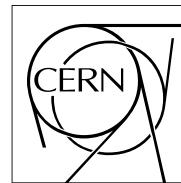


The Compact Muon Solenoid Experiment
Analysis Note

The content of this note is intended for CMS internal use and distribution only



24 January 2013 (v8, 06 August 2013)

Search for a BSM resonance decaying to W vector bosons in the semileptonic final state

C. Bernardes, A. Bonato, A. Hinzmann, Q. Li, S. Liu, M. Mozer, P. Mercadante, S. Novaes, M. Pierini, F. Santanastasio, A. Spiezia, T. Tomei

Abstract

A search for new particles decaying to two W bosons with subsequent decay to a final state containing two leptons and two quarks, X to WW to lν qqbar, is presented. Results are based on data corresponding to an integrated luminosity of 19.5 fb⁻¹ of proton-proton collisions at sqrt(s)=8 TeV and collected with the CMS detector at the CERN LHC. Jet sub-structure techniques are exploited for separating the signal from the SM background when the boost of the W causes the two quarks to merge into the same jet reconstructed in the detector. A model-independent statistical interpretation is applied to the selected dataset, setting limits on the cross section of a narrow resonance decaying to WW as a function of its mass.

CMS Draft Analysis Note

The content of this note is intended for CMS internal use and distribution only

2013/08/06

Head Id: 197036

<https://svn.cern.ch/repos/tdr2/notes/AN-13-045/trunk/AN-13-045.tex> Archive Id: 201186M

Archive Date: 2013/07/18

Archive Tag:

Search for a BSM resonance decaying to W vector bosons in the semileptonic final state

C. Bernardes⁵, A. Bonato¹, A. Hinzmann¹, Q. Li³, S. Liu³, Z. Xu³, M. Mozer², P. Mercadante⁵, S. Novaes⁴, M. Pierini¹, F. Santanastasio¹, A. Spiezia⁶, and T. Tomei⁴

¹CERN

²KIT

³Peking University

⁴SPRACE-UNESP

⁵SPRACE-UFABC

⁶Università di Perugia

Abstract

A search for new particles decaying to two W bosons with subsequent decay to a final state containing two leptons and two quarks, $X \rightarrow WW \rightarrow q\bar{q}\ell\nu$, is presented. Results are based on data corresponding to an integrated luminosity of 19.5 fb^{-1} of proton-proton collisions at $\sqrt{s} = 8 \text{ TeV}$ and collected with the CMS detector at the CERN LHC. Jet sub-structure techniques are exploited for separating the signal from the SM background when the boost of the W causes the two quarks to merge into the same jet reconstructed in the detector. A model-independent statistical interpretation is applied to the selected dataset, setting limits on the cross section of a narrow resonance decaying to WW as a function of its mass.

This box is only visible in draft mode. Please make sure the values below make sense.

PDFAuthor:

EXO-VV analysis team

PDFTitle:

Search for a BSM resonance decaying to vector bosons in the semileptonic final state Xto WWto "(lv+)"(qbarq")

PDFSubject:

CMS

PDFKeywords:

CMS, physics, exotica, graviton, boosted jets

Please also verify that the abstract does not use any user defined symbols

DRAFT

1 Contents

2	1	Preface	2
3	2	Introduction	2
4	3	Monte Carlo Samples and Data Sets	5
5	3.1	Signal MC	5
6	3.2	Background MC and cross-sections	9
7	3.3	Data samples	9
8	4	Event Reconstruction and Selection	12
9	4.1	Events Cleanup	12
10	4.2	Trigger and skim requirements	13
11	4.3	Lepton selection	14
12	4.4	Jet selection	15
13	4.5	E_T^{miss} selection	17
14	5	Reconstruction of the WW resonance	18
15	5.1	Leptonic W reconstruction	18
16	5.2	Hadronic W reconstruction	18
17	5.3	Reconstruction of the WW resonance	20
18	6	Event Selection	21
19	6.1	Kinematic criteria	21
20	6.2	Control plots in the hadronic W mass sideband	21
21	6.3	Optimization of the τ_{21} selection	29
22	6.4	Summary of final selection	34
23	6.5	Data vs MC comparison in signal region	35
24	7	Signal modeling: m_{WW} shapes and efficiency	38
25	7.1	Parametrization of the m_{WW} signal shape	38
26	7.2	Signal efficiency	39
27	8	Background estimate	43
28	8.1	Method 1 - α method	43
29	8.2	Method 2 - m_{WW} smoothness test	60
30	9	Systematic Uncertainties	62
31	9.1	Background Normalization and Shape	62
32	9.2	Signal Normalization	62
33	9.3	Signal Shape	65
34	10	Results	68
35	10.1	Tools for statistical analysis	68
36	10.2	Limits and p-values - Method 1	69
37	10.3	Limits and p-values - Method 2	74
38	11	Conclusions	79
39	A	B-tag veto optimization	80
40	B	F-test for background fit function	83
41	C	$t\bar{t}$ control plots	88
42	D	Closure test for Method 1	92

43 1 Preface

44 This analysis has followed a blinding policy agreed with the EXO conveners in November
 45 2012. The signal region was unblinded on May 6th, 2013.

46 The search has been performed by two different groups (with different analysis tools) working
 47 in close contact since January 2013 (i.e. group 1 = the authors of this analysis note and group
 48 2 = the authors of [1]). A synchronization of the object reconstruction and event selection be-
 49 tween the two groups was performed after the unblinding. We are still not fully synchronized
 50 due to known differences in the E_T^{miss} (few% level in the reconstructed p_T) which requires to
 51 reproduce all the PATtuples. For the PAS, we plan the following:

- 52 • use group 1 root-trees (since they implement the E_T^{miss} calculation consistent with
 53 the rest of the analysis);
- 54 • α method for background estimation (the group 2 implementation)
- 55 • cross-check of the α method using a smoothness test to m_{WW} distribution (a la “dijet”
 56 fit), using group 1 implementation.

57 2 Introduction

58 The Standard Model (SM) of particle physics has been very successful in describing high-
 59 energy physics phenomena investigated so far [2]. The discovery of the Higgs has completed
 60 the Standard Model (SM) as we know it [3, 4]. On the other hand, the SM is expected to be
 61 just an effective theory valid at energy scales up to ~ 1 TeV. The light Higgs mass leaves open
 62 the questions about the naturalness of the SM and leaves the door open to the presence of new
 63 physics to stabilize the Higgs vacuum. For larger scales, the theory predictions violate unitarity
 64 unless the constants of the SM are extremely fine tuned. This problem can be reformulated in
 65 terms of the large difference between the mass of the Higgs boson ($M_H \lesssim 1$ TeV/ c^2) and the
 66 Planck scale, M_{Pl} , where the gravitational force is expected to be of the same size of the other
 67 fundamental forces ($M_{Pl} \sim 10^{16}$ TeV/ c^2). This is usually called the *hierarchy problem*.

68 A particularly interesting example of physics Beyond the SM (BSM) predicts the presence of
 69 Extra-Dimensions (ED) beyond the $3 + 1$ dimensions experienced in our everyday life [5, 6].
 70 The space-time would then consist of $4 + N$ dimensions, where N is the number of additional
 71 ED. The boson carrier of the gravitational force, the *graviton*, can propagate also in the ED,
 72 effectively “diluting” the gravitational force. The actual scale where gravity becomes strong,
 73 $M_{Pl(4+N)}$, would then be much lower than the one effectively experienced in the $4D$ brane

$$M_{Pl(4+N)} = \frac{1}{2\pi R} (2\pi R M_{Pl})^{\frac{2}{N+2}} \quad (1)$$

74 Randall and Sundrum proposed a model where there is only one warped ED [7, 8]. Two $4D$
 75 branes (the Planck brane and the TeV brane) are separated by the fifth ED with size r_c . If the
 76 metric of the 5-th dimension is exponentially warped with a curvature k , the intensity of the
 77 coupling between the graviton and the SM particles will be dumped exponentially with an ex-
 78 ponential factor kr_c , explaining the weakness of the gravitational interaction. The relationship
 79 between the M_{pl} and the actual value on the Planck brane, M_5 is [7]

$$M_{Pl} = (1 - e^{-2kr_c}) \frac{M_5^3}{k} \quad (2)$$

80 This model, known as RS1, has two fundamental parameters: the mass of the first mode of
 81 the KK graviton, M_1 , and the dimensionless coupling strength $\tilde{k} = k/M_{Pl}$. The RS1 model
 82 predicts that the graviton couples preferentially to photons and SM fermions. Several searches
 83 for such signatures of the RS1 scenario have been pursued at the TeVatron [9–11] and at the
 84 LHC [12–19].

85 Different models with warped ED allow the SM fields to propagate in the 5-th dimension [20–
 86 22]. These models, generally defined as “*Bulk graviton*” models, can deal with the electroweak
 87 precision tests and the current limits on flavor changing neutral currents better than the RS1
 88 model. Furthermore they can give an explanation to the mass hierarchy in the SM: lighter
 89 fermions are those allowed to propagate farther in the ED, thus having weaker couplings to
 90 the Higgs boson localized near the TeV brane. These models are characterized by different
 91 couplings of the graviton to the SM fields, with a large branching ratio to SM vector boson pairs
 92 that can become dominant according to the parameters of the model. Therefore, the search
 93 for gravitons in the vector boson final state ($V=W,Z$) can be the “golden mode” for detecting
 94 evidences of this type of BSM physics due the large decay rate and the clean experimental
 signature [20].

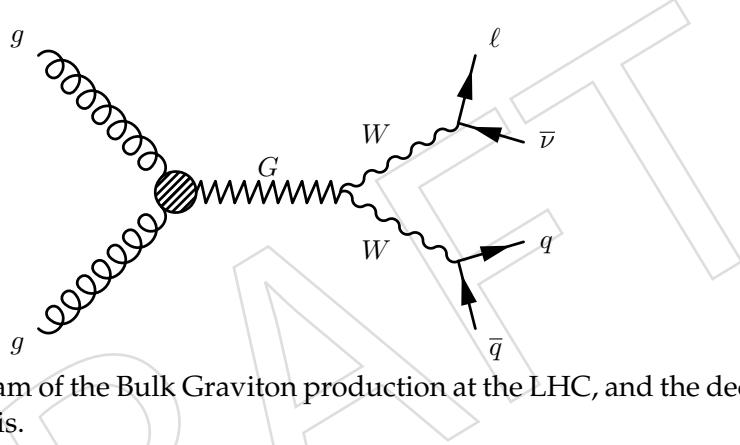


Figure 1: Feynman diagram of the Bulk Graviton production at the LHC, and the decay channel considered in this analysis.

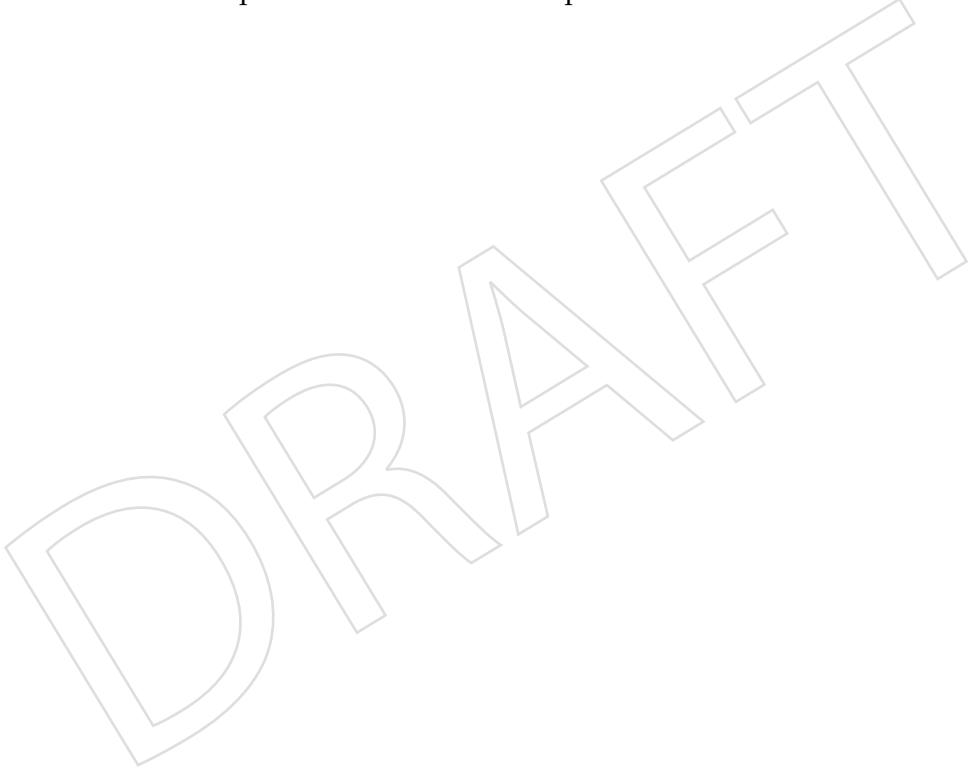
95
 96 Despite of the appeal of the specific Bulk graviton model, the interest for new physics mani-
 97 festing itself in vector bosons covers a much broader spectrum of scenarios. Many other BSM
 98 models propose the existence of new resonances coupling to pairs of vector bosons.

99 Ref. [23] considers a spin-1 particle decaying to ZZ/WW. Hidden valley models admit in prin-
 100 ciple BSM heavy particles of any spin state decaying to ZZ/WW [24]. Even if the final state
 101 is always composed by two vector bosons, the underlying physics and the kinematics of the
 102 events can differ a lot between the different cases. Even in the ED case, the kinematics of the
 103 decay products of a RS1 graviton and a bulk graviton are distinct because of different spin cor-
 104 relations among the final physics objects. These differences reflect in very different acceptances
 105 and signal efficiencies for a given set of event selections. This means that a search in this final
 106 state must be carried out with a sufficiently generic approach in order not to miss an a-priori
 107 open possibility. More targeted selections, exploiting the characteristic kinematics of the signal,
 108 can be done tailoring the event selection specifically for some benchmark models judged to be
 109 more appealing than others.

110 The lack of evidence for BSM physics in the searches carried out so far at colliders and elec-
 111 troweak precision tests (EWPT) push the mass of these resonances at the TeV scale, as con-
 112 firmed by the lack of a signal in the 7 TeV CMS searches [17–19]. In this mass regime, a large
 113 fraction of $X \rightarrow VV$ events are characterized by one boosted V decaying hadronically, where
 114 the two quarks merge into a single reconstructed jet.

In this note, a search for TeV resonances going to WW final states is presented. The focus is not exclusively on spin-2 models, although we use Bulk Graviton samples for studying the signal efficiency. The selections are kept voluntarily loose in order not to depend too much on the nature of the BSM resonance. We reconstruct the leptonic W candidates as $\ell\nu$ ($\ell = e, \mu$) pairs. The neutrino longitudinal momentum is calculated solving a second-order equation that makes use of the constrain from the known W mass. The hadronic W is reconstructed as a single, massive jet. The jet is identified as W if it passes selection criteria based on the jet mass and the sub-structure of the jet (V-tagging).

This note is organized as follows. In Sect. 3.1 and 3.2 the MC samples used are listed for the signal and background, respectively. In Sect. 3.3 the datasets analyzed are presented. The triggers used, the identification and selection of the leptons and jets, and the E_T^{miss} are described in Sect. 4. The reconstruction of the $X \rightarrow WW$ resonance and the heavy resonance tagging of the jets is described in Sect. 5. The final selection, including its optimization, is presented in Sect. 6 followed in Sect. 7 and 8 by the description of the signal and background modeling respectively. Sect. 9 addresses the estimation of the systematic uncertainties of the measurement. Finally, the statistical interpretation of the results is presented in Sect. 10.



131 3 Monte Carlo Samples and Data Sets

132 3.1 Signal MC

133 As signal hypothesis, the RS graviton [7, 8] and Bulk graviton [20] were considered using a rep-
 134 resentative set of couplings and masses. A set of simulated samples were privately produced
 135 using the full machinery available in CMSSW, following the conditions used for the official
 136 Summer12 campaign. For the Bulk graviton model, parton-level process files were generated
 137 in the LHE format with the JHU generator version 2.0.2 [25, 26], showering and hadronization
 138 were performed with PYTHIA version 6.426 [27], with the tune Z2star configuration, while
 139 the cross-sections were calculated with the CALCHEP generator [28], version 3.4. The cross-
 140 sections for the $pp \rightarrow G_{bulk} \rightarrow ZZ(WW) \rightarrow \ell\ell(\ell\nu) + qq$ process are shown in Fig. 2. For the
 141 RS graviton model, the generation was fully performed with PYTHIA. GEN-SIM files were
 142 produced with CMSSW version 5.2.6, which uses GEANT4 version 9.4.p03-cms for the full
 143 simulation of the CMS detector. The DIGI-RAW and RECO steps were produced with CMSSW
 144 version 5.3.2_patch4. These steps include an approximate description of pileup effects through
 145 the mixing of simulated minimum bias at the DIGI level. The simulated minimum bias sam-
 146 ples used are those from the dataset /MinBias_TuneZ2star_8TeV-pythia6/Summer12-
 147 START50_V13-v3/GEN-SIM, and the distribution of the number of pileup events superim-
 148 posed to the signal event was done according to the 2012_Summer_50ns_PoissonOOTPU con-
 149 figuration. The generated samples are summarized in Table 1.

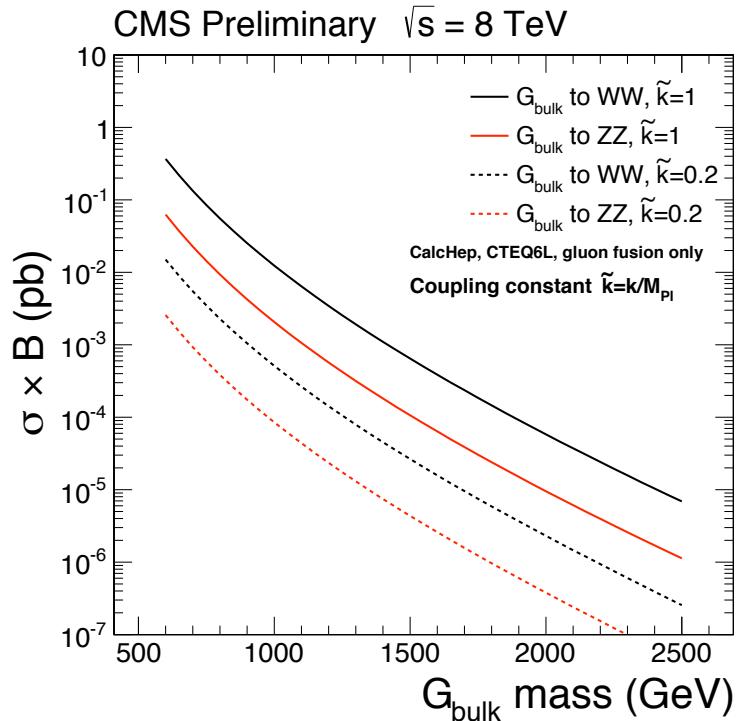


Figure 2: The cross-sections for the $pp \rightarrow G_{bulk} \rightarrow ZZ(WW) \rightarrow \ell\ell(\ell\nu) + qq$ process at a center-of-mass energy of 8 TeV, as calculated by CALCHEP.

150 In order to validate the private production, we performed a comparison between the privately
 151 produced sample /RSWW_1000_02_SIM_v2/qili-RSWW_1000_02_AODSIM_v2-c8f8ed334db8a7d6f
 152 56c62266b1dfa5b/USER, and the corresponding centrally produced sample (i.e. official Sum-
 153 mer12 production) /RSGravitonToWW_kMp102_M-1000_TuneZ2star_8TeV-pythia6/Summer12-
 154 DR53X-PU_S10_START53_V7A-v1/AODSIM. The private sample was produced with the same con-

figuration used in the official production. The comparison reported in Figs. 3 and 4 shows a good agreement between the two samples, thus validating our private production.

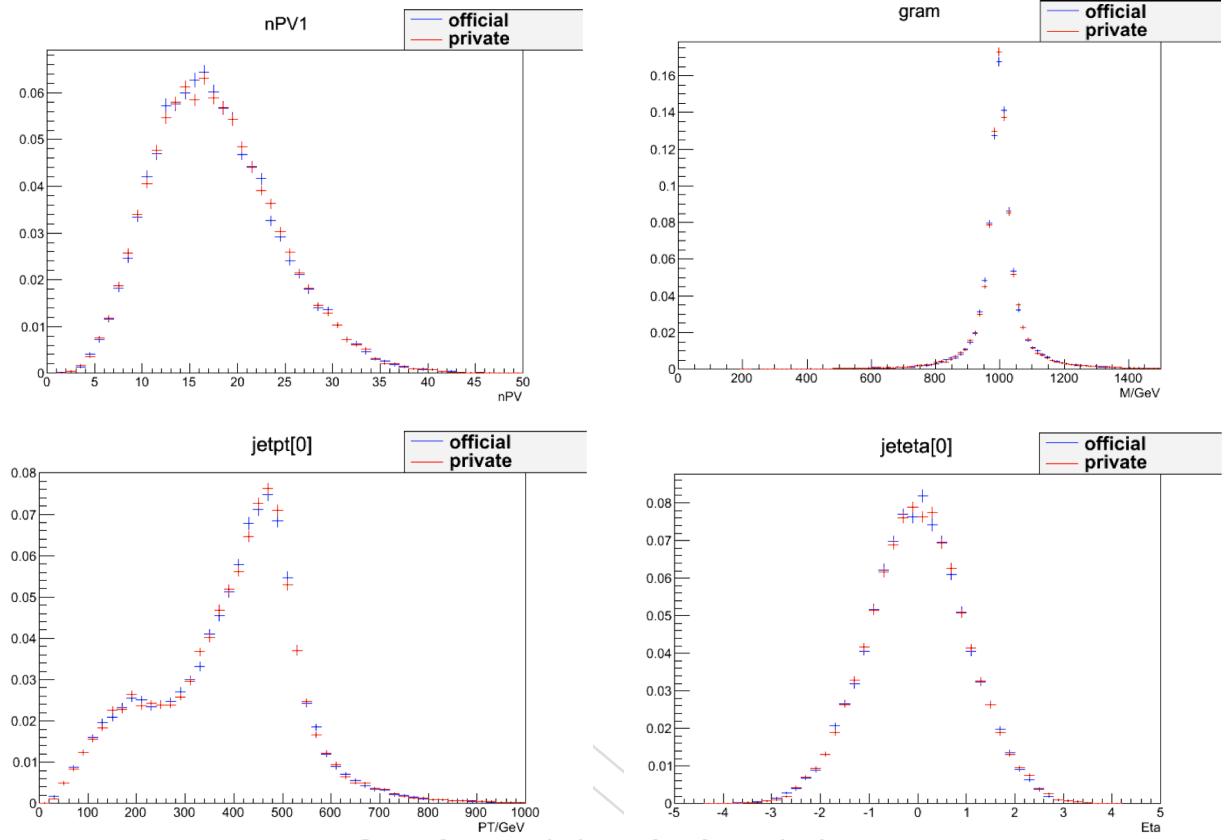


Figure 3: Comparison in between the official sample of RS graviton and a privately produced sample for the same model and same parameters, at RECO level (unless otherwise noted). The histograms are normalized to the same area. Top left: number of reconstructed vertices. Top right: generator level graviton mass. Bottom left: leading jet p_T . Bottom right: leading jet η .

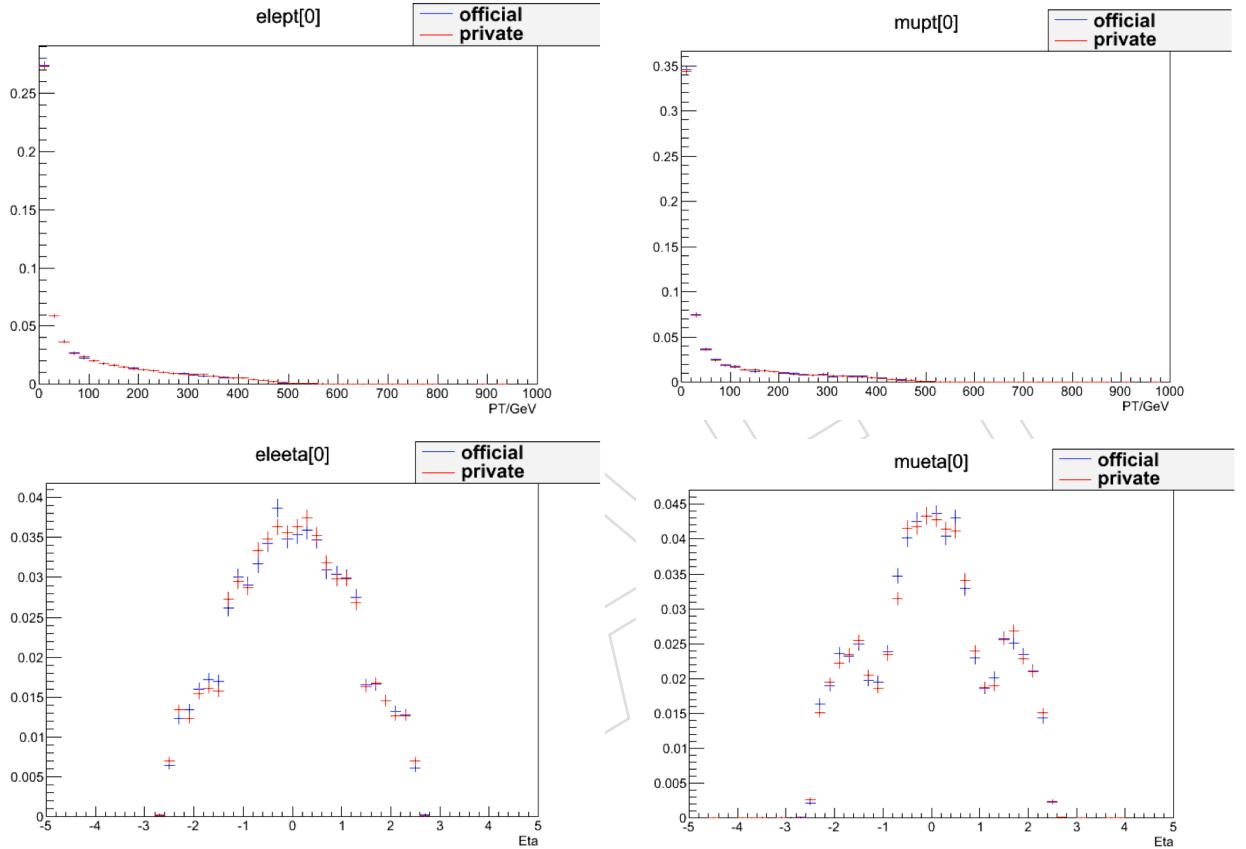


Figure 4: Comparison in between the official sample of RS graviton and a privately produced sample for the same model and same parameters, at RECO level. The histograms are normalized to the same area. Top left: leading electron p_T . Top right: leading muon p_T . Bottom left: leading electron η . Bottom right: leading muon η .

Table 1: Signal samples used in the analysis. The cross-sections quoted are for the full decay chain $pp \rightarrow X \rightarrow ZZ \rightarrow 2\ell 2q$, with $\ell = e, \mu$. The \tilde{k} parameter is shorthand for k/M_{Pl} . The samples can be found at `cms_dbs_ph_analysis_01`.

Sample	N	σ [pb]	\tilde{k}	M_G [TeV]
RS graviton				
/RSGraviton_WW_lvjj_kMpl02_M-600_TuneZ2star_8TeV-MG/	31875	8.42E+00	0.2	0.6
qili-RSWW_600_02_AODSIM-c8f8ed334db8a7d6f56c62266b1dfa5b/USER				
/RSGraviton_WW_lvjj_kMpl-005_M-1000_TuneZ2star_8TeV-MG/	35421	3.11E-02	0.05	1.0
qili-RSWW_1000_005_AODSIM-c8f8ed334db8a7d6f56c62266b1dfa5b/USER				
/RSGraviton_WW_lvjj_kMpl005_M-1000_TuneZ2star_8TeV-MG/	34541	4.81E-01	0.2	1.0
qili-RSWW_1000_02_AODSIM-c8f8ed334db8a7d6f56c62266b1dfa5b/USER				
/RSGraviton_WW_lvjj_kMpl02_M-1500_TuneZ2star_8TeV-MG/	23879	3.69E-02	0.2	1.5
Bulk graviton				
/BulkG_WW_lvjj_c0p2_M600-JHU-v2/qili-BulkG_WW_lvjj_c0p2_M600-JHU-AODSIM-	48949	0.015	0.2	0.6
c8f8ed334db8a7d6f56c62266b1dfa5b/USER				
/BulkG_WW_lvjj_c0p2_M700-JHU-v1/qili-BulkG_WW_lvjj_c0p2_M700-JHU-AODSIM-v1-	9832	0.005478	0.2	0.7
c8f8ed334db8a7d6f56c62266b1dfa5b/USER				
/BulkG_WW_lvjj_c0p2_M800-JHU-v1/qili-BulkG_WW_lvjj_c0p2_M800-JHU-AODSIM-	9506	0.00228	0.2	0.8
c8f8ed334db8a7d6f56c62266b1dfa5b/USER				
/BulkG_WW_lvjj_c0p2_M900-JHU-v1/qili-BulkG_WW_lvjj_c0p2_M900-JHU-AODSIM-	9506	0.001048	0.2	0.9
c8f8ed334db8a7d6f56c62266b1dfa5b/USER				
/BulkG_WW_lvjj_c0p2_M1000-JHU-v3/qili-BulkG_WW_lvjj_c0p2_M1000-JHU-AODSIM-	48796	0.0005114	0.2	1.0
c8f8ed334db8a7d6f56c62266b1dfa5b/USER				
/BulkG_WW_lvjj_c0p2_M1100-JHU-v1/qili-BulkG_WW_lvjj_c0p2_M1100-JHU-AODSIM-	9833	0.0002646	0.2	1.1
c8f8ed334db8a7d6f56c62266b1dfa5b/USER				
/BulkG_WW_lvjj_c0p2_M1200-JHU-v2/qili-BulkG_WW_lvjj_c0p2_M1200-JHU-AODSIM-	44952	0.00014120	0.2	1.2
c8f8ed334db8a7d6f56c62266b1dfa5b/USER				
/BulkG_WW_lvjj_c0p2_M1300-JHU-v1/qili-BulkG_WW_lvjj_c0p2_M1300-JHU-AODSIM-	9767	7.90E-05	0.2	1.3
c8f8ed334db8a7d6f56c62266b1dfa5b/USER				
/BulkG_WW_lvjj_c0p2_M1400-JHU-v1/qili-BulkG_WW_lvjj_c0p2_M1400-JHU-AODSIM-	9831	4.52E-05	0.2	1.4
c8f8ed334db8a7d6f56c62266b1dfa5b/USER				
/BulkG_WW_lvjj_c0p2_M1500-JHU-v1/qili-BulkG_WW_lvjj_c0p2_M1500-JHU-AODSIM-v1-	49996	2.65E-05	0.2	1.5
c8f8ed334db8a7d6f56c62266b1dfa5b/USER				
/BulkG_WW_lvjj_c0p2_M1600-JHU-v1/qili-BulkG_WW_lvjj_c0p2_M1600-JHU-AODSIM-v1-	45994	1.58E-05	0.2	1.6
c8f8ed334db8a7d6f56c62266b1dfa5b/USER				
/BulkG_WW_lvjj_c0p2_M1700-JHU-v1/qili-BulkG_WW_lvjj_c0p2_M1700-JHU-AODSIM-v1-	9899	9.57E-06	0.2	1.7
c8f8ed334db8a7d6f56c62266b1dfa5b/USER				
/BulkG_WW_lvjj_c0p2_M1800-JHU-v1/qili-BulkG_WW_lvjj_c0p2_M1800-JHU-AODSIM-	9833	5.87E-06	0.2	1.8
c8f8ed334db8a7d6f56c62266b1dfa5b/USER				
/BulkG_WW_lvjj_c0p2_M1900-JHU-v1/qili-BulkG_WW_lvjj_c0p2_M1900-JHU-AODSIM-	9897	3.67E-06	0.2	1.9
c8f8ed334db8a7d6f56c62266b1dfa5b/USER				
/BulkG_WW_lvjj_c0p2_M2000-JHU-v1/qili-BulkG_WW_lvjj_c0p2_M2000-JHU-AODSIM-v1-	49360	2.31E-06	0.2	2.0
c8f8ed334db8a7d6f56c62266b1dfa5b/USER				
/BulkG_WW_lvjj_c0p2_M2100-JHU-v1/qili-BulkG_WW_lvjj_c0p2_M2100-JHU-AODSIM-	9900	1.46E-06	0.2	2.1
c8f8ed334db8a7d6f56c62266b1dfa5b/USER				
/BulkG_WW_lvjj_c0p2_M2200-JHU-v1/qili-BulkG_WW_lvjj_c0p2_M2200-JHU-AODSIM-v1-	9702	9.40E-07	0.2	2.2
c8f8ed334db8a7d6f56c62266b1dfa5b/USER				
/BulkG_WW_lvjj_c0p2_M2300-JHU-v1/qili-BulkG_WW_lvjj_c0p2_M2300-JHU-AODSIM-	9899	6.03E-07	0.2	2.3
c8f8ed334db8a7d6f56c62266b1dfa5b/USER				
/BulkG_WW_lvjj_c0p2_M2400-JHU-v1/qili-BulkG_WW_lvjj_c0p2_M2400-JHU-AODSIM-	9899	3.91E-07	0.2	2.4
c8f8ed334db8a7d6f56c62266b1dfa5b/USER				
/BulkG_WW_lvjj_c0p2_M2500-JHU-v1/qili-BulkG_WW_lvjj_c0p2_M2500-JHU-AODSIM-	9898	2.55E-07	0.2	2.5
c8f8ed334db8a7d6f56c62266b1dfa5b/USER				

157 3.2 Background MC and cross-sections

158 The dominant background in the $G \rightarrow WW \rightarrow \ell\nu 2q$ analysis is the inclusive W production with
159 jets ($W+jets$). Other backgrounds considered are inclusive $t\bar{t}$ production, standard model dibo-
160 son production (WW , WZ and ZZ), single- t production, and the inclusive Drell-Yan production
161 with jets ($Z/\gamma/Z^*+jets$). All the background samples considered were produced in the frame of
162 the Summer12 production. The $W+jets$ and $DY+jets$ samples that we use have been produced
163 with MadGraph. The $t\bar{t}+jets$ and single- t samples have been produced with POWHEG. The
164 diboson samples were produced with Pythia6. A full list of the samples is found in Table 2.
165 The cross sections are taken from <https://twiki.cern.ch/twiki/bin/view/CMS/StandardModelCrossSectionsat8TeV>: the cross sections for $W+jets$ and $DY+jets$ samples
166 (generated in p_T bins) are obtained by rescaling the corresponding LO cross sections by a flat
167 NNLO/LO k-factor derived from the inclusive production; the cross sections for the remaining
168 processes are either NLO or approx. NNLO calculations.
169

170 3.3 Data samples

171 We use about 19.5 fb^{-1} of data collected in 2012 and prompt-reco'ed or re-reco'ed with CMSSW_5_3_X.
172 The data samples are listed in Table 3. We use only lumisections that have been declared good
173 for analysis by the central certification team, listed in the following official JSON file:

- 174 • Cert_190456-196531_8TeV_13Jul2012ReReco_Collisions12_JSON_v2.txt – for Run2012A
175 and Run2012B re-reconstruction.
- 176 • Cert_190782-190949_8TeV_06Aug2012ReReco_Collisions12_JSON.txt – for Run2012A
177 “recover” re-reconstruction.
- 178 • Cert_198022-198523_8TeV_24Aug2012ReReco_Collisions12_JSON.txt – for Run2012C
179 re-reconstruction.
- 180 • Cert_190456-203002_8TeV_PromptReco_Collisions12_JSON_v2.txt – for Run2012C
181 prompt reconstruction.
- 182 • Cert_190456-208686_8TeV_PromptReco_Collisions12_JSON.txt – for Run2012D
183 prompt reconstruction.

Table 2: Background samples used in the analysis.

Sample	Number of events	σ [pb]
MadGraph samples		
/WJetsToLNu_PtW-50To70_TuneZ2star_8TeV-madgraph/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	24.95M	1001.0
/WJetsToLNu_PtW-70To100_TuneZ2star_8TeV-madgraph/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	20.9M	529.3
/WJetsToLNu_PtW-100_TuneZ2star_8TeV-madgraph/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	12.1M	282.5
/WJetsToLNu_PtW-180_TuneZ2star_8TeV-madgraph-tarball/ Summer12_DR53X-PU_S10_START53_V7C-v1/AODSIM	9.73M	29.00
/DYJetsToLL_PtZ-50To70_TuneZ2star_8TeV_ext-madgraph-tarball/ Summer12_DR53X-PU_S10_START53_V7C-v1/AODSIM	4.82M	105.7
/DYJetsToLL_PtZ-70To100_TuneZ2star_8TeV_ext-madgraph-tarball/ Summer12_DR53X-PU_S10_START53_V7C-v1/AODSIM	4M	62.9
/DYJetsToLL_PtZ-100_TuneZ2star_8TeV_ext-madgraph-tarball/ Summer12_DR53X-PU_S10_START53_V7C-v1/AODSIM	2.5M	39.1
Powheg samples		
/TT_Chandle_TuneZ2star_8TeV-powheg-tauola/ Summer12_DR53X-PU_S10_START53_V7A-v2/AODSIM	21.7M	225.2
/T_s-channel_TuneZ2star_8TeV-powheg-tauola/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	0.209M	3.79
/T_t-channel_TuneZ2star_8TeV-powheg-tauola/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	3.75M	56.4
/T_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	0.498M	11.1
/Tbar_s-channel_TuneZ2star_8TeV-powheg-tauola/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	0.140M	1.76
/Tbar_t-channel_TuneZ2star_8TeV-powheg-tauola/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	1.90M	30.7
/Tbar_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	0.493M	11.1
Pythia samples		
/WW_TuneZ2star_8TeV_pythia6_tauola/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	3.87M	57.1
/WZ_TuneZ2star_8TeV_pythia6_tauola/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	1.91M	33.2
/ZZ_TuneZ2star_8TeV_pythia6_tauola/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	0.486M	8.06

Table 3: Data samples used in the analysis.

Sample	Run Range	$\mathcal{L} [\text{pb}]^{-1}$
/SingleMu/Run2012A-13Jul2012-v1/AOD	190456–193621	808.47
/SingleMu/Run2012A-recover-06Aug2012-v1/AOD	190782–190949	82.14
/SingleMu/Run2012B-13Jul2012-v1/AOD	193833–196531	4426
/SingleMu/Run2012C-24Aug2012-v1/AOD	198022–198913	495
/SingleMu/Run2012C-PromptReco-v2/AOD	198934–203746	6382
/SingleMu/Run2012C-EcalRecover_11Dec2012-v1/AOD	201191–201191	134.24
/SingleMu/Run2012D-PromptReco-v1/AOD	203768–208686	7210
TOTAL SingleMu	190456–208686	19538
/SingleElectron/Run2012A-13Jul2012-v1/AOD	190456–193621	808.5
/SingleElectron/Run2012A-recover-06Aug2012-v1/AOD	190782–190949	82.1
/SingleElectron/Run2012B-13Jul2012-v1/AOD	193833–196531	4422
/SingleElectron/Run2012C-24Aug2012-v1/AOD	198022–198913	495
/SingleElectron/Run2012C-PromptReco-v2/AOD	198934–203746	6386
/SingleElectron/Run2012C-EcalRecover_11Dec2012-v1/AOD	201191–201191	134.24
/SingleElectron/Run2012D-PromptReco-v1/AOD	203768–208686	7222
TOTAL SingleElectron	190456–208686	19550

184 4 Event Reconstruction and Selection

185 Events are processed with PAT in CMSSW_5_3_3_patch3 and CMGTools version V5_10_0.

186 We describe in the next sections the reconstruction of muons, electrons, jets, E_T^{miss} , $X \rightarrow WW$ resonances.
 187 The general strategy is to keep the selection requirements as loose as possible, in order not to
 188 bias the selected sample towards a specific BSM model. For convenience we summarize the
 189 selection requirements in Tab. 4

190 4.1 Events Cleanup

191 4.1.1 Noise Filters

192 Events are required to pass the following set of standard noise filters:

- 193 • primaryVertexFilterPath
- 194 • noscrapingFilterPath
- 195 • hcalLaserEventFilterPath
- 196 • HBHENoiseFilterPath
- 197 • trackingFailureFilterPath
- 198 • CSCTightHaloFilterPath
- 199 • eeBadScFilterPath
- 200 • EcalDeadCellTriggerPrimitiveFilterPath
- 201 • EcalLaserCorrectionFilterPath

202 The primaryVertexFilterPath filter requires the presence of at least one reconstructed
 203 vertex, passing the following requirements:

- 204 • obtained from a fit of a suitable set of tracks, i.e., !isFake
- 205 • number of degrees of freedom larger than 4
- 206 • absolute value of the longitudinal coordinate $\text{abs}(z)$ smaller than 24 cm
- 207 • transversal distance to the longitudinal axis ρ smaller than 2 cm.

208 In case more than one suitable primary vertex is found, the vertex with the highest tracks sum-
 209 p_T^2 is taken as the primary vertex of the event. This vertex serves as the reference to identify
 210 the tracks associated to the event that fired the trigger and it is used for PU removal in jet
 211 clustering.

212 4.1.2 Pileup removal

213 The presence of additional interactions with respect to the primary one, known as Pile-up (PU),
 214 is expected to affect this analysis in the following ways:

- 215 • additional energy from PU get added to the jets from the main interaction, modify-
 216 ing the energy and the position of the reconstructed jet;
- 217 • additional low p_T jets fully composed of PU energy get added to the event.

218 The amount of PU interaction per event in data and MC is shown in Fig. 5. In the following
 219 this distribution in MC has been re-weighted to match the data. In Fig. 6 we show consistency
 220 in the number of reconstructed vertexes between data and MC after MC samples have been
 221 re-weighted.

