Prospective results for vector-boson fusion-mediated Higgs-boson searches in the four lepton final state at the High Luminosity Large Hadron Collider

The ATLAS Collaboration

Abstract

The High Luminosity Large Hadron Collider is expected to be completed and operational in the second half of 2026, and will necessitate substantial upgrades to the ATLAS inner tracker detector. The impact of increased tracking coverage in the forward direction was investigated in terms of the separation of vector-boson fusion and gluon fusion-mediated Higgs-boson decays to four leptons in association with two jets. For an analysis dominated by statistical uncertainty, with vector-boson fusion production events treated as signal on top of gluon fusion background, the extension of jet tracking from pseudorapidity $|\eta| < 2.7$ to $|\eta| < 4.0$ improved the prospective vector-boson fusion discovery significance by 16%, while the relative uncertainty on the signal strength $\Delta \mu / \mu$ was reduced by 6%.
1. Introduction

In the Standard Model (SM) of Particle Physics, vector-boson fusion (VBF) is a production mechanism wherein two incoming quarks emit virtual $W$ or $Z$ bosons which undergo inverse pair decay to form a Higgs-boson [1]. As shown in Figure 1, contributions to leading order VBF production are made in the $s$, $t$, and $u$ channels. However, at circular hadron colliders such as the Large Hadron Collider (LHC), the $t$ and $u$ fusion channels are heavily favoured [2], as the partonic cross sections of these contributions rise logarithmically with the centre-of-mass energy of the subprocess ($\hat{\sigma} \propto \log \hat{s}/M_V^2$) [3], while the $s$ channel is suppressed by the application of cuts on hadronic decay products. It is worth noting, as well, that the $WW$ fusion contribution is the dominant term in the VBF cross section – a consequence of the larger coupling of the $W$ boson to fermions [3]. Contributions from higher-order quantum chromodynamics (QCD)-mediated diagrams are also on the order of $\sim 15\%$ [4, 5], meaning that the process cross section is dominated by the purely electroweak, Born-order diagrams.

![leading order diagrams](image)

Figure 1: Leading order diagrams for Higgs-boson production through vector-boson fusion. Here, $V = W^\pm$ or $Z$.

VBF Higgs-boson production is predicted to be the second largest contribution to the total Higgs-boson production cross section for a mass $m_H \sim 125\text{ GeV}$, though it is still an order of magnitude smaller than the production cross section for gluon fusion [6]. Unlike the gluon fusion process, though, the VBF production signature is highly distinctive, marked by the presence of two highly energetic final state quark jets at leading order, which tend to be found in opposite forward regions of cylindrical detectors such as ATLAS [2]. Furthermore, the resultant Higgs-boson tends to be produced in the central region, with low momentum, such that $E_H \sim m_H$. As such, the distinct kinematic properties of the dijet and Higgs systems can be used to simultaneously reduce backgrounds originating in QCD interactions, and identify VBF-like events.

1.1. The High Luminosity LHC and Phase-II upgrades

The High Luminosity LHC (HL-LHC) is expected to be completed and operational in the second half of 2026, and will increase the instantaneous luminosity of the collider to $\mathcal{L} = 5 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$, corresponding to an average number of pile-up interactions (or $\langle \mu \rangle$) of $\sim 140$ pp collisions per beam-crossing [7]. Furthermore, it is envisioned that an ultimate luminosity of $\mathcal{L} = 7.5 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ will be achievable, corresponding to $\langle \mu \rangle \sim 200$ collisions per pp bunch crossing. In addition, the physics programme associated with the HL-LHC aims to provide a total integrated luminosity of 3000 fb$^{-1}$ by 2035.

In order to cope with the substantial increase in noise and pile-up radiation resulting from the high-luminosity conditions, several upgrades to the ATLAS detector are targeted for the HL-LHC time scale starting in 2024. The planned upgrades span all detector systems within ATLAS, including the inner
tracker, trigger and data-acquisition systems, calorimeter systems, and the muon spectrometer. In fact, the inner tracker upgrades are planned to be the largest contribution to the upgrade scenario, comprising approximately 50% of the total upgrade cost [7]. In this note, the potential improvement in the VBF $H \to ZZ(\gamma*) \to 4l$ channel due to extended jet tracking was evaluated under three scenarios: The “Reference” scenario, which would extend tracking out to $|\eta| < 4.0$; the “Middle” scenario, which would extend jet tracking out to $|\eta| < 3.2$; and the “Low” scenario, wherein the same tracking coverage ($|\eta| < 2.7$) would exist as for the current ATLAS detector.

2. Monte Carlo samples

Hard-scattering events for both signal (VBF) and background (ggF) samples in this analysis were produced using the POWHEG BOX framework [8, 9], with the resultant Les Houches Event files interfaced to PYTHIA 8 [10] for parton showering. In particular, the signal Monte Carlo samples used in this analysis were generated using the POWHEG BOX VBF_H generator, which is a complete implementation of VBF Higgs-boson production at next-to-leading order (NLO) in QCD, in the POWHEG BOX framework. For this analysis, the generator was used to perform tree-level VBF Higgs-boson production calculations using exact NLO matrix elements.

Since the dominant background to VBF Higgs-boson production in the $H \to ZZ(\gamma*) \to 4l$ channel comes from ggF + 2 jets events, the POWHEG BOX HJJ generator [11] was used, which has NLO accuracy for gluon fusion-mediated $H + 2j$ calculations, i.e. gluon fusion-mediated Higgs-boson production with two final state jets. The use of this generator provides a notable improvement in jet modelling with respect to the gg_H generator [12], which was used in the previously-documented VBF $H \to 4l$ analysis [13]. In particular, the gg_H generator has NLO accuracy only for $0j$ production, therefore any two generated jets will necessarily come from the parton showering, rather than the hard scattering calculation. For both VBF and ggF samples CT10 PDFs were used for parton-level event generation while the AU2-CT10 tune was used in PYTHIA 8 for the showering [14, 15].

