



ATLAS NOTE

ATL-PHYS-PUB-2014-018
22nd October 2014



Studies of the VBF $H \rightarrow \tau_\ell \tau_{\text{had}}$ analysis at High Luminosity LHC conditions

The ATLAS Collaboration

Abstract

This note documents projections for the measurement of Standard Model $H \rightarrow \tau\tau$ in the vector boson fusion production channel for the High Luminosity LHC (HL-LHC). The projections are based on the Run-I analysis. It is assumed that the HL-LHC will deliver an integrated luminosity of 3000 fb^{-1} , a center-of-mass energy of 14 TeV and an average number of overlapping pp collisions of $\langle \mu \rangle = 140$. Only the VBF production is targeted, and the analysis requires one tau to decay leptonically and the other hadronically. The projected uncertainty on the signal strength is 24% when theory uncertainties are ignored and 10% (5%) background (signal) uncertainties are assumed. If the tracking coverage is extended, the projected uncertainty on the signal strength is as low as 8%.

1 Introduction

This note documents projections of the Standard Model $H \rightarrow \tau\tau$ analysis for the ECFA High Luminosity LHC Experiments Workshop 2014. The projection considers High Luminosity LHC (HL-LHC) running conditions with 14 TeV pp collisions, 3000 fb^{-1} delivered integrated luminosity, and an average number of overlapping pp collisions per bunch-crossing (pile-up) $\langle\mu\rangle = 140$. Only the VBF $\tau_\ell\tau_{\text{had}}$ ($\ell = e, \mu$) analysis category is considered here.

This projection is built from the existing 2012 analysis [1] by using the same Monte Carlo samples and multivariate analysis (MVA) techniques. It is projected to HL-LHC conditions by adding emulation of the harsher pile-up conditions and scaling the predictions by the ratios of cross-sections and integrated luminosity for HL-LHC versus 2012 conditions. The harsher pile-up conditions impacts jets and $E_{\text{T}}^{\text{miss}}$ significantly.

The brief object definitions and analysis preselection (reflecting the Run-I analysis) are described in Section 2. The pile up emulation is described in Section 3 and the analysis itself including the BDT description and input kinematic distributions can be found in Section 4.

The analysis considers possible extensions of the tracking volume to investigate the impact of tracking-based rejection of pile-up jets. The different investigated scenarios are described in Section 4.1. The results are given in Section 5.

2 Event selection

2.1 Final state objects

The selection of a VBF-like sample for the $\tau_\ell\tau_{\text{had}}$ final state discussed in this note relies on the identification of one lepton (electron or muon), one hadronic tau and at least two jets.

Muons are selected if they have transverse momentum higher than 26 GeV and are in the region $|\eta| < 2.4$. Quality criteria on the inner detector track associated to the muon are also applied.

Electron candidates are formed from a cluster in the electromagnetic calorimeters (transition regions, $1.37 < |\eta| < 1.52$, are excluded) that is matched with a track reconstructed within the inner detector, $|\eta| < 2.47$. Electrons with a transverse momentum higher than 26 GeV are selected and a “tight” identification working point is used.

For both muons and electrons, calorimeter and track-based isolation criteria with similar combined selection efficiency as the LHC Run-I analysis are assumed.

Jets are reconstructed using the anti- k_t algorithm [2] with a radius parameter $R = 0.4$, taking topological clusters in the calorimeters as inputs. Only jets with $|\eta| < 4.5$ are selected in this analysis. The general p_T threshold for jets is 30 GeV, and specific selection of the VBF analysis is mentioned later. Track-based pile-up suppression with jet-vertex fraction (JVF)[3] is applied in the range of the tracking volume (jet $|\eta| < 2.4$).

In the range $|\eta| < 2.5$, b -tagged jets are identified using the MV1 tagging algorithm based on the impact parameter information and on the reconstruction of the displaced vertices of the hadron decays inside the jets [4]. For this study, a b -tagging efficiency working point of 70% on average for b -jets with $p_T > 30 \text{ GeV}$ is used, which has an average efficiency of 0.5 – 1% for light-flavour jets.

This analysis uses tau candidates seeded by anti- k_t [2], $R = 0.4$ jets with $p_T > 20 \text{ GeV}$ whose calorimeter cluster and leading track must satisfy $|\eta| < 2.47$. The Boosted Decision Tree (BDT) tau identification method [5] is used, requiring that the tau candidate passes the “medium” tightness, corresponding to approximately 55–60% efficiency. A dedicated selection to reject fake tau candidates from electrons and muons is applied.

When different objects selected according to the above criteria overlap with each other geometrically (within $\Delta R < 0.2$), only one of them is considered for further analysis. The overlap is resolved by selecting muon, electron, τ_{had} and jet candidates in this order of priority.

