Universidad de Oviedo

Programa de Doctorado en Materiales

Estudio de procesos de producción de quarks top y energía faltante predichos por el Modelo Estándar y en sus extensiones, en estados finales leptónicos con el detector CMS en el LHC utilizando técnicas de aprendizaje automático

Study of the production processes of top quarks and missing energy predicted by the Standard Model and its extensions, in leptonic final states with the CMS detector at the LHC using machine learning techniques

TESIS DOCTORAL

Andrea Trapote Fernández

Diciembre 2023

Universidad de Oviedo

Programa de Doctorado en Materiales

Estudio de procesos de producción de quarks top y energía faltante predichos por el Modelo Estándar y en sus extensiones, en estados finales leptónicos con el detector CMS en el LHC utilizando técnicas de aprendizaje automático

Study of the production processes of top quarks and missing energy predicted by the Standard Model and its extensions, in leptonic final states with the CMS detector at the LHC using machine learning techniques

TESIS DOCTORAL

Directores de tesis

Dr. D. José Enrique Palencia Cortezón

Dr. D. Isidro González Caballero

Universidad de Oviedo

RESUMEN DEL CONTENIDO DE TESIS DOCTORAL

1.- Título de la Tesis	
Español: Estudio de procesos de producción de quarks top y energía faltante predichos por el Modelo Estándar y en sus extensiones, en estados finales leptónicos con el detector CMS en el LHC utilizando técnicas de aprendizaje automático.	Inglés: Study of the production processes of top quarks and missing energy predicted by the Standard Model and its extensions, in leptonic final states with the CMS detector at the LHC using machine learning techniques

2.- Autora	
Nombre: Andrea Trapote Fernández	
Programa de Doctorado: Materiales	
Órgano responsable: Centro Internacional de Postgrado	

RESUMEN (en español)

Durante el desarrollo de la tesis se han realizado medidas de procesos del Modelo Estándar (ME) y búsquedas de procesos más allá del ME relacionadas con el quark top en estados finales con leptones. Para ello, se han utilizado datos de colisiones protón-protón producidas en el LHC y recogidas por el experimento CMS del CERN.

Se ha realizado la primera medida de la sección eficaz del proceso de producción de pares quarks top-antitop con los primeros datos recogidos en el Run 3 del LHC a la nueva energía en el centro de masas de 13.6 TeV, obteniendo una sección eficaz de σ = 888 ± 34 (syst. + stat.) ± 20 (lumi) pb, consistente con las predicciones del ME y con una precisión comparable a la predicción teórica. También se ha desarrollado la metodología para medir la sección eficaz de este proceso a una energía en centro de masas de 5.02 TeV en estados finales con un leptón, con el objetivo de mejorar la precisión existente de dicha medida y combinarla con la medida realizada en estados finales con dos leptones. Se espera reducir la incertidumbre total respecto al resultado anterior a la mitad.
El conjunto de datos tomado a 5.02 TeV también se ha utilizado para medir, por primera a esta energía en centro de masas, la sección eficaz de procesos de producción asociados a dos bosones vectoriales masivos (WW, WZ y ZZ). Los resultados obtenidos con compatibles con las predicciones del ME.

Se han llevado a cabo dos búsquedas de procesos más allá del ME relacionadas con el quark top, ambas utilizando el último procesado de datos del Run 2 del LHC a 13 TeV. La primera es la búsqueda da la partícula supersimétrica del quark top (stop) asumiendo que se desintegra en un quark top y un neutralino, en un espacio de fases en el que la diferencia de masa entre el stop y el neutralino es muy similar a la masa del quark top. Esto hace que sea muy difícil de distinguir el proceso de señal de la producción de pares top quark y sea necesaria una búsqueda dedicada, ya que otras búsquedas no tienen sensibilidad en esta región. Haciendo uso de una red neuronal parámetrica, se excluye por primera vez la región de búsqueda completa a un nivel de confianda del 95%.
También se ha realizado una búsqueda bajo el marco de la teoría efectiva de campos de procesos asociados a la producción de quarks top en estados finales con leptones. Se estudia el efecto de un total de 26 operadores simultáneamente en 6 procesos de señal diferentes. En todos los casos se obtienen resultados consistentes con las predicciones del ME y se mejoraron los intervalos de confianza que había hasta el momento.

Por último, se ha desarrollado una técnica para identificar muones utilizando un modelo de aprendizaje automático que mejora los resultados de la estrategia utilizada hasta la fecha, ofreciendo resultados más robustos frente al número de vértices que se producen en las colisiones, que ha incrementado en el Run 3.

Universidad de Oviedo

RESUMEN (en Inglés)

In this thesis, measurements of Standard Model (SM) processes and searches for processes beyond the SM related to the top quark in final states with leptons are presented using data from proton-proton collisions produced at the LHC and collected by the CMS experiment at CERN.

The first measurement of the cross section of the top-antitop quark pair production process has been performed with the initial data of the Run 3 of the LHC at the new centre-of-mass energy of 13.6 TeV, yielding a cross-section of $\sigma = 888 \pm 34$ (syst. + stat.) $\pm$ 20 (lumi) pb. This result is consistent with SM predictions and achieves a precision comparable to theoretical expectations. Additionally, a methodology has been developed for measuring the cross section of this process at a centre-of-mass energy of 5.02 TeV in semilepton final states, aiming to enhance the precision of this measurement and combine it with the result obtained in dilepton final states. The goal is to reduce the total uncertainty compared to the previous result by half. The 5.02 TeV dataset has also been used to measure, for the first time at this energy, the cross section of processes associated with two massive vector bosons (WW, WZ, and ZZ). The obtained results are in agreement with SM predictions.

Two searches for processes beyond the SM related to the top quark are presented, both using the data collected during the Run 2 of the LHC at 13 TeV. The first is the search for the supersymmetric top quark (stop) assuming it decays into a top quark and a neutralino, in a phase space where the mass difference between the stop and the neutralino is very similar to the top quark mass. This makes it challenging to distinguish the signal process from top quark pair production, needing a dedicated search as other searches do not have sensitivity in this region. Using a parametric neural network, the entire search region is excluded for first time at a 95% confidence level. Another search has been performed under the framework of the effective field theory for processes associated with top quark production in final states with leptons. The effect of a total of 26 operators is studied simultaneously in six different signal processes. In all cases, the results are consistent with SM predictions, and the confidence intervals are improved compared to previous results.

Finally, a technique for muon identification has been developed using a machine learning model that improves the strategy used to date, providing more robust results against the number of vertices produced in collisions, which has increased in Run 3.

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

Agradecimientos

Es curioso que los agradecimientos siempre aparezcan al principio de las tesis, pero la mayoría de las veces es lo último que se escribe. En mi caso, así ha sido, y no porque no estuviera agredecida, sino porque esto marca el final de una etapa muy bonita y da vértigo dar los últimos pasos para cerrarla.

Me siento muy afortunada por haber podido formar parte del mejor grupo de investigación que me podría imaginar. Durante estos cuatro años he tenido la suerte de compartir despacho, pasillo y galletas con personas increíbles de las que he podido aprender y que me han ayudado a crecer, tanto personal como profesionalmente. Quiero comenzar agradeciéndole a Javi su confianza puesta en mí desde el principio, probablemente es uno de los culpables de que yo me embarcara en esta aventura ya que sus clases durante la carrera fueron la motivación principal para unirme a este grupo durante el TFG, y fue la mejor decisión que pude tomar. Con Santi y Bárbara he compartido varios análisis y de ellos he aprendido a cómo trabajar en grupo y comunicar tu trabajo al resto, superando cada revisión con elegancia y positividad. Agradecer, por supuesto, a Javier Cuevas por ser el mejor líder que un grupo puede tener, por darme la oportunidad de realizar el doctorado, por todo el tiempo que me ha dedicado, del cual he intentado absorber cada segundo de conocimiento y experiencia, y por su confianza y cercanía, así como por las lecciones al padelito (que espero seguir recibiendo). Por último, y quizás lo más importante, agradecer a mis directores, Kike e Isidro, por su orientación y consejos durante estos cuatro años, y por la paciencia que han tenido conmigo, sobre todo en este último mes en el que han peleado a mi lado para poder llegar a tiempo. Os agradezco enormemente el esfuerzo que habéis hecho. Quiero hacer mención especial a Kike, que ha sido la persona que me ha guiado y cuidado durante todo el doctorado. Ha sido un honor haber trabajo a tu lado y me parece imposible resumir en unas líneas todo lo que me has enseñado y has hecho por mí, de todo corazón, gracias.

El grupo no sería lo mismo sin sus estudiantes de doctorado, de hecho, la otra razón por la que yo acabé aquí lo era cuando lo conocí: Xuan. El entusiasmo con el que dabas las clases de FAEA hicieron que acabara de despertar en mí el interés por la física de partículas (y que se me acabara la tinta del boli...). Tuve la suerte de trabajar contigo desde el comienzo hasta casi casi el final, análisis tras análisis, código tras código, desde PAF hasta cafea, ¡tú mira si trabajo a gusto contigo que hasta te sigo a la empresa! Muchas gracias por todo lo que te has preocupado por mí y la paciencia que has tenido conmigo desde el principio, te has convertido en un amigo. Luego está el día cómico, Sergio y Carlos, que siempre están ahí cuándo los recesitas, tanto para hacerte reír cuándo estás desesperada como para resolverte el problema más enrevesado que se te ocurra de CMSSW. Espero recibir unas galletas cuándo ganéis el

Premio Nobel. Gracias a Pietro (sí, te he metido aquí para porque eres un jovenzuelo más) y Víctor por ayudarme siempre que lo he necesitado, sobre todo en estos últimos meses que han sido clave.

Clara, parece que fue ayer cuándo nos tocó juntas en las prímeras prácticas de la carrera y yo no fui a la primera clase. Ahí debiste de pensar "con qué desastre de persona me ha debido de tocar", ahora casi 10 años después ya has tenido tiempo para confirmar que estabas en lo cierto. Y es que yo siempre lo digo, no estaría aquí si no fuera por ti. Estoy muy agradecida por haber compartido esta etapa, de principio a fin, contigo. Hemos vivido hasta un confinamiento teletrabajando... Lo bueno es que justo después aparecieron los bros como un soplo de aire fresco: Alejandro y Carlos. Costó un poco enderezaros a cojinazo points, ¡pero mirad a dónde habéis llegado! ¡Que habéis sobrevivido dos meses solitos en Suiza! Espero poder haberos transmitido algo de mi experiencia durante este tiempo, aparte de mi buen gusto musical ;) Dejando las bromas a un lado, me alegro un montón de haber podido compartir despacho y aventuras con vosotros, vais a llegar muy lejos y yo espero estar ahí para verlo. Alejandro, espero que no me eches mucho de menos en las comidas, que cuides bien del cactus, y ya sabes, te quedas como responsable de mantener el buen legado paramés. Luego llegó Pelayo, que se pasa la vida en modo pro, atreviéndose a montar en coche conmigo el primer día que nos conocimos y casi morimos en el intento. Eres un crack, pero a ver a quién le vas a robar los chicles ahora. Por último, desearle mucha suerte a los Javis en esta nueva etapa que están comenzando, aprovechadla al máximo que se pasa más rápido de lo que creéis.

Aquí termina la mía. Han sido unos años muy intensos, pero muy bonitos. He estado muy a gusto y yo firmaba quedarme aquí haciendo esto para siempre, tanto investigar como docencia, ambas cosas me encantan. El problema es que para acabar aquí ahora tocaría irse unos años fuera de España, y eso no está hecho para mí. Por eso, aunque me ha costado muchísimo tomar la decisión, creo que es el momento de cerrar una puerta y abrir otra, y ya veremos que viene después.

Otra de las cosas buenas que me ha dado el doctorado es haber podido viajar y conocer a personas maravillosas, que ahora considero amigas. Agradecer al equipo IFCA/IFIC por todos los buenos momentos, en especial a Celia, Gorka, Pablo y Agustín. Con Celia he tenido la suerte de trabajar y aprender un montón de ella, pero también de compartir momentos duros y risas, siempre ahí. Gorka, Pablo y Agustín son la positividad personificada, siempre con una sonrisa en la cara dispuestos a hacer cualquier cosa, ¿para cuándo otro Negrita Music Fest o descenso del Sella? Me alegro un montón de haberos conocido y espero que sigamos haciendo esas quedadas Santander-Oviedo-Valencia muchos años más.

Gracias también a todas las personas que compartieron conmigo las estancias en Suiza: Dalia, Won, Adrián, Marina y... creo que se me olvida alguien... ¡Pedro! A veces la vida da segundas oportunidades, y creo que a nosotros nos volvió a cruzar los caminos

porque estábamos destinados a ser amigos (aunque seas un rancio y no lo admitas). Muchas gracias por haber cuidado de mí (y seguir haciéndolo), pero sobre todo gracias por todos los momentos felices que pasamos, ya fuera en el CERN, St. Genis, Ferney o en el Tapas Bar. Al final he tomado el camino más cobarde, pero de mí no te libras eh, ¡voy a verte a Villacarriedo si hace falta! Y el viaje a Creta sigue pendiente ;)

Let me switch to English to thank the brilliant people from CERN I was lucky to work with. A special mention goes to Francesca, thank you for your patience with me and for all that you have taught me. I will never forget: better an egg today than a chicken tomorrow!

Quiero hacer mención especial a dos personas que me cambiaron la vida sin saberlo. La primera es José Luis, que con solo 12 años creyó en esa "pequeña flor" y me dio la fuerza y confianza para llegar a ser lo que soy ahora. Hacen falta más profesores como tú, muchas gracias por todo. La segunda es Lara, que apareció a dar una charla sobre Ciencia de Datos en el momento en el que yo más perdida estaba y me descubrió un mundo desconocido del que ahora formo parte. Tuve la suerte de ser su alumna y me parece una mujer admirable, tanto personal como profesionalmente. Me hace muchísima ilusión que hayas aceptado estar en mi tribunal.

Muchas gracias a mi familia, por siempre creer en mí y proveerme de tapers y chori sin importar la distancia. Y mi mayor GRACIAS va para todos los amigos que me han apoyado (y aguantado) incondicionalmente siempre. Gracias Álvaro, Antonio, David, Estellica y María por haberos cruzado en mi vida hace ya casi 10 años y haberos convertido en una parte esencial de ella. Gracias Cris y María por haberme aguantado en ese año tan complicado, haciéndome sacar fuerzas a base de gatetes, pelis y palomitas. Cris, eres la mejor y lo sabes, fuiste como un ángel caído del cielo y me siento muy orgullosa de tenerte como amiga. Y como no, muchas gracias a los de siempre, a los que nunca fallan, a todos esos amigos de Zotes que me llevan aguantando toda la vida. En especial a Javi, porque salir de fiesta de vez en cuando es necesario, a Ane, porque pese a la distancia siempre hemos sabido que lo nuestro era inseparable, y a Sandra, por haber compartido la vida entera conmigo.

Ya solo me queda una personita, y es que es un poco especial. Muchas gracias Zurdy por, de una manera u otra, estar siempre ahí para mí cuando lo necesito. Eres la única persona que sabe decirme lo que necesito oír en cada momento y que me conoce más que yo a mí misma. Gracias por creer en mí incluso cuando yo no lo hacía, no creo que hubiera llegado aquí sin ti. Simplemente, gracias.

Para finalizar, gracias Viva Suecia, La La Love You, Shinova, Veintiuno, Nil Moliner y Dani Fernández por ponerle banda sonora a la escritura de esta tesis, porque...
"Haz que merezca la pena", "solo es necesaria una razón para cambiar el mundo", "que si la vida te da de más, comparte todo por si un día te vas", "que los mejores momentos son los que están por llegar".

Contents

List of Figures

List of Tables

Introduction

The Standard Model (SM) is the most successful theory of particle physics. It accurately describes all currently known elementary particles as well as the different interactions they have and the particles they give rise to. Even so, it is not a perfect model, it is not able to explain phenomena such as gravity and it fails to answer some questions that are still missing such as: What is dark matter made of? Why is there more matter than antimatter? Why do neutrinos have non-zero mass? Therefore, it is important to continue research using new techniques to see what lies beyond the SM (BSM).

In order to prove experimentally the existence of the SM particles and to know their properties, as well as to search for new physics, machines that accelerate particles to high energies to make them collide and give rise to new ones have been developed. The behavior of these particles is studied using complex detectors and sophisticated computational tools. The Large Hadron Collider (LHC) at CERN is the largest particle accelerator in the world. It is a circular accelerator that has achieved the most energetic proton-proton (pp) collisions up to date, producing a large number of new particles in each collision that allow us to study a multitude of processes thanks to the detectors placed surrounding the collision points. The LHC was built with two main goals: improving the precision of the fundamental parameters described by the SM and observing experimentally physics processes not observed so far.

We are living in what is commonly referred to as the "data age". Every passing second witnesses the generation and collection of an enormous volume of diverse data types, including data collected by the LHC detectors, and it is in our hands to take advantage of them. In recent years, a novel scientific discipline, known as data science, has emerged. It is based on the application of Machine Learning (ML) and Deep Learning techniques, such as Deep Neural Networks (DNN), to solve problems that would otherwise not be possible. These techniques allow us to analyze millions of data and train models that learn to "think" on their own. This remarkable progress has given rise to the Artificial Intelligence (AI) that is now so fashionable and that helps us in our daily lives, although we often do not realize. Notable examples include the Google search engine and personalized algorithms in social media platforms. These

AI techniques find applicability across a wide spectrum of disciplines, particle physics among them.

The results presented in this thesis are based on data taken by the Compact Muon Solenoid (CMS), one of the four large detectors of the LHC. Data from various periods of the LHC operation at different centre-of-mass energies ($\sqrt{s}$) are analyzed for cross section measurements and BSM searches.

The Run 3 of the LHC started in 2022 and the first pp collisions at $\sqrt{s} = 13.6$ TeV were produced. I have taken part of the first analysis at this unexplored centre-of-mass energy using a small amount of data: the top quark pair production ($t\bar{t}$) cross section measurement [1], which is the unique article published by CMS with Run 3 data up to date.

One of the main topics of this thesis is the measurement of the cross section of different processes at $\sqrt{s} = 5.02$ TeV using a small special dataset collected during 2017. This dataset was taken as a reference for heavy-ion collisions at that energy, but it is also useful to measure SM parameters. The $t\bar{t}$ inclusive cross section is first measured in dilepton final states and, some time later, in final states with one lepton in order to combine both results and to improve the precision. I have also contributed to the first analysis of diboson (VV) production using data from pp collisions at this centre-of-mass energy. These measurements are described in References [2, 3].

Once a parameter of the SM has been precisely measured, and thanks to the knowledge acquired, another main topic of this thesis is to prove the existence of BSM physics, which is another reason for the existence of the LHC. Having studied the previously described SM processes, I am in an ideal position to perform searches related to top quarks. In this thesis two BSM searches are presented related to the top quark production. One within the framework of Supersymmetry (SUSY) and another within the framework of Effective Field Theory (EFT), both of them using data collected during the Run 2 of the LHC at $\sqrt{s} = 13$ TeV. SUSY is a theory that predicts the existence of a new particle for each of the known SM particles. I have developed a search of the supersymmetric partner of the top quark, the stop quark, in a phase space where the mass difference between stop quark and neutralino is close to the top quark mass [4]. In this region, stop quark pair production and top quark pair production have similar kinematics and other searches do not have sensitivity. The main strategy of this search consists of the use of a DNN to separate the signal from $t\bar{t}$. On the other hand, I have also been involved in a search for new physics in top quark production with additional leptons in the final state within the context of the EFT framework [5]. This theory parameterizes potential new physics effects in terms of 26 dimension-six EFT operators. Kinematic variables are used in a simultaneous fit of the 26 EFT parameters to the data in order to observe the signal.

During the thesis I have contributed to the CMS muon group doing different activities at detector level and performance studies to improve the identification and isolation of the muons. In particular, making use of my expertise in muons, I have developed a new method to identify muons more efficiently than the previous one using a multivariate (MVA) technique [6]. This MVA is one of the identification tools to be used during the Run 3 and will be crucial for the analyses to select suitable muons.

The thesis is organized as follows. Chapter 1 includes a description of the SM, its limitations and its possible extensions in which BSM physics might be discovered. Chapter 2 provides details of the LHC, the CMS experiment and its different subdetectors, and a brief introduction of how the different particles are reconstructed in CMS. In Chapter 3, the data samples and the MC simulation used during the thesis are described. In Chapter 4, results of the first measurement of the top quark pair production cross section at $\sqrt{s} = 13.6$ TeV, together with the diboson production and top quark pair production cross sections at $\sqrt{s} = 5.02$ TeV are provided. Chapter 5 includes two BSM searches in the context of SUSY and EFT. A new technique that I developed to identify muons is described in Chapter 6. Finally, the conclusions are presented in the last chapter.

A summary of the published articles resulted from this thesis and the other main activities done during this period is presented in Figure 1.

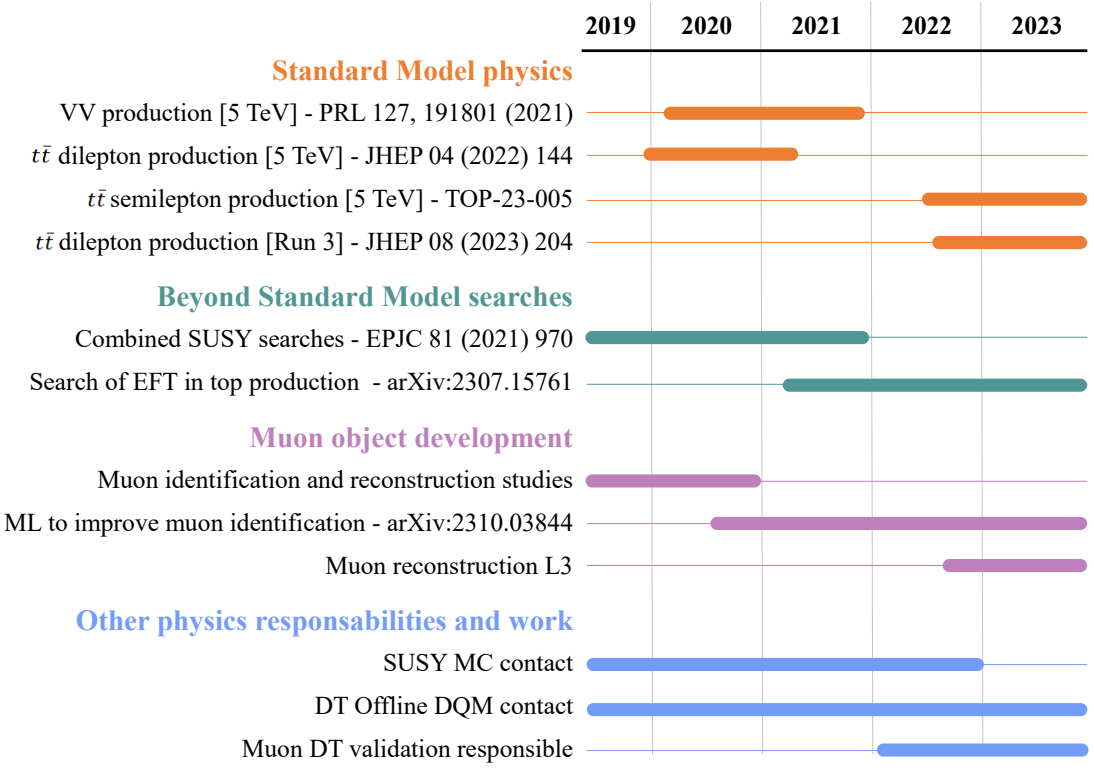

FIGURE 1: Timeline of the main analyses and tasks performed during the thesis. The time range in the articles goes from the beginning of the analysis to its publication.

Chapter 1

The Standard Model of particle physics and beyond

The SM of particle physics, established in the 1960s, stands as one of the most successful theories in the field of modern physics, providing a comprehensive framework to describe the fundamental particles and their interactions. However, some observations and theoretical considerations indicate the existence of phenomena BSM, which call for an extension of our current understanding. This chapter provides an introduction to the SM of particle physics and explore the motivations and implications of theories that go beyond its boundaries.

1.1 The Standard Model of particle physics

The SM is a Quantum Field Theory (QFT) that describes the fundamental particles that compose the matter and their interactions. QFT, which has been shown to accurately describe most of the particle physics phenomena ever observed, treats elementary particles as excited states of quantum fields and provides a framework for calculating their behavior. This framework includes not only the postulates of quantum mechanics, but also the ones from special relativity. It is a renormalizable and gauge invariant QFT, meaning it possesses symmetries, called gauge symmetries, which determine the interactions between particles, which are also represented by fields in this QFT. More complete descriptions of both SM and QFT can be found in References [7] and [8].

The elementary particles in the SM are categorized into two main classes, according to their spin: fermions (spin 1/2) and bosons (spin 0 or 1). A scheme of the particles that compose the SM is shown in Figure 1.1.

- **Fermions** are the constituents of matter and are divided into leptons (ℓ) and quarks (q), each of them split into 3 generations with identical quantum number

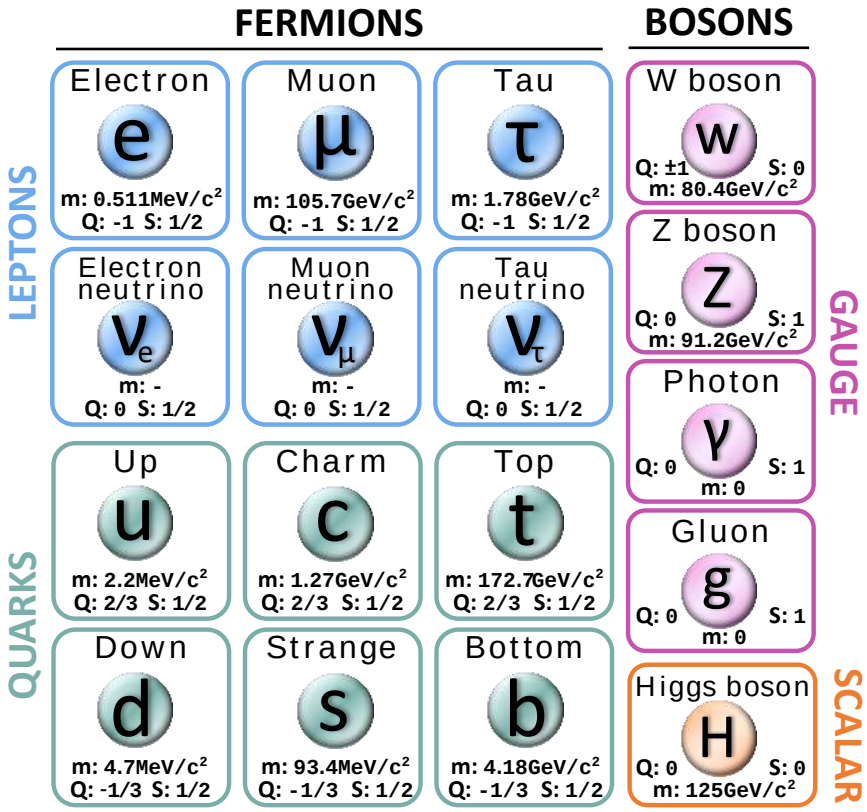

FIGURE 1.1: Elementary particles of the SM and their properties.

but with different masses. There are three electrically-charged leptons: electron (e), muon (μ) and tau (τ), with their corresponding neutrinos (ν_e, ν_μ, ν_τ). Leptons interact through the electromagnetic and the weak forces, while quarks also interact through the strong force given that they are colour-charged. There are six quarks: up (u), charm (c), and top (t), with electric charge of +2/3; and down (d), strange (s), and bottom (b), with electric charge of -1/2. Quarks are observed combined forming hadrons. Each fermion has a corresponding antiparticle that is an identical particle with the same mass, but with opposite electric charge.

- **Bosons** are divided in gauge bosons (spin 1) or scalar bosons (spin 0). Gauge bosons are the interaction mediators: photons (γ) are the mediator particles of the electromagnetic force, $W^{\pm}$ and Z bosons are the mediator particles of the weak force, and gluons (g) are the mediator particles of the strong force. Photons and gluons are massless particles, while $W^{\pm}$ and Z bosons are massive. Finally, the Higgs boson is the scalar boson responsible for giving mass to the fermions and the weak bosons.

The SM describes three of the four fundamental interactions: electromagnetic, weak, and strong. The first two can be grouped into a single one, known as electroweak (EWK) interaction. The gravity interaction is not described by the mathematical formulation of the SM up to date.

The SM interactions in QFT are defined through a Lagrangian density, $\mathcal{L}(\psi_i, \partial_\mu \psi_i, x^\mu)$, where ψ are the fields, $\partial_\mu \psi$ their derivatives, and x^μ the spacetime coordinates. Fermion fields are associated to spinors while gauge bosons fields are associated to vectors. The Higgs boson is represented by a scalar field. All fields are representations of the Lorentz group.

Furthermore, each SM interaction is represented by a different local gauge symmetry of the group $SU(3)_C \times SU(2)_L \times U(1)_Y$, where C is the colour charge, L is the weak isospin, and Y the weak hypercharge. The subsections 1.1.1 and 1.1.2 explain with more detail the EWK and strong interactions, respectively.

1.1.1 Electroweak interaction

The EWK interaction is a fundamental force that unifies the electromagnetic force, responsible for interactions between charged particles, and the weak force, responsible for processes like radioactive decays, under the $SU(2)_L \times U(1)_Y$ gauge group. It was introduced by Sheldon Glashow, Abdus Salam and Steven Weinberg in 1979 [9, 10].

In the context of the EWK interaction, fermions play a crucial role as the elementary particles that interact through this force. The EWK interaction involves the left-handed fermions ($\psi_L = \frac{1-\gamma^5}{2}\psi$), which are defined by their chirality in doublets, and the right-handed fermions ($\psi_R = \frac{1+\gamma^5}{2}\psi$), which are singlets. A generic representation of each component for leptons is shown in Equation 1.1, and an example of the representation for one quark generation is shown in Equation 1.2:

$$L = \begin{pmatrix} \ell \\ \nu_\ell \end{pmatrix}_L, \quad \ell_R, \tag{1.1}$$

$$L = \begin{pmatrix} u \\ d \end{pmatrix}_L, \quad u_R, \quad d_R. \tag{1.2}$$

The weak isospin third component (I_3) is zero for right-handed fermions and +1/2 (-1/2) for the upper (lower) of the left-handed doublets. Only the left-handed fermions interact through the weak force, as a consequence of the $SU(2)_L$ symmetry.

The interaction between fermions and the electroweak bosons is governed by the gauge couplings. These couplings describe the strength of the interaction between fermions and gauge bosons. In the electroweak theory, the strength of these couplings depends on the $SU(2)_L$ weak isospin, the $U(1)_Y$ weak hypercharge and the electroweak mixing angle. The weak hypercharge is related to the I_3 and the electric charge (Q_e) as: $Y = Q_e - I_3$.

This theory gives 4 massless electroweak bosons, but has been experimentally proven than W and Z bosons are massive, so the EWK symmetry would be broken. The solution is to introduce the masses by using the Higgs-Brout-Englert mechanism [11, 12], that covers the Spontaneous Symmetry Breaking (SSB) of the EWK symmetry. It consists on introducing an additional complex scalar Higgs field doublet which gives masses to the electroweak gauge bosons when it interacts with them. Photons are massless as they do not couple directly to the Higgs field, so the electromagnetic force remains unaffected.

1.1.2 Strong interaction

The quantum chromodynamics (QCD) is the theory that describes the strong force. It is a fundamental component of the SM of particle physics that explains the interactions between quarks and gluons [13].

QCD is a gauge theory symmetric under the $SU(3)_C$ group. Its associated conserved charge is the colour charge C, that can be blue, green or red. Only quarks and gluons have colour charge, therefore they are the only ones that interact through the strong force. Quarks are represented as colour triplets, whereas leptons are singlets.

There are 8 gluons, associated to 8 independent generators, the Gell-Mann matrices. Gluons are massless because they do not interact with the Higgs field and, therefore, they do not get mass from the SSB of the electroweak symmetry.

The strength of its coupling constant (α_s) depends on the energy scale of the interaction (Δ_{QCD}) and the momentum transferred (Q) involved:

$$\alpha_s(Q^2) \sim \frac{1}{ln(Q^2/\Delta^2_{QCD})}. \tag{1.3}$$

This dependence means that the strong interaction is stronger at low energy scales, with the consequence that quarks and gluons are not observed as free particles, they form colour neutral states called hadrons. These states are mainly observed in groups of three quarks (baryons) or as quark-antiquark pairs (mesons). The process by which quarks and gluons form hadrons is called hadronization. The top quark is the exception to this due to its high mass, it decays before hadronizing. A detailed description of the top quark and the physics surrounding it is presented in Subsection 4.1.1.

1.2 Limitations of the SM

The SM of particle physics has been remarkably successful in describing the fundamental particles and their interactions. However, it is not a complete theory and has

several limitations that point to the existence of physics beyond its predictions. Exploring these limitations and the open questions mentioned in the introduction is crucial for advancing our understanding of the fundamental nature of the universe. Some examples are listed bellow:

- **Gravitational interaction and unification of forces:** The SM does not incorporate gravity, which is described by Einstein's general theory of relativity. While electromagnetic, weak, and strong forces, are successfully described within the framework of the SM and QFT, the absence of a quantum theory of gravity remains a major challenge. Combining general relativity with the QFT framework of the SM is an open problem in theoretical physics.

- **Hierarchy problem:** The SM has the hierarchy problem, which is the large discrepancy between the weak scale and the Planck scale. The weak scale is much smaller than the Planck scale, raising questions about the stability of mass hierarchies. Resolving the hierarchy problem requires a deeper understanding of the underlying mechanisms responsible for mass generation, the quantum corrections to the Higgs boson mass and the potential existence of new physics at higher energy scales.

- **Neutrino properties:** Experimental observations have established that neutrinos have mass and undergo flavour oscillations. However, the SM assumes neutrinos are massless and does not account for their masses or mixing. The discovery of neutrino masses indicates the need for extensions to the SM, such as the inclusion of neutrino mass terms and new interactions.

- **Dark matter (DM) and dark energy:** Observational evidence supports the existence of DM, which constitutes a significant portion of the universe's mass. However, the particles that make up DM are not accounted for in the SM. The nature of DM remains a mystery, requiring new particles and theories BSM. Additionally, the accelerated expansion of the universe is attributed to dark energy, but the origin and nature of dark energy are not explained within the SM.

- **Matter-Antimatter Asymmetry:** The observed universe is predominantly composed of matter, while antimatter is relatively rare. The SM does not provide a satisfactory explanation for this matter-antimatter asymmetry. Understanding the origin of this imbalance requires new mechanisms that go BSM, such as theories involving violations of charge-parity (CP) symmetry or baryogenesis processes.

1.3 Physics Beyond the SM

Physics Beyond the SM refers to theoretical frameworks and extensions that aim to address the limitations and open questions in the SM of particle physics. This thesis explores two of these proposals: SUSY and EFT.

1.3.1 Supersymmetry

Supersymmetry, or SUSY, [14–21] is one of the most popular BSM theories that solves many of the open questions in the SM, such as the DM and the hierarchy problem, although no experimental evidence has yet been found. It extends the SM by introducing a symmetry between fermions and bosons. This symmetry is manifested through the introduction of supercharges and the construction of supermultiplets, which combine fermionic and bosonic states.

The simplest SUSY version, called the Minimal Supersymmetric SM (MSSM), for each known particle assigns a new SUSY partner or superpartner that differs by a half unit of spin. Gauge bosons are associated to spin 1/2 particles, named gauginos, and fermions are associated to scalar particles, named sfermions (sleptons and squarks). In addition, MSSM theory introduces two Higgs doublets which give rise to five physical Higgs bosons: two neutral scalar particles (h^0 and H^0), one neutral pseudoscalar particle (A^0), and a pair of charged Higgs bosons ($H^\pm$). Their supersymmetric partners are called higgsinos. Gauginos and higgsinos are mixed up together to form mass eigenstates: four neutral and four charged fermions, called neutralinos and charginos.

The convention to denote sfermions is adding a "s" before the SM name of their partners or adding a tilde on top of their particle symbol. For example, the stop quark is the SUSY partner of the top quark and is denoted as $\tilde{t}_1$. The partners of the SM bosons are named by adding an "ino" suffix to their name and also with the tilde on top. Thus, the gluino is the SUSY partner of the gluon and is denoted as $\tilde{g}$. Finally, neutralinos and charginos are denoted by $\tilde{\chi}^0_{1,2,3,4}$ and $\tilde{\chi}^\pm_{1,2}$ respecively, ordered by ascending masses.

The SM particles and their superpartners have the same quantum numbers, so to distinguish them a new quantum number called R-parity is introduced which is +1 for SM particles and -1 for SUSY particles. This number must be conserved in the interactions, therefore SUSY particles can only be produced in pairs and the lightest SUSY particle (LSP) has to be stable, since its decay is not kinematically allowed. The LSP is a viable DM candidate as it is potentially massive, stable and electrically neutral.

In this thesis, a search for stop quark pair production and neutralinos ($\tilde{\chi}_1^0$) in a parameter space where the kinematical properties of stop quark pair and top quark pair productions are very similar is presented in Section 5.1.

1.3.2 Effective Field Theory

There is no a priori reason to assume that particles will have a sufficiently low mass to be directly produced and detected at the LHC. Consequently, indirect methods play a crucial role in exploring higher energy scales. These methods aim to discover particles through their off-shell effects. The centre-of-mass energy for collisions at the LHC will not significantly increase throughout its remaining years of operation, so indirect approaches may provide an exciting opportunity to extend the discovery reach of the LHC.

Effective Field Theory, or EFT, [22–25] provides a flexible framework that encompasses a comprehensive description of the off-shell effects of the new physics phenomena at a higher mass scale, denoted as Λ. In this framework, the SM Lagrangian is considered as the lowest order term in an expansion of a more complete Lagrangian at a scale Λ. The expanded Lagrangian comprises a series of higher-dimensional operators, constructed from products of SM fields that respect the symmetries of the SM.

The EFT Lagrangian is expressed as:

$$\mathcal{L}_{\text{EFT}} = \mathcal{L}_{\text{SM}} + \sum_{d,i} \frac{c_i^d}{\Lambda^{d-4}} \mathcal{O}_i^d, \tag{1.4}$$

where $\mathcal{L}_{\text{SM}}$ is the SM Lagrangian, $\mathcal{O}_i^d$ are the EFT operators of dimension d that describe new physics interactions at a mass scale Λ, and c_i^d are the Wilson Coefficients (WCs) which control the strength of the EFT effects. Since each order in the expansion is scaled by an additional power of Λ, the terms in the lowest orders are expected to produce the largest expected deviations from the SM ($d \leq 4$) processes.

This thesis focuses on the top quark sector, specifically on operators that couple the top quark to leptons, bosons, and other heavy (top or bottom) quarks. The dominant SM contributions to these signatures arise from processes in which one or more top quarks are produced in association with a heavy boson or other top quarks. While each of these processes have been studied individually, the analysis presented in this thesis takes a more global approach, using the EFT framework to prove the potential effects of heavy new physics impacting these associated top quark processes simultaneously in leptonic final-states (multileptons)[5]. More details of this analysis can be found in Section 5.2.

Chapter 2

The LHC and the CMS experiment

All the analyses presented in this thesis have been performed with data obtained from pp collisions at the LHC and collected by the CMS experiment. The following sections provide an overview of both the LHC and CMS, as well as key concepts relevant to the subsequent chapters of this thesis. First, Section 2.1 offers a brief description of the LHC, including its fundamental operational parameters. Section 2.2 presents an overview of the CMS experiment, detailing its constituent subdetectors and the trigger system. Finally, Section 2.3 describes the techniques used in the reconstruction of the different physical observables.

2.1 The Large Hadron Collider

The Large Hadron Collider, or LHC [26], is the world's largest and most powerful circular particle accelerator. It was built by the European Centre of Nuclear Research (CERN) between 1998 and 2008 and is located on the French-Swiss frontier, close to Geneva, in a 27 km circumference tunnel, 100 m underground. In the LHC, two high-energy particle beams traveling at near-light speed in opposite directions are collided at a centre-of-mass energy up to 13.6 TeV. The most common collisions are pp collisions, but it also produces heavy-ion collisions like lead-proton or lead-lead collisions. The collisions take place at four different points along the LHC circunference in which the main experiments are located: A Toroidal LHC Apparatus (ATLAS), CMS, A Large Ion Collider Experiment (ALICE), and Large Hadron Collider beauty (LHCb) experiments.

The LHC is the last step of the CERN accelerator chain, shown in Figure 2.1. First, protons are extracted from hydrogen gas by applying an electric field. These protons are accelerated up to 50 MeV into the Linear Accelerator 2. The beam is then split into four beams that are accelerated and separeted in different bunches of 10^{11} protons in the Proton Synchrotron Booster, a circular accelerator of about 157 m, which makes

the protons reach energies of 1.4 GeV. These beams are recombined in two and passed to the Proton Synchrotron, where they reach an energy of 25 GeV. They are then injected into the Super Proton Synchrotron, a 7 km circular accelerator that accelerates the proton bunches to 450 GeV. Finally, these two beams are injected into the LHC ring with a bunch separation of 25 ns guided to rotate in two separate pipes in opposite directions. This rotation is facilitated by superconducting magnets, which generate a magnetic field of up to 8.3 T at temperatures below 1.9 K, using a liquid helium distribution system. Inside the LHC, protons are accelerated through a radio frequency system up to the target colliding energy and then the two beams are deviated slightly such that collision of proton bunches or bunch crossings (BX) occur at the collision points, where the experiments are located. The maximum design centre-of-mass energy in pp collisions of the LHC is 14 TeV, but has not yet been reached.

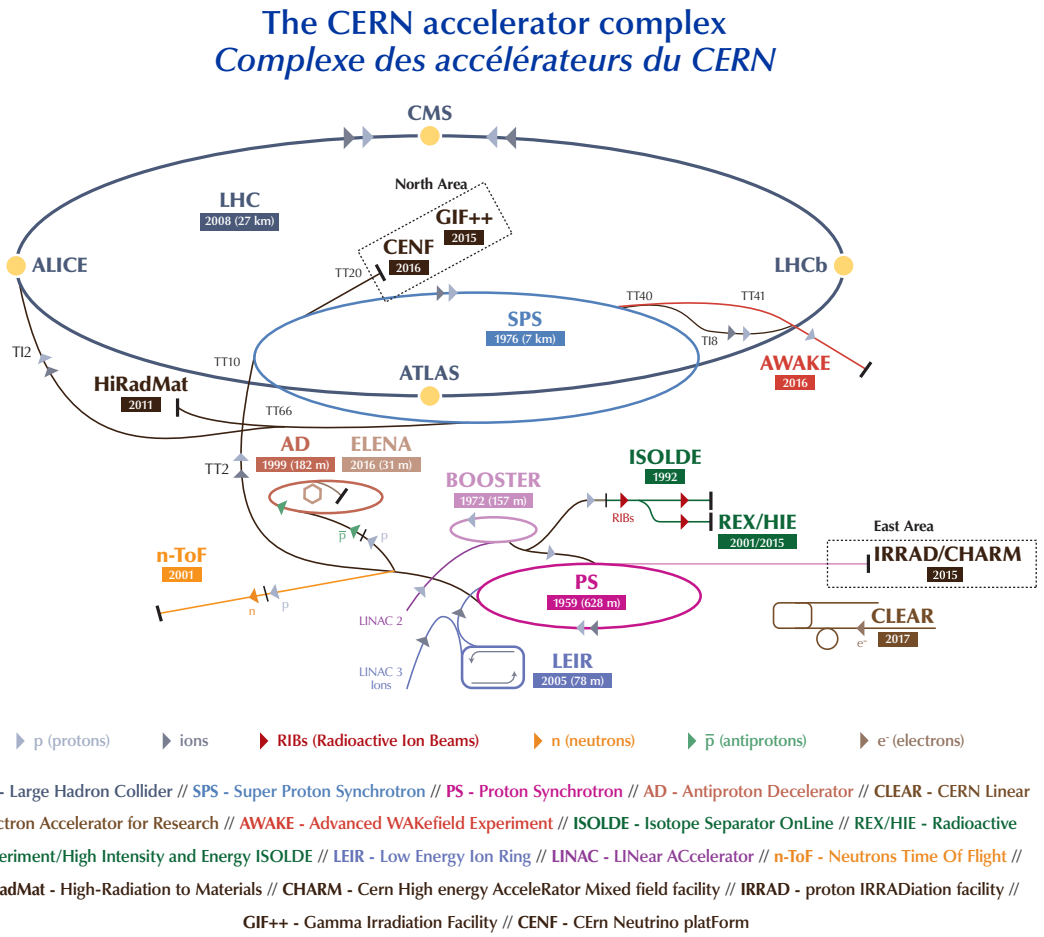

FIGURE 2.1: Scheme of the CERN accelerator chain. The LHC is the last ring (dark blue line). The four yellow dots indicate the collision points of the beams, where the 4 main detectors are located [27].

By design, on each BX more than one pp collision occur simultaneously, leading to the appearance of multiple interaction vertices, known as pileup or PU. Since the centre-of-mass collision energy has been increasing over the years, the PU interactions has

also increased.

The amount of data produced by the LHC is quantified with the instantaneous luminosity, L, defined as the number of pp collisions per unit of area and unit of time. This quantity is related with the probability of a certain physics process occurring, known as cross section (σ), and the differential event rate of that process ($\frac{dN}{dt}$), as:

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

The instantaneous luminosity only depends on the collider and the beam parameters and can be defined as:

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

where n_b is the number of colliding bunches (around 3000), N_b is the number of protons per bunch (around 10^{11}), f is the revolution frequency of the LHC, σ_x and σ_y are the bunches sizes along the transverse directions on collision, and R is a geometric factor to take into account the crossing angle of the colliding beams at the interaction point. Another representative quantity is the integral over time of the instantaneous luminosity, known as integrated luminosity, L_{int}, or just luminosity.

The first data-taking period, referred to as Run 1, started in 2010 with collisions at a centre-of-mass energy of 7 TeV and continued until the end of 2011, providing a delivered luminosity by the LHC over 6 fb^{-1}. In 2012, the beam energy was increased and $\sqrt{s} = 8$ TeV was reached, delivering an integrated luminosity around 23 fb^{-1}. Data collection was stopped in early 2013 and restarted again in 2015. The data-taking period between May 2015 and December 2018 is referred to as Run 2 and resulted in a total luminosity of 164 fb^{-1} delivered by the LHC in pp collisions at $\sqrt{s} = 13$ TeV. In 2015 and 2017, there were two additional short runs of pp collisions at $\sqrt{s} = 5.02$ TeV, delivering 29 and 334 pb^{-1} of data, respectively. These especial runs are motivated by the exploration of a new centre-of-mass energy that offers a low average of PU interactions and as a reference for the programmed heavy-ion collisions at that energy. In 2022, the Run 3 started with the first pp collisions at $\sqrt{s} = 13.6$ TeV, energy at which the LHC is currently operating. The accumulated luminosity evolution in time by the CMS experiment since 2015 is shown in Figure 2.2 (left). The delivered and recorded luminosity cumulative over all years is shown Figure 2.2 (right).

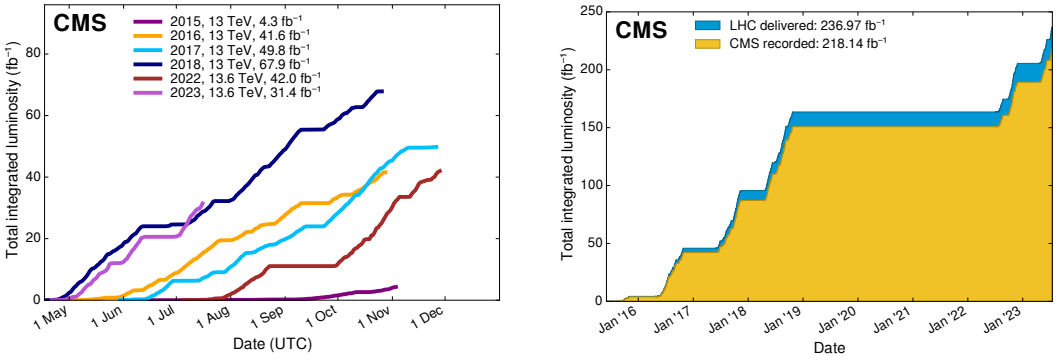

FIGURE 2.2: The cumulative integrated luminosity versus day delivered by CMS during stable beams for pp collisions at nominal centre-of-mass energy is shown on the left. This is shown for data-taking in 2015 (purple), 2016 (orange), 2017 (light blue), 2018 (navy blue), 2022 (brown), and 2023 (light purple). The cumulative luminosity delivered by the LHC (blue) and recorded by the CMS experiment (yellow) over all years is shown on the right. [28].

2.2 The Compact Muon Solenoid experiment

The CMS experiment [29] is located at one of the four points of the LHC ring where the beams collide, in an underground cavern near Cessy (France). It was designed with the general purpose of detecting a wide variety of processes produced at the energy scales of the LHC.

CMS is distributed in cylindrical layers around the LHC beam pipe with a length of 21.6 m and a diameter of 14.6 m. The whole structure is divided into two regions: the central part of the cylinder (barrel) and the bases of the cylinder (endcaps). It is composed of several subdetectors, each one designed to measure certain properties of different kind of particle, and a very strong superconducting solenoid magnet that creates a magnetic field of 3.8 T (100000 times the field of the earth). This large magnetic field is necessary to be able to bend charged particle trajectories and thus measure their momentum accurately through their curvature. More details of the CMS components can be found in the following subsections and a schematic view of them is represented in Figure 2.3.

In terms of its geometry, the coordinate system has its origin at the collision point, with the y-axis pointing vertically upward, the x-axis extending radially from the origin, and the z-axis aligned with the direction of the particle beams. The azimuth angle (ϕ) is measured from the x-axis in the x-y plane, which is perpendicular to the beam direction and is known as transverse plane; while the polar angle (θ) is measured from the z-axis in the x-z plane. Another significant angular parameter commonly used is the pseudorapidity (η). It is employed because differences in η remain invariant under Lorentz transformations along the z-axis for high-momentum particles. It is defined as a function of θ as:

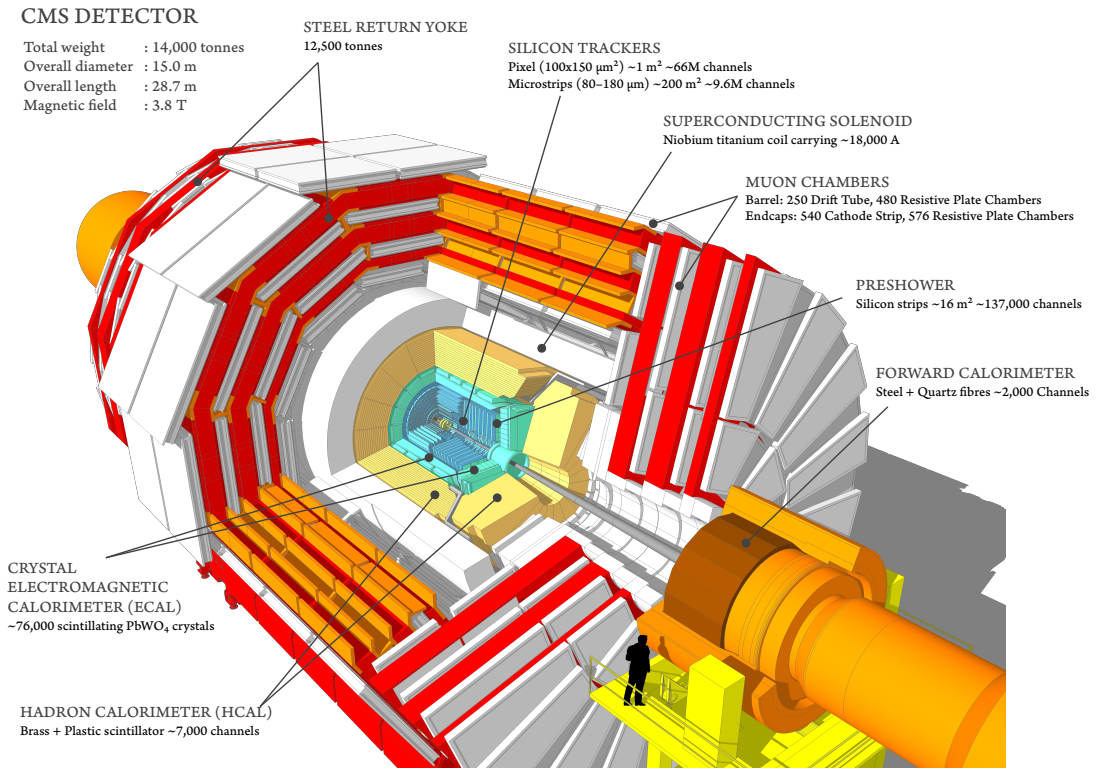

FIGURE 2.3: Schema of the CMS experiment and its subdetectors. [30].

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right). \tag{2.3}$$

Another widely used quantity is the projection of the particle momentum in the transverse plane, referred to as the transverse momentum or p_{T}, as there is no initial momentum in the transverse plane, so the transverse momentum of the final state particles must cancel one another out due to the momentum conservation law. Thus, the sum of the transverse momentum of all the particles produced in the collision should be 0, but if some of the particles produced in the collision are not detected, there would be a momentum imbalance that is known as missing transverse momentum and denoted by $\vec{p}_{\mathrm{T}}^{\,miss}$. The p_{T} is defined as:

$$p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2}. \tag{2.4}$$

where p_x and p_y are the particle momentum in the x and y axes, respectively. A summary of some of these quantities and the coordinate system of the CMS experiment are shown in Figure 2.4.

Once introduced the notation that is used during the thesis, in the following subsections a brief description of the subdetectors that form CMS is presented.

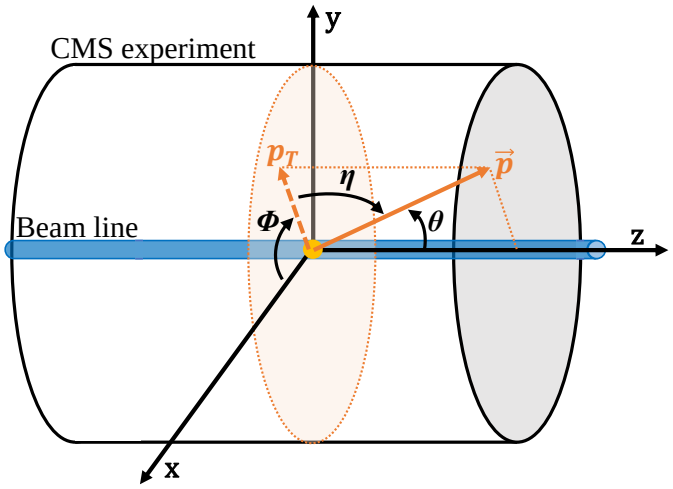

FIGURE 2.4: Coordinate system of the CMS experiment.

2.2.1 Tracker system

The tracker system [31] is the subdetector closest to the collision point, so it receives the largest volume of particles. Its function is to reconstruct accurately the trajectories (tracks) of charged particles and also the interaction vertices of the collision such as the primary vertex, the PU vertices, and the secondary vertices produced by the decay of large half-life particles, like the b or c quarks or τ leptons. The travel distance of these particles before decaying is known as impact parameter, and is well measured thanks to the tracker granularity.

The interaction of the particles with the tracker has to be very small in order not to disturb their trajectory. It is made entirely of silicon, a radiation-resistant material, with two types of detectors that can be distinguished according to their shape: the pixel detector and the silicon strip detector. Both of them are kept at very low temperatures so that they are not overheated by the energy of the particles they receive.

