Sede Amministrativa: Università degli studi di Padova
Dipartimento di Fisica e Astronomia

Corso di Dottorato di Ricerca in Fisica

XXX ciclo

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

Coordinatore: Ch.mo Prof. Gianguido Dall'Agata
Supervisore: Ch.mo Prof. Franco Simonetto

Dottoranda: Lisa Benato

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

(A. Einstein)

Contents

Abstract

This thesis presents a search for potential signals of new heavy resonances decaying into a pair of vector bosons, with masses between 1 TeV and 4 TeV, predicted by beyond standard model theories. The signals probed are spin-1 W', predicted by the Heavy Vector Triplet model, and spin-2 bulk gravitons, predicted by warped extra-dimension models. The scrutinized data are produced by LHC proton-proton collisions at a center-of-mass energy $\sqrt{s} = 13$ TeV during the 2016 operations, and collected by the CMS experiment, corresponding to an integrated luminosity of 35.9 fb^{-1}. One of the boson should be a Z, and it is identified through its invisible decay into neutrinos, while the other electroweak boson, consisting either into a W or into a Z boson, is required to decay hadronically into a pair of quarks. The decay products of heavy 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. The couple of neutrinos, escaping undetected, is reconstructed as missing momentum in the transverse plane of the CMS detector. The couple of quarks is reconstructed as one large-cone jet, with high transverse momentum, recoiling against the couple of neutrinos. Grooming algorithms are adopted in order to improve the jet mass resolution, by removing soft radiation components and spectator events from the particles clustered as the large-cone jet. The groomed jet mass is used to tag the hadronically decaying vector boson, to define the signal region of the search (close to the nominal mass of the W and Z bosons, between 65-105 GeV) and a signal-depleted control region, that is used for the background estimation. An hybrid data-simulation approach predicts the normalization and the shape of the main background, represented by a vector boson produced in association with jets, by taking advantage of the distribution of data in the signal-depleted control regions. Secondary backgrounds are predicted from simulations. Jet substructure techniques are exploited, in order to classify events into two exclusive purity categories, by distinguishing the couple of quarks inside the large-cone jet. This approach improves the background rejection and the discovery reach. The search is performed by scanning the distribution of the reconstructed mass of the resonance, looking for a local excess in data with regards to the prediction. Depending on the mass, upper limits on the cross-section of heavy spin-1 and spin-2 narrow resonances, multiplied by the branching fraction of the resonance decaying into Z and a W boson for a spin-1 signal, and into a pair of Z bosons for spin-2, are set in the range 0.9 – 63 fb and in the range 0.5 – 40 fb respectively. A W' hypothesis is excluded up to 3.11 TeV, in the Heavy Vector Triplet benchmark A scenario, and up to 3.41 TeV, considering the benchmark B scenario. A bulk graviton hypotesis, given the curvature parameter of the extra-dimension $\tilde{k} = 1.0$, is excluded up to 1.14 TeV.

Riassunto

Questa tesi presenta una ricerca di potenziali segnali di nuove risonanze pesanti, che decadono in una coppia di bosoni vettori, con masse comprese tra 1 TeV e 4 TeV, predette da teorie oltre il modello standard. I segnali indagati sono W' di spin 1, predette dal modello Heavy Vector Triplet, e gravitoni di spin 2, predetti da modelli che prevedono extra dimensioni ripiegate. I dati esaminati sono prodotti dalle collisioni protone-protone di LHC ad un'energia del centro di massa di $\sqrt{s} = 13$ TeV durante le operazioni del 2016, e raccolti dall'esperimento CMS, per una luminosità integrata di 35.9 fb^{-1}. Uno dei bosoni dev'essere una Z, che viene identificata dal suo decadimento invisibile in neutrini, mentre l'altro bosone elettrodebole, sia una W che una Z, deve decadere nel canale adronico in una coppia di quark. I prodotti di decadimento di risonanze pesanti sono generati con significativi boost di Lorentz; di conseguenza, ci si aspetta che i prodotti di decadimento dei bosoni (i quark e i neutrini) abbiano elevate energie e siano collimati. La coppia di neutrini, che sfugge alla rivelazione, viene ricostruita come momento mancante nel piano trasverso del rivelatore CMS. La coppia di quark viene ricostruita come un jet a largo cono, con elevato momento trasverso, che rincula contro la coppia di neutrini. Algoritmi di grooming sono impiegati per migliorare la risoluzione della massa del jet, rimuovendo la radiazione soffice e gli eventi spettatori dalle particelle clusterizzate come jet a largo cono. La massa ripulita del jet viene utilizzata per identificare il bosone vettore che decade in adroni, per definire la regione di segnale della ricerca (vicina alla massa nominale dei bosoni W e Z, nell'intervallo 65-105 GeV) e una regione di controllo svuotata dal segnale, che viene utilizzata per la stima dei fondi. Un approccio ibdrido dati-simulazione predice la normalizzazione e la forma del fondo principale, rappresentato da un bosone vettore prodotto in associazione con jet, sfruttando la distribuzione dei dati nelle regioni di controllo svuotate dal segnale. I fondi secondari sono predetti completamente con le simulazioni. Tecniche di sottostruttura del jet sono adoperate per classificare gli eventi in due categorie esclusive di purezza, distinguendo le coppie di quark dentro al jet a largo cono. Questo approccio migliora la soppressione del fondo e la potenzialità di scoperta. La ricerca viene fatta scansionando la distribuzione della massa ricostruita della risonanza, cercando un eccesso locale nei dati rispetto alle predizioni. In funzione della massa, limiti superiori sulla sezione d'urto per risonanze pesanti e strette di spin 1 e spin 2, moltiplicate per il rapporto di diramazione della risonanza che decade in Z e W per il segnale di spin 1, e in una coppia di bosoni Z per lo spin 2, sono fissati nell'intervallo 0.9 – 63 fb e nell'intervallo 0.5 – 40 fb rispettivamente. Un'ipotesi di W' è esclusa fino ad una massa di 3.11 TeV, nello scenario A di riferimento dell'Heavy Vector Triplet, e fino a 3.41 TeV, nello scenario B di riferimento. Un'ipotesi di gravitone, dato il parametro di curvatura della dimensione addizionale $\tilde{k} = 1.0$, è esclusa fino ad una massa di 1.14 TeV.

Chapter 1

Introduction

The discovery of the Higgs boson at the CERN Large Hadron Collider represents a milestone in the knowledge of the particle physics. The Higgs mechanism connects the theoretical formulation of the standard model of the particles to the current picture of the universe, as it is known: spin-1 weak bosons and standard model fermions are allowed to acquire masses, constituting the fundamental bricks of the known matter. Despite this successful achievement, some questions are still left unanswered; in order to solve the open problems, a plethora of new beyond standard model theories have been built.
Many of these theories hypothesize the existence of larger symmetries in the universe, or new extra-dimensions, that will result into the appearance of new heavy particles, expected to have a mass around 1 TeV. The Large Hadron Collider (LHC) is the ideal tool to investigate this unknown phase-space, given the fact that during the so-called LHC Run 2 era (started in 2015), the unprecedented center-of-mass energy of 13 TeV has been reached in the proton-proton collisions.
The CMS experiment, located in the northern part of the LHC ring, is a multi purpose detector, suitable to study highly energetic new phenomena. Its intense magnetic field, its sharp segmentation, its hermeticity and the interplay of many sophisticated reconstruction algorithms allow to measure with a very high precision the trajectories, the momenta and the energy deposits left by energetic particles.

This thesis presents a search for signals of heavy resonances that decay into a pair of vector bosons. The search is performed by using the 2016 data produced by proton-proton collisions of the LHC, and collected by the CMS detector. One Z boson is identified through its invisible decay in neutrinos, while the other vector boson is required to decay hadronically into a pair of quarks. Given the fact that the searched resonances have masses around the TeV, their decay products are expected to be produced with large Lorentz boosts. This leads to a non-trivial identification of the couple of quarks or leptons, coming from the vector bosons decays. In fact, they are expected to lie very close in angle. Dedicated algorithms and substructure techniques allow to distinguish a pair of quarks originating from a vector boson from the background processes, initiated by the strong interaction. The search is performed by scanning the distribution of the reconstructed mass of the resonance, looking for a local excess in data with regards to the predictions. The background estimation is performed with an hybrid data-simulation approach, by using the distribution of data in signal-depleted control regions.

The thesis is organized as follows.
In chapter 2, an overview of the theoretical motivations is presented. Two beyond standard model theories are considered: the Heavy Vector Triplet Model and the bulk graviton model, predicted by warped extra-dimensions theories.
In chapter 3, the CMS detector is briefly described, along with the physics objects exploited for the purpose of this search.
Chapter 4 is dedicated to the analysis: after a general introduction (sec. 4.1), the features of the data, signal and background samples used in the analysis are described in detail (sec. 4.2). Sec. 4.3 is dedicated to the selections applied, in order to reach the best signal-to-noise efficiency and to properly build the resonance candidate. The very first data-simulation comparison is performed in sec. 4.4. The background estimation technique, the final data-predicted background comparison and the signal modelling are included in sec. 4.5. Systematic uncertainties are listed in sec. 4.6. The final results, the statistical analysis and the physics interpretation are shown in sec. 4.7. Chapter 5 summarizes the conclusions.

Chapter 2

Theoretical motivation

The standard model (SM) of elementary particles represents, so far, the best available description of the particles and their interactions. It is the summation of two gauge theories: the electroweak interaction, that portrays the weak and electromagnetic interactions together, and the strong interaction, or Quantum Chromodynamics (QCD). Particles, namely quarks and leptons, are described as spin 1/2 fermions, whilst interactions are mediated by spin 1 bosons. The symmetry group of the standard model is:

$$SU_C(3) \times SU_L(2) \times U_Y(1), \tag{2.1}$$

where the first factor is related to strong interactions, whose mediators are eight gluons, while $SU_L(2) \times U_Y(1)$ is the electroweak simmetry group, whose mediators are the photon and Z-$W^{\pm}$ bosons.
In renormalizable theories, with no anomalies, all gauge bosons are expected to be massless, in contrast with our experimental knowledge. This inconsistency is solved by introducing a new scalar particle, the Higgs boson, that gives mass to weak bosons and fermions via the spontaneous symmetry breaking mechanism.

In the last decades, the standard model has been accurately probed by many experimental facilities (LEP, Tevatron. LHC), and the results lead to an impressive agreement between theoretical predictions and experiments [1]. The discovery of the Higgs boson at the CERN Large Hadron Collider, measured by both CMS and ATLAS collaborations [2–8], represents not only an extraordinary confirmation of the model, but also the latest biggest achievement in particle physics as a whole.

2.1 Beyond standard model theories

Even though the SM is the most complete picture of the universe of the particles, many questions are still left open. From a phenomenological point of view, some experimental observations are not included in the theory:

- in SM, neutrinos are massless (whilst the well established observation of neutrino flavour oscillation proves that neutrinos do carry mass);
- no candidates for dark matter are predicted;

- no one of the fields included in the SM can explain the cosmological inflation;
- the CP asymmetry embedded in the SM is not sufficient to explain the matter-antimatter asymmetry in the universe.

From a purely theoretical perspective, some issues are still relevant in the formulation of the model:

- *Flavour problem.*
  The standard model has 18 free parameters: 9 fermionic masses; 3 angular parameters in Cabibbo-Kobayashi-Maskawa matrix, plus 1 phase parameter; electromagnetic coupling α; strong coupling α_{strong}; weak coupling α_{weak}; Z mass; the mass of the Higgs boson. Such a huge number of degrees of freedom marks the SM as weakly predictive in the flavour sector.
- *Unification.*
  There is not a "complete" unification of strong, weak and electromagnetic interactions, since each one has its own coupling constant, behaving differently at different energy scales; not to mention the fact that gravitational interaction is completely excluded from the SM.
- *Hierarchy problem.*
  From Quantum Field Theory, it is known that perturbative corrections to the mass of the scalar bosons included in the theory tend to make it increase towards the energy scale at which the considered theory still holds [9]. If the standard model is seen as a low-mass approximation of a more general theory valid up to the Planck mass scale (*i.e.*, $\sim 1.2 \times 10^{19}$ GeV), a fine-tuning cancellation of the order of 1 over 10^{34} is needed in order to protect the Higgs mass at the electroweak scale ($\sim$ 100 GeV). Such an astonishing correction is perceived as very unnatural.

Numerous beyond standard model theories (BSM) have been proposed in order to overcome the limits of the SM.
Grand Unified Theories (GUT) aim at extending the symmetry group of the SM (eq. 2) into largest candidates, such as $S0(10)$, $SU(5)$ and $E(6)$. At GUT scale, approximately at 10^{16} GeV, non-gravitational interactions are expected to be ruled by only one coupling constant, α_{GUT}.
Super Symmetryc (SUSY) models [10] state that every fermion (boson) of the SM has a bosonic (fermionic) superpartner, with exactly the same quantum numbers, except the spin. If SUSY is not broken, each couple of partners and superpartners should have the same masses, hypothesis excluded by the non-observation of the s-electron. Super Symmetry represents a very elegant solution of the hierarchy problem of the Higgs boson mass, since the perturbative corrections brought by new SUSY particles exactly cancel out the divergences caused by SM particles corrections. A particular sub-class of SUSY models, minimal super symmetric standard models [11–13], is characterized by the introduction of a new symmetry, the R-parity, that guarantees the proton stability and also the stability of the lightest SUSY particle, a possible good candidate for dark matter.

Two other possible theoretical pictures are extensively described in sec. 2.2-2.3.

2.2 Heavy Vector Triplet

The heavy vector triplet model [14] provides a general framework aimed at studying new physics beyond the standard model, that can manifest into the appearance of new resonances.
The adopted approach is that of the simplified model, in which an effective Lagrangian is introduced, in order to describe the properties and interactions of new particles (in this case, a triplet of spin-1 bosons) by using a limited set of parameters, that can be easily linked to the physical observables at the LHC experiments. These parameters can describe many physical motivated theories (such as sequential extensions of the SM [15, 16] or Composite Higgs [17, 18]), built to solve the hierarchy problem of the SM.
Since a simplified model is not a complete theory, its validity is restricted to the on-shell quantities related to the production and decay mechanisms of the new resonances, that is how most of the LHC BSM searches are performed. Given these conditions, experimental results in the resonant region are sensitive to a limited number of the phenomenological Lagrangian parameters (or to a combination of those), whilst the remaining parameters tend to influence the tail of the distributions.
Limits on production cross-section times branching ratio ($\sigma\mathscr{B}$), as a function of the invariant mass spectrum of the probed resonance, can be extracted from experimental data. Given that $\sigma\mathscr{B}$ are functions of the simplified model parameters and of the parton luminosities, it is then possible to interpret the observed limits in the parameter space.

2.2.1 Simplified Lagrangian

The heavy vector triplet framework assumes the existence of an additional vector triplet, V_μ^a, $a = 1, 2, 3$, in which two spin-1 particles are charged and one is neutral:

$$\begin{aligned} V_\mu^\pm &= \frac{V_\mu^1 \mp i V_\mu^2}{\sqrt{2}}; \\ V_\mu^0 &= V_\mu^3. \end{aligned} \tag{2.2}$$

The triplet interactions are described by a simplified Lagrangian, that is invariant under SM gauge and CP symmetry, and accidentally invariant under the custodial symmetry $SU(2)_L \times SU(2)_R$:

$$\begin{aligned} \mathscr{L}_\mathcal{V} = &-\frac{1}{4}\left(D_\mu V_\nu^a - D_\nu V_\mu^a\right)\left(D^\mu V^{\nu\,a} - D^\nu V^{\mu\,a}\right) + \frac{m_V^2}{2} V_\mu^a V^{\mu\,a} \\ &+ i g_V c_H V_\mu^a \left(H^\dagger \tau^a D^\mu H - D^\mu H^\dagger \tau^a H\right) + \frac{g^2}{g_V} c_F V_\mu^a \sum_f \bar{f}_L \gamma^\mu \tau^a f_L \\ &+ \frac{g_V}{2} c_{VVV} \epsilon_{abc} V_\mu^a V_\nu^b \left(D^\mu V^{\nu\,c} - D^\nu V^{\mu\,c}\right) + g_V^2 c_{VVHH} V_\mu^a V^{\mu\,a} H^\dagger H - \frac{g}{2} c_{VVW} \epsilon_{abc} W^{\mu\nu\,a} V_\mu^b V_\nu^c. \end{aligned} \tag{2.3}$$

In the first line of the formula 2.3, V mass and kinematic terms are included, described with the covariant derivative $D_\mu V_\nu^a = \partial_\mu V_\nu^a + g\epsilon^{abc} W_\mu^b V_\nu^c$, where W_μ^a are the fields of the weak interaction and g is the weak gauge coupling. V_μ^a are not mass eigenstates, since they mix with the electroweak fields after the spontaneous symmetry breaking, therefore m_V isn't the physical mass of the V bosons.
The second line describes the interaction of the triplet with the Higgs field and the SM left-handed fermions; c_H describes the vertices with the physical Higgs and the three unphysical Goldstone bosons that, for the Goldstone equivalence theorem, are equivalent to the longitudinal polarization

of W and Z bosons at high-energy; hence, c_H is related to the bosonic decays of the resonances. c_F is the analogous parameter describing the V interaction with fermions, that can be generalized as a flavour dependent coefficent, once defined $J_F^{\mu\ a} = \sum_f \bar{f}_L \gamma^\mu \tau^a f_L$: $c_F V_\mu^a J_F^{\mu\ a} = c_\ell V_\mu^a J_\ell^{\mu\ a} + c_q V_\mu^a J_q^{\mu\ a} + c_3 V_\mu^a J_3^{\mu\ a}$.
The last part of the equation includes terms that are relevant only in strongly coupled scenarios (see sec. 2.2.2.2) through the V-W mixing, but it does not include vertices of V with light SM fields, hence it can be neglected while describing the majority of the LHC phenomenology, under the assumptions previously stated. Additional dimension four quadrilinear V interactions are non relevant for the processes discussed, otherwise their effects would be appreciated in electroweak precision tests and precise Higgs coupling measurements [19].

The parameters in the Lagrangian can be interpreted as follows: g_V describes the strenght of the interaction, that is weighted by c parameters. g_V ranges from $g_V \sim 1$ when the coupling is weak (sec. 2.2.3), to $g_V \sim 4\pi$ when the coupling is strong (sec. 2.2.4). c parameters are expected to be $c \sim 1$, except to c_H, that can be smaller for weak couplings. The combinations describing the vertices, $g_V c_H$ and $g^2/g_V c_F$, can be considered as the fundamental parameters, used to interpret the experimental results.

2.2.2 Mass eigenstates, mixing parameters and decay widths

The newly introduced $SU(2)_L$ triplet is expected to mix with the weak SM fields. The $U(1)_{em}$ symmetry is left unbroken by the new interaction, hence the massless combination of the electroweak fields, namely the photon, is the same as the SM:

$$A_\mu = B_\mu \cos\theta_W + W_\mu^3 \sin\theta_W, \tag{2.4}$$

with the usual definitions of the electroweak parameters:

$$\begin{aligned} \tan\theta_W &= \frac{g'}{g} \\ e &= \frac{g g'}{\sqrt{g^2 + g'^2}} \\ g &= e/\sin\theta_w \\ g' &= e/\cos\theta_w. \end{aligned} \tag{2.5}$$

The Z boson, on the other hand, mixes with the neutral component of the triplet, V^0, with a rotation parametrized with the angle θ_N:

$$\begin{pmatrix} \cos\theta_N & \sin\theta_N \\ -\sin\theta_N & \cos\theta_N d \end{pmatrix} \begin{pmatrix} Z \\ V^0 \end{pmatrix}. \tag{2.6}$$

The mass matrix of the rotated system is given by:

$$\mathbb{M}_N^2 = \begin{pmatrix} \hat{m}_Z^2 & c_H \zeta \hat{m}_Z \hat{m}_V \\ c_H \zeta \hat{m}_Z \hat{m}_V & \hat{m}_V^2 \end{pmatrix}, \tag{2.7}$$

where the parameters are defined as:

$$\begin{cases} \hat{m}_Z = \frac{e}{2\sin\theta_W \cos\theta_W} \hat{v} \\ \hat{m}_V^2 = m_V^2 + g_V^2 c_{VVHH} \hat{v}^2 \\ \zeta = \frac{g_V \hat{v}}{2\hat{m}_V} \\ \frac{\hat{v}^2}{2} = \langle H^\dagger H \rangle \end{cases}, \tag{2.8}$$

and $\hat{v}$, the vacuum expectation value of the Higgs field, can be different from the SM $v = 246$ GeV. The physical masses of Z and V^0, m_Z and M_0, and θ_N come from the matrix relations:

$$\begin{aligned}
&\mathrm{Tr}\left(\mathbb{M}_N^2\right) = \hat{m}_Z^2 + \hat{m}_V^2 = m_Z^2 + M_0^2 \\
&\left\|\mathbb{M}_N^2\right\| = \hat{m}_Z^2 + \hat{m}_V^2\left(1 - c_H^2\zeta^2\right) = m_Z^2 M_0^2 \\
&\tan 2\theta_N = \frac{2c_H\zeta\,\hat{m}_Z\hat{m}_V}{\hat{m}_V^2 - \hat{m}_Z^2}.
\end{aligned} \tag{2.9}$$

The $W^\pm$ bosons mix with the charged components of the triplet, $V^\pm$, leading to a mass matrix analogous to eq. 2.10:

$$\mathbb{M}_C^2 = \begin{pmatrix} \hat{m}_W^2 & c_H\zeta\,\hat{m}_W\hat{m}_V \\ c_H\zeta\,\hat{m}_W\hat{m}_V & \hat{m}_V^2 \end{pmatrix}, \tag{2.10}$$

where $\hat{m}_W$ is defined as:

$$\left\{\ \hat{m}_W = \tfrac{e}{2\sin\theta_W}\,\hat{v} = \hat{m}_Z\cos\theta_W \ \ ; \right. \tag{2.11}$$

the physical masses of W and $V^\pm$, m_W and $M_\pm$, and the angle θ_C parametrizing the rotation of the charged sector are described by:

$$\begin{aligned}
&\mathrm{Tr}\left(\mathbb{M}_C^2\right) = \hat{m}_W^2 + \hat{m}_V^2 = m_W^2 + M_\pm^2 \\
&\left\|\mathbb{M}_C^2\right\| = \hat{m}_W^2 + \hat{m}_V^2\left(1 - c_H^2\zeta^2\right) = m_W^2 M_\pm^2 \\
&\tan 2\theta_C = \frac{2c_H\zeta\,\hat{m}_W\hat{m}_V}{\hat{m}_V^2 - \hat{m}_W^2}.
\end{aligned} \tag{2.12}$$

The custodial symmetry of eq. 2.3 guarantees that:

$$\mathbb{M}_C^2 = \begin{pmatrix} \cos\theta_W & 0 \\ 0 & 1 \end{pmatrix} \mathbb{M}_N^2 \begin{pmatrix} \cos\theta_W & 0 \\ 0 & 1 \end{pmatrix}. \tag{2.13}$$

By taking the determinant of these matrices, a custodial relation among the masses can be extracted:

$$m_W^2 M_\pm^2 = \cos^2\theta_W\, m_Z^2 M_0^2. \tag{2.14}$$

The HVT model predicts the existence of new particles at the TeV scale, but it has to reproduce the SM parameters up to the current experimental accuracy. The scale of the electroweak masses ($m_Z \sim m_W \sim 100$ GeV) can be preserved in the model, without fine-tuning cancellations, if there is a very natural hierarchy among $\hat{m}_{(W,Z)}$ and $\hat{m}_V$:

$$\frac{\hat{m}_{(W,Z)}}{\hat{m}_V} \sim \frac{m_{(W,Z)}}{M_{(\pm,0)}} \ll 1. \tag{2.15}$$

No constraints on the strength of the interaction are necessary to guarantee the natural hierarchy, hence the parameter $\zeta = g_V\,\hat{v}/2\hat{m}_V$ can either be very small or close to unity (strong coupling). If the hierarchy applies, the second lines in eq. 2.9 and eq. 2.12 can be approximated as follows:

$$\begin{aligned}
m_Z^2 &= \hat{m}_Z^2\left(1 - c_H^2\zeta^2\right)\left(1 + \mathcal{O}(\hat{m}_Z^2/\hat{m}_V^2)\right) \\
m_W^2 &= \hat{m}_W^2\left(1 - c_H^2\zeta^2\right)\left(1 + \mathcal{O}(\hat{m}_W^2/\hat{m}_V^2)\right).
\end{aligned} \tag{2.16}$$

By definition (eq. 2.11), $\hat{m}_W = \hat{m}_Z\cos\theta_W$, hence the followinf relation holds to percent accuracy:

$$\frac{m_W^2}{m_Z^2} \approx \cos^2\theta_W. \tag{2.17}$$

The SM tree-level relation, $\rho = \frac{m_W^2}{m_Z^2} = 1$, is then reproduced if $\cos^2\theta_W$ is equivalent to the experimental measurement of the weak mixing angle, within 1% accuracy:

$$\cos\theta_W{}^2 \approx 1. - 0.23. \tag{2.18}$$

By combining the custodial relation with the mass hierarchy required to naturally reproduce the SM at the percent, another fundamental consequence can be derived, namely the mass degeneracy of the triplet (to percent accuracy):

$$M_\pm^2 = M_0^2 (1 + \mathcal{O}(\%)). \tag{2.19}$$

The degenerate mass will be called $M_V \approx M_\pm \approx M_0$; given 2.15, $M_V = \hat{m}_V$. The neutral and charged components will then be produced at similar rates.
Another implication of the mass hierarchy (2.15) is that the mixing angles $\theta_{(N,C)}$ between the electroweak fields and the triplet are small:

$$\theta_{(N,C)} \approx c_H \zeta \frac{\hat{m}_{(W,Z)}}{\hat{m}_V} \ll 1, \tag{2.20}$$

hence the couplings among SM particles are very close to the couplings predicted by the SM.

2.2.2.1 Decay widths into fermions

The couplings among the triplet and SM fermions are expressed as a function of the rotation angles $\theta_{(C,N)}$ and SM couplings (omitting the CKM matrix elements for quarks):

$$\begin{aligned} &\begin{cases} g_L^N = \frac{g^2}{g_V}\frac{c_F}{2}\cos\theta_N + \left(g_L^Z\right)_{SM}\sin\theta_N \approx \frac{g^2}{g_V}\frac{c_F}{2} \\ g_R^N = \left(g_R^Z\right)_{SM}\sin\theta_N \approx 0 \end{cases}, \\ &\begin{cases} g_L^C = \frac{g^2}{g_V}\frac{c_F}{2}\cos\theta_C + \left(g_L^W\right)_{SM}\sin\theta_N \approx \frac{g^2}{g_V}\frac{c_F}{2} \\ g_R^C = 0 \end{cases}, \end{aligned} \tag{2.21}$$

where $g_L^W = g/\sqrt{2}$; $g_{L,R}^{W,Z}$ are those predicted by the standard model. The V bosons interact with SM left fermions, and the strenght of the couplings with fermions is determined by $g^2/g_V c_F$, as stated in sec. 2.2.1. The decay width into fermions is then given by:

$$\Gamma_{V^\pm \to f\bar{f}'} \approx 2\Gamma_{V^0 \to f\bar{f}} \approx N_c \left(\frac{g^2 c_F}{g_V}\right)^2 \frac{M_V}{48\pi}, \tag{2.22}$$

where N_c is the number of colours (3 for quarks, 1 for leptons).

2.2.2.2 Decay widths into bosons

As a starting point, a proper choice of the gauge makes the derivation of the approximate decay widths easier. While the unitary gauge is very convenient in discussing the electroweak symmetry breaking mechanism, since it provides a basis in which the Goldstone components of the scalar fields of the theory are set to zero, it does not properly describe the logitudinally polarized bosons in high-enery regimes, since it introduces a dependence of the type E/m in the logitudinal polarization vector, not corresponding to the experimental results [20, 21]. This pathological behaviour can be overcome profiting of the equivalence theorem: while calculating the scattering amplitude of an high-energy process, the longitudinally polarized vectors are equivalent to their corresponding

Goldstone scalars. The scattering amplitude can therefore be calculated with Goldstone diagrams. In the so-called equivalent gauge [22], the Higgs doublet is then parametrized as:

$$H = \begin{pmatrix} i\pi_+ \\ \frac{\hat{h}+h-i\pi_0}{\sqrt{2}} \end{pmatrix}, \tag{2.23}$$

and the Goldstones π_0 and π_+ describe respectively W and Z longitudinal bosons; h is the physical Higgs boson. Rewriting the simplified Lagrangian 2.3 with 2.23 parametrization, two terms hold the information of the interaction of the Vs with the Goldstones:

$$\mathscr{L}_\pi = ... + c_H \zeta \hat{m}_V V_\mu^a \partial^\mu \pi^a + \frac{g_V c_H}{2} V_\mu^a \left(\partial^\mu h \pi^a - h \partial^\mu \pi^a + \epsilon^{abc} \pi^b \partial^\mu \pi^c\right) + ..., \tag{2.24}$$

that are ruled by the $c_H g_V$ parameters combination. When ζ parameter is $\zeta \approx 1$, the first term in eq. 2.24 becomes important, and it is absorbed by a redefinition of the V_μ^a and π^a fields,

$$\begin{aligned} V_\mu^a &\to V_\mu^a + \frac{c_H \zeta}{\hat{m}_V} \partial_\mu \pi^a \\ \pi^a &\to \frac{1}{\sqrt{1-c_H^2\zeta^2}} \pi^a; \; c_H^2 \zeta^2 < 1 \end{aligned} . \tag{2.25}$$

By properly taking into account all the terms of the simplified lagrangian in the equivalent gauge, the partial widths of the dibosonic decays are ($\hat{m}_V = M_V$):

$$\begin{aligned} \Gamma_{V^0 \to W_L^+ W_L^-} \approx \Gamma_{V^\pm \to W_L^\pm Z_L} &\approx \frac{g_V^2 c_H^2 M_V}{192\pi} \frac{\left(1 + c_H c_{VVV} \zeta^2\right)^2}{\left(1 - c_H^2 \zeta^2\right)^2} = \frac{g_V^2 c_H^2 M_V}{192\pi} \left(1 + \mathscr{O}(\zeta^2)\right) \\ \Gamma_{V^0 \to Z_L h} \approx \Gamma_{V^\pm \to W_L^\pm h} &\approx \frac{g_V^2 c_H^2 M_V}{192\pi} \frac{\left(1 - 4 c_H c_{VVV} \zeta^2\right)^2}{\left(1 - c_H^2 \zeta^2\right)^2} = \frac{g_V^2 c_H^2 M_V}{192\pi} \left(1 + \mathscr{O}(\zeta^2)\right). \end{aligned} \tag{2.26}$$

2.2.2.3 Decays into fermions and bosons: concluding remarks

From eq. 2.22-2.26, some important conclusions can be extracted.

- When the ζ parameter is small, all the triplet decays (both in fermions and in dibosons), branching fractions and productions are completely determined by $g^2 c_F / g_V$, $g_V c_H$, and the degenerate mass of the triplet M_V,
- c_{VVV}, c_{VVHH}, c_{VVW} can be neglected, as long as the interest is focused in narrow resonances.

The couplings of the new resonances to fermions and bosons depend in fact by several parameters; in the following paragraphs two simplified scenarios are discussed.

2.2.3 Benchmark model A: weak coupling scenario

Model A scenario aims at reproducing a simple generalization of the SM [15], obtained by extending the gauge symmetry group with an additional $SU(2)'$. The low-energy phenonemna are expected to be dominated by the SM, while the high-energy processes are relevant for the additional symmetry, bringing additional light vector bosons in play.

It can be shown that this kind of picture is portrayed by HVT when $c_H \sim -g^2/g_V^2$ and $c_F \sim 1$. This implies that:

$$\begin{aligned} g_V\, c_H &\approx g^2/g_V \\ g^2 c_F/g_V &\approx g^2/g_V, \end{aligned} \tag{2.27}$$

hence the partial decay widths into fermions (eq. 2.22) and bosons (eq. 2.26) differ only by a factor 2 and the colour factor (N_c). Branching fractions for the model A benchmark scenario ($g_V = 1$) are shown in fig. 2.1 (left); total widths are reported in fig.2.1 (right) for different coupling parameters g_V.

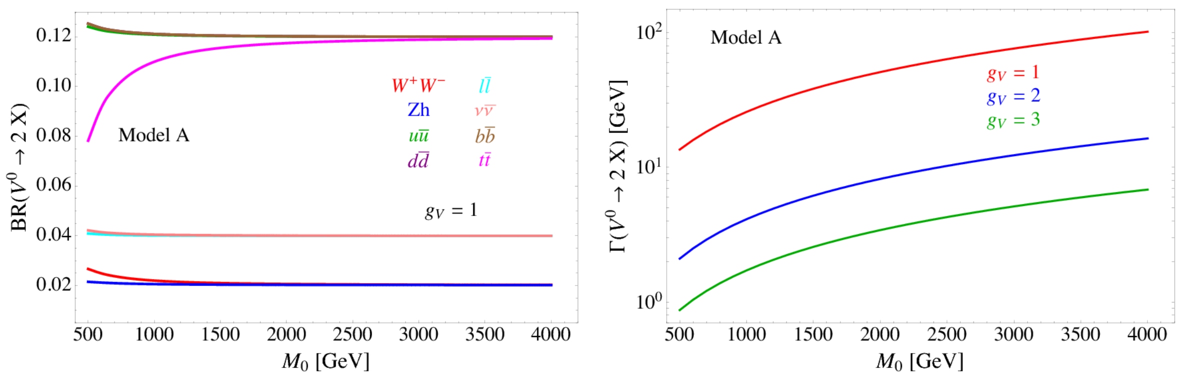


Figure 2.1: HVT model A scenario: branching fractions for fermionic and bosonic decays when $g_V = 1$ (left) as a function of the mass of the resonance M_0; total width of the resonance, as a function of its mass, considering different values of the parameter g_V (right).

2.2.4 Benchmark model B: strong coupling scenario

In composite Higgs models [17], the Higgs boson is the result of the spontaneous symmetry breaking of an $SO(5)$ symmetry to a $SO(4)$ group. New vector bosons are expected to appear, and the lightest ones can be represented by HVT model B when $c_H \sim c_F \sim 1$.

In this case:

$$\begin{aligned} g_V\, c_H &\approx -g_V \\ g^2 c_F/g_V &\approx g^2/g_V, \end{aligned} \tag{2.28}$$

hence the decay into bosons is not suppressed by g_V parameter. In the benchmark scenario $g_V = 3$, decays into dibosons are largely dominant, as it can be seen in fig. 2.2 (left); the total decay width increases for larger g_V (fig. 2.2, right). When the resonances start to be very broad, *i.e.* $\Gamma/M_V \gg 0.1$, the assumptions leading to the simplified model are no longer valid, hence higher order, non-resonant effects must be taken into account.

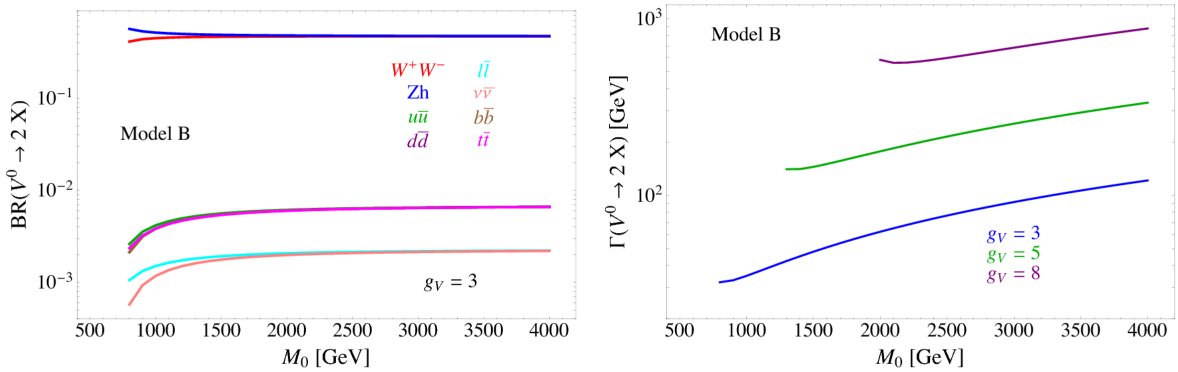


Figure 2.2: HVT model B scenario: branching fractions for fermionic and bosonic decays when $g_V = 3$ (left) as a function of the mass of the resonance M_0; total width of the resonance, as a function of its mass, considering different values of the parameter g_V (right).

2.2.5 HVT production

For resonance masses in the range of interest (∼1 TeV), the production mechanisms expected to be relevant are Drell-Yan (fig. 2.3) and Vector Boson Fusion (VBF) (fig. 2.4).

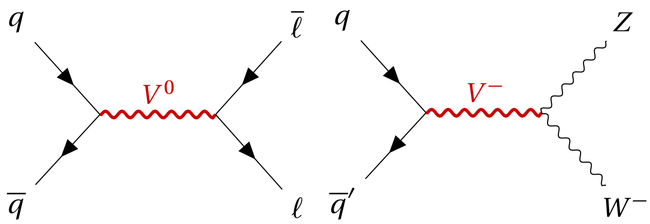


Figure 2.3: Examples of Drell-Yan production mechanism of a heavy V HVT boson: $q-\bar{q}$ quark scattering producing a neutral V^0 that decays leptonically (left); $q-\bar{q}'$ scattering producing a charged V^- that decays in a W and Z bosons (right).

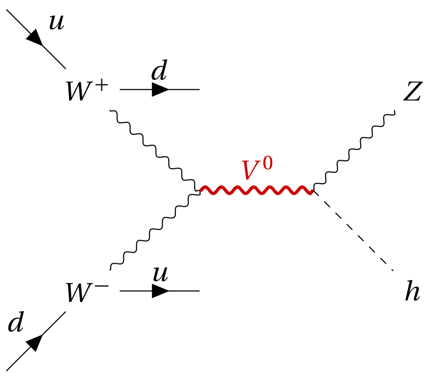


Figure 2.4: Example of VBF production mechanism of a heavy V HVT boson: a neutral V^0 boson is produced by a couple of W bosons, as a result of electroweak interactions of initial state u and d quarks. V^0 decays in a Z boson and a Higgs boson. The final state signature includes the presence of a pair of quarks, due to the primary interactions.

The cross-section of the production mechanisms is given by:

$$\sigma(pp \to V + X) = \sum_{i,j \in p} \frac{\Gamma_{V\to ij}}{M_V} f(J, S_i, S_j) g(C_i, C_j) \left. \frac{dL_{ij}}{ds} \right|_{s=M_V^2}, \tag{2.29}$$

where i, j are the partons involved in the hard interaction, $\Gamma_{V\to ij}$ is the partial width of the process $V \to ij$, $f(J, S_i, S_j)$ is a function of the spin of the resonance and of the partons, $g(C_i, C_j)$ is a function of the colour factors of each parton, s is the center-of-mass energy and $\frac{dL_{ij}}{ds}$ are the parton luminosities, that are independent from HVT model (that enters only in Γ_{ij}).
Parton luminosities, calculated for a center-of-mass energy of 14 TeV starting from quark and anti-quark parton distribution functions (PDF), are displayed in fig. 2.5 (Drell-Yan mechanism) and 2.6 (VBF mechanism). VBF luminosities are suppressed by the α_{weak} factor, therefore the process is relevant only when the bosonic decays of the triplet are dominant (strongly coupled scenario).

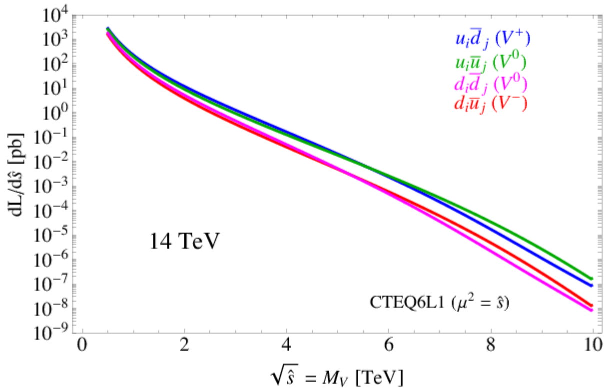


Figure 2.5: Parton luminosities for Drell-Yan process between i and j partons, as a function of the parton center-of-mass energy, for the LHC proton-proton collisions performed at 14 TeV.

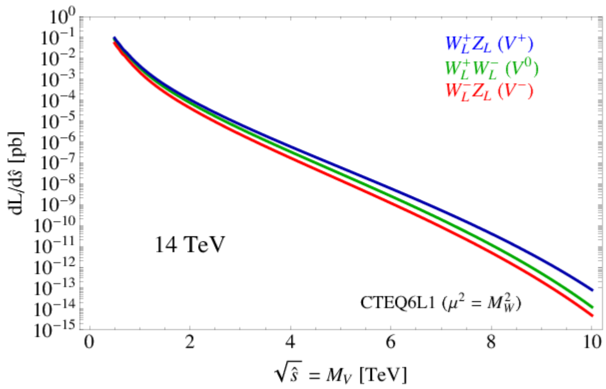


Figure 2.6: Parton luminosities for VBF process between i and j partons, as a function of the parton center-of-mass energy, for the LHC proton-proton collisions performed at 14 TeV.

2.2.6 Search for HVT resonances at LHC

No evidence of HVT resonances has been observed so far at LHC experiments. Data collected by ATLAS and CMS detectors are used to set limits on the HVT resonance masses and coupling parameters. Experimental results from proton-proton collisions performed at a center-of-mass energy of 8 TeV (Run 1 era) at LHC brought to the following conclusions. A weakly coupled resonance, in the context of benchmark model A ($g_V = 1$) was excluded up to 3 TeV by Run 1 data. By looking at parton luminosities in fig.2.5, in data produced by LHC proton-proton collision at 14 TeV, collected for an integrated luminosity of 300 fb^{-1}, the sensitivity is expected to increase up to $m_V \approx 6$ TeV. A strongly coupled resonance, in the context of benchmark model B ($g_V = 3$) is excluded up to 2 TeV by Run 1 data. Data produced by LHC at 14 TeV should increase the sensitivity up to $m_V \approx 3-4$ TeV. The most stringent limits are provided by the latest data produced by LHC at a center-of-mass energy of 13 TeV (Run 2 era).

Numerous searches for HVT triplet have been performed at CMS experiment in different final states: the most sensitive ones were those in all-hadronic topology. [23, 24] (search for WW, WZ, ZZ resonances in the $q\bar{q}q\bar{q}$ final state) excludes a W' with mass below 3.6 and a Z' with mass below 2.7 TeV in the model B scenario (fig. 2.7). [25, 26] (search for WH, ZH resonances in the $q\bar{q}b\bar{b}$ final state) excludes a W' lighter than 2.97 (3.15) TeV in the HVT model A (model B), and a Z' up to 1.67 (2.26) TeV in HVT model A (model B) (fig. 2.8). In fig. 2.9, results of [23, 24] (left) and [25, 26] (right) searches are interpreted as exclusion contours in the coupling parameter plane of the HVT model ($g_V c_H$ and $g^2 c_F / g_V$). In the grey shaded area, the narrow width approximation fails. The colored curves display the parameter exclusion for different mass hypoteses of the triplet. Colored dots show the model A and B benchmark scenarios.

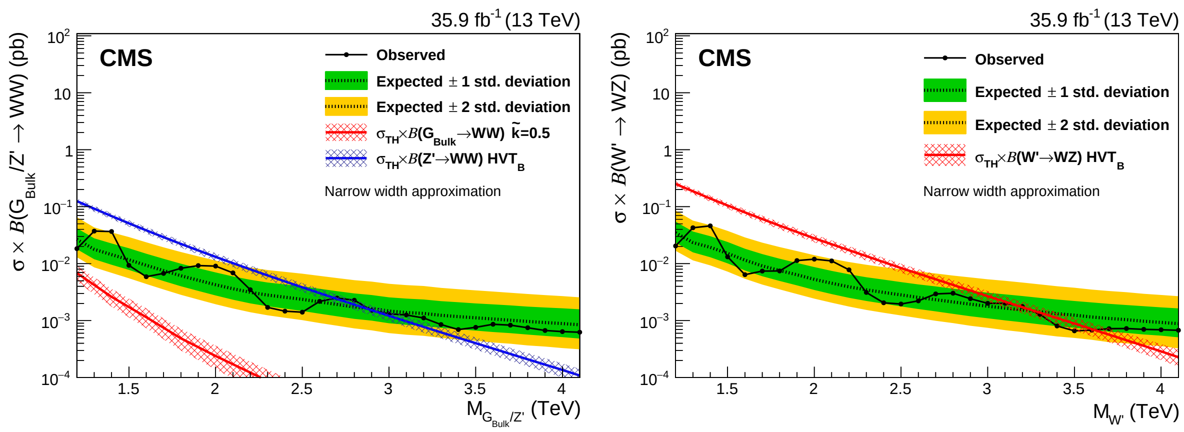


Figure 2.7: The observed and expected limits, with 68% and 95% uncertainty bands, on the product of the cross section and branching fraction $\sigma\mathcal{B}(Z' \to WW)$ for a spin-1 Z' (left) and $\sigma\mathcal{B}(W' \to WZ)$ for a spin-1 W' (right), as a function of the reconstructed mass of the diboson resonance. The colored lines show the theoretical predictions for the HVT model B.

Many other final states have been exploited at CMS: $ZW, ZZ \to \ell\bar{\ell}q\bar{q}$ [27]; $WH, ZH \to (\ell\bar{\ell}, \ell\nu, \nu\bar{\nu})b\bar{b}$ [28]; $WZ, WW \to \ell\nu q\bar{q}$ [29]. Finally, $ZW, ZZ \to \nu\bar{\nu}q\bar{q}$ [30] results will be extensively described in this thesis.

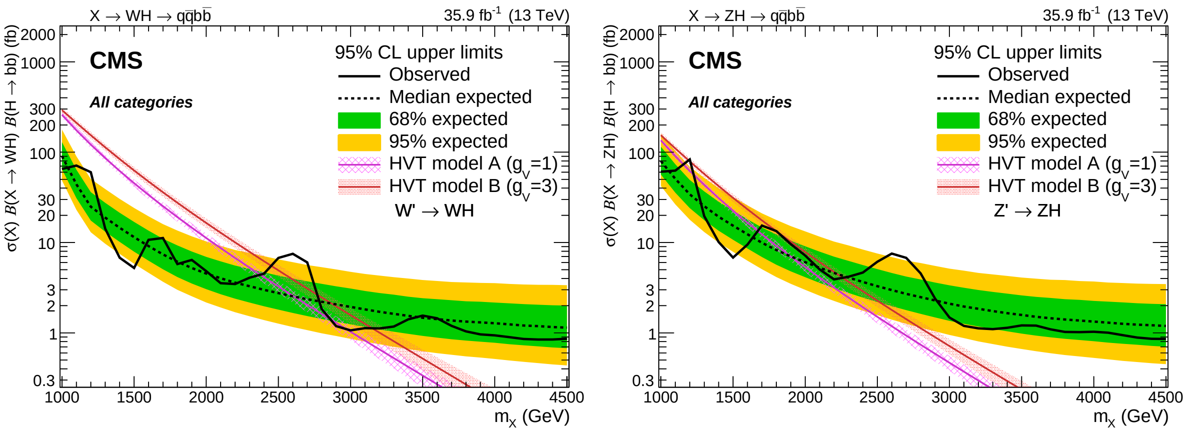


Figure 2.8: The observed and expected limits, with 68% and 95% uncertainty bands, on the product of the cross section and branching fraction $\sigma\mathscr{B}(W'\to WH)$ for a spin-1 W' (left) and $\sigma\mathscr{B}(Z'\to ZH)$ for a spin-1 Z' (right), as a function of the reconstructed mass of the diboson resonance. The colored lines show the theoretical predictions for the HVT model A and B.

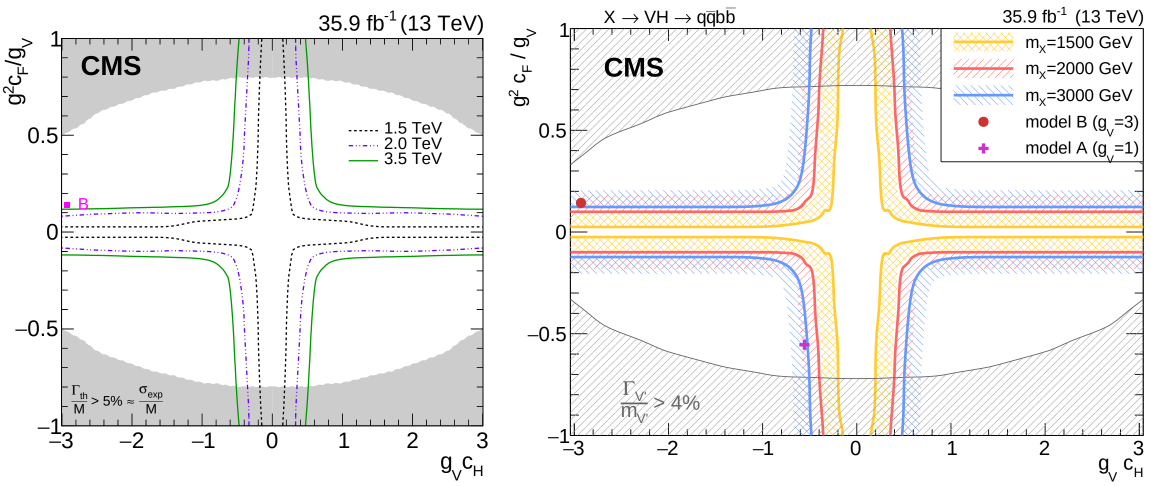


Figure 2.9: Exclusion contours in the coupling parameter plane of the HVT model ($g_V c_H$ and $g^2 c_F/g_V$).

Searches for HVT model B resonances have been performed at ATLAS experiment as well. Results for a $W' \to WZ$ reported in fig. 2.10 include the searches performed in $WW, WZ, ZZ \to q\bar{q}q\bar{q}$ final state [31]; $WZ, WW \to \ell \nu q\bar{q}$ final state [32]; $ZW, ZZ \to (\ell\bar{\ell}, \ell\nu, \nu\bar{\nu})q\bar{q}$ final state [33]. The all-hadronic final state has the best sensitivity and it excludes a W' resonance up to 3.3 TeV (model B scenario). Results for a $W' \to WH$ and for a $Z' \to ZH$ are displayed in fig. 2.11 (left and right respectively), and they include searches performed in $WH, ZH \to q\bar{q}b\bar{b}$ final state [34], and $WH, ZH \to \ell\bar{\ell}, \ell\nu, \nu\bar{\nu})b\bar{b}$ [35]. A W' is excluded up to 2.9 TeV and a Z' is excluded up to 2.8 TeV (in the model B scenario).

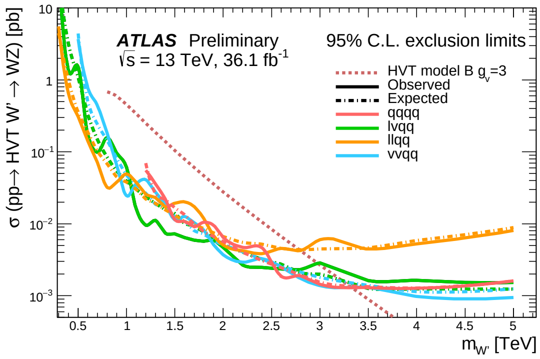


Figure 2.10: The observed and expected limits on the product of the cross section and branching fraction $\sigma\mathscr{B}(W' \to WZ)$ for a spin-1 W', as a function of the reconstructed mass of the diboson resonance. The dotted line shows the theoretical predictions for the HVT model B.

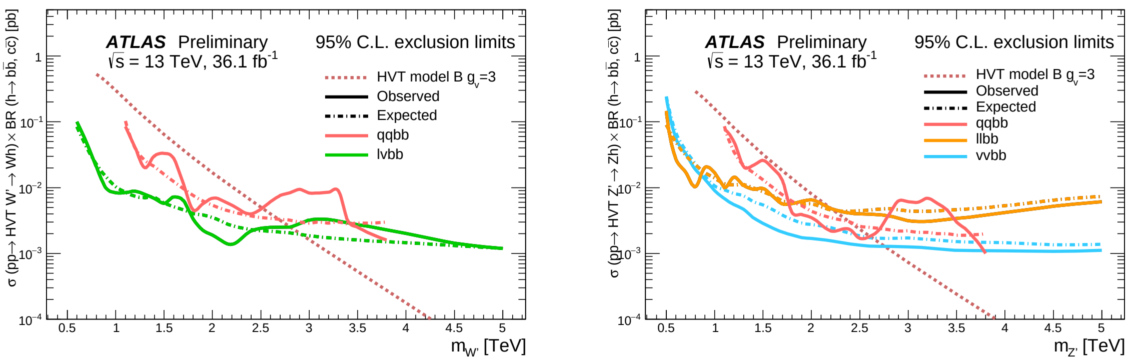


Figure 2.11: The observed and expected limits on the product of the cross section and branching fraction $\sigma\mathscr{B}(W' \to WH)$ for a spin-1 W' (left) and $\sigma\mathscr{B}(Z' \to ZH)$ for a spin-1 Z' (right), as a function of the reconstructed mass of the diboson resonance. The colored lines show the theoretical predictions for the HVT model B.

2.3 Warped extra dimension

The Randall-Sundrum model [36, 37] (RS1) proposes the introduction of one additional warped dimension in order to solve the hierarchy problem. The metric of the 5-dimensional space (a slice of AdS_5) generates an exponential hierarchy between the electroweak and Planck scales, associated respectively to the TeV three-brane, where the SM particles are confined, and the Planck three-brane. As a consequence of the new geometry, spin-2 massive gravitons are predicted to exist.
The bulk extension of the Randall-Sundrum model [38,39] states that the SM fields can propagate in the extra dimension. Light fermions are near the Planck brane, heavy fermions are close to the TeV brane, while the Higgs sector is confined in the TeV brane. Higgs couplings to the heavy fermions are therefore expected to be stronger: this naturally arising hierarchy of the masses of the SM fields gives a solution to the flavour problem. In this scenario, the fermionic decays of the bulk gravitons are suppressed, while the bosonic decays are preferred.