The signal events are characterized by true E_T^{miss} due to the presence of the neutrinos from the tau decays. In this analysis, the E_T^{miss} reconstruction [6] uses reconstructed high- p_T physics objects (electrons, photons, hadronically decaying τ -leptons, jets and muons) and a measurement of the soft term, which includes contributions from the underlying event, multi-parton interactions, and physics objects below analysis threshold.

2.2 Event preselection

Unless otherwise noted, the selection criteria are identical to the Run-I analysis. One lepton and one hadronically decaying tau are required, and there must be at least two jets with a significant separation in η as expected for VBF production. Some additional topological cuts are applied to suppress backgrounds while retaining most of the signal. The VBF category is intentionally defined loosely since the discrimination of signal from background is meant to be handled by the BDTs. The event selection is summarized in Table 1.

The Missing Mass Calculator (MMC) technique is used to estimate the di-tau mass [7].

The Run-I analysis cuts at $m_T(\ell, E_T^{\text{miss}}) < 70$ GeV. This cut is relaxed here to 100 GeV to avoid signal loss due to the degradation of the E_T^{miss} resolution at high $\langle\mu\rangle$.

Type	Selection
$\tau_\ell\tau_{\text{had}}$ preselection	exactly one identified and isolated lepton (e, μ) exactly one identified tau opposite sign lepton and tau no additional leptons passing loosened identification criteria no jets passing the b -tagging criteria $m_T(\ell, E_T^{\text{miss}}) < 100$ GeV
VBF categorization	leading jet with $p_T > 50$ GeV any additional jet with $p_T > 30$ GeV $ \Delta\eta(\text{lead jet, sublead jet}) > 3.0$ $m_{\tau\tau}^{\text{vis.}} > 40$ GeV

Table 1: Event selection and categorization criteria.

3 Emulation of High-Luminosity LHC conditions

The method of pile-up emulation follows the same procedure used by the $H \rightarrow WW^*$ projection for ECFA 2013 [8]. The approach is to port the existing Run-I analysis to HL-LHC conditions by overlaying pile-up jets on the 8 TeV samples, degrading the hard-scatter (HS) jet and E_T^{miss} resolution, and propagating the impact of this to the analysis.

3.1 Performance assumptions

A triggering efficiency for the previously described preselection similar to the one of the Run-I analysis is assumed.

The single lepton trigger thresholds are expected to rise monotonically as a function of instantaneous luminosity, but multi-object triggers can be used to recover the acceptance. In preparing for Run-II, a $\ell + \tau_{\text{had}} + \text{jet}$ trigger with lepton threshold 10 GeV below the single lepton trigger thresholds was found to have marginal signal acceptance loss at a small cost in bandwidth.

Additionally, the ATLAS trigger has many upgrades planned to mitigate the impact of higher instantaneous luminosity, including the New Small Wheel, hardware trackers, and finer L1 EM-calorimeter granularity. Improvements to the L1 granularity will especially improve triggers with electrons and taus, since shower shape variables can be built to discriminate against QCD jets better than existing isolation variables. It is then deemed unnecessary to consider scenarios of significant trigger efficiency loss in detail.

It is assumed that the lepton and tau reconstruction and identification efficiencies are equivalent to that observed in the 2012 data. This is a reasonable assumption as the detector upgrades for the HL-LHC aim for achieving a performance similar to Run-I despite the harsher pile-up conditions.

Hard-scatter jets from the 8 TeV samples are smeared to emulate the reconstructed jet resolution at HL-LHC conditions [9]. The jet smearing is propagated to the E_T^{miss} calculation.

The 8 TeV soft-term resolution is smeared to HL-LHC conditions : 33 MeV per unit of $\langle\mu\rangle$, which is derived from high pile-up Z/γ^* samples.

3.2 Pile-up jets

Insertion

Pile-up jets are inserted into the event according to rates recommended by the ATLAS projections [9]. For $\langle\mu\rangle = 140$ and jet $p_T \geq 30$ GeV, the rate is 2.4 additional pile-up jets per event. The kinematics of the inserted pile-up jets are derived from high pile-up Z/γ^* samples. Templates are built for p_T and η and are then randomly sampled for each inserted pile-up jet.

The hard-scattering jets are assumed to have a reconstruction efficiency of the track-confirmation algorithm recommended by the existing ATLAS projections. The pile-up jets are assumed to have the pile-up suppression efficiency of the 2012 JVF algorithm, which is 98% within the tracking volume [3]. No degradation of pile-up jet suppression is assumed because techniques already exist which outperform the JVF tagger. The pile-up suppression is assumed to not depend on p_T and η other than at the boundary of the tracker.

The insertion of pile-up jets is then propagated to the E_T^{miss} calculation.