The pixel detector is in the innermost zone. It is made up of 124 million pixels distributed in 4 concentric cylinders in the barrel and 3 disks in each endcap (from 2017, in 2016 only 3 cylinders and 4 disks in total). Each pixel is 10 x 15 μm in size, allowing to have a excellent resolution to determine with high precision the 3D particle trajectories and the impact parameter.

The silicon strip detector is composed of a total of 9.8 million strips of different sizes in order to detect the trajectory of the particles. It is divided in 10 concentric cylindrical layers, four of them form the inner barrel region and the other six the outer region, and 12 disks in each endcap. The resolutions along the z axis are between 200 and 500 μm, while those of positions on the transverse plane are between 20 and 50 μm.

The whole tracker has a good acceptance up to a $|\eta| < 2.5$ thanks to the endcaps.

2.2.2 Electromagnetic calorimeter

The electromagnetic calorimeter (ECAL) [32] is a scintillator detector consisting of around 80000 very dense lead tungstenate crystals ($PbWO_4$), each weighting 1.5 kg. These crystals detect photons and electrons and produce light proportional to their energy. These flashes of light are fast, short and well defined, which makes the detector very accurate and fast. Each crystal is connected to a photodetector that converts the light signal into an amplified electrical signal that is sent for analysis.

The ECAL system is located after the tracker system and is divided as well in the barrel region, with a coverage up to $|\eta| < 1.48$, and the endcaps, which extend the coverage up to $|\eta| < 3$. The crystals have a transverse section of 22×22 mm^2 in the barrel and of 26×26 mm^2 in the endcaps, with a length of around 230 mm.

In addition, there is another subdetector in the endcaps, known as ECAL preshower, which increases the spatial resolution allowing to improve the photon identification.

2.2.3 Hadronic calorimeter

Sorrounding the ECAL in the barrel region is the hadronic calorimeter (HCAL) [33], designed to measure the energy of hadrons using alternating layers of absorbing materials (brass strips) and fluorescent scintillators that produce a pulse of light when the particle passes. This light is collected by special optical fibers that pass it to photodetectors to amplify the signal.

The HCAL is divided in four subregions, each finely tuned to function optimally in distinct parts of the detector with specific conditions, covering up to $|\eta| < 5.2$.

2.2.4 Solenoid magnet

The solenoid [34] consists of a coil of superconducting fibres that creates a magnetic field up to 3.8 T when electricity passes through it. It is the largest superconducting magnet ever built, with a diameter of 6 m and a weight of 12000 tons. In addition, its design allows the two calorimeters to fit perfectly inside the coil. The magnetic field makes it possible to measure the curvature of the charged particles, their transverse momentum, and the sign of their electric charge.

2.2.5 Muon system

The muon system [35] is the outermost part of the CMS experiment and practically only muons and uncharged particles, such as neutrinos, arrive there. It is designed

to measure the trajectory and the momentum of the muons, which are the only particles that can be detected there as they are the unique charged particles capable of penetrating several meters of calorimeters and the solenoid.

There are three types of muon system detectors: the drift tubes (DTs), the resistive plate chambers (RPCs) in the barrel and endcaps, and the cathode strip chambers (CSCs). All of them are gaseous detectors, but each one has different features. The distribution of the different subsystems is shown in Figure 2.5.

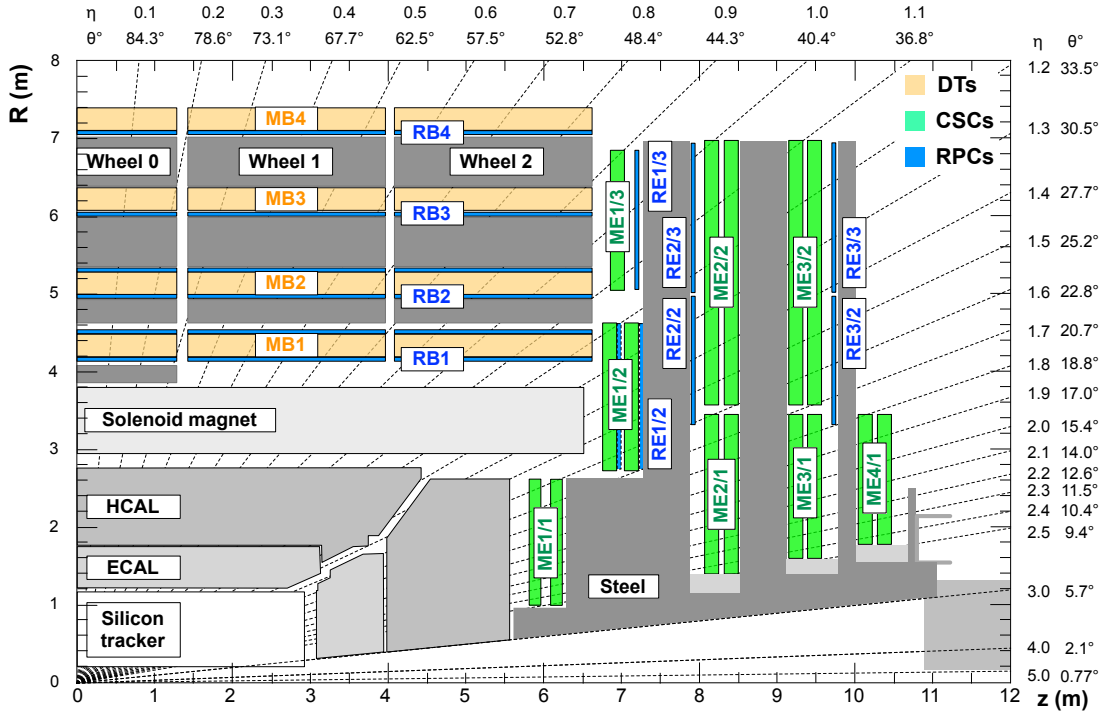

FIGURE 2.5: Layout of the CMS muon system together with the rest of subdetectors [36].

The drift tubes are located in the outermost part of the detector covering the barrel region ($|\eta| < 1.2$), where the muon rate is expected to be low. This subdetector is composed of 250 chambers organized in 5 wheels along the z axis. Each wheel is divided in 4 stations at the same radial distance in the x-y plane (r), which allow to measure the r-ϕ angle and the z direction of the muon trajectory. The chambers consists of groups of cells filled with gas, an anode wire and two electrode plates that create the drift electric field, so when a charged particle passes through the cell, it ionizes the atoms of the gas releasing electrons that are attracted to the anode, allowing to measure its drift time and to infer the position of the muon.

The CSCs are placed in the endcaps ($0.9 < |\eta| < 2.4$), where a larger contribution of neutron background is expected. There are 540 chambers organized in 4 disks or stations in each endcap. They are multiwire proportional chambers consisting of anode wires interleaved among several cathode strips within a gas. Wires and strips

are perpendicular, allowing to measure the radial and the ϕ component of the muon with high precision and resolution.

The RPCs cover the barrel and endcaps regions ($|\eta| < 1.9$) with 1056 gaseous parallel-plate detector that complement the DTs and CSCs. These chambers have a very good time resolution, which allows to measure the muon momentum very quickly so it can be used by the trigger system.

2.2.6 Trigger system

The trigger system plays a vital role in managing the vast amount of data generated during particle collisions. Its primary function is to select and record only the most interesting events for further analysis, as the LHC produces an enormous number of collisions, far more than can be processed in real time.

The CMS trigger system consists of two levels: the Level-1 Trigger (L1) and the High-Level Trigger (HLT). L1 is hardware-based and performs a rapid initial event selection based on simple criteria from the calorimeters and the muon system. It reduces the initial rate from 40 MHz to 100 kHZ using a very fast particle candidate reconstruction that decides in less than 4 μs if the event is stored or rejected. The selected events by the L1 triggers pass to the HLT. HLT is software-based and operates on a more detailed event data stream, applying complex reconstruction algorithms to make refined event selections with a typical output rate of around 1 kHz. It employs more sophisticated algorithms (paths) and extensive event reconstruction combining information from the whole detector to make precise selections based on physics criteria such as p_{T} or isolation of the physics objects. Events passing the HLT paths are recorded and stored for offline analyses.

The trigger system is highly adaptable and can be reconfigured to meet the specific needs of different physics analyses and data-taking conditions. Trigger algorithms are continuously optimized to ensure that important physics processes are captured while maintaining manageable data rates.

2.3 Event reconstruction in CMS

Once the data measured by the subdetectors have been collected, the different particles produced after the collisions can be identified according to their trajectory and its kinematic properties. This task is called event reconstruction and is performed by the Particle Flow (PF) algorithm [37] combining information from all the CMS subdetectors to provide a set of individual particle candidates corresponding to reconstructed muons, electrons, photons, charged hadrons and neutral hadrons.

The first step is the track reconstruction of charged particles using the inner part of the tracker. This process is done with an iterative algorithm joining the hits in the tracker to reconstruct the particle trajectory. Each iteration consists on three steps: initial seed generation with two or three compatible hits with the hypothetical particle trajectory, pattern recognition based on a combinatorial Kalman Filter method [38] including hits in the successive layers one by one, and finally, a fit of the predicted trajectory. In each iteration the complexity of the fitted tracks increases and the hits associated to those tracks are removed for the following iterations. This procedure is done in parallel for all the track candidates and is repeated until all the layers are explored or a stopping condition is satisfied.

Once the charged tracks are reconstructed, their iteration point with the beam axis can be extrapolated, known as vertex. The vertex with the highest quadratic sum of charged tracks p_{T} in each event is considered the primary interaction (primary vertex), while the rest appear due to pileup interactions, decays of short lived particles such as tau leptons, and the presence of long lived particles.

Then, another primary components used in the reconstruction are the clusters originated from both electronic and hadronic calorimeters. Clusters represent aggregations of energy depositions, each possessing energy and direction attributes. They play a crucial role in the processes of reconstructing and identifying electrons, photons, as well as neutral and charged hadrons.

The cluster reconstruction process begins with the derivation of seeds from individual cells within the calorimeters. These cells measure the energy deposited, and seeds are formed from cells with energy values exceeding those of their neighboring cells and surpassing a predefined threshold. Subsequently, topological clusters are constructed by grouping these seeds with adjacent energetic cells. Finally, an algorithm based on a Gaussian-mixture model is employed to determine the position and energy content of these clusters.

Tracks and clusters are associated geometrically through a linking algorithm, allowing the identification of the particles as each one interacts differently with each subdetector. Electrons are identified by inner tracks associated with clusters in the ECAL. Muons appear as inner tracks linked with tracks in the muon system. Photons and neutral hadrons only leave clusters in the ECAL and HCAL, respectively. Charged hadrons are identified by inner tracks associated with clusters in the calorimeters. A more detailed description of the reconstruction of these objects can be found in the following subsections.

2.3.1 Electron and photon reconstruction

Electron reconstruction at CMS combines tracks in the inner tracker with clusters in the ECAL [39]. The energy deposits in the ECAL are grouped in a supercluster (SC) which includes the principal electron and its radiated electrons and photons due to bremsstrahlung. On the other hand, radiation effects in the tracker are taken into account by reconstructing the electron trajectory with a Gaussian Sum Filtering (GSF) algorithm, resulting in GSF tracks. The SCs and the GSF tracks are associated to produce the PF electron candidates, that are the final reconstructed electrons.

Then, several quality criteria are applied to select well-reconstructed electrons and to reject possible objects misidentified as electrons. The proposal is to keep only prompt electrons coming from W, Z, Higgs bosons and τ lepton decays applying requirements in several variables such as isolation, the width of the reconstructed ECAL SC or $1/E - 1/\rho$, where E is the electron energy of the SC and ρ the momentum of the track. This observable tends to 0 for well-reconstructed electrons.

Photon candidates are reconstructed following the same ECAL clustering strategy, but without a GSF associated track since photons are neutral particles and do not leave hits in the tracker.

2.3.2 Muon reconstruction

Muon reconstruction combines information from the muon system and the tracker subdetector [36]. In an initial step, hits in the DTs and CSCs are matched to form muon segments, representing linear extrapolations of the muons trajectory based on the recorded hit positions.

There are three types of muons depending on the kind of reconstruction:

- **Standalone muons**: are muons reconstructed only with information from the muon system. Muon segments in the DTs and CSCs are combined as initial seed, and then compatible hits in all the muon subdetectors are added to the final trajectory (standalone track) using a Kalman Filter algorithm.
- **Global muons**: standalone tracks are extrapolated to the tracker and combined geometrically with a compatible inner track, resulting in a global track. The final global muon candidate is provided by the Kalman Filter method.
- **Tracker muons**: as opposed to global muons, tracker muons are reconstructed by extrapolating inner tracks up to the muon system and finding there a geometrically compatible muon segment or RPC hits.

Due to the way the algorithms are defined, a global muon can also be both a standalone muon and a tracker muon. When this happens, the three candidates are combined into one while maintaining references to the three distinct methods. About 99% of the muons are reconstructed as either global or tracker muons, and frequently as both. Tracker and global muons are also combined into one category if they have the same inner track.

Global muons have better momentum resolution at high p_T ($p_T > 200$ GeV) than the tracker muons, but for low p_T, tracker muons are more efficiently reconstructed as these muons may not leave enough hits in the muon chambers.

As in the electron reconstruction, the selection of high-quality muon candidates is crucial for the reliability of the reconstruction process. Criteria are applied to select candidates that meet specific requirements, such as a minimum number of hits, well-defined fit quality, and compatibility within tracks and segments. The muon isolation is also essential to reject misreconstructed muons such as charged hadrons that reach the muon system, or muons from heavy-flavor hadron decays.

2.3.3 Jets and b-jets tagging

Gluons and quarks, except the top quark, can not be observed as single states due to the strong interaction, as explained in Section 1.1.2. They appear as a collimated cluster of particles produced by hadronization in a certain direction, called jets.

The reconstruction of jets is performed by a specific clustering algorithm called anti-k_T [40]. This algorithm collects all the PF candidates and groups them in a cone-shaped jet with a given angular width $R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$, also known as distance parameter. This distance parameter is usually set to 0.4 and the resulting jets are named AK4 jets. If the charged PF particles coming from PU interactions are not considered, the reconstructed jets are known as pileup-per-particle identification (PUPPI) jets [41, 42]. The jet momentum is calculated as the sum of the momentum of its constituents.

Jets originated from the hadronization of b quarks (B hadrons) are known as **b jets** and, in contrast to the rest, they tend to have a larger impact parameter due to the long lifetime of B hadrons, and present a secondary vertex. This makes b jets distinguishable from the rest, which is very interesting because they appear in the decay of the top quarks.

There are several algorithms in CMS used to identify b jets, algorithms known as b tagging. These algorithms are typically based on MVA techniques that gives a probability as output. This probability evaluates whether the input jet originates from a b quark or not. Multiple sets of criteria, known as working points (WPs), are established to strike a balance between efficiency and purity. These categories, namely, loose,

medium, and tight, are chosen to maintain misidentification probabilities of 10%, 1%, and 0.1%, respectively. In the analyses developed in this thesis two algorithms are used: DeepCSV [43] and DeepJet [44].

2.3.4 Missing transverse momentum

The missing transverse momentum, as mentioned in Section 2.2, is the momentum imbalance produced in a pp collision in the transverse plane. This imbalance may be due to the presence of non-detected or invisible particles, but also may be due to misreconstructed particles, the presence of pileup interactions or instrumental effects affecting the resolution of the reconstruction of the other particles in the event. These effects are taken into account by applying different selections or filters to the data. It is defined as the negative sum of the p_T of all the reconstructed particles in the event:

$$\vec{p}_T^{\,miss} = \sum_i \vec{p}_T(i). \tag{2.5}$$

Chapter 3

Data samples and Monte Carlo simulation

Since data taking started, both the LHC and CMS have been evolving. Factors such as detector degradation, detector upgrades or changes in the conditions of the LHC makes data not uniform over time. Some of these changes, such as the increase in the centre-of-mass energy in the pp collisions, result in the data-taking being divided into different periods, known as Runs. All those effects are taken into account at analysis level with specific corrections and different simulations. Many scientists and experts are working to ensure that the detectors, the LHC, as well as the simulations, are ready for use. I am among them, having contributed to the CMS experiment over the last 4 years in different aspects described in Section 3.4.

3.1 Data-taking

In this thesis, data from three different periods of data-taking are used.

- **The Run 2** of the LHC, corresponding to pp collisions at $\sqrt{s} = 13\,\text{TeV}$, spans from 2016 to 2018, with a total recorded luminosity of $138\,\text{fb}^{-1}$. The pileup conditions changed even across the three years, as is shown in Figure 3.1, with a mean number of interactions per BX of 27, 38 and 39 for 2016, 2017 and 2018, respectively. This increase resulted in a higher integrated luminosity during 2017 and 2018.

  The radiation received by the different subdetectors over time may cause degradation on some of them. It is the case of the ECAL crystals at high $|\eta|$, which are progressively losing transparency leading to a shift in the ECAL timing. During 2016 and 2017 this shift was not propagated to the L1 calibration resulting in

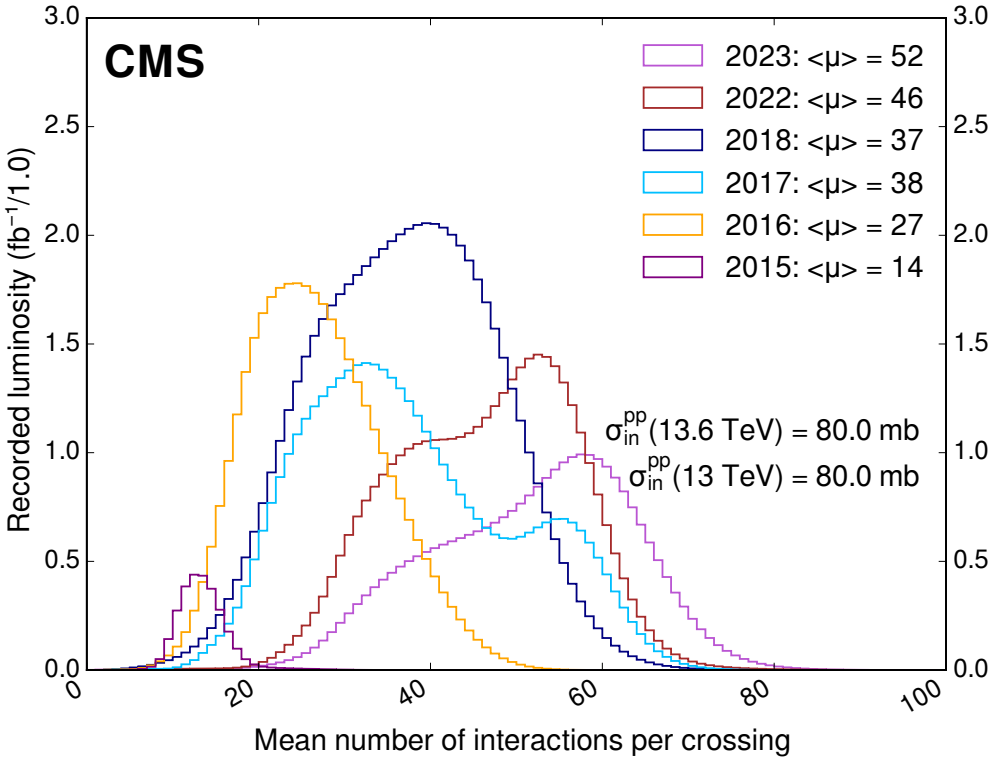

FIGURE 3.1: Distribution of the average number of interactions per crossing. This is shown for data-taking in 2015 (purple), 2016 (orange), 2017 (light blue), 2018 (navy blue), 2022 (brown), and 2023 (light purple) [28].

failures in the L1 reconstruction that sometimes assigned one object to the previous BX. This effect is known as L1 prefiring and was fixed in 2018. For 2016 an 2017 corrections were derived and implemented in the simulation.

Another important change in this period, between 2016 and 2017, was the addition of a new pixel layer to the pixel subdetector in its innermost side with the aim of improving the resolution and the reconstruction of the charged particles.

In addition, during the 2018 data-taking, a region of the HCAL endcap failed, making it impossible to collect information in that part until the end of the data-taking. This issue affects the jet and the electron reconstruction and analyses have to check if they are influenced by it, and, if so, data from that region should be vetoed.

- **The 5.02 TeV Run** was taken in 2017 during a week with a total integrated luminosity of 302 pb^{-1}, much less statistics than Run 2 but a factor 10 than what was delivered in a similar Run in 2015. This Run was taken as a reference for heavy-ion collisions, but it can also be used to perform SM physics measurements. Its low centre-of-mass energy means that the pileup is also greatly reduced, with a mean of 2 interactions per BX, as shown in Figure 3.2.

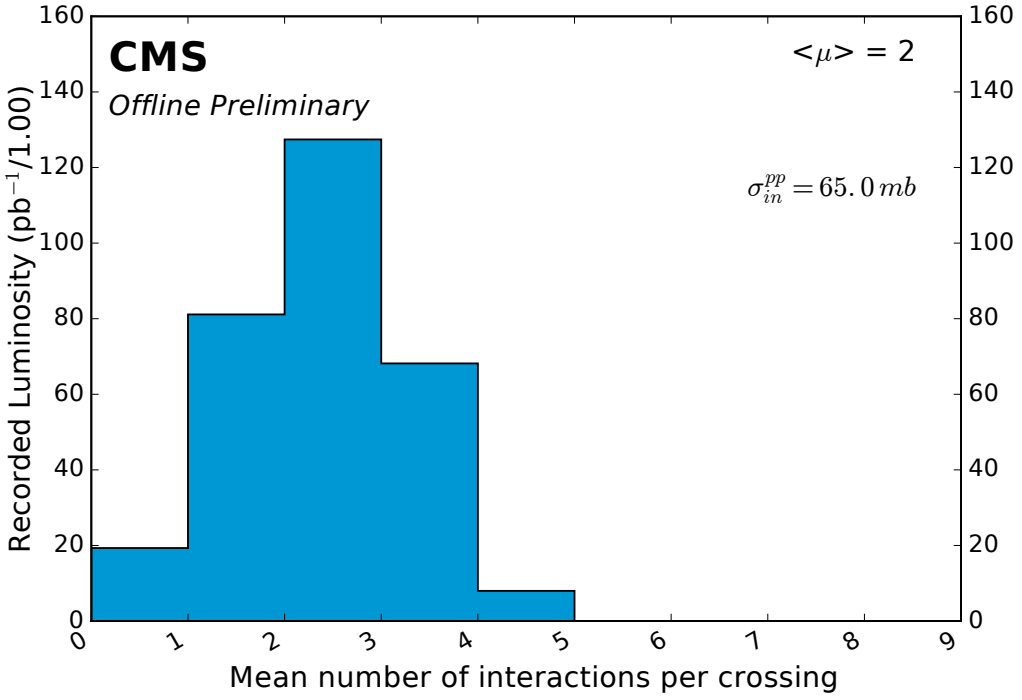

FIGURE 3.2: Mean number of interactions per bunch crossing for the 5.02 TeV pp collisions Run in 2017 [28].

- **The early Run 3** dataset, which means the initial part of the Run 3, was used to perform the first analysis at a new centre-of-mass-energy that is presented in Section 4.2. It started in Summer 2022 with pp collisions at $\sqrt{s} = 13.6$ TeV and includes a total integrated luminosity of 1.2 fb^{-1} with a mean of number of interactions per BX of 46.

3.2 Monte Carlo simulation

To compare experimental data with theory predictions, simulated events are generated for the different signal and background processes. Dedicated simulation samples are essential to provide precision measurements of the SM and to prove (or not) the existence of BSM processes. In CMS, Monte Carlo (MC) methods are used to simulate the pp collisions in the LHC and the interaction of the generated particles with the subdetectors.

The event simulation takes place in 2 main steps. First, the pp collision and the physical processes that occur are modelled in a step called generation. Then, the interaction of the generated particles with the different subdetectors is simulated using a detailed description of CMS.

3.2.1 Generation

In the generation step, the pp collision events are simulated. This is a complex process because protons are not elementary particles, they are hadrons composed by quarks and gluons (partons) glued by the strong interaction. As a consequence, in each collision more than 2 partons may interact.

The parton distribution functions (PDFs), $f_i(x_i)$, gives the probability of finding a parton in a proton with a given fraction of momentum $x_i = p_i/p_p$, where p_i is the parton momentum and p_p the proton momentum. The proton PDFs have been precisely measured in deep inelastic electron-proton scattering at HERA [45]. A related and widely used observable in particle physics is the cross section, σ, which is the probability of a particular process to be produced in a pp collision:

$$\sigma = \sum_{i,j} \int dx_i dx_j f_i(x_i, \mu_F) f_j(x_j, \mu_F) \sigma_{ij}(x_i x_j s; \mu_R, \mu_F), \tag{3.1}$$

where f_i and f_j are the PDFs of each parton and x_i and x_j their momentum fraction; μ_F is the factorization scale that regulates the energy at which the soft and hard interactions are separated; μ_R is the renormalization scale and is introduced in the partonic cross section, σ_{ij}, to cure divergences in the Feynman diagrams. Finally, s is the squared pp centre-of-mass energy.

The PDFs of the protons are used to simulate the hard-scattering through matrix elements (ME), as the initial step of the generation. The various software packages available are able to generate MC at different degrees of approximation in the perturbative expansion of the cross section: Leading Order (LO), Next-to-Leading Order (NLO) and Next-to-Next-to-Leading Order (NNLO). Higher orders of approximation give more accurate predictions. There are several generators with different levels of approximation, but the ones used in this thesis are MadGraph [46] at LO, and powheg [47] and MadGraph_aMC@NLO [46] at NLO.

Once the partons resulted from the hard-scattering are generated, the following step is to simulate the parton shower (PS) of the coloured particles radiated and their hadronization. This radiation may occur before the hard-scattering or after, and it is known as initial state radiation (ISR) or final state radiation (FSR). Furthermore, the partons that do not participate in the main collision can also interact. The result of this interaction is called Underlaying Event (UE). In the MC samples used in this thesis, the PS, hadronization and UE are simulated with pythia 8 [48].

3.2.2 Detector response

The generation step provides the simulation of all the particles produced in a pp collision, but it is necessary to take into account the interaction of these particles with the materials of the different subdetectors when they pass through CMS. This is simulated with the GEANT4 [49] software toolkit, which includes a detailed description of the CMS geometry, its materials and how each particle interacts with them. This software also computes the digitalization of the simulated signals in each subdetector that are then used in the reconstruction of the particle trajectories following the same procedure as in data, described in Section 2.3.

3.3 Corrections to simulation

Although the MC samples are produced with highest possible, it is necessary to apply corrections to compensate small differences with the observed data. The typical corrections that are applied in the analyses are described in the following subsections.

3.3.1 Lepton efficiency

Depending on the lepton selection of each analysis, scale factors (SFs) need to be applied to MC events to account for the differences between the lepton efficiency measured in data and in MC. The SF is the ratio between the lepton efficiency computed from data and from MC, and it is usually measured as a function of the p_{T} and η of the leptons.

The efficiency is computed separately for electrons and for muons, and for reconstruction, selection and isolation steps, but the technique is the same: tag-and-probe [36, 50]. Events from Z boson decays into 2 electrons or 2 muons with an invariant mass close to the Z boson mass are used. One of them, the tag lepton, is required to pass a strict criteria to reject backgrounds, and the other one, the probe lepton, is required to pass a looser selection. The denominator of the efficiency is the number of probe leptons and the numerator is the number of probe leptons passing the requirement to be measured. The leptons passing or failing the requirement are fitted separately to a signal-background model of the Z boson mass resonance using the invariant mass of the leptons. These fits are used to extract the number of passing leptons and to check that the background subtraction works correctly. The efficiency is computed as:

$$Eff = \frac{N_{probe}^{pass}}{N_{probe}^{pass} + N_{probe}^{fail}}. \tag{3.2}$$

The choice of the model, as well as the the invariant mass range of the fit, the MC sample used and the tag lepton criteria are taken into account as systematic uncertainties of the efficiencies.

Once the efficiencies are computed in data and in MC, the SFs can be calculated as the ratio of both. An example of the muon identification SFs as a function of the muon p_T and $|\eta|$ is shown in Figure 3.3.

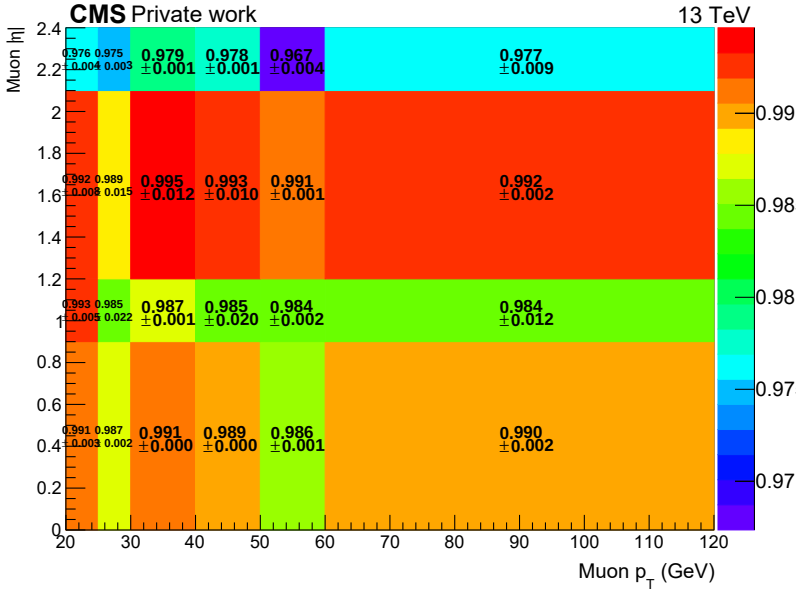

FIGURE 3.3: Muon identification SFs and their total uncertainty for the 2017 dataset.

3.3.2 Trigger efficiency

In order to more accurately model the data, SFs defined as the ratio of the trigger efficiencies in data and simulation are applied to MC events. To measure the trigger efficiency several methods exist. In particular, in the analyses presented in this thesis two of them are used.

Most most of the analyses compute the trigger efficiencies following the cross-trigger method. A set of orthogonal triggers to those used in the analysis are selected to measure the efficiency using the same offline lepton selection as in the analysis with additional minimal requirement based on the reference triggers. The efficiency is calculated then as:

$$\varepsilon = \frac{N_{Events\ passing\ lepton\ selection+orthogonal\ triggers+analysis\ triggers}}{N_{Events\ passing\ lepton\ selection+orthogonal\ triggers}}. \tag{3.3}$$

It is usually estimated as a function of the p_{T} or η of the leptons. Several uncertainties are considered including the statistical uncertainties of the data samples, the correlation between the reference and the analysis triggers and comparisons with other methods.

In the multilepton Run 2 analyses described in Section 5, the orthogonal triggers are $p_{\mathrm{T}}^{\mathrm{miss}}$ triggers. Then, in the analyses at 5.02 TeV, where the statistics is much lower, electron or muon triggers are used as cross-triggers in the muon or electron final states, respectively. Finally, in the early Run 3 analysis the trigger efficiencies are estimated with the tag-and-probe method for single lepton triggers. Figure 3.4 shows the $e\mu$ trigger SFs for the 2017 dataset used in the SUSY search presented in Section 5.1, as illustration.

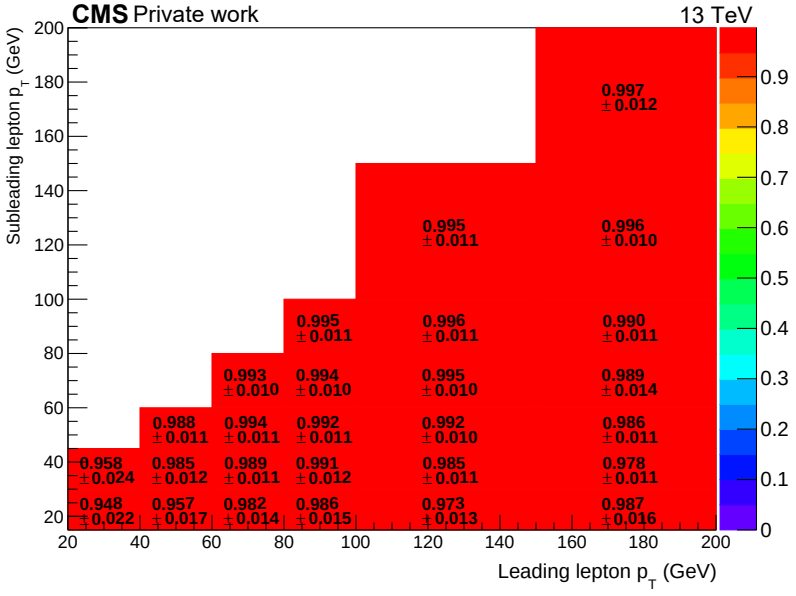

FIGURE 3.4: Trigger SFs and their total uncertainty for $e\mu$ events with the 2017 dataset.

3.3.3 Jet energy and resolution

Converting the measured jet energy into the true parton energy is not direct. Jets are calibrated following a set of tools to properly map the measured jet energy depositions to the parton energy, called jet energy corrections (JEC) [51].

The strategy consists on a chain of corrections applied sequentially, each of them taking care of a different effect that needs to be addressed. The corrections are SFs that depends on the jet p_{T}, η or flavour, and the output of each step is the input to the next one.

First, L1 pileup corrections are estimated from simulated events to remove the energy coming from pileup and UE particles and are applied differently to data and MC.

Second, jet response corrections are determined on simulated events by comparing the p_T from reconstruction and at particle-level as a function of the jet p_T and η in order to make the response uniform. These corrections are also applied to data and MC. Finally, residual corrections are applied only to MC to correct the remaining small differences within jet response in data and MC.

All these corrections have several associated systematic uncertainties related to the jet flavour, the simulation of the fragmentation and UE, the response of the ECAL and HCAL subdetectors and the statistical uncertainty in the determination of the p_T and η dependence. These uncertainties are propagated to p_T^{miss} at analysis level.

Furthermore, the jet energy resolution (JER) is corrected in MC to match the one observed in data.

3.3.4 b tagging

If b-tagged jets are used in the analysis, a SF needs to be applied to each b-tagged jet in MC events to correct the discriminant distribution and make it closer to the data distribution. SFs are centrally provided by CMS as a function of the jet p_T, η and flavour.

The b-tagging efficiency and misidentification rate are taken into account. The b-tagging efficiency is measured in samples enriched in b-tagged jets, typically by selecting multijet or $t\bar{t}$ events although there are different methods to measure it [43], while the misidentification rate is estimated in multijet events. Depending on the selected working point of the discriminant the efficiency values change.

Uncertainties in the b-tagging corrections are also considered in the analyses. They are propagated from the uncertainties of the efficiencies that include a variation of the contamination from light- and heavy-flavour jets in the estimation region and statistical fluctuations in both data and MC.

3.3.5 Pileup

Monte Carlo samples are simulated with the expected PU profile depending on the pp collision conditions, but it usually does not exactly match the measured data. A weight, computed as the ratio between the data and the MC distributions of the number of reconstructed PU vertices, is applied to MC events to correct the PU distribution. This weight is applied to each MC event as a function of the number of reconstructed PU vertices.

In addition, a systematic uncertainty is considered in the analyses by varying the cross section with which the PU MC distributions are derived.

An example of the effect of this correction is shown in Figure 3.5, corresponding to the Run 3 dataset used in the analysis described in Section 4.2. The figure shows the number of PU vertices for the observed data and the MC prediction on the left without the PU weights applied and on the right with the corrections applied. In the bottom panel the ratio between data and MC prediction can be seen. As expected the agreement between the observed data and the prediction from MC improves after applying the correction.

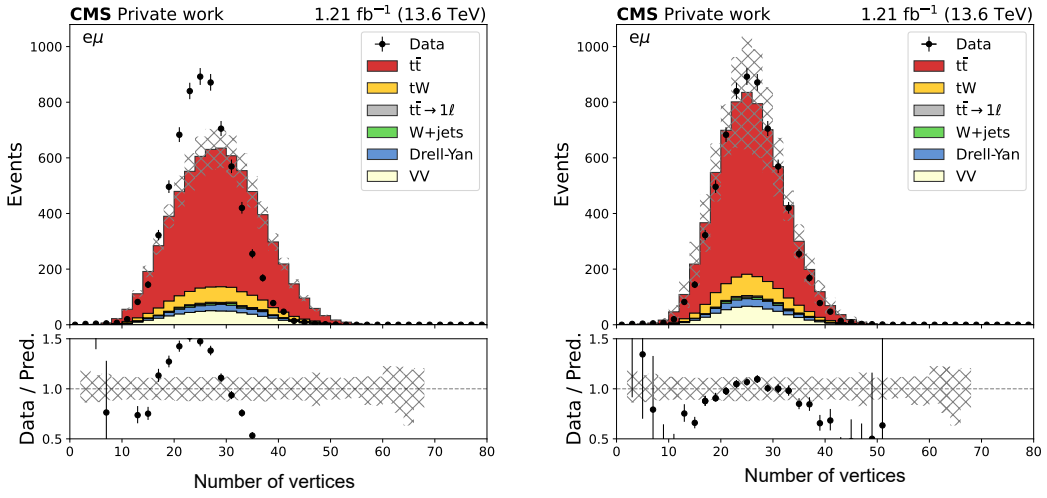

FIGURE 3.5: Left (right): Number of PU vertices in the Run 3 dataset for data and MC prediction without (with) the PU correction applied.

3.4 Personal contributions to the CMS experiment

During the course of my PhD, I have contributed in various aspects to the CMS Collaboration. Initially, in my first two years, I held the position of 'SUSY MC contact,' meaning that I was one of the people in charge of facilitating communication between different SUSY analyses and the generation of their needed MC samples, both of their signal and background samples. This role allowed me to gain extensive knowledge about the various SUSY models under exploration, as I was responsible for adjusting their simulation to produce signal processes with the appropriate final states. More generally, it also helped me to understand in depth how the MC event simulation described in Section 3.2 works, as I had to make use of different types of generators such as POWHEG and MADGRAPH.

Additionally, for the last four years, I have been involved in certifying the data collected by the Drift Tubes subdetector. My task was to verify the proper functioning of this subsystem during data-taking to ensure the quality of the data collected by CMS. This task is crucial, as any errors or misalignments in any of the CMS subsystems during data-taking could render the data unusable or require adjustments. In the last

two years, I have also been responsible for validating that, in each new version of the central software used by the entire CMS Collaboration, the part affecting the DTs remains in good condition.

Finally, in the last year, I held the position known as 'Muon reconstruction L3 convener,' meaning I was one of the two responsible people of the muon reconstruction in CMS. During this year, I performed tasks to improve the muon reconstruction part of the central CMS software, such as removing variables that were not in use or were obsolete. This is crucial because the software is extensive, and any improvement in memory and time is appreciated. This position helped me both to enhance my understanding of muon reconstruction and to comprehend how the central CMS software operates. Furthermore, it also put me in an ideal position to develop a new technique for muon identification, described in Chapter 6.

Chapter 4

Measurement of Standard Model cross sections

The measurement of SM cross sections is a fundamental aspect of particle physics research. Precise calculations of the production cross section of known processes are crucial for comparing experimental results with theoretical predictions from the SM. Any significant deviation between the measured and predicted cross sections could point to the existence of new particles.

In this chapter, three different measurements are described: top quark pair production cross section at $\sqrt{s}$ = 13.6 TeV in Section 4.2 and at $\sqrt{s}$ = 5.02 TeV in Section 4.3, and diboson production cross section also at $\sqrt{s}$ = 5.02 TeV in Section 4.4. But first, an introduction to the previous measurements of these processes and their importance in the LHC physics is presented in Section 4.1.

4.1 Introduction

4.1.1 Top quark

The top quark is the heaviest elementary particle of the SM and the unique quark that decays promptly in other particles, without suffering hadronization. This property makes it the only particle that can be used to probe QCD production. Consequently, investigations into top quark production and its properties offer a precise means of testing the SM. Moreover, numerous BSM theories postulate interactions involving the top quark. All of these factors establish top physics as one of the primary research domains in particle physics, and especially within the context of the LHC.

The top quark was discovered in 1995 at the Fermilab Tevatron by the D0 and CDF Collaborations. The measurements of its discovery were done in proton-antiproton

($p\bar{p}$) collisions at $\sqrt{s} = 1.8\,\text{TeV}$, obtaining a top quark mass of $m_t = 199 \pm 20\,(\text{stat.}) \pm 22\,(\text{syst.})\,\text{GeV}$ by D0 [52] and $m_t = 176 \pm 8\,(\text{stat.}) \pm 10\,(\text{syst.})\,\text{GeV}$ by CDF [53]. Later, the CMS and ATLAS experiments confirmed its existence at the LHC, with a measured mass of $m_t = 172.69 \pm 0.30\,\text{GeV}$ [54].

4.1.1.1 Production and decay modes

At hadron colliders, top quarks are mainly produced in quark anti-quark pairs ($t\bar{t}$). At the Tevatron, the $t\bar{t}$ production was dominated by quark-antiquark annihilation ($q\bar{q}$), while at the LHC it is dominated by gluon-gluon fusion (gg), with a probability of $\approx$ 85%, followed by $q\bar{q}$ with a probability of $\approx$ 15%. The LO Feynman diagrams of the $t\bar{t}$ production process are shown in Figure 4.1.

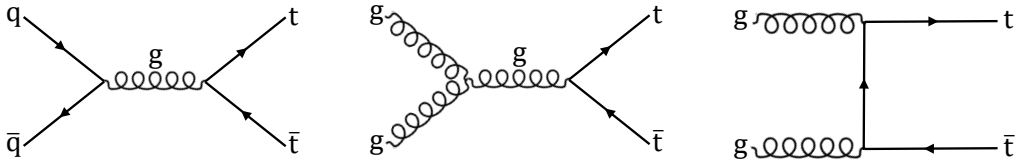

FIGURE 4.1: Feynman diagrams of $t\bar{t}$ production in pp collisions at LO.

The top quark decays almost 100% of the times to a W boson and a b jet. In turn, the W boson can decay hadronically to two quarks, or leptonically to a lepton and a neutrino. Thus, the $t\bar{t}$ production has 3 possible final states, depending on the number of leptons: hadronic (0), semilepton (1) and dilepton (2).

The dilepton final state has the lower production rate, but usually provides the most precise measurements. By requiring the presence of two leptons, a very clean signal region is achieved with a low background contamination. The semilepton final state has a higher branching ratio but also a higher background contamination from multijet (QCD) production. In this channel, the correct identification of b jets is essential to discriminate between $t\bar{t}$ and other backgrounds. Finally, the fully hadronic final state has the larger branching ratio, but is the less precise due to the overwhelming multijet background.

While not as common as the production of $t\bar{t}$ pairs, it is worth noting that top quarks can also be produced at the LHC through processes mediated by EWK interactions. In the majority of these production modes, the top quark is produce individually. At the LHC, the single top production is dominated by the tW-channel, the t-channel, in which the top quark is produced alone and is the dominant production mode, and the s-channel, which is very rare to be produced at the LHC as it requires the presence of sea quarks in the initial state.

4.1.1.2 State-of-the-art

The LHC is colloquially known as 'top factory' due to the high frequency at which top quarks are produced. A summary of the production cross sections of processes involving top quarks are shown in Figure 4.2. As mentioned above, the most common production mode is $t\bar{t}$, but there are many others covering several orders of magnitude and thanks to the large amount of luminosity delivered by the LHC in the last few years, it has been possible to prove the existence of all of them, including the four top quark production ($t\bar{t}t\bar{t}$) [55], which is the one with smaller production cross section.

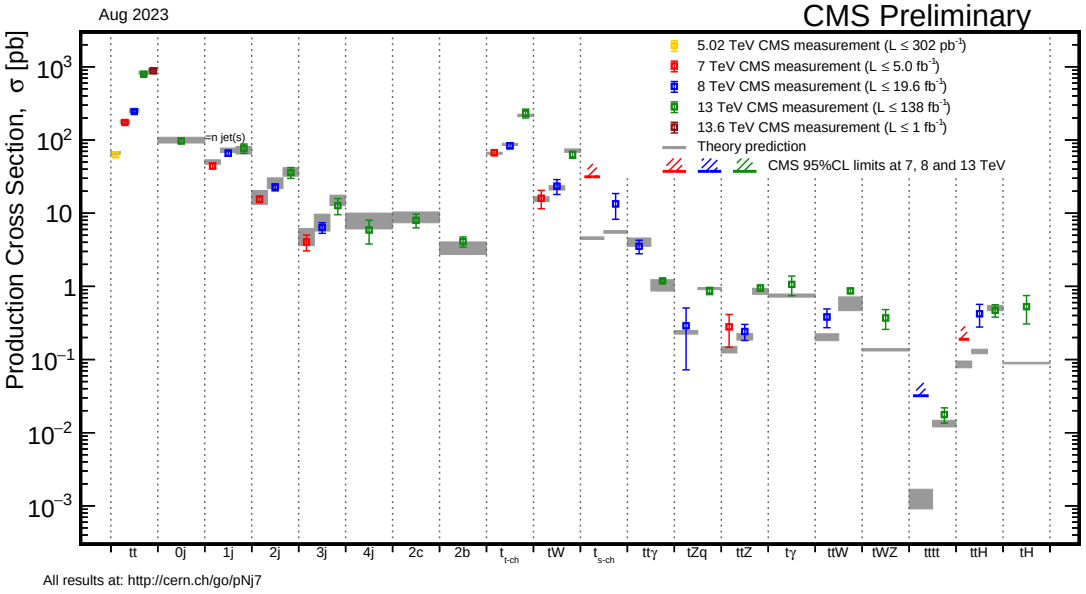

FIGURE 4.2: Summary of production cross sections involving top quarks processes. The theory prediction and CMS measurements are shown at different centre-of-mass energies. [56].

As $t\bar{t}$ has a larger production rate than many other SM processes, it appears as background in almost all the SM measurements and BSM searches and is crucial to have an accurate estimation of its cross section. Moreover, this is the first process that is always measured after a technical stop of the detectors or at the beginning of a Run because, as it has all the objects in its disintegration mode (leptons, jets and b jets), it provides a test of the proper operation of the detector.

CMS and ATLAS have measured the $t\bar{t}$ production cross section at different centre-of-mass energies and with different final states. A summary of the LHC and Tevatron measurements of the top pair production cross section as a function of the centre-of-mass energy up to June 2022 can be found in Figure 4.3, where all measurements are in exceptional accordance with the theoretical prediction at NNLO in QCD complemented with next-to-next-to-leading Logarithm (NNLL) prediction.

The first inclusive $t\bar{t}$ production measurement at $\sqrt{s} = 13\,\text{TeV}$ was done by the CMS Collaboration with the initial part of the Run 2 dataset [58], but it was statistically

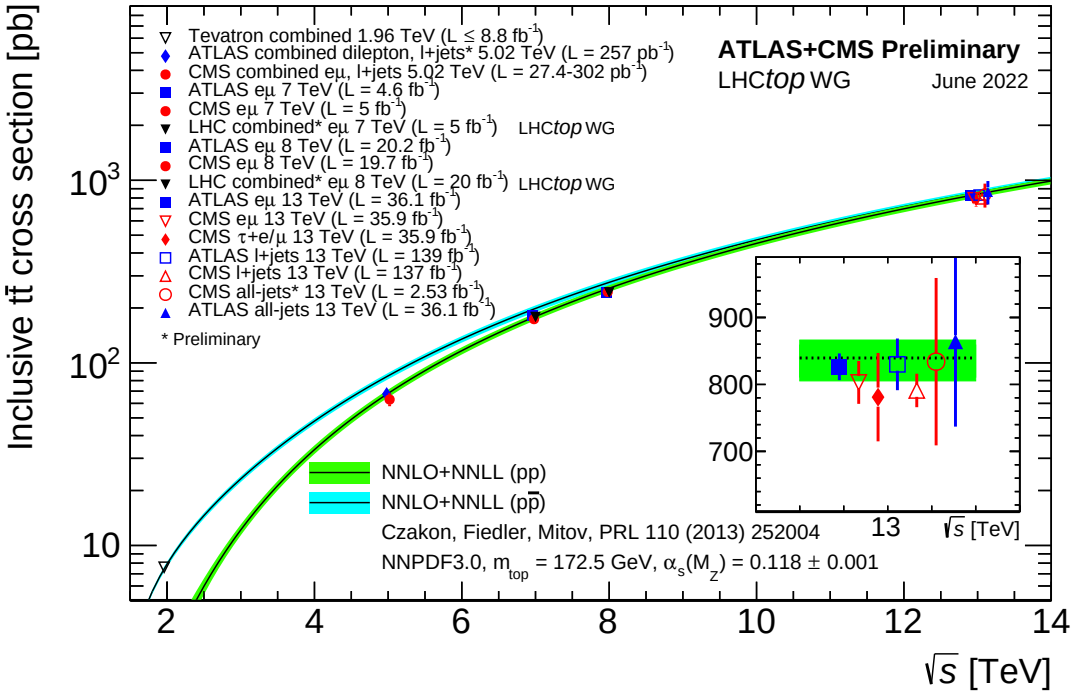

FIGURE 4.3: Summary of LHC and Tevatron measurements of the top quark pair production cross section in pp collisions as a function of the centre-of-mass energy compared to the NNLO+NNLL prediction [57].

limited. Then, ATLAS and CMS Collaborations published more precise results using the full Run 2 dataset in the semilepton final state with a total uncertainty of 4.6% by ATLAS [59] and 3.2% by CMS [60]. ATLAS also has a precise measurement in the electron-muon final state with a total uncertainty of 1.9% [61].

The first measurement at $\sqrt{s} = 13.6\,\text{TeV}$ was done by CMS with the initial part of the Run 3 dataset, using a luminosity of $1.2\,\text{fb}^{-1}$, and it was the $\text{t}\bar{\text{t}}$ production cross section in the dilepton and semilepton final states [1]. It was a complex analysis whose strategy consists on fitting several kinematic distributions and constraining uncertainties, achieving a total uncertainty of 3.4%. I contributed to this result by measuring the cross section with a cut-and-count approach, described in Section 4.2. Both results are compatible and prove the validity of the SM predictions at this previously unexplored energy. Then, ATLAS measured the cross section in the electron-muon final state using $12\,\text{fb}^{-1}$, achieving a similar total uncertainty [62].

During my thesis, I also measured the $\text{t}\bar{\text{t}}$ production cross section at $\sqrt{s} = 5.02\,\text{TeV}$ with the 2017 dataset, both in the dilepton final state [2] and the semilepton final state, which is currently in progress. The plan is to combine both results to be able to compare with the combined ATLAS result [63] and come up with a joint measure.

The $\text{t}\bar{\text{t}}$ measurements in which I have participated are described in detail in Sections 4.2 and 4.3.

4.1.2 Diboson

The $W^{\pm}$ and Z are the gauge bosons mediators of the EWK interaction. Their discovery at CERN with the Super Proton Synchrotron in the 1980s confirmed the unification of the electromagnetic and weak forces. They can be produced separately, but this thesis focuses on the production of boson pairs.

4.1.2.1 Production and decay modes

The on-shell production of massive gauge bosons pairs, commonly called diboson production or VV, where V = W or Z, constitutes other class of processes taking place at CERN. Diboson production encompasses WW, WZ and ZZ processes, which can involve triple gauge couplings (TGC), making them interesting for testing the EWK part of the SM. Some examples of the LO Feynman diagrams of the VV production processes through triple gauge couplings are shown in Figure 4.4.

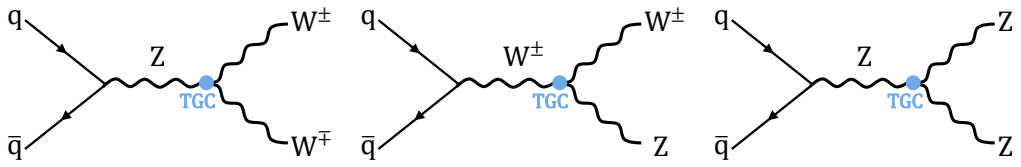

FIGURE 4.4: Feynman diagrams of WW (left), WZ (middle) and ZZ (right) production in pp collisions at LO.

Diboson production at the LHC is predominantly dominated by $q\bar{q}$ processes, with a probability around 95%. This dominance is attributed to the fact that W and Z bosons do not directly couple to gluons. Consequently, production modes in which one or both of the initial partons involved in the hard-scattering are gluons are relatively rare.

The W bosons decay to hadrons most of the times ($\approx$ 70%), with a small contribution from decays to lepton+neutrino ($\approx$ 30%). On the other hand, the most common decay mode of the Z boson is also to hadrons ($\approx$ 70%), but its second largest contribution comes from decays to invisible particles ($\approx$ 20%), which will lead to $p_{\mathrm{T}}^{\mathrm{miss}}$ in the final state. Finally, Z bosons can also decay with a smaller proportion to two or four leptons ($\approx$ 10%). This means that the most frequent final states of VV production are hadronic, but they are the most contaminated by background contributions from multijet processes. Lepton final states have lower branching ratio, but offer a cleaner phase space.

4.1.2.2 State-of-the-art

VV production is an irreducible background in the Higgs boson measurements and in many BSM searches, so have a precise measurement of its production cross section is essential.

The diboson production cross section is much smaller than the $t\bar{t}$ one, and it is different for each individual process. The highest is that of the WW process, followed by WZ, and by ZZ in last place. As discussed in the previous subsection, the most accurate measurements are given by final states with leptons, but the more leptons the smaller the branching rate. This makes VV measurements a challenge for the analyzers, which have to search for an intermediate phase space. A summary of the VV production cross section measurements by ATLAS at a centre-of-mass energies from 7 to 13 TeV is shown in Figure 4.5, where it can be seen that the ZZ production cross section is approximately 2 orders of magnitude smaller than that of $t\bar{t}$.

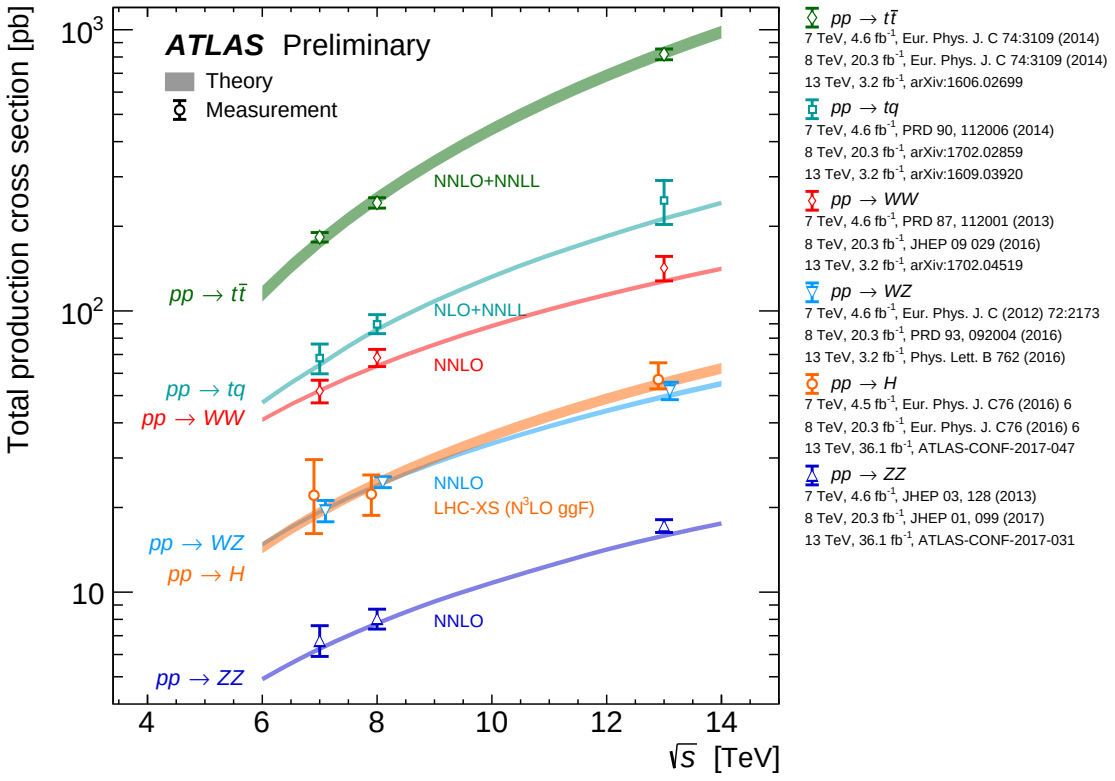

FIGURE 4.5: Summary of total production cross section measurements of several processes by ATLAS presented as a function of centre-of-mass energy. The theory predictions at NNLO are also shown [64].

