

Università degli studi di Padova
Dipartimento di Fisica e Astronomia

Tesi di Dottorato

Search for heavy resonances decaying into a Z boson and a vector boson in the $\nu\bar{\nu} q\bar{q}$ final state at CMS

Supervisor: Prof. Franco Simonetto
Candidate: Lisa Benato

Scuola di Dottorato di Ricerca, XXX ciclo

"I have no special talent. I am only passionately curious."

(A. Einstein)

Contents

1	Introduction	1
2	Theoretical motivation	3
3	The Large Hadron Collider and the CMS experiment	5
4	Search for diboson resonances in the $VZ \rightarrow q\bar{q} \nu\bar{\nu}$ final state	7
4.1	Analysis overview	7
4.2	Data and Monte Carlo simulations samples	8
4.2.1	Signal samples	8
4.2.2	Signal characterization	9
4.2.3	Background samples	16
4.2.4	Vector boson momentum corrections	18
4.2.5	Data samples	19
4.2.6	Trigger	19
4.3	Event selection	23
4.3.1	Vertex and Pile-up	23
4.3.2	Electrons	23
4.3.3	Photons	24
4.3.4	Muons	25
4.3.5	Taus	25
4.3.6	Jets	25
4.3.7	Jet mass	27
4.3.8	Jet substructure	31
4.3.8.1	Corrections induced by jet substructure variables	34
4.3.9	b-tagging	35
4.3.10	Missing Energy	35
4.3.11	Diboson candidate reconstruction	35
4.3.11.1	$V \rightarrow q\bar{q}$ reconstruction	35
4.3.11.2	$Z \rightarrow \nu\bar{\nu}$ reconstruction	36
4.3.11.3	Composite VZ candidate reconstruction	36
4.3.12	Final analysis selections	36
4.3.12.1	Z candidate selections	37

4.3.12.2	V candidate selections	37
4.3.12.3	Topology and event cleaning	37
4.4	Data and simulations comparison	40
4.5	Background estimation technique	51
4.5.1	Background normalization	51
4.5.2	Background shape	54
4.5.3	Alpha method validations	56
4.5.4	Signal modeling	63
4.5.4.1	Signal parametrization	63
4.6	Systematic uncertainties	68
4.6.1	Uncertainties affecting the data-driven main background estimation	68
4.6.1.1	Normalization	68
4.6.1.2	Shape	68
4.6.2	Uncertainties affecting the signal and the sub-dominant backgrounds	68
4.6.2.1	Trigger uncertainty	68
4.6.2.2	Jet momentum uncertainties	68
4.6.2.3	Jet mass uncertainties	69
4.6.2.4	V -tagging uncertainties	69
4.6.2.5	b-tagging uncertainties	70
4.6.2.6	Missing Energy uncertainties	71
4.6.2.7	Pile-up uncertainty	72
4.6.2.8	QCD renormalization and factorization scale uncertainties	72
4.6.2.9	PDF	73
4.6.3	Summary	73
5	Conclusions	77

1

Abstract

2

3

CONTENTS

Introduction

Theoretical motivation

The Large Hadron Collider and the CMS experiment

Search for diboson resonances in the $VZ \rightarrow q\bar{q} \nu\bar{\nu}$ final state

4.1 Analysis overview

- This analysis searches for potential signals of heavy resonances decaying into a pair of vector bosons, using the data collected by the CMS experiment during 2016, corresponding to an integrated luminosity of $\mathcal{L} = 35.9 \text{ fb}^{-1}$. One of the boson should be a Z , and it is identified through its invisible decay into a couple of neutrinos ($\nu\bar{\nu}$), while the other electroweak boson, labelled as V and consisting either in a W or in a Z boson, is required to decay hadronically into a pair of quarks ($q\bar{q}$). The decay products (the bosons) of heavy (around the TeV scale) resonances are produced with large Lorentz boosts; as a consequence, the decay products of the bosons (quarks and neutrinos) are expected to be highly energetic and collimated. In this regime, the standard jet reconstruction algorithms fail in distinguishing the two jets from the quarks, suggesting to look for a signature composed of a large-cone high- p_T jet, in which both q and \bar{q} lie, recoiling against a large amount of missing transverse momentum (\vec{p}_T^{miss}) due to the neutrinos escaping the detector. The hadronically decaying boson (Z, W) is then reconstructed as one large-cone jet, whose mass is used to define the signal region and signal-depleted control regions, the sidebands. Jet substructure techniques are exploited in order to suppress background contamination and to classify the events in two exclusive signal purity categories, allowing to improve the discovery reach.
- A general ZZ decay, predicted by the bulk graviton model (sec. ??), can be reconstructed both in final states with high signal purity but limited statistics (four charged leptons) and large statistics but overwhelming backgrounds (no charged leptons). The choice to look for one boson decaying hadronically and the other Z into neutrinos represents the best compromise between these two extremes. This topology can be also utilized to reconstruct a charged spin-1 vector boson W' decaying into an invisible Z and an hadronic W , predicted by the HVT model (sec. ??), making this analysis sensitive to a generic VZ final state.
- Signal events are collected with trigger paths requiring high \vec{p}_T^{miss} recoiling against jet activity. This signature is clearly a very challenging one in an environment with more than 50 primary collisions per bunch crossing. For this reason, the Particle-Flow algorithm is run at trigger level to obtain the highest possible resolution on the jets and thus on the \vec{p}_T^{miss} .
- The search is performed by examining the distribution of the diboson reconstructed transverse

44 mass of the resonance VZ (m_{VZ}^T) for a localized excess. The shape and normalization of the main
 45 background of the analysis (namely, the production of an electroweak boson in association with
 46 jets) are estimated with a data-simulation hybrid approach using the distribution of data in the
 47 sidebands, corrected for a function accounting for potential differences between the signal region
 48 and the sidebands. The predictions of the secondary background sources completely rely on simu-
 49 lations.

50 In fig. 4.1, a typical signal event of the $W' \rightarrow WZ \rightarrow q\bar{q}' \nu\bar{\nu}$ process detected by the CMS experiment
 51 is displayed; the mass of the W' is 2.5 TeV. The muon chambers in the barrel (DTs, in light red) and
 52 in the endcaps (CSCs, in light blue), along with the tracker detector (green) are shown in the (r, φ)
 53 transverse plane (left) and the (r, z) longitudinal plane (right). The large-cone jet, identifying the
 54 W hadronic decay, is displayed in red; the energy deposits in ECAL (light orange) and in HCAL (in
 55 violet) can be seen in the picture. The missing transverse energy, signature of the Z invisible decay,
 56 is represented as a blue arrow, lying in the transverse plane. The track multiplicity (green tracks) is
 57 shown in the center of the detector, where the tracker is installed.

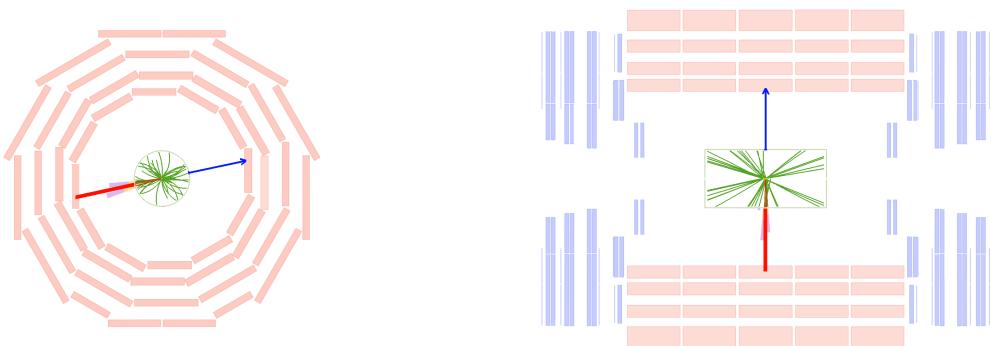


Figure 4.1: Left: representation of the decay of a W' of mass 2.5 TeV, in the transverse plane of the CMS detector. Right: representation of the decay of a W' of mass 2.5 TeV, in the longitudinal plane of the CMS detector.

58 4.2 Data and Monte Carlo simulations samples

59 4.2.1 Signal samples

60 Signal samples of a spin-2 (bulk graviton) decaying into a pair of Z bosons are exploited in the anal-
 61 ysis. To target the final state, one of the two Z bosons is forced to decay into neutrinos, while the
 62 other Z is forced to decay hadronically. The signal samples are produced in the narrow-width ap-
 63 proximation by setting the resonance width to 0.1% of its mass. Twelve mass points with 100000
 64 events each are simulated, with a m_G ranging from 600 GeV up to 4500 GeV.

65 Additionally, samples of a spin-1 HVT-like W' resonance decaying in a Z boson and a W boson
 66 are studied. The Z boson is forced to decay into neutrinos, and the W boson is forced to decay
 67 hadronically. Also in this case the signal samples are produced in the narrow-width approximation
 68 by setting the resonance width to 0.1% of its mass. Twelve mass points with 100000 events each are
 69 simulated, with a $m_{W'}$ ranging from 600 GeV up to 4500 GeV.

70 The signal samples are generated at leading-order (LO) with the MADGRAPH5_AMCATNLO v 2.2.2 [1]
 71 matrix element generator, while hadronization and fragmentation are handled by PYTHIA 8 [2] ver-
 72 sion 8.2121 with CUETP8M1 [3] tuning. A full detector simulation and event reconstruction has

4.2 Data and Monte Carlo simulations samples

- 73 been performed with GEANT4 [4] and CMSSW. The detector alignment scenario, calibrations and
 74 pile-up distributions are generated according to the expectations in 2016 data.
 75 All the signal samples used in the analysis and the related properties are reported in Tables 4.1-4.2.

Table 4.1: Spin-2 (bulk graviton) signal samples and production cross sections (assumed to be 1 pb) multiplied by the respective branching fractions of the Z decays considered ($\mathcal{B}(Z \rightarrow \nu\nu) = 0.20$, $\mathcal{B}(Z \rightarrow qq) = 0.6991$). A combinatorial factor of 2 is included in the cross-section calculation.

Signal process	m_G	Events	$\sigma \times \mathcal{B}$ (pb)
$G \rightarrow ZZ \rightarrow q\bar{q}\nu\bar{\nu}$	600 GeV	100000	0.27964
$G \rightarrow ZZ \rightarrow q\bar{q}\nu\bar{\nu}$	800 GeV	100000	0.27964
$G \rightarrow ZZ \rightarrow q\bar{q}\nu\bar{\nu}$	1000 GeV	100000	0.27964
$G \rightarrow ZZ \rightarrow q\bar{q}\nu\bar{\nu}$	1200 GeV	100000	0.27964
$G \rightarrow ZZ \rightarrow q\bar{q}\nu\bar{\nu}$	1400 GeV	100000	0.27964
$G \rightarrow ZZ \rightarrow q\bar{q}\nu\bar{\nu}$	1800 GeV	100000	0.27964
$G \rightarrow ZZ \rightarrow q\bar{q}\nu\bar{\nu}$	2000 GeV	100000	0.27964
$G \rightarrow ZZ \rightarrow q\bar{q}\nu\bar{\nu}$	2500 GeV	100000	0.27964
$G \rightarrow ZZ \rightarrow q\bar{q}\nu\bar{\nu}$	3000 GeV	100000	0.27964
$G \rightarrow ZZ \rightarrow q\bar{q}\nu\bar{\nu}$	3500 GeV	100000	0.27964
$G \rightarrow ZZ \rightarrow q\bar{q}\nu\bar{\nu}$	4000 GeV	100000	0.27964
$G \rightarrow ZZ \rightarrow q\bar{q}\nu\bar{\nu}$	4500 GeV	100000	0.27964

Table 4.2: Spin-1 (W') signal samples and production cross sections (assumed to be 1 pb) multiplied by the Z and W branching fraction ($\mathcal{B}(Z \rightarrow \nu\nu) = 0.2$, $\mathcal{B}(W \rightarrow qq) = 0.6760$).

Signal process	$m_{W'}$	Events	$\sigma \times \mathcal{B}$ (pb)
$W' \rightarrow WZ \rightarrow q\bar{q}'\nu\bar{\nu}$	600 GeV	100000	0.13482
$W' \rightarrow WZ \rightarrow q\bar{q}'\nu\bar{\nu}$	800 GeV	100000	0.13482
$W' \rightarrow WZ \rightarrow q\bar{q}'\nu\bar{\nu}$	1000 GeV	100000	0.13482
$W' \rightarrow WZ \rightarrow q\bar{q}'\nu\bar{\nu}$	1200 GeV	100000	0.13482
$W' \rightarrow WZ \rightarrow q\bar{q}'\nu\bar{\nu}$	1400 GeV	100000	0.13482
$W' \rightarrow WZ \rightarrow q\bar{q}'\nu\bar{\nu}$	1800 GeV	100000	0.13482
$W' \rightarrow WZ \rightarrow q\bar{q}'\nu\bar{\nu}$	2000 GeV	100000	0.13482
$W' \rightarrow WZ \rightarrow q\bar{q}'\nu\bar{\nu}$	2500 GeV	100000	0.13482
$W' \rightarrow WZ \rightarrow q\bar{q}'\nu\bar{\nu}$	3000 GeV	100000	0.13482
$W' \rightarrow WZ \rightarrow q\bar{q}'\nu\bar{\nu}$	3500 GeV	100000	0.13482
$W' \rightarrow WZ \rightarrow q\bar{q}'\nu\bar{\nu}$	4000 GeV	100000	0.13482
$W' \rightarrow WZ \rightarrow q\bar{q}'\nu\bar{\nu}$	4500 GeV	100000	0.13482

76 4.2.2 Signal characterization

- 77 This analysis is performed in a high mass region (from 1 TeV to 4.5 TeV). The MADGRAPH algorithm
 78 generates the hard process production in the collision with $p_T = 0$. In the next step of the simula-
 79 tion, during the hadronization, PYTHIA adds the QCD ISR (initial state radiation) and consequently a
 80 resonance p_T different from 0. Kinematical distributions at generator level are shown in fig. 4.2-4.4
 81 for spin-2 bulk graviton signal, and in fig. 4.5-4.7 for spin-1 HVT W' signal.

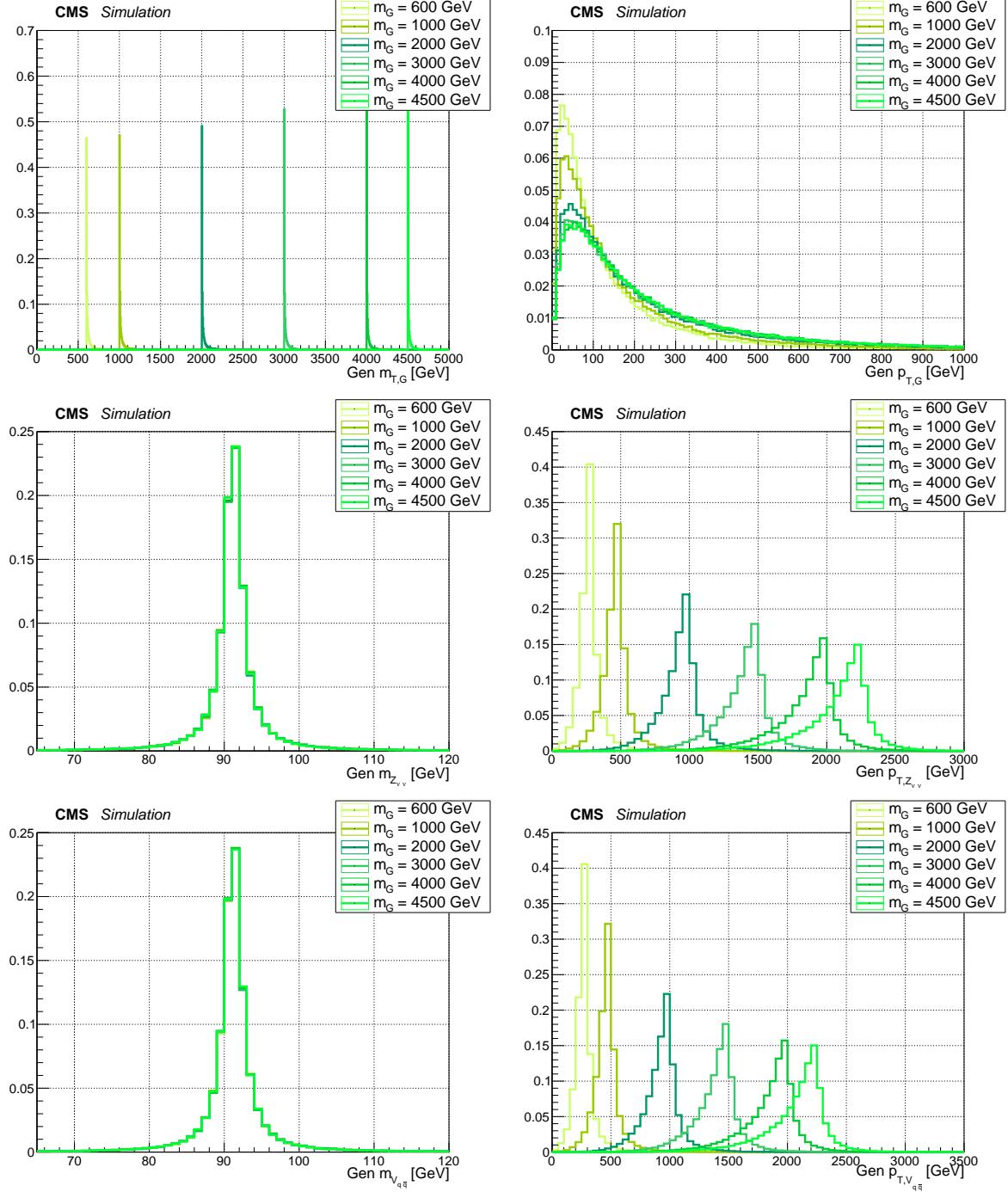


Figure 4.2: Main signal kinematic quantities at generation level after parton showering, for spin-2 bulk graviton signal, considering different mass hypotheses ($m_G = 0.6, 1, 2, 3, 4, 4.5 \text{ TeV}$). Top: graviton transverse mass and p_T distributions. Center: invisibly decaying Z mass and p_T . Bottom: hadronically decaying Z mass and p_T .

4.2 Data and Monte Carlo simulations samples

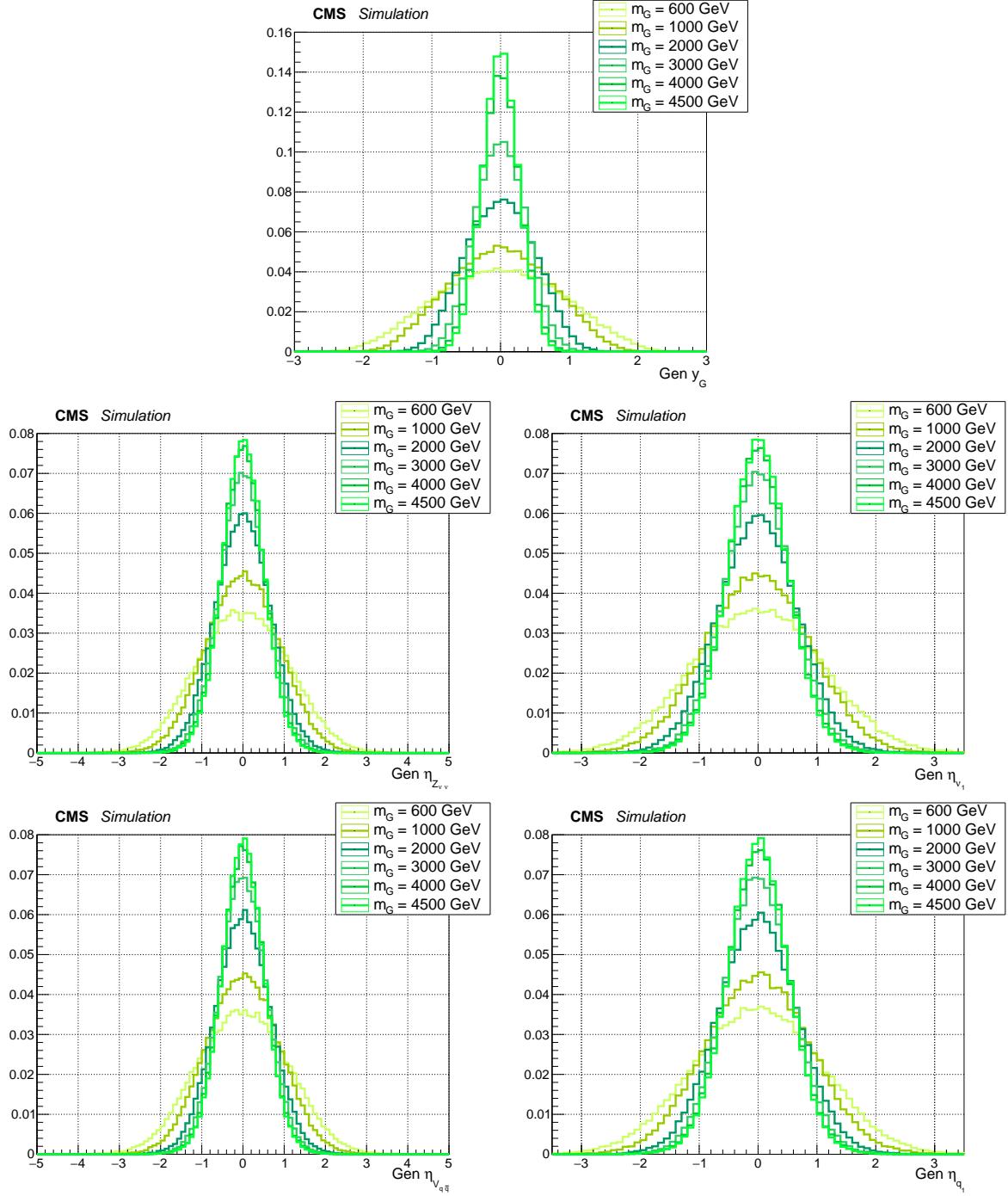


Figure 4.3: Main signal kinematic quantities at generation level after parton showering, for spin-2 bulk graviton signal, considering different mass hypotheses ($m_G = 0.6, 1, 2, 3, 4, 4.5$ TeV). Top: graviton rapidity \mathcal{Y} . Center: pseudorapidity η of the invisibly decaying Z , and pseudorapidity of the leading neutrino. Bottom: pseudorapidity η of the hadronically decaying Z , and pseudorapidity of the leading quark.

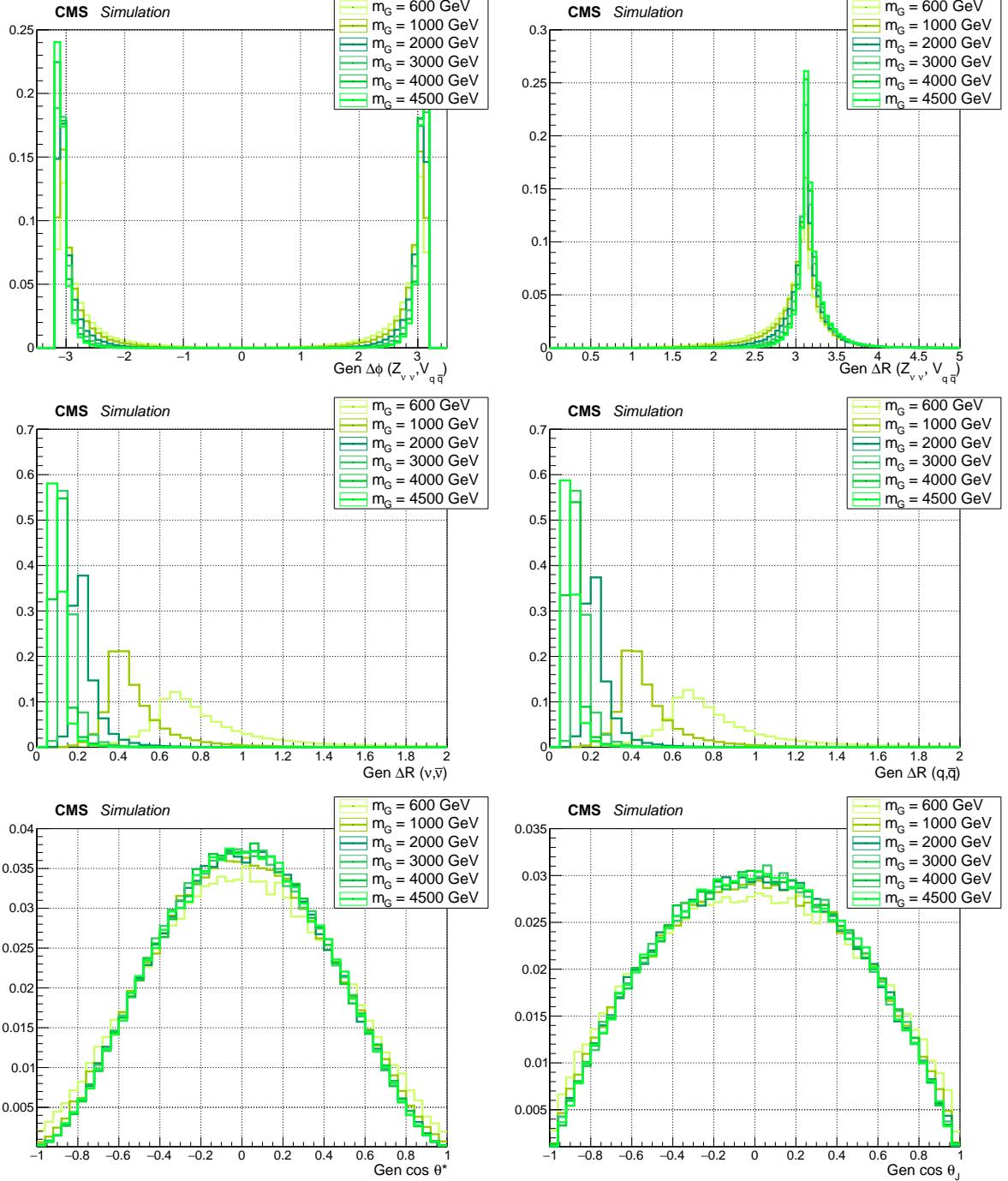


Figure 4.4: Main signal kinematic quantities at generation level after parton showering, for spin-2 bulk graviton signal, considering different mass hypotheses ($m_G = 0.6, 1, 2, 3, 4, 4.5$ TeV). Top: angular separation in the transverse plane $\Delta\varphi$ (left) and solid angle ΔR (right) between leptonic Z and hadronic Z . Center: solid angle between the neutrinos and the quarks. Bottom: distribution of $\cos \theta^*$ and $\cos \theta_j$ (described in text).

4.2 Data and Monte Carlo simulations samples

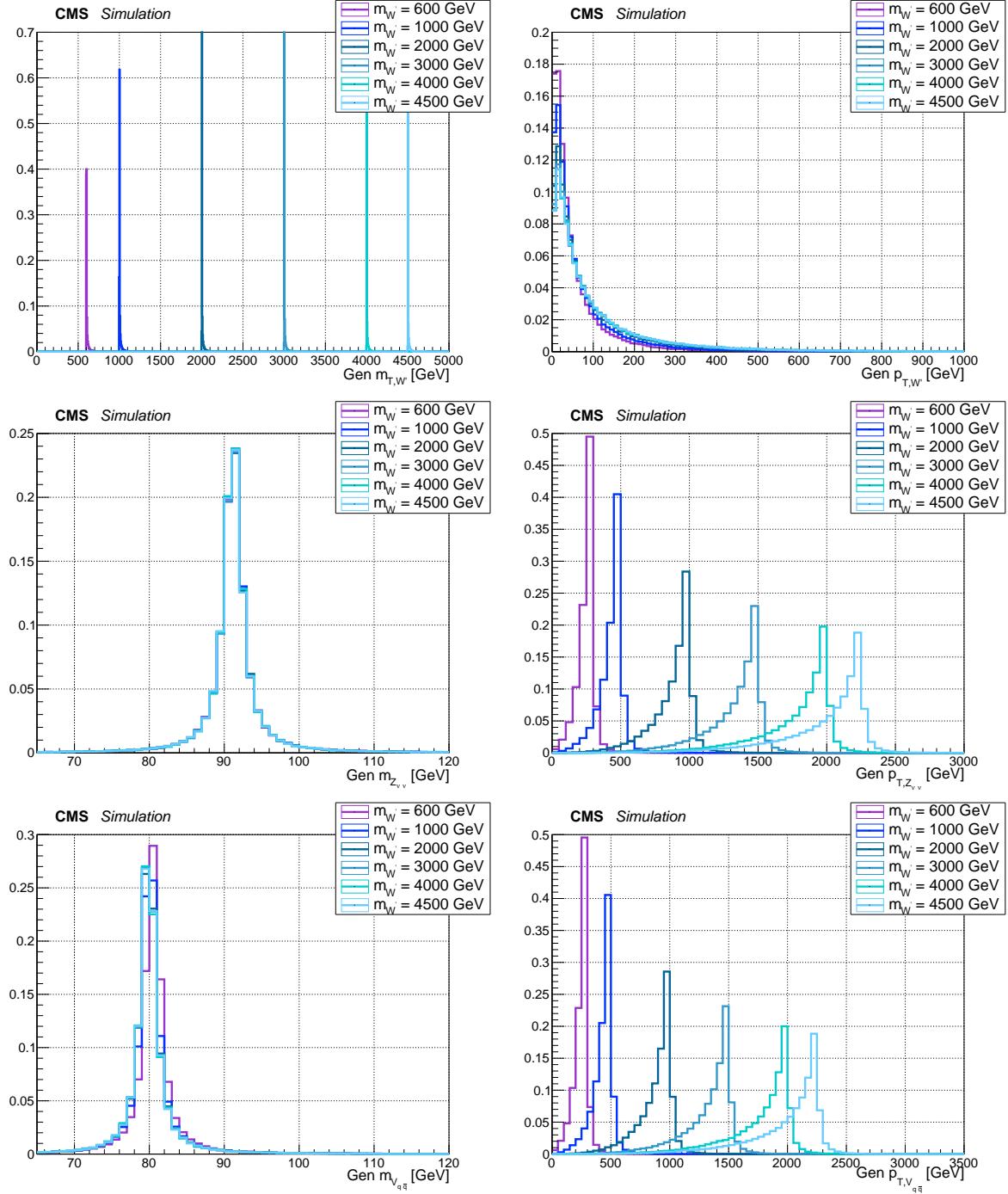


Figure 4.5: Main signal kinematic quantities at generation level after parton showering, for spin-1 W' signal, considering different mass hypotheses ($m_{W'} = 0.6, 1, 2, 3, 4, 4.5 \text{ TeV}$). Top: W' transverse mass and p_T distributions. Center: invisibly decaying Z mass and p_T . Bottom: hadronically decaying W mass and p_T .

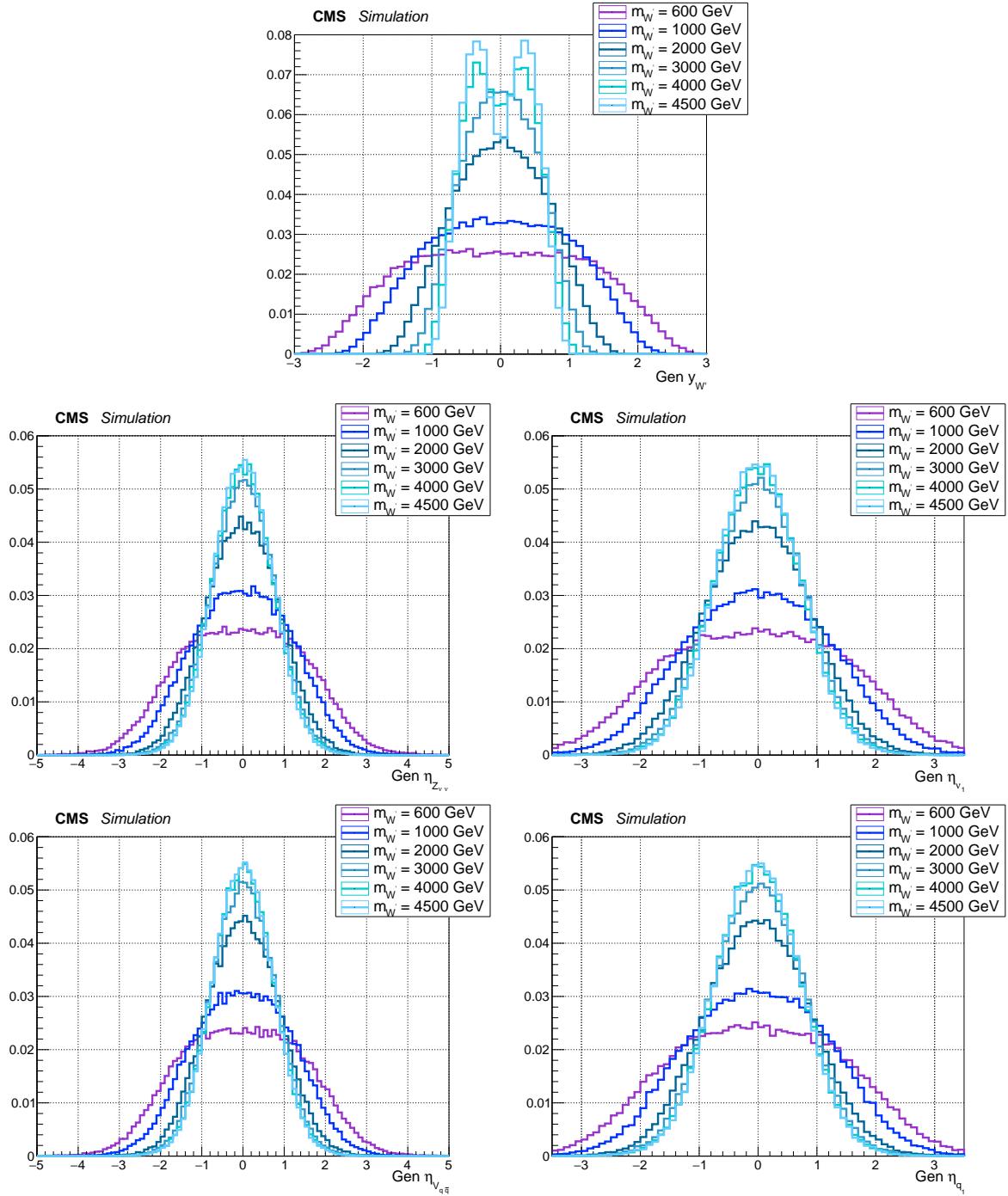


Figure 4.6: Main signal kinematic quantities at generation level after parton showering, for spin-1 W' signal, considering different mass hypotheses ($m_{W'} = 0.6, 1, 2, 3, 4, 4.5$ TeV). Top: W' rapidity \mathcal{Y} . Center: pseudorapidity η of the invisibly decaying Z , and pseudorapidity of the leading neutrino. Bottom: pseudorapidity η of the hadronically decaying W , and pseudorapidity of the leading quark.

