Search for Heavy Resonances Decaying into Two Higgs Bosons or into a Higgs Boson and a W or Z Boson in the $q\bar{q}$ $(b\bar{b})$ $\tau^+\tau^-$ Final State with the CMS Detector

---

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Camilla Galloni

aus

Italien

CERN-THESIS-2019-439
13/12/2018

Promotionskommission

Prof. Dr. Benjamin Kilminster (Vorsitz und Leitung der Dissertation)

Prof. Dr. Florencia Canelli

Prof. Dr. Laura Baudis

Prof. Dr. Ueli Straumann

Zürich, 2019

Abstract

This thesis presents a search for potential signals of new massive particles decaying to pairs of W, Z, and Higgs bosons that are predicted by beyond the standard model theories. The data analyzed have been collected with the CMS detector at the Large Hadron Collider (LHC) during pp collisions at center of mass energies of $\sqrt{s} = 8\,\mathrm{TeV}$ in 2012 (Run 1) and $\sqrt{s} = 13\,\mathrm{TeV}$ in 2016 (Run 2), corresponding to an integrated luminosity of 19.7 fb^{-1} and 35.9 fb^{-1} , respectively. Such new particles are the prominent feature of many theoretical models that aim to clarify some of the questions unanswered by the standard model, such as the apparently large difference between the electroweak and the gravitational scales. The final states analyzed in this work are compatible with the presence of a new massive resonance that decays to a Higgs boson and a W, Z, or Higgs boson, with one of the Higgs bosons further decaying to τ leptons. Since τ leptons are unstable, they can decay further into either lighter leptons (ℓ), electrons and muons, and neutrinos or into neutral and charged hadrons (τ_{h}) and neutrinos. Therefore, they can generate a plethora of final states. The other boson can be a W, Z, or H boson and is required to decay hadronically to a pair of quarks.

The first study is focused on the search for a resonance decaying to HH and subsequently to $\tau^{+}\tau^{-}\mathrm{b\overline{b}}$ in the final state where one of the tau leptons decays to hadrons and a neutrino; and the other τ to a lighter lepton, either an electron or a muon, and two neutrinos. This analysis is based on data recorded in Run 1 (2012) of the LHC. The second study searched for WH, ZH, or HH resonances decaying to quarks and τ leptons in data recorded during Run 2 (2016) of the LHC. In the Run 2 analysis, in order to extend the sensitivity, additional final states are considered in which both τ leptons from the H boson decay hadronically. In addition, the search is extended to consider hadronic decays of the W and Z bosons to $\mathrm{q\overline{q}}$.

These final states are particularly challenging because, for large resonance masses, the bosons are highly energetic and the final products from their decay are separated by a small angle in space. This collimation implies that the quarks from the hadronically boson decay are reconstructed in one large-cone jet. Novel jet-substructure techniques and dedicated algorithms for the mass reconstruction and flavor identification of the jets are applied to distinguish W, Z, and H bosons. The τ pair produced from the H boson decay has a high Lorentz boost and the final decay products are also collimated. Special techniques were developed as part of this doctoral work to correctly reconstruct and identify the τ lepton pairs in this particularly boosted topology.

The search is performed by scanning the distribution of the reconstructed mass of the resonance, looking for a local excess in data in comparison with the background prediction. Theoretical scenarios of new particles with a spin of 0, 1, and 2 are investigated, and upper limits are set on their cross-section as a function of mass. These are the first searches for heavy resonances decaying to bosons pair with τ leptons in the final state of Run 2 of the LHC.

Diese Doktorarbeit stellt eine Suche nach potenziellen Signalen neuartiger massiver Teilchen dar, welche zu Paaren von W-, Z- und Higgs-Bosonen zerfallen, die von Theorien Jenseits des Standardmodels der Teilchenphysik vorhergesagt werden. Die analysierten Daten wurden mit dem CMS-Detektor am Large Hadron Collider (LHC) bei *pp*-Kollisionen mit Massenschwerpunktsenergien von $\sqrt{s} = 8\,\text{TeV}$ im Jahr 2012 (Run 1) und $\sqrt{s} = 13\,\text{TeV}$ im Jahr 2016 (Run 2) gesammelt, was einer integrierten Leuchtstärke von 19.7 fb^{-1} bzw. 35.9 fb^{-1} entspricht. Solche neuartige Teilchen sind das Hauptmerkmal vieler theoretischer Modelle, die darauf abzielen, einige der vom Standardmodell unbeantworteten Fragen zu klären, wie beispielsweise den grossen Unterschied zwischen der elektroschwachen und der gravitativen Skala.

Die in dieser Arbeit analysierten Endzustände sind kompatibel mit dem Vorhandensein neuer massiver Resonanzen, die zu einem Higgs-Boson zerfallen, welches wiederum zu τ-Leptonen und zu einem zweiten W-, Z- oder Higgs-Boson zerfällt. Da τ-Leptonen instabil sind, können sie eine Vielzahl von Endzuständen erzeugen. Sie zerfallen entweder in leichtere Leptonen (ℓ), Elektronen und Myonen und Neutrinos oder in neutrale und geladene Hadronen (τ_h) und Neutrinos. Das andere Boson kann ein W-, Z- oder H-Boson sein und muss hadronisch zu einem Paar Quarks zerfallen. Die erste Studie basiert auf den im (Run 1) gesammelten Daten und konzentriert sich auf die Suche nach einer Resonanz, welche zu HH und anschliessend zu $\tau^+\tau^-\text{b}\overline{\text{b}}$ im Endzustand zerfällt, wobei eines der Tau-Leptonen zu Hadronen und einem Neutrino zerfällt. Das andere Tau-Lepton zerfällt zu einem leichteren Lepton (ein Elektron oder ein Myon), und zwei Neutrinos. Die zweite Studie sucht nach WH, ZH, oder HH Resonanzen, die zu Quarks und τ Leptonen zerfallen, wobei die LHC Daten des (Run 2) zu Grunde liegen. In der Run 2-Analyse werden zur Erweiterung der Empfindlichkeit zusätzliche Endzustände betrachtet, in denen beide τ Leptonen aus dem H-Boson hadronisch zerfallen. Die Suche wird darüber hinaus nach Teilchen erweitert, welche auch zu W- und Z-Bosonen zerfallen, indem für das hadronisch zerfallende Boson die inklusiven $\text{q}\overline{\text{q}}$-Zerfälle berücksichtigt werden. Diese Endzustände sind besonders herausfordernd, da bei hohen Resonanzmassen die Bosonen hochenergetisch sind und deren Zerfallsprodukte durch einen kleinen Raumwinkel getrennt sind.

Diese Kollimation impliziert, dass die Quarks aus dem hadronischen Bosonzerfall in einem einzelnen Grosskegelstrahl (Jet) rekonstruiert werden. Neuartige Jet-Substruktur-Techniken und spezielle Algorithmen für die Massenrekonstruktion und Flavour-Erkennung der Jets werden eingesetzt, um W, Z und Higgs Bosonen zu unterscheiden. Das aus dem H-Boson-Zerfall entstehende τ-Paar hat einen hohen Lorentz-Boost und die Zerfallsprodukte sind analog dazu ebenfalls stark kollimiert. Im Rahmen dieser Doktorar-

beit wurden spezielle Techniken entwickelt, um die τ-Leptonpaare in dieser besonders verzerrten Topologie korrekt zu rekonstruieren und zu identifizieren. Die Suche erfolgt durch das Rastern der rekonstruierten Massenverteilung der Resonanz, wobei nach einem lokalen Datenüberschuss im Vergleich zur Hintergrundprognose gesucht wird. Theoretische Szenarien neuartiger Teilchen mit einem Spin von 0, 1 und 2 werden untersucht, wobei Obergrenzen für ihren Wirkungsquerschnitt in Abhängigkeit der Masse festgelegt werden. Diese Arbeit repräsentiert die ersten LHC (Run 2) Nachforschungen, die nach schweren Resonanzen suchen, welche zu Bosonen-Paaren mit τ-Leptonen im Endzustand zerfallen.

Acknowledgements

It has been a unique experience and a pleasure to work at the University of Zurich and being part of the CMS collaboration. Many people helped me during this PhD journey and I would like to thank them.

I am sincerely grateful to Prof. Ben Kilminster for giving me the opportunity to work in his group on very interesting physics research topics, for the constructive discussions, his valuable guidance over these years, and for always supporting me in new projects. Thanks also to my PhD committee, Prof. Canelli, Prof. Baudis and Prof. Straumann for the useful suggestions during the review of my work.

A big thank goes to my colleagues and friends at the University of Zurich for the wonderful time spent together and all their help during these years.

In particular, I would like to thank Andreas Hinzmann and Clemens Lange for their valuable advise at the beginning of my work and throughout the end. I'd like to thank Danek Kotlinski, Stefanos Leontsinis and Lea Caminada for their endless patience with me and all my questions on the pixel detector and their guidance in understanding the new machine in 2017. A big thank to Claudia Seitz for the physics discussions and keeping me well fed with all the dinners organized at her place.

Thanks to Alberto Zucchetta, that supported me an infinity of times, first with fit convergence and later also in becoming more politically correct and serene. Thanks to Yuta Takahashi and Izaak Neutelings for the whole tau experience that made 2018 quite challenging, but in the end a success. A great thank also to Jan Staggeman, Isobel Ojalvo, Riccardo Manzoni, Aniello Spiezia, and Christian Veelken, for all their help during these years with tau leptons and their review of the practical details of the analysis.

A big thank to Jennifer and Thea who shared with me the PhD and programming experience since the very beginning, for the useful confrontation and discussion.

A big thank to Deborah, Giorgia, and Riccardo, in whom I found friends always ready to push me, argue, make me laugh, and think when needed. Thanks for being there when I needed it and I can only hope to be able to do my part in return.

Thanks also to the friends at CERN, for sharing lunch and coffee together, and sometimes trying to avoid too many physics discussions!

I am very thankful to Sandro for sharing the good moments and sustaining me in the difficult ones.

Last but not least, the biggest thanks go to my parents for all their support. They always encouraged me to do my best and to pursue my dreams. Thank you!

Contents

Chapter 1

Introduction

Humans always tried to give explanations to the natural processes surrounding them and wondered about the structure of the universe. In recent times, particle physics has linked together the most fundamental elements of nature, space, time and matter in its attempt to explain the laws that rule the universe. Two main theories are the basis of our present understanding: relativity and quantum physics. The standard model (SM) is the mathematical description of the processes of particle physics and was formalized in the 1970's, based on local gauge invariance of the Lagrangian under the group symmetries of the theory. It explains the electromagnetic and weak interactions in a common framework by introducing spin-1 bosons, such as the photons, the $W^{\pm}$ and Z bosons, which are the mediators of the interactions of the matter fields. A similar description can be extended to strong interactions, with gluons as mediators. An important addition to the SM is the Brout-Englert-Higgs-Hagen-Guralnik-Kibble mechanism that predicts the existence of a new field, the Higgs boson field, that explains how the $W^{\pm}$ and Z mediators can acquire mass, thus justifying the different energy scales of electromagnetism and weak interactions, as well as how fermions acquire mass through Yukawa interactions. In the last century many experiments confirmed SM predictions: the $W^{\pm}$ and Z bosons were discovered at the super proton synchrotron (SPS) at CERN in the 1980's, the top quark was discovered in the 1990's at the Tevatron at FNAL, and finally the Higgs boson was discovered in 2012 at the Large Hadron Collider (LHC) at CERN, a half century after its prediction. However, the SM has some important limitations. First of all, a quantum formulation of gravity is not incorporated in the SM. Some phenomena remain unexplained: the SM is lacking candidate fields for dark energy and dark matter, which are necessary to explain observations such as the expansion rate of the universe and the rotational velocity of galaxies. Moreover, the fact that the gravitational and electroweak interactions have such different scales is unexplained. All of this evidence indicate that the SM can be believed only as an approximation of a more complete theory. Many theories have been hypothesized in recent years to extend the SM either by increasing the number of symmetries or the number of spatial dimensions in the theory. Many beyond the SM (BSM) scenarios predict the appearance of new heavy particles, expected to have masses around the TeV scale. The unprecedented center-of-mass energy of 8 TeV in 2012 and 13 TeV in 2015 allows the study of this previously inaccessible and unknown phase space. An example of this quest is provided by this doctoral work. A search is presented for new heavy resonances decaying to WH, ZH, and HH performed with data collected by CMS detector, which features a multipurpose detector, suitable for studying highly energetic new phenomena.

The first result reported is a search for a HH resonance decaying to $\tau^+\tau^-\mathrm{b\overline{b}}$ based on data recorded in proton-proton (*pp*) collisions at a center-of-mass energy of $\sqrt{s} = 8\,\mathrm{TeV}$ during 2012 (Run 1 of the LHC). This is one of the first searches for new physics where the recently discovered Higgs boson is required in the final state as a tool to probe for new physics. The second result reported is a search for resonant production of WH, ZH or HH decaying to $\tau^+\tau^-\mathrm{q\overline{q}}$ or $\tau^+\tau^-\mathrm{b\overline{b}}$ in data corresponding to an integrated luminosity of 35.9 fb^{-1} of *pp* collisions collected in 2016 at $\sqrt{s} = 13\,\mathrm{TeV}$ (Run 2 of the LHC). Due to the large resonance masses considered the intermediate bosons are highly energetic, and the products from their decay can have a very small angular separation and overlap in the detector, making it difficult to resolve the components.

For the hadronic boson decays to a pair of quarks, novel jet reconstruction techniques, called "V tagging" (for a vector boson V = W or Z) and "H tagging", were developed during Run 1, which exploit the substructure of large-cone jets and help to resolve the collimated decay products. These dedicated algorithms allow a pair of quarks originating from a massive SM boson to be distinguished from the the background processes, initiated by the strong interaction. For bosonic decays to a tau lepton pair, special techniques were studied and developed during this doctoral work, in order to adapt the CMS identification algorithm for hadronic τ lepton decays to target this particular final state in which the two τ lepton decays happen within a small angular separation and exhibit a variety of different decays.

During Run 1, several searches for diboson resonances were carried out by both the ATLAS and the CMS collaborations and some small deviations from the SM expectation were observed at high mass, around 2 TeV [1], that could have indicated new physics. Therefore, when the LHC resumed physics collisions at higher energy in 2015, a major effort was put forth to further explode this high mass region of the excess with higher energy data. The integrated luminosity of data recorded in 2015 corresponds to 2.3 fb^{-1}, considerably less than the 2016 integrated luminosity of 35.9 fb^{-1}. Since one of the limiting factors of this analysis is the statistics of data in kinematic regions enriched in processes that are used for modeling the backgrounds, the larger 2016 dataset was used. As a result of the analysis, in the particular final states analyzed in this work, no deviation from the SM expectation was found, so that exclusion limits on the product of cross section of the new resonance and the decay to a dibosonic final state were set.

The thesis is organized in the follow way. In Chapter 2, an overview of the SM is presented. Two BSM theories are considered as benchmarks for the heavy resonance searches and are presented in Chapter 3: the warped extra dimension model and the Heavy Vector Triplet model. In Chapter 4, the experimental setup is described with an overview of the LHC and the CMS detector. Chapter 5 summarizes the methods used in CMS for event and physical object reconstruction. Jet substructure techniques and the identification methods of the boosted τ leptons pairs are presented as well. Chapter 6 reports the results form the HH analysis performed with the data collected during Run 1 of the LHC and contains the main steps of the analysis, including details on the final event selection, the estimation of the SM background, the main systematic uncertainties, and the interpretation of the results. Analogously, Chapter 7 is devoted to the search for resonant WH, ZH and HH production with data collected during Run 2 of the LHC with the increased center of mass energy of 13 TeV. For the statistical methods used in order to analyze the data, an overview is given in Appendix A. Finally, Chapter 8 provides a brief summary of this work.

[1]here and in this whole thesis work the light velocity c is assumed to be 1 by convention

Chapter 2

Theory

The known fundamental interactions of Nature are four: the electromagnetic, weak, strong and gravitational. For the first three there is a common formulation, the standard model (SM), that describes the known particles and their interactions with very high accuracy. Gravity is not included in the standard model formulation and gravitational effects are negligible at the subatomic scales. The standard model includes a mechanism of spontaneous symmetry braking, called the Higgs mechanism, due to the presence of a massive scalar boson, that explains how the known particles gain their masses. The SM is well-corroborated by experimental observations at collider experiments, and received further confirmations with the recent discovery of the Higgs boson. In 2012, evidence for the Higgs boson was found at the Tevatron and it was finally observed at the LHC. In this chapter, the main features of the standard model, the Higgs mechanism, and some of the observations that can't be explained in this framework are presented.

2.1 The standard model of particle physics

The mathematical formulation of the standard model is based on the symmetry group $\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ that explains the strong, weak and electromagnetic interactions. An $\mathrm{SU}(3)_C$ gauge invariance results in the presence of the mediators of the strong force, the *gluons* (g), which are color charged and self interacting, as described by Quantum Chromo-Dynamics (QCD). The $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ gauge group invariance was formulated by Glashow [1], Weimberg [2] and Salam [3] in the 1960's in order to describe the electroweak interactions: the charge-neutral massless photon is the mediator of the electrodynamic field, and the charged and neutral massive mediators of the electroweak interaction are the $W^{\pm}$ and Z bosons.

Matter is described by fermionic fields of spin 1/2, whose interactions are mediated by spin-1 bosonic fields. Twelve fermionic fields have been observed experimentally, six *lepton* fields and six *quark* fields. They are organized into three families made up of two leptons of electric charge -1 and 0 and two quarks of electric charge $+\frac{2}{3}$ and $-\frac{1}{3}$. Each particle has an antiparticle with identical identical, but opposite quantum numbers. The fermions in the different families have similar properties but different masses, which are generated through their unique coupling to the scalar field.

2.1.1 Leptons

The leptons may undergo only electromagnetic and weak interactions. The charged leptons are denoted as electron (e), muon (μ) and tau (τ). The electron is lightest charged lepton with a mass of 511 keV [4] and, thus, stable. The muon has a mass of 105.7 MeV, a lifetime of 2.2 μs [4], and eventually decays to an electron. Due to the high energy of particles produced at the LHC, the muon can be considered a stable particle in the detector, since its lifetime is sufficiently long. Taus are the only leptons heavy enough, with a mass of 1.78 GeV, to decay to hadrons, and its lifetime of $2.9 \cdot 10^{-13}$ s is short enough that only its decay products are observed in the detector. However, a tau with momentum of a few tens of GeV can travel a few millimeters before decaying, causing the decay products to be displaced from the primary interaction vertex, exhibiting among its decay features a track with large impact parameter or even a secondary vertex. The neutral leptons are the neutrinos, one for each family (ν_{e}, ν_{μ}, ν_{τ}), that are only subject to weak interactions and are not directly detectable in the experiment. Neutrinos are very light, but massive, as is evidenced by their observed oscillation between flavors, which can be explained if the mass eigenstates differ from the electroweak ones, as is done by the Pontecorvo-Maki-Nagakawa-Sakata (PMNS) model [5].

2.1.2 Quarks

Quarks undergo both electroweak and strong interactions. For the latter, they are said to have “color” charge. They are also grouped in three families. Ordinary matter is composed of electrons and the first family where the up (u) and down (d) quarks, with a mass of few MeV, are grouped. The second family consists of the charm (c) and strange (s) quarks, of masses 1.27 GeV and 96 MeV, respectively. Then, the third family includes the top (t) and bottom (b) quarks, which have masses of 173 GeV and 4.2 GeV, respectively. Quarks form hadrons, i.e. bound states, such as mesons (built of a quark and antiquark $q\bar{q}$) and baryons (built of three quarks qqq or three antiquarks $\bar{q}\bar{q}\bar{q}$). In the high-energy regime, quark and gluons interact freely, with asymptotic freedom, allowing a perturbative description of QCD, as in the case of the electroweak interaction. Whereas at small energies, the strength of the interaction increases with the distance, resulting in the quark (or color) confinement, for which a non-perturbative description is needed. These effects become important at energies close to the QCD scale $\Lambda_{\mathrm{QCD}} \sim 200$ MeV, near the light meson mass scale. When a quark or a gluon is produced through hard scattering, a process called "hadronization" happens on a time scale of 10^{-24} s: pairs of quarks and anti-quarks are produced from the interaction with the vacuum and combined with the original quark until colorless hadrons are formed. The hard scattering and hadronization phenomena can be treated separately thanks to the factorization of their effects. The top quark represents an exception in this sense, as its lifetime is so short ($\sim 0.5 \cdot 10^{-24}$ s) that it decays before bound states can be formed. Quark flavor is conserved in strong interactions, but not in the weak, because the quark mass eigenstates are not the same as the eigenstates of the weak interactions, such that the mixing is described by the Cabibbo-Kobayashi-Maskawa (CKM) matrix.

2.2 Standard model

The standard model provides a mathematical description of the interactions that occur in nature. Starting from the simplest case, the Lagrangian density of a free , i.e. non-interacting, Dirac field of spin $\frac{1}{2}$ is:

$$\mathcal{L}_{\mathrm{free}} = \bar{\psi}(x)\left(i\gamma^{\mu}\partial_{\mu} - m\right)\psi(x) \tag{2.1}$$

where $\psi(x)$ is the fermionic field at the space-time coordinate x, γ^{μ} are the Dirac matrices, satisfying $\{\gamma^{\mu}, \gamma^{\nu}\} = 2\eta^{\mu\nu}$, such that $\eta^{\mu\nu}$ is the Minkowski metric, $\bar{\psi}(x) = \psi^{\dagger}(x)\gamma^{0}$, ∂_{μ} is the derivative, and m the mass of the particle. If the particle undergoes interactions there are terms that can be added to the Lagrangian in order to describe the different kinds of processes.

2.2.1 Strong interactions

Inside hadrons, quarks are described as fermions with degrees of freedom that correspond to a spin $\frac{1}{2}$, with three values of color obeying the SU(3)$_C$ group that describes the strong interactions. This group is generated by the eight $\frac{\lambda}{2}$ generators, where the λ's are the Gell-Mann matrices [6]

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \ \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \ \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \ \lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$
$$\lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix}, \ \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \ \lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & -i & 0 \end{pmatrix}, \ \lambda_8 = \begin{pmatrix} \frac{1}{\sqrt{3}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{3}} & 0 \\ 0 & 0 & \frac{-2}{\sqrt{3}} \end{pmatrix} \tag{2.2}$$

that satisfy the commutation rule $\left[\frac{\lambda^a}{2}, \frac{\lambda^b}{2}\right] = if^{abc}\frac{\lambda^c}{2}$, with the structure constants f^{abc}. Under the SU(3)$_\mathrm{C}$ local gauge transformations, the fermion fields and the derivative term transform as

$$\begin{aligned} &\psi(x) \to e^{-ig_s \frac{\lambda^a}{2}\theta^a(x)}\psi(x) \\ &\partial_\mu\psi(x) = e^{-ig_s \frac{\lambda^a}{2}\theta^a(x)}(\partial_\mu\psi(x) - ig_s\frac{\lambda^a}{2}\partial_\mu\theta^a(x)\psi(x)) \end{aligned} \tag{2.3}$$

where g_s is the coupling constant of QCD. In order for the Lagrangian to respect the SU(3)$_\mathrm{C}$ invariance, the derivative term is changed to the covariant derivative:

$$\partial_\mu\psi(x)_i \to D_\mu\psi(x)_i = \partial_\mu\psi(x)_i + ig_sA^a_\mu(x)\frac{\lambda^a_{ij}}{2}\psi(x)_j \tag{2.4}$$

where A_μ is the connection field that correspond to the gluons, and i and j are the color indices. The gluon fields transform as:

$$A^a_\mu(x) \to A^a_\mu(x) + \partial_\mu\theta^a + g_sf^{abc}A^b_\mu\theta^c \tag{2.5}$$

The kinetic term for the gluon fields is:

$$-\frac{1}{4}F_a^{\mu\nu}F_{\mu\nu}^a \tag{2.6}$$

where

$$F_{\mu\nu}^a = \partial\mu A_\nu^a - \partial\nu A_\mu^a - g_s f^{abc} A_\mu^b A_\nu^c. \tag{2.7}$$

Then, the Lagrangian can be written as:

$$\mathcal{L}_{QCD} = -g_s\bar{\psi}(x)\gamma^\mu A_\mu^a(x)\frac{\lambda^a}{2}\psi(x) - \frac{1}{4}F_a^{\mu\nu}F_{\mu\nu}^a. \tag{2.8}$$

The first term represents the interaction of the quark with the vector gluon field A_μ, with the coupling strength g_s. Sometimes the interaction is written as a function of the strong coupling constant $\alpha_s = g_s^2/4\pi$. In the kinematic term of the gluon field, the $f^{abc}A_\mu^b A_\nu^c$ terms of Eq. (2.7) create self-interactions between the gluon fields of cubic and quartic order, due to the non commuting property of the generators of the SU$(3)_C$ non-abelian group. Requiring the local gauge invariance leads to the introduction of eight gauge bosons (the gluons) and to the description of their interactions with the fermionic fields of the quarks. The eight gluons differ by the color and anticolor charge that they carry, while the quarks can have three color eigenstates.

2.2.2 Weak interactions

Electroweak interactions are explained in the SM with a similar local gauge invariance as strong interactions, by imposing a symmetry under the SU$(2)_L$ $\times$ U$(1)_Y$ group. Experimental observations show that parity is violated by weak interactions, which is accounted for in the theoretical description by assigning different interactions to fermions of opposite chiralities. In the limit of a zero mass for the particles, the chirality corresponds to the helicity, which is defined as the normalized projection of the spin vector along the momentum direction. To define the chiral components ψ_L and ψ_R, the chirality projection operators with the $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ matrix are applied on the fields : $\psi_L = \frac{1-\gamma_5}{2}\psi$ and $\psi_R = \frac{1+\gamma_5}{2}\psi$.

The U$(1)_Y$ gauge group is abelian, with the gauge field B_μ resulting from the local invariance, and is associated with the weak hypercharge Y quantum number. This field interacts independently with both the chiral components of the spinor fields, according to the coupling constant g'.

The SU$(2)_L$ group is non-abelian and is associated with the weak isospin I_3, and the presence of 3 gauge fields, W_μ^i, that arise from the local gauge invariance. The generators of the group are the $T_i = \frac{\sigma_i}{2}$, where σ_i are the 3 Pauli matrices with null trace

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \ \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \ \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{2.9}$$

that obey the commutation relation, such that $\left[\frac{\sigma_i}{2}, \frac{\sigma_j}{2}\right] = i\epsilon_{ijk}\frac{\sigma_k}{2}$. The left chiral component of the fermion fields are doublets under the SU$(2)_L$ group, so they interact and mix, while the right components represents singlets, so they do not interact with the gauge fields. The strength of the weak interactions is defined by the coupling constant g. The representation of the fields can be written as:

$$\begin{aligned} \Psi_L &= \begin{pmatrix} \psi'_L \\ \psi''_L \end{pmatrix} = \frac{1-\gamma_5}{2} \begin{pmatrix} \psi' \\ \psi'' \end{pmatrix} \\ \psi'_R &= \frac{1+\gamma_5}{2}\psi' \\ \psi''_R &= \frac{1+\gamma_5}{2}\psi'' \end{aligned} \tag{2.10}$$

where the ψ' and ψ'' are the fermions of the same family, either the up- and down-quarks, or a neutrino and the corresponding lepton. The lepton and quark sectors are disjoint and their fields cannot mix through strong or electroweak interactions. The weak isospin and hypercharge are related to the electric charge by:

$$Q = I_3 + \frac{Y}{2}. \tag{2.11}$$

The covariant derivative for the SU$(2)_L \times$ U$(1)_Y$ gauge group is:

$$\partial_\mu \to D_\mu = \partial_\mu + igW_\mu^a(x)T_a + ig'\frac{Y}{2}B_\mu(x) \tag{2.12}$$

where the generator T_a couples with the gauge fields fields and the right handed components Ψ_L, while the hypercharge field couples with both the right- and left-handed components. The Lagrangian can be written as:

$$\mathcal{L}_{ewk} = i\bar{\Psi}_L \not{D}\Psi_L + i\bar{\psi}'_R \not{D}\psi'_R + i\bar{\psi}''_R \not{D}\psi''_R = \mathcal{L}_{\text{free}} + \mathcal{L}_{CC} + \mathcal{L}_{NC}, \tag{2.13}$$

where the $\mathcal{L}_{CC}$ and $\mathcal{L}_{NC}$ are the charged and neutral current terms. Expanding the calculations:

$$\mathcal{L}_{\text{free}} = i\bar{\Psi}_L \not{\partial}\Psi_L + i\bar{\psi}'_R \not{\partial}\psi'_R + i\bar{\psi}''_R \not{\partial}\psi''_R; \tag{2.14}$$

and

$$\begin{aligned}\mathcal{L}_{CC} &= ig\bar{\Psi}_L(\gamma^\mu W^1_\mu\frac{\sigma_1}{2}+\gamma^\mu W^2_\mu\frac{\sigma_2}{2})\Psi_L \\ &= i\frac{g}{\sqrt{2}}\bar{\Psi}_L(\gamma^\mu W^+_\mu\sigma_+ + \gamma^\mu W^-_\mu\sigma_-)\Psi_L \\ &= i\frac{g}{\sqrt{2}}\bar{\psi}'_L\gamma^\mu W^+_\mu\psi''_L + i\frac{g}{\sqrt{2}}\bar{\psi}''_L\gamma^\mu W^-_\mu\psi'_L,\end{aligned} \tag{2.15}$$

where

$$\begin{aligned}W^\pm_\mu &= \frac{1}{\sqrt{2}}(W^1_\mu \mp iW^2_\mu) \\ \sigma_\pm &= \frac{1}{\sqrt{2}}(\sigma_1 \pm i\sigma_2),\end{aligned} \tag{2.16}$$

and the σ matrices act on the indices of the left-chirality doublet. The charge current interaction couples the up- and down-type fields of the same family with charged boson fields $W^\pm_\mu$. A neutral current interaction also exists,

$$\begin{aligned}\mathcal{L}_{NC} =& ig\bar{\Psi}_L\gamma^\mu W^3_\mu\frac{\sigma_3}{2}\Psi_L \\ &+ ig'\frac{Y_L}{2}\bar{\psi}'_L B_\mu\psi'_L + ig'\frac{Y_L}{2}\bar{\psi}''_L B_\mu\psi''_L + ig'\frac{Y_{R'}}{2}\bar{\psi}'_R B_\mu\psi'_R + ig'\frac{Y_{R''}}{2}\bar{\psi}''_R B_\mu\psi''_R,\end{aligned} \tag{2.17}$$

although neither the W^3_μ nor the B_μ fields can be interpreted as the photon field since they couple also to neutral fields.

However it is possible to mix these two fields to obtain the the photon field (A_μ) and the neutral Z boson field (Z_μ), by introducing the Weinberg angle θ_W:

$$A_\mu = B_\mu\cos\theta_W + W^3_\mu\sin\theta_W \tag{2.18}$$

$$Z_\mu = B_\mu\sin\theta_W - W^3_\mu\cos\theta_W \tag{2.19}$$

where the Weinberg angle is related to g,g' and the electric charge e such that

$$g\ \sin\theta_W = g'\ \cos\theta_W = e. \tag{2.20}$$

Substituting these fields in Eq. (2.17) gives a coupling constant to the photon fields of:

$$gI_3\sin\theta_W + g'\frac{Y}{2}\cos\theta_W, \tag{2.21}$$

which is equivalent to the relation (2.11) for Q, which ties together the weak and

Table 2.1: Summary of the fermion field $SU(2)_3 \times SU(2)_L \times U(1)_Y$ representations for the left and right chiral components. The relation between the electroweak quantum numbers is $Q = I_3 + Y/2$.

Type	Family 1st	Family 2nd	Family 3rd	$SU(2)_L \times U(1)_Y$ I_3	Y	Q	$SU(3)_C$
Leptons	$\begin{pmatrix}\nu_{e,L}\\ e_L\end{pmatrix}$	$\begin{pmatrix}\nu_{\mu,L}\\ \mu_L\end{pmatrix}$	$\begin{pmatrix}\nu_{\tau,L}\\ \tau_L\end{pmatrix}$	$\begin{pmatrix}\frac{1}{2}\\ -\frac{1}{2}\end{pmatrix}$	-1	$\begin{pmatrix}0\\ -1\end{pmatrix}$	singlet
	$\nu_{e,R}$	$\nu_{\mu,R}$	$\nu_{\tau,R}$	0	0	0	
	e_R	μ_R	τ_R	0	-2	1	
Quarks	$\begin{pmatrix}u_L\\ d_L\end{pmatrix}$	$\begin{pmatrix}c_L\\ s_L\end{pmatrix}$	$\begin{pmatrix}t_L\\ b_L\end{pmatrix}$	$\begin{pmatrix}\frac{1}{2}\\ -\frac{1}{2}\end{pmatrix}$	$\frac{1}{3}$	$\begin{pmatrix}\frac{2}{3}\\ -\frac{1}{3}\end{pmatrix}$	triplet
	u_R	c_R	t_R	0	$\frac{4}{3}$	$\frac{2}{3}$	
	d_R	s_R	b_R	0	$-\frac{2}{3}$	$-\frac{1}{3}$	

electromagnetic quantum numbers. For the coupling coefficients of the gauge fields and the fermion fields, the values reported in Tab.2.1 can be chosen.

Similarly to the QCD case, for the SU(2)$_L$ and U(1)$_Y$ fields, the kinetic term tensors can be defined as:

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \tag{2.22}$$

$$W_{\mu\nu} = \partial_\mu W_\nu - \partial_\nu W_\mu - g\varepsilon^{abc} W^a_\mu W^b_\nu \tag{2.23}$$

Once inserted in the Lagrangian, they give rise to tri-linear and quadri-linear interactions of the kind: ZWW, γWW, ZZWW, $\gamma\gamma$WW, γZWW, and WWWW.

A summary of the properties of the fermion field as SM gauge symmetry representations is shown in Tab.2.1. Left and right chirality fields are respectively a doublet and a singlet of the $SU(2)_L$ group, so just the former is subject to the charged interaction, via the mediator W$^\pm$.

For the quarks, since the eigenstates of the strong interaction and weak interaction are different, the strength of weak interaction between flavor eigenstates is described by the coupling coefficients of the Cabibbo-Kobayashi-Maskawa (CKM) 3×3 unitary matrix [4, 7, 8]:

$$\mathrm{M_{CMK}} = \begin{pmatrix} 0.97446 \pm 0.00010 & 0.22452 \pm 0.00044 & 0.00365 \pm 0.00012 \\ 0.22438 \pm 0.00044 & 0.97359^{+0.00010}_{-0.00011} & 0.04214 \pm 0.00076 \\ 0.00896^{+0.00024}_{-0.00023} & 0.04133 \pm 0.00074 & 0.999105 \pm 0.000032 \end{pmatrix}. \tag{2.24}$$

The Z boson mediates the neutral weak interaction with both chiral components according to a different strength, due to the mixing of the gauge fields according to the Weinberg angle. The resulting vertex coupling can be written as:

$$\frac{-ie}{2\cos(\theta_W)\sin(\theta_W)}\gamma_\mu(c_V - c_A), \tag{2.25}$$

where c_V and c_A are the vectorial and axial couplings for the fermions reported in Tab. 2.2 for a value of $\sin(\theta_W) = 0.23$.

Table 2.2: Summary of the fermion vectorial and axial couplings to the Z boson field, where a value of $\sin(\theta_W) = 0.23$ is assumed.

Fermion	c_V	c_A
ν_e, ν_μ, ν_τ	0.50	0.50
e, μ, τ	0.04	-0.50
u, c, t	0.19	0.50
d, s, b	-0.35	-0.50

The photon is the mediator of the electromagnetic force and couples to fermions proportionally to their charge, which is related to the weak isospin and hypercharge. Interactions can change the quantum numbers of the fields through the charge carried by the mediators. Charged weak interactions change the weak isospin, thus the electric charge, whereas strong interactions change the color charge of quarks.

The electromagnetic mediator, the photon, is massless, but on the other hand the limited range of the weak interactions implies that their mediators are massive. The observations of the $W^\pm$ and Z bosons at the UA1 [9, 10] and UA2 [11, 12] experiments confirmed that they are not massless, being the $M_\mathrm{W} = 80.385 \pm 0.015$ GeV and $M_\mathrm{Z} = 91.187 \pm 0.0021$ GeV [13]. However, explicit mass terms of the gauge fields would break the gauge invariance. Direct fermion mass terms are also not allowed, because they are not invariant under the gauge transformations, being that $m\bar{\psi}\psi = m(\bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R)$, where the left and chiral components are linked together and transform differently between $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$. The solution needed to explain boson and fermion masses is provided by the Brout-Englert-Higgs-Guralnik-Hagen-Kibble mechanism [14–16], with a natural way of breaking the $SU(2)_L \times U(1)_Y$ symmetry to $U(1)_{em}$ without explicitly violating local gauge invariance.

2.3 Electroweak symmetry breaking and the Higgs boson

In 1964, theorists proposed a mechanism through which a complex scalar field with non-zero vacuum expectation value was introduced into the Lagrangian, resulting in the breaking of the electroweak symmetry. This mechanism is the Brout-Englert-Higgs-Guralnik-Hagen-Kibble mechanism and postulates the existence of a new scalar particle, called the Higgs boson. The Lagrangian term for this scalar field takes the form:

$$\mathcal{L} = T - V = (D_\nu \Phi)^\dagger (D^\nu \Phi) - (\mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2) \tag{2.26}$$

with $\lambda > 0$, and is invariant under the space rotation $\Phi \to e^{i\alpha}\Phi$. The Higgs field can be assumed to be a complex scalar isospin doublet and associated with a hypercharge equal to 1:

$$\Phi = \begin{pmatrix} \phi_+ \\ \phi_0 \end{pmatrix} = \begin{pmatrix} \phi_+ \\ \phi_0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}, \tag{2.27}$$

where the ϕ_i fields are real scalar fields.

If the potential parameter $\mu^2 > 0$ then the potential is simply a term of mass μ added to a term that has a four-linear vertex with coupling λ, so it is self-interacting. If $\mu^2 < 0$, then the potential has minima:

$$\frac{\partial V}{\partial \Phi} = 0 \Rightarrow \Phi^\dagger \Phi = -\frac{\mu^2}{2\lambda} = \frac{v^2}{2} \tag{2.28}$$

with infinite solutions on a circle of radius v, called thevacuum expectation value (VEV) of the scalar potential, and are connected through gauge transformations that change the phase of the field but not its modulus, i.e. rotations. Once a specific ground state is chosen, the symmetry is explicitly broken, but the Lagrangian is still gauge invariant with all the important consequences for the existence of gauge interactions. The covariant derivative is:

$$D_\mu = \partial_\mu - igW^i_\mu \frac{\sigma^i}{2} + \frac{ig'}{2} B_\mu, \tag{2.29}$$

where the scalar field is assumed to have hypercharge 1 and isospin -1/2, in a way that its electric charge is 0 and that it is invariant under $U(1)_{em}$ transformations in order to keep the photon massless. The perturbative expansion of the Higgs field can be done around the minima

$$\Phi(x) = \frac{1}{\sqrt{2}} e^{\frac{i\sigma^i \theta_i(x))}{v}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}, \tag{2.30}$$

where there are three massless fields $\theta_i(x)$ and a real scalar field $H(x)$ whose quanta correspond to a new physical massive particle, the Higgs boson (H). The presence of the former fields is expected as a consequence of the Goldstone theorem that states that the spontaneous breaking of a continuous symmetry generates as many massless bosons (Goldstone boson) as broken generators of the symmetry. These fields can be absorbed by the choice of a particular gauge, called the unitary gauge with a transformation to

$$\Phi(x) \to \Phi' = e^{\frac{-i\sigma^i \theta_i(x))}{v}} \Phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}. \tag{2.31}$$

By substituting the part of the Lagrangian with the coviariant derivative, the

$$(igW_\mu^i \frac{\sigma^i}{2}\Phi)^\dagger (igW_\mu^i \frac{\sigma^i}{2}\Phi) = \frac{g^2 v^2}{8}[(W_\mu^1)^2 + (W_\mu^2)^2 + (W_\mu^3)^2] \tag{2.32}$$

with these three terms being the mass terms for the bosons. The Lagrangian, in details, becomes

$$\begin{aligned} \mathcal{L} =& \frac{1}{2}\partial^\mu H \partial_\mu H - \frac{1}{2}(2\lambda v^2)H^2 \\ &+ \left[\left(\frac{gv}{2}\right)^2 W_\mu^+ W^{\mu -} + \frac{(g^2 + g'^2)v^2}{8} Z_\mu Z^\mu\right]\left(1 + \frac{H}{v}\right)^2 \\ &+ \lambda v H^3 + \frac{\lambda}{4}H^4 - \frac{\lambda}{4}v^4. \end{aligned} \tag{2.33}$$

The first line represents the free Lagrangian of the new field with a mass of $m_H = \sqrt{2\lambda v^2} = \sqrt{2}|\mu|$, which is a free parameter of the model. The terms in the second line that multiply the constant represents the mass terms of the weak bosons, which get masses:

$$m_W^2 = \left(\frac{gv}{2}\right)^2 \tag{2.34}$$

and

$$m_Z^2 = \frac{(g^2 + g'^2)v^2}{8} = \frac{m_W^2}{cos^2\theta_W}. \tag{2.35}$$

The Goldstone bosons removed with the unitary gauge transformation are then absorbed as additional degrees of freedom of the W and Z bosons, corresponding to their longitudinal polarizations. The second line of Eq. (2.33) describes the interactions of the scalar particle with the vectorial fields HWW, HZZ and HHWW, HHZZ. The third line of Eq. (2.33) shows that cubic and quartic self-interactions of the Higgs boson are predicted.

There are at this point two free parameters of the mechanism: the VEV v and the Higgs boson mass m_{H}. The first corresponds to the energy scale of the electroweak symmetry breaking and can be computed from the Fermi constant G_F as:

$$\frac{G_F}{\sqrt{2}} = \left(\frac{g}{2\sqrt{2}}\right)^2 \frac{1}{m_W^2} \;\Rightarrow\; v = \frac{1}{\sqrt{\sqrt{2}G_F}} \simeq 246\,\mathrm{GeV} \tag{2.36}$$

The mass of the fermions arises from Yukawa interactions of the Higgs boson with their left and right chiral components, having couplings y_f, such that

$$\mathcal{L}_{Yukawa} = -y_{f''}\left(\bar{\Psi}_L \phi \psi''_R + \bar{\psi}''{}_R \phi^\dagger \Psi_L\right) - y_{f'}\left(\bar{\Psi}_L \tilde{\Phi} \psi'_R + \bar{\psi}'{}_R \tilde{\Phi}^\dagger \Psi_L\right), \tag{2.37}$$

where

$$\tilde{\phi} = i\sigma_2 \phi * = \begin{pmatrix} \phi_0^* \\ -\phi_+^* \end{pmatrix} = \frac{1}{2}\begin{pmatrix} v + H(x) \\ 0 \end{pmatrix}, \tag{2.38}$$

and $\tilde{\phi}$ transforms as ϕ but has opposite hypercharge. The gauge invariance is ensured since the hypercharge of the Higgs field and of the fermionic fields satisfy the relation $Y_\Phi = Y_L - Y_R$.

Then the Lagrangian density for the fermion masses becomes:

$$\mathcal{L}_{Yukawa} = \sum_f m_f(\bar{\psi_L}\psi_R + \bar{\psi_R}\psi_L)\left(1 + \frac{H}{v}\right) \tag{2.39}$$

where $m_f = y_f v/\sqrt{2}$ and the sum runs over the up- and down- type fermions. Fermion masses are thus taken into account in the SM as the interaction of the fermion fields with the Higgs field, which changes the chirality of the fermions, but not the flavor. The strengths of the interactions are directly related to the fermion masses, and are free parameters of the theory. The SM does not explain the origin of these couplings and their values that can differ by many orders of magnitude, nor does it explain the hierarchy of the three fermion families.

2.4 Phenomenology of the Higgs boson and experimental status

As already pointed out, the Higgs boson mass and its Yukawa coupling to the fermions are not predicted, but are free parameters of the theory. However the SM predicts that at pp colliders such as the LHC, the main production mechanisms are the ones represented in Fig.2.1, with cross sections that vary as a function of the Higgs mass, which is another free parameter of the theory. As shown in Fig.2.1 (right), for pp

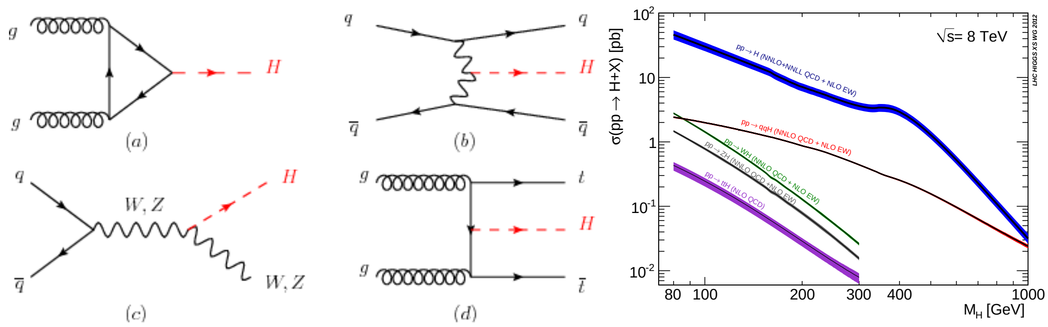


Figure 2.1: Left: representative Feynman diagrams for the Higgs production mechanisms: a) gluon fusion; b) vector boson fusion (V=W, Z); c) Higgs boson associated W and Z production; d) t$\bar{t}$ H associated production [4]. Right: Higgs boson production cross sections for different Higgs boson production processes for pp collisions with center of mass energy $\sqrt{s}$ = 8 TeV as a function of the Higgs boson mass [17].

collisions with a center of mass energy of $\sqrt{s}$ = 8 TeV of the center of mass energy, the gluon fusion (ggF) production has the highest production cross section, in which two gluons produce a loop of heavy quarks that then couple with the Higgs boson. The loop is necessary since the massless gluons don't couple directly with the Higgs field. The most dominant loop contribution is the one due to top quarks, since they are the most massive. One order of magnitude smaller than ggF is the vector boson fusion (VBF), where a pair of quark from the protons radiate heavy W or Z bosons, that then couple to the Higgs boson field. Together with the bosons, also a pair of quarks is created, produced close to the initial direction of the incoming quarks, in a way that the final state is expected to have two jets that are far apart in the two opposite forward regions, such that the two outgoing jets have a large invariant mass. The Higgs boson can also be produced in association with a single vector boson (VH, V = W$^{\pm}$ or Z). Both the VH and VBF production measurements allow the Higgs boson coupling to vector bosons to be probed. Finally, Higgs bosons can be produced in association with a pair of bottom or top quark (b$\bar{b}$H and t$\bar{t}$ H) or a single top quark (tH), with a rate depending on the magnitude of the Yukawa coupling y_b, y_t and their

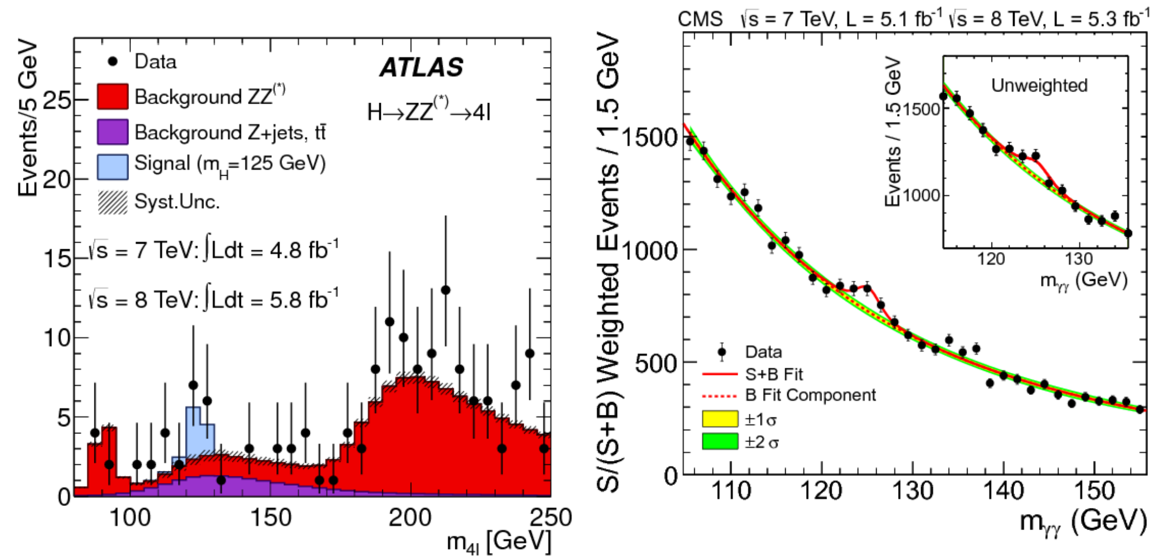


Figure 2.3: Distribution of the four lepton invariant mass, $m_{4\ell}$, for the combination of the $\sqrt{s} = 7\,\text{TeV}$ and $\sqrt{s} = 8\,\text{TeV}$ data collected by ATLAS [19]. The diphoton invariant mass distribution with each event weighted by the S/(S+B) value of its category for data collected by the CMS Collaboration at $\sqrt{s} = 7\,\text{TeV}$ and $\sqrt{s} = 8\,\text{TeV}$. The lines represent the fitted background and signal. The inset shows the central part of the unweighted invariant mass distribution [20].

sign, respectively.

The branching ratios ($\mathcal{B}$) depend on the Higgs boson mass, due to the kinematically allowed phase space, as depicted in Fig. 2.2.

In July 2012 experimental proof of the BEHGHL mechanism was deermined with the discovery of a new scalar boson of mass $\sim 125\,\text{GeV}$ announced by the ATLAS and CMS Collaborations [19–21] in data collected at $\sqrt{s} = 7\,\text{TeV}$ and $\sqrt{s} = 8\,\text{TeV}$. The sensitivity in the discovery was dominated by the $\text{H} \to \gamma\gamma$ and $\text{H} \to \text{ZZ}^* \to \ell^+\ell^-\ell'^+\ell'^-$ $(\ell = e, \mu)$ decay channels, even though they are among the lowest in terms of branching fraction, because they provide the highest purity and mass resolution, as shown in Fig. 2.3.

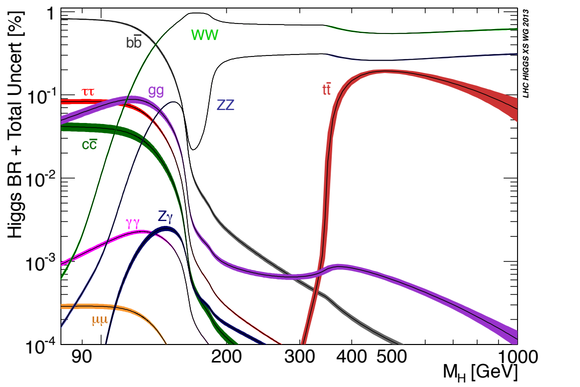


Figure 2.2: Branching ratios of the Higgs boson into SM particles for different values of the boson mass [18].

The Run I Higgs boson discovery was performed inclusively for all the production mechanisms. The combination of the ATLAS and CMS experiments results lead to a

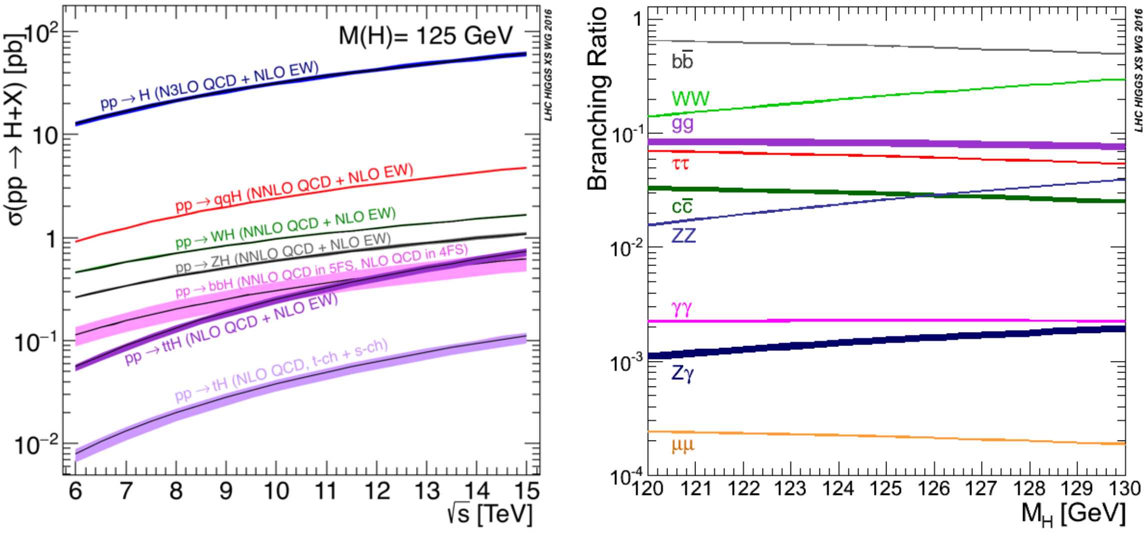


Figure 2.4: Cross sections for different Higgs boson production processes for pp collisions of various center of mass energies for a Higgs boson mass of 125 GeV (left). Branching ratios of the Higgs boson for a Higgs boson mass around 125 GeV (right) [17, 18].

precise determination of $m_H = 125.09 \pm 0.21(\text{stat.}) \pm 0.11(\text{syst.})\,\text{GeV}$, which is still the most precise to date [22].

For collisions with $\sqrt{s} = 13\,\text{TeV}$, the various contributions to the production cross section are shown in Fig.2.4 (left), together with the branching ratios for the Higgs boson decay (right).

Already in Run 1, the new particle was found to be compatible with being a scalar, with spin-parity of JP$= 0^+$ [23]. Furthermore, the combined measurement performed by the ATLAS and CMS experiments confirms the agreement with the SM predictions for the couplings with the fermions [24]. In Run 2, properties of the Higgs boson were further explored: the decay to $\tau^+\tau^-$ pairs was established by the CMS and ATLAS experiments independently [25, 26]. Evidence of the decay mode to bottom quark pairs was observed by both collaborations [27–29] and an upper limit on the the cross section times branching fraction of $H \to \mu\mu$ was obtained by CMS [30] of 2.9 times the SM value. The coupling between the Higgs boson and the SM particles was also studied with data collected in 2016 at $\sqrt{(s)} = 13\,\text{TeV}$, and found to be consistent with previous measurements and SM expectations [31].

The dependence of the coupling strength of the Higgs boson to the fermions and other SM boson has been probed extensively using a mass range that extend for three orders of magnitude, and found experimentally to be in perfect agreement with the standard model expectation. Also, the parity and spin of the new particle, checked with the

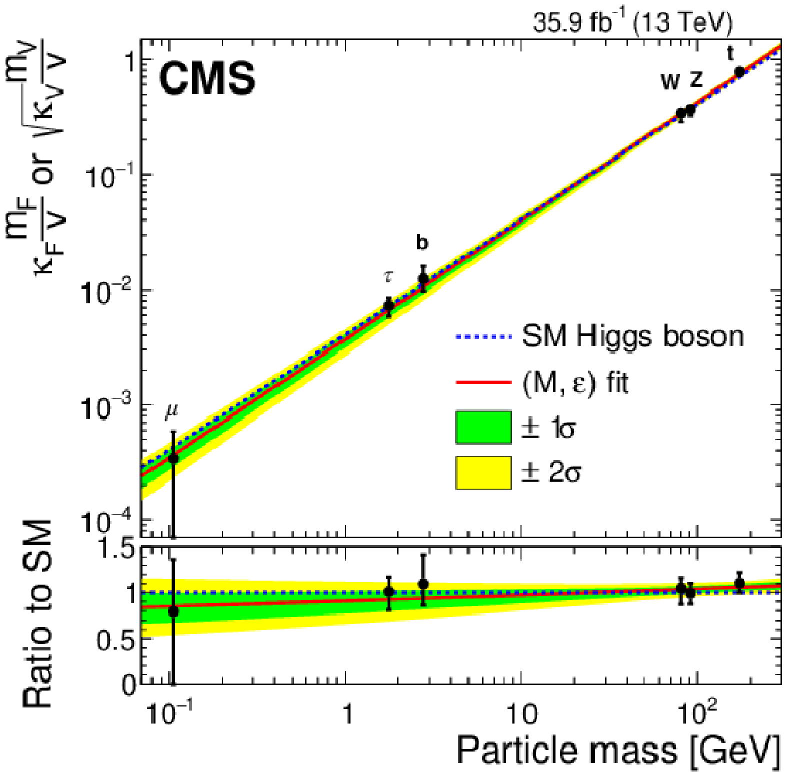


Figure 2.5: Couplings of the Higgs boson to the fermions measured by the CMS Collaboration with the data collected in 2016. The SM prediction is in agreement within the uncertainties [31].

angular distributions of the final decay products, confirm the compatibility of the observed particle with the scalar boson predicted by the SM.

Chapter 3

Beyond the standard model

Despite the incredible success of the SM in describing the data collected in the last decades by different experiments, there are phenomena that are not adequately explained in the SM picture. Some of the open points arise within the theory. For instance, the existence of three families of fermions identical but for their coupling to the scalar boson, is assumed, but not explained, and causes the masses of fermions to span over many orders of magnitude.

In the universe, there is a large asymmetry between matter and anti-matter ($\mathcal{O} = 10^9$), too large to be explained by the SM. In fact, in the SM, only the charged weak interaction distinguishes matter and anti-matter, with a very small asymmetry, while the other interactions produce or annihilate matter and antimatter with no asymmetry. Therefore, the size of CP violation observed so far in SM interactions is too small to account for the present-day matter-antimatter imbalance [32], suggesting the existence of additional as-yet-unknown sources of CP violation.

Furthermore, other compelling theoretical and experimental motivations suggest that the SM is not the ultimate theory able to completely describe the laws of nature, as is briefly described in the following.

From astrophysical observations, it appears that just 5% of the universe is made of the known matter that is part of the SM, whereas 26% of the universe is believed to be made of dark matter and the remaining 69% of dark energy. The former is postulated because the orbital velocity of galaxies within clusters is too high to be explained by the gravitational pull of visible matter alone [33, 34]. The observation of dark energy arises from the fact that the expansion of the universe is accelerating, with galaxies receding from each other at a rate that increases with distance [35]. The SM cannot provide particle fields compatible with the properties of either the dark matter or the dark energy.

Another shortcoming is that the SM does not include the fourth fundamental interaction, gravity. The gravitational force is relevant at energies on the order of the Planck mass ($M_P = \sqrt{\hbar c/G} \approx 22 \cdot 10^{-6}$g $\approx 1.2 \cdot 10^{19}$ GeV), while the electroweak interactions happen at energies around the GeV–TeV scale. This difference of many orders of magnitude in the scales is referred to as the hierarchy problem. In more detail, for the SM to be renormalizable, the Higgs boson mass is subject to radiative contributions via 1 loop diagrams with other SM particles that are quadratically divergent. If the SM is expected to be valid up to a cutoff energy scale Λ, the corrections to the Higgs boson mass are

$$\delta m_H^2 \approx \frac{3\Lambda^2}{8\pi^2 v^2}\left(2m_W^2 + m_Z^2 + m_H^2 - 4m_t^2\right) \approx -\frac{\Lambda^2}{25}. \tag{3.1}$$

If the SM holds its validity until the Planck scale ($\Lambda = M_P$) the radiative corrections are about 30 orders of magnitude bigger than M_H. Regardless of the Λ scale, the quadratic dependence of the divergence requires an extreme *fine-tuning* of the SM parameters at higher energy scales. This problem is one of the reasons to expect BSM physics at the TeV scale.

In this context, it would be natural to think that the SM is only the manifestation of a more extended theory beyond it, in which the standard model is valid for a given energy interval. In this way the presence of BSM physics could provide a solution to these problems by changing and enlarging the structure of the SM while preserving its remarkable success at describing the phenomenology of collider experiments until now. In the following, two BSM scenarios that are interesting for this work are presented with their motivation and predictions. In particular they predict the existence of new particles with masses in the TeV range, which can be produced at colliders.

In this sense, it appears interesting to use the Higgs boson itself to probe for possible kinds of new physical interactions, also profiting from the LHC increase of the center of mass energy from 8 TeV to 13 TeV that can allow to investigate a wider range of energies for the presence of new interactions and particles, as hints of a new theory, as it was done in this thesis considering two main theoretical frameworks.

3.1 Warped extra dimensions

In the 1920s Kaluza and Klein combined electromagnetism and gravity, the two known interactions at that time, by considering that nature could consist of additional dimensions [36–38]. Since then, many new physics models with additional dimensions have been proposed to attempt to unify the forces of nature, by combining into a single theory the electroweak and strong interactions together with the gravitational force. Kaluza introduced a further spatial dimension, in addition to the three spatial and the temporal one taken into account by general relativity. Klein suggested that this new dimension could be warped on itself, so that the new dimension would extend over a finite distance, in a way to be so compact that current experiments would not have detected it. These sets of models have in common the existence of one or more extra dimensions, that can be infinite, warped (WED) or also with a range over a finite intervals.

In the model proposed by Randall and Sundrum (RS) [39], the extra dimension y is delimited between two 3-branes, meaning that they have three spatial dimensions. In this picture, the electroweak and the strong interaction fields live in a brane called the "infra-red" or "TeV" brane ($y = L$), while gravity belongs to the "ultra-violet" or "Planck" brane ($y = 0$), and the region separating them is called "bulk". The metric in 5 dimensions is:

$$ds^2 = W(y)\eta_{\mu\nu}dx^\mu dx^\nu - dy^2 = e^{-2ky}\eta_{\mu\nu}dx^\mu dx^\nu - dy^2 = e^{-2kr_c\phi}\eta_{\mu\nu}dx^\mu dx^\nu - r_c^2 d\phi^2, \quad (3.2)$$

where k is the curvature and a change of coordinates is performed to introduce the compactification radius $r_c = y/\phi$, with $0 < \phi < \pi$. The warp factor $W(y)$ controls the Minkowski metric in each four-dimensional (4D) brane at each point of the 5th dimension.

In this framework, the energy scale of the five dimensions is related to the one of the 4D space and the volume of the compactified space, V_n , as

$$M_P^2 = M^{n-2}V_n = \frac{M_5^3}{k}(1 - e^{-2kr_c\phi}). \quad (3.3)$$

The Plank mass is explained as a function of the more general scale M_5 and the curvature of the 5 dimensional theory. Generally, any mass m_0 or scale v_0 in a 3-brane would transformto a mass m on another brane (or a point in the five-dimensional space) :

$$m = e^{-kr_c\phi}m_0 \qquad \text{and} \qquad v = e^{-kr_c\phi}v_0. \quad (3.4)$$

In this way, the Planck mass at the UV scale would acquire a factor $e^{-kr_c\pi}$ at the IR brane, and would translate into a value of few TeV or smaller for sufficiently high kr_c.

In other words, the weakness of the gravitational force at energies of the electroweak interactions could be explained by its propagation and the exponentially suppressing factor along the extra dimension. The small exponential factor above is the source of the large hierarchy between the observed Planck and weak scales. In this sense, the WED feature explains the electroweak-Planck scale hierarchy problem.

When these models are perturbatively expanded, two kinds of particles arise: from an expasion around the four-dimensional part of the metric, a spin-2 graviton ($G_{\mu\nu}(x, y)$), and from the expansion around the fifth dimension, a spin-0 radion (R or Φ),

$$g_{\mu\nu} = e^{-2ky}\eta_{\mu\nu} \;\rightarrow\; e^{-2ky+F(x,y)}(\eta_{\mu\nu} + G_{\mu\nu}(x, y)). \tag{3.5}$$

The fluctuation of the size of the extra dimension y, $F(x, y)$ can be expressed as a function of a 4D radion field Φ, as

$$F_{\mu\nu}(x, y) \propto e^{2ky}\Phi(x), \tag{3.6}$$

where $\Phi(x)$ is the 4D wave function at a given point of the fifth dimension. The fluctuation of the 4D space time corresponds to the $G_{\mu\nu}(x, y)$ graviton field.

In general, warped extra dimensional models predict a set of massive resonances, called a tower, for each particle propagating in the extra dimension. These are called Kaluza-Klein (KK) excitations and yield observable particles with specific masses. These can be seen to originate from the warped metric in the following way. For a simple massless scalar field, the action can be written as

$$S = \int d^5x\partial^M\varphi * \partial_M\varphi \text{ with } M = 0, 1, 2, 3, 4. \tag{3.7}$$

In warped extra dimension, since the fifth dimension is a circle, it is possible to do a Fourier expansion of the field as

$$\varphi(x^\mu, y) = \sum_{n=-\infty}^{\infty} \varphi_n(x^\mu)e^{iny/r} \tag{3.8}$$

The field equations for the massless scalar are:

$$\partial^M\partial_M\varphi = 0 \Rightarrow \sum_{n=-\infty}^{\infty} (\partial^\mu\partial_\mu\varphi_n - \frac{n^2}{r^2}\varphi_n)e^{iny/r} = 0. \tag{3.9}$$

This means that in four dimensions there is an infinite number of fields that satisfy the Klein-Gordon equations ($\partial^\mu\partial_\mu\varphi_n - \frac{n^2}{r^2}\varphi_n$) for massive fields of mass $m_n = \frac{n^2}{r^2}$. This feature can be generalized to the other SM fields in an analogous manner, such that there are "KK towers" for each of the SM fields, with the massless "0-modes" corresponding to the currently observed-SM fields. The massless graviton is the mediator of the gravitational force, while the radion is a field required to stabilize the size L of the extra dimension, with its ground state related to the size of the fifth dimension.

The first graviton KK excitation is:

$$G^{(1)}_{\mu\nu}(x, y) \propto e^{2ky} J_2(e^{2ky} \frac{m_G}{k}) G_{\mu\nu}(x), \tag{3.10}$$

where J_2 is the second Bessel function and m_G is the first KK excitation graviton mass of

$$m_G = kx_1 \Lambda_G / \bar{M}_P,$$

with $x_1 = 3.83$ being the first zero of the first Bessel function, the reduced Planck mass is $\bar{M}_P = M_P/\sqrt{8\pi} = 2.4 \cdot 10^{18}$ GeV and the UV cut-off scale is $\Lambda_G = e^{-kL}\bar{M}_P$, which is of the order of a few TeV. Similarly, the radion scale Λ_R relates to the graviton scale as $\Lambda_R = \sqrt{6}\Lambda_G$ [40]. These first massive modes are localized towards the IR brane.

The interactions of the lightest modes of the graviton and radions with the SM fields are given by:

$$\mathcal{L} = -\frac{c_i}{\Lambda_G} G^{\mu\nu(1)} T^i_{\mu\nu} - \frac{d_i}{\Lambda_R} \phi T^{\mu i}_{\mu} \tag{3.11}$$

where $T^i_{\mu\nu}$ are the energy-momentum tensors of the SM fields. The radion couples with the trace of the tensor, which vanishes for massless fields [41]. The couplings of gravitational modes to the SM fields are set by $\Lambda_{G,R}$ which are about the weak scale and not the Planck scale. The Kaluza-Klein excitations can thus be produced at energies reached by colliders and should be observable as spin-0 or -2 resonances that can be reconstructed from their SM decay products.

Moreover, the KK-graviton production cross section is larger than the corresponding radion cross section due to the fact that the radion coupling to gluons is loop-induced, mostly from top-quark loops, whereas the KK-graviton has tree-level couplings to gluons as shown in Fig.3.1. In the scenario called RS1, the $q\bar{q}$ annihilation contributes to the production, but it is suppressed in the bulk scenario where the light quarks are localized at the Plank scale and the gluons are allowed to propagate in the extra dimension. The graviton production cross section is proportional to $\tilde{k}^2$, for a volume suppression factor mildly dependent on $\tilde{k}^2 = k/\bar{M}_P$, and the radion cross section is proportional to $1/\Lambda_R^2$ [42].

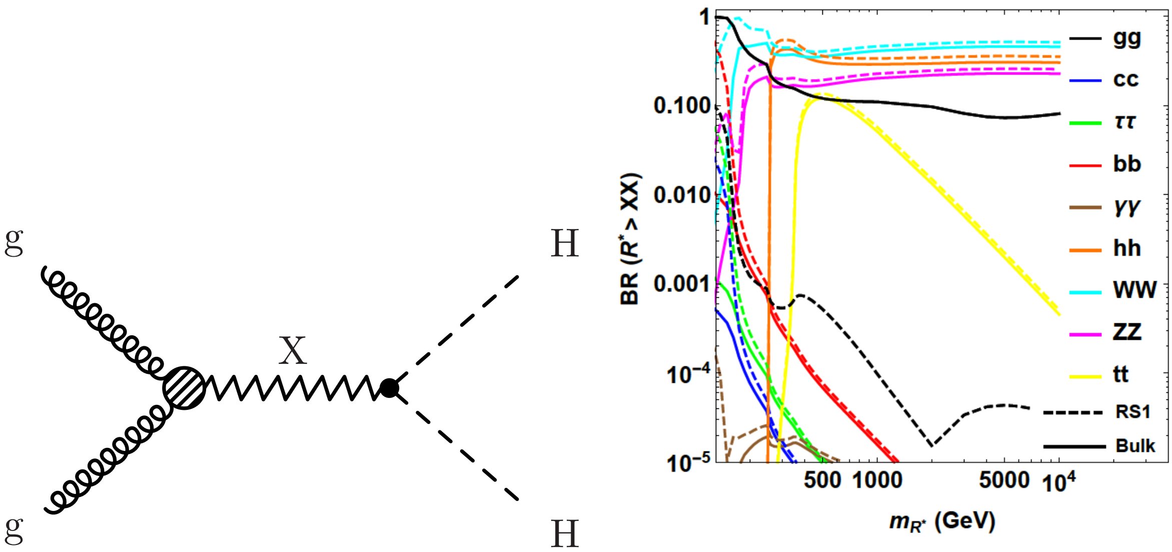


Figure 3.1: Feynman diagram of the dominant production mechanism of a radion or graviton and its decay to a pair of Higgs boson (left). Branching fractions of the lightest KK-radion decaying to SM particles as a function of its mass m_X in the RS1 (dashed line) and bulk (solid line) scenarios [42].

The most considered scenarios [41] are the RS1 scenario, where the SM particles are not allowed to propagate along the extra dimension, and the so-called bulk scenario where this constraint is removed [43].

