Thesis for the Degree of Ph.D.

Inclusive search for supersymmetry in pp collisions at $\sqrt{s}$ = 13 TeV using boosted object identification and razor kinematic variables in zero and one lepton final states

Changgi Huh

Department of Physics, Major in High Energy Physics

The Graduate School

December 2023

The Graduate School
Kyungpook National University

Inclusive search for supersymmetry in pp collisions at $\sqrt{s}$ = 13 TeV using boosted object identification and razor kinematic variables in zero and one lepton final states

Changgi Huh

Department of Physics, Major in High Energy Physics
The Graduate School

Supervised by Professor Prof. Sehwook Lee

Approved as a qualified thesis of Changgi Huh
for the degree of Ph.D. by the Evaluation Committee

December 2023

Chairman Prof. Hyon-suk Jo

Prof. Minsuk Kim

Prof. Chang-Seong Moon

Dr. Min Sang Ryu

Prof. Sehwook Lee

The Graduate School Council
Kyungpook National University

Abstract

A search for supersymmetry is conducted in final states consisting of hadrons and leptons where top quarks or W/Z/Higgs bosons are highly boosted. Proton-proton collision data at a center-of-mass energy of 13 TeV is utilized for the search, gathered by the CMS experiment at the LHC with an integrated luminosity of 138 fb^{-1}. Events that show potential for the boosted top quarks and W/Z/Higgs bosons undergoing hadronic or leptonic decays are identified through jet substructure techniques and then analyzed using the razor variables M^R and R^2 to identify any potential signal as a peak on a smoothly falling background. The observed event yields found in the signal regions are consistent with the predicted contributions from standard model processes that are made using control samples in the data. This search covers both R-parity conserving and violating supersymmetric models. The study findings are analyzed with respect to the production of gluino pairs, where gluinos disintegrate into top squarks and/or top quarks, or jets and vector bosons, all accompanied by neutralinos. In addition, the research explores direct production of top squark pairs, where top squarks disintegrate into top quarks and neutralinos, bottom squark pairs, where bottom squarks decay into b quarks and neutralinos, and neutralino and charigno pairs, where neutralino and charigno decay into vector bosons and neutralinos. The run2(138 fb1) expected limits shows we could reject gluino mass up to 2 TeV and top squark mass up to 1 TeV. We did unblinding 2017 (41.5 fb1) and observed and expected limits are agreed within 2σ.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The Standard Model (SM) of particle physics has successfully predicted and explained the existence of fundamental particles, such as the charm, bottom, top quarks, and tau leptons. The strong interaction, described by Quantum Chromodynamics (QCD), is explained through the exchange of gluons between quarks. The SM also predicts the existence of the W and Z bosons, which mediate the weak interaction, and the Higgs boson, responsible for endowing other particles with mass through the Higgs mechanism. The discovery of the Higgs boson in 2012 by the ATLAS and CMS experiments at the Large Hadron Collider (LHC) was a significant achievement.

However, despite these successes, the SM still leaves several unanswered questions. These include the matter-antimatter asymmetry in the universe, the nature of dark matter, neutrino oscillation phenomena, and the incorporation of gravity into the framework. Addressing these outstanding questions requires extending beyond the Standard Model (BSM) and exploring new theoretical frameworks.

One such extension is Supersymmetry (SUSY) [1], which introduces a new space-time symmetry called R-symmetry. SUSY predicts the existence of new particles known as ”superpartners,” with spins opposite to their corresponding SM particles (fermions have bosonic superpartners and vice versa). SUSY provides a framework that unifies all fundamental forces at very high energy scales.

In this thesis, I focus on the razor boost analysis, the primary research topic of

my Ph.D. The razor boost analysis specifically targets natural SUSY models and R-parity violations. It utilizes the razor kinematic variables to search for signatures of new physics in both hadronic and leptonic final states. These signatures include highly boosted W, Z, Higgs bosons, and top quarks.

The first chapter of this thesis provides an overview of the SM and introduces the concept of supersymmetry, with a particular emphasis on natural supersymmetry and its phenomenological implications. Chapter 2 provides an overview of the LHC and the CMS experiment, which was used to collect the data for this analysis. The simulation and reconstruction techniques employed in the analysis are described in Chapter 3. The details of the razor boost analysis itself are presented in Chapter 4, and a summary of the findings and prospects is provided in Chapter 5. Additional work conducted during my Ph.D. is documented in Appendix B.

1.1 Standard Model

The SM of particle physics represents our current understanding of elementary particles and their interactions. Developed over the past few decades and supported by numerous experimental observations, the SM has proven highly successful in predicting and explaining various physical phenomena. Figure 1.1 provides an overview of the particles described by the SM and their associated quantum numbers.

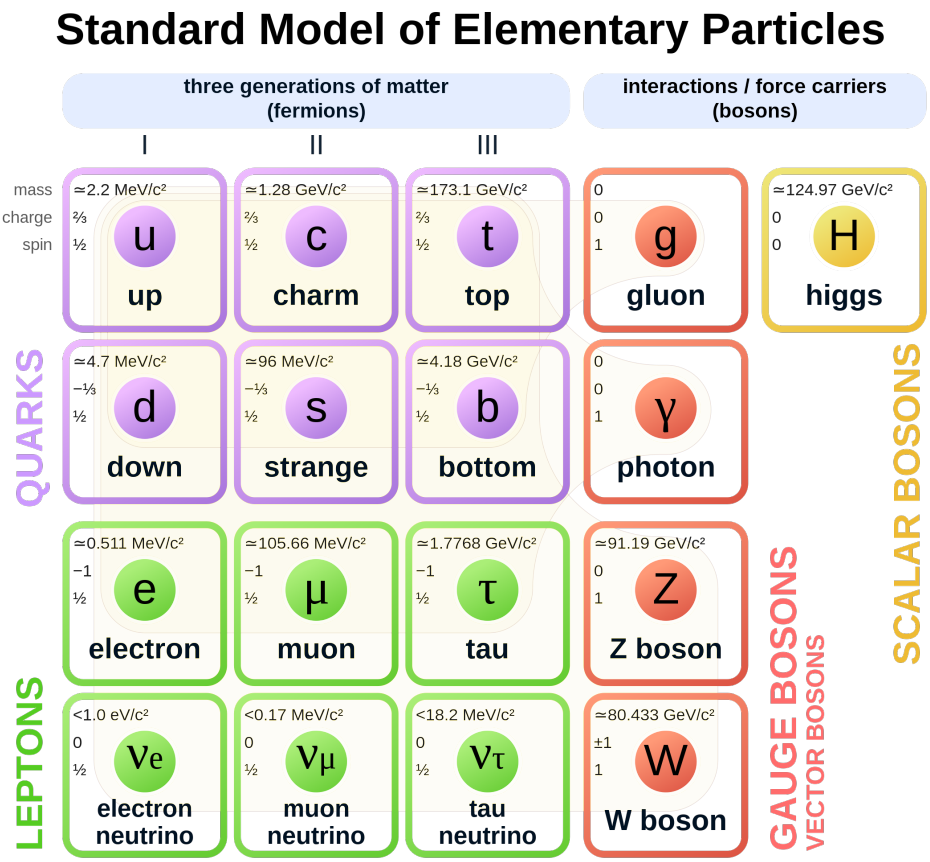


Figure 1.1: Overview of all particles described by the SM, including their respective quantum numbers. [2]

In the quantum field theory (QFT) framework, the Lagrangian of particle models encompasses the particles themselves and their interactions. QFT aims to elucidate

the fundamental nature of our universe and, to be consistent, must be renormalizable [3]. The interactions between particles in QFT are described using Feynman diagrams. Perturbation theory allows for calculating interaction probabilities to arbitrary orders for QFT with small coupling constants [4].

The underlying symmetry group of the SM is given by

$$SU(3)_C \times SU(2)_L \times U(1)_Y, \tag{1.1}$$

where $SU(3)_C$ represents the symmetry associated with the strong interaction, or quantum chromodynamics (QCD), which describes the color degree of freedom (C) of quarks. $SU(2)_L$ and $U(1)_Y$ describe the weak and electromagnetic interactions, respectively. $SU(2)_L$ symmetry acts on left-handed fermions (L) and is associated with weak isospin (T_3), while $U(1)_Y$ symmetry pertains to electromagnetic interactions and is characterized by the quantum number hypercharge (Y).

Fermions The fermions in the SM are fundamental particles that possess a spin of 1/2. These fermions can be divided into two categories: leptons and quarks. Each category consists of six particles, organized into three generations. While particles within the same generation exhibit similar properties, the mass of the particles increases with each successive generation. However, higher-generation particles tend to be unstable and can decay into corresponding particles from lower generations.

Leptons always comprise a negatively charged lepton and an electrically neutral neutrino, which, according to the SM, is considered to be massless. The first generation includes the electron (e) and the electron neutrino (ν_e), the second generation consists of the muon (μ) and the muon neutrino (ν_μ), and the third generation encompasses the tau (τ) and the tau neutrino (ν_τ).

Quarks, on the other hand, carry the color charge and experience strong interaction. Each generation of quarks comprises two types: one with an electric charge of +2/3 and the other with a charge of -1/3. The first generation includes the up (u) quark and the down (d) quark, the second generation encompasses the charm (c) quark and the strange (s) quark, and the third generation comprises the top (t) quark and the bottom (b) quark.

Bosons Bosons, characterized by integer spin values, are responsible for mediating the interactions between particles in the SM. The photon (γ) is the gauge boson associated with the electromagnetic force. The massive vector bosons mediate the weak force: the $W^{\pm}$ bosons and the Z boson. Additionally, the strong force, which acts on particles with color charge, is carried by eight massless gluons (g). Finally, the Higgs boson is a scalar boson with a spin of 0 and plays a crucial role in the mechanism through which particles acquire mass within the SM framework.

Quantum Electrodynamics Quantum Electrodynamics (QED) is a quantum field theory that describes the electromagnetic interaction between particles carrying electric charges. This interaction is mediated by massless and neutral gauge bosons known as photons (γ). The properties of photons play a crucial role in determining the structure and dynamics of all matter. At the microscopic level, QED explains the formation of atoms from nuclei and electrons, while at the macroscopic level, it defines the mechanical properties of various materials.

Building upon the concepts of QED, Glashow, Salam, and Weinberg formulated a unified theory that describes electromagnetic and weak forces within a single, consistent framework. This theory is based on the combined symmetry group $U(1)_Y \times SU(2)_L$. In the formalism of quantum field theory, the gauge bosons and their interactions naturally emerge. The $U(1)_Y$ symmetry group gives rise to a single gauge boson denoted as B, while the $SU(2)_L$ symmetry group requires the introduction of three gauge bosons: W^1, W^2, and W^3. These initial states are then mixed to obtain the physical bosons.

The charged weak bosons are obtained by combining W^1 and W^2 as follows:

$$\begin{pmatrix} \mathcal{W}^+ \\ \mathcal{W}^- \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} W^1 \\ -iW^2 \end{pmatrix}, \tag{1.2}$$

where $\mathcal{W}^+$ and $\mathcal{W}^-$ represent the charged weak bosons.

To obtain the photon (γ) and the Z boson, mixing between the gauge bosons B and W^3 is performed. This mixing is described by a rotation matrix:

$$\begin{pmatrix} \gamma \\ Z \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B \\ W^3 \end{pmatrix}, \tag{1.3}$$

where θ_W represents the weak mixing angle, also known as the Weinberg angle.

This mixing between the gauge bosons B and W^3 relates the strength of the electromagnetic interaction (characterized by the electric charge e) to the strength of the weak interaction (characterized by the coupling constant g) through the equation:

$$e = g \sin\theta_W . \tag{1.4}$$

Moreover, the electric charge (Q) of a particle can be determined by combining the weak isospin (T_3) and weak hypercharge (Y) according to:

$$Q = T_3 + \frac{Y}{2}. \tag{1.5}$$

In summary, the construction described above yields two electrically charged weak bosons ($\mathcal{W}^+$ and $\mathcal{W}^-$) that couple to the weak isospin. As in QED, the photon (γ) couples solely to the electric charge. The Z boson, on the other hand, couples to both the electric charge and the weak isospin with different strengths. Notably, all weak gauge bosons carry weak charges and can directly interact with one

another, the charged weak bosons can also interact with the photon due to their electric charge.

Quantum chromodynamics Quantum Chromodynamics (QCD) studies the strong sector in SM. In QCD, quark fields are charged under $SU(3)_C$ and mediated via massless gluons. Quark is a triplet of 3 colours, whereas the leptons are singlets. The symmetry group $SU(3)_C$ gives eight massless gluons that couple to all particles that carry color charge, including themselves. The three states of the color charge are typically denoted as red, green, and blue, whereas antiquarks carry the corresponding anticolors. Requiring local gauge invariance under a set of transformations that form an SU(3) group results in the following QCD Lagrangian :

$$\mathcal{L}_{QCD} = \sum_q \bar{\psi}_{q,a} \left(i\gamma^\mu \partial_\mu \delta_{a,b} - g_s \gamma^\mu t^C_{a,b} \mathcal{A}^C_\mu \right) \psi_{q,b} - \tfrac{1}{4} F^A_{\mu\nu} G^{A\mu\nu}. \tag{1.6}$$

Confinement refers to quarks exclusively observed in color-neutral, bound states, so-called hadrons. This can be bound states of three quarks of different colors, denoted as baryons or mesons, which contain a quark-antiquark pair with the corresponding (anti)color. However, these particles are only the valence quarks of the hadron. These permanent exchange gluons, which in turn exchange gluons themselves or create virtual quark-antiquark pairs.

The confinement phenomenon is a direct consequence of the self-interaction of the gluons and their zero mass, which implies that the coupling strength of the strong interaction increases with increasing distance. This means that if colored objects are separated from each other, it is energetically favorable to generate new colored particles until color-neutral states are produced. Asymptotic freedom refers to the opposite effect: at small distances, the strong coupling is small, and colored particles behave as free objects with respect to the strong interaction.

Higgs Mechanism The theoretical framework of the SM requires that gauge bosons are massless. Similarly, fermions of the same $SU(2)_L$ doublet, like an electron and the corresponding neutrino, have to have identical mass according to the theory, which is also not the case. This contradiction with respect to the experimental observations is solved by the *Brout-Englert-Higgs* mechanism. The mechanism postulates the Higgs field ϕ and the corresponding *Higgs potential* $V(\phi)$. The Higgs field is a $SU(2)_L$ doublet

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \tag{1.7}$$

where each component is a complex scalar field, the Higgs potential is symmetric with respect to the origin and has a non-trivial minimum. It is given by

$$V(\phi) = \mu^2|\phi|^2 + \lambda|\phi|^4, \tag{1.8}$$

with the parameters $-\mu^2$, $\lambda > 0$. The potential has a continuous minimum at the

so-called *vacuum expectation value*

$$v = |\phi|_{min} = \sqrt{\frac{-\mu^2}{2\lambda}}, \tag{1.9}$$

The two-dimensional sketch of the potential is shown in Fig. 1.2.

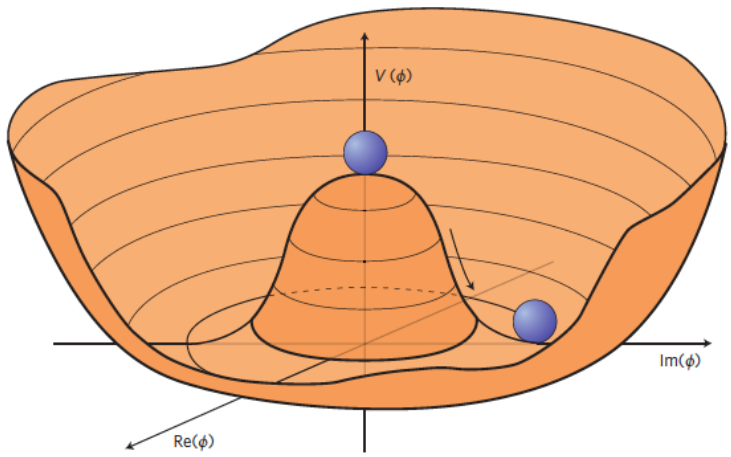


Figure 1.2: Two-dimensional sketch of the Higgs potential [5]. The ground state at $|\phi| = v$ does not have the symmetry of the potential.

The vacuum state of the potential is then chosen as

$$\phi = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\v\end{pmatrix}, \tag{1.10}$$

which spontaneously breaks the $SU(2)_L \times U(1)_Y$ symmetry of the electroweak model. Expanding the field in the radial direction with the parametrization

$$\phi = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\v+H(x)\end{pmatrix}, \tag{1.11}$$

and it generates the Higgs boson H with the mass

$$m_H = \sqrt{2}\mu. \tag{1.12}$$

Couplings of the W and Z bosons to the Higgs field generate terms that can be identified with the masses

$$\begin{aligned} m_W &= \tfrac{1}{2} g v \\ m_Z &= \tfrac{m_W}{cos\theta_W}, \end{aligned} \tag{1.13}$$

and terms describing the interaction between the Higgs and gauge bosons. The strength of these interactions is proportional to the mass of the gauge boson. Another consequence of the *Brout-Englert-Higgs* mechanism is self-interaction terms of the Higgs boson that are described by the coupling strength λ.

The Higgs boson can also provide mass terms for the fermions that are consistent with the theory, even though the approach can be criticized as somewhat arbitrary: For each fermion mass m_f a new coupling constant y_f is introduced, while no prediction about the values of those constants is made. This is done via so-called *Yukawa couplings*, which generate mass terms

$$m_f = y_f \frac{v}{\sqrt{2}}. \tag{1.14}$$

as well as couplings of the fermions and the Higgs boson. As for the gauge bosons, the strength of these interactions is proportional to the mass of the fermion.

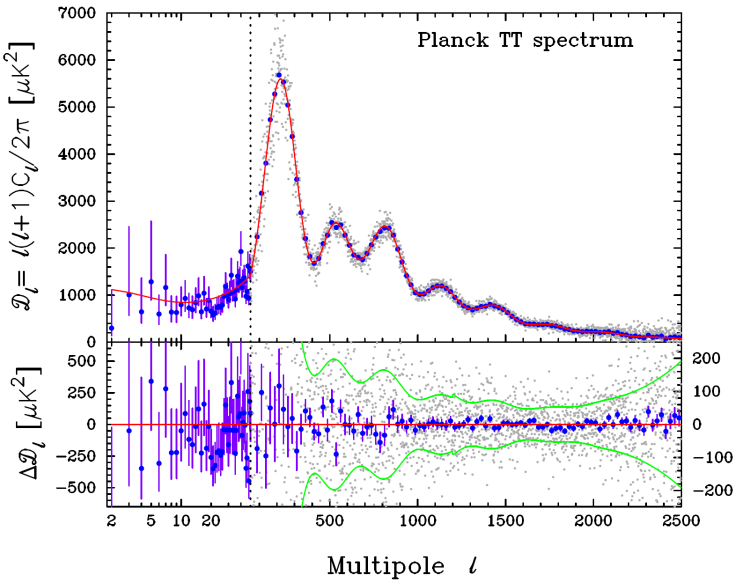


Figure 1.3: CMB measurement from the Planck Satellite. [6]

1.2 Dark Matter

Dark Matter (DM) continues to captivate scientists as the most perplexing puzzle in studying the universe. Recent measurements of the Cosmic Microwave Background radiation (CMB) by the Planck satellite experiment, as depicted in Figure 1.3, reveal that ordinary matter contributes a mere 4.5% to the total matter content. At the same time, an astonishing 26.5% is attributed to the matter of unknown origin, leaving us in the dark about a staggering 83% of the matter content.

In 1970, Vera C. Rubin revolutionized our understanding of galaxies by reporting the rotation velocity of our own Milky Way. Figure 1.4 presents two distinct curves: one depicting the expected rotation velocity based on visible matter distribution and the other representing the observed rotation velocity. Contrary to expectations, the observed data indicates a significant excess of matter towards the outer edges of galaxies. While we have identified some of these invisible halo objects as MAssive Compact Halo Objects (MACHOs), which account for about 20% of the halo mass, the majority of the dark matter remains elusive. MACHOs consist of non-luminous bodies, such as brown dwarfs.

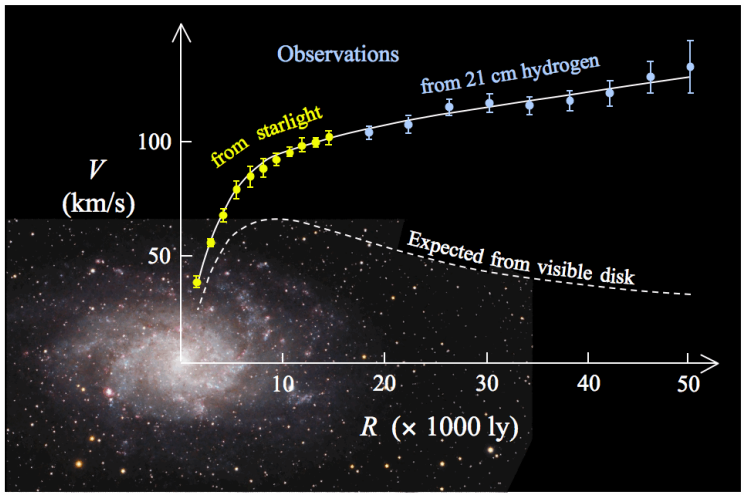


Figure 1.4: M33 galaxy rotation curve. [7]

Today, we have harnessed the power of gravitational lensing to detect MACHOs. The gravitational lensing effect, wherein massive objects such as stars or black holes act as lenses, allows us to observe the bending and focusing of light. By measuring the deflection angle of light, which is 1.7 arcseconds for the Sun, we can estimate the mass and distance of non-luminous bodies from Earth.

In the field of astrophysics, the leading candidate for dark matter is the Weakly Interacting Massive Particle (WIMP), a type of Cold Dark Matter characterized by its slow movement and weak interaction with electromagnetic forces while predominantly interacting through weak and gravitational interactions. Notably, WIMPs cannot be explained within the SM framework of particle physics. However, a stable particle with a mass at the weak scale (approximately 100 GeV) and a coupling strength of order one can account for the observed relic density of dark matter in the universe.

During the universe's early stages, particles within the SM could have constituted dark matter. However, as time progressed, the number density of dark matter particles diminished, leading to the situation, known as "freeze-out." This phenomenon can be resolved by applying the Friedman Equation, shedding light on today's abundance of dark matter.

The Hubble diagram presented in Figure 1.5, which incorporates data from type

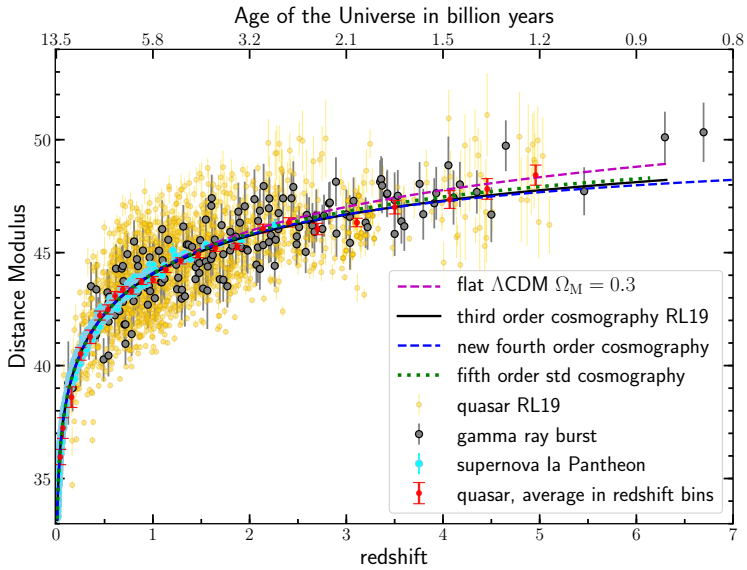


Figure 1.5: Hubble diagram illustrating the relationship between redshift and distance for type Ia supernovae (cyan points), quasars (yellow points), and gamma-ray bursts (black points) with 1σ uncertainties. The red points represent the mean distance modulus in narrow redshift bins for quasars. The dashed magenta line corresponds to a flat Λ_{CDM} model fit with $\Omega_M = 0.3$. [8]

Ia supernovae, quasars, and gamma-ray bursts, provides compelling evidence that the current matter fraction of the universe amounts to approximately 30%.

Detecting dark matter involves three primary approaches: direct search, indirect search, and collider experiments. Direct searches involve investigating WIMP-nucleon scattering to detect and measure the interactions between weakly interacting massive particles (WIMPs) and atomic nuclei. Indirect searches aim to identify secondary particles arising from WIMP interactions, such as gamma rays, neutrinos, and charged cosmic rays. Collider experiments, on the other hand, focus on studying high-energy particle collisions to generate and observe dark matter particles through pair production.

According to observations and measurements by the Planck satellite, approximately 22% of the universe consists of non-baryonic matter. Understanding the nature of this non-baryonic matter represents one of the most significant unresolved questions in modern astroparticle physics. Supersymmetry theories postulate the

existence of relic particles originating from the Big Bang, offering potential explanations for the mysterious dark matter component of the universe.

1.2.1 Supersymmetry

Supersymmetry (SUSY) is a theoretical framework in physics that proposes a fermionic spacetime symmetry, suggesting the existence of a partner particle for every known particle. In SUSY, each fermionic field is associated with a bosonic partner and vice versa. This symmetry is generated by supersymmetric transformations, which convert particles with different spins into each other. The transformations are described by the generator of transformations, denoted as Q.

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

The operator Q and its hermitian conjugate $Q^\dagger$ are fermionic operators, representing particles with a spin of $\frac{1}{2}$. This implies that SUSY is a spacetime symmetry rather than an internal or gauge symmetry. These operators satisfy specific commutation and anti-commutation relations, which are given by the following equations:

$$\begin{aligned} \{Q, Q^\dagger\} &= P^\mu \\ \{Q, Q\} &= \{Q^\dagger, Q^\dagger\} = 0 \\ [P^\mu, Q] &= [P^\mu, Q^\dagger] = 0 \end{aligned} \tag{1.16}$$

Here, P^μ represents the four-momentum generator of spacetime translations and is a bosonic operator [1]. The commutation and anti-commutation relations above define supersymmetry algebra or superalgebra. The fermions and bosons are arranged in irreducible representations of this algebra. These representations are known as supermultiplets, which consist of fermions and bosons. In each supermultiplet, the boson is the superpartner of the fermion.

A notable feature of supermultiplets is that they contain an equal number of fermions and bosons, denoted as $n_{\text{fermion}} = n_{\text{boson}}$. Additionally, particles within the same supermultiplet share the same gauge interaction. This means that they possess the same quantum numbers, such as color charge, electric charge, and the value of

Type	Spin	Mass eigenstate	↔	Weak/Gauge eigenstate
Squarks	0		$\tilde{u}_L, \tilde{u}_R, \tilde{d}_L, \tilde{d}_R$	
			$\tilde{s}_L, \tilde{s}_R, \tilde{c}_L, \tilde{c}_R$	
		$\tilde{t}_1, \tilde{t}_2, \tilde{b}_1, \tilde{b}_2$	↔	$\tilde{t}_L, \tilde{t}_R, \tilde{b}_L, \tilde{b}_R$
SLeptons	0		$\tilde{e}_L, \tilde{e}_R, \tilde{\nu}_e$	
			$\tilde{\mu}_L, \tilde{\mu}_R, \tilde{\nu}_\mu$	
		$\tilde{\tau}_1, \tilde{\tau}_2, \tilde{\nu}_\tau$	↔	$\tilde{\tau}_L, \tilde{\tau}_R, \tilde{\nu}_\tau$
Neutralinos	1/2	$\tilde{\chi}_1^0, \tilde{\chi}_2^0, \tilde{\chi}_3^0, \tilde{\chi}_4^0$	↔	$\tilde{B}^0, \tilde{W}^0, \tilde{H}_u^0, \tilde{H}_d^0$
Charginos	1/2	$\tilde{\chi}_1^\pm, \tilde{\chi}_2^\pm$	↔	$\tilde{W}^\pm, \tilde{H}_u^+, \tilde{H}_d^-$
Gluinos	1/2		$\tilde{g}$	

Table 1.1: The supersymmetric particles [1].

weak isospin coupling. The concept of supermultiplets allows for a deeper understanding of the connections between particles with different spins and properties within a unified framework.

R-parity

To address the strong experimental constraints on proton lifetime and incorporate certain important features, a multiplicative quantum number called R-parity is introduced for particles and sparticles in the framework of supersymmetry.

R-parity, denoted as P_R, is defined as:

$$P_R = (-1)^{3(B-L)+2S} \tag{1.17}$$

Here, S represents the spin of the particle or sparticle, while B and L correspond to the baryon number and lepton number, respectively. In the SM, all particles possess even R-parity, with $P_R = +1$, while supersymmetric particles have odd R-parity, with $P_R = -1$.

The conservation of R-parity has several important consequences [1]:

- In collider experiments, supersymmetric particles are always produced in pairs. This means that any decay involving supersymmetric particles must result in an odd number of daughter particles. Specifically, at any vertex in a decay process, an even number of supersymmetric particles is present.

- The lightest supersymmetric particle (LSP) is required to be stable. If the LSP is electrically neutral and weakly interacting, it becomes a viable candidate for dark matter. In fact, such a neutral and weakly interacting LSP could account for approximately 24% of the content of the universe [9].
- All decay chains involving non-LSP supersymmetric particles must result in an odd number (usually one) of LSPs.

The conservation of R-parity ensures the stability of the LSP and provides a potential dark matter candidate within the framework of supersymmetry. Additionally, it leads to distinctive experimental signatures, such as missing energy carried away by the LSP, enabling the search for supersymmetry in collider experiments and astrophysical observations.

Chapter 2

Large Hadron Collider (LHC) and Compact Muon Solenoid (CMS)

In recent decades, the field of particle physics has witnessed a relentless quest for new particles, prompting the construction of increasingly expansive and high-energy colliders. This pursuit reached its pinnacle with the remarkable achievement of discovering all the particles anticipated by the SM, including the elusive Higgs boson, at the Large Hadron Collider (LHC). Situated at the European Organization for Nuclear Research (CERN) in France-Switzerland, the LHC stands as the largest and most powerful particle collider worldwide, boasting a circumference of 27 kilometers. This colossal scientific instrument facilitates a multitude of experiments employing distinct detectors, each designed with a specific scientific objective in mind.

2.1 Large Hadron Collider (LHC)

The LHC, depicted in Figure 2.1, serves as the final stage in a series of accelerators. Its primary objective is to accelerate protons or heavy ions to center-of-mass energies reaching up to 13 TeV for protons and 2.76 TeV per nucleon for lead ions. This colossal machine encompasses multiple collision points strategically positioned with detectors.

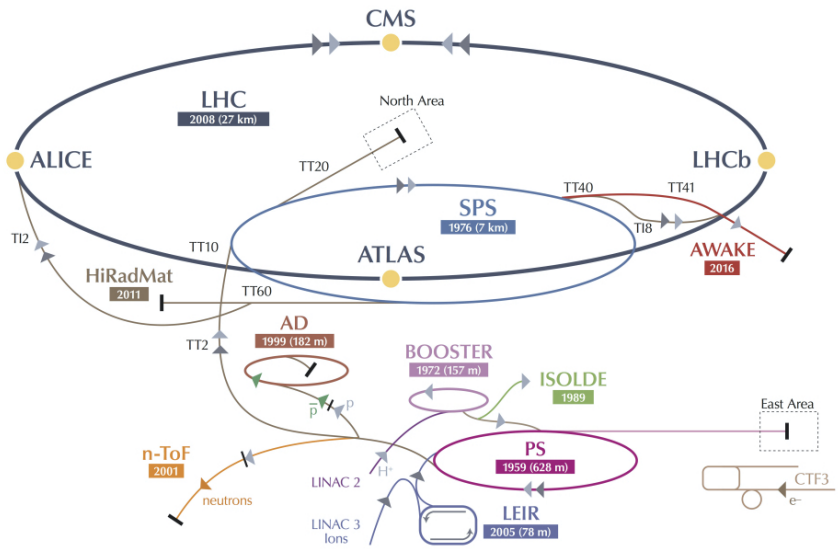


Figure 2.1: Representation of the CERN accelerators. [10]

The LHC is driven by fundamental questions in particle physics. One of its primary goals is to unravel the mysteries surrounding the origin of mass. Within the framework of the SM, the Higgs mechanism elucidates how particles acquire mass. Alternative models, such as the Higgsless model, propose electroweak symmetry breaking as the source of mass. The discovery of the Higgs boson would lend substantial support to the SM. Another vital pursuit at the LHC is the quest for dark matter, with particular attention directed towards supersymmetry. Additionally, the LHC explores other captivating avenues, including investigations into extra dimensions, monopoles, and unified theories.

The Large Hadron Collider (LHC) achieved a significant milestone with the discovery of the Higgs boson in July 2012. This elusive particle was detected through proton-proton collisions at center-of-mass energies of 7 and 8 TeV. The LHC then operated at 13 TeV proton-proton collision energy from 2015 to 2018, followed by a prolonged shutdown until 2021 to facilitate upgrades and enhancements to the collider and detectors in preparation for 13.6 TeV proton-proton collisions.

The luminosity of the LHC relies on various beam parameters and is defined by the equation:

$$L = \frac{N_b^2 n_b f_{\text{rev}} \gamma_r}{4\pi\varepsilon_n \beta^*} F, \tag{2.1}$$

Here, N_b represents the number of particles per bunch, n_b denotes the number of bunches per beam, f_{rev} signifies the revolution frequency of each bunch, γ_r corresponds to the relativistic factor, ε_n denotes the normalized beam emittance, β^* represents the β function at the collision point, and F is a geometrical factor inversely proportional to the crossing angle of the two beams at the collision point. The peak design value for the LHC's instantaneous luminosity is $L = 10^{34}\text{cm}^{-2}\text{s}^{-1}$. The beam parameters are summarized in Table 2.1.

Parameter	Injection	Collision
Beam energy [GeV]	450	7000
Relativistic gamma factor (γ_r)	479.6	7461
Beam emittance (ε_n) [μ rad]	3.5	3.75
Half crossing angle [μ rad]	±160	±142.5
β function (β^*) [m]	18	0.55
Revolution frequency (f_{rev}) [Hz]	11245	
Number of bunches (n_b)	2808	
Particles per bunch (N_b)	1.15×10^{11}	

Table 2.1: Beam parameters for the LHC at injection and collision energy [11].

2.2 Compact Muon Solenoid (CMS)

The Compact Muon Solenoid (CMS) is a vital component of the Large Hadron Collider (LHC) and plays a critical role in advancing particle physics research. Its purpose is to detect and precisely measure the properties of particles generated in proton-proton collisions, ensuring high efficiency and accuracy. The CMS detector incorporates several key features that enable its successful operation.

One distinguishing feature of CMS is its superconducting solenoid, which generates a high magnetic field. This magnetic field allows for the precise measurement of charged particle momenta by curving their trajectories. This capability is particularly important for muons, as CMS excels in muon identification and momentum

resolution across a wide range of momenta within the region $|\eta| < 2.5$. For muons with transverse momenta (p_T) less than 1 TeV, CMS can unambiguously determine the sign of their charge, achieving a dimuon mass resolution of approximately 1% at 100 GeV.

CMS includes a silicon-based tracking system that reconstructs the paths of charged particles as they move through the detector. This system allows for precise measurements of particle momenta and delivers outstanding reconstruction efficiency. Furthermore, CMS utilizes a scintillating crystal-based electromagnetic calorimeter that precisely measures the energy of electrons and photons. The electromagnetic calorimeter achieves excellent energy resolution for both photons and electrons. It also has mass resolutions for diphotons and dielectrons at 100 GeV of around 1%, in addition to ensuring precise photon direction measurement from the primary interaction vertex. Furthermore, it provides efficient photon and lepton isolation, even at high luminosities.

To precisely measure hadronic particles, CMS incorporates a hadronic calorimeter with broad geometric coverage ($|\eta| < 5$) and fine lateral segmentation ($\Delta\eta \times \Delta\phi < 0.1 \times 0.1$). This calorimeter enables good resolution for jets and missing transverse energy (E_T^{miss}). CMS also features a muon tracking system, which complements the muon measurements made by the solenoid. This system efficiently identifies and measures muons with high precision.

In addition to its sub-detectors, CMS employs a sophisticated triggering system and offline tagging algorithms to identify specific physics events of interest. These include the tagging of τ-leptons and b-jets, as well as the identification of various physics signatures. The CMS collaboration, consisting of thousands of scientists and engineers, collaboratively analyzes the data recorded by the detector to extract meaningful physics results.

Figure 2.2 provides a general view of the CMS detector, highlighting its various components and their geometric arrangement.

With its high-performance capabilities and comprehensive design, the CMS detector significantly contributes to our understanding of fundamental particles and their interactions. It played a pivotal role in the discovery of the Higgs boson and

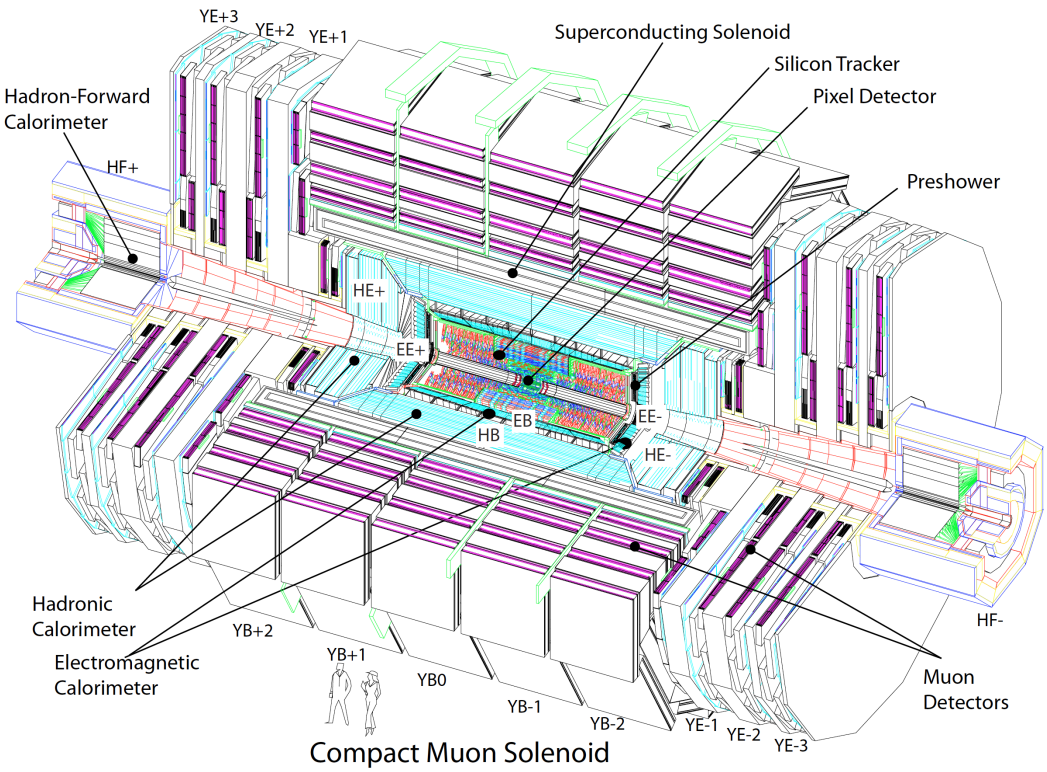


Figure 2.2: General view of the CMS detector [12].

continues to be instrumental in the search for new physics.

2.2.1 Magnet

The CMS detector incorporates a superconducting magnet that plays a crucial role in measuring the momentum of charged particles in the tracker. This magnet is an integral part of the CMS detector, and its strong magnetic field allows for precise curvature measurements of charged particle trajectories, which are essential for accurately determining particle momenta. Additionally, the magnet enables the tracker and calorimeters to be fully contained within its solenoid structure.

The CMS solenoid is constructed using a niobium-titanium (NbTi) alloy, which is a superconducting material. It operates at a current of 19 kA, generating a powerful magnetic field of 3.8 T. This strong magnetic field facilitates the bending of

charged particle trajectories, enabling precise momentum measurements. The inner radius of the magnet is designed to accommodate the tracker and calorimeters, ensuring that the paths of particles remain within the volume of the detector. The solenoid itself has a total weight of 11 kilotons, contributing to the overall mass of the CMS detector.

Figure 2.3 provides a longitudinal section of the CMS detector, illustrating the predicted magnetic field's value ($|B|$) and the corresponding field lines. The underground model depicted in the figure demonstrates a central magnetic flux density of 3.8 T.

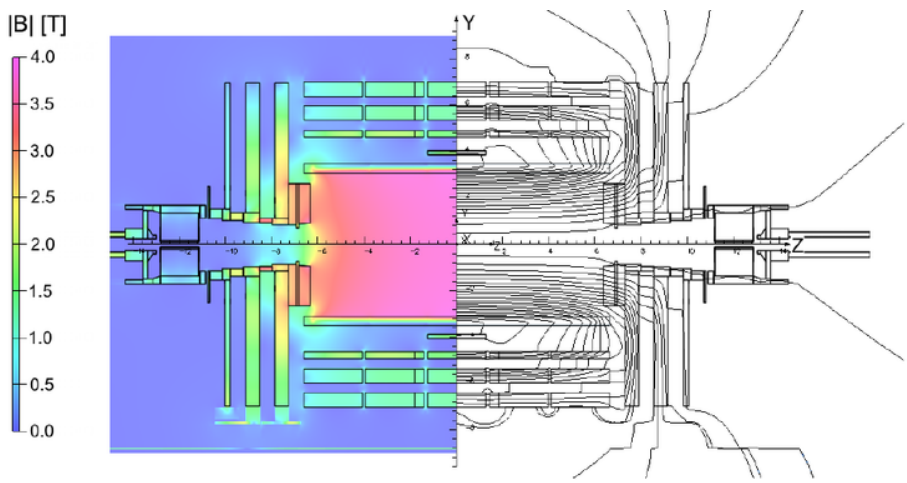


Figure 2.3: Value of $|B|$ (left) and field lines (right) are displayed on a longitudinal section of the CMS detector, demonstrating the central magnetic flux density of 3.8 T in the underground model [13].

2.2.2 Tracker

The CMS tracker is a critical component of the detector responsible for reconstructing the paths of charged particles and measuring their momenta with high precision. It consists of multiple layers of silicon detectors that enable precise position measurements and efficient particle tracking. The design of the CMS tracker allows for the reconstruction of charged particle trajectories in the region $|\eta| < 2.4$.

The tracker includes a silicon pixel tracker positioned closest to the beamline, which consists of a barrel and two pairs of endcap disks. The barrel contains three layers of silicon pixel detectors with radii of 4.4, 7.3, and 10.2 cm, while the endcap disks are located at distances of $|z|$ = 34.5 cm and $|z|$ = 46.5 cm from the interaction point. The pixel tracker provides three-dimensional position measurements, achieving a transverse coordinate position resolution of 10 μm and a longitudinal coordinate position resolution of 20-40 μm. In total, the barrel pixel tracker and endcap pixel tracker consist of 66 million pixels, covering an area of $1,\mathrm{m}^2$.

In addition to the pixel tracker, the CMS tracker comprises four subsystems of silicon microstrips. The tracker inner barrel (TIB) consists of four layers and covers the region from a radius of 20 cm to 55 cm. The tracker inner disks (TID) complement the TIB with three side disks, extending the coverage in the $|z|$ direction to ±118 cm. The outer tracker barrel (TOB) comprises six barrel layers, covering the region beyond 55 cm in radius, with the same $|z|$ coverage as the TID. Lastly, the tracker endcap (TEC) consists of nine disks on each side of the TIDs and TOB, providing coverage for the region $124 < |z| < 282$ cm.

The pitch, which is the distance between neighboring strips or pixels, varies in different parts of the tracker. For example, the pitch in the TIB ranges from 80 μm to 120 μm, while in the TID, it ranges from 81 μm to 158 μm. The TOB employs strips with pitches of either 183 μm or 122 μm, and the TEC has a position measurement resolution ranging from 18 μm to 47 μm.

Figure 2.4 presents a pictorial view of a tracker slice in the r-z plane, illustrating the various components of the tracker, including the pixel modules (red), single-sided strip modules (thin black lines), and strip stereo modules (thick blue lines).

The excellent spatial resolution and extensive coverage of the CMS tracker enable precise measurements of charged particle trajectories and momenta. For example, the single muon p_T resolution ranges from 0.65% to 1.5% at $\eta = 0$ and from 1% to 2% at $|\eta| = 1.6$ for muons with momenta ranging from 10 GeV to 100 GeV [15]. The high-performance tracking capabilities of the CMS tracker significantly contribute to the overall precision of the CMS detector.

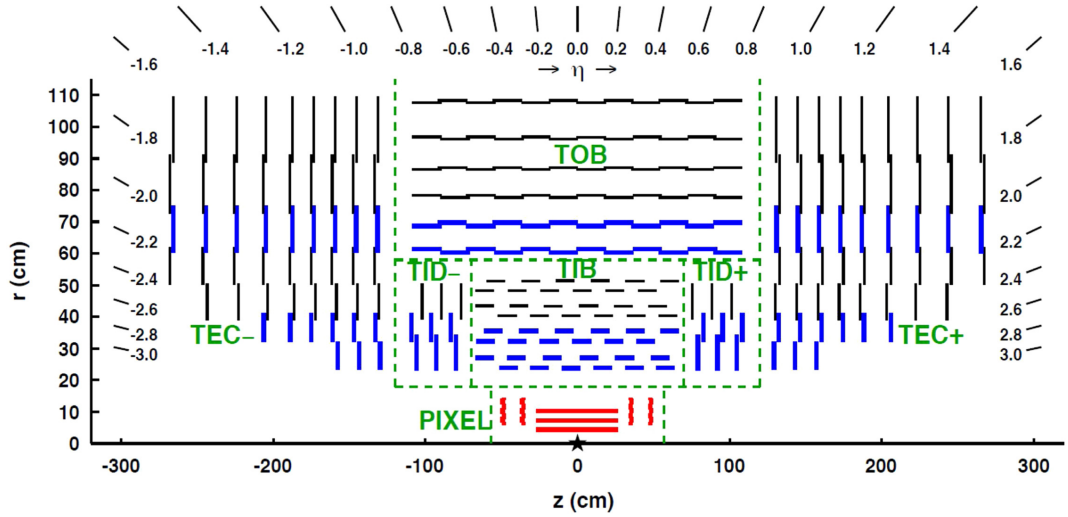


Figure 2.4: Pictorial view of a tracker slice in the r-z plane. Pixel modules are shown in red, single-sided strip modules are depicted as thin black lines, and strip stereo modules are shown in thick blue lines [14].

2.2.3 ECAL

The CMS electromagnetic calorimeter (ECAL) is a high-resolution calorimeter designed to precisely measure the energy of photons and electrons. It consists of 75,848 lead tungstate (PbWO4) crystals arranged in a barrel region and two endcap regions. The barrel region covers the range $|\eta| < 1.48$, while the endcap regions extend up to $|\eta| = 3.0$. The crystals in the barrel have a length of 23 cm and a front face of $2.2 \times 2.2, \text{cm}^2$, while those in the endcaps are 22 cm long with a front face of $2.86 \times 2.86, \text{cm}^2$. The crystals are organized into supermodules in the barrel and dees in the endcaps.

The ECAL also includes a preshower detector consisting of lead absorbers and silicon strip sensors. The preshower covers the region $1.65 \geq |\eta| \geq 2.6$ and helps improve the differentiation between the $\pi_0 \to \gamma\gamma$ process and the $h \to \gamma\gamma$ process.

Lead tungstate crystals were chosen for the ECAL due to their high density ($\rho = 8.28, \text{g/cm}^3$) and desirable properties such as a short radiation length (X_0) and a small Molière radius (R_M). The crystals absorb the energy of incoming electrons and photons, emitting light within 25 ns, with 80% of the energy emitted within that

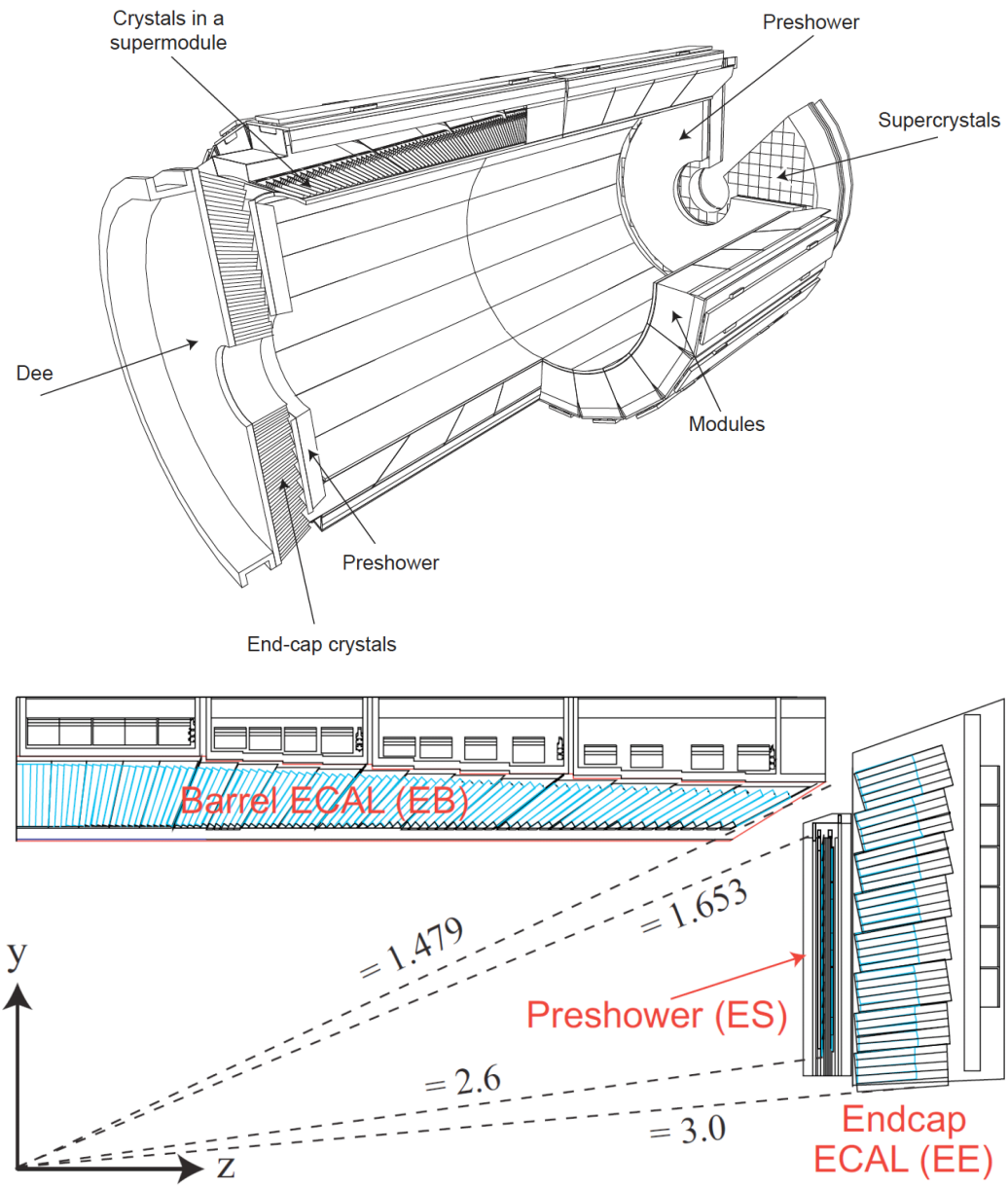


Figure 2.5: Schematic layout of the CMS electromagnetic calorimeter, presenting the arrangement of barrel supermodules, endcaps, and the preshower in front (top). Geometric view of one-quarter of the ECAL (bottom) [16].

time window.

The energy resolution of the ECAL has been measured and found to follow the equation:

$$\frac{\sigma_E}{E} = \frac{2.8\%}{\sqrt{E}} \oplus \frac{12\%}{E} \oplus 0.3\% \tag{2.2}$$

where σ_E is the energy resolution, E is the particle's energy, and the terms represent the stochastic term, the noise term, and the constant term, respectively.

The light produced in the scintillating crystals is collected by avalanche photodiodes (APDs) in the barrel and vacuum photodiodes (VPTs) in the endcaps. The signals from the APDs are pre-amplified and shaped by ASIC multi-gain pre-amplifier chips. The ECAL trigger towers are formed by grouping 5×5 crystals, and trigger primitives are generated based on the summed amplitudes of these crystal groups.

2.2.4 HCAL

The hadronic calorimeter (HCAL) in CMS surrounds the ECAL and is responsible for measuring the energy of hadrons and ensuring hermeticity for accurate missing transverse momentum ($\vec{p}_T^{\,\text{miss}}$) measurements. It consists of four subdetectors: the hadron barrel (HB), the hadron endcap (HE), the hadron outer (HO), and the hadron forward (HF) calorimeters.

The arrangement of the HCAL components is depicted in Figure 2.6, showing one-quarter of the CMS HCAL during the 2016 LHC operation. The HB covers the region up to $|\eta| < 1.3$, while the HE extends the coverage to $1.3 < |\eta| < 3.0$. The HF is responsible for the forward region, ranging from $3.0 < |\eta| < 5.0$. Steel is used as the absorber material in the very forward region, and quartz fibers serve as the active medium.

The HCAL employs alternating layers of plastic scintillators and non-magnetic brass as an absorbing material. These materials absorb the energy of incoming hadrons, causing the scintillators to emit light. The light signals are then read out by photomultiplier tubes (PMTs) and converted into electrical signals for further analysis.

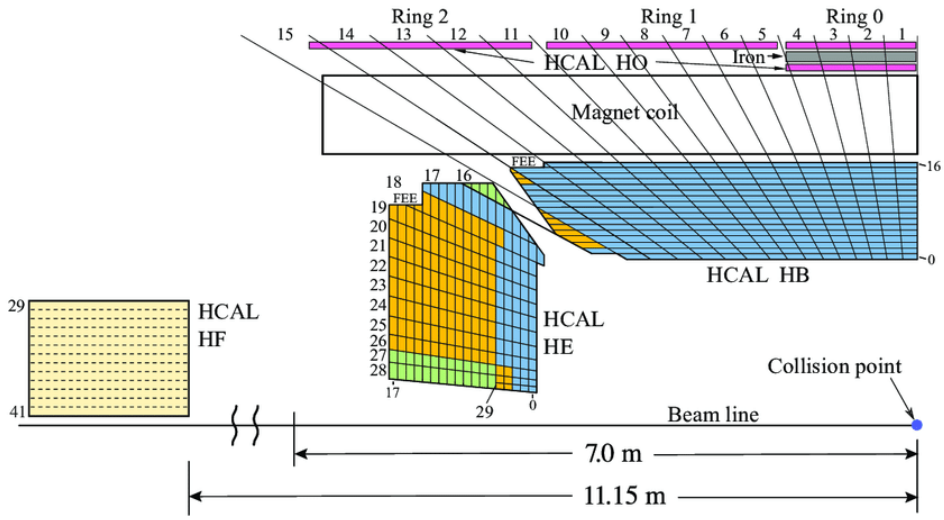


Figure 2.6: A schematic view of one-quarter of the CMS HCAL during 2016 LHC operation, showing the positions of its four major components: the hadron barrel (HB), the hadron endcap (HE), the hadron outer (HO), and the hadron forward (HF) calorimeters. The layers marked in blue are grouped as depth = 1, while the yellow, green, and magenta are combined as depths 2, 3, and 4, respectively. [17].

The energy resolution of the HCAL is given by:

$$\frac{\sigma_E}{E} = \frac{85\%}{\sqrt{E}} \oplus 7.4\% \tag{2.3}$$

Here, σ_E represents the energy resolution, E is the particle's energy, and the terms on the right side of the equation correspond to the stochastic term and the calibration uncertainties, respectively.

By combining the information from the HCAL with that from the ECAL, the CMS detector can provide a comprehensive energy measurement system for particles, enabling the reconstruction of jets, missing transverse energy, and other essential quantities utilized in particle physics analyses. The HCAL plays a crucial role in understanding the particles produced in high-energy collisions at the LHC.

2.2.5 Muon detector

The muon detector is a vital element of the CMS detector, meticulously engineered solely to measure muons. It plays a critical role in detecting, triggering, and precisely measuring muons' momentum and charge. Structured into a barrel region and two endcaps, it is interleaved with layers of the steel flux-return yoke.