2.1. Truth object smearing and pile-up simulation

The four momenta of the relevant final state truth physics objects in this analysis – namely, jets, electrons, and muons – underwent energy and $p_T$ smearing, using techniques developed for physics studies presented at the ECFA 2013 [16] and 2014 [17] workshops. More details on the tools which perform resolution smearing, simulated trigger effects, primary vertex identification efficiency (referred to otherwise as “jet tracking efficiency”), etc. are described in the ATLAS Phase-II Upgrade Scoping Document [7]. The package was also used to insert simulated, reconstruction-level pile-up jets on an event-by-event basis to reproduce the expected jet kinematic distributions of high pile-up conditions for both $\langle \mu \rangle = 140$ and $200$.

The tools also provided simulated “tracking confirmation” for jets – in essence, the use of tracking information to distinguish jets originating from the primary vertex from those from pile-up vertices. As described in Section 1.1, the “Reference”, “Middle”, and “Low” tracking detector layout scenarios were all considered. It is worth noting that in the “Low” scenario, simulated jet tracking confirmation only extends to $|\eta| < 2.4$, due to degraded performance in the track-based pile-up rejection between $2.4 < |\eta| < 2.7$. 

3
The working point for track confirmation was chosen such that the pile-up jet efficiency was fixed at 2%, meaning that only 2% of generated pile-up jets were randomly kept. Conversely, the hard scattering jet efficiency varied with jet $p_T$ and $|\eta|$, with the event-by-event efficiency ranging between 80% to 100%. It is important to note that these tracking efficiency cuts are only applied for jets with $p_T < 100\text{GeV}$, and only within the tracking volume under study. As such, the overall pile-up rejection efficiency would change for each layout, depending on its tracking coverage in $|\eta|$.

The event-by-event combination of smeared truth jets and simulated pile-up jets proceeded as follows: For each event, the truth-level jets from the hard scattering interaction had $p_T$ and $|\eta|$-based smearing applied to their four momenta, then underwent the kinematic selection described in Table 1. A uniformly distributed random number on $[0,1]$ was generated for each jet to determine whether it passed or failed the efficiency cut. Afterwards, a random number of simulated pile-up jets were generated for each event, with a given pile-up jet being rejected if a uniform distributed random number fell above the tracking efficiency cut. Those simulated pile-up jets passing the tracking efficiency cut were also required to pass the same jet selection cuts as the smeared hard scatter jets. Once complete, the two jet containers were combined, and the resultant container sorted by descending $p_T$, since in practice, true jets from the hard scatter process will be reconstructed alongside pile-up jets.

The initial number of pile-up jets generated was derived by sampling from a Poisson distribution with its mean set to the average number of pile-up jets expected for $\langle \mu \rangle = 200$ conditions at $\sqrt{s} = 14\text{TeV}$. The jet kinematics were derived from template distributions of $p_T$ and $\eta$, using simulations of reconstructed jets at $\sqrt{s} = 14\text{TeV}$. After taking into account jet tracking efficiency cuts, and jet selection, with forward jet tracking simulation to $|\eta| < 4.0$ enabled, $\sim 63\%$ of events have no pile-up jets passing both tracking acceptance and jet selection. However, for those events with a non-zero number of pile-up jets passing all cuts, on average an extra $\langle n_{PU} \rangle = 1.6$ jets are introduced. For jet tracking confirmation up to $|\eta| < 3.2$, these figures become $45\%$ and $\langle n_{PU} \rangle = 2.4$, and for tracking confirmation up to $|\eta| < 2.4$, they are $19\%$ and $\langle n_{PU} \rangle = 3.3$.

3. Analysis

The Higgs-boson candidate selection, performed on electrons and muons, followed the prescription of the most recent $H \rightarrow ZZ^{(*)} \rightarrow 4l$ analysis for $7 + 8\text{TeV}$ [18], and is summarized in Table 1. Tables 2 and 3, respectively, show the statistics of the VBF and ggF hard scattering-only samples used before and after kinematic and dijet selection, in terms of weighted event counts, and weighted event counts scaled to the appropriate process cross section, i.e. $\sigma \times B(H \rightarrow 4l)$, at $I = 3000\text{ fb}^{-1}$.

In the most recent $H \rightarrow 4l$ Higgs-dijet selection note [13], it was found that $S/B \sim 7$ for the VBF-enriched category, where $S$ was defined as the summed contribution of all the Higgs-boson production mechanisms, and the background $B$ included $qqZZ$, $t\bar{t}$, and $Z$+jets events. Therefore, the choice was made to focus on separating dijet VBF and gluon fusion events, as the contamination in the VBF-enriched category by dijet gluon fusion events was $\sim 44\%$ for the $7 + 8\text{TeV}$ combined results. Furthermore, using simulations of the single lepton (muon or electron) trigger used for the $H \rightarrow 4l$ analysis, the trigger efficiency was found to be $\geq 95\%$ for each sample. The nominal lepton kinematic cuts, described in Table 1, were also kept for each tracking scenario considered, as studies into extending lepton tracking coverage, while improving overall $4l$ acceptance, did not demonstrate any improvement in the discrimination of VBF signal itself.
Event pre-selection

Lepton selection
- Require all truth electrons to have $E_T > 7$ GeV and $|\eta| < 2.47$
- Require all truth muons to have $p_T > 6$ GeV and $|\eta| < 2.7$

Event selection

$H \rightarrow 4l$ kinematic selection
- Require at least one quadruplet of leptons consisting of two pairs of same-flavour opposite-charge leptons fulfilling the following requirements:
  - $p_T$ thresholds for three leading leptons in the quadruplet 20, 15, and 10 GeV
  - Select best quadruplet to be that with the leading, subleading dilepton masses closest to the Z mass
  - Leading dilepton mass requirement $50$ GeV $< m_{12} < 106$ GeV
  - Subleading dilepton mass requirement $m_{\text{Threshold}} < m_{34} < 115$ GeV
  - Remove quadruplet if alternative same-flavour opposite-charge dilepton gives $m_{ll} < 5$ GeV or $\Delta R(l, l') < 0.10 \ (0.20)$ for same (different) flavour leptons in the quadruplet

Dijet selection
- Require $p_T > 30$ GeV and $|\eta| < 4.5$ for each jet
- Remove jets overlapping leptons within $\Delta R < 0.2$
- Dijet mass $m_{jj} > 130$ GeV

| Table 1: Kinematic selection employed to isolate truth-level $H \rightarrow 4l + 2j$ events. |

