Chapter 3
Dataset and Simulated Event Samples

This chapter outlines the dataset used by the analysis described in this thesis and the trigger strategy adopted to collect data. The signal and background processes contributing to the analysis are described as well, together with the Monte Carlo generators used to model them.

3.1 Dataset and Trigger

The data used in the analysis described in this thesis were collected at a centre-of-mass energy $\sqrt{s} = 13$ TeV during the 2015–2018 running period, corresponding to an integrated luminosity of 139 fb^{-1}. During the Run 2, the mean number of interactions per bunch crossing $< \mu >$ is about 34. Events used in data analysis are selected only if all the relevant systems of the ATLAS detector are known to be in good operation condition.

In the analysis the events are categorized in the 0-lepton, 1-lepton and 2-lepton channels, aiming at selecting $ZH \rightarrow \nu\nu b\bar{b}$, $WH \rightarrow l\nu b\bar{b}$ and $ZH \rightarrow llb\bar{b}$ events. The trigger algorithms used to select such events exploit the signatures of the leptonic decay of the V boson, depending on the leptonic channel. Events in the 0-lepton channel are selected using a missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}}$ trigger because the events are characterized by the presence of two neutrinos from the Z boson decay. Selecting events that contain invisible particles is particularly difficult because such particles are not register in the detector. The strategy employed is to deduce the presence of invisible particles from the missing transverse momentum calculated from visible particles using only the calorimeter information.

In 1- and 2-lepton channels different triggers are used to select electron and muon decays of the W and Z bosons, respectively. In the electron sub-channel, a low-p_{T} threshold unprescaled single electron trigger is used, while in the muon sub-channel events are recorded using the same $E_{\mathrm{T}}^{\mathrm{miss}}$ triggers used in the 0-lepton channel. Muons are approximately invisible in the calorimeter so they contribute only marginally (i.e. they are treated like neutrinos) in the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculations. The choice of using the $E_{\mathrm{T}}^{\mathrm{miss}}$

G. Di Gregorio, *A Beauty-ful Boson*, Springer Theses,
https://doi.org/10.1007/978-3-031-20013-7_3

trigger is motivated by the low muon trigger efficiency for $|\eta| > 2.4$ region due to the TGC coverage and some gaps needed for the inner detector services. Additionally the $E_{\mathrm{T}}^{\mathrm{miss}}$ triggers effectively select muons since their contribution is not included in the online $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation. In the 1-lepton muon sub-channel, the $E_{\mathrm{T}}^{\mathrm{miss}}$ trigger provides an efficiency of 98% for signal events passing the offline selection, compared to a ∼80% efficiency of the single-muon trigger. In the 2-lepton muon sub-channel, the $E_{\mathrm{T}}^{\mathrm{miss}}$ trigger is preferred with respect to the single muon one because it is more efficient (∼5%).

3.1.1 E_T^{miss} Trigger

The $E_{\mathrm{T}}^{\mathrm{miss}}$ trigger algorithms used by ATLAS are based on the transverse momentum imbalance evaluated using only the calorimeter information. The first selection to build the $E_{\mathrm{T}}^{\mathrm{miss}}$ trigger algorithm is done at the L1 trigger level. Signals from calorimeter cells are used to form projective towers with granularity of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ [1]. Fixed threshold is then applied per tower: the energy E_i of tower i below this threshold is set to zero in the subsequent calculations. All the towers over threshold are summed into larger towers with granularity of $\Delta\eta \times \Delta\phi = 0.2 \times 0.2$. The $E_{\mathrm{T}}^{\mathrm{miss}}$ is then recalculated doing a vectorial sum of the contributions of the clustered towers. Events that are accepted by the L1 trigger are then transferred to the HLT where the $E_{\mathrm{T}}^{\mathrm{miss}}$ is recalculated.

At HLT level $E_{\mathrm{T}}^{\mathrm{miss}}$ is calculated using jets reconstructed in the HLT with the anti-k_t algorithm with $R = 0.4$. During the Run 2 period the $E_{\mathrm{T}}^{\mathrm{miss}}$ algorithm used in HLT has been modified to remove the pile-up effects and to improve the performance.

The efficiency of the L1 $E_{\mathrm{T}}^{\mathrm{miss}}$ trigger is measured using $Z \rightarrow \mu\mu$ events selected with a muon trigger. The muons have a little interaction with the calorimeter, so the transverse momentum of the dimuon system $p_{\mathrm{T}}(\mu\mu)$ provides a good proxy of the expected $E_{\mathrm{T}}^{\mathrm{miss}}$. Figure 3.1a shows the L1 trigger efficiency as a function of $p_{\mathrm{T}}(\mu\mu)$ for a L1 nominal threshold of 50 GeV. The algorithm achieves 90% efficiency for $p_{\mathrm{T}}(\mu\mu)$ ∼150 GeV and, at $p_{\mathrm{T}}(\mu\mu) > 200$ GeV it becomes fully efficient. Moreover it is observed that the efficiency is quite stable across the different $< \mu >$ values.

At the HLT level, the algorithm efficiency depends upon luminosity which has caused the increase of pile-up levels. The overall effect of the increase of the pile-up brings the degradation of the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution of the detector. To control the trigger rate due to the increasing level of pile-up, the HLT threshold was raised progressively from 70 GeV to 110 GeV during the Run 2. Figure 3.1b shows the overall (L1+HLT) $E_{\mathrm{T}}^{\mathrm{miss}}$ trigger efficiency for $Z \rightarrow \mu\mu$ events year-by-year as a function of $p_{\mathrm{T}}(\mu\mu)$. Due to the threshold increase, the efficiency remains quite stable within few percent in the full $p_{\mathrm{T}}(\mu\mu)$ range over the four years.