Pile-up jet suppression for $|\eta| > 2.4$

A new tracker in the forward region is being considered for Phase II upgrades. One benefit of a forward tracker would be the use of vertexing to reject forward pile-up jets.

To emulate the impact of pile-up jet suppression with a forward tracker, the analysis is re-run with forward pile-up jet rejection imposed by hand. Since the scope of the forward tracker is uncertain, a range of coverage and performance is considered: coverage of $|\eta| < 3.0$, 3.5 , and 4.0 and pile-up jet suppression of 50% , 75% , and 90% with negligible signal (hard-scattering jets) efficiency loss is evaluated. For comparison, a pile-up jet suppression of 90% with negligible signal efficiency loss is comparable to the performance of the JVF tagger within the existing tracker [3].

Additional pile-up emulation techniques

The method of randomly inserting pile-up jets into existing events does not consider correlations between pile-up jets. This could be problematic because the p_T imbalance of the event could be overstated, which would propagate to unphysical biases of quantities which are sensitive to balance like E_T^{miss} . The impact of this has been assessed by overlaying entire truth pile-up events onto the existing hard scatter events. The results were comparable.

3.3 Impact on observables

The VBF $H \rightarrow \tau\tau$ analysis relies on two jets with large $\Delta\eta$ to describe the VBF topology and on E_T^{miss} to describe di- τ decays. The presence of forward pile-up jets can then be expected to degrade the sensitivity of the analysis because they will cause migration of background events into the VBF category, and because they will bias the E_T^{miss} calculation.

The contamination of pile-up jets in the VBF category for the total background is significant. Most events (72%) have a sub-lead pile-up jet, and nearly half (42%) have a lead pile-up jet. These pile-up jets are especially problematic because they are typically forward, thus any event with pile-up lead and sub-lead jets in opposite hemispheres will have $|\eta_{j1} - \eta_{j2}| > 4.8$.

Degradation of the E_T^{miss} and MMC $m(\tau\tau)$ under the high-luminosity conditions is shown in Figure 1 for simulated VBF $H \rightarrow \tau\tau$. Both observables are degraded by jet and E_T^{miss} smearing and by the presence of forward pile-up jets biasing the E_T^{miss} calculation.

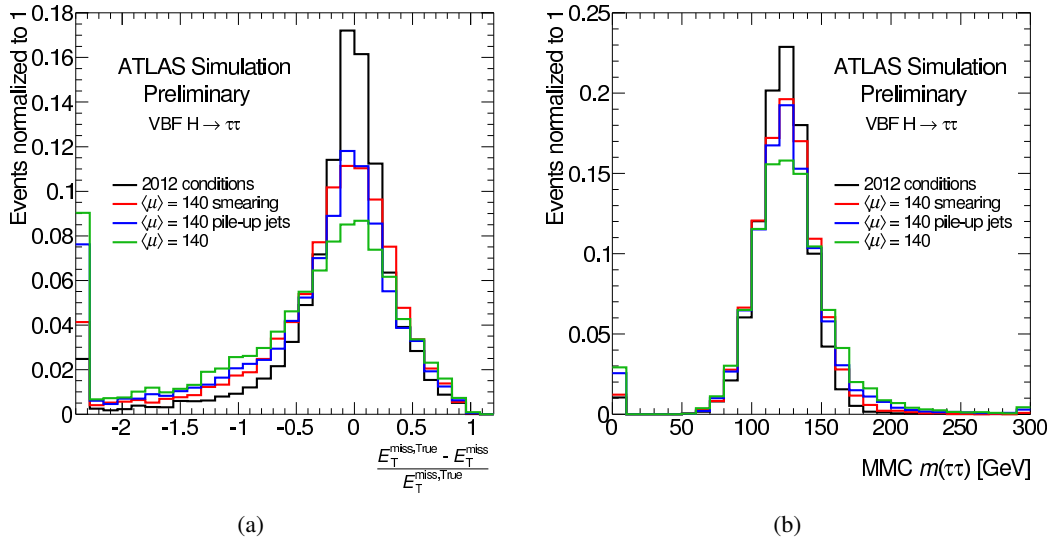


Figure 1: Degradation of E_T^{miss} -related observables at HL-LHC conditions for VBF $H \rightarrow \tau\tau$. The black line shows 2012 conditions, the red line shows 2012 conditions with HL-LHC smearing, the blue line shows 2012 conditions with HL-LHC pile-up jets, and the green line shows 2012 conditions with HL-LHC smearing and pile-up jets (i.e., full HL-LHC conditions). The first (last) bin contains the underflow (overflow) events. The underflow in the $m(\tau\tau)$ distribution shows the fraction of events which fail the MMC mass reconstruction.

4 Analysis

4.1 Boosted Decision Tree training

A multi-variate analysis approach is used by training BDTs to discriminate signal from background. It is trained using all backgrounds scaled to their respective cross-sections against the total (ggF + VBF) signal shapes. The same training parameters and input variables as the Run-I analysis are used, and the input variables are listed in Table 2.

