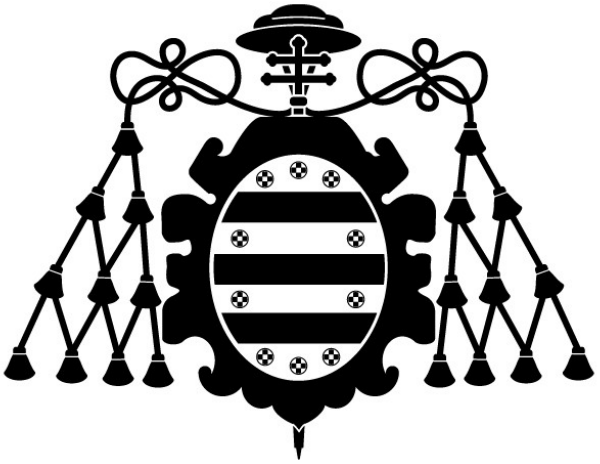


Universidad de Oviedo
Universidá d'Uviéu
University of Oviedo

Programa de Doctorado en Materiales

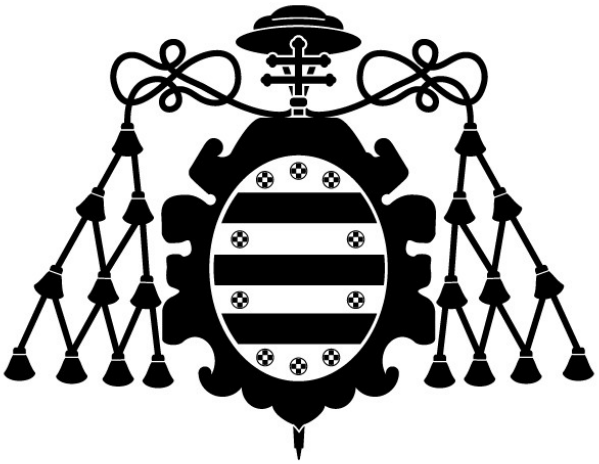
Búsqueda de nueva física en eventos con leptones de alto momento transverso con el detector CMS en el LHC

Search for new physics in events with high transverse momentum leptons with the CMS detector at the LHC

TESIS DOCTORAL

Sergio Sánchez Cruz

Junio 2020



Universidad de Oviedo

Universidá d'Uviéu

University of Oviedo

Programa de Doctorado en Materiales

Búsqueda de nueva física en eventos con leptones de alto momento transverso con el detector CMS en el LHC

Search for new physics in events with high transverse momentum leptons with the CMS detector at the LHC

TESIS DOCTORAL

Director de tesis

Dr. D. Francisco Javier Cuevas Maestro

Universidad de Oviedo
Universidá d'Uviéu
University of Oviedo

RESUMEN DEL CONTENIDO DE TESIS DOCTORAL

1.- Título de la Tesis	
Español: Búsqueda de nueva física en eventos con leptones de alto momento transverso con el detector CMS en el LHC	Inglés: Search for new physics in events with high transverse momentum leptons with the CMS detector at the LHC

2.- Autor	
Nombre: Sergio Sánchez Cruz	DNI:
Programa de Doctorado: Materiales	
Órgano responsable: Comisión Académica Programa Doctorado en Materiales	

RESUMEN (en español)

En esta tesis se presentan medidas de procesos del SM y búsquedas de procesos BSM en sucesos con leptones de alto momento medidos en el detector CMS del LHC. Los estudios se realizan con datos correspondientes a colisiones protón-protón a $\sqrt{s}$=13 TeV obtenidos durante el Run 2 del LHC.

Se han realizado medidas de la sección eficaz de producción de un quark top asociado a un bosón W, obteniendo una sección eficaz inclusiva de σ = 63.1 ± 1.8 (stat) ± 6.3 (syst) ± 2.1 (lumi) pb, consistente con las predicciones del SM. También se han realizado medidas de la sección eficaz diferencial de este proceso, obteniendo resultados compatibles con el SM, dentro de las incertidumbres.

Se han llevado a cabo búsquedas de procesos BSM en sucesos con dos leptones de carga opuesta y mismo sabor y momento transverso faltante. Esta elección de estado final está motivada por procesos dados en modelos SUSY, para los cuales se han diseñado regiones de señal. No se observan excesos significativos a la producción esperada en el SM. Por ello, se establecen límites superiores a la producción de partículas supersimétricas.

Por último, se ha realizado una medida de la producción de bosones de Higgs asociados a uno o dos quarks top en el canal multileptónico. Este canal se enfoca en las desintegraciones del bosón de Higgs a WW, ZZ y ττ . Se han diseñado regiones de señal enriquecidas en señal y los distintos fondos basándose en las distintas propiedades cinemáticas de los procesos. Los resultados permiten observar la producción de ttH con una significancia observada (esperada) de 4.7 (5.2) desviaciones estándar. La medida también se interpreta en términos de variaciones de los acoplamientos del Higgs, constriñendo κ_t a estar en $-1.15 < \kappa_t < -0.8$ o $0.9 < \kappa_t < 1.25$ con un nivel de confianza del 95 %, asumiendo que el resto de los acoplamientos tengan el valor predicho por el SM.

Universidad de Oviedo
Universidá d'Uviéu
University of Oviedo

RESUMEN (en Inglés)

This thesis presents both measurements of SM processes and searches for BSM physics in events with high momentum leptons collected in the CMS detector of the LHC. These studies use proton-proton collision data at $\sqrt{s}$=13 TeV collected during the Run 2 of the LHC.

Measurements of the production cross section of a top quark associated with a W boson have been performed. An inclusive cross section of σ = 63.1 ± 1.8 (stat) ± 6.3 (syst) ± 2.1 (lumi) pb has been measured, consistent with the SM prediction. Differential measurements of this process have also been conducted, and found to be consistent, within uncertainties, with the SM.

Searches for BSM physics have been performed in events with two opposite-sign same-flavor leptons and large momentum imbalance. The choice of this final state is motivated by several SUSY signatures, for which specific signal regions are built. No significant excess of data above the SM prediction is observed in the signal regions. Therefore, upper limits to sparticle production in the context of simplified SUSY models are set.

Finally, a measurement of the production of single and pairs of top quarks in association with a Higgs boson is presented in the multilepton channel. This channel targets the decay of the Higgs boson into WW, ZZ and ττ pairs. Regions enriched in signal and in the various backgrounds are constructed by classifying events exploiting kinematic differences between the two species. Results allow to observe ttH production with an observed (expected) significance of 4.7 (5.2) standard deviations. The measurement is also interpreted in the context of Higgs coupling modifiers, constraining κ_t to be in either one of the intervals $-1.15 < \kappa_t < -0.8$ and $0.9 < \kappa_t < 1.25$ at 95% confidence level, assuming the rest of the couplings to be those of the SM.

SR. PRESIDENTE DE LA COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO EN MATERIALES

Agradecimientos

He tenido suerte de hacer mi tesis en el mejor grupo que podría esperar. Me gustaría agradecer a mi director de tesis, Javier, por su guía y liderazgo y a todos los miembros del grupo: Javi, Chiqui, Kike, Bárbara, Santi, Carlos, Víctor, Clara y Andrea. Todos ellos han contribuído a la tesis de alguna forma u otra. También me gustaría agradecer a Pietro, a Xuan y a Nacho su apoyo y consejo durante los primeros años (y también los finales) de mi tesis. Ha sido un verdadero placer.

Let me switch to English to thank the brilliant people from CERN I was lucky to work with tWards this ttHesis. Joining the ttH multilepton group has been one of the cornerstones of my thesis, and it came out with the best possible outcome. I would like to thank all its members and, in particular, those with which I have worked more closely: Marco Peruzzi, Cristina Martín, Josh Thomas-Wilsker, Binghuan Li and Xanda Carvalho. It has been an amazing experience, and I would be definitely up for another observation!

I joined the Edge/Z group being a summer student. I would like specially thank Pablo Martínez and Marc Dünser, among many things, for making me part of the group. I would also like to thank Celia Fernández, Marius Teroerde, Mario Masciovecchio and Balaji Venkat for all the discussion and work that have taken us to a successful Run 2 analysis.

I would also like to thank my partners in the top trigger group: Carmen Díez, Marino Missiroli and Charis Koraka.

Finally, this thesis would have never been possible all the members of the CMS muon group. In particular, I have been able to work very closely with George Karathanasis, Dmytro Kovalskyi, Devin Taylor, Alice Florent, Pedro Fernández and Alicia Calderón.

Last, but certainly not least, I would like to thank Marina for always being there. También quiero agradecer a mis padres y a Alonso, sin su constante apoyo esta tesis definitivamente no habría sido posible.

Contents

Introduction

The Standard Model (SM) of particle physics is the current scientific paradigm that has been able to provide successful predictions for all the phenomena observed in particle physics experiments up to date. Formulated during the 1970s, it predicts a variety of fundamental particles and interactions, that have been observed in collider experiments, and their properties have been accurately measured. The construction of the SM culminated in 2012 with the discovery of the Higgs boson by the CMS and ATLAS Collaborations [1, 2] at the CERN Large Hadron Collider (LHC).

The LHC is the largest and most powerful collider built up to date, and is able to accelerate protons up to almost the speed of light and collide them. These collisions are studied by several experiments measuring the particles that emerge from them.

Data collected during Run 1 of the LHC (2009-2013) allowed physicists to claim the observation of the Higgs boson and measure some of its couplings to other fundamental particles. Additional advances were made in measurements of the properties of the SM as, for instance, precision physics in the top quark sector. Run 1 also allowed to impose constraints on some of the most straightforward models of new physics.

With an increase in center-of-mass energy and in luminosity, that corresponds to 137.2 fb^{-1} recorded in CMS, the Run 2 of the LHC (2015-2018) allows to study processes with even lower production rate and with higher precision. Results obtained during Run 2 are further establishing the validity of the SM at a higher energy scale, providing evidence, as an example, for the interaction of the Higgs boson with top and bottom quarks, and the τ fermion. Severe constraints were also put on many Beyond Standard Model (BSM) physics scenarios, by excluding for instance the presence of top squark with masses up to 1 TeV in several models, bringing natural Supersymmetry (SUSY) into question.

The research described in this thesis explores processes with at least two high momentum, well isolated leptons in the final state, using data produced in proton-proton (pp) collisions recorded by the CMS detector. Lepton identification and triggering are essential in this final state. These aspects must be commissioned for each period of the data taking, and new techniques must be developed to obtain the ultimate performance of the detector. During this thesis, additionally to the measurements described, I have been involved in these tasks. As a result, three public documents have been released, showing the performance of triggers associated to the top physics group [3] and the muon reconstruction and selection [4, 5].

The final state with at least two leptons provides a clear signature of interesting processes in a hadronic collider, as leptons may be produced in the decay of a Higgs or electroweak gauge boson. This channel is therefore suitable to measure the properties of the Higgs boson and of the top quark, since the decay of the latter is mediated by the electroweak interaction. Higgs and electroweak gauge bosons may also be produced in the decays of SUSY particles produced by the strong or electroweak interaction. This thesis aims to exploit the physics potential of this final state, covering topics on the three mentioned fields.

The top quark is the heaviest particle in the SM, being the only quark that decays before hadronizing. Processes involving top quark allow therefore to test perturbative Quantum Chromodynamics (QCD) at the LHC energy scale. Additionally, because of its high mass, which is a consequence of its natural coupling to the Higgs boson, the top quark plays an important role in many BSM models. The main top quark production channel at the LHC is the production of quark pairs ($\mathrm{t\bar{t}}$). During my thesis I performed measurements in this channel at a center-of-mass energy, $\sqrt{s}$, of 13 TeV [6], 8 and 7 TeV [7]. Previously to the start of the thesis, I had also contributed to the measurement described in reference [8]. The main background in $\mathrm{t\bar{t}}$ measurements is the associated production of a top quark and a W boson, tW. During the thesis, I have performed inclusive and differential cross-section measurements of this process [9, 10].

BSM searches are performed by looking for SUSY in events with two opposite-sign same-flavor leptons and large momentum imbalance. This signature allows to search for the production of colored, electroweak and lepton spartners. I have performed these searches with the data collected in 2015 [11] and 2016 [12]. The results obtained in the latter measurement were combined with other final states in [13]. In this thesis, the continuation of these searches with the dataset collected in Run 2 is described, for which a publication is in the final state of preparation, being scrutinized by the CMS Collaboration.

The large amount of data collected during Run 2 of the LHC also allows to measure for the first time the coupling of the top quark to the Higgs boson at tree level. During this thesis, I have performed measurements of the $\mathrm{t\bar{t}H}$ production in the multilepton channel, which is the most sensitive to the process with the luminosity collected up to date. The measurement performed with data collected during 2016 provided evidence for this process in the multilepton channel only [14]. The statistical combination of this measurement with other decay modes allowed to observe this process for the first time [15]. All the results obtained are consistent with the SM expectation, including the one obtained when using data collected in 2017 [16]. The measurement reported in this thesis uses the complete Run 2 dataset, and allows to observe this process in the multilepton channel only, for which a publication is also close to being released.

During this thesis, measurements and searches have been performed in increasingly large datasets as they were collected by the CMS detector, allowing for the usage of more elaborate analysis techniques and more precise calibrations. This, together with the decrease of statistical uncertainties, allows for an overall higher precision. In this document, the studies performed with the complete dataset are described when available, and associated publications to them will follow. Results obtained from partial datasets have also been published, as indicated. The main published articles that result from this thesis, together with the main responsibilities I have in the CMS Collaboration, are described in the timeline displayed in figure 1.

The thesis is organized as follows. Chapter 1 provides a brief review of the theoretical state-of-the-art of particle physics, including a description of the SM and its possible extensions. In chapter 2 the LHC and the CMS detector are described, together with the techniques used for the analysis of the collisions. These two chapters do not contain original research by the author, but aim to contextualize the thesis and describe the techniques used. Chapter 3 describes the lepton reconstruction used, including studies for a precise characterization of their efficiency. Chapter 4 describes the tW measurements performed in this thesis. In chapter 5, the searches for supersymmetry

with two opposite-sign same-flavor leptons in the final state are described. Finally, chapter 6 covers the measurement of the $t\bar{t}H$ production in the multilepton channel. The thesis is closed with a summary and conclusions.

Some of the analyses described in this thesis have been performed in collaboration with other institutions in the CMS Collaboration. The personal contributions of the author are highlighted in every chapter.

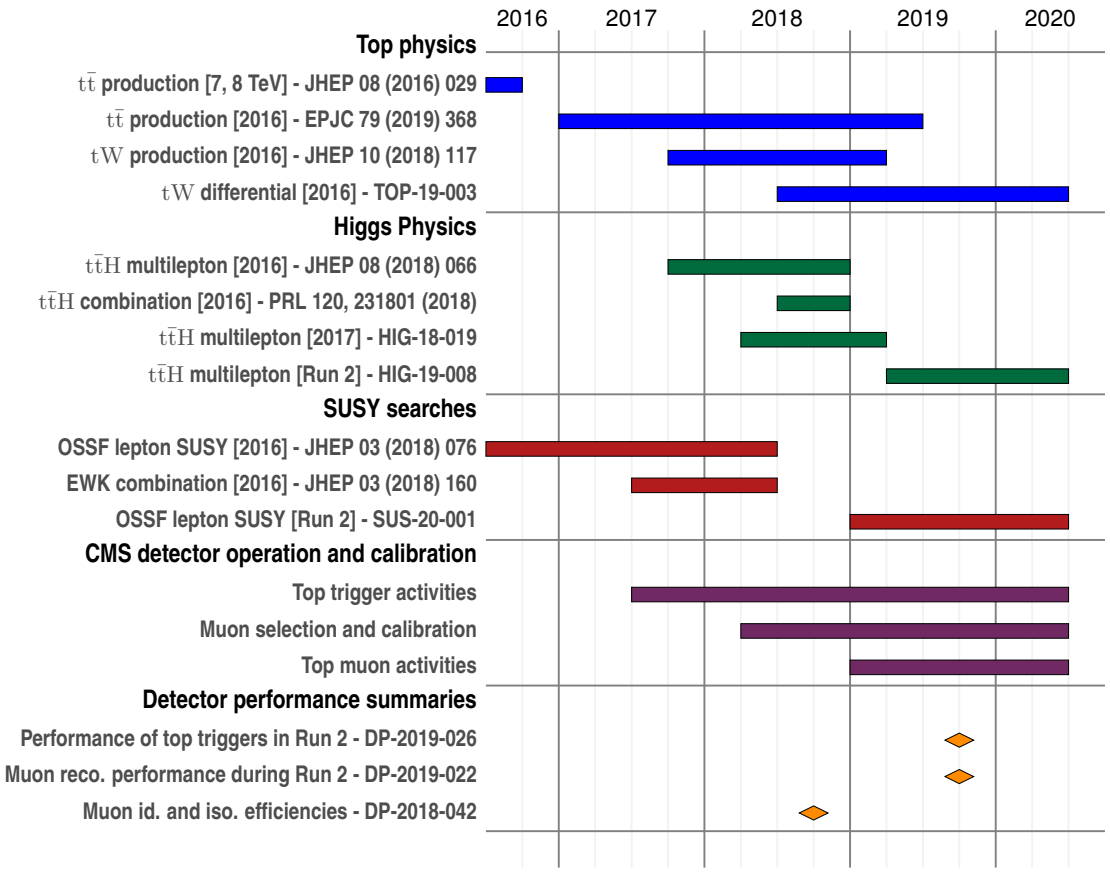


FIGURE 1: Timeline of the main tasks and analyses performed in the thesis. Time range for each article encompasses the time the analysis was started until when the associated article was published. Dataset associated to each publication are specified between brackets.

1

The Standard Model of particle physics and beyond

The universe as we know it is composed by elementary particles that interact among themselves following a set of natural rules. Our current understanding of these elementary particles is that they correspond to excited states of quantum fields, that are mathematically described with Quantum Field Theories (QFTs). The SM of particle physics is a renormalizable and gauge invariant quantum field theory that encompasses all the particles and their interactions observed up to date, providing accurate predictions to all the phenomena observed in particle physics, spanning over many orders of magnitude. Reviews on the current formulation of the SM can be found in reference [17]. As already mentioned, particles that conform matter and their interactions are represented by fields in this QFT.

Interactions are mediated by integer spin particles, called bosons, that arise when requiring that the theory is invariant until certain local gauge transformations. Gluons mediate the strong force, while the $Z^{\pm}$ and W bosons, and the photon mediate the electroweak interaction. Gravity is not described by the SM, but it is negligible at the experimentally accessible energy scales by particle physics experiments.

Matter particles have spin 1/2 and are referred to as fermions. Leptons and neutrinos do not interact with gluons, while quarks do. There are two types of quarks: up-type quarks, with electric charge of 2/3, and down-type quarks, with electric charge of -1/3. Three generations of leptons, neutrinos and quarks are present, having all generations identical quantum numbers, but different masses.

1.1 Matter content and symmetries of the SM

The propagation and interactions of fields in a QFT are fully determined by a Lagrangian density, $\mathcal{L}(\psi_i, \partial_\mu \psi_i, x^\mu)$, that depends on the fields ψ_i, their derivatives $\partial_\mu \psi_i$, and the spacetime coordinates, $x^\mu = (t, x, y, z)$.

Since this is a relativistic theory, fields must be representations of the Lorentz group. Fermion fields are represented by spinors, while interaction particles are represented by vectors. The Higgs boson, that is mentioned in following section, is represented by a scalar field.

The other ingredient of the SM are the local gauge symmetries. These symmetries, due to Noether's theorem, enforce the conservation of charges associated to the interactions. Imposing the theory to be invariant under these transformations requires the introduction of additional vector fields, that correspond to gauge bosons that mediate the interactions.

Two interactions are present in the SM: the Electroweak (EWK) interaction and the strong interaction. Each one of the two is represented by a different local gauge symmetry. QCD is built requiring the theory to be invariant under the 3-dimensional special unitary group $SU(3)_C$, while the theory of EWK interactions is invariant under $SU_L(2) \times U(1)_Y$

1.1.1 Quantum chromodynamics

QCD is the theory in the SM describing strong interactions. QCD is a gauge theory symmetric under the 3-dimensional special unitary group, $SU(3)_C$. The conserved charged associated to this symmetry is the color charge. Only quarks have color charge, therefore quark fields have a fundamental representation under this group, while the rest of the fermions have the trivial representation. The $SU(3)_C$ group is non-abelian and has 8 generators, T^a, the Gell-Mann matrices, that are related to the 8 gluons. The QCD Lagrangian is written as

$$\mathcal{L}_{QCD} = \bar{\psi}\left(i\gamma^\mu \partial_\mu - g_s \gamma^\mu T_a G^a_\mu - m\right)\psi - \frac{1}{4} G^a_{\mu\nu} G^{\mu\nu}_a, \tag{1.1}$$

with the strength field tensor $G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\mu G^a_\nu - g_s f^{abc} G^b_\mu G^c_\nu$, where f^{abc} are the structure constants of the SU(3) group. The a, b and c indices run over the 8 kinds of gluons. The third term in $G_{\mu\nu}$ must be introduced since the $SU(3)_C$ group is non-abelian, and adds self-interaction terms to the gluons, so they couple to each other. g_s is the coupling constant of the strong interaction, mostly referred to in this thesis as $\alpha_s = \frac{g_s^2}{4\pi}$.

The structure of the $SU(3)_C$ group conditions profoundly affects its phenomenology. At the level of one loop, the renormalization group equations predict an scaling of α_s as

$$\frac{d\alpha_s}{dQ^2} = -\frac{\alpha_s^2}{12\pi Q^2}(33 - 2n_f), \tag{1.2}$$

where n_f is the number of accesible flavors and Q the scale of the interaction. For $n_f = 3$ and below, α_s always decreases as a function of Q^2. This means that the strong interaction will be stronger at low energy scales, with the consequence that colored particles will form bounded colorless states, named hadrons. This phenomenon is called quantum confinement, and implies quarks or gluons will not be observed as free particles. The exception to this is the top quark which, due to its high mass, decays before hadronizing. At higher energies the opposite phenomenon occurs, asymptotic freedom, as the strength of the strong interaction is increasingly weaker.

1.1.2 Electroweak interaction

The description of the electroweak interaction in the SM was introduced by Glashow, Weinberg and Salam [18–20]. Since left and right-handed particles couple differently through the EWK interaction, the projection of matter fields to their respective chiral components is considered.

$$\psi_L = \frac{1}{2}\left(1-\gamma^5\right)\psi, \quad \psi_R = \frac{1}{2}\left(1+\gamma^5\right)\psi. \tag{1.3}$$

The electroweak theory is introduced by imposing a local gauge invariance of a $SU_L(2) \times U(1)_Y$ group. The first component of the group acts only on the left-handed components on fermions, hence they are represented as doublets, while the right-handed components are singlets:

$$L = \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \;\; u_R, \;\; d_R. \tag{1.4}$$

In the expression above only the representation for up and down quarks is written for simplicity, but a similar structure is followed for other quarks and leptons, with the exception of neutrinos, that do not have a right-handed component. This gauge group introduces three bosons W^i, associated to the generators of the $SU(2)_L$ group, which are Pauli matrices, σ_i. The associated conserved charge to the $SU(2)_L$ group is called weak isospin. For up-type left-handed fermions, this quantity is $+\frac{1}{2}$ and for down-type it is $-\frac{1}{2}$. For right-handed fermions, since they are singlets of $SU(2)_L$, their weak isospin component is zero.

Similarly to the strong interaction, the $SU(2)_L$ invariant Lagrangian can be written as

$$\mathcal{L} = -\frac{1}{4}W^i_{\mu\nu}W_i^{\mu\nu} + i\bar{u_R}\gamma^\mu\partial_\mu u_R + i\bar{d_R}\gamma^\mu\partial_\mu d_R + i\bar{L}\gamma^\mu D_\mu L, \tag{1.5}$$

where $W^i_{\mu\nu} = \partial_\mu W^i_\nu - \partial_\mu W^i_\nu - g_W\epsilon^{ijk}W^j_\mu W^k_\nu$ and $D_\mu = \partial_\mu + ig_W\frac{\sigma_i}{2}W^i_\mu$.

It should be noted that by imposing the invariance under $SU(2)_L$, this theory does not allow to contain mass terms in the Lagrangian, as $m\bar{\psi}\psi = m(\bar{\psi_R}\psi_L + \bar{\psi_L}\psi_R)$ is not invariant under $SU(2)_L$.

Inspired by the electromagnetic theory, and to explain the different couplings of the W and Z bosons, an additional $U(1)_Y$ group is introduced, with Lagrangian density

$$\mathcal{L} = \sum_{\psi=u_R,d_R,L}\bar{\psi}\left(i\gamma_\mu\partial^\mu - Yg_YB_\mu\right)\psi - \frac{1}{4}B^{\mu\nu}B_{\mu\nu}. \tag{1.6}$$

The conserved current for this symmetry is the hypercharge, Y, that is related with the third component of the isospin and the electric charge: $Y = 2(Q - I_3)$. The four degrees of freedom introduced by $SU(2)_L \times U(1)_Y$ local gauge symmetry can be related with the physical mass eigenstates corresponding to the W and Z bosons and the photons:

$$\mathrm{W}^{\pm} = \frac{1}{\sqrt{2}}\left(W^1 \mp iW^2\right) \tag{1.7}$$

$$\begin{pmatrix} \mathrm{Z} \\ \gamma \end{pmatrix} = \begin{pmatrix} \cos\theta_{\mathrm{W}} & -\sin\theta_{\mathrm{W}} \\ \sin\theta_{\mathrm{W}} & \cos\theta_{\mathrm{W}} \end{pmatrix} \begin{pmatrix} W^0 \\ B \end{pmatrix}, \tag{1.8}$$

where θ_{W} is the Weinberg angle, that depends only on g_W and g_Y.

1.1.3 Spontaneous symmetry breaking of the electroweak symmetry

As mentioned in the previous section, imposing the invariance under local gauge transformations of the $SU(3)_C \times SU(2)_Y \times U(1)_Y$ group does not allow for a theory with massive fermions. Similarly, massive bosons would also break this symmetry. This obviously contradicts experimental evidence in which some bosons and all fermions are massive.

Masses can be introduced in the theory by means of the Higgs-Brout-Englert mechanism [21, 22], that describes the Spontaneous Symmetry Breaking (SSB) of the EWK symmetry. An additional $SU(2)_L$ doublet of complex scalar fields, H, is introduced in the theory. These field propagate under a given potential

$$\mathcal{L}_{\mathrm{H}} = (D_\mu H)^\dagger (D^\mu H) + \mu^2 \phi^\dagger \phi + \lambda(\phi^\dagger \phi)^2. \tag{1.9}$$

A Yukawa interaction term is added between the fermions and the scalar doublet:

$$\mathcal{L}_{\text{Yukawa}} = -y_u \left(\bar{L} H u_R + \bar{u_R} H L\right) - y_d \left(\bar{L} H d_R + \bar{d_R} H L\right). \tag{1.10}$$

It should be taken into account that for leptons only the first term is present, since there are no right-handed neutrinos in the SM. Once these terms are added, the theory is still invariant under the local $SU(2)_L \times U(1)_Y$ symmetry. However, assuming $\mu^2 < 0$, the potential has a set of minima, and the H may acquire a vacuum expectation value, breaking the symmetry. The unitary gauge $H = \begin{pmatrix} 0 \\ \sqrt{\frac{-\mu^2}{\sqrt{2}\lambda}} + \mathrm{H}(x) \end{pmatrix}$ can assumed, where $\mathrm{H}(x)$ is the remaining degree of freedom after the symmetry breaking. This degree of freedom corresponds to a scalar, the Higgs boson. With this, terms appear that give rise to masses to the Z, W and H bosons, that take the following values:

$$\begin{aligned} m_{\mathrm{W}} &= \frac{gv}{2}, \\ m_{\mathrm{Z}} &= \frac{v\sqrt{g_{\mathrm{W}}^2 + g_{\mathrm{Y}}^2}}{2}, \\ m_{\mathrm{H}} &= \sqrt{2\lambda} v. \end{aligned} \tag{1.11}$$

The Yukawa couplings also introduce mass terms for the fermions, that are equal to $m_\psi = \frac{y_\psi v}{\sqrt{2}}$.

1.1.4 CKM matrix

While the SM contains three flavors of fermions, the construction considered so far only takes into account only one generation. This construction can be expanded by considering a family of quarks Q_{Li}, u_{Ri} and d_{Ri}, where i is an index running over the 3 flavors. The Yukawa interaction for quarks may then be written as

$$\mathcal{L}_{\text{Yukawa}} = -Y^d_{ij}\bar{Q}_{Li}Hd_{Rj} - Y^u_{ij}\bar{Q}_{Li}\epsilon H^* u_{Rj}, \tag{1.12}$$

where the Yukawa couplings, $Y^{u,d}$, become now 3×3 complex matrices, and ϵ is the 2×2 antisymmetric tensor. The indices i and j run over the three flavors. When the Higgs field acquires a vacuum expectation value, fermions still acquire mass terms. However, since the $Y^{u,d}$ matrices are not diagonal, the mass eigenstates do not correspond anymore to the gauge eigenstates. The former can be recovered by diagonalizing the $Y^{u,d}$ matrices. As a result, the interaction of the W bosons does not occur among fermions with the same flavor. Instead, couplings between different generations exist, proportionally to the elements of the Cabibbo Kobayashi Maskawa (CKM) matrix.

$$V_{CKM} = \begin{pmatrix} V_{\text{ud}} & V_{\text{us}} & V_{\text{ub}} \\ V_{\text{cd}} & V_{\text{cs}} & V_{\text{cb}} \\ V_{\text{td}} & V_{\text{ts}} & V_{\text{tb}} \end{pmatrix}. \tag{1.13}$$

The elements of the CKM matrix are obtained in the diagonalization of the $Y^{u,d}$ matrices, and are free parameters of the theory, that are to be measured experimentally. This matrix must be unitary by construction and, therefore, its elements are not independent. Measurements of these elements, together with the 6 resulting constrains are used to experimentally test the consistency of the SM.

As a conclusion, all the matter particles, with their corresponding quantum numbers are measured masses are shown in table 1.1. The mass eigenstates of the gauge bosons are the following. The photon and the gluons are massless, while the $\text{W}^\pm$ and the Z bosons have been measured with a mass of 80.385 ± 0.015 GeV and 91.1876 ± 0.0021 GeV respectively [17].

1.2 Open topics in the SM

The theory described in the previous section provides a complete description of all the phenomena observed so far in particle physics. The particles described in the previous section have been experimentally observed and measured. The construction of the SM was completed when the Higgs boson was observed by ATLAS and CMS. Despite of this striking excess, there are a few open topics for which the SM does not

	Q	I_3	Y	1st gen. mass	2nd gen. mass	3rd gen. mass
$(e, \mu, \tau)_L$	-1	-1/2	-1	511.0 keV	105.66 MeV	1.777 GeV
$(e, \mu, \tau)_R$	-1	0	-2			
$(\nu_e, \nu_\mu, \nu_\tau)_L$	0	+1/2	-1	-	-	-
$(u, c, t)_L$	+2/3	+1/2	+1/3	2.2 MeV	1.27 GeV	173 GeV
$(u, c, t)_R$	+2/3	0	+4/3			
$(d, s, b)_L$	-1/3	-1/2	+1/3	4.7 MeV	96 MeV	4.18 GeV
$(d, s, b)_R$	-1/3	0	-2/3			

TABLE 1.1: Quantum numbers of the matter particles in the SM. The three columns in the rate indicate the masses of the first, second and third generation particle, respectively, that have been taken from [17]. Uncertainties on the measured masses have been omitted.

have a complete explanation yet. In this section, some of these points are discussed. Additionally, the supersymmetric extensions of the SM are described.

Gravity The SM does not include an unified description of gravity as a QFT. However the contribution of gravity to processes at the accessible energy scales is negligible. Effects due to gravity are only assured to be relevant above the Planck scale, 15 orders of magnitude higher than the TeV scale. Attempts exist however to introduce gravity into the SM by means of a spin-2 mediator, the graviton.

Naturalness The quantum corrections to the Higgs boson mass serve to illustrate the hierarchy problem of the SM. Loop contributions from fermions and bosons to the Higgs propagator, like in the ones depicted in the left term of figure 1.1, can be significant. These radiative corrections are dominated by the top quark, which is the particle with the highest mass and therefore the largest coupling to the Higgs boson. This implies corrections to the mass that are of the order of

$$\Delta m_H^2 \propto \frac{y_t^2}{8\pi^2}\Lambda_{UV}^2 \tag{1.14}$$

where Λ_{UV} is a ultraviolet cut-off used to regulate the loop integral. Under the SM, the scale of Λ_{UV} can be as large as the Planck scale. Therefore, in absence from other BSM contributions besides gravity, these corrections can be large. Other particles can contribute to the quantum corrections of the Higgs boson mass and compensate those of the top quarks. Indeed, corrections from bosons contribute with the opposite sign to that of fermions, however their masses and couplings are different, therefore a complete cancellation is highly unexpected.

Neutrino masses In the SM formulation shown in the previous section, neutrinos do not acquire mass in the SSB mechanism. However, observations of neutrino mixings confirm they are indeed massive particles [23]. Indeed their masses are currently unknown, and also the mechanism through which they acquire them.

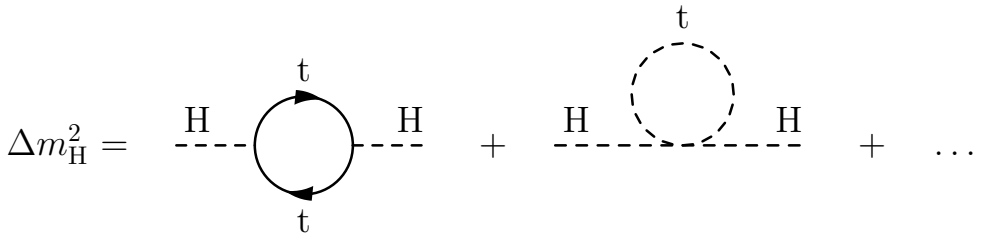


FIGURE 1.1: Corrections to the Higgs propagator induced by top quark loops (left term) and top squark loops (right term).

Dark matter The rotational curves of the galaxies cannot be explained only by the presence of observed baryionic dark matter. This hints to the presence of an known type of matter not included in the current formulation of the SM, and whose nature is unknown. Supporting evidence comes from observations of the cosmic microwave backgrounds [24].

1.2.1 Supersymmetry

Supersymmetric theories are introduced to solve some of the open issues in the SM, like the hierarchy problem and the nature of the dark matter constituents. These theories propose the existence of a new symmetry, SUSY, that relates fermions and bosons:

$$Q\,|\text{fermion}\rangle = |\text{boson}\rangle\,, \quad Q\,|\text{boson}\rangle = |\text{fermion}\rangle\,, \tag{1.15}$$

where Q is the generator this symmetry. In supersymmetric theories, particles are represented as multiplets of the associated algebra, named supermultiplets. Each supermultiplet contains the same amount of fermion and boson degrees of freedom, and therefore each particle has associated a superpartner. If the particle is a fermion, the superpartner must be a boson and viceversa.

Because of this property, supersymmetric theories predict a set of new particles that are not included in the SM. These particles have the same quantum numbers as their SM partners, but for their spin. Gauge bosons are associated to spin 1/2 particles, named gauginos and, conversely, fermions are associated to scalar particles, named sfermions. These new particles are usually referred to as sparticles. The Minimal Supersymmetric Standard Model (MSSM) includes only the needed particles to have a consistent SUSY theory containing the current particle content observed in the SM. Besides the spartners of the SM particles, an additional Higgs doublet must be added. One of the Higgs doublets will give masses to the down-type quarks, while the other will give mass to the up-type quark and the charged leptons. In total, three neutral and two charged Higgs bosons are present in the theory, together with the corresponding sparticles.

The nomenclature followed to denote sparticles is the following. Sparticles are usually denoted with the name of their SM partner with a tilde. Like that, the superpartner of the right-handed electron, $\mathrm{e_R}$, is denoted as $\widetilde{\mathrm{e}}_\mathrm{R}$. In many supersymmetric theories the spartners of the electroweak gauge bosons and the Higgs bosons do not correspond to the mass eigenstates, but to a mixture of them. The neutral mass eigenstates are named charginos, $\widetilde{\chi}_1^0$ and $\widetilde{\chi}_2^0$, while the neutral are named neutralinos, $\widetilde{\chi}_1^0$, $\widetilde{\chi}_2^0$, $\widetilde{\chi}_3^0$ and $\widetilde{\chi}_4^0$.

As anticipated, the introduction of these new particles compensates the quantum corrections to the Higgs boson propagation described in the previous section, as shown in the diagrams of figure 1.1. If SUSY was an unbroken symmetry, and since the quantum numbers of sparticles are the same as SM particles, this cancellation would be exact, since they would have the same masses as their SM partners.

However, such particles have not been observed. Therefore, if SUSY is realized in nature it must be broken at a certain scale Λ_{UV}. Then, the corrections to the Higgs boson mass are not canceled completely, however the Λ_{UV} scale is at the SUSY breaking scale and not at the Planck mass. The size of the corrections to the Higgs propagator would be acceptable if this scale would only be slightly above the electroweak scale. Because of this, it is very appealing to perform searches for SUSY particles in the current experiments at the LHC.

The phenomenology of a supersymmetric theory is conditioned by the masses of the associated partners, which are set by the mechanism through which the symmetry is broken. Several mechanisms have been proposed for this breaking, but in practice not that many constraints can be derived from this.

Additionally, the presence of SUSY particles introduces new interactions. In particular, the decay of the proton could be mediated by the superpartner of the right-handed strange or bottom quarks. This again contradicts the observations on the lifetime of the proton. This is typically handled by requiring the conservation of the *R*-parity, a quantum number that is defined as +1 for SM particles, and -1 for the SUSY particles. Besides protecting the proton from promptly decaying, this has phenomenological consequences. First, since this number must be conserved, sparticles can only be produced in pairs in colliders, since the initial state particles are SM particles. Secondly, R-parity conserving models predict the existence of a stable SUSY particle, since such its decay into sparticles would be kinematically forbidden and its decay into SM particles would violate *R*-parity conservation. This particle, known as Lightest Supersymmetric Particle (LSP), would be a suitable candidate to be the constituent of dark matter. In many supersymmetric models, the phenomenology of SUSY processes is set by the characteristics of the LSP and the Next-to-lightest Supersymmetric Particle (NLSP). Indeed not all the realistic SUSY models assume *R*-parity conservation, and in some of them a certain degree of violation is allowed.

Current searches for SUSY at the LHC make use of simplified models, in which the production of a certain kind of SUSY particles and an specific decay mode with 100% branching fraction are assumed. These models may not be realistic, but are nevertheless useful to define a framework to guide searches and set limit on the production of sparticles. Up to date, sparticles have not been detected. Instead, tight constrains are put by the LHC experiments on the production of sparticles. For instance, in the most straightforward models, stop masses below the TeV scale are excluded, questioning the existence of natural SUSY at the electroweak scale.

1.3 SM phenomenology at LHC

Many of the most important tests that physicists set the SM are collider experiments. Colliders are also used aiming to produce BSM particles, not accessible at the lower energy scales. Because of this and the because this thesis is focused in one of the LHC experiments, the rudiments used to draw testable predictions from the SM in hadronic colliders are described in this section. These techniques can also be used to produce predictions under the assumptions of BSM theories, provided they can be computed in perturbation theory.

1.3.1 Parton distribution functions

Any observable in a hadronic collider can be computed from a inclusive or differential cross section. Calculations of these cross sections are usually divided in two parts: the short distance interactions, that can be computed analytically using perturbation theory, and long distance QCD interactions. Perturbation theory cannot be used for the latter since the strong coupling becomes too large, and a dedicated treatment for them is needed. Cross-sections for the production of a given final state in a proton-proton (pp) collision can be computed at fixed order in perturbation theory as

$$\sigma = \sum_{i,j} \int dx_i dx_j f_i(x_j, \mu_F) f_j(x_j, \mu_F) \sigma_{ij}(x_1 x_2 s; \mu_R, \mu_F). \tag{1.16}$$

The first sum runs over the possible partons species in the two protons, for which all the momentum configurations are considered, parametrized by the Bjorken variables, $x_i = p_i/p$, where p_i is the momentum of the parton and p, that of the proton.

σ_{ij} is the partonic cross-section, that can be computed at a fixed order, using the Feynman rules that can be obtained from the Lagrangian densities described in the previous sections. The renormalization scale, μ_R, must be introduced in this calculation in order to cure the divergences in the loops of these Feynman diagrams. This scale enters through the running of the coupling constants, which become a function of it.

$f_i(x)$ are the Parton distribution functions (PDFs), that represent the probability for finding a parton in the proton with a given momentum fraction x. By adding this contribution, another scale, μ_F, is introduced that regulates the separation between the soft and hard interaction. These functions parametrize the soft interactions that occur in the initial state protons and therefore cannot be calculated from first principles with the current formulation of QCD. Instead, these functions have been measured in deep inelastic scattering and other experiments at certain scales. PDFs can be then propagated to the scale under study using the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations.

Fixed order calculations can give accurate predictions for the partonic final state, however this does not correspond to the final experimental observables. These particles may instead further decay or, if they are colored particles, they can radiate extra partons. Since this radiation may be soft, and also because the matrix elements can only be calculated up to a certain order, phenomenological Parton Shower (PS) models are used instead.

Besides parton showers, other effects must also be taken into account. In particular, the rest of the proton constituents may interact among each other, leading to secondary interactions. These interactions are known as Underlying Event (UE), and modeled using phenomenological models that take into account perturbative and non-perturbative effects.

1.3.2 Flavor schemes

When modeling the production involving third generation quarks, such as single top cross-section, two different approaches can be followed to treat initial state b quarks. The two approaches are called 4-flavor scheme (4FS) and 5-flavor scheme (5FS). In the 5FS the b quark is assumed to be a massless particle and therefore it is considered as part of the proton PDF. In the 4FS, the mass of the quark is considered and therefore, it can only be present in the initial state through gluon splitting, since its mass is larger than that of the proton. This nomenclature of "massive" and "massless" particle is only for convenience and does not imply that bottom quark mass effects are taken into account [25].

The two approaches are valid to make sensible predictions. In the 4FS the b quark emission is simulated exactly at fixed order in perturbation theory, while in the 5FS its emission comes from the parton shower so, in principle, the former should be more accurate. However, the gluon splitting in the former approach suffers from a collinear divergence, that may spoil the convergence of the perturbation series. This divergence is on the other hand absorbed by the PDFs in the 5FS.

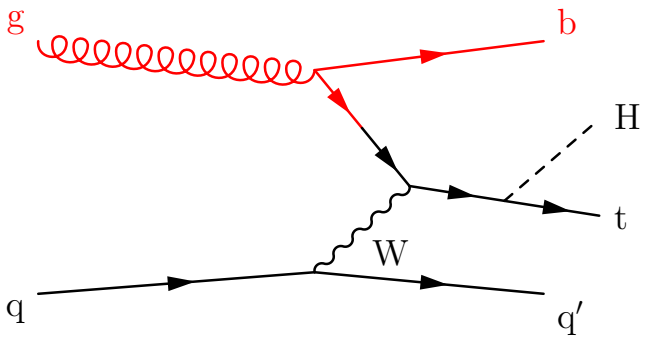


FIGURE 1.2: Feynman diagram for tH production in the 4FS and 5FS. The red lines are calculated at fixed order in the 4FS but are absorbed into the PDF in the 5FS.

1.3.3 Hadronization

Parton showers still only predict the radiation of final state partons. These partons are not the final physical observables as they still correspond to colored states. Colored particles hadronized to form bounded color neutral states, named hadrons. This corresponds to a fully non-perturbative process, that occurs below a given energy scale. Phenomenological models, such as PYTHIA [26] and HERWIG [27], that also model PS, exist for the treatment of these processes.

1.3.4 Monte Carlo generators

In practice, some of the calculations mentioned above are automated in the form of Monte Carlo (MC) event generators. These generators have the advantage that they allow to easily perform phase space integrations and, additionally, they allow to deliver predictions for arbitrary observables.

In this thesis, several analyses with different choices of MC models are described. Typically, the hard scattering is simulated at Leading Order (LO) or Next-to-leading Order (NLO) accuracy using the POWHEG [28], MADGRAPH or MADGRAPH MC@NLO [29] with a given choice of PDF. These calculations are then interfaced to PS generators. In particular, PYTHIA 8 is always used in this thesis. This generator also models the hadronization and the UE, that are tuned to reproduce quantities measured in data.

2

The LHC and the CMS experiment

The measurements and searches described in this thesis have been performed in the Compact Muon Solenoid (CMS) detector of the CERN LHC. The LHC facility and the CMS experiment are described in sections 2.1 and 2.2. Section 2.3 describes the reconstruction of events performed in the data collected by the CMS detectors. Section 2.4 describes the features of the datasets used in this thesis, and section 2.5 describes the corrections that are applied to simulations in order for them to reliably predict the features observed in data. Finally, section 2.6 is dedicated to the analysis techniques employed along this thesis.

2.1 The Large Hadron Collider

The LHC is the largest physics experimental facility. It is a collider that accelerates beams of hadrons up to a center-of-mass energy of up to $\sqrt{s} = 13$ TeV and collides them. These collisions take place in four interaction points where the main LHC experiments, A Toroidal LHC ApparatuS (ATLAS), CMS, Large Hadron Collider beauty (LHCb) and A Large Ion Collider Experiment (ALICE), are located. These experiments record and analyze these collisions, extracting information of many physical observables. Here only a brief description of the LHC and functioning is provided. A more detailed description can be found in reference [30].

The LHC profits from an accelerator chain that starts when protons, produced from a ionized hydrogen gas, are injected into a Radio Frequency Quadrupole, a cavity that accelerates and groups the protons in beams. Protons are then injected in the LINAC 2, a linear accelerator that accelerates them up to 50 MeV. Then they are injected into the Proton Synchrotron Booster, that accelerates them up to 1.4 GeV, and in which protons are grouped into bunches. Protons are then accelerated to an energy of 26 GeV in the Proton Synchrotron, where the bunch structure is set. Finally, the Super Proton Synchrotron accelerates protons until 450 GeV, energy at which protons are injected into the LHC.

The LHC is a circular collider situated along a 27 km underground tunnel near Geneva (Switzerland). Protons injected into the LHC are increasingly accelerated by a set of superconducting radio frequency cavities situated in one point of the LHC circumference. There are 16 cavities operating at a frequency of 400 MHz, each one of which provides a 2 MV acceleration voltage. The oscillating voltage is synchronized with the arrival time of the protons in a way that protons arriving earlier are accelerated less than the ones arriving later, collimating further the bunches.

Protons are kept in the circumference by bending their trajectory using a set of superconducting Nb-Ti magnets that generate a magnetic field of up to 8.3 T, that are situated in many sectors along the circumference. These magnets are operated at a temperature of 1.9 K, that is achieved with a circuit of liquid helium. Between the bending magnets, sets of quadrupole magnets focus the beam in the directions perpendicular to the beam axis, in order to avoid beam losses. Additional magnets further focus the beam and direct it to the collision points.

With this combination of radio frequency cavities and superconducting magnets, the LHC is able to accelerate and collide protons up to energies of $\sqrt{s} = 13$ TeV, the largest ones achieved in a human-made experiment. The other figure of merit relevant for experimental physicists is the instantaneous luminosity, $\mathcal{L}$. The instantaneous luminosity can be defined as the relation between the event rate and the cross-section of a process with a given cross-section σ:

$$\frac{dN}{dt} = \mathcal{L}\sigma. \tag{2.1}$$

This quantity depends only on the characteristics of the collider, and not on the process considered. It can be approximately computed as

$$\mathcal{L} = \frac{n_b N_p^2 f}{4\pi\sigma_x\sigma_y} R, \tag{2.2}$$

where N_p is the number of protons per bunch, n_b the number of colliding bunches, f is the beam revolution frequency, $\sigma_{x,y}$ are the beams sizes along the transverse directions, and R is a geometrical reduction factor to take into account the crossing angle at the interaction point.

A quantity associated to the instantaneous luminosity is the integrated luminosity, L, usually referred to in this thesis as luminosity, which is defined as the integral over time of the instantaneous luminosity

$$L = \int \mathcal{L} dt. \tag{2.3}$$

The luminosity integrated along the data taking periods of 2015, 2016 and 2017 by the CMS experiment is shown in figure 2.1.

The LHC was designed to deliver an instantaneous luminosity of about 10^{34} cm^{-2}s^{-1} during Run 2, but twice this value was achieve during the running period in 2017 and 2018. To achieve this high luminosity, many protons are collimated in bunches. Because of this, simultaneous collisions occurs at every bunch crossing. This phenomenon is called pile-up, and can be a limiting factor for the LHC experiments. Figure 2.1 shows distribution of the average number of interactions per bunch crossing. The average number of pile-up interactions was 27, 38 and 37 in the operation during 2016, 2017 and 2018, respectively, the data used in this thesis.

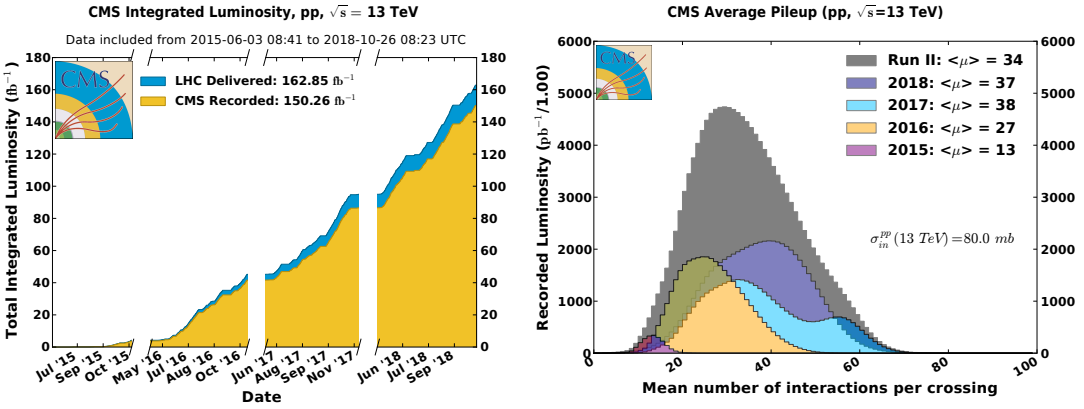


FIGURE 2.1: Left: integrated luminosity delivered and recorded by the CMS experiments during the Run 2 of the LHC. Data not recorded corresponds to periods in which the detector is not active. Right: Distribution of the number of interactions per bunch crossing in CMS during that period. Taken from [31].

2.2 The CMS experiment

The CMS experiment is, together with ATLAS, one of the two general-purpose detectors of the LHC. These detectors aim to record and study a wide variety of physical processes that may occur in pp at the LHC energy scale. In contrast, LHCb and ALICE are dedicated to the study of b quark physics and heavy ion collisions, respectively. In this section a brief description of the CMS detector, the experimental set-up of this thesis, is provided. A more complete description can be found in [32].

The CMS detector is a cylindrical device located around the LHC beam pipe in one of the interaction points. This detector has a length of 21.6 m, a diameter of 14.6 and weights 14500 tons. It is divided in several components: it features a superconducting solenoid magnet that generates a magnetic field of up to 3.8 T, a very granular tracking detector, an excellent muon detector and hermetic calorimeters. A general sketch of the CMS detector is shown in figure 2.2.

2.2.1 Coordinate system and notation

In order to describe the detector components, and also to quantify the physics observables in the measurements in this thesis a common coordinate system is established. The z axis is taken to point in the direction of the beam axis, pointing towards west. The y axis points in the vertical direction towards the earth surface, while the x axis points to the center of the LHC circumference. It is useful to define the $x-y$ plane as the transverse plane, which is perpendicular to the beam axis. The azimuthal angle, labeled as ϕ, is defined starting from the y axis. The projection of a particle momentum into the transverse plane is denoted transverse momentum, p_{T}. The polar angle in the $y-z$ plane, θ, is also defined starting in the y axis. This quantity is related to the pseudorapidity, η, defined as

$$\eta = -\log\tan\left(\frac{\theta}{2}\right). \tag{2.4}$$

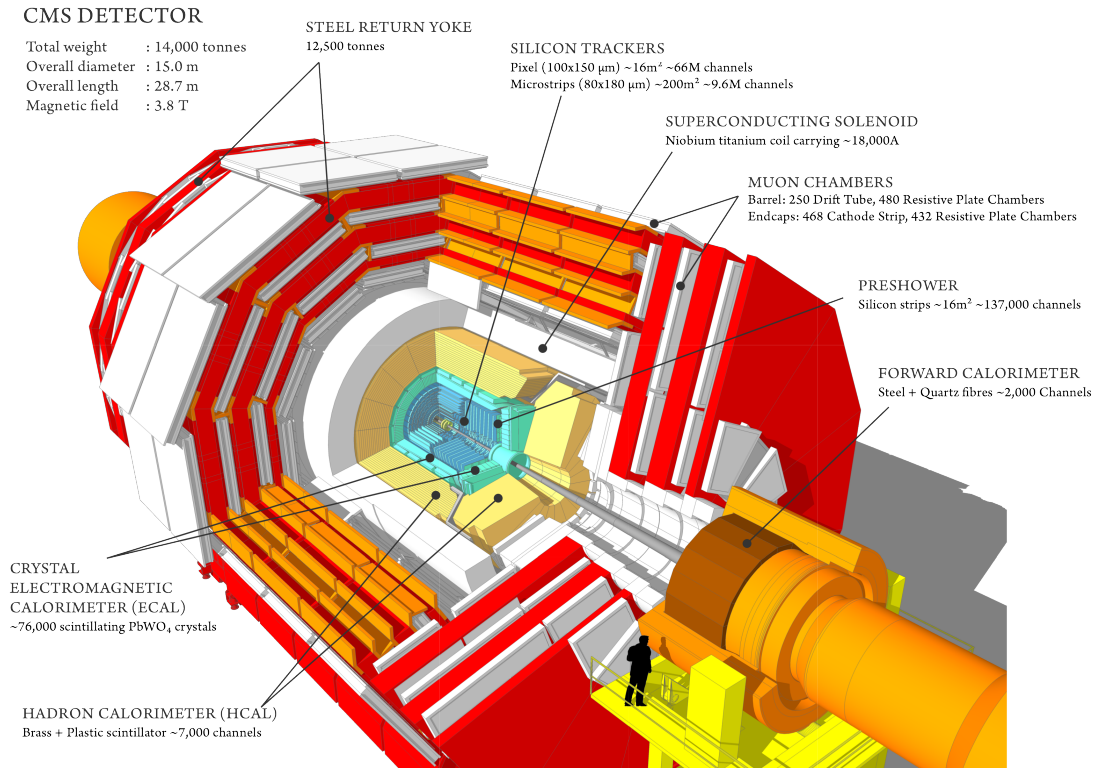


FIGURE 2.2: Layout of the CMS detector. Its main features and components are indicated. Taken from [33].

The difference in pseudorapidity between two massless particles is invariant under boosts along the z axis. Finally, angular distances are usually measured in units of as $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$.

Subdetectors in CMS are usually divided in two parts. The central part, surrounding the beam pipe is called the barrel. Two endcaps in the negative and positive z axis are located at the two sides of the barrel to increase the acceptance.

Finally, in this thesis, mathematical symbols in bold face are used to denote vectors.

2.2.2 Solenoid magnet

The solenoid magnet of the CMS detector is one of its key components. It is a superconducting Nb-Ti magnet, that generates a 3.8 T magnetic inside the solenoid and up to 2 T outside. Its is 13 m long and a has radius of 6 m, and is located in the center of the CMS detector, between the calorimeters and the muon system. The return yoke of the magnet is made of iron and interspersed with the chambers of the muon system. The magnetic field generated bends the trajectory of charged particles emerging from the collision, allowing to infer the transverse momentum of these particles from the curvature of their trajectory. This bending is larger the larger the magnetic field, so it is crucial to have an intense magnetic field for a precise momentum measurements, particularly for high momentum particles, for which the bending is smaller.

2.2.3 Tracking system

The tracking system allows to measure the track of charged particles produced in the collisions. It is located in the innermost part of the system, in order to make a precise measurement of the momentum and the impact parameter of the tracks. This allows to efficiently discriminate among the vertices corresponding to the several interactions that may occur in the same bunch crossing. A precise determination of the impact parameter is also crucial to tag b jets, which is an important observable in SM measurements and searches for new physics.

The tracking system is composed by two parts, the pixel detector and the silicon strip detector. Both are based on semiconductor technology but satisfy different granularity and size criteria.

The pixel detector is the innermost detector, situated a few centimeters away from the interaction point. Since a higher occupancy is expected in that region and since a high spacial resolution is needed closer to the interaction point for a precise measurement of the impact parameter, the pixel detector is very granular, with a spacial resolution of between 10 and 15 μm. The pixel detector is divided in 3 (4) layers in the barrel and 2 (3) disks in the endcap before 2017 (from 2017).

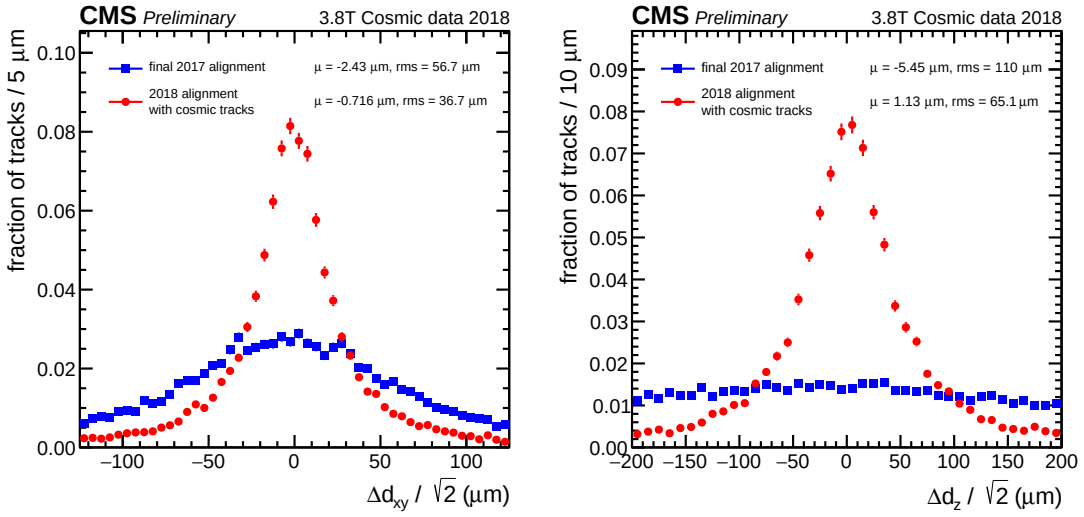


FIGURE 2.3: Impact parameter resolution in the transverse plane (left) and along the beam axis (right) measured in cosmic muons by comparing the tracks in the upper and lower part of the detector. Taken from [34].

The silicon strip system is the second layer of subdetector located 20 to 116 cm away from the interaction point. Silicon strips are located along the z direction in the barrel, and along the radial component in the endcap.

The performance of the CMS tracking system is shown in figure 2.3, in which the impact parameter resolution in the transverse and z components is measured.

2.2.4 Calorimeters

The aim of the calorimeters is to measure the energy of the particles produced in the collisions. Two types of calorimeters are present in the CMS detector: the Electromagnetic Calorimeter (ECAL) and Hadronic Calorimeter (HCAL). The two calorimeters are located between the tracking system and the solenoid magnet, with the ECAL being the innermost one. High resolution in the energy measurement and hermeticity are necessary to precisely measure the energy of neutral particles, as well as provide a good resolution in the total momentum imbalance of the event.

The ECAL is designed for a precise measurement of the energy of electrons and photons. It is a scintillation calorimeter made of $PbWO_4$ crystals. This material is suitable for this purpose due to its high density, short radiation length and small Moliere radius. The light produced by the scintillation is recorded by photodetectors, Avalanche Photodiodes (APDs), whose output can be read out. The ECAL energy resolution is shown in figure 2.4, as measured in $Z \rightarrow ee$ events.

The HCAL surrounds the ECAL and is designed to measure the energy of hadrons. It is the only element that allows to measure the energy of neutral hadrons and it provides a complementary measurement of the charged hadron measurements. Since the space between the ECAL and magnet is limited, part of the it is located outside the magnet solenoid. It is a sampling calorimeter in which brass layers are interspersed with plastic scintillators.

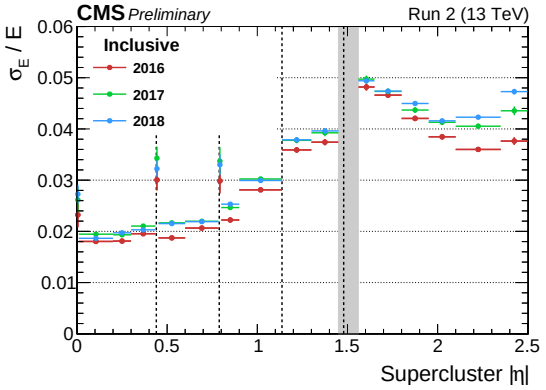


FIGURE 2.4: Energy resolution of the ECAL measured in Run 1 and Run 2 data as a function of η. The resolution is inferred from $Z \rightarrow ee$ events. Taken from [35].

2.2.5 Muon spectrometer

The muon system is located in the outermost part of the detector and is dedicated to the measurement of muons, which are the only particles, besides neutrinos, that are able to traverse the calorimeters and the solenoid. The detection and identification of muons is crucial for the CMS physics program, as their presence is one of the possible signature of many interesting processes. Besides the identification of muons, the muon system is designed to also be able to trigger based in their presence and to measure their momentum with a relatively high precision. The distribution of the

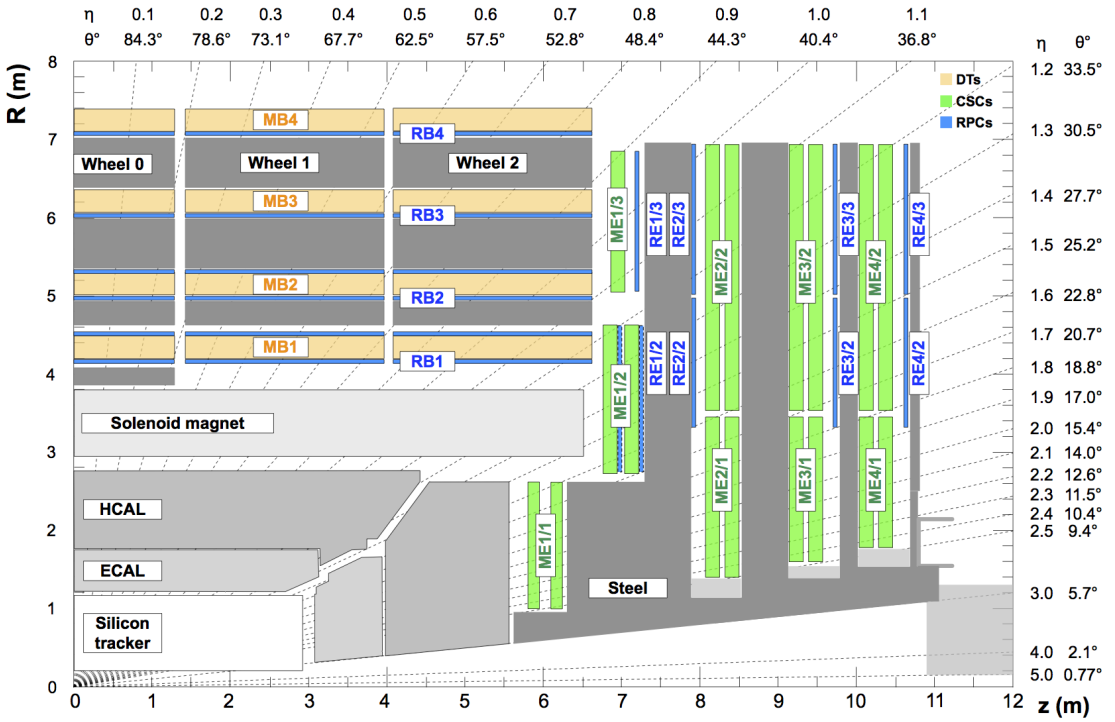


FIGURE 2.5: Outline of the CMS muon spectrometer. The plot shows an R-z cross section of a quadrant of the CMS, highlighting the muon systems. Taken from [36].

muon spectrometer in CMS is shown, together with the rest of the subdetectors, in figure 2.5.

Three kinds of gaseous muon detectors are used in the CMS detector: the Drift Tubes (DTs), Resistive Plate Chambers (RPCs) and Cathode Strip Chambers (CSCs). The three kind of detectors have distinctive features and are used in a complementary way.

DTs are located in the barrel of the CMS detector ($|\eta| < 1.2$), where the neutron-induced background is small and low muon rates are expected. It is organized in 5 wheels situated in perpendicular to the z axis. Each wheel consists of four stations interspersed among the layers of the magnet return yoke. Each station is a set of chambers located in a ring with the same radius, that is able to measure both the z direction of the muon trajectory as well as the $r - \phi$ bending angle. The outermost ring does not measure directions along the z direction. DTs are divided in several drift cells filled with gas, and the muon position is inferred by measuring the drift time to an anode wire of the charge produced by the ionization produced by the muon.

In the endcap regions of CMS, a larger contribution from background is expected. Moreover the magnetic field is larger and less uniform. CSCs are situated in $0.9 < |\eta| < 2.4$ and consist of four vertical stations in each endcap. CSCs operate as multiwire proportional chambers, with finely segmented cathode strips. The wires are pointing in the radial direction, which allowing to precisely measure the ϕ component of the muon interpolating among the charges induced in the strips.

RPCs are located in $|\eta| < 1.9$. They are a gaseous parallel-plate detector, that features a very good timing resolution, better than the minimum 25 ns between consecutive bunch crossings. CSCs complement the DTs and CSCs providing very good trigger capabilities.

The muon system has in general a very good coverage along the $|\eta| < 2.4$ range. The regions with $|\eta|$ around 0.25 and 0.8, that correspond to the transition region between the DT wheels and 1.2, in which the transitions between DTs and CSCs are located, show however a slightly lower efficiency. The excellent performance of the CMS muon system is described in chapter 3, in which the muon reconstruction algorithms are described.

2.2.6 Trigger system

The LHC collision rate can be as high as 40 MHz. This, together with the complexity of the CMS detector, require the need for a trigger system that promptly evaluates and selects the collisions that are interesting and must be stored, discarding the rest. The CMS trigger system consists of two logical layers: First Level Trigger (L1) and the High Level Trigger (HLT).

The L1 takes decisions based on the detector information obtained from a coarse readout of the calorimeters and the muon system. The tracking detectors are not used at this level since it is not possible to read out all the information at every bunch crossing. The L1 system allows for a latency up to 4 μs, during which events are buffered until they are either accepted or discarded.

Several algorithms are run in L1, that is implemented in hardware. Firstly each subdetector has its own local trigger, that promptly reconstructs the energy deposition in the calorimeters and muon hits in the subdetector. Then a two-layer trigger allows to reconstruct electrons, jets and hadronic τ from the calorimeter information. In parallel, three track finders combine the information from the different muon subdetectors to construct muon candidates. Finally the information from the reconstructed calorimeter and muon objects is received by a global trigger that makes the final L1 decision.

Most of the bandwidth of the L1 is dedicated to simple topologies, like single and double objects, as well as combinations of objects with certain momentum thresholds. As an example, a combined trigger may require the presence of an electron and a muon. However, trigger algorithms allow to calculate invariant masses and angular distances between objects, allowing for more complex topologies. The readout of the CMS detector imposes an upper limit of 100 kHz on the acceptance rate of the L1 trigger.

Upon a positive decision of the L1 trigger, the full event information can be read out and reconstructed by the HLT. At the HLT, events are reconstructed using a computing farm of commercial computers. The software used to do so is mainly written in C++ and integrated with the rest of the code used in offline event reconstruction, and in some occasions it coincides.