The trigger efficiency is 85–90% efficient at $E_{\mathrm{T}}^{\mathrm{miss}}$ =150 GeV and fully efficient above 180 GeV. The same values are obtained on simulated data applying the offline selection. Since the analysis described in this thesis exploits a phase space with $E_{\mathrm{T}}^{\mathrm{miss}}$

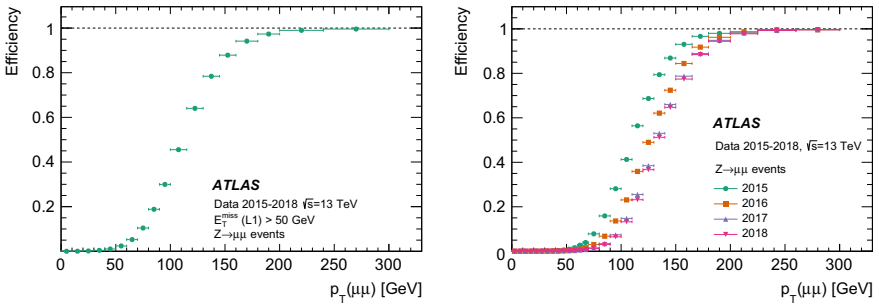

Fig. 3.1 **a** The L1 $E_{\mathrm{T}}^{\mathrm{miss}}$ trigger efficiency as a function of $p_{\mathrm{T}}(\mu\mu)$. **b** Full chain $E_{\mathrm{T}}^{\mathrm{miss}}$ trigger efficiencies for each year as a function of $p_{\mathrm{T}}(\mu\mu)$. In both plots the error bands correspond only to the statistical uncertainty [1]

≥ 250 GeV and the $E_{\mathrm{T}}^{\mathrm{miss}}$ trigger is completely efficient above 180 GeV, there is no need to apply data-to-MC scale factors to correct $E_{\mathrm{T}}^{\mathrm{miss}}$ trigger inefficiencies.

3.1.2 Single Electron Trigger

The single electron triggers select events with electrons from the huge amount of data produced by high energy *p-p* collisions at the LHC and only a fraction of these events can be recorded.

The L1 trigger for electrons uses calorimeter information to build an EM RoI consisting of 4×4 trigger towers with a granularity of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$. The energy of the trigger towers is calibrated at the electromagnetic energy scale and a nominal transverse energy threshold is applied. The threshold is η-dependent to account for energy losses and the geometry of the detector. Additionally, EM isolation requirements are applied. The isolation requirements are optimized to maintain a fixed L1 efficiency at the lowest possible rate. At the HLT, the full detector granularity is used within the RoI for the final trigger decision. Electrons are identified with EM clusters and with tracks matched to the clusters. As in the offline algorithm, the electron selection relies on a multivariate technique using a likelihood-based discriminant with four operating points. The working point used in the trigger is designed to be as close as possible to the offline version. Isolation requirements are also added at this stage.

The electron trigger efficiency is measured for electrons at the HLT using the tag-and-probe method [2]. The trigger efficiency is defined as the ratio of the number of triggered electron candidates to the number of produced electrons and it is evaluated considering $Z \rightarrow ee$ events. The evolution of the single-electron trigger efficiency in 2015–2018 is shown in Fig. 3.2a. The sharper efficiency turn-on as a function of the transverse energy E_{T} in 2015 is due to a looser identification requirement, a lower E_{T}

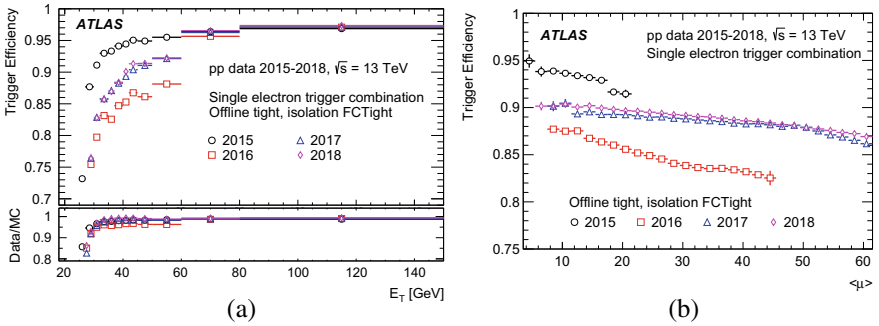

Fig. 3.2 **a** Evolution of the single-electron trigger efficiency as a function of the offline E_{T} during Run 2. In the bottom panel the ratios of data to MC simulation efficiencies are also shown. **b** Evolution of the single-electron efficiency as a function of pile-up during Run 2 [3]

threshold and no isolation requirement. Moreover 2016 data show some inefficiency at $E_{\mathrm{T}} < 60$ GeV due to too stringent online identification criteria.

Simulated events need to be corrected to reproduce as closely as possible the efficiencies measured in data. This is achieved applying scale factors defined as the ratio of the efficiency measurements in data to that one determined in simulated events. The data-to-MC scale factors, shown in the lower panel of Fig. 3.2a, are of the order of 18% close to the trigger E_{T} threshold and of the order of 4% above 40 GeV. The precision of these scale factors is of the order of 0.1%. Figure 3.2b shows the dependence of the trigger efficiency on pile-up, which is less pronounced at the second half of Run 2.