4.2 Data and Monte Carlo simulations samples

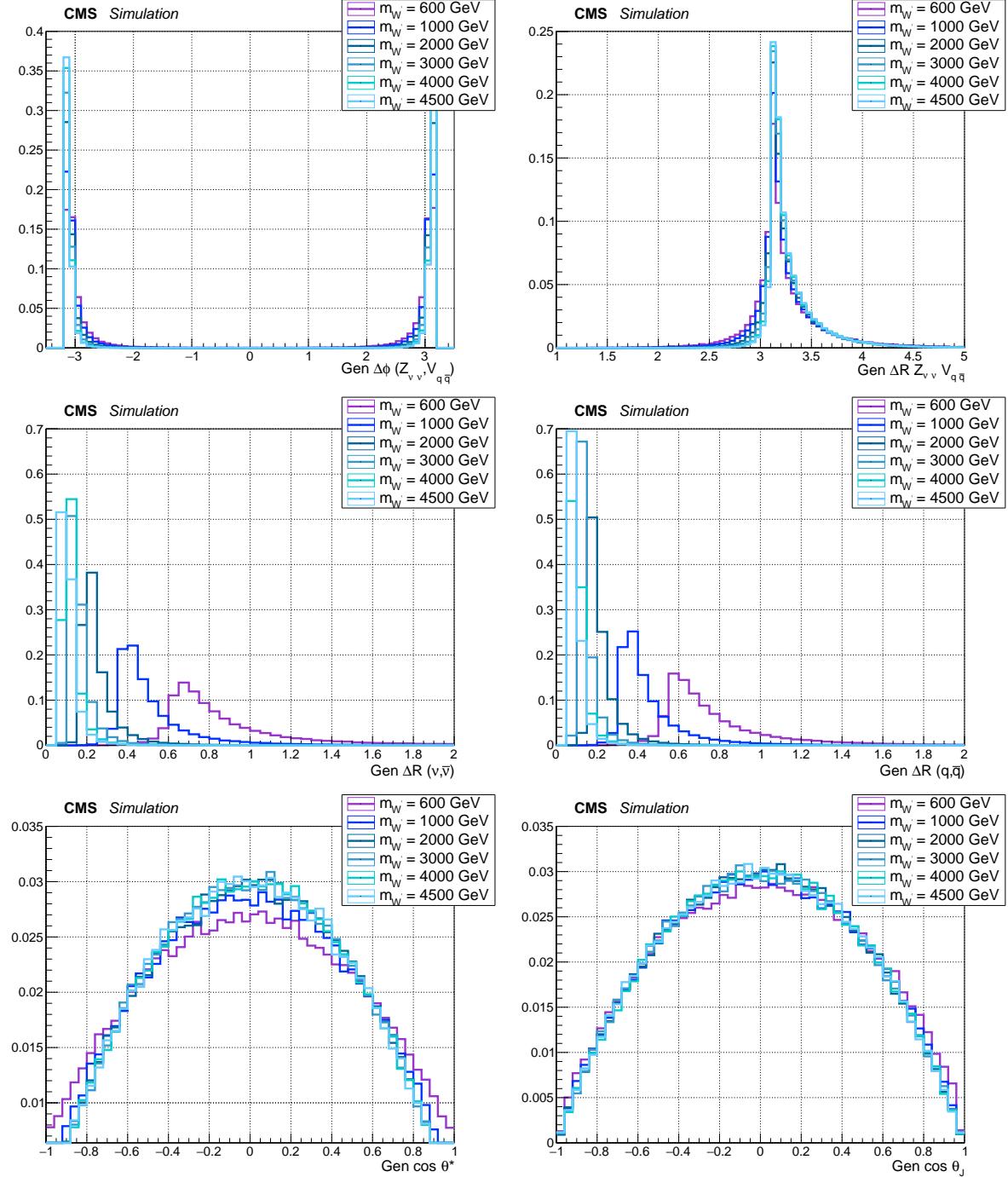


Figure 4.7: Main signal kinematic quantities at generation level after parton showering, for spin-1 W' signal, considering different mass hypotheses ($m_{W'} = 0.6, 1, 2, 3, 4, 4.5 \text{ TeV}$). Top: angular separation in the transverse plane $\Delta\varphi$ (left) and solid angle ΔR (right) between leptonic Z and hadronic W . Center: solid angle between the neutrinos and the quarks. Bottom: distribution of $\cos \theta^*$ and $\cos \theta_J$ (described in text).

82 Angular distributions are related to the spin, the polarization and the kinematics of the produced
 83 resonance; in particular:

- 84 • the ΔR among neutrinos and quarks reflect the boosted nature of the electroweak bosons:
 85 the more massive the resonance, the larger the boost, and hence the closer the fermions. By
 86 looking at fig. 4.4-4.7, with a jet clustering parameter of 0.8 (AK8 jet) it is possible to enclose
 87 the quarks produced by the decay of the V boson, for a resonance mass over 1 TeV;
- 88 • the $\cos \theta^*$, namely the cosine of the angle between the momentum of the V boson, calculated
 89 in the resonance rest frame, and the flight direction of the resonance itself in the laboratory
 90 frame. This variable depends on the spin of the diboson resonance (spin-2 and spin-1 distri-
 91 butions are different, fig. 4.4-4.7).
- 92 • the $\cos \theta_J$, the cosine of the angle between the momentum of the leading quark, calculated
 93 in the V rest frame, and the flight direction of the V boson in the laboratory frame. This vari-
 94 able depends on the polarization state of the decay bosons [5]; in both HVT and bulk graviton
 95 model, electroweak bosons are expected to be longitudinally polarized. When $\cos \theta_J \rightarrow 0$,
 96 quarks are produced very close in angle and hence it is difficult to disentangle the two sub-
 97 structures in the large-cone jet (sec. 4.3.8); when $\cos \theta_J \rightarrow \pi$ the quarks are emitted asymmet-
 98 rically (one is softer than the other).

99 4.2.3 Background samples

100 The physics processes yielding final states with two neutrinos in association with a pair of quarks
 101 are considered as sources of background; they are listed in tab. 4.3, along with the expected cross-
 102 sections at next-to-leading order (NLO) or next-to-next-to leading (NNLO). A summary of the stan-
 103 dard model cross-sections, measured by CMS, and their theoretical predictions is included in fig. 4.8-4.9.

- 104 • **$Z + \text{jets}$** : this process represents the main irreducible background for the signal. The produc-
 105 tion of a Z boson in association with one or more partons in the final state has topology that
 106 is similar to the signal. This $Z + \text{jets}$ background is produced in samples binned in p_T of the
 107 Z boson, starting from 100 GeV, with the AMC@NLO generator, with FXFX merging [6]. The
 108 contribution from events with $p_T < 100$ GeV is negligible after the requirement on the \vec{p}_T^{miss}
 109 to be greater than 200 GeV (sec. 4.3.12).
- 110 • **$W + \text{jets}$** : the leptonic decay of a W boson can be an irreducible background if the charged
 111 lepton escapes undetected (*i.e.* outside the detector acceptance) or fails the lepton identifi-
 112 cation requirements. The production of a W boson has a cross section larger by an order
 113 of magnitude with respect to the Z , and this makes the $W + \text{jets}$ a relevant background also
 114 when a lepton veto is applied. This $W + \text{jets}$ background is produced in samples binned in
 115 p_T of the W boson, starting from 100 GeV, with the AMC@NLO generator.
- 116 • **Top**: the pair and single production of top quarks represent a source of background, given
 117 the decay chain $t \rightarrow b W$, always including an electroweak W boson and a b-quark. The $t \bar{t}$
 118 pair production manifests as two b-jets and two W bosons in the final state, that can decay to
 119 leptons that escape the detector or fail to be identified as leptons. This analysis makes use of
 120 $t \bar{t}$ inclusive decays samples based on POWHEG v2 [7] NLO generator. Single-top and single-
 121 antitop samples are produced in the 5-flavours scheme using POWHEG v2 [8] NLO generator.
 122 Different production mechanisms are considered: $t W$ channel, when a top quark is produced
 123 in association with a W boson, due to a gluon-bottom quark scattering; s-channel, due to

4.2 Data and Monte Carlo simulations samples

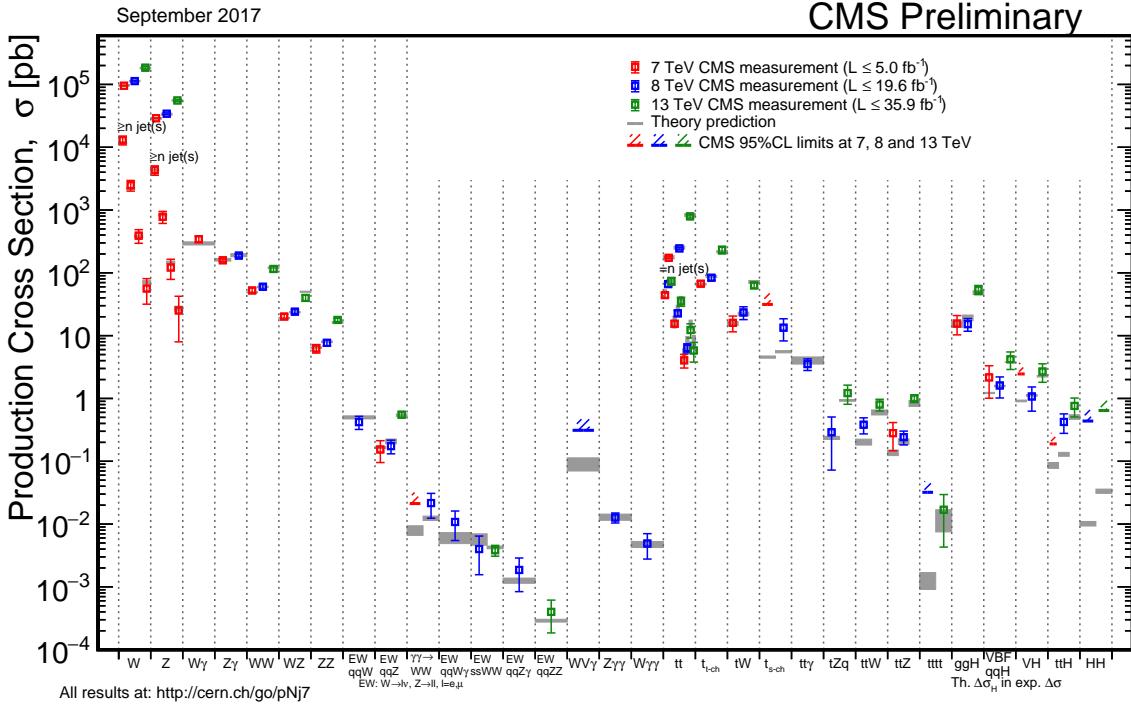


Figure 4.8: Production cross-sections of the main standard model processes, as measured by CMS, and theoretical predictions.

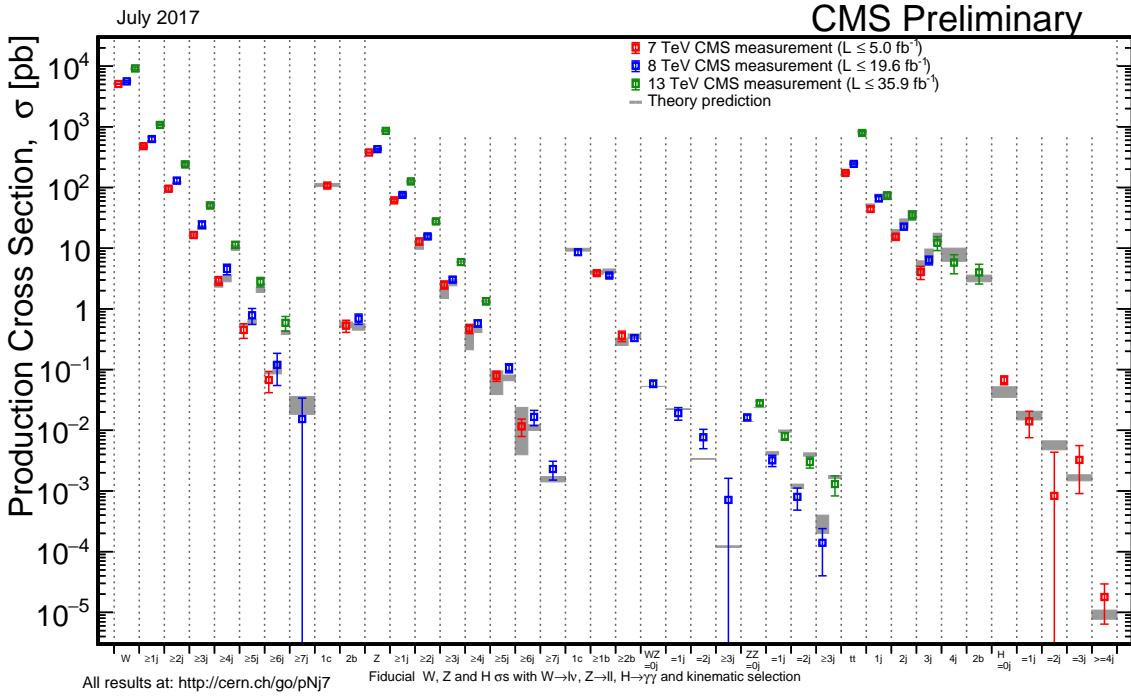


Figure 4.9: Production cross-sections of the standard model processes involving a vector boson in association with jets, as measured by CMS, and theoretical predictions. These phenomena represent the main background sources for the analysis.

124 quark-antiquark scattering, producing a top and an anti-bottom quark in the final state; t-
 125 channel, via a virtual W in a quark-b-quark scattering, resulting in a top quark and a quark
 126 jet in the final state.

- 127 • **Diboson:** the SM production of a pair of vector bosons is topologically close to the searched
 128 signal, by the way the cross-section of the process is low. The WW production is the most
 129 probable process, that imitates the signal when one of the W decays leptonically and the
 130 charged lepton fall outside the detector acceptance or it is mis-identified; the WZ and ZZ
 131 processes have smaller cross-sections but are topologically identical to the signal, except for
 132 the fact that the invariant mass of the diboson system has a smoothly falling spectrum, in
 133 contrast to the resonant signal distribution. Inclusive diboson production processes (WW ,
 134 WZ , ZZ) are considered; they are simulated at LO by PYTHIA generator.
 135 • **Multi-jet:** despite the production of multi-jet events, initiated by the strong interactions, has
 136 a very large cross-section, this source of background is suppressed by a dedicated selection
 137 and hence negligible for the analysis (sec. 4.3.12).

Table 4.3: Simulated Monte Carlo samples. The cross-section \times branching fraction for each process is shown in pb.

Signal process	Kinematical cuts	Generator	$\sigma \times \mathcal{B}$ [pb]	N. of events
$Z \rightarrow \nu\nu + \text{jets}$	$100 < p_{T,Z} < 250 \text{ GeV}$	amcatnloFXFX – Pythia8	170.4	10710313
$Z \rightarrow \nu\nu + \text{jets}$	$250 < p_{T,Z} < 400 \text{ GeV}$	amcatnloFXFX – Pythia8	6.636	2112619
$Z \rightarrow \nu\nu + \text{jets}$	$400 < p_{T,Z} < 650 \text{ GeV}$	amcatnloFXFX – Pythia8	0.9372	1101297
$Z \rightarrow \nu\nu + \text{jets}$	$p_{T,Z} > 650 \text{ GeV}$	amcatnloFXFX – Pythia8	0.1042	2047215
$W \rightarrow \ell \nu + \text{jets}$	$100 < p_{T,W} < 250 \text{ GeV}$	amcatnloFXFX – Pythia8	676.3	20178260
$W \rightarrow \ell \nu + \text{jets}$	$250 < p_{T,W} < 400 \text{ GeV}$	amcatnloFXFX – Pythia8	23.94	2001382
$W \rightarrow \ell \nu + \text{jets}$	$400 < p_{T,W} < 650 \text{ GeV}$	amcatnloFXFX – Pythia8	3.031	1939947
$W \rightarrow \ell \nu + \text{jets}$	$p_{T,W} < 650 \text{ GeV}$	amcatnloFXFX – Pythia8	0.4524	1974609
$t\bar{t}$ inclusive	-	Powheg – Pythia8	831.76	77229341
$t(\bar{t}W \text{ channel})$	-	Powheg – Pythia8	35.85	6952830
5f inclusive				
$\bar{t}(\bar{t}W \text{ channel})$	-	Powheg – Pythia8	35.85	6933094
5f inclusive				
t (s-channel)	-	amcatnloFXFX – Pythia8	3.344	622990
4f lepton decays				
t (t-channel)	-	Powheg – Madspin – – Pythia8	136.02	67240808
4f inclusive				
\bar{t} (t-channel)	-	Powheg – Madspin – – Pythia8	80.95	38811017
4f inclusive				
WW inclusive	-	Pythia8	118.7	7981136
WZ inclusive	-	Pythia8	47.2	3995828
ZZ inclusive	-	Pythia8	16.6	1988098

138 4.2.4 Vector boson momentum corrections

139 Corrections to the p_T spectrum of the V boson, due to NLO electroweak contributions, are en-
 140 hanced at TeV scale [9], and they become significant for the purpose of this search. These cor-

4.2 Data and Monte Carlo simulations samples

reactions are effectively applied on a per-event basis, depending on the p_T of the vector boson at generation level. Figure 4.10 shows the amount of the corrections for the W and Z bosons.

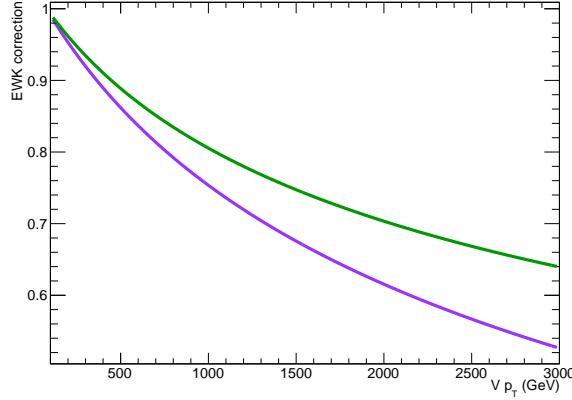


Figure 4.10: Electroweak corrections for the Z (green line) and W boson (purple line) as a function of the transverse momentum of the boson [9].

4.2.5 Data samples

Data samples used in this analysis have been collected during proton-proton collisions produced at LHC in 2016, at a center-of-mass energy of 13 TeV, with colliding bunches spaced by 25 ns, and with the magnetic field enabled. Three group of datasets have been considered:

- the MET dataset, where the analysis is performed, is collected by triggers requiring a large amount of \vec{p}_T^{miss} at HLT level in the event;
- the SingleMuon dataset, used to perform an unbiased trigger efficiency estimate, is collected by triggers requiring at least one well defined muon at HLT level;
- the SingleElectron dataset, used as cross-check for the trigger efficiency estimation, is collected by triggers requiring at least one well defined electron at HLT level.

Data selected for the analysis include all the runs certified as “good” for all subsystems. The corresponding integrated luminosity amounts to $35.9 \pm 0.9 \text{ fb}^{-1}$ [10]. In order to remove problematic or noise-dominated events, dedicated E_T^{miss} filters have been applied on data (and simulations).

4.2.6 Trigger

The most remarkable feature of the signal topology is the presence of a boosted Z decaying into neutrinos; the natural choice for the trigger requirement is to filter data firing at least one of the \vec{p}_T^{miss} trigger HLT paths listed in tab. 4.4, along with their corresponding L1 missing energy or jet seeds. PFMETNoMu indicates the E_T^{miss} (no μ) quantity, defined as the magnitude of the missing transverse momentum, reconstructed with the Particle-Flow algorithm at HLT, removing the muon candidates from the vector sum. PFMHTNoMu indicates the missing hadronic activity H_T^{miss} (no μ), defined as the magnitude of the vector sum of the transverse momenta of the jets, reconstructed with the Particle-Flow algorithm at HLT, once the muon candidates have been removed. PFMET indicates the pure E_T^{miss} calculated with Particle-Flow algorithm at HLT; different filters are applied

Table 4.4: HLT trigger paths used in the analysis.

HLT path	L1 seeds
HLT_PFMETNoMu90_PFMHTNoMu90_IDTight	L1_ETM70 OR L1_DoubleJetC56_ETM60 OR L1_ETM60 OR L1_ETM50
HLT_PFMETNoMu110_PFMHTNoMu110_IDTight	L1_ETM70 OR L1_DoubleJetC56_ETM60 OR L1_ETM60 OR L1_ETM50
HLT_PFMETNoMu120_PFMHTNoMu120_IDTight	L1_ETM70 OR L1_DoubleJetC56_ETM60 OR L1_ETM60 OR L1_ETM50
HLT_PFMET170_NoiseCleaned or	L1_ETM60 OR L1_ETM70
HLT_PFMET170_JetIdCleaned or	L1_ETM60 OR L1_ETM70
HLT_PFMET170_HBHECleaned	L1_ETM60 OR L1_ETM70

166 at HLT (cleaning events from noise in the detector). Different thresholds are applied to E_T^{miss} (no μ)
 167 and H_T^{miss} (no μ).

168 The approach adopted in this analysis consists in calculating the trigger efficiency on data, and
 169 applying the measured efficiency to Monte Carlo samples. Therefore the trigger is not required to
 170 have been fired in MC.

171 Given that the final state probed by the analysis consists into one AK8 jet, large E_T^{miss} and no charged
 172 leptons, an unbiased measurement of the E_T^{miss} trigger efficiency can be performed in an orthogo-
 173 nal dataset, collected with different triggers, and requiring events where a $W \rightarrow \ell \nu$ leptonic decay
 174 is taking place. This guarantees the presence of real \vec{p}_T^{miss} in the event, due to the neutrino; fur-
 175 thermore, the presence of a charged lepton guarantees that the leptonic W -like events are not over-
 176 lapped with the search region. The additional requirement to have at least one AK8 jet is applied in
 177 the trigger measurement, in order to probe a kinematical region similar to that of the signal region
 178 of the analysis.

179 The efficiency of the E_T^{miss} triggers is measured on SingleMuon dataset by selecting $W \rightarrow \mu \nu$ events
 180 using a logic or of single muon triggers HLT_IsoMu24 OR HLT_IsoTkMu24_v, namely, triggers
 181 asking for a PF muon reconstructed at HLT, with a p_T threshold of 24 GeV, that is isolated (in the
 182 whole reconstruction or at tracker level only). Offline selections consist in asking to have one iso-
 183 lated muon, with a suitable p_T threshold to be in the plateau of the muon trigger. The efficiency has
 184 been calculated as a function of the minimum quantity between the offline reconstructed E_T^{miss} (no
 185 μ):

$$E_T^{\text{miss}} (\text{no } \mu) = \left| \vec{p}_T^{\text{miss}} + \sum_i \vec{p}_T^{\mu,i} \right|, \quad (4.1)$$

186 where the contribution of all the offline PF muons is removed from the \vec{p}_T^{miss} computation as in the
 187 online algorithm, and the offline H_T^{miss} , defined as

$$H_T^{\text{miss}} = \left| \sum_j^{\text{n. of AK4 jets}} p_T^j \right|. \quad (4.2)$$

188 This approach guarantees to mimic the behaviour of the online L1 trigger seeds. The detailed se-
 189 lections are listed below:

- 190 • HLT_IsoMu24_v OR HLT_IsoTkMu24_v,

4.2 Data and Monte Carlo simulations samples

- 191 • 1 isolated muon $p_T > 35$ GeV, identified with tight requirements,
192 • at least one AK8 jet, $p_T > 170$ GeV, $|\eta| < 2.5$, identified with loose requirements,
193 • AK4 jets included in H_T^{miss} : $p_T > 30$ GeV, $|\eta| < 2.5$, identified with loose requirements.
- 194 The efficiency of the E_T^{miss} triggers has independently been measured also on SingleElectron
195 dataset, by selecting $W \rightarrow e \nu$ events using a single electron trigger (HLT_Ele27_WP Loose_Gsf
196 OR HLT_Ele27_WP Tight_Gsf OR HLT_Ele32_WP Tight_Gsf), asking to have one well identified
197 electron, with a suitable p_T threshold, and asking to the electron and the \vec{p}_T^{miss} to be separated in
198 the transverse plane (hence, in φ) in order to suppress fake jet events mis-identified as electrons at
199 trigger level ($\Delta\varphi > 0.5$). The detailed selections are listed below:
- 200 • HLT_Ele27_WP Loose_Gsf OR HLT_Ele27_WP Tight_Gsf OR HLT_Ele32_WP Tight_Gsf,
201 • 1 electron, $p_T > 35$ GeV, identified with tight requirements,
202 • at least one AK8 jet, $p_T > 170$ GeV, $|\eta| < 2.5$, identified with loose requirements,
203 • AK4 jets included in H_T^{miss} : $p_T > 30$ GeV, $|\eta| < 2.5$, identified with loose requirements.
- 204 All the available data have been employed to derive the efficiency. The final turn-on curves for the
205 E_T^{miss} triggers are shown in fig.4.11-4.12, measured in muon and electron dataset respectively. The
206 PFMETNoMu trigger efficiencies are displayed separately, together with their logic OR. The trigger
207 efficiency measured on SingleMuon dataset amounts to 96% at $E_T^{\text{miss}} = 200$ GeV; the trigger effi-
208 ciency measured on SingleElectron dataset amounts to 95% at $E_T^{\text{miss}} = 200$ GeV. The difference
209 needed to cover the gap between the two independent measurements is taken as trigger systematic
210 uncertainty, and it amounts to 1% at 200 GeV.

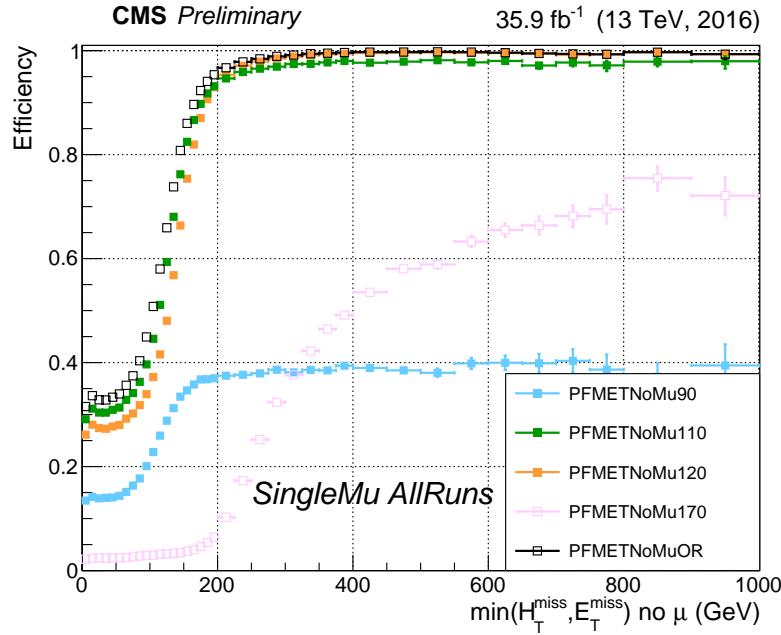


Figure 4.11: \vec{p}_T^{miss} trigger efficiency for the \vec{p}_T^{miss} trigger paths used in this analysis, calculated on SingleMuon dataset, as a function of the minimum of the variables E_T^{miss} (no μ) (eq. 4.1) and H_T^{miss} (eq. 4.2).

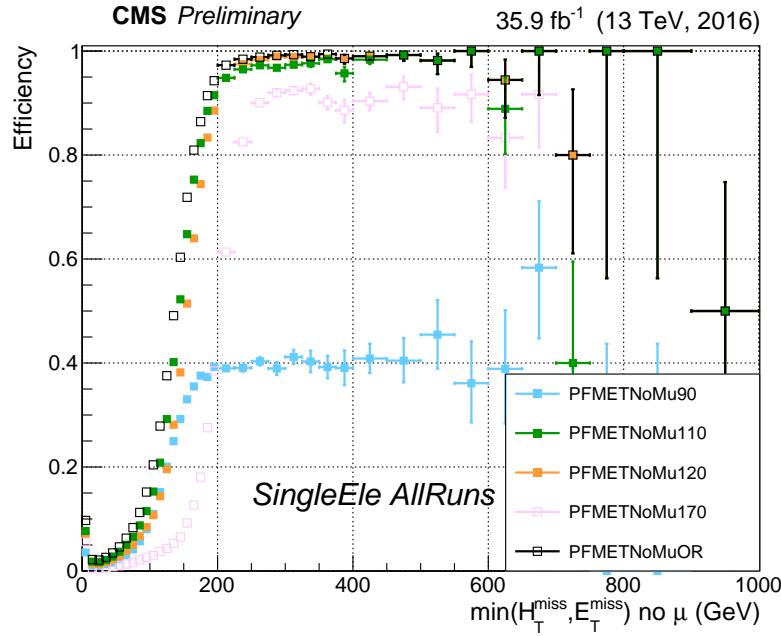


Figure 4.12: \vec{p}_T^{miss} trigger efficiency for the \vec{p}_T^{miss} trigger paths used in this analysis, calculated on SingleElectron dataset, as a function of the minimum of the variables E_T^{miss} (no μ) (eq. 4.1) and H_T^{miss} (eq. 4.2).

4.3 Event selection

In this section, the selections applied to the physics objects used in the analysis are presented and motivated by performance and validation plots. Background events are represented as coloured histograms: $Z + \text{jets}$ events in light blue, $W + \text{jets}$ events in violet, $t\bar{t}$ events in yellow, single-top events in orange, diboson (or VV) events in blue, multi-jet (QCD) events in gray. Background uncertainties are displayed as black shaded areas. Signal samples are represented as coloured shaded histograms: the kind of signal (graviton or W'), the mass and cross-section of the considered resonance are reported in the legend. Data are represented with black markers, with their corresponding Poissonian uncertainty bars. If data are displayed, the data-MC ratio is reported per each bin in the bottom panel, along with the overall data-MC ratio calculated in the whole spectrum and the scores of χ^2 and Kolmogorov-Smirnov goodness-of-fit tests.

4.3.1 Vertex and Pile-up

Due to the pile-up effect, several vertices are typically reconstructed in one event. The primary vertex of the event is defined as the one with the highest sum of transverse momenta $\sum p_T^2$ of clustered physics objects associated to it, which passes the following selections:

- number of degrees of freedom $N_{DoF} > 4$
- vertex position along the beampipe $|z_{vtx}| < 24\text{cm}$
- vertex distance with respect the beam pipe $d_0 < 2\text{cm}$

where z_{vtx} and d_0 are the distance along and perpendicular to the beam line of the vertex with respect the nominal interaction point $(0, 0, 0)$.

The Monte Carlo samples listed in sec. 4.2 are generated simulating the pile-up conditions, as expected in the 25 ns bunch crossing pile-up scenario. Nevertheless, the MC pile-up description does not match exactly the conditions in data, and there is therefore the need to reweight the simulated events in order to improve the agreement with the data.

The MC samples are reweighted assuming a total inelastic cross section of $\sigma_{in} = 69\,200\mu b$. The comparison between the distributions of primary vertices in data and MC after the pile-up reweighting is applied is shown in fig. 4.13 for an event selection (called inclusive selection, described in sec. 4.4) requiring large amount of \vec{p}_T^{miss} recoiling against an AK8 fat jet (tab. 4.12).

4.3.2 Electrons

Electrons considered in this analysis, reconstructed from energy deposits in the ECAL matched to tracks reconstructed in the silicon tracker, are required to pass the Particle-Flow criteria, and to fall in the ECAL pseudorapidity fiducial range ($|\eta| < 2.5$). The electron identification is defined with a “cut-based” approach. In the isolation definition, the effect of pile-up is considered by taking into account the energy deposits in the calorimeter, estimated through the so-called ρ -area method, by subtracting the median energy density in the event ρ multiplied by the electron energy deposits effective area. The isolation value is computed in a ΔR cone of 0.3 centered along the lepton direction.

Since in this analysis aims at a final state without any lepton, every electron identified with the looser cut-based criteria (*veto Id*) and transverse momentum $p_T > 10\text{ GeV}$ is vetoed. The detailed set of cuts to define a *veto Id* cut-based electron are reported in tab. 4.5; this set of selections allow to identify an electron with an efficiency of $\sim 95\%$. The supercluster width is indicated as $\sigma_{inj\eta}$; $\Delta\eta_{in}^{seed}$ and $\Delta\varphi_{in}$ are the difference in η and φ between the track position as it is measured in the inner layer,

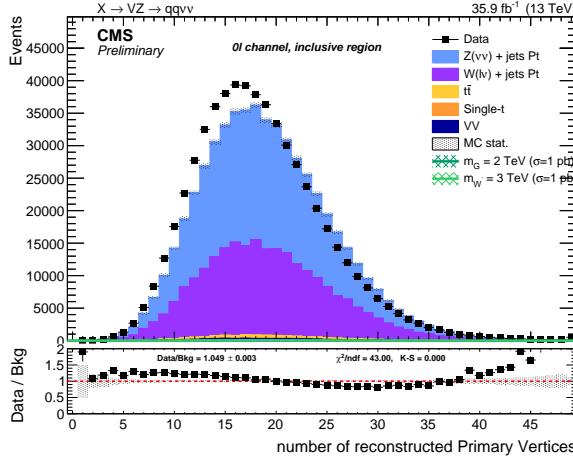


Figure 4.13: Primary vertices distributions in data and MC samples, after reweighting.

and then extrapolated to the interaction vertex and to the calorimeter, and the η of the seed cluster or the φ of the supercluster; H/E is the hadronic leakage, *i.e.* the ratio of the hadronic energy of the calorimetric towers to the electromagnetic energy of the electron supercluster; $relIso$ indicates the relative isolation calculated with the effective area approach; $1/E - 1/p$ is the difference of the inverse of the energy and the momentum; d_0 and d_z are the transverse and longitudinal impact parameters. A dedicated conversion veto is applied to mitigate the effects of electrons undergoing bremsstrahlung in the silicon detector.