222 Various algorithms are available to correct for PU effects. The charged particles coming from
 223 PU can also be removed before the jet clustering requiring that all the tracks come from the
 224 primary vertex with an algorithm called “Charged Hadron Subtraction” (CHS). In this case
 225 also the PU effect on jet angles is partially corrected¹. The bias to the jet kinematics coming
 226 from neutral hadrons of the PU is corrected with the so-called Fastjet method.

227 In the rest of the analysis the effect of the PU on the jets is corrected using both the CHS and the
 228 Fastjet algorithms. Following the recommendations of the MUO and EGM POGs, no attempt is
 229 done to correct the electrons and muons for PU for the dedicated high- p_T selection employed in
 230 this search. The final systematics due to PU on the signal acceptance is discussed in Section 9.

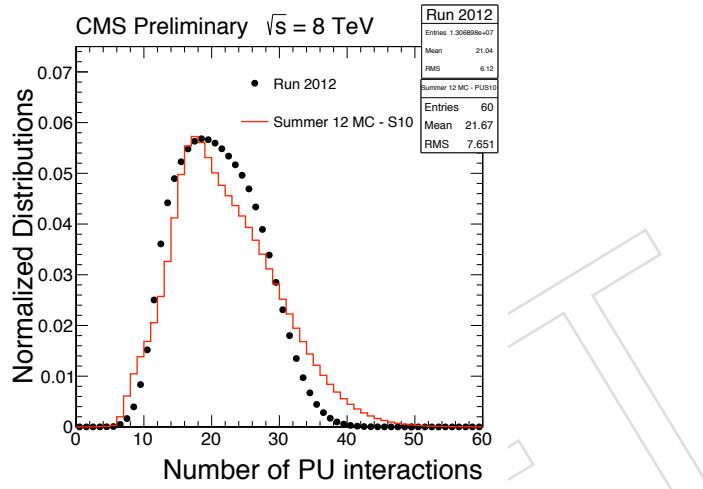


Figure 5: Number of PU interactions in data and MC. In the following the MC sample has been reweighted to match the data.

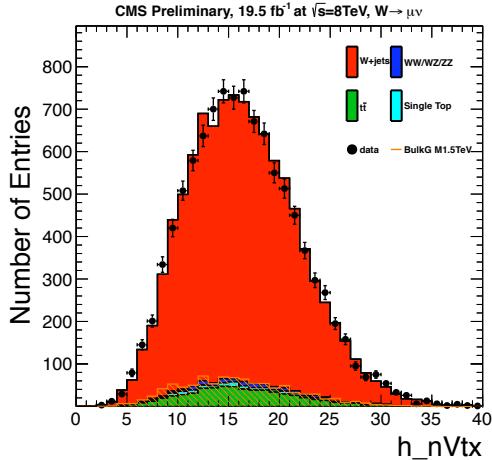


Figure 6: Number of reconstructed vertexes after the reweighting of the MC sample by the pile-up corrections. Points with error bars show data, histograms show contribution of dominant background channels, both after the selection described in Sect. 4.

231 4.2 Trigger and skim requirements

232 The results presented in this note are based on the SingleMu, and SingleElectron datasets. Each
 233 of these datasets contain at least one un-prescaled single lepton trigger with looser p_T require-

¹Alternatively, only jets with a sizable amount of tracks coming from primary vertex can be considered. This strategy is not used in this analysis.

234 ments than our offline selections. The lowest threshold un-prescaled, non-isolated triggers are
 235 used:

- 236 • HLT_Mu40_eta2p1_*

237 for the SingleMu dataset, and

- 238 • HLT_Ele80_CaloIdVT_TrkIdT_* (active for 5% of the data period)
 239 • HLT_Ele80_CaloIdVT_GsfTrkIdT_* (active for 95% of the data period)

240 for the SingleElectron dataset.

241 The trigger selection is not applied in the MC, instead the trigger efficiency is computed from
 242 data with a Tag-and-Probe technique and the Monte Carlo samples are correct accordingly by
 243 using the trigger efficiency as an event weight. The efficiency for these triggers with respect of
 244 the offline lepton selection employed in this analysis have been studied in detail in the context
 245 of the search for W' in lepton+neutrino channel [29] and they are summarized below.

- 246 • The efficiency of the single muon trigger is $0.94010 \pm (0.0 < |\eta| < 0.9)$, 0.84368
 247 $(0.9 < |\eta| < 1.2)$, and 0.82423 ($1.2 < |\eta| < 2.1$) depending on the muon η ;
- 248 • The efficiency of the single electron trigger is 0.991 ($0.0 < |\eta| < 1.4442$) and 0.976
 249 ($1.566 < |\eta| < 2.5$)² depending on the electron η .

250 The uncertainties on the trigger efficiency obtained from data are below 1%.

251 4.3 Lepton selection

252 4.3.1 Muon Selection

253 For the muon identification and reconstruction of muon momentum assignment we follow the
 254 recommendation of the MUON POG at
 255 <https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideMuonId>.

256 The old versions (before CMSSW_5_3_6_patch1) of the HighPT muon ID and the TuneP mo-
 257 mentum assignment are used³:

- 258 • muon identified both as TrackerMuon and GlobalMuon;
- 259 • number of muon hits larger than zero;
- 260 • number of matched muon stations larger than one;
- 261 • transversal impact parameter $d\eta$ smaller than 0.2 cm;
- 262 • longitudinal impact parameter $d\phi$ smaller than 0.5 cm;
- 263 • number of pixel hits larger than zero;
- 264 • number of tracker layer with hits larger than eight.

265 The last cut on number of tracker layer with hits is tightened with respect to the new version of
 266 HighPT muon ID in order to suppress muons with largely mis-measured p_T . This cut is known
 267 to introduce inefficiency of about 5% due to tracking changes in CMSSW 52X⁴

268 Further, the muon candidates are requested to pass a loose relative isolation cut, $Iso < 0.10$,
 269 where the subdetector-based isolation variable is defined as the ratio between the sum of tracks
 270 p_T in a cone of $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.3$ around the muon direction divided by the p_T of the

²The EB-EE transition region $1.4442 < |\eta| < 1.566$ is excluded from the analysis

³We plan to move to the new version of the HighPT muon ID when the new analysis PATtuples are produced.

⁴By using the new HighPT definition this efficiency loss would be recovered.

271 muon itself.

272 We select Tight muon candidates passing the HighPT muon ID+ISO with $p_T > 50$ GeV and
 273 $|\eta| < 2.1$ (to be in the plateau of the trigger efficiency). Loose muon candidates (used for the
 274 lepton veto selection) are required to pass the HighPT muon ID+ISO with $p_T > 20$ GeV and
 275 $|\eta| < 2.4$.

276 The muon ID+ISO efficiencies have been measured in data and MC [29]. Data/MC scale factors
 277 of ≈ 0.99 are used as event weights to correct the MC.

278 4.3.2 Electron Selection

279 For the electron identification and reconstruction of electron p_T assignment we follow the rec-
 280 ommendation of the EGM POG at

281 <https://twiki.cern.ch/twiki/bin/viewauth/CMS/HEEPElectronID> .

282 The latest version of the HEEP selection (v4.1) is used:

- 283 • supercluster pseudorapidity (η_{SC}) smaller than 1.442 (for barrel electrons) or in the
 284 range 1.56–2.5 (for endcap electrons);
- 285 • driven by calorimetric deposits (`isEcalDriven`);
- 286 • have $|\Delta\eta_{in}|$ smaller than 0.005 (0.007) for barrel electrons (endcap electrons);
- 287 • have $|\Delta\phi_{in}|$ smaller than 0.06 (both for barrel and endcap electrons);
- 288 • have the ratio of hadronic to electromagnetic energy smaller than 0.05;
- 289 • have $\sigma_{t\eta i\eta}$ smaller than 0.03 (for endcap electrons only);
- 290 • have either $E_{2\times 5}/E_{5\times 5}$ larger than 0.94 or $E_{1\times 5}/E_{5\times 5}$ larger than 0.83;
- 291 • have the number of inner layer lost hits smaller than 2;
- 292 • have $|d_{xy}|$ smaller than 0.02 (0.05) cm for barrel electrons (endcap electrons).

293 Further, the electron candidates are requested to pass the following detector-based isolation
 294 cuts (isolation cone $\Delta R = 0.3$):

- 295 • EM + Had Depth 1 Isolation $< 2 + 0.03 \cdot Et + 0.28 \cdot \rho$ for the barrel electrons;
- 296 • EM + Had Depth 1 Isolation $< 2.5 + 0.28 \cdot \rho$ for $Et < 50$, else $< 2.5 + 0.03 \cdot (Et - 50) +$
 297 $0.28 \cdot \rho$ for the endcap electrons;
- 298 • Trk Isolation (sum of tracks p_T) < 5 GeV for both barrel and endcap electrons.

299 We select Tight electron candidates passing the HEEP 4.1 electron ID+ISO with $p_T > 90$ GeV
 300 and $0.0 < |\eta| < 1.442$ OR $1.56 < |\eta| < 2.5$ (to be in the plateau of the trigger efficiency).
 301 Loose electron candidates (used for the lepton veto selection) are required to pass the HEEP
 302 4.1 electron ID+ISO with $p_T > 35$ GeV (this p_T cut is included by default in the HEEP 4.1
 303 selection) and the same η range of the tight electrons.

304 The electron ID+ISO efficiencies have been measured in data and MC [29]. Data/MC scale
 305 factors of ≈ 0.98 are used as event weights to correct the MC.

306 4.4 Jet selection

307 Jets are reconstructed from the list of Particle Flow (PF) candidates [30] reconstructed in the
 308 event. Charged hadrons originating from vertices other than the primary vertex were not used
 309 in the jet clustering with the Charged Hadron Subtraction (CHS) procedure. We use two jet
 310 collections:

- Cambridge-Aachen algorithm with distance parameter $R = 0.8$ (CA8) to select W candidates decaying to jets in boosted regime;
- $anti - k_T$ algorithm with distance parameter $R = 0.5$ (AK5) to require or veto the presence of b-tagged jets in the event.

4.4.1 CA8 Jet Selection

The selected PF candidates are clustered using a Cambridge-Aachen algorithm with distance parameter $R = 0.8$. The CA8 jet energy is corrected applying the L1L2L3 (L1L2L3+Residual) corrections for MC (data), derived for AK7 jets by the JetMET group [31]. This procedure was already validated in other analysis using wide jets [32, 33]. The global tag used for data (MC) is GR_P_V39_AN3 (START53_V15).

We select jet candidates with $p_T > 80$ GeV and $|\eta| < 2.4$. In addition, jets are subject to the following loose identification criteria ($> 99\%$ efficiency), JetID Loose:

- muon energy fraction smaller than 0.99
- photon energy fraction smaller than 0.99
- charged electromagnetic energy fraction smaller than 0.99
- neutral hadron energy fraction smaller than 0.99
- charged hadron energy fraction larger than 0
- number of constituent particles larger than 1

If a CA8 jet is within $\Delta R < 1.0$ of any tight electron or tight muon defined in the previous sections, the jet is not used in the analysis.

An additional filter on jets in the barrel-endcap transition region of the silicon tracker is applied, following the studies of Ref. [34]. Jets are rejected if their pseudorapidity falls in the region $1.0 < |\eta| < 1.5$ and the ratio between the multiplicity of charged PF candidates in the jet over neutral PF candidates in the jet, R_{CN} , is larger than 2. The goal of this filter is to remove jets where the tracking software behaves in an anomalous way, and many fake tracks are associated to the jet. The inefficiency of this cut for the signal events is below 1%.

4.4.2 AK5 Jet Selection and b-tagging

The selected PF candidates are clustered using an $anti - k_T$ algorithm with distance parameter $R = 0.5$. The jet energy is corrected applying the L1L2L3 (L1L2L3+Residual) corrections for MC (data), derived for the same algorithm by the JetMET group [31]. The global tag used for data (MC) is GR_P_V39_AN3 (START53_V15).

We select jet candidates with $p_T > 30$ GeV and $|\eta| < 2.4$, passing the JetID Loose criteria described above.

If an AK5 jet is within $\Delta R < 0.3$ of any tight electron or tight muon defined in the previous sections, the jet is not used in the analysis.

For b-tagging [35], we use the Combined Secondary Vertex algorithm with medium working point (CVSM). We apply b-tag scale factors using the method 1a (with event reweighting) described at <https://twiki.cern.ch/twiki/bin/viewauth/CMS/BTagSFMethods>. All MC samples used in the analysis are re-weighted accordingly with these factors. The MC-based b-tag efficiencies (for the calculation of the event weights) are obtained from the $t\bar{t}$ sample used in this analysis as a function of the jet p_T , jet η , and separately for light and heavy flavour jets. The same efficiencies are used also for the calculation of the weights in the $X \rightarrow WW$ signal sam-

353 ples. A negligible difference is found in the value of the event weights for signal samples when
 354 using either efficiencies derived from $t\bar{t}$ sample (as done in this analysis) or derived directly
 355 from the same signal samples ⁵.

356 **4.5 E_T^{miss} selection**

357 The missing transverse energy, E_T^{miss} , of the event is computed as the negative vector sum of all
 358 particle flow objects' transverse momenta.

359 We follow the recommendations of the JetMET POG reported at <https://twiki.cern.ch/twiki/bin/view/CMS/EXOMissingET2012> and <https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookMetAnalysis>.

362 Corrections are applied to the E_T^{miss} to account for jet energy corrections (Type-1). The mini-
 363 mum p_T of the jets used in the correction is 10 GeV; jets with EM fraction greater than 90% or
 364 identified as muons are not used.

365 A ϕ modulation of reconstructed E_T^{miss} is observed in Data as well as in Monte Carlo simulated
 366 events. The modulation is due to a systematic shift of the E_T^{miss} x/y components which increases
 367 linearly with the number of reconstructed vertices. The magnitude of the shift is different in
 368 Data and MC though. A correction for this ϕ modulation has been developed and it is applied
 369 in this analysis (MET x/y Shift Correction)

370 An initial requirement $E_T^{\text{miss}} > 40$ GeV is applied in the analysis.

DRAFT

⁵The distribution of weight1 – weight2 divided by weight1, calculated event-by-event, is a Gaussian-like distribution with mean $\approx 0.2\%$ and RMS $\approx 0.4\%$.

371 5 Reconstruction of the WW resonance

372 5.1 Leptonic W reconstruction

373 $W \rightarrow \ell\nu$ candidates are formed from all the charged leptons passing the selection requirements
 374 (electrons or muons) and the E_T^{miss} (representing the transverse component of the neutrino mo-
 375 mentum) reconstructed in the event. The longitudinal component of the neutrino momentum,
 376 $p_{z\nu}$, is calculated by solving a second-order equation that makes use of the known W mass:

$$m_\ell^2 + m_\nu^2 + 2 \cdot (E_\ell E_\nu - p_{x\ell} p_{x\nu} - p_{y\ell} p_{y\nu} - p_{z\ell} p_{z\nu}) = m_W^2 = (80.4)^2 \quad (3)$$

377 The solutions are:

$$p_{z\nu}^\pm = \frac{-B \pm \sqrt{\Delta}}{2A} \quad (4)$$

378 where:

$$A = 4(E_\ell E_\nu - p_{z\ell} p_{z\nu}); \quad (5)$$

$$B = -4a p_{z\ell} \quad (6)$$

$$C = 4E_\ell E_\nu(p_{x\nu}^2 + p_{y\nu}^2) - a^2 \quad (7)$$

$$a = m_W^2 - m_\ell^2 + 2p_{x\ell} p_{x\nu} + 2p_{y\ell} p_{y\nu} \quad (8)$$

$$\Delta = B^2 - 4AC \quad (9)$$

379 If $\Delta > 0$ there are two real solutions and $p_{z\nu}$ is set equal to the one with the smaller absolute
 380 value. If $\Delta < 0$ there are two complex solutions and $p_{z\nu}$ is set equal to the real part of the two
 381 solutions ($p_{z\nu} = -B/2A$).

382 5.2 Hadronic W reconstruction

383 In the $X \rightarrow W$ events, the Lorentz factor γ (boost factor) of the W boson is approximately
 384 $M_X/2M_W$ [36]. Due to the kinematics of the event, a large boost factor implies that the de-
 385 cay products will be predominantly emitted with a small $\Delta R \approx 2M_W/p_T^W$ [37]. This analysis
 386 searches for X resonances with mass greater than ≈ 1 TeV, for which the boosted topology is
 387 by far the dominant one. The $W \rightarrow q\bar{q}$ candidate is therefore reconstructed as a single massive
 388 CA8 jet. The pruned jet mass and the N-subjettiness (jet substructure reconstructed quantity)
 389 are used to discriminate against standard QCD jets.

390 The general points discussed below for the W boson would also be valid for hadronic Z or H
 391 decays (for example if one searches for $X \rightarrow WZ$ or $X \rightarrow WH$ resonances in the $\ell\nu +$ jet final
 392 state).

393 5.2.1 Jet grooming: pruning

394 *Jet grooming* techniques improve the resolution on the jet mass and help in reducing the effect
 395 of pile-up. The goal of jet grooming is to define a procedure for reclustering the jet constituents
 396 eliminating those originating from soft, large-angle QCD radiation or from other QCD pro-
 397 cesses different from the hard one where the W boson was created. Different jet grooming
 398 algorithms have been implemented in the CMSSW software and their performance on jets in
 399 QCD processes has been studied in detail in [38].

400 In this analysis, we chose to use the *jet pruning* algorithm because of its good capability in
 401 improving the jet mass resolution, the stability in presence of pileup interactions, and be-
 402 cause of the better operational experience, thanks to the many analysis that already used it
 403 at CMS [19, 32, 33, 38].

404 The jet pruning algorithm [39] recluster the jet constituents removing at the same time soft or
 405 large angle recombinations. The jet pruning algorithm that we use is based on the CA clustering
 406 algorithm with distance parameter, $R = 0.8$. In the reclustering procedure, recombinations of
 407 jet constituents characterized by soft energies are vetoed through a cut on the quantity z_{ij}

$$z_{ij} = \frac{\min(p_{T,i}, p_{T,j})}{p_{T,(i+j)}} < z_{cut} \quad (10)$$

408 where the indexes i and j refer to the two jet constituents considered in the recombination step
 409 and z_{cut} is a parameter of the pruning set to $z_{cut} = 0.1$. A second requirement is applied on the
 410 ΔR between the two constituents, ΔR_{ij}

$$\Delta R_{ij} > D_{cut} = \varkappa_{pr} \times \frac{m}{p_T} \quad (11)$$

411 In this case, m and p_T refer to the original ungroomed jet and the parameter $\varkappa_{pr} = 0.5$. If any
 412 of the conditions $z_{ij} < z_{cut}$ or $\Delta R_{ij} > D_{cut}$, is satisfied, the softest of the two branches in the
 413 recombination is taken out of the list of the jet constituents. Otherwise, the two constituents
 414 are recombined according to the normal CA procedure.

415 The resulting pruned jet mass (m_{jet}^{pruned}) is used for W-tagging, since it exhibits better resolution
 416 and smaller biases with respect to the ungroomed (standard) jet mass.

417 For reconstructing the kinematics of the $X \rightarrow WW$ resonance we use still the kinematics of the
 418 original jet, however. In this way we still use properly calibrated jets for reconstructing the
 419 kinematics of the event⁶.

420 5.2.2 Jet substructure: N-subjettiness

421 The single CA8 jet produced by a boosted $W \rightarrow q\bar{q}'$ decay exhibit a dipole-like substructure.
 422 This feature can be quantified by several *jet substructure* quantities proposed in phenomenology
 423 studies. We consider here a discrimination based on the *N-subjettiness* [40–42].

424 The original jet is reclustered with the k_T algorithm, stopping the reclustering when N subjects
 425 are found. The N-subjettiness, τ_N , is then defined as

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min((\Delta R_{1,k})^\beta, (\Delta R_{2,k})^\beta, \dots, (\Delta R_{N,k})^\beta) \quad (12)$$

426 where the index k runs over the jet constituents and the distances $\Delta R_{n,k}$ are calculated with re-
 427 spect to the axis of the $n-th$ subjet. The normalization factor d_0 is calculated as $d_0 = \sum_k p_{T,k} R_0^\beta$,
 428 setting R_0 to the jet radius of the original jet. In the analysis, the N-subjettiness is calculated
 429 from the ungroomed jets with the parameter $\beta = 1$. Jets with $\tau_N \approx 0$ have all their radiation
 430 aligned with the candidate subjet directions and therefore have N (or fewer) subjets. Jets with
 431 $\tau_N \gg 0$ have a large fraction of their energy distributed away from the candidate subjet
 432 directions and therefore have at least $N+1$ subjets.

433 The variable able to best discriminate between W -jets and QCD jets is the ratio of 2-subjettiness
 434 over 1-subjettiness, $\tau_{21} = \tau_2 / \tau_1$. Our W -jets should have low values of τ_{21} compared to QCD

⁶Jet energy scale corrections for pruned jets are not yet available in CMS.

435 jets. In the context of the analysis described in [33] and [1], more variables and combinations of
436 them were considered. The final outcome was that the gain in using more jet substructure vari-
437 ables at the same time is limited, at the price of a more complicated procedure, hence favoring
438 the usage of only τ_{21} for this version of the analysis.

439 **5.2.3 Selection criteria**

440 At this stage of the analysis, $W \rightarrow q\bar{q}'$ candidates are simply identified by CA8 jets passing
441 the selection criteria described in Section 4.4.1. No requirements on τ_{21} or m_{jet}^{pruned} are applied
442 here. The final cuts on the pruned jet mass and τ_{21} are optimized at the stage of the final event
443 selection. The pruned jet mass is also used to define control regions i) to cross check the MC
444 description of the data and ii) for the data-driven background estimation as described later.

445 **5.3 Reconstruction of the WW resonance**

446 We select events with one reconstructed `tight` (see Section 4.3) electron (or muon). We reject
447 events with two or more `loose` (see Section 4.3) electron or muons, since in signal events we
448 expect only one charged lepton. Thus, in each event there is only one reconstructed W leptonic
449 candidate. The W hadronic candidate is identified by the jet with the largest p_T in the event.

450 The WW resonance, a single candidate for each event, is then built from the leptonic W and the
451 hadronic W defined above.

DRAFT

452 6 Event Selection

453 6.1 Kinematic criteria

454 We consider two final states in this analysis: $X \rightarrow WW \rightarrow \ell\nu q\bar{q}'$, where $\ell = e$ or μ . We report
 455 below the final event selection for these two channels:

- 456 1. **one charged lepton**: at least one reconstructed `tight` (see Section 4.3) electron or muon;
- 457 2. **second lepton veto**: no additional `loose` (see Section 4.3) electron or muon;
- 458 3. **missing transverse energy**: an additional requirement $E_T^{\text{miss}} > 80$ GeV for the electron
 459 channel (40 GeV for the muon channel). This cut is increased with respect to the pre-
 460 selection in order to further suppress QCD-like background in the electron channel (QCD
 461 dijet events and γ +jet events);
- 462 4. **leptonic W p_T** : the p_T of the reconstructed leptonic W must be greater than 200 GeV. This
 463 cut is required to select the boosted W topology;
- 464 5. **hadronic W p_T** : the p_T of the reconstructed hadronic W (CA8 leading jet) must be greater
 465 than 200 GeV. This cut is required to select the boosted W topology;
- 466 6. **btag veto**: the event is required to have no AK5 jets (passing the selection in Section 4.4)
 467 tagged as b-jet accordingly with the CSV algorithm, medium working point (CSVM). The choice of the algorithm and the working point has been optimized, as described in
 468 Appendix A.
- 470 7. **dijet-like topology**: $\Delta R(\ell, W_{\text{had}}) > \pi/2$, $\Delta\phi(W_{\text{had}}, E_T^{\text{miss}}) > 2$, $\Delta\phi(W_{\text{had}}, W_{\text{lep}}) > 2$;

471 Furthermore, we define three orthogonal samples of events using the reconstructed pruned jet
 472 mass of the W hadronic candidate (leading CA8 jet):

- 473 • **Signal region**: $65 < m_{\text{jet}}^{\text{pruned}} < 105$ GeV;
- 474 • **Low-mass sideband region**: $40 < m_{\text{jet}}^{\text{pruned}} < 65$ GeV;
- 475 • **High-mass sideband region**: $105 < m_{\text{jet}}^{\text{pruned}} < 130$ GeV;

476 The signal region window was not optimized precisely for this analysis, we rather based our se-
 477 lection on what was done in a previous analysis targeting lower $X \rightarrow WW$ resonance masses [43].
 478

479 6.2 Control plots in the hadronic W mass sideband

480 We use the low and high sideband regions (events with $40 < m_{\text{jet}}^{\text{pruned}} < 65$ GeV or $105 <$
 481 $m_{\text{jet}}^{\text{pruned}} < 130$ GeV) to form a sample of data events with small signal contamination from a po-
 482 tential new $X \rightarrow WW$ resonance. In this control sample we compare data and MC simulation.
 483 All the backgrounds (except W+jets) are normalized to the cross sections described in Sec-
 484 tion 3.2. Instead, the W+jets cross section described in Section 3.2 is multiplied by a factor 1.475
 485 (1.460) for the muon (electron) channel, such that the total integral of the MC matches the num-
 486 ber of events in data in the low+high sideband region. For W+jets, the `WJetsToLNu_PtW-100-`
 487 `TuneZ2star_8TeV-madgraph` MC sample is used. Various distributions are shown for both
 488 electron and muon channels:

- Figure 7: lepton p_T and η ;
- Figure 8: E_T^{miss} and $p_{z\nu}$ (defined in Section 5.1);
- Figure 9: leptonic $W p_T$ and $m_T(\ell\nu)$;
- Figure 10: hadronic $W p_T$ and η ;
- Figure 11: $m_{\text{jet}}^{\text{pruned}}$ and τ_{21} ;
- Figure 12: m_{WW} (using the $p_{z\nu}$ defined in Section 5.1);

The agreement between data and simulation is not perfect, although the main kinematic features of these events in data are correctly described by the MC. The discrepancies are related to the fact that we select a sample dominated by $W+\text{jets}$ events in the tails of the pruned jet mass distribution events which is a quantity particularly sensitive to the parton shower algorithm and in general not well described by the simulation; in these events, the hadronic W candidate is in fact a massive jet originated from a light quark and not a real W decaying to pair of quarks. In this analysis we rely on data-driven methods for the $W+\text{jets}$ background estimation which allow to reduce this dependence on the final background prediction (see Section 8).

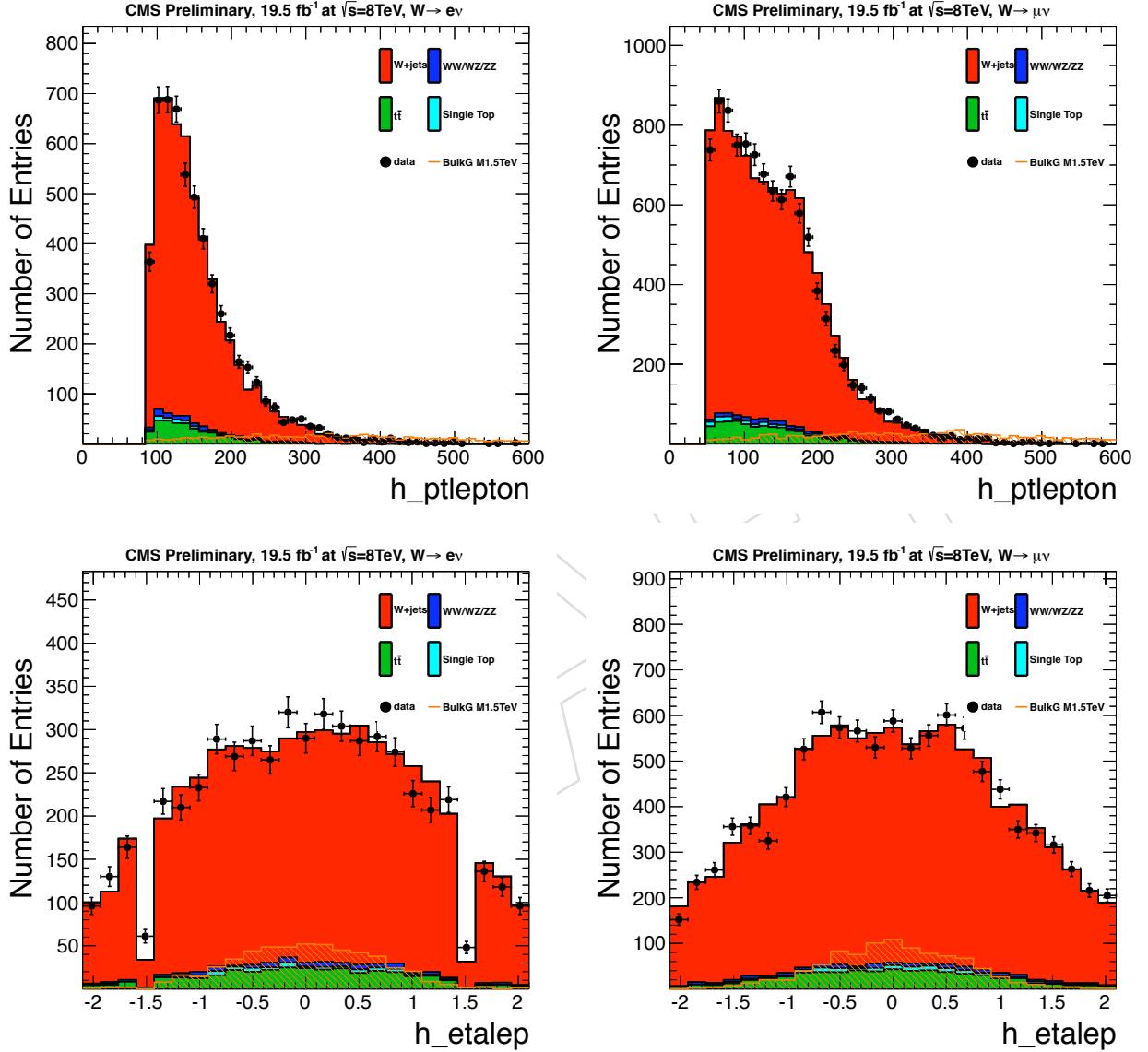


Figure 7: Lepton p_T and η for electron channel (left) and muon channel (right) for events with $40 < m_{\text{jet}}^{\text{pruned}} < 65 \text{ GeV}$ or $105 < m_{\text{jet}}^{\text{pruned}} < 130 \text{ GeV}$. The signal is scaled by 60000 (34800) for the muon (electron) channel.

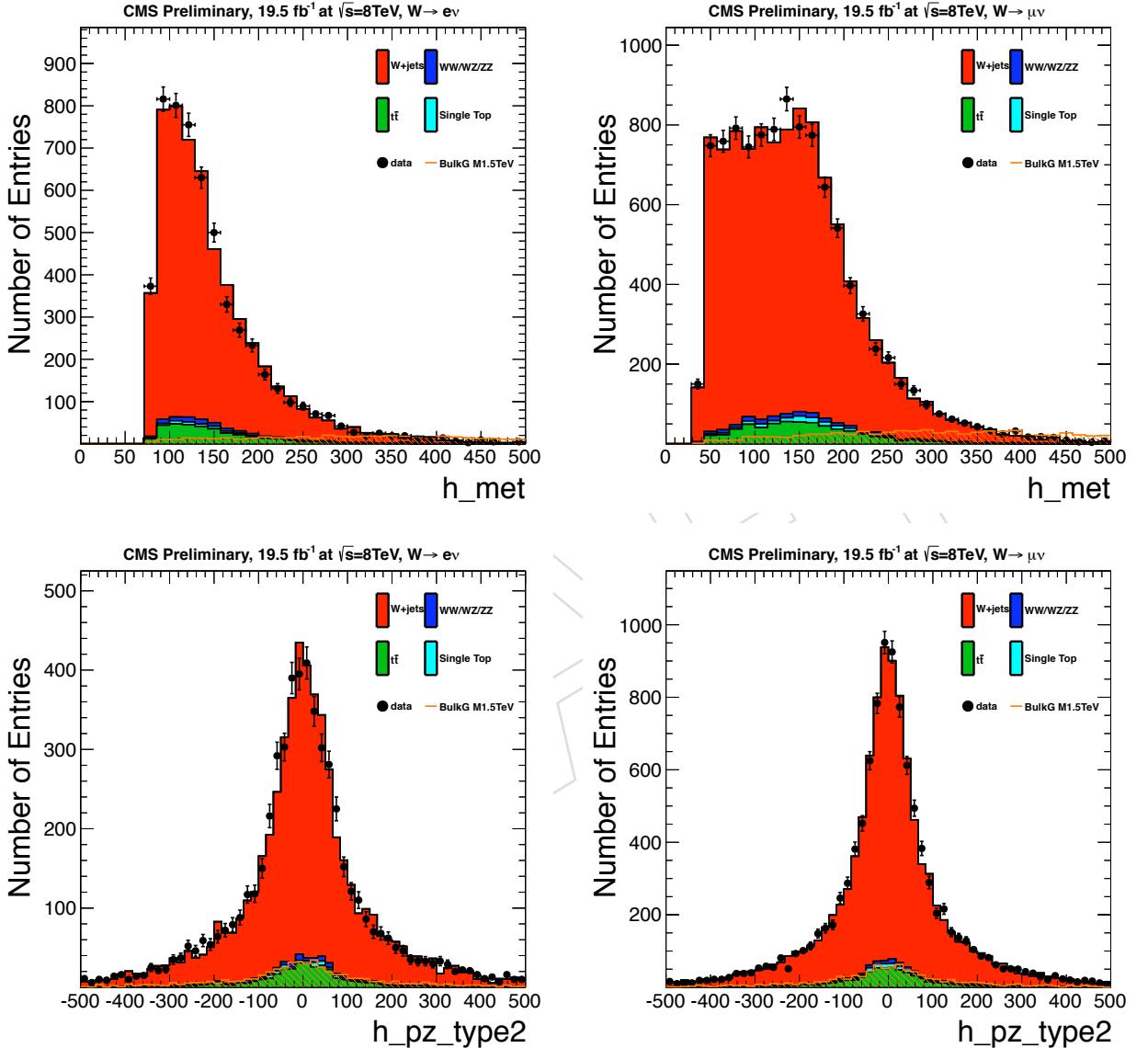


Figure 8: E_T^{miss} and $p_{z\nu}$ (defined in Section 5.1) for electron channel (left) and muon channel (right) for events with $40 < m_{\text{jet}}^{\text{pruned}} < 65$ GeV or $105 < m_{\text{jet}}^{\text{pruned}} < 130$ GeV. The signal is scaled by 60000 (34800) for the muon (electron) channel.

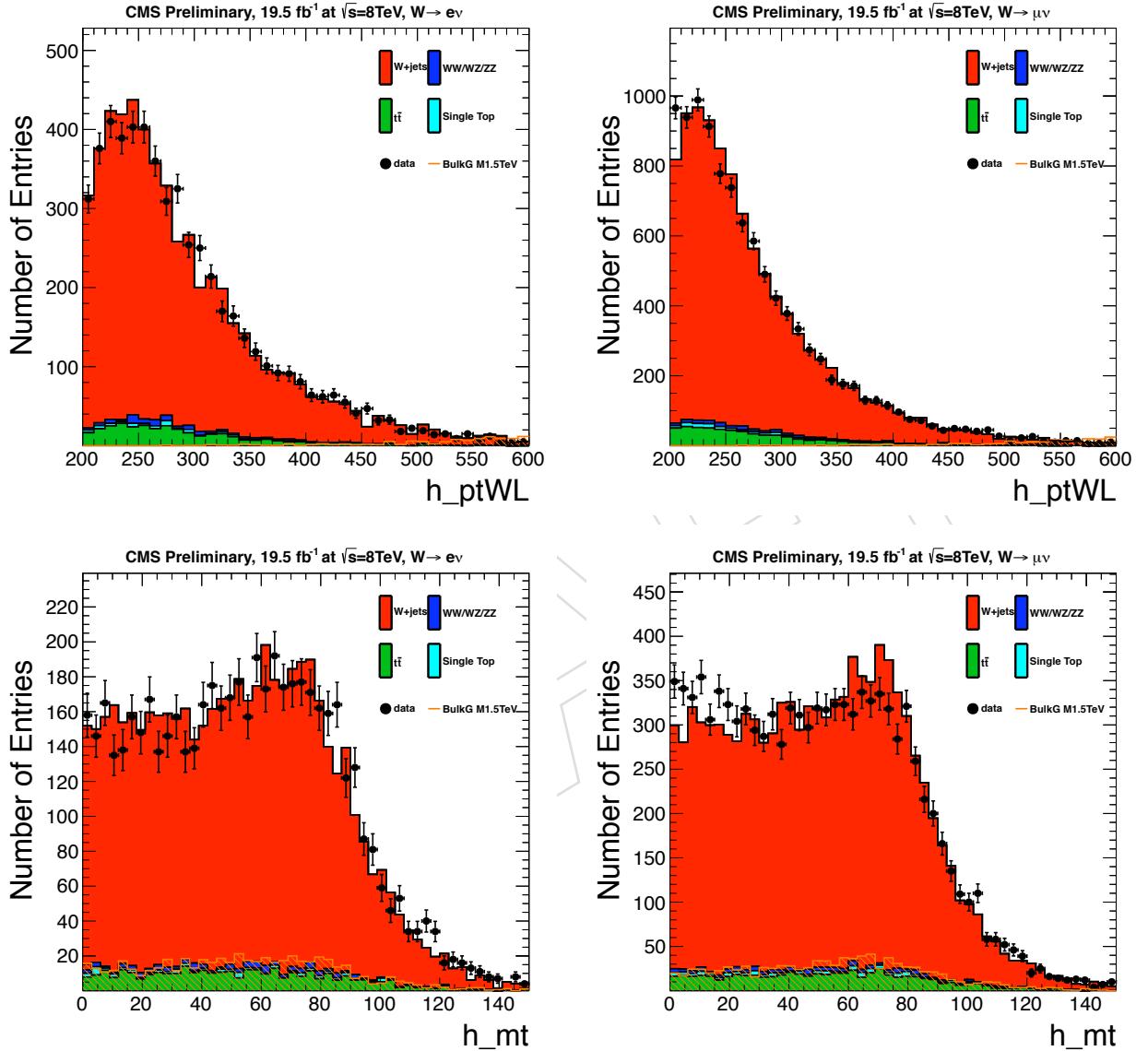


Figure 9: Leptonic W p_T and $m_T(\ell\nu)$ for electron channel (left) and muon channel (right) for events with $40 < m_{jet}^{pruned} < 65$ GeV or $105 < m_{jet}^{pruned} < 130$ GeV. The signal is scaled by 60000 (34800) for the muon (electron) channel.

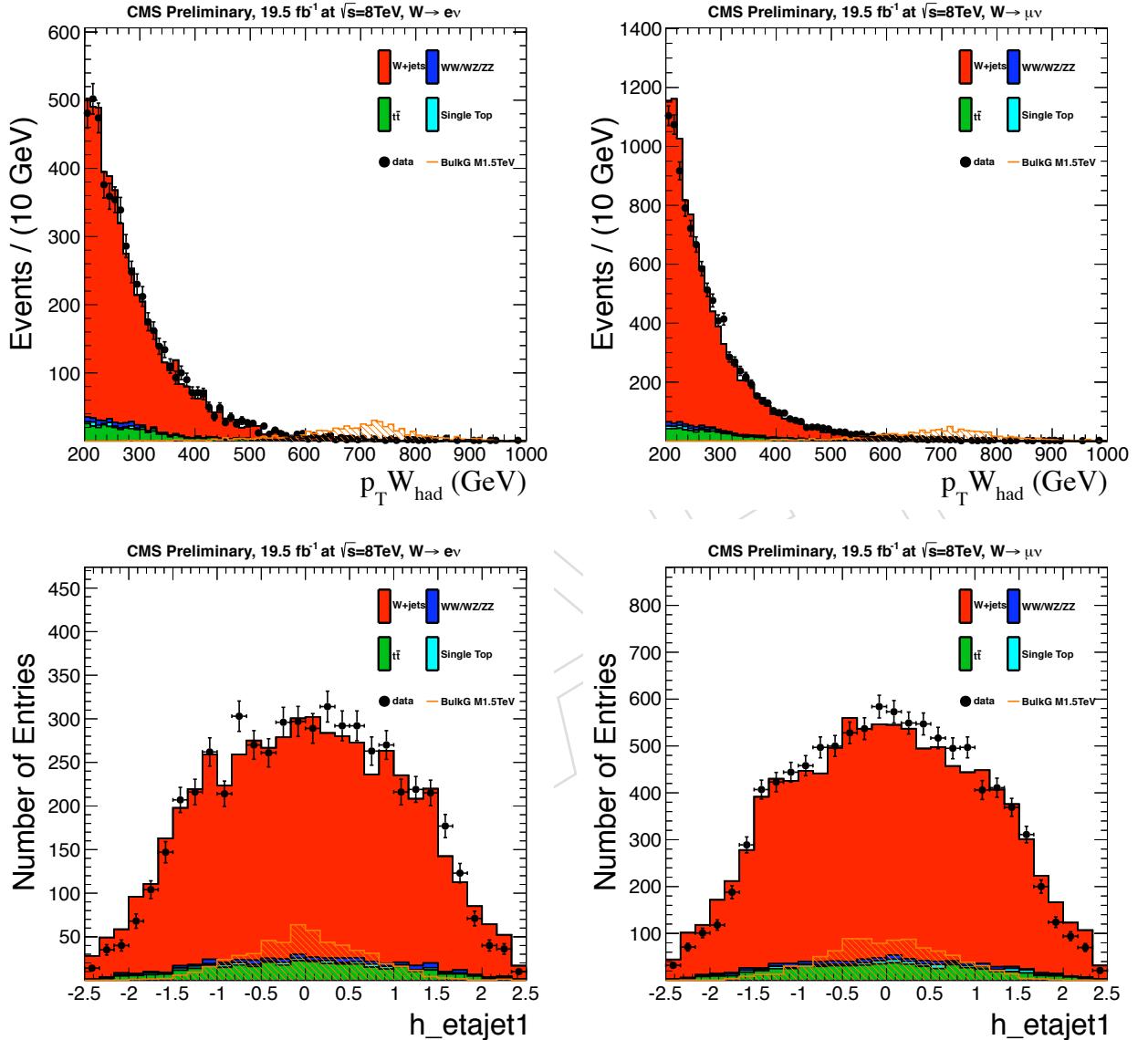


Figure 10: Hadronic W p_T and η for electron channel (left) and muon channel (right) for events with $40 < m_{jet}^{pruned} < 65$ GeV or $105 < m_{jet}^{pruned} < 130$ GeV. The signal is scaled by 60000 (34800) for the muon (electron) channel.