The barrel and endcap regions utilize gas ionization particle detectors to identify muons. Technical abbreviations are explained throughout the text. Drift tubes (DTs) are utilized in the barrel section to cover a region with $|\eta| \geq 1.2$, which is divided into 12 segments around the azimuthal angle (ϕ). DT chambers are composed of drift cells that contain a gold-plated stainless steel anode wire operating at 3.6 kV and filled with a gas mixture of 8% Argon and 15% CO_2. The mixture is responsible for the drift of released electrons by incident muons within cells. The drift time in the cells is roughly 400 ns. Cathode plates situated on the edges of the cells function at ±1.8 kV. As soon as a muon travels through the gas volume, it emits electrons that travel towards the anode cable and result in a cascading effect in the area near it, where the electric field power intensifies. Four layers of identical cells create a super layer (SL), and three of these SLs establish a chamber. The placement of one layer is perpendicular to the others, enabling precise assessments in the $r-\phi$ and $r-z$ directions.

Cathode strip chambers (CSCs) are utilized in the endcap regions of the muon system, which covers a region of $0.9 \geq |\eta| \geq 2.4$ with an expected higher rate of muons and neutron background compared to the barrel. Cathode strip chambers (CSCs) are utilized in the endcap regions of the muon system, which covers a region of $0.9 \geq |\eta| \geq 2.4$ with an expected higher rate of muons and neutron background compared to the barrel. Consistent citation, grammar and spelling will be maintained throughout the writing, with an objective, value-neutral and formal language, employing passive tone and imp-personal construction. A clear, concise and logical structure will be used with a balanced view, avoiding opinion without clearly marking if it is included. The technical vocabulary will be used if the meaning is more precise. CSCs are preferred for the endcaps because they respond quicker than DTs due to the shorter drift path while also being able to tolerate the higher

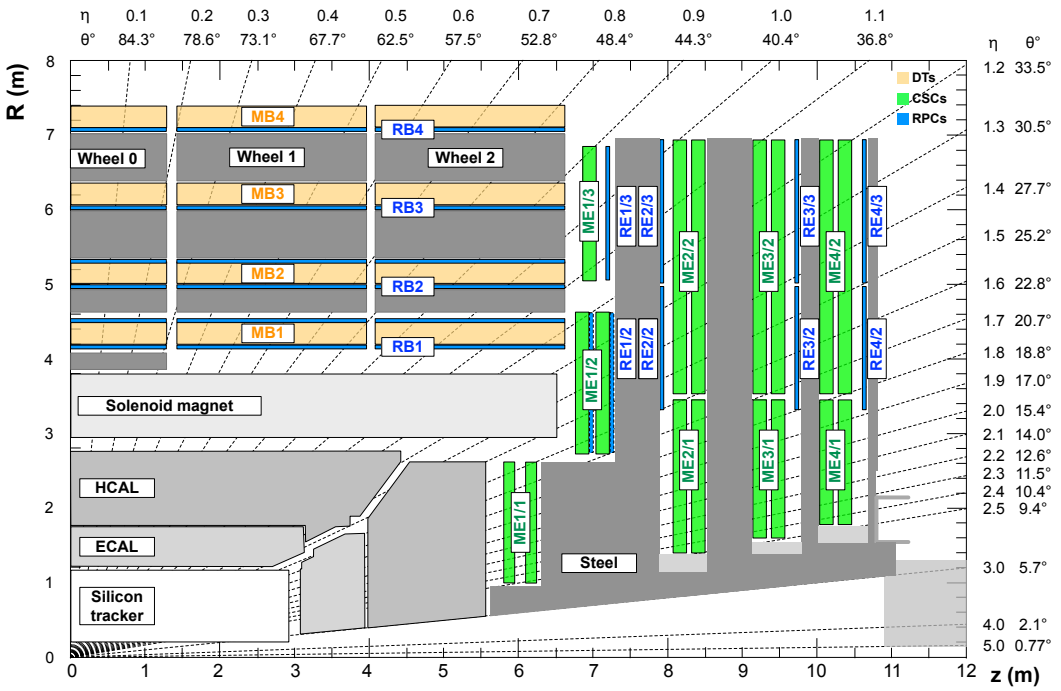


Figure 2.7: This figure illustrates a quadrant cross-section of the CMS detector, wherein the beam axis (z) runs horizontally and the radius (R) increases upwards from the lower left-hand interaction point. The diagram depicts the diverse muon stations and the steel flux-return disks (dark regions); the DTs are marked MB (”Muon Barrel”), and the CSCs are marked ME (”Muon Endcap”). Additionally, RPCs are mounted in the CMS barrel and endcaps, identified as RB and RE, respectively [18].

magnetic fields in the muon endcap regions. Technical abbreviations will be explained when first used. CSCs are preferred for the endcaps because they respond quicker than DTs due to the shorter drift path while also being able to tolerate the higher magnetic fields in the muon endcap regions. Each endcap of the detector is composed of four rings of chambers that are attached to the steel disks on the endcap's surface. One CSC chamber comprises six layers outfitted with 80 cathode strips placed radially outward and anode wires that run perpendicularly to the strips, which allow for position measurements in the $r-\phi$ plane. The chambers are filled with a gas mixture containing 50% CO2, 40% Argon, and 10% CF4. Depending on the ring, the anode wires operate at voltages ranging from 2.9 kV to 3.6 kV.

In the barrel region, Resistive Plate Chambers (RPCs) are positioned between the layers of DTs. RPCs offer swift and autonomous triggering at reduced transverse momentum (p_T) thresholds in the range $|\eta| \geq 1.6$. RPCs are double-gap chambers, each containing resistive Bakelite plates separated by a gas-filled volume of the same thickness. When a charged particle passes through an RPC, the gas in the gap volumes gets ionized, leading to an avalanche generated by the high electric field (9.6 kV). The image induced on the readout strips allows for the identification and triggering of muons.

The muon detector in CMS is composed of DTs, CSCs, and RPCs, which allow for the precise identification, triggering, and momentum measurement of muons across a wide range of pseudorapidity (η) values. This information is crucial for numerous physics analyses, such as the search for new particles and investigation into processes related to muons. The CMS detector's overall performance and capabilities rely heavily on the muon detector.

2.2.6 Trigger System

The CMS trigger system is an essential component for handling the high collision rate produced by the LHC and selecting collisions of interest for physics analysis. With a collision rate of approximately 40 MHz and limited data storage capability of about 1 kHz, the trigger system plays a crucial role in filtering and retaining the most relevant events.

The trigger system consists of two stages: the Level-1 trigger (L1) and the High-Level Trigger (HLT). The L1 trigger is implemented in hardware and operates on low-resolution information from the muon chambers and calorimeters, excluding the tracker data. The purpose of the L1 trigger is to quickly identify and select events with specific signatures, such as muons or high-energy deposits in the calorimeters. It consists of local, regional, and global components. The local trigger identifies object seeds within each sub-detector, which are then combined in the regional trigger to construct L1 objects such as muons, electrons, and jets. Finally, the global trigger applies programmable trigger requirements to select or reject events based on the L1 objects.

The calorimeter triggers derive their seeds from calorimetric towers, which provide information about transverse energy and the associated bunch crossing. The muon triggers utilize data from the muon sub-detectors, generating seeds based on individual chambers. The track information reconstructed in the muon system is combined and passed to the global trigger.

The HLT, implemented in software, is the second stage of the trigger system and operates on the complete detector information. It performs a more detailed reconstruction similar to the offline analysis but optimized to reduce the computing time per event. The HLT reconstruction consists of two stages: L2 and L3. In the L2 stage, the calorimeter and muon system data from the global trigger are used to reconstruct particles with higher precision than the L1 reconstruction. The selected events from L2 are then passed to the L3 stage, where the L2 particles are combined with information from the tracking detector to fully reconstruct tracks and vertices. The processing time for an event in the HLT is typically around 100 ms.

The L3 trigger applies stricter selection criteria to further refine the event selection. Events that pass the L3 trigger are categorized into different HLT paths, which are designed to select events with specific physical properties. Events classified in at least one trigger path are permanently stored for further analysis. These events are organized into non-exclusive datasets, grouping events with similar physical characteristics, such as the presence of a single muon or an electron-muon pair. The stored events serve as the basis for subsequent physics analysis.

Chapter 3

Supersymmetric particles search in CMS experiment

3.1 Motivation of analysis

The CERN LHC has provided sufficient data to conduct a large variety of searches for physics BSM. As explained in Chapter 2, supersymmetry is among the best-motivated candidates for new physics and predicts the existence of supersymmetric partners for each of the SM particles. Scenarios with non-degenerate supersymmetric particle spectra, with cross sections as low as ≈ 1 fb, have been explored in many final states. however, as yet, no traces of new physics have been found. Recently, the focus of searches has turned towards natural SUSY, in which the Higgs boson mass can be stabilized without excessive fine tuning. Natural SUSY requires the existence of a light top squark, $\tilde{t}_1$, and a somewhat light gluino, $\tilde{g}$, while accommodating mass scales for other supersymmetric particles that are beyond the direct reach of current LHC data. The possibility that the top squark could be light has motivated several searches by the CMS and ATLAS collaborations for the direct production of top squarks. The sensitivity of many of these searches, however, diminishes when the mass of the top squark approaches that of the lightest SUSY particle (LSP), assumed in the remainder of this thesis to be the lightest neutralino, $\tilde{\chi}_1^0$. Searches looking specifically for $\tilde{t}_1 \to t\tilde{\chi}_1^0$ also become less sensitive when the

mass difference, δm, between the top squark and the LSP is comparable to the top quark mass, m_t.

3.1.1 Target scenario

Final states with large mass differences and boosted decay products can occur in a wide variety of SUSY scenarios, in particular, motivated by Naturalness, which predicts gluinos and 3rd generation squarks at the O(TeV) scale. Figure 3.1 shows Feynman diagrams for several representative cases featuring heavy gluino, top squark, and electroweak gaugino production and subsequent decay to a much lighter stable lightest SUSY particle (LSP), in models with conserved R-parity (RPC). These decays, occurring directly, or via a chain, produce boosted heavy SM W/Z/Higgs bosons or top quarks. The final state also consists of lighter SM leptons and quarks and a moderate to large amount of missing transverse energy from the stable lightest neutralino. Another notable case is scenarios with R-parity violation (RPV), where the LSP is non-stable and decays to SM particles. In this case, the boosted object would be a high momentum RPV LSP significantly heavier than its SM decay products. Figure 3.2 shows two examples among many possible RPV configurations; for gluino pair production decaying to RPV top squarks and bottom quark pair production decaying to RPV neutralinos. In both cases, the RPV LSP decays into one or more quarks and leptons. As in the RPC case, boosted RPV final states additionally consist of lighter SM leptons and quarks. However, the missing transverse energy is low-to-moderate due to the absence of a neutral, stable heavy sparticle, and constitutes only SM neutrinos.

To be sensitive to a wide range of models, the analysis is designed to be inclusive and explores hadronic or leptonic final states with boosted hadronic W/Z/Higgs bosons, boosted hadronic top quarks or boosted leptonic jets targeting RPV models. The signal is discriminated from the SM backgrounds by the razor kinematic variables. The final state is divided into many bins based on boosted object, jet, b-quark jet, and lepton multiplicities to enhance sensitivity to different models. This study is an extension of an earlier study based on 36 fb^{-1} focusing only on final states with boosted hadronic W bosons and top quarks [19].

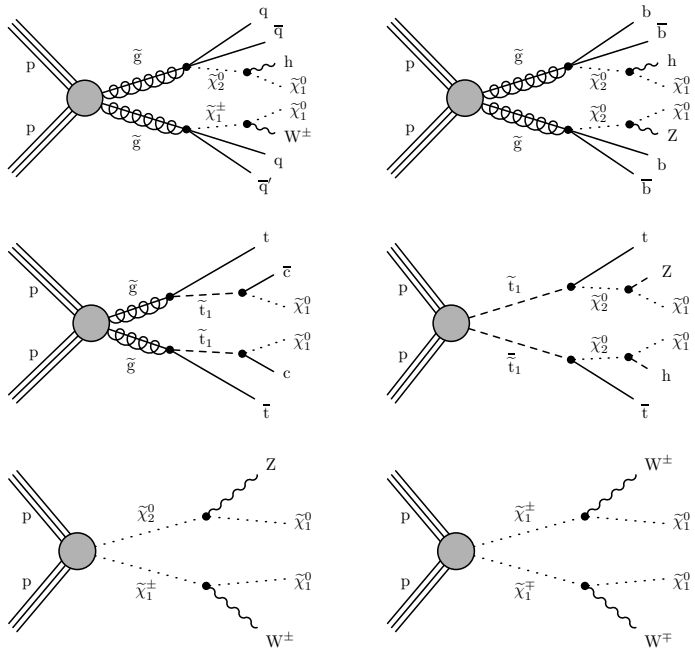


Figure 3.1: R-parity conserving signal models considered in this analysis: Gluino pair production T5qqqqWH (top left), T5bbbbZH (top right) and T5ttcc (middle left); top squark pair production T6ttZH (middle right); chargino-neutralino production TChiWZ (bottom left) and chargino-neutralino production TChiWW (bottom right).

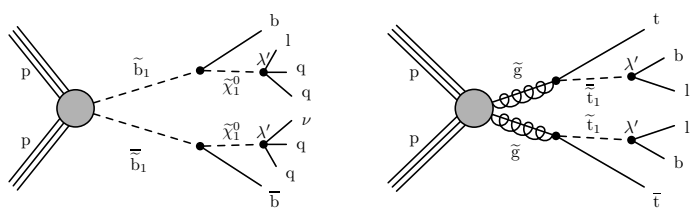


Figure 3.2: R-parity violating signal models considered in this analysis. Bottom squark pair production decaying to RPV neutralino, R2bbqqlv (left); gluino pair production decaying to RPV top squark, R5ttbl (right).

3.2 Data and Monte Carlo samples

This analysis is based on 138 fb^{-1} of 13 TeV pp collision data collected by the CMS detector in 2016, 2017 and 2018. Several datasets triggering different objects were used for different purposes, such as the selection of signal, control, and validation regions. The classes of datasets based on the objects triggered, and their usage are shown in Table 3.1. The full list of dataset names are given in Tables 3.2. We use NANOAODv9 in this analysis.

Table 3.1: Dataset classes with corresponding triggers and their usage purposes in the analysis

Dataset / trigger	2016	2017	2018	Purpose
JetHT / Hadronic	✓	✓	✓	All signal regions Hadronic control and validation regions Fake rate regions Lepton, photon trigger efficiency calculation
HTMHT / Hadronic, MET	✓	✓	–	All regions Hadronic control and validation regions Fake rate regions Lepton, photon trigger efficiency calculation
MET / MET	✓	✓	✓	All signal regions Hadronic control and validation regions Fake rate regions Lepton, photon trigger efficiency calculation
SingleMuon / SingleMuon	✓	✓	✓	Hadronic trigger efficiency calculation Some control regions with leptons
SingleElectron / SingleElectron	✓	✓	–	Hadronic trigger efficiency calculation Some control regions with leptons
SinglePhoton / SinglePhoton	✓	✓	–	Hadronic trigger efficiency calculation Photon control regions
EGamma / SingleElectron, SinglePhoton	–	–	✓	Hadronic trigger efficiency calculation Some control regions with leptons

Table 3.2: Collision datasets

Sample
/HTMHT/Run201*_MiniAODv2_NanoAODv9-v*/NANOAOD
/JetHT/Run201*_MiniAODv2_NanoAODv9-v*/NANOAOD
/MET/Run201*_MiniAODv2_NanoAODv9-v*/NANOAOD
/SingleElectron/Run201*_MiniAODv2_NanoAODv9-v*/NANOAOD
/SingleMuon/Run201*_MiniAODv2_NanoAODv9-v*/NANOAOD
/SinglePhoton/Run201*_MiniAODv2_NanoAODv9-v*/NANOAOD
/EGamma/Run201*_MiniAODv2_NanoAODv9-v*/NANOAOD

We also use Monte Carlo (MC) samples generated centrally by the CMS collaboration to study the characteristics of various background and signal processes which was followed by a reconstruction using `CMSSW_10_6_27`.

The CMS experiment utilizes a specialized software package called CMSSW to generate Monte Carlo simulation events. These events are essential for understanding the behavior of particles in the CMS detector. The process involves several steps, each contributing to the overall simulation, reconstruction, and analysis framework of CMS.

The primary physics processes are generated using programs like pythia and MadGraph. These programs produce information about the generated particles, which is then converted into the standard HepMC format. The detector simulation step, which utilizes the GEANT4 software, employs physics models known as ”physics lists” to simulate various types of particle interactions with detector materials. This step involves propagating particles through the detector based on these models.

Following the detector simulation, a dedicated digitization process is applied to simulate the response of the detector electronics. This process generates output signals that can be further processed. Subsequently, various reconstruction algorithms are applied to these signals to reconstruct the properties of the particles and their interactions.

Among the different steps involved in producing Monte Carlo samples, the detector simulation step is the most computationally expensive, consuming 40% of the total computing budget of CMS. The remaining steps, including event generation,

digitization, reconstruction, and analysis, collectively consume 45%, with reconstruction being the largest contributor.

Within the detector simulation step, evaluating the geometry and propagating the magnetic field accounts for 60% of the CPU usage. The electromagnetic (EM) physics models contribute 15%, while hadronic physics accounts for 10%. The remaining 15% is utilized by other components, including CMS-specific operations.

Background samples are listed in Tables 3.3, 3.4, and 3.5. For the background events, the effect of the CMS detector was simulated with a GEANT4 based full simulation.

To address the need for faster simulations, a framework called Fast Simulation has been developed and integrated into the CMS software. This Fast Simulation reproduces data samples in the same format as the Full Simulation and Reconstruction chain, which is based on GEANT4. Therefore, the output from the Fast Simulation can be seamlessly used in the analysis pipeline alongside other samples.

The Fast Simulation offers significant speed improvements compared to the Full Simulation, with event production rates being around 100 times faster. Despite the faster execution, the Fast Simulation maintains comparable accuracy for most physics objects considered in analyses. The Fast Simulation is particularly well-suited for physics analyses that would require extensive computational resources to perform using the Full Simulation and Reconstruction. Examples include analyses involving the exploration of extended parameter spaces of physics models like SUSY, or studies where background samples from high cross-section processes need to be selected based on final reconstructed objects.

The Fast Simulation is an object-oriented subsystem integrated into the general CMS C++ based software. It leverages many of the same modules as the Full Simulation and Reconstruction, with the reconstruction step accounting for most of the time spent in event production. The simulation part itself is typically 500-1000 times faster than the corresponding Full Simulation, though the exact speed improvement depends on the specific processes being simulated.

Signal samples are listed in Table 3.6. The Fast simulation was used for the signal events.

Table 3.3: List of SM MC samples used in the analysis - part1

Dataset	cross section [pb]
/DYJetsToLL_M-4to50_HT-100to200_TuneCP5_13TeV-madgraphMLM-pythia8/*-v2/NANOAODSIM	204.0
/DYJetsToLL_M-4to50_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8/*-v1/NANOAODSIM	45.743700
/DYJetsToLL_M-4to50_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8/*-v1/NANOAODSIM	45.743700
/DYJetsToLL_M-4to50_HT-400to600_TuneCP5_13TeV-madgraphMLM-pythia8/*-v2/NANOAODSIM	5.697
/DYJetsToLL_M-4to50_HT-600toInf_TuneCP5_13TeV-madgraphMLM-pythia8/*-v2/NANOAODSIM	1.85
/DYJetsToLL_M-4to50_HT-70to100_TuneCP5_13TeV-madgraphMLM-pythia8/*-v2/NANOAODSIM	145.50
/DYJetsToLL_M-50_HT-100to200_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/*-v2/NANOAODSIM	160.7
/DYJetsToLL_M-50_HT-1200to2500_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/*-v2/NANOAODSIM	0.1937
/DYJetsToLL_M-50_HT-200to400_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/*-v2/NANOAODSIM	48.63
/DYJetsToLL_M-50_HT-2500toInf_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/*-v2/NANOAODSIM	0.003514
/DYJetsToLL_M-50_HT-400to600_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/*-v2/NANOAODSIM	6.993
/DYJetsToLL_M-50_HT-600to800_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/*-v2/NANOAODSIM	1.761
/DYJetsToLL_M-50_HT-70to100_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/*-v2/NANOAODSIM	146.5
/DYJetsToLL_M-50_HT-800to1200_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/*-v2/NANOAODSIM	0.8021
/GJets_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8/*-v1/NANOAODSIM	8644.000000
/GJets_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8/*-v1/NANOAODSIM	2183.000000
/GJets_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8/*-v1/NANOAODSIM	260.200000
/GJets_HT-40To100_TuneCP5_13TeV-madgraphMLM-pythia8/*-v1/NANOAODSIM	18540.000000
/GJets_HT-600ToInf_TuneCP5_13TeV-madgraphMLM-pythia8/*-v1/NANOAODSIM	93.380000
/QCD_HT1000to1500_TuneCP5_PSWeights_13TeV-madgraph-pythia8/*-v1/NANOAODSIM	1127.000000
/QCD_HT100to200_TuneCP5_PSWeights_13TeV-madgraph-pythia8/*-v1/NANOAODSIM	23590000.000000
/QCD_HT1500to2000_TuneCP5_PSWeights_13TeV-madgraph-pythia8/*-v1/NANOAODSIM	119.900000
/QCD_HT2000toInf_TuneCP5_PSWeights_13TeV-madgraph-pythia8/*-v1/NANOAODSIM	21.980000
/QCD_HT200to300_TuneCP5_PSWeights_13TeV-madgraph-pythia8/*-v1/NANOAODSIM	1555000.000000
/QCD_HT300to500_TuneCP5_PSWeights_13TeV-madgraph-pythia8/*-v1/NANOAODSIM	324500.000000
/QCD_HT500to700_TuneCP5_PSWeights_13TeV-madgraph-pythia8/*-v1/NANOAODSIM	30310.000000
/QCD_HT50to100_TuneCP5_PSWeights_13TeV-madgraph-pythia8/*-v1/NANOAODSIM	187300000.000000
/QCD_HT700to1000_TuneCP5_PSWeights_13TeV-madgraph-pythia8/*-v1/NANOAODSIM	6444.000000
/ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-amcatnlo-pythia8/*-v1/NANOAODSIM	3.549000
/ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/*-v1/NANOAODSIM	69.09
/ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/*-v1/NANOAODSIM	115.3
/ST_tW_antitop_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8/*-v1/NANOAODSIM	32.510000
/ST_tW_top_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8/*-v1/NANOAODSIM	32.450000
/VBFHToBB_M-125_TuneCP5_13TeV-powheg-pythia8/*-v2/NANOAODSIM	3.873000
/VBFHToBB_M-125_TuneCP5_13TeV-powheg-pythia8/*-v2/NANOAODSIM	3.873000
/VBFHToWWTo2L2Nu_M-125_TuneCP5_13TeV-powheg-jhugen727-pythia8/*-v2/NANOAODSIM	3.892000

Table 3.4: List of SM MC samples used in the analysis - part2

Dataset	cross section [pb]
/WJetsToLNu_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8/*-v1/NANOAODSIM	1256.000000
/WJetsToLNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8/*-v1/NANOAODSIM	1.160000
/WJetsToLNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8/*-v1/NANOAODSIM	335.5
/WJetsToLNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8/*-v2/NANOAODSIM	0.008001
/WJetsToLNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8/*-v1/NANOAODSIM	45.25
/WJetsToLNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8/*-v1/NANOAODSIM	10.97
/WJetsToLNu_HT-70To100_TuneCP5_13TeV-madgraphMLM-pythia8/*-v1/NANOAODSIM	1264.0
/WJetsToLNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8/*-v1/NANOAODSIM	4.933
/WJetsToQQ_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8/*-v2/NANOAODSIM	2549.0
/WJetsToQQ_HT-400to600_TuneCP5_13TeV-madgraphMLM-pythia8/*-v2/NANOAODSIM	276.5
/WJetsToQQ_HT-600to800_TuneCP5_13TeV-madgraphMLM-pythia8/*-v2/NANOAODSIM	59.25
/WJetsToQQ_HT-800toInf_TuneCP5_13TeV-madgraphMLM-pythia8/*-v2/NANOAODSIM	28.75
/WWTo1L1Nu2Q_4f_TuneCP5_13TeV-amcatnloFXFX-pythia8/*-v1/NANOAODSIM	80.74000
/WWTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/*-v1/NANOAODSIM	11.09
/WWTo4Q_4f_TuneCP5_13TeV-amcatnloFXFX-pythia8/*-v1/NANOAODSIM	51.723000
/WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8/*-v1/NANOAODSIM	0.2158
/WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8/*_ext1-v1/NANOAODSIM	0.2158
/WWZ_4F_TuneCP5_13TeV-amcatnlo-pythia8/*-v1/NANOAODSIM	0.1707
/WWZ_4F_TuneCP5_13TeV-amcatnlo-pythia8/*_ext1-v1/NANOAODSIM	0.1707
/WZTo1L1Nu2Q_4f_TuneCP5_13TeV-amcatnloFXFX-pythia8/*-v1/NANOAODSIM	10.710000
/WZTo1L3Nu_4f_TuneCP5_13TeV-amcatnloFXFX-pythia8/*-v1/NANOAODSIM	3.033000
/WZTo3LNu_TuneCP5_13TeV-amcatnloFXFX-pythia8/*-v1/NANOAODSIM	5.213
/WZZ_TuneCP5_13TeV-amcatnlo-pythia8/*-v1/NANOAODSIM	0.05709
/WZZ_TuneCP5_13TeV-amcatnlo-pythia8/*_ext1-v1/NANOAODSIM	0.05709
/WminusH_HToBB_WToLNu_M-125_TuneCP5_13TeV-powheg-pythia8/*-v1/NANOAODSIM	0.033700
/WminusH_HToBB_WToQQ_M-125_TuneCP5_13TeV-powheg-pythia8/*-v1/NANOAODSIM	0.034922
/WminusH_WToQQ_HToInvisible_M125_TuneCP5_13TeV_powheg_pythia8/*-v2/NANOAODSIM	0.3674
/WplusH_HToBB_WToLNu_M-125_TuneCP5_13TeV-powheg-pythia8/*-v1/NANOAODSIM	0.053100
/WplusH_HToBB_WToQQ_M-125_TuneCP5_13TeV-powheg-pythia8/*-v1/NANOAODSIM	0.055037
/WplusH_WToQQ_HToInvisible_M125_TuneCP5_13TeV_powheg_pythia8/*-v2/NANOAODSIM	0.5884
/ggZH_HToBB_ZToBB_M-125_TuneCP5_13TeV-powheg-pythia8/*-v2/NANOAODSIM	0.043200
/ggZH_HToBB_ZToLL_M-125_TuneCP5_13TeV-powheg-pythia8/*-v1/NANOAODSIM	0.007800
/ggZH_HToBB_ZToNuNu_M-125_TuneCP5_13TeV-powheg-pythia8/*-v1/NANOAODSIM	0.014400
/ggZH_HToBB_ZToQQ_M-125_TuneCP5_13TeV-powheg-pythia8/*-v1/NANOAODSIM	0.04319
/ggZH_ZToQQ_HToInvisible_M125_TuneCP5_13TeV_powheg_pythia8/*-v1/NANOAODSIM	0.04319
/ttHJetToNonbb_M125_TuneCP5_13TeV_amcatnloFXFX_madspin_pythia8/*-v1/NANOAODSIM	0.507
/ttHJetTobb_M125_TuneCP5_13TeV_amcatnloFXFX_madspin_pythia8/*-v1/NANOAODSIM	0.211800

Table 3.5: List of SM MC samples used in the analysis - part3

Dataset	cross section [pb]
/TTGJets_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8/*-v1/NANOAODSIM	3.757000
/TTHH_TuneCP5_13TeV-madgraph-pythia8/*-v2/NANOAODSIM	0.0006655
/TTTT_TuneCP5_13TeV-amcatnlo-pythia8/*-v2/NANOAODSIM	0.008213
/TTTW_TuneCP5_13TeV-madgraph-pythia8/*-v2/NANOAODSIM	0.0007314
/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/*-v1/NANOAODSIM	88.29
/TTToHadronic_TuneCP5_13TeV-powheg-pythia8/*-v1/NANOAODSIM	377.96
/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8/*-v1/NANOAODSIM	365.34
/TTWH_TuneCP5_13TeV-madgraph-pythia8/*-v2/NANOAODSIM	0.001141
/TTWJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8/*-v1/NANOAODSIM	0.216000
/TTWJetsToQQ_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8/*-v1/NANOAODSIM	0.437700
/TTWW_TuneCP5_13TeV-madgraph-pythia8/*-v1/NANOAODSIM	0.007003
/TTWZ_TuneCP5_13TeV-madgraph-pythia8/*-v1/NANOAODSIM	0.002453
/TTZH_TuneCP5_13TeV-madgraph-pythia8/*-v2/NANOAODSIM	No value
/TTZToLLNuNu_M-10_TuneCP5_13TeV-amcatnlo-pythia8/*-v1/NANOAODSIM	0.2432
/TTZToQQ_TuneCP5_13TeV-amcatnlo-pythia8/*-v1/NANOAODSIM	0.5104
/TTZZ_TuneCP5_13TeV-madgraph-pythia8/*-v1/NANOAODSIM	0.001386
/ZH_HToBB_ZToBB_M-125_TuneCP5_13TeV-powheg-pythia8/*-v2/NANOAODSIM	0.554000
/ZH_HToBB_ZToLL_M-125_TuneCP5_13TeV-powheg-pythia8/*-v1/NANOAODSIM	0.048600
/ZH_HToBB_ZToNuNu_M-125_TuneCP5_13TeV-powheg-pythia8/*-v1/NANOAODSIM	0.089100
/ZH_HToBB_ZToQQ_M-125_TuneCP5_13TeV-powheg-pythia8/*-v1/NANOAODSIM	0.309931
/ZJetsToNuNu_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8/*-v1/NANOAODSIM	267.0
/ZJetsToNuNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8/*-v2/NANOAODSIM	0.2497
/ZJetsToNuNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8/*-v1/NANOAODSIM	73.08
/ZJetsToNuNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8/*-v2/NANOAODSIM	0.005618
/ZJetsToNuNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8/*-v2/NANOAODSIM	9.904
/ZJetsToNuNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8/*-v2/NANOAODSIM	2.413
/ZJetsToNuNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8/*-v2/NANOAODSIM	1.071
/ZJetsToQQ_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8/*-v2/NANOAODSIM	1012.0
/ZJetsToQQ_HT-600to800_TuneCP5_13TeV-madgraphMLM-pythia8/*-v2/NANOAODSIM	25.34
/ZJetsToQQ_HT-800toInf_TuneCP5_13TeV-madgraphMLM-pythia8/*-v2/NANOAODSIM	190.2
/ZZTo2L2Nu_TuneCP5_13TeV_powheg_pythia8/*-v1/NANOAODSIM	0.6008
/ZZTo2Nu2Q_5f_TuneCP5_13TeV-amcatnloFXFX-pythia8/*-v1/NANOAODSIM	4.040000
/ZZTo4L_TuneCP5_13TeV_powheg_pythia8/*-v1/NANOAODSIM	1.325
/ZZTo4Q_5f_TuneCP5_13TeV-amcatnloFXFX-pythia8/*-v1/NANOAODSIM	6.842000
/ZZZ_TuneCP5_13TeV-amcatnlo-pythia8/*-v1/NANOAODSIM	0.01398
/ZZZ_TuneCP5_13TeV-amcatnlo-pythia8/*_ext1-v1/NANOAODSIM	0.01398

Table 3.6: MC samples for the signal events

Dataset
/SMS-T5bbbbZH_TuneCP5_13TeV-madgraphMLM-pythia8/*/NANOAODSIM
/SMS-T5qqqqWH_TuneCP5_13TeV-madgraphMLM-pythia8/*/NANOAODSIM
/SMS-T5ttcc_TuneCP5_13TeV-madgraphMLM-pythia8/*/NANOAODSIM
/SMS-T6ttZH_TuneCP5_13TeV-madgraphMLM-pythia8/*/NANOAODSIM
/SMS-TChiWWpm-mNLSP200To1500_mLSP0To600_TuneCP5_13TeV-madgraphMLM-pythia8/*/NANOAODSIM
/SMS-TChiWZ-mNLSP200To1500_mLSP0To600_TuneCP5_13TeV-madgraphMLM-pythia8/*/NANOAODSIM
/SMS-TChiZZ-mNLSP200To1500_mLSP0To600_TuneCP5_13TeV-madgraphMLM-pythia8/*/NANOAODSIM
/SMS-T5ttbl_TuneCP5_13TeV-madgraphMLM-pythia8/*/NANOAODSIM
/SMS-T2bbqqlv_TuneCP5_13TeV-madgraphMLM-pythia8/*/NANOAODSIM

3.3 Object definition

Below, we present detailed information about the objects and event weighting applied in this analysis.

Jets

Jets are reconstructed based on a clustering of the PF candidates using FastJet with the anti-k_T algorithm [20]. The analysis uses two types of jets, which are reconstructed with different size parameters. Small radius jets, referred to as AK4 jets, used in event selection and calculating the razor variables, are reconstructed with the size parameter $R = 0.4$. Large radius jets, referred to as AK8 jets used for identifying boosted objects, are reconstructed with the size parameter $R = 0.8$.

Charged PF candidates associated with vertices other than the primary vertex are disregarded in the clustering to reduce the effect of pileup. The jet 4-momenta are corrected for residual pileup effects arising mainly from neutral particles and for detector response effects using the standard level-1, 2, and 3 corrections [21]. Table 3.7 shows the kinematic and identification criteria applied to AK4 and AK8 jets.

Table 3.7: Jet selection

NanoAOD collections: Jet, FatJet		
Variable	AK4 value	AK8 value
pt	≥ 30	≥ 200
abs(eta)	< 2.4	< 2.4
JetID	> 1 (tight lepton veto)	> 1 (tight)

To account for changing run-period conditions for the jet energy measurements, separate corrections are applied for each run period. Version `Summer19UL16APV_V7`, `Summer19UL16_V7` is the set of jet energy corrections used for the data collected in 2016, version `Summer19UL17_V5`, `Summer19UL18_V5` is used for the data collected in 2017 and 2018. Jet 4-momenta in simulated samples are smeared so the average jet energy resolution better matches with that of the observed data. Version

`Summer20UL16APV_JRV3`, `Summer20UL16_JRV3` of the jet energy resolution measurement is used for 2016, and version `Summer19UL17_JRV2`, `Summer19UL18_JRV2` for 2017 and 2018.

b-jet tagging

Jets originating from b quarks are identified using the "Deep Jet" algorithm. We use the recommended loose and medium working points from the BTag POG [22], shown in Table 3.8. The loose working point is used as a veto to identify $t\bar{t}$ (and signal) depleted control regions, while the medium working point is used to define signal enriched regions to perform the analysis. Event selection regions are defined based on the number of b-tagged jets. For various regions with a boosted W, Z or Higgs boson, tagged as to be described in Section 3.3, b-tagged jets are required to be separated from the boosted W, Z or Higgs boson by an angular distance of $\Delta R > 0.4$. The isolation requirement serves to remove overlaps with boosted top quarks or Higgs bosons.

Table 3.8: b tagging definitions

Input collection : AK4 jets (Table 3.7)						
Working point	Variable	2016preVFP	2016postVFP	2017	2018	$\Delta R(b,W/Z/H)$
Medium	DeepFlavB	> 0.2598	> 0.2489	> 0.3040	> 0.2783	–
Medium iso	DeepFlavB	> 0.2598	> 0.2489	> 0.3040	> 0.2783	> 0.8
Loose	DeepFlavB	> 0.0508	> 0.0480	> 0.0490	> 0.0532	–
Loose iso	DeepFlavB	> 0.0508	> 0.0480	> 0.0490	> 0.0532	> 0.8

To account for the discrepancies between data and MC simulation in b-tagged jet distribution, we apply appropriate Data/MC scale factors to the simulation. These scale factors and their prescription have been provided by the BTag POG [23]. Whenever an explicit selection based on the number of b-tagged jets is made, the corresponding b-tag scale factors are applied to MC events.

Muons

Muons are reconstructed with the PF algorithm and identified using different criteria [24, 25]. The analysis uses three different muon selections. The first two, called "veto iso" and "select iso" are muons identified with the veto and medium working points from the Muon POG. These are also required to be isolated and pass mini isolation criteria. The third selection consists of non-isolated muons, where the muon is required to be in the vicinity of a jet. Non-isolated muons are used for defining boosted leptonic tops or boosted leptonic jets. The non-isolation requirement between a lepton and AK4 jet is defined as "$\min(\Delta R(\ell, j)) > 0.4$ OR $p_T^{rel}(\ell, j) > 15$", where $p_T^{rel}(\ell, j) = (|\vec{p}_\ell|^2 - \vec{p}_\ell \cdot \vec{p}_j)/|\vec{p}_j|^2$. Table 3.9 lists the definitions of veto iso, select iso, and non isolated muons.

Table 3.9: Muon definitions

Input NanoAOD collection: Muon			
Variable/Method	Veto iso	Select iso	Noniso
softId	= 1	–	= 1
mediumPromptId	–	= 1	–
pt	≥ 5	≥ 10	≥ 30
abs(eta)	< 2.4	< 2.4	< 2.4
abs(d0)	< 0.2	< 0.05	< 0.05
abs(dz)	< 0.5	< 0.1	< 0.1
sigma3D	–	–	< 4
miniPFReliso_all	< 0.4	< 0.2	–
"B2G 2D isolation"	–	–	$\min(\Delta R(\mu, j)) > 0.4$ or $p_T^{rel}(\mu, j) > 15$

Muon scale factors were calculated with *Tag-and-Probe* method for the isolated leptons. The framework and detailed instructions are provided by the Muon POG. Tag and probe trees were produced and analyzed for the following UL MiniAODv2 datasets:

- Data: SingleMuon
- MC nominal: DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8

- MC alternative: DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8

The procedure is defined as follows:

- Trigger: HLT_IsoMu24_v* / HLT_IsoMu27_v* (2016,2018/2017)
- Tag selection: Tight ID, $p_T \geq 26(29)$ GeV (for 2017), $|\eta| < 2.4$, Isolation $(\Delta\beta) \leq 0.2$, $d_{xy} < 0.2$, $d_z < 0.5$
- Pair definition: Z mass window: $70 < m_{\mu\mu} < 130$ GeV, pair multiplicity = 1
- Systematic uncertainties were calculated with the alternative signal model, tag isolation shifts, histogram binning shifts, fit window shifts
- Pileup reweighting was applied

Figure 3.3 shows the resulting muon scale factors.

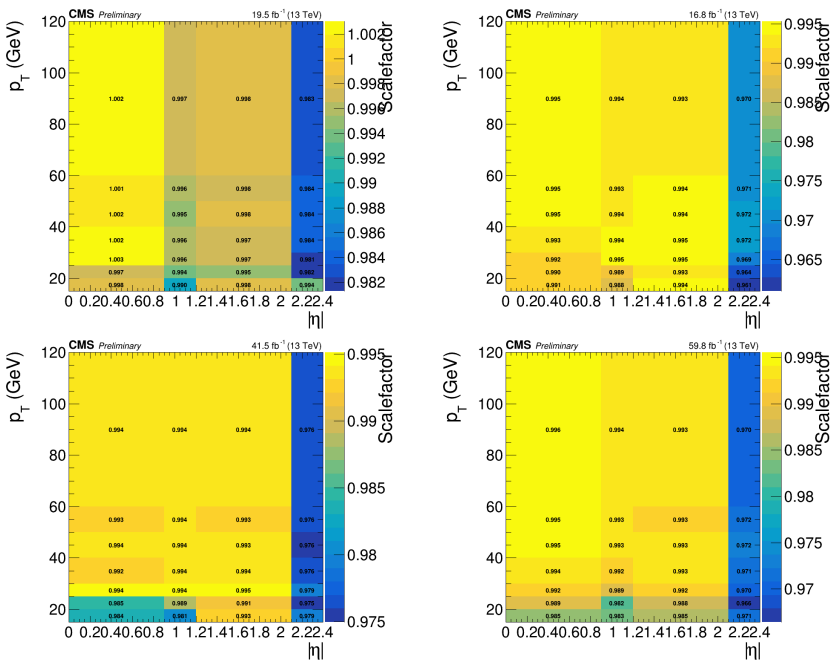


Figure 3.3: Muon scale factors for 2016preVFP (top left), 2016postVFP (top right). 2017 (bottom left) and 2018 (2018).

Electrons

Similar to muons, three electron selections are used in the analysis, as defined in Table 3.10. An MVA-based algorithm is used for electron identification.

Table 3.10: Electron definitions

Input NanoAOD collection: Electron			
Variable/Method	Veto iso	Select iso	Noniso
mvaFall17V2noIso_WPL	= 1	–	= 1
mvaFall17V2noIso_WP90	–	= 1	–
pt	≥ 5	≥ 10	≥ 30
abs(eta)	< 2.5	< 2.5	< 2.5
abs(eta)	–	![1.442 1.556]	![1.442 1.556]
abs(d0)	< 0.05	< 0.05	< 0.05
abs(dz)	< 0.1	< 0.1	< 0.1
miniPFReliso_all	< 0.4	< 0.1	–
"B2G 2D isolation"	–	–	$\min(\Delta R(e,j)) > 0.4$ or $p_T^{rel}(e,j) > 15$

Electron scale factors were calculated with *Tag-and-Probe* method for the isolated leptons. The framework and detailed instructions are provided by the EGamma POG. Tag and probe trees were produced and analyzed for the following UL MiniAODv2 datasets:

- Data: SingleElectron, EGamma
- MC nominal: DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8
- MC alternative: DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8

The procedure is defined as follows:

- Trigger: HLT_Ele27_WPTight_Gsf_v* / HLT_Ele32_WPTight_Gsf_L1DoubleEG_v* / HLT_Ele32_WPTight_Gsf_v* (2016/2017/2018)
- Tag selection: Tight cut-based ID, $p_T \geq 35$ GeV, $|\eta_{sc}| < 2.17$, opposite charge

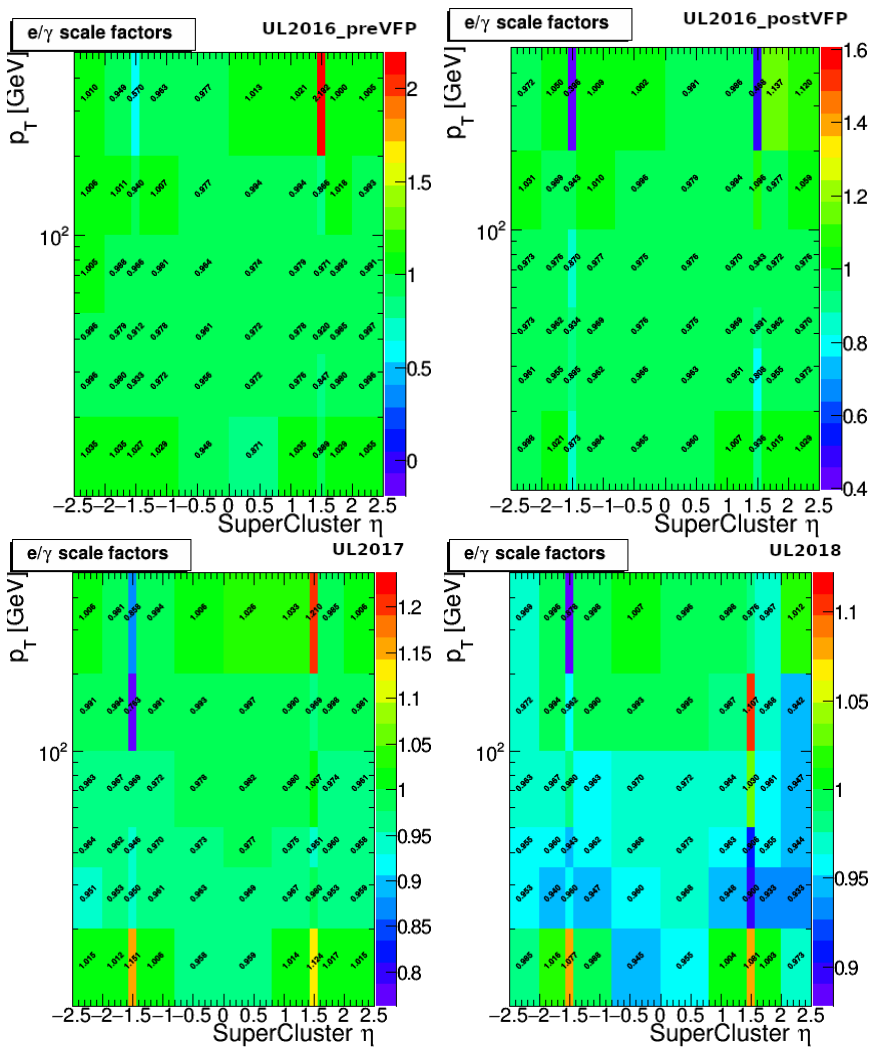


Figure 3.4: Electron scale factors for 2016preVFP (top left), 2016postVFP (top right). 2017 (bottom left) and 2018 (2018).

- Z mass window: 60-120 GeV for both Data and MC

Figure 3.4 shows the resulting electron scale factors.

Isolated tracks

We use the isolated track collection `IsoTrack` from NanoAODv9

Table 3.11: IsoTrack definition

NanoAOD collection: IsoTrack

Photons

Photons are used in a γ enriched region to estimate the $Z(\rightarrow \nu\nu)$ in the signal regions. Table 3.12 shows the photon selection and identification. The cut-based ID for photons is used to retain the capability of reverting a photon-ID variable, $\sigma_{i\eta i\eta}$ for the $Z(\rightarrow \nu\nu)$+jets background estimation procedure.

Table 3.12: Photon definition

NanoAOD collection: Photon	
Variable/Method	Value
pt	> 80
abs(eta)	< 2.5
cutBased	>= 1 (medium)
ele_veto	= 1
pixel_veto	= 1

Missing transverse energy

Missing transverse energy (E_T^{miss}) is calculated as the inverse vectorial sum of all PF candidate transverse momenta. No explicit selection is made on E_T^{miss}. However, it is correlated with the razor variable M_R^T introduced in Section A.1.

Boosted objects

The analysis makes use of both hadronic and leptonic boosted objects. Table 3.13 lists the identification criteria for these objects. The hadronic class includes hadronic boosted W/Z/Higgs bosons and top quarks, and are all based on AK8 jets defined in Section 3.3. ParticleNet tagging algorithm is used for identification. For boosted W/Z bosons and top quarks, no mass criterion is required, whereas for boosted Higgs bosons, a mass window of 90-140 GeV is selected. The p_T requirements are selected so that the mother particle's decay products are contained within an angular separation of 0.8. A novelty specific to this analysis is the use of leptonic boosted objects. These are defined by requiring the existence of a non-isolated lepton within an angular distance of $\Delta R < 0.8$ from an AK8 jet with $p_T > 300$ GeV and mass > 50 GeV. Apart from the explicit tagging of boosted objects, the analysis also uses mass-tagged AK8 jets, which are only required to pass an additional mass cut of $m \geq 50$ GeV, for the definition of various control and validation regions.

The boosted object discriminators, by default work against QCD multijet. Therefore, an AK8 jet can be classified to be consistent with more than one boosted particle. We resolve this overlap issue by checking that the object passes the tagger in the following order:

1. leptonic jet
2. hadronic Higgs tag
3. hadronic top tag
4. hadronic W OR Z tag

If an object is tagged both as a hadronic top and a hadronic Higgs boson, it is classified as a hadronic top quark.

Boosted object scale factors

Data / FullSim scale factors for hadronic boosted object taggers are provided centrally by the JetMET POG [26]. Scale factors for leptonic boosted object tagging and boosted object fake rates are calculated privately for this analysis.

Object	p_T	Mass	WP	non-iso lepton
Hadronic boosted objects: (input: AK8jets)				
top	≥ 300	50-220	ParticleNet WP3 (1%)	0
W	≥ 200	50-220	ParticleNet WP3 (1%)	0
Z	≥ 200	70-105	ParticleNet ZvsQCD>0.95	0
h	≥ 300	90-140	ParticleNet (MD) LP WP	0
Leptonic boosted objects: (input: AK8jets, noniso leptons)				
lepJet	≥ 300	≥ 50	–	≥1 within ΔR <0.8
Mass-tagged object: (input: AK8jets)				
m50	≥ 200	≥ 50	–	–

Table 3.13: Identification criteria for hadronic and boosted objects and mass-tagging.

Data and MC efficiencies and data / MC scale factors are derived in a fake enriched region defined by the following selection criteria:

- Hadronic triggers.
- = 1 or > 1 AK8 jet with $p_T > 200$ GeV.
- veto $e/\mu/\tau$
- veto b-tagged jets (not applied for Higgs tagger fake rate SFs)
- $\Delta\phi_{mega\,jets} < 2.8$
- $M_R > 800$, $R^2 > 0.08$

This selection is applied to both the data and MC, and efficiencies are calculated versus AK8 jet p_T. The denominator of the efficiencies is taken as the hardest AK8 jet with $p_T > 200$ GeV, while the numerator is taken as the hardest AK8 jet with $p_T > 200$ GeV that also passes the boosted object tagging criteria. The data and MC efficiencies are then divided to obtain the scale factors. All scale factors are calculated in the barrel and endcap regions separately. Figures 3.5, 3.6 and 3.7 show the fake rates and scale factors for hadronic W/Z tagging, hadronic top tagging, and

leptonic top tagging respectively. For the leptonic objects, non-isolated lepton SFs are used alongside AK8 jet corrections.

The efficiency scale factors are applied to MC when a tagged object is identified to be real by matching to a truth particle with $\Delta R(true, AK8Jet) < 0.8$. The fake rate scale factors are applied when the tagged object is not identified as a heavy boosted particle by matching.

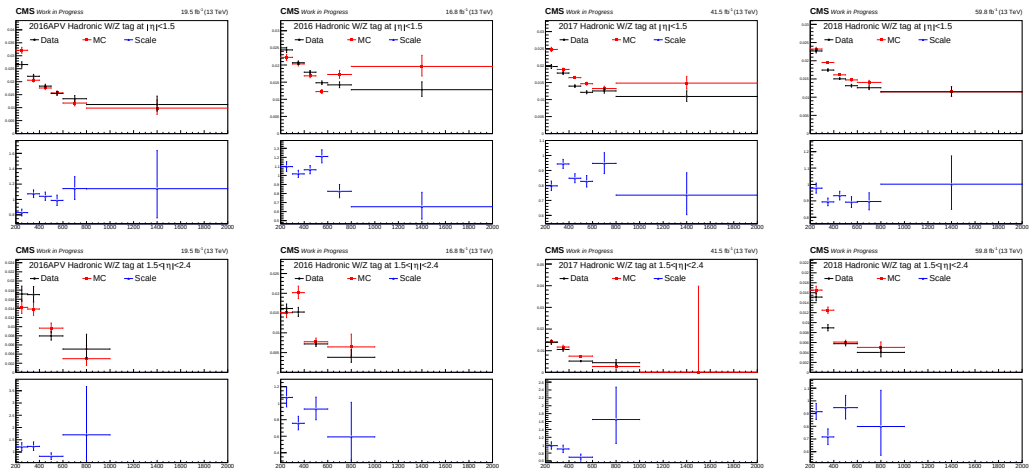


Figure 3.5: Hadronic W/Z tag data and MC fake rates and data/MC scale factors versus AK8 jet p_T for barrel (top) and endcap (bottom) for 2016APV, 2016 (left), 2017 (middle) and 2018 (right) data taking periods.

FullSim and FastSim efficiencies and FullSim / FastSim scale factors are derived in a preselection region defined by the following selection criteria:

- Hadronic triggers.
- = 1 or > 1 AK8 jet with $p_T > 200$ GeV.
- = 2 or > 2 AK4 jet.
- $M_R > 800$, $R^2 > 0.08$

This selection is applied to both the FullSim and FastSim, and efficiencies are calculated versus AK8 jet p_T. The denominator of the efficiencies is taken as the truth boosted objects, while the numerator is taken as the truth boosted objects that also

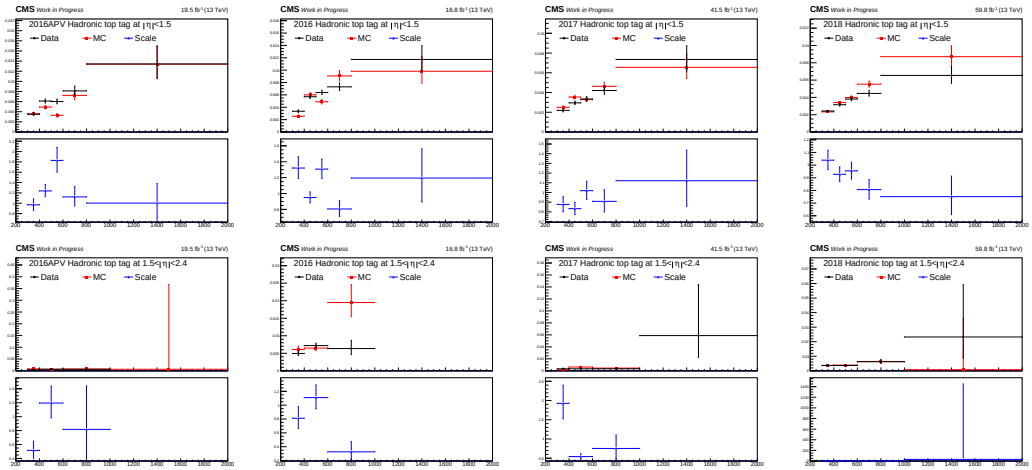


Figure 3.6: Hadronic top tag data and MC fake rates and data/MC scale factors versus AK8 jet p_T for barrel (top) and endcap (bottom) for 2016APV, 2016 (left), 2017 (middle) and 2018 (right) data taking periods.

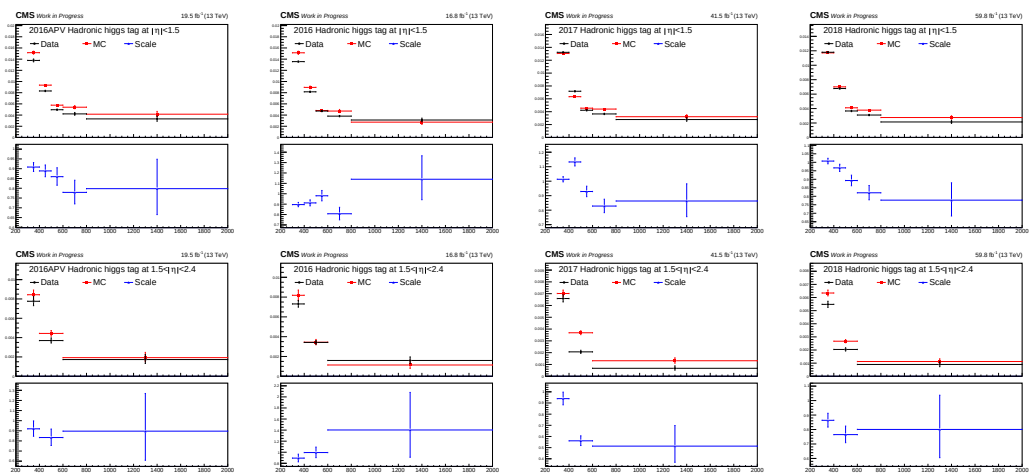


Figure 3.7: Hadronic higgs tag data and MC fake rates and data/MC scale factors versus AK8 jet p_T for barrel (top) and endcap (bottom) for 2016APV, 2016 (left), 2017 (middle) and 2018 (right) data taking periods.

pass the boosted object tagging criteria. The FullSim and FastSim efficiencies are then divided to obtain the scale factors. All scale factors are calculated in the barrel and endcap regions separately. Figures 3.8, 3.9, and 3.10 show the efficiencies and scale factors for hadronic W/Z tagging, hadronic top tagging, and hadronic higgs tagging respectively.