2.3.1 Randall-Sundrum original model (RS1)

The existence of additional n-dimensions implies that the effective Planck scale observed in 4-dimensions, $M_{PL} = 1.220910^{19}$ GeV, is related to the fundamental 4+n-dimensional Planck scale, M, via the geometry. M is expected to be of the order of the reduced $\overline{M}_{PL} = M_{PL}/2\pi$. If the 4-dimensional and the n additional metrics are factorizable, $\overline{M}_{PL}$ is the product of M and the volume of the compact space V_n:

$$\overline{M}_{PL}^2 = V_n M^{2+n}. \tag{2.30}$$

If $M \sim$ TeV, this implies that V_n must be very large, hence the compactification scale $\mu \sim 1/V_n^{1/n}$ is small (eV – MeV for n=2 – 7). Given the smallness of μ when compared to the electroweak scale, the effects of the extra dimensions should be evident in SM processes. Since they are not observed, SM particles are assumed to be confined in a 4-dimensional space, the TeV three-brane, while only gravity is allowed to propagate into the 4+n-dimensional space, the bulk. This mechanism solves the hierarchy of the Higgs scale but introduces a new hierarchy between μ and M.
In the Randall-Sundrum model [36, 37], only one additional dimension is added. The geometry of the 5-dimensional bulk is non-factorizable, and it is a slice of AdS_5 spacetime.[1] The 4-dimensional metric is multiplied by an exponential function of the fifth dimension (the "warp" factor):

$$ds^2 = e^{-2kr_c\phi}\eta_{\mu\nu}dx^\mu dx^\nu + r_c^2 d\phi^2; \tag{2.31}$$

x^μ are the usual 4-dimensional coordinates, $\eta_{\mu\nu} = diag(-1,1,1,1)$ is the Minkowski metric, k is a scale of order of $\overline{M}_{PL}$, ϕ is the coordinate of the extra dimension, $0 < |\phi| < \pi$, and r_c is the compactification radius of this finite interval. 4-dimensional mass scales are obtained by multiplying the bulk masses by $e^{-2kr_c\phi}$: given the exponential form of the warp factor, a small r_c suffices for generating a large hierarchy between Planck and Higgs scales.
Two 4-dimensional three-branes are located at the boundaries of the fifth dimension: the visible brane at $\phi = \pi$; the hidden brane at $\phi = 0$, and their metrics are obtained starting from the bulk 5-dimensional metric G_{MN}, where $M, N = \mu, \phi$:

$$\begin{aligned} g_{\mu\nu}^{\text{vis}}(x^\mu) &= G_{\mu\nu}\big(x^\mu, \phi = \pi\big) \\ g_{\mu\nu}^{\text{hid}}(x^\mu) &= G_{\mu\nu}\big(x^\mu, \phi = 0\big). \end{aligned} \tag{2.32}$$

[1] An n-dimensional anti-de Sitter space (AdS_n) is a maximally simmetric Lorentiaz mainfold, that solves the Einstein equation with a negative curvature (negative cosmological constant).

The classical action is given by:

$$
\begin{aligned}
S &= S_{\text{gravity}} + S_{\text{vis}} + S_{\text{hid}} \\
S_{\text{gravity}} &= \int d^4x \int_{-\pi}^{+\pi} d\phi \sqrt{-G}\left(-\Lambda + 2M^3\mathscr{R}\right) \\
S_{\text{vis}} &= \int d^4x \sqrt{-g_{\text{vis}}}\,(\mathscr{L}_{\text{vis}} - V_{\text{vis}}) \\
S_{\text{hid}} &= \int d^4x \sqrt{-g_{\text{hid}}}\,(\mathscr{L}_{\text{hid}} - V_{\text{hid}}),
\end{aligned}
\tag{2.33}
$$

where G (g) is the trace of the G_{MN} ($g_{\mu\nu}$) metric, Λ is the cosmological constant in the bulk, $\mathscr{R}$ is the 5-dimensional Ricci scalar, $\mathscr{L}$ and V are the lagrangian and the vacuum energy of the hidden and visible branes.
A 5-dimensional metric that preserves the 4-dimensional Poincaré invariance has the form:

$$ds^2 = e^{-2\sigma(\phi)}\eta_{\mu\nu}dx^\mu dx^\nu + r_c^2 d\phi^2. \tag{2.34}$$

The Poincaré invariance guarantees that r_c does not depend on x^μ. Given 2.34, the solution of the 5-dimensional Einstein's equations simplifies into:

$$\sigma = r_c\left|\phi\right|\sqrt{\frac{-\Lambda}{24M^3}}. \tag{2.35}$$

Furthermore, the Poincaré invariance imposes constraints to the vacuum energies and cosmological constant:

$$
\begin{aligned}
V_{\text{hid}} &= -V_{\text{vis}} = 24M^3k \\
\Lambda &= -24M^3k^2.
\end{aligned}
\tag{2.36}
$$

The final 5-dimensional metric is then:

$$ds^2 = e^{-2kr_c|\phi|}\eta_{\mu\nu}dx^\mu dx^\nu + r_c^2 d\phi^2. \tag{2.37}$$

A small r_c is considered, so the effects of the fifth dimension on 4-dimensional spacetime can't be appreciated. A 4-dimensional effective field theory approach is therefore motivated, and its mass parameters are related to the bulk parameters, M, k and r_c. In the Randall-Sundrum model, SM matter fields are confined in the TeV brane.
The massless gravitons, the mediators of the gravitational interaction in the effective field theory, are the zero modes ($h_{\mu\nu}$) of the quantum fluctuations of the classical solution (2.37):

$$ds^2 = e^{-2kT(x)|\phi|}\left(\eta_{\mu\nu} + h_{\mu\nu}(x)\right)dx^\mu dx^\nu + T^2(x)d\phi^2, \tag{2.38}$$

where the usual Minkowski metric has been replaced by $\overline{g}_{\mu\nu}(x) = \eta_{\mu\nu} + h_{\mu\nu}$; $h_{\mu\nu}$ are the tensor fluctuations around the Minkowski space, and represent both the physical graviton in 4-dimensions and the massless mode of the Kaluza-Klein decomposition of the bulk metric. r_c is the vacuum expectation value of $T(x)$.
By substituting eq. 2.38 in the classical action 2.33, an effective action can be extracted, and in particular the curvature term holds:

$$S_{\text{eff}} \sim \int d^4x \int_{-\pi}^{+\pi} d\phi\, 2M^3 r_c e^{-2kr_c|\phi|}\overline{\mathscr{R}}\sqrt{-\overline{g}}, \tag{2.39}$$

where $\overline{g}$ is the trace of $\overline{g}_{\mu\nu}$ and $\overline{\mathscr{R}}$ is the 4-dimensional Ricci scalar of $\overline{g}_{\mu\nu}$ metric. In this effective 4-dimensional action, the ϕ dependence can be integrated out, and the 4-dimensional Planck mass can be calculated:

$$\overline{M}_{PL}^2 = M^3 r_c \int_{-\pi}^{+\pi} d\phi \, e^{-2kr_c|\phi|} = \frac{M^3}{k}\left(1 - e^{-2kr_c\pi}\right). \tag{2.40}$$

It can be shown [36] that a field with a fundamental mass parameter m_0 in the bulk manifests in the visible three-brane with a physical mass m:

$$m = e^{-2kr_c\pi} m_0. \tag{2.41}$$

Scales $m \sim$ TeV are generated from $m_0 \sim \overline{M}_{PL}$ if $e^{kr_c\pi} \sim 10^{15}$. This relation stands still when Higgs field is introduced and confined in the visible three-brane:

$$v = e^{-2kr_c\pi} v_0, \tag{2.42}$$

where v is the Higgs vacuum expectation value in the TeV brane and v_0 is the 5-dimensional Higgs v.e.v.

The hierarchy problem is then solved by the exponential warp factor. The weakness of gravity in the TeV three-brane is motivated by the small overlap of the graviton wave function.

In order to calculate the mass spectrum of the graviton in the TeV brane, the tensor fluctuations of the Minkowski metric are expanded into a Kaluza-Klein (KK) tower $h_{\mu\nu}^{(n)}$:

$$h_{\mu\nu}(x,\phi) = \sum_{n=0}^{\infty} h_{\mu\nu}^{(n)}(x) \frac{\chi^{(n)}(\phi)}{\sqrt{r_c}}. \tag{2.43}$$

Once a suitable gauge is chosen, *i.e.* $\eta^{\mu\nu}\partial_\mu h_{\nu\alpha}^{(n)} = \eta^{\mu\nu} h_{\mu\nu}^{(n)} = 0$, the equation of motion of $h_{\mu\nu}^{(n)}$ becomes the Klein-Gordon relation, where $m_n^G \geq 0$:

$$\left(\eta^{\mu\nu}\partial_\mu\partial_\nu - (m_n^G)^2\right) h_{\mu\nu}^{(n)}(x) = 0. \tag{2.44}$$

By substituting eq. 2.43 into Einstein's equation, the solutions for $\chi^{(n)}(\phi)$ (commonly called "profiles") are [40, 41]:

$$\chi^{(n)}(\phi) = \frac{e^{2\sigma}}{N}\left[J_2(z_n^G) + \alpha_n Y_2(z_n^G)\right], \tag{2.45}$$

where J_2 and Y_2 are second order Bessel functions, N is the normalization of the wavefunction, α_n are coefficients and $z_n^G = m_n^G e^{\sigma(\phi)}/k$. m_n^G is the mass of the n-mode, and it depends on the roots of the Bessel functions $z_n^G = (3.83, 7.02, 10.17, 13.32, ...)$. In the limit $m_n^G/k \ll 1$ and $e^{kr_c\pi} \gg 1$:

$$m_n^G = k z_n^G(\pi) e^{-kr_c\pi}. \tag{2.46}$$

The interactions between the graviton KK modes and the matter fields in the TeV brane can be derived from the 4-dimensional effective Lagrangian, once $h_{\mu\nu}$ is replaced by its KK decomposition:

$$\mathscr{L} = -\frac{1}{\overline{M}_{PL}} T^{\mu\nu}(x) h_{\mu\nu}^{(0)} - \frac{1}{e^{-kr_c\pi}\overline{M}_{PL}} T^{\mu\nu}(x) \sum_{n=1}^{\infty} h_{\mu\nu}^{(n)}(x); \tag{2.47}$$

$T^{\mu\nu}$ is the space energy-momentum tensor of the matter fields. The zero mode of the gravitons coupling is $1/\overline{M}_{PL}$, while higher order KK modes couplings to all SM fields are suppressed by $e^{-kr_c\pi}\overline{M}_{PL}$, that is of the order of the TeV scale. Spin-2 KK masses and couplings are hence determined by the TeV scale, or, equivalently, KK gravitons are close to the TeV brane. This implies that KK gravitons can be produced via $q\bar{q}$ or gluon fusion, and that a leptonic decay of the resonance could represent a very clear signal signature.

2.3.2 Bulk extension of RS1: graviton production and decays

An extension of the original RS1 formulation has been proposed. It states that the usual SM fields are no longer confined in the TeV brane, but they are the zero modes of the corresponding 5-dimensional SM fields. If first and second generation fermions are close to the Planck brane, contribution to flavour changing neutral currents by higher-dimensional operators are suppressed. These contributions are excluded by electroweak precision tests, but they were not prevented in original RS1. The second motivation behind the choice is, as mentioned previously, the naturally arising flavour hierarchy: first and second generation quarks have small Yukawa couplings to the Higgs sector, confined in the TeV brane, while top quark and bosons have stronger Yukawa couplings.
In this picture, couplings between higher-order KK gravitons and light fermions are strongly suppressed, resulting into a negligible KK gravitons production via $q\bar{q}$, whilst gluon fusion production becomes dominant. KK gravitons decay into top quarks and Higgs bosons are dominant, given that both their profiles are near the TeV brane, while leptonic decays are negligible. Via the equivalence theorem, the Goldstone bosons are equivalent to the longitudinally polarized weak bosons, $W_L^{\pm}$ and Z_L, that have profiles close to the TeV brane. Decays of KK gravitons into weak dibosons (and production in VBF) are comparable to di-top and di-Higgs decays.

The KK decomposition and the KK mass spectrum of the graviton have already been presented in sec. 2.3.1. The KK decomposition of a massless 5-dimensional gauge field $A_M(x,\phi)$ is similarly performed [42]:

$$A_\mu(x,\phi) = \sum_{n=0}^{\infty} A_\mu^{(n)}(x) \frac{\chi^{(n)_A}(\phi)}{\sqrt{r_c}}. \tag{2.48}$$

The profiles for the gauge fields are:

$$\chi_A^{(n)}(\phi) = \frac{e^{\sigma}}{N_A}\left[J_1(z_n^A) + \alpha_n^A Y_1(z_n^A)\right], \tag{2.49}$$

where J_1 and Y_1 are first order Bessel functions. Similarly to eq. 2.50, the mass spectrum of the gauge field is:

$$m_n^A = k z_n^A(\pi) e^{-k r_c \pi}; \tag{2.50}$$

the first roots of the Bessel functions are $z_n^A = (2.45, 5.57, 8.70, 11.84, ...)$.

The Lagrangian expressing the interaction between the m and n modes of the bulk field F to the q KK gravitons mode G is [42]:

$$\mathscr{L}_{G-F} = \sum_{m,n,q} C_{mnq}^{FFG} \frac{1}{\overline{M}_{PL}} \eta^{\mu\alpha}\eta^{\nu\beta} h_{\alpha\beta}^{(q)}(x) T_{\mu\nu}^{(m,n)}(x), \tag{2.51}$$

C_{mnq}^{FFG} is the overlap integral of the profiles:

$$C_{mnq}^{FFG} = \int \frac{d\phi}{\sqrt{k}} e^{t\sigma} \frac{\chi_F^{(m)} \chi_F^{(n)} \chi_G^{(q)}}{\sqrt{r_c}}; \tag{2.52}$$

t depends on the type of field considered.
The coupling between gluons and the q KK graviton mode is given by:

$$C_{00q}^{AAG} = e^{k\pi r_c} \frac{2\left[1 - J_0(x_n^G)\right]}{k\pi r_c (x_n^G)^2 \left|J_2(x_n^G)\right|}. \tag{2.53}$$

Once eq. 2.53 is put in eq. 2.51, the most significant partial decay widths into the q KK graviton mode are:

$$\begin{aligned}
\Gamma(G\to t_R\bar{t}_R) &\sim N_c\frac{\left[\tilde{k}x_q^G\right]^2 m_q^G}{320\pi}\\
\Gamma(G\to hh) &\sim \frac{\left[\tilde{k}x_q^G\right]^2 m_q^G}{960\pi}\\
\Gamma\left(G\to W_L^+W_L^-\right) &\sim \frac{\left[\tilde{k}x_q^G\right]^2 m_q^G}{480\pi}\\
\Gamma(G\to Z_LZ_L) &\sim \frac{\left[\tilde{k}x_q^G\right]^2 m_q^G}{960\pi},
\end{aligned} \tag{2.54}$$

where $\tilde{k}=k/\overline{M}_{PL}$; the total decay width is:

$$\Gamma_G=\frac{13\left[\tilde{k}x_q^G\right]^2 m_q^G}{960\pi}. \tag{2.55}$$

Calculations, so far, have been performed considering $M\sim\overline{M}_{PL}$ and $k<M$, hypoteses under which the solution for the bulk metric (eq. 2.37) is valid. Hence, $\tilde{k}=k/\overline{M}_{PL}\leq 1$ is taken as a reference interval. This has also phenomenological consequences on the width of the resonance, as stated in eq. 2.55. The total decay width of the lightest KK graviton mode, compared to its mass, is shown as a function of $\tilde{k}$ in fig. 2.12 [43]. At $\tilde{k}=1$, in the bulk scenario, the KK graviton width is expected to be few % of its mass, up to 4 TeV (dotted red curve). The narrow width approximation holds, hence the resonance properties can be probed at the peak, neglecting the effects in the tails of the mass distribution.

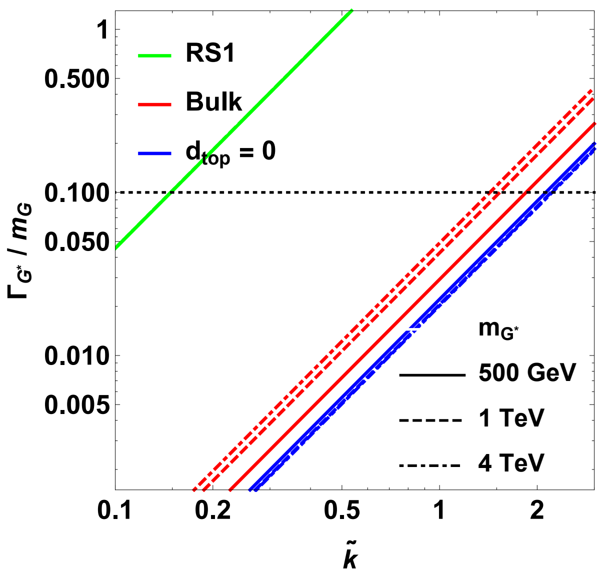


Figure 2.12: Width of the KK gravitons, in units of the mass of the resonance, as a function of the curvature parameter $\tilde{k}$. The red curves represent the bulk extension of RS1 original model for different mass hypoteses (from 500 GeV up to 4 TeV).

The total cross-section of a bulk graviton, produced at LHC in proton-proton interactions via gluon fusion (displayed in fig. 2.13), decaying into a couple of vector bosons (for the purpose of this thesis, a final state with two longitudinally polarized Z bosons is considered) is expressed as a function of the parton level cross-section $\hat{\sigma}$, the gluon parton distribution functions f_q, the momentum transfer $Q^2 \sim (m_q^G)^2$ and the center-of-mass energy s:

$$\sigma(pp \to ZZ) = \int dx_1 dx_2 f_g(x_1, Q^2) f_g(x_2, Q^2) \hat{\sigma}(x_1 x_2 s). \tag{2.56}$$

The differential parton level cross-section, averaged over colors and initial spin states, is (hatted quantities are calculated in the center-of-mass frame):

$$\frac{d\hat{\sigma}(gg \to ZZ)}{d\cos\hat{\theta}} \approx \frac{|\mathscr{M}_{+-00}|^2}{1024\pi\hat{s}}, \tag{2.57}$$

where $|\mathscr{M}_{+-00}|$ is the matrix element of the dominant contribution in $gg \to VV$ process (Γ_G is defined in eq. 2.55, a, b are color factors):

$$\mathscr{M}_{+-00}(g^a g^b \to VV) = -C_{00q}^{AAG} e^{-k\pi r_c} \left(\frac{x_n^G \tilde{k}}{m_n^G}\right)^2 \sum_n \frac{\delta_{ab} \mathscr{A}_{+-00}}{\hat{s} - m_n^{G\,2} + i\Gamma_G m_n^G}. \tag{2.58}$$

The relevant amplitudes taken account in the matrix element calculation are [38]:

$$\mathscr{A}_{+-00} = \mathscr{A}_{-+00} = \frac{\left(1 - 1/\beta_Z^2\right)\left(\beta_Z^2 - 2\right)\left[(\hat{t} - \hat{u})^2 - \beta_Z^2 \hat{s}^2\right]\hat{s}}{8M_Z^2}, \tag{2.59}$$

where $\beta_Z^2 = 1 - 4M_Z^2/\hat{s}$ and M_Z is the mass of the Z boson.

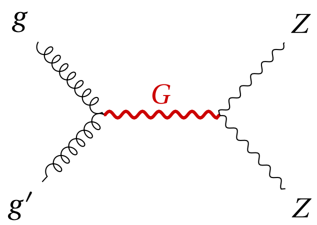


Figure 2.13: Gluon fusion production mechanism for a KK graviton that decays in a couple of Z bosons.

2.3.3 Search for KK bulk gravitons at LHC

No evidence of spin-2 bulk graviton resonances has been observed so far at LHC experiments. Data collected by ATLAS and CMS detectors are used to set limits on the graviton masses, generally considering different curvature parameter $\tilde{k}$ hypoteses, once assured the narrow width approximation is still valid (up to $\tilde{k} \sim 1$). The most stringent limits have been set with Run 2 data.

Many results of the diboson searches performed at CMS and already presented in sec. 2.2.6 are interpreted in the context of the bulk gravitons, together with the additional final states $WZ, ZZ \to \ell\bar{\ell}\nu\bar{\nu}$ [44] and $HH \to b\bar{b}b\bar{b}$ [45]. The most interesting limit is provided by [44], that, under the hypotesis $\tilde{k} = 0.5$, excludes a spin-2 bulk graviton with a mass lower than 800 GeV (fig. 2.14).

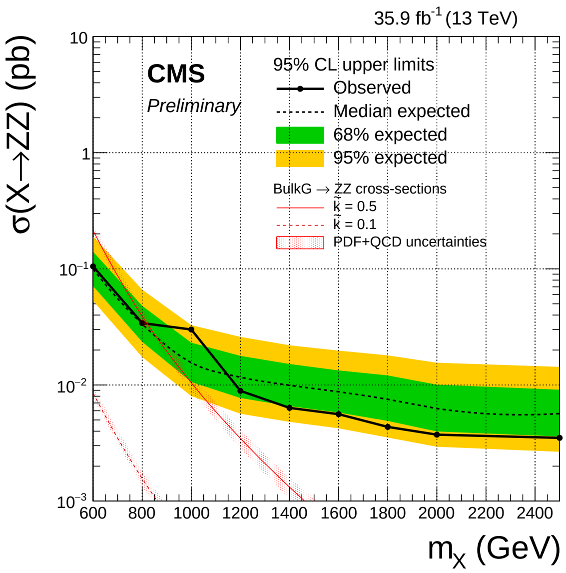


Figure 2.14: The observed and expected limits, with 68% and 95% uncertainty bands, on the product of the cross section and branching fraction $\sigma\mathscr{B}(G \to ZZ)$ for a spin-2 bulk graviton, as a function of the reconstructed mass of the diboson resonance. The colored lines show the theoretical predictions for $\tilde{k} = 0.1$ and 0.5.

Similarly for ATLAS experiment, searches for diboson resonances in sec. 2.2.6 have been interpreted in the graviton context. The most stringent limit is given by [32], where, under the assumption $\tilde{k} = 1$, a spin-2 bulk graviton with mass lower than 1.76 TeV is excluded (fig. 2.15).

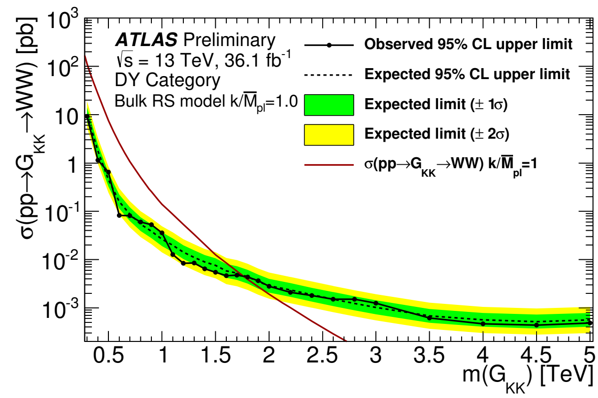


Figure 2.15: The observed and expected limits, with 68% and 95% uncertainty bands, on the product of the cross section and branching fraction $\sigma\mathscr{B}(G \to ZZ)$ for a spin-2 bulk graviton, as a function of the reconstructed mass of the diboson resonance. The colored lines show the theoretical predictions for $\tilde{k} = 1$.

Chapter

3

The Large Hadron Collider and the CMS experiment

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [46] is a 27 km ring structure designed for the acceleration and collision of protons and heavy ions. It is situated approximatively 100 m underground, between France and Switzerland, in the Geneva area, and it is part of the CERN research facilities. In order to reduce the cost of the project, approved in 1996, the LHC has been designed to fit the pre-existing underground tunnel of the Large Electron-Positron collider (LEP), built to accelerate electrons and positrons and running until the year 2000.

Moving from an electron-positron collider to an hadron collider allowed to reach higher energies in the center-of-mass frame, since the synchrotron radiation loss is inversely proportional to the fourth power of the mass of the particle involved: hence, the radiation is reduced by a factor $m_p/m_e \sim 10^3$. The choice of a proton-proton collider was driven also by the possibility to collect higher luminosities (and hence more statistics) with regards to, for example, a proton-antiproton collider, like Tevatron at Fermilab, in the USA.

In the LHC two identical beam pipes are designed to let protons circulate in opposite directions, in ultrahigh vacuum conditions (10^{-11}–10^{-10} mbar), to avoid spurious collisions with gas molecules. Given the reduced diameter of the tunnel (4 m), the two proton beams are magnetically coupled. The collider is composed by 8 arc sections (48 km) driving protons around the ring, and straight sections (6 km) where beam control systems and detectors are inserted. Proton beams collide in four interaction points, where the main LHC experiments are installed: ALICE, ATLAS, CMS, LHCb.

In fig. 3.1, a slice of the arc section is displayed. Around the beam pipes, two superconducting magnetic dipoles are located: they generate vertical magnetic fields in opposite directions. The superconducting coils are made of niobium-titanium, materials that are superconducting at very low temperature. At the LHC, they are kept at a temperature of 1.9 K (-271.3°C) by a closed liquid helium circuit. A current of 11850 A flows through the magnets, without any energy loss due to electrical resistance, generating a magnetic field of 8.33 T. Magnets of higher order in multipole expansion (quadrupoles, sextupoles, octupoles, etc.) are employed to optimize the proton trajectories; in particular, quadrupoles allow to focus and squeeze the beams. Along the LHC ring there are 9593 magnets; 1232 are dipoles, 392 are quadrupoles.

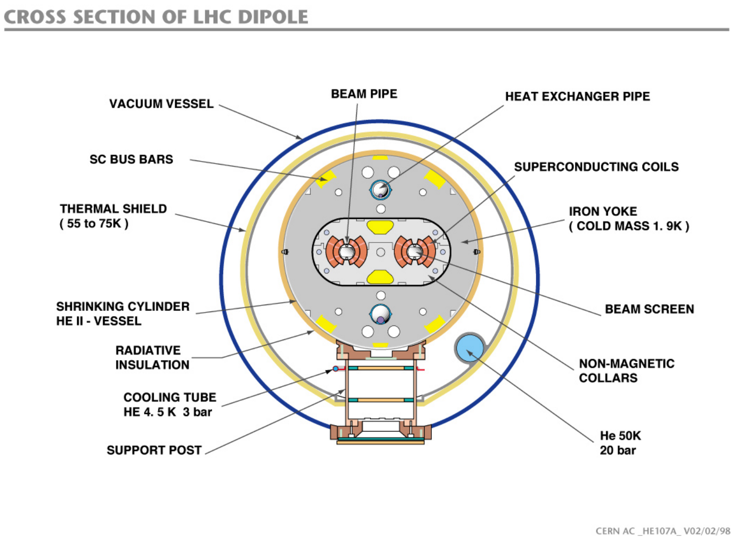


Figure 3.1: Section of the LHC dipole magnet structure.

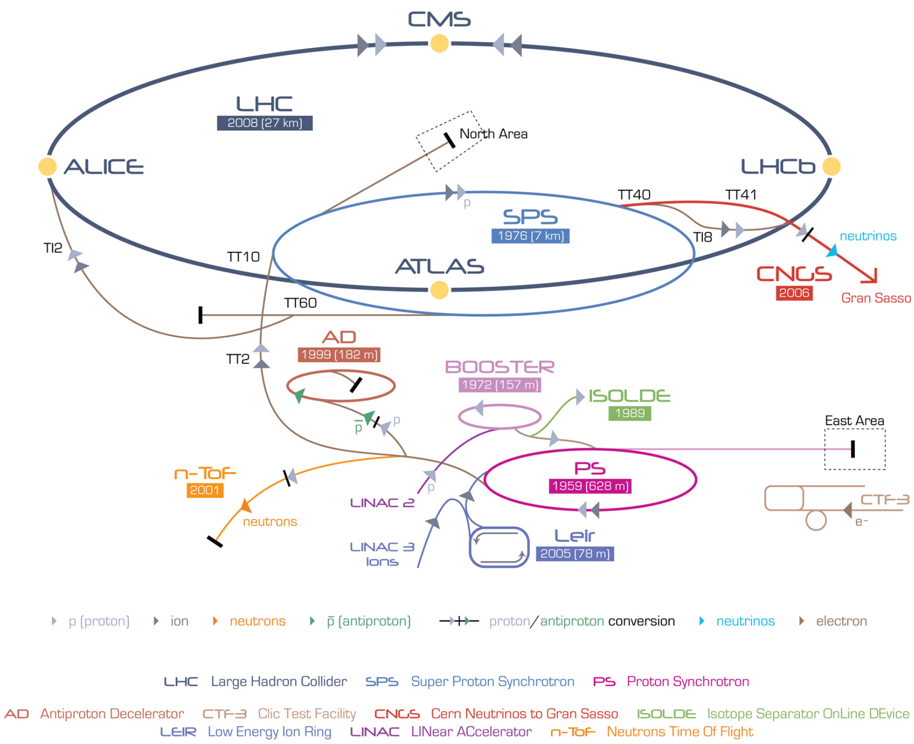


Figure 3.2: The CERN accelerator complex.

The LHC represents the final step of the CERN accelerator complex, shown in fig. 3.2. Protons are extracted from hydrogen atoms and inserted in the linear accelerator Linac2, that brings them to an energy of 50 MeV. They circulate around a little synchrotron, Proton Synchrotron Booster, reaching an energy of 1.4 GeV, and then in the Proton Synchrotron (PS), where their energy is increased to 25 GeV. The second to last step is the Super Proton Synchrotron, SPS, accelerating protons up to 450 GeV. They are finally injected in the Large Hadron Collider, where sixteen radiofrequency cavities (RF) accelerate protons inside each beam up to an energy of 6.5 TeV, corresponding to a center-of-mass energy of 13 TeV when colliding. The RF cavities provide an accelerating electromagnetic field up to 5 MV/m (maximum voltage of 2 MV), that oscillates with a frequency of 400 MHz. Like the magnets, the cavities are kept at low temperature (4.5 K, or -268.7°C) in order to allow superconducting conditions. The maximum beam energy can be reached in 15 minutes. After several hours of collisions ($\sim$ 10 hours), the quality of the beams deteriorates and they are extracted from the machine and dumped.

Protons circulate inside the LHC ring in bunches of $\sim 10^{11}$ particles each, 80 mm long. Focusing magnets allow to reduce the bunch diameter down to 16 μm. Different bunches are separated by 25 ns (or, $\sim$ 7.5 m), corresponding to a frequency of 40 MHz and an instantaneous (peak) luminosity (defined in eq. 3.1) of 1.2×10^{34} cm^{-2}s^{-1}. Given the structure of the beams, at every bunch crossing many protons interact simultaneously: this phenomenon is called pile-up. The designed maximum number of bunches per fill is 2808.

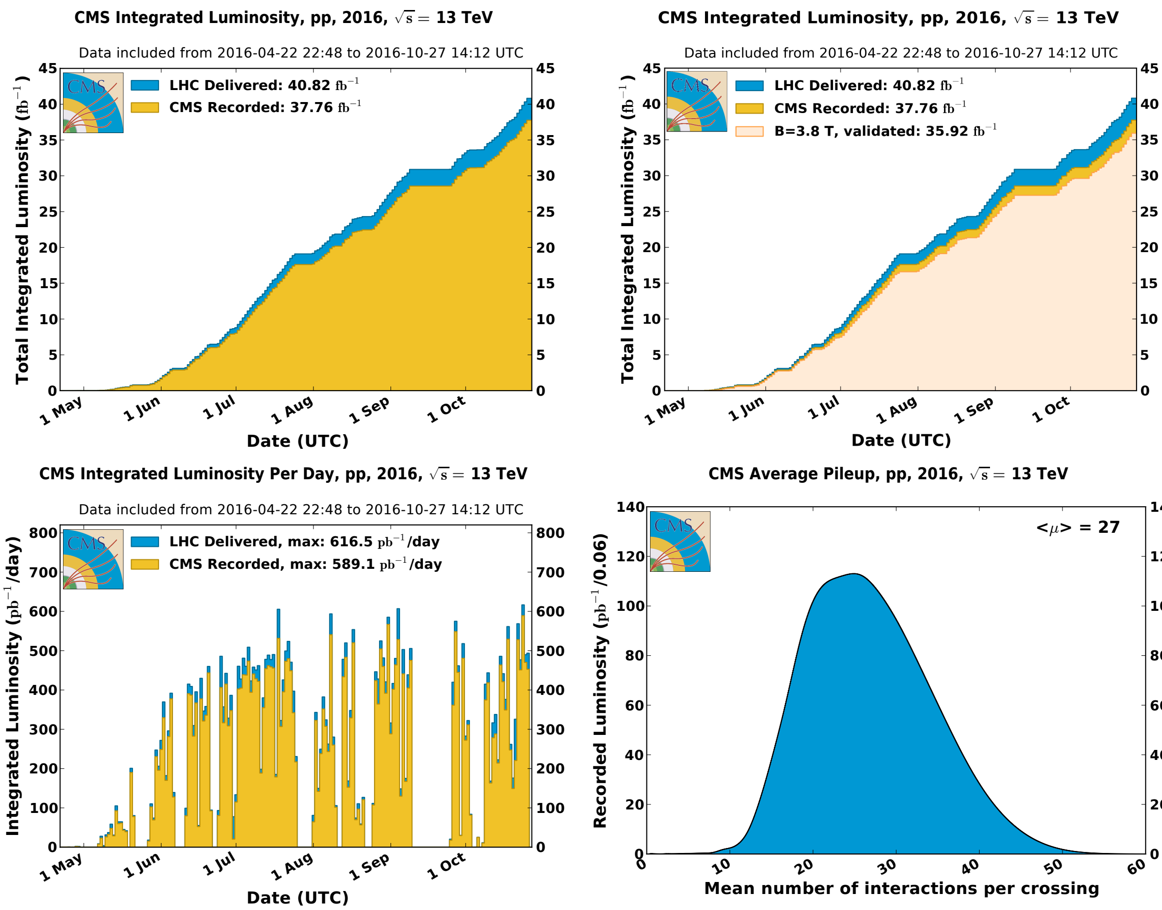


Figure 3.3: Luminosity in 2016 LHC data. Top-left plot: the cumulative integrated luminosity delivered by LHC (in blue) and recorded by CMS (in orange), as a function of the data taken period. Top-right plot: data recorded by CMS and declared as optimal for the physics analyses (in light orange), corresponding to a total integrated luminosity of 35.9 fb^{-1}. Bottom-left plot: maximum integrated luminosity per day. Bottom-right plot: number of proton interactions per bunch crossing (pile-up).

The main parameters characterizing an hadronic collider are the center-of-mass energy, corresponding to the sum of the energies of the beams, and the instantaneous luminosity, that describes the frequency of the interactions among the bunches in the beams. If the bunches in the first beam contain n_1 protons, and the bunches in the second beam contain n_2 protons, and if the colliding area is Σ, the frequency of complete turns around the ring is f, the instantaneous luminosity $\mathscr{L}_{\text{inst}}$ is:

$$\mathscr{L}_{\text{inst}} = f\frac{n_1 n_2}{\Sigma}. \tag{3.1}$$

If a generic physics process i has a cross-section of σ_i, the interaction rate R_i is:

$$R_i = \frac{dN_i}{dt} = \sigma_i \mathscr{L}_{\text{inst}}, \tag{3.2}$$

and the number of events recorded in the time interval $(0,\tau)$ is obtained from the integrated luminosity $\mathscr{L} = \int_0^\tau \mathscr{L}_{\text{inst}} dt$:

$$N_i = \sigma_i \int_0^\tau \mathscr{L}_{\text{inst}} dt. \tag{3.3}$$

In fig. 3.3, a summary of the luminosity measurement in 2016 data is presented. The luminosity delivered by LHC is represented in blue, the recorded by CMS is in orange. The mean number of interaction per bunch crossing (pile-up) is presented as well. The average number of interactions per collision is 27, the maximum is around 50 (in fig. 3.4, a record of 78 pile-up collisions was detected).

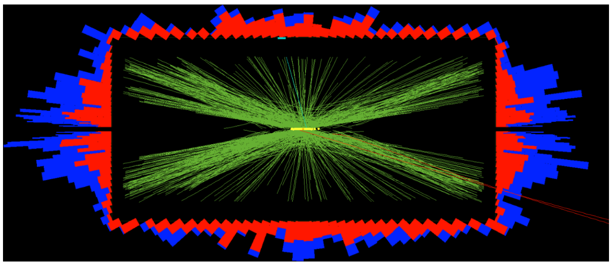


Figure 3.4: CMS collision event, where a record of 78 interactions per single bunch crossing were taking place simultaneously.

3.1.1 Proton-proton interactions

Proton-proton collisions allow to reach high energies and luminosities, but the drawback is the complexity of the events when compared to electron-positron collisions: not only because of the increasing backgrounds due to strong interactions among partons, but also because the momenta of the proton partons taking part in the interaction are unknown; not to mention the problem of disentangling the tracks of the particles coming from the interesting hard interactions from the spectator pile-up interactions.
The majority of the LHC events is represented by soft interactions, with low transverse momentum transfer, namely elastic and diffractive scatterings. In the so-called hard interactions, on the other hand, the transferred momentum among particles is high, allowing to produce massive resonant phenomena. These events manifest in peculiar final state signatures that can be distinguished from the soft background interactions.
At high momentum transfer (perturbative regime), a proton can be described as a collection of partons, each bringing a fraction x of the initial beam momentum, whose distribution is described by the parton distribution functions (PDF), $f(x, Q^2)$, as a function of the Bjorken's variable and of the

momentum transfer Q^2. At very high center-of-mass energies (13 TeV), the proton mass can be neglected; the available energy in the parton 1 – parton 2 scattering is unknown, $\sqrt{x_1 x_2 s}$. The total cross-section is given by:

$$\sigma = \int dx_1 f_1(x_1, Q^2) \int dx_2 f_2(x_2, Q^2) \sigma_{12}(x_1 p_1, x_2 p_2, Q^2), \tag{3.4}$$

where σ_{12} is the cross-section at parton level, and f_1, f_2 are the parton PDFs. In fig. 3.5, parton cross-sections of the main standard model processes are displayed, as a function of the center-of-mass energy.

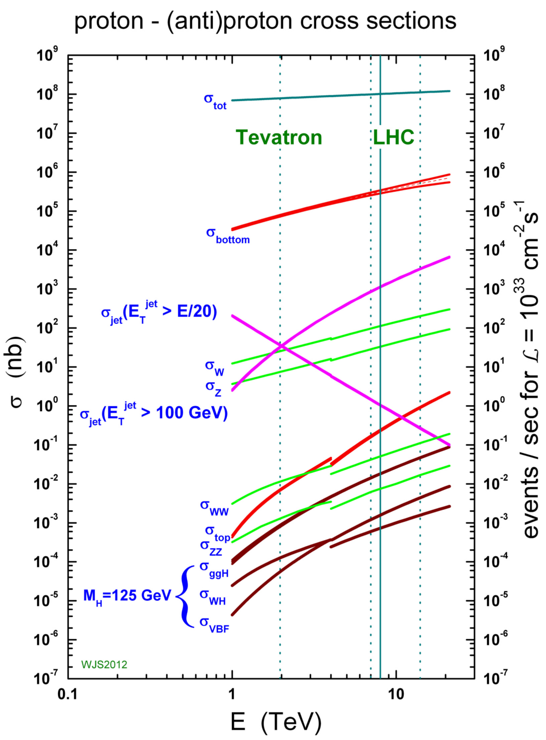


Figure 3.5: Cross-sections and number of expected events in proton-proton collisions, as a function of the center-of-mass energy. Rare phenomena, such as the Higgs boson production, can be observed at the LHC.

3.2 CMS detector

The Compact Muon Solenoid (CMS) is a multi-purpose detector built in the LHC ring. It is situated in a cavern 100 m underground, near Cessy, in France. It is a cylinder 22 m long, with a diameter of 15 m, and a weight of 12500 tons. Its physics programme includes the search for the Higgs boson (discovered in 2012), precision measurements of the Standard Model parameters and rare decays (physics of beauty quark), and search for new physics beyond the standard model (SUSY, exotic phenomena, dark matter, extra dimensions).
The CMS detector is structured in many layers of sub-detectors, giving different responses depending on the nature and the momentum of the particles passing through. The inner detectors have been finely segmented in order to afford the high radiation levels and particle multiplicity at the

interaction point, so that the reduced occupancy of each layer allows to measure and distinguish precisely the primary vertices of the hard interactions from the pile-up events. A very accurate time resolution is necessary to synchronize all the subsystems together.

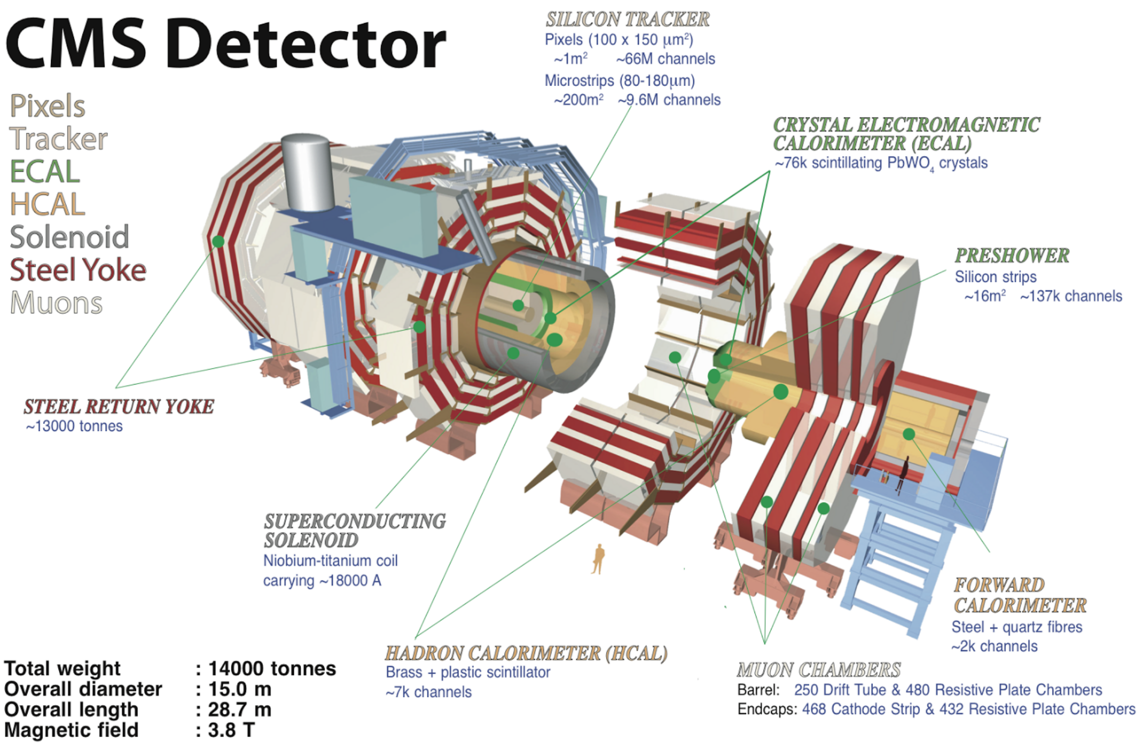


Figure 3.6: The CMS experiment.

Fig. 3.6 shows a sketch of the CMS detector. It is longitudinally segmented in the barrel region and two endcaps. In the forward region (over the endcaps), where the beam radiation is very intense, additional calorimeters have been placed. In fig. 3.7, the mean path of a specific particle through the sub-detectors is represented, depending on its flavour.
A detailed description of the CMS detector can be found in [47].

3.2.1 The coordinate system

The CMS coordinate system is depicted in fig. 3.8. x and y are the coordinates in the transverse plane, z is the longitudinal coordinate. The x axis points at the center of the LHC ring, the y axis points upward, the z axis is along the beam direction. The azimuthal angle φ lies in the transverse plane, and it is measured starting from the x axis; the radial coordinate is r. The polar angle θ lies in the plane rz. The transverse component of the 3-momentum, $\vec{p}_T$, is orthogonal to the beam axis and lies in the plane xy. The transverse energy is defined as the magnitude of $\vec{p}_T$: $E_T = E \sin\theta$.
Two other commonly used variables are the rapidity, $\mathscr{Y}$, and pseudorapidity, η, defined as functions of the particle energy E, the longitudinal component of the momentum p_z and the 3-momentum

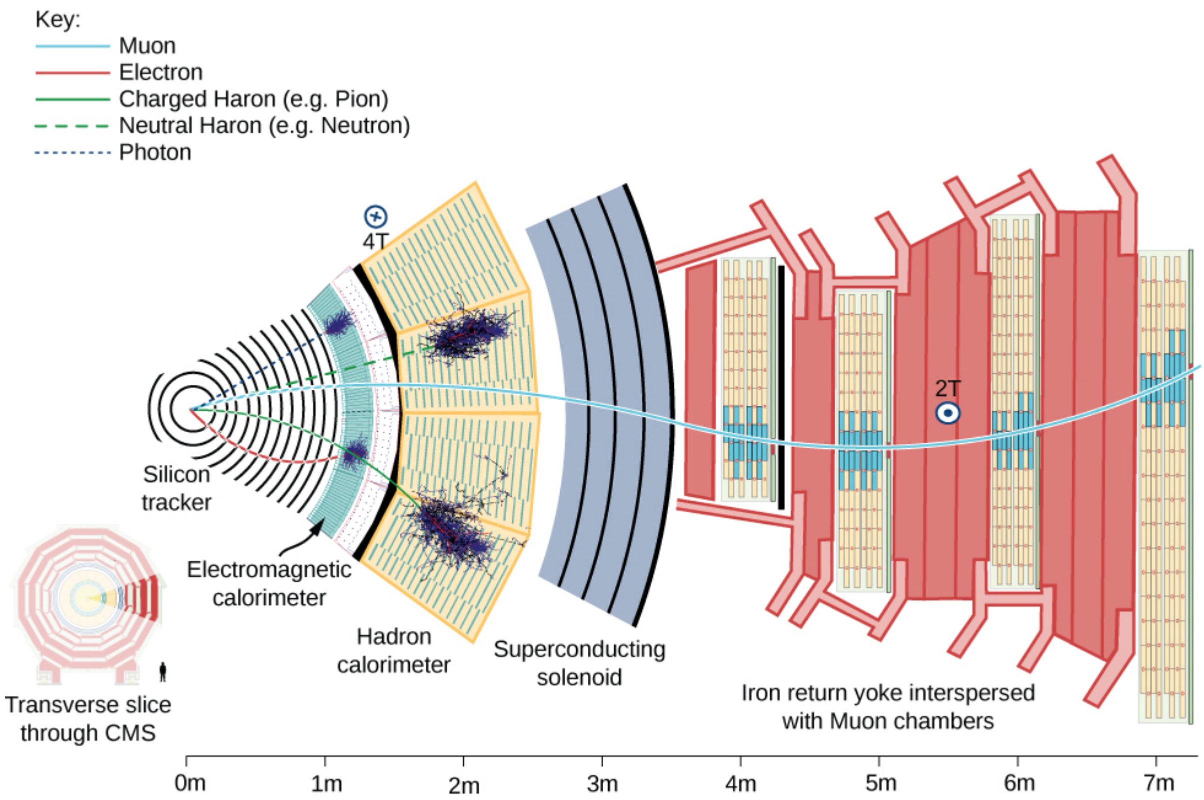


Figure 3.7: Mean path of a particle through the CMS detector. A muon, in light blue, passes through with a bended trajectory, depending on its momentum and charge, triggering signals in all the subsystems. An electron, in red, leaves a track in the silicon tracker and is absorbed by the electromagnetic calorimeter. A neutral or charged hadron, in green, stops inside the hadronic calorimeter. A photon, dotted blue line, showers in the electromagnetic calorimeter, without leaving any track in the silicon detector.

modulus:

$$\begin{aligned} \mathscr{Y} &= \frac{1}{2}\log\frac{E+p_z}{E-p_z} \\ \eta &= \frac{1}{2}\log\frac{|\vec{p}|+p_z}{|\vec{p}|-p_z} = -\log\tan\frac{\theta}{2}. \end{aligned} \tag{3.5}$$

When the considered particle is produced in the forward region, hence at $\theta = 0$, it means that $\eta \rightarrow \infty$. When the particle is produced in the transverse plane, hence $\theta = \pi/2$, $\eta = 0$. At high energies, when the masses can be neglected, rapidity and pseudorapidity coincide; these variables are largely used at colliders because they are invariant under Lorentz boosts along the beam direction.

3.2.2 The magnet

The CMS superconducting magnet is an hollow cylinder (13 m long, 6 m of diameter, shown in fig. 3.9). An electrical current of 19 kA flows through the niobium and titanium fibers that constitute the solenoid, providing a maximum magnetic field of 3.8 T and storing a maximum energy of 2.6 GJ. Superconducting conditions are mantained by a liquid helium cooling system, keeping the solenoid temperature at 4.5 K. In order to avoid stray fields, the magnetic field lines are closed by

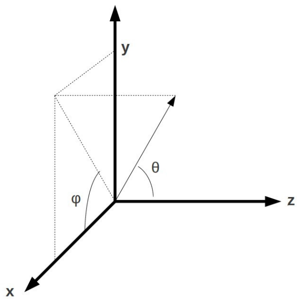


Figure 3.8: CMS coordinate system.

the return yoke, composed by 10 ktons of magnetized iron blocks, located in the outer part of CMS and alternated to the muon chambers. The homogeneus magnetic field inside the detector bends the trajectories of the charged particles, allowing the measurement of their momenta p, given the relation with the magnetic field strenght B and the radial coordinate r of the trajectory:

$$p[\mathrm{GeV}] = 0.3 \times B[\mathrm{T}] \times r[\mathrm{m}]. \tag{3.6}$$

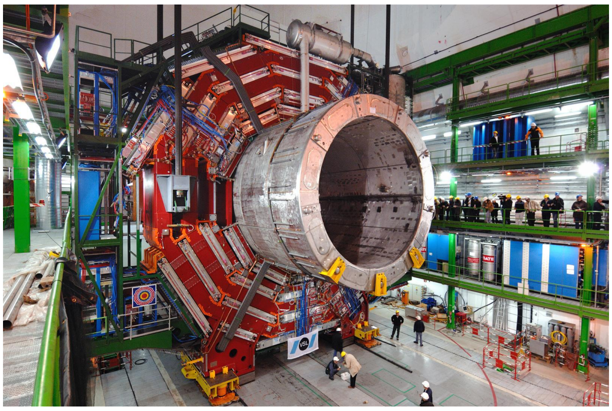


Figure 3.9: Installation of the superconducting solenoid in the CMS cavern.

3.2.3 The tracking system

The CMS tracking system [48, 49] is composed by a cylinder of silicon detectors (2.5 m of diameter and 5.8 m length). Their design guarantees a precise reconstruction of the tracks left by charged particles and of the interaction vertices, a fundamental tool to identify heavy quarks (charm, beauty) and leptons (taus). Tracker detectors cover a pseudorapidity region of $|\eta| < 2.5$ and have an active area of 210 m^2. The two sub-detectors of the tracking system are the pixel detector, installed close to the interaction point, and the strip detector, covering a radius of 0.2 – 1.2 m. The high granularity of the pixels and strips allows to keep the occupancy at acceptable levels, given the high multiplicity of the tracks ($\sim$1 MHz/mm^2). The silicon detectors and the electronic cables are cooled down to a temperature of $\sim 10°$ C. The structure of the tracking system is shown in fig. 3.10.

3.2.3.1 The pixel detector

The pixel detector is composed by 66 millions of silicon cells, whose dimensions are $100 \times 150\ \mu m^2$, 285 μm of thickness, placed in 1440 modules. Silicon cells are set in three layers in the barrel re-

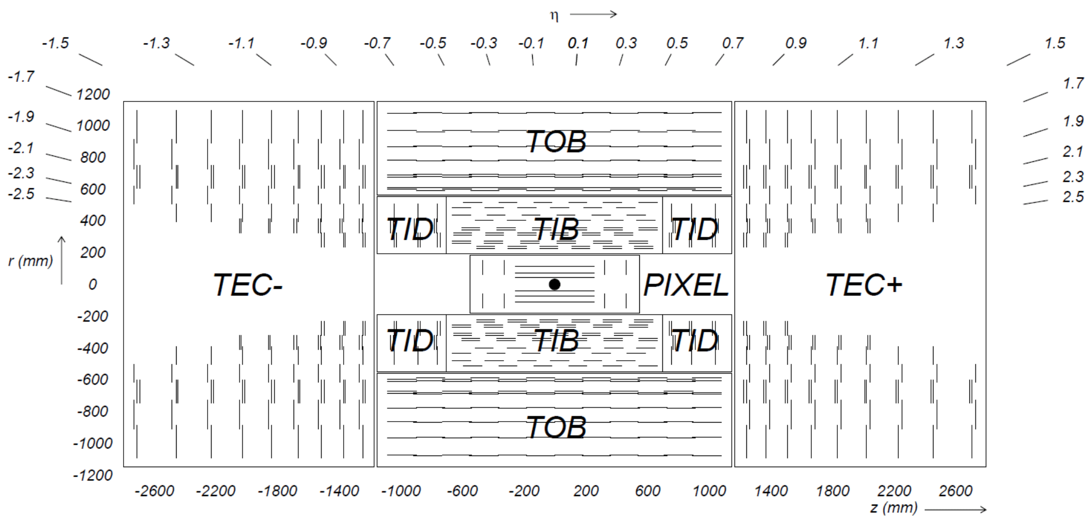


Figure 3.10: The CMS tracking system: the inner pixel detector, close to the interaction point, and the outer strip detector.

gion and in two disks at each endcap. Barrel modules are disposed parallel to the magnetic filed, whilst at the endcap they are tilted by about 20°. Pixels allow a spatial resolution of 10 μm in the transverse plane, and of ~20 μm along the longitudinal coordinate. Their reduced size guarantees an occupancy of 10^{-4} per pixel at each bunch crossing, in high luminosity regime.

3.2.3.2 The strip detector

The strip system is divided in the four-layered tracker inner barrel (TIB), covering a region $20 < r < 55$ cm with respect to the interaction point, the six-layered tracker outer barrel (TOB), located at $55 < r < 110$ cm, the three tracker inner disks (TID) and the nine tracker endcaps (TEC) at each cylinder base. Given the lower radiation level at higher radii (and hence a lower occupancy, around few percent), strips are bigger than the pixels. Silicon strips in TIB and TID are 320 μm thick, 10 cm long, and with a pitch ranging from 80 to 120 μm; strips in TOB and TEC are 25 cm long, with a different thickness (320 μm for TID, 500 μm for TEC) and pitch (97-184 μm). There are 15148 strip modules, and 9.3 million readout channels. The strip spatial resolution is about 20 – 50 μm in the transverse plane and about 200 – 500 μm along the longitudinal coordinate.

