PROCEEDINGS OF SCIENCE

Boosting Higgs Pair Production in the Final State $b\bar{b}b\bar{b}$ With Multivariate Techniques

D. Bortoletto*
University of Oxford
E-mail: daniela.bortoletto@cern.chl

The measurement of the Higgs pair production in the $b\bar{b}b\bar{b}$ final state at the Large Hadron Collider (LHC) and its future high-luminosity upgrade, the HL–LHC, is sensitive to new physics beyond the Standard Model and is critical for the extraction of the Higgs self-coupling. Herein, we present novel analysis that implements new multivariate techniques and optimises all possible Higgs decay topologies. The effect of pileup and all relevant backgrounds are included. We obtain a signal significance of $S/\sqrt{B} \approx 3$ for an integrated luminosity of $\mathscr{L} = 3ab^{-1}$.

38th International Conference on High Energy Physics
3-10 August 2016
Chicago, USA

*Speaker.

PoS(ICHEP2016)1032

1. Introduction

The measurement of Higgs pair production is one of the most important physics goals of the Large Hadron Collider (LHC) and its future high-luminosity upgrade, the HL-LHC which is expected to collect around $3ab^{-1}$ of data at or near its design energy of $\sqrt{s} = 14$ TeV [1, 2]. Current experimental results have shed light upon single Higgs production and therefore explored the minimum of the electroweak symmetry breaking potential. The production of Higgs bosons pairs is the golden channel to study the SM Higgs potential and to directly access the Higgs self-coupling. Furthermore di-Higgs production could be a sensitive probe for physics beyond the standard model. The measurement is extremely challenging since the Standard model (SM) cross-section for Higgs pair production in gluon fusion, which is the dominant production mechanism, is only 40 fb at next-to-next-to-leading order (NNLO) in QCD at $\sqrt{s} = 14$ TeV. Therefore the $b\bar{b}b\bar{b}$ final state is of particular interest for this measurement since the $H \to b\bar{b}$ with BR($H \to b\bar{b} = 0.57$) [3] is the largest of all Higgs boson decay channels. Unfortunately the $HH \to b\bar{b}b\bar{b}$ final state is particularly challenging due to the large the the background from multi-jet production. Previous studies of Higgs pair production in this final state [4, 5] concluded that a signal significance $S/\sqrt{B} \approx 2$ could be reached for an integrated luminosity of $\mathscr{L} = 3ab^{-1}$ at $\sqrt{s} = 14$ TeV. We have conducted a new feasibility study [6] bringing a novel strategy for measuring Higgs pair production from gluon fusion in the $b\bar{b}b\bar{b}$ final state that is based on the combination of a traditional cut-based approach and multivariate analysis (MVA) techniques. The new method uses all possible Higgs decay topologies to optimise the sensitivity for this measurement. In addition all relevant backgrounds, including $2b2j$ multi-jet, which was previously overlooked, is now included since it yields a non-negligible contribution due to the mis-identification of light-quarks and gluon jets. Our analysis strategy is also optimised for the HL-LHC environment where we expect a high number of multiple inelastic events in the same bunch crossing which is denoted as pile-up.

2. Modelling of signal and background processes

Higgs pair production is simulated at leading order (LO) with MADGRAPH5_AMCATNLO [7] with a dedicated model for double Higgs boson production via gluon-fusion. Mass effects from the exact form factors for top-quark triangle and box loops are taken into account [8]. The simulation is performed in the four-flavour scheme ($n_f = 4$). The renormalisation and factorisation scales are chosen to be $m_R = m_F = H_T = 2$. The NNPDF 3.0 $n_f = 4$ LO PDF set [9] with $\alpha_s(m_Z^2) = 0.118$, as provided in LHAPDF6 [10], is used. The simulated cross-section is rescaled to the total inclusive cross-section calculated at NNLO with corrections from soft-gluon resummation up to next-to-next-to-leading logarithmic accuracy (NNLL) [11]. The parton-level events are showered using PYTHIA8 [12] v8.201 with the Monash tune [13] and the NNPDF 2.3 LO PDF set [14]. The backgrounds from QCD multi-jet production are generated at LO with SHERPA[15] v2.1.1 with the same PDF set and scales as used for the signal processes. We consider QCD $4b$ multi-jet production, as well as QCD $2b2j$ and $4j$ production, and top quark pair production with fully hadronic final states. Multijet $4b$, $2b2j$ processes are simulated with MADGRAPH5_AMCATNLO while BLACKHAT [16] is used for $4j$. The cross sections for all these processes are then rescaled to NLO predictions. The simulated cross-section for the $t\bar{t}$ production is rescaled to NNLO+NNLL precision[17]. The values for the k-factors used in our analysis are documented in Ref. [6]. Other background processes, such as single Higgs production in the $Z(\to b\bar{b})H(\to b\bar{b})$, and $t\bar{t}H(\to b\bar{b})$ channels, and electroweak backgrounds, such as e.g $Z(\to b\bar{b})b\bar{b}$, have been simulated with MADGRAPH5_AMCATNLO but found to yield a significantly smaller contribution in the signal regions than the QCD backgrounds and therefore are not included in this analysis. Pile-up (PU) is simulated by overlaying n_{PU} Minimum Bias events, generated with PYTHIA8, on signal and background events. SOFTKILLER [18], as implemented in FASTJET [19], is used to subtract PU contaminations. We found the combined signal significances are similar for $n_{PU} = 80$ and $n_{PU} = 150$ and therefore we adopt the former as our baseline.