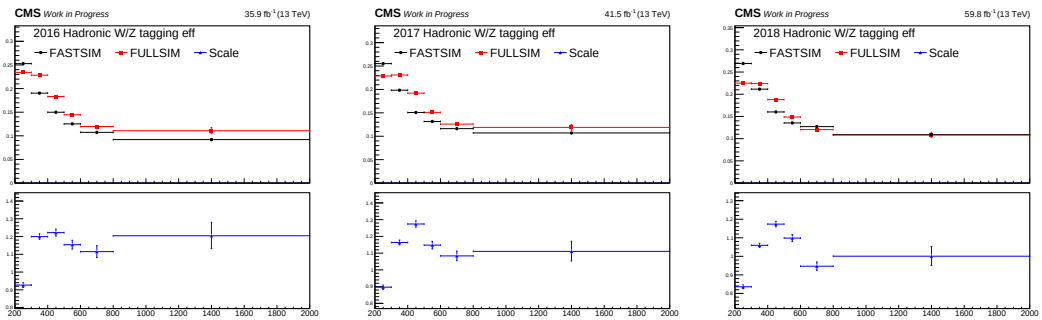


Figure 3.8: Truth matched hadronic W/Z tag fullsim and fastsim efficiencies and fullsim/fastsim scale factors versus AK8 jet p_T for 2016 (left), 2017 (middle), and 2018 (right) data taking periods.

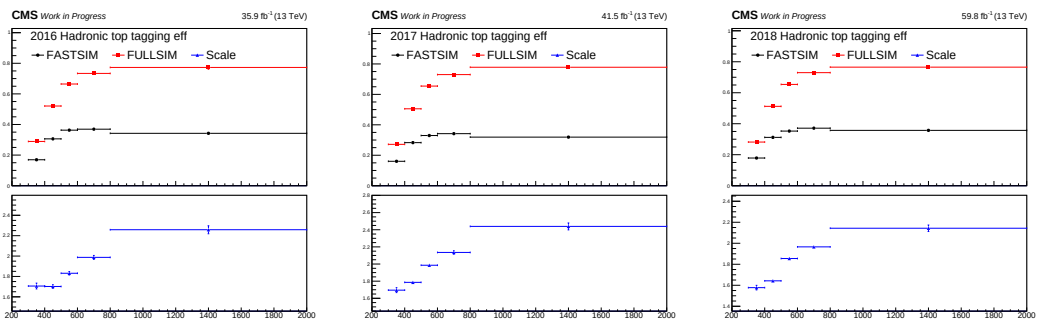


Figure 3.9: Truth matched hadronic top tag fullsim and fastsim efficiencies and fullsim/fastsim scale factors versus AK8 jet p_T for 2016 (left), 2017 (middle), and 2018 (right) data taking periods.

Events selected by this analysis usually have multiple real or fake boosted objects, therefore multiple scale factors need to considered per event. As in the case

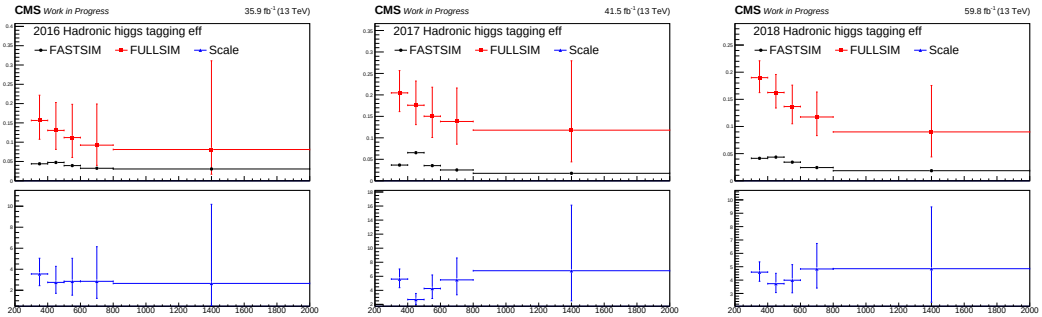


Figure 3.10: Truth matched hadronic higgs tag fullsim and fastsim efficiencies and fullsim/fastsim scale factors versus AK8 jet p_T for 2016 (left), 2017 (middle) and 2018 (right) data taking periods.

of b-tagging, a single event weight is computed taking into account all boosted object efficiencies and their scale factors. The event weight is obtained by dividing the SF-weighted sum of all boosted object efficiencies by the SF-unweighted sum of efficiencies. The uncertainties are taken into account by recalculating the event weight, this time using the up or down varied values of the SFs.

3.3.1 Trigger

The analysis uses different types of triggers that constitute different datasets for different purposes, such as event selection for the signal, control, validation and fake rate estimation regions, as was summarized in Table 3.1. The lists of specific triggers used based on primary datasets with their varying thresholds per data taking period are given in Table 3.14. Since the analysis targets events with boosted objects, triggers based on high jet p_T and high hadronic transverse energy H_T are used for selecting signal-like events. Triggers based on missing transverse energy (E_T^{miss}) and missing hadronic transverse momentum $H_{\mathrm{T}}^{\mathrm{miss}}$ are also used to complement the hadronic triggers.

Table 3.14: Triggers and thresholds for different data taking periods.

Dataset	Trigger	2016	2017	2018
JetHT	HLT_PFHTX, HLT_AK8PFJetY	900, 450	–	–
	HLT_PFHTX	–	1050	1050
HTMHT	HLT_PFHTX_PFMETY	300,110	–	–
MET	HLT_PFMET[NoMu]X_PFMHT[NoMu]Y_IDTight	110,110	–	–
	HLT_PFHTX_PFMETY_PFMHTZ_IDTight	–	500,100,100	500,100,100
			or 700,85,85	or 700,85,85
			or 800,75,75	or 800,75,75
	HLT_PFMET[NoMu]X_PFMHT[NoMu]Y_IDTight[_PFHTZ]	–	120,120,60	120,120,60
SingleMuon	HLT_Mu15_IsoVVVL_PFHTX	350 or 400	350 or 400	350 or 400
		or 450 or 600	or 450 or 600	or 450 or 600
	HLT_Iso[Tk]MuX	24	27	27
	HLT_[Tk]MuX	50 or 55	50 or 55	50 or 55
	HLT_[Old,Tk]MuX	–	100	–
SingleElectron	HLT_Ele15_IsoVVVL_PFHT	350 or 400	350 or 400	350 or 400
		or 450 or 600	or 450 or 600	or 450 or 600
	HLT_EleX_WPTight_Gsf	27 or 28 or 32	27 or 28 or 32	–
		or 35 or 38	or 35 or 38	–
	HLT_EleX_CaloIdVT_GsfTrkIdT	105 or 115	105 or 115	–
EGamma	HLT_EleX_WPTight_Gsf	–	–	27 or 28 or 32
				35 or 38
	HLT_EleX_CaloIdVT_GsfTrkIdT	–	–	105 or 115

The combined suite of hadronic and MET triggers used in the selection of the signal regions and the majority of control and validation regions is not fully efficient in the analysis regions characterized by $M_R > 800$ and $R^2 > 0.08$ (to be fully defined in Section 4.1). Therefore, we measure the trigger efficiencies from data and apply these efficiencies as weights to signal and background MC events. The efficiencies are measured and applied as a function of offline H_T and E_T^{miss} rather than the razor variables M_R and R^2 since the former allows a better statistical distribution, hence more accurate efficiency determination. The efficiencies are calculated using the orthogonal `SingleElectron`, `EGamma`, `SingleMuon` and `SinglePhoton` datasets after applying an offline hadronic preselection, which mainly consists of a reconstructed vertex, at least two regular jets (anti-k_T with $R = 0.4$), and one wide jet (anti-k_T with $\Delta R = 0.8$) having $p_T > 200$ GeV. For the `SingleElectron` / `EGamma` or `SingleMuon` datasets, we measure the efficiencies using only the events that pass the single electron triggers having online electron p_T thresholds of 23 or 27 GeV, or a single isolated muon trigger having an online muon p_T threshold of 24 GeV, respectively. We also require the event to have exactly one lepton passing our loose selection criteria (as defined in Section 3.3 and 3.3). For the `SinglePhoton` dataset, we use the OR of all single photon triggers (HLT_Photon) and require exactly one photon (as defined in Section 3.3) and zero leptons in the event.

Figure 3.11 and 3.11 shows the hadronic trigger efficiencies obtained from the `SingleElectron`, `SingleMuon`, `SinglePhoton` and MC simulation consisting of all background processes in the unrolled representation. The discrepancies seen between the efficiencies obtained from SingleLepton and SinglePhoton datasets would arise due to the online treatment of jets. Leptons and photons within jets contribute differently to the online jet p_T. To take such signature dependencies properly into account, we apply a signature based weighting of the MC events. The MC events having electrons, muons, or photons are weighted with the efficiencies obtained from the `SingleElectron` / `EGamma`, `SingleMuon` datasets, respectively. For the fully hadronic selections, the trigger efficiency weight is taken from the `SingleElectron` / `EGamma` dataset. This procedure improves the Data-MC agreement in the regions with leptons or photons.

The SingleElectron and SingleMuon triggers are also used for defining various control regions. The object thresholds for these triggers are higher than our offline selection in the regions where these triggers are used. Therefore, efficiencies are also calculated for these triggers as functions of electron or muon p_T, shown in Figure 3.13 and 3.14. Efficiencies are calculated using both the orthogonal JetHT and MET datasets and directly using simulation for cross check. No trigger requirement is applied to the denominator. Efficiencies calculated from the MET dataset are used for the leptonic triggers.

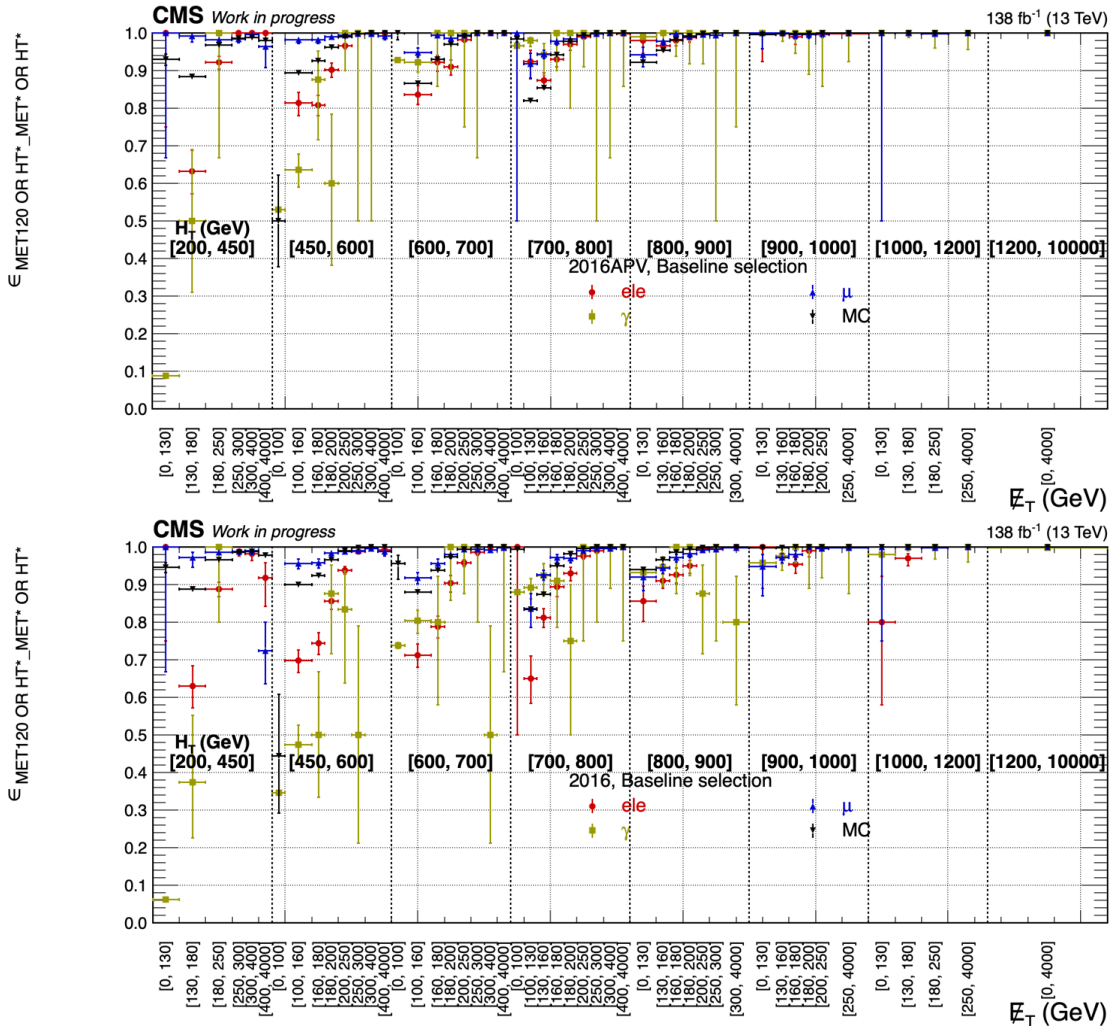


Figure 3.11: Unrolled trigger efficiencies for the JET, HT, and MET trigger suite versus H_T and E_T^{miss} obtained from the SingleElectron, SingleMuon, SinglePhoton, MET datasets, and simulation for 2016preVFP (top), 2016postVFP (bottom) data taking periods.

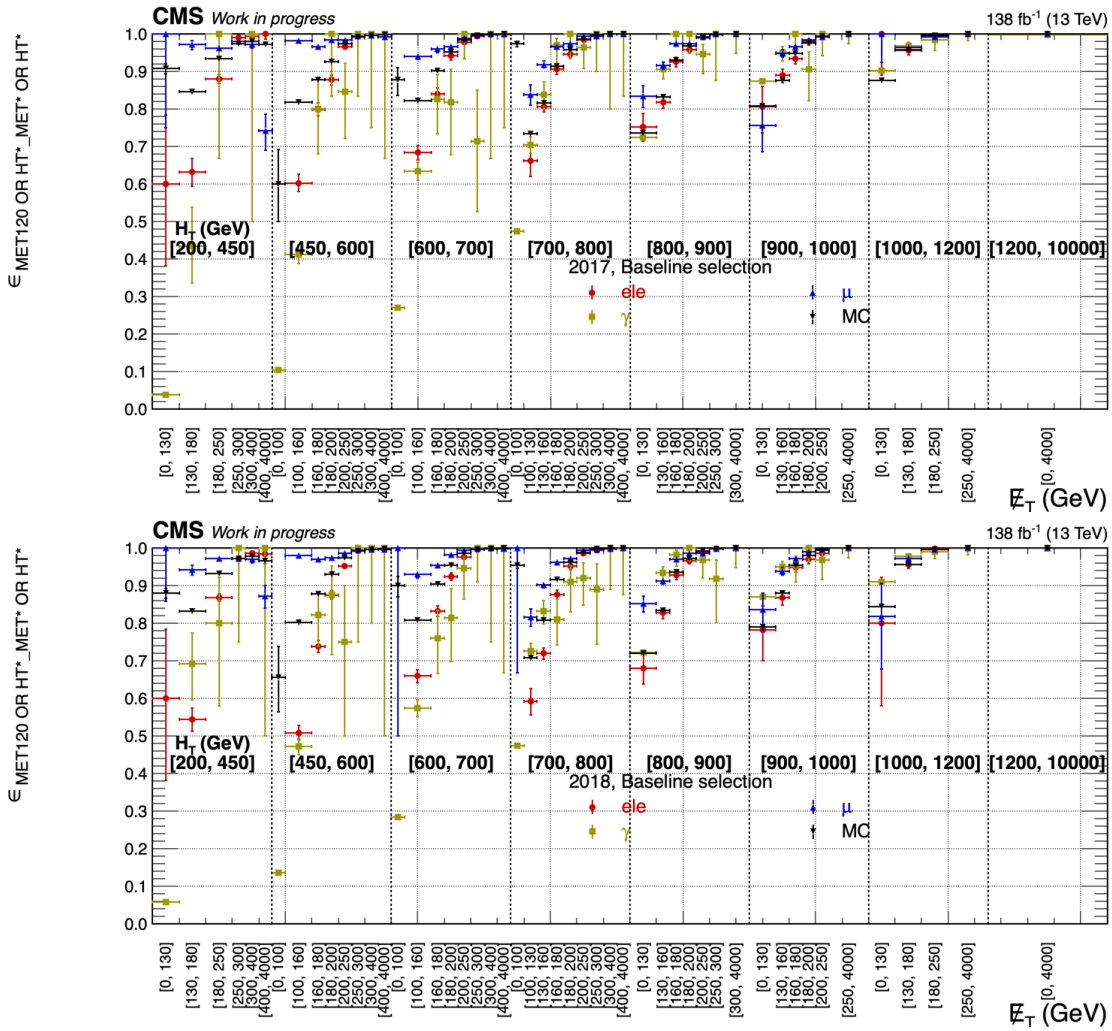


Figure 3.12: Unrolled trigger efficiencies for the JET, HT, and MET trigger suite versus H_T and E_T^{miss} obtained from the SingleElectron, SingleMuon, SinglePhoton, MET datasets, and simulation for 2017 (top), 2018 (bottom) data taking periods.

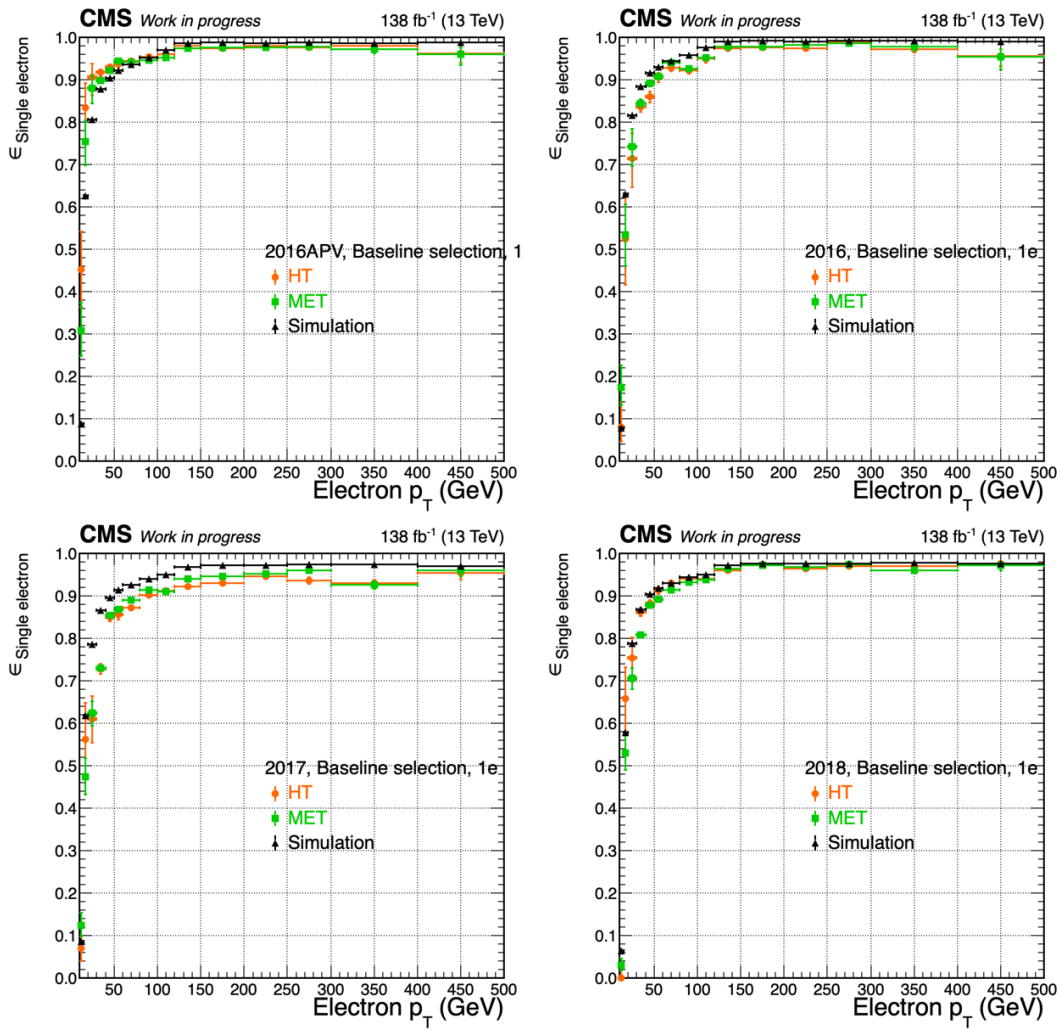


Figure 3.13: Efficiencies for the single electron triggers versus the relevant object p_T, obtained from the HT and MET datasets and simulation for the 2016APV (top left), 2016 (top right), 2017 (bottom left) and 2018 (bottom right) data taking periods.

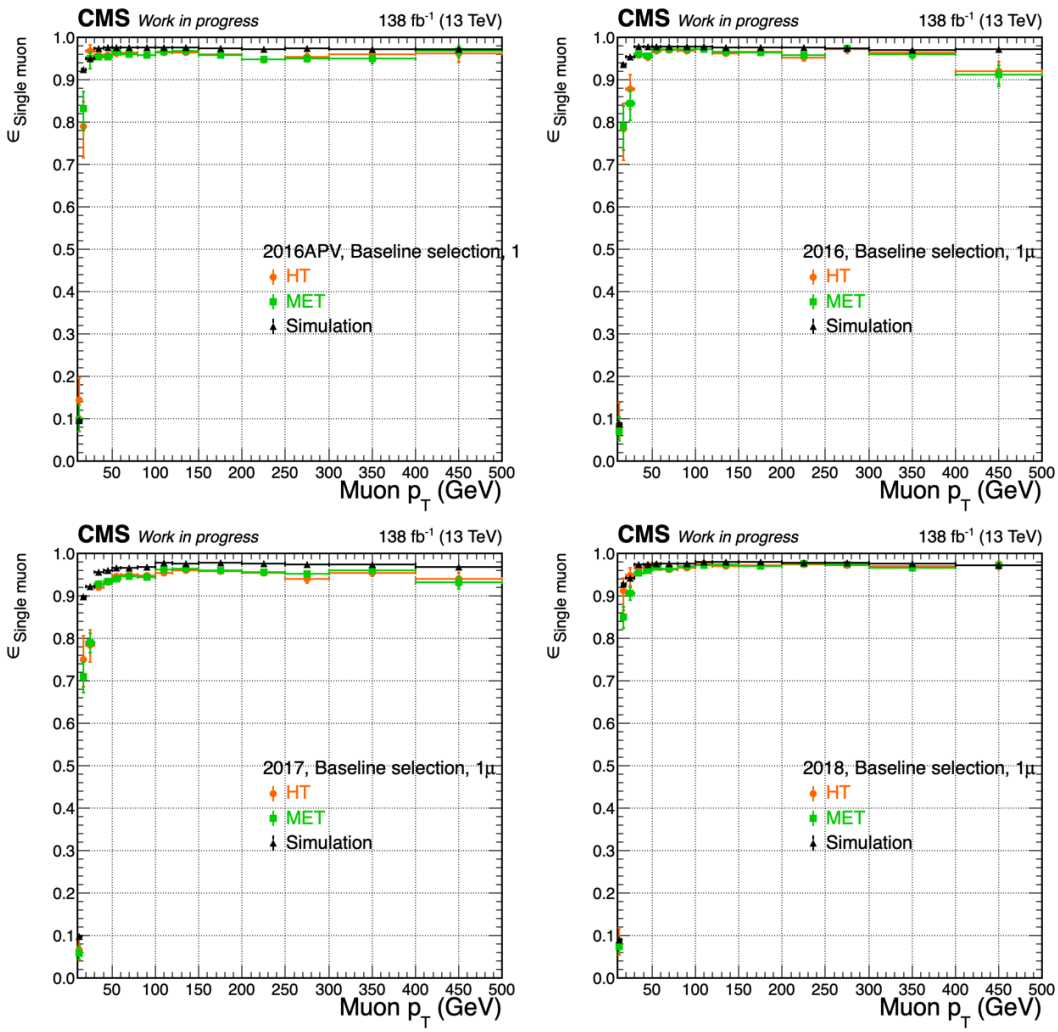


Figure 3.14: Efficiencies for the single muon triggers versus the relevant object p_T, obtained from the HT and MET datasets and simulation for the 2016APV (top left), 2016 (top right), 2017 (bottom left) and 2018 (bottom right) data taking periods.

Chapter 4

Analysis of Supersymmetric particles search in CMS experiment

4.1 Event Selection

This analysis looks for a discrepancy from the SM predictions in events with high M_R-R^2 in hadronic and leptonic final states with at least one boosted object. The boosted objects consist of boosted hadronic W/Z/Higgs bosons and top quarks, or boosted leptonic jets. Search regions are further classified into jet and b-jet multiplicity. All MC are corrected using various object data/MC scale factors and event reweighting, taking the associated systematic uncertainties into account, as described in Section 3.3.

Irreducible backgrounds in the signal search regions are mainly estimated using data control regions defined by reverting one or more selection variables and relaxing the boosted object selection criteria to only a mass greater than 50 GeV requirement. Data / MC correction factors versus $M_R \times R^2$ and various object multiplicities are derived in these control regions subsequently used to correct the background MC in the signal search regions. The correction factors are validated in validation regions(VR) with a similar selection to the control regions but with a completely tight boosted object selection. The validity of the background estimation is further checked via closure tests in validation regions. All signal, control, and validation

regions are binned in $M_R \times R^2$.

In this section, we define the baseline selection applied to data and MC events, then move on to the definitions of the three signal search regions used in this analysis.

4.1.1 Event Cleaning

Initial event cleaning is done according to the recommended procedure set out by the JetMET POG [27] by applying the following set of event filters. Unless stated otherwise, filters are applied to all data, background MC, and signal MC events.

- globalSuperTightHalo2016Filter
- HBHENoiseFilter
- HBHEIsoNoiseFilter
- eeBadScFilter
- EcalDeadCellTriggerPrimitiveFilter
- ecalBadCalibReduced$*$ (being finalized, will only apply to 2017 and 2018 data)
- BadPFMuonFilter
- BadPFMuonDzFilter
- hfNoisyHitsFilter
- goodVertices(primary vertex filter)

We also carried out studies of the L1 prefiring and HEM 15/16 HCAL FED in the latter portion of 2018 and applied recommended procedures.

4.1.2 Baseline preselection

All signal, control, and validation regions in this analysis are based on the following baseline preselection:

- Pass all detector- and beam-related filters
- Good primary vertices > 0 (via a vertex filter)
- ≥ 2 AK4 jets, with $p_T > 30$
- ≥ 1 AK8 jet, with $p_T \geq 200$ GeV
- $M_R > 800$ GeV, $R^2 > 0.08$
- $(M_R - 800) \times (R^2 - 0.08) > 30$ GeV
  - Fully hadronic preselection: Veto $e/\mu/\tau$
  - Isolated leptonic preselection: = 1 medium isolated lepton, = 0 jet
  - Non-isolated leptonic preselection: ≥ 1 non-isolated lepton

Figure 4.1 shows the distribution of several variables in the preselection region comparing data and MC. These variables will be used for defining selection criteria in the subsequent signal, control, and validation regions.

4.1.3 Signal regions

The signal regions(SR) in this analysis are characterized by the requirement of at least one of one or more types of the six types of tagged boosted object(s). Signal region selections are listed in Table 4.1. The selections are categorized into fully hadronic, isolated leptonic, or non-isolated leptonic final states. They are derived from the preselection regions described above, requiring 0 leptons, at least one isolated lepton, or at least one non-isolated lepton. Main event classification is based on the multiplicities of isolated leptons, jets, b jets, boosted $W/Z/h$ bosons, hadronic top quarks, and leptonic jets. Each selection increases sensitivity in different signal models featuring different SUSY particles, production, and decay mechanisms. In

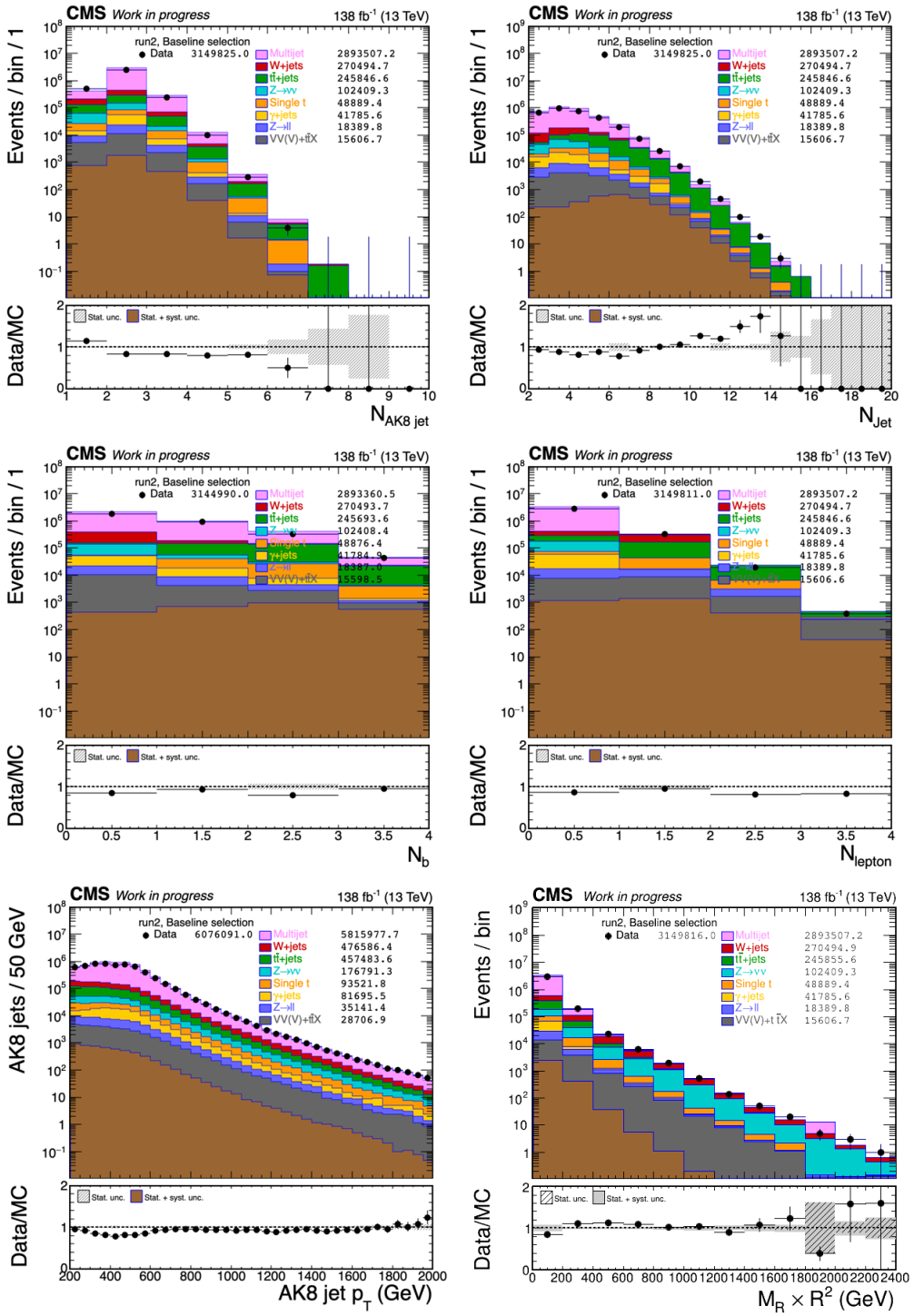


Figure 4.1: Data-MC comparison of several variables in the preselection region for the run2 data taking period.

the hadronic signal regions, to further reduce the SM backgrounds, especially those coming from multijet QCD production, we require the angular separation between the two megajets used for calculating the razor variables to be $\Delta\phi_{megajets} < 2.8$. In the leptonic regions, to enhance the signal over leptonically decaying top or W backgrounds, transverse mass is required to be $m_T > 120$ GeV. In the non-isolated leptonic signal regions, discrimination of leptonic W/top and multijet backgrounds is enhanced by requiring $m_T > 140$ and $\Delta\phi* > 0.8$ respectively, where $\Delta\phi*$ is defined as the angle between the leading boosted jet and the W boson which is defined as the leading non-isolated lepton $+p_{T,miss}$. All SRs are defined to be disjoint.

Each signal region then is binned further in $(M_R - 800) \times (R^2 - 0.08)$. The binning is defined as following steps.

1. Set the $(M_R - 800) \times (R^2 - 0.08)$ bins with [30, 40, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 3000]

2. Merge all the combinations of bins and find the bin that minimizes the following value. When values are calculated, B_k must be greater than a threshold value.

$$\sum_{i}^{SMSs} \sum_{j}^{mass\,bins} \sum_{k}^{searchbin} \frac{S_{ijk}}{\sqrt{S_{ijk} + B_k}} \tag{4.1}$$

The bin optimization in each signal region is shown in Table 4.2.

Signal regions are dominated with $\bar{t}t$+jets and single top processes for all categories followed by multijet production in the W categories and $Z(\rightarrow \nu\nu)$+jets production in the Top category.

Table 4.1: Signal regions and their selection criteria

	Category	Isolep	jets	b	V	h	hTop	lepJet	Extra
1	had 1hTop	0	≥ 5	–	–	0	1	0	$\Delta\phi_R < 2.8$
2	had 2hTop	0	≥ 2	–	–	0	≥ 2	0	$\Delta\phi_R < 2.8$
3	had V b 45j	0	4-5	≥ 1	≥ 1	0	0	0	$\Delta\phi_R < 2.8$
4	had V b 6j	0	≥ 6	≥ 1	≥ 1	0	0	0	$\Delta\phi_R < 2.8$
5	had 1V 0b 34j	0	3-4	0	1	0	0	0	$\Delta\phi_R < 2.8$
6	had 1V 0b 5j	0	≥ 5	0	1	0	0	0	$\Delta\phi_R < 2.8$
7	had 2V 0b 24j	0	2-4	0	≥ 2	0	0	0	$\Delta\phi_R < 2.8$
8	had 2V 0b 5j	0	≥ 5	0	≥ 2	0	0	0	$\Delta\phi_R < 2.8$
9	had H b 45j	0	4-5	≥ 1	–	≥ 1	–	0	$\Delta\phi_R < 2.8$
10	had H b 6j	0	≥ 6	≥ 1	–	≥ 1	–	0	$\Delta\phi_R < 2.8$
11	had HV b 6j	0	≥ 6	≥ 1	1	1	–	0	$\Delta\phi_R < 2.8$
12	had H 0b 34j	0	3-4	0	–	≥ 1	–	0	$\Delta\phi_R < 2.8$
13	had H 0b 5j	0	≥ 5	0	–	≥ 1	–	0	$\Delta\phi_R < 2.8$
14	had HV 0b 24j	0	2-4	0	≥ 1	≥ 1	–	0	$\Delta\phi_R < 2.8$
15	had HV 0b 5j	0	≥ 5	0	≥ 1	≥ 1	–	0	$\Delta\phi_R < 2.8$
16	lep 1hTop	1	≥ 3	–	–	0	1	0	$m_T > 120$
17	lep V b	1	≥ 4	≥ 1	≥ 1	0	0	0	$m_T > 120$
18	lep V 0b	1	≥ 2	0	≥ 1	0	0	0	$m_T > 120$
19	lep H b 3j	1	≥ 3	≥ 1	–	≥ 1	–	0	$m_T > 120$
20	lep H 0b 2j	1	≥ 2	0	–	≥ 1	–	0	$m_T > 120$
21	lepJet hTop	–	≥ 3	–	–	–	≥ 1	≥ 1	$m_T > 140, \Delta\phi* > 0.8$
22	lepJet 0V 24j	–	2-4	–	0	–	0	≥ 1	$m_T > 140, \Delta\phi* > 0.8$
23	lepJet 0V 5j	–	≥ 5	–	0	–	0	≥ 1	$m_T > 140, \Delta\phi* > 0.8$
24	lepJet 1V 24j	–	2-4	–	1	–	0	≥ 1	$m_T > 140, \Delta\phi* > 0.8$
25	lepJet 1V 5j	–	≥ 5	–	1	–	0	≥ 1	$m_T > 140, \Delta\phi* > 0.8$

Table 4.2: Signal regions and their binning

	category	Binning
1	had 1hTop	30, 40, 50, 75, 100, 150, 200, 250, 300, 400, 500, 3000
2	had 2hTop	30, 40, 50, 75, 100, 150, 200, 250, 3000
3	had V b 45j	30, 40, 50, 75, 100, 150, 200, 250, 300, 3000
4	had V b 6j	30, 40, 50, 75, 100, 150, 200, 250, 300, 3000
5	had 1V 0b 34j	30, 40, 50, 75, 100, 150, 200, 250, 300, 350, 3000
6	had 1V 0b 5j	30, 40, 50, 75, 100, 150, 200, 250, 3000
7	had 2V 0b 24j	30, 40, 50, 3000
8	had 2V 0b 5j	30, 40, 50, 3000
9	had H b 45j	30, 40, 50, 75, 100, 150, 200, 250, 3000
10	had H b 6j	30, 40, 50, 75, 100, 150, 3000
11	had HV b 6j	30, 40, 3000
12	had H 0b 34j	30, 40, 50, 75, 100, 3000
13	had H 0b 5j	30, 40, 75, 150, 3000
14	had HV 0b 24j	30, 40, 3000
15	had HV 0b 5j	30, 40, 3000
16	lep 1hTop	30, 40, 50, 75, 100, 150, 200, 250, 300, 3000
17	lep V b	30, 40, 50, 75, 100, 150, 200, 3000
18	lep V 0b	30, 40, 50, 75, 100, 150, 200, 3000
19	lep H b 3j	30, 40, 50, 75, 100, 3000
20	lep H 0b 2j	30, 40, 50, 3000
21	lepJet hTop	30, 40, 50, 75, 100, 150, 200, 3000
22	lepJet 0V 24j	30, 40, 50, 75, 100, 150, 200, 250, 300, 350, 3000
23	lepJet 0V 5j	30, 40, 50, 75, 100, 150, 200, 250, 3000
24	lepJet 1V 24j	30, 40, 50, 75, 100, 3000
25	lepJet 1V 5j	30, 40, 50, 75, 100, 3000

4.2 Background estimation

4.2.1 Sources of irreducible backgrounds

The signal region selections described above eliminate a large portion of the SM backgrounds. However, there still remains a sizeable amount of irreducible backgrounds in each signal region. These irreducible backgrounds are classified as follows:

- **Multijet:** QCD production of dijet or multijet processes. Mismeasurement of jets causes an imbalance of the visible momentum in the transverse direction. This reflects as missing transverse energy, and consequently leads to high R^2. The tagged boosted objects mostly result from mistagging light jets for these processes. On rare occasions, hadronically decaying W bosons and top quarks can be present and tagged as boosted objects. Leptonic boosted objects from non-isolated leptons can occur due to meson decays in the hadronic showers.

- **Lost lepton:** In events with leptonic decays of W bosons or top quarks, the lepton may escape detection. In these events, E_T^{miss} would constitute of a genuine component from the neutrino and the component the lost lepton. For these processes, the tagged boosted objects could either be genuine hadronically decaying top quarks or W bosons, or can also arise from mistagging light jets. The latter happens, e.g. in W+jets processes where the W boson decays leptonically. Genuine leptonic boosted objects can also appear here, e.g. from leptonic decays of boosted top quarks or from merging of the lepton from a W decay with one of the neighboring jets.

- $Z \to \nu\nu$**:** $Z(\to \nu\nu)$ events produced in association with jets lead to hadronic signatures with genuine E_T^{miss} , and constitute backgrounds in hadronic final states, in particular with zero b jets. As for the multijet case, such processes can also constitute backgrounds to the non-isolated leptonic category as leptonic boosted objects from non-isolated leptons can occur due to meson decays in the hadronic showers. The tagged boosted objects in this process entirely result from mistagging light jets.

Background characterization in signal regions

In order to understand the contributions of various residual SM background subprocesses to different SRs, and subsequently to devise a reliable procedure to estimate them, we performed a study at the MC truth level. We used the truth level information in the residual background events to determine which $t\bar{t}$+jets, single top and W+jets decay channels populate the various SRs. Tables 4.3 and 4.4 show the percentage of each decay channel for a given process populating the hadronic and isolated and non-isolated leptonic SRs, respectively. In the hadronic SRs, the dominant $t\bar{t}$+jets contribution comes from the semileptonic $t\bar{t}$ decays almost for all regions, while the hadronic decays usually contribute the least; whereas for single top and W+jets events, contributions from hadronic and leptonic decays are usually balanced. In the isolated leptonic SRs, for $t\bar{t}$+jets , semileptonic and dileptonic $t\bar{t}$ decays are balanced, while the hadronic decays contribute negligibly; whereas for single top and W+jets events, contributions from leptonic decays dominate. In the non-isolated leptonic SRs, the dominant $t\bar{t}$+jets contribution comes from the dileptonic $t\bar{t}$ decays almost for all regions, while the hadronic decays contribute negligibly; whereas for single top and W+jets events, similar to the isolated lepton SR case, contributions from leptonic decays dominate. This information is summarized in Table 4.5, and it is used to ensure defining control regions that correctly represent the backgrounds in the SRs.

4.2.2 Overview of the estimation strategy

The razor boost analysis, being highly inclusive, studies 25 signal regions with varying topologies. The composition of the irreducible backgrounds can largely vary among these signal regions, depending on the topology. Despite this fact, it is possible to employ a mostly unified strategy in estimating the irreducible backgrounds in these regions. In the following, we present an outline of the strategy, which details will follow in the next sections.

Background prediction is performed either using a MC assisted data-driven way, or directly using MC events. Choice of method depends on the abundance of the

Table 4.3: Percentage of contributions from different decay channels of $t\bar{t}$+jets , single top and W+jets processes to the hadronic signal regions

		tt+jets			Single top		W+jets	
	Category	had	1lep	2lep	had	lep	had	lep
1	had 1hTop	4.0	77.5	18.5	75.6	2.4	51.3	48.7
2	had 2hTop	12.3	78.9	8.8	33.3	66.7	52.1	47.9
3	had V b 45j	2.7	77.8	19.5	61.9	38.1	50.4	49.6
4	had V b 6j	7.1	76.6	16.3	67.1	32.9	49.3	50.7
5	had 1V 0b 34j	2.1	78.6	19.3	64.8	35.2	47.3	52.7
6	had 1V 0b 5j	6.6	76.4	17.0	76.2	23.8	54.1	45.9
7	had 2V 0b 24j	17.9	64.3	17.9	33.3	66.7	53.8	46.2
8	had 2V 0b 5j	21.6	70.6	7.8	0	100.0	57.1	42.9
9	had H b 45j	3.5	76.5	19.9	71.7	28.3	74.0	53.0
10	had H b 6j	8.3	77.5	14.2	77.8	22.2	59.5	40.5
11	had HV b 5j	9.6	80.7	9.6	100.0	0	77.8	22.2
12	had H 0b 34j	3.1	74.4	22.4	75.0	25.0	43.2	56.8
13	had H 0b 5j	3.3	78.0	18.7	66.7	33.3	43.2	56.8
14	had HV 0b 24j	0	78.4	21.6	–	–	41.9	58.1
15	had HV 0b 5j	4.3	82.6	13.0	–	–	42.9	57.1

Table 4.4: Percentage of contributions from different decay channels of $t\bar{t}$+jets , single top and W+jets processes to the isolated and non-isolated leptonic SRs

		tt+jets			Single top		W+jets	
	Category	had	1lep	2lep	had	lep	had	lep
16	lep 1hTop	0.9	67.0	32.0	54.0	46.0	1.5	98.5
17	lep V b	0.4	49.1	50.5	19.1	80.9	4.9	95.1
18	lep V 0b	0.1	43.5	56.3	38.8	61.2	1.3	98.7
19	lep H b 3j	0.5	51.7	47.8	–	–	0	100.0
20	lep H 0b 2j	0.9	42.1	57.0	–	–	1.5	98.5
21	lepJet 1hTop	2.1	44.0	53.9	13.3	86.7	13.9	86.1
22	lepJet 0V 24j	0.6	16.2	83.3	14.4	85.6	7.4	92.6
23	lepJet 0V 5j	1.5	23.0	75.5	19.3	80.7	19.0	81.0
24	lepJet 1V 24j	0.9	25.0	74.1	12.5	87.5	5.2	94.8
25	lepJet 1V 5j	0.9	25.0	74.1	12.5	87.5	5.2	94.8

Table 4.5: Summary of contributions from different decay channels of $t\bar{t}$+jets , single top and W+jets processes to the hadronic, leptonic and non-isolated leptonic signal region categories

	tt+jets			Single top		W+jets	
Category	had	1lep	2lep	had	lep	had	lep
had	~ 1 – 10%	dominant		balanced	balanced	balanced	balanced
isolep	negligible	balanced	balanced		dominant		dominant
nonisolep	negligible		dominant		dominant		dominant

background and the reliability of its MC modeling.

The **major backgrounds** with the largest contribution, such as tt+jets, W+jets, QCD multijets, $Z(\to \nu\nu)$+jets are estimated using the MC-assisted data-driven method. In this method, control regions (CRs) are determined in which the background of interest is largely dominant. Correction factors (CFs), defined as the ratio of data over MC in the distribution of the variable of interest are computed. The CFs are subsequently applied to the MC distribution of the corresponding background in the SRs.

For each background, multiple CFs are calculated versus different event variables iteratively. The iterative approach is favored over a multi-dimensional CF calculation, as the former would yield smaller statistical uncertainties. At each step, the obtained CF is applied to the CR, before calculating the next CF for the next variable. For all cases, CFs are calculated first versus $M_R \times R^2$, followed by the number of AK4 jets. For the special case of $Z(\to \nu\nu)$+jets estimation via a γ+ jets CR, a third iteration is done by calculating CFs versus number of AK8 jets with mass greater than 50 GeV. For the **minor backgrounds** with percent level contributions, prediction is based on direct MC counts. A potential correlation due to the order of iteration is checked by reverting the order of $M_R \times R^2$ and number of AK4 jets steps. The resulting distributions are compared and any discrepancy is added as a bin-by-bin systematic uncertainty.

Though the general strategy is similar, background prediction in SRs from different preselection groups has slight differences, as listed below. The differences mainly arise from the definition of CRs from which the CFs are derived.

1. Hadronic SRs with boosted W/Z/top/H:

    - **Multijet, top, W+jets backgrounds:** Define multijet, top, W-enriched CRs by reverting SR selections and simultaneously compute CFs from the 3 CRs.
    - **Z$(\to \nu\nu)$+jets:** Use γ+jets-enriched and DY-enriched control regions to find CFs. Cross-check with CFs computed from W- and T-enriched single lepton CRs.

- **Others:** Use MC counts.

2. Leptonic SRs with boosted W/Z/top/H:

   - **All BGs:** Use the same CFs as for the hadronic SRs.

3. Non-isolated leptonic SRs with boosted lepjets/leptops:

   - **top, W+jets BGs:** Use control regions obtained by reverting m_T and $\Delta\phi*$ to simultaneously compute data/MC correction factors.
   - **Multijets:** Use the CFs from hadronic SRs.
   - **Others:** Use MC counts.

Table 4.6 lists all the CRs used in this analysis, the type of background they estimate, their selection definitions, and the SR group for which they are used. CRs are defined by reverting the selections on one of more variables that define the SRs. For many cases, dedicated CRs are defined separately for events with 1 and greater than or equal to 2 AK8 jets with mass greater than 50. In several others, this classification is not done, in order to maintain statistics. The dependence on AK8 jets is rather taken into account by computing an additional CF versus $\mathrm{AK8_{m>50}}$ multiplicity. All CRs are by definition orthogonal to the SRs.

4.2.3 Simultaneous computation of the data / MC correction factors

Except for $\mathrm{CF}_{\mathrm{Zinv}}^{\gamma+\mathrm{jets}}$, i.e., the CF for $Z(\rightarrow \nu\nu)$+jets calculated from a γ+jets CR, all CFs are computed simultaneously for groups of 2 or 3 background processes from CRs. This is a new approach compared to the more traditional method used in the previous razor analyses, of first finding the purest CR, subtracting the backgrounds other than the background of interest, finding the CF for that background, applying it to the MC in other CRs, and proceeding iteratively to find CFs for other backgrounds. The simultaneous computation method has the advantages of both finding the CFs in one iteration and also of using information from multiple CRs in deriving the CFs.

Table 4.6: Control regions used for background estimation. The region names W4Z and T4Z are abbreviations for WforZ and TforZ. 'Photon added to E_T^{miss}, "Lepton added to E_T^{miss}, "'Leptons added to E_T^{miss} and $|m_{\ell\ell} - m_Z| < 10$ GeV.

			Selection										Used for SR		
	CR	BG	AK8$_{m50}$	ℓ_{niso}	b$_l$	b$_m$	ℓ_{veto}	ℓ_{sel}	γ	$\Delta\phi$	$\Delta\phi^*$	m_T	Had	lep	nIso
1	Q1	QCD	1	–	0	–	0	–	–	≥ 2.8	–	–	✓	✓	✓
2	Q2	QCD	≥ 2	–	0	–	0	–	–	≥ 2.8	–	–	✓	✓	✓
3	T1	Top	1	–	–	≥ 1	–	1	0	< 2.8	–	< 120	✓	✓	–
4	T2	Top	≥ 2	–	–	≥ 1	–	1	0	< 2.8	–	< 120	✓	✓	–
5	TnI	Top	≥ 1	≥ 1	–	≥ 1	–	–	0	–	< 0.8	< 140	–	–	✓
6	W1	W	1	–	0	–	–	1	0	< 2.8	–	30-100	✓	✓	–
7	W2	W	≥ 2	–	0	–	–	1	0	< 2.8	–	30-100	✓	✓	–
8	WnI	W	≥ 2	≥ 1	0	–	–	–	0	–	< 0.8	< 140	–	–	✓
9	G′	Zinv	≥ 1	–	–	–	0	–	≥ 1	< 2.8	–	–	✓	✓	–
10	DY‴	Zinv	≥ 1	–	0	–	–	2,OS	0	< 2.8	–	–	✓	✓	–
11	W4Z1″	Zinv	1	–	0	–	1	–	0	< 2.8	–	30-100	✓	✓	–
12	W4Z2″	Zinv	≥ 2	–	0	–	1	–	0	< 2.8	–	30-100	✓	✓	–
13	T4Z1″	Zinv	1	–	–	≥ 1	1	–	0	< 2.8	–	< 120	✓	✓	–
14	T4Z2″	Zinv	≥ 2	–	–	≥ 1	1	–	0	< 2.8	–	< 120	✓	✓	–

For the case of N backgrounds, first, N CRs are defined, ideally, where each is dominated by one of the N backgrounds. Data and MC distributions are obtained for the variable versus which the CF is to be derived. For each CR, first all the minor backgrounds to be estimated directly via MC, are subtracted from data, leaving only the contributions from the major backgrounds for which CFs are to be computed. Making the assumption that the CF for a given background is the same in each region, in each bin i of the distribution, the data yields, MC counts and CFs can be related as N equations with N unknowns, where the unknowns are the CFs. For example, for N=3, CF_{BGA}, CF_{BGB} and CF_{BGC}, as shown below:

$$\begin{aligned}
CF_{BGA}MC_{BGA}^{CRA} + CF_{BGB}MC_{BGB}^{CRA} + CF_{BGC}MC_{BGC}^{CRA} &= Data^{CRA} \\
CF_{BGB}MC_{BGA}^{CRB} + CF_{BGB}MC_{BGB}^{CRB} + CF_{BGC}MC_{BGC}^{CRB} &= Data^{CRB} \\
CF_{BGA}MC_{BGA}^{CRC} + CF_{BGB}MC_{BGB}^{CRC} + CF_{BGC}MC_{BGC}^{CRC} &= Data^{CRC}
\end{aligned}$$

We obtain the CFs by performing a simultaneous likelihood fit to data in the CRs. The likelihood is built for each bin separately, also incorporating the statistical

uncertainties as described in [28]. This method allows to the CF uncertainties taking into account the correlations between different processes.

An alternative approach which was used to cross check the robustness of the fit approach is to solve the set of equations using Cramer's rule to obtain the CFs. This method is straightforwardly extensible to estimating the uncertainties associated with the CFs. The central values and uncertainties are computed following the steps below: For each bin b:

- MC count means and uncertainties in each bin for each background process in each CR are used to construct Gaussians.
- 1M sets of counts are randomly sampled from each Gaussian.
- Eq. 4.2 is solved for each set for CF^{b}_{ProcX}, $X = A, B, C$. Distributions of solutions of each bin of each CF are obtained.
- The maxima of solution distributions are taken as the central values CF^{b}_{ProcX} while the 68% band around maxima gives the uncertainties $\sigma_{CF^{b}_{ProcX}}$.

The central values given by both approaches were very similar, while the uncertainties given by the fit method are overall slightly more conservative. The fit method is adopted throughout the rest of the analysis.

4.2.4 Validation strategy

CFs are validated by applying them to a series of validation regions (VRs) with different purposes and checking the resulting data-MC agreement. Table 4.7 lists all the validation regions, their selection criteria and their validation context.

A first set of VRs is used to check the validity of CFs in the presence of tagged boosted objects. This VRs are defined by requiring explicit boosted object tagging on top of the Q, T and W CR selections used for CF derivation, which do not have explicit criteria on the boosted object multiplicity, but rather are required to have AK8 jets with $m > 50$ GeV, in order to avoid limitations in statistics. Consequently,

the VRs are subsets of the CRs. These tagged boosted object VRs have similar background compositions with the CRs defined by mass-tagged AK8 jets. Therefore, ideally, the CFs versus $M_R \times R^2$ and jet multiplicity are expected to be independent of the boosted object criteria, and in case of sufficient statistics, would be equivalent. Tagged boosted object VRs validate the estimation in the hadronic and leptonic SRs.

A second set of VRs are used to check the validity of CFs for different kinematic conditions. They are obtained by taking a CR-like selection and reverting some selection criteria to obtain a more SR-like selection. They test the applicability of CFs derived in CRs in kinematics that are more alike SRs. These VRs validate the estimation in the hadronic and leptonic SRs.

Another VR exists for justifying the usage of top CFs, which are derived in a CR where the $t\bar{t}$ decays are via the single lepton channel, for SRs where $t\bar{t}$+jets backgrounds have contributions from both the single lepton and dilepton $t\bar{t}$ decays, as explained in Section 4.2.5.

A final set of VRs is used to check the validity of CFs for the nonisolated lepton regions. They work with a principle similar to the previous set. They validate the estimation in the nonisolated leptonic SRs.

4.2.5 Estimation of the multijet, W+jets and $t\bar{t}$+jets backgrounds

The CFs which will be used to correct the QCD multijet background in all SRs, and the $t\bar{t}$+jets and W+jets backgrounds in the hadronic and leptonic SRs are computed from the Q1, Q2; T1, T2 and W1, W2 CRs listed in Table 4.6, that are dominated by these 3 subsequent backgrounds. Two sets of CFs are calculated separately for cases with = 1 and ≥ 2 AK8 jets with mass greater than 50, using the CR sets Q1, T1, W1 and Q2, T2, W2. For a given variable and a bin, data and MC quantities in the CRs in each set $i = 1, 2$ are related with the following equations:

Table 4.7: Validation regions for the correction factors used in background estimation.