Figures 4.6 and 4.7 summarizes the VV production cross section measurements carried out by the CMS and ATLAS experiments, respectively, in comparison with the SM theory at centre-of-mass energies of 7, 8 and 13 TeV. All the experimental measurements are consistent with the SM predictions.

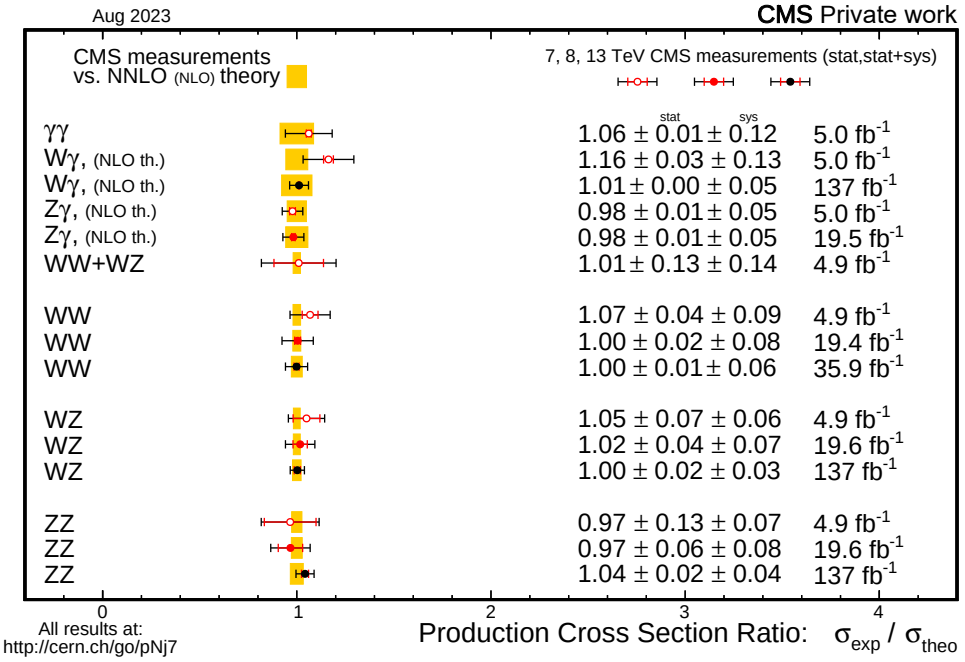

FIGURE 4.6: Summary of the diboson production cross section ratio between CMS measurements and theory at NNLO or NLO (when NNLO is not available) calculations.

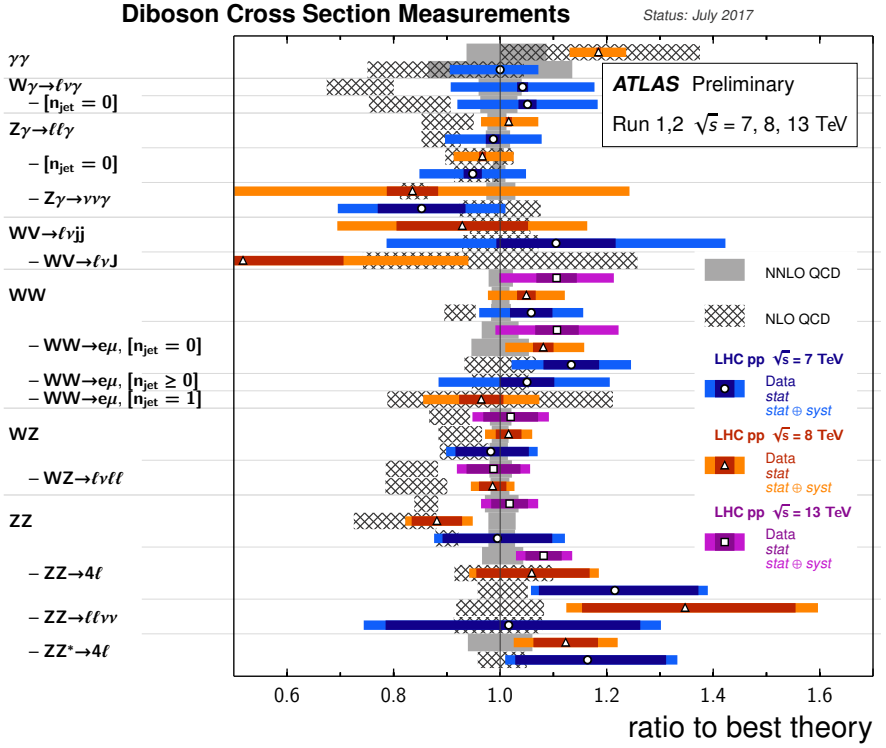

FIGURE 4.7: Summary of the diboson production cross section ratio between ATLAS measurements and theory at NNLO or NLO calculations. All theoretical expectations are shown using gray bars, hatched for NLO calculations and full for NNLO predictions. [64].

A more complete summary, which also incorporates measurements from D0 and CDF experiments, is presented in the Subsection 4.4.3 as part of the results presented in the CMS publication of the first VV measurement at 5.02 TeV [3], in which I participated. This measurement was performed using the 2017 dataset, with a total integrated luminosity of 302 fb^{-1} and selecting events with leptons in the final state. More details of this analysis can be found in Section 4.4.

Apart from the cross section, there are other interesting measurements related with diboson processes that can be made. An example is the charge asymmetry between W^+Z and W^-Z production, found to be $\frac{\sigma(pp\to W^+Z)}{\sigma(pp\to W^+Z)} = 1.41 \pm 0.04(\text{stat.}) \pm 0.01(\text{syst.}) \pm 0.01(\text{lumi.})$ in a measurement done with the full Run 2 dataset at $\sqrt{s}$ =13 TeV by the CMS Collaboration [65]. This publication also includes a study of the W and Z boson polarization, in which the longitudinally polarized W bosons was observed by the first time in the WZ channel, a search for anomalous values of the charged TGC WWZ, where strong constraints were provided, and an EFT interpretation in association with the WZ production.

4.2 Top quark pair production cross section at $\sqrt{s}$ = 13.6 TeV

In this section, a measurement of the $t\bar{t}$ inclusive cross section in final states with exactly one electron-muon pair with opposite sign of the electric charge and at least two jets is described [1]. The measurement is performed using a dataset of 1.21 $\pm$ 0.07 fb^{-1} recorded by the CMS detector during the first months of the Run 3 data-taking, with a centre-of-mass energy of 13.6 TeV.

4.2.1 Analysis strategy

The inclusive cross section, $\sigma_{t\bar{t}}$, is extracted following a cut-and-count approach, already used in previous CMS measurements [66, 67]. This strategy consists on counting the number of events that pass a certain selection in data and estimate the number of background events that survive to that selection, and extrapolating it to the full phase space following the equation:

$$\sigma_{t\bar{t}} = \frac{N - N_{\text{bkg}}}{BR \cdot \varepsilon \cdot \mathcal{A} \cdot L_{int}}, \tag{4.1}$$

where N is the number of observed events, N_{bkg} is the number of predicted background events, BR is the branching ratio of $t\bar{t}$ into the $e^{\pm}\mu^{\mp}$ final state, ε is the selection efficiency, $\mathcal{A}$ is the event acceptance, and L_{int} is the integrated luminosity.

The event acceptance is estimated from $t\bar{t}$ simulated events as the ratio between the number of generated events passing the event selection over the total number of generated events. The selection efficiency is also estimated from $t\bar{t}$ MC events after applying all the corrections to the efficiency (trigger, leptons and jets).

4.2.2 Object selection

The reconstruction of the different objects is based on the PF algorithm [37]. Then, further selection requirements are imposed.

4.2.2.1 Electrons

First, PF reconstructed electrons with p_T greater than 35 GeV and $|\eta| < 2.4$ are selected, excluding the region with $1.444 < |\eta| < 1.566$ to avoid the transition between barrel and endcap of the ECAL subdetector.

Additional identification (ID) criteria is applied to reject misidentified electrons following a sequential approach, known as cut-based ID [68], which includes several variables:

- The width of the ECAL shower in the η direction, $\sigma^2_{i\eta i\eta}$.
- The ratio between the deposited energy in the HCAL and the ECAL subdetectors, H/E.
- $1/E - 1/\rho$.
- The difference in η and ϕ directions between the ECAL SC and the extrapolated track, $\Delta\eta$ and $\Delta\phi$.
- The relative isolation of the electron, defined as the sum of the p_T of all the PF candidates inside a cone with $R < 0.3$ in the electron direction divided by the electron p_T.
- The number of expected missing inner tracker hits.
- The rejection of tracks from electrons that could result from photon interactions with the detector material, called conversion veto.

These variables are used to define several WPs with different efficiencies. In this analysis, the tight WP is selected and its requirements are summarized in Table 4.1.

Furthermore, electrons are required to have small impact parameters: the impact parameter in the transverse plane $|d_{xy}| < 0.05$ (0.10) cm and the longitudinal impact paramenter $|d_z| < 0.10$ (0.20) cm, in the barrel (endcap).

	$\lvert\eta\rvert \leq 1.479$	$\lvert\eta\rvert > 1.479$
$\sigma^2_{i\eta i\eta} <$	0.0104	0.0353
$H/E <$	0.026+1.15$/E_{SC}$+0.0324ρ/E_{SC}	0.0188+2.06$/E_{SC}$+0.183ρ/E_{SC}
$\lvert 1/E - 1/\rho\rvert <$	0.159	0.0197
$\lvert\Delta\eta\rvert <$	0.00255	0.00501
$\lvert\Delta\phi\rvert <$	0.022	0.0236
Relative isolation $<$	0.0287+0.506$/p_T$	0.0445+0.963$/p_T$
Expected missing inner hits $\leq$	1	1
Conversion veto	yes	yes

TABLE 4.1: Electron tight cut-based ID definition.

4.2.2.2 Muons

Reconstructed PF muons are initially selected if they have $p_T > 35\,\text{GeV}$ and $|\eta| < 2.4$. Then, to reject nonprompt muons, a cut-based ID criteria [69] is used, in particular, the tight WP, which consists of multiple identification requirements to suppress muons from decays in flight and hadrons detected in the muon system after traversing the calorimeters, known as punch-throughs. The muon variables used by the tight WP are:

- Muon reconstructed as global muon.
- Normalized χ^2 of the global muon track fit ($\chi^2/ndof$).
- Number of muon chamber hits included in the global muon track fit ($N_{\text{muon chamber hits}}$).
- Number of muon stations with muon segments used in the global muon track fit ($N_{\text{muon stations}}$).
- Number of pixel hits used to fit the global muon track ($N_{\text{pixel hits}}$).
- Number of tracker layers with hits used in the global muon track fit ($N_{\text{tracker layer hits}}$).
- Transverse and longitudinal impact parameters.

Table 4.2 summaries the set of criteria applied in the tight WP of the cut-based ID.

Variable	Requirement
Global muon	yes
$\chi^2/ndof$	< 10
$N_{\text{muon chamber hits}}$	≥ 1
$N_{\text{muon stations}}$	≥ 2
$N_{\text{pixel hits}}$	≥ 1
$N_{\text{tracker layer hits}}$	≥ 5
$\lvert d_{xy}\rvert$	< 0.02 cm
$\lvert d_z\rvert$	< 0.05 cm

TABLE 4.2: Muon tight cut-based ID definition.

In addition, a requirement in the relative muon isolation of 0.15 is applied taking in to account a cone in the muon direction with $R < 0.4$.

4.2.2.3 Jets

The jets are reconstructed using PF PUPPI candidates with the anti-k_T algorithm within an opening angle of $R < 0.4$. The initial selection criteria for jet candidates are $p_T > 30\,\text{GeV}$ and $|\eta| < 2.4$. Then, the recommended identification requirements are applied, corresponding to the tight ID, and summarized in Table 4.3.

Variable	Requirement
Fraction of neutral hadronic energy	< 0.9
Fraction of neutral electromagnetic energy	< 0.9
Fraction of charged hadronic energy	> 0.0
Fraction of charged electromagnetic energy	< 0.8
Charged multiplicity	> 0

TABLE 4.3: Jet tight ID definition.

The jet energy corrections and resolution are applied, derived specifically for this data-taking. In addition, a jet-lepton cleaning is considered by requiring the jet to be separated from any selected lepton by $R > 0.4$ to avoid double-counting of objects.

4.2.3 Event selection

The $t\bar{t}$ dilepton final state is typically characterized by the presence of a high-p_T isolated lepton (electron, muon) pair associated with p_T^{miss} and 2 b jets.

The trigger strategy of this analysis is based on double lepton triggers and single lepton triggers. The single lepton HLT selection requires the presence of an isolated electron (muon) reconstructed with $p_T > 32$ (24) GeV, while the dilepton HLT selection requires the presence of one isolated electron with $p_T > 23\,\text{GeV}$ and one isolated muon with $p_T > 8\,\text{GeV}$, or one isolated muon with $p_T > 23\,\text{GeV}$ and one isolated electron with $p_T > 12\,\text{GeV}$.

Then, events are selected if the two leading leptons of good quality are an electron and a muon, both of them with $p_T > 35\,\text{GeV}$. In a second step, events are rejected if these two leptons have not opposite electric charge. In addition, no more leptons in the event with $p_T > 10$ GeV are required. To reduce the contamination from the production of a Z boson or a virtual photon decaying into a lepton pair with opposite sign of the electric charge, process known as Drell-Yan or DY, the invariant mass of the lepton pair is required to be greater than 20 GeV. On top of that, at least two selected jets are required to be present in the event to further reduce background contributions

from processes such as VV, DY, or the production of a top quark in association with a W boson (tW).

4.2.4 Signal and background estimation

All the processes are estimated using MC simulation. The production of $t\bar{t}$ and single top quark events is simulated with the POWHEG (v2) [47] generator at NLO in QCD precision. The default value of the top quark mass is set to 172.5 GeV. DY and W+jets events are simulated at LO in QCD with MADGRAPH (v2.6.5) [46] generator. Diboson events are simulated at LO using PYTHIA 8 (v306). In all the samples, the ME calculation is simulated with the NNPDF3.1 [70] PDF set at NLO, the PS is simulated with PYTHIA 8 (v306) [48] with the CP5 [71] tune, and the CMS detector response with GEANT4 [49].

The main backgrounds of the analysis after the event selection arise from tW production and VV production in which at least two prompt leptons are produced from Z or W decays. The tW background is irreducible since one of its possible final states is very similar to that of $t\bar{t}$: 2 leptons, 2 neutrinos and 1 b jet. The Feynman diagram of this decay is shown in Figure 4.8. Other background sources are $t\bar{t}$ or W boson with additional jets ($t\bar{t} \rightarrow 1\ell$ or W+jets) events with decays into lepton+jets where at least one jet is incorrectly reconstructed as a lepton or a lepton from the decay of bottom or charm hadrons is reconstructed as a prompt lepton. Finally, DY contributes as a very residual background as it is suppressed by the selection.

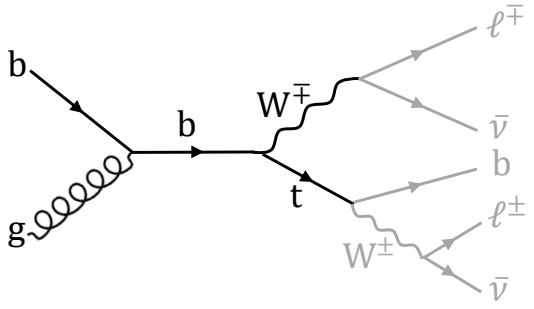

FIGURE 4.8: Example of a Feynman diagram of the tW production in pp collisions with further decays to two leptons, two neutrinos and 1 b jet.

Different SFs are applied to correct the difference in efficiencies in the reconstruction and identification of the events and the efficiency on the simulation. Dedicated lepton reconstruction, identification and isolation SFs are calculated using the tag-and-probe method separately for electrons and muons as a function of p_T and η.

Trigger efficiencies are also computed using the tag-and-probe method for single lepton triggers. Then, the SF is calculated for the electron-muon final state as the union of the muon and electron trigger efficiencies: $\epsilon = \epsilon_{elec} + \epsilon_{muon} - \epsilon_{elec} \times \epsilon_{muon}$.

Finally, MC events are reweighted to match the distribution of true interactions observed in data due to multiple collisions in the same bunch crossing. The target pileup distribution for data is generated using the instantaneous luminosity per bunch crossing for each luminosity section and the total pp inelastic cross section. For this early analysis, an experimental approach is employed. Events are reweighted based on the average of three different variables: the number of good primary vertices, the mean energy density calculated once using tracker input only and once only using calorimeter input.

Figure 4.9 shows several kinematic distributions in data and the prediction from MC for events passing the event selection before applying the 2 jets requirement, and Figure 4.10 after requiring the presence of 2 jets in the event. These figures also show the error band of the MC simulation, including the statistical and the systematic uncertainties described in Subsection 4.2.5. In all the variables, the distribution in data is well described by the MC prediction.

4.2.5 Systematic uncertainties

The measurement is affected by systematic uncertainties originated from detector effects and from theoretical assumptions. Each source of systematic uncertainty is assessed individually by suitable variations of the MC simulations or by variations of parameter values in the analysis within their estimated uncertainties.

4.2.5.1 Experimental uncertainties

- Uncertainties on the **lepton identification and isolation** are applied as a function of the lepton p_T and η and can be factorized in two groups: the contribution arising from the measurement of the efficiency using the tag-and-probe method in Z events in data, and the contribution of the potential topology differences between those events, used as reference, and signal events ($t\bar{t}$ events), which is 1% for electrons and 0.5% for muons.

- The uncertainties on the **trigger SFs** are propagated from the data and MC efficiencies. An additional uncertainty of 1% for electrons and 0.5% for muons is applied to take into account the difference in the phase space of the analysis and the events used to compute the efficiencies.

- As described in the previous subsection, simulation events are reweighted to match the data by taking the average of three different weights, corresponding to three distributions of **pileup-related** variables. An uncertainty is assigned based on the difference of the averaged pileup distribution and that based on the number of reconstructed primary vertices.

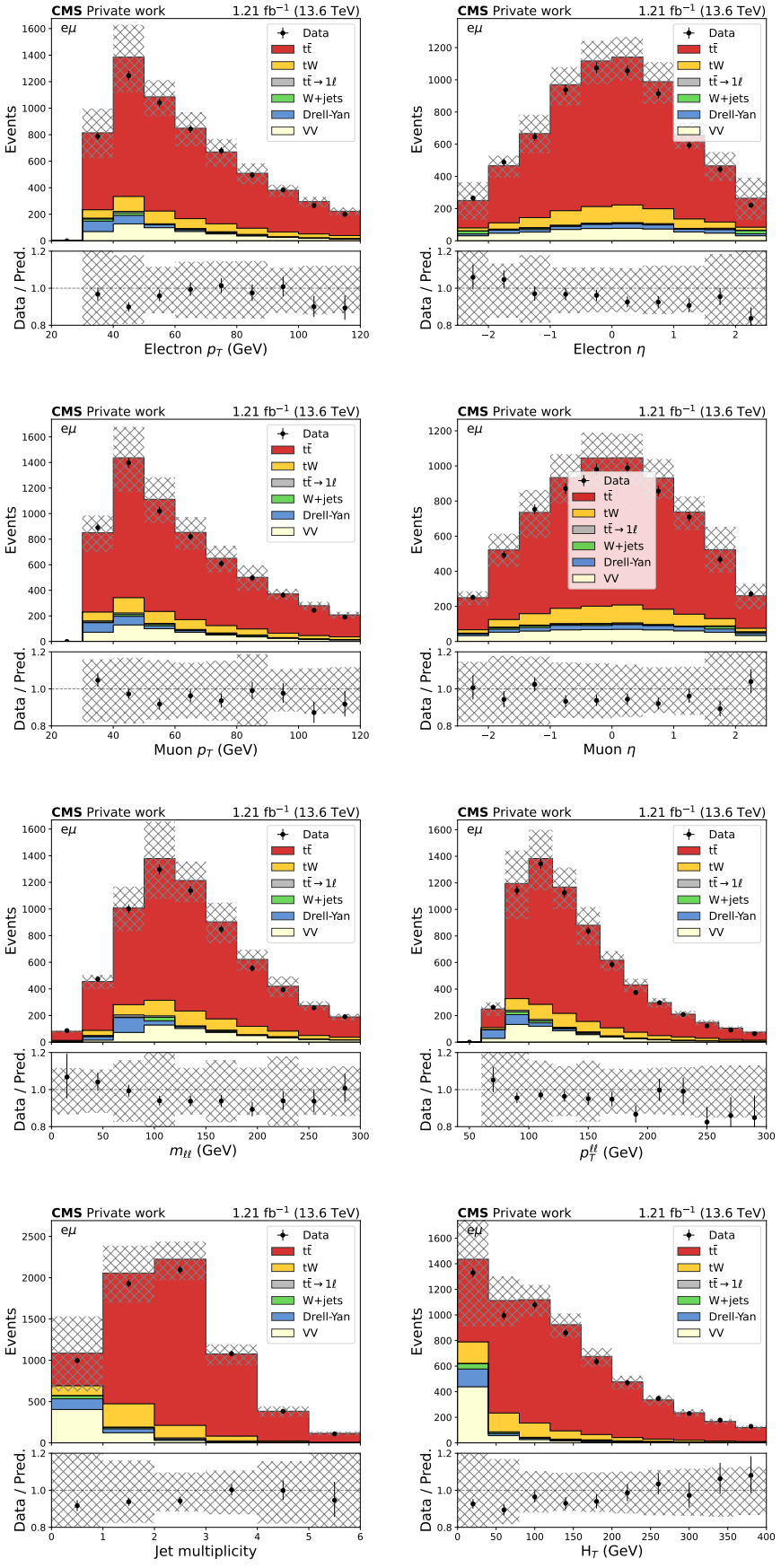

FIGURE 4.9: Distributions in data and predicted MC events passing the selection of an electron-muon pair with opposite sign of the electric charge. From top left to bottom right: electron p_T, electron η, muon p_T, muon η, invariant mass of the two leptons, dilepton p_T, jet multiplicity, and H_T.

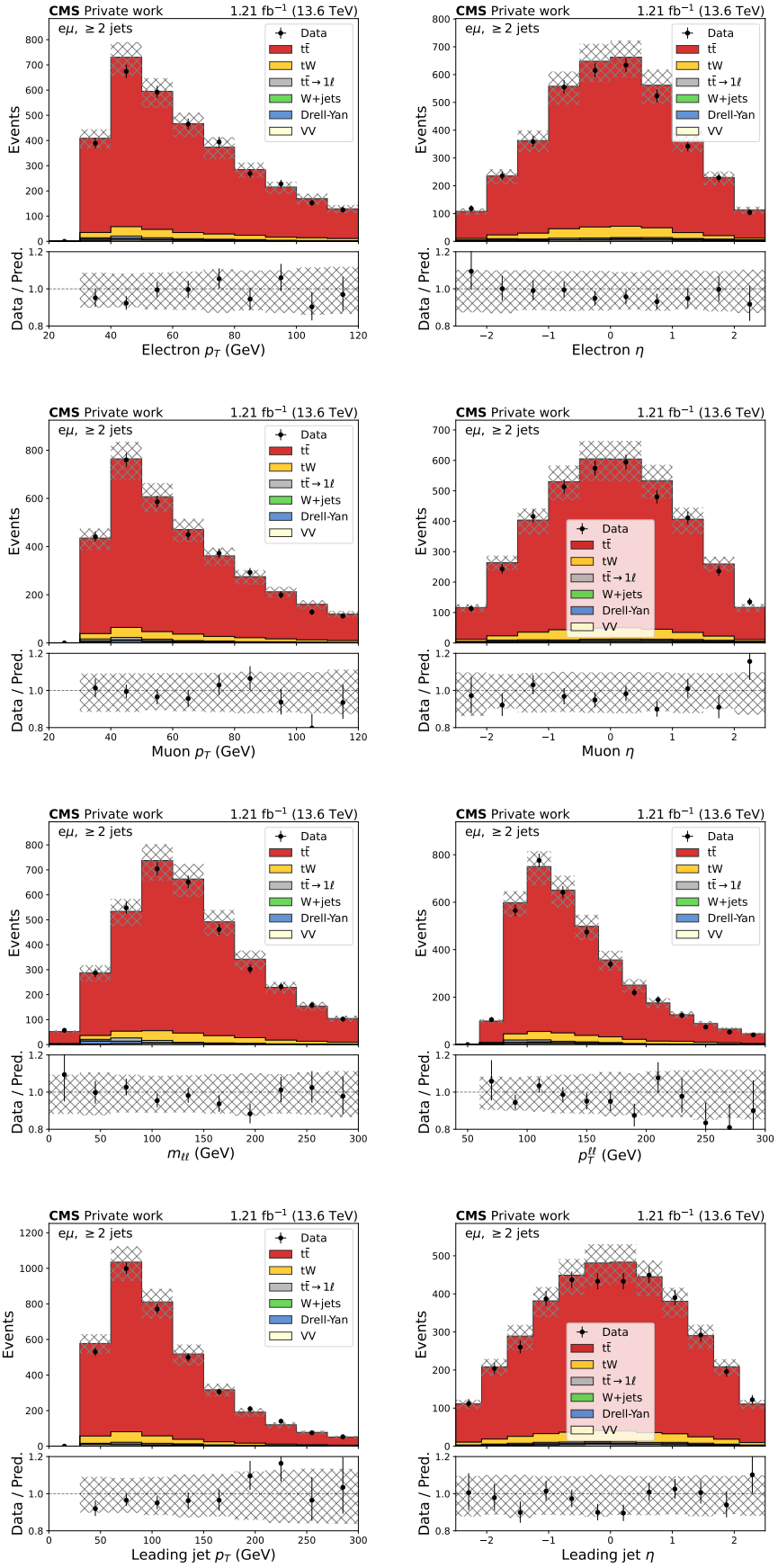

FIGURE 4.10: Distributions in data and predicted MC events passing the event selection. From top left to bottom right: electron p_T, electron η, muon p_T, muon η, invariant mass of the two leptons, dilepton p_T, leading jet p_T, and leading jet η.

- The uncertainty due to the limited knowledge of the **jet energy scale** is determined by changes implemented in the jet energy in bins of p_T and η, typically by a few percent [51]. The different uncertainty sources described in Section 3.3 are taken into account. Furthermore, the uncertainty due to the limited accuracy of the **jet energy resolution** is determined by changing the JER correction within their uncertainties [72].

- For tW **background normalization** an uncertainty of 15% [73] is applied. For the VV background a conservative normalization uncertainty of 30% [74] is assumed. For DY, 20% uncertainty is applied. Finally for the MC nonprompt contributions a 20% uncertainty for the $t\bar{t}$ semileptonic contribution and a 30% [75, 76] for the W+jets sample are considered.

- The uncertainty on the **integrated luminosity** is estimated to be 2.3%.

4.2.5.2 Modelling uncertainties

The modelling of the $t\bar{t}$ events by the MC simulation is an important ingredient in the measurement. The impact of theoretical assumptions in the modelling is determined by repeating the analysis and replacing the standard POWHEG + PYTHIA $t\bar{t}$ simulation by dedicated simulation samples with altered parameters and comparing the result with the $t\bar{t}$ sample used in the analysis.

- **Inicial and final state radiation**: the effect is evaluated by varying the parton shower scales by a factor of 2.

- **ME/PS matching (h_{damp})**: the impact of the matrix element and parton shower matching, which is parameterized by the POWHEG generator as $h_{damp} = 1.58^{+0.66}_{-0.59} m_t$ [71, 77], is calculated by varying this parameter within the uncertainties. This uncertainty is calculated using dedicated $t\bar{t}$ samples.

- **PDF+α_s**: the uncertainty from the choice of PDFs is determined by reweighting the sample of simulated $t\bar{t}$ events according to the 100 NNPDF3.1 [70] error PDF set and two extra weights for the strong coupling constant, α_s, uncertainties. All the individual uncertainties are summed in quadrature to give a single uncertainty.

- **μ_R and μ_F scales**: the ME scale uncertainties for the simulation of $t\bar{t}$ production are taken into account by varying the renormalization and factorization scales, μ_R and μ_F, in POWHEG by factors of 2 and 1/2, independently and simultaneously, but avoiding unphysical cases where $\mu_R/\mu_F = 0.25$ or 4. The maximum variation with respect to the central sample is taken as the uncertainty.

4.2.6 Results

The number of observed events in data and the expected signal and background events from simulation after the event selection are summarized in Table 4.4. The same information is also represented in Figure 4.11, where the comparison of the observed yields in data and the expectation from MC after the event selection shows a good agreement. Statistical and systematic uncertainties are shown in both.

This event selection offers a very clean region, with more than 91% of the MC simulated events coming from signal. The main background is tW, with a contribution of about 6%. The contribution of the other backgrounds is less than 3%.

	Number of events
tW	$230 \pm 3 \pm 35$
$t\bar{t} \to 1\ell$	$37 \pm 1 \pm 7$
W+jets	$0 \pm 0 \pm 0$
Drell–Yan	$26 \pm 2 \pm 5$
Diboson	$36 \pm 1 \pm 11$
Total background	$328 \pm 4 \pm 38$
Expected $t\bar{t}$ signal	$3512 \pm 5 \pm 109$
Observed	3715 ± 60

TABLE 4.4: Total number of events observed in data and the number of signal and background events expected after the event selection. Statistical and systematic uncertainties are shown.

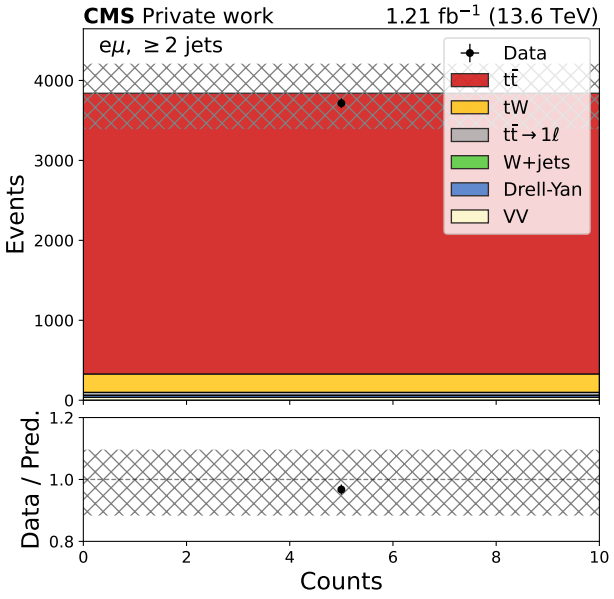

FIGURE 4.11: Observed yields in data compared to the expected ones from MC simulation after the final event selection. The grey band includes systematic and statistical uncertainties.

Table 4.5 summarizes the magnitude of the systematic and statistical uncertainties on the $\sigma_{t\bar{t}}$ measurement. They are computed by propagating the effect of the systematic uncertainties in the number of events to the cross section calculation. Adding the uncertainties in quadrature, the resulting total uncertainty is 4.5%, with the largest contribution arising from the systematic uncertainty of 3.4%. The dominant systematic uncertainties are those of the lepton efficiencies and those related with the jet energy scale. The uncertainty on the integrated luminosity gives also a not negligible contribution.

Source	$\Delta\sigma_{t\bar{t}}$ (pb)	$\Delta\sigma_{t\bar{t}}/\sigma_{t\bar{t}}$ (%)
Muon efficiences	7.1	0.8
Electron efficiences	19.6	2.2
Trigger efficiencies	7.3	0.8
Pileup reweighting	8.2	0.9
Jet energy resolution	0.24	0.03
Jet energy scale	11.5	1.3
Initial state radiation	1.4	0.2
Final state radiation	8.5	1.0
ME/PS matching	11.9	1.3
PDF+α_S	4.3	0.5
μ_R, μ_F scales	0.51	0.06
tW	9.0	1.0
$t\bar{t} \to 1\ell$	1.9	0.2
WJets	0.0	0.0
Drell–Yan	1.4	0.2
Diboson	2.8	0.3
Total systematic uncertainty	30.3	3.4
Integrated luminosity	20.4	2.3
Statistical uncertainty	16.0	1.8
Total	39.8	4.5

TABLE 4.5: Summary of the individual contributions to the systematic and statistical uncertainty on the $\sigma_{t\bar{t}}$ measurement. Numbers in the second column show the absolute uncertainties and the ones in the third column show the percentage.

The $t\bar{t}$ cross section is measured using the method discussed in Subsection 4.2.1, subtracting from the number of observed data events the estimated number of events from all background sources (from Table 4.4) and dividing the resulting number by the integrated luminosity of the data sample, the event acceptance, the selection efficiency and the branching ratio of the $e^{\pm}\mu^{\mp}$ $t\bar{t}$ final state. The event acceptance is found to be $A = 0.261 \pm 0.004$, while the selection efficiency is $\varepsilon = 0.38 \pm 0.01$. The branching ratio assumed of the $e^{\pm}\mu^{\mp}$ final state is BR =0.031938 [78], including decays through τ leptons.

Finally, following Equation 4.1 and assuming a top quark mass of m_t = 172.5 GeV, the measured inclusive cross section is:

$$\sigma_{t\bar{t}} = 888 \pm 34\,(\text{syst.} + \text{stat.}) \pm 20\,(\text{lumi.})\,\text{pb}.$$

This result is in excellent agreement with the cross section obtained with a different approach documented in [1] ($\sigma_{t\bar{t}} = 881 \pm 23\,(\text{syst.} + \text{stat.}) \pm 20\,(\text{lumi.})\,\text{pb}$), and with the theoretical cross section computed using the Top++ 2.0 program [79] at NNLO in QCD complemented with NNLL terms, assuming a top quark mass of $m_t = 172.5\,\text{GeV}$ and using the PDF4LHC21 set of PDFs [80], which is $\sigma_{t\bar{t}} = 924^{+32}_{-40}\,\text{pb}$. The measurement is already included in the cross section summary plot provided by the LHC top working group, shown in Figure 4.12 (updated version of Figure 4.3).

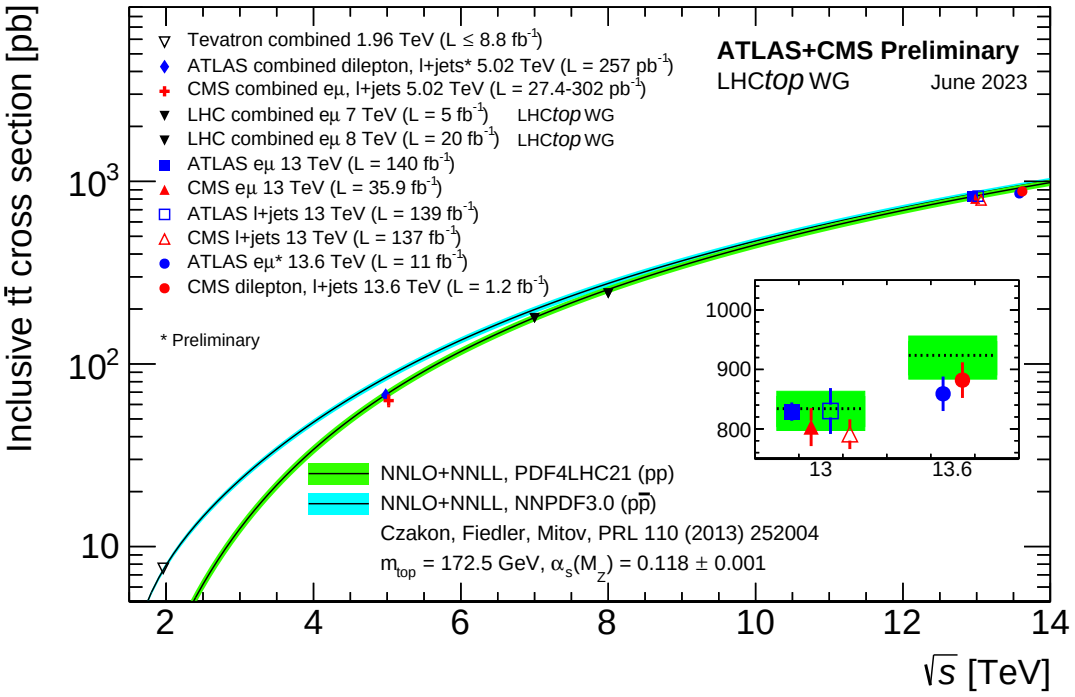

FIGURE 4.12: Summary of LHC and Tevatron measurements of the top quark pair production cross section in pp collisions as a function of the centre-of-mass energy compared to the NNLO QCD calculation complemented with NLL prediction [57].

.

The dependence of the $t\bar{t}$ cross section on the top quark mass is studied with two alternative $t\bar{t}$ samples simulated with top quark masses of 169.5 and 175.5 GeV. A fit to the results is done using a linear function and the slope value obtained is -0.0130 $\pm$ 0.0002, which can be used to linearly extrapolate to other mass values. Figure 4.13 shows the $t\bar{t}$ cross section measured for the different top masses with the statistical uncertainties together with the result of the fit.

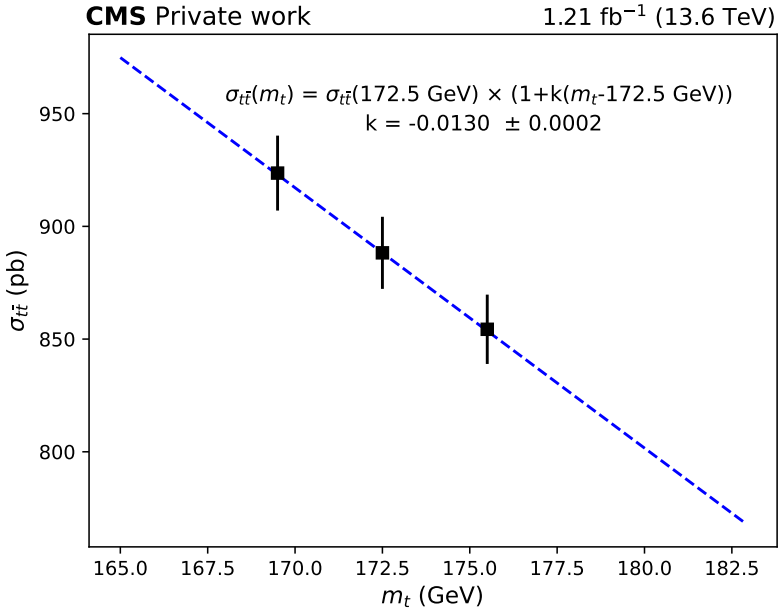

FIGURE 4.13: Dependence of the measured $t\bar{t}$ cross section as a function of the top quark mass. Uncertainty bars correspond to the statistical error.

4.3 Top quark pair production cross section at $\sqrt{s}$ = 5.02 TeV

The CMS and ATLAS Collaborations have measured the $t\bar{t}$ cross section at several centre-of-mass energies, but the 5.02 TeV is of special interest because it has the lowest average of PU interactions, one order of magnitude below than that of the 13 TeV pp collisions. Furthermore, as the energy available for the hard-scattering process is reduced, the longitudinal momentum carried by the gluon will be smaller [81] and thus can provide complementary information on the gluon distribution.

The first measurement of the $t\bar{t}$ production cross section at $\sqrt{s} = 5.02$ TeV was performed by the CMS experiment in the dilepton and semilepton final states using a data sample collected in 2015 that corresponds to an integrated luminosity of $27.4 \pm 0.6\,\text{pb}^{-1}$. It obtains a combined cross section of $\sigma_{t\bar{t}} = 69.5 \pm 6.1\,(\text{stat.}) \pm 5.6\,(\text{syst.}) \pm 1.6\,(\text{lumi.})$ pb, dominated by the statistical uncertainty [82].

Then, a dilepton measurement in the electron-muon final state was done using $302 \pm 5\,\text{pb}^{-1}$ collected in 2017 at 5.02 TeV [2], an order of magnitude more data than in the 2015 dataset. The analysis follows the same cut-and-count strategy described in the previous section. The measured $t\bar{t}$ production cross section is $\sigma_{t\bar{t}} = 60.7 \pm 5.0\,(\text{stat.}) \pm 2.8\,(\text{syst.}) \pm 1.1\,(\text{lumi.})$ pb. The result is still dominated by the statistical uncertainty, which is about 8%. To improve it, a combination with the 2015 dataset measurement in the semilepton final state is performed. The semilepton analysis uses the $27.4\,\text{pb}^{-1}$ and obtains a $\sigma_{t\bar{t}} = 68.9$ pb with a total uncertainty of 13%, also

dominated by the statistical contribution. The combined cross section is $\sigma_{t\bar{t}} = 63.0 \pm 4.1\,(\text{stat.}) \pm 3.0\,(\text{syst.} + \text{lumi})$ pb.

Both results are in agreement with the the theoretical SM prediction at NNLO+NNLL assuming a top quark mass of 172.5 GeV and using the NNPDF3.1 PDF set: $\sigma_{t\bar{t}} = 66.8 \pm 3$ pb [78]. The comparison of the CMS and ATLAS experimental results with the theoretical predictions are summarized in Figure 4.14.

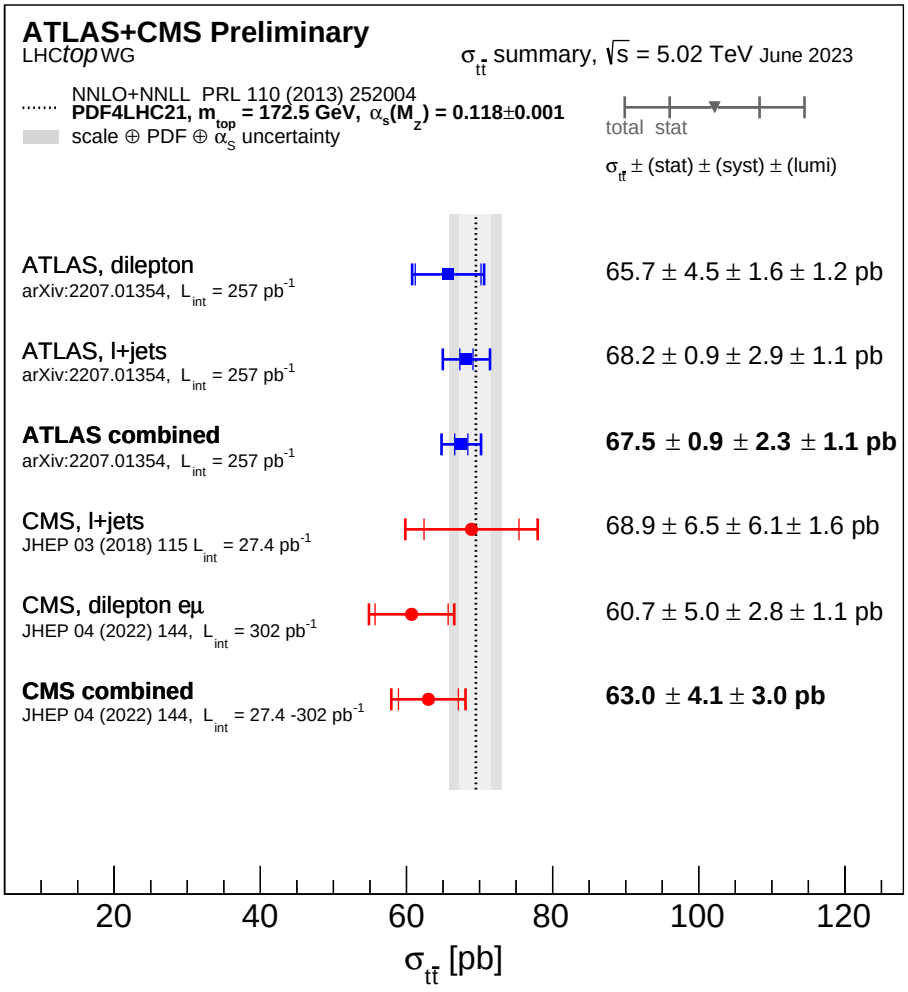

FIGURE 4.14: Summary of the CMS and ATLAS $t\bar{t}$ cross section measurement in pp collisions at 5.02 TeV in the separate semilepton and dilepton final states along with the combined measurements. The theoretical prediction is also shown [57].

The statistical uncertainty is reduced from 8.2% to 6.5% with the combination, but it is still the dominant source of uncertainty. To further improve this result I am working on measuring the $t\bar{t}$ cross section in the semilepton final state with the 2017 dataset, and its combination with the dilepton measurement. This analysis is described below.

4.3.1 Object and event selection

This measurement is done in the semilepton final state of the $t\bar{t}$ production, which implies selecting events with multiple jets and b-tagged jets. This will give important background contributions from QCD or W+jets processes, which will be reduced using sophisticated lepton identification techniques and a MVA specifically trained for this analysis to separate $t\bar{t}$ from W+jets.

Leptons are identified using a MVA algorithm, in particular a boosted decision tree (BDT), that was designed to separate leptons coming from W, Z, and Higgs bosons or τ decays (prompt leptons) from those originating from charm or bottom hadron decays (nonprompt leptons). This algorithm was initially developed for the $t\bar{t}$ production is association with a Higgs boson ($t\bar{t}H$) multilepton analysis [83], but has been already used in several analysis involving leptons [3, 55, 84–91]. The details of the algorithm are documented in a very recent publication in which I have participated [6].

In addition to the lepton MVA requirements, muons and electrons candidates are required to have p_T >20 GeV and $|\eta|$ <2.4. A summary of the electron and muon selection criteria can be found in Table 4.6, where SIP_{3D} is the ratio between the 3D impact parameter and the uncertainty in its estimation, and the mini-isolation is a kind of isolation constructed by defining a cone whose radius varies with the p_T of the lepton: $R = 0.2$ for leptons with $p_T < 50$ GeV, R =10 GeV/p_T for lepton p_T between 50 and 200 GeV, and $R = 0.05$ in other cases. Mini-isolation is very useful to select isolated leptons in boosted topologies where leptons can overlap with jets.

	Electrons	Muons
$\|d_{xy}\|$ <	0.02 cm	0.05 cm
$\|d_z\|$ <	0.1 cm	0.1 cm
SIP_{3D} <	8	8
Mini-isolation <	0.085	0.325
Lepton MVA >	0.125	0.55
Lost hits =	0	-
Conversion veto	True	-

TABLE 4.6: Selection criteria for electrons and muons.

Jets are reconstructed using PF candidates with the anti-k_T algorithm in a cone with $R < 0.4$, and specific jet energy corrections developed for this special Run are applied. The jet identification requirements are the same described in Section 4.2.2.3, corresponding to the tight ID. Identified jets are selected for the analysis if they have $p_T >$ 25 GeV, $|\eta| < 2.4$ and if no selected lepton is found among their constituents.

Then, the DeepCSV b-tagging algorithm [43] is applied on each jet. Jets are identified as b jets using the medium working point, corresponding to a misidentification rate of 1% and an efficiency around 75%.

In this analysis, single lepton triggers with low lepton p_T threshold are used, which is 17 GeV for muons and 20 GeV for electrons.

The event selection is designed to maximize the signal contribution over the expected background contamination. Events are required to contain a single selected electron or muon with p_T greater than 20 GeV and at least 3 jets. A requirement of $p_T^{miss} > 30$ GeV is applied to reduce the background contamination from QCD events.

Events are further categorized into 12 categories depending on the number of jets, the number of b-tagged jets, and the lepton flavour (electron or muon):

Number of b-tagged jets	3 jets	4 jets	$\geq$ 5 jets
1	3j1b	4j1b	$\geq$5j1b
$\geq$ 2	3j$\geq$2b	4j$\geq$2b	$\geq$5j$\geq$2b

TABLE 4.7: Categorization of the events. They are further divided according to the lepton flavour.

4.3.2 Signal and background estimation

Both $t\bar{t}$ signal and background processes are estimated with MC simulation, with the exception of QCD events, which are estimated from data with a technique described in detail in Subsection 4.3.2.1.

The simulated $t\bar{t}$ events are generated at NLO in QCD using POWHEG (v2) interfaced to PYTHIA 8 (v230) for PS with the CP5 underlying event tune. A nominal top quark mass of 172.5 GeV and NNPDF3.1 NNLO PDFs are used. A similar setup is used for the simulation of the single top quark production, *t*-channel, and in association with a W boson, tW. DY and W+jets events are generated at NLO with the use of MADGRAPH_AMC@NLO (v2.4.2) interfaced to PYTHIA 8 (v230). The simulation includes lepton pairs with invariant mass $m_{\ell\ell} > 20$ GeV and up to two extra partons at matrix element level. All the simulations include an emulation of the full detector response, based on GEANT4.

After the event selection, the main background contributions come from single top (tW and t-channel) and W+jets events in their lepton+jets final state, and QCD multijets events when a jet is misidentified as a lepton. Drell-Yan process gives a very small contribution when one of the leptons in the final state is misidentified as a jet.

Figure 4.15 shows the number of observed events in data and predicted events in each of the categories. The category with the higher number of events is 3j1b, but it has an important contribution from W+jets. Categories with at least two b-tagged jets are the purest, with practically all the events from $t\bar{t}$ signal and a very small contribution from single-lepton processes.

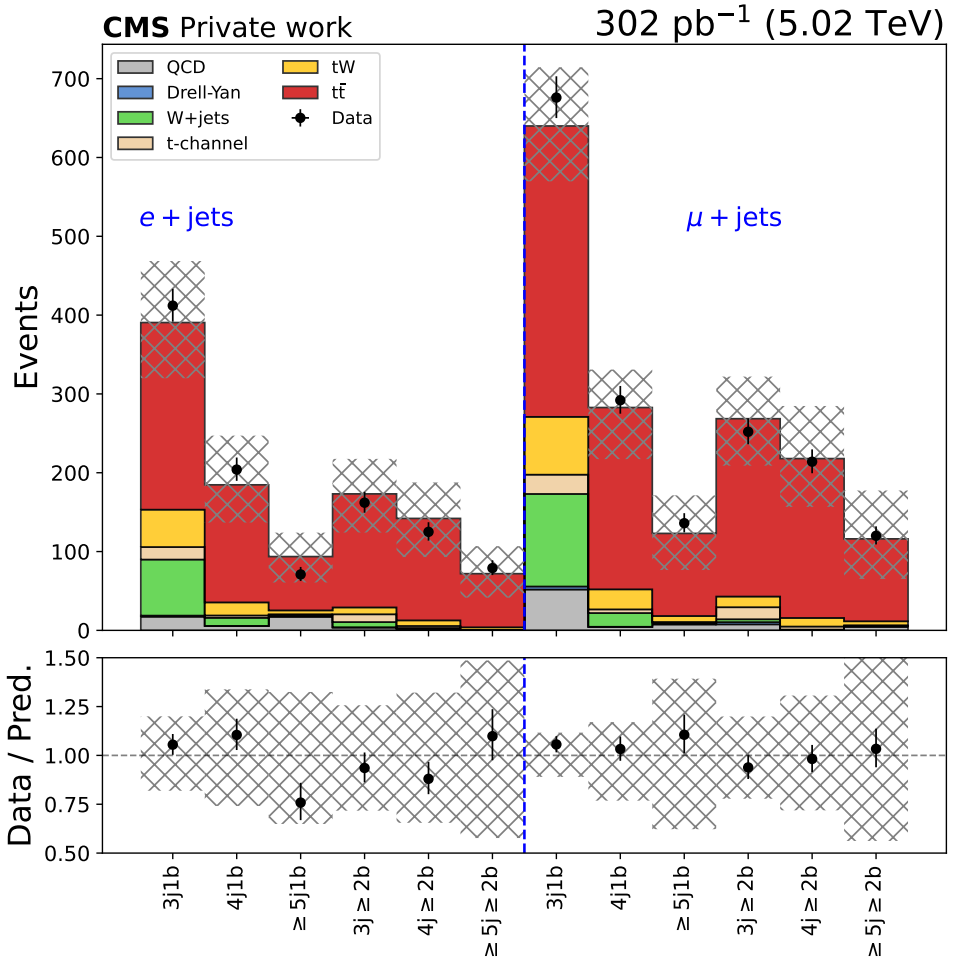

FIGURE 4.15: Number of observed and predicted events in the different categories of the analysis.

Simulated samples have an average pileup of 2, same as data, so no pileup reweighting is applied. The uncertainty on the inelastic pp collision cross section used to estimate the number of pileup events barely modifies the average pileup of 2, so its effect is not considered in the analysis. The rest of corrections described in Section 3.3 are applied to MC events. Lepton SFs are computed using the tag-and-probe method for this special dataset. Due to the low luminosity of this dataset, the trigger efficiency is measured in data and MC for muons and electrons independently using a cross-trigger technique with $e^{\pm}\mu^{\mp}$ events in two η bins and using the muon trigger and the electron trigger of the analysis as orthogonal triggers. The electron trigger efficiency is around 85% and the muon trigger efficiency around 95%. The b-tagging SFs derived for the 2017 dataset at 13 TeV are used assuming that the detector conditions are the same. Finally, L1 prefire weights are also applied in order to suppress the prefiring issue of the ECAL L1 trigger.

4.3.2.1 QCD background estimation

It is common to employ data-driven techniques to estimate QCD backgrounds in analysis with high number of jets in the final state to avoid modelling complications. In general, this methods consist of defining a control region (CR) independent of the signal region, extracting there the QCD contribution from data and then extrapolating it to the signal region.

In this case, the CR is defined as the event selection (signal region or SR) but inverting the mini-isolation and lepton MVA requirements presented in Table 4.6. This CR was chosen in order to reduce contamination from $t\bar{t}$ and W+jet events while maintaining a significant number of events in the sample for the estimation of the QCD background. The QCD events are estimated in the CR as the number of events in data ($N^{CR}(obs)$) minus the number of events coming from non-QCD MC ($N^{CR}(MC)$).

To estimate the amount of QCD events in the signal region ($N^{SR}(QCD)$), an extrapolation factor is calculated from the CR region. The factor is calculated using events with $p_T^{miss} < 20\,GeV$, where a large contribution of QCD background and low signal contamination are expected. The following formula is then used to calculate the QCD estimation in the signal region:

$$N^{SR}(QCD) = \Big(N^{CR}(obs) - N^{CR}(MC)\Big) \times \frac{N^{SR}_{low-p_T^{miss}}(obs) - N^{SR}_{low-p_T^{miss}}(MC)}{N^{CR}_{low-p_T^{miss}}(obs) - N^{CR}_{low-p_T^{miss}}(MC)}. \quad (4.2)$$

The estimation is done separately for each of the 12 categories of the analysis. A normalization uncertainty on the QCD background is also estimated separately by varying the requirement in p_T^{miss} ± 5 GeV and calculating the normalization factor. The result is that a flat normalization uncertainty of 30% covers most of the variations in the normalization factor for all the categories and it is assigned to the QCD estimation in the analysis.

The electron and muon p_T distributions for observed data and predicted events are shown in Figures 4.17 and 4.16 in the different categories of the analysis. Data distributions are well described by the SM prediction in all the categories.