A boosted decision tree-based (BDT) approach was used to separate VBF from gluon fusion + 2 jet Higgs-boson production, where training was performed using VBF $H \rightarrow 4l$ events as “signal” and dijet gluon fusion $H \rightarrow 4l$ events as “background”. To determine the optimal set of training variables to distinguish the two processes, an iterative scan was performed over a number of sets of single jet, dijet, and Higgs-dijet kinematic variables which were found to have sensitivity to the different Higgs-boson production modes. The optimized permutation of variables was that which produced a BDT training that maximized the approximate significance $Z_0$ of VBF signal events (denoted $n_{\text{VBF}}$) over ggF events (denoted $n_{\text{ggF}}$), written as $Z_0 \approx n_{\text{VBF}}/\sqrt{n_{\text{VBF}} + n_{\text{ggF}}}$.

Ultimately, the set of kinematic variables which was found to optimize discrimination of VBF and gluon fusion production was,

$$m_{ij}, \Delta \eta_{ij}, p_T^{H jj}, p_T^{(\text{jet} \ 1)}, p_T^{(\text{jet} \ 2)}, \eta_{z H}^{\text{Zapp}}.$$  

Here, $p_T^{(\text{jet} \ 1)}$ and $p_T^{(\text{jet} \ 2)}$ are the transverse momenta of the leading and subleading jets, while $m_{ij}$ and $\Delta \eta_{ij}$ are the dijet mass and pseudorapidity separation, respectively. The kinematic variables related to the Higgs-dijet system are the Higgs-dijet system transverse momentum $p_T^{H jj}$ and $\eta_{z H}^{\text{Zapp}}$, defined as,

$$\eta_{z H}^{\text{Zapp}} = \eta_{H} - \left( \eta_{(\text{jet} \ 1)}, \eta_{(\text{jet} \ 2)} \right).$$  \hspace{1cm} (1)

Distributions of the optimal training variables are shown in Appendix A for each detector layout scenario considered in this note. It was found that when tracking coverage in the forward region decreased, the additional acceptance of events with $\geq 1$ forward pile-up jets in the gluon fusion and VBF samples increased appreciably, leading to a loss of discriminating power in commonly used VBF search variables like $m_{ij}$ and
<table>
<thead>
<tr>
<th>Sample</th>
<th>Full sample events</th>
<th>Full event weights</th>
<th>Weights after 4l cuts</th>
<th>After dijet cuts</th>
</tr>
</thead>
<tbody>
<tr>
<td>ggF (POWHEG)</td>
<td>4 919 000</td>
<td>2.23 × 10⁸</td>
<td>4.09 × 10⁷</td>
<td>3.25 × 10⁶</td>
</tr>
<tr>
<td>VBF (POWHEG)</td>
<td>9.8 × 10⁵</td>
<td>9.8 × 10⁵</td>
<td>193 036</td>
<td>95 768</td>
</tr>
</tbody>
</table>

Table 2: Weighted numbers of events before and after inclusive and dijet event selection. Note that these values are derived from the truth-level samples before smearing or pile-up simulation has been applied.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Expected events after 4l selection</th>
<th>Expected events after dijet selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>ggF (POWHEG)</td>
<td>7812</td>
<td>621</td>
</tr>
<tr>
<td>VBF (POWHEG)</td>
<td>720</td>
<td>358</td>
</tr>
</tbody>
</table>

Table 3: Expected event counts before and after inclusive and dijet event selection at \( I = 3000 \text{ fb}^{-1} \) and \( \sqrt{s} = 14 \text{ TeV} \), assuming a Higgs-boson mass of \( m_H = 125.5 \text{ GeV} \).

\( \Delta \eta jj \) which rely on the presence of a two-jet system highly separated in rapidity. However, discriminating variables that characterize the momentum balance and angular relationship between the Higgs-boson and dijet systems are less affected, as the additional information on the Higgs-boson kinematics contained in these quantities provides a greater degree of pile-up robustness.

In order to avoid the presence of large QCD scale variation uncertainties in the gluon fusion + 2 jet cross section, the BDT was prevented from making cuts on \( p_{THj} \) below 50 GeV. As calculated using the Stewart-Tackmann method [19], this requirement limits the absolute size of the potential QCD scale variation uncertainty to \( \sim 15\% \), as detailed in Section 3.3.

### 3.1. Statistical analysis

To assess the prospective significance (\( Z_0 \)) and error on the signal strength (\( \Delta \mu/\mu \)) resultant from discriminating dijet VBF and gluon fusion Higgs production, the negative log likelihood (NLL) minimization technique was used [20], such that the smeared samples with simulated pile-up were used to perform the NLL scans using the VBF \( H \to 4l \) sample as signal, and the ggF \( H \to 4l \) sample as background.

The statistical significance of the VBF excess over gluon fusion was computed using the background-only \( p \)-value (\( p_0 \)), which is evaluated using the test statistic \( q_\mu \),

\[
q_\mu = -2 \ln \frac{L(\mu, \theta)}{L(\hat{\mu}, \hat{\theta})},
\]

such that \( q_0 \) represents the background-only hypothesis, or \( \mu = \sigma / \sigma_{\text{SM}} = 0 \). Here, \( \mu \) is the signal strength parameter, while \( \theta \) represents the set of nuisance parameters for signal and background. The terms with single circumflexes (\( \hat{\mu} \) and \( \hat{\theta} \)) denote the unconditional maximum likelihood estimates of each parameter.

The value of \( p_0 \) is defined to be the probability to obtain a value of \( q_0 \) larger than the observed value under the background-only hypothesis. In particular, the value of \( p_0 \) can be expressed as,

\[
p_0 = \int_{q_{0,\text{obs}}}^{\infty} f(q_0 | 0, \hat{\theta}_0) \, dq_0.
\]
The $p_0$ value is then converted to the corresponding number of standard deviations ($\sigma$) in a one-sided Gaussian test. This equivalent formulation is referred to as the local significance, $Z_0$.