Event reconstruction at the HLT is seeded by a positive decision of one or more L1 triggers. One or several algorithms (or paths) are ran, targeting specific topologies,

and are seeded by one L1 decision targeting that topology. Each path performs a increasingly complex object reconstruction, aiming to promptly reject spurious events in order to reduce the computing time taken by each event, and preforming a more precise reconstruction for likely candidates. The average processing time of an event depends on the running conditions, such as the pile-up, but it is usually around 150 ms, as it can be seen in figure 2.6. It can also be seen that the bulk of the events are processed quite rapidly, since for them a complete reconstruction is not needed to make a decision, while for a few, in the tails, it is needed to run more complex algorithms that can take up to one second. The HLT accepts a rate of around 1000 Hz, integrated over several hours, that can be stored and further processed.

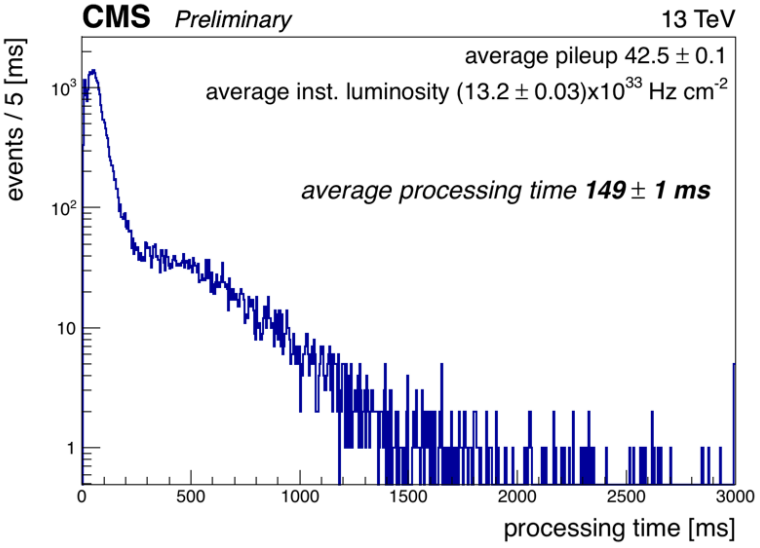


FIGURE 2.6: Event processing time at the LHC in a run taken during 2016. Taken from [37].

2.2.7 Computing system

The amount of luminosity delivered by the LHC and collected by the associated experiments corresponds to an overwhelming amount of data that must be stored and analyzed. Besides data itself, event simulations of physical processes are performed that emulate the complete detector response, which is a computationally intensive task. These simulations also need for significant storage resources. These kind of computational and storage capabilities are not available in any single existing institution. Instead, CMS implements a computing model based on the Worldwide LHC Computing Grid (WLCG) [38], a network that interconnects computer based in different locations in the world.

Computing centers around the world are classified in types with different roles. A single Tier-0 cluster, located physically at CERN and in Budapest, is used for raw data acquisition and triggering, its storage and its prompt reconstruction. The Tier-0 consists of around 230K computing cores and 800 PB of storage. The next layer is composed by 13 Tier-1 facilities, to which raw and reconstructed data is transferred. Larger scale data reconstruction is also performed on Tier-1 clusters. Tier-2 are used

for simulation of collision events, and they provide computing time and data storage for users of the institutions. Finally, Tier-3 usually correspond to smaller computing infrastructures that are used by users of institutions to perform analysis tasks.

This thesis has been performed profiting from the data reconstruction and simulations produced in the computing network described. Additionally, a significant part of the computing resources needed for the analyses in chapters 4 to 6 have been performed in the Tier-2 facility in IFCA (Santander) and the Tier-3 cluster in the University of Oviedo. In total, this thesis has used an order of a few tens of years of CPU time in the Tier-3.

2.3 Event reconstruction in CMS

2.3.1 Particle Flow

Event reconstruction of data aims to identify the collision products and measure their kinematic properties, to ultimately construct the relevant physical observables. The Particle Flow (PF) algorithm [39] performs this task by combining the information from all the subdetectors described in the previous section.

The inputs to the PF algorithm are the tracks measured in the tracking system, the energy depositions in the calorimeters, as well as information of the muon system. Information on the electron and muon reconstruction is also introduced as an input, to exploit the peculiarities of these objects, as described in chapter 3.

The tracking of charged particles in the inner detectors is performed with the means of an iterative tracking algorithm. The algorithm runs several tracking iterations trying to keep an increasingly high efficiency but very high purity. Hits associated to a track in a given iteration are then masked for the following iterations in order to avoid them being associated to a different track. At each step of the iteration, different quality criteria are applied on the track seeds, the χ^2 as well as the probability from the coming from one of the reconstructed vertices, all of them adapted to the track p_{T}, $|\eta|$ and the number of hits. From the reconstructed tracks, primary vertices corresponding to several interactions can be identified. In this thesis, the primary vertex with the highest quadratic sum p_{T} of their tracks is taken as the vertex of the hard scattering.

Clustering of energy depositions in the calorimeters is done separately for each subdetector, seeded by local maxima of energy depositions. Then clusters are built by aggregating neighboring depositions. This algorithm aims for a high efficiency when detecting low energy particles and high spacial resolution.

The PF algorithm gathers these inputs using a linking algorithm that connects them. This connection is made geometrically, either extrapolating the reconstructed tracks to the calorimeter cells or through the position the calorimeter deposits. Tracker tracks may also be linked to hits in the muon system. Since particles interact differently with each one of the subdetectors, this linkage allows for particle identification. Muons and electrons are built by the association of tracks with segments in the muon systems and deposition in the ECAL, respectively, as described in chapter 3. Charged hadrons produced a tracker track and depositions in the calorimeters. Photons and neutral hadrons only deposit their energy in the ECAL and HCAL, respectively.

Missing transverse momentum

The complete reconstruction of all the particles in the collision, as well as their assignment to a specific primary vertex allows to infer the presence of invisible particles in the final state, such as neutrinos but also BSM particles, like the SUSY LSP. The missing transverse momentum, $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$, is defined as

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} = -\sum_{i} \vec{p}_{\mathrm{T}}(i), \tag{2.5}$$

where the sum runs over all the observed particles from the selected primary vertex. The magnitude of $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ is usually denoted as $p_{\mathrm{T}}^{\mathrm{miss}}$. Due to momentum conservation, if only visible particles were present in the final state and in absence of resolution effects, this quantity would be zero. Therefore, zero values of $p_{\mathrm{T}}^{\mathrm{miss}}$ hint to the presence of an invisible particle.

Several algorithms are used to identify and reject events with spurious $p_{\mathrm{T}}^{\mathrm{miss}}$ due to failures of the reconstruction algorithm or the detectors. These algorithms [40], or filters, identify several of these sources, such detector noise in the calorimeters, background induced by the beams or poorly reconstructed high p_{T} muons.

In figure 2.7, the $p_{\mathrm{T}}^{\mathrm{miss}}$ resolution during the 2016 data taking period is shown in $\mathrm{Z} \to \mu\mu$ events, also compared with simulations, showing a reasonably good agreement. In $\mathrm{Z} \to \mu\mu$ events all the final state objects are visible, so non-zero values are due to instrumental effects. The effect of the $p_{\mathrm{T}}^{\mathrm{miss}}$ filters is also shown in a sample of multijet events in the same figure.

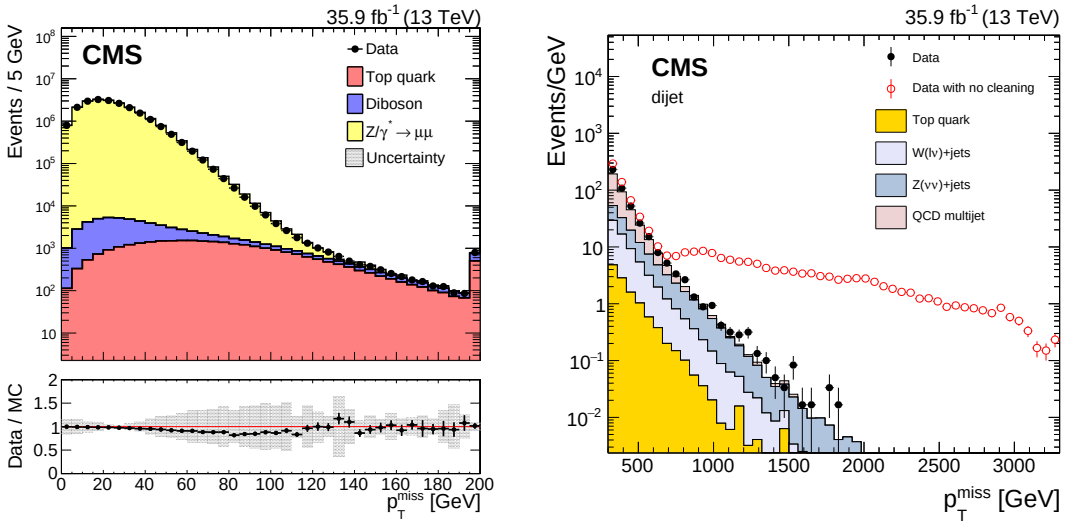


FIGURE 2.7: Left: distribution of instrumental $p_{\mathrm{T}}^{\mathrm{miss}}$ in DY+jets events. Right: $p_{\mathrm{T}}^{\mathrm{miss}}$ distribution in multijet events passing the $p_{\mathrm{T}}^{\mathrm{miss}}$ filters (black markers) and not applying them (red markers). Taken from [40].

In addition, in some occasions the H_T^{miss} variable is considered. It is defined in the same way as $p_{\mathrm{T}}^{\mathrm{miss}}$, but only objects (jets, light leptons and hadronic τ) passing specific selection criteria in the analysis are used. This variable is less sensitive to the

presence of invisible particles, but is more robust to energy mismeasurements and spurious signals. In the $\mathrm{t\bar{t}H}$ multilepton analysis described in chapter 6, a linear combination of the two is considered, $p_{\mathrm{T}}^{\mathrm{miss}}$ LD, that provides a good compromise between discrimination power and robustness.

2.3.2 Jets

Due to the quantum confinement described in chapter 1, colored final states are not observable. Instead, quarks and gluons radiate partons, that eventually hadronize into hadrons, which are bounded, colorless states. The observable for a parton emission is then a spray of particles, named jet, that is more or less collimated in a given direction. From a phenomenological point of view, jets may be defined by clustering all particles in the final state. However, clustering algorithms must be collinear and infrared safe in order to be a suitable to be compared with theoretical predictions: the additional emission of soft or collinear partons should not affect the result of the clustering.

In this thesis, jets are defined using the anti-k_{T} algorithm [41], that is run on all the particles identified by the PF algorithm to come from the selected from the selected primary vertex. Usually a distance parameter of 0.4 is taken, however in the search for SUSY jets with a distance parameter of 0.8 are also used, to reconstruct large radius jets produced in the hadronic decay of boosted Z and W bosons.

In order to reject jets originating from detector noise and misidentified particles, very mild selection criteria are applied to reconstructed jets in order for them to be applied in the analysis. They are simple requirements in the minimum number of constituents, as well as minimum hadron and electromagnetic energy fractions [42].

2.3.3 b-jet tagging

It is usually interesting to identify the flavor of the partons that have produced a jet. The identification of jets produced by b quarks, referred to as b jets, is particularly interesting, since they appear in the decay of top quarks and other massive particles. b jets have distinctive features due to the presence of hadrons containing b. These hadrons usually have a longer lifetime that other light flavor hadrons. Because of this, b-jets usually present a secondary vertex, corresponding to the decay of the hadron after it has flown a given distance. Additionally, b quarks have a larger mass than light flavor quarks and gluons, so particles in the decay have a larger momentum relative to the jet axis than other constituents. Finally, these decays may also give raise to electrons and muons.

Several algorithms are being used in CMS to identify b jets [43, 44]. The field of flavor tagging has significantly evolved during the time this thesis was made. Therefore three different algorithms are employed: the DeepCSVv2 algorithm, the DeepCSV and the DeepJet algorithm. The three of them are neural networks exploiting similar kinematic features, however the technical complexity is increasingly large, resulting in a improved performance. Three Working points (WPs) are established to tag b jets: loose, medium and tight, targeting to a 10, 1 and 0.1% of light jet misidentification probability, approximately.

The performance of the DeepCSV and DeepJet algorithms are shown in figure 2.8 in simulations, compared to the performance in simulations calibrated by the efficiency and mistag rate observed in data for the mentioned working points.

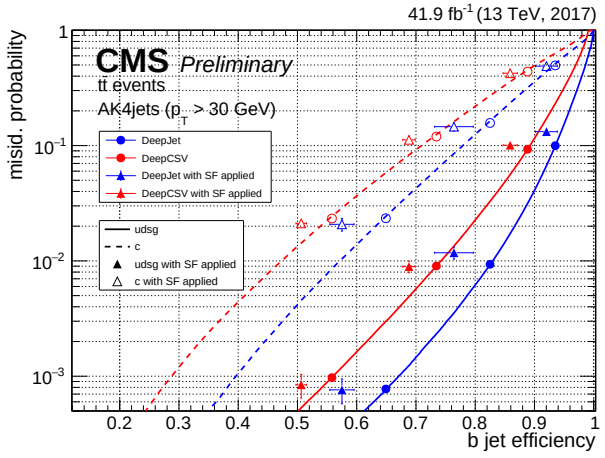


FIGURE 2.8: Performance of the DeepCSV and DeepJet tagging algorithms in $\mathrm{t\bar{t}}$ events. Misidentification probability for light jets (lines) and c jets (dotted lines) as a function of the b tagging efficiency. Taken from [45].

Finally, other discriminators exist that allow to discriminate also between jets produced by quarks and gluons [46], exploiting features such as the higher multiplicity and a softer fragmentation in gluons jets.

2.4 Datasets: conditions and performance

Due to the complexity of the CMS detector and the LHC, the conditions under which it is run change with the time. As two examples, it was already shown in the previous sections that the geometry of the CMS pixel detector was changed from the 2017, adding a new tracking layer, but also the pile-up conditions were increased during the 2017 and 2018. These effects are carefully taken into account when analyzing data, thanks to dedicated corrections or specific simulations. In this section, the most important features of the data taking are described.

Pile-up conditions As shown in figure 2.1, the number of simultaneous interaction per bunch crossing was significantly larger during the 2017 and 2018 data taking. This resulted in an increase of the instantaneous and integrated luminosity during those years, that allows to explore even lower cross-section processes. However, this also results in a slightly degraded performance during those years in the measurement of $p_{\mathrm{T}}^{\mathrm{miss}}$, jet energy and lepton isolation. Additionally, the increase of the instantaneous luminosity also meant a raise in the momentum thresholds of the triggers, in order to keep the rate under control.

Pixel performance The CMS pixel detector was upgraded during 2017. The new pixel detector has an improved readout and has four layers and three disks, ensuring

a 4-hit coverage across all the tracker acceptance. Additionally, the innermost layer is closer to the interaction point, allowing for a better impact parameter resolution. However, during the 2017 data taking some of the pixel modules failed due to the radiation-induced damage in components of their power supply. This resulted in a slight degradation of the performance, which is partially recovered by the redundancy of the system. The failing components were fully replaced before the start of the 2018 data taking, and no significant failures have been observed since.

L1 prefiring The progressive transparency loss of the ECAL crystals led to a gradual shift in the timing of the ECAL. This shift was not propagated to the L1 calibration during the 2016 and 2017 data taking and, because of this, a fraction of the ECAL objects at high $|\eta|$ firing the trigger were associated to the previous bunch crossing. When this occurs only the previous bunch crossing is accepted by the trigger, because the L1 logic does not allow two consecutive bunch crossings to fire the trigger. Then, the previous bunch crossing is also rejected in the HLT, since the event is empty. This effect is studied by selecting events two bunch crossings after a trigger has been fired. These events cannot be affected by the pre-firing because of the trigger rules, and the prefiring rate can be measured. This effect was found to be of the order of a few percent. It was accounted for in the L1 calibrations for the 2018 data taking, so no calibration is needed.

CMS Hadron Endcap issue During 2018 data taking a sector of the HCAL endcap failed and could not be recovered. This results in a miscalibration of jets and electrons in a small region of the phase space. In the SUSY analysis described in chapter 5 this is accounted for vetoing all events with a jet, electron or muon in that region. In the $\mathrm{t\bar{t}H}$ analysis described in chapter 6 an additional uncertainty is applied to account for a possible miscalibration.

2.5 Corrections to simulations

Monte Carlo simulations are used to estimate the production rate of events with given kinematic properties. Dedicated simulations are typically used for signal and background processes, in a way that physics analyses can be optimized, and also to interpret them in the context of the SM or any BSM theory. Generators described in section 1.3 quite reliably model most of the physical processes that can happen in pp collisions, but the interaction of final state particles with the detectors must also be modeled in order to have a complete prediction. In CMS, the GEANT [47] software is used to model the complete detector geometry, the interaction of the final state particles with each one of the subcomponents, as well as the digitalization of the detected signals. Then these signals are analyzed as if they were actual collision data, running all the reconstruction algorithms on them. These simulations are referred to as full simulations. Since the complete simulation of the detector is a computationally intensive process, for the modeling of several processes, the CMS Fast Simulation package [48] is used instead, that allows for a much faster emulation of the components, with a simplified geometry and emulating the response of the subdetectors with what is observed in full simulations.

Despite of the complete simulation of the CMS detector, slight differences between data and full and fast simulations are observed in the detector response. This ultimately affects the estimation of the object identification and reconstruction performance. Since it is aimed to precisely estimate all the observables, this performance is carefully measured in data and these small discrepancies are corrected. In this section, some of the methods used for this are described.

2.5.1 Lepton efficiency corrections

The efficiency of the lepton selection and reconstruction is corrected using the tag-and-probe method [49]. Since a focus is put in lepton reconstruction and selection in this thesis, this method is reviewed in greater detail in chapter 3.

2.5.2 Trigger efficiency corrections

The trigger efficiency can be measured using different methods, depending on the nature of the trigger. In this thesis it is made use of different combinations of single and double lepton triggers. Single lepton trigger efficiencies can be easily measured using the tag-and-probe technique. This technique could be in principle extrapolated to dilepton triggers, but it gets increasingly complicated when using more elaborated trigger strategies or when the efficiencies of the two leptons cannot be factorized.

In this thesis, an alternative approach, called orthogonal trigger method, is used instead. Efficiency is measured in events collected using a trigger whose decision is independent from the presence of the objects we aim to trigger on. Triggers based on the momentum imbalance of the event are used. These triggers are indeed not fully independent from single and di-lepton triggers, since a lepton mismeasurement could yield to instrumental $p_{\mathrm{T}}^{\mathrm{miss}}$. However, these correlations are found to be small and are usually assigned as a systematic uncertainty in the measurements. Efficiency for a given trigger in events passing a offline selection is measured as

$$\epsilon = \frac{\mathrm{N}(\text{trigger} \wedge \text{offline} \wedge \text{orthogonal})}{\mathrm{N}(\text{offline} \wedge \text{orthogonal})}. \tag{2.6}$$

These measurements can be performed, in the case of dileptonic events, as a function of the lepton kinematics. Several uncertainty sources may be considered: the comparison with the measurements using other orthogonal triggers, or with the ones obtained with the tag-and-probe method, the correlations between orthogonal and measured triggers in simulations, residual kinematic dependencies, as well as the statistical uncertainty of the data samples.

2.5.3 Jet energy and resolution corrections

Jets reconstructed are calibrated in order to provide a more accurate estimation of their momenta and in order for simulations to properly describe data. The set of corrections is fully described in [50]. Using simulated events, corrections are derived

to subtract contributions to the jet energy from pile-up and UE particles. Similarly, other corrections are applied in order to make the measured jet energy correspond to that of the generator level jets. Finally, residual corrections are applied to simulations correct for the small discrepancies in data and simulations. These corrections are determined by studying the momentum imbalance in dijet events, between jets and electron and muon pairs in DY+jets events, and between jets and photons in γ+jets events.

These calibrations, as well as the corresponding uncertainties, are propagated to the calculation of $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$.

2.5.4 b tagging corrections

Both the efficiency to identify b jets as such and the misidentification probability of light flavor jets as b jets must be taken into account when calibrating the b tagging algorithms. These measurements are fully described in [43].

Misidentification probability is measured in an inclusive sample of multijet events. The so-called negative (positive) taggers are built by constructing the b-jet tagger using only tracks and secondary vertices with negative (positive) impact parameter and flying distance. These taggers are symmetric for light flavor jets, so the mistag rate for a given working point is determined as the proportion of jets that pass that working point of the negative tagger. The ratio between the misidentification probability of light jets and the misidentification rate of the negative tagger in all jets is taken into account using simulated events.

Jet b-tagging efficiency is measured in samples of jets enriched in b jets, either by selecting suitable samples of multijet events or selecting $\mathrm{t\bar{t}}$ events. Several complementary approaches are followed to obtain a efficiency for each working point.

In some occasions, it is necessary to calibrate the efficiency and misidentification rate as a function of the value of the discriminator. In these cases, an iterative approach is followed, selecting $\mathrm{t\bar{t}}$ and DY+jets events and deriving calibrations independently. In the two cases, contamination from b and light flavor jets exist, so the determination of one rate affects the determination of the other. This procedure is followed iteratively until a convergence is achieved.

2.6 Statistical analysis techniques

When performing physics analyses, the usage of statistical tools is necessary in order to improve the performance of the analysis itself, but also to interpret the measured observables in terms of physical quantities. Several of such techniques are used along this thesis, and are introduced in this section.

2.6.1 MVA discriminants

In this thesis, Multivariate (MVA) discriminators are used to classify events and objects between two or more categories using a set of input variables. Usually, these

two categories are signal and background, however in the $t\bar{t}H$ multilepton analysis described in chapter 6, classification among several background and signal species is used. These algorithms achieve, by construction, better discrimination performance than usual techniques using sequential cuts.

Two types of supervised learning algorithms are used in this thesis: Artificial Neural Networks (ANNs) and Boosted Decision Trees (BDTs). This kind of algorithms are usually trained in sets of simulated samples, in which the species of each one of them are known a priori. A more detailed description of these methods can be found in the usual literature [51, 52].

BDTs are built from decision trees. Decision trees are classifiers that consist of the application of several sequential criteria. These criteria are set sequentially, by applying first a classification based on the most discriminating variable, dividing the training set into two subsets or nodes, one of them more enriched in signal and the other in background. Then the same procedure is applied iteratively to each one of the subsets separately, until a reasonable purity is achieved in each of the nodes. The maximum number of such decisions that are made is denoted as the depth of the tree. Decision trees can perform very well when considering a small number of input variables and large training datasets. However they are very sensitive to fluctuations in the training dataset, easily leading to overtraining. If this occurs, the classifier does not generalize to other datasets different from the training one.

In order to avoid overtraining, a technique called boosting is used, in which an ensemble of weak learners is trained. In this case weak learners are shallow trees. Intuitively, the first weak learner is trained using the approach above, and the rest are focused to properly classify samples for which the previous learners have failed. Several boosting algorithms have been developed. In this thesis, the implementation of gradient boosting in the TMVA software [53] is used.

Many types of ANNs have been developed in the past years [51]. In this thesis, only feed-forward neural networks are considered. They are mappings built in the following way. Given an activation function f, a neuron is $f(\sum a_i x_i + b_i)$, where x_i are input variables, and a_i and b_i are the weights and biases, parameters of the model. An set of neurons $\mathbf{L}$ is defined to be a layer of the network, which is formed by the composition of several layers, $\mathbf{L_1} \circ \cdots \circ \mathbf{L_d}$. The network, which is said to have a depth d, corresponds to a mapping $\mathbb{R}^n \to \mathbb{R}^m$, where n is the number of input variables and k any integer.

It can be proven, under mild assumptions on the activation functions, that any bounded $\mathbb{R}^n \to \mathbb{R}^m$ function can be approximated in a compact $\mathbb{R}^n$ with a neural network with enough neurons and with a suitable choice of weights and biases [54]. The parameters of the network are trained in a simulated dataset of the different signal and background species, choosing the set of weights and biases minimizing the cross-entropy variable. In the context of this thesis, neural networks make use of rectified linear units as activation functions. The minimization is performed using the batch gradient descent algorithm implemented in the Tensorflow [55] package, with the Keras [56] interface.

2.6.2 Statistical treatment of the results

The three physics analysis described in this thesis define categories, in which event counting experiments are performed. Then, inference is performed in these observables to obtain relevant physical information. In the case of the tW inclusive production cross-section and $\mathrm{t\bar{t}H}$ production measurements, confidence intervals are drawn on strength parameters. In the search for SUSY production, since no signal is observed, upper limits are drawn on the SUSY signal strength for each model. In this section, the statistical framework used is described. This framework coincides with the one described in [57].

Given a set of counting experiments observations, x_i, a likelihood function is defined as

$$\mathcal{L}(\boldsymbol{x} \mid \boldsymbol{\mu},\ \boldsymbol{\theta}) = \prod_i \mathcal{P}\left(x_i,\ \boldsymbol{\mu}\boldsymbol{s_i}(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta})\right) \prod_j p_j(\tilde{\theta}_j \mid \theta_j). \tag{2.7}$$

In the equation, $\boldsymbol{\mu}$ are the signal strengths for the various signals, $\boldsymbol{s_i}$, their contribution to a given category i and b_i the contribution from background. The latter two are function of $\boldsymbol{\theta}$, the nuisance parameters, that parametrize the systematic uncertainties affecting to their estimation. The symbol $\mathcal{P}$ denotes the probability function of the Poisson distribution.

The functions $p_j(\tilde{\theta}_j \mid \theta_j)$ are the a-priori expectations for the nuisance parameters, and constrain them. This function is a log-normal probability density function in nuisances affecting the normalization of processes, and a gamma function if the uncertainty is of statistical origin. In the $\mathrm{t\bar{t}H}$ multilepton analysis, the approach described in [58] is followed for the latter type of uncertainties. Systematic uncertainties affecting also the shape of the distributions are incorporated, following the morphing technique described in [59], and their prior is Gaussian. In some occasions, the normalization of given backgrounds is left unconstrained in the fit.

The best fit for $\boldsymbol{\mu}$ is obtained by maximizing the likelihood, $\mathcal{L}$. Confidence intervals and upper limits are obtained by considering the profile likelihood ratio t-statistic:

$$q(\boldsymbol{\mu}) = -2\log \frac{\mathcal{L}(\boldsymbol{x} \mid \boldsymbol{\mu},\ \hat{\hat{\boldsymbol{\theta}}}(\boldsymbol{\mu}))}{\mathcal{L}(\boldsymbol{x} \mid \hat{\boldsymbol{\mu}},\ \hat{\boldsymbol{\theta}})}, \tag{2.8}$$

where $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\theta}}$ are the best fit values, and $\hat{\hat{\boldsymbol{\theta}}}(\boldsymbol{\mu})$, the best fit values for a given choice of $\boldsymbol{\mu}$. Due to the Wilk's theorem, in the limit of large number of observations, $q(\boldsymbol{\mu})$ is distributed as a χ^2_n distribution under the hypothesis of $\boldsymbol{\mu}$ being the true value, where n is the dimension of $\boldsymbol{\mu}$. Then confidence intervals can be constructed by considering the crossings of $\boldsymbol{\mu}$ with the quantiles of the χ^2 distribution.

Upper limits on the production of a signal can also be obtained by considering the same t-statistic as

$$\mathrm{CL}_{s+b}(\boldsymbol{\mu}) = P(q(\boldsymbol{\mu}) > q^{obs}(\boldsymbol{\mu}) \mid \boldsymbol{\mu}). \tag{2.9}$$

A possible way to define the exclusion limit would be with the Neyman construction, taking $\mathrm{CL}_{s+b}(\boldsymbol{\mu}) < \alpha$ as an upper limit at confidence level (CL) α. However, by construction, $\mathrm{CL}_{s+b}(0) < \alpha$ will occur in an $1-\alpha$ proportion of the times, i. e., this will happen a 5% of the times if we aim for upper limits at 95% CL. A modification of this upper limits is therefore considered. Instead, the upper limit is defined as $\mathrm{CL}_s(\boldsymbol{\mu}) < \alpha$, defining

$$\mathrm{CL}_s(\boldsymbol{\mu}) = \frac{\mathrm{CL}_{s+b}(\boldsymbol{\mu})}{\mathrm{CL}_b(\boldsymbol{\mu})}, \tag{2.10}$$

where $\mathrm{CL}_b(\boldsymbol{\mu}) = P(q(\boldsymbol{\mu}=0) > q^{obs}(\boldsymbol{\mu}=0) \mid \boldsymbol{\mu}=0)$. By construction, the resulting limits are one-sided, and overcover the true value, since they are strictly higher than the CL_{s+b} ones. These upper limits can be computed using asymptotic formulae [60] for the limit with large number of observations. In the cases in which this approximation is not valid, they can be estimated drawing a set of pseudoexperiments.

2.6.3 Unfolding

In some occasions, it is interesting to study the production cross section as a function of a given observable of interest. In the case of inclusive measurements it is trivial to remove the effects due to the detector resolution by computing the efficiency. However, when considering differential distributions one should also consider migrations among bins of the distributions. The method to do so is called unfolding. A more complete review on the different unfolding techniques can be found in reference [61]. This section covers the method followed in the differential cross section measurement in tW, described in chapter 4.

2.6.3.1 Response matrix

The problem can be stated in terms of x, a given variable in the before any detector effect is applied, and y, the reconstructed variable. Usually these measurements are done over binned datasets, although this binning can typically be chosen. We denote x_j the number of signal events in a given bin of x and y_i the number of signal events in a bin of y.

To model the limited resolution of the detectors, the response matrix, $\mathcal{R}$, parametrizes the signal event migrations between the detector-level and particle-level variables. The matrix is defined in a way $y_i = \mathcal{R}_{ij}x_j$, and is constructed using signal simulated events, after they have been corrected by the corresponding calibrations and scale factors. With these simulations, the response matrix can be computed as

$$R_{ij} = \frac{\text{signal generated in bin } j \text{ and reconstructed in bin } i}{\text{signal generated in bin } j}.$$

This matrix takes into account signal efficiency as well as migrations between bins of the distribution. When applying this method in actual data, a contribution from

background events may appear in the signal region. In order to account for this, the estimated backgrounds are subtracted from data

$$\hat{y}_i = N_i - N_i^{bkg} = \sum_{j=1} \mathcal{R}_{ij}\hat{x}_j. \tag{2.11}$$

where $\hat{y}_i$ and $\hat{x}_j$ represent estimators for y_i and x_j, respectively.

If the response matrix is inverted, equation 2.11 can be used to obtain an unbiased estimator for x_j. However, as it will be shown below, the non-diagonal terms in the response matrix may increase significantly the covariance of this estimator. Because of this, it is crucial to choose a binning of the generator and reconstructed-level variables that ensures that the matrix is as diagonal as possible.

To quantify the levels of non-diagonality of the response matrix, the condition number of the matrix is used. The condition number of a given matrix $\mathcal{R}$ is defined as $\text{cond}(\mathcal{R}) = \sigma_{\text{max}}/\max(0, \sigma_{\text{min}})$, where $\sigma_{\text{max(min)}}$ are the largest (smallest) singular values of $\mathcal{R}$. If the condition number is small, of the order of 10, the problem is well-conditioned and can be solved using using the maximum likelihood estimator. If it is large, the problem is ill-conditioned and regularization techniques may be needed.

The binning is optimized to achieve the maximum stability and purity. These quantifies are defined, respectively as:

$$s_i = \frac{\mathcal{R}_{ii}}{\sum_j \mathcal{R}_{ij}},\ p_j = \frac{\mathcal{R}_{jj}}{\sum_i \mathcal{R}_{ij}}.$$

2.6.3.2 Regularization

As mentioned above, an unbiased estimator for y_i would be obtained by inverting the response matrix. However, this inversion may amplify the statistical uncertainty of data, leading to large uncertainties in the estimated unfolded distribution. This estimator has a variance equal to the Frechet-Cramer-Rao bound [61], hence it can only be improved by introducing a bias to reduce its variance.

The approach followed in this analysis is to consider Tikhanov regularization [62], as it is implemented in TUnfold [63]. This approach is performed with the following construction. A solution to the problem can be obtained by minimizing the following cost function:

$$\begin{aligned}
\mathcal{L} &= \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3 && (2.12)\\
&= (\mathbf{y} - \mathcal{R}\mathbf{x})^T V_y (\mathbf{y} - \mathcal{R}\mathbf{x}) && (2.13)\\
&+ \tau^2 \mathbf{x}^T (L^T L)\mathbf{x} && (2.14)\\
&+ \lambda(Y - e^T\mathbf{x}), && (2.15)
\end{aligned}$$

where V_y is the covariance matrix associated to $\mathbf{y}$, $Y = \sum_i y_i$ and $e_j = \sum_i \mathcal{R}_{ij}$. The first term $\mathcal{L}_1$ corresponds to the likelihood function that is maximized to obtain the maximum likelihood estimator, which corresponds with the solution associated to the matrix inversion. The $\mathcal{L}_2$ and $\mathcal{L}_3$ terms introduce a bias to this solution, that allows to reduce the variance of the estimator.

$\mathcal{L}_2$ is a regularization term that penalizes larges curvatures in $\mathbf{y}$, associated to numerical instabilities, and is modulated by a term, τ, that regulates the size of the bias introduced. Different L matrices can be applied, depending on the instabilities to be penalized. In this case, the bias aims to penalize large curvatures or second derivatives, hence $L_{ij} = 2\delta_{ij} - \delta_{ij-1} - \delta_{ij+1}$ is taken, where δ represents the Kronecker δ.

$\mathcal{L}_3$ is an area constraint, which a Lagrangian multiplier λ, that forces the estimated total number of events in the particle-level space to be consistent with the reconstructed space.

The optimal choice of the τ^2 parameter is determined using the L-curve method. Low values for values of τ^2 allow for larger fluctuations in data, while high values introduce a larger bias. The optimal value is then fixed by scanning the L-curve, which is defined as (L_x, L_y) with $L_x = \log \mathcal{L}_1$ and $L_y = \log \mathcal{L}_2/\tau^2$. For large τ values, higher values of $\mathcal{L}_2$ are penalized and high values of $\mathcal{L}_1$ are allowed. On the contrary, for small τ values $\mathcal{L}_1$ takes small values and $\mathcal{L}_2$ takes large values.

In the analysis shown in chapter 4, the scan in the L-curve was performed for all the studied variables, and the best τ value was tested. No need for regularization terms was needed in this analysis, and the Maximum Likelihood Estimator (MLE) is used.

3

Lepton reconstruction and identification at CMS

Leptons are the main objects used in this thesis. This chapter covers lepton reconstruction at CMS, with a focus on muon reconstruction and identification, since it was one of the main tasks I contributed to during this thesis. This is described in section 3.1. Electron reconstruction and identification, and the main isolation variables used in CMS are described in sections 3.2 and 3.3. Then, an ad-hoc discriminator developed for the $t\bar{t}H$ analysis is described in section 3.4. A complete description of lepton efficiency measurements and the associated uncertainties is shown in section 3.5. Finally, a short summary of the identification of hadronic τ is made in section 3.6.

3.1 Muon reconstruction and identification

In this section, the muon reconstruction methods used in CMS detector and the identification criteria used are described. A focus is put on the identification criteria that are used in the following chapters of the thesis. A complete description of the algorithms used can be found in [36]. The performance of the muon selection and reconstruction in Run 2 data is also shown.

3.1.1 Muon reconstruction

Muons are reconstructed in CMS by combining the information from the tracking detectors and the muon spectrometer, using three complementary algorithms.

Standalone muon tracks are constructed starting from a track built from groups of DT or CSC segments, and are propagated and updated using RPC, CSC and DT segments using a Kalman filter algorithm [64]. Tracker muon tracks are built starting from all reconstructed tracks in the inner tracker, which are propagated to the muon system. The propagated track is matched geometrically to segments in the muon system. A tracker muon is reconstructed if at least one segment is matched to the propagated track. Global muons are built starting from standalone tracks, that are propagated to the inner detector and matched to tracker tracks. A global fit is then performed with the Kalman filter using both the standalone and the associated tracker track. Tracker and global muons are merged into a single candidate if they share the same inner track.

About 99% of muons produced the geometrical acceptance of the muon system are reconstructed as a global or tracker muon. The global algorithm is designed to reconstruct muons that transverse the muon system with a very high purity. Tracker muons are instead used to recover efficiency in the regions where the muon system is not very instrumented.

Muons reconstructed by these algorithms are then used as an input of the PF algorithm. All global isolated muons matched to DT and CSC segments are selected as PF algorithm, while additional quality criteria are imposed on tracker muons or non-isolated global muons.

The default algorithm to measure the muon momentum takes the information from the global fitted trajectory and the tracker-only trajectory. The latter is used for muons with $p_{\mathrm{T}} > 200$ GeV if the charge-momentum ratio, q/p, agrees within two standard deviations with the tracker-only fit. In the rest of the cases, the inner track is used [65]. This approach is followed in the analyses described in this thesis. There are other more refined algorithms to estimate with higher precision the momentum of high p_{T} muons, for which tracker tracks will be almost straight and a significant amount of showering is expected in the muon system [66].

3.1.2 Muon identification

Several quality criteria are imposed to select muons at the analysis level. Working points are also defined targeting for several levels of purity and efficiency, and also different sources of muons. In this section, the three main selections used in this thesis are described. The $\mathrm{t\bar{t}H}$ multilepton analysis described in chapter 6 makes use of the prompt-lepton MVA described in section 3.4. There are other selections, not described in this thesis, that are tailored for high p_{T} and low p_{T} muons, that do not make use of the PF algorithm, and are described in [66].

The loose muon identification criteria aim to select muons produced in prompt decays in the primary vertex, but also those produced in light and heavy flavor decays, with a very high efficiency. This working point is used to efficiently reject charged hadrons that are reconstructed as muons. Loose muons are required to be selected by the PF algorithm and also to be either a tracker or a global muon.

The medium muon identification criteria aim to select muons produced in prompt decays and in heavy flavor decays. Medium muons are required to pass the loose selection and to use more than 80% of the hits in the tracker track. Additional quality criteria are based on the following quantities. The segment compatibility evaluates the consistency between the tracker trajectory and the segments in the muon system, as it would be expected from a minimum ionizing particle, returning a value between 0 and 1, with 1 representing the highest degree of compatibility. A kink-finding algorithm is used to evaluate the quality of the combined track. The algorithm divides the track in two at several points in the trajectory, and evaluates the compatibility between the resulting tracks, with a large χ^2 value indicating a large difference between the two. For global muons with a global track with a $\chi^2/\mathrm{dof} < 3$, the position match between the tracker and standalone track must have $\chi^2 < 12$, and with the kink-finding algorithm returning a score less than 20, segment compatibility is required to be higher than 0.303. This cut is relaxed to 0.451 otherwise.

The tight muon selection aims to reject muons from decays in flight and hadronic punch-through. A tight muon is a loose muon reconstructed as a global muon with global track $\chi^2/\text{dof} < 10$, and a tracker track using at least 6 tracker layers and at least one pixel hit. The track must be matched to at least two muon stations to suppress hadronic punch-through. Impact parameter cuts of 2 mm and 5 mm in the transverse plane and along the beam direction are applied, respectively.

3.1.3 Muon selection performance

In this section, the performance of some of the selections defined above are shown in pp collisions recorded with the CMS detector during years 2016, 2017 and 2018. During my thesis, I coordinated the efforts in the muon group of the CMS Collaboration to commission and evaluate the performance of these selections, as well as tuning them. These studies are documented in these public documents [4, 5]. The efficiency figures shown in this section were done in collaboration with the IFCA group, while the resolution performance studies were done by the Rochester group.

The efficiencies are shown for muons to pass the loose, medium and tight selection criteria is shown in figure 3.1, measured in $Z \rightarrow \mu^{\pm}\mu^{\mp}$ events collected with single muon triggers in CMS during Run 2 of the LHC. Efficiency is measured as a function of the muon η using the tag-and-probe method [49], with the parameters described in section 3.5. The efficiency of the three selection criteria is above 95% for the three selections, with the exception of the tight identification criteria in regions where the muon system is less instrumented. Results also show the robustness of the muon system and reconstruction algorithms across the different periods of the Run 2 data taking, with different detector geometry and pile-up conditions.

The momentum scale and resolution achieved by the CMS detector is shown in figure 3.2, as it is measured in $Z \rightarrow \mu^{\pm}\mu^{\mp}$ events corrected by the so-called Rochester method [67]. Events are collected with a single muon trigger, and muons are required to pass the tight identification criteria. The figure shows that scale of the Z resonance is determined with a high precision, and the associated resolution achieved, which has a mild dependence on the η of the muon.

3.2 Electron reconstruction and identification

The electron reconstruction at CMS, described in greater detail in [68], combines the information from the calorimeters and tracker to build electron candidates. Since electrons radiate photons via bremsstrahlung, their trajectory will change as they traverse the tracker. Also because of this, when an electron arrives to the ECAL, it may not be composed by a single object, but by a combination of electrons and photons that have been produced. The reconstruction algorithm aims to reconstruct all the associated object into an electron candidate.

Three types of seeds are used to reconstruct electrons. A dedicated algorithm is used to cluster ECAL deposits corresponding to a candidate electron and associated electrons and photons into a Super Cluster (SC). Additionally, doublets of tracker hits matched geometrically to the ECAL SC are used as seeds to the algorithm. The other

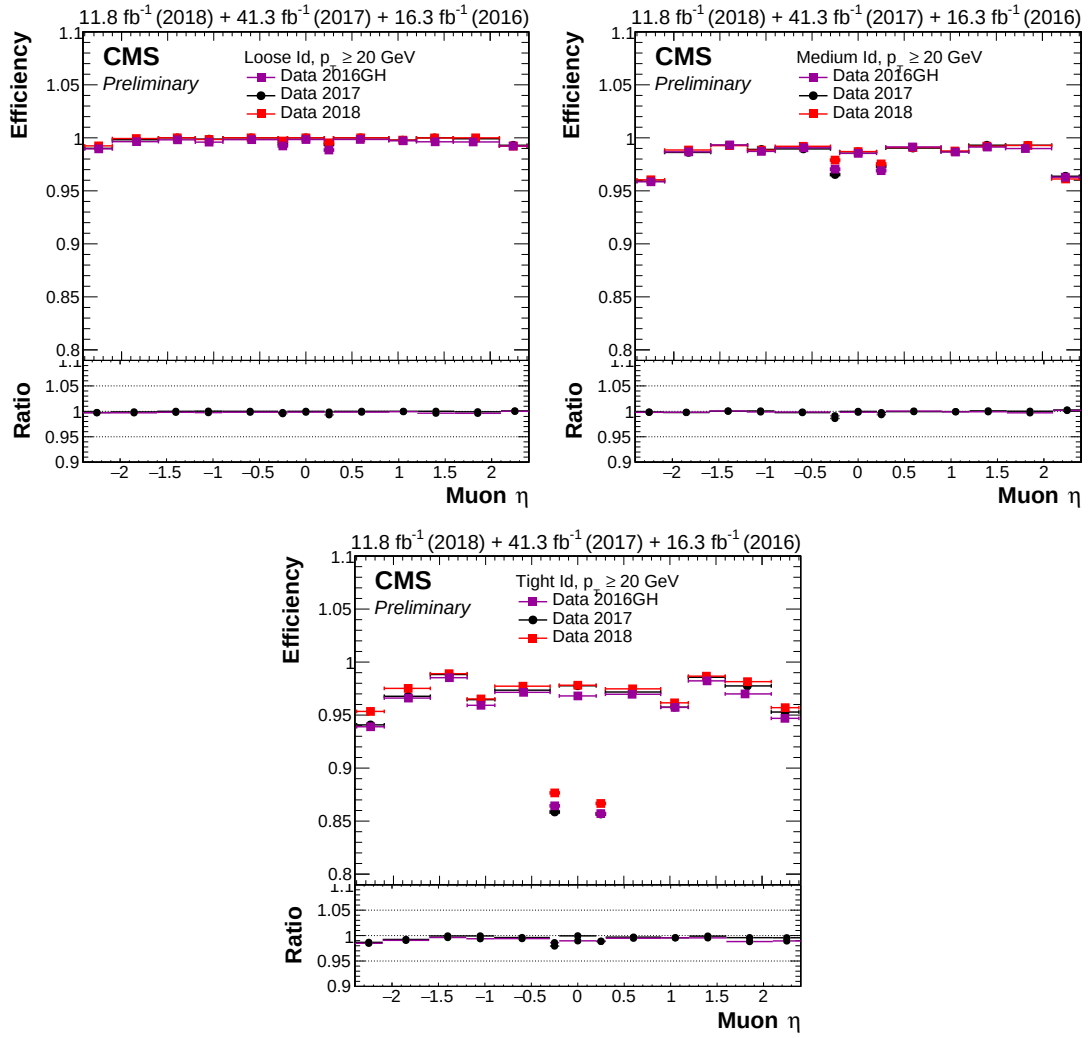


FIGURE 3.1: Efficiency for muons to pass the loose (top left), medium (top right) and tight (bottom) identification criteria. Efficiencies are measured using the tag-and-probe method. Taken from [4].

kind of seeds are tracker tracks, that are tested to be compatible with an electron. The two latter seeds are fed into a dedicated tracking algorithm, that takes into account the effect of photon emission in the tracker in the electron trajectory, using Gaussian Sum Filtering (GSF) [69]. These tracks are referred to as GSF tracks.

The ECAL clusters, the GSF tracks, and SC associated to tracker tracks are fed into the PF algorithm, which produces electron and photon candidates. Electron candidates are constructed from the association of a GSF track and an ECAL SC.

Reconstructed electrons are applied quality criteria to reject jets misidentified as leptons, keeping only those produced in the prompt decays of W, Z, τ and, very optimistically, H decays. Several variables are considered for this purpose. The $\sigma^{5\times5}_{i\eta i\eta}$ variable is defined as

$$\sigma^{5\times5}_{i\eta i\eta} = \frac{\sum (\eta_i - \eta)^2 w_i}{\sum w_i},$$

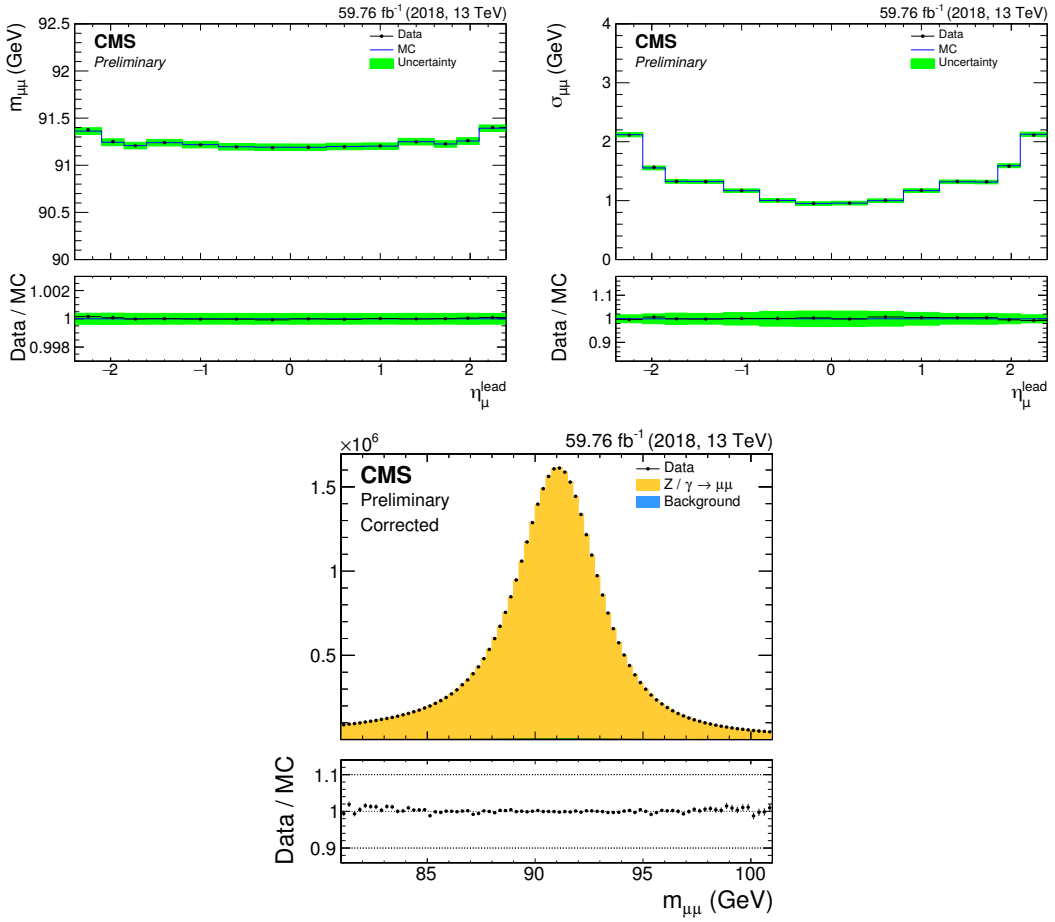


FIGURE 3.2: Momentum resolution in the CMS detector measured in $Z \rightarrow \mu^{\pm}\mu^{\mp}$ events. Top row: mean (left) and standard deviation (right) of the $m_{\mu^{\pm}\mu^{\mp}}$ distribution, as a function of the η of the leading lepton. Bottom row: $m_{\mu^{\pm}\mu^{\mp}}$ distribution in data, compared with the MC simulations. Data and simulations have been corrected using the Rochester methods. Taken from [5].

where the sum runs over the 5×5 matrix of ECAL clusters around the highest energy crystal of the glsSC, and w_i is a weight that depends logarithmically on the energy deposited in the crystal. Additionally, the ratio between the energy deposited in the HCAL and the ECAL, H/E is also considered. These variables allow to discriminate genuine electrons from jets identified as electrons.

The distance in pseudorapidity between the ECAL SC and the extrapolated track, $\Delta\eta$ and $\Delta\phi$, the analogous quantity in the ϕ directions are also used. The η and ϕ positions of the SC are defined as the energy-weighed position of the clusters. Finally, another useful variable is the difference between the inverse of the SC energy and the inverse of the track momentum, $\frac{1}{E_{SC}} - \frac{1}{p}$. Additional criteria may be also applied to further suppress the photon conversion.

In this thesis, most of these variables are combined in different manners to identify electrons. In the tW analysis described in chapter 4, selections on each one of these variables are applied sequentially as outlined in table 3.1. The criteria are applied depending on the η of the electron supercluster, η_{SC}. This is denoted as the cut-based

tight identification. The remaining two analyses use these variables combined in a BDT discriminator, that improves the discrimination performance. The BDT is trained in DY+jets simulated samples using prompt electrons as signal and jets as background.

Several working points are defined, depending on the efficiency. Two working points are used in the $\mathrm{t\bar{t}H}$ multilepton analysis (see chapter 6), the loose and the WP-80 working points. The first aims for a very high efficiency, while the second aims for a 80% efficiency. Two working points are also used in the search for SUSY (chapter 5), for which the cut in the BDT score is tuned as a function of the p_{T} and η_{SC} to achieve the optimal performance

	$\lvert\eta_{SC}\rvert \leq 1.479$	$\lvert\eta_{SC}\rvert > 1.479$
$\sigma^{5\times5}_{i\eta i\eta}$	< 0.00998	< 0.0292
$\Delta\eta$	< 0.00308	< 0.00605
$\Delta\phi$	< 0.0816	< 0.0394
H/E	< 0.0414	< 0.0641
Rel. PF isolation	< 0.0588	< 0.0571
$\lvert\frac{1}{E_{\mathrm{SC}}} - \frac{1}{p}\rvert$	< 0.0129	< 0.0129
Expected missing inner hits	≤ 1	≤ 1
Conversion veto	yes	yes

TABLE 3.1: Definition of the electron cut based tight identification criteria. Criteria are applied depending on the η_{SC} of the electron candidate. Isolation is described in section 3.3.

3.3 Lepton isolation

In many cases, it is necessary to discriminate between leptons produced in prompt decays from those coming from other sources. These sources may be dominated by genuine leptons produced in heavy flavor decays, for which usual identification variables are not enough. Since non-prompt leptons are usually produced inside jets, the detection of other objects surrounding the lepton candidate can be used to discriminate them. Isolation variables are used for that purpose, and built considering a cone in ΔR around the reconstructed lepton. The PF relative isolation is then computed as

$$I = \sum p^{i}_{\mathrm{T}} / p^{\mathrm{lep}}_{\mathrm{T}},$$

where the sum runs over all the PF candidates with a p^{i}_{T} included in the cone size and coming from the primary vertex, and $p^{\mathrm{lep}}_{\mathrm{T}}$ corresponds to the p_{T} of the lepton. While the charged particle component of the isolation is very well measured due to the tracking system, additional corrections are needed in the neutral part, composed by neutral hadrons and photons, since the contribution from pile-up cannot be directly subtracted, because no tracking information is available to discriminate between the different primary vertices. Two different approaches are followed: the $\Delta\beta$ corrections, and the effective area corrections.

The $\Delta\beta$ corrections are applied to the neutral component of the isolation by computing the contribution from charged hadrons coming from pile-up in the cone. This quantity is scaled by a factor of 0.5, corresponding to the ratio between charged and neutral particle production rate at inelastic pp collisions, as observed in simulations. The relative isolation is then computed as

$$I = \left(\sum_{\text{ch. had.}} p_{\mathrm{T}}^{i} + \max \left(\sum_{\text{ph.}} p_{\mathrm{T}}^{i} + \sum_{\text{neu. had.}} p_{\mathrm{T}}^{i} - 0.5 \sum_{\text{ch. had. (pile}-\text{up)}} p_{\mathrm{T}}^{i} \right) \right) / p_{\mathrm{T}}. \quad (3.1)$$

The effective area correction is similar to the jet areas method used to subtract the pile-up contribution from jets [70]. In this case, the pile-up correction is assumed to be $p_{\mathrm{T}}^{\mathrm{PU}} = \rho A_{eff}$. ρ is the average density of an event, defined as the median of the energy density of particles within the area of any jet, reconstructed with a R parameter of 0.6. A_{eff} corresponds to the dependence of isolation as a function of the number of reconstructed vertices in the collision.

Different isolation cone sizes are used depending on the object and on the analysis. By default, a fixed cone size of 0.4 is used for muons, while a 0.3 cone size is used for electrons. However, for several analyses it is interesting to consider a variable cone size that shrinks as a function of p_{T}. This allows to recover efficiency to topologies in which the leptons are produced in the decay of boosted objects, and are emitted colinearly with the rest of the decay products. In those cases, mini-isolation is considered, for which a cone size with a radius between 0.05 and 0.2, depending on the lepton p_{T} is used.

In this thesis, both fixed cone size and mini-isolation are considered in the different analyses. Muon fixed-cone isolation is corrected with $\Delta\beta$ corrections, while mini-isolation and electron fixed cone isolations are corrected with the effective areas approach.

3.4 Prompt-lepton MVA identification

The three analyses described in this thesis use the presence of two or more leptons in the final state as a signature to trigger and identify the presence of interest. These signal leptons are usually discriminated from those coming from non-prompt sources using isolation variables. However, many analyses still have a significant contribution from non-prompt sources in their signal regions. This is the case of the $\mathrm{t\bar{t}H}$ multilepton analysis, fully described in chapter 6, that looks for a rare SM signature, comparable to the rate of non-prompt leptons to pass the usual isolation requirements.

To enhance the separation power between prompt and non-prompt leptons, this analysis makes use of a multivariate discriminator that takes several aspects of the leptons into account. This method has been used in several searches for supersymmetry and Higgs measurements [14, 71]. A retraining of this method was used for recent $\mathrm{t\bar{t}Z}$ measurements [72] and the tZq observation [73]. During my thesis, I maintained and improved this discriminator, and tuned it to achieve the best performance for the $\mathrm{t\bar{t}H}$ multilepton analysis.

The discriminator is a BDT trained in simulated events, separately for electrons and muons. Two trainings for each lepton flavor are done in simulations reflecting the detector conditions during the 2016 and 2017 data taking, to account for the different detector geometry and pile-up conditions. No significant changes were present between 2018 and 2017, so the training performed in 2017 simulations is used for 2018 data and simulation.

Signal leptons are leptons coming from the prompt decay of a W boson or a τ in a $\mathrm{t\bar{t}H}$ simulated sample. Even if the impact parameter of τ decays is slightly different from that of prompt W decays, it is important to keep them as signal, in order not to lose performance when considering $\mathrm{H} \to \tau\tau$ decay in the $\mathrm{t\bar{t}H}$ measurement. Backgrounds are leptons in a semileptonic $\mathrm{t\bar{t}}$ sample that are not matched to a prompt W boson, τ decay or any other source of prompt leptons.

Three types of variables are considered: kinematic variables, isolation variables, b tagging variables, identification variables, and impact parameter variables. Several of these variables are defined considering the jet associated to the lepton. This jet is the one in which the PF candidate that conforms the lepton is included. Only jets with p_T greater than 15 GeV are considered. If such jet does not exist, the variables are assigned a sentinel value, that is defined in the list of input variables below.

The variables that the prompt-lepton MVA uses as an input are the following.

- Lepton p_T and $|\eta|$.
- Charged component of the mini-isolation variable, defined as $I_\mathrm{l}^\mathrm{charged} = \sum_\mathrm{charged} p_\mathrm{T}$.
- Neutral component of the mini-isolation variable, corrected for PU effects with the effective areas method, defined as $I_\mathrm{l}^\mathrm{neutrals} = \max\left(0, \sum_\mathrm{neutrals} p_\mathrm{T} - \rho\,\mathcal{A}\,\left(\frac{R}{0.3}\right)^2\right)$.
- Lepton-to-jet p_T ratio variable, $p_\mathrm{T}^\mathrm{ratio}$: the ratio of the transverse momentum of the lepton to the transverse momentum of the nearest jet, $p_\mathrm{T}^\mathrm{l}/p_\mathrm{T}^\mathrm{jet}$. If no jet associated to the lepton is present, this variable is set to $\frac{1}{1+\mathrm{I_{rel}}}$. Related to this variable is the jet relative isolation variable, defined as $\frac{1}{p_\mathrm{T}^\mathrm{ratio}} - 1$, that is related to the lepton isolation.
- Lepton relative p_T variable: the component of the lepton momentum in direction transverse to the jet, $p_\mathrm{T}^\mathrm{rel} = p_\mathrm{l} \sin\theta$, where θ denotes the angle between the lepton and jet momentum vectors. If no jet associated to the lepton is present, this variable is set to zero.
- Jet b-tagging score: the discriminant value of the Deep Jet b-tagging algorithm of the matched jet, described in section 2.3.3. When such a jet is not present, the variable is set to zero.
- Jet charged constituents: the number N_charged of charged particles within the matched jet. Tracks associated to those particles must be within a $\Delta R < 0.4$ of the lepton and are required to come from the primary vertex to enter the counting. Minimal track quality, p_T and impact parameter criteria are also applied. If such a jet does not exist this variable is set to zero.

- Impact parameters: the impact parameters of the lepton track with respect to the PV in the transverse direction, d_{xy}, and in the direction along the beam direction d_z. The logarithm of these variables is used as the input, to enhance the training of the BDT.
- Significance of the impact parameter: the impact parameter, in three dimensions, of the lepton track with respect to the PV, divided by its uncertainty, which corresponds to its significance d/σ_d.
- Electron MVA ID discriminator: the output of the BDT that separates electrons from jets, described in section 3.2. This variable is only used in electrons.
- Muon segment compatibility: the compatibility of track segments in the muon system with the pattern expected for a minimum ionizing particle. This variable is only used by muons.

The performance of this discriminator is evaluated in $\mathrm{t\bar{t}}$ simulated events, using prompt electrons and muons as signal and non-prompt leptons as background. It is compared to the selections used in the WW cross section analysis [74] and the search for $\mathrm{t\bar{t}t\bar{t}}$ [75], both by the CMS Collaboration. Non-prompt leptons contribute significantly to the two analyses, however the source and p_{T} spectrum of these non-prompt leptons is not necessarily the one of the $\mathrm{t\bar{t}H}$ multilepton analysis. Therefore they are shown here as benchmark points to compare the performance of the prompt-lepton MVA with other approaches with rectangular cuts.

The WW measurement uses the tight muon and tight cut-based electron identification criteria, with tighter cuts on the impact parameter of the leptons with respect to the primary vertex, and relative isolation using fixed cone sizes. The $\mathrm{t\bar{t}t\bar{t}}$ analysis uses medium muon identification criteria and a custom working point of the electron MVA-based identification criteria, with tight impact parameter cuts. Rectangular cuts are also imposed in the mini-isolation, p_{T}^{rel} and p_{T}^{ratio}.

Efficiency for signal and background muons passing the loose identification criteria and signal and background reconstructed electrons, all them with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$, and are shown in figure 3.3. The $\mathrm{t\bar{t}H}$ multilepton curve includes the requirements used in the $\mathrm{t\bar{t}H}$ analysis described in chapter 6, for different cuts of the prompt-lepton MVA. The approach followed in the $\mathrm{t\bar{t}H}$ multilepton analysis reduces the background acceptance rate by more than a factor of 5 for muons and almost a factor of 2 for electrons. This shows significant gains can be obtained by using MVA techniques in lepton identification.

3.5 Lepton efficiency characterization

The algorithms described in the previous section allow to select muons from specific sources with very high purity and efficiency. However, once the suitable selection is chosen, the muons passing it must be properly characterized. Lepton efficiency is typically calibrated by correcting the efficiency predicted by simulated events with the one measured in data. This procedure brings an uncertainty that can be significant. These uncertainties are of the order of 1-2% per each lepton, however they can be

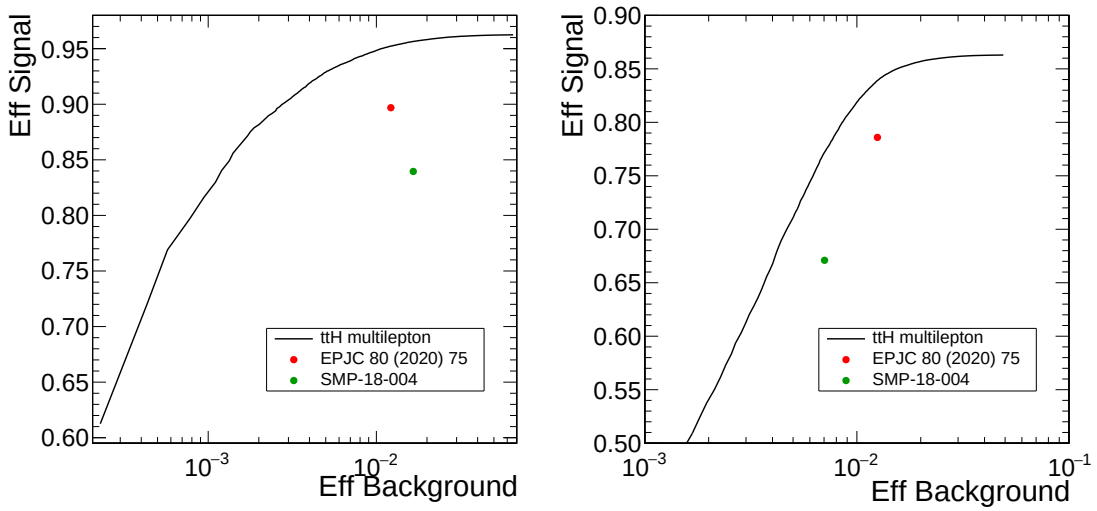


FIGURE 3.3: Signal lepton efficiency as a function of the non-prompt lepton efficiency for muons (left) and electrons (right). Several discriminators are considered: the $\mathrm{t\bar{t}H}$ multilepton selection for different working points of the prompt-lepton MVA, and the selection used in the $\mathrm{t\bar{t}t\bar{t}}$ analysis (red) and WW analysis (green).

higher for selections that are dependent on the topology of the event, as isolation variables or the prompt-lepton MVA described in the previous section. For example, the uncertainty associated to this quantity in the tW analysis described in chapter 4 is among the leading ones. Additionally, this uncertainty was the dominant in the $\mathrm{t\bar{t}H}$ measurement in the analysis performed with 2016 data [14].

This section provides a comprehensive review of the methods employed to measure these efficiencies, focusing on the different sources of uncertainty that affect the measurement. An emphasis is put in isolation, since all the analyses described in this thesis make use of it to discriminate prompt leptons from other sources, and is more sensitive to the topology of the event.

Lepton efficiency calibrations are usually performed in two steps. In the first one, the DY+jets events are used as a standard candle to measure the lepton efficiency using the tag-and-probe method [49]. Then, efficiencies measured in data are compared to the prediction by DY+jets simulations to obtain a set of scale factors. In the second step, simulated events in the relevant regions of the analysis are corrected by these scale factors.

Each one of the steps has associated systematic uncertainties stemming from the several assumptions that are made.

3.5.1 Measurement in DY events

In the tag-and-probe method, the efficiency is measured in DY+jets events. The method is described with more detail in reference [49]. In the following, we will mostly focus on the systematic uncertainties.

The method profits from the fact that DY+jets events show a resonance in the Z boson mass, and this can be used to discriminate from the various backgrounds. Events are

collected using single lepton trigger. The lepton firing the trigger, referred to as tag, is required to pass quality criteria to reject backgrounds. Another lepton, the probe, is required to be reconstructed passing looser selection criterion, which is going to be the denominator of the efficiency measurement. Efficiency for the probe to pass a given selection can then be obtained by counting the number of events for which the probe pass or fail the selection.

However, possible backgrounds in the measurement must be subtracted, since their efficiency may be significantly smaller than that of signal, biasing the measurement. To account for this, a fit is performed to the m_{ll} distribution to disentangle the signal peak from the non-resonant background. The integral of the signal in the passing and failing categories can then be used. However, a model for both signal and background must be chosen.

This choice introduces a source of systematic uncertainty, as slightly different results could be obtained with a different signal-background model. Different approaches are followed in the electron and muon groups of the CMS Collaboration. The muon group uses the sum of two Voigtian functions as signal hypothesis, and an error function with a multiplicative exponential term at high m_{ll} from background [5]. The efficiencies for electrons are measured using a template constructed using DY+jets simulated events for signal, and an analytical function for background [76].

Systematic uncertainties are considered by performing different variations of these models, and also by varying the signal-background composition of the sample. The former is performed considering different analytic functions or different event generators when the templates are obtained from simulations. The latter is done by considering different selections to the tag lepton or by changing the m_{ll} range that is fitted.

Other subtle effects can affect the measurement. Leptons may radiate photons as they are produced. This is effect is more prominent in leptons with low p_T, for which a secondary peak is observed in the m_{ll} distribution. When measuring isolation efficiencies, this effect can be significant, as these leptons are usually less isolated because of the presence of the radiated photon itself. This radiation can be recovered and reassigned into the lepton 4-momentum, or the peak can be fitted and accounted for as signal or background, depending on which analysis the measurement is going to be made.

3.5.2 Phase space extrapolation

The following step in the correction for the data and MC discrepancies is to apply the scale factors derived using the tag-and-probe method in the measurement regions. This application is only valid if the distribution of the identification and isolation variables employed in the lepton selection are the same in leptons produced in the Z boson decay and signal leptons present in the measurement regions.

Identification variables are usually robust enough to be independent from the topology of the event. There are exceptions to this, since in cases in which two muons are produced colinearly to each other the reconstruction may fail. Analyses treated in the thesis are not very sensitive to this problem, since leptons are required to be

well separated. However, isolation variables can be affected by the presence of relatively close-by jets, yielding to a bias due to the application of the scale factor in a different region to what has been measured. This effect is referred to as phase space extrapolation.

Phase space extrapolation uncertainty can affect analyses in different manners. Analyses performed in a Drell-Yan-like topology the bias introduced by this effect are more reduced, while for analyses performed in boosted topologies, in which different objects can be merged with the lepton, this effect can be large. This effect be therefore assessed in a analysis-by-analysis or topology-by-topology basis, and also depending on the specific lepton selection.

3.5.3 Measurement of the phase space extrapolation in the $\mathrm{t\bar{t}H}$ selection

The lepton selection used in the $\mathrm{t\bar{t}H}$ analysis is particularly sensitive to the topology of the event for two reasons. First, the prompt-lepton MVA requirements applied make use of isolation and variables related to close-by jets. Second, the analysis profits from a very pure selection requirements, so quite strict criteria are applied. This could enhance any phase space extrapolation effect.