3.2 Simulated Samples

Simulated samples of signal and background processes are produced and studied for the optimization of the analysis criteria and to determine the expected signal and background distributions in the different regions of the analysis. Furthermore, simulations are also used to estimate the systematic uncertainties caused by different sources. The QCD multi-jet process is treated differently because the production cross-section is large (almost eight order of magnitude larger than the Higgs cross-section) and the rejection power of such events by the analysis selection is high. Such background is estimated using data-driven techniques, as described in Sect. 6.3.7. Simulated events are obtained with a multiple step process. The samples are produced with state-of-art Monte Carlo event generators. The simulation begins with a hard process in which the constituents of the colliding particles interact at high momentum scale to produce outgoing fundamental objects as the SM quarks, leptons, bosons

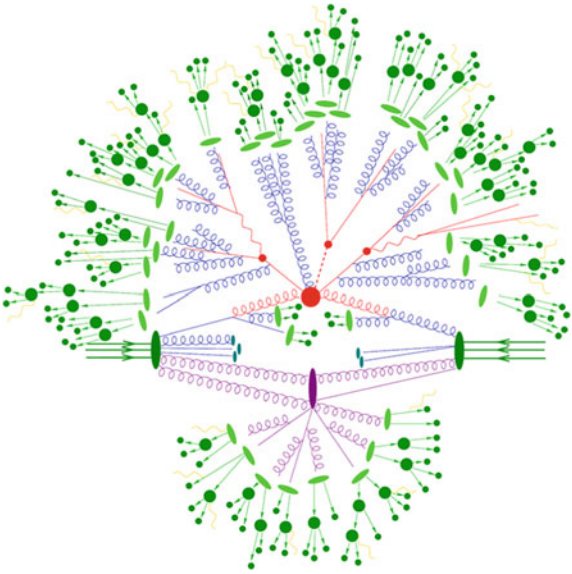

Fig. 3.3 Schematic diagram of the steps of the event simulation. The simulation of the hard process is shown in red while the parton shower is represented with blue lines. The production of hadrons and their decay is represented by green lines and dots. The purple lines represent the simulation of the underlying events [5]

or some new particles of BSM theories. The partons from the hard process radiate gluons as they evolve, leading to the formation of parton shower. The constituents of the parton showers hadronise to form colourless hadrons triggering the process of hadronisation. Many of the produced hadrons are unstable, so another step of the simulation is the hadron decay. In hadron-hadron collisions, the other constituent partons of the incoming hadrons undergo multiple interactions which produce the underlying events. The final step is a detailed simulation of the ATLAS detector response based on Geant 4 [4]. All the simulation steps, except for the detector simulation response, are schematically shown in Fig. 3.3 and briefly summarised in the following.

- **Hard process simulation:** Many LHC processes involve large momentum transfers such the production of jets with high transverse momenta. The simulation of these processes is the core of any simulation of collider events through Monte Carlo generators. Such interactions can be described by perturbation theory since QCD quanta are asymptotically free. The cross-section for a scattering process of two hadrons $h1$ and $h2$ at hadron colliders can be computed as:

$$\sigma_{h1h2\to n} = \sum_{a,b} \int_0^1 \int_0^1 dx_a dx_b \int f_a^{h1}(x_a, \mu_F) f_b^{h2}(x_b, \mu_F) d\hat{\sigma}_{ab\to n}(\mu_F, \mu_R) \tag{3.1}$$

where:

– $f_a^{h1}(x_a, \mu_F)$ ($f_b^{h2}(x_b, \mu_F)$) is the parton distribution function[1] (PDF) which depends on the momentum fraction x_a (x_b) of a parton a (b) with respect to its parent hadron $h1$ ($h2$) and on the factorization scale μ_F;
– $\hat{\sigma}_{ab\to n}$ denotes the parton-level cross-section for the production of the final state n from the initial partons a and b which depends on the momenta given by the final phase space Φ_n, on the factorization scale μ_F and on the renormalisation scale μ_R.

Equation 3.1 can be re-written as:

$$\sigma = \sum_{a,b} \int_0^1 \int_0^1 dx_a dx_b \int d\Phi_n f_a^{h1}(x_a, \mu_F) f_b^{h2}(x_b, \mu_F) \times \frac{1}{2\hat{s}} |\mathcal{M}_{ab\to n}(\Phi_n; \mu_F, \mu_R)|^2 \tag{3.2}$$

where $\mathcal{M}_{ab\to n}$ is the matrix element, $d\Phi_n$ is the differential phase space element over the n final-state particles and $1/(2\hat{s}) = 1/(2x_a x_b s)$ is the parton flux which depends on the hadronic squared centre-of-mass energy s.
The cross-section is fully specified only for a given PDF set and a given choice of the factorization μ_F and renormalisation μ_R scales. These scales are unphysical and their impact on the cross-section decreases in higher order terms of perturbation theory which are taken into account.