4.3.3 Analysis strategy

A maximum-likelihood fit is performed to determine the best value of $\sigma_{t\bar{t}}$ and, for that, it is necessary to use the distribution of a variable with a good discrimination between signal and background. Different kinematic variables were tested to be used in the fit for each category such as the invariant mass of the jets, the difference in distance

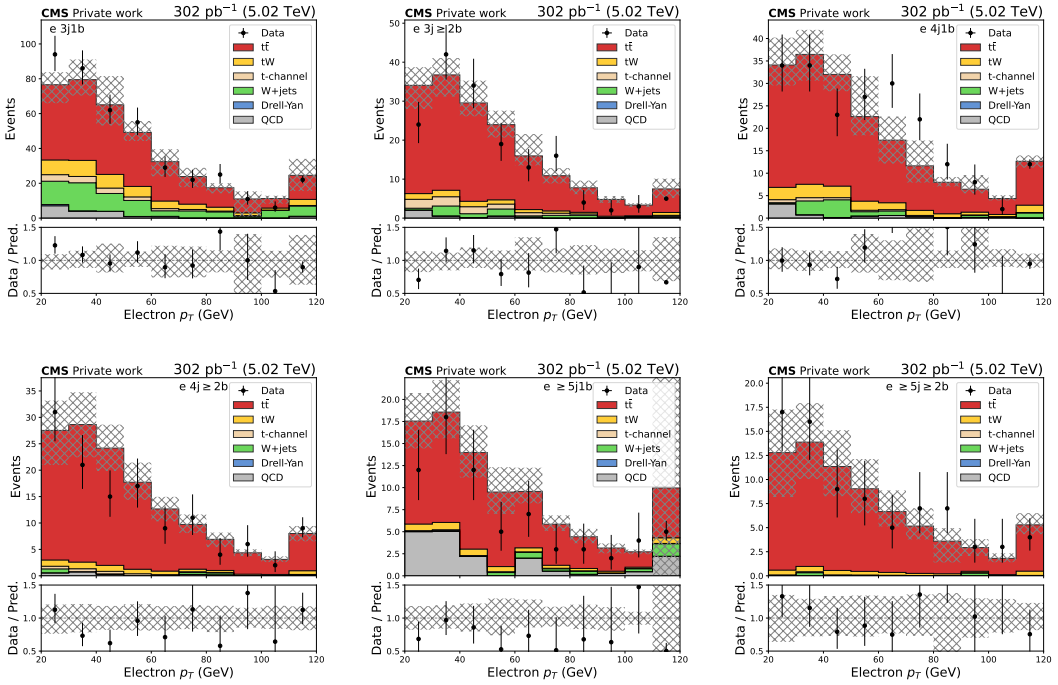

FIGURE 4.16: Electron p_T distribution for data and MC prediction in the different categories of the analysis for the e+jets final state.

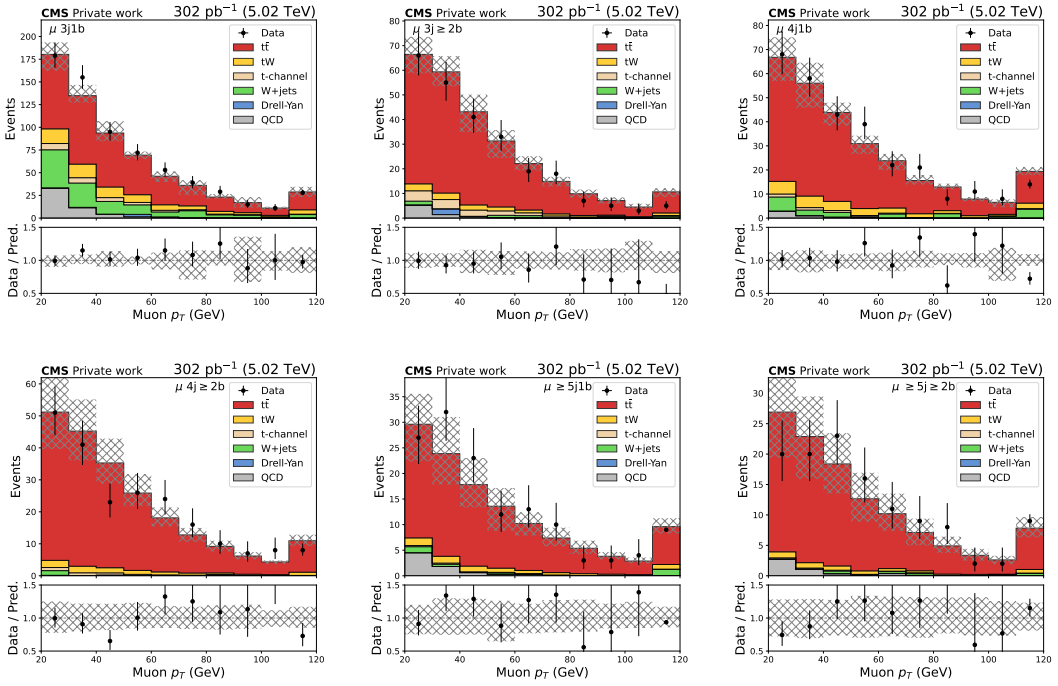

FIGURE 4.17: Muon p_T distribution for data and MC prediction in the different categories of the analysis for the μ+jets final state.

parameter, ΔR, between two jets, or the minimum and median value of those variables. Finally, the median $\Delta R(j, j)$ distribution was selected motivated by the fact that it is the most sensitive to the $t\bar{t}$ production mechanism and minimizing the statistical and

systematic uncertainties. This distribution is used in the fits in every region except in the 3j1b category.

The 3j1b category contains an important amount of signal events that would contribute to the cross section measurement, however, we expect a large number of W+jets events in this category. Therefore, a MVA is trained to separate the $t\bar{t}$ signal from the W+jets background. Several input variables were tested, and those with the most discrimination power were selected:

- H_T: The scalar sum of the p_T of all the jets in the event.
- $p_T(j_0)$: The p_T of the leading jet.
- $m(u,u)$: The invariant mass of the two non-b-tagged jets (u).
- $\Delta R(u,u)$: The ΔR between the two non-b-tagged jets.
- $m_{min}(j,j)$: The minimum invariant mass of all possible combinations of two jets.
- $\Delta R_{med}(j,j)$: The median ΔR between all possible combinations of two jets.
- $\Delta R(\ell,b)$: The minimum ΔR between the lepton and the b-tagged jet.
- $m(\ell,b)$: The invariant mass of the lepton and the b-tagged jet.

Normalized distributions of the input variables for signal and background events are shown in Figure 4.18, where a good discrimination is found. The data and MC prediction distributions of these variables in the signal region can be found in Figure 4.19. In all of them, SM prediction describes reasonably well the data.

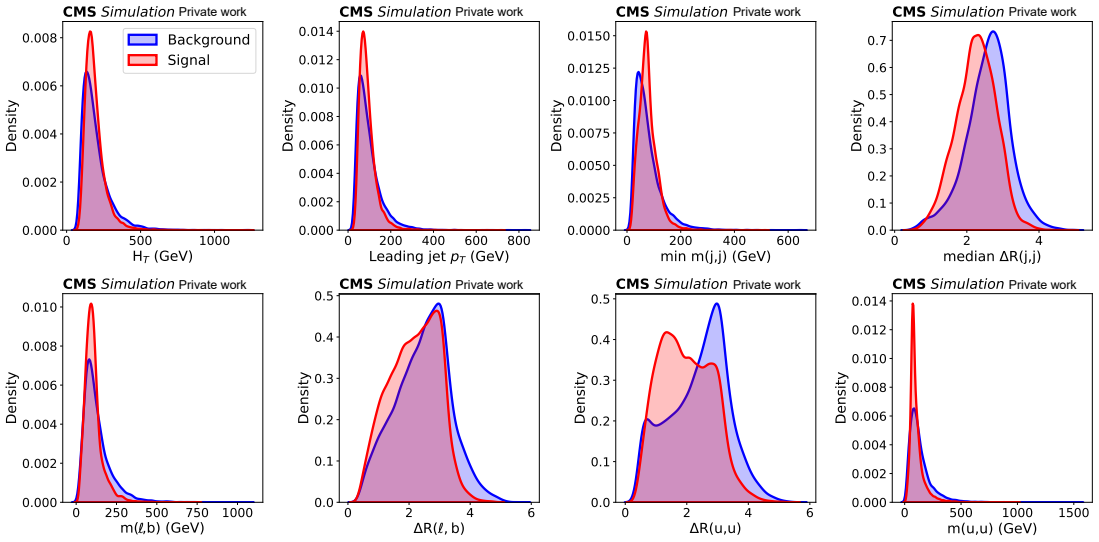

FIGURE 4.18: Normalized distributions of the MVA input variables in blue for signal events and in red for background events.

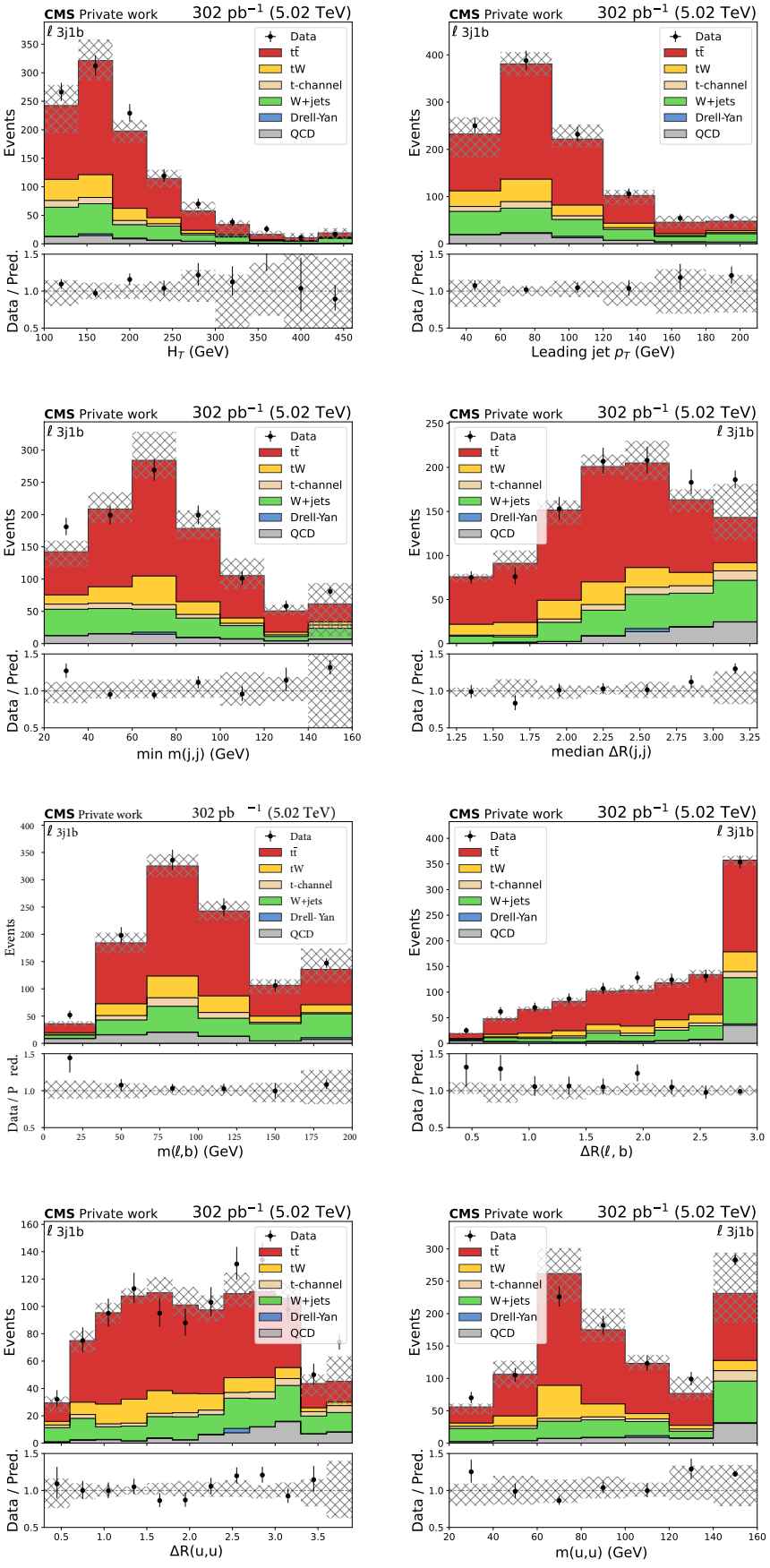

FIGURE 4.19: Distribution for data and MC prediction of the input variables used for the training of the MVA in the 3j1b category.

Several ML models were tested such as BDTs or neural networks, and the best performance in this particular classification task is achieved by a random forest [92]. This model combines multiple decision trees to make predictions. It introduces randomness by using random subsets of the training sample (bootstrapping) and random feature selection when training each decision tree. The final prediction is made by averaging or voting on the predictions of individual trees. Random forests are very robust and have a good generalization performance, avoiding overfitting problems. That is why it is the model that works best in this task, in which the biggest challenge was the class-imbalance: $\approx$ 100000 $t\bar{t}$ events and $\approx$ 10000 W+jets events. The $t\bar{t}$ and W+jets samples used for the training are independent of those used in the rest of the analysis. They are then split in 70% for training and 30% for testing. The model is trained using the Scikit-learn package [93]. The parameters of the model (hyperparameters) are optimized to give the highest accuracy as possible without overfitting: 500 trees with maximum depth of 6.

The receiver operator characteristic (ROC) curve together with the area under the curve (AUC) for the train and test samples are shown in Figure 4.20 (left). It shows the proportion of correctly predicted signal cases out of all actual signal cases (true positive rate or sensitivity) in front of the proportion of actual background incorrectly predicted as signal cases out of all actual background cases (false positive rate). A perfect classifier has an AUC of 1, while a random classifier has an AUC of 0.5. A reasonable performance is obtained with an AUC of about 0.77. Figure 4.20 (right) shows the MVA score distribution, which is the probability of an event to be classified as signal, for signal and background samples. A good agreement between the results with the test and train datasets is observed, showing a negligible amount of overfitting in the model.

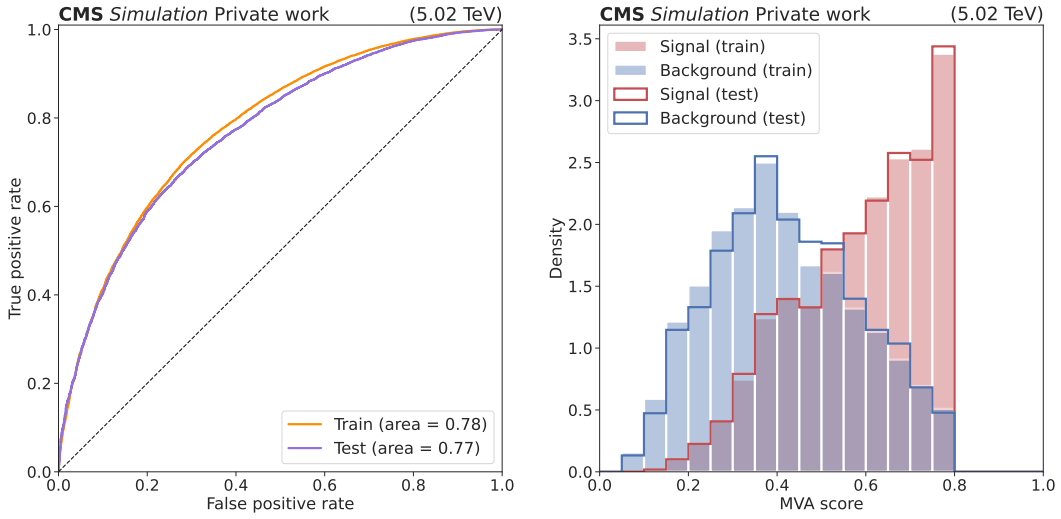

Figure 4.20: ROC curve of the model in orange for train dataset and in purple for test dataset (left), and MVA score for signal events in blue and for background events in red (right).

The MVA score distribution for data and predicted events is shown in Figure 4.21 for the e+jets and μ+jets final states. It can be observed how $t\bar{t}$ events populate the last bins of the distribution, while the proportion of W+jets is higher in the first bins.

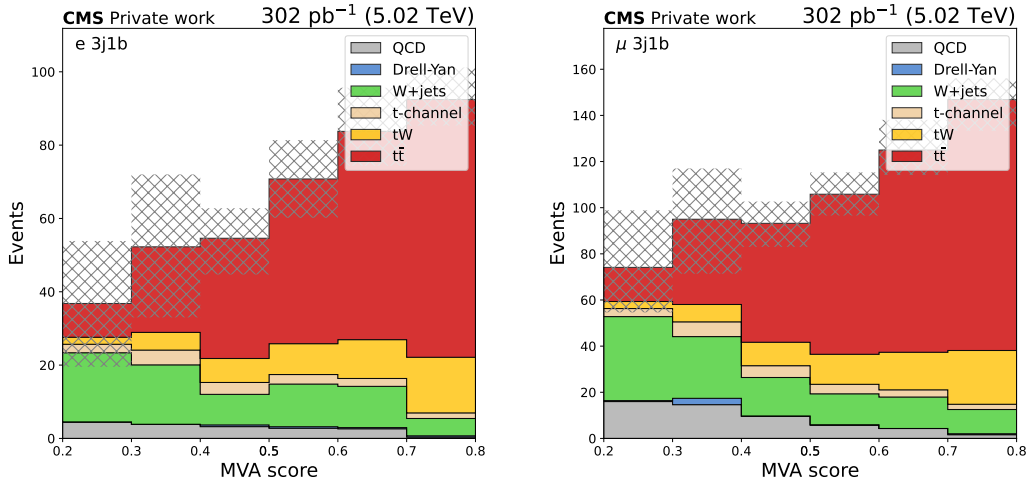

FIGURE 4.21: MVA score for data and predicted events in the e+jets (left) and μ+jets final states (right).

4.3.4 Results

This analysis is currently under review by the CMS Collaboration, and to prevent introducing bias in the measurement, only expected results are presented. The cross section is obtained by fitting the median $\Delta R(j,j)$ distribution in the different categories and the MVA score distribution in the 3j1b category. Figure 4.22 shows these distributions for e+jets and μ+jets final states.

Systematic uncertainties on the lepton, trigger and b-tagging efficiencies as well as jet and $p_{\mathrm{T}}^{\mathrm{miss}}$ related uncertainties are taking into account in the measurement. Modelling uncertainties described in Section 4.2.5.2 are also included. The normalization uncertainty is taken to be 5.6% [94] in the tW and 2% [95] in the t-channel, based on the theoretical uncertainties. A normalization uncertainty of 30% is applied to the QCD background, as discussed in Section 4.3.2.1, to account for the systematic uncertainties in the QCD estimation. A 20% normalization uncertainty is assumed on the W+jets and DY backgrounds. Finally, the uncertainty in the measurement of the integrated luminosity is estimated to be 1.9% [96].

A maximum-likelihood fit is done simultaneously to the distributions in the 12 categories. The signal strength (r), defined as the ratio between the measured and the predicted signal cross sections, is extracted and uncertainties are constrained. The best fit value of the signal strength is computing by minimizing the negative log-likelihood (NLL) function, introducing the systematic uncertainties as nuisance parameters (NPs)

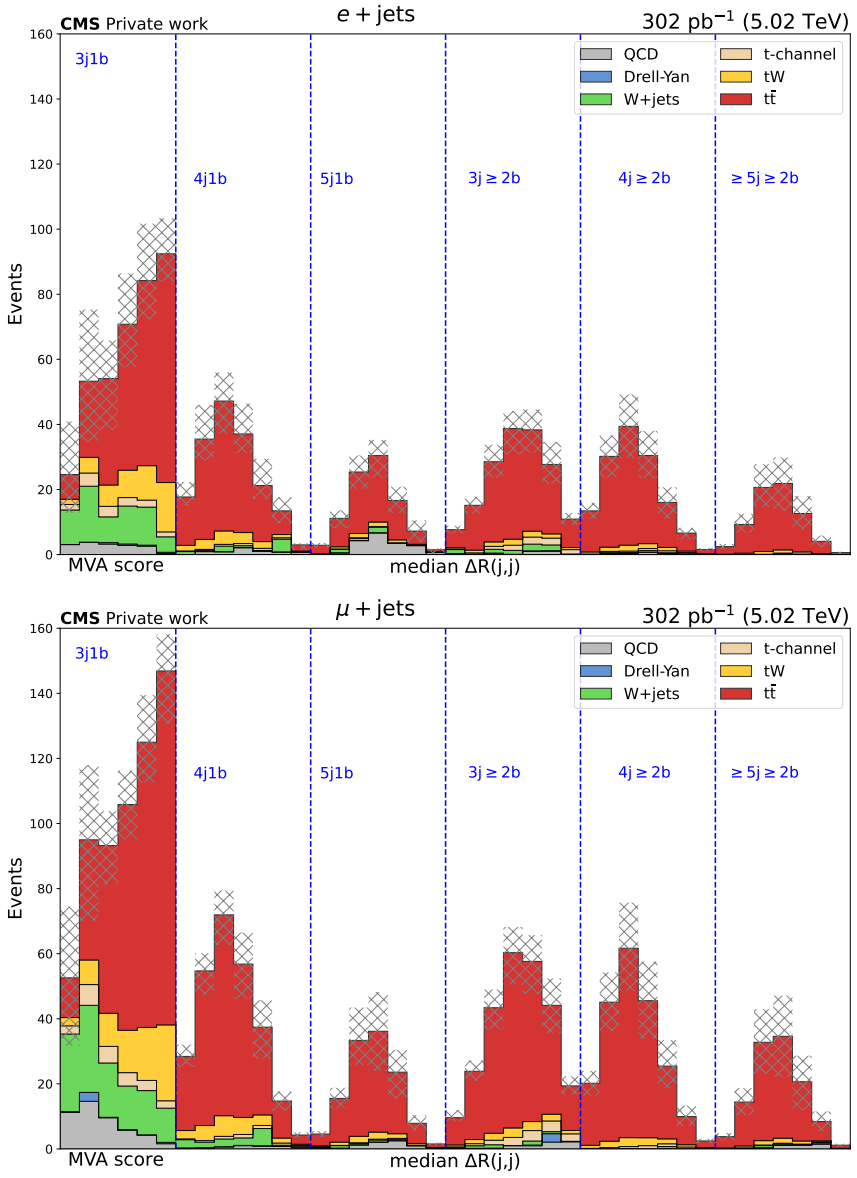

FIGURE 4.22: Final distributions for the e+jets (top) and μ+jets (bottom) final states: MVA score bins for the 3j1b category and $\Delta R(j,j)$ bins for the other categories.

modelled using Gaussian distributions. The signal strength is found to be r =1.0. The uncertainty on the cross section is extracted by finding the r values where the twice the difference in the NLL (-2ΔNLL) curves cross one. The scan of the fit is shown in Figure 4.23. The systematic uncertainty is expected to be $^{+4.4}_{-4.1}$% and the statistical uncertainty is expected to be 2.5%, resulting in a total expected uncertainty of 5%. This result is a significant improvement with respect to the previous measurement in this final state, which presents a total uncertainty of 13%. Moreover, this result is no longer dominated by statistical uncertainty, as was the case in the previous measurements.

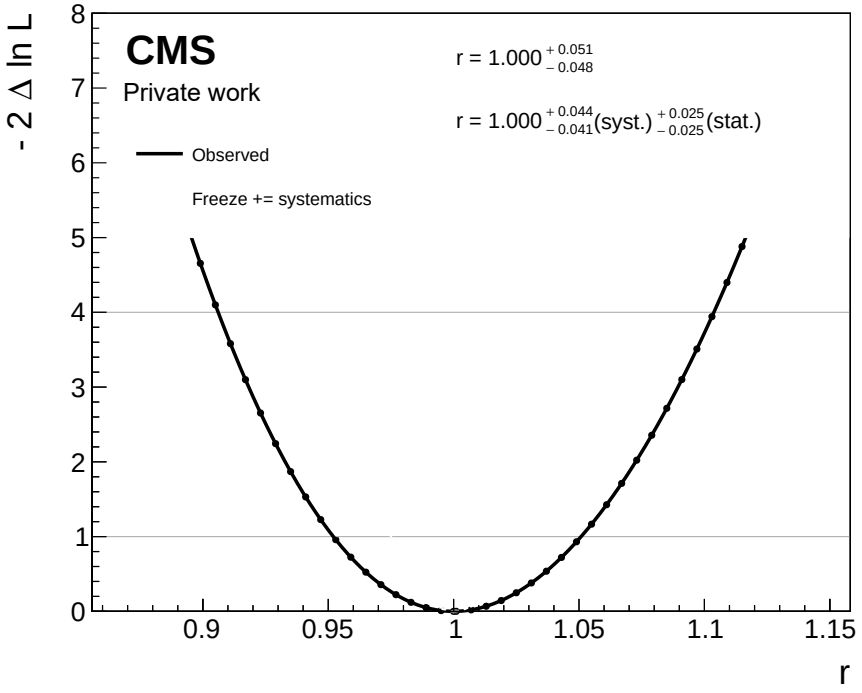

FIGURE 4.23: -2ΔNLL values of the likelihood fit.

4.3.4.1 Combination with dilepton result

The semilepton analysis presented above is combined with the results in the dilepton measurement with the 2017 dataset, detailed in Reference [2]. The dilepton analysis follows a cut-and-count strategy similar to the the one described in Section 4.2.1. For the combination, the number of events of each process and the uncertainties are added to the semilepton final distribution (shown in Figure 4.22) as a new category.

The sources of systematic uncertainty are considered as fully correlated between the two analyses, with the exception of few uncertainties treated as uncorrelated either because are estimated differently, as jet energy scale and PDFs, or they were not considered in the dilepton analysis, as b tagging, $p_{\mathrm{T}}^{\mathrm{miss}}$ related uncertainties, QCD, and t-channel normalization.

A maximum-likelihood fit is performed to the final distribution with the extra bin from the dilepton analysis and the signal strength is found to be r =1. The measured systematic uncertainty is expected to be $^{+4.2}_{-3.9}$% and the statistical uncertainty is expected to be 2.4%, resulting in a total expected uncertainty of 4.7%. This is the most accurate measurement of the $\mathrm{t\bar{t}}$ cross section at 5.02 TeV to date by CMS, almost halving the precision of the previous combined measurement, which is 8.4%.

4.4 Diboson production cross section at $\sqrt{s}$ = 5.02 TeV

The D0 and CDF Collaborations have measured the diboson production cross section in $p\bar{p}$ collisions at a centre-of-mass energy of 1.8 TeV, and CMS and ATLAS Collaborations have measured them in pp collisions at 7, 8 and 13 TeV. The result presented in this section is the first measurement of the diboson production done at $\sqrt{s} = 5.02$ TeV, finally reducing the gap between 1.8 TeV and 7 TeV. This measurement is very interesting as the 5.02 TeV dataset is characterized by the relatively low PU contribution. Additionally, the overall energy available for the hard-scattering process is constrained and, as a consequence, the p_T thresholds of the leptons can be relaxed allowing the study of bosons with a lower momentum than the studied at 13 TeV.

The same dataset described in the previous section (Section 4.3) at 5.02 TeV with a integrated luminosity of 302 pb^{-1} is used to measure the inclusive production cross section of W^+W^-, $W^\pm Z$, and ZZ processes with purely lepton decays [3].

These processes have a small branching ratio, so the main challenge of the analysis is to have clean signal regions with this statistically limited dataset. Electrons and muons are the key objects and their efficient identification and selection is crucial. In this special energy regime with low contamination from PU interaction, unusual requirements can be implemented such as relaxing the standard requirement in the lepton p_T to select a higher number of events.

The MC samples used are the same as those used in the $t\bar{t}$ semilepton measurement, with the addition of the VV signal samples, generated at NLO in QCD using POWHEG (v2). The signal cross sections are scaled from NLO to NNLO using k-factors extracted from MATRIX calculations [97]. The lepton and jet corrections as well as the systematic uncertainties applied are also the same as the ones described in Subsections 4.3.2 and 4.3.4.

The analysis strategy and the obtained results are presented in detail in the following subsections.

4.4.1 Object selection

The strategy of the analysis is to select prompt leptons coming from the W and Z bosons decays ($W^\pm \rightarrow \ell^\pm\nu$ and $Z \rightarrow \ell^+\ell^-$). For that, the lepton MVA described in Subsection 4.4.1 is used to reduce the nonprompt lepton background contribution from heavy-flavor quark decays and jet misidentified as leptons.

Two lepton categories are defined: loose and tight leptons. The loose identification is used as an initial preselection and consist of basic requirements, similar to the ones used in the $t\bar{t}$ analysis (Table 4.6) but without the lepton MVA and with a more

relaxed requirement on the mini-isolation of < 0.4. Apart from that, loose electrons are required to pass the loose WP of a BDT discriminant [68] designed to reject electrons coming from misreconstructed jets. Muons are required to pass the medium WP of the cut-based ID [69], which is designed to be highly efficient for selecting prompt muons and muons from heavy quark decays using variables such as the $\chi^2/ndof$ of the global muon track fit or the compatibility of the inner track with the muon segment. Electrons with $|\eta| < 2.5$ and muons with $|\eta| < 2.4$ are preselected if they also have a $p_T > 8\,\text{GeV}$.

Then, the tight lepton criteria is defined starting from the loose leptons but with additional cuts in mini-isolation (< 0.085 for electrons and < 0.325 for muons) and in the lepton MVA (> 0.125 for electrons and > 0.55 for muons) to reduce the nonisolated lepton background contribution.

For jets, the selection criteria is the same as in the $t\bar{t}$ analysis at 5.02 TeV: tight ID, $p_T > 25\,\text{GeV}$ and $|\eta| < 2.4$, removing the jets that overlap with the selected leptons.

4.4.2 Event selection

The trigger strategy is that events at least pass one single lepton trigger with p_T thresholds of 17 GeV for electrons or 12 GeV for muons, the lowest available p_T thresholds.

Several signal regions are defined, each specifically optimized for the processes and final states to be measured in this analysis. All of them have in common that they require events with at least two loose leptons with a minimal invariant mass of any lepton pair higher than 12 GeV. The SRs for each VV process are defined below.

4.4.2.1 WW signal region

For the W^+W^- measurement one signal region is defined as follows. Exactly two tight leptons with opposite sign of the electric charge and different-flavour are required, the leading with $p_T > 20\,\text{GeV}$ and the subleading with $p_T > 10\,\text{GeV}$. To reject DY events, the p_T of the dilepton system ($p_T(\ell\ell')$) is required to be greater than 20 GeV and their azimuthal separation ($\Delta\phi_{\ell\ell'}$) greater than 2.8, requirements that can be justified by Figure 4.24, which shows the distribution of these variables before applying the requirements and it can be seen as signal events populate the region with high values of the $p_T(\ell\ell')$ and $\Delta\phi_{\ell\ell'}$ distributions. The transverse mass of the lepton and p_T^{miss} pairs is required to be greater than 20 GeV to further reduce nonprompt lepton and DY backgrounds. In addition, a veto to events with jets is also applied to reject $t\bar{t}$ events.

The dilepton p_T and the invariant mass of the leptons distributions for events passing the SR requirements are shown in Figure 4.25, where the MC prediction describes

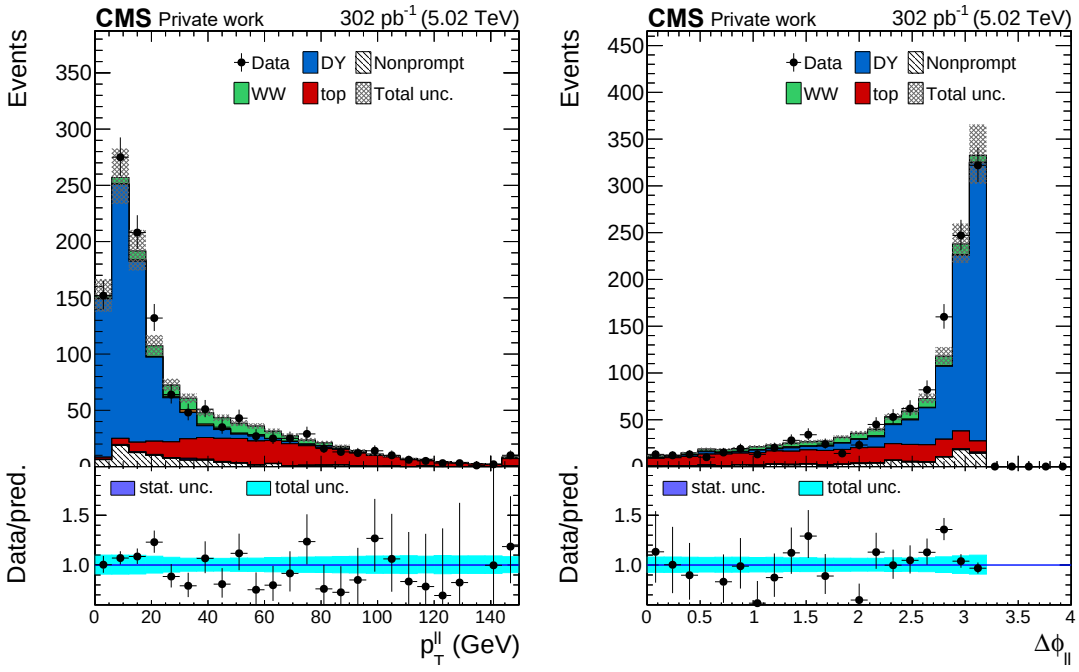

FIGURE 4.24: The dilepton p_{T} on the left and the dilepton azimuthal separation on the right for data and MC prediction in a WW relaxed region defined as the signal region but without $p_{\mathrm{T}}(\ell\ell')$, $\Delta\phi_{\ell\ell'}$ and transverse mass of the leptons and $p_{\mathrm{T}}^{\mathrm{miss}}$ requirements.

accurately the distributions observed in data and a great purity of the final selection can be observed.

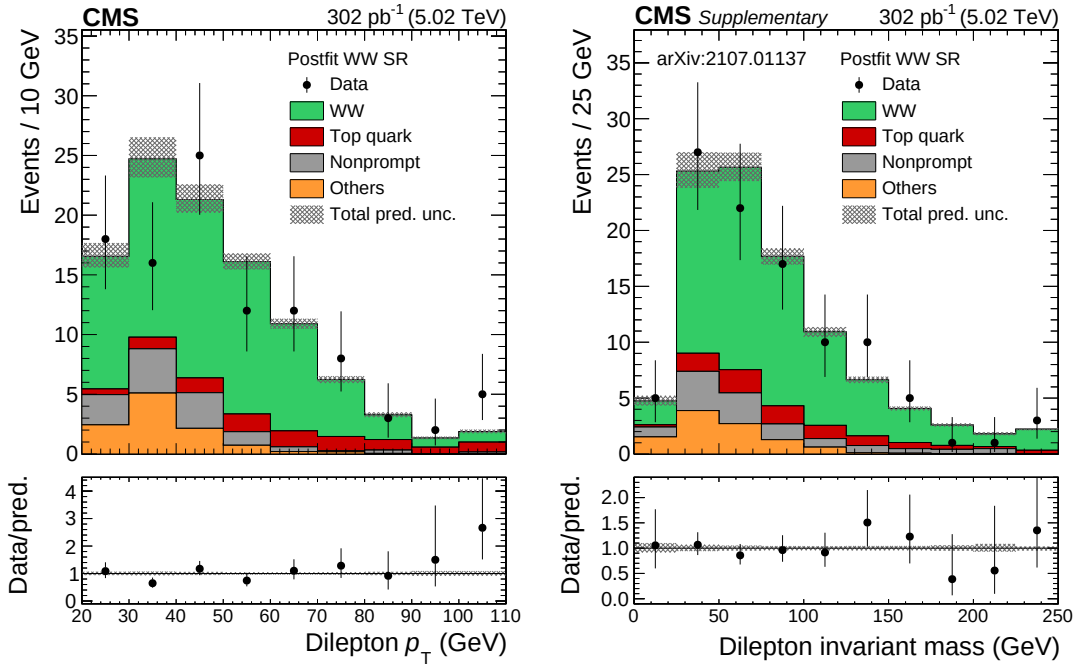

FIGURE 4.25: The dilepton p_{T} on the left and the invariant mass of the leptons on the right for data and MC prediction in the WW signal region. Events from DY, conversions, and diboson processes are grouped into the 'Others' category. The WW contribution is scaled to the measured cross section (see Section 4.4.3) [3].

This is a very pure signal region, where the main background contributions come from $t\bar{t}$ and nonprompt leptons. Both the signal and most of the backgrounds ($t\bar{t}$, single top, DY, VV and conversions) are estimated from MC. The photon conversions are estimated from DY, W+jets, diboson and top MC samples where at least one of the leptons in the event is matched at generator level to a photon.

The nonprompt lepton background is estimated using the tight-to-loose data-driven method [98]. In this case, the method selects events which pass the SR selection criteria except that the leptons are required to pass the loose instead of the tight object selection criteria. This is the application region (AR) of the method. The estimation of the nonprompt background in the signal region is obtained by applying appropriately chosen weights, known as fake-rates, to the observed events selected in the AR. The applied fake-rates are derived from a $t\bar{t}$ MC sample in a lepton+jet region (QCD enriched) separately for electrons and muons, and are parameterized as function of η and p_T of the lepton. The fake-rates measure the probability for a loose selected lepton to pass the tight criteria. The contamination from irreducible backgrounds with prompt leptons and conversions in the AR is taken from simulation and subtracted from the nonprompt estimation to avoid double-counting in the SR. The nonprompt estimation is compared with a direct MC prediction as a cross-check of the method and an uncertainty is extracted from the comparison of the estimation and the MC prediction, that is 30% for electrons and 15% for muons.

4.4.2.2 WZ signal regions

In the $W^{\pm}Z$ measurement two exclusive signal regions are defined, one with two muons with the same electric charge (2μSS) and another with three leptons (3ℓ).

For the 2μSS category the two muons with the same electric charge are required to pass the tight identification criteria, the leading lepton p_T to be higher than 20 GeV and the subleading lepton p_T higher than 10 GeV. Also, the leptons in this category must pass the tight charge requirement, which corresponds to a relative uncertainty in the curvature of the muon track lower than 20%. In addition, a minimal requirement of $p_T^{miss} > 25$ GeV is included and also a veto to events with jets to reduce the nonprompt background contribution. This is a very clean signal region with a small contribution of nonprompt background, which is estimated following the same procedure as in the WW selection. The rest of the background contributions are estimated from MC.

For the 3ℓ category, exactly three loose leptons with at least one pair with opposite sign of the electric charge and same flavor (OSSF) are required. To exploit the characteristic kinematics of on shell $W^{\pm}Z$ production, an assignation algorithm is applied to tag the two leptons from the Z boson decay (ℓ_{Z1} and ℓ_{Z2}) and that of the W boson decay (ℓ_W). If only one OSSF lepton pair is found in the event, the leptons corresponding to it are tagged as ℓ_{Z1} and ℓ_{Z2} while the different flavor one is tagged as ℓ_W. If multiple OSSF

pairs are found, the OSSF lepton pair with the closest invariant mass to the Z boson is selected for the ℓ_{Z1} - ℓ_{Z2} pair. Once leptons have been tagged, additional selection criteria are applied in order to increase the purity of $W^{\pm}Z$ events. The invariant mass of the ℓ_{Z1} and ℓ_{Z2} leptons has to be close to the Z boson mass, $|m_{\ell_{Z1}I,\ell_{Z2}} - 91.2| < 30\,\text{GeV}$. The two leptons with the same sign of the electric charge are required to pass the tight lepton requirements and the other just the loose lepton criteria to increase the efficiency. The ℓ_W is required to have a p_T >20 GeV and the invariant mass of the three lepton system is required to be $m_{\ell_{Z1},\ell_{Z2},\ell_W} > 100\,\text{GeV}$.

In the 3ℓ signal region, the main background contribution comes from Z+jets events. All backgrounds are estimated from MC, including the nonprompt lepton background, which is well modelled in the three lepton final state. It is validated in a control region defined by inverting the tight lepton criteria and a 20% of normalization uncertainty is estimated.

Figure 4.26 shows the observed data and MC predictions in the 3ℓ WZ signal region of the W boson transverse mass constructed with ℓ_W and p_T^{miss}, and the dilepton p_T of ℓ_{Z1} and ℓ_{Z2}. The tails of the distributions present a very clean signal region, but the number of events there is small. The 2μSS signal region was designed to improve the sensitivity, as it is a very pure region, but only provides around 3 signal events more.

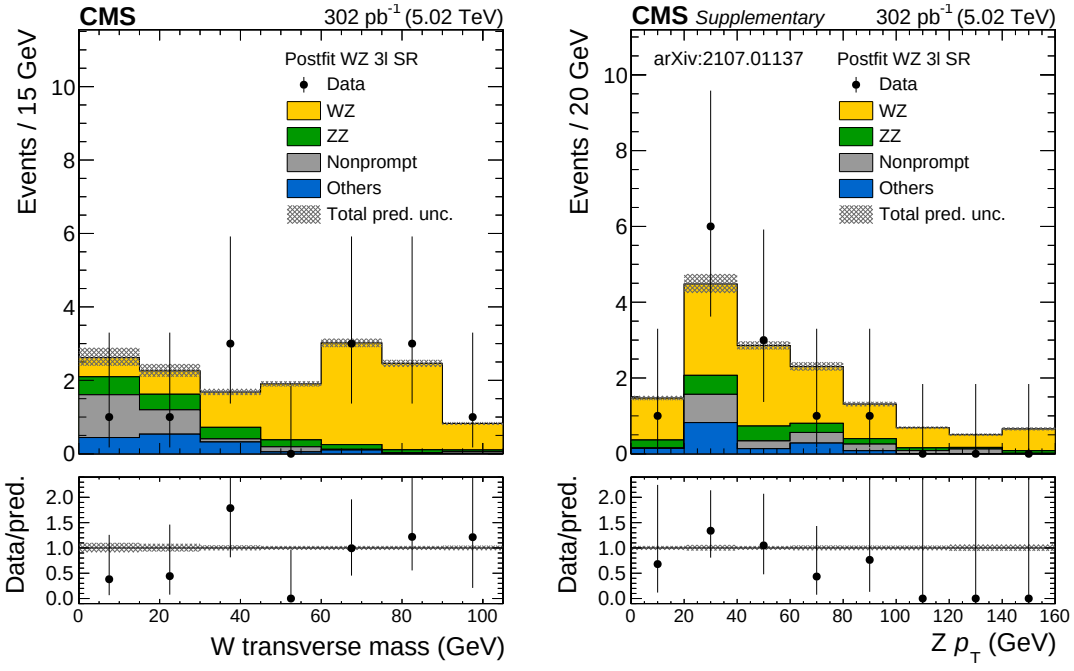

FIGURE 4.26: The W boson transverse mass on the left and the Z boson p_T on the right for data and MC prediction in the WZ 3ℓ signal region. Events from conversions, and DY processes are grouped into the 'Others' category. The WZ contribution is scaled to the measured cross sections (see Section 4.4.3) [3].

4.4.2.3 ZZ signal regions

The ZZ measurement also has two exclusive categories, one with four leptons (4ℓ) and another with two leptons ($2\ell 2\nu$).

For the $2\ell 2\nu$ signal region, exactly two OSSF tight leptons are required, the leading lepton with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$, the subleading with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$, and the invariant mass of them has to be close to the Z boson mass, $|m_{\ell,\ell} - 91.2| < 10\,\mathrm{GeV}$. Events with jets are rejected to reduce $\mathrm{t\bar{t}}$ background contamination. To suppress WW a DY events, the presence of the Z boson decaying into the neutrino pair is used by requiring the axial $p_{\mathrm{T}}^{\mathrm{miss}}$ in the event ($-p_{\mathrm{T}}^{\mathrm{miss}} \times cos(\Delta\phi(p_{\mathrm{T}}^{\mathrm{miss}}, p_T^{\ell\ell}))$ [99]), which expresses the projection of the transverse momentum of the neutrino pair of the invisibly decaying Z boson onto the direction of the transverse momentum of the other Z boson decaying to charged leptons, to be higher than 50 GeV. The relative difference between the $p_{\mathrm{T}}^{\mathrm{miss}}$ and the $p_T^{\ell\ell}$ is required to be smaller than 0.3.

This region has an important background contribution from WW process, but it still helps to the measurement by providing a nonnegligible amount quantity of signal events. The data and predicted MC distributions of the $p_{\mathrm{T}}^{\mathrm{miss}}$ and the dilepton invariant mass are shown in Figure 4.27.

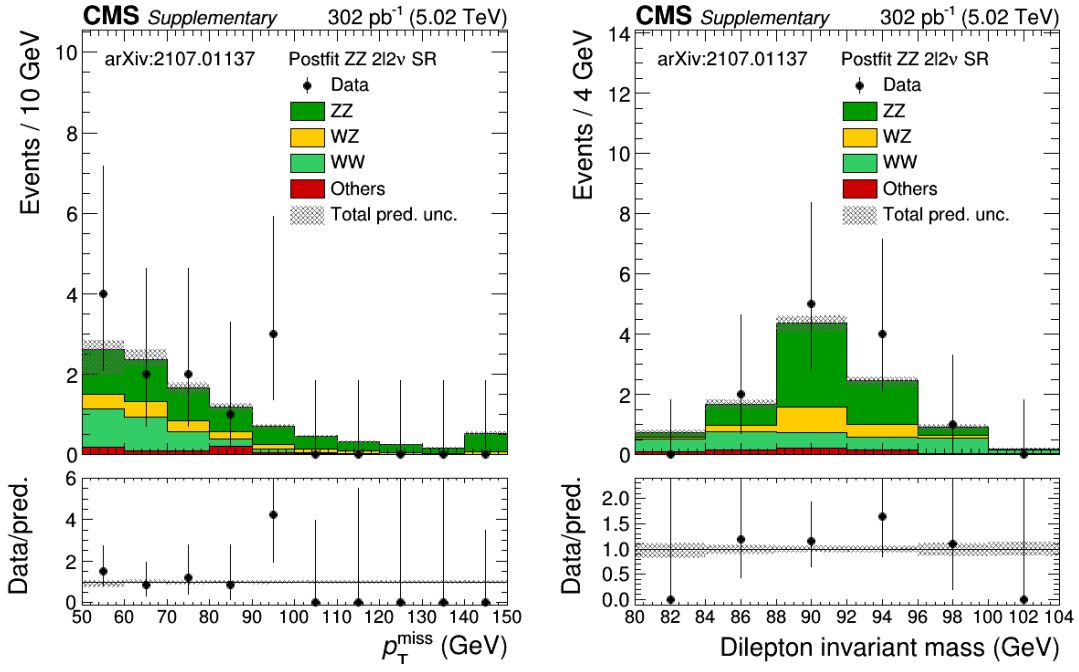

FIGURE 4.27: The $p_{\mathrm{T}}^{\mathrm{miss}}$ on the left and the dilepton invariant mass on the right for data and MC prediction in the ZZ $2\ell 2\nu$ signal region. Events from top, and DY processes are grouped into the 'Others' category. The ZZ contribution is scaled to the measured cross sections (see Section 4.4.3) [3].

To improve the sensitivity, the 4ℓ signal region is defined by requiring exactly 4 loose leptons with invariant mass of all the possible lepton pairs greater than 12 GeV. With

just these requirements, a very clean region is achieved with a very small contribution from nonprompt leptons, estimated from MC. In Figure 4.28 the distributions of the p_T and the invariant mass of the ZZ system are presented, where it can be seen that practically all the events passing the signal region are from signal. The caveat with this region is the low amount of events surviving the selection, with less than 3 signal events. That is why the $2\ell 2\nu$ region is also used, which although it has more background contribution, provides more signal events and both are compensated.

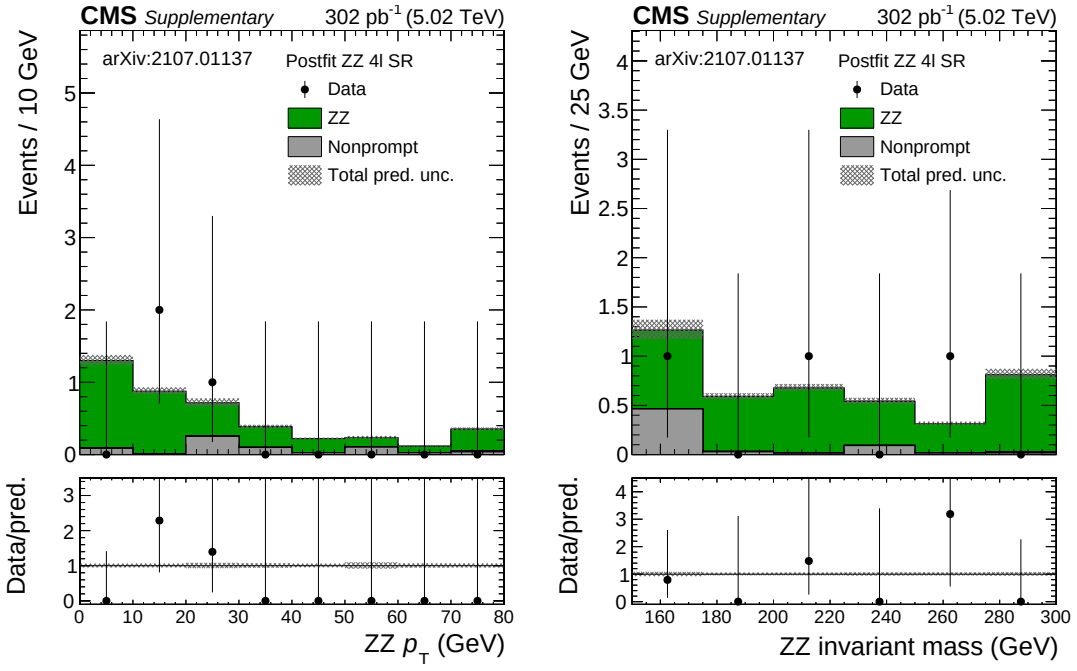

FIGURE 4.28: The p_T of the ZZ system on the left and the invariant mass on the right for data and MC prediction in the ZZ 4ℓ signal region. The ZZ contribution is scaled to the measured cross sections (see Section 4.4.3) [3].

4.4.3 Results

The cross section measurements are completely dominated by the statistical limitation of the size of the dataset, therefore the impact of the different sources of systematic uncertainties is small. The background normalization uncertainties are 10% for $t\bar{t}$ and DY, 30% for conversions, 20% for VV, and 30% (20%) for nonprompt electrons and 15% (20%) for nonprompt muons in the signal regions with 2 leptons (more than 2 leptons). A summary of all the uncertainties considered, described in the previous sections, with their relative effect on each measurement is shown in Table 4.8.

The number of observed events in data, and expected signal and background events in each of the signal regions together with their statistical and systematic uncertainties are summarized in Table 4.9.

Source	W^+W^-	$W^\pm Z$	ZZ
Lepton efficiencies	1% (2%) for μ (e)	1-3%	∼2%
Trigger efficiencies	5% ($p_T < 15$ GeV), 1%	3% (2μSS), <0.5% (3ℓ)	1% ($2\ell 2\nu$), <0.5% (4ℓ)
JES	∼1-2%	∼1% (2μss)	∼1% ($2\ell 2\nu$)
JER	∼1-2%	∼1% (2μss)	∼1% ($2\ell 2\nu$)
L1 Prefiring	< 2%	1-2%	1%
Luminosity	1.9%	1.9%	1.9%
PDF+ α_S	< 1%	< 1%	∼2%
μ_R, μ_F scales	< 3%	< 3%	∼3-4%
Nonprompt e	30%	30% (2μSS), 20% (3ℓ)	15% ($2\ell 2\nu$), 20% (4ℓ)
Nonprompt μ	15%	15% (2μSS), 20% (3ℓ)	15% ($2\ell 2\nu$), 20% (4ℓ)
Misidentified charge	–	20% (2μSS)	–
Conversions	30%	30%	30%
Top quark	10%	–	10%($2\ell 2\nu$)
Drell–Yan	10%	–	10%($2\ell 2\nu$)
Other VV	20%	20%	20%($2\ell 2\nu$)

TABLE 4.8: Sources of systematic uncertainties considered in each VV measurement and their relative input values.

	W^+W^-	$W^\pm Z$ 2μSS	$W^\pm Z$ 3ℓ	ZZ $2\ell 2\nu$	ZZ 4ℓ
Nonprompt ℓ	$11.2 \pm 1.3 \pm 3.4$	$0.5 \pm 0.1 \pm 0.1$	$1.2 \pm 0.5 \pm 0.2$	–	$0.5 \pm 0.2 \pm 0.1$
Conversions	$2.7 \pm 0.7 \pm 0.7$	–	$0.8 \pm 0.3 \pm 0.2$	–	–
Top quark	$9.0 \pm 0.1 \pm 0.1$	–	–	$0.5 \pm 0.0 \pm 0.1$	–
Drell–Yan	$1.8 \pm 0.5 \pm 0.2$	–	–	$0.2 \pm 0.3 \pm 0.1$	–
Other VV	$5.6 \pm 1.0 \pm 1.1$	$0.1 \pm 0.0 \pm 0.0$	$1.7 \pm 0.0 \pm 0.1$	$4.3 \pm 0.1 \pm 0.6$	–
Total background	$30.3 \pm 1.9 \pm 3.9$	$0.6 \pm 0.1 \pm 0.1$	$4.0 \pm 0.6 \pm 0.4$	$4.8 \pm 0.3 \pm 0.7$	$0.5 \pm 0.2 \pm 0.1$
Expected signal	$55.2 \pm 0.3 \pm 1.8$	$3.2 \pm 0.8 \pm 0.2$	$14.8 \pm 0.1 \pm 0.6$	$4.0 \pm 0.0 \pm 0.2$	$2.7 \pm 0.0 \pm 0.2$
Observed	101	4	12	12	3

TABLE 4.9: Expected and observed number of events in each of the signal regions. The uncertainties correspond to the statistical and systematic component, respectively.

The analysis strategy is to measure the diboson cross sections using a cut-and-count approach, similar to the one described in Section 4.2 (Eq. 4.1) but extracting the number of signal events, N^{SR}_{signal}, with a likelihood fit to the observed events in the signal regions. For each region, the likelihood fit is done with a single free floating parameter, the signal strength r, corresponding to the normalization of the signal process. In the case of having several signal regions, the measurements are combined into another likelihood fit over all the joint bins defined in the exclusive categories. The number of signal events is extracted from the best fit parameter value.

Then, the cross section is computed as:

$$\sigma_{VV} = \frac{N^{SR}_{signal}}{BR(V) \cdot BR(V) \cdot \varepsilon \cdot L_{int}}. \tag{4.3}$$

The BR depends on the decay of the W or Z boson of each process: $BR(W \to \ell\nu) = 0.3258$ and $BR(Z \to \ell\ell) = 0.1086$ [78]. The efficiency is computed in MC separately for each process as the ratio of the number of events that pass the SR requirements over those that pass a more relaxed region, called total region, without any acceptance

requirement. The total region is different for each process, but all of them are defined at generator level by selecting events with lepton dressed with photons within a cone of $R < 0.1$ and excluding events containing any OSSF lepton pair with invariant mass below 4 GeV. In processes with Z bosons, an additional kinematic requirement of $60\,\text{GeV} < m(\ell_1, \ell_2) < 120\,\text{GeV}$ is imposed to remain on the Z boson peak.

The measured cross section for each diboson process and the theoretical prediction computed with MATRIX at NNLO QCD × NLO EW are:

Process	Estimation	Cross section (pb)
W^+W^-	Measured	$37.0^{+5.5}_{-5.2}\,(\text{stat.})^{+2.7}_{-2.6}\,(\text{syst.})$
	MATRIX	$29.8^{+0.7}_{-0.6}\,(\text{scale})$
$W^\pm Z$	Measured	$6.4^{+2.5}_{-2.1}\,(\text{stat.})^{+0.5}_{-0.3}\,(\text{syst.})$
	MATRIX	$11.3^{+0.2}_{-0.2}\,(\text{scale})$
ZZ	Measured	$5.3^{+2.5}_{-2.1}\,(\text{stat.})^{+0.5}_{-0.4}\,(\text{syst.})$
	MATRIX	$3.9^{+0.1}_{-0.1}\,(\text{scale})$

TABLE 4.10: Measured and predicted diboson cross sections. The predictions are estimated with MATRIX at NNLO QCD × NLO EW.