In the RS1 scenario all the particles are localized at the TeV brane. Therefore the strength of the couplings between KK graviton and SM matter are democratic between each field, whereas the bulk scenario predicts that the SM fields, the Higgs, W, and Z bosons, are peaked towards the IR brane. The light fermions would be localized near the UV brane, explaining in this way their smaller masses and coupling to the Higgs boson. Consequently, the graviton and radion would couple predominantly to the Higgs boson, the top quark, and the longitudinal components of the W and the Z boson, whereas the photon and the gluon coupling would be suppressed by a factor $\sim 1/kL$. The branching ratios of the first massive KK-radion and graviton fields are displayed in Fig. 3.1(right) and Fig. 3.2 for the two scenarios.

Mixing between the radion and the Higgs boson is possible but is not taken into account here. As shown in Fig.3.1 the radion has one of its largest branching fractions into a pair of Higgs bosons, around 24%, that is constant as a function of the mass once kinematically possible and a width of the decay to bosons that is proportional to m_R^4/Λ_R^2, while the width to fermions (mostly the top quark) goes like m_R^2/Λ_R^2, so that the coupling to bosons is enhanced. The graviton branching fraction to a pair of Higgs bosons depends on the model, especially on how the top quark is localized, but with

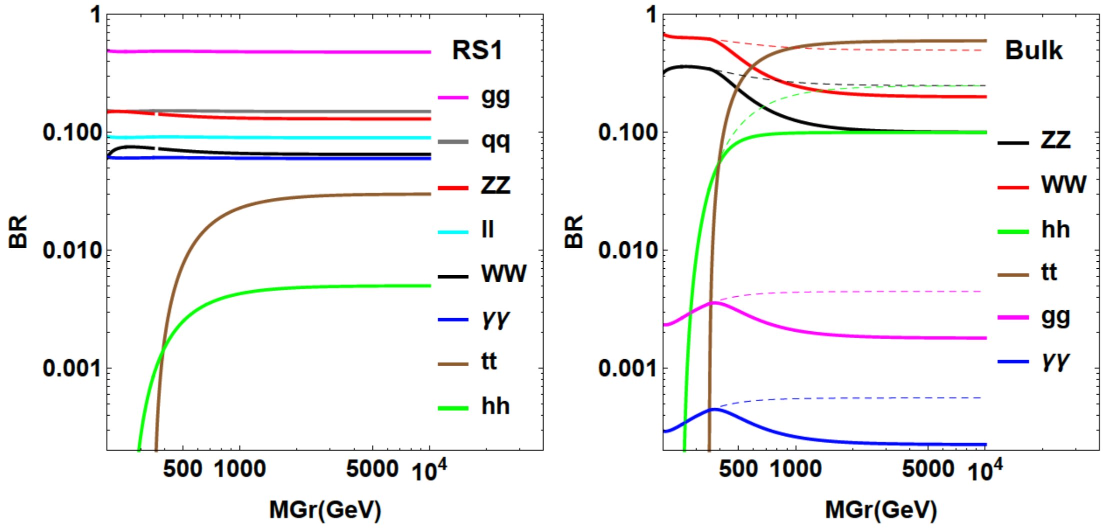


Figure 3.2: Branching ratios of the lightest KK-graviton decaying to the SM particles as a function of it mass m_X in the RS1 (left) and bulk (right) scenarios [42].

the parameters proposed here [41] it is around 10%, with a total width below 5%, for the bulk scenario with $\tilde{k} = 0.5$.

This kind of models are particularly interesting because they allows for a Higgs sector at the TeV scale and at the same time the unification of gauge couplings at high energy and provide a natural hierarchy of masses.

3.2 Heavy vector triplet

Another set of theoretical models predict the existence of spin-1 resonances as a manifestation of new physics. Such theories are mainly split into two classes: extended gauge [44, 45] or composite Higgs models [46, 47]. These models usually have a large number of free parameters that describe the dynamics, but the part concerning the on-shell production of a resonance has just a few important parameters: the mass and the couplings to the other fields that control its production and decay. Therefore, it is convenient to adopt an approach using a simplified model with an effective Lagrangian that describes the properties and interactions of the new particles using a limited set of parameters. The phenomenological parameters can be easily linked to physical observables at the LHC experiments. The simplified approach chosen to describe such a large class of models is the Heavy Vector Triplet framework [48], where only the relevant couplings and mass parameters are retained. The production cross section times branching fraction ($\sigma\mathcal{B}$) can be probed as a function of the invariant mass of the resonance, and be interpreted in the simplified model parameter space, which can then be applied to different models by computing relations between the parameters..

In this framework, a vector of real fields in the $SU(2)_L$ representation is introduced describing one neutral and two oppositely charged spin-1 fields

$$\mathrm{V}'^{\pm}_{\mu} = \frac{\mathrm{V}'^{1}_{\mu} \mp i\mathrm{V}'^{2}_{\mu}}{\sqrt{2}} \qquad \text{and} \qquad \mathrm{V}'^{0}_{\mu} = \mathrm{V}'^{3}_{\mu}. \tag{3.12}$$

The interactions of the new fields with the SM particles are presented in a phenomenological Lagrangian

$$\begin{aligned}\mathcal{L}_V = &-\frac{1}{4}(D_\mu \mathrm{V}'^a_\nu - D_\nu \mathrm{V}'^a_\mu)(D^\mu \mathrm{V}'^{\nu a} - D^\nu \mathrm{V}'^{\mu a}) + \frac{m^2_{\mathrm{V}'}}{2}\mathrm{V}'^a_\mu \mathrm{V}'^{\mu a} \\ &+ \frac{g^2}{g_V} c_F \mathrm{V}'^a_\mu \sum_f \bar{\Psi}_L \gamma^\mu \tau^a \Psi_L + i g_V c_H \mathrm{V}'^a_\mu (H^\dagger \tau^a D^\mu H - D^\mu H^\dagger \tau^a H) \\ &+ \frac{g_V}{2} c_{VVV} \epsilon_{abc} \mathrm{V}'^a_\mu \mathrm{V}'^b_\nu (D^\mu \mathrm{V}'^{\nu c} - D^\nu \mathrm{V}'^{\mu c})(D_\mu \mathrm{V}'^c_\nu - D_\nu \mathrm{V}'^c_\mu) \\ &+ g_V^2 c_{VVHH} \mathrm{V}'^a_\mu \mathrm{V}'^{\mu a} H^\dagger H - \frac{g}{2} c_{VVW} \epsilon_{abc} W^{\mu\nu a} \mathrm{V}'^b_\mu \mathrm{V}'^c_\nu .\end{aligned} \tag{3.13}$$

The first line represents the kinetic and mass terms of the heavy vector triplet bosons and their trilinear and quadri-linear interactions with the W and Z bosons of the SM $SU(2)_L$, where g is the SM $SU(2)_L$ coupling constant and g_V is the coupling constant of the new physics, with the covariant derivative:

$$D_\mu \mathrm{V}'^a_\nu = \partial_\mu \mathrm{V}'^a_\nu + g\epsilon^{abc} W^b_\mu \mathrm{V}'^c_\nu . \tag{3.14}$$

The V'^a fields are not the mass eigenstates because they couple and mix with the W_μ fields after the electroweak symmetry breaking, therefore $m_{\mathrm{V}'}$ doesn't exactly correspond to the physical mass of the new resonance.

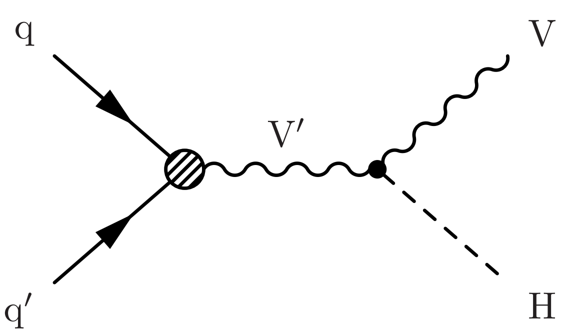


Figure 3.3: Feynman diagram for the $q\bar{q}$ production of a heavy vector boson V′ (W′ or Z′) that decays to a SM vector boson V and a Higgs boson H.

The second line indicates the coupling to the fermionic fields and the Higgs boson. The parameter c_F describes the interaction between the V′ boson and the fermions and is responsible for fermionic decays as well as the for $q\bar{q}$ production mode depicted in Fig. 3.3, and fermionic decays. For simplicity, a universal coupling to fermions is assumed, but in principle the coupling could be different for leptons, light and heavy flavor quarks. The term with the coupling coefficient c_H describes the vertices with the physical Higgs boson and the three unphysical Goldstone bosons that, because of the Goldstone equivalence theorem, represent the longitudinal polarizations, $W_L^{\pm}$ and Z_L, of the physical vector bosons. Therefore, c_H regulates the decay of the new resonances to the SM bosons.

The third line of the equation contains new operators and free parameters, which regulate the V′-Wmixing. However, this is of marginal effect, and these terms do not contain other interactions with SM fields, so they are irrelevant for the energies reached at colliders. Therefore, to a first approximation they can be neglected.

The parameters of the Lagrangian can be interpreted in a simplified description. The free parameter g_V is the typical strength of V′ interactions and can vary over an order of magnitude depending on the model, as they are $g_V \sim 1$ in the weakly coupled scenario and $g_V \sim 4\pi$ in the extremely strong limit. The dimensionless coefficient c are usually of order ~ 1 and parametrize the departure from the typical strength: although the coefficient c_F is of order one in most of the explicit models, the parameter c_H is of order one in the strongly-coupled scenario, but can be reduced in a weakly coupled case. For the purpose of analyzing and presenting experimental results, the combinations $g_V c_H$ and $g^2 c_F / g_V$ that enter in the vertices are instead treated as fundamental parameters, as they control production and decay rates.

After electroweak symmetry breaking, the heavy vector acquires mass and it is found that the charged and neutral W′ and Z′ bosons are expected to be practically degenerate ($M_{\pm} \simeq M_0 \simeq M_{\mathrm{V'}}$), which implies that they have comparable production and decay rates at the hadron collider. The partial widths of the resonance to fermions and bosons are:

$$
\begin{aligned}
\Gamma_{\mathrm{V}'^{\pm} \to f\bar{f}'} &\simeq 2\Gamma_{\mathrm{V}'^{0} \to f\bar{f}} \simeq N_c[f] \left(\frac{g^2 c_F}{g_V} \right)^2 \frac{M_{\mathrm{V'}}}{48\pi} \\
\Gamma_{\mathrm{V}'^{\pm} \to \mathrm{WZ}} &\simeq \Gamma_{\mathrm{V}'^{0} \to \mathrm{WW}} \simeq \frac{g_V^2 c_H^2 M_{\mathrm{V'}}}{192\pi} \\
\Gamma_{\mathrm{V}'^{\pm} \to \mathrm{WH}} &\simeq \Gamma_{\mathrm{V}'^{0} \to \mathrm{ZH}} \simeq \frac{g_V^2 c_H^2 M_{\mathrm{V'}}}{192\pi} \,,
\end{aligned}
\tag{3.15}
$$

where $N_c[f]$ is the number of colors (3 for quarks, 1 for leptons).

In general, the couplings of the new resonances to fermions and bosons can depend on several parameters of the specific theoretical models. In the following, two simplified scenarios (A and B) are discussed, exemplifying a broader classes of models. Scenario A considers different ranges of g_V, relatively small values $g_V \lesssim 3$ and represents the weakly-coupled extensions of the SM gauge group. On the contrary, the B scenario considers $g_V \gtrsim 3$ and describes the strongly-coupled scenarios. Usually in these scenarios two benchmark models are used. The benchmark for model A corresponds to $g_V = 1$ and represent the sequential model in [44], where a generalization of the SM is done by extending the gauge symmetry with an additional $SU(2)$. In this HVT scenario, the couplings are $c_F \sim 1$ and $c_H \sim -g^2/g_V^2$ so that

$$g_V c_H \sim g^2/g_V \qquad \text{and} \qquad g^2 c_F/g_V \sim g^2/g_V, \tag{3.16}$$

which means that the coupling to fermions and bosons is of the same order of magnitude. The total width of the new resonances in Model A goes as g^2/g_V.

The benchmark of model B corresponds to $g_V = 3$ and represents composite Higgs model in [46]. The Higgs boson in this case is the result of spontaneous symmetry breaking of an $SO(5)$ symmetry to a $SO(4)$ group. In this case, c_H is unsuppressed, and the couplings are

$$g_V c_H \sim -g_V \qquad \text{and} \qquad g^2 c_F/g_V \sim g^2/g_V. \tag{3.17}$$

Therefore, the dominant branching fractions are to dibosons, whereas the fermionic decays are extremely suppressed, between around one percent and one per mill. The total width of the new resonances in Model B increases with g_V.

The branching fractions of the new neutral spin-1 resonance differ in the two scenarios and are shown in Fig. 3.4 for the different decay modes as a function of the resonance mass in the benchmark models of scenario A ($g_V = 1$) and B ($g_V = 3$). Similarly, the behavior of the width of the resonance as a function of the mass is depicted in Fig. 3.5, for different values of the parameter g_V.

When the resonances start to be broad, i.e. $\Gamma/M_{\mathrm{V'}} \sim 10\%$, the assumptions leading to the simplified model are no longer valid. In fact, higher order and non-resonant effects have to be taken into account and are not included in this simplified framework since they might contribute to the tail and substantially change the prediction of the model. In the same way, from the empirical point of view, experimental results in the resonant region are sensitive to the limited number of the phenomenological Lagrangian parameters while the results of an experimental search which is sensitive to the tail

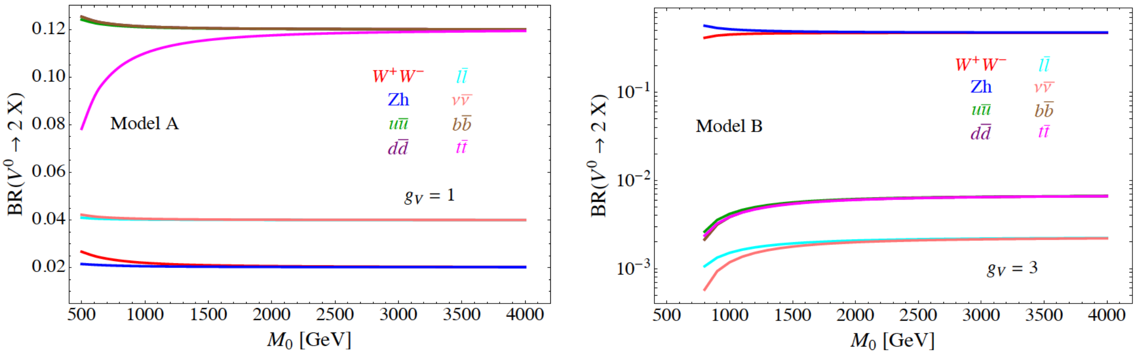


Figure 3.4: Branching fractions for the different decay channels of the neutral spin-1 resonance Z′ (V'^0) for the benchmarks A ($g_V = 1$) (left) and B ($g_V = 3$) (right), as a function of the resonance mass [48].

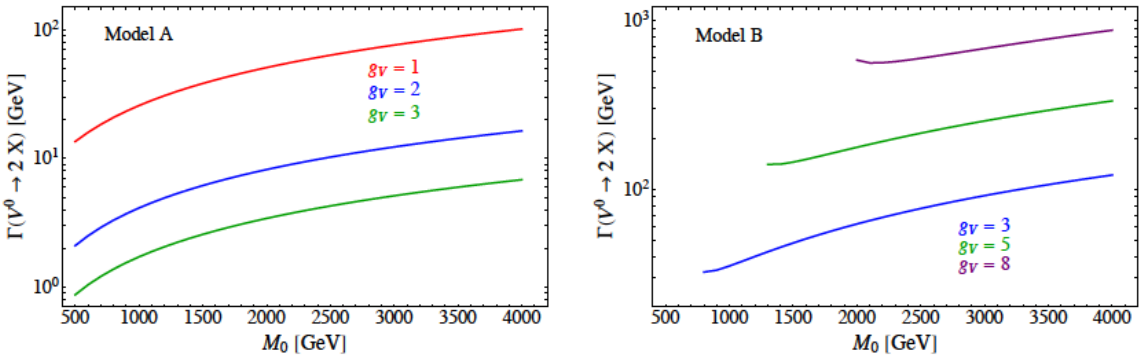


Figure 3.5: Width of the neutral Z′ (V'^0) resonance as a function of the resonance mass for different g_V in model A (left) and model B (right) [48].

of the distribution cannot be easily translated into bounds on the phenomenological parameter space. Since a simplified model is not a complete theory, its validity is restricted to the quantities related to the on-shell production and decay mechanisms of the new resonances, which is how most of the LHC BSM searches are performed.

3.3 Searches at the LHC

Diboson resonances can be studied in many decay channels and have a rich phenomenology at the LHC. Depending on the theoretical models the intermediate boson can have a low or a high transverse momentum. Therefore, searches at the LHC need to explore several decay channels and to make use of complementary reconstruction and various analysis techniques to be sensitive to this large variety of signals. Of primary importance in every search for new resonances is the reconstruction of a variable that could discriminate between signal and background events. Usually the invariant mass of the final decay products or the transverse mass in the case of a partial decay to invisible particles such as neutrinos are used, because they would manifest a signal as an excess or bump over a smoothly falling background spectrum. Another important aspect is the criteria of the event selection: usually the cross sections of these BSM processes are very small, so channels with large branching ratios are preferred, which usually coincide with a hadronic final state for the diboson searches because of the large branching ratios of the W, Z, H bosons, which are about 67%, 70%, and 58%, respectively. However these final states are also very populated by standard model background processes with large cross section, such as the overwhelming multijet QCD production. For this kind of background, usually data-driven background estimations are used, meaning that the prediction is done based on data itself using one or more control regions that are enriched in such processes, because simulations are not found to be as reliable in such particularly boosted phase-space. Final states with leptons offer a compromise between branching fractions and reduced background contamination, which is due mainly to electroweak processes such as top quark pair production and the production of a vector boson (W or Z) with additional jets.

After the Higgs boson was discovered during Run 1 of the LHC, it became possible to use the new boson itself to probe for the existence of physics beyond the standard model. The experimental challenges are very different depending on the final state adopted. The exploitation of the H decay to b-quark pairs relies on the capability to distinguish the jets originating from b quark from jets originating from light quarks and gluons, which can be misidentified as b-quark jets due to instrumental effects.

Another frequent Higgs boson decay is to τ lepton pairs, which happens about 6.3% of the cases. Since τ leptons are unstable and they decay to hadrons and leptons in association with neutrinos, multiple final states are produced. In the fully hadronic channel, special criteria need to be applied to ensure that the genuine hadronic tau leptons decays are discerned from possible misidentified gluon- and quark-originated jets. Moreover, the presence of neutrinos prevents a complete kinematic reconstruction of the event. With the background mainly being from irreducible electroweak processes, the $\tau\tau$ final states profit from a lower background contamination than in the $b\bar{b}$ case. As a balance between background contamination and signal efficiency, the focus of this doctoral work are diboson final states with a Higgs boson decaying to tau leptons and another SM boson decaying to quark pairs.

At the time this work was started, searches for Higgs boson pair production in pp collisions had been performed during Run 1 of the LHC by both the ATLAS and

CMS Collaborations. A combination of resonant production searches of the ATLAS Collaboration is presented in Fig. 3.6 [49].

Apart from a modest excess of events corresponding to 2.4 standard deviations from the background-only hypothesis, localized around 300 GeV in the resonance mass spectrum in the ATLAS search for HH $\to$ $b\bar{b}\gamma\gamma$, no significant deviation from the standard model expectation was found. A combination of the results in the $b\bar{b}\tau^+\tau^-$, $\gamma\gamma WW^*$, $\gamma\gamma b\bar{b}$, and $b\bar{b}b\bar{b}$ channels was performed, and upper limits on the resonant and non resonant production of Higgs boson pairs was set at 95% confidence level.

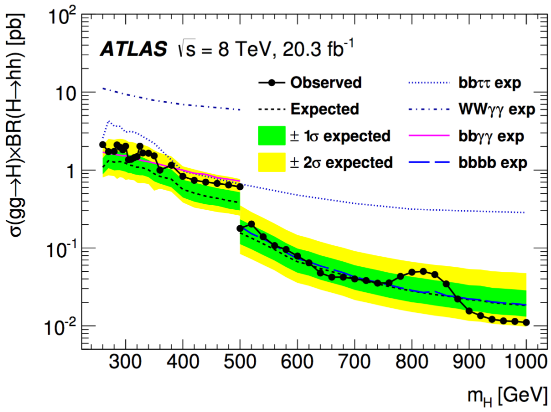


Figure 3.6: Upper limits on the resonant production of Higgs bosons from searches in the $b\bar{b}\tau\tau$, $\gamma\gamma WW^*$, $b\bar{b}\gamma\gamma$, $b\bar{b}b\bar{b}$ with the data recorded by ATLAS during Run 1 [49].

This excess was not confirmed with data acquired at the beginning of Run 2 of the LHC in 2015 [50].

The CMS Collaboration also explored different channels: $b\bar{b}\tau\tau$, $b\bar{b}\gamma\gamma$ and $b\bar{b}b\bar{b}$ and, at the end of Run 1, performed a combination of the searches for resonant production 3.7 [51]. These searches found that data was in very good agreement with the standard

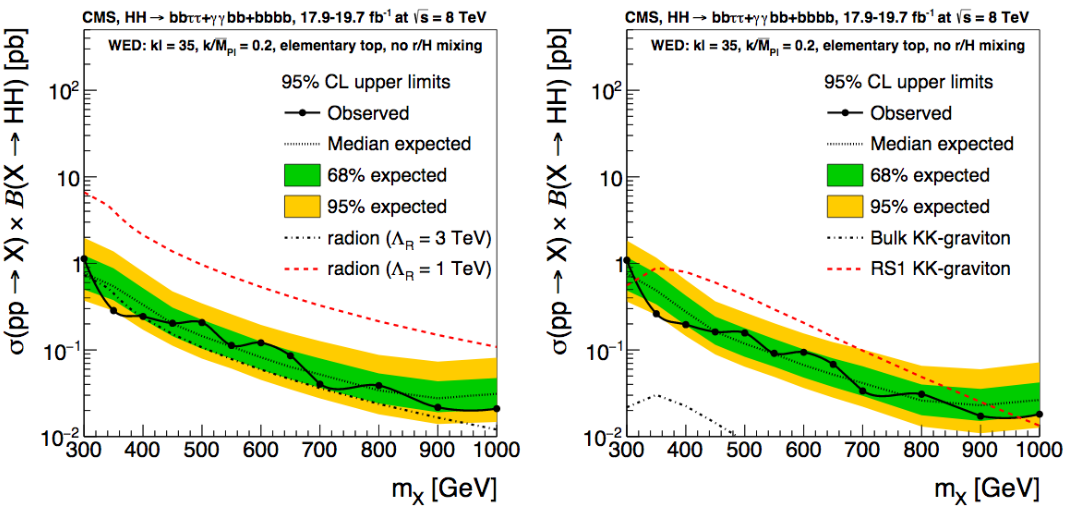


Figure 3.7: Upper limits on the resonant production of Higgs bosons from searches in the $b\bar{b}\tau\tau$, $b\bar{b}\gamma\gamma$, $b\bar{b}b\bar{b}$ with the data recorded by CMS during Run 1 [51].

model expectations and, thus, proceeded to set limits on the production cross section of new resonances in the spins of 0 and 2, compatible with a radion or a graviton. The phase space investigated coincides with low and intermediate resonance masses. In this case, the final decay products of the resonance are well separated and the standard algorithms for the reconstruction can be used.

Towards the end of Run 1, also in preparation for the following Run 2 of the LHC, with increased center of mass energy, a series of techniques were developed to target boosted-object reconstruction and identification. Previously precluded phase space, such as with resonance masses above 1 TeV, became accessible with these novel techniques. For resonant production of two Higgs bosons, the $\mathrm{b\bar{b}b\bar{b}}$ final state was investigated by the ATLAS and CMS Collaborations, by deploying special b tagging techniques to identify energetic large-cone jets originating from Lorentz-boosted b-quark pairs. These searches extended the HH resonances masses spectrum up to 3 TeV [52, 53]. The focus of this work was to extend the phase space analyzed to higher resonance masses into the TeV scale for the $\tau^+\tau^-\mathrm{b\bar{b}}$ final state. Thus the final state considered was consistent with a high-momentum Higgs boson decaying to tau leptons and another boson, either a W, a Z, or a Higgs, decaying to quark pairs. Therefore, a special reconstruction for the boosted τ lepton pairs was developed, as described in Sec. 5.3.7.5, and used in the search described in Chapter 6.

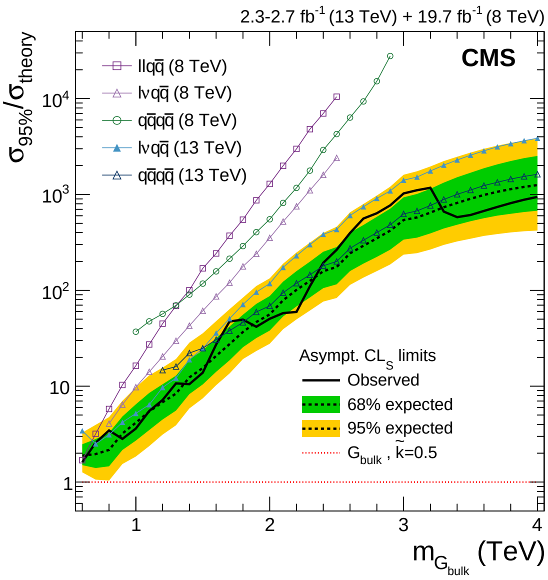


Figure 3.8: Exclusion limits at 95% CL on the signal strength in the bulk graviton model with $\tilde{k} = 0.5$, as a function of the resonance mass, obtained by combining the 8 and 13 TeV diboson searches. The signal strength is defined as the ratio of the excluded cross section to the theoretical prediction. The curves with symbols refer to the different inputs to the combination. The thick solid (dashed) line represents the combined observed (expected) limits. [51].

Both CMS and ATLAS searched for diboson resonances in a variety of final states. The CMS collaboration combined the various results using 19.7 fb^{-1} of luminosity recorded during Run 1 of the LHC and 2.7 fb^{-1} collected in 2015 from Run 2 of the LHC [54]. The signal hypotheses considered were a spin-2 graviton in the bulk scenario (Fig. 3.8), and spin-1 W′, Z′, or generic V′ bosons in the HVT benchmark models A and B. The latter is shown in Fig.3.9, where the results are also interpreted to set constraints on the HVT phenomenological coupling of the new resonance with the SM boson and fermion fields, $g_V c_H$ and $g^2 c_F/g_V$. Even with a luminosity ten times smaller, the searches conducted with 2015 data have a comparable sensitivity with the searches from Run 1 of the LHC, due to the increase of center of mass energy to 13 TeV.

The second part of this work, then, is focused on the 35.9 fb^{-1} of data recorded by the CMS experiment during 2016, that with about 15 times the integrated luminosity of 2015, is expected to further increase the sensitivity and reach of the search. A larger variety of final states are analyzed with respect to Run 1. In the second analysis presented, d-τ final states where both tau leptons decay hadronically ($\tau_\mathrm{h}\tau_\mathrm{h}$) are considered, in addition to the states with $\ell\tau_\mathrm{h}$ (where $\ell = e, \mu$) that were used in the first analysis presented. Also, in the second analysis presented, bosons are permitted to decay also to light-quark jets, in addition to the b-quark jets considered in the first analysis. This allow the search to be sensitive to WH and ZH final states predicted by the HVT model.

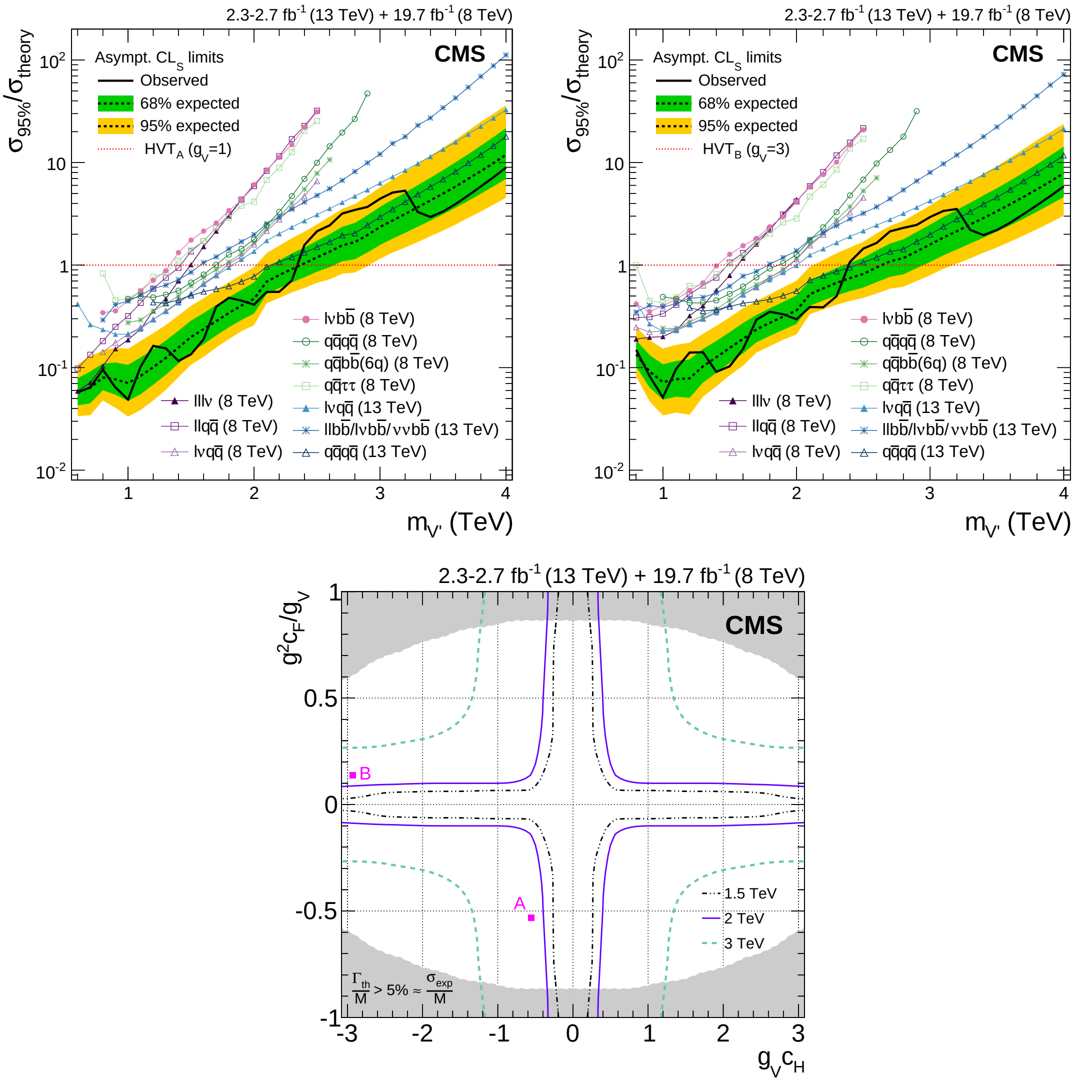


Figure 3.9: Exclusion limits at 95% CL for HVT models A (upper left) and B (upper right) on the signal strengths for the mass degenerate triplet V′ as a function of the resonance mass, obtained by combining the 8 and 13 TeV analyses. The signal strength is defined as the ratio between the excluded cross section and the theoretical prediction. The curves with symbols refer to the different final states used in the combinations. The thick solid (dashed) line represents the combined observed (expected) limits. In the lower plot, exclusion regions are shown in the plane of the HVT-model couplings ($g_V c_H$, $g^2 c_F/g_V$) for three resonance masses of 1.5, 2.0, and 3.0 TeV, where g denotes the weak gauge coupling. The points A and B of the benchmark models used in the analysis are also shown. The boundaries of the regions excluded in this search are indicated by the solid, dashed, and dashed-dotted lines. The areas indicated by the solid shading correspond to regions where the resonance width is predicted to be more than 5% of the resonance mass, in which the narrow-resonance assumption is not satisfied [51].

Chapter 4

The LHC and the CMS experiment

The data that is analyzed in this thesis was collected by the Compact Muon Solenoid (CMS) experiment, located at the Large Hadron Collider (LHC) at the Swiss-French border near Geneva. The LHC is an accelerator designed to collide protons and ions at a center-of-mass energy of 14 TeV, to test the SM, look for the Higgs boson, and search for new physics. In this chapter the LHC and CMS will be introduced, together with a description of the experimental data acquisition system.

4.1 The LHC

The LHC [55] is a circular collider designed to collide protons at beam energy of 7 TeV, thus at a center-of-mass energy of 14 TeV. Additionally, the LHC collides heavy ions (Pb^{82+}) at an energy of 574 TeV per nucleus. Data taking started in 2010 with a center-of-mass energy of 7 TeV, and was increased to 8 TeV in 2012. In 2015, the machine reached a center-of-mass energy of 13 TeV and the same energy was kept throughout 2016, for the so-called Run 2 data taking period. While the data collected before (after) the long shut down in 2013 are generally referred to as Run 1 (Run 2), in this thesis, Run 1 data refers to the subset of data collected in 2012, and Run 2 refers to the subset of data collected in 2016.

The LHC is located at the European Organization for Nuclear Research (CERN) laboratories, and is composed of an accelerator facility and a storage ring located between 45 and 170 meters underground in a 27-km long tunnel that previously hosed the Large Electron Positron (LEP) accelerator.

Before being stored and accelerated in the LHC, particles are produced and go through a series of pre-accelerating stages, as shown in Fig.4.1. Electrons are stripped from

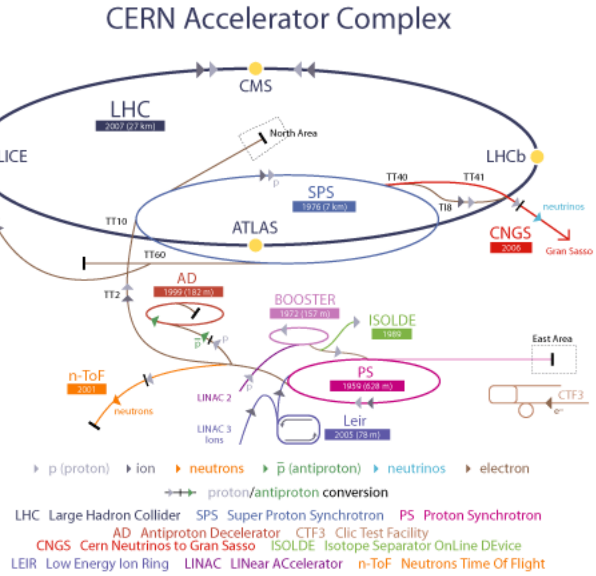


Figure 4.1: Schematic overview of the injection chain, the LHC ring and of the experiments at the interaction points [60].

hydrogen atoms and the resulting protons are accelerated up to 50 MeV by the LINAC 2, before being injected into the Proton Synchroton Booster (PSB), where they reach an energy of 1.4 GeV. Afterwards, they are accelerated up to an energy of 26 GeV in the Proton Synchrotron and then up to 450 GeV in the Super Proton Synchrotron, the last step before being injected in the LHC. The LHC consists of 1232 superconducting dipole magnets that bend two opposite beams of protons with a magnetic field of 8.3 T, operating at a temperature of 1.9 K. The acceleration is achieved with a series of high-frequency (HF) cavities with an oscillation frequency of 400 MHz. In order to be accelerated, the protons are required to be synchronized with it and, thus, are grouped in so-called bunches with a designed inter-bunch distance of 25 ns.

Quadrupole, sextuple, and octapole magnets focus the particle beams to increase the interaction probability in the four collision points. The LHC is designed to have four interaction points where the two beams, made up of 2808 bunches, each consisting of 10^{11} protons, collide with a design instantaneous luminosity of $10^{34}\text{cm}^{-2}\text{s}^{-1}$.

Four main experiments are installed in the interaction points of the particle beams: the A Large Ion Collider Experiment (ALICE) [56], with a detector especially built to analyze interactions between heavy nuclei; the Large Hadron Collider beauty (LHCb) [57], a b-physics experiment with a one-sided detector in the forward direction; and the two multipurpose experiments, A Toroidal LHC ApparatuS (ATLAS) [58] and the Compact Muon Solenoid (CMS) [59].

4.1.1 LHC operations during CMS data taking

The rate of events for a particular physics process depends on many parameters, the most important being its cross section, at a given center-of-mass energy, and the instantaneous luminosity, which is proportional to the number of interacting particles in an interval of time. Specifically, the rate dN_{events}/dt of a process is:

$$\frac{dN_{events}}{dt} = \mathcal{L} \cdot \sigma$$

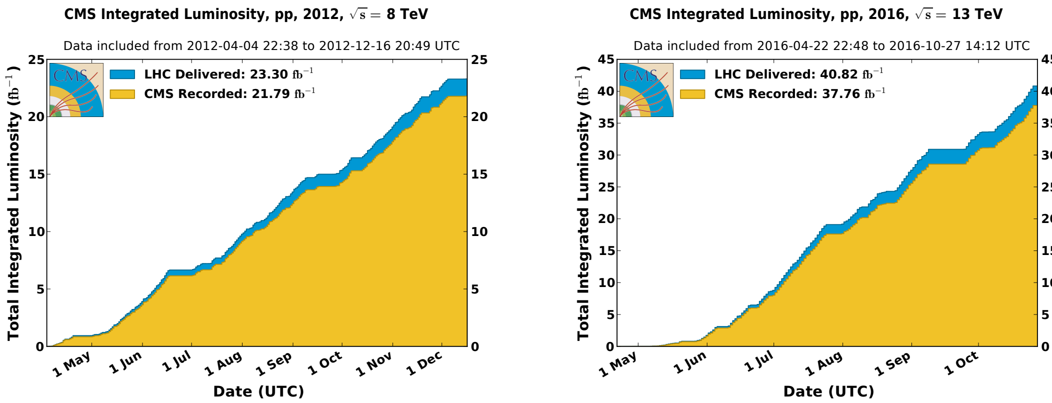
where σ is the cross section of the physic process, which depends on the center-of-mass energy, and the instantaneous luminosity $\mathcal{L}$, which depends on several LHC parameters, such as the number of bunches, the number of protons in each bunch, and the sizes of the beam profiles at the interaction point. The integrated luminosity, defined as $L = \int \mathcal{L}\ dt$, measures the amount of data delivered by the LHC.

From 2010 to 2011, the LHC collided protons at a center-of-mass energy of 7 TeV. In 2012 and 2015, the center-of-mass energy reached 8 TeV and 13 TeV, respectively. In the 2012 and 2015 periods, the collected datasets correspond to integrated luminosities of 19.7 fb^{-1} and 2.2 fb^{-1}, respectively. The collected data is certified to be good for physics analysis by checking that all sub-detectors are operational and that the reconstruction efficiency of physics objects achieves the expected performance. In 2016, 35.9 fb^{-1} of luminosity was certified as good.

The data analyzed in this thesis corresponds to the complete dataset delivered by LHC and acquired by the CMS experiment during 2012 at $\sqrt{s} = 8$ TeV and 2016 at $\sqrt{s} = 13$ TeV, as shown in Fig. 4.2.

4.2 CMS detector

CMS is a multipurpose detector built to identify and measure various types of particles in order to study known SM processes and new physics extensions. It consists of various subsystems with different purposes and characteristics. The name of the experiment pays tribute to central features of the detector, namely a superconducting 3.8 T solenoid magnet used to bend the trajectories of charged particles emerging from the collisions, and a powerful system for reconstructing muons. A schematic view of the onion-like CMS apparatus is shown in Fig. 4.3: the several subdetector enclose each other in order to provide hermetic spatial coverage around the interaction point. Therefore, the detector and its subsystems are divided into a cylindrical central part, referred to as a



(a) (b)

Figure 4.2: Integrated luminosity delivered by the LHC (blue curve) and recorded by the CMS experiment (yellow curve) in 2012 (a) and 2016 (b) during stable beams and for *pp* collisions at 8 TeV and 13 TeV centrr-of-mass energy, respectively. The luminosity is determined from counting rates as measured by the luminosity detectors after offline validation [61, 62].

barrel, and two forward disc-like parts, or *endcaps*.

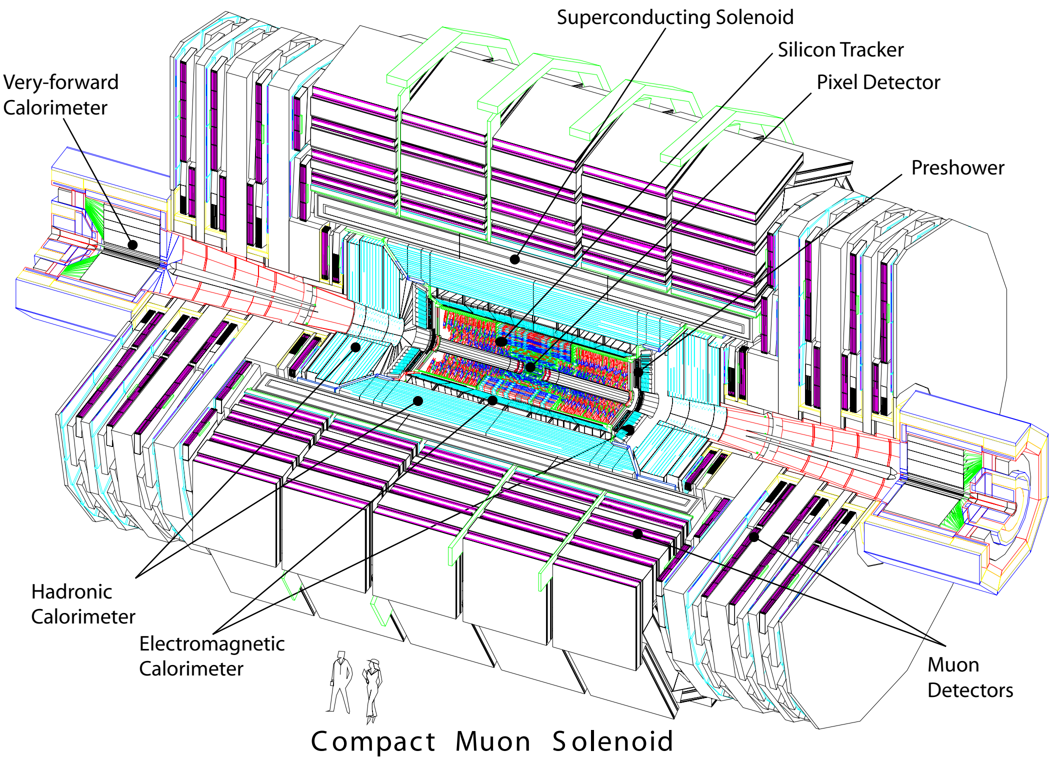


Figure 4.3: A view of the CMS apparatus with a cut out cross section and components annotated, as taken from [63]. From the inside out: the inner tracking system with the pixel and the silicon strip detector, the electromagnetic and hadronic calorimeters, and the muon system embedded in the iron return yoke of the solenoid.

In order to describe the position and kinematic properties of particles within the detector, a coordinate system is defined as follows. The origin of the coordinates is identified by the nominal collision point at the center of the detector. The x-axis is taken to be horizontal and oriented towards the center of the LHC ring, while the y-axis points vertically upwards. The z-axis is oriented anti-clockwise along the beam direction. The xy-plane is called the transverse plane and is perpendicular to the beam direction. The azimuthal angle ϕ is measured in the xy-plane with respect to the x-axis. The polar angle θ is defined as the angle formed with respect to the z-axis, and is used to define the pseudorapidity variable $\eta = -\ln\tan(\theta/2)$. With these quantities it is possible to define a Lorentz invariant spatial angle

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}.$$

In the following, the most important CMS subsystems will be introduced and described, starting from the innermost part of the detector. The three main subdetectors are a tracking system embedded in a magnetic field, an electromagnetic and a hadronic calorimeter, and a muon system, as shown in Fig. 4.4.

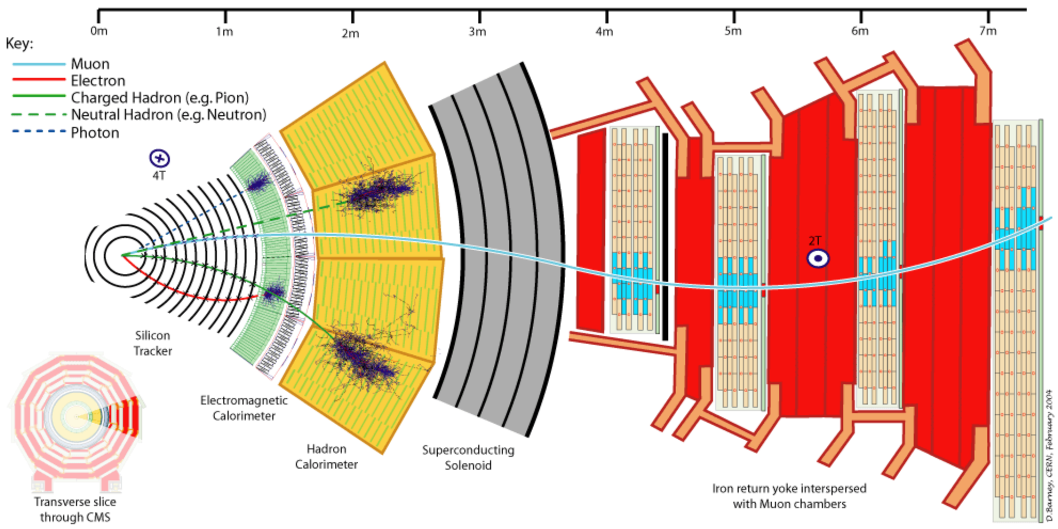


Figure 4.4: Schematic view of a transverse slice of the CMS detector. Also shown are how the different long-lived particles interact with each subdetector. This information is exploited to distinguish the different types of particles [64].

Given the high LHC bunch crossing rate, an online trigger system is required to process and store a fraction of events interesting for physics analysis. This will be described at the end of the chapter. Further details about the CMS detector and data acquisition system can be found in [65–67].

4.2.1 Magnet

The CMS detector is embedded in a 3.8 Tesla magnetic field parallel to the beam pipe generated by a 13 m long superconducting solenoid with an inner bore of 6 m [66]. The inner diameter is large enough to accommodate the tracking system, the electromangnetic and hadronic calorimeters. The return field of the solenoid is large enough to saturate the 1.5 m of iron in the outer muon systems. A large magnetic field is crucial to measure with high precision the transverse momenta of cherged particles due to the curvature of their trajectories. Charged particles subject to a magnetic fields move in helical trajectories. The deflection angle θ in the plane transverse to the beampipe is approximated by $\theta = \rho/L$, where ρ is the bending radius and L is the path length inside the solenoid [68]. From the radius of curvature of a particle of charge qe, the component of the momentum in the plane transverse to the beampipe (p_{T}) is obtained as [4]: $p_{\mathrm{T}}[\mathrm{GeV}] = 0.3\ q\ B[\mathrm{T}]\ \rho[\mathrm{m}]$. The associated relative uncertainty on the momentum $\sigma(p_{\mathrm{T}})/p_{\mathrm{T}}$ depends on the number N of measurement points or hits of the particle with the tracker system:

$$\frac{\sigma(p_{\mathrm{T}})}{p_{\mathrm{T}}} = \frac{\sigma(x)\ p_{\mathrm{T}}}{0.3BL^2}\sqrt{\frac{720}{N+4}}$$

where $\sigma(x)$ is the spatial hit resolution [68]. The CMS magnetic field is designed to provide a momentum resolution for charged particles of typically 1% (5%) for $p_{\mathrm{T}} = 100$ GeV ($p_{\mathrm{T}} = 1$ TeV) [69], with unambiguous charge identification for muons up to a momentum of 1 TeV.

4.2.2 Inner tracking detectors

The CMS inner tracking system [70–72], as previously mentioned, allows for charged particles to be recognized and their transverse momenta measured, due to their curvature in the solenoidal magnetic field. The system can also be used to reconstruct the primary and secondary vertices in the interaction. Fine granularity and fast readout are required given that a flux of thousands of charged particles go though the detector at each bunch crossing. Since the detector material can cause multiple scattering of the particles and a degradation of the spatial resolution, the tracker detector has to be lightweight in its so called *material budget*, which is quantified to be between 0.4 and 1.8 radiation lengths (X_0), depending on the η region. The tracker features silicon pixels and microstrip detectors. The tracker system has a total length of 5.8 m and a

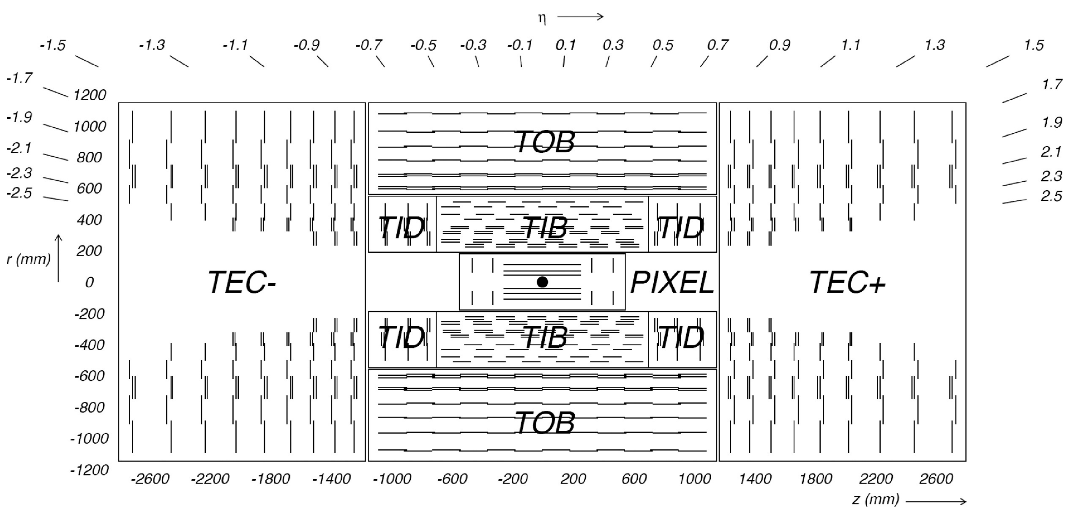


Figure 4.5: Schematic overview of the CMS tracker detector in the y-z plane [65].

diameter of 2.6 m, and it covers a pseudorapidity region up to $|\eta| < 2.5$. In Figure 4.5 a schematic view of the tracker system is shown.

4.2.2.1 Pixel detector

Close to the interaction point (IP), where the flux is the highest, sits the pixel detector, which constitutes the innermost system of the tracker. There are 66 million, rectangular, silicon pixels of size $100 \times 150\ \mu\mathrm{m}^2$, which are grouped in modules with an area of $2 \times 8\ \mathrm{cm}^2$, divided for the barrel between three layers, positioned at radii $r = 4.4$, 7.3, and 10.2 cm, and for the endcaps in two disks, positioned at $z = \pm 34.5$ and ± 46.5 cm. Pixel sensors and front-end electronics are arranged on top of each other and are kept at a stable temperature with a liquid mono-phase C_6F_{14} cooling system. Reading out analog pulse-height information associated with each pixel hit and exploiting charge sharing between adjacent pixels, a spatial resolution of about $10\,\mu$m in the $r - \phi$ plane and of about $25\,\mu$m in the z-axis is achieved. The spatial resolution of a given detector layer is measured with the so-called *triplet method*. For every track a *residual* is defined as the difference between the reconstructed *hit*, i.e. the position of the pixel cluster, on the layer and the hit extrapolated by computing a new track where all the detector information is used except the hit on the layer that the resolution is measured for. From the width of residual distributions for the different layers, e.g. reported in Fig. 4.6 for the intermediate pixel layer in 2016, it is then possible to compute the expected intrinsic hit resolution of the detector. This configuration allows to precisely reconstruct primary and secondary vertices ,which enables the identification of B hadrons and τ leptons that have relatively long lifetimes, but typically decay in the beam pipe.

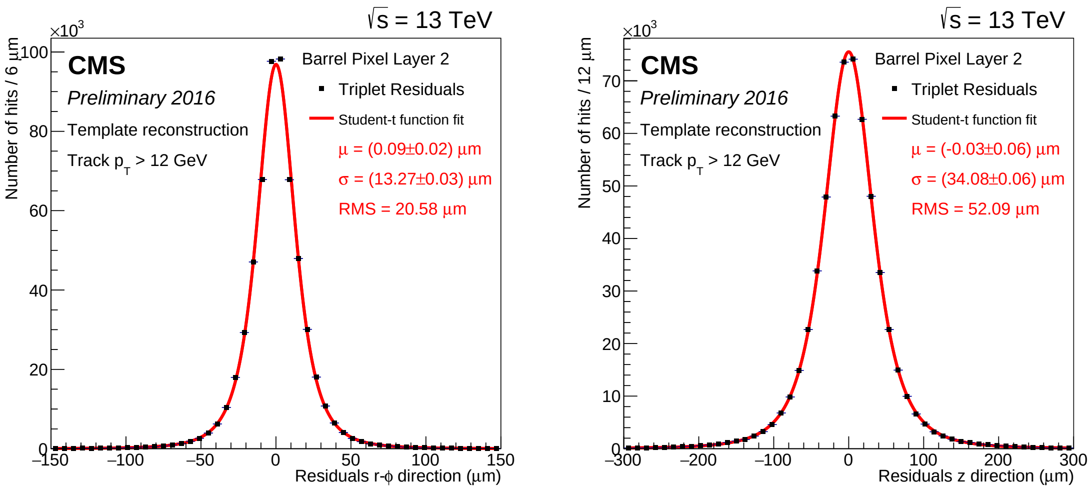


Figure 4.6: Residual distributions of the second layer of the pixel detector as measured in Run 284043 in 2016, along the r-ϕ direction (left) and the z direction(right)

Part of the work of this thesis was to measure and monitor the intrinsic resolution during the various data acquisition periods, as shown in Fig.4.7, in order to ensure the correct operation of the detector.

4.2.2.2 Strip detector

While the silicon pixel detector uses sensors that are segmented in two dimensions, the sensors at large radii are implemented as one dimensional strip. Two dimensional spacial resolution can be achieved by arranging in different layers sensors with rotated strip orientation on top of each other. Silicon microstrips are grouped in three larger subsystems: Tracker Inner Barrel and Disks (TIB/TID), Tracker Outer Barrel (TOB), and Tracker End Cap (TEC). Ranging from 80 μm to in the first layer of the TIB to 184 μm in the TEC, the strip pitch decreases at increasing radii. Silicon microstrips have a resolution between 22 μm and 55 μm in the radial direction depending on the part of the detector. While along the other coordinate, due to the strip modules mounted at a stereo angle of 100 mrad, the resolution varies from 230 μm to 530 μm.

The silicon strip detector provides nine additional measurement points, extending the lever arm for p_{T} measurements to a radius of 1.1 m. As further explained in Sec. 7.3. the tracker momentum resolution improves greatly the capability of the overall momentum reconstruction also for objects like muons, that are usually identified by matching segments in the outer muon detectors.

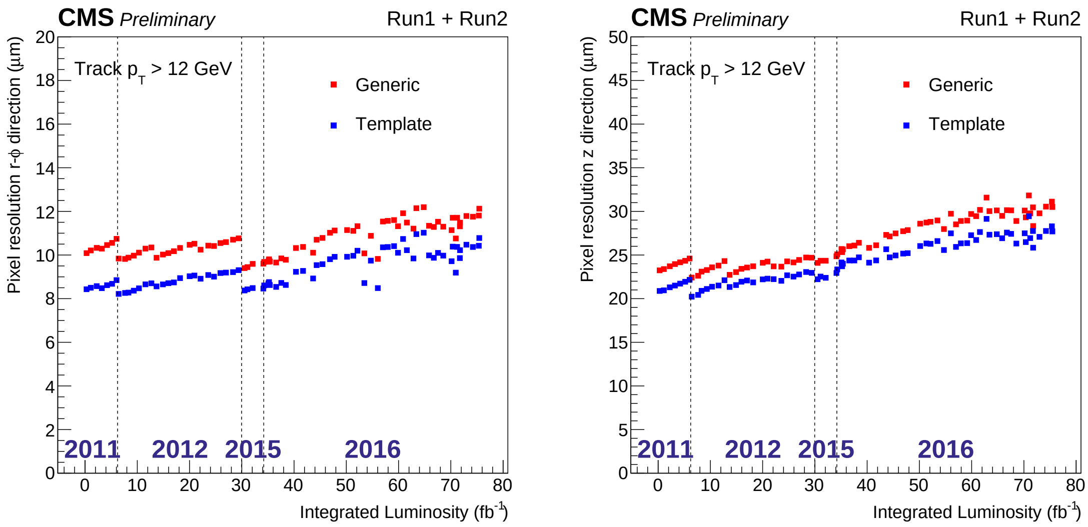


Figure 4.7: Residual distributions of the second layer of the pixel detector, along the r-ϕ (left) and the z direction (right). Two track reconstruction algorithms are considered: the generic, used predominantly at the trigger level, is in red and the template, used in the offline reconstruction, is in blue.

4.2.3 Calorimeter

The calorimeter system measures the energies of particles through their interactions with matter. It also allows the energy of the neutral particles, which don't leave a signal in the tracking system, to be measured. To reduce energy losses due to interactions with passive detector material, both the electromagnetic (ECAL) and hadronic calorimeter (HCAL) are situated inside the magnet. In the electromagnetic calorimeter, electrons, positrons, and photons are absorbed and their energy is measured. The hadronic sampling calorimeter is composed of layers of brass absorber and plastic scintillator tiles and measures the energy deposited by hadron-induced showers. At large pseudorapidities the very-forward hadronic calorimeter complements the system, making CMS an almost hermetic detector, an essential requirement for reconstructing missing transverse momentum arising from particles that escape detection, e.g. neutrinos.

The information from the calorimeter system is also exploited by the trigger system to identify events interesting for physics analyses.

Electromagnetic calorimeter

The electromagnetic calorimeter [73] is a homogeneous calorimeter made of $PbWO_4$ crystals. The choice of this material is motivated by the short radiation length of 0.89 cm and a small Moliere radius of 2.2 cm, which allows for the construction of a

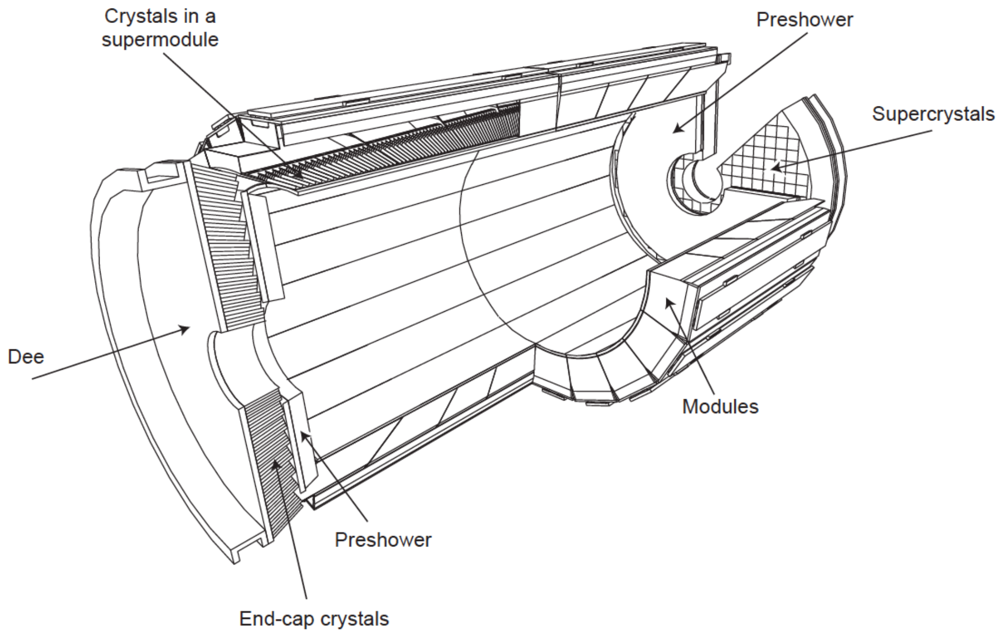


Figure 4.8: Layout of the CMS ECAL showing the arrangement of crystal modules, supermodules and endcaps, with the preshower in front. [65].

very compact detector with a total length of 25.8 X_0. The ECAL is divided into a barrel (EB) and two endcap (EE) regions. Figure 4.8 presents a schematic view of the electromagnetic calorimeter configuration.

The barrel part covers a pseudorapidity interval up to $|\eta| < 1.479$, while both endcaps cover a range of $1.479 < |\eta| < 3.0$. The crystals used in the barrel and endcaps have a cross-sectional area of 22×22 mm^2 and 28.6×28.6 mm^2, respectively [73].

Between the tracking system and the EE, in the pseudorapidity region of $1.653 < |\eta| < 2.6$, a preshower detector is installed, made of two lead disk absorbers and two silicon sensors planes. Because of its finer granularity (~ 2 mm pitch silicon sensors) [73], it permits closely-spaced photon showers from π^0 decay to be distinguished from single photons. At high energies (above 500 GeV), the resolution of the calorimeter is degraded because of the shower leakage outside the calorimeter, while for lower energies, the energy resolution of the ECAL can be described as

$$\frac{\sigma(E)}{E} = \frac{a}{\sqrt{E}} \oplus b \oplus \frac{c}{E}.$$

The parameters were determined in test beam measurements [74] as $a = 2.8\%$ GeV$^{1/2}$ for the stochastic term, $b = 0.30\%$ for the constant term, and $c = 0.127$ GeV for the

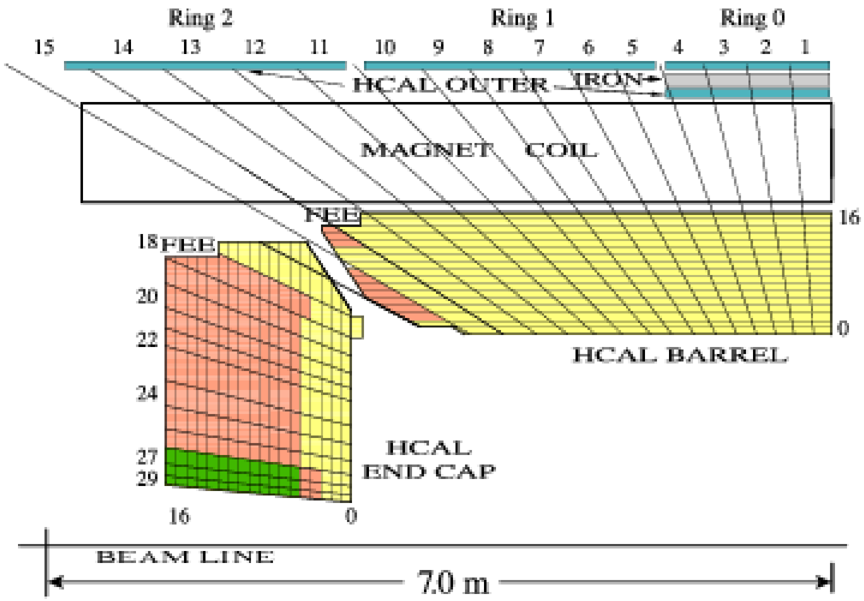


Figure 4.9: The HCAL tower segmentation in the rz-plane for one-fourth of the HB, HO, and HE detectors [65].

noise term. The ECAL resolution is in the range from 0.4% to 1.5% for energies in the interval from 10 to 250 GeV [75].

Hadronic calorimeter

Because the ECAL represents only about 1.1 nuclear interaction lengths, hadrons predominantly reach the hadronic calorimeter, where they deposit the bulk of their energy after causing hadronic showers in the brass absorbers. The energy released is measured in plastic scintillators that are interleaved with the absorbing plates and read out by wavelength shifting fibers and photon detectors. The HCAL is divided into the barrel, made of an inner (HB) and an outer (HO) part, the endcap (HE), and the forward (HF) detectors. A schematic view of the HCAL configuration is shown in Fig. 4.9.

The barrel extends to $|\eta| < 1.3$, while each endcap covers the range $1.3 < |\eta| < 3.0$, whereas the HF continues to $|\eta| < 5.2$ to maximize the solid angle coverage, in order to reconstruct the missing transverse momentum of particles that do not interact with the detector, e.g. neutrinos.

Due to the available space between the outer extent of the electromagnetic calorimeter ($R = 1.77$ m) and the inner part of the magnet coil ($R = 2.95$ m), the HB thickness is limited to 5.8 (10) hadronic interaction lengths at $|\eta| = 0$ (1.2). The remaining part of the energy is measured with layers of scintillators of the HO, which is situated outside the magnet and exploits the solenoid as additional absorbing material. In this way, the depth of the calorimeter reached is in total equivalent to at least 11.8 interaction lengths.

Given the high fluence of hadrons in the very forward region, the HF calorimeter requires the use of radiation-hard materials. In fact, while in the rest of the detector an energy of about 100 GeV is expected to be released per bunch crossing on average at $\sqrt{s} = 14$ TeV, this amounts to 760 GeV in the very forward region [65]. For this reason, steel is used as an absorber and radiation-tolerant quartz fibers are inserted as an active medium. The HF is also used to monitor the instantaneous luminosity delivered to CMS [76].

The combined measurements of the ECAL and HCAL have a resolution σ_E that can be parametrized as a function of the energy with [77]:

$$\frac{\sigma_E}{E} = \frac{0.847\ \mathrm{GeV}^{1/2}}{\sqrt{E}} \oplus 0.074,$$

which holds for energies E in the range 30 GeV $-$ 1 TeV. The energy resolution for the endcap has a similar behavior as a function of the energy, but with different parameters: 1.98 and 0.09, instead of 0.847 and 0.074. For typical jet-energy thresholds on the order of 40 GeV, the energy resolution is on the order of $10\% - 20\%$, decreasing for higher energies. However, the resolution can be improved by combining the HCAL measurements with data from other subdetectors using a so-called particle flow (PF) algorithm, which will be summarized in Chapter 5 [78].

4.2.4 Muon system

The muon system [65,79] is of vital importance in detecting muons and measuring their momenta. Muons are relative long-lived particles with a lifetime of 2.2 μs, have quite a large mass compared to electrons, and do not interact strongly, so they can travel distances much longer that the dimensions of the detector, interacting minimally with the ECAL, HCAL and the solenoid. They are the only measurable particles able to reach the muon system that is the outermost subdetector, being placed outside the solenoid. The muon systems are divided into a barrel region and two endcap parts, as shown in Figure 4.10. The barrel region covers a pseudorapidity range up to $|\eta| < 1.2$ and features four layers of drift tube (DT) chambers embedded in the rings of the magnetic field return yoke. The choice of DT as gaseous particle detectors is due to the low muon rate, the small neutron-induced background, and the homogeneity of the

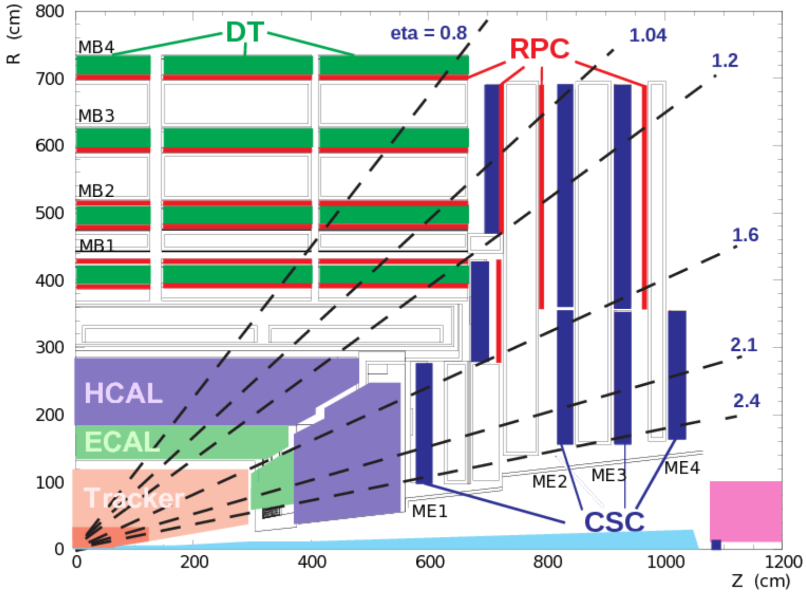


Figure 4.10: Schematic view of the CMS muon system [80].

magnetic field in this region. The magnetic return flux in the iron plates allows for the possibility of an independent momentum measurement with respect to the one in the tracker system. A different orientation of the sensing wires in the muon chambers allows for measurements of the $r - \phi$ and z coordinates.

Both endcap regions cover the pseudorapidity range $0.9 < |\eta| < 2.4$. In this part, the particle flux is higher and the magnetic field is large and non-uniform. Therefore, a faster response time, finer segmentation and higher radiation resistance is required. For these reasons cathode strip chambers (CSC) are used as detectors.

Resistive plate chambers (RPCs) are placed both in the barrel and endcap regions and cover a pseudorapidity interval $|\eta| < 1.6$. RPCs are operated in avalanche mode and have coarser position resolution than the DTs or CSCs, but they have an excellent timing resolution, useful for fast-trigger response at high rates. The RPC can also be used to resolve ambiguities when building a track from multiple hits in a chamber.

The resolution achieved for single point measurements is about $80 - 120\,\mu$m for drift tubes and $40 - 150\,\mu$m for cathode strip chambers, as measured in pp collisions at 7 TeV [81]. The efficiency of detecting a muon is above 95% for both detectors. The timing resolution reachable with the RPC is 3 ns, much lower than the bunch spacing of 25 ns of the LHC design. resolution for a muon as a function of its transverse momentum is shown in Figure 5.2. The standalone p_{T} resolution of the muon system is less than 10% for energies up to 100 GeV and degrades to 40% for 1 TeV muons at high η, being mainly limited by multiple scattering in the detector material. This can be improved by about an order of magnitude by combining with the information from the other subdetectors, mainly the tracking system, in the global fit of the PF algorithm.