3.2.4 The elctromagnetic calorimeter

The CMS electromagnetic calorimeter (ECAL, shown in fig. 3.11) [50] is a homogeneous detector composed by lead tungstate ($PbWO_4$) scintillating crystals, designed to measure the energy deposits of photons and electrons through their electromagnetic showers. $PbWO_4$ is transparent and dense (8.3 gr/cm^3); it has a fast time response (the 85% of the scintillating light is emitted at every bunch crossing, namely 24 ns), high scintillating efficiency and radiation resistance; it has a radiation length is $X_0 = 0.89$ cm and a Molière radius of 2.19 cm. The ECAL is divided in the barrel region ($\eta < 1.479$, at a radius of 1.3 m) and the endcaps ($1.479 < \eta < 3$). The 61200 crystals employed in the barrel region, whose size is (22×22) mm$^2 \times 23$ cm, have a radiation length of $25.8X_0$; the 7324 crystals in the endcaps, 28.6×28.6 mm$^2 \times 22$ cm, have a radiation length of $24.7X_0$. Before the endcaps, on each side, a pre-shower detector is installed: it is composed by two disks of lead absorber and two layers of silicon strips, of a radiation length up to $3X_0$. The pre-shower calorimeter has been designed to distinguish the photons coming from the π^0 decay, from the photons produced in the rare Higgs decay $H \to \gamma\gamma$. The readout and amplification of the scintillating light, performed by avalanche photodiodes in the barrel and by vacuum phototriodes in the endcaps, requires a stable temperature of 18° C, mantained by a water cooling system.

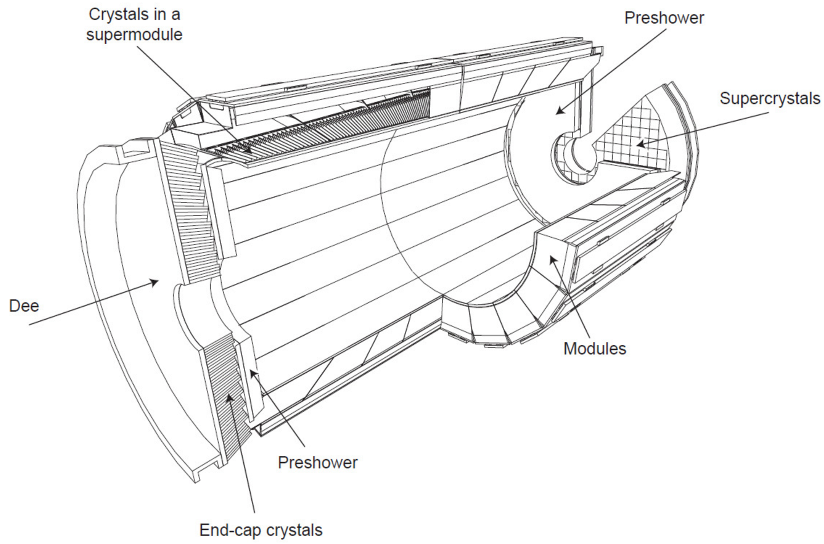


Figure 3.11: The CMS electromagnetic calorimeter.

The energy resolution of the calorimeter is parametrized as:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2, \tag{3.7}$$

where $S = 0.018\ \mathrm{GeV}^{\frac{1}{2}}$ is the stochastic term, $N = 0.04$ GeV is related to noise contribution, and $C = 0.005$ is a constant term depending on the calibration.

3.2.5 The hadronic calorimeter

The hadronic calorimeter (HCAL, displayed in fig. 3.12) [51] is a sampling calorimeter, composed by brass and plastic scintillator layers. It has been designed to guarantee a good hermeticity, allowing to perform a precise measurement of the missing transverse energy. It is located within the electromagnetic calorimeter and the solenoid, covering a region of $|\eta| < 1.3$ in the barrel, and $1.3 < |\eta| < 3$ in the endcaps. Brass is non-magnetic and has short interaction length (16.4 cm): the 60 mm thick absorber layers used in the barrel reach 5.6 interaction lengths at $\eta = 0$ and 10.8 interaction lenghts at $\eta = 1.3$; the 80 mm thick layers in the endcaps reach 11 interaction lenghts. An additional calorimetric layer has been installed out of the solenoid, in order to reach 11.8 interaction lenghts in the barrel region. The scintillation light, typically in the blue-violet region of the electromagnetic spectrum, is collected by wavelenght-shifter fibers, translated and amplified by multi-channel hybrid photodiodes, proportionally to the magnitude of the energy deposits. An additional hadronic calorimeter (HF) has been placed in the forward region, $3 < |\eta| < 5.2$, at 11.2 m from the interaction point. It has beeen studied to afford the high levels of radiations: it is composed by 55 mm thick absorber layers of stainless-steel, and quartz fibers, able to detect the Cherenkov scintillating light of the charged particles of the hadronic showering. A longitudinally segmentation allow to distinguish hadronic particles from electromagnetic components. The energy resolution of the hadronic calorimeter is:

$$\left(\frac{\sigma}{E}\right) \approx \frac{a}{\sqrt{E}} \oplus b\%, \tag{3.8}$$

where $a = 65\%$ in the barrel region, 85% in the endcaps, 100% in the forward region, and $b = 5\%$.

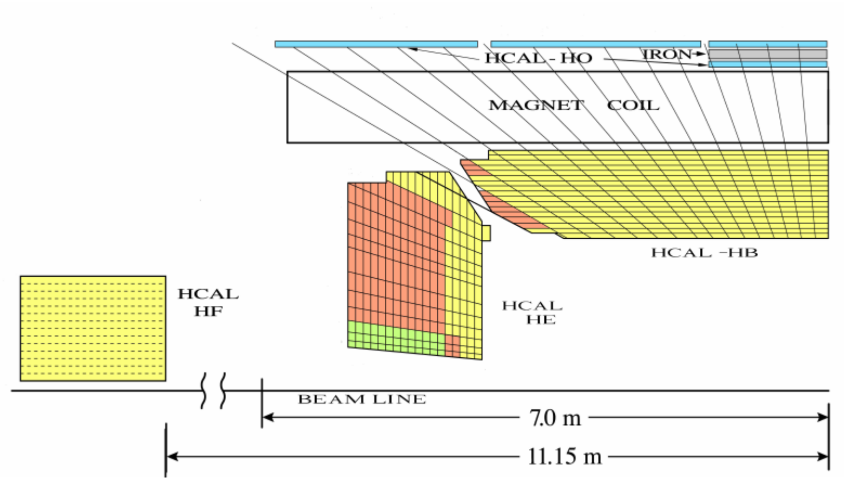


Figure 3.12: The CMS hadronic calorimeter.

3.2.6 The muon system

The outer system of the CMS experiment consists into gas detectors for identifying muons [52], that are located between the iron return yokes, designed to close the magnetic field generated by the solenoid. In the barrel region, where a smaller number of muons is expected and the magnetic field is less strong, Drift Tubes (DT) detectors are installed. In the endcaps, where the flux of particles is larger, Cathod Strip Chambers (CSC) are used, and disposed in three disks. CSCs are designed to allow faster responses, higher granularity and radiation resistance. Resistive Plate Chambers (RPC) are installed both in the barrel and in the endcaps as additional triggering system. The geometry of the muon system is shown in fig. 3.13; it consists of 250 DTs, 530 CSCs, 610 RPCs, and it covers a region $|\eta| < 2.4$.

3.2.6.1 The Drift Tubes

Drift Tube detectors cover a region of $|\eta| < 1.2$ and are arranged in four stations, segmented along the beam line in five wheels. The basic element of the detector is the cell, that has a size of 42×13 mm^2. Each cell is filled with a gas mixture (85% argon, 15% CO_2), in which the process of ionization takes places; the ionization electrons drift from the 50 μm thick steel anodic wire, located in the center of the cell, towards the aluminium cathodic strips, located at its edge. Additional electrodes on the surface of the cells allows to shape the electric field, in order to make the drift speed of the electrons uniform: the muon position is then extrapolated from the measurement of the drift time. Every station is composed by three cells superlayers. In the inner and the outer superlayers, the cells are oriented such in a way that the anodic wire is located along the z axis, to measure the φ coordinate. In the intermediate superlayer, wires are parallel to the radial coordinate, hence they can measure the z position. The spatial resolution of the system is 100 μm in the (r, φ) plane, 1 mrad in the φ coordinate, and 150 μm in the longitudinal z coordinate.

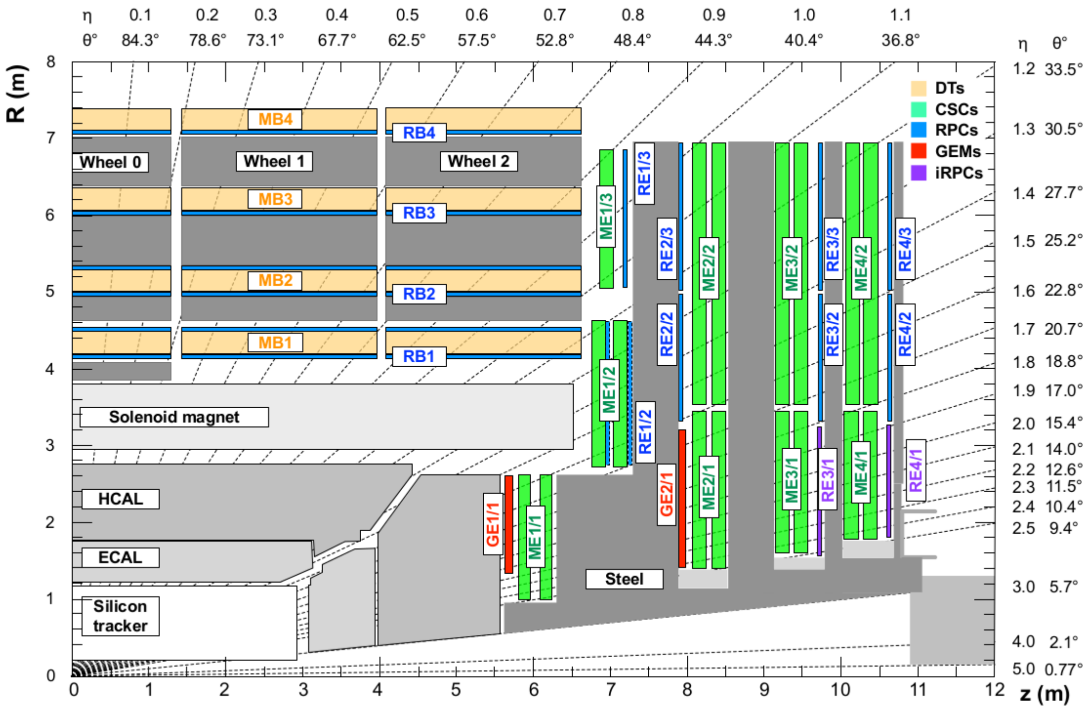


Figure 3.13: Section of CMS detector, in the plane (r, z), parallel to the beamline, that emphasizes the location of the muon detectors, in particular: Drift Tubes (DT, in yellow); Cathode Strip Chambers (CSC, in green); Resistive Plate Chambers (RPC, in blue).

3.2.6.2 The Cathode Strip Chambers

Cathode Strip Chambers cover a region of $0.9 < |\eta| < 2.4$, overlapping with the DT in the pseudorapidity range $0.9 < |\eta| < 1.2$. The anodic wires inside each CSC are installed in six planes, with the aim of measuring the radial coordinate; the wire planes are perpendicularly crossed by cathodic strips, disposed along the radial direction to measure the φ coordinate. Ionization electrons produced by muons passing through the gas mixture in the chambers migrate from the anodes, inducing a charge distribution on the cathodes, from which the azimuthal coordinate can be reconstructed. The spatial resolution in the r coordinate is 200 μm, and it is 75 – 150 μm in the (r, φ) plane. CSCs are arranged in four disks and in three concentric rings.

3.2.6.3 The Resistive Plate Chambers

Resistive Plate Chambers are located both in the barrel (disposed in six layers) and in the endcap region (three layers), up to a pseudorapidity of $|\eta| < 1.6$. These gas detectors are charged at very high voltages, in order to work in the avalanche ionization mode. The plastic resitive plates are equipped with readout strips. The spatial resolution of the detector is low (1-2 cm), but the fast timing response (2-3 ns) and good time resolution (1 ns) allow to employ RPCs as an additional triggering system and to profit of a precise measurement of the bunch-crossing time.

3.2.7 The trigger system and data acquisition

The CMS trigger system [53] has been designed considering the high instantaneous luminosity, such that it can provide a fast response and it allows to reduce the nominal event rate of 40 MHz in proton proton collision. The complexity of the CMS detector and the very high number of readout channels result into a huge amount of data per event, approaching the order of few MB per bunch crossing, hence 40 TB per second. The processes of handling and recording data are currently limited by the employed technology to a frequency of ~100 Hz. Applying online selections to skim the events that are going to be written on tape, without rejecting interesting signals of hard processes and rare phenomena, is therefore a crucial and challenging point for every data analysis. Events are filtered by trigger selections at different levels: the Level-1 (L1) trigger is an hardware device, that allows to reduce the event rate from 40 MHz to the order of 100 kHz; the High Level Trigger (HLT) is a set of software algorithms that skims the event rate down to few hundred Hz. Once the trigger decisions are taken, the final events are handled by the Data Acquisition System (DAQ), that collects the informations coming from to the sub-detectors and sends them to the storage unities.

3.2.7.1 The Level-1 trigger

The L1 trigger is an hardware device composed by customized electronics, and it accesses the informations coming from the calorimeters and the muon system, while the tracker is not considered given the excessively large bandwidth needed by its readout channels. The L1 trigger performs a first raw local reconstruction of each object, called "trigger primitive". The L1 trigger is composed by three subsystems: the calorimeter trigger, the muon trigger (divided in three independent sub-subsystems for each muon sub-detector, DTs, RPCs and CSCs), and the global trigger, that combines the informations of the former subsystems. The best quality trigger primitives reconstructed by the calorimeter and muon detectors (namely, roughly reconstructed electrons, photons, muons, jets, jets coming from the hadronic decays of tau leptons, and missing transverse energy) are handled by the global trigger, which takes the decision of discarding or keeping the event every 3.2 μs. The simplest trigger selections require the presence of a single object, whose energy or transverse momentum is higher than a certain threshold; more complicated triggers involve multiple objects or geometrical selections, that can be performed in parallel up to 128 simultaneous requirements.

3.2.7.2 The High Level Trigger

The HLT skims the L1 output rate down to few hundreds of Hz by applying a set of algorithms implemented in the same software used for the offline analysis, consisting in an event reconstruction performed exploiting the whole informations coming from all sub-detectors. The computing time is still a crucial factor, hence selections applied to HLT physics objects are generally less accurate than those of the offline analysis; furthermore, HLT can discard the event even before its full reconstruction (*i.e.* by looking only at certain region of the detectors). Events filtered by the HLT decisions are assigned to precise trigger paths and recorded in different categories of datasets.

3.2.7.3 Data acquisition, computing and storage

The DAQ system deals with the storage, transfer and handling of the data collected by CMS; it also supports and stores the data simulations and calibrations of the sub-detectors. The CMS computational resources are located in worldwide distributed data nodes, called Tiers. The CMS software (CMSSW) is based on an object oriented architecture (mainly C++). The basic unity of every data, both real and simulated ones, is the Event, that could contain very rough informations (RAW data format) or higher level refined objects (AOD, Analysis Object Data) where all the calibrations and corrections needed to properly deal with the final physics objects are already in place. Data are handled by C++ or python modules, and the outputs are written in ROOT [54] files.

3.2.8 Particle Flow event reconstruction

The particle flow (PF) algorithm [55] aims at identifying and reconstructing each particle produced by the proton-proton collisions, combining the informations coming from all the CMS sub-detectors. It is particularly suitable to improve the reconstruction of jets, missing transverse momentum (used to identify neutrinos) and hadronically decaying tau leptons.
The association of the informations is performed at different stages. The reconstruction of the charged particles in the silicon detector is performed with an iterative algorithm, and the reconstructed object is called a tracker track. Then, a clustering algorithm is performed to collect and combine the energy deposits in the calorimeters, in such a way to distinguish neutral from charged particles, reconstruct their directions, and improve the energy measurement of the very energetic charged particles, whose tracks are less bended by the magnet and hence less precisely determined. The last informations are provided by the hits collected in the muon system. The three sets of reconstructions are then combined with a link algorithm, that aims at associating tracker tracks to calorimeter clusters and muon hits with geometrical criteria. A track in the silicon detector is linked to a calorimeter cluster if the extrapolated position lies in the cluster itself. Similarly, clusters in different calorimeters are linked when the position in the more granular calorimeter (*i.e.* ECAL) lies in the envelope of the clusters in the less granular calorimeter (*i.e.* HCAL). The decision of linking a tracker track to a muon track is based on the χ^2 of a global fit between the two tracks.
The particle flow algorithm then interprets the collected and linked informations as different particles. Muons are identified by the combination of a track in the silicon detectors and a track in the muon chambers. Photons are determined directly by ECAL clusters. Electrons energies and positions are measured by ECAL clusters, linked to a corresponding tracker track, and considering all the energy clusters produced by the bremsstrahlung photons radiated while interacting with the detector material. The hadrons are identified by the tracks (if charged) linked to the corresponding ECAL and HCAL clusters. The hadron energy resolution, 10% at 100 GeV combining ECAL and HCAL, is such that neutral hadrons can be distinguished as an energy calorimetric excess when overlapped by a charged hadron occupying the same calorimetric towers. Finally, the missing transverse momentum is defined as the negative sum of the transverse momenta of all the particles identified by the PF algorithm.

3.2.9 Physics objects

3.2.9.1 Track reconstruction

The reconstruction of the trajectories of the charged particles passing through the CMS detector is performed by multiple iterations of the Combined Track Finder algorithm, that is based on a Kalman filter approach [56]; given the high multiplicity of particles produced at each bunch crossing and the multiple scatterings in the detector materials, tracking represents a challenging task. The CTF algorithm builds a track starting from the so-called seeds, namely triplets of hits collected in the pixel detector inner layers, or couples of hits if the track originates from the interaction point. The initial guess of the track given by the seeds is then extrapolated to the outer layers: if other hits are found to be compatible with the trajectory hypotesis (χ^2-based hypotesis test), they are added to the track. Once the outer layers are reached, another reconstruction is performed backward, in order to clean the track from spurious hits and enhance the tracking efficiency. The final collected hits are re-fitted with Kalman filter and more precise algorithms, in order to improve the quality of the measurement. If two tracks share more than a half of their hits, the worst quality track is rejected. The track reconstruction efficiency for particles with $p_T > 0.9$ GeV is 94% in the barrel and 85% in the endcap region [49].

3.2.9.2 Vertices reconstruction

The reconstruction of the vertices at each bunch crossing is performed in steps. Primary vertices are originating from the proton proton collisions, whilst secondary vertices are due to long-lived particles (heavy quarks and τ leptons). The starting point of the procedure is clustering the reconstructed tracks originating from the primary vertex; the decision is taken by the deterministic annealing algorithm [57], taking as input the longitudinal impact parameter. The algorithm allows to distinguish vertices separated down to 1 mm. The second step is run by the adaptive vertex fitter [58], that measures the position of the vertex for the chosen set of tracks. The algorithm is based on an iterative re-weighted Kalman filter, that down-weights the wrongly associated tracks not compatible with the considered vertex. The primary vertex is selected as the vertex where the sum of the p_T^2 of the associated tracks is the largest. The spatial resolution on the vertex position is 10-40 μm in the (r,φ) plane, and 15-50 μm in the longitudinal coordinate.

3.2.9.3 Electrons and photons reconstruction

Electrons are reconstructed [59] combining a track with the energy deposits clustered in the ECAL, due to the showering of the electron through the detector and the emission of bremsstrahlung photons. The combination can proceed both from the silicon detector in the outgoing direction and in the opposite way: the tracker seeding as starting point is suitable for low energy electrons, whose trajectories are less bended and hence more accurately measured by the tracker system; the grouping of ECAL clusters (called superclusters) followed by a consecutive track extrapolation, performed by taking into account the electron interaction with the detector material, is more efficient in case of high energetic electrons, due to the higher resolution of the ECAL scintillating crystals. A Gaussian-sum filter algorithm (GSF) [60] allows to properly take into account the effects of the bremsstrahlung radiation, that is distributed not as a single Gaussian (standard Kalman filters) but rather as a sum of Gaussian functions.

The identification of an electron relies on three groups of variables: observables built by combining measurements performed in the silicon detectors and in the calorimeter; purely calorimetric observables; purely tracking informations. Different selections are used for electron candidates found in the barrel or in the endcaps, and they can vary from loose criteria (high detection efficiency but less purity, namely more contamination from object mis-identified as electrons) to tight criteria. Data and Monte Carlo simulations reproducing Z, Υ and J/Ψ decays in e^+e^- are used to study the optimal working points, each one targetting at a different purity.

The electron energy is determined correcting the raw energy measurement of the ECAL superclusters by taking into account the effects of the losses due to radiation or gaps between the calorimeter modules, and the pile-up contribution. The electron momentum resolution has been measured in $Z \to e^+e^-$ decays in Run 1 LHC data, and it varies from 1.7% to 4.5% depending on the pseudorapidity range [61]. The electron isolation variable is defined as the p_T sum of the charged and neutral particles lying in a cone of $\Delta R = 0.3$ around the electron trajectory, divided by the transverse momentum of the electron itself:

$$I^e_{\Delta R=0.3} = \frac{\sum_{\text{char. hadrons}} p_T + max\left[0, \sum_{\text{neut. hadrons}} p_T + \sum_{\text{photons}} p_T - 0.5 \sum_{\text{pile-up char. hadrons}} p_T\right]}{p_T^e}; \tag{3.9}$$

the contribution of the pile-up charged particles is removed. The isolation variable is used to distinguish electrons coming from the leptonic decays of electroweak bosons (low $I^{e}_{\Delta R=0.3}$) from electrons coming from the decays of heavy fermions, when they are more likely produced in association with light flavour jets and hence topologically close to calorimetric deposits due to hadrons (high $I^{e}_{\Delta R=0.3}$).

Photons are reconstructed with the ECAL clusters only. Given their importance in the discovery of the Higgs boson, dedicated studies have been performed both in data and in Monte Carlo simulations reproducing the $H \to \gamma\gamma$ process. Particular care has been taken in the treatment of the photon conversions into electron-positron pairs while interacting with the tracker detector. Dedicated selections allow to define different photon identification working points. Similarly to the case of the electrons, the photon isolation variable can be defined. The photon energy resolution varies from 1% to 3%, depending on the η range [62].

3.2.9.4 Muon reconstruction

A muon candidate can be built exploiting the hits collected in the silicon tracker (track) and in the muon system (standalone muon) [63]. Each muon sub-detector (DTs, RPCs and CSCs) performs a local reconstruction of the particle candidate; the informations from the three muon chambers are combined with a Kalman filter approach.
Three different strategies are adopted to define a muon candidate in the CMS detector. A standalone muon is reconstructed by using only the local reconstruction in the muon chambers. A tracker muon is built starting from a track in the silicon detector, that is extrapolated up to the muon chambers, taking into account the multiple scattering and the energy loss through the material. The tracker muon is defined if at least one segment, *i.e.* a short track built with CSCs or DTs hits, is matched to the starting track. This technique is the most efficient for the reconstruction of low energetic muons. A global muon is built starting from a standalone muon, and then its trajectory is extrapolated towards the inner layer of the silicon detector and eventually matched to a track; this approach is suitable for high energetic muons ($p_T > 200$ GeV).
Different algorithms are used to assign a momentum to the muon candidate, in order to mitigate the effects of bremsstrahlung, that becomes significant when the muon approaches energies of the order of 1 TeV. The radiated photons generate spurious hits in the chambers and larger occupancy, significantly deteriorating the momentum measurement.
Starting from 2016 LHC Run, the muon reconstruction takes into account the Alignment Position Errors, namely the uncertainties due to the position of the muon chambers with respect to the silicon detectors. The final resolution on the muon momentum measurement depends on the p_T and η of the candidate, and ranges from 1% for very low momenta, up to ~7% ($|\eta| < 0.9$) – 10% ($1.2 < |\eta| < 2.4$) [64].
The muon isolation $I^{\mu}_{\Delta R=0.4}$ is defined similarly to the electron isolation, but by taking into account a larger cone $\Delta R = 0.4$ around the muon direction.

3.2.9.5 Jet reconstruction

The nature of the strong interaction is such that coloured partons, namely quarks and gluons, are forced to aggregate to form a color-neutral hadron, in the process called hadronization. Therefore, partons cannot be observed as free particles in a detector, but rather as collimated jets of hadronic particles.
Jets are reconstructed starting by the PF candidates in the event. The charged hadron subtraction algorithm (CHS) removes candidates not associated to the primary vertex in order to suppress pile-up

contributions [65]. The remaining particles are used as input to jet clustering algorithms to reconstruct particle-flow jets. The jets are clustered using the FASTJET package [66] with the anti-k_T jet sequential clustering algorithm [67]. A sequential clustering algorithm is designed to be infrared and collinear safe, namely, if the final state particles undergo a soft emission or a collinear gluon splitting, the number and shapes of the jets should not change. The starting point of a sequential clustering algorithm is the definition of the distances bewtween two particles i and j, and the distance of a given particle i from the beam-spot B:

$$\begin{aligned} d_{ij} &= \min\left(p_{T,i}^{2a} p_{T,j}^{2a}\right) \frac{R_{ij}^2}{R_0^2}, \\ d_{iB} &= p_{T,i}^{2a} \end{aligned} \tag{3.10}$$

where $p_{T(i,j)}$ are the transverse momenta of the particles, $R_{ij}^2 = \left(\mathscr{Y}_i - \mathscr{Y}_j\right)^2 + \left(\varphi_i - \varphi_j\right)^2$ is the angular distance between the particles, a is an exponent depending on the clustering algorithm chosen, and R_0 is the clustering parameter. The algorithm then operates as follows:

- it computes all the possible combinations of distances d_{ij} and d_{iB} and it finds the minimum;
- if the minimum is d_{ij}, the four-momenta of the particles i and j are summed up in one candidate ij; i and j are removed from the list of available particles, the distances are updated, and the algorithm proceeds to re-calculate all the possible remaining d_{ij};
- the clustering stops when the smallest quantity is d_{iB}: i particle is defined as one jet, and it is removed from the list of particles;
- this process is repeated until all the particles are assigned to a jet, that must be separated from another jet at least by a distance $R_{ij} > R_0$.

If the anti-k_T algorithm is applied, the exponent $a = -1$. This means that it tends to cluster high p_T particles first, given that the hard term dominates d_{ij} in equation 3.10. Since the soft particles have lower impacts, the shape of the jet is not sensitive to the soft radiation and rather stable against the softer pile-up contributions.
In this analysis, clustering parameters of $R_0 = 0.8$ and $R_0 = 0.4$ will be used to define the “fat”-jets or AK8 jets, and the “standard”-jets or AK4 jets. In order to avoid double-counting of PF candidates, AK4 jets are considered only if the angular separation from the leading AK8 jet is larger than $R_0 > 0.8$. Since the detector response to different particles is non-linear, particular care should be taken in the assignement of the measured momentum of the clustered jet to the corresponding true value of the original parton [68]. A set of jet energy corrections (JECs) are applied sequentially and with a fixed order. Each correction constists in a rescaling of the jet four-momentum, and it takes into account different effects that are factorized.

- The L1 JECs remove the effect of the pile-up; they consist into an offset correction of the jet p_T. They are determined from Monte Carlo (MC) simulations of di-jet events produced by strong interaction with and without pile-up events on top, and parametrized as a function of kinematical variables (jet area, pseudorapidity and p_T) and of the average p_T density per unit area, ρ. Residual differences between data and the detector simulation are evaluated in data collected with a random trigger, called zero bias, applying the only requirement of the beam crossing happening. Pile-up offset corrections are displayed in fig. 3.14 (top left), as a function of the jet pseudorapidity.

- The simulated response of the detector is not uniform over jet p_T and η. This effect is mitigated by the L2L3 MC-truth corrections. They are calculated in MC simulations of di-jet events, by taking into account the discrepancy between the reconstructed p_T of the jet and the true p_T at particle generator level (*i.e.*, before simulating the interaction of the parton showers with the detector), as a function of jet p_T and η. L2L3 scale factors describing the simulated jet response are reported in fig. 3.14 (top right), as a function of the jet pseudorapidity.

- The small data-MC discrepancies (~1%) left after applying the previous set of JECs are corrected by the L2 and L3 residual corrections. The L2Residuals are calculated in di-jet events, as a function of p_T. The L3Residuals are calculated in $Z \to (\mu\mu, ee)$ + jet events, photon + jet events and multijet events, as a function of η and p_T [68]. Data-MC scale factors for L2L3Residuals are displayed in fig. 3.14 (bottom), as a function of the jet η and p_T.

- An optional correction, not used in this analysis, is the L5 flavour-dependent correction, that is extracted from MC simulations.

Each jet energy correction is determined with an uncertainty, and reported in fig. 3.15 for 2015 data, as a function of p_T and η of the jet. The total uncertainty for jets with p_T larger than 30 GeV (100 GeV) is smaller than 3% (1%) in the barrel, and up to 5% (3%) in the endcaps [69].
An additional effect that must be taken in account in the analysis is the discrepancy in the jet energy resolution (JER) observed in data and in Monte Carlo samples. A smearing procedure is applied in MC simulations (described in detail in sec. 4.3.6), in order to restore a better agreement. Jet energy resolutions in Monte Carlo simulations are displayed in fig. 3.16 (top), as a function of the jet p_T and the average number μ of reconstructed primary vertices, considering central (left) and forward (right) jets. The resolution is stable against the pile-up for jet p_T >100 GeV, and it ranges from 10% at 100 GeV, down to 4% at 1 TeV [69]. In fig. 3.16 (bottom), data-MC smearing scale factors are reported as a function of η.

3.2.9.6 Tau reconstruction

Tau leptons have a very small lifetime ($\sim 3 \times 10^{-13}$ s), hence they decay before reaching the pixel detector and they can only be reconstructed through their decay products. Approximatively 60% of the times, τ leptons decay in hadrons, hence they are reconstructed as small collimated jets in the CMS detector. The main decay modes of the hadronic tau, τ_h, are one or three charged mesons (mainly $\pi^{\pm}$), also in association with a π^0 decaying in a couple of photons, and a τ neutrino. Hence, photons and charged hadrons are the main ingredients of dedicated algorithms to perform the τ_h reconstruction and identification, in order to distinguish them from quark and gluon-initiated jets. The main CMS τ_h reconstruction algorithm, Hadron Plus Strips (HPS) [70], is particle-flow based. HPS builds the tau candidate from a PF jet, clustered with the anti-k_T algorithm with $R_0 = 0.5$, and it reconstructs the $\pi^0 \to \gamma\gamma$ decays within the jet cone, by taking into account the photon conversions in the silicon detector. The exploitation of the PF informations is such that the HPS algorithm shows stable performances in the reconstruction of the τ_h energy as a function of the energy itself. The τ_h candidate is required to be isolated, namely no energy deposits other than the τ decay products should be present in the tau cone. Depending on the low threshold set to consider the surrounding particles as included in the cone, different isolation working points can be defined. With the looser working point, the probability of mis-identifying a quark or gluon jet as a tau is around 1% [70].

3.2.9.7 b-jets tagging

The bottom quark plays a fundamental role in numerous standard model processes, *i.e.* the physics related to the top quark (that decays into a W and a b-quark with a branching fraction of 100%) and

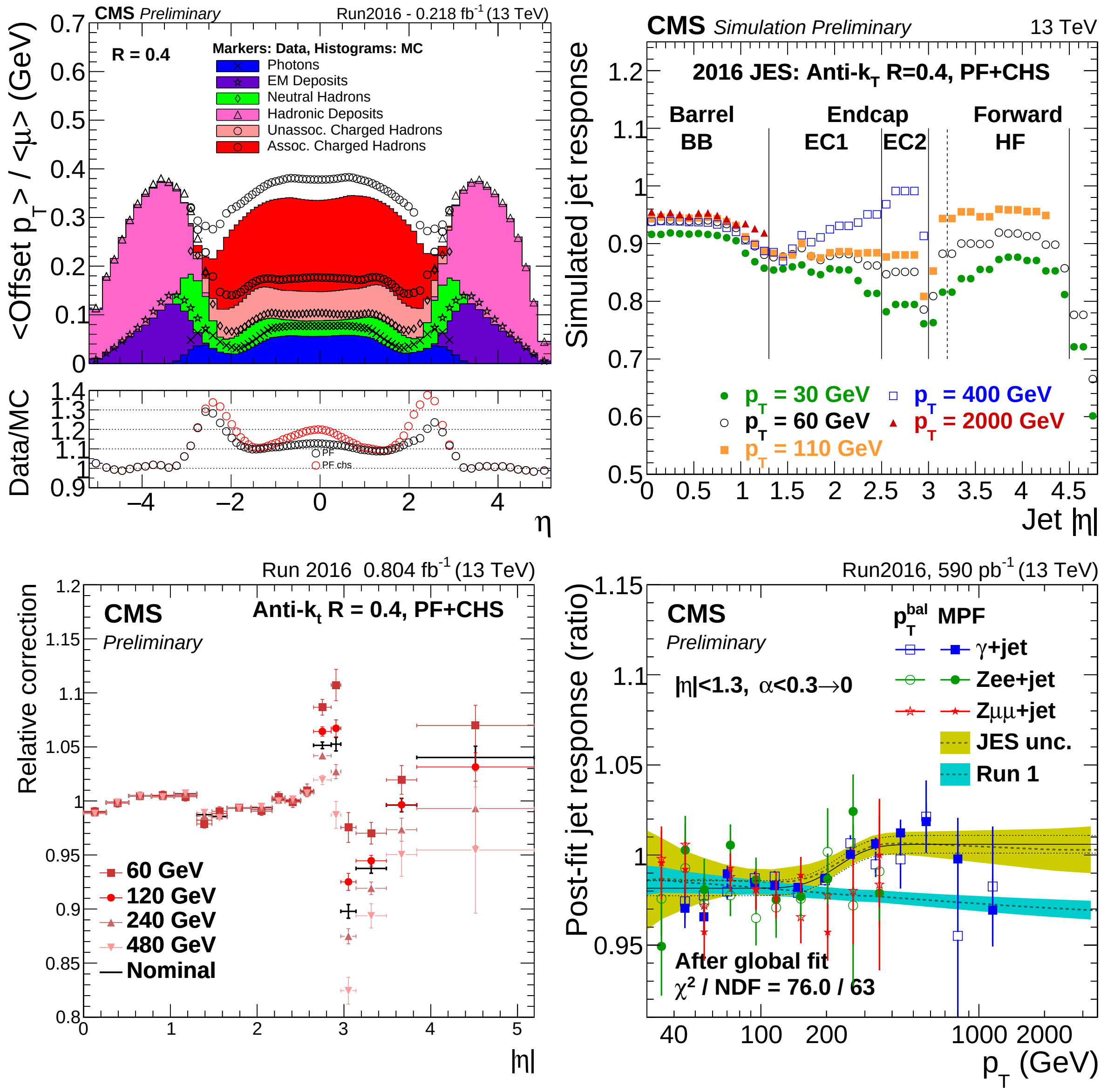


Figure 3.14: Top left: average p_T offset due to additional pile-up events, measured both in data and in MC simulations, as a function of the jet pseudorapidity. Top right: simulated jet response (L2L3 MC-truth corrections), as a function of the jet pseudorapidity. Bottom left: L2L3 residual data-MC corrections, evaluated on di-jet events, as a function of the jet η. Bottom right: L2L3 residual data-MC corrections, evaluated on di-jet and $Z/\gamma+$jet events, as a function of the jet p_T.

the Higgs boson (decaying into $b\bar{b}$ with a branching fraction ∼60%). Many algorithms have been exploited by the CMS collaboration, with the aim of distinguishing a b-quark initiated jet and jets originating from light quarks or gluons [71]. The most remarkable feature of the b-quark is the long lifetime (∼ 1.5 ps), that has the experimental consequence of a displaced decay (few mm) with respect to the primary vertex. The direct leptonic decays of the b-quark (into μ and e) or the cascade leptonic decays involving c-quarks give an additional handle to its identification.
Given the high spatial resolution of the silicon detector, track reconstruction is a key point of the b-tagging procedure. Tracks inside a jet candidate must satisfy criteria related not only to their quality but also on their distance from the interaction point. The track impact parameter is the distance

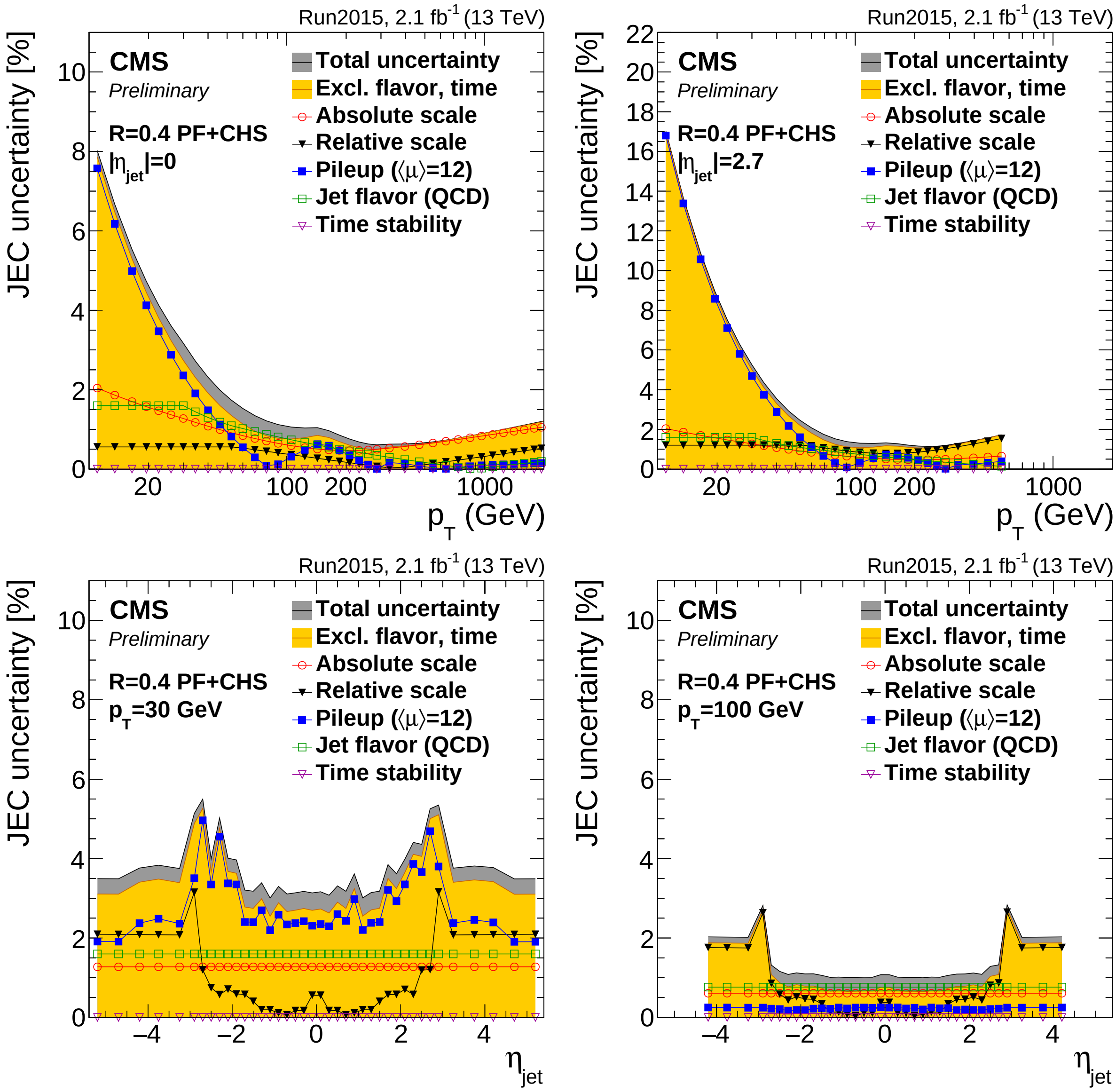


Figure 3.15: Jet energy corrections uncertainties, as a function of jet p_T (top) and η (bottom), calculated in 2015 data. The yellow histograms report the convolution of the uncertainties applied in the analysis.

between the primary vertex and the coordinate of closest approach. Tracks that are too far from the interaction point are discarded, in order to suppress the pile-up contributions. The Combine Secondary Vertex (CSV) algorithm [72] sorts jet candidates in categories, based on the number of reconstructed secondary vertices (one reconstructed secondary vertex, no secondary vertices but two tracks with high impact parameters, and the remaining cases). A multivariate approach allows to train the algorithm over the categories, considering as discriminating variables both tracking informations (numbers and properties of the tracks) and their relations with the secondary vertex reconstruction (impact parameters; angular, linear, 2D and 3D distances of the vertex from the tracks and the jet axis; invariant mass of the charged particles associated to the secondary vertex). By tuning the selections, working points with different efficiencies have been set. The loose working point, used in this analysis, has a 90% signal efficiency and a 40% mis-identification rate. The

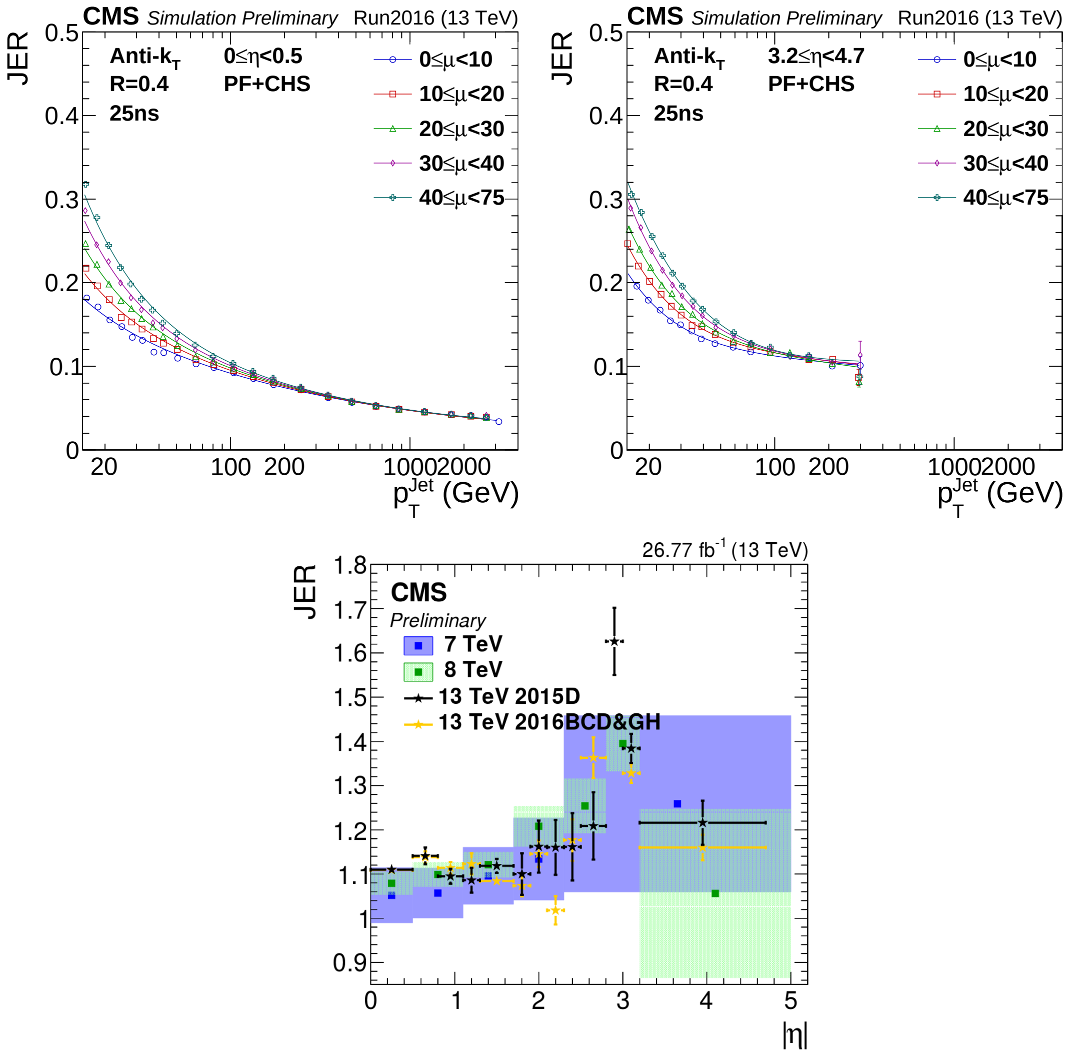


Figure 3.16: Top: jet energy resolution in MC simulations, as a function of the jet p_T. Different curves represent a different average number of primary vertices per event (μ). Bottom: data-MC scale factors, as a function of the jet η, measured in 2016 data (yellow dots).

b-tagging efficiency is different in data and in simulations. Multiplicative scale factors are calculated in events enriched in b-quark jets.

3.2.9.8 Missing transverse energy reconstruction

Neutrinos can interact with the other particles only via the electroweak interactions; hence, when a neutrino is produced in the proton-proton collisions, it passes through the CMS experiment, undetected. Its only experimental signature is the momentum imbalance ($\vec{p}_T^{\,\text{miss}}$) in the transverse plane (r, φ). The magnitude of $\vec{p}_T^{\,\text{miss}}$ vector is also called missing transverse energy, E_T^{miss}. Given its definition, it is evident that E_T^{miss} is a delicate variable to deal with, since it depends on all the other objects, on their imperfect measurements, on the detector noise and the pile-up events.
The PF E_T^{miss} is the negative sum of the transverse momenta of the PF candidates reconstructed in the event. Inefficiencies in the tracker reconstruction and non-linear responses of the calorimeters

can be corrected by propagating the jet energy corrections to $\vec{p}_T^{\text{miss}}$ [73]:

$$\vec{p}_T^{\text{miss,corr}} = \vec{p}_T^{\text{miss}} - \sum_{j \in \text{jets}} \left(\vec{p}_{T,j}^{\text{corr}} - \vec{p}_{T,j}^{\text{raw}} \right), \tag{3.11}$$

where "corr" ("raw") is related to corrected (raw) p_T of the considered jet. This correction is known as the "Type-I" correction to E_T^{miss}. Jets included in the calculation are AK4 with CHS algorithm applied to remove the pile-up contribution, they must have $p_T > 15$ GeV and less than 90% of their energy deposited in the electromagnetic calorimeter. If a muon lies in the jet cone, it is subtracted from the jet and added after the p_T correction. A similar correction is performed to correct $\vec{p}_T^{\text{miss}}$ at trigger level; in this case, a jet p_T threshold of 35 GeV is chosen.
The E_T^{miss} uncertainty depends on the topology of the final state. It is calculated per-event by factorizing $\vec{p}_T^{\text{miss}}$ in components: electrons, photons, muons, taus, jets, jets with p_T <10 GeV and all the remaining PF candidates that are not clustered inside jets, called unclustered energy. The momentum of every object is varied within its uncertainties (namely, the energy scale and resolution), and the effects are propagated to $\vec{p}_T^{\text{miss}}$. The most significant contributions in the unclustered energy is due to neutral PF hadrons and hadrons reconstructed in the forward hadronic calorimeter. The effects related to jet energy scale and unclustered energy scale are measured on simulation, in events with a top and an anti-top quarks, and amounts to 5% and 30% respectively [73].

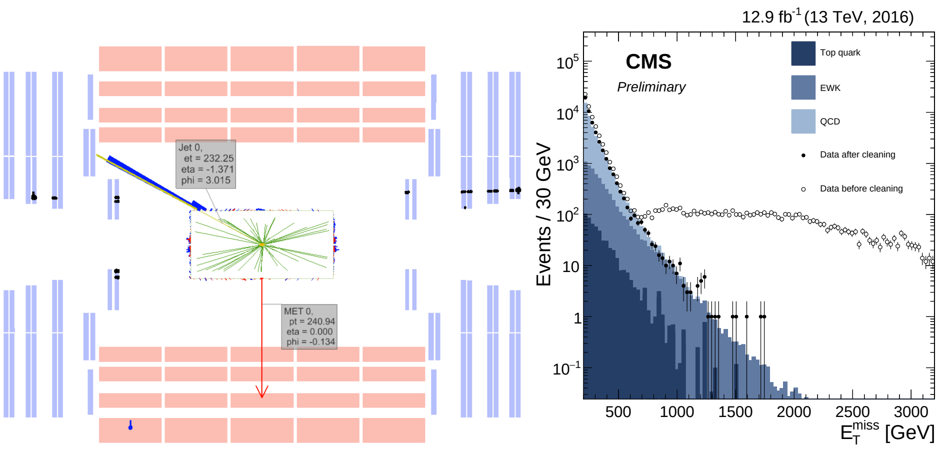


Figure 3.17: Left: event display of beam halo particles hitting the CSC detector. Right: comparison of data and simulations (histograms) when di-jet events are selected, before (open markers) and after (filled markers) anomalous E_T^{miss} cleaning algorithms have been applied on data.

Many instrumental effects can give rise to anomalous E_T^{miss} determination: they have been studied in detail during Run1 [74,75] and Run2 [73], and they are mainly caused by ECAL and HCAL. In ECAL, anomalous $\vec{p}_T^{\text{miss}}$ is caused by particles hitting the sensors of the photodetectors, or by beam halo particles (namely, particles produced in spurious proton interactions before reaching the interaction point in the detector) showering inside the calorimeter, or by losses due to ECAL dead cells. An event display representing beam halo muons hitting the CSC detector is shown in fig. 3.17 (top). In HCAL, spurious $\vec{p}_T^{\text{miss}}$ can be related to noise in the hybrid photodiodes and readout frontend. In HF, missing p_T can be related to particles lost in the light guides and photomultipliers. Additional anomalous E_T^{miss} can be produced by low quality muon tracks, that are not linked to segments reconstructed in the muon chambers by the PF algorithm. These tracks are then classified as charged hadrons, taken into account in the $\vec{p}_T^{\text{miss}}$ calculation, and result into a large amount of fake E_T^{miss}.

Dedicated algorithms have been designed to identify and reject events with anomalous E_T^{miss}, and they are consistently applied on data and simulations. In fig. 3.17 (right), Monte Carlo simulations (coloured histograms) are compared to data before the algorithms removing the anomalous E_T^{miss} have been applied (open markers) and after the cleaning (filled markers): the spurious high-$\vec{p}_T^{\text{miss}}$ tail has been suppressed.

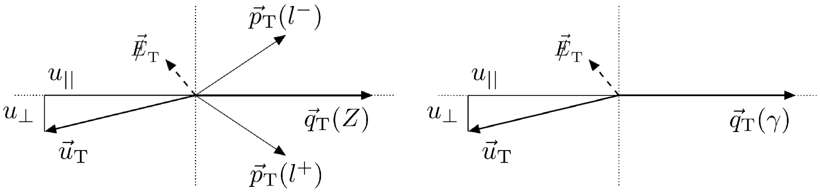


Figure 3.18: Left (Right): kinematics of $Z \to \ell\ell$ (photon) events in the (r, φ) plane; $\vec{u}_T$ is the hadronic recoil, $\vec{q}_T$ is the transverse momentum of the considered boson.

Performance of E_T^{miss} reconstruction are studied in events with a leptonic decay of a Z boson (in two muons or in two electrons) or an isolated photon. The distributions of E_T^{miss} are shown in fig. 3.19, separately for the three event categories. The hadronic recoil $\vec{u}_T$ is defined in the transverse plane as the vectorial sum of all the particle transverse momenta, except the momentum $\vec{q}_T$ of the vector boson considered (Z or γ). From the momentum conservation, the following relation holds:

$$\vec{q}_T + \vec{p}_T^{\text{miss}} + \vec{u}_T = 0. \tag{3.12}$$

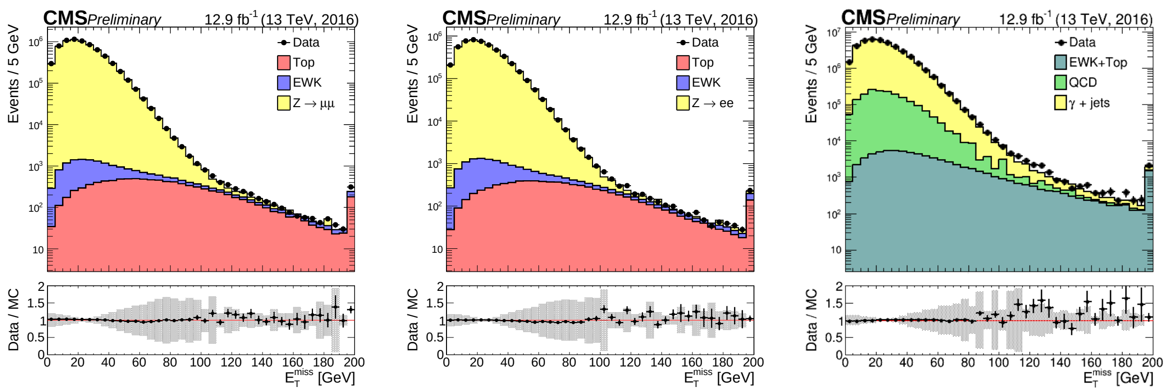


Figure 3.19: Data (black markers) and Monte Carlo (histograms) distributions of E_T^{miss} variable, in events reconstructing respectively a $Z \to \mu\mu$ decay (left), a $Z \to ee$ decay (center), an isolated photon (right).

The hadronic recoil is projected in the parallel and perpendicular directions with regards to $\vec{q}_T$. The components $u_{||}$ and $u_{\perp}$, along with the vectors described in eq.3.12, are schematically represented in fig. 3.18. The E_T^{miss} response, defined as $-\langle u_{||}\rangle / \langle q_T\rangle$, is calculated as a function of q_T in data and simulations (fig. 3.20, left). The distributions of the two components of the hadronic recoil, $u_{||} + q_T$ and $u_{\perp}$, are modelled as Voigtian functions (the convolution of a Gaussian with a Breit-Wigner). The resolution of each component is calculated as the full width at half maximum of the corresponding Voigtian, and it is displayed in fig. 3.20 (center and right plots), as a function of q_T.