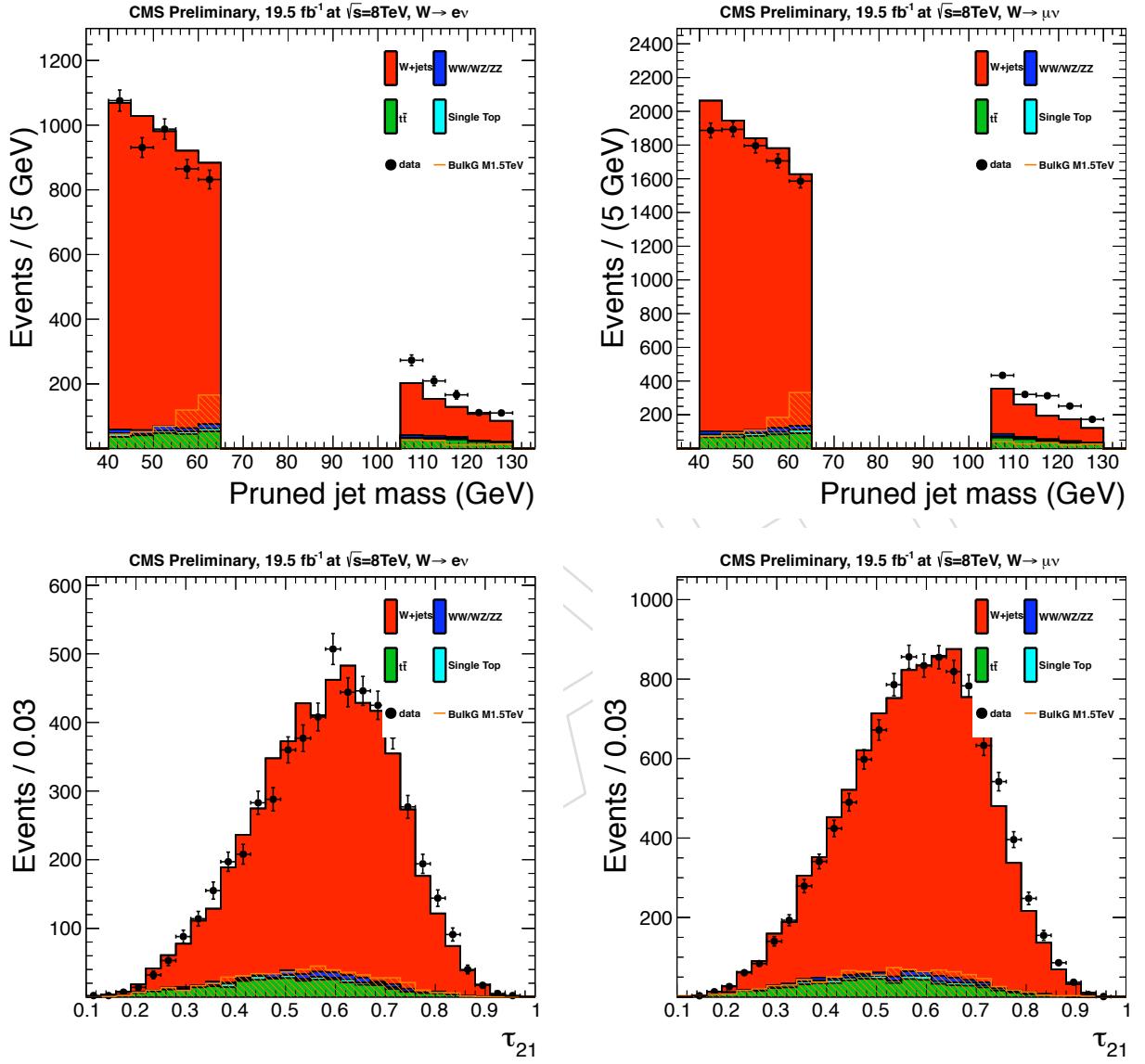


Figure 11: m_{jet}^{pruned} and τ_{21} for electron channel (left) and muon channel (right) for events with $40 < m_{jet}^{pruned} < 65$ GeV or $105 < m_{jet}^{pruned} < 130$ GeV. The signal is scaled by 60000 (34800) for the muon (electron) channel.

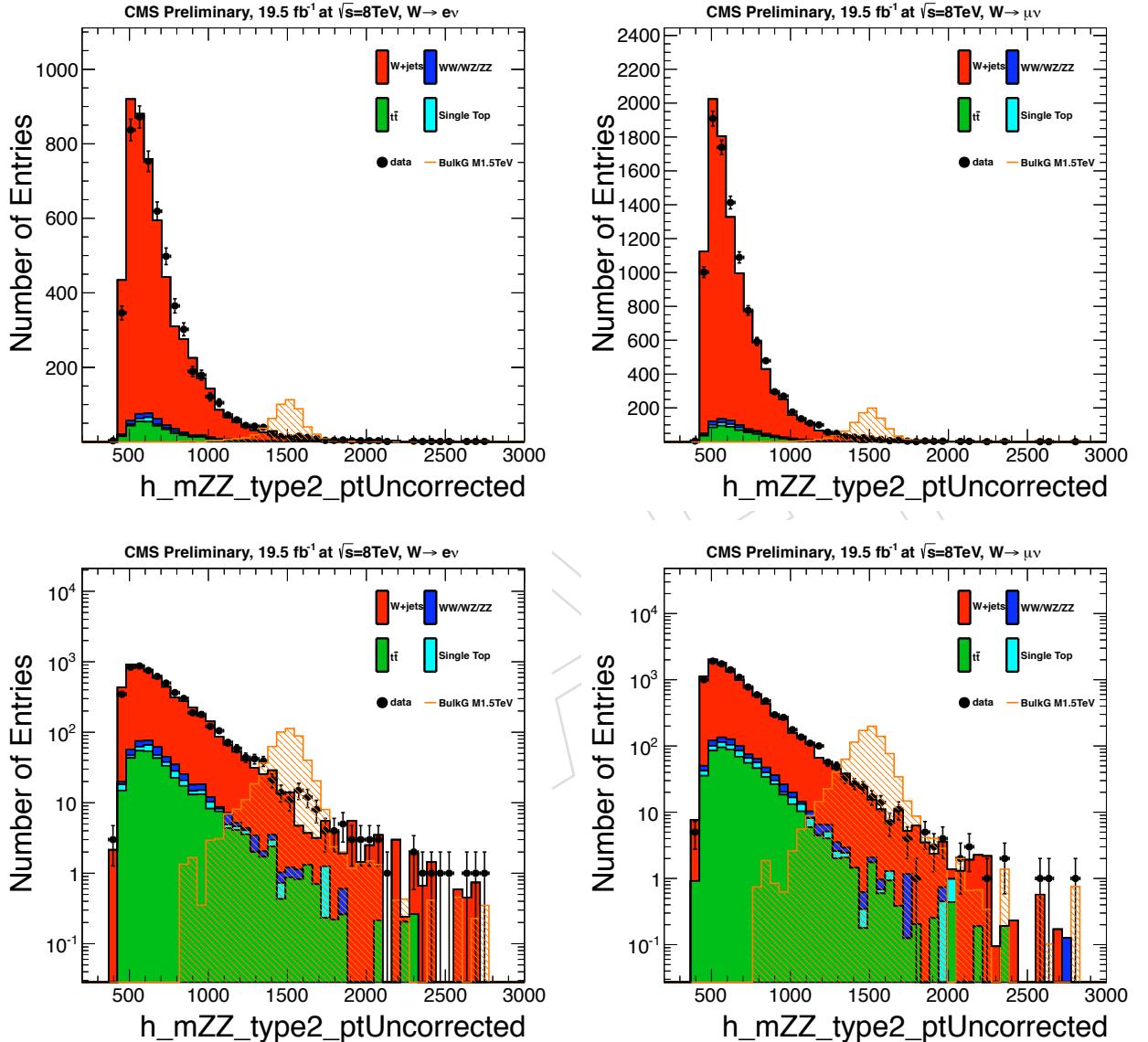


Figure 12: m_{WW} (using the $p_{\tau\tau}$ defined in Section 5.1) in linear and log scale for electron channel (left) and muon channel (right) for events with $40 < m_{jet}^{pruned} < 65$ GeV or $105 < m_{jet}^{pruned} < 130$ GeV. The signal is scaled by 60000 (34800) for the muon (electron) channel.

503 **6.3 Optimization of the τ_{21} selection**

504 The 2- to 1-subjettiness ratio, τ_{21} , is a key variable able to discriminate between jets coming
 505 from hadronic decays of boosted W and QCD. In this section we study the optimization of
 506 the selection on this variable on top of the selection criteria listed in Section 6.1. Although the
 507 variable has already lost a part of its separation power due to the previous requirement on the
 508 jet pruned mass (signal region: $65 < m_{jet}^{pruned} < 105$ GeV), there is still a degree of separation in
 509 between signal and background. We choose as a figure of merit the quantity suggested in Ref.
 510 [44] (“Punzi’s significance” with $a = b = 2$, from Eq. 7 of the reference cited):

$$P = \frac{\epsilon_S}{1 + \sqrt{B}} \quad (13)$$

511 where ϵ_S is the signal selection efficiency and B is the remaining background. We perform the
 512 following optimization procedure:

- 513 • set a window of $\pm 15\%$ around the target resonance mass;
- 514 • histogram the τ_{21} variable for signal events (for the chosen mass) and for background
 515 events which pass all the final selection requisites including the signal region m_{jet}^{pruned}
 516 cut and also have a reconstructed resonances mass m_{WW} in the window defined
 517 above.
- 518 • set an upper threshold τ_{21}^{\max} and integrate the τ_{21} distributions of signal and back-
 519 ground up to the threshold. The value obtained for the signal sample is proportional
 520 to the signal efficiency ϵ_S , while the value obtained for the background sample is an
 521 estimate of B ;
- 522 • calculate the significance P .

523 The procedure above is repeated for values of τ_{21}^{\max} ranging from 0.05 from 0.95, in steps of
 524 0.05. The W+jets sample `WJetsToLNu_PtW-180_TuneZ2star_8TeV-madgraph-tarball`
 525 is used to increase the MC statistics at high m_{WW} values. The W+jets background is normalized
 526 using the same scale factors introduced in the previous section. Since the sensitivities to new
 527 physics of the electron and muon channels are quite similar, we merge the two samples for this
 528 optimization study (consistent conclusions are obtained if one performs the optimization for
 529 the two channels separately).

530 The results are shown in Figs. 13 , 14, 15, and 16 for target resonance masses of $M_G = 1000, 1500,$
 531 2000, and 2500 GeV. It can be seen how there is a lack of statistics in the simulated background
 532 samples for the case of 2.5 TeV resonance mass. In general, a flat threshold of $\tau_{21}^{\max} = 0.5$ is
 533 a reasonable choice. With this flat cut, the analysis is optimal over a broad range of nominal
 534 masses, while the loss in performances at high masses is small and hard to quantify because
 535 of the limited MC statistics. Furthermore, adopting a m_{WW} independent selection avoids any
 536 possible biases and sculpting on the m_{WW} distribution itself, simplifying the extraction of the
 537 background. We show the variation of the optimized threshold on τ_{21} with the signal mass in
 538 Fig. 17, along with the difference in the Punzi figure of merit when choosing the flat $\tau_{21} < 0.5$
 539 requirement as opposed to the optimized one. Above 2 TeV of resonance, the optimization
 540 seems to suggest a gradually looser cut on τ_{21} . This is due to the loss of signal efficiency when
 541 the W is very boosted and the quarks are very close in space; in this regime genuine signal
 542 events look like more and more similar to background events. The inefficiency of the $\tau_{21} < 0.5$
 543 at high mass is recovered by introducing a low-purity category as described below. It is also
 544 important to note that the signal distributions are essentially depopulated for values of τ_{21}

545 above 0.75.

546 Based on these considerations, we classify events in the signal region in two categories:

- 547 • **High-Purity (HP) category:** events in the signal region, with τ_{21} in the $[0, 0.50]$ range.
- 548 • **Low-Purity (LP) category:** events in the signal region, with τ_{21} in the $[0.50, 0.75]$ range.

550 Given the presence of two categories, the precise choice of the τ_{21} threshold for the HP-LP
 551 transition (i.e. a uniform 0.5 for all the resonance masses rather than a mass-dependent cut) is
 552 less important for the final results than in the case when only one category is considered.

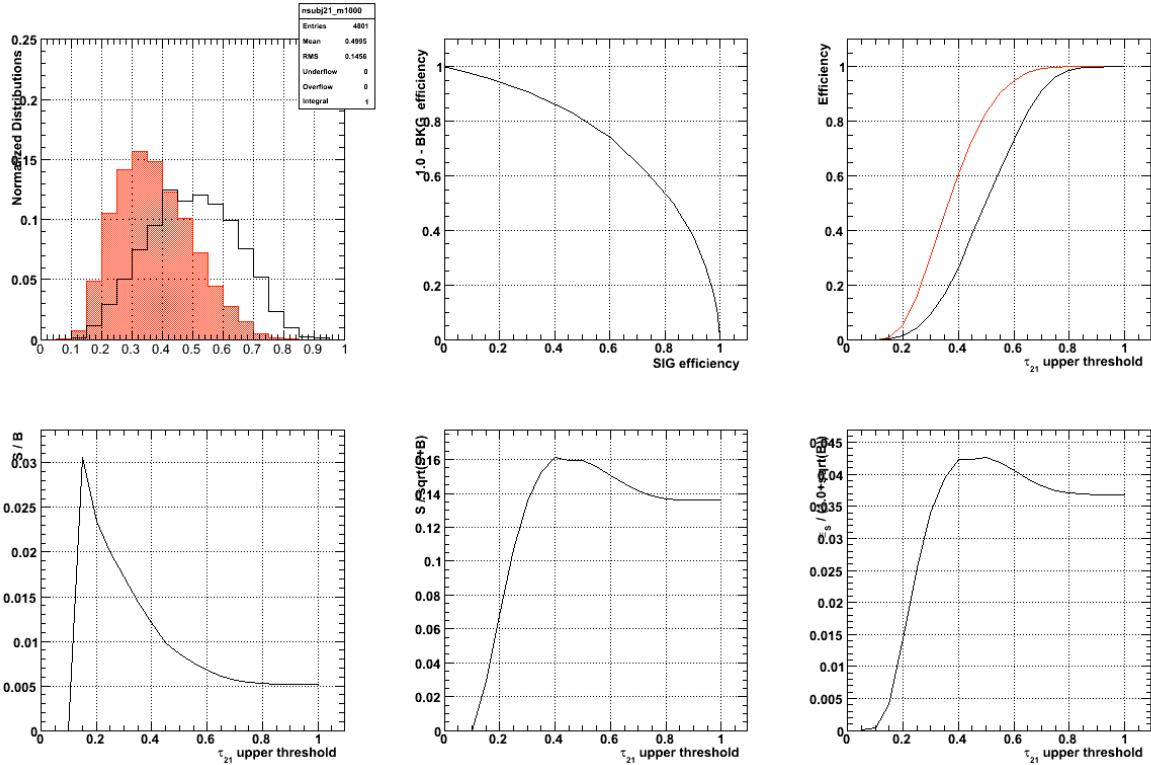


Figure 13: Steps of optimization for $M_G = 1000$ GeV. Top left: distributions of τ_{21} for signal and background. Top middle: signal efficiency \times background rejection curve (ROC curve). Top right: signal (red) and background (black) efficiencies as function of τ_{21} cut. Bottom left: signal S over background B as function of τ_{21} cut. Bottom middle: $S/\sqrt{S+B}$ as function of τ_{21} cut. Bottom right: Punzi's significance (as defined in Eq.13) as function of τ_{21} cut.

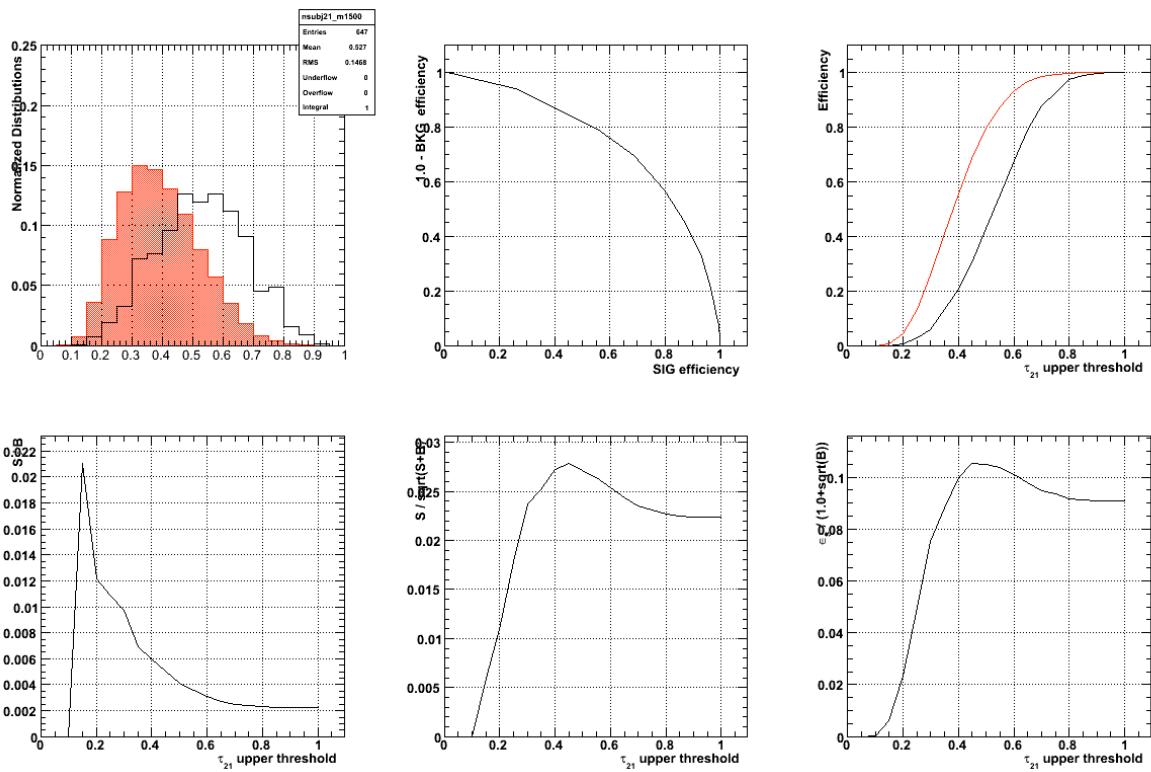


Figure 14: Steps of optimization for $M_G = 1500$ GeV. Top left: distributions of τ_{21} for signal and background. Top middle: signal efficiency \times background rejection curve (ROC curve). Top right: signal (red) and background (black) efficiencies as function of τ_{21} cut. Bottom left: signal S over background B as function of τ_{21} cut. Bottom middle: $S / \sqrt{S+B}$ as function of τ_{21} cut. Bottom right: Punzi's significance (as defined in Eq.13) as function of τ_{21} cut.

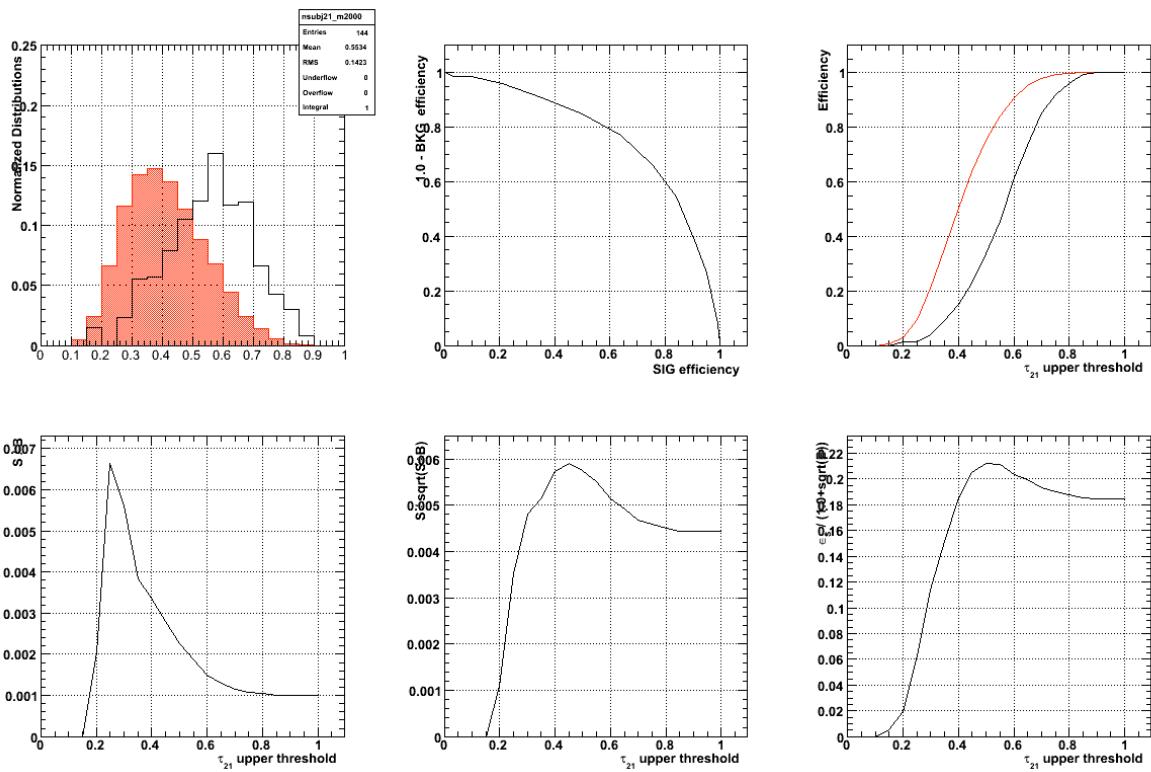


Figure 15: Steps of optimization for $M_G = 2000$ GeV. Top left: distributions of τ_{21} for signal and background. Top middle: signal efficiency \times background rejection curve (ROC curve). Top right: signal (red) and background (black) efficiencies as function of τ_{21} cut. Bottom left: signal S over background B as function of τ_{21} cut. Bottom middle: $S/\sqrt{S+B}$ as function of τ_{21} cut. Bottom right: Punzi's significance (as defined in Eq.13) as function of τ_{21} cut.

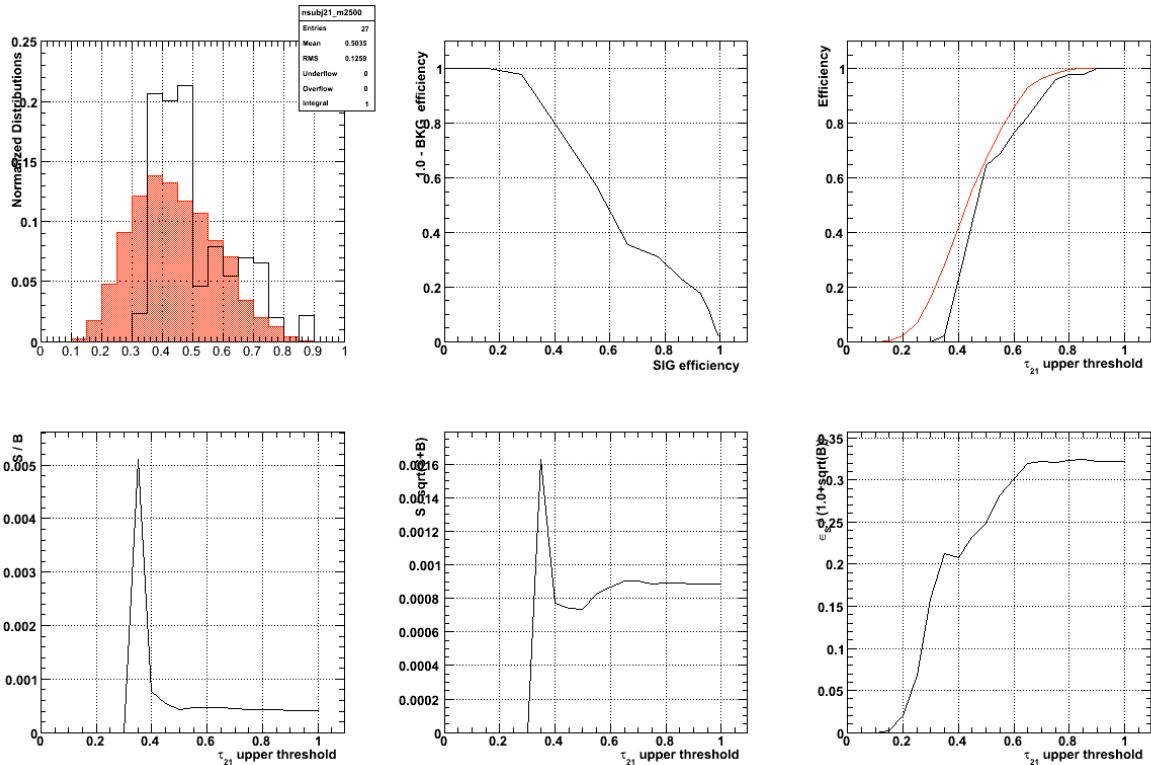


Figure 16: Steps of optimization for $M_G = 2500$ GeV. Top left: distributions of τ_{21} for signal and background. Top middle: signal efficiency \times background rejection curve (ROC curve). Top right: signal (red) and background (black) efficiencies as function of τ_{21} cut. Bottom left: signal S over background B as function of τ_{21} cut. Bottom middle: $S/\sqrt{S+B}$ as function of τ_{21} cut. Bottom right: Punzi's significance (as defined in Eq.13) as function of τ_{21} cut.

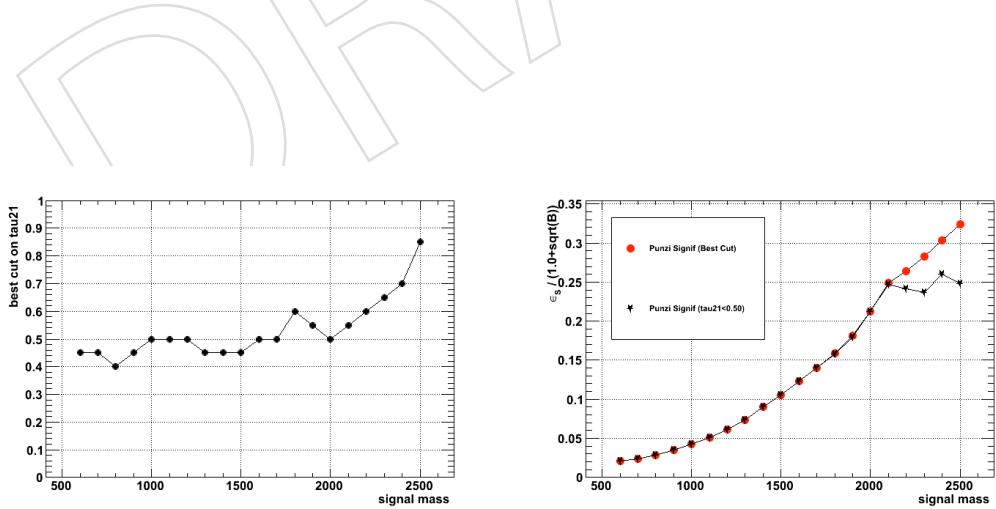


Figure 17: Left: variation of the τ_{21} cut with signal mass. Right: comparison of the fully optimized (“best cut”) and flat ($\tau_{21} < 0.5$) thresholds for the Punzi’s figure of merit used for the optimization.

553 **6.4 Summary of final selection**

554 A summary of the final kinematic and jet substructure selection criteria is presented in Table 4.

Table 4: Summary of final selection.

Selection	Value	Comments
Tight Lepton selection		
Electron p_T	$p_T > 90 \text{ GeV}$	
Muon p_T	$p_T > 50 \text{ GeV}$	
Electron η	$ \eta _{\text{SC}} < 2.5$ except [1.4442, 1.566] range	Avoid the ECAL gap.
Muon η	$ \eta < 2.1$	
Loose Lepton selection		
Electron p_T	$p_T > 35 \text{ GeV}$	
Muon p_T	$p_T > 20 \text{ GeV}$	
Electron η	$ \eta _{\text{SC}} < 2.5$ except [1.4442, 1.566] range	Avoid the ECAL gap.
Muon η	$ \eta < 2.4$	
CA8 jet selections		
Jet p_T	$p_T > 80 \text{ GeV}$	
Jet η	$ \eta < 2.4$	Used for hadronic W reconstruction
AK5 jet selections		
Jet p_T	$p_T > 30 \text{ GeV}$	
Jet η	$ \eta < 2.4$	Used for b-tag jet selection
E_T^{miss} selections		
E_T^{miss} (electron ch.)	$E_T^{\text{miss}} > 80 \text{ GeV}$	
E_T^{miss} (muon ch.)	$E_T^{\text{miss}} > 40 \text{ GeV}$	
Boson selections		
Pruned jet mass (signal)	$65 < m_{\text{jet}}^{\text{pruned}} < 105 \text{ GeV}$	
Pruned jet mass (low-mass sideband)	$40 < m_{\text{jet}}^{\text{pruned}} < 65 \text{ GeV}$	
Pruned jet mass (high-mass sideband)	$105 < m_{\text{jet}}^{\text{pruned}} < 130 \text{ GeV}$	
Leptonic W p_T	$p_T > 200 \text{ GeV}$	
Hadronic W p_T	$p_T > 200 \text{ GeV}$	
Back-to-back topology	$\Delta R(\ell, W_{\text{had}}) > \pi/2$, $\Delta\phi(W_{\text{had}}, E_T^{\text{miss}}) > 2$ $\Delta\phi(W_{\text{had}}, W_{\text{lep}}) > 2$	
Veto		
Number of loose electrons	0	in addition to tight lepton
Number of loose muons	0	in addition to tight lepton
Number of b-tag jets	0	CVS medium working point
Diboson selections		
2- to 1-subjettiness ratio (high purity)	$\tau_{21} < 0.50$	
2- to 1-subjettiness ratio (low purity)	$0.50 \leq \tau_{21} < 0.75$	

555 **6.5 Data vs MC comparison in signal region**

556 Figure 18 show the m_{WW} distribution in data after full selection compared with the MC pre-
557 diction. The 4 analysis categories are shown separately. In this specific set of plots, the W+jets
558 background have been scaled in order that the total background would match the number
559 of events in data. The data/MC scale factors for W+jets are (eleLP,eleHP,muLP,muHP) =
560 (1.67,1.25,1.73,1.42). All the other backgrounds are fixed to the cross section values specified
561 in Section 3.2

562 No significant excess of data above this MC background prediction is visible across the entire
563 WW mass spectra.

564 It should be noted that the final results of this analysis are not derived from these MC back-
565 ground predictions. The final background estimate is obtained using two different data-driven
566 methods described in Section 8.1 and Section 8.2, respectively.

567 Finally, we report in Figure 19 a few control plots for electron and muon channel (all purity
568 categories) in the signal region only. Some of these plots will be included in the PAS.

DRAFT

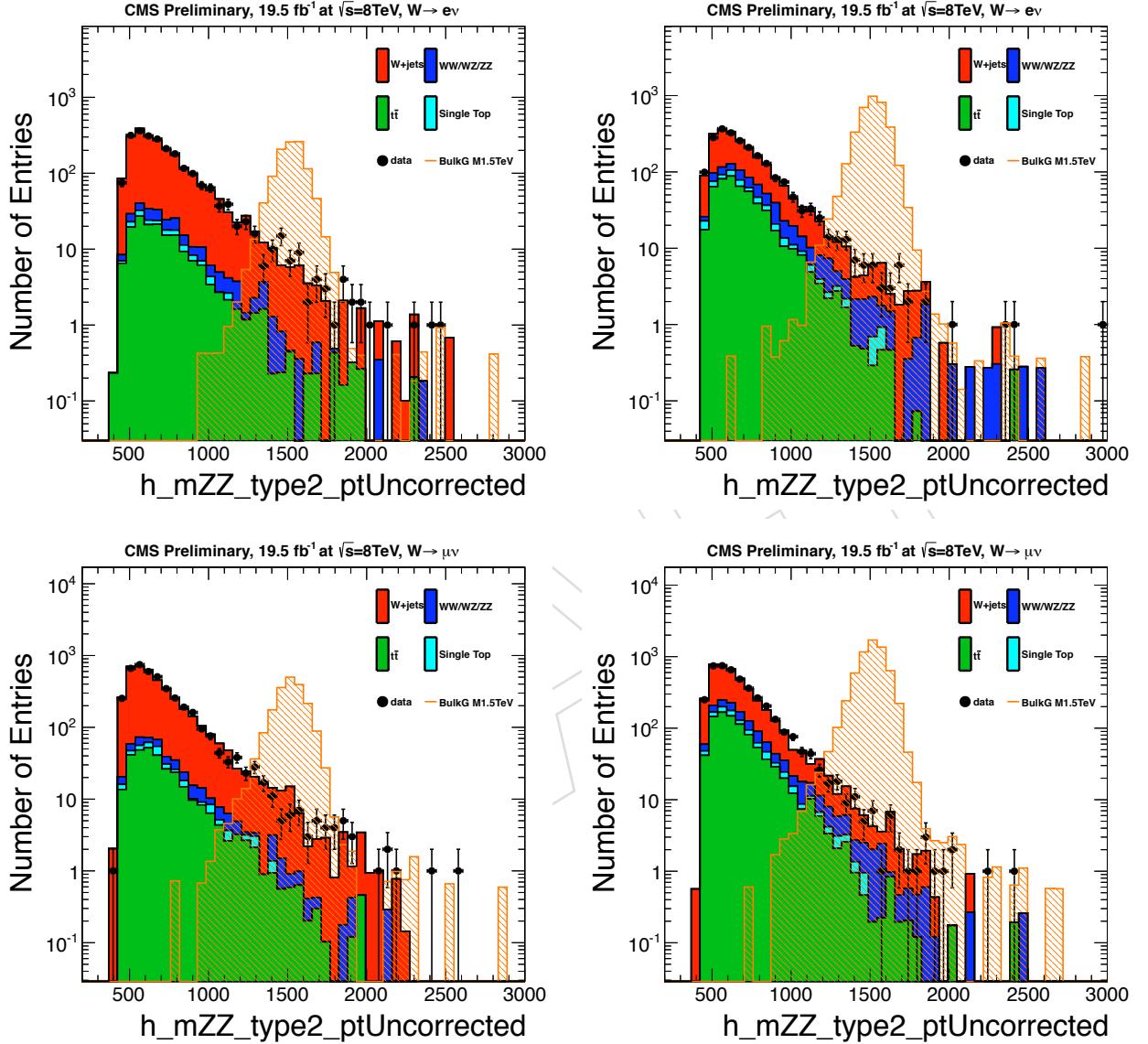


Figure 18: m_{WW} (using the $p_{z\nu}$ defined in Section 5.1) for electron LP (top-left), electron HP (top-right), muon LP (bottom-left), muon HP (bottom-right) for events passing the full selection with $65 < m_{jet}^{pruned} < 105$ GeV. The signal is scaled by 60000 (34800) for the muon (electron) channel.

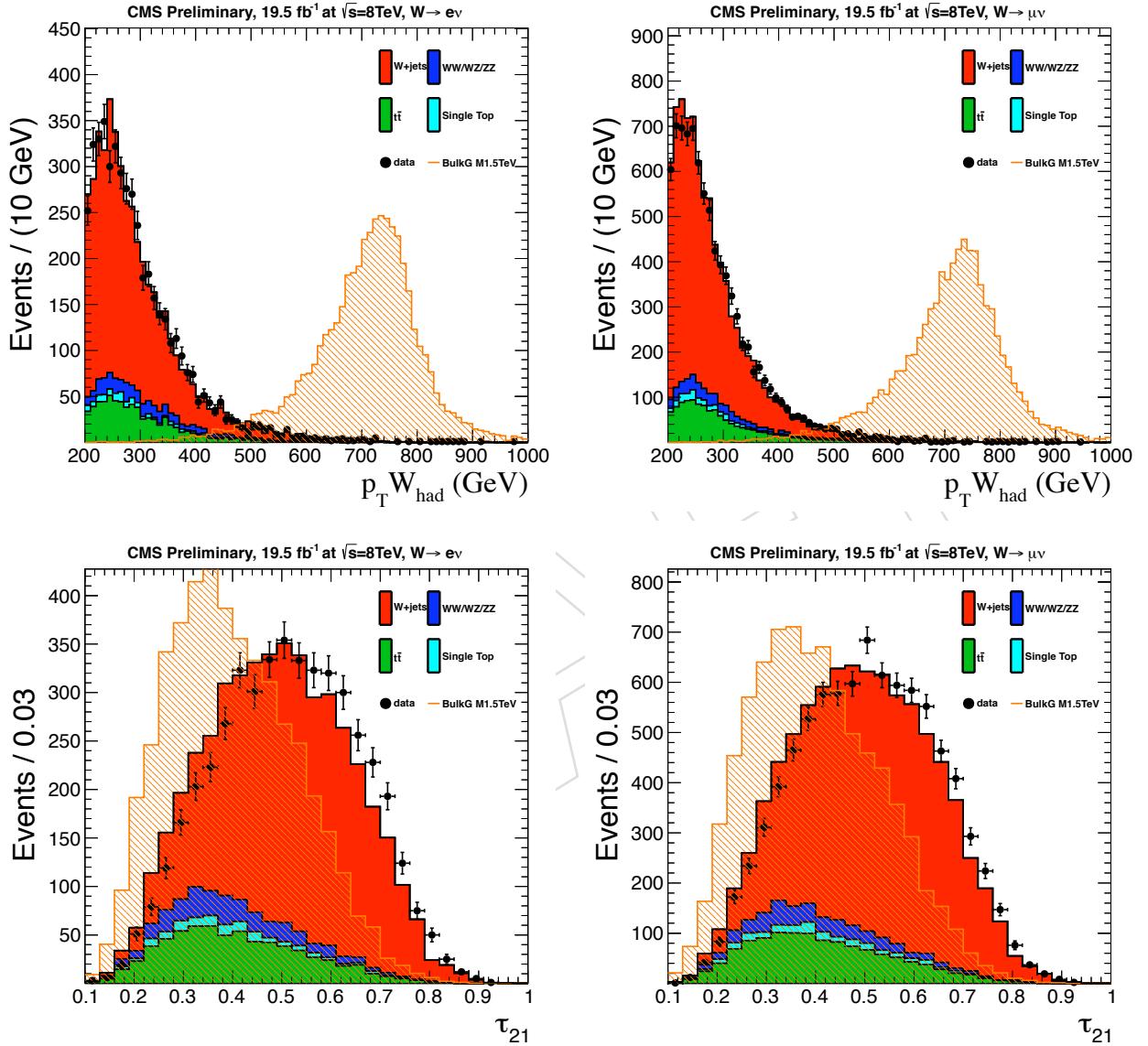
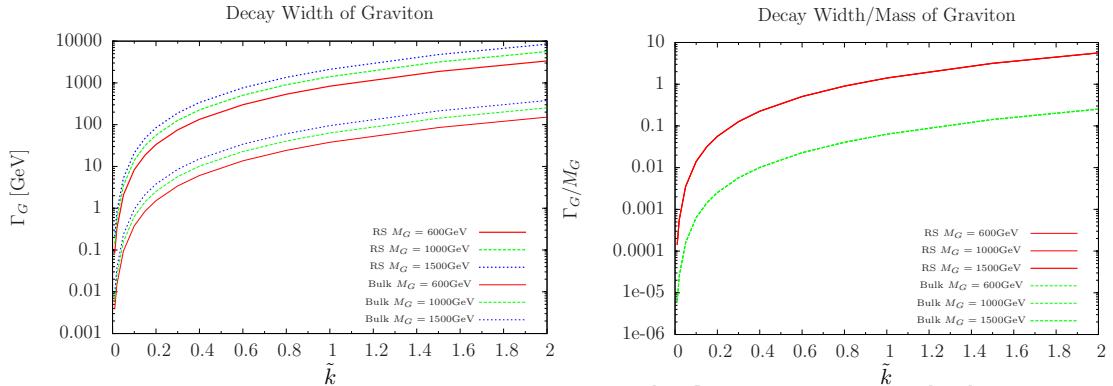


Figure 19: Hadronic W p_T and τ_{21} for electron channel (left) and muon channel (right) for events with $65 < m_{\text{jet}}^{\text{pruned}} < 105$ GeV. The signal is scaled by 60000 (34800) for the muon (electron) channel.

569 7 Signal modeling: m_{WW} shapes and efficiency

570 7.1 Parametrization of the m_{WW} signal shape

571 We used the bulk graviton samples with $\tilde{k} = 0.2$ listed in Sect. 3.1 for modeling the shape of the
 572 reconstructed m_{WW} distribution of the signal. Fig.20 shows the dependence of the width of the
 573 graviton line shape at the parton level (taken from [20]). The natural width of the resonance
 574 obtained using the chosen value of \tilde{k} is sufficiently small to be neglected when compared to
 575 the detector resolution. This makes our modeling of the detector effects on the signal shape
 576 independent of the actual model used for generating the events and allows us to apply it to a
 577 model-independent search for narrow-resonances.



578 Figure 20: Dependence of the natural width of the graviton (GeV) as a function of the coupling
 579 constant \tilde{k} in the RS1 and Bulk graviton models (left). The relative width (width divided by the
 580 mass) is shown on the right.

578 The m_{WW} spectrum of the signal is modeled with a Crystal-Ball function (i.e. a Gaussian core
 579 with powerlaw tails on both sides, `RooDoubleCB` implemented in `RooFit`). The m_{WW} signal
 580 distribution varies depending on the resonance mass hypothesis. Several examples are shown
 581 in Fig. 21 for electron and muon channels independently on the V-tag purity.

