \quad \text{with} \quad |\Omega_k| = \binom{n}{k}.\tag{7}$$

When the jth reconstructed tau is promoted, $\Omega_A \in \Omega_k$ becomes $\Omega_B = \Omega_A \cup \{j\}$ with $\Omega_B \in \Omega_{k+1}$. The correct probability $P(\Omega_B)$ for the event with index set Ω_B is $P(\Omega_B)$ as given by Eq. 6. The probability to observe an event with index set Ω_B after promotion is, however,

$$\tilde{P}(\Omega_B) = \sum_{\Omega_{A'}} P(\Omega_{A'} \to \Omega_B)\, \omega(\Omega_{A'} \to \Omega_B) \left(\prod_{i \in \Omega_{A'}} \epsilon_i \right)\left(\prod_{i \notin \Omega_{A'}} (1 - \epsilon_i) \right),\tag{8}$$

where the sum runs over all possible events with an index set $\Omega_{A'} \in \{\Omega_B \setminus \{j\} : j \in \Omega_B\}$, in which promotion of one tau lepton j would yield Ω_B. The probability of going from $\Omega_{A'}$ to Ω_B, $P(\Omega_{A'} \to \Omega_B)$, is a constant given by $P(\Omega_{A'} \to \Omega_B) = (n - |\Omega_{A'}|)^{-1} = (n - |\Omega_A|)^{-1}$, as picking any non-signal tau lepton in the event to be promoted has the same probability by definition of the method.

The reweighting factors $\omega(\Omega_{A'} \to \Omega_B)$ that have already been introduced in Eq. 8 are to be chosen such that $\tilde{P}(\Omega_B) = P(\Omega_B)$ for any given values of the fake rates $\{\epsilon_i\}$. Due to this latter condition, the reweighting factors must be of the form

$$\omega(\Omega_{A'} \to \Omega_B) = C \cdot \frac{\left(\prod_{i \in \Omega_B} \epsilon_i \right)\left(\prod_{i \notin \Omega_B} (1 - \epsilon_i) \right)}{\left(\prod_{i \in \Omega_{A'}} \epsilon_i \right)\left(\prod_{i \notin \Omega_{A'}} (1 - \epsilon_i) \right)} = C \cdot \frac{\epsilon_j}{1 - \epsilon_j},\tag{9}$$

where C is a constant that is independent of the tau fake rates. Inserting this in Eq. 8 and canceling the efficiency terms coming from $P(\Omega_B)$ then leads to the requirement

$$1 \overset{!}{=} \frac{P(\Omega_B)}{\tilde{P}(\Omega_B)} = \frac{1}{\sum_{\Omega_{A'}} (n - |\Omega_{A'}|)^{-1} \cdot C},\tag{10}$$

which allows C to be determined as

$$C = \frac{n - |\Omega_A|}{\sum_{\Omega_{A'}} 1} = \frac{n - k}{k + 1},\tag{11}$$

---

[5] All of these counts exclude truth-matched tau leptons. It is $k \leq n - 1$, otherwise no tau lepton is left to be promoted.

as $|\{\Omega_{A'}\}| = |\Omega_B| = k + 1$ by definition.

The reweighting factor ω thus depends on the fake rate ϵ_j and includes a combinatorial factor C, which can be written as

$$C = \frac{\text{number of non-signal reconstructed tau leptons before promotion}}{\text{number of signal tau leptons after promotion}}, \tag{12}$$

where the numerator is just the number of candidates for the promotion. This can be interpreted as a combinatorial factor arising from the flow of events through a directed graph of nodes with fixed n and k, i. e. a multiplication with the number of outgoing states and a division by the number of incoming states.

References

[1] ATLAS Collaboration, *Search for direct stau production in events with two hadronic tau leptons in* $\sqrt{s}$ = 13 *TeV pp collisions with the ATLAS detector*, ATLAS-CONF-2019-018, 2019, https://cds.cern.ch/record/2676595.

[2] ATLAS Collaboration, *ATLAS simulation of boson plus jets processes in Run 2*, ATL-PHYS-PUB-2017-006, 2017, https://cds.cern.ch/record/2261937.

[3] T. Gleisberg et al., *Event generation with SHERPA 1.1*, JHEP **02** (2009) 007, arXiv:0811.4622 [hep-ph].

[4] S. Höche, F. Krauss, M. Schönherr, and F. Siegert, *QCD matrix elements + parton showers. The NLO case*, JHEP **04** (2013) 027, arXiv:1207.5030 [hep-ph].

[5] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, JHEP **12** (2008) 039, arXiv:0808.3674 [hep-ph].