### 3.2. Definition of the VBF signal region

The signal region for this analysis was to be defined by cuts on the BDT classifier, with the metric of improvement being the resultant VBF $H \rightarrow 4l$ significance ($Z_0$). In particular, iterative likelihood scans were performed by incrementally tightening the cut on the BDT classifier between $0 < \text{(BDT response)} < 1$, where the requirement of $\text{(BDT response)} > 0$ was made to limit the effect of pile-up in the signal region. The step size in the scan was taken to be $\Delta\text{(BDT response)} = 0.1$, as finer step sizes were found to have a negligible effect on the final result. An integrated luminosity of $\mathcal{L} = 3000 \text{ fb}^{-1}$ was also assumed. At each step of the iteration, a signal region was defined by the events passing the BDT classifier cut, and $Z_0$ was recalculated, with the optimal choice of cut being that which maximized $Z_0$.

This iterative scan process was then repeated to determine a second optimal cut lying between the first cut and $\text{(BDT response)} > 0$. Therefore, three signal categories between $0 < \text{(BDT response)} < 1$, across which a simultaneous likelihood fit was performed. In particular, categories defined by $\text{BDTG} \in [0, 0.6, 0.8, 1]$ were found to be the optimal choice for all three detector layout scenarios for $\langle \mu \rangle = 200$. Here, the tightest category $0.8 < \text{BDTG} < 1$ corresponded to the highest purity of VBF signal, driving the final significance, while looser cuts corresponded to lower concentrations of the desired signal.

The BDT classifier distributions for VBF and ggF samples are shown in Figure 2 for each detector layout, for $\langle \mu \rangle = 200$ conditions and pile-up efficiency of 0.02. The presence of events with $\geq 1$ pile-up jets induces an appreciably large secondary peak around (BDT score) $\sim 0$ in both the VBF and ggF classifier distributions. Since the BDT used training samples with pile-up overlaid, the subset of events with $\geq 1$ pile-up jets would appear as a secondary bump in approximately the same score region for both VBF and ggF distributions. However, it is also clear that this feature is both reduced in size, and pushed further into the negative score region, with increased tracking coverage in $|\eta|$. This effect is a consequence of the lower pile-up contamination, particularly in the VBF signal sample. That is to say, the BDT is able to use a larger sample of true dijet VBF signal events for training, which improves rejection of both ggF background events, and any form of event with pile-up jets.

### 3.3. QCD scale variation uncertainty in gluon fusion + 2 jets

The uncertainty on the gluon fusion cross section due to missing higher-order calculation terms is typically evaluated by varying the QCD renormalization ($\mu_R$) and factorization ($\mu_F$) scales. For the purposes of simplicity, as well, the scales are often assumed to vary at the same rate, using a single scale variation parameter $\mu = \mu_F = \mu_R$, where $\mu$ is often varied as $\mu = m_H/2$ or $2 \times m_H$ in comparison to the nominal choice, $\mu = m_H$. Therefore, the symmetrized difference in dijet yield between the nominal and upward/downward scale variations gives the uncertainty on the cross section $\Delta \sigma$, with the relative uncertainty $\Delta \sigma / \sigma$ propagated through the statistical analysis.

However, problems arise in the 2-jet gluon fusion cross section calculation when tight cuts are imposed on variables which tend to isolate events with Born-order VBF-like kinematics, such as $\Delta \phi_{Hj}$ or $p_{Tj}$. In particular, the inclusive dijet gluon fusion phase space is separated into bins containing mainly 2-jet and $\geq 3$-jet events, and when scale variations are applied, the migration of events between these bins...
leads to error bands which no longer envelop the nominal distribution, hence provide unrealistically small estimates on the cross section uncertainty.

In order to provide more realistic estimates of $\Delta \sigma /\sigma$, the Stewart-Tackmann (S-T) method was employed to estimate the QCD scale variation uncertainty for the region of events passing a given cut on $p_{THjj}$ [19, 21]. In particular, the cross sectional variance in this region was estimated as $\Delta \sigma_{2}^2 = \Delta \sigma_{2}^2 + \Delta \sigma_{2}^2$, where $\Delta \sigma_{2}^2$ is the symmetrized uncertainty for the inclusive dijet region, and $\Delta \sigma_{2}^2$ is the symmetrized uncertainty for the region of events failing the cut on $p_{THjj}$. Figure 3 shows a distribution of the cumulative, exclusive 2-jet differential cross section as a function of the cut on $p_{THjj}$, along with the calculated and S-T error bands. As the cut on $p_{THjj}$ grows tighter, the relative uncertainty increases, reaching a value of $\Delta \sigma /\sigma \sim 100\%$ by requiring approximately $p_{THjj} < 20$ GeV. The relative uncertainties $\Delta \sigma_{2} /\sigma_{2}$ in bins of 10 GeV cuts are found in Table 4.
Figure 3: Cumulative differential cross sections for gluon fusion + 2 jet events in the exclusive 2-jet region, as a function of cut on $p_{THjj}$. The distribution of cumulative cross section values is shown with error bands calculated using the nominal QCD scale variation uncertainty approach (dotted lines), and the corrected Stewart-Tackmann approach.

<table>
<thead>
<tr>
<th>$p_{THjj}$ cut (GeV)</th>
<th>$\Delta\sigma_2/\sigma_2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>96.8</td>
</tr>
<tr>
<td>30</td>
<td>35.6</td>
</tr>
<tr>
<td>40</td>
<td>20.7</td>
</tr>
<tr>
<td>50</td>
<td>15.2</td>
</tr>
<tr>
<td>60</td>
<td>12.8</td>
</tr>
<tr>
<td>70</td>
<td>11.7</td>
</tr>
<tr>
<td>80</td>
<td>11.2</td>
</tr>
<tr>
<td>90</td>
<td>10.9</td>
</tr>
</tbody>
</table>

Table 4: Relative uncertainties $\Delta\sigma_2/\sigma_2$ in bins of 10 GeV rectangular cuts on $p_{THjj}$, computed using the Stewart-Tackmann method.