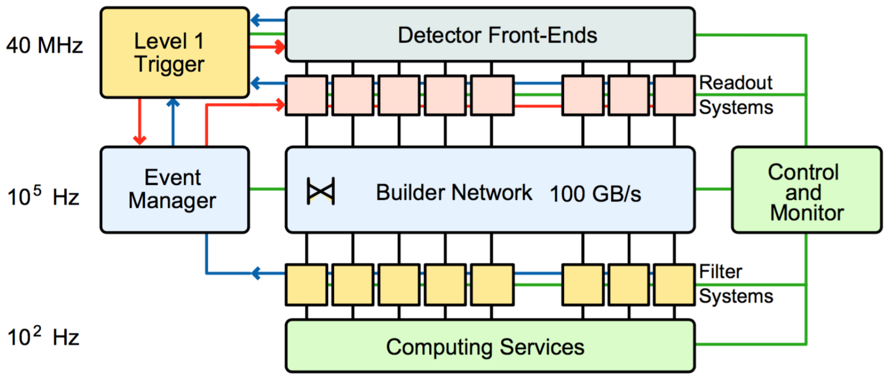
4.2.5 Readout system

The design of the LHC assumes for pp collisions a bunch crossing frequency of 40 MHz and an average of 20 simultaneous pp collisions occurring per bunch crossing. Considering that the combined raw data of all the subdetectors amounts to 1.5 MB/event, this translates to an enormous amount of data to be processed and stored. The current technology used to process and store data allows a frequency of 100 Hz. Moreover, only some events are of interest for the physics analyses, while the vast majority is well understood inelastic and elastic proton scattering. Therefore, it is very important to filter the events online in order to reject most of the processed events and retaining just events compatible with signals of hard processes and rare phenomena of interest for physics analyses.

This is achieved by a two-stage trigger system [82]: a hardware trigger named Level-1 (L1) and a software trigger called High Level Trigger (HLT). The overall rate reduction of the two stages is designed to be a factor of 10^6 or higher [65]. After the data has fulfilled the requirements of the L1 trigger, the Data Acquisition (DAQ) system builds the event data from the data fragments of each subdetector, and transfers it to the HLT trigger, in which computers uses software algorithms to analyze the entire event data in order to make a final decision on whether events are interesting enough to be stored [83]. A diagram showing the DAQ system, including the trigger systems, is shown in Fig. 4.11.

Level-1 Trigger

At every bunch crossing the L1 trigger analyzes coarse information from the calorimeters and the muon system and makes a decision in order to select events as fast as possible. The maximum allowed latency is limited by the buffering capabilities of the front-end electronics of the subdetectors, where the full data are kept while the trigger decision is pending. A decision is computed in less than 3.2 μs. The event rate is reduced to about 100 kHz, which is suitable for further processing by the HLT.



(a)

Figure 4.11: Diagrams of the CMS DAQ architecture [65].

High Level Trigger

The data that satisfy the L1 trigger requirements are transferred to the central processing units that run the HLT software on dedicated computer farms. The HLT has a more complex criteria to reduce the event rate with respect to L1. A fast and sometimes partial reconstruction of the events is done. Different requirements regarding physical objects can be made, organized into trigger paths that establish if an event is discarded or accepted. Each trigger path probes the event for a range of properties that can make it interesting for further examination, e.g., large jet or lepton multiplicities, large energy deposits, objects of high transverse momentum, or large missing transverse momentum. Events can be "prescaled" by a certain factor P, meaning that only one of every P events is saved and written to disk. Changing the prescale scheme during data acquisition allows the HLT output rate to remain approximately constant at about 1 kHz, independent of the instantaneous luminosity provided by the LHC.

Data acquisition and computing

The computing system allows the data collected by the experimental apparatus to be stored, handled, and then analyzed all around the world. During CMS data taking, the output of the trigger system is stored at the Tier-0 computing center located at CERN, which provides also a fast first reconstruction of the events in order to give feedback for the monitoring of the experiment.

Other computing centers (Tier 1, Tier 2, Tier 3) located all over the globe support and store the real and simulated data. Tier-1 centers perform the full reconstruction of the events from the raw format and ensure data availability for the Tier-2 and Tier-3 where the final analysis of the data is carried out. The coordination and interconnection of these sites happens through the Worldwide LHC Computing Grid [84], that ensures a fairly distributed execution of computing tasks and data access to all the different institutes around the world.

Chapter 5

Object reconstruction

The reconstruction and identification of stable particles created during a collision in the CMS experiment is performed by the so-called particle-flow (PF) algorithm [78], [85], which combines the information of all the CMS subdetectors, allowing for the reconstruction of collision products at the particle level in order to achieve an optimal determination of their direction, energy, and type. In this chapter, the track and vertex reconstruction will be presented together with the the particle flow reconstruction, in Section 5.3.

5.1 Tracks

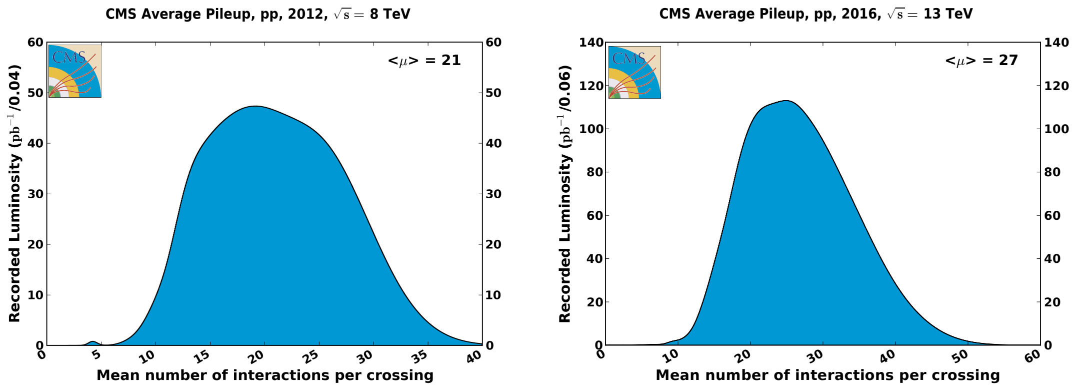
The innermost part of the CMS detector is dedicated to the reconstruction of charged particle trajectories (tracks) from hits measured in the pixel and tracking systems and to provide measurements of the particle momenta and directions. Track reconstruction is performed by the so-called Combinatorial Tracker Finder (CTF) algorithm, which is based on a Kalman filter technique [86–88], through an iterative tracking process [?, 89]. The CTF algorithm consists of four different steps, and after each iteration, hits associated to the high-quality track candidates are removed from the input list. The first step is the *seed generation*, which builds the initial track using triplets of hits if there is a hit in the inner pixel layer or a pair of hits if no inner pixel hit is found with the assumption that the track originates from the interaction region. Recognition begins with trajectory seeds created in the inner region of the tracker, from which the helix parameters are estimated and facilitating the reconstruction of low-momentum tracks. Then the *track finding* step associates to the initial track hits in the next outer layer and updates the track parameters. Once the outer layer is reached, another reconstruction is performed backwards starting from the outermost hit in the detector layers order to improve tracking efficiency and remove from the track spurious hits. Afterwards the *track fitting* is done by re-fitting the trajectory with Kalman Filters and smoothing techniques, in order to improve the accuracy of the measured parameters. The last iteration consists of the *track filter* that maximizes the efficiency for rejecting of fake tracks by applying quality requirements, such as the transverse impact parameter and the number of layers in which hits are found. If two tracks share more than half of their hits, the track with the worst quality is rejected.

The average reconstruction efficiency for tracks, measured measured in simulation and verified in data [90], for promptly-produced charged particles with transverse momenta of $p_{\mathrm{T}} > 0.9$ GeV, is 94% for pseudorapidities of $|\eta| < 0.9$ and 85% for $0.9 < |\eta| < 2.5$. The inefficiency is caused mainly by hadrons that undergo nuclear interactions in the tracker material. For isolated muons, the corresponding efficiencies are > 99%. The typical resolution is around 10μm and 30 μm for the transverse and longitudinal impact parameters, and mostly independent of η.

5.2 Primary vertex and pileup

The vertex reconstruction [91] identifies and measures the location, and the associated uncertainty, of all proton-proton interaction vertices in each LHC bunch crossing (primary vertices), and the ones originating from heavy-flavor and long-lived particles (secondary vertices), using all the reconstructed tracks in an event.

The reconstruction consists of three different steps. An initial *track selection* is done for tracks consistent with being produced promptly in the primary interaction region by checking the significance of the transverse impact parameter relative to the center of the beam spot, the number of hits in the tracker, and the normalized χ^2 associated with each track. Then in the *track clustering* step, the selected tracks are clustered using the deterministic annealing algorithm (DA) [92] using as primary criteria their impact parameter along the z coordinate. Finally, during the *fitting for vertex position*, the adaptive vertex fitter [93] takes into account all the candidate vertices with at least two tracks. For each candidate, a weight w_i close to 1 (0) is assigned to each track i which reflects the likelihood that it genuinely belongs to the vertex. Tracks consistent with the position of the reconstructed vertex have a weight close to 1, whereas tracks that lie more than a few standard deviations from the vertex have smaller weights. The performance of the fit is then evaluated from the number of degrees of freedom



(a) (b)

Figure 5.1: Number of interactions per bunch crossing in *pp* collisions recorded by the CMS experiment in 2012 at $\sqrt{s} = 8$ TeV (left) and in 2016 at $\sqrt{s} = 13$ TeV (right) [94].

n_{dof}, defined as:

$$n_{\mathrm{dof}} = -3 + 2 \sum_{i=1}^{\#tracks} w_i.$$

The value of n_{dof} is therefore strongly correlated with the number of tracks that are compatible with arising from the interaction region.

The primary vertex where the hard scattering originates is chosen as the reconstructed vertex with the highest value of $\sum_i p_{\mathrm{T},i}^2$ where $p_{\mathrm{T},i}$ is the transverse momentum of all tracks for Run 1, and of all the physical objects, including jets and leptons originating from the vertex in the Run 2 algorithm. The resolution depends on the event topology and is typically between 10–40 μm in the transverse plane and 12–50 μm in the z-direction.

The additional pp interactions occurring in the same bunch crossing, or out of time, but with signals overlapping with the considered bunch crossing, are called pileup (PU) interactions. In the 2012 and 2016 data-taking periods, the number of PU interactions in the same bunch crossing was on average 21 and 27, respectively. The distributions of the average number of PU events per bunch crossing in the 2012 and 2016 data-taking periods are shown in Figure 5.1.

It is possible to identify secondary vertices from decays of long-lived particles. In fact their signature in the detector is compatible with vertices displaced with respect to the primary interaction, but consistent with the momentum direction of the tracks associated to the PV.

In this analysis, reconstructed vertices are selected if they are consistent with the expected interaction point. This is ensured by requiring the vertex to be less than 2 (24) cm away in the x, y (z) direction from the interaction point and have $n_{\mathrm{dof}} > 4$.

Due to PU events, particle tracks and energy deposits not associated with the primary interaction are measured in the detector and reconstructed offline. Simulations of the physical processes of interest for this work are generated by simulating the pileup conditions as expected in the 2012 and 2016 data taking. Since, however, the simulated pileup description does not completely replicate the data conditions, corrections are computed to improve the agreement with data, as reported in Sec.7.3.

5.3 Particle-flow event reconstruction

In order to reconstruct the main physical objects linked to the stable particles produced in the collision, the PF technique links together three main elements (PF elements): charged-particle tracks, calorimetric clusters, and tracks from the muon chambers [**?**], [85], [89].

The algorithm is used for both the HLT level and the final offline reconstruction of the events that have been stored on tape, with slight differences. The former kind of reconstruction is a simplified version of the latter with a shorter processing time, where the low-p_{T} tracks which make the reconstruction time-consuming are omitted. The common principle is to combine the information from all subdetectors in order to improve the limited energy resolution of the HCAL by taking advantage of the more precise energy and momentum resolution of the ECAL and the tracking system.

Relating tracks from charged particles to energy deposits in the calorimeter permits the energy of charged hadrons to be determined with much better resolution than compared to calorimeter-only based measurements, allowing a decomposition of the jet constituents down to the particle level.

Each particle that leaves a trace of its passage in the detector has one or more PF elements associated to it, corresponding to its interactions with the various subdetectors. The elements belonging to the same particle are grouped together in blocks by a linking algorithm [78], [85]. The PF algorithm then proceeds to identify the candidates in the following order: muons, electrons, charged hadrons, photons and neutral hadrons. Then these fundamental constituents are combined to reconstruct jets, tau leptons and calculate the missing energy in the transverse plane. The reconstructed jets are the manifestation of the of the quarks or gluons produced in the collision, while from the missing transverse energy the energies and directions of particles that do not interact with the detector, e.g. neutrinos, can be inferred.

Particle flow candidates are also used to reconstruct and identify the location of all pp interaction vertices.

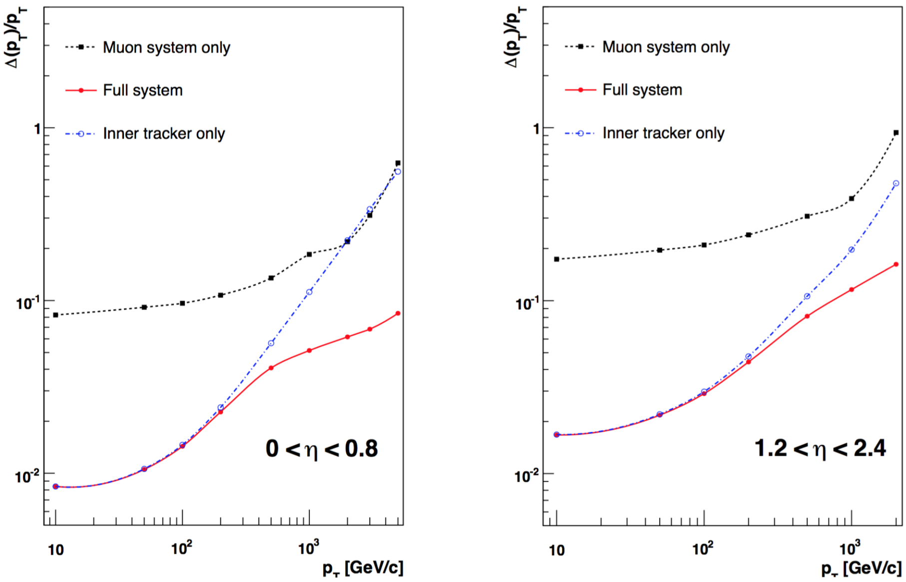
5.3.1 Muons

In the CMS reconstruction software, there are three possible ways to define a muon candidate. A *standalone muon* is reconstructed by using only the local reconstruction in the muon chambers. A *tracker muon* is a candidate with a track from the tracker of $p_{\mathrm{T}} > 0.5$ GeV and p > 2 GeV, whose extrapolation to the muon system, taking into account the average expected energy losses and multiple scattering in the detector material, is compatible with at least one muon segment (i.e. a short track stub made of DT or CSC hits). This kind of reconstruction is targeted especially to low-momentum muons. The third possibility is *global muon* reconstruction: for each standalone muon, the trajectory is extrapolated to the inner tracker detector and a matching track in the tracker is found. The hits in both subdetectors are then re-fitted. At the matching stage any ambiguity is solved by a global χ^2 fit that is used to select a unique global muon. As shown in Fig. 5.2 the information from the inner tracking system can improve the expected momentum resolution for a muon for $p_{\mathrm{T}} < 200$ GeV with respect to the momentum as determined solely from the muon chambers, while for highly energetic muons the measurement in the muon system improves the momentum resolution [59, 95].

The PF algorithm selects global muon candidates if their combined momentum is compatible with that determined only from the tracker within three standard deviations in order to lower the misidentification of charged hadrons as muons. Starting from the LHC Run in 2016, the alignment position errors, namely the uncertainties due to the position of the muon chambers with respect to the silicon detectors, is taken into account in the muon reconstruction. The final resolution on the muon momentum measurement depends on the p_{T} and η of the candidate, and ranges and ranges from 1% for very low momenta in the central region, up to 7% and 10% for higher momenta in the region $|\eta| < 0.9$ and $1.2 < |\eta| < 2.4$, respectively [96].

Muon identification selection

At the analysis level, the purity of the muon candidates sample is increased by applying selection criteria based on:



(a) (b)

Figure 5.2: The muon transverse momentum resolution as a function of the transverse momentum p_{T} using the muon system only, the inner tracking only, for both the $|\eta| < 0.8$ (a) and $1.2 < |\eta| < 2.4$ regions [65].

- χ^2/dof, the quality of the global-muon track fit,
- *number of hits* in the tracker system and in the muon spectrometer,
- d_{xy} and d_z, the transverse and longitudinal impact parameters w.r.t. the PV,
- $\sigma(p_{\mathrm{T}})/p_{\mathrm{T}}$, the relative uncertainty on the muon track transverse momentum,
- the *relative isolation* $\mathrm{I}/p_{\mathrm{T}}$, which depends on the energy activity around the muon candidate trajectory. The isolation I is computed as:

$$\mathrm{I} = \mathrm{I}_{\Delta\mathrm{R}=0.4} = \sum p_{\mathrm{T}}^{\mathrm{ch\ had}} + \max(0, \left(\sum p_{\mathrm{T}}^{\mathrm{neut\ had}} + \sum p_{\mathrm{T}}^{\gamma} - p_{\mathrm{T}}^{\mathrm{PU,\ ch\ had}}\right) \qquad (5.1)$$

considering all PF candidates reconstructed within a cone of radius $\Delta R = 0.4$ around the momentum direction of the muon candidate. In Equation 5.1, the first term refers to charged PF candidates originating from the primary vertex, the second to neutral hadrons, and the third to photons. The last term corrects the isolation for the energy associated to PU interactions. The contribution of the muon candidate is excluded from the isolation computation. The relative isolation is also defined as $\mathrm{I}/p_{\mathrm{T}}$.

	Variable	high-p_{T} cat.	loose cat.
Reconstruction	global muon	yes	–
	global or tracker muon	–	yes
	PF muon	yes	yes
Identification	$\chi^2/d.o.f.$	< 10	–
	muon chamber hits in global fit	> 0	–
	segments with two muon stations	> 1	–
	inner tracker (pixel) hits	> 0	–
	tracker layers with hit	> 5	–
	d_{xy}	< 0.2 cm	–
	d_z	< 0.5cm	–
	$\sigma(p_{\mathrm{T}})/p_{\mathrm{T}}$ of the best muon track	< 0.3	–
Isolation	$\mathrm{I}_{\Delta\mathrm{R}=0.4}/p_{\mathrm{T}}$	< 0.2	< 0.25

Table 5.1: Muon identification selection for the high-p_{T} and loose categories. The former is used for the analysis performed on the data recorded in 2012, whilst the latter for the 2016 data.

The selection applied on these variables depends on the identification efficiency and purity desired for the analysis. In this work the so-called *high-p_{T}* and *loose* identification criteria are used to maximize signal efficiency for the two analyses performed with the Run 1 and Run 2 datasets respectively. The two selection criteria for the 2012 and 2016 data-taking periods are presented in Table 5.1.

Reconstruction and identification efficiencies are estimated in data with a tag-and-probe technique in $\mathrm{Z} \to \mu^+\mu^-$ events [98]. The loose identification efficiency in data and simulation is presented in Figure 5.3 as a function of the muon $|\eta|$.

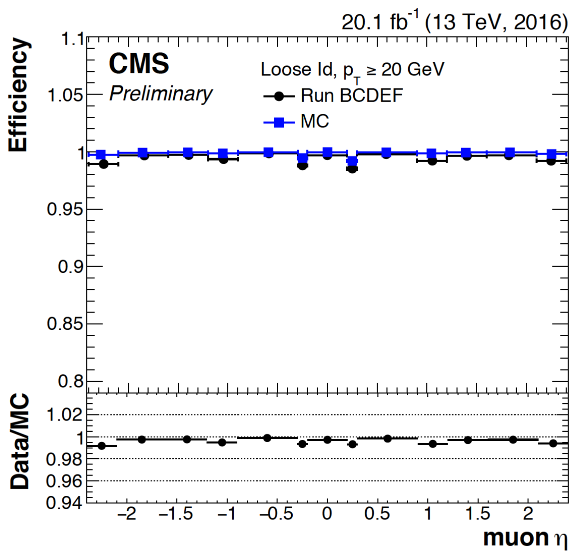


Figure 5.3: Muon loose category identification efficiency in data collected during 2016 in pp collisions at $\sqrt{s} = 13$ TeV (black point) and simulation (blue square) as a function of the muon $|\eta|$ [97].

The efficiency results are compared in data and simulated events and data-to-simulation corrective factors are derived based on their differences.

5.3.2 Electrons

While traversing the innermost part of the detector, the electrons produced in collisions lose between 33–86% of their initial energy due to bremsstrahlung [99], following a non-gaussian distribution that depends on the material encountered before reaching the ECAL [100]. The energy deposits in the ECAL crystals are spread mainly in the ϕ direction, because of the motion of the electrons in the magnetic field. Electron tracks can be reconstructed with the standard Kalman filter track procedure used for all charged particles. However, the large radiative losses for electrons in the tracker material compromise this procedure and lead in general to a reduced hit collection efficiency (hits might be lost when the change in curvature is large because of bremsstrahlung), as well as to a poor estimation of track parameters. In addition, the energy of radiated photons has to be taken into account to precisely reconstruct the electron initial energy. Creation of additional electrons from photon conversion can also occur.

These difficulties require a specific reconstruction procedure for electrons, that can proceed in two ways [99]. With *Tracker seeding*, the track is fitted from the triplets or doublets of hits from the tracker and later the calorimeter information is added. This procedure targets especially the reconstruction of low-p_{T} electrons in jets. The second method is *ECAL seeding*, which starts from measurements of ECAL superclusters (SC) position and energy and then extrapolates the electron trajectory towards the collision vertex with the helix corresponding to the initial electron energy, propagated through the magnetic field without emission of radiation. The SC are selected requiring some minimal energy deposit thresholds and low hadronic activity in the HCAL towers in a region close to the ECAL deposit ($\Delta R < 0.15$), to avoid jet-induced backgrounds. Then, the trajectory from superclusters is propagated inward assuming both the positive and negative charge hypothesis and matched to pairs or triplets of hits in the inner tracker layers (track seeds) compatible with being generated by an electron. Trajectories of the electron candidates are then globally refitted using a dedicated Gaussian sum filter (GSF) [100] algorithm that takes into account the radiative energy losses for the electrons. The bremsstrahlung energy loss distribution due to photons emitted in a direction tangent to the electron trajectory in the tracker material is modeled by a sum of Gaussians rather than by a single Gaussian, improving the resolution on reconstructed momentum with respect to the standard Kalman filter.

To prevent uncorrelated tracker hits or jets from being reconstructed as electrons, identification requirements based on the ECAL shower shape and the track-ECAL cluster matching are usually applied.

Different selections are used for electron candidates found in the barrel or in the endcaps, because of detector differences in these regions. Data and Monte Carlo simulations reproducing Drell-Yan decays into electron pairs are used to define various working points, each one targeting a different efficiency and misidentification rate. The electron energy is obtained by correcting the raw energy measurement from the ECAL superclusters for the imperfect containment of the clustering algorithm, taking into account the losses due to radiation, the interaction with the upstream ECAL, and the gaps between the calorimeter modules. The contributions from the average energy from the pileup radiation are removed.

The momentum scale is calibrated in Run 1 with an uncertainty smaller than 0.3%. The momentum resolution for electrons produced in Z boson decays depends on the electron pseudorapidity and ranges from 1.7% to 4.5%, [99].

Electron identification selection

Electrons considered in this work are selected with $|\eta| < 2.5$, which corresponds to the pseudorapidity coverage of the tracker.

Electrons are selected from reconstructed PF candidates based on a *cut-based* definition that utilizes the following variables [99, 101]:

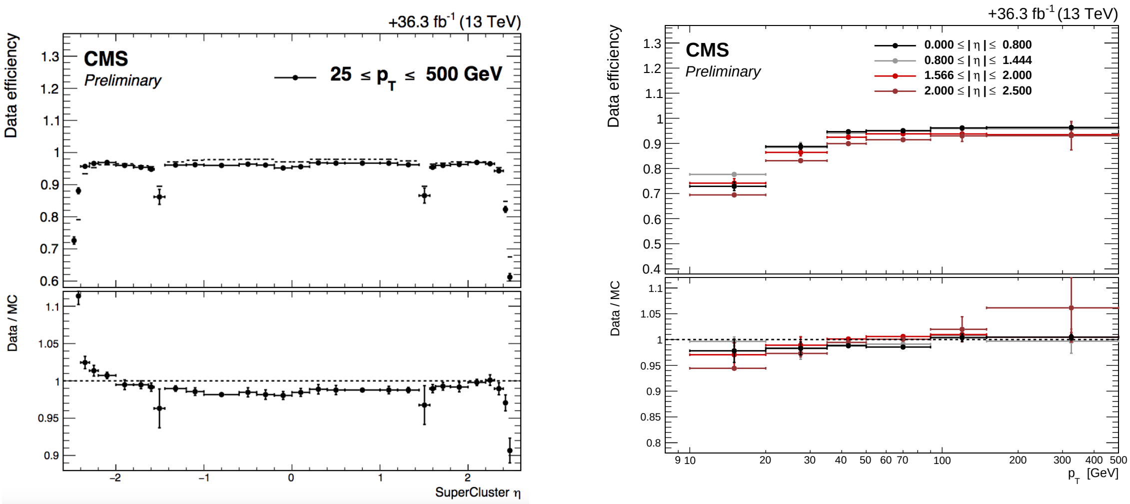
- $|\Delta\eta_{in}|$ and $|\Delta\phi_{in}|$, are the differences in η and ϕ between the track extrapolation and the supercluster position,
- σ_{inin}, the width of the supercluster along the η direction,
- d_{xy} and d_z,the transverse and longitudinal impact parameters with respect to the PV,
- *missing hits*, electrons from conversion are characterized by missing hits in the innermost tracker layers,
- *conversion veto*, a special veto is in place to mitigate the identification of reconstructing a track from a conversion vertex,
- $\left|\frac{1}{E} - \frac{1}{p}\right|$, which represents the difference between the inverse of the energy as measured from the ECAL and the momentum as measured form the tracker,
- H/E, a hadronic leakage variable that measures the energy fraction deposited in the HCAL,
- the isolation I or, alternatively, the relative isolation $\mathrm{I}/p_{\mathrm{T}}$: the main source of the misidentification of primary electrons comes from jets and electrons from

	Variable name	tight cat.		veto cat.	
		barrel	endcap	barrel	endcap
Identification	$\lvert\Delta\eta_{in}\rvert$	< 0.004	< 0.005	< 0.00749	< 0.00895
	$\lvert\Delta\phi_{in}\rvert$	< 0.03	< 0.02	< 0.228	< 0.213
	σ_{inin}	< 0.01	< 0.03	< 0.0115	< 0.037
	d_{xy}	< 0.02 cm	< 0.02 cm	-	-
	d_z	< 0.1 cm	< 0.1 cm	-	-
	missing hits	0	0	≤ 2	≤ 3
	conversion veto	false	false	false	false
	$\left\lvert\frac{1}{E} - \frac{1}{p}\right\rvert$	< 0.05	< 0.05	< 0.299	< 0.159
	H/E	< 0.12	< 0.10	< 0.356	< 0.211
Rel iso.	$\mathrm{I}/p_{\mathrm{T}}$	< 0.10	< 0.10	< 0.175	< 0.159

Table 5.2: Electron identification selection for tight and veto categories. The tight category is used in the Run 1 analysis for the data collected at $\sqrt{s} = 8$ TeV, whilst the analysis with the Run 2 data collected at $\sqrt{s} = 13$ TeV makes use of the veto selection. [104]

semileptonic quark decays. In this case, a higher energy activity close to the electron trajectory is registered. The $\mathrm{I}/p_{\mathrm{T}}$ variable is defined as for the muon in Equation 5.1, using an isolation cone of $\Delta R = 0.3$ centered along the lepton direction. In the isolation definition, the neutral pileup contribution is considered by taking into account the energy deposits in the calorimeter, estimated through the so-called ρ area method, by subtracting the median energy density ρ in the event multiplied by the electron energy deposits effective area. The isolation value is computed in a cone of $\Delta R = 0.3$.

Various selection criteria for electron identification are defined: veto, loose, medium, and tight. Based on the selection requirement applied on the previously listed variables average electron identification efficiencies of 95, 90, 80 or of 70% are achieved. In this work, the tight and veto categories are used for the Run 1 and Run 2 analyses respectively and their associated selections are presented in Table 5.2. In Run 1, the ECAL barrel-endcap overlap region, $1.4442 < |\eta| < 1.566$, was excluded because the reconstruction of an electron object in this region was not optimal. Reconstruction and identification efficiencies are estimated in data using a tag-and-probe technique on $\mathrm{Z} \to \mathrm{e}^{+}\mathrm{e}^{-}$ events [102], [99]. The results are compared between data and simulation and correction factors are derived which correct the simulation to approximate the data to within uncertainties. The reconstruction and veto electron identification efficiencies in 2016 data and simulation are presented in Figure 5.4 as a function of p_{T} and $|\eta|$.



(a) (b)

Figure 5.4: Electron reconstruction (a) and veto category identification efficiency (b) in data collected during 2016 pp collisions at $\sqrt{s} = 13$ (top pad) and data to simulation efficiency ratios (bottom pad) as a function of $|\eta|$ and p_{T} (for different $|\eta|$ regions) of the electrons. The large reconstruction data/MC scale factors at high psudorapidity region are due to different beam spot positions in data and simulations [103].

5.3.3 Modified lepton isolation

As already pointed out, an important feature of the lepton identification is the isolation requirement that allows physical processes with real leptons to be discriminated from processes like QCD multijet production, where the jets can be misidentified as leptons. In the particular final states used in these analyses, in order to not lose lepton identification capability due to the proximity of the hadronic tau, a special isolation computation has been studied in [106](Sec.3.2). The products of the reconstructed and identified (see section 5.3.7.5) hadronic tau lepton decay that enter in the lepton isolation cone are not considered for the isolation computation. Then standard isolation

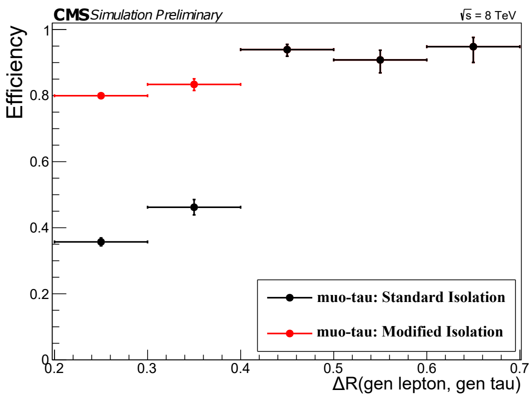


Figure 5.5: Comparison between the efficiency of the standard lepton isolation and the modified one, where the PF candidates from an identified hadronic tau lepton decay are removed, as a function of the angular distance between the lepton and the hadronic tau decay, in the $\mu\tau_{\mathrm{h}}$ final states for simulated events with $\mathrm{Z}' \to \mathrm{ZH} \to \mathrm{q\bar{q}}\tau^{+}\tau^{-}$ with a Z' mass of 2.5 TeV. [105]

criteria are applied to this modified isolation. The improvement on the signal efficiency is shown in Fig. 5.5.

5.3.4 Photons

Photons are reconstructed through ECAL clusters only. Variables related to the cluster shape are used to discriminate between photons that undergo conversion into an electron-positron pair early or late in the detector material in order to have a more precise energy measurement. Significant improvements in energy resolution are obtained by correcting the initial sum of energy deposits forming the supercluster for the variation of shower shape and containment in the clustered crystals and for the shower losses of photons that convert before reaching the calorimeter, depending on the photon rapidity. The photon energy resolution varies from 1% to 3%, depending on the pseudorapidity range [107].

5.3.5 Jets

5.3.5.1 Reconstruction

Most of the processes of interest at the LHC contain quarks or gluons in the final state. Partons are not directly observable and manifest themselves by undergoing hadronization and forming collimated jets of stable particles that are detected in the tracker and the calorimetric systems.

The energy and momentum of partons produced in the primary interaction is reconstructed and inferred using jet algorithms to cluster the particles coming from their hadronization. On average, charged particles, photons, and neutral hadrons share, respectively, 65%, 25% and 10% of the jet energy. Since the highest energy fraction is carried by charged particles, jets are reconstructed more precisely by including the information from the tracking system in addition to calorimetric clusters. The tracks from charged hadrons identified by PF are linked to calorimetric deposits if the particle $p_{\mathrm{T}} > 750$ MeV, and the best energy determination is obtained by the combination of the tracker and calorimetric measurements. A photon or a neutral hadron is defined if the calorimetric energy deposits are not linked to any track or exceed the associated track momentum. Instead, tracks with momenta that exceed considerably the calorimetric energy deposits, are associated to minimum ionizing particles, consistent with muon candidates.

All these PF particles (muons, electrons, photons, charged and neutral hadrons) are used as input for the jet (hence called *PF jets*) reconstruction. This type of jet has a better energy precision than jets built using the HCAL and ECAL information only, called *Calo jets* [78].

The jet energy response and direction resolution are presented in Figure 5.6 - 5.7 for PF and Calo jets, showing the better performance of the former which uses all PF constituents as inputs to the reconstruction [78].

Particle candidates are merged together into jets by sequential clustering algorithms, implemented in the FastJet package [108]. Multiple algorithms are used by the CMS collaboration. The sequential clustering algorithm is designed to be infrared and collinear safe if the final state particles undergo a soft emission or a collinear gluon splitting, so that the number and shapes of the jets should not change. These algorithms reconstruct jets considering two input objects and defining their relative distance d_{ij} and their distance to the beam-spot d_{iB}:

$$d_{ij} = min\left(p_{\mathrm{T_i}}^{2a}, p_{\mathrm{T_j}}^{2a}\right)\frac{\Delta R_{ij}^2}{R_0^2}, \tag{5.2}$$

$$d_{iB} = p_{\mathrm{T_i}}^{2a}, \tag{5.3}$$

where $p_{\mathrm{T_i}}$ is the transverse momentum of the i-particle, a is the parameter of the clustering algorithm chosen, ΔR_{ij}^2 is the angular distance between the two particles, and R_0 is a parameter of the algorithm that quantifies the jet size. All possible combinations of particles are evaluated and the minimum between all d_{ij} and all d_{iB} values is searched for until no particles are left on the input list of the algorithm. I f the minimum value corresponds to the:

- *beam distance (d_{iB})*, the corresponding particle i is removed from the input list and defined as a jet candidate,

- *distance between objects (d_{ij})*, the i and j particles momenta are merged into a new constituent that is created and added to the algorithm list.

The process terminates when all the particles are assigned to a jet that is separated from the others by a distance greater than R_0.

If the minimum value corresponds to the beam distance d_{iB}, then the corresponding particle i is removed from the input list and defined as a jet candidate, whereas if the minimum value corresponds to the distance between objects d_{ij}, then the i and j particles are merged into a new constituent that is created and added to the algorithm list

The $a = 1$ choice corresponds to the inclusive-k_{T} algorithm [109], while $a = 0$, to the *Cambridge-Aachen* algorithm. The exponent for the anti-k_{T} algorithm [110] is $a = -1$, meaning that soft-p_{T} particles are added first to the higer-p_{T} particles in their proximity. Instead two soft particles with the same ΔR_{ij}^2 will have a larger d_{ij} distance. Therefore soft constituents will cluster with hard ones before they cluster among themselves. If no hard particles are found within a distance of $2R_0$, all particles within a radius R_0 are accumulated in a canonical jet. The anti-k_{T} algorithm is *infrared safe*, namely, because soft radiation doesn't modify the shape of the jets that is instead determined by hard radiation. It is also *collinear safe* meaning that if a hard constituent is split into two or more softer-p_{T} collinear candidates, the resulting jet does not change. In the 13 TeV data analysis, clustering parameters of $R_0 = 0.8$ and $R_0 = 0.4$ are used to define the *large* or *wide cone* jets or AK8 jets, and the *standard* jets or AK4 jets. For the analysis of the 8 TeV data, the large cone jets used are defined with the Cambridge-Aachen algorithm and a $R_0 = 0.8$ (CA8), while the standard jets are clustered with $R_0 = 0.5$ (AK5).

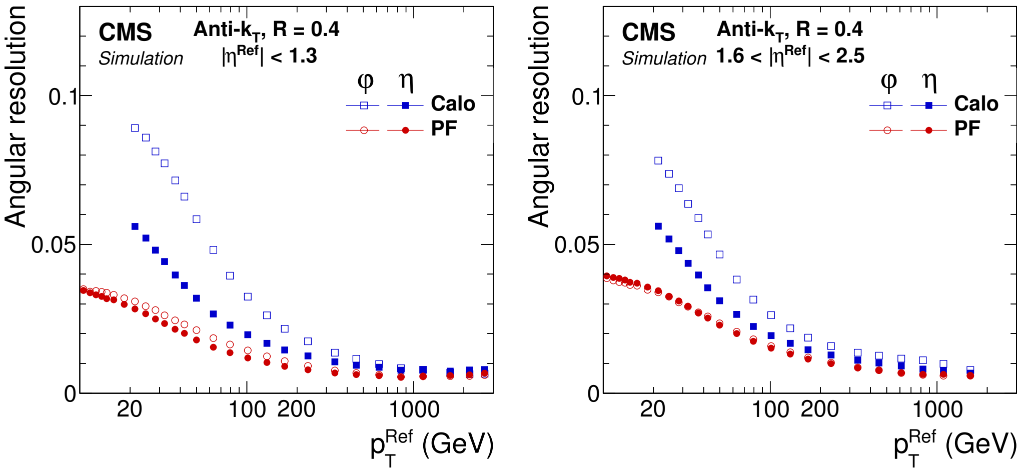
In order to reduce the contribution from pileup, which could bias especially the jet energy reconstruction, charged hadron candidates whose impact parameter is not compatible with the primary vertex are removed from the constituent of the jets, by the so-called *charged hadron subtraction* (CHS) method [111].

5.3.5.2 Energy calibration

Several levels of jet energy corrections are applied to the momentum of the clustered (*raw*) jets in order to obtain an energy value that is closer to the energy of the initial parton, correcting for effects due to non-linear detector response to different particles, detector segmentation, electronic noise, and noise due to other interactions during the same bunch crossing. In particular, a correcting factor $Corr$ is applied to the four-momentum vector p_i^{raw} of each particle clustered in the jet to obtain the corrected value p_i^{corr} [112, 113]:

$$p_i^{corr} = Corr \cdot p_i^{raw}.$$

The $Corr$ factor consists of different components derived from data and simulation and



(a)

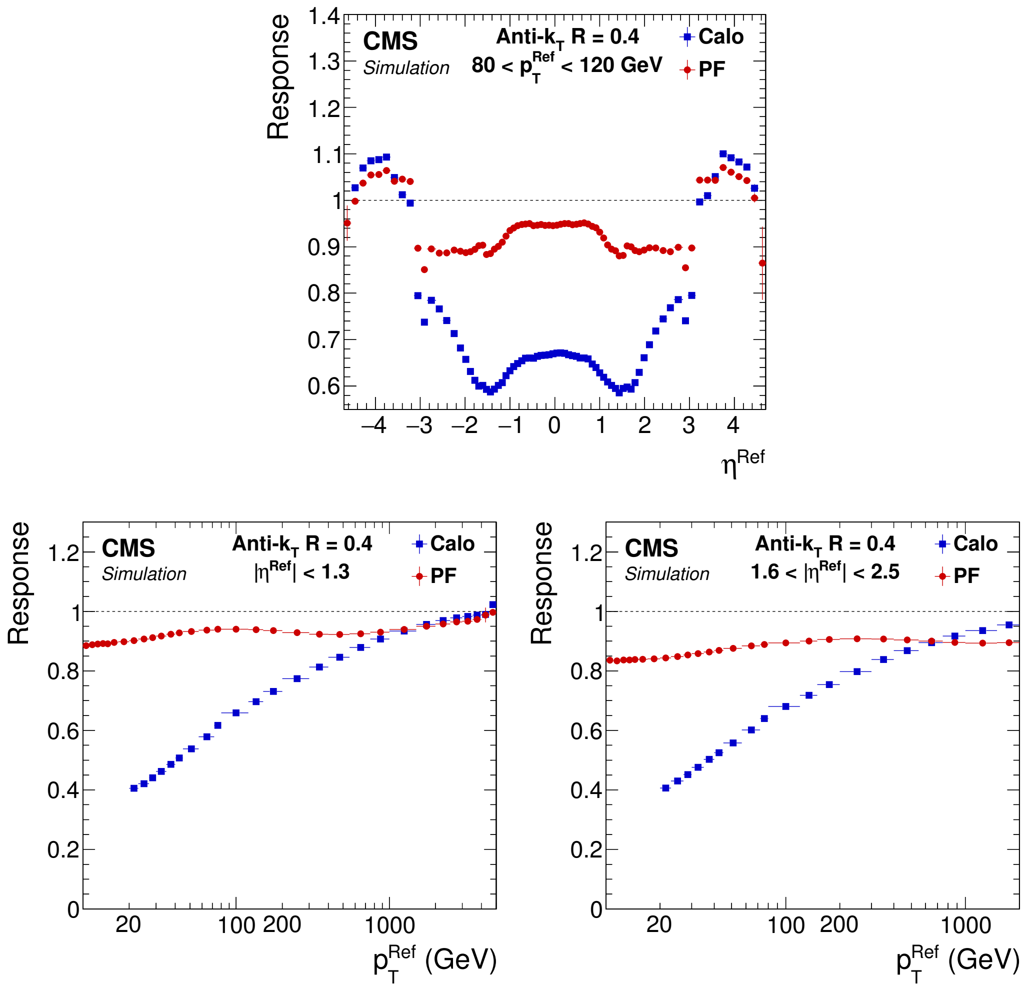
Figure 5.6: Jet $|\eta|$ and ϕ resolutions as a function of p_T in the the central (left) and forward (right) regions. Performances for reconstructed calo-jets (squares) and for PF-jets (circles) in simulation are presented in [85].

applied sequentially to p_i^{raw}, i.e the output of each step is the input to the next one. Residual corrections are determined from data to account for differences between the jet response in data and simulation.

The different components included in $Corr$ are presented in the following, and in Figure 5.8 a schematic view of the correction application is represented [112]:

- *L1 offset*, to subtract electronic noise and remaining PU contributions (referred to as "offset"). These corrections depend on the kinematical variables the jet (raw p_T, η, area) and on the event average p_T density per unit area, ρ. They are determined using QCD multijet simulations with and without pileup events. Relative differences with data are determined in zero-bias events, i.e. events not containing hard interactions and selected using random triggers, applying the only requirement of bunch crossing [114]. The offset corrections are reported in Fig. 5.9 (left) as a function of the jet pseudorapidity.

- *L2L3 corrections*, to correct for the jet response dependence on η and p_T due to non-uniformities in the ECAL and HCAL cluster energies response and detector properties [113]. The jet response corrections are determined in QCD di-jet simulations, by comparing the reconstructed p_T to the generated particle-level one (Note that particle-level jets do not include energy from neutrino contributions). The corrections are derived as a function of jet p_T and η and make the response



(a)

Figure 5.7: Jet energy response as a function of η (top) and p_{T} in te central (bottom left) and forward (bottom right) regions. Performances for reconstructed calo-jets (squares) and for PF-jets (circles) in simulation are presented in [85].

uniform over these two variables, as is displayed in Fig. 5.9(right). This correction is applied to both simulation and data;

- *L2L3 residuals*, to correct for remaining differences on the order of a few percent in the jet response in data as a function of η and p_{T}. The L2Residuals are η-dependent and calculated in di-jet events, while L3Residuals are calculated in $Z \to (\mu\mu, ee)+$ jet events, photon + jet events and multijet events, as a function of η and p_{T}.

After these factorized corrections, the average jet response, called the jet energy scale

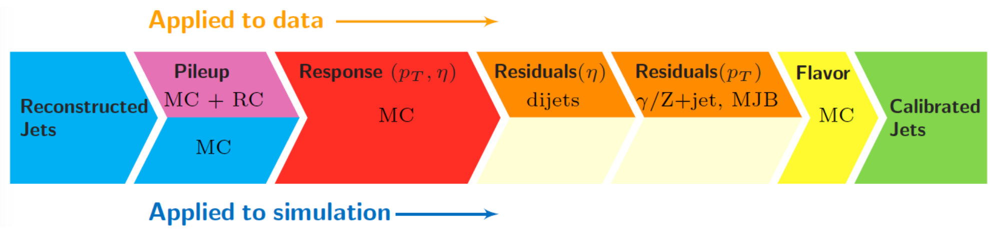


Figure 5.8: he scheme of how jet energy corrections are applied to data and simulation [115].

(JES), is compatible within uncertainties with unity. The total uncertainty is dependent on η and p_{T} and is smaller than 3% in the central region and 5% in the endcaps [98]. An additional effect that must be taken into account in the analysis is the discrepancy between the jet energy resolution (JER) observed in data and simulations. A smearing procedure is applied to simulated events in order to achieve a better agreement with data. The scaling method rescales the jet p_{T}, after jet energy scale corrections have been applied, by a factor that depends on the generator level p_{T}, as in:

$$c_{\mathrm{JER}} = 1 + (sf_{\mathrm{JER}} - 1)\frac{p_{\mathrm{T}} - p_{\mathrm{T}}^{gen}}{p_{\mathrm{T}}}, \tag{5.4}$$

where sf_{JER} is the data-to-simulation correction factor for the resolution. The match between simulated, reconstructed jets and generator level jets is done based on the spatial direction ($\Delta R < R_0/2$) and transverse momentum ($|p_{\mathrm{T}} - p_{\mathrm{T}}^{gen}|/p_{\mathrm{T}} < 3 \cdot \sigma_{\mathrm{JER}}$), where the σ_{JER} is the resolution of the energy scale in the simulations. When there is no matching, the stochastic smearing is used and the jet four momentum is scaled by a factor

$$c_{\mathrm{JER}} = 1 + Gaus(0, \sigma_{\mathrm{JER}})\sqrt{(\max(\mathrm{sf_{JER}}^2 - 1), 0)}, \tag{5.5}$$

where $Gaus(0, \sigma_{\mathrm{JER}})$ is a random number extracted from a Gaussian distribution of mean 0 and variance $\sigma_{\mathrm{JER}}{}^2$. The scaling factor c_{JER} is positively defined. This combination of the scaling and stochastic procedures is called the hybrid method. The jet energy resolution in the central region ranges from 10% for a jet with $p_{\mathrm{T}} > 100$ GeV to 4% for a jet with $p_{\mathrm{T}} \sim 1$ TeV [98].

5.3.5.3 Pileup mitigation on jet observables

In place of the CHS algorithm defined in 5.3.5.1 and used in Run 1, in Run 2 the algorithm chosen to mitigate the pileup effect on the main jet observables used in the analysis, jet mass and N-subjettines, is the pileup per particle identification (PUPPI) [117].

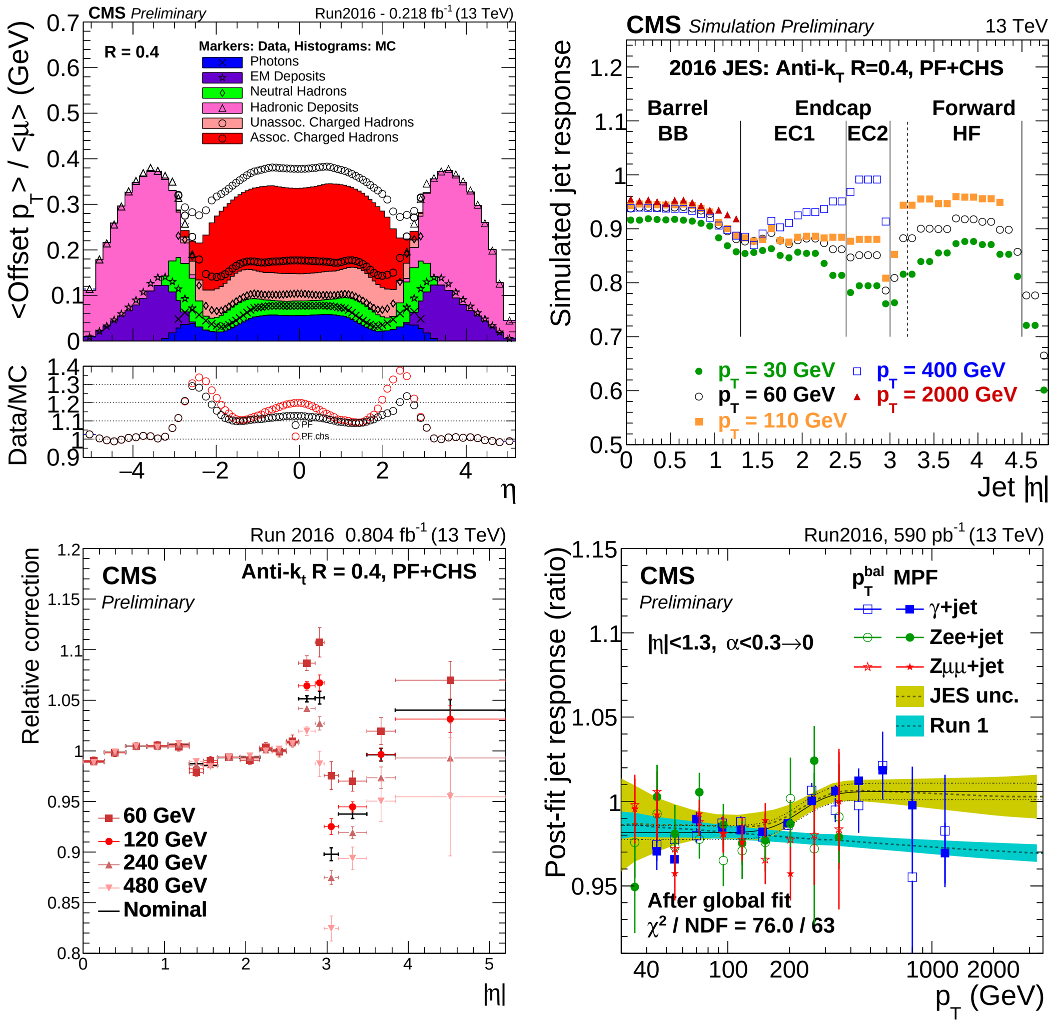


Figure 5.9: Offset of the energy measurement in data (markers) and simulation (histograms) normalized by the average number of interactions, separated by type of PF candidate. The ratio of data over simulation, representing the scale factor applied for the pileup offset in data, is also shown for PF and PF+CHS (left). The simulated jet response values for the L2L3 corrections (top right). The η-dependent L2Residual correction and PT-dependent L3Residual correction are shown in the bottom row in the left and right plots, respectively [116].

This method uses event pileup properties, tracking information, and shape of the local candidate distribution in order to distinguish parton shower radiation from pileup-like radiation and to compute for each constituent a weight that reflects its likelihood to originate from pileup. This weight is also used to rescale its four-momentum. A local variable α, constructed to differentiate the faster-falling p_{T} spectrum of pileup radiation as compared to that from the primary vertex, is computed using the distribution of charged pileup as a proxy for all pileup, in order to calculate a weight for each particle on an event-by-event basis.

Different definitions of α are used for the central ($|\eta| < 2.5$) and forward ($|\eta| > 2.5$) regions of the detector, where the tracking information is not available. In the central region, for a given particle i, α_i is defined as:

$$\alpha_i = log \sum_{j \in C, PV, j \neq i} \left(\frac{p_{\mathrm{T},j}}{\Delta R_{i,j}} \right)^2 \Theta(R_0 - \Delta R_{i,j}), \tag{5.6}$$

where Θ is the step function, i is the particle in question, and j ranges over the neighboring charged particles from the primary vertex within a cone of radius R_0. Charged particles are considered to be from the primary vertex if their track is associated with the primary vertex or its distance of closest approach to the leading vertex, along the beam pipe direction, is $d_z < 0.3$ cm. In the forward region, outside of the tracker coverage, the sum is taken over all particles. Due to the collinear singularity of the parton shower, a particle i from the hard physics process is likely to be near other particles from the same process so that α_i tends to be larger, while it is smaller for pileup particles. In order to define weights for the four-momenta, a χ^2 approximation is defined as:

$$\chi_i^2 = \frac{(\alpha_i - \bar{\alpha}_{\mathrm{PU}})^2}{RMS_{\mathrm{PU}}^2} \tag{5.7}$$

where $\bar{\alpha}_{\mathrm{PU}}$ is the median of the α_i distribution for pileup particles in the event and RMS_{PU} is the corresponding RMS of the distribution: in the central region they are calculated using the charged hadrons, while in the forward region all the particles in the event. Particles are then given a weight of $w_i = F_{\chi^2, NDF=1}(\chi_i^2)$, where $F_{\chi^2, NDF=1}$ is the cumulative distribution functions of the χ^2 with one degree of freedom. By construction, charged particles not originating from the primary vertex are assigned a weight of 0, similar to the charged hadron subtraction algorithm.

The algorithm parameter choices are close to what is recommended in Ref. [117]. Particles with $w_i < 0.01$ are rejected [118]. The p_{T} of every particle is corrected by its

Puppi weight, e.g $p_{\mathrm{T}} \leftrightarrow w_i \cdot p_{\mathrm{T},i}$. A selection dependent on the number of vertices is applied on the minimum scaled p_{T} of the neutral particles: $w_i \cdot p_{\mathrm{T},i} > (A + B \times nPV)$ GeV, where nPV is the number of reconstructed vertices in the event, and A and B are tunable parameters which are tuned separately in different pseudorapidity bins. For pseudorapidity $|\eta| < 3$ the parameters are tuned in order to optimize the resolution on the jet mass and p_{T}, while for $|\eta| > 3$ they are chosen such that the missing energy resolution is optimized. No additional pileup corrections are applied to jets clustered from these weighted inputs.

5.3.5.4 Jet mass

The jet mass is the main observable in distinguishing a jet due to a Vector (V) Boson (W/Z) or a Higgs boson from a jet produced by colored interactions (QCD jets). Jet-grooming techniques improve the suppression of uncorrelated underlying event and pileup radiation and soft radiation from the jet. They improve the discrimination by correcting the jet mass for QCD jets towards lower values, while maintaining the jet mass for boson jets near their original values.

Two main approaches were followed in the Run 1 and Run 2 analyses. The main grooming tool in Run 1 was “pruning” on CA8 jets [119, 120], while for Run 2 the “soft-drop” algorithm [121] with AK8 PUPPI jets. Typically, jet clustering algorithms with large radius (R = 0.8) cone (CA8 or AK8) are used in order to entirely contain the decay products of the highly Lorentz-boosted standard model bosons.

The pruning algorithm aims at removing soft and wide-angle radiation contributions to jets during the reclustering by checking two conditions for the protojets:

$$z = \frac{\min(p_{\mathrm{T},i}, p_{\mathrm{T},j})}{p_{\mathrm{T},p}} > z_{\mathrm{cut}} \text{ and } \Delta R(i,j) < R_{\mathrm{max}}, \tag{5.8}$$

where the soft threshold parameter z_{cut} is set to 0.1 and the maximal angular separation threshold is $R_{\mathrm{max}} = 0.5 \times m_{\mathrm{jet}}/p_{\mathrm{T,jet}}$, where jet means the original jet. If the previous conditions are not satisfied the protojet with the lower p_{T} is ignored [122]. Figure 5.10 (left) shows the pruning algorithm performance on simulated events containing jets originating from W boson and QCD.

Following theoretical work [123, 124] that aimed at understanding and calculating jet mass observables in QCD events, a new algorithm was developed to accomplish infrared safe jet grooming, called the soft-drop algorithm [121]. Like any grooming method, soft-drop declustering removes wide-angle soft radiation from a jet in order to mitigate the effects of contamination from initial state radiation (ISR), the underlying event (UE), and pileup. Subsequent to the AK8 clustering, the constituents of these jets are reclustered with the Cambridge-Aachen algorithm and the soft-drop procedure removes the softer constituents unless:

$$z = \frac{\min(p_{T,i}, p_{T,j})}{p_{T,p}} > z_{cut} \left(\frac{\Delta R_{ij}}{R_0} \right)^{\beta}, \tag{5.9}$$

where the soft threshold parameter z_{cut} is set to 0.1. The soft-drop algorithm depends on an angular exponent β that is set to 0 in the algorithm used in Run 2. Since the soft-drop algorithm is primarily aimed at separating W-jets from q/g-jets, it does not fully reject contributions from the underlying event or the pileup. Therefore it is used in combination with PUPPI. Since from studies on data and simulations , it is observed that the soft-drop puppi mass does not peak at the nominal boson mass by default, and the mean of the distributions are shifted with the jet p_{T}, a set of corrections, derived from data and simulations, are applied.

- Generator level: in simulations, at generator level, the soft-drop mass of the boson has a shift that depends on the p_{T}: a correction of 3–5% (especially for $p_{\mathrm{T}} < 500$ GeV, smaller for higher p_{T}) is applied to data and simulations.

- Reconstructed: the peak of the reconstructed soft-drop mass is not centered around the nominal value. This is due to p_{T} and η dependent reconstruction effects and it is similar in magnitude to the L2L3 corrections that were applied to the pruned mass in Run 1 (5–12%).

- Scale: the scale of the soft-drop mass is verified in data in a $\mathrm{t\bar{t}}$-enriched region, selected with one lepton and a jet compatible with a boosted W boson decay. The fitted mean of the peak in simulation is found to be in good agreement with data, and a scale factor 1.0000 ± 0.0094 is applied to simulated events to match the scale measured in data.

- Resolution: the soft-drop mass resolution is estimated with the same method as the scale, and the ratio between the resolution in data and simulation is found to be 1.00 ± 0.20. The jet mass in simulated events is smeared by this factor.

In Fig. 5.10 (right) the Run 2 soft-drop jet mass distribution is shown for simulated event with jets originating from W, Z or Higgs bosons and data, dominated by QCD

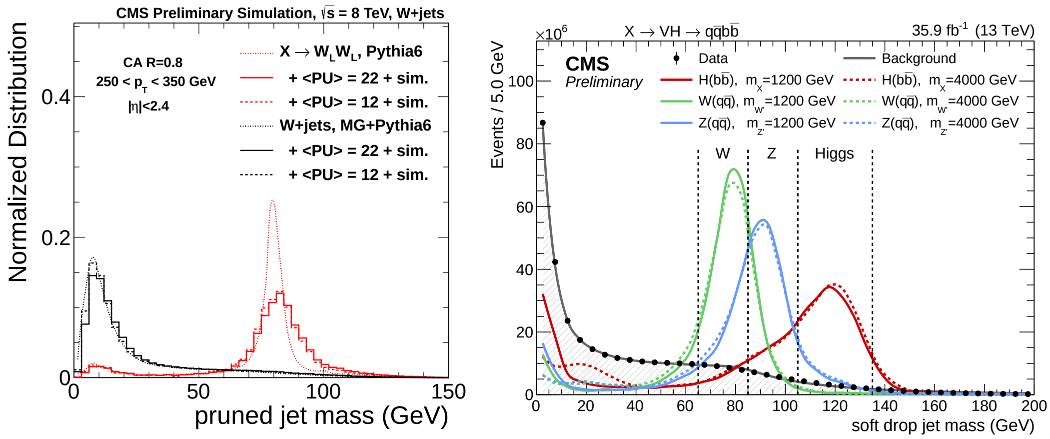


Figure 5.10: Pruned jet mass distribution (left) in simulated samples of boosted W bosons and inclusive QCD jets in the W+jet topology. MG denotes the MADGRAPH5 generator. The thick dashed lines represent the generator predictions without pileup interactions and without CMS simulation. The histograms are the distributions after CMS simulation with two different pileup scenarios corresponding to an average number of interactions of 12 and 22 [122]. soft-drop mass distribution (right) in simulated events for jets originated from W(green)/Z(blue)/H(red) bosons compared to jets in data that are predominantly produced by quark and gluons. Solid and dotted lines represent different p_{T} regimes: $p_{\mathrm{T}} \lesssim 1$ TeV and $p_{\mathrm{T}} \gtrsim 1$ TeV, respectively [125].

multijet production. Three regions of the soft-drop mass spectrum are depicted: the Wmass window, from 65 to 85 GeV, the Z mass window, from 85 to 105 GeV, and the Higgs mass window, from 105 to 135 GeV.

5.3.5.5 Jet substructure

In addition to the groomed mass, another tool to discriminate QCD jets from jets originating from SM boson decay is the inner structure of the jet. Jets originating from one parton have a different substructure than jets originating from more than one parton, and this can be studied through the spatial distribution of the jet constituents relative to the jet axis. The *N-subjettiness* variable [126] , calculated from a jet, is extensively used to identify boosted vector bosons that decay hadronically. This observable measures the distribution of jet constituents relative to candidate subjet axes in order to quantify how well the jet can be divided into N subjets. The transverse momentum and angular distance of each jet constituent with respect to the closest of the N subjet axes are then used to compute the N-subjettiness τ_N variable, as

$$\tau_N = \frac{1}{d_0} \sum_k p_{\mathrm{T},k} \times \min(\Delta R_{1,k}, \Delta R_{2,k}, ..., \Delta R_{N,k}), \tag{5.10}$$

where the normalization factor is $d_0 = \sum_k p_{\mathrm{T},k} \times R_0$, with R_0 being the clustering

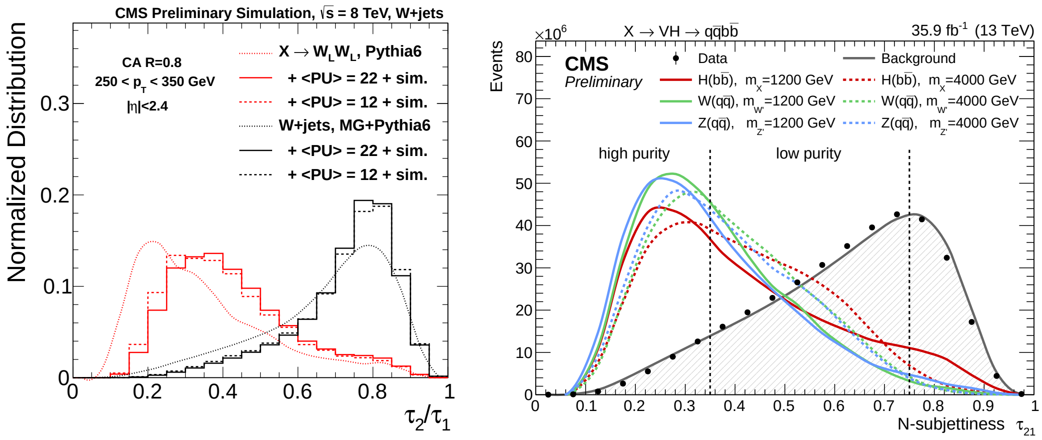


Figure 5.11: N-subjettiness distribution (left) in simulated samples of boosted W bosons and inclusive QCD jets in the W+jet topology. MG denotes the MADGRAPH5 generator. Thick dashed lines represent the generator predictions without pileup interactions and without CMS simulation. The histograms are the distributions after CMS simulation with two different pileup scenarios corresponding to an average number of interactions of 12 and 22 [122]. N-subjettiness distribution (right) in simulated events for jets originated from W(green)/Z(blue)/H(red) bosons compared to jets in data that are predominantly produced by quark and gluons. Solid and dotted lines represent different p_T regimes: $p_T \lesssim 1$ TeV and $p_T \gtrsim 1$ TeV, respectively [125].

radius parameter of the original jet ($R_0 = 0.8$ for this work), $p_{T,k}$ is the p_T of the k-th jet constituent, and $\Delta R_{N,k}$ is its distance from the N-th subjet. Particularly effective in distinguishing jets originating from boosted W, Z or H bosons from jets originated from quarks and gluons, is the ratio of the "2-subjettiness" and the "1-subjettiness" ($\tau_{21} = \tau_2/\tau_1$. In Run 1 analyses the N-subjettiness was calculated with the CA8 jets with the CHS pileup removal algorithm, while in Run 2 AK8 jets with the PUPPI algorithm for pileup removal are used. The distributions of τ_{21} are shown in Fig. 5.11. Scale factors to correct the τ_{21} efficiency are calculated form a separate sample of semileptonic $t\bar{t}$ events using boosted $W \rightarrow q\bar{q}\prime$ decays in data for the 0.4 and 0.75 working points. For events having a high-purity value of PUPPI τ_{21}, a scale factor 1 ± 0.6 is applied to the boson identification efficiency in signal samples. In the low purity category, the scale factor is 1.03 ± 0.23. Extrapolation uncertainties on the τ_{21}-selection due to propagation to higher momenta (than the $p_T \sim 200 - 300$ GeV in the reference $t\bar{t}$ sample) are estimated by comparing the selection efficiency in samples with W/Z bosons of $p_T \sim 200$ GeV from different generators (Pythia and Herwig), with simulated samples containing W/Z bosons with p_T in the range of interest.

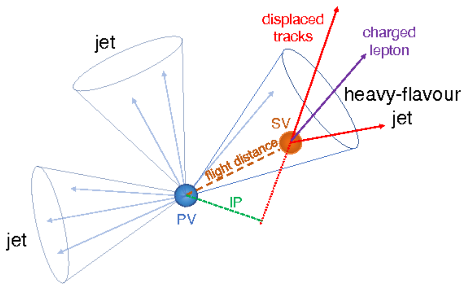


Figure 5.12: Illustration of a heavy flavor jet structure with a secondary vertex (SV): the products of the decay are charged particle tracks (possibly also lower p_{T} leptons) that are displaced from the primary vertex (PV), and hence have a large impact parameter (IP) [127].

5.3.5.6 Jets originating from bottom quarks

It is of vital importance for the analyses presented in this work, as well as for many new physics searches and SM measurements, to distinguish or *tag* jets coming from the hadronization of b quarks (b jets) from jets arising from c quarks (c jets), and light quarks and gluons (udsg jets).

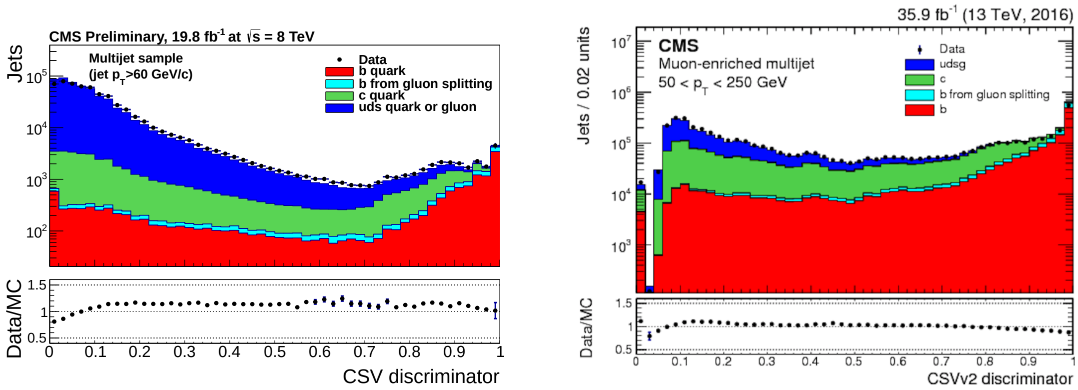
Tagger algorithms rely on the different properties of B hadrons with respect to other hadrons and gluon jets: larger masses, longer lifetimes ($\tau \sim 1.5$ ps, $c\tau \sim 450$ μm), large semileptonic branching ratio ($\sim 40\%$), and daughter particles with larger momentum with respect to other flavor hadrons. These unique properties are exploited by several algorithms to identify b jets, utilizing the reconstruction of lower level physics objects such as tracks, vertices and jets. Efficient track reconstruction and good spatial resolution close to the interaction point are necessary for all b tagging algorithms [127].

Tracks inside a jet candidate must satisfy criteria related not only to their quality, but also to their distance from the interaction point. The track impact parameter is the distance between the primary vertex and the coordinate of closest approach of the track helix. Since heavy-flavor hadrons have a long lifetime, the visible products of the decay are mainly charged hadrons and sometimes leptons, which are displaced from the primary vertex, hence having a large impact parameter, as illustrated in Fig. 5.12.

Appropriate selection criteria are applied to the tracks used as inputs for calculating variables used by the b-tagging algorithm. In particular, to ensure a good momentum and impact parameter resolution, tracks are required to have $p_{\mathrm{T}} > 1$ GeV, a χ^2 value of the trajectory fit normalized to the number of degrees of freedom below 5, and at least one hit in the pixel layers of the tracker detector. The last of these requirements is less stringent than the requirement used for b jet identification in Run 1, where at least eight hits were required in the pixel and strip tracker combined, of which at least two were hits in the pixel detector. In the first part of 2016, in fact, saturation effects were observed due to a high occupancy in the readout electronics of the silicon strip tracker, leading to a tracking inefficiency. The requirement on the number of hits was relaxed to recover tracking efficiency and after the fix in the electronics, this requirement was left because it had no impact on the b-tagging performance. The tracks that are too far from the interaction vertex are discarded to suppress contributions from pileup: the transverse (longitudinal) impact parameter of a selected track is required to be smaller than 0.2 (17) cm and the distance between the track and the jet axis at their point of closest approach is required to be less than 0.07 cm.

Several b-tagging algorithms have been deployed by the CMS collaboration, each one with its own peculiarities, but they usually can be separated into two main categories: the *track-based* algorithms are the ones that exploit the long B-hadron lifetime looking for displaced tracks and the *vertex-based* ones that reconstruct secondary vertices within the jets. At the start of the Run 2 of the LHC, the *inclusive vertex finding* (IVF) algorithm was adopted as the standard secondary vertex reconstruction algorithm used to define variables for heavy-flavor jet tagging. In contrast with the Run 1 *adaptive vertex reconstruction* (AVR) algorithm [128], which uses tracks clustered in the reconstructed jets as input, the IVF algorithm uses as input all the tracks in the event with $p_{\mathrm{T}} > 0.8$ GeV and impact parameter < 0.3 cm. This is well-suited for B hadron decays within relatively small angles giving rise to overlapping, or completely merged, jets, that are of interest for the analyses in this work. The AVR algorithm achieves this by running on clusters of tracks, without requiring them to be reconstructed as jets and by relaxing some requirements in order to maximize the secondary vertex reconstruction efficiency. In particular, secondary vertices are rejected when they share 80% or more of their tracks, and when the 2D flight distance significance is less than 2 (1.5) for secondary vertices used in b- (c-) tagging algorithms. The remaining secondary vertices are then associated with the jets by requiring the angular distance between the jet axis and the secondary vertex flight direction to satisfy $\Delta R < 0.3$. For jets with $p_{\mathrm{T}} > 20$ GeV in $\mathrm{t\bar{t}}$ events, the efficiency for reconstructing a secondary vertex for b (udsg) jets using the IVF algorithm is about 75% (12%), which is 10% (7%) higher than the efficiency for reconstructing a secondary vertex with the AVR algorithm.

