



Università degli studi di Padova
Dipartimento di Fisica e Astronomia

Tesi di Dottorato

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

Supervisor: Prof. Franco Simonetto
Candidate: Lisa Benato

Scuola di Dottorato di Ricerca, XXX ciclo

"I have no special talent. I am only passionately curious."
(A. Einstein)

Contents

1	Introduction	1
2	Theoretical motivation	3
3	The Large Hadron Collider and the CMS experiment	5
3.1	Background estimation technique	5
3.1.1	Background normalization	6
3.1.2	Background shape	9
3.1.3	Background prediction	12
3.1.4	Alpha method validation	13
3.1.5	Signal modeling	15
3.1.5.1	Signal parametrization	20
4	Conclusions	25

Abstract

Introduction

Theoretical motivation

The Large Hadron Collider and the CMS experiment

3.1 Background estimation technique

The goal of the analysis is to look for localised excesses in the m_{VZ}^T spectrum. The α -method is used in searches for heavy resonances since the LHC Run 1 [1], 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.

The α function is defined as the ratio of the two functions describing the simulated m_{VZ}^T shape in the SR and SB:

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

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:

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

Notice that in the above description does not include the definition of the SB and SR. Ideally, the best choice would be a variable such that the distribution of m_{VZ}^T in the Signal and Sideband regions are similar. In this analysis, the softdrop+PUPPI corrected jet mass m_j (see Section ??) is chosen as the control variable, and the cut values are those reported in Section ???. All the selections used in the α method background prediction are the same reported in Section ???. In order to respect the blinding policy of the diboson VV searches, the Higgs mass window is not used neither for the background estimation (it can be potentially contaminated by a signal) nor as signal region.

In a real case scenario, the background is not purely composed of one single process neither in the SR nor in the SB. The background composition is assumed to be dominated by one single process (V +jets) whose modeling in simulation is considered not to be trustworthy. Other subdominant backgrounds ($t\bar{t}$ and single- t production (Top), and VV) generally have smaller contributions (of the order of 5% for VV , and 9% for Top), and are considered quite well understood and modeled by MC generators.

EXPLAIN WHY Z+jet and W+jet are together

The shape and normalization of the diboson and Top production is taken from the simulation. The sub-dominant backgrounds are then subtracted from the V +jets contribution when fitting this template to data.

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.

3.1.1 Background normalization

The first step in the background prediction consists in a proper estimation of the background normalization. The three main backgrounds (V +jets, Top, and VV) are considered separately due to the different shape in the jet mass distribution. The three contributions are described with functional forms determined by fits on the simulated backgrounds. The number of expected events in the SR is extracted through the same equation:

$$N_{SR}^{data} = \left[N_{SB}^{data} - N_{SB}^{Top} - N_{SB}^{VV} \right] \times \left[\frac{N_{SR}^{Vjet}}{N_{SB}^{Vjet}} \right] + N_{SR}^{Top} + N_{SR}^{VV}$$

where in this case N are the number of events, and not functions.

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_j spectrum shows a peak, corresponding to the vector boson hadronic decay reconstruction, that is modelled as a gaussian. For the jet mass spectrum of the Top backgrounds, the peaks corresponding to the W and top mass can be modelled as gaussian functions, superimposed to a falling exponential background.

The functional forms chosen to build the jet mass templates are:

ErfExp: an “error function”, that consists of an exponential multiplied by an Erf:

$$F_{\text{ErfExp}}(x) = e^{ax} \cdot \frac{1 + \text{Erf}((x - b)/w)}{2}$$

ErfExpTail: an ErfExp modified to account for an additional term in the exponential:

$$F_{\text{ErfExpTail}}(x) = e^{-x/(a+b*x)} \cdot \frac{1 + \text{Erf}((x - o)/w)}{2}$$

ErfPow2: an Erf multiplied by a power law:

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

3.1 Background estimation technique

Gaus3: three gaussians:

$$F_{\text{Gaus3}}(x) = f_0 \cdot e^{2(x-a)^2/b} + f_1 \cdot e^{2(x-c)^2/d} + (1-f_0-f_1) \cdot e^{2(x-e)^2/g}$$

ExpGaus: an exponential plus one gaussian:

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

ErfExpGaus: an error function plus one gaussian:

$$F_{\text{ErfExpGaus}}(x) = f_0 \cdot F_{\text{ErfExp}}(x, a, b, c) + (1-f_0) \cdot e^{2(x-d)^2/e}$$

ErfExpGaus2: an error function plus two gaussians:

$$F_{\text{ExpGaus2}}(x) = f_0 \cdot F_{\text{ErfExp}}(x, a, b, c) + f_1 \cdot e^{2(x-d)^2/e} + (1-f_0-f_1) \cdot e^{2(x-f)^2/g}$$

55 The choice of the functions is channel-dependent, and it depends on the background shape and
56 the available statistics, and is summarized in Table 3.1.

category	V +jets	alt. V +jets	Top	VV
2ν , low-purity	ErfPow2	ExpGaus	ErfExpGaus2	ExpGaus
2ν , high-purity	ExpGaus	ErfExpGaus	ErfExpGaus2	ExpGaus

Table 3.1: Chosen functions to fit the jet mass distribution for each channel.

57 The following plots (Figure 3.3-3.3) show the mass fits to the jet mass in the different channels.
58 Table 3.2 summarizes the expected background yield in the signal region, that is in agreement with
59 observation in both the purity categories.

region	category	Expected	Stat.	Syst.	Alt. function	Observed
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

Table 3.2: Expected background yield in the SB ($30 < m_j < 65\text{GeV}$, $m_j > 135\text{GeV}$) and in the SR ($65 < m_j < 105\text{GeV}$) and relative uncertainties.

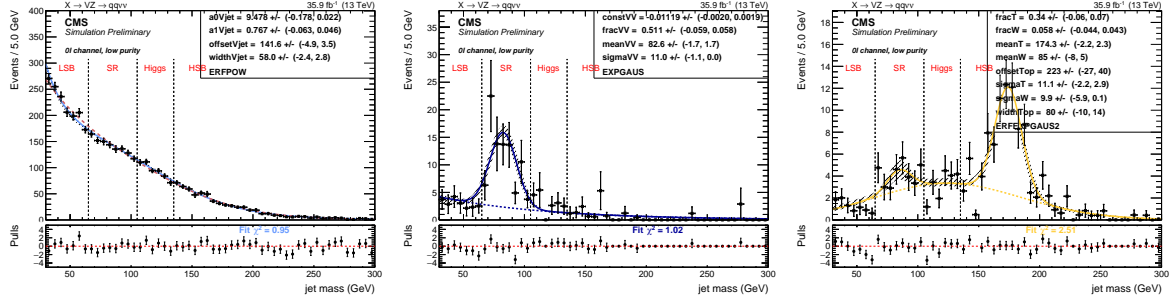


Figure 3.1: Fit to the simulated m_j in the low-purity category for the three backgrounds: V +jets (left), VV (center), Top (right). For the main background prediction, the alternative function is displayed with a dotted red line, superimposed to the main choice (continuous light blue curve).

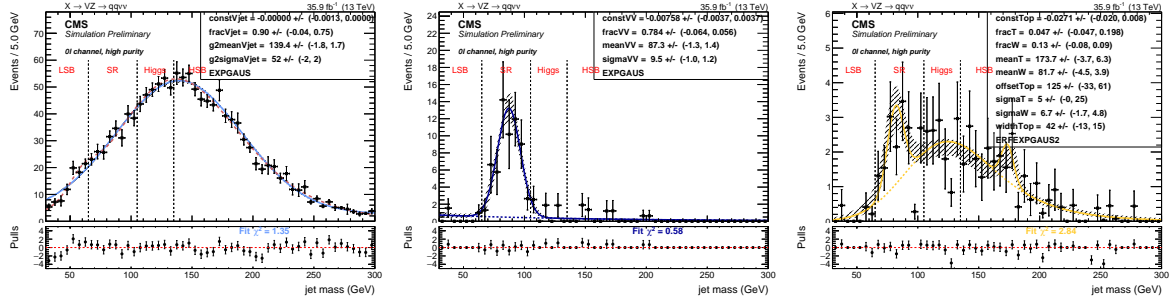


Figure 3.2: Fit to the simulated m_j in the high-purity category for the three backgrounds: V +jets (left), VV (center), Top (right). For the main background prediction, the alternative function is displayed with a dotted red line, superimposed to the main choice (continuous light blue curve).