Variable	Definition
$\Delta R(\tau, \ell)$	Separation of the lepton and τ_{had}
m_T	Transverse mass of the lepton and E_T^{miss}
E_T^{miss}	Centrality of the E_T^{miss} between the lepton and τ_{had}
ϕ -centrality	$\tau\tau$ mass estimator
MMC mass	Invariant mass of the 2 leading jets
$m_{j1,j2}$	Product of the η s of the two leading jets
$\eta_{j1} \times \eta_{j2}$	Absolute difference η s of the two leading jets
$ \eta_{j1} - \eta_{j2} $	Centrality of the lepton between the two leading jets
ℓ η -centrality	$ \vec{p}_T^\ell + \vec{p}_T^{\tau_h} + \vec{p}_T^{j1} + \vec{p}_T^{j2} + \vec{E}_T^{\text{miss}} $
p_T^{Total}	

Table 2: Discriminating variables used for the BDT training.

Forward tracking scenarios

BDTs are trained for a variety of forward tracker coverages ($|\eta| < 3$, $|\eta| < 3.5$ and $|\eta| < 4$) and pile-up rejection values (50%, 75% and 90%). Figure 2 shows the efficiency for rejecting the background versus the efficiency for selecting the signal for the scenario of 90% forward pile-up rejection. For a given signal efficiency, the background rejection improves with larger coverage.

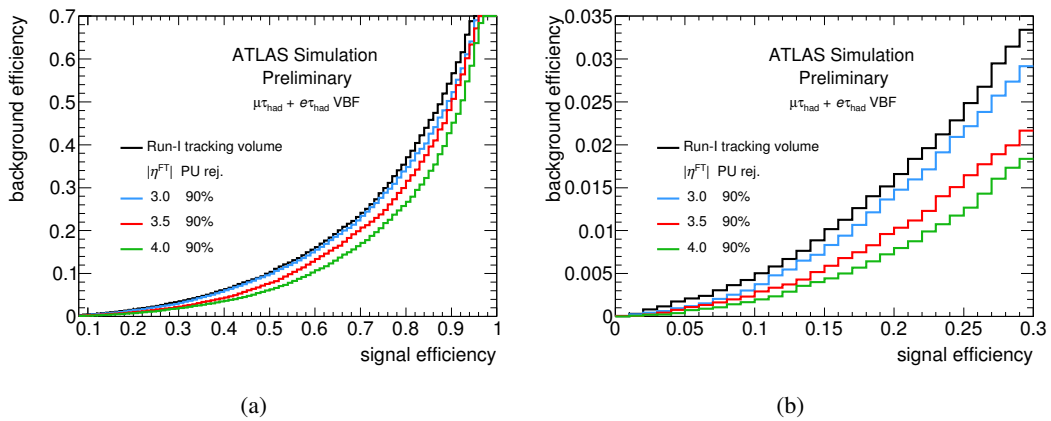


Figure 2: Signal efficiency versus background efficiency for scenarios of generic forward tracker coverage and rejection power. The right plot is a zoom of the left plot. A BDT is trained in the VBF category for each scenario.

4.2 Kinematic distributions

Predicted signal and background BDT input distributions as well as basic object kinematics are shown in Figures 3, 4 and 5. Signal and background are predicted with the same methods as the Run-I analysis [1], including the data-driven prediction of the dominant $Z \rightarrow \tau\tau$ and fake backgrounds.

Figures 3 (a) and (b) present the leading and sub-leading jet p_T , η -related variables are in Figures 3 (c) and (d) (jet η), (e) ($\Delta\eta_{jj}$) and (g) ($\eta_{lead\,jet} \times \eta_{sub-lead\,jet}$). The jet invariant mass (m_{jj}) is in Figure 3 (f), Figure 3 (h) presents the E_T^{miss} . The three dominant backgrounds are distinguished by colors, the rest of the backgrounds is summed into “Others”. Lepton variables are in Figure 4: (a) and (b) present the p_T of the hadronic τ and lepton, (c) and (d) their η and (e) presents the $\Delta R(\tau_h, l)$.

Mass-related variables are shown in Figures 4 (f) - (h): the output of the MMC in (f), the visible mass (m_{vis}) in (g), the transverse mass (m_T) reconstructed from lepton and E_T^{miss} in (h). Variables combining jet and lepton variables are presented in Figure 5: (a) $E_T^{miss}\phi$ -centrality, (b) ℓ η -centrality and (c) p_T^{Total} .

The resulting BDT score is presented in Figure 6 (a) in the full range. Figure 6 (b) presents the zoom of the high BDT score region (0.8 – 1.0).