	VR	AK8$_{m50}$	V	hTop	H	ℓ_{niso}	b$_l$	b$_m$	ℓ_{veto}	γ	$\Delta\phi$	$\Delta\phi^*$	m_T
Check validity of CFs in presence of boosted objects - Validate had and lep SR BG estimate													
1	QV	–	≥ 1	–	–	0	0	–	0	0	≥ 2.8	–	–
2	Qtop	–	–	≥ 1	–	0	0	–	0	0	≥ 2.8	–	–
3	TV	–	≥ 1	–	–	0	–	≥ 1	1	0	< 2.8	–	< 120
4	Ttop	–	–	≥ 1	–	0	–	≥ 1	1	0	< 2.8	–	< 120
5	TH	–	–	–	≥ 1	0	–	≥ 1	1	0	< 2.8	–	< 120
6	WV	–	≥ 1	–	–	0	0	–	1	0	< 2.8	–	30-100
7	Wtop	–	–	≥ 1	–	0	0	–	1	0	< 2.8	–	30-100
Check validity of CFs for different kinematic conditions - Validate had and lep SR BG estimate													
8	S'V	–	≥ 1	–	–	0	0	–	0	0	≥ 2.8	–	–
9	S'Top	–	–	≥ 1	–	0	0	–	0	0	≥ 2.8	–	–
10	Q'	≥ 1	0	0	0	0	0	–	0	0	< 2.8	–	–
Check validity of the Top CF for dilepton backgrounds													
11	Tll	≥ 1	–	–	–	–	–	≥ 1	2	0	< 2.8	–	$M_{T2} < 200$
Check validity of CFs for the noniso lepton regions - Validate noniso lep SR BG estimate													
12	low$d\phi^*$	≥ 1	–	–	–	≥ 1	–	–	–	0	–	< 0.8	> 140
13	lowM_T	≥ 1	–	–	–	≥ 1	–	–	–	0	–	> 0.8	< 140

$$\mathrm{CF}^{\mathrm{i}}_{\mathrm{QCD}}\mathrm{MC}^{\mathrm{Qi}}_{\mathrm{QCD}} + \mathrm{CF}^{\mathrm{i}}_{\mathrm{ttjets}}\mathrm{MC}^{\mathrm{Qi}}_{\mathrm{ttjets}} + \mathrm{CF}^{\mathrm{i}}_{\mathrm{Wjets}}\mathrm{MC}^{\mathrm{Qi}}_{\mathrm{Wjets}} = \mathrm{Data}^{\mathrm{Qi}} \quad (4.2)$$

$$\mathrm{CF}^{\mathrm{i}}_{\mathrm{QCD}}\mathrm{MC}^{\mathrm{Ti}}_{\mathrm{QCD}} + \mathrm{CF}^{\mathrm{i}}_{\mathrm{ttjets}}\mathrm{MC}^{\mathrm{Ti}}_{\mathrm{ttjets}} + \mathrm{CF}^{\mathrm{i}}_{\mathrm{Wjets}}\mathrm{MC}^{\mathrm{Ti}}_{\mathrm{Wjets}} = \mathrm{Data}^{\mathrm{Ti}} \quad (4.3)$$

$$\mathrm{CF}^{\mathrm{i}}_{\mathrm{QCD}}\mathrm{MC}^{\mathrm{Wi}}_{\mathrm{QCD}} + \mathrm{CF}^{\mathrm{i}}_{\mathrm{ttjets}}\mathrm{MC}^{\mathrm{Wi}}_{\mathrm{ttjets}} + \mathrm{CF}^{\mathrm{i}}_{\mathrm{Wjets}}\mathrm{MC}^{\mathrm{Wi}}_{\mathrm{Wjets}} = \mathrm{Data}^{\mathrm{Wi}}. \quad (4.4)$$

$$(4.5)$$

where we solve for CF^{i}_{QCD}, $CF^{i}_{tt\,jets}$ and $CF^{i}_{W\,jets}$.

The first step is to calculate CFs versus $M_R \times R^2$. Figures 4.2, 4.4 and 4.6 show data and MC distributions for $M_R \times R^2$ for the Q1, T1, W1, Q2, T2 and W2 CRs for the 2016, 2017 and 2018 data taking periods. The resulting CFs and their uncertainties versus $M_R \times R^2$ are shown in Figure 4.3, 4.5, and 4.7. The correlation matrix of CFs are shown in Figure 4.8.

Next, we apply the $M_R \times R^2$ CFs to the relevant SM MC event-by-event and obtain data and MC distributions versus AK4 jet multiplicity for each CR, which are shown in Figures 4.9, 4.11 and 4.13. We obtain CFs for AK4 jet multiplicity from these distributions, as shown in Figures 4.10, 4.12, and 4.14.

The correctness of the procedure is checked in a closure test by applying the $M_R \times R^2$ and AK4 jet multiplicity CFs event-by-event to the MC in the Q1, T1, W1, Q2, T2 and W2 control regions and checking the data-MC agreement. As shown in Figures 4.18, 4.19, and 4.20 for the 2016, 2017, and 2018 data taking periods, data and corrected MC are in precise agreement.

Validating the use of QCD, W+jets and $t\bar{t}$+jets correction factors in the isolated lepton signal regions

The BG characterization study described in Section 4.2.1, checking the proportions of different BG decay channels in SRs, showed a non-negligible amount of dileptonic contribution in the residual $t\bar{t}$+jets backgrounds. In particular, in the isolated lepton SRs, the dilepton channel contribution is similar to that of single lepton. On the other hand, the T1 and T2 CRs, which are used to determine the $t\bar{t}$+jets CFs are dominated with events arising from single lepton decays. Therefore, the validity of

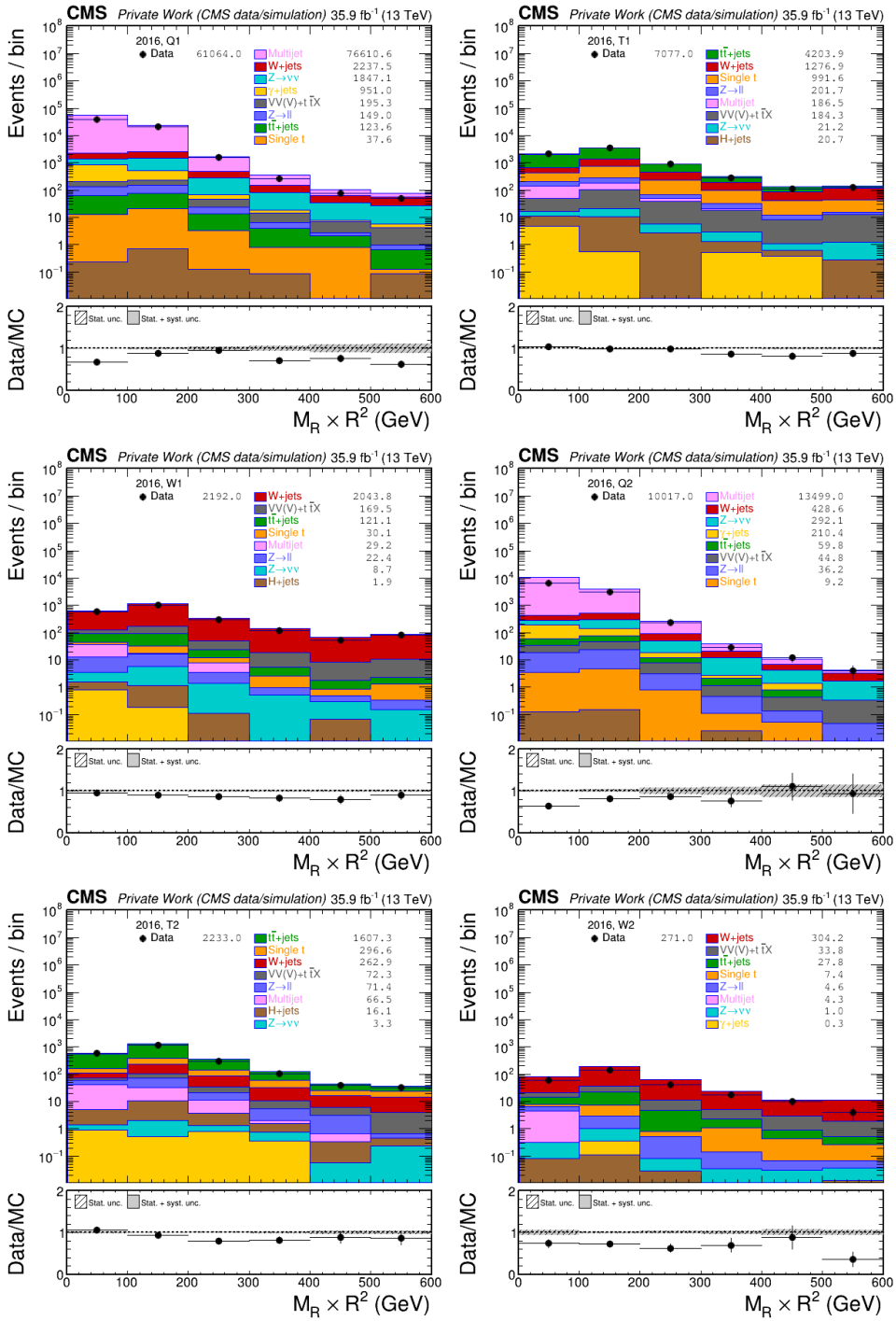


Figure 4.2: Data and MC distributions for $M_R \times R^2$ for the Q1, T1, W1, Q2, T2 and W2 control regions defined in Table 4.6 for the 2016 data taking period.

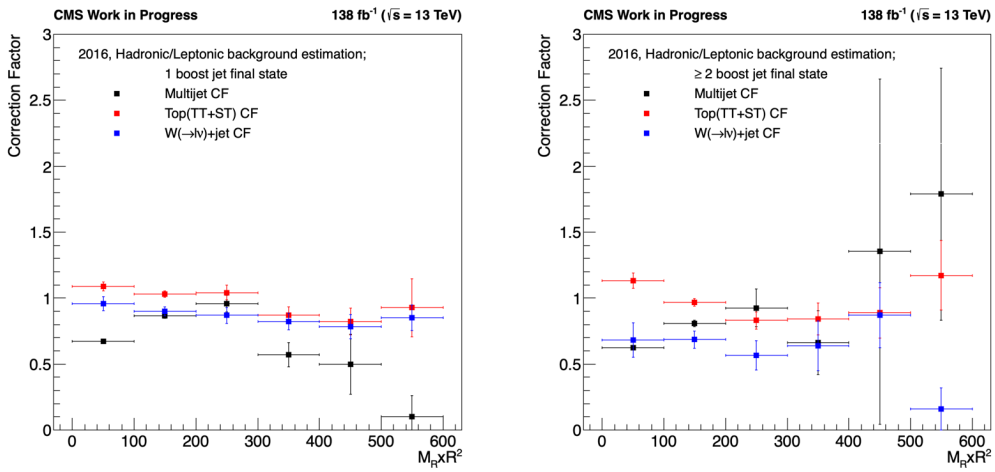


Figure 4.3: Distributions of $M_R \times R^2$ correction factors for QCD, $t\bar{t}$+jets and W+jets MC for cases with exactly 1 or at least 2 mass-tagged AK8 jets.

applying the $t\bar{t}$+jets CFs in SRs with a significant contribution of dilepton events needs to be established. For this purpose, we define a dilepton VR Tll dominated by the dilepton decay mode, whose selection is shown in Table 4.7. We apply the $t\bar{t}$+jets CFs to this VR, and check the data-MC agreement. Figure 4.21 shows the Tll VR before and after applying the $t\bar{t}$+jets CFs where a data-MC agreement is visible.

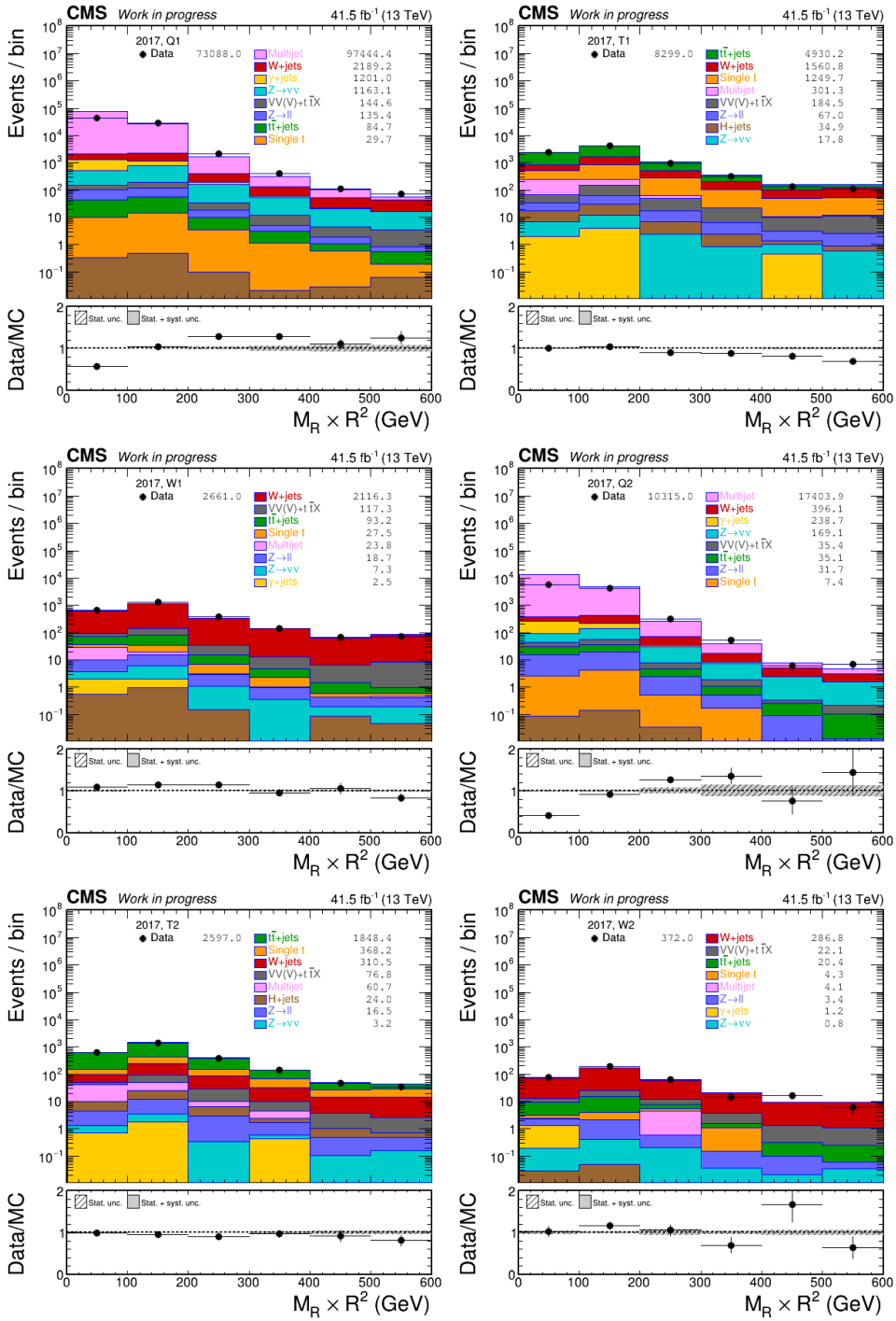


Figure 4.4: Data and MC distributions for $M_R \times R^2$ for the Q1, T1, W1, Q2, T2 and W2 control regions defined in Table 4.6 for the 2017 data taking period.

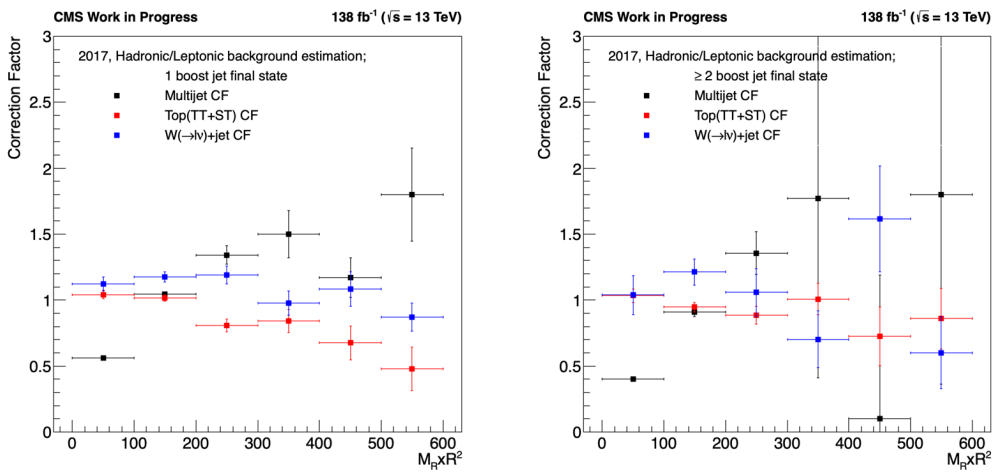


Figure 4.5: Distributions of $M_R \times R^2$ correction factors for QCD, $t\bar{t}$+jets and W+jets MC for cases with exactly 1 or at least 2 mass-tagged AK8 jets.

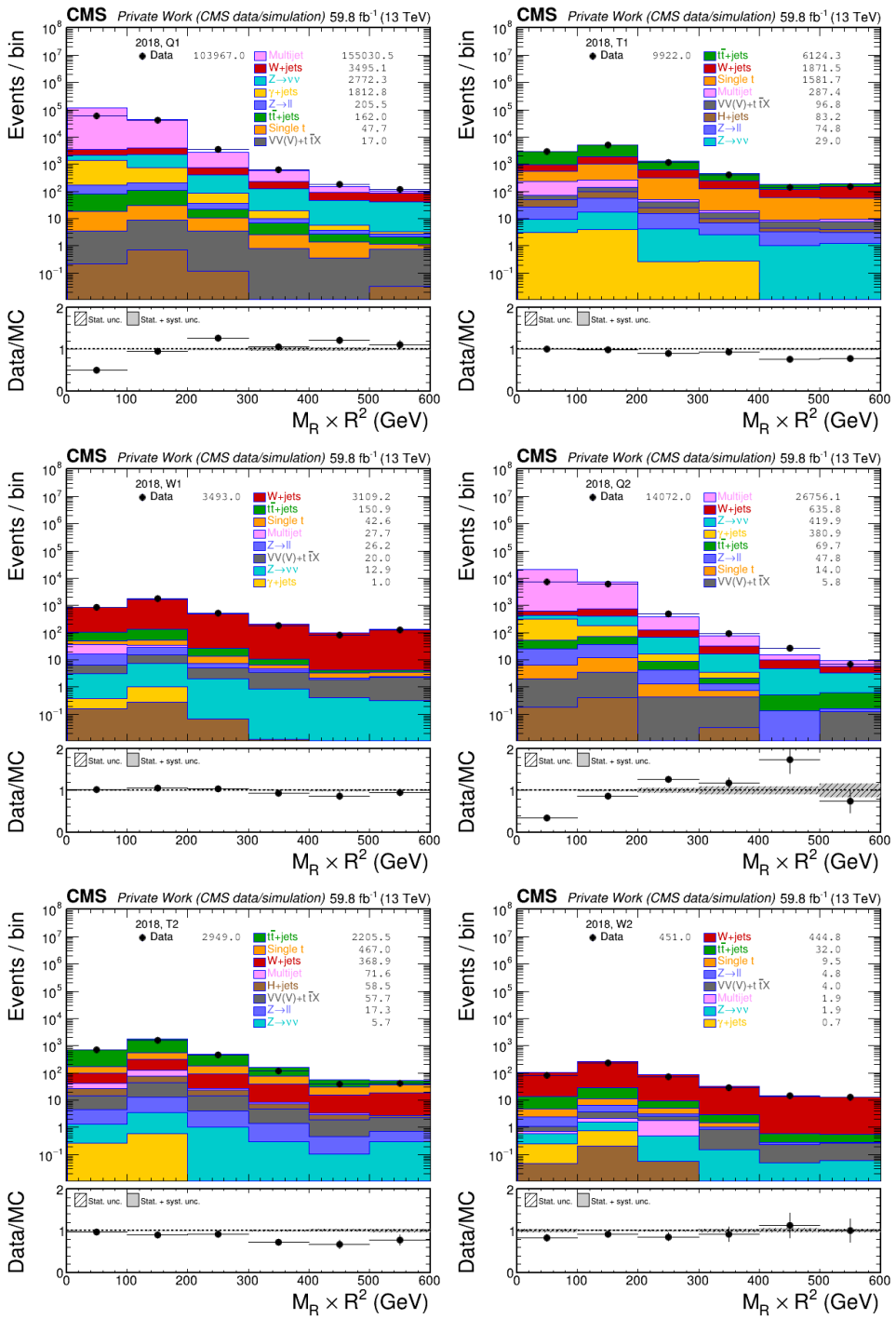


Figure 4.6: Data and MC distributions for $M_R \times R^2$ for the Q1, T1, W1, Q2, T2 and W2 control regions defined in Table 4.6 for the 2018 data taking period.

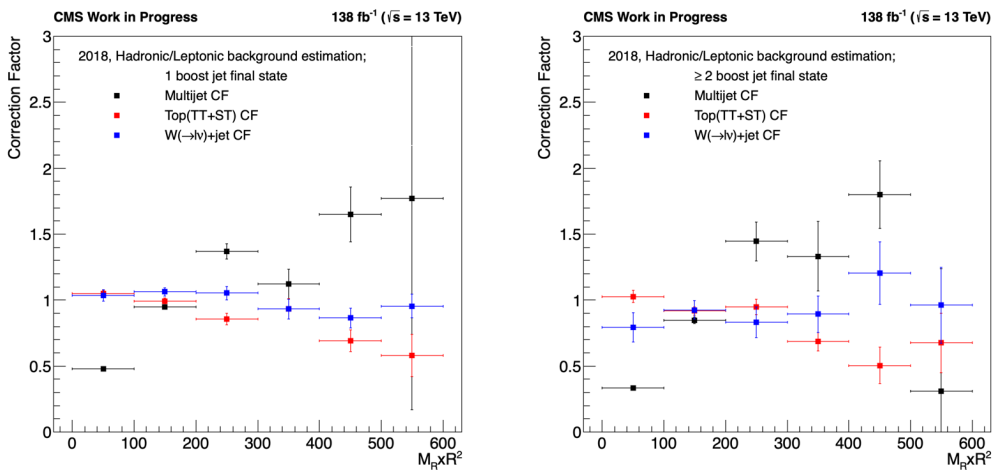


Figure 4.7: Distributions of $M_R \times R^2$ correction factors for QCD, $t\bar{t}$+jets and W+jets MC for cases with exactly 1 or at least 2 mass-tagged AK8 jets.

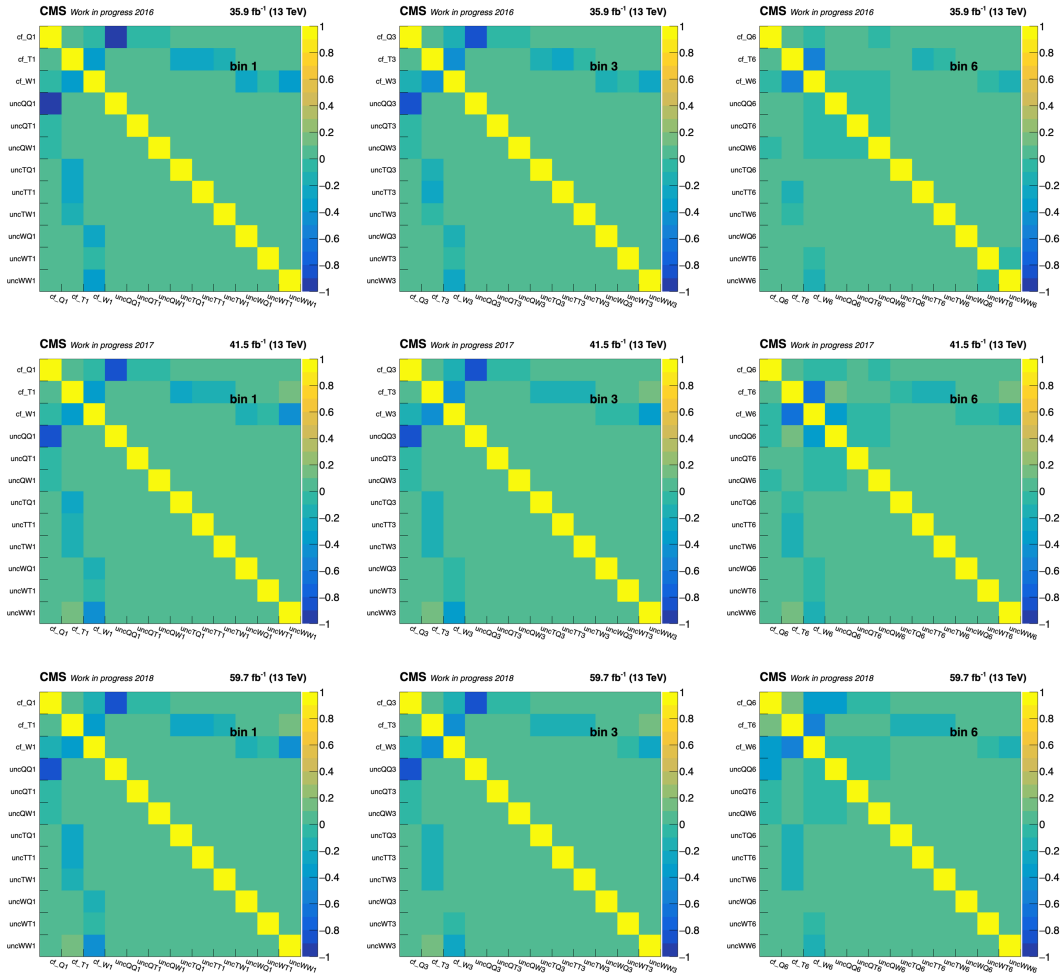


Figure 4.8: $M_R \times R^2$ correlation matrix for QCD, $t\bar{t}$+jets and W+jets MC for 3 different $M_R \times R^2$ bins in 2016(top), 2017(middle), 2018(bottom).

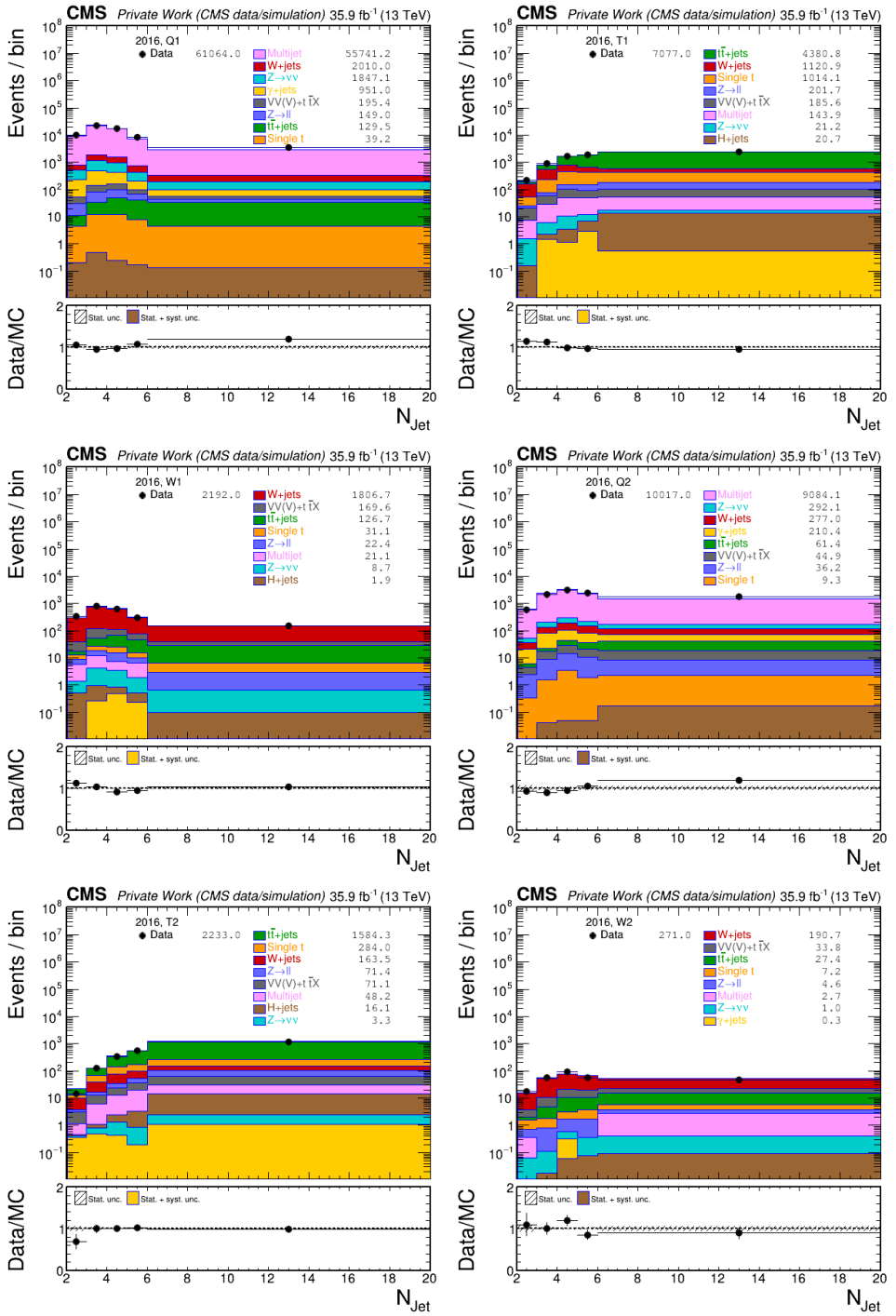


Figure 4.9: Data and MC distributions for AK4 jet multiplicity for the Q1, T1, W1, Q2, T2 and W2 control regions defined in Table 4.6 after event-by-event application of the $M_R \times R^2$ correction factors.

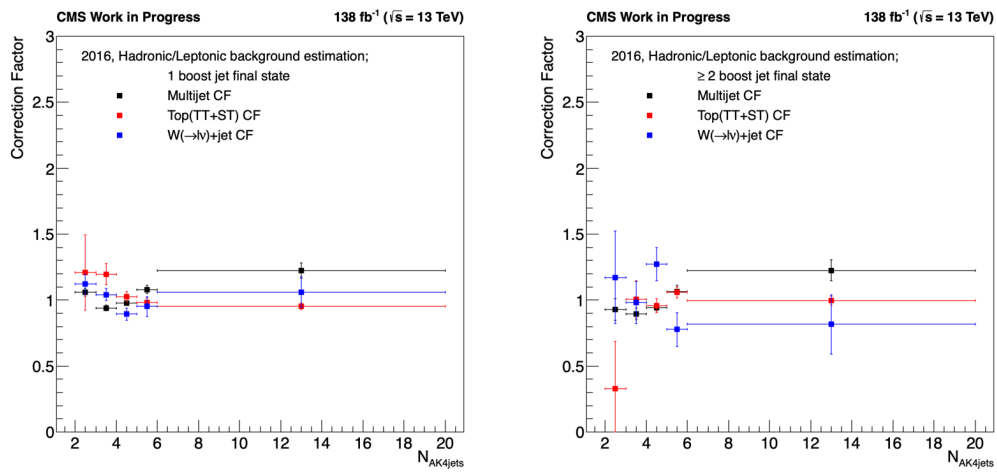


Figure 4.10: Distributions of AK4 jet multiplicity correction factors for QCD, $t\bar{t}$+jets and W+jets MC for cases with exactly 1 or at least 2 mass-tagged AK8 jets.

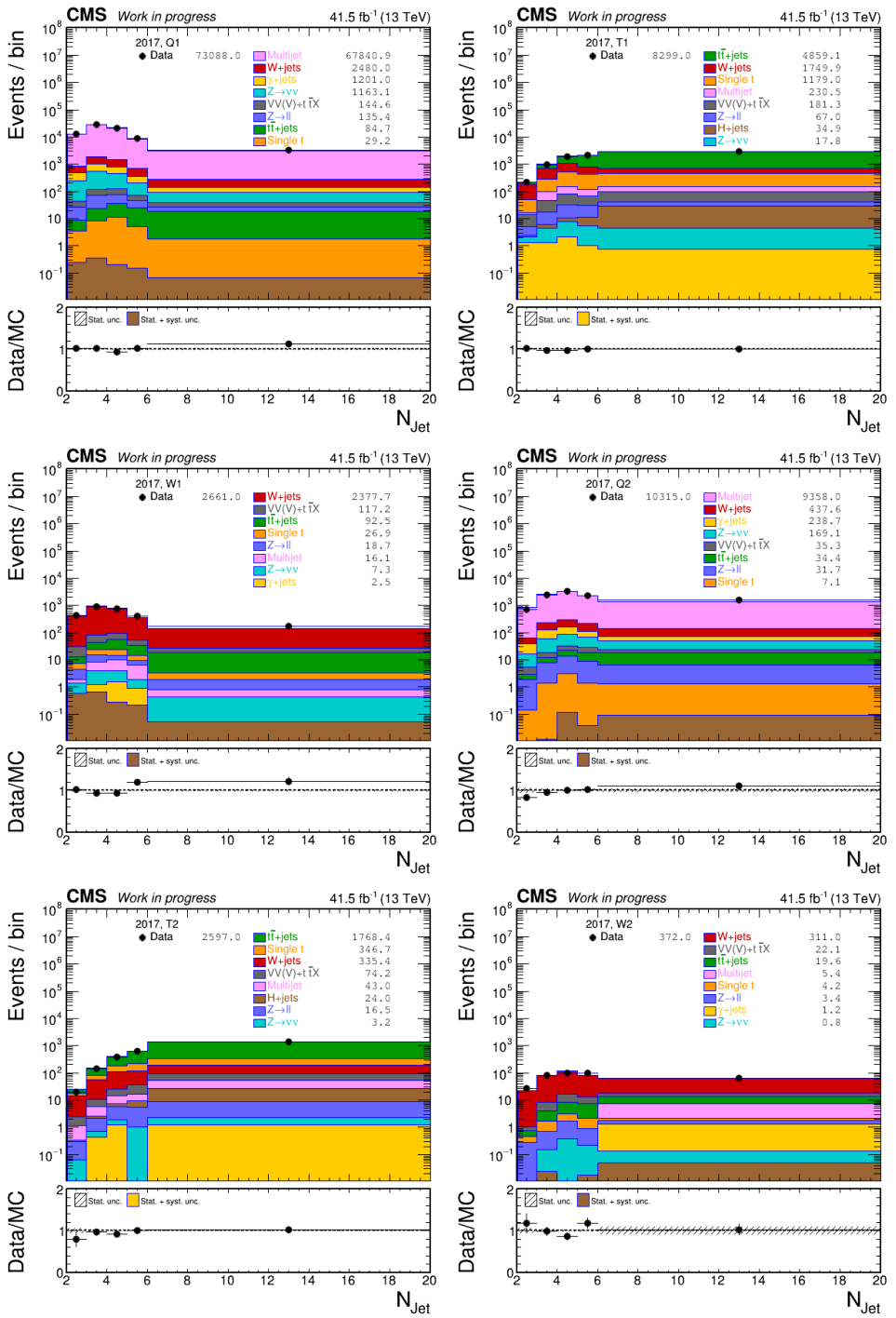


Figure 4.11: Data and MC distributions for AK4 jet multiplicity for the Q1, T1, W1, Q2, T2 and W2 control regions defined in Table 4.6 after event-by-event application of the $M_R \times R^2$ correction factors.

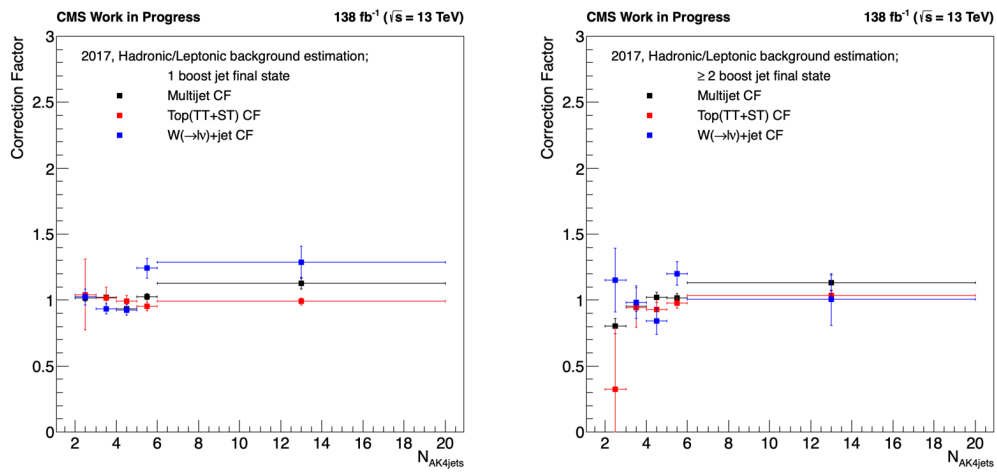


Figure 4.12: Distributions of AK4 jet multiplicity correction factors for QCD, $t\bar{t}$+jets and W+jets MC for cases with exactly 1 or at least 2 mass-tagged AK8 jets.

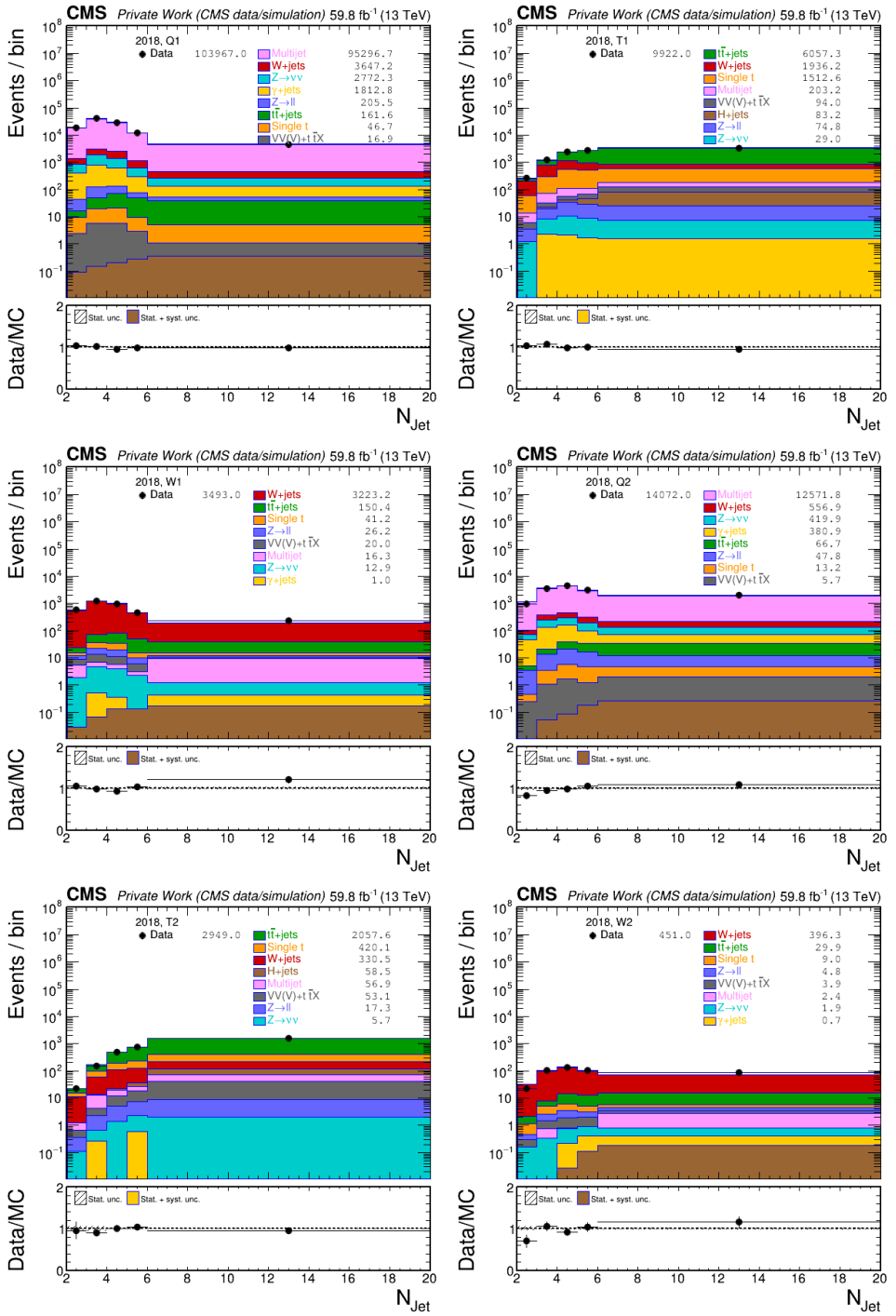


Figure 4.13: Data and MC distributions for AK4 jet multiplicity for the Q1, T1, W1, Q2, T2 and W2 control regions defined in Table 4.6 after event-by-event application of the $M_R \times R^2$ correction factors.

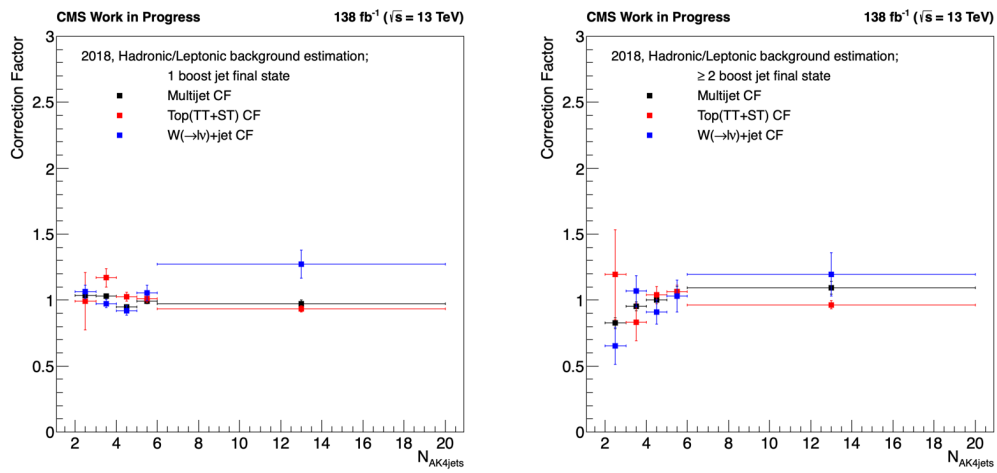


Figure 4.14: Distributions of AK4 jet multiplicity correction factors for QCD, $t\bar{t}$+jets and W+jets MC for cases with exactly 1 or at least 2 mass-tagged AK8 jets.

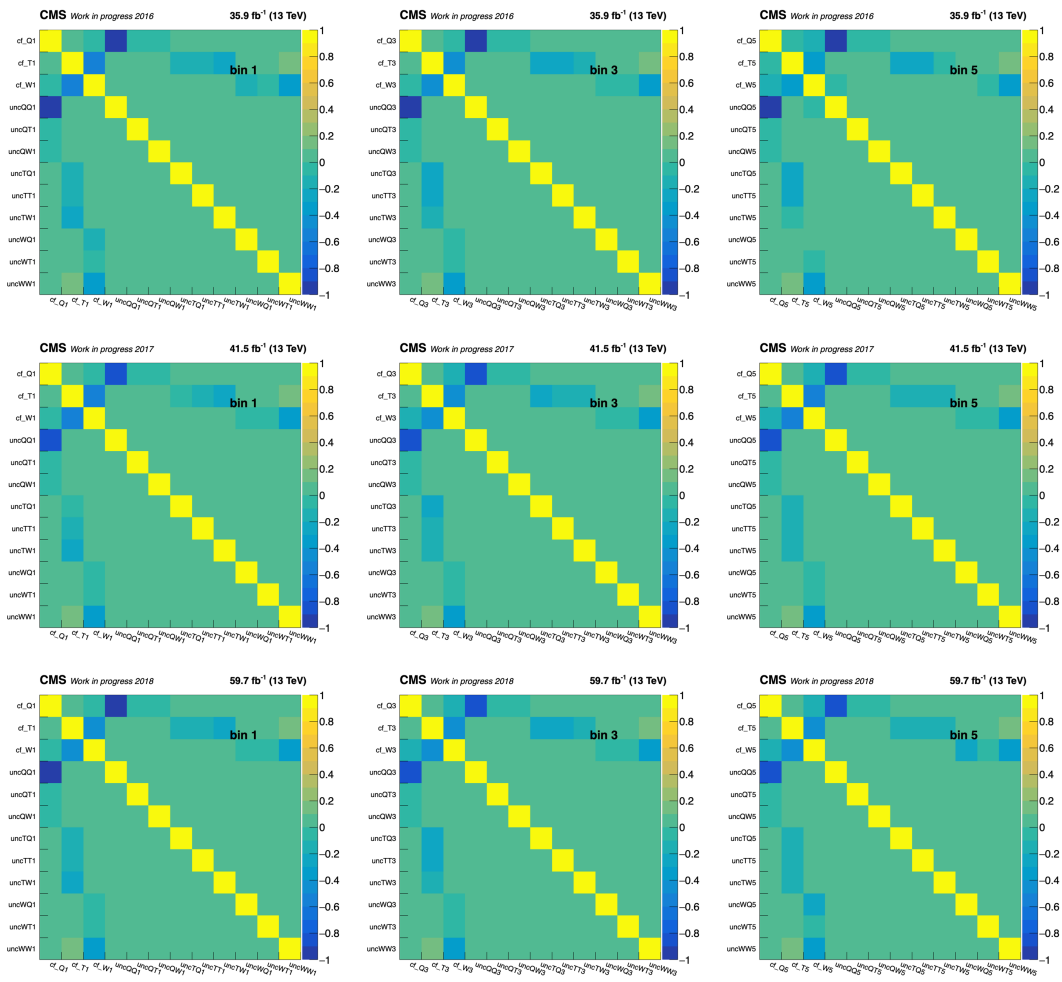


Figure 4.15: AK4 jet multiplicity correlation matrix for QCD, $t\bar{t}$+jets and W+jets MC for 3 different AK4 jet multiplicity bins in 2016(top), 2017(middle), 2018(bottom).

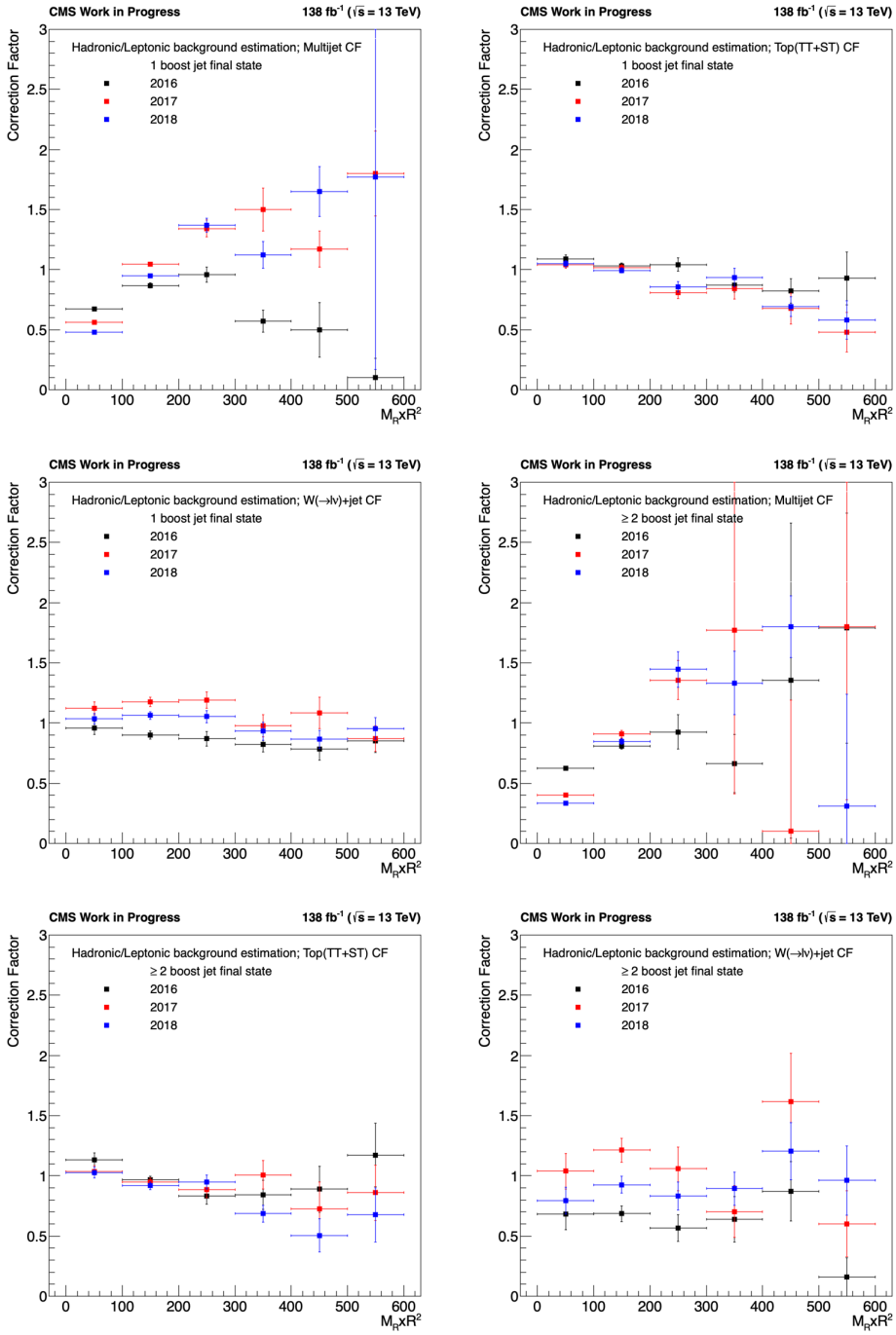


Figure 4.16: Distributions of $M_R \times R^2$ correction factors for QCD, $t\bar{t}$+jets and W+jets MC for cases with exactly 1 or at least 2 mass-tagged AK8 jets in each year.

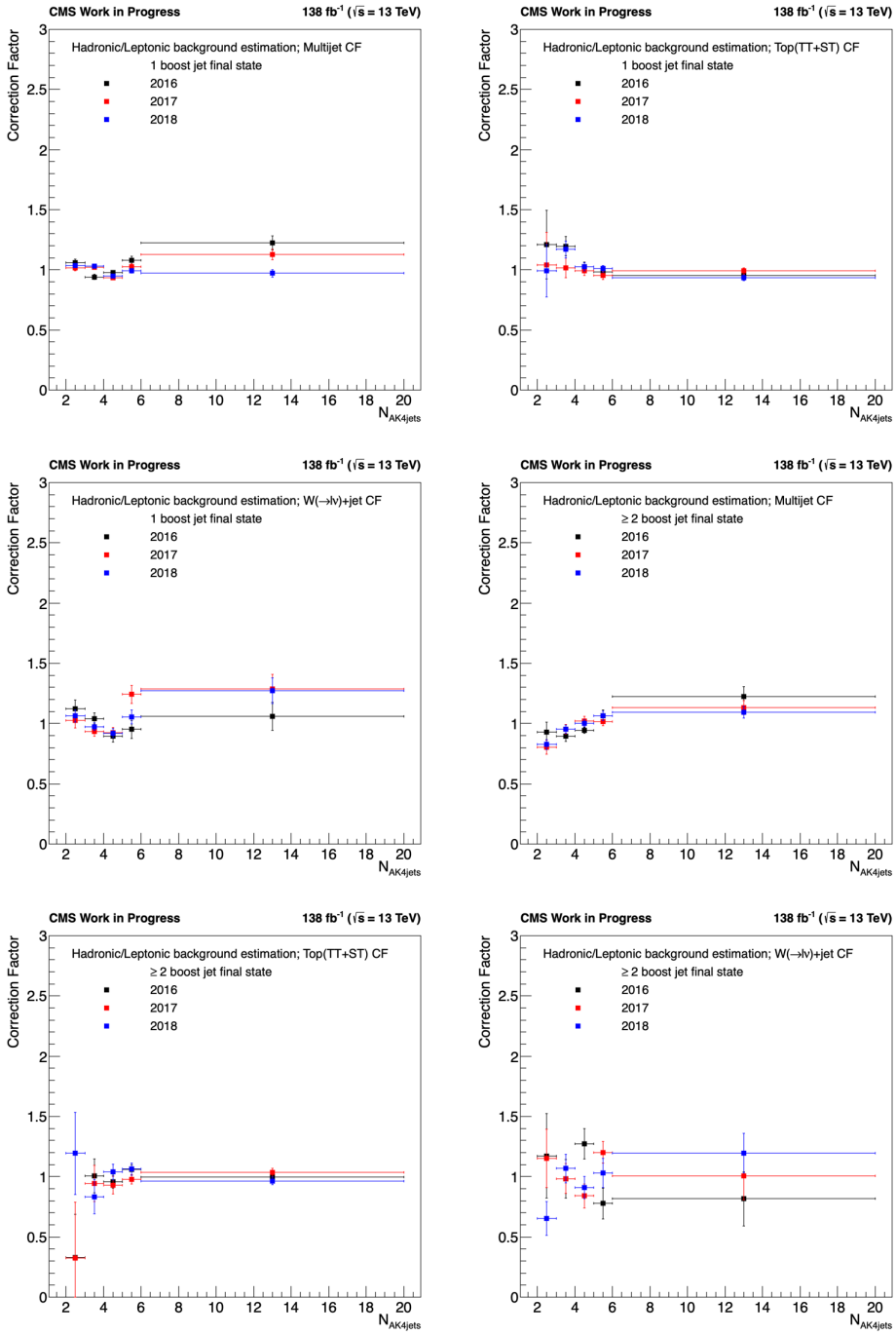


Figure 4.17: Distributions of AK4 jet multiplicity for QCD, $t\bar{t}$+jets and W+jets MC for cases with exactly 1 or at least 2 mass-tagged AK8 jets in each year.

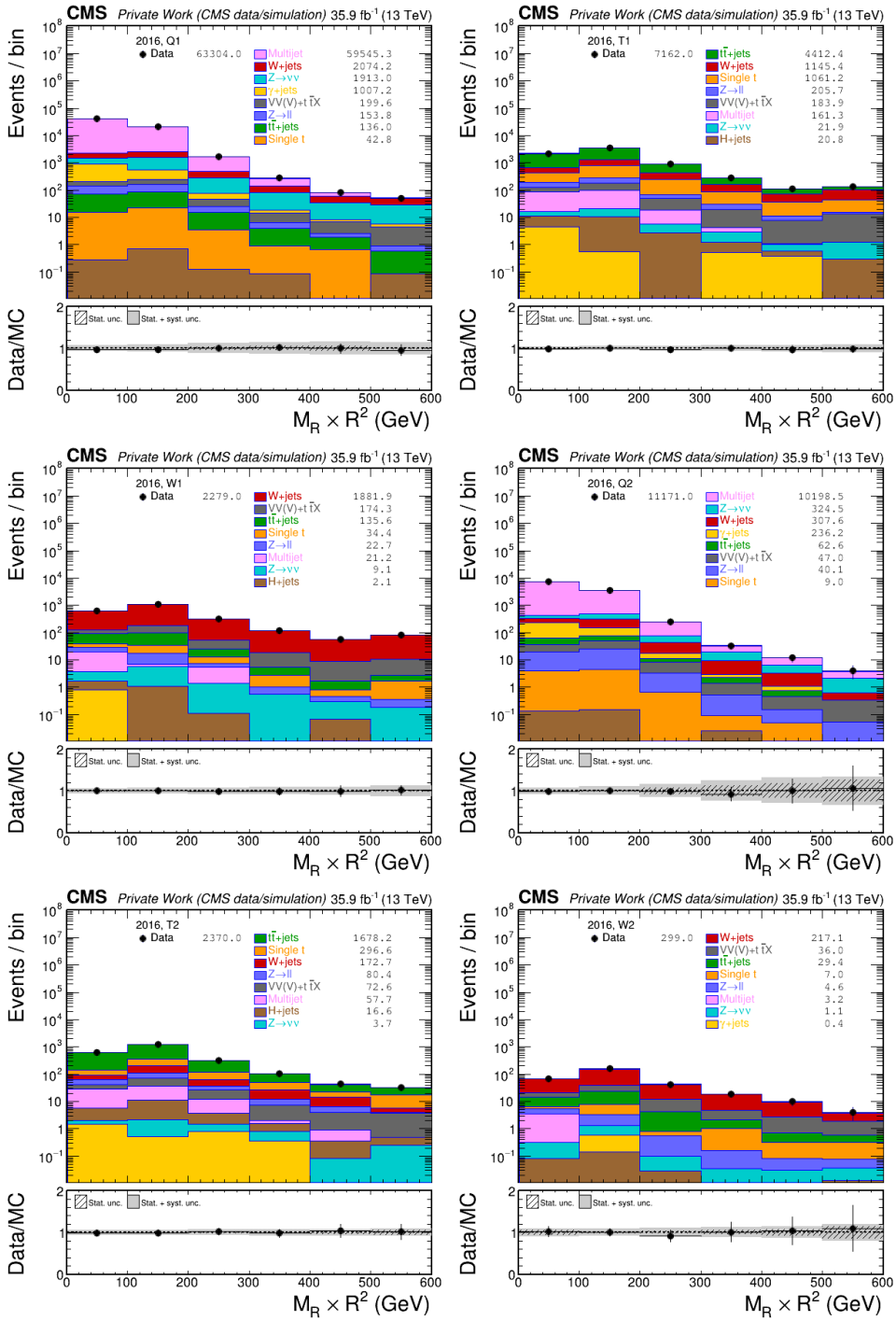


Figure 4.18: Data and MC distributions for $M_R \times R^2$ for the Q1, T1, W1, Q2, T2 and W2 control regions defined in Table 4.6 after event-by-event application of the $M_R \times R^2$ and AK4 jet multiplicity correction factors, shown for the 2016 data taking period.

