IL NUOVO CIMENTO **42 C** (2019) 161
DOI 10.1393/ncc/i2019-19161-3

Colloquia: IFAE 2018

Study of the modelling of $W/Z + HF$ as background for $VH(H \to b\bar{b})$ at ATLAS

G. Cabras([1])([2])

([1]) *INFN, Sezione di Bologna - Bologna, Italy*
([2]) *Alma Mater Studiorum, Dipartimento di Fisica, Università di Bologna - Bologna, Italy*

received 31 January 2019

Summary. — The Higgs boson properties has been measured at the ATLAS experiment since its discovery, obtaining a good agreement with the Standard Model predictions. Evidence of $H \to b\bar{b}$ is crucial for constraining the overall Higgs boson decay width and to confirm the Yukawa coupling to quark sector. The most promising production mode is the VH, where the Higgs boson is in association with a vector boson. The vector boson decaying into leptons helps in the extraction of the signal from the multijet background which dominates the $H \to b\bar{b}$. One of the main sources of background for the Standard Model $VH(H \to b\bar{b})$ search is the production of V ($V = W$ or Z) bosons in the association with jets. This contribution describes the ATLAS simulation setup used to model this irriducible background in proton-proton collisions at centre-of-mass energy of 13 TeV. Several Monte Carlo generators are studied in different phase-space regions and compared to unfolded data distribution, with additional requirements on the heavy-flavour content of the associated jets.

1. – Analysis strategy

The Higgs boson was discovered by ATLAS [1] and CMS in 2012. It is important to measure its couplings to the SM particles to find out if the 125 GeV boson is the Standard Model Higgs. The decay with the largest branching ratio is $H \to b\bar{b}$, 58% in the Standard Model but it has been only recently observed with a low statistical significance. The main problem is that if it is produced in gluon fusion, the signal is overwhelmed by multi-jet QCD background. To improve the signal-to-background ratio, the most promising production mode to study is the associated production with a vector boson, VH, thanks to the clear signature due to leptonic decay of V.

There are three signal (fig. 1) processes depending on the number of charged leptons in the final state: 0, 1 and 2. The kinematic phase space is splitted by the number of charged leptons, jets, b-jets and the transverse momentum of the vector boson V (in the

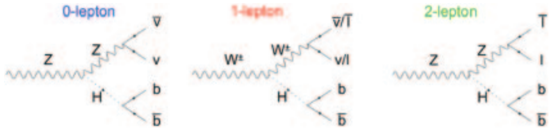


Fig. 1. – Signal processes.

TABLE I. – *Event selection.*

Quantity	0-lepton	1-lepton	2-lepton
Number of jets	2, 3	2, 3	2, $\geqslant 3$
p_T^V region 1	$\geqslant 150$ GeV	$\geqslant 150$ GeV	$\geqslant 150$ GeV
p_T^V region 2	–	–	$\geqslant 75$ GeV

case of the 0-leptons channel, it corresponds to the missing E_T), see table I. In all the selection the presence of exactly 2 tagged b-jets is required. The invariant mass $m_{b\bar{b}}$ of the two b-jets is the best signal-to-background discriminant, since the signal is peaked around 125 GeV. To better discriminate, a multivariate tecnique is used, training a boosted decision tree (BDT), using as input $m_{b\bar{b}}$ and other kinematic event variables.

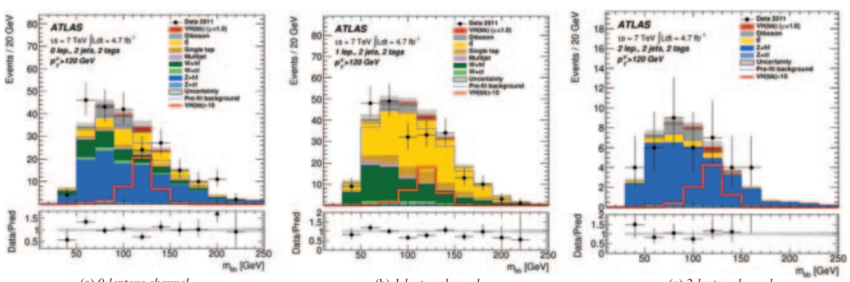
The signal is then extracted with a Profile Likelihood Fit of the BDT output [2].

2. – Background modelling: $V+$ jets

The main sources of background are $t\bar{t}$ and $V+$ jets, as can be seen in fig. 2.

While the $t\bar{t}$ source is reducible improving b-tagging techniques, $V+$ jets source is irreducible. Modelling uncertainties cover three areas: normalisation, acceptance differences between analysis regions and differential distributions of the most important kinematic variables. The strategy is to compare different Monte Carlo generators and, where it is possibile, to perform data-driven studies.

The nominal Monte Carlo used is Sherpa 2.2.1 (5 Flavour Number Scheme), which does the matrix-element generation and the parton showering with ME+PS@NLO. The $V+0$, 1, 2 jets is generated at NLO, while $V+3$ and 4 jets are generated at LO. As can be seen from fig. 3, which shows the comparison with the 7 TeV data, all the genera-



(a) *0-leptons channel.* (b) *1-lepton channel.* (c) *2-lepton channel.*

Fig. 2. – Main sources of background for the three analysis channels.