Table 4.5: Electron cut-based selection for 25 ns bunch spacing conditions. EB: barrel cuts ($|\eta_{\text{supercluster}}| \leq 1.479$); EE: endcap cuts ($|\eta_{\text{supercluster}}| > 1.479$)

Electrons	<i>Veto Id</i>	
	EB	EE
$\sigma_{i\eta i\eta}$	< 0.0115	0.037
$\Delta\eta_{in}^{seed}$	< 0.00749	0.00895
$\Delta\varphi_{in}$	< 0.228	0.213
H/E	< 0.356	0.211
$relIso$ (Effective Area)	< 0.175	0.159
$ 1/E - 1/p $	< 0.299	0.15
$ d_0 $	< 0.05	0.10
$ d_z $	< 0.10	0.20
missing hits	\leq 2	3
conversion veto	yes	yes

4.3.3 Photons

As in the case of electrons, a photon veto is applied in the analysis both for the signal and the control regions. Events are rejected if they contain one (or more) photon with $p_T > 15$ GeV, $|\eta| < 2.5$, passing the *loose* cut-based photon Id, whose definition is reported in tab. 4.6. The isolation cuts (using the ρ -area method for the mitigation of the pile-up) and conversion-safe veto are applied. The isolation value is computed in a ΔR cone of 0.3 and it is corrected for pile-up by subtracting the event-by-event energy density (ρ) times the photon energy deposits effective area.

4.3 Event selection

Table 4.6: Photon cut-based selection for 25 ns bunch spacing conditions. EB: barrel cuts ($|\eta_{\text{supercluster}}| \leq 1.479$); EE: endcap cuts ($|\eta_{\text{supercluster}}| > 1.479$)

Photons	<i>Loose Id</i>	
	EB	EE
H/E	< 0.0597	0.0481
$\sigma_{i\eta i\eta}$	< 0.01031	0.03013
PF ch.had.iso.(ρ -corr)	< 1.295	1.011
PF neu.had.iso.(ρ -corr)	< $10.910 + 0.0148 p_T + 0.000017 p_T^2$	$5.931 + 0.0163 p_T + 0.000014 p_T^2$
PF photon iso.(ρ -corr)	< $3.630 + 0.0047 p_T$	$6.641 + 0.0034 p_T$
conversion veto	yes	yes

267 4.3.4 Muons

268 The minimal criteria to define a muon is that it must be identified by the Particle-Flow algorithm,
269 and should be reconstructed either as a global muon or as a tracker muon (sec. ??). The muon
270 isolation is defined in a cone with a radius of $\Delta R = 0.4$ centered along the lepton direction. In the
271 analysis event selection, all muons identified with the loosest criteria previously described, p_T over
272 10 GeV, PF isolation below 0.25, $\eta < |2.4|$ are vetoed.

273 4.3.5 Taus

274 The presence of hadronically decaying taus acts as a veto for the events both in the signal and in
275 the control regions, in order to suppress electroweak backgrounds. The selection criteria for taus
276 are $p_T > 18$ GeV and $|\eta| < 2.3$. Loose identification criteria of the hadronic tau reconstruction algo-
277 rithms are required and applied in order to identify possible tau candidates.

278 4.3.6 Jets

279 In this analysis, jets are considered if the corrected p_T is larger than 30 GeV for AK4 jets, and larger
280 than 200 GeV for AK8 jets, and lie in the tracker acceptance ($|\eta| < 2.4$). The requirement on AK8
281 jets transverse momentum is motivated by the fact that $p_T = 200$ GeV is the minimum kinematical
282 threshold ensuring to enclose the lighter hadronically decaying vector boson (namely, the W bo-
283 son) in the jet cone. Additionally, AK4 jets are required to pass *loose* jet identification requirements,
284 AK8 are required to pass *tight* jet identification requirements defined in tab. 4.7. AK8 jets are used
285 to reconstruct the hadronically decaying electroweak boson candidate, whilst AK4 jets are used to
286 suppress the contribution of top and QCD background events. Jet energy corrections are applied
287 to AK4 and AK8 CHS jets. Fig. 4.14- 4.16 show the data/simulation comparison after the analysis
288 selections (tab. 4.12 without Top cleaning and Event cleaning).

289 Since it has been measured that the jet energy resolution (JER) is not the same in data and MC,
290 an additional smearing is applied in simulation, in order to get a better agreement. There are two
291 independent ways to get the smearing. The scaling method rescales the corrected four-momentum
292 of a reconstructed jet by a factor

$$c_{\text{JER}} = 1 + (s_{\text{JER}} - 1) \frac{p_T - p_T^{\text{gen}}}{p_T}, \quad (4.3)$$

293 where p_T is the transverse momentum of the jet, p_T^{gen} is the transverse momentum of the genera-
294 tor level particle corresponding to the reconstructed jet, and s_{JER} is the data-simulation resolution

Table 4.7: *Loose* and *Tight* jet identification requirements for 25 ns bunch spacing conditions.

Particle-Flow jet ID	<i>Loose</i>	<i>Tight</i>
Neutral Hadron Fraction	< 0.99	< 0.90
Neutral EM Fraction	< 0.99	< 0.90
Number of Constituents	> 1	> 1
Muon Fraction	-	-
Additionally, for $ \eta < 2.4$		
Charged Hadron Fraction	> 0	> 0
Charged Multiplicity	> 0	> 0
Charged EM Fraction	< 0.99	< 0.99

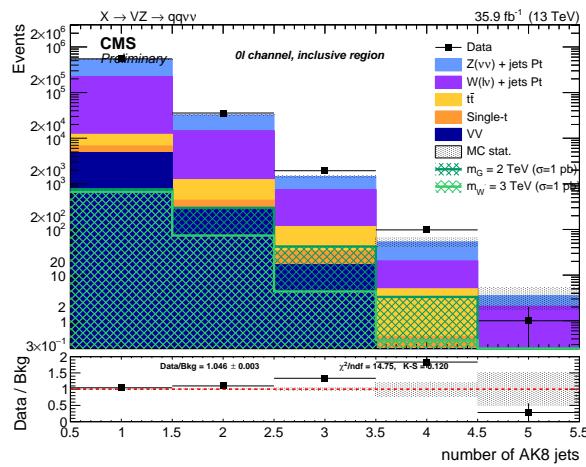
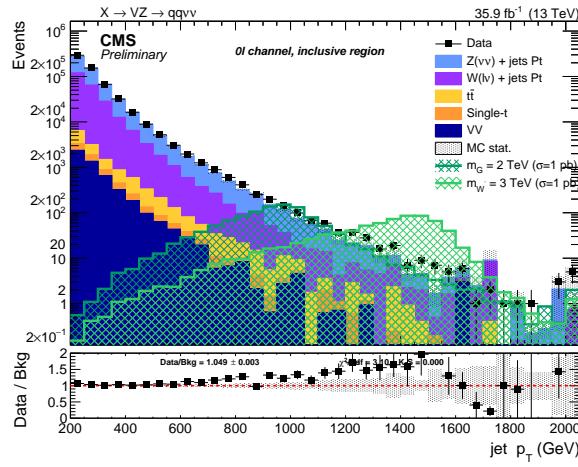


Figure 4.14: Number of reconstructed AK8 jets after inclusive selections.


 Figure 4.15: Leading AK8 jet p_T spectrum after inclusive selections.

295 scale factor. The factor c_{JER} is positively defined, hence, when negative, it is set equal to zero. The

4.3 Event selection

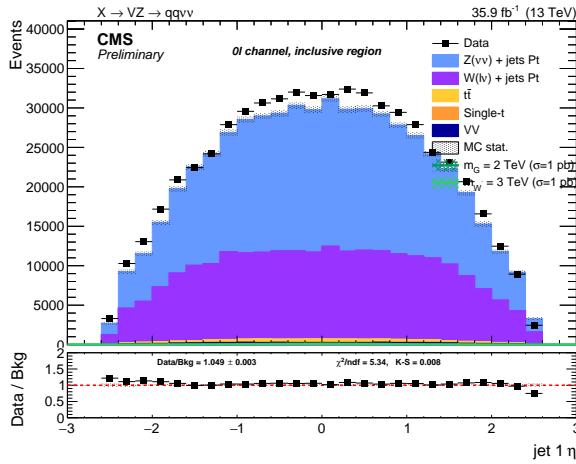


Figure 4.16: Leading AK8 jet η spectra after inclusive selections.

296 generator level particle and a reconstructed jet are defined as matched if:

$$\Delta R < R_0/2, \quad (4.4)$$

$$|p_T - p_T^{\text{gen}}| < 3 \times \sigma_{\text{JER}} \times p_T,$$

297 where R_0 is the jet clustering parameter and σ_{JER} is the relative p_T resolution measured in simula-
298 tion.

299 The alternative approach is the stochastic smearing, and it does not require the matching with the
300 generator level particle. The jet four-momentum is rescaled by a factor

$$c_{\text{JER}} = 1 + \mathcal{N}(0, \sigma_{\text{JER}}) \sqrt{\max(s_{\text{JER}}^2 - 1, 0)}, \quad (4.5)$$

301 where σ_{JER} is the relative p_T resolution in simulation, s_{JER} is the data-simulation scale factor, and
302 $\mathcal{N}(0, \sigma)$ is a random number extracted from a gaussian normal distribution, whose mean is zero
303 and variance σ^2 . Scaling factor c_{JER} is positively defined.

304 The smearing procedure adopted in this analysis is the hybrid method: when a matching jet at
305 generator level is found, the scaling method is adopted, else the stochastic smearing is chosen. The
306 smearing coefficients (scale factors, SF) as a function of the jet η and their uncertainties are reported
307 in tab. 4.8 for 2016 data [11].

308 4.3.7 Jet mass

309 The jet mass is the main observable in distinguishing a V jet from a jet produced by colour inter-
310 action (QCD jets). Jet grooming procedure consists in the suppression of uncorrelated underlying
311 event, pile-up and soft radiation from the jet: it improves the signal and background discrimina-
312 tion, by pushing the jet mass for QCD jets towards lower values of the spectrum, while maintaining
313 the jet mass for V -jets around the electroweak boson mass window.
314 The grooming technique of the analysis relies on the “soft drop declustering” algorithm, a jet sub-
315 structure technique that recursively removes soft wide-angle radiation from a jet [12], in order to
316 mitigate the contaminations from initial state radiation, along with pile-up and multiple scatter-
317 ings.
318 The soft drop algorithm starts with a jet clustered with the anti- k_T algorithm with a parameter R_0 ;
319 the jet is then reclustered with the Cambridge-Aachen method [13], whose definition is included in

Table 4.8: Data-simulation jet smearing coefficients and their corresponding uncertainties.

Jet η	Smearing SF
0.0–0.5	1.109 ± 0.008
0.5–0.8	1.138 ± 0.013
0.8–1.1	1.114 ± 0.013
1.1–1.3	1.123 ± 0.024
1.3–1.7	1.084 ± 0.011
1.7–1.9	1.084 ± 0.011
1.9–2.1	1.140 ± 0.047
2.1–2.3	1.067 ± 0.053
2.3–2.5	1.177 ± 0.041
2.5–2.8	1.364 ± 0.039
2.8–3.0	1.857 ± 0.071
3.0–3.2	1.328 ± 0.022
3.2–5.0	1.16 ± 0.029

320 eq. ??, once the exponent α is set $\alpha = 0$. The soft drop algorithm is ruled by two parameters, a soft
 321 threshold z_{cut} , that cuts on the energy fraction of soft radiation, and an angular exponent β . The
 322 procedure is the following:

- 323 • the jet is declustered into two subjets, j_1 and j_2 , by reverting the final step of Cambridge-
 324 Aachen algorithm;
- 325 • if j_1 and j_2 respect the soft drop condition (eq. 4.6), j is defined as the groomed jet;
- 326 • if they don't pass the condition, the leading subjet in p_T is redefined as the new j ;
- 327 • if j can't be declustered anymore, it is defined as the groomed jet.

328 The parameters $z_{\text{cut}} = 0.1$ and $\beta = 0$ are set in the soft drop condition:

$$\frac{\min(p_T^1, p_T^2)}{p_T^1 + p_T^2} > z_{\text{cut}} \left(\frac{\Delta R_{12}}{R_0} \right)^\beta, \quad (4.6)$$

329 where p_T^1 and p_T^2 are the momenta of the constituents, ΔR_{12} is their angular distance. z_{cut} and β
 330 parameters affect the degree of jet grooming: if $\beta \rightarrow \infty$ the jet remains ungroomed, while the more
 331 β approaches zero, the more soft collinear radiation is removed.

332 The net effect of the soft drop algorithm is studied in Monte Carlo simulations of a W hadronic
 333 decay process (signal), in association with jets, and of a multi-jet QCD process (background). Jets
 334 are clustered with the anti- k_T algorithm with a parameter $R_0 = 1$ and asked to have $p_T > 500$ GeV
 335 and $|\mathcal{Y}| < 4$. The parameter z_{cut} is chosen such in a way that the number of events falling in the W
 336 mass window ($[70, 90]$ GeV) is the 35% of the total number of events. The results before (black curve)
 337 and after the application of soft drop algorithm (coloured curves, depending on the value of β) are
 338 presented in fig. 4.17 [12]. In particular, by comparing the ungroomed jet mass (in black) with the
 339 mass groomed with a parameter $\beta = 0$ (adopted in this analysis and displayed with a green curve),
 340 the soft drop mass of the leading jet is a very narrow distribution peaking around the nominal W
 341 window in the signal sample, whilst it is pushed at lower values in the background sample.

342 The soft drop algorithm is used in association with the Pile Up Per Particle Identification algorithm
 343 (PUPPI) [14], designed to combine detector informations in order to compute a local metric α , that

4.3 Event selection

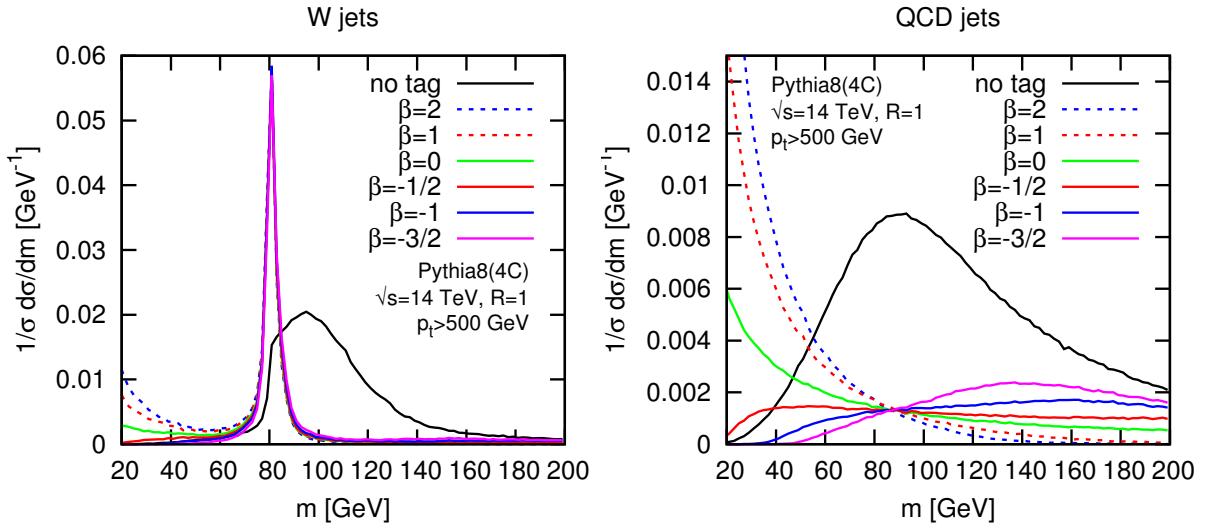


Figure 4.17: Distributions of the jet mass in $W +$ jeta signal simulations (left) and multi-jet QCD background (right), before (in black) and after applying soft drop algorithm. Each curve corresponds to a different value of the parameter β . [12]

describes with a weight how likely it is that one particle is coming from the primary vertex or from a pile-up event. A fundamental feature exploited by the algorithm is the p_T spectrum of the primary vertex particles, expected to be harder than that of the pile-up ones.

The local shape α is defined as:

$$\alpha_i = \log \sum_{j \in \text{event}} \frac{p_{T,j}}{\Delta R_{ij}} \Theta(R_{\min} \leq \Delta R_{ij} \leq R_0), \quad (4.7)$$

where Θ is the Heaviside step function, ΔR_{ij} is the angular distance between the considered i particle and the neighbour j particle, laying in a cone $R_0 = 0.4$ centered around i direction, within a minimum distance $R_{\min} = 0.0001$. Given the softer p_T spectra of pile-up particles, α_i is smaller when i particle does not originate from the primary vertex.

The function

$$\chi^2_i = \Theta(\alpha_i - \bar{\alpha}_{PU}) \frac{(\alpha_i - \bar{\alpha}_{PU})^2}{\sigma_{PU}^2} \quad (4.8)$$

estimates how much α_i fluctuates from the median of the pile-up local shape $\bar{\alpha}_{PU}$ (that has a variance σ_{PU}^2), and it is distributed like a χ^2 with 1 degree of freedom. The PUPPI weight is defined as the cumulative χ^2 distribution $F_{\chi^2, 1 \text{ d.o.f.}}$,

$$w_i = F_{\chi^2, 1 \text{ d.o.f.}}(\chi^2_i). \quad (4.9)$$

If the local metric of a particle is distributed closely to the expected distribution of the pile-up, its weight is $w = 0$. Large fluctuations are more likely related to non pile-up particles, and they receive a weight close to 1. All the particles whose weights are smaller than 0.01 are removed from the jet clustering procedure.

The default soft drop PUPPI jet mass suffers from a systematic shift from the expected value of about $\sim 10\%$, and from some residual dependence on the jet p_T . Further corrections to the jet mass have been applied:

363 1. a p_T -dependent correction to account for a small shift in the generated vector boson mass,
 364 applied only on simulated samples,

365 2. a p_T - η -dependent correction to the reconstructed jet mass, applied separately for jets in the
 366 barrel and endcaps regions.

367 Fig. 4.18- 4.19 show the jet mass for hadronically decaying W or Z bosons in bulk graviton and
 368 W' signal samples, before and after the correction, without applying any selections. In fig. 4.20,
 369 the distribution of soft drop PUPPI jet mass is shown for the expected backgrounds of the analysis
 370 (coloured histograms) and data (black markers), before and after the corrections.

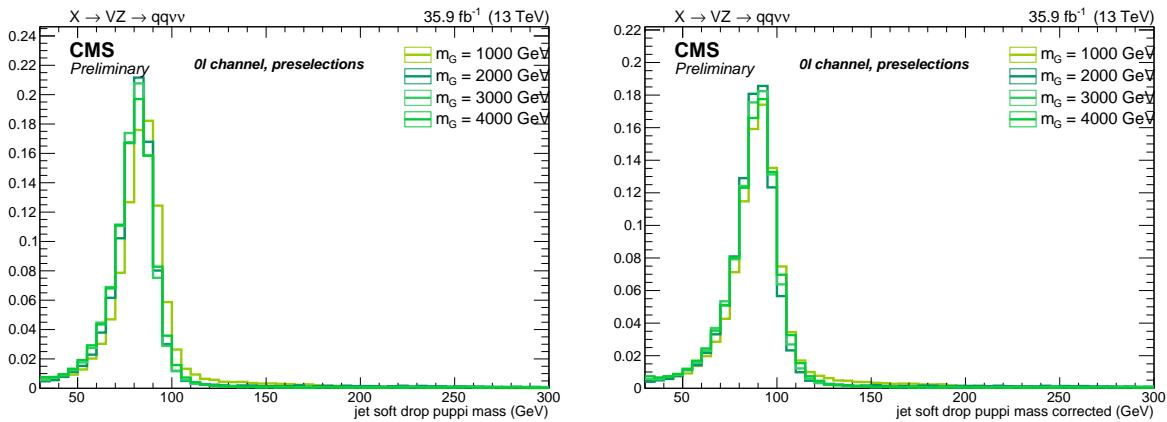


Figure 4.18: Soft drop PUPPI mass of AK8 jet reconstructed for different bulk graviton signal samples, before corrections (left) and after corrections (right).

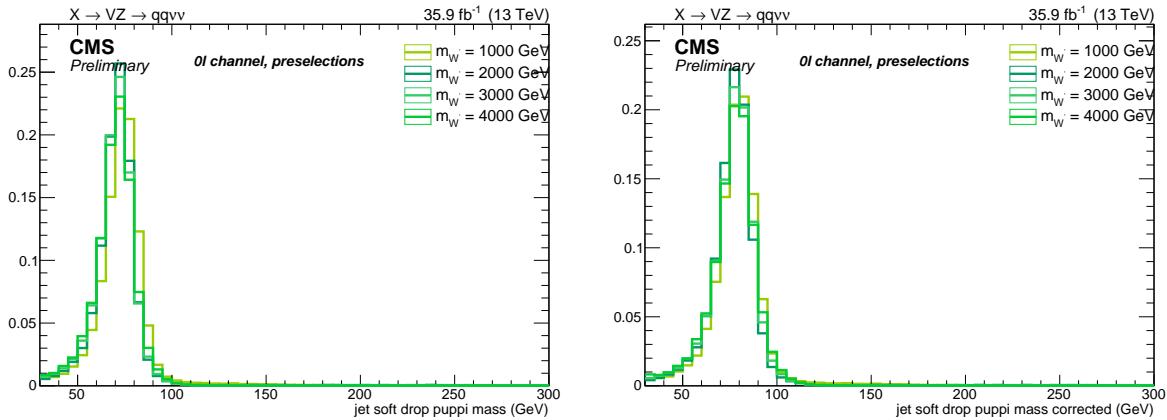


Figure 4.19: Soft drop PUPPI mass of AK8 jet reconstructed for different W' signal samples, before corrections (left) and after corrections (right).

371 In order to obtain a better data-Monte Carlo agreement, a smearing procedure has been applied to
 372 the soft drop PUPPI jet mass, by using the stochastic method, with a constant smearing coefficient
 373 (1.00 ± 0.20), that does not depend on jet pseudorapidity, if it is restricted to $|\eta| < 2.5$.

374 The selection applied on the jet mass is a crucial step of the analysis, and it has to fulfill three pur-

4.3 Event selection

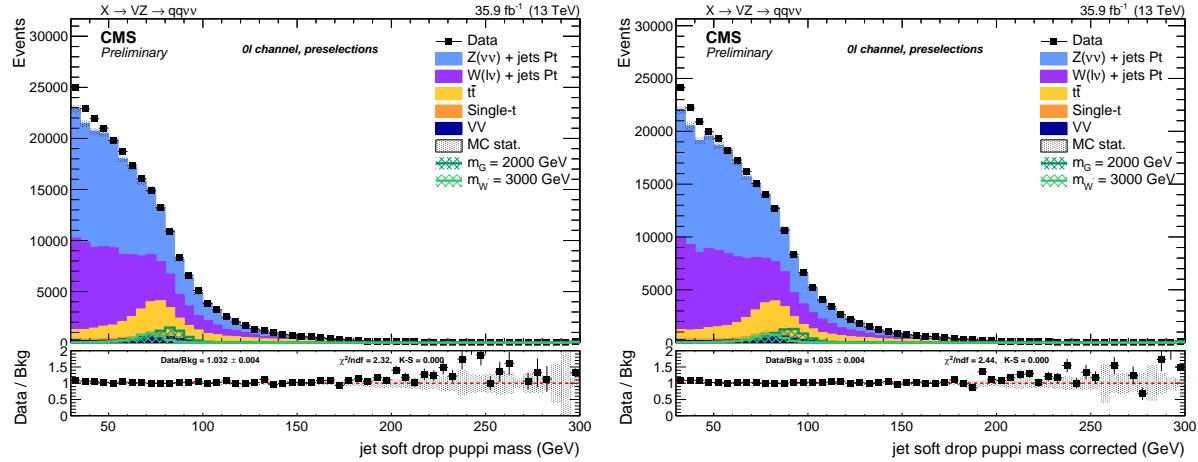


Figure 4.20: Soft drop PUPPI mass of AK8 jet; left: before corrections. right: after corrections.

poses: it has to provide the maximum signal significance (best compromise between signal efficiency and background reduction), it has to avoid overlaps with the Higgs boson mass window, and it has to provide a sufficient data and simulation statistics for the control regions (the regions outside the mass cut). The soft drop PUPPI mass variable is used to define the following regions:

Table 4.9: Mass regions defined for the analysis.

	low-sideband	V-region	H-region	high-sideband
M_J	30-65 GeV	65-105 GeV	105-135 GeV	> 135 GeV

The "signal region" (SR) refers to the V-region, where the largest signal yield is expected. The "sidebands" (SB) refer to the low-sideband and high-sideband, where a negligible amount of signal is expected. Events with a jet mass value lower than 30 GeV are discarded, because of the high background contamination. The jet mass distribution of the V candidate, in the sidebands and in the signal region, is shown in fig. 4.21. If the soft drop, PUPPI corrected mass of a large-con jet falls into the V-region, the jet is defined as V-tagged.

4.3.8 Jet substructure

In order to further discriminate signal from background, the inner structure of the jet is investigated. Studying the distribution of the jet constituents with respect to the jet axis allows to test the hypothesis of the existence of multiple substructures, that could be an evidence of jets originated by more than one parton. The constituents of the considered jet are clustered again with the k_T algorithm, and it is forced to return n subjets. The n -subjettiness [15], τ_n , is defined as

$$\tau_n = \frac{1}{d_0} \sum_k p_{T,k} \min(\Delta R_{1,k}^\beta, \Delta R_{2,k}^\beta, \dots, \Delta R_{n,k}^\beta), \quad (4.10)$$

where k labels the particles included in the jet, $p_{T,k}$ is the corresponding transverse momentum of the k constituent, and $\Delta R_{i,k}$ is the solid angle between the k constituent and the i subjet candidate.

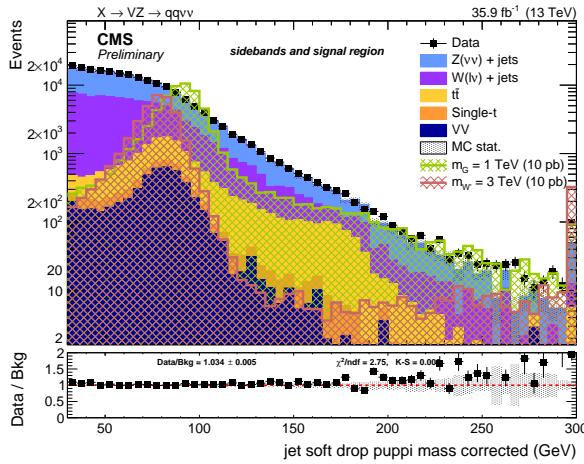


Figure 4.21: Distribution of the soft drop PUPPI corrected mass of the leading AK8 jet, selected as the hadronically decaying V candidate, in the sidebands and control region of the analysis, for expected SM background, bulk graviton signal, W' signal, and data.

393 The parameter d_0 is a normalization factor:

$$d_0 = \sum_k p_{T,k} R_0, \quad (4.11)$$

394 where R_0 is the clustering parameter of the considered jet. The τ_n variable describes to what degree
395 a jet can be considered as composed by n substructures; smaller values of τ_n correspond to higher
396 compatibility to the n -prong hypothesis. A large-cone jet generated by the hadronic decay of an
397 electroweak boson is expected to be a 2-prong object, whilst light flavour and gluon jets generated
398 by colour interaction have a 1-prong monolithic structure. The τ_2 or the τ_1 alone, by the way, do
399 not provide an optimal signal and background discrimination, as shown in fig. 4.22 (left and center);
400 by looking at fig. 4.22 (right), it is clear that the most powerful discriminating variable is their ratio
401 $\tau_{21} = \tau_2/\tau_1$:

$$\tau_{21} = \frac{\frac{1}{d_0} \sum_k p_{T,k} \min(\Delta R_{1,k}, \Delta R_{2,k})}{\frac{1}{d_0} \sum_k p_{T,k} \Delta R_{1,k}}. \quad (4.12)$$

402 In fig. 4.23, the distributions of the τ_{21} variable are displayed for background and data, after applying
403 the PUPPI algorithm (left), and for different bulk graviton mass hypotheses (right). The signal
404 distribution is expected to peak at low values of the τ_{21} subjettness variable.

405 The τ_{21} variable is used to classify the events into two exclusive categories, in order to improve
406 the signal discovery reach. Events are included in either the high-purity ($\tau_{21} < 0.35$) or low-purity
407 ($0.35 < \tau_{21} < 0.75$) category.

408 The choice of the τ_{21} categorization listed above is based on a study of the analysis sensitivity. An-
409 other τ_{21} categorization is probed, according to which events are grouped into different high-purity
410 ($\tau_{21} < 0.40$) and low-purity ($0.40 < \tau_{21} < 0.75$) categories. This different set of τ_{21} cuts has been
411 tested, along with that chosen for this analysis. Two figures of merit are considered: the discovery
412 reach, namely the bulk graviton signal significance (displayed in fig. 4.24), and the expected exclu-
413 sion limit on cross-section times branching fraction at 95% CL (displayed in fig. 4.25), as a function
414 of the reconstructed transverse mass of the resonance. To this purpose, the entire analysis workflow

4.3 Event selection

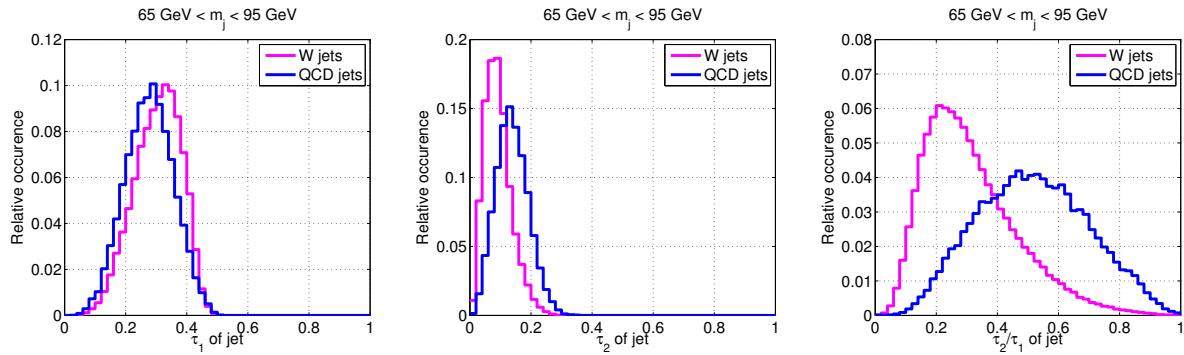


Figure 4.22: Distribution of τ_1 (left), τ_2 (center), and τ_{21} (right) variables, in simulations of a W plus jets process (in pink) and for a multi-jet QCD originated process (in blue). A selection on the leading jet mass is applied: $65 < m_j < 95 \text{ GeV}$; jets are clustered with a parameter $R_0 = 0.6$, $p_T > 300 \text{ GeV}$, $|\eta| < 1.3$ [15].

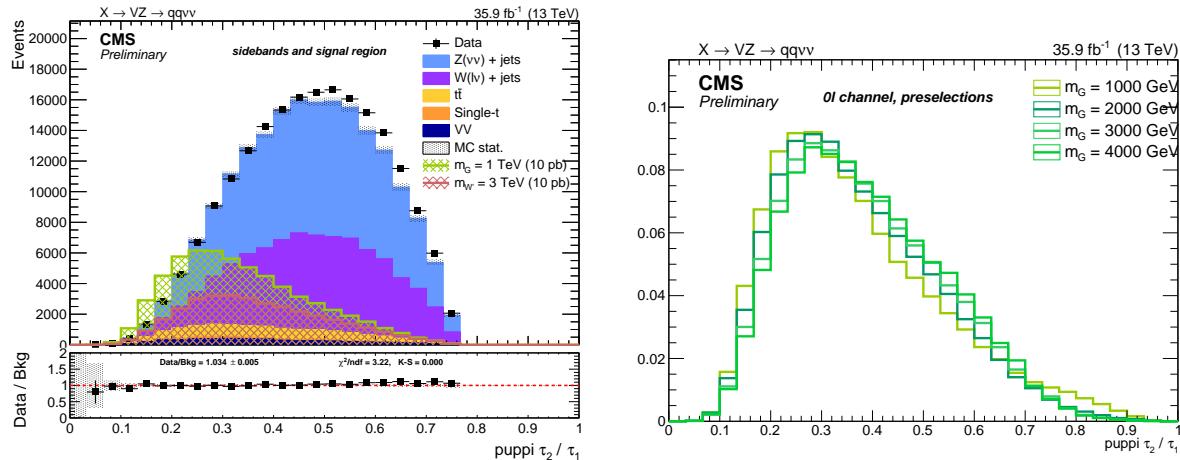


Figure 4.23: Distribution of the τ_{21} subjetteness of the leading AK8 jet, selected as the hadronically decaying V candidate, for expected SM background and data (left), and for bulk graviton signal (right).