To achieve a higher efficiency and reduced misidentification rate, the b tagger used in this work is the *Combined Secondary Vertex* (CSV) algorithm. For Run 2 of the LHC, a slightly improved version of the algorithm used for Run 1 was developed [129], and the multivariate technique was deployed on a larger set of input variables (CSVv2). Events are divided into categories based on the number of vertices: one reconstructed



(a) (b)

Figure 5.13: Discriminator distribution for the CSV algorithm shown shown for a selection of data enriched in multijet QCD events collected at $\sqrt{s} = 8$ TeV (a) [114], and for a similar sample collected with $\sqrt{s} = 13$ TeV that is more enriched in b-quark jets by requiring a muon from semi-leptonic decay in the opposite jet (b) [127].

secondary vertex, no secondary vertices but two tracks with large impact parameters, and the remaining cases. The training of the artificial neural network used in the CSV relies on quantities related to both the tracks and vertices, such as: number of tracks, p_{T} and η distributions of the tracks relative to the axis defining the local cluster of tracks, impact parameters, angular and linear 2D and 3D distances of the vertex from the tracks and the jet axis, and energy and invariant mass of the charged particles associated to the secondary vertex.

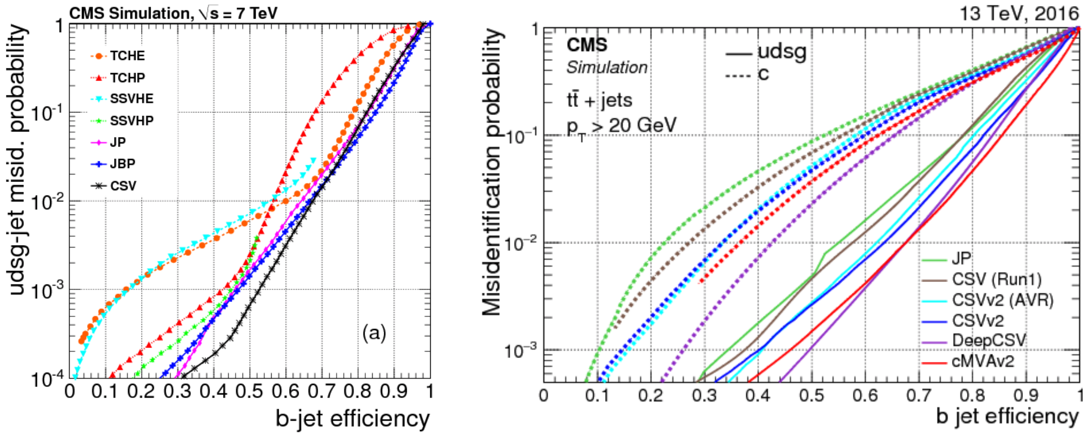
The algorithm is able to distinguish between b-quark and c-quark jets and also between b-quark and light-quark or gluon jets. The algorithm provides a continuous output between 0 and 1, where values close to 1 indicate a jet likely to arise from the hadronization of a b quark, as shown in Figure 5.13 in QCD multijet data for Run 1 and Run 2.

5.3.5.7 Efficiency of b-tagging

The performance of the b-tagging algorithms is based on the probabilities ϵ_b, ϵ_c and ϵ_q to tag correctly a b jet or to misidentify c jets or q (udsg) jets as b jets, respectively. These efficiencies are defined as:

$$\epsilon_i = \frac{\#\text{identified as b} - \text{jets}}{\#i\ - \text{jets}};\ \ i = b, c, q$$

and are shown in Figure 5.14 for simulation as obtained with the CSV algorithm.



(a) (b)

Figure 5.14: Performance of the CSV b-jet tagging algorithm compared to other algorithms in terms of identification efficiency and mistag rate are shown in black (a) for Run 1 [132] and in blue (b) for Run 2 simulations [127].

Based on the percentage of misidentified light-flavor jets, three different selections for the CSV discriminant are defined: loose ($\epsilon_q \sim 10\%$), medium ($\epsilon_q \sim 1\%$) and tight ($\epsilon_q \sim 0.1\%$) [130, 131].

The "loose" working point of the CSV algorithm is used in the analysis with the data collected in Run 1 of the LHC, for which the b-tagging efficiency is about 85% with mis-tagging probabilities of 40% for charm-quark jets and 10% for light-quark and gluon jets at jet $p_{\mathrm{T}} \sim 80$ GeV [131]. For the Run 2 analysis, two working points are used: the "medium", that has an efficiency for bottom-quark jets of 63% with a mis-tagging probability on charm-quark jets of 12% and 1% on light-flavor jets, and the "tight" which whose b-tagging efficiency, c-tagging, and light-flavor tagging efficiency are 41%, 2.2%, and 0.1%, respectively [127]. Corrections are applied to simulated data to cover differences in efficiency with respect to that of the data.

5.3.5.8 Identification of b jets in boosted topologies

Special techniques have been deployed in order to identify particles decaying to b quarks with a large Lorentz boost, analogously to what is done for W and Z boson tagging, using wide cone jets. Jet substructure techniques can then be applied to resolve the subjets that are associated with the partons from the boson decay, as is described in Section 5.3.5.5. B tagging can be applied either on the CA8 (AK8) jet or on its subjets, obtained with pruning or soft-drop, as is used with Run 1 and Run 2 analyses, respectively. In Run 2, for both the wide cone jet and the subjet tagging cases, the

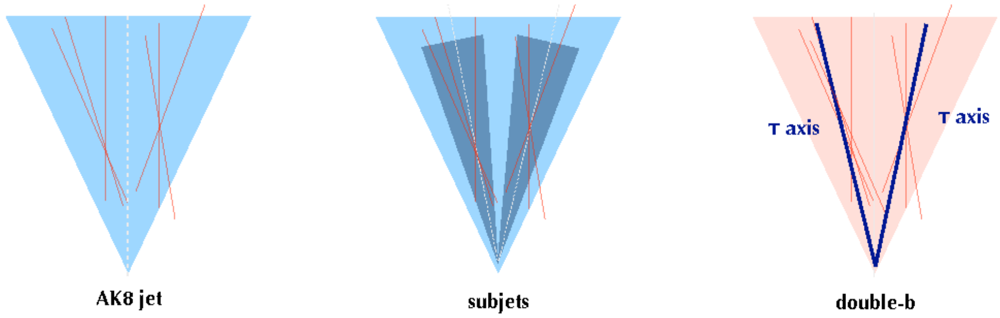


Figure 5.15: Schematic illustration of the AK8 jet (left) and subjet (middle) b-tagging approaches, and of the double-b tagger approach (right) [127].

CSVv2 algorithm is used. In the first approach the CSVv2 algorithm is applied to the AK8 jet, but using looser requirements for the track-to-jet and vertex-to-jet association criteria, consistently with the $R = 0.8$ parameter, whereas in the second approach, the CSVv2 algorithm is applied to the subjets, as depicted in Figure5.15.

Figure 5.16 shows the efficiency of identifying a $H \to \mathrm{b\bar{b}}$ jet versus the misidentification probability of different backgrounds: inclusive multijet events, $g \to \mathrm{b\bar{b}}$ jets and single b-quark jets. When b tagging is applied to the subjets, both subjets are required to be tagged. As will be explained later, the main backgrounds for the analyses presented here are top-quark pair production and the production of V bosons in association with jets. For these analyses, the most relevant plots of Figure 5.16 are the lower ones. As is shown in these plots, the subjet b-tagging algorithm performs better. The lower misidentification probability at the same efficiency is explained by the fact that for the subjet b tagging, the two subjets are required to be tagged. Requiring both subjets to be tagged while there is only one b hadron present in the background jets results in a lower misidentification probability. The double-b tagger algorithm [127] has a similar performance, but is not used in these analyses because the training and the validation of this tool, which uses multivariate analysis techniques, was restricted to jets with pruned masses ranging from 50 to 200 GeV. For the Run 2 data analyses, it was chosen to opt for a categorization on the multiplicity of subjets with a b tag. This provides a high purity region enriched in signal with 2 b-tagged subjets and a low purity, but higher statistics region to to maintain signal efficiency due to possible b-tagging failure at high masses.

For the Run 1 analyses, following the same principle of maximizing the signal acceptance, a combination of subjet b tagging and large cone jet b tagging was used with CA8 jets and pruned CA8 subjets [114]. The performance is shown in Figure 5.17.

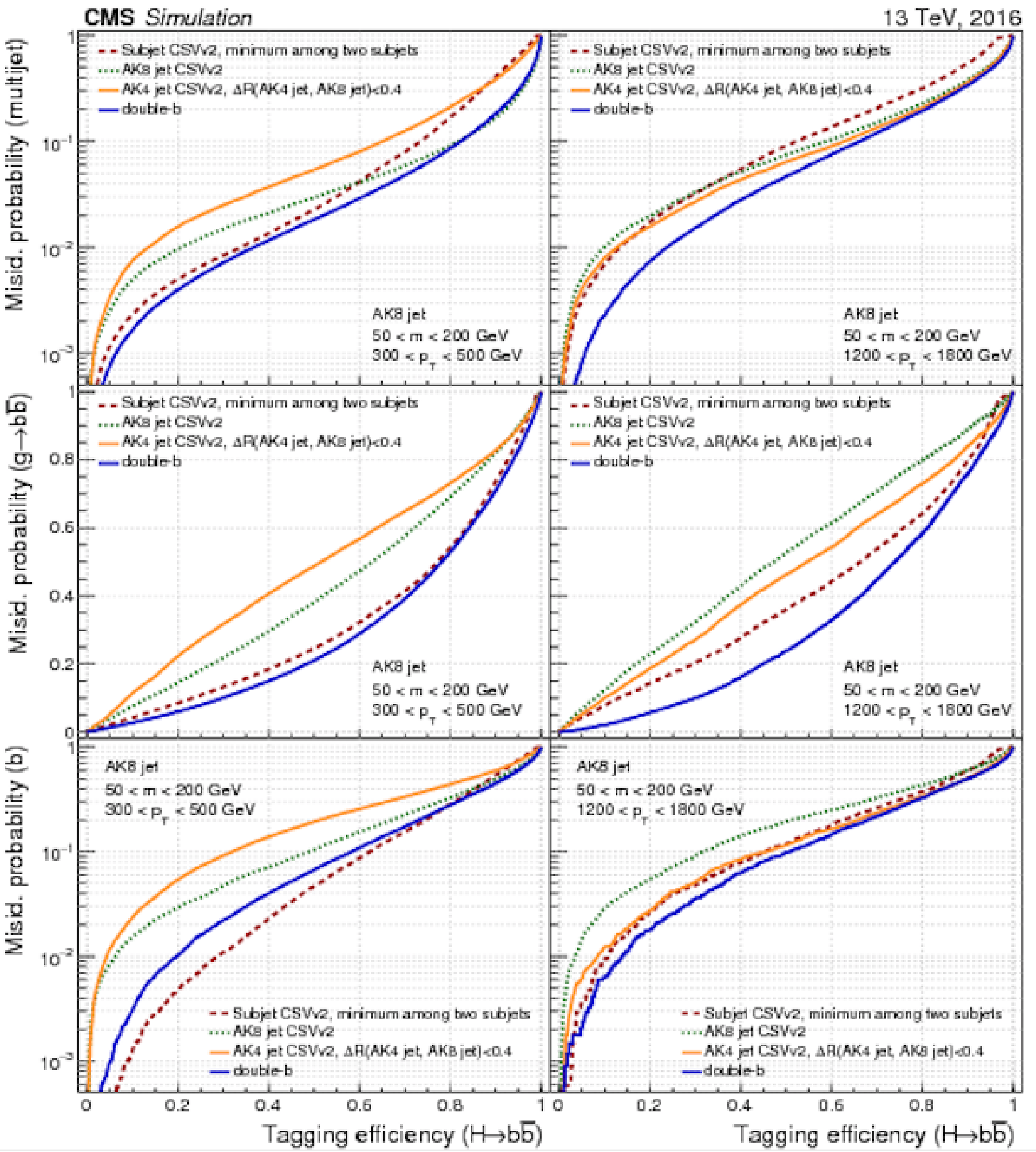


Figure 5.16: Misidentification probability using jets in a multijet sample (upper), for $g \to b\bar{b}$ jets (middle), and for single b jets (lower), versus the efficiency to correctly tag $H \to b\bar{b}$ jet, when the CSVv2 algorithm is applied to different kinds of jets: AK8 jets, their subjets and AK4 jets matched to the AK8 jets. For the subjet b-tagging curves, both subjets are required to be tagged. The double-b tagger is also applied to AK8 jets. The AK8 jets are selected to have a pruned jet mass between 50 and 200 GeV, with $300 < p_{\mathrm{T}} < 500$ GeV (left) and $1.2 < p_{\mathrm{T}} < 1.8$ TeV (right). [127]

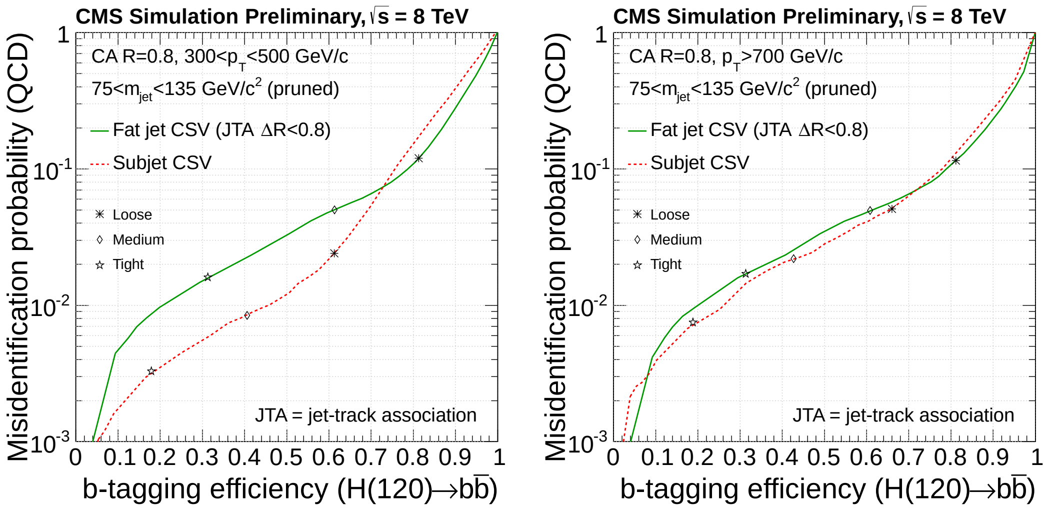


Figure 5.17: Misidentification probability, using jets in a multijet sample versus the efficiency to correctly tag $H \to b\bar{b}$ jet, when the CSVv2 algorithm is applied to different kinds of jets: CA8 jets and their subjets. For the subjet b-tagging curves, both subjets are required to be tagged. The CA8 jets are selected to have a pruned jet mass between 75 and 135 GeV, with $300 < p_T < 500$ GeV (left) and $p_T > 700$ GeV (right). [114]

The use of a fixed-size jet-track association cone leads to track sharing between the subjets of fat jets once their angular separation becomes similar or smaller than the size of the association cone. Because of track sharing, the b-tagging probabilities for individual subjets become increasingly correlated as the fat jet transverse momentum increases. For boosted Higgs jets, where both subjets are required to be b-tagged, this finally results in the subjet b-tagging performance approaching the fat jet b-tagging performance at large fat-jet transverse momenta, as reported in Figure 5.17. The final b-tagging procedure consists of tagging the CA8 subjets if the angular distance between them is greater that the track association cone ($\Delta R(sj_1, sj_2) > 0.3$), or tagging the fat jet if otherwise.

5.3.5.9 Jet selection

In this analysis, reconstructed candidates are considered as jets of large cone or small cone if they satisfy the following requirements: $|\eta| < 2.5$, neutral hadron and electromagnetic energy fraction smaller than 90% and charged electromagnetic energy deposits smaller than 99% of the candidate total energy. The reconstructed jets needs to be associated to at least one charged hadron.

The main selection requirements, including b-tagging requirements, as well as differences between Run 1 and Run 2, are summarized and highlighted in Tab. 5.3.

		Run 1	Run 2
small-cone jets	algorithm	AK 5	AK 4
	kinematic selection	$\|\eta\| < 2.4$ $p_{\mathrm{T}} > 20$	$\|\eta\| < 2.4$ $p_{\mathrm{T}} > 20$
large-cone jets	algorithm	CA8	AK 8
	kinematic selection	$\|\eta\| < 1$ $p_{\mathrm{T}} > 400$	$\|\eta\| < 2.4$ $p_{\mathrm{T}} > 200$
	pileup subtraction	CHS	CHS, puppi
	grooming	pruning	soft-drop
	H tagging	τ_{21} and either 2 b-tagged subjets or 1 b-tagged CA8	1 or 2 b-tagged subjets
	V tagging	—	τ_{21}: HP and LP

Table 5.3: Jet selection requirements used in the Run 1 analysis for the data collected at $\sqrt{s} = 8$ TeV and in the Run 2 analysis with data collected at $\sqrt{s} = 13$ TeV.

5.3.6 Missing transverse energy

The missing transverse momentum vector $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ can be defined as the imbalance in the transverse momentum of all the particles that interact in the detector. Because of momentum conservation, $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ correlates to the transverse momentum that is carried by weakly interacting particles, such as neutrinos. The $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ is defined as the negative vectorial sum of the transverse momenta p_{T} of all the PF particles reconstructed in the event [78] so it is also called PF $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$:

$$\vec{p}_{\mathrm{T}}^{\mathrm{miss}} = -\sum \vec{p}_{\mathrm{T}}.$$

The magnitude of this vector is known as the missing transverse momentum $p_{\mathrm{T}}^{\mathrm{miss}}$. Equivalently, the missing transverse momentum can also be referred to as missing transverse energy ($\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$), whose magnitude is $E_{\mathrm{T}}^{\mathrm{miss}}$.

The estimation of this variable strongly depends on the correct energy and momentum measurements for all the PF objects. Minimum energy thresholds in the calorimeters, inefficiencies in the tracker, and nonlinearities in the response of the calorimeter for hadronic particles could lead to a mis-measurement of this quantity.

To reduce possible biases on the $p_{\mathrm{T}}^{\mathrm{miss}}$, *Type-1* corrections are applied: they replace the vector sum of transverse momenta of particles which can be clustered as jets with the vector sum of the transverse momenta of the jets to which jet energy corrections is applied. Quality filters are further applied in the analysis to remove events where the $p_{\mathrm{T}}^{\mathrm{miss}}$ is severely mis-reconstructed.

The performance of missing energy reconstruction (scale and resolution) is studied in events with identified Z bosons decaying to leptons (electrons and muons) or isolated photons. The momenta of muons originating from Z bosons is known with a resolution of 1–6% [133], 1–4% for electrons and photons [107], while for jets it is around 5–15% [114]. Therefore the $p_{\mathrm{T}}^{\mathrm{miss}}$ resolution is dominated by the hadronic activity in the event.

The performance is evaluated by comparing the momentum of the vector boson to the one of the hadronic recoiling system. For momentum conservation:

$$\overrightarrow{q}_{\mathrm{T}} + \overrightarrow{u}_{\mathrm{T}} + \overrightarrow{p}_{\mathrm{T}}^{\mathrm{miss}} = 0,$$

where $\overrightarrow{q}_{\mathrm{T}}$ is the momentum of he vector boson in the transverse plane and the transverse momentum of the hadronic recoil ($\overrightarrow{u}_{\mathrm{T}}$) is defined as the vector momentum sum of all the particles that are not the boson decay products.

The hadronic recoil momentum $\overrightarrow{u}_{\mathrm{T}}$ can be further projected onto 2 components, parallel ($u_{\parallel}$) and perpendicular ($u_{\perp}$), with respect to the boson direction. The $< u_{\parallel} > / < q_T >$ is called the $\overrightarrow{p}_{\mathrm{T}}^{\mathrm{miss}}$ response and is closely related to the jet energy corrections. The resolution instead is evaluated considering the widths of the $< u_{\parallel} >$ and $< u_{\perp} >$ distributions.

The response curves as a function of q_{T} extracted from data and simulation in $\mathrm{Z} \to \mu^+\mu^-$, $\mathrm{Z} \to e^+e^-$, and photon events is shown in Fig. 5.18, and it can be seen that there is good agreement in all the different channels: the missing transverse momentum is able to fully recover the hadronic recoil activity corresponding to a Z boson with q_{T} of 50 GeV, whereas the uncorrected, unclustered energy contributions are dominant compared to the corrected energy of the recoiling jets below 50 GeV [134].

Figure 5.19 shows the resolution of the transverse and parallel components extracted from data and simulation in the same events. The resolutions measured in the different samples are in good agreement and are found to be increasing with the q_{T} . The isotropic nature of energy fluctuations, such as detector noise and the underlying event, causes the perpendicular component of the recoil energy to have a more stable resolution compared to the parallel component.

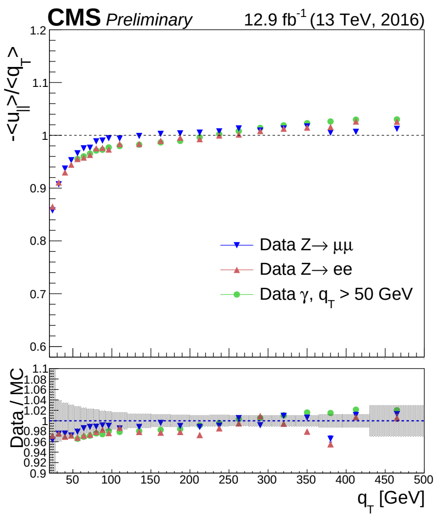


Figure 5.18: The $p_{\mathrm{T}}^{\mathrm{miss}}$ response as a function of the vector boson momentum q_{T} for different data samples (upper frame). The ratio of data to simulation with the error band displaying the systematic uncertainty of the simulations (lower frame) [134].

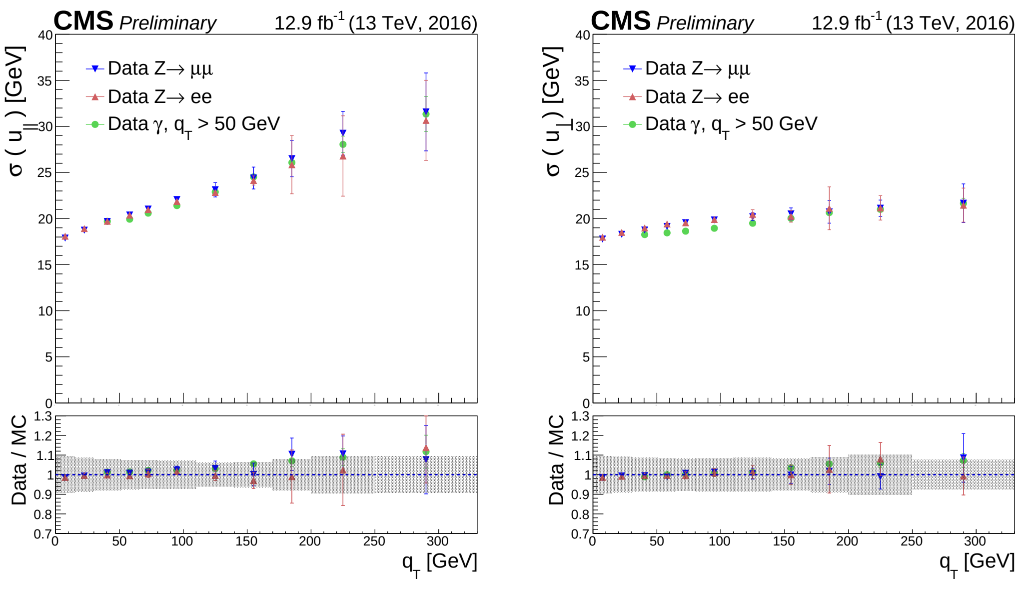
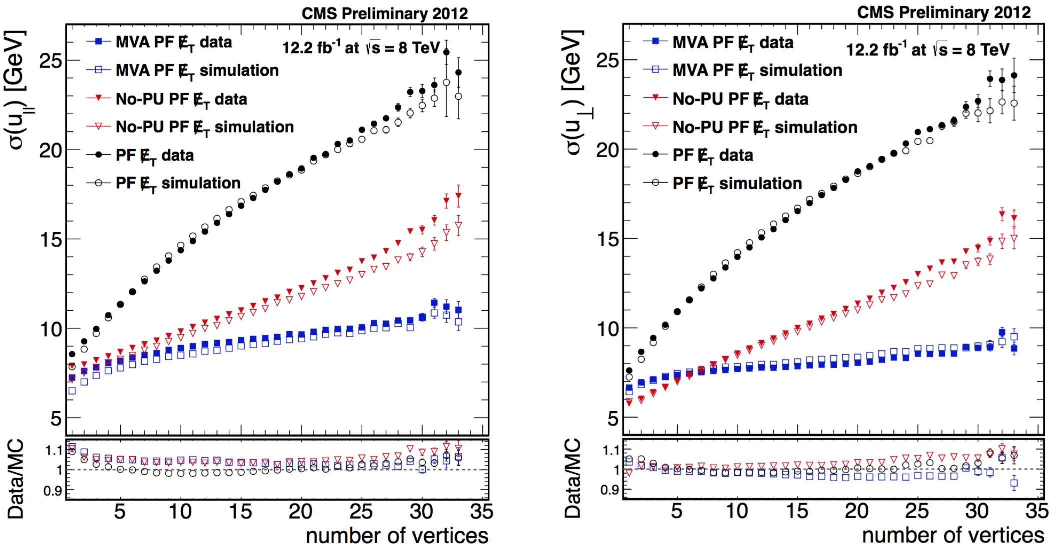


Figure 5.19: The $p_{\mathrm{T}}^{\mathrm{miss}}$ resolution in the parallel and perpendicular directions as a function of the vector boson momentum q_{T} for different data samples (upper frame). The ratio of data to simulation with the error band displaying the systematic uncertainty of the simulation (lower frame) [134].



(a) (b)

Figure 5.20: Parallel (a) and perpendicular (b) $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution as a function of the number of reconstructed vertices for $Z \to \mu^+\mu^-$ events in data (black circle) and simulation (white circles) [135]. The $E_{\mathrm{T}}^{\mathrm{miss}}$ is reconstructed with the PF algorithm (PF $\not{E}_T$ in the plot).

The $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution as a function of PU is shown in Fig. 5.20 for the 2012 data-taking period [135]. Similar performance is observed in the 2016 data, as shown in Fig. 5.21.

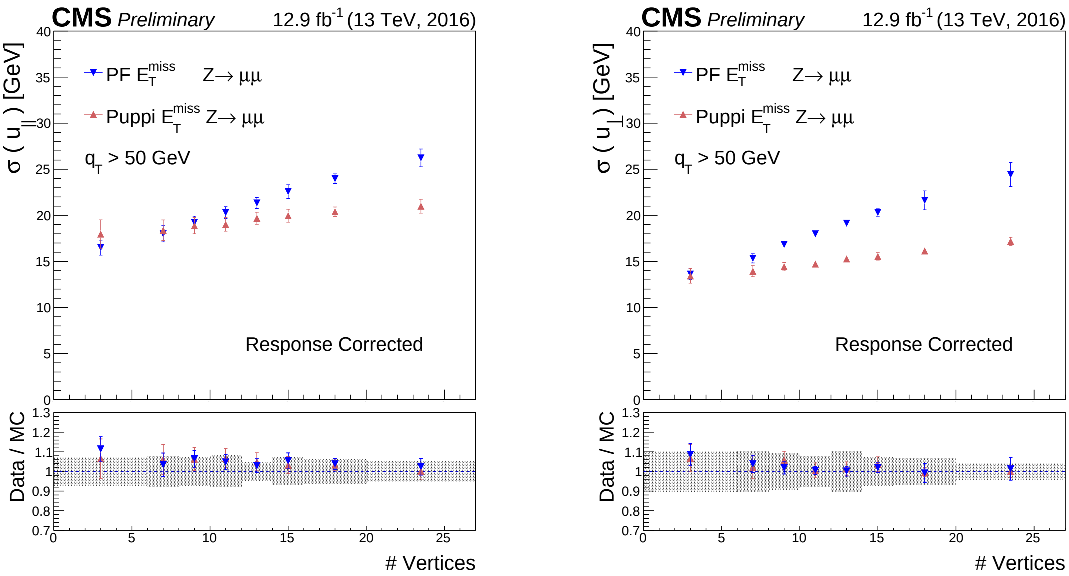
An important variable, beside the $p_{\mathrm{T}}^{\mathrm{miss}}$, is the missing transverse energy significance (S) which is defined as:

$$S = 2\ln\left(\frac{\mathcal{L}(\overrightarrow{\epsilon} = \sum \overrightarrow{\epsilon}_i)}{\mathcal{L}(\overrightarrow{\epsilon} = 0)}\right), \tag{5.11}$$

where $\overrightarrow{\epsilon}$ is the *true* $\overrightarrow{E}_T^{\mathrm{miss}}$ and $\sum \overrightarrow{\epsilon}_i$ is the *observed* $\overrightarrow{E}_T^{\mathrm{miss}}$. At the numerator, there is the likelihood that the true value of the $\overrightarrow{E}_T^{\mathrm{miss}}$ equals the observed value, while the denominator corresponds to the *null hypothesis* that the true $\overrightarrow{E}_T^{\mathrm{miss}}$ is zero. With a very good approximation, the likelihood $\mathcal{L}(\overrightarrow{\epsilon})$ has a Gaussian distribution and the significance can be written as

$$S = \left(\sum \overrightarrow{\epsilon}_i\right)^{\dagger} V^{-1} \left(\sum \overrightarrow{\epsilon}_i\right), \tag{5.12}$$

where V is the 2×2 $E_{\mathrm{T}}^{\mathrm{miss}}$ covariance matrix, that models the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution smearing in each event. It is constructed by propagating the individual resolutions of the objects used for the calculation of the $E_{\mathrm{T}}^{\mathrm{miss}}$, primarily the hadronic components of the event. Jets with $p_{\mathrm{T}} > 15$ GeV and all objects with $p_{\mathrm{T}} < 15$ GeV enter the calculation, the



(a) (b)

Figure 5.21: Parallel (a) and perpendicular (b) $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution as a function of the number of reconstructed vertices for $Z \to \mu^+\mu^-$ events in data for PF $\not{E}_T$ (blue) and MET computed in events with the PUPPI pileup removal, called PUPPI $\not{E}_{\mathrm{T}}$. The $E_{\mathrm{T}}^{\mathrm{miss}}$ is reconstructed with the PF algorithm (PF $\not{E}_T$ in the plot) [134].

former in the form:

$$V = R(\phi) \begin{bmatrix} \sigma_{p_{\mathrm{T}}}^2 & 0 \\ 0 & p_{\mathrm{T}}^2 \sigma_{\phi}^2 \end{bmatrix} R(\phi)^{-1}, \tag{5.13}$$

where $R(\phi)$ is the rotation matrix from the jet reference frame to the detector one, and $\sigma_{p_{\mathrm{T}}}$, and σ_{ϕ} are the jet momentum and angular resolution measured differentially in p_{T} and η using simulations and retuned with data-based techniques. For the objects with $p_{\mathrm{T}} < 15$ GeV, low p_{T} jets and unclustered hadronic activity, the sum of all the constituents associated with the unclustered energy is calculated:

$$\overrightarrow{p}_{\mathrm{T}} = \sum_i \overrightarrow{p}_{T,i} \sigma_{uc}^2 = \sigma_0^2 + \sigma_s^2 \sum_i |\overrightarrow{p}_{T,i}| \tag{5.14}$$

with σ_0 and σ_s determined by data-driven techniques as explained in [136]. Then their contribution to the covariance matrix is taken to be isotropic and equals

$$V_{uc} = \begin{bmatrix} \sigma_{uc}^2 & 0 \\ 0 & \sigma_{uc}^2 \end{bmatrix}. \tag{5.15}$$

Electrons and muons are assumed to have perfect resolution with respect to the hadronic component of the event and make no contribution to the covariance matrix V.

5.3.7 τ leptons

Taus are the heaviest leptons with a mass of 1776.8 ± 0.1 MeV [4]. Therefore, they play an important role in those scenarios, such as in Higgs physics or BSM physics, where the coupling of new particles is directly proportional to the mass of the fermions. They have a very small lifetime ($\sim 3 \times 10^{-13}$ s), hence they decay most of the time before reaching the innermost layer of the detector and can be reconstructed just through they decay products. Taus can decay either leptonically to an electron or a muon, plus neutrinos, or hadronically to a combination of charged and neutral hadrons, most commonly pions. The branching ratios are reported in Table 5.4.

Decay mode	Meson resonance (MeV)	$\mathcal{B}$[%]
$\tau^- \to e^-\bar{\nu}_e\nu_\tau$		17.8
$\tau^- \to \mu^-\bar{\nu}_\mu\nu_\tau$		17.4
$\tau^- \to h^-\nu_\tau$		11.5
$\tau^- \to h^-\pi^0\nu_\tau$	$\rho(770)$	26.0
$\tau^- \to h^-\pi^0\pi^0\nu_\tau$	$a_1(1260)$	9.5
$\tau^- \to h^-h^+h^-\nu_\tau$	$a_1(1260)$	9.8
$\tau^- \to h^-h^+h^-\pi^0\nu_\tau$		4.8
Other hadronic decays		3.2
All hadronic decays		64.8

Table 5.4: Approximate branching ratios ($\mathcal{B}$) of different τ decay modes. Pions and kaons are listed as generic hadrons (h). Charge conjugation invariance is assumed [137].

Leptons from the leptonic decays are reconstructed through the standard electron and muon reconstruction, while the hadronic tau lepton decays τ_h are reconstructed and identified by the *hadron-plus-strip* (HPS) algorithm [137]. The major challenge of the algorithm is to distinguish between genuine taus and quark or gluon jets from the copious QCD multijet production.

The τ_h reconstruction performed by the HPS algorithm is described in the next section. The basic features of the algorithm are identical between Run 1 [137] and Run 2, except for the improvement in the strip reconstruction that was implemented for the Run 2 analysis [138]. The main identification criteria will also be presented.

5.3.7.1 Hadron-plus-strip algorithm

Starting from the PF constituents of a reconstructed jet (AK4 jets for Run 2 and AK5 jets for Run 1), the HPS algorithm aims at reconstructing the hadronic decays of the tau leptons, with the different charged and neutral hadron combinations. The neutral pions decay promptly into a pair of photons, which have a high probability to undergo conversion to an electron and a positron while they traverse and interact with the detector. These pairs separate in the high magnetic fields of CMS, giving rise to deposits in the ECAL calorimeter. In order to reconstruct the full neutral energy release by the photon and electron candidates, the ECAL energy deposits are clustered in “strips" of rectangular shape elongated in ϕ direction and narrow in η direction.

Charged candidates (“prongs”) used for the tau reconstruction are required to have $p_{\mathrm{T}} > 0.5$ GeV and to be compatible with the primary vertex of the event by applying a loose criterion on the transverse inverse parameter $d_{xy} < 0.1$ cm, in order not to reject genuine taus with long decay length.

Given a set of strips reconstructed for a jet, using the charged constituents of the jet, the HPS algorithm produces all combinations possible for the following decay modes: $h^{\pm}$, $h^{\pm}\pi^{0}$, $h^{\pm}h^{\mp}h^{\pm}(+\pi^{0})$. The invariant mass of the constituents has to be compatible with the intermediate resonances typical of the tau hadronic decay $\rho(770)$ and $a_1(1260)$, for the $h^{\pm}\pi^{0}$ and the $h^{\pm}\pi^{0}\pi^{0}$ or $h^{\pm}h^{\mp}h^{\pm}$ decay modes, respectively.

The algorithm proceeds from charged candidates and defines signal cones of radius $R_{\mathrm{sig}} = 3.0\ \mathrm{GeV}/p_{\mathrm{T}}$, where the p_{T} is the transverse momentum of the hadronic system, with a cone radius between 0.05 and 0.10. In case of multiple tau hypotheses for the same jet, the one with the largest p_{T} is selected, resulting in a single τ_{h} reconstructed per jet.

The main change to the HPS algorithm introduced in Run 2 is in regards to the strip reconstruction. Multiple photons coming from tau decays, such as $h^{\pm}\pi^{0}\pi^{0}$ are reconstructed from the Run 2 dynamic strip reconstruction into one bigger strip, thus these events are accounted for in the $h^{\pm}\pi^{0}$ category.

5.3.7.2 Dynamic strip reconstruction

Photon as well as electron constituents of the jets that seed the τ reconstruction are clustered into strip in the $\eta - \phi$ plane, which are used to collect all ECAL energy deposits from neutral pions produced in the hadronic tau decay. The size of the strip used in the reconstruction is set to a fixed value of 0.20×0.05 in the algorithm used for Run 1 analyses [137]. Since electrons and positrons from low-p_{T} photons can bend substantially in the magnetic field or scatter against the detector material and end outside the fixed size strip, an incomplete cluster could be used in the tau reconstruction, meaning that other deposits would be instead accounted for in the isolation computation, thus spoiling the identification of a genuine tau. Conversely, if the tau lepton is highly energetic, the decay products tend to be very collimated and in this case a smaller strip size helps in reducing the background contributions to the strip.

In Run 2 an improved version of the algorithm is deployed to solve this problem, by taking into account the possibility of enlarging (reducing) the size of the strip depending on the lower (higher) momentum associated with it.

The clustering of electrons and photons into strips is an iterative procedure. The highest p_{T} electron or photon not yet included into any strip is used as a seed for a new strip. The initial position of the strip is set to the η and ϕ of the seed e/γ.

The next most energetic e/γ deposit within:

$$\Delta\eta = f(p_{\mathrm{T}}^{e/\gamma}) \,\text{with}\, f = 0.20 * p_{\mathrm{T}}^{-0.66} \tag{5.16}$$

and

$$\Delta\phi = g(p_{\mathrm{T}}^{e/\gamma}) \,\text{with}\, g = 0.35 * p_{\mathrm{T}}^{-0.71} \tag{5.17}$$

is merged into the strip. The dimensionless functions f and g are determined from simulations of single τ lepton generated with uniform p_{T} in the range from 20 to 400 GeV, such that 95% of the e/γ candidates that arise from τ_h decays are contained within one strip. The upper limits on the strip size are set to 0.15 in $\Delta\eta$ and 0.3 in $\Delta\phi$, while the lower limit is 0.05 in both directions.

The position of the strip is recomputed as the energy-weighted average of the initial deposits

$$\eta_{\mathrm{strip}} = \frac{1}{p_{\mathrm{T}}^{\mathrm{strip}}} \sum p_{\mathrm{T}}^{e/\gamma} * \eta^{e/\gamma} \tag{5.18}$$

and

$$\phi_{\mathrm{strip}} = \frac{1}{p_{\mathrm{T}}^{\mathrm{strip}}} \sum p_{\mathrm{T}}^{e/\gamma} * \phi^{e/\gamma}, \tag{5.19}$$

and the transverse momentum is recomputed as $p_{\mathrm{T}}^{strip} = \sum p_{\mathrm{T}}^{e/\gamma}$.

If no further e/γ candidate is found within these boundaries, the procedure ends for the given strip and goes on to the other e/γ candidates associated with the initial jet, in order to start the reconstruction of a new strip.

The charged candidates are combined with the strips to form a τ decay hypothesis. The compatibility of a given combination with a genuine tau decay is checked by reconstructing the mass of the visible hadronic constituents and requiring it to correspond to either a $\rho(770)$ or an $a_1(1260)$ hadron. The size and the position of the mass window are adjusted in order to take into account effects due to the energy of the strip [138].

5.3.7.3 Relaxed τ_h reconstruction

A relaxed τ_h decay mode has been included in the reconstruction specifically for events with high-p_T taus that might decay to $\tau \to h^\pm h^\mp h^\pm \nu_\tau$ or $\tau \to h^\pm h^\mp h^\pm \pi^0 \nu_\tau$. In fact for a very boosted τ lepton the probability to fail the reconstruction of one of the charged hadrons increases due to the possible tracks overlapping or missing hits in the tracker detector. The criteria for the identification is relaxed from three charged hadrons to two, and the charge of the most energetic track is assigned to the reconstructed hadronic tau, with an incorrect charge assignment of $\sim 20\%$, while for the cases with one or three prongs, the incorrect charge assignment is $< 1\%$ [139].

5.3.7.4 τ_h identification

The primary handle on reducing the misidentification probability for a jet to fake a tau is the isolation requirement. Two types of isolation discriminants are deployed in CMS: a cut-based one and one based on a multi-variate analysis (MVA).

The cut-based isolation is computed from the PF candidates inside an isolation cone of radius $\Delta R = 0.5$, with the expression:

$$I_{\tau_h} = \sum p_\mathrm{T}^{charged}(d_z < 0.2\mathrm{cm}) + \max(0, \sum p_\mathrm{T}^{\gamma} - \Delta\beta \sum p_\mathrm{T}^{charged}(d_z > 0.2\mathrm{cm})). \quad (5.20)$$

Charged candidates and photons with $p_\mathrm{T} > 0.5$ GeV, which are not tau constituents, but are in the isolation cone, are taken into account. PU contributions are removed by requiring the charged candidates to originate from the hadronic tau production vertex ($d_z < 0.2$ cm). The PU contribution to the photon energy in the isolation cone is estimated from the charged particles within a cone of $R = 0.8$ not compatible with the tau production vertex ($d_z > 0.2$ cm), properly rescaled with the $\Delta\beta$ factor. This factor represents the ratio between the energy carried by the neutral and charged particles in the inelastic collision, with a correction for the different cone sizes used in the isolation. For Run 1 analyses, this empiric factor has a value of 0.46 [137], while in Run 2 a $\Delta\beta = 0.2$ is used. This value can be seen as an approximation of the neutral to charge hadron production rate (0.5), corrected for the difference of the isolation cone size and the cone size used for the charged PU component : $0.5 * (0.5/0.8)^2 \sim 0.195$. In Run 2 the further requirement on the fraction of the energy carried by the photons used for the hadronic tau reconstruction, but located in strips outside of the signal cone ($p_{\mathrm{T}}^{strip,outer}$), to be less than 10% helps in further reducing by 20% the jet probability of being misidentified as an hadronic tau.

The quantities related to the isolation deposits are further combined with variables related to the non-negligible τ lifetime information in order to provide the best possible discriminator. A Boosted Decision Tree (BDT) is trained using variables such as:

- multiplicity of electron and photon candidates in the signal and isolation cones;
- differential $p_{\mathrm{T}}^{strip,outer}$ and $p_{\mathrm{T}}-$ weighted angular distance (ΔR, $\Delta\eta$, $\Delta\phi$) distributions of the photon and electrons strip deposits inside or outside the signal cone;
- τ lifetime related variables such as the leading track transverse and 3D impact parameters in the case of 1 prong decays and the secondary vertex information in the 3 prong case, and their respective significances.

Different working points are defined to have isolation efficiencies between 40% and 90% in steps of 10%, relative to the reconstructed tau candidates, with misidentification probabilities smaller than $\mathcal{O}(10^{-2})$. In the analyses presented here, the MVA isolation is used.

Electrons and muons can also be misidentified as hadronic tau leptons decays. Electrons can mimic the one-prong τ_{h} decay, since they have a charged track and can emit additional bremsstrahlung radiation that can be misidentified as π^0s. A BDT-based discriminator trained with ECAL and HCAL cluster distributions is used: typically the requirements are 75% efficient, with a misidentification rate of 10^{-2}–10^{-3}. Discriminators for muon misidentification have also an efficiency of 95%–100%, with a misidentification rate of 10^{-3} –10^{-4}: τ_{h} candidates with matching segments in the muon detectors are rejected.

5.3.7.5 Energetic di- τ pairs

One of the main features of this thesis work is the development, commissioning and validation of a new technique for the reconstruction of boosted tau pairs such as the ones that might be originating from the decay of highly energetic Z or Higgs bosons. In such cases, the final-state τ leptons can be emitted very close to each other and therefore be reconstructed in a single highly energetic jet. As shown in Fig. 6.4, the ΔR between the two τ leptons shifts towards smaller values with the increasing of the resonance mass and, thus, the momentum of the di-τ system. The standard tau reconstruction for this topology has a poor performance, since it is designed to reconstruct one tau candidate per jet. Moreover, particles originating from the neighboring tau can enter in the jet around the other tau, thus spoiling its reconstruction. These features mean that the usual τ reconstruction and lepton identification techniques need to be modified in order to cope correctly with the PF products coming from the nearby τ or lepton.

Cleaning technique for energetic tau lepton pairs

In Run 1, two different approaches were developed in order to reconstruct tau pairs in semileptonic $\mu\tau_{\mathrm{h}}$ and $e\tau_{\mathrm{h}}$, and fully hadronic $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ events.

The first strategy to avoid these issues is to clean τ decay products from the spurious lepton. Since the PFTau collection is constructed starting from the AK5 PFJet collection, a new jet collection (*cleaned jet*) is defined: the electrons and muons that satisfy minimal identification criteria are removed from the jet constituents [106]. Such *cleaned jets* are then used as seeds to the HPS algorithm and the standard tau isolation criteria are applied.

The final set of requirements for the selection of the τ_{h} is: $p_{\mathrm{T}} > 20$ GeV, and $|\eta| < 2.3$. Identification criteria to reject muon and electron faking τ_{h} also applied together with the isolation requirement.

The global τ_{h} selection efficiency after requesting all the chosen criteria listed above (p_{T}, η and discriminants) is shown in figure 5.22. For a $\mathrm{Z}' \to \mathrm{ZH} \to \mathrm{q\bar{q}}\tau^+\tau^-$ signal of mass of 2500 GeV, the efficiency of the tau reconstruction is shown as a function of the p_{T} of the τ_{h} and varies from $\sim 50\%$ to $\sim 80\%$ in the $\mu\tau_{\mathrm{h}}$ final state and from $\sim 40\%$ to $\sim 80\%$ in the $\mathrm{e}\tau_{\mathrm{h}}$ final state, when using the cleaned τ reconstruction instead of the standard one. Looking at Fig.5.22, even if the recovery in efficiency is significant for both the channels, in the $\mathrm{e}\tau_{\mathrm{h}}$ we observe a recovery slightly worse than the $\mu\tau_{\mathrm{h}}$ channel, because while the muon has a clean detector signature, the electron is a more complex object (due to the emission of photons caused by bremsstrahlung). This results in a slightly worse recovery for the case of a nearby electron.

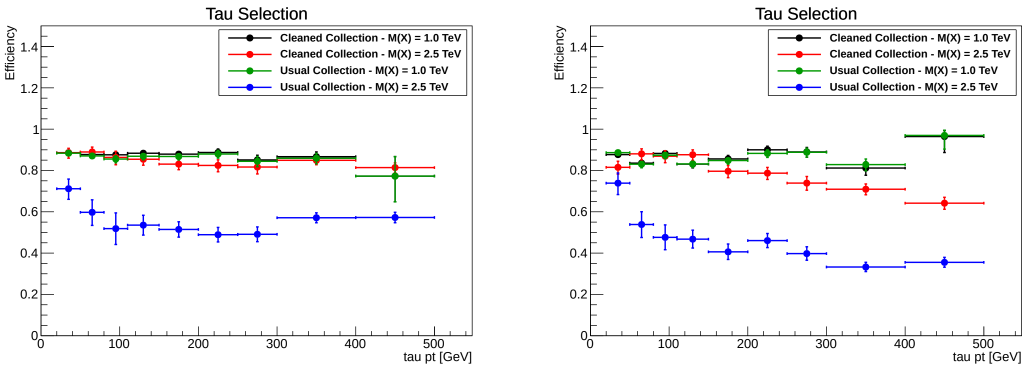


Figure 5.22: Global τ selection efficiency for the $\mu\tau_h$ channel (left) and $e\tau_h$ channel (right).

Boosted technique for energetic tau lepton pairs

For the fully hadronic final state of a τ lepton pair, the so-called "boosted" reconstruction was developed in Run 1. For the boosted reconstruction, large cone CA8 jets are used as inputs and a subjet searching technique is applied: the last step of the jet clustering is undone to identify the two parent subjets of the final jet, ordered by mass. The parent subjets are expected to coincide with the two τ leptons. To reduce possible misidentification both subjets are required to have $p_T > 10$ GeV and the mass of the heaviest subjet to be less than 2/3 of the mass of the original jet, in order to avoid cases in which one of the subjets is just originating from the soft emission of the other one. If the two subjets are selected, they are used as inputs to the HPS algorithm. If the pair is discarded, the subjet finding procedure is repeated for the heaviest subjet that is then split into two subjets. If no subjets are found within a given large cone jet, no tau reconstruction is performed for it and the subjet algorithm proceeds to the declustering of the other CA8 jets in the event. In case of leptonic tau decay, the subjet finding algorithm can reconstruct the lepton as a subjet, but at the analysis level, standard lepton identification criteria are applied to identify it as a real electron or muon. After the HPS reconstruction, the MVA-based isolation discriminators are applied to the τ_h candidate, taking into account in the isolation computation just the PF candidates that belong to the area of the subjet seed, instead of the usual cone of $R = 0.5$, in order to reduce the jet to tau misidentification probability without suffering from the proximity of the second tau decay in the event. The decay mode criteria are relaxed and tau candidates with two charged hadrons are accepted, in order to recover events with tracking inefficiency due to the dense environment of a high-p_{T} jet.

Between Run 1 and Run 2 the "boosted" reconstruction algorithm was further developed and adopted also for the semileptonic final states, in order to simplify and unify the approach in different channels, to the similar performance in the cleaned and boosted reconstruction [140].

In Figure 5.23, the efficiency of the standard and boosted τ_{h} reconstruction are compared for τ_{h} in simulated events of $X \to \mathrm{HH} \to \mathrm{b\bar{b}}\tau^{+}\tau^{-}$ decays, for a resonance mass of 2.5 TeV. The two reconstruction exhibit similar performance for various decay modes (DMs) in $\mu\tau_{\mathrm{h}}$ events, but show better efficiency for the boosted reconstruction in $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ events.

In Figure 5.24, the efficiencies of the standard reconstruction and isolation identification of highly-boosted τ lepton pairs in simulated events of $X \to \mathrm{HH} \to b\bar{b}\tau\tau$ decays in $\tau_e\tau_{\mathrm{h}}$, $\tau_\mu\tau_{\mathrm{h}}$, and $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ final states are shown. While the efficiency in $\ell\tau_{\mathrm{h}}$ final states is computed only for the τ_{h} candidate, in $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ final states it is computed for one of the τ_{h} candidates, by considering how often a reconstructed τ_{h} is found in a cone of 0.1 around each of the generated τ_{h}, independently on the other τ_{h}, and once relative to the τ_{h} candidate pair, by requiring both generated τ_{h} to be matched to reconstructed τ_{h}.

Furthermore, the expected probability for broad jets to be misidentified as τ_{h} pairs is shown for events of simulated multijet production. The τ_{h} candidates are selected by requiring $p_{\mathrm{T}} > 20$ GeV, $|\eta| < 2.3$, and the very loose WP of the MVA-based isolation.

Muons and electrons are identified with loose identification and isolation requirements as described in Section 5.3.1 - 5.3.2 . PF candidates originating from the τ_{h} decay are not taken into account for the computation of the lepton isolation.

The increase in the efficiency in the reconstruction is substantial for bosons of transverse momentum greater than 0.5 TeV, with the dedicated τ_{h} reconstruction compared to the standard one. The increase in the misidentification probability is sustainable because of the high signal purity in the phase space of the searches with boosted di-τ pairs.

In Fig. 5.25, the misidentification probabilities of the standard and the boosted reconstruction of taus are compared in simulated background events as a function of the distance between the subjets of the wide jet. QCD multijet events are used for the fully hadronic channel (left), while for the semileptonic channel a simulated sample of top-quark pair production is used in two ways: inclusively (center) and by just selecting fully hadronic $\mathrm{t\bar{t}}$ events without generated leptons (right). In the first case the misidentification probability includes also a component due to leptons faking taus,

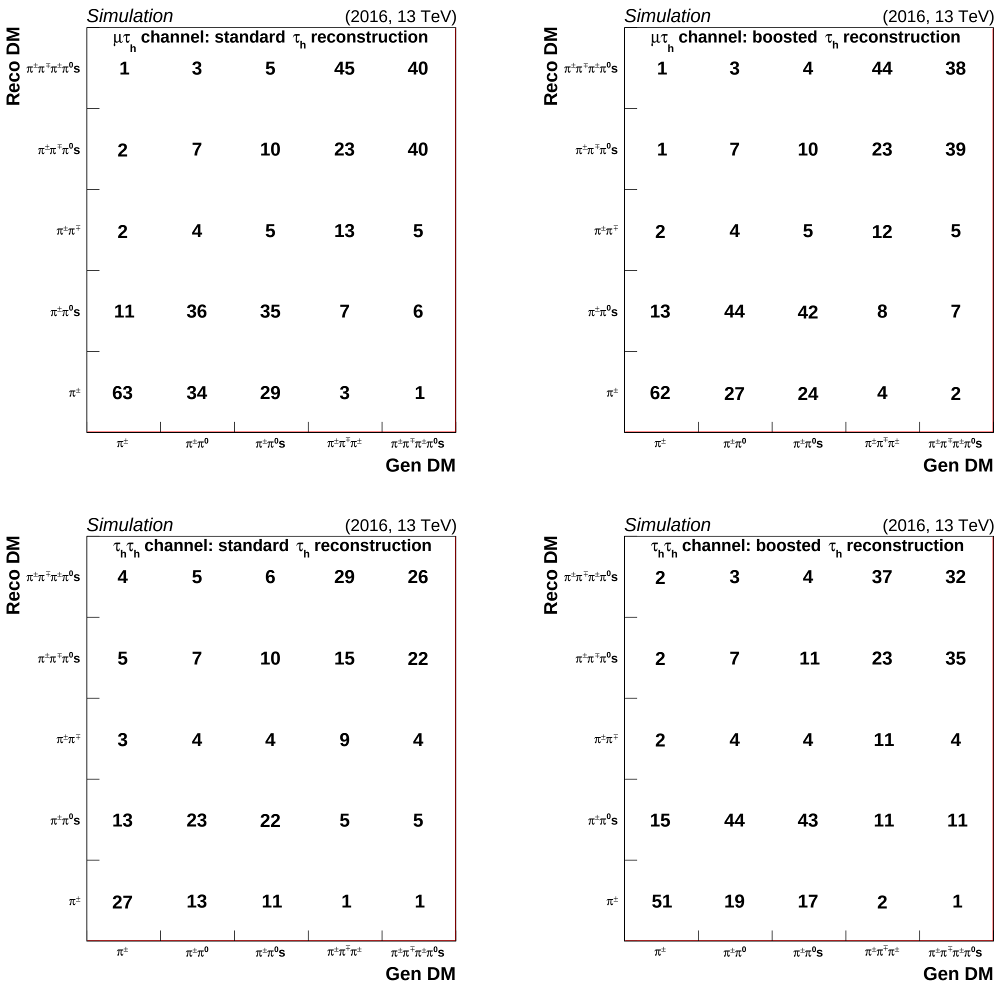


Figure 5.23: Decay mode reconstruction efficiency and migration for the standard (left) and boosted (right) τ_h reconstructions in simulated $\mu\tau_\mathrm{h}$ (top) and $\tau_\mathrm{h}\tau_\mathrm{h}$ events of $X \to \mathrm{HH} \to b\bar{b}\tau\tau$ decays, for a resonace mass of 2.5 TeV. .

while in the latter the main contribution is due to quark-originated jets faking taus. It can be seen especially for the fully hadronic $\tau_\mathrm{h}\tau_\mathrm{h}$ channel that the standard reconstruction efficiency is smaller that the boosted one for $\Delta R(sj1, sj2) < 0.4$, while they are comparable for higher subjet distances.

The performance of the boosted di-tau pair reconstruction is checked in 2016 data appropriately selected to have a pure sample of tau lepton pairs from boosted Z bosons and the scale factors are found to be compatible with unity within the uncertainties, proving that the reconstruction in data is well-modeled by simulations.

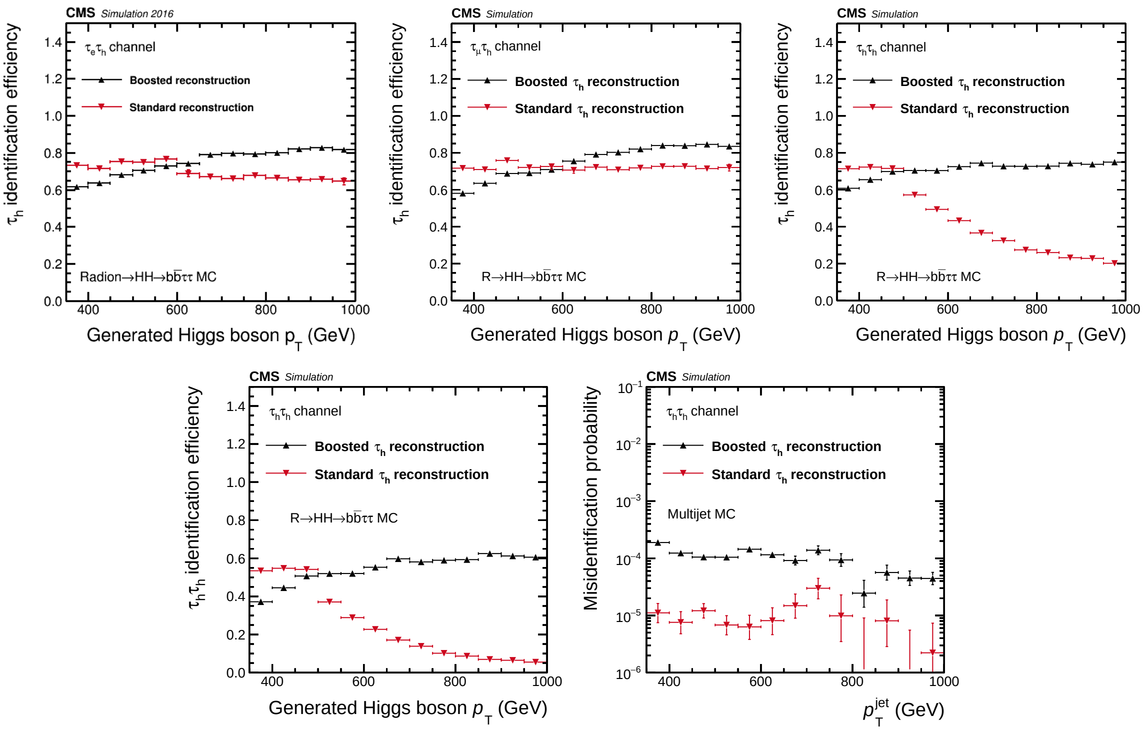


Figure 5.24: Reconstruction and identification efficiency of the τ_h in $\tau_e\tau_\mathrm{h}$ (top left), $\tau_\mu\tau_\mathrm{h}$ (top center), and $\tau_\mathrm{h}\tau_\mathrm{h}$ (top right) final states, and of the $\tau_\mathrm{h}\tau_\mathrm{h}$ system in $\tau_\mathrm{h}\tau_\mathrm{h}$ final states (bottom left), as a function of the generated transverse momentum of the Higgs boson, as well as the probability for broad jets in multijets events to be misidentified as $\tau_\mathrm{h}\tau_\mathrm{h}$ final states (bottom right), as a function of the broad jet p_T [138].

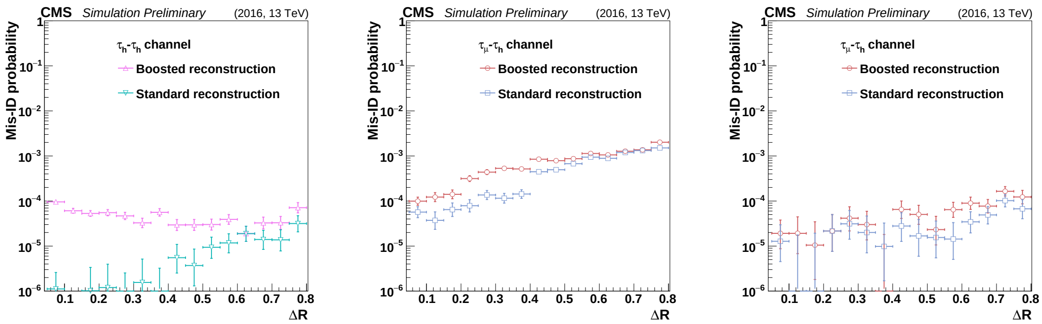


Figure 5.25: Misidentification probability of the di-τ system reconstruction as a function of the distance between the subjets of the AK8 jets in the $\tau_h\tau_h$(left) and in the $\mu\tau_h$ (center and right) channels. For the $\tau_h\tau_h$ channel QCD multijets events are used, while for the semileptonic $\mu\tau_h$ case, simulated events are used for top-quark pair production either inclusively (center) or by selecting just the fully hadronic $t\bar{t}$ final state (right).

5.3.8 Di- τ system kinematic reconstruction

The determination of the kinematic properties of the di-tau initial system is challenging because in every τ lepton decay one or more neutrinos are produced. A choice can be to just use the visible particles in the final decay to reconstruct the system. In this case, the visible mass, m_{vis}, of the di-τ system is defined as the invariant mass of all detectable products of the two τ decays.

Another possibility is to use for the reconstruction of the di-τ system the SVFIT algorithm [141–143], which combines the $\vec{p}_T^{\,\text{miss}}$ and the covariance E_T^{miss} matrix with the visible momenta of the tau candidate to calculate a more precise estimator of the kinematics of the parent boson.

In Figs. 5.26 - 5.27 the SVFIT-reconstructed Higgs boson mass is compared to the *visible mass* of the di-τ system. The SVFIT algorithm shows a better capability to reconstruct the original Higgs boson mass. The resolution on SVFIT mass is best for a W′ signal, while for Z′, radion and graviton signals the distribution are wider because of the SVFIT algorithm assumption that the p_T^{miss} in the event comes entirely from the neutrinos in the tau decays, while in events with hadronic Z and H decays, neutrinos can be produced from leptonic decays of B hadrons.

In the same way, the whole four-vector of the di-τ system can be estimated. In Fig. 5.28, the Higgs p_T spectra, in simulated bulk radion events with a center of mass energy of 8 TeV, are shown as determined by the SVFIT algorithm, after the event selection of leptons, τ_h and jets. The lower threshold of 200 GeV is imposed by the kinematics of the rest of the event.

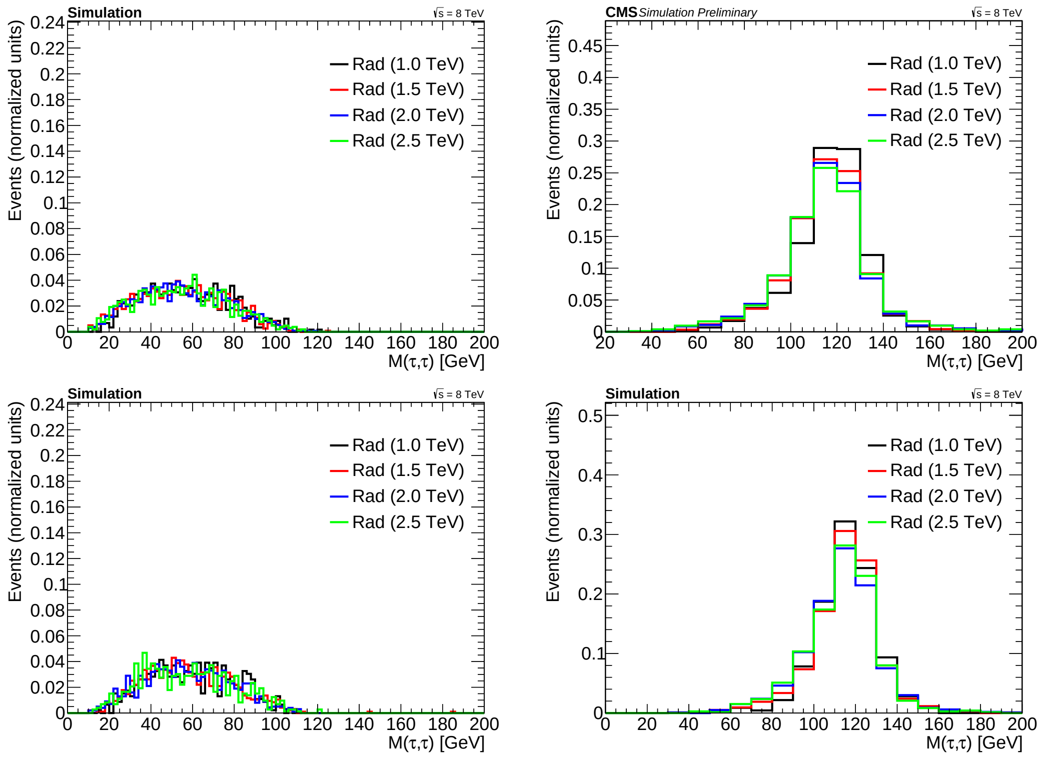


Figure 5.26: Methods for the reconstruction of the ditau system, visible mass(left) and SVFit masses (right), in simulated events for four different radion masses in the $\mu\tau_{\mathrm{h}}$ (upper plots) and $e\tau_{\mathrm{h}}$ (lower plots) channels.

In Fig. 5.29 we report the Higgs p_{T} as determined by the SVFIT, after pre-selection of muons, electrons, τ and jets, for the W′, Z′, graviton, and radion signals at a center of mass energy of 13 TeV.

The resonance mass is then reconstructed from the sum of the four-momentum of the wide-cone jet and the di-τ system four-momentum, estimated with SVFIT. The signal resonance mass shapes obtained with the visible and SVFIT procedures are shown in Fig. 5.30, in simulated bulk radion events with a center of mass energy of 8 TeV.

The signal shapes obtained from this procedure are shown in Fig. 5.31 for the W′, Z′, and radion signals at a center of mass energy of 13 TeV.

Typical resolution on the di-τ system mass reconstructed by the SVFIT algorithm varies from 10% to 14%, for resonance masses between 1 TeV and 2.5 TeV, while the resolution on reconstructed resonance mass is stable between 6% and 7%.

The algorithm performs similarly in data and simulations, as it is shown in Fig. 5.32, with the events collected in 2012 pp collisions at $\sqrt{s} = 8\,\mathrm{TeV}$ that satisfy the requirements defined in Sec.6.3. The SVFIT algorithm di-τ system kinematic reconstruction

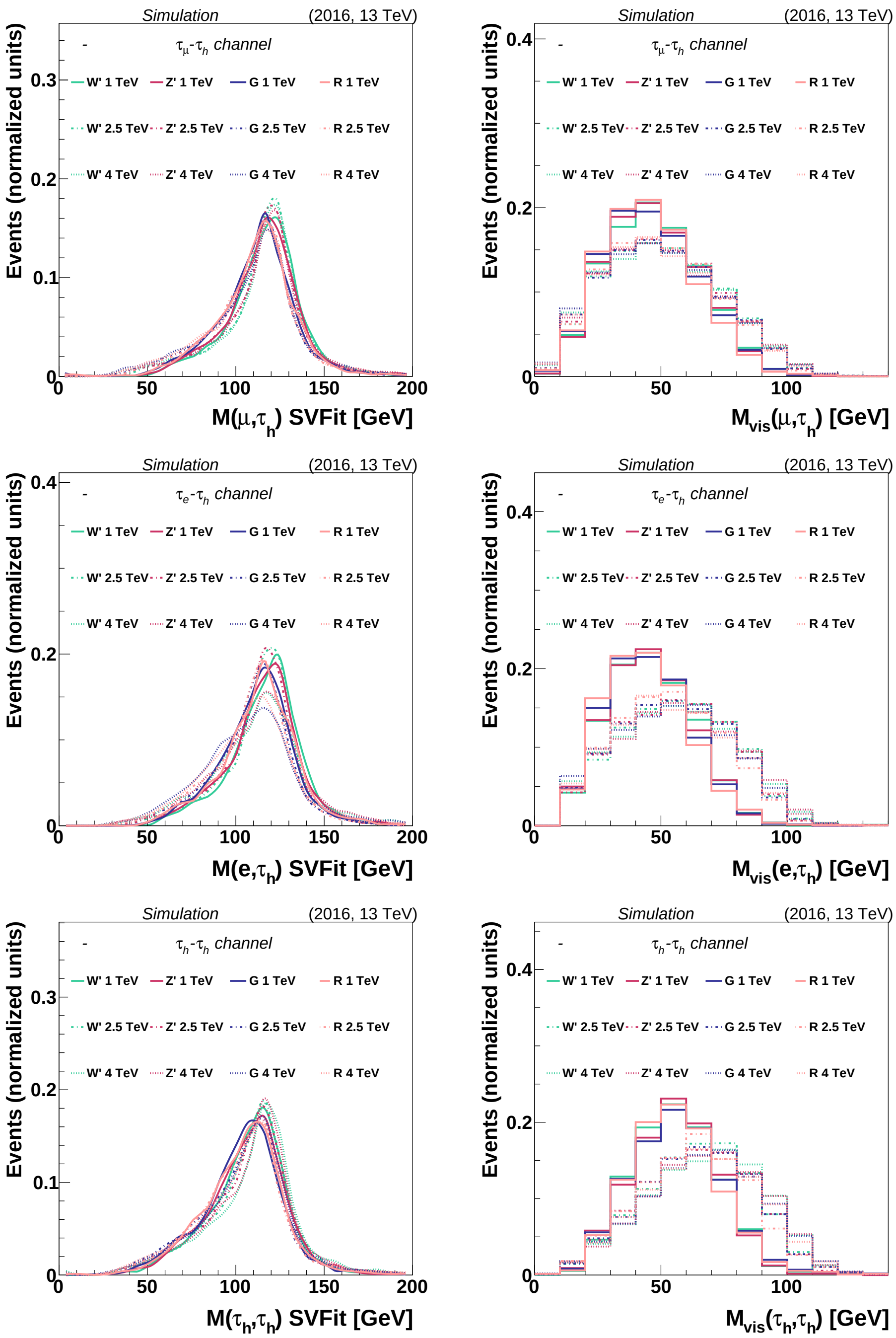


Figure 5.27: Methods for the reconstruction of the ditau system, visible mass and SVFIT mass, for four signal samples in the $\mu\tau_\mathrm{h}$ (upper plots), $e\tau_\mathrm{h}$ (middle plots) channels, $\tau_\mathrm{h}\tau_\mathrm{h}$ (middle plots) channels.

is also reported in Figs.5.33 –5.35 for 13 TeV data.