### 3.4. Flattening the distribution of $p_{THjj}$

Since the trained BDT in this analysis used $p_{THjj}$ as a discriminating variable, it was important to prevent cuts being made in the region of the $p_{THjj}$ phase space that could induce large S-T uncertainties. Otherwise, cuts that induce QCD scale variation uncertainties $O(100\%)$ would be possible in the ggF cross section. Therefore, in order to prevent the BDT from making such cuts, the distribution of $p_{THjj}$ was ‘flattened’ below a threshold value of 50 GeV, such that,

$$p'_{THjj} = \begin{cases} 
  p_{THjj} & \text{if } p_{THjj} > 50 \text{ GeV} \\
  50 \text{ GeV} & \text{if } p_{THjj} < 50 \text{ GeV} 
\end{cases}$$

(4)

Therefore, with a tight cut on the BDT classifier, the dijet $gg \rightarrow H$ QCD scale variation uncertainty would, at most, be equivalent to the Stewart-Tackmann uncertainty for such a rectangular cut on $p_{THjj}$. 
3.5. S-T uncertainties for multiple BDT categories

When defining a signal region by a cut on the BDT classifier, one implicitly defines non-linear cuts on the variables used to train the BDT. That is to say, unlike a typical rectangular cut-based analysis, one makes potentially several cuts on the same variable, which can be looser or tighter, depending on where the cut occurs in a given decision tree. In the context of an IR-sensitive variable like $p_{T}^{jj}$, this complicates the assessment of the S-T uncertainty described in Section 3.3, as one must consider that the uncertainty will either be enlarged or reduced, depending on the final $p_{T}^{jj}$ spectrum which defines the signal region.

The first step to mitigate the size of the uncertainty is to flatten the distribution of $p_{T}^{jj}$, as described in Section 3.4. Doing so will ultimately put a “ceiling” on the size of the relative uncertainty, preventing it from growing to 100% of the cross section. It is then possible to derive a more accurate assessment of the total uncertainty by employing the binned, post-flattening kinematic spectrum of $p_{T}^{jj}$ in the signal region, along with the bin-by-bin S-T uncertainties. In particular, for a histogram of the $p_{T}^{jj}$ kinematic spectrum, divided up into $N$ bins, one can define the total uncertainty as,

$$\Delta\sigma_{2}^{\text{Tot}} = \sum_{i,j=1}^{N} \text{cov}(\sigma_{2}^{i}, \sigma_{2}^{j}) \frac{\Delta\sigma_{2}^{i}}{\sigma_{2}^{i}} \frac{\Delta\sigma_{2}^{j}}{\sigma_{2}^{j}} \times h(x_{i})h(x_{j}).$$  \hspace{1cm} (5)$$

Here, $i, j$ are bin indices, and $h$ is the histogram of $p_{T}^{jj}$ for events within a particular range of BDT scores, with bin centres $x_{i}$. Accordingly, $\sigma_{2}^{i}$ is the 2-jet exclusive cross section for events passing a cut $p_{T}^{jj} < x_{i}$ (derived from samples generated with MCFM), and $\Delta\sigma_{2}^{i}$ is the S-T uncertainty (in units of $\sigma$) for the $i^{th}$ bin, as shown in Figure 3.

The covariance terms $\text{cov}(\sigma_{2}^{i}, \sigma_{2}^{j})$ represent the correlations of the exclusive cross sections defined by cuts at $x_{i}$ and $x_{j}$. In general, these are not explicitly known for $i \neq j$, and therefore must be modelled. For the purposes of this study, however, off-diagonal covariances were neglected, with the net effect being an overall larger total uncertainty. This calculation is, in effect, a weighted average of the different S-T uncertainties, dependent on how frequently the corresponding cuts are used in defining the signal region. Furthermore, in the limit of a histogram with only a single populated bin, one retrieves the equivalent S-T uncertainty for a rectangular cut.

4. Results

The expected signal significance $Z_{0}$ and signal strength uncertainty $\Delta\mu/\mu$, calculated for $\langle \mu \rangle = 200$ conditions and an integrated luminosity of $I = 3000$ fb$^{-1}$, are shown in Table 5 for each jet tracking scenario (Low, Middle, Reference). In addition, the expected number of events with (BDTG response) $> 0$ for each scenario is given for both the VBF and gluon fusion dijet samples, wherein the jet tracking confirmation (hence pile-up rejection) extends to $p_{T} < 100$ GeV.

Events were divided up into three categories defined by cuts on the BDT classifier, with bin edges [1, 0.8, 0.6, 0]. For these likelihood fits, only the statistical uncertainty factored into the calculations of $Z_{0}$ and $\Delta\mu/\mu$. Although with current theoretical knowledge, the QCD scale variation uncertainty for the selected gluon fusion + 2j events is non-trivial ($\Delta\sigma/\sigma \leq 15\%$), it does not change appreciably with detector layout, since the shape differences in $p_{T}^{jj}$ between VBF and ggF events are relatively insensitive
to pile-up. Therefore, it was initially neglected in the fits to change in VBF signal discrimination between the different jet tracking scenarios.

For each statistical test, a 4l candidate mass window of $115 < m_{4l} < 130$ GeV was employed to emulate conditions used in previous iterations of the VBF $H \rightarrow ZZ^{(*)} \rightarrow 4l$ analysis. It was found that, with respect to the Reference scenario, the signal significance $Z_0$ degraded by 5% and 14% when moving to the Middle and Low scenarios, respectively. Additionally, it was found that the signal strength uncertainty $\Delta \mu/\mu$ increased by 2% and 6% by moving to the Middle and Low layouts. In Table 7 the pile-up contamination of the VBF and ggF Higgs production signals is shown. The contamination is reduced significantly as one moves from the Low to the Middle and Reference layouts, resulting in a purer VBF signal, particularly in the bin corresponding to the tightest BDT score cut.

Estimates including the QCD scale variation uncertainty were also made using the Stewart-Tackmann approach [19], with the results shown in Table 5. In each table, it is evident that the final significance and $\Delta \mu/\mu$ values are degraded by approximately the same amount for each layout, leading to approximately the same relative improvement as before when extending tracking coverage. This result is somewhat expected, as the magnitude of the QCD scale variation uncertainty does not change appreciably between detector layouts.