[6] F. Cascioli, P. Maierhöfer, and S. Pozzorini, *Scattering Amplitudes with Open Loops*, Phys. Rev. Lett. **108** (2012) 111601, arXiv:1111.5206 [hep-ph].

[7] A. Denner, S. Dittmaier, and L. Hofer, Collier: *A fortran-based complex one-loop library in extended regularizations*, Comput. Phys. Commun. **212** (2017) 220–238, arXiv:1604.06792 [hep-ph].

[8] S. Schumann and F. Krauss, *A parton shower algorithm based on Catani-Seymour dipole factorisation*, JHEP **03** (2008) 038, arXiv:0709.1027 [hep-ph].

[9] NNPDF collaboration, *Parton distributions for the LHC run II*, JHEP **04** (2015) 040, arXiv:1410.8849 [hep-ph].

[10] S. Catani, L. Cieri, G. Ferrera, D. de Florian, and M. Grazzini, *Vector Boson Production at Hadron Colliders: A Fully Exclusive QCD Calculation at Next-to-Next-to-Leading Order*, Phys. Rev. Lett. **103** (2009) 082001, arXiv:0903.2120 [hep-ph].

[11] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823, arXiv:1005.4568 [physics.ins-det].

[12] S. Agostinelli et al., Geant4 — *a simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250.

[13] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv:0710.3820 [hep-ph].

[14] ATLAS Collaboration, *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model*, ATL-PHYS-PUB-2016-017, 2016, https://cds.cern.ch/record/2206965.

[15] R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. B **867** (2013) 244, arXiv:1207.1303 [hep-ph].

[16] ATLAS Collaboration, *Vertex Reconstruction Performance of the ATLAS Detector at* $\sqrt{s}$ = 13 *TeV*, ATL-PHYS-PUB-2015-026, 2015, https://cds.cern.ch/record/2037717.

[17] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. C **77** (2017) 490, arXiv:1603.02934 [hep-ex].

[18] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-k_T jet clustering algorithm*, JHEP **04** (2008) 063, arXiv:0802.1189 [hep-ph].

[19] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet user manual*, Eur. Phys. J. C **72** (2012) 1896, arXiv:1111.6097 [hep-ph].

[20] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. D **96** (2017) 072002, arXiv:1703.09665 [hep-ex].

[21] ATLAS Collaboration, *Selection of jets produced in* 13 *TeV proton–proton collisions with the ATLAS detector*, ATLAS-CONF-2015-029, 2015, https://cds.cern.ch/record/2037702.

[22] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, Eur. Phys. J. C **76** (2016) 581, arXiv:1510.03823 [hep-ex].

[23] ATLAS Collaboration, *Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV*, JHEP **08** (2018) 089, arXiv:1805.01845 [hep-ex].

[24] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. (2019), arXiv:1902.04655 [hep-ex].

[25] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **76** (2016) 292, arXiv:1603.05598 [hep-ex].

[26] ATLAS Collaboration, *Identification and energy calibration of hadronically decaying tau leptons with the ATLAS experiment in pp collisions at $\sqrt{s} = 8$ TeV*, Eur. Phys. J. C **75** (2015) 303, arXiv:1412.7086 [hep-ex].

[27] ATLAS Collaboration, *Reconstruction of hadronic decay products of tau leptons with the ATLAS experiment*, Eur. Phys. J. C **76** (2016) 295, arXiv:1512.05955 [hep-ex].

[28] ATLAS Collaboration, *Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at $\sqrt{s} = 13$ TeV*, ATLAS-CONF-2017-029, 2017, https://cds.cern.ch/record/2261772.

[29] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **78** (2018) 903, arXiv:1802.08168 [hep-ex].

[30] ATLAS Collaboration, *E_T^{miss} performance in the ATLAS detector using 2015–2016 LHC pp collisions*, ATLAS-CONF-2018-023, 2018, https://cds.cern.ch/record/2625233.

[31] C. G. Lester and D. J. Summers, *Measuring masses of semi-invisibly decaying particles pair produced at hadron colliders*, Phys. Lett. B **463** (1999) 99, arXiv:hep-ph/9906349.

[32] A. Barr, C. Lester, and P. Stephens, *A variable for measuring masses at hadron colliders when missing energy is expected; m_{T2}: the truth behind the glamour*, J. Phys. G **29** (2003) 2343, arXiv:hep-ph/0304226.

[33] A. Samara, *Improving the modelling of the associated production of W-bosons and jets in the ATLAS detector*, Master's thesis, LMU München, Feb, 2018. `https://www.etp.physik.uni-muenchen.de/publications/theses/download/master_asamara.pdf`.