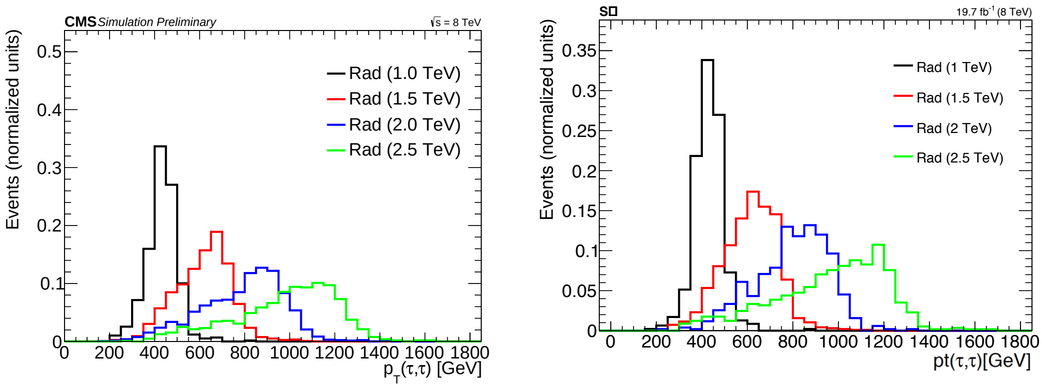


Figure 5.28: Reconstructed p_{T} of the Higgs boson for HH signal MC in the $\mu\tau_{\mathrm{h}}$ (left) and $e\tau_{\mathrm{h}}$ (right) events at $\sqrt{s} = 8\,\mathrm{TeV}$.

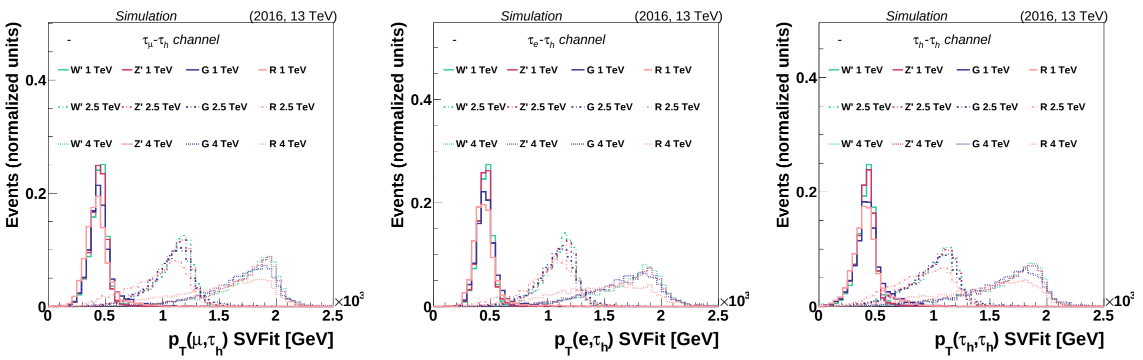


Figure 5.29: Reconstructed p_{T} of the Higgs boson for HH signal MC for H H signal in the $\mu\tau_{\mathrm{h}}$ (left), $e\tau_{\mathrm{h}}$ (center), and $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ (right) events at $\sqrt{s} = 13\,\mathrm{TeV}$.

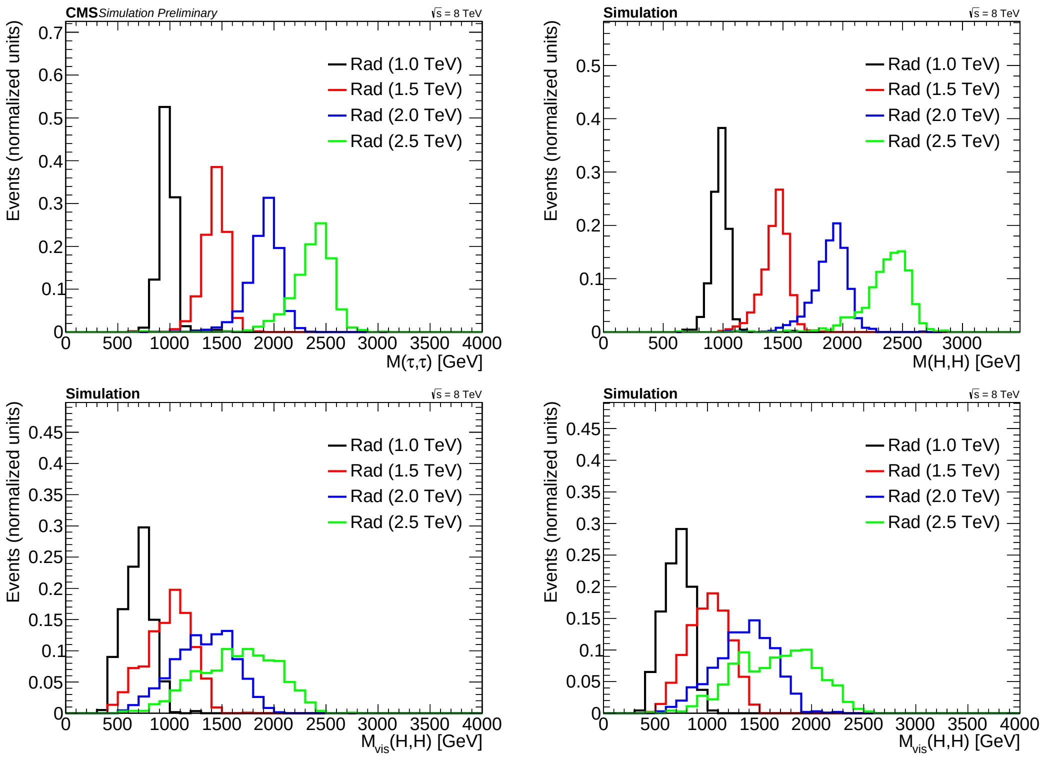


Figure 5.30: Invariant mass of the reconstructed diboson system for signal of different masses in the $\mu\tau_\mathrm{h}$ (left) and $e\tau_\mathrm{h}$ (right) channels using the SVFIT algorithm (top) and throught just the visible products (bottom).

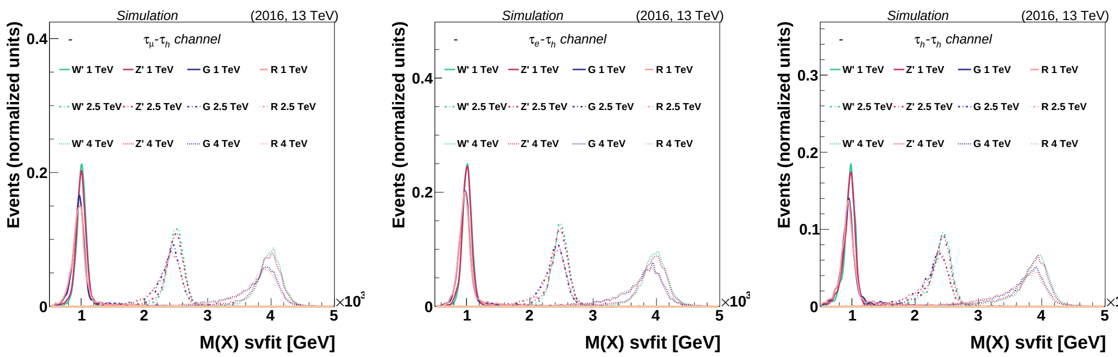


Figure 5.31: Reconstructed resonance mass for signal of different masses in the $\mu\tau_\mathrm{h}$ (left), $e\tau_\mathrm{h}$ (center) and $\tau_\mathrm{h}\tau_\mathrm{h}$ (right) channels.

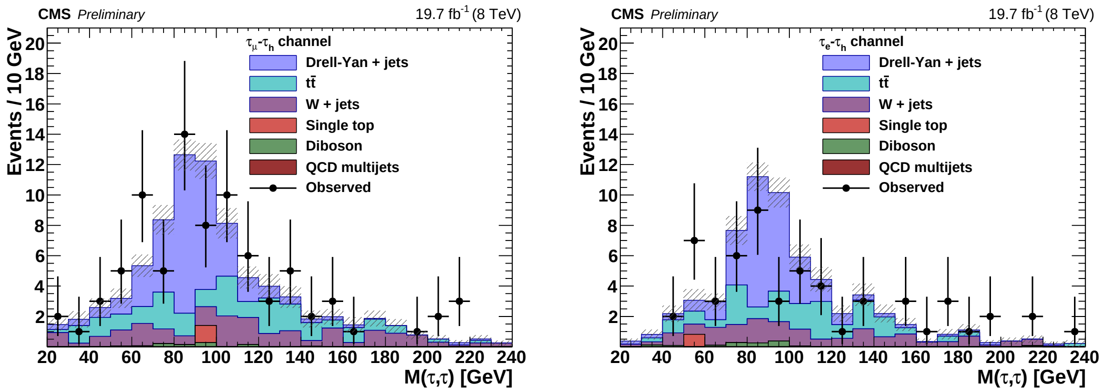


Figure 5.32: The number of background and data events in the di-τ mass spectrum as reconstructed bythe SVFIT algorithm in the $\mu\tau_{\mathrm{h}}$ (left) and $e\tau_{\mathrm{h}}$ (right) channels. The complete signal selection, except for the pruned mass window and the Higgs b-tagging requirements, is applied.

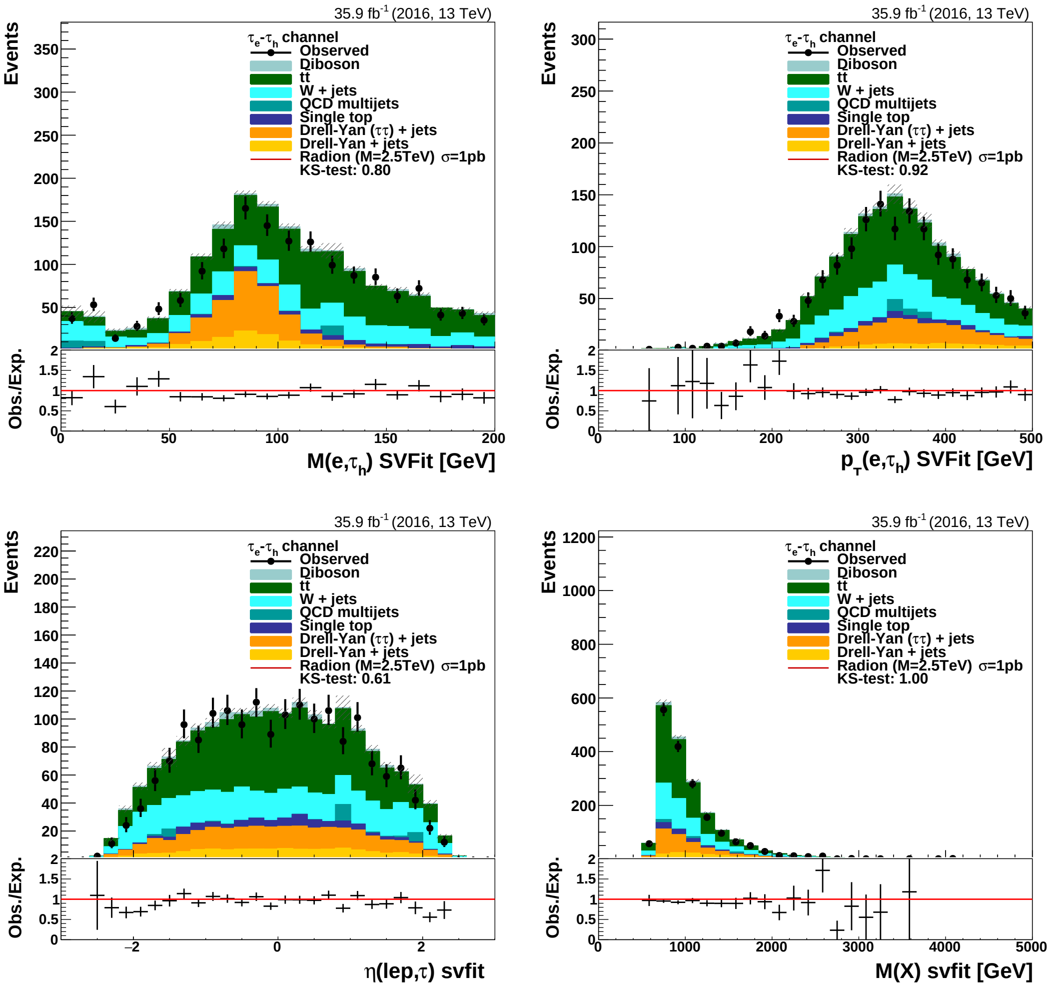


Figure 5.33: Comparison between data and expected simulated events for the $e\tau_{\mathrm{h}}$ channel for the following variables: di-τ mass and p_{T} (top), and di-τ η and resonance mass (bottom).

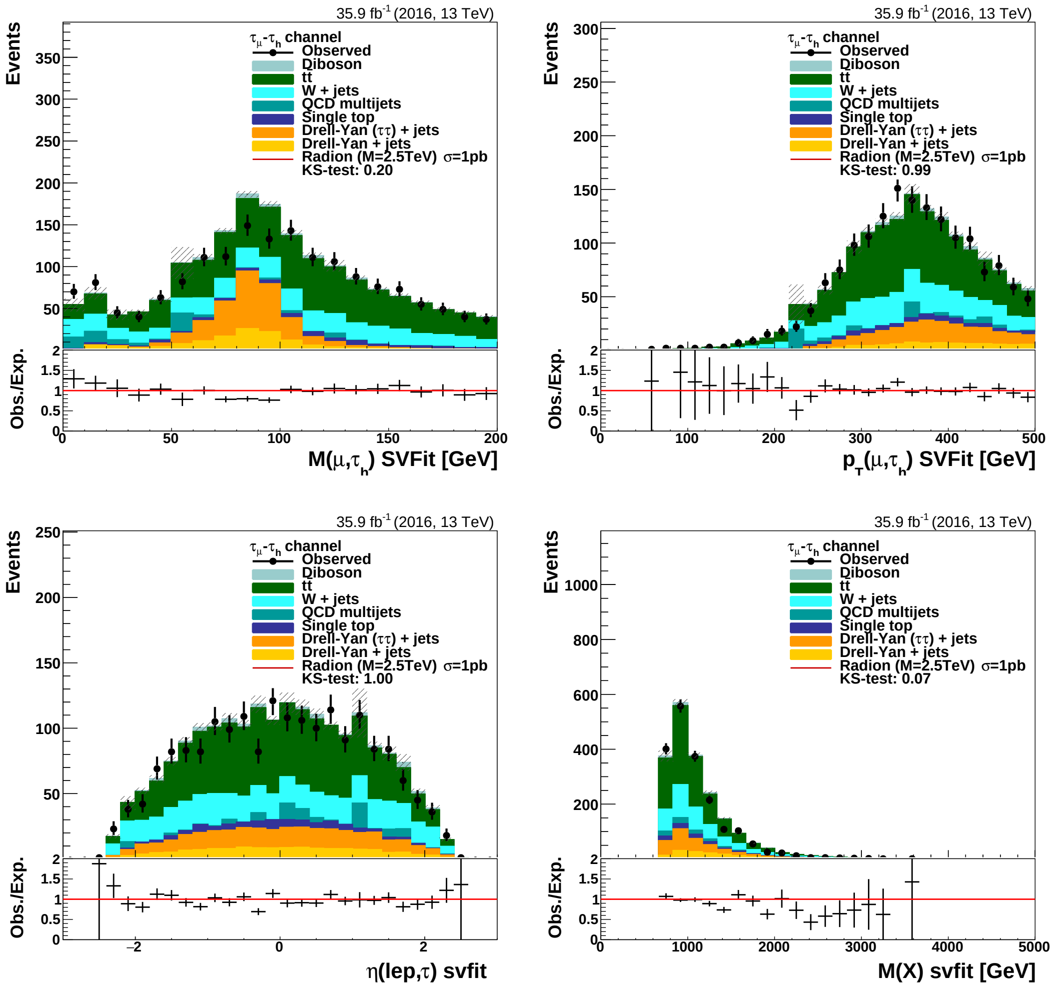


Figure 5.34: Comparison between data and expected simulated events for the $\mu\tau_{\mathrm{h}}$ channel for the following variables: di-τ mass and p_{T} (top), and di-τ η and resonance mass (bottom).

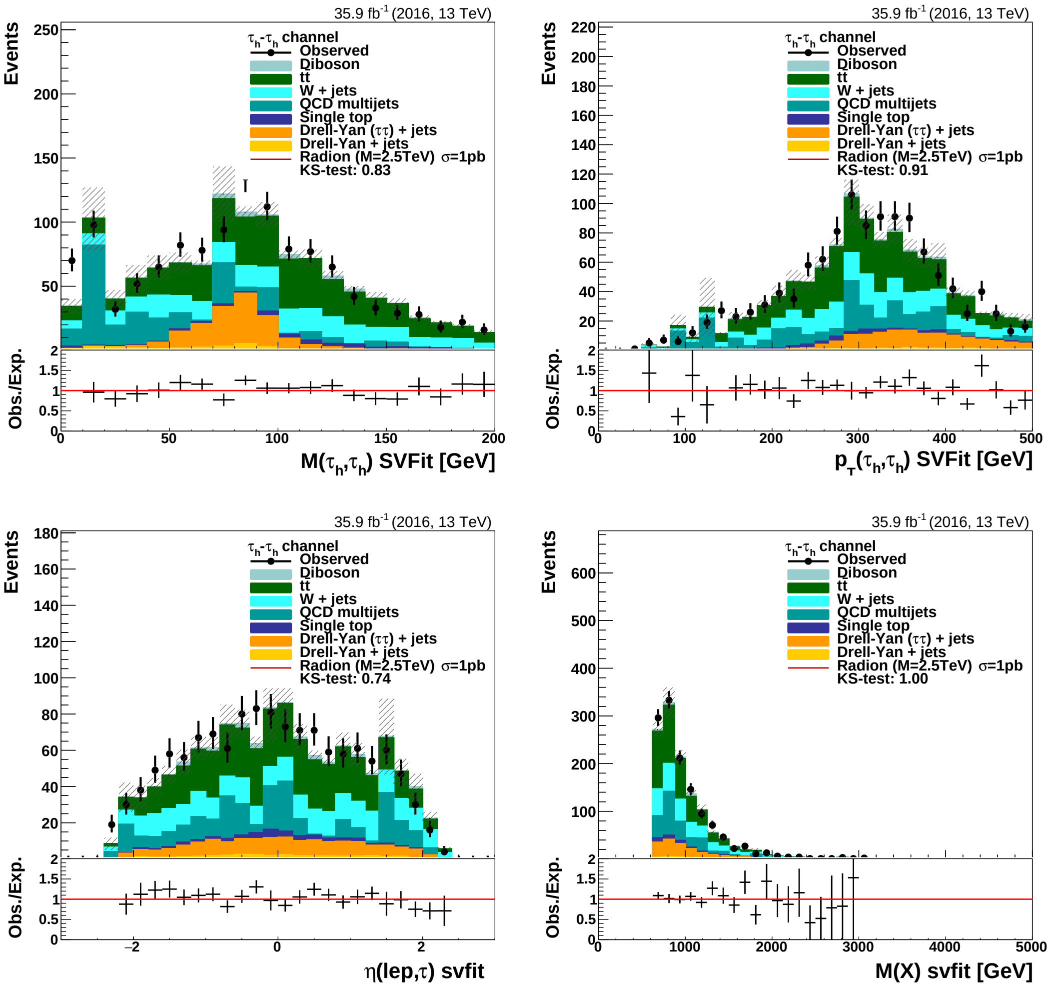


Figure 5.35: Comparison between data and expected simulated events for the $\tau_h\tau_h$ channel for the following variables: di-τ mass and p_T (top), and di-τ η and resonance mass (bottom).

Chapter 6

Search for heavy resonances in Run 1

A search for a signal compatible with a spin-0 massive resonance decaying into a pair of Higgs bosons is performed using the proton-proton collisions data sample collected at a center-of-mass energy of 8 TeV at CMS in 2012, corresponding to an integrated luminosity of 19.7 fb^{-1}.

In general HH events can be reconstructed in many different final states, either with large statistics and overwhelming backgrounds (e.g. jets originating from four bottom quarks) or high signal purity and limited background statics (e.g. four τ leptons). A good compromise between these two extremes is to look for one boson decaying hadronically and the other decaying to tau leptons.

In this analysis, one of the Higgs bosons decays to a pair of τ leptons, while the other is required to decay hadronically into a pair of bottom quarks. For a high-mass ($\gtrsim$ 1 TeV) resonance, the intermediate H bosons are produced with a large Lorentz boost; hence the decay products of the bosons are expected to be highly energetic and collimated. The hadronization products of the bottom quarks coming from one of the two intermediate H bosons give rise to the presence of one single "merged" large-cone jet, of high p_{T} which can be identified through a study of its substructure, consistent with the presence of two bottom quarks. This system is recoiling against the two tau leptons, produced by the other intermediate Higgs boson, that are similarly energetic.

In the analysis, events in the semileptonic final state, i.e. with one hadronic tau lepton decay and one leptonic tau lepton decay into an electron or a muon, are considered. The cleaning techinque, developed in [106] and described in Section 5.3.7.5, is used for the reconstruction and identification of the hadronic tau lepton decays. Events are collected with trigger paths requiring a highly energetic jet or large hadronic activity. The mass of the large-cone jet is used to define the signal region and signal-depleted control regions, which are sidebands (SBS). The usage of jet substructure techniques improves the background suppression, enhancing the sensitivity of the search. Various control regions in data are defined in order to estimate the background contribution to the signal region. The search is performed by examining the diboson invariant mass for a localized excess, in a spectrum that extends from 800 to 2500 GeV.

6.1 Data sample and simulation

The process $pp \to X \to \mathrm{HH} \to b\bar{b}\tau\tau$ is simulated at parton level using MadGraph 5 1.4.5 [144] in the narrow-width approximation, which is compatible with a spin-0 bulk radion model. Narrow-width approximation hereby means that the predicted resonance width is smaller than the experimental resolution. Five signal samples are generated with masses between 0.8 and 2.5 TeV.

The SM processes are generated using MC simulation with MadGraph 5 1.3.30 (Z/γ+jets and W+jets with leptonic decays), powheg 1.0 r1380 ($\mathrm{t\bar{t}}$ and single top quark production) [145–148], and pythia 6.426 [149] (SM diboson production and QCD multijet events). Showering and hadronization are performed with pythia and τ_h decays are simulated using tauola 1.1.5 [150] for all simulated samples. Geant4 [151] is used for the simulation of the CMS detector. Events are selected online by a trigger that requires the presence of at least one of the following: either a hadronic jet reconstructed by the anti-k_T algorithm [110] with a distance parameter of 0.5, $p_\mathrm{T} > 320$ GeV, and $|\eta| < 5.0$; or a total hadronic transverse energy, $\mathrm{H_T}$, defined as the scalar sum of the transverse energy of all the jets of the event, larger than 650 GeV.

It has been verified in events selected with an independent trigger requiring a muon of $p_\mathrm{T} > 18$ GeV that the efficiency of this jet trigger combination after applying the offline event selection is above 99%, as shown in Figure 6.1. The difference from 100% is considered as a systematic uncertainty.

6.2 Signal characterization

This analysis is performed in a high mass region (from 800 GeV to 2.5 TeV). The MadGraph algorithm is used to generate the hard process production in the collision, while in the next step of the simulation, during the hadronization, the QCD initial state

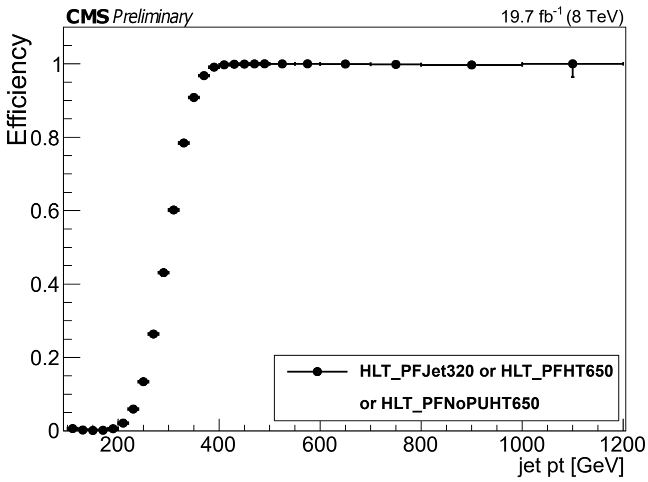


Figure 6.1: Trigger efficiency as a function of the wide-jet transverse momentum computed in events selected with an independent trigger that requires a muon with transverse momentum of at least 18 GeV.

radiation is added by PYTHIA. The event centrality can be seen from the generator level kinematic distributions of the intermediate Higgs bosons produced in the spin-0 radion decay as shown in Fig. 6.2.

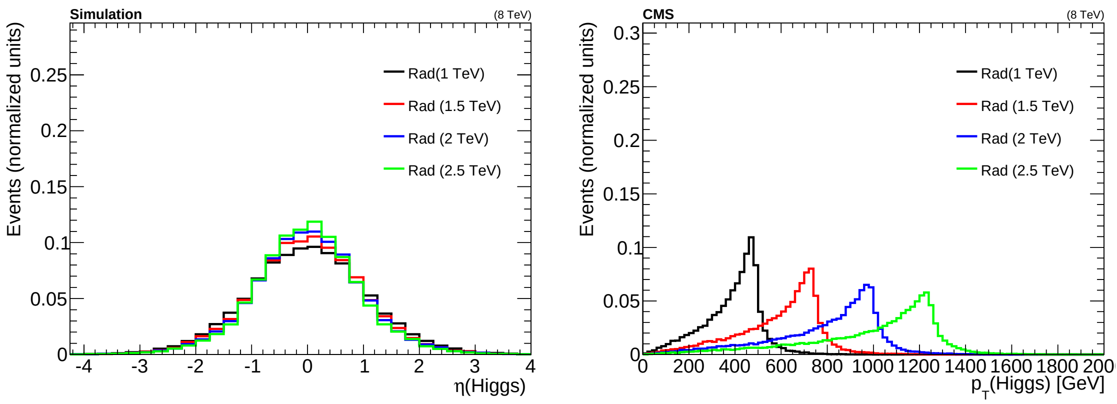


Figure 6.2: Kinematic distributions of the Higgs bosons at generator level: η (left) and generated transverse momentum (right).

In Fig. 6.3, the η angular distribution of the b quarks and the tau leptons from the intermediate Higgs bosons are shown. In Fig. 6.4, the distance between the b quarks and the tau leptons is shown: the higher the resonance mass, the higher the boost of the intermediate Higgs boson, and thus, the smaller the angular separation between the final decay products.

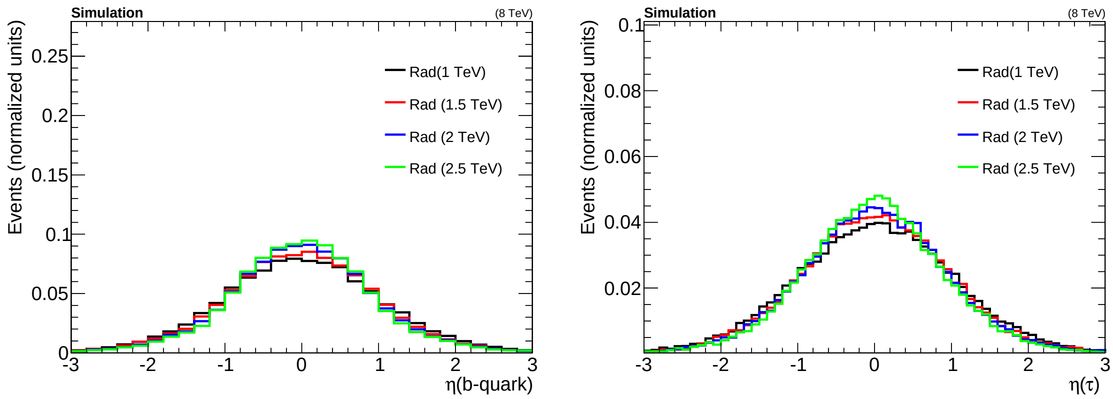


Figure 6.3: Eta distributions of b quarks (left) and τ leptons (right) produced in the Higgs boson decay.

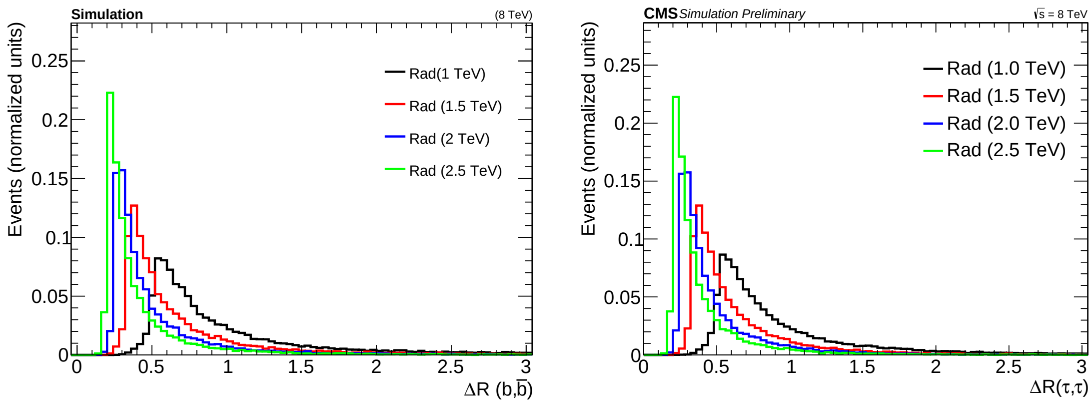


Figure 6.4: Distance between the Higgs boson decay products: b quark (left) and τ leptons (right).

In Fig.6.5, the generator-level muon and electron p_{T} spectra for different resonance masses are shown in the $\mu\tau_{\mathrm{h}}$ and the $e\tau_{\mathrm{h}}$ final states, as well as the momentum of the visible products of the τ_{h}. It can be seen that the lepton p_{T} spectra are softer than the τ_{h} p_{T} spectrum, because of the large fraction of energy carried away by the neutrinos in the case of a leptonic tau decay.

6.3 Event reconstruction and selection

The PF algorithm [?] is used to identify and to reconstruct candidate charged hadrons, neutral hadrons, photons, muons, and electrons produced in proton-proton collisions. Jets and τ_{h} candidates are then reconstructed using the PF candidates, as explained in Chapter 5.

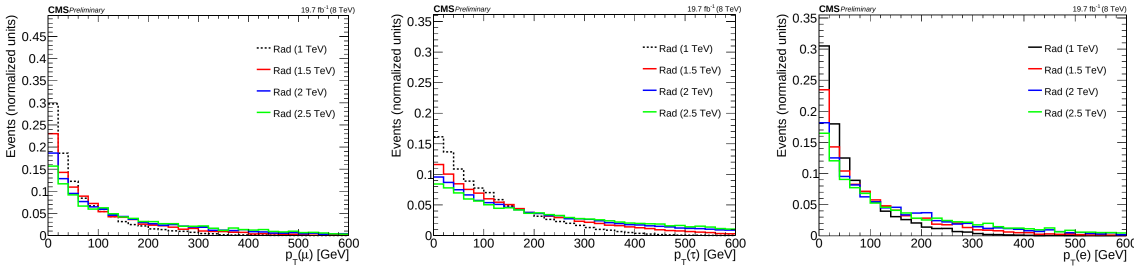


Figure 6.5: Generator-level muon (left), τ_{h} (center) and electron (right) transverse visible momentum distributions.

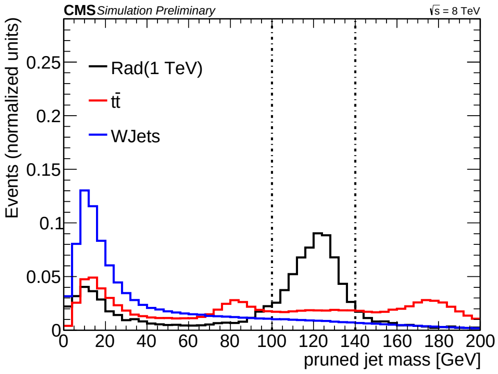
Jets

Hadronically-decaying boosted bosons are reconstructed with the CA8 jet algorithm with the CHS pileup mitigation procedure, as described in Sec.5.3.5.1. In order for the offline selection to match the trigger requirementsI and to avoid inefficiencies close to the threshold, at least one jet in the event is required to have $p_{\mathrm{T}} > 400$ GeV. The leading jet in the event is required to have $|\eta| < 1.0$ to ensure optimal tracking performance. This requirement does not worsen the sensitivity of the analysis, due to the central topology of signal events, as shown in Fig.6.3.

The pruned jet mass (m_{jet}^{P}) and the τ_{21} subjettiness ratio are used to discriminate between Higgs-boson jets and quark/gluon jets, as shown in Fig.6.6, for the most relevant physical processes that satisfy our event selection. Events with $\tau_{21} > 0.75$ are rejected since they are compatible with jets originating from quarks or gluons. The hadronic Higgs-boson jet candidate is identified by requiring $100 < m_{\mathrm{jet}}^{P} < 140$ GeV, while events with jets of masses $m_{\mathrm{jet}}^{P} < 100$ GeV or $m_{\mathrm{jet}}^{P} > 140$ GeV are considered as background-dominated sidebands (SBs) and used as control regions.

In order to tag jets from H$\rightarrow b\bar{b}$ decays, the pruned subjets are used as the basis for b tagging: if ΔR is larger than 0.3, the CSV b-tagging algorithm [152] is applied to both of the subjets, while it is applied to the whole CA8 jet otherwise, following the recipe outlined in 5.3.5.6. The “loose” working point of the CSV algorithm [152] is chosen for both subjet and large-cone-jet b tagging. It has a b-tagging efficiency of about 85%, with mistagging probabilities of ≈40% for charm-quark jets and ≈10% for light-quark and gluon jets at jet p_{T} near 80 GeV.

Jets originated from single quarks or gluons are reconstructed with the AK5 jet algorithm, not considering the ones that overlap with the Higgs-boson wide jet and with the leptons. The number of b jets in the event provides a useful criterion to reduce the $\mathrm{t\bar{t}}$ background. Events are separated depending on the number of additional b-tagged



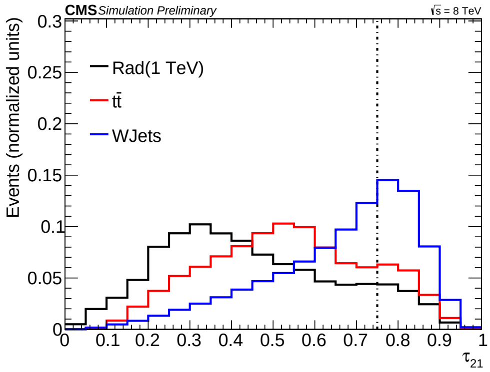


Figure 6.6: Pruned jet mass spectrum (left) and τ_{21} (right) distribution of the most energetic wide jet in simulated events of different processes: top quark pair production, W+ jet production and Radion decaying to Higgs bosons.

AK5 jets: no b-tagged jets are required in the signal region selection, while at least 1 b-tagged jet is required for the $t\bar{t}$-enriched event control region.

Missing transverse energy

All particles reconstructed with the PF algorithm are used to determine the missing transverse momentum, $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$, with a procedure described in Section 5.3.6. Type-I corrected $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ is used in the analysis, along with dedicated filters to remove detector noise and events with faulty reconstruction.

Leptons: electron, muon and hadronic taus

Electrons with $p_{\mathrm{T}} > 10$ GeV and $|\eta| < 2.5$ are selected if they satisfy the tight requirements reported in Sec. 5.3.2. Muons are required to have $p_{\mathrm{T}} > 10$ GeV and $|\eta| < 2.4$ and to pass the requirements on the quality of the track, as explained in Sec.5.3.1. Electron and muon candidates have to satisfy particle-flow based isolation criteria that require low activity in a cone around the lepton, the isolation cone, after the removal of particles due to additional pileup interactions and the possible presence of a nearby hadronic tau lepton decay, as described in Sec. 5.3.3.

The reconstruction of τ_{h} candidates is done with the cleaning technique, explained in Sec. 5.3.7.5, and starts from AK5 jets where electrons and muons, identified by looser criteria than the nominal ones used in the analysis, are removed from the list of particles used in the clustering. HPS-reconstructed τ_{h} candidates with $p_{\mathrm{T}} > 35$ GeV and $|\eta| < 2.3$ are considered in the analysis. Electrons and muons misidentified as τ_{h}

candidates are suppressed using dedicated criteria based on the consistency between the measurements in the tracker, the calorimeters, and the muon detectors. Finally, loose MVA-based isolation criteria are applied to the τ_h candidates, not considering electrons or muons in the τ_h isolation cone. In the analysis, various requirements on the MVA-based isolation are applied to select different regions. A τ_h candidate is defined as isolated if this isolation variable is > -0.7. We define as "Intermediate Isolation" the region between -0.95 and -0.7. This region provides more statistics with respect to the signal region for the main backgrounds of the analysis, thus will be used as control region. The output of the tau isolation MVA is shown in Fig. 6.7.

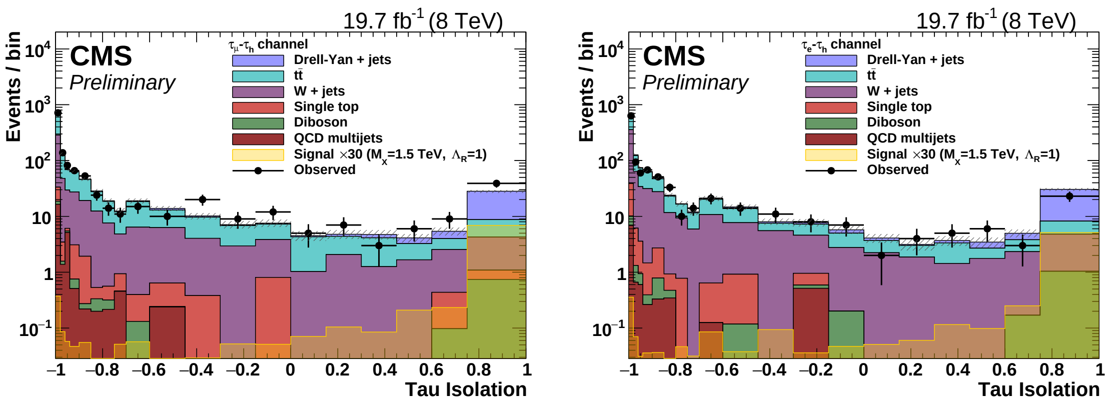


Figure 6.7: MVA-based τ_h isolation in $\mu\tau_\mathrm{h}$ and $e\tau_\mathrm{h}$ events. The distribution for a radion signal with a mass of 1.5 TeV is also shown.

6.3.1 Additional requirements

Additional selection requirements are applied to remove backgrounds from low-mass resonances and avoid overlaps between τ_h and other leptons: the visible mass $m_\mathrm{vis}(\ell, \tau_\mathrm{vis}) >$ 10 GeV, $\Delta R_{\ell,\tau\mathrm{vis}} > 0.1$ (where $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ and ℓ denotes the electron or muon), $|\vec{p}_\mathrm{T}^\mathrm{miss}| > 50$ GeV, and $p_{\mathrm{T},\tau\tau} > 100$ GeV, as estimated from the SVFit procedure. An upper cut is placed on $\Delta R_{\ell,\tau\mathrm{vis}} < 1.0$ in order to reject W+jets events, where a jet misidentified as a τ_h lepton is usually well-separated in space from the isolated lepton. A summary of the selection is given in Tab. 6.1.

In Figs.6.8–6.10 a comparison between data and expected background processes from simulations is shown, applying the complete set of requirements, except for the pruned jet mass window and the Higgs b-tagging selection. The prediction from simulations is able to describe data well within the statistical uncertainty.

Table 6.1: Summary of the optimized event selection for the signal region. The selection variables are explained in the text. The label ℓ refers to electrons and muons.

Selection
$p_{\mathrm{T},\tau_{\mathrm{h}}} > 35$ GeV , $p_{\mathrm{T},\ell} > 10$ GeV
$0.1 < \Delta R_{\ell,\tau\mathrm{vis}} < 1.0$, $m_{\mathrm{vis}}(\ell,\tau_{\mathrm{h}}) > 10$ GeV
$\lvert\overrightarrow{p}_{\mathrm{T}}^{miss}\rvert > 50$ GeV
$p_T(\tau\tau)$ from SVFit > 100 GeV
$p_{T,\mathrm{jet}} > 400$ GeV and $\lvert\eta_{\mathrm{jet}}\rvert < 1$
$100 < m_{\mathrm{jet}}^{P} < 140$ GeV , $\tau_{21} < 0.75$
Higgs-b-tagging: 1 CSVL-tagged fat jet if ΔR(sj1,sj2) < 0.3
or 2 CSVL-tagged subjets if ΔR(sj1,sj2) > 0.3
$N_{\text{extra b-tagged jets}} = 0$

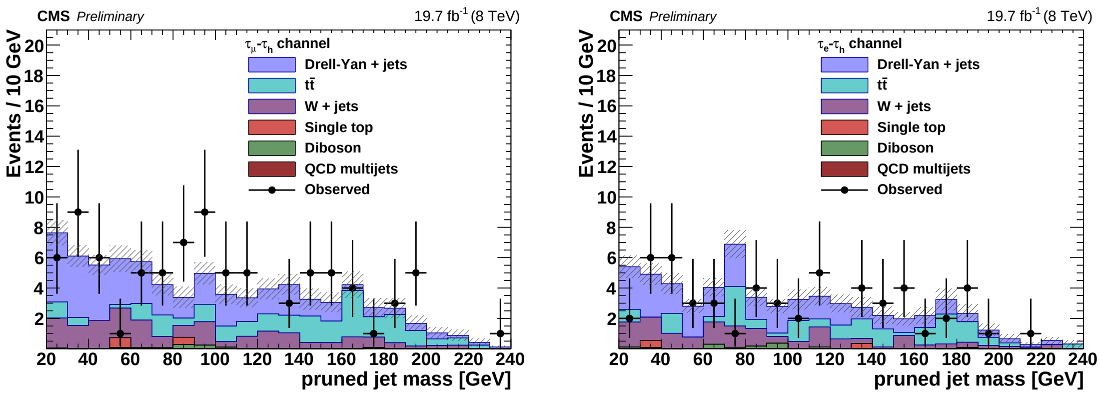


Figure 6.8: Number of background and data events in the pruned jet mass spectrum in the $\mu\tau_{\mathrm{h}}$ (left) and $e\tau_{\mathrm{h}}$ (right) channels. The complete signal selection except the pruned mass window and the Higgs b-tagging requirements is applied.

In Fig. 6.11 the distributions of the pruned jet mass and SVFit mass are shown in $\mu\tau_{\mathrm{h}}$ and $e\tau_{\mathrm{h}}$ events that satisfy the complete signal selection (including H tagging), but with relaxed a requirement on the pruned jet mass. The Higgs b-tagging requirement reduces considerably the statistics of events that satisfy the complete selection, in both data and simulated samples.

6.4 Background estimation

The main backgrounds are $t\bar{t}$, W+jets and Drell-Yan +jets events, while the other background components are negligible.

In order to rely as little as possible on the simulated events, that satisfy the complete signal region selection, which, due to the statistics, are rather inaccurate, a predomi-

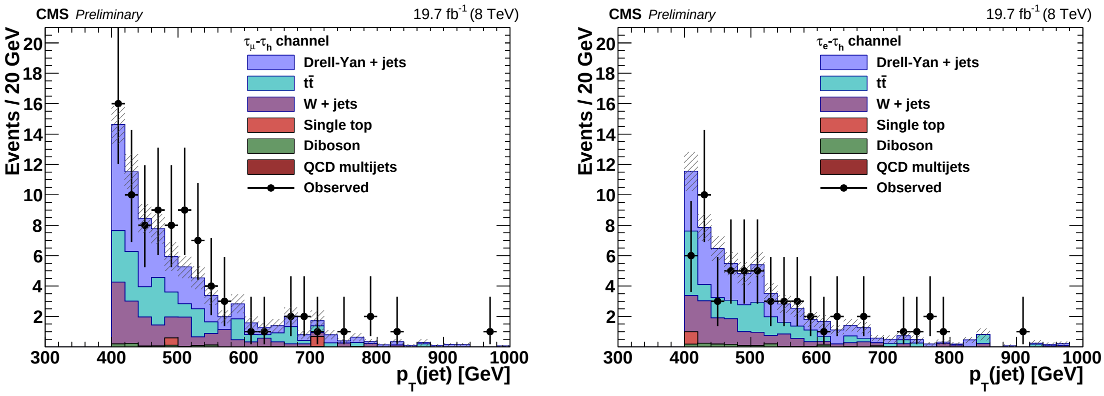


Figure 6.9: Number of background and data events in the jet transverse momentum spectrum in the $\mu\tau_{\mathrm{h}}$ (left) and $e\tau_{\mathrm{h}}$ (right) channels. The complete signal selection except the pruned mass window and the Higgs b-tagging requirements is applied.

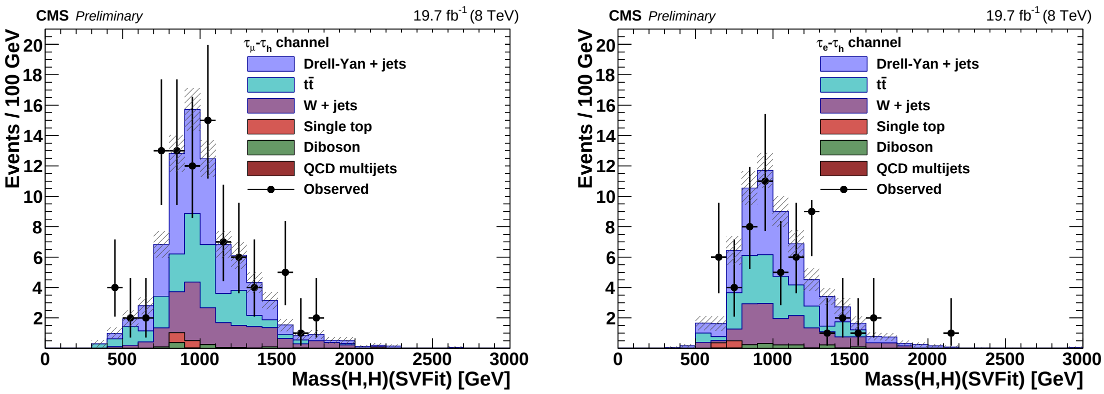


Figure 6.10: Number of background and data events in the invariant mass of the di-H system spectrum as reconstructed by SVFIT in the $\mu\tau_{\mathrm{h}}$ (left) and $e\tau_{\mathrm{h}}$ (right) channels. The complete signal selection except the pruned mass window and the Higgs b-tagging requirements is applied.

nantly data-driven prediction is performed. The background shapes are modeled with background simulation from events in which the Higgs b-tagging requirement is not applied, while the background yields and background composition are estimated from observed b-tagged events with a pruned mass in the signal sidebands. Since the number of expected background events for our selection is below one event, we use sidebands with a similar selection, but a significantly higher number of expected background events, to determine with a higher precision the number of events in the signal region due to background processes. Sideband regions are defined in the analysis in a way to differ from the signal region by either changing the pruned jet mass requirement, removing the Higgs-b-tagging requirement ("untagged"), or varying requirements on the MVA-based τ_{h} isolation.

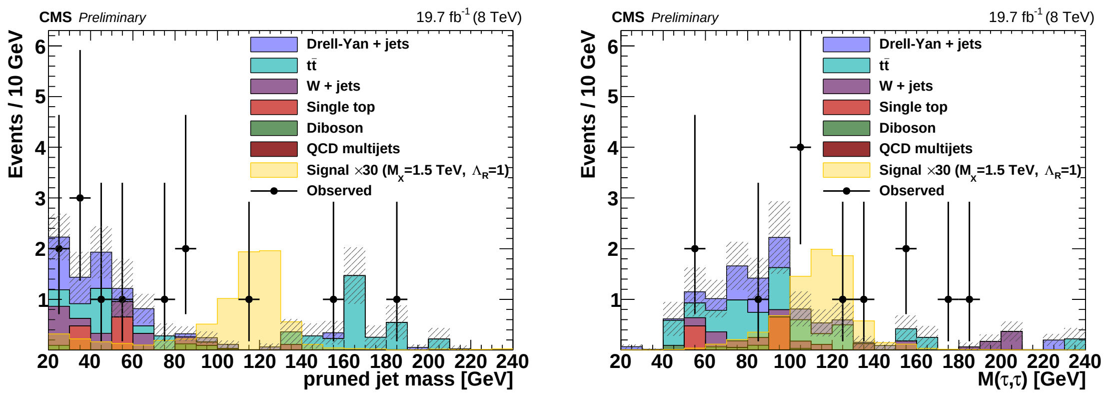


Figure 6.11: Number of background and data events in the pruned jet mass (left) and reconstructed SVFit mass (right) spectrum in the $\mu\tau_{\mathrm{h}}$ and $e\tau_{\mathrm{h}}$ channels combined. The complete signal selection except the pruned mass window requirement is applied. The distributions for a radion signal with a mass of 1.5 TeV are also shown.

For this purpose, an intermediate isolation region $[-0.95, -0.7]$ and an inverted isolation region $[-1, -0.7]$ are defined (the loose isolation region used in the analysis is $[-0.7, 1]$); agreement between data and background simulation is shown in Fig. 6.7. An overview of the various sidebands used in the background estimation procedure and its cross check is given in Tab. 6.2 together with the expected number of background events and signal efficiency in each sideband. The usage of each sideband and the corresponding assumptions that lead to systematic uncertainties on the background estimate are listed as well. In the following, the procedure to determine the scale factors is explained. Then the background estimation method, both for yields and shapes, is presented together with its validation.

Events that satisfy the intermediate isolation region requirement without Higgs-boson b tagging and having a jet with pruned mass of 20 GeV $< m^{P}_{\mathrm{jet}} <$ 240 GeV are used to estimate the data-to-simulation corrective factors for the background normalization. The distribution of the background samples in this region is shown in Fig. 6.12 (a, b). Data is divided by considering two contributions: one from $t\bar{t}$ events, whose distribution shows a peak around the top mass around 170 GeV, and all the other backgrounds together, which have a falling distribution and whose main contribution is from QCD production of a W boson in association with jets. A fit is performed to get the overall background normalization considering just these two contributions to derive data-to-simulation scale factors ($\xi^{Untagged}_{IntermediateIso}$) as shown in Fig. 6.12 (a, b).

In order to check dependencies of the scale factors on the b-tagging requirement, we use a higher-statistics sample by inverting the isolation cut. The two background

Region	Pruned jet mass [GeV]	Higgs-b-tagging	Tau isolation
Signal	100–140	b-tagged	loose
IntermediateIso	20–240	untagged	intermediate
	derive data-to-simulation scale factors for backgrounds		
	assumption: data-to-simulation agreement same as with loose isolation		
InvertedIso	20–240	untagged b-tagged	inverted
	extrapolate data-to-simulation agreement from untagged to b-tagged sample		
	assumption: b-tag scale factor same as with loose isolation		
Mass sideband	20–100, 140–240	untagged	loose
	cross check background estimation procedure		
	assumption: pruned jet mass spectrum correctly described		
Untagged	100–140	untagged	loose
	obtain simulated background shapes for signal region		
	assumption: background shape not changed by b-tag requirement		

Table 6.2: Summary of the signal and sideband regions, their purpose, and assumptions made.

components are then fit again in this inverted isolation region as shown in Fig. 6.12 (c, d, e, f) and data-to-simulation scale factors for untagged ($\xi^{Untagged}_{InvertedIso}$) and b-tagged ($\xi^{b-tagged}_{InvertedIso}$) samples are computed. The results found for each sample are in agreement within their uncertainties for all the different regions with and without the b-tagging requirement. To account for a possible dependence on the b-tagging requirement, we determine the signal region scale factor (ξ_{SR}) as the product of the factor found in the intermediate isolation region and the ratio of the b-tagged to untagged scale factors found in the inverted isolation sideband:

$$\xi_{SR} = \xi^{Untagged}_{IntermediateIso} \times \xi^{b-tagged}_{InvertedIso} / \xi^{Untagged}_{InvertedIso}. \tag{6.1}$$

These overall scale factors are applied to the simulated samples in the signal region and are reported in the last column of Tab. 6.3.

We perform a consistency check for the scale factors by computing them also for isolated τ_{h} candidates in the low (20 GeV $< m^P_{\mathrm{jet}} <$ 100 GeV) and high ($m^P_{\mathrm{jet}} >$ 140 GeV) pruned jet-mass sidebands together, i.e. in events with the full selection applied, τ isolation requirement included, but vetoing in the fit the pruned jet mass signal region [100, 140] GeV and removing the b-tag requirement on the Higgs jet. In the $\mu\tau_{\mathrm{h}}$ channel, for $\mathrm{t\bar{t}}$ a $\xi^{Untagged}$ is measured to be 0.86 ± 0.24, while for the other backgrounds

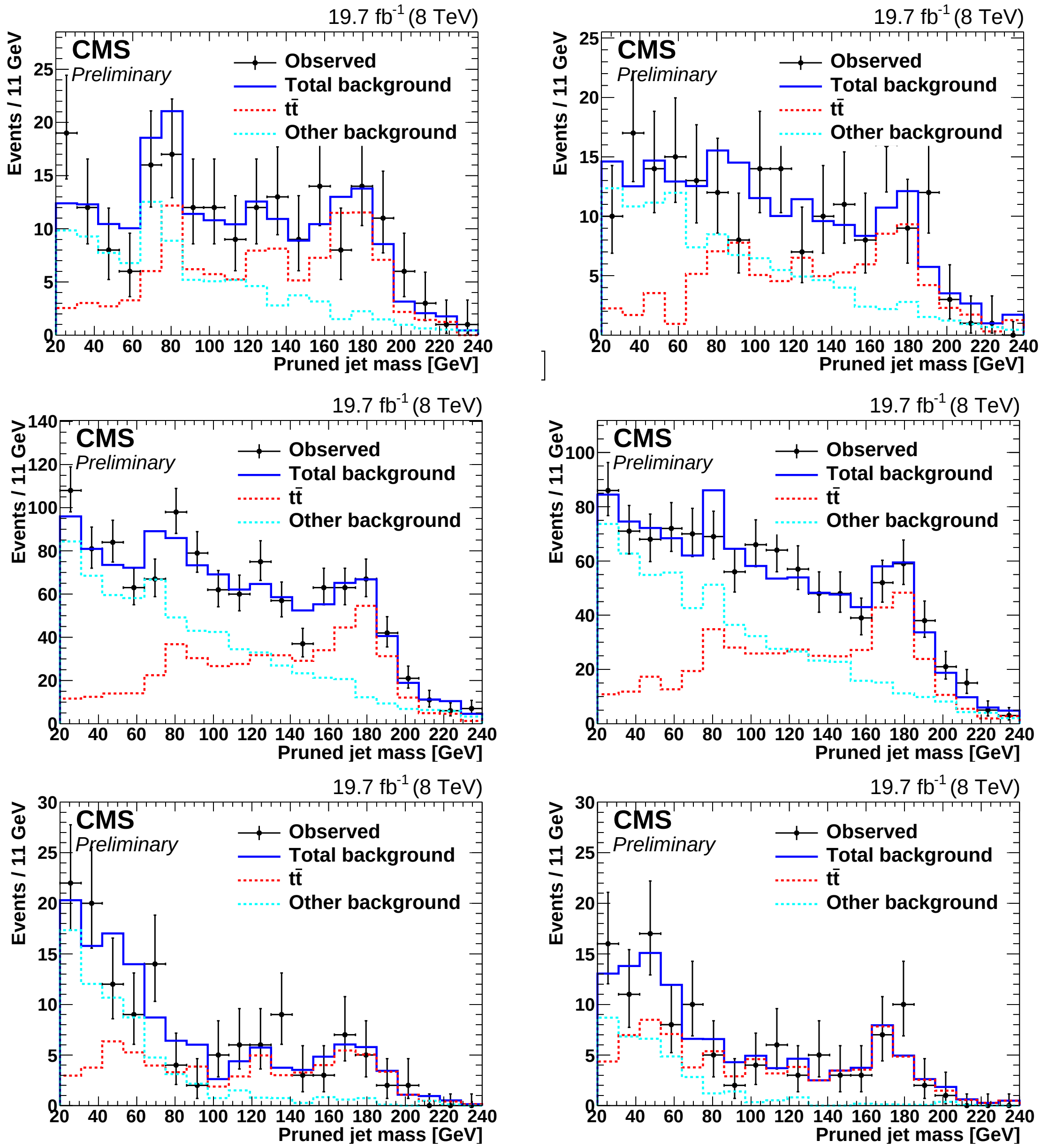


Figure 6.12: Post-fit normalization for the background components: (a, b) in the intermediate isolation region sideband without b tagging, (c, d) in the inverted isolation region sideband without b tagging, (e, f) in the inverted isolation region sideband with b tagging, in the (a, c, e) $\mu\tau_{\rm h}$ channel and (b, d, f) $e\tau_{\rm h}$ channel.

$\xi^{Untagged} = 0.77 \pm 0.13$ in agreement with $\xi^{Untagged}_{IntermediateIso}$ from the background estimation procedure.

Since the expected number of background events is very small after signal selection, the untagged sideband in simulation is used to estimate the background distribution shapes. Then, those are scaled to the signal region yields obtained from simulation

Channel	**Background**	$\xi^{Untagged}_{IntermediateIso}$	$\xi^{b-tagged}_{InvertedIso}$	$\xi^{Untagged}_{InvertedIso}$	ξ_{SR}
$\mu\tau_h$	$t\bar{t}$	1.04 ± 0.17	0.65 ± 0.12	0.74 ± 0.06	0.91 ± 0.24
	other	0.68 ± 0.13	0.94 ± 0.17	0.93 ± 0.06	0.69 ± 0.19
$e\tau_h$	$t\bar{t}$	0.95 ± 0.18	0.86 ± 0.12	0.78 ± 0.07	1.05 ± 0.28
	other	1.02 ± 0.16	0.62 ± 0.18	0.98 ± 0.07	0.65 ± 0.22

Table 6.3: Summary of scale factors, ξ, obtained for background samples in the different regions

and corrected in the overall normalization with data-to-simulation scale factors. This procedure is legitimate, since the dependence of the simulated shapes on the b-tagging is found to be negligible, which is confirmed in two steps using simulations. First, by dropping the jet $|\eta| < 1$ and the H-tagging requirement to increase the statistics, the background distributions for $\mathrm{t\bar{t}}$ and other backgrounds are checked to be similar in the isolation and intermediate isolation regions. Secondly, in the intermediate isolation region, the distribution for $\mathrm{t\bar{t}}$ and the other backgrounds are found compatible before and after the b-tagging requirement.

6.5 Systematic uncertainties

The sources of systematic uncertainty in this analysis, which affect either the background estimation or the signal efficiencies, are described below.

For the signal efficiency, the uncertainties on the integrated luminosity (2.6%) [153], and the uncertainty on the modeling of pileup (additional interactions occurring in the same LHC bunch crossing) (0.2–1.4%) are taken into account. The scale factors for lepton identification are derived from dedicated analyses of observed and simulated $Z \to \ell^+\ell^-$ events, using the "tag-and-probe" method [133,154]. The uncertainties in these factors are taken as systematic uncertainties and amount to about 2% for electrons, 1% for muons, and 12% for hadronic tau decays. The jet and lepton four-momenta are varied over a range given by the energy scale and resolution uncertainties [155]. In this process, variations in the lepton and jet four-momenta are propagated consistently to $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$.

Additional uncertainties come from the procedure of removing nearby tracks and leptons used in the hadronic τ reconstruction, and from the isolation variable computation in the case of boosted topologies. The variation of the identification efficiency due to the cleaning procedure and the modified lepton isolation with respect to the standard ones, as measured in simulations, is assigned as a systematic uncertainty, corresponding to 1–16% for τ reconstruction and 1–21% for lepton isolation. The jet trigger efficiency has an uncertainty of < 1%, as determined from a less selective trigger. Following the method derived for vector boson identification in merged jets [156], a scale factor of 0.94 ± 0.06 is used for the efficiency of the pruning and subjet searching techniques applied on the CA jet, where the uncertainty is included in the estimation of the overall systematic uncertainty. For the b tagging, data-to-MC corrections [152] derived from several control samples are applied and the uncertainties on these corrections are propagated as systematic uncertainties in the analysis (6–9%).

The uncertainties in the background estimate are dominated by the limited number of simulated events and sideband data events. For the background normalization, we assign as systematic uncertainty the statistical error coming from the combination of the uncertainties of the scale factors quoted in the last column in Table 6.3, which amount to 26–33% for the $t\bar{t}$ and the other background components. An additional uncertainty of 50% is assigned to the $t\bar{t}$ yields in the signal region since the number of MC-simulated events available to estimate its contribution is limited.

In Table 6.4 a summary of all the systematic uncertainties evaluated for the signal in the analysis is shown for each channel, where the minimum and the maximum values between the five signal masses are present.

6.6 Results

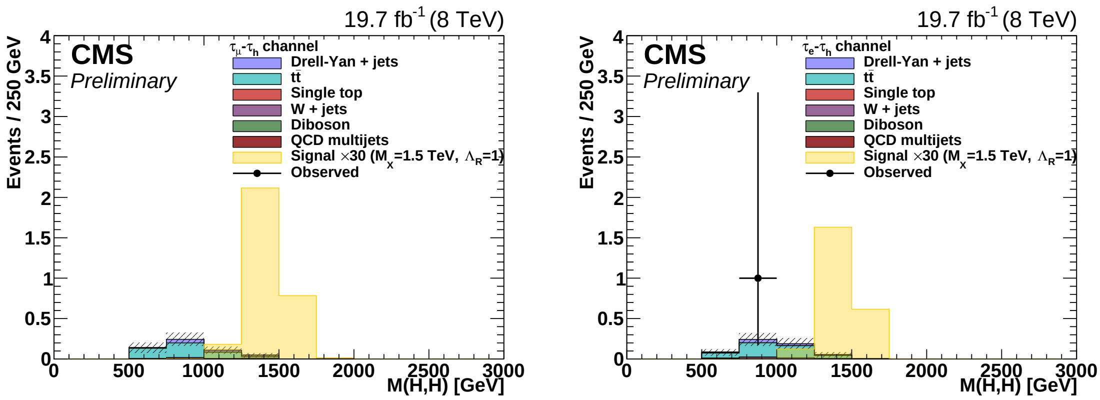


Figure 6.13: Number of background and signal (1.5 TeV) events with complete selection applied in the $\mu\tau_{\rm h}$ (left) and $e\tau_{\rm h}$ (right) channels.

The expected background and the data distributions for the final selection in the signal region are reported in Fig. 6.13. Table 6.5 shows the signal efficiencies, while the background expectation and the number of observed events are shown in Table 6.6.

Source	$\mu\tau_h$ **channel**	$e\tau_h$ **channel**
Luminosity	2.6%	2.6%
Pile-up	0.2–1.4%	0.7–1.2%
Mass window and τ_{21}	8.9%	8.9%
Higgs-b-tag	2.4–10%	2.4–10%
b-tag for veto	6.2–8.8%	6.0–8.5%
Jet energy scale	2.2–2.4%	1.1–2.2%
Jet energy resolution	< 0.5–1.1%	0.5–1.2%
Electron ID	-	1.3–1.8%
Electron energy resolution	-	0.2–0.7%
Electron energy scale	-	0.1–0.4%
Muon ID	0.8–0.9%	-
Muon momentum resolution	< 0.5%	-
Muon momentum scale	< 0.5–0.8%	-
Lepton modified iso	1.2–14.3%	3.5–20.8%
Tau ID	8.9–12.4%	8.5–11.9%
Tau Scale	< 0.5–1.1%	< 0.5–2.4%
Tau-jet cleaning	0.4–7.0%	0.5–15.7%
MET	Included in lepton and jet uncertainties	
Total	16–27%	21–34%

Table 6.4: Summary of the systematic uncertainties on the signal. Minimum and maximum values between the signal masses are reported.

No significant deviation is found between the observed number of events from the expected background.

Table 6.5: Summary of the signal efficiencies (including acceptances and tau pair branching ratios $\mathcal{B}(\tau\tau)$). Statistical and systematic uncertainties are included. The branching ratio of the τ pair in the final state is also shown.

	Mass [TeV]	$\mu\tau_h$	$e\tau_h$
$\mathcal{B}(\tau\tau)$		22.6%	23.1%
$\varepsilon_{sig}(\%)$	0.8	0.19 ± 0.03(stat) ± 0.03(syst)	0.14 ± 0.03(stat) ± 0.03(syst)
	1.0	1.70 ± 0.09(stat) ± 0.31(syst)	1.10 ± 0.07(stat) ± 0.20(syst)
	1.5	3.16 ± 0.13(stat) ± 0.63(syst)	2.44 ± 0.11(stat) ± 0.48(syst)
	2.0	5.07 ± 0.17(stat) ± 1.17(syst)	3.95 ± 0.15(stat) ± 0.91(syst)
	2.5	4.02 ± 0.14(stat) ± 1.09(syst)	2.60 ± 0.11(stat) ± 0.88(syst)

Upper limits on the production cross section of a new resonance in the di-Higgs boson final state are set. The $\mathrm{CL_s}$ criterion [157, 158] is used to extract upper bounds on the product of the cross section and the branching ratio of a spin-0 radion signal decaying into a pair of Higgs bosons, combining both event categories. The test statistic is a profile likelihood ratio [159] and the systematic uncertainties are treated as nuisance

Table 6.6: Summary of the number of observed and expected background events in the mass window around the considered resonance masses for the two channels. Statistical and systematic uncertainties are included. Just one data event in the mass window [680, 920] GeV is observed in the $e\tau_{\mathrm{h}}$ channel.

	Mass [TeV](N_{obs})	$\mu\tau_{\mathrm{h}}$	$e\tau_{\mathrm{h}}$
N_{bkg}	[0.68,0.92] (1)	0.20 ± 0.09	0.22 ± 0.10
	[0.85,1.15] (0)	0.25 ± 0.11	0.25 ± 0.10
	[1.25,1.75] (0)	0.05 ± 0.02	0.06 ± 0.02
	[1.70,2.30] (0)	0.005 ± 0.002	0.006 ± 0.002
	[2.10,2.90] (0)	0.001 ± 0.0003	0.002 ± 0.0005

parameters. The nuisance parameters are described with log-normal prior probability distribution functions, except for those related to the extrapolation from sideband events, which are expected to follow a Γ distribution [159]. A more detailed explanation of the statistical treatment and the handling of the nuisance parameters is reported in Appendix A. For each resonance mass hypothesis, only events in a region corresponding to ±2.5 times the expected mass resolution around each mass point are considered in the likelihood, thus a shape analysis by counting events in these regions is performed.

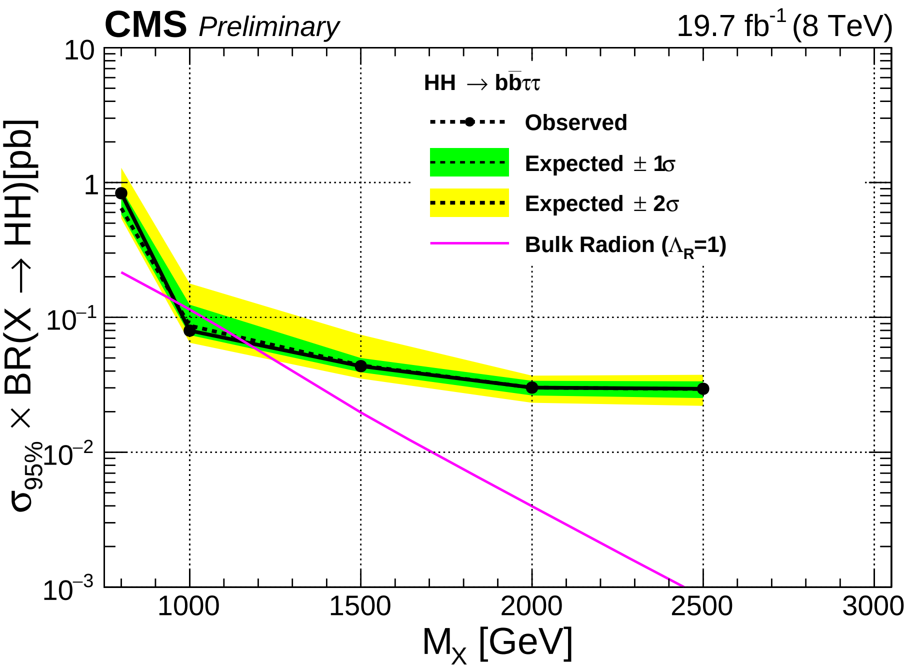


Figure 6.14: Expected and observed 95% CL upper limits on the cross section of a bulk radion resonance decaying into Higgs bosons ($\sigma(X \to \mathrm{HH})$): $e\tau_{\mathrm{h}}$ and $\mu\tau_{\mathrm{h}}$ channel combination.

In Fig. 6.14 the expected and observed upper limits for the production cross section of a spin-0 resonance in proton-proton collisions decaying to HH are shown combining the

$\mu\tau_h$ and $e\tau_h$ channels. They are compared to the expected values for the production cross section of a Radion $\to$ HH for which the ultraviolet mass scale parameter as defined in Ref. [41] has been set to $\Lambda_R = 1$ TeV.

The analysis sets 95% CL upper limits on the cross section of a spin-0 resonance ranging from 850 to 30 fb for resonances of masses between 800 to 2500 GeV and radions (with $\Lambda_R = 1$ TeV) are excluded between 950 and 1150 GeV. While other searches have looked for resonances decaying into a pair of Higgs bosons with τ leptons and bottom quarks in the final state, or also in other final states (Fig. 6.15), this is the first search in the high-mass regime ($\gtrsim 1$ TeV), where the two b quarks coming from one of the two intermediate Higgs bosons give rise to the presence of one single "merged" jet and the two τ leptons from the other intermediate Higgs boson traverse the detector very close to each other and require advanced reconstruction techniques. It is important to note that different reconstruction and analysis techniques extend nicely the results towards higher values of the resonance mass and that the searches in the $b\bar{b}\tau\tau$ and $b\bar{b}b\bar{b}$ channels have a comparable sensitivity despite the smaller branching ratio of the former.