As described in chapter 6, several working points are used in the analysis. Because of this, the efficiency is measured in two steps: the loose selection efficiency and the tight selection efficiency in leptons passing the loose selection are measured separately. The first step is not expected to have a significant phase space dependence, since only a mild selection requirement is applied at that level. The effect of the phase space extrapolation is accounted for in the tight selection efficiency measured in loose leptons.

The approach followed is the following one. Since signal regions require higher jet multiplicity than DY+jets events, the efficiencies are measured in a region enriched in $\mathrm{t\bar{t}}$ events. This allows to check effect of additional jets in the event in a region with a reasonably large event rate. The nominal corrections in the analysis are however determined with the usual tag-and-probe measurement in DY+jets events, and the efficiency measurement in $\mathrm{t\bar{t}}$ is used to estimate the systematic uncertainty due to the phase space extrapolation.

Events are collected using single lepton triggers. A reconstructed $\mathrm{e\mu}$ pair must be present in the event. When measuring muon efficiencies, the muon is required to pass the loose $\mathrm{t\bar{t}H}$ selection (probe), and the electron is required to pass the tight $\mathrm{t\bar{t}H}$ selection (tag). Conversely, when measuring electron efficiencies, the electron and the muon are required to pass the loose and tight $\mathrm{t\bar{t}H}$ requirements, respectively.

In order to have a selection enriched in $\mathrm{t\bar{t}}$ events, events are also required to have at least two reconstructed jets, out of which at least one of them must be b-tagged, following the selections described in chapter 6. Similarly to the $\mathrm{t\bar{t}H}$ multilepton analysis, contribution low mass resonances is rejected by requiring $m_{\mathrm{ll}} > 12\,\mathrm{GeV}$ for all pairs of loose leptons in the event.

Efficiency can be measured by comparing the number of events for which the probe passes and does not pass the tight selection criteria. However, the presence of non-prompt or misidentified leptons in the region must be taken into account. This presence can be very significant in events in which the probe lepton fails the selection. In order to do this, efficiency is measured after the contribution from non-prompt leptons has been subtracted from the numerator and denominator:

$$\epsilon = \frac{N^{\text{data}}_{\text{passing}} - N^{\text{non-prompt}}_{\text{passing}}}{N^{\text{data}}_{\text{total}} - N^{\text{non-prompt}}_{\text{total}}}. \tag{3.2}$$

The estimation of non-prompt events is then carried out as usually done in several top physics analyses [6–8]. This approach profits from the fact that the sign of non-prompt leptons are usually uncorrelated from the sign of the prompt lepton in semileptonic $\mathrm{t\bar{t}}$ events, which are the main source of non-prompt leptons in this region. Then, the contribution of non-prompt leptons can be written as

$$N^{\text{non-prompt}}_{OS} = \left(N^{\text{data}}_{SS} - N^{\text{prompt}}_{SS}\right) \times R^{OS}_{SS}, \tag{3.3}$$

where the subscript SS (OS) denotes the number of events with same (opposite) leptons. R^{OS}_{SS} is defined as the ratio between opposite-sign and same-sign events in processes with non-prompt leptons. This quantity and the subtraction of processes with prompt leptons is taken from simulated events.

Since the final aim is to measure at which level the scale factors obtained using the tag-and-probe measurements are suitable to be applied in $\mathrm{t\bar{t}}$ events, the measurement of the efficiency in data is compared to the efficiency predicted by $\mathrm{t\bar{t}}$ simulations in that region, after it has been corrected by the scale factors measured in tag-and-probe measurements.

The measurement is performed separately for electrons and muons as a function of p_{T}. Unfortunately the event rate in the same-sign sideband is not large enough to make a measurement over η and p_{T} simultaneously. The results of the measurement are shown in figure 3.4, separately for electrons and muons, as a function of p_{T}, and separately for each year of the data taking. The difference between the two is used as the systematic uncertainty for the tight lepton selection in the $\mathrm{t\bar{t}H}$ multilepton analysis described in chapter 6.

3.6 Hadronic τ reconstruction

Hadronically decaying τ fermions, τ_{h}, are reconstructed using the hadron-plus-strips algorithm [77]. This algorithm reconstructs τ_{h} in its various decay modes: one-prong τ decays ($h^{\pm}$, $h^{\pm} + \pi^0$, $h^{\pm} + 2\pi^0$), two-prong τ decays ($h^{\pm}h^{\mp}$, $h^{\pm}h^{\mp} + \pi^0$, $h^{\pm}h^{\mp} + 2\pi^0$) and three-prong τ decays ($h^{\pm}h^{\mp}h^{\pm}$, $hh^{\pm}h^{\mp}h^{\pm} + \pi^0$), where h denotes a charged pion or kaon. Neutral pions decay into photons that may convert into $\mathrm{e^+e^-}$ pairs with a high probability. Therefore pions are reconstructed by clustering the photon and electron constituents of the jet. τ_{h} candidates are constructed by combining the reconstructed pions with the charged components of the jet, according to the modes described.

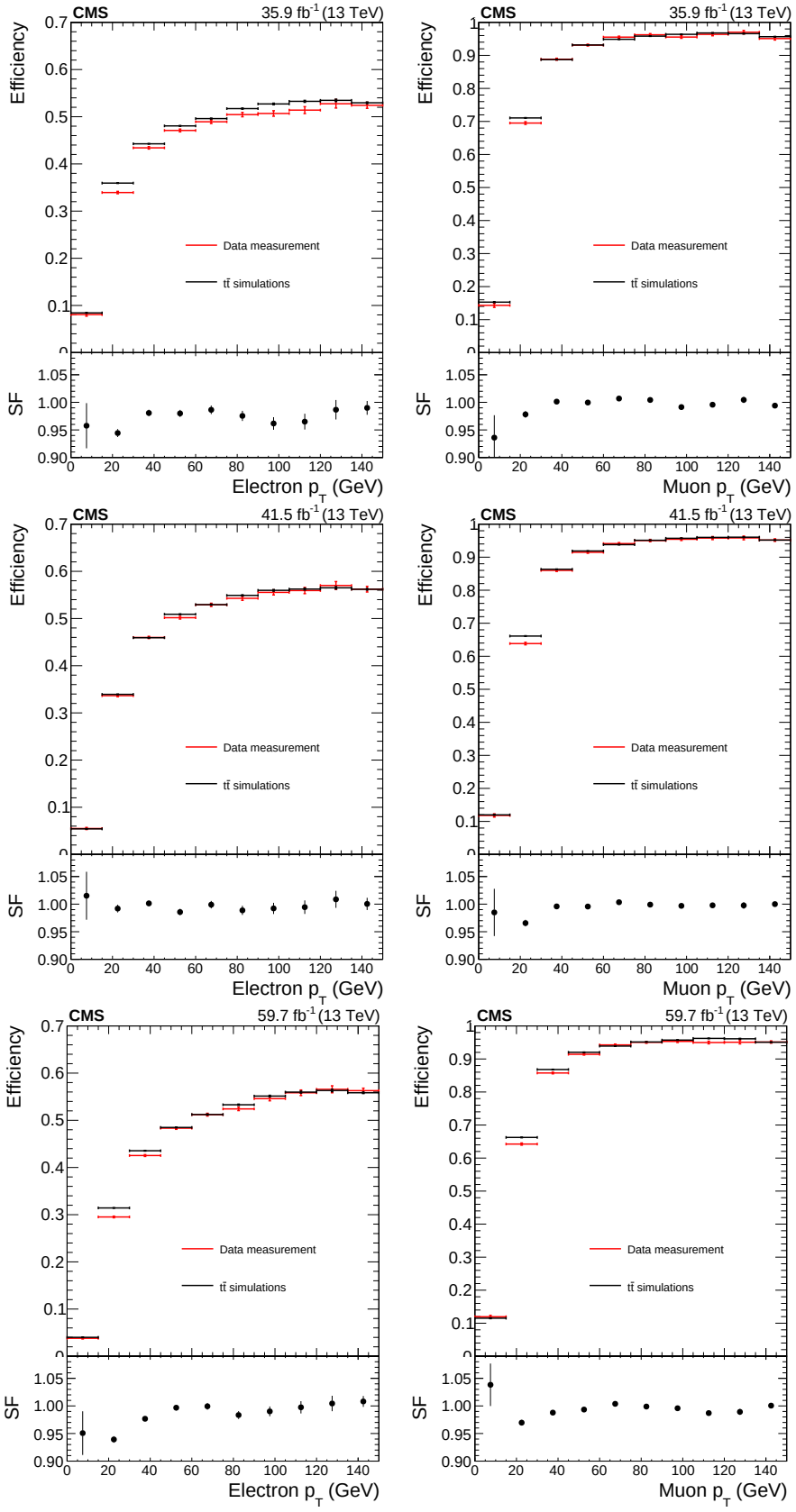


FIGURE 3.4: Comparison of the in-situ efficiency measurement in $t\bar{t}$ data events (red) and the prediction of simulations corrected by DY+jets scale factors (black). Left column shows the results for electrons and right column for muons. Each row shows the results for each year of the Run 2 data taking. The ratio panel shows the quotient between the two.

τ_h are then discriminated from quark and gluon jet background, muons and electrons using a convolutional Deep Neural Network (DNN) [78], referred to as "Deep Tau v2.1". This algorithm combines the usage of high-level features of the reconstructed τ_h and low level information from the particle flow candidates within the τ_h isolation cone. This approach achieves a significantly better performance than other methods used in CMS. The DNN has been trained using $t\bar{t}$ and W+jets simulated samples. The outputs of the DNN are three nodes discriminating against jets, electrons and muons. Different working points are defined based on the score of each node: 8 in the discrimination against jets, 4 against muons and 8 against electrons. Working points in the discriminator against jets are labeled very-very-very, very-very, very loose, loose, medium, tight, very tight and very-very tight, according to the required purity. The performance of this algorithm, compared to other algorithms used, described in [79], are shown in figure 3.5, in which the working points are also shown.

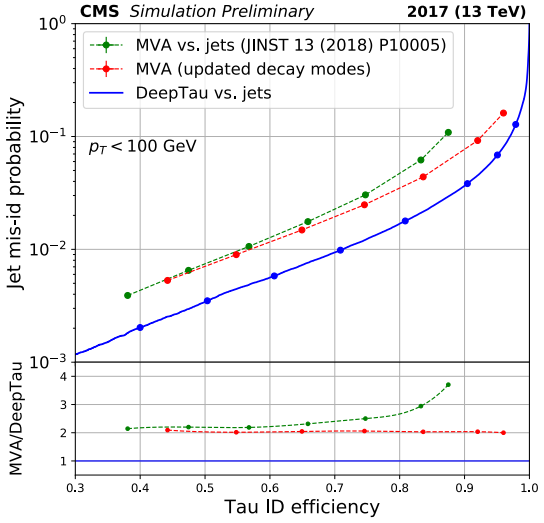


FIGURE 3.5: Performance of the Deep Tau tagger in simulated events, comparing to other discriminators. The jet misidentification probability is shown as a function of the efficiency. Taken from [78].

4

Measurement of tW production at CMS

In this chapter, measurements are presented of the single top production in association with a W boson (tW) performed at CMS with pp collision data at $\sqrt{s}$ = 13 TeV collected by the CMS experiment during the year 2016. This dataset allowed for high precision measurement of this production mode, and opens the way for measurements of this process in other kinematic regimes.

This chapter is organized as follows. In section 4.1, the measurements of tW production will be motivated in the general context of measurements of top quark properties at the LHC. A review on the current status of theory calculations for these processes is also shown. The following sections will cover the measurements of the inclusive tW production cross section and as a function of several kinematic variables. The event selection for these measurements is described in section 4.2, and the MC models used to estimate the signal and background contribution to these regions are described in section 4.3. Section 4.4 reviews the systematic uncertainties that are taken into account in these measurements. Section 4.5 shows the methodology used in the measurement for the inclusive tW production cross section and the results obtained, while section 4.6 covers the differential measurements. The chapter is closed with the conclusions in section 4.7.

4.1 Top physics and the tW process

The study of the properties of the top quark is, by its own right, one of the main fields of study in particle physics. The top quark was predicted by Kobayashi and Maskawa to explain Charge Parity (CP) violation [80], but it was not until 1995, when it was observed by the D0 and CDF experiments [81, 82]. With a mass, m_{t} of 172.9 $\pm$ 0.4 GeV, as measured by the CMS and ATLAS experiments at LHC and the experiments at Tevatron [17], it is the heaviest fundamental particle in the SM.

Its large mass has two implications that motivate the studies of top quark physics. Firstly, due to the large mass difference between the top quark and its decay products, and despite the fact that this decay is mediated by the weak force, the top quark decays promptly, before any hadronization can take place. Because of this, it is the only quark that can be studied in a free state, and one of the few available probes for perturbative QCD (pQCD).

Additionally, its high mass is a direct consequence of its large coupling to the Higgs boson. This Yukawa coupling is the only natural one in the SM. Because of this, the

top quark plays a special role in the spontaneous breaking of the EWK symmetry. This behavior also implies that the top quark induces large radiative corrections to the mass of the Higgs boson, and may also affect the stability of the vacuum at high energy scales [83], as described in chapter 1. Because of this, the top quark also plays an important role in many BSM models that aim to enforce the naturalness of the SM. For instance, most scenarios of SUSY predict a scalar top with a mass in the electroweak scale to cancel the large radiative corrections to the Higgs mass.

4.1.1 Top quark production modes at the LHC

The top quark is produced in different modes in pp collisions at the LHC energy scales, each one of which has a different production rate and allow to study different physical aspects.

4.1.1.1 Top quark pair production

The dominant production mode is the production of quark anti-quark pairs ($\mathrm{t\bar{t}}$), for which some of the leading order diagrams are shown in figure 4.1. The cross section for this production in pp and $\mathrm{p\bar{p}}$ collisions as a function of $\sqrt{s}$ is shown in figure 4.2. This high cross section allows to measure the properties of this process very precisely. The production was dominated by quark-anti quark annihilation ($\mathrm{q\bar{q}}$) in $\mathrm{p\bar{p}}$ collisions at the Tevatron energy scales, while at the LHC the gluon-gluon fusion (gg) dominates, with a contribution of 85%, followed by $\mathrm{q\bar{q}}$ interactions.

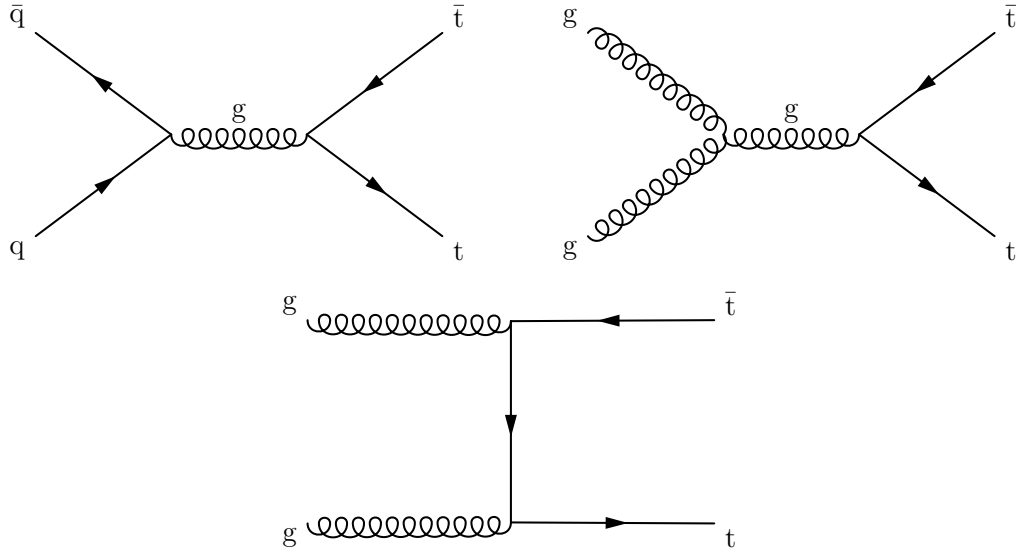


FIGURE 4.1: Leading order Feynman diagrams for $\mathrm{t\bar{t}}$ production in pp or $\mathrm{p\bar{p}}$ collisions. Diagram in the top left shows the $\mathrm{q\bar{q}}$ production mode, and the remaining two, gluon fusion.

Top pair production cross section has been calculated at NNLO accuracy in series of α_s for all the production modes: qq (including qq, $\mathrm{q\bar{q}}$, $\mathrm{q'q}$ and $\mathrm{q'\bar{q}}$), qq and gg. To take into account the radiation of additional soft partons at all orders, which lead to

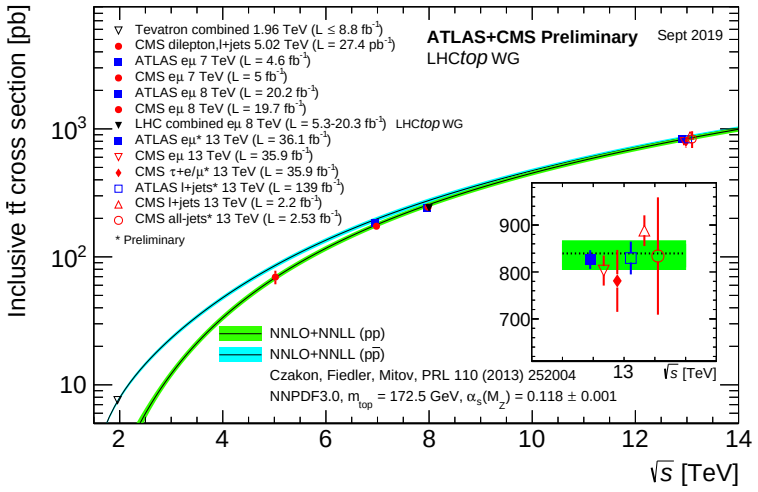


FIGURE 4.2: Measurements of $t\bar{t}$ cross sections in pp and $p\bar{p}$ collisions as a function of $\sqrt{s}$ by the Tevatron and LHC experiments, together with the Next-to-next-to-leading Order (NNLO)+Next-to-next-to-leading Logarithm (NNLL) prediction. Taken from [84].

large logarithmic terms, these are resummed at NNLL. At $\sqrt{s} = 13$ TeV, and taking $\mu_R = \mu_F = m_t$, the NNLO+NNLL prediction for the $t\bar{t}$ cross section in pp collisions is 832 pb, using the CT14nnlo PDF [85] to model the structure of the proton. The uncertainty on the calculation due to missing higher order calculations is assessed by considering independent variations of the μ_R and μ_F scales between $m_t/2$ and $2m_t$, leading to an uncertainty between 2-3%. The uncertainties associated to the modeling of the structure of the proton are also propagated to the cross section, leading to an uncertainty of 4% [86].

Predictions of the differential $t\bar{t}$ cross section are available in the form of various high precision calculations. These predictions are necessary for many phenomenological studies on proton collider physics, including the study of top quark properties, but also searches for BSM physics or Higgs physics, among others. Most of these predictions rely on the Narrow Width Approximation (NWA), in which the width of the top quark is assumed to be zero, allowing to factorize production and decays of the top quark pair. There are automatic MC generators that generate $t\bar{t}$ events at NLO accuracy, which allow to predict a broad variety of observables [28, 87–89]. Additionally, calculations at NNLO accuracy for stable top quarks exist for a few set of observables [90]. For decayed top quarks, the best prediction at fixed order include the combination of NNLO production and NNLO decay calculations, that are usually quoted as $\hat{\text{N}}$NLO [91].

Several observables are sensitive to off-shell effects, that are not taken into account under the NWA, such as those aiming to reconstruct the top mass [92]. Fixed order calculations at NLO exist for this process, including the top quark decays. A full NLO prediction matched to parton showers is not trivial to obtain, and has only been obtained for the dilepton channel [93, 94]. This aspect, of great importance when discussing the interference between tW and $t\bar{t}$ processes, is further discussed in section 4.1.2.

The top pair production mode led to the discovery of the top quark at the Tevatron collider by the D0 and CDF Collaborations [81, 82]. Both at Tevatron and at the LHC, this process is studied depending on the different decays of the top quark. The top quark can either decay either hadronically ($\mathrm{qq'b}$) or to a triplet of lepton, neutrino and b-quark (lvb). These decays are always mediated by a W boson and, since the V_{tb} element dominates over V_{ts} and V_{td}, a b-quark appears in the decay in most of the cases. Measurements of this process are performed aiming for the different decay modes of the top quark.

The dilepton decay mode provides a clean signature with two leptons in the final state, two jets due to the two b quarks produced in the top decays, and missing transverse momentum due to the two neutrinos. This is a clean topology that can be easily triggered on because of the presence of the two leptons, keeping a reasonably large fiducial region. Backgrounds in this topology include tW production, which is an irreducible background, and Drell-Yan events, that can be suppressed by requiring jets, b tagged jets, vetoing the Z boson peak or studying events with different flavor leptons.

The single lepton final state corresponds to cases in which one of the top quarks decays leptonically and the other one hadronically. This signature, that has a larger background contamination from multijet production, has a higher branching ratio and allows for a complete unambiguous reconstruction of the $\mathrm{t\bar{t}}$ system, due to the presence of only one invisible particle in the final state. This allows for precise differential measurements as a function of the top kinematics.

Finally, the fully hadronic channel is the one with the highest cross section. Although all the objects in the final state are visible, the overwhelming multijet background makes precision measurements in this channel challenging.

When I joined the CMS Collaboration, I contributed to measurements of the inclusive cross section in pp collisions at a $\sqrt{s}$ at 7 and 8 TeV in the dilepton channel [7], performing a cross-check of the main analysis, following the signal extraction used in [95] by the ATLAS Collaboration. This measurement had a precision comparable to that of the main CMS analysis, and was more precise than the current state-of-the-art calculations. I also contributed to the first measurement of the inclusive $\mathrm{t\bar{t}}$ production cross section performed in the Run 2 of the LHC at $\sqrt{s} = 13$ TeV [8]. This result was one of the first results in the LHC establishing the validity of the SM at this energy scale, that was then accessible for the first time.

The inclusive production cross section was then measured in this channel with a precision of 4% by the CMS Collaboration [6] and 2.5% by the ATLAS Collaboration [96], both compatible with the state-of-the-art prediction, and with a higher precision.

Additionally, CMS and ATLAS have performed precision measurements of differential observables in the single lepton and dilepton channels [96–101].

These measurements show a generally good agreement with respect to the NLO simulations. However these models fail to predict the top p_{T} spectrum, which could be partially explained by missing high orders in the calculations, as the discrepancy diminishes when considering higher order corrections [102].

Additionally, the production of top anti-top pairs also allows to measure the top mass, the α_s coupling [6], and is affected by the Yukawa coupling of the Higgs boson to the top quark [103].

4.1.1.2 Single top production

Top quarks can also be produced at the LHC in processes mediated by the electroweak interactions. In most of these production modes, the top quark is produced singly. At the LHC, the single top production is dominated by the so-called *t*-channel, *s*-channel and tW-channel. The production cross section for these processes is one order of magnitude lower than that of $\mathrm{t\bar{t}}$, since these processes are mediated by the electroweak interaction and partons from the sea of quarks are present in the initial state. Leading order diagrams can be seen for these processes in figure 4.3. These processes involve the tWb vertex. Contributions from the tWd and tWs vertices are present but their effect is negligible. Contributions from Flavor Changing Neutral Currents (FCNC) would also affect the production of some of these processes, but are highly suppressed in the SM. Because of this, they may be an excellent probe for BSM physics.

***t*-channel** *t*-channel production is the single top process with the highest cross section at the LHC, with a value of $217.0^{+6.6}_{-4.6}$ (scale) $\pm$ 3.5 (PDF + α_s) at NLO in the 5FS at $\sqrt{s}$ =13 TeV [104]. Measurements of this process can help constrain the PDFs of the proton by performing inclusive and differential cross section measurements. Further constrains can be achieved with measurements of the ratio between the production rate of t and $\mathrm{\bar{t}}$ processes. Finally, since top quarks are expected to be almost 100% polarized through the *t*-channel, measurements of this process are sensitive to any BSM contribution that may alter the Lorentz structure of the coupling between top, b quarks and W bosons.

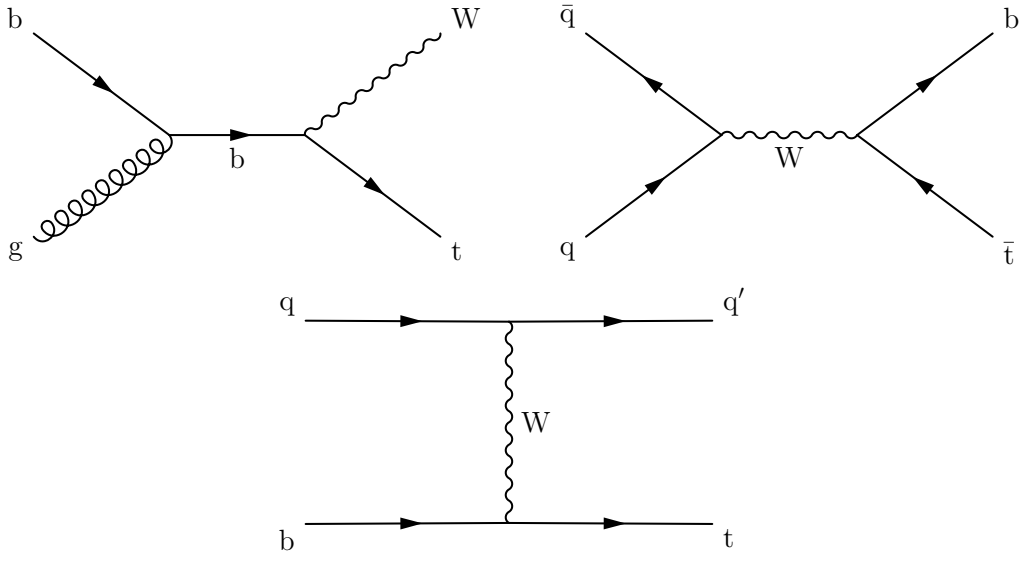


FIGURE 4.3: Leading order diagrams for the main modes of single top production: tW-channel (top left), *s*-channel (top right) and *t*-channel (bottom).

tW-channel The second production channel in terms of production rate at the LHC, is the associated production of a top quark and a W boson, with an expected rate of 71.7 $\pm$ 1.8 (scale) $\pm$ 3.4 (PDF) pb at $\sqrt{s} = 13$ TeV, calculated at approximate NNLO [105]. The measurement of this process is also a probe to the V_{tb} coupling. However, one of the most important features of this process is its interference with $\mathrm{t\bar{t}}$ production at NLO. Therefore this process is only well defined at Born level. This process is the main topic of this chapter and is described in greater detail in section 4.1.2.

***s*-channel** The s-channel production mode is very rare in the LHC, due to the presence of quarks from the sea of quarks in the initial state, and the searches performed at the time this thesis was written have not allowed for an observation of the process in pp collisions. However, this process was observed at the Tevatron, as its production cross section in $\mathrm{p\bar{p}}$ collisions is higher. The process is sensitivity to new particles such as a charged Higgs boson or an extra W' production.

4.1.1.3 Rare processes

Besides the main top quark production modes described above, there are other modes that have a smaller cross section but their study remains interesting as they are probes to the couplings of the top quark to other particles and, potentially, to BSM physics. Thanks to the amount of luminosity delivered by the LHC during Run 1 and, particularly, Run 2, the study of these processes has become possible in the past years.

These processes include the associated production of a top anti-top quark pair with a boson ($\mathrm{t\bar{t}H}$, $\mathrm{t\bar{t}Z}$, $\mathrm{t\bar{t}W}$ and $\mathrm{t\bar{t}}\gamma$), the production of a single top quark associated with a Z boson, tZq, with a Higgs boson, tH, and with a Higgs and a W boson, tHW. $\mathrm{t\bar{t}H}$, tH and tHW will be covered in greater detail in chapter 6. Similarly, $\mathrm{t\bar{t}W}$ and $\mathrm{t\bar{t}Z}$ will also be covered in that chapter, as they represent the largest irreducible backgrounds in measurements of $\mathrm{t\bar{t}H}$, tH and tHW production.

tZq production has been observed for the first time by the ATLAS [106] and CMS [73] Collaborations in collisions at $\sqrt{s} = 13$ TeV. This process includes contribution from the coupling of the top quark to the Z boson, but also of the WWZ coupling. Additionally, it is also sensitive to FCNC contributions that may appear in BSM scenarios [107].

The study of $\mathrm{t\bar{t}t\bar{t}}$ production is very valuable because it allows to test calculations involving several orders of QCD. Additionally, this process is also affected by the Yukawa coupling of the top quark to the Higgs boson. Both ATLAS and CMS have performed searches for this process in collision data, however its tiny cross section of 12 fb makes the studies very challenging. Therefore, with the current amount of data collected, the precision of this studies is not enough to claim evidence for this process [75].

4.1.2 The tW process

The associated production of a single top quark with a W boson is the main topic of this chapter. This process is mediated by the electroweak interaction and is affected by the V_{tb} element of the CKM matrix. The process is trivially defined at Born level,

however it interferes with $\mathrm{t\bar{t}}$ production at NLO, and removal methods must be applied to obtain a consistent definition of this process at NLO. This aspect is discussed in section 4.1.3.

Evidence for tW production at 7 TeV was reported by the CMS and ATLAS experiments [108, 109] and its observation was achieved with $\sqrt{s} = 8$ TeV pp colisions by those collaborations [110, 111]. Both the ATLAS and CMS Collaboration performed measurements of this process in the dilepton channel, obtaining results consistent with the SM. At $\sqrt{s}$ = 13 TeV, the ATLAS Collaboration has performed a measurement of the inclusive cross section of the process with a luminosity of 3.2 fb^{-1} [112], and a differential measurement with 36.1 fb^{-1} [113], both compatible with the approximate NNLO predictions. The CMS Collaboration has measured this process in collisions at $\sqrt{s}$ = 13 TeV with collected luminosity of 35.9 fb^{-1}. The production cross section for this process has been measured inclusively [9] and differentially [10]. These measurements constitute the contribution of this thesis to the study of the tW process and are described in sections 4.2- 4.6.

The tW process can also be used to probe for new interactions at higher energy scales. The measurement done in [114], in the same topology and using similar analysis techniques to those of the analysis shown in this chapter, allows to put limits on the Wilson coefficients for additional terms in the SM Lagrangian.

4.1.3 Interference between $\mathrm{t\bar{t}}$ and tW

One of the most remarkable properties of the tW process is its interference with $\mathrm{t\bar{t}}$ events, which is described in this section. A more detailed description can be found in [115].

The interference occurs when computing observables at NLO accuracy in perturbation theory for the tW process. At NLO, the $\mathrm{t + W + b}$ final state is accessible, which coincides with the final state of the $\mathrm{t\bar{t}}$ process. This is shown in figure 4.4 in which diagrams associated to $\mathrm{t\bar{t}}$ production and tW production associated with an additional b quark are depicted. In the former, the two top quark lines are resonant, while in the latter only one of the two quarks is resonant. The contribution of the first kind of diagrams is larger than the second one, so this introduces a difficulty to the modeling of the tW process at NLO, since the NLO corrections are going to be typically larger than the Born ones.

Three approaches exist to remove the overlapping between doubly and singly-resonant contributions, and have a consistent definition of the tW process, separated from $\mathrm{t\bar{t}}$. The complete squared matrix element can be written as

$$|\mathcal{M}_{\mathrm{bb2l2}\nu}|^2 = |\mathcal{M}_{singly}|^2 + |\mathcal{M}_{doubly}|^2 + 2\Re(\mathcal{M}^*_{singly}\mathcal{M}_{doubly}), \tag{4.1}$$

where $\mathcal{M}_{singly\ (doubly)}$ represents the matrix element associated to the singly (doubly)-resonant diagrams.

In the Diagram Removal (DR) method [115], only the $\mathcal{M}_{singly}$ is kept, neglecting the contribution from interference terms and doubly-resonant diagrams. This approach

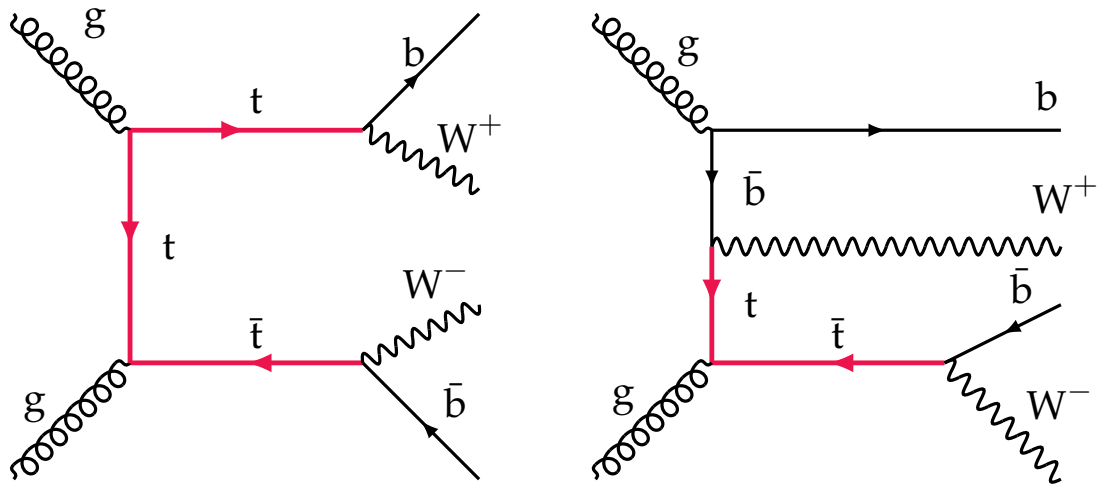


FIGURE 4.4: Interfering Feynman diagrams between $t\bar{t}$ and tW production. $t\bar{t}$ production (left) shows two resonant top quarks (doubly resonant), while tW production (right) shows only one (single resonant).

violates the gauge invariance, however, in practice, little dependence on the gauge choice is observed. An alternative approach, called Diagram Removal 2 (DR2) [116], the terms $|\mathcal{M}_{singly}|^2 + 2\Re(\mathcal{M}^*_{singly}\mathcal{M}_{doubly})$ are used to define the tW process. A third approach exists, called Diagram Subtraction (DS) [115, 117], in which an additional term, $\mathcal{M}_{DS}$ is added at the level of squared matrix element. This term is chosen so that $\mathcal{M}_{DS} - \mathcal{M}_{doubly}$ will vanish when $m^2_{bW} \to m^2_t$. This provides a gauge invariant construction, but only allows for a local subtraction of the interference.

In order to have a complete description of this interference, NLO calculations of the $pp \to l^+\nu l^-\nu b\bar{b}$ process, dubbed bb4l, taking into account off-shell effects are needed. The current state of the art for the bb4l process is a 4FS calculation, that includes off-shell contributions, NLO corrections to production, decay and their interference, and matching to parton showers. This calculation has been implemented in the POWHEG-BOX-RES framework, and is fully described in reference [93, 94]. This calculation provides an exact treatment of the interference at NLO accuracy. Additionally, since the calculation is performed in the 4FS it allows to model phase space regions with jet vetoes.

From the experimental viewpoint, $t\bar{t}$ and tW measurements are typically designed to be robust to these effects, which are evaluated comparing the different methods to handle the interference[1]. However, BSM searches for scalar partners of the top quark aim for phase spaces with a significant contribution from off-shell top quarks. In such phase spaces, the interference effects cannot be avoided and may constitute a significant source of systematic uncertainty.

Dedicated measurements of the quantum interference effects have been performed by the ATLAS Collaboration [118] in pp collisions. The measurement is performed in the $e\mu$ channel in topology with jets, and an observable sensitive to off-shell effects is explored. Figure 4.5 shows the distribution of the observable in data compared to the full bb4l prediction in 4.5 and the $t\bar{t}$ +tW predictions with the DS, DR and DR2

[1] At least until the 4FS NLO bb4l calculation was released.

methods. All the generators model well data in the bulk of the distributions, where interference effects are not expected to be dominant. The $\mathrm{t\bar{t}}$ +tW prediction slightly deviate from data in the phase space where interference effects are relevant, while this phase space is properly modeled by the bb4l prediction. This measurement can also be exploited to measure the top quark width with high precision [119].

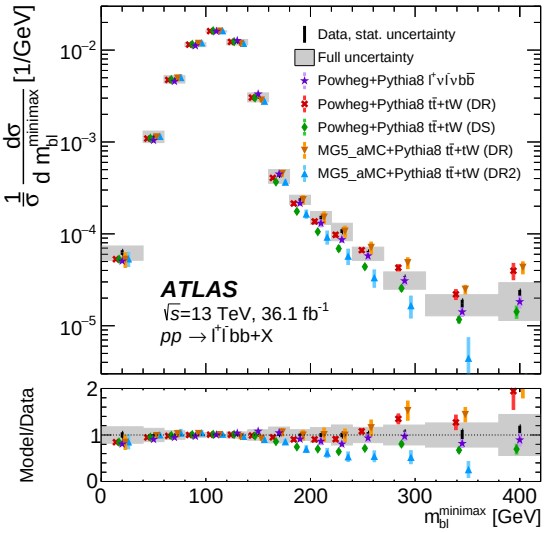


FIGURE 4.5: Normalized differential cross section as a function of the $m_{\mathrm{bl}}^{\mathrm{minmax}}$ variable in dileptonic $\mathrm{t\bar{t}}$ events. $m_{\mathrm{bl}}^{\mathrm{minmax}}$ is defined as $\min(\max(m_{\mathrm{b}_1,\mathrm{l}_1}, m_{\mathrm{b}_2,\mathrm{l}_2}), \max(m_{\mathrm{b}_1,\mathrm{l}_2}, m_{\mathrm{b}_1,\mathrm{l}_2}))$, and is expected to have a kinematic endpoint in m_t for on-shell production. Taken from [118].

4.2 Analysis strategy

In this section, the general analysis strategy is described. Two different measurements are considered in this chapter: the inclusive measurement and the differential measurement of the tW process. Both analyses are performed in the dilepton $\mathrm{e}^{\pm}\mu^{\mp}$ channel, which is the one with the smallest background contribution.

In this final state, signal is characterized by the presence of an electron, a muon and a b-tagged jet in the final state. The main challenge of the analysis is the overwhelming presence of $\mathrm{t\bar{t}}$ events across all the phase space.

The top quark pairs are produced with a cross-section more than 10 times higher than signal, and present a similar final state, which contains only an additional b jet, comparing to signal. A significant portion of $\mathrm{t\bar{t}}$ events will lose one b jet due to the limited acceptance of the detector or because the jet being below the energy threshold.

Different approaches are followed to tackle with this background in the inclusive and differential analysis: the inclusive analysis makes use of a likelihood fit to several signal and control regions, defined by the jet and b-tagged jet multiplicity of the event and two BDT scores, while the differential analysis is designed as a counting experiment performed in a region enriched as much as possible in signal. However, both

analysis rely on the different jet multiplicity of the two processes, and variables derived from this fact, as the main discriminating variable.

4.2.1 Object selection

Muons identified by the PF algorithm are applied additional identification criteria corresponding to the "tight" selection, described in section 3.1.2, to efficiently select muons produced prompt W, Z or τ decays. Additionally, in order to reject muons not produced in W, Z or τ decays, muons are required to be isolated with a relative isolation variable of less than 15%, with an isolation cone size of 0.4.

Reconstructed electrons are required to pass the "cut-based tight" identification criteria, described in section 3.2. Electron candidates in the transition region, fulfilling $1.442 < |\eta_{SC}| < 1.5660$, are rejected, where η_{SC} denotes the pseudorapidity of the ECAL supercluster.

Both electrons and muons in the event are required to have small impact parameters measured with respect to the primary vertex and to be isolated with the rest of the physics objects to reject leptons produced in quark decays.

Jets follow the definition in 2.3.2 and are required to have a p_{T} of at least 30 GeV. Since leptons are usually clustered as jets by the anti-k_{T} algorithm, jets within a $\Delta R < 0.4$ of a selected leptons are not included in the counting. b tagged jets are identified using the CSVv2 algorithm, described in section 2.3.3. Finally, $p_{\mathrm{T}}^{\mathrm{miss}}$ is constructed using all the particles reconstructed by the PF algorithm and coming from the primary vertices. Momentum corrections to the energy of the jets is propagated to the $p_{\mathrm{T}}^{\mathrm{miss}}$.

In order to be able to capture jets that are slightly below the energy acceptance, and achieve better discrimination between $\mathrm{t\bar{t}}$ and tW events, "loose" jets, defined as those jets fulfilling the criteria above but with a p_{T} threshold of 25 GeV and a p_{T} below 30 GeV, are used.

4.2.2 Event selection

Events are collected with a set of double and single lepton triggers to maximize the signal efficiency. These triggers require the presence of a muon (electron) with p_{T} greater than 23 (12) GeV and an electron (muon) with a p_{T} greater than 12 (8) GeV. In addition, other triggers that only require the presence of one muon (electron) with a p_{T} greater than 24 (27) GeV are used. This strategy is followed in order to increase the precision on the measurement of the trigger efficiency, that is performed following the orthogonal trigger method, described in section 2.5.2.

Events collected by these triggers are required to have a pair of leptons with a p_{T} greater than 25 (20) GeV for the leading (sub-leading) lepton with opposite charge. The two leading leptons are required to be an electron and a muon. Events with $m_{\mathrm{ll}} < 20$ GeV are rejected to suppress the contribution from low mass resonances. Specific filters are applied to reject events with anomalous $p_{\mathrm{T}}^{\mathrm{miss}}$ due to detector noise, cosmic rays and other sources.

The MC modeling of several kinematic variables is checked in events fulfilling the selection described above. Figure 4.6 shows the data distributions of several kinematic variables, compared to the MC predictions with its corresponding uncertainties, with a good agreement between the two.

After this selection, categories are considered based on the jet (n_j) and b-tagged jet (n_b) multiplicity of the events. The (n_j, n_b) distribution of selected events is shown in figure 4.7, compared to the SM predictions. Each bin contains a different amount of signal and background events. Therefore, these bins are used to define the several regions of interest in the analysis. In particular, the region with exactly one jet that is b-tagged (1j1b) is the one that is enriched the most in signal. The region with exactly two jets, out of which one is a b-tagged jet (2j1b) is less pure but keeps a significant part of the signal. On the other hand, the region with exactly two jets that are b-tagged (2j2b) is fully enriched in $\mathrm{t\bar{t}}$ events, and is used to constrain this background.

The inclusive measurement is performed simultaneously with events in the 1j1b and 2j1b region. The 2j2b region is also included in the signal extraction fit to constrain the $\mathrm{t\bar{t}}$ background and the uncertainties associated to it.

The differential measurement is performed only in events in the 1j1b region. Additionally, in order to obtain a region purer in signal, only events with no additional "loose" jets are considered. The distribution of number of "loose" jets is shown in figure 4.7 for events in the 1j1b region, showing the discrimination power of this variable.

4.3 Background and signal estimation

Simulated events are used to predict the data yields in the measurement regions of these analyses.

The tW signal is simulated at NLO using POWHEG v1 [120], with the NNPDF 3.0 PDF set [121]. The DR approach described in section 4.1.3 is used to take into account the interference with $\mathrm{t\bar{t}}$. A set of simulated events using the DS approach is also used to determine the uncertainty associated to the modeling of the interference.

$\mathrm{t\bar{t}}$ events are simulated using POWHEG v2 [87], which is also used to infer the dependency on the PDFs and the factorization and renormalization scales. This sample is normalized to the NNLO+NNLL calculation obtained in [86].

Drell–Yan and W+jets background events are simulated at NLO with MADGRAPH MC@NLO v2.2.2 with NNPDF 3.0 PDFs. These processes are simulated with up to two additional partons and the FxFx scheme is used for merging [89]. Contributions from WW, WZ and ZZ (denoted as VV) are simulated with PYTHIA v8.205 [26] at LO. Finally, $\mathrm{t\bar{t}W}$ and $\mathrm{t\bar{t}Z}$ production are simulated at NLO precision with MADGRAPH MC@NLO.

All these samples, except $\mathrm{t\bar{t}}$, are interfaced to PYTHIA v8.205 with CUETP8M1 underlying event tune [122] to simulate parton shower and hadronization. For $\mathrm{t\bar{t}}$ events, the CUETP8M2T4 tune [123] is used instead.

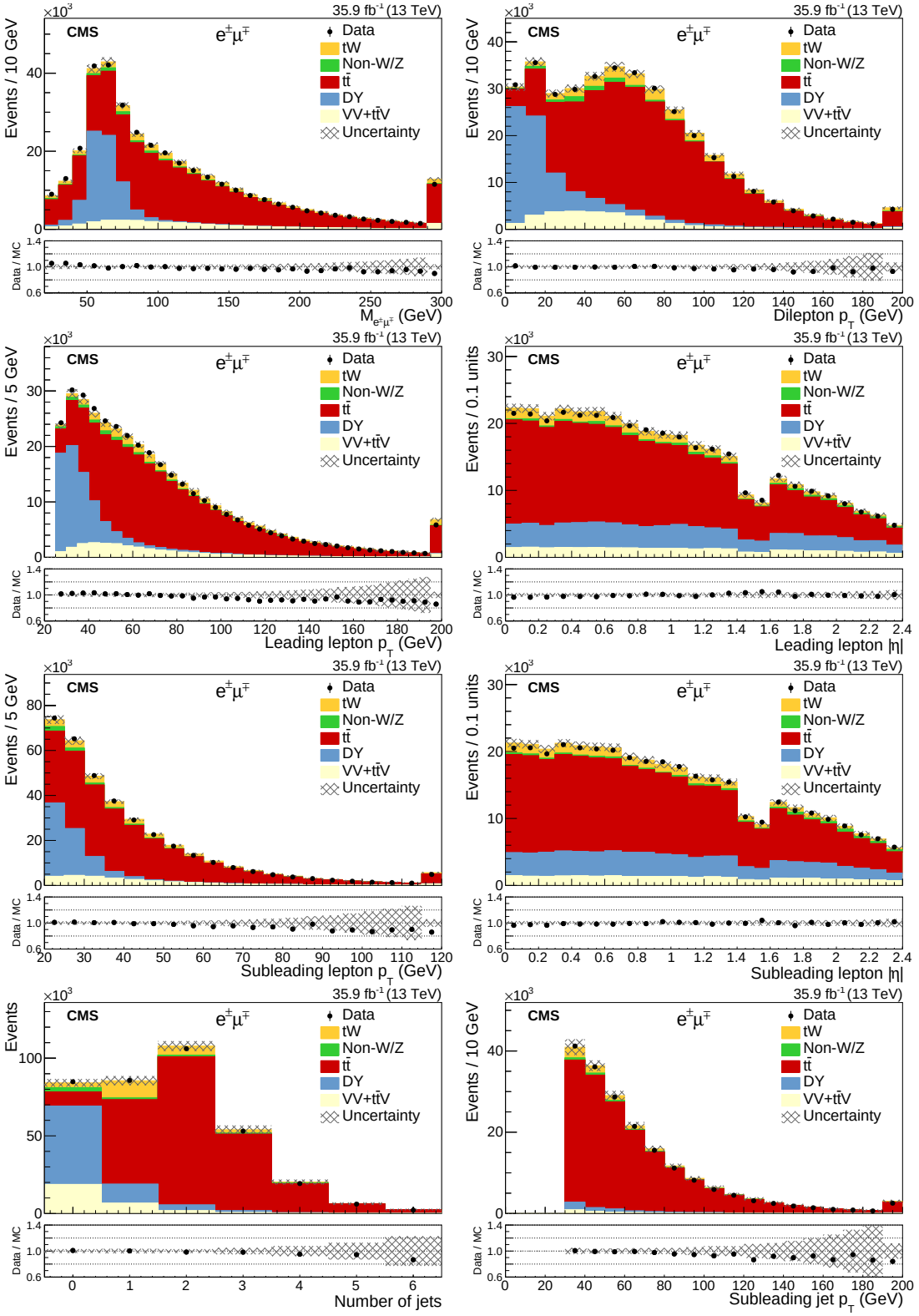


FIGURE 4.6: Observed events as a function of several kinematic variables, compared to the predictions in events with a $e^{\pm}\mu^{\mp}$ pair. Column in the left (right) shows, from top to bottom, the $m_{e^{\pm}\mu^{\mp}}$, the p_T of the leading lepton, the p_T of the sub-leading lepton and the jet multiplicity (p_T of the dilepton system, $|\eta|$ of the leading lepton,$|\eta|$ of the sub-leading lepton and the p_T of the sub-leading lepton).

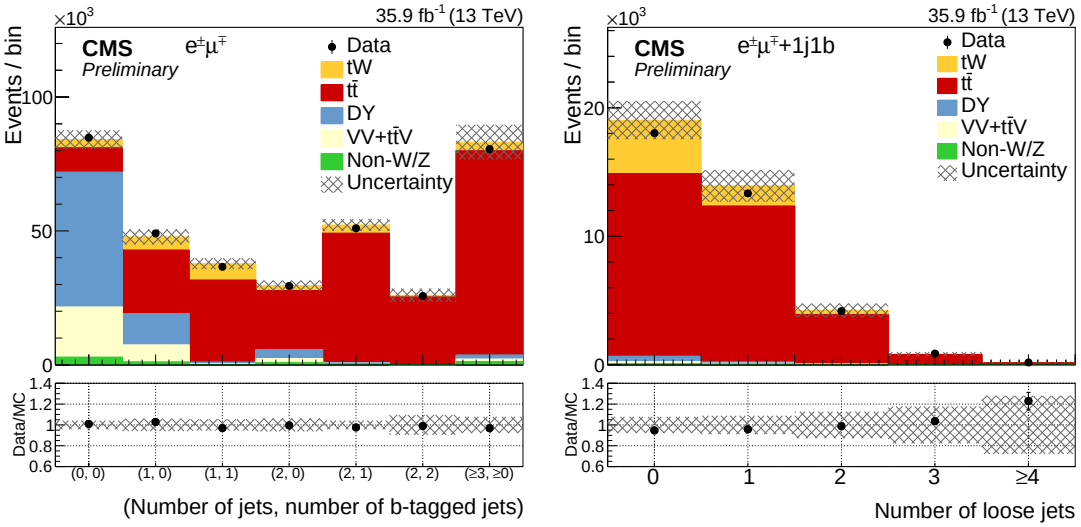


FIGURE 4.7: Distribution of the (n_{jet}, $n_{\mathrm{b}-jet}$) variable in events with a $\mathrm{e}^{\pm}\mu^{\mp}$ pair (left) and the "loose" jet multiplicity in events in the 1j1b region (right).

Contribution from leptons not coming from W, Z or τ decays is expected to be dominated by $\mathrm{t\bar{t}}$ and W+jets events. The simulations described above are used to model them.

4.4 Systematic uncertainties

Measurements of the tW cross section are affected by several sources of systematic uncertainty. Each source is evaluated by performing consistent variations of the signal and background estimations by the estimated uncertainty. Two types of systematic uncertainties are taken into account: those stemming from the calibration or characterization of the reconstructed physics objects, and those due to the unknowns in the MC models, either due to their intrinsic assumptions or to the limited knowledge of fundamental parameters.

4.4.1 Experimental uncertainties

- Uncertainties due to the trigger and lepton reconstruction efficiencies are estimated by varying the scale factors by their corresponding uncertainties, that are obtained following the method described in section 2.5.2. Three variations are applied independently: trigger, electron and muon efficiency uncertainties.
- The uncertainty due to the limited knowledge of the Jet Energy Scale (JES) and Jet Energy Resolution (JER) is performed by varying the scale and resolution of the jets within its uncertainties as measured in bins of p_{T} and η [46]. Even if there are several sources of uncertainty affecting the jet energy scale, the quadratic sum of all the possible uncertainties is taken into account as a single variation. This assumption is valid, since all jets selected in the analysis, both for signal and $\mathrm{t\bar{t}}$, belong to the same kinematic regime.

- Uncertainties resulting from b tagging efficiency and misidentification rate are determined by varying the b tagging data-to-simulation scale factors for b and light jets, respectively [43].
- The inelastic cross section is varied by its uncertainty, $\pm$ 4.6% [124]. This cross section is used to estimate the distribution of the number of pile-up interactions in data, to account for that in the MC.
- The uncertainty in the total integrated luminosity is estimated to be 2.5% [125].

4.4.2 Modeling uncertainties

Simulated events used to estimate signal and background contain underlying assumptions, that may have an impact in the analysis. This is assessed by considering dedicated simulated samples of the $\mathrm{t\bar{t}}$ and tW processes. These samples are generated varying the relevant parameters from those of the standard POWHEG+PYTHIA simulations.

The uncertainty on the missing higher orders in the calculation of the hard-production process is assessed by considering variations of μ_R and μ_F by factors of 2 and 0.5 relative to their nominal value. The uncertainty associated to the knowledge of the proton structure is obtained by reweighing the sample of simulated $\mathrm{t\bar{t}}$ events by the 100 NNPDF3.0 replicas. For each bin of the fitted distributions, the root-mean-square of the 100 replicas is taken as this source of uncertainty. These two variations are used to estimate the acceptance and distributions of signal and backgrounds. The total production rate for $\mathrm{t\bar{t}}$ is a assigned a 5% uncertainty due to the two effects [86].

The interference between tW and $\mathrm{t\bar{t}}$ production is handled by considering the difference between the DR and DS prescriptions.

To account for the uncertainties in the parton shows and jet fragmentation modeling, several aspects are taken into account:

- Underlying event: several PYTHIA parameters are tuned according to measurements of the underlying event [123]. These parameters are varied in simulated $\mathrm{t\bar{t}}$ events according to their uncertainties.
- Matching between matrix element and parton shower. The uncertainty associated to the matching between the matrix element and parton shower calculations in POWHEG is considered by varying the damping parameter, $h_{damp} = 1.58^{+0.66}_{-0.59} m_t$, in simulated $\mathrm{t\bar{t}}$ events [123].
- Initial State Radiation (ISR) and Final State Radiation (FSR). The PS scale used to simulate the initial and final state radiation is varied up and down by a factor of two.
- Color reconnection: the effect of multiple parton interactions and the parameterization of color reconnection has been studied in [123] and is varied in simulated $\mathrm{t\bar{t}}$ events. Two alternative models of color reconnection are also compared with the default PYTHIA models. The first is a model with strong formation beyond leading color [126] and the second a model in which the gluons can be moved

to another string [127]. All the models are tuned to measurements of the color reconnection. The uncertainty is taken as the largest difference among the variations for each bin of the distributions.

- Semileptonic B hadron decays: the semileptonic B hadron branching fraction is varied depending on the differences between PYTHIA semileptonic branching fractions for B^0, $\mathrm{B}^{\pm}$, B_s^0, Λ_b and the Particle Data Group values [17].
- B hadron fragmentation: the fragmentation into B hadrons is varied within the uncertainties of the Bowler-Lund fragmentation function, tuned to data measured by ALEPH and DELPHI Collaborations [128, 129]. In addition, the difference between the Bowler-Lund and Peterson fragmentation function. The largest difference between these two variations is taken as the uncertainty.

The mismodeling of the top p_T spectrum in $\mathrm{t\bar{t}}$ described in section 4.1 is also taken into account by considering simulated events with the nominal POWHEG model. These events are reweighed in order for the top p_T spectrum to match that of data. The difference between this prediction and the nominal POWHEG prediction is considered. This effect was observed to have no impact in the inclusive analysis, and is not taken into account. In the differential analysis, this correction is relevant when determining the p_T distributions of the leptons. In that case, the reweighed distribution is taken as the central value, and the difference with respect to the nominal as the systematic uncertainty.

Finally, other backgrounds different from $\mathrm{t\bar{t}}$ are assigned a 50% uncertainty. This uncertainty accounts for the limited knowledge of the cross section for these processes, as well as extrapolations from the full phase space to the phase space in which this analysis takes place.

In the inclusive analysis, these uncertainties are taken into account as independent nuisance parameters in the fit. For those uncertainties affecting the signal, the effect on the extrapolation from the full to the fiducial phase space is factorized from the effect on efficiency and distribution shapes. This effect is taken into account as separated nuisance parameter, that cannot be constrained in the fit. The number of expected $\mathrm{t\bar{t}}$ and tW events in these regions is shown in table 4.1.

Region	tW	$\mathrm{t\bar{t}}$
1j1t	6147 ± 442	30622 ± 1862
2j1t	3125 ± 294	48484 ± 1984
2j2t	725 ± 85	25052 ± 2411

TABLE 4.1: Number of expected $\mathrm{t\bar{t}}$ and tW events in the various signal and control regions. Uncertainties include both the systematic and statistical components.

4.5 Inclusive measurement of the tW cross section

As noted above, the signal regions defined in the analysis have a dominant contribution of $\mathrm{t\bar{t}}$ events, and only the 1j1b and 2j1b regions have a significant contribution of

signal. The signal purity of these two regions is of the order of 20% and 5%, respectively.

The distinct signature between $\mathrm{t\bar{t}}$ and tW in the 1j1b region is that in $\mathrm{t\bar{t}}$ events one jet is outside the acceptance or does not pass the p_{T} or identification requirements. Therefore the "loose" jet multiplicity distribution can be employed as a discriminating variable between signal and background. However, as seen in figure 4.7, its discriminating power does not allow to construct a region fully enriched in signal.

The topology of the 2j1b region is the one expected in $\mathrm{t\bar{t}}$ events in which one of the b jets has not been tagged as such. This can occur in a non negligible amount of cases due to the limited efficiency of the tagging methods. However, there is a significant contribution from signal, for cases in which tW is produced in association with additional partons. In these cases, the additional jet is expected to be softer, since it is expected to come from radiation or from a diagram line corresponding to an off-shell top quark.

Since in the 1j1b and 2j1b regions there is no single variable that is expected to easily discriminate between $\mathrm{t\bar{t}}$ and tW, dedicated multivariate discriminators are employed in each of the 1j1b and 2j1b categories to obtain regions purer in signal. Even with the usage of these techniques, the presence of $\mathrm{t\bar{t}}$ events is still dominant over signal, and the uncertainties associated to the estimation of this background are the dominant ones in the analysis. To further constrain them, the 2j2b region is also included in the analysis. The total cross section measurement is then performed by making a likelihood fit to event yields in bins of the discriminator distribution in the 1j1b and 2j1b region, exploiting the discrimination power of the whole shape, and the yields on the distribution of the sub-leading jet p_{T} in the 2j2b region, which allow to slightly constrain the uncertainties associated to the jet energy scale.

4.5.1 BDT for background discrimination

To improve the discrimination, BDTs with gradient boosting are used. The input variables used for the 1j1b region are:

- p_{T} of the leading "loose" jet, set to 0 for events with no "loose" jets;
- magnitude of the vector sum of the p_{T}'s of the leptons, the jet and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ (p_{T}^{sys});
- p_{T} of the jet;
- ratio of the scalar sum of the p_{T} of the leptons to the scalar sum (H_T) of the p_{T}'s of the leptons, the jet and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$;
- number of "loose" jets;
- centrality (ratio between the scalar sum of the p_{T} and the total momentum) of the jet and two leptons;
- magnitude of the vector sum of the p_{T} of the jet and leptons;
- H_T;

- ratio of p_{T}^{sys} and H_T of the event;
- invariant mass of the combination of the leptons, jet and $p_{\mathrm{T}}^{\mathrm{miss}}$;
- number of b-tagged "loose" jets.

These distributions are used aiming to discriminate profiting from the different topology and kinematics expected in $\mathrm{t\bar{t}}$ and tW events. Variables related to the "loose" jet multiplicity or kinematics are used to recover the jet that has not been selected in $\mathrm{t\bar{t}}$ events. Other variables, such as p_{T}^{sys} aim to be sensitive to this missing jet via the momentum imbalance that would appear in the total system. Finally, others, such as the p_{T} of the jet or H_T, are sensitive to the higher energy that is present in the $\mathrm{t\bar{t}}$ system in comparison with the tW system. The distribution of these variables is shown in figures 4.8- 4.9.

For the 2j1b region, the following variables are used as input for the BDT:

- ΔR between the dilepton and dijet systems, $\Delta R(\mathrm{e}^{\pm}\mu^{\mp}, j_1 j_2)$;
- ΔR between the dilepton system and the system formed by the two jets and $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$, $\Delta R(\mathrm{e}^{\pm}\mu^{\mp}, j_1 j_2 \vec{p}_{\mathrm{T}}^{\mathrm{miss}})$;
- p_{T} of the sub-leading jet;
- ΔR between the leading lepton and the leading jet $\Delta R(\mathrm{l}_1, j_1)$.

The sub-leading jet p_{T} is included in the BDT because this distribution is expected to be softer in signal than in $\mathrm{t\bar{t}}$ events. ΔR distributions in the various systems are expected to have slightly different between signal and $\mathrm{t\bar{t}}$ events due to the higher total energy in the $\mathrm{t\bar{t}}$ system.

The distribution of these variables is shown in figure 4.10, featuring a good agreement between data and prediction.

Some of the hyperparameters in the two BDTs trained are shown in table 4.2. The normalized distribution of the BDT scores in training and testing simulated events is presented in figure 4.11, showing the discrimination power of the variables, and its generalization from training to test sample. The presented distributions do not show any sign of overtraining.

The distribution of the BDTs score in data is compared to predictions in figure 4.12, showing a good agreement between data and the predictions.

Hyperparameter	2j1t	1j1t
Number of trees	200	2000
Shrinkage	0.05	0.01
Maximum tree depths	4	4

TABLE 4.2: Hyperparameters used in the training of the BDT discriminants used in the 2j1b and 1j1b regions.

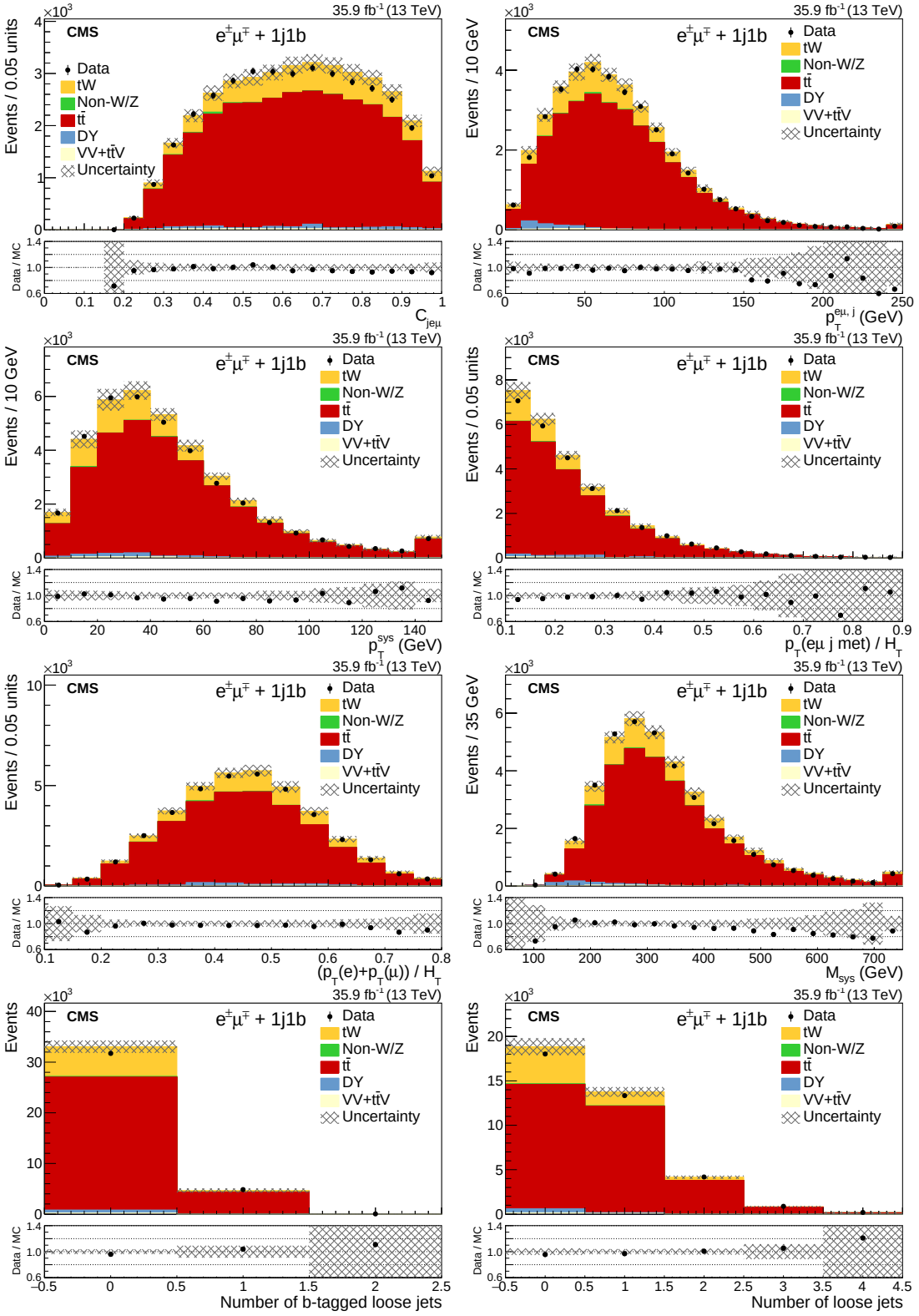


FIGURE 4.8: Distributions of observed and expected events as function of some of the BDT input variables in the 1j1b region.

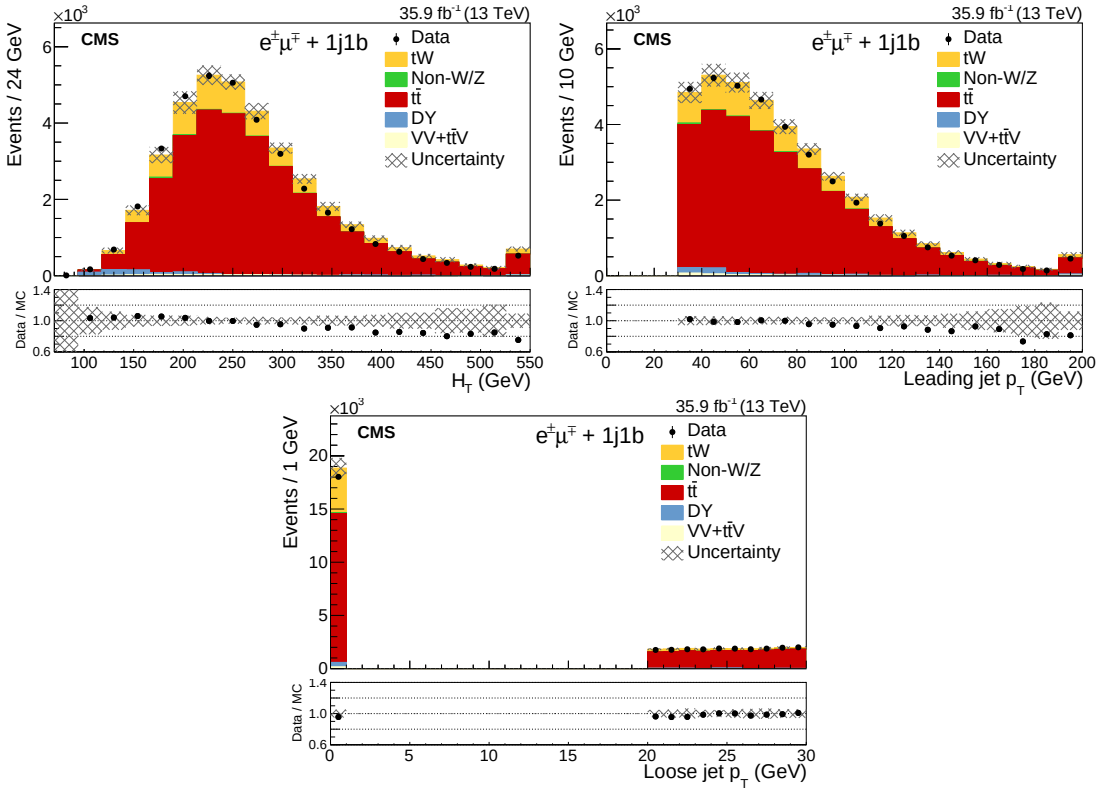


FIGURE 4.9: Distributions of observed and expected events as function of some of the BDT input variables in the 1j1b region.