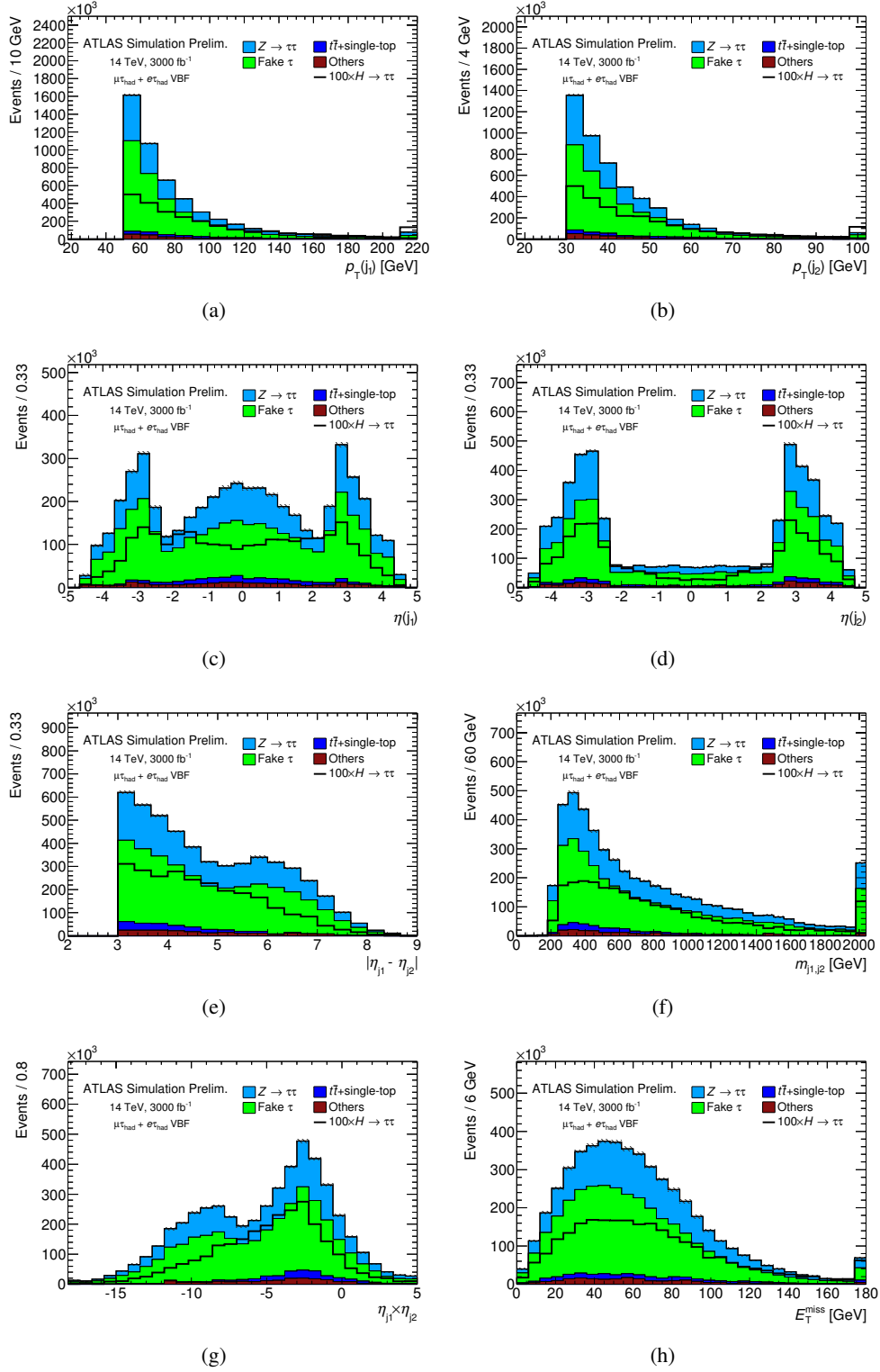


Figure 3: Signal (solid line, multiplied by 100 for visualisation purposes) and background (colored areas) HL-LHC predictions of (a) leading jet p_T , (b) sub-leading jet p_T , (c) leading jet η , (d) sub-leading jet η , (e) $\Delta\eta_{jj}$, (f) m_{jj} , (g) $\eta_{\text{lead jet}} \times \eta_{\text{sub-lead jet}}$ and (h) E_T^{miss} . The last bin contains the overflow events.