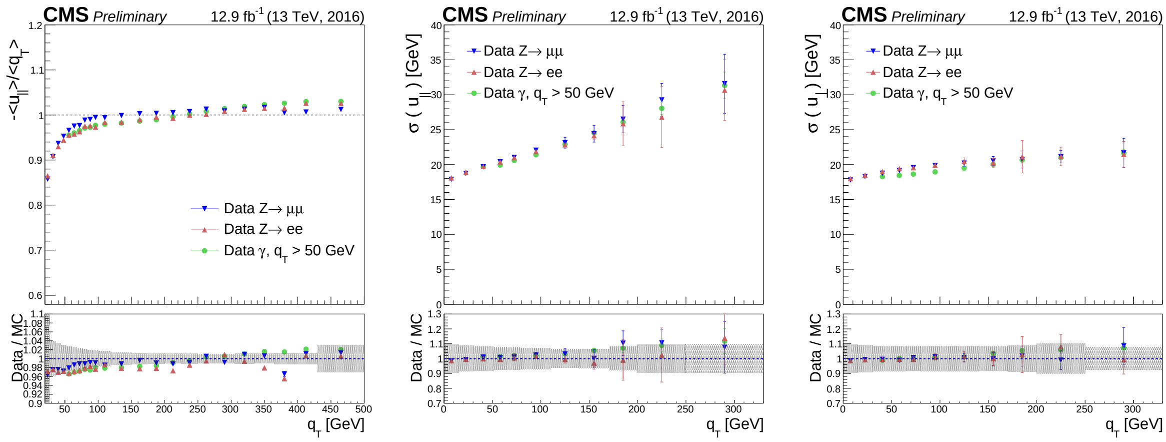


Figure 3.20: Left: E_T^{miss} response, as a function of the transverse momentum q_T of the vector boson considered in the event (Z decaying in $\mu\mu$, Z in ee or a photon), measured on data. Center and right: resolution on the measurement of the parallel and perpendicular hadronic recoil in data, as a function of q_T.

3.3 ATLAS, ALICE, LHCb detectors

3.3.1 ATLAS

ATLAS (A Toroidal LHC ApparatuS) [76] is a multi-purpose experiment, that shares the same scientifical aims of CMS. The simultaneous observation of an Higgs boson-like particle at the two experimental facilities represented an irrefutable proof of the discovery of the Higgs boson.
ATLAS has a cylindrical shape (diameter of 25 m, length of 46 m) and weights 7000 tons. Like CMS, ATLAS is composed by many sub-detectors: trackers, calorimeters and muon system. The ATLAS magnetic field is provided by a solenoid, located inside the cylinder, and a big toroid, located outside the sub-detectors, able to reach a magnetic field of 2 T at the interaction point. The main differences among the two experiments are listed below.

- *Tracker* – the CMS tracker has a better p_T resolution (mainly due to the higher magnetic field): $\sigma_{p_T}/p_T \approx 5 \cdot 10^{-4} p_T + 0.01$ at ATLAS; $\sigma_{p_T}/p_T \approx 1.5 \cdot 10^{-4} p_T + 0.005$ at CMS.

- *Electromagnetic calorimeter* – the CMS electromagnetic calorimeter is completely enclosed inside the solenoid, whilst ATLAS calorimeter is outside of the solenoid. The particles going through the solenoid suffer an energy loss and a consequent deterioration of the energy resolution. The CMS ECAL has an energy resolution of $\sigma_E/E \approx 3\%/\sqrt{E}$; the ATLAS calorimeter has a sandwich structure (liquid argon and lead layers) and a resolution of $\sigma_E/E \approx 10\%/\sqrt{E}$.

- *Hadronic calorimeter* – the CMS HCAL is partly inside the solenoid, partly outside, depauperating the resolution. The ATLAS hadronic calorimeter (made of iron and plastic scintillator tiles) has an energy resolution $\sigma_E/E \approx 50\%/\sqrt{E} + 0.03$ GeV; CMS HCAL has a resolution of $\sigma_E/E \approx 100\%/\sqrt{E} + 0.05$ GeV.

- *Muon system* – the peculiar geometry of the ATLAS muon system allows a better resolution of the standalone measurement of the muon momenta (*i.e.*, without using tracker and calorimeters), that is around 10% at 1 TeV. CMS has better performance when combining the informations coming from the inner detectors (7% at 1 TeV against the 35% for the standalone measurement).

3.3.2 ALICE

ALICE (A Large Ion Collider Experiment) [77] studies the heavy ion collisions (lead-lead) or proton-ion in order to explore the physics of the hadrons in high density (or temperature) regimes, when a new state of matter appears, the so-called quark-gluon plasma (QGP). The QGP played a crucial role in the very first instants of the life of the universe.

3.3.3 LHCb

LHCb (Large Hadron Collider beauty) [78] is a detector designed to study the b-quark properties, in particular the CP violation and other rare phenomena related to B hadrons. The final aim of these measurements is trying to solve the matter-antimatter asymmetry problem.

The three detectors are depicted in fig. 3.21.

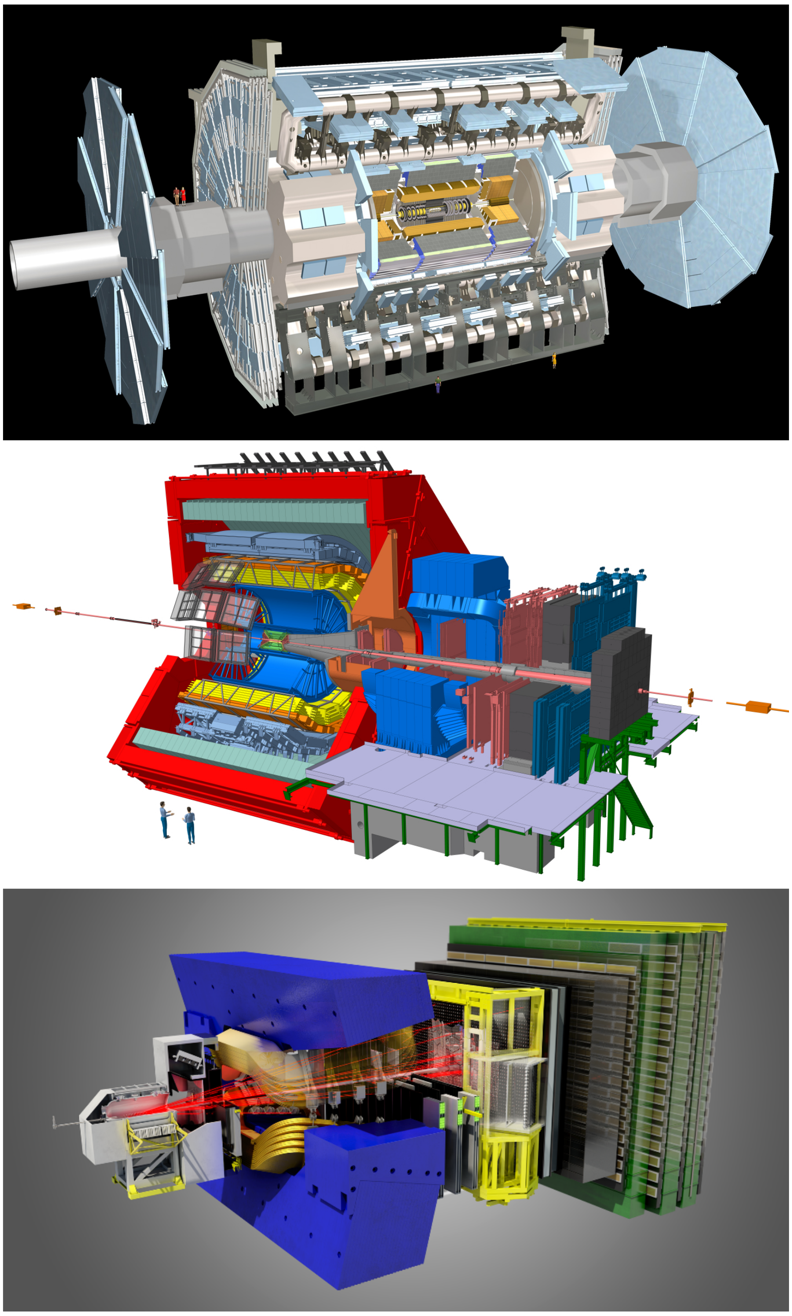


Figure 3.21: Top: the ATLAS experiment. Center: the ALICE experiment. Bottom: the LHCb experiment.

Chapter 4

Search for diboson resonances in the $\mathrm{VZ} \to \mathrm{q\bar{q}}\nu\bar{\nu}$ 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 $\mathscr{L} = 35.9\ \mathrm{fb}^{-1}$. 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 ($\vec{p}_T^{\,\mathrm{miss}}$) 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. 2.3.2), 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. 2.2), making this analysis sensitive to a generic VZ final state.

Signal events are collected with trigger paths requiring high $\vec{p}_T^{\,\mathrm{miss}}$ 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 $\vec{p}_T^{\,\mathrm{miss}}$.

The search is performed by examining the distribution of the diboson reconstructed transverse mass of the resonance VZ (m^T_{VZ}) for a localized excess. The shape and normalization of the main background of the analysis (namely, the production of an electroweak boson in association with jets) are estimated with a data-simulation hybrid approach using the distribution of data in the sidebands, corrected for a function accounting for potential differences between the signal region and the sidebands. The predictions of the secondary background sources completely rely on simulations.

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

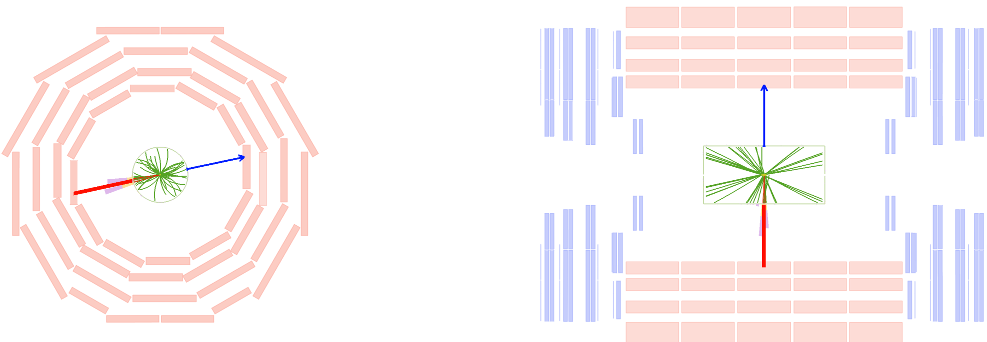


Figure 4.1: Left: representation of the decay of a W' of mass 2.5 TeV, in the transverse plane of the CMS detector. Right: representation of the decay of a W' of mass 2.5 TeV, in the longitudinal plane of the CMS detector.

4.2 Data and Monte Carlo simulations samples

4.2.1 Signal samples

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

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

The signal samples are generated at leading-order (LO) with the MADGRAPH5_AMCATNLO V 2.2.2 [79] matrix element generator, while hadronization and fragmentation are handled by PYTHIA 8 [80] version 8.2121 with CUETP8M1 [81] tuning. A full detector simulation and event reconstruction has

been performed with GEANT4 [82] and CMSSW. The detector alignement scenario, calibrations and pile-up distributions are generated according to the expectations in 2016 data.
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 ($\mathscr{B}(Z \to \nu\nu) = 0.20$, $\mathscr{B}(Z \to qq) = 0.6991$). A combinatorial factor of 2 is included in the cross-section calculation.

Signal process	m_G	Events	$\sigma \times \mathscr{B}$ (pb)
$G \to ZZ \to q\bar{q}\nu\bar{\nu}$	600 GeV	100000	0.27964
$G \to ZZ \to q\bar{q}\nu\bar{\nu}$	800 GeV	100000	0.27964
$G \to ZZ \to q\bar{q}\nu\bar{\nu}$	1000 GeV	100000	0.27964
$G \to ZZ \to q\bar{q}\nu\bar{\nu}$	1200 GeV	100000	0.27964
$G \to ZZ \to q\bar{q}\nu\bar{\nu}$	1400 GeV	100000	0.27964
$G \to ZZ \to q\bar{q}\nu\bar{\nu}$	1800 GeV	100000	0.27964
$G \to ZZ \to q\bar{q}\nu\bar{\nu}$	2000 GeV	100000	0.27964
$G \to ZZ \to q\bar{q}\nu\bar{\nu}$	2500 GeV	100000	0.27964
$G \to ZZ \to q\bar{q}\nu\bar{\nu}$	3000 GeV	100000	0.27964
$G \to ZZ \to q\bar{q}\nu\bar{\nu}$	3500 GeV	100000	0.27964
$G \to ZZ \to q\bar{q}\nu\bar{\nu}$	4000 GeV	100000	0.27964
$G \to ZZ \to q\bar{q}\nu\bar{\nu}$	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 ($\mathscr{B}(Z \to \nu\nu) = 0.2$, $\mathscr{B}(W \to qq) = 0.6760$).

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

4.2.2 Signal characterization

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

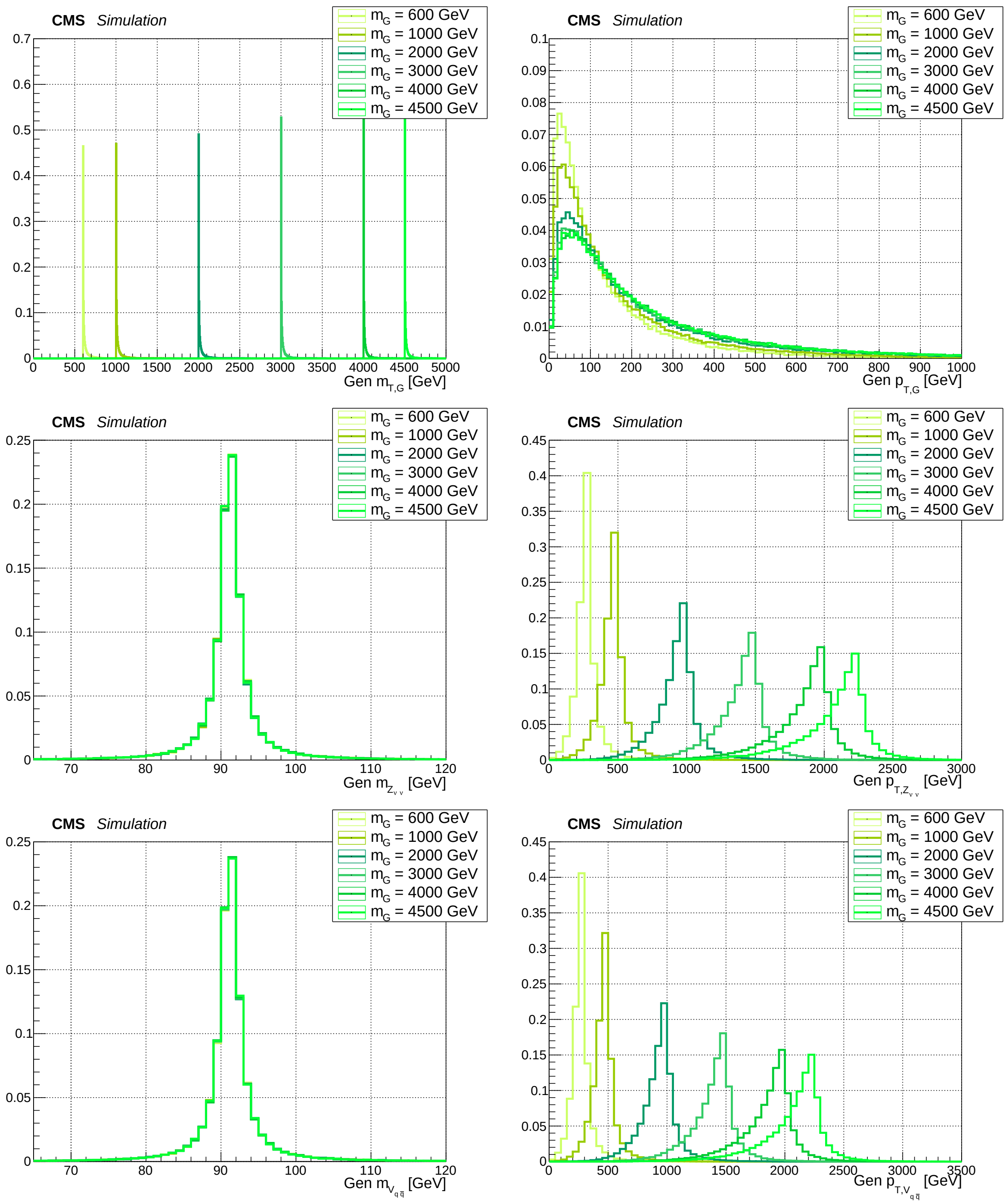


Figure 4.2: Main signal kinematic quantities at generation level after parton showering, for spin-2 bulk graviton signal, considering different mass hypoteses ($m_G = 0.6, 1, 2, 3, 4, 4.5$ TeV). Top: graviton transverse mass and p_T distributions. Center: invisibly decaying Z mass and p_T. Bottom: hadronically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

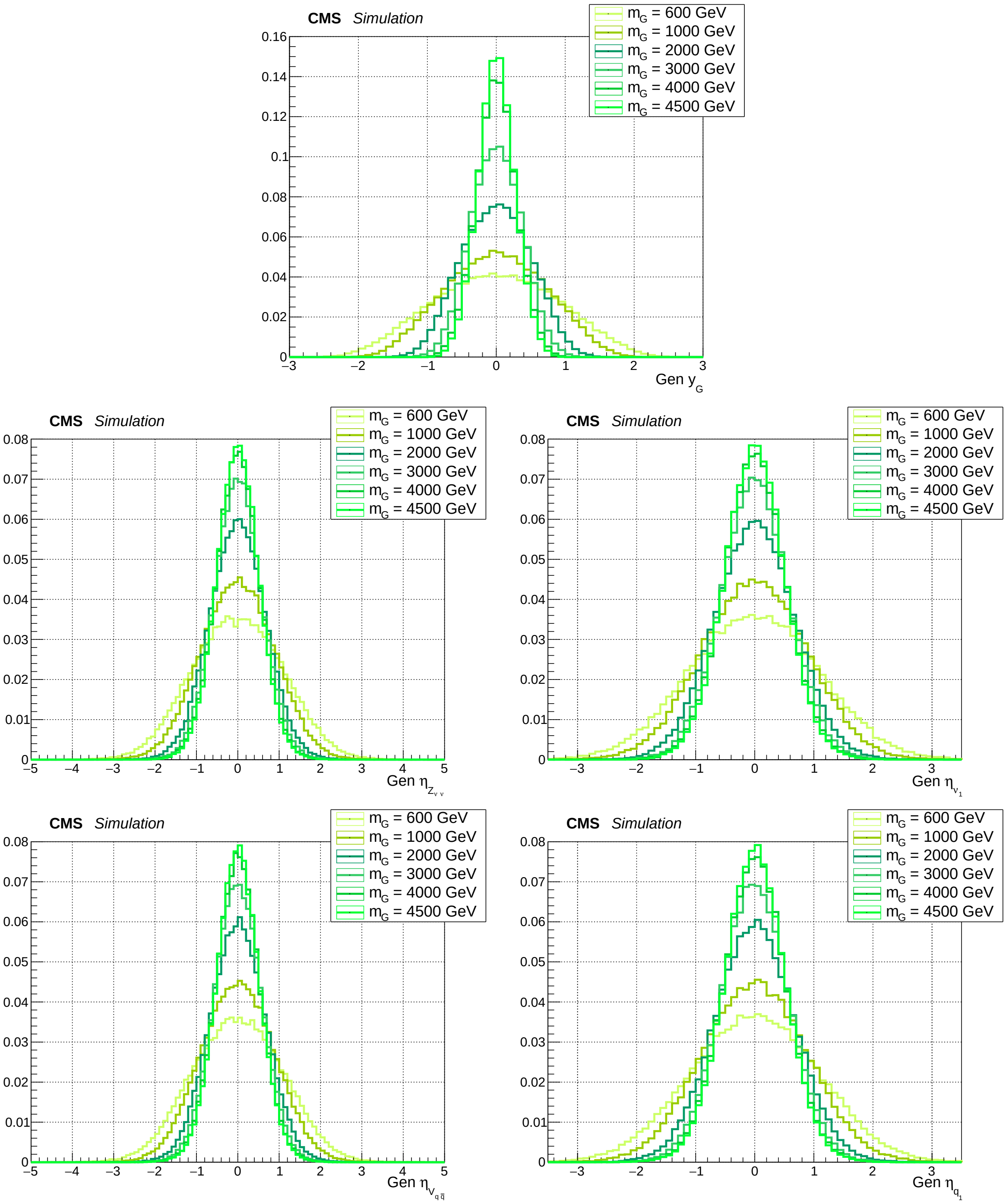


Figure 4.3: Main signal kinematic quantities at generation level after parton showering, for spin-2 bulk graviton signal, considering different mass hypoteses ($m_G = 0.6, 1, 2, 3, 4, 4.5$ TeV). Top: graviton rapidity $\mathscr{Y}$. Center: pseudorapidity η of the invisibly decaying Z, and pseudorapidity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

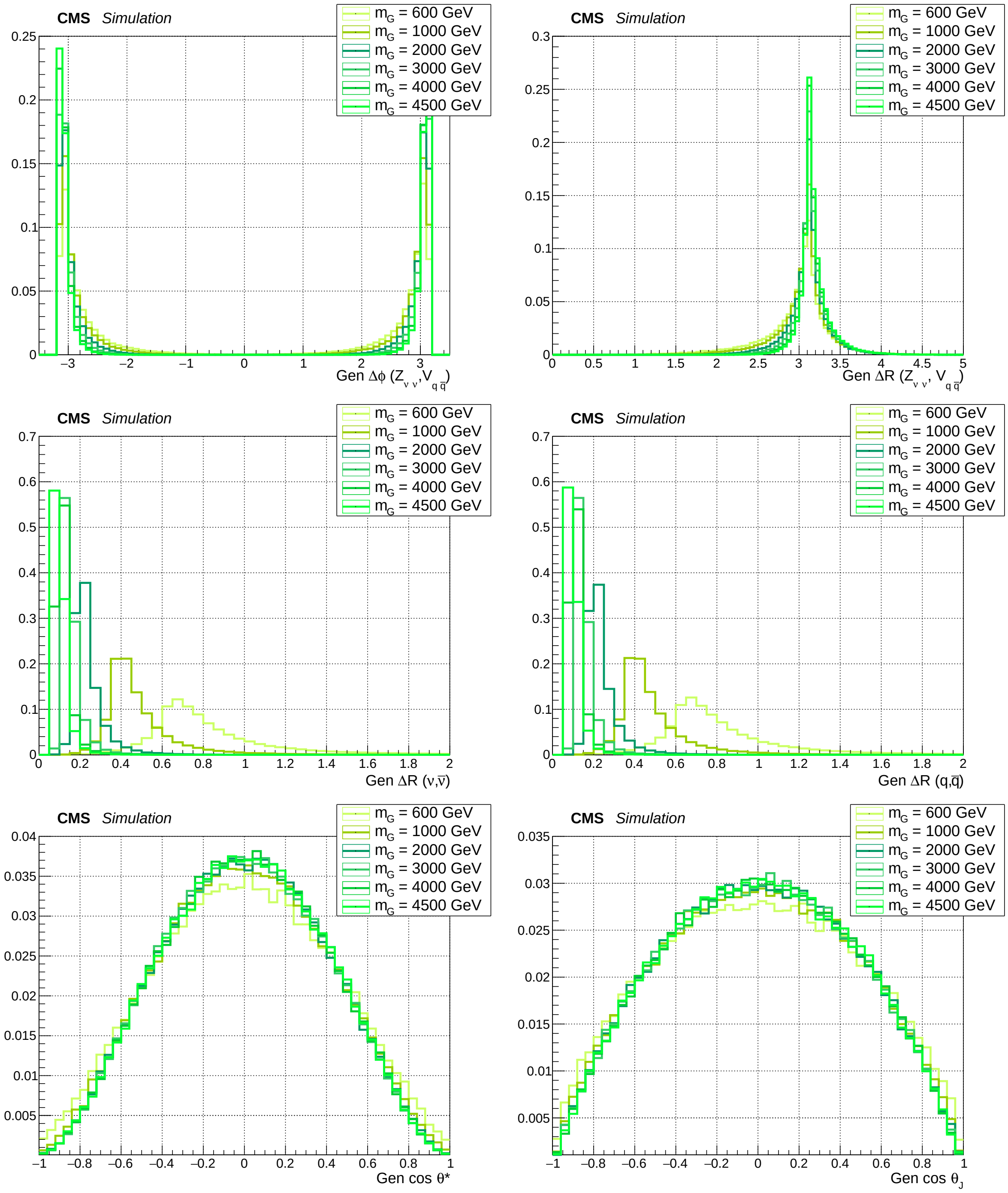


Figure 4.4: Main signal kinematic quantities at generation level after parton showering, for spin-2 bulk graviton signal, considering different mass hypoteses ($m_G = 0.6, 1, 2, 3, 4, 4.5$ TeV). Top: angular separation in the transverse plane $\Delta\varphi$ (left) and solid angle ΔR (right) between 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).

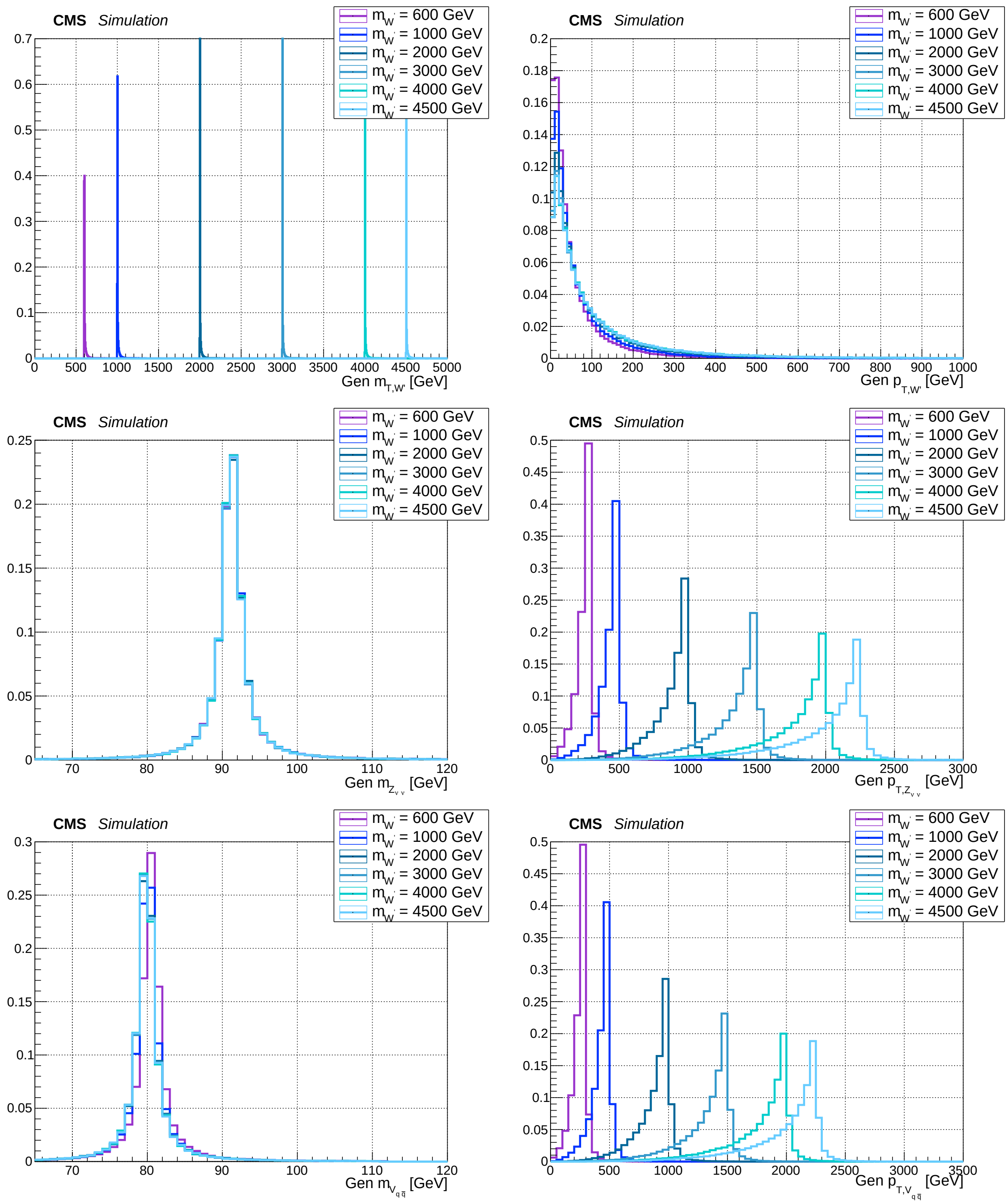


Figure 4.5: Main signal kinematic quantities at generation level after parton showering, for spin-1 W' signal, considering different mass hypoteses ($m_{W'} = 0.6, 1, 2, 3, 4, 4.5$ TeV). Top: W' transverse mass and p_T distributions. Center: invisibly decaying Z mass and p_T. Bottom: hadronically decaying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

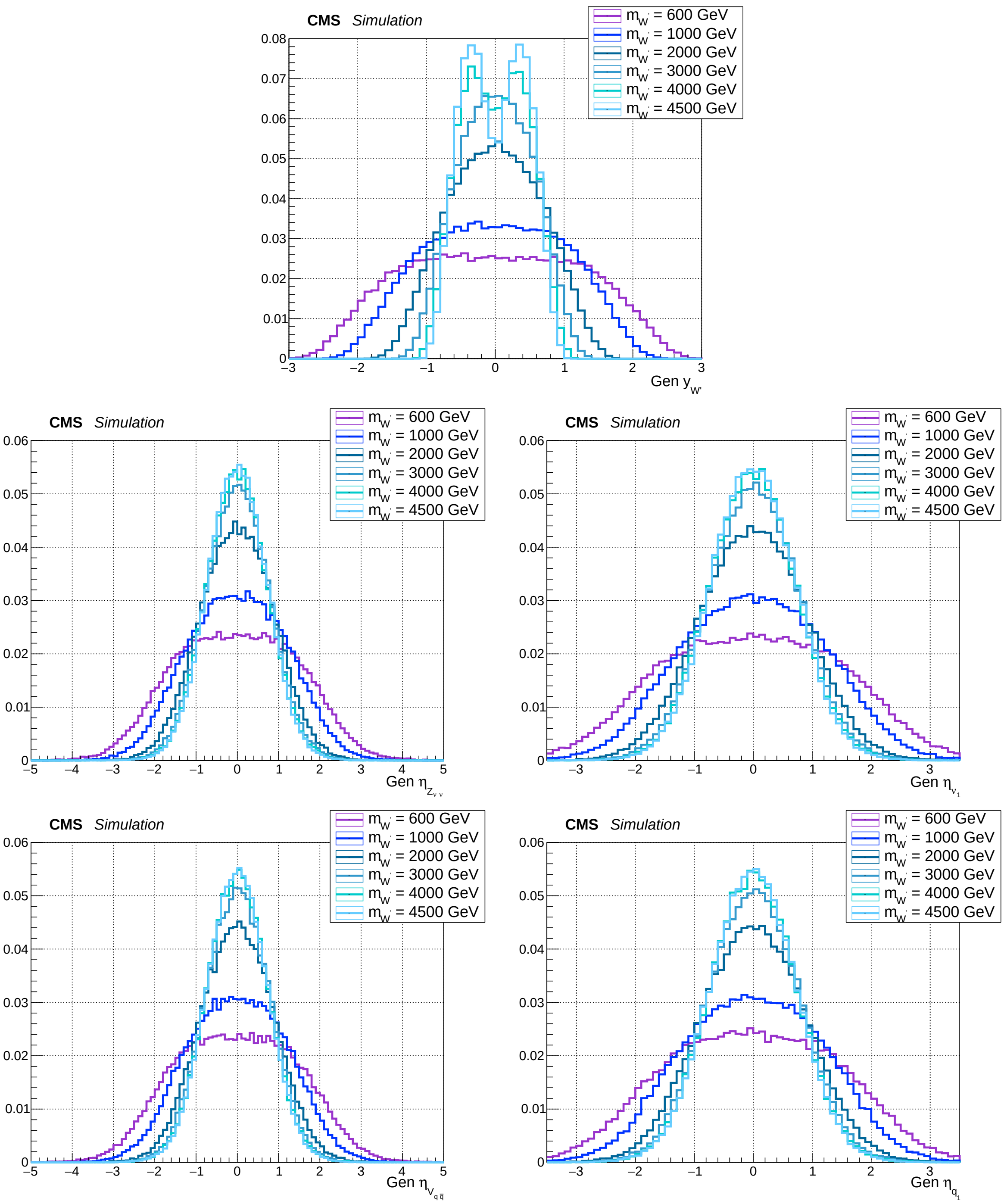


Figure 4.6: Main signal kinematic quantities at generation level after parton showering, for spin-1 W' signal, considering different mass hypoteses ($m_{W'} = 0.6, 1, 2, 3, 4, 4.5$ TeV). Top: W' rapidity $\mathscr{Y}$. Center: pseudorapidity η of the invisibly decaying Z, and pseudorapidity of the leading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

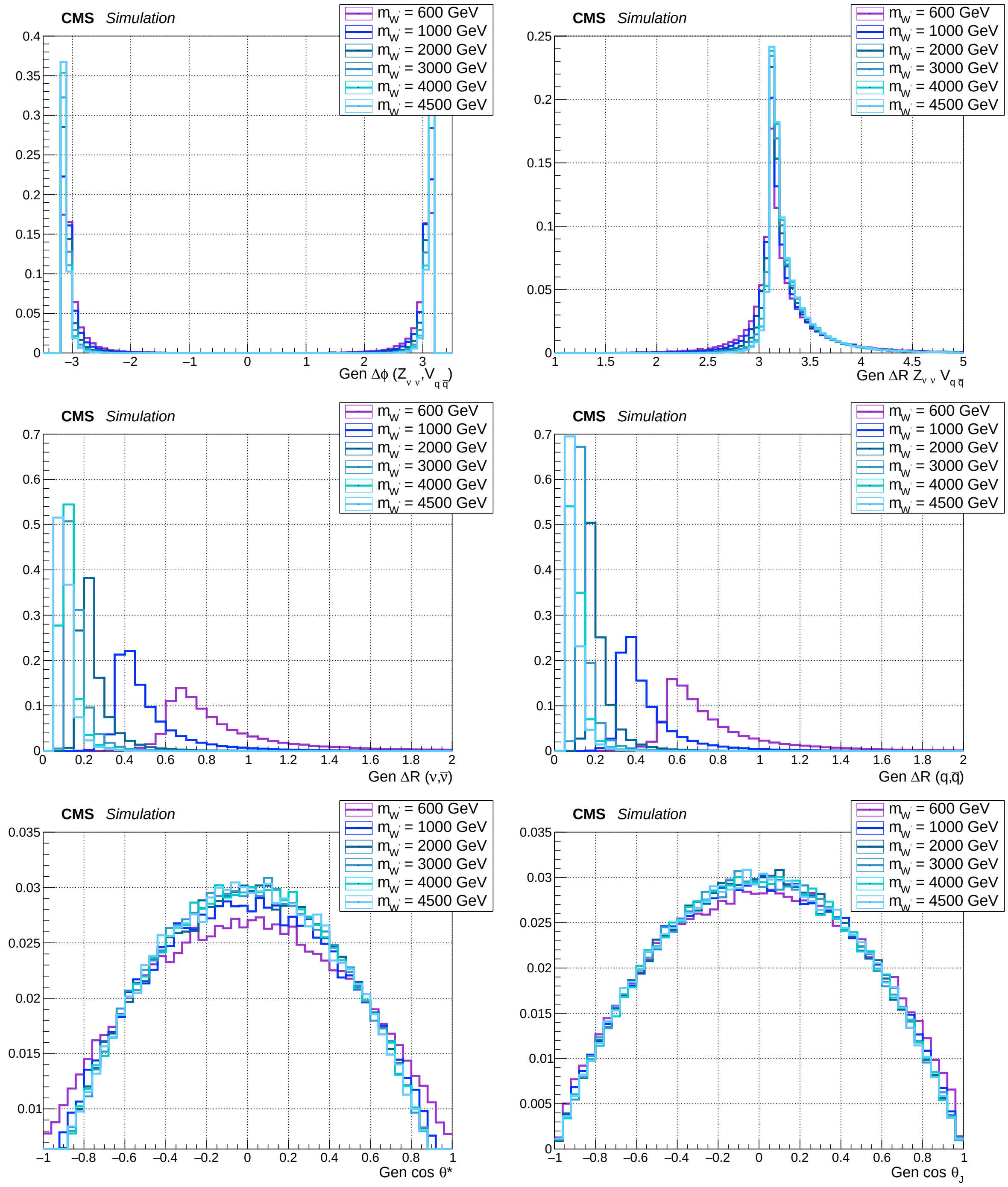


Figure 4.7: Main signal kinematic quantities at generation level after parton showering, for spin-1 W' signal, considering different mass hypoteses ($m_{W'} = 0.6, 1, 2, 3, 4, 4.5$ TeV). Top: angular separation in the transverse plane $\Delta\varphi$ (left) and solid angle ΔR (right) between leptonic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

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

- the ΔR among neutrinos and quarks reflect the boosted nature of the electroweak bosons: the more massive the resonance, the larger the boost, and hence the closer the fermions. By looking at fig. 4.4-4.7, with a jet clustering parameter of 0.8 (AK8 jet) it is possible to enclose the quarks produced by the decay of the V boson, for a resonance mass over 1 TeV;

- the $\cos\theta^*$, namely the cosine of the angle between the momentum of the V boson, calculated in the resonance rest frame, and the flight direction of the resonance itself in the laboratory frame. This variable depends on the spin of the diboson resonance (spin-2 and spin-1 distributions are different, fig. 4.4-4.7).

- the $\cos\theta_J$, the cosine of the angle between the momentum of the leading quark, calculated in the V rest frame, and the flight direction of the V boson in the laboratory frame. This variable depends on the polarization state of the decay bosons [83]; in both HVT and bulk graviton model, electroweak bosons are expected to be longitudinally polarized. When $\cos\theta_J \to 0$, quarks are produced very close in angle and hence it is difficult to disentangle the two substructures in the large-cone jet (sec. 4.3.8); when $\cos\theta_J \to \pi$ the quarks are emitted asymmetrically (one is softer than the other).

4.2.3 Background samples

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

- Z **+ jets**: this process represents the main irreducible background for the signal. The production of a Z boson in association with one or more partons in the final state has topology that is similar to the signal. This Z + jets background is produced in samples binned in p_T of the Z boson, starting from 100 GeV, with the aMC@NLO generator, with FXFX merging [84]. The contribution from events with $p_T < 100$ GeV is negligible after the requirement on the $\vec{p}_T^{\,\text{miss}}$ to be greater than 200 GeV (sec. 4.3.12).

- W **+ jets**: the leptonic decay of a W boson can be an irreducible background if the charged lepton escapes undetected (*i.e.* outside the detector acceptance) or fails the lepton identification requirements. The production of a W boson has a cross section larger by an order of magnitude with respect to the Z, and this makes the W + jets a relevant background also when a lepton veto is applied. This W + jets background is produced in samples binned in p_T of the W boson, starting from 100 GeV, with the aMC@NLO generator.

- **Top**: the pair and single production of top quarks represent a source of background, given the decay chain $t \to bW$, always including an electroweak W boson and a b-quark. The $t\bar{t}$ pair production manifests as two b-jets and two W bosons in the final state, that can decay to leptons that escape the detector or fail to be identified as leptons. This analysis makes use of $t\bar{t}$ inclusive decays samples based on POWHEG v2 [85] NLO generator. Single-top and single-antitop samples are produced in the 5-flavours scheme using POWHEG v2 [86] NLO generator. Different production mechanisms are considered: tW channel, when a top quark is produced in association with a W boson, due to a gluon-bottom quark scattering; s-channel, due to

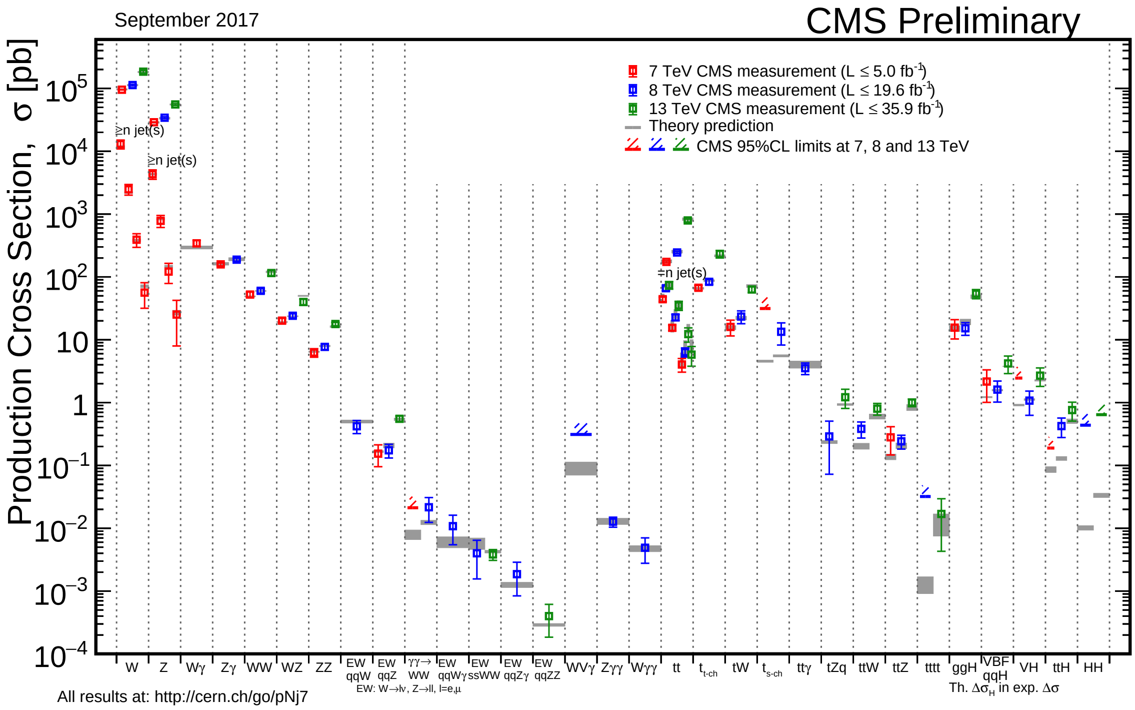


Figure 4.8: Production cross-sections of the main standard model processes, as measured by CMS, and theoretical predictions.

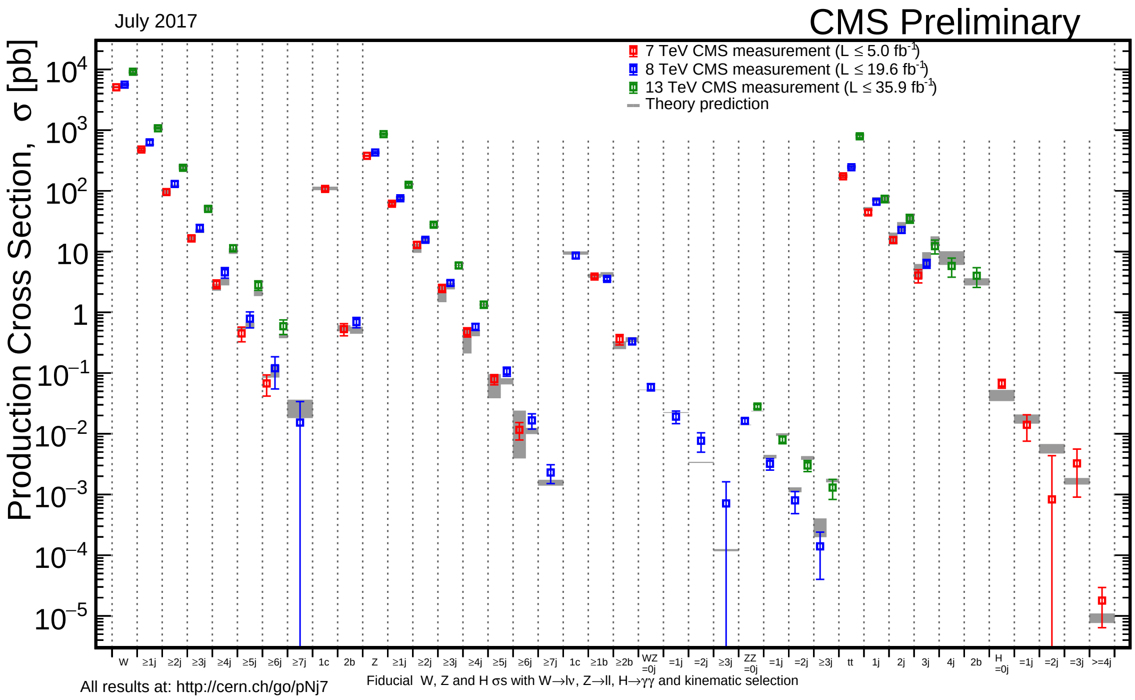


Figure 4.9: Production cross-sections of the standard model processes involving a vector boson in association with jets, as measured by CMS, and theoretical predictions. These phenomena represent the main background sources for the analysis.

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

- **Diboson**: the SM production of a pair of vector bosons is topologically close to the searched signal, by the way the cross-section of the process is low. The WW production is the most probable process, that imitates the signal when one of the W decays leptonically and the charged lepton fall outside the detector acceptance or it is mis-identified; the WZ and ZZ processes have smaller cross-sections but are topologically identical to the signal, except for the fact that the invariant mass of the diboson system has a smoothly falling spectrum, in contrast to the resonant signal distribution. Inclusive diboson production processes (WW, WZ, ZZ) are considered; they are simulated at LO by PYTHIA generator.

- **Multi-jet**: despite the production of multi-jet events, initiated by the strong interactions, has a very large cross-section, this source of background is suppressed by a dedicated selection and hence negligible for the analysis (sec. 4.3.12).

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

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

4.2.4 Vector boson momentum corrections

Corrections to the p_T spectrum of the V boson, due to NLO electroweak contributions, are enhanced at TeV scale [87], and they become significant for the purpose of this search. These corrections 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.

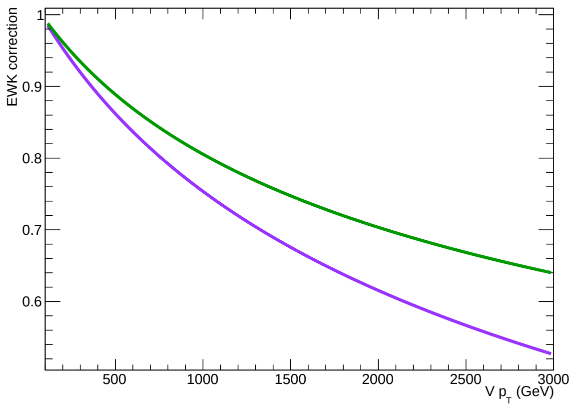


Figure 4.10: Electroweak corrections for the Z (green line) and W boson (purple line) as a function of the transverse momentum of the boson [87].

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 $\vec{p}_T^{\text{miss}}$ 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 $35.9 \pm 0.9\ \text{fb}^{-1}$ [88]. 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 $\vec{p}_T^{\text{miss}}$ 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

at HLT (cleaning events from noise in the detector). Different thresholds are applied to E_T^{miss} (no μ) and H_T^{miss} (no μ).
The approach adopted in this analysis consists in calculating the trigger efficiency on data, and applying the measured efficiency to Monte Carlo samples. Therefore the trigger is not required to have been fired in MC.
Given that the final state probed by the analysis consists into one AK8 jet, large E_T^{miss} and no charged leptons, an unbiased measurement of the E_T^{miss} trigger efficiency can be performed in an orthogonal dataset, collected with different triggers, and requiring events where a $W \to \ell\,\nu$ leptonic decay is taking place. This guarantees the presence of real $\vec{p}_T^{\text{miss}}$ in the event, due to the neutrino; furthermore, the presence of a charged lepton guarantees that the leptonic W-like events are not overlapped with the search region. The additional requirement to have at least one AK8 jet is applied in the trigger measurement, in order to probe a kinematical region similar to that of the signal region of the analysis.
The efficiency of the E_T^{miss} triggers is measured on `SingleMuon` dataset by selecting $W \to \mu\,\nu$ events using a logic or of single muon triggers `HLT_IsoMu24 OR HLT_IsoTkMu24_v`, namely, triggers asking for a PF muon reconstructed at HLT, with a p_T threshold of 24 GeV, that is isolated (in the whole reconstruction or at tracker level only). Offline selections consist in asking to have one isolated muon, with a suitable p_T threshold to be in the plateau of the muon trigger. The efficiency has been calculated as a function of the minimum quantity between the offline reconstructed E_T^{miss} (no μ):

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

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

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

This approach guarantees to mimic the behaviour of the online L1 trigger seeds. The detailed selections are listed below:

- `HLT_IsoMu24_v OR HLT_IsoTkMu24_v`,

- 1 isolated muon $p_T > 35$ GeV, identified with tight requirements,
- at least one AK8 jet, $p_T > 170$ GeV, $|\eta < |2.5$, identified with loose requirements,
- AK4 jets included in H_T^{miss}: $p_T > 30$ GeV, $|\eta < |2.5$, identified with loose requirements.

The efficiency of the E_T^{miss} triggers has independently been measured also on `SingleElectron` dataset, by selecting $W \to e\,\nu$ events using a single electron trigger (`HLT_Ele27_WPLoose_Gsf OR HLT_Ele27_WPTight_GsfOR HLT_Ele32_WPTight_Gsf`), asking to have one well identified electron, with a suitable p_T threshold, and asking to the electron and the $\vec{p}_T^{\text{miss}}$ to be separated in the transverse plane (hence, in φ) in order to suppress fake jet events mis-identified as electrons at trigger level ($\Delta\varphi > 0.5$). The detailed selections are listed below:

- `HLT_Ele27_WPLoose_Gsf OR HLT_Ele27_WPTight_GsfOR HLT_Ele32_WPTight_Gsf`,
- 1 electron, $p_T > 35$ GeV, identified with tight requirements,
- at least one AK8 jet, $p_T > 170$ GeV, $|\eta < |2.5$, identified with loose requirements,
- AK4 jets included in H_T^{miss}: $p_T > 30$ GeV, $|\eta < |2.5$, identified with loose requirements.

All the available data have been employed to derive the efficiency. The final turn-on curves for the E_T^{miss} triggers are shown in fig.4.11-4.12, measured in muon and electron dataset respectively. The `PFMETNoMu` trigger efficiencies are displayed separately, together with their logic OR. The trigger efficiency measured on `SingleMuon` dataset amounts to 96% at E_T^{miss} =200 GeV; the trigger efficiency measured on `SingleElectron` dataset amounts to 95% at E_T^{miss} =200 GeV. The difference needed to cover the gap between the two independent measurements is taken as trigger systematic uncertainty, and it amounts to 1% at 200 GeV.

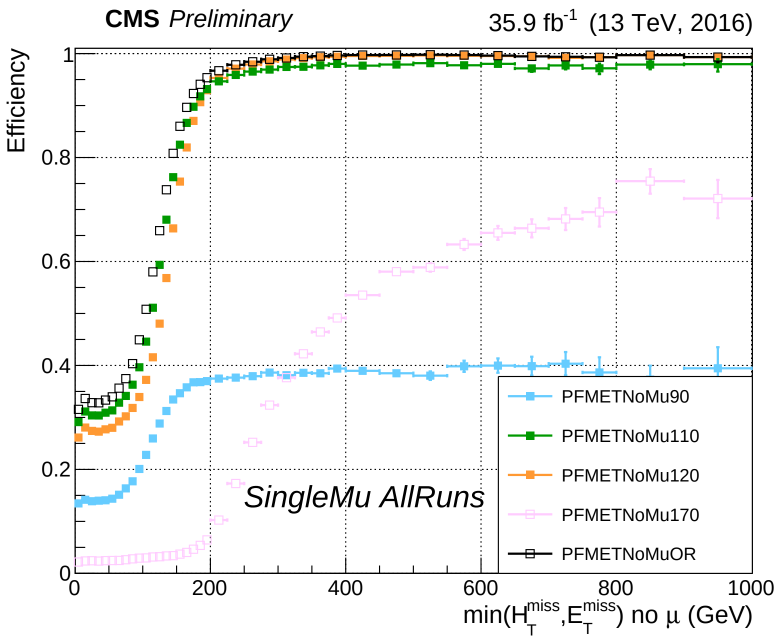


Figure 4.11: $\vec{p}_T^{\text{miss}}$ trigger efficiency for the $\vec{p}_T^{\text{miss}}$ trigger paths used in this analysis, calculated on `SingleMuon` dataset, as a function of the minimum of the variables E_T^{miss} (no μ) (eq. 4.1) and H_T^{miss} (eq. 4.2).

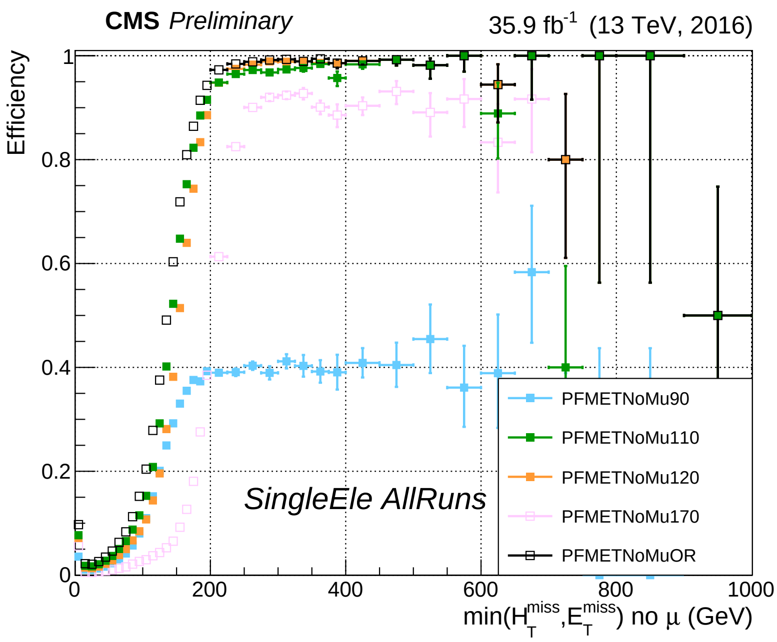


Figure 4.12: $\vec{p}_T^{\text{miss}}$ trigger efficiency for the $\vec{p}_T^{\text{miss}}$ trigger paths used in this analysis, calculated on `SingleElectron` dataset, as a function of the minimum of the variables E_T^{miss} (no μ) (eq. 4.1) and H_T^{miss} (eq. 4.2).