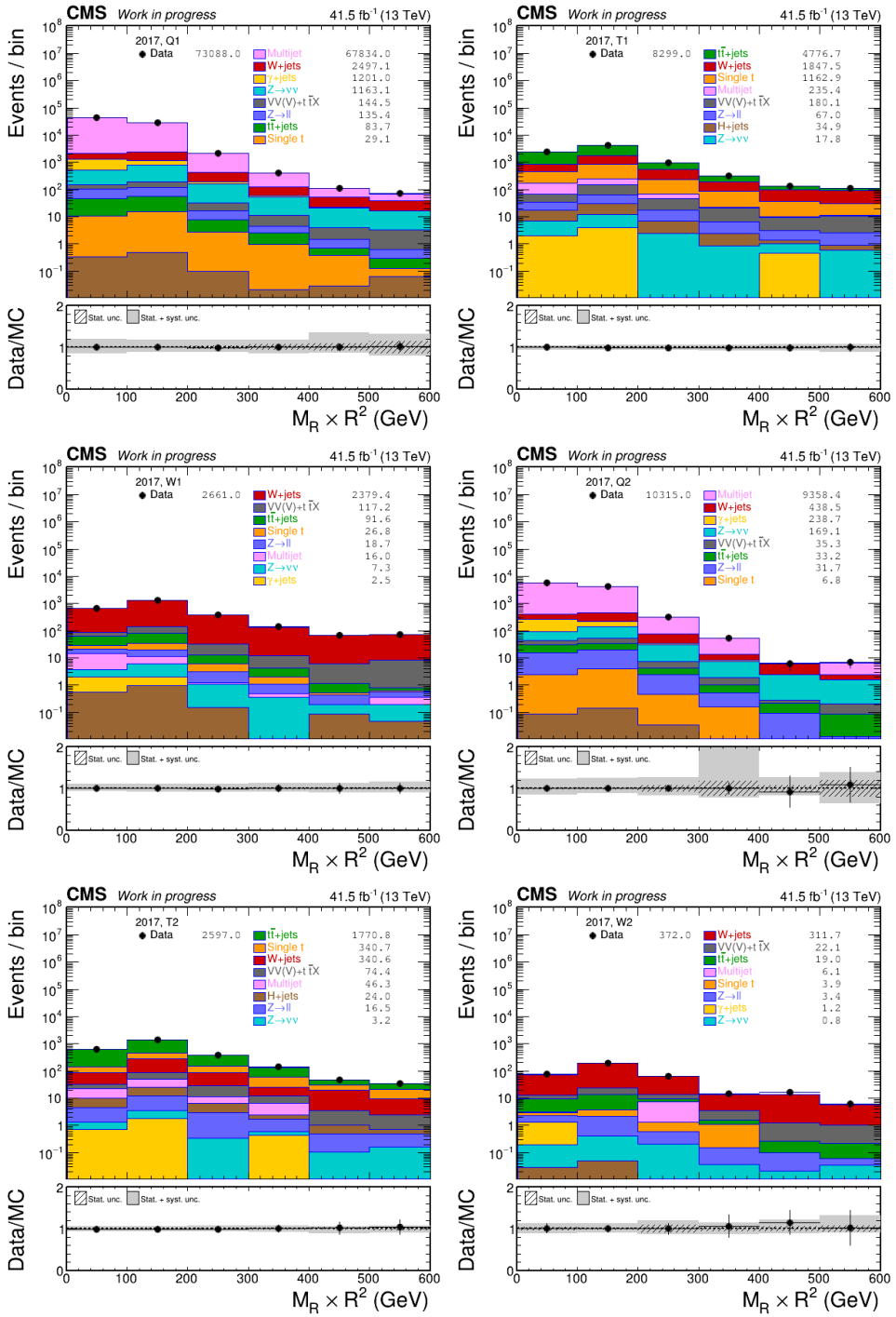


Figure 4.19: Data and MC distributions for $M_R \times R^2$ for the Q1, T1, W1, Q2, T2 and W2 control regions defined in Table 4.6 after event-by-event application of the $M_R \times R^2$ and AK4 jet multiplicity correction factors, shown for the 2017 data taking period.

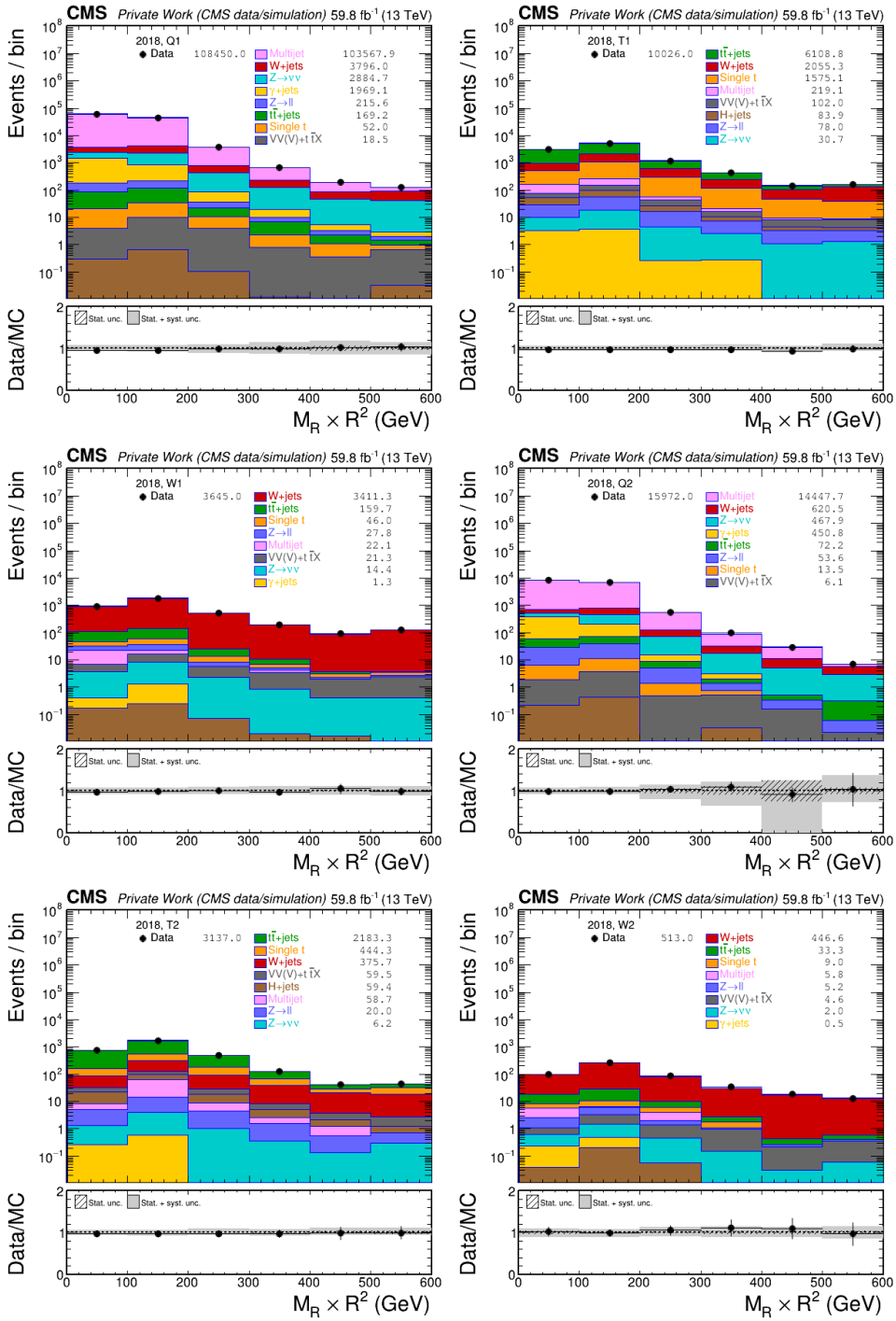


Figure 4.20: Data and MC distributions for $M_R \times R^2$ for the Q1, T1, W1, Q2, T2 and W2 control regions defined in Table 4.6 after event-by-event application of the $M_R \times R^2$ and AK4 jet multiplicity correction factors, shown for the 2018 data taking period.

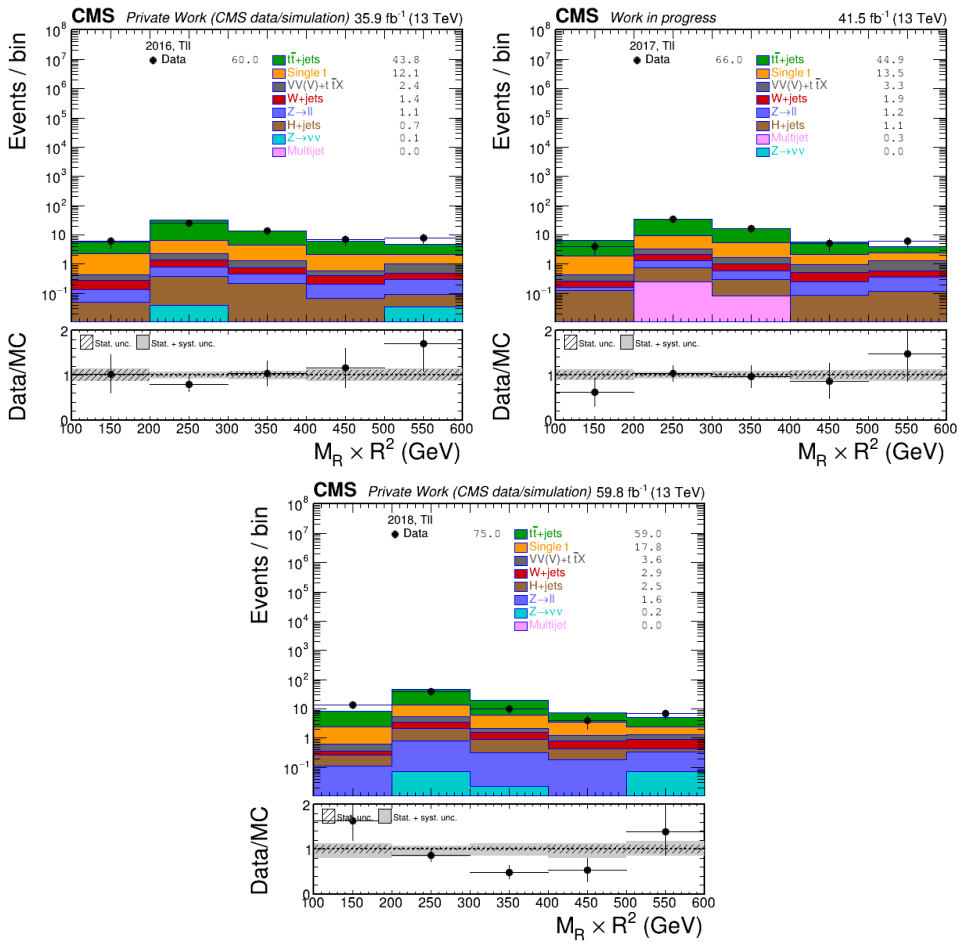


Figure 4.21: Data-MC comparison in the dilepton VR Tll after applying the $t\bar{t}$+jets CFs.

4.2.6 $Z(\rightarrow \nu\nu)$+jets background estimation

There is no selection that can isolate the $Z(\rightarrow \nu\nu)$+jets process which could directly serve as a CR to do an estimate for the other SM backgrounds. However there are other processes that sufficiently resemble $Z(\rightarrow \nu\nu)$+jets , which will be used to derive the CFs for this process. We will follow an established procedure where γ+jets-enriched and DY-enriched CRs are used to derive the CFs to correct the SRs, and where additional W-enriched single lepton CRs are used to calculate a second set of CFs for cross-checks and calculating systematic uncertainties. The following sections present the procedure and results in detail.

Deriving CF_{zinv} from γ+jets and dilepton CRs

The main estimate for the correction factor to be applied to $Z(\rightarrow \nu\nu)$+jets MC events

$$CF_{zinv} = CF_{\gamma} \cdot R^{data/MC}_{Z_{ll}/\gamma} = CF_{\gamma} \cdot \frac{|N^{DY,data}_{zll}|/|N^{DY,MC}_{zll}|}{|N^{G,data}_{\gamma}|/|N^{G,MC}_{\gamma}|} \tag{4.6}$$

is the product of a correction factor CF_{γ} obtained from a photon-enriched CR, and a double ratio $R^{data/MC}_{Z_{ll}/\gamma}$ calculated from a DY-enriched CR and the photon-enriched CR. The double ratio is a constant that corrects for the cross section difference in the γ+jets and $Z(\rightarrow \nu\nu)$+jets processes. CF_{γ}, whose derivation will be explained in detail below, is computed iteratively versus first $M_R \times R^2$, then AK4 jet multiplicity and finally mass-tagged AK8 jet multiplicity. The CF_{γ} to be applied to each event becomes

$$CF_{\gamma} = CF_{\gamma}^{M_R \times R^2} \cdot CF_{\gamma}^{N_{AK4}} \cdot CF_{\gamma}^{N_{AK8(m>50)}} \tag{4.7}$$

Selection for the photon-enriched CR G is listed in Table 4.6. The photon is vectorially added to E_T^{miss} in order to mimic the kinematic properties of $\nu\nu$ coming from the $Z \rightarrow \nu\nu$ decays. Furthermore, jets are removed when they overlap with the photon within $\Delta R < 0.4$ and have a similar p_{T} to that of the photon, i.e. when the photon to jet p_{T} ratio is between 0.5 and 2. This procedure aims to avoid double counting fake jets which would otherwise not have been clustered in the absence of the photon. The modified E_T^{miss} and jets are used as input to M_R and R^2 calculation.

Figure 4.22 shows the $M_R \times R^2$ distribution in the G region. The main process that contributes to the G region apart from γ+jets is QCD multijet, where a jet is misidentified as a photon. The γ+jets events contain both direct photons and fragmentation photons. Instead of relying on simulation for the subtraction of the multijet background, a data-driven measurement of the photon purity is performed. Data in the G regions are scaled with the purity to exclude the multijet background with fake photons and the fraction of fragmentation photons, f, to obtain a pure direct γ+ jets sample. Consequently, CF_γ can be expressed as

$$CF_\gamma = \frac{N^{G,data} \cdot P_\gamma \cdot f}{N_\gamma^{G,MC}}. \tag{4.8}$$

We predict the QCD multijet background contamination in the G region by performing a template fit to the shape of the photon charged isolation variable. The fits are performed in bins of the variable for which the CF is being calculated, i.e. in the first step, for $M_R \times R^2$, then for AK4 jet multiplicity, and finally for mass-tagged AK8 jet multiplicity. Fits are performed separately for the barrel and endcap regions, up to $|\eta| < 2.5$. Templates for the QCD multijet process are obtained for each variable bin from data by inverting the $\sigma_{i\eta i\eta}$ cut (i.e., requiring $\sigma_{i\eta i\eta} > 0.01015(0.0272)$ for the barrel (endcap) photons respectively), in photon identification selection, while applying all the other photon selection cuts. Event selection criteria are taken to be the same as those in the G region. Templates for prompt photons are obtained from MC simulation for the γ+jets process.

Subsequently, purity for each $M_R \times R^2$ bin is computed as

$$P_\gamma = \frac{\text{promt}}{\text{prompt} + \text{fake}}. \tag{4.9}$$

Next, we obtain the fragmentation photon fraction directly from the γ+jets MC. Direct prompt photons are defined as prompt photons with $\Delta R > 0.4$ with respect to any status 23 quarks or gluons, and fragmentation prompt photons are all prompt photons that are not direct. The direct photon distribution thus obtained from the G sample is then used for the $Z(\to \nu\nu)$+jets estimation. We apply a 10% systematic

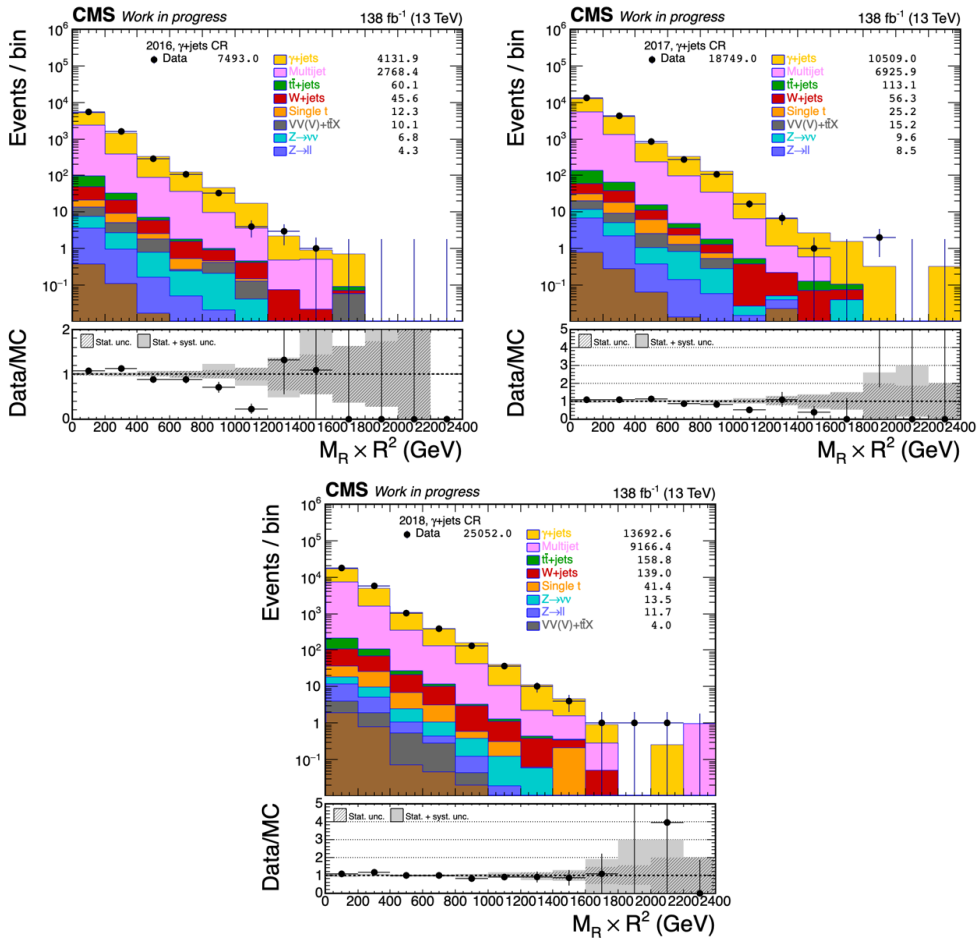


Figure 4.22: Data and MC distributions for $M_R \times R^2$ for the G control region defined in Table 4.6.

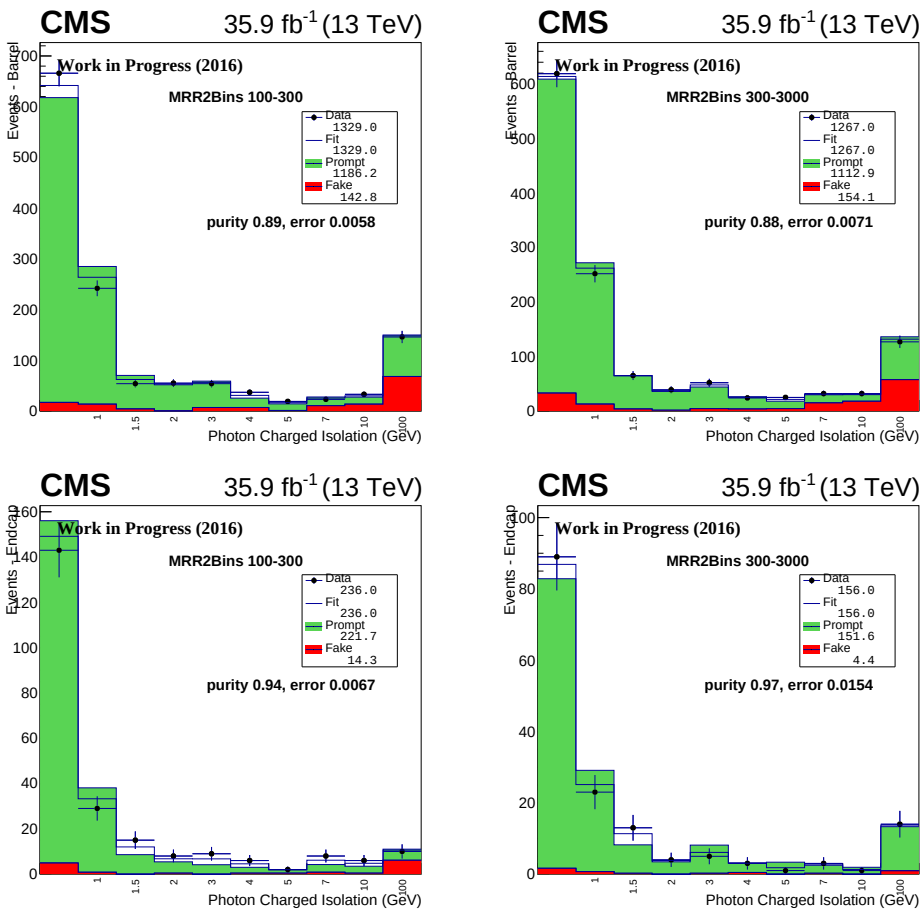


Figure 4.23: Template fits to photon charged isolation in the G region in bins of $M_R \times R^2$ for photons in the barrel (top) and endcap (bottom) section of ECAL, shown for the 2016 data taking period.

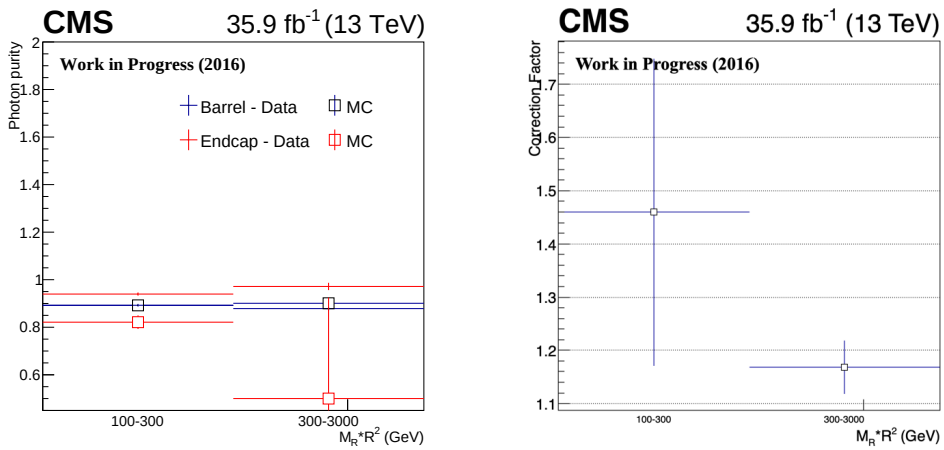


Figure 4.24: Photon purity for the ECAL barrel and endcap sections (right) and γ+jets correction factor versus $M_R \times R^2$ for 2016.

on the fragmentation photon fraction values.

Example fits for barrel and endcap sections are shown in Figure 4.23 for different $M_R \times R^2$ bins. The bin-by-bin values are used for scaling G region data, and a systematic uncertainty of 10% is taken to account for the uncertainty in the purity estimation. The derived purity values and the resulting $CF_\gamma^{M_R \times R^2}$ distributions versus $M_R \times R^2$ are shown in Figure 4.24. They are applied to photon + jets MC. Calculation of purity and CF is then performed in bins of AK4 jet multiplicity and mass-tagged AK8 jet multiplicity, respectively. Figures 4.29 and 4.30 show the template fits, purity and $CF_\gamma^{N_{AK4}}$ for bins of AK4 jet multiplicity. Figures 4.35 and 4.36 show the template fits, purity and $CF_\gamma^{N_{AK8(m>50)}}$ for bins of mass-tagged AK8 jet multiplicity.

We next use a $Z \to ll$ control region DY to correct CF_γ. The DY CR, whose definition is listed in Table 4.6, requires 2 opposite charge same flavor leptons, whose invariant mass lies within 10 GeV of the Z boson mass. For the e^+e^- case, any extra veto muons are vetoed, while for the $\mu^+\mu^-$ case, any extra veto electrons are vetoed. Momentum of the dilepton system is added to E_T^{miss}, which in turn is used for a modified $R^2_{\ell\ell}$ calculation. The number of loosely b-tagged jets is required to be 0 in order to eliminate $t\bar{t}$+jets backgrounds.

Figure 4.41 shows data-MC comparison of the dilepton invariant mass $M_{\ell\ell}$ dis-

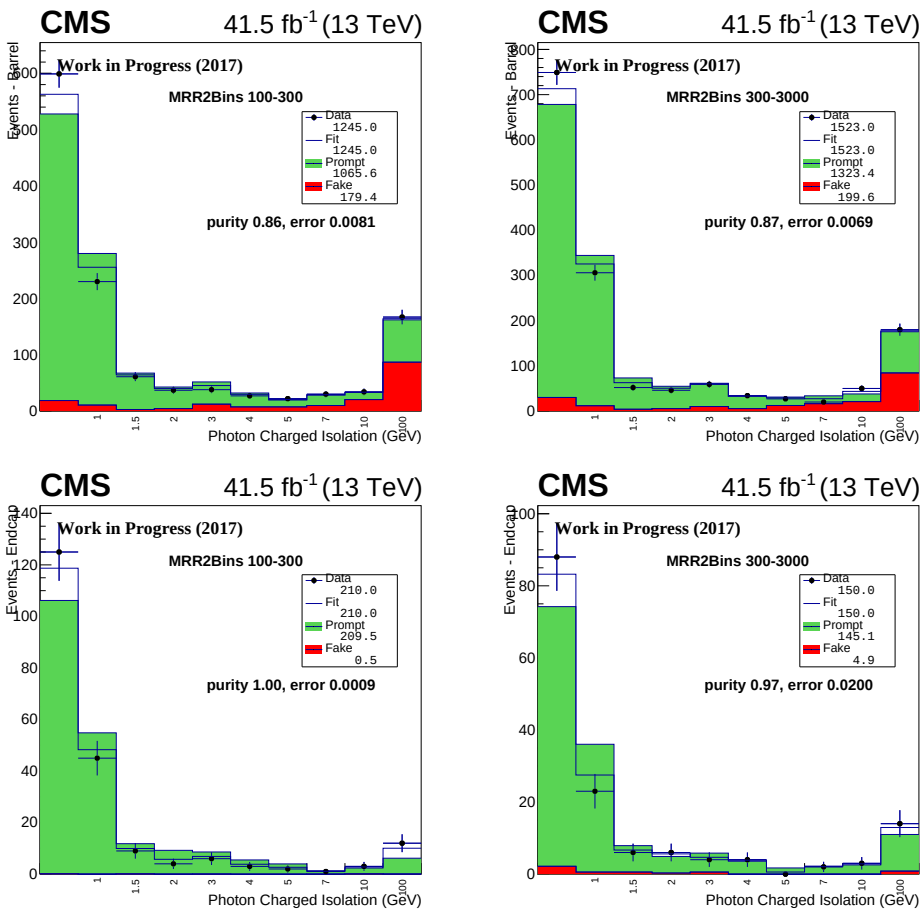


Figure 4.25: Template fits to photon charged isolation in the G region in bins of $M_R \times R^2$ for photons in the barrel (top) and endcap (bottom) section of ECAL, shown for the 2017 data taking period.

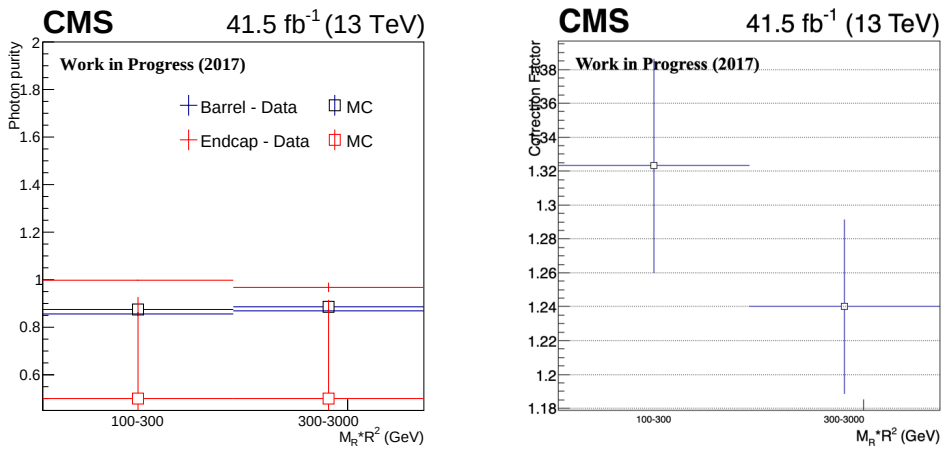


Figure 4.26: Photon purity for the ECAL barrel and endcap sections (right) and γ+jets correction factor versus $M_R \times R^2$ for 2017.

tributions (including both electrons and muons) before the $M_{\ell\ell} - m_Z < 10$ GeV cut. and the $M_R \times R^2_{\ell\ell}$ distribution. We use the data and MC distributions here together with the corresponding distributions from the G region to compute the double ratio

$$R^{data/MC}_{Z_{ll}/\gamma} = \frac{|N^{DY,data}_{zll}|/|N^{DY,MC}_{zll}|}{|N^{G,data}_{\gamma}|/|N^{G,MC}_{\gamma}|} \tag{4.10}$$

The resulting double ratios and uncertainties are as follows:

- 2016: 0.8208±0.16616
- 2017: 1.2931±0.21644
- 2018: 1.0299±0.17469

The data-MC discrepancies are reflected into the double ratio as a systematic uncertainty, which is calculated as the average of the deviations in the M_R and 5 R^2 bins.

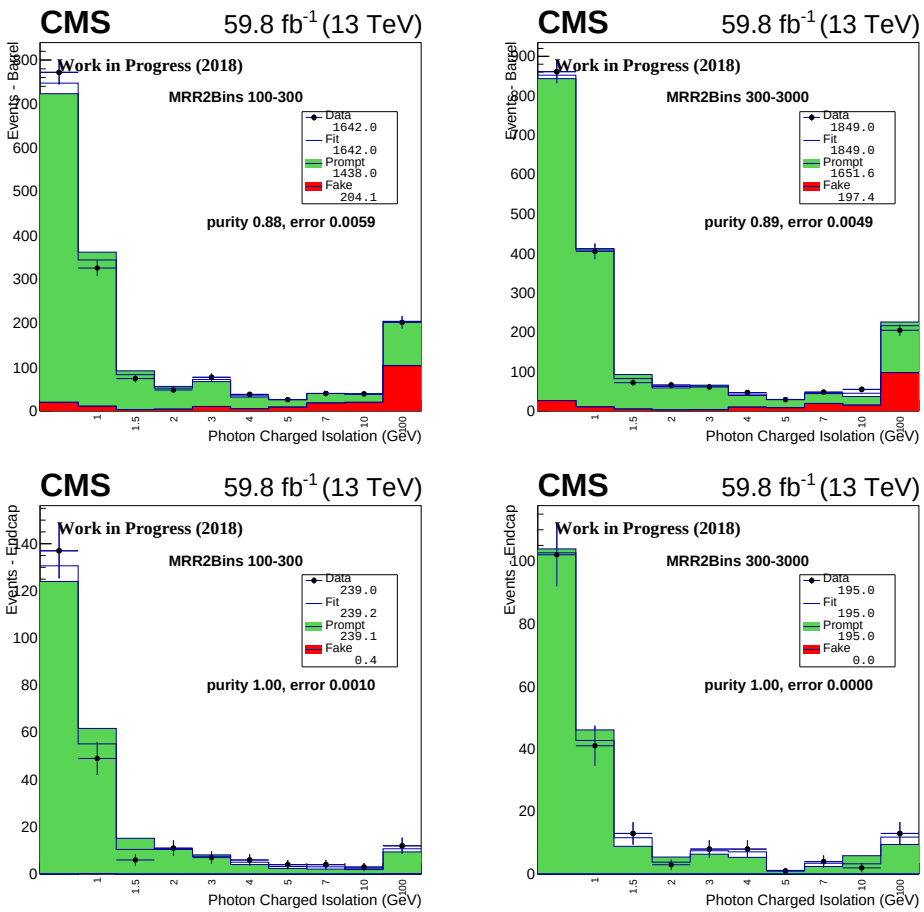


Figure 4.27: Template fits to photon charged isolation in the G region in bins of $M_R \times R^2$ for photons in the barrel (top) and endcap (bottom) section of ECAL, shown for the 2018 data taking period.

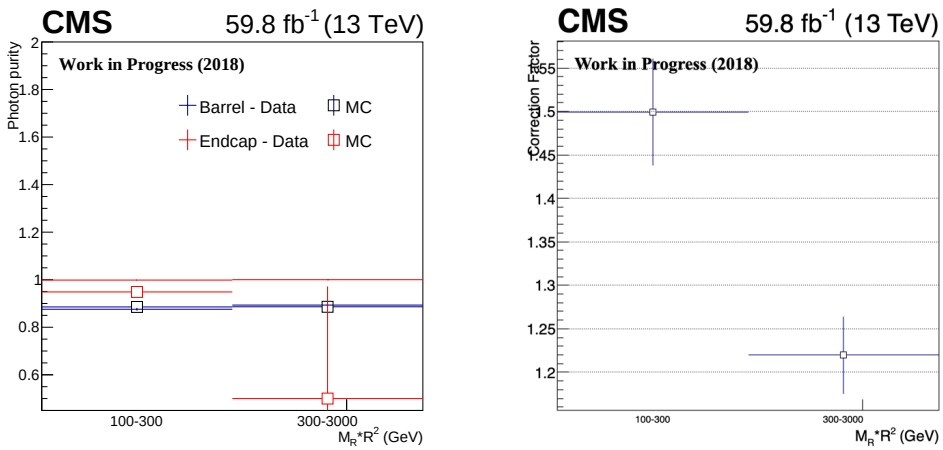


Figure 4.28: Photon purity for the ECAL barrel and endcap sections (right) and γ+jets correction factor versus $M_R \times R^2$ for 2018.

$Z(\to \nu\nu)$+jets background estimation: cross check with $W(\to l\nu)$+jets enriched CRs

In order to validate the estimate from G and DY regions, we perform an alternative $Z(\to \nu\nu)$+ jets estimate from a system of $W(\to l\nu)$+ jets and semileptonic $t\bar{t}$+ jets enriched control regions W4Z and T4Z (symbols being abbreviations of WforZ and TforZ). These CRs, defined separately for cases of = 1 or ≥ 1 mass-tagged AK8 jet, are listed in Table 4.6, and require = 1 lepton and either the existence or veto of a b-tagged jet. The m_T variable is computed with the genuine E_T^{miss} in the event. After $W(\to l\nu)$+jets events are isolated with this selection, the lepton is added to E_T^{miss} to mimic the contribution of the second ν from $Z \to \nu\nu$ decays. The modified E_T^{miss} is then used for recomputing R^2. These selections are the same as the W and T region selections, except for the difference introduced by using a veto lepton and the modified R^2. Distributions of $M_R \times R^2$ are shown for both regions for = 1 and ≥ 1 mass-tagged AK8 jet in Figures 4.42, 4.43 and 4.44 for 3 years of data taking.

The $CF_{Zinv,W}$ is to be obtained from the correction to the W+jets MC. However, since the W4Z regions have visible contribution from top events, we perform a simultaneous calculation using the W4Z and T4Z regions. Two sets of CFs are calculated separately for cases with = 1 and ≥ 2 AK8 jets with mass greater than 50,

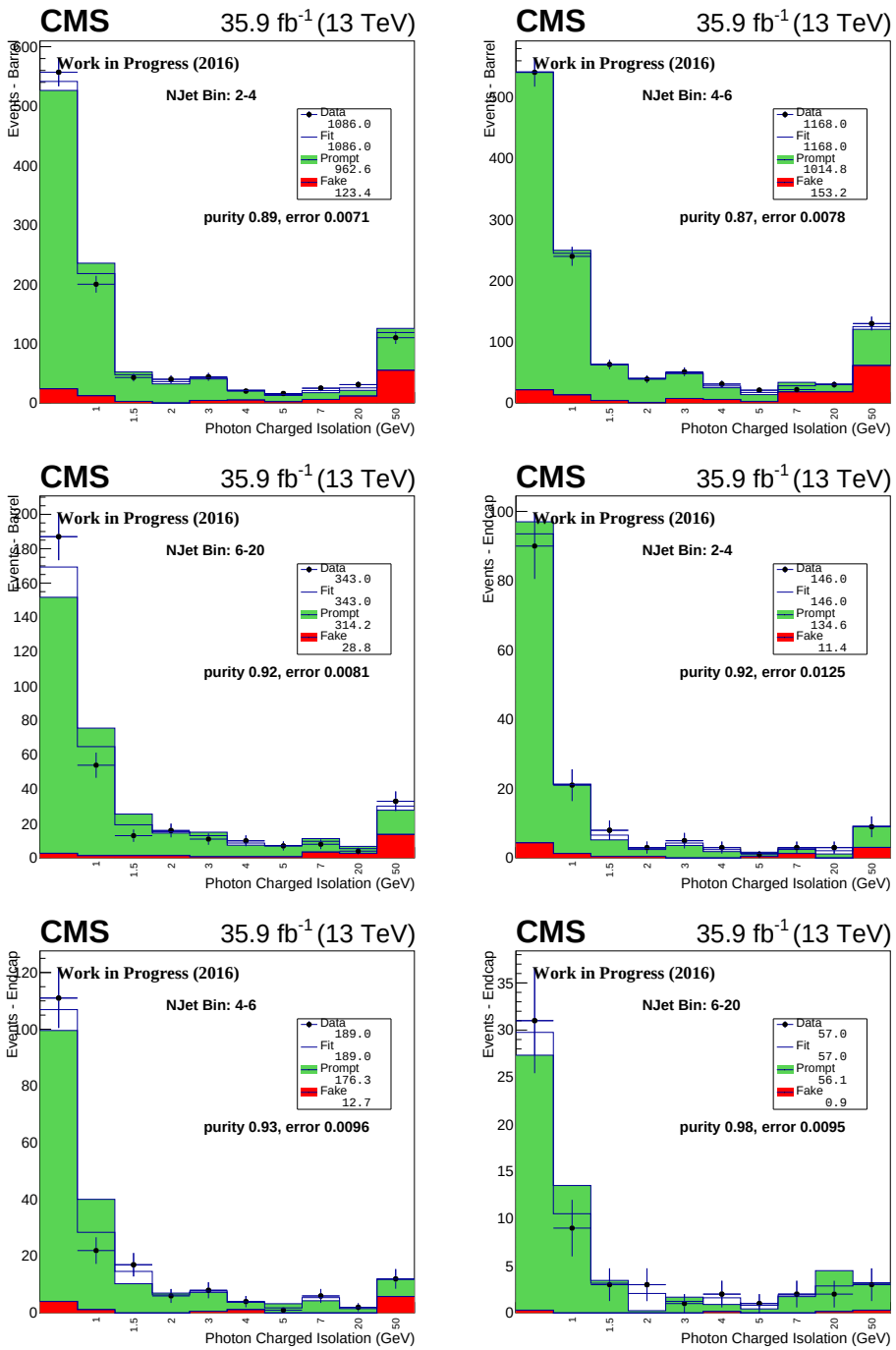


Figure 4.29: Template fits to photon charged isolation in the G region in bins of AK4 jet multiplicity for photons in the barrel (top) and endcap (bottom) section of ECAL, shown for the data taken period 2016.

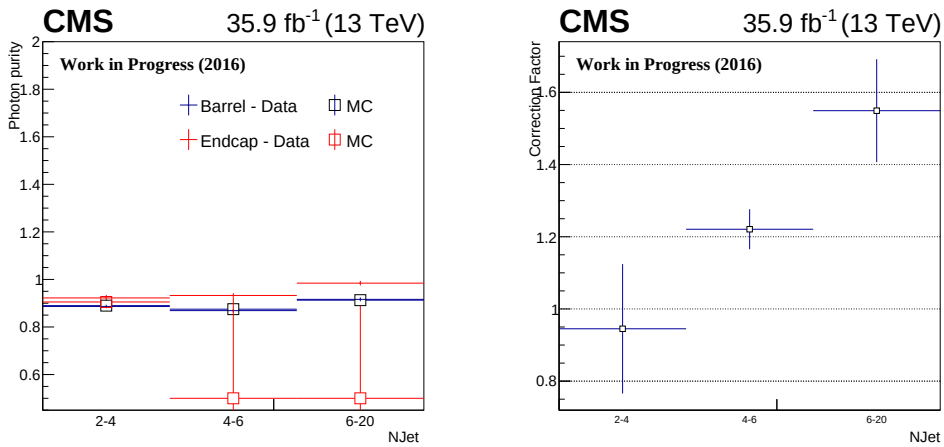


Figure 4.30: Photon purity for the ECAL barrel and endcap sections (right) and γ+jets correction factor versus AK4 jet multiplicity.

using the CR sets W4Z1, T4Z1 and W4Z2, T4Z2. For a given variable and a bin, data and MC quantities in the CRs in each set $i = 1,2$ are related with the following equations:

$$\mathrm{CF}^{\mathrm{i}}_{\mathrm{Zinv,W}}\mathrm{MC}^{\mathrm{W4Zi}}_{\mathrm{Wjets}} + \mathrm{CF}^{\mathrm{i}}_{\mathrm{ttjets}}\mathrm{MC}^{\mathrm{W4Zi}}_{\mathrm{ttjets}} = \mathrm{Data}^{\mathrm{W4Zi}}. \quad (4.11)$$

$$\mathrm{CF}^{\mathrm{i}}_{\mathrm{Zinv,W}}\mathrm{MC}^{\mathrm{T4Zi}}_{\mathrm{Wjets}} + \mathrm{CF}^{\mathrm{i}}_{\mathrm{ttjets}}\mathrm{MC}^{\mathrm{T4Zi}}_{\mathrm{ttjets}} = \mathrm{Data}^{\mathrm{T4Zi}} \quad (4.12)$$

$$(4.13)$$

where we simultaneously solve for $CF^{i}_{Zinv,W}$ and $CF^{i}_{tt\,jets}$. Figure 4.45 shows the correction factors versus $M_R \times R^2$. Next, these CFs are applied to the MC in the CRs and CFs are computed versus AK4 jet multiplicity. Figures 4.48, 4.49, and 4.50 show the AK4 jet multiplicity in the 4 CRs, and Figure 4.51 shows the resulting AK4 jet CFs. We emphasize that $CF_{Zinv,W}$ is the CF to be used for correcting the $Z(\rightarrow \nu\nu)$+jets MC for the cross check. $CF_{tt\,jets}$ is only a product of the simultaneous use of W and top-enriched CRs, and is not used in the analysis.

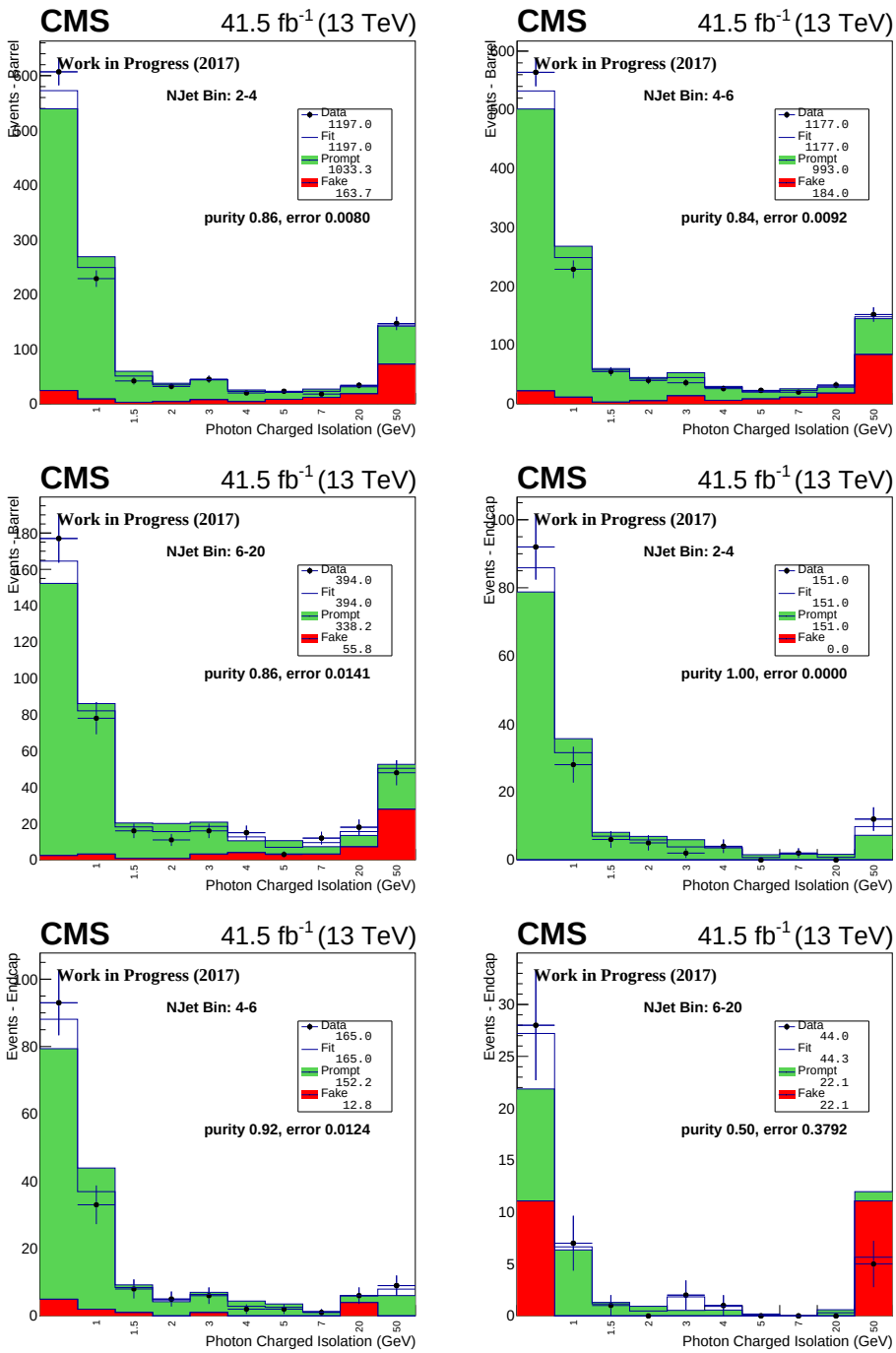


Figure 4.31: Template fits to photon charged isolation in the G region in bins of AK4 jet multiplicity for photons in the barrel (top) and endcap (bottom) section of ECAL, shown for the data taken period 2017.

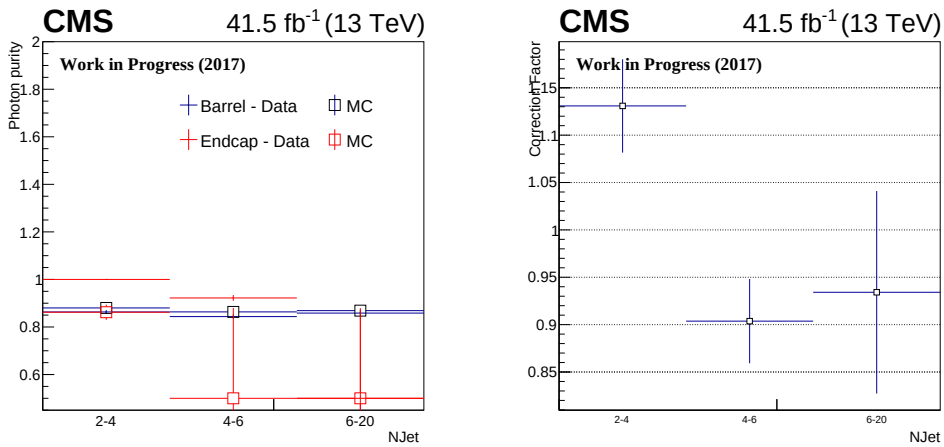


Figure 4.32: Photon purity for the ECAL barrel and endcap sections (right) and γ+jets correction factor versus AK4 jet multiplicity for the data taken 2017.

$Z(\rightarrow \nu\nu)$+jets background estimation results

Figure 4.54 shows the collective results of the $Z(\rightarrow \nu\nu)$+jets estimate. It compares estimates from the G and Z regions, L region and directly from MC simulation. Overall, there is good agreement between the data driven estimates. Estimate from the G and Z regions will be used as a central value, whereas its difference with the estimate from the L region will be used as a systematic.

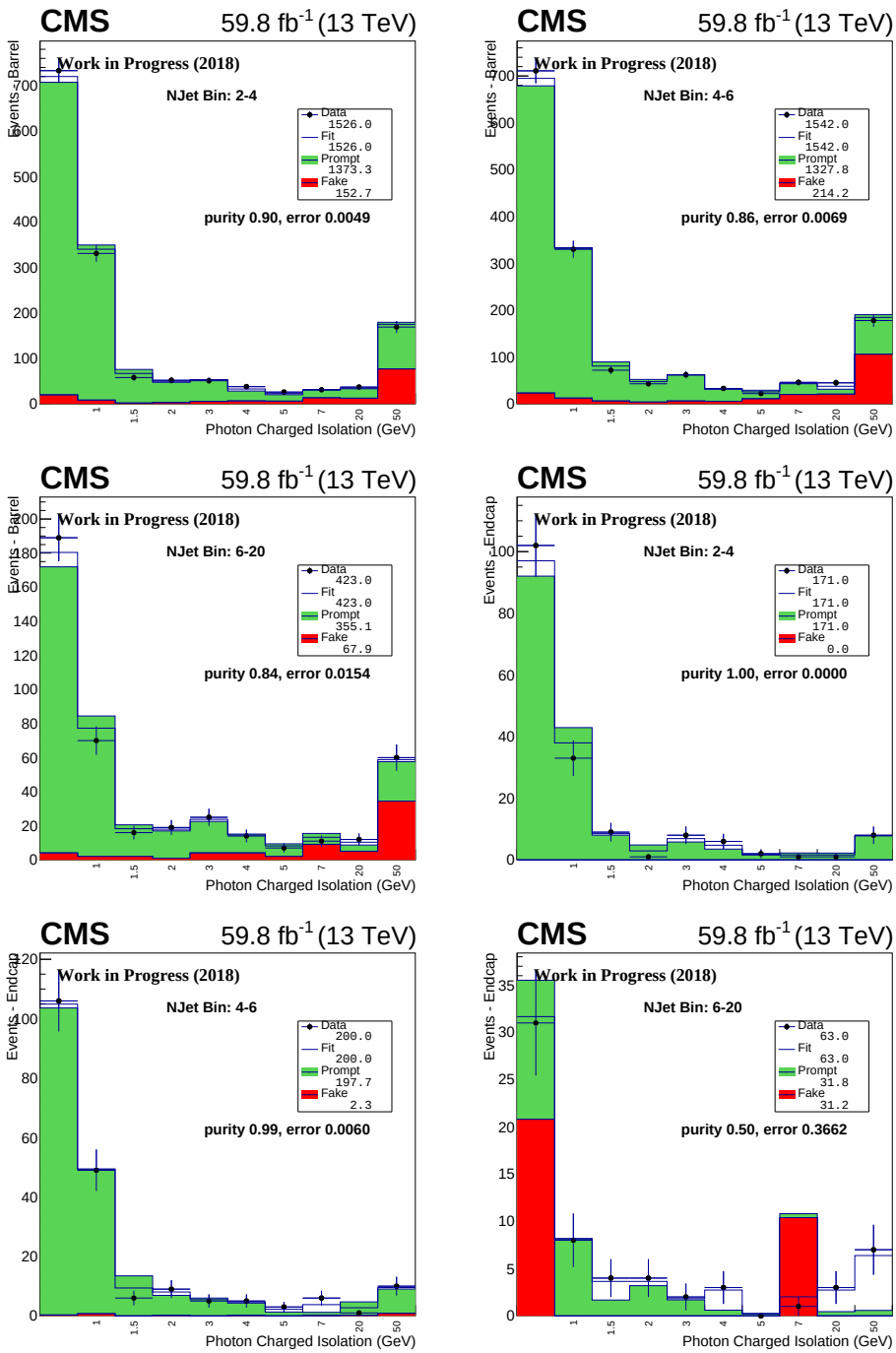


Figure 4.33: Template fits to photon charged isolation in the G region in bins of AK4 jet multiplicity for photons in the barrel (top) and endcap (bottom) section of ECAL, shown for the data taken period 2018.

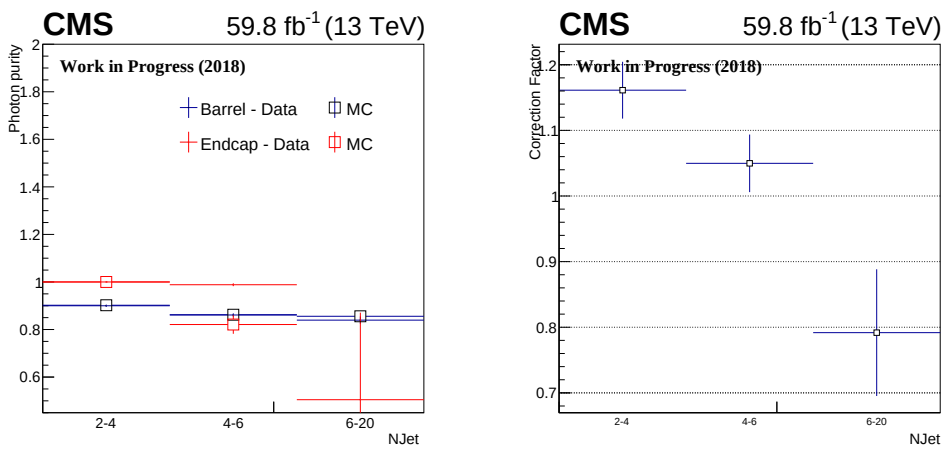


Figure 4.34: Photon purity for the ECAL barrel and endcap sections (right) and γ+jets correction factor versus AK4 jet multiplicity for the data taken 2018.

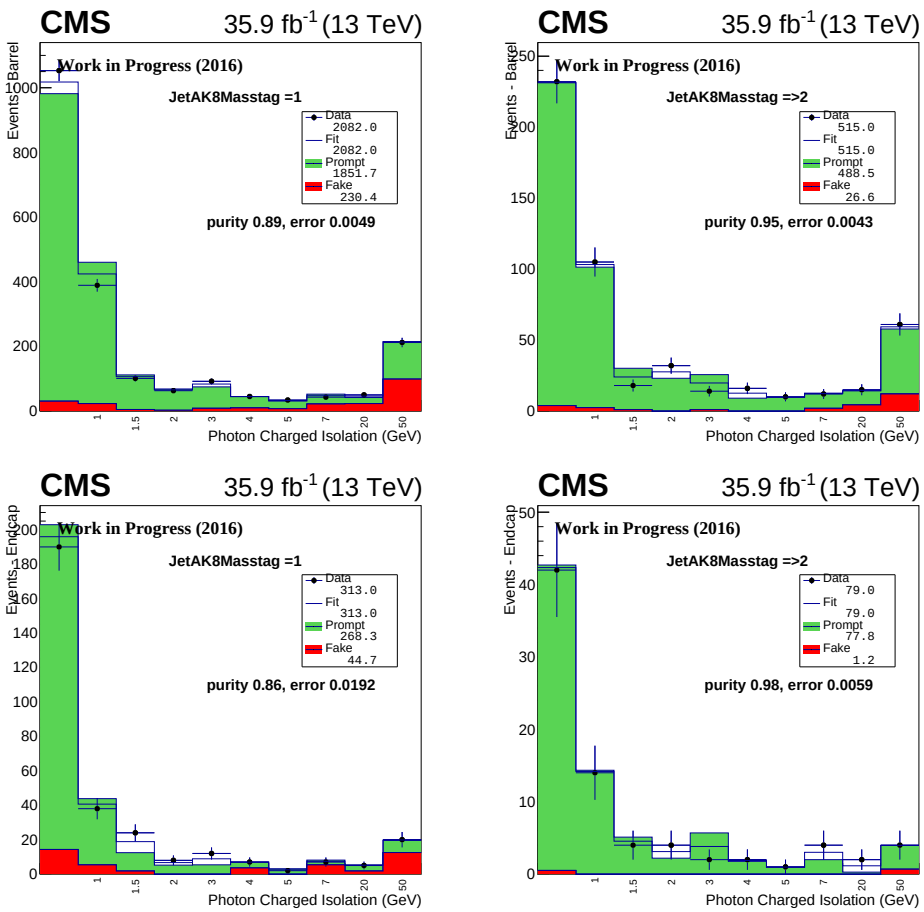


Figure 4.35: Template fits to photon charged isolation in the G region in bins of AK8 jet multiplicity for photons in the barrel (top) and endcap (bottom) section of ECAL, shown for the 2016 data taking period.