Even if the amount of observed events is not expected to be a limitation for this analysis, the amount of simulated events that are used is limited. In particular, the alternative samples used to estimate the systematic uncertainties are significantly smaller than the nominal ones. Fluctuations of these samples could yield to an overestimation of the uncertainty as well as instabilities in the signal extraction fit. In order to ensure that each bin of the signal extraction contains enough simulated events for a precise estimation of both the expected yields, as well as the of associated uncertainties, a specific binning is chosen for these distributions. In particular, for the BDT distributions, the quantiles of the background distribution are taken as the bin limits. This ensures that all bins contain approximately a similar amount of simulated background events, which is the optimal way to reduce the statistical uncertainties associated to the limited amount of simulated events. For the sub-leading jet p_{T}, the binning is chosen according to the expected detector resolution.

4.5.2 Signal extraction

The observed data in the distributions used for the signal extraction, is shown in figure 4.13. The SM expectations are also shown in the plot, together with the uncertainties associated to it.

The signal is then extracted to yields in these distributions. Besides the nuisance parameters, that parametrize the systematic uncertainties described in section 4.4, the fit model includes a signal-strength parameter, $\mu_{\mathrm{tW}} = \sigma_{\mathrm{tW}} / \sigma_{\mathrm{tW}}^{exp}$, that is unconstrained

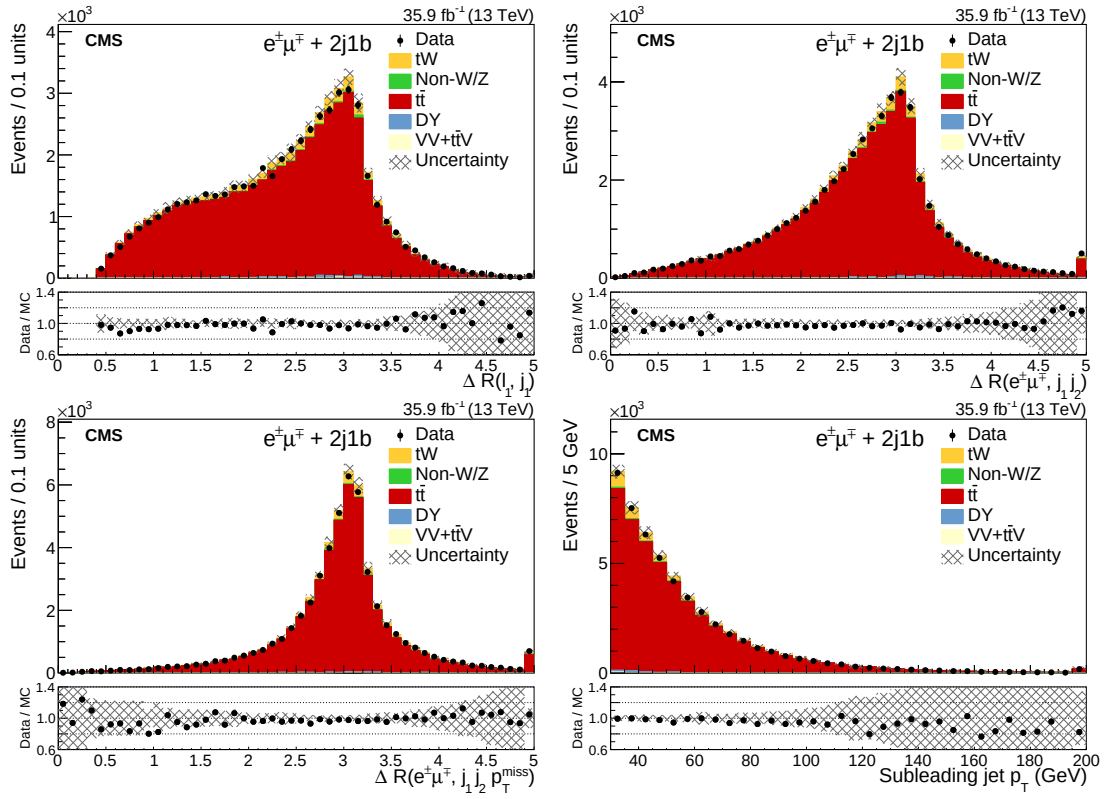


FIGURE 4.10: Distributions of observed and expected events as function of some of the BDT input variables in the 2j1b region.

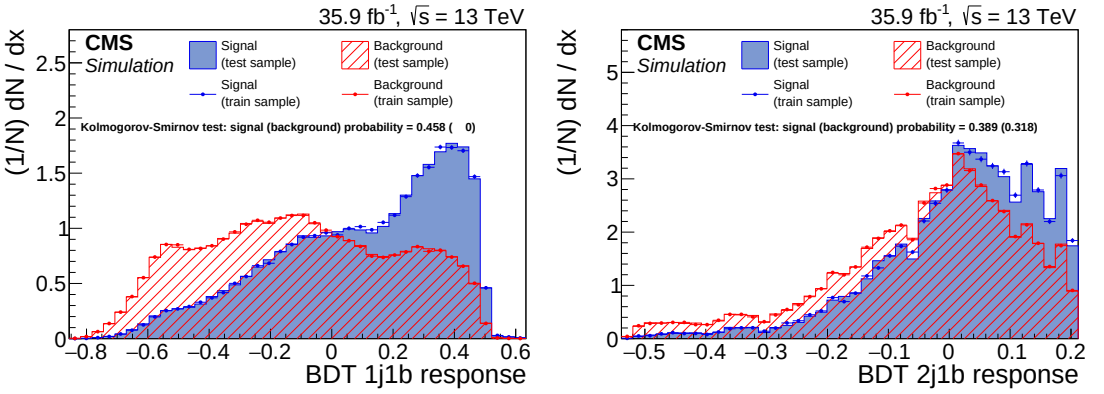


FIGURE 4.11: Normalized distributions of the BDT scores in $t\bar{t}$ (red) and signal (blue) simulated events used for training (dots) and testing (filled). Left shows the 1j1b discriminator, while right shows the 2j1b discriminator.

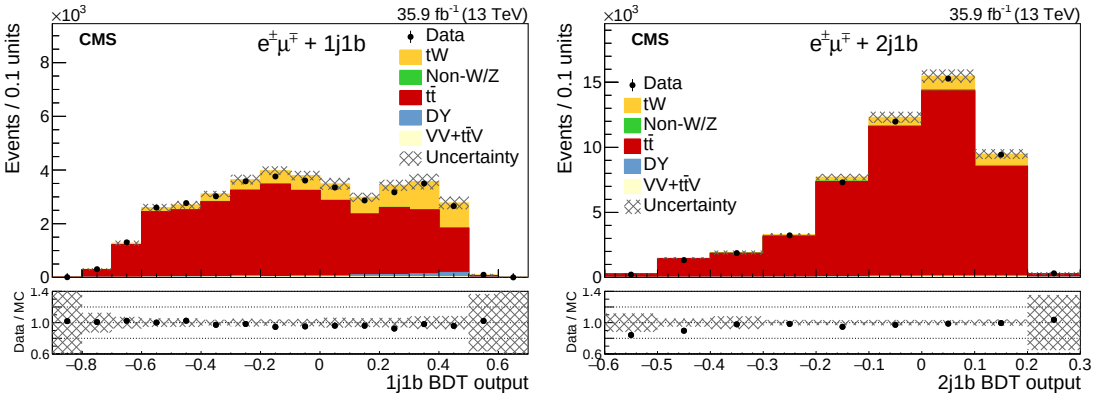


FIGURE 4.12: Distribution of events in the 1j1b (left) and 2j1b region (right) as a function of the BDT score in each respective region.

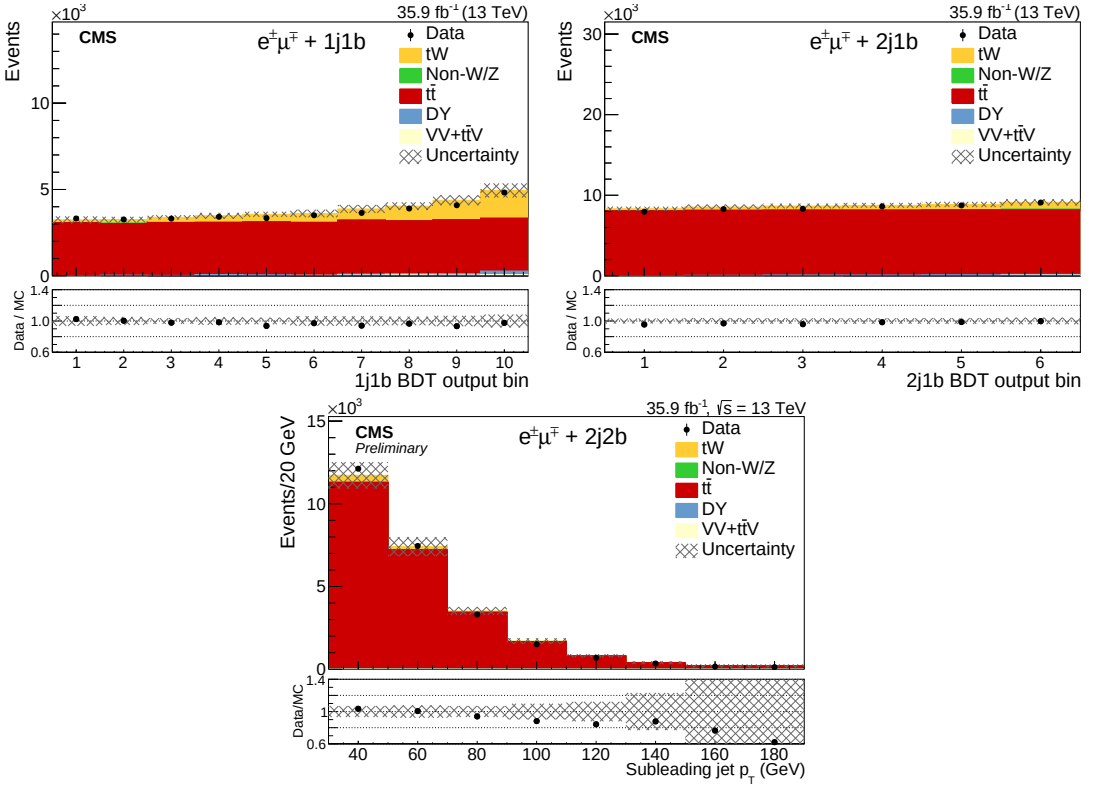


FIGURE 4.13: Signal extraction variables in the 1j1b (top left), 2j1b (top right) and 2j2b (bottom), with the respective chosen binning, before any fit is performed.

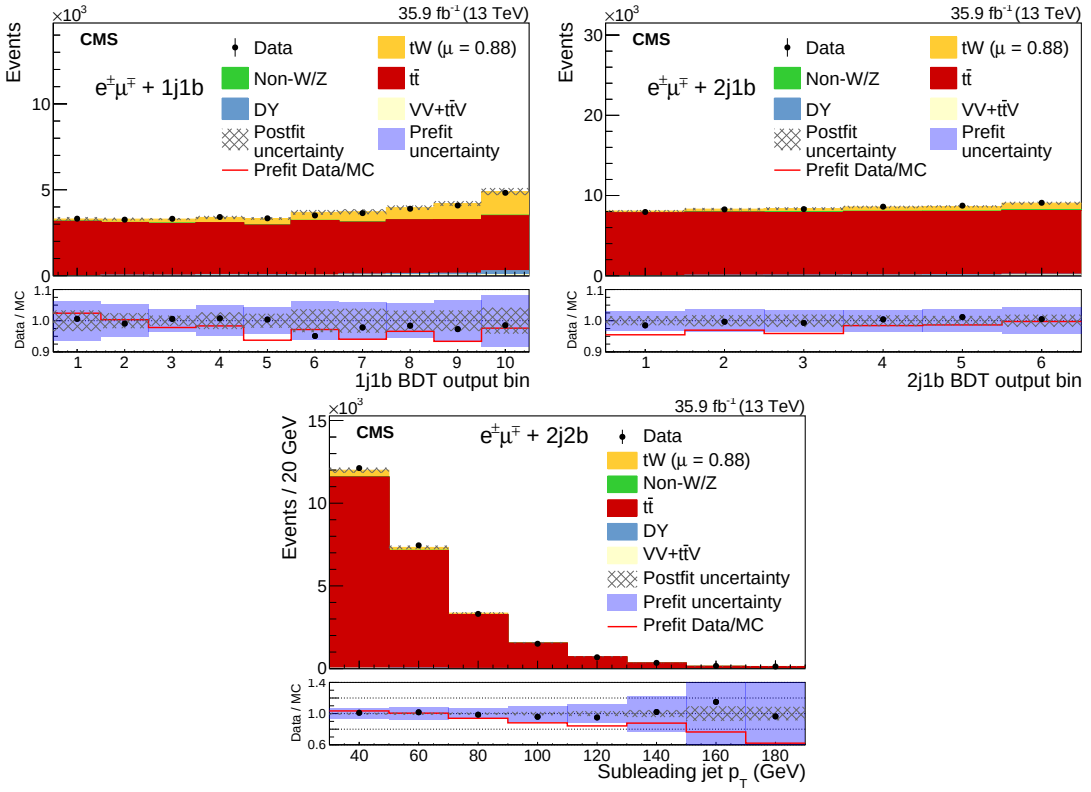


FIGURE 4.14: Signal extraction variables in the 1j1b (top left), 2j1b (top right) and 2j2b (bottom), with the respective chosen binning, after the fit is performed. Red line and blue color indicates the prefit data/MC agreement and uncertainties.

in the fit. This parameter of the model defines the scaling of the signal with respect to the value predicted by the SM. The best fit for μ_{tW} is obtained by maximizing the likelihood. The 68% confidence interval is obtained by considering variations of the test statistic described in section 2.6.2 by one unit from its minimum.

4.5.3 Results

The best fit for the tW signal-strength parameter is 0.88 $\pm$ 0.02 (stat) $\pm$ 0.09 (syst) $\pm$ 0.03 (lumi), corresponding to a measured cross section of 63.1 $\pm$ 1.8 (stat) $\pm$ 6.3 (syst) $\pm$ 2.1 (lumi) pb, consistent with the SM expectation.

Figure 4.14 shows the data observed in the various signal regions as well as the predictions after the fit was performed, setting the μ_{tW} parameter to its postfit value, showing a good compatibility between data and the statistical model.

Table 4.3 shows the impact of each one of the sources of systematic uncertainty in the analysis. The impact of a given uncertainty is obtained by comparing the uncertainty of the nominal fit with that of a fit performed fixing the associated nuisances to their postfit value. The alternative fit will have, by construction, the same postfit values as the nominal one, but the uncertainty will be smaller since the fixed parameters do not play a role. The quadratic difference between the uncertainty of the nominal and

the uncertainty of the alternative fit is taken as the impact of the source. Statistical uncertainty is obtained by fixing all the nuisance parameters to their postfit value.

Source	Uncertainty (%)
Experimental	
Trigger efficiencies	2.7
Electron efficiencies	3.2
Muon efficiencies	3.1
JES	3.2
Jet energy resolution	1.8
b tagging efficiency	1.4
Mistag rate	0.2
Pileup	3.3
Modeling	
$\mathrm{t\bar{t}}$ μ_R and μ_F scales	2.5
tW μ_R and μ_F scales	0.9
Underlying event	0.4
Matrix element/PS matching	1.8
Initial-state radiation	0.8
Final-state radiation	0.8
Color reconnection	2.0
b fragmentation	1.9
Semileptonic b decay	1.5
PDFs	1.5
DR-DS	1.3
Background normalization	
$\mathrm{t\bar{t}}$	2.8
VV	0.4
Drell–Yan	1.1
Non-W/Z leptons	1.6
$\mathrm{t\bar{t}V}$	0.1
MC finite sample size	1.6
Full phase space extrapolation	2.9
Total systematic (excluding integrated luminosity)	10.1
Integrated luminosity	3.3
Statistical	2.8
Total	11.1

TABLE 4.3: Breakdown of the systematic and statistical uncertainty associated to the signal extraction fit.

The uncertainty is dominated by trigger and lepton efficiency and luminosity. The size of these uncertainties is due to the dominant presence of background in signal and control regions: a small uncertainty in $\mathrm{t\bar{t}}$ events is amplified when propagating it to signal events.

4.6 Differential measurements of the tW cross section

This section describes measurements of the tW production differential cross section. This measurement is performed in the 1j1b region described in section 4.2.2, which is the one with the highest signal purity. While other differential cross section distributions in phase spaces dominated by background have made use of multivariate discriminants [113, 130], this measurement uses uniquely object multiplicity cuts to define the measurement region.

This approach has three advantages. First, it does not introduce strong assumptions on the distribution of the signal in a multivariate discriminant, reducing the model dependency of the measurement. Secondly, this allows to define a fiducial region consistent with the regions employed in the measurement, and assumptions on the extrapolation to the full phase space are made explicit on the method. Finally, this method allows to measure the cross section as a function of any variable, provided that the signal purity in all the bins is large enough for a sensitive measurement.

4.6.1 Observables under study

The following variables are studied in this analysis:

- the p_{T} of the highest p_{T} lepton;
- the jet p_{T};
- the difference in the azimuthal angle of the muon and the electron, $\Delta\phi(\mathrm{e}^{\pm},\mu^{\mp})$;
- the longitudinal momentum component of the system formed by the muon, the electron and the jet of the event, $p_z(\mathrm{e}^{\pm},\mu^{\mp},j)$;
- the invariant mass of the system formed by the electron, the muon and the jet $m(\mathrm{e}^{\pm},\mu^{\mp},j)$;
- the transverse mass of the system formed by the electron, the muon, the jet and the missing transverse momentum $m_T(\mathrm{e}^{\pm},\mu^{\mp},j,p_{\mathrm{T}}^{\mathrm{miss}})$.

The first three variables provide general information regarding the kinematic properties of the tW system. In particular, the first two are sensitive to the mismodeling of the top quark p_{T}. The $\Delta\phi(\mathrm{e}^{\pm},\mu^{\mp})$ variable allows to explore correlations between the two top quarks and to measure spin-related properties. The $p_z(\mathrm{e}^{\pm},\mu^{\mp},j)$ is a proxy of the total boost to the system, and provides sensitivity to the production mechanism. The last two, the invariant and transverse masses are sensitive to the total energy and mass of the tW system.

4.6.2 Selection and fiducial space definition

In order to achieve the maximum signal purity, the measurement is performed in events in the 1j1b region, described in section 4.2.2, without reconstructed "loose" jets.

The observed number of events as a function of the observables under study is shown in figure 4.15, together with the SM prediction of the tW prediction from the POWHEG model.

Since this selection still has a significant contribution from $\mathrm{t\bar{t}}$ events, it is important to check the modeling of this process in a phase space that is close to the measurement region. To do so, a control region is defined by considering events in the 1j1b region with additional "loose" jets in the final state. The observed distributions in this region are compared with the predictions in figure 4.16, showing a good agreement between data and the predictions.

The fiducial region is constructed using particle level objects. A summary of the main requirements applied to the objects is discussed here. The complete definition is shown in a dedicated reference [131]. Particle level objects are required to have a lifetime greater than 30 ps.

Particle level charged leptons are defined as those produced in prompt decays of W, Z, or τ. In particular, leptons produced in the decay of heavy hadrons are not taken into account. These leptons are dressed with photons with a size of $\Delta R < 0.1$.

Particle level jets are defined by clustering all stable particles with the anti-k_{T} algorithm [41] with a jet cone parameter of $R = 0.4$. Neutrinos are excluded from this clustering, as well as prompt leptons and photons. b jets are defined as those jets that contain a decayed B hadron.

The fiducial region is defined as events with one particle-level muon with $p_{\mathrm{T}} > 20$ and $|\eta| < 2.4$, one electron with $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.4$ and $|\eta < 1.442|$ or $|\eta| > 1.566$. Events are also required to have one particle-level jet with $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.4$, and no additional jets with $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.4$. This definition of particle objects and the definition of the fiducial region are summarized in tables 4.4 and 4.5, respectively.

Muons	
p_{T} (GeV)	> 20
$\lvert\eta\rvert$	< 2.4
Electrons	
p_{T} (GeV)	> 20
$\lvert\eta\rvert$	< 2.4 && (< 1.4442 || > 1.5660)
Jets	
p_{T} (GeV)	> 30
$\lvert\eta\rvert$	< 2.4
Loose jets	
p_{T} (GeV)	$> 20 \wedge < 30$
$\lvert\eta\rvert$	< 2.4

TABLE 4.4: Selection requirements on particle level objects.

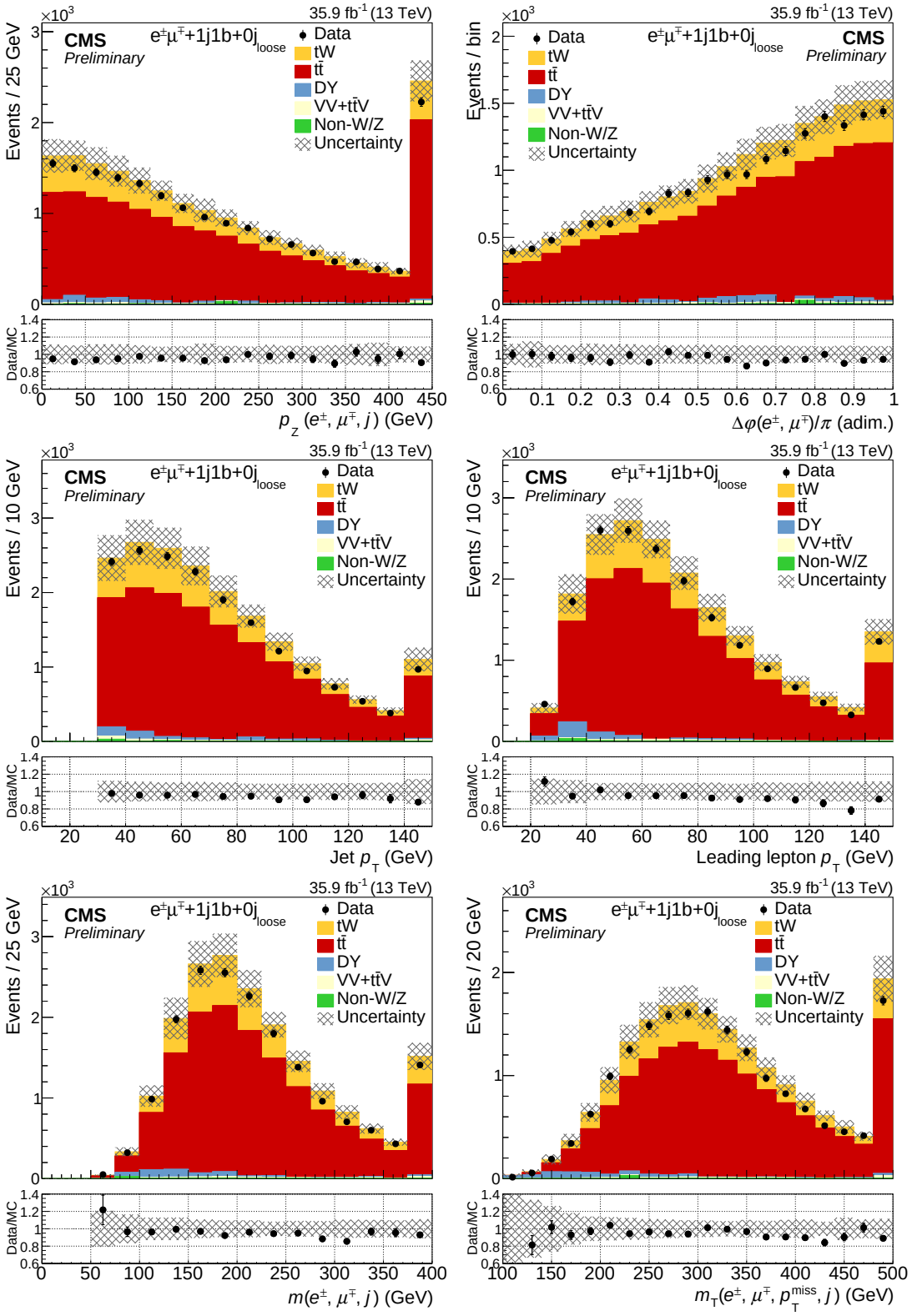


FIGURE 4.15: Distribution of the variables under study for events in the measurement region.

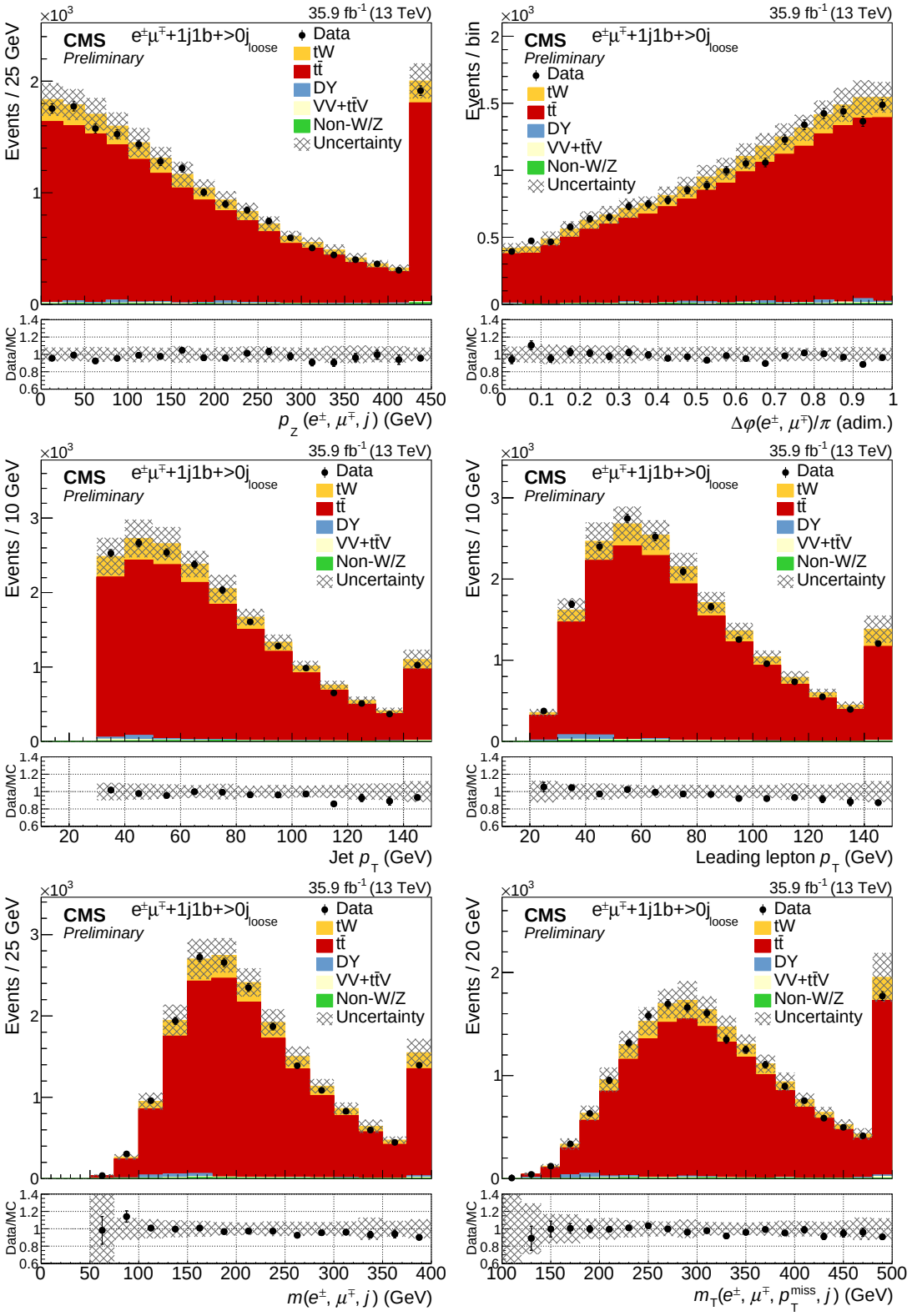


FIGURE 4.16: Distribution of the variables under study for events in the $t\bar{t}$ control region.

Number of leptons	≥ 2
p_{T} (ℓ_1)	> 25 GeV
$m_{e\mu}$	> 20 GeV
Number of jets	1
Number of loose jets	0
Number of b jets	1

TABLE 4.5: Definition of the fiducial region.

4.6.3 Unfolding to particle level

To unfold the results to particle level, the approach described in section 2.6.3 is followed.

The response matrices are obtained using the tW signal simulations described in section 4.3. The usage of response matrices based in simulations introduces uncertainties associated to the calibrations derived for them and mismodeling effects. To account for this, replicas of the response matrix are considered varying each one of the systematics uncertainties, described in section 4.4, affecting the signal. The response matrices of the variables under study are shown in figure 4.17. As mentioned in section 2.6.3, the binning of these matrices has been optimized to reduce the stability and purity. The condition number for the matrices is shown in table 4.6, which is of the order of magnitude of the unity. The choice for τ, the regularization parameter, is made by scanning the L-curve as described in 2.6.3. No significant difference is seen in the result between applying and not applying the regularization terms, which is consistent with the condition numbers of the response matrices. Because of this, the unbiased maximum likelihood estimator is used.

For a given variable X, the absolute differential tW cross section for a given bin j of X is computed from the number of unfolded signal events $N_j^{sig,unf}$ as

$$\left(\frac{d\sigma}{dX}\right)_j = \frac{1}{\mathcal{L}} \frac{N_j^{sig,unf}}{\Delta_j}, \tag{4.2}$$

where Δ_j is the bin width of bin j. In order to profit from the cancellation of systematic uncertainties, the normalized differential cross section is obtained, dividing by the fiducial cross section (the sum over all bins of the absolute differential cross section).

Uncertainties associated to the background subtraction are taken into account by considering suitable variations of the systematic uncertainties. To propagate them to the cross section measurement, the measurement is performed using variations of the systematic uncertainties both in the background subtraction and the response matrix. Response matrices and background subtraction are varied simultaneously for systematics that affect both signal and backgrounds in a correlated way.

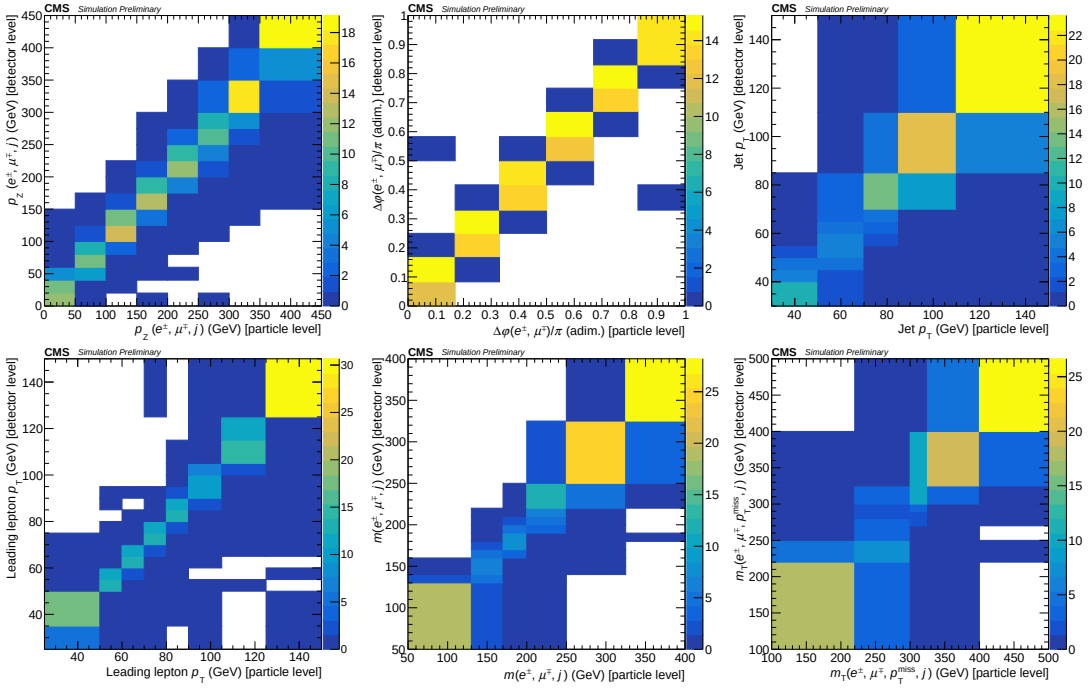


FIGURE 4.17: Response matrices for the variables under study, obtained in tW simulated events in the signal region.

Variable	Condition number
Leading lepton p_T (GeV)	2.05
Jet p_T (GeV)	4.42
$\Delta\varphi(e^{\pm}, \mu^{\mp})/\pi$	1.03
$p_Z(e^{\pm}, \mu^{\mp}, j)$ (GeV)	2.05
$m_T(e^{\pm}, \mu^{\mp}, j, p_T^{miss})$ (GeV)	5.58
$m(e^{\pm}, \mu^{\mp}, j)$ (GeV)	3.34

TABLE 4.6: Condition number for reponse matrices of the variables under studied. It has been verified that condition number of replicas corresponding to variations of the systematic uncertainties are of the same order as the nominal ones.

4.6.4 Results

The normalized differential tW cross section as a function of the observables under study are shown in figure 4.18. Fair agreement, within the uncertainties, is observed with respect to the POWHEG DR, POWHEG DS and MADGRAPH5_AMC@NLO. The leading systematic uncertainties affecting to each bin of the distributions are shown in figure 4.19. The leading uncertainties in these measurements are due to JES and JER, whose impact is large due to its effect in the background estimation.

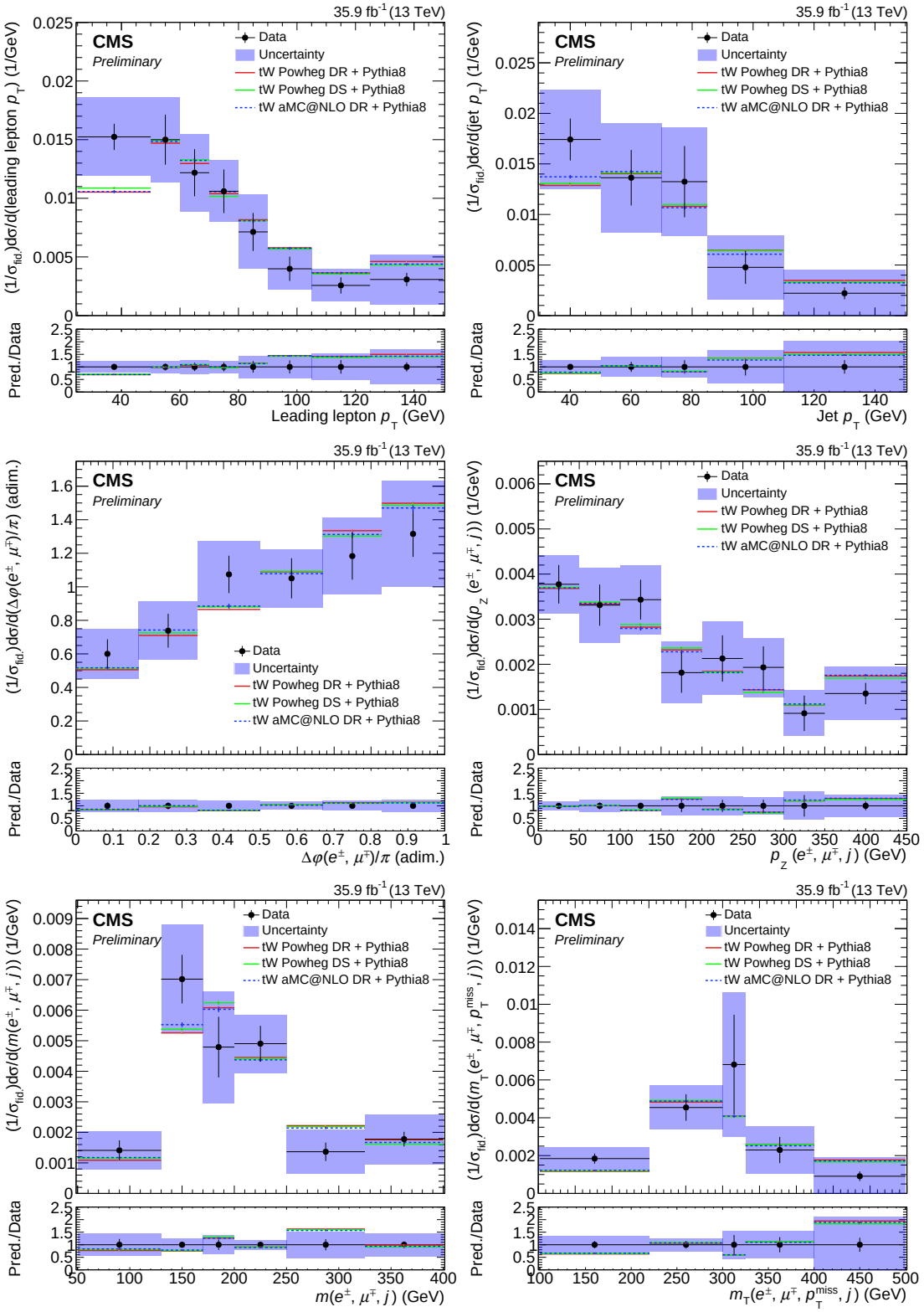


FIGURE 4.18: Observed distribution of the variables under study unfolded to particle level. Solid bands represent the total uncertainty of the measurement. Predictions of MADGRAPH MC@NLO (blue), POWHEG DR (red) and POWHEG DS (green) are also shown.

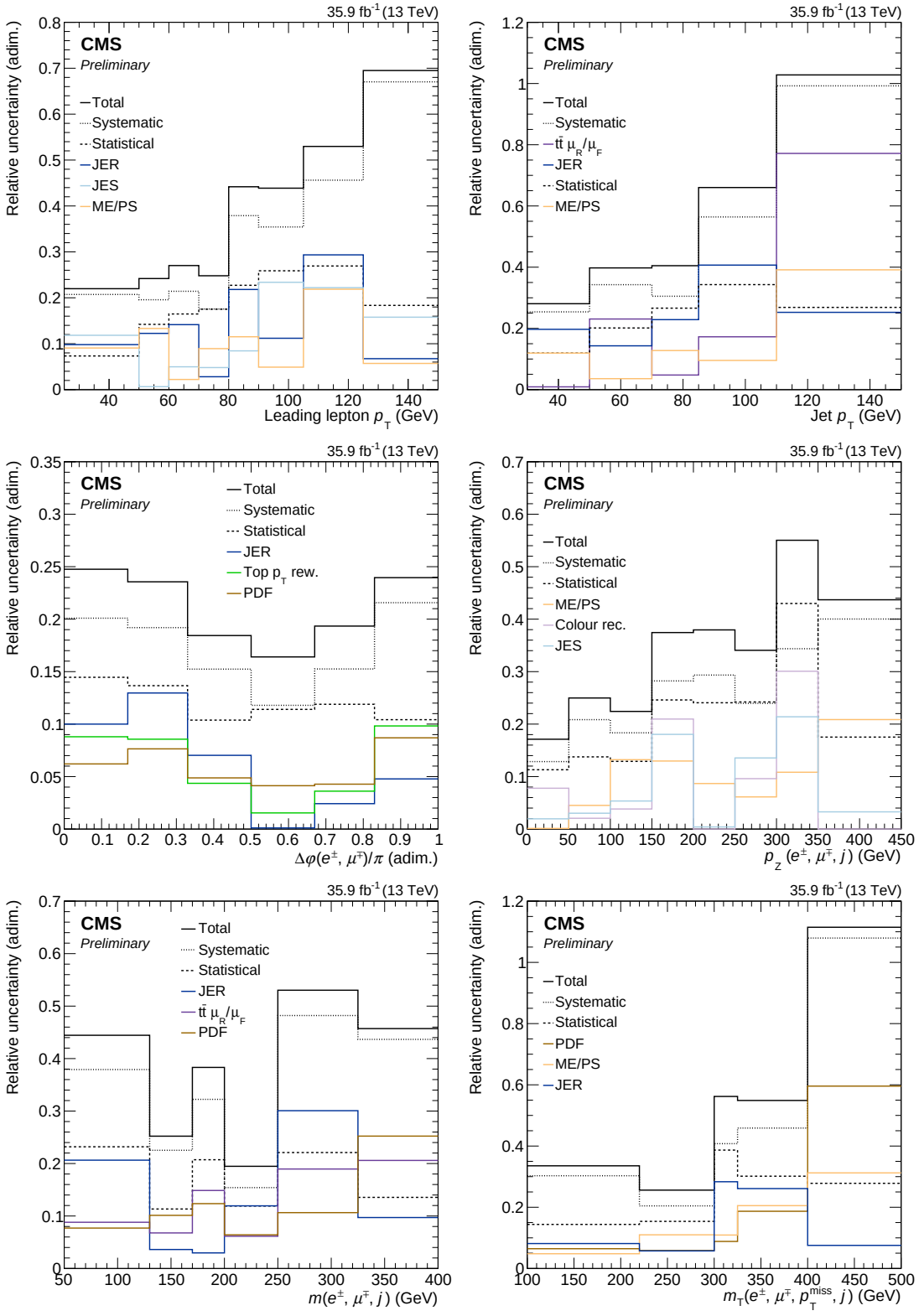


FIGURE 4.19: Breakdown of the systematic uncertainties affecting to the measurement of the normalized differential cross section.

4.7 Conclusions

Measurements of the tW production cross section have been performed in pp collision events at $\sqrt{s} = 13$ TeV collected with the CMS detector during 2016, corresponding to an integrated luminosity of 35.9 fb^{-1}. The measurements are performed in events with a $\mathrm{e}^{\pm}\mu^{\mp}$ pair and at least one b-tagged jet.

These measurements are intimately related with high precision measurements of the $\mathrm{t\bar{t}}$ process, since the two processes interfere and are therefore irreducible backgrounds to each other. In the case of tW measurements, a precise modeling of $\mathrm{t\bar{t}}$ events is crucial, due to its larger cross section.

The inclusive tW production cross section is measured making use of multivariate techniques, that allow to discriminate signal from $\mathrm{t\bar{t}}$ events. The signal is extracted performing a maximum likelihood fit to the distribution of events in the score of these discriminants. The inclusive cross section is measured to be 63.1 $\pm$ 1.8 (stat) $\pm$ 6.3 (syst) $\pm$ 2.1 (lumi) pb, consistent with the SM prediction. This measurement has been published in [9].

Normalized differential tW production cross sections are also measured in a fiducial region enriched in signal events. The measurements are performed as a function of several properties of the event: the transverse momentum of the leading lepton; the transverse momentum of the jet; the difference in the ϕ angle of the muon and the electron; the longitudinal momentum of the muon, the electron and the jet; the invariant mass of the muon, the electron and the jet; and the transverse mass of the electron, the muon, the jet, and the $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$.

The dominating uncertainties to these measurements are experimental uncertainties, such as jet energy scale and trigger and lepton efficiency, that have a significant impact in the final measurements due to their effect on the modeling of $\mathrm{t\bar{t}}$ events.

Both the inclusive and differential measurements are consistent with the prediction from the various NLO predictions. This result has been released in [10].

5

Search for new physics in events with two opposite-sign same-flavor leptons

This chapter covers searches for new physics in events with two opposite-sign same-flavor (OSSF) leptons and missing transverse momentum in the final state. The search is performed in pp collision evens at $\sqrt{s} = 13$ TeV recorded by the CMS detector in 2016, 2017 and 2018.

The chapter is organized as follows. Section 5.1 will cover the characteristics of the topology in which the search is made, as well as the potential signals featuring this topology. The status of these searches at the moment in which this thesis was written will also be shown. The simulated datasets used in the search are briefly described in section 5.2. The event selection and Signal Region (SR) definitions will be described in section 5.3, and the background estimation methods will be covered in section 5.4. The results of the search will be shown in section 5.5 and their interpretation on the context of supersymmetric simplified models, in section 5.6.

Due to the complexity of the research performed in high energy physics, the work shown in this chapter has been done in collaboration with groups of UCSD, RTWH Aachen, ETH Zürich, CERN and IFCA, all within the CMS Collaboration. Besides leading the development of the full Run 2 analysis, my personal contributions to the analysis include the development of the $\mathrm{t\bar{t}}$ likelihood discriminant, and the definition and tuning of various signal regions in both the electroweak and strong production modes. I also redesigned the factorization method to estimate the flavor-symmetric background in the full Run 2 analysis, and performed the signal extraction for some of the considered models.

5.1 Signal models and previous searches

This search looks for supersymmetric processes in which an opposite-sign same-flavor pair of leptons and missing transverse momentum are produced. As mentioned in chapter 1, momentum imbalance is one of the characteristics of R-parity conserving SUSY models, in which an invisible LSP is present in the final state. In such models, SUSY particles are produced in pairs, either via the strong interaction, producing squarks or gluinos, or via the electroweak interaction, producing charginos, neutralinos and sleptons.

Leptons may appear in the decay chains of these SUSY particles. Although the branching ratios to final states with leptons are typically smaller than hadronic final states in many SUSY models, this topology is relatively easy to trigger on and are affected by limited SM backgrounds. In particular, searches with leptons in the final state are the most sensitive for electroweak SUSY production and have at least similar sensitivity for strong searches in most of the cases.

An OSSF pair can be produced in the three-body-decay of a second neutralino

$$\widetilde{\chi}_2^0 \to l^{\pm}\widetilde{l}^{\mp} \to l^{\pm}l^{\mp}\widetilde{\chi}_1^0$$
$$\widetilde{\chi}_2^0 \to Z\widetilde{\chi}_1^0 \to l^{\pm}l^{\mp}\widetilde{\chi}_1^0.$$

A similar topology may occur in Gauge Mediated Supersymmetry Breaking (GMSB) scenarios, in which the gravitino, the spartner of the hypothetical graviton, is the LSP:

$$\widetilde{\chi}_1^0 \to Z\widetilde{G} \to l^{\pm}l^{\mp}\widetilde{G}.$$

Different kinematic features occur in each one of the decay modes, and depend on the mass splitting of the neutralinos. In the decay modes mediated by a Z boson, if this mass splitting is larger than the Z boson mass, the resonant contribution dominates and an excess of events containing a Z boson is observed. Otherwise, a resonance with a kinematic endpoint depending on the mass splitting of the neutralinos is observed. This resonance is referred to as a kinematic edge.

The $\widetilde{\chi}_2^0$ may have been produced via direct production or in the decays of strongly produced SUSY particles. This will condition the topology of the event, as additional objects may be produced in the decay of these particles. In this search, several signal regions are defined targeting different topologies.

Another possible topology with OSSF leptons is the following. Since SUSY particles in R-parity conserving models are produced in pairs, the pair of opposite-sign same flavor pair can occur in the direct pair production of sleptons:

$$pp \to \widetilde{l}^{\pm}\widetilde{l}^{\mp} \to l^{\pm}l^{\mp}\widetilde{\chi}_1^0\widetilde{\chi}_1^0.$$

In this case, since the two leptons are produced in different decay chains, the two leptons are mostly uncorrelated and no resonance-like features are observed in the m_{ll} distribution.

Several simplified models involving these topologies are considered to define and optimize the signal regions of the analysis, as well as to interpret the results.

5.1.1 Models with a Z candidate

As mentioned above, processes with SUSY particles and a Z boson in the final state can occur in the decay of a neutralino, that may have been produced with the strong and electroweak interaction.

To study the strong interaction, a simplified model inspired in GMSB SUSY is considered. This model will be referred to as the GMSB scenario, and shows direct pair gluino production. Each one of the gluinos decays into a pair of quarks and a $\widetilde{\chi}_1^0$, which further decays into a Z boson and the massless gravitino. Considering the leptonic decay of one of the Z bosons and the hadronic decay of the remaining one, events in this topology feature an OSSF lepton pair, 6 jets, and missing transverse momentum due to the gravitinos. The kinematic properties of these objects depend on the gluino and $\widetilde{\chi}_1^0$ masses, which are the free parameters of the model. The Feynman diagram for this process is shown in figure 5.1.

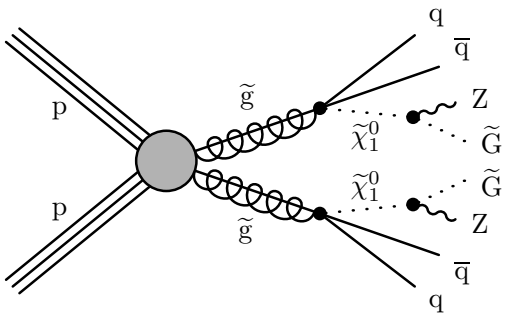


FIGURE 5.1: Feynman diagram for the simplified model of gluino pair production, considered as a benchmark for searches of strong production SUSY with a Z candidate.

Two electroweak production models are considered: direct neutralino pair production ($\widetilde{\chi}_1^0\widetilde{\chi}_1^0$) and the associated production of a neutralino and a chargino ($\widetilde{\chi}_2^0\widetilde{\chi}_1^\pm$).

In the first case, a GMSB model with a massless gravitino is considered [132, 133], in which the phenomenology of the model is fully determined by the NLSP. The topology we are interested in appears when the NLSP is a higgsino-like $\widetilde{\chi}_1^0$, as a gaugino-like would decay into a $\gamma\widetilde{\text{G}}$ with a high rate. We consider a set of higgsino-like $\widetilde{\chi}_1^0$, $\widetilde{\chi}_2^0$ and $\widetilde{\chi}_1^\pm$, that are approximately degenerate, while the rest of sparticles are decoupled. Under this model, $\widetilde{\chi}_2^0$ and $\widetilde{\chi}_1^\pm$ would decay promptly into $\widetilde{\chi}_1^0$ emitting additional low energy partons, that can be neglected. Then, the effective total cross section for $\widetilde{\chi}_1^0\widetilde{\chi}_1^0$ is dominated by $\text{pp} \to \widetilde{\chi}^0{}_i\widetilde{\chi}^0{}_j$ with $i,j = 1,2$ and $\text{pp} \to \widetilde{\chi}_1^+\widetilde{\chi}_1^-$, with all the modes contributing similarly.

The branching fractions of the $\widetilde{\chi}_1^0$ depend on its higgsino content, as well as its mass. Below and slightly above the kinematic threshold for Z or H production, the decay is dominated by photons. In this measurement, we restrict ourselves instead to extreme cases. In the first of them, the neutralino always decays into a Z boson and the massless gravitino. In this case, the final state would present an OSSF lepton pair and two jets produced in the decay of the two Z bosons and missing transverse momentum due to the gravitino. In the second scenario, the neutralino decays 50% of the times into a Z boson and a gravitino, and 50% of the times into a H boson and a gravitino. In this scenario, 50% of the events will have a HZ final state, 25% a ZZ final state, and in the remaining 25%, a HH final state, all of them with $p_{\text{T}}^{\text{miss}}$ in the final state. Feynman diagrams for these two scenarios are shown in figure 5.2. The free parameter of this model is the mass of the neutralino.

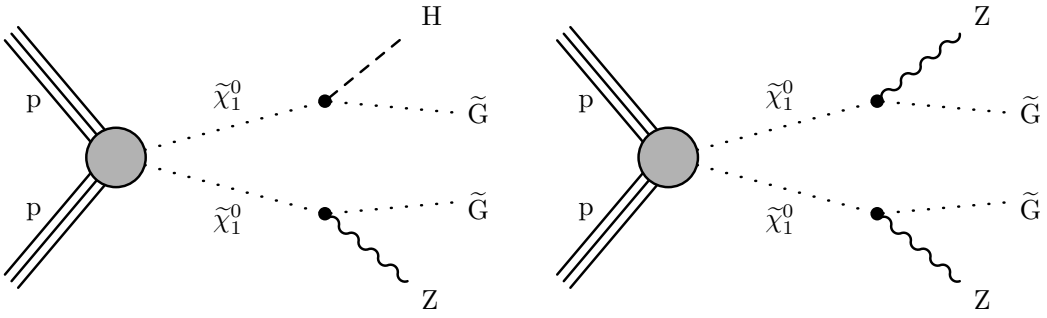


FIGURE 5.2: Feynman diagram for the simplified model of neutralino pair production, considered as one of the benchmarks for searches of electroweak production SUSY with a Z candidate. Left: one of the neutralinos decays into a $Z\widetilde{G}$ and the other to $H\widetilde{G}$. Right: both neutralinos decay to $Z\widetilde{G}$.

In the $\widetilde{\chi}_2^0\widetilde{\chi}_1^\pm$ production model, the second neutralino decays into the lightest neutralino and a Z boson, while the chargino decays into a W boson and a lightest neutralino. Targeting the hadronic decay of the W boson, the final state present shows an OSSF lepton pair, two jets and p_T^{miss}. A feynman diagram for this model is shown in figure 5.3. The parameters of this model are the mass of the chargino and second neutralino, that are assumed to be degenerate, and the mass of the first neutralino.

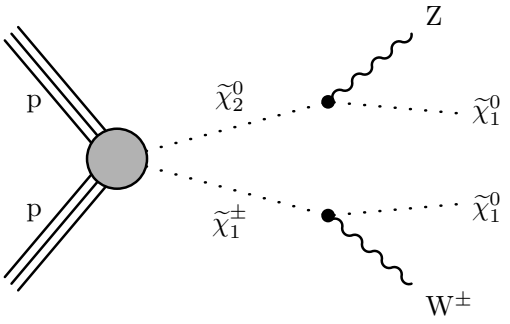


FIGURE 5.3: Feynman diagram for the simplified model of neutralino chargino production, considered as one of the benchmarks for searches of electroweak production SUSY with a Z candidate.

5.1.2 Models with a resonant edge

The direct production of an squark and its subsequent decay into a quark and a neutralino can induce the resonant edge topology:

$$\widetilde{q} \to q\widetilde{\chi}_2^0 \to ql^\pm l^\mp \widetilde{\chi}_1^0. \tag{5.1}$$

The decay of the $\widetilde{\chi}_2^0$ can occur in the two ways described above. If the on-shell Z boson is not accessible because of the small mass splitting between $\widetilde{\chi}_2^0$ and $\widetilde{\chi}_1^0$, two situations can occur.

If $m_{\widetilde{\chi}_2^0} - m_{\widetilde{\chi}_1^0} < m_Z$ and $m_{\widetilde{\chi}_2^0} - m_{\widetilde{\chi}_1^0} < m_{\widetilde{l}}$, the only possible decay mode is through an off-shell Z boson. Then, the m_{ll} distribution shows a kinematic endpoint at, m_{ll}^{edge}, given by

$$m_{ll}^{edge} = m_{\widetilde{\chi}_2^0} - m_{\widetilde{\chi}_1^0}.$$

If $m_{\widetilde{l}} < m_{\widetilde{\chi}_2^0} - m_{\widetilde{\chi}_1^0} < m_Z$, the decay mediated by an slepton is kinematically allowed and the decay occurs through two sequential two-body decays. In that case, the kinematic endpoint is given by

$$m_{ll}^{edge} = \frac{\sqrt{\left(m_{\widetilde{\chi}_2^0}^2 - m_{\widetilde{l}}^2\right)\left(m_{\widetilde{l}}^2 - m_{\widetilde{\chi}_1^0}^2\right)}}{m_{\widetilde{l}}}.$$

These kind of models are particularly interesting because, if a signal was to be measured, the position of the edge resonance would give direct information on the masses of the sparticles produced in the decay chain. Figure 5.4 shows the Feynman diagram for the simplified model considered, the so-called slepton-edge model. It features the production of an squark pair, with the subsequent decay described in the reaction 5.1. In the search, two variations of this model are considered, one in which the squark is an sbottom and another in which it is a light flavor squark. In this model the mass of the $\widetilde{\chi}_1^0$ is assumed to be 100 GeV and the mass of the slepton to be $0.5(m_{\widetilde{\chi}_2^0} + m_{\widetilde{\chi}_1^0})$. The squark and $\widetilde{\chi}_2^0$ masses are free parameters of the model.

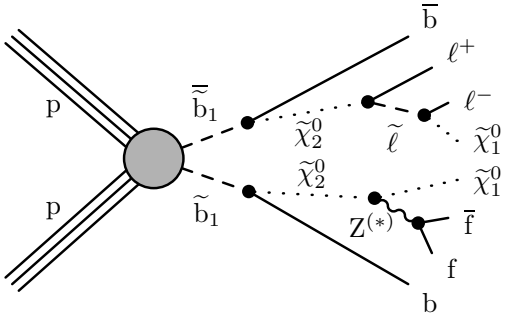


FIGURE 5.4: Feynman diagram for the slepton edge model

5.1.3 Non-resonant models

This study also considers the search for a direct production of a pair of sleptons (selectrons and smuons). In the model considered, a pair of left-handed or right-handed sleptons is produced. Each one of them subsequently decays into a lepton of the same

flavor and a $\widetilde{\chi}_1^0$, the LSP in this model. A Feynman diagram for this process is shown in figure 5.5. Both the slepton and the $\widetilde{\chi}_1^0$ masses are free parameters of the model.

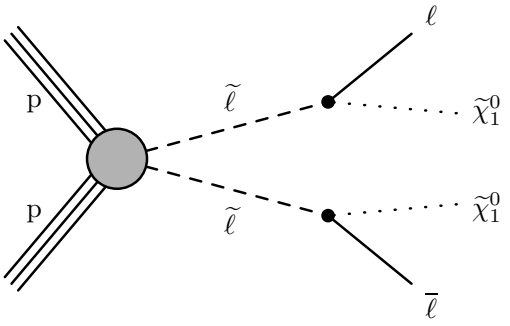


FIGURE 5.5: Feynman diagram for the direct slepton production model considered.

5.1.4 Previous searches

I have contributed to three publications of SUSY searches performed in this final state with Run 2 data. The first set of pp collisions at 13 TeV, corresponding to 2.3 fb^{-1}, collected during 2015 with CMS allowed to search for strong production models, profiting from the increase in cross section at the high energy achieved [11]. Then this search was performed with the luminosity collected during 2016, and regions targeting including electroweak production models [12]. The electroweak production limits obtained in the latter publications were combined with other decay channels, resulting in an increased sensitivity [13].

Additionally, CMS released a search for direct production of sleptons with the same dataset [134], using similar selections and background estimation techniques.

Similar searches were also performed in CMS with Run 1 data. At 8 TeV searches for strong [135] and electroweak [136] production of SUSY in these topologies were performed. A combination of electroweak production searches was also performed in [137].

The searches covered in this chapter are based on the searches that I contributed to and are mentioned above, as well as the direct slepton pair production search. The search is performed in pp collisions collected across the 2016, 2017 and 2018 years, with a total integrated luminosity of 137.2 fb^{-1}. This luminosity allows for higher sensitivity to lower cross section signals because of the reduction of the statistical uncertainties and the refinement of the background estimation and rejection techniques.

5.2 Signal and background simulations

Although the background estimation techniques for the main backgrounds are based on data, simulated events for specific processes are also used. These are needed to

validate the data-driven background prediction techniques, to model smaller backgrounds not covered by these techniques, and to estimate the contribution from potential signals in the analysis.

Most background processes are simulated using the same generators as described in section 4.3. Dedicated simulated samples are used for each one of the years of the data taking, in order to account for the different detector geometry and pile-up conditions. The PYTHIA tune used for 2017 and 2018 simulations is CP5 [138], while the one used in 2016 is CUETP8M1 for $\mathrm{t\bar{t}}$ events and CUETP8M2T4 for the rest. Contribution from γ+jets events in the photon control regions is simulated at LO with the MADGRAPH generator.

For this analysis, since larger contribution from ZZ processes is expected, a more careful treatment is performed of this process. $\mathrm{q\bar{q}} \rightarrow \mathrm{ZZ}$ and $\mathrm{gg} \rightarrow \mathrm{ZZ}$ are modeled separately. $\mathrm{q\bar{q}} \rightarrow \mathrm{ZZ}$ is simulated at NLO using POWHEG 2.0. Generator-level p_{T} dependent *k*-factors are assigned to take into account NNLO/NLO differences [139]. $\mathrm{gg} \rightarrow \mathrm{ZZ}$ is simulated using MCFM 7.0 [140] at LO, and is normalized to the NLO calculations [141].

Signals are simulated at LO precision with the MADGRAPH MC@NLO generator with up to two additional partons in the matrix element calculation. Events are then interfaced to PYTHIA v8.212 for fragmentation and hadronization. The detector simulation is performed using the CMS fast simulation package [48]. Signal simulations are normalized to NLO+Next-to-leading logarithm (NLL) calculations [142–144].

5.3 Signal and control region definitions

In the section, the definition of signal and control regions is described.

5.3.1 Object selection

Muons and electrons are the most important objects for this analysis. The driving principle of the muon and electron selection is to keep a high efficiency while having a similar muon and electron efficiency within a reasonable purity. This is done to enhance the performance of the flavor-symmetric background estimation methods, described in section 5.3.

Muons are required to pass the "medium" selection criteria, described in section 3.1.2. Electrons are required to pass a MVA-based selection (described in section 3.2), designed to keep a high efficiency and low acceptance rates of electrons in jets. Muons and electrons in the transition region of the ECAL are rejected, to ensure their reconstruction efficiencies are similar.

Tracks associated to electrons and muons are required to have an impact parameter of less than 0.5 mm in the transverse plane and less than 1 mm in the direction along the beam direction. They are also required to be isolated from the rest of PF candidates in the event. In order to do that, the mini-isolation variable is used, which is required to be less than 10% of the lepton p_{T} for electrons and 20% for muons.

Jets are selected using the criteria described in section 2.3.2, and are required to have a p_T greater than 35 GeV. This threshold is relaxed to 25 GeV to veto jets and b jets. Since leptons are usually clustered as jets by the anti-k_T algorithm, jets within a $\Delta R < 0.4$ of selected leptons are not included in the counting. Jets produced in the parton showers of b quarks are tagged using the deepCSV algorithm, described in section 2.3.3.

Some of the signal regions look for the presence of hadronically decaying Z or W bosons, whose decay products are collimated into a single jets. For that purpose, jets are clustered with the anti-k_T algorithm and a radius parameter of 0.8. Z/W candidates are then required to have a soft-drop mass between 65 and 105 GeV, and $\tau_2/\tau_1 < 0.4$ (0.45) in 2016 (2017 and 2018) data. This way only jets are selected that have a mass consistent with the W and Z bosons and have the characteristic 2-prong substructure expected in their decays.

Additionally, since the presence of additional leptons must be vetoed in some signal regions, in order to build suitable control regions or to be independent with other CMS analyses. In order to do that, events with isolated PF candidates are rejected. In order to be considered for this veto, the track-based relative isolation with a cone size of $\Delta R < 0.3$ is required to be less than 20% of the track momentum and less than 5 GeV. Additionally, the track is required to have an impact parameter with respect to the primary vertex less than 0.1 cm and 0.2 cm relative to the primary vertex.

Photons are required to pass identification criteria based on the shape of the cluster in the ECAL and the fraction of energy deposited in the ECAL [145]. Photons are required to have $p_T > 50$ GeV, excluding the transition region of the ECAL.

5.3.2 Discriminating variables

The M_{T2} variable and a $t\bar{t}$ likelihood discriminator are used to select and classify events. These two variables are built to reject $t\bar{t}$ events, and are described in this section.

5.3.2.1 M_{T2}

The M_{T2} variable [146, 147] is used in many searches to reject backgrounds, as it aims to reconstruct the mass of a heavy particle, that has been produced in pairs, and decays into a visible and an invisible object. Given a pair of visible objects v_1 and v_2, and p_T^{miss} due to the presence of the two invisible objects, M_{T2} $(v_1 v_2)$ is defined as

$$M_{T2} = \min_{\vec{p}_T^{miss(1)} + \vec{p}_T^{miss(2)} = \vec{p}_T^{miss}} \left[\max \left(M_T \left(v_1, \vec{p}_T^{miss(1)} \right), M_T \left(v_2, \vec{p}_T^{miss(2)} \right) \right) \right], \tag{5.2}$$

where $\vec{p}_T^{miss(1,2)}$ are hypothesis vectors in the transverse plane, that consider all the possible configurations for the invisible pair of particles consistent with the observed $\vec{p}_T^{miss}$. $M_T \left(v_i, \vec{p}_T^{miss(i)} \right)$ is the transverse mass of the system composed by the visible object and the hypothesis for the invisible particle momentum.

If v_1 and v_2 are produced in association with two invisible objects in a decay of a pair of particles with mass M, $\mathrm{M_{T2}}$ ($v_1 v_2$) has a kinematic endpoint at M. In particular, in an event with two leptons and $p_{\mathrm{T}}^{\mathrm{miss}}$ one can consider the $\mathrm{M_{T2}}$ variable evaluated in the two leptons, $\mathrm{M_{T2}(ll)}$. This variable has a kinematic endpoint in the W mass for dileptonic $\mathrm{t\bar{t}}$ and WW events. Of course, due to the limited resolution of the detector, as well as off-shell effects of the produced particle, a significant contribution from those background events may contribute with values above the kinematic endpoint.

Another related variable is the $\mathrm{M_{T2}(lblb)}$ variable, which is constructed in events with a pair of leptons ($\mathrm{l_1}$ and $\mathrm{l_2}$) and a pair of b-tagged jets ($\mathrm{b_1}$ and $\mathrm{b_2}$). To construct it, the two possible pairings between leptons and jets are considered, and $\mathrm{M_{T2}(l_1 + b_1, l_2 + b_2)}$ and $\mathrm{M_{T2}(l_1 + b_2, l_2 + b_1)}$ are calculated. The minimum of the two is defined as $\mathrm{M_{T2}(lblb)}$. This variable has, for dileptonic $\mathrm{t\bar{t}}$ events, a kinematic endpoint at the mass of the top quark.

5.3.2.2 $\mathrm{t\bar{t}}$ discriminator

A $\mathrm{t\bar{t}}$ likelihood discriminator is used to enhance the sensitivity of slepton-edge search. This discriminator is a naive Bayes classifier, that categorizes events as $\mathrm{t\bar{t}}$-like and non-$\mathrm{t\bar{t}}$ like. Since no signal is used in the design of the discriminator, it can be considered signal-agnostic.

Several variables, that are characteristics of the $\mathrm{t\bar{t}}$ topology are chosen as inputs to the multivariate discriminator:

- $p_{\mathrm{T}}^{\mathrm{miss}}$. Even if invisible particles are produced both in $\mathrm{t\bar{t}}$ and signal, SUSY models with large mass splitting predict a harder $p_{\mathrm{T}}^{\mathrm{miss}}$ distribution than that of $\mathrm{t\bar{t}}$ events.
- The $\sum m_{\mathrm{lb}}$ variable. This variable is constructed for events with two leptons and at least two jets, and is defined as the sum of the invariant masses of two pairs, each one composed by a lepton and a jet. The first one of the pairs is selected by considering all possible pairings between the two leptons and the b jets in the event. If there are no b-tagged jets in the event, all the jets are considered instead. The pair with the smallest invariant mass is taken. To select the second pair, the procedure is repeated using the lepton and jets not included in the first pair.
- $p_{\mathrm{T}}^{\mathrm{ll}}$. This variable is typically large for signal events, since the two leptons are produced in the decay chain of the sbottom, which typically has a high mass and, therefore, its decay products are usually collimated. In $\mathrm{t\bar{t}}$ events, in the other hand, the two leptons belong to different decay chains and are then mildly correlated.
- $|\Delta\phi^{\mathrm{ll}}|$, due to the boost of system mentioned above, leptons in signal events are also expected to be quite close-by.

The correlation matrix of these variables in $\mathrm{t\bar{t}}$ simulations is shown in figure 5.6. Mild correlations among $|\Delta\phi^{\mathrm{ll}}|$, $p_{\mathrm{T}}^{\mathrm{ll}}$ and $\sum m_{\mathrm{lb}}$ are observed, while $p_{\mathrm{T}}^{\mathrm{miss}}$ is quite uncorrelated with respect to the others.

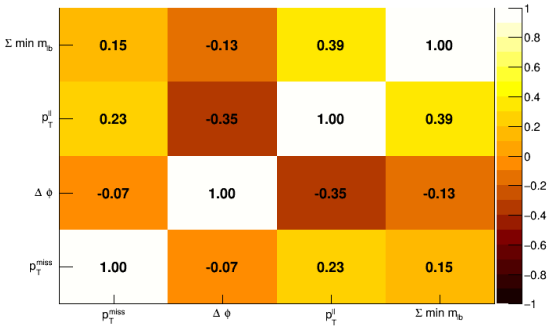


FIGURE 5.6: Correlation matrix of the input variables to the likelihood discriminator variables in simulated $\mathrm{t\bar{t}}$ events.