415 has been applied, performing an unbinned shape analysis with the analysis background estimation
 416 method, taking into account all the systematic uncertainties. In each figure, on the left, the figure of
 417 merit is plotted separately for each purity category, while in the right part of the figures the low and
 418 purity categories are combined together. Significance has been computed with a limited number of
 419 toys (100), hence the curves are non perfectly smooth, while the exclusion limit has been computed
 420 with the asymptotic formula. The procedures to extract signal significance and exclusion limits are
 421 described in sec. ???. Considering that the search region is 1-4 TeV, the choice of 0.35-0.75 τ_{21} working
 422 points is legitimated.

423 When doing the τ_{21} categorization, V -tagging scale factors have been taken into account to cor-
 424 rect data and simulation discrepancies introduced by the n -subjettiness. They are described in
 425 sec. 4.3.8.1.

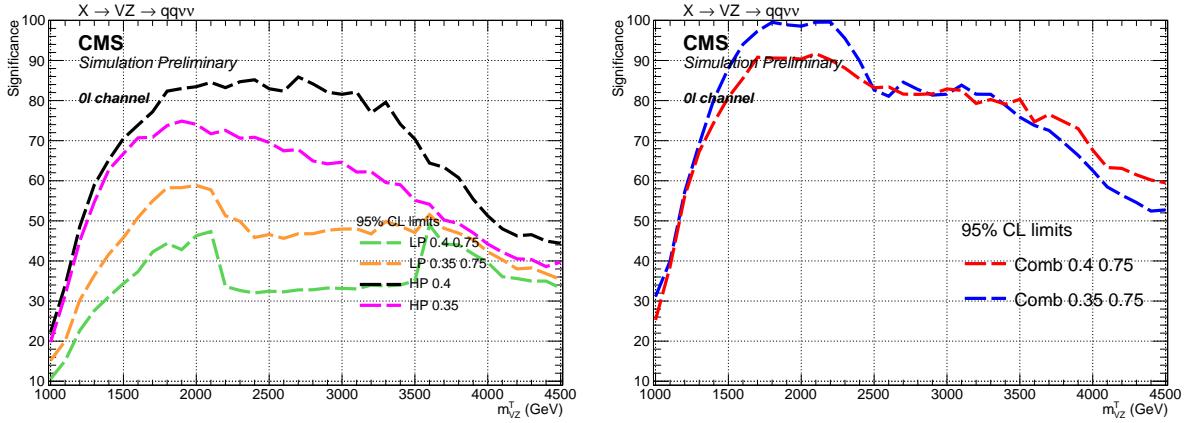


Figure 4.24: Analysis sensitivity to bulk graviton signals, computed by applying different τ_{21} categorizations, considering the categories separately (left) and combining them together (right), as a function of the resonance mass.

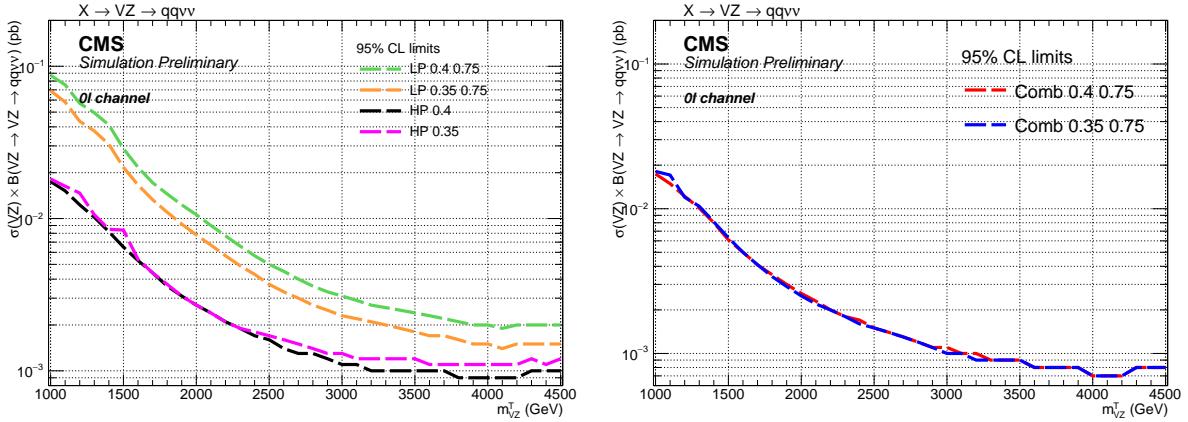


Figure 4.25: Exclusion limit on cross-section time branching fraction at 95% CL of bulk graviton signals, computed by applying different τ_{21} categorizations, considering the categories separately (left) and combining them together (right), as a function of the resonance mass.

4.3.8.1 Corrections induced by jet substructure variables

By applying a selection on the jet τ_{21} , the jet mass spectrum is sculpted, hence the effects of the V -tagging procedure shall take into account both the selections on mass and on substructure simultaneously. The distributions of the groomed jet mass and τ_{21} subjetiness have been compared in data and simulations, by selecting samples of di-jet, $t\bar{t}$ and $W +$ jets events, and a significative discrepancy has been observed (10%) [5]. Scale factors are extracted by selecting a $t\bar{t}$ sample in data, because an high p_T W boson is produced by the top quark decay. The hadronically decaying W boson is tagged by choosing events where the soft drop mass of a large-cone jet lies in a window centered around the nominal W mass. The jet mass distributions of events passing and failing the selection on the τ_{21} variable ($\tau_{21} < 0.35$ and $0.35 < \tau_{21} < 0.75$, considered separately) are fitted simultaneously, both in data and in simulations. The V -tagging scale factors are defined as the ratio of the τ_{21} categorization efficiencies in data and MC, and they are summarized in tab. 4.10.

4.3 Event selection

438 The systematic uncertainties depend on the simulation of the $t\bar{t}$ process, they cover the discrepancies
439 observed while using different Monte Carlo simulations, and due to the choice of the fitting
440 function.

Table 4.10: Data-simulation scale factors, calculated on $t\bar{t}$ samples, that correct the discrepancies related to the τ_{21} categorization.

τ_{21} selection	Purity category	Data-MC scale factor
$\tau_{21} < 0.35$	high-purity	0.99 ± 0.11
$0.35 < \tau_{21} < 0.75$	low-purity	1.03 ± 0.23

441 **4.3.9 b-tagging**

442 The presence of a b-tagged quark can be an hint to identify the top quark decays, representing a potential background to the search. The CSV b-tagging algorithm [16] is applied to the AK4 jets.
443 The jet is considered as tagged if the CSV discriminator value is above a threshold value; the b-tag efficiency is defined as the number of jets fulfilling this requirement, divided by the total number of jets. Since the purpose of the b-tagging is to reject the top quark events, the working point with the largest efficiency is chosen; the threshold of the CSV multivariate discriminant is listed in tab. 4.3.9.
446

Table 4.11: Working point for CSV b-tagging algorithm.

Working point	CSV discriminant threshold	tagging efficiency	mis-tag probability
CSVL (Loose)	> 0.5426	~ 85%	~ 10%

448 Events where an AK4 jet, not laying in the AK8 jet cone, is b-tagged with the *loose* working point
449 threshold, are rejected. This veto allows to suppress the single-top events and $t\bar{t}$ events by one half.
450 The b-tagging efficiency is not the same in data and MC. In order to take into account this difference,
451 b-tagging scale factors for b-jets and mis-tagged light jets, measured for different physics processes,
452 are calculated. A weight is extracted on a per-event basis, as a function of the b-tagging status of the
453 jets and their kinematic variables [17].

454 **4.3.10 Missing Energy**

455 As pointed out in sec. ??, Type-I corrected E_T^{miss} is used in the analysis, along with dedicated filters
456 to remove detector noise and events with bad reconstruction. In order to lie in the plateau of the
457 trigger efficiency, $E_T^{\text{miss}} > 200\text{GeV}$. Fig. 4.26 shows the E_T^{miss} distribution for data and Monte Carlo
458 after the corrections and filters.

459 **4.3.11 Diboson candidate reconstruction**

460 **4.3.11.1 $V \rightarrow q\bar{q}$ reconstruction**

461 The identification of jets produced by the hadronic decays of one vector boson is based on the two
462 concepts:

- 463 • *Jet mass*: jets produced by the decay of a massive particle should have an invariant mass
464 around the nominal mass of the original particle. Oppositely, jets originated by QCD radiation
465 are produced by the emission of quarks or gluons and typically have smaller invariant mass.
466 This effect is further enhanced by the grooming techniques (sec. 4.3.7).

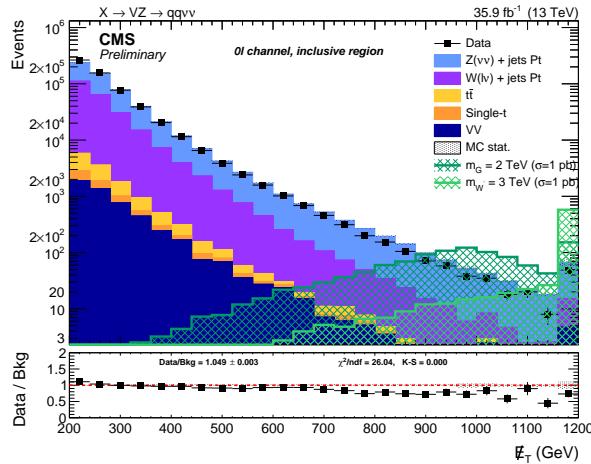


Figure 4.26: Type-1 corrected E_T^{miss} distribution after inclusive selections.

- *Jet substructure:* looking inside the structure of jets gives an handle in discriminating the original seed of the jet. Z and W -jets are produced by two partons merged into a single large-cone jet.

The leading AK8 jet respecting the jet mass and jet substructure selections is tagged as the V candidate.

4.3.11.2 $Z \rightarrow \nu\bar{\nu}$ reconstruction

If the Z boson decays into a pair of neutrinos, no product is visible in the detector, hence the invisible decay of the Z boson is determined only by its transverse component, namely by the E_T^{miss} .

4.3.11.3 Composite VZ candidate reconstruction

Given that the longitudinal component of the Z boson is unknown, a simple and effective solution is to consider the transverse mass of the VZ candidate, using the jet and \vec{p}_T^{miss} kinematics, defined by the following formula:

$$m_{VZ}^T = \sqrt{2E_T^V E_T^{\text{miss}} \cdot (1 - \cos \Delta\varphi(V, \vec{p}_T^{\text{miss}}))}, \quad (4.13)$$

where E_T^V is the transverse energy of the V candidate (defined in sec. ??), and $\Delta\varphi$ is the angle between the V and the Z candidates in the transverse plane.

4.3.12 Final analysis selections

Events considered in this analysis have to pass a certain number of selections before being considered as suitable signal candidates, both in data and in simulations. The selections are reported below and in tab. 4.12. The selections applied to group the events in purity category, defined on the PUPPI corrected τ_{21} subjetiness variable (sec. ??), and into signal or control region, defined on the PUPPI corrected soft drop mass (sec. 4.3.7) are reported in tab. 4.13. The final signal efficiency is shown separately in purity categories in fig. 4.3.12.3, for both spin-2 and spin-1 signal hypotheses.

4.3 Event selection

4.3.12.1 Z candidate selections

- *Trigger:* HLT_PFMETNoMu90_PFMHTNoMu90_IDTight or HLT_PFMETNoMu110_PFMHTNoMu110_IDTight or HLT_PFMETNoMu120_PFMHTNoMu120_IDTight or HLT_PFMET170_NoiseCleaned or HLT_PFMET170_JetIdCleaned or HLT_PFMET170_HBHECleaned (required in data only);
- $E_T^{miss} > 200 \text{ GeV};$
- *Corrections:* Type-I, noise filters.

4.3.12.2 V candidate selections

- p_T : at least one AK8 Particle-Flow jet with $p_T > 200 \text{ GeV}$;
- η : $|\eta| < 2.4$;
- *Identification:* tight Particle-Flow Id;
- *charged hadron fraction:* chf > 0.2 ;
- *neutral hadron fraction:* nhf < 0.9 ;
- *Mass:* soft drop PUPPI corrected mass $> 30 \text{ GeV}$;
- *Substructure:* PUPPI corrected τ_{21} subjetttines, depending on the category $\tau_{21} < 0.35$ for high-purity, $0.35 < \tau_{21} < 0.75$ for low-purity.

4.3.12.3 Topology and event cleaning

504 Minimal requirements are applied to objects that are vetoed:

- *Veto on electrons:*
 - p_T : $p_T > 10 \text{ GeV}$;
 - η : $|\eta| < 2.5$;
 - *Id: veto* cut-based working point;
- *Veto on muons:*
 - p_T : $p_T > 10 \text{ GeV}$;
 - η : $|\eta| < 2.4$;
 - *Id: loose Id*;
 - *Isolation:* Particle-Flow Isolation < 0.25 ;
- *Veto on hadronic taus:*
 - p_T : $p_T > 18 \text{ GeV}$;
 - η : $|\eta| < 2.4$;
 - *Id: loose Id*;
- *Veto on photons:*
 - p_T : $p_T > 15 \text{ GeV}$;

- 520 – $\eta: |\eta| < 2.5$;
- 521 – *Id: loose cut-based working point.*
- 522 Further selections are applied to suppress spurious events.
- 523 • *Event cleaning:* events where the V and the Z candidates are collinear are rejected:
 524 $\Delta\varphi(V, \vec{p}_T^{\text{miss}}) > 2$.
- 525 • *Top rejection:* as discussed in sec. 4.3.9, a b-tag veto is imposed on AK4 jets lying outside the
 526 AK8 cone; this reduces the top quark background contamination by 50%.
- 527 • *QCD rejection:* a minimum angular separation $\Delta\varphi > 0.5$ is imposed in the transverse plane be-
 528 tween the \vec{p}_T^{miss} vector and the momenta of all the AK4 jets in the event, lying outside the AK8
 529 cone and not tagged as b-quark initiated jets. The effect of this cut is to suppress the multi-jet
 530 QCD background: it has been studied by considering additional QCD simulated samples to
 531 the analysis backgrounds. As it can be inferred by looking at the distribution of the minimum
 532 azimuthal separation between \vec{p}_T^{miss} and the AK4 jets, shown in fig. 4.27 (where looser selec-
 533 tions are applied w.r.t. the nominal selections of the analysis, *i.e.*, no QCD event cleaning is
 534 performed), if a minimum $\Delta\varphi = 0.5$ threshold is imposed, the QCD contribution is reduced
 535 from 32% to 5%. In the final signal region, the QCD event yield amounts to 2%, and hence it
 536 is negligible (3% in low-purity, less than 1% in high-purity).

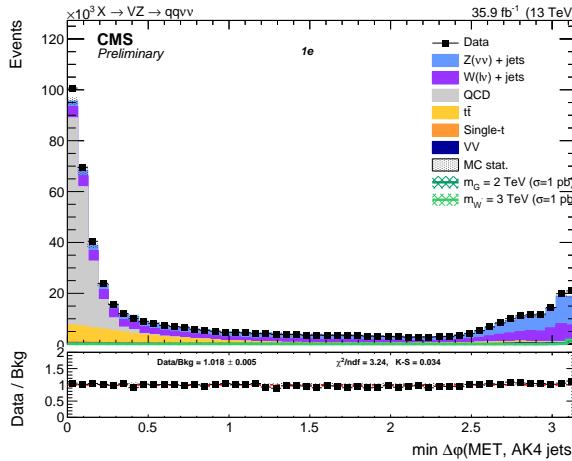


Figure 4.27: Distribution of the minimum azimuthal separation bewteen \vec{p}_T^{miss} and the momenta of all the AK4 jets present in each event. By imposing $\min\Delta\varphi > 0.5$, the QCD background (in gray) is suppressed.

- 537 The final selections of the analysis are summarized in tab. 4.12-4.13. The detection efficiencies due
 538 to each cut sequentially applied to bulk graviton signal samples (fig. 4.3.12.3, left) and W' signal
 539 samples (fig. 4.3.12.3, right) are shown. The signal efficiency for bulk graviton ranges from $\sim 30\%$
 540 at 1 TeV, up to 20% at 4.5 TeV for low-purity category, whilst it's around 20% for the high-purity
 541 category in the whole mass range. The signal efficiency for W' ranges from $\sim 40\%$ at 1 TeV, up to 25%
 542 at 4.5 TeV for low-purity category, whilst it's around 25% for the high-purity category in the whole
 543 mass range. The different detection efficiencies for the two signals are related to their production
 544 mechanisms: the graviton is produced in gluon fusion, hence more hadronic activity is expected

4.3 Event selection

Table 4.12: Summary of the selection cuts for the $VZ \rightarrow \nu\nu qq$ analysis.

	$Z \rightarrow \nu\nu$
Trigger	HLT_PFMETNoMu90_PFMHTNoMu90_IDTight or HLT_PFMETNoMu110_PFMHTNoMu110_IDTight or HLT_PFMETNoMu120_PFMHTNoMu120_IDTight or HLT_PFMET170_NoiseCleaned or HLT_PFMET170_JetIdCleaned or HLT_PFMET170_HBHECleaned
E_T^{miss}	Type-I corrected $> 200 \text{ GeV}$
Veto	e, μ, τ, γ
V	$p_T > 200 \text{ GeV, tight Id}$ $\text{nhf} < 0.8; \text{chf} > 0.2$
QCD cleaning	$\min\Delta\varphi(\text{AK4 jets}, \vec{p}_T^{\text{miss}}) > 0.5$
Top cleaning	veto on b-tagged AK4 jets outside the AK8 cone, <i>loose</i> working point (< 0.460)
Event cleaning	$\Delta\varphi(V, \vec{p}_T^{\text{miss}}) > 2$

Table 4.13: Cuts to categorize the $VZ \rightarrow \nu\nu qq$ analysis events into low- and high-purity categories, and into signal region and sidebands.

	$Z \rightarrow \nu\nu$
V mass	Signal Region: $65 < m_V < 105$ Side Bands: $30 < m_V < 65, m_V > 135 \text{ GeV}$
$V \tau_{21}$	$0.35 < \tau_{21} < 0.75$ for low-purity $\tau_{21} < 0.35$ for high-purity

545 around the VZ decay process, and this results as a loss of efficiency when the QCD rejection cut is
 546 applied.

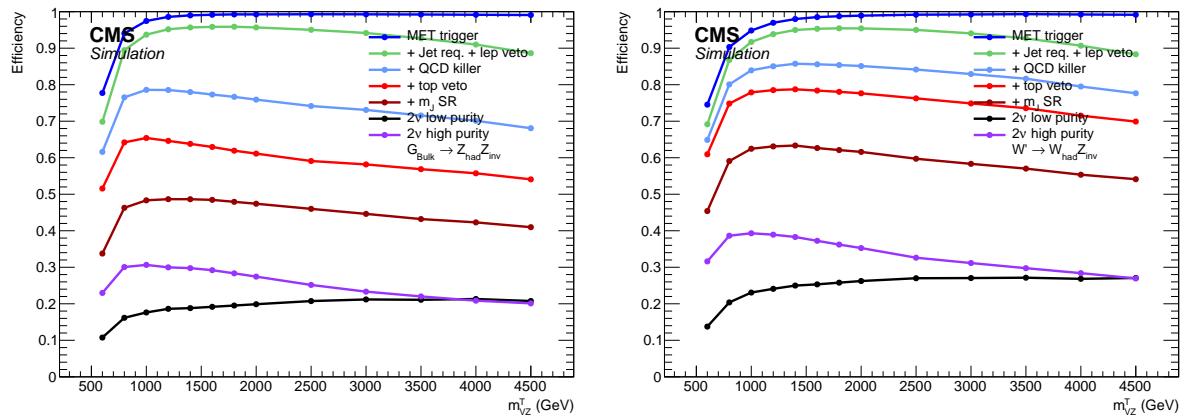


Figure 4.28: Signal efficiency for a spin-2 bulk graviton decaying into a pair of Z bosons (left), and for a spin-1 W' decaying into a W and a Z bosons (right), as a function of the mass of the heavy particle. The efficiencies are separated by purity category after the signal region selections.

547 4.4 Data and simulations comparison

548 In this section, a comparison between data and simulation is reported for various kinematic ob-
 549 servables. It can be seen that the dominant background contribution comes from the $Z + \text{jets}$ and
 550 $W + \text{jets}$ production, while sub-leading contributions from top ($t\bar{t}$ and single-top) production and
 551 dibosons can be minor yet non-negligible.

552 In the following plots (fig. 4.29-4.38), the comparison is performed in three different regions. On
 553 top of the selections defined above, additional criteria are defined:

- 554 • *Inclusive*: no selection is performed on top of the selections, except for a veto on the jet mass
 555 $65 < m_V < 135 \text{ GeV}$ to avoid potential signal contamination from VZ signals;
- 556 • *Sidebands (SB)*: only events in the sidebands, defined the interval between $30 < m_V < 65 \text{ GeV}$
 557 and $m_V > 135 \text{ GeV}$ are considered. This region can be considered as signal-depleted. The
 558 main difference with the previous regions is that the bulk of the jet mass distribution, peaking
 559 at $m_V \sim 20 \text{ GeV}$, is not included. The region selected is thus much closer kinematically to the
 560 signal region.
- 561 • *Signal region (SR)*: it represents the phase space where is signal is expected.

562 A summary of the number of expected events from Monte Carlo simulations, per each sample, along
 563 with the number of events observed in data in each category is reported in tab. 4.14. No significant
 564 excess is observed in data distributions with regards to simulation predictions in signal region.

Table 4.14: Expected background yields and number of events observed in data.

cut	inclusive	SB low-purity	SB high-purity	SR low-purity	SR high-purity
data	586318.00	107363.00	13967.00	44989.00	23074.00
$Z + \text{jets}$	320996.11	57551.99	7774.40	22933.14	10763.87
	57%	56%	56%	53%	45%
$W + \text{jets}$	224607.51	40447.51	5197.74	16248.78	7428.42
	40%	40%	37%	38%	31%
$t\bar{t}$	6308.09	2599.53	670.29	2482.38	3035.21
	1%	3%	5%	6%	13%
VV	5168.06	1075.75	206.54	1283.63	2053.19
	1%	1%	1%	3%	9%
single-top	1968.65	431.28	79.27	329.71	461.84
	<1%	<1%	1%	1%	2%
BkgSum	559048.42	102106.07	13928.25	43277.64	23742.54

4.4 Data and simulations comparison

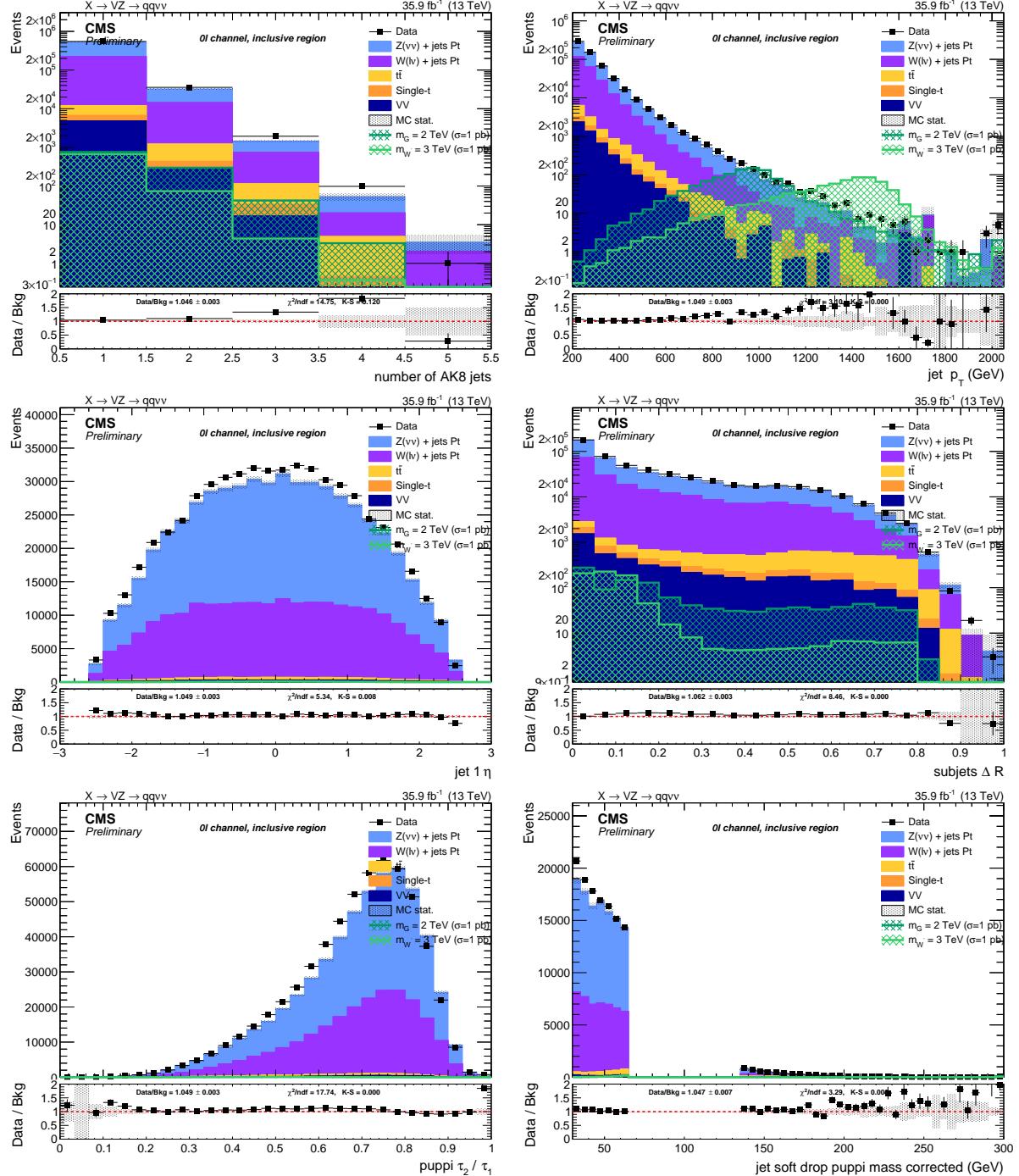


Figure 4.29: Top: number of AK8 jets in the event (left) and V jet candidate p_T (right). Center: V jet candidate η (left) and angular separation ΔR between the constituents leading subjets (right). Bottom: V jet candidate τ_{21} subjetiness after PUPPI correction (left) and V jet candidate soft drop PUPPI mass (right). Events are selected with the *inclusive* selection, and simulated backgrounds are normalized to luminosity.

Search for diboson resonances in the $VZ \rightarrow q\bar{q} \nu\bar{\nu}$ final state

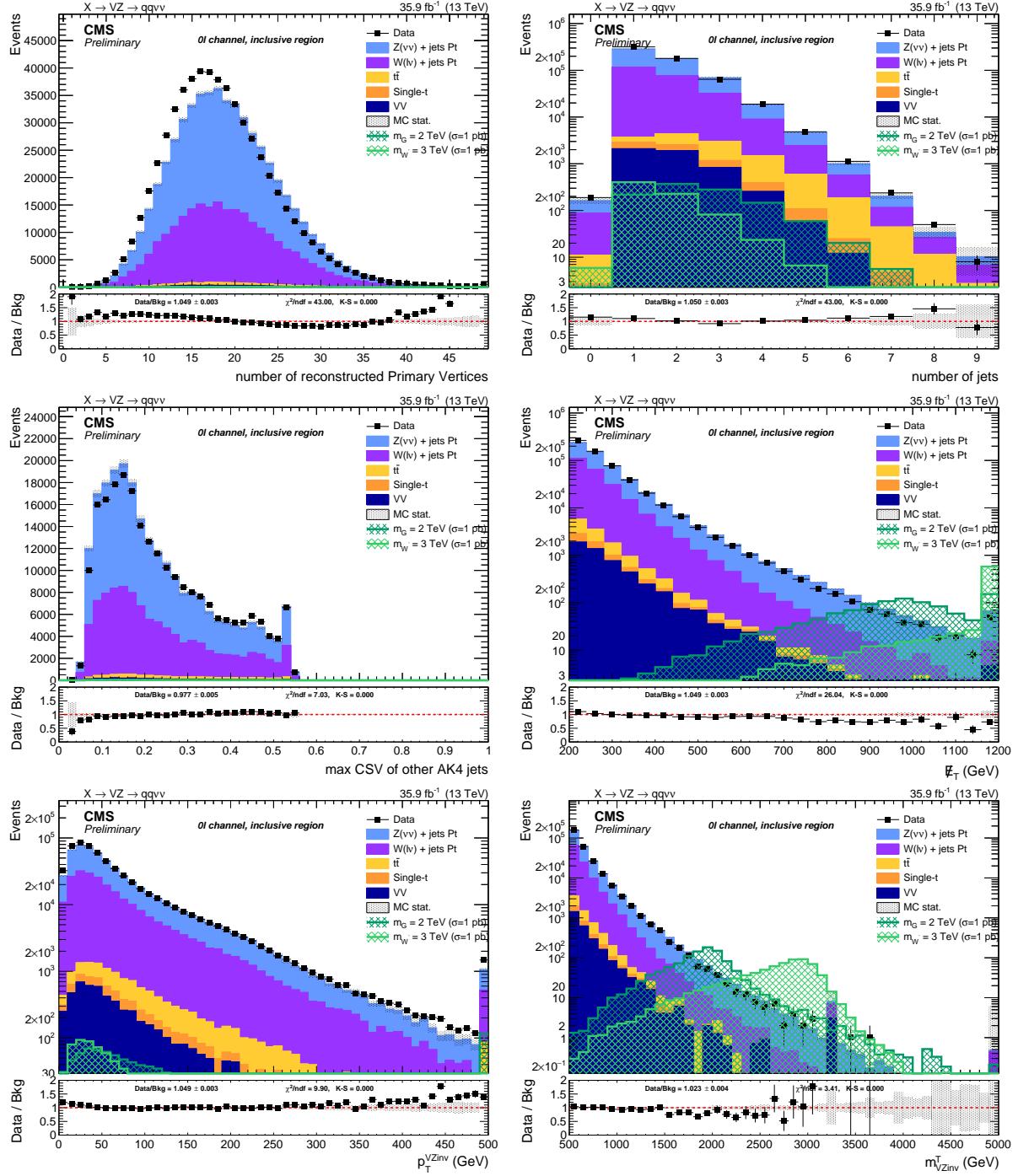


Figure 4.30: Top: number of reconstructed primary vertices (left) and number of AK4 jets in the event (right). Center: distribution of the b-tagging multivariate discriminant for the AK4 jets not included in the V jet cone (left) and E_T^{miss} distribution (right). Bottom: p_T of the VZ candidate (left) and transverse mass of the VZ candidate (right). Events are selected with the *inclusive* selection, and simulated backgrounds are normalized to luminosity.

4.4 Data and simulations comparison

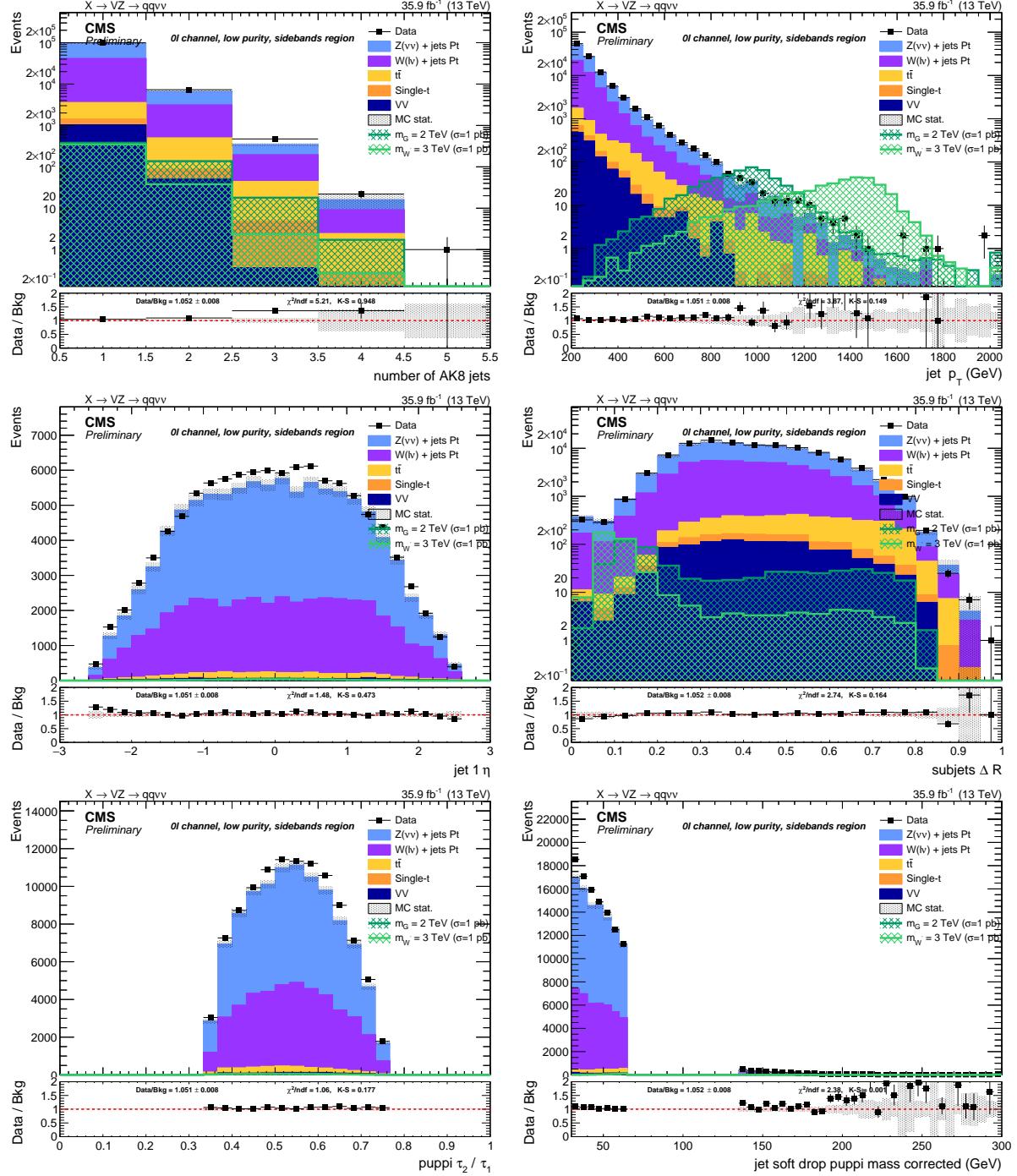


Figure 4.31: Top: number of AK8 jets in the event (left) and V jet candidate p_T (right). Center: V jet candidate η (left) and angular separation ΔR between the constituents leading subjets (right). Bottom: V jet candidate τ_{21} subjetiness after PUPPI correction (left) and V jet candidate soft drop PUPPI mass (right). Events are selected with the *low-purity sidebands* selection, and simulated backgrounds are normalized to luminosity.