The three experimental results at 5.02 TeV are in agreement with the predictions, but they are dominated by the statistical uncertainty due to the limited size of the available dataset. These values are shown in Figure 4.29. This figure summarizes the VV production cross section measurements at different centre-of-mass energies from 2 to 13 TeV obtained by CMS [74, 100–106], ATLAS [99, 107–114], CDF [115, 116], and D0 [117–119] Collaborations. It also shows the NNLO QCD x NLO EWK and NLO predictions, computed with MATRIX [97]. All the experimental measurements are consistent with the SM predictions.

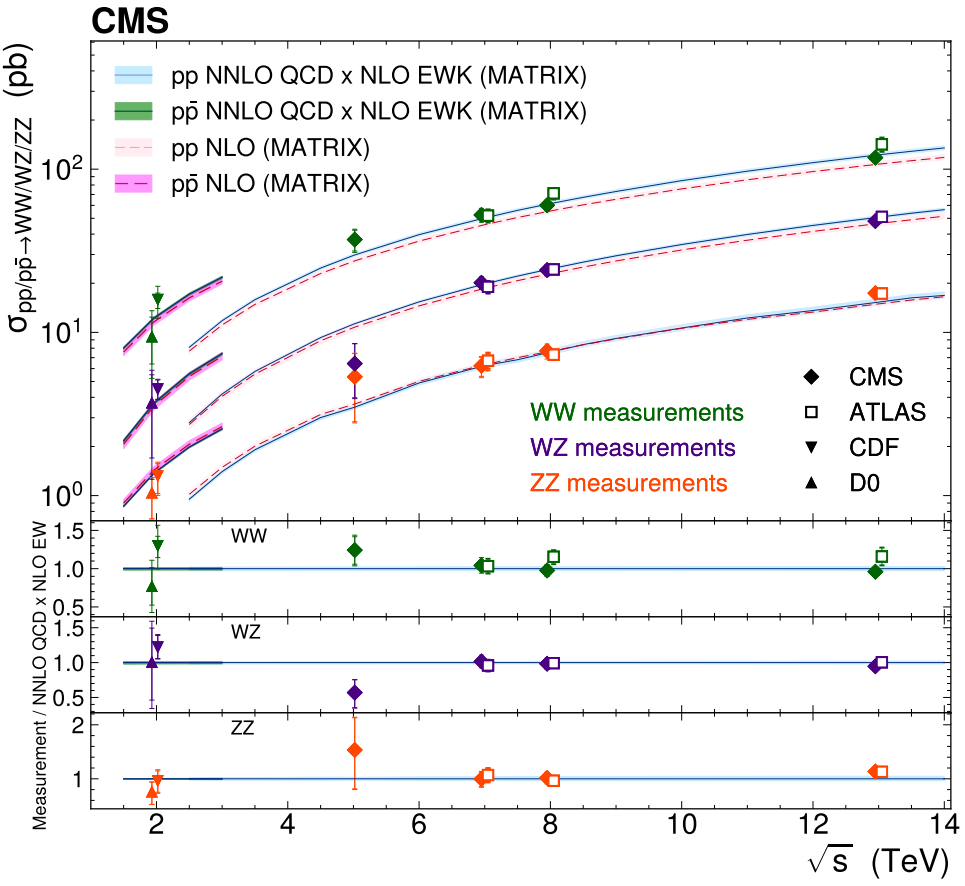

FIGURE 4.29: Summary of LHC and Tevatron measurements of the WW (green), WZ (purple) and ZZ (orange) production cross section as a function of the centre-of-mass energy compared to the NNLO QCD x NLO EWK and NLO predictions [3].

Chapter 5

Beyond Standard Model searches

An important proportion of the analyses being carried out today in particle physics, are searches for new physics, processes not described by the SM. This was one of the main goals for which the LHC was built, in addition to improving the experimental accuracy of the parameters predicted by the SM. In the LHC, BSM processes are expected to be observed for the first time.

Top quarks, as the heaviest known elementary particles, play a crucial role in the exploration of physics BSM. Their unique properties, including their large mass and coupling to the Higgs boson, make them an ideal target for the search for new physics.

In the previous chapter I have described my contribution to the SM precision measurements, mainly focused on the $t\bar{t}$ process. The knowledge and expertise gained doing those analyses on $t\bar{t}$ physics place me in an ideal position to search for BSM in scenarios related with top quarks. In this chapter, two distinct searches in which I have participated are presented, both of them using the full Run 2 dataset: a search of the supersymmetric partner of the top quark in Section 5.1, and a search under the EFT framework in Section 5.2.

5.1 Search for supersymmetric partners of the top quark

In this section, a search targeting a signal model in which a pair of stop and anti-stop quarks are produced decaying into a top and anti-top quark pair, and two lightest neutralinos, as shown in Figure 5.1, with the top and anti-top quarks further decaying into a b jet and a W boson, is presented [4]. The signal models used are the so-called "T2tt" models from the Simplified Model Spectra (SMS) [120, 121], in which a branching ratio of 100% is assumed for the stop to top + $\tilde{\chi}_1^0$ decay.

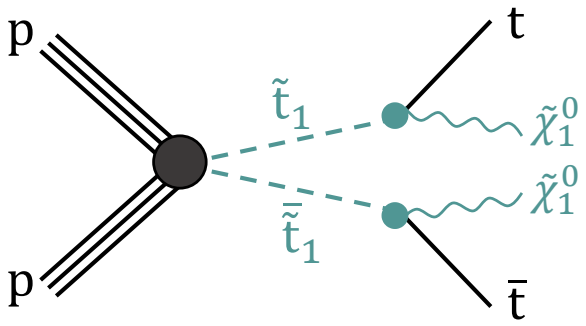

FIGURE 5.1: Illustration of the stop pair production with the stop quarks further decaying into a top quark and the lightest neutralino.

Several searches of stop quark pair production with the full Run 2 dataset at $\sqrt{s} = 13\,\text{TeV}$, corresponding to a total integrated luminosity of $137 \pm 2\,\text{fb}^{-1}$, have been performed by the ATLAS [122–125] and CMS [126–128] Collaborations in different final states with 0, 1 and 2 leptons , but none of them is sensitive to the region where the mass difference between the stop quark and the neutralino is close to the top quark mass. This analysis focuses on that special region, known as "top corridor". A summary of the stop exclusion limits at 95% confidence level (CL) obtained by the CMS Collaboration is shown in Figure 5.2 as a function of the stop quark and neutralino masses. Stop masses up to approximately 1300 GeV are excluded. The covered region in the lower left corner corresponds to the top corridor, which is not excluded by any of the searches.

The top corridor requires special care because the signal and $t\bar{t}$ background events have similar kinematics, especially at low neutralino masses, and signal events can only be detected as an excess on the $t\bar{t}$ cross section. The previous dedicated search [129] in this region was carried out using the 2016 dataset with an integrated luminosity of $35\,\text{fb}^{-1}$. This search excluded the presence of a SUSY signal up to $m_{\tilde{t}_1} = 208\,\text{GeV}$ for a massles neutralino, as shown in Figure 5.3.

In this section, a dedicated search of the stop quark pair production in the top corridor region using the full Run 2 dataset is presented. It includes signal models with higher neutralino masses than the previous one, up to about 110 GeV. Another improvement with respect to the previous search, which was based in a single variable, is the use of a MVA algorithm that takes advantage of the extra $p_{\text{T}}^{\text{miss}}$ in the event and the small kinematic differences with respect to $t\bar{t}$, introduced by the massive neutralinos in the event.

The new result is combined with three previous CMS searches in different final states [126–128], as described in Subsection 5.1.5, improving the existing exclusion limits at 95% CL. These searches are also interpreted in an alternative DM signal model whose results are presented in Subsection 5.1.5.2.

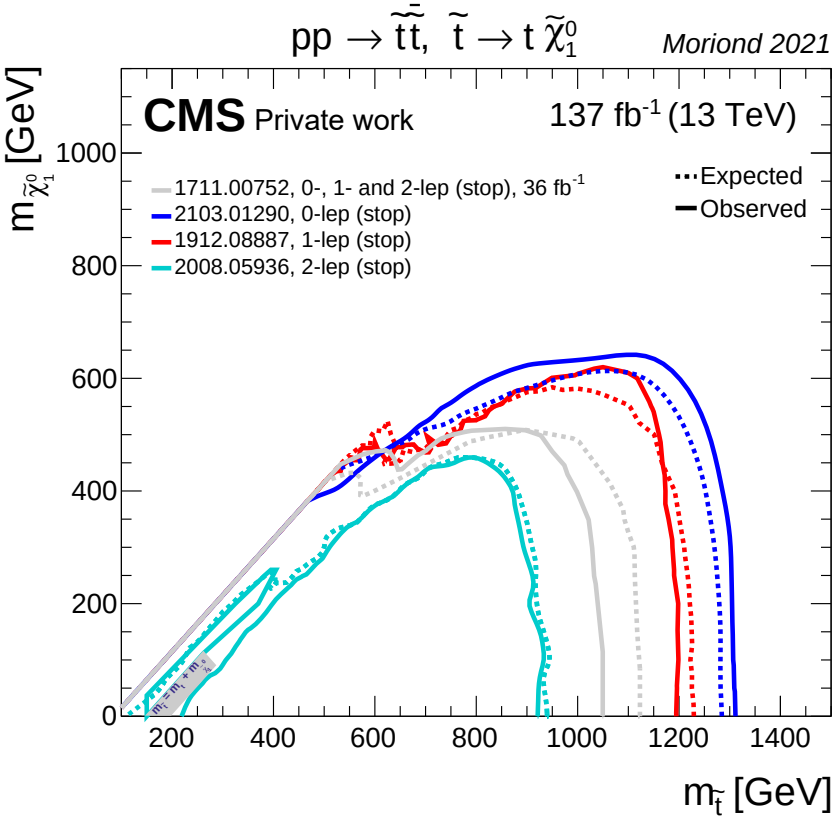

FIGURE 5.2: Mass limits at 95% CL obtained in the context of a simplified model of the stop pair production with the stop decays to a on- or off-shell top quark and the LSP. The solid (dashed) lines correspond to the observed (expected) limits of the different analyses.

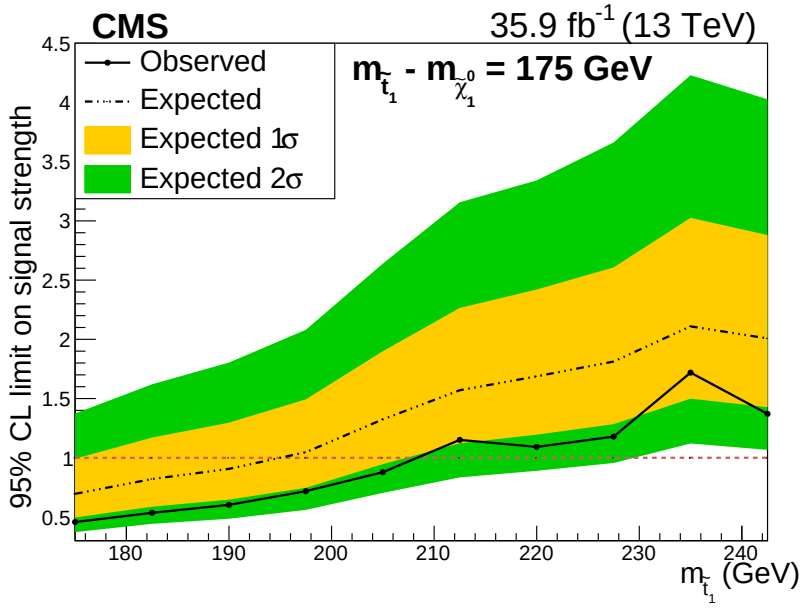

FIGURE 5.3: Expected and observed upper limits at 95% CL on the signal production cross section as a function of the stop quark mass asumming a massless neutralino. The green and yellow bands represent the regions containing 68 and 95%, respectively, of the distribution of limits expected under the background-only hypothesis [129].

All the details of the top corridor analysis are discussed below.

5.1.1 Object and event selection

This analysis focuses on the dilepton decays of the top quark pair, giving final states with two hight-p_T isolated leptons (electrons or muons), 2 b jets, 2 neutrinos and 2 neutralinos. They are characterized by the presence of an extra amount of p_T^{miss} arising from the neutralinos, which is the key to differentiate signal events from $t\bar{t}$ events.

The reconstruction of the different objects is based on the PF algorithm. Electrons and muons are selected if they have $p_T > 20\,GeV$ and $|\eta| < 2.4$. The identification and isolation requirements for electrons and muons are exactly the same as those of the $t\bar{t}$ cross section measurement at 13.6 TeV, described in Subsections 4.2.2.1 and 4.2.2.2, respectively, corresponding to the tight WP of the cut-based ID approaches.

Jets reconstructed with the anti-k_T algorithm within an opening angle of 0.4, $p_T > 30\,GeV$ and $|\eta| < 2.4$ are selected. The ID requirements are the same described in Subsection 4.2.2.3. Then, as the signal is enriched in jets originating from a b quark, the DeepJet tagger [44] is applied. The medium WP is used in order to optimize the efficiency and background rejection without introducing larger uncertainties. This WP has a b-tagging efficiency of about 70% and a misidentification rate of 1%.

The trigger strategy of this analysis is based on double lepton and single lepton triggers. The single muon triggers require the presence of an isolated muon reconstructed with 24 GeV, and the single electron triggers have p_T thresholds of 27, 35 and 32 GeV for 2016, 2017 and 2018 datasets, respectively. The dilepton HLT selection requires the presence of one isolated electron with $p_T > 23\,GeV$ and one isolated muon with $p_T > 8$ or 12 GeV, or one isolated muon with $p_T > 23\,GeV$ and one isolated electron with $p_T > 12\,GeV$, or one isolated muon with $p_T > 17\,GeV$ and other with $p_T > 8\,GeV$, or one isolated electron with $p_T > 23\,GeV$ and other with $p_T > 8\,GeV$.

The baseline selection is very similar to that of the $t\bar{t}$ analyses, as the final state is very similar. It consists in selecting events with two leptons with opposite sign of the electric charge and an invariant mass of $m_{\ell\ell} > 20\,GeV$ and the leading lepton with $p_T > 25\,GeV$. Besides, if the two leading leptons in the event are two electrons or two muons, the invariant mass of the leptons has to be at least 15 GeV away from the Z boson mass in order to further reduce DY contamination. Furthermore, events are required to have at least two selected jets and at least one b-tagged jet. Events are classified according to the flavour of the two leading leptons into $e^{\pm}\mu^{\mp}$, $\mu^+\mu^-$ and e^+e^- categories, and to the data-taking period (2016, 2017 and 2018).

5.1.2 Signal and background estimation

For the signal events simulation, the T2tt model from the SMS is used, assuming that top quarks are unpolarized and a branching fraction of 100% for the stop quark decaying into a top quark and a neutralino. The MadGraph generator at LO is used to generate signal samples with the same PDF set and MC tune as for the simulation of the $t\bar{t}$ background (NNPDF3.0 set and tune CUETP8M2T4 [77] for 2016 sample, and NNPDF3.1 set and tune CP5 for 2017 and 2018 samples). A T2tt scan of 79 different combinations of $m_{\tilde{t}_1}$ and $m_{\tilde{\chi}_1^0}$ was generated, requiring $|m_{\tilde{t}_1} - m_{\tilde{\chi}_1^0}| - m_t = 0, \pm 10, \pm 20, \pm 30$, with $m_{\tilde{\chi}_1^0}$ from 0 to 120 GeV in steps of 10 GeV. The distribution of the number of generated events in each mass point is shown in Figure 5.4.

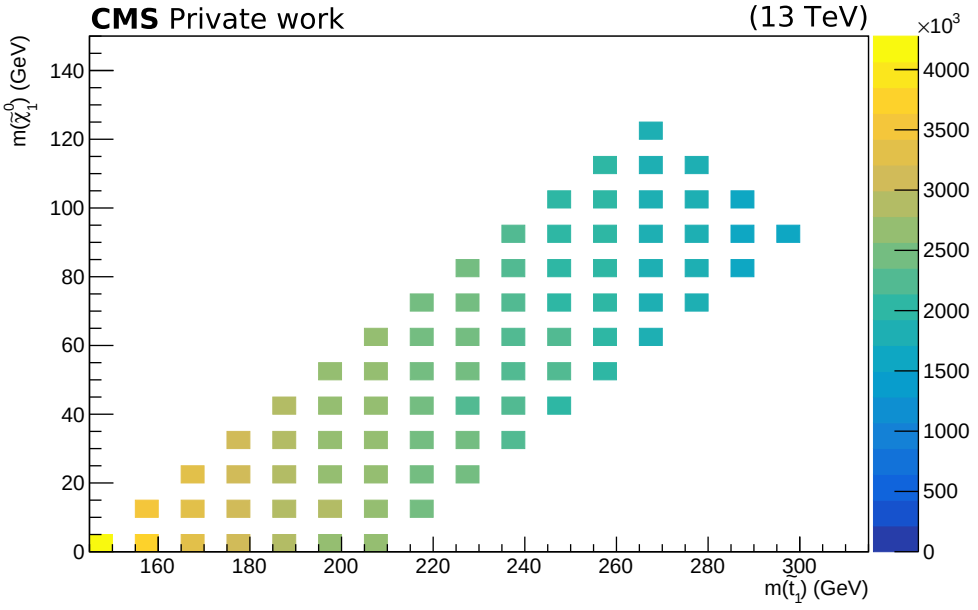

Figure 5.4: Distribution of signal points for each scanned stop quark and neutralino masses. The z-scale (colours) shows the total number of generated events.

In this analysis, where signal and $t\bar{t}$ are very similar, an accurate estimation of the $t\bar{t}$ process is crucial to have enough sensitivity. The $t\bar{t}$ process is simulated using the powheg (v2) genearator at NLO in QCD, and the same for tW events. The DY process is generated at LO order, and the production of W and Z bosons in association with $t\bar{t}$ events (referred to as $t\bar{t}W$ and $t\bar{t}Z$, or both together as $t\bar{t}V$), are generated at NLO using the MadGraph_aMC@NLO (v2.2.2) generator. The contributions from diboson processes are simulated at LO using pythia 8 (v226).

The NNPDF3.1 PDF set is used for all the MC samples in 2017 and 2018. Samples in 2016 use NNPDF3.0 PDF set [130]. Parton showering and hadronization are handled by pythia using the underlying event tune CP5 for all SM $t\bar{t}$, background and signal events in 2017 and 2018; and the CUETP8M2T4 tune is used for all events in 2016.

The response of the CMS detector is simulated for all the generated events with the GEANT4 package.

Simulated events are normalized according to the integrated luminosity and the theoretical cross section of each process. For the normalization of the simulated $t\bar{t}$ sample, the full NNLO+NLL accuracy calculation [131] is used, performed with the TOP++ 2.0 program [79], to obtain a $t\bar{t}$ production cross section of 832^{+20}_{-29} (scale) $\pm$ 35 (PDF+α_S) pb, assuming $m_t = 172.5\,\text{GeV}$.

All the SM processes are estimated using the MC samples described above, with the exception of the nonprompt leptons background. The events with nonprompt leptons are estimated from MC, including the contribution of events with jets misidentified as leptons or with leptons coming from the decay of a b quark mistakenly identified as coming from the hard process.

The same corrections used for the analyses previously described, explained in detail in Section 3.3, are applied to all simulated events, including lepton, trigger and b-tagging SFs, jet energy corrections, PU reweighting and L1 prefiring corrections.

5.1.3 Analysis strategy

After applying the baseline selection described in Subsection 5.1.1, it is found that about 98% of the background contribution comes from $t\bar{t}$, tW and DY events. The distributions of several variables for data and MC events passing the baseline selection are shown in Figure 5.5, where the distributions in data are well described by the MC predictions.

To improve the discrimination between signal and those backgrounds, a signal region is defined as the baseline selection plus two other selection requirements: $p_T^{miss} > 50\,\text{GeV}$ and $m_{T2}(\ell\ell) > 80\,\text{GeV}$. The $m_{T2}(\ell\ell)$ variable is defined as:

$$m_{T2}(\ell\ell) = \min_{\vec{p}_{T,1}^{miss}+\vec{p}_{T,2}^{miss}=\vec{p}_T^{miss}} \left(\max\left[m_T(\vec{p}_T^{\ell 1}, \vec{p}_{T,1}^{miss}), m_T(\vec{p}_T^{\ell 2}, \vec{p}_{T,2}^{miss})\right]\right), \tag{5.1}$$

where m_T is the transverse mass, and $\vec{p}_{T1}^{miss}$ and $\vec{p}_{T2}^{miss}$ are the estimated p_T of two neutrinos that are presumed to determine the total p_T^{miss} of the event. The transverse mass is calculated for each lepton-neutrino pair for different assumptions of the neutrino p_T. This computation of m_{T2} is carried out the algorithm described in Reference [132].

This variable is the key of the analysis to reject $t\bar{t}$ events because, if the p_T^{miss} and the leptons are well measured, they have a kinematic endpoint at the W boson mass, given that the p_T^{miss} is due to the presence of the neutrinos. For signal events, the missing energy associated to the neutralinos contribute to the p_T^{miss} and the $m_{T2}(\ell\ell)$ variable

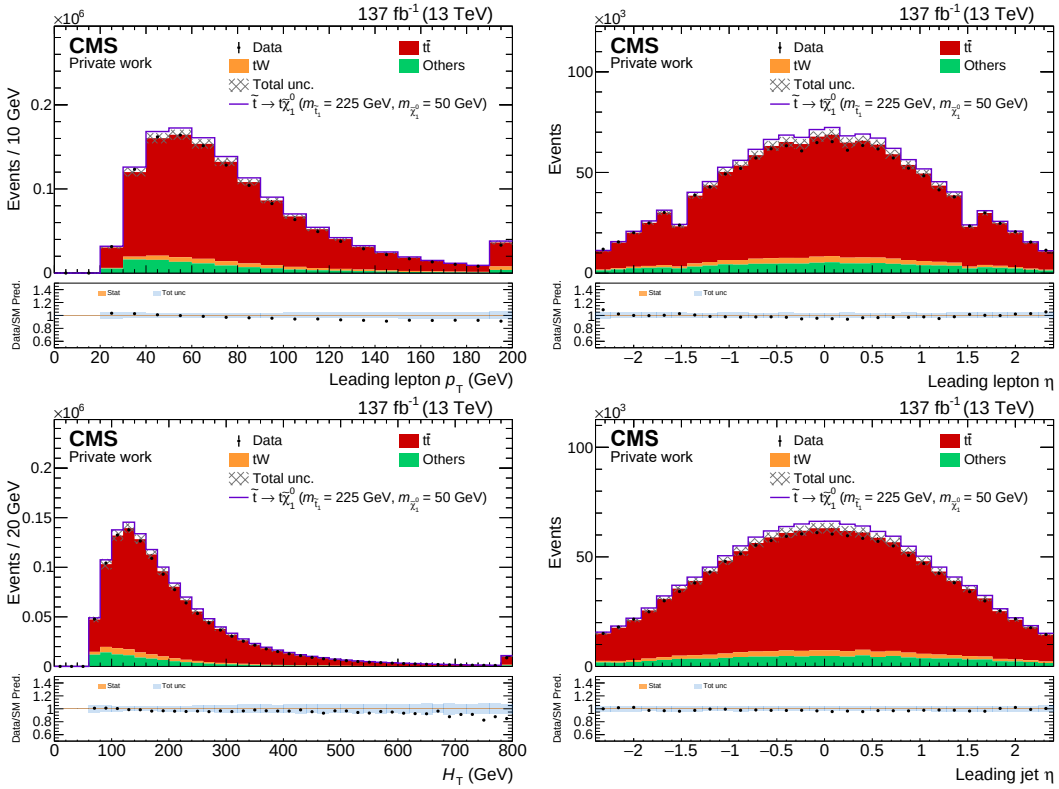

FIGURE 5.5: Distributions of data and MC predicted for events passing the baseline selection. A signal model of $m_{\tilde{t}_1} = 225\,\text{GeV}$ and $m_{\tilde{\chi}_1^0} = 50\,\text{GeV}$ is stacked on top of the background prediction. From top left to bottom right: leading lepton p_T, leading lepton η, H_T and leading jet η. Events from DY, nonprompt, $t\bar{t}W$, $t\bar{t}Z$ and diboson processes are grouped into the "Others" category.

can have higher values than the W boson mass, so no kinematic endpoint is found for these events. That is the reason why the signal region is defined by applying a requirement in $m_{T2}(\ell\ell) > 80\,\text{GeV}$. The $m_{T2}(\ell\ell)$ distribution for events passing the baseline selection can be seen on Figure 5.6 (left).

The minimum p_T^{miss} requirement is set to 50 GeV in order to avoid loosing a lot of signal events in models with low neutralino masses, however further discrimination between signal and $t\bar{t}$ events is expected at larger values of p_T^{miss} for models with large neutralino masses (of up to 120 GeV in the target models). The p_T^{miss} distribution for events passing just the baseline selection is shown on Figure 5.6 (right).

Figure 5.7 shows the data and MC prediction of some distributions for events passing the signal region requirements, practically populated by $t\bar{t}$ events ($\approx 90\%$). In this region the $t\bar{t}$ contribution is mostly due to events with mismeasured jets, which translates into mismeasured p_T^{miss}, or, in a smaller proportion, $t\bar{t}$ dilepton events where one of the leptons is missed and a nonprompt lepton, mainly from a b quark decay, is taken as the second prompt lepton. The addition of the p_T^{miss} and $m_{T2}(\ell\ell)$ requirements to the baseline selection improves the ratio of signal to background events, achieving a purer region.

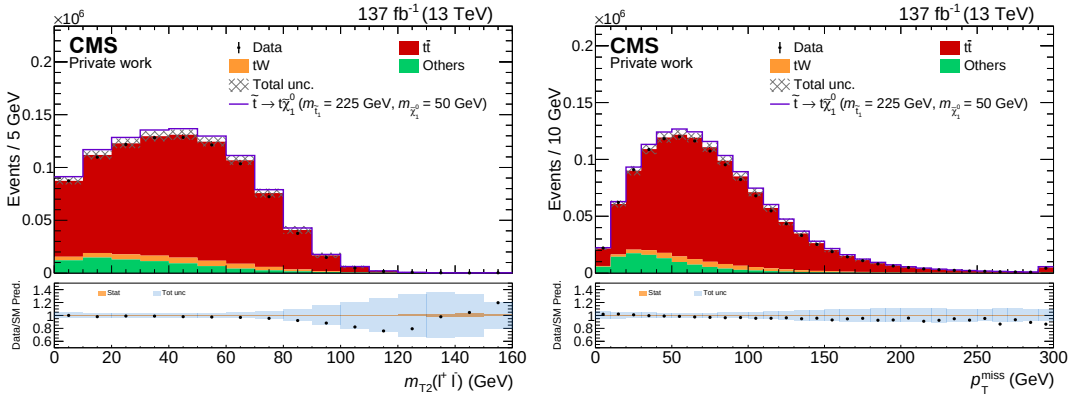

FIGURE 5.6: $m_{T2}(\ell\ell)$ (left) and p_T^{miss} (right) distributions of data and MC predicted for events passing the baseline selection. A signal model of $m_{\tilde{t}_1} = 225\,GeV$ and $m_{\tilde{\chi}_1^0} = 50\,GeV$ is stacked on top of the background prediction. Events from DY, nonprompt, t$\bar{t}W$, t$\bar{t}Z$ and diboson processes are grouped into the "Others" category.

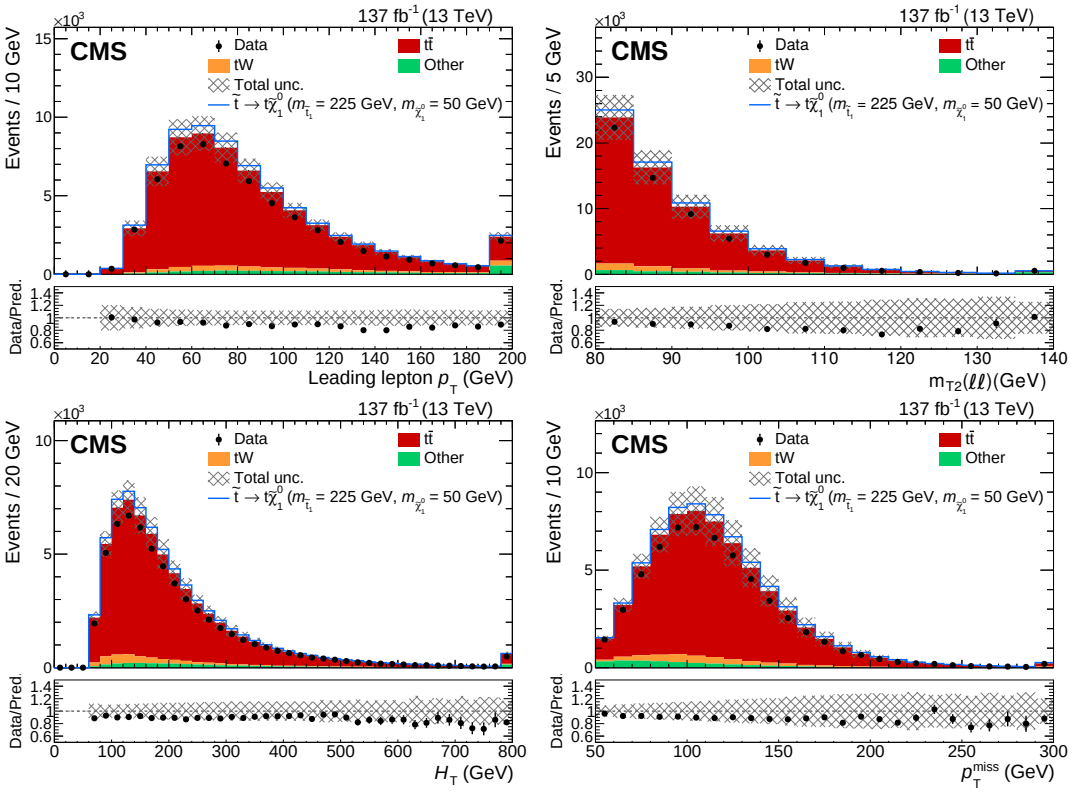

FIGURE 5.7: Distribution of data and MC predicted for events passing the signal region. A signal model of $m_{\tilde{t}_1} = 225\,GeV$ and $m_{\tilde{\chi}_1^0} = 50\,GeV$ is stacked on top of the background prediction. From top left to bottom right: leading lepton p_T, $m_{T2}(\ell\ell)$, H_T and p_T^{miss}. Events from DY, nonprompt, t$\bar{t}W$, t$\bar{t}Z$ and diboson processes are grouped into the "Others" category [4].

As mentioned above, the biggest challenge of this search is to separate signal events from the $t\bar{t}$ background, as just making a selection of events or using a single variable is not enough. In order to maximize the sensitivity to the signal process, the strategy of this analysis is based on a DNN [133] to exploit the differences between signal and $t\bar{t}$ events of several observables and achieve a better final discriminator.

5.1.3.1 DNN training

DNNs, inspired by the biological neural networks found in the brain, consist of interconnected units called neurons that receive, process, and transmit information to other neurons via a network of connections. The fundamental structure of a DNN, as depicted in Figure 5.8, comprises layers of neurons interlinked. Each connection is associated with a weight, reflecting the significance of information transfer between neurons. Information flows from left to right within the network. Initially, network weights are typically initialized, often randomly. The first layer of neurons corresponds to the input variables. When all neurons in one layer are linked to those in the next layer, each neuron in the subsequent layer receives as input the sum of information from each neuron in the preceding layer, weighted by their associated values. This input then undergoes a non-linear activation function, and the processed information goes to subsequent layers. If the DNN is used for a classification task, as is this case, the final layer uses an activation functions that allows to interpret the output numbers as class probabilities. The accuracy of predictions is assessed by comparing the predicted class with the real one. The weights are updated layer by layer in the opposite direction using back-propagation and the gradient descent of the loss function to minimize the error and achieve the optimal values. More details of how a DNN works can be found in Reference [134]. As mentioned in the introduction, DNNs are used to take advantage of all kinds of data, including particle physics data.

In this analysis, a DNN is trained using $t\bar{t}$ events as background and events from the 79 different neutralino-stop masses assumptions as signal. Furthermore, the shape of the distribution in some variables depends on the signal point (see Figures 5.11 and 5.12) and, to improve the sensitivity, a different model should be trained for each signal point, or, what is better, train just one model with a different output for each signal point. That is achieved by introducing the stop quark and neutralino masses in the training and is what is known as parametric DNN: training a unique parametric DNN gives a specific model for each signal point. Figure 5.8 represents the structure of a parametric DNN.

For $t\bar{t}$ background events, a couple of random mass values are assigned to each event from the signal masses in the way that these two variables do not have a high importance in the classification and avoid introducing correlations.

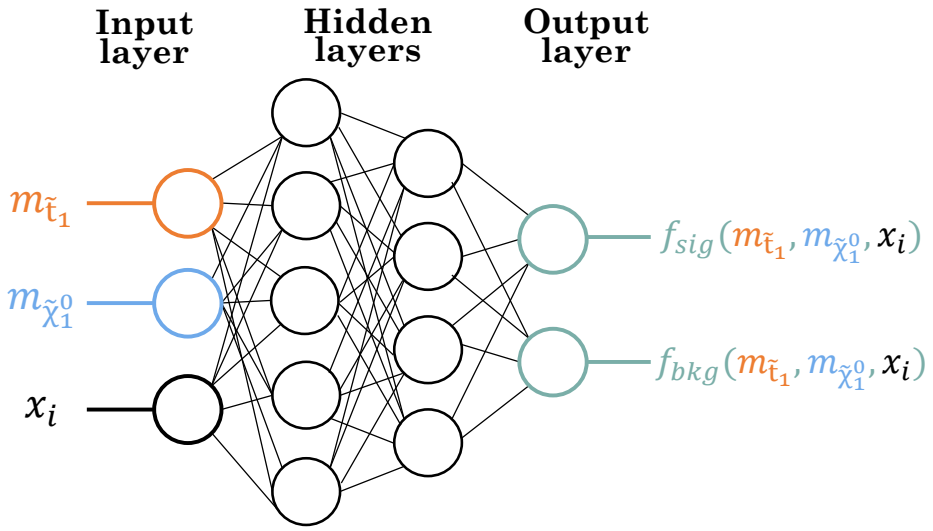

FIGURE 5.8: Structure of a parametric deep neutral network. x_i are the input variables and $f_{sig/bkg}$ the probability of an event to be classified as signal/background.

The training is done using events passing the baseline selection in order to use the separation power of different observables over a large phase space, and only events in the $e^{\pm}\mu^{\mp}$ category are used, since this is the channel with a larger amount of events. Several variables were tested and, finally, a total of 13 training variables were selected in which the distribution in data is well described by the MC prediction, shown in Figures 5.9 and 5.10: $m_{\tilde{t}_1}$, $m_{\tilde{\chi}_1^0}$, $p_T^{e\mu}$, $\Delta\phi$, $\Delta\eta$, $p_T^{\ell_1}$, η^{ℓ_1}, $p_T^{\ell_2}$, η^{ℓ_2}, p_T^{miss}, $m_{e\mu}$, $m_{T2}(e\mu)$ and H_T.

Figures 5.11 and 5.12 show the normalized distributions of these variables for events passing the baseline selection in the $e^{\pm}\mu^{\mp}$ category for three signal points and for $t\bar{t}$. It can be seen that p_T^{miss} and $m_{T2}(e\mu)$ are the two variables that better discriminate signal from background events. These figures also show that, in some variables, the shape of the distributions is different among the signal points. That is way the parametric DNN is useful as it profits advantage of those differences.

For the training, 40% of the $t\bar{t}$ and signal available events were used, leaving the rest for the analysis to ensure that no bias is introduced. This corresponds to a total of 7 million events, 3.5 million from $t\bar{t}$ background and 3.5 million from signal. These events are further split, as usual in ML, in 60% for training the model, 15% for validation and 25% for test. The training was performed with TensorFlow [135] using the Keras interface [136]. DNNs have several parameters, known as hyperparameters, that can be modified to achieved the best performance. The optimization of the hyperparameters was done manually, but a grid-search strategy was also tested. Finally, the DNN structure selected consists of 7 hidden layers with ReLu activation function [136] (13, 300, 200, 100, 100, 100, 100, 10 neurons) and an output layer with 2 neurons with softmax normalization function [136], which allows to interpret the output numbers as probabilities. The optimizer selected was Adam [137] with a learning rate of 0.0001.

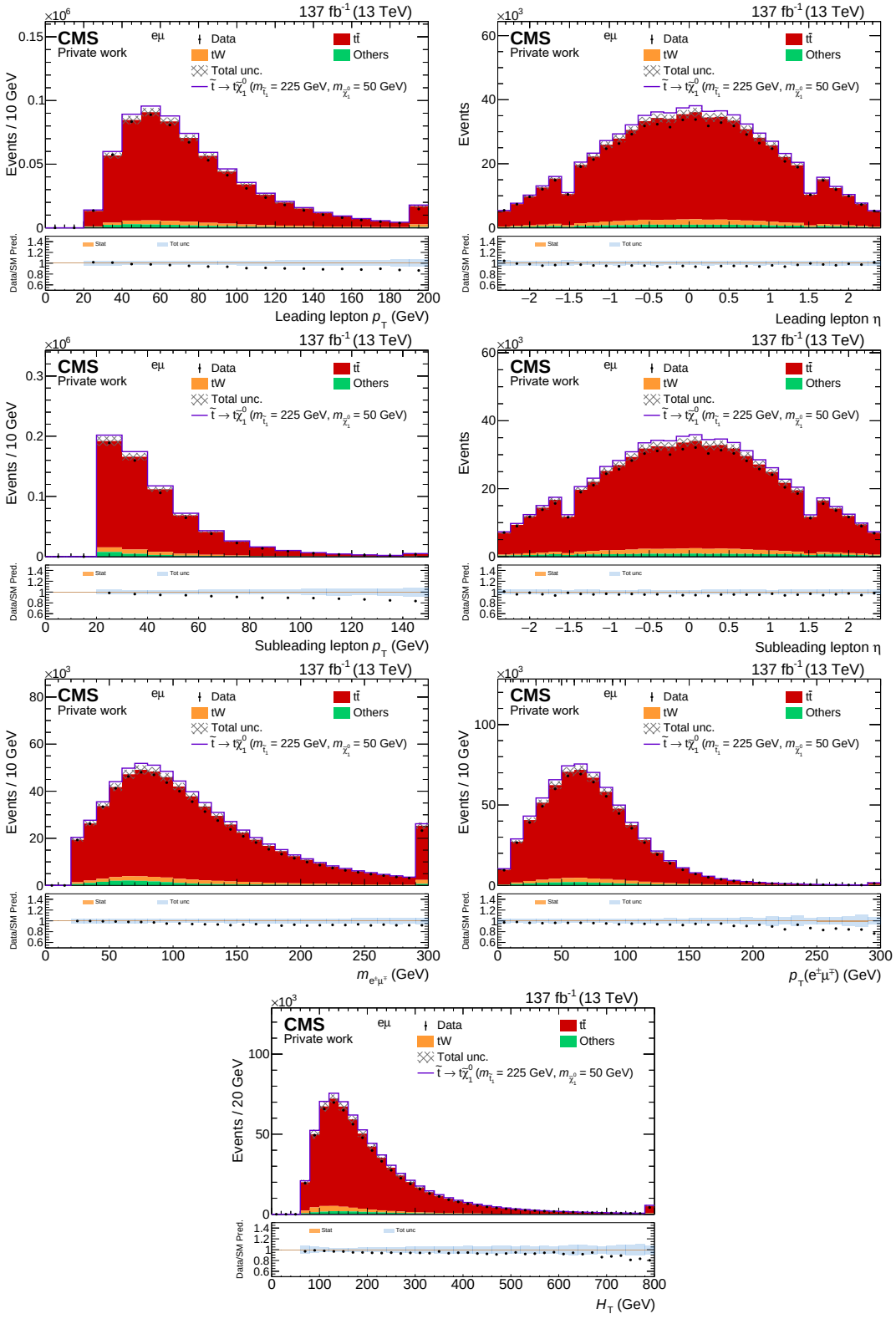

FIGURE 5.9: Distributions of data and MC predicted for events passing the baseline selection in the $e^{\pm}\mu^{\mp}$ category with a signal model of $m_{\tilde{t}_1} = 225\,\text{GeV}$ and $m_{\tilde{\chi}_1^0} = 50\,\text{GeV}$ stacked on top of the background prediction. From top left to bottom right: leading lepton p_T, leading lepton η, subleading lepton p_T, subleading lepton η, invariant mass of the leptons, p_T of the dilepton system and H_T. Events from DY, nonprompt, $t\bar{t}W$, $t\bar{t}Z$ and diboson processes are grouped into the "Others" category.

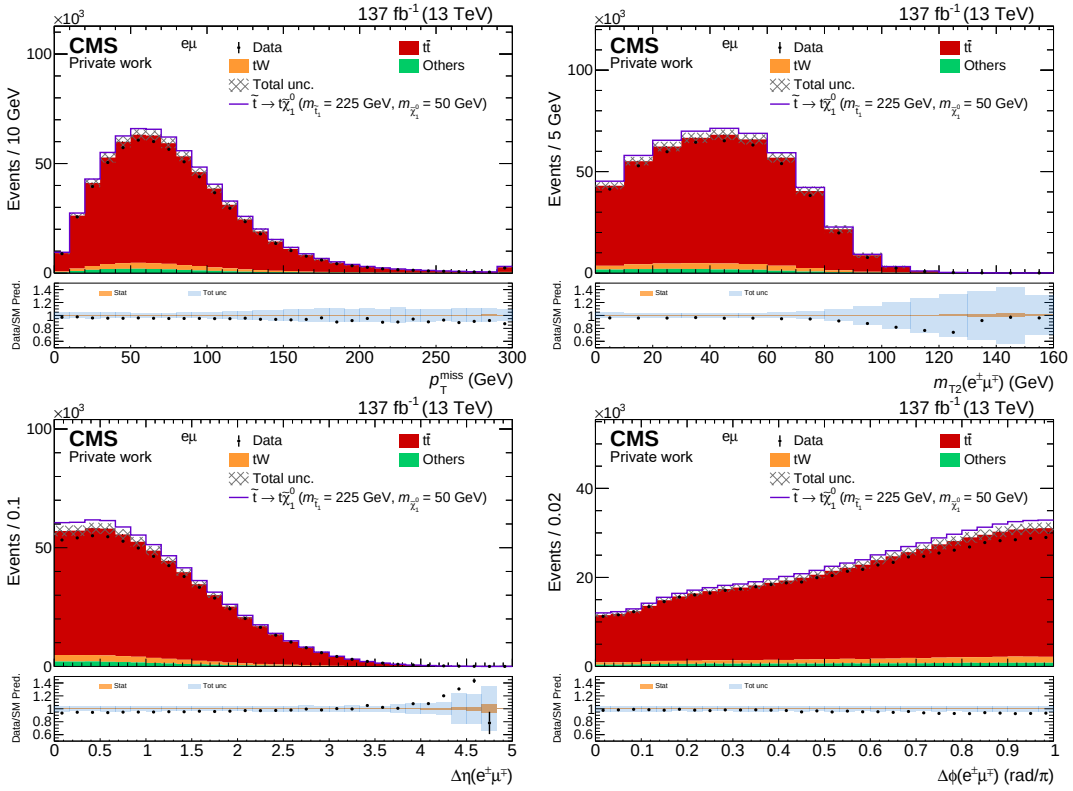

FIGURE 5.10: Distributions of data and MC predicted for events passing the baseline selection in the $e^{\pm}\mu^{\mp}$ category with a signal model of $m_{\tilde{t}_1} = 225\,\text{GeV}$ and $m_{\tilde{\chi}_1^0} = 50\,\text{GeV}$ stacked on top of the background prediction. From top left to bottom right: p_T^{miss}, $m_{T2}(e\mu)$, $\Delta\eta$ and $\Delta\phi$ between the leptons. Events from DY, nonprompt, $t\bar{t}W$, $t\bar{t}Z$ and diboson processes are grouped into the "Others" category.

Figure 5.13 shows the ROC curves for train and test datasets together with their AUCs, both around 0.65, and Figure 5.14 shows the probability of an event to be classified as signal, which is the DNN score. These plots prove that there is no overfitting in the model and that signal and $t\bar{t}$ events populate different regions.

The DNN is applied to events passing the signal region requirements in order to maximize the discrimination power between signal and background. Figure 5.15 shows the DNN score for two distinct mass points in the signal region for signal and $t\bar{t}$ background. The DNN score shape varies for both the background and the signal since stop quark and neutralino masses are introduced in the training. This figure also demonstrates how well the DNN score discriminates between the signal and the background, particularly at higher values of the distribution.

The agreement of the DNN score between data and MC simulated events was tested for the $e^{\pm}\mu^{\mp}$ category in a control region defined as the signal region but inverting the p_T^{miss} and $m_{T2}(e\mu)$ requirements, where the signal contribution is expected to be small. In this region, the presence of $t\bar{t}$ and tW events is predominant, and the data is found to be well described by MC predictions in all the hypothetical mass points. Figure 5.16 shows the DNN score distribution for two of the signal points.

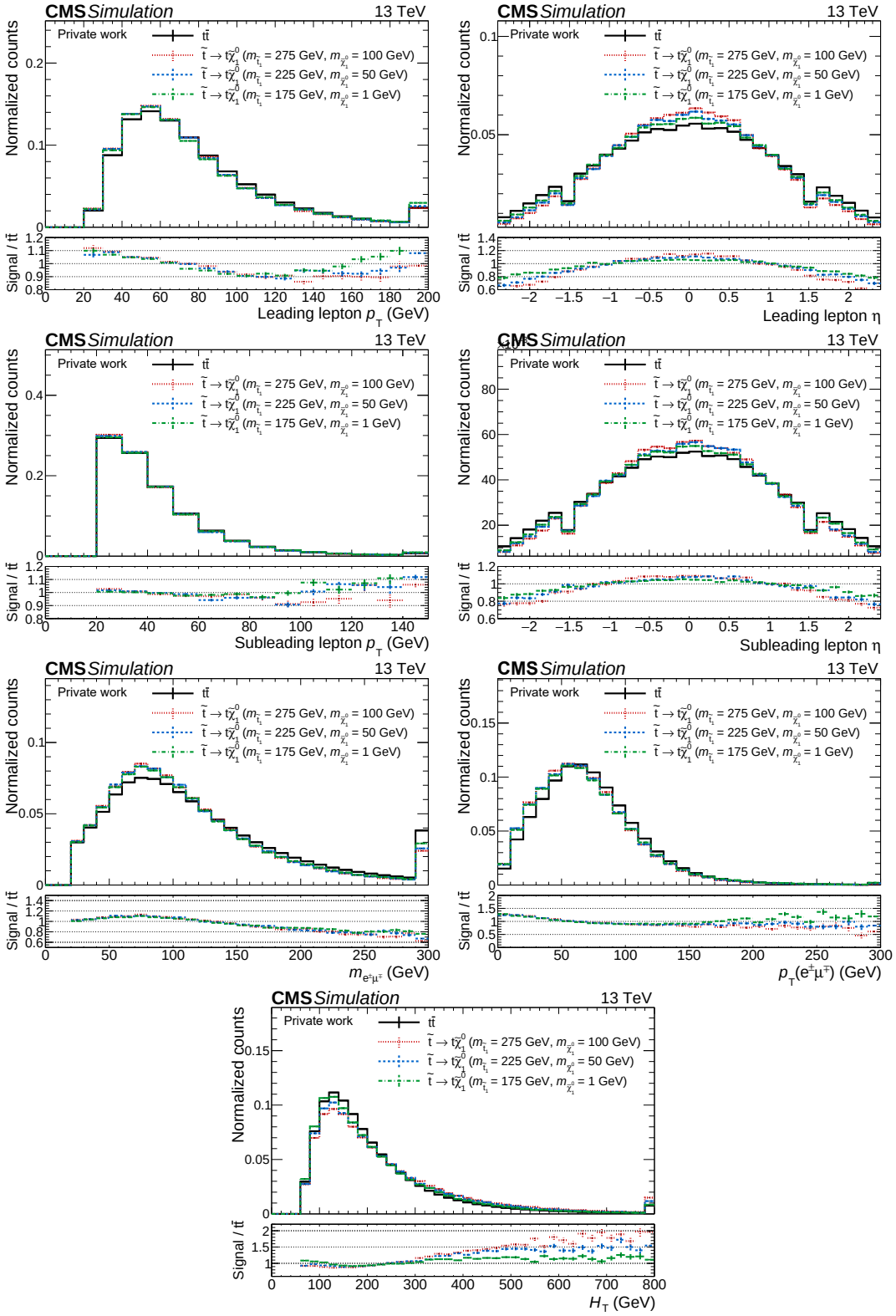

FIGURE 5.11: Normalized distribution of the DNN input variables after the baseline selection in the $e^{\pm}\mu^{\mp}$ category. Three different signal models and $t\bar{t}$ are compared. From top left to bottom right: leading lepton p_T, leading lepton η, subleading lepton p_T, subleading lepton η, invariant mass of the leptons, p_T of the dilepton system and H_T.

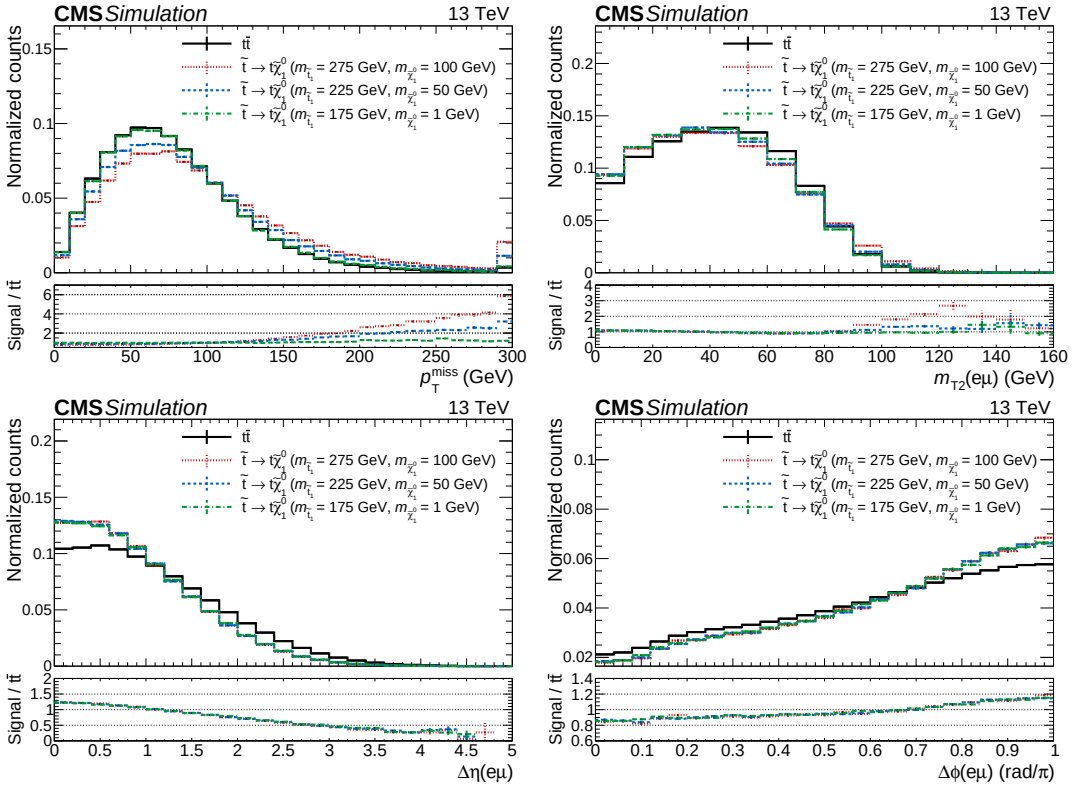

FIGURE 5.12: Normalized distribution of the DNN input variables after the baseline selection in the $e^{\pm}\mu^{\mp}$ category. Three different signal models and $t\bar{t}$ are compared. From top left to bottom right: p_T^{miss}, $m_{T2}(e\mu)$, $\Delta\eta$ and $\Delta\phi$ between the leptons [4].

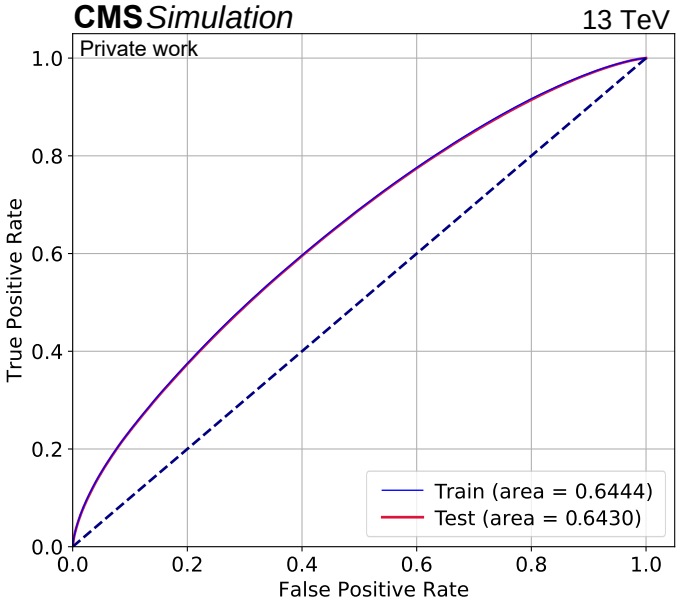

FIGURE 5.13: ROC curve of the DNN for the train dataset in blue and for the test dataset in red.

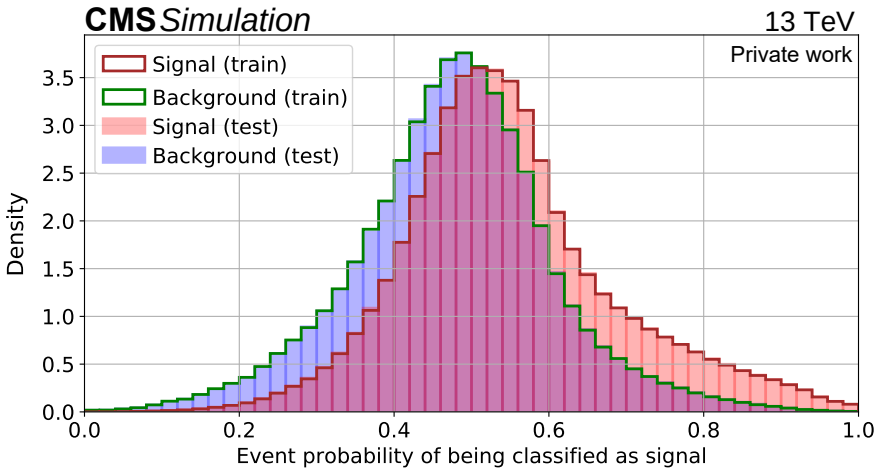

FIGURE 5.14: Normalized DNN score for signal events in red and for $t\bar{t}$ background events in blue. The empty histograms show the distribution with the train dataset and the fill histograms with the test dataset.