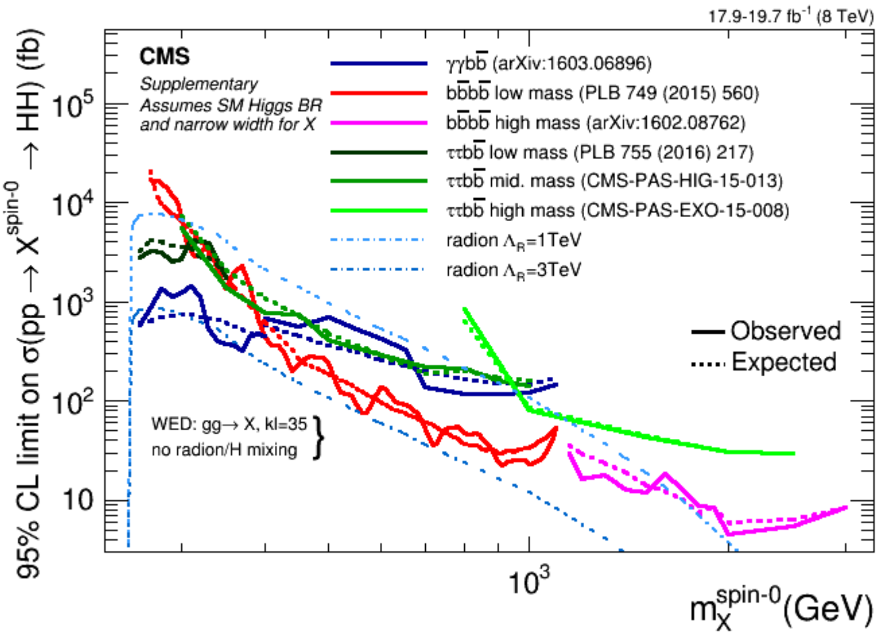


Figure 6.15: Expected and observed limits on the production of a spin-0 resonance that decays to a pair of Higgs bosons for the analyses performed by the CMS collaboation with the 2012 data.

Chapter 7

Search for heavy resonances in Run 2

7.1 Overview

After 3 years of a shutdown and machine development, in 2015 the LHC started again its physics program with collisions at 13 TeV. The higher energy of the collisions meant an increase on the partonic luminosities, i.e. higher luminosity for the partonic interactions, as shown in Fig.7.1. The cross section for the production of gluon-originated resonances, such as radions or gravitons, for a resonance mass of 1 TeV is almost higher by a factor of 6 and rapidly increasing with the resonance mass. Instead the production of a heavy vector boson, such as a W' or a Z', which is produced in quark interactions, gains a factor of about 3 in the cross section for a resonance of mass of 1 TeV and increases more mildly with the resonance mass with respect to the gluon-originated case.

However, this increase affects also the production of the SM processes like $t\bar{t}$ and V+jets, that constitute the main source of backgrounds: the $t\bar{t}$ cross section increases by a factor of $\sim$ 3-4 and the V+jets by $\sim$ 2–3.

Overall this results in a better sensitivity for analyses that search for resonances at high masses that can extend their reach to a higher value in the tail of the mass spectrum. The total integrated luminosity recorded in 2016 by CMS is $35.9\,\text{fb}^{-1}$, almost twice the amount from Run 1, providing a further improvement for analyses that search for high-mass resonances that are usually limited by the amount of data in the control regions.

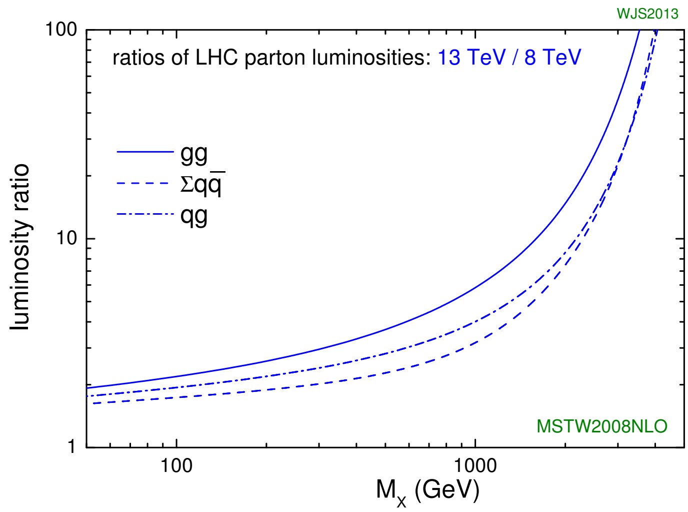


Figure 7.1: Ratio beween the partonic luminosities at 13 and 8 TeV for interactions of gluons (solid line), quarks (dashed lines) and quark with gluons (dotted-dashed line) [160].

With respect to the 8 TeV case, the analysis features the usage of a different trigger requirement based on large missing transverse momentum and a different background estimation technique, more reliant on control regions in data, which have larger statistics at this higher center-of-mass energy, a more complex categorization of the final states, and additional signal models: in addition to the spin-0 resonance, spin-1 and spin-2 resonances are also considered, with parameters consistent with the ones predicted by either the bulk radion or graviton or V′ (W′ or Z′) models. The search looks for resonant production of either a Higgs boson pair or a Higgs boson and a W or Z boson. The (one) Higgs boson is assumed to decay to τ leptons, while the other boson decays to quarks.

7.2 Data sample and simulation

The data sample analyzed in this search has been collected during 2016 pp collisions at a center-of-mass energy of 13 TeV, in 25 ns bunch spacing runs and with the magnetic field enabled. They correspond to an integrated luminosity of 35.9 fb^{-1}. The signal processes $pp \to X \to VH \to q\bar{q}\tau^+\tau^-$ and $pp \to X \to HH \to b\bar{b}\tau^+\tau^-$ are simulated at leading order (LO) using the MadGraph5_aMC@NLO v2.2.2 [161] Monte Carlo (MC) event generator, for resonance masses between 900 and 4000 GeV, where the Higgs boson is forced to decay to τ pairs and the other boson to a pair of quarks. The signal processes where pp $\rightarrow$ X $\rightarrow$ HH $\rightarrow$ $\mathrm{b\bar{b}}$VV and pp $\rightarrow$ X $\rightarrow$ VH $\rightarrow$ $\mathrm{q\bar{q}}$VV are also considered, in which VV $\rightarrow 2\ell 2\nu$ or $2\tau 2\nu$, as they can yield final states similar to those of the primary signal process. The natural width of the resonance is assumed to be smaller than the experimental resolution of its reconstructed mass, as consistent with with parameters consistent with the ones predicted"the benchmark radion, graviton and HVT models. The samples are produced assuming a width of 0.1% of the resonance mass.

The SM background processes are generated using MC simulation. The Z $/\gamma^*$+jets events and the W+jets events are simulated at LO with the MadGraph5_aMC@NLO generator. The POWHEG v2 generator is used to simulate $\mathrm{t\bar{t}}$ and single top quark production at next-to-leading order [145–148]. The LO PYTHIA 8.205 [162] generator is used for SM diboson (WW, WZ, or ZZ) and multijet events. For all signal and background samples, showering and hadronization are modeled using PYTHIA, τ lepton decays are described using TAUOLA 1.1.5 [163], and the response of the detector is simulated using Geant4 [151].

Additional collisions in the same or adjacent bunch crossings (pileup) are superimposed onto the hard scattering processes, with the pileup vertex multiplicity distribution adjusted to match that of data.

7.2.1 Trigger

The $p_{\mathrm{T}}^{\mathrm{miss}}$ primary dataset, where a missing energy trigger is required, is used for the analysis. Also, datasets triggered by single isolated leptons are used as control regions to check the trigger selection efficiency in data. The data samples used for the analysis, after being recorded, are reprocessed with the latest calibrations for the 2016 data taking. Data are moreover filtered from events that are problematic or noise dominated due to partial failures in the detector subsystems.

Events are selected on-line by the two-stage trigger described in Sec. 4.2.5. For the higher center-of-mass energy in the collisions of the LHC, the thresholds of many triggers used in Run 1 were tightened for the 2015 and 2016 data taking. Due to the higher multijet production rate, the single jet and H_{T} (the scalar sum of physics objects) trigger threshold were almost doubled. A different choice of triggers was done for the 2016 data analysis, not to penalize the signal efficiency.

Since signal events contain neutrinos coming from tau decays are characterized by large missing transverse momentum final states, pure $E_{\mathrm{T}}^{\mathrm{miss}}$ triggers or triggers are adopted that require $p_{\mathrm{T}}^{\mathrm{miss}}$ or $H_{\mathrm{T}}^{\mathrm{miss}}$ larger than 90 GeV, in combination with additional requirements, such as the presence of a jet with $p_{\mathrm{T}} > 80$ GeV.

The $E_{\mathrm{T}}^{\mathrm{miss}}$ triggers are the logic OR of different trigger quantities, with thresholds on both the $E_{\mathrm{T}}^{\mathrm{miss}}$ and the $H_{\mathrm{T}}^{\mathrm{miss}}$ computed using particle flow objects. The efficiency of the $E_{\mathrm{T}}^{\mathrm{miss}}$ triggers is measured by selecting $\mathrm{W} \to \mu\nu$ events using a trigger that requires the presence of an isolated muon of $p_{\mathrm{T}} > 24$ GeV. To ensure the single muon trigger efficiency, the requirement on the muon p_{T} is tightened to 30 GeV and the additional presence of a large-cone jet of $p_{\mathrm{T}} > 170$ GeV is required. And among these events, it is checked how many pass the OR of the triggers as a function of the missing transverse energy in the event. The turn-on curve for the $E_{\mathrm{T}}^{\mathrm{miss}}$ trigger used in this analysis is shown in Figure 7.2 as a function of the offline reconstructed missing transverse momentum. Simulated events are selected by requiring the missing energy to be higher than 200 GeV. Since at 200 GeV the $E_{\mathrm{T}}^{\mathrm{miss}}$ trigger is 95% efficient, the turn-on efficiency is applied as a weight to simulations, depending on the missing transverse energy in the event. A 2% systematic uncertainty from a fit to the trigger turn-on is considered as an additional systematic in analysis.

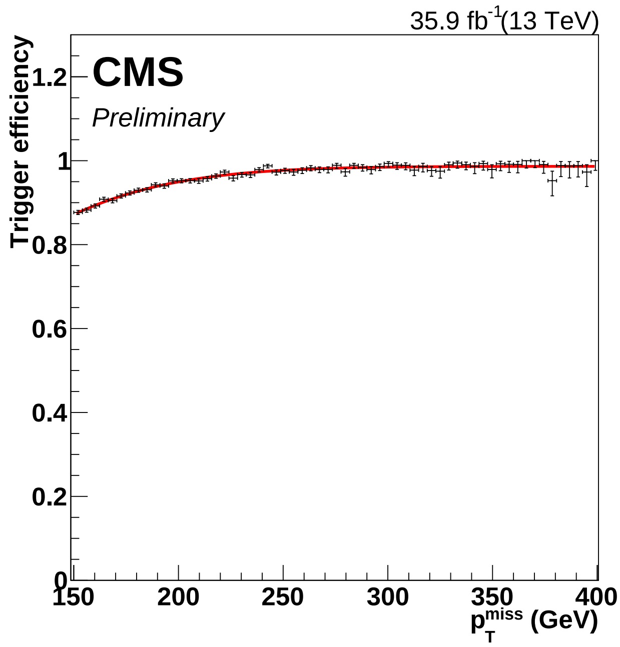


Figure 7.2: Trigger efficiency for the `OR` of the HLT paths as a function of the offline $E_{\mathrm{T}}^{\mathrm{miss}}$ in 2016 data events that pass the single muon trigger.

7.3 Event reconstruction and selection

In this section, a list of the physics objects used in the analysis is briefly presented, together with the modeling of the main properties of these objects and the simulation description of the data. A more extended description of the object reconstruction is provided in Chap.5.

Vertex and pileup

The reconstructed vertex with the largest value of summed physics-object p_{T}^{2} is taken to be the primary interaction vertex. The physics objects are the jets, clustered using the jet finding algorithm [164, 165] with the tracks assigned to the vertex as inputs, and the associated missing transverse momentum, taken as the negative vector sum of the p_{T} of those jets.

Although the simulation used in this analysis are generated with 25 ns bunch crossing scenario, the pileup description does not match exactly the one in data, so the simulations are reweighed to match the data, assuming an inelastic minimum bias cross section of 69.2 mb. Comparisons between the distributions of the primary vertices in data and simulations after the pileup reweighting procedure are shown in Fig. 7.3 for the event selection presented in Tab.7.1.

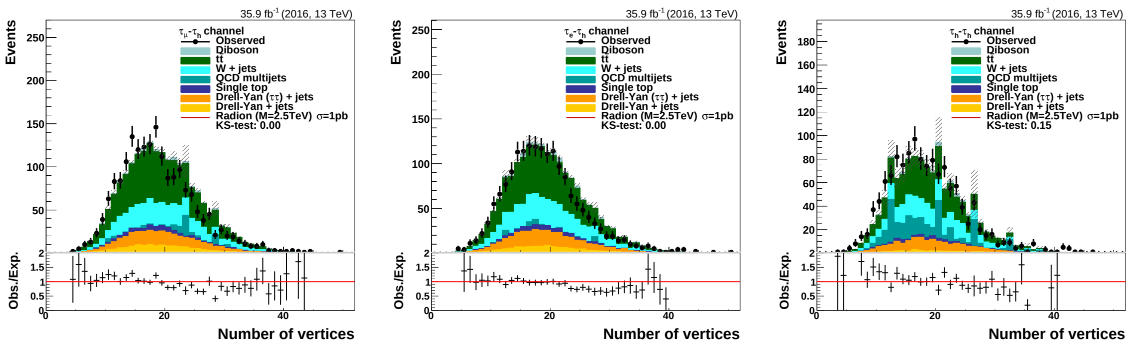


Figure 7.3: Disributions showin the number of vertices in data and simulations with PU corrections applied in $\mu\tau_{\mathrm{h}}$ (left), $e\tau_{\mathrm{h}}$ (center), and $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ (right) events.

Muon

Muons are identified with loose identification criteria and required to have $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ and $|\eta| < 2.4$. To avoid a loss of signal efficiency due to the proximity of possible products from the decay of the other tau lepton, muons are required to be isolated by imposing a limit on the magnitude of the p_T sum of all the PF candidates (excluding the muon) within $\Delta R < 0.4$ around the muon direction, after the contributions from particles associated with reconstructed τ_h candidates within the isolation cone are removed, as reported in Sec. 5.3.3. Data to simulation scale factors are used to correct the selection efficiency in the simulation in order to match the one in data.

Electron

Electrons are reconstructed and identified with veto selection requirements as described in Sec. 5.3.2, requiring $p_\mathrm{T} > 10\,\mathrm{GeV}$ and $|\eta| < 2.5$. The PF isolation is calculated considering the PF candidates within a cone of 0.3, after the contributions from pileup and particles associated with reconstructed τ_h candidates within the isolation cone are removed. To take into account differences in the selection efficiency between data and simulations, a set of scale factors are applied to simulations depending on the electron p_T and η.

Hadronically decaying τ

The hadronically decaying tau leptons are reconstructed and identified with the "boosted" technique described in Sec.5.3.7, not only in the $\tau_\mathrm{h}\tau_\mathrm{h}$ channel, but also in semileptonic $\mu\tau_\mathrm{h}$ and $e\tau_\mathrm{h}$ events. The τ_h candidates selected through the HPS algorithm are then required to have $|\eta| < 2.3$ and $p_\mathrm{T} > 20\,\mathrm{GeV}$, have a number of charged and neutral constituents consistent with a tau decay (decay modes) and to satisfy the multivariate-discriminator isolation requirement, in order to discriminate between genuine τ_h and quark- or gluon- initiated jets. If no τ_h candidates are identified with this method, then the procedure is repeated using AK4 jets as seeds, with similar selection requirements.

The τ_h candidate of highest p_T is required to satisfy a medium isolation requirement that corresponds to a 50–60% efficiency in the considered topology. If two τ_h candidates are identified, as in the $\tau_\mathrm{h}\tau_\mathrm{h}$ channel, the isolation requirement on the τ_h of second highest p_T is relaxed to achieve a 70–80% efficiency. The probability to misidentify a large-cone jet as an $\mathrm{H} \to \tau\tau$ decay is below 0.1% after these selection criteria.

Jets

For the reconstruction of the hadronically decaying boson, the approach described in Sec.5.3.5.1 is used. The AK8 jets are required to have $p_{\mathrm{T}} \geq 200$ GeV and $|\eta| \leq 2.4$ and satisfy tight quality criteria in order to remove spurious jet-like features originating from noise patterns in the calorimeters or the tracker. The CHS is used for the pileup mitigation for the kinematic variable of the jets, while PUPPI is used for the jet observables such as the soft-drop mass and the τ_{21} subjettiness ratio. The jet is considered as a boson jet candidate if the soft-drop m_{jet} falls in the range $[30, 250]$ GeV. As can be seen from Fig. 7.4, the soft-drop jet mass is peaked at the mass of the bosons for the signals.

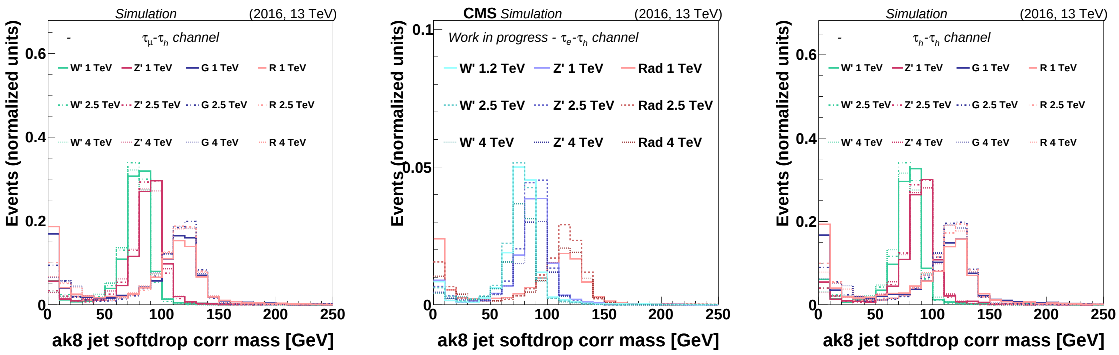


Figure 7.4: Jet soft-drop mass distributions for signals in the $\mu\tau_{\mathrm{h}}$ (left), $e\tau_{\mathrm{h}}$ (center), $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ (right) channels.

The two-prong hadronic decays of W and Z boson candidates are used to discriminate against jets initiated from single quarks and gluons. Small values of the tau τ_{21} correspond to a high compatibility with the hypothesis that the jet is produced by two partons from the decay of a massive object, rather than arising from a single parton. The distributions of the τ_{21} for the boson jet candidate in the signal simulations are shown in Fig.7.5. V boson jet candidates with $\tau_{21} \geq 0.75$ are rejected, while other candidates are categorized into high-purity (HP) jets with $\tau_{21} \leq 0.4$ and low-purity (LP) jets with $0.4 \leq \tau_{21} \leq 0.75$ in order to enhance the sensitivity of the analysis.

Jets originating from the dominant $\mathrm{b\overline{b}}$ decays of Higgs bosons are likely to have two displaced vertices because of the long lifetime and large mass of the b quarks. The inclusive, combined secondary-vertex b tagging algorithm [127] is applied to the two subjets, which are considered as b-tagged if they pass a working point that provides a misidentification rate of ≈10% while maintaining an 85% efficiency. Higgs candidates

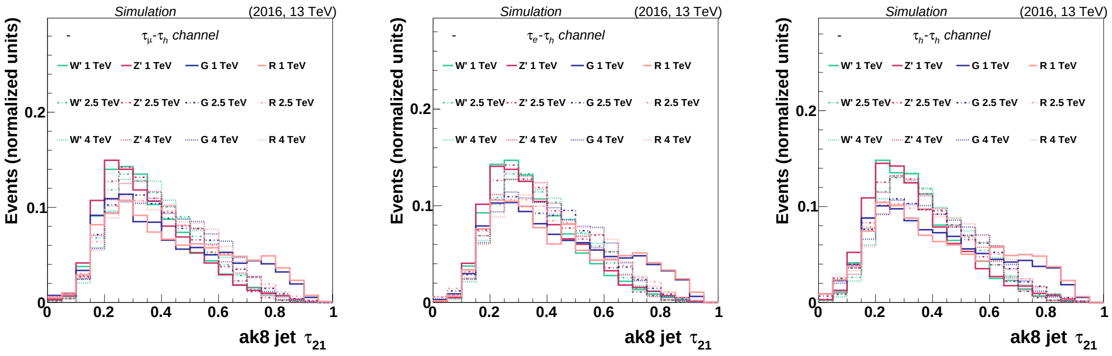


Figure 7.5: τ_{21} distributions for signals in the $\mu\tau_\mathrm{h}$ (left), $e\tau_\mathrm{h}$ (center) channels, $\tau_\mathrm{h}\tau_\mathrm{h}$ (right) channels.

are divided into 2 categories in order to enhance the sensitivity of the analysis: events with 2 b-tagged subjets and events with 1 b-tagged subjet.

To remove backgrounds containing top-quark decays, events with AK4 jets that do not overlap with the AK8 jet and the identified leptons are subjected to a veto based on the same b tagging algorithm, but with a working point corresponding to an efficiency of ≈70% for identifying jets originating from b quarks and a ≈1% misidentification rate.

Missing Transverse Energy

To account for the presence of neutrinos, the particle flow $E_\mathrm{T}^\mathrm{miss}$ is considered, computed as the negative vector sum of transverse momenta of all the PF candidates, as reported in Sec.5.3.6. To ensure the full efficiency of the trigger requirements, events are selected with a minimum requirement of $E_\mathrm{T}^\mathrm{miss} > 200\,\mathrm{GeV}$.

Higgs to $\tau\tau$ reconstruction

In order to reconstruct the kinematics of the H boson in its $\tau\tau$ decay the SVfit algorithm [141–143] is used. As explained in Sec. 5.3.8, the SVfit algorithm is based on a likelihood approach and estimates the di-τ system mass using the measured momenta of the visible decay products of both τ leptons, the reconstructed $\vec{p}_\mathrm{T}^\mathrm{miss}$, and the $\vec{p}_\mathrm{T}^\mathrm{miss}$ resolution, obtained from the $E_\mathrm{T}^\mathrm{miss}$ covariance matrix as explained in Sec. 5.3.6.

7.3.1 Comparison of data and simulations with loose selection

In this section, the main kinematical variables of jets and leptons are presented when a loose inclusive selection, in Tab. 7.1, is applied to events in data and simulations, in order to identify the main backgrounds and the level of agreement of the simulation description with data.

	$e\tau_{\mathrm{h}}$	$\mu\tau_{\mathrm{h}}$	$\tau_{\mathrm{h}}\tau_{\mathrm{h}}$
Leading τ_h	boosted or standard τ_h DMs, VL MVA isolation $p_{\mathrm{T}}(\tau_{\mathrm{h}}) > 20\,\mathrm{GeV}$, $\lvert\eta\rvert(\tau_{\mathrm{h}}) < 2.3$		
Second lepton	e: Veto cut-based ID correcter iso. veto $p_{\mathrm{T}}(e) > 10\,\mathrm{GeV}$ $\lvert\eta\rvert(e) < 2.5$	μ:L ID corrected iso. L $p_{\mathrm{T}}(\mu) > 10\,\mathrm{GeV}$ $\lvert\eta\rvert(\mu) < 2.4$	τ_{h}: new DMs MVA isolation VL $p_{\mathrm{T}}(\tau_{\mathrm{h}}) > 20\,\mathrm{GeV}$ $\lvert\eta\rvert(\tau_{\mathrm{h}}) < 2.3$
Jet	$p_{\mathrm{T}} > 200\,\mathrm{GeV}$, $\lvert\eta\rvert < 2.4$		
$E_{\mathrm{T}}^{\mathbf{miss}}$	$> 200\,\mathrm{GeV}$		
N(Medium b-tagged AK4 jets(p_{T} >20 GeV))	< 1		
V-**tagging**	—		
H(bb)-tagging	—		

Table 7.1: Summary of the event loose selection requirements.

In Figures 7.6 - 7.15 the comparison is shown for $\mu\tau_{\mathrm{h}}$, $e\tau_{\mathrm{h}}$, and $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ events for the loose selection selection. Data to simulation corrective factors are applied. Neither τ_{21} nor b-tagging requirements are applied, to ensure that there is negligible signal contamination in the region.

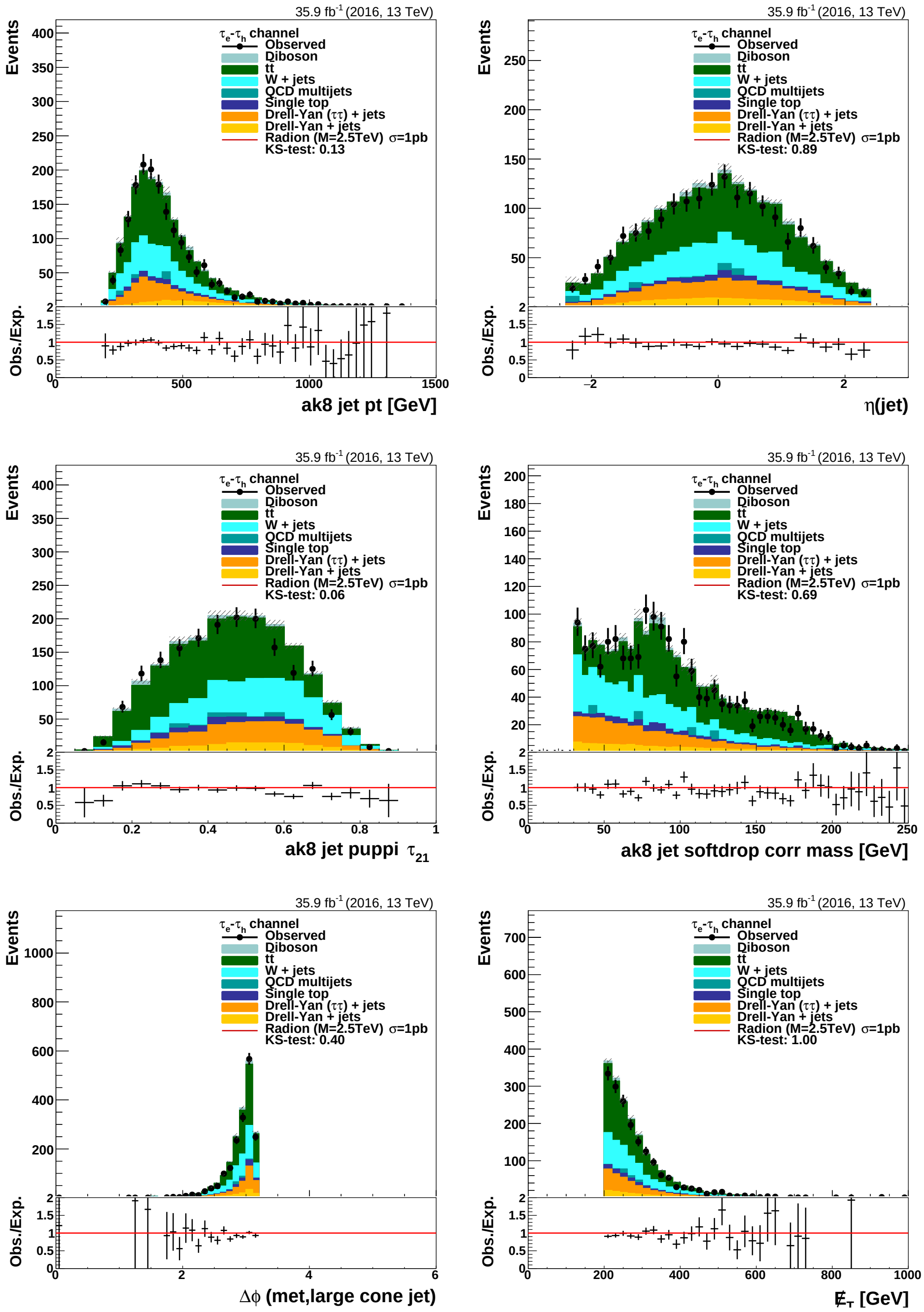


Figure 7.6: Comparison between data and expected simulated events for the $e\tau_\mathrm{h}$ channel for the following variables: jet p_T and η (top), jet τ_{21} and soft-drop mass distributions (middle), angular distance between the large cone jet and missing transverse momentum vector and missing transverse energy (bottom).

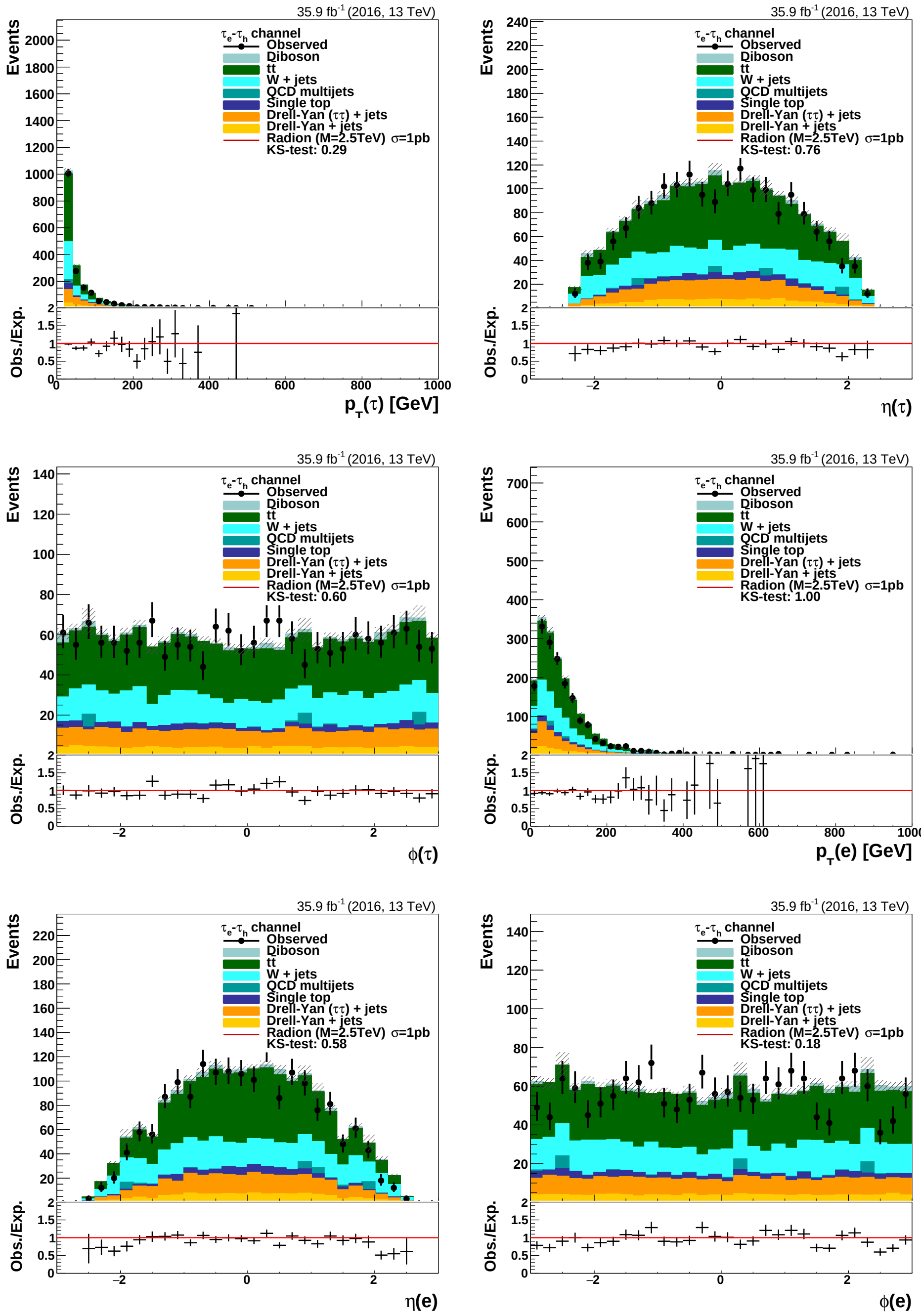


Figure 7.7: Comparison between data and expected simulated events for the $e\tau_{\mathrm{h}}$ channel for the following variables: τ_{h} p_{T} and η (top), τ_{h} ϕ and electron p_{T} distributions (middle), electron η and ϕ (bottom).

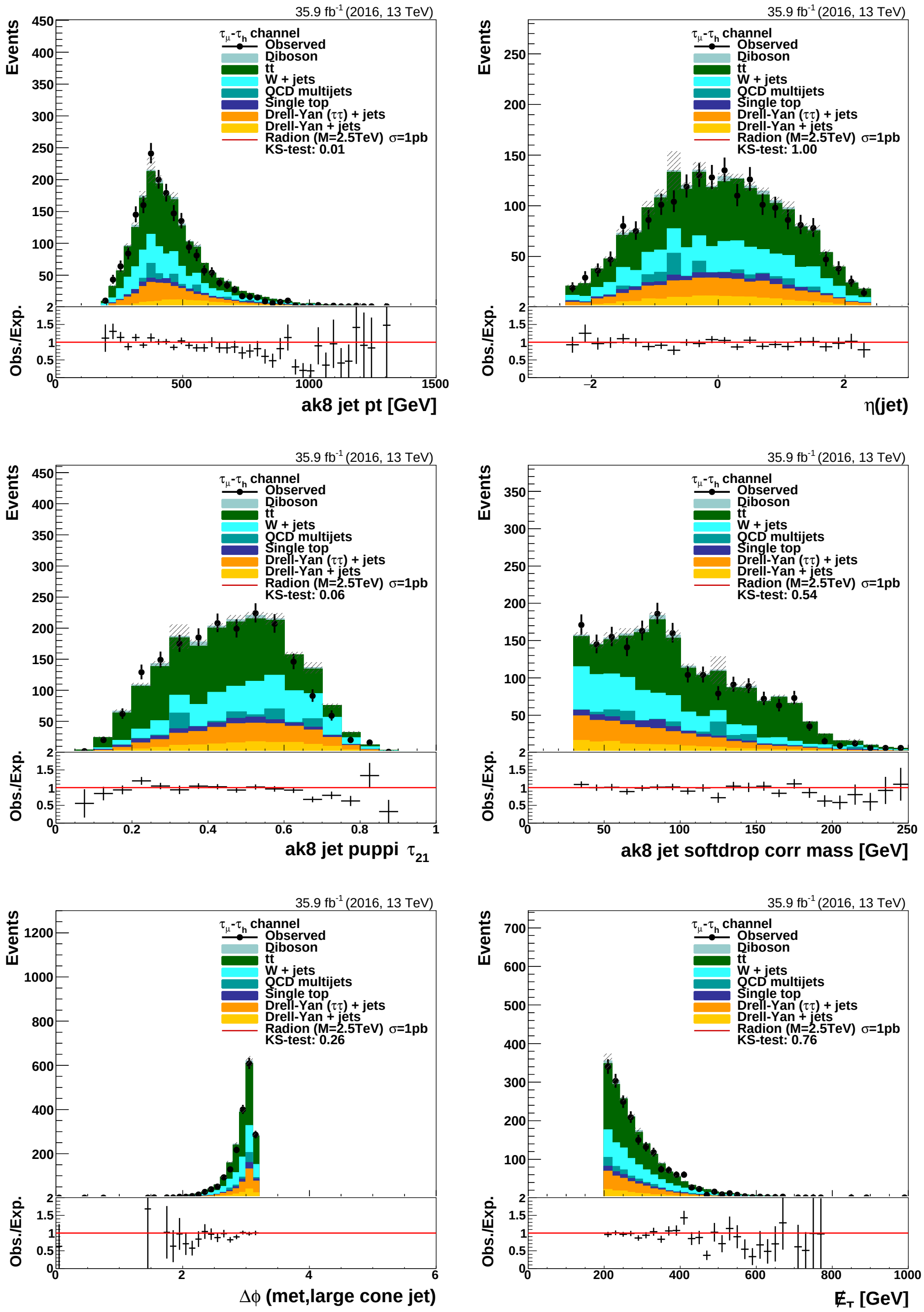


Figure 7.8: Comparison between data and expected simulated events for the $\mu\tau_{\mathrm{h}}$ channel for the following variables: jet p_{T} and η (top), jet τ_{21} and soft-drop mass distributions (middle), angular distance between the large cone jet and missing transverse momentum vector and missing transverse energy (bottom).

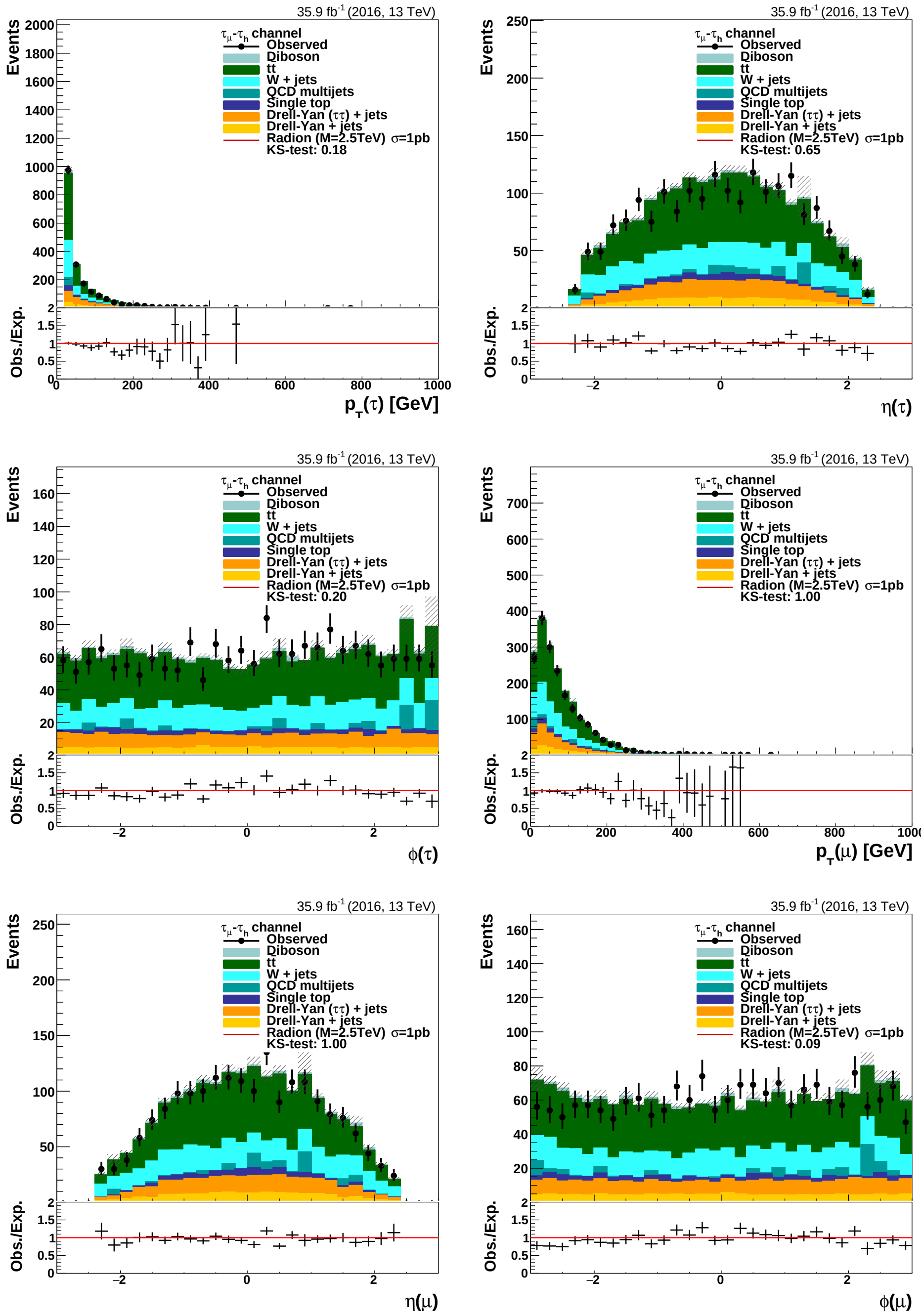


Figure 7.9: Comparison between data and expected simulated events for the $\mu\tau_\mathrm{h}$ channel for the following variables: tau p_T and η (top), tau ϕ and muon p_T distributions (middle), muon η and ϕ (bottom).

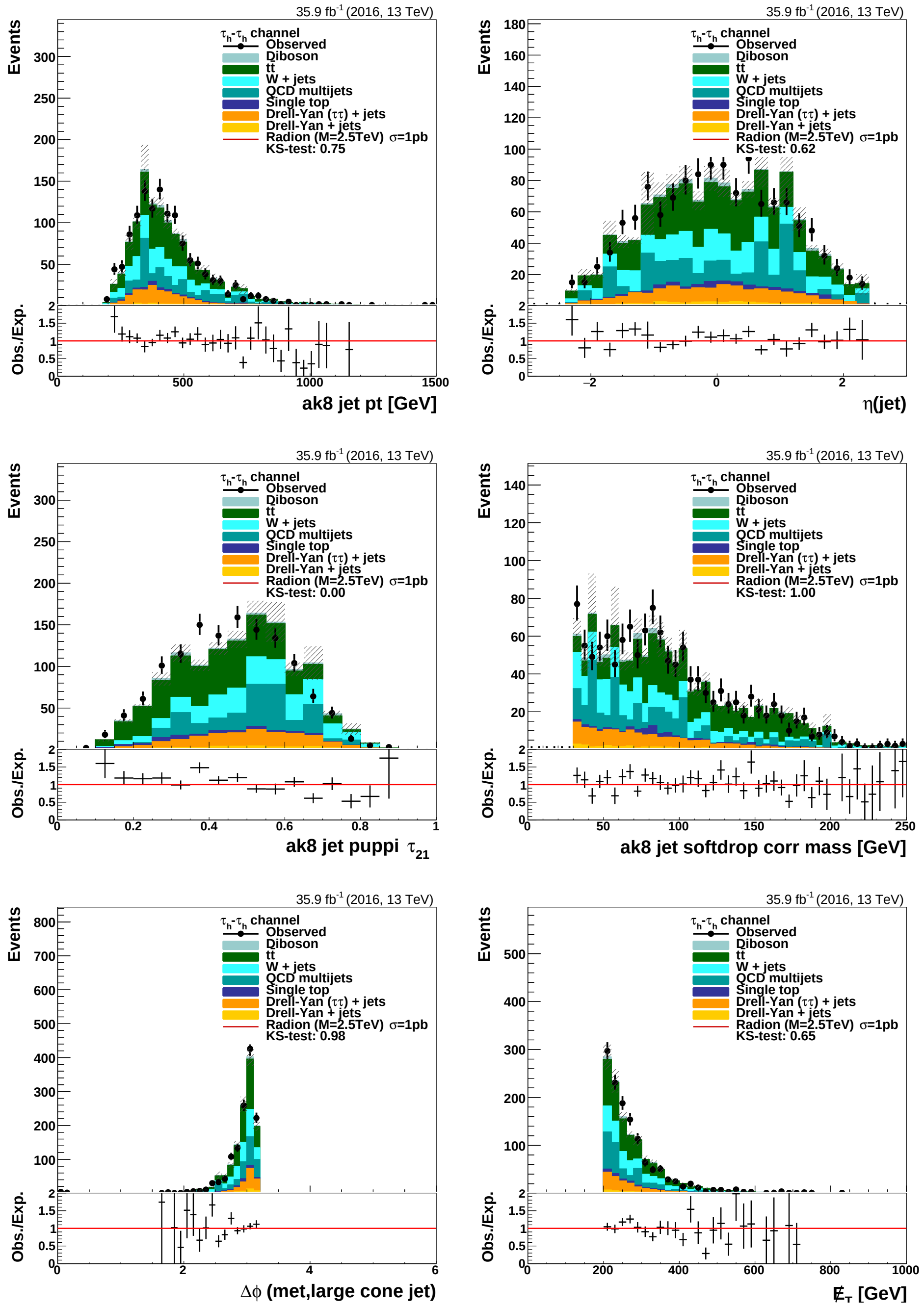


Figure 7.10: Comparison between data and expected simulated events for the $\tau_\mathrm{h} - \tau_\mathrm{h}$ channel for the following variables: jet p_T and η (top), jet τ_{12} and soft-drop mass distributions (middle), angular distance between the large cone jet and missing transverse momentum vector and missing transverse energy (bottom).

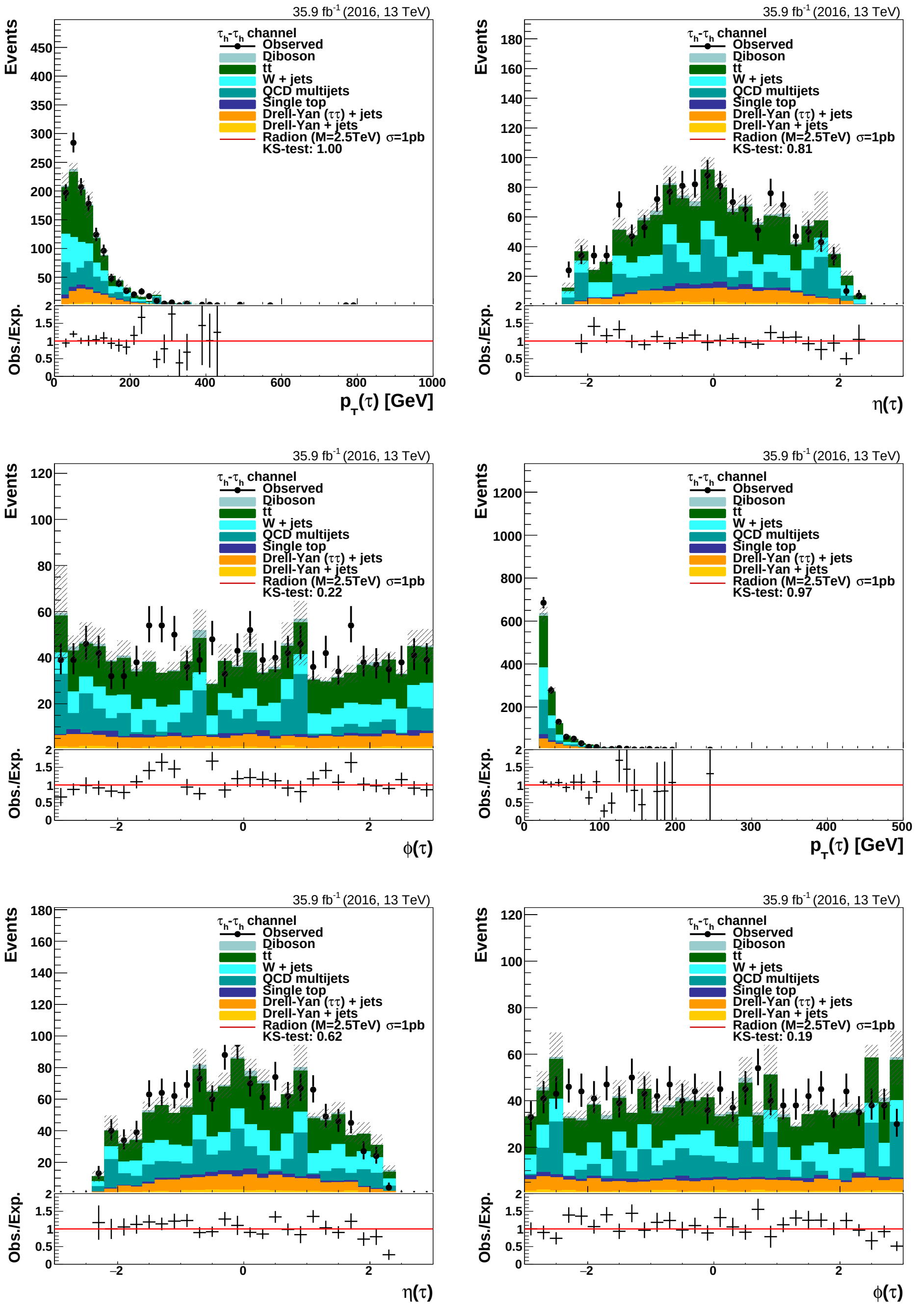


Figure 7.11: Comparison between data and expected simulated events for the $\tau_h\tau_h$ channel for the following variables: leading τ_h p_T and η (top), leading τ_h ϕ and second leading τ_h p_T distributions (middle), second leading τ_h η and ϕ (bottom).

7.3.2 Semileptonic channel $\ell\tau_h$

In the analysis, $e\tau_h$ and $\mu\tau_h$ events are merged and analyzed together in the so-called semileptonic channel $\ell\tau_h$. In this way the background estimation profits from higher statistics in the signal and control regions. This is possible because the muons and electrons are selected with the same kinematic requirements. As shown for the inclusive selection in Figs.7.12–7.13, in the combined channel there are no discontinuities due to different selection criteria and background compositions, for the main variables used in the analysis that are the jet mass and the resonance mass.

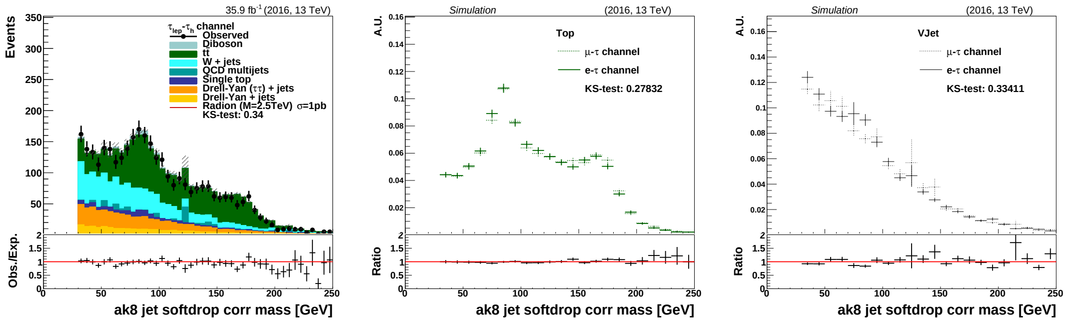


Figure 7.12: Comparison of data and simulated distribution of the jet soft-drop mass (left) and comparison of this distribution between the channels for the main background components: Top (including $t\bar{t}$ and single-t) (center) and V+jets (Drell-Yan, W+jets, QCD and Diboson) (right). Inclusive selection is applied.

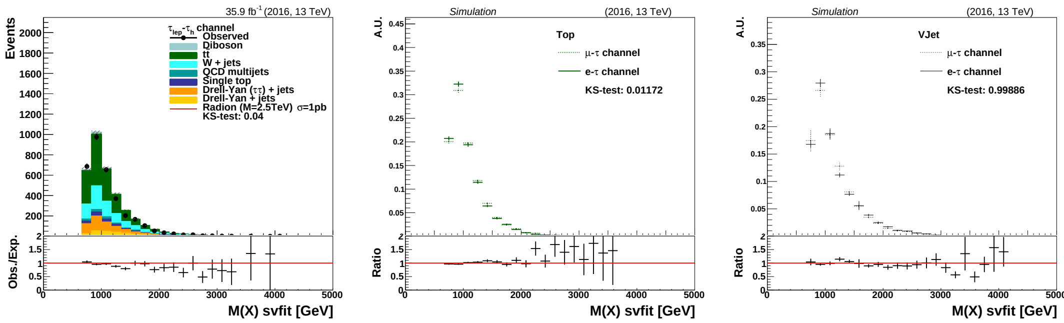


Figure 7.13: Comparison of data and simulated distribution of the resonance mass (left) and comparison of this distribution between the channels for the main background components: Top (including $t\bar{t}$ and single-t) (center) and V+jets (Drell-Yan, W+jets, QCD and Diboson) (right). Inclusive selection is applied.

7.3.3 Final selection and analysis regions

In this section the final selection requirements used in the analysis to provide optimal signal sensitivity are summarized. The different analysis signal and control regions are also described.

To reject events where the τ_h is mimicked by a jet, the leading τ_h is required to satisfy the MVA-based isolation requirement with the medium working point.

Several selection requirements are applied to remove SM backgrounds, such as meson and baryon resonances, Z+jets, W+jets, and $\mathrm{t\bar{t}}$ and single top quark production. The angular distance $\Delta R_{\tau\tau}$ should be smaller than 1.5, in order to reject W+jets events in which a jet misidentified as a τ lepton is typically spatially well-separated from the genuine lepton, as shown in the upper plots of Figs. 7.14–7.16. To further increase the signal purity, the di-τ mass, as estimated from the SVFIT algorithm, should be between 50 and 150 GeV, as shown in the lower plots of Figs. 7.14–7.16. Events with top quark pairs or single top quarks are suppressed by removing events in which any AK4 jet not overlapping with the AK8 jet is b-tagged, as shown in the middle plots of Figs. 7.14–7.16.

Events that pass these selection requirements are further divided into categories depending on the AK8 jet soft-drop mass: the soft-drop jet mass must be in the interval of 30–250 GeV. If the mass is in the range 65–85 GeV, the candidate is classified as a W boson, if it is in the range 85–105 GeV it is classified as a Z boson, and if it is the range 105–135 GeV it is considered to be a Higgs boson. Events with soft-drop mass smaller than 65 GeV or greater than 135 GeV are used as control regions (mass sidebands (SB)) for the background estimation. A jet is V tagged if it fulfills the soft-drop jet mass and τ_{21} requirements. Higgs boson jet candidates are classified according to the number of subjets (1 or 2) that pass the b tagging selection. Subjet b tagging is not used for jets compatible with W or Z candidates and no N-subjettiness requirement is applied to the Higgs boson candidate jet. If neither the N-subjettiness nor the b tagging requirements are satisfied, the event is rejected.

Finally, the resonance candidate mass m_X, defined as the invariant mass of the $\mathrm{H} \to \tau\tau$ candidate and the hadronically decaying boson jet, is required to be larger than 750 GeV in order to ensure full reconstruction efficiencies.

The final selection requirements are listed in Tab. 7.2, while the signal and control regions considered in the analysis are summarized in Tab 7.3.

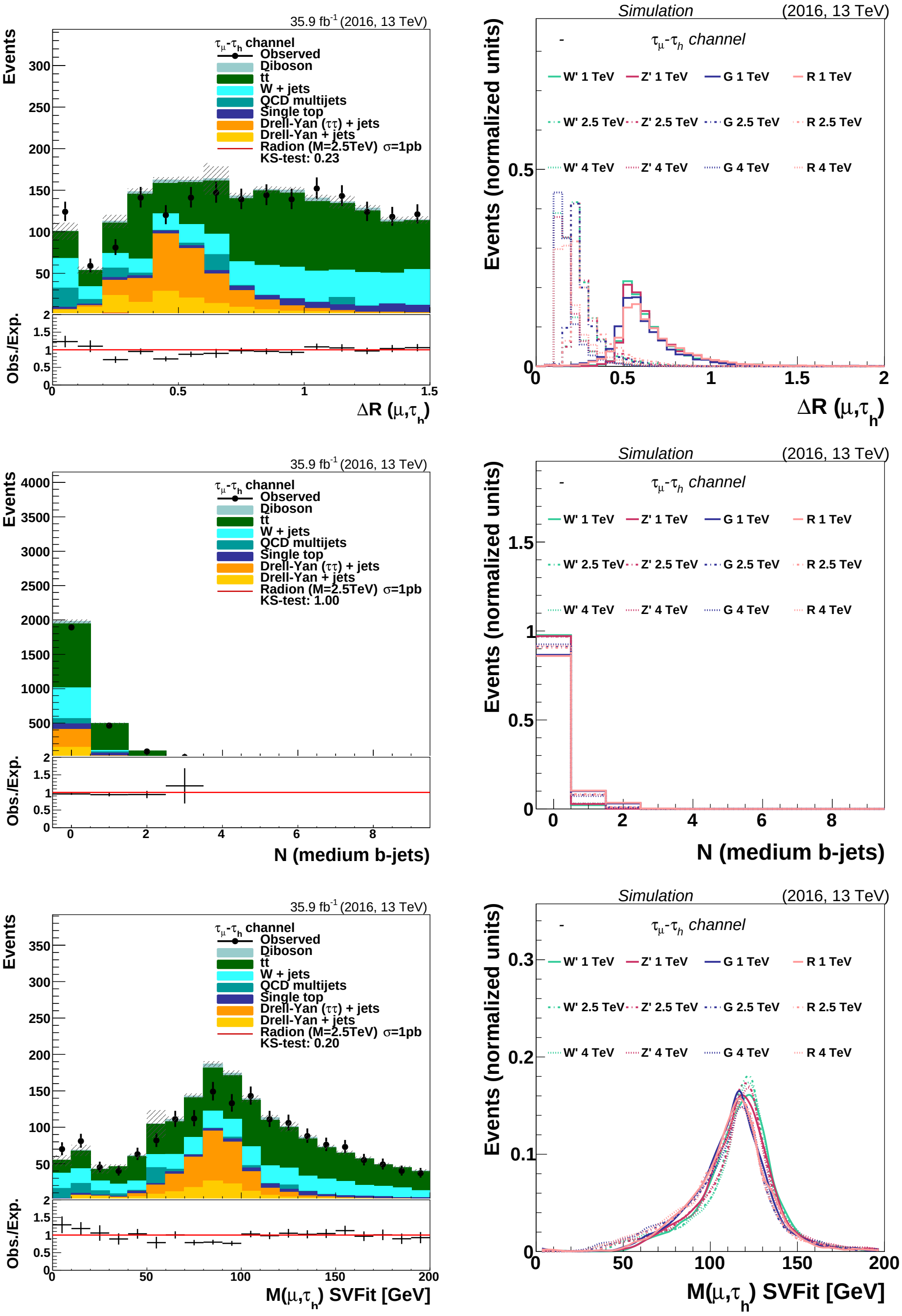


Figure 7.14: Comparison for the $\mu\tau_{\mathrm{h}}$ channel for background and signal processes of the following variables: distance between the lepton and the hadronic tau (top), multiplicity of b-tagged AK4 jets (Medium CSVv2)(middle) and SVFIT-reconstructed Higgs boson mass (bottom).

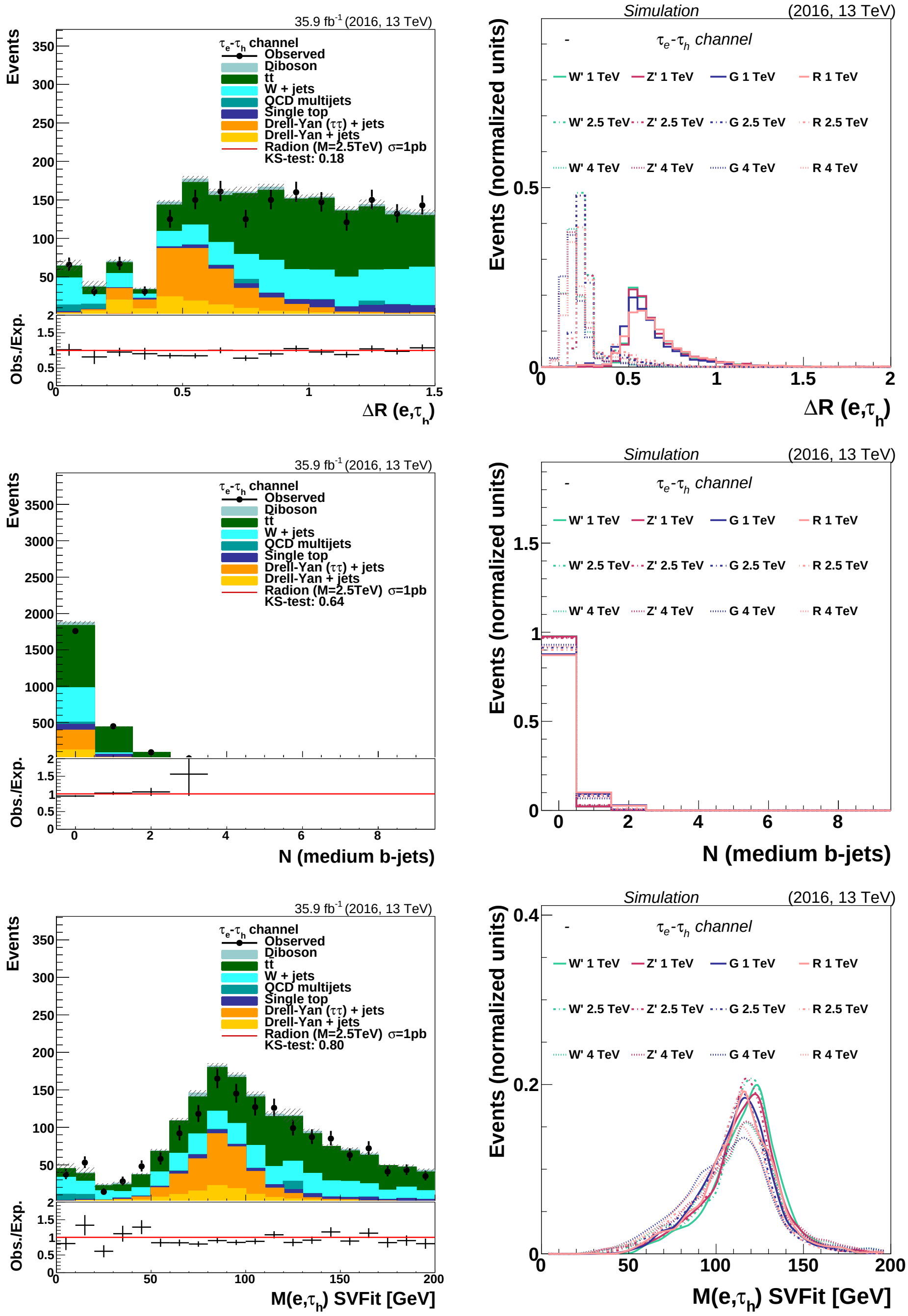


Figure 7.15: Comparison for the $e\tau_h$ channel for background and signal processes of the following variables: distance between the lepton and the hadronic tau (top), multiplicity of b-tagged AK4 jets (Medium CSVv2)(middle) and SVFIT-reconstructed Higgs boson mass (bottom).

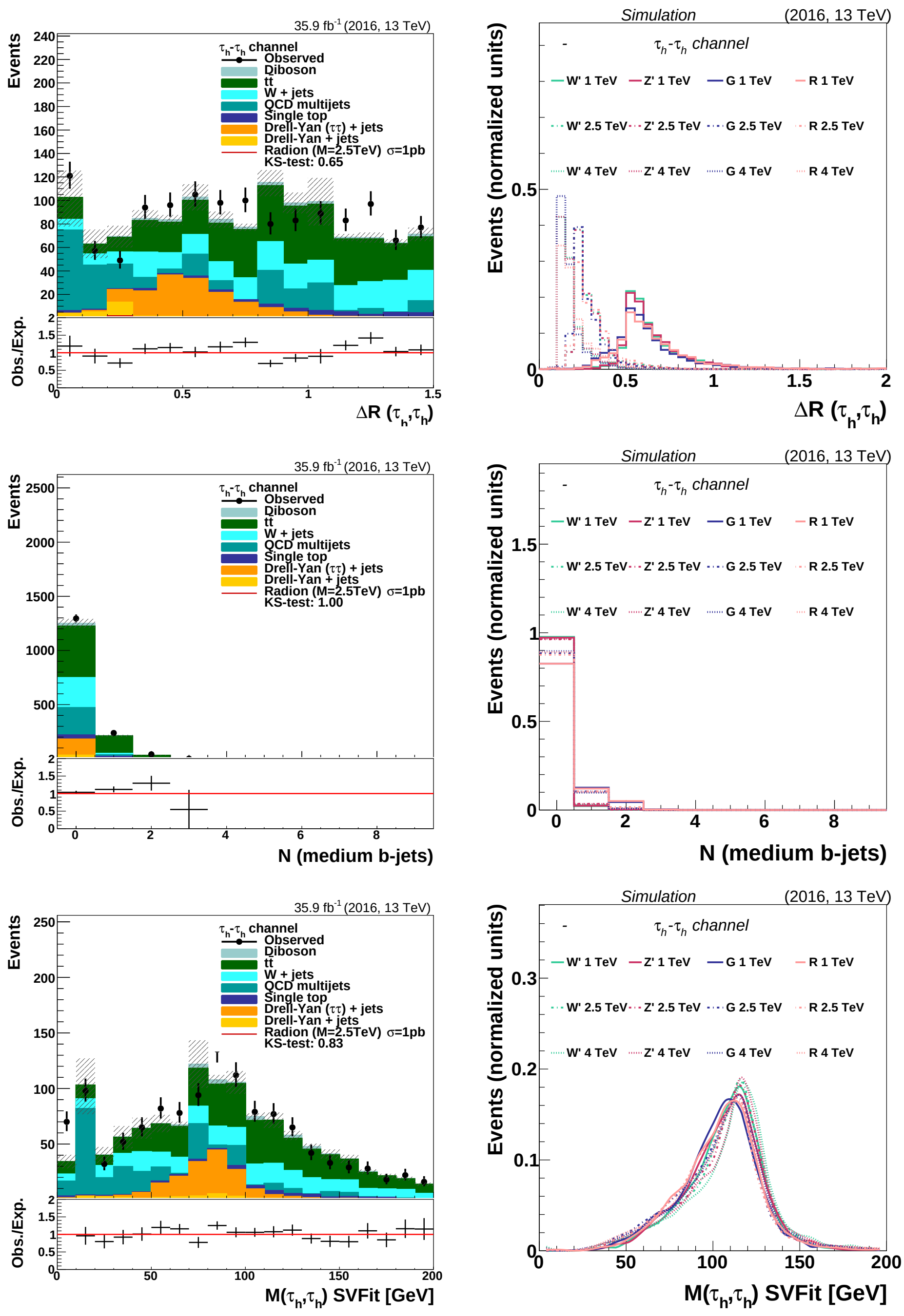


Figure 7.16: Comparison for the $\tau_h\tau_h$ channel for background and signal processes of the following variables: distance between the lepton and the hadronic tau (top), multiplicity of b-tagged AK4 jets (Medium CSVv2)(middle) and SVFIT-reconstructed Higgs boson mass (bottom).

	$e\tau_h$ channels	$\mu\tau_h$ channel	$\tau_h\tau_h$ channel
Leading τ_h	boosted or standard τ_h Medium MVA isolation $p_T(\tau) > 20\,\text{GeV}$, $\lvert\eta(\tau)\rvert < 2.3$		
Second lepton	e: Veto cut-based ID correcter iso. veto $p_T(e) > 10\,\text{GeV}$ $\lvert\eta(e)\rvert < 2.5$	μ:L ID corrected iso. L $p_T(\mu) > 10\,\text{GeV}$ $\lvert\eta(\mu)\rvert < 2.4$	τ_h: new DMs MVA isolation VL $p_T(\tau_h) > 20\,\text{GeV}$ $\lvert\eta(\tau_h)\rvert < 2.3$
Jet	$p_T > 200\,\text{GeV}$, $\lvert\eta\rvert < 2.4$		
E_T^{miss}	$> 200\,\text{GeV}$		
$\Delta R(\ell, \ell)$	< 1.5		
H **(di-τ)**	$50\,\text{GeV} < \text{MassSVFIT}(\tau, \tau) < 150\,\text{GeV}$		
N(Medium b-tagged AK4 jets(p_T >20 GeV))	< 1		
V-tagging	τ_{21} <0.40 0.40< τ_{21} <0.75		
H(bb)-tagging	1 CSVL b-tagged subjet 2 CSVL b-tagged subjets		

Table 7.2: Summary of the final selection requirements applied in the analysis.

Table 7.3: Summary of the analysis regions. The selection requirements applied to the events in the control regions, i.e. V tagging or H tagging, depend on the kind of signal under consideration.

Category	soft-drop jet mass window (GeV)				
	[30,65]	[65,85]	[85,105]	[105,135]	[135,250]
HP	CR	$\tau_{21} < 0.4$	$\tau_{21} < 0.40$	2 b-tagged subjets CSVv2 L	CR
LP	CR	0.4< τ_{21} <0.75	0.40 < τ_{21} <0.75	1 b-tagged subjet CSVv2 L	CR
Signals		W′	Z′	radion/graviton	

7.4 Background estimation

The main sources of background events originate from top quark pair production and from the production of a vector boson in association with jets, while minor contributions arise from single top quark, diboson, and multijet production. In the background estimation, the background contribution of W+jets and Z+jets are considered together in the V+jets component because with this analysis selection, the large cone jets in those events are likely to be quark- or gluon-initiated jets. Instead events originating from either $\mathrm{t\bar{t}}$ and single top quark production ($\mathrm{t\bar{t}}$, t) likely have jets that contain the entire top quark decay or a W boson from a top quark decay with a genuine 3-prong or 2-prong substructure. Since the jet mass spectrum of these background components is different, it is possible to disentangle them and obtain the V+jets prediction through the soft-drop jet mass fit in data. The diboson background gives just a minor contribution and is added to the V+jets background and estimated from data. The QCD multijet contribution is also rather small due to the requirements on the $p_{\mathrm{T}}^{\mathrm{miss}}$ and on isolated leptons and τ_{h} and it is added to the V+jets background and estimated from data. The background contributions are thus split into either the $\mathrm{t\bar{t}}$, t processed, or into V+jets production. The first is estimated from simulation and validated in control regions in data enriched in $\mathrm{t\bar{t}}$ events. The latter includes Z+jets and W+jets, multijet, and SM diboson production.

The α-ratio method is used as explained in Sec. 7.4.2 to obtain both the number of events and the resonance mass spectrum expected of the V+jets background in the signal regions. Due to poor statistics after the whole selection, the electron and muon semileptonic channels are merged together into the $\ell\tau_{\mathrm{h}}$ channel for the background estimation, while the fully hadronic channel is considered on its own because of the different kinematic requirements on the second τ_{h} candidate.

7.4.1 $\mathrm{t\bar{t}}$ and single top quark production estimation

The shape of the distribution of the top quark pair and single top quark background is determined from simulation for both the jet mass and resonance mass modeling. Especially after the b-tagging requirement, there is a significant contamination of $\mathrm{t\bar{t}}$ and single top quark production events. Thus, the top background description has to be validated on data first. Control regions having a purity larger than 80% for top quarks are selected by inverting the b tag veto on the AK4 jets and tightening the b tagging criteria. Events are separated according to the requirements of large-cone jet identification. Data are found to be well-described by simulations in terms of the jet and dijet resonance mass distributions, as shown in Figs. 7.17, for the events in the top-enriched region when the LP τ_{21} requirements is applied.