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 + jets events in light blue, W + 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 $|z_{vtx}| < 24$cm
- vertex distance with respect the beam pipe $d_0 < 2$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 $\vec{p}_T^{\text{miss}}$ 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 $p_T > 10$ 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 $\sigma_{i\eta i\eta}$; $\Delta\eta_{in}^{seed}$ and $\Delta\varphi_{in}$ are the difference in η and φ between the track position as it is measured in the inner layer,

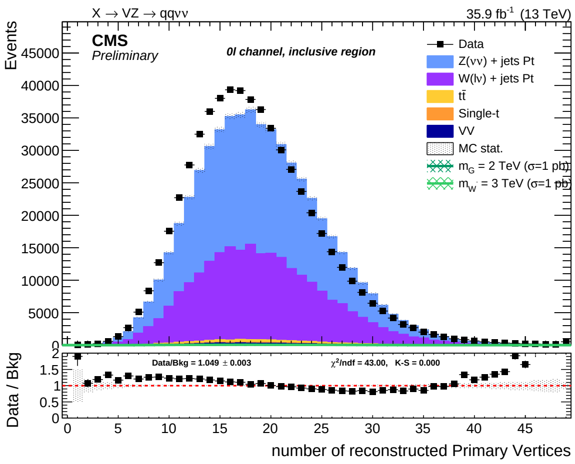


Figure 4.13: Primary vertices distributions in data and MC samples, after reweighting.

1363 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		*Veto Id*	
		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	≤	2	3
conversion veto		yes	yes

1370 ### 4.3.3 Photons

1371 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 contains 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.

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		*Loose Id*	
		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

4.3.4 Muons

The minimal criteria to define a muon is that it must be identified by the Particle-Flow algorithm, and should be reconstructed either as a global muon or as a tracker muon (sec. 3.2.9). The muon isolation is defined in a cone with a radius of $\Delta R = 0.4$ centered along the lepton direction. In the analysis event selection, all muons identified with the loosest criteria previously described, p_T over 10 GeV, PF isolation below 0.25, $\eta < |2.4|$ are vetoed.

4.3.5 Taus

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

4.3.6 Jets

In this analysis, jets are considered if the corrected p_T is larger than 30 GeV for AK4 jets, and larger than 200 GeV for AK8 jets, and lie in the tracker acceptance ($|\eta| < 2.4$). The requirement on AK8 jets transverse momentum is motivated by the fact that $p_T = 200$ GeV is the minimum kinematical threshold ensuring to enclose the lighter hadronically decaying vector boson (namely, the W boson) in the jet cone. Additionally, AK4 jets are required to pass *loose* jet identification requirements, AK8 are required to pass *tight* jet identification requirements defined in tab. 4.7. AK8 jets are used to reconstruct the hadronically decaying electroweak boson candidate, whilst AK4 jets are used to suppress the contribution of top and QCD background events. Jet energy corrections are applied to AK4 and AK8 CHS jets. Fig. 4.14- 4.16 show the data/simulation comparison after the analysis selections (tab. 4.12 without Top cleaning and Event cleaning).
Since it has been measured that the jet energy resolution (JER) is not the same in data and MC, an additional smearing is applied in simulation, in order to get a better agreement. There are two independent ways to get the smearing. The scaling method rescales the corrected four-momentum of a reconstructed jet by a factor

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

where p_T is the transverse momentum of the jet, p_T^{gen} is the transverse momentum of the generator 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	*Loose*	*Tight*
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

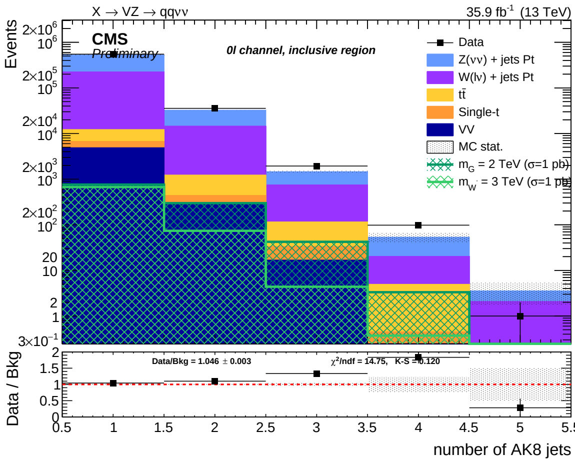


Figure 4.14: Number of reconstructed AK8 jets after inclusive selections.

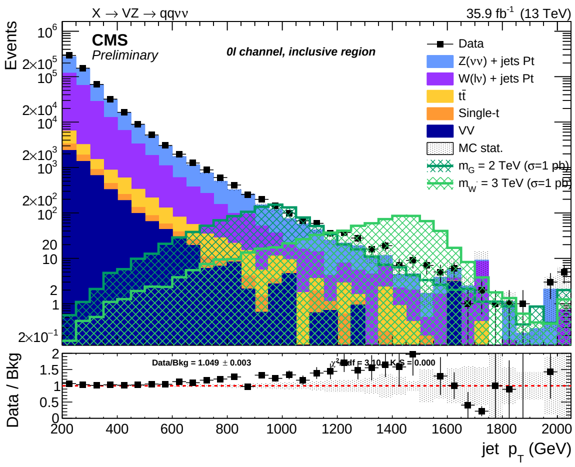


Figure 4.15: Leading AK8 jet p_T spectrum after inclusive selections.

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

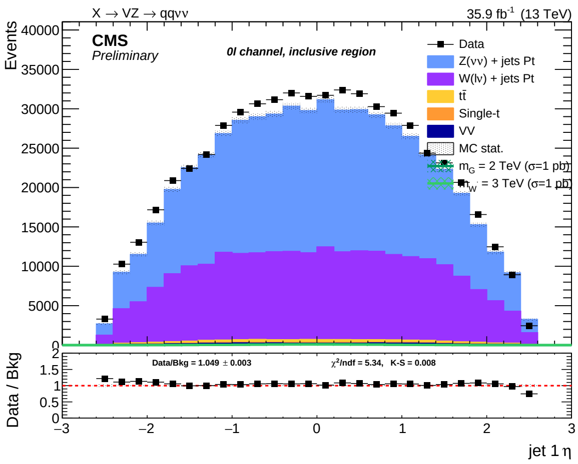


Figure 4.16: Leading AK8 jet η spectra after inclusive selections.

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

$$\begin{aligned} &\Delta R < R_0/2, \\ &|p_T - p_T^{\text{gen}}| < 3 \times \sigma_{\text{JER}} \times p_T, \end{aligned} \tag{4.4}$$

where R_0 is the jet clustering parameter and σ_{JER} is the relative p_T resolution measured in simulation.

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

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

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

The smearing procedure adopted in this analysis is the hybrid method: when a matching jet at generator level is found, the scaling method is adopted, else the stochastic smearing is chosen. The smearing coefficients (scale factors, SF) as a function of the jet η and their uncertainties are reported in tab. 4.8 for 2016 data [69].

4.3.7 Jet mass

The jet mass is the main observable in distinguishing a V jet from a jet produced by colour interaction (QCD jets). Jet grooming procedure consists in the suppression of uncorrelated underlying event, pile-up and soft radiation from the jet: it improves the signal and background discrimination, by pushing the jet mass for QCD jets towards lower values of the spectrum, while maintaining the jet mass for V-jets around the electroweak boson mass window.

The grooming technique of the analysis relies on the “soft drop declustering” algorithm, a jet substructure technique that recursively removes soft wide-angle radiation from a jet [89], in order to mitigate the contaminations from initial state radiation, along with pile-up and multiple scatterings.

The soft drop algorithm starts with a jet clustered with the anti-k_T algorithm with a parameter R_0; the jet is then reclustered with the Cambridge-Aachen method [90], 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

eq. 3.10, once the exponent a is set $a = 0$. The soft drop algorithm is ruled by two parameters, a soft threshold z_{cut}, that cuts on the energy fraction of soft radiation, and an angular exponent β. The procedure is the following:

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

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

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

where p_T^1 and p_T^2 are the momenta of the constituents, ΔR_{12} is their angular distance. z_{cut} and β parameters affect the degree of jet grooming: if $\beta \rightarrow \infty$ the jet remains ungroomed, while the more β approaches zero, the more soft collinear radiation is removed.
The net effect of the soft drop algorithm is studied in Monte Carlo simulations of a W hadronic decay process (signal), in association with jets, and of a multi-jet QCD process (background). Jets are clustered with the anti-k_T algorithm with a parameter $R_0 = 1$ and asked to have $p_T > 500$ GeV and $|\mathscr{Y}| < 4$. The parameter z_{cut} is chosen such in a way that the number of events falling in the W mass window ([70, 90] GeV) is the 35% of the total number of events. The results before (black curve) and after the application of soft drop algorithm (coloured curves, depending on the value of β) are presented in fig. 4.17 [89]. In particular, by comparing the ungroomed jet mass (in black) with the mass groomed with a parameter $\beta = 0$ (adopted in this analysis and displayed with a green curve), the soft drop mass of the leading jet is a very narrow distribution peaking around the nominal W window in the signal sample, whilst it is pushed at lower values in the background sample.

The soft drop algorithm is used in association with the Pile Up Per Particle Identification algorithm (PUPPI) [91], designed to combine detector informations in order to compute a local metric α, that

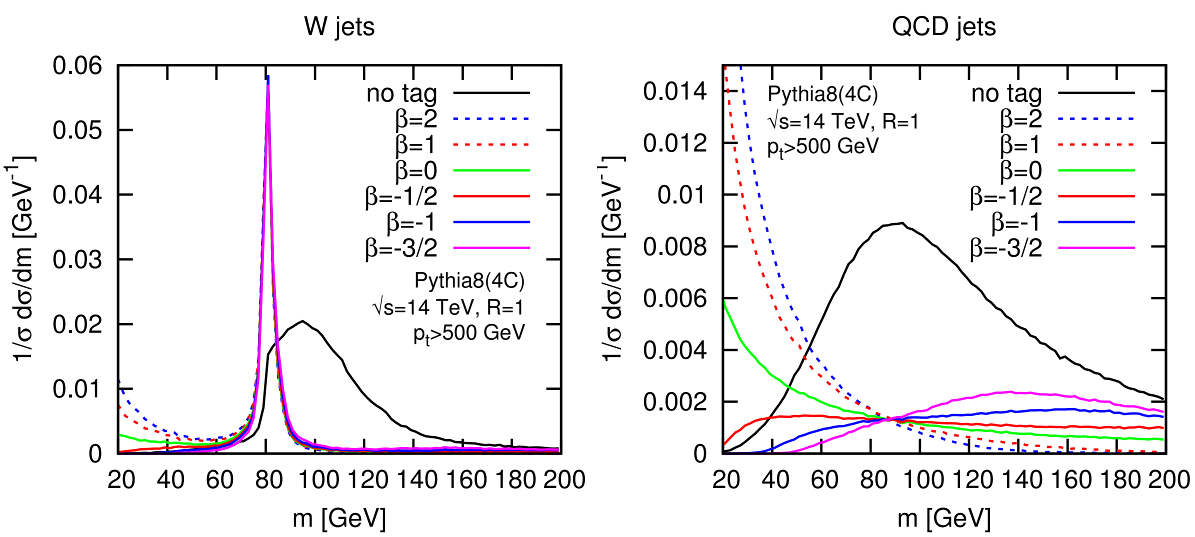


Figure 4.17: Distributions of the jet mass in W + jeta signal simulations (left) and multi-jet QCD background (right), before (in black) and after applying soft drop algorithm. Each curve corresponds to a different value of the parameter β. [89]

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:

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

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

$$\chi_i^2 = \Theta(\alpha_i - \bar{\alpha}_{PU}) \frac{(\alpha_i - \bar{\alpha}_{PU})^2}{\sigma_{PU}^2} \tag{4.8}$$

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 $F_{\chi^2,\,1\text{ d.o.f.}}$,

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

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:

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

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

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

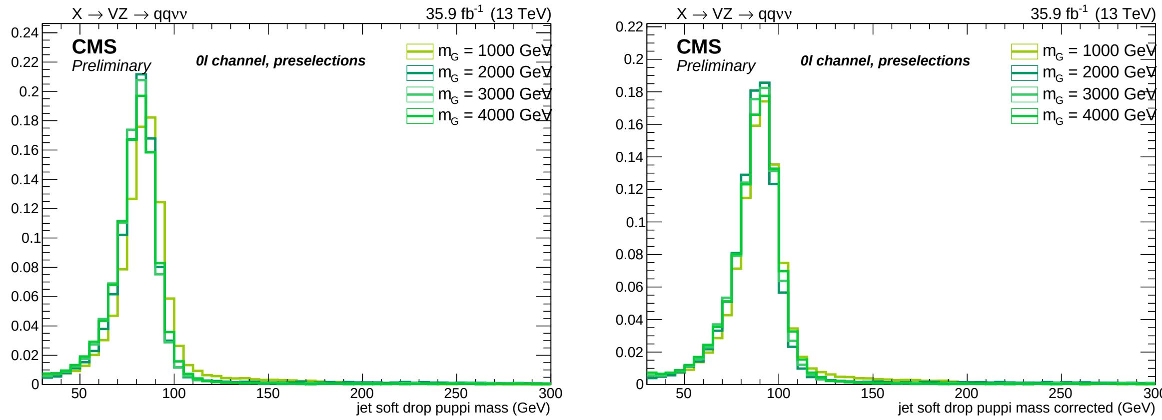


Figure 4.18: Soft drop PUPPI mass of AK8 jet reconstructed for different bulk graviton signal samples, before corrections (left) and after corrections (right).

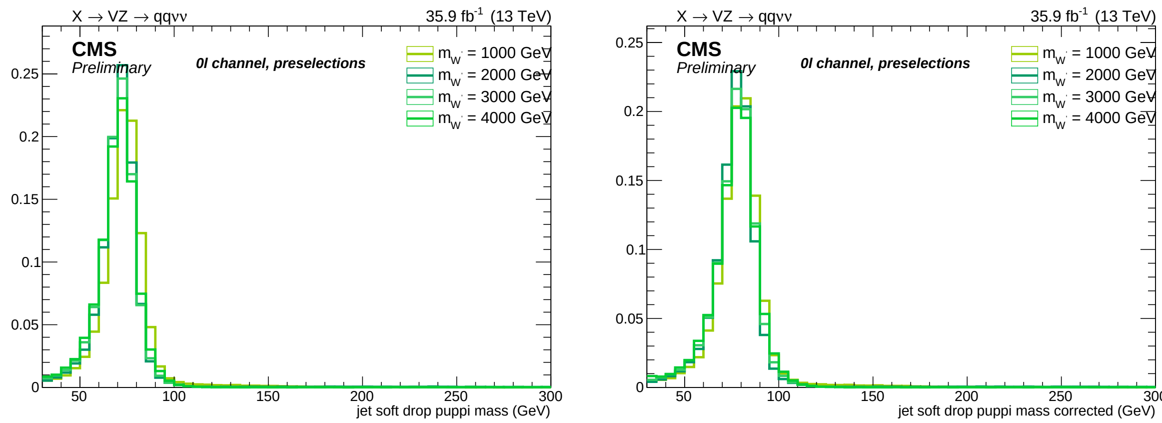


Figure 4.19: Soft drop PUPPI mass of AK8 jet reconstructed for different W' signal samples, before corrections (left) and after corrections (right).

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

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

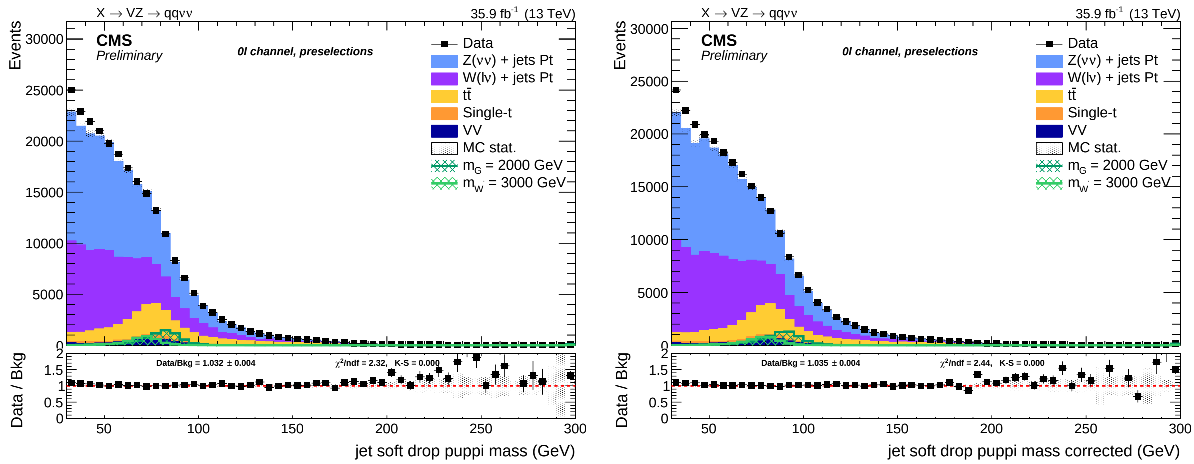


Figure 4.20: Soft drop PUPPI mass of AK8 jet; left: before corrections. right: after corrections.

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 [92], τ_n, is defined as

$$\tau_n = \frac{1}{d_0} \sum_k p_{T,k} \min\left(\Delta R^{\beta}_{1,k}, \Delta R^{\beta}_{2,k}, \ldots, \Delta R^{\beta}_{n,k}\right), \tag{4.10}$$

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

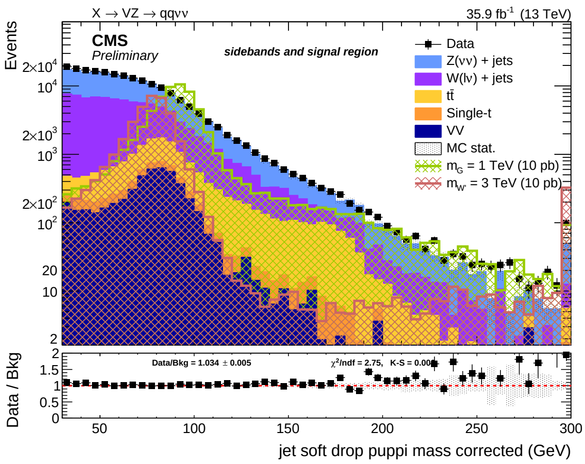


Figure 4.21: Distribution of the soft drop PUPPI corrected mass of the leading AK8 jet, selected as the hadronically decaying V candidate, in the sidebands and control region of the analysis, for expected SM background, bulk graviton signal, W' signal, and data.

The parameter d_0 is a normalization factor:

$$d_0 = \sum_k p_{T,k} R_0, \tag{4.11}$$

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

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

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

The τ_{21} variable is used to classify the events into two exclusive categories, in order to improve the signal discovery reach. Events are included in either the high-purity ($\tau_{21} < 0.35$) or low-purity ($0.35 < \tau_{21} < 0.75$) category.
The choice of the τ_{21} categorization listed above is based on a study of the analysis sensitivity. Another τ_{21} categorization is probed, according to which events are grouped into different high-purity ($\tau_{21} < 0.40$) and low-purity ($0.40 < \tau_{21} < 0.75$) categories. This different set of τ_{21} cuts has been tested, along with that chosen for this analysis. Two figures of merit are considered: the discovery reach, namely the bulk graviton signal significance (displayed in fig. 4.24), and the expected exclusion limit on cross-section times branching fraction at 95% CL (displayed in fig. 4.25), as a function of the reconstructed transverse mass of the resonance. To this purpose, the entire analysis workflow

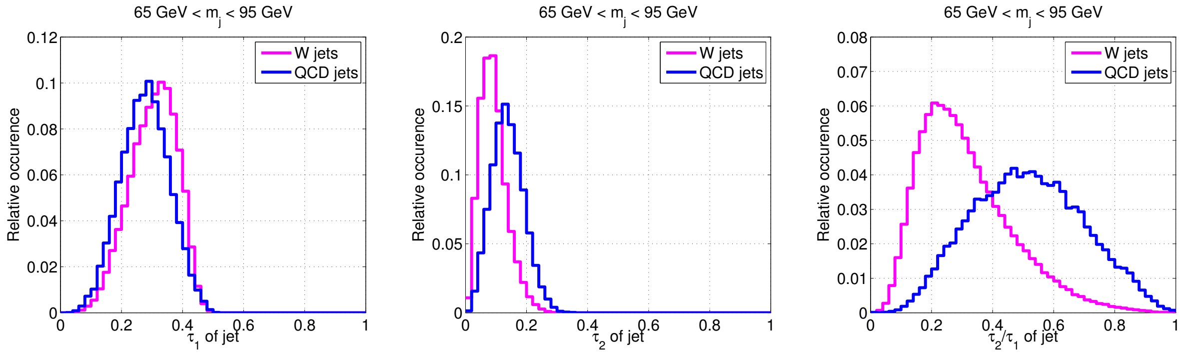


Figure 4.22: Distribution of τ_1 (left), τ_2 (center), and τ_{21} (right) variables, in simulations of a W plus jets process (in pink) and for a multi-jet QCD originated process (in blue). A selection on the leading jet mass is applied: $65 < m_j < 95$ GeV; jets are clustered with a parameter $R_0 = 0.6$, $p_T > 300$GeV, $|\eta| < 1.3$ [92].

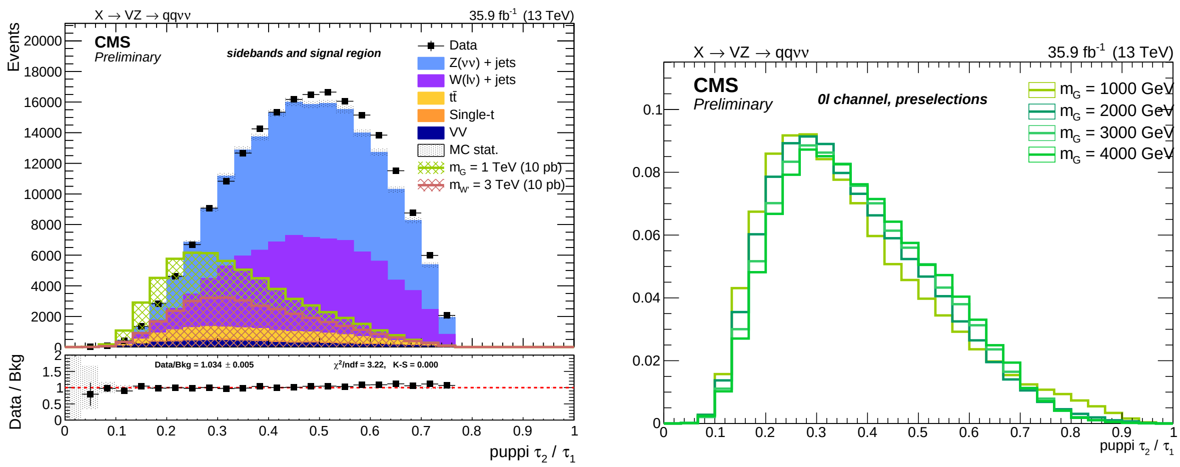


Figure 4.23: Distribution of the τ_{21} subjettiness of the leading AK8 jet, selected as the hadronically decaying V candidate, for expected SM background and data (left), and for bulk graviton signal (right).

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

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

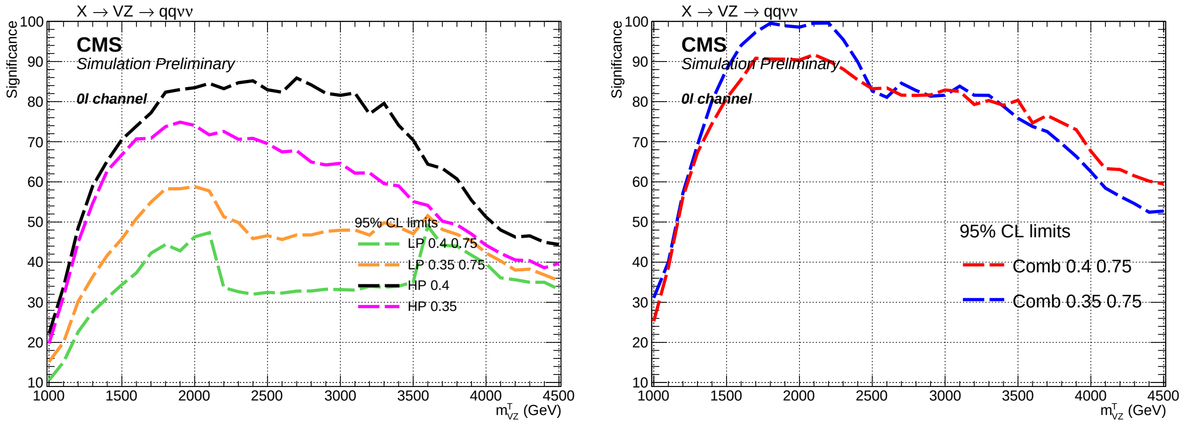


Figure 4.24: Analysis sensitivity to bulk graviton signals, computed by applying different τ_{21} categorizations, considering the categories separately (left) and combining them together (right), as a function of the resonance mass.

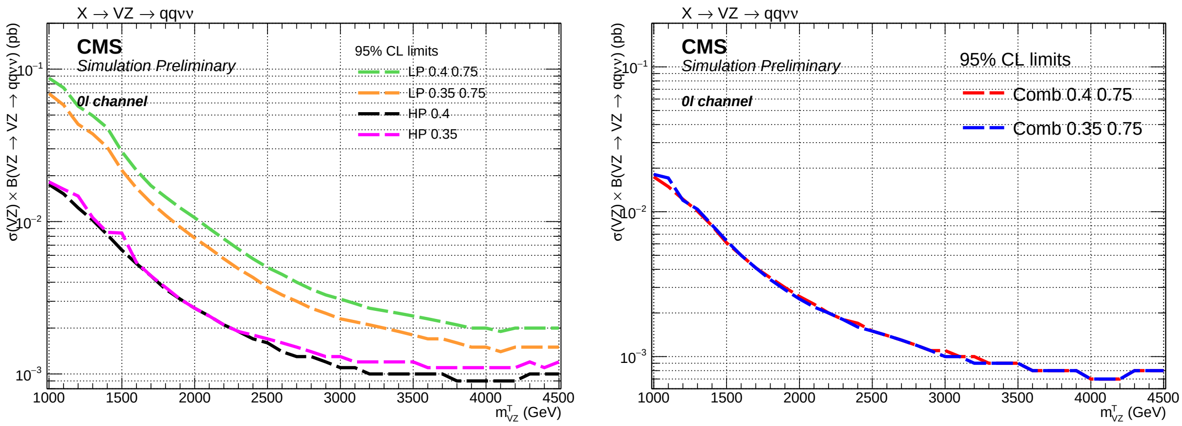


Figure 4.25: Exclusion limit on cross-section time branching fraction at 95% CL of bulk graviton signals, computed by applying different τ_{21} categorizations, considering the categories separately (left) and combining them together (right), as a function of the resonance mass.

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} subjettiness 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%) [83]. 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.

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

4.3.9 b-tagging

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 [72] is applied to the AK4 jets. 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.

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

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

Events where an AK4 jet, not laying in the AK8 jet cone, is b-tagged with the *loose* working point threshold, are rejected. This veto allows to suppress the single-top events and $t\bar{t}$ events by one half. The b-tagging efficiency is not the same in data and MC. In order to take into account this difference, b-tagging scale factors for b-jets and mis-tagged light jets, measured for different physics processes, are calculated. A weight is extracted on a per-event basis, as a function of the b-tagging status of the jets and their kinematic variables [93].

4.3.10 Missing Energy

As pointed out in sec. 3.2.9.8, Type-I corrected E_T^{miss} is used in the analysis, along with dedicated filters to remove detector noise and events with bad reconstruction. In order to lie in the plateau of the trigger efficiency, $E_T^{\text{miss}} > 200\text{GeV}$. Fig. 4.26 shows the E_T^{miss} distribution for data and Monte Carlo after the corrections and filters.

4.3.11 Diboson candidate reconstruction

4.3.11.1 $\mathrm{V} \to \mathrm{q\bar{q}}$ reconstruction

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

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

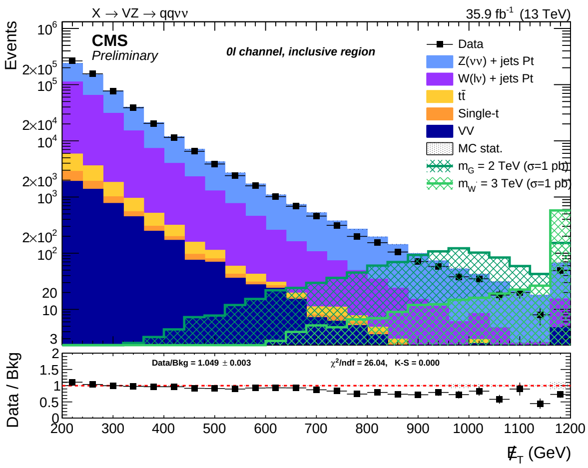


Figure 4.26: Type-1 corrected E_T^{miss} distribution after inclusive selections.

- *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 $\mathbf{Z} \rightarrow \nu\bar{\nu}$ reconstruction

If the Z boson decays into a pair of neutrinos, no product is visible in the detector, henche the invisble 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 $\vec{p}_T^{\text{miss}}$ kinematics, defined by the following formula:

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

where E_T^V is the transverse energy of the V candidate (defined in sec. 3.2.1), 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} subjettines 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 hypoteses.

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 GeV;
- *Corrections*: Type-I, noise filters.

4.3.12.2 V candidate selections

- p_T: at least one AK8 Particle-Flow jet with $p_T > 200$ 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 GeV;
- *Substructure*: PUPPI corrected τ_{21} subjettines, 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

Minimal requirements are applied to objects that are vetoed:

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

- η: $|\eta| < 2.5$;
- *Id*: *loose* cut-based working point.

Further selections are applied to suppress spurious events.

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

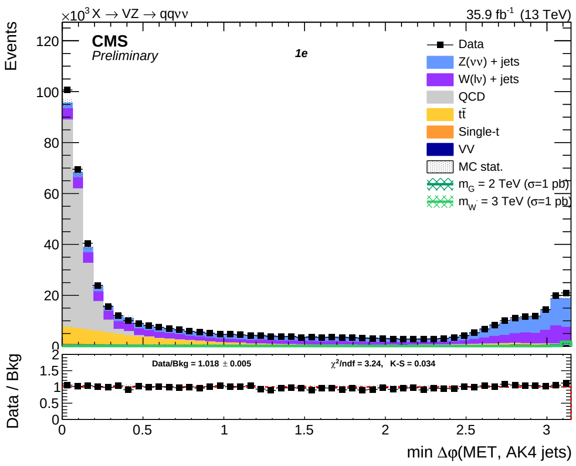


Figure 4.27: Distribution of the minimum azimuthal separation bewteen $\vec{p}_T^{\,\text{miss}}$ and the momenta of all the AK4 jets present in each event. By imposing min$\Delta\varphi > 0.5$, the QCD background (in gray) is suppressed.

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

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

	$Z \to \nu\nu$
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 GeV
Veto	e, μ, τ, γ
V	$p_T > 200$ GeV, *tight* Id nhf<0.8; chf>0.2
QCD cleaning	min$\Delta\varphi$(AK4jets, $\vec{p}_T^{\text{miss}}) > 0.5$
Top cleaning	veto on b-tagged AK4 jets outside the AK8 cone, *loose* working point (< 0.460)
Event cleaning	$\Delta\varphi(V, \vec{p}_T^{\text{miss}}) > 2$

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

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

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

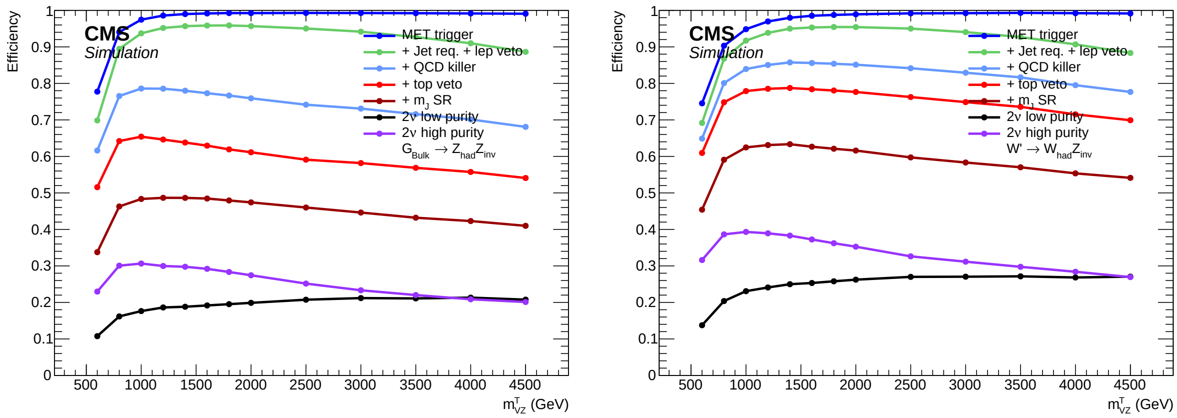


Figure 4.28: Signal efficiency for a spin-2 bulk graviton decaying into a pair of Z bosons (left), and for a spin-1 W' decaying into a W and a Z bosons (right), as a function of the mass of the heavy particle. The efficiencies are separated by purity category after the signal region selections.

4.4 Data and simulations comparison

In this section, a comparison between data and simulation is reported for various kinematic observables. It can be seen that the dominant background contribution comes from the Z + jets and W + jets production, while sub-leading contributions from top ($t\bar{t}$ and single-top) production and dibosons can be minor yet non-negligible.
In the following plots (fig. 4.29-4.38), the comparison is performed in three different regions. On top of the selections defined above, additional criteria are defined:

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

A summary of the number of expected events from Monte Carlo simulations, per each sample, along with the number of events observed in data in each category is reported in tab. 4.14. No significant 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 + jets	320996.11	57551.99	7774.40	22933.14	10763.87
	57%	56%	56%	53%	45%
W + 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

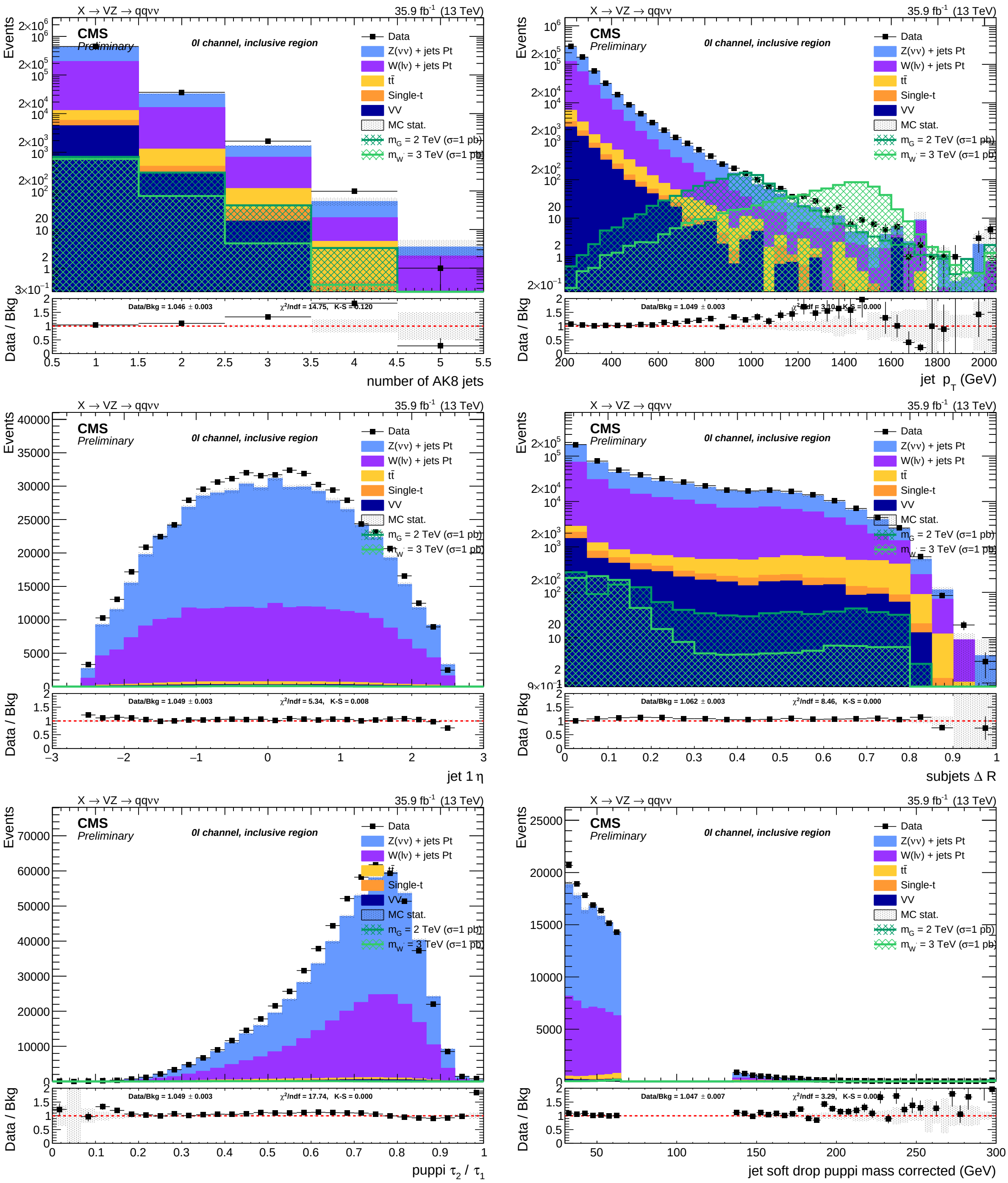


Figure 4.29: Top: number of AK8 jets in the event (left) and V jet candidate p_T (right). Center: V jet candidate η (left) and angular separation ΔR between the constituents leading subjets (right). Bottom: V jet candidate τ_{21} subjettiness after PUPPI correction (left) and V jet candidate soft drop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

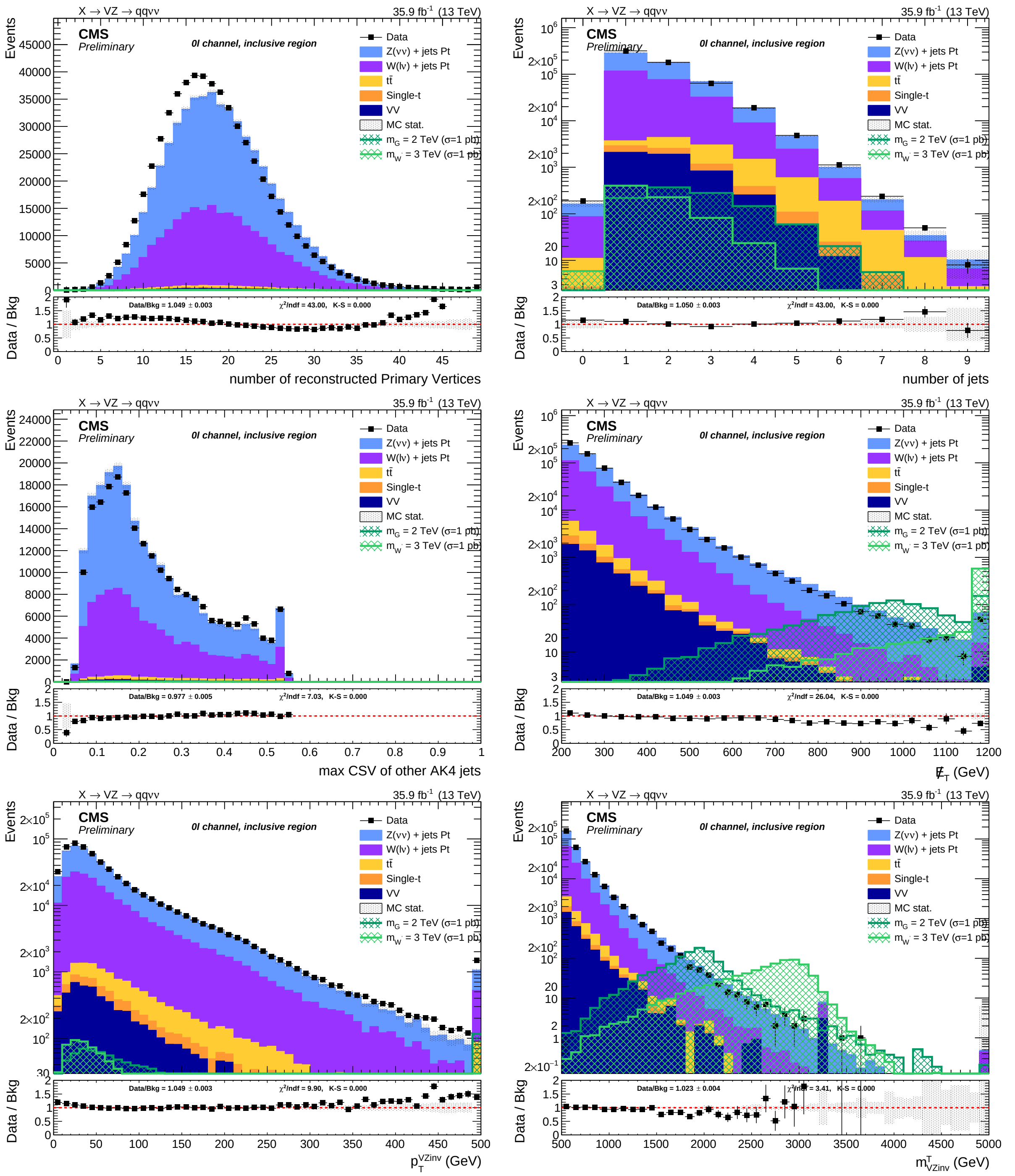


Figure 4.30: Top: number of reconstructed primary vertices (left) and number of AK4 jets in the event (right). Center: distribution of the b-tagging multivariate discriminant for the AK4 jets not included in the V jet cone (left) and E_T^{miss} distribution (right). Bottom: p_T of the VZ candidate (left)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

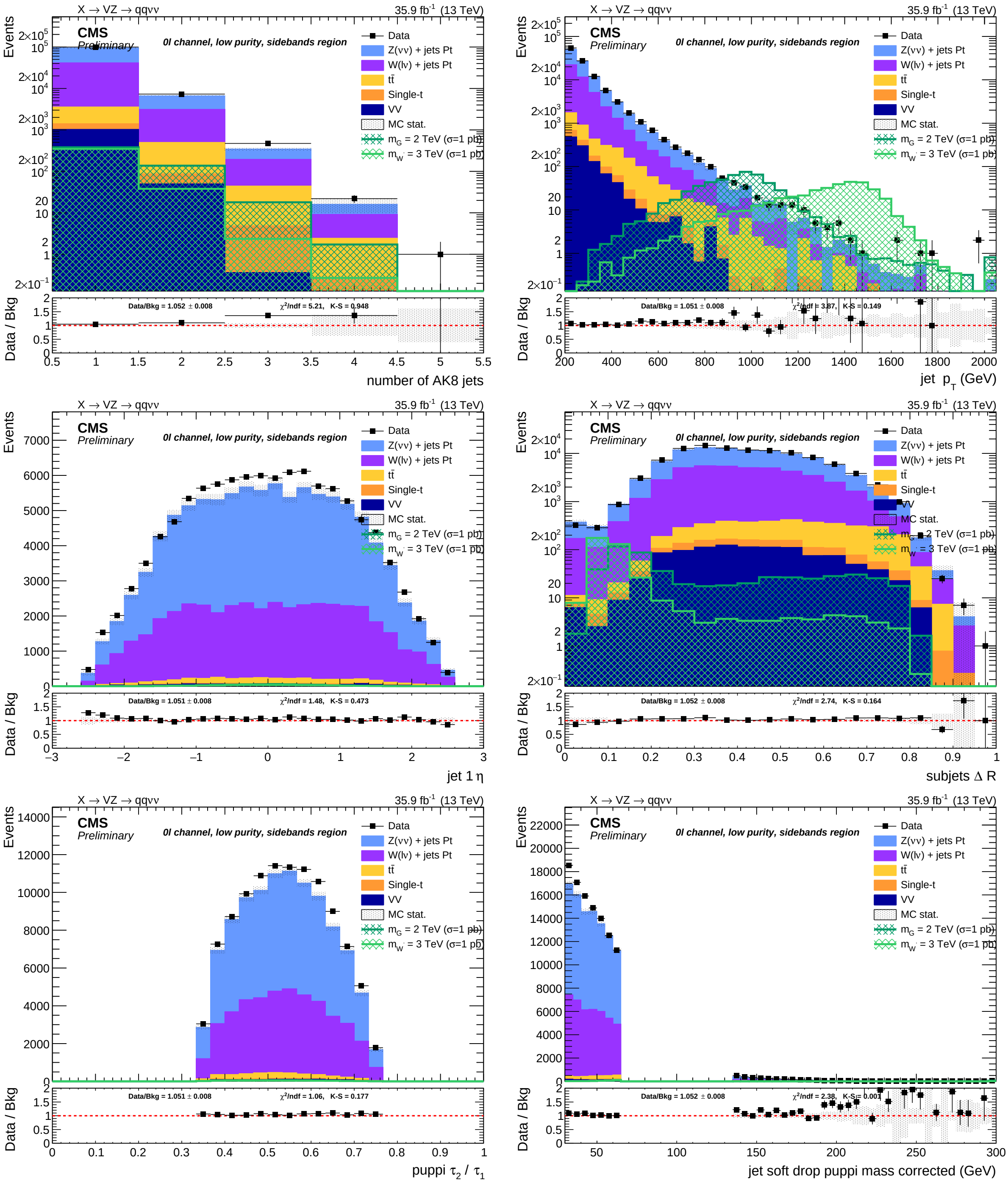


Figure 4.31: Top: number of AK8 jets in the event (left) and V jet candidate p_T (right). Center: V jet candidate η (left) and angular separation ΔR between the constituents leading subjets (right). Bottom: V jet candidate τ_{21} subjettiness after PUPPI correction (left) and V jet candidate soft drop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

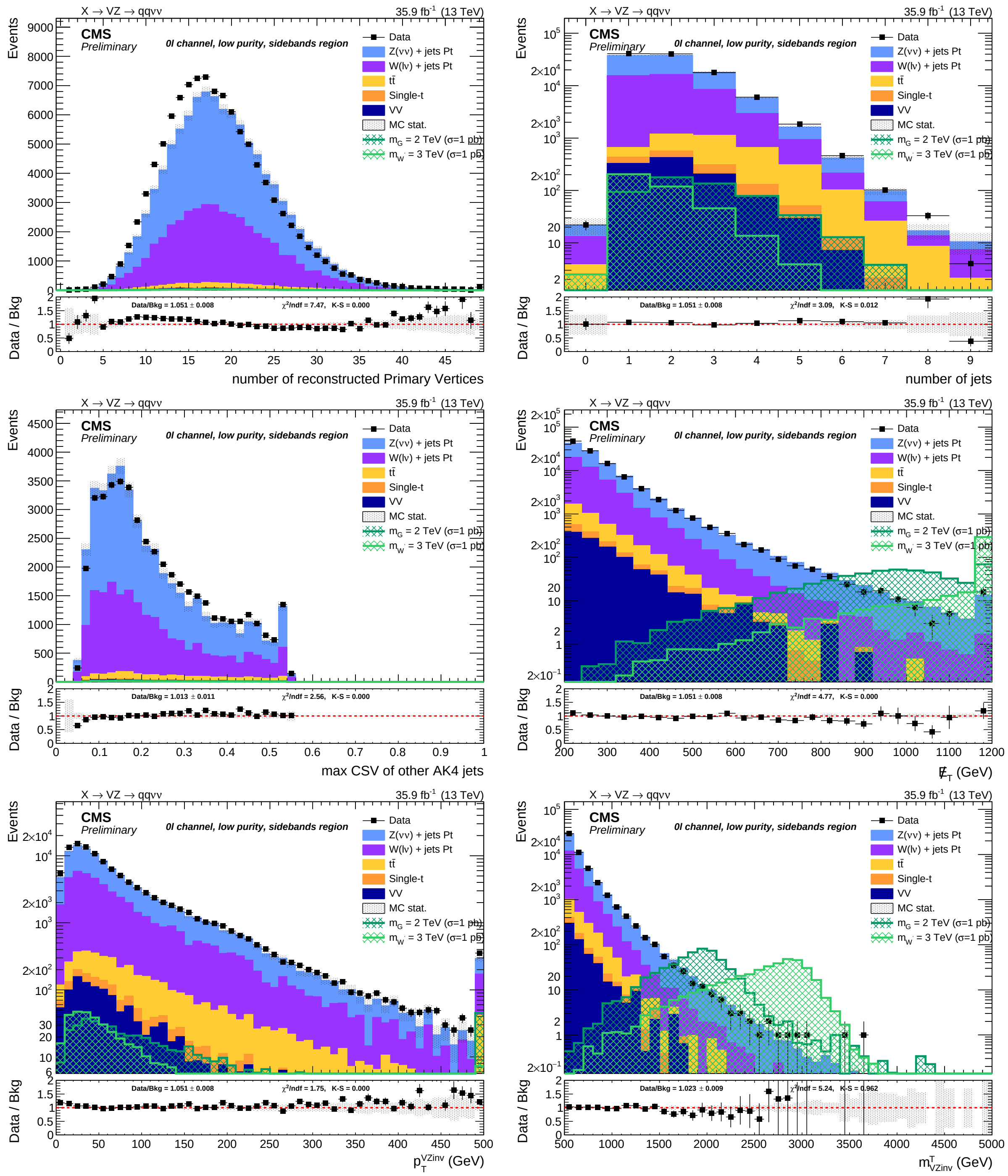


Figure 4.32: Top: number of reconstructed primary vertices (left) and number of AK4 jets in the event (right). Center: distribution of the b-tagging multivariate discriminant for the AK4 jets not included in the V jet cone (left) and E_T^{miss} distribution (right). Bottom: p_T of the VZ candidate (left)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

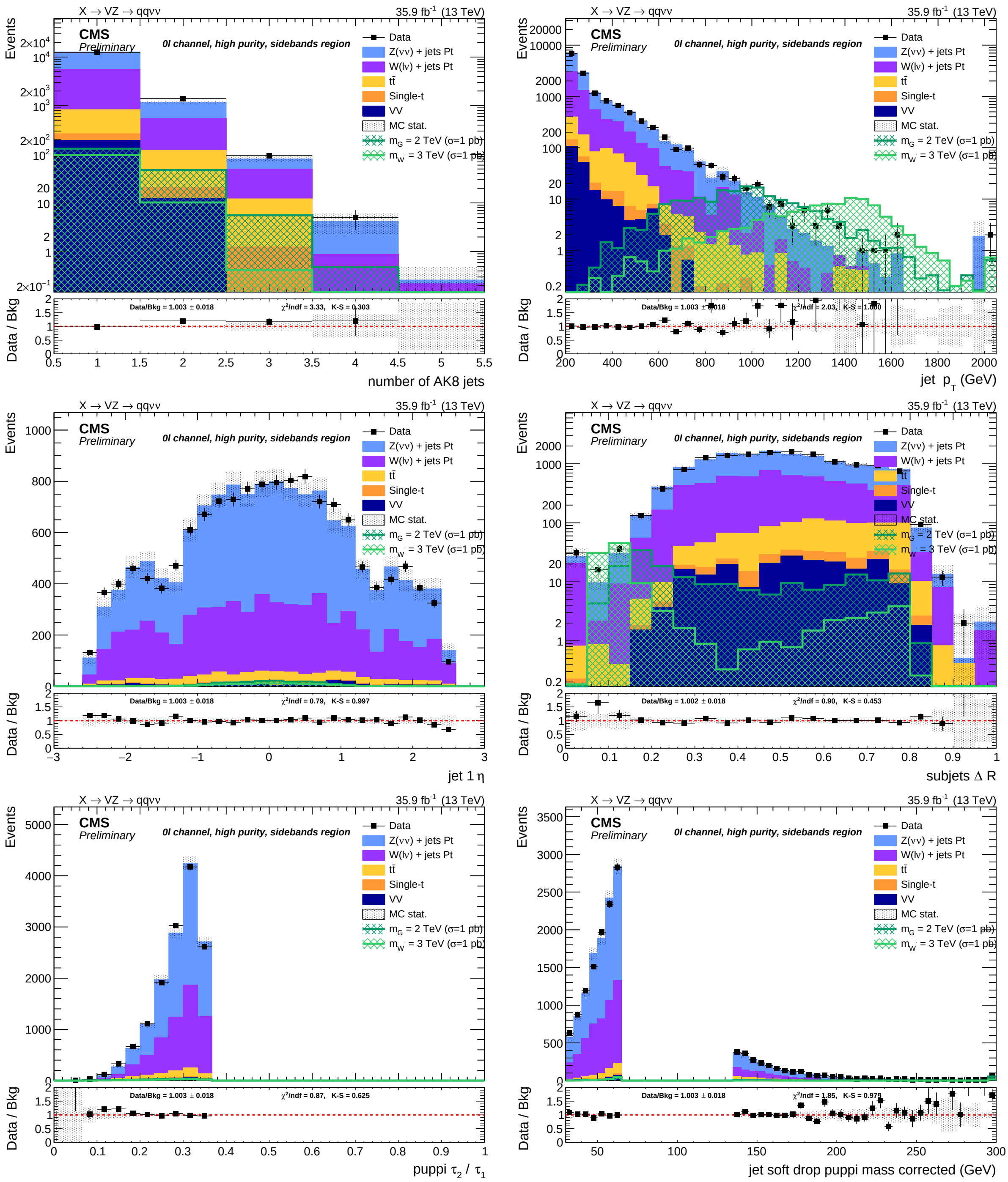


Figure 4.33: Top: number of AK8 jets in the event (left) and V jet candidate p_T (right). Center: V jet candidate η (left) and angular separation ΔR between the constituents leading subjets (right). Bottom: V jet candidate τ_{21} subjettiness after PUPPI correction (left) and V jet candidate soft drop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