Search for diboson resonances in the $VZ \rightarrow q\bar{q} \nu\bar{\nu}$ final state

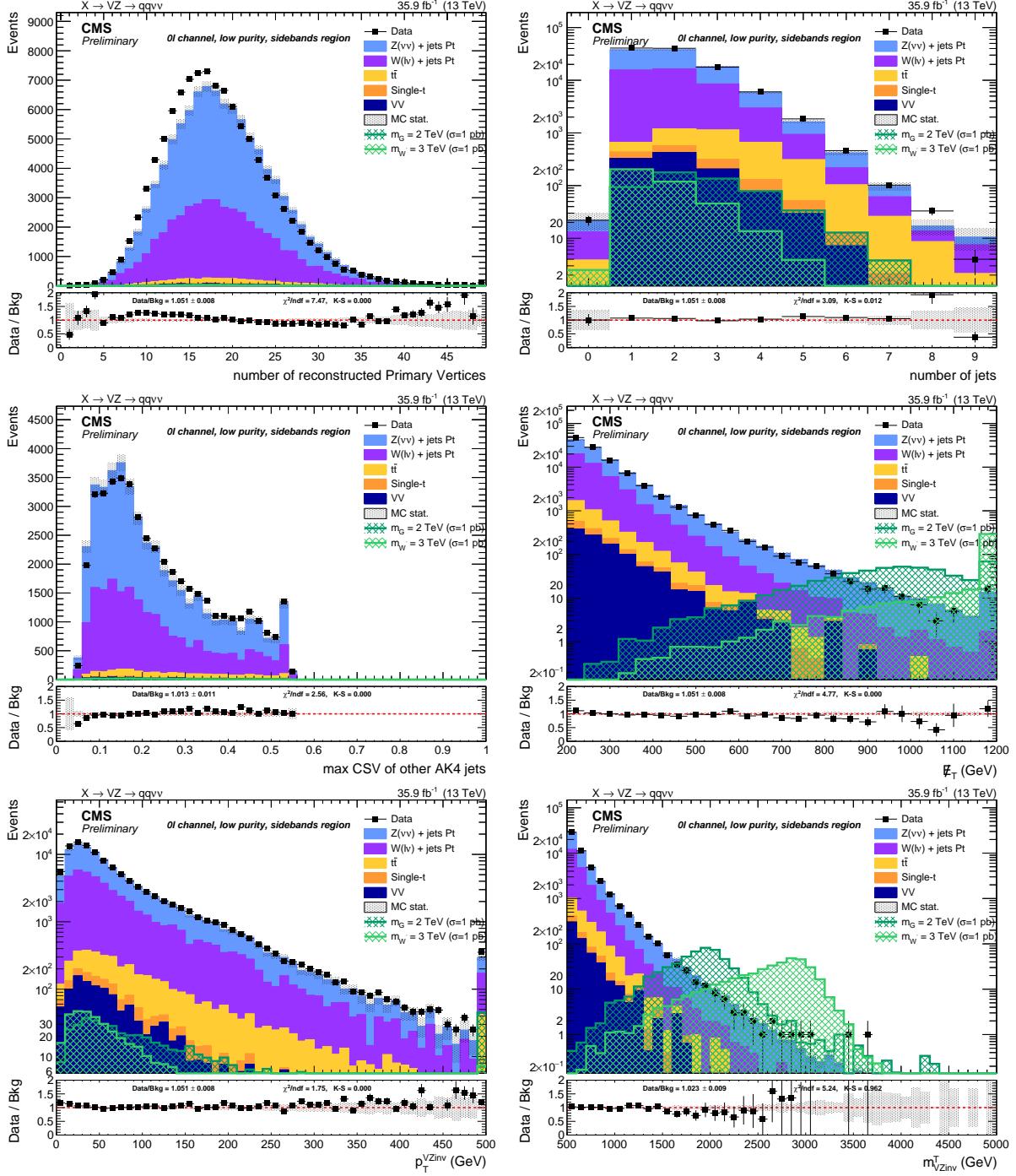


Figure 4.32: Top: number of reconstructed primary vertices (left) and number of AK4 jets in the event (right). Center: distribution of the b-tagging multivariate discriminant for the AK4 jets not included in the V jet cone (left) and E_T^{miss} distribution (right). Bottom: p_T of the VZ candidate (left) and transverse mass of the VZ candidate (right). Events are selected with the *low-purity sidebands* selection, and simulated backgrounds are normalized to luminosity.

4.4 Data and simulations comparison

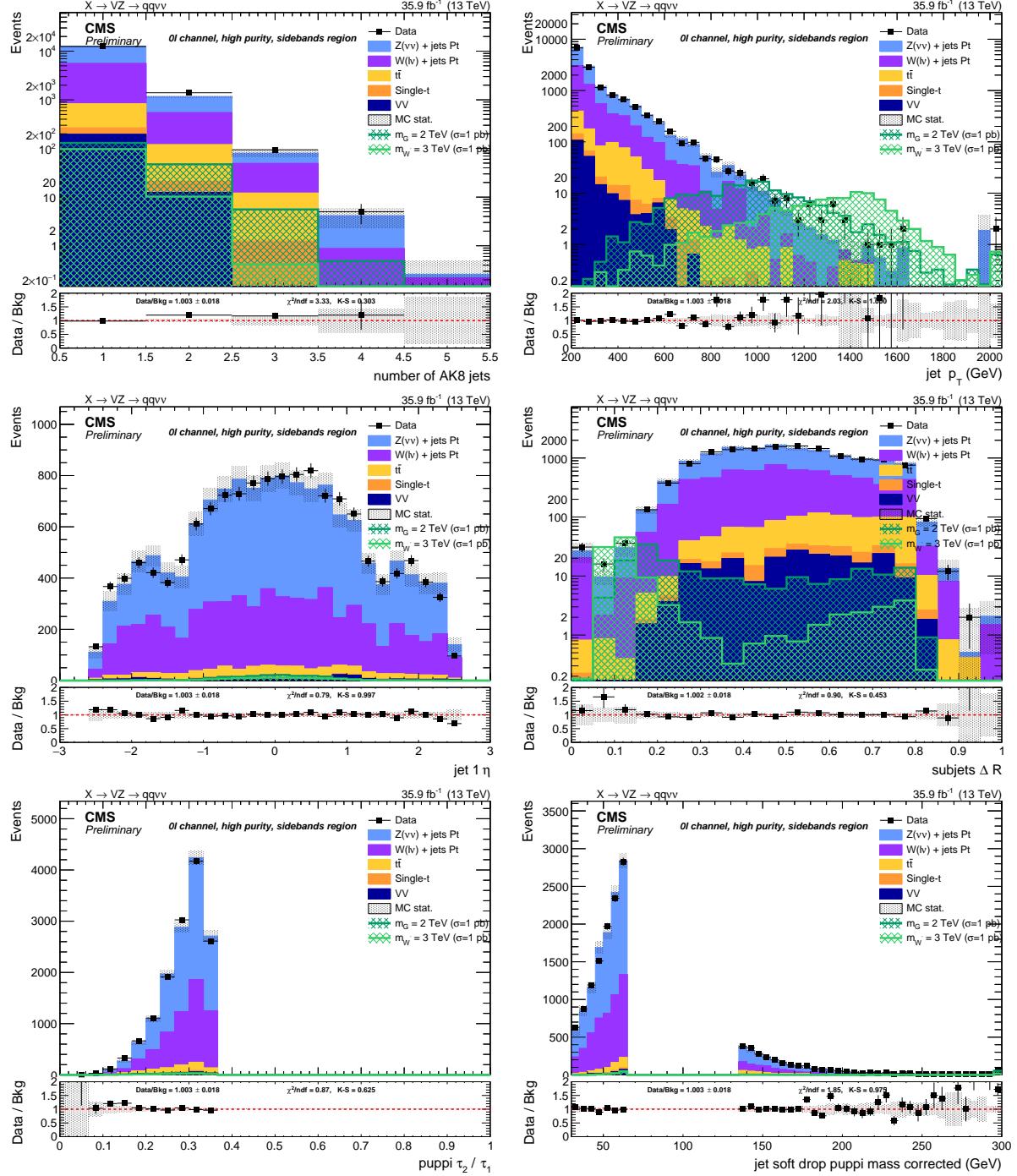


Figure 4.33: Top: number of AK8 jets in the event (left) and V jet candidate p_T (right). Center: V jet candidate η (left) and angular separation ΔR between the constituents leading subjets (right). Bottom: V jet candidate τ_{21} subjetiness after PUPPI correction (left) and V jet candidate soft drop PUPPI mass (right). Events are selected with the *high-purity sidebands* selection, and simulated backgrounds are normalized to luminosity.

Search for diboson resonances in the $VZ \rightarrow q\bar{q} \nu\bar{\nu}$ final state

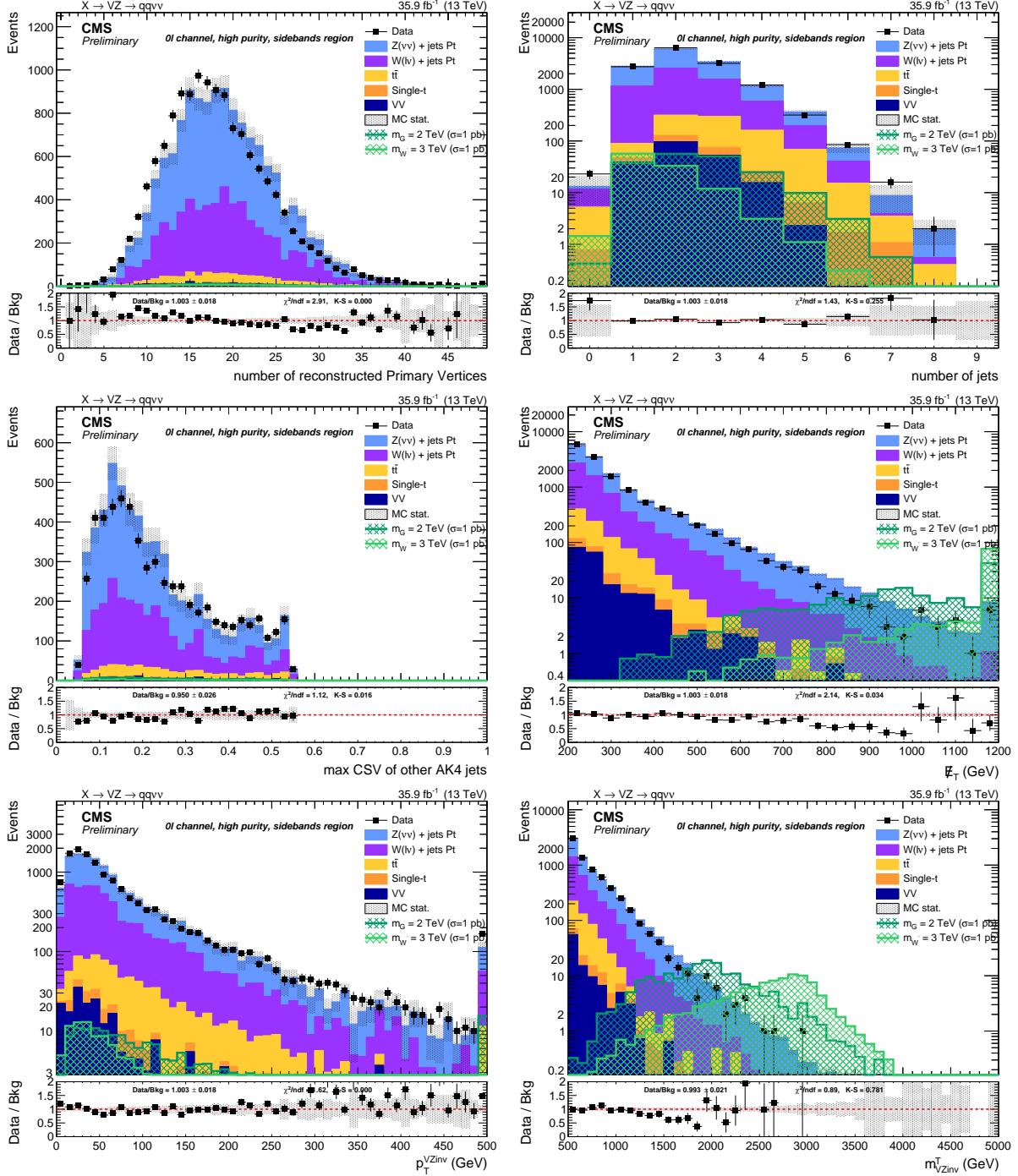


Figure 4.34: Top: number of reconstructed primary vertices (left) and number of AK4 jets in the event (right). Center: distribution of the b-tagging multivariate discriminant for the AK4 jets not included in the V jet cone (left) and E_T^{miss} distribution (right). Bottom: p_T of the VZ candidate (left) and transverse mass of the VZ candidate (right). Events are selected with the *high-purity sidebands* selection, and simulated backgrounds are normalized to luminosity.

4.4 Data and simulations comparison

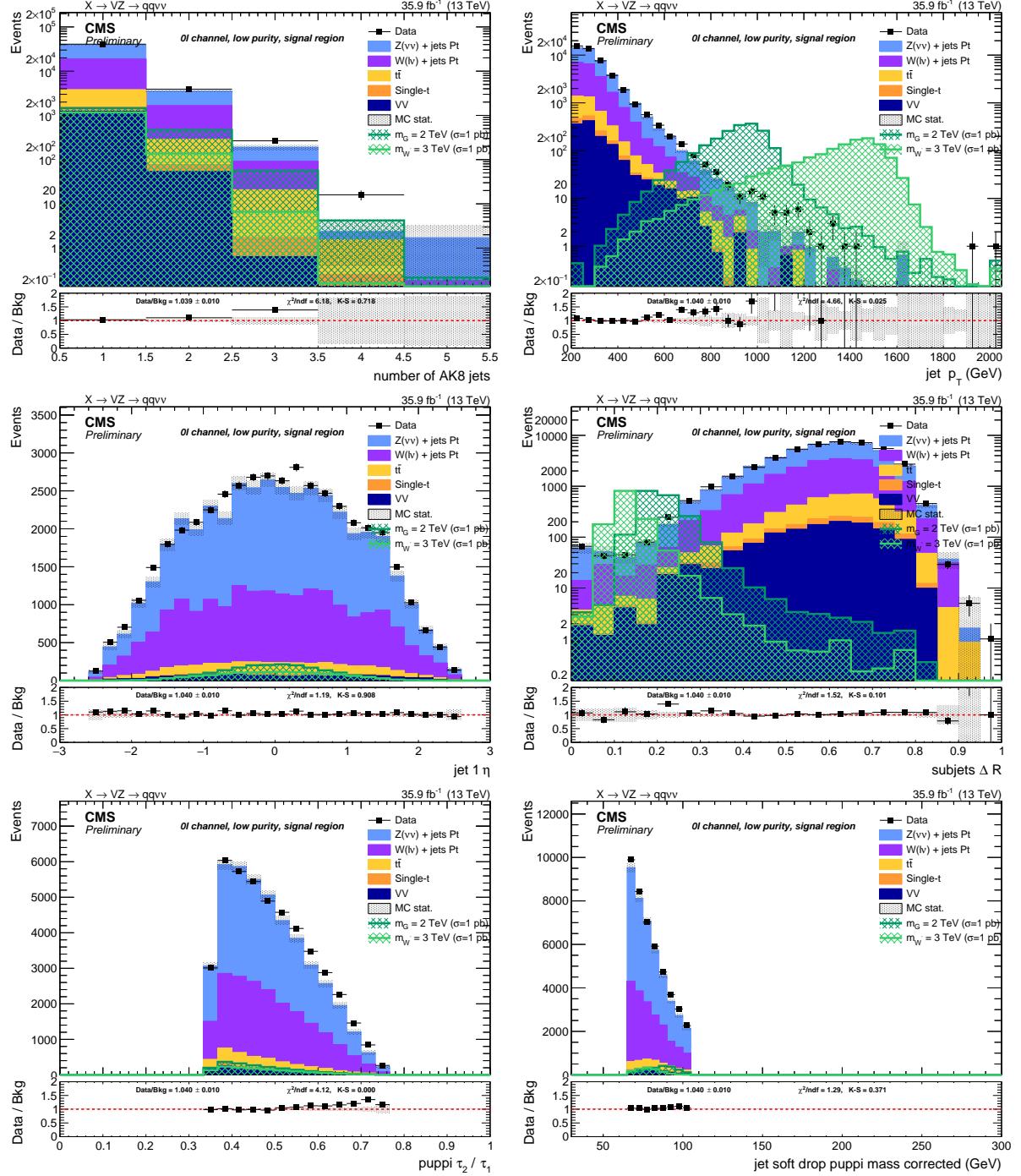


Figure 4.35: Top: number of AK8 jets in the event (left) and V jet candidate p_T (right). Center: V jet candidate η (left) and angular separation ΔR between the constituents leading subjets (right). Bottom: V jet candidate τ_{21} subjettness after PUPPI correction (left) and V jet candidate soft drop PUPPI mass (right). Events are selected with the *low-purity signal region* selection, and simulated backgrounds are normalized to luminosity.

Search for diboson resonances in the $VZ \rightarrow q\bar{q} \nu\bar{\nu}$ final state

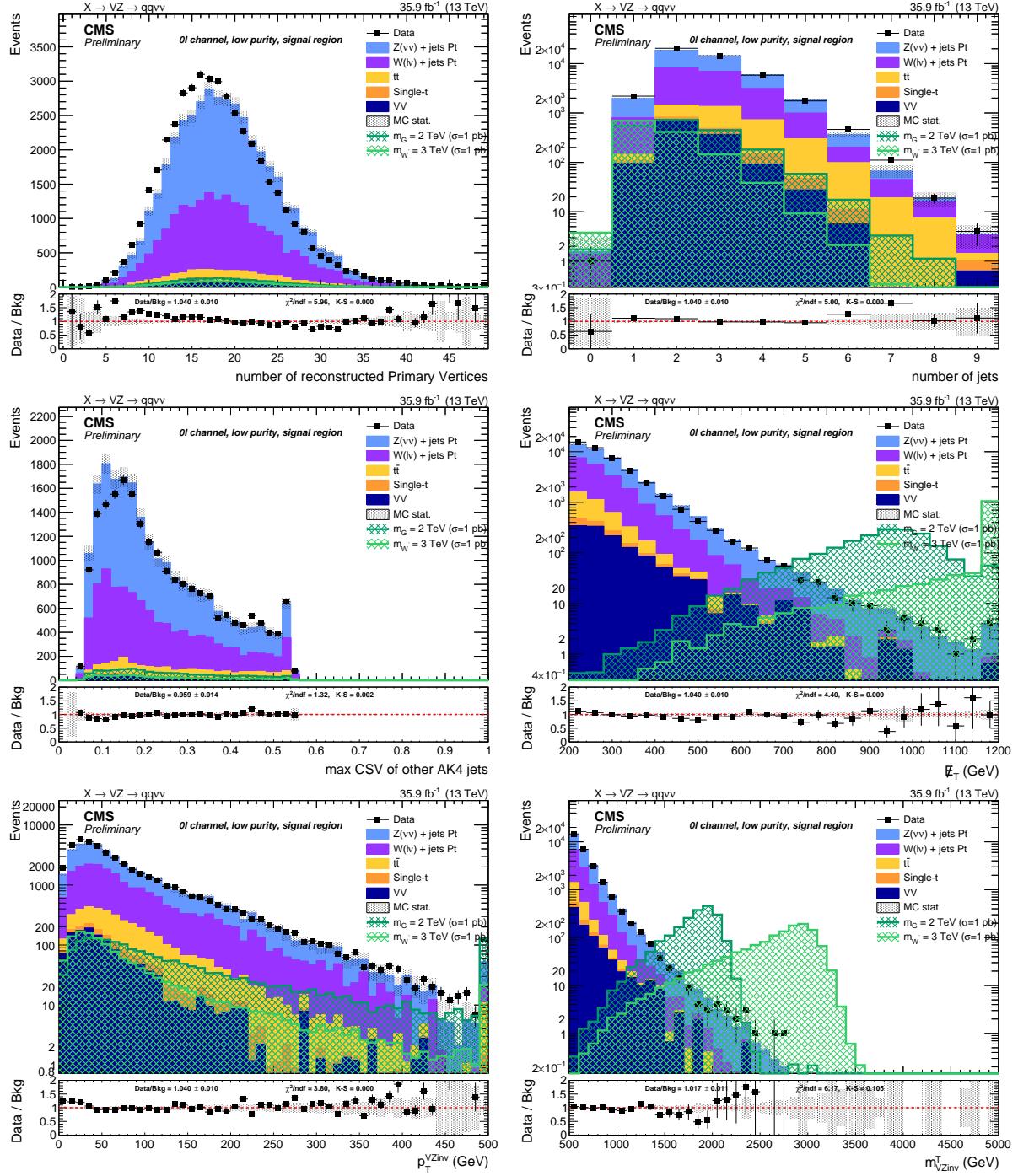


Figure 4.36: Top: number of reconstructed primary vertices (left) and number of AK4 jets in the event (right). Center: distribution of the b-tagging multivariate discriminant for the AK4 jets not included in the V jet cone (left) and E_T^{miss} distribution (right). Bottom: p_T of the VZ candidate (left) and transverse mass of the VZ candidate (right). Events are selected with the *low-purity signal region* selection, and simulated backgrounds are normalized to luminosity.

4.4 Data and simulations comparison

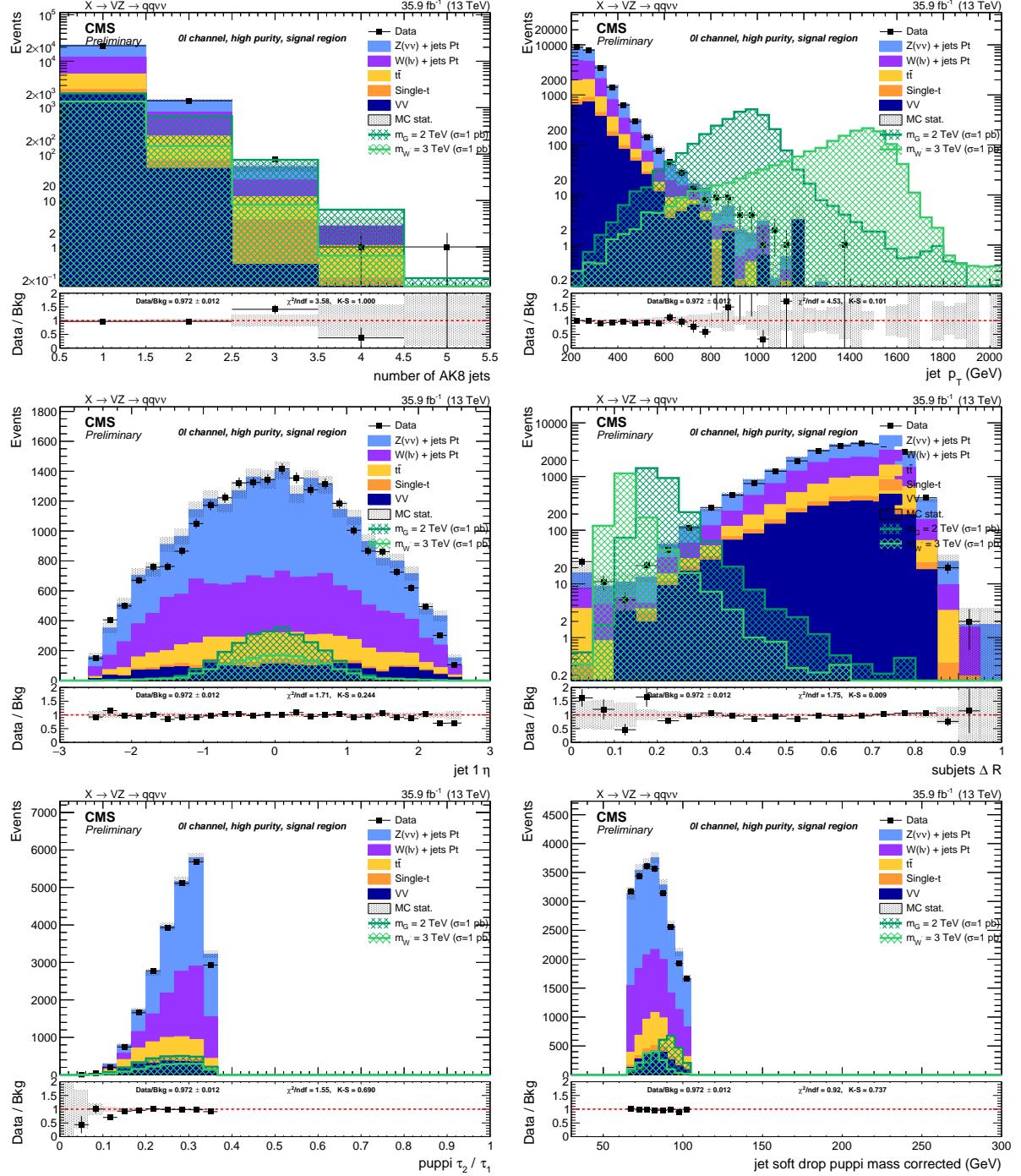


Figure 4.37: Top: number of AK8 jets in the event (left) and V jet candidate p_T (right). Center: V jet candidate η (left) and angular separation ΔR between the constituents leading subjets (right). Bottom: V jet candidate τ_{21} subjetiness after PUPPI correction (left) and V jet candidate soft drop PUPPI mass (right). Events are selected with the *high-purity signal region* selection, and simulated backgrounds are normalized to luminosity.

Search for diboson resonances in the $VZ \rightarrow q\bar{q} \nu\bar{\nu}$ final state

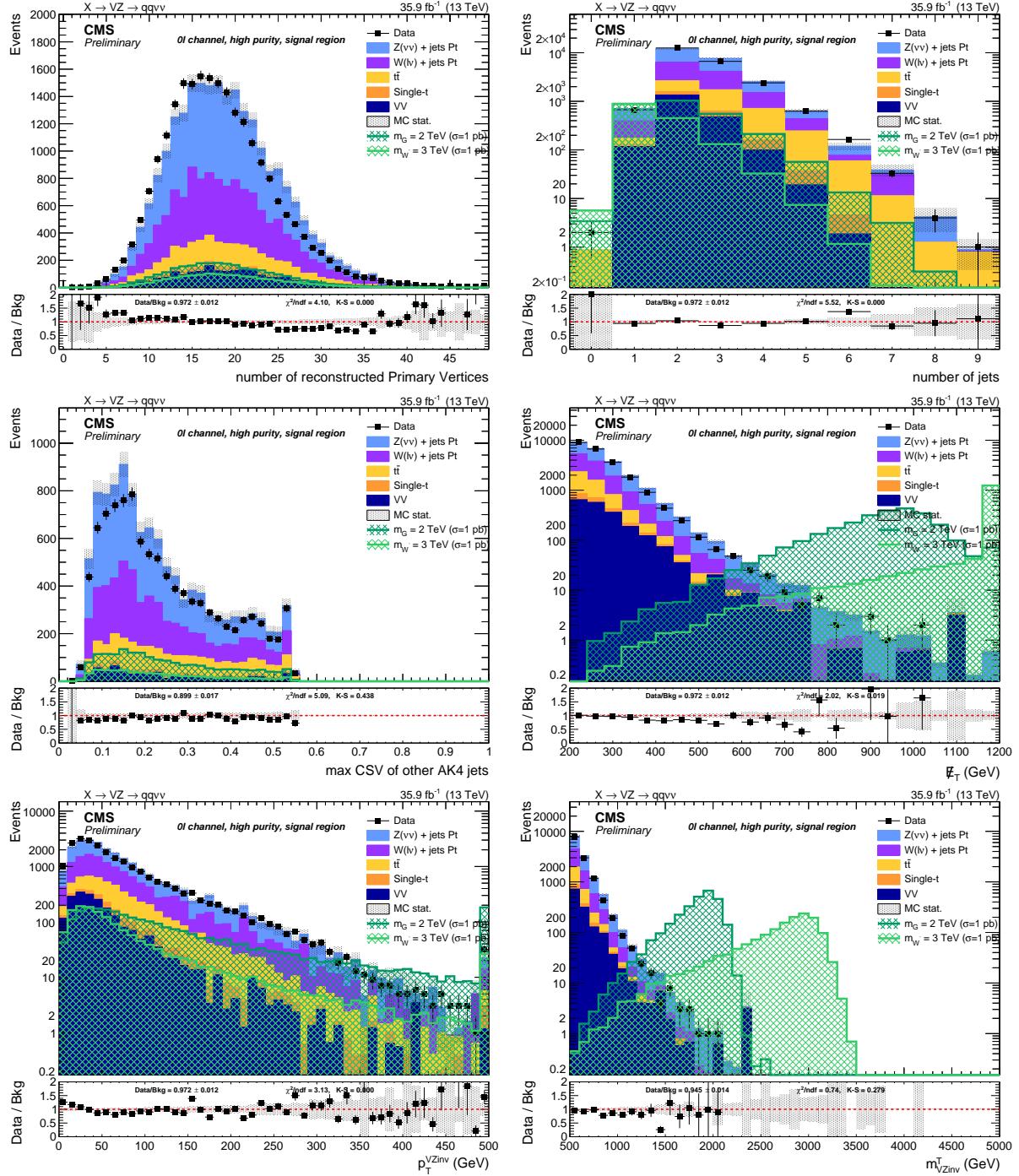


Figure 4.38: Top: number of reconstructed primary vertices (left) and number of AK4 jets in the event (right). Center: distribution of the b-tagging multivariate discriminant for the AK4 jets not included in the V jet cone (left) and E_T^{miss} distribution (right). Bottom: p_T of the VZ candidate (left) and transverse mass of the VZ candidate (right). Events are selected with the *high-purity signal region* selection, and simulated backgrounds are normalized to luminosity.

565 4.5 Background estimation technique

566 The goal of the analysis is to look for localized excesses in the m_{VZ}^T spectrum. The α method is used
 567 in searches for heavy resonances since Run 1 [18], and it has been introduced to be less dependent
 568 on the MC simulation for the background m_{VZ}^T estimation, due to the many sources of systematic
 569 uncertainties that are hard to understand and control. The two exclusive regions, *signal region* (SR)
 570 and *sidebands region* (SB), define a signal enriched or signal depleted phase space, respectively.
 571 First, the background normalization is extracted from data in the SB. Then, the α method extracts
 572 a predicted shape from the data in the SB to the SR using a transfer function (the α function) de-
 573 rived from simulation. The method relies on the assumption that the correlation between m_{VZ}^T and
 574 the groomed jet mass is reasonably well reproduced by the MC. The α -ratio is deemed to be more
 575 trustworthy since many systematic uncertainties would approximately cancel in the ratio.
 576 Let's assume that, in the simplest case, only one dominant background is present. The α function
 577 is defined as the ratio of the two functions describing the simulated m_{VZ}^T shape in the SR and SB:

$$\alpha(m_{VZ}^T) = \frac{f_{\text{SR}}^{\text{MC,bkg}}(m_{VZ}^T)}{f_{\text{SB}}^{\text{MC,bkg}}(m_{VZ}^T)}, \quad (4.14)$$

578 and the background distribution in the SR is thus estimated as the product of $\alpha(m_{VZ}^T)$ with the shape
 579 in the data SB:

$$f_{\text{bkg}}(m_{VZ}^T) = f_{\text{SB}}(m_{VZ}^T) \times \alpha(m_{VZ}^T) \quad (4.15)$$

580 In the above description, no definition of the SB and SR is included. Ideally, the best choice would
 581 be a variable such that the distribution of m_{VZ}^T in the signal region and sidebands are similar. In this
 582 analysis, the soft drop PUPPI corrected jet mass m_V (sec. 4.3.7) is chosen as the control variable, and
 583 the cut values are those reported in tab. 4.3.7. All the selections used in the α method background
 584 prediction are the same reported in sec. 4.3.11.
 585 In a real case scenario, the background is not purely composed of one single process neither in the
 586 SR nor in the SB. As already pointed out in sec. 4.2.3 and confirmed in sec. 4.4, the background
 587 composition is dominated by two processes, $Z + \text{jets}$ ($\sim 50\%$ in the whole SR) and $W + \text{jets}$ ($\sim 35\%$
 588 in the whole SR), grouped together as $V + \text{jets}$, whose modeling in simulation is considered not to
 589 be trustworthy. Other subdominant backgrounds, $t\bar{t}$ and single-t production, grouped as Top, and
 590 diboson (VV), generally have smaller contributions (of the order of 5% for VV , and 9% for Top,
 591 in the whole SR), and are considered quite well understood and modeled by MC generators. The
 592 justification of merging $W + \text{jets}$ and $Z + \text{jets}$ together as a single $V + \text{jets}$ background is provided
 593 in sec. 4.5.3.
 594 The shape and normalization of the VV and Top production are taken from the simulation. The
 595 shape and normalization of the main background are evaluated with the α approach. The V can-
 596 didate mass variable is used to perform the normalization prediction, the VZ candidate transverse
 597 mass variable is used for the shape prediction.
 598 A different background prediction is derived for each category separately, thus dividing low- and
 599 high-purity categories, and it is calculated in a transverse mass range $950 < m_{VZ}^T < 4750$ GeV.