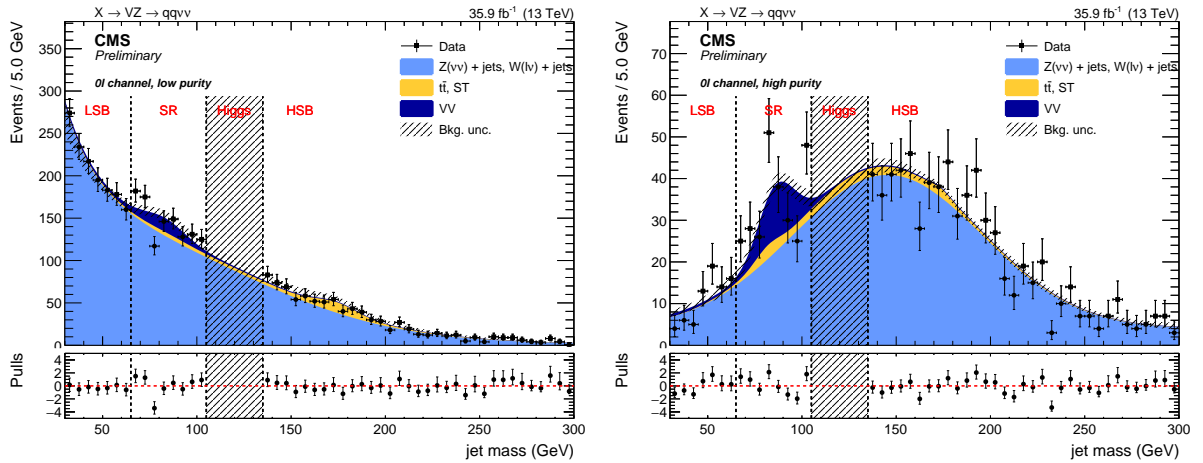


Figure 3.3: Fit to data m_j in the low- (left) and high-purity category (right).

3.1 Background estimation technique

3.1.2 Background shape

The mass of the resonance candidate (m_{VZ}^T) is parametrized separately for the V+jets ($N_{SR}^{Vjet}(m_{VZ}^T)$, $N_{SB}^{Vjet}(m_{VZ}^T)$), Top production ($N_{SR}^{Top}(m_{VZ}^T)$, $N_{SB}^{Top}(m_{VZ}^T)$), and dibosons ($N_{SR}^{VV}(m_{VZ}^T)$, $N_{SB}^{VV}(m_{VZ}^T)$). These functions are extracted fitting the simulated m_{VZ}^T spectra in SR and SB, respectively. The top and the diboson are normalized to luminosity. The V+jets functions are used to extract the α -function:

$$\alpha(m_{VZ}^T) = \frac{N_{SR}^{Vjet}(m_{VZ}^T)}{N_{SB}^{Vjet}(m_{VZ}^T)}$$

The main background is extracted through a fit to data in the SB, after subtracting the corresponding the 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 contribution in the SR is added to the main background estimation.

In formulas, the procedure used to extract the total background prediction is the following:

$$N_{SR}^{main}(m_{VZ}^T) = N_{SR}^{main}(m_{VZ}^T) \times \alpha(m_{VZ}^T)$$

$$N_{SR,SB}^{bkg}(m_{VZ}^T) = N_{SR,SB}^{main}(m_{VZ}^T) + N_{SR,SB}^{Top}(m_{VZ}^T) + N_{SR,SB}^{VV}(m_{VZ}^T)$$

$$N_{SR}^{data}(m_{VZ}^T) = \left[N_{SB}^{data}(m_{VZ}^T) - N_{SB}^{Top}(m_{VZ}^T) - N_{SB}^{VV}(m_{VZ}^T) \right] \times \left[\frac{N_{SR}^{Vjet}(m_{VZ}^T)}{N_{SB}^{Vjet}(m_X)} \right] + N_{SR}^{Top}(m_{VZ}^T) + N_{SR}^{VV}(m_{VZ}^T)$$

The functions used to parametrize the m_{VZ}^T distributions are:

ExpN: a product of two exponentials:

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

ExpTail: a modified exponential function with an additional parameter to model the exponential tails:

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

The functions chosen to parametrize the main background and extract the α -function are reported in Table 3.3 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. Table 3.3 reports both the main function and the alternative function.

category	Main bkg function	Main bkg alternative	diboson	top
2 ν , low-purity	ExpN	ExpTail	ExpTail	ExpTail
2 ν , high-purity	ExpTail	ExpN	ExpTail	ExpTail

Table 3.3: Main and alternative functions chosen to parametrize the main background contribution in the m_{VZ}^T distribution for each channel.

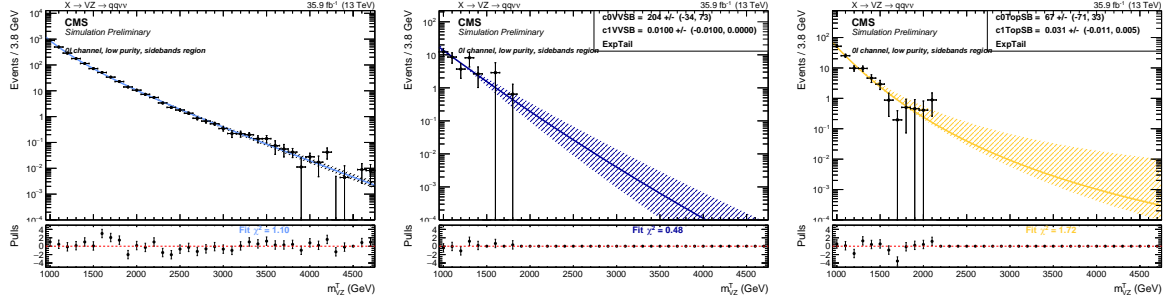


Figure 3.4: 2 neutrinos, low-purity channel. Fits to the simulated background components V +jets (left), VV (center), Top (right) in the sidebands (SB).

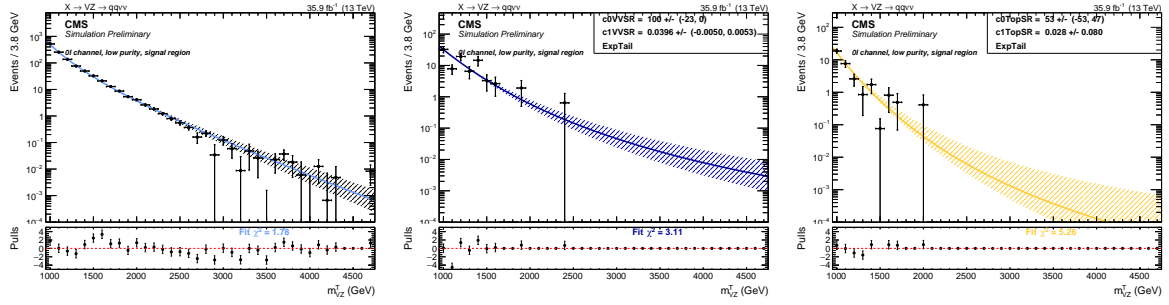


Figure 3.5: 2 neutrinos, low-purity channel. Fits to the simulated background components V +jets (left), VV (center), Top (right) in the signal region (SR).