A complimentary figure of merit to the signal strength uncertainty is the amount of additional integrated luminosity (hence data) required to achieve the levels of precision in $\Delta \mu/\mu$ shown in the Reference scenario at $I = 3000$ fb$^{-1}$. In Table 6, this figure is quoted with and without theoretical uncertainty for the Middle and Low scenarios. It is clear from these values that the extended jet tracking of the Reference scenario can lead to better results with significantly less data. In particular, when considering current estimates of the S-T uncertainty, an extra $\Delta I = 1$ ab$^{-1}$ would be necessary with the Low scenario to achieve the same result as the Reference case, which suggests that the latter scenario is close to the limits of possible precision for such detector operating conditions.

<table>
<thead>
<tr>
<th>Scoping scenario</th>
<th>VBF + 2j events</th>
<th>ggF + 2j events</th>
<th>$Z_0$ (VBF vs. ggF)</th>
<th>$\Delta \mu/\mu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td>237 (206)</td>
<td>324 (159)</td>
<td>11.4</td>
<td>±0.134</td>
</tr>
<tr>
<td>Middle</td>
<td>270 (205)</td>
<td>520 (177)</td>
<td>10.9</td>
<td>±0.137</td>
</tr>
<tr>
<td>Low</td>
<td>325 (198)</td>
<td>917 (211)</td>
<td>9.8</td>
<td>±0.142</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scoping scenario</th>
<th>VBF + 2j events</th>
<th>ggF + 2j events</th>
<th>$Z_0$ (VBF vs. ggF)</th>
<th>$\Delta \mu/\mu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td>237</td>
<td>324</td>
<td>7.6</td>
<td>±0.167</td>
</tr>
<tr>
<td>Middle</td>
<td>270</td>
<td>520</td>
<td>7.5</td>
<td>±0.174</td>
</tr>
<tr>
<td>Low</td>
<td>325</td>
<td>917</td>
<td>6.8</td>
<td>±0.186</td>
</tr>
</tbody>
</table>

Table 5: Expected signal and background event counts for different jet tracking coverage scenarios at $I = 3000$ fb$^{-1}$ and $\langle \mu \rangle = 200$ in the region (BDT score) $> 0$ and $115 < m_{4l} < 130$ GeV. Shown also is the VBF signal significance and signal strength precision for each scenario. Background is composed solely of ggF events, and scenarios are shown where only the statistical uncertainty (top), and also Stewart-Tackmann uncertainties (bottom) are considered in the fit. Bracketed terms represent the number of events in the signal region with two selected jets from the primary vertex.

Various small studies and refinements to the analysis methodology described in this note were implemented after the production of the Scoping Document. In particular, simulations of four lepton triggering effects,
<table>
<thead>
<tr>
<th>Scoping scenario</th>
<th>Without theo. unc.</th>
<th>With theo. unc.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\Delta \mu/\mu$</td>
<td>$\Delta \mu/\mu$</td>
</tr>
<tr>
<td>Reference</td>
<td>$\pm 0.134$</td>
<td>$\pm 0.167$</td>
</tr>
<tr>
<td>Middle</td>
<td>$\pm 0.137$</td>
<td>125</td>
</tr>
<tr>
<td>Low</td>
<td>$\pm 0.142$</td>
<td>425</td>
</tr>
</tbody>
</table>

Table 6: Additional integrated luminosity required to attain equivalent precision to Reference layout at $I = 3000 \text{ fb}^{-1}$, for cases with and without theoretical uncertainty.

<table>
<thead>
<tr>
<th>Scoping scenario</th>
<th>BDTG $&gt; 0.8$</th>
<th>$0.6 &lt; \text{BDTG} &lt; 0.8$</th>
<th>$0 &lt; \text{BDTG} &lt; 0.6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>VBF Sample</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reference</td>
<td>2.0</td>
<td>4.6</td>
<td>13.1</td>
</tr>
<tr>
<td>Middle</td>
<td>3.0</td>
<td>6.4</td>
<td>23.6</td>
</tr>
<tr>
<td>Low</td>
<td>5.2</td>
<td>12.0</td>
<td>38.7</td>
</tr>
<tr>
<td>ggF Sample</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reference</td>
<td>23.2</td>
<td>37.9</td>
<td>52.1</td>
</tr>
<tr>
<td>Middle</td>
<td>24.0</td>
<td>43.4</td>
<td>65.0</td>
</tr>
<tr>
<td>Low</td>
<td>41.2</td>
<td>59.4</td>
<td>76.2</td>
</tr>
</tbody>
</table>

Table 7: Pile-up contamination in the dijet VBF and gluon fusion samples for the various scoping scenarios in the 3 BDT score bins as defined in the text.

and estimates of $q\bar{q}ZZ$ background levels were implemented, mainly to justify the assumptions made in the original analysis which neglected these small, albeit non-trivial, effects. In addition, the statistical analysis was performed for a lower pile-up scenario ($\langle \mu \rangle = 140$), in order to evaluate the effects of extended jet tracking on more favourable luminosity conditions. Lastly, in order to account for the presence of $q\bar{q}ZZ$ background, the three-category fit was limited to a mass window of $120 < m_{4l} < 130 \text{ GeV}$, to emulate the range used in the $7 + 8 \text{ TeV}$ data analyses. Details of the analysis refinements, and assessments of their impact on the final results, are given below.

4.1. Lepton trigger simulation

It is expected that efficiency losses due to triggering should be a small effect in the $H \rightarrow 4l$ channel. In particular, the most recent $H \rightarrow 4l$ analysis performed on the combined $7 + 8 \text{ TeV}$ data sets showed a $\geq 97\%$ trigger efficiency for the $4\mu/2\mu2e/2e2\mu$ channels, and a close to 100% efficiency in the $4e$ channel [13]. However, in order to provide the closest possible emulation of data, trigger effects were integrated using simulated electron and muon efficiency tools to emulate the single lepton and dilepton triggers used in the $H \rightarrow 4l$ analysis. In particular, these effects were approximated on an event-by-event
basis by computing the total trigger inefficiency as,

\[ \epsilon_{4l} = \prod_{e, \mu} (1 - (\text{Single lep. trigger eff.})) \prod_{ee, \mu\mu} (1 - (\text{Dilepton trigger eff.})) \]

such that an event would be rejected if \( X < \epsilon_{4l} \), where \( X \) is a uniformly generated random number. As previously mentioned, using such trigger simulations resulted in an overall trigger efficiency of \( \geq 95\% \) for each sample.