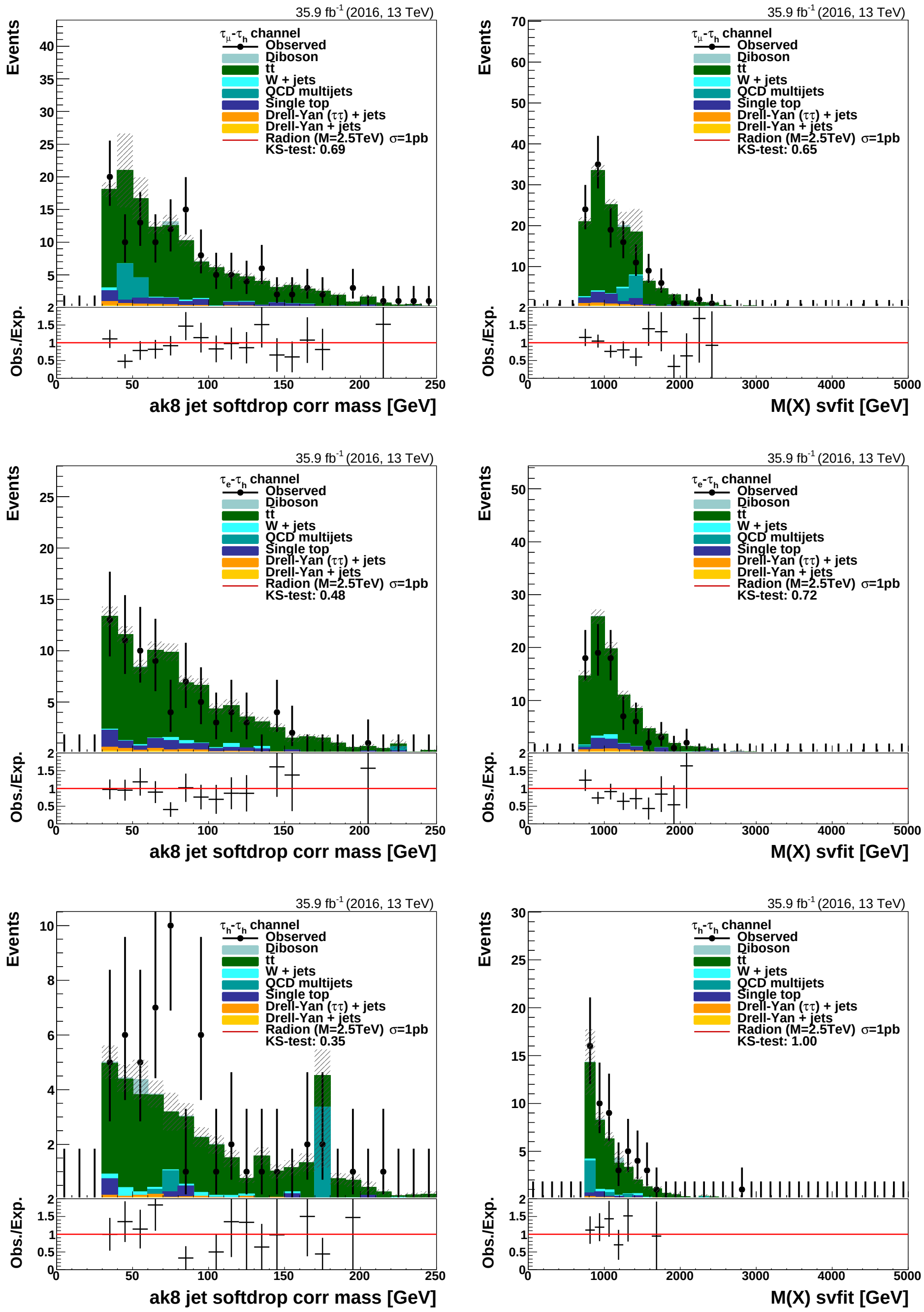


Figure 7.17: Comparison between simulation and data distributions of the jet soft-drop mass (left) and resonance mass (right) are shown in the top-quark control region in the $\mu\tau_h$ (upper row), $e\tau_h$ (middle), and $\tau_h\tau_h$ (lower row) channels. τ_{21} LP selection applied.

Multiplicative scale factors are derived in each category to correct for the difference in normalization of data and simulation in the control regions, after subtracting the other background contributions. Scale factors obtained in control regions of $\ell\tau_{\mathrm{h}}$ events are applied also to the $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ channels, where there are fewer events. The normalization of top quark production processes in each region is also corrected for using the scale factors reported in Table 7.4. The top quark background scale factors are affected by the statistical uncertainty in data, and by systematic uncertainties from the event reconstruction and modeling from simulation. These effects account for the lepton and b tag efficiency uncertainty.

Table 7.4: Normalization scale factors for top quark production for different event categories, depending on the V tagging and H tagging requirement applied. Uncertainties are due to the limited number of events in the control regions and the uncertainty in the b tagging efficiency.

Channel	τ_{21} LP	τ_{21} HP	1 b-tagged subjet	2 b-tagged subjets
$\ell\tau_{\mathrm{h}}$	0.96 ± 0.04	1.06 ± 0.06	1.00 ± 0.06	1.11 ± 0.15

7.4.2 α-ratio background estimation

The aim of the analysis is to look for localized excesses in the resonance mass spectrum m_X. The α-ratio method is a background estimation technique used since Run 1 [166], to rely marginally on simulation for the background estimation, due to the many sources of systematic uncertainties that are hard to understand and control in the boosted regime. Two exclusive regions, named the *signal region* (SR) and the *sideband region* (SB), are defined in order to select a signal-enriched or signal-depleted phase space, respectively. First, the background normalization is extracted from data in the SB. Then, the alpha method predicts the shape of the data in the SR starting from the distribution of the data in the SB, using a transfer function (the α function) derived from simulations. This method relies on the assumption that the correlation between m_X and the soft-drop jet mass is reasonably well-reproduced by simulations. The α ratio is considered to be more trustworthy than a pure simulation-based background prediction because systematic uncertainties would approximately cancel in the ratio.

The shape and normalization of the $\mathrm{t\bar{t}}$ and single top quark production is taken from the simulation with corrective normalization factors from control regions in data. The shape and normalization of the main V+jets background are evaluated with the α-approach. The jet soft-drop mass variable is used to perform the normalization prediction and the resonance mass variable is used for the shape prediction. A different background prediction is derived for each category separately and it is calculated in the resonance mass range from 850 to 4000 GeV.

7.4.2.1 Background normalization

The background normalization is the goal of the first step of the background prediction. The backgrounds are split in two categories: V+jets, and $\mathrm{t\bar{t}}$ or single top quark production. The estimated contribution from the V+jets background is based on data, in regions defined by applying the complete signal selection apart from the jet mass requirements. Two jet mass sidebands are defined with jet masses in the range of 30–65 GeV for the low sideband (LSB), or above 135 GeV for the high sideband (HSB), and used to predict the background contribution in the signal regions.

The V+jets and $\mathrm{t\bar{t}}$ background components have different shapes in the jet mass distribution and are described with functional forms determined by fits on the simulated backgrounds. Since for the V+jets component the jet is likely to originate from a single quark or gluon, the shape of the jet soft-drop mass is smoothly decreasing, even if a requirement on the τ_{21} such as the one of the HP category, can modify the shape of the spectrum eliminating events towards smaller jet mass values. The $\mathrm{t\bar{t}}$, t component shows a smoothly decreasing spectrum with two peaks in the proximity of the W boson and top quark masses, for the jets that originate from single quarks, merged W and t quarks, respectively. The $\mathrm{t\bar{t}}$ component, shown in Figs. 7.19–7.20, is normalized to the expected yields from simulation, with the corrective factors measured in the data regions enriched in $\mathrm{t\bar{t}}$ and single top quark production events, reported in Tab. 7.4.

The shape of the jet soft-drop mass distribution for the V+jets background component is fitted with two functions: a main one, used to extract the number of V+jets events in the signal region, and a second alternative function, used to estimate a systematic uncertainty due to the choice of the functional form. In Fig.7.18 indicative fits for the main (left) and alternative (right) functions are shown for the HP τ_{21} category of $\ell\tau_{\mathrm{h}}$ events.

The functional forms chosen to represent the jet soft-drop mass (m_j) templates are:

- **Exp-** an exponential function: $F_{\mathrm{Exp}}(x) = e^{ax}$
- **Pol-** a third order polynomial: $F_{\mathrm{Pol}}(x) = a_0 \cdot x + a_2 \cdot x^2 + a_3 \cdot x^3$
- **ErfExp-** a modification of the standard "error function", that is multiplied by an exponential function: $F_{\mathrm{ErfExp}}(x) = e^{ax} \cdot \frac{1+\mathrm{Erf}((x-b)/w)}{2}$
- **Gaus2-** two gaussians: $F_{\mathrm{Gaus2}}(x) = f_0 \cdot e^{2(x-a)^2/b} + (1-f_0) \cdot e^{2(x-c)^2/d}$
- **Gaus3-** three gaussians:
$F_{\mathrm{Gaus3}}(x) = f_0 \cdot e^{2(x-a)^2/b} + f_1 \cdot e^{2(x-c)^2/d} + (1-f_0-f_1) \cdot e^{2(x-e)^2/g}$

- **ExpGaus2**- an exponential plus two gaussians:
  $F_{\text{ExpGaus2}}(x) = f_0 \cdot e^{ax} + f_1 \cdot e^{2(x-b)^2/c} + (1 - f_0 - f_1) \cdot e^{2(x-d)^2/e}$
- **Voig**- a Voigtian function, which is the convolution of a gaussian and a Lorentzian distribution: $F_{\text{Voig}}(x) = f_0 \cdot \int_{-\infty}^{\infty} Gaus(x', \mu, \sigma) Lorentzian(x', \mu, \gamma) dx'$
- **CrysBall**- a Crystal ball function, i.e. a gaussian core with an exponential tail (controlled by the n parameter) that starts at the α-th sigma of the gaussian.
- **Gaus2Voig**- a Voigtian function plus two gaussians:
  $F_{\text{Gaus2Voig}}(x) = f_0 \cdot F_{\text{Voig}}(x) + f_1 \cdot e^{2(x-b)^2/c} + (1 - f_0 - f_1) \cdot e^{2(x-d)^2/e}$

The choice of the functions is channel-dependent, chosen based on a χ^2 figure of merit, and it depends on the background shape and the available statistics and is summarized in Table 7.5.

category		V+jets	alt. V+jets	$t\bar{t}$
τ_{21} HP	$\ell\tau_h$	ErfExp	Voig	Gaus3
	$\tau_h\tau_h$	ErfExp	CrysBall	Gaus2Voig
τ_{21} LP	$\ell\tau_h$	CrysBall	Pol	ExpGaus2
	$\tau_h\tau_h$	Pol	Exp	ExpGaus2
2 b tag	$\ell\tau_h$	Exp	Pol	Gaus2
	$\tau_h\tau_h$	Pol	Exp	Gaus3
1 b tag	$\ell\tau_h$	Pol	Exp	Gaus3
	$\tau_h\tau_h$	Pol	Exp	Gaus2

Table 7.5: Functional form used to model the jet mass distribution for each category.

The $t\bar{t}$ and V+jets templates are summed together, maintaining their relative weights and finally fitted to the jet soft-drop mass spectrum m_j of data in the SBs.

The number of expected events in the SR is then obtained by integrating the background components in the jet mass window of each signal region. The procedure is repeated for the alternative functions used for modeling the V+jets jet mass, and the observed difference in the normalization is taken to be the associated systematic uncertainty.

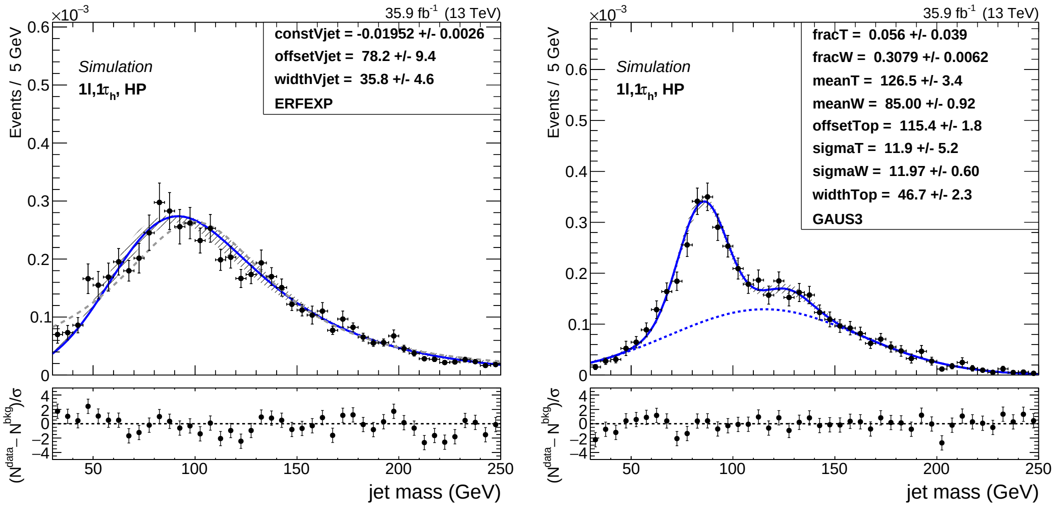


Figure 7.18: Fit to the simulated m_j in $\ell\tau_{\mathrm{h}}$ events that satisfy the τ_{21} HP requirements, for the V+jets background (left): the main (solid) and alternative (dashed) functions are displayed. On the right the simulated $\mathrm{t\bar{t}}$, t background m_j spectrum is fitted in $\ell\tau_{\mathrm{h}}$ events that satisfy the τ_{21} HP requirements.

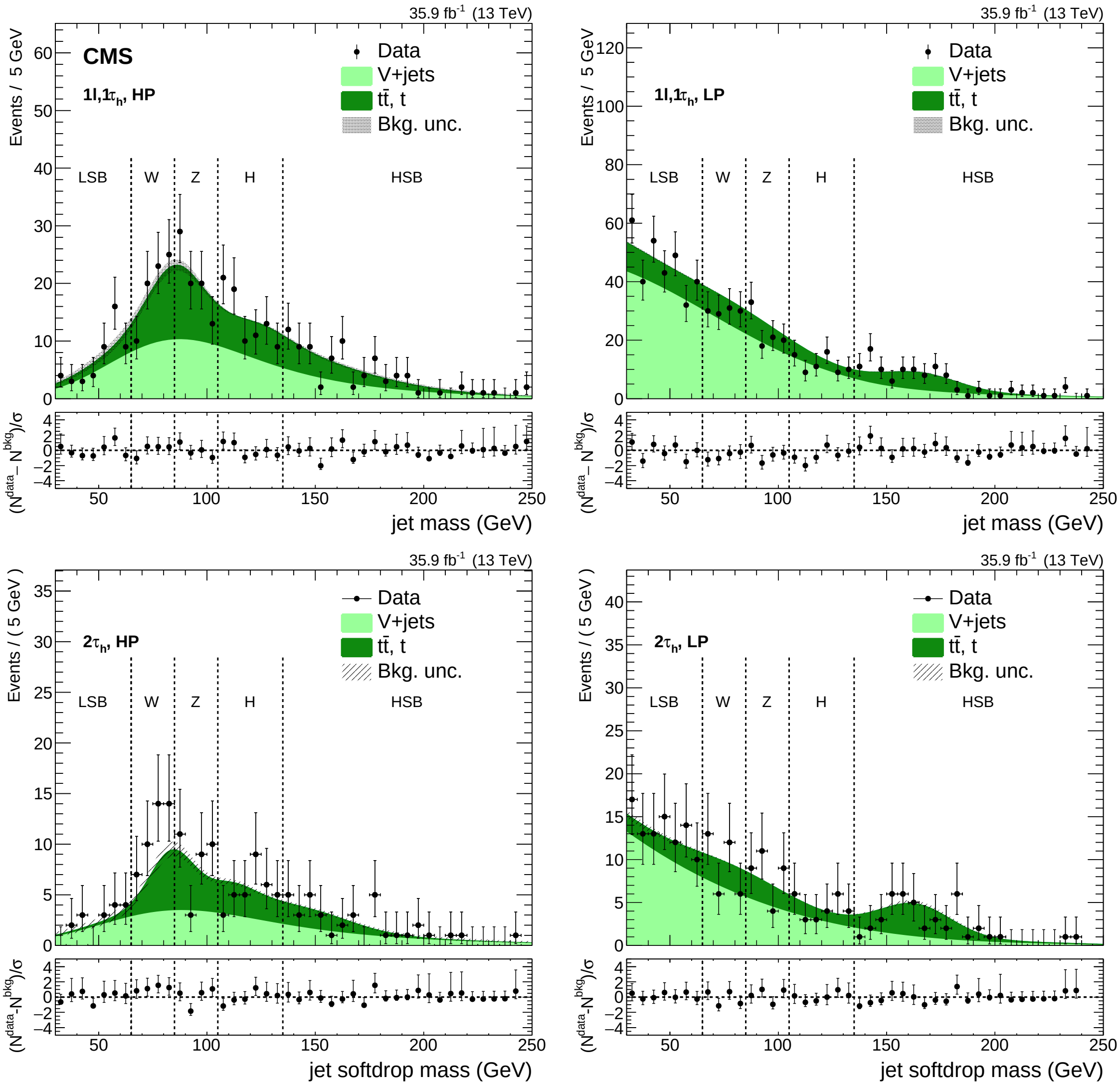


Figure 7.19: Soft-drop jet mass distribution in data in the HP (left) and LP (right) categories for the $\ell\tau_{\mathrm{h}}$ (upper row) and $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ (lower row) channels, together with the background prediction (fitted to the data in the SBs as explained in the text).

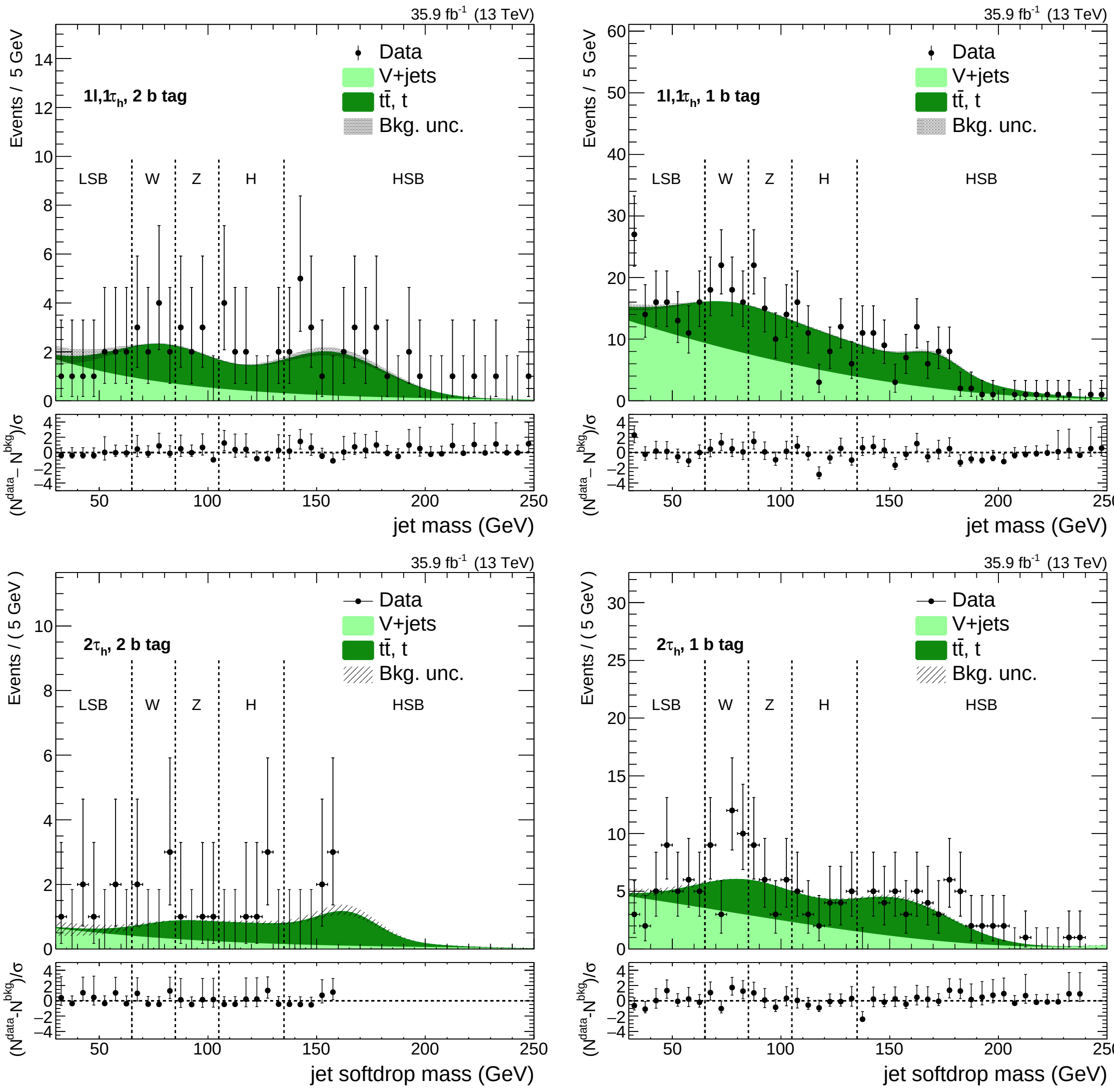


Figure 7.20: Soft-drop jet mass distribution in data in the 2 b-tagged subjets (left) and the 1 b-tagged subjet (right) categories for the $\ell\tau_\mathrm{h}$ (upper row) and $\tau_\mathrm{h}\tau_\mathrm{h}$ (lower row) channels, together with the background prediction (fitted to the data in the SBs as explained in the text).

The jet soft-drop mass distributions in data are reported in Figs. 7.19–7.20. The expected number of background events in each signal region is reported in Table 7.6. The quoted uncertainties are calculated as:

- the statistical uncertainty of the fit to the V+jets background performed on the SB in data and the systematic uncertainty due to propagation of the uncertainties of the fits on the $\mathrm{t\bar{t}}$, t backgrounds performed on simulations, to the fit performed on the data SB to extract the V+jets parameters

- the alternative function uncertainty is the difference in the background yields in the SR depending on the choice of the function used to describe the V+jets background.

Table 7.6: Predicted number of background events and the observed number in the signal region, for all event categories. The regions denoted by W, Z and H are intervals in the jet soft-drop mass distribution that range from 65 to 85 GeV, from 85 to 105 GeV, and from 105 to 135 GeV, respectively. Separate sources of uncertainty in the expected number are reported as the statistical uncertainty in the V+jets contribution from the fitting procedure (fit), the difference between the nominal and alternative function form chosen for the fit (alt), and the uncertainty in the background from top quarks from the fit to the simulated jet mass spectrum.

Category			V+jets (± fit)(± alt)	$t\bar{t}$, t	Total exp. events	Data
W region	HP	$\ell\tau_h$	$38 \pm 7 \pm 12$	37.8 ± 0.6	76 ± 14	78
		$\tau_h\tau_h$	$13.0 \pm 3.2 \pm 0.2$	16.0 ± 1.8	29.0 ± 3.7	45
	LP	$\ell\tau_h$	$105.3 \pm 6.8 \pm 9.0$	34.2 ± 0.9	140 ± 11	120
		$\tau_h\tau_h$	$27.0 \pm 3.3 \pm 3.0$	12.3 ± 0.6	39.3 ± 4.5	37
Z region	HP	$\ell\tau_h$	$39.9 \pm 6.1 \pm 7.9$	42.4 ± 1.0	82 ± 10	82
		$\tau_h\tau_h$	$13.7 \pm 3.0 \pm 2.5$	18.0 ± 1.8	31.6 ± 4.3	33
	LP	$\ell\tau_h$	$73.5 \pm 4.8 \pm 6.1$	29.1 ± 1.9	102.6 ± 8.0	92
		$\tau_h\tau_h$	$19.1 \pm 2.3 \pm 2.5$	10.4 ± 0.8	29.5 ± 3.5	33
H region	2 b tag	$\ell\tau_h$	$2.4 \pm 0.9 \pm 0.4$	6.9 ± 0.6	9.2 ± 1.2	10
		$\tau_h\tau_h$	$1.1 \pm 0.6 \pm 0.1$	3.8 ± 1.8	4.9 ± 1.9	5
	1 b tag	$\ell\tau_h$	$29.3 \pm 3.5 \pm 6.6$	37.3 ± 1.2	66.6 ± 7.5	56
		$\tau_h\tau_h$	$11.5 \pm 2.2 \pm 2.6$	15.4 ± 1.7	26.9 ± 3.8	23

7.4.2.2 Background shape

The second and final step of the background estimation consists of the prediction of the shape of the resonance mass spectrum in the signal region.

The distribution of the V+jets background resonance mass (m_X) in the SR is estimated from the data in the SBs through the use of a transfer function $\alpha(m_X)$, computed from simulations which accounts for the small kinematical differences and the correlations involved in the interpolation from the SBs to the SR. The systematic uncertainties that affect the simulated V+jets spectra cancel out in the ratio and do not affect the predicted background shape in the SR.

Each resonance mass spectrum in the SR and SB regions is parametrized separately for the V+jets background ($N_{SR}^{MC,V+jets}(m_X)$ and $N_{SB}^{MC,V+jets}(m_X)$) and the $t\bar{t}$,t background ($N_{SR}^{t\bar{t},t}(m_X)$ and $N_{SB}^{t\bar{t},t}(m_X)$), fitting the simulated resonance mass distributions. The α

function is determined from simulations as:

$$\alpha(m_{\mathrm{X}}) = \frac{\mathrm{N}_{\mathrm{SR}}^{\mathrm{MC,V+jets}}(m_{\mathrm{X}})}{\mathrm{N}_{\mathrm{SB}}^{\mathrm{MC,V+jets}}(m_{\mathrm{X}})}, \tag{7.1}$$

Depending on the category, the functional forms used to parametrize the m_X distributions are:

- **Exp**: a simple exponential function: $F_{\mathrm{Exp}}(x) = e^{ax}$
- **ExpN**: a product of two exponentials: $F_{\mathrm{ExpN}}(x) = e^{ax+b/x}$
- **Pow**: a power function: $F_{\mathrm{Pow}}(x) = 1/(x/\sqrt{s})^{a})$

The functions chosen to parametrize the main background and extract the α-function are reported in Table 7.7 for each category.

The distribution of the V+jets background in the SR is then estimated by fitting an analytic function to data in the SB in simulations and data, after subtracting the top quark background estimated from simulation, and multiplying by the $\alpha(m_{\mathrm{X}})$ transfer function. Figures 7.21 and 7.22 show the fits of the V+jets (left) and $\mathrm{t\bar{t}}$, t (right) background simulations in SR and SB, respectively, for the $\ell\tau_{\mathrm{h}}$ events in the W signal region HP category. The normalization of the V+jets is determined from the fit to the jet mass, as reported in Table 7.6.

The resonance mass distribution in the SB is shown in Fig. 7.23 (left) with the different V+jets and $\mathrm{t\bar{t}}$, t components, as well as the alpha function shown in Fig. 7.23 (center). The functions chosen for the modeling of the resonance mass spectra in each category are listed in Table 7.7.

The overall background in the SR is then expected to be:

$$\mathrm{N}_{\mathrm{SR}}^{\mathrm{data}}(m_{\mathrm{X}}) = \alpha(m_{\mathrm{X}})[\mathrm{N}_{\mathrm{SB}}^{\mathrm{data}} - \mathrm{N}_{\mathrm{SB}}^{\mathrm{t\bar{t},t}}](m_{\mathrm{X}}) + \mathrm{N}_{\mathrm{SR}}^{\mathrm{t\bar{t},t}}(m_{\mathrm{X}}), \tag{7.2}$$

where $\mathrm{N}_{\mathrm{SB}}^{\mathrm{t\bar{t},t}}$ and $\mathrm{N}_{\mathrm{SR}}^{\mathrm{t\bar{t},t}}$ are the distributions for the top quark process in the SB and SR, respectively. The shape and the normalization of the $\mathrm{t\bar{t}}$, t distribution are fixed from simulation, with the latter corrected using the appropriate scale factors in Table 7.4.

As a check, the background shape prediction in the signal region is performed again using the alternative functions listed in Table 7.7. The main and alternative functions predictions for the resonance mass spectrum are found compatible within their uncertainties that are dominated by the statistical uncertainty of the fits, as shown in Fig. 7.23 (right).

category			Main bkg function	Main bkg alternative	$t\bar{t}$, t
$\ell\tau$	τ_{21} HP	W region	ExpN	Exp	ExpN
		Z region	ExpN	Exp	ExpN
	τ_{21} HP	W region	ExpN	Exp	ExpN
		Z region	ExpN	Exp	ExpN
	2b tags	H region	Exp	Pow	Exp
	1b tag	H region	Exp	Pow	ExpN
$\tau\tau$	τ_{21} HP	W region	ExpN	Exp	Exp
		Z region	ExpN	Exp	Exp
	τ_{21} HP	W region	Exp	Pow	Exp
		Z region	ExpN	Exp	Exp
	2b tags	H region	Exp	Pow	Exp
	1b tag	H region	ExpN	Pow	Exp

Table 7.7: Main and alternative functions chosen to parametrize the V+jets background contribution in the m_X distribution for each channel.

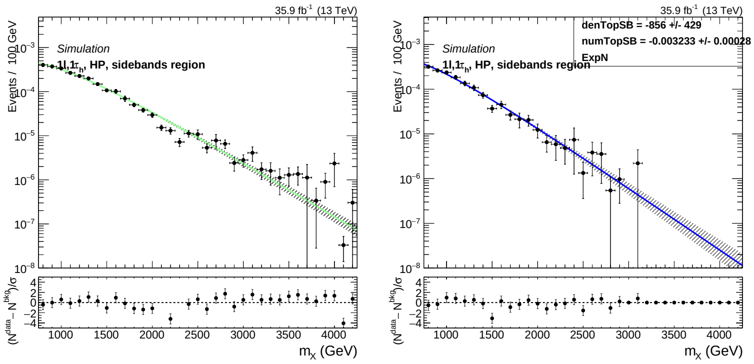


Figure 7.21: τ_{21} HP $\ell\tau_{\mathrm{h}}$ channel, W boson mass window SR. Fits to the simulated background components V+jets (left), $\mathrm{t\bar{t}}$, t (right) in the sideband (SB).

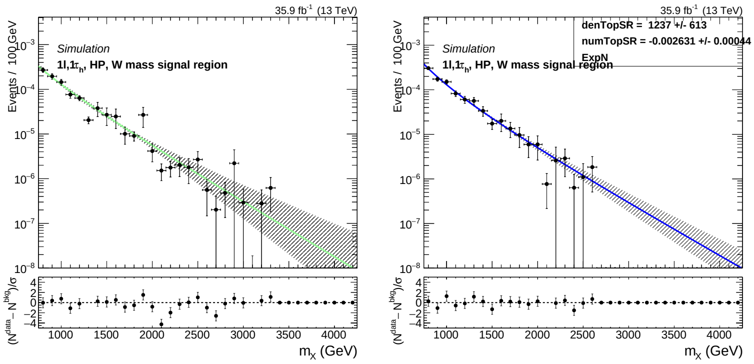


Figure 7.22: τ_{21} HP $\ell\tau_{\mathrm{h}}$ channel, W boson mass window SR. Fits to the simulated background components V+jets (left), $\mathrm{t\bar{t}}$, t (right) in the signal region (SR).

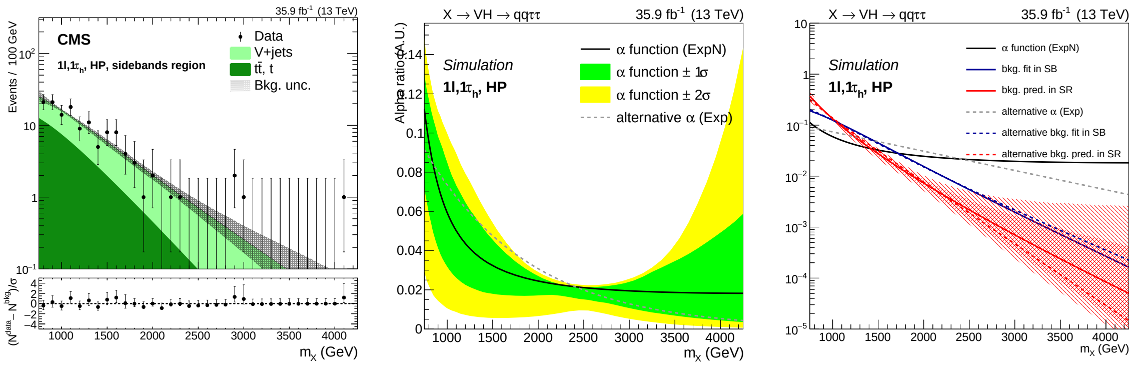


Figure 7.23: τ_{21} HP $\ell\tau_{\mathrm{h}}$ channel, W boson mass window SR. Fit to data in the SB (left), alpha function (center), and alpha function compared to the background shape in both SB and SR (right). The black line, with the corresponding 1σ (green) and 2σ (yellow) uncertainty bands, represents the α-function. The blue and red lines represent the estimated background in the SB and SR, respectively, with the main (solid line) and alternative (dotted line) parametrizations.

7.4.3 Validation of the background prediction method

7.4.3.1 Prediction of the background in the low mass side band

To validate the background estimation technique, the α-method is performed to predict the number of events and their resonance mass distribution for events where contributions from the signals considered in the analysis are very small. However, for the validity of the test, a region with a kinematic and flavor selection close to the analysis signal region is chosen. The low mass sideband is further split in two regions: a *test low mass sideband* with jet soft-drop mass between 30 and 50 GeV and the *test signal region* with a jet mass between 50 and 65 GeV. The background prediction is performed with the α-method using the test low mass and high mass sidebands, as shown in Fig.7.24 for one category. With these parameters, the prediction of the background in the SR region is estimated from the fit to the LSB and HSB regions and checked with data. In Table 7.8 the background expectations for the test low mass sideband for the different categories are presented together with the observed number of events. The alpha method is found to be able to predict both the number and the shape of the expected background events.

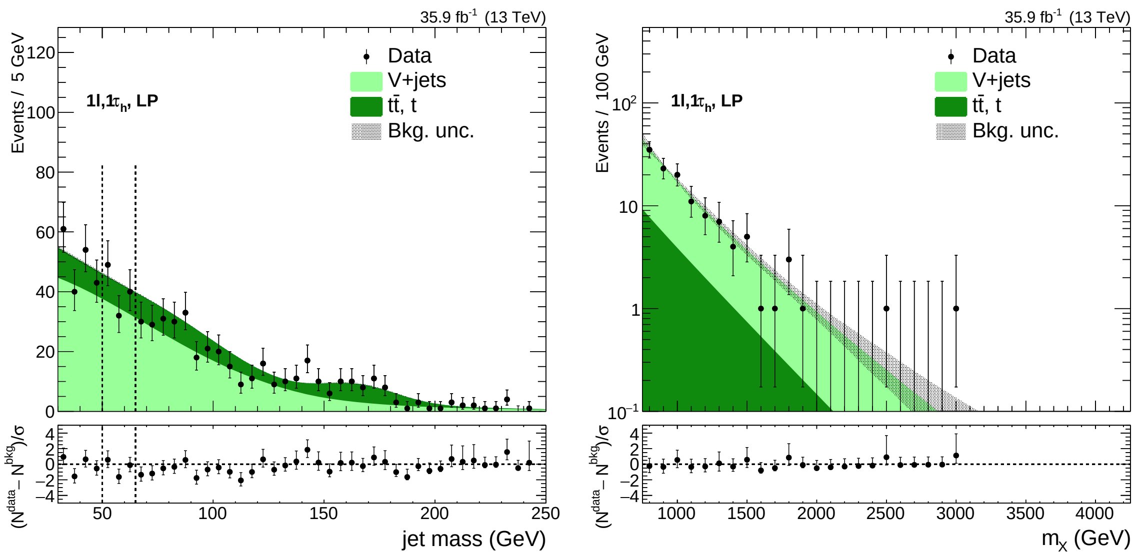


Figure 7.24: Fit to data of the jet soft-drop mass m_j in the SB region of the $\ell\tau_h$ LP τ_{21} category events, used to predict expected yields in the jet mass window from 50 to 65 GeV jet, which is treated as a signal region in the closure test (left). The shape of the backgrounds predicted with the α-ratio method for the closure test signal region is found to be in good agreement with data (right).

Table 7.8: Predicted number of background events and the observed number in the test signal region (50–65 GeV), for all event categories. Separate sources of uncertainty in the expected number are reported as the statistical uncertainty in the V+jets contribution from the fitting procedure (fit), the difference between the nominal and alternative function form chosen for the fit (alt), and the uncertainty in the background from top quarks from the fit to the simulated jet mass spectrum.

Category		V+jets (± fit)(± alt)	$t\bar{t}$, t	Total exp. events	Data
$\ell\tau_h$	HP	$13.9 \pm 4.9 \pm 2.8$	9.3 ± 0.6	23.2 ± 5.7	34
	LP	$103.6 \pm 8.1 \pm 7.9$	25.3 ± 1.4	129 ± 11	121
	2 b tags	$3.3 \pm 1.5 \pm 0.2$	3.0 ± 0.3	6.3 ± 1.5	6
	1 b tag	$33.7 \pm 4.4 \pm 3.3$	17.1 ± 0.9	50.8 ± 5.6	40
$\tau_h\tau_h$	HP	$6.0 \pm 2.5 \pm 0.8$	1.8 ± 0.2	7.9 ± 2.7	11
	LP	$26.2 \pm 3.9 \pm 2.4$	8.0 ± 0.5	34.2 ± 4.6	36
	2 b tags	$1.3 \pm 0.8 \pm 0.1$	0.6 ± 0.1	1.9 ± 0.8	2
	1 b tag	$11.5 \pm 2.6 \pm 1.3$	5.0 ± 0.5	16.5 ± 3.0	16

7.4.3.2 α-ratio shapes for different components in the V+jets background

A further check of the background estimation consists in making sure that the relative contributions and the resonance mass shapes of the various backgrounds, that may be different, do not impact the background prediction. This holds if the α-ratio shape of these backgrounds is similar. Binned α-ratio functions are made with simulated samples of Z+jets, W+jets, QCD, Dibosons and the inclusive V+jets. No significant differences are found in the shapes of the background contributions.

7.4.3.3 Impact of the jet $\rightarrow \tau_h$ fake rate on the background prediction

The two main components of the V+jets background differ in the fact that in the W+jets events that satisfy the analysis selection criteria, either the lepton or the τ_h are likely to be misidentified jets, while in the Z+jets events the leptons are genuine and recoiling against a jet. A check is performed to ensure that the overall V+jets normalization is not affected by a possible mismodeling of the rate of jets faking taus in the simulations. To test the effect of a possible the jet to τ_h misidentification probability mismodeling on the overall V+jets prediction, the expected number of events from the W+jets simulations is changed by a factor of 0.5 and 2. Then the α-ratio background normalization procedure is repeated for the different value of the prefit W+jets normalization in the HP category in the semileptonic and fully hadronic channels.

Since the probability of a jet faking the tau affects just the leptonic part of the event, it is independent of the selection applied on the large-cone jet, so the check is performed for the V-tagged HP category in the W mass window, in the semileptonic and fully hadronic channels. Results are shown in Tab. 7.9 and Figs.7.25–7.27. The prefit normalization changes only slightly the fit functions, and after the fit to data, the results for the V+jets normalization and the resonance mass spectra agree within the uncertainties, for all the scenarios of the different W+jets prefit normalization: nominal, half of the nominal and double the nominal value. Similar results are found also for the $\tau_{\rm h}$ $\tau_{\rm h}$ channel.

W HP category	V+jets (± fit)(± alt)	Top yields	Total yields	Obs. events
$\ell\tau_{\rm h}$ (W+jets nom)	$37.9 \pm 6.5 \pm 12.2$	37.8 ± 0.6	75.7 ± 13.8	
$\ell\tau_{\rm h}$ (W+jets nom/2)	$35.5 \pm 6.1 \pm 9.6$	37.8 ± 0.6	73.2 ± 11.4	78
$\ell\tau_{\rm h}$ (W+jets nom*2)	$37.3 \pm 6.3 \pm 11.9$	37.8 ± 0.6	75.1 ± 13.5	
$\tau_{\rm h}\tau_{\rm h}$ (W+jets nom)	$13.0 \pm 3.2 \pm 0.2$	16.0 ± 1.8	29.0 ± 3.7	
$\tau_{\rm h}\tau_{\rm h}$ (W+jets nom/2)	$13.3 \pm 3.3 \pm 0.5$	16.0 ± 1.8	29.2 ± 3.8	45
$\tau_{\rm h}\tau_{\rm h}$ (W+jets nom*2)	$12.1 \pm 3.0 \pm 0.1$	16.0 ± 1.8	28.0 ± 3.5	

Table 7.9: Expected background yields and data events in W HP region for different W+jets prefit normalizations in the $\ell\tau_h$ and $\tau_h\tau_h$ channels.

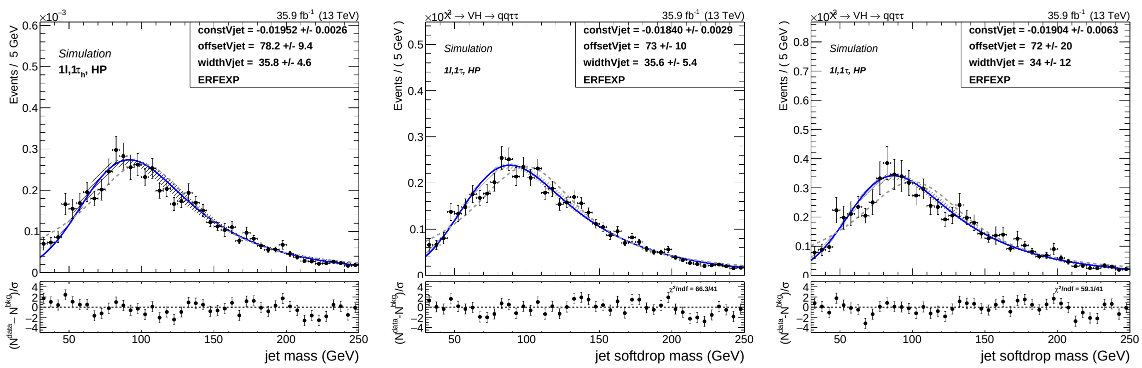


Figure 7.25: Fit to the simulated jet soft-drop mass m_j in $\ell\tau_{\rm h}$ events that satisfy the τ_{21} HP requirements, for the V+jets background, with the W+jets component normalized to its nominal value(left), half of its nominal value(center) and double of its nominal value(right). The main background functions are the solid lines while the alternative ones are dashed.

7.5 Signal characterization

The search is performed by looking for a localized excess in resonance mass spectrum compatible with the signals under study. Functional forms are used to model the resonance mass spectra of the simulated signals. The simulated signal samples, for

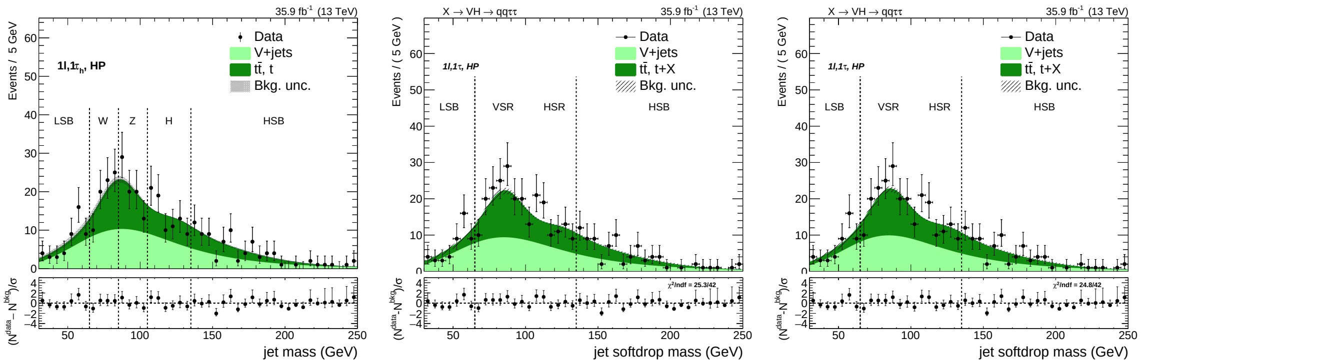


Figure 7.26: Fit to data m_j in $\ell\tau_h$ HP τ_{21} category events with the W+jet component normalized to its nominal value (left), half of its nominal value (center) and double of its nominal value (right).

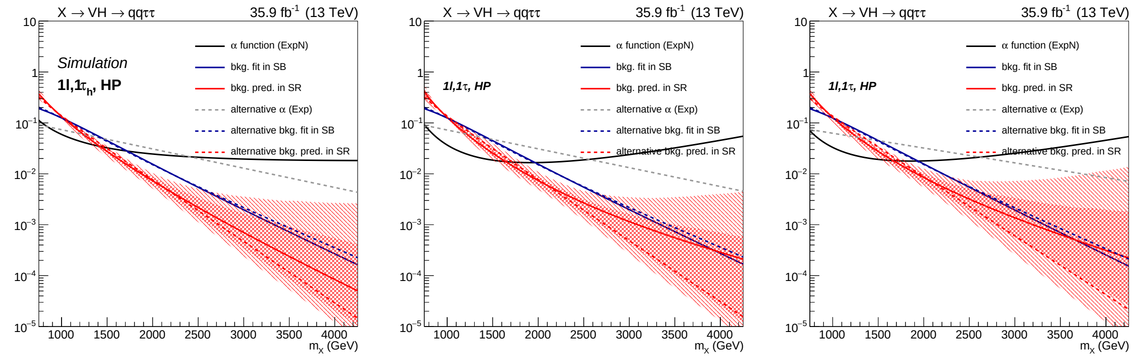


Figure 7.27: τ_{21} HP $\ell\tau_h$ channel, W boson mass window SR. The blue and red lines represent the α-ratio estimated background in the SB and SR, respectively, with the main (solid line) and alternative (dotted line) parametrization, for different initial W+jets normalization: the nominal normalization (left), half of the nominal normalization (center), and double the nominal value(right).

different mass hypotheses are fitted in the SR with empirical functions in order to get the shape used in the unbinned likelihood fit for the final signal extraction. In order to model the signal shape, Crystal Ball functions are utilized: these functions have four parameters and consist of a Gaussian core convolved with a power-law tail. For each sample a fit is performed to the distribution of the resonance mass, as represented in Fig.7.28 (left) for a W′ signal of mass 2.5 TeV in the τ_{21} HP category in the W mass region.

The signal is parametrized by interpolating the fitted parameters for each signal category in order to have a continuity of the signal shape for every possible mass values in the range. A linear fit is performed to parametrize the Crystall Ball parameters.

The signal normalizations, i.e. the product of the sample acceptance and efficiency and the branching ratios, are taken as the integral of the resonance mass distributions and reported in Fig. 7.29 for the different signals, channels, and categories.

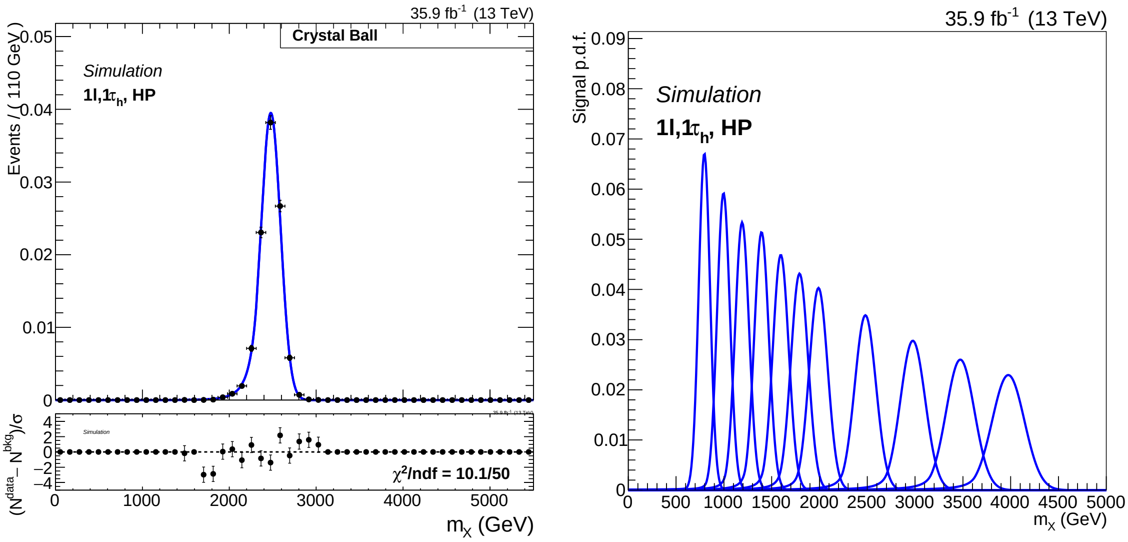


Figure 7.28: Crystal Ball functional form fit to the the resonance mass spectrum of a W′ signal of mass 2.5 TeV in the τ_{21} HP category in the W mass region (left). Fitted distributions of the W′ signal resonance mass distribution in the W mass region HP category.

Since signal samples were simulated for a certain set of masses, in order to calculate the expected normalization for other mass points, functional forms are used to parametrize the normalization as a function of the resonance mass. Polynomial functions are used to fit the normalization of the available samples as a function of the resonance mass and then used to interpolate for the non-simulated mass points.

Effects due to the main systematic uncertainties described in Sec. 7.6 are considered on the normalization of the expected signal events and the shape of the signal resonance mass distribution for both the mean and the sigma of the Gaussian core.

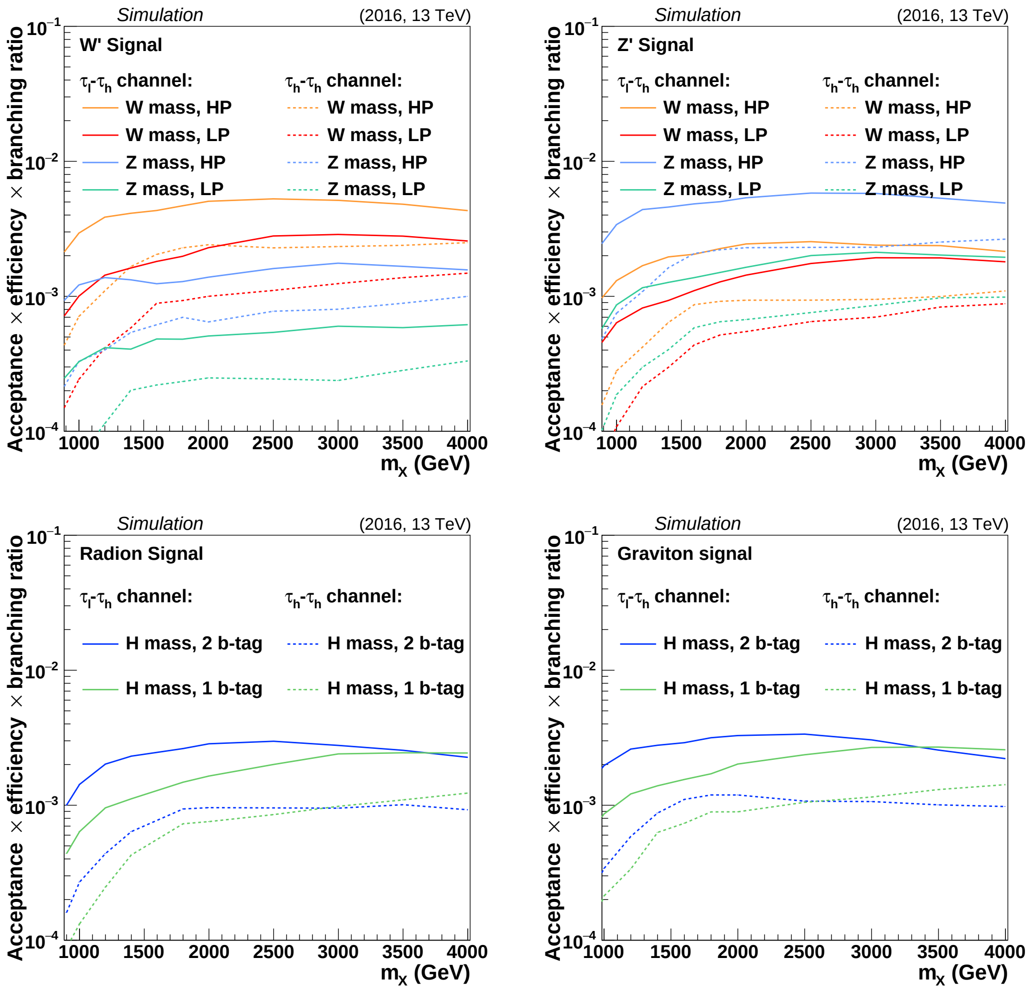


Figure 7.29: Product of the acceptance, the selection efficiency, and the branching fractions for the signals considered in the analysis, in the different categories.

7.6 Systematic uncertainties

In this section the systematic uncertainties, resulting from experimental and theory sources, that may affect both the normalization and shape of the signal and background distributions, are evaluated and reported. The impact of the systematic contribution varies depending on the channel and the resonance mass, and the main ones are related to the background prediction, due to the low statistics in the SB both in data and in the simulated samples. The most important systematic contributions on the signal predictions are due to V tagging, H tagging and the τ_h identification.

The principal background is V+jets, and its modeling represents the largest uncertainty in the analysis. The systematic uncertainty in the V+jets background is dominated by the statistical uncertainty associated with the number of events in the jet mass distribution SBs in data and simulation. An additional uncertainty is related to the choice of model used for the jet mass in the V+jets background. The latter is evaluated from the differences in the expected background yields obtained when using the alternative fitting functions, e.g as shown in Fig. 7.18. A systematic uncertainty due to the parameterization and fit on the $\mathrm{t\bar{t}}$, t background in the SBs is also propagated to the uncertainty of the V+jets background. The uncertainties in the shape of the V+jets distribution are estimated from the covariance matrix of the fit to m_X in the sidebands and the uncertainties in the $\alpha(m_\mathrm{X})$ ratio, which depend on the number of events in data and simulation, respectively. For the top-quark processes, uncertainties from normalization and shape in the parameterization are propagated to the final background estimation. The single top quark and top quark pair production normalization uncertainty arises predominantly from the limited number of events in the CRs.

The uncertainties in the trigger efficiency, and in the electron and muon reconstruction, identification, and isolation efficiencies are obtained by varying the corresponding scale factors by their uncertainty, and each is found to be 1–2% [99, 133]. For the τ_h reconstruction and identification, the uncertainties vary between 6 and 8% and between 10 and 13%, depending on the resonance mass, in the $\ell\tau_\mathrm{h}$ and the $\tau_\mathrm{h}\tau_\mathrm{h}$ channels, respectively [154]. A separate uncertainty due to the extrapolation of the reconstruction and identification ($^{+5\%}_{-35\%} * (p_\mathrm{T}/1000\,\mathrm{GeV})$) of τ_h leptons at large p_T has an impact on the signal normalization of 18% in the $\ell\tau_\mathrm{h}$, and 30% in the $\tau_\mathrm{h}\tau_\mathrm{h}$ channels, for a 4 TeV signal hypothesis. This uncertainty is responsible for an increase of 1% in the width of the signal distribution.

For the τ energy scale contribution to the systematics uncertainty, shape and normalization have been evaluated after varying the tau energy scale y 3%. In the $\ell\tau_\mathrm{h}$ channel the change in the normalization and the signal width is about 1%. In the $\tau_h\tau_h$ channel the normalization varies from about 5% at 1 TeV to 3% at 4 TeV, while the signal width varies by about 3%.

Jet energy scale and resolution uncertainties affect both the selection efficiencies and the shape of distributions. The jet energy scale uncertainty is evaluated simultaneously on jets and p_T^{miss} and accounts for a variation in signal efficiency of 1–3%. The jet energy resolution effect is evaluated by smearing the jet p_T using the η-dependent coefficient uncertainties with the hybrid method 5.3.5.1 and has an impact of 1–2%. The effect on the resonance mass distribution is at the level of 1–2% for the mean and the width of the signal distribution. The corrections to the jet mass scale and resolution are also taken into account, and result in a variation of 1–8% in the expected number of signal events. Event migrations between the mass windows due to the effect of jet mass scale and resolution variations are estimated to be between 2 and 15%, depending on the signal and the vector boson mass region.

Scale factors for V tagging and b tagging represent one of the largest source of normalization uncertainty for the signal. Uncertainties in normalization correspond to 6 and 11% in the HP and LP categories, respectively. An additional uncertainty from the extrapolation of the W tagging from the $t\bar{t}$ scale to larger values of jet p_T is estimated using an alternative HERWIG [167] shower model. It is parametrized as a function of the jet p_T to be $A \cdot \log(p_T/200\,\mathrm{GeV})$, where A =8.5% for the high-purity category and A = 3.9% for low-purity category. This amounts to an uncertainty that varies from 2 to 18% for the 0.9–4 TeV mass hypotheses and the two V tag categories. In addition, the contribution to the signal normalization uncertainty from the b tagging uncertainty varies between 3 (4)% to 7 (5)% for the 2 (1) b-tagged subjet categories, estimated by varying the data-to-simulations corrective factor by their uncertainties. Effects due to the AK4 jets b-tagging efficiency uncertainty amount to 3% on the normalization of the $t\bar{t}$,t background and 1% on the signal yields.

Normalization uncertainties from the choice of the parton distribution function (PDF) grow larger with higher resonance mass, and are in general larger for gluon-initiated processes than for quark-initiated processes. For W′ and Z′ production, which are sensitive to quark PDFs, effects range from 6 to 37%, while radion and graviton production depend on gluon PDFs, and result in a variation of 10 to 64% in the number of expected signal events. Uncertainties of similar magnitude arise from factor-of-two independent variations in the factorization and renormalization scales, resulting in 3 to 13% variations for W′ and Z′, and 10 to 19% for radion and graviton production. While these normalization uncertainties are not considered in setting limits on production, effects on signal acceptance are propagated to the final fit, amounting to 0.5–2% for PDF uncertainties, depending on the resonance mass.

Other systematic uncertainties affecting the normalization of signal and minor backgrounds considered in the analysis include pileup contributions (0.5%), estimated by varying the minimum bias cross section by 5%, and integrated luminosity (2.5%) [168]. A list of the main systematic uncertainties is given in Table 7.10.

Table 7.10: Summary of systematic uncertainties for the background and signal events. Uncertainties marked with "shape" are propagated also to the shape of the distributions, and those marked with † are not included in the limit bands, but instead reported in the theory band. The dash symbol is reported where the uncertainty is not applicable to a certain signal or background. The symbols qq' and gg refer to quark-initiated and gluon-initiated processes, respectively.

	V+jets	$\mathrm{t\bar{t}}$, t	Signal
α-function	shape	—	—
Bkg. normalization	11–60%	2–38%	—
Top quark scale factors	—	5–14%	—
Jet energy scale	—	—	shape
Jet energy resolution	—	—	shape
Jet mass scale	—	—	1%
Jet mass resolution	—	—	8%
V tagging	—	—	6% (HP)–11% (LP)
V tagging extrapol.	—	—	8–18% (HP), 2–8% (LP)
b tagging	—	—	3–7% (1b), 4–5% (2b)
b-tagged jet veto	—	3%	1%
Trigger	—	—	2%
Lepton identification, isolation	—	—	2%
τ lepton identification	—	—	6–8% ($\ell\tau_\mathrm{h}$), 10–13% ($\tau_\mathrm{h}\tau_\mathrm{h}$)
τ lepton identification p_T extrapol.	—	—	0.5–18% ($\ell\tau_\mathrm{h}$), 0.2–30% ($\tau_\mathrm{h}\tau_\mathrm{h}$), shape
τ lepton energy scale	—	—	1% ($\ell\tau_\mathrm{h}$), 3–5% ($\tau_\mathrm{h}\tau_\mathrm{h}$), shape
Pileup	—	—	0.5%
Renorm./fact. scales†	—	—	2.5–12.5%(qq'), 10–19%(gg)
PDF yield†	—	—	6–37%(qq'), 10–64%(gg)
PDF acceptance	—	—	0.5–2%
Integrated luminosity	—	—	2.5%

In the final fits of the resonance mass spectrum in data, the systematic uncertainties are treated as nuisance parameters and described by a probability density function as described in App. A. In the scope of this search log-normal priors are adopted since they are usually associated with positively defined parameters. Uncertainties that are partially correlated, like the ones associate with the α-method, are decorrelated through linear transformations along the eigenvalues of the covariance matrix.

7.6.1 Fit diagnostics: nuisance parameters

As a consistency check of the systematic uncertainty treatment, the nuisance parameters ($\hat{\theta}$) are profiled, i.e. post-fit, are compared to the value before the fit (θ_0), normalized with respect to the pre-fit uncertainty ($\Delta\theta$). The nuisance *pulls* are defined as ($\hat{\theta} - \theta_0/\Delta\theta$) and are computed with a fit to data in both the background-only (black) and signal+background (red) hypotheses, as shown in Fig. 7.30 for a W′ signal of 2 TeV of mass in the $\ell\tau_\mathrm{h}$ and $\tau_\mathrm{h}\tau_\mathrm{h}$, HP and LP , W and Z soft-drop mass regions combined. The distributions of the pulls doesn't show any unexpected behavior, since they are distributed around 0, within the pre-fit uncertainties (green and yellow bands), so compatible with the pre-fit values. The post-fit uncertainties are also close in magnitude to the pre-fit values. In some cases the fitted values are farther from the pre-fit, but still within the 2σ band and this happens because the final fit on data is able to constrain further some parameters.

The nuisance parameters are divided in main groups: the ones that regulate the shape and normalization of the V+jets background (indicated with the strings "_eig" and "_norm"), the ones for the $\mathrm{t\bar{t}}$, t background (with the strings "Top_*_fit" and its normalization due to the scale factor computed in the $\mathrm{t\bar{t}}$ enriched control region "sf_Top"), and other uncertainties affecting the normalization of the signal. The latter are related to the identification efficiencies ("eff"), the jet energy scale and resolution ("jes" and "jer"), as well as the migration due to the jet mass scale and resolution ("scale_j_m_migr" and "res_j_m_migr").

In a similar way, also *impacts* can be defined as the shift that one nuisance parameter induces in the signal strength (r), which represents the product of the cross section and branching ratios for the particular final state, when it is changed by the post-fit $+1\sigma$ and -1σ uncertainty around the post-fit value, while the other nuisance parameters are set to their post-fit values. Figure 7.31 represents the impacts of the most relevant nuisance parameters for a W′ signal of 2 TeV of mass in the $\ell\tau_\mathrm{h}$ and $\tau_\mathrm{h}\tau_\mathrm{h}$, HP and LP, W and Z soft-drop mass regions combined. The most relevant uncertainty for the signal strength are related to the background estimation and the jet energy calibration and the V-tagging and tau-identification efficiency uncertainties. No unexpected behavior is observed.

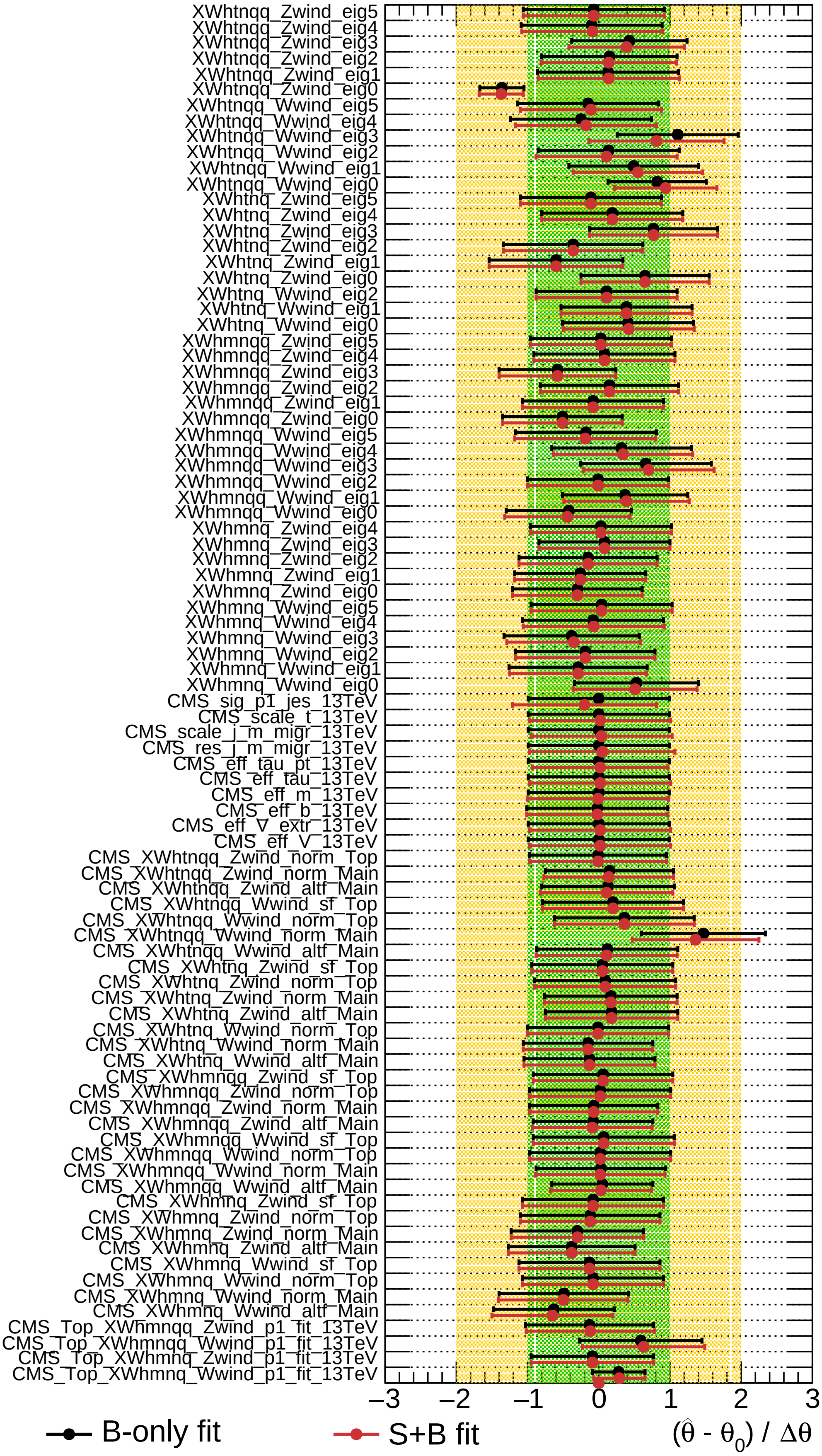


Figure 7.30: Distribution of the pulls for a 2 TeV W′ resonance search in the $\ell\tau_\mathrm{h}$ and $\tau_\mathrm{h}\tau_\mathrm{h}$, W and Z mass regions, with τ_{21} HP and LP categories combined.

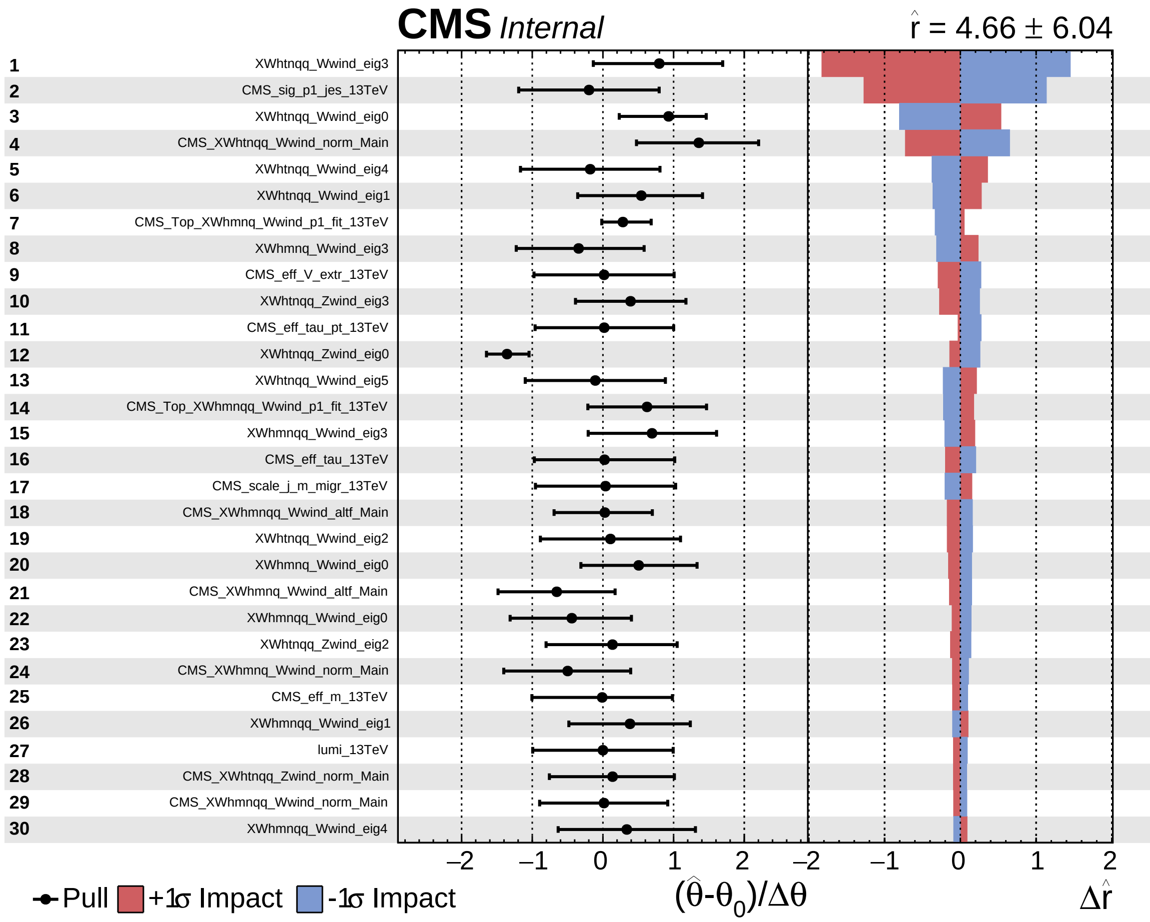


Figure 7.31: Distribution of the impacts for a 2 TeV W′ resonance search in the $\ell\tau_\mathrm{h}$ and $\tau_\mathrm{h}\tau_\mathrm{h}$, W and Z mass regions, with τ_{21} HP and LP categories combined.

7.7 Results

Results are obtained from a combined fit of the signal and background to the unbinned distribution of the resonance mass in data, based on a profile likelihood, where the systematic uncertainties are considered as nuisance parameters [157, 169]. The background-only hypothesis is tested against the signal + background hypothesis in the different categories, simultaneously. No evidence of significant deviations from the background expectation are found. The data in the SR and the background predictions before and after the final fit in the SR are shown in Figs. 7.32 and 7.33.