The probability density function (pdf) for these variables is extracted from data events with an $\mathrm{e\mu}$ pair, which is expected to be enriched in $\mathrm{t\bar{t}}$ events. An ad-hoc model is used, assuming an analytic expression for these variables. The validity of these models is corroborated by a Kolmogorov-Smirnov goodness-of-fit test to data. $p_{\mathrm{T}}^{\mathrm{miss}}$ is modeled using a sum of exponential functions, a Crystal ball for $\sum m_{\mathrm{lb}}$ and $p_{\mathrm{T}}^{\mathrm{ll}}$, and a third order polynomial for $|\Delta\phi^{\mathrm{ll}}|$. The pdfs for the variables have been determined separately for each one of the three years of the data taking, and the $\mathrm{t\bar{t}}$ likelihood discriminator is constructed separately for each year.

The discriminator is defined, for a given event, as

$$\mathrm{nll} = -\sum_i \log f_i(x_i),$$

where x_i are the input variables described above and f_i, the fitted pdf in $\mathrm{t\bar{t}}$ data.

Figure 5.7 shows the distribution of the discriminator score in $\mathrm{t\bar{t}}$ simulated events and signal for several sparticle masses. The distributions of the score in simulations reproducing the conditions of each year of the data taking have a good agreement among the three, as it is shown in the figure, which shows the robustness of the method.

Based on the $\mathrm{t\bar{t}}$ discriminator, events are classified as $\mathrm{t\bar{t}}$-like or as non $\mathrm{t\bar{t}}$-like if the discriminator score is larger or smaller than 24, respectively. This working point has been chosen to provide the largest signal sensitivity to the slepton-edge model, while keeping enough amount of data events in the control regions.

5.3.3 Signal regions

Events in the signal and dileptonic control regions are collected with a set of triggers requiring the presence of a pair of leptons passing mild isolation criteria and a given p_{T} threshold, which depends on the lepton flavor and the period of the data taking. For the leading lepton, this threshold ranges between 23 and 17 GeV, while for the subleading lepton, the threshold ranges between 12 and 8 GeV. To recover efficiency

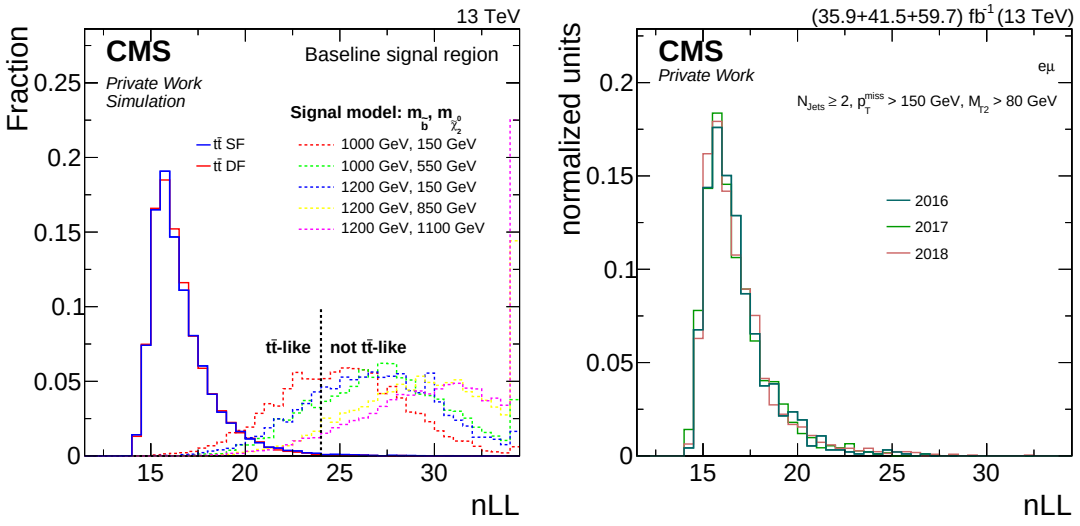


FIGURE 5.7: Left: distribution of the $t\bar{t}$ likelihood discriminant in $t\bar{t}$ SF (blue) and DF (red) events, as well as for various signal models (dashed lines). Right: distribution of the $t\bar{t}$ likelihood in simulations of the 2016, 2017 and 2018 datasets.

in cases in which the leptons are emitted close-by to other objects, which is the typical signature in boosted topologies, another set of triggers without isolation requirement is also used. The momentum thresholds for the leading and sub-leading leptons range between 37 and 25 GeV and 33 and 8 GeV, respectively. No single lepton triggers are used, since it is desired to keep a symmetry between electron and muon reconstruction efficiencies.

Events collected by these triggers are also required to pass a minimal dileptonic event selection. Events are required to have at least two reconstructed leptons passing the requirements described in section 5.3.1. The leading of these leptons is required to have $p_{\mathrm{T}} > 25$ GeV and the subleading $p_{\mathrm{T}} > 20$ GeV. The two leading leptons are required to have opposite sign. Then, events are classified based on the flavor of these two leptons between same-flavor (SF) and different-flavor (DF). SF events are used to define the signal regions, while DF events are only used for some of the control regions.

The p_{T} of the system formed by the two leading leptons, $p_{\mathrm{T}}^{\mathrm{ll}}$, is required to be larger than 50 GeV, in order to be consistent with the p_{T} cut applied to the photons, to ensure the consistency of the $p_{\mathrm{T}}^{\mathrm{miss}}$ "templates" method, described in section 5.4. Events for which the azimuthal angle between the two leptons, $|\Delta\phi^{\mathrm{ll}}|$, is less than 0.1 are rejected, to keep similar electron and muon isolation efficiencies. The invariant mass of the two leptons, m_{ll}, is required to be greater than 20 GeV, to reject contributions from low mass resonances.

Signal regions are defined aiming to be sensitive to the simplified models described in the previous section. However, it is aimed to keep sensitivity to other models besides those.

5.3.3.1 On-Z regions

Events are required, in order to enter in the on-Z regions, to have an m_{ll} consistent with the production of a Z boson, between 86 GeV and 96 GeV. Events containing additional isolated PF candidates passing the criteria described for the veto are rejected. To reject events with contributions from instrumental $p_{\mathrm{T}}^{\mathrm{miss}}$, the two jets with the highest p_{T} are required to be separated from $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ in ϕ by at least 0.4 radians.

Six disjoint SRs are defined in the search for strong production of SUSY with a Z candidate. SRA, SRB and SRC are composed by events with 2-3 jets, 4-5 jets and 6 or more reconstructed jets, respectively. In the GMSB scenario, up to 6 jets are expected, however SRA and SRB provide additional sensitivity to events in which some of the jets are outside the acceptance. These regions are further divided between those containing at least one b jet or those that do not have b jets. Events in regions without b jets are required to have $H_{\mathrm{T}} > 500$ GeV and $\mathrm{M_{T2}(ll)} > 80$ GeV, while the ones with b jets are required to have $H_{\mathrm{T}} > 200$ GeV and $\mathrm{M_{T2}(ll)} > 100$ GeV, to more efficiently suppress $\mathrm{t\bar{t}}$ background. These criteria are summarized in table 5.1.

Region	n_j	n_{b}	H_{T} [GeV]	$\mathrm{M_{T2}(ll)}$ [GeV]	$p_{\mathrm{T}}^{\mathrm{miss}}$ binning [GeV]
SRA b veto	2–3	=0	>500	>80	[100,150,230,300,>300]
SRB b veto	4–5	=0	>500	>80	[100,150,230,300,>300]
SRC b veto	≥6	=0	—	>80	[100,150,250,>250]
SRA b tag	2–3	≥1	>200	>100	[100,150,230,300,>300]
SRB b tag	4–5	≥1	>200	>100	[100,150,230,300,>300]
SRC b tag	≥6	≥1	—	>100	[100,150,250,>250]

TABLE 5.1: Summary of the strong on-Z regions. Events are required to have $86 < m_{\mathrm{ll}} < 96$ GeV.

For the electroweak production, three regions are built. The first two target the WZ and ZZ topologies, in which the W and one of the Z bosons decay hadronically, and the other one, leptonically. The third one targets the HZ final state, with the Z boson decaying leptonically and the Higgs boson into a $\mathrm{b\bar{b}}$ pair.

The first region, named "VZ resolved", aims for the cases in which the two jets produced in the Z or W decays are reconstructed individually. For this region, events are required to have at least two jets. Events with b jets or $\mathrm{M_{T2}(ll)} < 80$ GeV are rejected, to reduce the contribution from $\mathrm{t\bar{t}}$ events. Then, the mass of the two jets that are closer in $\Delta\phi$, m_{jj}, is required to be smaller than 110 GeV, to be consistent with the hadronic decay of the W and Z boson.

The second region, referred to as "VZ boosted", recovers sensitivity to cases in which the W or Z boson is produced with a large boost in the transverse plane and the two jets produced in the decay are merged into a single one. For this region, a W/Z candidate with $p_{\mathrm{T}} > 200$ GeV is required in the final state. The W/Z candidate is required to be separated from $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ with a $\Delta\phi > 0.8$. Events with b-tagged jets are also rejected. In addition, events with at least two jets are rejected in order to avoid overlap with the resolved region.

The third region, the HZ region, is designed to be sensitive to the cases in which the Z boson decays leptonically and the H boson decays through the $\mathrm{b\bar{b}}$ mode, which is the one with the dominant branching ratio. Events in this category are required to have at least two b-tagged jets with an invariant mass, m_{bb}, less than 150 GeV. In order to reduce the presence of $\mathrm{t\bar{t}}$ events, $\mathrm{M_{T2}(lblb)}$ is required to be greater than 200 GeV.

All the electroweak regions are split in $p_{\mathrm{T}}^{\mathrm{miss}}$ bins. The selection criteria for these regions is summarized in table 5.2.

Region	n_j (n_j^{fat})	n_{b}	Dijet mass [GeV]	$\mathrm{M_{T2}}$ [GeV]	$p_{\mathrm{T}}^{\mathrm{miss}}$ binning [GeV]
Boosted VZ	($\geq$1)	=0	—	—	[100,200,300,400,500,>500]
Resolved VZ	$\geq$2	=0	$m_{jj} < 110$	$\mathrm{M_{T2}(ll)} > 80$	[100,150,250,350,>350]
HZ	$\geq$2	=2	$m_{\mathrm{bb}} < 150$	$\mathrm{M_{T2}(lblb)} > 200$	[100,150,250, >250]

TABLE 5.2: Summary of the electroweak on-Z regions. Events are required to have 86 $< m_{\mathrm{ll}} <$ 96 GeV. Events in the resolved VZ region are vetoed from the boosted VZ region.

5.3.3.2 Off-Z regions

Two different signal regions are constructed aiming for the slepton-edge model and direct slepton pair production, respectively.

For the slepton-edge signal region, events are expected to have at least two jets produced in the decay of the squarks, plus two additional leptons produced in their decay chain. Events are required to have at least two jets and $p_{\mathrm{T}}^{\mathrm{miss}} > 150$ GeV. To suppress events with instrumental $p_{\mathrm{T}}^{\mathrm{miss}}$, it is required that the two jets with the highest p_{T} have a $\Delta\phi$ with respect to $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ greater than 0.4. In order to suppress the dominant background, composed by $\mathrm{t\bar{t}}$ events, $\mathrm{M_{T2}(ll)}$ is required to be larger than 80 GeV.

Then events are divided in 28 exclusive signal regions. Events are classified depending on their m_{ll} value in 7 categories: [20-60], [60-86], [96-150], [150-200], [200-300], [300-400] and greater than 400 GeV. Events with m_{ll} between 86 and 96 GeV are not included in any of these categories, and they are not taken into account in the analysis. For each m_{ll} bin, events are classified according to the b-tagged jet multiplicity: one category is built for events with at least one b-tagged jet and another for events with no b-tagged jets. Events are also classified as $\mathrm{t\bar{t}}$-like and non $\mathrm{t\bar{t}}$-like based on the likelihood discriminant described in the previous section.

Regions aiming for direct slepton production are expected to have two OSSF leptons and missing transverse momentum in the final state, and no or little hadronic activity. In these regions, events with additional isolated PF candidates are vetoed. In order to reject contribution from Drell–Yan events, the leading lepton is required to have $p_{\mathrm{T}} >$ 50 GeV. Besides, m_{ll} must be lower than 65 GeV or greater than 120 GeV. To reduce the presence of both Drell–Yan, $\mathrm{t\bar{t}}$ and WW events, events are required to have both $\mathrm{M_{T2}(ll)}$ and $p_{\mathrm{T}}^{\mathrm{miss}}$ larger than 100 GeV.

On top of this selection, two types of regions are defined. Most of the sensitivity is expected to come from events without jets in the final state. Therefore, a region

with no reconstructed jets is defined. In order to keep sensitivity to signal events in which low energy jets are present due to ISR radiation, another category with jets is built. In this category, at least one jet with $p_{\mathrm{T}} > 25$ GeV must be present, but the quotient $p_{\mathrm{T}}^{lep2}/p_{\mathrm{T}}^{jet1}$ is required to be greater than 1.2. The leading jet is required to have $\Delta\phi(j, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}) > 0.4$ in this region. These selection criteria are summarized in table 5.3.

Region	n_j	n_{b}	$p_{\mathrm{T}}^{lep2}/p_{\mathrm{T}}^{jet1}$	$\mathrm{M_{T2}}$ [GeV]	$p_{\mathrm{T}}^{\mathrm{miss}}$ binning [GeV]
Slepton jet-less	=0	=0	—	$\mathrm{M_{T2}}(\mathrm{ll})$ >100	[100,150,225,300,>300]
Slepton with jets	>0	=0	> 1.2	$\mathrm{M_{T2}}(\mathrm{ll})$ >100	[100,150,225,300,>300]

TABLE 5.3: Summary of the regions for the direct slepton pair production search. Events are required to have m_{ll} smaller than 65 GeV or larger than 120 GeV.

5.3.4 Dileptonic control regions

Dileptonic control regions are defined to estimate some of the backgrounds. These regions are defined on top of the dilepton event selection defined in the previous section.

5.3.4.1 Different-flavor control regions

One of the main background processes are the so-called flavor-symmetric processes, those in which DF and SF lepton pairs are produced at the same rate. To estimate them, control regions are built by applying the same selection criteria as signal regions, but requiring a DF lepton pair instead of a SF lepton pair. As will be described in section 5.4, these events are used to estimate the contribution of flavor-symmetric processes to the signal region, after they have been corrected by the transfer factor.

Since DF and SF pairs are produced at the same rate, the transfer factor is expected to be close to the unity. Because of that, it is necessary to increase the statistical power of the control regions in cases in which the rate of flavor-symmetric processes is expected to be low. This is the case in regions aiming for SUSY with a Z candidate. In those regions, the amount of flavor-symmetric events in the control region is increased by removing the requirement on m_{ll} to be between 86 and 96 GeV.

5.3.4.2 DY+jets and $\mathrm{t\bar{t}}$ control regions

To develop some of the background estimation methods, it is necessary to consider data regions enriched in DY+jets events. Three different control regions are considered. One region is defined to derive the transfer factors between SF and DF events. For this region, leptons are required to have a SF lepton pair, at least two jets, $p_{\mathrm{T}}^{\mathrm{miss}}$ less than 50 GeV and with m_{ll} between 60 and 120 GeV. The other two regions are used to extrapolate DY+jets contribution from regions on-Z to signal regions off-Z. For this purpose a region is defined for the slepton-edge search, requiring events with at least

two jets, p_T^{miss} less than 50 GeV and $M_{T2}(ll)$ greater than 80 GeV. For the slepton regions, the control region is defined on top of their definition by removing the p_T^{miss} cut, the jet veto cut and the m_{ll} cut.

To validate the estimation of flavor-symmetric backgrounds, a $t\bar{t}$-enriched control region in the SF channel is defined. This region requires exactly two jets in the event, $m_{ll} < 70$ GeV or $m_{ll} > 110$ and $100 < p_T^{miss} < 150$ GeV, to reject contributions from DY+jets events.

5.3.4.3 p_T^{miss} templates validation regions

To validate the DY+jets estimation in the on-Z regions, in which it is the dominant background, DY+jets enriched regions are defined. These regions are built by inverting the cut that rejects events for which one of the two leading jets is closer in ϕ than 0.4 to $\vec{p}_T^{miss}$. In the case of the boosted VZ region, the inverted cut requires ϕ to be more than 0.8 between the W/Z candidate and $\vec{p}_T^{miss}$. Inverting this cut builds regions with a significant contribution from DY+jets events that have significantly large p_T^{miss} due to mismeasured jets.

5.3.5 Photon control regions

In signal regions with jets, the contribution from DY+jets events is estimated using a control region with reconstructed photons. In particular, these γ+jets events are used to determine the p_T^{miss} distribution of DY+jets, as in both cases their spectrum is expected to be driven by limited resolution of the jet momentum.

These events are collected using a set of single photon triggers, that select events containing a photon with at least 50 GeV. Only a fraction of these events is accepted by the trigger with an acceptance factor dependent on the p_T threshold, to keep their acceptance rate under an acceptable value. Events collected by these triggers are weighed by this fraction in order for their effective luminosity to match the luminosity collected in the rest of the regions.

Events with additional isolated charged PF candidates are vetoed, to suppress the contribution from electroweak processes, such as $W\gamma$, that may significantly populate the tails of the p_T^{miss} distribution in the control region.

The p_T^{miss} distribution may depend on the jet multiplicity and kinematics of the event, as well as on the presence of b jets in the event, since they may contain neutrinos produced in the hadron decays. Therefore a photon control region is built for each one of the signal regions applying to the jets in the event the same requirements that are applied in the signal region definition.

In regions with requirements on M_{T2}, it is not possible to directly apply the cuts on the γ+jets samples, since since the variable is constructed with two visible particles. In order to have a consistent definition of these variables in the γ+jets samples, the decay of the γ boson is emulated. This decay is performed assuming a mother particle with the mass of a Z boson, and a momentum corresponding to the reconstructed photon. We consider a system in which the mother particle is at rest. Then the decay of the

mother particle to leptons is simulated taking into account the angular correlation between them due to spin correlations in the matrix element. Analysis requirements on η and p_{T} of the leptons are also applied to the leptons obtained in the decay.

5.3.6 Control regions for WZ, ZZ and $\mathrm{t\bar{t}Z}$

The contribution of WZ, ZZ and $\mathrm{t\bar{t}Z}$ processes to signal regions is estimated using simulations. In order to verify the good modeling of these processes in simulations, dedicated control regions enriched in them are built.

Two types of control regions are constructed. The first set aims to check the modeling of the three processes in regions with jets, while the second kind measures the contribution from ZZ and WZ to the slepton regions, that only allow reduced hadronic activity.

In regions with jets, events are required to pass the minimal dilepton selection criteria, described in section 5.3.3, and to contain at least two jets with $|\Delta\phi(j_{1,2}, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})| > 0.4$. For the three regions, an OSSF lepton pair with $86 < m_{\mathrm{ll}} < 96$ GeV is required.

Events in the WZ region are required to have:

- At least 3 leptons. A Z candidate is built by picking the OSSF pair closer to the Z mass. The remaining lepton, l^{W}, is the W candidate.
- $p_{\mathrm{T}}^{\mathrm{miss}} > 70$ GeV,
- $\mathrm{M_T}(\mathrm{l}^{W}, p_{\mathrm{T}}^{\mathrm{miss}}) > 50$ GeV,
- no b-tagged jets with $p_{\mathrm{T}} > 25$ GeV in the loose working point,

For the $\mathrm{t\bar{t}Z}$ background, events are required to have $p_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV and at least two b-tagged jets with the medium working point.

For the ZZ background, events are required to have two OSSF pairs of leptons. The two pairs of leptons are required to have $m_{\mathrm{ll}} > 40$ GeV to reject low mass resonances. Events with b-tagged jets passing the loose working point are rejected.

In the regions without jets, only ZZ and WZ are relevant. In the WZ region, events are required to have

- at least 3 leptons. A Z candidate is built picking the OSSF pair closer to the Z mass. The remaining lepton, l^{W}, is the W candidate.
- The Z candidate is required to have m_{ll} between 76 and 106 GeV,
- $p_{\mathrm{T}}^{\mathrm{miss}} > 70$ GeV,
- $\mathrm{M_T}(\mathrm{l}^{W}, p_{\mathrm{T}}^{\mathrm{miss}}) > 50$ GeV,
- no jets with $p_{\mathrm{T}} > 25$ GeV,

- the invariant mass of the three leptons to be greater than 110 GeV to reject contributions from conversions.

For the ZZ region, the events are required to have:

- at least 4 leptons
- two OSSF pairs. Z_1 is the pair with an invariant mass closer to the Z boson and Z_2 the remaining one.
- m_{Z_1} between 76 and 106 GeV,
- m_{Z_2} between 50 and 130 GeV,
- no jets with a p_T greater than 25 GeV.

5.4 Background predictions

Three types of backgrounds are affecting this analysis. The first kind are processes in which events with a SF lepton pair and events with a DF lepton pair are produced with the same rate. Another contribution comes from DY+jets events. Even though all the signal regions of this analysis require the presence of p_T^{miss} in the final state, a significant number of events may leak into the signal region because of instrumental p_T^{miss}. In the third type of processes, referred to as Z+ν processes, the two leptons are produced in a Z/γ^* decay and genuine p_T^{miss} is present in the final state. The estimation of the first kind of processes is fully based on data, while remaining contributions from the other are done using simulations. This section will cover the methods used in this search to predict each one of them.

5.4.1 Flavor-symmetric processes

Flavor-symmetric processes are dominated by $t\bar{t}$ events, although contributions from WW, $Z \to \tau\tau$ and tW production events are expected. In particular, in some of the slepton signal regions, WW events are expected to contribute dominantly the flavor symmetric component.

As mentioned in the previous section, these backgrounds are estimated in control regions that are built with the same requirements as the signal regions, requiring DF instead of SF leptons. By construction, the control regions will only be populated by flavor-symmetric events, that are expected to contribute equally to signal and control regions.

However, due to reconstruction, isolation and trigger efficiency differences between electrons and muons, the flavor-symmetric yield in the DF and SF regions may differ. In order to take into account this difference, a transfer factor between signal and control regions, $R_{SF/OF}$, is computed. This factor is derived with the so-called factorization method. In the analysis performed in reference [12], the factor was derived with a combination of this method and a direct measurement in the $t\bar{t}$ control region

defined in the previous section. In this iteration of the analysis, the larger dataset allows for a more precise determination of $R_{SF/OF}$, dominated by the factorization method, so the direct measurement is not used. The $t\bar{t}$ control region is used instead to check the validity of the method in data.

In the factorization method, it is assumed that the selection efficiencies for the two leptons are independent. Then, the efficiency of the dilepton selection can be written as the product of the efficiencies of each lepton separately, $\epsilon_{l_1 l_2} = \epsilon_{l_1}\epsilon_{l_2}$.

To derive the method, the following nomenclature is adopted: $N_{ee(\mu\mu)}$ denotes the number of $e^{\pm}e^{\mp}$ ($\mu^{\pm}\mu^{\mp}$) events. The ^{hard} superscript means the quantity refers to the generator-level quantity, before any trigger or reconstruction effects are taken into account. When the superscript "*" is included, only offline reconstruction and identification efficiencies are taken into account, and trigger efficiencies have been factored out. Finally, trigger efficiencies in the differnet channels are denoted as ϵ^T_{ee}, $\epsilon^T_{e\mu}$ and $\epsilon^T_{\mu\mu}$.

The factor $r_{\mu/e}$ is defined as the ratio of muon and electron offline and trigger efficiencies, ϵ_μ/ϵ_e. $r_{\mu/e}$ is one of the coefficients to be measured in data, and can be computed as

$$r_{\mu/e} = \sqrt{\frac{N_{\mu\mu}}{N_{ee}}} = \sqrt{\frac{\epsilon^T_{\mu\mu}\epsilon^*_\mu(l_1)\epsilon^*_\mu(l_2)}{\epsilon^T_{ee}\epsilon^*_e(l_1)\epsilon^*_e(l_2)}}. \tag{5.3}$$

The number of SF events can be written as

$$\begin{aligned} N_{ee} &= \epsilon^T_{ee}N^*_{ee} = \epsilon^T_{ee}\epsilon^*_e(l_1)\epsilon^*_e(l_2)N^{hard}_{ee} && (5.4)\\ &= \frac{1}{2}\epsilon^T_{ee}\epsilon^*_e(l_1)\epsilon^*_e(l_2)N^{hard}_{DF} && (5.5)\\ &= \frac{1}{2}\epsilon^T_{ee}\epsilon^*_e(l_1)\epsilon^*_e(l_2)\frac{N^*_{DF}}{\epsilon^*_e(l_1)\epsilon^*_\mu(l_2)} && (5.6)\\ &= \frac{1}{2}\frac{\epsilon^T_{ee}}{\epsilon^T_{DF}} = \frac{1}{2}\frac{1}{r_{\mu/e}(l_2)}\frac{\sqrt{\epsilon^T_{ee}\epsilon^T_{\mu\mu}}}{\epsilon^T_{DF}}N_{DF}, && (5.7)\end{aligned}$$

where in step 5.6 the convention has been adopted that l_2 is the electron in the DF sample and l_1 is the muon. Similarly, for muons,

$$N_{\mu\mu} = r_{\mu/e}(l_1)\frac{\sqrt{\epsilon^T_{ee}\epsilon^T_{\mu\mu}}}{\epsilon^T_{DF}}N_{DF}, \tag{5.8}$$

where again the convention that l_1 is the muon has been adopted. Then the total flavor symmetric can be written, defining $R_T = \frac{\sqrt{\epsilon^T_{ee}\epsilon^T_{\mu\mu}}}{\epsilon^T_{DF}}$ as

$$N_{SF} = \frac{1}{2}\left(r_{\mu/e}(\mu) + \frac{1}{r_{\mu/e}(e)}\right)R_T N_{DF}. \tag{5.9}$$

In this calculation, we have kept track of which lepton each one of the $r_{\mu/e}$ factors have to be evaluated in. In previous iterations of the analysis [12], only the dependence of the $R_{SF/OF}$ factor as a function of the p_T of one of the leptons was kept. In the current parameterization, and thanks to the higher collected luminosity and a deeper understanding of the detector response, it is possible to evaluate $r_{\mu/e}$ as as a function of the lepton 3-momentum. Then, the $R_{SF/OF}$ transfer factor becomes a function of the p_T and η of the two leptons.

$r_{\mu/e}$ factor is measured in the DY+jets control region defined in the previous section, using the relation $r_{\mu/e} = \sqrt{\frac{N_{\mu\mu}}{N_{ee}}}$. A parameterization of this factor as a function of the lepton p_T and η using an ad-hoc functional form, that has been empirically found to describe well data:

$$r_{\mu/e}(l) = r^0_{\mu/e} \cdot f\left(p_T(l)\right) \cdot g\left(\eta(l)\right) \tag{5.10}$$

$$f(p_T) = (a_1 + b_1/p_T) \tag{5.11}$$

$$g(\eta) = a_2 + \begin{cases} 0 & |\eta| < 1.6 \\ c_1 \cdot (\eta - 1.6)^2 & \eta > 1.6 \\ c_2 \cdot (\eta + 1.6)^2 & \eta < -1.6 \end{cases}, \tag{5.12}$$

where $r^0_{\mu/e}$, a_1, a_2, b_1, c_1 and c_2 are constants that are determined from data and parametrize the dependencies of $r_{\mu/e}$ as a function of the lepton η and p_T. The dependencies on p_T and η are determined performing two separate fits, one for each dependency. In order to derive $f(p_T)$ and $g(\eta)$, the $r^2_{\mu/e}(l^+,l^-) = r_{\mu/e}(l^+)r_{\mu/e}(l^-)$ is considered. When considering $r^2_{\mu/e}(l^+,l^-)$ as a function of a given variable, $r^2_{\mu/e}(l^+)$, one must marginalize over the remaining variables, the kinematics of the remaining lepton and η of that lepton:

$$r^2_{\mu/e}(l^+) = r_{\mu/e}(l^+) \int r_{\mu/e}(l^-) dl^- = r_{\mu/e}(l^+) \bar{r_{\mu/e}}, \tag{5.13}$$

having assumed that the two leptons are independent, and η and p_T of the leptons are independent. It should be noted that the choice of the positive lepton is just a way to randomize the choice. Picking the leading or sub-leading lepton to obtain the parameterization would break the independence assumption.

In a first step, the marginalized distribution $r^2_{\mu/e}(l^+)$ as a function of the p_T allows to obtain the distribution of $f(p_T)$. However, since there is a degeneracy between a_1, $r^0_{\mu/e}$ and $\bar{r_{\mu/e}}$, the fit only allows to determine the ratio b_1/a_1. The overall magnitude of $r^0_{\mu/e}$ will be determined at a later step. The result of this fit is shown in figure 5.8.

In the second step, the η dependencies of the $r_{\mu/e}$ factor are determined using the same methodology. In order to avoid correcting twice the p_T dependencies due to possible correlations between p_T and η, each dielectron event is weighed by $f(p_T^{l_1})f(p_T^{l_2})$. Again a degeneracy among the various parameters is present, and only c_1/a_2 and c_2/a_2 can be obtained from this fit. The result of this fit is shown in figure 5.9.

Since the fits described above do not to fully determine the $r_{\mu/e}$ function, only its p_T and η dependencies have been obtained. $r^0_{\mu/e}$ is then determined inclusively, weighing

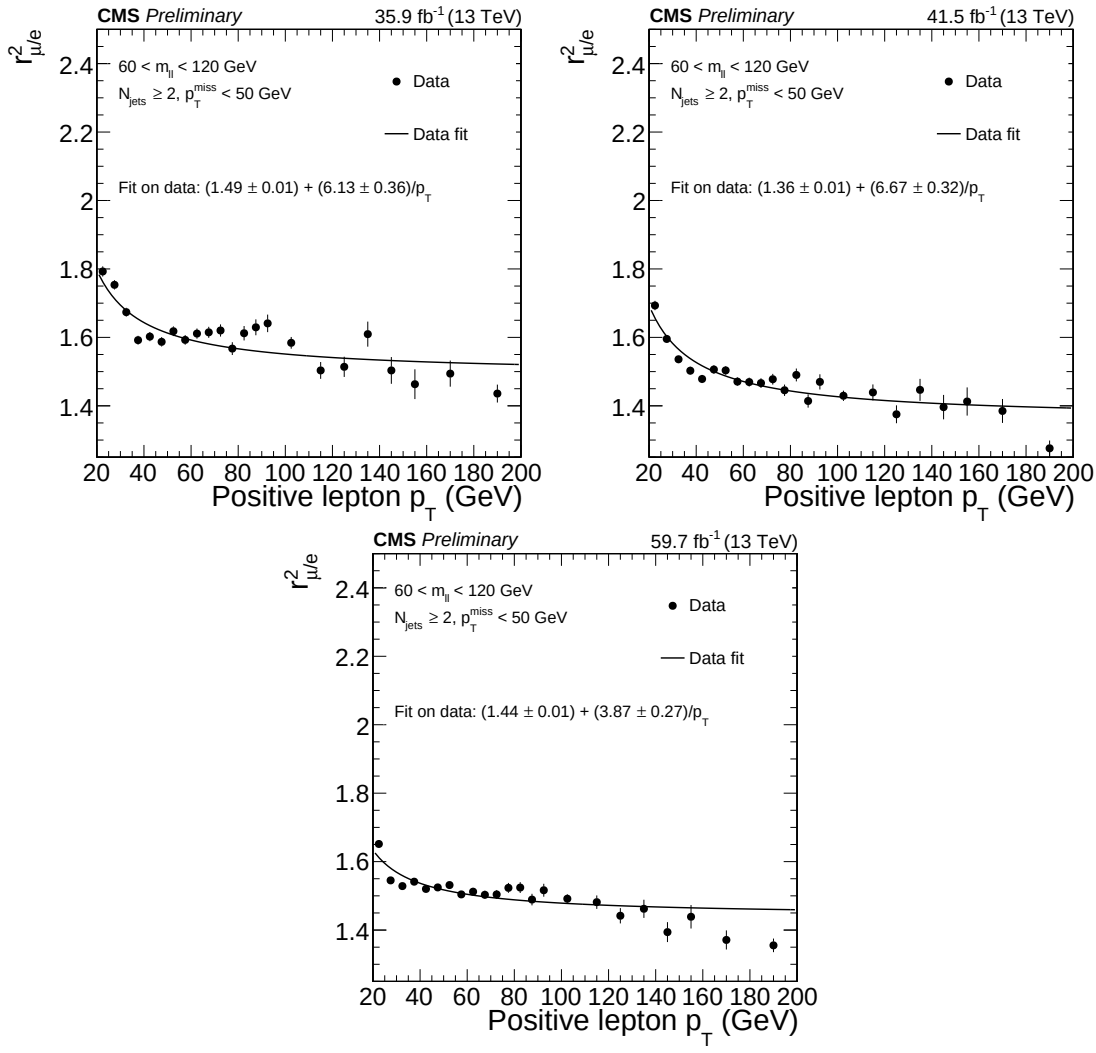


FIGURE 5.8: Determination of the p_{T} dependency of the $r_{\mu/e}$ factor. $r^2_{\mu/e}$ as a function of the positive lepton p_{T} is shown in data collected during 2016 (top left), 2017 (top right) and 2018 (bottom). A fit to an hyperbola is performed to the distribution.

dielectron events by $f(p_{\mathrm{T}}^{l_1})f(p_{\mathrm{T}}^{l_2})g(\eta^{l_1})g(\eta^{l_2})$, to fully account for the observed η and p_{T} dependencies. The resulting values of the $r_{\mu/e}$ parameterization are shown in table 5.4. A significant non-zero value is obtained for c_1 and c_2, that paremeterize the η dependency. This dependency is present due to the L1 ECAL prefiring described in chapter 2.

	$r^0_{\mu/e}$	a_1	b_1	a_2	c_1	c_2
2016	1.277±0.001	1.493±0.008	6.135±0.364	0.600±0.001	0.356±0.022	0.476±0.024
2017	1.226±0.001	1.356±0.008	6.665±0.325	0.647±0.002	0.462±0.024	0.690±0.027
2018	1.234±0.001	1.437±0.006	3.870±0.266	0.653±0.001	0.097±0.015	0.099±0.015

TABLE 5.4: Fitted values for the parameterization of $r_{\mu/e}$, obtained in the DY+jets control regions with consecutive fits to correct the p_{T} and η dependencies.

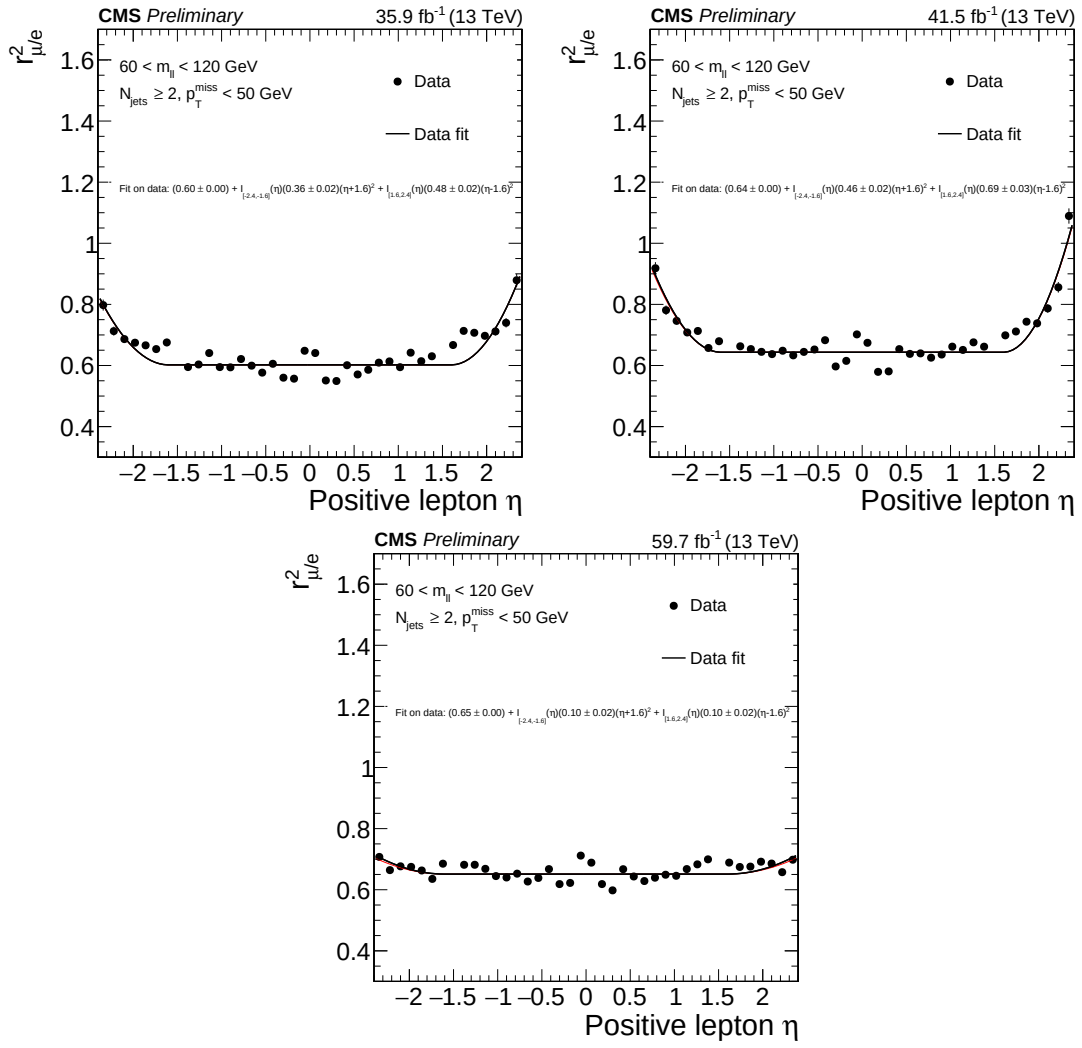


FIGURE 5.9: Determination of the η dependency of the $r_{\mu/e}$ factor. $r^2_{\mu/e}$ as a function of the positive lepton η is shown in data collected during 2016 (top left), 2017 (top right) and 2018 (bottom). ee events have been weighted by $f(p_T^{e_1})f(p_T^{e_2})$ to correct for the p_T dependency. A fit to the function in equation 5.12 is shown.

Once the $r_{\mu/e}$ parameterization has been completely determined, the distributions of $\sqrt{N_{\mu\mu}/N_{ee}}$ are considered after dielectron events have been reweighed by the $r_{\mu/e}(l_1)r_{\mu/e}(l_2)$ as a function of several observables. These distributions are shown in figure 5.10 for the three years of the data taking. No significant trends are observed in the distributions, and residual trends are used to estimate the systematic uncertainty of the method. These residual trends can be due to correlations between the kinematics of the two leptons and trends not covered by the ad-hoc parameterization. Three sources of uncertainties are considered for $r_{\mu/e}$: a 5% flat uncertainty, a 5% uncertainty that modulates its p_T dependency, and a 5% modulating its η dependency. As seen in figure 5.10, these uncertainties cover the residual dependencies.

The remaining ingredient is the R_T factor, that corrects the remaining effects due to the residual differences in trigger efficiencies. This factor can be determined by measuring the trigger efficiencies using the orthogonal trigger method. In previous iterations of

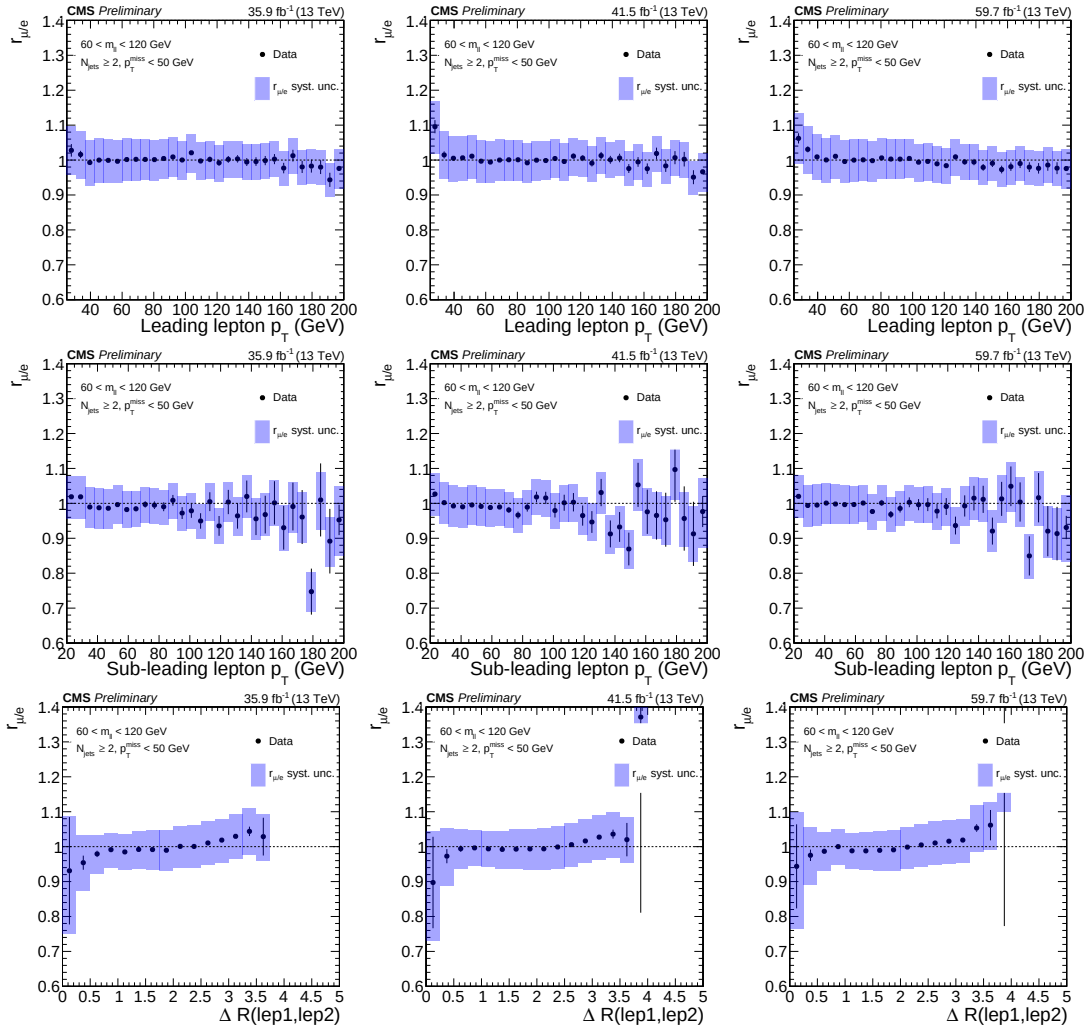


FIGURE 5.10: $r_{\mu/e}$ after it has been corrected by the p_T and η dependencies as a function of the leading lepton p_T (top), sub-leading lepton p_T (middle) and ΔR between the two leptons. Blue band shows the systematic uncertainty associated to $r_{\mu/e}$.

the analysis, these measurements were performed with triggers that collect events with high H_T. However, the dedicated bandwidth for these triggers was significantly decreased for the 2017 and 2018 data taking, and the amount of data collected by them was not enough to make a precise measurement. Because of this, triggers selecting events with large momentum imbalance are used.

The efficiencies for each one of the channels are measured in dilepton events that are not entering the signal regions or their equivalent DF control regions. Since the trigger reconstruction algorithm and menu changed during Run 2, the measurement is performed separately for each one of the years of the data taking. These efficiencies are shown in in table 5.5, together with the corresponding R_T value. This measurement has also been performed as a function of several kinematic variables. No significant trends on this variable have been found, and a 4% systematic uncertainty is assigned to this factor to account for small statistical fluctuations in the measurements.

The whole method is validated in the control region described in the previous section,

Data-taking era	2016	2017	2018
ee eff.	0.916±0.003	0.906±0.003	0.923±0.003
μμ eff.	0.949±0.003	0.873±0.004	0.925±0.003
eμ eff.	0.890±0.003	0.858±0.003	0.898±0.002
R_T	1.048±0.043	1.037±0.044	1.029±0.042

TABLE 5.5: Trigger efficiencies for each of the channels and year of the data taking, as measured with the orthogonal trigger method, and the corresponding R_T value. The uncertainty shown is statistical for the trigger efficiencies, and includes the systematic uncertainty for R_T.

that is enriched in $t\bar{t}$ events. Data in this region is compared to the SM prediction by the data-driven estimate of the flavor symmetric background and simulations for the remaining backgrounds, as shown in figure 5.11. Data agrees well with predictions within the statistical and systematic uncertainties of the method.

5.4.1.1 Flavor-symmetric processes in on-Z regions

Since the expected rate of flavor-symmetric events in the on-Z regions is expected to be small, the statistical power of the DF flavor control regions could be a limiting factor of the analysis. In order to tackle this, the requirements on these control regions are relaxed by removing the $86 < m_{ll} < 96$ GeV cut. To account for this, an additional factor, κ, is needed

$$\kappa = \frac{N^{\mathrm{DF}(86<m_{ll}<96\,\mathrm{GeV})}}{N^{\mathrm{DF}(m_{ll}>20\,\mathrm{GeV})}}.$$

κ is estimated data in a set of regions with DF lepton pairs, separately for different sets of signal regions. It is estimated in the DF control region associated to the following regions:

- three regions: SRA, SRB and SRC, merging the two b jet multiplicity bins of each one of them,
- "VZ resolved" region;
- "VZ boosted" region, removing the "VZ resolved" veto;
- HZ signal region.

These measurements are shown in figure 5.12. The κ values for the six defined control regions are found to be consistent with simulations within the statistical uncertainty. However, the difference in SRC between the measured value and the one predicted by the simulations is taken as a systematic uncertainty, once the statistical uncertainty is subtracted.

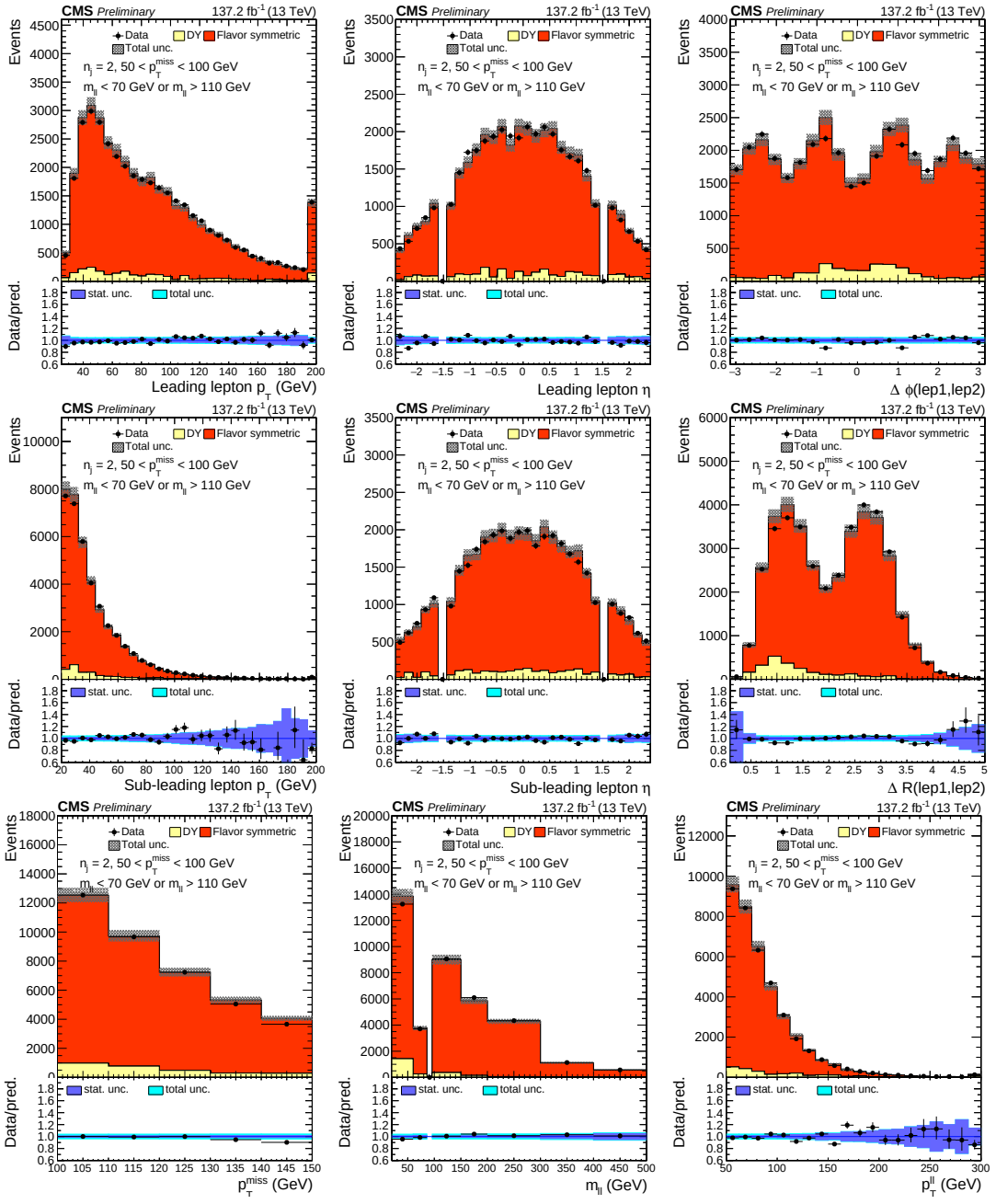


FIGURE 5.11: Data in the $t\bar{t}$ control region as a function of several kinematic variables compared to the estimates of the SM backgrounds.

The dependency of κ on $p_{\mathrm{T}}^{\mathrm{miss}}$ and the b jet multiplicity is also studied. In order to do that, events passing the minimal dilepton cut, with at least two jets and $M_{T2} > 80$ GeV. Events are also required to have $\Delta\phi$ between $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ and the two leading jets greater than 0.4. κ is measured in events with and without b-tagged jets, and as a function of $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$. These results are shown in figure 5.12.

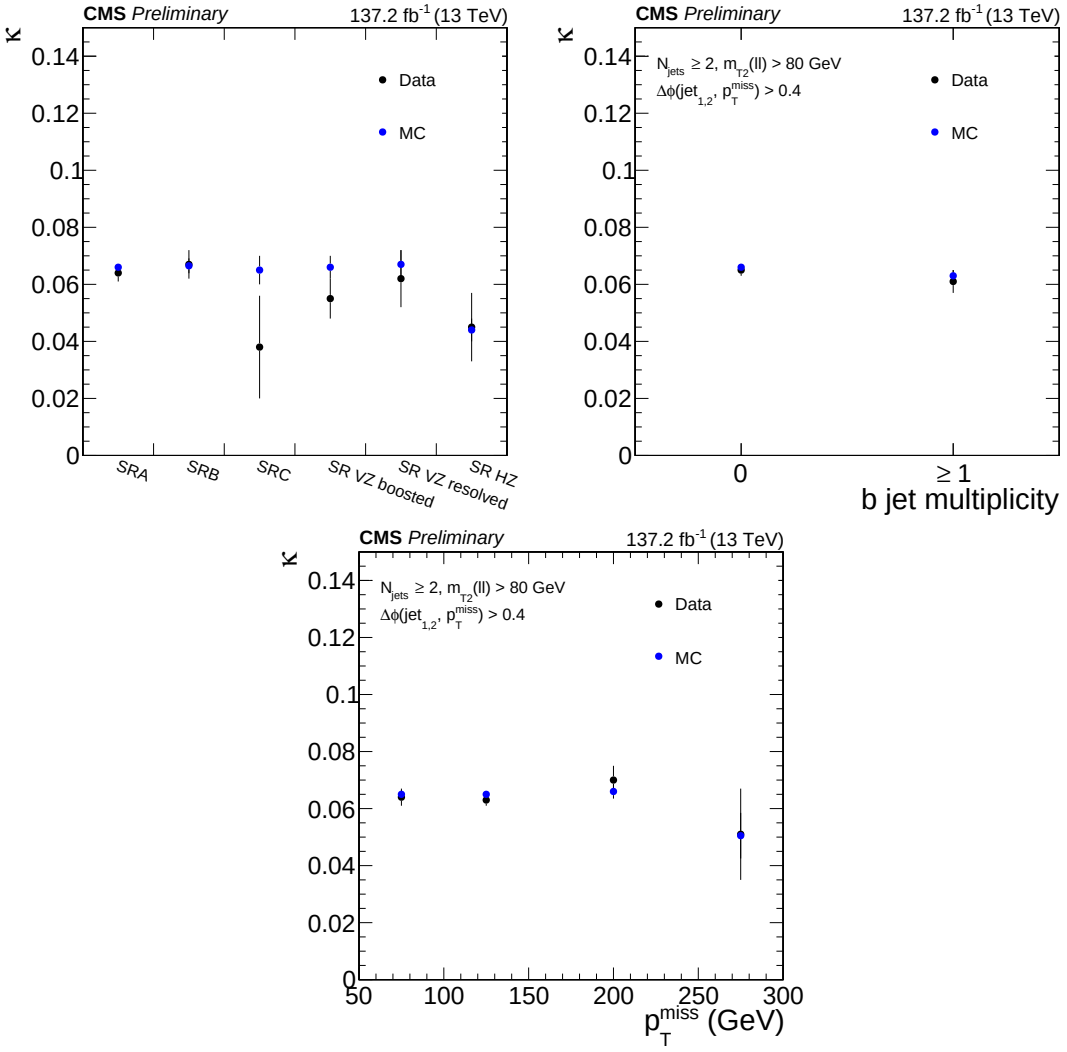


FIGURE 5.12: Measurement of the κ factor in data (black) and simulations (blue). The κ factor is measured for different control regions (top left), as a function of the b jet multiplicity (top right), and as a function of $p_{\mathrm{T}}^{\mathrm{miss}}$ (bottom).

The following systematic uncertainties are considered for κ: the statistical uncertainty in the measurement and a 20% to cover the trends observed as a function of $p_{\mathrm{T}}^{\mathrm{miss}}$. The dependencies as a function of the b multiplicity is found to be negligible compared to the others and is not taken into account.

5.4.2 DY+jets backgrounds

DY+jets events can enter the signal regions due to the limited energy resolution and acceptance of the detector, that can lead to instrumental $p_{\mathrm{T}}^{\mathrm{miss}}$. This $p_{\mathrm{T}}^{\mathrm{miss}}$ is usually

small, however due to the large cross section of the DY+jets process, it can contribute with a significant number of events to the signal regions, specially the ones with lower $p_{\mathrm{T}}^{\mathrm{miss}}$ requirements.

In signal regions with large or moderate jet multiplicity requirements, i.e. the on-Z regions and the slepton-edge search, instrumental $p_{\mathrm{T}}^{\mathrm{miss}}$ is dominated by energy mismeasurements of the jets. Therefore the shape of the $p_{\mathrm{T}}^{\mathrm{miss}}$ distribution is determined from a γ+jets data sample, using the so-called "$p_{\mathrm{T}}^{\mathrm{miss}}$ templates" method.

In slepton regions, in which only events with small hadronic activity are allowed, this assumption may not be valid anymore, and the "$p_{\mathrm{T}}^{\mathrm{miss}}$ templates" method cannot be used. In these regions, the $65\ \mathrm{GeV} < m_{ll} < 105\ \mathrm{GeV}$ requirement is inverted to obtain a region enriched in DY+jets, that can be used to infer its presence.

5.4.2.1 "$p_{\mathrm{T}}^{\mathrm{miss}}$ templates" method

The "$p_{\mathrm{T}}^{\mathrm{miss}}$ templates" method relies on the fact that energy mismeasurements of the jets contribute more dominantly to the instrumental $p_{\mathrm{T}}^{\mathrm{miss}}$ than photons or leptons, which are expected to be measured with higher precision. Under that assumption, and assuming the same event topology for DY+jets and γ+jets processes, the $p_{\mathrm{T}}^{\mathrm{miss}}$ distribution should be equivalent between the two. The distribution of $p_{\mathrm{T}}^{\mathrm{miss}}$ is then inferred from the photon control regions defined in section 5.3.5. This method has been developed within the UCSD group and it is only briefly described here for completeness.

Processes with genuine $p_{\mathrm{T}}^{\mathrm{miss}}$ can contribute to the γ+jets samples. This effect is expected to be more significant in regions with large $p_{\mathrm{T}}^{\mathrm{miss}}$ requirement. In order to account for it, the contribution from these events is subtracted using simulations. The contribution from $\mathrm{W}+\gamma$, $\mathrm{t\bar{t}}\gamma$ and $\mathrm{Z}(\rightarrow \nu\nu)\gamma$ is considered. The modeling of these processes is checked in a control region with a reconstructed photon and a muon, which is dominated in $\mathrm{W}\gamma$ events. Simulations are found to overpredict data by a 30%, therefore the subtraction is performed by scaling simulations by 0.7. This value is also assigned as a systematic uncertainty across all the signal regions.

Additionally, due to the different boson mass, the boson p_{T} distribution is different between γ+jets and dilepton samples. To account for this, γ+jets data events are reweighed so their p_{T} spectrum matches the expected one in DY+jets events. This reweighing is derived separately for each region using DY+jets and γ+jets simulations in dedicated samples.

Finally, the validity of the method is checked using DY+jets and γ+jets simulations in each signal region separately. The prediction of the procedure applied to the γ+jets simulations is compared to the prediction of the DY+jets simulations. The difference between the two is assigned as a systematic uncertainty of the method. These differences range between 20% and 100%, the greater of which come from regions with low number of simulated events.

This procedure allows to obtain an estimate of the $p_{\mathrm{T}}^{\mathrm{miss}}$ distribution of the DY+jets process in the signal region. This distribution is normalized using events with $50 < p_{\mathrm{T}}^{\mathrm{miss}} < 100\ \mathrm{GeV}$, after the remaining backgrounds have been subtracted. Since a

small contribution from signal is expected in some of these regions, the normalization for the templates is parametrized as a freely floating parameter in the signal extraction fit. The region with $50 < p_{\mathrm{T}}^{\mathrm{miss}} < 100$ GeV is also included in the fit to constrain this parameter. This way the signal hypothesis is taken into account in the background estimation.

In summary, the following uncertainties are considered in the method:

- the statistical uncertainty of data in the control sample,
- the uncertainty associated to the closure test in simulations,
- the uncertainty associated to the electroweak subtraction.

The statistical uncertainty of the $50 < p_{\mathrm{T}}^{\mathrm{miss}} < 100$ GeV region, that is used to normalize the templates, is included by considering the normalization parameter in the fit. The validity of this method has been checked in the validation regions described in the previous sections, finding the prediction to agree well with data within the uncertainties.

5.4.2.2 DY+jets in slepton-edge regions

The contribution from DY+jets in the slepton-edge regions can also be estimated using the $p_{\mathrm{T}}^{\mathrm{miss}}$ templates method. It is estimated in a region with $86 < m_{\mathrm{ll}} < 96$ GeV, that is more enriched in this process, and is suitable to apply the "$p_{\mathrm{T}}^{\mathrm{miss}}$ templates" method. Then, the leakage of DY+jets into the off-Z signal regions must be taken into account. In order to do so, a transfer factor, $r_{\mathrm{in/out}}$, is constructed. The transfer factor is defined as

$$r_{\mathrm{in/out}}^{[m_{\mathrm{ll}}^1, m_{\mathrm{ll}}^2]} = \frac{N(m_{\mathrm{ll}}^1 < m_{\mathrm{ll}} < m_{\mathrm{ll}}^2)}{N(86\,\mathrm{GeV} < m_{\mathrm{ll}} < 96\,\mathrm{GeV})}$$

for each signal region. This factor is measured in the DY+jets control region with jets defined in section 5.3.4.2. The data yield, together with the contribution from other backgrounds other than DY+jets is shown in figure 5.13. The contribution from these backgrounds is subtracted in the $r_{\mathrm{in/out}}$ calculation. These measurements are also performed as a function of $p_{\mathrm{T}}^{\mathrm{miss}}$ and $\mathrm{M_{T2}(ll)}$, showing some trends. In order to take them into account, a 50% uncertainty is assigned for regions below the Z peak and 100% for regions above the Z peak. The results are shown in table 5.6.

5.4.2.3 DY+jets in slepton regions

Since in the slepton regions only low energy jets are required, the "$p_{\mathrm{T}}^{\mathrm{miss}}$ templates" method cannot be used for two reasons. First, a significant contribution of instrumental $p_{\mathrm{T}}^{\mathrm{miss}}$ from lepton energy mismeasurement can take place. Secondly, contributions to the instrumental $p_{\mathrm{T}}^{\mathrm{miss}}$ can come from objects outside the detector acceptance or

not passing our object definition, that cannot be handled in a controlled way at the analysis level.

Because of this, and of the fact that the contribution of DY+jets to the slepton regions is subdominant in the most sensitive regions, the contribution from DY+jets events is estimated directly from a control region defined by inverting the Z boson mass veto. This is done separately for each signal region and for each $p_{\mathrm{T}}^{\mathrm{miss}}$ bin. The contribution from the on-Z region is extrapolated to the off-Z region with another $r_{\mathrm{in/out}}$ transfer factor, dubbed $r_{\mathrm{in/out}}^{\mathrm{slepton}}$, that is determined from data in the DY+jets control region without jets defined in section 5.3.4.2. The data yield for the measurement is shown in figure 5.13. Since slepton regions are not classified in m_{ll}, only one transfer factor (off-Z/on-Z) is computed. Systematic uncertainties are applied to this procedure from a closure test on DY+jets simulations, resulting on a 50% uncertainty, which is dominated by the statistical uncertainty of the test. The results of the measurements are shown in table 5.6.

A small contribution from signal is expected in the on-Z regions, which may be particularly significant in the high $p_{\mathrm{T}}^{\mathrm{miss}}$ bins. Because of that, these regions are included in the signal extraction, and the contribution of DY+jets is modulated as a freely floating parameter in the fit. In total, 8 parameters are added to account for the DY+jets contribution in the two regions with and without jets, and in the four $p_{\mathrm{T}}^{\mathrm{miss}}$ bins. In that way, the DY+jets contribution is constrained by data in the on-Z region, taking into account any potential signal, as well as the presence of other backgrounds.

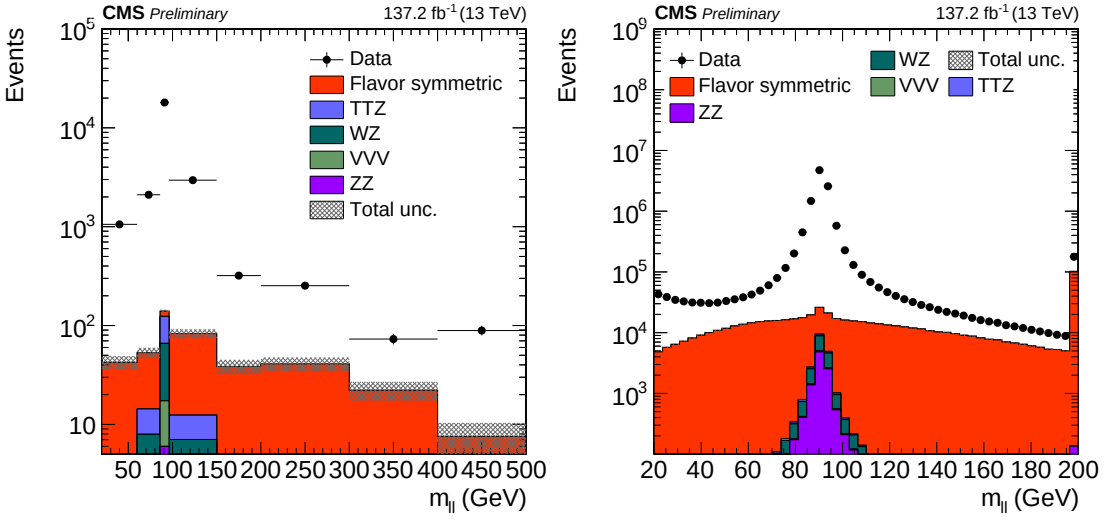


FIGURE 5.13: Measurements of the $r_{\mathrm{in/out}}$ value for the slepton-edge regions (left) and the slepton regions (right). Only two regions are considered for the slepton measurement: on-Z and off-Z.

5.4.3 Z+ν backgrounds

Processes with genuine $p_{\mathrm{T}}^{\mathrm{miss}}$ and in which the two leptons come from the same Z/γ^{*} boson are not included in any of the two groups above. The contribution from them

Transfer factor	Value
$r^{\text{slepton}}_{\text{in/out}}$	0.08 ± 0.04
$r^{[20,60]}_{\text{in/out}}$	0.057 ± 0.029
$r^{[60,86]}_{\text{in/out}}$	0.117 ± 0.059
$r^{[86,150]}_{\text{in/out}}$	0.159 ± 0.079
$r^{[150,200]}_{\text{in/out}}$	0.015 ± 0.015
$r^{[200,300]}_{\text{in/out}}$	0.011 ± 0.011
$r^{[300,400]}_{\text{in/out}}$	0.003 ± 0.003
$r^{[400,\infty]}_{\text{in/out}}$	0.005 ± 0.005

TABLE 5.6: Measured values for $r_{\text{in/out}}$ in the different control regions for the slepton-edge and the slepton analysis. The uncertainty shown is the assigned systematic one.

is taken into account using simulations. Given that some of them have a flavor-symmetric components, generator information is used to select simulated events in which the two leptons come from the decay of the same Z/γ^* boson.

The contribution from these processes is dominated by WZ, ZZ and $\mathrm{t\bar{t}Z}$ processes in the case of the regions with jets. Only diboson processes contribute significantly to the slepton regions. The agreement of these processes is checked in the control regions defined in section 5.3.6.

In the regions without jets, an good agreement is observed between simulations and the predictions. For WZ a systematic uncertainty of 6% is taken on its normalization, as measured by the latest CMS measurement [148]. A mild trend is observed when considering the vector sum of the lepton associated to the W decay and the $\vec{p}_{\mathrm{T}}^{\,\text{miss}}$, which is a proxy of the modeling of the $p_{\mathrm{T}}^{\text{miss}}$ distribution in regions with two leptons. A shape uncertainty is assigned to account for this difference. For ZZ a small dependency of the normalization is seen in the 4e, 4μ and 2e2μ channels, that is due to the uncertainties in the lepton efficiency calibration. A systematic uncertainty of 20% is used to account for this. Additionally, the shape difference between applying and not applying the NNLO/NLO *k*-factors is taken as a source of systematic uncertainty for this process.

In regions with jets, mild excesses are observed with respect to data, depending on the data taking era. The differences are used to correct simulations to match data. The scale factors applied are shown in table 5.7. It should be noted that the uncertainties associated to the scale factors include statistical uncertainty only. In particular, $\mathrm{t\bar{t}Z}$ normalization is performed with respect to the NLO cross section and performed in a particular region of the phase space valid for this analysis. A more thorough and precise study of this process is performed in chapter 6. A systematic uncertainty of 50%, 30% and 50% is applied to the ZZ, WZ and $\mathrm{t\bar{t}Z}$ processes, respectively, to account for kinematic dependencies observed.

Process	WZ	ZZ	$t\bar{t}Z$
2016	1.04 ± 0.17	0.91 ± 0.50	1.22 ± 0.23
2017	1.14 ± 0.17	1.14 ± 0.30	1.56 ± 0.19
2018	1.28 ± 0.13	1.34 ± 0.24	1.51 ± 0.15

TABLE 5.7: Scale factors for the WZ, ZZ and $t\bar{t}Z$ processes for each year of the data taking.

On top of these modeling uncertainties described above, the impact of experimental uncertainties, such as lepton and trigger efficiencies, jet energy scale and resolution, and b tagging are taken into account.

5.5 Results

In this section, the number of observed events in each one of the signal regions is compared to the SM predictions provided by the estimation methods discussed along this chapter.