- **Parton shower:** In the event simulation, the hard process is described with the lowest-order matrix elements. The higher orders of the process can be approximate through a parton shower algorithm, which is formulated as an evolution in momentum transfer from the high energy scales, associated to the hard process, to the low scales (of the order of 1 GeV), associated to the confinement of partons into hadrons. The parton shower represents the connection between the scattered partons and the final hadrons, and it aims at modelling this extra radiation before hadronisation.
- **Hadronisation:** As the momentum decreases perturbation theory can not be used any more to describe the QCD interactions. At this stage, the perturbative evolution must be replaced by a non-perturbative hadronisation model. This model describes how the quarks and gluons from hard scattering simulation, parton showers and multiple scattering simulations are transformed into colorless final states. Individual parton hadronisation is not allowed, instead color-connected systems of partons hadronise collectively.
- **Decays:** Following the hadronisation phase of the event generation, unstable hadrons which decay into lighter hadrons are produced. Unstable hadrons decay into particles that are considered stable on the detector timescales. This is an important part of the event simulation because the observed final-state hadrons result from a convolution of hadronisation and decay processes.
- **Underlying events:** The underlying event simulation contains all the events not coming from the primary hard scattering process. The underlying event includes

[1] The parton distribution function is used to described the probability to find a parton carrying a fraction x of the total momentum. Different parton distribution functions are realised by several collaborations.

contribution from initial and final state radiation, beam-beam remnants or multiple parton interactions. The energy scale of the underlying process is expected to be smaller than the one of the hard scattering event, the underlying event has a uniform activity represented by low energy hadrons. The simulation of pile-up is an independent step of the simulation. This means that the simulation of the hard scattering event is overlaid with these simulated pile-up events.

- **Pile-up events:** At LHC there is high chance that multiple inelastic interactions in the same bunch (in-time pile-up) and neighbouring bunch crossing (out-of-time pile-up) happen simultaneously at the collision point. Each component of the pile-up has a dedicated simulation. The pile-up simulation is achieved by overlaying simulated minimum bias events to signal events prior to the conversion of energy deposits to detector signals [6].
- **Detector simulation:** The final step in the Monte Carlo chain is the simulation of interactions between the final state particles created by the event generators and the ATLAS detector. The detector simulation describes particle-matter interactions, particles trajectories and particle decays within the ATLAS detector volume. This simulation is based on Geant 4 and events are reconstructed with the ATLAS reconstruction software [7]. The detector simulation follows the detector state evolutions in term of active subdetectors, dead channels and misalignment. The configuration of the detector, including the misalignment and distortions, can be set at run time. The energies deposited in the active portions of the detector are recorded as *hits*, containing the total energy deposition, position and time, and then they are written to a simulation output file. The detector simulation is quite slower than the event generation, and for this reason it is convenient to simulate the interaction with the ATLAS detector only if the events pass the full analysis chain. *MC filters* are applied between the generation and simulation steps to select events in a particular phase space.

3.2.1 Monte Carlo Generators

Several programs are available to simulate the events. There are some multi-purpose generators and some specialized programs. The former type of generator is able to simulate the full event from the matrix element to the hadronisation, while the latter type generates only part of the event. In this case an interface for the specialized programs is needed. The MC event generators used in the analysis are described in the following.

- **`Pythia`:** `Pythia` [8] is a multi-purpose generator providing parton shower, underlying event and matrix element calculations. The latter calculation is given at leading order (LO), while both parton shower and underlying event models are tuned on existing measurements. Several PDFs are hardcoded in this event generator. This generator can be used to simulate just certain parts of the event. This feature is useful since the accuracy of the matrix element calculation is just

at LO. However, the parton shower model exhibits good agreement with data therefore Pythia is often used to provide the parton shower description for a simulation model.

- **Herwig:** Herwig [9] is a multi-purpose particle physics event generator but it is not used as a stand-alone generator any more. However it is still used for the parton shower description of a simulation model. Many simulations that use Pythia to describe the parton shower are also generated with Herwig, the comparison between the two generators allows to evaluate the systematic effects of the parton shower model on the simulation.
- **Sherpa:** Sherpa [10] is a general-purpose event generator, capable of simulating the physics of hadron-hadron collisions. It covers both the matrix element calculation and the parton shower description. It provides next-to-leading order (NLO) calculation for a variety of processes. It is the preferred generator for processes with additional radiated jets, which are directly included in the matrix-element calculation. Differently from Pythia and Herwig, there are no "hybrid" versions of Sherpa, therefore Sherpa can not be used only to generate part of the event.
- **Powheg:** Powheg [11] is a generator able to simulate the matrix elements with NLO calculation for a variety of processes. It is used together with other generators that simulate the parton shower.
- **MadGraph:** MadGraph [12] is a generator that provides only the matrix element calculations at LO and NLO. In most of the case it is not used as default matrix element generator but it is used to study systematic effects.

All ATLAS simulated dataset model pile-up events with Pythia 8.1 [8]. In addition, all simulations, except Sherpa, use the EvtGen [13] program to describe the decay of hadrons containing bottom or charm quarks.

3.3 Signal Process Simulation

The VH signals studied in this thesis include three main Higgs processes: $ZH \to \nu\nu b\bar{b}$, $WH \to l\nu b\bar{b}$ and $ZH \to llb\bar{b}$ where $l = e, \mu, \tau$. The tree-level Feynman diagrams of the three different processes are shown in Fig. 3.4.

The ZH production mode is furthermore split into two contributions depending on the partonic initial state, $qq \to ZH$ or $gg \to ZH$. Two LO Feynman diagrams for the $gg \to ZH$ process are shown in Fig. 3.5.