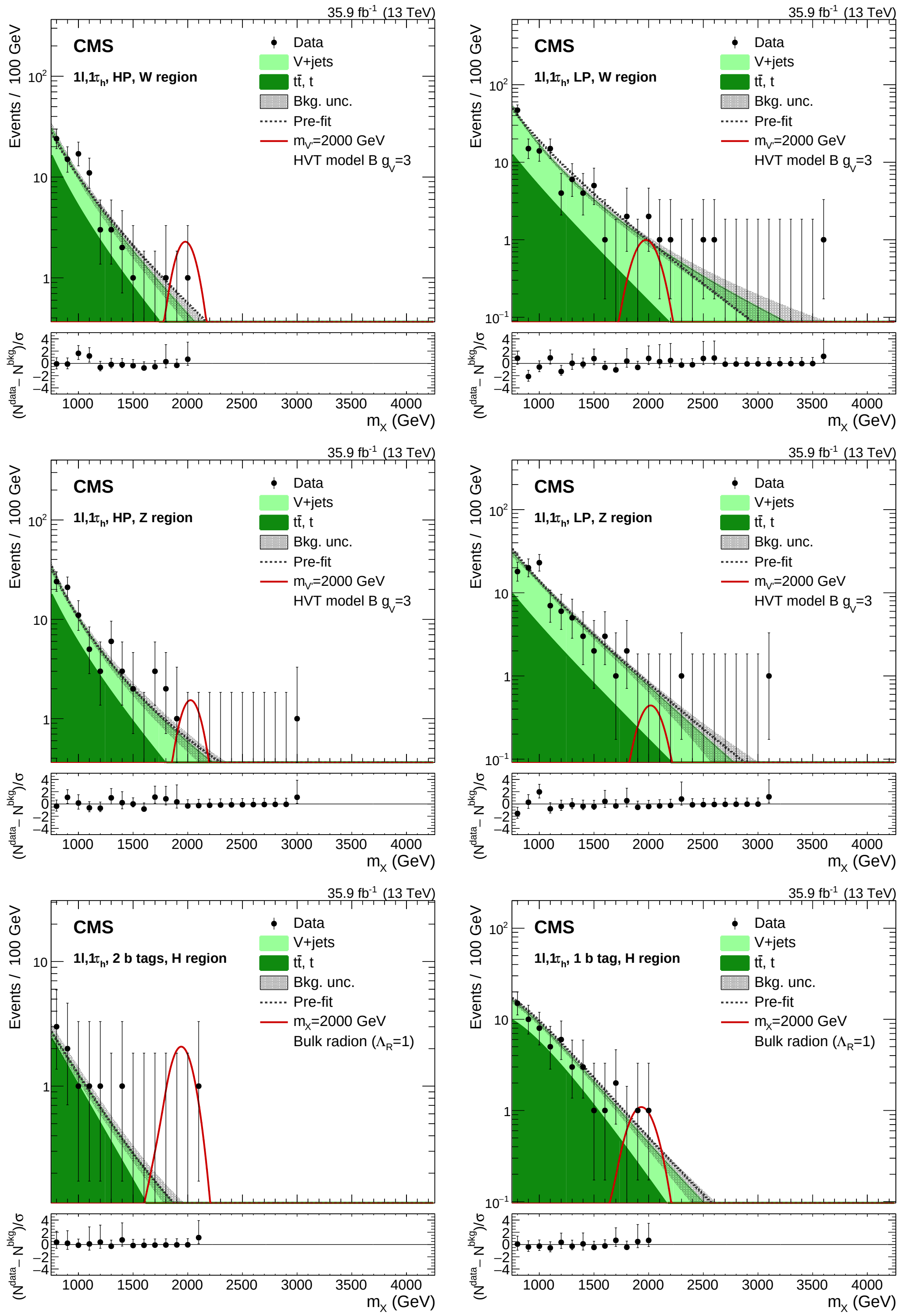


Figure 7.32: Data and expected backgrounds in the $\ell\tau_{\mathrm{h}}$ channel. The W mass window is shown in the HP (upper left) and LP (upper right) categories, the Z mass window for the HP (middle left) and LP (middle right) categories, and the H mass window for the two b-tagged subjet (lower left) and one b-tagged subjet (lower right) categories. The lower panels depict the pulls in each bin, $(\mathrm{N_{data}} - \mathrm{N_{bkg}})/\sigma$, where σ is the statistical uncertainty in data, as given by the Garwood interval [170], and provide estimates of the goodness of fit. Signal contributions are shown, assuming benchmark HVT model B for the V' and $\Lambda_{\mathrm{R}} = 1$ for the radion.

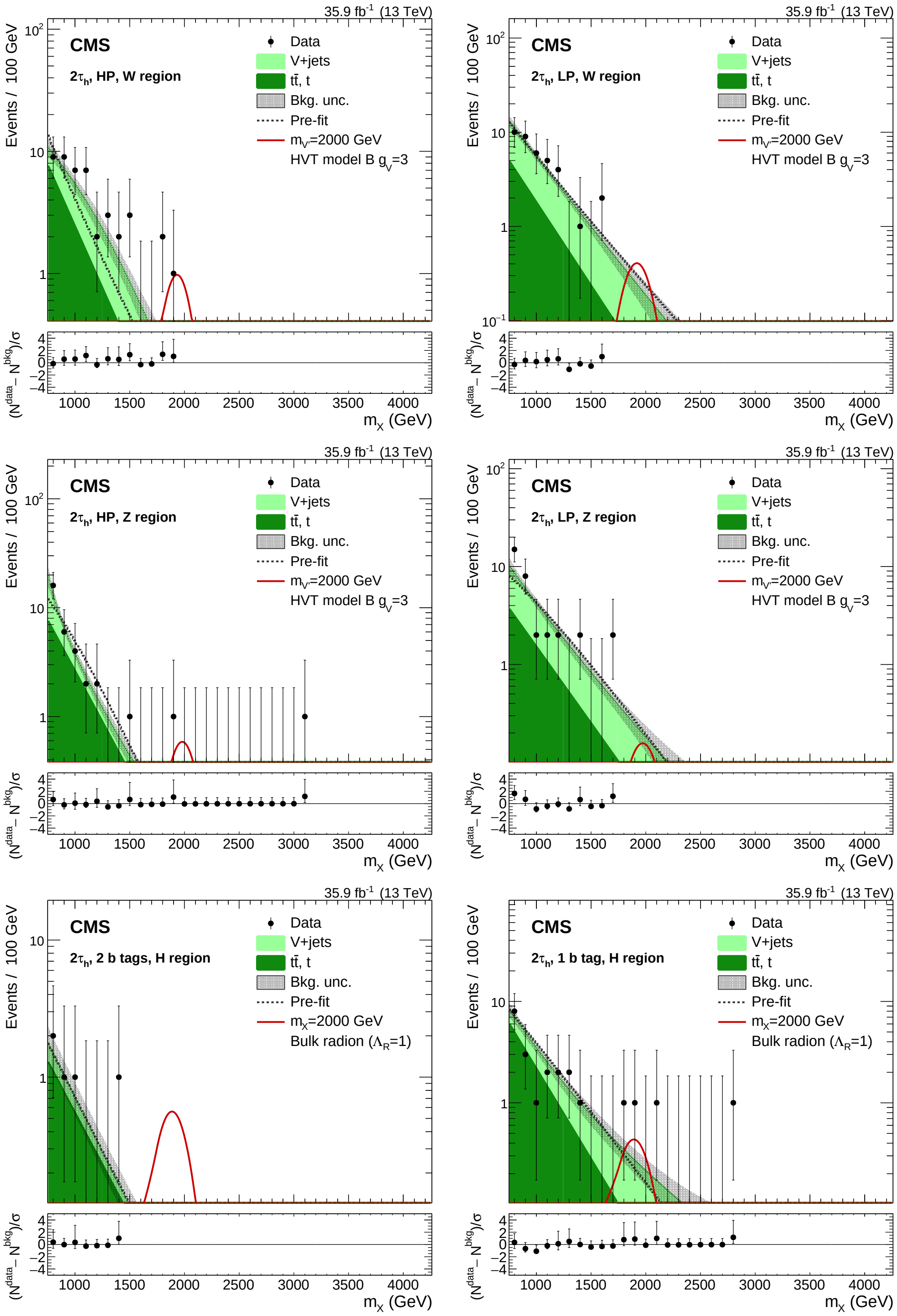


Figure 7.33: Data and expected backgrounds in the $\tau_h\tau_h$ channel. The W mass window is shown in the HP (upper left) and LP (upper right) categories, the Z mass window for the HP (middle left) and LP (middle right) categories, and the H mass window for the two b-tagged subjet (lower left) and one b-tagged subjet (lower right) categories. The lower panels depict the pulls in each bin, $(N_{data} - N_{bkg})/\sigma$, where σ is the statistical uncertainty in data, as given by the Garwood interval [170], and provide estimates of the goodness of fit. Signal contributions are shown, assuming benchmark HVT model B for the V′ and $\Lambda_R = 1$ for the radion.

7.7.1 Expected limits

Since no significant discrepancy between the data and the background expectation is found, the $\mathrm{CL_s}$ criterion is used to determine the 95% confidence level (CL) limit on the signal contribution in the data, with the asymptotic approximation method [157, 158, 171]. In the limit setting, the signals are assumed to have narrow widths, i.e. widths that are negligible compared to the resonance-mass resolution of approximately 7%. The limits are obtained on the product of the cross section and branching fraction for a heavy resonance (X) that decays to HH, WH, or ZH as a function of the resonance mass. For all the signals the different purity categories are considered simultaneously. For the WH and ZH final states, the W and Z boson mass regions are combined because there are contributions from both signals to the two mass regions. The $\ell\tau_\mathrm{h}$ and $\tau_\mathrm{h}\tau_\mathrm{h}$ combined limits, together with the $\pm1\sigma$ and $\pm2\sigma$ bands, are shown as reported in Figs. 7.34–7.35. Resonance spins of 0 and 2 are considered for the HH final state, while the resonance spin is assumed to be 1 for the WH and ZH final states. The exclusion limit ranges from 80 to 5 fb for resonances of spin 0 and 2, and from 180 to 5 fb for spin-1 resonances.

The predictions from the bulk radion and graviton models are superimposed on the exclusion limits assuming $\Lambda_\mathrm{R} = 1\,\mathrm{TeV}$ and $\tilde{k} = 0.5$. With this assumption for the theory parameters, a radion resonance with mass below 2.7 TeV is excluded at 95% CL. For a spin-1 signal, the results are interpreted in the context of the simplified HVT benchmark models A and B, and both the predictions are shown on the limit plots. As shown in Fig. 7.35, a W′ (Z′) resonance of mass lower than 2.6 (1.8) TeV is excluded at 95% CL in the HVT benchmark model B. The HVT benchmark model A is also reported for completeness. The expected and observed limits on the V′ resonance are shown in Fig. 7.36 (left), for the mass-degenerate spin-1 triplet hypothesis, again with the benchmark model A and B predictions.

7.7.2 HVT interpration

For the spin-1 signals, the results are also interpreted in the context of the simplified HVT model with heavy vector bosons ($\mathrm{V}^\pm$, V^0), which are mass degenerate. The model is parametrized in terms of a new interaction of strength g_V, the coupling to the H boson or the longitudinally-polarized SM vector boson c_H, and the coupling to fermions c_F.

The exclusion limit shown in Fig. 7.36 (left) can be interpreted as a limit in the space of

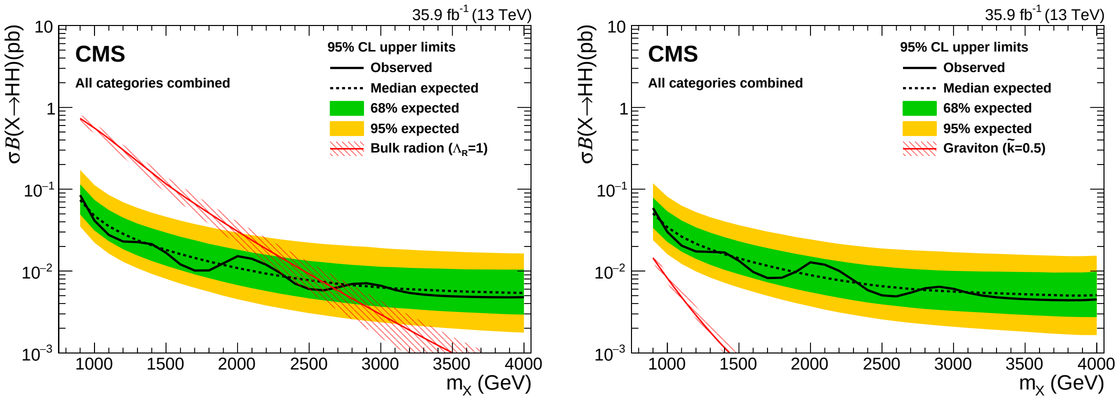


Figure 7.34: Observed 95% CL upper limits on $\sigma\mathcal{B}(\mathrm{X(spin\text{-}0)} \to \mathrm{HH})$ (left) and $\sigma\mathcal{B}(\mathrm{X(spin\text{-}2)} \to \mathrm{HH})$ (right). Expected limits are shown with ±1 and ±2 standard deviation uncertainty bands. The $\ell\tau_\mathrm{h}$ and $\tau_\mathrm{h}\tau_\mathrm{h}$ final states, and the one and two b-tagged sub-jet categories, are combined to obtain the limits. The solid red lines and the red dashed areas correspond to the cross sections predicted by the bulk radion and graviton and their corresponding uncertainties, as reported in Table 7.10.

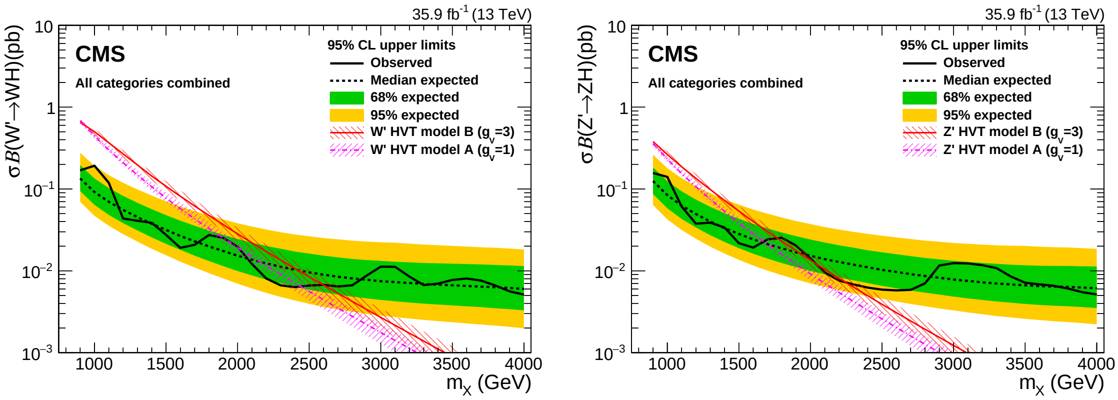


Figure 7.35: Observed 95% CL upper limits on $\sigma\mathcal{B}(\mathrm{W'} \to \mathrm{WH})$ (left) and $\sigma\mathcal{B}(\mathrm{Z'} \to \mathrm{ZH})$ (right). Expected limits are shown with ±1 and ±2 standard deviation uncertainty bands. The $\ell\tau_\mathrm{h}$ and $\tau_\mathrm{h}\tau_\mathrm{h}$ final states, for the HP and LP τ_{21} categories, and the W and Z boson mass signal regions, are combined to obtain the limits. The solid lines and the relative dashed areas in magenta and red correspond to the cross sections predicted by the HVT models A and B, respectively, and their corresponding uncertainties, as reported in Table 7.10.

the HVT model parameters $[g_V c_H, g^2 c_F/g_V]$. The excluded region in such a parameter space for narrow resonances is shown in Fig. 7.36 (right). The region of parameter space where the natural resonance width is larger than the typical experimental resolution of 7%, for which the narrow width assumption is not valid, is shaded.

A comparison between the 2016 results of diboson resonant production searches in different final states is done in Fig.7.37 of hypotheses of W′ and Z′, for the HVT model in the benchmark model B, and of a bulk graviton (bottom plot), with $\tilde{k} = 0.5$

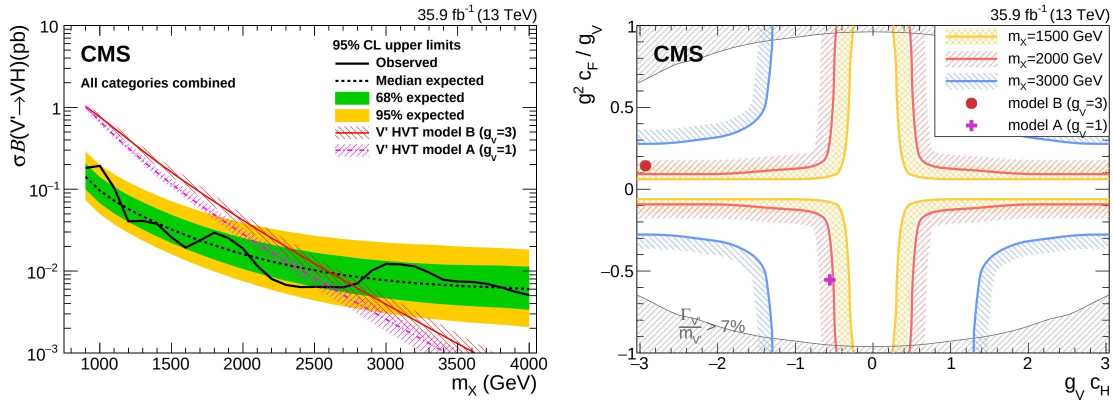


Figure 7.36: Expected (with $\pm1(2)\sigma$ bands) and observed 95% CL upper limit on $\sigma\times$BR(X$\to$VH) (left) in the $\ell\tau_h$ and $\tau_h\tau_h$, τ_{21} HP and LP categories, W and Z mass signal window regions combined. Observed exclusion limit (right) of the space of the HVT model parameters $[g_V c_H, g^2 c_F/g_V]$ for three different mass hypothesis (1.5, 2, and 3 TeV). The parameter g_V represents the coupling strength of the new interaction; c_H is the coupling between the HVT resonance and the Higgs boson or longitudinally-polarized SM vector bosons; and c_F is the coupling between the HVT resonance and the SM fermions. The region of parameter space where the natural resonance width is larger than the typical experimental resolution of 7%, for which the narrow width assumption is not valid, is shaded in grey.

resonances that decay to diboson final states.

The searches with leptons in the final states have higher sensitivity in the low-mass resonance region, because of the higher efficiency of rejecting background events, while the hadronic final states have high sensitivity in the higher mass tail, already depleted of background events, where the higher branching ratio maximizes the signal expectation. Some channels show localized excesses of data with respect to the standard model prediction, although none are significant. As a final remark, it can be noted that different searches have results and sensitivity that are comparable with each other, thus justifying the current effort ongoing in the CMS Collaboration of performing a statistical combination of the results. The ATLAS collaboration already combined searches for resonant production of diboson, dileptons, and lepton plus missing momentum with 2015 and 2016 data [172], providing the best limits to date: a heavy vector-boson triplet is excluded with mass below 5.5 TeV in a weakly coupled scenario (HVT model A) and 4.5 TeV in a strongly coupled scenario (Model B), as well as a Kaluza-Klein bulk graviton with mass below 2.3 TeV, for $\tilde{k} = 1$. For both the collaborations, these results will be further improved by the inclusion of the data acquired in 2017 and 2018, in a grand combination of the data of Run 2 of the LHC, which will allow this unique phase space for heavy diboson resonances to be probed with unprecedented capabilities.

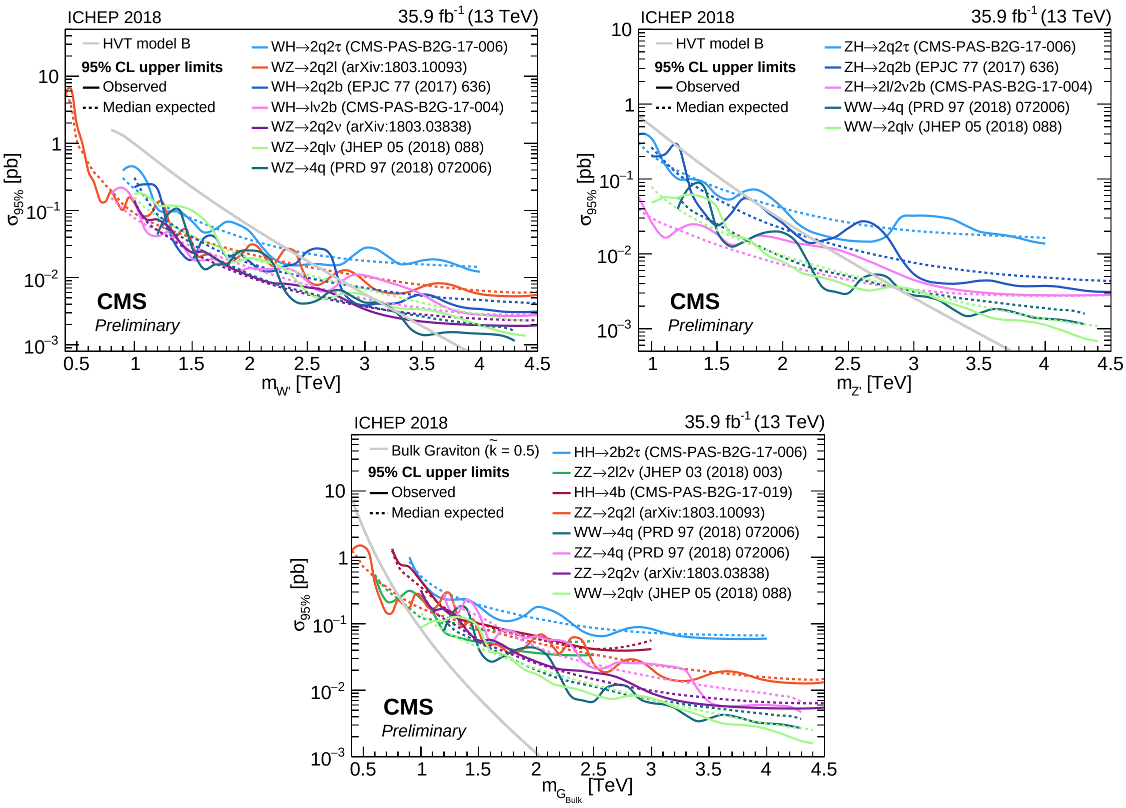


Figure 7.37: Expected and observed 95% CL upper limit on the production cross section of W′ (upper left) and Z′ resonances (upper right), both in the HVT benchmark model B, and of a bulk graviton resonance (bottom plot), with $\tilde{k} = 0.5$, decaying to diboson final states.

Chapter 8

Conclusions

A search for new massive resonances decaying to a pair of Higgs bosons (HH) or to a Higgs boson and a W or Z boson (WH or ZH) in final states with a large-cone jet and the decay products of a τ lepton pair has been presented. In particular, two analyses performed with data collected in pp collisions at different center-of-mass energies have been described.

The first analysis is performed with pp collision data at $\sqrt{s} = 8\,\text{TeV}$ collected in 2012, and is focused on the final state given by the resonant production of two Higgs bosons where one H boson decays to $\tau\tau \rightarrow \ell\tau_\text{h}$, with $\ell = e$ or μ and neutrinos, while the other Higgs boson decays to a pair of bottom quarks. The second analysis is performed with pp collision data at $\sqrt{s} = 13\,\text{TeV}$ collected in 2016, in a final state consistent with a $\text{H} \rightarrow \tau\tau$ decay and with a second SM boson decaying into quarks, with the second boson being a W, Z, or Higgs boson.

Specialized methods are studied and developed to reconstruct and identify the visible decay of the highly Lorentz-boosted di-τ pair produced by the decay of the Higgs boson candidate. In each event, the visible di-τ system is combined with the missing momentum reconstructed in the event from the neutrinos generated in the τ decay, in order to reconstruct the kinematics of the H boson candidate. Recoiling against it, a large-cone jet with a compatible mass is identified with advanced techniques, referred to as V tagging and H tagging, that help distinguish hadronic decays of massive bosons and achieve a large suppression of background from the QCD mulijet and W+jets processes, based on the spatial distribution of the jet constituents and the jet mass. In particular, the H-tagging algorithm combines jet-substructure information with identification techniques based on the peculiarities of jets with multiple b-quarks, such as the presence of displaced tracks or a secondary vertices.

The search is then performed combining the two boson candidates and computing the invariant mass of the system. The signal of a new resonance would manifest itself as a localized excess over the smoothly-falling background distribution.

In this analysis, data are found in agreement with the standard model background expectations and then exclusion limits are set for the product of the new resonance production and its branching ratio to a pair of bosons. Warped extra dimensions models and heavy vector triplet models that predict spin-0, spin-1 and spin-2 resonances are considered as benchmark scenarios for the result interpretation. The HH Run 1 analysis set 95% confidence level (CL) upper limits on the product of the cross section and the branching ratio of a spin-0 resonance decaying to a pair of Higgs bosons, from 850 to 30 fb for resonances with masses between 800 to 2500 GeV, and excluded bulk radions (with $\Lambda_R = 1$ TeV) for masses between 950 and 1150 GeV.

In the analogous final state, in Run 2, 95% CL upper limits for the resonant production of Higgs boson pairs are set and range between 80 and 5 fb, for resonance masses ranging for between 900 and 4000 GeV, with spin-0 and spin-2 hypotheses. In the benchmark bulk-radion model with $\Lambda_R = 1$ TeV, the exclusion of the radion resonance decaying to HH was extended to 2.7 TeV.

For a spin-1 signal, the upper limits at 95% CL range from 180 to 5 fb for resonance masses between 900 GeV and 4000 GeV. The results are interpreted in the context of the simplified HVT benchmark model and W′, Z′, or mass degenerate V′ resonances are excluded at 95% CL for masses lower than 2.6 TeV, 1.8 TeV and 2.8 TeV, respectively, in the HVT benchmark model B.

This analysis is part of a set of searches for heavy resonances decaying into dibosons. The sensitivity in this channel is found to be comparable to searches performed in other diboson final states, therefore the best results would be provided by a statistical combination of all searches. This combination is at the moment being performed for the 2016 data analyses and will be updated at the end of Run 2, with 2017 and 2018 recorded data, achieving a target integrated luminosity of 150 fb^{-1}, providing an unprecedented ability for probing such a unique phase-space.

Appendix A

Statistical approach

The method used to extract limits on the signal strength is the modified frequentist approach, also known as the CLs criterion [157, 158]. The method is characterized by the statistical uncertainty treatment and the test statistics, which are based on a profile likelihood ratio. A description of the CLs method is reported here, whereas more information can be found in the description from the LHC Higgs Combination group [171]. The parameters adopted to build the statistical model of the data distribution are the number of signal events s, as predicted by the theory model that is tested, the yield of background processes b, the signal strength modifier μ, and nuisance parameters θ, that account for the systematic uncertainties that affect the expectations for signal and background, $s(\theta)$ and $b(\theta)$. All systematic uncertainties are taken either fully correlated (100% - positive or negative) or fully uncorrelated (independent).

The likelihood function is built starting from a Poissonian probability function $\mathcal{L}(data \mid \mu, \theta)$:

$$\mathcal{L}(data \mid \mu, \theta) = \text{Poisson}(data \mid \mu \cdot s(\theta) + b(\theta))p(\bar{\theta} \mid \theta), \tag{A.1}$$

where data represents the measurement observation or pseudo-data, θ represents the full suite of nuisance parameters and $p(\bar{\theta} \mid \theta)$ are the *probability distribution functions* (*pdfs*) of the nuisance parameters. Following Bayes' theorem, also posterior *pdfs* can be defined as:

$$\rho(\theta \mid \bar{\theta}) \sim p(\bar{\theta} \mid \theta)\pi_\theta(\theta), \tag{A.2}$$

where $(\pi_\theta(\theta))$ are hyper-priors for those measurements, chosen usually to be uniformly distributed. With this choice, if $p(\bar{\theta} \mid \theta)$ is a normal, $\rho(\theta \mid \bar{\theta})$ is a normal or a log-normal distribution, while if $p(\bar{\theta} \mid \theta)$ is a Poissonian, $\rho(\theta \mid \bar{\theta})$ is a gamma distribution. The type of source of uncertainty determines the assumption of the prior. If no assumption can be made on a parameter from prior measurements or consideration, the proper distribution for the prior is uniform. A Gaussian function is indicated for parameters that can assume both positive and negative values, while a log-normal distribution is suited for parameters that can assume only positive values, like cross sections, selection efficiency, and luminosity. Gamma distributions are used for uncertainties of statistical nature, e.g. parameters were the primary uncertainty source is the statistics of events in a control region.

The nuisance *pdfs* can be used to constrain the likelihood of the main parameters or to construct sampling distributions of the test statistics. Consider an unbinned likelihood, with k observed events,

$$\text{Poisson}(data \mid \mu \cdot s(\theta) + b(\theta)) = k^{-1}\Pi_i(\mu S f_s(x_i) + B f_b(x_i))e^{-(\mu S+B)}, \qquad \text{(A.3)}$$

where f_s and f_b are the signal and background *pdfs*, relative to the observable x_i, while S and B are the expected number of signal and background events.

A test statistics $\tilde{q}_\mu$ can be build to test the compatibility of the data with the background-only or signal+background hypotheses, where the signal is allowed to be scaled by some strength factor μ, based on the profile likelihood ratio:

$$\tilde{q}_\mu = -2\ln\frac{\mathcal{L}(data \mid \mu, \hat{\theta}_\mu)}{\mathcal{L}(data \mid \hat{\mu}, \hat{\theta})}, \qquad \text{with the constraint} \qquad 0 \leq \hat{\mu} \leq \mu, \qquad \text{(A.4)}$$

where $\hat{\theta}_\mu$ is the conditional maximal estimator of θ, given the signal strength parameter μ, while $\hat{\mu}$ and $\hat{\theta}$ are the parameter estimators the correspond to the global maximum of the likelihood. The maximum likelihood estimator of the signal strength $\hat{\mu}$ is defined to be positive (signal rate is positive) and has an upper boundary $\leq \mu$ imposed by hand to guarantee a one-sided confidence interval, which from the physics point of view means that upward fluctuations of the data such that $\hat{\mu} \geq \mu$ are not considered against the signal hypothesis (a signal with strength μ).

Given the μ hypothesis, the test statistic is measured in data, $\tilde{q}_\mu^{obs}$, as well as the values of the nuisance parameters best describing the observed data (i.e. maximizing the likelihood) in the background-only $\hat{\theta}_0^{obs}$ or signal + background $\hat{\theta}_\mu^{obs}$ hypotheses.

Toy Monte Carlo pseudo-data are then generated to construct the test statistics *pdfs* $f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{obs})$, and $f(\tilde{q}_\mu|0, \hat{\theta}_0^{obs})$ assuming a signal strength μ in the signal + background hypothesis and the background-only hypothesis ($\mu = 0$). The nuisance parameters are fixed to the measured values in data $\hat{\theta}_0^{obs}$ and $\hat{\theta}_\mu^{obs}$ while generating the pseudo-experiment, but they are allowed to float in the fits that are required to evaluate the test statistics $\tilde{q}_\mu$. The p-values associated to the signal plus background and background-only hypothesis are defined as:

$$p_\mu = P(\tilde{q}_\mu > \tilde{q}_\mu^{obs}|\text{signal} + \text{background}) = \\ P(\tilde{q}_\mu \geq \tilde{q}_\mu^{obs}|\mu s(\theta_\mu^{obs}) + b(\theta_\mu^{obs})) = \int_{\tilde{q}_\mu^{obs}}^{\infty} f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{obs}) d\tilde{q}_\mu, \quad \text{(A.5)}$$

$$p_0 = 1 - p_b = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{obs}|\text{background only}) \\ = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{obs}|b(\theta_0^{obs})) = \int_{\tilde{q}_0^{obs}}^{\infty} f(\tilde{q}_\mu|0, \hat{\theta}_0^{obs}) d\tilde{q}_\mu. \quad \text{(A.6)}$$

The CLs is defined as the ratio of the p-values:

$$\mathrm{CL_s}(\mu) = \frac{\mathrm{CL_{s+b}}}{\mathrm{CL_b}} = \frac{p_\mu}{1 - p_b} = \frac{p_\mu}{p_0}. \quad \text{(A.7)}$$

Given the confidence level α, if CLs $< \alpha$, a model with signal strength μ is excluded at $(1-\alpha)$ confidence level (CL). E.g. the 95% CL observed upper limit on the theoretical model is set by solving the equation CLs$(\mu) = 0.05$ for μ. Similarly, upper expected limits, along with the 1 and 2σ uncertainty bands, can be extracted by generating pseudo-data under the background-only hypothesis, and by calculating the corresponding CLs and 95% upper limit for each of the pseudo-data. A cumulative distribution of the calculated upper limits is then constructed: the 50% quantile corresponds to the median expected, the 2.5% and 97.5% quantiles correspond respectively to the $\pm 2\sigma$ (95%) uncertainty bands, and the 16% and 84% quantiles to $\pm 1\sigma$ (68%) uncertainty bands.

A.1 Profile likelihood asymptotic approximation

Without the physical requirement $\hat{\mu} \geq 0$, the test statistic is $\tilde{q_\mu} = q_\mu$, with

$$q_\mu = -2\ln\frac{\mathcal{L}(data \mid \mu, \hat{\theta}_\mu)}{\mathcal{L}(data \mid \hat{\mu}, \hat{\theta})}, \text{with the constraint } \hat{\mu} \leq \mu. \quad \text{(A.8)}$$

Following Wilks' theorem [173], in the asymptotic regime, q_μ is expected to follow the distribution of a $\frac{1}{2}\chi^2$ with one degree of freedom (taken as the difference between the degrees of freedom of the numerator and the denominator of the likelihood ratio), since the hypothesis tested is to have a signal with strength (any positive real number - one degree of freedom) with respect to having a signal of strength 0 (one single point - 0 degree of freedom).

Since $\mathrm{CL_s}(\mu) = \frac{\mathrm{CL_{s+b}}}{\mathrm{CL_b}}$, the value of μ that makes $\frac{1}{2}q_\mu = 1.92$ corresponds to a $\mathrm{CL_{s+b}} = 0.025$, which for cases where the observation is equal to the expectations from the background $\mathrm{CL_b} = 0.5$, yields to $\mathrm{CL_s} = 0.05$.

However with the physical requirement $\hat{\mu} \geq 0$, the test statistic $\tilde{q}_\mu$ does not follow exactly a $\frac{1}{2}\chi^2$, yet, it follows the formula [174]:

$$f(\tilde{q}_\mu|\mu) = \frac{1}{2}\delta(\tilde{q}_\mu) + \begin{cases} \frac{e^{-\tilde{q}_\mu/2}}{2\sqrt{2\pi\tilde{q}_\mu}}, & for\ 0 < \tilde{q}_\mu < \mu^2/\sigma^2 \\ \frac{1}{(2\mu/\sigma)\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(\tilde{q}_\mu + (\mu/\sigma)^2)^2}{(2\mu/\sigma)^2}}, & for\ \tilde{q}_\mu > \mu^2/\sigma^2, \end{cases} \tag{A.9}$$

where $\sigma^2 = \mu^2/q_{\mu,A}$ is the test statistic evaluated with the Asimov data set, i.e. the data set of the expected background and nominal nuisance parameters (setting all fluctuations to zero).

The function $f(\tilde{q}_\mu|b)$ can be used to extract expected limits and 1 and 2 σ bands without generating toy Monte Carlo experiments. It is demonstrated in [174] that $\tilde{q}_\mu$ and q_μ are equivalent in the asymptotic limit. The upper limits can be also extracted from the equation:

$$\mathrm{CL}_s = 0.05 = \frac{1 - \Phi(\sqrt{q_\mu})}{\Phi(\sqrt{q_{\mu,A}} - \sqrt{q_\mu})} \tag{A.10}$$

where Φ is the cumulative of a standard Gaussian function and Φ^{-1} is the quantile. Then the median and expected error band can be computed as

$$\mu_{up+N} = \sigma \cdot (\Phi^{-1}(1 - \alpha\Phi(N)) + N) \tag{A.11}$$

with $\alpha = 0.05$ and $\mu \equiv \mu_{up}^{med}$ in the calculation of σ, so that the median expected CL_s is obtained for $N = 0$:

$$\mu_{up}^{med} = \sigma \cdot (\Phi^{-1}(1 - \alpha 0.5)) = \sigma \cdot \Phi^{-1}(0.975). \tag{A.12}$$

The asymptotic is a good approximation of the full CLs method, but possible biases can arise in application cases with a small number of events.

A.2 Quantification of data excess

In the case of observing an excess in data events with respect to what was expected from the background prediction, the characterization begins with the p-value calculation if the upward fluctuation of the background-only hypothesis. This is done by tossing pseudo-data in the background-only hypothesis and building up the corresponding parton distribution function of the test statistics.

$$\mathrm{p-value} = P(q_0 > q_0^{obs}) = \int_{q_0^{obs}}^{\infty} f(q_\mu|0, \hat{\theta}_0^{obs}) dq_0 \tag{A.13}$$

where q_0^{obs} is the observed test statistic value in data calculated for $\mu = 0$ withe the only assumption that $\hat{\mu} \geq 0$, i.e. data deficits are not used against the background hypothesis and treated differently than excesses. An estimation of the p-value, known as *local* can be calculated as [171]:

$$\mathrm{p-value}^{estimate} = \frac{1}{2}\left[1 - \mathrm{erf}(\sqrt{q_0^{obs}/2}\right] \tag{A.14}$$

A more accurate characterization of the local p-value should take into account the effect of the choice in possible values of invariant mass, known as the $look - elsewhere$ effect [169].

Bibliography

[1] S. L. Glashow, “Partial Symmetries of Weak Interactions”, *Nucl. Phys.* **22** (1961) 579–588, `doi:10.1016/0029-5582(61)90469-2`.

[2] S. Weinberg, “A Model of Leptons”, *Phys. Rev. Lett.* **19** (Nov, 1967) 1264–1266, `doi:10.1103/PhysRevLett.19.1264`.

[3] A. Salam, “Weak and Electromagnetic Interactions”, *Conf. Proc.* **C680519** (1968) 367–377.

[4] Particle Data Group Collaboration, “Review of Particle Physics”, *Chin. Phys.* **C40** (2016), no. 10, 100001, `doi:10.1088/1674-1137/40/10/100001`.

[5] Z. Maki, M. Nakagawa, and S. Sakata, “Remarks on the Unified Model of Elementary Particles”, *Progress of Theoretical Physics* **28** (1962), no. 5, 870–880, `doi:10.1143/PTP.28.870`.

[6] M. Gell-Mann, “Symmetries of Baryons and Mesons”, *Phys. Rev.* **125** (Feb, 1962) 1067–1084, `doi:10.1103/PhysRev.125.1067`.

[7] N. Cabibbo, “Unitary Symmetry and Leptonic Decays”, *Phys. Rev. Lett.* **10** (Jun, 1963) 531–533, `doi:10.1103/PhysRevLett.10.531`.

[8] M. Kobayashi and T. Maskawa, “CP Violation in the Renormalizable Theory of Weak Interaction”, *Prog. Theor. Phys.* **49** (1973) 652–657, `doi:10.1143/PTP.49.652`.

[9] T. U. Collaboration, “Experimental observation of isolated large transverse energy electrons with associated missing energy at s=540 GeV”, *Physics Letters B* **122** (1983), no. 1, 103 – 116, `doi:https://doi.org/10.1016/0370-2693(83)91177-2`.

[10] T. U. Collaboration, "Experimental observation of lepton pairs of invariant mass around 95 GeV/c2 at the CERN SPS collider", *Physics Letters B* **126** (1983), no. 5, 398 – 410, doi:https://doi.org/10.1016/0370-2693(83)90188-0.

[11] T. U. Collaboration, "Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider", *Physics Letters B* **122** (1983), no. 5, 476 – 485, doi:https://doi.org/10.1016/0370-2693(83)91605-2.

[12] T. U. Collaboration, "Evidence for $Z_0 \to e^+e^-$ at the CERN pp collider", *Physics Letters B* **129** (1983), no. 1, 130 – 140, doi:https://doi.org/10.1016/0370-2693(83)90744-X.

[13] C. Rubbia, "Experimental observation of the intermediate vector bosons W^+, W^-, and Z^0", *Rev. Mod. Phys.* **57** (Jul, 1985) 699–722, doi:10.1103/RevModPhys.57.699.

[14] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.* **13** (Aug, 1964) 321–323, doi:10.1103/PhysRevLett.13.321.

[15] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (Oct, 1964) 508–509, doi:10.1103/PhysRevLett.13.508.

[16] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global Conservation Laws and Massless Particles", *Phys. Rev. Lett.* **13** (Nov, 1964) 585–587, doi:10.1103/PhysRevLett.13.585.

[17] LHC Higgs Cross Section Working Group Collaboration, "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector", doi:10.23731/CYRM-2017-002, arXiv:1610.07922.

[18] LHC Higgs Cross Section Working Group et al., "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties", *CERN-2013-004* (CERN, Geneva, 2013) doi:10.5170/CERN-2013-004, arXiv:1307.1347.

[19] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett.* **B716** (2012) 1–29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

[20] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett.* **B716** (2012) 30–61, `doi:10.1016/j.physletb.2012.08.021`, `arXiv:1207.7235`.

[21] CMS Collaboration, "Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV", *JHEP* **06** (2013) 081, `doi:10.1007/JHEP06(2013)081`, `arXiv:1303.4571`.

[22] ATLAS, CMS Collaboration, "Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments", *Phys. Rev. Lett.* **114** (2015) 191803, `doi:10.1103/PhysRevLett.114.191803`, `arXiv:1503.07589`.

[23] CMS Collaboration, "Study of the Mass and Spin-Parity of the Higgs Boson Candidate Via Its Decays to Z Boson Pairs", *Phys. Rev. Lett.* **110** (2013), no. 8, 081803, `doi:10.1103/PhysRevLett.110.081803`, `arXiv:1212.6639`.

[24] ATLAS, CMS Collaboration, "Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV", *JHEP* **08** (2016) 045, `doi:10.1007/JHEP08(2016)045`, `arXiv:1606.02266`.

[25] CMS Collaboration, "Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector", *Phys. Lett.* **B779** (2018) 283–316, `doi:10.1016/j.physletb.2018.02.004`, `arXiv:1708.00373`.

[26] ATLAS Collaboration, "Cross-section measurements of the Higgs boson decaying to a pair of tau leptons in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", Technical Report ATLAS-CONF-2018-021, CERN, Geneva, Jun, 2018.

[27] CMS Collaboration, "Inclusive search for a highly boosted Higgs boson decaying to a bottom quark-antiquark pair", *Phys. Rev. Lett.* **120** (2018), no. 7, 071802, `doi:10.1103/PhysRevLett.120.071802`, `arXiv:1709.05543`.

[28] ATLAS Collaboration, "Evidence for the $H \to b\bar{b}$ decay with the ATLAS detector", Technical Report ATLAS-CONF-2017-041, CERN, Geneva, Jul, 2017.

[29] CMS Collaboration, "Evidence for the decay of the Higgs Boson to Bottom Quarks", Technical Report CMS-PAS-HIG-16-044, CERN, Geneva, 2017.

[30] CMS Collaboration, “Search for the Higgs boson decaying to two muons in proton-proton collisions at $\sqrt{s}$ =13 TeV”, *Submitted to: Phys. Rev. Lett.* (2018) arXiv:1807.06325.

[31] CMS Collaboration, “Combined measurements of Higgs boson couplings in proton-proton collisions at $\sqrt{s}$ = 13 TeV”, arXiv:1809.10733.

[32] A. D. Sakharov, “Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe”, *Pisma Zh. Eksp. Teor. Fiz.* **5** (1967) 32–35, doi:10.1070/PU1991v034n05ABEH002497. [Usp. Fiz. Nauk161,no.5,61(1991)].

[33] F. Zwicky, “Die Rotverschiebung von extragalaktischen Nebeln”, *Helvetica Physica Acta* **6** (1933) 110–127.

[34] V. C. Rubin and W. K. Ford, Jr., “Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions”, *The Astrophysical Journal* **159** (February, 1970) 379, doi:10.1086/150317. "http://dx.doi.org/10.1086/150317".

[35] Supernova Cosmology Project Collaboration, “Measurements of Omega and Lambda from 42 high redshift supernovae”, *Astrophys. J.* **517** (1999) 565–586, doi:10.1086/307221, arXiv:astro-ph/9812133.

[36] T. Kaluza, “Zum Unitätsproblem in der Physik”, *Sitzungsber. Preuss. Akad. Wiss. Berlin,.(Math. Phys.)* (1921) 966–97.

[37] O. Klein, “The Atomicity of Electricity as a Quantum Theory Law”, *Nature* **118** (1926) 516 EP, doi:10.1143/PTP.28.870.

[38] O. Klein, “Quantentheorie und fünfdimensionale Relativitätstheorie”, *Zeitschrift für Physik* **37** (Dec, 1926) 895–906, doi:10.1007/BF01397481.

[39] L. Randall and R. Sundrum, “A Large mass hierarchy from a small extra dimension”, *Phys. Rev. Lett.* **83** (1999) 3370–3373, doi:10.1103/PhysRevLett.83.3370, arXiv:hep-ph/9905221.

[40] C. Csaki, M. L. Graesser, and G. D. Kribs, “Radion dynamics and electroweak physics”, *Phys. Rev.* **D63** (2001) 065002, doi:10.1103/PhysRevD.63.065002, arXiv:hep-th/0008151.

[41] M. Gouzevitch et al., "Scale-invariant resonance tagging in multijet events and new physics in Higgs pair production", *JHEP* **07** (2013) 148, doi:10.1007/JHEP07(2013)148, arXiv:1303.6636.

[42] A. Oliveira, "Gravity particles from Warped Extra Dimensions, predictions for LHC", arXiv:1404.0102.

[43] A. L. Fitzpatrick, J. Kaplan, L. Randall, and L.-T. Wang, "Searching for the Kaluza-Klein Graviton in Bulk RS Models", *JHEP* **09** (2007) 013, doi:10.1088/1126-6708/2007/09/013, arXiv:hep-ph/0701150.

[44] V. Barger, W. Y. Keung, and E. Ma, "Gauge model with light W and Z bosons", *Phys. Rev. D* **22** (Aug, 1980) 727–737, doi:10.1103/PhysRevD.22.727.

[45] C. Grojean, E. Salvioni, and R. Torre, "A weakly constrained W' at the early LHC", *JHEP* **07** (2011) 002, doi:10.1007/JHEP07(2011)002, arXiv:1103.2761.

[46] R. Contino, D. Pappadopulo, D. Marzocca, and R. Rattazzi, "On the effect of resonances in composite Higgs phenomenology", *Journal of High Energy Physics* **2011** (Oct, 2011) 81, doi:10.1007/JHEP10(2011)081.

[47] B. Bellazzini, C. Csáki, and J. Serra, "Composite Higgses", *Eur. Phys. J.* **C74** (2014), no. 5, 2766, doi:10.1140/epjc/s10052-014-2766-x, arXiv:1401.2457.

[48] D. Pappadopulo, A. Thamm, R. Torre, and A. Wulzer, "Heavy Vector Triplets: Bridging Theory and Data", *JHEP* **09** (2014) 060, doi:10.1007/JHEP09(2014)060, arXiv:1402.4431.

[49] ATLAS Collaboration, "Searches for Higgs boson pair production in the $hh \to \mathrm{b\bar{b}}\tau\tau, \gamma\gamma\mathrm{WW}^*, \gamma\gamma\mathrm{b\bar{b}}, bbbb$ channels with the ATLAS detector", *Phys. Rev.* **D92** (2015) 092004, doi:10.1103/PhysRevD.92.092004, arXiv:1509.04670.

[50] "Search for Higgs boson pair production in the $b\bar{b}\gamma\gamma$ final state using pp collision data at $\sqrt{s} = 13$ TeV with the ATLAS detector", Technical Report ATLAS-CONF-2016-004, CERN, Geneva, Mar, 2016.

[51] CMS Collaboration, "Search for Higgs boson pair production in the $bb\tau\tau$ final state in proton-proton collisions at $\sqrt{s} = 8$ TeV", *Phys. Rev.* **D96** (2017), no. 7, 072004, doi:10.1103/PhysRevD.96.072004, arXiv:1707.00350.

[52] CMS Collaboration, "Search for heavy resonances decaying to two Higgs bosons in final states containing four b quarks", *Eur. Phys. J.* **C76** (2016), no. 7, 371, doi:10.1140/epjc/s10052-016-4206-6, arXiv:1602.08762.

[53] B. Cooper, N. Konstantinidis, L. Lambourne, and D. Wardrope, "Boosted $hh \to b\bar{b}b\bar{b}$: A new topology in searches for TeV-scale resonances at the LHC", *Phys. Rev.* **D88** (2013), no. 11, 114005, doi:10.1103/PhysRevD.88.114005, arXiv:1307.0407.

[54] CMS Collaboration, "Combination of searches for heavy resonances decaying to WW, WZ, ZZ, WH, and ZH boson pairs in proton–proton collisions at $\sqrt{s} = 8$ and 13 TeV", *Phys. Lett.* **B774** (2017) 533–558, doi:10.1016/j.physletb.2017.09.083, arXiv:1705.09171.

[55] L. Evans and P. Bryant, "LHC Machine", *JINST* **3** (2008) S08001, doi:10.1088/1748-0221/3/08/S08001.

[56] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *JINST* **3** (2008) S08002, doi:10.1088/1748-0221/3/08/S08002.

[57] LHCb Collaboration, "The LHCb Detector at the LHC", *JINST* **3** (2008) S08005, doi:10.1088/1748-0221/3/08/S08005.

[58] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *Journal of Instrumentation* **3** (2008), no. 08, S08003. http://stacks.iop.org/1748-0221/3/i=08/a=S08003.

[59] CMS Collaboration, "CMS physics: Technical design report",.

[60] CERN, "CERN accelerator complex". http://te-epc-lpc.web.cern.ch/te-epc-lpc/machines/pagesources/Cern-Accelerator-Complex.jpg.

[61] CMS Collaboration, "2012 Proton-Proton Collisions: Luminosity versus day". http://cms-service-lumi.web.cern.ch/cms-service-lumi/publicplots/int_lumi_per_day_cumulative_pp_2012.png.

[62] CMS Collaboration, "2016 Proton-Proton 13 TeV Collisions: Luminosity versus day". https://cms-service-lumi.web.cern.ch/cms-service-lumi/publicplots/int_lumi_per_day_cumulative_pp_2016.pdf.

[63] C. Collaboration, "Detector Drawings", (Mar, 2012). CMS Collection.

[64] CMS Collaboration, "Heavy Ion Physics Program in CMS Experiment", *Nucl. Phys.* **A830** (2009) 97C–104C, arXiv:0907.4492. "10.1016/j.nuclphysa.2009.09.014".

[65] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004. http://dx.doi.org/10.1088/1748-0221/3/08/S08004.

[66] CMS Collaboration, "CMS Physics: Technical Design Report Volume 1: Detector Performance and Software", technical report, Geneva, 2006. "https://cds.cern.ch/record/922757".

[67] CMS Collaboration, "CMS technical design report, volume II: Physics performance", *J. Phys.* **G34** (2007), no. 6, 995–1579, doi:10.1088/0954-3899/34/6/S01.

[68] C. Grupen and B. Schwartz, "Particle detectors". Cambridge, UK: Cambridge Univ. Pr. (2008) 651 p, 2008.

[69] CMS Collaboration, "The CMS magnet project: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 1997.

[70] CMS Collaboration, "The CMS tracker system project: Technical Design Report", technical report, Geneva, 1997. "https://cds.cern.ch/record/368412".

[71] CMS Collaboration, "The CMS tracker: addendum to the Technical Design Report", technical report, Geneva, 2000. "https://cds.cern.ch/record/490194".

[72] M. Weber, "The CMS tracker", Technical Report CMS-CR-2004-046, CERN, Geneva, Oct, 2004. https://cds.cern.ch/record/800402.

[73] CMS Collaboration, "The CMS electromagnetic calorimeter project: Technical Design Report", technical report, Geneva, 1997. https://cds.cern.ch/record/349375.

[74] P. Adzic, "Energy resolution of the barrel of the CMS Electromagnetic Calorimeter", *Journal of Instrumentation* **2** (2007), no. 04, P04004.

[75] CMS Collaboration, "CMS Physics: Technical Design Report Volume 1: Detector Performance and Software". Technical Design Report CMS. CERN, Geneva, 2006.

[76] CMS Collaboration, "Measurement of CMS Luminosity ", Technical Report CMS-PAS-EWK-10-004, CERN, 1900. Geneva, 1900.

[77] CMS Collaboration, "Performance of the CMS Hadron Calorimeter with Cosmic Ray Muons and LHC Beam Data", *JINST* **5** (2010) T03012, `doi:10.1088/1748-0221/5/03/T03012`, `arXiv:0911.4991`. `http://dx.doi.org/10.1088/1748-0221/5/03/T03012`.

[78] CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET", CMS Physics Analysis Summary CMS-PAS-PFT-09-001, CERN, 2009.

[79] CMS Collaboration, "The CMS muon project: Technical Design Report", technical report, Geneva, 1997. `https://cds.cern.ch/record/343814`.

[80] CMS Collaboration, "CMS reconstruction improvement for the muon tracking by the RPC chambers", *PoS* **RPC2012** (2012) 045, `doi:10.1088/1748-0221/8/03/T03001`, `arXiv:1209.2646`. JINST8,T03001 (2013).

[81] CMS Collaboration, "The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s} = 7$ TeV at the LHC", *JINST* **8** (2013) P11002, `doi:10.1088/1748-0221/8/11/P11002`, `arXiv:1306.6905`.

[82] CMS Collaboration, "CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems", technical report. `"http://cds.cern.ch/record/706847"`.

[83] CMS Collaboration, "CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project", technical report, Geneva, 2002. `"http://cds.cern.ch/record/578006"`.

[84] C. e. a. Eck, “LHC computing Grid: Technical Design Report”, technical report, Geneva, 2005.

[85] CMS Collaboration, “Particle-flow reconstruction and global event description with the CMS detector”, *JINST* **12** (2017) P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.

[86] P. Billoir, “Progressive track recognition with a Kalman like fitting procedure”, *Comput. Phys. Commun.* **57** (1989) 390–394, doi:10.1016/0010-4655(89)90249-X. http://dx.doi.org/10.1016/0010-4655(89)90249-X.

[87] R. Fruhwirth, “Application of Kalman filtering to track and vertex fitting”, *Nucl. Instrum. Meth.* **A262** (1987) 444–450, doi:10.1016/0168-9002(87)90887-4. http://dx.doi.org/10.1016/0168-9002(87)90887-4.

[88] P. Billoir and S. Qian, “Simultaneous pattern recognition and track fitting by the Kalman filtering method”, *Nucl. Instrum. Meth.* **A294** (1990) 219–228, doi:10.1016/0168-9002(90)91835-Y. http://dx.doi.org/10.1016/0168-9002(90)91835-Y.

[89] CMS Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, *JINST* **9** (2014), no. 10, P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569. http://dx.doi.org/10.1088/1748-0221/9/10/P10009.

[90] CMS Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, *JINST* **9** (2014), no. 10, P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.

[91] T. Speer et al., “Vertex Fitting in the CMS Tracker”, Technical Report CMS-NOTE-2006-032, CERN, Geneva, Feb, 2006.

[92] K. Rose, “Deterministic Annealing for Clustering, Compression, Classification, Regression, and Related Optimization Problems”, in *Proceedings of the IEEE*, pp. 2210–2239. 1998.

[93] R. Frühwirth, W. Waltenberger, and P. Vanlaer, “Adaptive Vertex Fitting”, Technical Report CMS-NOTE-2007-008, CERN, Geneva, Mar, 2007. https://cds.cern.ch/record/1027031.

[94] CMS Collaboration, "CMS Luminosity Public results". https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.

[95] CMS Collaboration, "Performance of CMS Muon Reconstruction in Cosmic-Ray Events", *JINST* **5** (2010) T03022, doi:10.1088/1748-0221/5/03/T03022, arXiv:0911.4994.

[96] CMS Collaboration, "Performance of muon reconstruction including Alignment Position Errors for 2016 Collision Data",.

[97] CMS Collaboration, "CMS Muon POG work in progress". https://twiki.cern.ch/twiki/bin/view/CMS/MuonWorkInProgressAndPagResults.

[98] CMS Collaboration, "Muon performance in early 2016 collision data",.

[99] CMS Collaboration, "*Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s}$ = 8 TeV*", *JINST* **10** (2015), no. 06, P06005, doi:10.1088/1748-0221/10/06/P06005, arXiv:1502.02701. http://dx.doi.org/10.1088/1748-0221/10/06/P06005.

[100] W. Adam, R. Frühwirth, A. Strandlie, and T. Todorov, "Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC", *Journal of Physics G: Nuclear and Particle Physics* **31** (2005), no. 9, N9.

[101] CMS Collaboration, "Electron reconstruction and identification at $\sqrt{s} = 7$ TeV", Technical Report CMS-PAS-EGM-10-004, CERN, Geneva, 2010. https://cds.cern.ch/record/1299116.

[102] CMS Collaboration, "Electron and Photon performance using data collected by CMS at $\sqrt{s} = 13$ TeV and 25ns", technical report, Dec, 2015. https://cds.cern.ch/record/2118397.

[103] CMS Collaboration, "Electron and photon performance in CMS with the full 2016 data sample.",.

[104] CMS Collaboration, "CMS Electron recommendation for Run 2". https://twiki.cern.ch/twiki/bin/view/CMS/CutBasedElectronIdentificationRun2.

[105] CMS Collaboration, "Search for resonant pair production of Higgs bosons decaying to $b\bar{b}$ and $\tau^+\tau^-$ in proton-proton collisions at $\sqrt{s} = 8$ TeV", Technical Report CMS-PAS-EXO-15-008, CERN, Geneva, 2015.

[106] CMS Collaboration, "Search for narrow high-mass resonances in proton-proton collisions at $\sqrt{s} = 8$ TeV decaying to Z and Higgs bosons", *Phys. Lett. B* **748** (Feb, 2015) 255. 23 p. Replaced with published version. Added journal reference and DOI.

[107] "Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV", *Journal of Instrumentation* **10** (2015), no. 08, P08010.

[108] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet user manual", *The European Physical Journal C* **72** (Mar, 2012) 1896, doi:10.1140/epjc/s10052-012-1896-2.

[109] S. D. Ellis and D. E. Soper, "Successive combination jet algorithm for hadron collisions", *Phys. Rev. D* **48** (Oct, 1993) 3160–3166, doi:10.1103/PhysRevD.48.3160.

[110] M. Cacciari, G. P. Salam, and G. Soyez, "The Anti-k(t) jet clustering algorithm", *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189. http://dx.doi.org/10.1088/1126-6708/2008/04/063.

[111] CMS Collaboration, "Pileup Removal Algorithms", Technical Report CMS-PAS-JME-14-001, CERN, Geneva, 2014.

[112] CMS Collaboration, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS", *JINST* **6** (2011) P11002, doi:10.1088/1748-0221/6/11/P11002, arXiv:1107.4277. http://dx.doi.org/10.1088/1748-0221/6/11/P11002.

[113] CMS Collaboration, "Performance of jets at CMS", *J. Phys.: Conf. Ser.* **587** (2015), no. 1, 012004. 6 p. https://cds.cern.ch/record/2158943.

[114] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", *JINST* **12** (2017), no. 02, P02014, doi:10.1088/1748-0221/12/02/P02014, arXiv:1607.03663.

[115] CMS Collaboration, "Jet Energy Scale pubblic page", technical report, 2018. https://twiki.cern.ch/twiki/bin/view/CMS/JetEnergyScale.

[116] "Jet energy scale and resolution performances with 13TeV data",.

[117] D. Bertolini, P. Harris, M. Low, and N. Tran, “Pileup Per Particle Identification”, *JHEP* **10** (2014) 059, doi:10.1007/JHEP10(2014)059, arXiv:1407.6013.

[118] CMS Collaboration, “Jet algorithms performance in 13 TeV data”, Technical Report CMS-PAS-JME-16-003, CERN, Geneva, 2017.

[119] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, “Techniques for improved heavy particle searches with jet substructure”, *Phys. Rev.* **D80** (2009) 051501, doi:10.1103/PhysRevD.80.051501, arXiv:0903.5081.

[120] 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.

[121] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, “Soft Drop”, *JHEP* **05** (2014) 146, doi:10.1007/JHEP05(2014)146, arXiv:1402.2657.

[122] CMS Collaboration, “Identifying Hadronically Decaying Vector Bosons Merged into a Single Jet”, Technical Report CMS-PAS-JME-13-006, CERN, Geneva, 2013.

[123] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, “Towards an understanding of jet substructure”, *JHEP* **09** (2013) 029, doi:10.1007/JHEP09(2013)029, arXiv:1307.0007.

[124] M. Dasgupta, A. Fregoso, S. Marzani, and A. Powling, “Jet substructure with analytical methods”, *Eur. Phys. J.* **C73** (2013), no. 11, 2623, doi:10.1140/epjc/s10052-013-2623-3, arXiv:1307.0013.

[125] CMS Collaboration, “Search for heavy resonances decaying into a vector boson and a Higgs boson in hadronic final states with 2016 data”, Technical Report CMS-PAS-B2G-17-002, CERN, Geneva, 2017.

[126] J. Thaler and K. Van Tilburg, “Maximizing Boosted Top Identification by Minimizing N-subjettiness”, *JHEP* **02** (2012) 093, doi:10.1007/JHEP02(2012)093, arXiv:1108.2701.

[127] CMS Collaboration, “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”, arXiv:1712.07158.

[128] CMS Collaboration, "Measurement of $B\bar{B}$ Angular Correlations based on Secondary Vertex Reconstruction at $\sqrt{s} = 7$ TeV", *JHEP* **03** (2011) 136, doi:10.1007/JHEP03(2011)136, arXiv:1102.3194.

[129] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *Journal of Instrumentation* **8** (2013), no. 04, P04013. http://stacks.iop.org/1748-0221/8/i=04/a=P04013.

[130] CMS Collaboration, "Identification of b quark jets at the CMS Experiment in the LHC Run 2", Technical Report CMS-PAS-BTV-15-001, CERN, Geneva, 2016. http://cds.cern.ch/record/2138504.

[131] CMS Collaboration, "Performance of b tagging at $\sqrt{s} = 8$ TeV in multijet, ttbar and boosted topology events", Technical Report CMS-PAS-BTV-13-001, CERN, Geneva, 2013. "http://cds.cern.ch/record/1581306".

[132] CMS Collaboration, "Identification of b-quark jets with the CMS experiment. Identification of b-quark jets with the CMS experiment", *JINST* **8** (Nov, 2012) P04013. 69 p. Comments: Submitted to the Journal of Instrumentation.

[133] CMS Collaboration, "Performance of CMS muon reconstruction in *pp* collision events at $\sqrt{s} = 7$ TeV", *JINST* **7** (2012) P10002, doi:10.1088/1748-0221/7/10/P10002, arXiv:1206.4071. http://dx.doi.org/10.1088/1748-0221/7/10/P10002.

[134] CMS Collaboration, "Performance of missing energy reconstruction in 13 TeV pp collision data using the CMS detector", Technical Report CMS-PAS-JME-16-004, CERN, Geneva, 2016.

[135] CMS Collaboration, "Performance of MET reconstruction and pileup mitigation techniques in CMS", *Nucl. Part. Phys. Proc.* **273-275** (2016) 2512–2514, doi:10.1016/j.nuclphysbps.2015.09.442, arXiv:1502.05207.

[136] CMS Collaboration, "Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ TeV", *JINST* **10** (2015), no. 02, P02006, doi:10.1088/1748-0221/10/02/P02006, arXiv:1411.0511.

[137] CMS Collaboration, "Reconstruction and identification of τ lepton decays to hadrons and ν_τ at CMS", *JINST* **11** (2016), no. 01, P01019, doi:10.1088/1748-0221/11/01/P01019, arXiv:1510.07488.

[138] CMS Collaboration, “Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV”, arXiv:1809.02816.

[139] CMS Collaboration, “Tau Reconstruction and identification performance in run 2”, technical report, 2015. http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2015/229.

[140] CMS Collaboration, “Tau identification in boosted topologies”, CMS Detector Performance Summary CMS-DP-2016-038, CERN, 2016.

[141] CMS Collaboration, “Search for Neutral MSSM Higgs Bosons Decaying to Tau Pairs in *pp* Collisions at $\sqrt{s} = 7$ TeV”, *Phys. Rev. Lett.* **106** (2011) 231801, doi:10.1103/PhysRevLett.106.231801, arXiv:1104.1619.

[142] L. Bianchini, J. Conway, E. K. Friis, and C. Veelken, “Reconstruction of the Higgs mass in H$\tau\tau$ Events by Dynamical Likelihood techniques”, *Journal of Physics: Conference Series* **513** (2014), no. 2, 022035.

[143] L. Bianchini et al., “Reconstruction of the Higgs mass in events with Higgs bosons decaying into a pair of τ leptons using matrix element techniques”, *Nucl. Instrum. Meth.* **A862** (2017) 54–84, doi:10.1016/j.nima.2017.05.001, arXiv:1603.05910.

[144] J. Alwall et al., “MadGraph 5 : Going Beyond”, *JHEP* **06** (2011) 128, doi:10.1007/JHEP06(2011)128, arXiv:1106.0522.

[145] P. Nason, “A New method for combining NLO QCD with shower Monte Carlo algorithms”, *JHEP* **11** (2004) 040, doi:10.1088/1126-6708/2004/11/040, arXiv:hep-ph/0409146.

[146] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method”, *JHEP* **11** (2007) 070, doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.

[147] S. Alioli, P. Nason, C. Oleari, and E. Re, “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”, *JHEP* **06** (2010) 043, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.

[148] S. Alioli, S.-O. Moch, and P. Uwer, "Hadronic top-quark pair-production with one jet and parton showering", *JHEP* **01** (2012) 137, doi:10.1007/JHEP01(2012)137, arXiv:1110.5251.

[149] T. Sjöstrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 physics and manual", *JHEP* **05** (2006) 026, doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.

[150] S. Jadach, J. H. Kuhn, and Z. Was, "TAUOLA: A Library of Monte Carlo programs to simulate decays of polarized tau leptons", *Comput. Phys. Commun.* **64** (1990) 275, doi:10.1016/0010-4655(91)90038-M.

[151] S. Agostinelli et al., "Geant4—a simulation toolkit", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** (2003), no. 3, 250 – 303. "http://www.sciencedirect.com/science/article/pii/S0168900203013688".

[152] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *JINST* **8** (2013) P04013, doi:10.1088/1748-0221/8/04/P04013, arXiv:1211.4462.

[153] CMS Collaboration, "CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update", CMS Physics Analysis Summary CMS-PAS-LUM-13-001, CERN, 2013.

[154] CMS Collaboration, "Performance of tau-lepton reconstruction and identification in CMS", *JINST* **7** (2012) P01001, doi:10.1088/1748-0221/7/01/P01001, arXiv:1109.6034.

[155] CMS Collaboration, "Determination of jet energy calibration and transverse momentum resolution in CMS", *JINST* **6** (2011) P11002, doi:10.1088/1748-0221/6/11/P11002, arXiv:1107.4277.

[156] CMS Collaboration, "Identification techniques for highly boosted W bosons that decay into hadrons", *JHEP* **12** (2014) 017, doi:10.1007/JHEP12(2014)017, arXiv:1410.4227.

[157] A. L. Read, "Presentation of search results: The CL(s) technique", *J. Phys.* **G28** (2002) 2693, doi:10.1088/0954-3899/28/10/313.

[158] T. Junk, "Confidence level computation for combining searches with small statistics", *Nucl. Instrum. Meth.* **A434** (1999) 435, `doi:10.1016/S0168-9002(99)00498-2`, `arXiv:hep-ex/9902006`.

[159] ATLAS and CMS Collaborations, "Procedure for the LHC Higgs boson search combination in summer 2011", Technical Report ATL-PHYS-PUB-2011-011, CMS-NOTE-2011-005, CERN, 2011.

[160] I. C. L. James Stirling, "Parton Luminosities and Cross Sections". `http://www.hep.ph.ic.ac.uk/~wstirlin/plots/plots.html`.

[161] J. Alwall et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP* **07** (2014) 079, `doi:10.1007/JHEP07(2014)079`, `arXiv:1405.0301`.

[162] T. Sjötrand et al., "An Introduction to PYTHIA 8.2", *Comput. Phys. Commun.* **191** (2015) 159, `doi:10.1016/j.cpc.2015.01.024`, `arXiv:1410.3012`.

[163] N. Davidson et al., "Universal Interface of TAUOLA Technical and Physics Documentation", *Comput. Phys. Commun.* **183** (2012) 821, `doi:10.1016/j.cpc.2011.12.009`, `arXiv:1002.0543`.

[164] M. Cacciari, G. P. Salam, and G. Soyez, "The Anti-k(t) jet clustering algorithm", *JHEP* **04** (2008) 063, `doi:10.1088/1126-6708/2008/04/063`, `arXiv:0802.1189`.

[165] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet User Manual", *Eur. Phys. J. C* **72** (2012) 1896, `doi:10.1140/epjc/s10052-012-1896-2`, `arXiv:1111.6097`.

[166] CMS Collaboration, "Search for a Higgs boson in the mass range from 145 to 1000 GeV decaying to a pair of W or Z bosons", *JHEP* **10** (2015) 144, `doi:10.1007/JHEP10(2015)144`, `arXiv:1504.00936`.

[167] M. Bahr et al., "Herwig++ Physics and Manual", *Eur. Phys. J. C* **58** (2008) 639, `doi:10.1140/epjc/s10052-008-0798-9`, `arXiv:0803.0883`.

[168] CMS Collaboration, "CMS Luminosity Measurement for the 2016 Data Taking Period", CMS Physics Analysis Summary CMS-PAS-LUM-17-001, CERN, 2017.

[169] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J. C* **71** (2011) 1554, doi:10.1140/epjc/s10052-011-1554-0,10.1140/epjc/s10052-013-2501-z, arXiv:1007.1727. [Erratum: *Eur. Phys. J. C* **73** (2013) 2501].

[170] F. Garwood, "(i) Fiducial Limits for the Poisson Distribution", *Biometrika* **28** (1936), no. 3-4, 437–442, doi:10.1093/biomet/28.3-4.437.

[171] The ATLAS Collaboration, The CMS Collaboration, The LHC Higgs Combination Group Collaboration, "Procedure for the LHC Higgs boson search combination in Summer 2011", Technical Report CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, CERN, Geneva, Aug, 2011.

[172] ATLAS Collaboration, "Combination of searches for heavy resonances decaying into bosonic and leptonic final states using 36 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Phys. Rev. D* **98** (Sep, 2018) 052008, doi:10.1103/PhysRevD.98.052008.

[173] S. S. Wilks, "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses", *Ann. Math. Statist.* **9** (03, 1938) 60–62, doi:10.1214/aoms/1177732360.

[174] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *The European Physical Journal C* **71** (Feb, 2011) 1554, doi:10.1140/epjc/s10052-011-1554-0.