582 A summary of the natural width and the standard deviation of the Gaussian core of the matched
 583 part of the signal is presented for several mass points in Table 5 for the electron and muon
 584 channels. The electron and muon shapes are very similar, even at high m_{WW} , since the mass
 585 resolution is dominated anyway by the E_T^{miss} contribution.

586 No significant dependence in the m_{WW} shape on the V-tag purity category has been found in
 587 our studies.

Resonance mass [GeV]	Total Natural Width, Γ [GeV]	Mean of CB function [GeV]	σ of CB function [GeV]
electron channel			
1000	2.5 (13.0)	1021	61
1500	3.75 (19.5)	1522	81
2000	5.0 (26.0)	2027	99
2500	6.25 (32.5)	2526	110
muon channel			
1000	2.5 (13.0)	1017	64
1500	3.75 (19.5)	1521	83
2000	5.0 (26.0)	2026	102
2500	6.25 (32.5)	2524	115

Table 5: The main parameters (mass/mean and width/ σ) of the signal shape for different mass points at the generator and reconstructed level for the electron and muon channels. The total natural width, Γ , refers to bulk graviton samples with $\tilde{k} = 0.2$ (or $\tilde{k} = 0.5$ in parenthesis) and is taken from [20]. The MC samples used for fitting the signal shape parameters are the ones described in Sect. 3.1 (i.e. generated with $\tilde{k} = 0.2$).

588 7.2 Signal efficiency

589 The signal reconstruction efficiency in each of the 4 channels (electron HP/LP, muon HP/LP)
 590 is defined as the ratio:

$$\varepsilon_{SIG} = \frac{\text{Nr. of events passing the full selection in that category}}{\text{Total nr. of events generated}} \quad (14)$$

591 where the total number of events generated includes both electron and muon decays
 592 ($X \rightarrow WW \rightarrow \ell\nu q\bar{q}'$ events with $\ell = e$ or μ , no τ decays).

593 The signal efficiencies (as defined above) for the final selection criteria reported in Section 6.4
 594 (signal region, i.e. $65 < m_{jet}^{pruned} < 105$ GeV) are summarized in Table 6.

595 The efficiency is parameterized as a function of the nominal graviton mass and it is interpolated
 596 to all the mass points considered in this search. The efficiency parameterization is shown in
 597 Fig. 22.

598 The HP efficiency drops at high values of the resonance mass mainly due to the inefficiency of
 599 the $\tau_{21} < 0.50$ cut. This inefficiency is partially recovered in the LP category ($0.50 \leq \tau_{21} < 0.75$).

600 At high mass also the m_{jet}^{pruned} reduces the efficiency due to a degradation of the pruned jet mass
 601 distribution (wider and with more low-mass tails). The rest of the overall inefficiency is due to
 602 the kinematic requirements on leptons, E_T^{miss} , jets, and bosons, as well as on the lepton ID and
 603 ISO requirements.

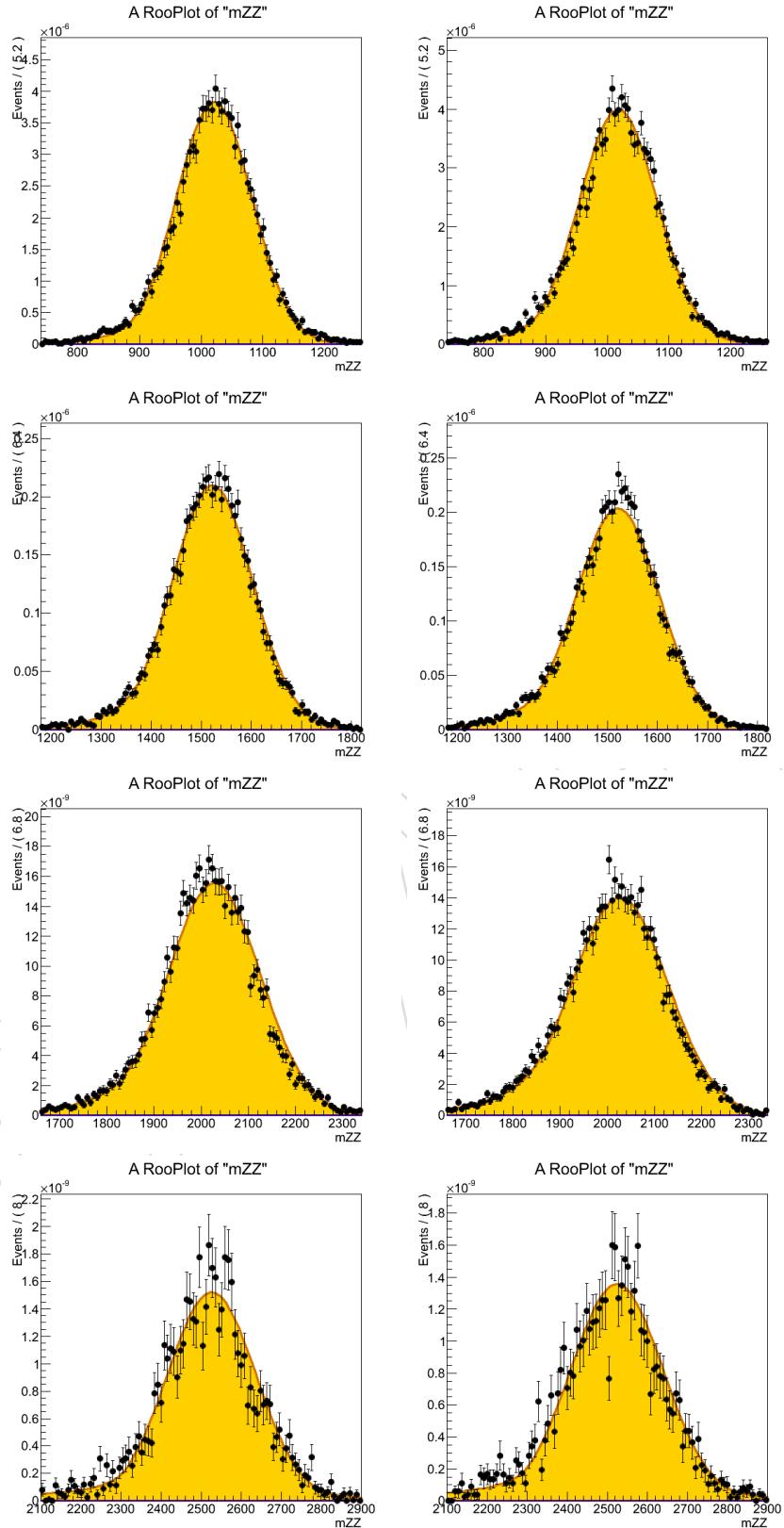


Figure 21: Distribution of the invariant mass, m_{WW} , and probability parameterization for MC bulk graviton with events with $\tilde{\chi} = 0.2$ and resonance mass of 1, 1.5, 2, and 2.5 TeV. Events reconstructed in electron (muon) channel are shown in the left (right) column.

M_G [GeV]	Signal Efficiency			
	ELE HP	MU HP	ELE LP	MU LP
600	0.090	0.130	0.017	0.026
700	0.126	0.158	0.025	0.032
800	0.159	0.176	0.034	0.042
900	0.170	0.185	0.037	0.047
1000	0.182	0.198	0.045	0.049
1100	0.189	0.198	0.049	0.054
1200	0.193	0.205	0.054	0.053
1300	0.201	0.204	0.062	0.053
1400	0.204	0.205	0.059	0.057
1500	0.200	0.200	0.059	0.062
1600	0.202	0.194	0.064	0.061
1700	0.201	0.184	0.068	0.062
1800	0.193	0.192	0.069	0.067
1900	0.190	0.184	0.071	0.067
2000	0.186	0.176	0.073	0.066
2100	0.184	0.174	0.076	0.067
2200	0.170	0.164	0.079	0.071
2300	0.155	0.154	0.076	0.073
2400	0.157	0.144	0.076	0.073
2500	0.140	0.132	0.079	0.074

Table 6: Expected signal efficiency for high-purity (HP), and low-purity (LP) categories, in electron and muon channel separately, for several resonance mass hypothesis.

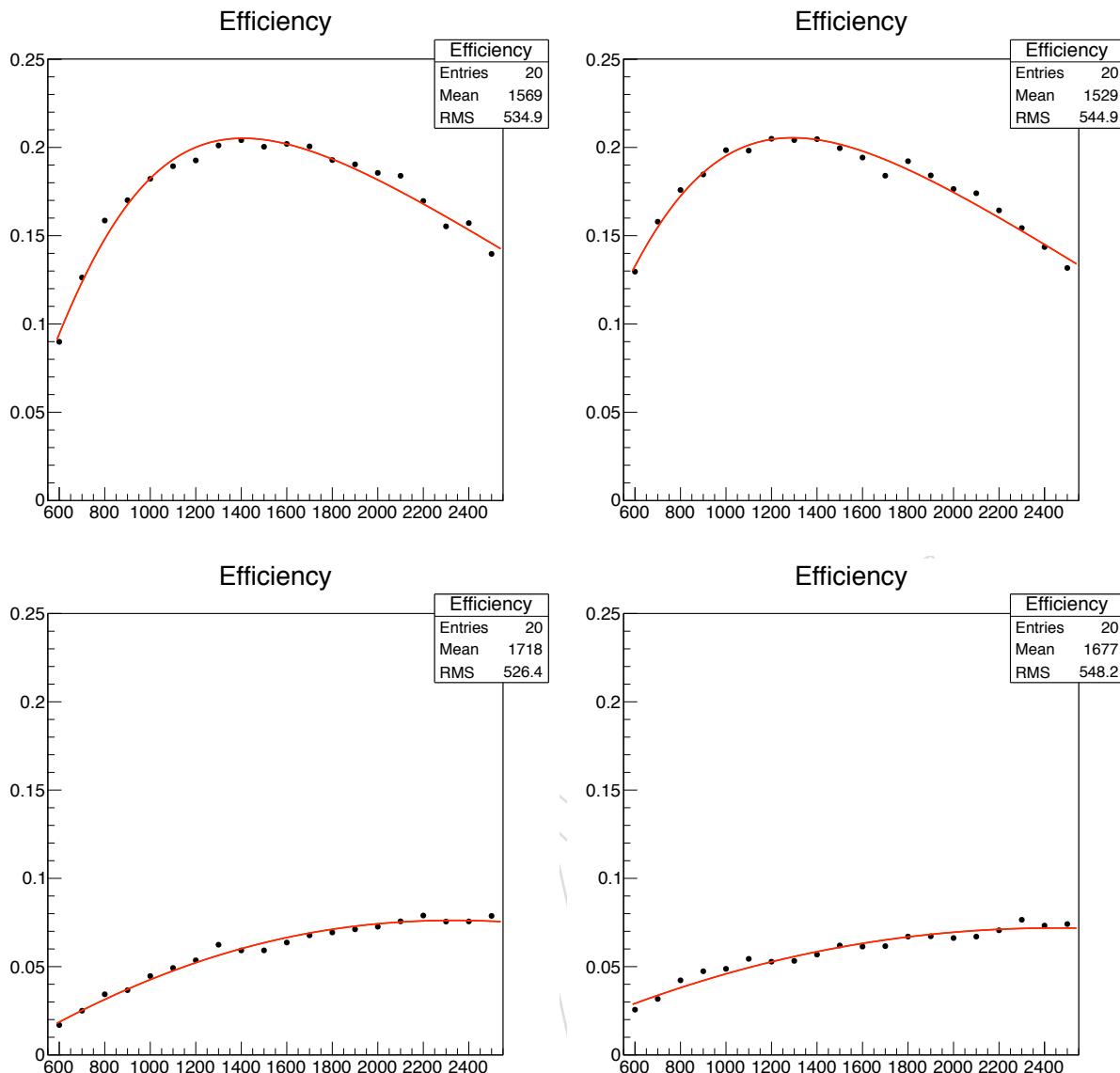


Figure 22: Parameterization of signal efficiency as a function of Graviton mass hypothesis in high-purity (top) and low-purity category (middle), in the electron (left) and muon (right) channels.

604 8 Background estimate

605 We search for bumps in the m_{WW} spectrum due to new, heavy $X \rightarrow WW$ resonances. This sec-
 606 tion describes the methods used to derive both the normalization and the shape of the m_{WW} dis-
 607 tribution for SM background events.

608 The background in the signal region is estimated from data (with the aid of MC for cross-checks
 609 and corrections) using two different methods:

- 610 • In the first method (α method), the background prediction in the signal region is
 611 obtained from the data in the low-mass sideband region ($40 < m_{jet}^{pruned} < 65$ GeV),
 612 rescaled for an appropriate transfer function from the sideband to the signal region
 613 derived from simulation.
- 614 • In the second method, we perform directly a “smoothness test” to the m_{WW} distribu-
 615 tion of data in the signal region ($65 < m_{jet}^{pruned} < 105$ GeV), and we search for bumps
 616 on top of an exponentially falling background spectrum, which is described by an
 617 empirical parameterization.

618 The analysis is performed using one or the other method (in parallel) and the final results (i.e.
 619 limits and p-value) are compared in Section 10.

620 These two methods are described in the following sections.

621 8.1 Method 1 - α method

622 The method presented in this section is similar to the one already used in previous CMS anal-
 623 ysis to search for an heavy Higgs (mass < 1 TeV) to WW [33] and X to ZZ [45], in the semi-
 624 leptonic channels.

625 As shown in Fig. 18, the main backgrounds after final selection are $W+jets$, $t\bar{t}$, and SM VV
 626 processes. While in the low-purity category the $W+jets$ dominates by far, in the high-purity
 627 category the 3 contributions are comparable, in particular at high values of m_{WW} . The three dif-
 628 ferent background sources are studied separately as described in the following sections. There
 629 is a small contribution single top process which is derived entirely from MC. There is also a
 630 small $Z+jets$ contribution which is effectively estimated together with the $W+jets$ (in the fol-
 631 lowing we’ll refer at the sum of the two background simply as “ $W+jets$ ”, since the contribution
 632 from $Z+jets$ is marginal).

633 The total background prediction (normalization and shape) as a function of the reconstructed
 634 m_{WW} is obtained by this formula, separately for each analysis category (ele LP, ele HP, mu LP,
 635 mu HP):

$$\begin{aligned} N_{tot}^{signal}(m_{WW}) &= N_{W+jets}^{signal}(m_{WW}) + N_{OB}^{signal}(m_{WW}) \\ &= N_{data}^{sideband}(m_{WW}) \times (1 - R_0(m_{WW})) \times \alpha^{MC}(m_{WW}) \times F_{W+jets} \\ &\quad + N_{OB}^{signal}(m_{WW}) \end{aligned} \tag{15}$$

636 where

- 637 • $N_{tot}^{signal}(m_{WW})$ is the total background prediction in the signal region;
- 638 • $N_{W+jets}^{signal}(m_{WW})$ is the $W+jets$ (plus the small $Z+jets$) background prediction in the
 639 signal region;

- $N_{OB}^{signal}(m_{WW}) = N_{VV}^{signal}(m_{WW}) + N_{t\bar{t}}^{signal}(m_{WW}) + N_{singletop}^{signal}(m_{WW})$ is the background prediction in the signal region for the sum of the non-W+jets backgrounds (OB = Other Backgrounds = VV, $t\bar{t}$, single top);
- $N_{data}^{sideband}(m_{WW})$ is the m_{WW} distribution in data for the low-mass sideband region ($40 < m_{jet}^{pruned} < 65$ GeV);
- $R_0(m_{WW}) = N_{OB}^{sideband}(m_{WW}) / N_{data}^{sideband}(m_{WW})$ is the fraction of non-W+jets backgrounds (VV, $t\bar{t}$, single top) in the sideband region with respect to the total background in the sideband region (i.e. data). The $1 - R_0(m_{WW})$ multiplicative term represents the subtraction of the non-W+jets contribution from the data in the sideband region.
- $\alpha^{MC}(m_{WW}) = N_{W+jets}^{signal}(m_{WW}) / N_{W+jets}^{sideband}(m_{WW})$ is the W+jets background prediction in the signal region divided by the one in the sideband region, calculated from MC. It represents the transport function (from sideband to signal region) used to correct the data in the sideband region and extract the W+jets m_{WW} shape.
- F_{W+jets} is an overall scale factor used to set the normalization of the W+jets background prediction.

The terms reported in the formula above are discussed in details in the following sections.

8.1.1 $t\bar{t}$ background

The $t\bar{t}$ process can have two real Ws one decaying in leptons and the other hadrons, thus generating the same final state of the $X \rightarrow WW$ signal. In addition this background have b quarks produced from the top decays, while b quarks are not expected in signal events. In the analysis, we use a veto on the number of b-tag jets (see Section 4.4.2) to suppress this contribution. Still some residual background is present after final selection and must be included in the final background estimation.

The $t\bar{t}$ background is estimated from MC (for the shapes of the distributions) with a data-driven normalization. The MC sample is initially rescaled to the integrated luminosity of the data using the NLO cross sections listed in Section 3.2. We then select a control sample dominated by $t\bar{t}$ events (top control sample) where we can check the agreement between data and MC both in shape and in the number of events. The selection of the top control sample is almost the same as the one described in Section 6.1 for the main signal sample, with the following changes:

- the hadronic W candidate is by default the CA8 jet with the highest p_T (as in the signal sample selection). If the of the jet is greater than 200 GeV (and there is more than one jet candidate) then the jet with the highest pruned mass is used to form the hadronic W candidate of the event;
- instead of a veto, we require that there is at least one jet tagged as b-jet accordingly with the CSV algorithm. We require that this b-tag jet is far from the W hadronic candidate ($\Delta R > 0.8$). This is done to ensure that we can use the existing b-tag scale factors which are derived in the event topologies with resolved jets.⁷.

Figures 23, 24 show the distributions of some reconstructed quantities employed in the analysis for the events in the top control sample. More plots are shown in Appendix C. The sample is dominated by $t\bar{t}$ events. In Fig. 23 we observe differences in the distributions of both the

⁷For top events, it should be noted that, if the b quark and the quarks from W decay are in the same reconstructed jet, the mass of the jet will be close to the top mass and therefore will be rejected by our pruned jet mass cuts.

682 pruned jet mass and the τ_{21} reconstructed quantity. On the other hand, the m_{WW} distribution is
 683 reasonably well described apart from an overall scale factor, as shown in Fig.24. In these plots,
 684 the minor W+jets contribution has been normalized using the same scale factors employed in
 685 Section 6.2 (1.475 for muon, 1.460 for electron).

686 Since pruned mass and τ_{21} are used in the W-tagging procedure we decided to correct the $t\bar{t}$
 687 simulation in the following way. Looking at the pruned jet mass distribution we can see that
 688 the events in the signal region $65 < m_{jet}^{pruned} < 105$ GeV follow two different distributions:
 689 a gaussian-like part (W signal) and a combinatorial component (for example due to partial
 690 overlap between the b-quark from top and the quarks from W in the same CA8 jet). Both
 691 components contribute to the background in the signal region. So, for each purity category,
 692 we count the number of events with $65 < m_{jet}^{pruned} < 105$ GeV (signal region) both in data and
 693 MC, and calculate their ratio. These data/MC scale factors are summarized in Table 7 and they
 694 are used to correct the $t\bar{t}$ MC prediction in the signal region ($N_{t\bar{t}}^{signal}(m_{WW})$) for the rest of the
 695 analysis.

696 We plan to put the two plots in Fig. 25 in the PAS.

Table 7: Scale factors for $t\bar{t}$ background MC derived from the top control sample.

Electron LP	Electron HP	Muon LP	Muon HP
1.30 ± 0.07	0.97 ± 0.03	1.23 ± 0.05	0.96 ± 0.02

697 8.1.2 VV background

698 Both WW and WZ processes from SM represent an irreducible background for this search. The
 699 contribution of the SM VV background becomes more and more important at high value of
 700 m_{WW} (for the high purity category) as shown in Fig. 18.

701 The SM VV background in the signal region ($N_{VV}^{signal}(m_{WW})$) is estimated from MC using NLO
 702 cross sections reported in Section 3.2. The efficiency for tagging a real hadronic W (or Z) is
 703 corrected using the data/MC scale factors introduced in Section 9.2.8.

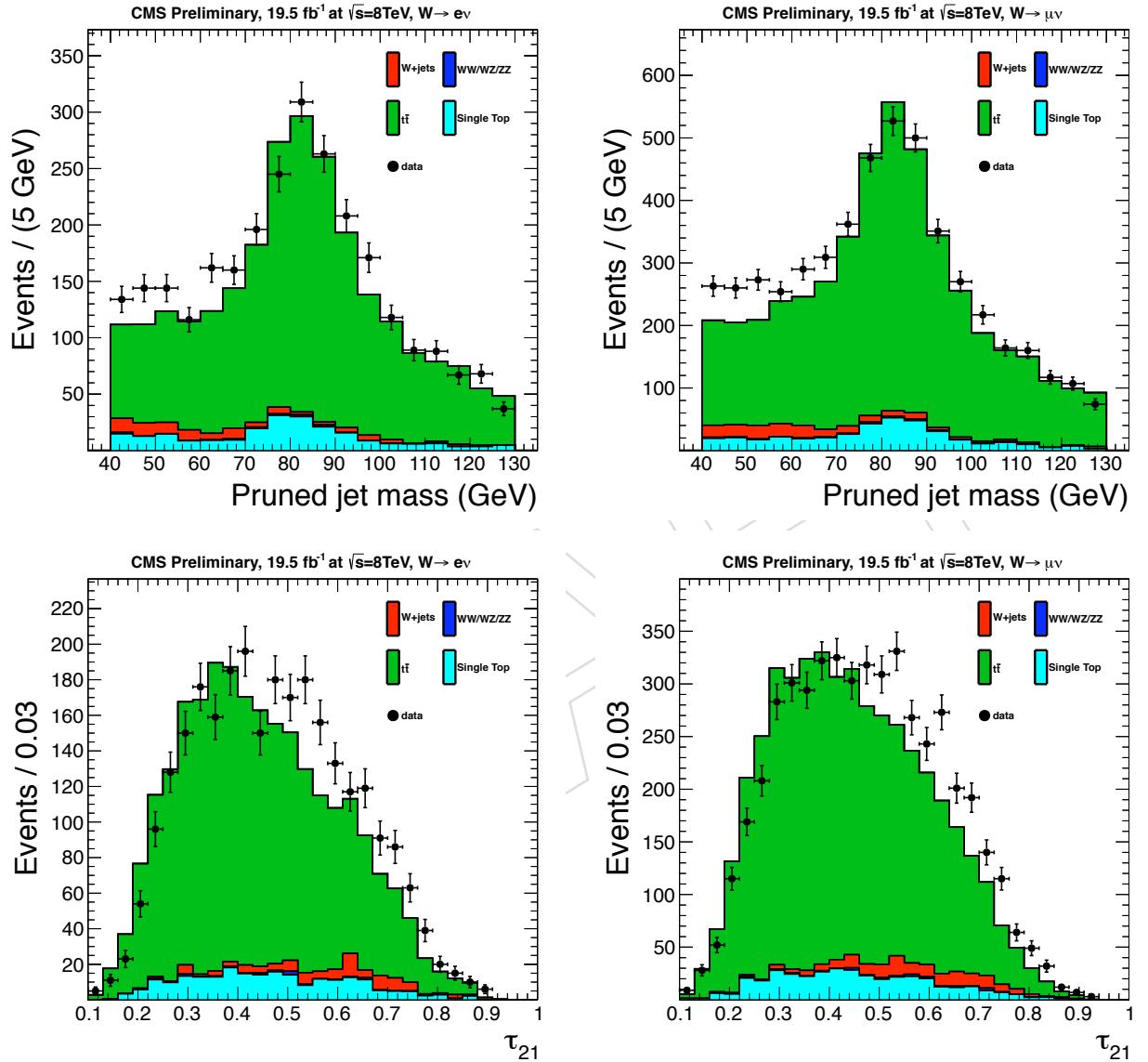


Figure 23: m_{jet}^{pruned} and τ_{21} for electron channel (left) and muon channel (right) for events with $40 < m_{jet}^{pruned} < 130 \text{ GeV}$ in the top control sample.

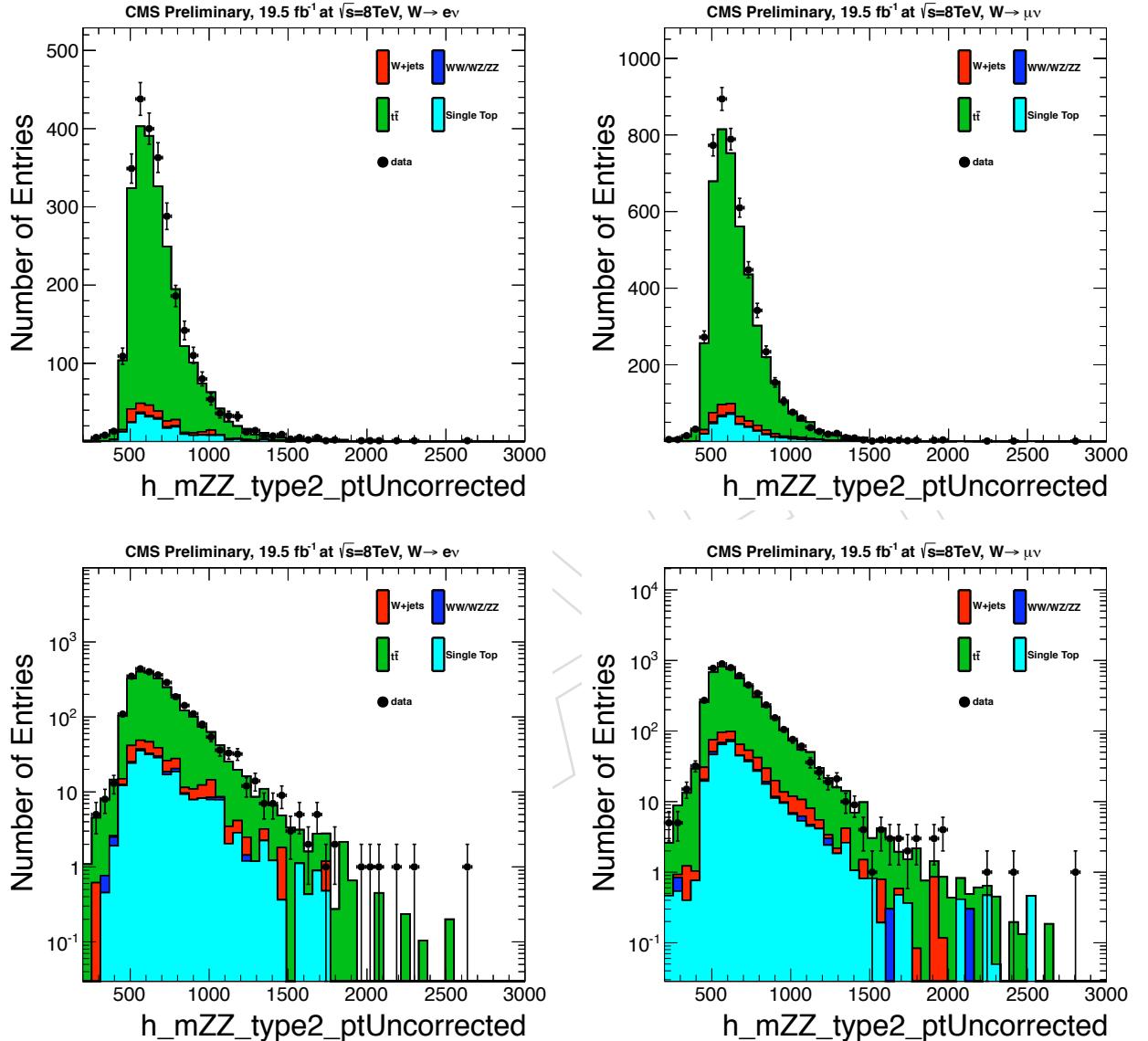


Figure 24: m_{WW} (using the $p_{z\nu}$ defined in Section 5.1) in linear and log scale for electron channel (left) and muon channel (right) for events with $40 < m_{jet}^{pruned} < 130$ GeV in the top control sample.

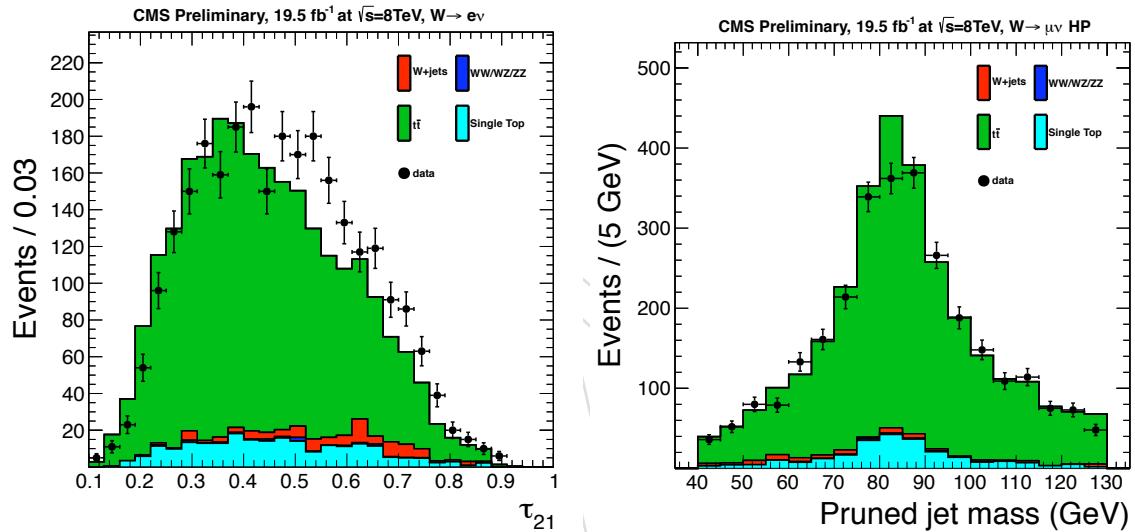


Figure 25: The τ_2/τ_1 distribution in the top-enriched control sample is shown on the left for the muon channel. The pruned jet mass distribution in the top-enriched control sample is shown on the right after applying the $\tau_2/\tau_1 < 0.5$ requirement.

704 **8.1.3 W+jets background**

705 The W+jets background is the dominant (reducible) background of this search in all the analysis
 706 categories. The real $W \rightarrow \ell\nu$ is reconstructed, while the hadronic W candidate is a light-quark
 707 initiated jet which gets a large mass due to QCD radiation in the fragmentation process (fake
 708 W). The pruned jet mass and τ_{21} are not perfectly modeled in the simulation, as visible in
 709 Figure 11. For this reason we use a data-driven method to estimate both the normalization and
 710 the m_{WW} shape of the W+jets background.

The W+jets background prediction has been introduced in Equation 15:

$$N_{W+jets}^{signal}(m_{WW}) = N_{data}^{sideband}(m_{WW}) \times (1 - R_0(m_{WW})) \times \alpha^{MC}(m_{WW}) \times F_{W+jets} \quad (16)$$

711 We proceed to describe the calculation of each term of this equation.

- F_{W+jets} - The normalization of the W+jets background is obtained using the same fit procedure employed in Ref. [1] (the actual software implementation is also exactly the same⁸). We select events passing the full selection (including τ_{21} selection) plus and additional requirement on the minimum reconstructed $m_{WW} > 800$ GeV (corresponding to the region where we perform the search for $X \rightarrow WW$ resonances). We then take the pruned jet mass distribution of the data in the range $40 < m_{jet}^{pruned} < 130$ GeV for each analysis category, as shown in Figure 26. The high-purity categories shows a clear peak around 80 GeV corresponding to the presence of real Ws decaying to hadrons in this sample. The background contributions for VV, $t\bar{t}$, and single top are obtained by fitting the individual MC predictions with various functions describing the combinatorial background and the Gaussian core of these distributions (more details can be found in Section 5.4 of [1]); the fit parameters (both normalization and shape) for these backgrounds are fixed by the MC prediction. The W+jets component is also parametrized with a combinatorial-like function which is found to well describe the MC (more details can be found in Section 5.4 of [1]). We finally perform a fit to the data in the *low-mass plus high-mass sideband regions* ($40 < m_{jet}^{pruned} < 65$ GeV plus $105 < m_{jet}^{pruned} < 130$ GeV) using the background parameterizations introduced above for each component. In this fit, the only free parameters are associated to the normalization and shape of the W+jets background, while the other background components are fixed from the simulation. The fit result for the W+jets background is extrapolated in the signal region ($65 < m_{jet}^{pruned} < 105$ GeV) to provide the normalization of the W+jets background (i.e. the scale factors F_{W+jets}). The total number of W+jets events expected in the signal region is reported in Table 8 for each analysis category.

Table 8: The total number of W+jets events expected in the signal region for each analysis category, derived from the fit to the pruned jet mass distribution.

Electron LP	Electron HP	Muon LP	Muon HP
547 ± 27	341 ± 20	836 ± 32	538 ± 25

- $R_0(m_{WW})$ - We use the data in the sideband of the pruned jet mass to estimate the W+jets background in the signal region as shown in Eq. 16. Figure 26 shows that W+jets is dominant background in the sideband region, but some non-negligible

⁸Thanks to the authors of [1] for providing their code.

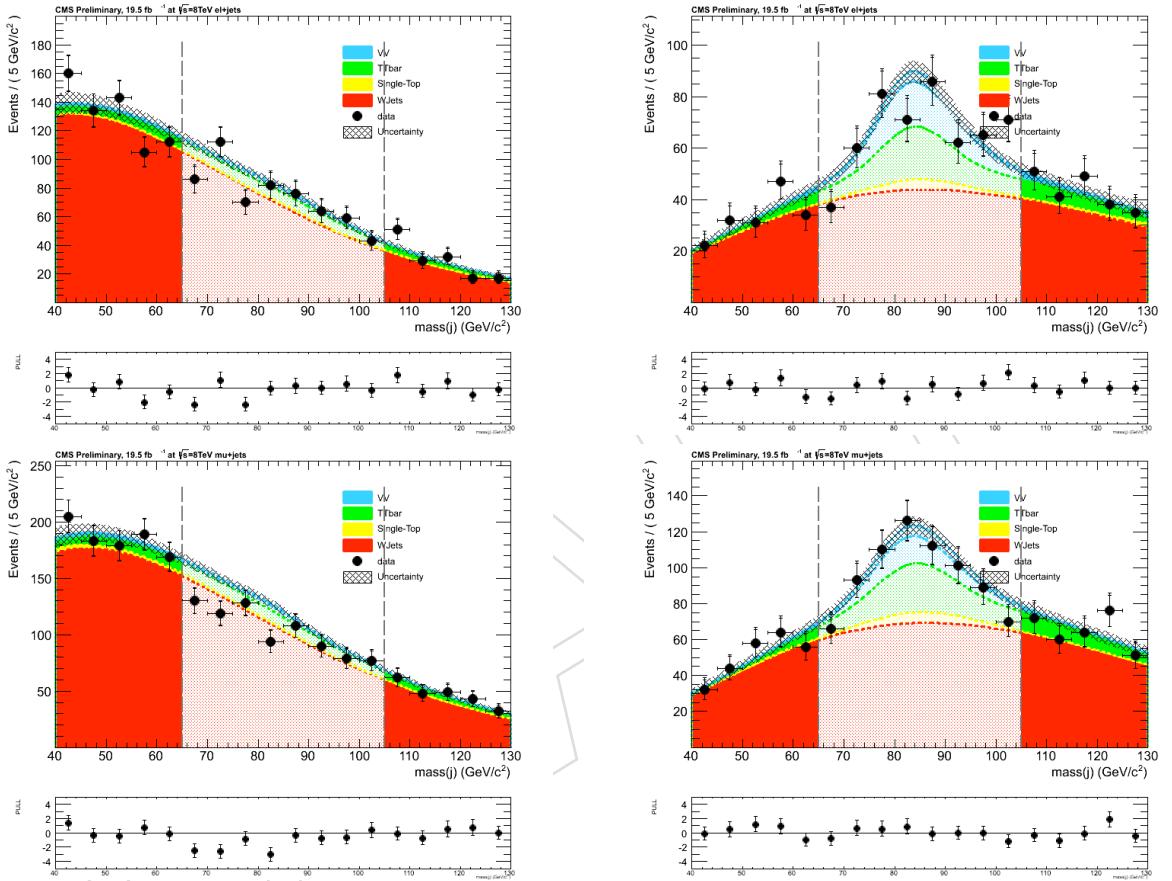


Figure 26: Pruned jet mass distributions for the electron LP (top-left), electron HP (top-right), muon LP (bottom-left), and muon HP (bottom-right) categories for events with $40 < m_{jet}^{pruned} < 130$ GeV passing the full selection plus the $m_{WW} > 800$ GeV requirement. The different background components, results of the fit to the low-mass and high-mass sideband regions in data, are shown.

contamination is present from other processes as well (mostly $t\bar{t}$). So we need to subtract the non-W+jets background from the data in the sideband before estimating the W+jets background. This “subtraction” is done via the $(1 - R_0(m_{WW}))$ multiplicative factor of Eq. 16, where $R_0(m_{WW}) = N_{OB}^{sideband}(m_{WW}) / N_{data}^{sideband}(m_{WW})$ is the fraction of non-W+jets backgrounds (VV, $t\bar{t}$, single top) in the sideband region with respect to the total background in the sideband region (i.e. data). The $R_0(m_{WW})$ functions for each analysis category are shown in Figure 27.

- $\alpha^{MC}(m_{WW}) = N_{W+jets}^{signal}(m_{WW}) / N_{W+jets}^{sideband}(m_{WW})$ is the W+jets background prediction in the signal region divided by the one in the sideband region, and it is calculated from MC. It represents the transport function (from sideband to signal region) used to correct the m_{WW} spectrum of the data in the sideband region and extract the W+jets m_{WW} shape in the signal region, as shown in Eq. 16. This method only provides the W+jets shape, while the normalization has been discussed before in this Section. Most of the systematic uncertainties associated to the MC cancels in this ratio. The method relies on the assumption that the correlation between m_{WW} and the pruned jet mass for W+jets events in data is reasonably well reproduced by the MC. The procedure has been successfully validated performing a closure test with data in the sideband region, as described in Appendix D. Figure 28 shows the $\alpha^{MC}(m_{WW})$ functions for the different analysis categories.

8.1.4 Total background

The total background prediction (both m_{WW} shape and normalization) is obtained using Eq. 15 (reported also below for simplicity):

$$\begin{aligned} N_{tot}^{signal}(m_{WW}) &= N_{W+jets}^{signal}(m_{WW}) + N_{OB}^{signal}(m_{WW}) \\ &= N_{data}^{sideband}(m_{WW}) \times (1 - R_0(m_{WW})) \times \alpha^{MC}(m_{WW}) \times F_{W+jets} \\ &\quad + N_{OB}^{signal}(m_{WW}) \end{aligned} \tag{17}$$

The dataset in the sideband region is corrected by the scale factors and transport functions discussed so far (RooDataset with event weights). The total background prediction is shown for the 4 analysis categories in Figures 29 (linear scale) and 30 (log scale), together with the actual data in the signal region. A good agreement between the data and the background prediction is observed.

A fit with the following leveled-exponential function is performed to the dataset representing the background prediction in the range $m_{WW} > 800$ GeV:

$$f(m_{WW}) = Norm. \cdot e^{-\frac{m_{WW}}{p_0 + p_1 m_{WW}}} \tag{18}$$

The studies reported in Appendix B show that this function is appropriate to describe the total background in the signal region. This fit result provides the final background model which is used to search for $X \rightarrow WW$ resonances.

Table 9 shows the observed number of data events, the expected background yield, and the expected signal yield for a bulk graviton models in each of the 4 channels (eleLP, eleHP, muLP, muHP).

8.1.5 Background uncertainty

We consider two type of uncertainties for the total background prediction:

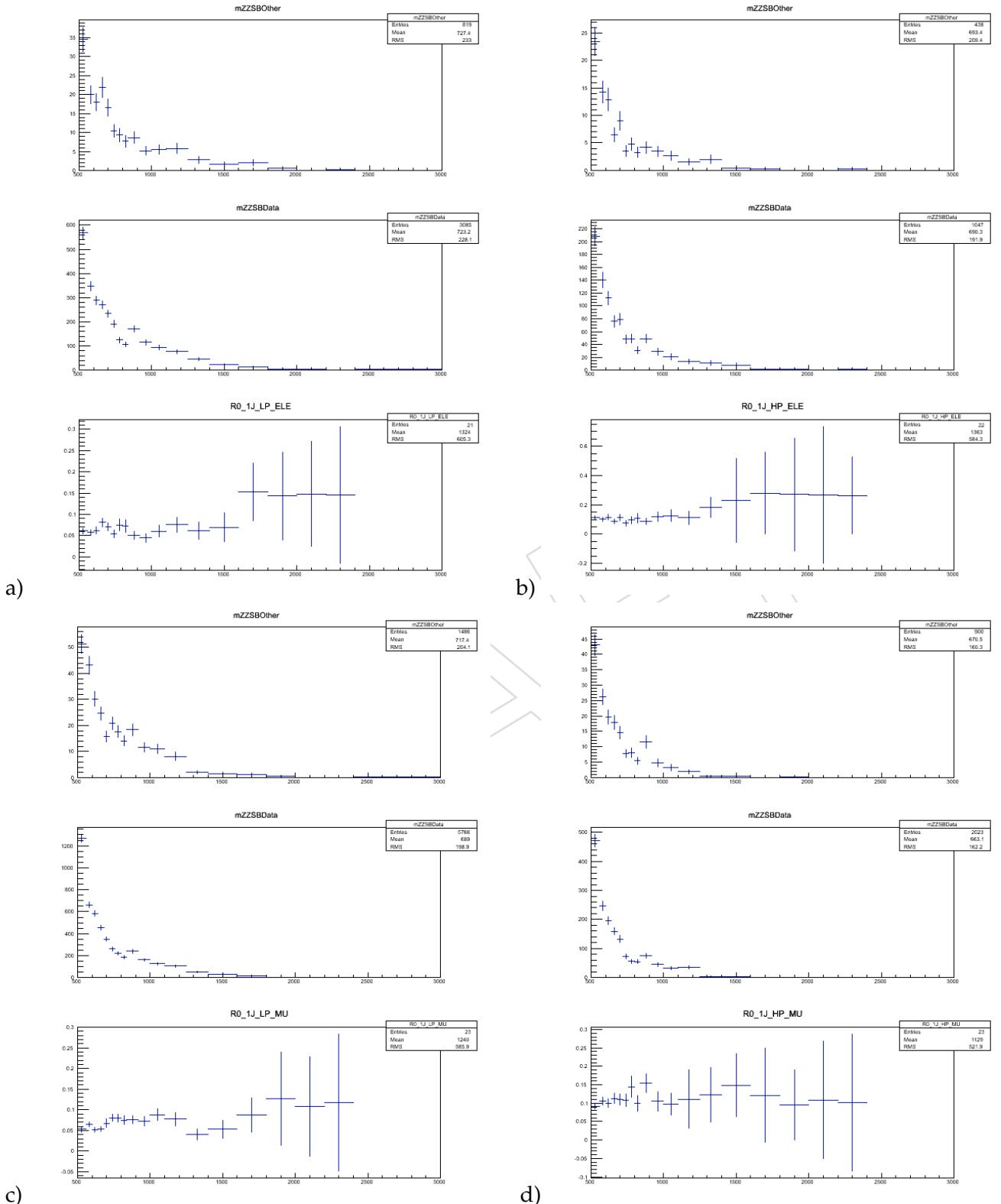


Figure 27: $R_0(m_{WW})$ functions vs m_{WW} for the electron LP (a), electron HP (b), muon LP (c), and muon HP (d) categories. Each box shows three plots (from the top to the bottom): $N_{OB}^{sideband}(m_{WW})$, $N_{data}^{sideband}(m_{WW})$, and their ratio $R_0(m_{WW})$.