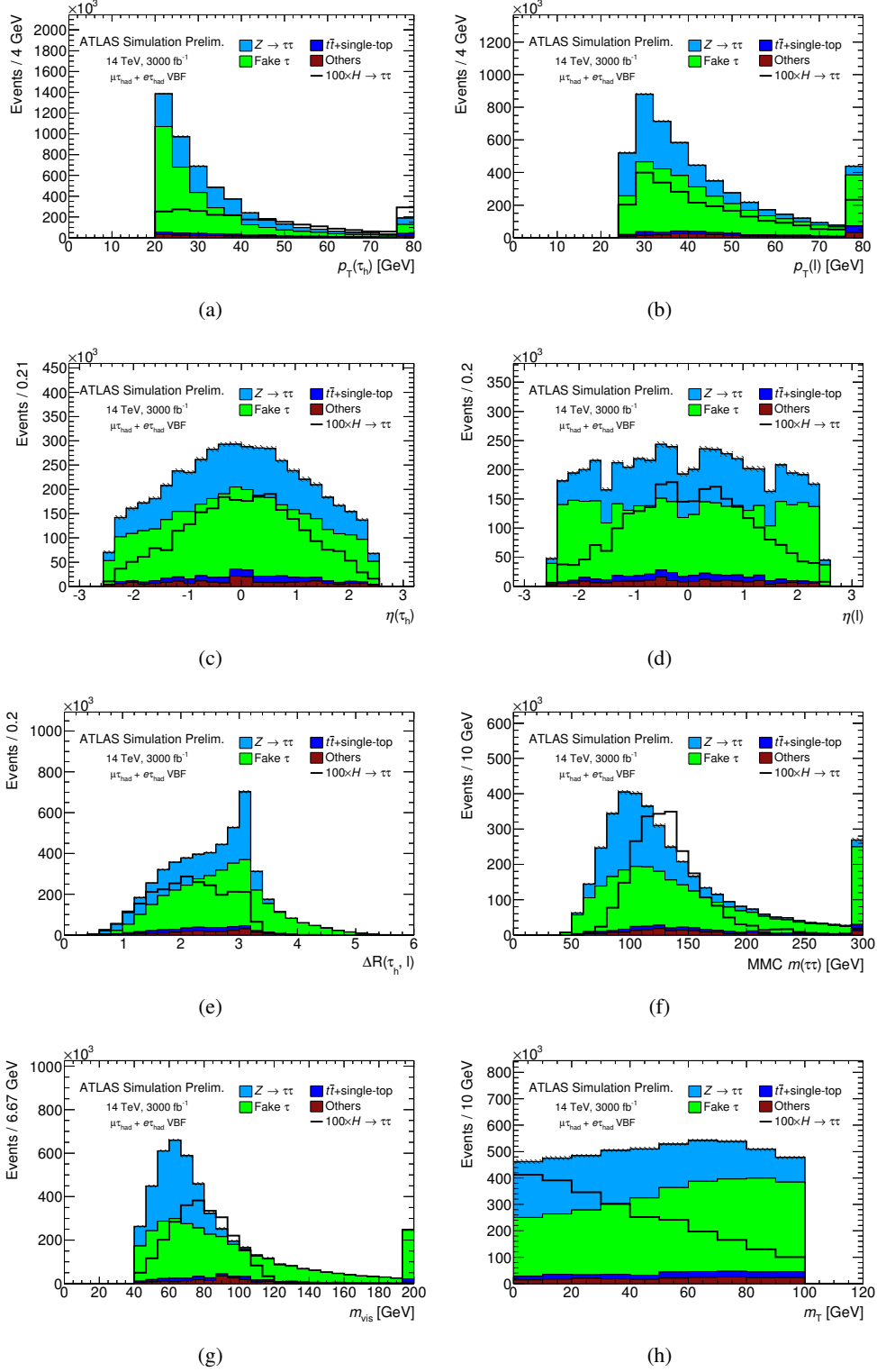


Figure 4: Signal (solid line, multiplied by 100 for visualisation purposes) and background (colored areas) HL-LHC predictions of (a) $p_T(\tau_{\text{had}})$, (b) $p_T(l)$, (c) $\eta(\tau_{\text{had}})$, (d) $\eta(l)$, (e) $\Delta R(\tau_h, l)$, (f) MMC (g) m_{vis} and (h) $m_T(l, E_T^{\text{miss}})$. The last bin contains the overflow events.