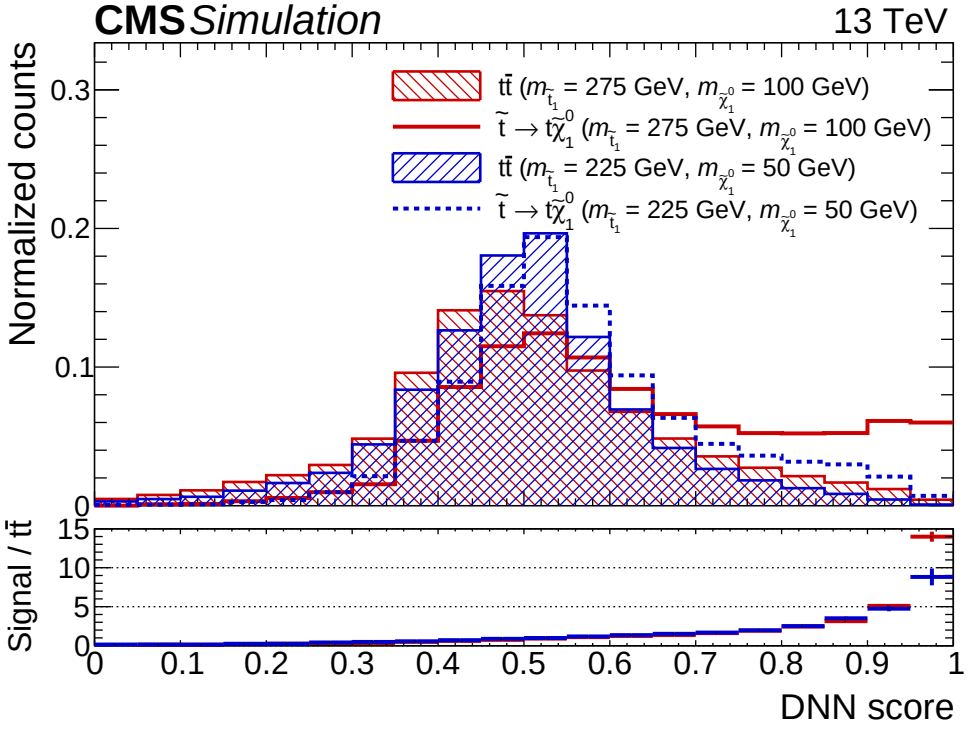

FIGURE 5.15: Normalized DNN score distribution in the signal region for signal and $t\bar{t}$ background in two mass points [4].

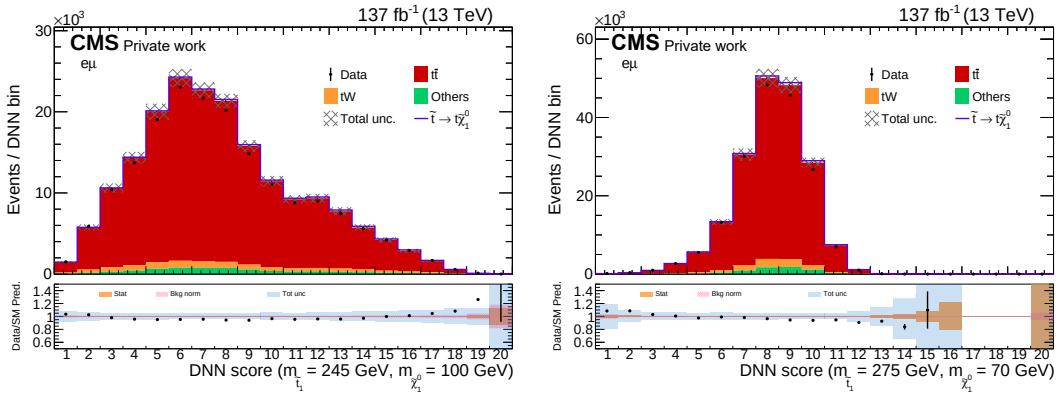

FIGURE 5.16: DNN score distributions of data and MC predicted for events passing the control region requirements in the $e^{\pm}\mu^{\mp}$ category for different stop and neutralino masses assumptions. Events from DY, nonprompt, $t\bar{t}W$, $t\bar{t}Z$ and diboson processes are grouped into the "Others" category.

5.1.4 Systematic uncertainties

The background and signal estimations are affected by several systematic uncertainties in the acceptance, efficiency, and normalization. The effect of uncertainties in the trigger efficiencies, lepton reconstruction, identification and isolation efficiencies, jet energy scale and resolution, muon energy scale, pileup reweighting, and b-tagging efficiency are considered in the estimation of background and signal yields by varying the nominal correction values by their uncertainties. The uncertainty in $p_{\mathrm{T}}^{\mathrm{miss}}$ from the contribution of unclustered energy is evaluated based on the momentum resolution of the different PF candidates, according to their classification. Details on the procedure can be found in References [37, 138, 139].

Because of the large impact of the $t\bar{t}$ background prediction in this search, various modelling systematic uncertainties are assigned, reflecting the limited knowledge of the main theoretical parameters used in the simulation. The ranges of variation of these parameters were set in several previous CMS analyses [77] and the modelling of the $t\bar{t}$ background has been shown to accurately describe several kinematic variables within the systematic uncertainties [140]. The modelling uncertainties taken into account are the initial and final state radiation, PDF+α_S and μ_F/μ_R scales by varying different parameters in the MC generator within their uncertainties, as described in Subsection 4.2.5.2. Apart from those, there are other modelling uncertainties that are calculated with dedicated samples such as the ME/PS matching, underlying event, and the top quark mass. An uncertainty due to the known mismodelling of the top quark p_{T} is also considered. More information can be found in Reference [77].

The MADGRAPH modelling of the initial state radiation in signal events is improved by scaling the p_{T} distribution of the initial state radiation jets in MC, according to a correction derived using $t\bar{t}$ events. An uncertainty is applied by considering variations of half the difference between the corrections and unity. The effect of this uncertainty

on the signal yields amounts to about 1%. The effect on the acceptance of the uncertainties in the μ_F/μ_R scales is also taken into account in signal events, resulting in an uncertainty of the order of 0.5-1.0%.

A summary of the experimental and modelling sources of uncertainty applied in the analysis with their contribution to the DNN score distribution for $t\bar{t}$ background and signal can be found in Table 5.1. The values correspond to the average of the variations in the DNN score bins in each of the 79 signal points.

Source	Uncertainties (%)	
	$t\bar{t}$	signal
Electron efficiency	1.5	
Muon efficiency	0.5	
Trigger efficiency	1.2	
Muon energy scale	1.4	
b tagging efficiency	3.0	
Jet energy scale	7.5	5.7
Jet energy resolution	16.0	7.0
Unclustered energy	4.2	5.0
Pileup reweighting	3.2	1.5
Size of the MC sample	3.0	25.0
Initial-state radiation	0.6	1.0
Final-state radiation	3.4	–
ME/PS matching	2.0	–
Underlying event	1.5	–
PDFs and α_S	1.0	–
μ_F, μ_R scales	3.8	0.7
Top p_T	1.3	–
Top quark mass	1.5	–

TABLE 5.1: Summary of the contributions of the experimental and modeeling uncertainties in the DNN score distribution for signal and the $t\bar{t}$ background in the signal region.

Furthermore, a normalization uncertainty is applied to each background and signal process. The $t\bar{t}$ normalization uncertainty is taken from the NNLO+NLL estimation of the cross section and additionally varying the top quark mass by ± 1 GeV, leading to an uncertainty of 6%. For tW a normalization uncertainty of 15% is applied and, for other backgrounds, including dibosons, nonprompt leptons, DY and $t\bar{t}V$, a normalization uncertainty of 30% is assigned. A 6.5% uncertainty in the signal normalization is assigned, according to the uncertainties in the predicted cross section of signal models in the stop quark mass range of the analysis [141].

Finally, the uncertainty in the integrated luminosity, which affects both signal and backgrounds normalization, is 1.6% [142–144].

The statistical uncertainties affect background and signal predictions due to the limited size of the MC samples in each bin of the distributions. They are introduced through the Barlow–Beeston approach [145].

5.1.5 Results

As no significant excess has been observed, the results are obtained in the form of exclusion limits by testing the SUSY hypothesis against the SM-only hypothesis. For each signal point, the DNN output distribution in the signal region is computed separately for each data-taking period (2016, 2017 and 2018) and each category ($e^{\pm}\mu^{\mp}$, $\mu^{+}\mu^{-}$ and $e^{+}e^{-}$) in order to maximize the sensitivity to the signal, and then a binned profile likelihood fit is performed simultaneously, with the systematic uncertainties acting as NPs modelled using Gaussian distributions. The systematic uncertainties are treated as correlated among bins, processes, data-taking period and category. On the opposite, statistical uncertainty is treated as uncorrelated among bins.

The post-fit DNN score distribution for 4 different signal points are shown in Figure 5.17. The fit function includes the background and the signal prediction scaled by the post-fit signal strength. The two plots in the top row show the distribution of two signal points of the central diagonal of the top corridor, where signal and $t\bar{t}$ are more similar since they assume a mass difference between neutralino and stop quark of 175 GeV, very close to the top quark mass, and they are more difficult to separate. The other plots, in the bottom row, are for signal points away from that diagonal and it can be seen that the signal has a larger contribution at the tail of the distribution and $t\bar{t}$ is flatter.

No excess is observed and the upper limits on the production cross section of stop quark pairs are extracted using the asymptotic approximation [146–149] at 95% CL. The expected and observed upper limits for the three central diagonals of the top corridor with $\Delta m(\tilde{t}_1, \tilde{\chi}_1^0)$ = 165, 175 and 185 GeV as a function of the stop mass are shown in Figure 5.18. The expected ± 1 and $\pm 2\sigma$ bands are also shown.

The observed upper limits on the signal cross section in the full search region are shown in Figure 5.19. Both the observed and expected cross section limits exclude the model over the full top corridor region for first time. This far improves the results of the previous analysis performed with 36 fb^{-1} (only 2016 dataset) using the m_{T2} distribution, which managed to exclude stop quark masses up to 210 GeV in the central diagonal.

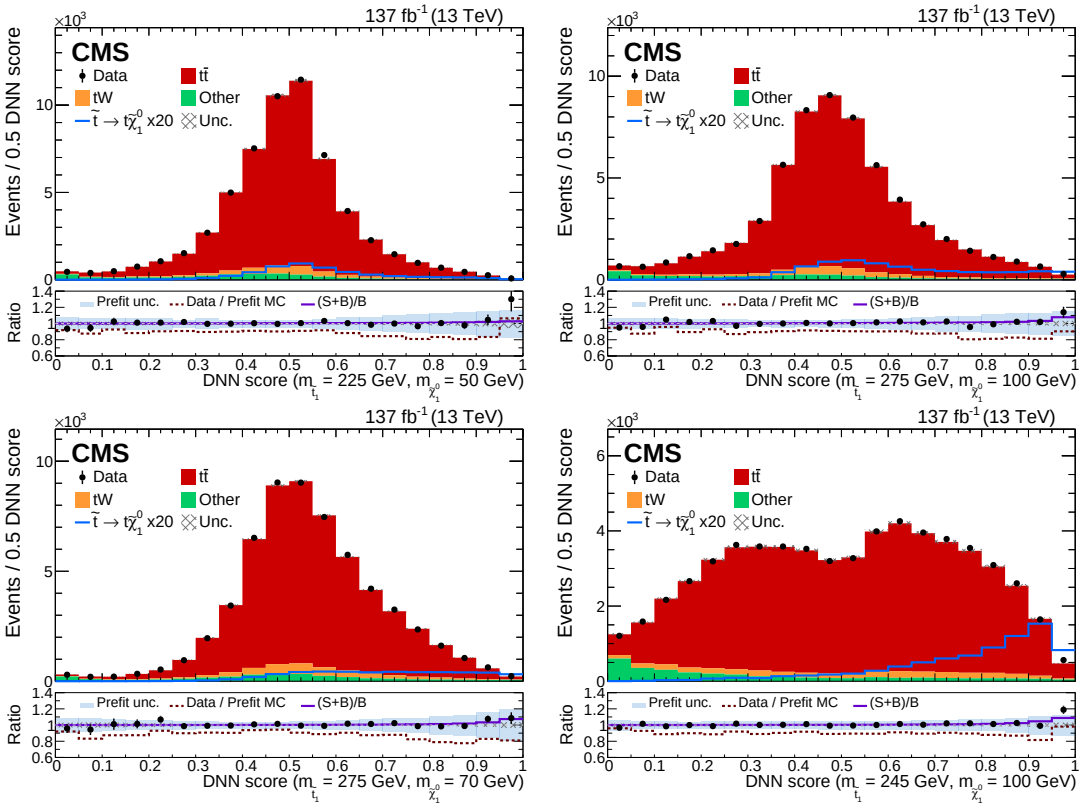

FIGURE 5.17: Post-fit DNN score distributions for data and MC predicted events passing the signal region for different signal points. The superimposed signal prediction is scaled by the post-fit signal strength and, in the upper panels, it is also multiplied by a factor 20 for better visibility. The post-fit uncertainty band (crosses) includes statistical, background normalization, and all systematic uncertainties. Events from $t\bar{t}V$, DY, nonprompt leptons, and diboson processes are grouped into the 'Other' category. The lower panel contains the data-to-prediction ratio before the fit (dotted brown line) and after (dots), each of them with their corresponding band of uncertainties (blue band for the pre-fit and crosses band for the post-fit). The ratio between the sum of the signal and background predictions and the background prediction (purple line) is also shown [4].

5.1.5.1 Combination

A statistical combination of the results of the three searches in final states with 0, 1 and two leptons described in detail in References [126–128] is performed outside the top corridor area in order to provide interpretations in the context of different signal scenarios. The result of the top corridor analysis is also added to the final result, but without combining because the other analyses are not sensitive to that region.

Figure 5.20 shows the combination upper exclusion limits for the T2tt model at 95% CL, where stop quark masses up to 1325 GeV for a massless neutralino, and a neutralino mass of 700 GeV for a stop quark mass of 1150 GeV, are excluded. The combination improve expected limits by over 50 GeV in comparison with the fully hadronic analysis, which is the one with the best result. This plot also illustrates that the top corridor is completely excluded.

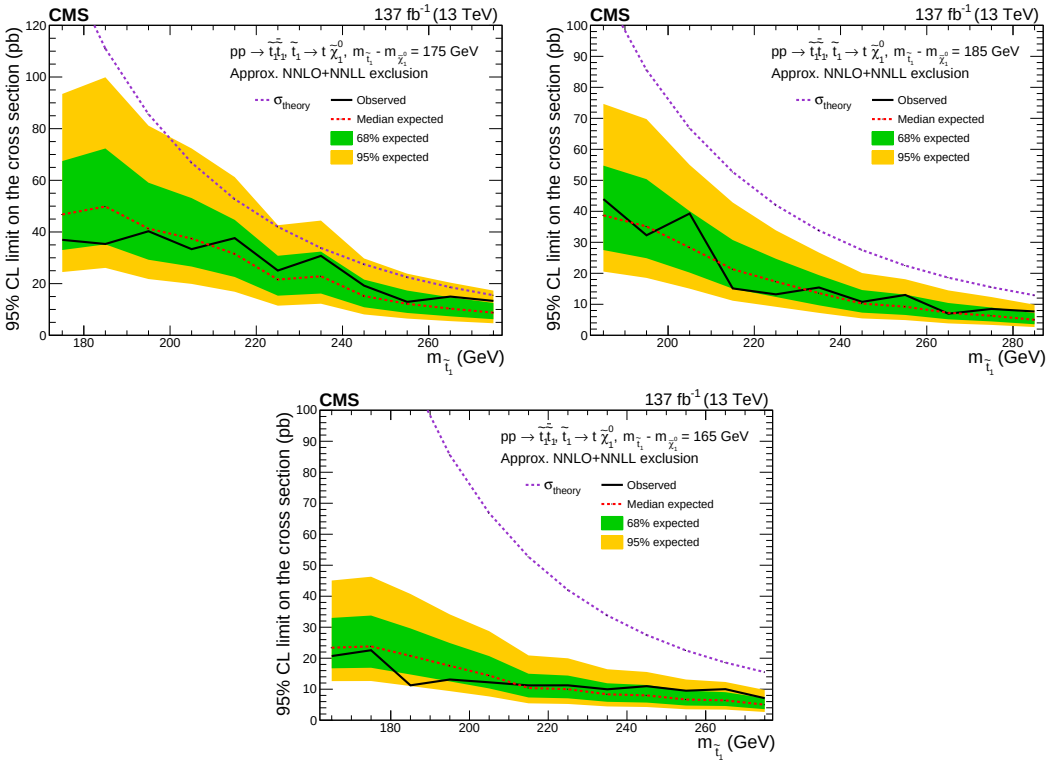

FIGURE 5.18: Expected and observed upper limits at 95% CL on the signal production cross section as a function of the stop quark mass for the three central diagonals of the top corridor region. The green and yellow bands represent the regions containing 68 and 95%, respectively, of the distribution of limits expected under the background-only hypothesis. The purple dotted line indicates the approximate NNLO+NNLL production cross section [4].

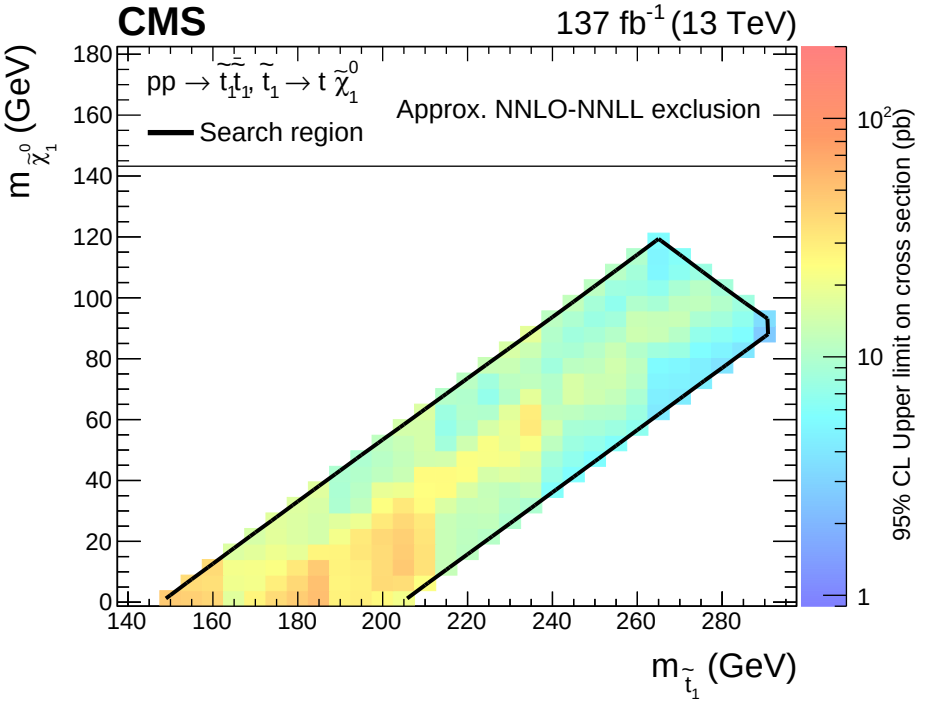

FIGURE 5.19: Observed upper limits at 95% CL on the signal cross section as a function of the stop quark and the neutralino masses [4].

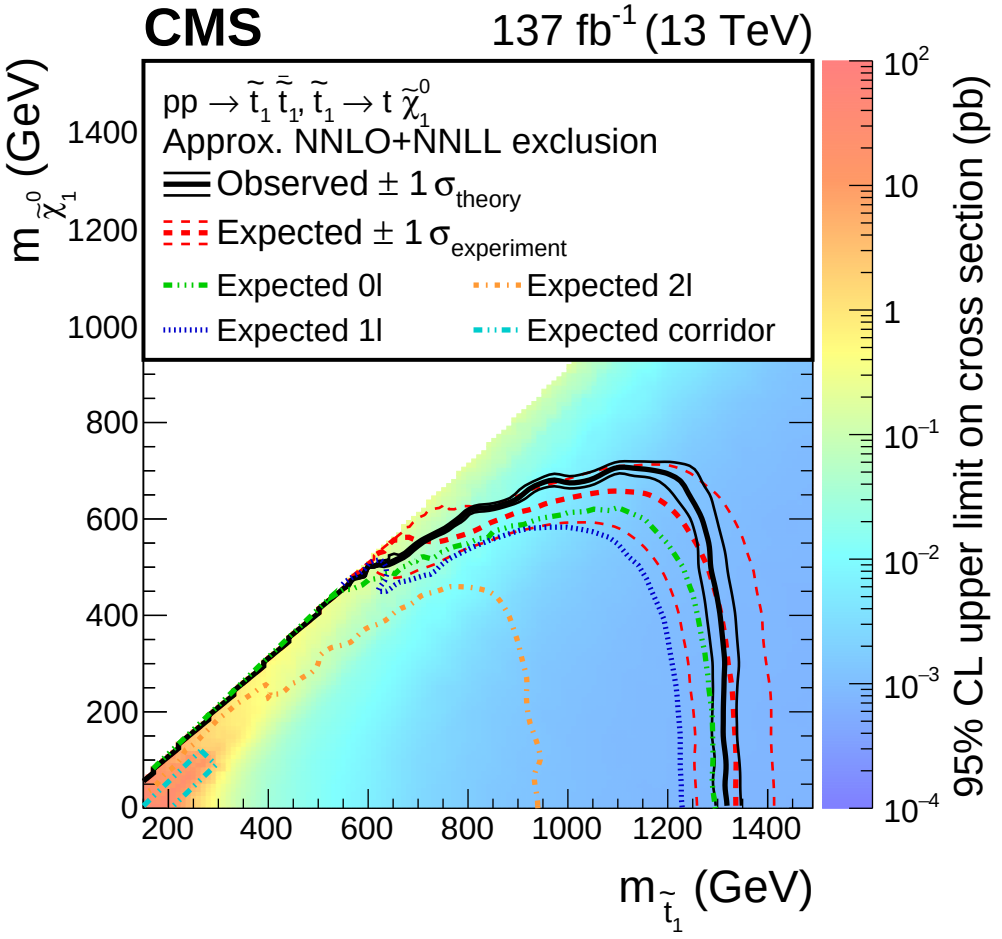

FIGURE 5.20: Expected and observed upper limits on the T2tt signal cross section as a function of the stop quark and the neutralino masses. The color indicates the 95% CL upper limit on the cross section at each point in the plane. The area below the thick black curve represents the observed exclusion region at 95% CL, while the dashed red lines indicate the expected limits at 95% CL and the region containing 68% of the distribution of limits expected under the background-only hypothesis of the combined analyses. The thin black lines show the effect of the theoretical uncertainties in the signal cross section [4].

5.1.5.2 Dark matter interpretation

The results of the stop quark searches included in the combination can further be interpreted in simplified models of associated production of DM particles with a top quark pair. Numerous astronomical observations suggest the existence of a substance that permeates the universe and appears to interact significantly with ordinary matter through gravity. Despite attempts to directly observe it, this mysterious DM has not been proven. Nevertheless, estimates indicate that DM makes up over a quarter of the total mass-energy in the universe. The possibility that DM interacts with ordinary matter means that gravitation remains open. This suggests that DM might be directly produced by the LHC. If produced, DM particles are practically invisible to the detector, revealing their presence through missing energy in the event.

The simplest and most relevant model for DM production in association with $t\bar{t}$ proposes a spin-0 interaction between DM and SM particles. In this simplified model, a scalar (ϕ) or pseudoscalar (a) particle mediates the interaction between SM quarks and a new Dirac fermion (χ), which is the DM candidate particle [150–154]. Under the assumption of minimal flavour violation [155, 156], the heavy-flavour quark production is favored and the spin-0 mediators couple to quarks having mass m_q with couplings, similar to those of the SM, proportional to $g_q m_q$, where the coupling strength g_q is taken to be 1. The coupling strength of the mediator to the DM particles g_{DM} is also set to 1. The Feynman diagram of the $t\bar{t}$ +DM model at LO is presented in Figure 5.21.

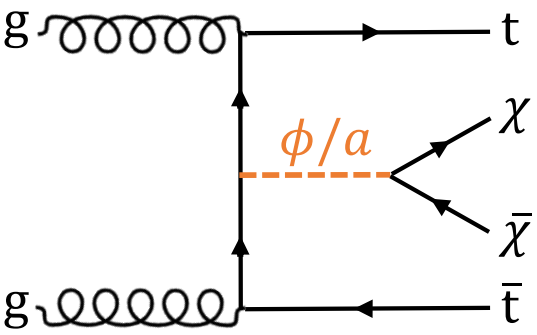

FIGURE 5.21: Feynman diagram at LO of the direct DM production through ϕ or a mediator particles in association with a top quark pair.

Previous searches by the ATLAS and CMS Collaborations excluded scalar and pseudoscalar mediator particles with a mass of up to 290 and 300 GeV, respectively [157–161].

Signal samples of the signal model [162] are generated using MADGRAPH (v2.4.2) at LO interfaced with PYTHIA 8 (v230) using the CP5 tune. The NNPDF3.1 set of PDFs at NNLO is used. Samples for mediator masses of 50, 100, 200, 300, and 500 GeV are considered for both ϕ and a models. The mass of the DM particle is set to 1 GeV.

A maximum-likelihood fit is done including all the search regions of the different analyses included in the stop combination. The results are shown in Figure 5.22 in terms of exclusion limitson $\sigma(pp \to t\bar{t}\chi\bar{\chi})/\sigma_{theory}$ at 95% CL as a function of the DM mediators mass. Scalar and pseudoscalar mediators with masses up to 400 and 420 GeV are excluded, respectively. Previous results are improved by more than 100 GeV [159, 160] and the sensitivity extends beyond $m_{\phi/a} > 2m_t$ for the first time.

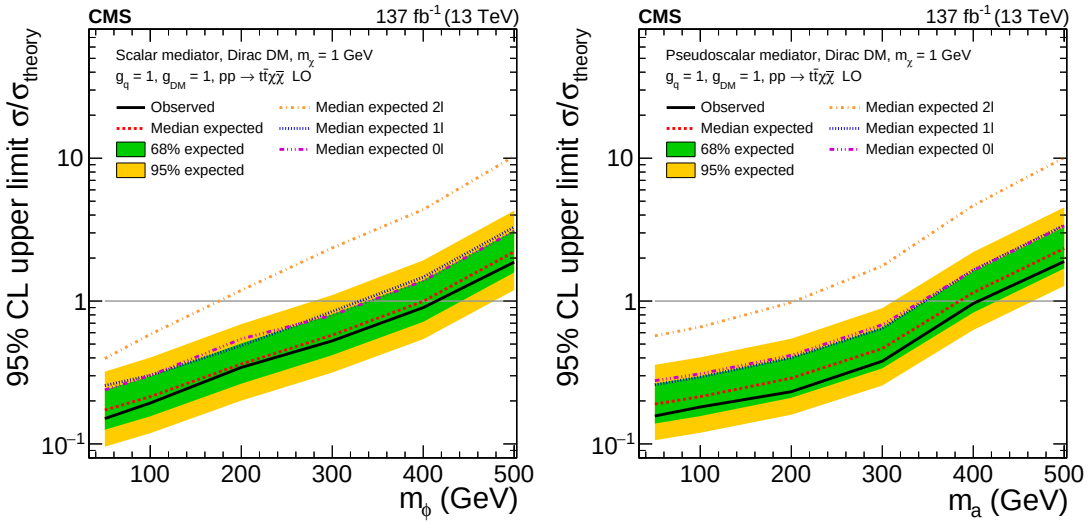

FIGURE 5.22: The 95% CL expected (dashed line) and observed limits (solid line) on $\sigma/\sigma_{\text{theory}}$ for a DM particle with $m_\chi = 1\,\text{GeV}$, as a function of the mediator mass for a scalar (left) and pseudoscalar (right). The expected individual results of the searches with 0 (pink), 1 (purple) and 2 (orange) leptons in the final state as well as their combination (red) are shown. The green and yellow bands represent the regions containing 68 and 95%, respectively, of the distribution of limits expected under the background-only hypothesis [4].

5.2 Search for EFT in top quark production with additional leptons

The EFT is an indirect flexible framework that allows to search new physics at higher energy scales through WCs and EFT operators characterizing the different possible interactions. I have participated in a search targeting processes in which one or more top quarks are produced in association with additional leptons [5]. In the SM, these signatures arise primarily from associated top quark production involving a W, Z, or Higgs bosons, but many EFT operators can also contribute to these processes through dimension-six vertices involving top quarks and either bosons, leptons, or other quarks.

This analysis focuses on the leptonic decays of the top quark, which provide cleaner final states than the hadronic decays. The effect of 26 dimension-six operators that significantly impact associated top processes are studied simultaneously in 6 signal processes: $\text{t}\bar{\text{t}}\text{H}$ [83], $\text{t}\bar{\text{t}}$ in association with one lepton and one neutrino ($\text{t}\bar{\text{t}}\ell\nu$, representing $\text{t}\bar{\text{t}}\text{W}$) [89], $\text{t}\bar{\text{t}}$ in association with two leptons ($\text{t}\bar{\text{t}}\ell\bar{\ell}$, representing $\text{t}\bar{\text{t}}\text{Z}$) [90], top quark production in association with two leptons and a quark ($\text{t}\ell\bar{\ell}q$, representing top quark production in association with a Z boson and a quark tZq) [91], top quark production in association with a Higgs boson and a quark (tHq) [83] and $\text{t}\bar{\text{t}}\text{t}\bar{\text{t}}$ [163]. This extends the search signatures with respect to the previous related EFT analysis by CMS [164], in which 16 operators were studied. Furthermore, this analysis is performed using the $138\,\text{fb}^{-1}$ of integrated luminosity, corresponding to the full Run 2 dataset.

One of the main challenges is the parameterization of the WCs in the MC simulated events (see Subsection 5.2.2). The 26 operators are summarized in Table 5.2, classified in 4 groups according to the involved particles: four heavy quarks (4hq), two heavy quarks and two light quarks (2hq2ℓq), two heavy quarks and two charged or neutral leptons (2hq2ℓ), and two heavy quarks and gauge or Higgs bosons (2hqV). The definitions of the WCs and their corresponding operators can be found in Table 1 of Reference [165].

Operator category	Wilson coefficients
Two-heavy (2hqV)	$c_{t\varphi}$, $c^{-}_{\varphi Q}$, $c^{3}_{\varphi Q}$, $c_{\varphi t}$, $c_{\varphi tb}$, c_{tW}, c_{tZ}, c_{bW}, c_{tG}
Two-heavy-two-lepton (2hq2ℓ)	$c^{3(\ell)}_{Q\ell}$, $c^{-(\ell)}_{Q\ell}$, $c^{(\ell)}_{Qe}$, $c^{(\ell)}_{t\ell}$, $c^{(\ell)}_{te}$, $c^{S(\ell)}_{t}$, $c^{T(\ell)}_{t}$
Two-heavy-two-light (2hq2lq)	c^{31}_{Qq}, c^{38}_{Qq}, c^{11}_{Qq}, c^{18}_{Qq}, c^{1}_{tq}, c^{8}_{tq}
Four-heavy (4hq)	c^{1}_{QQ}, c^{1}_{Qt}, c^{8}_{Qt}, c^{1}_{tt}

TABLE 5.2: List of WCs included in this analysis grouped in 4 categories.

Some Feynman diagrams of how WCs from each of these groups can affect associated top quark processes are displayed in Figure 5.23.

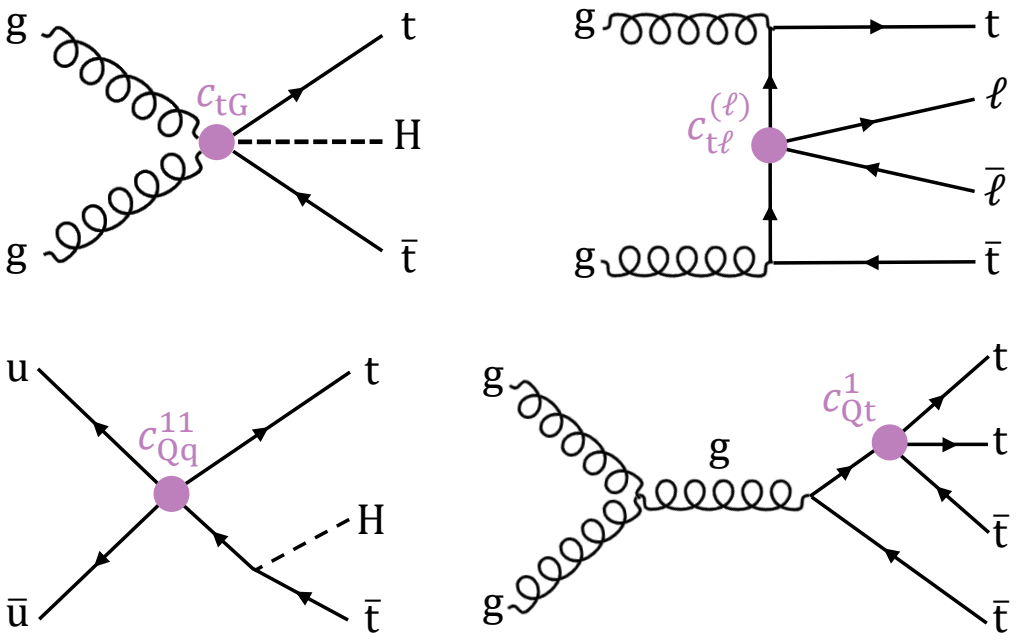

FIGURE 5.23: Feynman diagrams with vertices associated to one WC from each category: from top left to bottom right: c_{tG}, $c^{(\ell)}_{t\ell}$, c^{11}_{Qq}, and c^{1}_{Qt}.

The 95% CL observed limits on most of the listed EFT WCs, assuming new physics contributions from one specific operator at a time, obtained in several searches performed by the ATLAS and CMS Collaborations are shown in Figure 5.24.

A detailed description of the analysis and its results is provided below.

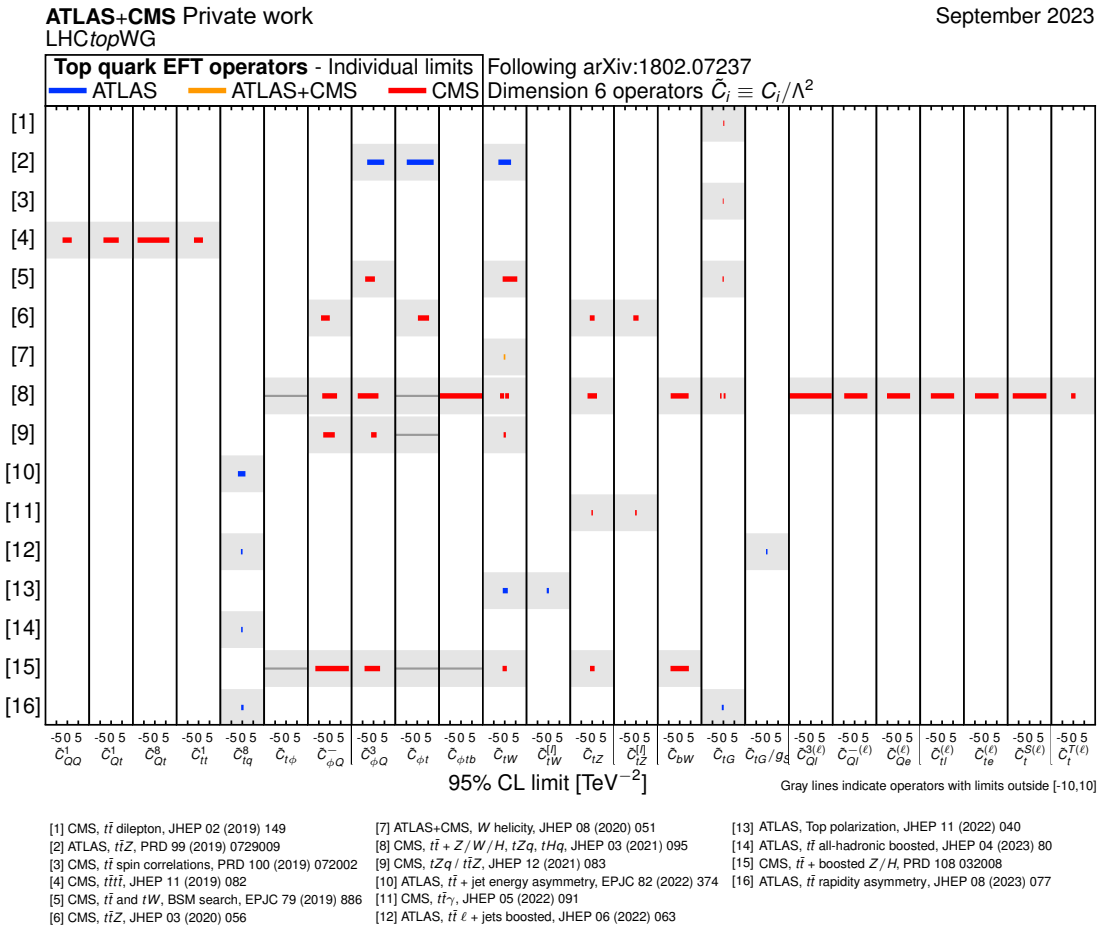

FIGURE 5.24: Summary of the observed limits at 95% CL on the EFT WCs of the dimension-6 operators related to interactions involving top quarks obtained by the ATLAS and CMS Collaborations. The results are reported as individual constraints assuming new physics contributions from one specific operator at a time. Gray lines indicate operators with limits outside [-10,10].

5.2.1 Object and event selection

This search targets final states with several leptons, so their efficient identification and selection is crucial. PF reconstructed electrons and muons are required to pass a tight selection criteria, mainly based on the lepton MVA, to reject nonprompt leptons. Both are selected if they have $p_{\mathrm{T}} > 10\,\mathrm{GeV}$, $|d_{xy}| < 0.05$ cm, $|d_z| < 0.1$ cm, $\mathrm{SIP}_{3\mathrm{D}} < 8$, and mini-isolation < 0.4. Then, electrons (muons) are required to pass the loose WP of the electron BDT (medium WP of the cut-based ID), $|\eta| < 2.5$ (2.4) and the lepton MVA $>$ 0.9 (0.85).

Jets reconstructed with the anti-k_T algorithm with distance parameter $R = 0.4$ are required to pass the tight ID criteria, p_{T} greater than 30 GeV and $|\eta| < 2.4$. To identify b jets, the loose and medium WPs of the DEEPJET b-tagging algorithm are used, corresponding to an efficiency of about 85% and 70% and a mistagging of 10% and 1%, respectively.

As an initial selection of the events, single, double, and triple lepton triggers with p_T thresholds similar to those of the analysis described in the previous section are used. Then, different categories are defined targeting the final states of the signal processes: 2 leptons with the same sign of the electric charge (2ℓSS), three leptons (3ℓ) and four leptons (4ℓ).

- **2ℓSS**: this category targets the $t\bar{t}H$, $t\bar{t}\ell\nu$, and $t\bar{t}t\bar{t}$ signal events. Events in this category must contain two tight leptons with the same electric charge, the leading lepton must have $p_T > 25\,\text{GeV}$ and the subleading $p_T > 15\,\text{GeV}$. The same sign of the electric charge criteria aims to reject the $t\bar{t}$ +jets background, but for further suppression a precise measurement of the electric charges of the electrons and muons is required. After the conditions applied to ensure this, which are described in [166], the charge misidentification rates for electrons are vastly reduced. For muons, the muon misidentification charge rate in the CMS detector is negligible. The 2ℓSS events are required to have a jet multiplicity ≥ 4. These events are divided in two categories according to the b-tagged jet multiplicity: at least 3 medium b-tagged jets (with the aim of having a $t\bar{t}t\bar{t}$ enriched region), and at least 2 loose b-tagged jets with at least one of them passing the medium WP. Finally, events are categorized based on the total lepton charge to account to the fact that $t\bar{t}W^+$ cross section is expected to be almost twice that of $t\bar{t}W^-$.

- **3ℓ**: this category is mostly comprised of $t\bar{t}\ell\nu$, $t\bar{t}\ell\ell$, $t\ell\bar{\ell}q$, and $t\bar{t}H$ signal events. Events must have exactly three leptons passing the tight object selection criteria. In the descending order of p_T, the lepton must have $p_T > 25\,\text{GeV}$, $p_T > 15\,\text{GeV}$, and $p_T > 10\,\text{GeV}$. In case of the third lepton being an electron, this requirement is $p_T > 15\,\text{GeV}$ to reject contributions from nonprompt electrons. The presence of at least 2 jets is also required. Events are separated depending on whether or not they contain a same-flavour pair of leptons with opposite sign of the electric charge such that $|m(Z) - m(\ell^+\ell^-)| < 10\,\text{GeV}$ in order to obtain enhanced contribution from the on-shell Z decay in the $t\bar{t}Z$ process. Any 3ℓ event that does not fall within this region is further separated based on the total lepton charge. In all 3ℓ subcategories, events are further separated based on b-tagged jet multiplicity: exactly 1, and at least 2 jets passing the medium b-tagging WP. The contribution from $t\ell\bar{\ell}q$ is enhanced in the exactly 1 b-tagged jet case, while the other case helps to distinguish $t\ell\bar{\ell}q$ and $t\bar{t}\ell\ell$ events.

- **4ℓ**: this category targets $t\bar{t}\ell\ell$ and $t\bar{t}H$ events. The presence of at least four leptons passing the tight object selection criteria is required. Lepton of first, second, third, and fourth highest p_T must have $p_T > 25$, 15, 10, and 10 GeV, respectively. In case the third and fourth leptons are electrons, the p_T thresholds are instead $> 15\,\text{GeV}$. The presence of at least 2 jets is also required, with at last one of them passing the medium b-tagging WP and the other passing the loose WP.

Events in all categories are then divided into jet multiplicity bins, giving a total of 43 unique categories. A summary of the event categorization is shown in Figure 5.25.

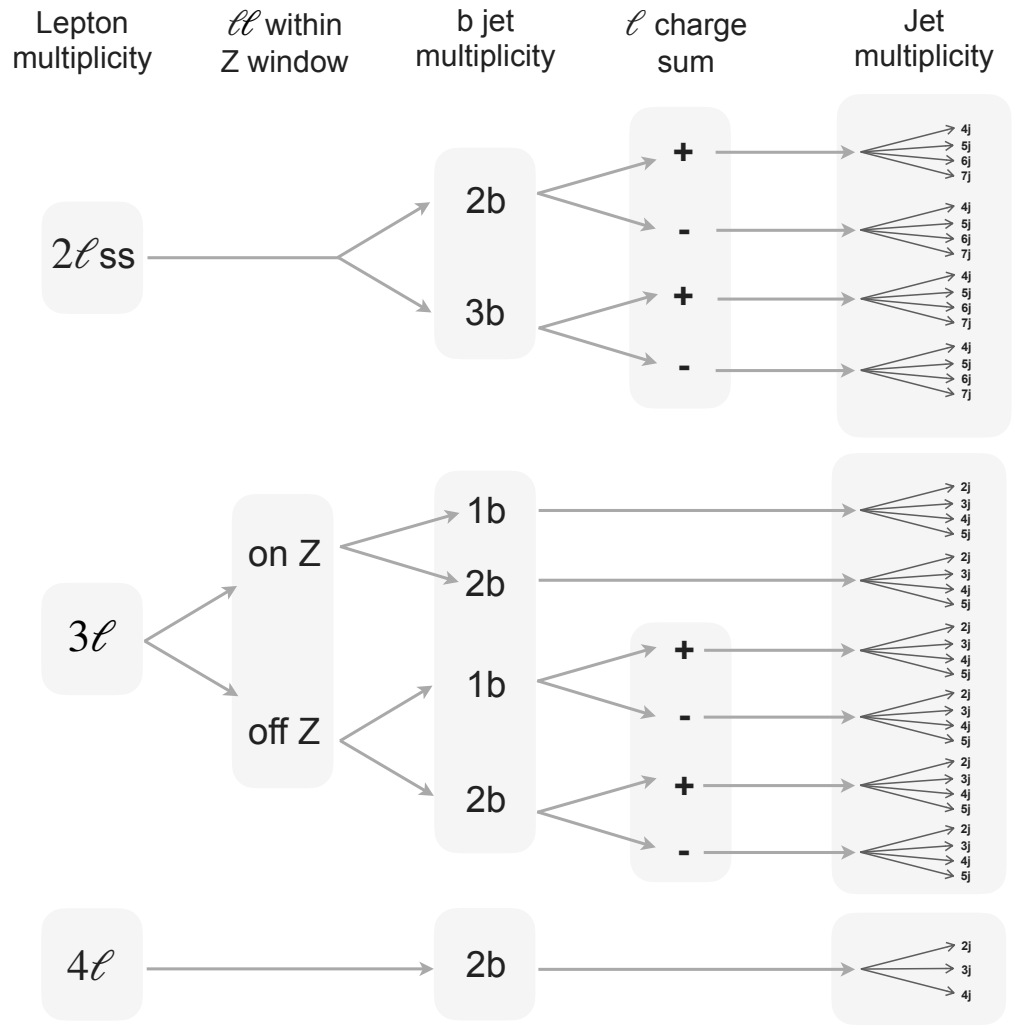

FIGURE 5.25: Summary of the 43 event categories of the analysis [5].

5.2.2 Signal and background estimation

5.2.2.1 Signal samples

Processes involving one or more top quarks produced in association with additional charged leptons constitute the signal for this analysis. These processes include $t\bar{t}H$, $t\bar{t}\ell\nu$, $t\bar{t}\ell\bar{\ell}$, $t\ell\bar{\ell}q$, tHq, and $t\bar{t}t\bar{t}$. The signal samples are produced at LO with the MADGRAPH (v2.6.5) event generator, using the DIM6TOP model [165] to incorporate the EFT effects. Focusing on operators that give rise to interactions involving top quarks, listed in Table 5.2, this model uses the Warsaw basis [24] of gauge-invariant dimension-six operators, and is capable of producing only tree-level modelling of their effects. In order to allow MADGRAPH to properly handle the emission of gluons from the vertices involving the c_{tG} WC, which impacts interactions involving top quarks, gluons, and

the Higgs boson, an extra factor of the strong coupling g_s is applied to this WCs, as explained in Reference [167]. Furthermore, to avoid overlap between the $t\bar{t}\ell\bar{\ell}$ and $t\bar{t}H$ samples, and between $t\ell\bar{\ell}q$ and tHq, $t\bar{t}\ell\bar{\ell}$ and $t\ell\bar{\ell}q$ processes are specifically required in MadGraph to not include an intermediate Higgs boson.

The simulations use the NNPDF3.1 set of PDFs. Parton showering and hadronization for the samples is performed with the pythia 8 (v240) generator with the CP5 tune, which also handles the decays of the top quark and the Higgs boson. The top quark mass used in the simulation is 172.5 GeV. All simulated signal processes are normalized with their respective NLO cross sections, which may include QCD and EWK corrections. They are listed in Table 5.3.

Process	Cross section (pb)	Accuracy	Reference
$t\bar{t}H$	$0.5071 \pm 2.4\%\text{(PDF)} \, ^{+7.6\%}_{-7.1\%}\text{(QCD)}$	NLO (QCD + EWK)	[168]
$t\bar{t}\ell\bar{\ell}$ ($m_{\ell\ell} > 10\,\text{GeV}$)	$0.281 \, ^{+12\%}_{-10\%}\text{(QCD)}$	NLO (QCD + EWK)	[168]
$t\bar{t}\ell\nu$	$0.235 \, ^{+10\%}_{-11\%}\text{(QCD)}$	NLO (QCD + EWK)	[169]
$t\ell\bar{\ell}q$ ($m_{\ell\ell} > 30\,\text{GeV}$)	$0.076 \pm 2.7\%\text{(PDF)} \pm 2.0\%\text{(QCD)}$	NLO QCD	[46, 170, 171]
tHq	$0.071 \pm 5.1\%\text{(PDF)} \, ^{+6.5\%}_{-15\%}\text{(QCD)}$	NLO QCD	[168]
$t\bar{t}t\bar{t}$	$0.01337 \pm 6.9\%\text{(PDF)} \, ^{+3.6\%}_{-11\%}\text{(QCD)}$	NLO (QCD + EWK) + NLL′	[172]

Table 5.3: Theoretical cross sections at NLO used for normalization of simulated signal samples.

The six signal processes include diagrams with zero EFT vertices, corresponding to the SM contributions, and diagrams with one EFT vertex, accounting to the possible new physics contribution. In order to compute the predicted yields in terms of the WCs, it is first necessary to understand how the cross section depends on the WCs. Starting with the ME, it is defined for a given process as the sum of the SM and new physics components:

$$\mathcal{M} = \mathcal{M}_{\text{SM}} + \sum_i \frac{c_i}{\Lambda^2} \mathcal{M}_i, \tag{5.2}$$

where $\mathcal{M}_{\text{SM}}$ is the SM ME, $\mathcal{M}_i$ are the MEs corresponding to the new physics components, and c_i are the WCs.

Since the cross section is proportional to the square of the ME, it will depend quadratically on the WCs, and as the weight of each generated event corresponds to the events contribution to the inclusive cross section, each event weight will also depend quadratically on the 26 WCs.

The MadGraph event reweighting [173] method is used to determine the 26-dimensional quadratic parameterization. The quadratic function describing the weight of event i in terms of the WCs can be written as follows:

$$w_i\left(\frac{\vec{c}}{\Lambda^2}\right) = s_{0i} + \sum_j s_{1ij} \frac{c_j}{\Lambda^2} + \sum_j s_{2ij} \frac{c_j^2}{\Lambda^4} + \sum_{j,k} s_{3ijk} \frac{c_j}{\Lambda^2} \frac{c_k}{\Lambda^2}, \tag{5.3}$$

where $\vec{c}$ represents the set of WCs, the sum over j and k corresponds to the sum over the WCs, and the values s_0, s_1, s_2, and s_3 represent the coefficients in the quadratic parameterization of the weight from the SM, the interference between EFT and SM, the pure EFT contribution, and the interference between two different EFT contributions, respectively. Since the sum of multiple quadratic functions is also quadratic, the dependence of any observable bin can be found by simply summing the quadratic weight functions for every event that passes the selection criteria for the given bin.

For the $t\bar{t}X$ processes ($t\bar{t}H$, $t\bar{t}\ell\nu$, and $t\bar{t}\ell\bar{\ell}$), an additional final-state parton is included in the ME generation to improve the modelling at high jet multiplicities. This can also significantly impact the dependence of the $t\bar{t}X$ processes on the WCs [167]. The single top processes ($t\ell\bar{\ell}q$ and tHq) and the $t\bar{t}t\bar{t}$ sample are not produced with an additional parton due to technical complications associated with correctly performing the jet matching between the ME and the PS for t-channel and single top processes. An additional uncertainty is applied to these processes to cover this effect.

5.2.2.2 Background estimation

The simulated samples used in the analysis to estimate the MC backgrounds include diboson production, triboson production and tWZ processes. Additional samples are also used to simulate SM background processes that are estimated from data, and are used for validation purposes. These samples include Z+jets, W+jets, $t\bar{t}$+jets, DY, and some single top processes (s-, t-channel). $t\bar{t}$ and $t\bar{t}\gamma$ samples are also included in the analysis as they can contribute via photon conversions.

The MC samples used in the analysis are estimated using simulations with MadGraph (triboson, tWZ, s-channel, Z+jets, W+jets, $t\bar{t}$ +jets, $t\bar{t}\gamma$, DY), powheg (ZZ, WW, $t\bar{t}$, t-channel, tW) and MCFM [174] (gg) generators. pythia 8 generator with the CP5 tune is used in all of them.

The background contributions can be divided in reducible and irreducible backgrounds. A background is considered to be irreducible if all final state leptons are genuine prompt leptons. The contribution from irreducible backgrounds in the singal regions is dominated by diboson (WZ and ZZ) production, but a subleading contribution from triboson and tWZ production processes is expected.

Three types of reducible backgrounds are distinguished arising from the misreconstruction or misidentification of objects, and are estimated with data-driven methods.

The primary contribution is due to processes with nonprompt leptons that pass the tight selection criteria. It is estimated following the same technique described in Subsection 4.4.2.1, where the AR is defined as the SR but with a looser selection criteria of the leptons. The probability for a nonprompt lepton that passes the looser lepton selection to pass the tight selection is measured in a sample of multijet events

collected by a set of low-p_T lepton triggers as a function of the lepton p_T and η, separately for electrons and muons. These weights are then applied to the SR leptons. The nonprompt lepton background estimations is validated with $t\bar{t}$ and multijet events in dedicated 2ℓSS CRs.

The second leading contribution to reducible backgrounds is in the 2ℓSS SRs due to cases in which the charge of one of the leptons in an event with a lepton pair with opposite sign of the electric charge is measured incorrectly. This contribution, which is primarily dominated by $t\bar{t}$ events, is estimated by selecting events passing the 2ℓSS SR but inverting the same sign of the electric charge criteria on the dilepton pair. The probability of each of the leptons to have their charge measured incorrectly is used to weight events in this region. The measurement of the probability is performed with DY and $t\bar{t}$ events by counting the number of electrons that pass the tight requirements that have their charges mismeasured according to the MC information. This probability is negligible for muons, but for electrons is larger. The validity of the measured probabilities is checked in a CR with 2 tight leptons within 30 GeV of the Z peak and less than 4 jets (orthogonal to the 2ℓSS SR), where the misidentified charge contribution is estimated by scaling events with a lepton pair with opposite sign of the electric charge in this region by the probabilities. Comparing the prediction to the actual data with the same sign of the electric charge in the CR, a SF is derived to account to the found differences, and a 30% of uncertainty is applied to the estimation.

Finally, a small contribution arises from the conversion of photons interacting with the detector material. This contribution is estimated with $t\bar{t}\gamma$ events.

In order to improve the modelling of the data, the standard corrections (see Section 3.3) are applied to simulated events in terms of: PU reweighthing, lepton reconstruction, identification and isolation efficiencies derived specifically for this analysis using the tag-and-probe method, trigger efficiencies also derived specifically for this analysis using the cross-trigger method with p_T^{miss} triggers, b-tagging efficiencies, jet energy scale and jet energy resolution corrections and L1 prefiring correction.

The number of observed data and MC simulated events in each of these categories is shown in Figure 5.26. Here it can be seen that each category is populated by different signal processes, being the ones with a larger amount of events 2ℓSS with 2 b-tagged jets and 3ℓ on-shell Z.

5.2.3 Analysis strategy

In each of the 43 categories, a kinematic variable is used for the signal extraction to get the highest sensitivity as possible. Several variables were tested keeping in mind that the WCs cannot be fully isolated and cannot be associated with a single event category, so it is not possible to chose a particular variable for each WC. Nevertheless,

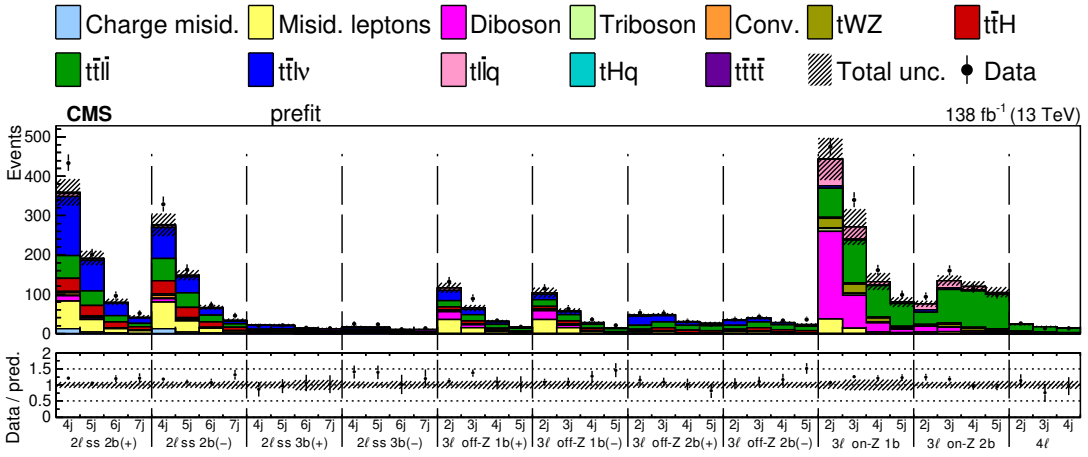

FIGURE 5.26: Number of data and MC predicted events in the 43 categories of the analysis [5].

since some WCs impact certain event categories more than others, it is possible to target WCs by choosing specific variables for event categories that may be particularly sensitivity to the given WCs.