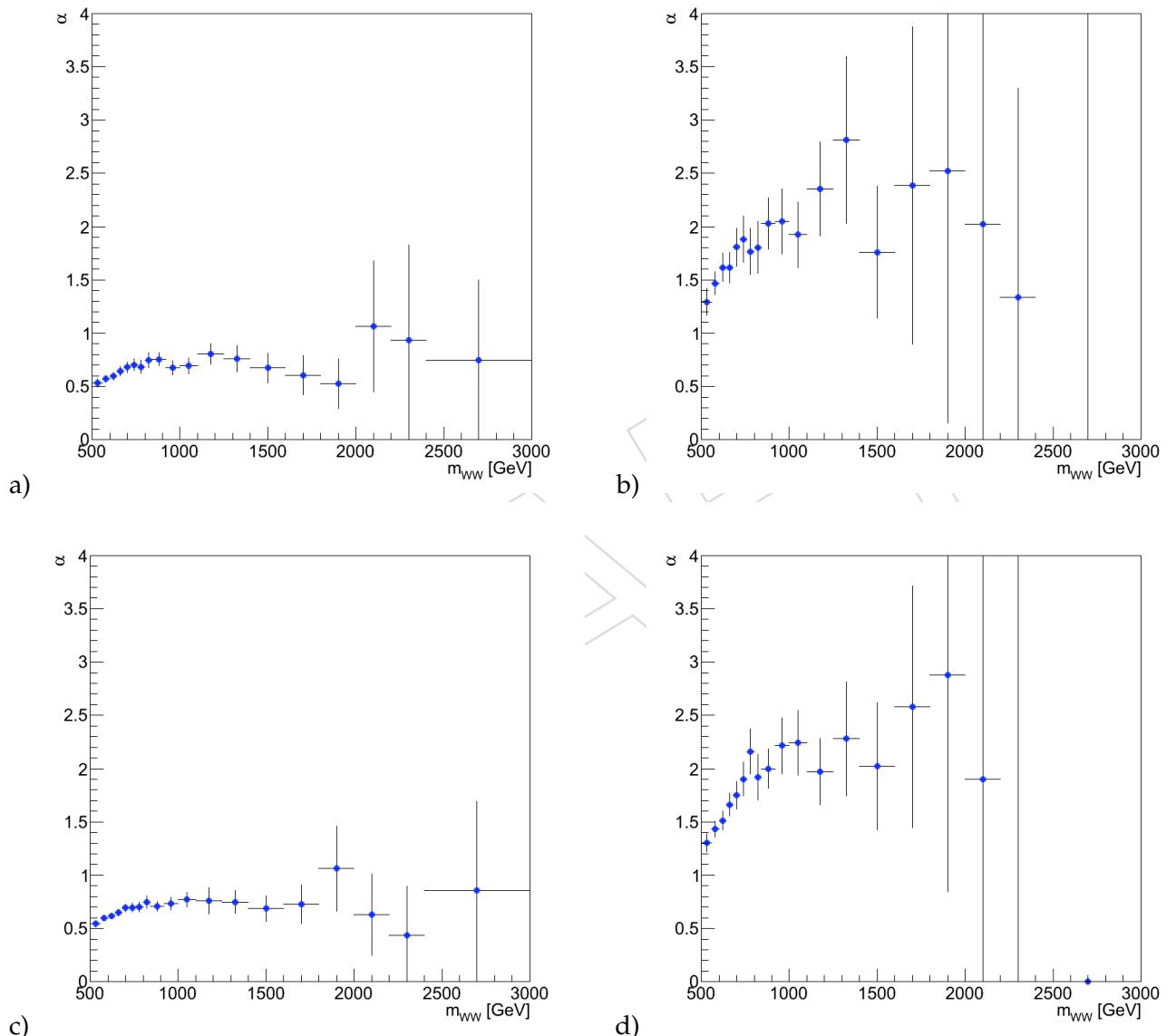


Figure 28: $\alpha^{MC}(m_{WW})$ functions vs m_{WW} for the electron LP (a), electron HP (b), muon LP (c), and muon HP (d) categories.

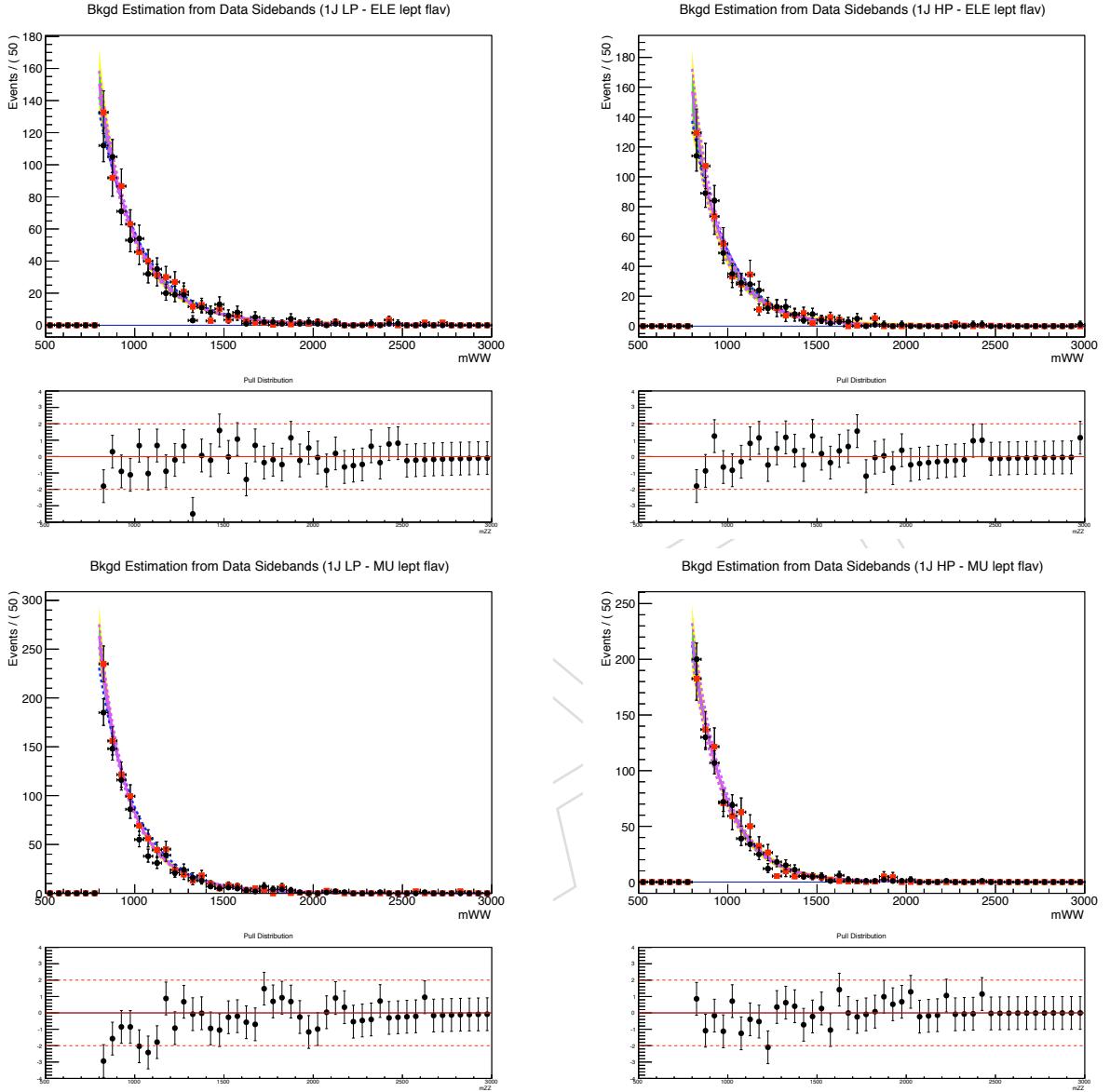


Figure 29: m_{WW} distributions for data (black points) in the signal region after full selection for the four event categories used in this analysis: electron LP (top-left), electron HP (top-right), muon LP (bottom-left), muon HP (bottom-right). The red point represents the background prediction from the data in the sideband region corrected by the scale factors and transport functions described in the text. The fit to the corrected data in the sideband (using a leveled exponential function with two shape parameters) is also shown together with error bands from fit parameters. The blue line represents a simple exponential fit (with only one shape parameter) not used in the analysis.

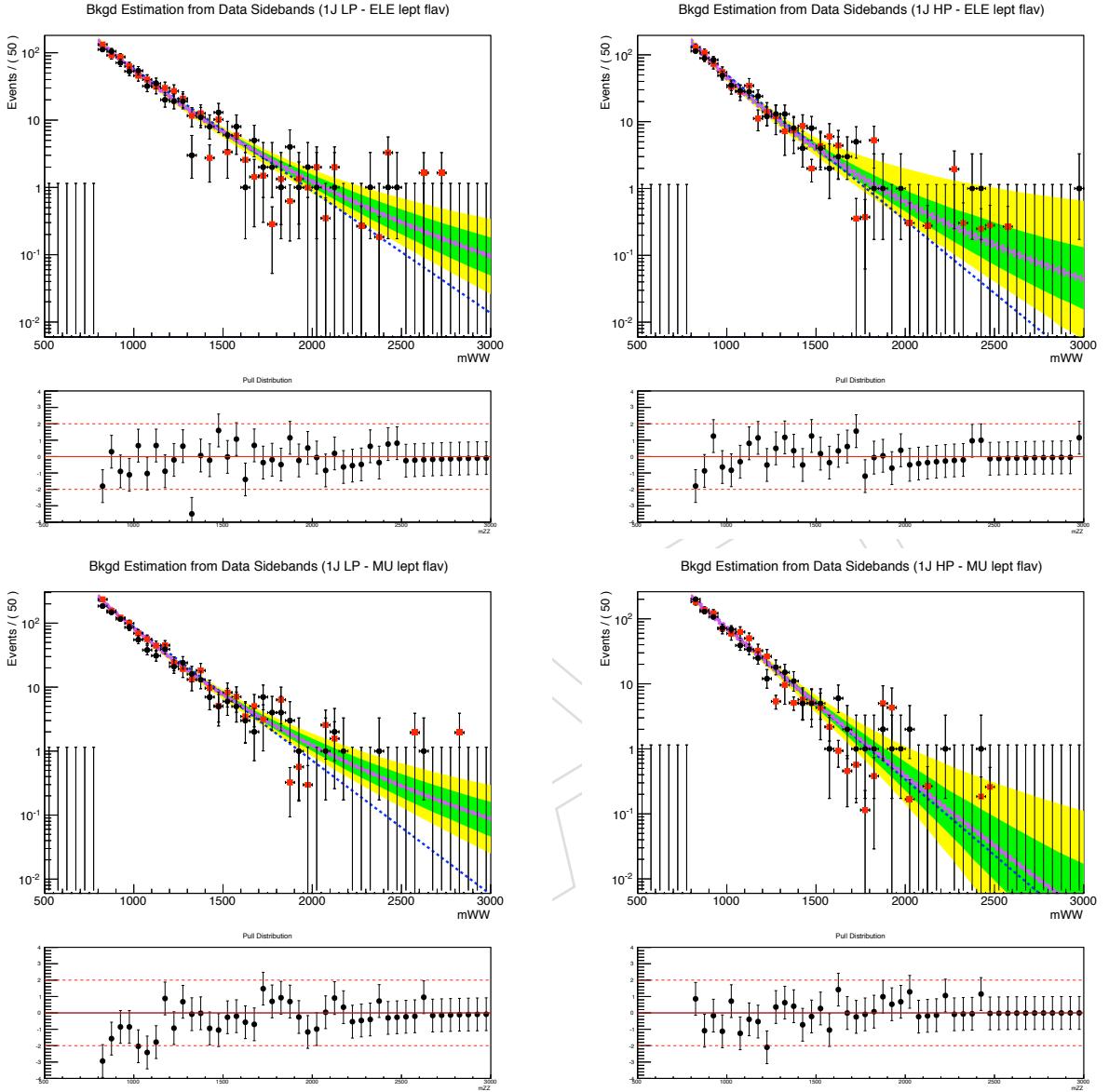


Figure 30: m_{WW} distributions for data (black points) in the signal region after full selection for the four event categories used in this analysis: electron LP (top-left), electron HP (top-right), muon LP (bottom-left), muon HP (bottom-right). The red point represents the background prediction from the data in the sideband region corrected by the scale factors and transport functions described in the text. The fit to the corrected data in the sideband (using a leveled exponential function with two shape parameters) is also shown together with error bands from fit parameters. The blue line represents a simple exponential fit (with only one shape parameter) not used in the analysis.

	Ele LP	Ele HP	Mu LP	Mu HP
Observed Data	592	533	825	767
Expected Background (data-driven, α method)	637 ± 34	548 ± 52	959 ± 41	793 ± 60
Expected Signal (MC)				
Bulk Graviton ($\tilde{k} = 0.5$)	M ₁ = 1000 GeV	2.4	11.3	2.6
				12.2

Table 9: Observed and expected yields with 19.5 fb^{-1} of data compared to the predictions from the MC samples and from the data-driven background estimation with the $m_{\text{jet}}^{\text{pruned}}$ sideband method. The yields are quoted in the range $800 < m_{WW} < 3000 \text{ GeV}$ for each category. Notice that signal extraction requires further analysis of the m_{WW} spectrum exploiting the shapes of the expected signal. The errors on the expected background comes only from statistical uncertainty in the data of the sideband.

774 • Uncertainty on background fit parameters (shape and normalization)

775 We use a leveled exponential fit function with 2 shape parameters and 1 normaliza-
 776 tion parameter. Each fit returns uncertainties on all the fitted parameter, which
 777 are in general correlated. We want to include systematic uncertainties on the fit pa-
 778 rameters in the limit/p-value calculation. However, from a technical point of view,
 779 the statistical tool that we use for the limit calculation (see Section 10) is not able
 780 to properly deal with partially correlated parameters. For this reason we redefine
 781 the parameters by diagonalizing the covariance matrix in order to decorrelate them.
 782 In this procedure, the new parameters are defined in such a way to be centered at
 783 zero and with error equal to unity⁹. The background fit parameterization is then
 784 redefined as a function of these new, uncorrelated parameters. This new fit function
 785 (and its related uncertainties on fit parameters) is used to describe the background
 786 in the limit/p-value calculation of Section 10.

787 • Uncertainty on $\alpha^{\text{MC}}(m_{WW})$

788 In addition to the statistical uncertainties from the number of events in the sidebands
 789 of data, there are statistical uncertainties on the extrapolation factors $\alpha^{\text{MC}}(m_{WW})$ due
 790 to the limited size of the W+jets MC sample. These uncertainties are propagated
 791 to the final background parameterization in the following way: a large number of
 792 alternative sets of m_{WW} distributions, for both numerator and denominator of the
 793 $\alpha^{\text{MC}}(m_{WW})$ ratio, are randomly generated from the original histograms, according to
 794 their statistical uncertainties in each bin. We generated 500 pseudo-datasets. A new
 795 $\alpha^{\text{MC}}(m_{WW})$ ratio is calculated for each dataset. The fit to the extrapolated sidebands
 796 is repeated for each $\alpha^{\text{MC}}(m_{WW})$ and the modified extrapolated m_{WW} shape is stored,
 797 including the fit results.

798 The distributions of the shape parameters of the fit function derived from these alter-
 799 native fits are shown in Figures 31 and 32, including the diagonalization procedure
 800 described in the previous bullet. The systematic uncertainty that we assign at the
 801 end of this procedure corresponds to the standard deviation of the Gaussian fit to the
 802 distributions of the fitted values. The values of these uncertainties are of the same
 803 order of magnitude of the errors coming from the covariance matrix, emphasizing
 804 the importance of including this uncertainty in the final statistical analysis. The fact
 805 that the varied parameter follows a Gaussian distribution indicates that the width

⁹the shift and renormalization of the parameter is a somewhat arbitrary convention present in the diagonaliza-
 tion tool that we use.

806 of these Gaussian distributions may be added in quadrature to the data statistical
807 errors to include the uncertainty due to limited W+jets MC statistics for $\alpha^{MC}(m_{WW})$
808 calculation.

809 Studies reported at [1] have shown that $\alpha^{MC}(m_{WW})$ can change due to different MC
810 with different parton showering algorithm considered. Herwig++ and Pythia were
811 compared in this study, showing that the variation on $\alpha^{MC}(m_{WW})$ is of the same
812 size of the statistical uncertainties on the MC. For this reason the uncertainties on
813 $\alpha^{MC}(m_{WW})$ have been inflated by a factor $\sqrt{2}$ to take this effect into account.

814 The uncertainty on $R_0(m_{WW})$ due to the MC prediction in the sideband region (MC statistics,
815 cross section of the non-W+jets processes, etc..) is negligible compared to the other uncertainties
816 reported above, since the non-W+jets contamination in the sideband region is quite small (≈ 5 -
817 10% depending on the category, see Figure 27).

DRAFT

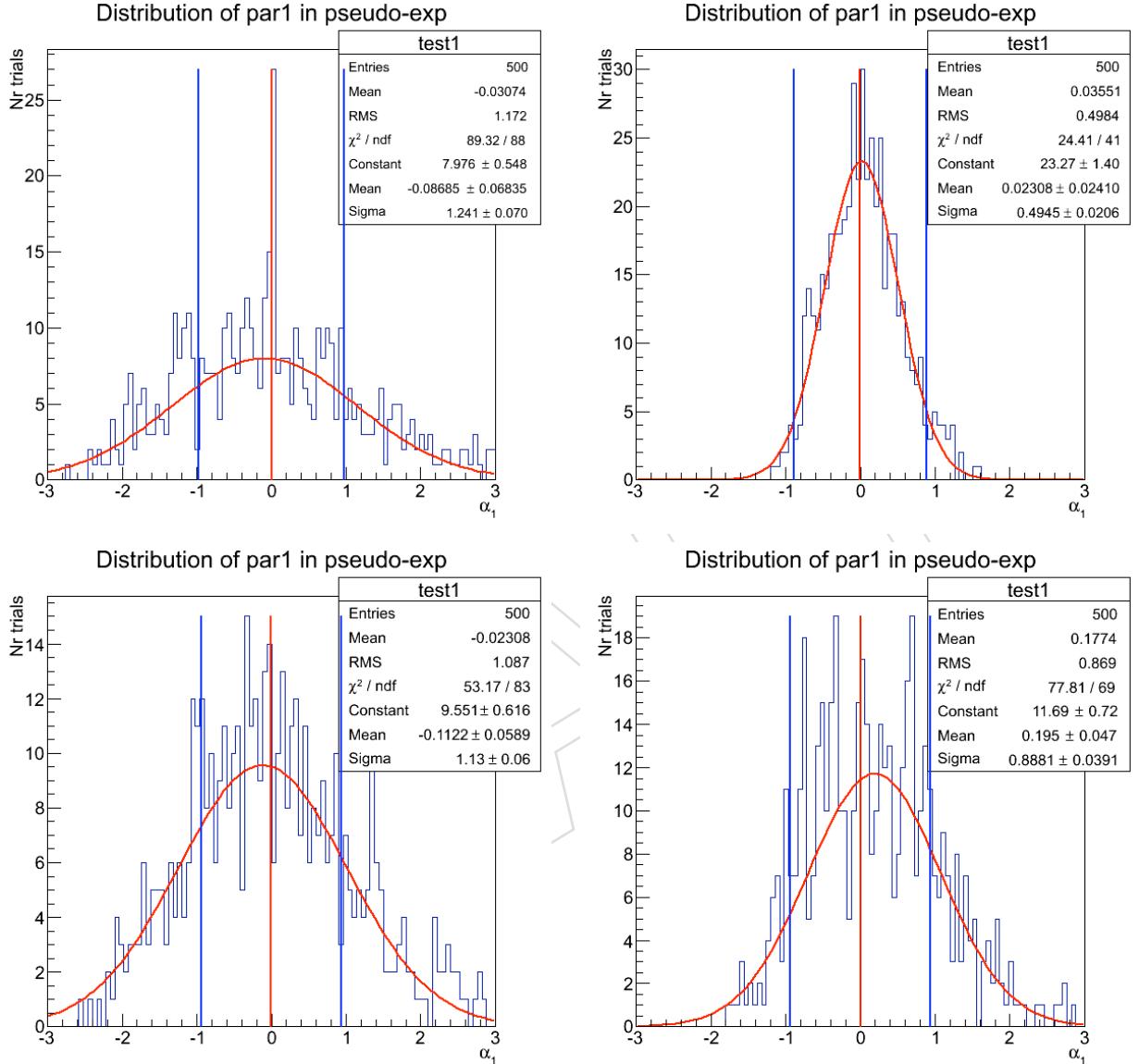


Figure 31: Background shape parameter 1 (after diagonalization procedure described in the text) as derived from the alternative extrapolation factors (histogram), compared to the nominal values (red line) and the uncertainty on the nominal values calculated from the covariance matrix of the fit (blue lines). The distributions are presented for the eleLP (top-left), eleHP (top-right), muLP (bottom-left), muHP (bottom right) categories, respectively.

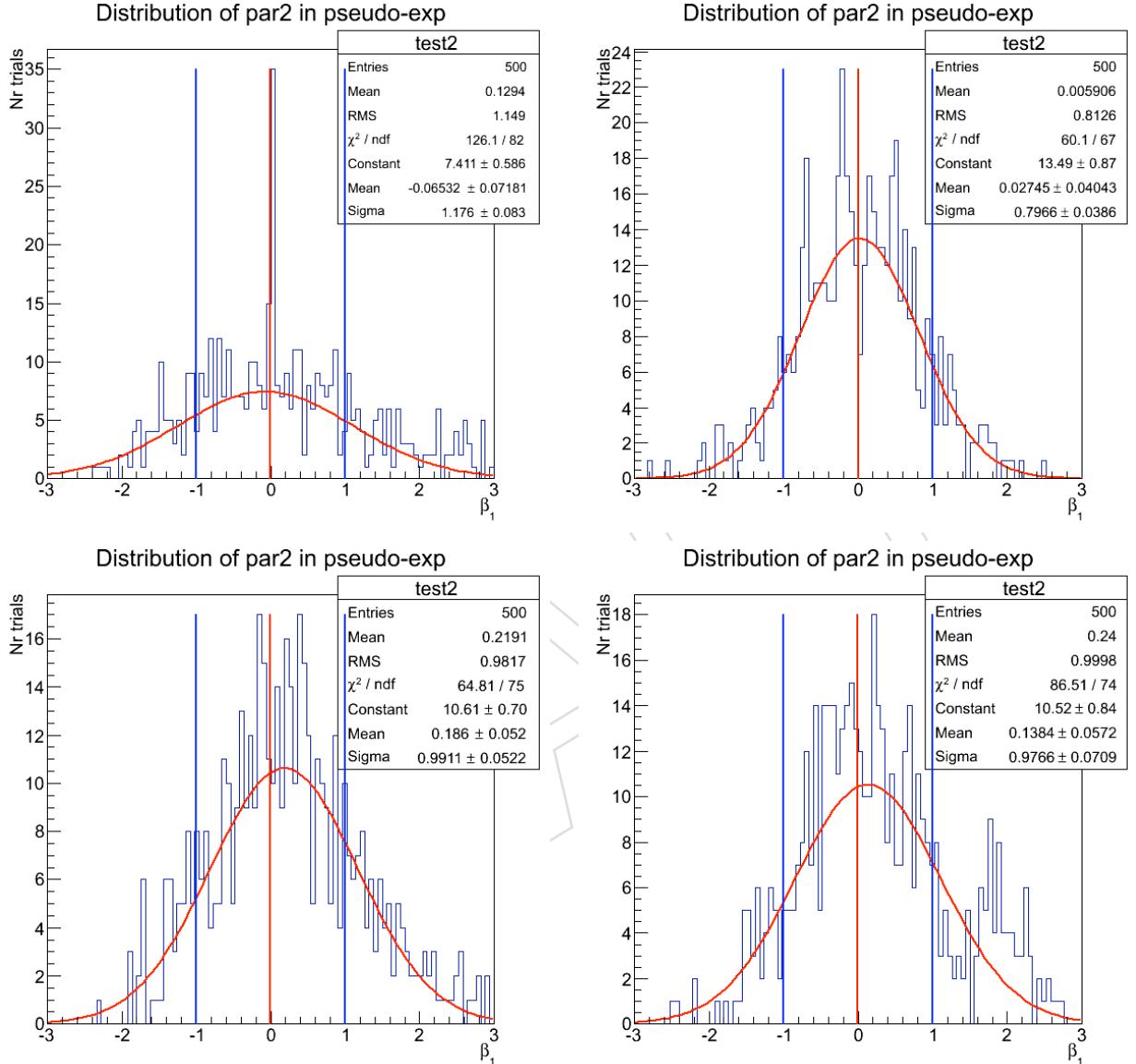


Figure 32: Background shape parameter 2 (after diagonalization procedure described in the text) as derived from the alternative extrapolation factors (histogram), compared to the nominal values (red line) and the uncertainty on the nominal values calculated from the covariance matrix of the fit (blue lines). The distributions are presented for the eleLP (top-left), eleHP (top-right), muLP (bottom-left), muHP (bottom right) categories, respectively.

818 **8.2 Method 2 - m_{WW} smoothness test**

819 The method presented in this section is similar to the one already used in previous CMS anal-
 820 ysis to search for bumps produced by heavy resonances in a steeply falling reconstructed mass
 821 distribution of data [19, 46].

We take the m_{WW} distributions in the signal region after the final selection for each of the 4 categories considered (electron HP/LP, muon HP/LP). For each category we perform a fit to the data using the following leveled-exponential parameterization:

$$f(m_{WW}) = \text{Norm.} \cdot e^{-\frac{m_{WW}}{p_0 + p_1 m_{WW}}} \quad (19)$$

822 From this functional form, one can see that for the values of the fit parameters $p_1 = 0$, one
 823 recovers the simple exponential fit.

824 We studied in a quantitative way the quality of the fits and how many free parameters are
 825 effectively needed in the fit function for properly describing the background (see Appendix B).

826 The outcome of these studies is that the 2-degree leveled exponential function introduced above
 827 is able to describe well the m_{WW} distribution for all the categories.

828 The results of the background-only fits to the m_{WW} distributions in data (starting from 700 GeV,
 829 as in the PAS) for the 4 categories are shown in Figure 33. A fix binning of 100 GeV is used to
 830 show the mass spectra but the fit is unbinned. The fit has a chi-squared per degree of freedom
 831 (χ^2/ndf) of 13.3/14, 7.6/13, 19.8/14, and 8.2/13 for the electron LP, electron HP, muon LP,
 832 muon HP categories, respectively¹⁰.

833 By visual inspection of the pull distribution, no deviations that are statistically significant and
 834 consistent with a WW resonance peak, are observed between the distribution of the data points
 835 and the smooth fit through all the m_{WW} spectrum. The same functional form used here to
 836 perform a pure background-only fit to the data is used also in Section 10 as background model
 837 to set limits on the cross section of new $X \rightarrow WW$ resonances and to calculate the p-value
 838 (significance) of potential excesses of events observed in data.



¹⁰The χ^2 is calculated in the fit range using the same binning of the plot, standard ROOT errors (Sumw2), and skipping bins with zero entries

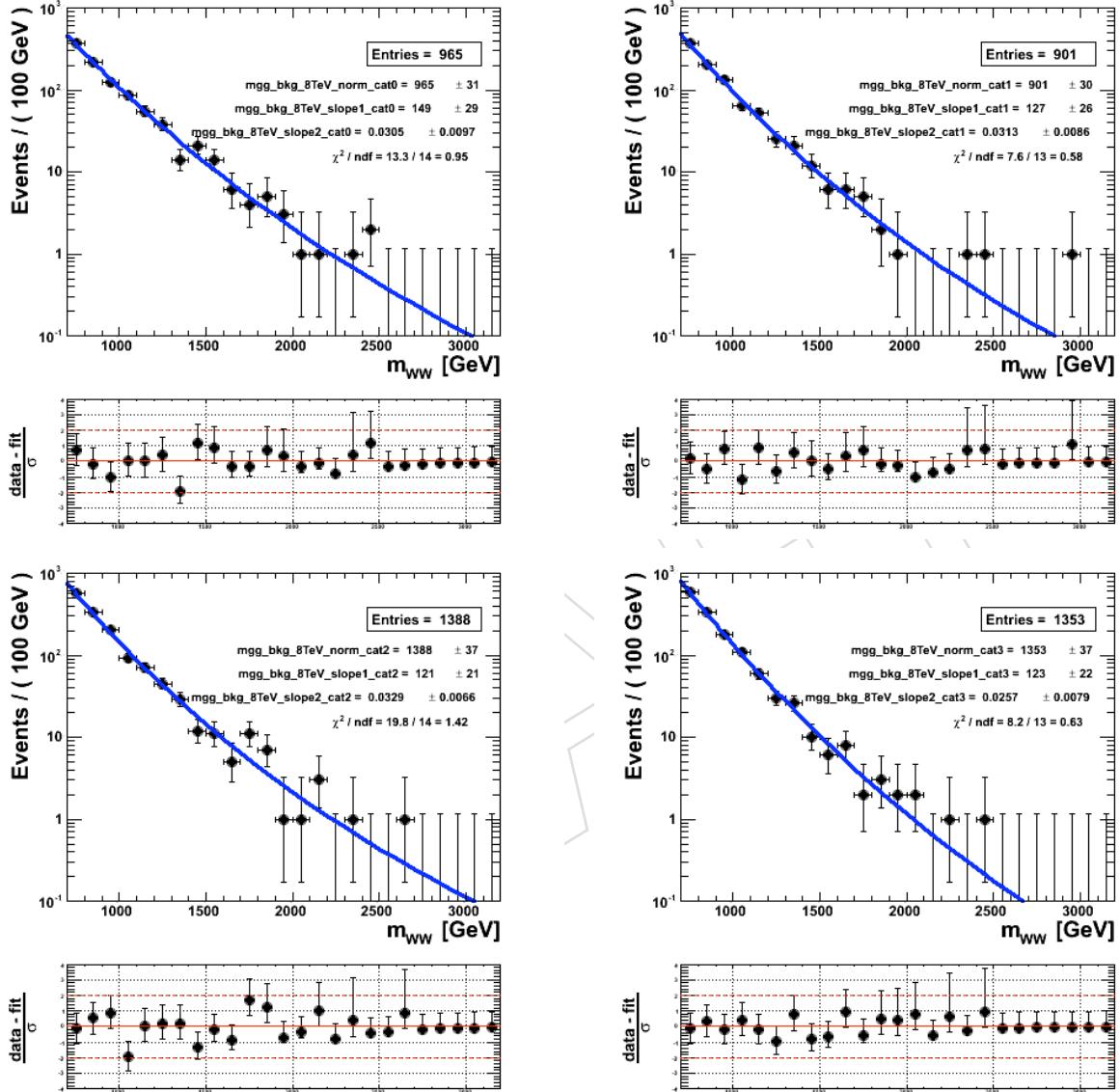


Figure 33: m_{WW} distributions for data in the signal region after full selection for the four event categories used in this analysis: electron LP (top-left), electron HP (top-right), muon LP (bottom-left), muon HP (bottom-right). Background-only fits using a 2nd-order levelled exponential are shown, together with the pull distribution.

839 9 Systematic Uncertainties

840 This section describes the systematic uncertainties on signal ad background used for the calcu-
 841 lation of the limits on cross section and the significance/p-value of potential excesses of events
 842 observed in the data.

843 The uncertainties described below are all included as nuisance parameters in the statistic tool
 844 described in Section 10.

845 9.1 Background Normalization and Shape

846 The systematics uncertainties on the parameters describing the background shape and back-
 847 ground normalization differ between method 1 (α , Section 8.1) band method 2 (m_{WW} smooth-
 848 ness test, Section 8.2).

849 9.1.1 Background uncertainties - Method 1

850 The main uncertainties on the background modeling (both shape and normalization) come
 851 from the number of data events in the m_{jet}^{pruned} sideband region and the size of the MC sample
 852 used to estimate the transport function $\alpha^{MC}(m_{WW})$. More details on how these uncertainties are
 853 evaluated are discussed in Section 8.1.5. In the statistical analysis described in Section 10, the
 854 background parameters are nuisance parameters floating within the range specified by their
 855 uncertainties using a Gaussian constraint (“param” option in Higgs combination tool).

856 9.1.2 Background uncertainties - Method 2

857 In this method, the background is estimated by fit the data in the signal region using a smoothly
 858 falling function. In the statistical analysis described in Section 10, the background parameters
 859 (both shape and normalization) are nuisance parameters floating without limits (“flatParam”
 860 option in Higgs combination tool).

861 9.2 Signal Normalization

862 The following sources of systematic uncertainties affect the number of signal events expected
 863 after the final selection (either due to signal efficiency or integrated luminosity uncertainties).
 864 The signal efficiency and the integrated luminosity are used as input in the limit and signifi-
 865 cance calculation.

866 The uncertainties are calculated for the LP+HP category, separately for electron and muon
 867 channel.

868 9.2.1 Integrated luminosity

869 The uncertainty on the knowledge of the integrated luminosity of the data sample (4.4%) [47]
 870 introduce an uncertainty on the number of signal events passing the final selection.

871 This uncertainty is fully correlated in all channels.

872 9.2.2 Trigger

873 The uncertainty on the single lepton trigger efficiency introduces an uncertainty in the signal
 874 efficiency. Uncertainties of 1% for both electron channel and muon channel are taken from this
 875 Ref. [29].

876 This uncertainty is fully correlated in LP and HP channels for the same lepton flavour, and
 877 uncorrelated for different lepton flavours.

878 **9.2.3 Lepton energy/momentum scale**

879 The uncertainty on the lepton energy / momentum scale introduces an uncertainty in the signal
 880 efficiency. We scale up and down the electron energy by 0.6% for barrel and 1.5% for endcap,
 881 taken from Ref. [48], propagate the change to the E_T^{miss} reconstructed in the event, and re-run the
 882 entire analysis for the signal samples. We check the variation in the number of selected signal
 883 events from the nominal value and take the largest relative difference as systematic uncertainty.
 884 The same procedure is used also for the muon channel, by scaling up and down the muon
 885 momentum by 0.2% (or 5% if the muon p_T is greater than 200 GeV). A flat value of 0.2 (0.7)%
 886 for electron (muon) channel is used for all the resonance mass hypotheses.

887 This uncertainty is fully correlated in LP and HP channels for the same lepton flavour, and
 888 uncorrelated for different lepton flavours.

889 **9.2.4 Lepton energy/momentum resolution**

890 The uncertainty on the lepton energy / momentum resolution introduces an uncertainty in the
 891 signal efficiency. We smear the electron energy momentum 1.3% for barrel and 2.8% for endcap,
 892 taken from Ref. [49], propagate the change to the E_T^{miss} reconstructed in the event, and re-run
 893 the entire analysis for the signal samples. We check the variation in the number of selected
 894 signal events from the nominal value and take the largest relative difference as systematic un-
 895 certainty. The variation is found to be of the order of 0.1% (or smaller) for both electron and
 896 muon channels for all mass hypotheses and therefore negligible for this analysis.

897 This uncertainty is fully correlated in LP and HP channels for the same lepton flavour.

898 **9.2.5 Lepton identification/isolation efficiency**

899 The uncertainty on the lepton identification and isolation efficiency introduces an uncertainty
 900 in the signal efficiency. Uncertainties of 3% for electron channel and 1% for muon channel are
 901 taken from this Ref. [29].

902 This uncertainty is fully correlated in LP and HP channels for the same lepton flavour, and
 903 uncorrelated for different lepton flavours.

904 **9.2.6 Jet energy scale**

905 The uncertainty on the jet energy scale introduces an uncertainty in the signal efficiency. We
 906 scale up and down the CA8 jet energy accordingly with the uncertainties in the p_T - and η -
 907 dependent jet energy corrections (accessible via the global tag). At the same time, we scale
 908 up and down (in the same direction of the CA8 jet case) also the energy of all the jets in the
 909 AK5 collection with $p_T > 10$ GeV, $EMEnergyFraction < 0.9$, and $MuonEnergyFraction < 0.9$
 910 (the latter two cuts to clean the collection from electrons and jets) and propagate the change to
 911 the E_T^{miss} (the change on the E_T^{miss} is calculated only using variations of AK5 jets; this is done in
 912 order not to double-count the contribution from CA8 jets matching with AK5 jets). We then re-
 913 run the entire analysis for the signal samples using the new jet p_T 's and new E_T^{miss} . We check the
 914 variation in the number of selected signal events from the nominal value and take the largest
 915 relative difference as systematic uncertainty. The resulting systematic uncertainty depends on
 916 the resonance mass (M_{res}) accordingly with this linear function, $\sigma = 0.37\% + 1.05 \cdot 10^{-3}\% \cdot M_{\text{res}}$.
 917 for both electron and muon channels.

918 This uncertainty is fully correlated in all channels.

919 9.2.7 Jet energy resolution

920 The uncertainty on the jet energy resolution introduces an uncertainty in the signal efficiency.
 921 We smear the CA8 jet energy by the data/MC scale factors for jet energy resolution as a function
 922 of jet η following the recommendations reported at <https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution>. At the same time, we scale up and down (in the same direction of the CA8
 924 jet case) also the energy of all the jets in the AK5 collection with $p_T > 10 \text{ GeV}$, $EMEnergyFraction <$
 925 0.9 , and $MuonEnergyFraction < 0.9$ (the latter two cuts to clean the collection from electrons
 926 and jets) and propagate the change to the E_T^{miss} (the change on the E_T^{miss} is calculated only using
 927 variations of AK5 jets; this is done in order not to double-count the contribution from CA8 jets
 928 matching with AK5 jets). We then re-run the entire analysis for the signal samples using the
 929 new jet p_T 's and new E_T^{miss} . We check the variation in the number of selected signal events be-
 930 tween the nominal value and the value obtained after the correction for the different jet energy
 931 resolution in data and MC. The relative variation is taken as systematic uncertainty. A flat value
 932 of 0.3% for both electron and muon channel is used for all the resonance mass hypotheses.

933 This uncertainty is fully correlated in all channels.

934 9.2.8 W-tagging efficiency

935 The uncertainty on the W-tagging efficiency introduces an uncertainty in the signal efficiency.
 936 Data/MC scale factors and uncertainties on the W-tagging has been evaluated using a control
 937 sample of $t\bar{t}$ events in the semi-leptonic channel. More details can be found in Ref. [1]. The
 938 data/MC scale factor is 0.93 for high purity with an uncertainty of 9%, and 1.10 for low purity
 939 with an uncertainty of 27%. These scale factors are applied on the signal and SM VV predictions
 940 from MC.

941 This uncertainty is fully anti-correlated in LP and HP channels for the same lepton flavour, and
 942 uncorrelated for different lepton flavours.

943 9.2.9 E_T^{miss} scale and resolution

Uncertainties in the energy/momentum scale and resolution of leptons, jets in the event are propagated to an uncertainty on the E_T^{miss} determination, and therefore introduce an uncertainty in the signal efficiency. The procedure used to include the E_T^{miss} uncertainties is described in the sub-sections above for each source of systematic uncertainty. The generic formula for the E_T^{miss} after variations is:

$$\vec{E}_T^{\text{miss}}(\text{mod.}) = \vec{E}_T^{\text{miss}} + \sum_{\text{recoobjects}} [\vec{p}_T - \vec{p}_T(\text{mod.})] \quad (20)$$

944 where “mod.” indicates the new vector after the variations have been applied.

945 9.2.10 b-tag identification efficiency

946 The uncertainty on the b-tag identification efficiency for light-flavour (u,d,s) and heavy-flavour
 947 (c and b) quarks introduces an uncertainty in the signal efficiency. We scale up and down the
 948 b-tag scale factors (separately for light flavour and heavy flavour jets) accordingly with the
 949 uncertainties reported at <https://twiki.cern.ch/twiki/bin/viewauth/CMS/BTagSFMethods>. We then re-run the entire analysis for the signal samples using the new b-tag scale factors.
 951 We check the variation in the number of selected signal events between the nominal value and

952 the value obtained with the new scale factors. The relative variation is taken as systematic un-
 953 certainty. The variation is found to be smaller than 0.2% for both electron and muon channels,
 954 both light flavour and heavy flavour scale factors, and for all mass hypotheses. Therefore it is
 955 considered negligible for this analysis.