600 4.5.1 Background normalization

601 The first step in the background prediction consists in a proper estimation of the background nor-
 602 malization. The jet mass distributions of the three backgrounds ($V + \text{jets}$, Top, and VV) are de-
 603 scribed with functional forms determined by fits on the simulated backgrounds. The so-built tem-
 604 plates are summed together, maintaining the relative weights between the three, and finally fitted

605 to the data in the jet mass sidebands. During the fit to data SB, the parameters of the $V + \text{jets}$ back-
 606 ground are left free to float and adapt to the data distribution. The integral of the final sum of the
 607 fitted functions over the SR jet mass range represents the background yield prediction in the SR.
 608 The empirical functional forms for each background are chosen to reflect the physics properties of
 609 the samples. In the low-purity category, the $V + \text{jets}$ background is a falling background with no
 610 peaks, modelled as a power law, while in the high-purity category the $V + \text{jets}$ background compo-
 611 nent is characterized by a broad distribution roughly centered at m_V , modelled as a gaussian, with
 612 an exponential tail at high mass values. The exponential falling $VV m_V$ spectrum shows a peak,
 613 corresponding to the reconstruction of a vector boson hadronic decay. The hadronic decays of W
 614 and Z bosons cannot be distinguished, hence they are modelled together as a gaussian. For the
 615 jet mass spectrum of the Top backgrounds, two peaks corresponding to the W and top quark mass
 616 can be observed; they are modelled as gaussian functions, superimposed to a falling exponential
 617 background.
 618 An extended likelihood fit is performed, hence the functional forms chosen to build the jet mass
 619 templates are normalized to unity (becoming probability density functions) through normalization
 620 factors (f_0, f_1):

- ErfPow2: an error function (Erf) multiplied by a power law, that is a function of the center-of-mass energy $\sqrt{s} = 13$ TeV. It depends on 4 parameters (the power law parameters c_0, c_1 , and the error function offset o and width w):

$$F_{\text{ErfPow2}}(x) = \left(\frac{x}{\sqrt{s}}\right)^{-c_0+c_1 \log(x/\sqrt{s})} \cdot \frac{1 + \text{Erf}((x-o)/w)}{2};$$

- ExpGaus: an exponential plus one gaussian. It depends on 4 parameters (the normalization f_0 , the exponential parameter a , the gaussian mean b and variance c):

$$F_{\text{ExpGaus}}(x) = f_0 \cdot e^{ax} + (1-f_0) \cdot e^{2(x-b)^2/c};$$

- ErfExpGaus: an error function, multiplied to an exponential, plus one gaussian. It depends on 6 parameters (the normalization f_0 , the exponential parameter a , the gaussian mean b and variance c , the error function offset o and width w):

$$F_{\text{ErfExpGaus}}(x) = f_0 \cdot e^{ax} \cdot \frac{1 + \text{Erf}((x-o)/w)}{2} + (1-f_0) \cdot e^{2(x-b)^2/c};$$

- ErfExpGaus2: an error function, multiplied to an exponential, plus two gaussians. It depends on 9 parameters (the normalization factors f_0 and f_1 , the exponential parameter a , the two gaussians means $b-d$ and variances $c-e$, the error function offset o and width w):

$$F_{\text{ErfExpGaus2}}(x) = f_0 \cdot e^{ax} \cdot \frac{1 + \text{Erf}((x-o)/w)}{2} + f_1 \cdot e^{2(x-b)^2/c} + (1-f_0-f_1) \cdot e^{2(x-d)^2/e}.$$

621 The choice of the functions is category-dependent, and it is summarized in tab. 4.15. In order to
 622 make the background evaluation less dependent as possible from the choice of the function de-
 623 scribing the jet mass of the main $V + \text{jets}$ background, an alternative function has been used to
 624 fit the $V + \text{jets}$ mass spectrum. The absolute difference bewteen the number of expected events
 625 calculated with the main $V + \text{jets}$ function and the alternative is taken as systematic uncertainty.
 626 The following plots (fig. 4.39-4.40) show the fits to the jet mass distributions in Monte Carlo samples,
 627 in the different categories; the alternative functions for the main background are displayed with

4.5 Background estimation technique

Table 4.15: Chosen functions to fit the jet mass distributions for each category.

Category	$V + \text{jets}$	alt. $V + \text{jets}$	Top	VV
low-purity	ErfPow2	ExpGaus	ErfExpGaus2	ExpGaus
high-purity	ExpGaus	ErfExpGaus	ErfExpGaus2	ExpGaus

Table 4.16: Expected background yield in the SB ($30 < m_V < 65 \text{ GeV}$, $m_V > 135 \text{ GeV}$) and in the SR ($65 < m_V < 105 \text{ GeV}$) and the respective systematic and statistical uncertainties.

Region	Category	Expected events	Statistical uncertainty	Systematic uncertainty	Alternative function uncertainty	Observed events
SB	low-purity	2356.6	± 52.5	± 16.0	± 1.1	2314
SR	low-purity	1093.2	± 48.1	± 16.4	± 49.1	1153
SB	high-purity	779.8	± 29.1	± 13.1	± 0.3	774
SR	high-purity	254.4	± 15.3	± 17.9	± 7.8	271

628 dotted lines. The background estimation, after the fit to data SB, is shown in fig. 4.41. The bottom
629 panels of each plot display the fit pulls (per-bin), namely, the number of events observed in data
630 (or in Monte Carlo simulations) minus the number of events predicted by the fit, divided by the
631 uncertainty in the data (or simulations). Table 4.16 summarizes the expected background yield in
632 the signal region, that is in agreement with observation in both the purity categories. The quoted
633 uncertainties are calculated as follows:

- 634 • the statistic uncertainty is the uncertainty of the fit to the $V + \text{jet}$ background performed on
635 data SB;
- 636 • the systematic uncertainty is the propagation of the uncertainties of the fits to the VV and
637 Top backgrounds performed on simulations, to the fit performed on data SB to extract the $V + \text{jets}$ functional parameters;
- 638 • the alternative function uncertainty describes the discrepancy in the background yield in SR
639 depending on the choice of the function to describe the $V + \text{jets}$ background.

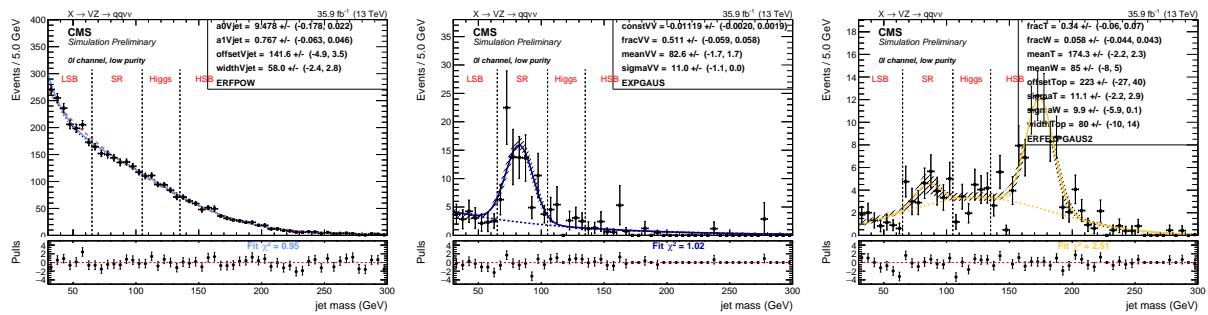


Figure 4.39: Fit to the simulated m_V in the low-purity category for the three backgrounds: $V + \text{jets}$ (left), VV (center), Top (right). For the main background prediction, the alternative function is displayed with a dotted red line, superimposed to the main choice (continuous light blue curve).

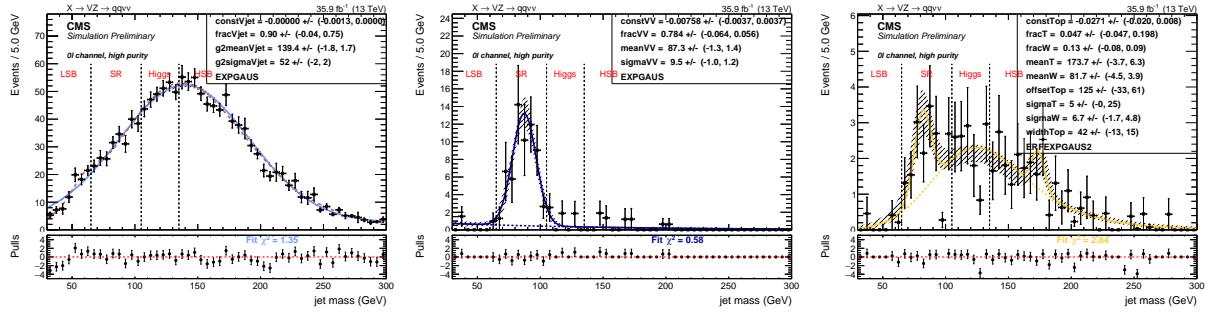


Figure 4.40: Fit to the simulated m_V in the high-purity category for the three backgrounds: $V + \text{jets}$ (left), VV (center), Top (right). For the main background prediction, the alternative function is displayed with a dotted red line, superimposed to the main choice (continuous light blue curve).

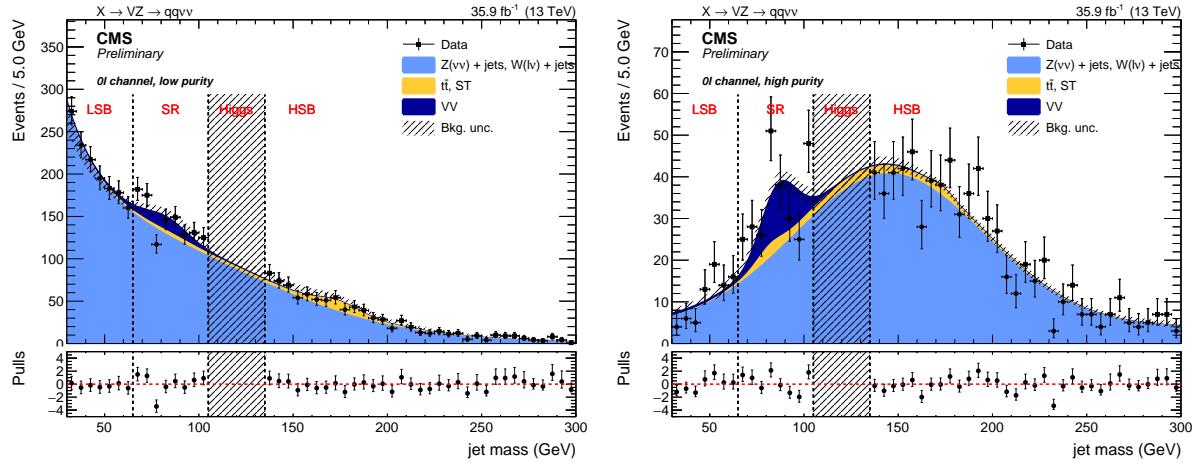


Figure 4.41: Background yield prediction in the signal region, after the fit to data sidebands, in the low- (left) and high-purity category (right). Data and predictions are in agreement.

4.5.2 Background shape

The second part of the background prediction consists in estimating the background shape of the transverse mass of the diboson candidate, m_{VZ}^T . Each transverse mass spectrum is parametrized separately for the $V + \text{jets}$ background ($f_{SR}^{\text{MC}, V + \text{jets}}(m_{VZ}^T)$, $f_{SB}^{\text{MC}, V + \text{jets}}(m_{VZ}^T)$), Top production ($f_{SR}^{\text{MC}, \text{Top}}(m_{VZ}^T)$, $f_{SB}^{\text{MC}, \text{Top}}(m_{VZ}^T)$), and diboson background ($f_{SR}^{\text{MC}, VV}(m_{VZ}^T)$, $f_{SB}^{\text{MC}, VV}(m_{VZ}^T)$). The parameters of these functions are extracted by fitting the simulated m_{VZ}^T spectra in SR and SB, respectively. The top and the diboson spectra are normalized to luminosity; the $V + \text{jets}$ spectrum is normalized according to the data-driven prediction obtained in sec. 4.5.1. The functions describing the $V + \text{jets}$ background, calculated from simulations, are used to define the α function, that has the purpose to take into account the kinematical differences of the SR compared to SB:

$$\alpha(m_{VZ}^T) = \frac{f_{SR}^{\text{MC}, V + \text{jets}}(m_{VZ}^T)}{f_{SB}^{\text{MC}, V + \text{jets}}(m_{VZ}^T)}. \quad (4.16)$$

The parameters describing the main background are then left free to float and extracted through a fit to data in the SB, after subtracting the corresponding Top and VV contribution from data. The resulting shape is then multiplied by the α function in order to get the main background expectation

4.5 Background estimation technique

- 654 in the SR. Finally, the Top and diboson contributions in the SR are added to the main background
 655 estimation.
 656 In formulas, the procedure used to extract the total background prediction is the following:

$$f_{\text{SR}}^{\text{data}}(m_{VZ}^T) = \left(f_{\text{SB}}^{\text{data}}(m_{VZ}^T) - f_{\text{SB}}^{\text{MC, Top}}(m_{VZ}^T) - f_{\text{SB}}^{\text{MC, VV}}(m_{VZ}^T) \right) \times \left[\frac{f_{\text{SR}}^{\text{MC, } V + \text{jets}}(m_{VZ}^T)}{f_{\text{SB}}^{\text{MC, } V + \text{jets}}(m_{VZ}^T)} \right] + f_{\text{SR}}^{\text{MC, Top}}(m_{VZ}^T) + f_{\text{SR}}^{\text{MC, VV}}(m_{VZ}^T), \quad (4.17)$$

- 657 where the expression in brackets represents the main background evaluation in data SB; the α -ratio
 658 is the expression enclosed in square brackets.
 659 The functions probed to parametrize the m_{VZ}^T distributions are smoothly falling exponential func-
 660 tions:

- ExpN: a product of two exponentials. It depends on two parameters a, b :

$$F_{\text{ExpN}}(x) = e^{ax+b/x}$$

- ExpTail: a modified exponential function with an additional parameter to model the expo-
 nential tails. It depends on two parameters a, b :

$$F_{\text{ExpTail}}(x) = e^{-x/(a+bx)}$$

Table 4.17: Main and alternative functions chosen to parametrize the background contributions in the m_{VZ}^T distribution for each category.

Category	Main bkg function	Main bkg alternative	Diboson	Top
low-purity	ExpN	ExpTail	ExpTail	ExpTail
high-purity	ExpTail	ExpN	ExpTail	ExpTail

- 661 The functions chosen to parametrize the backgrounds and extract the α function are reported in
 662 tab. 4.17, for each category. As a cross-check for the main α function used in the background esti-
 663 mation, an additional α function is extracted with alternative function choices for the $V + \text{jets}$ back-
 664 ground. Table 4.17 reports both the main function and the alternative function. In fig. 4.42 (4.44),
 665 the fits to each simulated background are reported for sidebands and signal region respectively, for
 666 low- (high-) purity categories. In fig. 4.43 (4.45), the results of the fit to data SB are presented for
 667 the low- (high-) purity categories: the expected background distribution in SB, where parameters
 668 describing the $V + \text{jets}$ background are extracted according to data distribution (left); the α -ratio
 669 function, calculated with the main function to describe the $V + \text{jets}$ background (black solid line)
 670 and the alternative function (gray dotted line) (center); the full background estimation performed
 671 with the main and alternative functions for describing the $V + \text{jets}$ background: the background
 672 shape in SB (blue solid curve for the main function, light blue dotted curve for the alternative) and
 673 the final background shape in SR (red solid line for the main function, green dotted line for the al-
 674 ternative) (right). A proof to the compatibility of the two predictions in SR is presented in sec. 4.5.3.
 675 The bottom panels in the plots display the fit pulls (per-bin), namely, the number of events observed
 676 in data (or in Monte Carlo simulations) minus the number of events predicted by the fit, divided by
 677 the uncertainty in the data (or simulations).
 678 Fig. 4.46 summarizes the final background predictions as a function of the search variable, the trans-
 679 verse mass. Data and predictions are in agreement in both the categories.

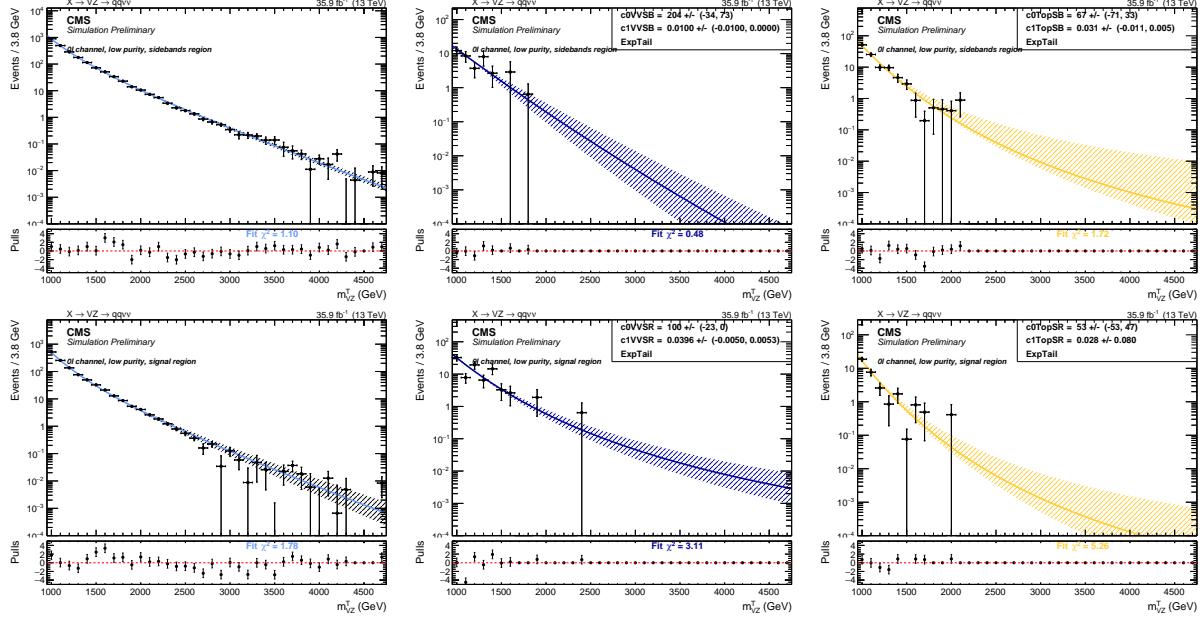


Figure 4.42: Low-purity category. Top: fits to the simulated background components $V + \text{jets}$ (left), VV (center), Top (right) in the sidebands (SB). Bottom: fits to the simulated background components $V + \text{jets}$ (left), VV (center), Top (right) in the signal region (SR).

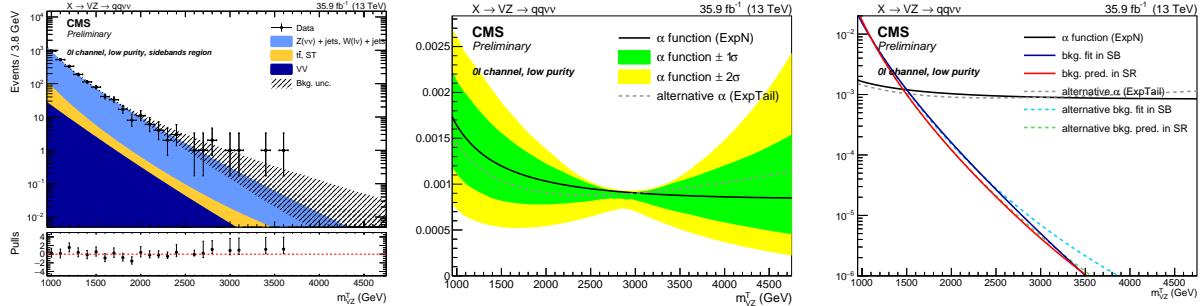


Figure 4.43: Low-purity category. Result of the fit to data in the SB (left), α -ratio function (center), and α function compared to the background shape in both SB and SR (right). The black line, with the corresponding 1σ (green) and 2σ (yellow) uncertainty bands, represents the α function. The gray line is the alternative α function. The blue and red solid lines represent the estimated background in the SB and SR, respectively, with both the main (solid line) and alternative (dotted line) parametrizations.

4.5.3 Alpha method validations

The first required validation is performed in order to legitimate the choice of putting the $Z + \text{jets}$ and the $W + \text{jets}$ backgrounds together while performing the background estimation. The full procedure has been repeated, by keeping the two background contributions separated. Fit results performed in SB (top plots) and SR (bottom plots) in MC samples are displayed in fig. 4.47 (4.48) for low- (high-) purity category, for $Z + \text{jets}$ and $W + \text{jets}$ background, separately, and for the combination of the two. In fig. 4.49, the α functions calculated for $Z + \text{jets}$ background (red dotted line) and for $W + \text{jets}$ background (blue dotted line) are in agreement with the α function used in the analysis (black solid line), calculated by merging together the two backgrounds, both in low- (left plot) and

4.5 Background estimation technique

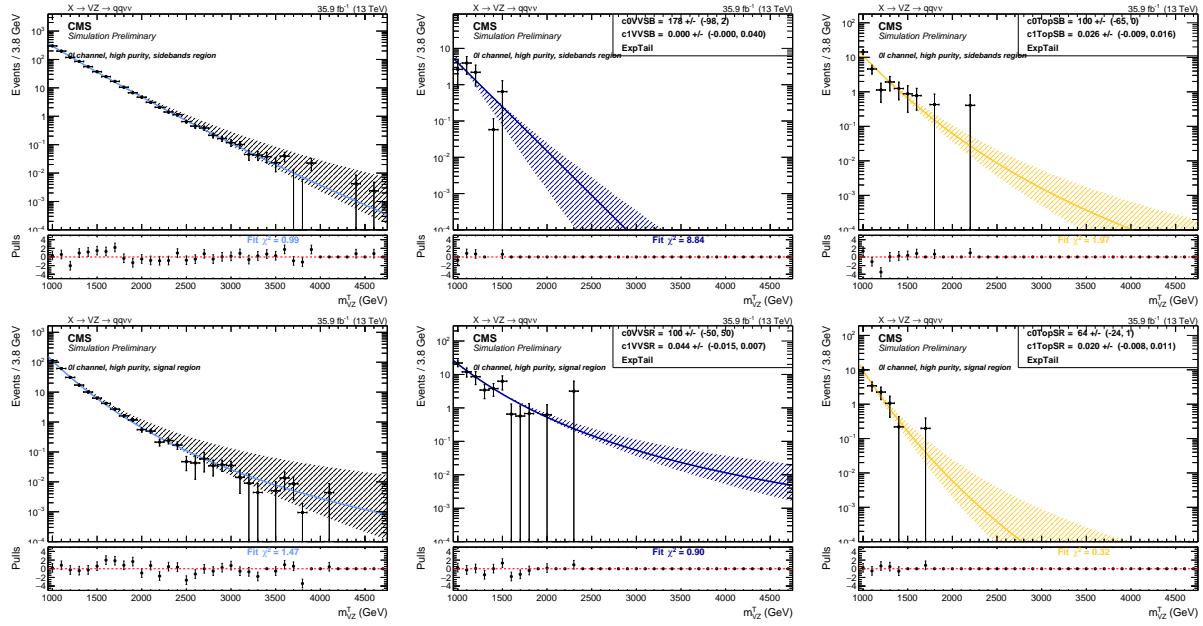


Figure 4.44: High-purity category. Top: fits to the simulated background components $V + \text{jets}$ (left), VV (center), Top (right) in the sidebands (SB). Bottom: fits to the simulated background components $V + \text{jets}$ (left), VV (center), Top (right) in the signal region (SR).

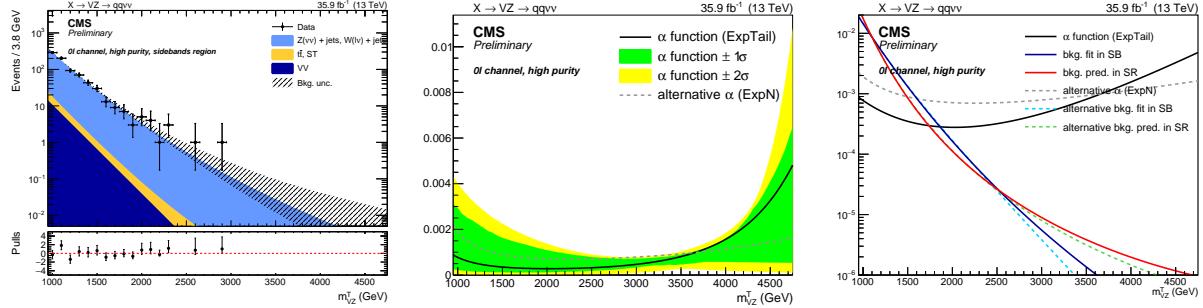


Figure 4.45: High-purity category. Result of the fit to data in the SB (left), α -ratio function (center), and α function compared to the background shape in both SB and SR (right). The black line, with the corresponding 1σ (green) and 2σ (yellow) uncertainty bands, represents the α function. The gray line is the alternative α function. The blue and red solid lines represent the estimated background in the SB and SR, respectively, with both the main (solid line) and alternative (dotted line) parametrizations.

689 high-purity category (right plot).

690 As a robustness check of the α -ratio method, a closure test is performed on data. Instead of pre-
691 dicting the background in the real SR from both the lower and the upper jet mass sidebands, the SB
692 and SR are redefined for the purposes of this test. The low sideband is splitted into two sub-regions:
693 30 – 50 GeV (LSB) and 50 – 65 GeV (SR). The former is considered as the new low sideband, while
694 the latter is exploited as a pseudo-signal region. The high sideband is instead effectively used in the
695 fit without any modifications with respect to the standard α -ratio method. With this configuration,
696 the prediction of the background in the SR region is estimated from the fit to the LSB region and

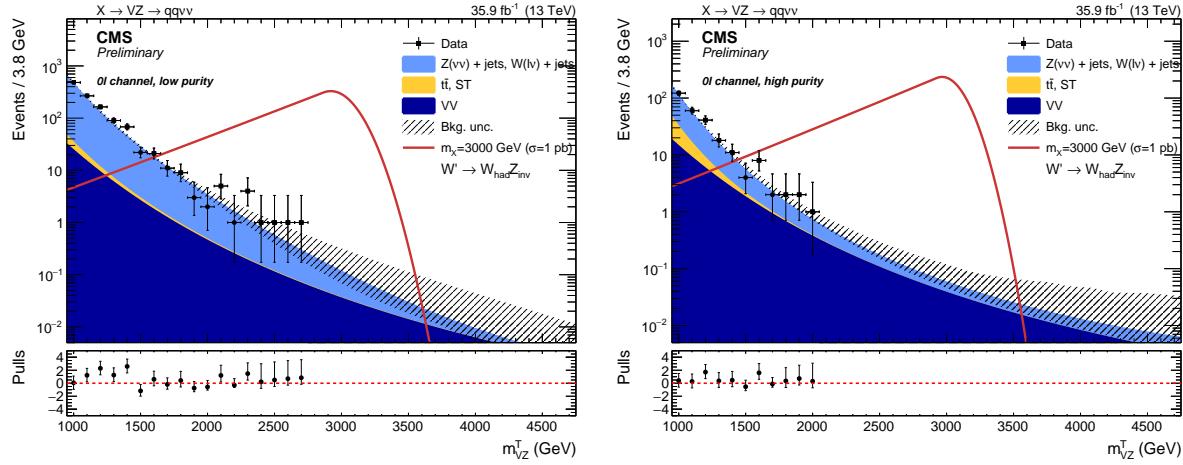


Figure 4.46: Expected background predicted with the α method in the low- (left) and high-purity category (right), compared to observations (black markers) and a signal hypothesis of a spin-1 W' of mass 3 TeV.

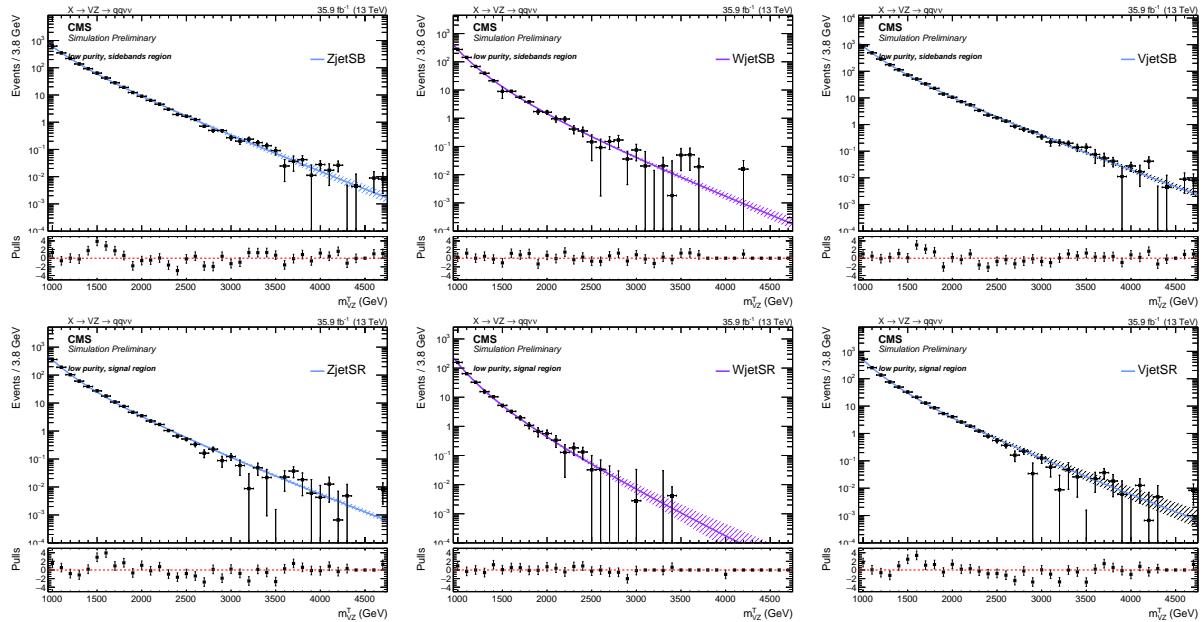


Figure 4.47: Validation of the α method, low-purity category. Top: fits to the simulated background components $Z + \text{jets}$ (left), $W + \text{jets}$ (center), and their combination $V + \text{jets}$ (right), in the sidebands (SB). Bottom: fits to the simulated background components $Z + \text{jets}$ (left), $W + \text{jets}$ (center), and their combination $V + \text{jets}$ (right), in the signal region.

the high-sidebands, and checked with data for both shape and normalization. This test has been performed before the unblinding of the signal region of the analysis.

In fig. 4.50 and tab. 4.18, the predicted shapes and normalizations are compared to the observed ones in data. A good overall agreement both in normalization and shape is obtained. There is a bit of tension in normalization for high-purity category, due to an upper fluctuation in data around 60 GeV. This cross check confirms that the method to extract the $V + \text{jets}$ background is reliable and can be used to model the background in the search for potential excesses in the signal region

4.5 Background estimation technique

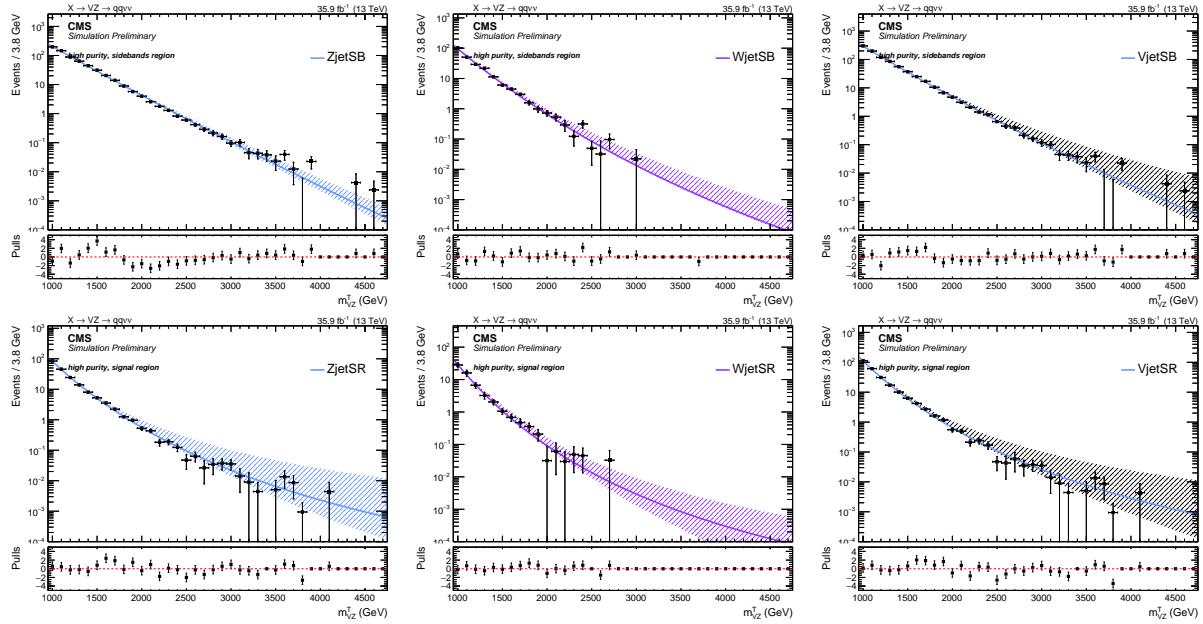


Figure 4.48: Validation of the α method, high-purity category. Top: fits to the simulated background components $Z + \text{jets}$ (left), $W + \text{jets}$ (center), and their combination $V + \text{jets}$ (right), in the sidebands (SB). Bottom: fits to the simulated background components $Z + \text{jets}$ (left), $W + \text{jets}$ (center), and their combination $V + \text{jets}$ (right), in the signal region.