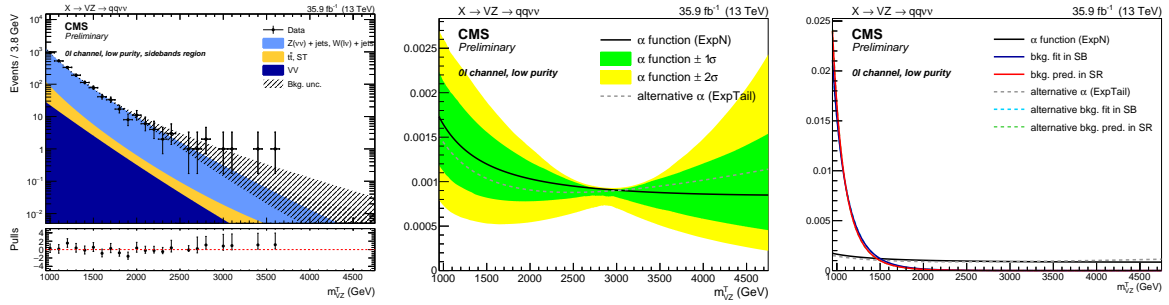


Figure 3.6: 2 neutrinos, low-purity channel. Fit to data in the SB (left), alpha function (center), and alpha function compared to the background shape in both SB and SR (right). The black line, with the corresponding 1σ (green) and 2σ (yellow) uncertainty bands, represents the α -function. The gray line is the alternative α -function. The blue and lines represent the estimated background in the SB and SR, respectively, with both the main (solid line) and alternative (dotted line) parametrization.

3.1 Background estimation technique

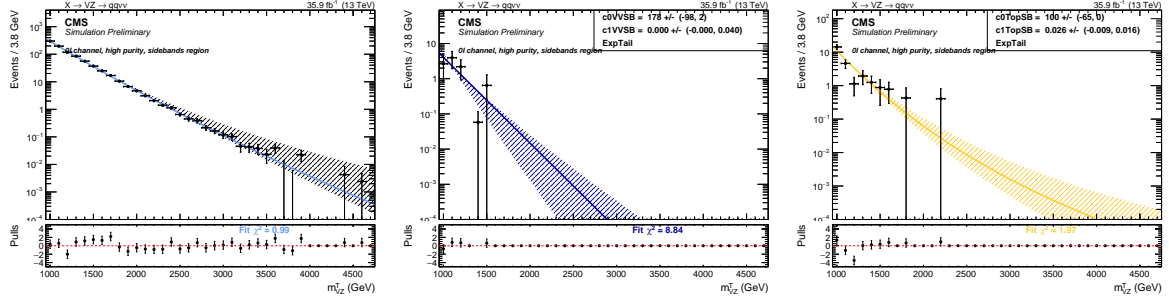


Figure 3.7: 2 neutrinos, high-purity channel. Fits to the simulated background components V +jets (left), VV (center), Top (right) in the sidebands (SB).

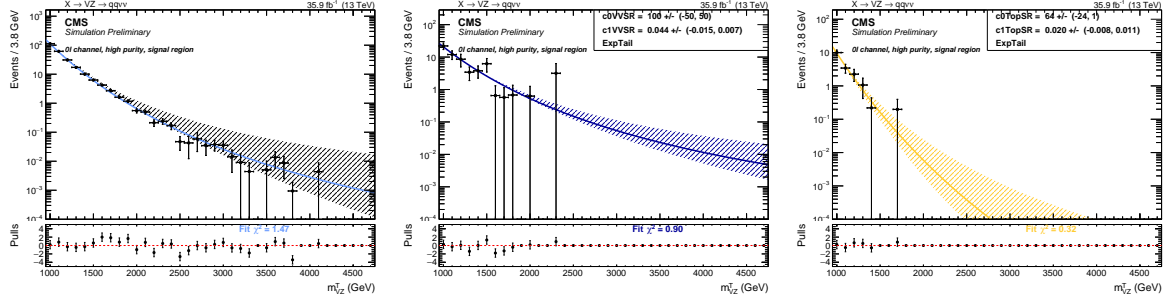


Figure 3.8: 2 neutrinos, high-purity channel. Fits to the simulated background components V +jets (left), VV (center), Top (right) in the signal region (SR).

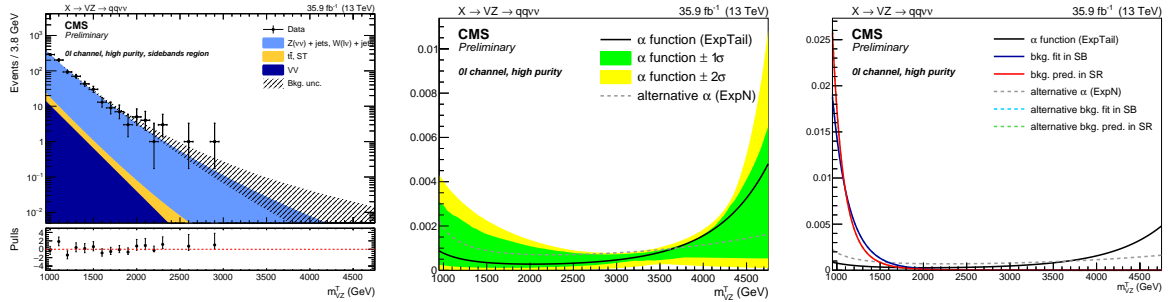


Figure 3.9: 2 neutrinos, high-purity channel. Fit to data in the SB (left), alpha function (center), and alpha function compared to the background shape in both SB and SR (right). The black line, with the corresponding 1σ (green) and 2σ (yellow) uncertainty bands, represents the α -function. The gray line is the alternative α -function. The blue and lines represent the estimated background in the SB and SR, respectively, with both the main (solid line) and alternative (dotted line) parametrization.

3.1.3 Background prediction

Fig. 3.10 summarizes the final background predictions as a function of the search variable, the transverse mass. Data and predictions are in agreement in both channels.

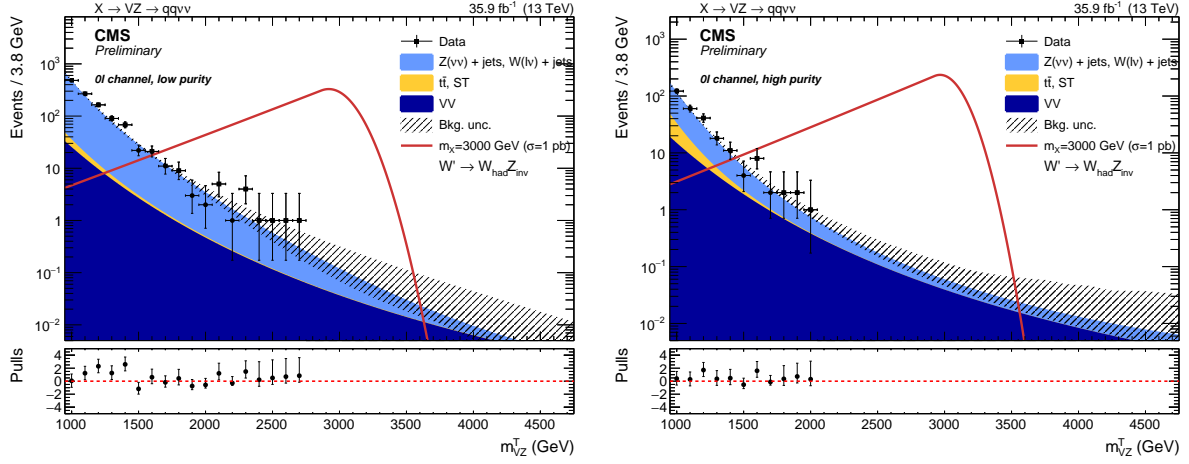


Figure 3.10: Expected background with the α method in the low- (left) and high-purity category (right), compared to observations (black dots) and a signal hypothesis of a spin-1 W' of mass 3 TeV.

3.1 Background estimation technique

3.1.4 Alpha method validation

As a validation of the α -ratio method described in this Section, a closure test is performed on data. Instead of predicting the background in the real V SR form 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 in two sub-regions: $30 - 50$ GeV (LSB) and $50 - 65$ GeV (SR). The first is considered as the low sideband for the purpose of this test, while the latter is exploited as a pseudo-signal region (and thus not kept blind). 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 the high-sidebands, and checked with data for both shape and normalization.