All the qq-initiated production processes are simulated using the Powheg generator with the MiNLO (Multiscale Improved Next-to-Leading Order) procedure [14] applying AZNLO tune [15] with NNPDF3.0NLO PDF set [16] for the matrix-element. The events are also interfaced to Pythia 8.212 [17] MC generator for the modelling of the parton shower, underlying event and multiple parton interactions. The simulation of qq-initiated production processes includes also qg-initiated

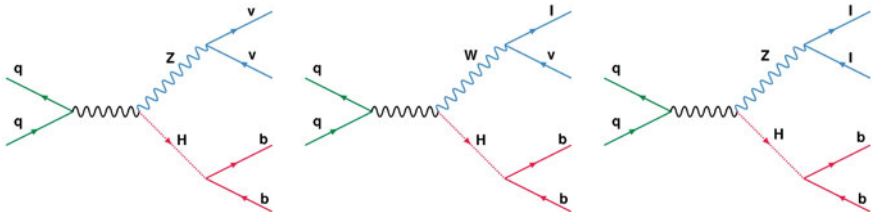

Fig. 3.4 Leading order Feynman diagram for $ZH \to \nu\nu b\bar{b}$, $WH \to l\nu b\bar{b}$ and $ZH \to llb\bar{b}$ processes with $l = e, \mu, \tau$. These diagrams represent the 0-, 1-, and 2-lepton channels

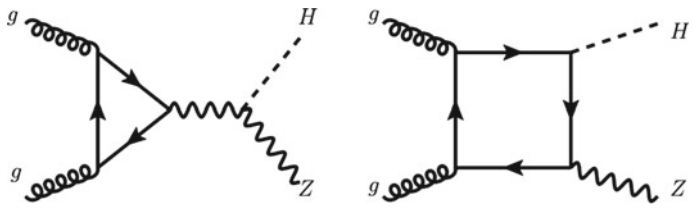

Fig. 3.5 Two leading order Feynman diagrams of the $gg \to ZH$-process

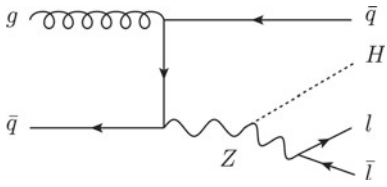

Fig. 3.6 One of the leading order Feynman diagram of the qg-initiated process

events. An example of a Feynman diagram of the qg-initiated process is shown in Fig. 3.6.

The gg-initiated events are simulated using the `Powheg` generator using NNPDF3.0NLO PDF set for the matrix-element. Furthermore the events are interfaced to `Pythia 8.212` generator for the modelling of parton shower, underlying event and multiple parton interactions. The gg-initiated events are generated at LO with NLO corrections.

All the samples are normalised to the best available theoretical prediction for the cross-section of each process at $\sqrt{s} = 13$ TeV. The signal cross-sections are evaluated following the prescriptions of the LHC Higgs Cross-Section Working Group for a Higgs boson mass of $m_H = 125$ GeV [18, 19], considering the $b\bar{b}$ branching ratio fixed to 58%. The WH signal samples are normalised to the production cross-section calculated at next-to-next-to-leading order (NNLO) in QCD with electroweak (EW) corrections at NLO accuracy. The $gg \to ZH$ cross-section is calculated at NLO in QCD, while for the $qq \to ZH$ process the cross-section is evaluated from the difference between the total ZH cross-section, calculated at NNLO in QCD with NLO EW corrections, and the gluon-induced ZH production. This is done to avoid double

Table 3.1 Monte Carlo samples used for the signal processes and the cross-section times branching ratio (BR) used to normalise the different processes at $\sqrt{s} = 13$ TeV l is inclusive of e, μ, τ leptons

Process	Generator	$\sigma \times$ BR [fb]
$qq \to ZH \to \nu\nu b\bar{b}$	Powheg MiNLO + Pythia 8.212	89.08
$qq \to WH \to l^+\nu b\bar{b}$	Powheg MiNLO + Pythia 8.212	164.6
$qq \to WH \to l^-\nu b\bar{b}$	Powheg MiNLO + Pythia 8.212	104.5
$qq \to ZH \to llb\bar{b}$	Powheg MiNLO + Pythia 8.212	44.84
$gg \to ZH \to \nu\nu b\bar{b}$	Powheg + Pythia 8.212	14.30
$gg \to ZH \to l^-l^+b\bar{b}$	Powheg + Pythia 8.212	7.23

counting when merging $qq \to ZH$ and $gg \to ZH$ samples together. A summary of the signal process and the cross-section times branching ratio values are listed in Table 3.1. The branching ratios correspond to the product of the Higgs and vector boson decay. Table 3.1 contains also the information about the generators used to produce the samples used in the analysis.

3.4 Background Process Simulation

3.4.1 Vector Boson + Jets Production

The production of the V ($V = Z$ or W) boson in association with jets is one of the main backgrounds in all the lepton channels considered in the analysis. Figure 3.7 shows the lowest order Feynman diagrams for Z+jets and W+jets background.

The production of high statistic V+jets MC samples is important for this and many other ATLAS analyses. Efforts have been made by the ATLAS Collaboration to update the simulations to the latest V+jets MC generators, moving from LO to NLO generator, and updating parameter tunings for the Run 2 data taking [20]. The V+jets

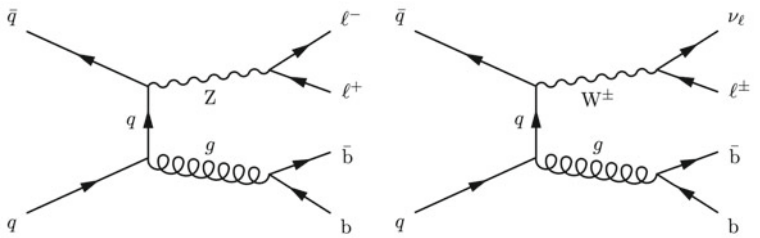

Fig. 3.7 Lowest order Feynman diagrams for the quark induced Z+jets (**a**) and W+jets (**b**) processes