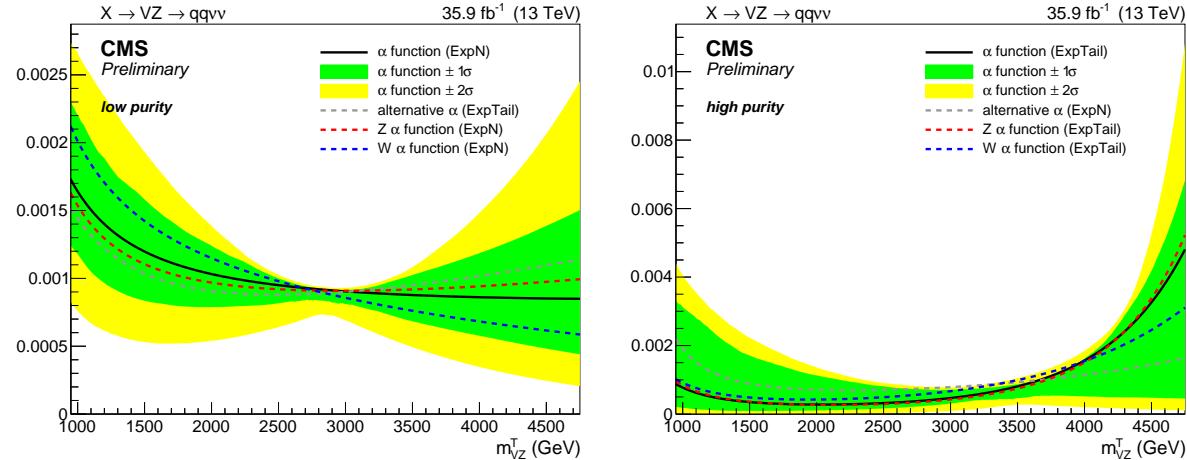


Figure 4.49: Validation of the α method: α functions calculated for $Z + \text{jets}$ background (red dotted line) and for $W + \text{jets}$ background (blue dotted line) separately, and α function for the total $V + \text{jets}$ background (black solid line). Left: low-purity category; right: high-purity category.

704 defined in the analysis.

705 The last check performed is a study of the impact of the choice of the function to describe the V
 706 + jets background on the very last result of the analysis, namely the exclusion limit on the signal
 707 cross-section times branching fraction. The procedure of the limit extraction is discussed in detail
 708 in sec. ???. The main and alternative functions chosen to parametrize the dominant background
 709 depend on the purity category and are listed in tab. 4.17. In fig. 4.51 (top), the fit results of the

Table 4.18: Expected and observed background yield in the pseudo-SR jet mass region ($50 < m_V < 65$ GeV), predicted from the LSB one ($30 < m_V < 50$ GeV) and high-sideband ($m_V > 135$ GeV).

Region	Category	Expected	Observed
SB	low-purity	1841.3 ± 45.7	1793
pseudo-SR	low-purity	529.9 ± 37.8	521
SB	high-purity	728.5 ± 29.9	725
pseudo-SR	high-purity	39.3 ± 5.2	49

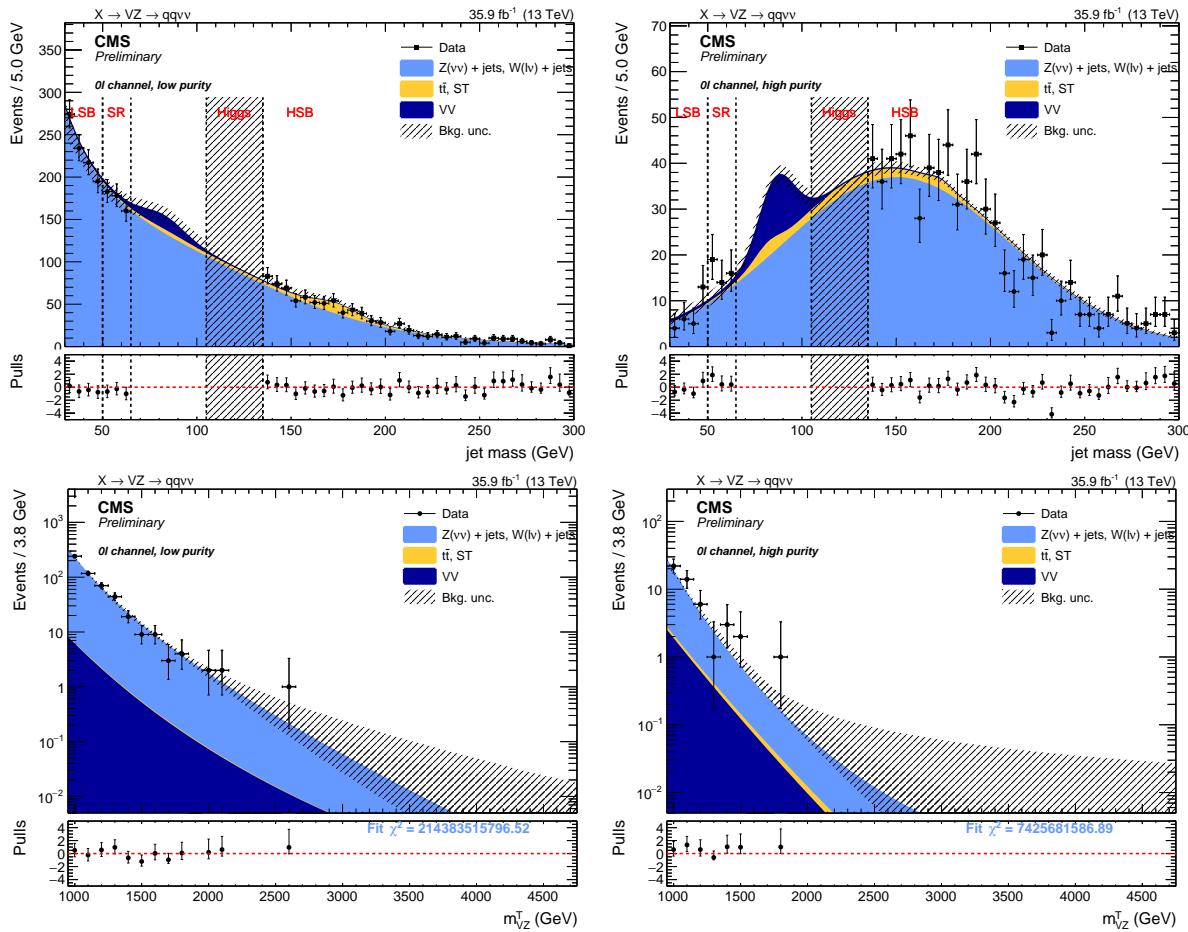


Figure 4.50: Top: results of the fit to the m_V spectrum in data, in the sidebands defined for the α method validation: low-sideband ($30 < m_V < 50$ GeV) and high-sideband ($m_V > 135$ GeV) (left: low-purity category, right: high-purity category). Bottom: results of the fits to the m_{VZ}^T spectrum, in the pseudo-signal region ($50 < m_V < 65$ GeV) defined for the α method validation (left: low-purity category, right: high-purity category). Both the true signal region and the Higgs regions are kept blind.

710 background shape prediction of the transverse mass are displayed. They are obtained by choosing
 711 the main function to describe the main background (red curve) and the alternative function (green
 712 curve); the two predictions are in agreement and very close to each other, for both low- (left) and
 713 high- (right) purity category. In fig. 4.51 (center), the 95% CL exclusion limits on cross-section times
 714 branching fraction are displayed for a spin-2 bulk graviton hypothesis, as a function of the mass of the

4.5 Background estimation technique

resonance. The same figure of merit is shown in fig. 4.51 (bottom), considering a spin-1 W' hypothesis. In the plots, the exclusion limits are calculated by choosing the main function to describe the $V + \text{jets}$ background (left plots: green curve for low-purity category alone and black curve for high-purity category alone, right plot: red curve for the combination of the categories) or the alternative function (left plots: orange curve for low-purity category alone and pink curve for high-purity category alone, right plot: blue curve for the combination of the categories). The impact of the choice of the function is negligible ($<< 1\%$).

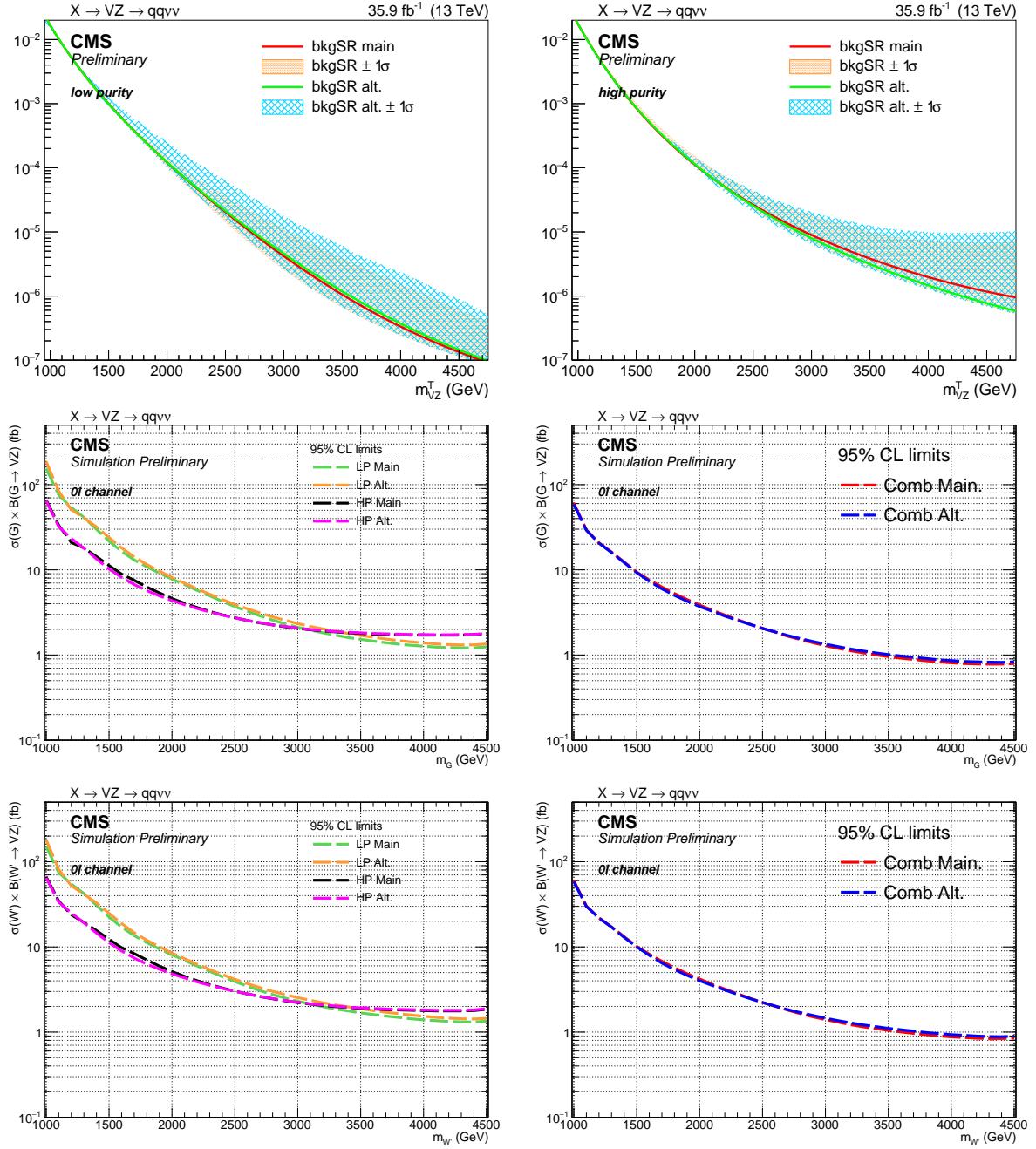


Figure 4.51: Validation of the α method: impact of the choice of the function to describe the dominant $V + \text{jets}$ background. Top: fit results of the background shape prediction in the SR obtained with the main function (red curve) and the alternative function (green curve), for low- (left) and high- (right) purity categories. Center: exclusion limits on cross-section times branching fraction for a spin-2 bulk graviton hypothesis, as a function of the mass of the resonance, calculated by choosing the main function (left plots: green curve for low-purity category alone and black curve for high-purity category alone, right plot: red curve for the combination of the categories) or the alternative function (left plots: orange curve for low-purity category alone and pink curve for high-purity category alone, right plot: blue curve for the combination of the categories). Bottom: exclusion limits on cross-section times branching fraction for a spin-1 W' hypothesis.

4.5 Background estimation technique

4.5.4 Signal modeling

The simulated signal samples, with different resonance mass hypotheses, are fitted in the SR with an empirical function in order to be able to perform an unbinned likelihood fit for the signal extraction. The function chosen to model the signal samples is a *Crystal Ball* function [19, 20], which is composed by a gaussian-like core convoluted to two power-law tails. Both spin-2 (fig. ??) and spin-1 (fig. 4.53) signal samples are fitted.

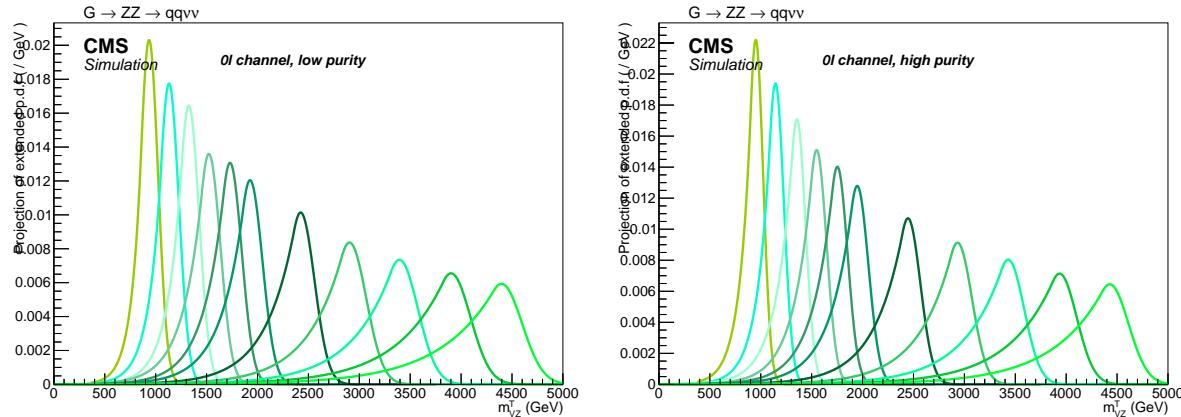


Figure 4.52: Interpolation of the signal as a function of the resonance transverse mass m_{VZ}^T , for a spin-2 (bulk graviton) signal hypothesis with an arbitrary cross section of 1 pb in the low- (left) and high-purity category (right).

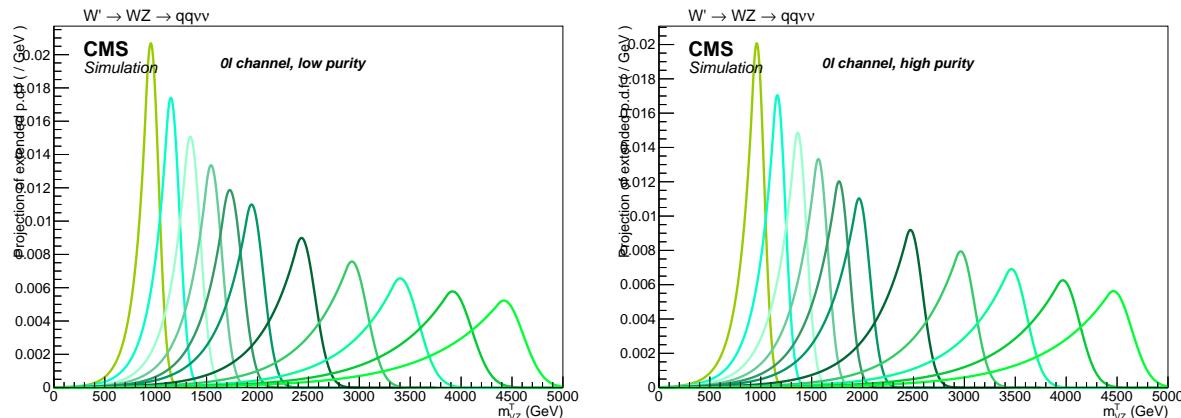


Figure 4.53: Interpolation of the signal as a function of the resonance transverse mass m_{VZ}^T , for a spin-1 (W') signal hypothesis with an arbitrary cross section of 1 pb in the low- (left) and high-purity category (right).

4.5.4.1 Signal parametrization

The signal is parametrized by interpolating the fitted parameters separately for each category in order to have a continuous variation of the signal shape for every possible m_{VZ}^T value within the range. A linear fit is performed on the mean and the width of the gaussian core of the Crystal Ball functions. The interpolations are shown in fig. 4.54- 4.55 for the spin-2 signal model, and in fig. 4.56- 4.57 for

733 the spin-1 signal model. Shape systematic uncertainties, as described in sec. 4.6, are taken into ac-
 734 count while describing the mean and sigma of the gaussian core, and they are related to the effects
 735 of the jet mass scale and resolution. Other shape parameters describing the tails of the Crystal Ball
 736 are fitted as 3rd degree polynomial.

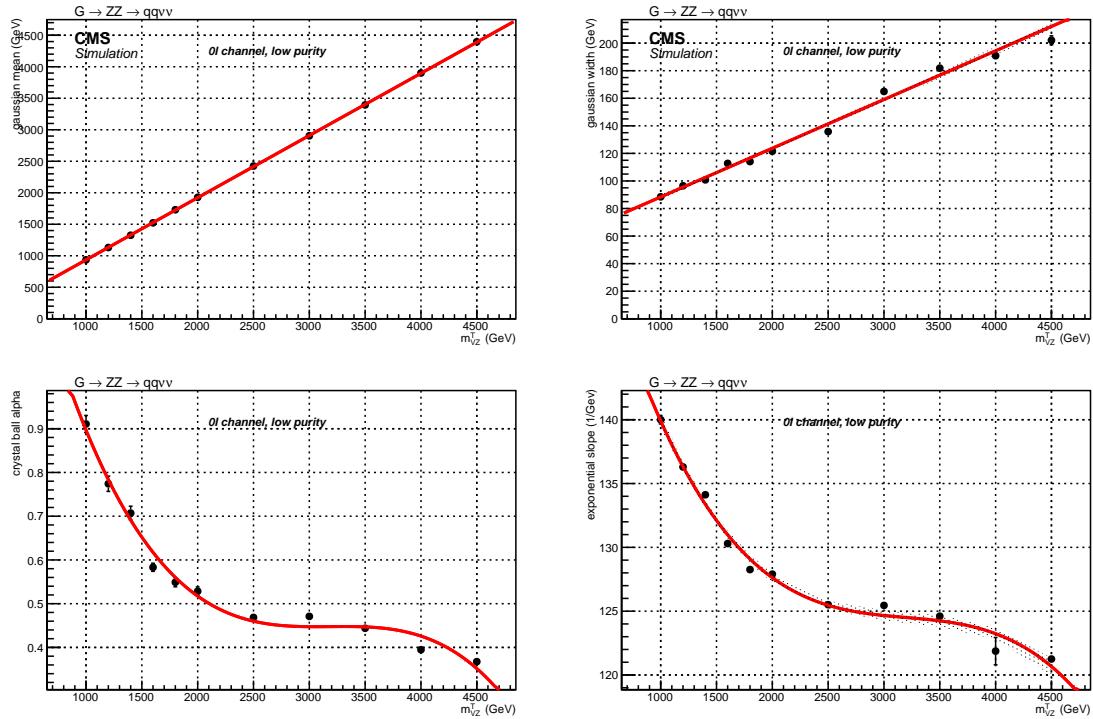


Figure 4.54: Interpolation of the fitted parameters as a function of the resonance transverse mass m_{VZ}^T , for a spin-2 (bulk graviton) signal hypothesis, low-purity category.

737 The normalization of the signal samples is extrapolated from the fitted integral of the Crystal Ball
 738 functions. The points are then connected with a line, in order to have an acceptable description
 739 of the normalization as a function of m_{VZ}^T . The interpolations are shown in fig. 4.58 for the spin-2
 740 signal model, and in fig. 4.59 for the spin-1 signal model.

4.5 Background estimation technique

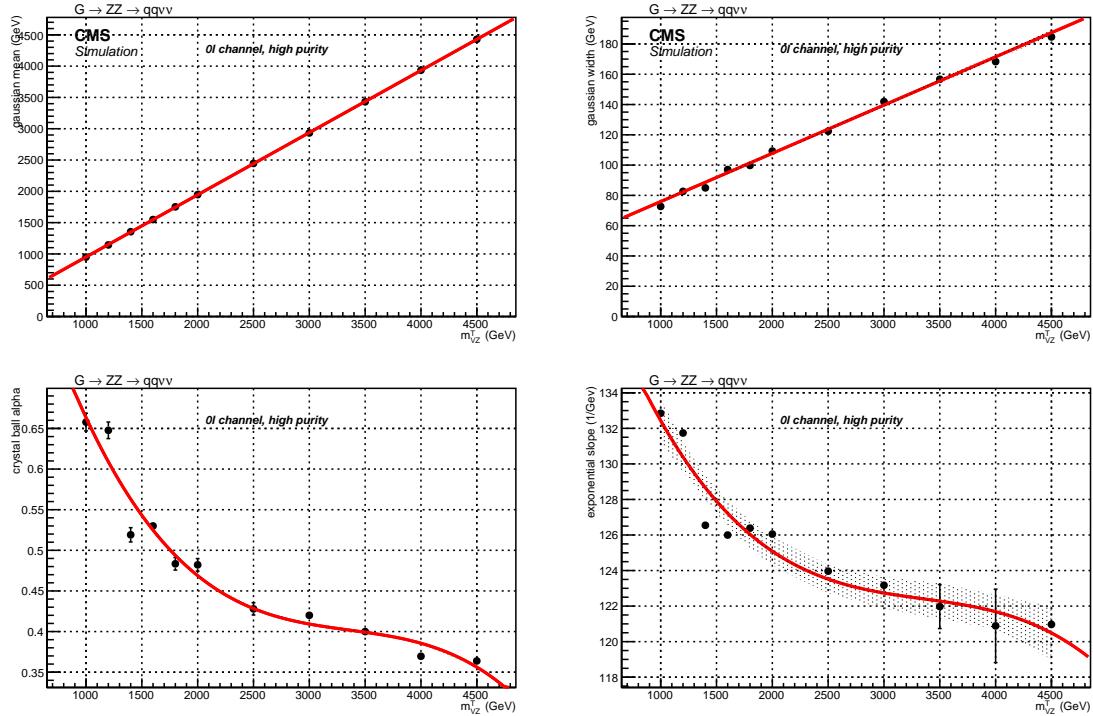


Figure 4.55: Interpolation of the fitted parameters as a function of the resonance transverse mass m_{VZ}^T , for a spin-2 (bulk graviton) signal hypothesis, high-purity category.

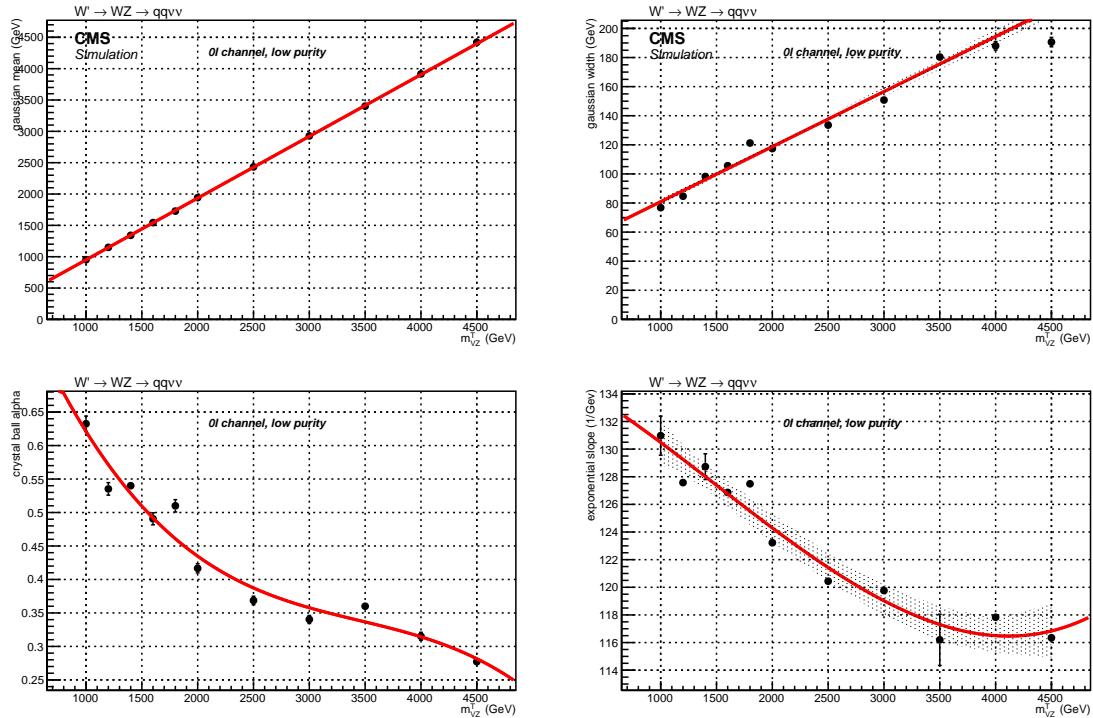


Figure 4.56: Interpolation of the fitted parameters as a function of the resonance transverse mass m_{VZ}^T , for a spin-1 (W') signal hypothesis, low-purity category.

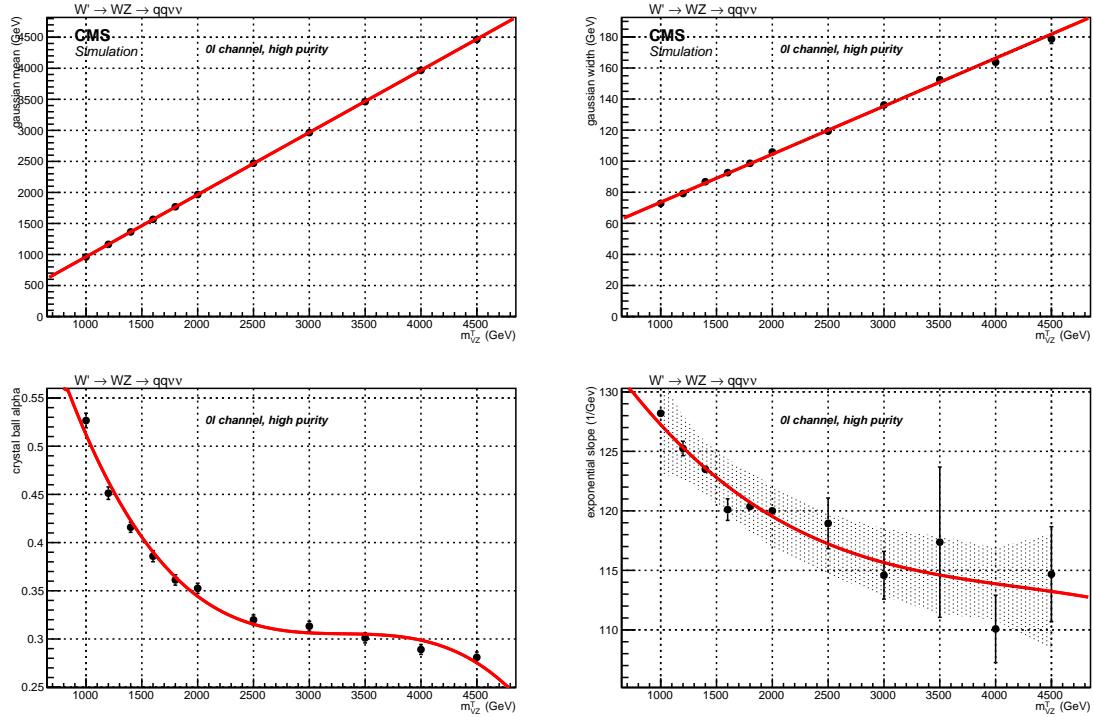


Figure 4.57: Interpolation of the fitted parameters as a function of the resonance transverse mass m_{VZ}^T , for a spin-1 (W') signal hypothesis, high-purity category.

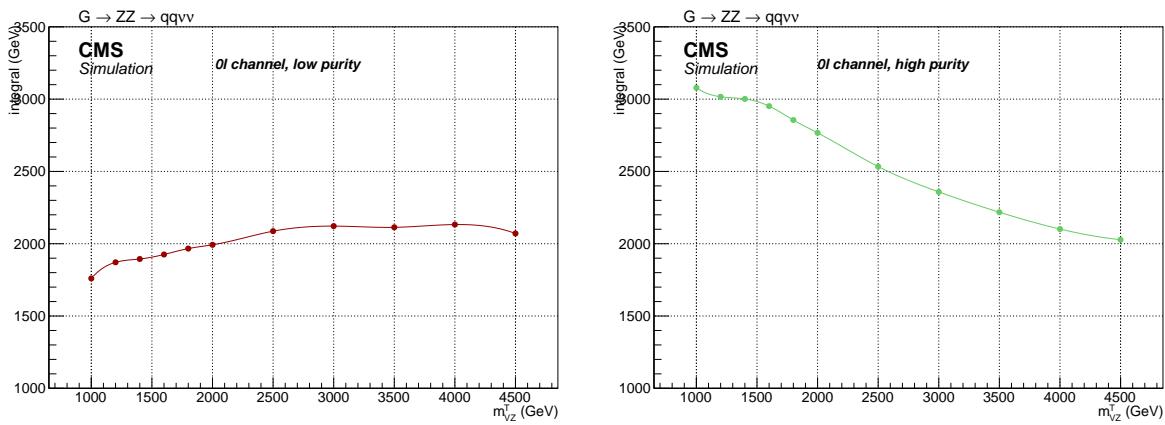


Figure 4.58: Interpolation of the signal normalization as a function of the resonance transverse mass m_{VZ}^T , for a spin-2 (bulk graviton) signal hypothesis. From left to right: low-purity, high-purity.

4.5 Background estimation technique

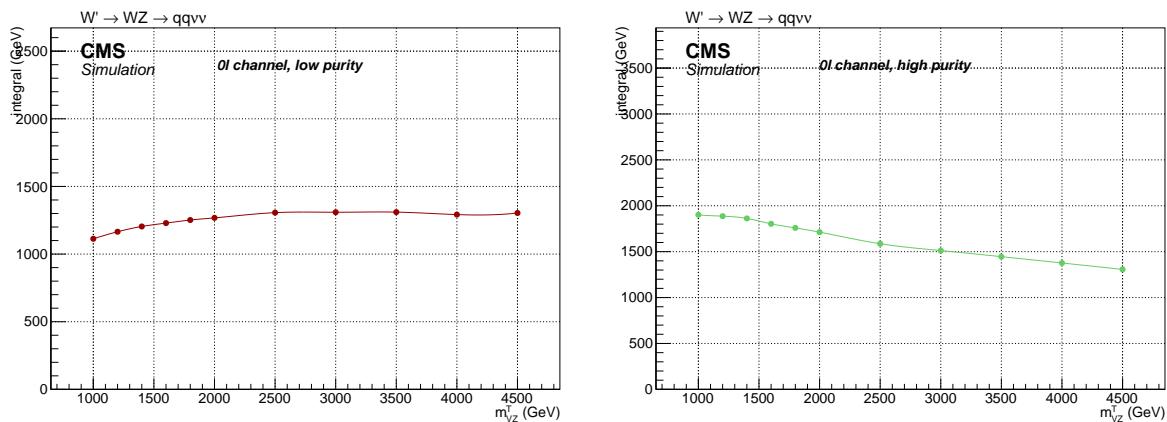


Figure 4.59: Interpolation of the signal normalization as a function of the resonance transverse mass m_{VZ}^T , for a spin-1 (W') signal hypothesis. From left to right: low-purity, high-purity.

741 **4.6 Systematic uncertainties**

742 The background and signal predictions are affected by systematic uncertainties that have to be es-
743 timated and taken into account in the signal extraction procedure. This section includes a list of the
744 relevant systematic uncertainties for this analysis and how they are estimated.

745 **4.6.1 Uncertainties affecting the data-driven main background estimation**

746 **4.6.1.1 Normalization**

747 The prediction of the normalization and shape of main background, $V + \text{jets}$, are both taken from
748 data. The normalization is extracted from fits to the jet mass sidebands with arbitrary functions
749 tested on simulation. The effects related to the contribution of the sub-dominant backgrounds are
750 also taken into account, for both the normalization and the shape.

751 The uncertainties on the sub-dominant backgrounds normalization, namely the uncertainties on
752 the parameters describing the jet mass spectra obtained with the fits performed on simulations,
753 are propagated to the main background yield prediction. An additional uncertainty on the main
754 background yield comes from the fit with the alternative function. In this case, the difference in
755 the predicted number of events due to the function choice is taken as a systematic uncertainty.
756 The limited number of events in data in the sidebands is treated separately as a source of statistical
757 uncertainty. Numerical values are reported in tab. 4.16.

758 **4.6.1.2 Shape**

759 The shape uncertainties on the main background are determined with the α method, discussed in
760 sec. 4.5.2. The uncertainties on the parameters of the main background prediction in the signal
761 regions are affected by the parameter uncertainties of the fit to m_{VZ}^T in data in the jet mass side-
762 bands, and by the parameter uncertainties of the two components of the α function (numerator
763 and denominator), that are the m_{VZ}^T fits to the simulated $V + \text{jets}$ distributions in SR and SB. These
764 uncertainties are propagated to the shape of the main background in the signal region. Before being
765 provided to the likelihood fit, these parameters are decorrelated through a linear transformation.

766 **4.6.2 Uncertainties affecting the signal and the sub-dominant backgrounds**

767 **4.6.2.1 Trigger uncertainty**

768 Trigger uncertainty is evaluated shifting by one standard deviation (*i.e.* 1%, as discussed in sec. 4.2.6)
769 the E_T^{miss} trigger efficiency calculated on data, that is applied as per-event weight to MC samples.
770 The impact has been studied in signal and secondary background samples: it amounts to 0.7–0.5%
771 for signal samples, depending on the mass hypothesis, whilst it affects by 1% the top and diboson
772 normalization. No effect can be appreciated in signal and background shapes.