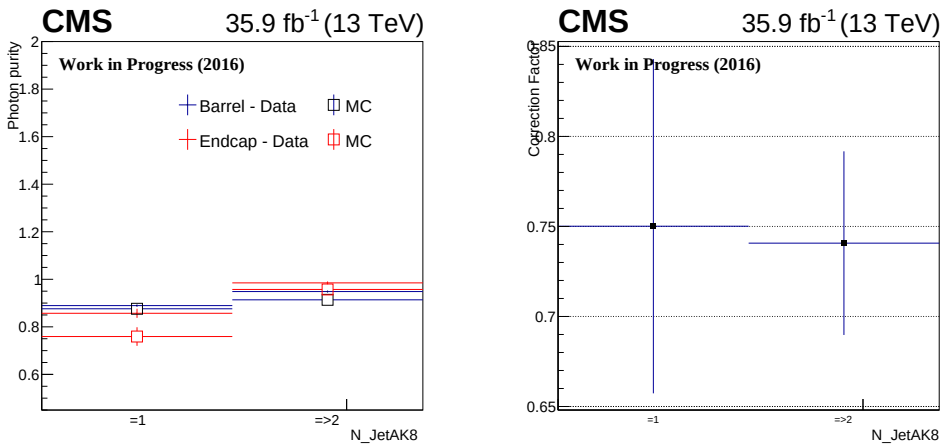


Figure 4.36: Photon purity for the ECAL barrel and endcap sections (right) and γ+jets correction factor versus AK8 jet multiplicity for the 2016 data taking period.

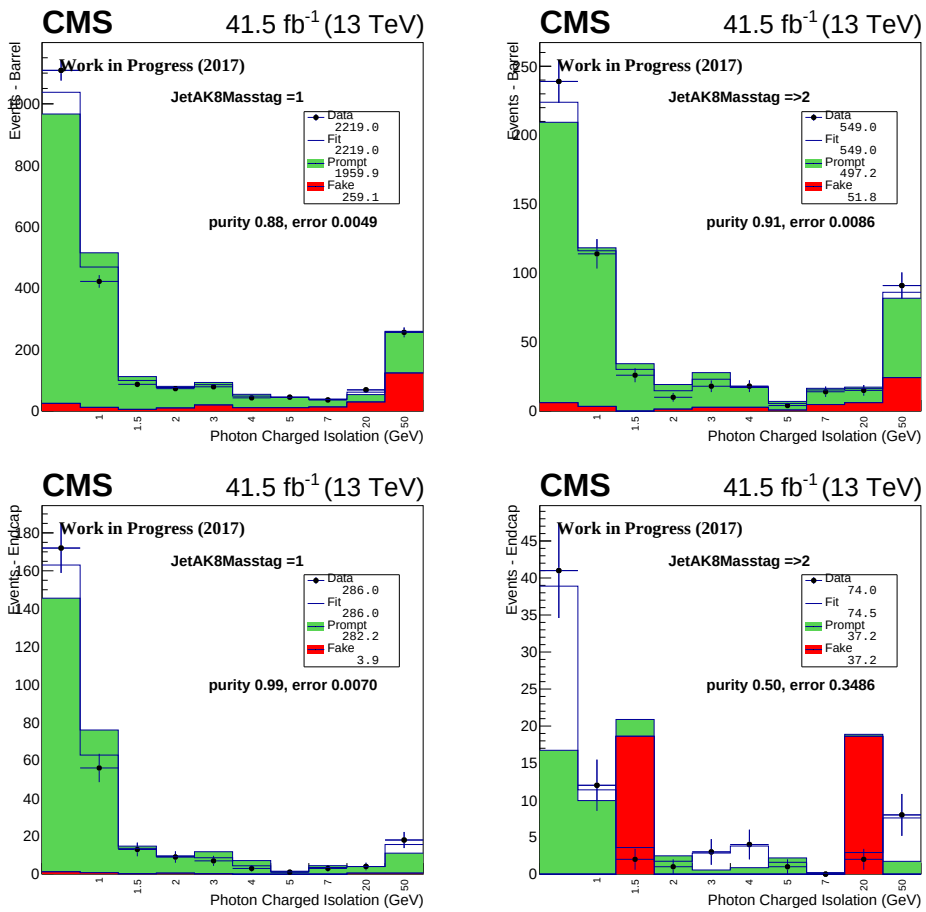


Figure 4.37: Template fits to photon charged isolation in the G region in bins of AK8 jet multiplicity for photons in the barrel (top) and endcap (bottom) section of ECAL, shown for the 2017 data taking period.

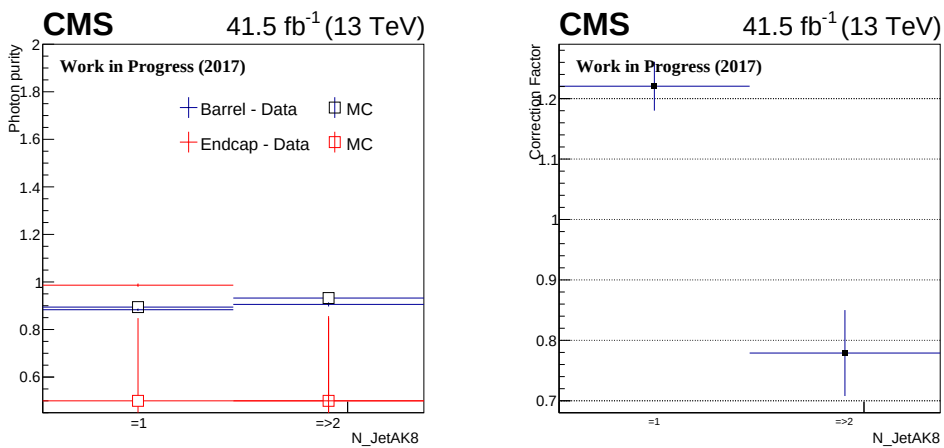


Figure 4.38: Photon purity for the ECAL barrel and endcap sections (right) and γ+jets correction factor versus AK8 jet multiplicity for 2017 data taking period.

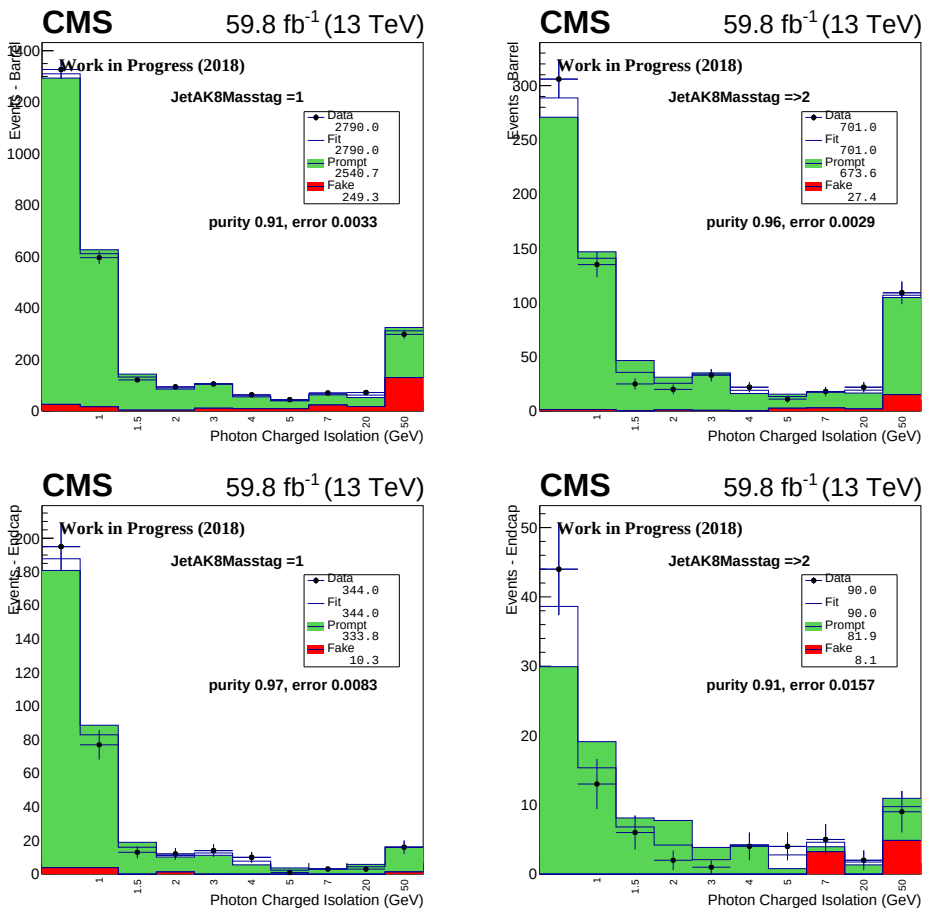


Figure 4.39: Template fits to photon charged isolation in the G region in bins of AK8 jet multiplicity for photons in the barrel (top) and endcap (bottom) section of ECAL, shown for the 2018 data taking period.

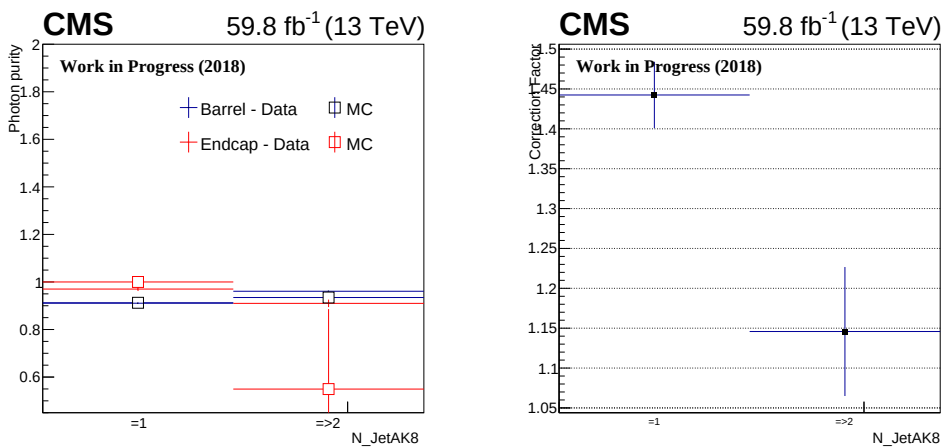


Figure 4.40: Photon purity for the ECAL barrel and endcap sections (right) and γ+jets correction factor versus AK8 jet multiplicity for 2018 data taking period.

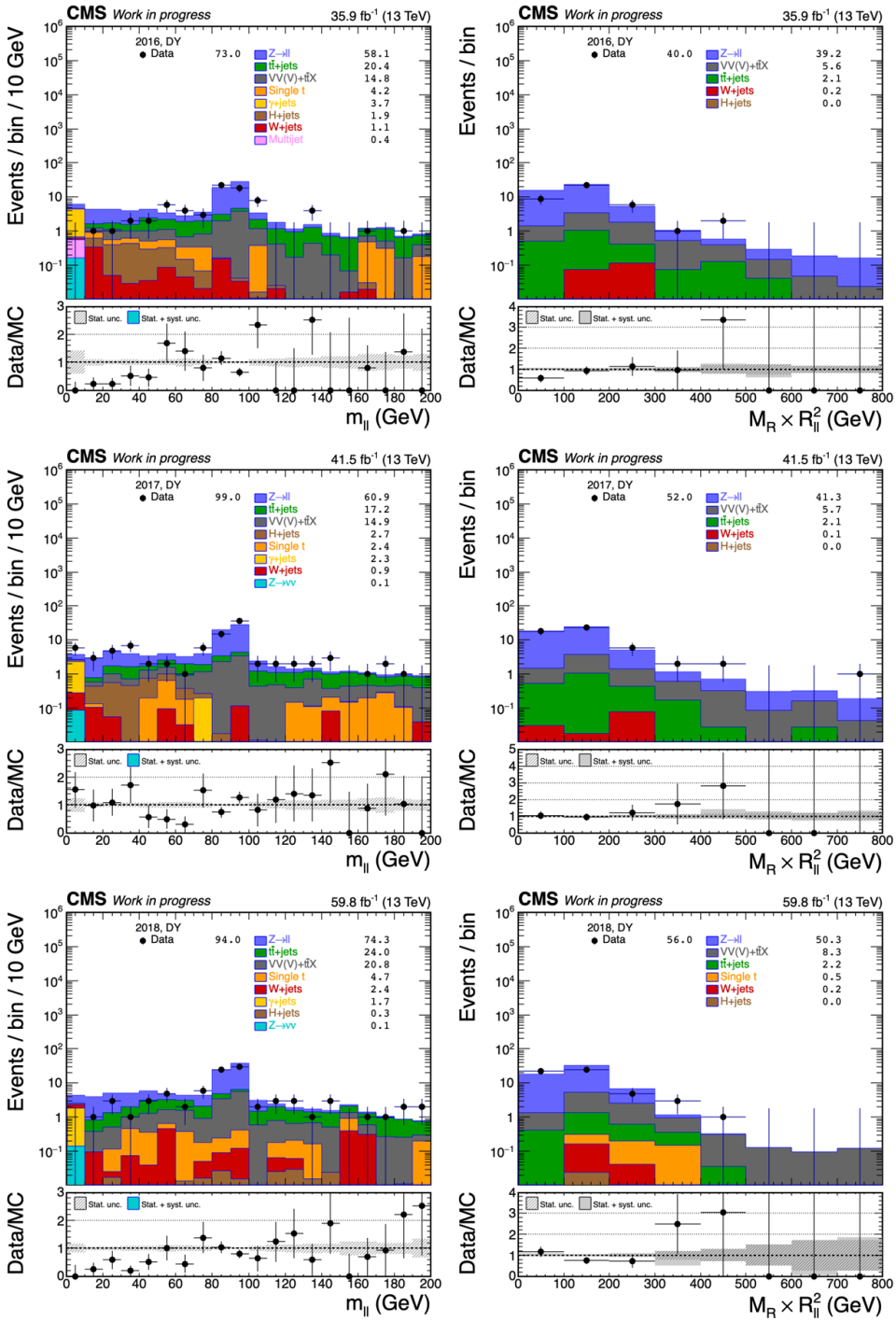


Figure 4.41: Data and MC $N-1$ distributions for the dilepton invariant mass $M_{\ell\ell}$ (left) and $M_R \times R^2$ (right) in the DY control region defined in Table 4.6.

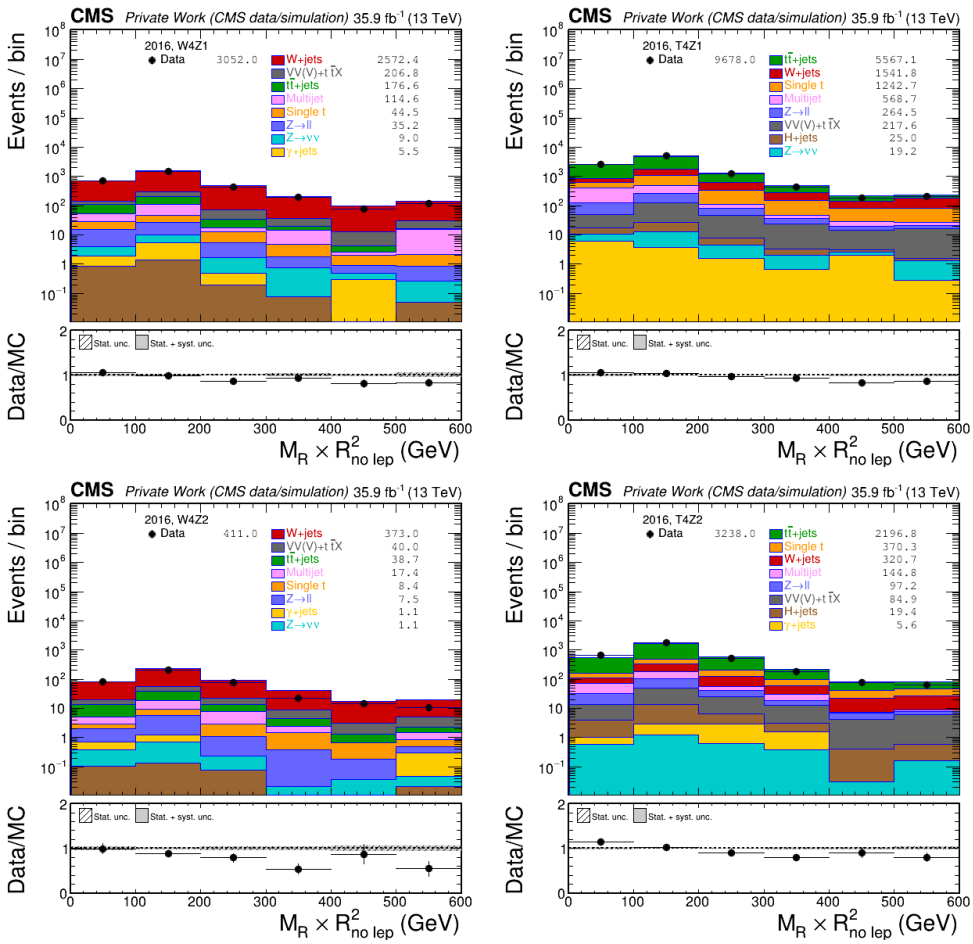


Figure 4.42: Data and MC distributions for $M_R \times R^2$ for the W4Z1, T4Z1, W4Z2, T4Z2 control regions defined in Table 4.6 for the 2016 data taking period.

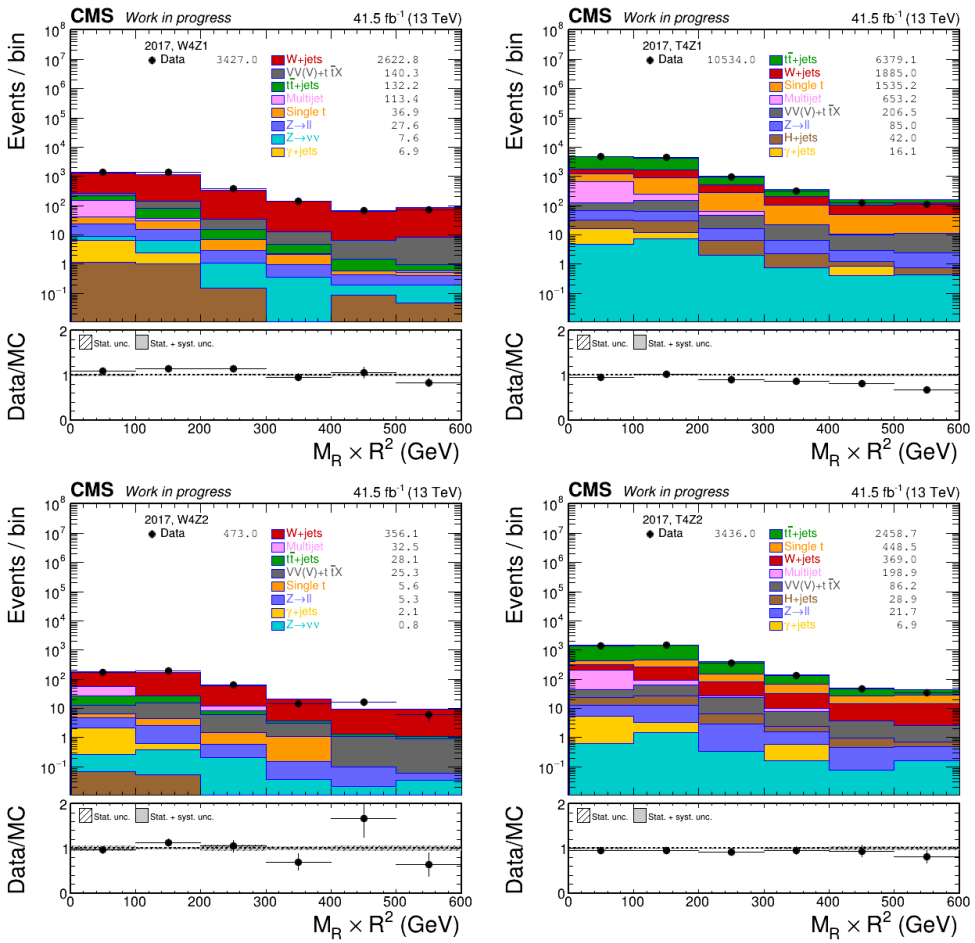


Figure 4.43: Data and MC distributions for $M_R \times R^2$ for the W4Z1, T4Z1, W4Z2, T4Z2 control regions defined in Table 4.6 for the 2017 data taking period.

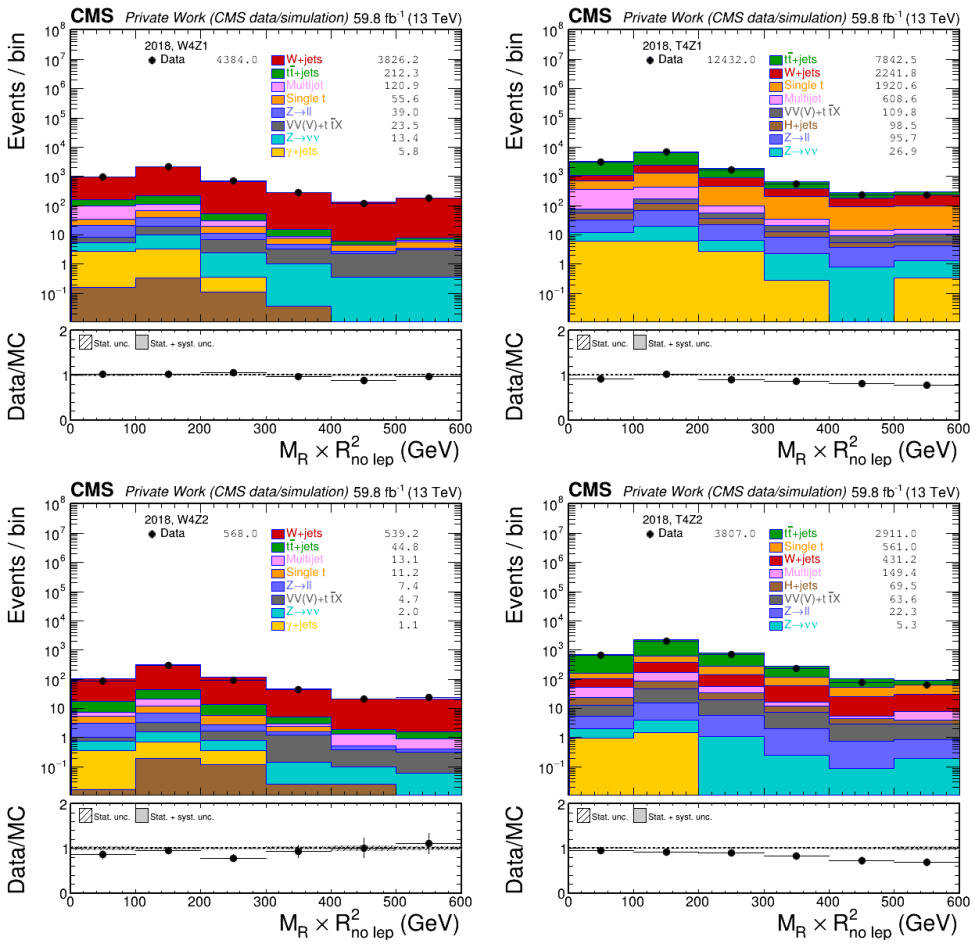


Figure 4.44: Data and MC distributions for $M_R \times R^2$ for the W4Z1, T4Z1, W4Z2, T4Z2 control regions defined in Table 4.6 for the 2018 data taking period.

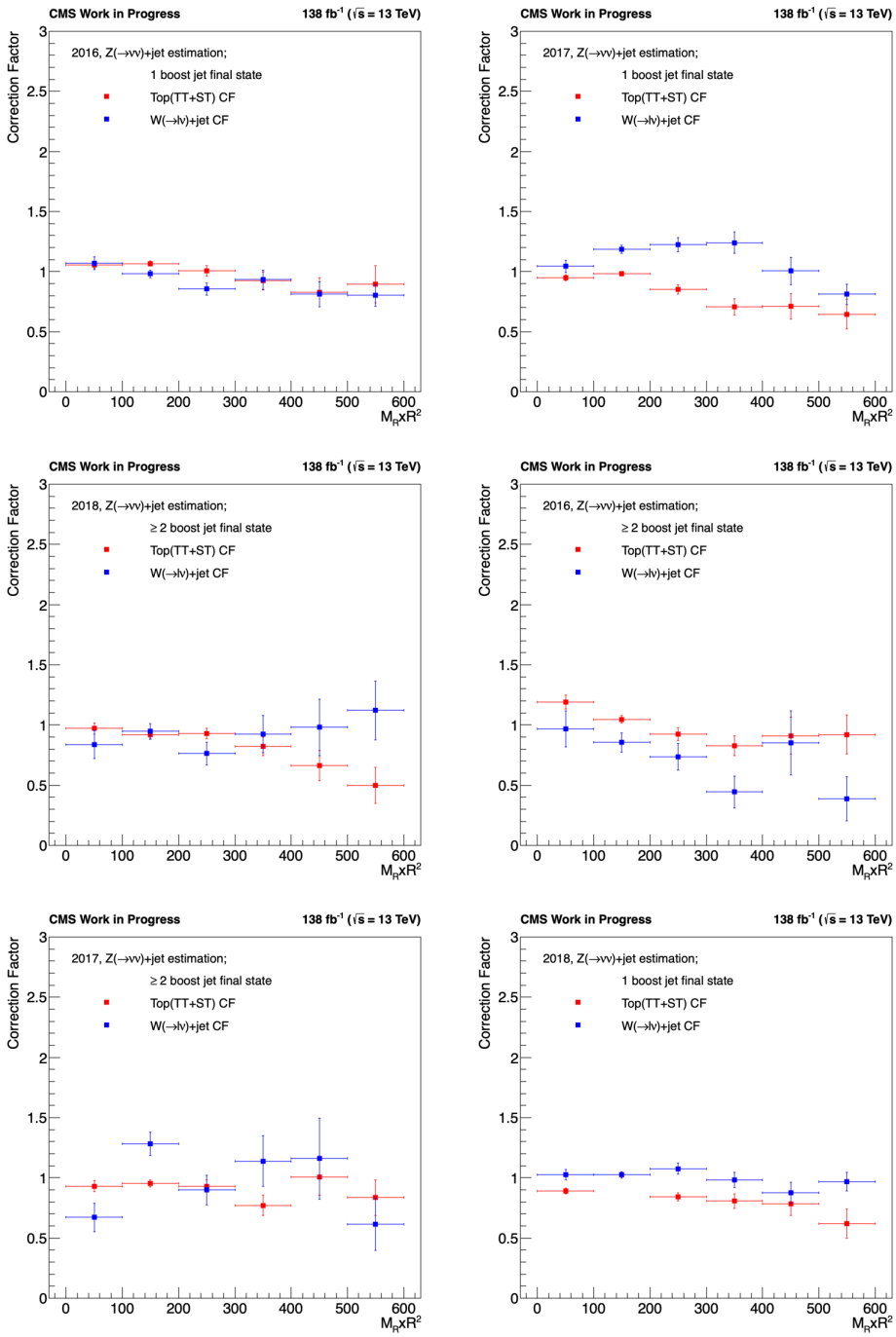


Figure 4.45: Distributions of $M_R \times R^2$ correction factors for $t\bar{t}$+jets and W+jets MC for cases with exactly 1 or at least 2 mass-tagged AK8 jets.

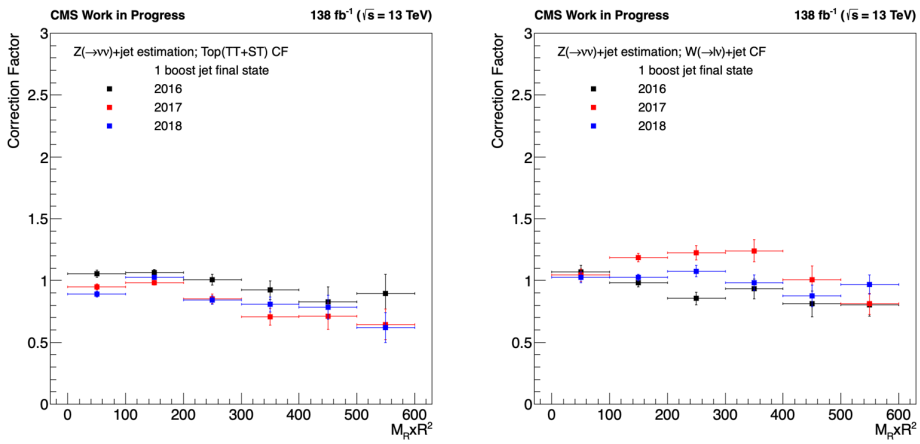


Figure 4.46: Yearly distributions of $M_R \times R^2$ correction factors for $t\bar{t}$+jets and W+jets MC for cases with exactly 1 or at least 2 mass-tagged AK8 jets.

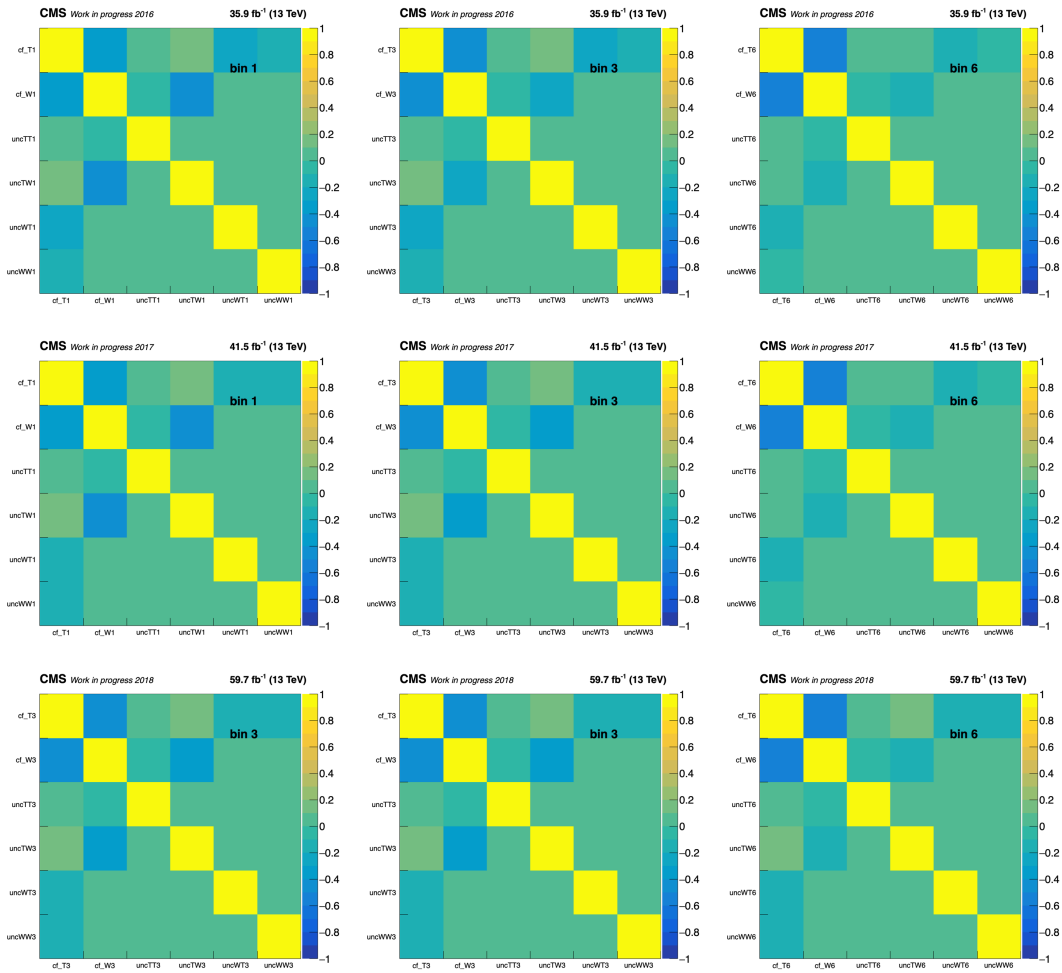


Figure 4.47: $M_R \times R^2$ correlation matrix for $t\bar{t}$+jets and W+jets MC for 3 different $M_R \times R^2$ bins in 2016(top), 2017(middle), 2018(bottom).

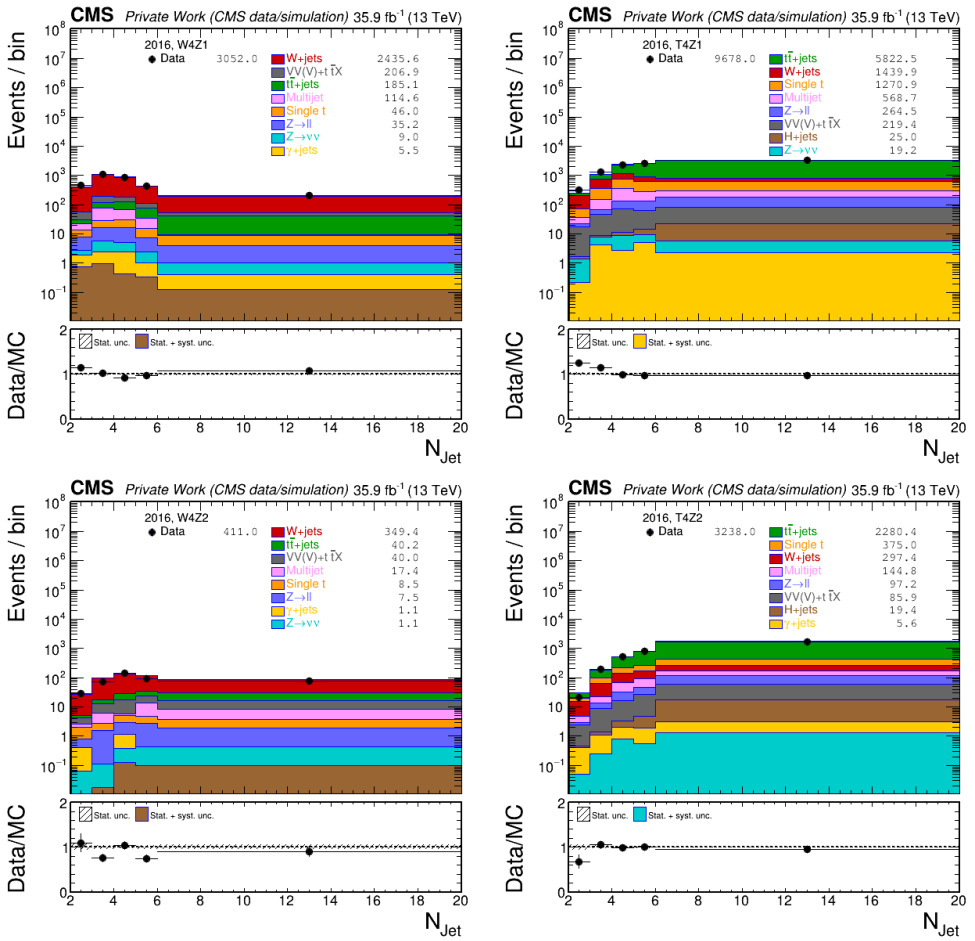


Figure 4.48: Data and MC distributions for AK4 jet multiplicity for the W4Z1, T4Z1, W4Z2, T4Z2 control regions defined in Table 4.6 after event-by-event application of the $M_R \times R^2$ correction factors, shown for the 2016 data taking period.

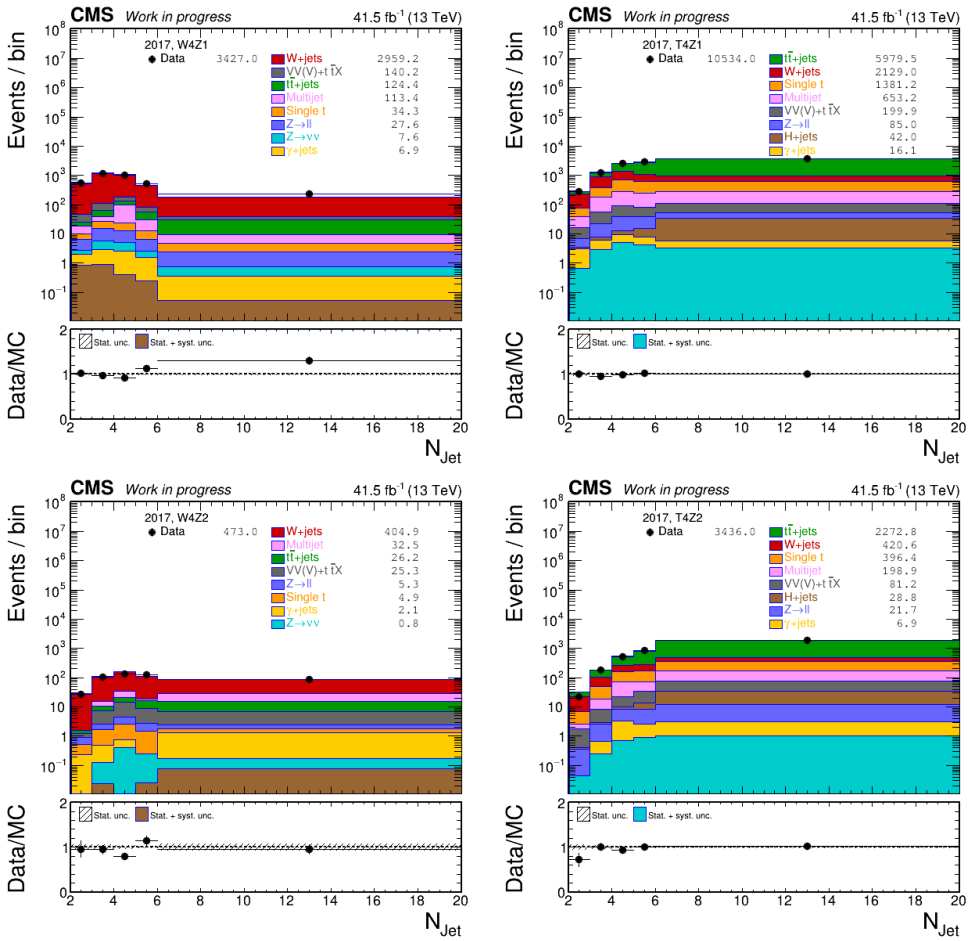


Figure 4.49: Data and MC distributions for AK4 jet multiplicity for the W4Z1, T4Z1, W4Z2, T4Z2 control regions defined in Table 4.6 after event-by-event application of the $M_R \times R^2$ correction factors, shown for the 2017 data taking period.

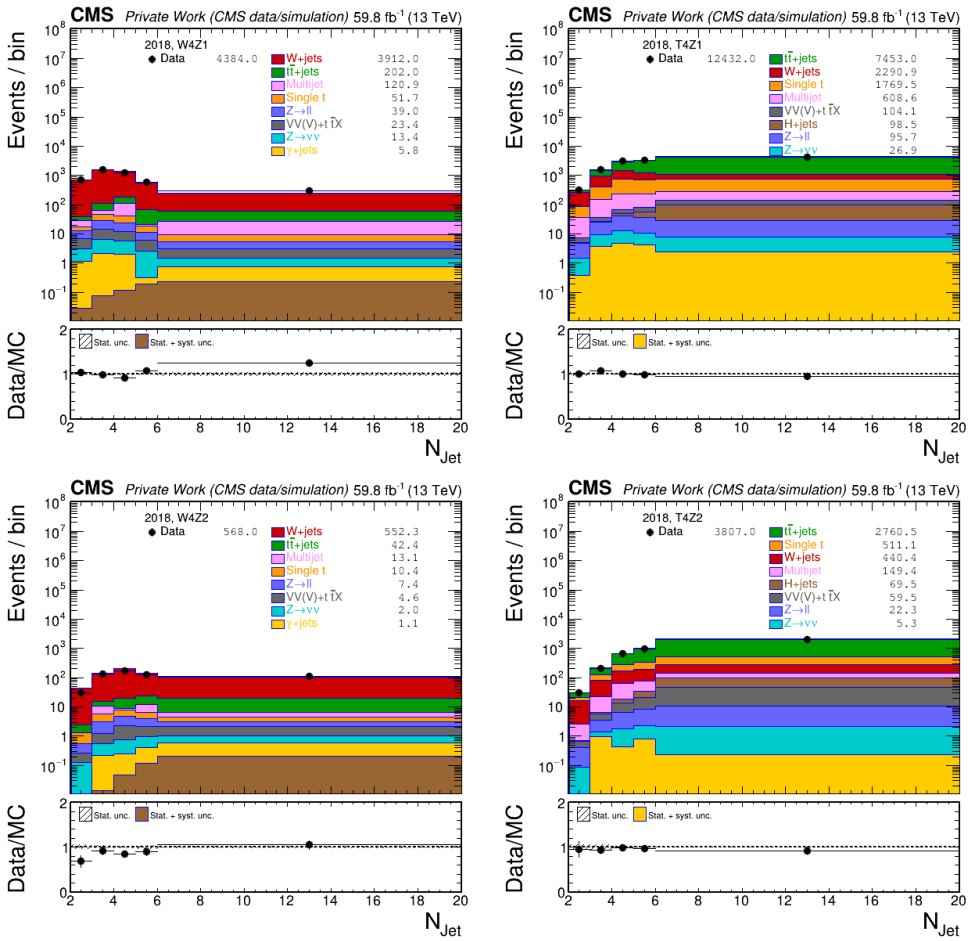


Figure 4.50: Data and MC distributions for AK4 jet multiplicity for the W4Z1, T4Z1, W4Z2, T4Z2 control regions defined in Table 4.6 after event-by-event application of the $M_R \times R^2$ correction factors, shown for the 2018 data taking period.

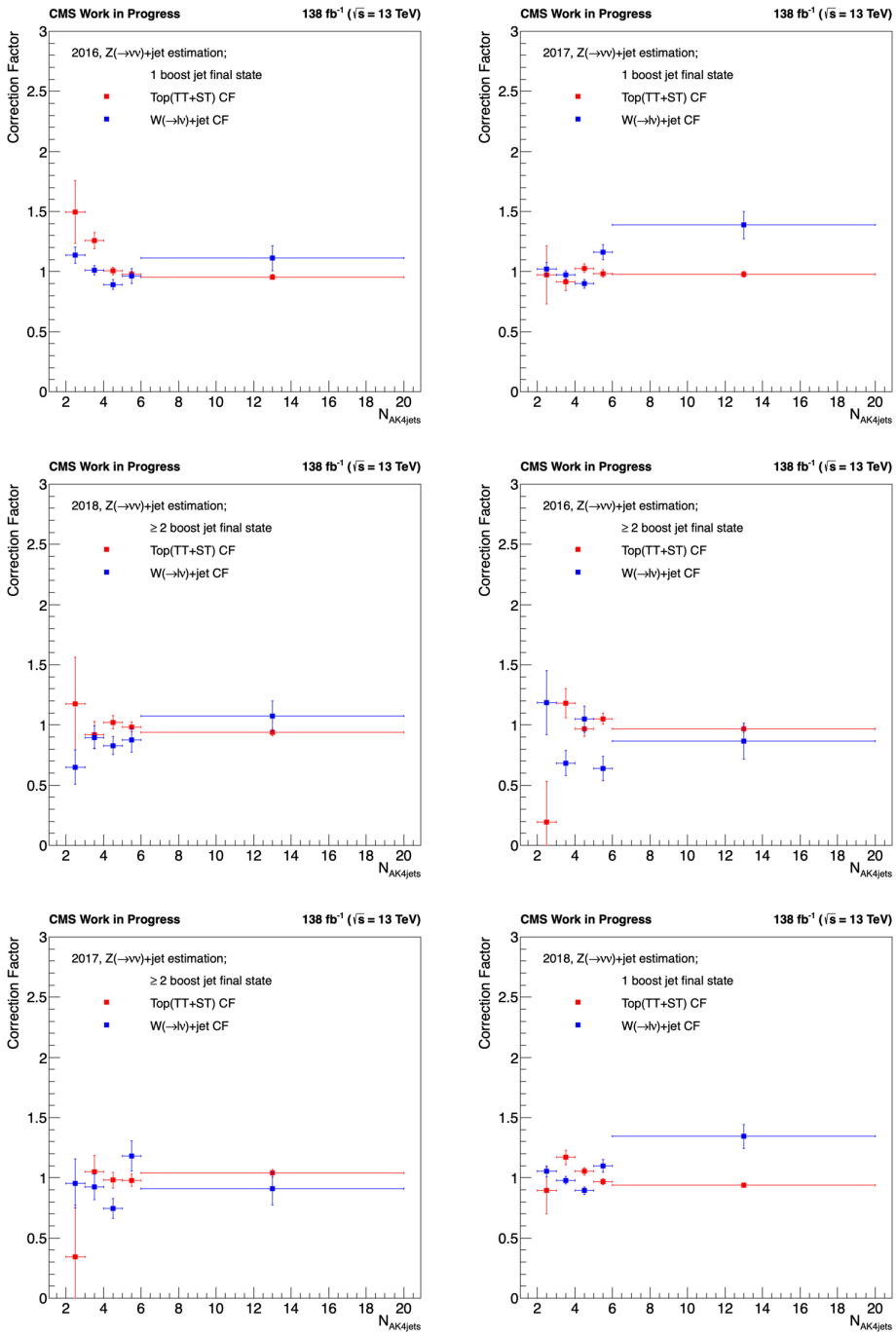


Figure 4.51: Distributions of AK4 jet multiplicity correction factors for $t\bar{t}$+jets and W+jets MC for cases with exactly 1 or at least 2 mass-tagged AK8 jets.

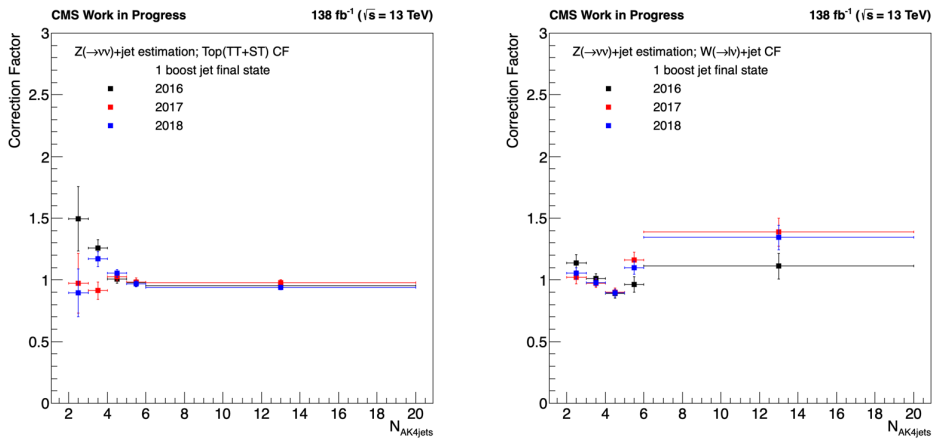


Figure 4.52: Yearly distributions of AK4 jet multiplicity correction factors for $t\bar{t}$+jets and W+jets MC for cases with exactly 1 or at least 2 mass-tagged AK8 jets.

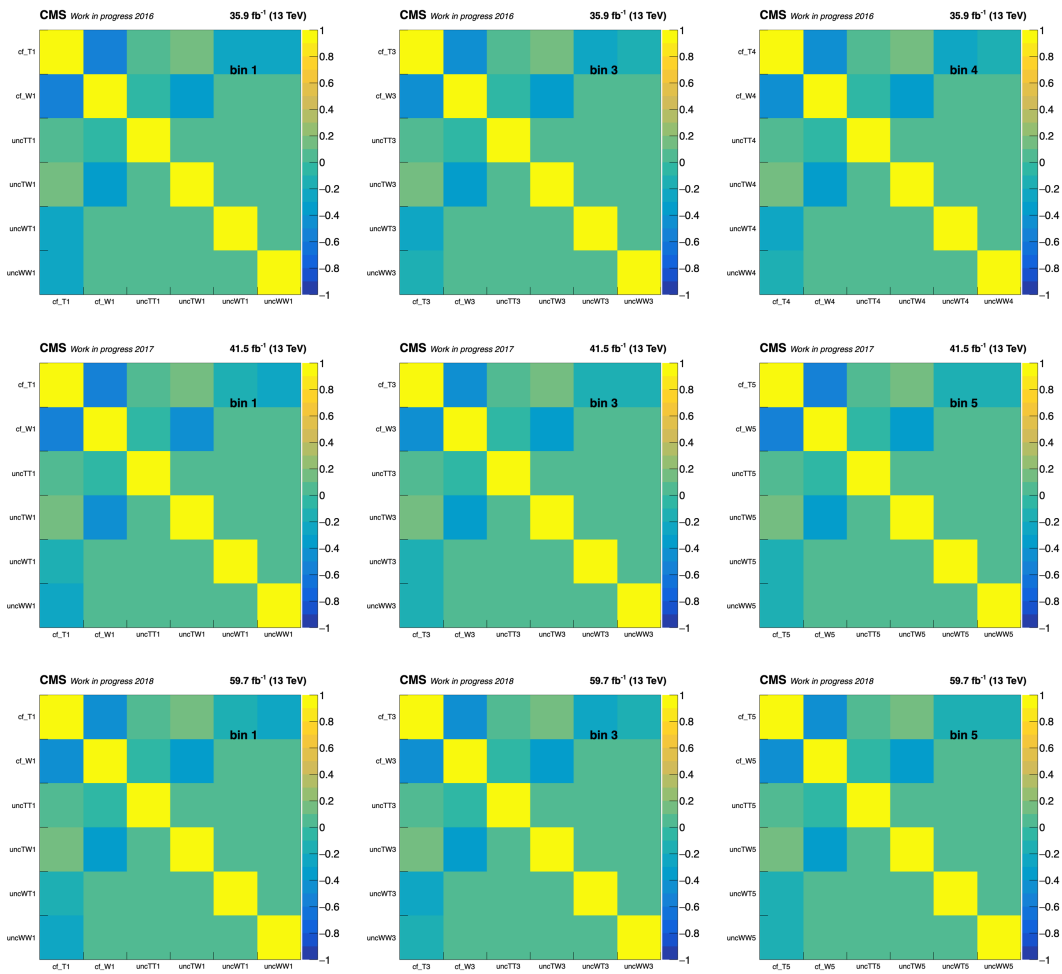


Figure 4.53: AK4 jet multiplicity correlation matrix for QCD, $t\bar{t}$+jets and W+jets MC for 3 different AK4 jet multiplicity bins in 2016(top), 2017(middle), 2018(bottom).

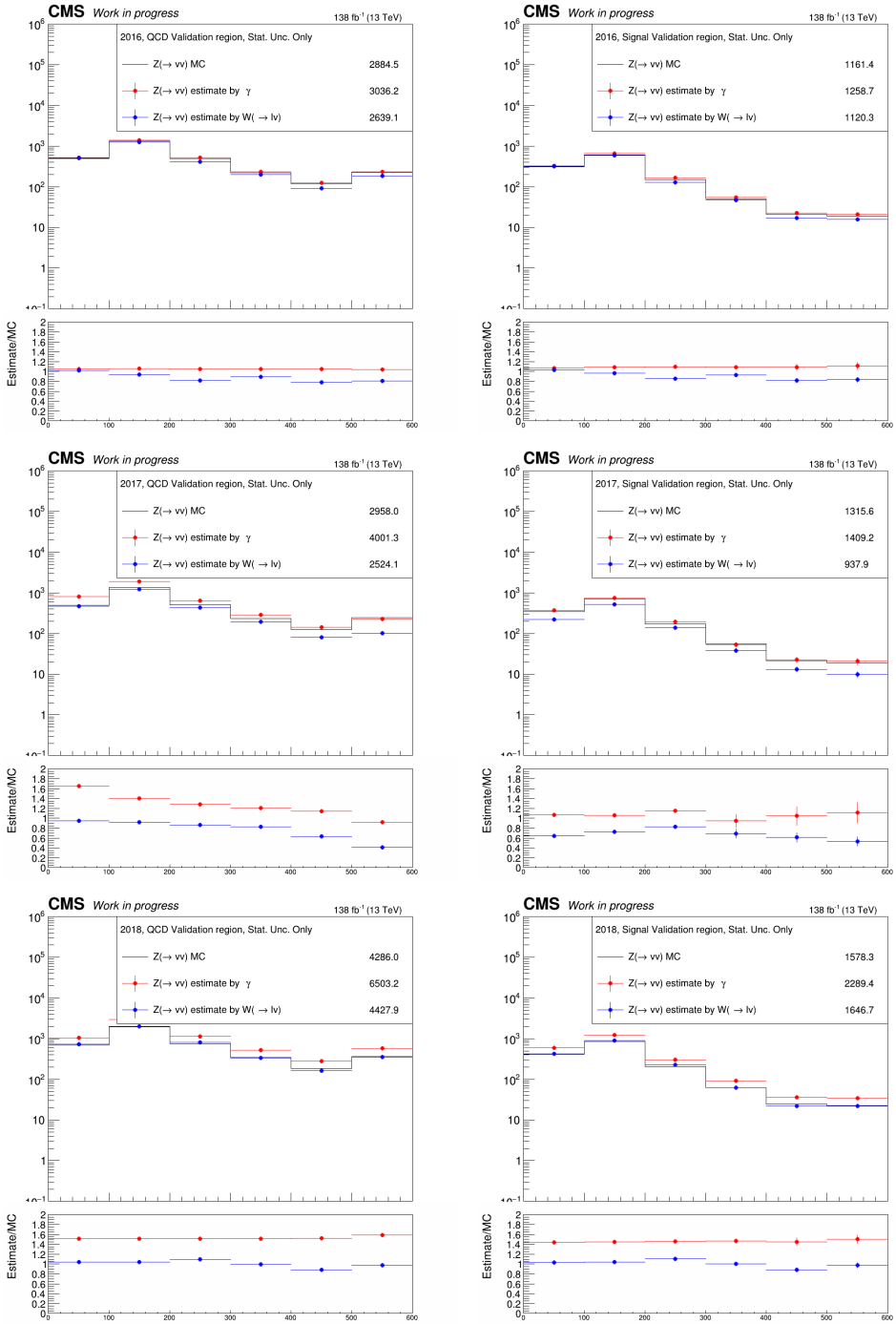


Figure 4.54: $M_R \times R^2$ distribution with $Z(\rightarrow \nu\nu)$ MC, background estimation from $\gamma+jets$ and dilepton CRs. The results are 2016, 2017, 2018, run2 from top.

4.2.7 Background estimation for the non-isolated lepton regions

Signal regions in the non-isolated lepton category are dominated by top backgrounds, followed by multijet and W+jets . We obtain the correction factors for the $t\bar{t}$+jets and W+jets backgrounds in the non-isolated category by the simultaneous derivation method. The multijet correction factor is taken directly as the one derived in Section 4.2.5. We revert the variables $\Delta\phi*$ and m_T in different ways with respect to the signal region definitions to construct control and validation regions, as seen in Figure 4.55. The low $\Delta\phi*$, low-m_T selection $\Delta\phi* < 0.8$, $m_T < 140$ defines the control region. This selection is further divided into two regions, Tnl, with ≥ 1 b-tagged jets and Wnl, with zero b-tagged jets, to constitute CRs for $t\bar{t}$+jets and W+jets CF derivation, respectively. The selections for these are shown in Table 4.6 and the $M_R \times R^2$ distributions are shown in Figure 4.56.

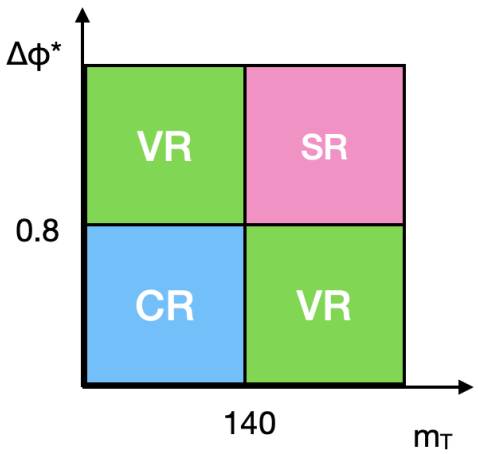


Figure 4.55: Signal, control and validation regions for the non-isolated lepton category in the $\Delta\phi*$-m_T phase space.