Table 3.2 Monte Carlo samples used for the V+jets processes and the cross-section times branching ratio used to normalise the different processes at $\sqrt{s} = 13$ TeV. l is inclusive of e, μ, τ leptons

Process	Generator	$\sigma \times$BR [pb]
$Z(\rightarrow \nu\nu)$ + jets	`Sherpa 2.2.1`	1914
$W(\rightarrow l\nu)$ + jets	`Sherpa 2.2.1`	20080
$Z/\gamma^*(\rightarrow ll)$ + jets [$m_{ll} > 40$ GeV]	`Sherpa 2.2.1`	2107

processes are simulated with `Sherpa 2.2.1` using NNPDF3.0NNLO PDF set [16] with dedicated parton shower tuning developed by `Sherpa` authors. Events with many jets produced in association with the W or the Z boson largely contribute to the background in the phase space of the analysis. `Sherpa 2.2.1` provides a combination of different matrix elements with different parton multiplicities. Events with zero or one additional partons are generated at NLO in the ME calculation, while events with two or three additional partons are included at LO in the ME. Samples are normalised using cross-sections calculated at NNLO accuracy [21]. The list of the V+jets samples used and the relative cross-sections is given in Table 3.2.

The analysis treated in this thesis includes different regions of phase space corresponding to high p_{T} of the vector boson (p_{T}^V). To generate sufficient high p_{T}^V events, the V+jets samples are split depending on p_{T}^V and the scalar p_{T} sum of all parton-level jets with $p_{\mathrm{T}} > 20$ GeV called H_T. Samples are produced in slices of $\max(H_T, p_{\mathrm{T}}^V)$ according to the following intervals:

$$[0-70, 70-140, 140-280, 280-500, 500-1000, >1000]\,\mathrm{GeV}$$

Final states with b-tagged jets are studied in this analysis. It is important to make sure that the MC statistics in this specific heavy-flavour enriched phase space are large enough to provide a robust MC prediction to be compared to the data. For this reason the V+jets samples are generated applying different filters to select the flavour composition of the jets produced in association with the vector boson V. Three filters are used:

- *BFilter*: at least a b-hadron with $|\eta| < 4$ is required;
- *CFilterBVeto*: at least a c-hadron with $p_{\mathrm{T}} > 4$ GeV and $|\eta| < 3$ is required, and no overlap with the BFilter sample;
- *CVetoBVeto*: events which pass the previous two filters are rejected.

3.4.2 Top Pair Production

The top pair process is a main background for both the 0-lepton and 1-lepton channels. Figure 3.8 shows an example of lowest order Feynman diagrams for the top pair background.

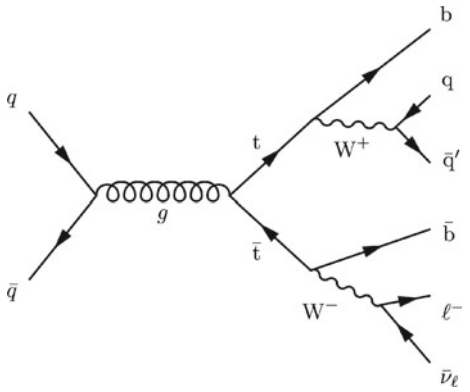

Fig. 3.8 Lowest order Feynman diagram for the $t\bar{t}$ process

The MC configuration for the $t\bar{t}$ process is `Powheg + Pythia 8.230`, which utilities the `Powheg` NLO matrix element generator [22, 23] interfaced to `Pythia 8.230` using the A14 tune [24] for the parton shower, hadronisation, underlying event and multiple parton interactions. Additionally the sample uses NNPDF3.0NLO PDF set in the matrix element calculation. The cross-section used to normalised the $t\bar{t}$ sample is calculated at NNLO in QCD including resummation of next-to-next-to-leading logarithmic (NNLL) soft gluon terms [25]. Table 3.3 briefly summaries the generators used for the top pair production together with the cross-section value. To have enough statistics, the samples used in the $VH(b\bar{b})$ analysis are filtered at generator stage using $E_{\mathrm{T}}^{\mathrm{miss}}$ filters.[2]

3.4.3 Single-Top Production

The single-top process is a relevant background for both 0-lepton and 1-lepton channels. Three different channels are generated: s-channel, t-channel and W-associated production (Wt-channel). Among the three channels, Wt-channel is the one which gives the most relevant contribution in the analysis. Figure 3.9 shows the lowest order Feynman diagrams for the three channels of single-top production.

The single-top background events are generated separately for the different channels, using the `Powheg` generator for the hard scattering process with the

Table 3.3 Monte Carlo samples used for the $t\bar{t}$ process and the cross-section times branching ratio (BR) used to normalise the different processes at $\sqrt{s} = 13$ TeV

Process	Generator	$\sigma \times$ BR [pb]
$t\bar{t}$	`Powheg + Pythia 8.230`	831.76

[2] The $E_{\mathrm{T}}^{\mathrm{miss}}$ values used for the different slices are defined using the truth level information.

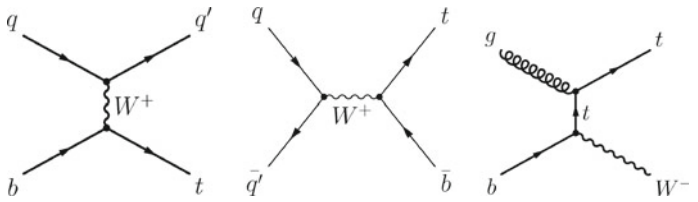

Fig. 3.9 Lowest order Feynman diagrams for the three channels of the single top production: t-channel (left), s-channel (middle), Wt-channel (right)

NNPDF3.0NLO PDF set. The events are generated at NLO in the ME calculation and interfaced with `Pythia 8.230` algorithm for the simulation of parton shower, underlying event and multiple parton interactions. Single top samples in t- and s-channels are generated applying a lepton filter to require the leptonic decay of the W boson, while the Wt-channel samples are generated applying either a dilepton filter (both W bosons decay leptonically) or without applying any filter. In the analysis, both Wt-channel samples are used, after removing the overlap between the two samples.