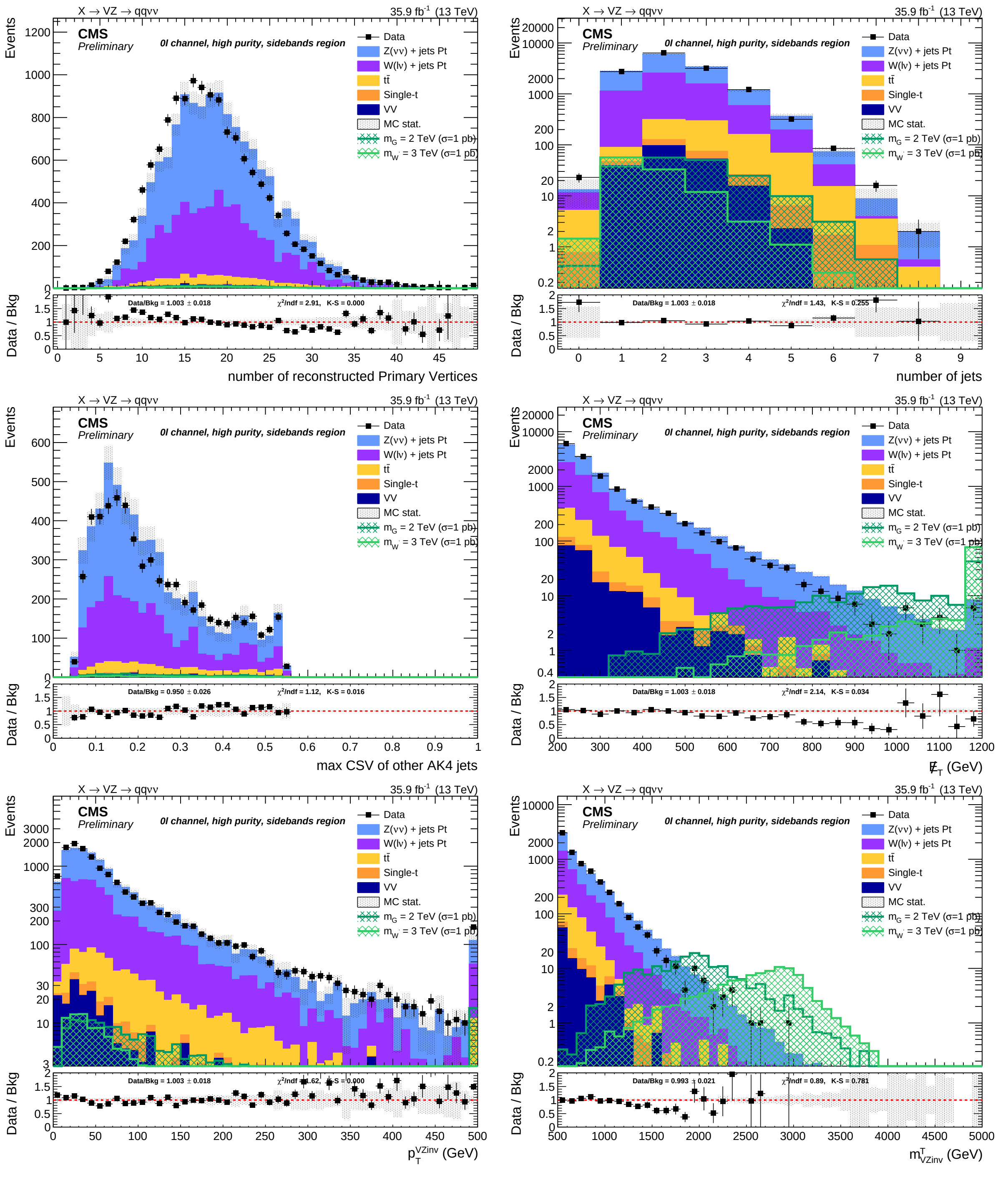


Figure 4.34: Top: number of reconstructed primary vertices (left) and number of AK4 jets in the event (right). Center: distribution of the b-tagging multivariate discriminant for the AK4 jets not included in the V jet cone (left) and E_T^{miss} distribution (right). Bottom: p_T of the VZ candidate (left)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

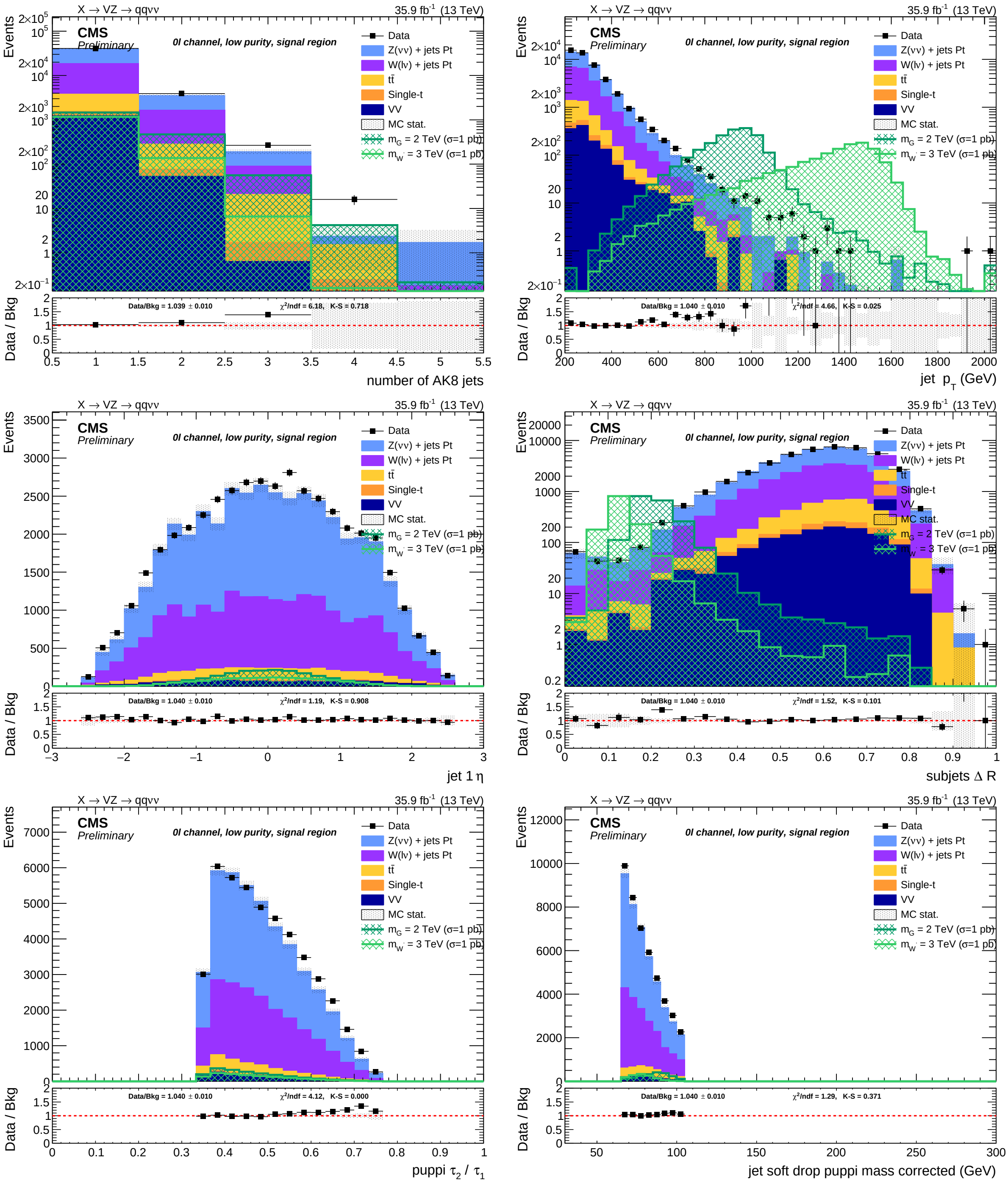


Figure 4.35: Top: number of AK8 jets in the event (left) and V jet candidate p_T (right). Center: V jet candidate η (left) and angular separation ΔR between the constituents leading subjets (right). Bottom: V jet candidate τ_{21} subjettiness after PUPPI correction (left) and V jet candidate soft drop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

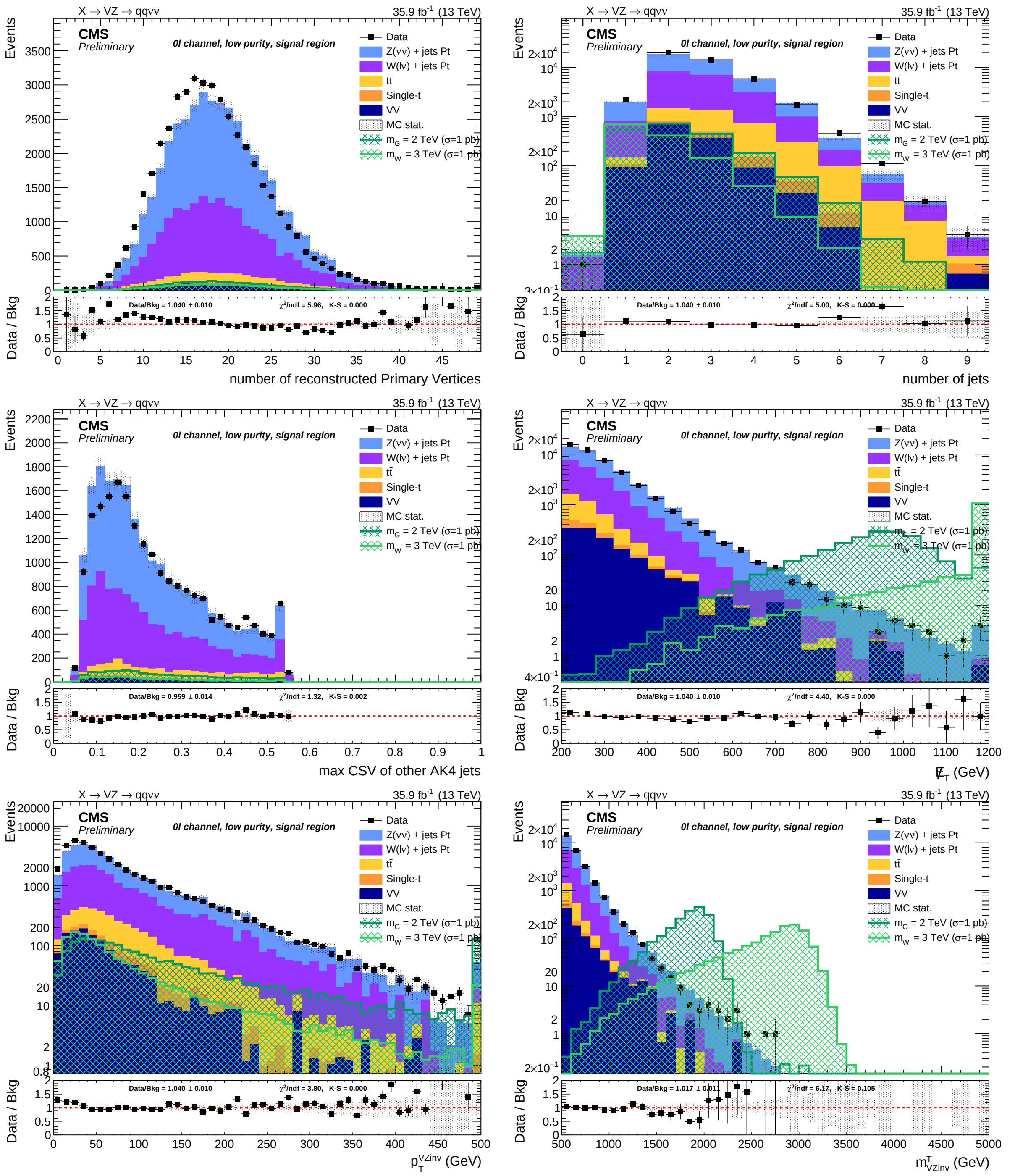


Figure 4.36: Top: number of reconstructed primary vertices (left) and number of AK4 jets in the event (right). Center: distribution of the b-tagging multivariate discriminant for the AK4 jets not included in the V jet cone (left) and E_T^{miss} distribution (right). Bottom: p_T of the VZ candidate (left) 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.

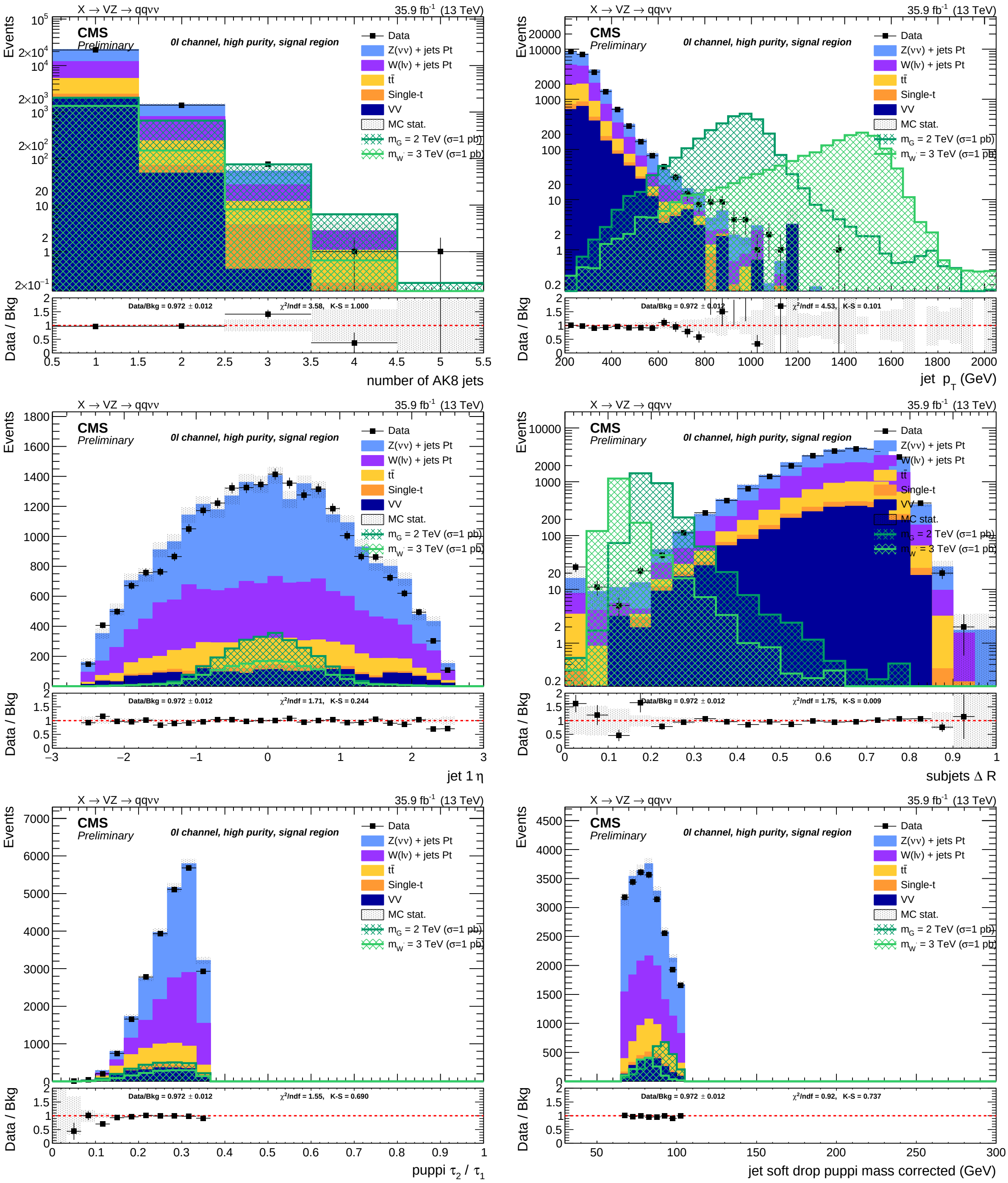


Figure 4.37: Top: number of AK8 jets in the event (left) and V jet candidate p_T (right). Center: V jet candidate η (left) and angular separation ΔR between the constituents leading subjets (right). Bottom: V jet candidate τ_{21} subjettiness after PUPPI correction (left) and V jet candidate soft drop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

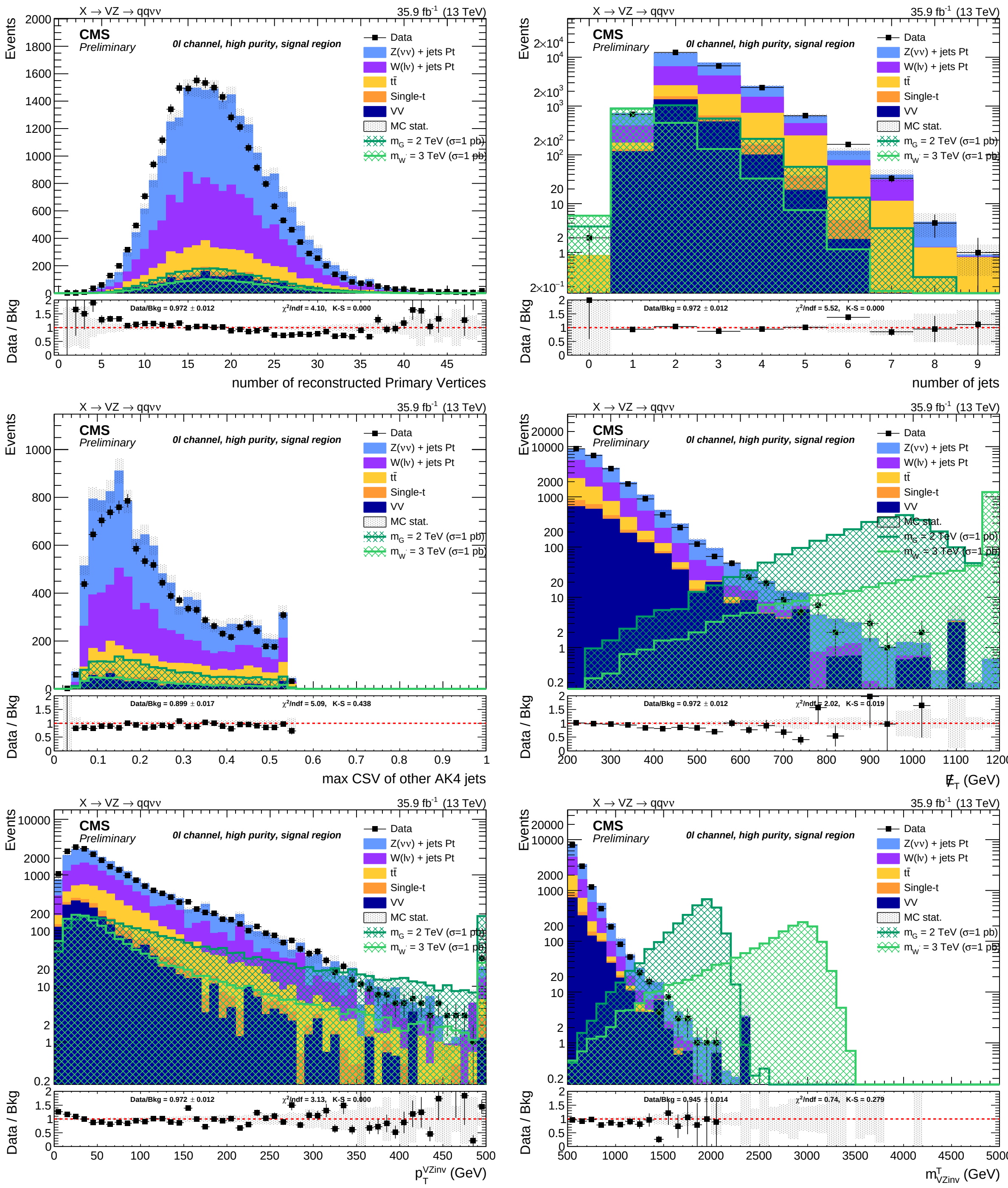


Figure 4.38: Top: number of reconstructed primary vertices (left) and number of AK4 jets in the event (right). Center: distribution of the b-tagging multivariate discriminant for the AK4 jets not included in the V jet cone (left) and E_T^{miss} distribution (right). Bottom: p_T of the VZ candidate (left) 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.

4.5 Background estimation technique

The goal of the analysis is to look for localized excesses in the m_{VZ}^T spectrum. The α *method* is used in searches for heavy resonances since Run 1 [94], and it has been introduced to be less dependent on the MC simulation for the background m_{VZ}^T estimation, due to the many sources of systematic uncertainties that are hard to understand and control. The two exclusive regions, *signal region* (SR) and *sidebands region* (SB), define a signal enriched or signal depleted phase space, respectively. First, the background normalization is extracted from data in the SB. Then, the α method extracts a predicted shape from the data in the SB to the SR using a transfer function (the α function) derived from simulation. The method relies on the assumption that the correlation between m_{VZ}^T and the groomed jet mass is reasonably well reproduced by the MC. The α-ratio is deemed to be more trustworthy since many systematic uncertainties would approximately cancel in the ratio.
Let's assume that, in the simplest case, only one dominant background is present. The α function is defined as the ratio of the two functions describing the simulated m_{VZ}^T shape in the SR and SB:

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

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

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

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

4.5.1 Background normalization

The first step in the background prediction consists in a proper estimation of the background normalization. The jet mass distributions of the three backgrounds (V + jets, Top, and VV) are described with functional forms determined by fits on the simulated backgrounds. The so-built templates are summed together, maintaining the relative weights between the three, and finally fitted

to the data in the jet mass sidebands. During the fit to data SB, the parameters of the V + jets background are left free to float and adapt to the data distribution. The integral of the final sum of the fitted functions over the SR jet mass range represents the background yield prediction in the SR. The empirical functional forms for each background are chosen to reflect the physics properties of the samples. In the low-purity category, the V + jets background is a falling background with no peaks, modelled as a power law, while in the high-purity category the V + jets background component is characterized by a broad distribution roughly centered at m_V, modelled as a gaussian, with an exponential tail at high mass values. The exponential falling VV m_V spectrum shows a peak, corresponding to the reconstruction of a vector boson hadronic decay. The hadronic decays of W and Z bosons cannot be distinguished, hence they are modelled together as a gaussian. For the jet mass spectrum of the Top backgrounds, two peaks corresponding to the W and top quark mass can be observed; they are modelled as gaussian functions, superimposed to a falling exponential background.
An extended likelihood fit is performed, hence the functional forms chosen to build the jet mass templates are normalized to unity (becoming probability density functions) through normalization 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 anf f_1, the exponential parameter a, the two gaussians means b-d and variances c-e, the error function offset o and width w):

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

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

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

Category	V + jets	alt. V + jets	Top	VV
low-purity	ErfPow2	ExpGaus	ErfExpGaus2	ExpGaus
high-purity	ExpGaus	ErfExpGaus	ErfExpGaus2	ExpGaus

Table 4.16: Expected background yield in the SB ($30 < m_V < 65$ GeV, $m_V > 135$ GeV) and in the SR ($65 < m_V < 105$ 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

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

- the statistic uncertainty is the uncertainty of the fit to the V + jet background performed on data SB;
- the systematic uncertainty is the propagation of the uncertainties of the fits to the VV and Top backgrounds performed on simulations, to the fit performed on data SB to extract the V + jets functional parameters;
- the alternative function uncertainty describes the discrepancy in the background yield in SR depending on the choice of the function to describe the V + jets background.

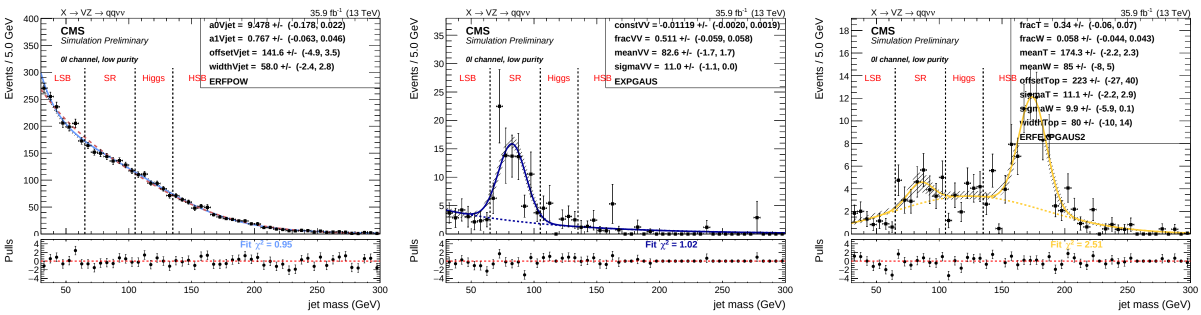


Figure 4.39: Fit to the simulated m_V in the low-purity category for the three backgrounds: V + jets (left), VV (center), Top (right). For the main background prediction, the alternative function is displayed with a dotted red line, superimposed to the main choice (continuous light blue curve).

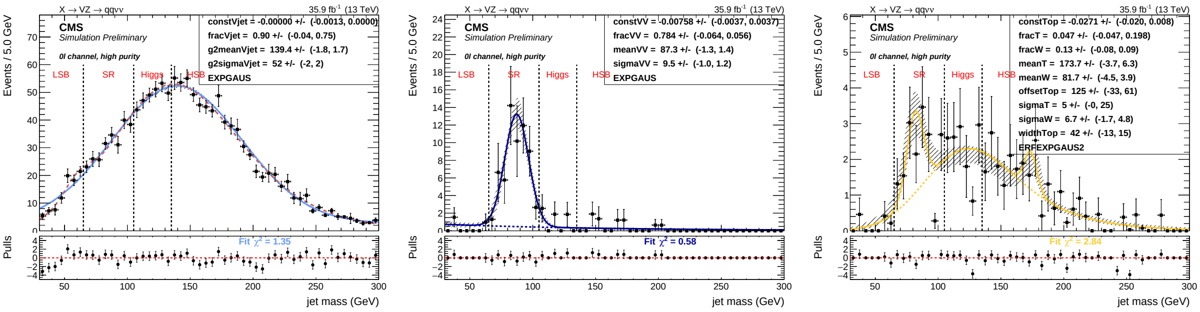


Figure 4.40: Fit to the simulated m_V in the high-purity category for the three backgrounds: V + jets (left), VV (center), Top (right). For the main background prediction, the alternative function is displayed with a dotted red line, superimposed to the main choice (continuous light blue curve).

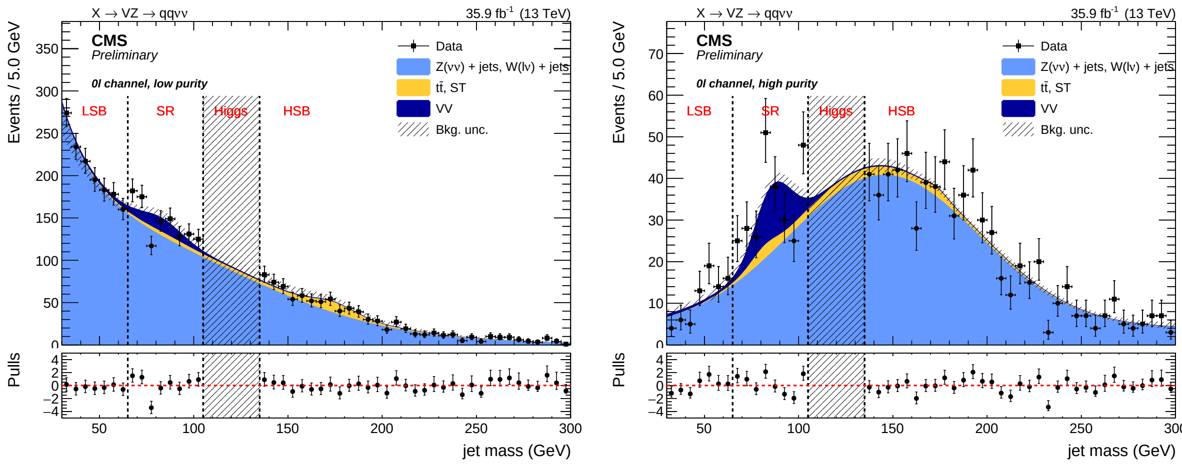


Figure 4.41: Background yield prediction in the signal region, after the fit to data sidebands, in the low- (left) and high-purity category (right). Data and predictions are in agreement.

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 + jets background ($f_{\text{SR}}^{\text{MC, } V\text{ + jets}}(m_{VZ}^T)$, $f_{\text{SB}}^{\text{MC, } V\text{ + jets}}(m_{VZ}^T)$), Top production ($f_{\text{SR}}^{\text{MC, Top}}(m_{VZ}^T)$, $f_{\text{SB}}^{\text{MC, Top}}(m_{VZ}^T)$), and diboson background ($f_{\text{SR}}^{\text{MC, VV}}(m_{VZ}^T)$, $f_{\text{SB}}^{\text{MC, VV}}(m_{VZ}^T)$). 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 + jets spectrum is normalized according to the data-driven prediction obtained in sec. 4.5.1. The functions describing the V + 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:

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

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

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

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

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

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

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

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

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

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

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

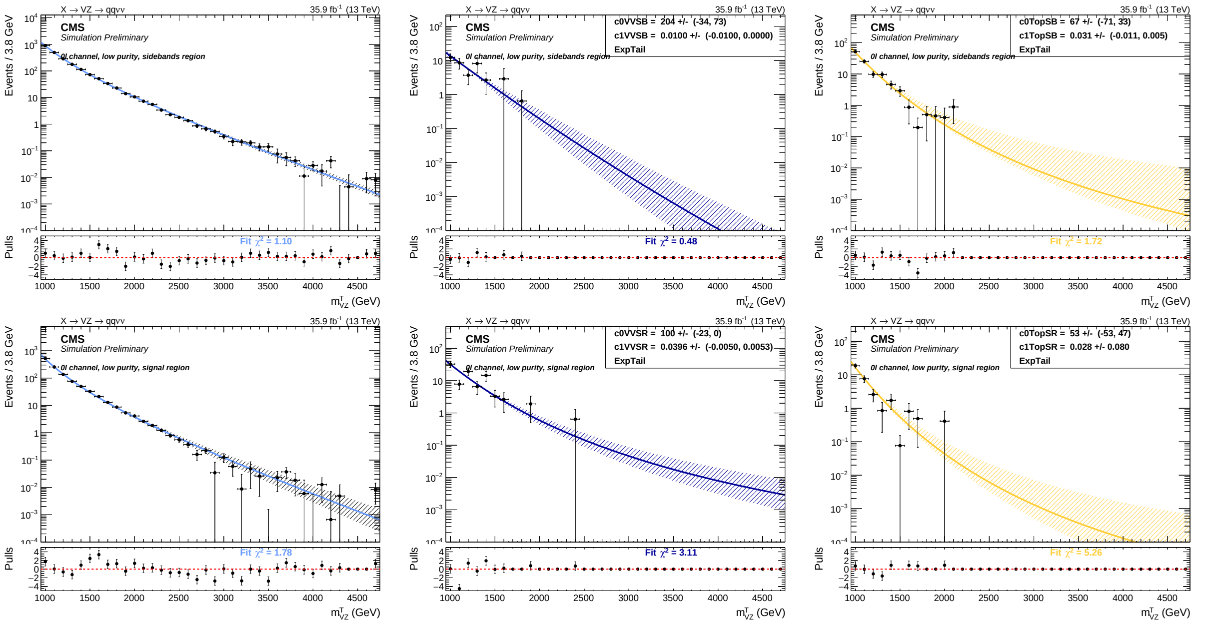


Figure 4.42: Low-purity category. Top: fits to the simulated background components V + jets (left), VV (center), Top (right) in the sidebands (SB). Bottom: fits to the simulated background components V + jets (left), VV (center), Top (right) in the signal region (SR).

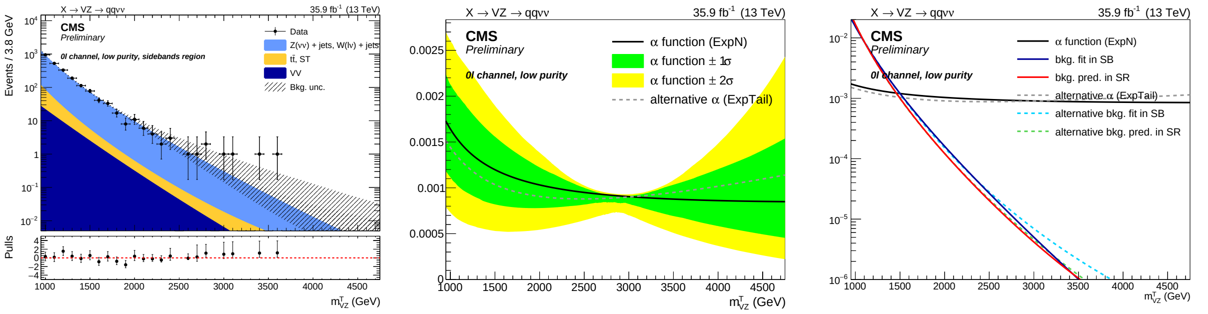


Figure 4.43: Low-purity category. Result of the fit to data in the SB (left), α-ratio function (center), and α function compared to the background shape in both SB and SR (right). The black line, with the corresponding 1σ (green) and 2σ (yellow) uncertainty bands, represents the α function. The gray 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 + jets and the W + 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 + jets and W jets background, separately, and for the combination of the two. In fig. 4.49, the α functions calculated for Z + jets background (red dotted line) and for W + 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

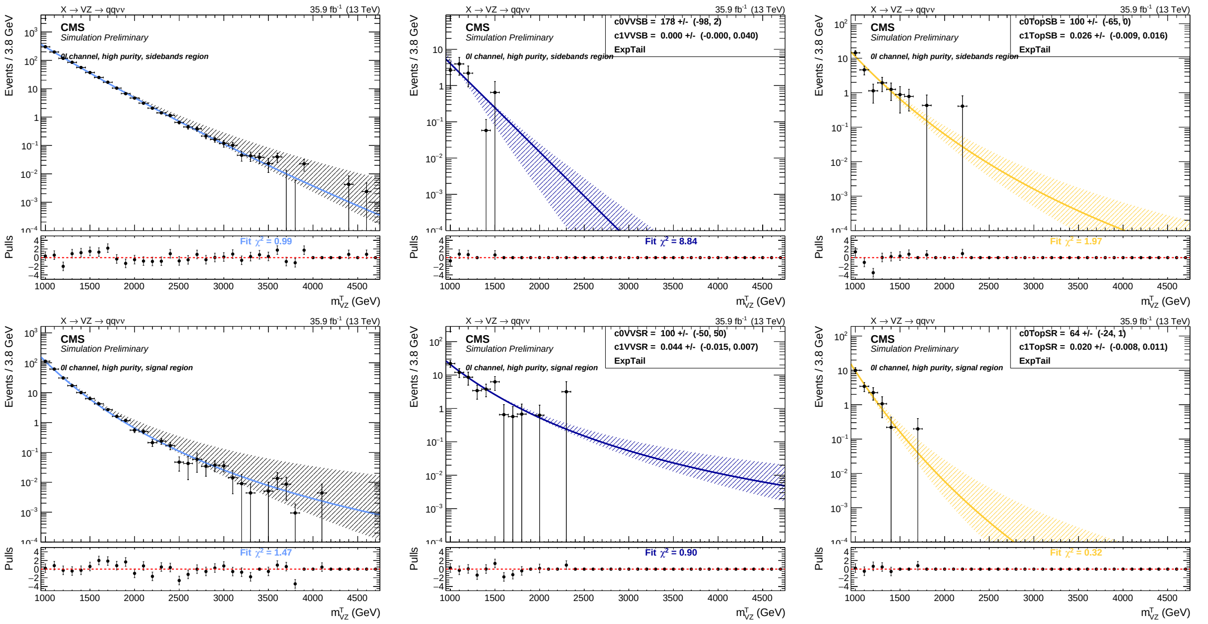


Figure 4.44: High-purity category. Top: fits to the simulated background components V + jets (left), VV (center), Top (right) in the sidebands (SB). Bottom: fits to the simulated background components V + jets (left), VV (center), Top (right) in the signal region (SR).

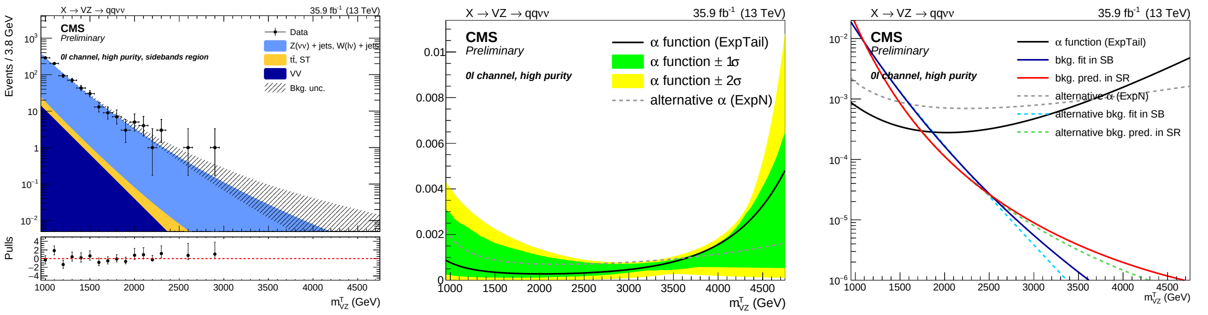


Figure 4.45: High-purity category. Result of the fit to data in the SB (left), α-ratio function (center), and α function compared to the background shape in both SB and SR (right). The black line, with the corresponding 1σ (green) and 2σ (yellow) uncertainty bands, represents the α function. The gray 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.

high-purity category (right plot).

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

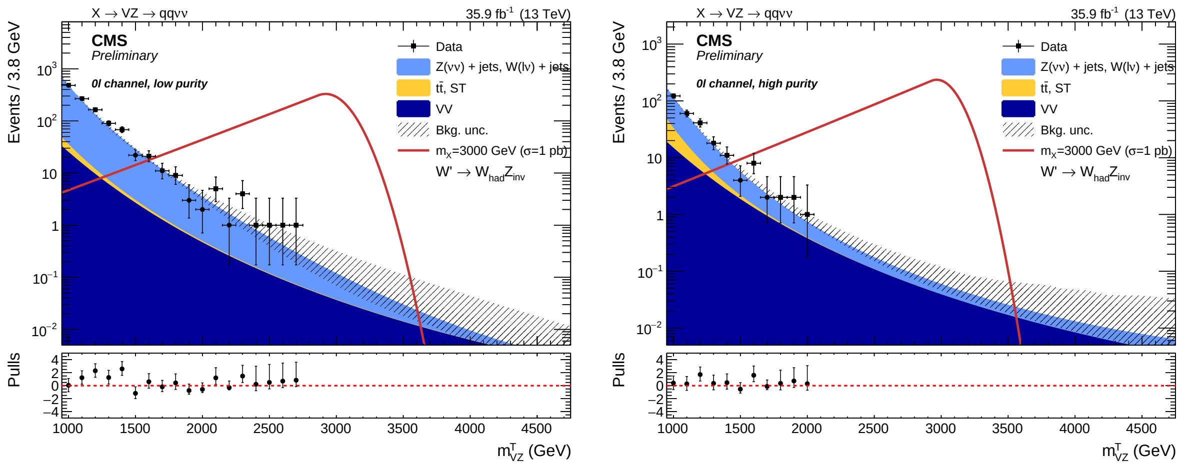


Figure 4.46: Expected background predicted with the α method in the low- (left) and high-purity category (right), compared to observations (black markers) and a signal hypotesis of a spin-1 $W^{'}$ of mass 3 TeV.

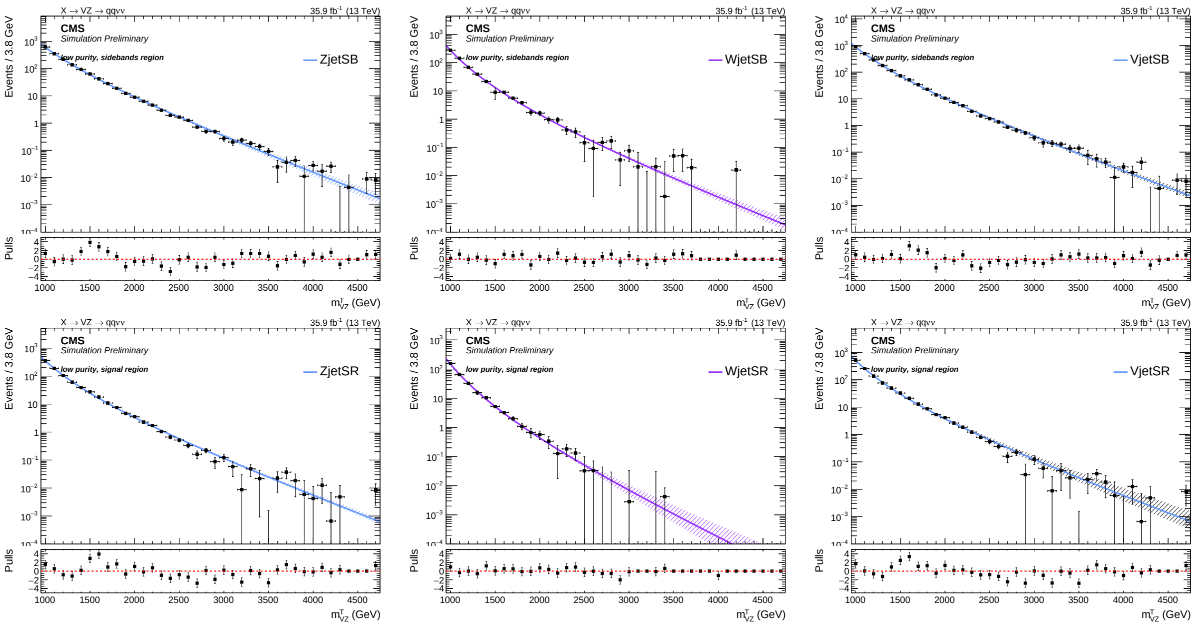


Figure 4.47: Validation of the α method, low-purity category. Top: fits to the simulated background components Z + jets (left), W + jets (center), and their combination V + jets (right), in the sidebands (SB). Bottom: fits to the simulated background components Z + jets (left), W + jets (center), and their combination V + 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 + jets background is reliable and can be used to model the background in the search for potential excesses in the signal region

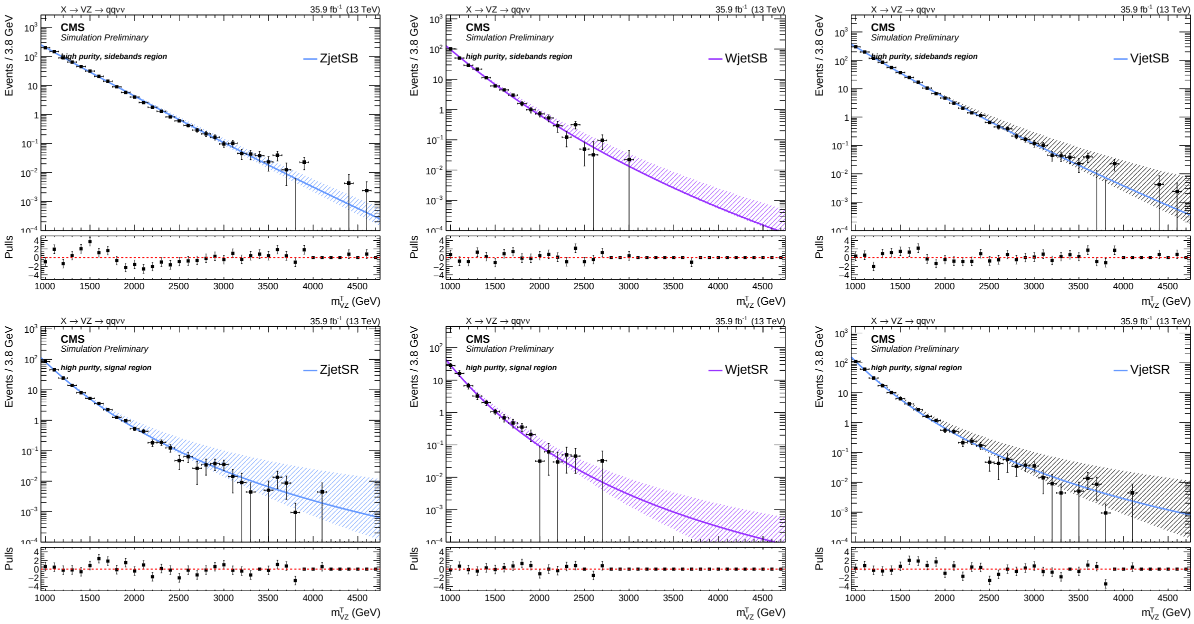


Figure 4.48: Validation of the α method, high-purity category. Top: fits to the simulated background components Z + jets (left), W + jets (center), and their combination V + jets (right), in the sidebands (SB). Bottom: fits to the simulated background components Z + jets (left), W + jets (center), and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

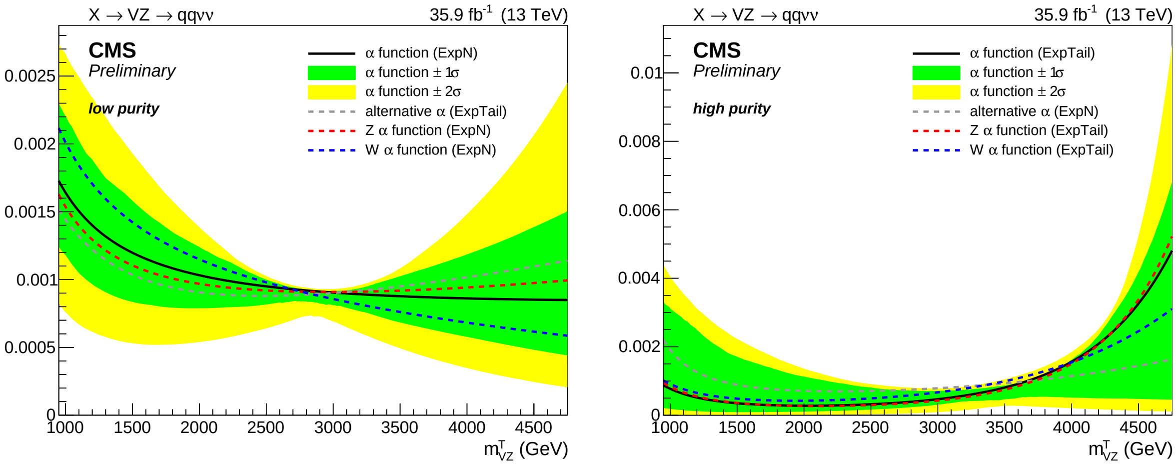


Figure 4.49: Validation of the α method: α functions calculated for Z + jets background (red dotted line) and for W + jets background (blue dotted line) separately, and α function for the total V + jets background (black solid line). Left: low-purity category; right: high-purity category.

defined in the analysis.

The last check performed is a study of the impact of the choice of the function to describe the V + jets background on the very last result of the analysis, namely the exclusion limit on the signal cross-section times branching fraction. The procedure of the limit extraction is discussed in detail in sec. 4.7. The main and alternative functions chosen to parametrize the dominant background 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

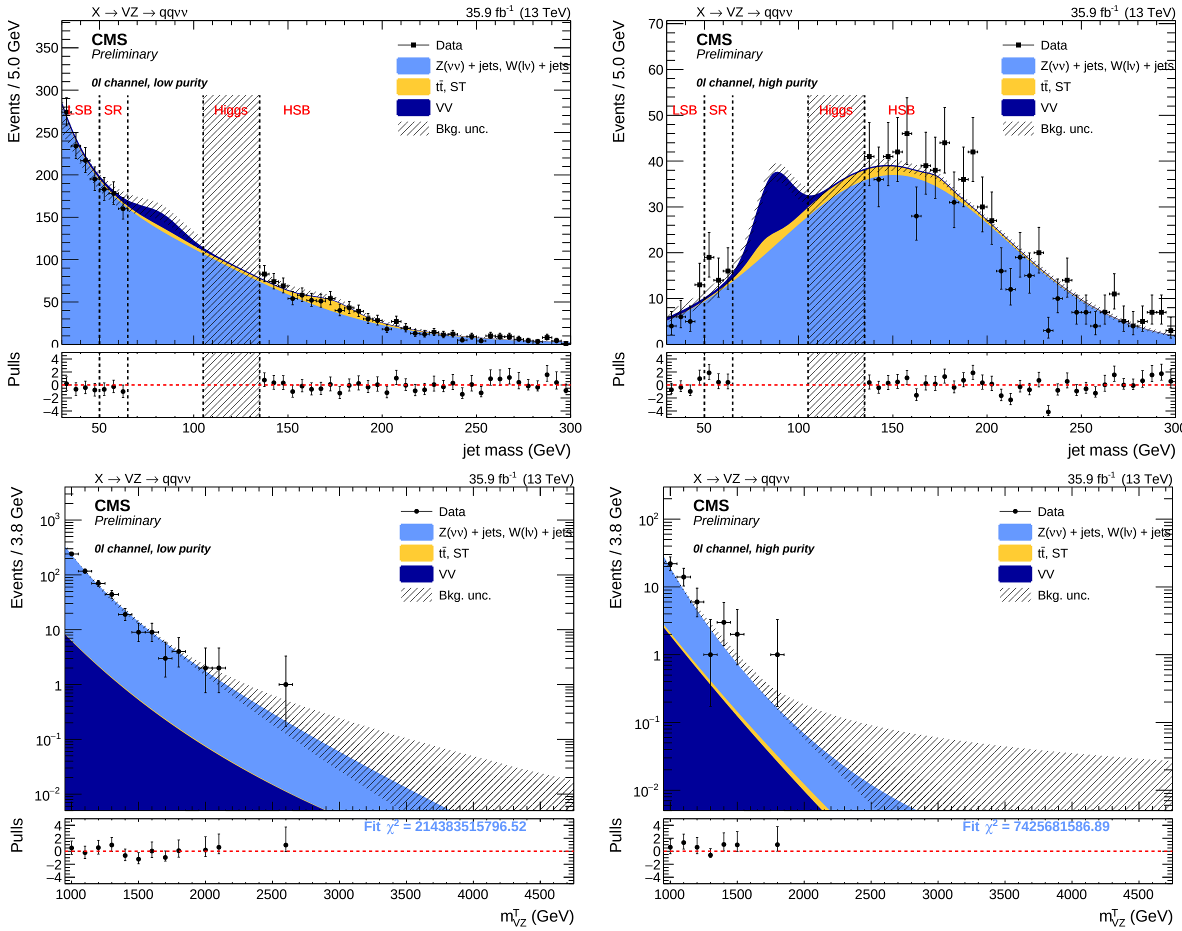


Figure 4.50: Top: results of the fit to the m_V spectrum in data, in the sidebands defined for the α method validation: low-sideband ($30 < m_V < 50$ GeV) and high-sideband ($m_V > 135$ GeV) (left: low-purity category, right: high-purity category). Bottom: results of the fits to the m^T_{VZ} spectrum, 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.

background shape prediction of the transverse mass are displayed. They are obtained by choosing the main function to describe the main background (red curve) and the alternative function (green curve); the two predictions are in agreement and very close to each other, for both low- (left) and high- (right) purity category. In fig. 4.51 (center), the 95% CL exclusion limits on cross-section times branching fraction are displayed for a spin-2 bulk graviton hypotesis, as a function of the mass of the resonance. The same figure of merit is shown in fig. 4.51 (bottom), considering a spin-1 W' hypotesis. In the plots, the exclusion limits are calculated by choosing the main function to describe the $V+$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 ($<< 1\%$).

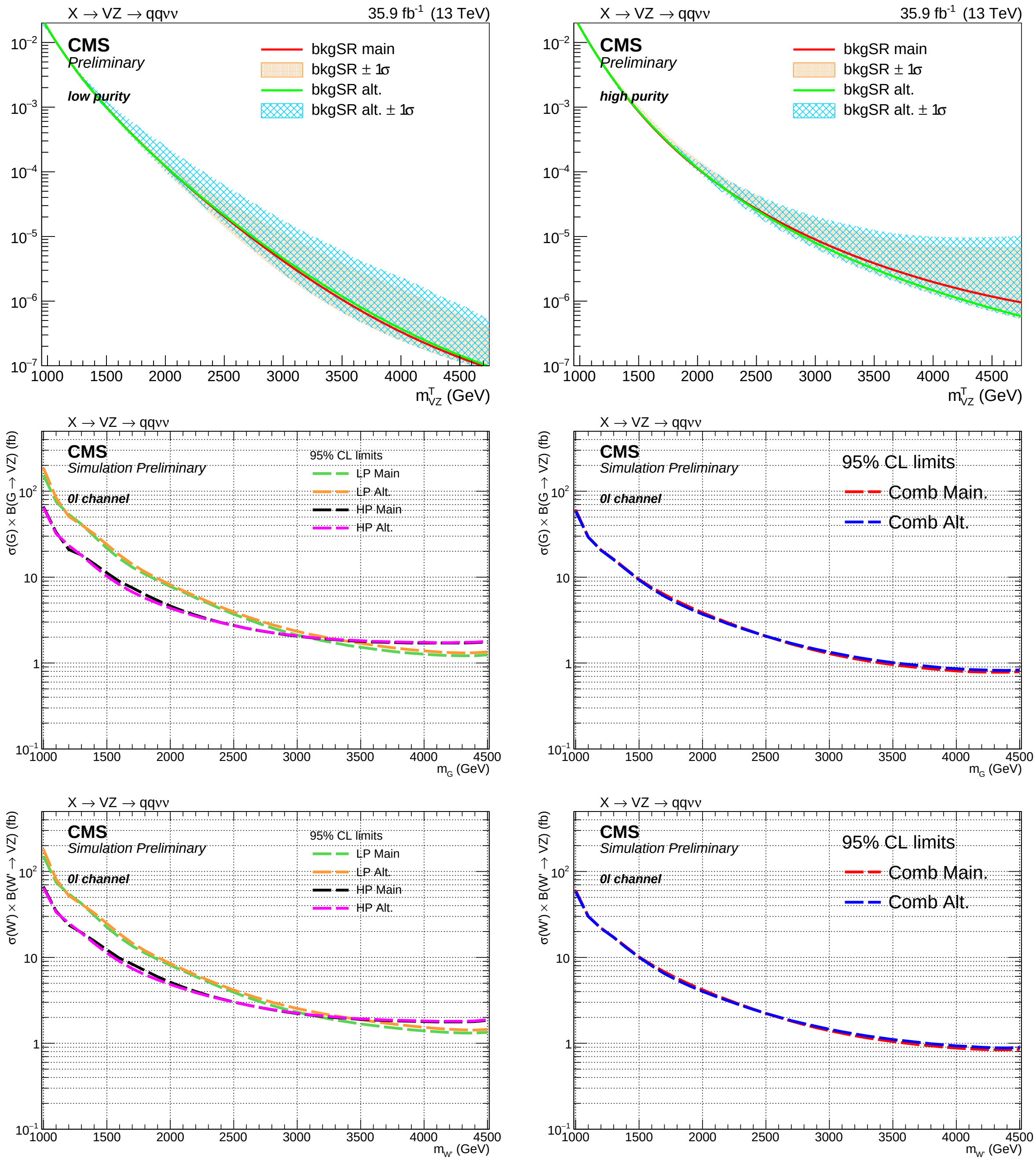


Figure 4.51: Validation of the α method: impact of the choice of the function to describe the dominant V + jets background. Top: fit results of the background shape prediction in the SR obtained with the main function (red curve) and the alternative function (green curve), for low- (left) and high- (right) purity categories. Center: exclusion limits on cross-section times branching fraction for a spin-2 bulk graviton hypotesis, 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' hypotesis.