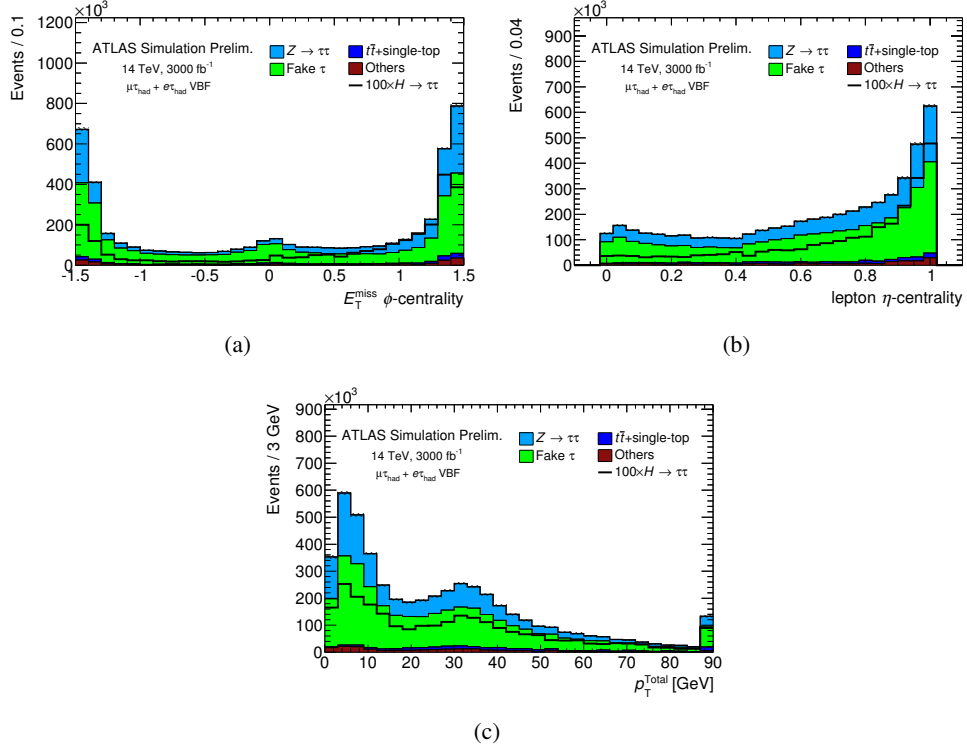


Figure 5: Signal (solid line, multiplied by 100 for visualisation purposes) and background (colored areas) HL-LHC predictions of (a) $E_T^{\text{miss}} \phi$ -centrality, (b) $\ell \eta$ -centrality and (c) p_T^{Total} . The last bin contains the overflow events.

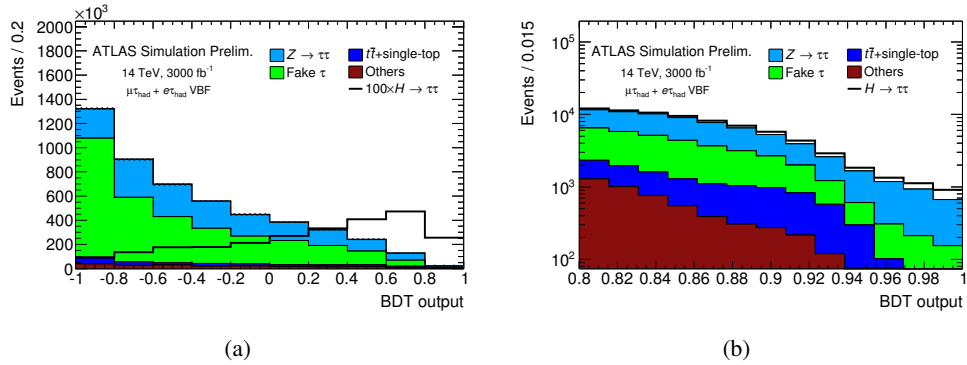


Figure 6: Signal (solid line, multiplied by 100 in (a) for visualisation purposes) and background (colored areas) HL-LHC predictions of the BDT spectrum in the (a) full range and (b) highest bins range. Signal and background are overlaid in (a) and stacked in (b).

5 Results

5.1 Yields

Yields for signal and background in the high BDT score bins are shown in Table 3. As in the Run-I analyses, $Z \rightarrow \tau\tau$ and fakes are the dominant backgrounds in the most signal-like regime. The binning of the BDT is optimized to maximize the expected sensitivity.

process	VBF category	third highest bin	second highest bin	highest bin
VBF $H \rightarrow \tau\tau$	8970	114	147	206
ggF $H \rightarrow \tau\tau$	16410	44	46	39
$Z \rightarrow \tau\tau$	1682400	875	720	514
fakes	2959800	205	190	155
$t\bar{t}$	191400	100	20	< 20
other	198600	< 20	< 20	< 20
signal	25380	158	193	245
background	5032200	1180	930	669

Table 3: Yields for signal and background in the VBF category and in the most sensitive BDT bins, as shown in Figure 6.

5.2 Uncertainties assumptions

When calculating the sensitivity of the analysis, three scenarios of background uncertainties and two scenarios of theory uncertainties are considered. The theory uncertainties are varied from no theory uncertainties to Run-I theory uncertainties, which are as large as 6% (30%) for the VBF (ggF) Higgs production modes [1]. The experimental signal uncertainty is fixed at 5% accounting for experimental sources such as jet energy scale uncertainties. The experimental background uncertainties are varied to 10% and 5% of the prediction, and they are treated as uncorrelated between backgrounds and between bins of the BDT score. This assumes the uncertainties can be reduced compared to the Run-I analysis.