956 9.2.11 Pileup modeling

957 The MC signal events are reweighted such that the number of pileup interactions matches with
 958 what present in the data. The number of pileup interactions in data can be estimated using
 959 measurements of instantaneous luminosity and knowing the total inelastic pp cross section at
 960 8 TeV. The uncertainties on these two quantities introduce an uncertainty on the number of
 961 pileup interactions considered in the simulation, and therefore on the signal efficiency. We
 962 followed the official recipe (which just amounts to changing the total inelastic cross-section by
 963 +/- 6%) and doing the analysis all over again. The relative change in the efficiency is found to
 964 be around 0.5% for all the resonance mass hypotheses.

965 9.3 Signal Shape

966 The following sources of systematic uncertainties affect the shape of the m_{WW} distribution for
 967 signal events. The m_{WW} signal shapes are used in the limit and significance calculation.

968 The uncertainties on the m_{WW} signal shape are evaluated at the same time of the uncertainties
 969 on the signal normalization described above (including the propagation of the effects to the
 970 E_T^{miss}). In all cases, for each variation (lepton or jet energy/resolution) we create the histograms
 971 of m_{WW} signal shapes for all the resonance mass hypotheses considered after the full selection.
 972 We then check the variation in the mean and rms of those histograms. The mean and rms are
 973 calculated in a 5rms range around the mean. We take the largest relative difference compared to
 974 nominal value for mean and rms of the histograms as systematic uncertainty on the mean and
 975 width of the DoubleCB function introduced in Section 7.1 to perform the m_{WW} signal shape
 976 fits. These shape uncertainties are valid for both HP and LP categories since we did not observe
 977 any significant variation in the m_{WW} shapes between the two cases.

978 9.3.1 Lepton energy/momentum scale

979 The uncertainty on the lepton energy/momentum scale introduces an uncertainty in the position
 980 and the width of the m_{WW} signal shape. A systematic uncertainty of < 0.1% (0.1%) on the
 981 mean (width) of the DoubleCB function describing the m_{WW} signal shape are used for the elec-
 982 tron channel. A systematic uncertainty of < 0.1% (0.5%) on the mean (width) of the DoubleCB
 983 function describing the m_{WW} signal shape are used for the muon channel.

984 9.3.2 Lepton energy/momentum resolution

985 The uncertainty on the lepton energy/momentum resolution introduces an uncertainty in the position
 986 and the width of the m_{WW} signal shape. A systematic uncertainty of < 0.1% (0.1%) on the
 987 mean (width) of the DoubleCB function describing the m_{WW} signal shape are used for the elec-
 988 tron channel. A systematic uncertainty of < 0.1% (0.1%) on the mean (width) of the
 989 DoubleCB function describing the m_{WW} signal shape are used for the muon channel.

990 9.3.3 Jet energy scale

991 The uncertainty on the jet energy scale introduces an uncertainty in the position and the width
 992 of the m_{WW} signal shape. A systematic uncertainty of 1.3% ($\sigma = 1.68\% + 5.3 \cdot 10^{-4}\% \cdot M_{\text{res.}}$) on

993 the mean (width) of the DoubleCB function describing the m_{WW} signal shape are used for both
 994 the electron and muon channels.

995 **9.3.4 Jet energy resolution**

996 The uncertainty on the jet energy resolution introduces an uncertainty in the position and the
 997 width of the m_{WW} signal shape. A systematic uncertainty of < 0.1% (3%) on the mean (width)
 998 of the DoubleCB function describing the m_{WW} signal shape are used for both the electron and
 999 muon channels.

1000 **9.3.5 Unclustered energy scale**

1001 We have evaluated this uncertainty with the following procedure: select all the jets (NOT used
 1002 for the JES and JER systematics) with $p_T < 10$ GeV and MuFraction < 0.9; for each event per-
 1003 form the vectorial sum of these jets; the new created jet is assumed to be the contribution from
 1004 the unclustered energy and a JEC uncertainty of 10% is assigned to it; the energy of this jet is
 1005 varied accordingly with its uncertainty and the MET is recalculated; the variation with respect
 1006 to the default value is used as systematic uncertainty. This studies showed that the uncertainty
 1007 on the unclustered energy has some impact on the width of the mWW distribution (3%-1% for
 1008 mass 600GeV-2500GeV), while it is negligible for the peak postion of the mWW distribution
 1009 and for the signal efficiency (< 1%).

1010 **9.3.6 E_T^{miss} scale and resolution**

1011 Uncertainties in the energy/momentum scale and resolution of leptons, jets in the event are
 1012 propagated to an uncertainty on the E_T^{miss} determination, and therefore introduce an uncer-
 1013 tainty in the m_{WW} signal shape. The procedure used to include the E_T^{miss} uncertainties is de-
 1014 scribed in the sub-sections above for each source of systematic uncertainty.

1015 Table 10 summarize the impact of each source of systematic uncertainty on both the signal
 1016 yields after full selection (both due to uncertainty on signal efficiency or integrated luminosity
 1017 of the data sample) and on the reconstructed m_{WW} signal shape. The numbers indicated in this
 1018 table are included in the limit and significance calculation for all the resonance mass hypotheses
 1019 and analysis categories considered, unless otherwise noted by a variable range.

Table 10: Summary of sources of systematic uncertainties and their impact on the number of signal events expected after full selection and on the reconstructed m_{WW} signal shape (peak position and width). The total uncertainty is reported as the sum in quadrature of the different terms. See text for more details.

Source of Syst. Uncert.	Number of Signal Events [%]	m_{WW} Signal Shape	
		Mean [%]	Width [%]
Electron energy (Muon momentum) scale	0.2 (0.7)	< 0.1	0.1 (0.5)
Electron energy (Muon momentum) resolution	< 0.1	< 0.1	0.1
Jet energy scale	1–3	1.3	2–3
Jet energy resolution	0.3	< 0.1	3
Unclustered energy scale	0.1	< 0.1	3–1
E_T^{miss} scale and resolution	Included in lepton/jet/unclust.en. uncert.		
Trigger	1	-	-
Electron (Muon) identification/isolation efficiency	3 (1)	-	-
W-tagging efficiency HP (LP)	9 (27)	-	-
b-tag identification efficiency	< 0.2	-	-
Pileup modeling	0.5	-	-
Integrated luminosity	4.4	-	-
Total Syst. Uncert. HP (LP)	10–11 (27–28)	1.3	4.5

DRAFT

10 Results

1021 In this section, we use the signal m_{WW} shapes presented in Section 7.1, the signal efficiency
 1022 presented in Section 7.2, the background parameterization introduced in Sections 8.1 and 8.2
 1023 and studied in Appendix B, and the systematic uncertainties discussed in Section 9 to perform
 1024 a search for new resonances in the m_{WW} spectrum of data selected accordingly with the criteria
 1025 summarized in Section 6.4.

1026 We present the upper limits on $X \rightarrow WW$ resonance cross section and the p-value/significance
 1027 of potential excesses of events observed in data, separately for the two background estimation
 1028 methods considered, i.e. Method 1 (α , Section 8.1) and Method 2 (Smoothness Test, Section 8.2).

1029 10.1 Tools for statistical analysis

1030 For the limits we followed the modified frequentest prescription described in [50, 51] (CL_S).
 1031 We use Asymptotic CL_S [52]. The results with full CL_S were cross-checked by the other group
 1032 working on this analysis and they are reported in Appendix of [1]. In summary, looking at
 1033 the total combination, observed and expected limits, derived with Asymptotic and full CLs,
 1034 agree within 2-3% for $m = 0.8, 1.2, 1.6$ TeV, while they disagree by 5% at 2.4 TeV. This variation
 1035 is however significantly smaller than the size of the 68% interval of the expected limit. For this
 1036 reason we think that the Asymptotic approximation is valid in this case.

1037 Systematic uncertainties are treated as nuisance parameters and profiled in the statistical in-
 1038 terpretation. The shape analysis that we pursued uses simultaneously the four categories of
 1039 events introduced in Section 6.3 (HP and LP in the electron and muon channel separately) us-
 1040 ing shapes and normalization predicted for the signal. We tested 16 different mass hypotheses
 1041 of the Bulk graviton model separated scanning the mass range $m_{WW} \in [1000, 2500]$ GeV at regu-
 1042 lar steps of 100 GeV. This choice covers the mass range where previous CMS analysis stopped
 1043 focusing on the interesting region above the TeV scale where the boosted topology is by far the
 1044 dominant one. For the implementation of the statistical analysis, we use the Higgs combination
 1045 tool [53] by calling the following example command:

- 1046 • Observed and expected limits:

```
1047 combine -M Asymptotic -n ${label} -m $mass -s $randomseed
1048 -d ${datacard}.txt -H ProfileLikelihood
1049 --rMax $maxBoundary --rMin $minBoundary
```

1050 In order to assess quantitatively the significance of the potential excesses of events observed
 1051 in the m_{WW} distributions of data, one can study the p-value as a function of the mass. The
 1052 p-value is defined as the probability for the background-only hypothesis to fluctuate by the
 1053 same amount (or more) of the observed data. We also use the Higgs combination tool for this
 1054 calculation by calling the following commands:

- 1055 • Observed significance:

```
1056 combine -M ProfileLikelihood -n ${label}ObsSignif -m $mass
1057 --signif --pvalue -d $WORKDIR/"${datacard}.txt"
```

- 1058 • Expected significance :

```
1059 combine -M ProfileLikelihood -n ${label}ExpSignif -m $mass
1060 -s $randomseed --signif --pvalue --expectSignal=1
1061 -t -1 --toysFreq -d $WORKDIR/"${datacard}.txt"
```

1062 Low p-values indicate large incompatibilities with the background-only hypothesis.

1063 10.2 Limits and p-values - Method 1

1064 The background estimation Method 1 is described in Section 8.1.

1065 Figure 34 shows the 95% CL upper limits on the cross section of the process $pp \rightarrow X \rightarrow WW$
1066 combining the four event categories as a function of the resonance mass. For comparison, the
1067 production cross sections times branching ratio of $G^* \rightarrow WW$ are presented for the benchmark
1068 Bulk graviton model [20] with values of the coupling strength $\tilde{k} = k/M_{Pl}$ equal to 0.2 and 0.5.
1069 We are not able to exclude any of these two specific models in the mass range considered. The
1070 p-values of the observation are presented in Fig. 35.

1071 Figure 36 (Figure 37) shows the limits (p-values) separately for the 4 categories considered
1072 in this analysis. Figure 38 (Figure 39) shows the limits (p-values) for the combination of the
1073 electron LP+HP (muon LP+HP) categories.

1074 By comparing the combination of the LP and HP categories for a given lepton flavour and
1075 the limits for the LP and HP separately, it can be appreciated how the sensitivity of the HP
1076 category is the dominant one. The LP one contributes only marginally to the final sensitivity of
1077 the analysis and only at high values of the resonance mass hypothesis (above 2 TeV).

DRAFT

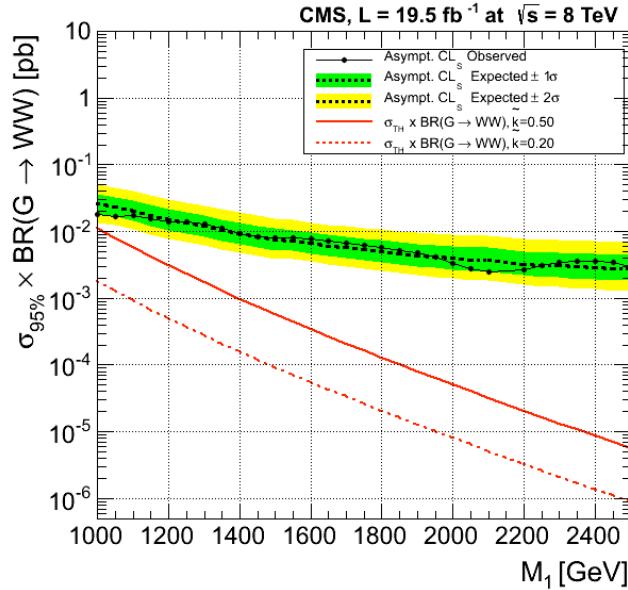


Figure 34: Method 1 - All analysis categories combined - Observed (dashed) and expected (solid) 95% CL upper limit on the product of the graviton production cross section and the branching fraction of $G^* \rightarrow WW$ using 19.5 fb^{-1} of data. The 68% and 95% ranges of expectation are also shown with green and yellow bands. The theoretical cross section times BR for $G^* \rightarrow WW$ is shown as a red dashed (solid) curve for $\tilde{k} = 0.2$ ($\tilde{k} = 0.5$).

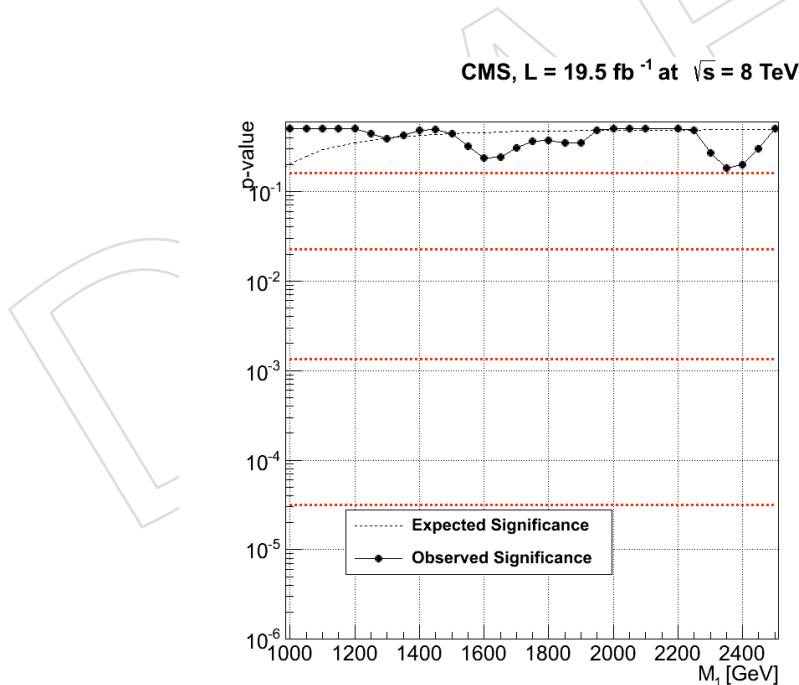


Figure 35: Method 1 - All analysis categories combined - Observed p-values as a function of the nominal signal mass (black dots and solid black line) using 19.5 fb^{-1} of data. The expected p-value in presence of a Bulk graviton ($\tilde{k} = 0.2$) is shown by the dashed line. Conversions of the p-value to number of standard deviations of a two-tail Gaussian are drawn as dashed horizontal red lines.

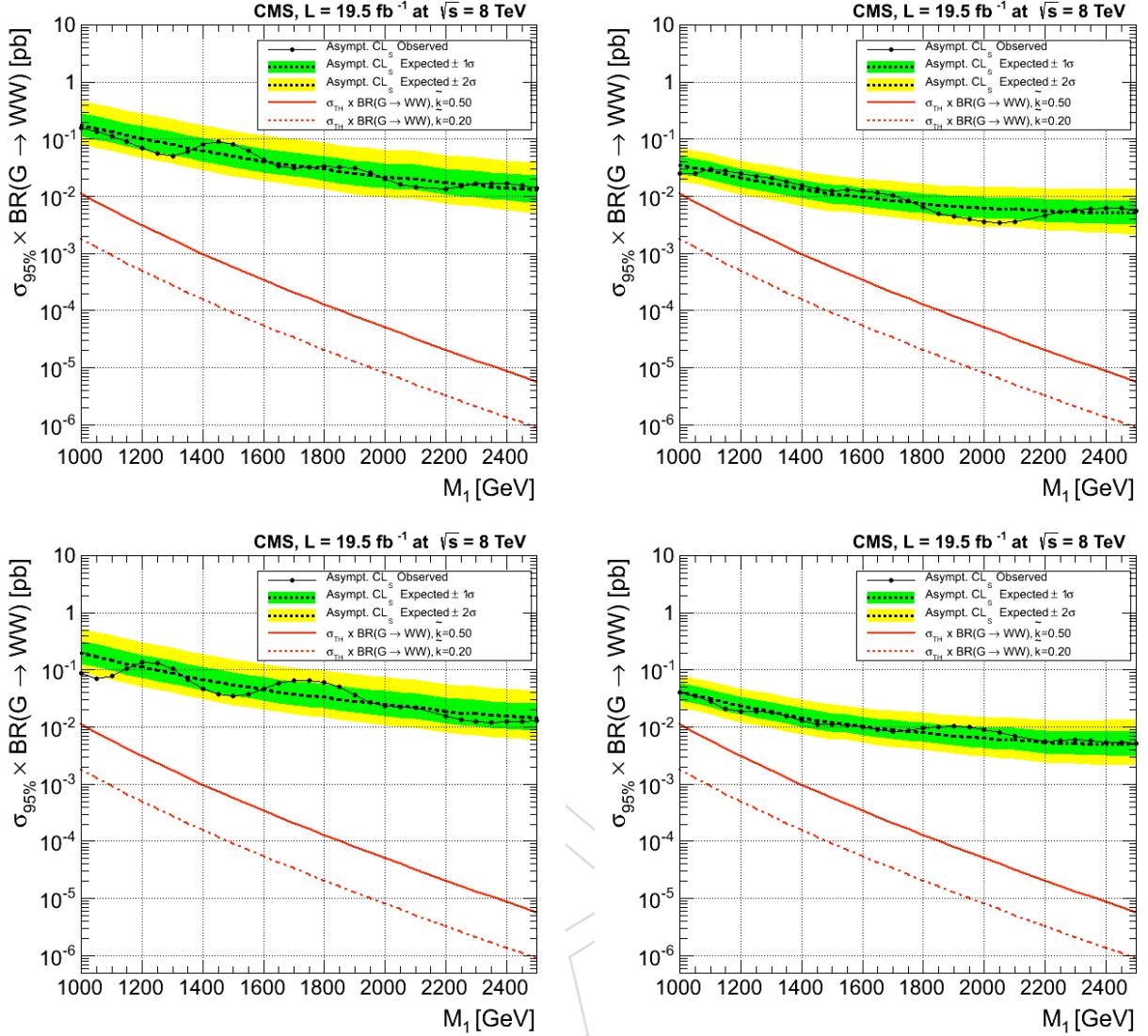


Figure 36: Method 1 - EleLP (top-left), EleHP (top-right), MuLP (bottom-left), MuHP (bottom-right) - Observed (dashed) and expected (solid) 95% CL upper limit on the product of the graviton production cross section and the branching fraction of $G^* \rightarrow WW$ using 19.5 fb^{-1} of data. The 68% and 95% ranges of expectation are also shown with green and yellow bands. The theoretical cross section times BR for $G^* \rightarrow WW$ is shown as a red dashed (solid) curve for $\tilde{k} = 0.2$ ($\tilde{k} = 0.5$).

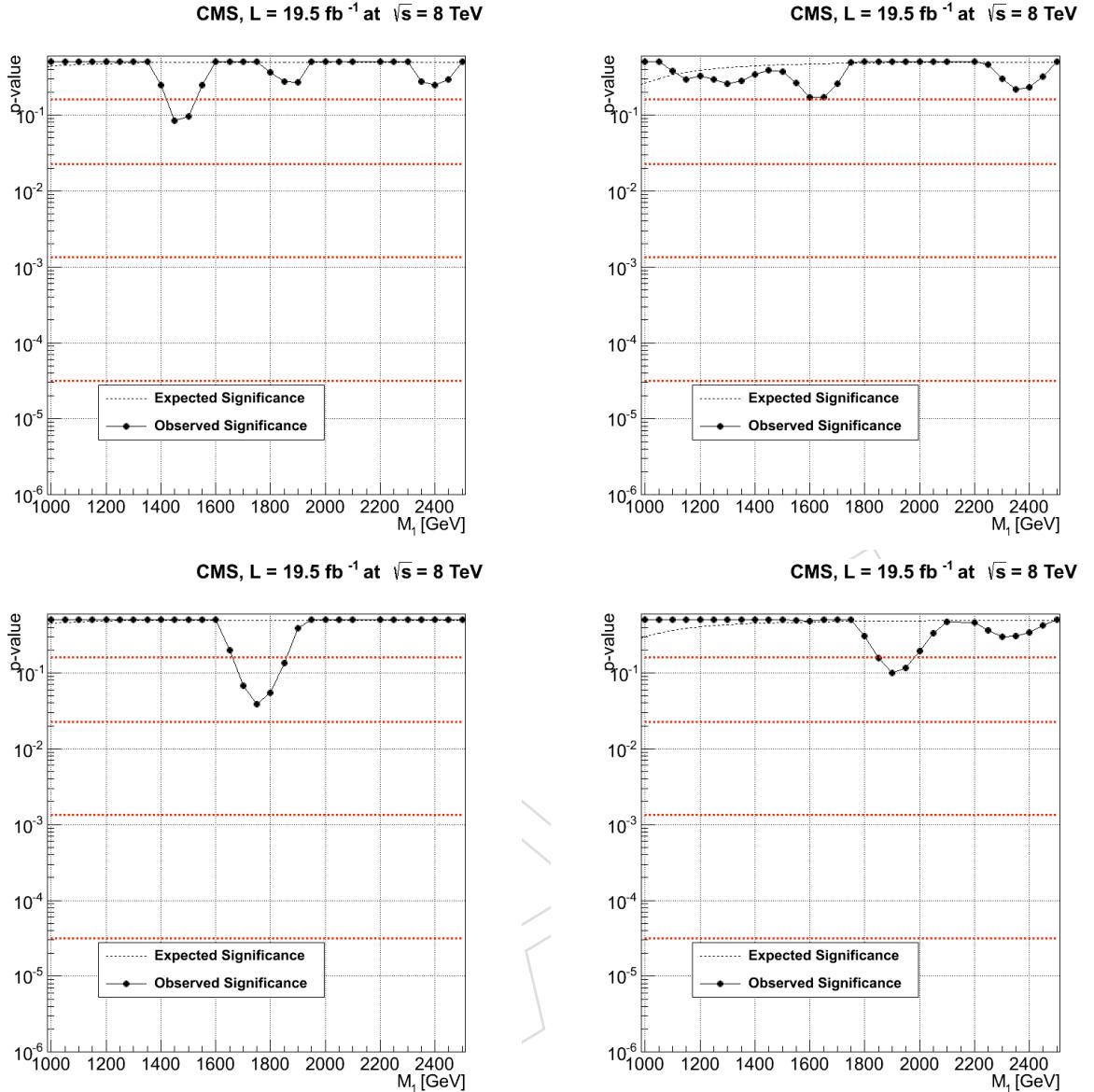


Figure 37: Method 1 - EleLP (top-left), EleHP (top-right), MuLP (bottom-left), MuHP (bottom-right) - Observed p-values as a function of the nominal signal mass (black dots and solid black line) using 19.5 fb^{-1} of data. The expected p-value in presence of a Bulk graviton ($\tilde{k} = 0.2$) is shown by the dashed line. Conversions of the p-value to number of standard deviations of a two-tail Gaussian are drawn as dashed horizontal red lines.

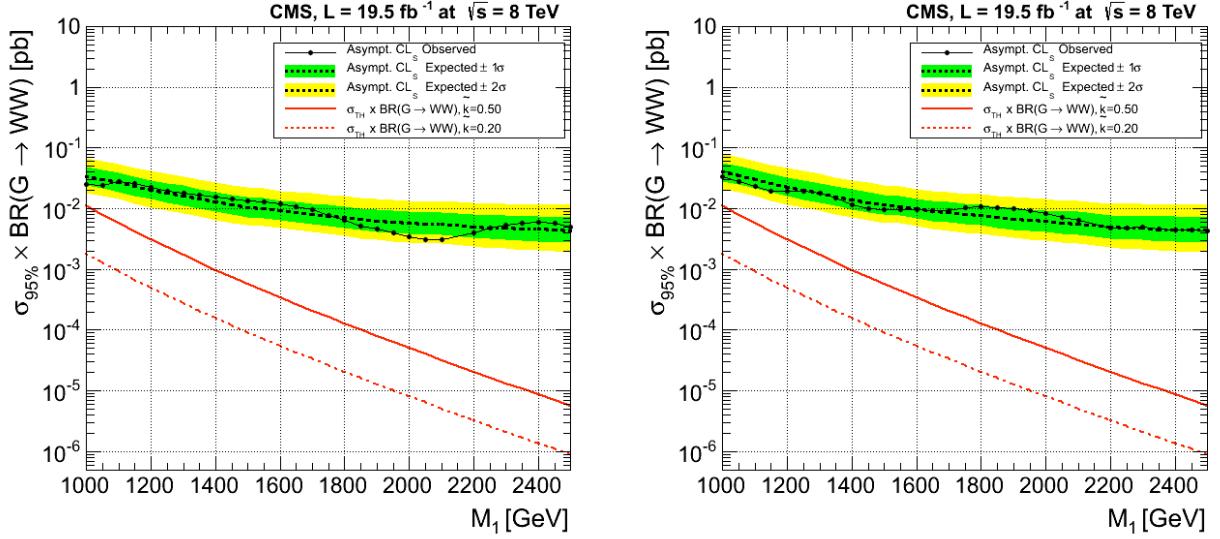


Figure 38: Method 1 - Ele LP+HP (left), Mu LP+HP (right) - Observed (dashed) and expected (solid) 95% CL upper limit on the product of the graviton production cross section and the branching fraction of $G^* \rightarrow WW$ using 19.5 fb^{-1} of data. The 68% and 95% ranges of expectation are also shown with green and yellow bands. The theoretical cross section times BR for $G^* \rightarrow WW$ is shown as a red dashed (solid) curve for $\tilde{k} = 0.2$ ($\tilde{k} = 0.5$).

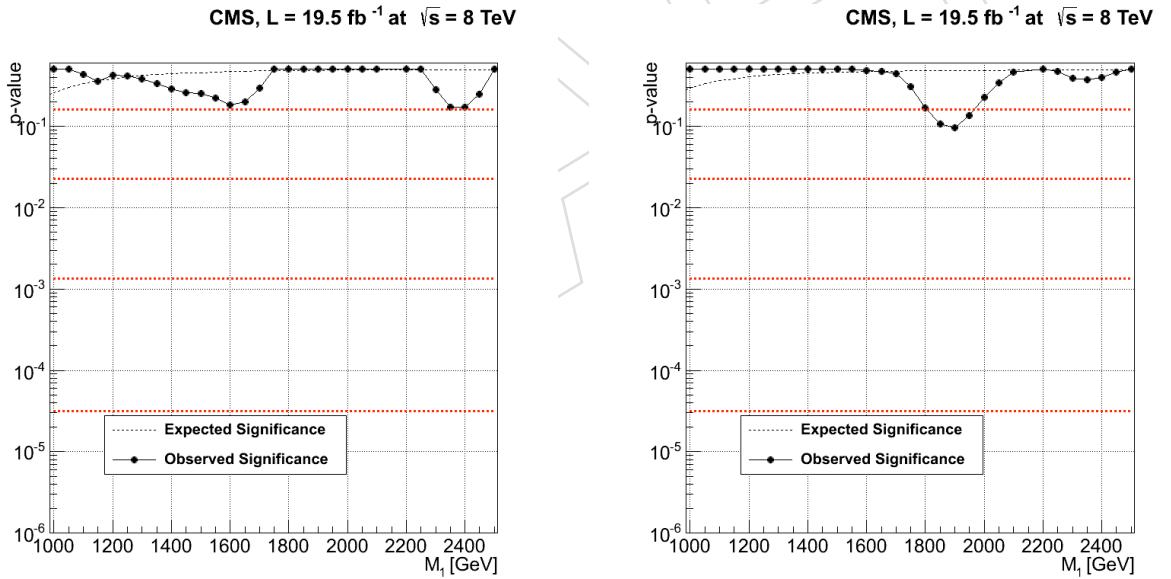


Figure 39: Method 1 - Ele LP+HP (left), Mu LP+HP (right) - Observed p-values as a function of the nominal signal mass (black dots and solid black line) using 19.5 fb^{-1} of data. The expected p-value in presence of a Bulk graviton ($\tilde{k} = 0.2$) is shown by the dashed line. Conversions of the p-value to number of standard deviations of a two-tail Gaussian are drawn as dashed horizontal red lines.

1078 10.3 Limits and p-values - Method 2

1079 The background estimation Method 2 is described in Section 8.2.

1080 Figure 40 shows the 95% CL upper limits on the cross section of the process $pp \rightarrow X \rightarrow WW$
1081 combining the four event categories as a function of the resonance mass. For comparison, the
1082 production cross sections times branching ratio of $G^* \rightarrow WW$ are presented for the benchmark
1083 Bulk graviton model [20] with values of the coupling strength $\tilde{k} = k/M_{Pl}$ equal to 0.2 and 0.5.
1084 We are not able to exclude any of these two specific models in the mass range considered. The
1085 p-values of the observation are presented in Fig. 41.

1086 Figure 42 (Figure 43) shows the limits (p-values) separately for the 4 categories considered
1087 in this analysis. Figure 44 (Figure 45) shows the limits (p-values) for the combination of the
1088 electron LP+HP (muon LP+HP) categories.

1089 By comparing the combination of the LP and HP categories for a given lepton flavour and
1090 the limits for the LP and HP separately, it can be appreciated how the sensitivity of the HP
1091 category is the dominant one. The LP one contributes only marginally to the final sensitivity of
1092 the analysis and only at high values of the resonance mass hypothesis (above 2 TeV).

DRAFT

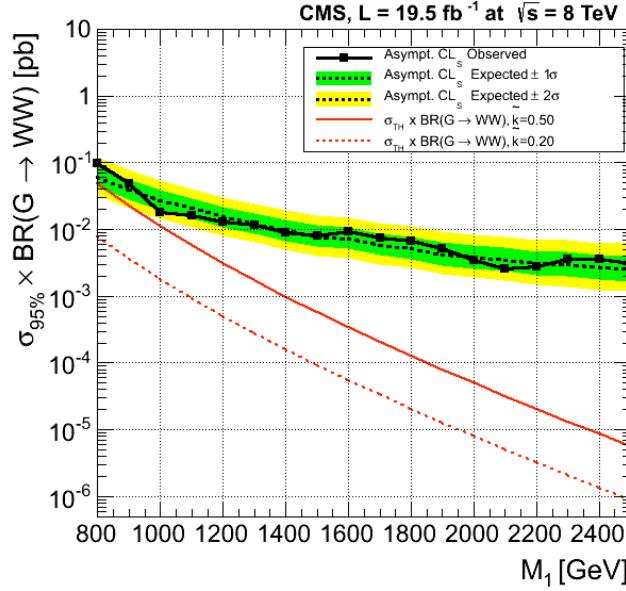


Figure 40: Method 2 - All analysis categories combined - Observed (dashed) and expected (solid) 95% CL upper limit on the product of the graviton production cross section and the branching fraction of $G^* \rightarrow WW$ using 19.5 fb^{-1} of data. The 68% and 95% ranges of expectation are also shown with green and yellow bands. The theoretical cross section times BR for $G^* \rightarrow WW$ is shown as a red dashed (solid) curve for $\tilde{k} = 0.2$ ($\tilde{k} = 0.5$).

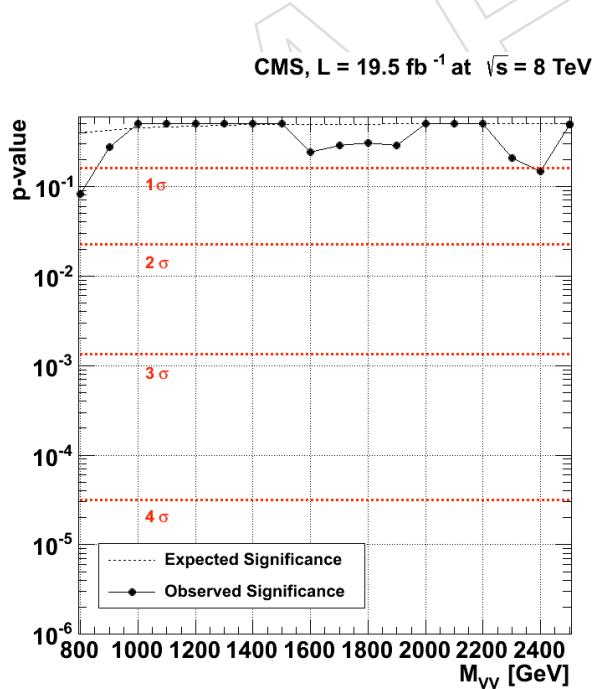


Figure 41: Method 2 - All analysis categories combined - Observed p-values as a function of the nominal signal mass (black dots and solid black line) using 19.5 fb^{-1} of data. The expected p-value in presence of a Bulk graviton ($\tilde{k} = 0.2$) is shown by the dashed line. Conversions of the p-value to number of standard deviations of a two-tail Gaussian are drawn as dashed horizontal red lines.

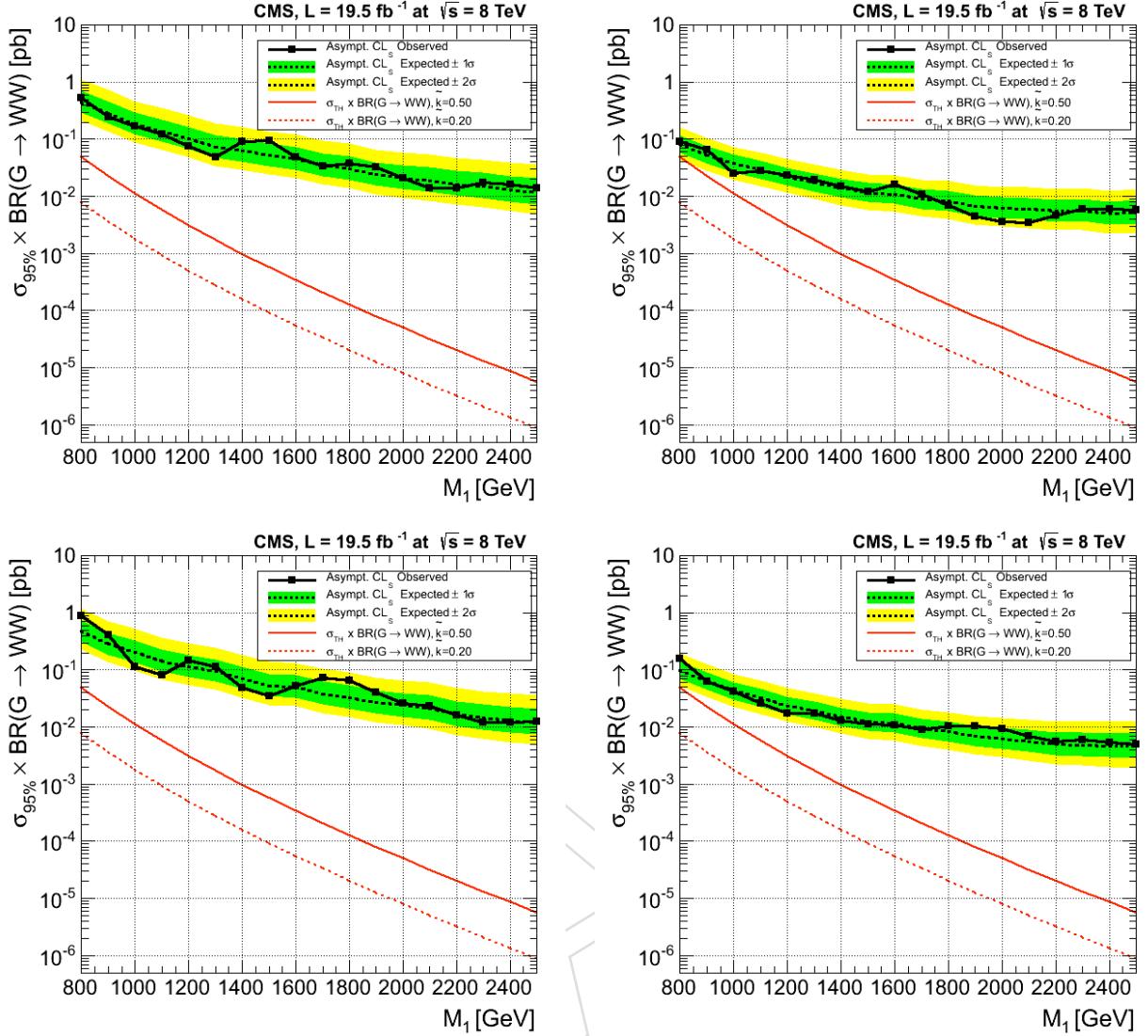


Figure 42: Method 2 - EleLP (top-left), EleHP (top-right), MuLP (bottom-left), MuHP (bottom-right) - Observed (dashed) and expected (solid) 95% CL upper limit on the product of the graviton production cross section and the branching fraction of $G^* \rightarrow WW$ using 19.5 fb^{-1} of data. The 68% and 95% ranges of expectation are also shown with green and yellow bands. The theoretical cross section times BR for $G^* \rightarrow WW$ is shown as a red dashed (solid) curve for $\tilde{k} = 0.2$ ($\tilde{k} = 0.5$).

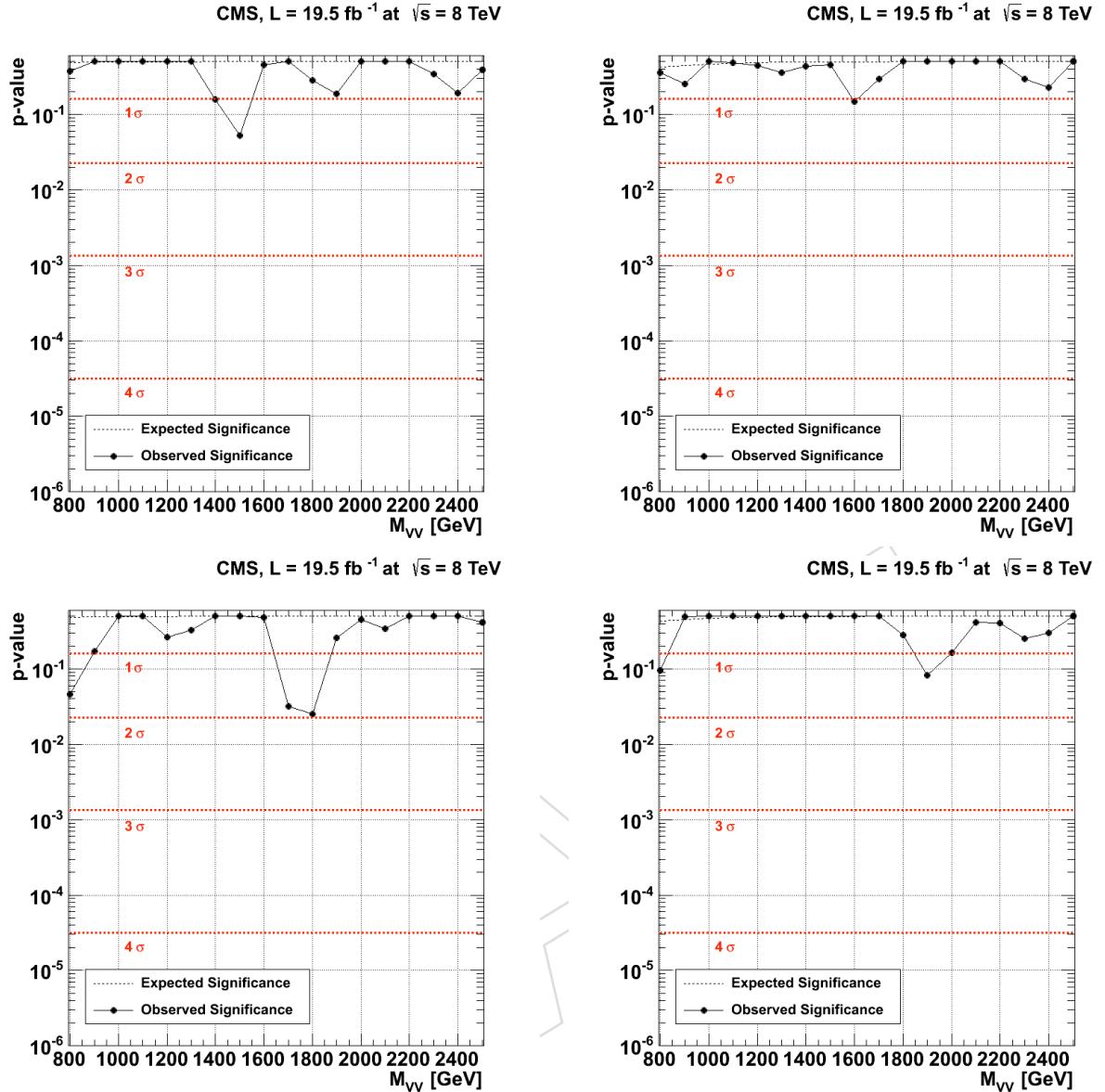


Figure 43: Method 2 - EleLP (top-left), EleHP (top-right), MuLP (bottom-left), MuHP (bottom-right) - Observed p-values as a function of the nominal signal mass (black dots and solid black line) using 19.5 fb^{-1} of data. The expected p-value in presence of a Bulk graviton ($\tilde{k} = 0.2$) is shown by the dashed line. Conversions of the p-value to number of standard deviations of a two-tail Gaussian are drawn as dashed horizontal red lines.