4.5.4 Signal modeling

The simulated signal samples, with different resonance mass hypoteses, 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 [95, 96], 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.

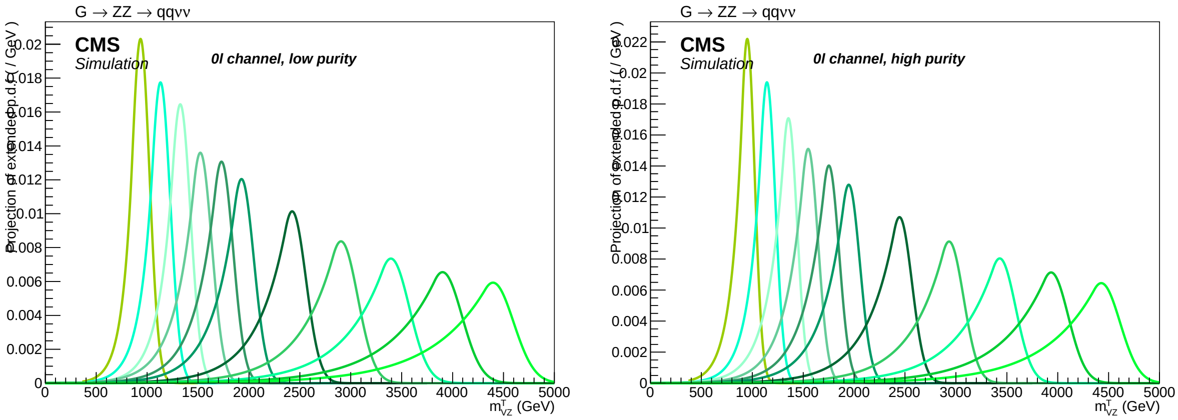


Figure 4.52: Interpolation of the signal as a function of the resonace transverse mass m^T_{VZ}, for a spin-2 (bulk graviton) signal hypothesis with an arbitrary cross section of 1 pb in the low- (left) and high-purity category (right).

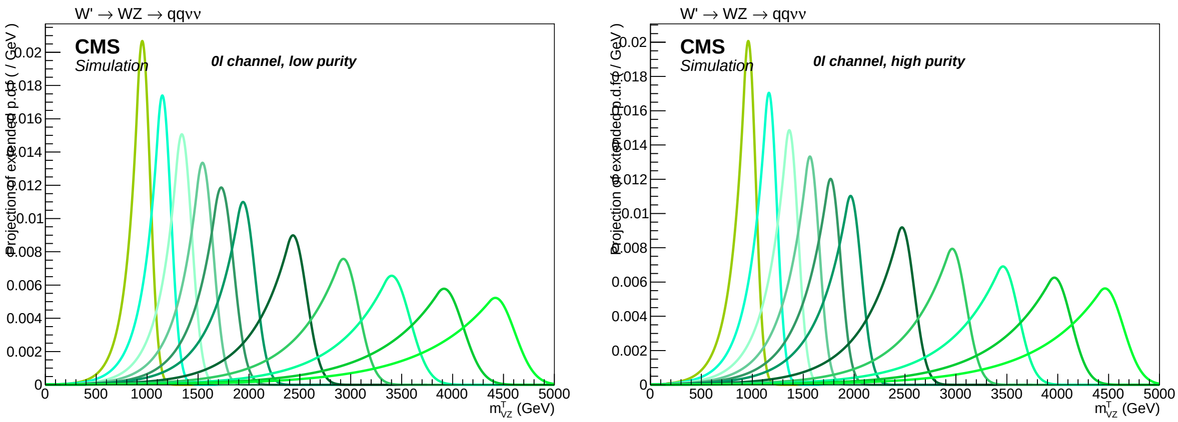


Figure 4.53: Interpolation of the signal as a function of the resonace transverse mass m^T_{VZ}, for a spin-1 ($W^{'}$) signal hypothesis with an arbitrary cross section of 1 pb in the low- (left) and high-purity category (right).

4.5.4.1 Signal parametrization

The signal is parametrized by interpolating the fitted parameters separately for each category in order to have a continous variation of the signal shape for every possible m^T_{VZ} 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 the spin-1 signal model. Shape systematic uncertainties, as described in sec. 4.6, are taken into account while describing the mean and sigma of the gaussian core, and they are related to the effects of the jet mass scale and resolution. Other shape parameters describing the tails of the Crystal Ball are fitted as 3$^{\text{rd}}$ degree polynomial.

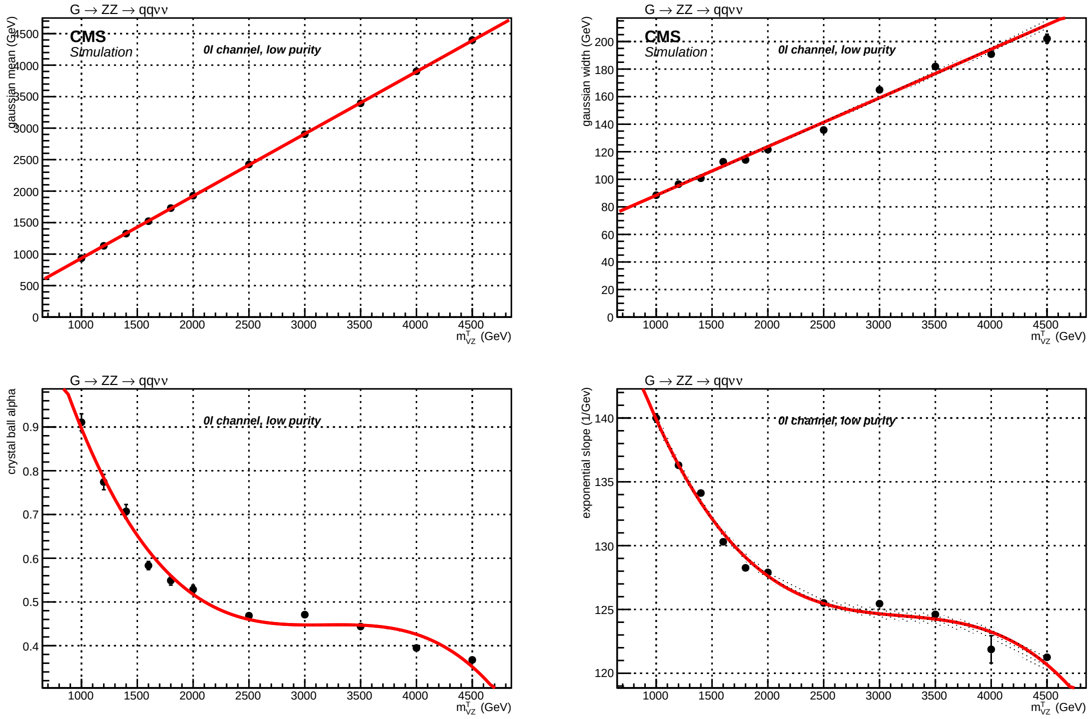


Figure 4.54: Interpolation of the fitted parameters as a function of the resonance transverse mass m^T_{VZ}, for a spin-2 (bulk graviton) signal hypothesis, low-purity category.

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

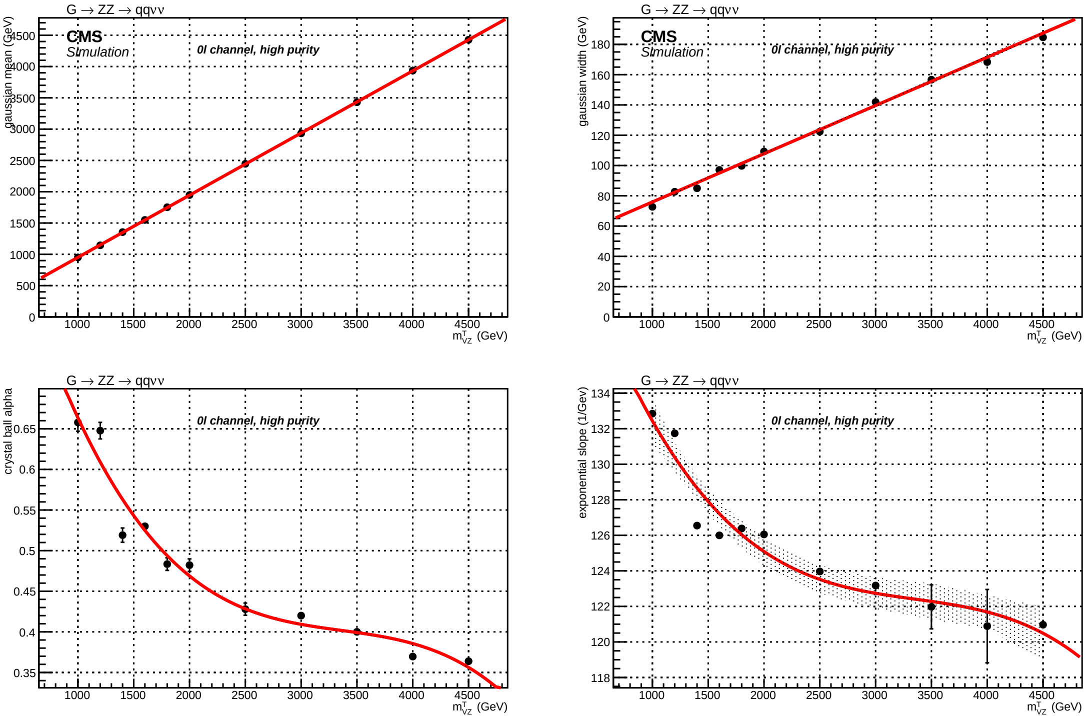


Figure 4.55: Interpolation of the fitted parameters as a function of the resonace transverse mass m^T_{VZ}, for a spin-2 (bulk graviton) signal hypothesis, high-purity category.

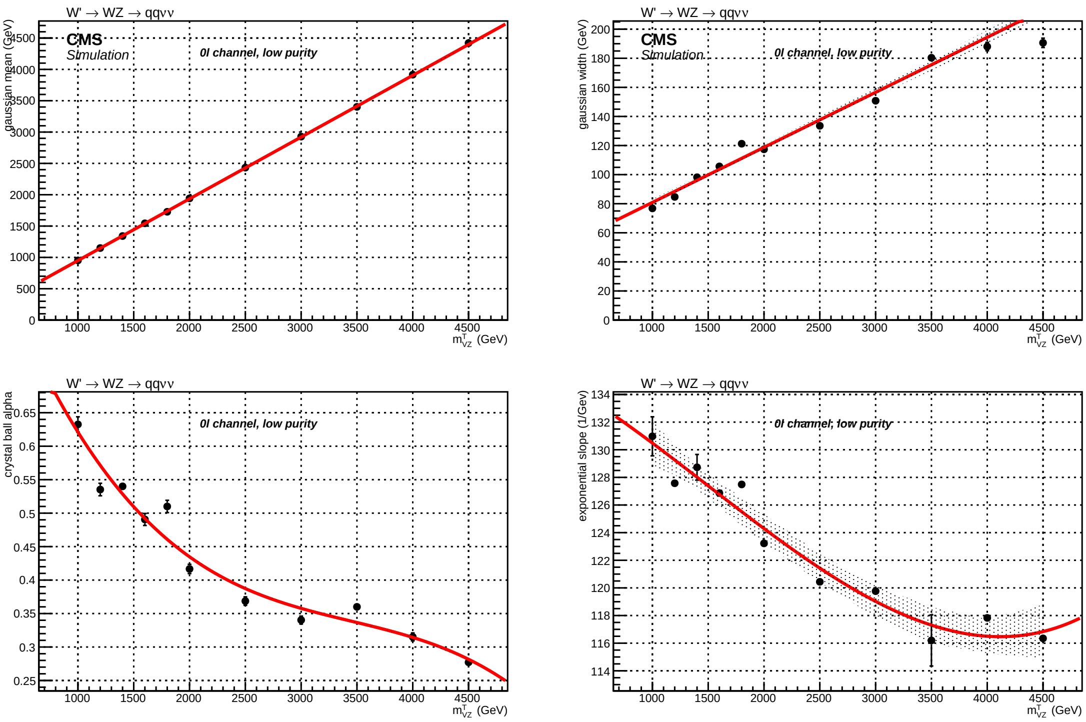


Figure 4.56: Interpolation of the fitted parameters as a function of the resonace transverse mass m^T_{VZ}, for a spin-1 (W') signal hypothesis, low-purity category.

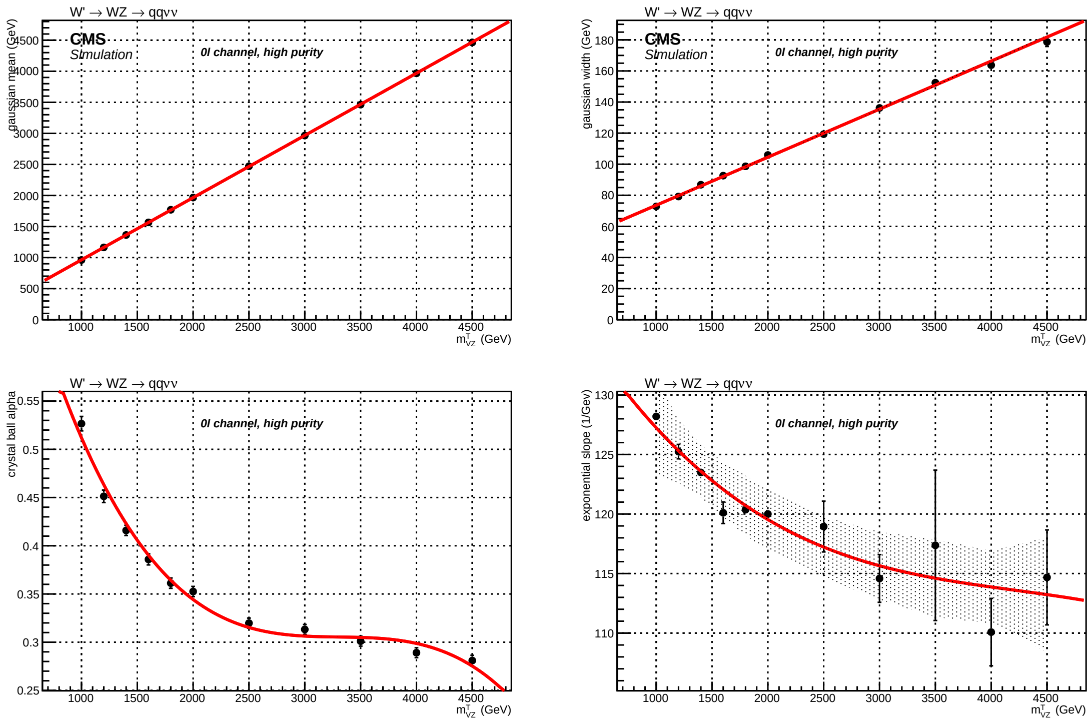


Figure 4.57: Interpolation of the fitted parameters as a function of the resonace transverse mass m_{VZ}^{T}, for a spin-1 (W') signal hypothesis, high-purity category.

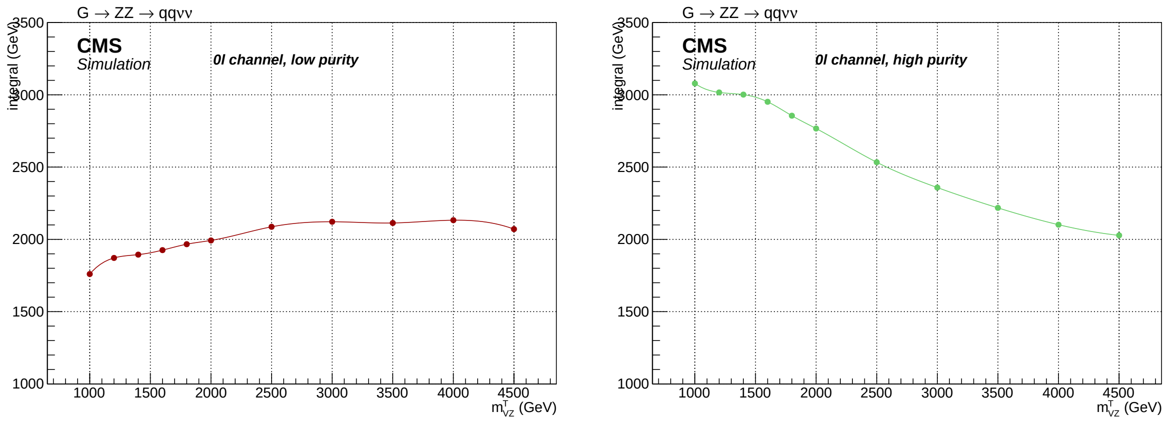
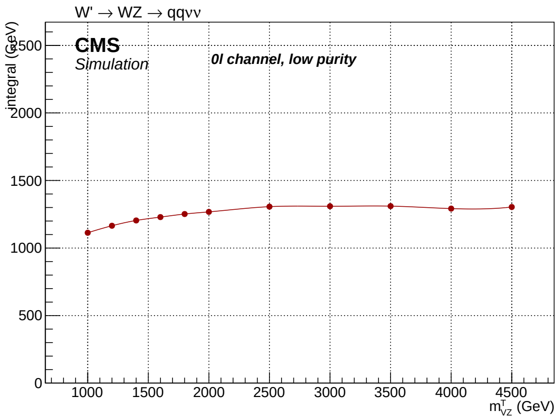


Figure 4.58: Interpolation of the signal normalization as a function of the resonace transverse mass m_{VZ}^{T}, for a spin-2 (bulk graviton) signal hypothesis. From left to right: low-purity, high-purity.



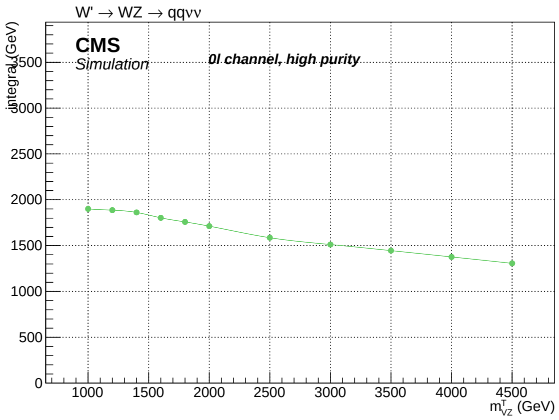


Figure 4.59: Interpolation of the signal normalization as a function of the resonace transverse mass m^T_{VZ}, for a spin-1 (W') signal hypothesis. From left to right: low-purity, high-purity.

4.6 Systematic uncertainties

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

4.6.1 Uncertainties affecting the data-driven main background estimation

4.6.1.1 Normalization

The predictiond of the normalization and shape of main background, V + jets, are both taken from data. The normalization is extracted from fits to the jet mass sidebands with arbitrary functions tested on simulation. The effects related to the contribution of the sub-dominant backgrounds are also taken into account, for both the normalization and the shape.

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

4.6.1.2 Shape

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

4.6.2 Uncertainties affecting the signal and the sub-dominant backgrounds

4.6.2.1 Trigger uncertainty

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

4.6.2.2 Jet momentum uncertainties

Jet uncertainties are evaluated in the signal regions by moving up and down by one standard deviation the source of the uncertainty. The two sources are the uncertainty on the jet energy correction, also identified as jet energy scale (JES) [68], [69], and the uncertainty due to the different jet momentum resolution (JER) [69].

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

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

4.6.2.3 Jet mass uncertainties

The soft drop PUPPI corrected jet mass is affected by two different uncertainties sources.
Soft drop jet mass calibration is varied within $\pm\sqrt{\left(\text{JES}^2_{\text{unc.}}+\text{JMS}^2_{\text{unc.}}\right)}$, where $\text{JES}_{\text{unc.}}$ is the uncertainty of the JES, described aboce, and $\text{JMS}_{\text{unc.}} = 0.0094$ is a constant coefficent (4.3.7, [68], [69]). The impact is calculated on signal and secondary backgrounds, both in normalization and shape.
As regarding the smearing, the soft drop PUPPI corrected jet mass of the signal samples and sub-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^T_{VZ}	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%

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^T_{VZ}	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%

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

4.6.2.4 V-tagging uncertainties

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

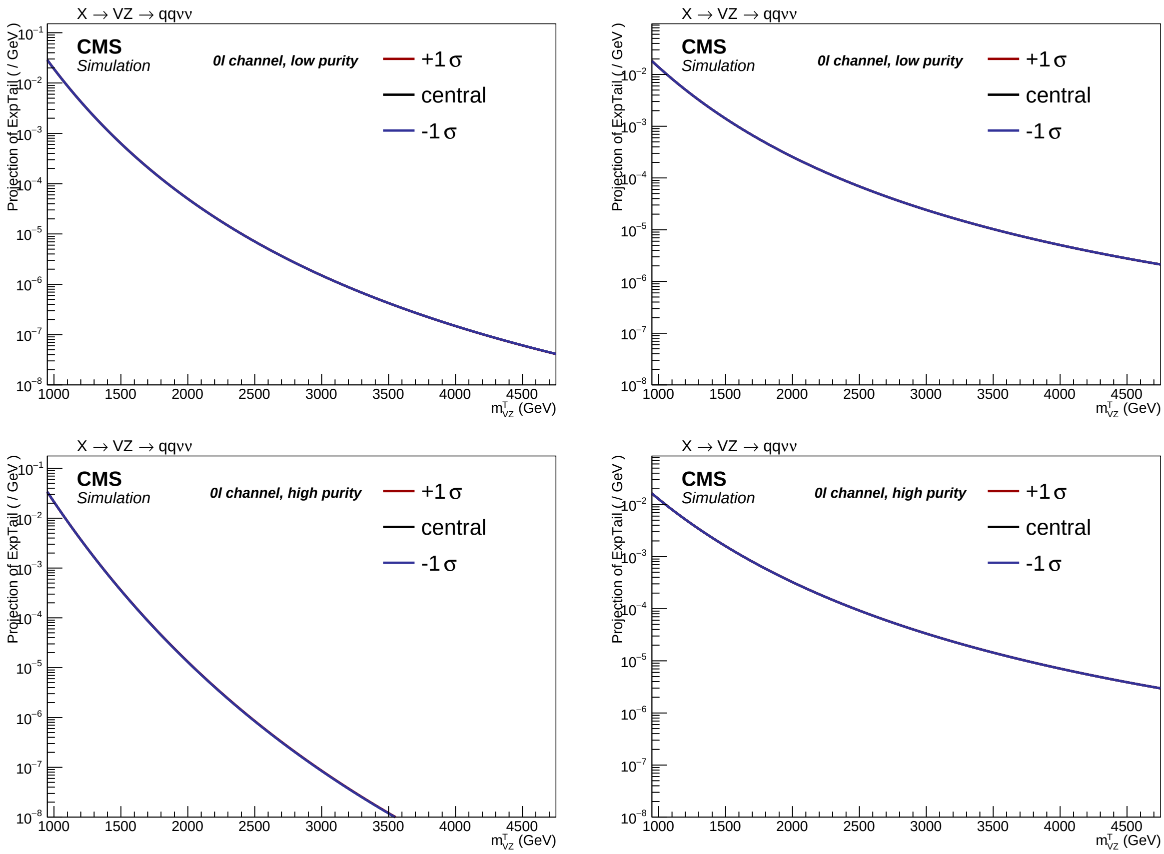


Figure 4.60: Shape variations due to jet mass calibration corrections obtained in the Top (left) and diboson (right) backgrounds, in the low-purity (top) and high-purity (bottom) category.

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 [97]). 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 [93].
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.

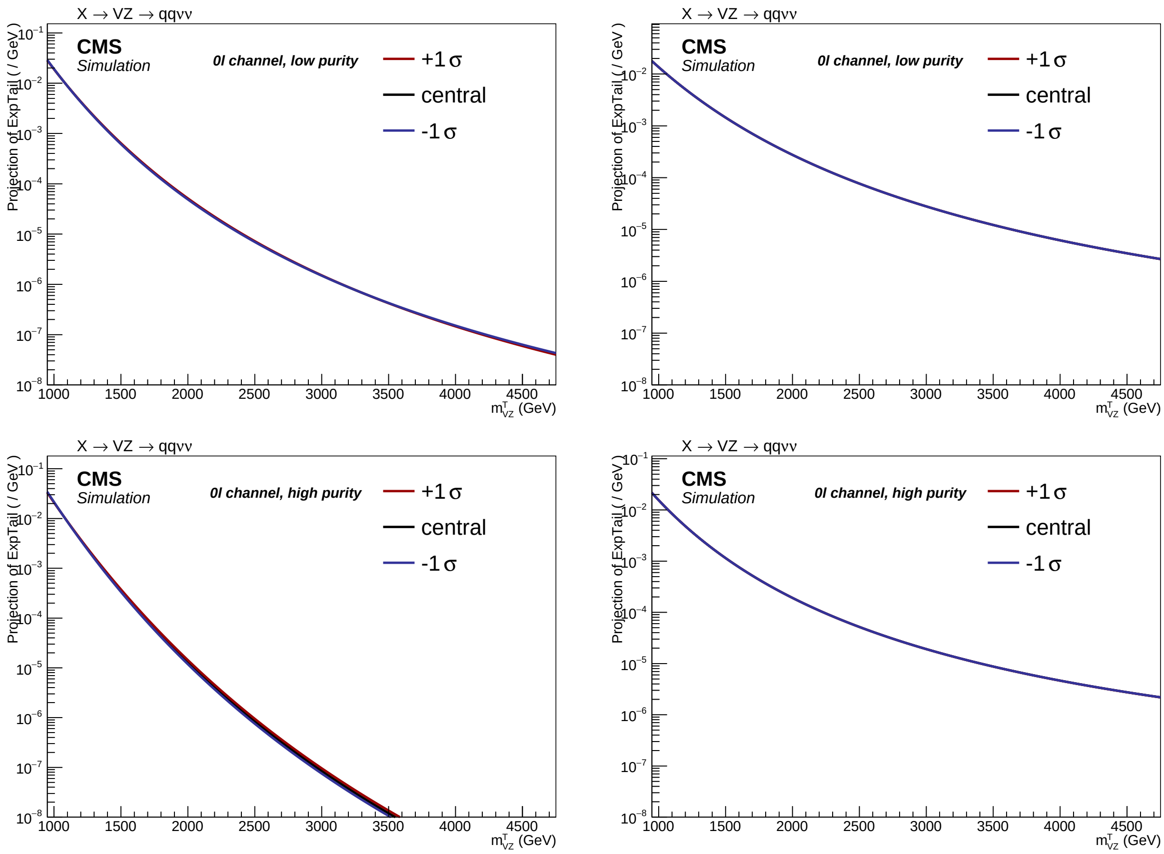


Figure 4.61: Shape variations due to jet mass resolution obtained in the Top (left) and diboson (right) backgrounds, in the low-purity (top) and high-purity (bottom) category.

4.6.2.6 Missing Energy uncertainties

As described in sec. 3.2.9.8, 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 $\vec{p}_T^{\text{miss}}$ 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 neglibile 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 $\vec{p}_T^{\text{miss}}$ 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.

4.6.2.7 Pile-up uncertainty

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

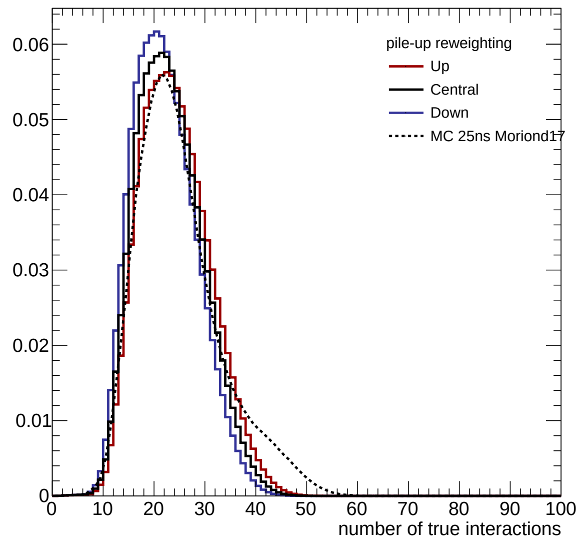


Figure 4.62: Pile-up scenario in 2016 data (black curve), and scenarios obtained by shifting up (red curve) or down (blue curve) the central value of the total inelastic cross-section (69 200 mb), compared to pile-up distribution simulated in Monte Carlo samples (dotted curve).

4.6.2.8 QCD renormalization and factorization scale uncertainties

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

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

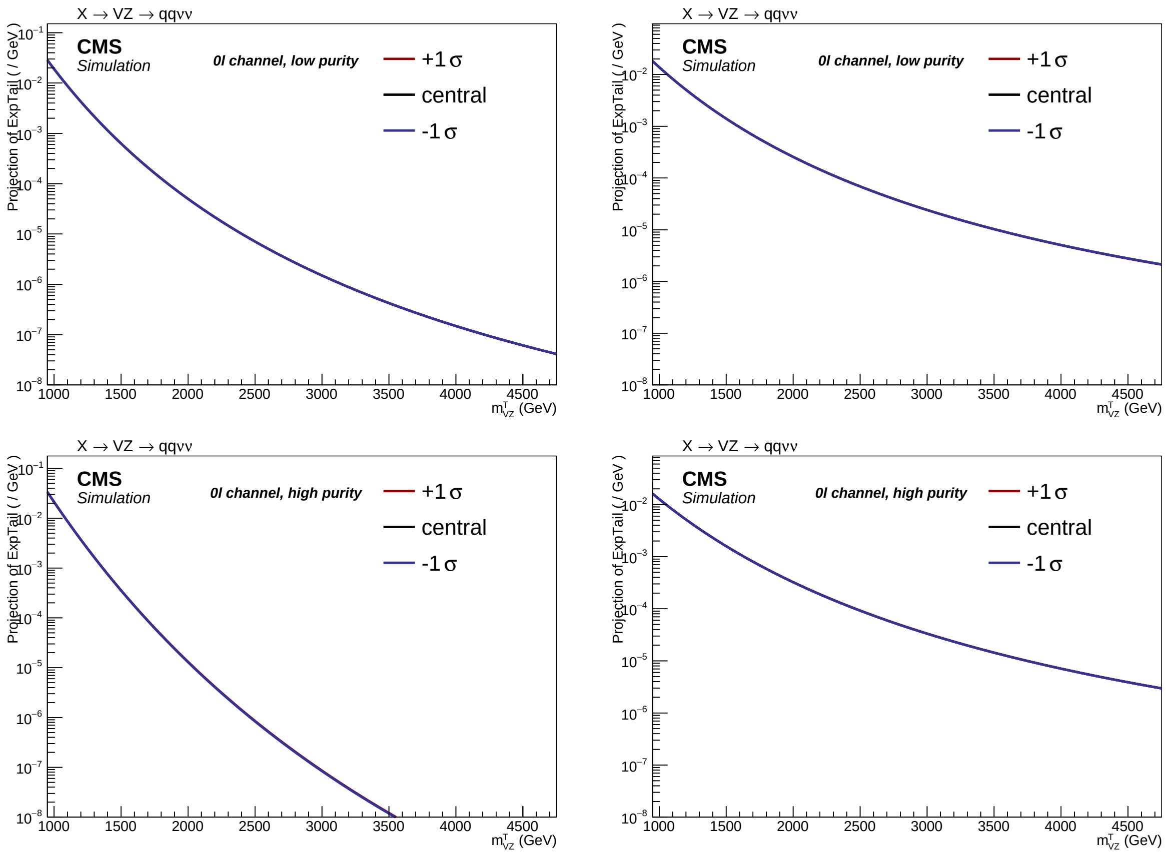


Figure 4.63: Shape variations due to pile-up uncertainty obtained in the Top (left) and diboson (right) backgrounds, in the low-purity (top) and high-purity (bottom) category.

4.6.2.9 PDF

Systematic uncertainties related to the PDFs parameters are estimated according to the PDF4LHC prescriptions [98], and using the NNPDF3.1 [99] 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^T_{VZ} 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 [88].

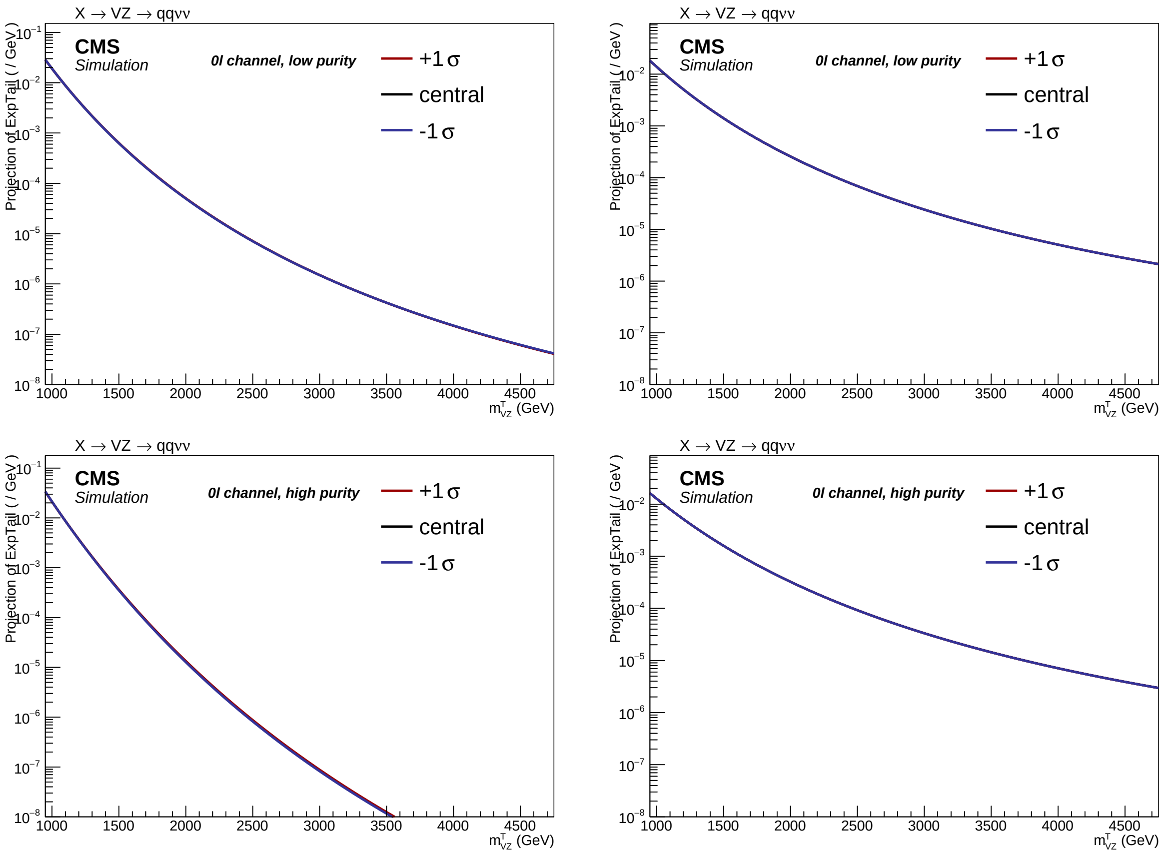


Figure 4.64: Shape variations due to QCD factorization in the Top (left) and diboson (right) backgrounds, in the low-purity (top) and high-purity (bottom) category.

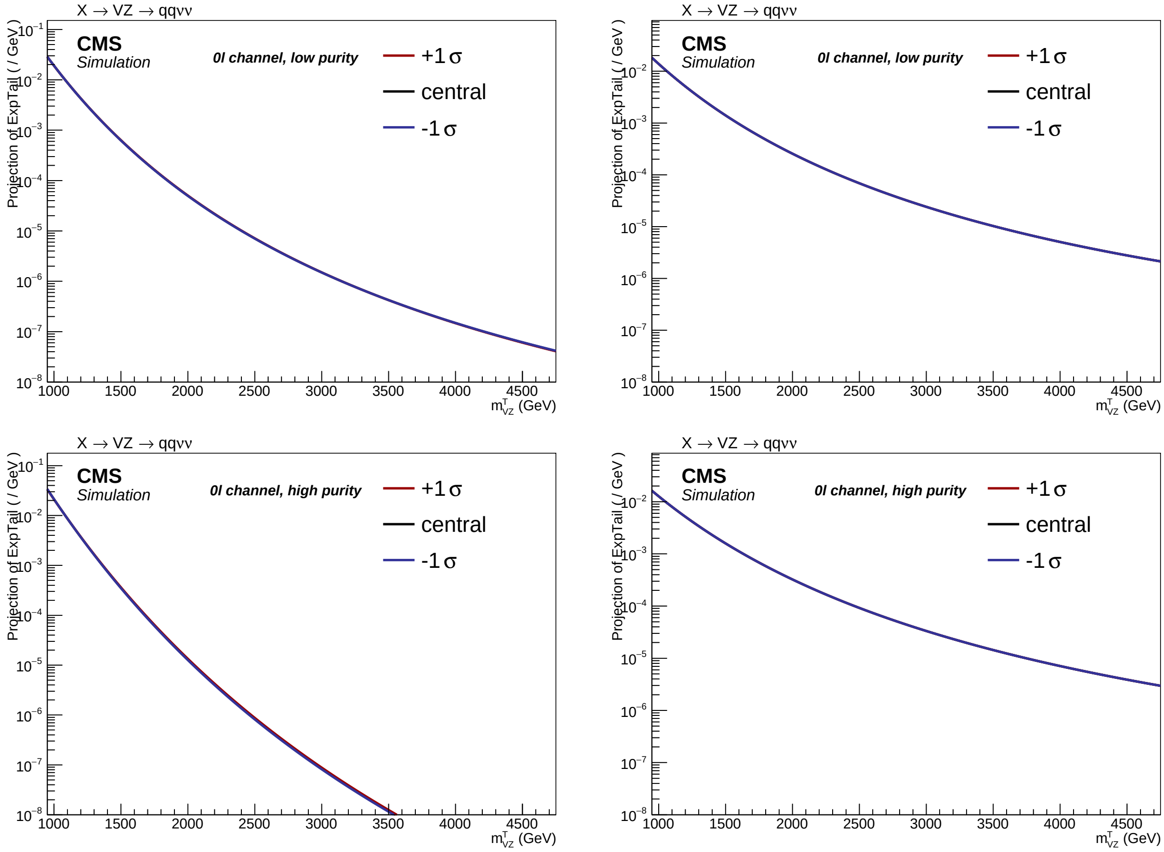


Figure 4.65: Shape variations due to PDF scale in the Top (left) and diboson (right) backgrounds, in the low-purity (top) and high-purity (bottom) category.

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 + jets	Top	VV	Signal
α-function	✓	✓	-	-	-
Bkg. normalization		4.8%(LP)	68.2%(LP)	11.4%(LP)	-
(fit)		14.7%(HP)	47.7%(HP)	19.1%(HP)	-
Bkg. normalization		4.9%(LP)	-	-	-
(alternative function)		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)	1.7% (LP)	3.2-9.4% (LP)
	-	-	2.8% (HP)	3.3% (HP)	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%

4.7 Results and interpretation

4.7.1 Statistical approach

4.7.1.1 The modified frequentist approach: asymptotic formulae to extract an upper limit on signal strength

The modified frequentist approach, also known as CL_s criterion [100–102], is used to determine the 95% confidence level upper limit on the signal contribution in the data.
The parameters used to model the data distribution are the background event yield, b, the signal event yield s, predicted by the theoretical model, the signal strength modifier μ, parametrizing how much the signal yield deviates from the model expectation s, and the nuisance parameters θ, namely, the uncertainties affecting the signal and background yields, that can be seen as functions of the nuisances: $b(\theta)$, $s(\theta)$. In this approach, the uncertainties are considered either as fully correlated (100%) or uncorrelated.
The likelihood function is built starting from a Poissonian probability density function:

$$\mathfrak{L}(\text{ data }|\mu,\theta) = \text{Poisson (data }|\mu\cdot s(\theta)+b(\theta))\cdot p(\tilde{\theta}|\theta), \tag{4.18}$$

where "data" can either be real or generated pseudo-data, whilst $p(\tilde{\theta}|\theta)$ is the probability distribution of the nuisance parameters, inferred through an independent dataset $\tilde{\theta}$. Considering an unbinned likelihood, where k events have been observed,

$$\text{Poisson (data }|\mu\cdot s(\theta)+b(\theta)) = \frac{1}{k}\prod_i\left(\mu S f_s(x_i)+B f_b(x_i)\right)\times e^{-(\mu S+B)}, \tag{4.19}$$

where f_s and f_b are the probability density functions for signal and background for an observable x, and S and B are the total expected signal and background event yields.
The measurement of the compatibility of data with the signal plus background or the background-only hypoteses is performed by defining a likelihood ratio test statistics $\tilde{q}_\mu$ [103],

$$\begin{gathered}\tilde{q}_\mu = -2\log\frac{\mathfrak{L}(\text{data }|\mu,\hat{\theta}_\mu)}{\mathfrak{L}(\text{data }|\hat{\mu},\hat{\theta})},\\ 0\le\hat{\mu}\le\mu.\end{gathered} \tag{4.20}$$

The quantities $\hat{\mu}$ and $\hat{\theta}$ are global maxima of the likelihood, while $\hat{\theta}_\mu$ is the conditional maximum, given μ. The signal strength $\hat{\mu}$ is defined positive, the upper boundary $\hat{\mu}\le\mu$ is set in order to avoid to consider upward fluctuations in data (namely, when the global maximum is larger than the hypotesis μ) as an incompatibility with the signal hypotesis (μ).
Given the μ hypotesis, the test statistic value is measured on data, and labelled as $\tilde{q}_\mu^{\text{obs.}}$. Parameters $\hat{\theta}_0^{\text{obs.}}$ and $\hat{\theta}_\mu^{\text{obs.}}$ are calculated by maximizing the likelihood function 4.18. Toy Monte Carlo pseudo-data are then generated to build the probability density functions $f(\tilde{q}_\mu|\mu,\hat{\theta}_\mu^{\text{obs.}})$ (signal with μ strength hypotesis) and $f(\tilde{q}_\mu|0,\hat{\theta}_0^{\text{obs.}})$ (background-only hypotesis). Nuisance parameters are fixed to their values measured on data, $\hat{\theta}_\mu^{\text{obs.}}$ and $\hat{\theta}_0^{\text{obs.}}$, but left free to float in fits that are required to evaluate $\tilde{q}_\mu$.
The p-values associated to signal plus background and background-only hypoteses are defined as:

$$\begin{aligned} p_\mu &= \mathscr{P}\left(\tilde{q}_\mu\ge\tilde{q}_\mu^{\text{obs.}}|\text{ signal + background }\right) = \int_{\tilde{q}_\mu^{\text{obs.}}}^{\infty} f(\tilde{q}_\mu|\mu,\hat{\theta}_\mu^{\text{obs.}})d\tilde{q}_\mu,\\ 1-p_b &= \mathscr{P}\left(\tilde{q}_\mu\ge\tilde{q}_\mu^{\text{obs.}}|\text{ background-only }\right) = \int_{\tilde{q}_\mu^{\text{obs.}}}^{\infty} f(\tilde{q}_\mu|0,\hat{\theta}_0^{\text{obs.}})d\tilde{q}_\mu.\end{aligned} \tag{4.21}$$

The CL_s is defined as the ratio of the above p-values:

$$CL_s = \frac{p_\mu}{1-p_b}. \tag{4.22}$$

Given the a-priori confidence level α, if $CL_s \leq \alpha$, a model with signal strength μ is excluded at $(1-\alpha)$ confidence level (C.L.). The 95% C.L. *observed* upper limit on the theoretical model is set by extracing μ from the equation $CL_s = 0.05$.
Similarly to the observed limit, an upper *expected* limit, along with the $\pm1\sigma$ and $\pm2\sigma$ uncertainty bands, can be extracted by generating pseudo-data under the background-only hypotesis, and by calculating the CL_s and 95% upper limit for each of the pseudo-data. A cumulative distribution of the calculated upper limits is then constructed: the 50% quantile corresponds to the median expected, the 2.5%, 16%, 84%, 95.5% quantiles correspond respectively to -2σ, -1σ, $+1\sigma$, $+2\sigma$ uncertainty bands.
Generating a large number of pseudo-data, however, can be a very expensive computational effort. This problem is overcome by profiting of asymptotic formulae [103], derived through Wilk's [104] and Wald's [105] theorems. The set of pseudo-data is replaced by only one dataset, the Asimov dataset: it corresponds to a dataset where the statistical fluctuations are suppressed, and hence every parameter is set to its expectation value. These values are then equivalent to the outcomes of a large sample of Monte Carlo simulations. The *expected* limit can therefore be calculated from the Asimov dataset.
By using the asymptotic formulae, the distribution of the test statistic $\tilde{q}_\mu$ is given by:

$$f(\tilde{q}_\mu|\mu) = \frac{1}{2}\delta(\tilde{q}_\mu) + \begin{cases} \frac{1}{2\sqrt{2\pi}}\frac{1}{\tilde{q}_\mu}e^{-\tilde{q}_\mu/2} & 0 < \tilde{q}_\mu \leq \mu^2/\sigma^2 \\ \frac{1}{2\sqrt{2\pi}}\frac{1}{2\mu/\sigma}e^{-\frac{1}{2}\frac{\left(\tilde{q}_\mu+\mu^2/\sigma^2\right)^2}{(2\mu/\sigma)^2}} & \tilde{q}_\mu > \mu^2/\sigma^2 \end{cases}; \qquad \sigma^2 = \frac{\mu^2}{\tilde{q}_{\mu,A}}, \tag{4.23}$$

where the test statistic $\tilde{q}_{\mu,A}$ is evaluated in the Asimov dataset. Once defined the inverse of the cumulative Gaussian distribution Φ, the asymptotic expression of the CL_s simplifies into:

$$CL_s = \frac{1-\Phi\left(\sqrt{\tilde{q}_\mu}\right)}{\Phi\left(\sqrt{\tilde{q}_{\mu,A}}-\sqrt{\tilde{q}_\mu}\right)}. \tag{4.24}$$

The expected upper limit and its N uncertainty bands are given by:

$$\begin{aligned} \mu_{up} &= \sigma\cdot\Phi^{-1}(1-0.5\alpha), \\ \mu_{up+N} &= \sigma\cdot\left[\Phi^{-1}(1-\alpha\Phi(N))+N\right]. \end{aligned} \tag{4.25}$$

4.7.1.2 Treatment of the systematic uncertainties

The nuisance parameters θ, introduced to describe the systematic uncertainties, are expected to have their own probability density function, $\rho(\theta)$, called *prior*, that is inferred by an additional set of measurements $\tilde{\theta}$, used to define the mean, the shape and the width of each uncertainty. The distribution of the priors depends on the type of uncertainty considered. Flat priors (namely, a constant value) are assigned to nuisances unconstrained a-priori; gaussian priors are assigned to nuisances allowed to assume both negative and positive values; log-normal priors are used for positively defined nuisances (such as cross-sections, efficiencies, luminosity, scale factors). For the purpose of this search, log-normal priors are being adopted. Partially correlated uncertainties, *i.e.* those associated to the α method parameters, are decorellated through linear transformations.

4.7.1.3 Computation of local p-values

The discovery of a signal can be inferred from data if a p-value that is incompatible with the background-only hypotesis is observed. The discovery test statistics is defined as:

$$q_0 = -2\log\frac{\mathfrak{L}(\text{data}\,|0,\hat{\theta}_0)}{\mathfrak{L}(\text{data}\,|\hat{\mu},\hat{\theta})}, \qquad \hat{\mu} \geq 0. \tag{4.26}$$

The boundary $\hat{\mu} \geq 0$ is motivated by the fact that an underfluctuation of the background is not considered as an evidence against the background-only hypotesis. The distribution $f(q_0|0,\hat{\theta}_0^{obs})$ is again built with pseudo-data, generated under the background-only hypotesis with nuisances $\hat{\theta}_0^{obs}$. The exact p-value is therefore:

$$p_0 = \mathscr{P}\left(q_0 \geq q_0^{\text{obs.}}|\text{ background-only }\right) = \int_{q_0^{\text{obs.}}}^{\infty} f(q_0|0,\hat{\theta}_0^{obs})\,dq_0, \tag{4.27}$$

that can be converted into a significance Z, once the convention of the one-sided Gaussian tail is adopted:

$$p_0 = \int_Z^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx. \tag{4.28}$$

By taking advantage of the Wilk's theorem, the p-value can be approximated as:

$$p_0^{\text{appr.}} = \frac{1}{2}\left[1-\text{erf}\left(\sqrt{q_0^{\text{obs.}}/2}\right)\right]. \tag{4.29}$$

Since the p-value depends on the phase-space considered (specifically, on the resonance mass hypotesis), eq. 4.27 is known as the *local p-value*. A scan of the local p-values is a measurement of a local departure from the background-only hypotesis. In case of a local excess, the global significance is computed by correcting the local significance with trial factors, that take into account the so-called *look-elsewhere* effect [106], namely, the probability to observe the same excess anywhere in the whole mass range.

4.7.2 Signal extraction strategy for the analysis

The background prediction, estimated with the α method (sec. 4.5), the signal parametrization (sec. 4.5.4.1), and the observed data are used as inputs for the signal extraction procedure. An unbinned maximum likelihood fit is performed on each purity category, and on the combination of the categories, in order to present, for each theoretical model taken into account, a global limit on the production cross-section times branching ratio, that is the parameter describing the signal yield and defining the signal strength r (equivalent to the signal strength μ discussed in the previous section).

4.7.2.1 Fit diagnostics: nuisances pulls and impacts

The systematic uncertainties, treated as log-normal nuisance parameters, are allowed to vary around their nominal values and are profiled during the maximum likelihood estimation of the signal strength. As a diagnostic, the profiled values (post-fit) of the nuisance parameters $\hat{\theta}$ are compared to their a-priori expectations (pre-fit) θ_0, in unities of the width of the gaussian core of the nuisance parameter $\Delta\theta$. The quantities $(\hat{\theta}-\theta_0)/\Delta\theta$ are called nuisance pulls, and they have been computed both

in the background-only hypotesis (blue bars) and in the signal plus background hypotesis (green bars), for the low- (fig. 4.66) and high-purity (fig. 4.67) categories. In fig. 4.66-4.67, the signal of a spin-2 bulk graviton with a mass of 3 TeV is considered. The distribution of pulls does not show any anomaly, since pulls are centered around zero (no discrepancies with the a-priori expectations) and their widths are around one (no strong deviations from the original assumption on the width of the nuisance distributions), for both the background-only and signal plus background hypoteses. The only pulls with mean values a bit shifted from zero or with widths smaller than one are related to α method parameters, that are under control.
The impacts of a nuisance parameter θ are defined as the shifts induced in the signal strength (r, the cross-section times branching fraction in this case) as θ is fixed and brought to its $+1\sigma$ or -1σ post-fit values, while all the other nuisance parameters are simultaneously profiled as log-normal. In fig. 4.68, impacts are calculated by combining the two purity categories, assuming a signal hypotesis of a spin-2 bulk graviton of mass 2.5 TeV. As expected a-priori (sec. 4.6), the most relevant systematic impacting the determination of the signal strength is represented by the uncertainty on the V-tagging procedure. No pathological behaviour can be observed.

Figure 4.66: Nuisance pulls for the low-purity category, calculated under both the background-only (blue bars) and signal plus background hypotheses (green bars). A signal hypotesis of a spin-2 bulk graviton of mass 3 TeV is considered.

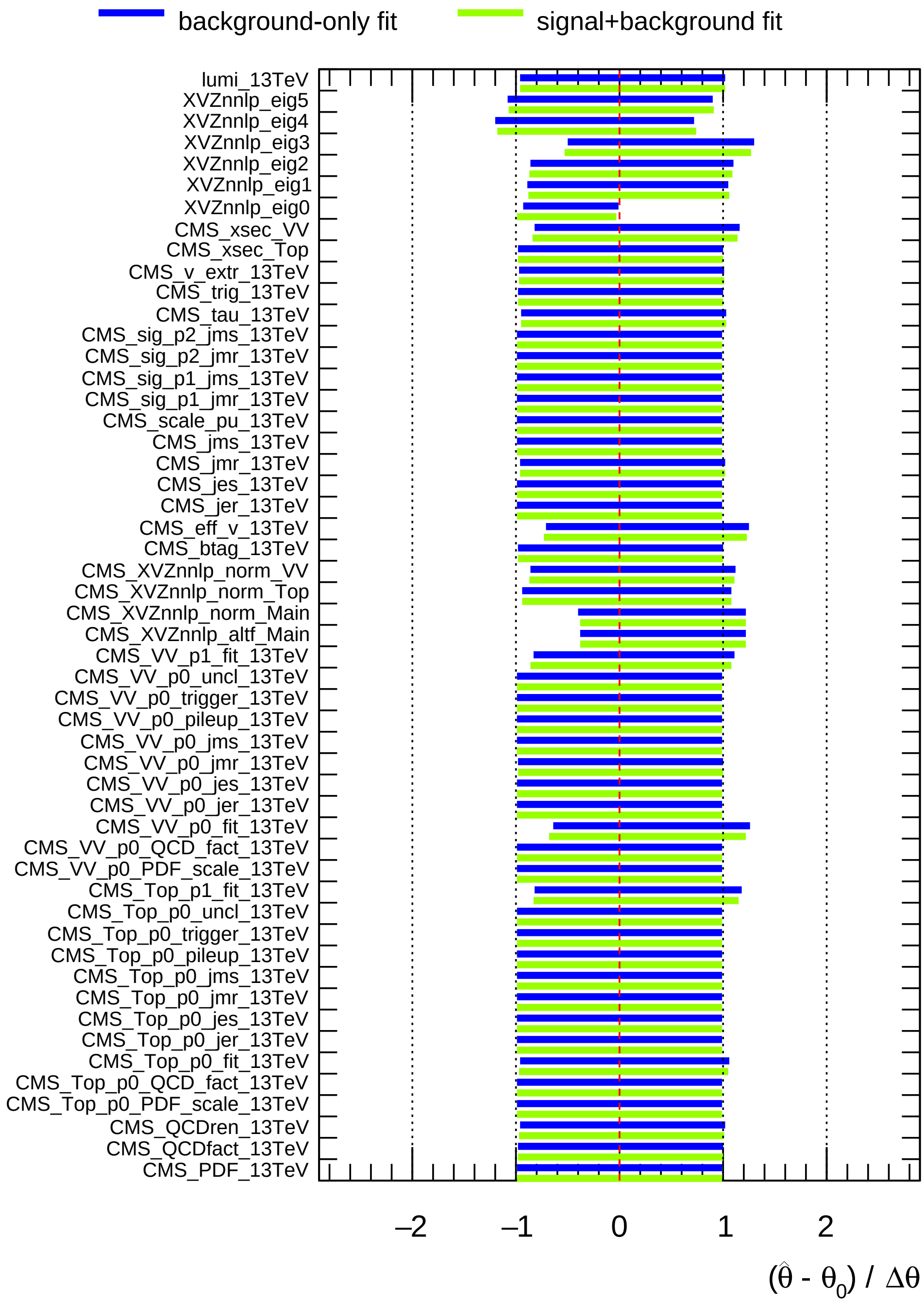


Figure 4.67: Nuisance pulls for the high-purity category, calculated under both the background-only (blue bars) and signal plus background hypotheses (green bars). A signal hypotesis of a spin-2 bulk graviton of mass 3 TeV is considered.