5.3 Sensitivity

The projected sensitivity is shown in Table 4. The two scenarios of background uncertainties and two scenarios of theory uncertainties are shown. The sensitivity of the projection is driven by the uncertainty on the background prediction. For $\sigma_B^{\text{syst.}} = 10\%$, the projected uncertainty on μ with no signal theory uncertainties is 0.24. For $\sigma_B^{\text{syst.}} = 5\%$, this projected uncertainty is 0.13.

5.4 Improvement potential with a forward tracking

The impact of pile-up jet rejection in the forward region is also evaluated as an example of the impact of a forward tracker, and results are given in Table 5. Multiple scenarios of $|\eta|$ coverage and pile-up jet rejection are considered.

		current $\sigma_S^{\text{theo.}}$	no $\sigma_S^{\text{theo.}}$
$\sigma_B^{\text{syst.}}$	$\sigma_S^{\text{syst.}}$	$\Delta\mu$	$\Delta\mu$
10%	5%	0.25	0.24
5%	5%	0.16	0.13

Table 4: Uncertainty on the signal strength ($\Delta\mu$) for different scenarios of background uncertainties and signal theory uncertainties.

forward pile-up jet rejection	50%	75%	90%
forward tracker coverage	$\Delta\mu$		
Run-I tracking volume	0.24		
$ \eta < 3.0$	0.18	0.15	0.14
$ \eta < 3.5$	0.18	0.13	0.11
$ \eta < 4.0$	0.16	0.12	0.08

Table 5: Uncertainty on the signal strength ($\Delta\mu$) for different scenarios of forward tracking. Negligible loss of HS jets to forward pile-up jet rejection is assumed. A 10% systematic uncertainty is assumed for backgrounds, a 5% experimental systematic uncertainty is assumed for signals, and theoretical uncertainties on signals are ignored.

6 Conclusions

The projection of the Standard Model $H \rightarrow \tau\tau$ analysis to the High Luminosity LHC (HL-LHC) running conditions with 14 TeV pp collisions, 3000 fb $^{-1}$ delivered integrated luminosity, and an average number of overlapping pp collisions $\langle\mu\rangle = 140$ is performed. The VBF $\tau_\ell\tau_{\text{had}}$ ($\ell = e, \mu$) analysis category is considered, and the uncertainty on the signal strength (μ) is projected to be 24% when theory uncertainties are ignored and 10% (5%) background (signal) uncertainties are assumed. The projected uncertainty could be reduced significantly if pile-up jets outside the current tracking volume could be rejected similar to pile-up jet rejection within the tracking volume in 2012. The uncertainty on μ is projected to be 8–18% depending on the scenario of forward tracker coverage and pile-up jet rejection.

References

- [1] The ATLAS Collaboration, *Evidence for Higgs boson Yukawa couplings in the $H \rightarrow \tau\tau$ decay mode with the ATLAS detector*, ATLAS-CONF-2014-061 (2014), atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2014-061.
- [2] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-kt jet clustering algorithm*, JHEP **04** (2008) 063, [arXiv:0802.1189 \[hep-ph\]](https://arxiv.org/abs/0802.1189).
- [3] The ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*, ATLAS-CONF-2014-018 (2014), atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2014-018.
- [4] The ATLAS Collaboration, *Calibration of the performance of b-tagging for c and light-flavour jets in the 2012 ATLAS data*, ATLAS-CONF-2014-046 (2014), atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2014-046.
- [5] The ATLAS Collaboration, *Identification of Hadronic Decays of Tau Leptons in 2012 Data with the ATLAS Detector*, ATLAS-CONF-2013-064 (2013), atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2013-064.
- [6] The ATLAS Collaboration, *Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 7$ TeV with ATLAS*, The European Physical Journal C - Particles and Fields **72** (2012) 1–35. [10.1140/epjc/s10052-011-1844-6](https://arxiv.org/abs/10.1140/epjc/s10052-011-1844-6).
- [7] A. Elagin et al., *A new mass reconstruction technique for resonances decaying to di-tau*, Nucl. Instrum. Meth. A **654** (2011) 481, [arXiv:1012.4686 \[hep-ex\]](https://arxiv.org/abs/1012.4686).
- [8] The ATLAS Collaboration, *Projections for measurements of Higgs boson cross sections, branching ratios and coupling parameters with the ATLAS detector at a HL-LHC*, ATL-PHYS-PUB-2013-014 (2013), atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2013-014.
- [9] The ATLAS Collaboration, *Performance assumptions for an upgraded ATLAS detector at a High-Luminosity LHC*, ATL-PHYS-PUB-2013-004 (2013), atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2013-004.