In Figure 3.11 and Table 3.4, 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 defined in the analysis.

region	category	Expected	Observed
SB	2ν , low-purity	1841.3 ± 45.7	1793
SR	2ν , low-purity	529.9 ± 37.8	521
SB	2ν , high-purity	728.5 ± 29.9	725
SR	2ν , high-purity	39.3 ± 5.2	49

Table 3.4: Expected and observed background yield in the SR jet mass region ($50 < m_j < 65\text{GeV}$), predicted from the LSB one ($30 < m_j < 50\text{GeV}$) and high-sideband ($m_j > 135$ GeV).

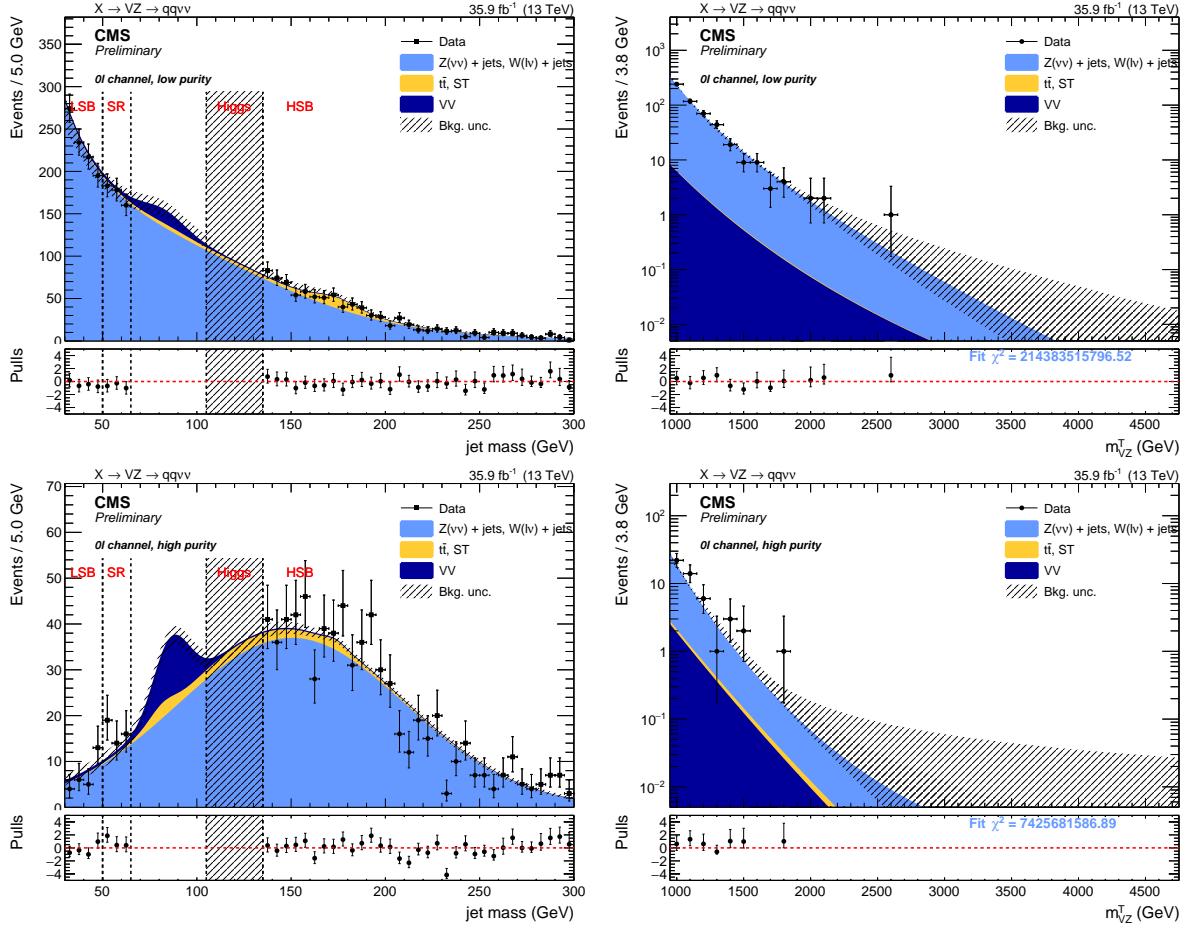


Figure 3.11: Top: fit to the m_j spectrum in data in the sideband defined for the Alpha method validation in the low-purity category: LSB ($30 < m_j < 50$ GeV) and high-sideband ($m_j > 135$ GeV). Bottom: fit to the m_j spectrum in data in the sideband defined for the Alpha method validation in the high-purity category: LSB ($30 < m_j < 50$ GeV) and high-sideband ($m_j > 135$ GeV). Both the signal region and the Higgs regions are kept blind, while the pseudo-signal is shown in the SR region ($50 < m_j < 65$ GeV).

91 3.1.5 Signal modeling

92 The simulated signal mass points are fitted in the SR with an empiric function in order to be able to
93 perform an unbinned likelihood fit for the signal extraction.

94 The function chosen to model the signal samples is a *Crystal ball* function, which is composed
95 by a gaussian-like peak plus a tail towards lower values. Both spin-2 and spin-1 signal samples are
96 fitted.

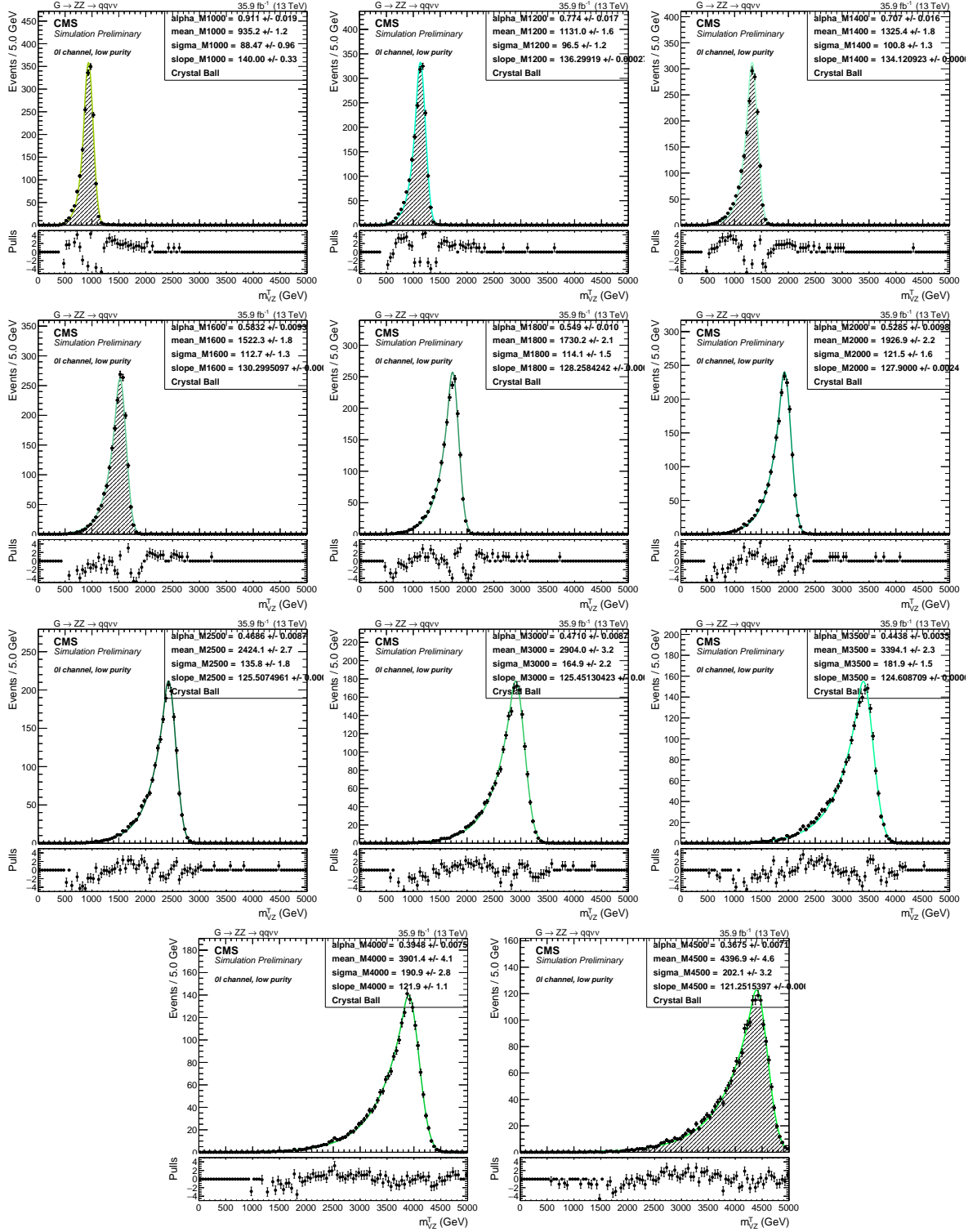


Figure 3.12: Fit to the spin-2 (Bulk Graviton) signal samples in the low-purity category.