4.2. Inclusion of \( qqZZ \) background

Estimates of the VBF discovery significance and \( \Delta\mu/\mu \) were also computed including truth Monte Carlo \( qq \rightarrow ZZ + X \) events simulated at \( \sqrt{s} = 14 \text{ TeV} \), where \( X \) represents 0, 1, or 2 hard scatter jets. The presence of such events (referred to henceforth as ‘\( qqZZ \)’) contribute a small but present source of non-resonant background for the \( H \rightarrow 4l \) decay, with smaller contributions from \( \bar{t}t \) and \( Z+\text{jets} \) events not considered in this analysis. The leptons and jets in the \( qqZZ \) events were subject to the same smearing as the signal Monte Carlo samples, and simulated pile-up jets were also introduced. BDT classifier distributions of Higgs-boson signal and \( qqZZ \) background are shown in Figure 4 for each detector layout at \( \langle \mu \rangle = 200 \) conditions. It is clear that the shapes of the BDT classifier distributions for \( qqZZ \) events in each layout are highly reminiscent of the ggF “background”.

In fact, the BDT classifier distributions for \( qqZZ \) events have a larger concentration of events near to -1, as events with \( \geq 1 \) pile-up jets make up the largest fraction (\( \sim 95\% \)) of the dijet \( qqZZ \) sample. Since the BDT is trained with pile-up included, and owing to the high pile-up contamination in ggF + 2 jets events, most of the \( qqZZ \) events passing selection will therefore strongly resemble background in their kinematics. Consequently, a cut of BDTG > 0, combined with the low production cross section in the nominal \( m_{4l} \) mass window of \( 120 < m_{4l} < 130 \text{ GeV} \), will lead to an overall small contribution of \( qqZZ \) events to the overall process background, as detailed in Sections 4.3 and 4.4. In effect, the reduction of the lower edge of the mass window from 110 to 120 GeV leads to a larger change in VBF significance than the contribution of \( qqZZ \) background.

4.3. Predictions for \( \langle \mu \rangle = 200 \) with trigger, \( qqZZ \) background

The statistical analysis was re-performed including trigger effects, \( qqZZ \) background, and the narrower mass window, with the results for \( \langle \mu \rangle = 200 \) conditions given in Table 8 with \( qqZZ \) background excluded, and included in Table 9, where values including the Stewart-Tackmann uncertainty are also shown. It was found that the relative improvement between tracking scenarios was overall similar to that without the additional physics effects, due to the smallness of their overall contribution to the final result. The largest change was due to the narrowing of the \( m_{4l} \) mass window which mainly affects the discovery significance \( Z_0 \) due to the smaller signal acceptance. In particular, it was found that moving from the Reference to Low scenarios led to a degradation of 14\% in \( Z_0 \), and an increase of \( \sim 8\% \) in \( \Delta\mu/\mu \), which is nearly the same result as was predicted without the additional background and trigger effects.
Figure 4: BDT classifier distributions for Low, Middle, and Reference detector layouts at $\langle \mu \rangle = 200$ and pile-up efficiency of 0.02. Distributions for both VBF and ggF $H \rightarrow 4l$ events are shown, along with $qqZZ$ background, for a mass window of $120 < m_{4l} < 130$ GeV.

<table>
<thead>
<tr>
<th>Scoping scenario</th>
<th>VBF + 2 $j$ events</th>
<th>ggF + 2 $j$ events</th>
<th>$Z_0$</th>
<th>$\Delta \mu / \mu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td>192</td>
<td>287</td>
<td>10.4</td>
<td>0.149</td>
</tr>
<tr>
<td>Middle</td>
<td>218</td>
<td>454</td>
<td>9.8</td>
<td>0.153</td>
</tr>
<tr>
<td>Low</td>
<td>259</td>
<td>803</td>
<td>8.9</td>
<td>0.161</td>
</tr>
</tbody>
</table>

Table 8: Expected signal and background event counts for different jet tracking coverage scenarios at $I = 3000$ fb$^{-1}$ and $\langle \mu \rangle = 200$ in the region (BDT score) $> 0$ and $120 < m_{4l} < 130$ GeV. Shown also is the VBF signal significance and signal strength precision for each scenario. Background is composed solely of ggF events, and only the statistical uncertainty is considered in the fit. Bracketed terms represent the number of events in the signal region with two selected jets from the primary vertex.
Table 9: Expected signal and background event counts for different jet tracking coverage scenarios at $I = 3000 \text{ fb}^{-1}$ and $\langle \mu \rangle = 200$ in the region (BDT score) $> 0$ and $120 < m_{4l} < 130 \text{ GeV}$. Shown also is the VBF signal significance and signal strength precision for each scenario. Background is composed of both ggF and $qqZZ$ events, and scenarios are shown where only the statistical uncertainty (top), and also Stewart-Tackmann uncertainties (bottom) are considered in the fit. Bracketed terms represent the number of events in the signal region with two selected jets from the primary vertex.

### 4.4. Predictions for $\langle \mu \rangle = 140$ with trigger, $qqZZ$ background

The BDT training and statistical analysis was also re-performed including trigger effects, $qqZZ$ background, and the narrower mass window for $\langle \mu \rangle = 140$ conditions. It is important to note that these results do not represent a fully accurate simulation of the experimental conditions; while the amount of simulated pile-up jets was reduced to match the expected number of $pp$ bunch crossings, the tracking confirmation employed was the same as for $\langle \mu \rangle = 200$ conditions. Therefore, the estimates presented are likely more pessimistic than expected from a more realistic reproduction of operating conditions.