5.5.1 Results of the on-Z SRs

The results for the SRs of the on-Z strong-production search are shown in figure 5.14, and for the electroweak search are shown in figure 5.15. The $p_{\mathrm{T}}^{\mathrm{miss}}$ template prediction for each SR is normalized to events with 50-100 GeV, therefore data and prediction agree in the first bin of each distribution by construction.

Data in these regions shows a fairly good agreement with the SM prediction, and no remarkable differences are seen. The largest discrepancy occurs in the highest $p_{\mathrm{T}}^{\mathrm{miss}}$ bins of the resolved VZ region, where 2 events are observed, while 6.3 ± 2.2 where predicted. This difference corresponds to a local significance of -1.2 standard deviations (s.d.), very compatible with a statistical fluctuation.

5.5.2 Results of the edge search

The results for the 28 signal regions in the search for an edge are shown in figure 5.16. There is a general good agreement between data and the SM expectations. The largest discrepancy appears in the non $t\bar{t}$-like region with b-tagged jets. In the bin with $20 < m_{\mathrm{ll}} < 60$ GeV 2 events are observed while $7.4^{+3.7}_{-2.5}$ were expected, corresponding to a local significance of -2.4 standard deviations.

5.5.3 Results of the slepton search

The results for the slepton search are shown in figure 5.17. The DY+jets estimation has been obtained by performing a background-only fit to the on-Z region only. An

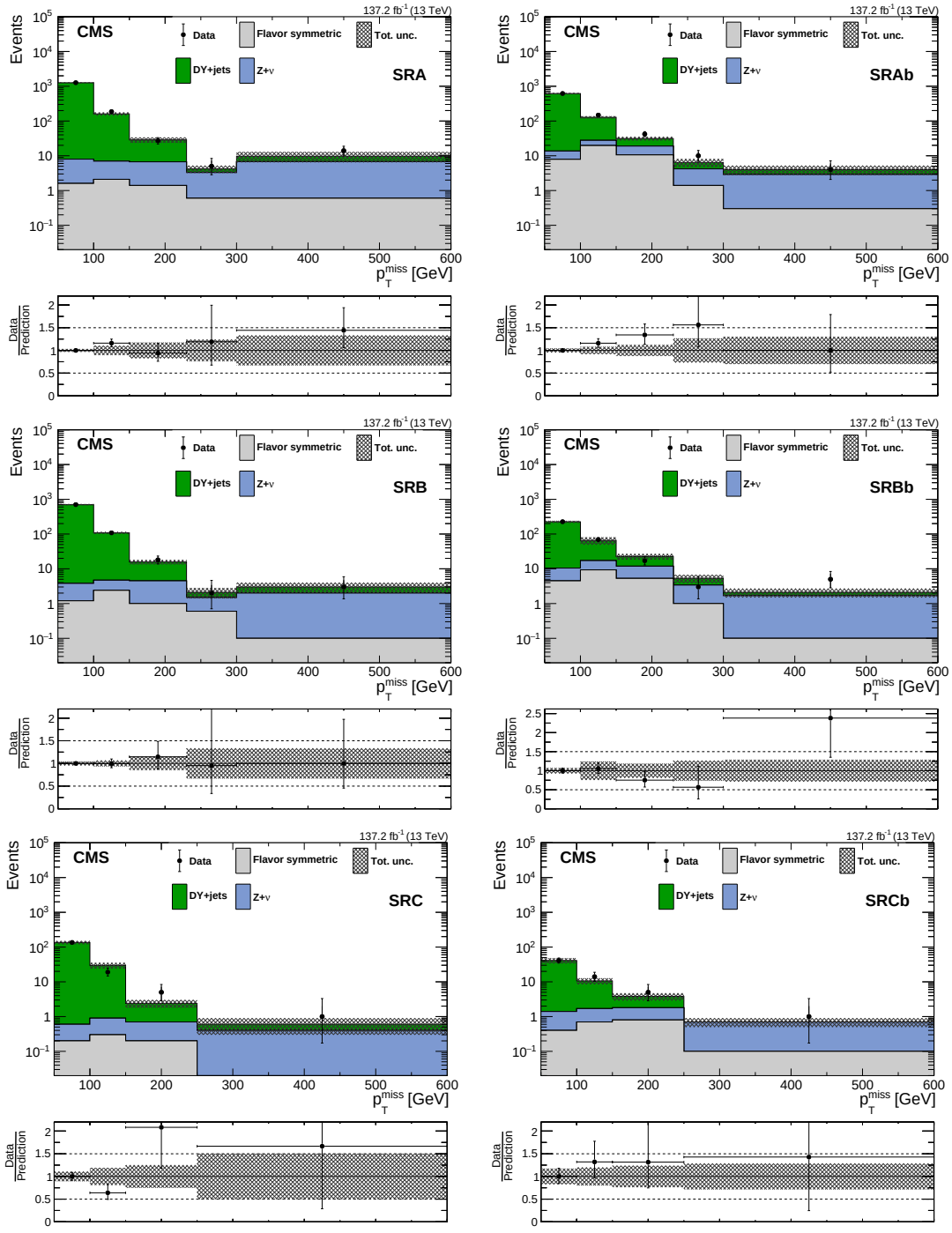


FIGURE 5.14: The $p_{\mathrm{T}}^{\mathrm{miss}}$ distribution in the strong on-Z signal regions with no b-tagged jets (left) and with b-tagged jets (right). The rows correspond to the SRA (upper), SRB (middle), and SRC (lower).

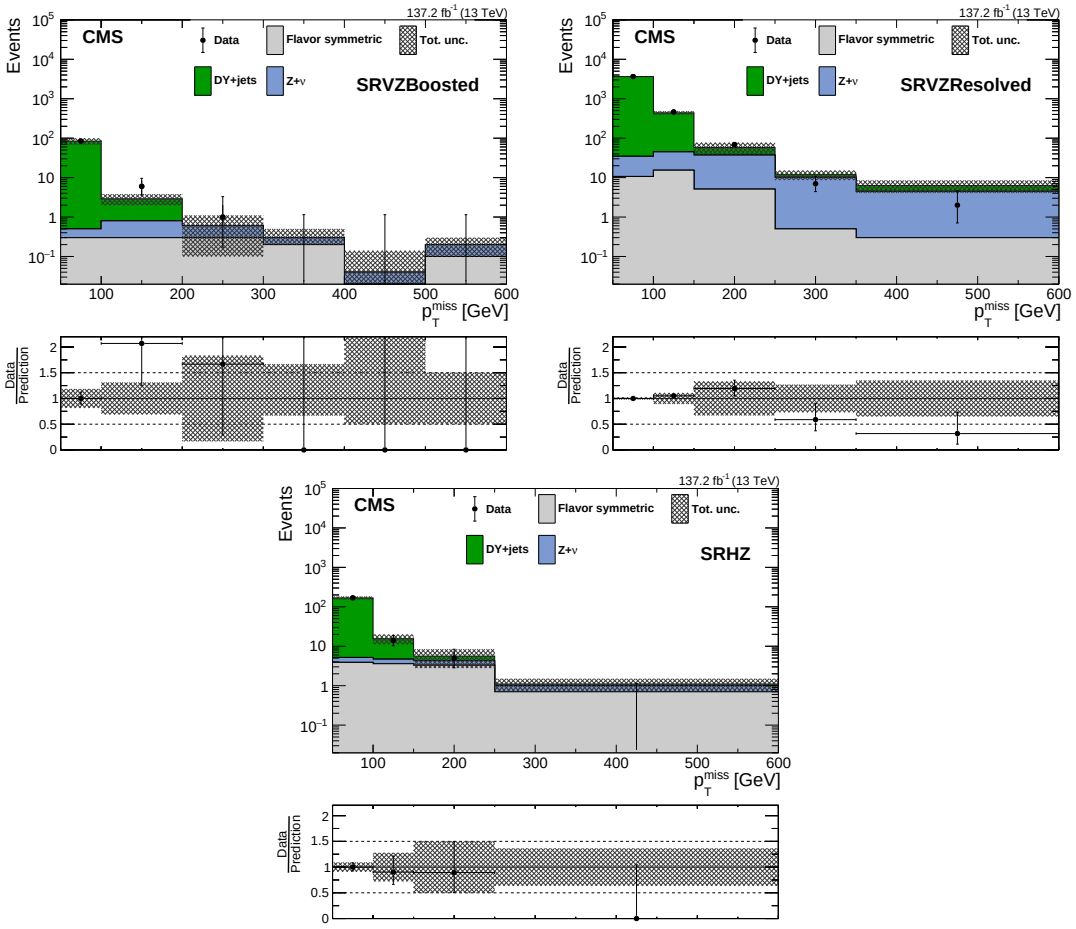


FIGURE 5.15: The $p_{\mathrm{T}}^{\mathrm{miss}}$ distribution in the electroweak on-Z signal regions. Top row shows the VZ boosted (left) and resolved (right) signal regions, and the bottom row shows the HZ signal region.

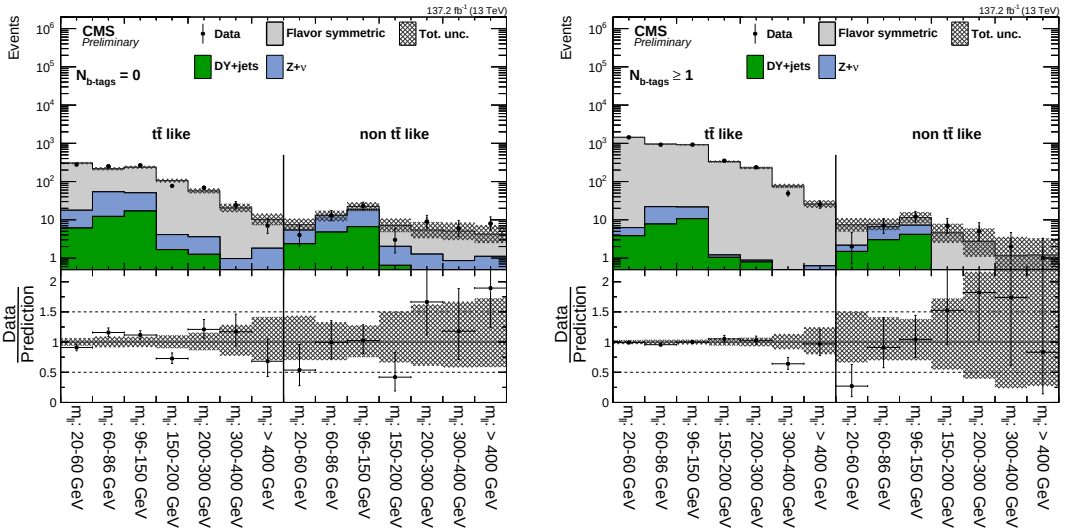


FIGURE 5.16: Results of the search for an edge, compared to the SM predictions, in the regions with without b-tagged jets (left) and with at least one b-tagged jet.

overall good agreement between data and the SM expectation is observed. The largest discrepancy appears in the region without jets, in the highest $p_{\mathrm{T}}^{\mathrm{miss}}$ bin, in which 17 events were observed and 9.3 ± 2.3 were predicted.

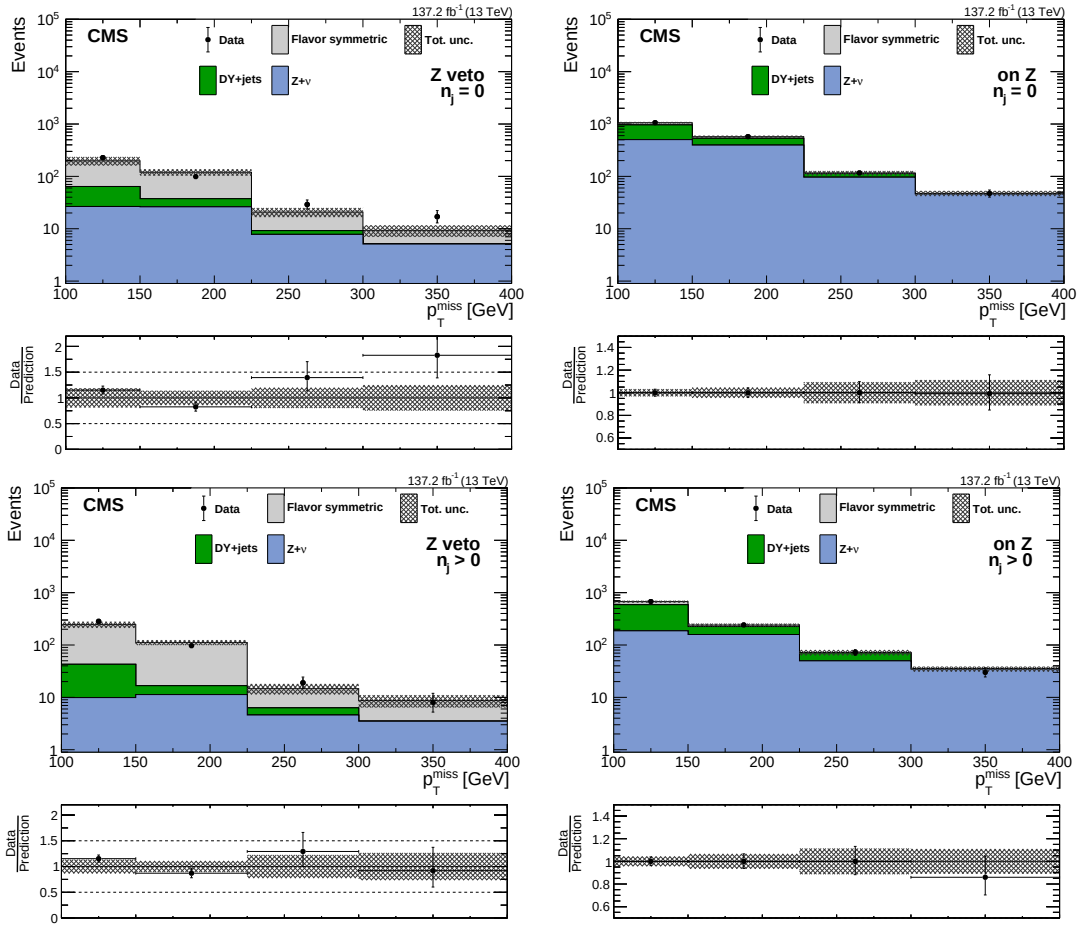


FIGURE 5.17: Distribution of the $p_{\mathrm{T}}^{\mathrm{miss}}$ in data events compared to the SM predictions in the slepton SRs (left) and the on-Z control regions used to determine the contribution from DY+jets events (right). Top row shows the region without jets, and bottom row shows the region with jets.

5.6 Interpretation

Since no significant differences between data and SM predictions have been observed in the searches, upper limits are set in terms of the production cross section of the simplified models considered for this search. These upper limits have been calculated at 95% confidence level (CL) using the $\mathrm{CL_S}$ criterion in the asymptotic formulation, described in section 2.6.2. However, these limits have been compared for a few significant points in each simplified model with the ones obtained using the $\mathrm{CL_S}$ criterion with toys, described in section 2.6.2. The two are found to agree within 1-5%, which validates the approach used.

The systematic uncertainties associated to backgrounds have been described in section 5.4. For signal, besides the uncertainties associated to the calibration of lepton

and trigger efficiencies, jet energy scale and resolution, b jet tagging, additional uncertainties are added to account for the additional calibration needed in fast simulations. Additional uncertainties are taken into account for the mismodeling of ISR in MadGraph samples at LO accuracy. The uncertainties on the signal cross section due to the PDF choice and missing diagrams in the calculation, which are estimated varying the renormalization and factorization scales, are propagated directly to the observed exclusion contours.

Upper limits on the GMSB scenario are obtained from the interpretation of the strong on-Z regions. These upper limits are shown in figure 5.18 as a function of the gluino and $\widetilde{\chi}_1^0$ masses. The exclusion contours are defined as those points in which the upper limit matches the expected theoretical cross section, and the observed and expected ones are shown. The results in these regions allow to exclude gluino masses between 1600 and 1850 GeV, depending on the $\widetilde{\chi}_1^0$ mass. Upper limits are weaker in the regions with lower mass difference between gluino and $\widetilde{\chi}_1^0$, since a lower efficiency for these models is expected because jets produced in the decay of the gluino will be softer. Also weaker exclusion is expected for low $\widetilde{\chi}_1^0$ masses since less $p_{\mathrm{T}}^{\mathrm{miss}}$ will be present.

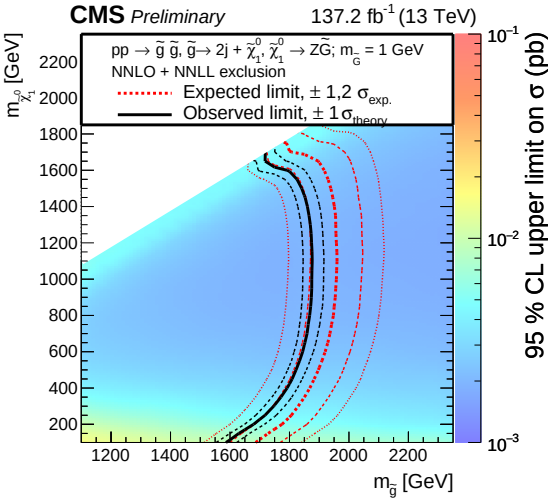


FIGURE 5.18: Observed upper limit at 95% CL for the GMSB scenario as a function of the $\widetilde{\mathrm{g}}$ and $\widetilde{\chi}_1^0$ masses. Observed (black) and expected (red) exclusion contours are also shown. The thin red dotted lines indicate the regions containing 68% and 95% of the expected limits under the background-only hypothesis. The thin dotted black line indicates the change in observed limit when varying the signal cross section by its uncertainty.

The results of the electroweak signal regions are interpreted in terms of the electroweak models. The VZ resolved and boosted regions drive the sensitivity for the $\widetilde{\chi}_2^0\widetilde{\chi}_1^{\pm}$ production.

For the $\widetilde{\chi}_1^{\pm}\widetilde{\chi}_1^{\pm}$ production models the sensitivity is given both for the VZ and HZ regions. For the cases in which the $\widetilde{\chi}_1^{\pm}$ only decays to the Z boson, most of the sensitivity is expected in the VZ regions, however some efficiency is recovered by the HZ regions in the cases in which one of the Z bosons decays hadronically. In the scenario in which the $\widetilde{\chi}_1^{\pm}$ decays to Z and H bosons with a 50% probability more sensitivity is given by

the HZ regions, however it also contributes with a 25% to the ZZ final state. The exclusion limits for the electroweakino models are shown in figure 5.19 as a function of the NLSP and LSP masses.

In the $\widetilde{\chi}_2^0\widetilde{\chi}_1^\pm$ production model, results allow to exclude $\widetilde{\chi}_2^0$ (or $\widetilde{\chi}_1^\pm$) masses up to 750 GeV. The observed limit is stronger than the expected to the observed yields being smaller than the SM prediction in two of the $p_{\mathrm{T}}^{\mathrm{miss}}$ bins of the resolved VZ region.

For the $\widetilde{\chi}_1^0\widetilde{\chi}_1^0$ models, neutralino masses up to 700 GeV are excluded in the first scenario, and up to 500 GeV in the second scenario.

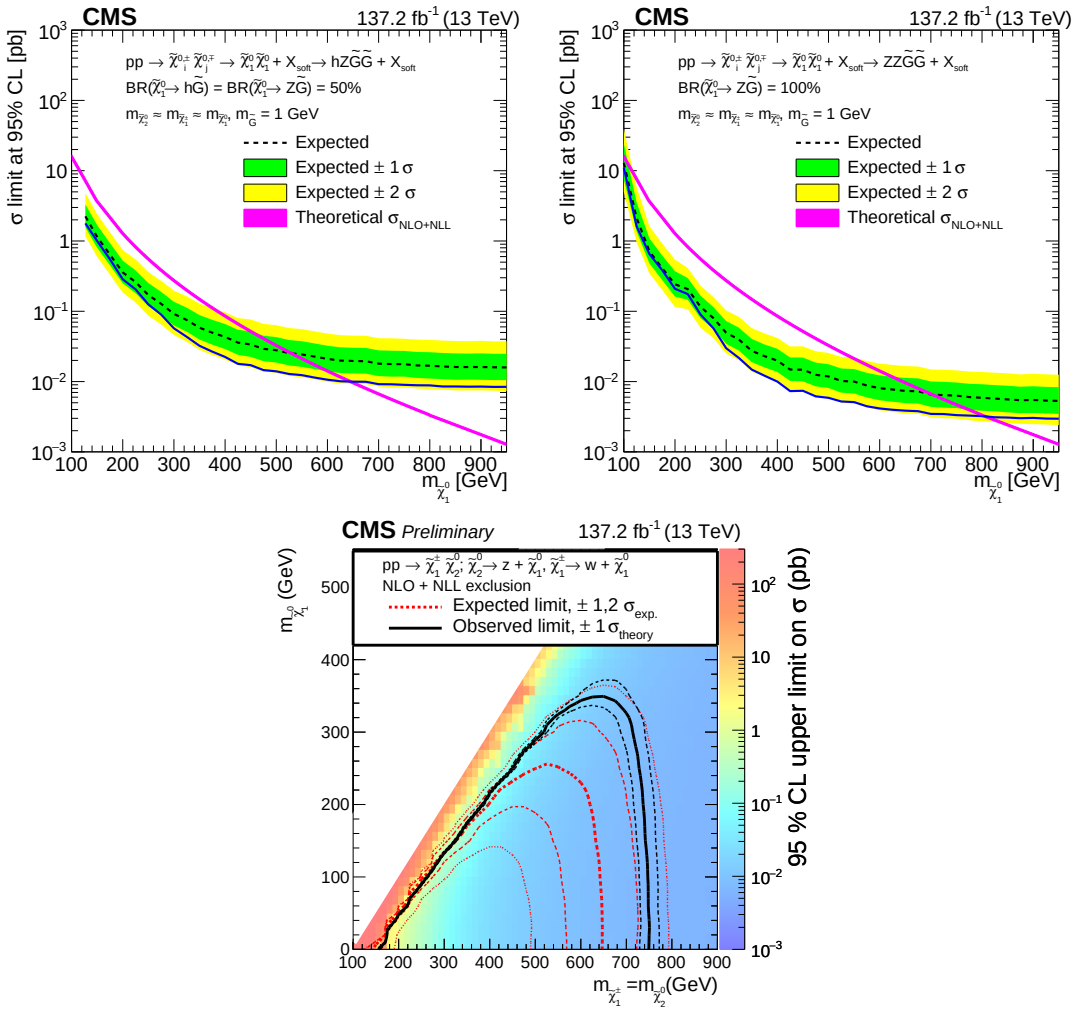


FIGURE 5.19: Upper row: observed upper limit on the signal cross section as a function of the $\widetilde{\chi}_1^0$ mass for the $\widetilde{\chi}_1^0\widetilde{\chi}_1^0$ model in the scenario with 50% branching ratio to Z and H bosons, and in the ZZ topology (right). The magenta line shows the theoretical cross section for the signal. Green (yellow) band shows the region expected to contain 68% (95%) of the limits expected under the background-only hypothesis. Bottom row: upper limits on the $\widetilde{\chi}_2^0\widetilde{\chi}_1^\pm$ production model as a function of the $\widetilde{\chi}_1^\pm$ and $\widetilde{\chi}_1^0$ masses. Observed (black) and expected (red) exclusion contours are also shown. The thin red dotted lines indicate the regions containing 68% and 95% of the expected limits under the background-only hypothesis. The thin dotted black line indicates the change in observed limit when varying the signal cross section by its uncertainty.

The results of the slepton edge region are interpreted in terms of upper limits in sbottom and squark production in the context of the slepton-edge models. The upper limits are shown in figure 5.20 as a function of the squark mass and the $\widetilde{\chi}_2^0$ mass. These results allow to probe sbottom masses between 1300 and 1600 GeV, and light flavor squark masses between 1600 and 1700 GeV, depending on the $\widetilde{\chi}_2^0$ mass. As described in section 5.1, the position of the edge resonance depends on the $\widetilde{\chi}_2^0$ mass and therefore exclusion limits are stronger in regions with low $\widetilde{\chi}_2^0$, which correspond to edge position of 20-60 GeV, where a slight deficit of data is seen.

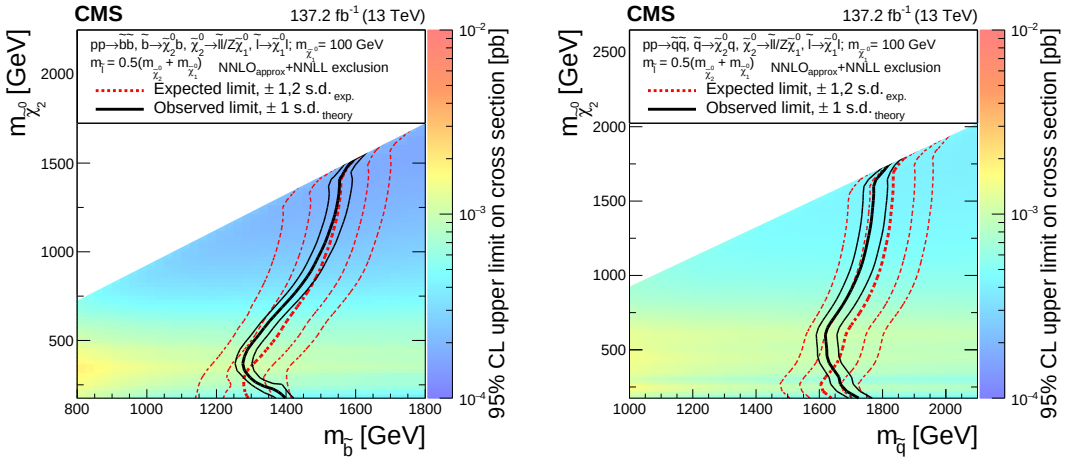


FIGURE 5.20: Observed upper limit at 95% CL for the slepton edge scenario for sbottom (left) and squark (right) production, as a function of the squark and $\widetilde{\chi}_2^0$ masses. Observed (black) and expected (red) exclusion contours are also shown. The thin red dotted lines indicate the regions containing 68% and 95% of the expected limits under the background-only hypothesis. The thin black line indicates the change in observed limit when varying the signal cross section by its uncertainty.

Finally, the results of the slepton signal regions are interpreted in the context of slepton pair production. Here, both the on-Z and off-Z regions are included in the signal extraction fit, to account for any potential signal contamination in the DY+jets estimation. Upper limits on slepton pair production are shown in figure 5.21 as a function of the slepton and $\widetilde{\chi}_1^0$ mass, showing that the results allow to probe slepton masses up to 650 GeV for low $\widetilde{\chi}_1^0$ mass.

5.7 Conclusions

A search for new physics in events with two opposite-sign same-flavor leptons has been shown in this chapter. The search presented has been performed with the data collected during the full Run 2 of the LHC, but it has been based on similar searches performed with data collected during 2015 and 2016.

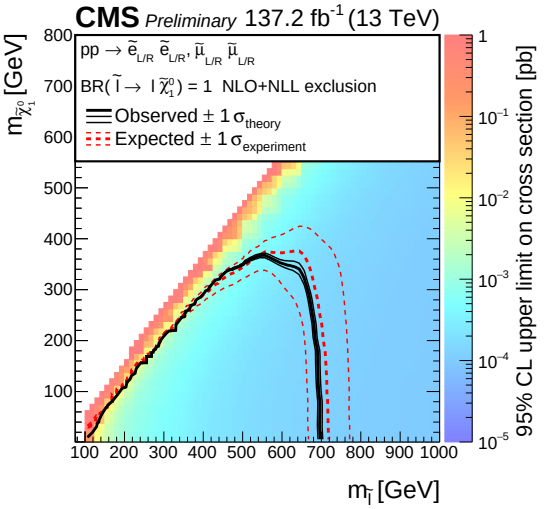


FIGURE 5.21: Observed upper limit at 95% CL for the direct slepton pair production model as a function of the slepton and $\widetilde{\chi}_1^0$ masses. Observed (black) and expected (red) exclusion contours are also shown. The thin red dotted lines indicate the regions containing 68% of the expected limits under the background-only hypothesis. The thin dotted black line indicates the change in observed limit when varying the signal cross section by its uncertainty.

This search is sensitive to a variety of SUSY models involving both strongly and electroweak produced SUSY. Ad-hoc signal regions have been defined targeting topologies inspired by a set of simplified SUSY models. Data-driven background estimation methods have been developed to estimate the contribution from the main backgrounds. Other backgrounds have been estimated using simulations, validated in dedicated control regions.

The results of the search have been found to be compatible with SM predictions, and are used to impose upper limits on the production of the simplified models.

6

Measurements of $\mathrm{t\bar{t}H}$ and tH production in the multilepton channel

In this chapter, the measurements of the production of the $\mathrm{t\bar{t}H}$ and tH processes are described. The chapter is outlined as follows. Section 6.1 contextualizes the measurement, describing the importance and features of signals and the main backgrounds. The general analysis strategy is described in section 6.2. Object and event selections are described in sections 6.3 and 6.4. Background estimation methods and signal models are described in section 6.5. Further categorization techniques are used to to perform the signal extraction. These techniques as well as the statistical model employed are described in section 6.6. Finally, results and their interpretation in the context of modified couplings of the Higgs boson are shown in sections 6.7 and 6.8.

6.1 The $\mathrm{t\bar{t}H}$ and tH processes

After the discovery of the Higgs boson, it is important to measure its properties and check their compatibility with the SM. Together with its spin, its coupling to massive bosons and fermions is one of the first properties to be checked. The main decay modes of the Higgs boson into some of the most massive particles WW, $\tau\tau$, $\mathrm{b\bar{b}}$ and ZZ allow to directly measure this coupling with relatively high precision. The coupling to second generation particles, such as muons or c quarks is too small to obtain a precise measurement with the available data.

6.1.1 Top Yukawa sector

The Yukawa coupling of the top quark to the Higgs boson cannot be studied directly in Higgs boson decays, since the decay into two top quarks is not allowed kinematically. It can instead be explored by measuring the Higgs boson production via gluon fusion, which is the dominant production mode at the LHC, or in its decay into a $\gamma\gamma$ pair. In these processes, for which some of their Feynman diagrams are shown in figure 6.1, the top quark contributions appear in loops. These loops can however receive contributions from BSM particles. Therefore it is desirable to measure this coupling in processes in which this coupling appears at tree level.

These measurements can be interpreted in the κ framework [149]. In this framework, only modifications of the coupling strengths by BSM effects are considered, assuming

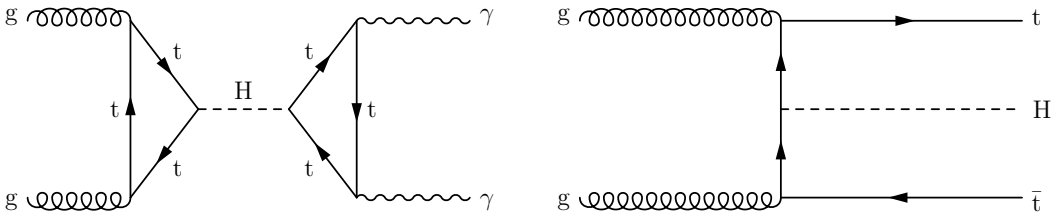


FIGURE 6.1: Left: Feynman diagram for the production of a Higgs boson in the gluon fusion mode and its decay into a photon pair. Right: leading order Feynman diagram for the production of a Higgs boson in association with a top quark pair.

the Lorentz structure of the coupling to be the same to the SM case. These coupling modifiers, κ_i, are defined in a way that cross section and partial widths associated to the interaction of the particle i and Higgs boson scale with a factor κ_i^2 at leading order. At higher orders, this scaling property is lost, however QCD NLO corrections usually factorize with this scaling.

In some cases, such as κ_γ or κ_g are used to scale the effective couplings of the Higgs to photons or gluons. These couplings can be resolved taken into account the contributions from the loops, assyming the SM matter content.

One of the processes that allows to measure κ_t with the highest precision is the associated production of a Higgs boson and a $t\bar{t}$ pair, $t\bar{t}H$. One of the leading order Feynman diagrams for this process is shown in figure 6.1. The study of $t\bar{t}H$ is the most precise way of measuring the Yukawa coupling at the LHC at tree level. At leading order, the cross section of this process is proportional to $|\kappa_t|^2$.

The $t\bar{t}H$ production cross section has been calculated at NLO in QCD with NLO electrowewak corrections [150]. The LO contribution features $\mathcal{O}(\alpha_S^2\alpha)$, while the NLO QCD contribution adds additional terms of the order $\mathcal{O}(\alpha_S^3\alpha)$, involving the $q\bar{q}$, gg and gq channels. The additional electroweak corrections comprise terms proportional to $\mathcal{O}(\alpha^3)$, $\mathcal{O}(\alpha_s\alpha^2)$, and $\mathcal{O}(\alpha_s^2\alpha^2)$. $\mathcal{O}(\alpha_s\alpha^3)$ and $\mathcal{O}(\alpha^4)$ are not included in the calculations as they are expected to be small.

The other main process sensitive to the Yukawa coupling at leading order is the associated production of one top quark and a Higgs boson, tH. This process also has the interesting feature of being affected by the relative sign of y_t and the coupling to the W boson to the Higgs. Two production modes occur for this process: the t-channel and the tW-channel. The s-channel is usually not considered due to its lower cross section.

The t-channel and tW-channel production modes are well defined up to LO in the 5FS. For higher orders or in the 4FS some of the processes interfere. For instance, t-channel interferes with tW and s-channels at higher orders in the hadronic decay of the W boson. Nevertheless, this interference is small and calculations at NLO exist in the 4FS and 5FS for the t-channel. The tW channel, in the other hand, interferes with $t\bar{t}H$ production at NLO. Calculations exist [151] that tackle this interference using the DR and DS schema covered in chapter 4.

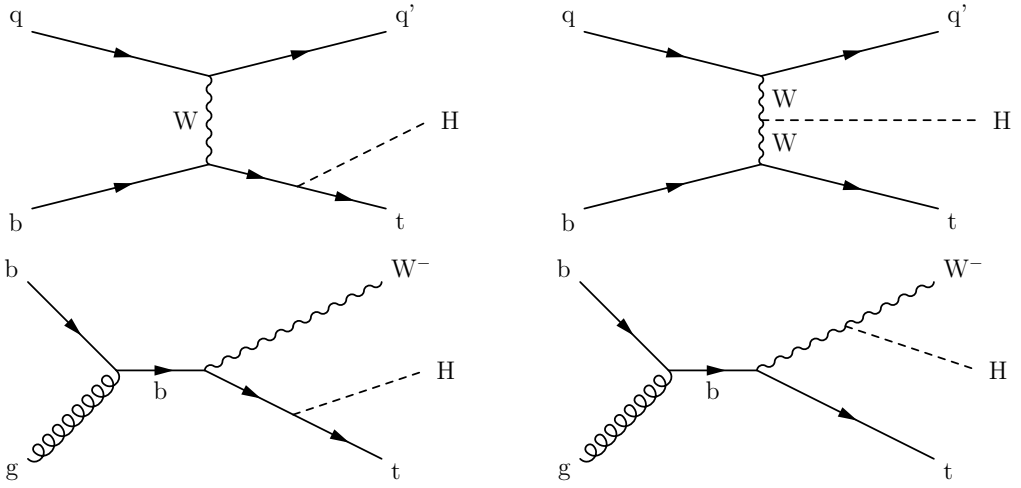


FIGURE 6.2: Leading order diagrams for tH production for the *t*-channel (top row) and tW-channel (bottom row), proportional to κ_t (left column) and to κ_W (right column).

These processes contribute with diagrams that scale proportionally to κ_t or with κ_W as shown in figure 6.2. Since the initial and final state is the same in the diagrams involving κ_t and κ_W, these diagrams interfere. At leading order and at $\sqrt{s} = 13$ TeV, the cross section for these processes can be written as a function of the coupling modifiers as [152]:

$$\sigma_{tHq} = (2.63\kappa_t^2 + 3.58\kappa_W^2 - 5.21\kappa_t\kappa_W)\sigma_{tHq}^{SM} \tag{6.1}$$

$$\sigma_{tHW} = (2.91\kappa_t^2 + 2.31\kappa_W^2 - 4.22\kappa_t\kappa_W)\sigma_{tHW}^{SM}. \tag{6.2}$$

In the SM case, the interference is destructive and the cross section for these processes is small. On the opposite case, with $|\kappa_t| = |\kappa_W| = 1$ and $\kappa_t/\kappa_W = -1$, the interference between the diagrams in constructive and the cross section for the *t*-channel and tW-channel is 11 and 9 times higher than in the SM prediction, respectively. This scenario is referred to as the Inverted Top Coupling (ITC) and makes tH measurements particularly interesting since it does not affect the cross section for $t\bar{t}H$ production.

The study of these processes can be used to probe CP violation in the Higgs sector. Even if the pseudoscalar Higgs boson is disfavored by experimental data [153], CP violation in the top-Higgs interaction is still allowed. In particular, beyond the κ framework, the Lagrangian of the top Yukawa sector can be generalized as follows [154, 155]:

$$\mathcal{L} = \bar{\psi}_t \left(\cos(\alpha)\kappa_{Htt} + \sin(\alpha)\kappa_{Att}\gamma^5\right) \frac{y_t}{\sqrt{2}}\psi_t\phi, \tag{6.3}$$

where α is a CP-mixing phase, and κ_{Htt} and κ_{Att} are dimensionless rescaling parameters. ψ_t corresponds to the top quark spinor, and ϕ is a scalar representing the Higgs field. The SM case is recovered when $\alpha = 0$, $\kappa_{Htt} = 1$ and $\kappa_{Att} = 0$. Instead $\alpha = \pi$ corresponds to the ITC case. These two scenarios correspond to the absence of CP

violation, while cases with $0 < \alpha < \pi$ allow some degree of violation. The pure CP odd case occurs with $\alpha = \pi/2$.

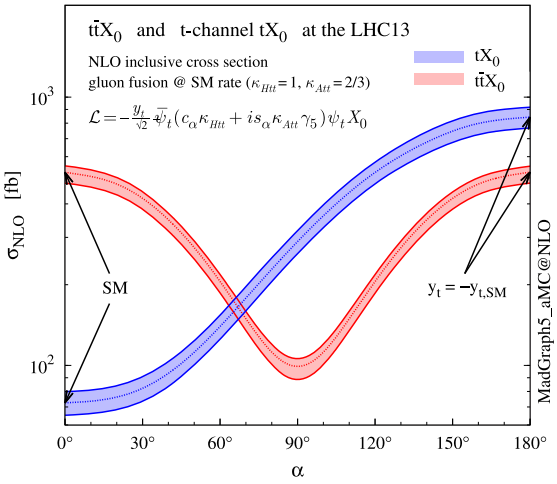


FIGURE 6.3: Dependence of $t\bar{t}H$ and tH production as a function of the α mixing angle. Taken from [155].

Of course, contributions to this Lagrangian are significantly constrained by measurements of the Higgs production cross section via gluon fusion and its $\gamma\gamma$ decay. However, the SM cross section for these processes is recovered when $\kappa_{Htt} = 1$ and $\kappa_{Att} = 2/3$. For these choices, all possible values of α will result in a gluon fusion cross section consistent with the SM. However, the cross sections for $t\bar{t}H$ and tH production are affected by this CP phase, as shown in figure 6.3.

$t\bar{t}H$ and tH production can also be sensitive to some of the terms of the Higgs potential [156]. As seen in the diagram in figure 6.4, higher order electroweak corrections to this process include diagrams proportional to κ_λ, the modifier of the trilinear term of the Higgs potential. Similar diagrams appear in the tH production, also shown in figure 6.4. This figure also shows the effect of the κ_λ correction with respect to the LO calculation, confirming the potential sensitivity of differential $t\bar{t}H$ and tH measurements to κ_λ.

6.1.2 Experimental status

The production rate of $t\bar{t}H$ in pp collisions is shown as a function of $\sqrt{s}$ in figure 6.5, compared to other Higgs production modes. The $t\bar{t}H$ cross section is one order of magnitude lower than the main Higgs production modes, and the tH cross section is one order of magnitude lower than $t\bar{t}H$. This makes the study of this process very challenging. Additionally, the different decays of the top quarks and the Higgs boson lead to a large variety of final states with very different topologies.

The $b\bar{b}$ decay of the Higgs is the dominant mode in terms of rate, but is affected by large backgrounds in LHC measurements. It is searched for in fully hadronic, single lepton and dilepton channels. The fully hadronic channel is largely affected by multijet background, while the $t\bar{t}$ + jets background affects all the channels. The latter includes contributions from $t\bar{t}$ in association with light jets, with c-jets, and the irreducible contribution from $t\bar{t}$ + $b\bar{b}$. The latest results in this decay mode were obtained

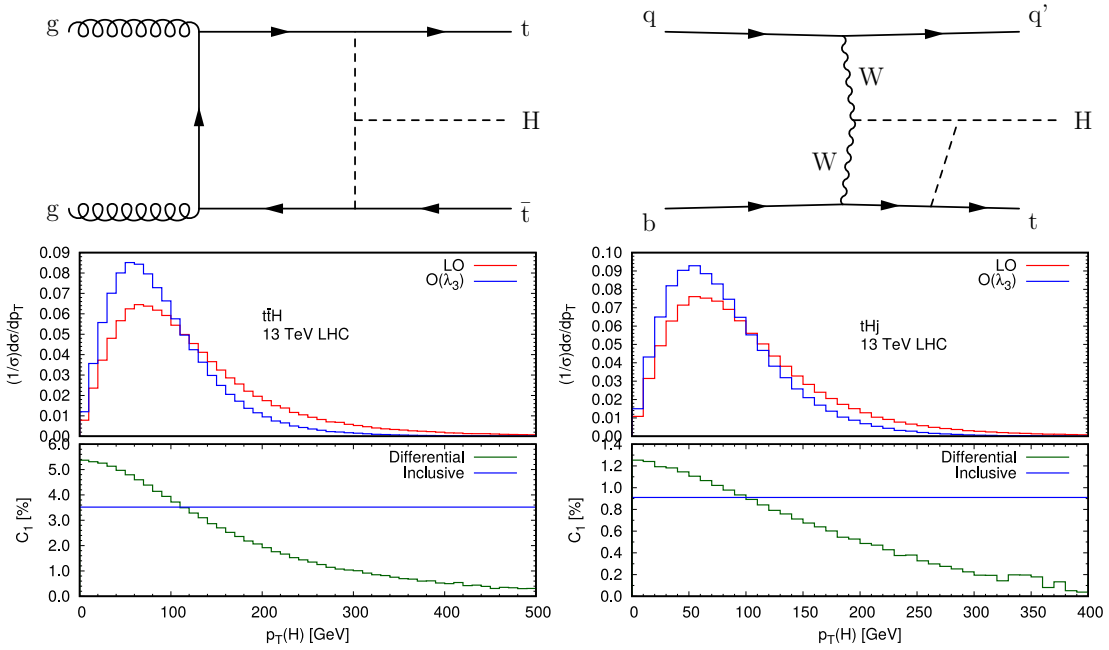


FIGURE 6.4: Top: Feynman diagrams showing the contributions of the κ_λ coupling to $t\bar{t}H$ (left) and tH production (right). Bottom: Effect of the κ_λ corrections in the distribution of the Higgs p_T in $t\bar{t}H$ (left) and tH (right) production. Taken from [156].

by the CMS Collaboration with 77.4 fb^{-1} of pp collisions [157], and with 36.1 fb^{-1} by the ATLAS Collaboration [158]. Both analyses make use of multivariate analysis techniques to discriminate signal from background events. The CMS analysis achieves a sensitivity of 3.9 standard deviations to the $t\bar{t}H$ production, while the analysis by ATLAS achieves a sensitivity with 1.4 standard deviations with the single lepton and dilepton categories only.

The $\gamma\gamma$ decay mode of the Higgs boson allows for a cleaner environment with more reduced background. This final state also allows to easily resolve the Higgs system, which makes this channel suitable for differential studies. Additionally, the available datasets recorded by the ATLAS and CMS experiment make measurements in this channel competitive to those of the rest of the channels. Both Collaborations have released results with the full Run 2 datasets [159, 160], reaching an observation of the process with 5.2 and 6.6 standard deviations, compared to 4.4 and 4.7 expected, respectively. The analysis by ATLAS also includes a measurement of tH production, but still without enough sensitivity to provide evidence for this process. Both analyses include a study of the CP violation in the top Yukawa sector, excluding the pure CP-odd model at 3.9 and 3.2 σ, respectively.

This chapter of the thesis is dedicated to the measurement of $t\bar{t}H$ and tH production in the multilepton channel, so it will be described in greater detail in the following sections. This channel probes the WW and $\tau\tau$ decay channels, although it has a small contribution from the ZZ decays. These decay modes have a moderately large branching fraction, and searches for these processes are made in final states in which only small SM backgrounds are expected, such as regions with at least same-sign leptons and moderately large jet multiplicity. The search shown in this chapter corresponds to the CMS measurement performed with the full run 2 dataset. Previous searches to

which I have contributed to are the ones released with the 2016 and 2017 dataset [14, 16], obtaining a signal strength for $t\bar{t}H$ production of $0.96^{+0.34}_{-0.31}$. The latest result by the ATLAS Collaboration was shown in [161], measuring a signal strength for $t\bar{t}H$ of $0.58^{+0.36}_{-0.33}$.

Measurements of the ZZ decay of the Higgs boson [162, 163] also typically include dedicated categories targeting the $t\bar{t}H$ production modes. However, despite of the interest of these final states, their branching ratio is still too small for a measurement of the process with the available datasets.

In addition to the recent observations of this process in the $\gamma\gamma$ channel, the $t\bar{t}H$ production had already been observed in the datasets collected during 2016 by the ATLAS and CMS Collaborations [15, 164]. These observations were made by combining the most sensitive of the channels described above, including the one described in this thesis.

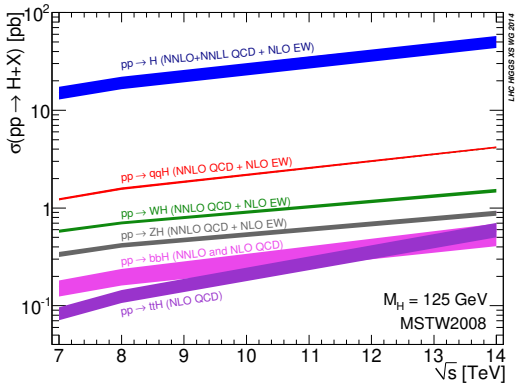


FIGURE 6.5: Production cross section for the main Higgs production modes in pp collisions as a function of $\sqrt{s}$. Taken from [165].

6.1.3 Irreducible backgrounds in the multilepton channel

On top of the importance of the $t\bar{t}H$ process itself for the reasons mentioned above, some of its backgrounds in the multilepton channels are interesting themselves as studies of the SM consistency or as probes to BSM physics. This is the case for $t\bar{t}Z$ and $t\bar{t}W$ production, which are the leading contribution of the irreducible backgrounds to the measurement. Their production cross section is small, however it is comparable to that of the signals.

$t\bar{t}Z$ production is interesting as it gets contributions from diagrams in which the Z boson couples to the top quarks at leading order. Therefore it is one of the best probes to test this coupling in the SM. Additionally, both $t\bar{t}Z$ and $t\bar{t}W$ are sensitive to BSM physics affecting this and other couplings, that can be parametrized as additional terms in the SM Lagrangian.

$t\bar{t}Z$ and $t\bar{t}W$ production cross sections have been calculated at NLO accuracy with electroweak corrections in reference [150], which has been used as a reference in recent results. These corrections, similarly to those in $t\bar{t}H$, do not include $\mathcal{O}(\alpha_s\alpha^3)$ terms. However, the contribution from this correction is sizable for $t\bar{t}W$ production, as it

contains the $gq \to t\bar{t}W^{\pm}q'$ real emission channel that includes $tW \to tW$ scattering, and represents a 12% correction with respect to the LO cross section [166, 167]. The full NLO+NNLL cross section including electroweak corrections is reported in this reference [168]. Additionally, the presence of NLO QCD corrections with one extra parton can be sizable, of the order of an additional 10% [29]. Finally, it was recently suggested that off-shell effects in the production of the top quarks yield to negligible corrections to the inclusive cross-section, but may have a sizable effect in kinematic distributions [169].

The ATLAS and CMS Collaborations have performed measurements of these processes at $\sqrt{s} = 8$ and 13 TeV. The most recent results on the topic reach high level of precision for $t\bar{t}Z$ [72, 170], showing a fairly good compatibility with the state-of-the-art predictions. On $t\bar{t}W$ both ATLAS and CMS have released results in which data shows a preference for higher values of the cross section for this process than the reference cross section [161, 171].

6.2 Analysis strategy

In the following sections, the measurement of the $t\bar{t}H$ and tH production cross section in the multilepton channel performed with data collected with the CMS detector is described. This analysis is performed in evens with at least two light leptons and hadronically decaying τ in the final state. Event categories are built depending on the lepton multiplicity and flavor. Categories with requirements on hadronic τ (τ_h) provide sensitivity to the $H \to \tau\tau$ decay mode, while the remaining ones target the WW and ZZ decays, although they provide some sensitivity to the $\tau\tau$ channel thanks to the leptonic decays of the τ.

The topologies selected in the analysis require relatively large lepton multiplicity or lepton pairs with equal charge, and are chosen in such a way that only rare SM processes, such as the signal, are accepted. Because of this, and since the signal cross section is very small, typical SM backgrounds with cross section orders of magnitude larger but with a different final state topology may have a sizable contribution due to events with misidentified leptons or with leptons coming from non-prompt decays.

Because of this, one of the key ingredients of the analysis is object identification. This is crucial in the case of leptons, as large discrimination is needed to reject leptons coming from non-prompt decays of partons, while keeping a high signal efficiency. For that purpose, the MVA techniques for electron and muon identification have been developed, as shown in chapter 3. Additionally, state-of-the-art identification algorithms are used to discriminate τ_h from jets.

On top of the lepton multiplicity selections, subcategories are defined by applying additional criteria are applied to the jet multiplicity and kinematic variables describing the global features of the event. Then, events are further classified according to multivariate discriminators taking as inputs kinematic variables of the reconstructed objects. These classifiers are trained to discriminate signal from the various backgrounds in each category. Additionally, a control analysis is performed to check the consistency of the result in the most sensitive categories without using multivariate discriminators trained in simulations.

Despite of the usage of advanced techniques for background discrimination, signal regions have a sizable contribution from backgrounds. This contribution needs to be accounted for in an accurate way in order to achieve a high precision in the measurement of the signal. Two types of background processes are considered: irreducible and reducible backgrounds.

Irreducible backgrounds are those in which the reconstructed leptons are genuine leptons coming from prompt decays, and reconstructed τ_h are produced in hadronic τ decays. They are dominated by $t\bar{t}W$ and $t\bar{t}Z$ production in the main categories, but with significant contributions from tZq and WZ. $t\bar{t}$ and Drell-Yan events also appear as irreducible backgrounds in some categories. These processes are estimated using state-of-the-art simulations, and are normalized to the most accurate cross section prediction available, as indicated in the following sections. Additionally, control regions are defined whenever is possible to check the level of agreement between data and these models.

Reducible backgrounds are those which enter the signal region, but their topology does not fit the signal region definition. Here it is distinguished between non-prompt leptons, flips and conversions. Charge flips are events with leptons in which the sign of their charge has been measured with the opposite sign, while conversions refer to electrons which have been produced in the conversion of a photon. The contribution from charge flips and events with non-prompt leptons in the final state are estimated using dedicated data-driven methods, while conversions are estimated with simulations.

The work described in this chapter has been performed in collaboration with other groups in the CMS Collaboration: CERN, IHEP, LLR/CNRS (Ecole Polytechnique), NICPB (Tallin) and UCL. My contributions to the analysis focus on the categories without τ_h. I have contributed on the development of the electron and muon identification criteria and their optimization, to the estimation of the non-prompt lepton background, to the development of new control regions for irreducible, and to the signal extraction fit.

6.3 Object selection

As mentioned above, the selection of objects is designed to achieve the maximum purity, particularly on leptons, which are used to define the topology of event candidates.

6.3.1 Light lepton selection

Three types of light lepton (e and μ) selections are defined. "Loose" leptons are the minimal selection on top of which the other two definitions are constructed. It is designed to be very inclusive, keeping a large signal efficiency, but also a quite large acceptance of leptons coming from quark decays. Loose leptons are only used to remove duplicate leptons that are close-by and to reject low mass resonances or efficiently reduce events with a Z candidate.

Loose leptons are required to have at least 5 GeV and to be within the acceptance, $|\eta| < 2.4$ for muons and $|\eta| < 2.5$ for electrons. Additionally, they are required to

have an impact parameter with respect to the primary vertex of less than 0.05 cm in the transverse plane and 0.1 cm in the direction along the beam axis. The significance of this impact parameter, defined as the distance over its associated uncertainty, is also required to be less than 8 standard deviations. A mild isolation cut is applied to these leptons by requiring mini-isolation to be less than 40% the p_{T} of the lepton. Muons are required to pass the "loose" working point 3.1.2. For electrons, the p_{T} requirement is raised to $p_{\mathrm{T}} > 7$ GeV, and are required to have at most one missing hit from its expected trajectory in the tracker, and to pass the "loose" MVA selection defined in 3.2.

Leptons passing the "fakeable" lepton selection are used to estimate the contribution from non-prompt leptons in data samples. This selection is constructed to keep a high acceptance rate for non-prompt leptons, but it is tuned in order to have a good closure of the non-prompt estimation method in simulations. Since these leptons are used to estimate the contribution from non-prompt leptons, all the regions are constructed applying cuts on the fakeable object multiplicity. Jets that have been clustered together with a fakeable lepton are not taken into account for the event selection, they are only used for building the prompt lepton MVA.

Fakeable leptons are required to pass the loose identification criteria. Additionally, the jet associated to the lepton is required to fail the medium working point of the DeepFlavor discriminant. This cut is applied to avoid removing jets passing this criterion, which are likely to be b jets in which a non-prompt lepton is produced. Additionally, electrons are required to have no missing hits from the expected trajectory, and to pass a selection designed to be tighter than the requirements at the HLT. This selection is imposed on the ratio of the energy in the HCAL to the energy in the ECAL, $H/E < 0.1$; the difference between the inverse of the electron cluster energy and the inverse of the track momentum, $\frac{1}{E} - \frac{1}{p} > -0.04$, and the width of the electron cluster in the η direction, $\sigma_{i\eta i\eta} < 0.011(0.03)$ for $\eta_{SC} < 1.479$ ($\eta_{SC} > 1.479$). Electrons are also rejected if they are associated with a successfully reconstructed conversion vertex. Finally, electrons that are closer to a loose muon than 0.3 in ΔR are discarded.

Then, additional requirements are applied to leptons not passing the tight prompt-lepton MVA identification criteria described below. For electrons, the jet relative isolation variable, defined in section 3.4, is required to be less than 0.7. For muons, this variable is required to be less than 0.5 and the Deep Jet score of the associated jet is required to fail a working point that ranges between the loose and the medium, depending on the muon cone-p_{T}.

The cone-p_{T} variable is designed to be, in non-prompt leptons, a proxy of the p_{T} of the parton that has originated the jet in which the lepton is produced and, in prompt-leptons, a proxy of the actual lepton p_{T}. It is therefore defined as the p_{T} for leptons passing the tight selection, and as $0.9p_{\mathrm{T}}(1 + I_{jet})$ for the rest, where I_{jet} is the jet relative isolation. This variable is more suitable to parametrize the fake-rate, therefore the event-level variables are defined on top of it. In this chapter, cone p_{T} is referred to as p_{T} for fakeable leptons, unless explicitly mentioned. Fakeable leptons are required to have cone-$p_{\mathrm{T}} > 10$ GeV.

Finally, the "tight" selection is used to achieve the maximum prompt lepton purity. Leptons in this category are required to have a prompt-lepton MVA score, defined in section 3.4, greater than 0.85 (0.8) for electrons (muons). Muons are additionally required to pass the medium criteria described in section 3.1.2. The working point of

the cut on the prompt-lepton MVA has been carefully tuned to achieve the maximum sensitivity in the analysis. All the selection criteria in electrons and in muons are summarized in tables 6.1 and 6.2, respectively.

Observable	Loose	Fakeable	Tight
(Cone-)p_{T}	$p_{\mathrm{T}} > 7$ GeV	> 10 GeV	> 10 GeV
$\lvert\eta\rvert$	< 2.5	< 2.5	< 2.5
$\lvert d_{xy}\rvert$	< 0.05 cm	< 0.05 cm	< 0.05 cm
$\lvert d_z\rvert$	< 0.1 cm	< 0.1 cm	< 0.1 cm
d/σ_d	< 8	< 8	< 8
I_{e}	$< 0.4 \times p_{\mathrm{T}}$	$< 0.4 \times p_{\mathrm{T}}$	$< 0.4 \times p_{\mathrm{T}}$
$\sigma_{i\eta i\eta}$	–	$<$ { 0.011 / 0.030 }[1]	$<$ { 0.011 / 0.030 }[1]
H/E	–	< 0.10	< 0.10
1/E - 1/p	–	> -0.04	> -0.04
Conversion rejection	–	✓	✓
Missing hits	≤ 1	$= 0$	$= 0$
MVA ID	>WP-loose[2]	>WP-80 (>WP-loose)[2]†	>WP-loose[2]
Deep Jet of nearby jet	–	<WP-medium	<WP-medium
Jet relative isolation[4]	–	< 0.7 (–) †	–
Prompt-e MVA	–	< 0.80 (> 0.80)	> 0.80

Table 6.1: Loose, fakeable and tight selection criteria for electrons. [1]The requirement $\sigma_{i\eta i\eta}$, is varied as function of η_{SC} of the electron candidate.
[2] WP-loose and WP-80 make reference to the identification criteria defined in section 3.2.
† Fails (passes) the requirement prompt-e MVA > 0.80.
A hyphen (–) indicates selection criteria that are not applied.

Observable	Loose	Fakeable	Tight
p_{T}	> 5 GeV	> 10 GeV	> 10 GeV
$\lvert\eta\rvert$	< 2.4	< 2.4	< 2.4
$\lvert d_{xy}\rvert$	< 0.05 cm	< 0.05 cm	< 0.05 cm
$\lvert d_z\rvert$	< 0.1 cm	< 0.1 cm	< 0.1 cm
d/σ_d	< 8	< 8	< 8
I_{μ}	$< 0.4 \times p_{\mathrm{T}}$	$< 0.4 \times p_{\mathrm{T}}$	$< 0.4 \times p_{\mathrm{T}}$
PF muon	>WP-loose[1]	>WP-loose[1]	>WP-medium[1]
Deep Jet of nearby jet	–	<WP-interp. (<WP-medium)[2]	<WP-medium[2]
Jet relative isolation[3]	–	<0.5 (–) †	–
Prompt-μ MVA	–	< 0.85 (> 0.85)	> 0.85

Table 6.2: Loose, fakeable and tight selection criteria for muons.
[1] Loose and medium ID make reference to the criteria described in chapter 3.
[2]Upper cut on the Deep Jet score defined with a linear interpolation from Deep Jet WP-medium at cone-p_{T} 20 GeV to Deep Jet WP-loose at cone-p_{T} 45 GeV.
† Fails (passes) the requirement prompt-μ MVA > 0.85.
A hyphen (–) indicates selection criteria that are not applied.

6.3.2 τ_h selection

τ_h are reconstructed using the method described in section 3.6, and the working points defined in the Deep Tau v2.1 discriminators are used. Similarly to the selection of light leptons and with the same purposes, three levels of τ_h identification criteria are defined for τ_h: loose, fakeable and tight.

Loose τ_h are required to have $p_T > 20$ GeV and $|\eta| < 2.3$. Impact parameter is required to be less than 0.2 along the beam direction, respectively. The Deep Tau discriminator against jets is required to pass the VVLoose working point.

Fakeable τ_h are additionally required to pass the VLoose Deep Tau discriminator against muons and the VVVLoose Deep Tau discriminator against electrons. Only τ_h that have not been reconstructed as the 2-prong+π^0 decay mode are considered. Tight τ_h implement additional cuts on the Deep Tau discriminator against jets, that depend on the category. All these selections are summarized in table 6.3.

Observable	Loose	Fakeable	Tight
p_T	> 20 GeV	> 20 GeV	> 20 GeV
$\|\eta\|$	< 2.3	< 2.3	< 2.3
$\|d_z\|$	< 0.2 cm	< 0.2 cm	< 0.2 cm
Decay mode finding	New	New	New
Decay modes	All	All except 2-prong($+\pi^0$)	All except 2-prong($+\pi^0$)
DeepTau vs. jets	$>$ WP-VVLoose	$>$ WP-VVLoose	Channel-dependent
DeepTau vs. muons	–	$>$ WP-VLoose	$>$ WP-VLoose
DeepTau vs. electrons	–	$>$ WP-VVVLoose	$>$ WP-VVVLoose

TABLE 6.3: Loose, fakeable and tight selection criteria for hadronic τ decays. A hyphen (–) indicates selection criteria that are not applied.

6.3.3 Jets and b-tagging

Jets in the analysis are reconstructed following the procedure outlined in section 2.3.2. They are required to have $|\eta| < 5$ and $p_T > 25$ GeV, maximizing the acceptance to account for the spectator quark in tH events, which is expected to be emitted in the forward direction with a significant rate. Jets within the tracker acceptance are referred to as "jets", while the rest are dubbed as "forward jets". Jets with $2.7 < |\eta| < 3$ are required to have $p_T > 60$ GeV, as this region is problematic due to noise in the calorimeters, which introduces a large bias in the p_T estimation of soft jets.

Since all PF candidates are introduced in the jet clustering algorithm, leptons may be also identified as jets. Since we are only interested in jets produced in the parton shower of a colored particle, jets that include the PF candidate associated to a fakeable electron or muon, or are closer in $\Delta R < 0.3$ to a loose τ_h.

Jets are considered as b-tagged if they pass the medium or loose working point of the Deep Jet discriminant. Some of the MVA variables used in the analysis make use of the score of the discriminator. Ad-hoc calibrations are applied to simulations in order for the shape of the discriminator to be well modeled.

Finally, we also define "light jets" as the set jets with $|\eta| < 2.4$ that fail the loose b-tag jet requirement or have $|\eta| > 2.4$.

6.4 Event selection and classification

Events are recorded using a set of single-, double- and triple-lepton triggers, lepton+τ_h triggers, and double τ_h triggers. Triggers are required to be consistent with the minimum lepton requirement on each category. For instance, events in a category with only two leptons are required to have fired a single- or double-lepton trigger, while no requirements are applied on the triple-lepton triggers or triggers with τ_h. These triggers require leptons and τ_h to have a p_T above a certain threshold, in order to keep their acceptance rate within acceptable values. These thresholds vary by a few GeV depending on the data taking period, to account for its conditions. Trigger paths, together with their thresholds during the Run 2 data taking period are shown in table 6.4.

Trigger path	Threshold (GeV)
Single μ	22-27
Single e	25-35
Double e	23 (leading) 12 (sub-leading)
Double μ	17 (leading) 8 (sub-leading)
μ +e	23 (leading) 12-8 (sub-leading)
Triple lepton	16-8 (leading) 12-9 (sub-leading) 9-5 (third)
μ + τ_h	19-20 (μ) 20-27 (τ)
e + τ_h	24 (e) 30-20 (τ)
Double τ_h	35-40

TABLE 6.4: Typical p_T threshold for the triggers used in the analysis. When two numbers are included, the threshold varied between those numbers during the data taking era.

As mentioned in the previous sections, several signal regions are defined according to the lepton and τ_h multiplicity of the event. Each one of these categories aims for a specific decay mode of the top quarks and the Higgs boson, therefore specific cuts are applied for each signal region separately. All events that have a pair of loose leptons with invariant mass smaller than 12 GeV are rejected, in order to avoid contributions from low mass resonances.

6.4.1 Signal regions

Signal regions used in the analysis are defined in this section. Ten disjoint signal regions are defined according to their light lepton and τ_h multiplicity, targeting different decay modes of the Higgs boson and the top quarks. Three of the defined regions, the $2l_{SS} + 0\tau_h$, $3l + 0\tau_h$ and $2l_{SS} + 1\tau_h$ aim to be sensitive to both $t\bar{t}H$ and tHq processes, while the rest of them target $t\bar{t}H$ production only. The targeted $t\bar{t}H$ decay modes are described in table 6.5.

Signal region	Targeted $t\bar{t}H$ decay mode
$2l_{SS} + 0\tau_h$	$t \to blν$, $t \to bqq'$ with $H \to WW \to lνqq'$
$3l + 0\tau_h$	$t \to blν$, $t \to blν$ with $H \to WW \to lνqq'$ $t \to blν$, $t \to bqq'$ with $H \to WW \to lνlν$ $t \to blν$, $t \to bqq'$ with $H \to ZZ \to llqq'$ or $llνν$
$2l_{SS} + 1\tau_h$	$t \to blν$, $t \to bqq'$ with $H \to ττ \to lννν\tau_h ν$
$4l + 0\tau_h$	$t \to blν$, $t \to blν$ with $H \to WW \to lνlν$ $t \to blν$, $t \to blν$ with $H \to ZZ \to llqq'$ or $llνν$
$1l + 2\tau_h$	$t \to blν$, $t \to bqq'$ with $H \to \tau^+\tau^- \to \tau_h ν\tau_h ν$
$2l + 2\tau_h$	$t \to blν$, $t \to blν$ with $H \to \tau^+\tau^- \to \tau_h ν\tau_h ν$
$3l + 1\tau_h$	$t \to blν$, $t \to blν$ with $H \to ττ \to lννν\tau_h ν$
$0l + 2\tau_h$	$t \to bqq'$, $t \to bqq'$ with $H \to ττ \to \tau_h ν\tau_h ν$
$1l + 1\tau_h$	$t \to bqq'$, $t \to bqq'$ with $H \to ττ \to lννν\tau_h ν$
$2l_{OS} + 1\tau_h$	$t \to blν$, $t \to bqq'$ with $H \to \tau^+\tau^- \to lννν\tau_h ν$

TABLE 6.5: Targeted $t\bar{t}H$ decay modes in each one of the signal regions.

Leptons in each region are required to have p_T and η within the detector acceptance and fulfilling the thresholds of the trigger requirements. These requirements for each region are described in table 6.6. The requirements on the tight τ_h selection working point are also outlined in the table. The charge of leptons and τ_h are required to be consistent with the expectation in each decay mode. Some of the signal regions implement specific requirements in the charge of the two leptons. For instance, the $2l_{SS} + 0\tau_h$ region selects only events with two same-sign leptons. Although signal is yields to both same and opposite-sign pairs, the same-sign topology is very rare in SM processes, yielding to a more favorable signal-to-background ratio.

The set of kinematic selections applied to each of the signal regions is outlined in table 6.7. All signal regions reject events with a pair of loose leptons with an invariant mass, m_{ll}, smaller than 12 GeV, in order to suppress contributions from low mass resonances. To reject events from processes with on-shell Z bosons in the final state, a Z-veto selection is applied to some of the regions. This selection rejects events with an opposite-same same-flavor pair of loose leptons fulfilling $|m_{ll} - m_Z| < 10$ GeV. For regions in which a same-sign lepton pair is required, $2l_{SS} + 0\tau_h$ and $2l_{SS} + 1\tau_h$, this veto is also applied to the same-sign lepton pair if the two leptons are electrons, in order

Region	Lep. p_T (GeV)	Lep. $\|\eta\|$	τ_h p_T (GeV)	τ_h $\|\eta\|$	τ_h id. WP
$2lss + 0\tau_h$	$> 25/15$	<2.5 (2.4) for e (μ)	–	–	–
$3l + 0\tau_h$	$> 25/15/15$	<2.5 (2.4) for e (μ)	–	–	–
$2lss + 1\tau_h$	$> 25/15$	<2.5 (2.4) for e (μ)	> 20	< 2.3	vloose
$4l + 0\tau_h$	$> 25/15/15/10$	<2.5 (2.4) for e (μ)	–	–	–
$1l + 2\tau_h$	> 30 (25) e (μ)	<2.1	30/20	< 2.1	medium
$2l + 2\tau_h$	$> 25/15$	<2.5 (2.4) for e (μ)	20	< 2.3	medium
$3l + 1\tau_h$	$> 25/15/15$	<2.5 (2.4) for e (μ)	20	< 2.3	vloose
$0l + 2\tau_h$	–	–	40	< 2.1	loose
$1l + 1\tau_h$	> 30 (25) e (μ)	<2.1	30	< 2.1	medium
$2los + 1\tau_h$	$> 25/15$	<2.5 (2.4) for e (μ)	> 20	< 2.3	tight

TABLE 6.6: Outline of the kinematic selections for light leptons and τ_h in each one of the signal regions. The rightmost column shows the identification working point applied to the τ_h discriminator against jets.

to reject contributions from DY+jets events in which their charge has been measured incorrectly.