3. Event Selection and Reconstruction

The final state particles obtained after the parton shower are clustered with jet reconstruction algorithms implemented in FASTJET v3.1.0. Small-R jets are reconstructed using the anti-k_t algorithm [20] with size parameter R=0.4. Only small-R jets with transverse momentum $p_T^{jet} > 40$ GeV and pseudorapidity $|\eta^{jet}| < 2.5$ are considered in the analysis. Large-R jets are reconstructed with the same algorithm with $R = 1.0$. A trimming [21] procedure with parameters $R_{trim} = 0.2$ and $p_T^{frac} = 0.05$ is applied to mitigate the effects of PU on the large-R jets properties. Large-R jets are required to satisfy the requirements $p_T^{jet} > 200$ GeV and $|\eta^{jet}| < 2.0$. In addition, they must satisfy the BDRS mass-drop tagger (MDT)[22] conditions with parameters $\mu_{mdt} = 0.67$ and $y_{mdt} = 0.09$.

The identification of jets from b-quarks (b-tagging) plays a key role in achieving a high signal purity in the $b\bar{b}b\bar{b}$ final state. A small-R jet is b-tagged with a probability of $f_b = 0.80$ if there is at least one b-quark with $p_T > 15$ GeV among its constituents. A small-R jet with no b-quark but at least one c-quark with $p_T > 15$ GeV among its constituents is mis-tagged as a b-jet with probability $f_c = 0.10$. The mis-tag probability for jets from light quarks (d,u,s) and gluons is assumed to be $f_l = 0.01$. Only jets that have at least one constituent with $p_T > 15$ GeV can be b-tagged and only the four leading small-R jets in an event are considered for b-tagging to reduce the background due to mis-tagged jets. Large-R jets are considered b-tagged if they have at least two matching anti-k_t $R = 0.3$ subjets that are b-tagged by the same criteria as used for small-R jets [23].

Events are selected into three mutually exclusive categories based on the Higgs decay topology. First, events with at least two selected large-R jets, the leading two of which are taken as the Higgs candidates, are assigned to the boosted category. Events with exactly one selected large-R jet, taken as the leading Higgs candidate, and at least two b-tagged small-R jets with an angular separation $\Delta R > 1.2$ from the large-R jet are assigned to the intermediate category. The remaining events may be classified into the resolved category if they contain at least four b-tagged small-R jets. In this case, the two Higgs candidates are chosen to be the two dijet combinations with the smallest mass difference. In all categories, only events for which the invariant mass of each Higgs candidate lies within a symmetric mass window of width 80 GeV around 125 GeV are considered. The selection criteria are loose since the final selection is determined by the a Multi Variate Analysis (MVA).

PoS(ICHEP2016)1032

4. Multivariate analysis

All selected events are processed through a multi-layer feed-forward artificial neural network (ANN), known as a perceptron or deep neural network, to optimise the separation between signal and background. The optimisation is performed separately for boosted, intermediate, and resolved events. The ANN follows a $N_{var} \times 5 \times 3 \times 1$ architecture, where N_{var} denotes the number of input variables. Input variables to the ANN include kinematic properties of the reconstructed Higgs boson candidates as well as a number of substructure variables in the case of the intermediate and boosted categories. The full list of input variables, along with their relevance for the discrimination between signal and background, as obtained from the trained ANNs in a fully automated way, are found in Ref. [6]. Events are classified as signal or background based on a cut on the ANN output denoted as y_{cut}, which is optimised according to the signal significance of the event topology.

5. Results

The number of signal and background events obtained in the three categories after applying a selection on the ANN output are given in Table 1, along with the corresponding signal significances $S/\sqrt{B}$ for an integrated luminosity of $\mathscr{L} = 3ab^{-1}$. The signal significance for the combination of the three categories is derived by adding those for the individual categories in quadrature.

Category	PU strategy	y_{cut}	N_{signal}	$N_{background}$	$S/\sqrt{B}$
Boosted	no PU	0.80	290	1.2×10^4	2.7
Boosted	PU80 +SK +Trim	0.80	290	3.7×10^4	1.5
Intermediate	no PU	0.75	130	3.1×10^3	2.3
Intermediate	PU80 +SK +Trim	075	140	5.6×10^3	1.9
Resolved	no PU	0.50	630	1.1×10^5	1.9
Resolved	PU80 +SK +Trim	0.60	640	1.0×10^5	2.0
Combined	no PU				4.0
Combined	PU80 +SK +Trim				3.1

Table 1: Number of signal, background events, and $S/\sqrt{B}$ in the three event categories after applying a selection on the ANN output y_{cut} for an integrated luminosity of $\mathscr{L} = 3ab^{-1}$. The results are quoted for the baseline analysis with nPU = 80 (PU80+SK+Trim) and for no pile-up. The combined results are also provided.