The $t\bar{t}$+jets and W+jets CFs are obtained by solving the following two equations:

$$\mathrm{CF}^{\mathrm{i}}_{\mathrm{ttjets,nl}}\mathrm{MC}^{\mathrm{Tnl}}_{\mathrm{ttjets}} + \mathrm{CF}^{\mathrm{i}}_{\mathrm{Wjets,nl}}\mathrm{MC}^{\mathrm{Tnl}}_{\mathrm{Wjets}} = \mathrm{Data}^{\mathrm{Tnl}}. \quad (4.14)$$

$$\mathrm{CF}^{\mathrm{i}}_{\mathrm{ttjets,nl}}\mathrm{MC}^{\mathrm{Wnl}}_{\mathrm{ttjets}} + \mathrm{CF}^{\mathrm{i}}_{\mathrm{Wjets,nl}}\mathrm{MC}^{\mathrm{Wnl}}_{\mathrm{Wjets}} = \mathrm{Data}^{\mathrm{Wnl}} \quad (4.15)$$

$$(4.16)$$

Figure 4.57 shows the correction factors versus $M_R \times R^2$. Next, these CFs are applied to the MC in the CRs and CFs are computed versus AK4 jet multiplicity. Figure 4.59 shows the AK4 jet multiplicity in the CRs, and Figure 4.60 shows the resulting AK4 jet CFs.

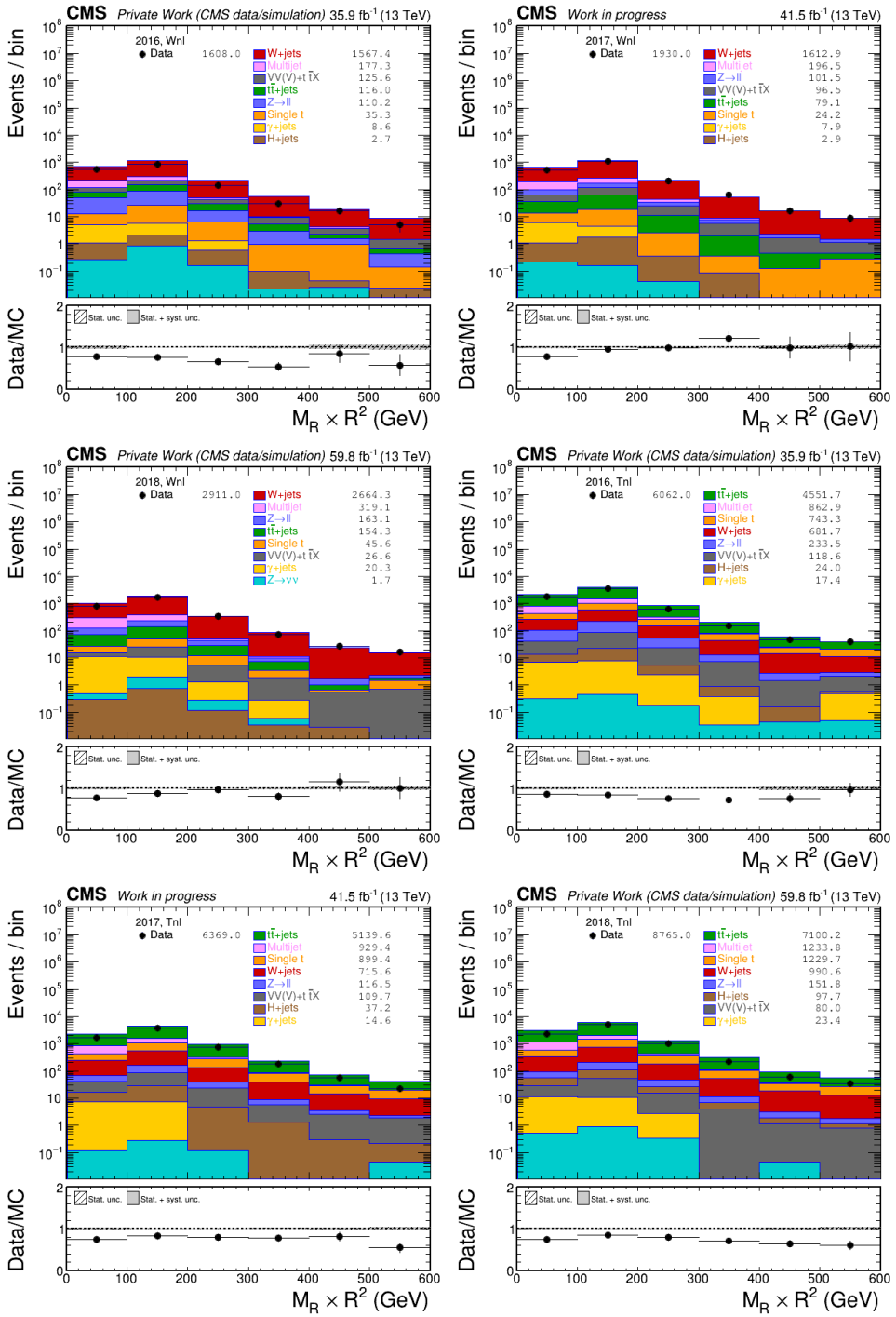


Figure 4.56: Data and MC distributions for $M_R \times R^2$ for the Tnl and Wnl control regions defined in Table 4.6 for the 2016, 2017, 2018 data taking periods.

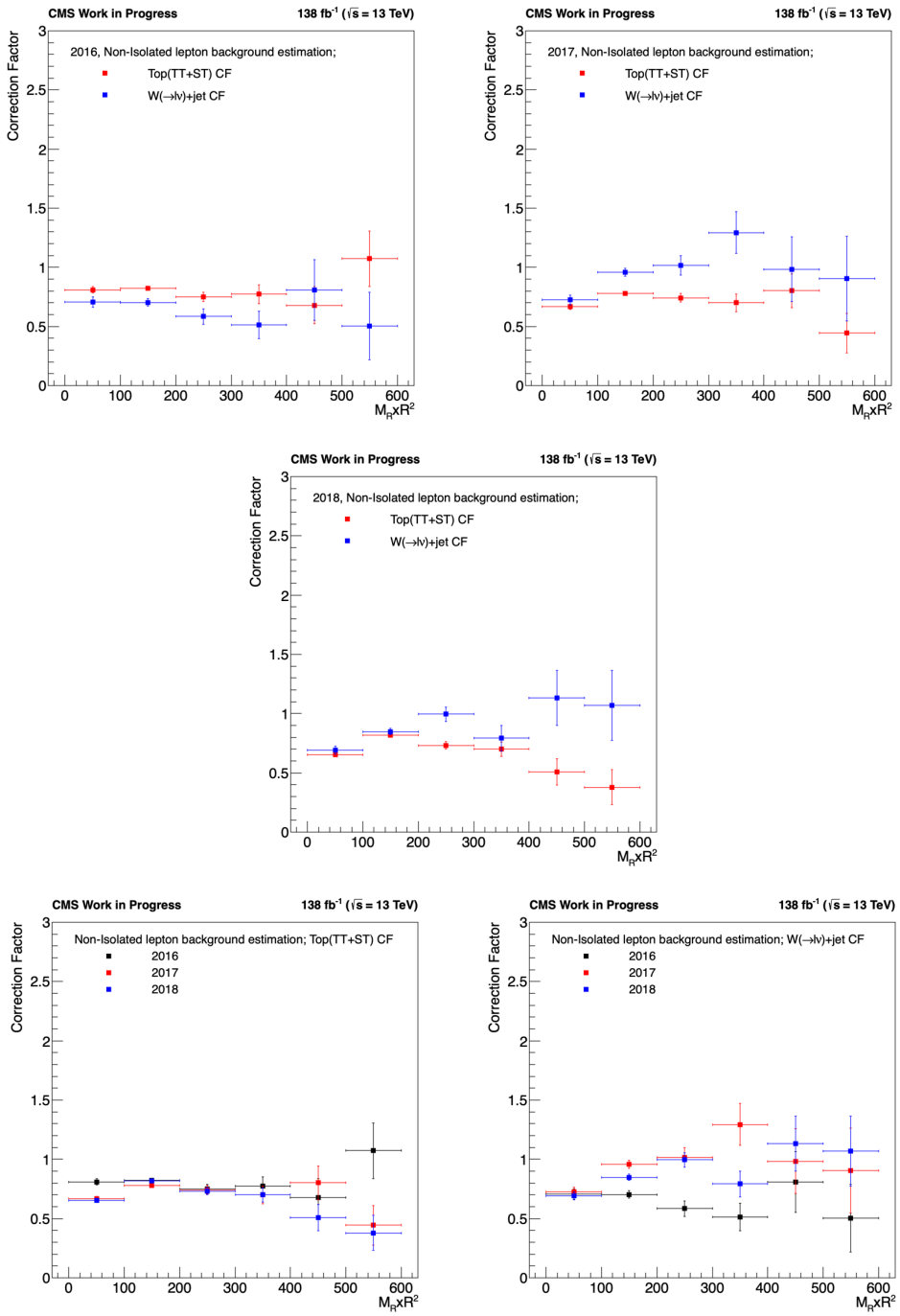


Figure 4.57: Distributions of $M_R \times R^2$ correction factors for $t\bar{t}$+jets and W+jets MC for the non-isolated category.

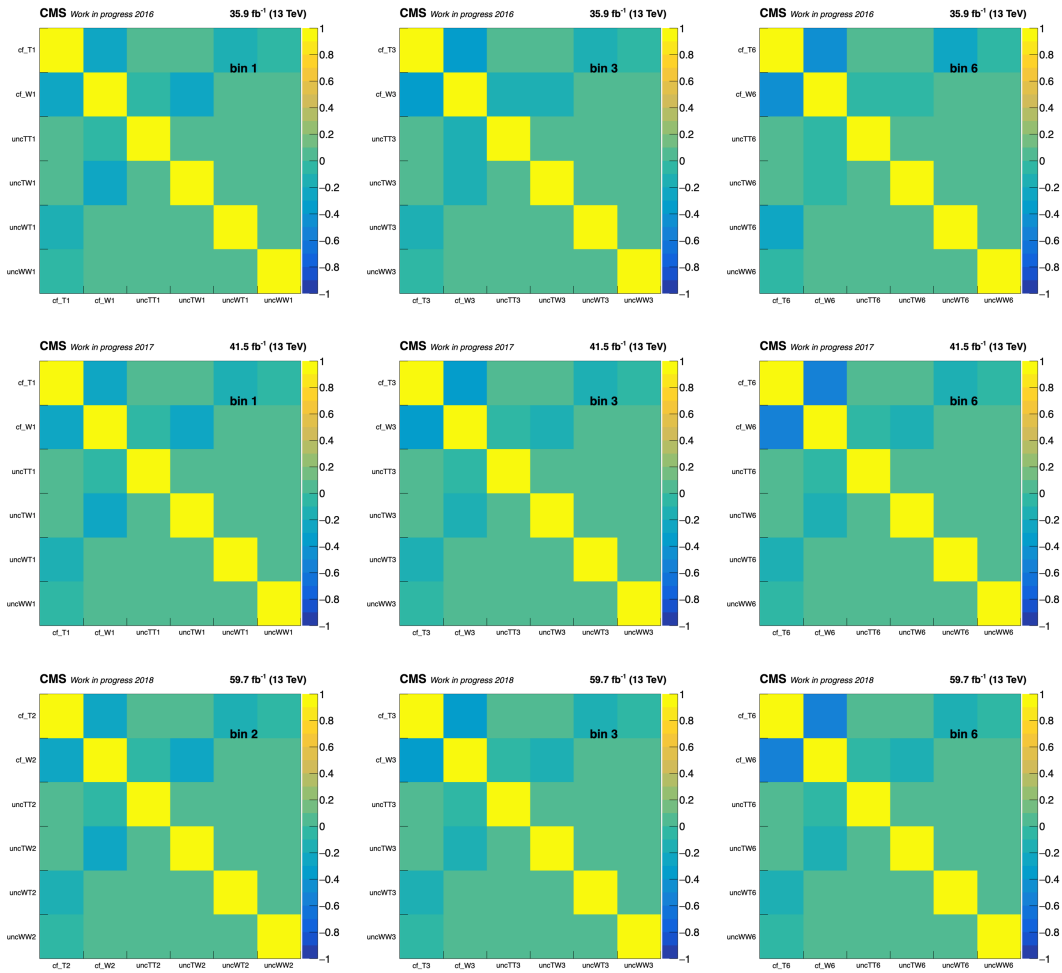


Figure 4.58: $M_R \times R^2$ correlation matrix for $t\bar{t}$+jets and W+jets MC for 3 different $M_R \times R^2$ bins in 2016(top), 2017(middle), 2018(bottom).

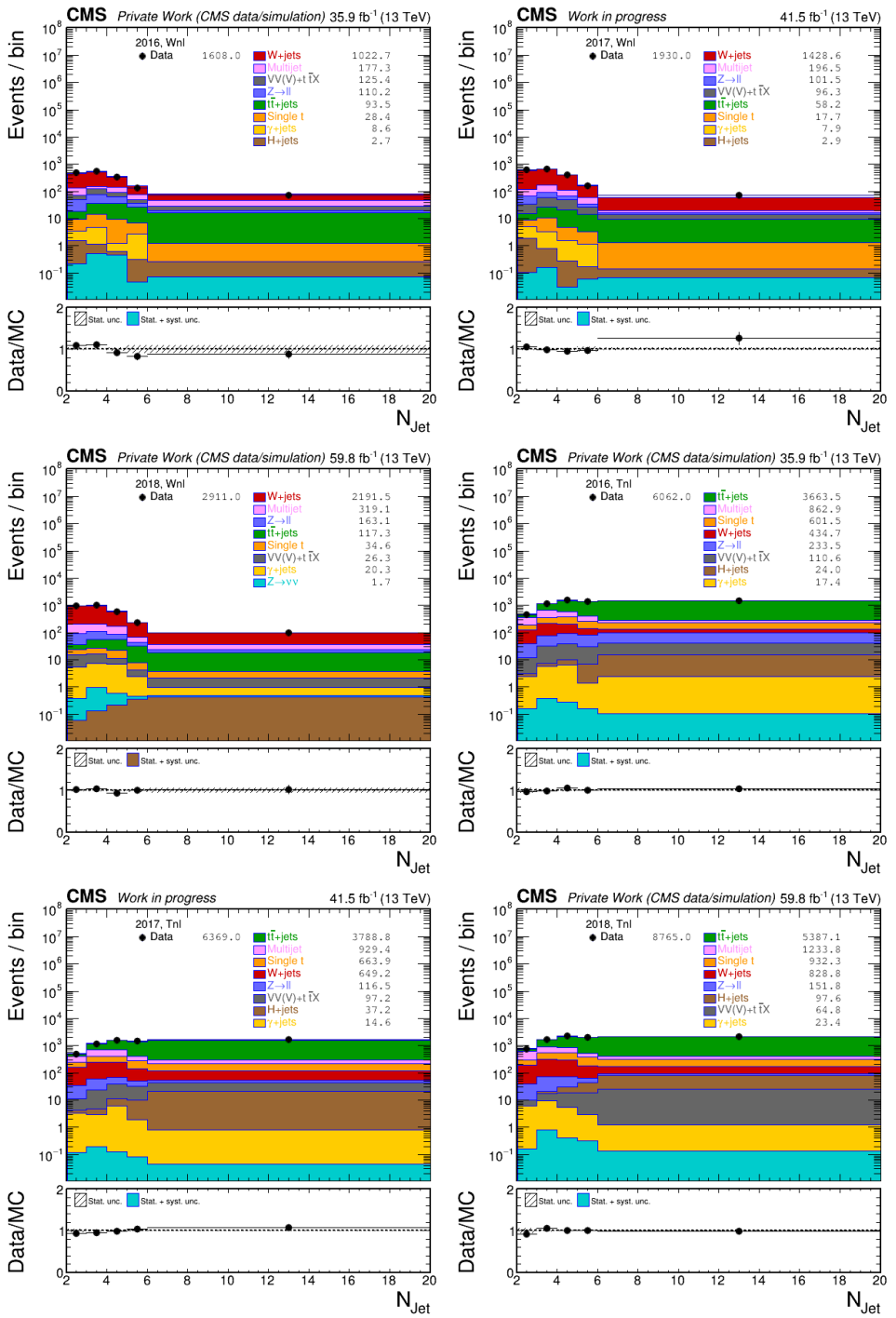


Figure 4.59: Data and MC distributions for AK4 jet multiplicity for the Tnl and Wnl control regions defined in Table 4.6 after event-by-event application of the $M_R \times R^2$ correction factors, shown for the 2016, 2017 and 2018 data taking periods.

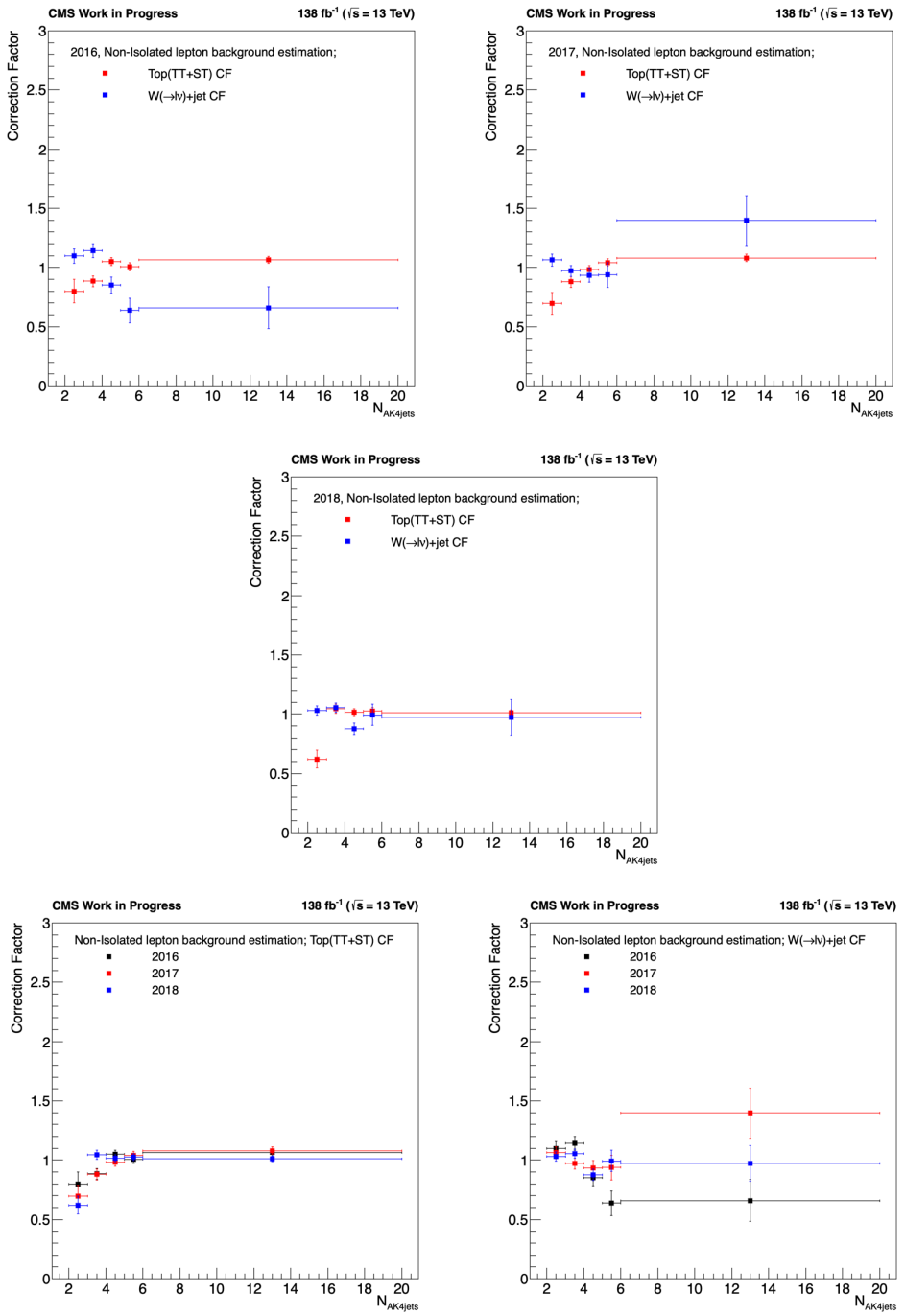


Figure 4.60: Distributions of AK4 jet multiplicity correction factors for $t\bar{t}$+jets and W+jets MC for the non-isolated category.

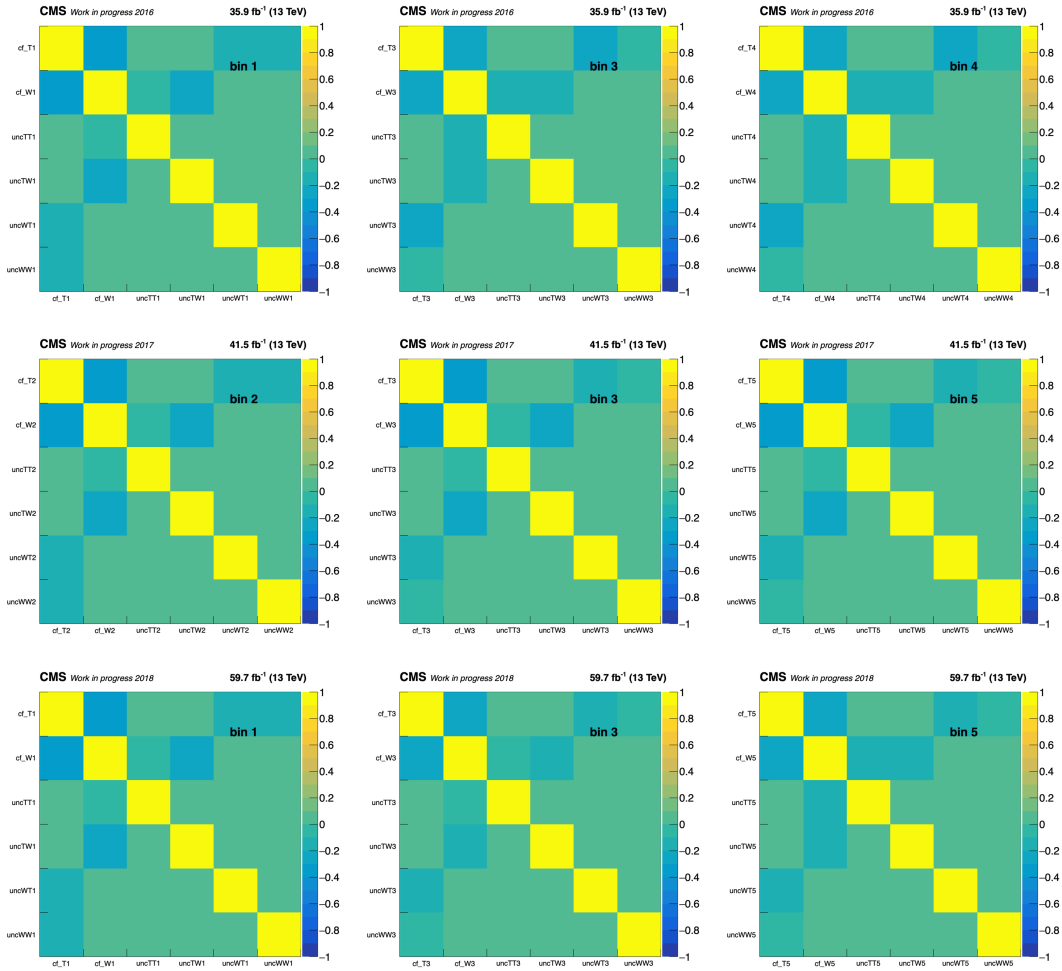


Figure 4.61: AK4 jet multiplicity correlation matrix for $t\bar{t}$+jets and W+jets MC for 3 different AK4 jet multiplicity bins in 2016(top), 2017(middle), 2018(bottom).

4.2.8 Validation of background estimation

Background estimation as outlined in Section 4.2.4 using the validation regions listed in Table 4.7. Correction factors in the relevant category are applied to the corresponding MC, and data-MC agreement is checked.

First, we check the validity of the CFs calculated for hadronic and isolated leptonic categories in validation regions with explicit W/Z, hadronic top, and Higgs tagging, listed as no 1-9 in Table 4.7. Figures 4.62, 4.63, 4.64, 4.65, 4.66, 4.67, show the comparison of data-MC distributions versus $M_R \times R^2$ in these VRs for 2016, 2017 and 2018 data-taking periods. Next, we check the validity of working with a multijet control region defined by reverting ΔR via the VRs no 10 and 11 in Table 4.7. Figure 4.68, 4.69, and 4.70 shows the comparison of data-MC distributions versus $M_R \times R^2$ in these VRs for the 2016, 2017, and 2018 data-taking periods. Finally, we check the validity of CFs defined in the non-isolated lepton control regions defined by reverting both $\Delta\phi*$ and m_T in validation regions with reverted $\Delta\phi*$ or reverted m_T, listed as no 13 and 14 in Table 4.7. Figure 4.71 shows the comparison of data-MC distributions versus $M_R \times R^2$ in these VRs for the 2016, 2017, and 2018 data-taking periods.

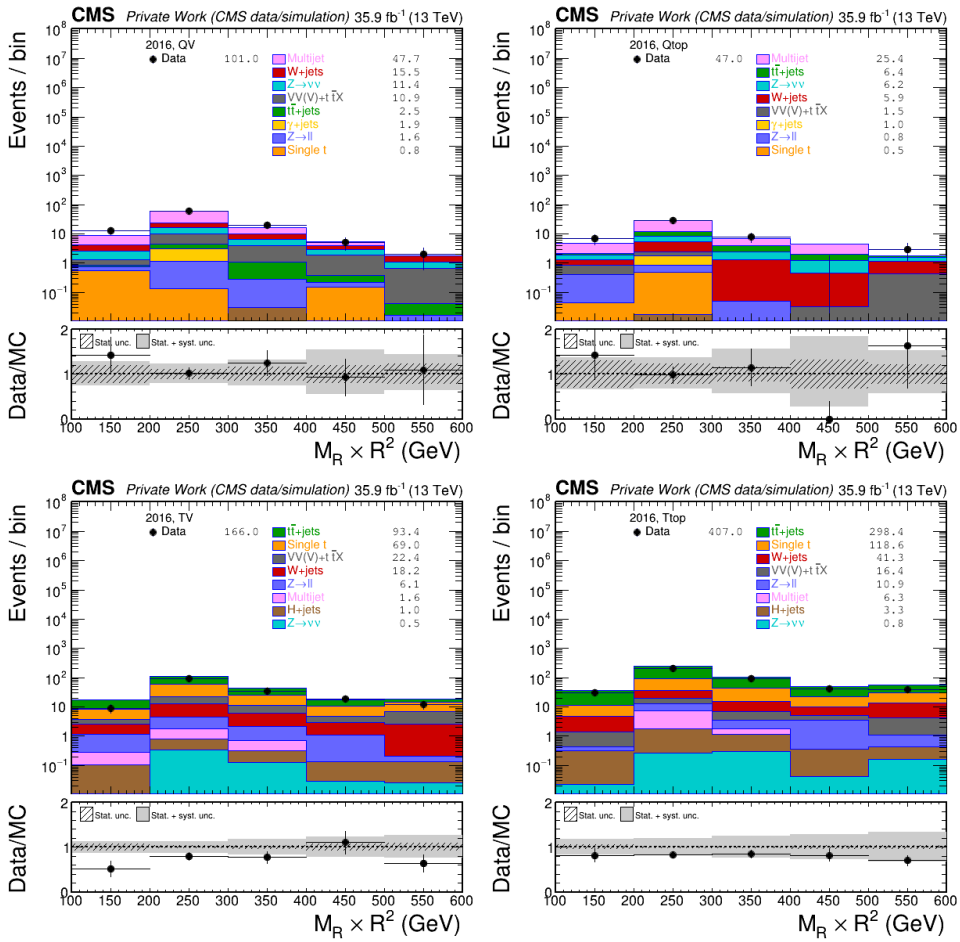


Figure 4.62: Data and MC distributions for $M_R \times R^2$ in the validation regions no 1-4 with explicit boosted object tagging defined in Table 4.7 after event-by-event application of all correction factors, shown for the 2016 data taking period.

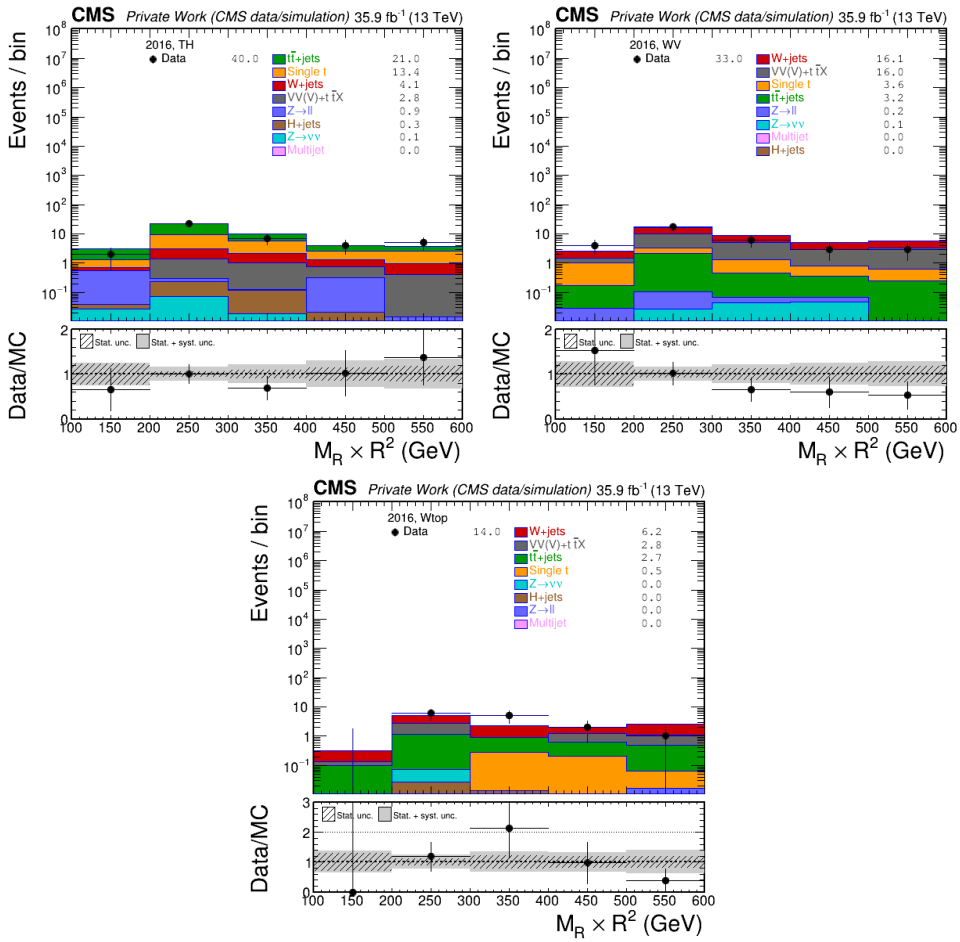


Figure 4.63: Data and MC distributions for $M_R \times R^2$ in the validation regions no 5-7 with explicit boosted object tagging defined in Table 4.7 after event-by-event application of all correction factors, shown for the 2016 data taking period.

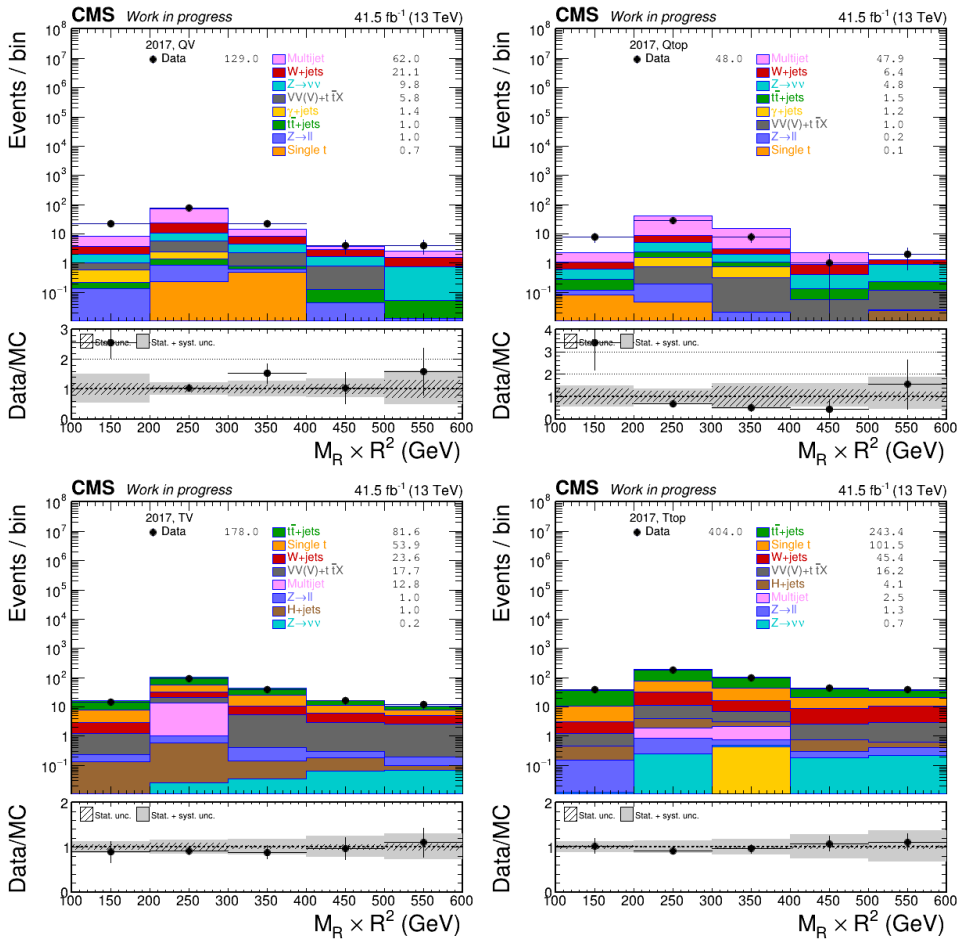


Figure 4.64: Data and MC distributions for $M_R \times R^2$ in the validation regions no 1-4 with explicit boosted object tagging defined in Table 4.7 after event-by-event application of all correction factors, shown for the 2017 data taking period.

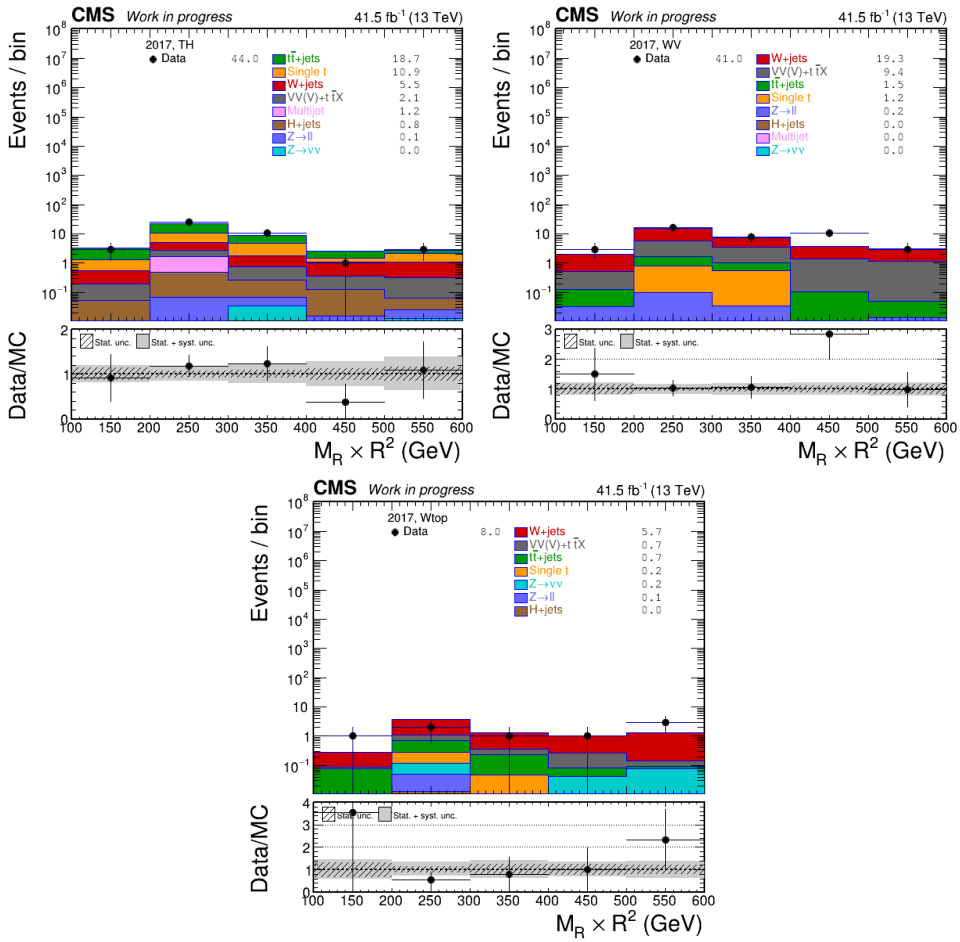


Figure 4.65: Data and MC distributions for $M_R \times R^2$ in the validation regions no 5-7 with explicit boosted object tagging defined in Table 4.7 after event-by-event application of all correction factors, shown for the 2017 data taking period.

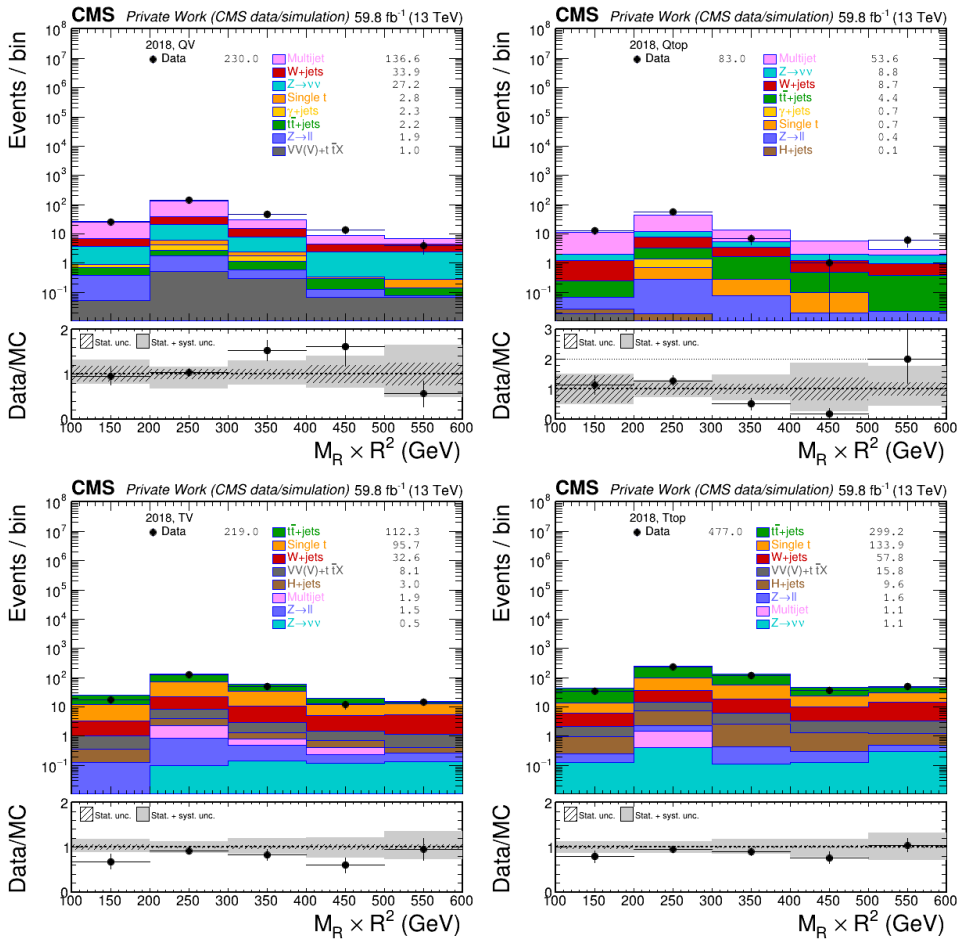


Figure 4.66: Data and MC distributions for $M_R \times R^2$ in the validation regions no 1-4 with explicit boosted object tagging defined in Table 4.7 after event-by-event application of all correction factors, shown for the 2018 data taking period.

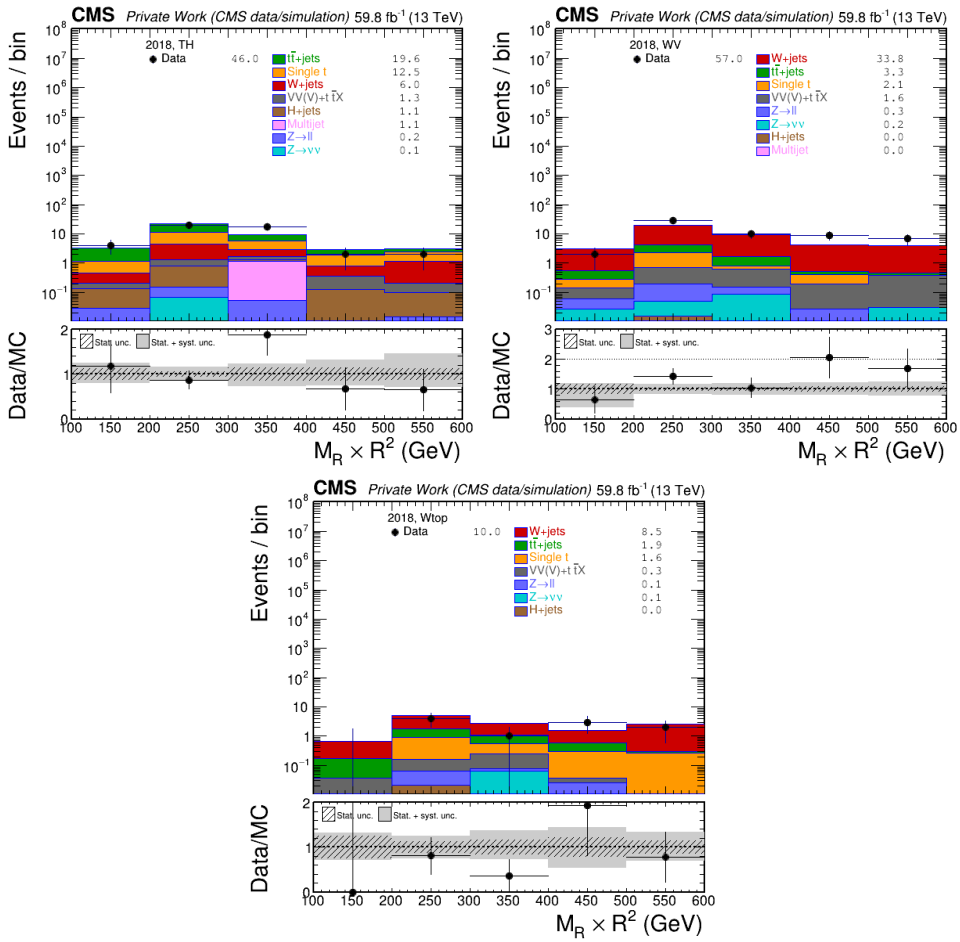


Figure 4.67: Data and MC distributions for $M_R \times R^2$ in the validation regions no 5-7 with explicit boosted object tagging defined in Table 4.7 after event-by-event application of all correction factors, shown for the 2018 data taking period.

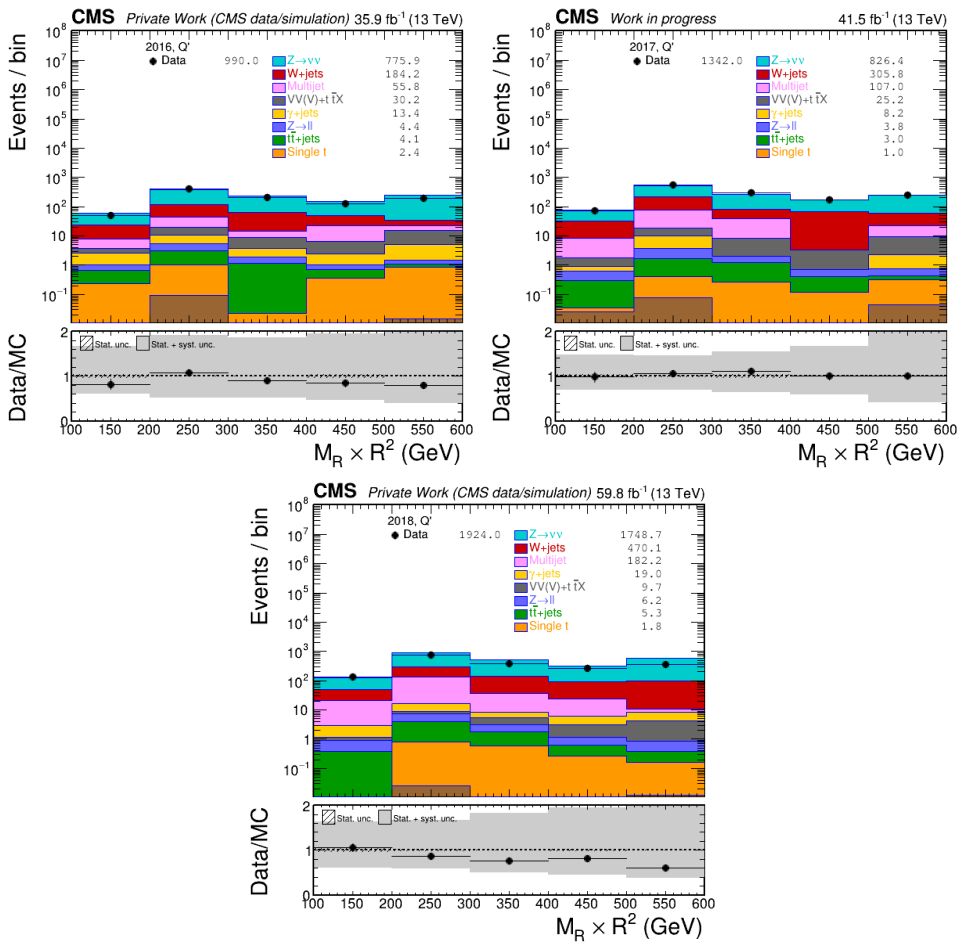


Figure 4.68: Data and MC distributions for $M_R \times R^2$ in the validation regions no 10 for testing different kinematic conditions, defined in Table 4.7 after event-by-event application of all correction factors, shown for 2016, 2017, and 2018 data taking period.

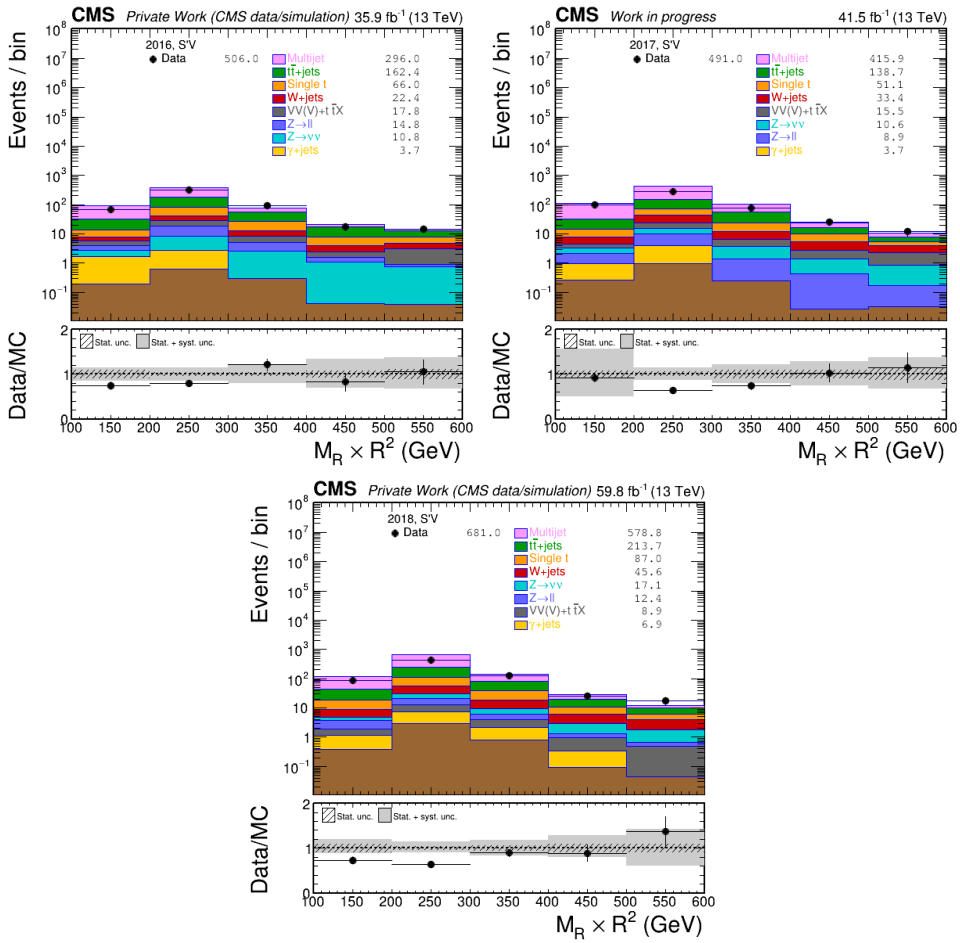


Figure 4.69: Data and MC distributions for $M_R \times R^2$ in the validation regions no 11 for testing different kinematic conditions, defined in Table 4.7 after event-by-event application of all correction factors, shown for 2016, 2017, and 2018 data taking period.

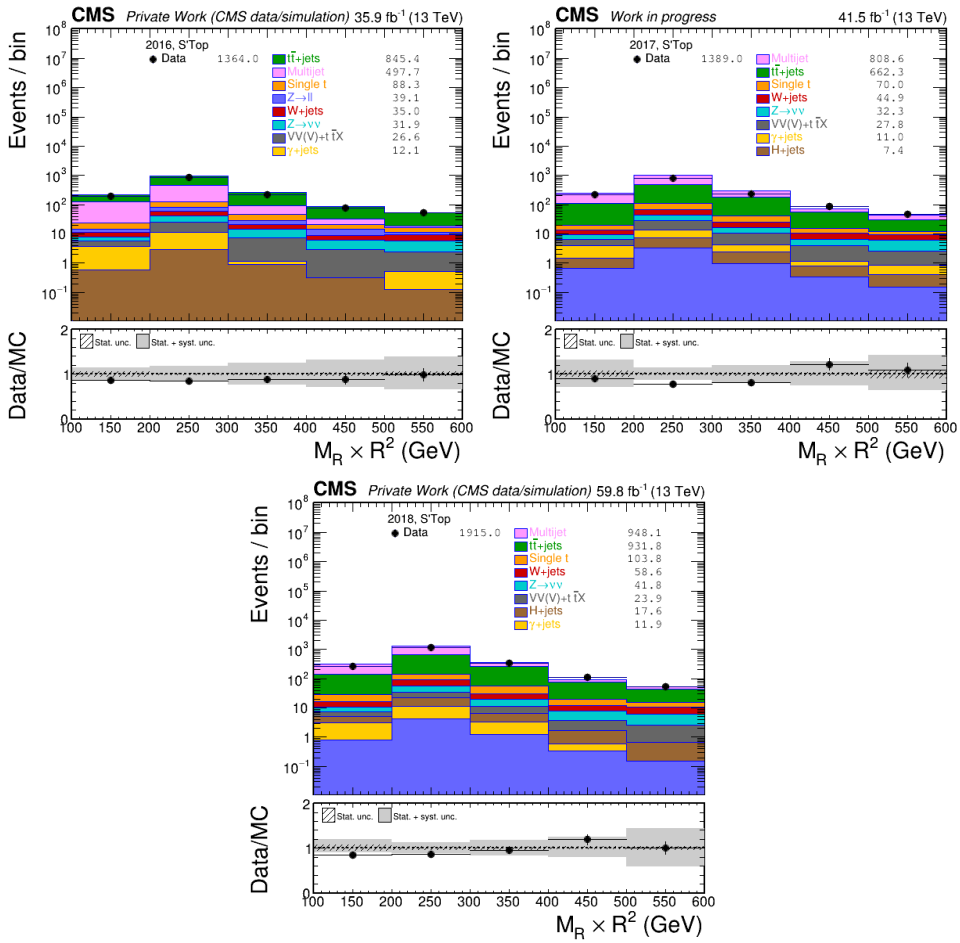


Figure 4.70: Data and MC distributions for $M_R \times R^2$ in the validation regions no 12 for testing different kinematic conditions, defined in Table 4.7 after event-by-event application of all correction factors, shown for 2016, 2017, and 2018 data taking period.

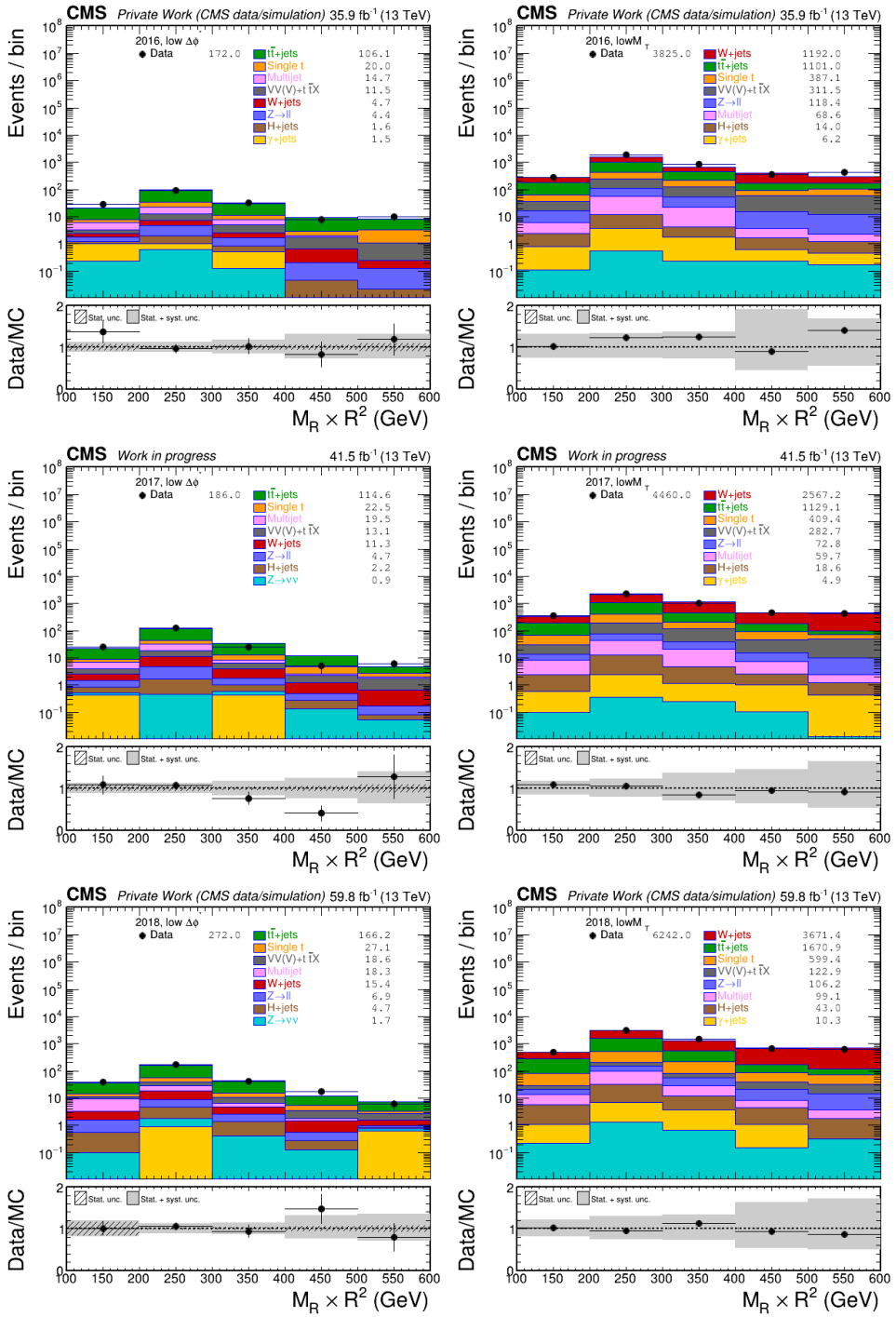


Figure 4.71: Data and MC distributions for $M_R \times R^2$ in the validation regions no 13,14 for non-isolated leptonic regions, with reverted $\Delta\phi*$ or m_T, defined in Table 4.7 after event-by-event application of all correction factors, shown for the 2016, 2017 and 2018 data taking period.