Additionally to the lepton and τ_h selection, criteria are applied in the minimum central jet multiplicity. This requirement is applied taking into account the expected jet multiplicity in each decay mode, but also aims to select events in which several jets are outside the acceptance. Additionally, a b-tagged jet multiplicity requirement is applied for each signal regions targeting $t\bar{t}H$ production. This selection requires events to contain at least two loose or one medium b-tagged jets. For the regions targeting tHq production, a requirement on the number of light jets and medium b-tagged jets is applied, taking into account that the spectator quark may be emitted in the forward direction and only one b-tagged jet is expected.

Some of the regions also implement selections on p_T^{miss} LD, defined in section 3.6, aiming to reject events without neutrinos in the final state. The threshold of this selection, which is different for each signal region, also depends on the jet multiplicity of the event, the lepton flavor and charge. Finally, in order to ensure all signal regions are disjoint among themselves, events that contain more leptons than required, and passing tight identification criteria are not included in the signal region. In order to also be disjoint from the dedicated $H \to ZZ$ analysis [163], events with at least four leptons and $m_{4l} < 140$ GeV are rejected.

6.4.2 Control regions

To check the modeling of some of the main irreducible backgrounds, control regions are defined to check their modeling in simulations. The 3l control region is used to validate WZ and $t\bar{t}Z$ events, while the 4l control region is built to check the agreement of ZZ and $t\bar{t}Z$ events.

These regions are built by relaxing and inverting cuts from the $3l + 0\tau_h$ $t\bar{t}H$ and the $4l + 0\tau_h$ region, respectively.

	Lep. mult.	τ_h mult.	Jet mult.	b jet multiplicity	Light jets	b-tagged jets	Z veto	p_T^{miss} LD (GeV)	$\sum_{i=l,\tau} Q_i$
$2lss + 0\tau_h$ $t\bar{t}H$	2 (same-sign)	0	≥ 3	✓			✓	>30 if ee	+2,-2
$2lss + 0\tau_h$ tHq	2 (same-sign)	0			$\geq$1	≥ 1	✓	>30 if ee	+2,-2
$3l + 0\tau_h$ $t\bar{t}H$	3	0	≥ 2	✓			✓	>30 (45)[1]	+1, -1
$3l + 0\tau_h$ tHq	3	0			$\geq$1	≥ 1	✓		+1, -1
$2lss + 1\tau_h$ $t\bar{t}H$	2 (same-sign)	1	≥ 3	✓			✓	>30 if ee	+1, -1
$2lss + 1\tau_h$ tHq	2 (same-sign)	1			$\geq$1	≥ 1	✓	>30 if ee	+1, -1
$4l + 0\tau_h$	4	0	≥ 2	✓			✓	>30 (45)[1]	0
$1l + 2\tau_h$	1	2	≥ 3	✓					
$2l + 2\tau_h$	2	2	≥ 2	✓			✓		0
$3l + 1\tau_h$	3	1	≥ 2	✓			✓	>30 (45)[1]	0
$0l + 2\tau_h$	0	2	≥ 4	✓					0
$1l + 1\tau_h$	1	1	≥ 4	✓					
$2los + 1\tau_h$	2 (opposite-sign)	1	≥ 3	✓			✓	>30[2]	

TABLE 6.7: Definition of signal regions in the main analysis. The $2lss + 0\tau_h$, $3l + 0\tau_h$ and $2lss + 1\tau_h$ regions are constructed by the union of events in the $t\bar{t}H$ and tHq regions. [1] The selection is 30 (45) GeV for events with (without) an opposite-sign same-flavor lepton pair, and is not applied in events with at least 4 jets. [2] This cut is applied only to events in which the two selected leptons have the same flavor. In the $2lss + 0\tau_h$ and $2lss + 0\tau_h$ the Z veto is also applied to the same-sign lepton pair if the two leptons are electrons.

3l control region The three lepton control region is defined by inverting the veto on an opposite-sign same-flavor pair of loose leptons with an invariant mass close to the Z boson mass of the $3l + 0\tau_h$ $t\bar{t}H$ region. Additionally, no explicit cut on the jet or b jet multiplicity is applied. Instead, events are classified in bins of jet and b jet multiplicity. All of the classes require at least one reconstructed jet. Classes with low (high) jet multiplicity are expected to be more enriched in WZ ($t\bar{t}Z$) events.

4l control region The four lepton control region is defined by inverting the Z veto of the $4l + 0\tau_h$ region. Again, no explicit cuts on the jet multiplicity are applied. In this case, events are classified depending on the presence of a second opposite-sign same-flavor pair, the jet and b jet multiplicity.

The categories in the 3l and 4l regions are outlined in table 6.8. The event yield in these categories is used in the signal extraction fit to constrain the irreducible backgrounds.

Category	Second on-Z OSSF pair	Jet multiplicity	b jet multiplicity
		3l region	
1	–	= 1	= 0
2	–	= 2	= 0
3	–	= 3	= 0
4	–	>3	= 0
5	–	= 2	= 1
6	–	= 3	= 1
7	–	= 4	= 1
8	–	> 4	= 1
9	–	= 2	> 1
10	–	= 3	> 1
11	–	= 4	> 1
12	–	>4	> 1
		4l region	
1	✓	–	–
2	Veto	= 0	–
3	Veto	> 0	=1
4	Veto	> 0	>1

TABLE 6.8: Categories defined in the control regions. A hyphen (–) is written when no cut is applied.

6.5 Signal and background estimation

In this section, the estimation of the various backgrounds and signal is described.

6.5.1 Reducible backgrounds

Reducible backgrounds are constituted by different effects: non-prompt leptons and misidentified τ_h, lepton charge mismeasurement and photon conversions. For the first two, data-driven approaches are followed, while simulations are used for photon conversions.

6.5.1.1 Non-prompt leptons

Thanks to the high purity of the lepton selection in this analysis, the probability for a non-prompt lepton or a jet to be identified as a prompt lepton is very small. However, since the cross section of many backgrounds, such as $\mathrm{t\bar{t}}$ events, is very large compared to that of signal, a significant contribution from processes containing non-prompt leptons may be present in the selection. This contribution is dominated in the most sensitive regions by $\mathrm{t\bar{t}}$ events. For instance, semileptonic $\mathrm{t\bar{t}}$ events may enter in the $2l ss + 0\tau_h$ category if a lepton with the same sign to the prompt lepton is produced in one of the b jets. Dileptonic $\mathrm{t\bar{t}}$ may also contribute to the $3l + 0\tau_h$ region if a non-prompt lepton is produced.

To estimate this background a data-driven approach is followed, taking a sideband from each signal region. This region is referred to as Application Region (AR), and its definition is identical to that of the signal regions, but from the fact that the lepton identification criteria are loosened to require leptons only to pass the fakeable criteria. The AR is enriched in events with non-prompt leptons, that can be used to estimate the contribution of this process in the signal region with a suitable transfer factor, denoted as fake-rate. The contribution of processes with prompt leptons in the AR is subtracted before the transfer factor is applied.

This approach has the main advantage with respect to simulation-based approaches that no assumptions are needed on the wellness of the modeling of the event variables. For instance, signal regions in this analysis require a higher jet multiplicity to what is expected in $\mathrm{t\bar{t}}$ events. The generation of these additional jets is usually difficult to model and sensitive to several modeling effects, including the tune of the underlying event parameters [123].

Additionally, since the transfer factor between the SR and the AR is determined on data, the approach does not make strong assumptions on the modeling of the rate of non-prompt fakeable leptons passing the tight identification. However, since this method is also validated using simulations, and the fact that similar approaches could be followed to calibrate the fake-rate of simulated events, this latter aspect does not represent a significant gain with respect to simulation-based approaches.

The method has two steps: the measurement of the fake-rate in a region enriched in non-prompt leptons, the Measurement Region (MR), and the application of the fake-rate to events in the AR to obtain an estimate of the non-prompt background in the SR. The whole procedure is then validated with a closure test in simulated events, that is used to asses any potential bias due to variations of the fake-rate between the MR and the AR.

Fake-rate measurement

The fake-rate is measured as a function of the cone-p_{T} and in two $|\eta|$ regions, separately for electrons and muons. For that purpose, a sample of multijet data events enriched in non-prompt leptons is used. This sample is constructed by requiring the presence of exactly one electron or muon passing the fakeable selection criteria. Additionally, at least one recoiling jet is required in the event, which is required to be separated by the lepton with $\Delta R > 0.7$.

While this sample is dominated by multijet events, non-prompt contribution to the signal regions comes mostly from $\mathrm{t\bar{t}}$ events. A different flavor composition and momentum spectrum of the partons yielding to the production of non-prompt leptons is therefore expected. To account for this, the definition of the fakeable selection criteria was carefully tuned to have a similar flavor composition between the two samples. The difference in p_{T} distribution of the partons between samples is handled by parameterizing the fake-rate as a function of the cone-p_{T}, instead of the reconstructed p_{T}.

Events are collected using a set of single lepton triggers with different p_{T} thresholds and no isolation criteria. These triggers have been designed specifically for the purpose of measuring this fake-rate and are prescaled to keep their acceptance rate under control. Electron triggers require the presence of an additional jet. This criterion is only applied in muon triggers for low p_{T} thresholds.

Leptons are required to have a p_{T} above a certain threshold in order for them to be above the turn-on of the trigger. Additionally, these events are only used to measure the fake-rate above a certain cone-p_{T} value. Otherwise, since cone-p_{T} is a correction based on the isolation of the lepton, selecting leptons with cone-p_{T} close to the p_{T} threshold would contaminate the selection in leptons that are isolated, introducing a bias in the measurement. Thanks to the isolation cut of less than 0.4 p_{T}, selecting leptons with cone-p_{T} above twice the p_{T} threshold effectively removes this bias. An additional bias could be introduced by the trigger selection, since the electron triggers impose quality cuts on the electron. This is avoided by applying the trigger emulation cuts described in section 6.3.1.

Passing and failing categories are defined depending on whether the lepton passes the tight selection or not. Then, the fake-rate f can be computed from the number of non-prompt leptons in each of the categories, N_{pass} and N_{fail}:

$$f = \frac{N_{\mathrm{pass}}}{N_{\mathrm{pass}} + N_{\mathrm{fail}}}. \tag{6.4}$$

Contribution from other processes, such as W +jets events, DY+jets, or $\mathrm{t\bar{t}}$ events may contribute to this region with prompt leptons, particularly in the numerator of the measurement. For the measurement of the fake-rate this contribution must be subtracted. To discriminate between non-prompt and prompt the $\mathrm{M_T}^{\mathrm{fix}}$ variable is defined as

$$\mathrm{M_T}^{\mathrm{fix}} = \sqrt{2 p_{\mathrm{T}}^{\mathrm{fix}} p_{\mathrm{T}}^{\mathrm{miss}} (1 - \cos\Delta\phi)}, \tag{6.5}$$

where $\Delta\phi$ is the azimuthal angle between the lepton momentum and the $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$, and $p_{\mathrm{T}}^{\mathrm{fix}}$ is a constant value set to 35 GeV. This is a variation of the $\mathrm{M_T}$ variable, in which the p_{T} of the lepton is set to 35 GeV to reduce its correlation with the p_{T}, since the fake-rate may have a p_{T} dependence.

The contribution from non-prompt leptons in the two categories is determined by performing a fit to the $\mathrm{M_T}^{\mathrm{fix}}$ shape simultaneously in the passing and failing categories. The fitting distributions are templates obtained from multijet simulations and simulations from the processes leading to prompt-leptons. The normalizations of the templates are free parameters of the fit. Additionally, nuisance parameters are added to the fit that allow to modify the shape of these templates, considering a linear deformation and a stretching of the template. Statistical uncertainties of the templates are also nuisance parameters of the fit.

Alternative fits are considered to estimate the systematic uncertainties associated to this subtraction of prompt processes. One source of uncertainty is obtained by fitting the overall normalization of prompt leptons processes, and applying it to all bins. Another alternative considered is to divide each bin of the measurement in two regions with high and low values of $\mathrm{M_T}^{\mathrm{fix}}$. This gives two independent measurements of the fake-rate with no subtraction of the prompt lepton component. Taking the ratio of events with prompt leptons with high and low values of $\mathrm{M_T}^{\mathrm{fix}}$ from simulations, the fake-rate measurements in the two categories can be unfolded to obtain an unbiased estimate of the fake-rate.

Fake-rate application

Once the fake-rates have been obtained, a transfer factor is built to extrapolate from the AR to the SR. The expression for the transfer factor depends on the lepton multiplicity required in the signal region definition. In a region with two leptons, the expected number of non-prompt events with two leptons passing the tight criteria, N_{pp}, can be written as

$$N_{\mathrm{pp}} = \sum_{\mathrm{pf}} \frac{f_1}{1-f_1} + \sum_{\mathrm{fp}} \frac{f_2}{1-f_2} - \sum_{\mathrm{ff}} \frac{f_1}{1-f_1}\frac{f_2}{1-f_2}, \tag{6.6}$$

where f_i is the fake-rate evaluated in the i-th lepton, sorted by cone-p_{T}. The sums $\sum_{\mathrm{pf,fp,ff}}$ run over the events in the application region having a specific combination of leptons passing and failing the tight selection. For instance, $\sum_{\mathrm{fp}}$ runs over all the events in the application region for which the leading lepton has failed the tight criteria and the subleading has passed them.

For regions with three leptons, the estimation in the signal region can be written as

$$
\begin{aligned}
N_{\text{ppp}} = & \sum_{\text{fpp}} \frac{f_1}{1-f_1} + \sum_{\text{pfp}} \frac{f_2}{1-f_2} + \sum_{\text{ppf}} \frac{f_3}{1-f_3} \\
& - \sum_{\text{ffp}} \frac{f_1}{1-f_1}\frac{f_2}{1-f_2} - \sum_{\text{fpf}} \frac{f_1}{1-f_1}\frac{f_3}{1-f_3} - \sum_{\text{pff}} \frac{f_2}{1-f_2}\frac{f_3}{1-f_3} \\
& + \sum_{\text{fff}} \frac{f_1}{1-f_1}\frac{f_2}{1-f_2}\frac{f_3}{1-f_3}.
\end{aligned}
\tag{6.7}
$$

A similar formula can be written for signal regions with four leptons. The contribution from prompt processes to the AR is estimated using simulations, as it is done in the SR, and it is subtracted before applying the transfer factor to data.

Closure test

In order to check the validity of the method, a closure test is performed in simulated events passing the signal region requirements. This method ensures that no bias is being introduced due to the different flavor composition of the multijet data sample and the signal region, for which the non-prompt contribution is enriched in $\mathrm{t\bar{t}}$ events. It also takes into account uncovered kinematic dependencies of the fake rate as a function of the signal extraction variables.

The closure test is performed separately for electrons and muons in events in the $2lss + 0\tau_{\text{h}}$ region, which is the one expected to have a most significant contribution of $\mathrm{t\bar{t}}$ events producing non-prompt leptons. The event yield predicted by simulations in the signal region is compared to the estimation of the method applied to $\mathrm{t\bar{t}}$ simulated events in the AR. The latter is done with fake-rates computed in $\mathrm{t\bar{t}}$ and multijet simulated samples separately. The comparison between these two fake-rates allows to evaluate the dependency of the fake-rate between the two samples, while the comparison between the $\mathrm{t\bar{t}}$ yield in the signal region and the estimation performed using the fake-rate obtained with a multijet sample is an estimation of the total bias introduced by the method.

The closure is shown in figure 6.6 for muons and electrons separately. A very good closure is observed for muons, while a 30-40% non-closure is observed for electrons. This non-closure is due to the different flavor composition between the multijet and the $\mathrm{t\bar{t}}$ samples, and is used to correct the fake-rate. It is also taken into account as a source of systematic uncertainty.

6.5.1.2 τ_{h} misidentification rate

The estimation from misreconstructed gluon and quark jets as τ_{h} is estimated using a similar method to non-prompt leptons, with the only difference that the fake-rate is measured in $\mathrm{t\bar{t}}$ event enriched data sample. The selection for this is constructed by requiring an opposite-sign $\mathrm{e\mu}$ pair with $p_{\text{T}} > 25$ and 15 GeV, for the leading and subleading lepton. At least two jets are required, and the event must pass the usual b tag multiplicity requirement. Events with $m_{\mathrm{e\mu}} < 12$ GeV are rejected to suppress

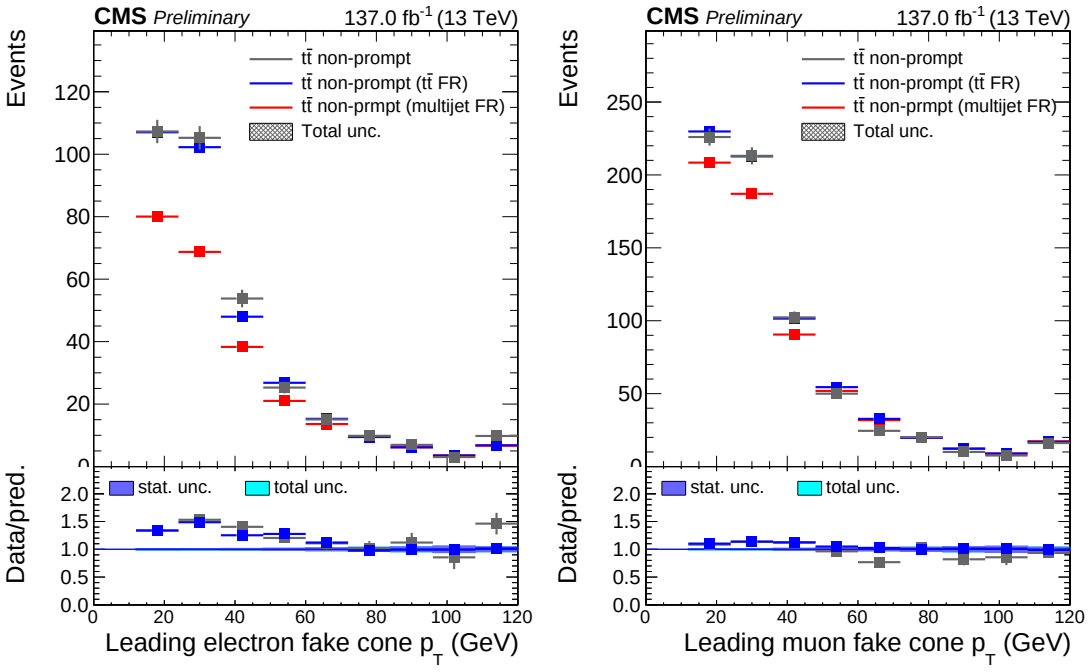


FIGURE 6.6: Closure test of the fake rate method for electrons (left) and muons (right) in $t\bar{t}$ simulations as a function of the lepton cone-p_T. Gray points shows the yield predicted in the signal region, and is compared to the yield simulations in the AR corrected by a fake-rate derived in $t\bar{t}$ (blue) and multijet (red) simulations. Only statistical uncertainties are shown.

contributions from low mass resonances. Finally, events are required to have a τ_h passing the fakeable selection. The efficiency of this fakeable τ_h to pass the different tight selection criteria is evaluated after subtracting the contribution from genuine τ_h as estimated from simulations.

This fake-rate is applied using the same formulae as it is used for non-prompt leptons, and a closure test is also applied to estimate the systematic uncertainty of the method.

6.5.1.3 Charge flips

The charge sign of electrons and muons may be measured incorrectly, and background events may enter in the $2lss + 0\tau_h$ and $2lss + 1\tau_h$ regions. This rate is negligible for muons, once the requirement on the p_T measurement uncertainty is applied. However, electrons may radiate significantly and the sign of the curvature may be measured incorrectly with a higher probability.

This mismeasurement rate is determined by measuring the number of reconstructed same-sign and opposite-sign $Z \to e^{\pm}e^{\mp}$ events in data. The number of $Z \to e^{\pm}e^{\mp}$ events is obtained by performing a fit to the Z peak. The shapes for the m_{ll} distribution are templates obtained from simulated events.

The measurement is performed in bins of p_T and η, with a total of 6 categories. The number of same-sign events can be calculated as $N_{SS} = (q_1 + q_2)N_{OS}$, where q_i is the charge flip rate of the i-th lepton and $q \ll 1$ has been assumed. 21 such equations are

written, corresponding to the possible categories the two leptons can be in, since no ordering of the leptons is taken. These 21 equations, together with their 6 unknowns, that correspond to the flip-rate for each region, constitute an overconstrained problem, that can be solved using regularization techniques.

6.5.1.4 Conversions

A contribution from processes in which an electron is produced in the conversion of a photon. In this analysis, this contribution is dominated by $\mathrm{t\bar{t}\gamma}$ events, and are estimated using simulations.

6.5.2 Signal and irreducible backgrounds

Signals and irreducible backgrounds are estimated using simulations. The normalization of processes in which it is possible to have a control region enriched in them is left freely floating in the analysis. This way, no a-priori assumption is made on their cross section, only on its acceptance and in the shapes of the distributions, which are expected to be less affected by higher order corrections in perturbation series. However, we quote postfit scale factors with respect to the cross section calculations described below, so they can be compared with other predictions. In this analysis, the normalization of the $\mathrm{t\bar{t}Z}$, $\mathrm{t\bar{t}W}$, WZ and ZZ processes are determined in situ in the signal extraction fit.

$\mathrm{t\bar{t}H}$ signal events are simulated at NLO accuracy with MadGraph mc@nlo, and are normalized to the NLO calculation with electroweak corrections presented in [150]. For the BSM interpretation, this cross section is scaled proportionally to κ_t, as expected at LO precision in electroweak in perturbation theory.

tH events are simulated at LO accuracy using MadGraph. These samples contain information to reweigh events according to different BSM scenarios, with different values for κ_t and κ_W. This allows to take into account the effect of variations of these couplings in the event kinematics. The 4FS is employed for the t-channel production, as it is expected to describe better the additional b quark in the gluon splitting, that would be treated within the parton shower in the 5FS. For the tW-channel, the 5FS is used, since it is the only way to technically reweigh events for different κ_t and κ_W values. The cross section to normalize these samples is obtained separately for each choice of κ_t and κ_W at NLO accuracy in the 5FS for the *t*-channel [150] and at NLO accuracy in the 5FS, using the DR2 scheme to handle the interference with $\mathrm{t\bar{t}H}$ events [151].

$\mathrm{t\bar{t}W}$ events are simulated at NLO accuracy with MadGraph mc@nlo, and are normalized to the NLO calculations with electroweak corrections presented in [150]. Additionally, on top of the obtained distributions, the $\mathcal{O}(\alpha_s\alpha^3)$ corrections, section 6.1.3, are simulated with MadGraph mc@nlo and included in the $\mathrm{t\bar{t}W}$ production. This correction accounts for both the total $\mathrm{t\bar{t}W}$ production rate as well as the differences in shape. $\mathrm{t\bar{t}Z}$ events are simulated in the phase space with a Z mass higher than 1 GeV at NLO accuracy with MadGraph mc@nlo, and are normalized according to the NLO cross section with electroweak corrections shown in [150].

The cross sections used for these processes are summarized in table 6.9.

WZ and $qq \to ZZ$ events are simulated at NLO using POWHEG v2 [87]. They are normalized to NLO calculations, although their normalization is measured in data. $gg \to ZZ$ events are simulated using the MCFM 7.0 generator at LO and normalized to a NLO calculation [172]. tZq events are relevant in the categories with leptons and in the control region for $t\bar{t}Z$. They are simulated at NLO accuracy in the 4FS with the MADGRAPH MC@NLO package, and normalized to the cross section provided by that generator.

Contributions from rarer processes like $t\bar{t}t\bar{t}$, tWZ, ttWH or triboson processes are also taken into account with dedicated simulations.

Process	Cross section (fb)
$t\bar{t}H$	$507.1^{+29.4}_{-47.1} \pm 18.2$
tH	$74.3^{+2.9}_{-4.7} \pm 2.1$
tHW	$15.6^{+0.7}_{-0.9} \pm 0.4$
$t\bar{t}Z$	$839.3^{+94.0}_{-80.5} \pm 28.5$
$t\bar{t}W$	$600.8^{+77.5}_{-69.1} \pm 24.0$

TABLE 6.9: List of cross section used to normalize the simulated samples used to estimate contributions from the main irreducible backgrounds and signal events, together with their corresponding uncertainties. The first uncertainty corresponds to the variation of the factorization scales, while the second corresponds to the propagation of the PDF uncertainties.

6.6 Signal extraction

As mentioned in the previous sections, two complementary approaches are followed to extract the signals in this measurement. The main analysis uses events in all the categories and extracts the signal by fitting the event yields in several subcategories. These subcategories are built using multivariate discriminators that are constructed using kinematic variables. For the control analysis, only events in the most sensitive categories are used. In the categories without τ_h, simple kinematic distributions are fitted, while in the $2lss + 1\tau_h$, the score of a matrix element discriminant is used.

The first approach provides higher sensitivity to the process, while the second allows to establish a baseline analysis, and for an easier interpretation of the results.

6.6.1 Main analysis

Two different approaches are followed in the main analysis. The $2lss + 0\tau_h$, $3l + 0\tau_h$ and $2lss + 1\tau_h$ are the most sensitive categories in the analysis, as they are quite pure in signal and a reasonably large number of expected signal events. Additionally, they are constructed to accept a significant number of tH events. For these regions, the classification is made using multiclass ANN. These discriminators produce several output

variables, that can be interpreted as the probability of an event to have been produced by a given process. This allows to built selections that are pure in the $\mathrm{t\bar{t}H}$ and tH signals separately, but also on the main backgrounds. This is done to obtain a higher signal sensitivity, and also to constrain these backgrounds from data. Additionally, the large number of expected events in the $2l_{ss} + 0\tau_h$ and $3l + 0\tau_h$ categories allow to further classify based in other variables, such as lepton flavor or b tag jet multiplicity to enhance the sensitivity.

Some of the remaining signal regions, $0l + 2\tau_h$, $1l + 1\tau_h$, $1l + 2\tau_h$ and $2l_{os} + 1\tau_h$, have a significant contribution from non-prompt leptons or misidentified τ_h, one order of magnitude higher than that of signal. The remaining ones, $2l + 2\tau_h$, $3l + 1\tau_h$ and $4l + 0\tau_h$, correspond to final state with four leptons, receiving a negligible contribution from tH and an overall small event yield. In these cases, little sensitivity gain is expected from the multiclass approach, so BDTs are used to construct categories enriched in signal.

The discriminators used in this chapter are inputs from the IHEP and NICPB groups, which I have used to perform the signal extraction. For this reason, only the input variables and the relevant details of their training are covered.

Some of these discriminators use others as input variables: the hadronic top tagger and the Higgs jet tagger, that allow to improve the sensitivity.

Hadronic top tagger It allows to find triplets of jets that correspond to the hadronic decay of a top quark. It is a BDT trained on $\mathrm{t\bar{t}}$ simulated events using jet triplets coming from the top quark decay as signal and other combinations, i.e. jet triplets in that have not been produced in the same top quark decay, as background. The discriminator uses 16 input variables, including the score of the b-tagging discriminators and the quark-gluon likelihood of the jets, angular variables distances and masses of several combinations of jets. All the possible jet triplets of an event are considered, and the triplets with the highest and second highest score are kept.

Higgs jet tagger It allows to identify, in the $2l_{ss} + 0\tau_h$ region, the jet that has been produced in the Higgs boson decay. It uses 5 input variables including the kinematic properties of the jet, its b tagging score and quark gluon likelihood, as well as the angular distances between the jet and the leptons. All jets in the event, but the triplet identified by the hadronic top tagger are evaluated, and the one with the highest score is kept.

Signal regions with ANNs

The ANNs are trained with samples of simulated events, different from those used in the signal extraction. As input variables, a combination of low level variables, namely the three-momenta of the objects, and high level variables (invariant masses and angular distances between the various objects) are used. This combined strategy allows to improve significantly the discrimination with respect to the BDT based approaches.

The ANN in the $2l_{ss} + 0\tau_h$ region features 36 input variables and returns four scores, corresponding to the probability of an event probability to be a $\mathrm{t\bar{t}H}$, tH, $\mathrm{t\bar{t}W}$ or $\mathrm{t\bar{t}}$

event. 41 and 37 variables are used in the $2\mathrm{l}ss + 1\tau_\mathrm{h}$ and $3\mathrm{l} + 0\tau_\mathrm{h}$ regions, respectively. In these two categories, the ANNs have three nodes, corresponding to $\mathrm{t\bar{t}H}$, tH and background processes. The list of input variables in these ANNs is shown in table 6.10.

Input variable	$2\mathrm{l}ss + 0\tau_\mathrm{h}$	$2\mathrm{l}ss + 1\tau_\mathrm{h}$	$3\mathrm{l} + 0\tau_\mathrm{h}$
Number of electrons in the category definition	✓	✓	✓
Cone-p_T of selected leptons	2	2	3
η of selected leptons	2	2	3
ϕ of selected leptons	2	2	3
Transverse mass of selected leptons	2	2	
Leading lepton charge	✓		
p_T of selected tau		1	
η of selected tau		1	
ϕ of selected tau		1	
Transverse mass of $\mathrm{l}(+\tau_\mathrm{h})$ system		2	
Invariant mass lepton+tau		2	
Charge sum of leptons+taus		✓	✓
Maximum lepton η	✓		
Minimum ΔR leptons$(+\tau_\mathrm{h})$		✓	✓
Minimum ΔR leptons+jets	2	2	3
Minimum ΔR tau+jets		1	
Has OSSF lepton pair?			✓
Number of jets	✓	✓	✓
p_T of the jets	4	3	3
η of the jets	4	3	3
ϕ of the jets	4	3	3
Jets average ΔR	✓	✓	✓
Number of forward jets	✓	✓	✓
p_T of leading forward jet	✓	✓	✓
$\lvert\eta\rvert$ of leading forward jet	✓	✓	✓
Minimum $\Delta\eta$ bt. leading forward jet and central jets		✓	✓
Number of loose b-jets	✓	✓	✓
Number of medium b-jets	✓	✓	✓
Invariant mass of loose b-jets		✓	✓
Invariant mass of medium b-jets	✓		
$p_\mathrm{T}^\mathrm{miss}$ LD	✓	✓	✓
Hadronic top tagger	✓	✓	✓
Hadronic top p_T	✓	✓	✓
Higgs jet tagger	✓		
Number of variables	36	41	37

TABLE 6.10: List of variables used for the multiclass ANN in the $2\mathrm{l}ss + 1\tau_\mathrm{h}$, $2\mathrm{l}ss + 0\tau_\mathrm{h}$ and $3\mathrm{l} + 1\tau_\mathrm{h}$ regions. A check mark indicates that that variable is used in that category, while the number corresponds to the number of variables used as an input, corresponding to that entry.

Events in each category are divided in subcategories based on the output scores of the ANN. One subcategory per each output node is built with events for which the score of that particular node is higher than that of the other nodes. This defines four categories in the $2l ss + 0\tau_h$ region and three in the other two. Events in the $2l ss + 0\tau_h$ region are further categorized depending on the flavor of the selected leptons in the ee, $e\mu$ and $\mu\mu$ channels. Events in the $3l + 0\tau_h$ region are also classified according to the presence or not of two medium b-tagged jets, and according to the lepton flavor in the eee, $ee\mu$ $e\mu\mu$ and $\mu\mu\mu$. Not all the subcategorization is performed to all the nodes. Finally, each subcategory is divided bins according to the maximum ANN output score to maximize the sensitivity.

Events in the $3l$ and $4l$ control regions are also added in the fit in both the main and the control analysis, to determine the normalization of the $t\bar{t}Z$ and diboson processes in situ.

Signal regions with BDT

The remaining signal regions are either dominated by one background source or have too small event yield to profit from the multiclass approach. For these regions BDT are used, and events are classified depending on the score of these BDTs in several bins.

The input variables used in these BDTs are summarized in table 6.11.

6.6.2 Control analysis

The control analysis is designed to be a robust cross-check of the analysis, that serves to consolidate the measurement. Only the most sensitive categories to the $t\bar{t}H$ signal are used: the $2l ss + 0\tau_h$ $t\bar{t}H$ region, the $3l + 0\tau_h$ $t\bar{t}H$ region, the $2l ss + 1\tau_h$ $t\bar{t}H$ region and the $4l + 0\tau_h$ region. Then, events are categorized as follows.

$2l ss + 0\tau_h$ Events in this region are categorized in three classes, depending on the jet multiplicity. Three subcategories are built depending on the flavor of the two leading leptons, and two additional subcategories are built depending on the charge of the leptons. In total 18 categories are built. Then, events are further classified in 9 bins of m_{ll}.

$3l + 0\tau_h$ Events in this region are categorized in two classes, depending on the jet multiplicity. Two subcategories are built, depending on the sign of the sum of lepton charges. Then, events are further classified in 5 bins of m_{3l}.

$2l ss + 1\tau_h$ Events in this category are classified based on the Matrix Element Method (MEM). Two categories are considered: events with three jets, in which one of the jets for signal has been lost, and events with at least four jets. Several hypotheses are considered, corresponding to the decay chains expected for events coming from the

Input variable	$0l+2\tau_h$	$1l+1\tau_h$	$1l+2\tau_h$	$2los+1\tau_h$	$2l+2\tau_h$	$3l+1\tau_h$	$4l+0\tau_h$
Cone-p_T of leptons		1	1	2		3	4
p_T of τ_h('s)	2	1	2	1	2	1	
Transverse mass of leptons		1	1	2			
Transverse mass of τ_h('s)	2	1					
Transverse mass of lepton+τ_h's system			✓				
Visible mass of leading τ_h/lepton and τ_h pair	✓	✓	✓	✓	✓	✓	
SVFit mass [173] of leading τ_h/lepton and τ_h pair	✓	✓					
Charge sum of lepton and τ_h		✓					
Maximum η of leptons / τ_h('s)	✓	✓	✓	✓			
ΔR between leptons				✓			
ΔR between τ_h's	✓		✓				
ΔR between leptons and τ_h		✓					
ΔR between OS leptons and τ_h			✓	✓			
ΔR between SS leptons and τ_h			✓	✓			
Minimum ΔR between leptons and τ_h('s)	✓			✓			
Minimum ΔR between leptons and jets		✓	✓	✓	✓		
Minimum ΔR between τ_h('s) and jets	✓			✓			
cos*θ of τ_h/leptons and τ_h pair	✓	✓	✓		✓		
Number of jets		✓					
Average ΔR between jets	✓	✓	✓	✓	✓		
Number of loose b-jets		✓					
Number of medium b-jets	✓						
Invariant mass of loose b-jets	✓	✓	✓	✓	✓		
p_T^{miss} LD	✓	✓	✓	✓	✓	✓	✓
Has OSSF lepton pair?						✓	✓
Minimum invariant mass of loose lepton pairs						✓	✓
Hadronic top tagger	✓	✓	✓	✓			
Second Hadronic top tagger		✓					
Hadronic top p_T	✓		✓	✓			
Second Hadronic top p_T	✓						
Number of variables	18	18	17	18	9	9	7

TABLE 6.11: List of variables used for the regions in which a BDT is used. A check mark indicates that that variable is used in that category, while the number corresponds to the number of variables used as an input, corresponding to that entry

different processes in this region. These hypotheses are shown in table 6.12. Then, the MEM weight for an event with reconstructed-level observables, $\mathbf{y}$, can be computed as

$$w_\Omega(\mathbf{y}) = \frac{1}{\sigma_\Omega}\sum_p \int d\mathbf{x} dx_a dx_b \frac{f_i(x_a,Q) f_j(x_b,Q)}{x_a x_b s} \delta(x_a P_a + x_b P_b - \sum p_k) |\mathcal{M}_\Omega(\mathbf{x})|^2 W(\mathbf{y} \mid \mathbf{x}), \tag{6.8}$$

where σ_Ω is a normalization term fixed by requiring $\int d\mathbf{y} w_\Omega(\mathbf{y})$. $\sum_p$ runs over the possible partons in the initial state, and x_a, x_b, correspond to the Bjorken fractions. $f_i(x,Q)$ are the PDF associated to a parton with flavor i, Bjorken fraction x and a given scale, Q, of the process, and $\delta(x_a P_a + x_b P_b - \sum p_k)$ imposes the energy and momentum conservation. $\mathcal{M}_\Omega(\mathbf{x})$ is the matrix element for a given process hypothesis, and is computed at LO using MADGRAPH. $W(\mathbf{y} \mid \mathbf{x})$ is a transfer function from parton-level quantities $\mathbf{x}$ to detector-level quantities, determined from simulations.

$4l + 0\tau_h$ Events are categorized in three categories based on the invariant mass of the four leptons, m_{4l}.

Hypothesis	Decay modes
$t\bar{t}H$	$t \to blν$, $t \to bqq'$ $H \to ττ \to lντ_h$
$t\bar{t}Z$ irreducible	$t \to blν$, $t \to bqq'$ $Z \to ττ \to lντ_h$
$t\bar{t}Z$ reducible	$t \to blν$, $t \to bqq'$ $Z \to ll \to lτ_h$
$t\bar{t}$ non-prompt	$t \to blν$ $t \to bτν \to b[\to l]τ_hν$

TABLE 6.12: Background and signal hypotheses for which the MEM weights are calculated, together with the associated decay chains. In the $t\bar{t}Z$ reducible hypothesis, one of the leptons is misidentified as a τ_h, while in the $t\bar{t}$ non-prompt, one of the leptons is produced in a meson decay in the b jet.

6.6.3 Systematic uncertainties

Similarly to the other analyses presented in this thesis, systematic uncertainties are present due to imprecisely measured or simulated effects. These uncertainties may affect the observed number of events, and they are parametrized as nuisance parameters in the signal extraction fit.

Auxiliary measurements used to validate and correct the simulations, and imprecisions the data-driven estimation methods are referred to as experimental uncertainties. Modeling uncertainties raise from the limited accuracy of the theory predictions, mainly dominated by the missing higher-order corrections in the cross section calculations and the simulation models, and from the uncertainties on the PDF.

Experimental uncertainties

- The main categories of the analysis make use of leptons to characterize the topology of the signal. Because of that, since ad-hoc lepton identification criteria are used, their precise characterization is very important. The efficiency measurement for electrons and muons and the estimation of associated uncertainties is outlined in chapter 3, leading to an uncertainty amounting between 1% and 2% in the bulk of the distribution, depending on the lepton flavor and its p_T and η. Efficiencies for the loose selection and tight selection are measured separately. The uncertainty to the former is estimated by considering different shape hypotheses for the m_{ll} distribution in DY+jets and background, as the phase space extrapolation is expected to be negligible in this case. The uncertainty in the tight selection is expected to be more dominated by this effect, and the uncertainty for it is assigned with the $t\bar{t}$ measurement described in chapter 3. In total, four nuisance parameters are used to take into account these two effects separately in electrons and muons.

- Efficiency of the trigger selection is measured with the orthogonal trigger method using triggers collecting events with large momentum imbalance. These efficiencies are measured as a function of the lepton multiplicity for the corresponding set of triggers and as a function of the relevant kinematic variables. Additional dependencies of the trigger efficiency are taken into account as systematic uncertainties.

- τ_h efficiency and energy scale are measured in $Z \to \tau_h\tau_h$ events with an uncertainty of 5 and 1.2%, respectively [79].

- Several uncertainties are considered in the estimation of the non-prompt lepton background, which is performed with the data-driven estimation described in section 6.5.1.1. The measurement of the fake-rate is performed with three different methods, taking as the nominal value the center value of the two and the envelope as the uncertainty. Three nuisances are used, separately for muons: an overall variation of the fake-rate across all p_T and η bins, and two variations that affect the fake-rate p_T and η dependency, but keep the total number of estimated non-prompt events in the $2lss + 0\tau_h$ constant. An additional uncertainty for any non-closure observed in simulations, which is of the order of 5-10% for muons and 25-40% for electrons, depending on the data taking era.

- The uncertainty associated to the τ_h fake-rate is also propagated to the analysis.

- Uncertainties associated to the calibration of the jet energy scale and resolution are taken into account. Eleven variations of the jet energy scale are considered, taking into account several uncertainty sources associated to their measurement. Six separate components of the jet energy resolution uncertainties are used to account for different η regions in the detector and different p_T regimes. An additional uncertainty is used to account for any potential miscalibration of jets due to the Hadron Endcap issue, described in chapter 2. Finally, the uncertainty associated to the unclustered energy is propagated to the computation of p_T^{miss}.

- The observed transparency loss in the ECAL endcap yields to a prefiring of electron and jets trigger, leading to an overall trigger inefficiency, as shown in chapter 2. This inefficiency is taken into account by applying corrections determined on data to event simulations. The size of these corrections is varied by its corresponding uncertainty as an additional systematic uncertainty.

- b tagging efficiency and b tagging rate of light and c jets in simulations is also calibrated using scale factors derived in data. These scale factors also allow to correct the shape dependence of the Deep Jet discriminant. Several components are taken into account as separate uncertainty, taking into account the systematic and statistical uncertainty of these measurements. Additionally, the effect of the jet energy scale uncertainty is also propagated to the b tagging calibrations.

- The uncertainty in the luminosity amounts to 2.5% for data recorded in 2016 and 2018 and to 2.3% for data recorded in 2017. Correlations among the different measurements are taken into account in the signal extraction.

Modeling uncertainty

- The effect of the missing orders in perturbation theory in the cross section calculation and the simulation models is taken into account by considering independent and correlated uncertainties of the renormalization and factorization scales of the process. For each process, the effect on the total cross section and on the acceptance and shapes of the distributions are taken into account as separated uncertainties. The former is taken into account in the fit even if the normalization of the process is determined from the fit. Additionally, the uncertainty of the PDF is propagated to the calculation to the cross section of all the simulated processes.
- A mismodeling of the ZZ and WZ processes is observed at high jet multiplicity, so a 30% uncertainty for simulated events of these processes is assigned.
- Other low cross section processes are assigned a 50% uncertainty to account for the limited knowledge of their cross section and their extrapolation to the phase space of the analysis. Conversions are assigned a 30-50% uncertainty, based on the agreement observed in control regions by previous analyses [14].
- Uncertainty on the calculations of the Higgs boson branching fractions are also propagated to the analysis [150].
- The statistical uncertainty of the simulations and the application region of the fake-rate are also taken into account using the Barlow-Beeston method [58].

6.7 Results

In this section, observed number of events in each subcategory of the analysis is compared to the expectations from the signal and background models. The signal extraction fit is also described and its results are also shown.

6.7.1 Main analysis

Signal is extracted by performing a maximum likelihood to the observed yields in all the categories. Besides the systematic uncertainties, that are parametrized as nuisance parameters of the fit, the strength $\mu = \sigma/\sigma^{SM}$ for the $t\bar{t}H$, tH, $t\bar{t}W$, $t\bar{t}Z$, WZ and ZZ processes are free unconstrained parameters of the fit. The rate of the $t\bar{t}WW$ background is constrained to scale by the same factor as $t\bar{t}W$ production.

The number of observed events in the signal regions without τ_h are shown in figures 6.7 and 6.8 after the fit to all signal and control regions of the main analysis is performed. A good agreement is seen between the observation and the statistical model across all the signal regions.

Assuming the discriminating observables for the tH and $t\bar{t}H$ signals to conform to their SM expectation the production rate for the $t\bar{t}H$ signal is measured to be $\mu_{t\bar{t}H} = 0.92 \pm 0.19\,(\text{stat.})^{+0.17}_{-0.13}$ (syst.+lumi.) times the SM expectation, equivalent to a $t\bar{t}H$ production cross section of 464.9 ± 96.3 (stat.) ± 38.2 (syst.+lumi.) fb, and that

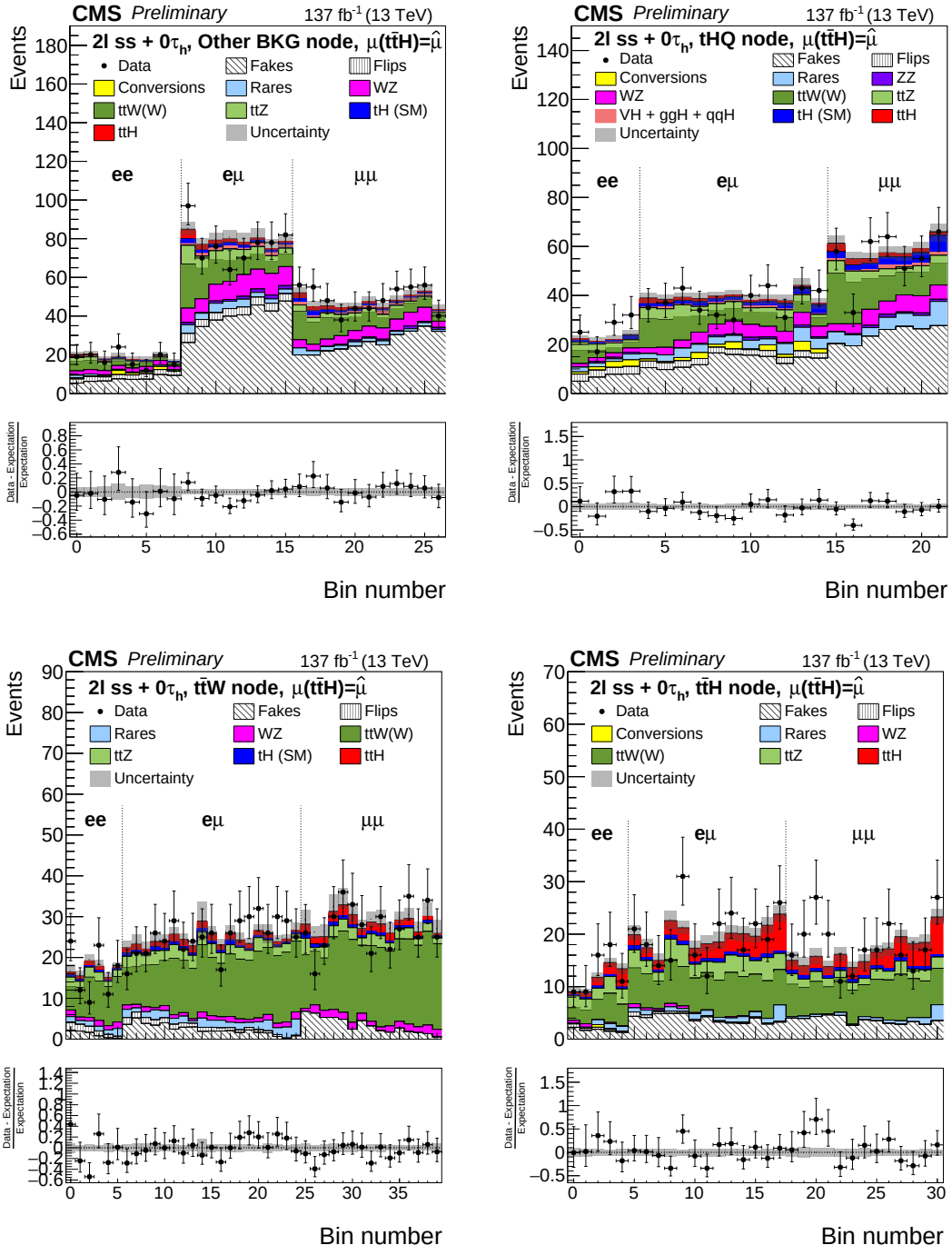


FIGURE 6.7: Observed event yield in the $2l ss + 0\tau_h$ region for events in the other category (top left), $t\bar{t}H$ category (top right), $t\bar{t}W$ category (bottom left) and $t\bar{t}H$ category (bottom right), compared to the results of the fit to the statistical model.

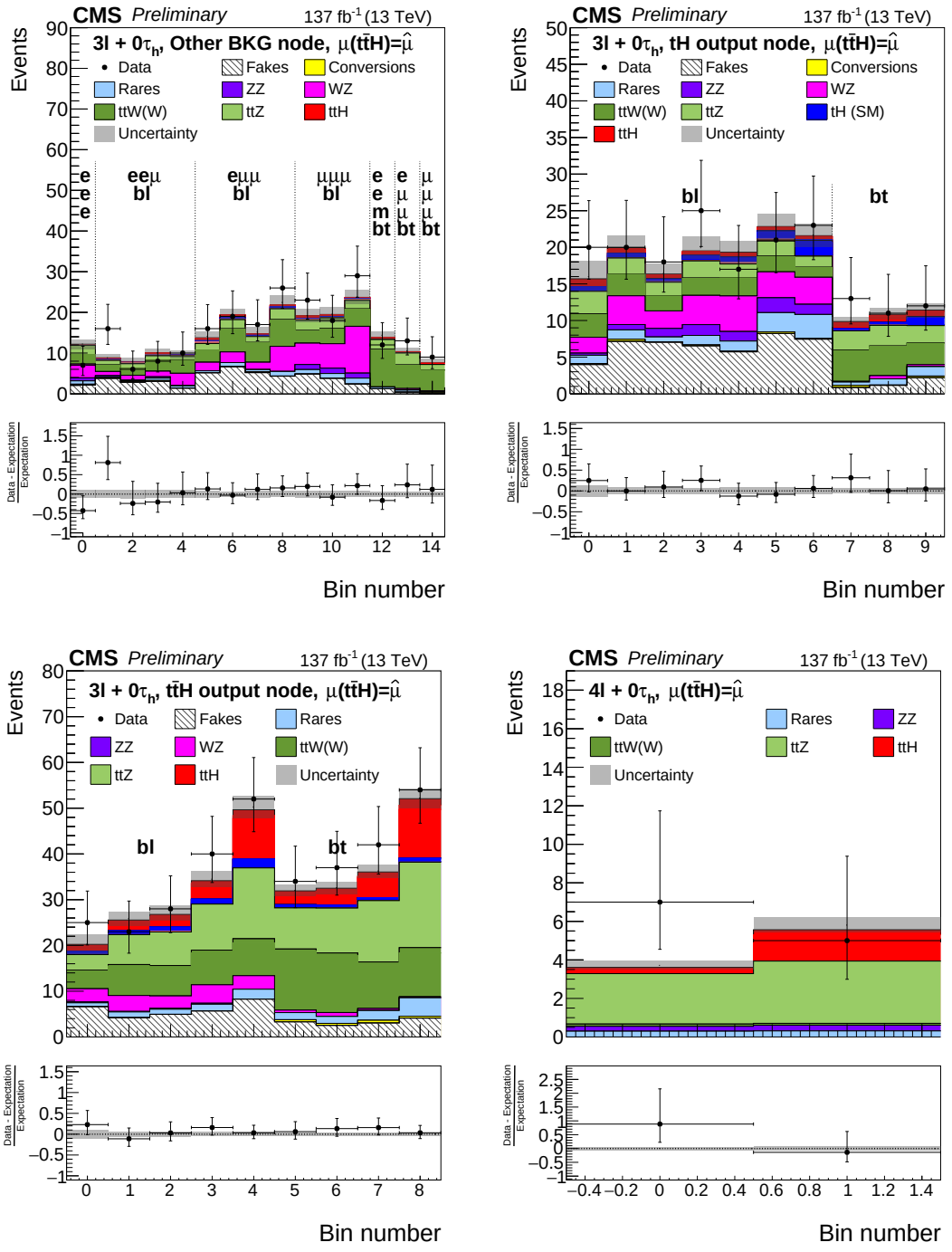


FIGURE 6.8: Observed event yield in the $3l + 0\tau_h$ region for events in the other category (top left), tH category (top right), and $t\bar{t}H$ category (bottom left), and for the $4l + 0\tau_h$ category (bottom right), compared to the results of the fit to the statistical model.

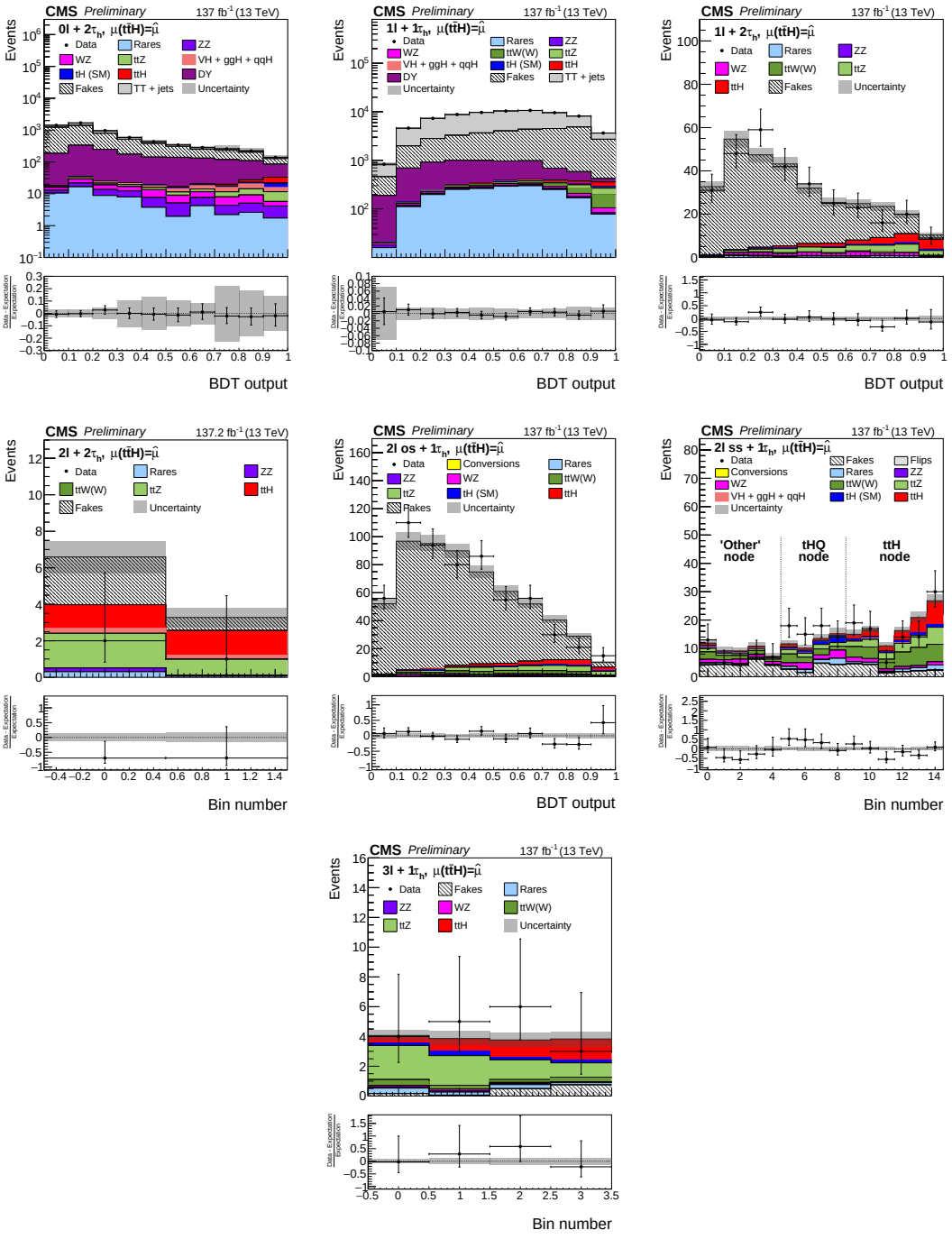


FIGURE 6.9: Observed event yields in the categories with τ_h in the final state. Top row shows the $0l + 2\tau_h$ (left), $1l + 1\tau_h$ (center) and $1l + 2\tau_h$ (right) categories, middle row shows the $2l + 2\tau_h$ (left), $2l os + 1\tau_h$ (middle) and $2l ss + 1\tau_h$ (right) categories, and the bottom row, the $3l + 1\tau_h$ category.

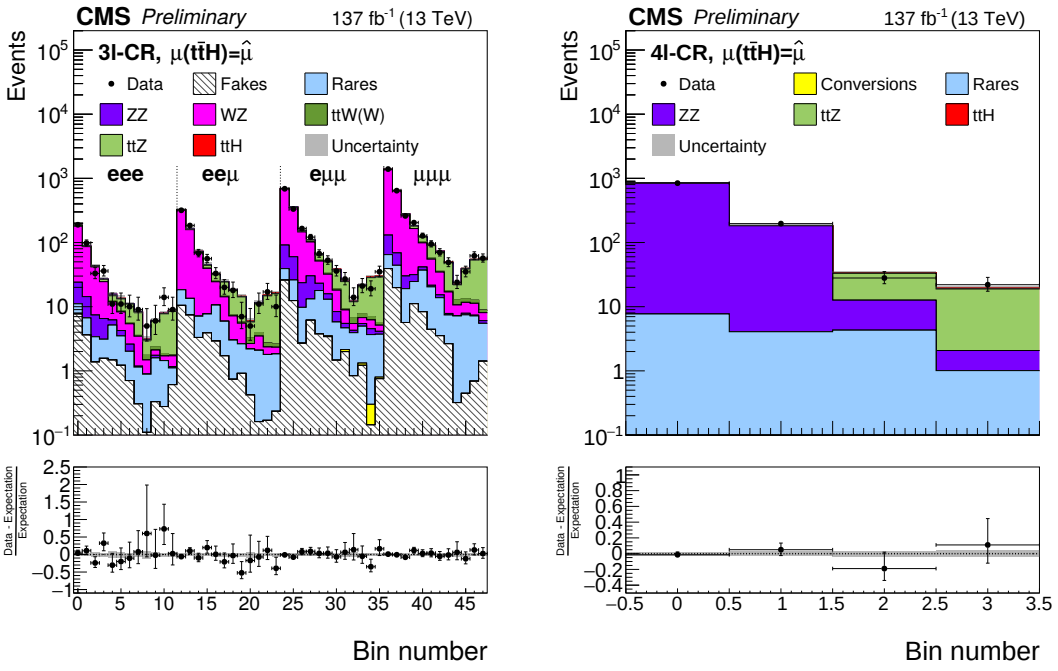


FIGURE 6.10: Observed event yield in the subcategories in the 3l (left) and 4l (right), after the fit .

of the tH signal is measured to be $\mu_{tH} = 5.8 \pm 2.7$ (stat.) ± 3.0 (syst.+lumi.) times the SM expectation for this production rate, equivalent to a cross section for tH production of 431 ± 200 (stat.) ± 222 (syst.+lumi.) fb. The production rate for the main unconstrained backgrounds is measured to be $\theta_{t\bar{t}Z} = 1.03 \pm 0.14$ (stat.+syst.+lumi.) and $\theta_{t\bar{t}W} = 1.43 \pm 0.21$ (stat.+syst.+lumi.) times their SM expectation. The rate of $t\bar{t}W$ is above the SM expectation, due to an excess of events in the $t\bar{t}W$ category of the $2lss + 0\tau_h$ region, which is quite pure in $t\bar{t}W$ events. This behavior shows a similar trend to other $t\bar{t}H$ and $t\bar{t}W$ analyses by both the CMS and ATLAS collaborations [161, 171], and could be at least partially explained by the missing higher order corrections in the reference $t\bar{t}W$ cross section, as described in section 6.1.3.

Assuming a tH production rate equivalent to that of the SM, these results correspond to an observed (expected) significance for the $t\bar{t}H$ signal amounting to 4.7 (5.2) standard deviations. Assuming a $t\bar{t}H$ production equivalent to that of the SM, the observed and expected significance for tH production is 1.4 and 0.3.

Figure 6.11 also shows the confidence regions for simultaneous measurements of the $t\bar{t}W$ and $t\bar{t}H$, $t\bar{t}Z$ and $t\bar{t}H$, and tH and $t\bar{t}H$ for 0.68 and 0.95 confidence levels. The rest of the parameters, including the other two parameters of interest not plotted, are profiled. They show the level of agreement with SM expectation and also the level of correlation of the measured parameters of interest.

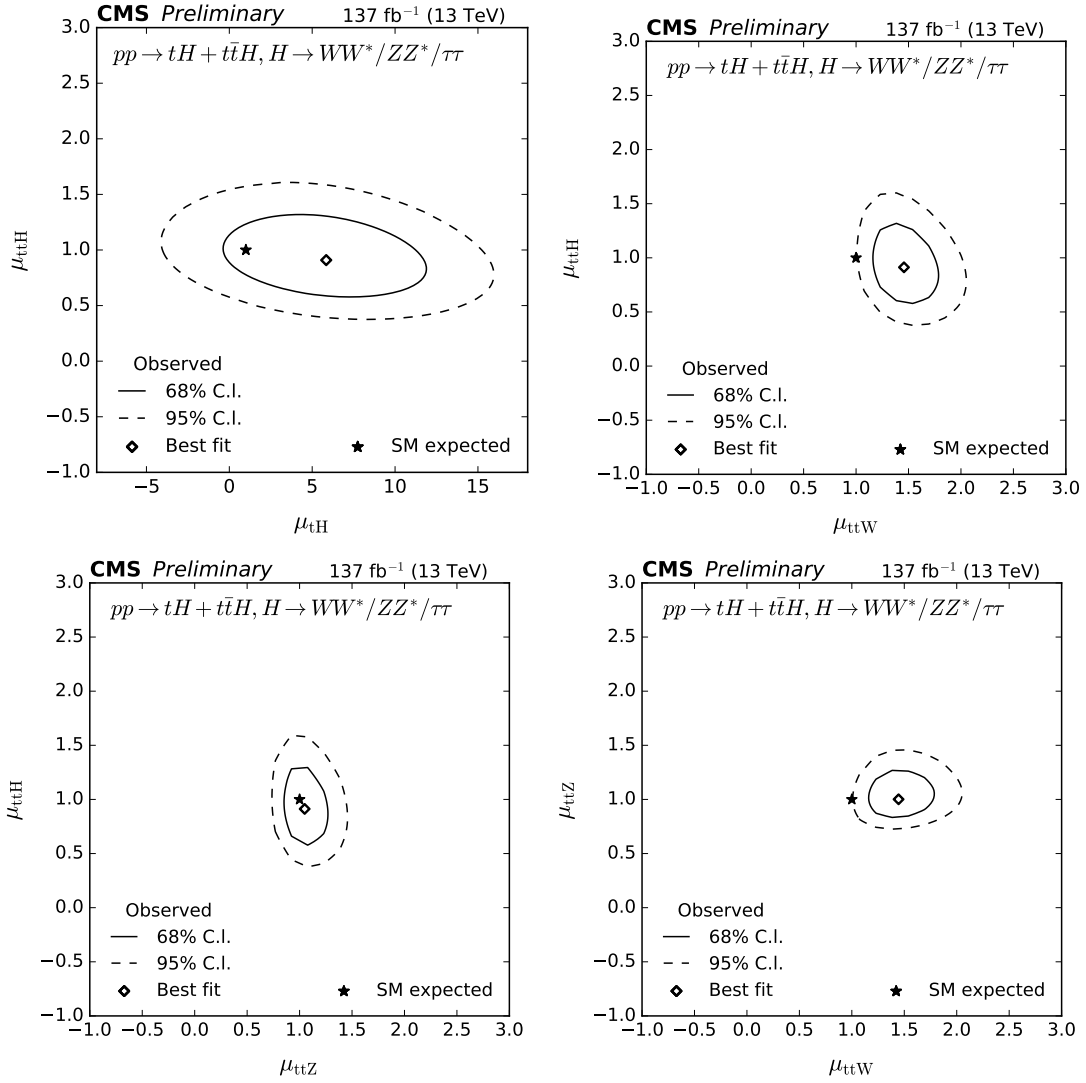


FIGURE 6.11: Confidence regions on the signal strength of pairs of $\mathrm{t\bar{t}H}$, tH, $\mathrm{t\bar{t}W}$ and $\mathrm{t\bar{t}Z}$ when scanning over two of them and profiling the others as obtained in the main analysis. The best fit point (square) and the SM prediction (star) are also plotted.

6.7.2 Control analysis

Signal extraction is performed in the control analysis using a similar likelihood fit, keeping the $\mathrm{t\bar{t}H}$, $\mathrm{t\bar{t}Z}$ and $\mathrm{t\bar{t}H}$ production rates as free parameters of the fit. tH is fixed to the SM prediction within its associated uncertainties. The observed events in each signal region, together with the best approximation of the statistical model are shown in figures 6.12-6.14. The production rate for the floating processes is measured to be $\mu_{\mathrm{t\bar{t}H}} = 0.5 \pm 0.3$ (stat.+syst.+lumi.), $\mu_{\mathrm{t\bar{t}H}} = 1.3 \pm 0.5$ (stat.+syst.+lumi.), $\mu_{\mathrm{t\bar{t}H}} = 0.8 \pm 0.4$ (stat.+syst.+lumi.), and $\mu_{\mathrm{t\bar{t}H}} = 1.5 \pm 1.5$ (stat.+syst.+lumi.) times the SM expectation, in the $2lss + 0\tau_{\mathrm{h}}$, $3l + 0\tau_{\mathrm{h}}$, $2lss + 1\tau_{\mathrm{h}}$, and $4l + 0\tau_{\mathrm{h}}$ channel, respectively. A signal strength of $\mu_{\mathrm{t\bar{t}H}} = 0.91 \pm 0.21$ (stat.) ± 0.18 (syst.+lumi.) is obtained for the simultaneous ML fit of all four channels. The corresponding observed (expected) significance of the $\mathrm{t\bar{t}H}$ signal in this analysis amounts to 3.8 (4.0) standard deviations.

The observed $t\bar{t}W$ and $t\bar{t}Z$ production rates amount to $\mu_{t\bar{t}W+t\bar{t}WW} = 1.081^{+0.207}_{-0.182}$ and $\mu_{t\bar{t}Z} = 0.894^{+0.148}_{-0.130}$ times the SM expectation.

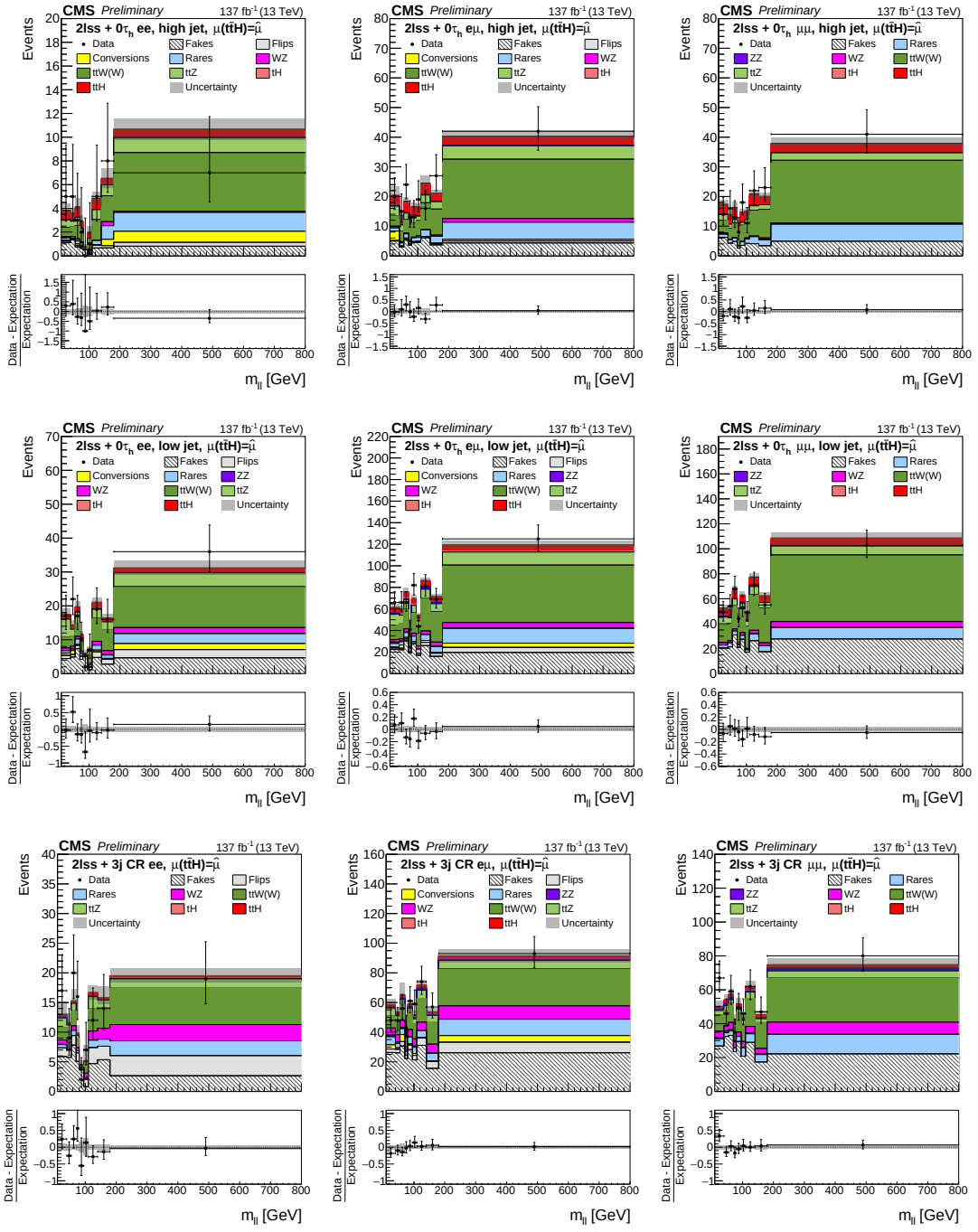


FIGURE 6.12: Observed number of events in the $2lss + 0\tau_h$ $t\bar{t}H$ category for the control analysis. Left column shows the ee channel, middle column shows the $e\mu$ channel, and the right column shows the $\mu\mu$ channel. Events are also classified according to the jet multiplicity. Events with at least 6 jets, between 3 and 5 jets or 3 jets are shown in the top, middle and bottom rows, respectively.