The predicted t-channel single-top cross-section for p-p collisions at center-of-mass energy $\sqrt{s} = 13$ TeV is $\sigma_t = 136.02^{+5.4}_{-4.6}$ pb for the top quark and $\sigma_{\bar{t}} = 80.95^{+4.1}_{-3.6}$ pb for the anti-top quark. In the same condition the s-channel single-top cross-section is $\sigma_t = 6.35^{+0.2}_{-0.2}$ pb for the top quark and $\sigma_{\bar{t}} = 3.97^{+0.19}_{-0.17}$ pb for the anti-top quark. The Wt-channel single-top cross-section is $\sigma_t = 71.7^{+3.8}_{-3.8}$ pb. The cross-sections of the t- and s-channels are calculated at NLO while the cross-section of the Wt-channel is computed at NNLO. In both cases, the top mass value used in the calculation is $m_t = 172.5$ GeV. All the single-top samples used in the analysis are normalised to the cross-sections from the higher order calculation. The full list of single-top samples used in the analysis, together with the generator information and the cross-section value times branching ratio, is shown in Table 3.4.

Table 3.4 Monte Carlo samples used for the single top processes and the cross-section times branching ratio (BR) used to normalise the different processes at $\sqrt{s} = 13$ TeV

Process	Generator	$\sigma \times$ BR [nb]
Single top t-channel, t	`Powheg + Pythia 8.230`	0.0370
Single top t-channel, $\bar{t}$	`Powheg + Pythia 8.230`	0.0222
Single top s-channel, t	`Powheg + Pythia 8.230`	0.0020
Single top s-channel, $\bar{t}$	`Powheg + Pythia 8.230`	0.0013
Single top Wt-channel, t	`Powheg + Pythia 8.230`	0.0380
Single top Wt-channel, $\bar{t}$	`Powheg + Pythia 8.230`	0.0380

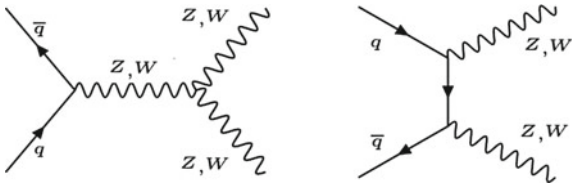

Fig. 3.10 Lowest order Feynman diagrams for the diboson process

3.4.4 Diboson Production

The production of two vector bosons is a sub-dominant background process but important because it has a final state similar to the one of the signal. The diboson background consists of final states generated by WW, WZ and ZZ events. Several diboson processes contribute in the analysis phase space: $ZZ \to b\bar{b}\nu\bar{\nu}$ in 0-lepton channel, $ZW \to b\bar{b}l\nu$ in 1-lepton channel and $ZZ \to b\bar{b}l\bar{l}$ in 2-lepton channel. Additional processes as $WW \to q\bar{q}l\nu$ can give a smaller contribution arising from the fakes reconstruction of one of the leptons or from the mistagging of a jet from the W decay. Figure 3.10 shows the leading order Feynman diagrams for the diboson production.

Quark-induced diboson processes are simulated using `Sherpa 2.2.1` generator for both the matrix element and parton shower calculation with the NNPDF3.0NNLO PDF set. `Sherpa` provides a combination of different matrix elements with different parton multiplicities: processes with zero or one additional partons are

Table 3.5 Monte Carlo samples used for the diboson process and the cross-section times branching ratio (BR) used to normalise the different processes at $\sqrt{s} = 13$ TeV

Process	Generator	$\sigma \times$ BR [nb]
$qq \to W^+W^- \to l\nu qq$	`Sherpa 2.2.1`	24.710
$qq \to W^+W^- \to qql\nu$	`Sherpa 2.2.1`	24.73
$qq \to WZ \to l\nu qq$	`Sherpa 2.2.1`	11.4
$qq \to WZ \to qqll$	`Sherpa 2.2.1`	3.44
$qq \to WZ \to qq\nu\nu$	`Sherpa 2.2.1`	6.80
$qq \to ZZ \to qqll$	`Sherpa 2.2.1`	15.56
$qq \to WZ \to l\nu b\bar{b}$	`Sherpa 2.2.1`	2.50
$qq \to ZZ \to llbb$	`Sherpa 2.2.1`	15.54
$qq \to ZZ \to \nu\nu bb$	`Sherpa 2.2.1`	15.57
$qq \to ZZ \to qq\nu\nu$	`Sherpa 2.2.1`	15.56
$gg \to W^-W^+ \to l\nu qq$	`Sherpa 2.2.2`	0.6224
$gg \to W^+W^- \to l\nu qq$	`Sherpa 2.2.2`	0.6225
$gg \to ZZ \to llqq$	`Sherpa 2.2.2`	0.923
$gg \to ZZ \to \nu\nu qq$	`Sherpa 2.2.2`	0.925

generated at NLO in the matrix element, while two or three additional partons are generated at LO in the matrix element. Gluon-induced samples ($gg \rightarrow VV$, with $V = Z$ or W boson) are generated using Sherpa 2.2.2 with NNPDF3.0NNLO PDF set. In this case the gg-initiated samples are generated at LO in the matrix element.