Comparing the expected exclusion limits obtained by several variables, the best option was found to be the variable $p_{\mathrm{T}}(\ell j)_{max}$ in most of the categories. This variable is defined as the p_{T} of the leading pair of all possible combinations of leptons and jets selected in the event, thus, it may represent the p_{T} of two leptons, two jets, or a lepton and a jet, making it very useful because it has good sensitivity to most of the studied WCs. For the EFT operators involving Z bosons, the sensitivity is enhanced by using the p_{T} of the same-flavour lepton pair with opposite sign of the electric charge associated with the Z boson, denoted as $p_{\mathrm{T}}(\mathrm{Z})$. This variable is used in all the 3ℓ on-shell categories, except of those with two or three jets and two b-tagged jets, where the $p_{\mathrm{T}}(\ell j)_{max}$ is used instead. An example of the $p_{\mathrm{T}}(\ell j)_{max}$ distribution for events in the 3ℓ off-shell categories with 1 b-tagged jet and negative sum of the leptons charge is shown in Figure 5.27 (left), and an example of the $p_{\mathrm{T}}(\mathrm{Z})$ for events in the 3ℓ on-shell categories with 1 b-tagged jet is shown in Figure 5.27 (right).

The final distribution used to do the fit is shown in Figure 5.28. Four bins are used for the categories that are binned in $p_{\mathrm{T}}(\ell j)_{max}$ and five bins are used for those binned in $p_{\mathrm{T}}(\mathrm{Z})$, which often have more statistics. This gives a distribution with a total of 178 bins. A reasonable description of the data by the SM prediction is observed in all the bins, within the uncertainty bands.

The number of observed events in each bin is treated as an independent Poisson measurement [175], and a maximum likelihood fit is performed over all of the bins. The 26 WCs, which are the parameters of interest in the likelihood fit, are used to parameterize the total yield in each bin as a quadratic function. This yield is affected by the

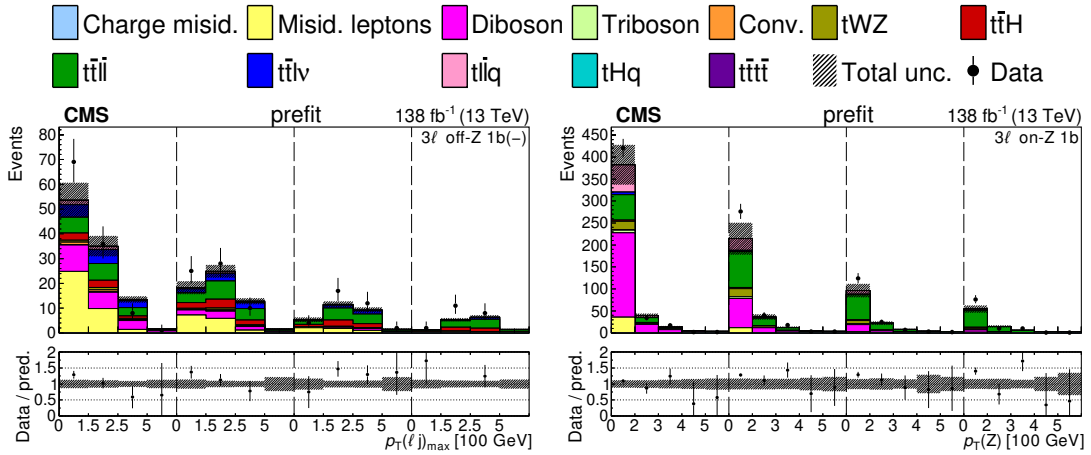

FIGURE 5.27: The $p_T(\ell j)_{max}$ distribution for data and SM predicted events in the 3ℓ off-Z 1b $(-)$ categories (left), and the $p_T(Z)$ distribution for 3ℓ on-Z 1b categories (right). The jet subcategories are arranged from low jet multiplicity to high jet multiplicity from left to right [5].

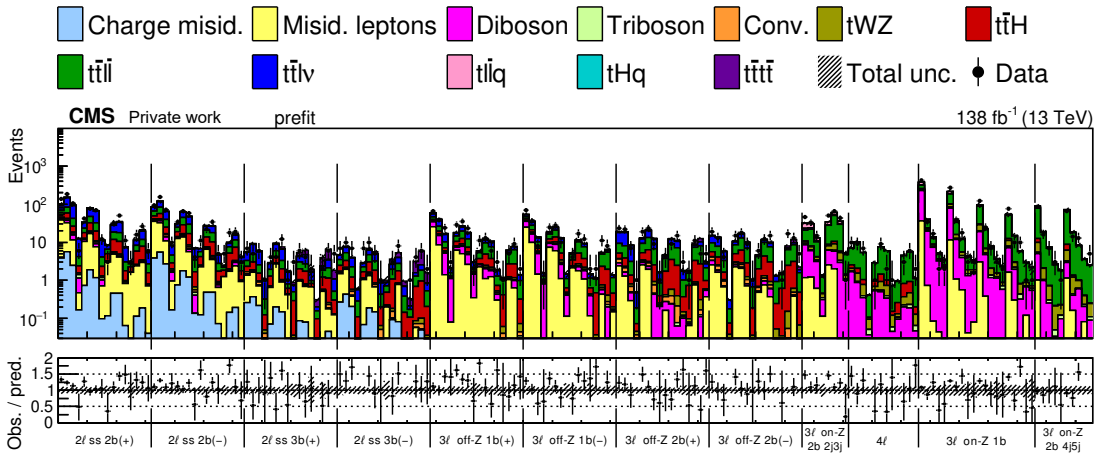

FIGURE 5.28: Final distribution for data and SM predicted events including the selected kinematic variable of each category. Each event category is subdivided into its jet multiplicity components. For example, the first four bins of the 2ℓSS 2b (+) category are the $p_T(\ell j)_{max}$ variable for four jets, the next four bins are the $p_T(\ell j)_{max}$ variable for 5 jets, etc.

systematic uncertainties (enumerated in Section 5.2.4), which are handled as NPs in the fit.

Two separate likelihood fits are performed over each WC to find the values of the other WCs or NPs which minimize the NLL function, denoted as 1D scans: one where the other 25 WCs are treated as unconstrained NPs (profiled), and one where the other 25 WCs are fixed to their SM values of zero. The 1σ and 2σ limits are extracted by finding the WC values where the twice the difference in the NLL (2ΔNLL) curves cross one and four, respectively. These values of the 2ΔNLL curve correspond to the 68.27% and 95.45% confidence intervals (CIs), respectively.

Additionally, simultaneous scans are carried out for a particular subset of WC pairs, referred to as 2D scans. The 2D scans are performed in a similar manner to the 1D scans where distinct set of points of the WCs pair are studied, while the other 24 WCs are either profiled or fixed to their SM values of zero.

5.2.4 Systematic uncertainties

As in the rest of the analyses presented in this thesis, systematic uncertainties are taken into account, all of them treated as 100% correlated across all processes and event categories.

The standard experimental uncertainties are considered: uncertainties on the lepton identification and isolation, trigger and b-tagging efficiencies, the jet energy resolution uncertainty and the jet energy scale uncertainty, the uncertainty on the PU reweighting and the uncertainty due to the L1 prefiring correction.

An uncertainty of 30% is assigned to the charge-flip estimation to account for the differences observed between the MC prediction and data observed in the charge-flip control region. The uncertainty on the nonprompt lepton background is on average 3%, including statistics and systematic uncertainties of the method used to do the estimation. An additional uncertainty due to the jet mismodelling observed in some of the diboson CRs with high jet multiplicities is derived ad the difference between data and simulation, and its value is on average 7%.

The modelling uncertainties included in this analysis are the initial- and final-state radiation in the PS simulation uncertainty, uncertainties in the matrix element generators due to the μ_R and μ_F scales, and an uncertainty due to limitations in the number of final-state partons generated with MadGraph in the $t\ell\bar{\ell}q$ and tHq LO EFT samples. For each jet bin, an uncertainty equal to the difference between the $t\ell\bar{\ell}q$ LO and $t\bar{t}Z$ NLO samples which are not covered by all other systematic uncertainties is added.

A summary of the experimental and modelling uncertainties and their relative average variation on the nominal SM predicted yields can be found in Table 5.4.

Signal and background expected yields are normalized using theoretical cross sections produced at NLO or greater and uncertainties in the normalization due to the PDF and QCD scale choices are accounted for. The average uncertainty across all the final distribution bins due to the scales is 1–4% of the total nominal predicted rate, while that from the PDFs is 1%.

The uncertainty of the integrated luminosity is 1.6%, as in the stop quark search presented in the previous section.

Source	Uncertainties (%)
Electron efficiency	2
Muon efficiency	1
Trigger efficiency	$\leq$1%
b-tagging efficiency	1%
Jet energy scale and resolution	1%
Misidentified-lepton rate	3%
Charge misreconstruction rate	1%
Jet mismodelling	7%
Pileup reweighting	1%
L1 prefiring	1%
Initial- and final-state radiation	1–2%
Additional radiation	7%
μ_F, μ_R scales	3%

TABLE 5.4: Summary of the systematic uncertainties. The average variations in the SM prediction yield along the bins is shown.

5.2.5 Results

The likelihood fits described in Subsection 5.2.3 are done and the post-fit final distribution obtained by fitting simultaneously the 26 WCs and the NPs minimizing the NNL function are shown in Figure 5.29 (top). The post-fit yields of each category is also shown in that figure (bottom). In all cases, data are found to be consistent with SM expectations.

The 1σ and 2σ CIs are extracted from the likelihood fits for each WC and shown in Figure 5.30. The limits represented in black correspond to the case where the other WCs are profiled, while the limits represented in red correspond to the case where the other WCs are fixed to their SM values of 0. All the CIs are consistent with the SM hypothesis.

The disjoint 1σ intervals observed in certain scans where the other 25 WCs are fixed to zero, in particular the 4hq WCs, arise due to the inherent quadratic nature of the EFT parameterization. In principle, this inherent degeneracy could apply to all WCs, but the degeneracy can be resolved when the contributions from multiple processes across various bins result in one of the two minima demonstrating significantly better agreement with the observed data. However, when profiling across the other 25 WCs, the interferences between these WCs can compensate for each other effects within the range between the two minima. This compensation leads to the emergence of a single, elongated minimum instead of two disjoint minima. This would be the case, for example, of the c^1_{Qt} WC, whose 1D scan is shown in Figure 5.31 (left). Figure 5.31 (right) shows an example of a 1D scan with just one minima, corresponding to the $c^{(\ell)}_{\mathrm{Qe}}$ WC.

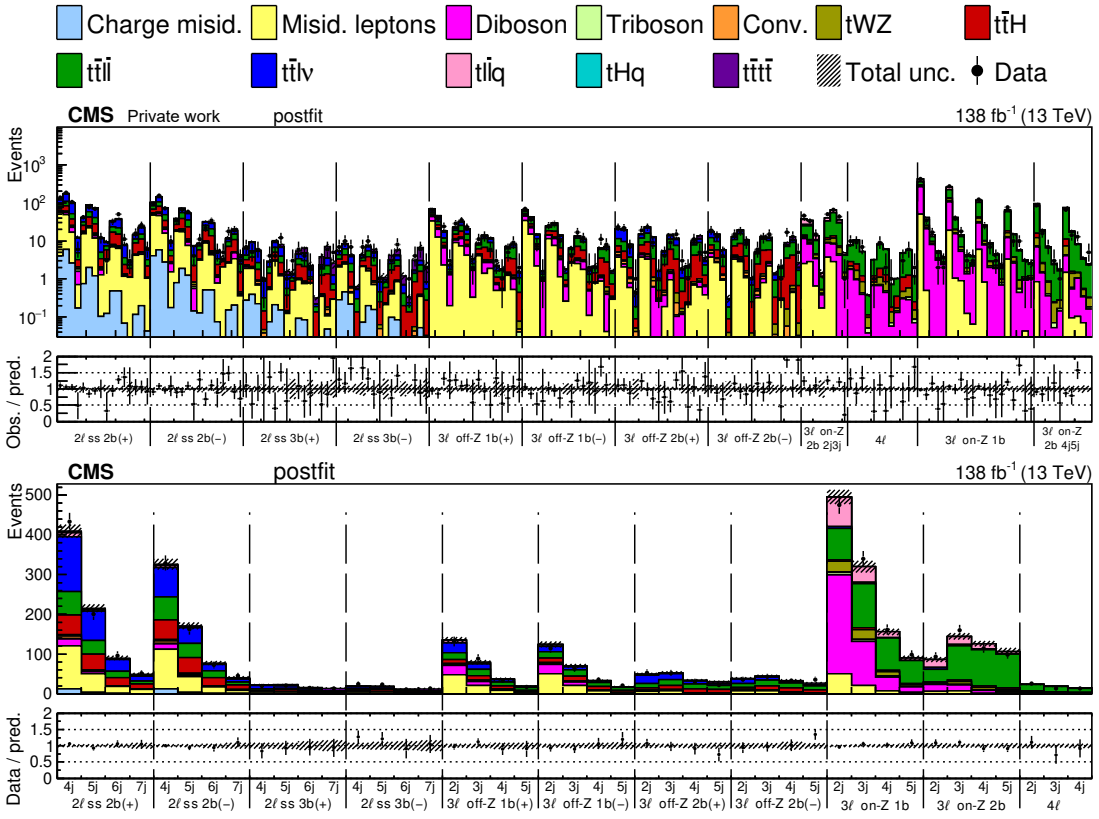

FIGURE 5.29: Distributions for data and SM predicted events in the post-fit scenario, including the selected kinematic variable of each category (top), or only the yields (bottom) [5]. The post-fit values are obtained by simultaneously fitting all 26 WCs and the NPs.

This result significantly improves the CIs obtained by the previous analysis, described in Reference [164], enhancing it by factors of around 2 to 6, depending on the WCs.

The CIs obtained from both the profiled and fixed to SM scans are very similar in most of the WCs. This can be attributed to the absence of significant correlations between the considered WCs. The correlation between the WCs is explored with 2D scans described in Subsection 5.2.3. Several pairs were studied and negligible correlations were found in most of the cases. In the ones where nonnegliglible correlations appear, the 2D scans in which the other 24 WCs are profiled are very similar to the 2D scans in which the other WCs are fixed to zero, indicating that while the correlations between the given pair of WCs are important, the correlations with the other 24 are less significant. An example of this can be found in Figure 5.32, where the 2D scans over the c_{tZ} and c_{tW} WCs with the other 24 WCs fixed to their SM values of zero (on the left) or profiled (on the right) are shown. The correlation between both WCs is evident.

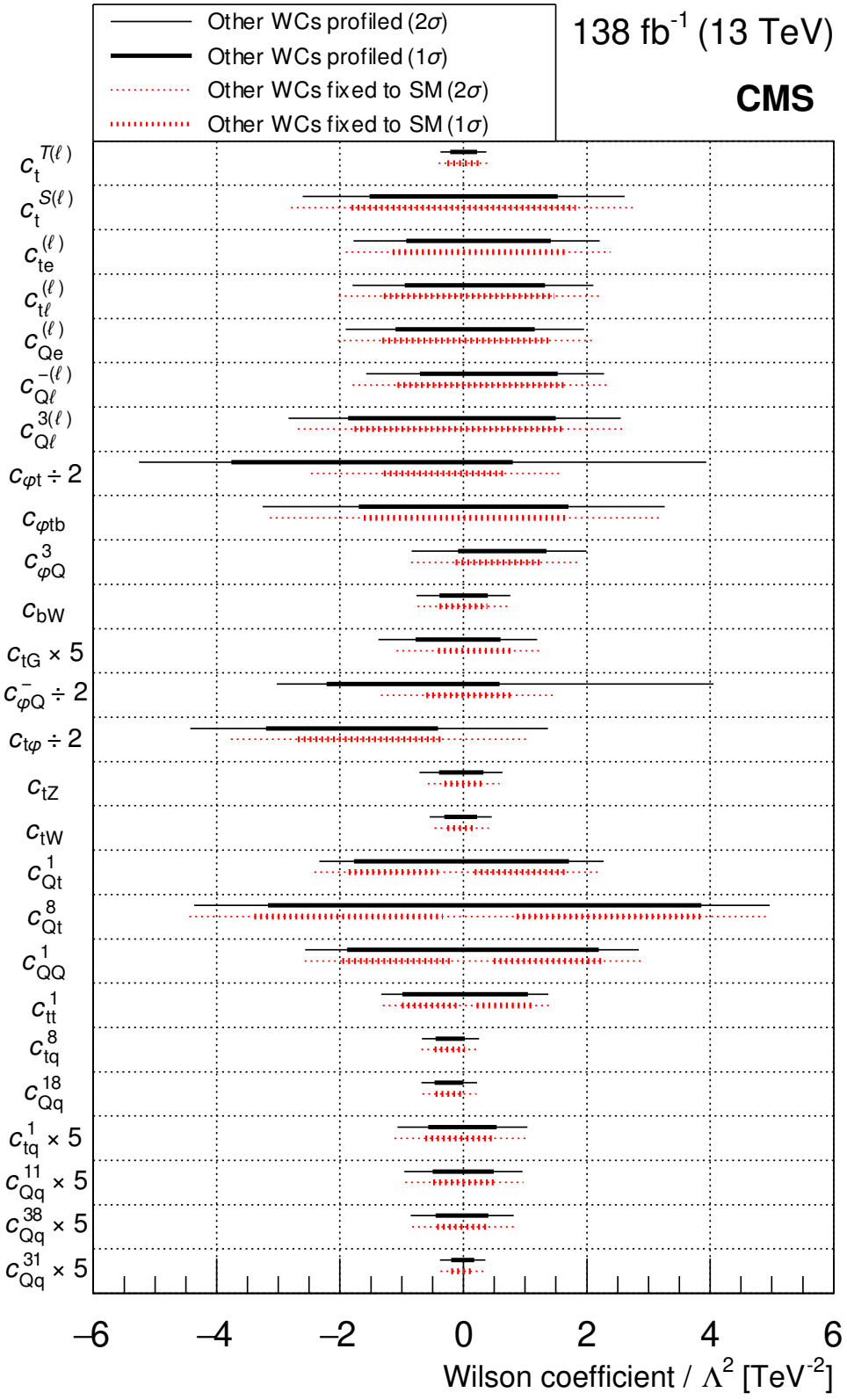

FIGURE 5.30: CIs extracted from the likelihood profiled and fixed to SM fits. The WC 1σ (thick line) and 2σ (thin line) CIs are shown. To make the figure more readable, the intervals for $c_{t\varphi}$, $c_{\varphi t}$, and $c_{\varphi Q}^{-}$ were scaled by 0.5, and the intervals for c_{tG}, c_{tq}^{1}, c_{Qq}^{11}, c_{Qq}^{38}, and c_{Qq}^{31} were scaled by 5 [5].

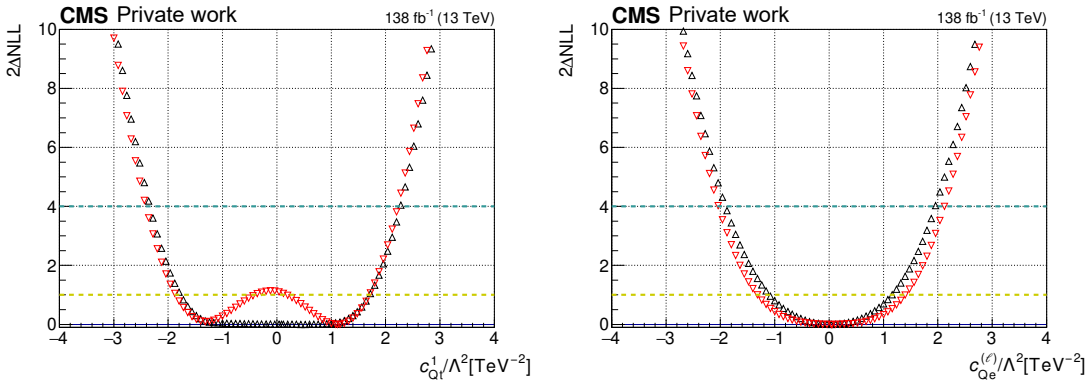

FIGURE 5.31: 1D scans over the c^{1}_{Qt} (left) and $c^{(\ell)}_{Qe}$ (right) WCs. The red points correspond to the case where the other WCs are fixed to their SM values of zero, while the black points correspond to the case where the other WCs are profiled. Green and blue dashed lines indicates the 1σ and 2σ CIs, respectively.

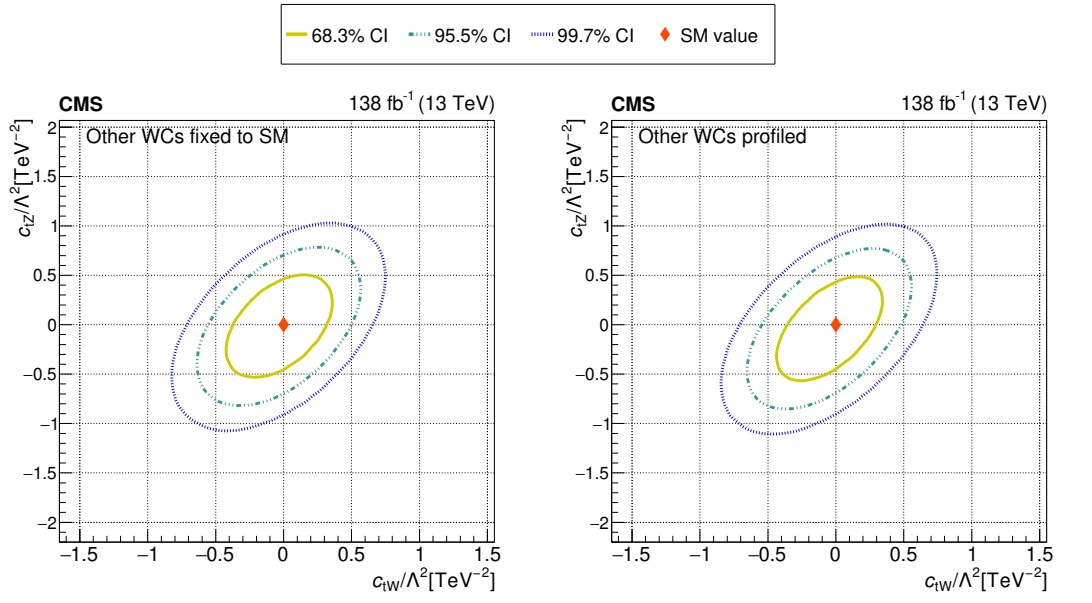

FIGURE 5.32: The observed 68.3, 95.5, and 99.7% confidence level contours of a 2D scan for c_{tW} and c_{tZ} with the other WCs fixed to their SM values (left), and profiled (right). Diamond markers show the SM prediction [5].

Chapter 6

Machine Learning techniques to improve muon identification in Run 3

The presence of leptons is key for interesting physical events, as has been shown throughout the analyses presented in this thesis where the lepton identification is a crucial ingredient in all of them. In particular, analyses where W, Z, Higgs bosons, or τ leptons are involved need to apply a requirement to select prompt leptons. Besides, certain analyses target topologies where muons are expected to come from the decay of B or C hadrons. However, muons coming from spurious hits or decays in flight from light-flavour hadrons (pions or kaons) can be reconstructed as muons and, hence contaminate measurements and searches. This may be problematic, especially if the signal cross section is significantly small. Because of this, the discrimination of signal leptons from misidentified leptons becomes crucial.

The strategy used in the muon identification during Run 2 relied on a cut-based approach using a set of requirements on individual variables related to the information from the tracker and the muon systems. This approach, as has been discussed in the course of the thesis, defines three WPs [69], loose, medium and tight, each of them tuned to have different balances between efficiency and purity. Their efficiencies, measured with Run 2 data by selecting muons with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$, are 99.75, 98.25, and 96.00%, respectively, with statistical uncertainties between 0.02 and 0.03% for muons with $|\eta| < 0.9$. Then, for muons with $|\eta| > 0.9$ the efficiencies for each WP are 99.77, 98.55, and 97.46% with the statistical uncertainties between 0.02 and 0.04% [69]. The most commonly used in analysis are medium and tight WPs, which offer a good compromise between efficiency and nonprompt leptons rejection.

The cut-based ID approach has been proven to work successfully during Run 2. However, the performance of muon identification worsens as the number of PU interactions

increases, so the use of more sophisticated techniques is crucial to maintain an acceptable compromise between efficiency and purity for Run 3 collision conditions, where the average of PU interactions is over 50 in comparison with 37 PU interactions during Run 2 (see Figure 3.1). With this aim, a MVA approach is designed, referred to as "muon MVA ID" [6]. The purpose of the muon MVA ID is to provide a general and robust discriminant for muon identification that improves the performance of the cut-based ID approach. In addition, this MVA offers another advantage, given that it provides a continuous variable, each analysis could select the most suitable trade-off between efficiencies and nonprompt rates, offering more flexibility than just 3 WPs.

It has already been established that such techniques yield favorable performance in the selection of objects, as in the case of the previously mentioned lepton MVA to separate prompt leptons from nonprompt leptons. This MVA was initially developed in $t\bar{t}H$ multilepton analysis [83], in which the use of the lepton MVA leads to a reduction in the nonprompt lepton background from $t\bar{t}$ events by a factor of 4 with respect to the use of an equivalent cut-based technique, and has also been successfully used in several analyses involving leptons. Some examples are other $t\bar{t}H$ measurements [84, 85], the measurement of the WZ process [86, 87] or the multiboson production at $\sqrt{s} = 5.02\,\text{TeV}$ [3] where the use of this MVA significantly reduces the poorly modelled nonprompt lepton background, the search for electroweak production of supersymmetric particles [88], the observation of tZq production [176] and $t\bar{t}t\bar{t}$ [55], and in other precision top quark measurements [89–91].

The muon MVA ID has not yet been used in any analysis, but is expected to give results as satisfactory as the lepton MVA. In this chapter the details of the muon MVA ID training and the performance obtained are presented.

6.1 Data and simulated samples

Both the MVA training and the performance studies are based on the data recorded in 2018, belonging to the Run 2 data-taking period of pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with a total integrated luminosity of $59.7 \pm 1.5\,\text{fb}^{-1}$ [144] recorded with the CMS detector. A single muon trigger is used requiring the presence of a muon with $p_T > 24\,\text{GeV}$ fulfilling loose isolation requirements [177].

For the efficiency studies, apart from the data events, a DY MC sample with dilepton events simulated at NLO in QCD with the MadGraph_aMC@NLO (v2.6.1) generator is employed. For the training and the nonprompt rate measurement, $t\bar{t}$ events including semileptonic decays generated with POWHEG (v2) at NLO accuracy in QCD are used. For all the samples, the NNPDF3.1 NNLO PDF set is used, and the PS and hadronization are simulated with PYTHIA 8 (v.240) using the UE CP5 tune.

Two independent $t\bar{t}$ samples are used, one for the MVA training and the other for performance studies. The difference between them is that the last one includes generation-level information for the particles produced in the PU interactions, whereas the training sample does not. To determine the source of a given reconstructed muon in a simulated sample, geometrical coincidences between the hits that make their trajectory in the muon system and the hits produced by the interaction of a given simulated particle with the detector are looked for, establishing a one-to-one correspondence between reconstructed muons and simulated particles which allows to know the generated parent particle. In the DY and the $t\bar{t}$ training samples, only generated particles corresponding to the primary vertex are considered. However, to enhance the definition of the background sources, a dedicated $t\bar{t}$ sample that includes generation information for particles from all primary vertices (PU matching) are also used to check the nonprompt lepton rate. This sample was created specifically for muon studies, not for use in analysis, which is the reason why it has a number of events an order of magnitude less than the other one.

6.2 Muon MVA ID development

6.2.1 Initial selection

The choice of a $t\bar{t}$ sample for the training provides a wide variety of sources of genuine muons, including muons from prompt decays and from heavy-flavour hadron decays.

The training is carried out with muons passing the loose cut-based ID WP, p_T greater than 10 GeV and $|\eta| < 2.4$. The loose muons are particles identified as a muon by the PF algorithm and that are also reconstructed either as a global muon or as a tracker muon, this helps to avoid poorly reconstructed muons with almost 100% efficiency. The requirement of $p_T > 10$ GeV is imposed to avoid low-p_T muons coming from low-mass resonances that have different features than muons from Z boson decays, for example. To identify these low-p_T muons, another special MVA discriminator already exists known as "soft MVA ID" [178]. The muon MVA will target medium-p_T muons.

Then, muons passing the preselection described above are divided in signal and background muons according to the particle from which they are originated, using generation information. As the muon MVA ID aims to be used in different analyses, not only muons from prompt decays of W, Z, and Higgs bosons are considered as signal. Isolated muons from τ leptons decays and nonisolated muons from heavy-flavour decays are also taken as signal. On the other hand, background muons come from the decays of pions and kaons or other light-flavour hadrons, and from spurious signatures in the detector misreconstructed as muons, such as punch-throughs. Reconstructed muons that can not be matched to any generated particle are also considered as background.

For the training, the $t\bar{t}$ sample without PU matching was selected because its number of events is much higher than the one with PU matching. As a result, real muons from PU interactions are included as background in the category where the reconstructed muon is not matched to a generated particle, which leads into a signal contamination. A cross-check was done comparing training performed with each of the samples and found them to be consistent, as expected from the studies presented in Reference [179].

The $t\bar{t}$ training sample composition of muons passing the preselection is shown in Table 6.1. Approximately 1 million muons pass the preselection, but of them only 2% are background muons. The class-imbalance between signal and background muons is the main challenge of the MVA development, therefore it is important to use the sample with higher statistics for the training.

Classification	Muons from...	Composition (%)
Signal	bosons	60.4
	τ leptons	7.7
	B hadrons	20.1
	B to C hadrons	7.3
	C hadrons	2.6
Background	light-flavour hadrons	1.0
	not matched particle	0.7
	punch-throughs	0.2

TABLE 6.1: $t\bar{t}$ training sample composition in terms of MC generation matching.

6.2.2 Input variables

The muon MVA ID is trained with the identification variables used to define the cut-based ID WPs, which have already been mentioned in previous chapters of the thesis (for instance, see Subsections 4.2.2.2 and 4.4.1). The impact parameter variables (d_{xy} and d_z), which are part of the tight WP criteria, are excluded of the training as non-prompt muons produced in heavy-flavour hadron decays should be selected by the MVA. Thus, the training is more general, only to identify muons, and to discard non-prompt muons from heavy-flavour hadrons further isolation requirements or a custom impact parameter selection criteria should be applied at the analysis level. Another option would be to use a combination of the muon MVA ID and the lepton MVA, which does include isolation variables.

Apart from the identification variables, the p_T and η of the muons are also introduced in the training as the various signal and background contributions have different p_T and η profiles, as can be seen in Figure 6.1. Although the contribution of each muon source is fairly constant through the η distribution, different p_T profiles appear being the low-p_T region enriched in heavy-flavour hadron muons and the higher-p_T region dominated by prompt muons from boson decays, as expected.

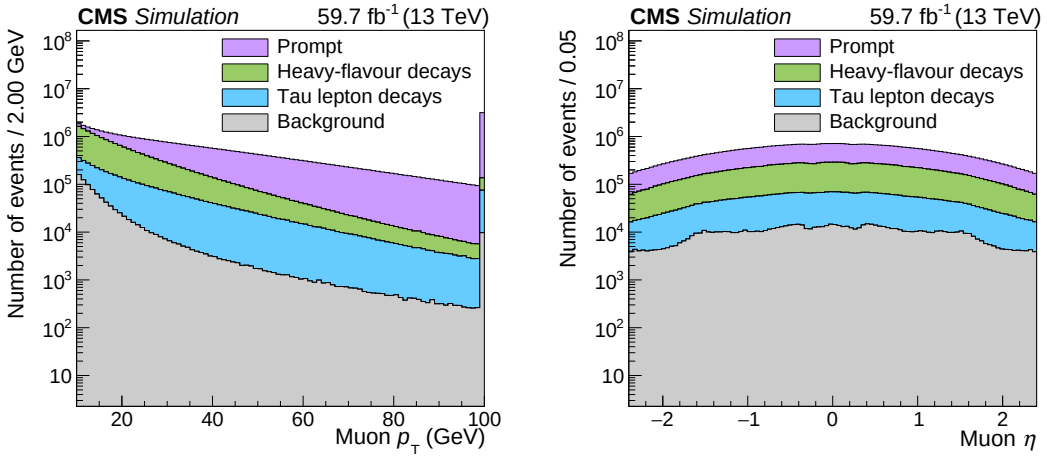

FIGURE 6.1: Composition of the $t\bar{t}$ training sample after the muon preselection as a function of p_T (left) and η (right) [6].

In the training, p_T and η are weighted to have the same distributions in both signal and background samples to avoid introducing any kinematic bias in the muon classification. Thus, the MVA can learn from these variables but will not use them to separate signal from background.

The complete list of the 12 input variables is the following:

- The p_T and η of the muon (weighted).
- A flag checking whether the muon passes the global-muon reconstruction criteria or not, as the preselected muons can be also tracker-muons.
- Number of hits in the pixel detector used to fit the muon track.
- Number of tracker layers used in the muon track fit.
- Fraction of tracker hits used for the fit of the inner track.
- Number of muon stations with hits included in the muon track fit.
- Number of matched stations: number of segments (one per station) reconstructed in the muon chambers and used in the global muon track fit.
- Segment compatibility: the compatibility of track segments in the muon system with the pattern expected for a minimum ionizing particle.
- Normalized χ^2 of the muon track fit.
- Local χ^2: χ^2 of the position matching between the inner and standalone tracks. If there is no standalone track, it is set to 0.

- χ^2 of the kink-finder algorithm on the inner track defined in Reference [69]. The algorithm splits the inner track in two parts in several places along the trajectory and compares them. A large χ^2 indicates that the two parts are not compatible with being a single track.

The signal and background normalized distribution of the input variables, except p_T, η and the global muon flag, are shown in Figure 6.2. Good discrimination is observed between signal and background in most of the variables, especially in the segment compatibility and the number of matched stations where signal muons present higher values than background muons: signal muons have more stations with segments used in the muon track fit and the compatibility of those segments is better.

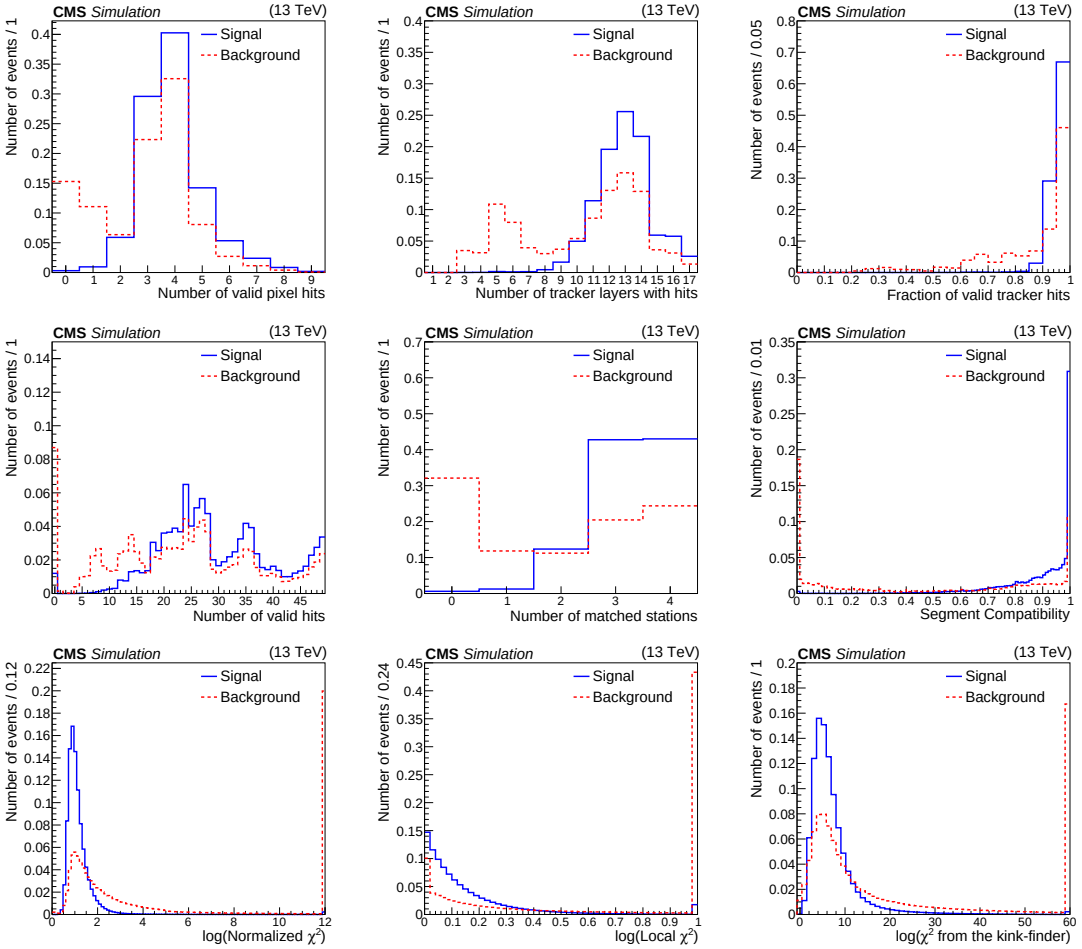

FIGURE 6.2: Normalized distributions in the $t\bar{t}$ training sample of the input variables for signal muons in blue and for background muons in red. For the local, normalized, and kink-finder χ^2 the logarithm of the variables is presented for better visualization [6].

6.2.3 Training

Several ML algorithms were tested as, for instance, BDTs, K-nearest neighbors, neural networks, but finally the best option was to train a random forest. As already mentioned before, random forest is a model that works very well in class-imbalance cases. The training is carried out with SCIKIT-LEARN package. The $t\bar{t}$ training sample is divided in two parts: 60% for training and 40% for testing. The main challenge of this classification problem is the class-imbalance, there are many more signal muons than background muons, therefore the "balanced" mode is used to automatically adjust weights inversely proportional to class frequencies in the input data and, in this way, try to avoid the overfitting.

A random forest has several hyperparameters which need to be optimized in order to achieve an accurate model. The most important ones are the number of trees and the maximum depth of those trees. The greater the depth of the trees, the greater the complexity of the model and the greater the risk of overfitting. For this parameter, maximum depths between 2 and 50 were tested with different values of the number of trees, between 100 and 5000, and the selected option was 200 trees with a maximum depth of 8. The memory usage was also taken into account when looking for the optimal model because this model is in the CMS central software so that the whole CMS Collaboration can use it.

The importance of the training variables can be extracted from the model to understand how the classification is carried out. The importance of a feature is computed as the normalized total reduction of the criterion brought by that feature. The importance (or mean decrease in impurity) of a feature is calculated as the sum over the number of splits (across all tress) that include the feature, proportionally to the number of samples it splits. The most important input variables in the muon MVA ID are the local and normalized χ^2, followed by the segment compatibility and the number of matched stations. On the other hand, the variables with the least importance were p_T and η, so they have not influence in the classification, as expected.

The ROC curve of the model for the train and test datasets are shown in Figure 6.3 and the output of the MVA, referred to as MVA score, is shown in Figure 6.4. It represents the probability of one muon to be classified as signal muon, in blue for signal muons and in red for background muons, so signal muons are expected to have values close to 1 and background muons close to 0. Furthermore, these figures not only prove that the random forest is a good discriminant, they also prove that there is no overfitting since the distributions of the train and test dataset are equal.

The MVA score of the $t\bar{t}$ sample with PU matching for muons passing the preselection is shown on the top left part of Figure 6.5. The shape is similar to that obtained with the training sample, which demonstrates that the training is robust. A study to understand the peaks that appear on the tail of the distribution was done and it

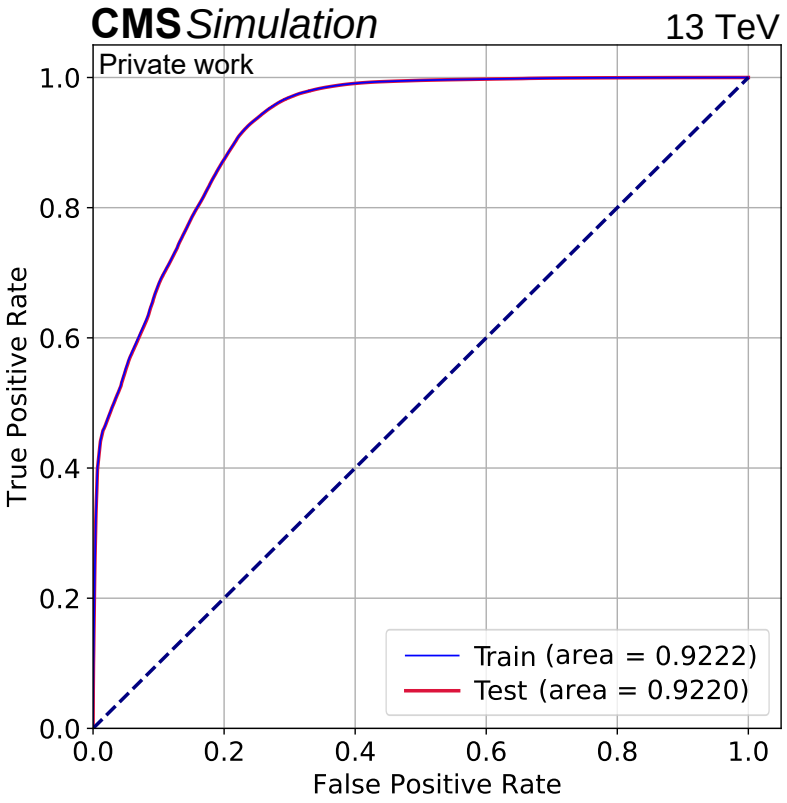

FIGURE 6.3: ROC curve of the MVA for the train dataset in blue and for the test dataset in red.

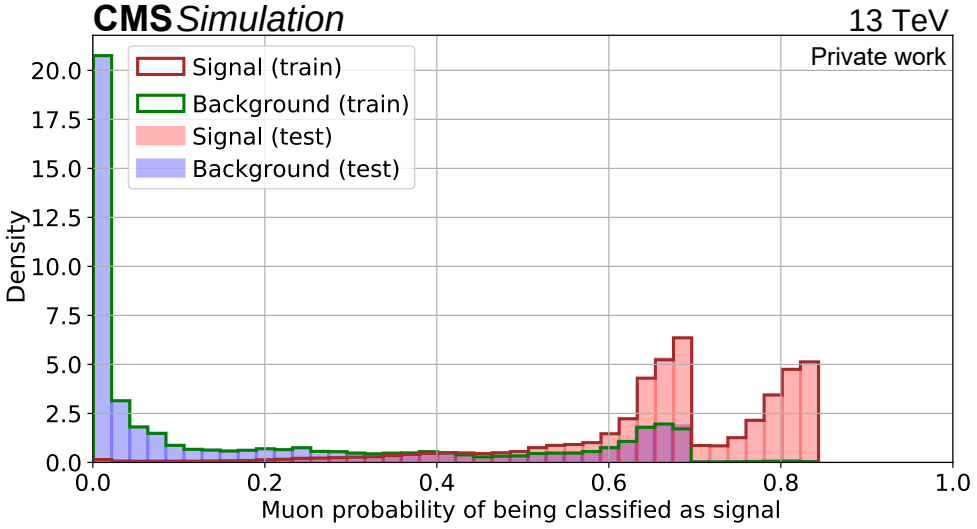

FIGURE 6.4: Normalized MVA score for signal muons in red and for background muons in blue.

was found that the peak in the middle of the distribution corresponds to muons with lower-p_T. This makes sense since low-p_T muons do not have the same characteristics as medium- or high-p_T muons and the model is able to learn this. Figure 6.5 (right) shows the MVA score of the preselected muons but moving the p_T requirement from $> 10\,\text{GeV}$ to $> 20\,\text{GeV}$, and the peak in the middle of the distribution becomes smaller. Then, plots on the bottom of the figure show the MVA score distribution for muons with $p_T > 25\,\text{GeV}$ (left) and $> 30\,\text{GeV}$ (right), where the central peak disappears. This means that for the muon p_T range used in the majority of analyses, which usually require $p_T > 20\,\text{GeV}$, the discrimination between signal and background muons is even better.

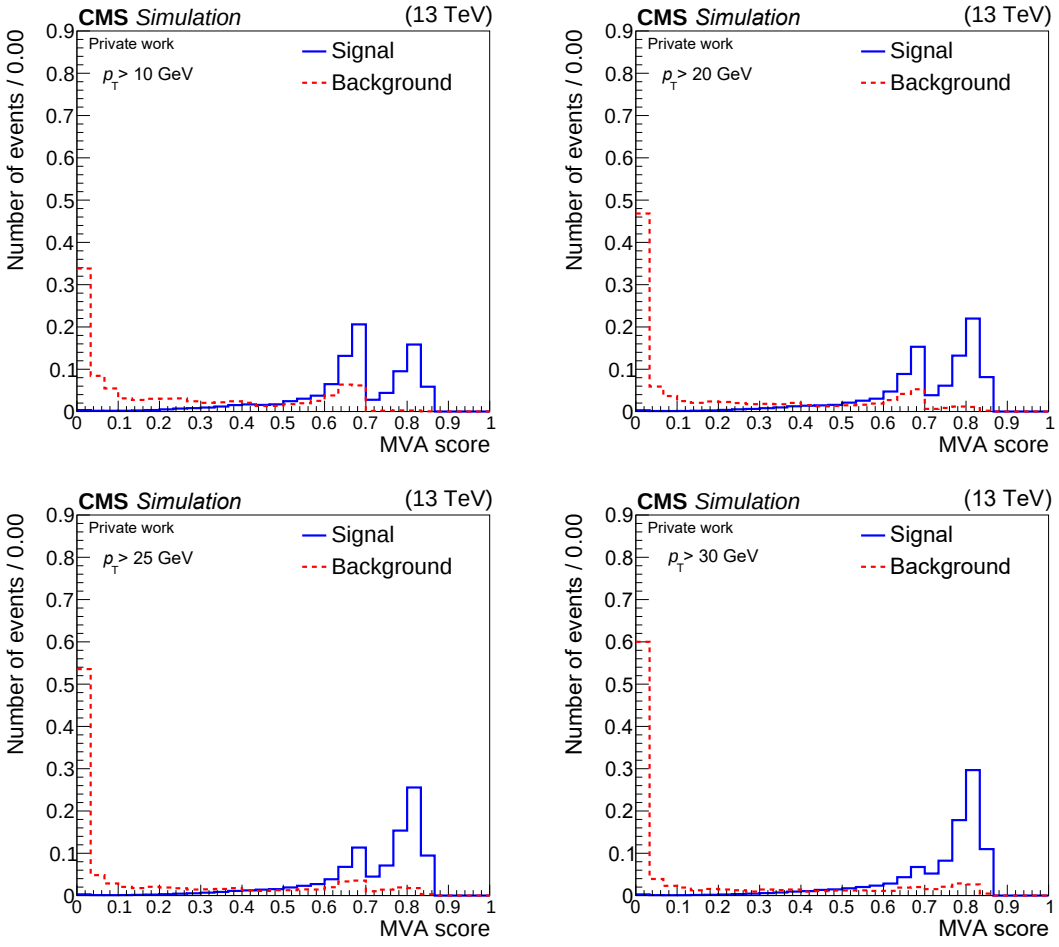

FIGURE 6.5: Normalized MVA score for signal muons in blue and for background muons in red with the $t\bar{t}$ sample with PU matching. For preselected muons with $p_T > 10\,\text{GeV}$, (top left), $p_T > 20\,\text{GeV}$ (top right), $p_T > 35\,\text{GeV}$, (bottom left), and $p_T > 30\,\text{GeV}$ (bottom right).

6.3 Results

The trained muon MVA ID discriminator provides a continuous variable that gives flexibility in choosing the best trade-off between signal and background efficiencies for each analisys. In addition, medium and tight WPs are defined to offer a simple way to check the MVA performance. The medium WP is defined as the value in the MVA score that yields the same background contamination as the medium WP of the cut-based ID, calculated for muons with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ (typical requirement of the analyses) using the $\mathrm{t\bar{t}}$ sample with PU matching. The background contamination, calculated as the number of background muons passing the identification criteria divided by the total number of background muons, is around 60%. For the tight WP, a higher threshold in the MVA score is selected in order to further reduce the background contamination by 10% with respect to the medium WP. The selected thresholds are MVA score > 0.08 and 0.20 for the medium and tight WPs, respectively.

Figure 6.6 shows the ROC curve of the muon MVA ID and its defined WPs using the $\mathrm{t\bar{t}}$ sample with PU matching for muons passing the preselection. The medium and tight WPs of the cut-based ID and the ROC curve of the soft MVA ID are also shown for reference.

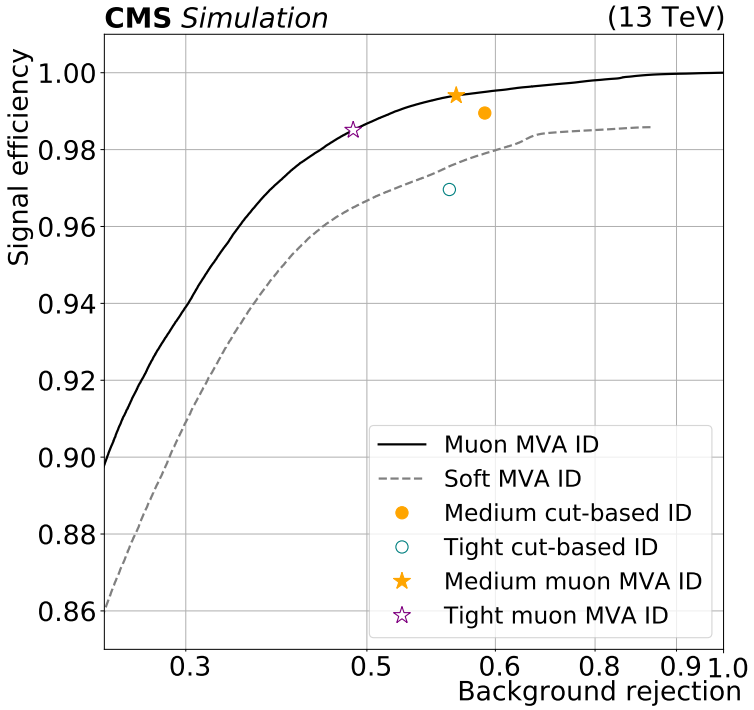

FIGURE 6.6: ROC curve for the muon MVA ID in black with the selected medium and tight WPs shown as orange and purple starts, respectively. Orange and blue points show the medium and tight WPs of the cut-based ID. The ROC curve of the soft MVA ID is shown in grey [6].

The performance of the muon MVA ID is much better than that of the soft MVA ID, which is expected since the latter was specifically designed for low-p_{T} muons. The medium WPs are similar, being the efficiency of the muon MVA ID slightly higher, although both were already very high, around 99%. Then, the tight muon MVA ID reduces the background muon contribution by 10% while keeping the efficiency still very high, over 98%. The tight WPs can not be fairly compared, since the tight WP of the cut-based ID includes impact paramenter requirements and those variables are not used in the MVA training.

In order to check the efficiency as a function of p_{T}, η, and the number of PU vertices in data and DY simulation, the tag-and-probe method described in Section 3.3 is used. Tag muons are required to pass the cut-based tight WP and to have fired a single-muon trigger, to avoid bias in the efficiency measurement, and probe muons are required to pass the loose cut-based ID criteria. Only statistical uncertainties are considered, as systematic uncertainties will depend on each analysis.

The measured efficiencies for each muon MVA ID WPs as a function of p_{T} are shown in Figure 6.7 in two η regions. For the medium WP, the cut-based ID efficiency is also included for comparison. Even though both efficiencies are very high, around 99.5%, the one achieved by the muon MVA ID is systematically 0.5% larger. For the tight WP, the efficiency is about 99% for muons with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and 1–2% smaller for muons with p_{T} between 10 and 30 GeV. This efficiency is just 1% smaller than that of the medium WP while its background rejection is 10% better, which makes it very useful for analyses with background muons contamination. The efficiencies of the muon MVA ID are similar in both η regions.

Figure 6.8 shows the efficiency as a function of the number of vertices. It is kept over 99.5% (98.0%) for the medium (tight) muon MVA ID WP, even at PU conditions with up to 60 PU vertices. The muon MVA ID is shown to be more robust as a function of PU than the cut-based ID, whose efficiency decreases. This may be useful in the context of changing data-taking conditions or high PU scenarios expected in Run 3.

In general, the efficiency in DY simulation is systematically higher than the efficiency in data in both approaches as a result of small imperfections in the modelling. The discrepancy observed is around 0.5–1.0% for the medium cut-based ID criteria, while for the medium MVA ID it is reduced due to the correlation between variables learned by the model.

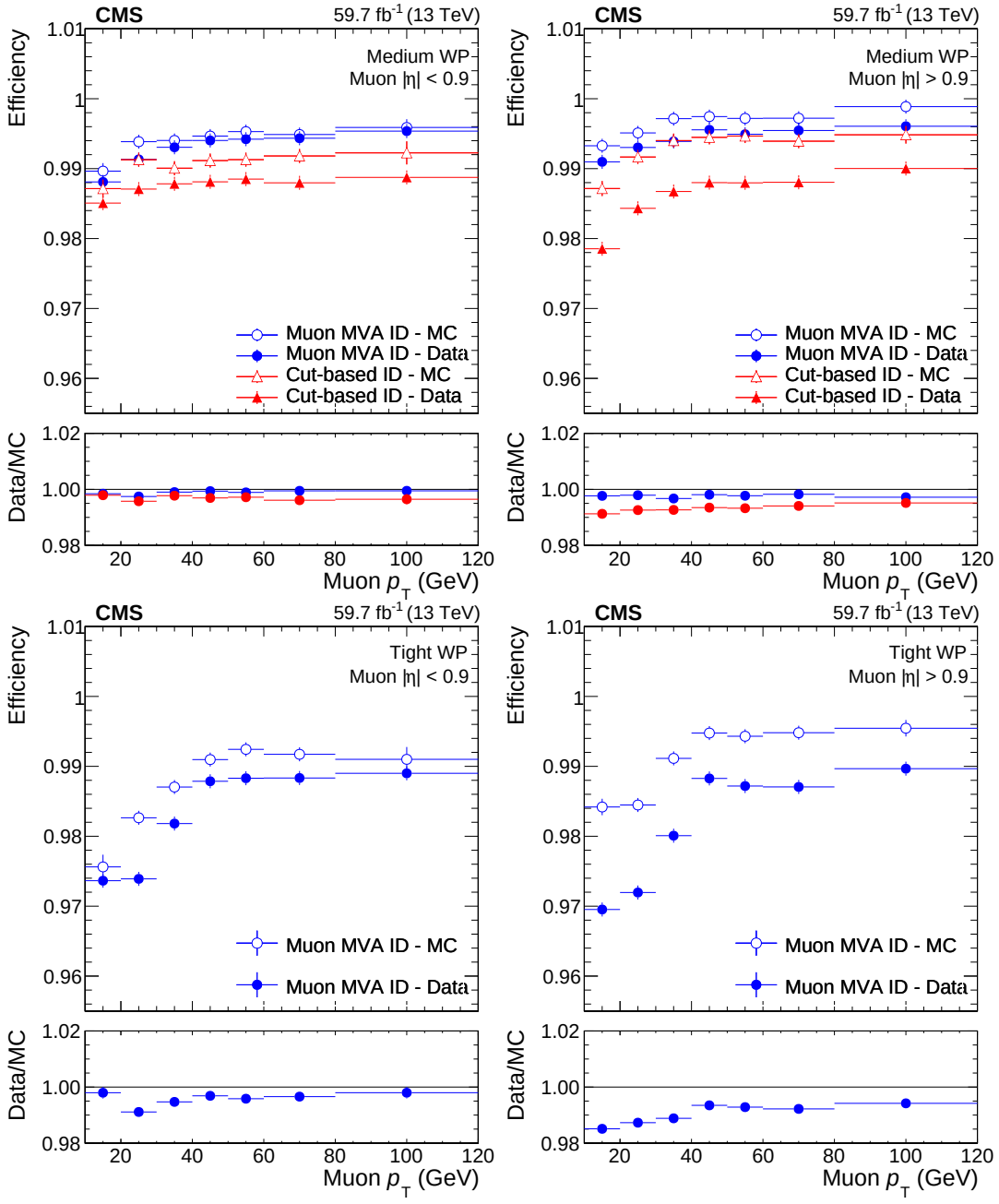

FIGURE 6.7: Efficiency for the medium (top) and tight (bottom) WPs as a function of p_T for muons with $|\eta| < 0.9$ (left) and $|\eta| > 0.9$ (right). Blue dots show the muon MVA ID performance both for the 2018 dataset and DY simulation, whereas red triangles show the efficiency of the medium cut-based ID. The ratio of efficiencies in data and MC is also shown [6].