For the baseline scenario, PU80+SK+Trim, with all backgrounds included, we obtain a signal significance of $S/\sqrt{B} \approx 3.1$ for the combination of all categories. A signal significance of $S/\sqrt{B} \approx 4.7$, close to the threshold for claiming observation, is found if all backgrounds but the irreducible $4b$ component are neglected, indicating that a reduction of the mistag rates for jets from light and charm quarks is important. Further significance improvements can be achieved through more effective PU mitigation techniques, as indicated by the comparison between the baseline result and the one obtained when the PU is neglected. We have also found that the sensitivity of the analysis depends strongly on the Higgs mass resolution. Hence jet energy and mass resolution, should be maintained or improved in the high-PU environment of the HL-LHC .

6. Conclusions

We have presented a novel strategy for measuring Higgs pair production in the $b\bar{b}b\bar{b}$ final state that employs MVA techniques and uses optimally different signal topologies. The analysis demonstrates that the use of PU reduction methods, such as trimming and SOFTKILLER is imperative in the environment of the HL-LHC. The resulting signal significance of $S/\sqrt{B} \approx 3.1$ represents a considerable improvement over previous results, which did not include PU effects. We also show that improvements in the b-tagging performance and PU mitigation could lead to the observation of Higgs pair production in the $b\bar{b}b\bar{b}$ final state alone if a low enough trigger threshold can be maintained during the HL-LHC.

References

[1] ATLAS Collaboration, arXiv:1307.7292 [hep-ex].

[2] CMS Collaboration, arXiv:1307.7135 [hep-ex].

[3] S. Dittmaier et al., arXiv:1201.3084 [hep-ph].

[4] D. Wardrope et al., Eur. Phys. J. C 75 (2015) no.5, 219 [arXiv:1410.2794].

[5] D. E. Ferreira de Lima et al., JHEP 1408 (2014) 030 [arXiv:1404.7139].

[6] J. K. Behr et al., Eur. Phys. J. C 76 (2016) no.7, 386 [arXiv:1512.08928].

[7] J. Alwall et al., JHEP 1407 (2014) 079 [arXiv:1405.0301]; F. Maltoni et al., JHEP 1411 (2014) 079 [arXiv:1408.6542].

[8] T. Plehn et al., Nucl. Phys. B 479 (1996) 46 [hep-ph/9603205].

[9] R. D. Ball et al. [NNPDF Collaboration], JHEP 1504 (2015) 040 [arXiv:1410.8849].

[10] A. Buckley et al., Eur. Phys. J. C 75 (2015) 132 [arXiv:1412.7420].

[11] D. de Florian and J. Mazzitelli, Phys. Rev. Lett. 111 (2013) 201801 [arXiv:1309.6594]; D. de Florian and J. Mazzitelli, JHEP 1509 (2015) 053 [arXiv:1505.07122].

[12] T. Sjostrand et al., Comput. Phys. Commun. 178 (2008) 852 [arXiv:0710.3820];T. Sjostrand et al., Comput. Phys. Commun. 191 (2015) 159 [arXiv:1410.3012].

[13] P. Skands et al., Eur. Phys. J. C 74 (2014) no.8, 3024 [arXiv:1404.5630].

[14] R. D. Ball et al., Nucl. Phys. B 867 (2013) 244 [arXiv:1207.1303]; R. D. Ball et al. [NNPDF Collaboration], Nucl. Phys. B 877 (2013) 290 [arXiv:1308.0598].

[15] T. Gleisberg et al., JHEP 02 (2009) 007, [arXiv:0811.4622].

[16] Z. Bern et al., Phys.Rev.Lett. 109 (2012) 042001, [arXiv:1112.3940].

[17] M. Czakon et al., Phys. Rev. Lett. 110 (2013) 252004, [arXiv:1303.6254].

[18] M. Cacciari et al., Eur. Phys. J C 72(2015) 59.

[19] M. Cacciari et al., Eur. Phys. J. C72 (2012) 1896 arXiv:1111.6097 [hep-ph].

[20] M. Cacciari et al., JHEP 0804 (2008) 063, [arXiv:0802.1189].

[21] D. Krohn et al., JHEP 02 (2010) 084, [arXiv:0912.1342].

[22] J. M. Butterworth et al., Phys.Rev.Lett. 100 (2008) 242001, [arXiv:0802.2470].

[23] ATLAS Collaboration, Eur. Phys. J. C 75 (2015) no.9, 412 [arXiv:1506.00285].