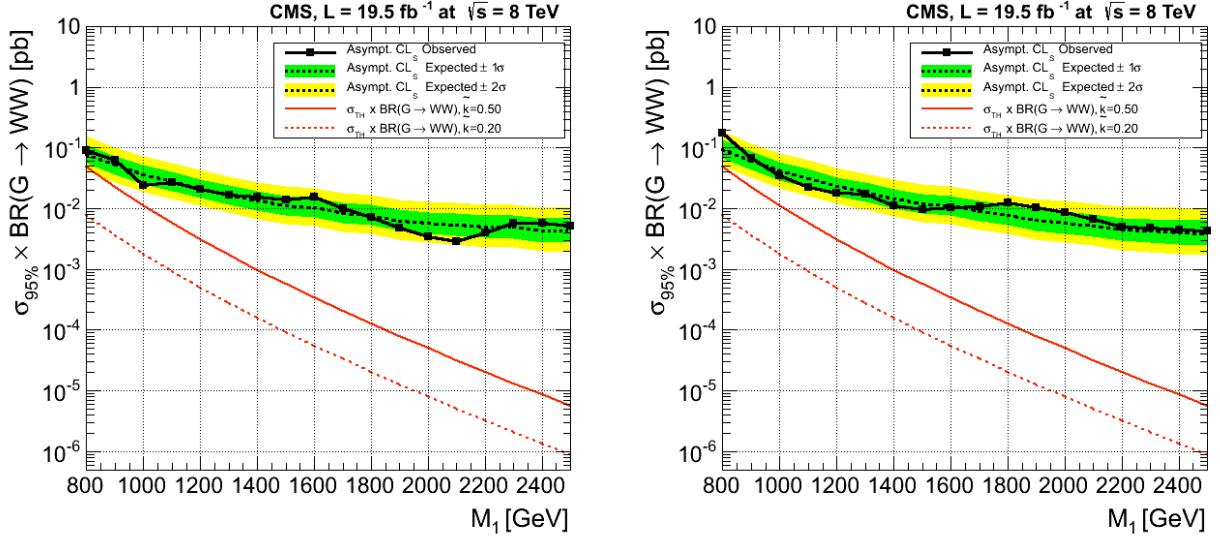


Figure 44: Method 2 - Ele LP+HP (left), Mu LP+HP (right) - Observed (dashed) and expected (solid) 95% CL upper limit on the product of the graviton production cross section and the branching fraction of $G^* \rightarrow WW$ using 19.5 fb^{-1} of data. The 68% and 95% ranges of expectation are also shown with green and yellow bands. The theoretical cross section times BR for $G^* \rightarrow WW$ is shown as a red dashed (solid) curve for $\tilde{k} = 0.2$ ($\tilde{k} = 0.5$).

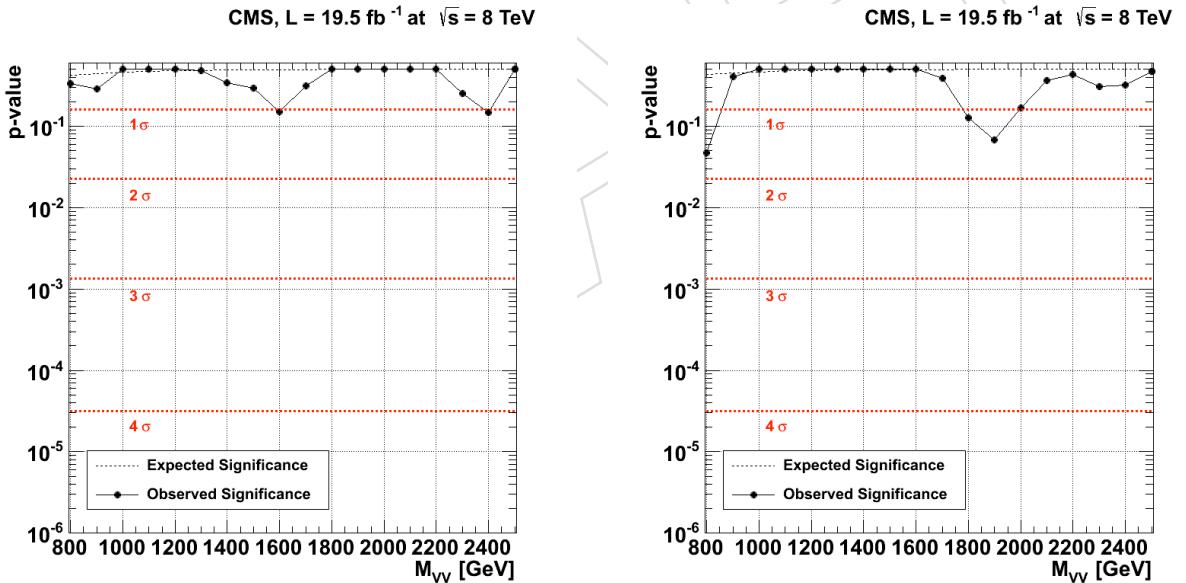
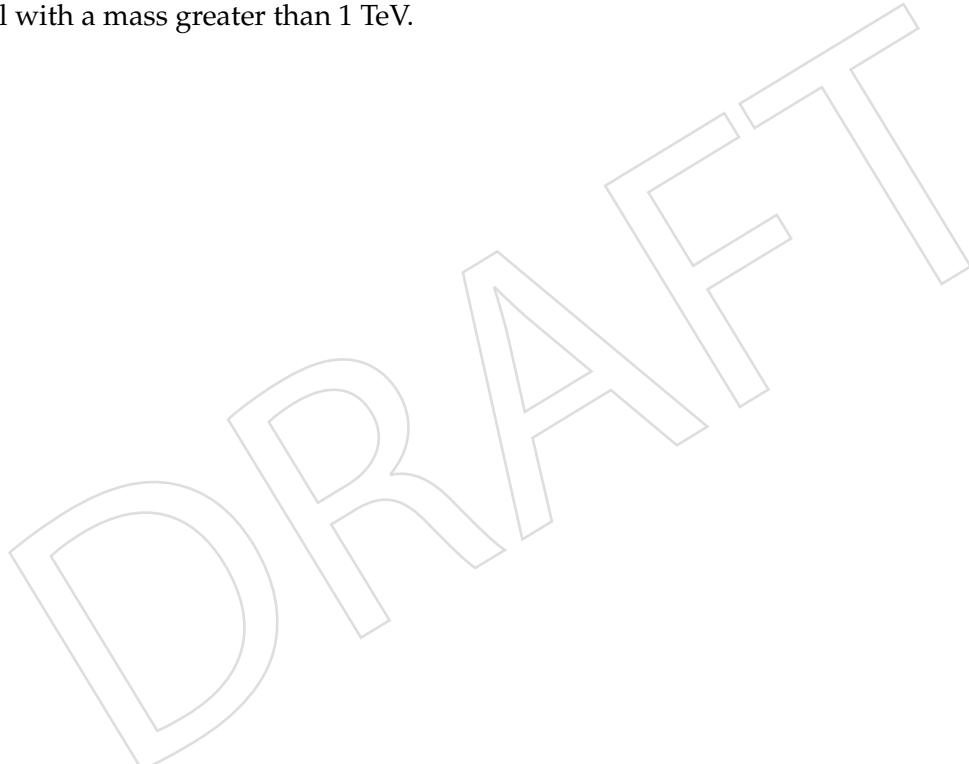


Figure 45: Method 2 - Ele LP+HP (left), Mu LP+HP (right) - Observed p-values as a function of the nominal signal mass (black dots and solid black line) using 19.5 fb^{-1} of data. The expected p-value in presence of a Bulk graviton ($\tilde{k} = 0.2$) is shown by the dashed line. Conversions of the p-value to number of standard deviations of a two-tail Gaussian are drawn as dashed horizontal red lines.

1093 11 Conclusions

- 1094 A search for new particles decaying to two W bosons with subsequent decay to a final state
1095 containing one charge lepton, one neutrino, and two quarks, $X \rightarrow WW \rightarrow q\bar{q}\ell\nu$, is presented.
1096 Results are based on data corresponding to an integrated luminosity of 19.5 fb^{-1} of proton-
1097 proton collisions at $\sqrt{s} = 8 \text{ TeV}$ and collected with the CMS detector at the CERN LHC.
- 1098 Jet sub-structure techniques are exploited for separating the signal from the SM background
1099 when the boost of the W causes the two quarks to merge into the same jet reconstructed in the
1100 detector. Two different data-driven background estimate techniques have been used providing
1101 consistent results.
- 1102 No excess above the expected background is found. We set upper limits at 95% CL on the
1103 $G^* \rightarrow WW$ cross section, which can be interpreted as generic limits on narrow resonances X
1104 decaying into the same final state. The integrated luminosity of the sample is not large enough
1105 to allow to set limits on Bulk Graviton models with $\tilde{k} = 0.2$ or $\tilde{k} = 0.5$.
- 1106 This analysis represents the first search in CMS for $X \rightarrow WW$ resonances in the semi-leptonic
1107 channel with a mass greater than 1 TeV.

A large, faint watermark is printed diagonally across the page. It contains the word "DRAFT" in a bold, sans-serif font, with each letter's top stroke pointing towards the bottom-left corner of the page. The watermark is light gray and serves as a draft indicator.

1108 A B-tag veto optimization

1109 Each event passing the final selection is required to have no AK5 jets (passing the selection
 1110 in Section 4.4) tagged as b-jet accordingly with the CSV algorithm, medium working point
 1111 (CSVM). The reason for this choice is discussed below.

1112 We considered different options for the btag veto using the CSV algorithm:

- 1113 • different working points (loose L, medium M, tight T);
- 1114 • different jet selection: either all the selected AK5, or only those that are far from the
 1115 W hadronic candidate in the event by $\Delta R = 0.8$ (the latter option is identified by the
 1116 label “clean” in the following plots).

Similarly to what is done for the τ_{21} optimization in Section 6.3, we choose as a figure of merit the quantity suggested in Ref. [44] (“Punzi’s significance”):

$$P = \frac{\epsilon_S}{1 + \sqrt{B}} \quad (21)$$

1117 where ϵ_S is the signal selection efficiency and B is the remaining background after all the cuts
 1118 have been applied. The signal and background is evaluated in a region of $\pm 15\%$ around the
 1119 target resonance mass. The studies are performed for LP+HP categories combined, since the
 1120 presence of bjets in the event is independent on the τ_{21} value of the leading jet.

1121 The significance calculated for the different b-tag veto options and for different resonance
 1122 masses (0.8, 1, 1.5, and 2 TeV) is reported in Figures 46 and 47. For all resonance masses in-
 1123 vestigated, the optimal working point is found to be the CSVM (using all selected AK5 jets).

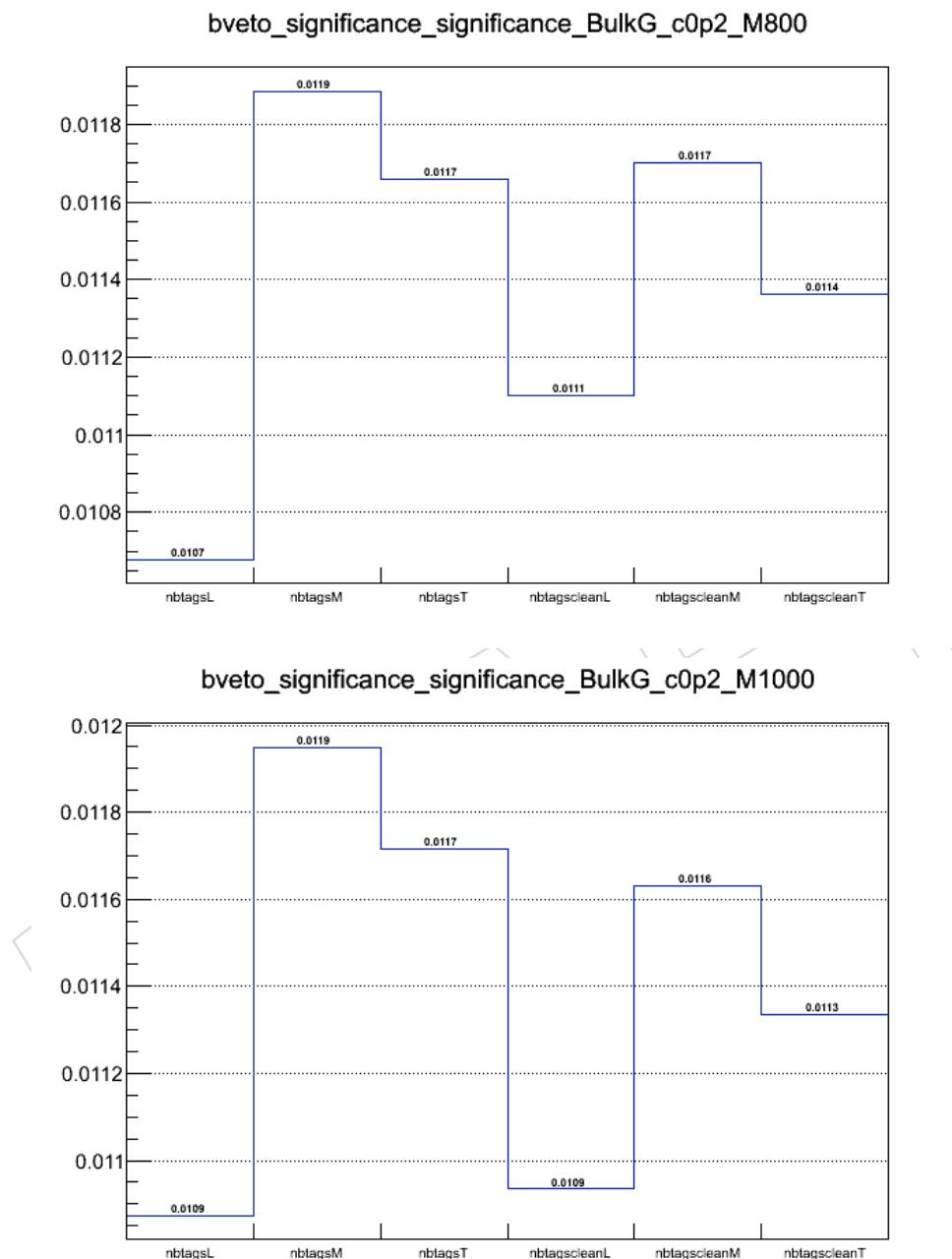


Figure 46: Punzi significance for various choices of b-tag veto and for different resonances masses (0.8 TeV top, 1 TeV bottom).

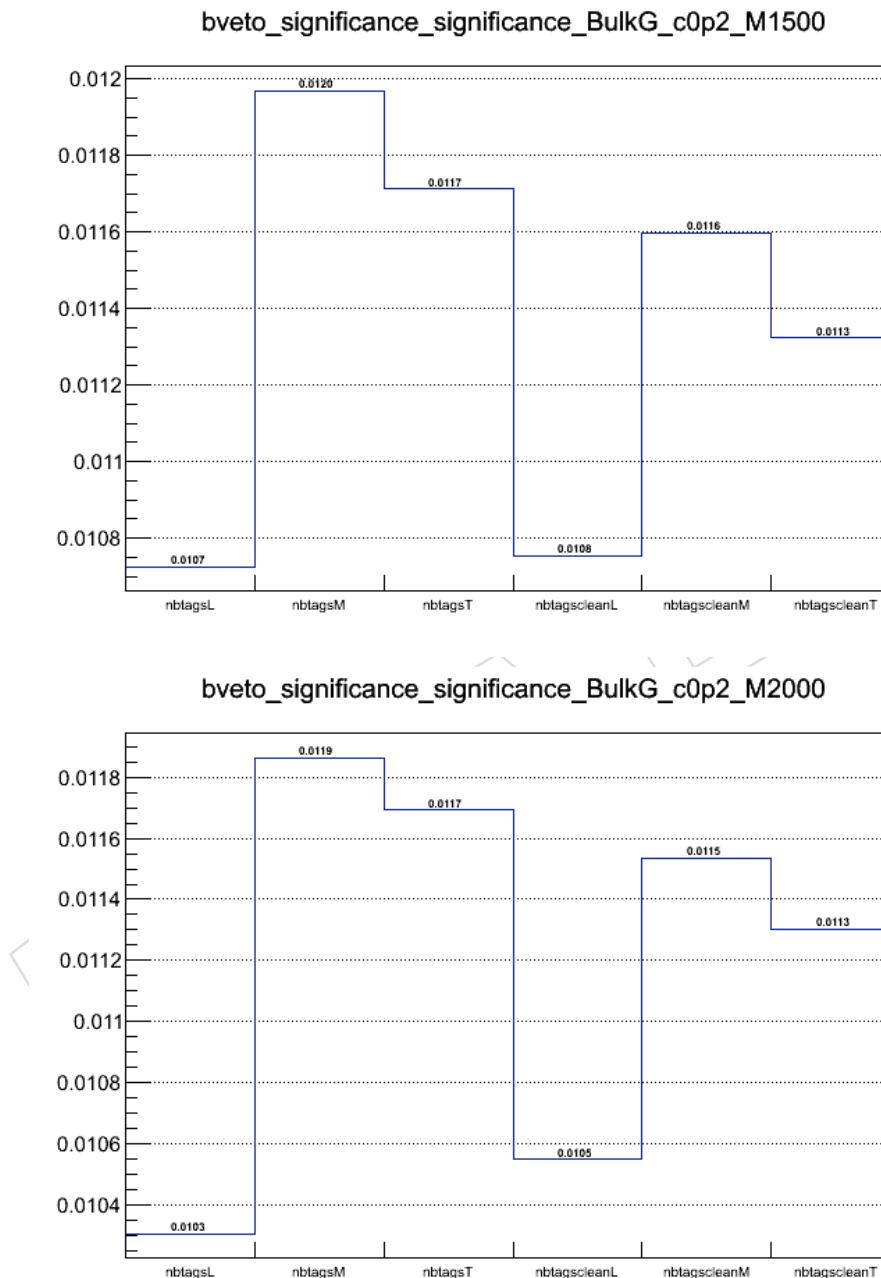


Figure 47: Punzi significance for various choices of b-tag veto and for different resonances masses (1.5 TeV top, 2 TeV bottom).

1124 B F-test for background fit function

1125 In this section we report studies on the choice of the fit function for the background estimation
 1126 method described in Section 8.2 (Method 2 - m_{WW} smoothness test). The studies are performed
 1127 using both MC (total background rescaled to integrated luminosity of the data) and data.

We take the m_{WW} distributions (in both MC and data) in the signal region after the final selection for each of the 4 categories considered (electron HP/LP, muon HP/LP). For each category we perform an unbinned fit using the following leveled-exponential parameterization ¹¹:

$$f(m_{WW}) = \text{Norm.} \cdot e^{-\frac{m_{WW}}{p_0 + p_1 m_{WW} + p_2 m_{WW}^2}} \quad (22)$$

1128 We perform 3 different set of fits with different number of floating parameters starting from a
 1129 reconstructed WW mass of 700 GeV:

- 1130 • **par1 function** - 1 slope parameter : p_0 floating, $p_1 = p_2 = 0$;
- 1131 • **par2 function** - 2 slope parameters : p_0 and p_1 floating, $p_2 = 0$;
- 1132 • **par3 function** - 3 slope parameters : p_0 , p_1 , and p_2 floating.

1133 The fits to the data and the MC (total background) for the 4 categories are shown in Figures 48,
 1134 and 49.

1135 Overall, the fit quality for the data and MC is reasonably good.

1136 Qualitatively from the plots, the par2 and par3 exponential functions provide a better description
 1137 of the tails than the par1 exponential function. Still, we want to derive (in a quantitative
 1138 way) how many free parameters are effectively needed in the fit function for properly describing
 1139 the background.

1140 This can be done using an F-test, as suggested by the CMS Statistics Committee during the
 1141 review of a similar search in fully hadronic final state [54]. The generic procedure is described
 1142 below.

- 1143 • One considers two models, i.e. model1 and model2, where model1 is “nested” with
 1144 the model2. That is, model 1 has n1 parameters, and model 2 has n2 parameters,
 1145 where $n2 > n1$, and for any choice of parameters in model 1, the same regression
 1146 curve can be achieved by some choice of the parameters of model 2. The model
 1147 with more parameters will always be able to fit the data at least as well as the model
 1148 with fewer parameters (within statistical uncertainties on the fit parameters). Thus
 1149 typically model 2 will give a better (i.e. lower error) fit to the data than model 1. But
 1150 one often wants to determine whether model 2 gives a *significantly* better fit to the
 1151 data.
- If there are N data points to estimate parameters of both models from, then one can
 calculate the F statistic, given by

$$F_{obs}^{21} = \frac{\frac{RSS_1 - RSS_2}{n2 - n1}}{\frac{RSS_2}{N - n2}} \quad (23)$$

1152 where RSS_i is the residual sum of squares of model i ($\sum_{bins} (data_i - fit_i)^2$, skipping
 1153 bins with 0 entries, and with fit_i value calculated integrating the function over the
 1154 bin range and dividing by the bin width).

¹¹From this functional form, one can see that for the values of the fit parameters $p_1 = p_2 = 0$, one recovers the simple exponential fit.

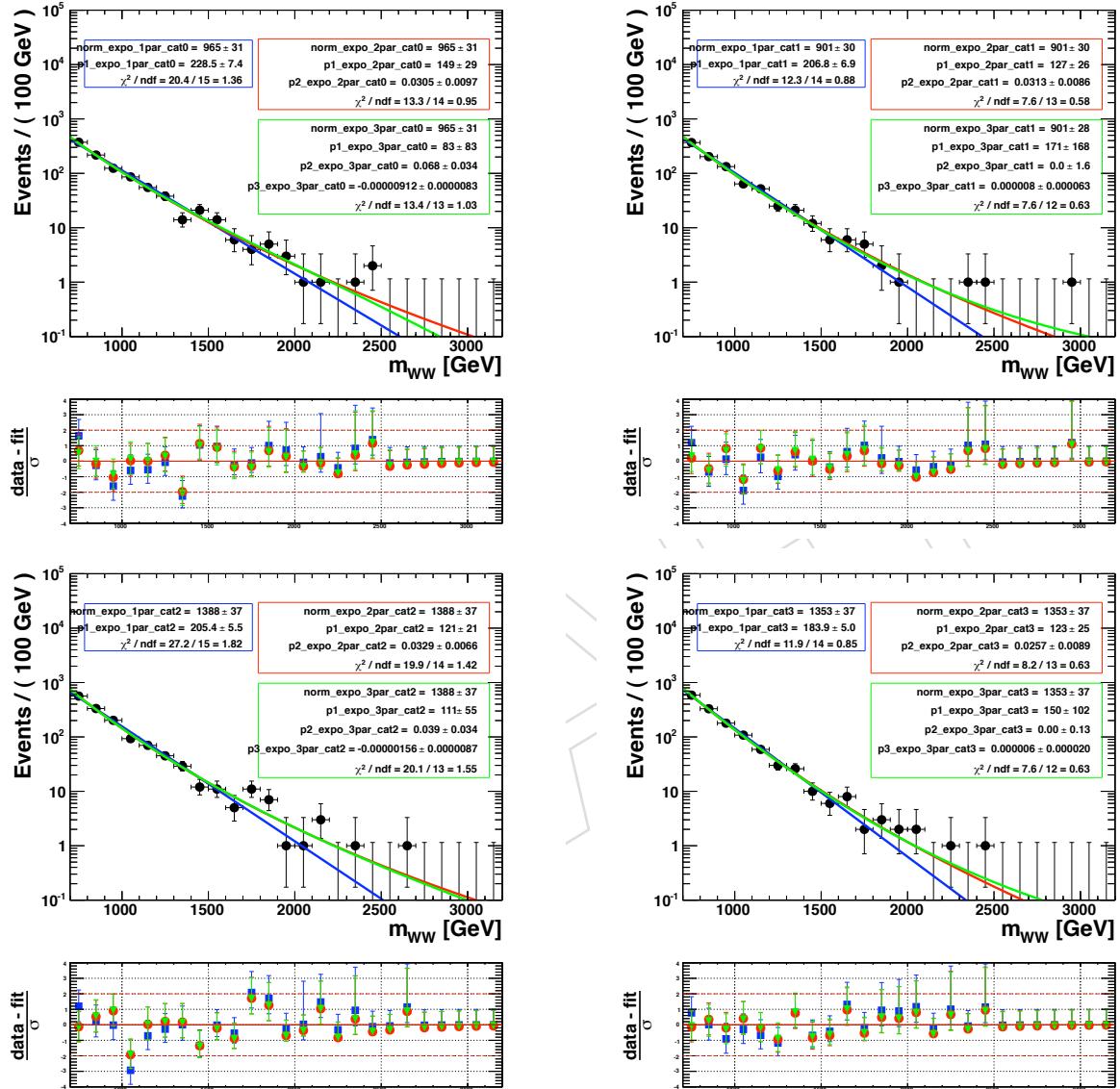


Figure 48: m_{WW} distributions for data in the signal region after full selection for the four event categories used in this analysis: electron LP (top-left), electron HP (top-right), muon LP (bottom-left), muon HP (bottom-right). Background-only fits using 1st-, 2nd-, and 3rd- order levelled exponential are shown, together with the pull distribution.

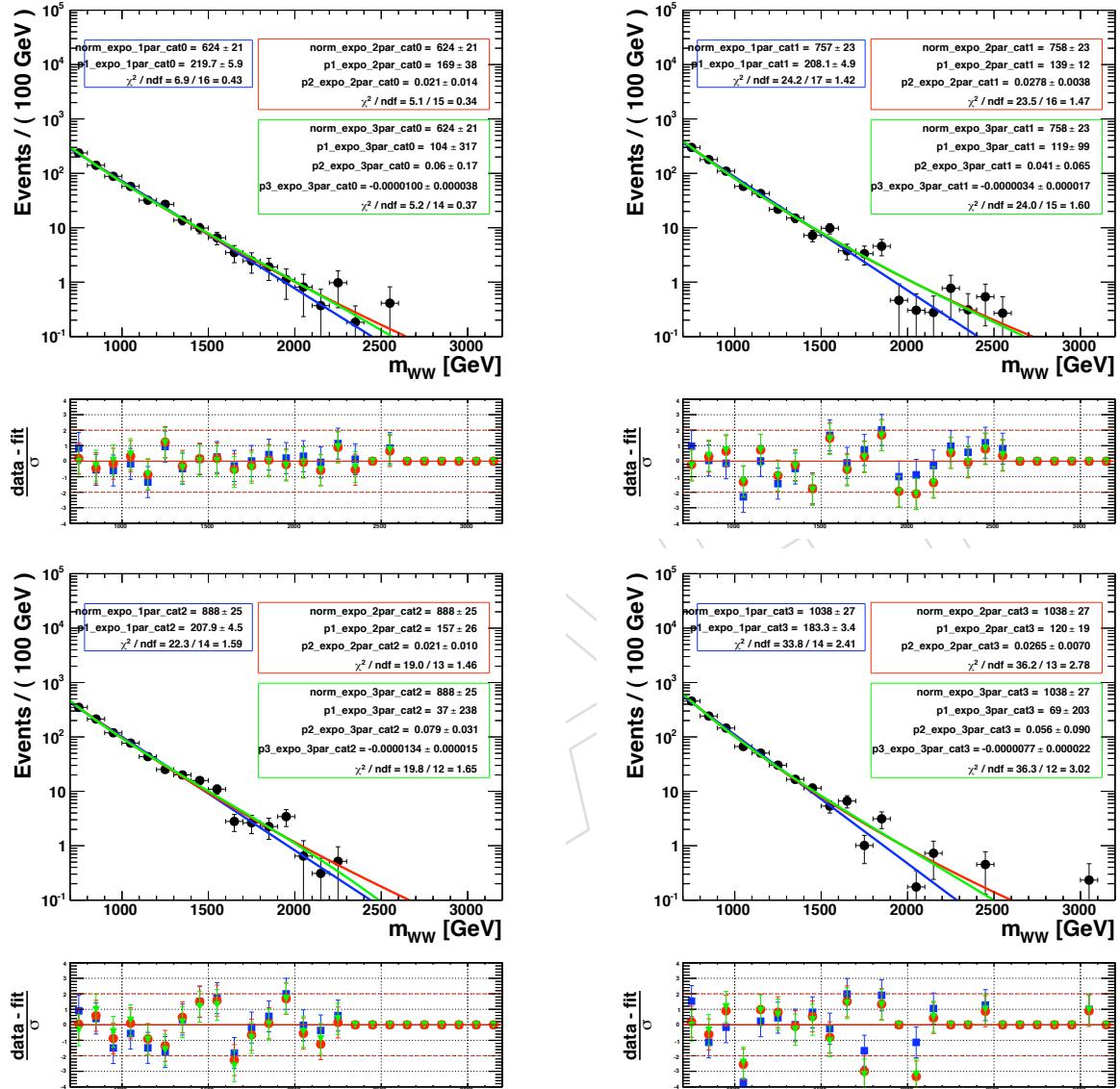


Figure 49: m_{WW} distributions for MC (total background) in the signal region after full selection for the four event categories used in this analysis: electron LP (top-left), electron HP (top-right), muon LP (bottom-left), muon HP (bottom-right). Background-only fits using 1st-, 2nd-, and 3rd- order levelled exponential are shown, together with the pull distribution.

- 1155 • Under the null hypothesis that model 2 does not provide a significantly
 1156 better fit than model 1, the F variable will have and F -distribution with $(n_2 - n_1, N - n_2)$
 1157 degrees of freedom. The null hypothesis is rejected if the observed con-
 1158 fidence level, $CL_{obs}^{21} = 1 - \int_{-\infty}^{F_{obs}^{21}} F\text{-distribution}(n_2 - n_1, N - n_2)$, is smaller than a
 1159 desired probability α (set to 0.05).

1160 In the particular case of this analysis, we use the F-test recursively to understand how many
 1161 parameters are needed to fit the data.

- 1162 1. Set `model1 = par1` function and `model2 = par2` function. Calculate F_{obs}^{21} and CL_{obs}^{21} . If
 1163 $CL_{obs}^{21} > \alpha \rightarrow$ `par1` is sufficient to describe the data, else go to 2)
 1164 2. Set `model1 = par2` function and `model2 = par3` function. Calculate F_{obs}^{32} and CL_{obs}^{32} . If
 1165 $CL_{obs}^{32} > \alpha \rightarrow$ `par2` is sufficient to describe the data, else go to 3)
 1166 3. etc.. until the condition $CL > \alpha$ is satisfied. However, in our case, we stop here and
 1167 just say that “at least 3 parameters” are needed to described the data (since we did not
 1168 consider functions with more than 3 slope parameters in this study).

1169 The results of the F-test are summarize in Table 11 for data and MC, for the 4 categories, and for
 1170 two different binning settings for the m_{WW} variable (50 GeV or 100 GeV). The results suggest
 1171 that the `par2` function (with 2 slope parameters) is sufficient to describe the data and the MC
 1172 background in all cases, with a very few exceptions over the phase space considered in this
 1173 study.

1174 We decide to use the `par2` function (2 slope parameters) to describe the background in all the 4
 1175 categories considered.

Table 11: Results of F-test for background fit function.

Type	Binning [GeV]	Value	ELE LP	ELE HP	MU LP	MU HP
DATA	50	$F_{obs}^{21} =$ $F_{obs}^{32} =$ $CL_{obs}^{21} =$ $CL_{obs}^{32} =$ Result	1.12399 -1.2979 0.29812 1 par1	18.8009 -0.541061 0.000224766 1 par2	0.0821246 -0.706755 0.776626 1 par1	4.61824 1.85274 0.041118 0.185609 par2
DATA	100	$F_{obs}^{21} =$ $F_{obs}^{32} =$ $CL_{obs}^{21} =$ $CL_{obs}^{32} =$ Result	32.52 4.70039 5.4604e-05 0.05 par2	25.07 -1.32545 0.000239972 1 par2	21.1344 -0.378005 0.000414651 1 par2	38.8384 2.44742 3.05813e-05 0.143694 par2
MC	50	$F_{obs}^{21} =$ $F_{obs}^{32} =$ $CL_{obs}^{21} =$ $CL_{obs}^{32} =$ Result	15.6475 -0.318859 0.00045115 1 par2	8.57912 -2.21259 0.00669063 1 par2	19.2861 -0.652957 0.000145924 1 par2	35.0135 0.475678 3.04864e-06 0.496739 par2
MC	100	$F_{obs}^{21} =$ $F_{obs}^{32} =$ $CL_{obs}^{21} =$ $CL_{obs}^{32} =$ Result	28.7664 3.7947 7.89586e-05 0.0717518 par2	23.2805 -1.27424 0.000186583 1 par2	20.45 -3.11277 0.000573562 1 par2	27.2775 0.905574 0.000164445 0.360066 par2

1176 C $t\bar{t}$ control plots

1177 Plots from the $t\bar{t}$ control region defined in Section 8.1.1.

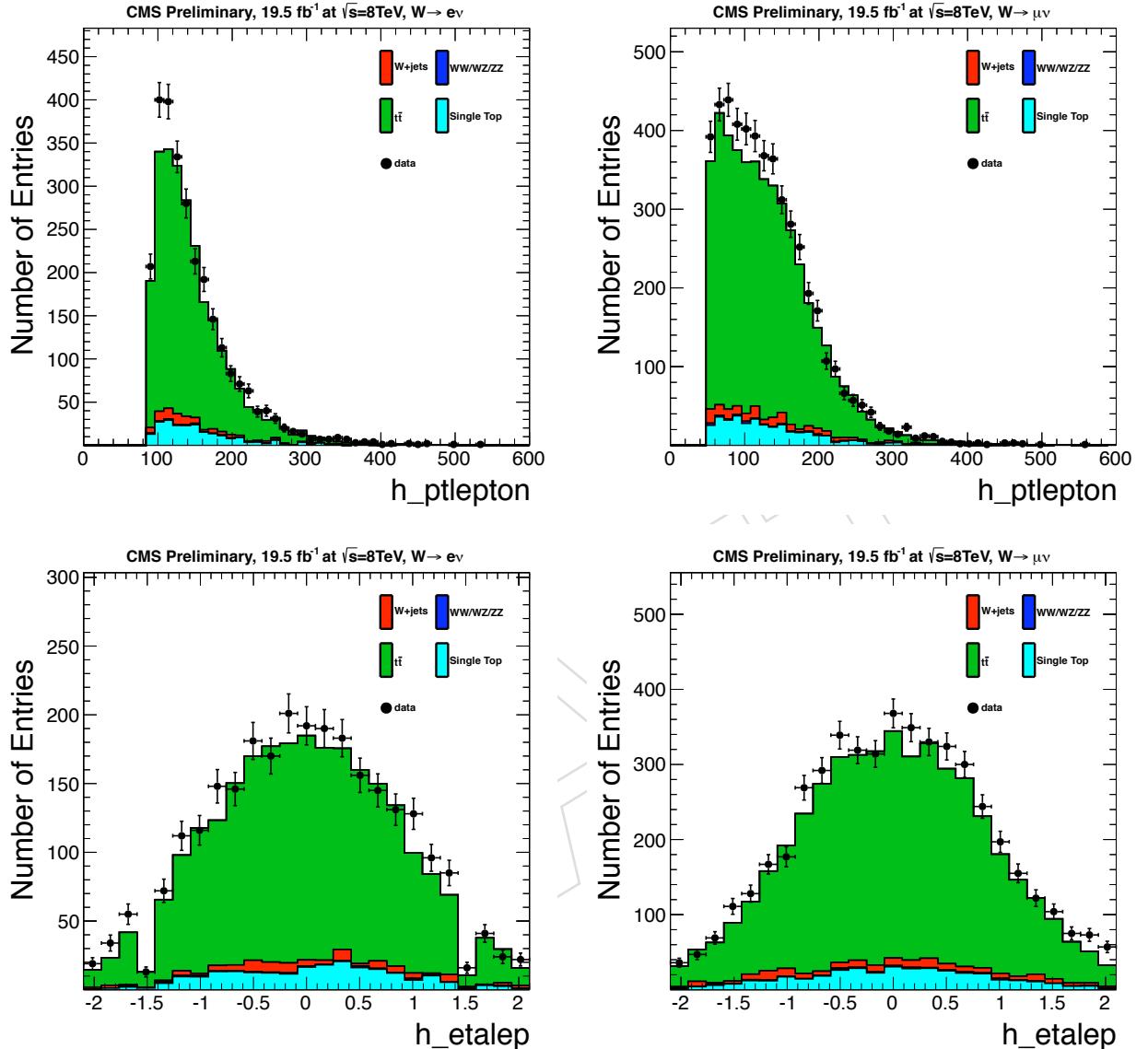


Figure 50: Lepton p_T and η for electron channel (left) and muon channel (right) for events with $40 < m_{jet}^{pruned} < 130 \text{ GeV}$ in the top control sample.

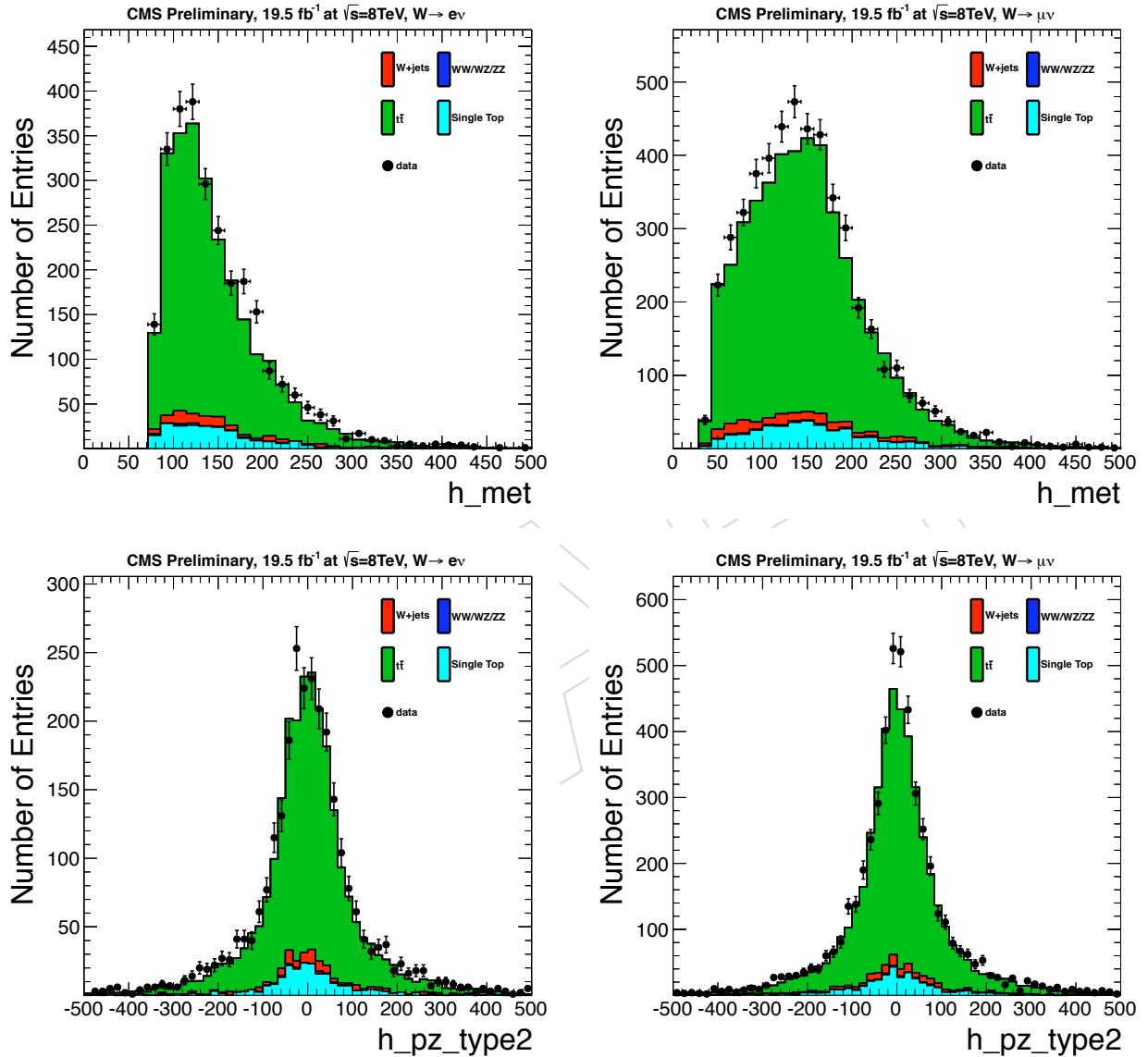


Figure 51: E_T^{miss} and $p_{z\nu}$ (defined in Section 5.1) for electron channel (left) and muon channel (right) for events with $40 < m_{\text{jet}}^{\text{pruned}} < 130$ GeV in the top control sample.

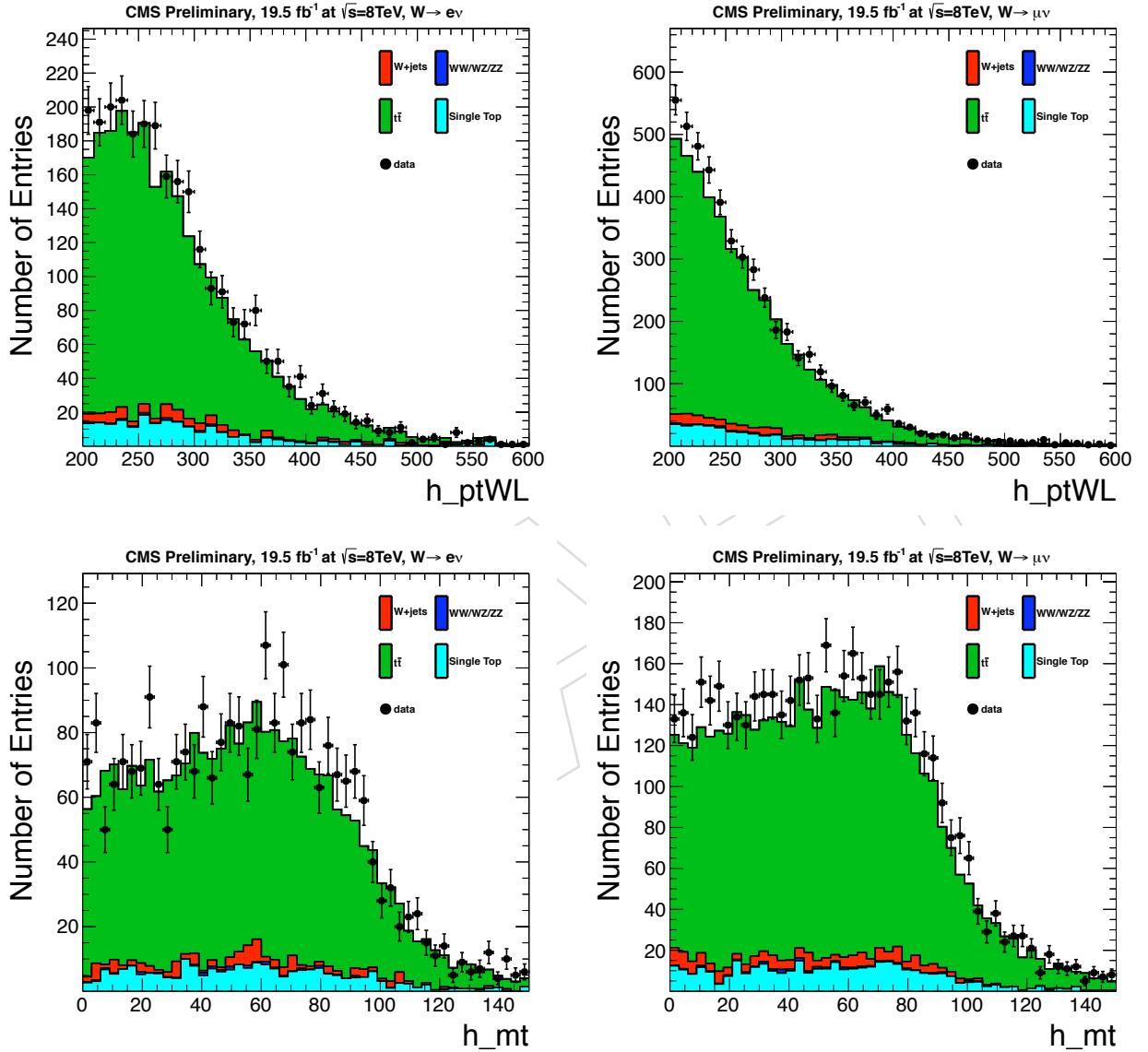


Figure 52: Leptonic $W p_T$ and $m_T(\ell\nu)$ for electron channel (left) and muon channel (right) for events with $40 < m_{jet}^{pruned} < 130$ GeV in the top control sample.