The results for these tests are given in Table 10, where values including the Stewart-Tackmann uncertainty are also shown. Plots of the BDT response for each detector layout are also shown in Figure 5. Due to the lower overall amount of forward pile-up, the performance improvement gained by extended tracking is smaller between equivalent tracking scenarios at $\langle \mu \rangle = 200$ and 140. In particular, it was found that moving from the Reference to Low tracking scenarios under $\langle \mu \rangle = 140$ conditions led to a degradation of 8% in $Z_0$, and an increase of 2% in $\Delta \mu/\mu$, in contrast to the 14% decrease in $Z_0$ and 8% increase seen in $\Delta \mu/\mu$ for $\langle \mu \rangle = 200$.

### 4.5. Predictions for $\langle \mu \rangle = 140$ with $I = 2 \text{ ab}^{-1}$ with trigger, $qqZZ$ background

Using the BDT-based analysis and statistical tests defined in Section 4.4 for $\langle \mu \rangle = 140$ conditions, the VBF significance and $\Delta \mu/\mu$ were also computed for an integrated luminosity of $I = 2 \text{ ab}^{-1}$, a scenario which is possible if the running time of the HL-LHC is kept constant. The results are shown in Table 11, where values including the Stewart-Tackmann uncertainty are also shown.
5. Conclusions

The impact of increased jet tracking coverage in the forward direction has been investigated in terms of the separation of vector-boson fusion and gluon fusion-mediated Higgs-boson decays to four leptons in association with two jets. It was found that for simulated HL-LHC-like conditions, at an integrated luminosity of $\mathcal{L} = 3000 \text{ fb}^{-1}$ and an average number of pile-up interactions of $\langle \mu \rangle = 200$, a reduction of the range of jet tracking confirmation from $|\eta| < 4.0$ to $|\eta| < 3.2$ reduces the expected significance $Z_0$ by 4% and increases the relative uncertainty on the signal strength $\Delta \mu / \mu$ by 3%, while further limiting the range to $|\eta| < 2.4$ reduces the significance by 14%, and increases $\Delta \mu / \mu$ by 9%. If considering the $\sim 15\%$ dijet gluon fusion QCD theoretical uncertainty calculated using the Stewart-Tackmann method, the increase in $\Delta \mu / \mu$ is approximately 14% in reducing coverage from $|\eta| < 4.0$ to $|\eta| < 2.4$.

Although the VBF production mechanism will likely have been discovered in the $H \rightarrow 4l$ channel by the timescale of the HL-LHC, significantly high signal event counts will still be required to perform precision measurements of Higgs-boson properties. Of particular importance are measurements of differential cross sections of VBF production, which justify the emphasis given in this analysis on obtaining a VBF
Table 10: Expected signal and background event counts for different jet tracking coverage scenarios at $I = 3000 \text{ fb}^{-1}$ and $\langle \mu \rangle = 140$ in the region (BDT score) $> 0$ and $120 < m_4l < 130 \text{ GeV}$. Shown also is the VBF signal significance and signal strength precision for each scenario. Background is composed of both ggF and qqZZ events, and scenarios are shown where only the statistical uncertainty (top), and also Stewart-Tackmann uncertainties (bottom) are considered in the fit. Bracketed terms represent the number of events in the signal region with two selected jets from the primary vertex.

Table 11: Expected signal and background event counts for different jet tracking coverage scenarios at $I = 2000 \text{ fb}^{-1}$ and $\langle \mu \rangle = 140$ in the region (BDT score) $> 0$ and $120 < m_4l < 130 \text{ GeV}$. Shown also is the VBF signal significance and signal strength precision for each scenario. Background is composed of both ggF and qqZZ events, and scenarios are shown where only the statistical uncertainty (top), and also Stewart-Tackmann uncertainties (bottom) are considered in the fit.

signal region with the smallest possible contamination. This focus is also relevant to the measurement of Higgs-boson couplings and the potential probing of BSM physics in the Higgs sector, both of which will be increasingly relevant as more high luminosity data is acquired.

These measurements will require a clear separation of the VBF and ggF production mechanisms, and high suppression of pile-up-related background, both of which motivated the development of this BDT-based analysis. Techniques to probe the Higgs-boson couplings are also being developed in parallel, which will be tested both with Run-II data and using simulation with HL-LHC conditions. Such work is beyond the scope of this document, but of potential relevance to future studies of Higgs-boson production at the HL-LHC.
Appendix

A. Training variable kinematics for different detector layouts

The training variable kinematics, as defined in Section 3, are shown for each detector layout in Figures 6, 7, and 8 for Low, Middle, and Reference layouts, where jet tracking confirmation is extended to $p_T < 100$ GeV. The effects of extended jet tracking in pile-up rejection are evident from these three distributions, as reducing tracking coverage leads to an increase in the acceptance of forward jets. As a consequence of this increased acceptance, kinematic variables which are highly correlated with the presence of forward jets, namely $m_{jj}$ and $\Delta \eta_{jj}$, see a significant decrease in shape separation between VBF and ggF events.

Figure 6: BDT training variable kinematics for the Low detector layout. The solid blue distributions represent VBF-mediated Higgs-boson events, while the hatched red distributions represent 2-jet gluon fusion-mediated Higgs-boson events.
Figure 7: BDT training variable kinematics for the Middle detector layout. The solid blue distributions represent VBF-mediated Higgs-boson events, while the hatched red distributions represent 2-jet gluon fusion-mediated Higgs-boson events.

Figure 8: BDT training variable kinematics for the Reference detector layout. The solid blue distributions represent VBF-mediated Higgs-boson events, while the hatched red distributions represent 2-jet gluon fusion-mediated Higgs-boson events.
References


[17] ECFA High Luminosity LHC Experiments Workshop - Physics and Technology Developments
Summary submitted to ECFA2014. 96th Plenary ECFA meeting, (2014),
url: https://cds.cern.ch/record/1983664.

[18] ATLAS Collaboration, ‘Event selection and background estimation for the measurement of the
properties of the Higgs particle in the four lepton decay channel with the ATLAS detector’,

[19] S. Gangal and F. J. Tackmann,
Next-to-leading-order uncertainties in Higgs + 2 jets from gluon fusion,
Phys. Rev. D 87 (9 2013) 093008.


[21] I. W. Stewart and F. J. Tackmann,
Theory Uncertainties for Higgs and Other Searches Using Jet Bins,