3.1 Background estimation technique

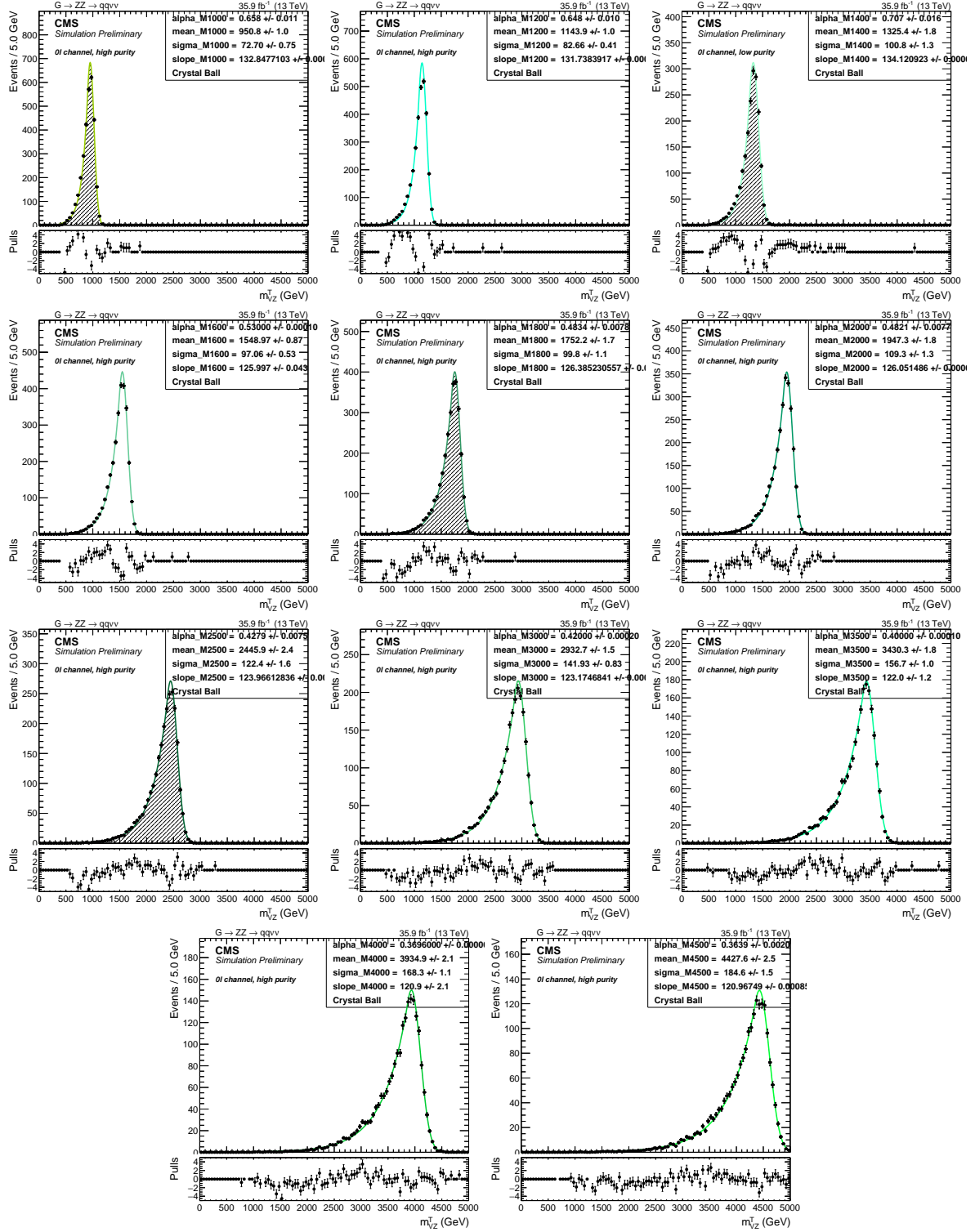
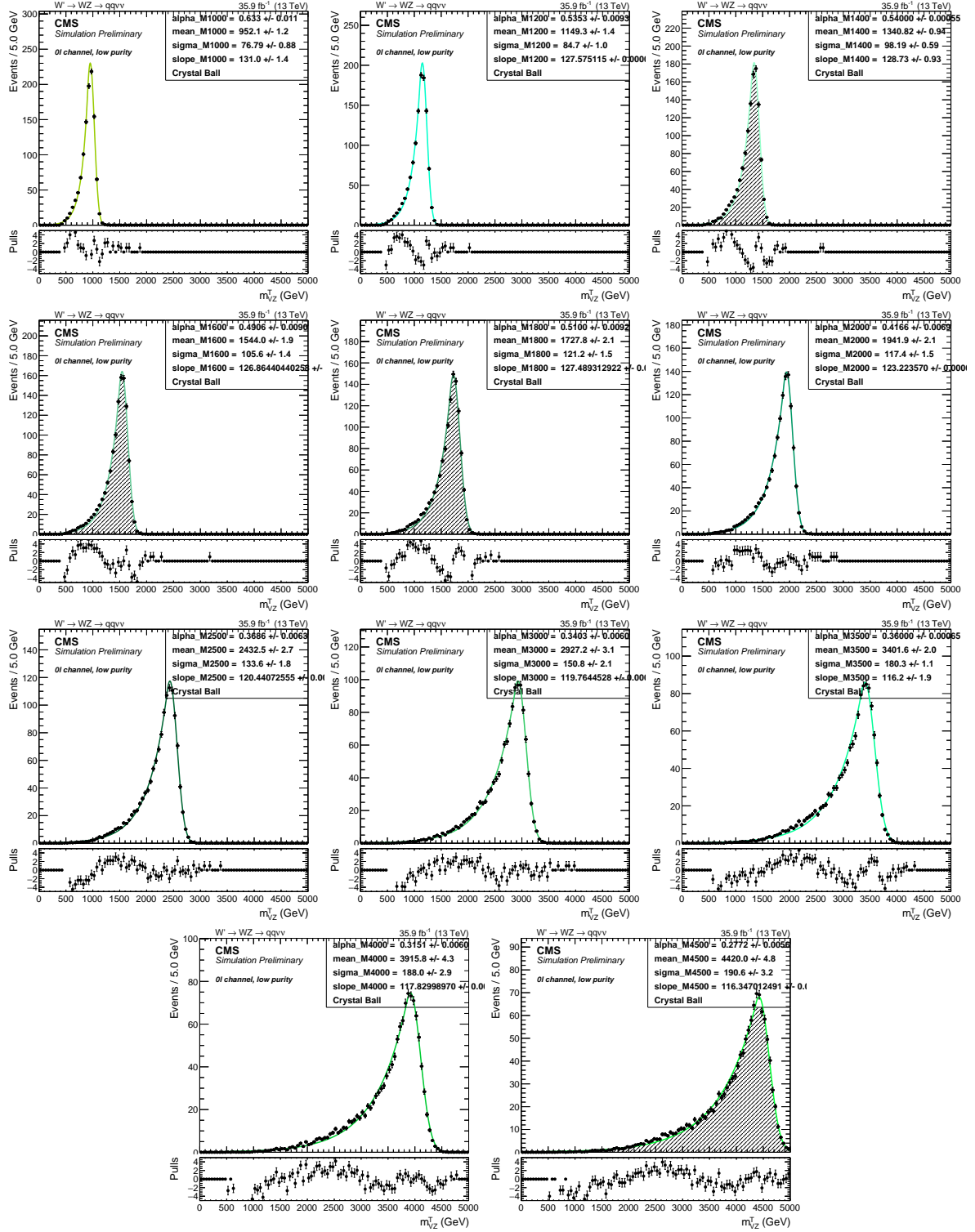


Figure 3.13: Fit to the spin-2 (Bulk Graviton) signal samples in the high-purity category.