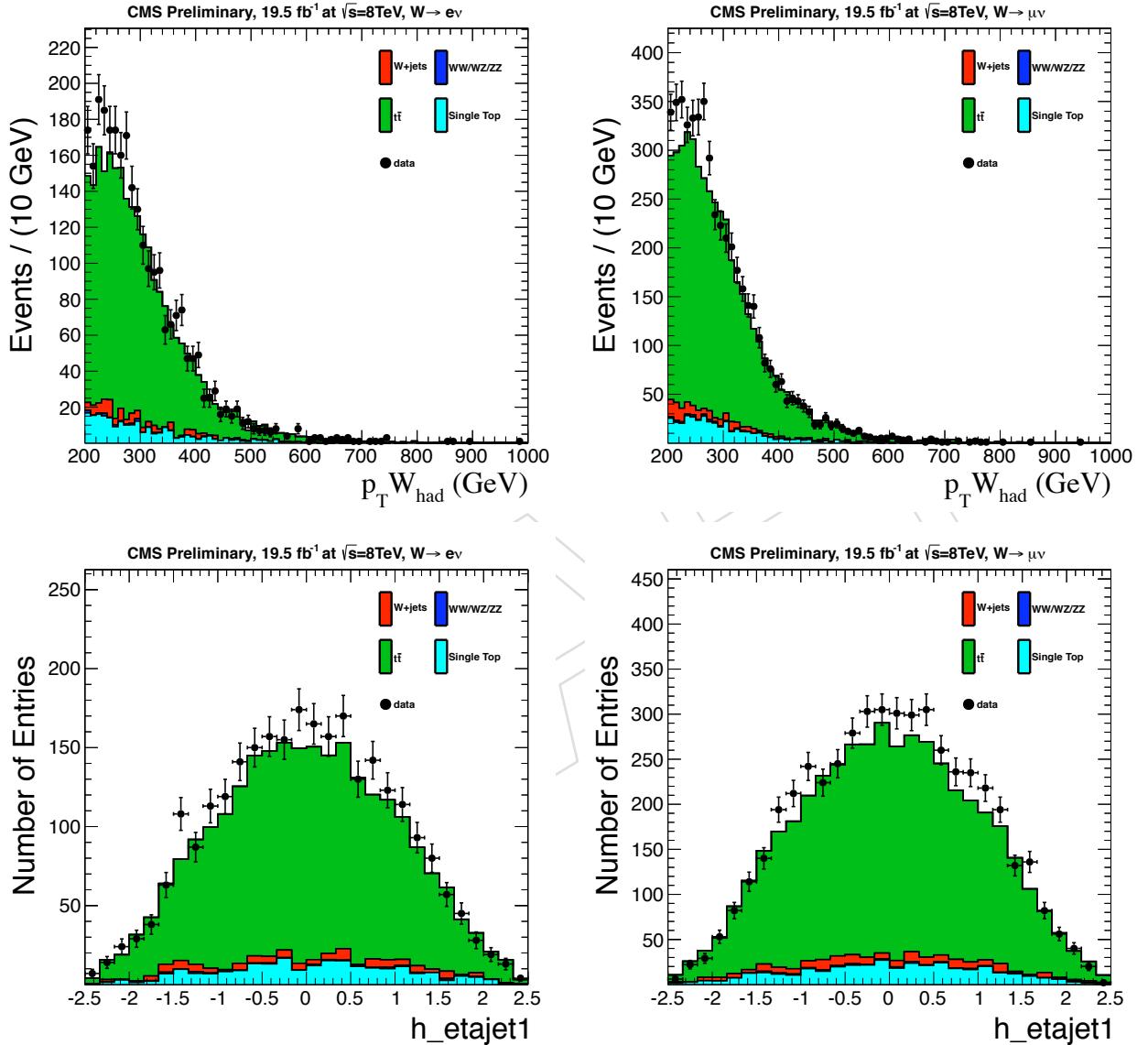


Figure 53: Hadronic W p_T and η for electron channel (left) and muon channel (right) for events with $40 < m_{jet}^{pruned} < 130$ GeV in the top control sample.

1178 D Closure test for Method 1

1179 We perform a validation of the background estimation method described in Section 8.1 (Method
 1180 1, α), in particular in Section 8.1.3 for the W+jets background estimation.

1181 In the main analysis, we use the data in the low-mass sideband region ($40 < m_{jet}^{pruned} < 65$ GeV)
 1182 to predict the total background in the signal region ($65 < m_{jet}^{pruned} < 105$ GeV).

$$\begin{aligned} N_{tot}^{signal}(m_{WW}) &= N_{W+jets}^{signal}(m_{WW}) + N_{OB}^{signal}(m_{WW}) \\ &= N_{data}^{sideband}(m_{WW}) \times (1 - R_0(m_{WW})) \times \alpha^{MC}(m_{WW}) \times F_{W+jets} \\ &\quad + N_{OB}^{signal}(m_{WW}) \end{aligned} \quad (24)$$

1183 This method relies on the assumption that the correlation between the reconstructed m_{WW} and
 1184 m_{jet}^{pruned} are reasonably well modeled by the simulation. The validity of the method assumptions
 1185 can be tested using real data in a control region dominated by W+jets events, with a small
 1186 expected signal contamination.

1187 For this closure test, we define new sideband and signal regions:

- 1188 • low-mass sideband region ($40 < m_{jet}^{pruned} < 65$ GeV);
- 1189 • high-mass sideband region ($115 < m_{jet}^{pruned} < 130$ GeV);
- 1190 • “signal region” ($100 < m_{jet}^{pruned} < 115$ GeV).

1191 What we call “signal region” above is still dominated by W+jets background events, with a
 1192 very small contamination from a possible new physics signal. The goal of the closure test is to
 1193 predict the number of data events in the signal region using the data in the low-mass sideband
 1194 region, following the same procedure employed in the main analysis (Method 1, Section 8.1).

- 1195 • The W+jets normalization (F_{W+jets}) is obtained from fits to the m_{jet}^{pruned} distribution
 1196 as described in Section 8.1.3. The fit is performed using both low- and high- mass
 1197 sideband regions defined above, and the number of W+jets in the “signal” region is
 1198 obtained. Figure 54 show the m_{jet}^{pruned} distributions for the 4 categories with the fit
 1199 results.
- 1200 • The $R_0(m_{WW})$ functions, to subtract the non-W+jets events in the sideband region to
 1201 the data, are shown in Figure 55 for the 4 categories.
- 1202 • The $\alpha^{MC}(m_{WW})$ functions are shown in Figure 56 for the 4 categories.
- 1203 • The total background prediction is shown for the 4 analysis categories in Figures 57
 1204 (linear scale) and 58 (log scale), together with the actual data in the signal region. A
 1205 good agreement between the data and the background prediction is observed, both
 1206 in the shape and the normalization of the distributions.

1207 This cross check confirms that the method 1 to extract the W+jets background is reliable and
 1208 can be used to search for new $X \rightarrow WW$ resonances.

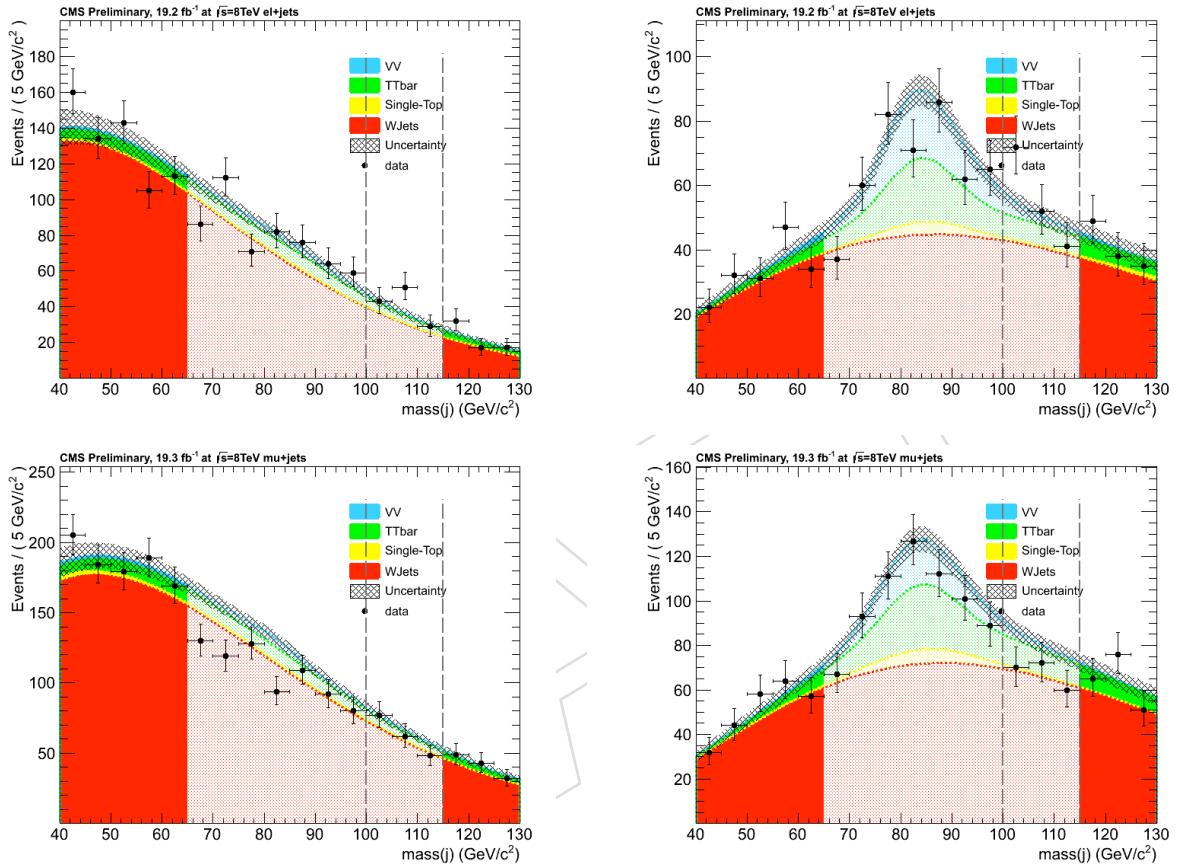


Figure 54: Closure test - Pruned jet mass distributions for the electron LP (top-left), electron HP (top-right), muon LP (bottom-left), and muon HP (bottom-right) categories for events with $40 < m_{jet}^{pruned} < 130$ GeV passing the full selection plus the $m_{WW} > 800$ GeV requirement. The different background components, results of the fit to the low-mass and high-mass sideband regions in data, are shown. In the closure test, the “signal region” is identified by the events in the range $100 < m_{jet}^{pruned} < 115$ GeV .

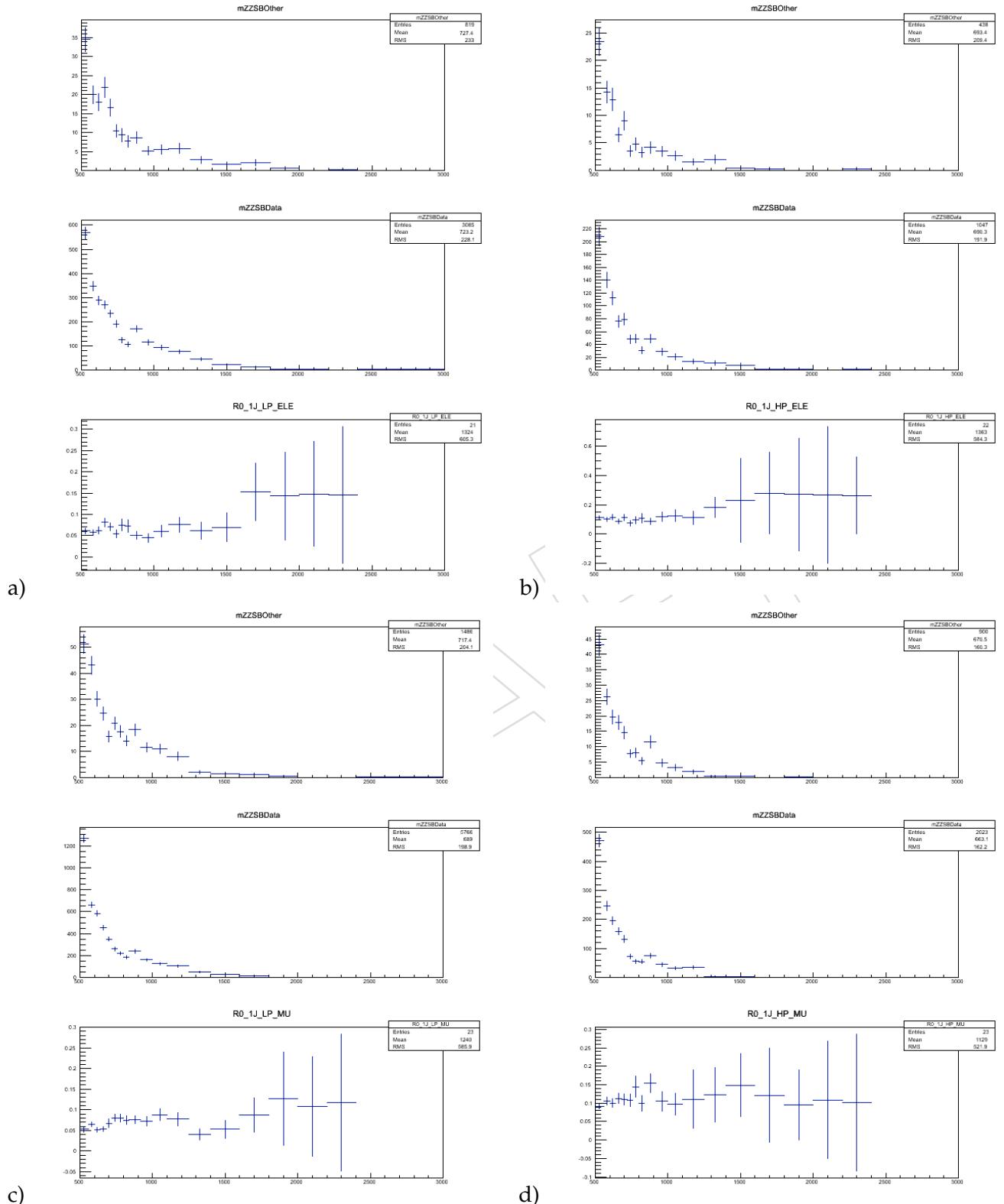


Figure 55: Closure test - $R_0(m_{WW})$ functions vs m_{WW} for the electron LP (a), electron HP (b), muon LP (c), and muon HP (d) categories. Each box shows three plots (from the top to the bottom): $N_{OB}^{\text{sideband}}(m_{WW})$, $N_{data}^{\text{sideband}}(m_{WW})$, and their ratio $R_0(m_{WW})$.

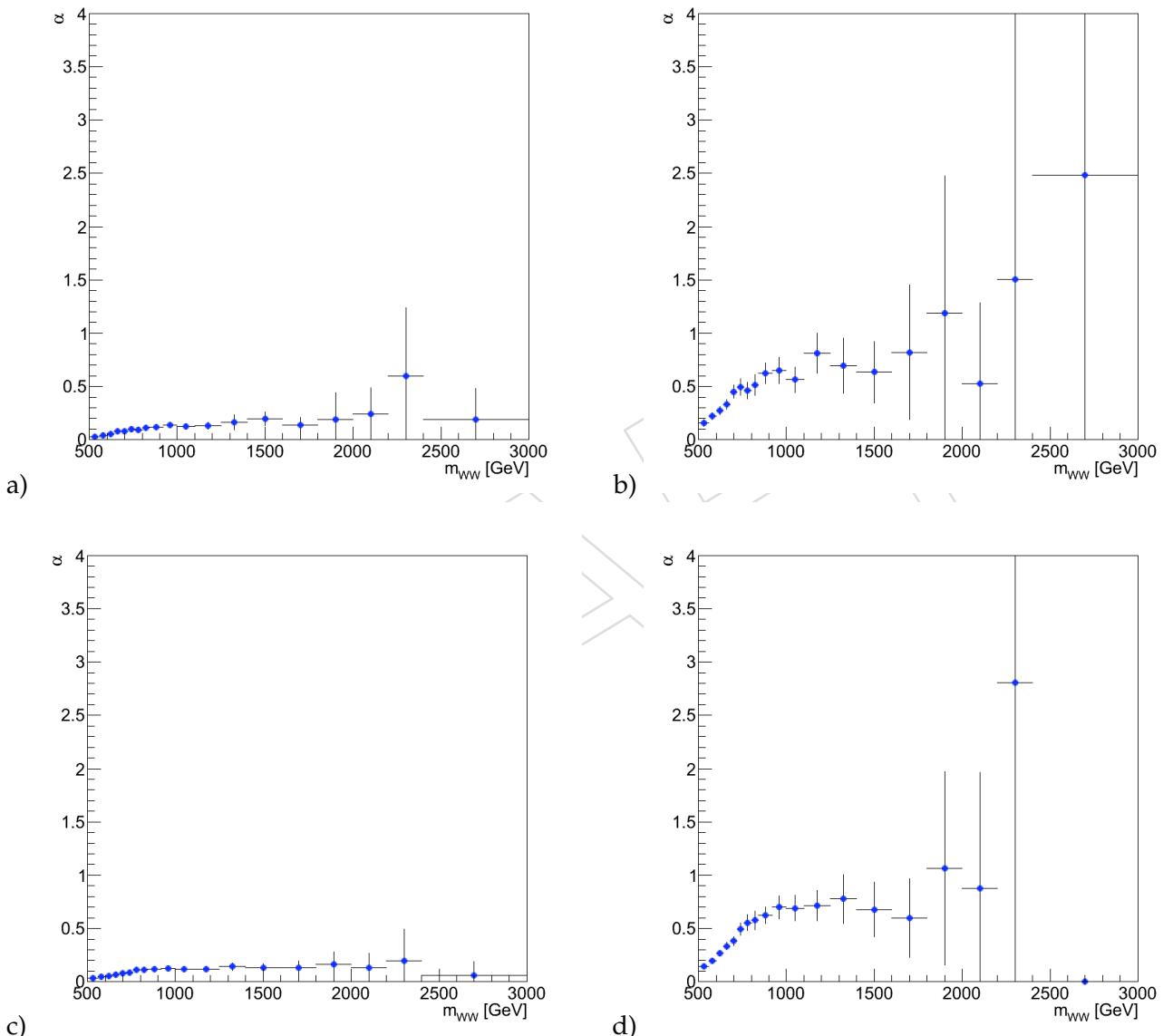


Figure 56: Closure test - $\alpha^{MC}(m_{WW})$ functions vs m_{WW} for the electron LP (a), electron HP (b), muon LP (c), and muon HP (d) categories.

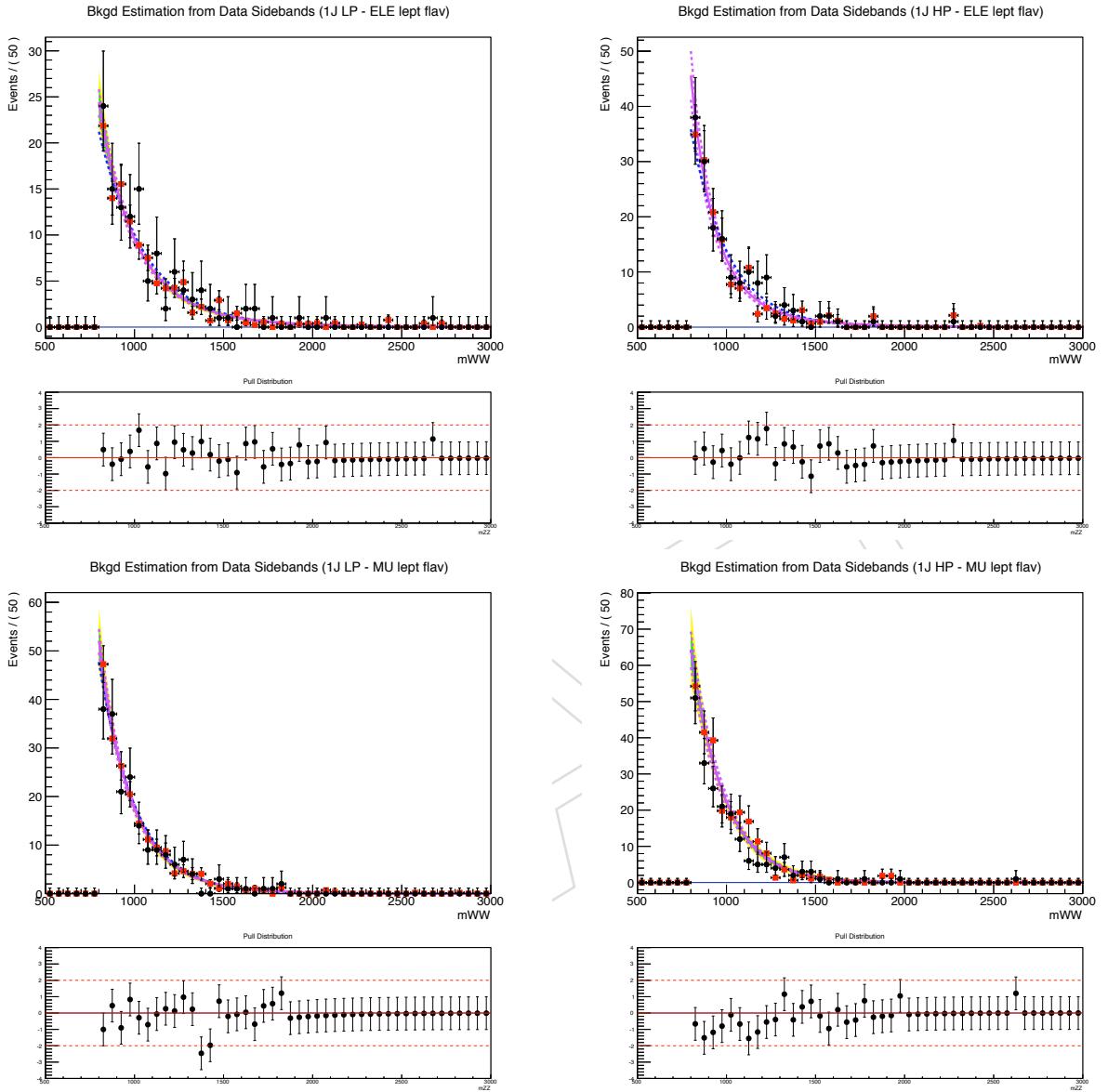


Figure 57: Closure test - m_{WW} distributions for data (black points) in the signal region after full selection for the four event categories used in this analysis: electron LP (top-left), electron HP (top-right), muon LP (bottom-left), muon HP (bottom-right). The red point represents the background prediction from the data in the sideband region corrected by the scale factors and transport functions described in the text. The fit to the corrected data in the sideband (using a leveled exponential function with two shape parameters) is also shown together with error bands from fit parameters. The blue line represents a simple exponential fit (with only one shape parameter) not used in the analysis.

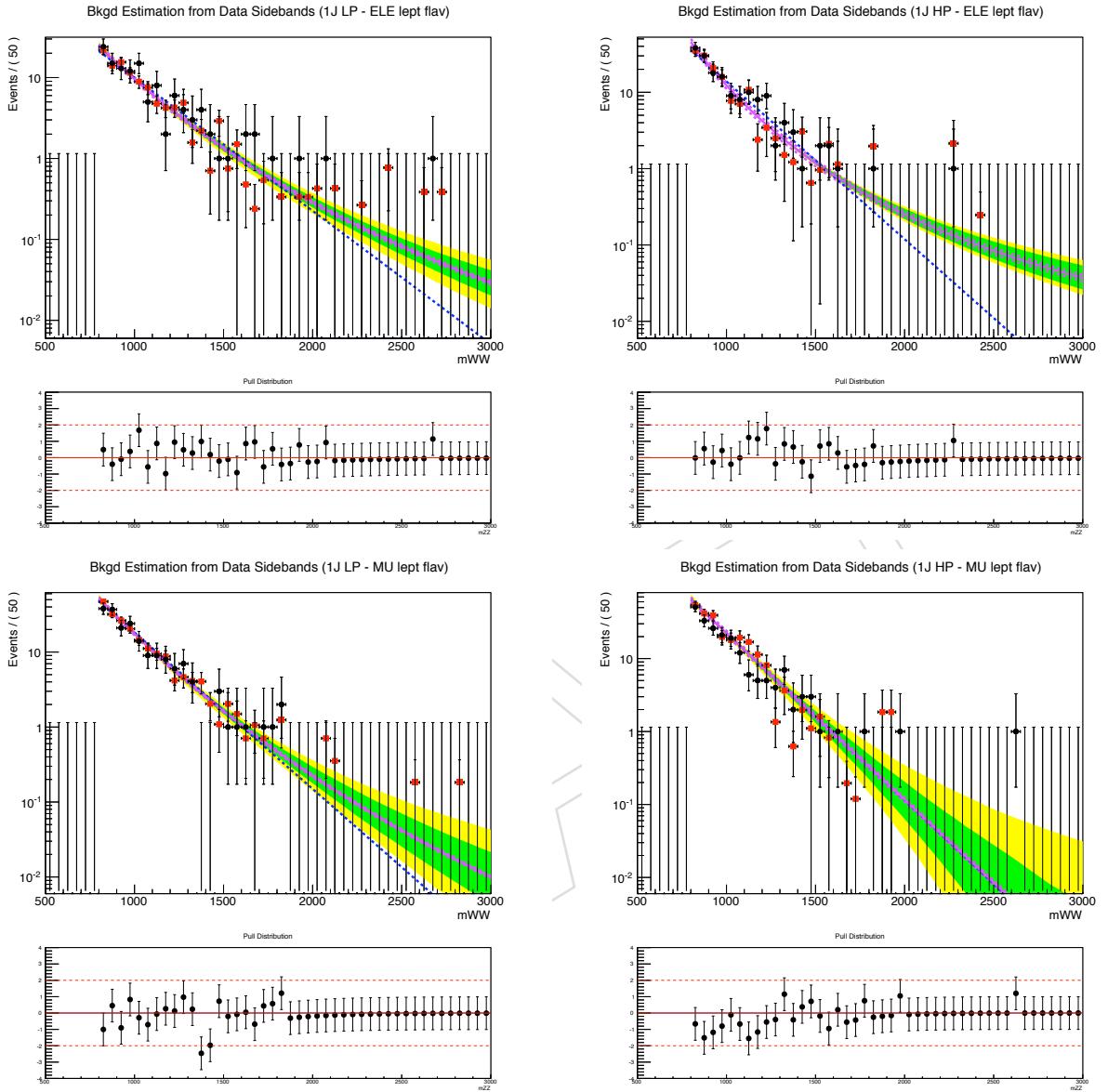


Figure 58: Closure test - m_{WW} distributions for data (black points) in the signal region after full selection for the four event categories used in this analysis: electron LP (top-left), electron HP (top-right), muon LP (bottom-left), muon HP (bottom-right). The red point represents the background prediction from the data in the sideband region corrected by the scale factors and transport functions described in the text. The fit to the corrected data in the sideband (using a leveled exponential function with two shape parameters) is also shown together with error bands from fit parameters. The blue line represents a simple exponential fit (with only one shape parameter) not used in the analysis.

1209 References

- 1210 [1] CMS Collaboration, ““Search for high mass exotic resonances decaying to WW in the
1211 semi-leptonic channel””, CMS Analysis Note CMS AN-13/139, (2013).
- 1212 [2] ALEPH, CDF, D0, DELPHI, L3, OPAL, SLD Collaborations, the LEP Electroweak
1213 Working Group, the Tevatron Electroweak Working Group, and the SLD Electroweak
1214 and Heavy Flavour Groups, “Precision electroweak measurements and constraints on the
1215 Standard Model”, (2010). arXiv:1012.2367.
- 1216 [3] ATLAS Collaboration Collaboration, “Observation of a new particle in the search for the
1217 Standard Model Higgs boson with the ATLAS detector at the LHC”, *Physics Letters B* **716**
1218 (2012), no. 1, 1, doi:10.1016/j.physletb.2012.08.020.
- 1219 [4] CMS Collaboration Collaboration, “Observation of a new boson at a mass of 125 GeV
1220 with the {CMS} experiment at the {LHC}”, *Physics Letters B* **716** (2012), no. 1, 30,
1221 doi:10.1016/j.physletb.2012.08.021.
- 1222 [5] G. D. N. Arkani-Hamed, S. Dimopoulos, “The hierarchy problem and new dimensions at
1223 a millimeter”, *Phys. Lett.* **B429** (1998) 263–272,
1224 doi:doi:10.1016/S0370-2693(98)00466-3, arXiv:9803315.
- 1225 [6] S. D. e. I. Antoniadis, N. Arkani-Hamed, “New dimensions at a millimeter to a Fermi and
1226 superstrings at a TeV”, *Phys. Lett.* **B436** (1998) 257–263,
1227 doi:doi:10.1016/S0370-2693(98)00860-0, arXiv:9804398.
- 1228 [7] L. Randall and R. Sundrum, “A large mass hierarchy from a small extra dimension”,
1229 *Phys. Rev. Lett.* **83** (1999) 3370, doi:10.1103/PhysRevLett.83.3370,
1230 arXiv:hep-ph/9905221.
- 1231 [8] L. Randall and R. Sundrum, “An alternative to compactification”, *Phys. Rev. Lett.* **83**
1232 (1999) 4690, doi:10.1103/PhysRevLett.83.4690, arXiv:hep-th/9906064.
- 1233 [9] CDF Collaboration, “Search for Randall-Sundrum gravitons in the diphoton channel at
1234 CDF”, *Phys. Rev. D* **83** (2011) 011102, doi:10.1103/PhysRevD.83.011102,
1235 arXiv:1012.2795.
- 1236 [10] CDF Collaboration, “Search for New Dielectron Resonances and Randall-Sundrum
1237 Gravitons at the Collider Detector at Fermilab”, *Phys. Rev. Lett.* **107** (2011) 051801,
1238 doi:10.1103/PhysRevLett.107.051801, arXiv:1103.4650v2.
- 1239 [11] D0 Collaboration, “Search for Randall-Sundrum Gravitons in the Dielectron and
1240 Diphoton Final States with 5.4 fb^{-1} of Data from pp Collisions at $\sqrt{s} = 1.96 \text{ TeV}$ ”, *Phys.*
1241 *Rev. Lett.* **104** (2010) 241802, doi:10.1103/PhysRevLett.104.241802,
1242 arXiv:1004.1826.
- 1243 [12] CMS Collaboration, “Search for Signatures of Extra Dimensions in the Diphoton Mass
1244 Spectrum at the Large Hadron Collider”, *Phys. Rev. Lett.* **108** (2012) 111801,
1245 doi:10.1103/PhysRevLett.108.111801, arXiv:1112.0688.
- 1246 [13] CMS Collaboration, “Search for narrow resonances in dilepton mass spectra in pp
1247 collisions at $\sqrt{s} = 7 \text{ TeV}$ ”, *Phys. Lett. B* **714** (2012) 158,
1248 doi:10.1016/j.physletb.2012.06.051, arXiv:1206.1849.

- [14] CMS Collaboration Collaboration, "Search for a narrow spin-2 resonance decaying to a pair of Z vector bosons in the semileptonic final state", *Phys.Lett.* **B718** (2013) 1208–1228, doi:10.1016/j.physletb.2012.11.063, arXiv:1209.3807.
- [15] ATLAS Collaboration, "Search for strong gravity signatures in same-sign dimuon final states using the ATLAS detector at the LHC", *Phys. Lett. B* **709** (2012) 322, doi:10.1016/j.physletb.2012.02.049, arXiv:1111.0080.
- [16] ATLAS Collaboration, "Search for Dilepton Resonances in pp Collisions at $\sqrt{s} = 7 \text{ TeV}$ with the ATLAS Detector", *Phys. Rev. Lett.* **107** (2011) 272002, doi:10.1103/PhysRevLett.107.272002, arXiv:1108.1582.
- [17] CMS Collaboration, "Search for a narrow, spin-2 resonance decaying to a pair of Z bosons in the final state", *Physics Letters B* **718** (2013), no. 4–5, 1208 – 1228, doi:10.1016/j.physletb.2012.11.063, arXiv:1209.3807. CMS-PAS-EXO-11-102.
- [18] CMS Collaboration, "Search for exotic resonances decaying into WZ/ZZ in pp collisions at $\sqrt{s} = 7 \text{ TeV}$ ", *JHEP* (2012) doi:10.1007/JHEP02(2013)036, arXiv:1211.5779. CMS-PAS-EXO-12-014 ; CMS-PAS-EXO-11-081; CMS-PAS-EXO-11-061.
- [19] S. Chatrchyan et al., "Search for heavy resonances in the W/Z-tagged dijet mass spectrum in pp collisions at 7 TeV", (2013). arXiv:1212.1910. CMS-PAS-EXO-11-095 ; submitted to Phys. Lett. B.
- [20] K. Agashe, H. Davoudiasl, G. Perez, and A. Soni, "Warped Gravitons at the LHC and Beyond", *Phys. Rev. D* **76** (2007) 036006, doi:10.1103/PhysRevD.76.036006, arXiv:hep-ph/0701186.
- [21] A. L. Fitzpatrick, J. Kaplan, L. Randall, and L.-T. Wang, "Searching for the Kaluza-Klein Graviton in Bulk RS Models", *JHEP* **09** (2007) 012, doi:10.1088/1126-6708/2007/09/013, arXiv:hep-ph/0701150.
- [22] O. Antipin, D. Atwood, and A. Soni, "Search for RS gravitons via $W_L W_L$ decays", *Phys. Lett. B* **666** (2008) 155, doi:10.1016/j.physletb.2008.07.009, arXiv:0711.3175.
- [23] W.-Y. K. V. Barger, Y. Gao, "T-Anomaly Induced LHC Signals", *Phys. Lett. B* **655** (2007) 228, doi:doi:10.1016/j.physletb.2007.08.090, arXiv:0707.3648.
- [24] J. E. Juknevich, D. Melnikov, and M. J. Strassler, "A Pure-Glue Hidden Valley. States and Decays", *J. High Energy Phys.* **0907** (2009) 055, doi:doi:10.1088/1126-6708/2009/07/055, arXiv:0903.0883.
- [25] Y. G. et al., "Spin-Determination of Single-Produced Resonances at Hadron Colliders", *Phys. Rev. D* **81** (2010) doi:10.1103/PhysRevD.81.075022, arXiv:1001.3396.
- [26] S. B. et al., "On the Spin and Parity of a Single-Produced Resonance at the LHC", *Phys. Rev. D* **86** (2012) 095031, doi:10.1103/PhysRevD.86.095031, arXiv:1208.4018.
- [27] T. Sjöstrand, S. Mrennan and P. Z. Skands, "PYTHIA 6.4 Physics and Manual", *JHEP* **05** (2006) 026, doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.

- [28] A. A.Belyaev, N.Christensen, "CalcHEP 3.4 for collider physics within and beyond the Standard Model", *Computer Physics Communications* **184** (2013) 1729–1769, doi:10.1016/j.cpc.2013.01.014, arXiv:1207.6082.
- [29] CMS Collaboration, "Search for new physics in the single lepton + MET final states with full 2012 dataset at $\sqrt{s}=8$ TeV", CMS Analysis Note CMS-AN-2012/423, (2013).
- [30] CMS Collaboration, "Commissioning of the Particle-Flow Reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV", CMS Physics Analysis Summary CMS-PAS-PFT-10-002, (2010).
- [31] CMS Collaboration, "Jet Energy Corrections determination at 7 TeV", CMS Physics Analysis Summary CMS-PAS-JME-10-010, (2010).
- [32] CMS Collaboration, "Search for anomalous $t\bar{t}$ production in the highly-boosted all-hadronic final state", *JHEP* **09** (2012) 029, doi:10.1007/JHEP09(2012)029, arXiv:1204.2488. CMS-PAS-EXO-11-006.
- [33] CMS Collaboration, "Searches for new physics in the $WW \rightarrow l\nu j$ final state with hadronic W bosons merged into a single jet", CMS Analysis Note CMS-AN-2012/381, (2012).
- [34] CMS Collaboration, "Search for BSM $t\bar{t}$ production in the boosted all-hadronic final state using pp collisions at $\sqrt{s} = 8$ TeV", CMS Analysis Note CMS-AN-2012/179, (2013).
- [35] CMS Collaboration, "b-jet Identification in the CMS Experiment", CMS Physics Analysis Summary CMS-PAS-BTV-11-004, (2012).
- [36] M. Gouzevitch et al., "Scale-invariant resonance tagging in multijet events and new physics in Higgs pair production", arXiv:1303.6636.
- [37] ATLAS Collaboration, "Performance of large-R jets and jet substructure reconstruction with the ATLAS detector", Technical Report ATLAS-CONF-2012-065, CERN, (Jul, 2012).
- [38] S. Chatrchyan et al., "Studies of jet mass in dijet and $W/Z+jet$ events", (2013). arXiv:1303.4811. CMS-PAS-SMP-12-019 ; submitted to JHEP.
- [39] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, "Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches", *Phys.Rev.* **D81** (2010) 094023, doi:10.1103/PhysRevD.81.094023, arXiv:0912.0033.
- [40] J. Thaler and K. Van Tilburg, "Identifying Boosted Objects with N-subjettiness", *JHEP* **1103** (2011) 015, doi:10.1007/JHEP03(2011)015, arXiv:1011.2268.
- [41] J. Thaler and K. Van Tilburg, "Maximizing Boosted Top Identification by Minimizing N-subjettiness", *JHEP* **1202** (2012) 093, doi:10.1007/JHEP02(2012)093, arXiv:1108.2701.
- [42] I. W. Stewart, F. J. Tackmann, and W. J. Waalewijn, "N-Jettiness: An Inclusive Event Shape to Veto Jets", *Phys.Rev.Lett.* **105** (2010) 092002, doi:10.1103/PhysRevLett.105.092002, arXiv:1004.2489.
- [43] CMS Collaboration, "Search for a Standard Model-like Higgs boson decaying into WW to $l\nu q\bar{q}$ in pp collisions at $\sqrt{s} = 8$ TeV", CMS Physics Analysis Summary CMS-PAS-HIG-13-008, (2013).

- 1329 [44] G. Punzi, "Sensitivity of searches for new signals and its optimization", in *proceedings to*
1330 *PhyStat 2003 Conference, Stanford, Ca, USA, September 2003, PSN MODT002*.
1331 [arXiv:physics/0308063v2](https://arxiv.org/abs/physics/0308063v2).
- 1332 [45] CMS Collaboration, ""Search for a BSM resonance decaying to Z vector bosons in the
1333 semileptonic final state"", CMS Analysis Note CMS-AN-2013/040, (2013).
- 1334 [46] CMS Collaboration, ""Search for Narrow Resonances using the Dijet Mass Spectrum
1335 with 19.6 fb-1 of pp Collisions at $\sqrt{s} = 8$ TeV"", CMS Physics Analysis Summary
1336 CMS-PAS-EXO-12-059, (2013).
- 1337 [47] CMS Collaboration, ""CMS Luminosity Based on Pixel Cluster Counting - Summer 2012
1338 Update"", CMS Physics Analysis Summary CMS-PAS-LUM-12-001, (2012).
- 1339 [48] CMS Collaboration, ""ECAL Detector Performance, 2011 Data"", CMS Performance Note
1340 CMS-DP-2012/007, (2012).
- 1341 [49] CMS Collaboration, ""Search for High Mass Resonances Decaying to Electron Pairs at 8
1342 TeV with the Full 2012 dataset"", CMS Analysis Note CMS-AN-2012/415, (2013).
- 1343 [50] A. L. Read, "Presentation of search results: The CL(s) technique", *J. Phys.* **G28** (2002)
1344 2693–2704, doi:10.1088/0954-3899/28/10/313.
- 1345 [51] A. L. Read, "Modified frequentist analysis of search results (the CLs method)", technical
1346 report, (2000).
- 1347 [52] G. Cowan, K. Cranmer, E. Gross, O. Vitells, "Asymptotic formulae for likelihood based
1348 tests of new physics", *Eur. Phys. J.* **C71** (2011) 1554, [arXiv:physics/1007.1727v2](https://arxiv.org/abs/physics/1007.1727v2).
- 1349 [53] CMS Collaboration, ""Combined results of searches for the standard model Higgs boson
1350 in pp collisions at $\sqrt{s} = 7$ TeV"", *Phys. Lett. B* (2012), no. 710, [arXiv:1202.1488](https://arxiv.org/abs/1202.1488).
- 1351 [54] Tommaso Dorigo, ""Suggestions on F-test from CMS Statistics Committee"",
1352 <https://hypernews.cern.ch/hypernews/cms/get/exo-11-095/29/1/1/1.html>.