All the samples are normalised to the NLO inclusive cross-section. The inclusive cross-sections for the various diboson processes are summarized in Table 3.5.

References

1. Aad G et al. (2020) Performance of the missing transverse momentum triggers for the ATLAS detector during Run-2 data taking. arXiv:2005.09554 [hep-ex]
2. ATLAS Collaboration (2017) Electron efficiency measurements with the ATLAS detector using 2012 LHC proton-proton collision data. Eur Phys J C 77:195. https://doi.org/10.1140/epjc/s10052-017-4756-2. arXiv:1612.01456 [hep-ex]
3. ATLAS Collaboration (2019) Performance of electron and photon triggers in ATLAS during LHC Run 2. arXiv:1909.00761 [hep-ex]
4. Agostinelli S et al. (2003) GEANT4 - a simulation toolkit. Nucl Instrum Meth A 506:250. https://doi.org/10.1016/S0168-9002(03)01368-8
5. Simulations. https://theory.slac.stanford.edu/our-research/simulations
6. Marshall Z (2014) Simulation of pile-up in the ATLAS experiment. J Phys Conf Ser 513. ed. by D. Groep and D. Bonacorsi 022024. https://doi.org/10.1088/1742-6596/513/2/022024
7. ATLAS Collaboration (2010) The ATLAS simulation infrastructure. Eur Phys J C 70:823. https://doi.org/10.1140/epjc/s10052-010-1429-9. arXiv:1005.4568 [physics.ins-det]
8. Sjostrand T, Mrenna S, Skands PZ (2008) A brief introduction to PYTHIA 8.1. Comput Phys Commun 178:852. https://doi.org/10.1016/j.cpc.2008.01.036. arXiv:0710.3820 [hep-ph]
9. Bellm J et al. (2016) Herwig 7.0/Herwig++ 3.0 release note. Eur Phys J C 76:196. https://doi.org/10.1140/epjc/s10052-016-4018-8. arXiv:1512.01178 [hep-ph]
10. Gleisberg T, Höche S, Krauss F, Schönherr M, Schumann S et al. (2009) Event generation with SHERPA 1.1. JHEP 02:007. https://doi.org/10.1088/1126-6708/2009/02/007. arXiv:0811.4622 [hep-ph]
11. Alioli S, Nason P, Oleari C, Re E (2010) A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. JHEP 06:043. https://doi.org/10.1007/JHEP06(2010)043. arXiv:1002.2581 [hep-ph]
12. Alwall J et al. (2014) The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. JHEP 07:079. https://doi.org/10.1007/JHEP07(2014)079. arXiv:1405.0301 [hep-ph]
13. Lange DJ (2001) The EvtGen particle decay simulation package. Nucl Instrum Meth A 462:152. https://doi.org/10.1016/S0168-9002(01)00089-4
14. Luisoni G, Nason P, Oleari C, Tramontano F (2013) $HW^{\pm}$/HZ+ 0 and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO. JHEP 10:083. https://doi.org/10.1007/JHEP10(2013)083. arXiv:1306.2542 [hep-ph]
15. ATLAS Collaboration (2011) Measurement of the transverse momentum distribution of Z/λ* bosons in proton-proton collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector. Phys Lett B 705:415. https://doi.org/10.1016/j.physletb.2011.10.018. arXiv:1107.2381 [hep-ex]
16. Ball RD et al. (2013) Parton distributions with LHC data. Nucl Phys B 867:244. doi: 10.1016/j.nuclphysb.2012.10.003. arXiv:1207.1303 [hep-ph]
17. Sjöstrand T et al. (2015) An introduction to PYTHIA 8.2. Comput Phys Commun 191:159. https://doi.org/10.1016/j.cpc.2015.01.024. arXiv:1410.3012 [hep-ph]
18. LHCSW Group, SM Higgs production cross sections at 13 TeV. https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt13TeV

19. de Florian D et al. (2016) Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector, 2/2017. https://doi.org/10.23731/CYRM-2017-002. arXiv:1610.07922 [hep-ph]
20. ATLAS Collaboration (2016) Monte Carlo generators for the Production of a W or Z/λ* Boson in association with jets at ATLAS in Run 2, ATL-PHYS-PUB-2016-003. https://cds.cern.ch/record/2120133
21. Catani S, Cieri L, Ferrera G, de Florian D, Grazzini M (2009) Vector boson production at hadron colliders: a fully exclusive QCD calculation at NNLO. Phys Rev Lett 103:082001. https://doi.org/10.1103/PhysRevLett.103.082001.arXiv:0903.2120 [hep-ph]
22. Frixione S, Nason P, Ridolfi G (2007) A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction. JHEP 09:126. https://doi.org/10.1088/1126-6708/2007/09/126. arXiv:0707.3088 [hep-ph]
23. Nason P (2004) A new method for combining NLO QCD with shower Monte Carlo algorithms. JHEP 11:040. https://doi.org/10.1088/1126-6708/2004/11/040. arXiv:hep-ph/0409146
24. ATLAS Collaboration (2014) ATLAS Pythia 8 tunes to 7 TeV data, ATL-PHYS-PUB-2014-021. https://cds.cern.ch/record/1966419
25. Beneke M, Falgari P, Klein S, Schwinn C (2012) Hadronic top-quark pair production with NNLL threshold resummation. Nucl Phys B 855:695. https://doi.org/10.1016/j.nuclphysb.2011.10.021. arXiv:1109.1536 [hep-ph]