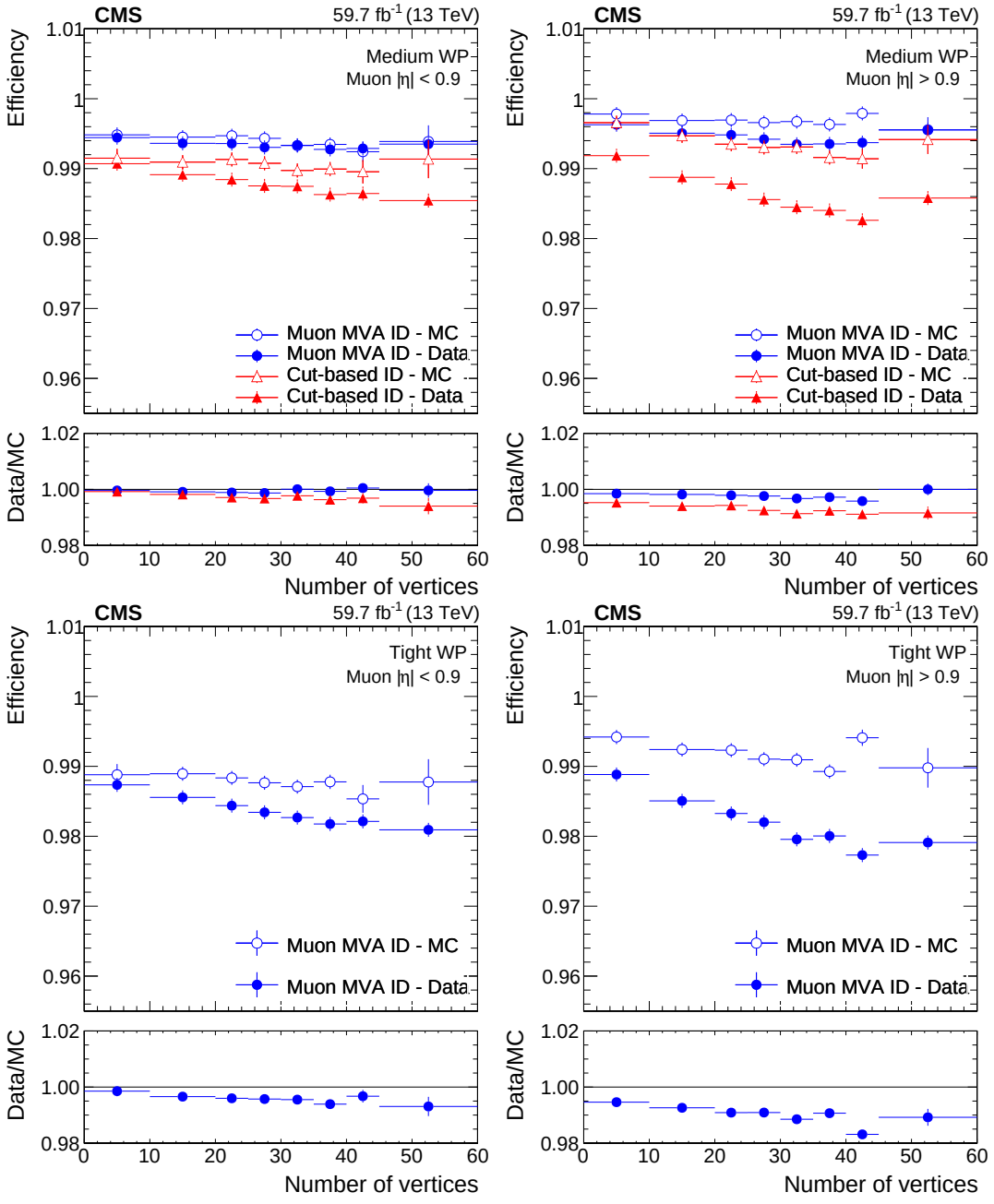

FIGURE 6.8: Efficiency for the medium (top) and tight (bottom) WPs as a function of number of vertices for muons with $|\eta| < 0.9$ (left) and $|\eta| > 0.9$ (right). Blue dots show the muon MVA ID performance both for the 2018 dataset and DY simulation, whereas red triangles show the efficiency of the medium cut-based ID. The ratio of efficiencies in data and MC is also shown [6].

Summary and conclusions

In this thesis, results from both precision measurements of the SM and searches for new physics BSM, which are the two main scientific objectives for which the LHC was built, are presented. All the analyses included in this thesis use data collected by the CMS detector in pp collisions produced at the LHC.

Precision measurements of the SM have been performed using data from pp collisions at two different centre-of-mass energies: 13.6 TeV and 5.02 TeV. The first measurement of the cross section of the $t\bar{t}$ process in the dilepton final state at 13.6 TeV was carried out using a small amount of data, 1.2 fb^{-1}, collected at the beginning of the Run 3. This result is, by the time this thesis is finished, the only one published by the CMS Collaboration at the highest, and thus unexplored, centre-of-mass energy ever achieved by the LHC. This measurement is also a way to validate the good functioning of the CMS detector, as it exhibits good agreement with the SM prediction ($\sigma_{t\bar{t}} = 924^{+32}_{-40}$, pb). The final value obtained for the $t\bar{t}$ cross section has an uncertainty comparable to that of the theoretical predictions, as shown in Figure 4.12:

$$\sigma_{t\bar{t}} = 888 \pm 34\,(\text{syst.} + \text{stat.}) \pm 20\,(\text{lumi.})\,\text{pb}.$$

Currently, the analysis methodology to measure the $t\bar{t}$ cross section in the semilepton final state has been also developed using the dataset collected in 2017 at a centre-of-mass energy of 5.02 TeV, with an integrated luminosity of 302 pb^{-1}. This dataset is of particular interest because its number of interactions from PU is an order of magnitude smaller than the Run 2 dataset at 13 TeV, enabling the possibility to use lower momentum particles. The measurement of the $t\bar{t}$ cross section in the dilepton final state resulted in $\sigma_{t\bar{t}} = 60.7 \pm 5.0\,(\text{stat.}) \pm 2.8\,(\text{syst.}) \pm 1.1\,(\text{lumi.})$ pb. The plan is to combine this result with the measurement in the semilepton decay channel, reducing the statistical uncertainty and improving the precision of the measurement. The total uncertainty is expected to be reduced to half of the previous result.

The 5.02 TeV dataset is also used to measure, for the first time at this centre-of-mass energy, the production cross section of massive boson pairs, including WW, WZ, and

ZZ. All measurements are carried out in final states with multiple leptons, and the measured cross sections are the following:

$$\sigma_{W^+W^-} = 37.0^{+5.5}_{-5.2}\,(\text{stat.})^{+2.7}_{-2.6}\,(\text{syst.})\text{pb},$$
$$\sigma_{W^\pm Z} = 6.4^{+2.5}_{-2.1}\,(\text{stat.})^{+0.5}_{-0.3}\,(\text{syst.})\text{pb},$$
$$\sigma_{ZZ} = 5.3^{+2.5}_{-2.1}\,(\text{stat.})^{+0.5}_{-0.4}\,(\text{syst.})\text{pb}.$$

These results are in agreement with the SM predictions, as shown in Figure 4.29.

The second part of the thesis encompasses two searches for physics processes not observed so far, both related to top quarks and using the full Run 2 dataset at 13 TeV. The first one is a search for the pair production of the supersymmetric partner of the top quark, the stop quark. The target signal process involves a stop-antistop quark pair decaying into a top-antitop quark pair and neutralinos, with a mass difference between the stop quark and the neutralino very close to the top quark mass. This similarity makes it challenging to distinguish the signal from $\mathrm{t\bar{t}}$, and most of the existing searches have not been sensitive to the presence of this process. The analysis presented in this thesis is a dedicated search in this region using a DNN to exploit small differences between the signal and the $\mathrm{t\bar{t}}$ background in various observables, achieving a sensitivity that would otherwise not be possible. With this strategy, the entire search region was excluded at a 95% CL, as shown in Figure 5.19.

This thesis also presents a search within the framework of the EFT in top quark production in association with leptons. It simultaneously studies the effect of 26 WCs in 6 signal processes: $\mathrm{t\bar{t}}H$, $\mathrm{t\bar{t}}\ell\nu$, $\mathrm{t\bar{t}}\ell\bar{\ell}$, $t\ell\bar{\ell}q$, tHq, and $\mathrm{t\bar{t}t\bar{t}}$. By defining dedicated selections and utilizing variables that offer the highest sensitivity to the signal, CIs are extracted for each of the WCs through likelihood fits, with the option of either leaving the other coefficients free or fixing them to their SM values of zero. In all cases, the data are found to be consistent with the SM prediction, and improvements are achieved in the CIs compared to those available up to date.

Lastly, a new technique for muon identification has been developed using a machine learning model, which, as shown in Figure 6.6, provides better results than techniques based on various set of requirements on individual variables used to date. Furthermore, this model performs equally well under high PU conditions, demonstrating its robustness and suitability for usage in various PU conditions, including Run 3, where the number of PU interactions is higher than in Run 2.

In summary, the results presented in this thesis have contributed to the advance of knowledge in the field of particle physics, using a variety of innovative data analysis techniques, including machine learning.

Resumen y conclusiones

En esta tesis se presentan resultados tanto de medidas de precisión del Modelo Estándar (ME) como de búsquedas de nueva física más allá de sus predicciones, que son los dos objetivos científicos principales para los cuales el LHC fue construido. Todos los análisis recogidos en esta tesis se llevan a cabo haciendo uso de datos tomados por el detector CMS en colisiones de protones producidas en el LHC.

Las medidas de precisión del ME se han realizado usando datos de colisiones de protones a dos energías en centro de masas diferentes: 13.6 TeV y 5.02 TeV. La primera medida de la sección eficaz del proceso $t\bar{t}$ en el estado final con dos leptones a 13.6 TeV se realizó usando una pequeña cantidad de datos, $1.2\,\text{fb}^{-1}$, recogidos al comienzo del Run 3. Este resultado es, en el momento que se finalizó la tesis, el único publicado por CMS a la más alta, y por tanto inexplorada, energía en centro de masas alcanzada por el LHC. Esta medida sirve también para validad el buen funcionamiento del detector CMS, ya que presenta un buen acuerdo con la predicción del ME ($\sigma_{t\bar{t}} = 924^{+32}_{-40}\,\text{pb}$). El valor final obtenido de la sección eficaz de $t\bar{t}$ tiene una incertidumbre comparable a la de la predicción teórica, como se puede ver en la Figura 4.12:

$$\sigma_{t\bar{t}} = 888 \pm 34\,(\text{syst.} + \text{stat.}) \pm 20\,(\text{lumi.})\,\text{pb}.$$

En la actualidad, se ha desarrollado la metodología del análisis para medir la sección eficaz de $t\bar{t}$ en el estado final con un leptón con el conjunto de datos tomado en 2017 a una energía en centro de masas de 5.02 TeV, utilizando una luminosidad integrada de $302\,\text{pb}^{-1}$. Este conjunto de datos es interesante porque su número de interacciones provenientes de PU es un orden de magnitud menor que el que tiene el conjunto de datos del Run 2 a 13 TeV y permite estudiar rangos del momento de las partículas más bajos. La medida de la sección eficaz de $t\bar{t}$ en el estado final con dos leptones obtuvo una $\sigma_{t\bar{t}} = 60.7 \pm 5.0\,(\text{stat.}) \pm 2.8\,(\text{syst.}) \pm 1.1\,(\text{lumi.})\,\text{pb}$. El plan es combinar este resultado con el obtenido por la medida en el canal de desintegración con un leptón y así reducir la incertidumbre estadística y mejorar la precisón de la medida. Se espera reducir la incertidumbre total a la mitad respecto al resultado anterior.

El conjunto de datos de 5.02 TeV también se utilizó para medir, por primera vez a esta energía en centro de masas, la sección eficaz de producción de un par de bosones masivos, incluyendo WW, WZ y ZZ. Todas las medias se llevaron a cabo en los estados finales con varios leptones, y las secciones eficaces medidas son las siguientes:

$$\sigma_{W^+W^-} = 37.0^{+5.5}_{-5.2}\,(\text{stat.})^{+2.7}_{-2.6}\,(\text{syst.})\text{pb},$$
$$\sigma_{W^\pm Z} = 6.4^{+2.5}_{-2.1}\,(\text{stat.})^{+0.5}_{-0.3}\,(\text{syst.})\text{pb},$$
$$\sigma_{ZZ} = 5.3^{+2.5}_{-2.1}\,(\text{stat.})^{+0.5}_{-0.4}\,(\text{syst.})\text{pb}.$$

Estos resultados están en concordancia con los predichos por el ME, como se puede ver en la Figura 4.29.

La segunda parte de la tesis engloba dos búsquedas de procesos de nueva física más allá del ME todavía no observados, ambas relacionadas con quarks top y utilizando el conjunto de datos completo del Run 2 a 13 TeV. La primera es una búsqueda de la producción en pares de la partícula supersimétrica del quark top, el stop. El proceso de señal que se busca es dos quarks stop-antistop que se desintegran en un par top-antitop y neutralinos con una diferencia de masas entre el stop y el neutralino muy similar a la del quark top, lo que hace que separar la señal de $t\bar{t}$ sea muy complicado y la mayoría de las búsquedas existentes hasta la fecha no lograban ser sensibles a la presencia de este proceso. El análisis presentado en esta tesis es una búsqueda dedicada en esta región haciendo uso de una red neuronal paramétrica para explotar las pequeñas diferencias en diferentes variables que hay entre señal y $t\bar{t}$ y conseguir tener sensibilidad, que de otro modo no sería posible. Con esta estrategia, se excluyó la región de búsqueda entera a un nivel de confianza del 95%, tal y como se muestra en la Figura 5.19.

En esta tesis también se presenta una búsqueda bajo el marco de la teoría efectiva de campos en la producción de quarks top en asociación con leptones. En ella, se estudian simultaneamente el efecto de 26 WCs en 6 procesos de señal: $t\bar{t}H$, $t\bar{t}\ell\nu$, $t\bar{t}\ell\bar{\ell}$, $t\ell\bar{\ell}q$, tHq and $t\bar{t}t\bar{t}$. Haciendo uso de la definición de diferentes selecciones dedicadas y de las variables que ofrecían mayor sensibilidad a la señal, los intervalos de confianza fueron extraídos para cada uno de los WCs haciendo un ajuste de verosimilitud, tanto dejando libres el resto de coeficientes como ajustándoles a su valor del ME. En todos los casos, se encontró que los datos eran consistentes con la predicción del ME y se consiguieron mejorar los intervalos de confianza que había hasta la fecha.

Por último, también se ha desarrollado una nueva técnica para la identificación de muones utilizando un modelo de aprendizaje automático que, como se muestra en la Figura 6.6, ofrece mejores resultados que las técnicas basadas en diferentes requisitos en variables individuales utilizadas hasta el momento. Además, este modelo funciona igual de bien a alto PU, lo que significa que es robusto y podrá usarse en diferentes

condiciones de PU, incluyendo el Run 3, donde el número de interacciones de PU es mayor que en el Run 2.

En resumen, con los resultados presentados en esta tesis se ha contribuido al avance del conocimiento en el campo de física de partículas uitilzando una variedad de técnicas de análisis de datos novedosas, entre las que se encuentra el aprendizaje automático.

Bibliography

[1] CMS Collaboration. "First measurement of the top quark pair production cross section in proton-proton collisions at $\sqrt{s} = 13.6\,\text{TeV}$". *JHEP* 08 (2023), p. 204. DOI: 10.1007/JHEP08(2023)204. arXiv: 2303.10680.

[2] CMS Collaboration. "Measurement of the inclusive $\mathrm{t\bar{t}}$ production cross section in proton-proton collisions at $\sqrt{s} = 5.02\,\text{TeV}$". *JHEP* 04 (2022), p. 144. DOI: 10.1007/JHEP04(2022)144. arXiv: 2112.09114.

[3] CMS Collaboration. "Measurements of the weak diboson production cross sections in leptonic decays at $\sqrt{s} = 5.02\,\text{TeV}$ with the CMS experiment". *Phys. Rev. Lett* 127 (2021), p. 191801. DOI: 10.1103/PhysRevLett.127.191801. arXiv: 2107.01137.

[4] CMS Collaboration. "Combined searches for the production of supersymmetric top quark partners in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$". *EPJC* 81 (2021), p. 970. DOI: 10.1140/epjc/s10052-021-09721-5. arXiv: 2107.10892.

[5] CMS Collaboration. "Search for physics beyond the standard model in top quark production with additional leptons in the context of effective field theory". *Submitted to JHEP* (2023). arXiv: 2307.15761.

[6] CMS Collaboration. "Muon identification using multivariate techniques in the CMS experiment in proton-proton collisions at $\sqrt{s} = 13$ TeV". *Submitted to JINST* (2023). arXiv: 2310.03844.

[7] C. Patrignani et al. "Review of Particle Physics". *Chin. Phys.* C40.10 (2016). DOI: 10.1088/1674-1137/40/10/100001.

[8] M. E. Peskin and D. V. Schroeder. "An Introduction to quantum field theory" (1995). URL: http://www.slac.stanford.edu/~mpeskin/QFT.html.

[9] S.L. Glashow. "Partial-symmetries of weak interactions". *Nuclear Physics* 22.4 (1961), pp. 579–588. ISSN: 0029-5582. DOI: https://doi.org/10.1016/0029-5582(61)90469-2.

[10] S. Weinberg. "A Model of Leptons". *Phys. Rev. Lett.* 19 (1967). DOI: 10.1103/PhysRevLett.19.1264.

[11] P. W. Higgs. "Spontaneous Symmetry Breakdown without Massless Bosons". *Phys. Rev.* 145 (4 1966), pp. 1156–1163. DOI: 10.1103/PhysRev.145.1156.

[12] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". *Phys. Rev. Lett.* 13 (1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321.

[13] H. Fritzsch, M. Gell-Mann, and H. Leutwyler. "Advantages of the color octet gluon picture". *Physics Letters B* 47 (1973), p. 4. DOI: https://doi.org/10.1016/0370-2693(73)90625-4.

[14] P. Ramond. "Dual theory for free fermions". *Phys. Rev. D* 3 (1971). DOI: 10.1103/PhysRevD.3.2415.

[15] Yu. A. Golfand and E. P. Likhtman. "Extension of the algebra of Poincaré group generators and violation of P invariance". *JETP Lett.* 13 (1971). URL: http://www.jetpletters.ac.ru/ps/1584/article_24309.pdf.

[16] A. Neveu and J. H. Schwarz. "Factorizable dual model of pions". *Nucl. Phys. B* 31 (1971). DOI: 10.1016/0550-3213(71)90448-2.

[17] D. V. Volkov and V. P. Akulov. "Possible universal neutrino interaction". *JETP Lett.* 16 (1972). URL: http://www.jetpletters.ac.ru/ps/1766/article_26864.pdf.

[18] J. Wess and B. Zumino. "A Lagrangian model invariant under supergauge transformations". *Phys. Lett. B* 49 (1974). DOI: 10.1016/0370-2693(74)90578-4.

[19] J. Wess and B. Zumino. "Supergauge transformations in four-dimensions". *Nucl. Phys. B* 70 (1974). DOI: 10.1016/0550-3213(74)90355-1.

[20] P. Fayet. "Supergauge invariant extension of the Higgs mechanism and a model for the electron and its neutrino". *Nucl. Phys. B* 90 (1975). DOI: 10.1016/0550-3213(75)90636-7.

[21] H. P. Nilles. "Supersymmetry, supergravity and particle physics". *Phys. Rept.* 110 (1984). DOI: 10.1016/0370-1573(84)90008-5.

[22] A. Pich. "Effective Field Theory". *Les Houches Summer School in Theoretical Physics, Session 68: Probing the Standard Model of Particle Interactions* (1998).

[23] W. Buchmuller and D. Wyler. "Effective Lagrangian analysis of new interactions and flavor conservation". *Nucl. Phys. B* 268 (1986), p. 621. DOI: 10.1016/0550-3213(86)90262-2.

[24] B. Grzadkowski et al. "Dimension-six terms in the standard model Lagrangian". *JHEP* 10 (2010), p. 085. DOI: 10.1007/JHEP10(2010)085. arXiv: 1008.4884.

[25] A. Falkowski and R. Rattazzi. "Which EFT". *JHEP* 10 (2019), p. 255. DOI: 10.1007/JHEP10(2019)255. arXiv: 1902.05936.

[26] L. Evans and P. Bryant. "LHC Machine". *JINST* 3 (2008), S08001. DOI: 10.1088/1748-0221/3/08/S08001.

[27] J. Haffner. *The CERN accelerator complex*. Available in URL: http://cds.cern.ch/record/2636343. 2018.

[28] *CMS Luminosity Public Results*. Available in URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults. 2023.

[29] CMS Collaboration. "The CMS Experiment at the CERN LHC". *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[30] T. Sakuma. *Cutaway diagrams of CMS detector*. Available in URL: "https://cds.cern.ch/record/2665537. 2019.

[31] CMS Collaboration. "The CMS tracker system project: Technical Design Report". Technical Design Report CMS (1997). URL: https://cds.cern.ch/record/368412.

[32] CMS Collaboration. "The CMS electromagnetic calorimeter project: Technical Design Report". Technical Design Report CMS (1997). URL: https://cds.cern.ch/record/349375.

[33] CMS Collaboration. "The CMS hadron calorimeter project: Technical Design Report". Technical Design Report CMS (1997). URL: https://cds.cern.ch/record/357153.

[34] CMS Collaboration. "The CMS magnet project: Technical Design Report". Technical Design Report CMS (1997). URL: https://cds.cern.ch/record/331056.

[35] CMS Collaboration. "The CMS muon project: Technical Design Report". Technical Design Report CMS (1997). URL: https://cds.cern.ch/record/343814.

[36] CMS Collaboration. "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$". *JINST* 13 (2018), P06015. DOI: 10.1088/1748-0221/13/06/P06015. arXiv: 1804.04528.

[37] CMS Collaboration. "Particle-flow reconstruction and global event description with the CMS detector". *JINST* 12 (2017), P10003. DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965.

[38] R. Fruhwirth. "Application of Kalman filtering to track and vertex fitting". *Nucl. Instrum. Meth.* A262 (1987). DOI: 10.1016/0168-9002(87)90887-4.

[39] CMS Collaboration. *Electron reconstruction in CMS*. Tech. rep. CMS-NOTE-2006-040. CERN, 2006. URL: http://cds.cern.ch/record/934070.

[40] M. Cacciari, G. P. Salam, and G. Soyez. "The anti-k_T jet clustering algorithm". *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189.

[41] CMS Collaboration. "Pileup mitigation at CMS in 13 TeV data". *JINST* 15 (2020), P09018. DOI: 10.1088/1748-0221/15/09/P09018. arXiv: 2003.00503.

[42] D. Bertolini et al. "Pileup per particle identification". *JHEP* 10 (2014), p. 059. DOI: 10.1007/JHEP10(2014)059. arXiv: 1407.6013.

[43] CMS Collaboration. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". *JINST* 13 (2018), P05011. DOI: 10.1088/1748-0221/13/05/P05011. arXiv: 1712.07158.

[44] E. Bols et al. "Jet flavour classification using DeepJet". *JINST* 15 (2020), P12012. DOI: 10.1088/1748-0221/15/12/P12012. arXiv: 2008.10519.

[45] H1 and Zeus Collaborations. "Combined Measurement and QCD analysis of the inclusive $e^{\pm}p$ scattering cross sections at HERA". *JHEP* 01 (2010). DOI: 10.1007/JHEP01(2010)109. arXiv: 0911.0884.

[46] 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.

[47] S. Frixione, P. Nason, and C. 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.

[48] T. Sjöstrand et al. "An introduction to PYTHIA 8.2". *Comput. Phys. Commun.* 191 (2015), p. 159. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012.

[49] S. Agostinelli et al. "Geant4—a simulation toolkit". *Nucl. Instrum. Meth. A* 506 (2003). DOI: 10.1016/S0168-9002(03)01368-8.

[50] CMS Collaboration. "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8\,\text{TeV}$". *JINST* 10 (2015), P06005. DOI: 10.1088/1748-0221/10/06/P06005. arXiv: 1502.02701.

[51] CMS Collaboration. "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV". *JINST* 12 (2017), P02014. DOI: 10.1088/1748-0221/12/02/P02014. arXiv: 1607.03663.

[52] D0 Collaboration. "Observation of the top quark". *Phys. Rev. Lett.* 74 (1995). DOI: 10.1103/PhysRevLett.74.2632. arXiv: 9503003.

[53] CDF Collaboration. "Observation of top quark production in $\bar{p}p$ collisions". *Phys. Rev. Lett.* 74 (1995), p. 2626. DOI: 10.1103/PhysRevLett.74.2626.

[54] Particle Data Group, R. L. Workman, et al. "Review of particle physics". *Prog. Theor. Exp. Phys.* 2022 (2022), p. 083C01. DOI: 10.1093/ptep/ptac097.

[55] CMS Collaboration. "Observation of four top quark production in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ ". Submitted to *Phys. Lett. B*. 2023. arXiv: 2305.13439.

[56] *CMS Standard Model summary plots*. Available in URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsCombined. 2023.

[57] *LHC Top Working Group*. Available in URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCTopWGSummaryPlots. 2023.

[58] CMS Collaboration. "Measurement of the top quark pair production cross section in proton-proton collisions at $\sqrt{s} = 13$ TeV". *Phys. Rev. Lett.* 116 (2016). DOI: 10.1103/PhysRevLett.116.052002. arXiv: 1510.05302.

[59] ATLAS Collaboration. "Measurement of the $t\bar{t}$ production cross-section in the lepton+jets channel at $\sqrt{s} = 13$ TeV, with the ATLAS experiment". *Phys. Lett. B* 810 (2020). DOI: 10.1016/j.physletb.2020.135797. arXiv: 2006.13076.

[60] CMS Collaboration. "Measurement of differential $t\bar{t}$ production cross sections in the full kinematic range using lepton+jets events from proton-proton collisions at $\sqrt{s} = 13$ TeV". *Phys. Rev. D* 104 (2021). DOI: 10.1103/physrevd.104.092013. arXiv: 2108.02803.

[61] ATLAS Collaboration. "Inclusive and differential cross-sections for dilepton $t\bar{t}$ production measured in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector". *JHEP* 07 (2023), p. 141. DOI: 10.1007/jhep07(2023)141. arXiv: 2303.15340.

[62] ATLAS Collaboration. *Measurement of $t\bar{t}$ and Z-boson cross sections and their ratio using pp collisions at $\sqrt{s} = 13.6$ TeV with the ATLAS detector*. Tech. rep. CERN, 2023. URL: https://cds.cern.ch/record/2854834.

[63] ATLAS Collaboration. "Measurement of the $t\bar{t}$ production cross-section in pp collisions at $\sqrt{s} = 5.02$ TeV TeV with the ATLAS detector". *JHEP* 06 (2023), p. 138. DOI: 10.1007/jhep06(2023)138. arXiv: 2207.01354.

[64] *ATLAS Standard Model summary plots*. Available in URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/SM/. 2023.

[65] CMS Collaboration. "Measurement of the pp $\rightarrow$ WZ inclusive and differential cross sections, polarization angles and search for anomalous gauge couplings at $\sqrt{s} = 13$ TeV". *JHEP* 07 (2022), p. 032. DOI: 10.1007/JHEP07(2022)032.

[66] CMS Collaboration. "Measurement of the top quark pair production cross section in proton-proton collisions at $\sqrt{s} = 13$ TeV". *Phys. Rev. Lett.* 116 (2016), p. 052002. DOI: 10.1103/PhysRevLett.116.052002. arXiv: 1510.05302.

[67] CMS Collaboration. "Measurement of the $t\bar{t}$ production cross section using events in the eμ final state in pp collisions at $\sqrt{s} = 13$ TeV". *Eur. Phys. J.* 77 (2017), p. 172. DOI: 10.1140/epjc/s10052-017-4718-8. arXiv: 1611.04040.

[68] CMS Collaboration. "Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC". *JINST* 16 (2021), P05014. DOI: 10.1088/1748-0221/16/05/P05014. arXiv: 2012.06888.

[69] CMS Collaboration. "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV". *JINST* 13 (2018), P06015. DOI: 10.1088/1748-0221/13/06/P06015. arXiv: 1804.04528.

[70] R. D. Ball et al. "Parton distributions from high-precision collider data". *Eur. Phys. J. C* 77 (2017), p. 663. DOI: 10.1140/epjc/s10052-017-5199-5. arXiv: 1706.00428.

[71] CMS Collaboration. "Extraction and validation of a new set of CMS PYTHIA 8 tunes from underlying-event measurements". *Eur. Phys. J. C* 80 (2020), p. 4. DOI: 10.1140/epjc/s10052-019-7499-4. arXiv: 1903.12179.

[72] CMS Collaboration. *Jet algorithms performance in 13 TeV data*. CMS Physics Analysis Summary CMS-PAS-JME-16-003. CERN, 2017. URL: http://cds.cern.ch/record/2256875.

[73] CMS Collaboration. "Observation of tW production in the single-lepton channel in pp collisions at $\sqrt{s} = 13$ TeV". *JHEP* 11 (2021), p. 111. DOI: 10.1007/JHEP11(2021)111. arXiv: 2109.01706.

[74] CMS Collaboration. "W^+W^- boson pair production in proton-proton collisions at $\sqrt{s} = 13$ TeV". *Phys. Rev. D* 102 (2020), p. 092001. DOI: 10.1103/PhysRevD.102.092001. arXiv: 2009.00119.

[75] CMS Collaboration. "Measurement of the production cross section of a W boson in association with two b jets in pp collisions at $\sqrt{s} = 8$ TeV ". *Eur. Phys. J. C* 77 (2017), p. 92. DOI: 10.1140/epjc/s10052-016-4573-z. arXiv: 1608.07561.

[76] CMS Collaboration. "Measurement of the inclusive and differential WZ production cross sections, polarization angles, and triple gauge couplings in pp collisions at $\sqrt{s} = 13$ TeV ". *JHEP* 07 (2022), p. 032. DOI: 10.1007/JHEP07(2022)032. arXiv: 2110.11231.

[77] CMS Collaboration. *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*. CMS Physics Analysis Summary CMS-PAS-TOP-16-021. CERN, 2016. URL: http://cds.cern.ch/record/2235192.

[78] Particle Data Group, P. A. Zyla, et al. "Review of Particle Physics". *Prog. Theore. Expt. Phys.* 2020 (2020), p. 083C01. ISSN: 2050-3911. DOI: 10.1093/ptep/ptaa104.

[79] M. Czakon and A. Mitov. "Top++: A program for the calculation of the top-pair cross-section at hadron colliders". *Comput. Phys. Commun.* 185 (2014). DOI: 10.1016/j.cpc.2014.06.021. arXiv: 1112.5675.

[80] D. B. Richard et al. "The PDF4LHC21 combination of global PDF fits for the LHC Run III". *J. Phys. G* 49 (2022), p. 080501. DOI: 10.1088/1361-6471/ac7216. arXiv: 2203.05506.

[81] CMS Collaboration. "Measurement of double-differential cross sections for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV and impact on parton distribution functions". *Eur. Phys. J. C* 77 (2017), p. 459. DOI: 10.1140/epjc/s10052-017-4984-5. arXiv: 1703.01630.

[82] CMS Collaboration. "Measurement of the inclusive $t\bar{t}$ cross section in pp collisions at $\sqrt{s} = 5.02$ TeV using final states with at least one charged lepton". *JHEP* 03 (2018), p. 115. DOI: 10.1007/JHEP03(2018)115. arXiv: 1711.03143.

[83] CMS Collaboration. "Measurement of the Higgs boson production rate in association with top quarks in final states with electrons, muons, and hadronically decaying tau leptons at $\sqrt{s} = 13$ TeV". *Eur. Phys. J. C* 81 (2021), p. 378. DOI: 10.1140/epjc/s10052-021-09014-x. arXiv: 2011.03652.

[84] CMS Collaboration. "Evidence for associated production of a Higgs boson with a top quark pair in final states with electrons, muons, and hadronically decaying leptons at $\sqrt{s} = 13$ TeV". *JHEP* 08 (2018), p. 066. DOI: 10.1007/JHEP08(2018)066. arXiv: 1803.05485.

[85] CMS Collaboration. "Observation of $t\bar{t}H$ production". *Phys. Rev. Lett.* 120 (2018), p. 231801. DOI: 10.1103/PhysRevLett.120.231801. arXiv: 1804.02610.

[86] CMS Collaboration. "Measurement of the inclusive and differential WZ production cross sections, polarization angles, and triple gauge couplings in pp collisions at $\sqrt{s} = 13$ TeV ". *JHEP* 07 (2022), p. 032. DOI: 10.1007/JHEP07(2022)032. arXiv: 2110.11231.

[87] CMS Collaboration. "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.

[88] CMS Collaboration. "Search for electroweak production of charginos and neutralinos in proton-proton collisions at $\sqrt{s} = 13$ TeV". *JHEP* 04 (2022), p. 147. DOI: 10.1007/JHEP04(2022)147. arXiv: 2106.14246.

[89] CMS Collaboration. "Measurement of the cross section of top quark-antiquark pair production in association with a W boson in proton-proton collisions at $\sqrt{s} = 13$ TeV". *JHEP* 07 (2023), p. 219. DOI: 10.1007/JHEP07(2023)219. arXiv: 2208.06485.

[90] CMS Collaboration. "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.

[91] CMS Collaboration. "Inclusive and differential cross section measurements of single top quark production in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV". *JHEP* 02 (2022), p. 107. DOI: 10.1007/JHEP02(2022)107. arXiv: 2111.02860.

[92] A. Prinzie and D. Van den Poel. "Random multiclass classification: Generalizing random forests to random MNL and random NB". *Proc. 18th International Conference on Database and Expert Systems Applications*. 2007, p. 349. DOI: 10.1007/978-3-540-74469-6_35.

[93] F. Pedregosa et al. "Scikit-learn: Machine learning in Python". *J. Mach. Learn. Res.* 12 (2011), p. 2825. URL: https://jmlr.org/papers/v12/pedregosa11a.html.

[94] N. Kidonakis. "Theoretical results for electroweak-boson and single-top production". *PoS* DIS2015 (2015). arXiv: 1506.04072.

[95] J. Campbell, T. Neumann, and Z. Sullivan. "Single-top-quark production in the t-channel at NNLO". *JHEP* 02 (2021), p. 040. DOI: 10.1007/JHEP02(2021)040. arXiv: 2012.01574.

[96] CMS Collaboration. *Luminosity measurement in proton-proton collisions at 5.02 TeV in 2017 at CMS*. CMS Physics Analysis Summary CMS-PAS-LUM-19-001. 2021. URL: http://cds.cern.ch/record/2765655.

[97] M. Grazzini, S. Kallweit, and M. Wiesemann. "Fully differential NNLO computations with MATRIX". *Eur. Phys. J. C* 78 (2018), p. 537. DOI: 10.1140/epjc/s10052-018-5771-7. arXiv: 1711.06631.

[98] CMS Collaboration. *Performance of Methods for Data-Driven Background Estimation in SUSY Searches*. CMS Physics Analysis Summary CMS-PAS-SUS-10-001. 2010. URL: https://cds.cern.ch/record/1279147.

[99] ATLAS Collaboration. "Measurement of the ZZ production cross section in proton-proton collisions at $\sqrt{s} = 8$ TeV using the $ZZ \to \ell^-\ell^+\ell'^-\ell'^+$ and $ZZ \to \ell^-\ell^+\nu\bar{\nu}$ channels with the ATLAS detector". *JHEP* 01 (2017), p. 099. DOI: 10.1007/JHEP01(2017)099. arXiv: 1610.07585.

[100] CMS Collaboration. "Measurement of the W^+W^- cross section in pp collisions at $\sqrt{s} = 7$ TeV and limits on anomalous $WW\gamma$ and WWZ couplings". *Eur. Phys. J. C* 73 (2013), p. 2610. DOI: 10.1140/epjc/s10052-013-2610-8. arXiv: 1306.1126.

[101] CMS Collaboration. "Measurement of the W^+W^- cross section in pp collisions at $\sqrt{s} = 8$ TeV and limits on anomalous gauge couplings". *Eur. Phys. J. C* 76 (2016), p. 401. DOI: 10.1140/epjc/s10052-016-4219-1. arXiv: 1507.03268.

[102] CMS Collaboration. "Measurement of the WZ production cross section in pp collisions at $\sqrt{s} = 7$ and 8 TeV and search for anomalous triple gauge couplings at $\sqrt{s} = 8$ TeV". *Eur. Phys. J. C* 77 (2017), p. 236. DOI: 10.1140/epjc/s10052-017-4730-z. arXiv: 1609.05721.

[103] CMS Collaboration. "Measurements of the pp $\to$ 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.

[104] CMS Collaboration. "Measurement of the ZZ production cross section and search for anomalous couplings in $2\ell 2\ell'$ final states in pp collisions at $\sqrt{s} = 7$ TeV". *JHEP* 01 (2013), p. 063. DOI: 10.1007/JHEP01(2013)063. arXiv: 1211.4890.

[105] CMS Collaboration. "Measurement of the pp $\to$ ZZ production cross section and constraints on anomalous triple gauge couplings in four-lepton final states at $\sqrt{s}$ = 8 TeV". *Phys. Lett. B* 740 (2015), p. 250. DOI: 10.1016/j.physletb.2014.11.059. arXiv: 1406.0113.

[106] CMS Collaboration. "Measurements of pp $\rightarrow$ ZZ production cross sections and constraints on anomalous triple gauge couplings at $\sqrt{s} = 13\,\text{TeV}$". *Eur. Phys. J. C* 81 (2021), p. 200. DOI: 10.1140/epjc/s10052-020-08817-8. arXiv: 2009.01186.

[107] ATLAS Collaboration. "Measurement of W^+W^- production in pp collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector and limits on anomalous WWZ and WWγ couplings". *Phys. Rev. D* 87 (2013), p. 112001. DOI: 10.1103/PhysRevD.87.112001. arXiv: 1210.2979.

[108] ATLAS Collaboration. "Measurement of total and differential W^+W^- production cross sections in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector and limits on anomalous triple-gauge-boson couplings". *JHEP* 09 (2016), p. 029. DOI: 10.1007/JHEP09(2016)029. arXiv: 1603.01702.

[109] ATLAS Collaboration. "Measurement of the W^+W^- production cross section in pp collisions at a centre-of-mass energy of $\sqrt{s}$ = 13 TeV with the ATLAS experiment". *Phys. Lett. B* 773 (2017), p. 354. DOI: 10.1016/j.physletb.2017.08.047. arXiv: 1702.04519.

[110] ATLAS Collaboration. "Measurement of WZ production in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector". *Eur. Phys. J. C* 72 (2012), p. 2173. DOI: 10.1140/epjc/s10052-012-2173-0. arXiv: 1208.1390.

[111] ATLAS Collaboration. "Measurements of $W^{\pm}Z$ production cross sections in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector and limits on anomalous gauge boson self-couplings". *Phys. Rev. D* 93 (2016), p. 092004. DOI: 10.1103/PhysRevD.93.092004. arXiv: 1603.02151.

[112] ATLAS Collaboration. "Measurement of $W^{\pm}Z$ production cross sections and gauge boson polarisation in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". *Eur. Phys. J. C* 79 (2019), p. 535. DOI: 10.1140/epjc/s10052-019-7027-6. arXiv: 1902.05759.

[113] ATLAS Collaboration. "Measurement of ZZ production in pp collisions at $\sqrt{s} = 7$ TeV and limits on anomalous ZZZ and $ZZ\gamma$ couplings with the ATLAS detector". *JHEP* 03 (2013), p. 128. DOI: 10.1007/JHEP03(2013)128. arXiv: 1211.6096.

[114] ATLAS Collaboration. "$ZZ \rightarrow \ell^+\ell^-\ell'^+\ell'^-$ cross-section measurements and search for anomalous triple gauge couplings in 13 TeV pp collisions with the ATLAS detector". *Phys. Rev. D* 97 (2018), p. 032005. DOI: 10.1103/PhysRevD.97.032005. arXiv: 1709.07703.

[115] CDF Collaboration. "Measurement of the WW and WZ production cross section using final states with a charged lepton and heavy-flavor jets in the full CDF Run II data set". *Phys. Rev. D* 94 (2016), p. 032008. DOI: 10.1103/PhysRevD.94.032008. arXiv: 1606.06823.

[116] CDF Collaboration. "Measurement of the ZZ production cross section using the full CDF II data set". *Phys. Rev. D* 89 (2014), p. 112001. DOI: 10.1103/PhysRevD.89.112001. arXiv: 1403.2300.

[117] D0 Collaboration. "Measurements of WW and WZ production in W+jets final states in $p\bar{p}$ collisions". *Phys. Rev. Lett.* 108 (2012), p. 181803. DOI: 10.1103/PhysRevLett.108.181803. arXiv: 1112.0536.

[118] D0 Collaboration. "A measurement of the WZ and ZZ production cross sections using leptonic final states in 8.6 fb^{-1} of $p\bar{p}$ collisions". *Phys. Rev. D* 85 (2012), p. 112005. DOI: 10.1103/PhysRevD.85.112005. arXiv: 1201.5652.

[119] D0 Collaboration. "Measurement of the ZZ production cross section and search for the standard model Higgs boson in the four lepton final state in $p\bar{p}$ collisions". *Phys. Rev. D* 88 (2013), p. 032008. DOI: 10.1103/PhysRevD.88.032008. arXiv: 1304.5422.

[120] Johan Alwall, Philip Schuster, and Natalia Toro. "Simplified models for a first characterization of new physics at the LHC". *Phys. Rev. D* 79 (2009). DOI: 10.1103/PhysRevD.79.075020. arXiv: 0810.3921.

[121] D. Alves. "Simplified models for LHC new physics searches". *J. Phys. G* 39 (2012). Ed. by Nima Arkani-Hamed et al. DOI: 10.1088/0954-3899/39/10/105005. arXiv: 1105.2838.

[122] ATLAS Collaboration. "Search for top squarks in final states with one isolated lepton, jets, and missing transverse momentum in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector". *Phys. Rev. D* 94 (2016). DOI: 10.1103/PhysRevD.94.052009. arXiv: 1606.03903.

[123] ATLAS Collaboration. "Search for direct top-squark pair production in final states with two leptons in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector". *JHEP* 06 (2014), p. 124. DOI: 10.1007/JHEP06(2014)124. arXiv: 1403.4853.

[124] ATLAS Collaboration. "Search for top squark pair production in final states with one isolated lepton, jets, and missing transverse momentum in $\sqrt{s} = 8$ TeV pp collisions with the ATLAS detector". *JHEP* 11 (2014), p. 118. DOI: 10.1007/JHEP11(2014)118. arXiv: 1407.0583.

[125] ATLAS Collaboration. "Search for direct top squark pair production in final states with two leptons in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector". *Eur. Phys. J. C* 77 (2017). DOI: 10.1140/epjc/s10052-017-5445-x. arXiv: 1708.03247.

[126] CMS Collaboration. "Search for top squarks and dark matter particles in opposite-charge dilepton final states at $\sqrt{s} = 13$ TeV". *Phys. Rev. D* 97 (2018). DOI: 10.1103/PhysRevD.97.032009. arXiv: 1711.00752.

[127] CMS Collaboration. "Search for direct production of supersymmetric partners of the top quark in the all-jets final state in proton-proton collisions at $\sqrt{s} =$ 13 TeV". *JHEP* 10 (2017), p. 005. DOI: 10.1007/JHEP10(2017)005. arXiv: 1707.03316.

[128] CMS Collaboration. "Search for top squark pair production in pp collisions at $\sqrt{s} = 13$ TeV using single lepton events". *JHEP* 10 (2017), p. 019. DOI: 10.1007/JHEP10(2017)019. arXiv: 1706.04402.

[129] CMS Collaboration. "Search for the pair production of light top squarks in the $e^{\pm}\mu^{\mp}$ final state in proton-proton collisions at $\sqrt{s} = 13$ TeV". *JHEP* 03 (2019), p. 101. DOI: 10.1007/JHEP03(2019)101. arXiv: 1901.01288.

[130] R. 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.

[131] M. Czakon, P. Fiedler, and A. Mitov. "Total top-quark pair-production cross section at hadron colliders through $O(\alpha_S^4)$". *Phys. Rev. Lett.* 110 (2013). DOI: 10.1103/PhysRevLett.110.252004. arXiv: 1303.6254.

[132] H. Cheng and Z. Han. "Minimal kinematic constraints and m_{T2}". *JHEP* 12 (2008), p. 063. DOI: 10.1088/1126-6708/2008/12/063. arXiv: 0810.5178.

[133] J. Schmidhuber. "Deep Learning". *Scholarpedia* 10.11 (2015), p. 32832. DOI: 10.4249/scholarpedia.32832.

[134] Y. LeCun, Y. Bengio, and G. Hinton. "Deep Learning". *Nature* 521 (2015), 436–444. DOI: https://doi.org/10.1038/nature14539.

[135] M. Abadi et al. "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems". Software available from https://www.tensorflow.org/. 2016. arXiv: 1603.04467.

[136] F. Chollet et al. *Keras*. Software available from https://keras.io. 2015.

[137] P. Kingma Diederik and B. Jimmy. "Adam: A Method for Stochastic Optimization". Software available from https://keras.io/api/optimizers/adam/. 2014. arXiv: 1412.6980 [cs.LG].

[138] CMS Collaboration. "Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV". *JINST* 10 (2015), P08010. DOI: 10.1088/1748-0221/10/08/P08010. arXiv: 1502.02702.

[139] CMS Collaboration. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". *JINST* 9 (2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009. arXiv: 1405.6569.

[140] CMS Collaboration. "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.

[141] C. 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 (2014). DOI: 10.1140/epjc/s10052-014-3174-y. arXiv: 1407.5066.

[142] CMS Collaboration. "Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS". *Eur. Phys. J. C* 81 (2021), p. 800. DOI: 10.1140/epjc/s10052-021-09538-2. arXiv: 2104.01927.

[143] CMS Collaboration. *CMS luminosity measurements for the 2017 data-taking period*. CMS Physics Analysis Summary CMS-PAS-LUMI-17-004. 2017. URL: https://cds.cern.ch/record/2621960.

[144] CMS Collaboration. *CMS luminosity measurements for the 2018 data-taking period at $\sqrt{s} = 13$ TeV*. CMS Physics Analysis Summary CMS-PAS-LUMI-18-002. 2019. URL: https://cds.cern.ch/record/2676164.

[145] R. Barlow and C. Beeston. "Fitting using finite Monte Carlo samples". *Comput. Phys. Commun.* 77 (1993), p. 219. DOI: 10.1016/0010-4655(93)90005-W.

[146] G. Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". *Eur. Phys. J. C* 71 (2011). DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727.

[147] T. Junk. "Confidence level computation for combining searches with small statistics". *Nucl. Instrum. Meth. A* 434 (1999). DOI: 10.1016/S0168-9002(99)00498-2.

[148] Alexander L. Read. "Presentation of search results: The CL_s technique". *J. Phys. G* 28 (2002). DOI: 10.1088/0954-3899/28/10/313.

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

[150] T. Lin, Edward W. Kolb, and L. Wang. "Probing dark matter couplings to top and bottom quarks at the LHC". *Phys. Rev. D* 88 (2013), p. 063510. DOI: 10.1103/PhysRevD.88.063510. arXiv: 1303.6638.

[151] M. R. Buckley, . Feld, and D. Goncalves. "Scalar simplified models for dark matter". *Phys. Rev. D* 91 (2015), p. 015017. DOI: 10.1103/PhysRevD.91.015017. arXiv: 1410.6497.

[152] U. Haisch and E. Re. "Simplified dark matter top-quark interactions at the LHC". *JHEP* 06 (2015), p. 078. DOI: 10.1007/JHEP06(2015)078. arXiv: 1503.00691.

[153] A. Chiara et al. "A comprehensive approach to dark matter studies: exploration of simplified top-philic models". *JHEP* 11 (2016), p. 111. DOI: 10.1007/JHEP11(2016)111. arXiv: 1605.09242.

[154] D. Abercrombie et al. "Dark matter benchmark models for early LHC Run-2 searches: report of the ATLAS/CMS dark matter forum". *Phys. Dark Univ.* 27 (2020). Ed. by Antonio Boveia et al., p. 100371. DOI: 10.1016/j.dark.2019.100371. arXiv: 1507.00966.

[155] G. D'Ambrosio et al. "Minimal flavor violation: An effective field theory approach". *Nucl. Phys. B* 645 (2002), p. 155. DOI: 10.1016/S0550-3213(02)00836-2. arXiv: /0207036.

[156] G. Isidori and D. M. Straub. "Minimal flavour violation and beyond". *Eur. Phys. J. C* 72 (2012), p. 2103. DOI: 10.1140/epjc/s10052-012-2103-1. arXiv: 1202.0464.

[157] ATLAS Collaboration. "Search for dark matter produced in association with bottom or top quarks in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector". *Eur. Phys. J. C* 78 (2018), p. 18. DOI: 10.1140/epjc/s10052-017-5486-1. arXiv: 1710.11412.

[158] ATLAS Collaboration. "Search for new phenomena in events with two opposite-charge leptons, jets and missing transverse momentum in pp collisions at $\sqrt{s} =$ 13 TeV with the ATLAS detector". *JHEP* 04 (2021), p. 165. DOI: 10.1007/JHEP04(2021)165. arXiv: 2102.01444.

[159] CMS Collaboration. "Search for dark matter particles produced in association with a top quark pair at $\sqrt{s} = 13$ TeV". *Phys. Rev. Lett.* 122 (2019), p. 011803. DOI: 10.1103/PhysRevLett.122.011803. arXiv: 1807.06522.

[160] CMS Collaboration. "Search for dark matter produced in association with a single top quark or a top quark pair in proton-proton collisions at $\sqrt{s} = 13$ TeV". *JHEP* 03 (2019), p. 141. DOI: 10.1007/JHEP03(2019)141. arXiv: 1901.01553.

[161] CMS Collaboration. "Search for production of four top quarks in final states with same-sign or multiple leptons in proton-proton collisions $\sqrt{s} = 13$ TeV". *Eur. Phys. J. C* 80 (2020), p. 75. DOI: 10.1140/epjc/s10052-019-7593-7. arXiv: 1908.06463.

[162] O. Mattelaer and E. Vryonidou. "Dark matter production through loop-induced processes at the LHC: the s-channel mediator case". *Eur. Phys. J. C* 75 (2015), p. 436. DOI: 10.1140/epjc/s10052-015-3665-5. arXiv: 1508.00564.

[163] CMS Collaboration. "Evidence for four-top quark production in proton-proton collisions at $\sqrt{s} = 13$ TeV". *Phys. Rev. Lett.* 844 (2023). DOI: /10.1016/j.physletb.2023.138076. arXiv: 2303.03864.

[164] CMS Collaboration. "Probing effective field theory operators in the associated production of top quarks with a Z boson in multilepton final states at $\sqrt{s} = 13$ TeV". *JHEP* 12 (2021), p. 083. DOI: 10.1007/JHEP12(2021)083. arXiv: 2107.13896.

[165] J. A. Aguilar-Saavedra et al. *Interpreting top-quark LHC measurements in the standard-model effective field theory*. LHC TOP WG note CERN-LPCC-2018-01. 2018. arXiv: 1802.07237. URL: http://cds.cern.ch/record/2305783.

[166] CMS Collaboration. "Search for new physics in top quark production with additional leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV using effective field theory". *JHEP* 03 (2021), p. 095. DOI: 10.1007/JHEP03(2021)095. arXiv: 2012.04120.

[167] R. Goldouzian et al. "Matching in pp $\to t\bar{t}W/Z$/+jet SMEFT studies". *JHEP* 06 (2021), p. 151. DOI: 10.1007/JHEP06(2021)151. arXiv: 2012.06872.

[168] LHC Higgs Cross Section Working Group, D. de Florian, et al. *Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector*. CERN Report CERN-2017-002-M. 2016. DOI: 10.23731/CYRM-2017-002. arXiv: 1610.07922.

[169] R. Frederix and I. Tsinikos. "On improving NLO merging for $t\bar{t}W$ production". *JHEP* 11 (2021), p. 029. DOI: 10.1007/JHEP11(2021)029. arXiv: 2108.07826.

[170] R. Frederix and S. Frixione. "Merging meets matching in ". *JHEP* 12 (2012), p. 061. DOI: 10.1007/JHEP12(2012)061. arXiv: 1209.6215.

[171] J. Alwall et al. "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions". *Eur. Phys. J. C* 53 (2008), p. 473. DOI: 10.1140/epjc/s10052-007-0490-5. arXiv: 0706.2569.

[172] M. van Beekveld, A. Kulesza, and L. Moreno Valero. "Threshold resummation for the production of four top quarks at the LHC". 2022. arXiv: 2212.03259.

[173] O. Mattelaer. "On the maximal use of Monte Carlo samples: re-weighting events at NLO accuracy". *Eur. Phys. J. C* 76 (2016), p. 674. DOI: 10.1140/epjc/s10052-016-4533-7. arXiv: 1607.00763.

[174] J. M. Campbell and R. K. Ellis. "An update on vector boson pair production at hadron colliders". *Phys. Rev. D* 60 (1999), p. 113006. DOI: 10.1103/PhysRevD.60.113006.

[175] W. Verkerke and D. Kirkby. "The RooFit toolkit for data modeling". *Proc. 13th Int. Conf. on Computing in High Energy and Nuclear Phys. (CHEP 2003): La Jolla CA, United States, March 24–28, 2003*. [eConf C0303241 (2003) MOLT007]. 2003. arXiv: physics/0306116. URL: https://www.slac.stanford.edu/econf/C0303241/proc/papers/MOLT007.PDF.

[176] CMS Collaboration. "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 (2019), p. 132003. DOI: 10.1103/PhysRevLett.122.132003. arXiv: 1812.05900.

[177] CMS Collaboration. "Performance of the CMS muon trigger system in proton-proton collisions at $\sqrt{s} = 13$ TeV". *JINST* 16 (2021), P07001. DOI: 10.1088/1748-0221/16/07/P07001. arXiv: 2102.04790.

[178] CMS Collaboration. "Measurement of properties of $B_s^0 \to \mu^+\mu^-$ decays and search for $B^0 \to \mu^+\mu^-$ with the CMS experiment". *JHEP* 04 (2020), p. 188. DOI: 10.1007/JHEP04(2020)188. arXiv: 1910.12127.

[179] E. M. Metodiev, B. Nachman, and J. Thaler. "Classification without labels: Learning from mixed samples in high energy physics". *JHEP* 10 (2017), p. 174. DOI: 10.1007/JHEP10(2017)174. arXiv: 1708.02949.