Figure 6.15 shows the confidence region corresponding to a 68% and 95% confidence level on the simultaneous extraction of the $t\bar{t}H$ and $t\bar{t}Z$, $t\bar{t}Z$ and $t\bar{t}W$, and $t\bar{t}H$ and

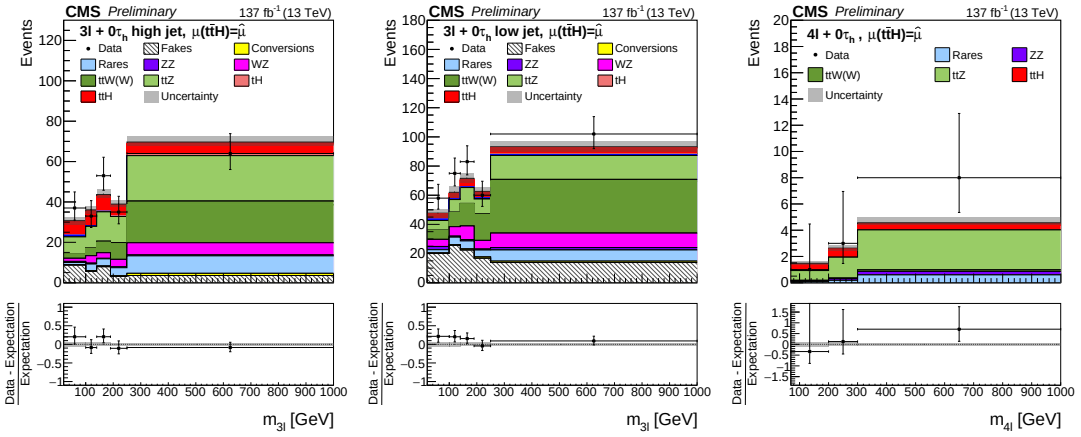


FIGURE 6.13: Observed number of events in the $2lss + 0\tau_h$ $t\bar{t}H$ category (left and middle) and the $4l + 0\tau_h$ (right) for the control analysis. Plot in the left shows the high jet multiplicity category, and the plot in the middle the low jet multiplicity category.

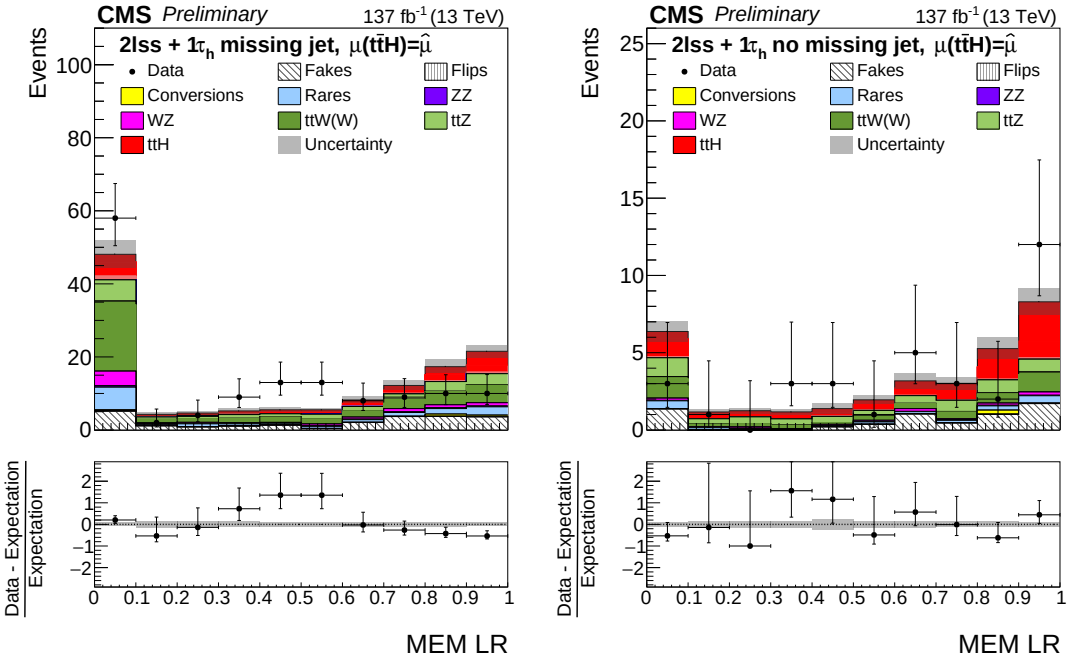


FIGURE 6.14: Observed number of events in the $2lss + 1\tau_h$ $t\bar{t}H$ category for the control analysis. Left shows event with a missing jet, and right events with at least four jets.

$\mathrm{t\bar{t}W}$ strengths.

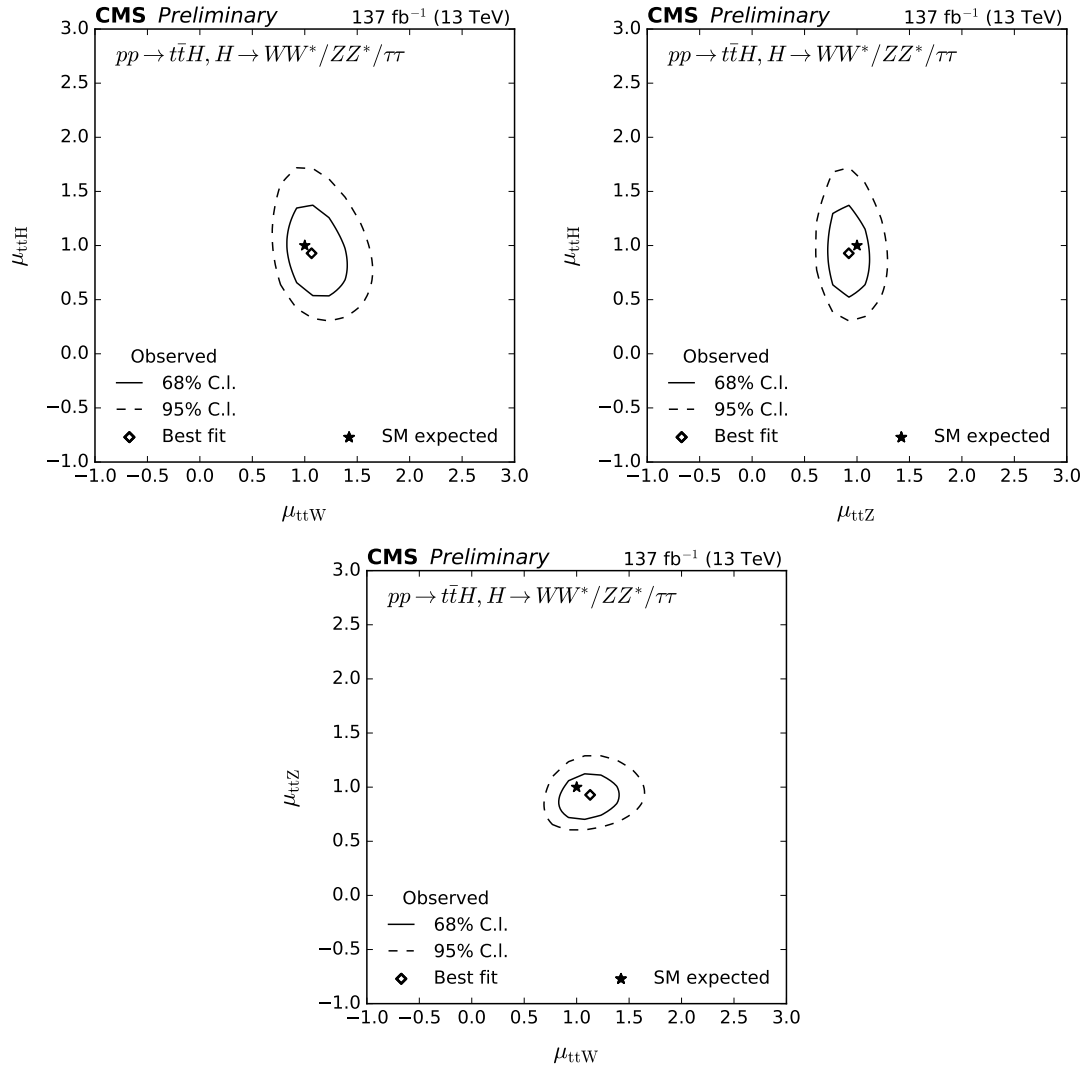


FIGURE 6.15: Confidence regions on the signal strength of pairs of $\mathrm{t\bar{t}H}$, $\mathrm{t\bar{t}W}$ and $\mathrm{t\bar{t}Z}$ when scanning over two of them and profiling the others as obtained in the control analysis. The best fit point (square) and the SM prediction (star) are also plotted.

6.8 Interpretation

The results of the main analysis are interpreted in terms of the coupling modifiers κ_{t} and κ_{W}. The former affects both the $\mathrm{t\bar{t}H}$ and tH production, while κ_{W} only affects the production of tH. The effect of κ_{t} and κ_{W} in the event kinematics of tH are already implemented in the samples, as described in section 6.5.2. No kinematic dependence with κ_{t} is present at leading order in $\mathrm{t\bar{t}H}$ events, so only the scaling of the cross section as κ_{t}^2 is considered. The effects of the variations of κ_{W} and κ_{t} in the branching ratios of the Higgs boson are also considered.

In order to perform this interpretation, the statistical model used for the signal extraction of the main analysis is used, fixing the signal strengths to the SM values. The

likelihood function is evaluated profiling all the nuisances and the strength modifiers for the irreducible backgrounds for several hypotheses of κ_t and κ_W. The point with the maximum likelihood is taken as the best fit for κ_t and κ_W.

Figure 6.16 shows $-2\Delta\log\mathcal{L}$ value, where $\mathcal{L}$ is the profiled likelihood, for several values of κ_t, assuming a κ_W equivalent to the standard model. Both the expected and observed scans are shown in the plot. Limits on κ_t can be set by finding the crossing of $-2\Delta\log\mathcal{L}$ with the quantiles of a χ^2_1 distribution. The current results constrain κ_t to be within either of the two intervals $-1.15 < \kappa_t < -0.8$ and $0.9 < \kappa_t < 1.25$ at 95% confidence level. Figure 6.16 also shows confidence regions at 68% and 95% confidence level, obtained by finding the crossings of $-2\Delta\log\mathcal{L}$ with the quantiles of a χ^2_2 distribution.

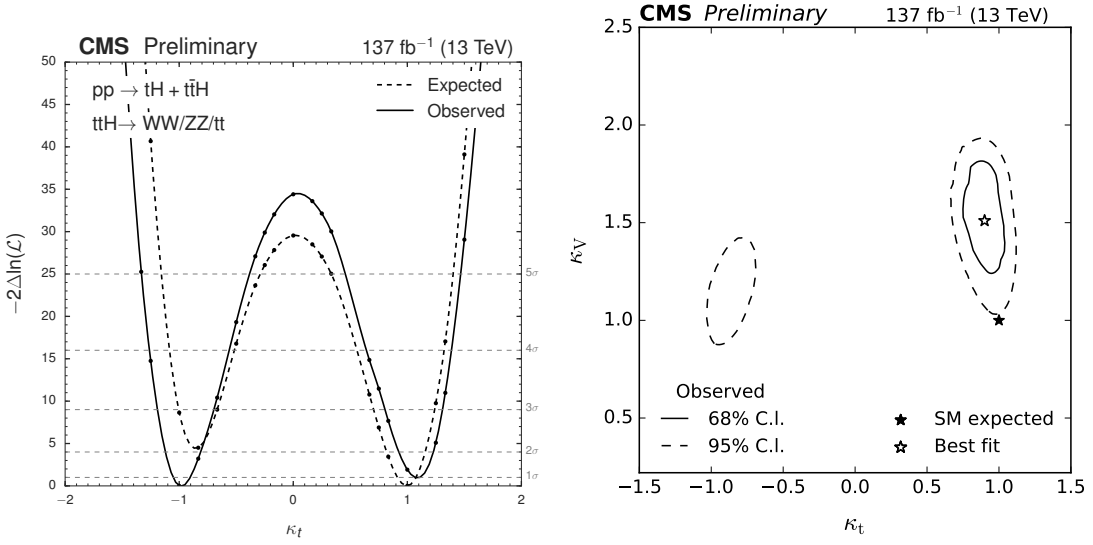


FIGURE 6.16: Interpretation of the $t\bar{t}H$ main analysis as a function in the context of the κ framework. Left: shows $-2\Delta\log\mathcal{L}$ for different hypotheses of κ_t, keeping κ_W to its SM values. Horizontal lines corresponding to 1, 2, 3, 4 and 5 standard deviations are also shown. Right: credible regions at 68 and 95% confidence level in the (κ_t, κ_W) plane. This plot does not include the $t\bar{t}W$ $\alpha_s\alpha^3$ electroweak corrections.

6.9 Conclusions

A measurement of $t\bar{t}H$ and tH production in the multilepton channel is performed. This measurement aims to be sensitive to the WW, ZZ and $\tau\tau$ decay modes of the Higgs boson, and both hadronic and leptonic decays of the top quark. Signal regions are built aiming for different combinations of these decays, and selections are applied according to the topology expected. Events are categorized among these signal regions, and are further categorized by the event kinematics to select regions enriched in signal and background events. Two approaches are followed to do so: one approach using MVA techniques, and another using simple kinematic variables and a matrix element discriminant.

The observed number of events in the regions defined in the first approach are shown in figure 6.17, together with the number of background and signal events predicted by the statistical model. This approach allows to measure $\mathrm{t\bar{t}H}$ with a production rate of $\mu_{\mathrm{t\bar{t}H}} = 0.92 \pm 0.19\ (\mathrm{stat.})^{+0.17}_{-0.13}$ (syst.+lumi.) times the SM expectation. The production rate for tH signal is measured to be $\mu_{\mathrm{tH}} = 5.8 \pm 2.7\ (\mathrm{stat.}) \pm 3.0$ (syst.+lumi.) times the SM expectation. These results correspond to an observed (expected) significance of 4.7 (5.2) standard deviations for $\mathrm{t\bar{t}H}$ and 1.4 (0.3) for tH production, in both cases fixing the other signal to the SM value.

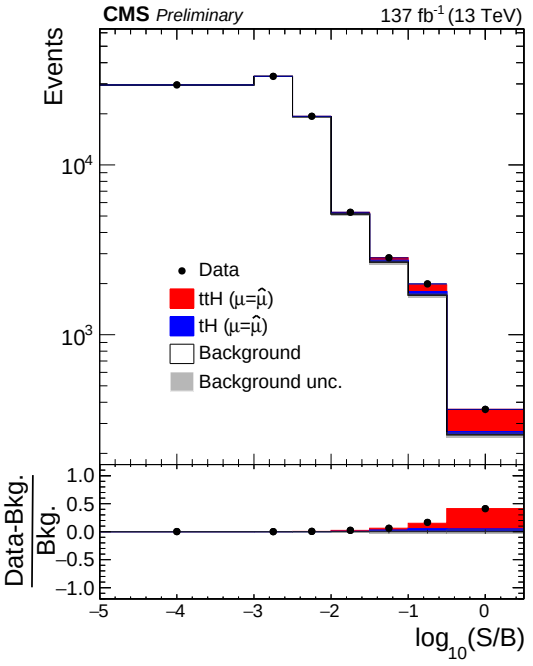


FIGURE 6.17: Number of observed events in the main analysis signal regions, compared to the number of background and signal events predicted by the fit. All the bins used in the signal extraction are grouped according to its signal-to-background ratio.

The second approach allows to measure the $\mathrm{t\bar{t}H}$ signal strength with a value of $\mu_{\mathrm{t\bar{t}H}} = 0.91 \pm 0.21\ (\mathrm{stat.}) \pm 0.18$ (syst.+lumi.). The observed (expected) significance to the $\mathrm{t\bar{t}H}$ signal to be different from zero in this approach amounts to 3.8 (4.0) standard deviations.

These results allow to claim the observation of $\mathrm{t\bar{t}H}$ production in the multilepton channel.

Additionally, results have been interpreted in the context of modifiers of the Higgs boson couplings to the top quark and W bosons. Assuming the coupling of the rest of the couplings to be that of the SM, κ_{t} is constrained to be in either of the two intervals $-1.15 < \kappa_{\mathrm{t}} < -0.8$ and $0.9 < \kappa_{\mathrm{t}} < 1.25$ at 95% confidence level.

Summary and conclusions

The unprecedented amount of collision data recorded by the CMS experiment during the Run 2 of the LHC has opened the way for the study of processes with lower cross-section, allowing for a more precise determination of SM quantities and to perform searches for BSM physics processes. In this thesis, both precise measurements of SM parameters and searches for BSM physics are presented in pp collisions recorded by the CMS detector during the Run 2 of the LHC. These studies have been done in events with two or more high p_{T} leptons in the final state, signature that is indicative of the production of heavy particles.

A measurement of the tW production cross-section in $\mathrm{e}^{\pm}\mu^{\mp}$ events has been presented. This measurement is the complement to measurements of $\mathrm{t\bar{t}}$ production, for which tW production is the main background. Events are categorized according to their jet and b-tagged jet multiplicity and MVA discriminators are used to obtain regions pure in signal. The cross-section is measured to be

$$\sigma_{\mathrm{tW}} = 63.1 \pm 1.8\ (\mathrm{stat}) \pm 6.3\ (\mathrm{syst}) \pm 2.1\ (\mathrm{lumi})\ \mathrm{pb},$$

consistent with the SM predictions of 71.7 $\pm$ 1.8 (scale) $\pm$ 3.4 (PDF) pb with approximate NNLO accuracy [105].

A differential measurement has also been performed in the same channel, in events with exactly one jet that is b-tagged. Despite of the large contribution of $\mathrm{t\bar{t}}$ events in the signal region, the differential cross-section for several variables of interest is measured consistently with the SM predictions with a precision between 20% and 100%, depending on the bin of the distribution. These measurements open the way for more precision measurements, as well as measurements of the interference between the $\mathrm{t\bar{t}}$ and tW processes.

Searches for the production of SUSY particles have also been performed in events with two opposite-sign same-flavor leptons, and missing transverse momentum. This signature is sensitive to different models of gluino, squark, electroweak spartners and slepton production. Different signal regions were defined for each one of the models, and main backgrounds are estimated using data-driven techniques. These measurements allow to exclude several ranges of sparticle masses under the assumption of simplified models, not excluded by previous searches.

Finally, a measurement of $\mathrm{t\bar{t}H}$ and tH production has also been performed. This measurement is performed in the multilepton channel, aiming for the WW, ZZ and $\tau\tau$ decay modes of the Higgs boson. Events are categorized based on the lepton and τ_{h} multiplicity, and further selection criteria in jets and b-tagged jet multiplicity are applied. Signals are then extracted in two complementary ways. In the main analysis, events are categorized based on the output of multivariate classifiers and other kinematic properties of the event. A control analysis is performed by means of a categorization based on the kinematic properties of the event and a matrix element discriminator. Special care is taken in the discrimination of the background due to non-prompt leptons, and the characterization of selected leptons. Similarly, state-of-the-art models are used to estimate the irreducible backgrounds.

These analysis provide a measurement of the $\mathrm{t\bar{t}H}$ signal strength of

$$\mu_{\mathrm{t\bar{t}H}} = 0.91 \pm 0.19\ (\text{stat.})^{+0.17}_{-0.13}\ (\text{syst.+lumi.})$$

The cross-section of this process is measured with an uncertainty around 20%. The analysis presented has an observed (expected) sensitivity of 4.7 (5.2) standard deviations in the main analysis, and of 3.8 (4.0) standard deviations in the control analysis. The observed (expected) significance of the main analysis to tH production is 1.4 (0.3).

The main analysis is also interpreted in the context of anomalous couplings of the Higgs boson to the top quark and the W boson, that affect both the $\mathrm{t\bar{t}H}$ and tH cross-sections. Kinematic effects of these anomalous couplings in tH production are taken into account in the analysis. The analysis constrains the κ_{t} coupling modifier to be in either of the two intervals $-1.15 < \kappa_{\mathrm{t}} < -0.8$ and $0.9 < \kappa_{\mathrm{t}} < 1.25$ at 95% confidence level, under the assumption that the Higgs boson coupling to the W boson is the predicted by the SM. Confidence regions are also set in the $(\kappa_{\mathrm{t}}, \kappa_{\mathrm{W}})$ plane of the coupling modifiers.

In summary, this thesis has exploited the potential of final states with two or more leptons to perform both SM precision measurements and searches for BSM physics. All the results obtained are consistent with SM expectations. Similarly, under this hypothesis, a measurement of the $\mathrm{t\bar{t}H}$ production in the multilepton channel has been performed with enough sensitivity for an observation.

Resumen y conclusiones

Los datos recopilados por el experimento CMS durante el Run 2 del LHC han dado paso al estudio de una serie de procesos con baja sección eficaz, permitiendo la determinación de cantidades predichas por el SM y la realización de búsquedas de procesos BSM. En esta tesis se realizan tanto medidas de precisión de parámetros del SM como búsquedas de física BSM, realizadas en colisiones pp medidas en el detector CMS durante el Run 2 del LHC. Estas medidas se han llevado a cabo en eventos con dos o más leptones de alto p_{T} en el estado final, que son característicos en la producción de partículas pesadas.

Se presenta una medida de la sección eficaz de producción de tW en sucesos $\mathrm{e}^{\pm}\mu^{\mp}$. Esta medida es complementaria a medidas de producción de $\mathrm{t\bar{t}}$, para las cuales tW es el proceso de fondo dominante. Los sucesos se clasifican de acuerdo a su multiplicidad de jets y de jets identificados como b-jets, y se hace uso de discriminadores multivariable para obtener regiones enriquecidas en la señal. El valor medido de la sección eficaz es

$$\sigma_{\mathrm{tW}} = 63.1 \pm 1.8\ (\mathrm{stat}) \pm 6.3\ (\mathrm{syst}) \pm 2.1\ (\mathrm{lumi})\ \mathrm{pb},$$

en acuerdo con las predicciones del SM, que son de 71.7 $\pm$ 1.8 (scale) $\pm$ 3.4 (PDF) pb calculadas con precisión aproximada a NNLO [105].

Se ha realizado también una medida diferencial en este canal, con sucesos con exactamente un jet identificado como b-jet. A pesar de la contribución significativa de sucesos de $\mathrm{t\bar{t}}$ a la región de señal, se ha medido sección eficaz diferencial con una precisión entre 20 y 100%, dependiendo de la distribución, consistentemente con las predicciones del SM. Estas medidas son un primer paso a medidas más precisas, así como medidas de la interferencia entre los procesos $\mathrm{t\bar{t}}$ y tW.

Se han llevado a cabo búsquedas de partículas SUSY en sucesos con dos leptones de distinta carga y mismo sabor y momento transverso faltante. Esta topología permite ser sensible a distintos modelos de producción de gluino, squarks, sleptones y compañeros supersimétricos de los bosones electrodébiles. Se han definido distintas regiones de señal para cada uno de los modos de producción considerados, en los cuales se han estimado las contribuciones de los distintos procesos de fondo usando técnicas basadas en datos. Estas medidas permiten excluir ciertos rangos masas de partículas supersimétricas no excluidos por búsquedas previas, asumiendo modelos simplificados de producción.

Por último, se ha realizado una medida de la producción de $\mathrm{t\bar{t}H}$ y tH. La medida se ha realizado en el canal multileptónico, enfocándose en los modos de desintegración del bosón de Higgs a WW, ZZ y $\tau\tau$. Los sucesos se clasifican de acuerdo al número de leptones y τ_{h} reconstruidos, aplicando selecciones en el número de jets y de jets identificados como b-jets. Los resultados se obtienen de dos formas complementarias. En el análisis principal, los eventos se clasifican de acuerdo a un clasificador multivariable y otras propiedades cinemáticas de los eventos. Se realiza también un análisis de control basándose en las propiedades cinemáticas de los sucesos y un discriminador basado

en los elementos de matriz. El fondo debido a leptones no producidos en desintegraciones de bosones W y Z o leptones τ se suprime eficientemente por medio de técnicas dedicadas. Los fondos irreducibles se estiman utilizando los modelos basados en el SM más precisos.

Estos análisis miden la tasa de producción de $t\bar{t}H$ con un valor de

$$\mu_{t\bar{t}H} = 0.92 \pm 0.19 \text{ (stat.)}^{+0.17}_{-0.13} \text{ (syst.+lumi.)}$$

veces la predicción del SM. La sección eficaz de este proceso se mide con una incertidumbre alrededor de 20%. El análisis principal presentado tiene una significancia observada (esperada) de 4.7 (5.2) desviaciones estándar para la producción de $t\bar{t}H$. Para el análisis de control la significancia observada (esperada) es de 3.8 (4.0). La significancia del análisis principal a la producción de tH es de 1.4 y 0.3, respectivamente.

El análisis principal también se interpreta en el contexto de variaciones en los acoplamientos del bosón de Higgs al quark top y el bosón W, afectando a las tasas de producción de $t\bar{t}H$ y tH. Los efectos cinemáticos de estos acoplamientos en la producción de tH se tienen en cuenta en el análisis. El análisis permite constreñir el modificador del acoplamiento κ_t a estar en los intervalos $-1.15 < \kappa_t < -0.8$ and $0.9 < \kappa_t < 1.25$, con un nivel de confianza del 95%, asumiendo que el acoplamiento del bosón W al Higgs es el predicho por el SM. Se establecen regiones de confianza en el plano (κ_t, κ_W) de los modificadores del acoplamiento.

En conclusión, esta tesis aprovecha el potencial de los estados finales con dos o más leptones para realizar medidas de precisión del SM y búsqueda de física BSM. Todos los resultados obtenidos son consistentes con las predicciones del SM. Además, de acuerdo con la hipótesis del SM, se ha realizado una medida de la producción de $t\bar{t}H$ en el canal multileptónico, suficientemente sensible para una observación del proceso.

List of acronyms

4FS 4-flavor scheme
5FS 5-flavor scheme
ALICE A Large Ion Collider Experiment
ANN Artificial Neural Network
APD Avalanche Photodiode
AR Application Region
ATLAS A Toroidal LHC ApparatuS
BDT Boosted Decision Tree
BSM Beyond Standard Model
CKM Cabibbo Kobayashi Maskawa
CMS Compact Muon Solenoid
CP Charge Parity
CSC Cathode Strip Chamber
DGLAP Dokshitzer-Gribov-Lipatov-Altarelli-Parisi
DNN Deep Neural Network
DR Diagram Removal
DR2 Diagram Removal 2
DS Diagram Subtraction
DT Drift Tube
ECAL Electromagnetic Calorimeter
EWK Electroweak
FCNC Flavor Changing Neutral Currents
FSR Final State Radiation
GMSB Gauge Mediated Supersymmetry Breaking
GSF Gaussian Sum Filtering
HCAL Hadronic Calorimeter
HLT High Level Trigger
ISR Initial State Radiation
ITC Inverted Top Coupling
JER Jet Energy Resolution
JES Jet Energy Scale
L1 First Level Trigger
LHC Large Hadron Collider
LHCb Large Hadron Collider beauty
LO Leading Order
LSP Lightest Supersymmetric Particle
MC Monte Carlo
MEM Matrix Element Method
MLE Maximum Likelihood Estimator
MR Measurement Region
MSSM Minimal Supersymmetric Standard Model
MVA Multivariate
NLL Next-to-leading logarithm
NLO Next-to-leading Order
NLSP Next-to-lightest Supersymmetric Particle
NNLL Next-to-next-to-leading Logarithm
NNLO Next-to-next-to-leading Order
NWA Narrow Width Approximation
PDF Parton distribution function
pdf probability density function
PF Particle Flow
PS Parton Shower
QCD Quantum Chromodynamics
QFT Quantum Field Theory
RPC Resistive Plate Chamber
SC Super Cluster
SM Standard Model
SR Signal Region
SSB Spontaneous Symmetry Breaking
SUSY Supersymmetry
UE Underlying Event
WLCG Worldwide LHC Computing Grid
WP Working point

Bibliography

[1] Serguei Chatrchyan et al. "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC". *Phys. Lett. B* 716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[2] Georges Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". *Phys. Lett. B* 716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[3] "Performance of top triggers in Run 2" (2019). URL: https://cds.cern.ch/record/2687014.

[4] "Muon identification and isolation efficiencies with 2017 and 2018 data" (2018). URL: https://cds.cern.ch/record/2629364.

[5] "Muon reconstruction performance during Run II" (2019). URL: http://cds.cern.ch/record/2682902.

[6] Albert M Sirunyan et al. "Measurement of the $\mathrm{t\bar{t}}$ production cross section, the top quark mass, and the strong coupling constant using dilepton events in pp collisions at $\sqrt{s} = 13$ TeV". *Eur. Phys. J. C* 79.5 (2019), p. 368. DOI: 10.1140/epjc/s10052-019-6863-8. arXiv: 1812.10505.

[7] Vardan Khachatryan et al. "Measurement of the $\mathrm{t\bar{t}}$ production cross section in the e-mu channel in proton-proton collisions at $\sqrt{s}$ = 7 and 8 TeV". *JHEP* 08 (2016), p. 029. DOI: 10.1007/JHEP08(2016)029. arXiv: 1603.02303.

[8] Vardan Khachatryan et al. "Measurement of the top quark pair production cross section in proton-proton collisions at $\sqrt{(s)} = 13$ TeV". *Phys. Rev. Lett.* 116.5 (2016), p. 052002. DOI: 10.1103/PhysRevLett.116.052002. arXiv: 1510.05302.

[9] Albert M Sirunyan et al. "Measurement of the production cross section for single top quarks in association with W bosons in proton-proton collisions at $\sqrt{s} = 13$ TeV". *JHEP* 10 (2018), p. 117. DOI: 10.1007/JHEP10(2018)117. arXiv: 1805.07399.

[10] CMS Collaboration. "Measurement of differential cross sections for single top quark production in association with a W boson at $\sqrt{s} = 13$ TeV" (Mar. 2020).

[11] Vardan Khachatryan et al. "Search for new physics in final states with two opposite-sign, same-flavor leptons, jets, and missing transverse momentum in pp collisions at sqrt(s) = 13 TeV". *JHEP* 12 (2016), p. 013. DOI: 10.1007/JHEP12(2016)013. arXiv: 1607.00915.

[12] Albert M Sirunyan et al. "Search for new phenomena in final states with two opposite-charge, same-flavor leptons, jets, and missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV". *JHEP* 03 (2018), p. 076. DOI: 10.1007/s13130-018-7845-2. arXiv: 1709.08908.

[13] A.M. Sirunyan et al. "Combined search for electroweak production of charginos and neutralinos in proton-proton collisions at $\sqrt{s} = 13$ TeV". *JHEP* 03 (2018), p. 160. DOI: 10.1007/JHEP03(2018)160. arXiv: 1801.03957.

[14] Albert M Sirunyan et al. "Evidence for associated production of a Higgs boson with a top quark pair in final states with electrons, muons, and hadronically decaying τ leptons at $\sqrt{s} = 13$ TeV". *JHEP* 08 (2018), p. 066. DOI: 10.1007/JHEP08(2018)066. arXiv: 1803.05485.

[15] Albert M Sirunyan et al. "Observation of $t\bar{t}H$ production". *Phys. Rev. Lett.* 120.23 (2018), p. 231801. DOI: 10.1103/PhysRevLett.120.231801. arXiv: 1804.02610.

[16] "Measurement of the associated production of a Higgs boson with a top quark pair in final states with electrons, muons and hadronically decaying τ leptons in data recorded in 2017 at $\sqrt{s} = 13$ TeV" (Nov. 2018).

[17] M. Tanabashi et al. "Review of Particle Physics". *Phys. Rev. D* 98 (3 2018), p. 030001. DOI: 10.1103/PhysRevD.98.030001. URL: https://link.aps.org/doi/10.1103/PhysRevD.98.030001.

[18] S.L. Glashow. "Partial Symmetries of Weak Interactions". *Nucl. Phys.* 22 (1961), pp. 579–588. DOI: 10.1016/0029-5582(61)90469-2.

[19] Steven Weinberg. "A Model of Leptons". *Phys. Rev. Lett.* 19 (21 1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264. URL: https://link.aps.org/doi/10.1103/PhysRevLett.19.1264.

[20] S. L. Glashow, J. Iliopoulos, and L. Maiani. "Weak Interactions with Lepton-Hadron Symmetry". *Phys. Rev. D* 2 (7 1970), pp. 1285–1292. DOI: 10.1103/PhysRevD.2.1285. URL: https://link.aps.org/doi/10.1103/PhysRevD.2.1285.

[21] Peter W. Higgs. "Spontaneous Symmetry Breakdown without Massless Bosons". *Phys. Rev.* 145 (4 1966), pp. 1156–1163. DOI: 10.1103/PhysRev.145.1156. URL: https://link.aps.org/doi/10.1103/PhysRev.145.1156.

[22] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". *Phys. Rev. Lett.* 13 (9 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.321.

[23] B. Pontecorvo. "Neutrino Experiments and the Problem of Conservation of Leptonic Charge". *Sov. Phys. JETP* 26 (1968), pp. 984–988.

[24] Gianfranco Bertone, Dan Hooper, and Joseph Silk. "Particle dark matter: Evidence, candidates and constraints". *Phys. Rept.* 405 (2005), pp. 279–390. DOI: 10.1016/j.physrep.2004.08.031. arXiv: hep-ph/0404175.

[25] Fabio Maltoni, Giovanni Ridolfi, and Maria Ubiali. "b-initiated processes at the LHC: a reappraisal". *JHEP* 07 (2012). [Erratum: JHEP 04, 095 (2013)], p. 022. DOI: 10.1007/JHEP04(2013)095. arXiv: 1203.6393 [hep-ph].

[26] Torbjörn Sjöstrand et al. "An Introduction to PYTHIA 8.2". *Comput. Phys. Commun.* 191 (2015), pp. 159–177. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012.

[27] Johannes Bellm et al. "Herwig 7.0/Herwig++ 3.0 release note". *Eur. Phys. J. C* 76.4 (2016), p. 196. DOI: 10.1140/epjc/s10052-016-4018-8. arXiv: 1512.01178 [hep-ph].

[28] Stefano Frixione, Paolo Nason, and Carlo Oleari. "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method". *JHEP* 11 (2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092.

[29] 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), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph].

[30] "LHC Machine". *JINST* 3 (2008). Ed. by Lyndon Evans and Philip Bryant, S08001. DOI: 10.1088/1748-0221/3/08/S08001.

[31] CMS Collaboration. *CMS Luminosity - Public results*. 2020. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults (visited on 05/20/2020).

[32] G L Bayatian et al. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical Design Report CMS. Geneva: CERN, 2006. URL: https://cds.cern.ch/record/922757.

[33] CMS Collaboration. *CMS detector design*. 2020. URL: http://cms.web.cern.ch/news/cms-detector-design (visited on 05/20/2020).

[34] "CMS Tracker Alignment Performance Results 2018" (2018). URL: http://cds.cern.ch/record/2650977.

[35] "ECAL 2016 refined calibration and Run2 summary plots" (2020). URL: https://cds.cern.ch/record/2717925.

[36] A.M. Sirunyan et al. "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV". *JINST* 13.06 (2018), P06015. DOI: 10.1088/1748-0221/13/06/P06015. arXiv: 1804.04528.

[37] CMS Collaboration. *Plots approved for CHEP2016, San Francisco*. 2020. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/HLTplotsCHEP2016 (visited on 05/20/2020).

[38] CERN. *LHC Computing Grid*. 2020. URL: http://www.cern.ch/lcg/.

[39] A.M. Sirunyan et al. "Particle-flow reconstruction and global event description with the CMS detector". *JINST* 12.10 (2017), P10003. DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965 [physics.ins-det].

[40] Albert M Sirunyan et al. "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector". *JINST* 14.07 (2019), P07004. DOI: 10.1088/1748-0221/14/07/P07004. arXiv: 1903.06078 [hep-ex].

[41] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The anti-k_t jet clustering algorithm". *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

[42] Albert M Sirunyan et al. "Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques" (Apr. 2020). arXiv: 2004.08262 [hep-ex].

[43] A.M. Sirunyan et al. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". *JINST* 13.05 (2018), P05011. DOI: 10.1088/1748-0221/13/05/P05011. arXiv: 1712.07158 [physics.ins-det].

[44] "CMS Phase 1 heavy flavour identification performance and developments" (2017). URL: http://cds.cern.ch/record/2263802.

[45] "Performance of the DeepJet b tagging algorithm using 41.9/fb of data from proton-proton collisions at 13TeV with Phase 1 CMS detector" (2018). URL: http://cds.cern.ch/record/2646773.

[46] *Jet algorithms performance in 13 TeV data*. Tech. rep. CMS-PAS-JME-16-003. Geneva: CERN, 2017. URL: http://cds.cern.ch/record/2256875.

[47] M. Asai. "Geant4-a simulation toolkit". *Trans. Amer. Nucl. Soc.* 95 (2006), p. 757.

[48] S. Abdullin et al. "The fast simulation of the CMS detector at LHC". *J. Phys. Conf. Ser.* 331 (2011). Ed. by Simon C. Lin, p. 032049. DOI: 10.1088/1742-6596/331/3/032049.

[49] Serguei Chatrchyan et al. "Measurement of the Inclusive W and Z Production Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV". *JHEP* 10 (2011), p. 132. DOI: 10.1007/JHEP10(2011)132. arXiv: 1107.4789.

[50] Vardan Khachatryan et al. "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV". *JINST* 12.02 (2017), P02014. DOI: 10.1088/1748-0221/12/02/P02014. arXiv: 1607.03663 [hep-ex].

[51] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016.

[52] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. New York, NY, USA: Springer New York Inc., 2001.

[53] Andreas Hoecker et al. "TMVA: Toolkit for Multivariate Data Analysis". *PoS* ACAT (2007), p. 040. arXiv: physics/0703039.

[54] Moshe Leshno et al. "Multilayer feedforward networks with a nonpolynomial activation function can approximate any function". *Neural Networks* 6.6 (1993), pp. 861 –867. ISSN: 0893-6080. DOI: https://doi.org/10.1016/S0893-6080(05)80131-5. URL: http://www.sciencedirect.com/science/article/pii/S0893608005801315.

[55] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: https://www.tensorflow.org/.

[56] François Chollet et al. *Keras*. https://keras.io. 2015.

[57] *Procedure for the LHC Higgs boson search combination in Summer 2011*. Tech. rep. CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11. Geneva: CERN, 2011. URL: https://cds.cern.ch/record/1379837.

[58] Roger J. Barlow and Christine Beeston. "Fitting using finite Monte Carlo samples". *Comput. Phys. Commun.* 77 (1993), pp. 219–228. DOI: 10.1016/0010-4655(93)90005-W.

[59] J.S. Conway. "Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra". *PHYSTAT 2011*. 2011, pp. 115–120. DOI: 10.5170/CERN-2011-006.115. arXiv: 1103.0354 [physics.data-an].

[60] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". *Eur. Phys. J. C* 71 (2011). [Erratum: Eur.Phys.J.C 73, 2501 (2013)], p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an].

[61] Glen Cowan. *Statistical Data Analysis*. Oxford: Clarendon (Oxford), 1998.

[62] A. N. Tikhonov. "Solution of incorrectly formulated problems and the regularization method". *Soviet Math. Dokl.* 4 (1963), pp. 1035–1038.

[63] Stefan Schmitt. "TUnfold: an algorithm for correcting migration effects in high energy physics". *JINST* 7 (2012), T10003. DOI: 10.1088/1748-0221/7/10/T10003. arXiv: 1205.6201.

[64] R Frühwirth. "Application of Kalman filtering to track and vertex fitting". *Nucl. Instrum. Methods Phys. Res., A* 262 (1987), 444. 19 p. URL: https://cds.cern.ch/record/178627.

[65] Serguei Chatrchyan et al. "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.

[66] Albert M Sirunyan et al. "Performance of the reconstruction and identification of high-momentum muons in proton-proton collisions at $\sqrt{s} = 13$ TeV". *JINST* 15.02 (2020), P02027. DOI: 10.1088/1748-0221/15/02/P02027. arXiv: 1912.03516.

[67] A. Bodek et al. "Extracting Muon Momentum Scale Corrections for Hadron Collider Experiments". *Eur. Phys. J. C* 72 (2012), p. 2194. DOI: 10.1140/epjc/s10052-012-2194-8. arXiv: 1208.3710.

[68] Vardan Khachatryan et al. "Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s}$ = 8 TeV". *JINST* 10.06 (2015), P06005. DOI: 10.1088/1748-0221/10/06/P06005. arXiv: 1502.02701.

[69] Wolfgang Adam et al. *Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC*. Tech. rep. Jan. 2005.

[70] Matteo Cacciari and Gavin P. Salam. "Pileup subtraction using jet areas". *Phys. Lett. B* 659 (2008), pp. 119–126. DOI: 10.1016/j.physletb.2007.09.077. arXiv: 0707.1378.

[71] A.M. Sirunyan et al. "Search for electroweak production of charginos and neutralinos in multilepton final states in proton-proton collisions at $\sqrt{s} = 13$ TeV". *JHEP* 03 (2018), p. 166. DOI: 10.1007/JHEP03(2018)166. arXiv: 1709.05406.

[72] Albert M Sirunyan et al. "Measurement of top quark pair production in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV". *JHEP* 03 (2020), p. 056. DOI: 10.1007/JHEP03(2020)056. arXiv: 1907.11270.

[73] Albert M Sirunyan et al. "Observation of Single Top Quark Production in Association with a Z Boson in Proton-Proton Collisions at $\sqrt{s}$ =13 TeV". *Phys. Rev. Lett.* 122.13 (2019), p. 132003. DOI: 10.1103/PhysRevLett.122.132003. arXiv: 1812.05900.

[74] *Studies of W^+W^- production at $\sqrt{s} = 13$ TeV*. Tech. rep. CMS-PAS-SMP-18-004. Geneva: CERN, 2020. URL: http://cds.cern.ch/record/2714766.

[75] Albert M Sirunyan et al. "Search for the production of four top quarks in the single-lepton and opposite-sign dilepton final states in proton-proton collisions at $\sqrt{s}$ = 13 TeV". *JHEP* 11 (2019), p. 082. DOI: 10.1007/JHEP11(2019)082. arXiv: 1906.02805.

[76] "Electron and Photon performance in CMS with the full 2017 data sample and additional 2016 highlights for the CALOR 2018 Conference" (2018). URL: https://cds.cern.ch/record/2320638.

[77] *Performance of reconstruction and identification of tau leptons in their decays to hadrons and tau neutrino in LHC Run-2*. Tech. rep. CMS-PAS-TAU-16-002. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2196972.

[78] "Performance of the DeepTau algorithm for the discrimination of taus against jets, electron, and muons" (2019). URL: http://cds.cern.ch/record/2694158.

[79] A.M. Sirunyan et al. "Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV". *JINST* 13.10 (2018), P10005. DOI: 10.1088/1748-0221/13/10/P10005. arXiv: 1809.02816 [hep-ex].

[80] Makoto Kobayashi and Toshihide Maskawa. "CP-Violation in the Renormalizable Theory of Weak Interaction". *Progress of Theoretical Physics* 49.2 (Feb. 1973), pp. 652–657. ISSN: 0033-068X. DOI: 10.1143/PTP.49.652. eprint: https://academic.oup.com/ptp/article-pdf/49/2/652/5257692/49-2-652.pdf. URL: https://doi.org/10.1143/PTP.49.652.

[81] F. Abe et al. "Observation of top quark production in $\mathrm{p\overline{p}}$ collisions". *Phys. Rev. Lett.* 74 (1995), pp. 2626–2631. DOI: 10.1103/PhysRevLett.74.2626. arXiv: hep-ex/9503002.

[82] S. Abachi et al. "Observation of the top quark". *Phys. Rev. Lett.* 74 (1995), pp. 2632–2637. DOI: 10.1103/PhysRevLett.74.2632. arXiv: hep-ex/9503003.

[83] Stephen P. Martin. "A Supersymmetry primer". *Perspectives on supersymmetry. Vol.2*. Vol. 21. 2010, pp. 1–153. DOI: 10.1142/9789812839657_0001. arXiv: hep-ph/9709356.

[84] LHC Top Working Group. *LHCTopWG Summary Plots*. 2020. URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCTopWGSummaryPlots.

[85] Sayipjamal Dulat et al. "New parton distribution functions from a global analysis of quantum chromodynamics". *Phys. Rev. D* 93.3 (2016), p. 033006. DOI: 10.1103/PhysRevD.93.033006. arXiv: 1506.07443.

[86] Michal Czakon and Alexander Mitov. "Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders". *Comput. Phys. Commun.* 185 (2014), p. 2930. DOI: 10.1016/j.cpc.2014.06.021. arXiv: 1112.5675.

[87] Simone Alioli et al. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX". *JHEP* 06 (2010), p. 043. DOI: 10.1007/JHEP06(2010)043. arXiv: 1002.2581.

[88] Paolo Nason. "A New method for combining NLO QCD with shower Monte Carlo algorithms". *JHEP* 11 (2004), p. 040. DOI: 10.1088/1126-6708/2004/11/040. arXiv: hep-ph/0409146.

[89] Rikkert Frederix and Stefano Frixione. "Merging meets matching in MC@NLO". *JHEP* 12 (2012), p. 061. DOI: 10.1007/JHEP12(2012)061. arXiv: 1209.6215.

[90] Michal Czakon, David Heymes, and Alexander Mitov. "High-precision differential predictions for top-quark pairs at the LHC". *Phys. Rev. Lett.* 116.8 (2016), p. 082003. DOI: 10.1103/PhysRevLett.116.082003. arXiv: 1511.00549.

[91] Jun Gao and Andrew S. Papanastasiou. "Top-quark pair-production and decay at high precision". *Phys. Rev. D* 96.5 (2017), p. 051501. DOI: 10.1103/PhysRevD.96.051501. arXiv: 1705.08903.

[92] Andrew S. Papanastasiou. "Top Quark Pair Production: theory overview". *10th International Workshop on Top Quark Physics*. Jan. 2018. arXiv: 1801.01020.

[93] Tomáš Ježo and Paolo Nason. "On the Treatment of Resonances in Next-to-Leading Order Calculations Matched to a Parton Shower". *JHEP* 12 (2015), p. 065. DOI: 10.1007/JHEP12(2015)065. arXiv: 1509.09071.

[94] Tomáš Ježo et al. "An NLO+PS generator for $\mathrm{t\bar{t}}$ and Wt production and decay including non-resonant and interference effects". *Eur. Phys. J. C* 76.12 (2016), p. 691. DOI: 10.1140/epjc/s10052-016-4538-2. arXiv: 1607.04538.

[95] Georges Aad et al. "Measurement of the $t\bar{t}$ production cross-section using $\mathrm{e\mu}$ events with b-tagged jets in pp collisions at $\sqrt{s}$ = 7 and 8 TeV with the ATLAS detector". *Eur. Phys. J. C* 74.10 (2014). [Addendum: Eur.Phys.J.C 76, 642 (2016)], p. 3109. DOI: 10.1140/epjc/s10052-016-4501-2. arXiv: 1406.5375.

[96] Georges Aad et al. "Measurement of the $\mathrm{t\bar{t}}$ production cross-section and lepton differential distributions in $\mathrm{e\mu}$ dilepton events from pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector" (Oct. 2019). arXiv: 1910.08819.

[97] Albert M Sirunyan et al. "Measurements of $\mathrm{t\bar{t}}$ differential cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV using events containing two leptons". *JHEP* 02 (2019), p. 149. DOI: 10.1007/JHEP02(2019)149. arXiv: 1811.06625.

[98] Albert M Sirunyan et al. "Measurement of the top quark polarization and $\mathrm{t\bar{t}}$ spin correlations using dilepton final states in proton-proton collisions at $\sqrt{s} =$ 13 TeV". *Phys. Rev. D* 100.7 (2019), p. 072002. DOI: 10.1103/PhysRevD.100.072002. arXiv: 1907.03729.

[99] A.M. Sirunyan et al. "Measurements of differential cross sections of top quark pair production as a function of kinematic event variables in proton-proton collisions at $\sqrt{s} = 13$ TeV". *JHEP* 06 (2018), p. 002. DOI: 10.1007/JHEP06(2018)002. arXiv: 1803.03991.

[100] Albert M Sirunyan et al. "Measurement of $\mathrm{t\bar{t}}$ normalised multi-differential cross sections in pp collisions at $\sqrt{s} = 13$ TeV, and simultaneous determination of the strong coupling strength, top quark pole mass, and parton distribution functions" (Apr. 2019). arXiv: 1904.05237.

[101] Georges Aad et al. "Measurements of top-quark pair differential and double-differential cross-sections in the l+jets channel with pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector". *Eur. Phys. J. C* 79.12 (2019), p. 1028. DOI: 10.1140/epjc/s10052-019-7525-6. arXiv: 1908.07305.

[102] Alexander Mitov. "Dilepton fiducial region definitions". LHC TOP WG meeting. 2020. URL: https://indico.cern.ch/event/888620/timetable/#9-dilepton-fiducial-region-def.

[103] Albert M Sirunyan et al. "Measurement of the top quark Yukawa coupling from $\mathrm{t\bar{t}}$ kinematic distributions in the lepton+jets final state in proton-proton collisions at $\sqrt{s} = 13$ TeV". *Phys. Rev. D* 100.7 (2019), p. 072007. DOI: 10.1103/PhysRevD.100.072007. arXiv: 1907.01590.

[104] P. Kant et al. "HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions". *Comput. Phys. Commun.* 191 (2015), pp. 74–89. DOI: 10.1016/j.cpc.2015.02.001. arXiv: 1406.4403.

[105] Nikolaos Kidonakis. "Theoretical results for electroweak-boson and single-top production". *PoS* DIS2015 (2015), p. 170. DOI: 10.22323/1.247.0170. arXiv: 1506.04072.

[106] Georges Aad et al. "Observation of the associated production of a top quark and a Z boson in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector" (Feb. 2020). arXiv: 2002.07546.

[107] John Campbell, R. Keith Ellis, and Raoul Röntsch. "Single top production in association with a Z boson at the LHC". *Phys. Rev. D* 87 (2013), p. 114006. DOI: 10.1103/PhysRevD.87.114006. arXiv: 1302.3856.

[108] Serguei Chatrchyan et al. "Evidence for Associated Production of a Single Top Quark and W Boson in pp Collisions at $\sqrt{s}$ = 7 TeV". *Phys. Rev. Lett.* 110 (2013), p. 022003. DOI: 10.1103/PhysRevLett.110.022003. arXiv: 1209.3489.

[109] Georges Aad et al. "Evidence for the associated production of a W boson and a top quark in ATLAS at $\sqrt{s} = 7$ TeV". *Phys. Lett. B* 716 (2012), pp. 142–159. DOI: 10.1016/j.physletb.2012.08.011. arXiv: 1205.5764.

[110] Serguei Chatrchyan et al. "Observation of the associated production of a single top quark and a W boson in pp collisions at $\sqrt{s}$ =8 TeV". *Phys. Rev. Lett.* 112.23 (2014), p. 231802. DOI: 10.1103/PhysRevLett.112.231802. arXiv: 1401.2942.

[111] Georges Aad et al. "Measurement of the production cross-section of a single top quark in association with a W boson at 8 TeV with the ATLAS experiment". *JHEP* 01 (2016), p. 064. DOI: 10.1007/JHEP01(2016)064. arXiv: 1510.03752.

[112] Morad Aaboud et al. "Measurement of the cross-section for producing a W boson in association with a single top quark in pp collisions at $\sqrt{s} = 13$ TeV with ATLAS". *JHEP* 01 (2018), p. 063. DOI: 10.1007/JHEP01(2018)063. arXiv: 1612.07231.

[113] Morad Aaboud et al. "Measurement of differential cross-sections of a single top quark produced in association with a W boson at $\sqrt{s} = 13$ TeV with ATLAS". *Eur. Phys. J. C* 78.3 (2018), p. 186. DOI: 10.1140/epjc/s10052-018-5649-8. arXiv: 1712.01602.

[114] Albert M Sirunyan et al. "Search for new physics in top quark production in dilepton final states in proton-proton collisions at $\sqrt{s}$ = 13 TeV". *Eur. Phys. J. C* 79.11 (2019), p. 886. DOI: 10.1140/epjc/s10052-019-7387-y. arXiv: 1903.11144.

[115] Stefano Frixione et al. "Single-top hadroproduction in association with a W boson". *JHEP* 07 (2008), p. 029. DOI: 10.1088/1126-6708/2008/07/029. arXiv: 0805.3067.

[116] Wolfgang Hollik, Jonas M. Lindert, and Davide Pagani. "NLO corrections to squark-squark production and decay at the LHC". *JHEP* 03 (2013), p. 139. DOI: 10.1007/JHEP03(2013)139. arXiv: 1207.1071.

[117] Tim M. P. Tait. "tW^- mode of single top quark production". *Phys. Rev. D* 61 (3 1999), p. 034001. DOI: 10.1103/PhysRevD.61.034001. URL: https://link.aps.org/doi/10.1103/PhysRevD.61.034001.

[118] Morad Aaboud et al. "Probing the quantum interference between singly and doubly resonant top-quark production in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". *Phys. Rev. Lett.* 121.15 (2018), p. 152002. DOI: 10.1103/PhysRevLett.121.152002. arXiv: 1806.04667.

[119] Christian Herwig, Tomáš Ježo, and Benjamin Nachman. "Extracting the Top-Quark Width from Nonresonant Production". *Phys. Rev. Lett.* 122.23 (2019), p. 231803. DOI: 10.1103/PhysRevLett.122.231803. arXiv: 1903.10519.

[120] Emanuele Re. "Single-top Wt-channel production matched with parton showers using the POWHEG method". *Eur. Phys. J. C* 71 (2011), p. 1547. DOI: 10.1140/epjc/s10052-011-1547-z. arXiv: 1009.2450.

[121] Richard D. Ball et al. "Parton distributions for the LHC Run II". *JHEP* 04 (2015), p. 040. DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849.

[122] Vardan Khachatryan et al. "Event generator tunes obtained from underlying event and multiparton scattering measurements". *Eur. Phys. J. C* 76.3 (2016), p. 155. DOI: 10.1140/epjc/s10052-016-3988-x. arXiv: 1512.00815.

[123] *Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of $t\bar{t}$ at $\sqrt{s} = 8$ and 13 TeV*. Tech. rep. CMS-PAS-TOP-16-021. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2235192.

[124] M. Aaboud et al. "Measurement of the Inelastic Proton-Proton Cross Section at $\sqrt{s} = 13$ TeV with the ATLAS Detector at the LHC". *Phys. Rev. Lett.* 117.18 (2016), p. 182002. DOI: 10.1103/PhysRevLett.117.182002. arXiv: 1606.02625 [hep-ex].

[125] *CMS Luminosity Measurements for the 2016 Data Taking Period*. Tech. rep. CMS-PAS-LUM-17-001. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2257069.

[126] Jesper R. Christiansen and Peter Z. Skands. "String Formation Beyond Leading Colour". *JHEP* 08 (2015), p. 003. DOI: 10.1007/JHEP08(2015)003. arXiv: 1505.01681.

[127] Spyros Argyropoulos and Torbjörn Sjöstrand. "Effects of color reconnection on $\mathrm{t\bar{t}}$ final states at the LHC". *JHEP* 11 (2014), p. 043. DOI: 10.1007/JHEP11(2014)043. arXiv: 1407.6653.

[128] A. Heister et al. "Study of the fragmentation of b quarks into B mesons at the Z peak". *Phys. Lett. B* 512 (2001), pp. 30–48. DOI: 10.1016/S0370-2693(01)00690-6. arXiv: hep-ex/0106051.

[129] J. Abdallah et al. "A study of the b-quark fragmentation function with the DELPHI detector at LEP I and an averaged distribution obtained at the Z Pole". *Eur. Phys. J. C* 71 (2011), p. 1557. DOI: 10.1140/epjc/s10052-011-1557-x. arXiv: 1102.4748.

[130] Albert M Sirunyan et al. "Measurement of differential cross sections and charge ratios for *t*-channel single top quark production in proton-proton collisions at $\sqrt{s} = 13$ TeV" (July 2019). arXiv: 1907.08330.

[131] CMS Collaboration. *Object definitions for top quark analyses at the particle level*. Tech. rep. CMS-NOTE-2017-004. CERN-CMS-NOTE-2017-004. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2267573.

[132] Konstantin T. Matchev and Scott D. Thomas. "Higgs and Z boson signatures of supersymmetry". *Phys. Rev. D* 62 (2000), p. 077702. DOI: 10.1103/PhysRevD.62.077702. arXiv: hep-ph/9908482.

[133] Patrick Meade, Matthew Reece, and David Shih. "Prompt Decays of General Neutralino NLSPs at the Tevatron". *JHEP* 05 (2010), p. 105. DOI: 10.1007/JHEP05(2010)105. arXiv: 0911.4130.

[134] CMS Collaboration. "Search for selectrons and smuons at $\sqrt{s} = 13$ TeV" (Dec. 2017).

[135] Vardan Khachatryan et al. "Search for Physics Beyond the Standard Model in Events with Two Leptons, Jets, and Missing Transverse Momentum in pp Collisions at *sqrts* = 8 TeV". *JHEP* 04 (2015), p. 124. DOI: 10.1007/JHEP04(2015)124. arXiv: 1502.06031.

[136] Vardan Khachatryan et al. "Searches for electroweak production of charginos, neutralinos, and sleptons decaying to leptons and W, Z, and Higgs bosons in pp collisions at 8 TeV". *Eur. Phys. J. C* 74.9 (2014), p. 3036. DOI: 10.1140/epjc/s10052-014-3036-7. arXiv: 1405.7570.

[137] Vardan Khachatryan et al. "Searches for electroweak neutralino and chargino production in channels with Higgs, Z, and W bosons in pp collisions at 8 TeV". *Phys. Rev. D* 90.9 (2014), p. 092007. DOI: 10.1103/PhysRevD.90.092007. arXiv: 1409.3168.

[138] Albert M Sirunyan et al. "Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements". *Eur. Phys. J. C* 80.1 (2020), p. 4. DOI: 10.1140/epjc/s10052-019-7499-4. arXiv: 1903.12179.

[139] Julien Baglio, Le Duc Ninh, and Marcus M. Weber. "Massive gauge boson pair production at the LHC: a next-to-leading order story". *Phys. Rev. D* 88 (2013). [Erratum: Phys.Rev.D 94, 099902 (2016)], p. 113005. DOI: 10.1103/PhysRevD.94.099902. arXiv: 1307.4331.

[140] John M. Campbell and R.K. Ellis. "MCFM for the Tevatron and the LHC". *Nucl. Phys. B Proc. Suppl.* 205-206 (2010). Ed. by Johannes Blümlein, Sven-Olaf Moch, and Tord Riemann, pp. 10–15. DOI: 10.1016/j.nuclphysbps.2010.08.011. arXiv: 1007.3492.

[141] John M. Campbell et al. "Two loop correction to interference in $gg \to ZZ$". *JHEP* 08 (2016), p. 011. DOI: 10.1007/JHEP08(2016)011. arXiv: 1605.01380.

[142] Benjamin Fuks et al. "Gaugino production in proton-proton collisions at a center-of-mass energy of 8 TeV". *JHEP* 10 (2012), p. 081. DOI: 10.1007/JHEP10(2012)081. arXiv: 1207.2159.

[143] Benjamin Fuks et al. "Precision predictions for electroweak superpartner production at hadron colliders with Resummino". *Eur. Phys. J. C* 73 (2013), p. 2480. DOI: 10.1140/epjc/s10052-013-2480-0. arXiv: 1304.0790.

[144] Christoph Borschensky et al. "Squark and gluino production cross sections in pp collisions at $\sqrt{s}$ = 13, 14, 33 and 100 TeV". *Eur. Phys. J. C* 74.12 (2014), p. 3174. DOI: 10.1140/epjc/s10052-014-3174-y. arXiv: 1407.5066.

[145] Vardan Khachatryan et al. "Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at sqrt(s) = 8 TeV". *JINST* 10.08 (2015), P08010. DOI: 10.1088/1748-0221/10/08/P08010. arXiv: 1502.02702.

[146] C.G. Lester and D.J. Summers. "Measuring masses of semiinvisibly decaying particles pair produced at hadron colliders". *Phys. Lett. B* 463 (1999), pp. 99–103. DOI: 10.1016/S0370-2693(99)00945-4. arXiv: hep-ph/9906349.

[147] Alan Barr, Christopher Lester, and P. Stephens. "m(T2): The Truth behind the glamour". *J. Phys. G* 29 (2003), pp. 2343–2363. DOI: 10.1088/0954-3899/29/10/304. arXiv: hep-ph/0304226.

[148] Albert M Sirunyan et al. "Measurements of the pp $\rightarrow$ WZ inclusive and differential production cross section and constraints on charged anomalous triple gauge couplings at $\sqrt{s} = 13$ TeV". *JHEP* 04 (2019), p. 122. DOI: 10.1007/JHEP04(2019)122. arXiv: 1901.03428.

[149] J R Andersen et al. "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties" (July 2013). Ed. by S Heinemeyer et al. DOI: 10.5170/CERN-2013-004. arXiv: 1307.1347.

[150] D. de Florian et al. "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector" (Oct. 2016). DOI: 10.23731/CYRM-2017-002. arXiv: 1610.07922.

[151] Federico Demartin et al. "tWH associated production at the LHC". *Eur. Phys. J. C* 77.1 (2017), p. 34. DOI: 10.1140/epjc/s10052-017-4601-7. arXiv: 1607.05862.

[152] Albert M Sirunyan et al. "Combined measurements of Higgs boson couplings in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$". *Eur. Phys. J. C* 79.5 (2019), p. 421. DOI: 10.1140/epjc/s10052-019-6909-y. arXiv: 1809.10733.

[153] Vardan Khachatryan et al. "Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV". *Phys. Rev. D* 92.1 (2015), p. 012004. DOI: 10.1103/PhysRevD.92.012004. arXiv: 1411.3441.

[154] P. Artoisenet et al. "A framework for Higgs characterisation". *JHEP* 11 (2013), p. 043. DOI: 10.1007/JHEP11(2013)043. arXiv: 1306.6464.

[155] Federico Demartin et al. "Higgs production in association with a single top quark at the LHC". *Eur. Phys. J. C* 75.6 (2015), p. 267. DOI: 10.1140/epjc/s10052-015-3475-9. arXiv: 1504.00611.

[156] Fabio Maltoni et al. "Trilinear Higgs coupling determination via single-Higgs differential measurements at the LHC". *Eur. Phys. J. C* 77.12 (2017), p. 887. DOI: 10.1140/epjc/s10052-017-5410-8. arXiv: 1709.08649.

[157] "Measurement of $\mathrm{t\bar{t}H}$ production in the $\mathrm{H} \rightarrow \mathrm{b\bar{b}}$ decay channel in $41.5\,\text{fb}^{-1}$ of proton-proton collision data at $\sqrt{s} = 13\,\text{TeV}$" (May 2019).

[158] Morad Aaboud et al. "Search for the standard model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector". *Phys. Rev. D* 97.7 (2018), p. 072016. DOI: 10.1103/PhysRevD.97.072016. arXiv: 1712.08895.

[159] Georges Aad et al. "Study of the CP properties of the interaction of the Higgs boson with top quarks using top quark associated production of the Higgs boson and its decay into two photons with the ATLAS detector at the LHC" (Apr. 2020). arXiv: 2004.04545.

[160] Albert M Sirunyan et al. "Measurements of $\mathrm{t\bar{t}H}$ production and the CP structure of the Yukawa interaction between the Higgs boson and top quark in the diphoton decay channel" (Mar. 2020). arXiv: 2003.10866.

[161] "Analysis of $\mathrm{t\bar{t}H}$ and $\mathrm{t\bar{t}W}$ production in multilepton final states with the ATLAS detector" (Oct. 2019).

[162] Georges Aad et al. "Higgs boson production cross-section measurements and their EFT interpretation in the 4l decay channel at $\sqrt{s}$ = 13 TeV with the ATLAS detector" (Apr. 2020). arXiv: 2004.03447.

[163] "Measurements of properties of the Higgs boson in the four-lepton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV" (Mar. 2019).

[164] M. Aaboud et al. "Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector". *Phys. Lett. B* 784 (2018), pp. 173–191. DOI: 10.1016/j.physletb.2018.07.035. arXiv: 1806.00425.

[165] LHC Higgs Cross Section WG. *LHC Higgs Cross Section WG Picture Gallery*. 2020. URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWGCrossSectionsFigures (visited on 05/20/2020).

[166] Jeff Asaf Dror et al. "Strong tW Scattering at the LHC". *JHEP* 01 (2016), p. 071. DOI: 10.1007/JHEP01(2016)071. arXiv: 1511.03674.

[167] Rikkert Frederix, Davide Pagani, and Marco Zaro. "Large NLO corrections in $\mathrm{t\bar{t}W^{\pm}}$ and $\mathrm{t\bar{t}t\bar{t}}$ hadroproduction from supposedly subleading EW contributions". *JHEP* 02 (2018), p. 031. DOI: 10.1007/JHEP02(2018)031. arXiv: 1711.02116.

[168] Alessandro Broggio et al. "Top-quark pair hadroproduction in association with a heavy boson at NLO+NNLL including EW corrections". *JHEP* 08 (2019), p. 039. DOI: 10.1007/JHEP08(2019)039. arXiv: 1907.04343.

[169] Giuseppe Bevilacqua et al. "The simplest of them all: $t\bar{t}W^{\pm}$ at NLO accuracy in QCD" (May 2020). arXiv: 2005.09427 [hep-ph].

[170] Morad Aaboud et al. "Measurement of the $\mathrm{t\bar{t}Z}$ and $\mathrm{t\bar{t}W}$ cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". *Phys. Rev. D* 99.7 (2019), p. 072009. DOI: 10.1103/PhysRevD.99.072009. arXiv: 1901.03584.

[171] Albert M Sirunyan et al. "Measurement of the cross section for top quark pair production in association with a W or Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV". *JHEP* 08 (2018), p. 011. DOI: 10.1007/JHEP08(2018)011. arXiv: 1711.02547.

[172] Simone Alioli et al. "ZZ production in gluon fusion at NLO matched to parton-shower". *Phys. Rev. D* 95.3 (2017), p. 034042. DOI: 10.1103/PhysRevD.95.034042. arXiv: 1609.09719.

[173] Lorenzo 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. A* 862 (2017), pp. 54–84. DOI: 10.1016/j.nima.2017.05.001. arXiv: 1603.05910 [hep-ex].