 Figure 3.14: Fit to the spin-1 (W') signal samples in the low-purity category.

3.1 Background estimation technique

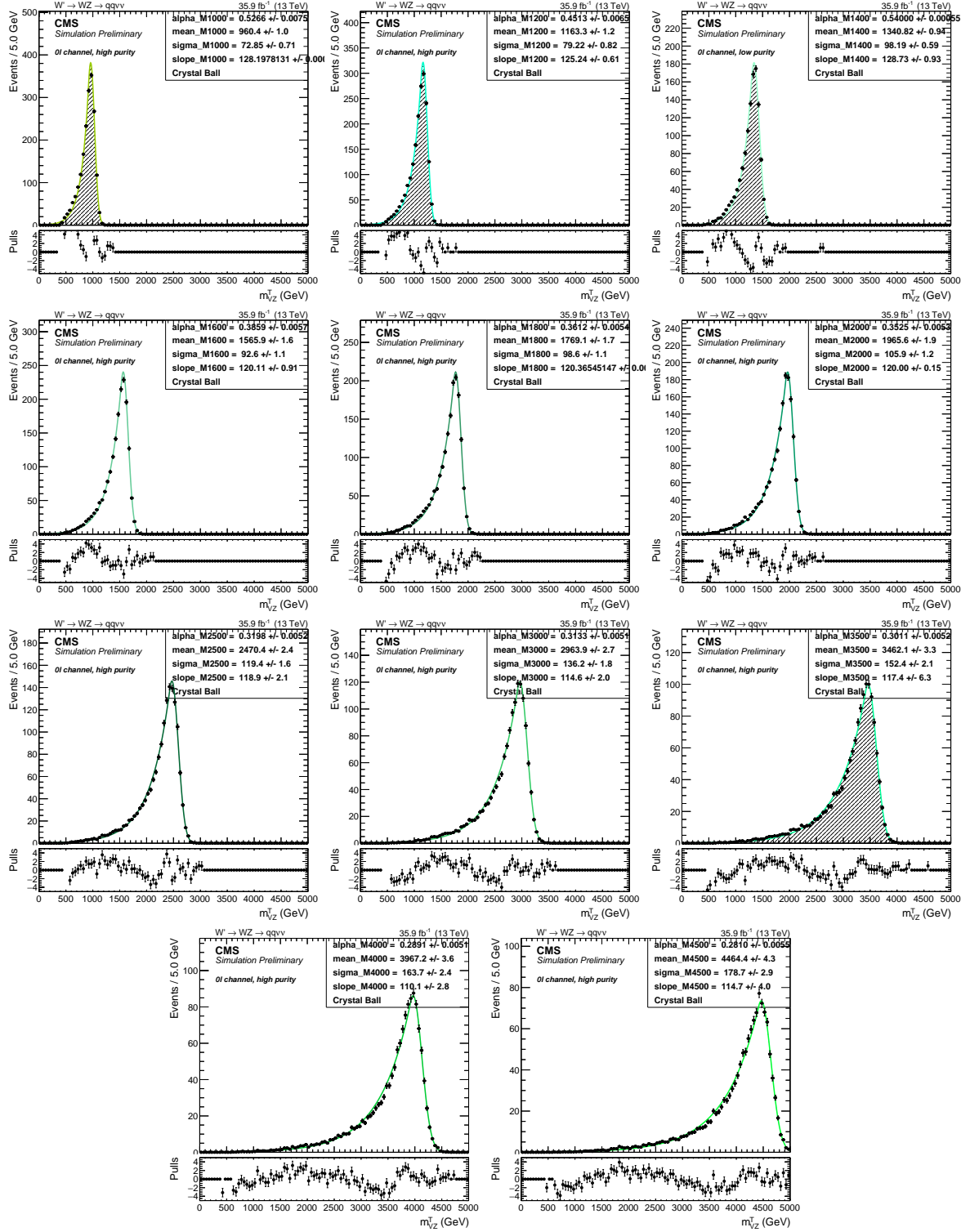


Figure 3.15: Fit to the spin-1 (W') signal samples in the high-purity category.

3.1.5.1 Signal parametrization

The signal is parametrized by interpolating the fitted parameters separately for each channel in order to have a continuous variation of the signal shape for every possible m_X value within the range. A linear fit is performed on the mean and the width of the Gaussian. The interpolations are shown in Fig. 3.16- 3.17 for the spin-2 signal model, and in Fig. 3.18- 3.19 for the spin-1 signal model. Shape systematic uncertainties, as described in section ??, 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 3rd degree polynomial.

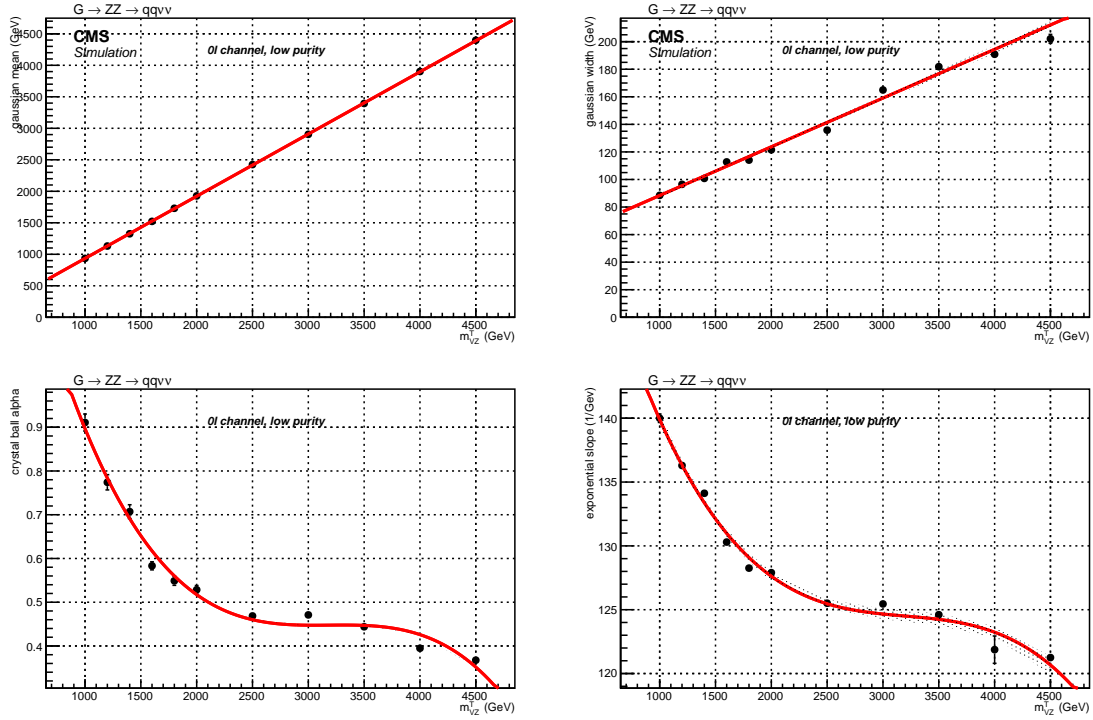


Figure 3.16: Interpolation of the fitted parameters as a function of the generated mass m_{VZ}^T , for a spin-2 (Bulk Graviton) signal hypothesis, low-purity category.

The normalization of the signal samples is extrapolated from the fitted integral of the Gaussian functions. The points are then connected with a line, in order to have an acceptable description of the normalization as a function of m_{VZ}^T . The interpolations are shown in Fig. 3.20 for the spin-2 signal model, in Fig. 3.21 for the spin-1 signal model.