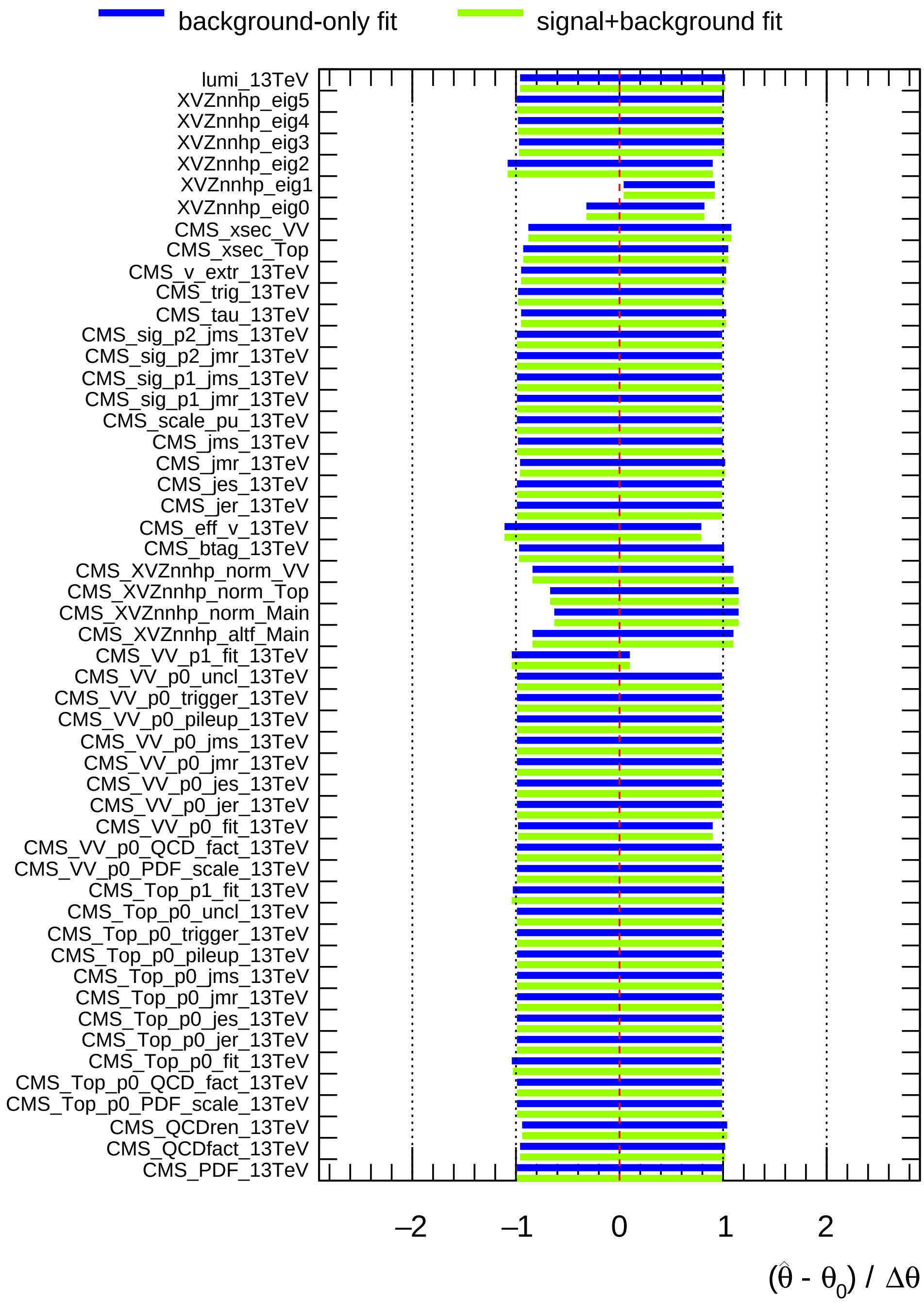
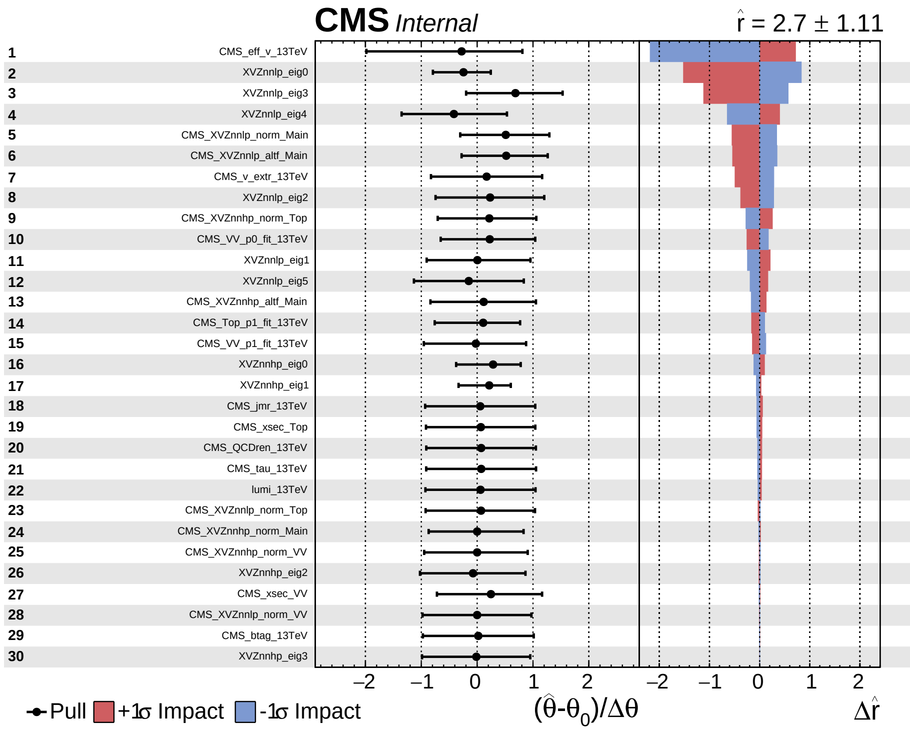
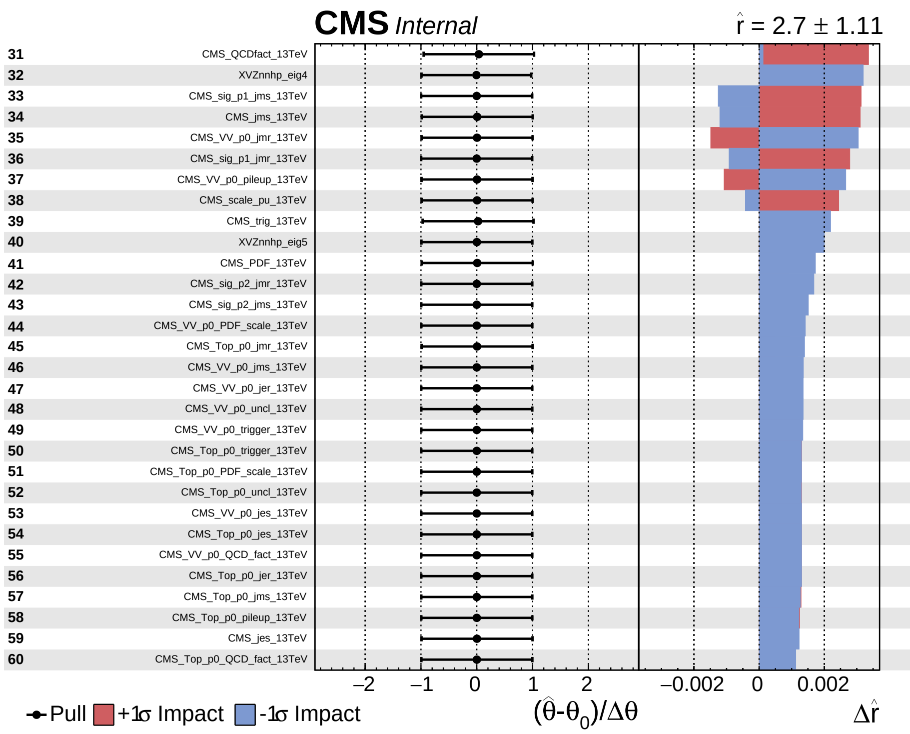


Figure 4.68: Impacts of the nuisance parameters on the signal strength estimation, for the combination of the low- and high-purity categories. A signal hypotesis of a spin-2 bulk graviton of mass 2.5 TeV is considered. θ_0 is the pre-fit value of the nuisance parameter taken into account; $\hat{\theta}$ is the value of the nuisance parameter after the maximum likelihood fit; $\Delta\hat{r}$ represents the impact, *i.e.* the shift induced in the parameter of interest (in this case, r, the cross-section times branching fraction, describing the signal strength) as the θ parameter is fixed and brought to its +1σ or -1σ post-fit values, with all the other nuisance parameters profiled as log-normal.





4.7.2.2 Results: expected and observed limits

The observed upper limit on the resonance cross-section times branching fraction $\sigma\mathscr{B}(X \rightarrow V_{\text{had}}Z_{\text{inv}})$, as well as the expected limit and its relative 68% and 95% uncertainty bands, are reported as a function of the resonance mass. The limits are obtained by considering separately a spin-2 bulk graviton and a spin-1 (W') heavy resonances in the narrow-width approximation. For the spin-2 case (fig. 4.69), data are compared to theoretical predictions on $\sigma\mathscr{B}(G \rightarrow Z_{\text{had}}Z_{\text{inv}})$, obtained by imposing a curvature parameter of the fifth extra-dimension $\tilde{k} = 0.5$ (red curve) and $\tilde{k} = 1.0$ (blue curve). In case of spin-1 hypotesis (fig.4.70), HVT model A (red curve) and model B (blue curve) theoretical predictions on $\sigma\mathscr{B}(W' \rightarrow W_{\text{had}}Z_{\text{inv}})$ are reported.
Given the fact that the background prediction is performed in a transverse mass range $950 < m_{VZ}^{T} < 4750$ GeV of the resonance, and given that the higher the nominal mass of the resonance, the more the Crystal Ball functions, parametrizing the m_{VZ}^{T} distributions of both spin-1 and spin-2 signals, tend to have low-mass tails (sec. 4.5.4.1), a safe conservative criterion is to set limits in the resonance mass range 1 TeV– 4 TeV.
No significant excess is observed in data with respect to the background-only hypotesis, neither in the low-purity, nor in the high-purity category. As it can be inferred by fig. 4.69-4.70, low-purity category has a larger sensitivity to both spin-1 and spin-2 signals in the high mass region, whilst high-purity category is more sensitive at low masses. This reflects the different signal efficiencies of the two categories, as discussed in sec. 4.3.12 (fig. 4.3.12.3). By combining the two categories, the best exclusion limits can be determined. Upper limits on the $\sigma\mathscr{B}(X \rightarrow V_{\text{had}}Z_{\text{inv}})$ of heavy spin-2 and spin-1 narrow resonances are set in the range 0.5 – 40 fb and in the range 0.9 – 63 fb respectively. A spin-2 bulk-graviton, once assumed a curvature parameter $\tilde{k} = 1.0$, is excluded up to 1.14 TeV. A spin-1 W', predicted by the model A scenario ($g_V = 1$), is excluded up to a mass of 3.11 TeV. A spin-1 W', predicted by the model B scenario ($g_V = 3$), is excluded up to a mass of 3.41 TeV.

4.7.2.3 Results: local p-value scan

Scans of the local significance (left plots) and of the local p-values (right plots), as a function of the resonance mass, are presented in fig. 4.71 (spin-2 signal) and in fig. 4.72 (spin-1 signal). No significant deviation is observed with regards to the background-only hypotesis. The maximum deviation is observed in the low-purity category, around 1.3 and 2.5 TeV, and it amounts to $\sim 2\sigma$. For the combination of the categories, data are compatible with the background-only hypotesis within 1σ in the whole mass spectrum.

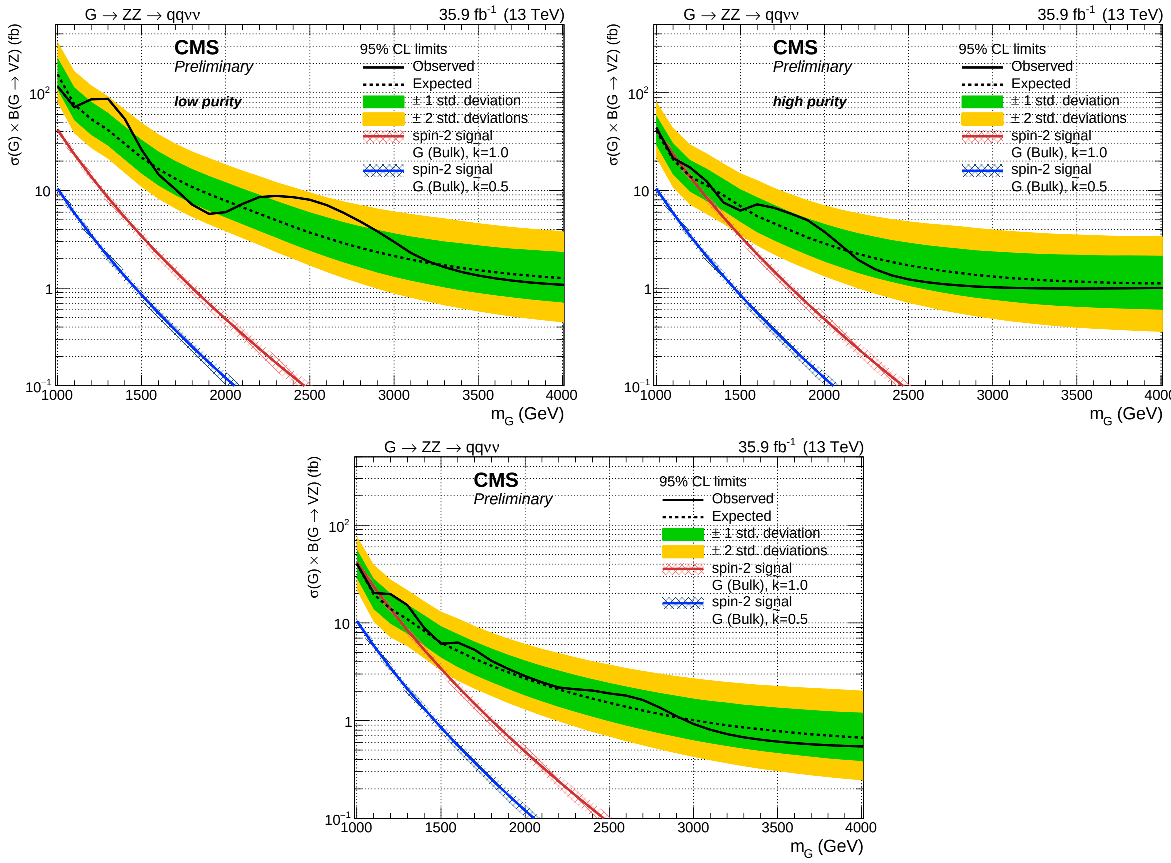


Figure 4.69: Top: observed and expected (with $\pm1(2)\sigma$ band) 95% C.L. upper limit on $\sigma\mathscr{B}(G \rightarrow Z_{\text{had}}Z_{\text{inv}})$ for a spin-2 (bulk graviton) signal, for low-purity (left) and high-purity (right) categories, including all statistical and systematics uncertainties. Background predictions are extracted with the α method. Bottom: observed and expected (with $\pm1(2)\sigma$ band) 95% C.L. upper limit on $\sigma\mathscr{B}(G \rightarrow Z_{\text{had}}Z_{\text{inv}})$ for a spin-2 (bulk graviton) signal, combining the two purity categories.

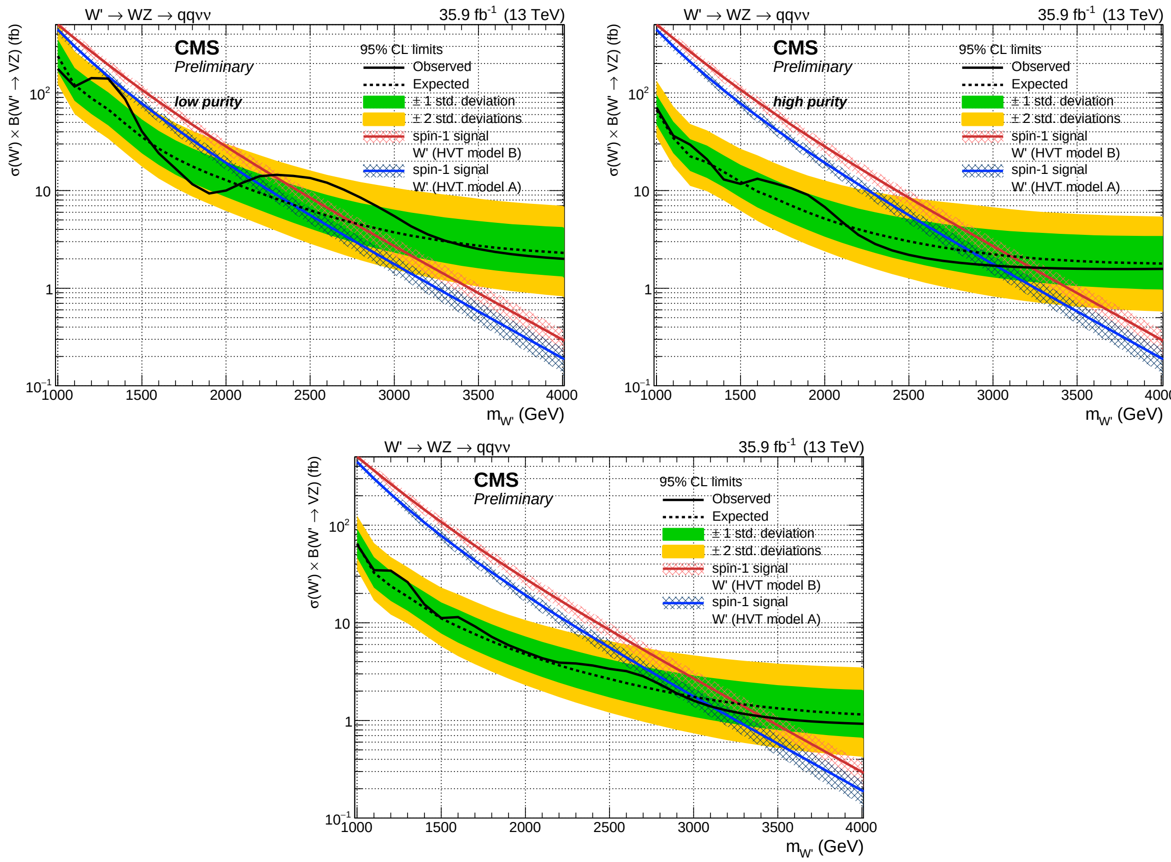


Figure 4.70: Top: observed and expected (with $\pm 1(2)\sigma$ band) 95% C.L. upper limit on $\sigma\mathscr{B}(W' \to W_{\mathrm{had}}Z_{\mathrm{inv}})$ for a spin-1 (HVT) signal, for low-purity (left) and high-purity (right) categories, including all statistical and systematics uncertainties. Background predictions are extracted with the α method. Bottom: observed and expected (with $\pm 1(2)\sigma$ band) 95% C.L. upper limit on $\sigma\mathscr{B}(W' \to W_{\mathrm{had}}Z_{\mathrm{inv}})$ for a spin-1 (HVT) signal, combining the two purity categories.

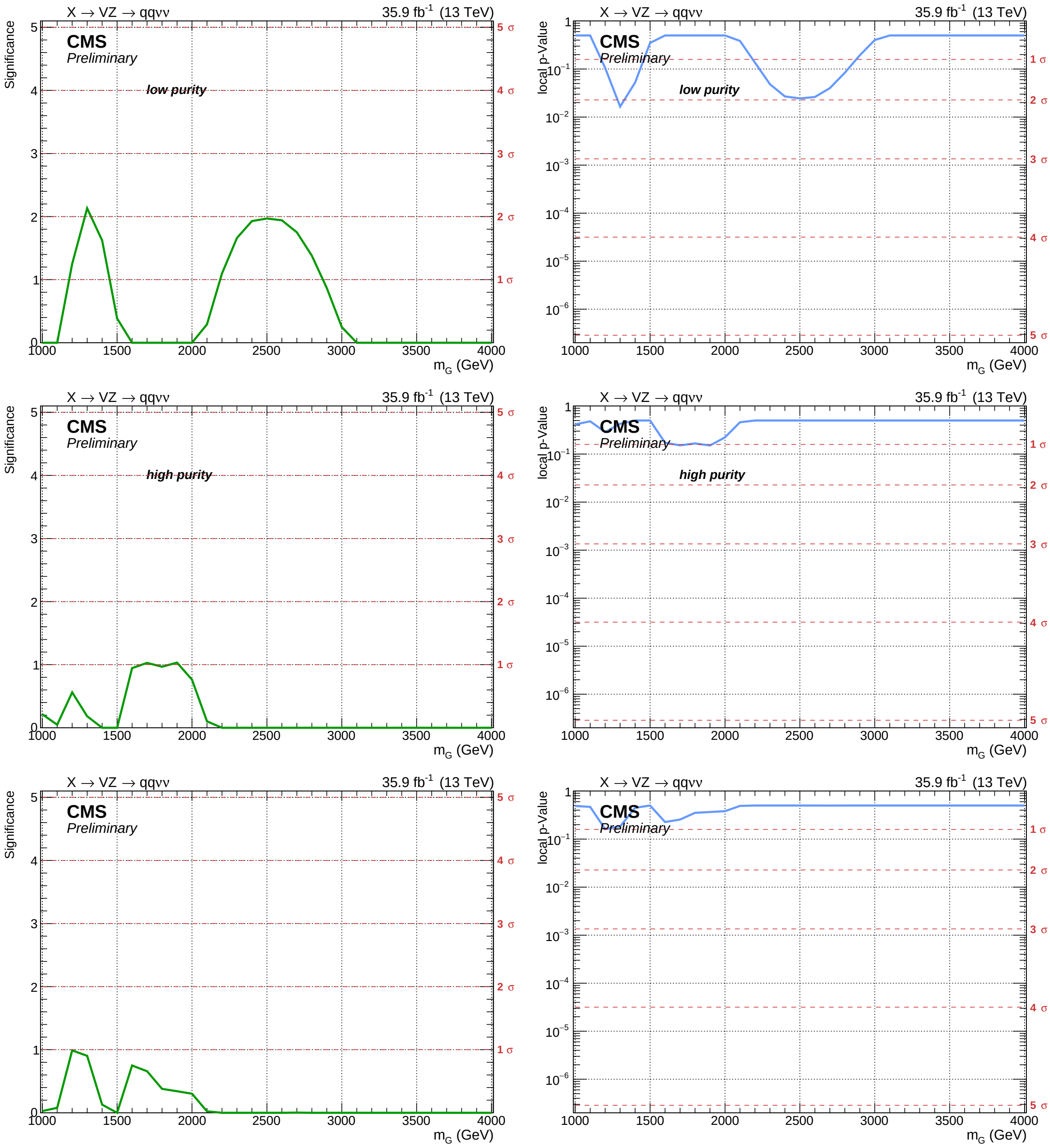


Figure 4.71: Local significances (left plots) and local p-values (right plots) as a function of the resonance mass, for a spin-2 bulk graviton hypotesis, in the low- (top), high-purity categories (center), and in the combination of the categories (bottom).

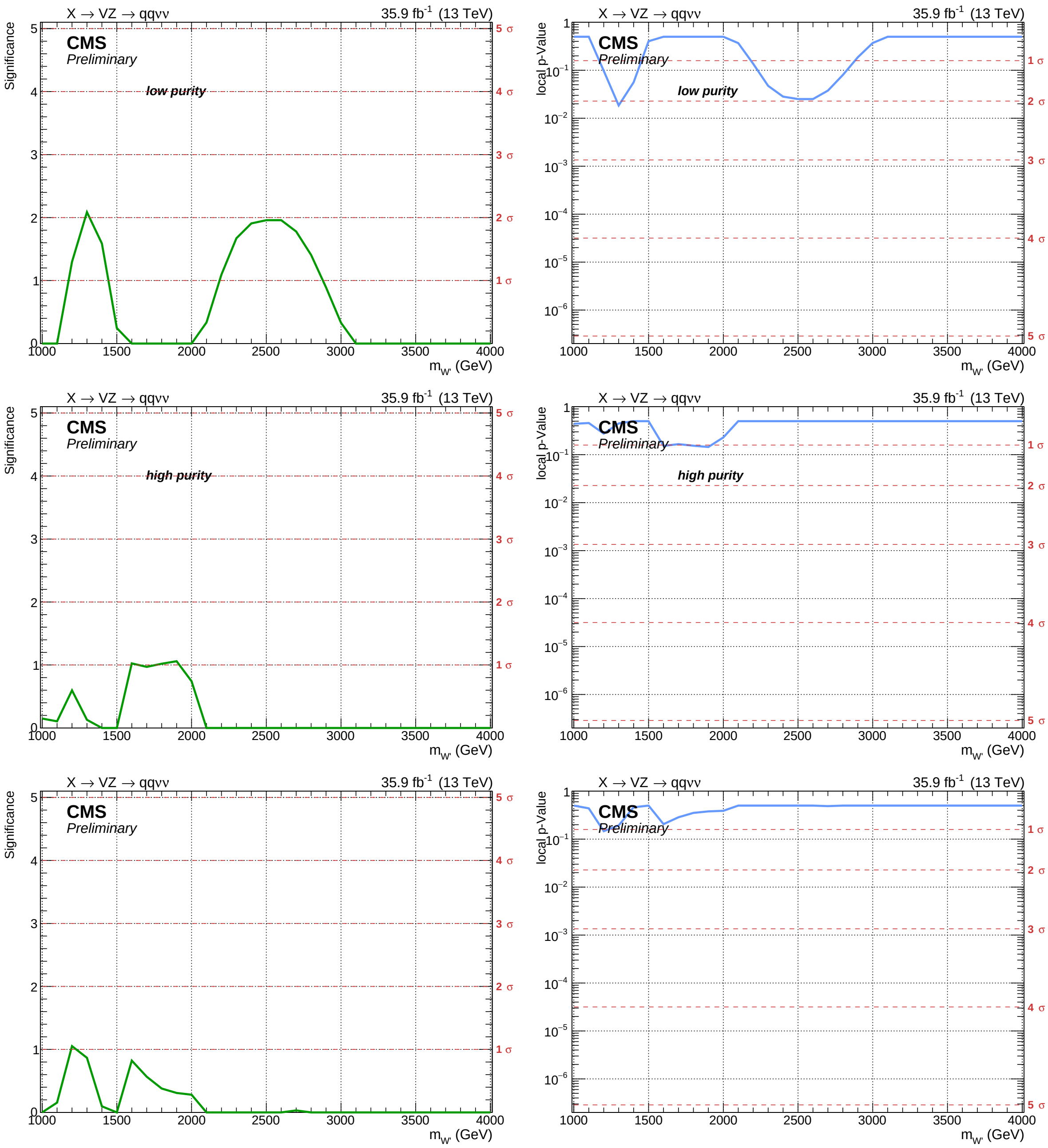


Figure 4.72: Local significances (left plots) and local p-values (right plots) as a function of the resonance mass, for a spin-1 W' hypotesis, in the low- (top), high-purity categories (center), and in the combination of the categories (bottom).

4.7.3 Interpretation of the results in the HVT model

For the HVT signal models, upper limits on the cross-section times branching fraction can be interpreted in the parameter space of the model (sec. 2.2), $\left(g_V c_H, g^2 c_F/g_V\right)$, where c_H describes the coupling of the heavy triplet to SM bosons, c_F the coupling of the triplet to SM fermions, g_V is the strength of the interaction, and g is the weak gauge coupling (sec. 2.2.1).
The benchmark model A is realized when $\left(g_V = 1, c_H = -0.556, c_F = -1.316\right)$; benchmark model B scenario is realized when $\left(g_V = 3, c_H = 0.976, c_F = 1.024\right)$ [14].
This search is sensitive to the charged components of the vector triplet, namely to $(W^{+'}, W^{-'})$. The excluded parameter space is shown in fig. 4.73. Since in the benchmark model A and model B all parameters are fixed, they are represented as a blue and a red marker respectively. The coloured curves represent the contours of the parameter space excluded by the observations in data, by considering a signal hypotesis of mass 1.5 TeV (in orange), 2 TeV (in green), 3 TeV (in violet). Currently, upper limits suggest an exclusion up to 3 TeV. The shaded gray area indicates the parameter space where the narrow width approximation fails; namely, the resonance intrinsic width becomes comparable to the experimental resolution, that amounts to 6% in this analysis (sec. 4.5.4.1).

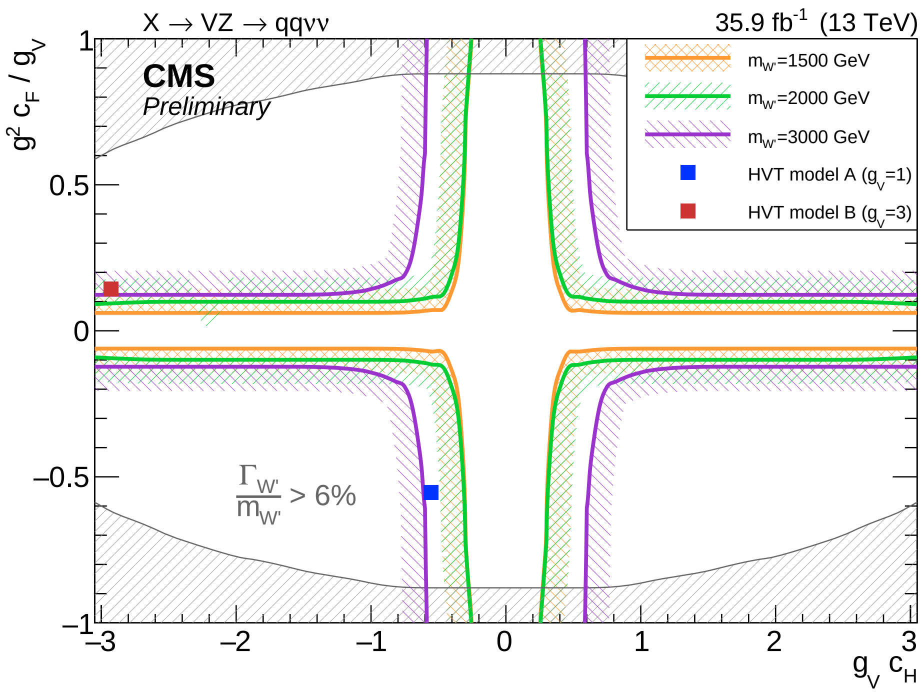


Figure 4.73: Exclusion limits on the parameter space of the HVT model. Coloured curves represent the contours of the parametric region excluded by observations in data, considering a spin-1 W' resonance of mass 1.5 TeV (in orange), 2 TeV (in green), 3 TeV (in violet). Benchmark model A and model B are represented as blue and red markers. The shaded gray area indicates the parameter space where the narrow width approximation fails.

Chapter 5

Conclusions

This thesis presented a search for heavy resonances with masses between 1 TeV and 4 TeV, decaying into a pair of vector bosons, predicted by beyond standard model theories. The data produced by LHC proton-proton collisions, at a center-of-mass energy $\sqrt{s} = 13$ TeV during the 2016 operations, and collected by the CMS experiment, corresponding to an integrated luminosity of 35.9 fb^{-1}, are analyzed. The probed final state includes the invisible decay modes of one Z boson, reconstructed as a large amount of missing transverse momentum, and the hadronic decay of the other vector boson (Z, W), reconstructed as a large-cone jet. The collected events are divided into two purity categories, based on the substructure of the hadronically decaying V boson. No significant excesses over the expected background are observed in the entire mass range probed by the analysis. Depending on the resonance mass, 95% C.L. upper limits on the cross-section of heavy spin-1 and spin-2 narrow resonances, multiplied by the branching fraction of the resonance decaying into Z and a W boson for a spin-1 signal, and into a pair of Z bosons for spin-2, are set in the range 0.9 – 63 fb and in the range 0.5 – 40 fb respectively. A W' hypotesis is excluded up to 3.11 TeV, in the context of the Heavy Vector Triplet model A scenario, and up to 3.41 TeV, considering the model B scenario. A bulk graviton hypotesis, given the curvature parameter $\tilde{k} = 1.0$, is excluded up to 1.14 TeV.

This is the first search for $VZ \to q\bar{q}\nu\bar{\nu}$ performed by the CMS Collaboration at $\sqrt{s} = 13$ TeV. This analysis is part of a set of searches for heavy resonances decaying into dibosons. The future perspectives of the analysis consist both in the combination of this final state with other diboson searches sharing the same treatment of one boson hadronic decay (namely, the same definition of the sidebands and signal regions), and in the combination of the 2016 data with the newly collected 2017 data. The luminosity planned to be delivered by the LHC collider in 2017 is comparable to what was collected in 2016 (~ 40 fb^{-1}). By doubling the statistics, marginal improvements are foreseen; hence, a larger enhancement can be achieved by decreasing the impacts of the systematic uncertainties. New jet substructure techniques are currently being tested, in order to improve the jet mass resolution (recursive soft drop), suppress the pile-up contribution (PUPPI associated to SoftKiller algorithm [107]), exploit the jet substructure and tag the nature of a large-cone jet (originating from W, Z, Higgs boson or top quark) with machine learning techniques. Preliminary results on these new methods seem to be promising.

Bibliography

[1] M. Baak and R. Kogler, *The global electroweak Standard Model fit after the Higgs discovery*, in *Proceedings, 48th Rencontres de Moriond on Electroweak Interactions and Unified Theories: La Thuile, Italy, March 2-9, 2013*, pp. 349–358, 2013, 1306.0571, https://inspirehep.net/record/1236809/files/arXiv:1306.0571.pdf.

[2] ATLAS Collaboration, ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012) 1, [1207.7214].

[3] CMS Collaboration, CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Phys. Lett. B* **716** (2012) 30, [1207.7235].

[4] CMS Collaboration, CMS Collaboration, *Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s}$ = 7 and 8 TeV*, *JHEP* **06** (2013) 081, [1303.4571].

[5] ATLAS Collaboration, ATLAS Collaboration, *Evidence for the spin-0 nature of the Higgs boson using ATLAS data*, *Phys. Lett. B* **726** (2013) 120, [1307.1432].

[6] CMS Collaboration, CMS Collaboration, *Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV*, *Eur. Phys. J. C* **75** (2015) 212, [1412.8662].

[7] ATLAS Collaboration, ATLAS Collaboration, *Measurement of the Higgs boson mass from the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4\ell$ channels in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector*, *Phys. Rev. D* **90** (2014) 052004, [1406.3827].

[8] CMS and ATLAS Collaborations, "Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments." 2015.

[9] G. Degrassi, S. Di Vita, J. Elias-Miro, J. R. Espinosa, G. F. Giudice, G. Isidori et al., *Higgs mass and vacuum stability in the Standard Model at NNLO*, *JHEP* **08** (2012) 098, [1205.6497].

[10] S. P. Martin, *A Supersymmetry primer*, hep-ph/9709356.

[11] C. Csaki, *The Minimal supersymmetric standard model (MSSM)*, *Mod. Phys. Lett.* **A11** (1996) 599, [hep-ph/9606414].

[12] D. J. Castano, E. J. Piard and P. Ramond, *Renormalization group study of the Standard Model and its extensions. 2. The Minimal supersymmetric Standard Model*, *Phys. Rev.* **D49** (1994) 4882–4901, [`hep-ph/9308335`].

[13] H. E. Haber, *Introductory low-energy supersymmetry*, in *Proceedings, Theoretical Advanced Study Institute (TASI 92): From Black Holes and Strings to Particles: Boulder, USA, June 1-26, 1992*, pp. 589–686, 1993, `hep-ph/9306207`.

[14] D. Pappadopulo, A. Thamm, R. Torre and A. Wulzer, *Heavy vector triplets: bridging theory and data*, *Journal of High Energy Physics* **2014** (2014) 1–50.

[15] V. D. Barger, W.-Y. Keung and E. Ma, *A Gauge Model With Light W and Z Bosons*, *Phys. Rev.* **D22** (1980) 727.

[16] C. Grojean, E. Salvioni and R. Torre, *A weakly constrained W' at the early LHC*, *JHEP* **07** (2011) 002, [`1103.2761`].

[17] R. Contino, D. Pappadopulo, D. Marzocca and R. Rattazzi, *On the effect of resonances in composite higgs phenomenology*, *Journal of High Energy Physics* **2011** (2011) 1–50.

[18] B. Bellazzini, C. Csáki and J. Serra, *Composite Higgses*, *Eur. Phys. J.* **C74** (2014) 2766, [`1401.2457`].

[19] G. F. Giudice, C. Grojean, A. Pomarol and R. Rattazzi, *The Strongly-Interacting Light Higgs*, *JHEP* **06** (2007) 045, [`hep-ph/0703164`].

[20] ATLAS Collaboration, G. Aad et al., *Evidence for Electroweak Production of $W^{\pm}W^{\pm}jj$ in pp Collisions at $\sqrt{s} = 8$ TeV with the ATLAS Detector*, *Phys. Rev. Lett.* **113** (2014) 141803, [`1405.6241`].

[21] CMS Collaboration, V. Khachatryan et al., *Study of vector boson scattering and search for new physics in events with two same-sign leptons and two jets*, *Phys. Rev. Lett.* **114** (2015) 051801, [`1410.6315`].

[22] A. Wulzer, *An Equivalent Gauge and the Equivalence Theorem*, *Nucl. Phys.* **B885** (2014) 97–126, [`1309.6055`].

[23] CMS Collaboration, A. M. Sirunyan et al., *Search for massive resonances decaying into WW, WZ, ZZ, qW, and qZ with dijet final states at sqrt(s) = 13 TeV*, `1708.05379`.

[24] CMS Collaboration, CMS Collaboration, *Search for massive resonances decaying into WW, WZ, ZZ, qW and qZ in the dijet final state at $\sqrt{s} = 13$ TeV*, CMS Physics Analysis Summary CMS-PAS-B2G-17-001, CERN, Geneva, 2017. http://cds.cern.ch/record/2256663.

[25] CMS Collaboration, CMS Collaboration, *Search for heavy resonances decaying into a vector boson and a Higgs boson in hadronic final states with 2016 data*, .

[26] CMS Collaboration, CMS Collaboration, *Search for heavy resonances decaying into a vector boson and a Higgs boson in hadronic final states with 2016 data*, CMS Physics Analysis Summary CMS-PAS-B2G-17-002, CERN, Geneva, 2017. http://cds.cern.ch/record/2256742.

[27] CMS Collaboration, CMS Collaboration, *Search for heavy resonances decaying into a Z boson and a W boson in the $\ell^+\ell^-\mathrm{q\bar{q}}$ final state*, CMS Physics Analysis Summary CMS-PAS-B2G-16-022, CERN, Geneva, 2017. http://cds.cern.ch/record/2242955.

[28] CMS Collaboration, V. Khachatryan et al., *Search for heavy resonances decaying into a vector boson and a Higgs boson in final states with charged leptons, neutrinos, and b quarks*, *Phys. Lett.* **B768** (2017) 137–162, [1610.08066].

[29] CMS Collaboration, CMS Collaboration, *Search for new resonances decaying to* WW/WZ $\to \ell\nu$qq, CMS Physics Analysis Summary CMS-PAS-B2G-16-020, CERN, Geneva, 2016. http://cds.cern.ch/record/2205880.

[30] CMS Collaboration, CMS Collaboration, *Search for heavy resonances decaying into a Z boson and a vector boson in the $\nu\nu\, q\bar{q}$ final state*, CMS Physics Analysis Summary CMS-PAS-B2G-17-005, CERN, Geneva, 2017. http://cds.cern.ch/record/2273910.

[31] ATLAS Collaboration, M. Aaboud et al., *Search for diboson resonances with boson-tagged jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, 1708.04445.

[32] ATLAS Collaboration, ATLAS Collaboration, *Search for WW/WZ resonance production in $\ell\nu qq$ final states in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, ATLAS Conference Note ATLAS-CONF-2017-051, CERN, Geneva, Jul, 2017. http://cds.cern.ch/record/2273867.

[33] ATLAS Collaboration, ATLAS Collaboration, *Searches for heavy ZZ and ZW resonances in the llqq and vvqq final states in pp collisions at sqrt(s) = 13 TeV with the ATLAS detector*, ATLAS Conference Note ATLAS-CONF-2016-082, CERN, Geneva, Aug, 2016. https://cds.cern.ch/record/2206275.

[34] ATLAS Collaboration, M. Aaboud et al., *Search for heavy resonances decaying to a W or Z boson and a Higgs boson in the $q\bar{q}^{(\prime)}b\bar{b}$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, 1707.06958.

[35] ATLAS Collaboration, ATLAS Collaboration, *Search for heavy resonances decaying to a W or Z boson and a Higgs boson in final states with leptons and b-jets in 36.1 fb^{-1} of pp collision data at $\sqrt{s} = 13$ TeV with the ATLAS detector*, ATLAS Conference Note ATLAS-CONF-2017-055, CERN, Geneva, Jul, 2017. http://cds.cern.ch/record/2273871.

[36] L. Randall and R. Sundrum, *A Large mass hierarchy from a small extra dimension*, *Phys. Rev. Lett.* **83** (1999) 3370–3373, [hep-ph/9905221].

[37] L. Randall and R. Sundrum, *An Alternative to compactification*, *Phys. Rev. Lett.* **83** (1999) 4690–4693, [hep-th/9906064].

[38] K. Agashe, H. Davoudiasl, G. Perez and A. Soni, *Warped Gravitons at the LHC and Beyond*, *Phys. Rev.* **D76** (2007) 036006, [hep-ph/0701186].

[39] A. L. Fitzpatrick, J. Kaplan, L. Randall and L.-T. Wang, *Searching for the Kaluza-Klein Graviton in Bulk RS Models*, *JHEP* **09** (2007) 013, [hep-ph/0701150].

[40] H. Davoudiasl, J. L. Hewett and T. G. Rizzo, *Phenomenology of the Randall-Sundrum Gauge Hierarchy Model*, *Phys. Rev. Lett.* **84** (2000) 2080, [hep-ph/9909255].

[41] T. Gherghetta and A. Pomarol, *Bulk fields and supersymmetry in a slice of AdS*, *Nucl. Phys.* **B586** (2000) 141–162, [hep-ph/0003129].

[42] H. Davoudiasl, J. L. Hewett and T. G. Rizzo, *Experimental probes of localized gravity: On and off the wall*, *Phys. Rev.* **D63** (2001) 075004, [hep-ph/0006041].

[43] A. Oliveira, *Gravity particles from Warped Extra Dimensions, predictions for LHC*, 1404.0102.

[44] CMS Collaboration, CMS Collaboration, *Search for diboson resonances in the $2l2\nu$ final state*, CMS Physics Analysis Summary CMS-PAS-B2G-16-023, CERN, Geneva, 2017. http://cds.cern.ch/record/2264700.

[45] CMS Collaboration, CMS Collaboration, *Search for heavy resonances decaying to a pair of Higgs bosons in the four b quark final state in proton-proton collisions at $\sqrt{s} = 13$ TeV*, CMS Physics Analysis Summary CMS-PAS-B2G-16-026, CERN, Geneva, 2017. http://cds.cern.ch/record/2264684.

[46] L. Evans and P. Bryant, *LHC Machine*, *JINST* **3** (2008) S08001.

[47] CMS Collaboration, CMS Collaboration, *The CMS experiment at the CERN LHC*, *JINST* **3** (2008) S08004.

[48] CMS Collaboration, CMS Collaboration, *The CMS tracker system project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997, https://cds.cern.ch/record/368412.

[49] CMS Collaboration, CMS Collaboration, *Description and performance of track and primary-vertex reconstruction with the CMS tracker*, *JINST* **9** (2014) P10009, [1405.6569].

[50] CMS Collaboration, CMS Collaboration, *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997, https://cds.cern.ch/record/349375.

[51] CMS Collaboration, CMS Collaboration, *The CMS hadron calorimeter project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997, https://cds.cern.ch/record/357153.

[52] CMS Collaboration, CMS Collaboration, *The CMS muon project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997, https://cds.cern.ch/record/343814.

[53] CMS Collaboration, CMS Collaboration, *CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project*. Technical Design Report CMS. CERN, Geneva, 2002, https://cds.cern.ch/record/578006.

[54] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, *Nucl. Instrum. Meth.* **A389** (1997) 81–86.

[55] CMS Collaboration, A. M. Sirunyan et al., *Particle-flow reconstruction and global event description with the cms detector*, *JINST* **12** (2017) P10003, [1706.04965].

[56] R. Fruhwirth, *Application of Kalman filtering to track and vertex fitting*, *Nucl. Instrum. Meth.* **A262** (1987) 444–450.

[57] K. Rose, *Deterministic annealing for clustering, compression, classification, regression, and related optimization problems*, in *Proceedings of the IEEE*, vol. 86, pp. 2210 – 2239, 12, 1998.

[58] R. Fruhwirth, W. Waltenberger and P. Vanlaer, *Adaptive vertex fitting*, *J. Phys.* **G34** (2007) N343.

[59] CMS Collaboration, CMS Collaboration, *Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in pp Collisions at $\sqrt{s} = 7$ TeV*, *JINST* **8** (2013) P09009, [1306.2016].

[60] W. Adam, R. Frühwirth, A. Strandlie and T. Todorov, *RESEARCH NOTE FROM COLLABORATION: Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC*, *Journal of Physics G Nuclear Physics* **31** (Sept., 2005) N9–N20, [physics/0306087].

[61] CMS Collaboration, CMS Collaboration, *Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at s = 8 TeV*, *JINST* **10** (2015) P06005, [1502.02701].

[62] CMS Collaboration, CMS Collaboration, *Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at sqrt(s) = 8 TeV*, *JINST* **10** (2015) P08010, [1502.02702].

[63] CMS Collaboration, CMS Collaboration, *Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV*, *JINST* **7** (2012) P10002, [1206.4071].

[64] CMS Collaboration, CMS Collaboration, *Performance of muon reconstruction including Alignment Position Errors for 2016 Collision Data*, CMS Performance Note CMS-DP-2016-067, CERN, Geneva, Nov, 2016. http://cds.cern.ch/record/2229697.

[65] CMS Collaboration, CMS Collaboration, *Pileup Removal Algorithms*, Tech. Rep. CMS-PAS-JME-14-001, CERN, Geneva, 2014. https://cds.cern.ch/record/1751454.

[66] M. Cacciari, G. P. Salam and G. Soyez, *FastJet user manual*, *Eur. Phys. J. C* **72** (2012) 1896, [1111.6097].

[67] M. Cacciari, G. P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, *JHEP* **04** (2008) 063, [0802.1189].

[68] CMS Collaboration, CMS Collaboration, *Determination of jet energy calibration and transverse momentum resolution in cms*, *Journal of Instrumentation* **6** (2011) P11002.

[69] CMS Collaboration, CMS Collaboration, *Jet energy scale and resolution performances with 13TeV data*, CMS Performance Note CMS-DP-2016-020, CERN, Geneva, Jun, 2016. http://cds.cern.ch/record/2160347.

[70] CMS Collaboration, CMS Collaboration, *Performance of tau-lepton reconstruction and identification in CMS*, *JINST* **7** (2012) P01001, [1109.6034].

[71] CMS Collaboration, CMS Collaboration, *Identification of b quark jets at the CMS Experiment in the LHC Run 2*, CMS Physics Analysis Summary CMS-PAS-BTV-15-001, CERN, Geneva, 2016. https://cds.cern.ch/record/2138504.

[72] CMS Collaboration, CMS Collaboration, *Identification of b-quark jets with the CMS experiment*, *JINST* **8** (2013) P04013, [1211.4462].

[73] CMS Collaboration, CMS Collaboration, *Performance of missing energy reconstruction in 13 TeV pp collision data using the CMS detector*, .

[74] CMS Collaboration, CMS Collaboration, *Missing transverse energy performance of the CMS detector*, *JINST* **6** (2011) P09001, [1106.5048].

[75] CMS Collaboration, CMS Collaboration, *Performance of Missing Transverse Momentum Reconstruction Algorithms in Proton-Proton Collisions at p s = 8 TeV with the CMS Detector*, .

[76] ATLAS Collaboration, ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08003.

[77] ALICE Collaboration, ALICE Collaboration, *The ALICE experiment at the CERN LHC*, *JINST* **3** (2008) S08002.

[78] LHCb Collaboration, LHCb Collaboration, *The LHCb Detector at the LHC*, *JINST* **3** (2008) S08005.

[79] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079, [1405.0301].

[80] Torbjorn Sjostrand and Stephen Mrenna and Peter Skands, *A brief introduction to pythia 8.1*, *Comput.Phys.Commun.* **178** (2008) 852–867, [0710.3820].

[81] CMS Collaboration, CMS Collaboration, *Event generator tunes obtained from underlying event and multiparton scattering measurements*, *Eur. Phys. J.* **C** (2016) 155, [1512.00815].

[82] S. Agostinelli, J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Arce et al., *Geant4a simulation toolkit*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** (2003) 250 – 303.

[83] CMS Collaboration, V. Khachatryan et al., *Identification techniques for highly boosted W bosons that decay into hadrons*, *JHEP* **12** (2014) 017, [1410.4227].

[84] Rikkert Frederix and Stefano Frixione, *Merging meets matching in mc@nlo*, *JHEP* **12** (2012) 061, [1209.6215].

[85] Emanuele Re, *Single-top wt-channel production matched with parton showers using the powheg method*, *Eur.Phys.J.* **C71** (2011) 1547, [1009.2450].

[86] John M. Campbell, R. Keith Ellis, Paolo Nason, Emanuele Re, *Top-pair production and decay at nlo matched with parton showers*, *JHEP* **04** (2015) 114, [1412.1828].

[87] S. Kallweit, J. M. Lindert, S. Pozzorini, M. Schönherr and P. Maierhöfer, *NLO QCD+EW automation and precise predictions for V+multijet production*, in *Proceedings, 50th Recontres de Moriond, QCD and high energy interactions*, pp. 121–124, 2015, 1505.05704, http://inspirehep.net/record/1372103/files/arXiv:1505.05704.pdf.

[88] CMS Collaboration Collaboration, *CMS Luminosity Measurements for the 2016 Data Taking Period*, CMS Physics Analysis Summary CMS-PAS-LUM-17-001, CERN, Geneva, 2017. https://cds.cern.ch/record/2257069.

[89] A. J. Larkoski, S. Marzani, G. Soyez and J. Thaler, *Soft Drop*, *JHEP* **05** (2014) 146, [1402.2657].

[90] Y. L. Dokshitzer, G. D. Leder, S. Moretti and B. R. Webber, *Better jet clustering algorithms*, *JHEP* **08** (1997) 001, [hep-ph/9707323].

[91] D. Bertolini, P. Harris, M. Low and N. Tran, *Pileup per particle identification*, *Journal of High Energy Physics* **2014** (2014) 59.

[92] J. Thaler and K. Van Tilburg, *Identifying boosted objects with n-subjettiness*, *Journal of High Energy Physics* **2011** (2011) 15.

[93] CMS Collaboration, *Performance of b-Tagging Algorithms in Proton Collisions at 13 TeV using the 2016 Data*, Physics Analysis Summary CMS-DP-2016-042, CERN, Geneva, Jul, 2016. https://cds.cern.ch/record/2202967.

[94] CMS Collaboration, V. Khachatryan et al., *Search for a higgs boson in the mass range from 145 to 1000 GeV decaying to a pair of W or Z bosons*, *JHEP* **10** (2015) 144, [1504.00936].

[95] M. Oreglia, *A Study of the Reactions $\psi' \to \gamma\gamma\psi$*, Ph.D. thesis, SLAC, 1980.

[96] T. Skwarnicki, *A study of the radiative CASCADE transitions between the Upsilon-Prime and Upsilon resonances*, Ph.D. thesis, Cracow, INP, 1986.

[97] M. Bahr, S. Gieseke, M. A. Gigg, D. Grellscheid, K. Hamilton, O. Latunde-Dada, S. Platzer, P. Richardson, M. H. Seymour, A. Sherstnev, J. Tully, B. R. Webber, *Herwig++ physics and manual*, *Eur.Phys.J.* **C58** (2008) 639–707, [0803.0883].

[98] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, *J. Phys.* **G43** (2016) 023001, [1510.03865].

[99] NNPDF Collaboration, R. D. Ball et al., *Parton distributions from high-precision collider data*, *Eur. Phys. J.* **C77** (2017) 663, [1706.00428].

[100] A. L. Read, *Presentation of search results: the CL_s technique*, *J. Phys.* **G28** (2002) 2693.

[101] T. Junk, *Confidence level computation for combining searches with small statistics*, *Nucl. Instrum. Meth.* **A434** (1999) 435, [9902006].

[102] The ATLAS Collaboration, The CMS Collaboration, The LHC Higgs Combination Group, *Procedure for the LHC Higgs boson search combination in Summer 2011*, Tech. Rep. CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, CERN, Geneva, Aug, 2011. https://cds.cern.ch/record/1379837.

[103] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *The European Physical Journal C* **71** (2011) .

[104] S. S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, *The Annals of Mathematical Statistics* **9** (03, 1938) 60–62.

[105] A. Wald, *Tests of statistical hypotheses concerning several parameters when the number of observations is large*, *Transactions of the American Mathematical Society* **54** (1943) 426–482.

[106] E. Gross and O. Vitells, *Trial factors for the look elsewhere effect in high energy physics*, *The European Physical Journal C* **70** (Nov, 2010) 525–530.

[107] M. Cacciari, G. P. Salam and G. Soyez, *SoftKiller, a particle-level pileup removal method*, *Eur. Phys. J.* **C75** (2015) 59, [1407.0408].