773 **4.6.2.2 Jet momentum uncertainties**

774 Jet uncertainties are evaluated in the signal regions by moving up and down by one standard devia-
775 tion the source of the uncertainty. The two sources are the uncertainty on the jet energy correction,
776 also identified as jet energy scale (JES) [21], [11], and the uncertainty due to the different jet mo-
777 mentum resolution (JER) [11].

778 Considering the jet energy scale, the transverse momenta of the jets are shifted by the uncertainty
779 value of the corresponding jet energy correction. The impact on the normalization due to the jet

4.6 Systematic uncertainties

- 780 energy correction is evaluated in the signal region, by taking into account its effect on jets and on
 781 E_T^{miss} simultaneously, in a correlated fashion.
 782 The JER effect is evaluated (together with its impact on E_T^{miss}) by smearing the jet p_T by the η -
 783 dependent coefficients listed in tab. 4.8, up and down by one standard deviation, using the hybrid-
 784 method (sec. 4.3.6).
 785 The impact of JEC uncertainties is evaluated also on the signal and background shapes. The result-
 786 ing normalization and shape uncertainties are reported in sec. 4.6.2.6.

787 **4.6.2.3 Jet mass uncertainties**

- 788 The soft drop PUPPI corrected jet mass is affected by two different uncertainties sources.
 789 Soft drop jet mass calibration is varied within $\pm \sqrt{(\text{JES}_{\text{unc.}}^2 + \text{JMS}_{\text{unc.}}^2)}$, where $\text{JES}_{\text{unc.}}$ is the uncertainty
 790 of the JES, described above, and $\text{JMS}_{\text{unc.}} = 0.0094$ is a constant coefficient (4.3.7, [21], [11]). The
 791 impact is calculated on signal and secondary backgrounds, both in normalization and shape.
 792 As regarding the smearing, the soft drop PUPPI corrected jet mass of the signal samples and sub-
 793 dominant backgrounds has been smeared up or down by a smearing coefficient (described in sec. 4.3.7),
 that is $\text{JMR} = 1.00 \pm 0.20$.

Table 4.19: Summary of jet mass energy corrections systematic uncertainties (JMS). The symbol Δ indicates the variation for each variable, due to the considered systematic uncertainty shift.

m_{VZ}^T	1 TeV	4 TeV
Δ events	1.0%	1.0%
Δ mean	0.1%	0.1%
Δ RMS	<0.1%	0.4%
secondary background	VV	Top
Δ events	0.1%	0.7%
Δ slope	<0.1%	0.2%

794

Table 4.20: Summary of et mass resolution corrections systematic uncertainties (JMR). The symbol Δ indicates the variation for each variable, due to the considered systematic uncertainty shift.

m_{VZ}^T	1 TeV	4 TeV
Δ events	5.2%	4.9%
Δ mean	0.1%	0.1%
Δ RMS	0.4%	0.3%
secondary background	VV	Top
Δ events	2.0%	3.1%
Δ slope	1.0%	4.0%

- 795 Results are presented in detail in tab. 4.19-4.20, for JMS and JMR uncertainties. Shape uncertainties
 796 on signal are evaluated as the variation in the mean and variance of the transverse mass distribution.
 797 Shape uncertainties on top and diboson backgrounds are quoted as the relative variation in the
 798 slope of the exponential falling distribution of m_{VZ}^T , and their effects are shown in fig. 4.60-4.61.

799 **4.6.2.4 V-tagging uncertainties**

- 800 Data-Monte Carlo V-tagging scale factors are applied to the signal and secondary background yields
 801 (sec. 4.3.8.1), and their uncertainty is taken as systematic. The contribution of the uncertainty is 11%

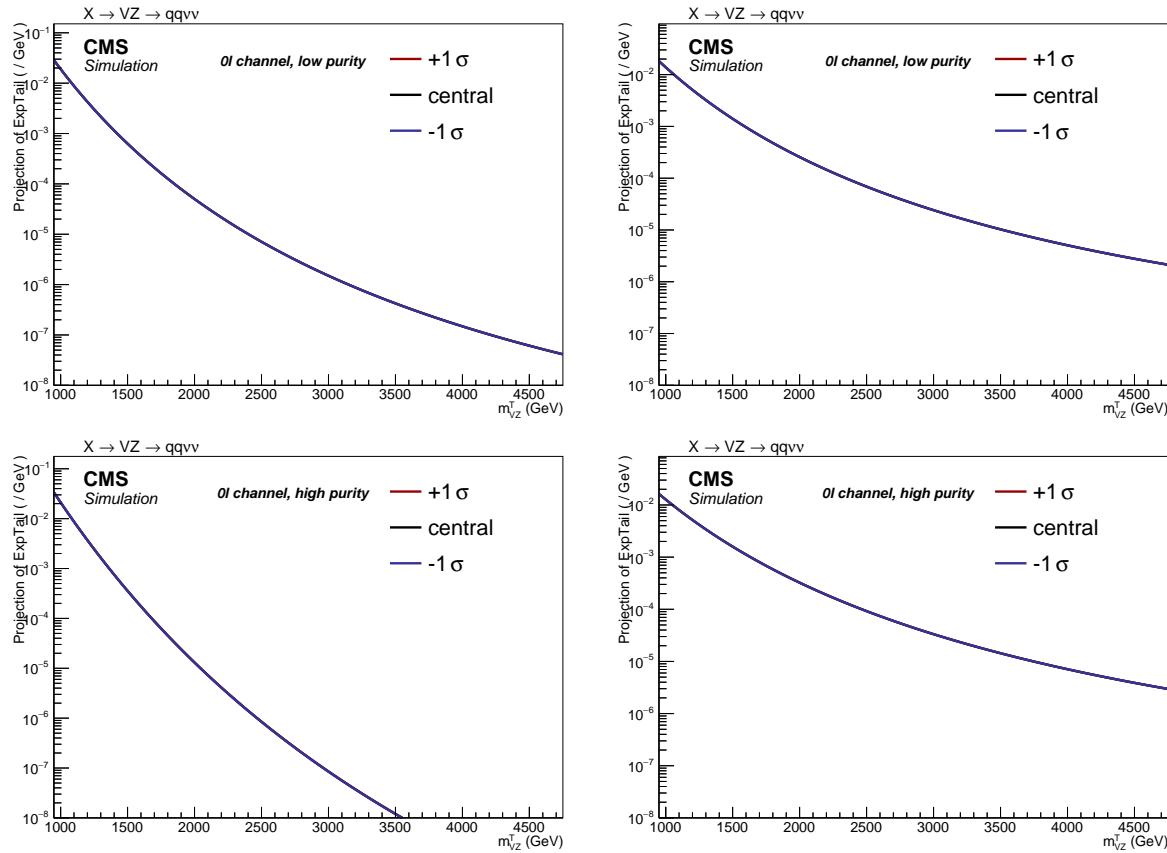


Figure 4.60: Shape variations due to jet mass calibration corrections obtained in the Top (left) and diboson (right) backgrounds, in the low-purity (top) and high-purity (bottom) category.

for the high-purity and 23% for low-purity category, applied on signal and secondary backgrounds. While combining the categories, V -tagging uncertainties are considered as anti-correlated. The V -tagging scale factors are measured in $t\bar{t}$ samples, hence at p_T values generally not larger than 200–300 GeV. An uncertainty due to the V -tagging extrapolation at higher momenta is considered by using an alternative showering scheme (HERWIG [22]). It is parametrized as a function of the jet p_T : $X \times \log(p_T/200\text{GeV})$, where $X = 0.085$ for the high-purity category and $X = 0.039$ for low-purity category. It amounts to 9–20%, depending on the mass of the signal sample considered, to 2–3% for VV and Top backgrounds in high-low purity category. While combining the categories, V -tagging extrapolation uncertainties are considered as correlated.

4.6.2.5 b-tagging uncertainties

The assigned b-tagging uncertainty, related to the b-tag veto applied to AK4 jets that lie outside the V jet cone, with the aim of suppressing the top quark induced background, is the relative difference in shape and normalization, calculated in signal and secondary background events, obtained by shifting up or down the event weight through the envelope of the data-MC b-tagging scale factors uncertainties [17]. The impact of this systematic uncertainty on signal normalization ranges from 0.7% at 1 TeV, up to 1.0% at 4 TeV. The impact on VV background normalization is 0.3%, whilst on Top it is 2.2%. Effects on signal and background shapes are negligible.

4.6 Systematic uncertainties

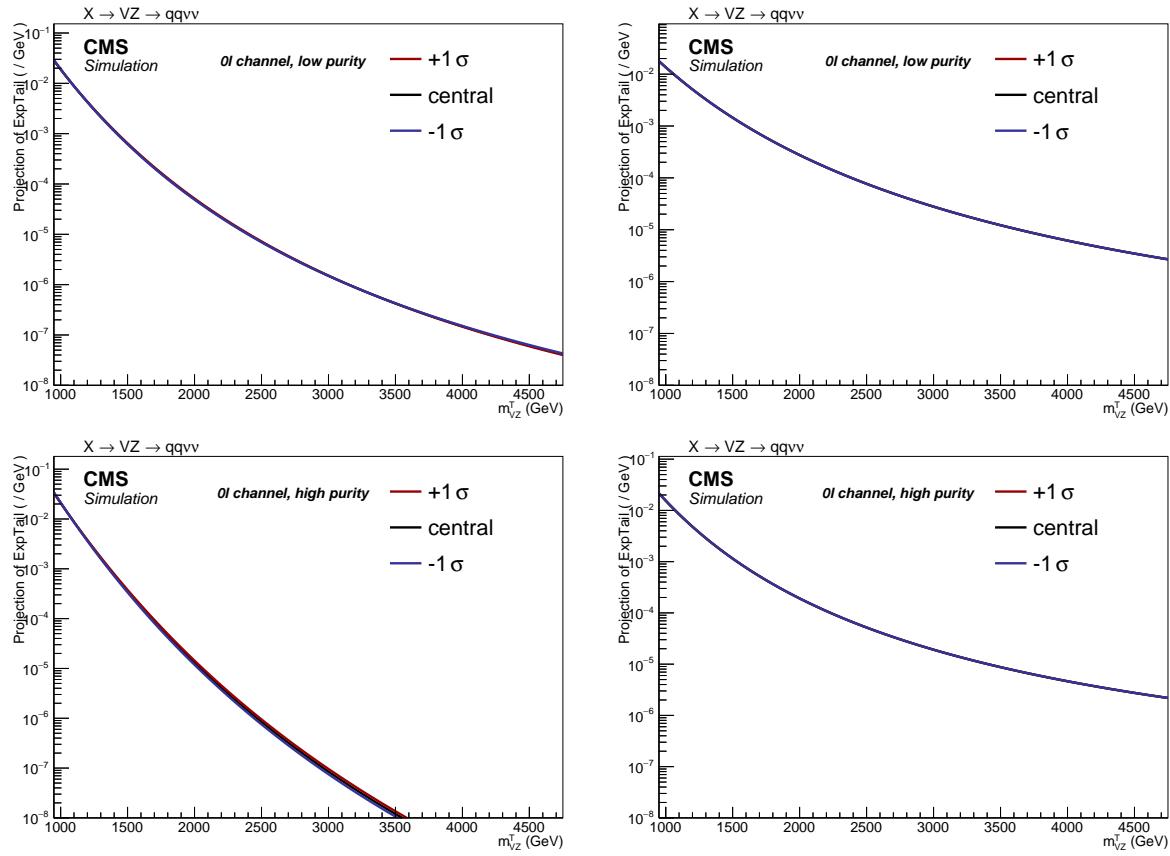


Figure 4.61: Shape variations due to jet mass resolution obtained in the Top (left) and diboson (right) backgrounds, in the low-purity (top) and high-purity (bottom) category.

4.6.2.6 Missing Energy uncertainties

- As described in sec. ??, the E_T^{miss} evaluation depends on all the reconstructed particles in the event, and on their uncertainties. Missing energy uncertainties are calculated by factorizing \vec{p}_T^{miss} in components: electrons, photons, muons, taus, jets and unclustered energy. Dedicated uncertainties are derived by propagating the original object scales and resolutions to the E_T^{miss} itself.
- In this analysis, a leptonic veto is applied, hence the E_T^{miss} uncertainties are due to jets and unclustered energy. The effect of JES is evaluated on E_T^{miss} in a correlated way with jets, by scaling up or down the central value of JES by one sigma, both on E_T^{miss} and on jets p_T . The result is a negligible uncertainty on signal normalization, 0.2% and less than 0.1% uncertainty on top and diboson normalizations, negligible impact on signal, top and diboson shapes.
- The same procedure applies for the uncertainties related to jet JER, that are varied up and down by one sigma in both jets and \vec{p}_T^{miss} at the same time. The result is a negligible uncertainty on signal and diboson normalizations, 0.3% uncertainty on top normalization, and negligible effects on signal and background shapes.
- The last contribution in E_T^{miss} uncertainty is related to unclustered energy, whose impact is evaluated scaling up or down the central value by its own resolution, depending on the particle type. The uncertainty is negligible on signal and background normalizations and shape.

837 4.6.2.7 Pile-up uncertainty

838 An additional source of systematic uncertainty is the limited knowledge of the total proton-proton
 839 inelastic cross-section at 13 TeV, used to get the expected number of vertices distribution for the
 840 pile-up reweighting procedure. A 4.6% uncertainty is assumed for the default value of 69 200 mb,
 841 and the vertices distributions are varied accordingly (fig. 4.62). Changing the pile-up weight varies
 842 also the MC normalizations in the signal region, and the relative difference is estimated to be 0.2%
 843 for the diboson background, 0.3% for top processes, and 0.4-0.7% for signal samples. Pile-up im-
 844 pacts on signal shapes are negligible, and it affects by 0.8% and 0.4% the diboson and top shapes
 845 (fig. 4.63).

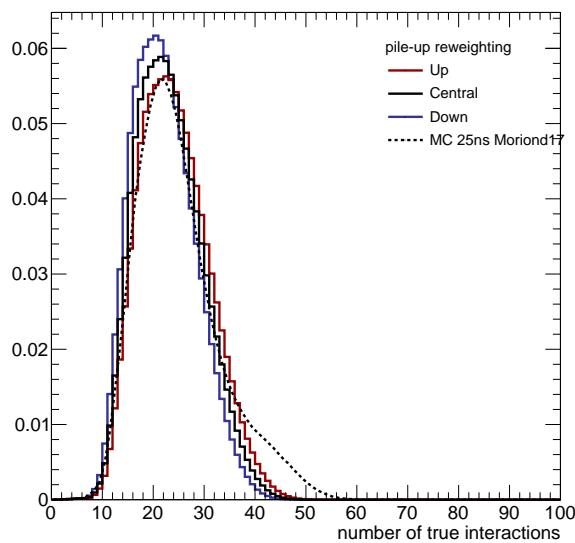


Figure 4.62: Pile-up scenario in 2016 data (black curve), and scenarios obtained by shifting up (red curve) or down (blue curve) the central value of the total inelastic cross-section (69 200 mb), compared to pile-up distribution simulated in Monte Carlo samples (dotted curve).

846 4.6.2.8 QCD renormalization and factorization scale uncertainties

847 Divergencies appearing in perturbative QCD calculation, used to predict the cross-sections and the
 848 spectra of the observables in Monte Carlo simulations, are absorbed in the renormalization and
 849 factorization scales, μ_R and μ_F . Per-event weights are calculated for a variation of these scales by
 850 a factor 2. The two scales can be varied separately and independently, or together assuming 100%
 851 correlation; the first approach is adopted. The weight is propagated up to the final distributions,
 852 accounting for normalization and shape uncertainties.

853 The QCD variations have negligible effect on signal acceptance and on the mean and sigma of the
 854 gaussian core of the Crystal Ball functions. The QCD factorization has an impact on top background
 855 shape (1.1%) and normalization (3.1%), and on diboson normalization (0.9%). The QCD renorm-
 856 alization affects the top normalization (7.3%) and diboson normalization (1.3%).

4.6 Systematic uncertainties

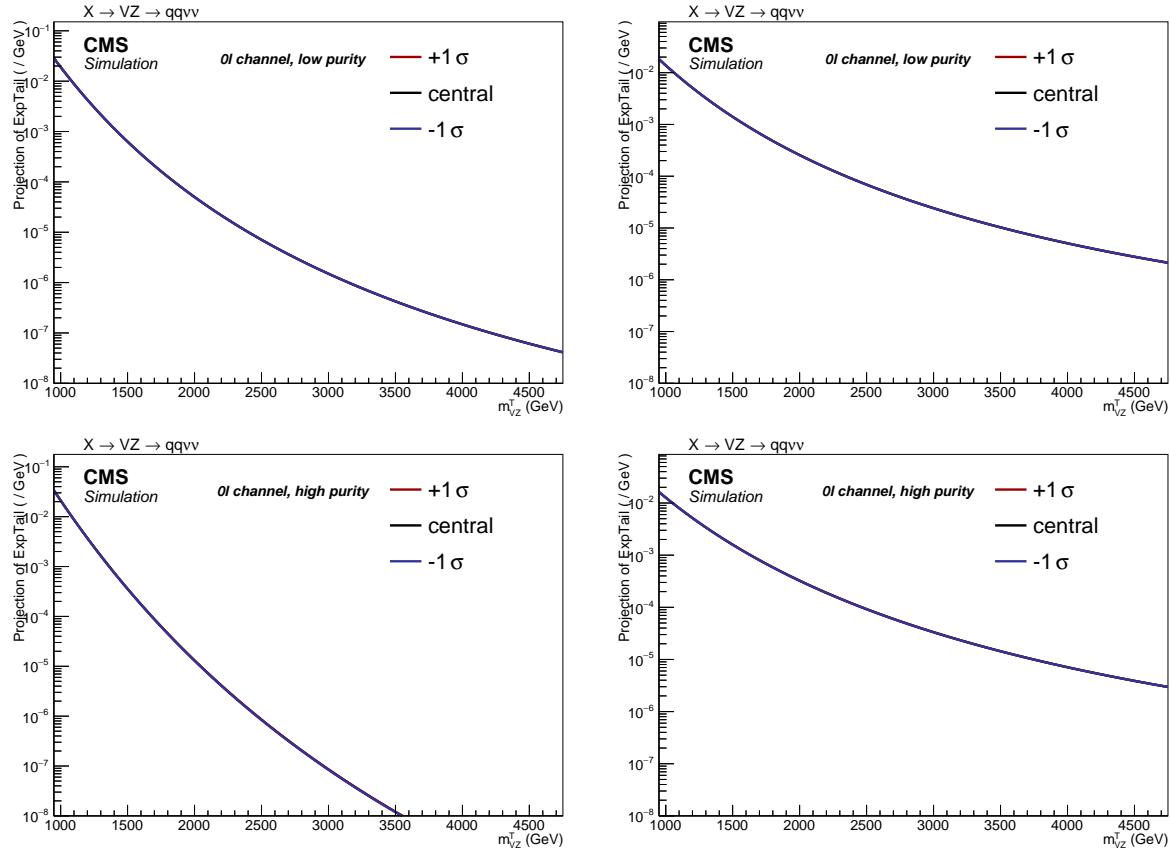


Figure 4.63: Shape variations due to pile-up uncertainty obtained in the Top (left) and diboson (right) backgrounds, in the low-purity (top) and high-purity (bottom) category.

4.6.2.9 PDF

Systematic uncertainties related to the PDFs parameters are estimated according to the PDF4LHC prescriptions [23], and using the NNPDF3.1 [24] set. Each parameter describing the PDFs is varied within its uncertainty, resulting in a set of per-event weights. The 100 shifted weights have been considered together, by calculating the effect of their envelope, compared to their central values, on the expected event yield and on the m_{VZ}^T distributions, and propagated as a normalization or shape uncertainty. The effect of the PDF uncertainty on the signal acceptance is found to be negligible, and it amounts to 10.3% for top background normalization and 2.1% for diboson background normalization. PDF uncertainties affect top background shape by 1.2%.

4.6.3 Summary

A summary of all the systematic uncertainties is listed in tab. 4.21. In addition to those described in the previous sections, an uncertainty of 10% on top background normalization is assumed, that is the uncertainty on the top production cross-sections obtained from CMS measurements (sec. 4.2.3), and an uncertainty of 15% is assigned to the diboson background normalization, due to the uncertainty on the cross-section measurements performed by CMS. An additional 3% covers the uncertainty related to the tau veto, and an uncertainty of 2.5% is assigned to the data integrated luminosity [10].

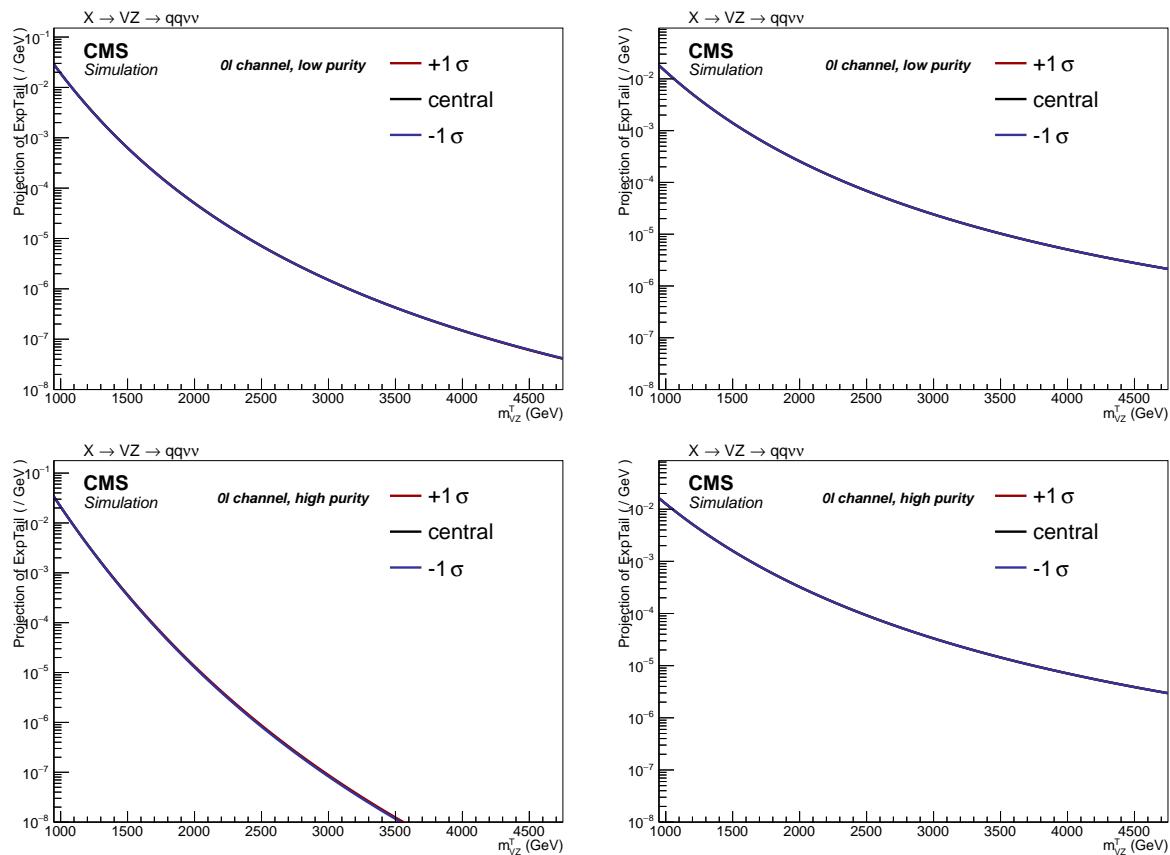


Figure 4.64: Shape variations due to QCD factorization in the Top (left) and diboson (right) backgrounds, in the low-purity (top) and high-purity (bottom) category.

4.6 Systematic uncertainties

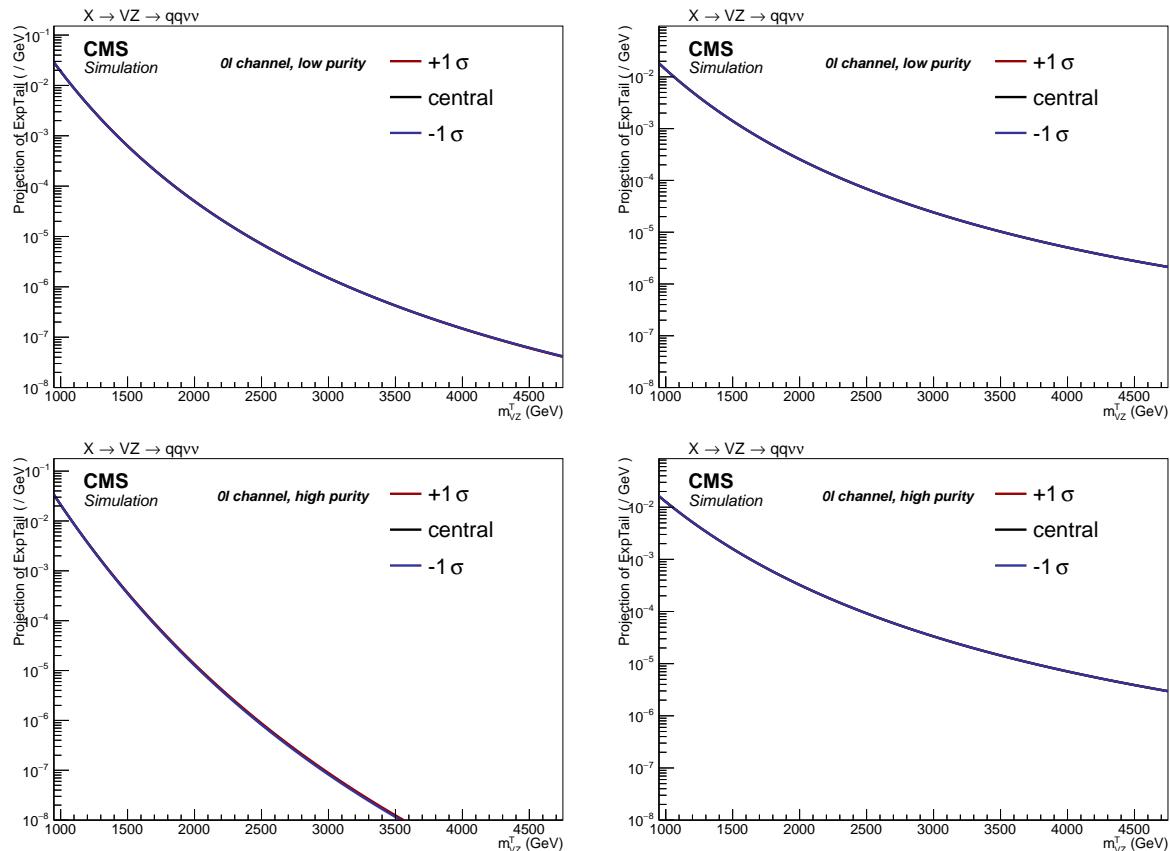


Figure 4.65: Shape variations due to PDF scale in the Top (left) and diboson (right) backgrounds, in the low-purity (top) and high-purity (bottom) category.

Table 4.21: Summary of the systematic uncertainties for the backgrounds and signal samples. LP and HP indicate the uncertainty assigned for each purity category, low- and high-purity, respectively.

	shape	$V + \text{jets}$	Top	VV	Signal
α -function	✓	✓	-	-	-
Bkg. normalization (fit)		4.8%(LP) 14.7%(HP)	68.2%(LP) 47.7%(HP)	11.4%(LP) 19.1%(HP)	-
Bkg. normalization (alternative function)		4.9%(LP) 4.4%(HP)	- -	- -	-
jet energy scale	-	-	0.2%	0.1%	<0.1%
jet energy resolution	-	-	0.3%	<0.1%	<0.1%
unclustered energy	-	-	<0.1%	<0.1%	<0.1%
jet mass scale	✓	-	0.7%	0.1%	1.8%
jet mass resolution	✓	-	3.1%	2.0%	5.1%
trigger	-	-	1.0%	0.9%	0.7-0.5%
V boson tagging (τ_{21})	-	-		11% (HP), 23% (LP)	
V tagging extrapolation	-	-	1.4% (LP) 2.8% (HP)	1.7% (LP) 3.3% (HP)	3.2-9.4% (LP) 6.9-20.6% (HP)
b-tag veto	-	-	2.2%	0.3%	0.7-1.0%
pile-up	✓	-	0.3%	0.2%	0.4-0.7%
QCD renormalization	✓	-	7.3%	1.3%	<0.1%
QCD factorization	✓	-	3.1%	0.9%	<0.1%
PDF	✓	-	10.3%	2.1%	10.4-18.9% (scale)
luminosity	-	-	2.5%	2.5%	2.5%
cross section	-	-	10%	15%	-
tau veto	-	-	3%	3%	3%

874

Chapter

5

Conclusions

875

876

877 Plot di Clemens

878

Bibliography

879

880

- 881 [1] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential*
882 *cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079,
883 [[1405.0301](#)].
- 884 [2] Torbjorn Sjostrand and Stephen Mrenna and Peter Skands, *A brief introduction to pythia 8.1*,
885 *Comput.Phys.Commun.* **178** (2008) 852–867, [[0710.3820](#)].
- 886 [3] CMS Collaboration, CMS Collaboration, *Event generator tunes obtained from underlying event*
887 *and multiparton scattering measurements*, *Eur. Phys. J. C* (2016) 155, [[1512.00815](#)].
- 888 [4] S. Agostinelli, J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Arce et al., *Geant4a simulation*
889 *toolkit*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators,*
890 *Spectrometers, Detectors and Associated Equipment* **506** (2003) 250 – 303.
- 891 [5] CMS Collaboration, V. Khachatryan et al., *Identification techniques for highly boosted W*
892 *bosons that decay into hadrons*, *JHEP* **12** (2014) 017, [[1410.4227](#)].
- 893 [6] Rikkert Frederix and Stefano Frixione, *Merging meets matching in mc@nlo*, *JHEP* **12** (2012)
894 061, [[1209.6215](#)].
- 895 [7] Emanuele Re, *Single-top wt-channel production matched with parton showers using the*
896 *powheg method*, *Eur.Phys.J. C* **71** (2011) 1547, [[1009.2450](#)].
- 897 [8] John M. Campbell, R. Keith Ellis, Paolo Nason, Emanuele Re, *Top-pair production and decay*
898 *at nlo matched with parton showers*, *JHEP* **04** (2015) 114, [[1412.1828](#)].
- 899 [9] S. Kallweit, J. M. Lindert, S. Pozzorini, M. Schönherr and P. Maierhöfer, *NLO QCD+EW*
900 *automation and precise predictions for V+multijet production*, in *Proceedings, 50th Rencontres*
901 *de Moriond, QCD and high energy interactions*, pp. 121–124, 2015, [1505.05704](#),
902 <http://inspirehep.net/record/1372103/files/arXiv:1505.05704.pdf>.
- 903 [10] CMS COLLABORATION Collaboration, *CMS Luminosity Measurements for the 2016 Data Taking*
904 *Period*, CMS Physics Analysis Summary CMS-PAS-LUM-17-001, CERN, Geneva, 2017.
905 <https://cds.cern.ch/record/2257069>.

- 906 [11] CMS Collaboration, CMS Collaboration, *Jet energy scale and resolution performances with*
907 *13TeV data*, CMS Performance Note CMS-DP-2016-020, CERN, Geneva, Jun, 2016.
908 <http://cds.cern.ch/record/2160347>.
- 909 [12] A. J. Larkoski, S. Marzani, G. Soyez and J. Thaler, *Soft Drop*, *JHEP* **05** (2014) 146, [1402.2657].
- 910 [13] Y. L. Dokshitzer, G. D. Leder, S. Moretti and B. R. Webber, *Better jet clustering algorithms*, *JHEP*
911 **08** (1997) 001, [hep-ph/9707323].
- 912 [14] D. Bertolini, P. Harris, M. Low and N. Tran, *Pileup per particle identification*, *Journal of High*
913 *Energy Physics* **2014** (2014) 59.
- 914 [15] J. Thaler and K. Van Tilburg, *Identifying boosted objects with n-subjettiness*, *Journal of High*
915 *Energy Physics* **2011** (2011) 15.
- 916 [16] CMS Collaboration, CMS Collaboration, *Identification of b-quark jets with the CMS*
917 *experiment*, *JINST* **8** (2013) P04013, [1211.4462].
- 918 [17] CMS Collaboration, *Performance of b-Tagging Algorithms in Proton Collisions at 13 TeV using*
919 *the 2016 Data*, Physics Analysis Summary CMS-DP-2016-042, CERN, Geneva, Jul, 2016.
920 <https://cds.cern.ch/record/2202967>.
- 921 [18] CMS Collaboration, V. Khachatryan et al., *Search for a higgs boson in the mass range from 145*
922 *to 1000 GeV decaying to a pair of W or Z bosons*, *JHEP* **10** (2015) 144, [1504.00936].
- 923 [19] M. Oreglia, *A Study of the Reactions $\psi' \rightarrow \gamma\gamma\psi$* , Ph.D. thesis, SLAC, 1980.
- 924 [20] T. Skwarnicki, *A study of the radiative CASCADE transitions between the Upsilon-Prime and*
925 *Upsilon resonances*, Ph.D. thesis, Cracow, INP, 1986.
- 926 [21] CMS Collaboration, CMS Collaboration, *Determination of jet energy calibration and*
927 *transverse momentum resolution in cms*, *Journal of Instrumentation* **6** (2011) P11002.
- 928 [22] M. Bahr, S. Gieseke, M. A. Gigg, D. Grellscheid, K. Hamilton, O. Latunde-Dada, S. Platzer, P.
929 Richardson, M. H. Seymour, A. Sherstnev, J. Tully, B. R. Webber, *Herwig++ physics and*
930 *manual*, *Eur.Phys.J.* **C58** (2008) 639–707, [0803.0883].
- 931 [23] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, *J. Phys.* **G43** (2016) 023001,
932 [1510.03865].
- 933 [24] NNPDF Collaboration, R. D. Ball et al., *Parton distributions from high-precision collider data*,
934 *Eur. Phys. J.* **C77** (2017) 663, [1706.00428].