3.1 Background estimation technique

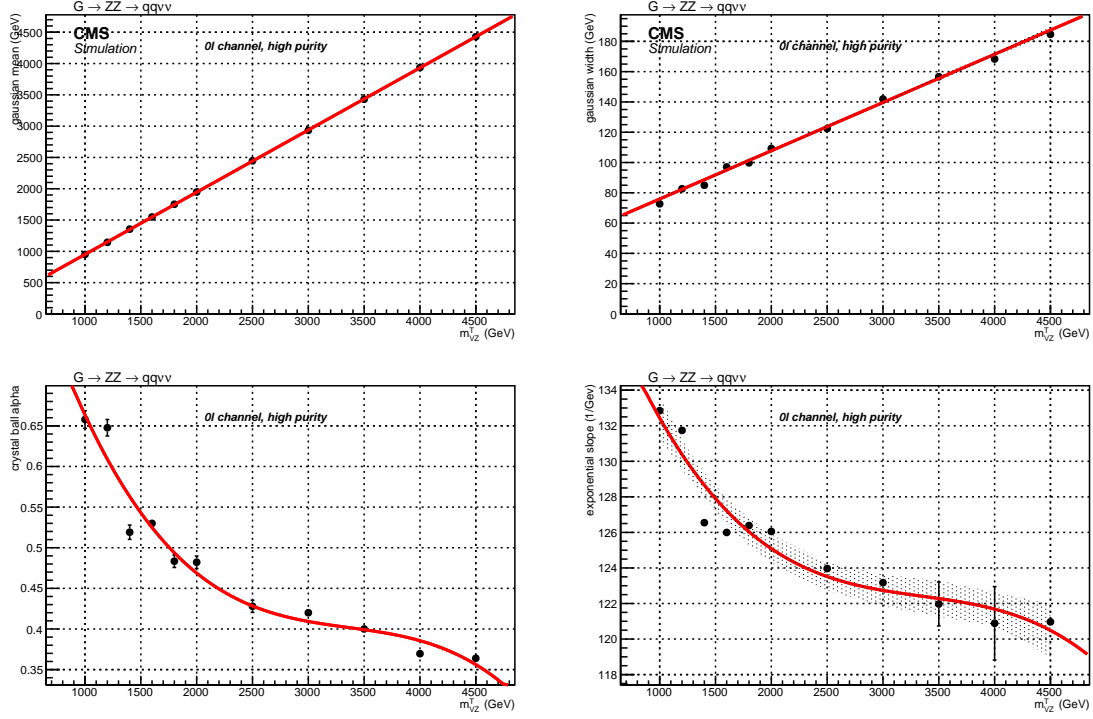


Figure 3.17: Interpolation of the fitted parameters as a function of the generated mass m_{VZ}^T , for a spin-2 (Bulk Graviton) signal hypothesis, high-purity category.

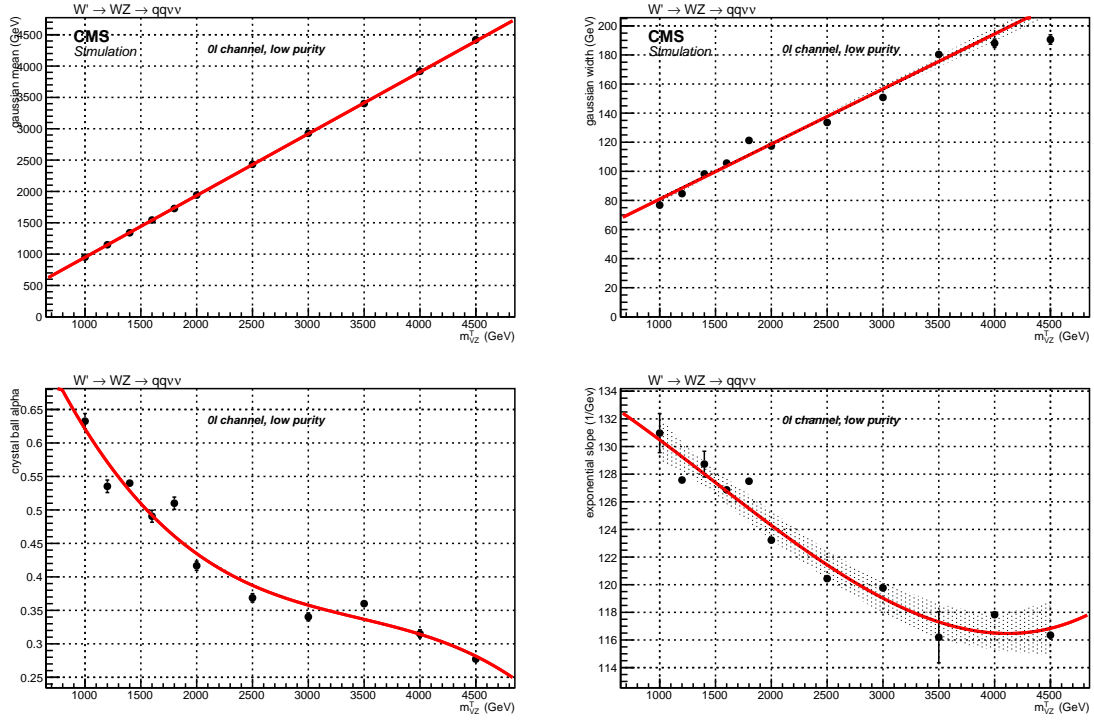


Figure 3.18: Interpolation of the fitted parameters as a function of the generated mass m_{VZ}^T , for a spin-1 (W') signal hypothesis, low-purity category.

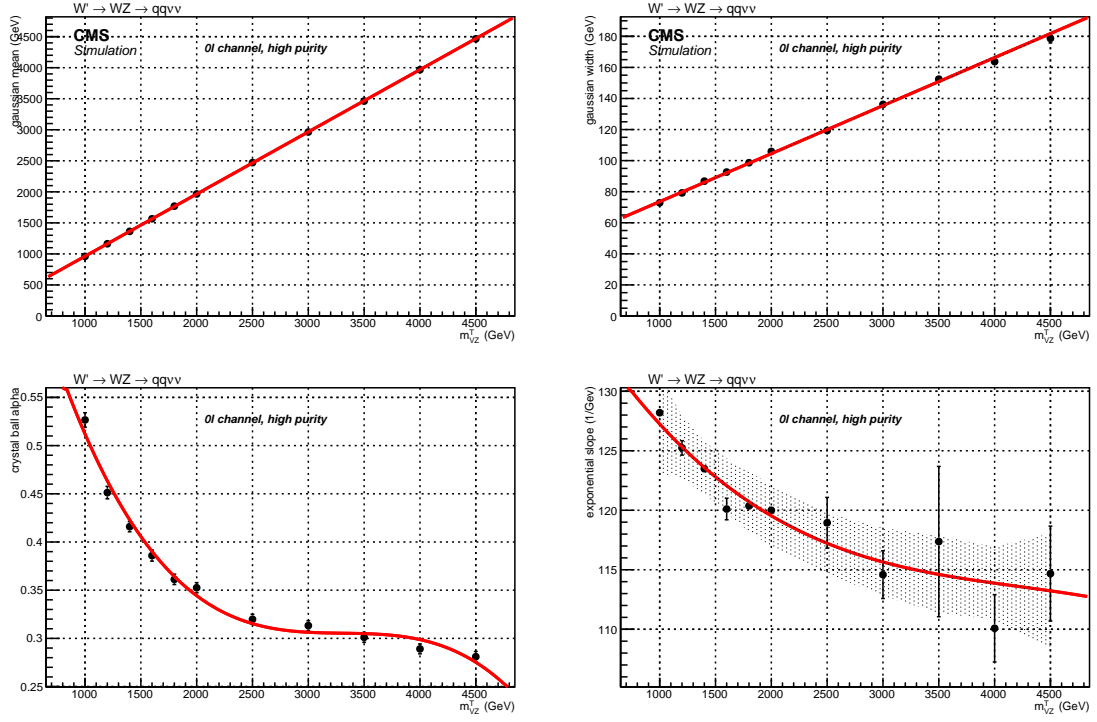


Figure 3.19: Interpolation of the fitted parameters as a function of the generated mass m_{VZ}^T , for a spin-1 (W') signal hypothesis, high-purity category.