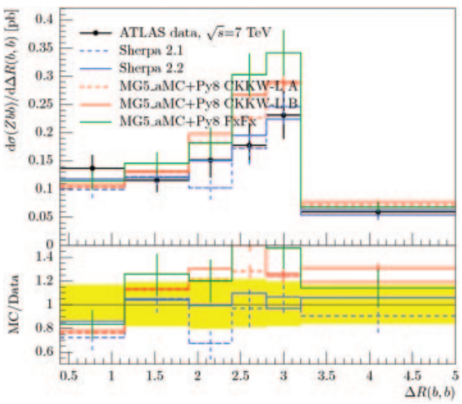
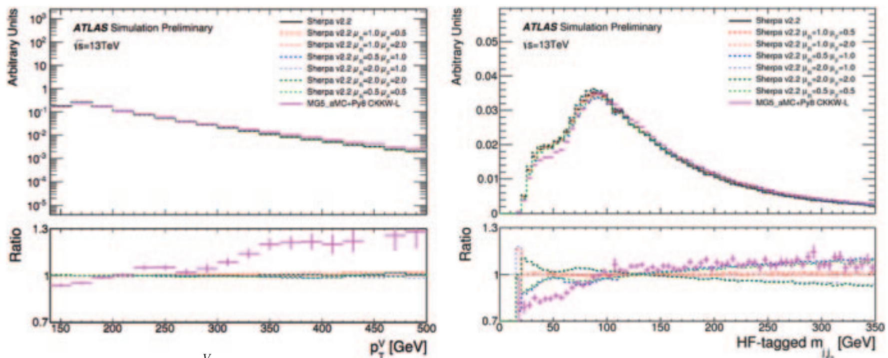


Fig. 3. – Comparison between Sherpa, MadGraph and data at $\sqrt{s} = 7\,\text{TeV}$.

tors reasonably well describe the rates and shapes but Sherpa 2.1 slightly under-predicts the rate of high-p_T b-jets, which is improved in Sherpa 2.2, [3]. The alternative Monte Carlo sample is MadGraph 5 with Pythia 8 for the modelling of the parton shower and the underlying event. MadGraph is at LO up to 4 jets, while the higher jet multiplicities are modelled by the parton shower algorithm.

The normalisation and acceptance uncertainties are calculated by adding the difference between the nominal Sherpa 2.2.1 sample and its systematic variations in quadrature: renormalisation scale by factors of 0.5 and 2; factorisation scale by factors of 0.5 and 2; CKKW merging scale from 30 GeV to 15 GeV; parton shower/resummation scale by factors of 0.5 and 2. The difference between the nominal and alternative sample is added in quadrature to the rest to obtain the total uncertainty.

2˙1. *W+ jets.* – For the $W+$ jets background due to the limited number of events in the dedicated control region, normalisation, acceptance and shape systematic uncertainties are estimated with internal studies on Sherpa and its comparison with MadGraph. The systematic uncertainties are dominated by the comparison of the two different Monte



(a) p_T^V for W+jets. (b) m_{bb} for Z+jets.

Fig. 4. – $V+$ jets: shape comparisons of the distribution of the transverse momentum of the vector boson for $W+$ jets (left) and the angular separation between the two b-jets (right) for $Z+$ jets.

Carlo samples. The difference between MadGraph and Sherpa is larger than Sherpa scale variations. The systematic uncertainty which has the major impact on the signal extraction is the one associated to p_T^V, due to the mismodelling at high transverse momentum, fig. 4(a) [4].

2·2. *Z+ jets*. – While for normalisation and acceptance uncertainties the approach is the same of W+jets, uncertainties in the shapes of m_{bb} and p_T^V distributions are estimated in this case by comparing Z+ jets background to data in signal regions with a high Z+ jets purity (1- and 2-tag regions for the 2-lepton channel), excluding the m_{bb} region around the Higgs boson mass and removing the $t\bar{t}$ contamination by the requirement that $E_T^{miss}/\sqrt{S_T} < 3.5\sqrt{\text{GeV}}$. The systematic uncertainty which has the major impact on the signal extraction is the one associated to $m_{b\bar{b}}$, due to the mismodelling at low energies, fig. 4(b) [4].

3. – Conclusions

The measured signal strength with respect to the Standard Model prediction is $mu = 1.21^{+0.24}_{-0.11}$ (stat.) $^{+0.22}_{-0.19}$. A possibility to improve the results is to have control regions with more statistics and to tune Monte Carlo data with the measurement of the V+ jets cross-section at 13 TeV.

REFERENCES

[1] The ATLAS Collaboration, *JINST*, **3** (2008) S08003.
[2] The ATLAS Collaboration, *JHEP*, **12** (2017) 024.
[3] The ATLAS Collaboration, *Monte Carlo Generators for the Production of a W or Z/γ* Boson in Association with Jets at ATLAS in Run 2*, ATL-PHYS-PUB-2016-003.
[4] The ATLAS Collaboration, *ATLAS simulation of boson plus jets processes in Run 2*, ATL-PHYS-PUB-2017-006.