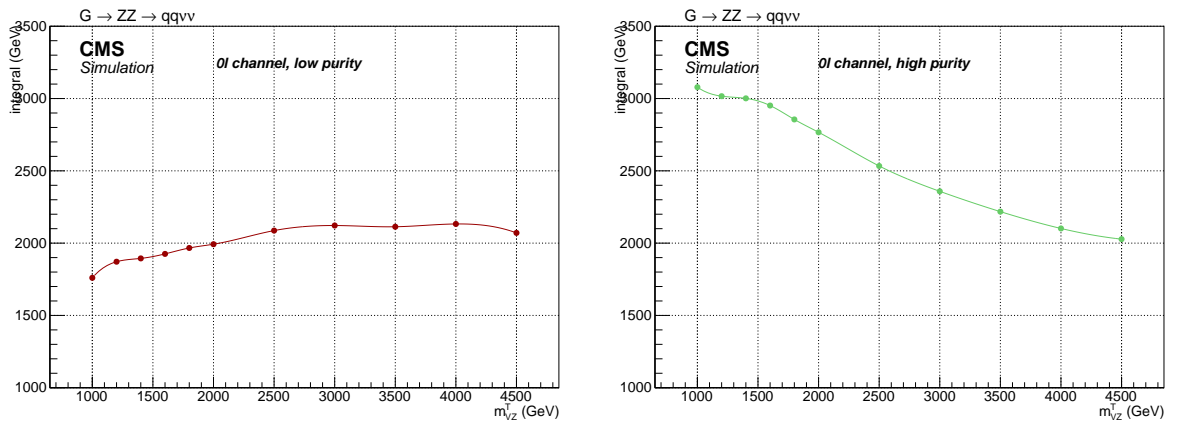


Figure 3.20: Interpolation of the signal normalization as a function of the generated mass m_{VZ}^T , for a spin-2 (Bulk Graviton) signal hypothesis. From left to right: low-purity, high-purity.

3.1 Background estimation technique

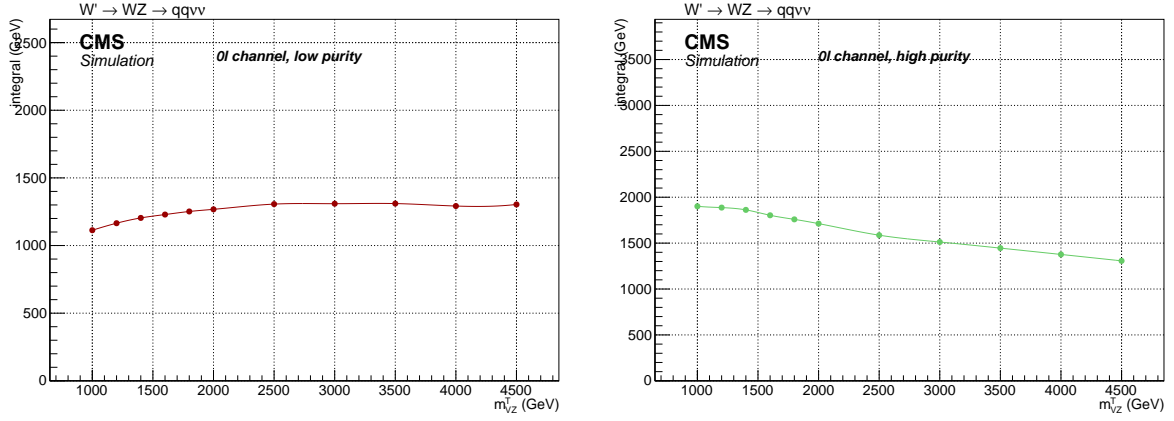


Figure 3.21: Interpolation of the signal normalization as a function of the generated mass m_{VZ}^T , for a spin-1 (W') signal hypothesis. From left to right: low-purity, high-purity.

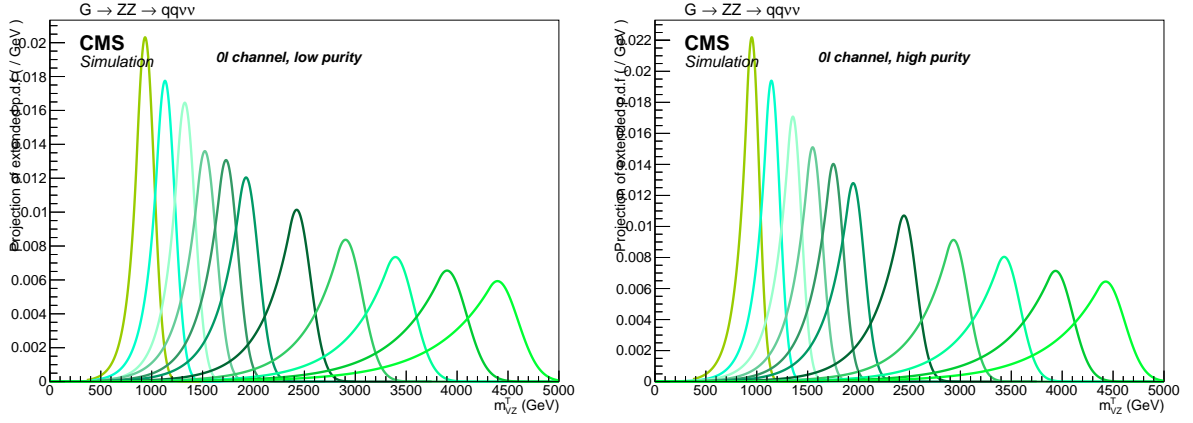


Figure 3.22: Interpolation of the signal as a function of the generated mass m_{VZ}^T , for a spin-2 (Bulk Graviton) signal hypothesis with an arbitrary cross section of 1 pb in the low- (left) and high-purity category (right).

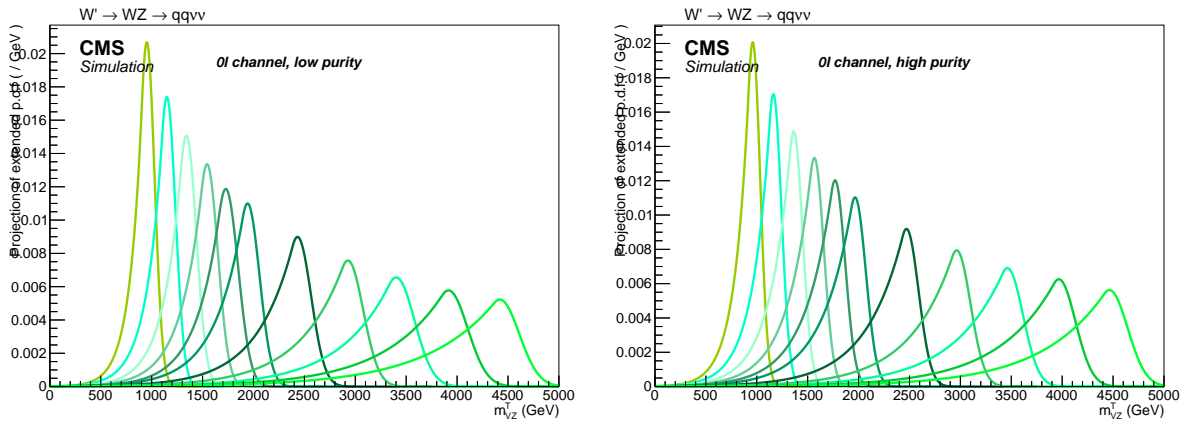


Figure 3.23: Interpolation of the signal as a function of the generated mass m_{VZ}^T , for a spin-1 (W') signal hypothesis with an arbitrary cross section of 1 pb in the low- (left) and high-purity category (right).

110

Chapter

4

Conclusions

111

112

113 Plot di Clemens

114

Bibliography

115

116

- 117 [1] CMS Collaboration, V. Khachatryan et al., *Search for a higgs boson in the mass range from 145 to*
118 *1000 GeV decaying to a pair of W or Z bosons, JHEP* **10** (2015) 144, [1504.00936].