4.3 Systematic uncertainties

The following systematic uncertainties are applied to take into account the inconsistencies of MC events with respect to data.

4.3.1 Experimental sources of systematic uncertainties

- **Uncertainties on the trigger efficiency:** The per-bin statistical uncertainties on the trigger efficiency measurements presented in Section 3.3.1 (from 68% Clopper-Pearson intervals) are propagated to the MC sample yields, uncorrelated across all bins.

- **Luminosity:** Uncertainties are applied on the luminosity, separated into uncorrelated between the 2016 (1%), 2017 (2%), and 2018 (1.5%) datasets, correlated between the 2017 (0.2%) and 2018 datasets (0.6%), and correlated between the 2016 (0.6%), 2017 (0.9%), and 2018 (2%) datasets.

- **Pileup reweighting:** Pileup is applied as a shape uncertainty using up/down (ultra-legacy) variations provided by the LUMI POG [29], uncorrelated across the 2016, 2017, and 2018 datasets.

- **L1 prefiring uncertainty:** Uncertainties on the L1 ECal and muon prefiring rates for the 2016 and 2017 datasets are included as shape uncertainties on all MC samples using the up/down variations in NanoAODv9, which are derived by the recipe described in [30].

- **Jet energy scale/jet energy resolution:** The jet energy scale and resolution uncertainties are applied as shape uncertainties on all MC samples, using the corrections derived by the JetMET group [30, 31]. JES corrections are applied correlated, and JER uncorrelated across all years. Detailed JES/JER information was presented in Section 3.3.

- **Heavy jet mass scales and resolution (JMS/JMR):** The uncertainties on the ParticleNet regressed mass scale and resolution are applied as shape uncer-

tainties on all MC samples, uncorrelated across all years, according to up/-down variations provided by the JMAR group.

- **b tagging scale factor uncertainties:** Data/FullSim and FullSim/FastSim scale factors for b tagging and mistagging are applied as recommended by the BTV POG [23] and as described in Section 3.3. The uncertainties in SFs are applied as a shape systematic, by computing the event weights by varying the SFs up and down by the SF uncertainties.

- **Boosted object tagging / fake rate uncertainties:** Boosted object data/FullSim and FullSim/FatSim scale factors and fake rates were presented in Section 3.3. These SFs are applied with the same method used for applying b-tag SFs. The uncertainties on SFs are incorporated as a shape systematic similarly, by varying the SF up and down and recalculating the event weight.

- **Lepton scale factor uncertainties:** Uncertainties on electron and muon scale factors, shown in Sections 3.3 and 3.3 are applied as event weights.

- **Lost lepton uncertainty:** All hadronic regions in this analysis contain leptonically decaying events from top and W processes where the lepton is lost. The uncertainties arising from lost leptons are also propagated into the final results. These systematic uncertainties are derived from the data-MC discrepancy in the p_T and h distributions of a signal region-like selection with an additional veto lepton, m_T <100, and no $\Delta\phi$ cut. We first compute an average data-MC discrepancy vs. lepton p_T, which amounts to 12%. Then we compute a further h-dependent uncertainty combined with the former. These uncertainties are applied as event weights to MC events with one GEN lepton not selected with the loose lepton ID.

- **BG shape correction factor uncertainties:** The uncertainties on the correction factors obtained from the simultaneous fit are applied as an event-by-event weight variation as a shape systematic to consider the MC modeling inaccuracies. Additionally, the uncertainties to check for applying the correction factors iteratively, which are obtained by reverting the order of iterations,

are applied.

- $Z(\to \nu\nu)$+**jets estimation uncertainty:** As described in Section 4.2.6, two alternative methods were used for obtaining the $Z(\to \nu\nu)$+jets background estimate. The difference between the $M_R \times R^2$ distributions obtained by the two methods for each region is incorporated as a systematic on the $Z(\to \nu\nu)$+jets estimation in that region.

The systematic effects on yields for Top backgrounds are shown in Figure 4.72.

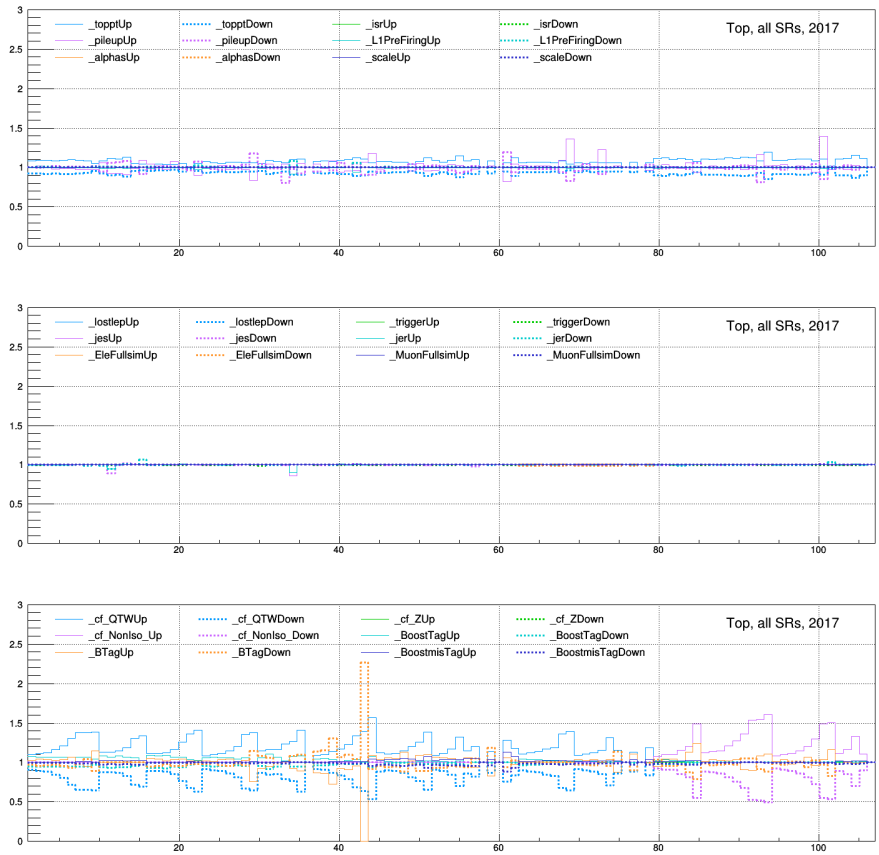


Figure 4.72: The systematic effects on yields for top backgrounds.

4.3.2 Theoretical sources of systematic uncertainties

- **top p_T reweighting:** The uncertainties associated with this reweighting are defined as follows: the one sigma down variation corresponds to no reweighting at all; the one sigma up variation corresponds to applying the reweighting twice. These uncertainties are propagated to the Data/MC ratio used in the estimation of the TTjets background.

- **ISR reweighting:** The initial state radiation (ISR) reweighting recipe was applied for the signal samples as recommended in [31]. Take 1% systematics to signal FastSim.

- **QCD renormalization and factorization scale uncertainty:** The variation around the central renormalization and factorization scale choice by a factor of 2 and 1/2 with respect to the default value.

4.3.3 Impacts

The blind impacts for background only fit and background+signal fit for two different signals(T5qqqqWH, R2bbqqlv) are shown in Figure 4.73, 4.74. The unblind impacts are shown in Figure 4.75.

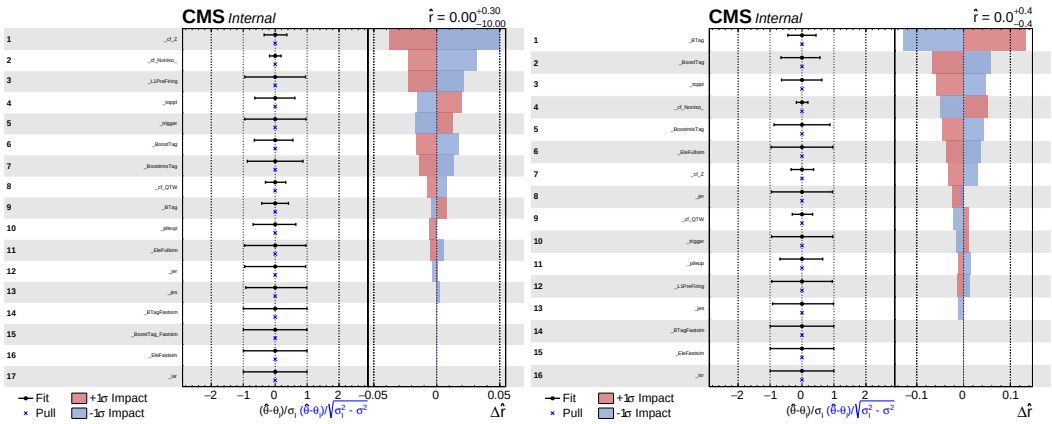


Figure 4.73: Blinded impacts for background only fit with T5qqqqWH($\tilde{g}:1600, \tilde{\chi}_0^1:500$) and R2bbqqlv($\tilde{b}:700, \tilde{\chi}_0^1:500$).

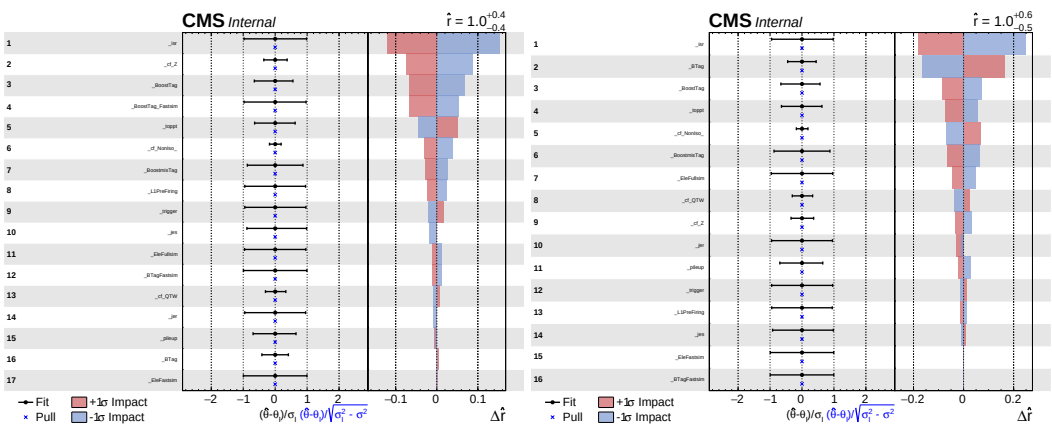


Figure 4.74: Blinded impacts for background+signal fit with T5qqqqWH($\tilde{g}:1600, \tilde{\chi}_0^1:500$) and R2bbqqlv($\tilde{b}:700, \tilde{\chi}_0^1:500$).

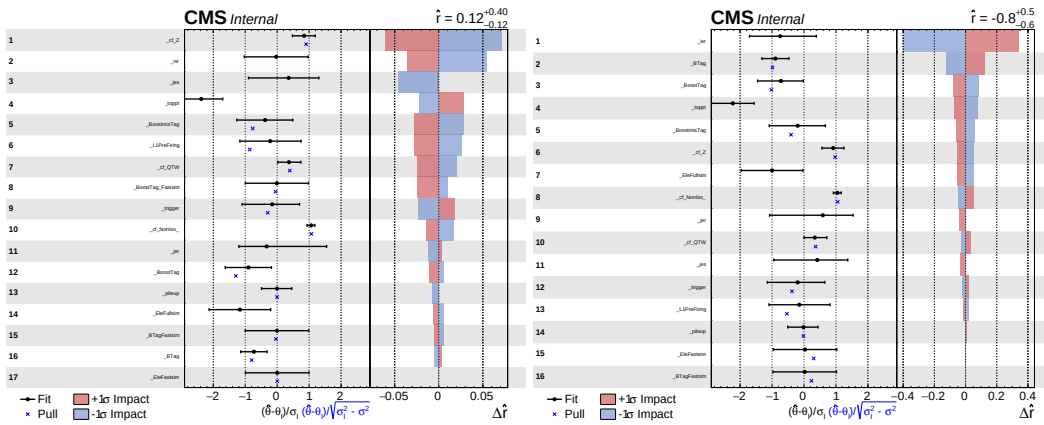


Figure 4.75: Unblinded impacts with T5qqqqWH($\tilde{g}$: 1600, $\tilde{\chi}_0^1$: 500) and R2bbqqlv($\tilde{b}$: 700, $\tilde{\chi}_0^1$: 500).

4.4 Results and Interpretation

We present the results of the background estimation along with predicted yields from several signal points for all search regions in Figure 4.76 with binning with Table 4.2

The Figure 4.77 shows the 41.5 fb^{-1} data is allowed to unblind the observed data with predicted yields from several signal points for all search regions.

4.5 Interpretation with run2

We interpret the analysis results using the SMSs shown in Figures 3.1 and 3.2.

Limits are calculated through a binned maximum likelihood fit, using the binning shown in Section 4.4 for various search categories and for their combination. Experimental and theoretical systematic uncertainties are considered as the nuisance parameters of the likelihood functions, for which log-normal a prior distributions are assumed. Currently only the search regions are considered in the fit. Limits are shown in Figures 4.78- 4.85.

4.6 Interpretation with 2017 data

The 2017 Limits are calculated through a binned maximum likelihood fit, using the binning shown in Section 4.4 for various search categories and for their combination. Experimental and theoretical systematic uncertainties are considered as the nuisance parameters of the likelihood functions, for which log-normal a prior distributions are assumed. The observed and expected limits are shown in Figures 4.86-4.91.

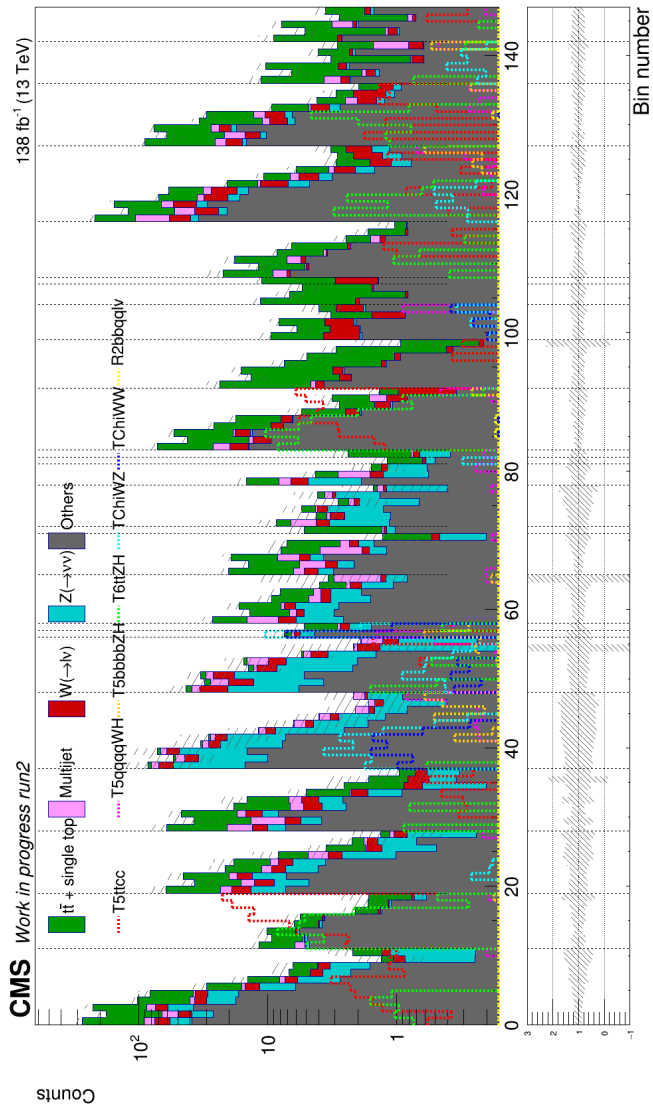


Figure 4.76: The $(M_R - 800) \times (R^2 - 0.08)$ distribution observed in data is shown along with the background prediction obtained for the SRs. Data/background prediction ratio is shown in the lower panels, where the gray band is the total uncertainty on the background prediction. Signal benchmark point: T5ttcc($\tilde{g}$: 1700, $\tilde{\chi}_0^1$: 200), T5qqqqWH, T5bbbbZH($\tilde{g}$: 2000, $\tilde{\chi}_0^1$: 500), T6ttZH($\tilde{t}$: 800, $\tilde{\chi}_0^1$: 100), TChiWZ, TChiWW($\tilde{\chi}_\pm^1$: 800, $\tilde{\chi}_0^1$: 100)

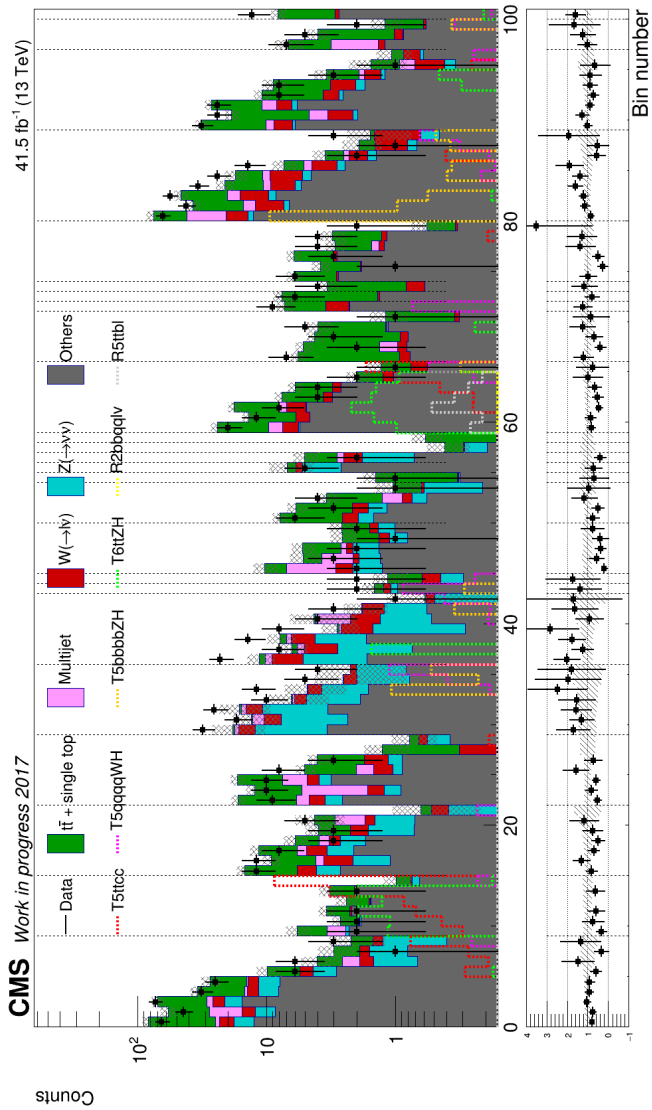


Figure 4.77: The $(M_R - 800) \times (R^2 - 0.08)$ distribution observed in data with 2017 is shown along with the background prediction obtained for the SRs. Data/background prediction ratio is shown in the lower panels, where the gray band is the total uncertainty on the background prediction. The difference between Data/background prediction is under investigation. Signal benchmark point: T5ttcc, T5qqqqWH, T5bbbbZH($\tilde{g}:1800, \tilde{\chi}_0^1:300$), T6ttZH($\tilde{t}:900, \tilde{\chi}_0^1:100$), R2bbqqlv, R5ttbl($\tilde{t}:1400, \tilde{\chi}_0^1:1000$).

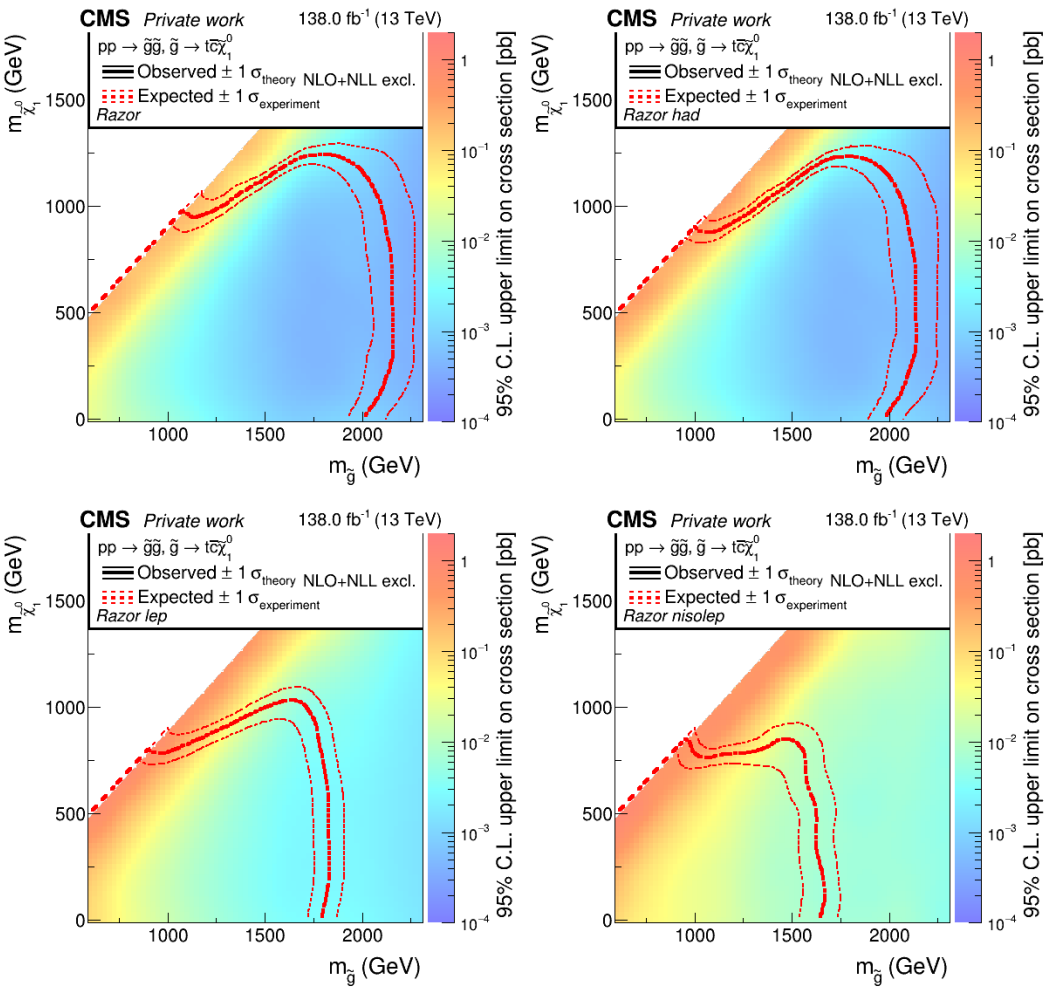


Figure 4.78: Observed 95% and expected upper limits on the signal cross sections using asymptotic CLs versus gluino and neutralino masses for the T5ttcc models with All SRs(top left), hadronic SRs(top right), isolated leptonic SRs(top left), and nonisolated SRs(top right). Also shown are the contours corresponding to the observed and expected lower limits, including their uncertainties, on the gluino and neutralino masses.

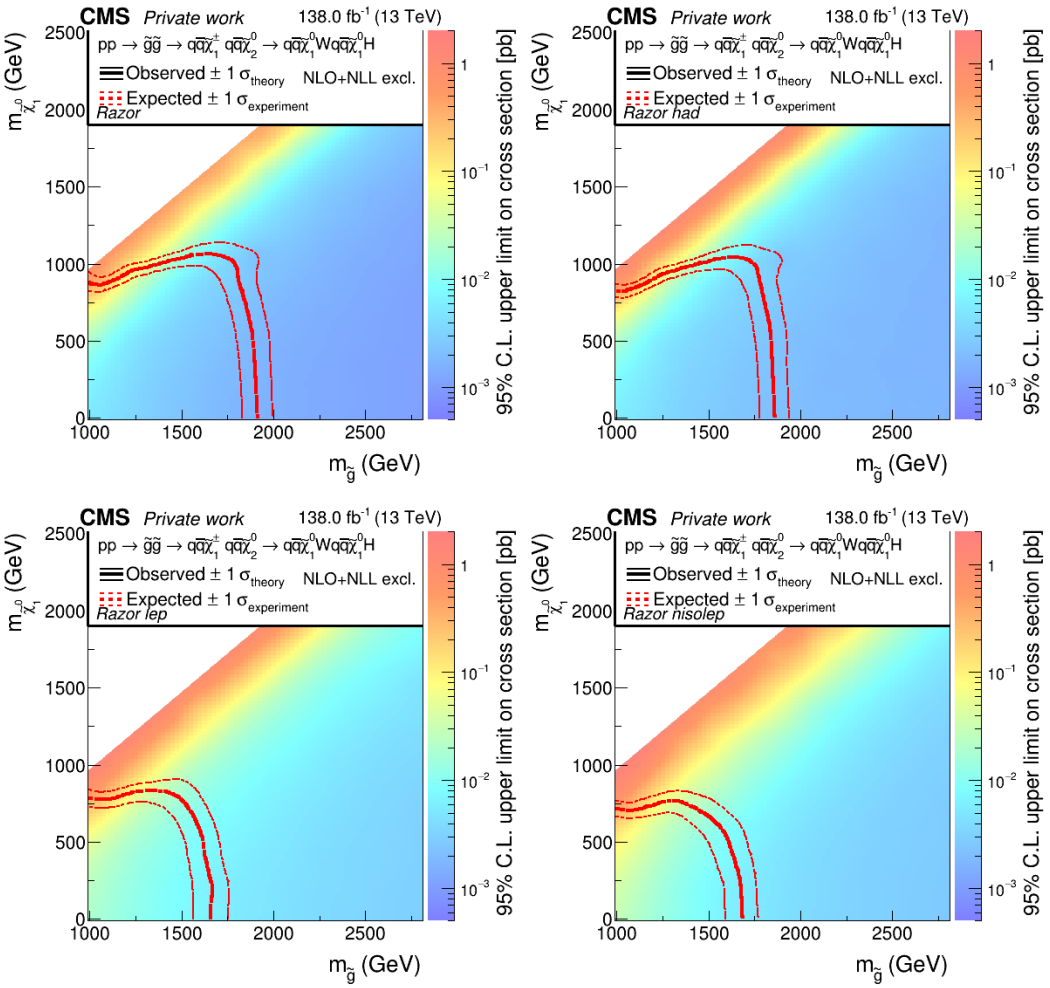


Figure 4.79: Observed 95% and expected upper limits on the signal cross sections using asymptotic CLs versus gluino and neutralino masses for the T5qqqqWH models with All SRs(top left), hadronic SRs(top right), isolated leptonic SRs(top left), and nonisolated SRs(top right). Also shown are the contours corresponding to the observed and expected lower limits, including their uncertainties, on the gluino and neutralino masses.

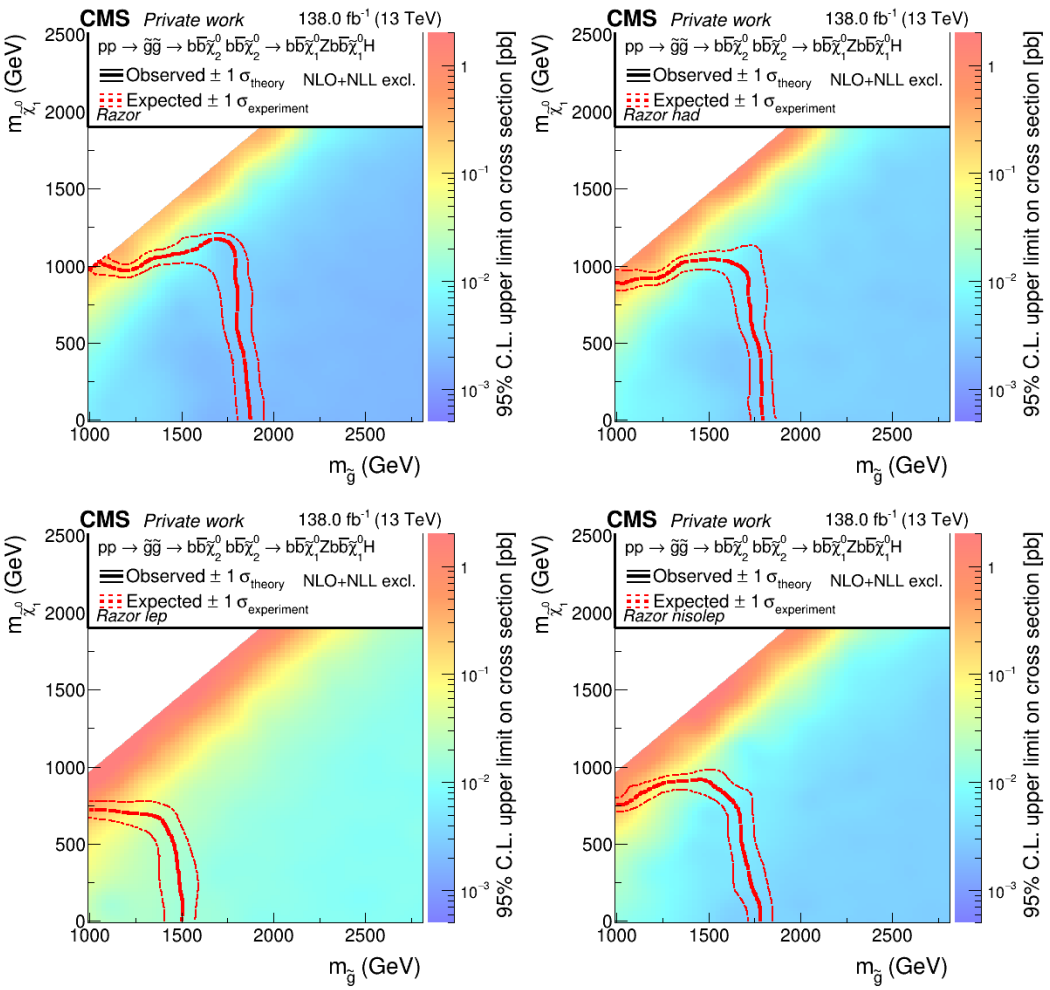


Figure 4.80: Observed 95% and expected upper limits on the signal cross sections using asymptotic CLs versus gluino and neutralino masses for the T5bbbbZH model with All SRs(top left), hadronic SRs(top right), isolated leptonic SRs(top left), and nonisolated SRs(top right). Also shown are the contours corresponding to the observed and expected lower limits, including their uncertainties, on the gluino and neutralino masses.

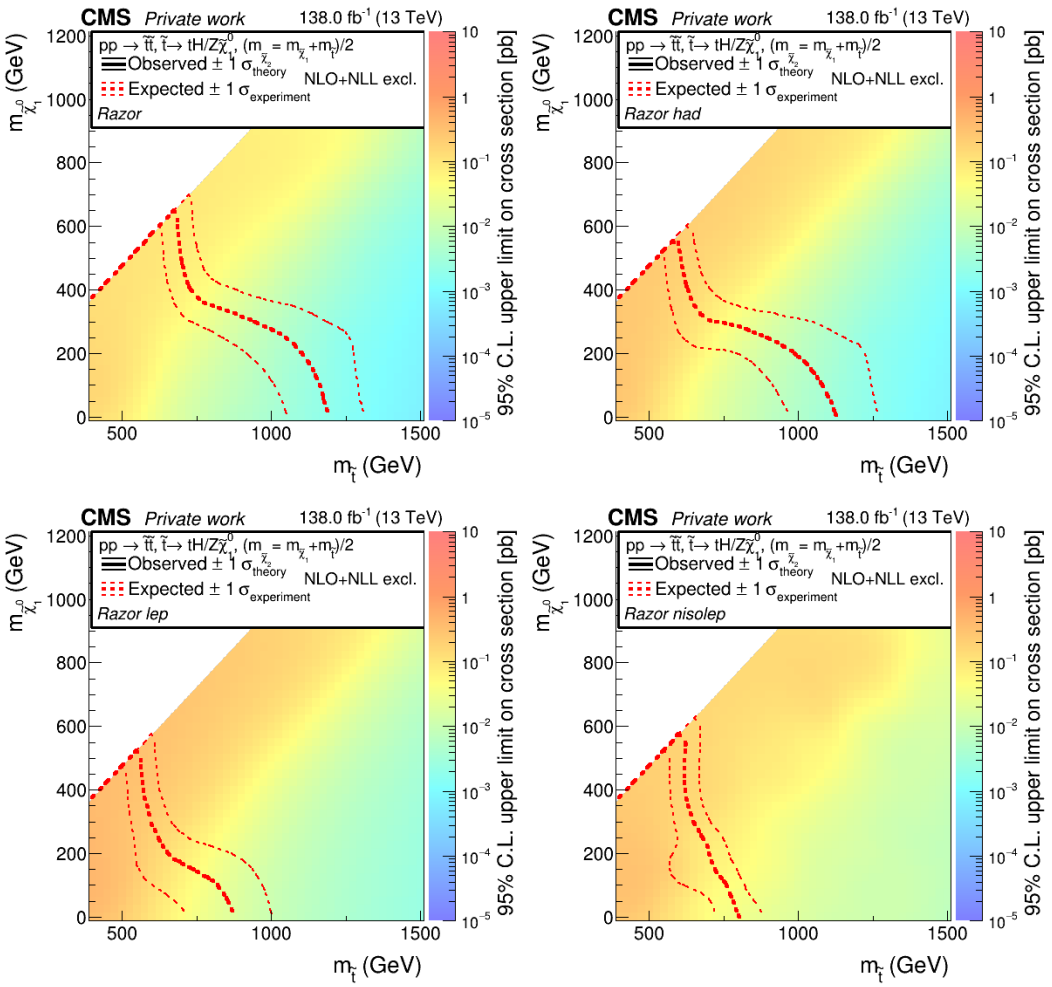


Figure 4.81: Observed 95% and expected upper limits on the signal cross sections using asymptotic CLs versus top squarks and neutralino masses for the T6ttZH model with All SRs(top left), hadronic SRs(top right), isolated leptonic SRs(top left), and nonisolated SRs(top right). Also shown are the contours corresponding to the observed and expected lower limits, including their uncertainties, on the top squarks and neutralino masses.

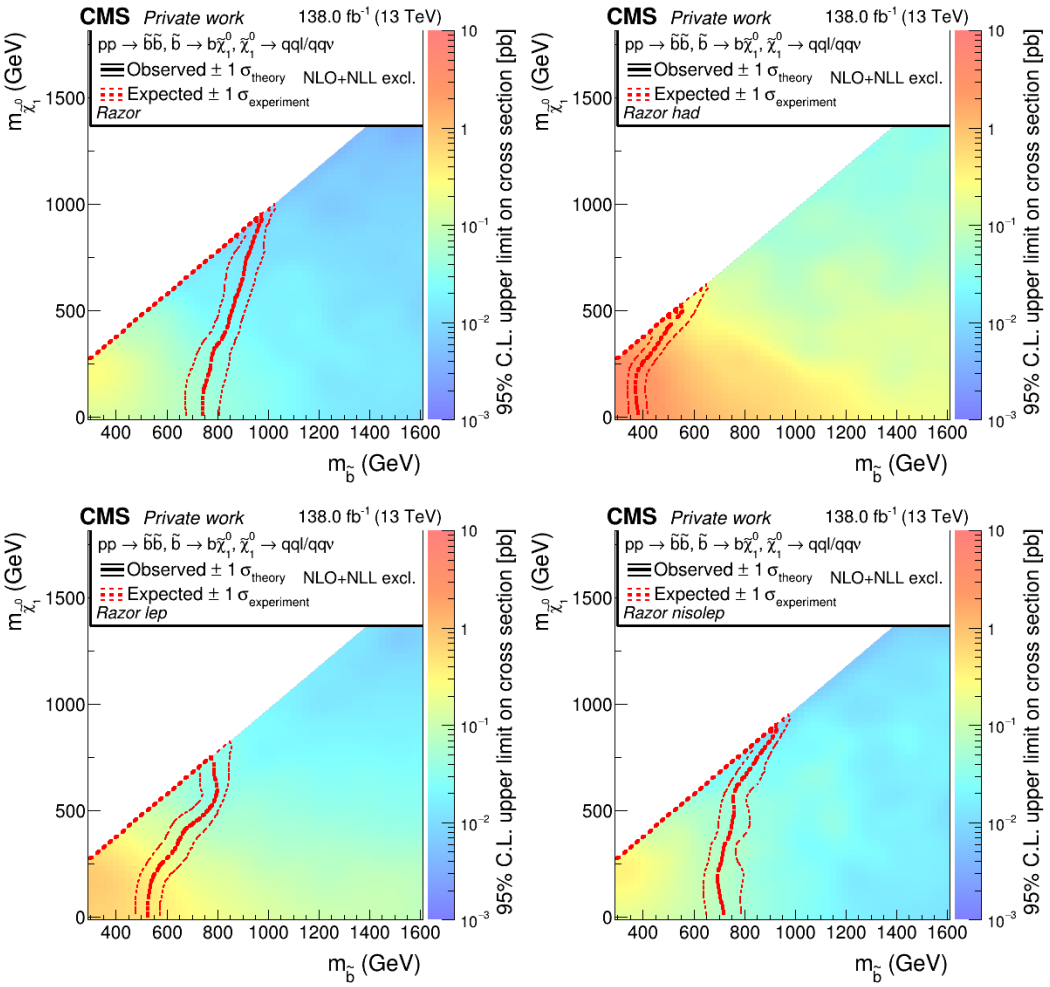


Figure 4.82: Observed 95% and expected upper limits on the signal cross sections using asymptotic CLs versus top squarks and neutralino masses for the R2bbqqlv model. Also shown are the contours corresponding to the observed and expected lower limits, including their uncertainties, on the top squarks and neutralino masses.

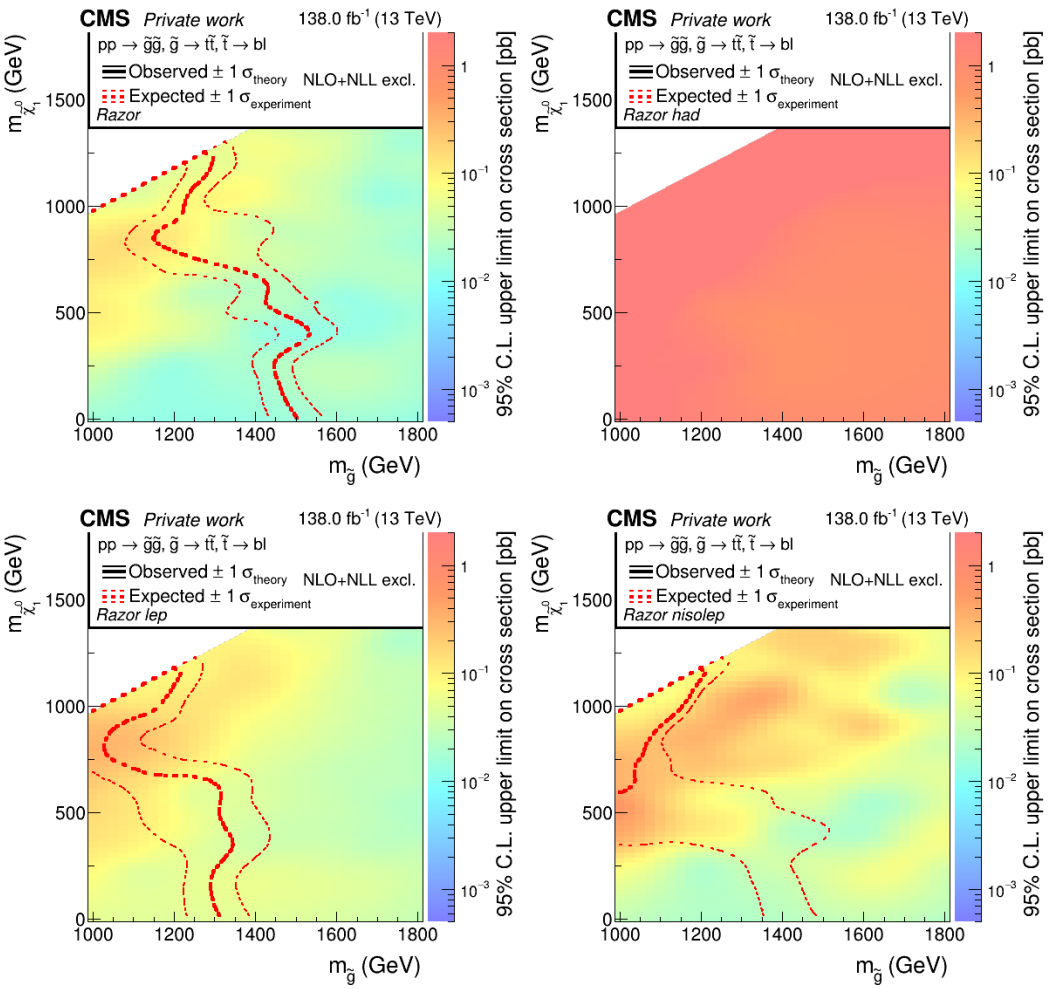
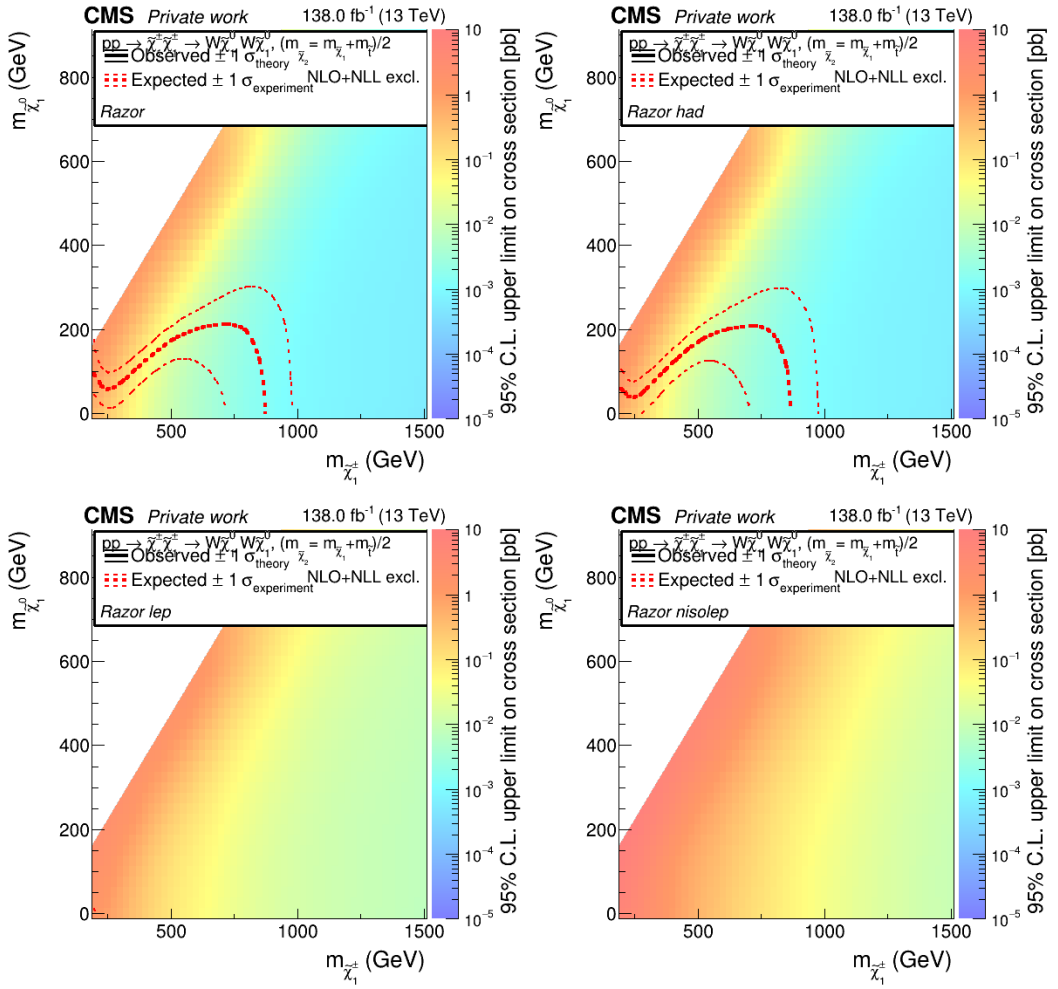
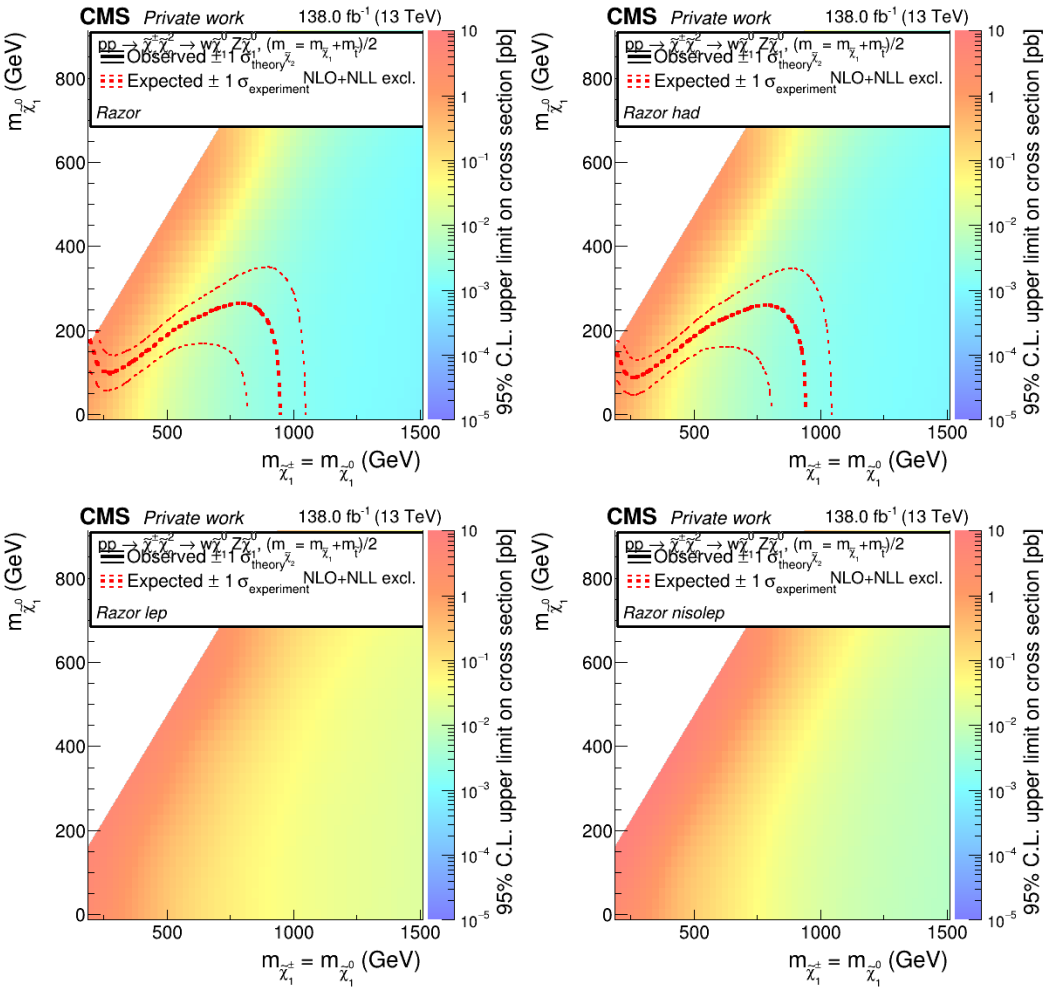


Figure 4.83: Observed 95% and expected upper limits on the signal cross sections using asymptotic CLs versus top squarks and neutralino masses for the R5ttbl model. Also shown are the contours corresponding to the observed and expected lower limits, including their uncertainties, on the top squarks and neutralino masses.



d

Figure 4.84: Observed 95% and expected upper limits on the signal cross sections using asymptotic CLs versus chargino and neutralino masses for the TChiWW model with All SRs(top left), hadronic SRs(top right), isolated leptonic SRs(top left), and nonisolated SRs(top right). Also shown are the contours corresponding to the observed and expected lower limits, including their uncertainties, on the chargino and neutralino masses.



d

Figure 4.85: Observed 95% and expected upper limits on the signal cross sections using asymptotic CLs versus chargino and neutralino masses for the TChiWZ model with All SRs(top left), hadronic SRs(top right), isolated leptonic SRs(top left), and nonisolated SRs(top right). Also shown are the contours corresponding to the observed and expected lower limits, including their uncertainties, on the chargino and neutralino masses.

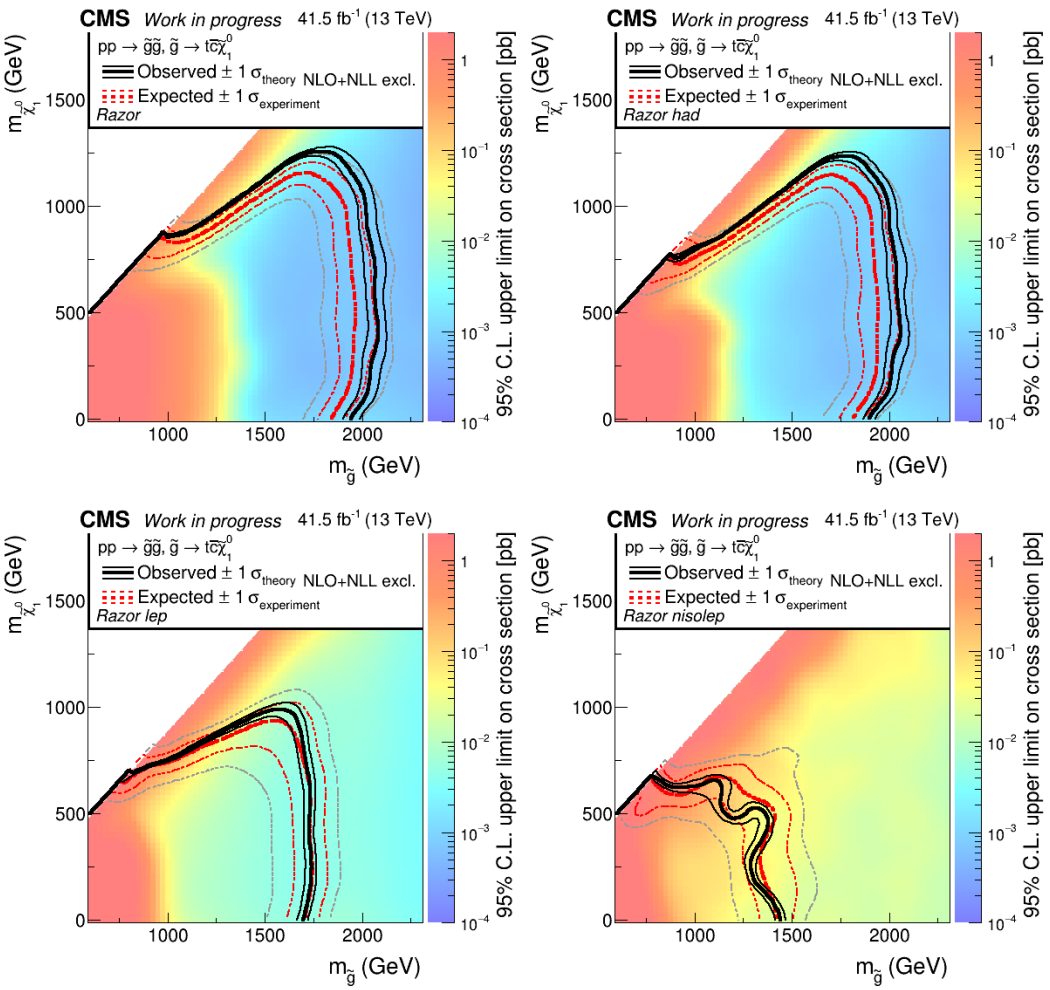


Figure 4.86: Observed 95% CL upper limits on the signal cross sections versus the gluino and neutralino mass for the T5ttcc model with All SRs (top left), hadronic SRs (top right), isolated leptonic SRs (top left), and nonisolated SRs (top right) for 2017 data. Also shown are the contours corresponding to the observed and expected lower limits, including their uncertainties, on the gluino and neutralino masses.

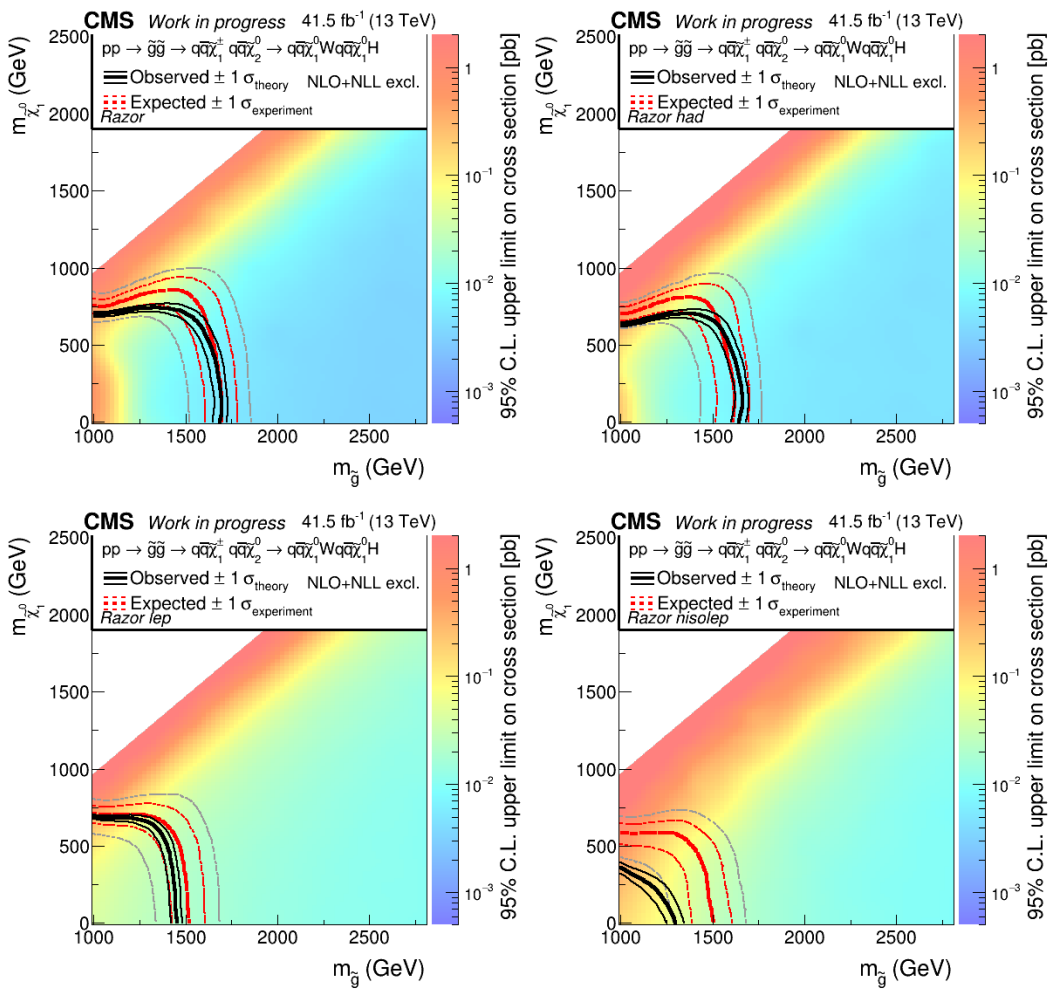


Figure 4.87: Observed 95% CL upper limits on the signal cross sections versus the gluino and neutralino mass for the T5qqqqWH model with All SRs (top left), hadronic SRs (top right), isolated leptonic SRs (top left), and nonisolated SRs (top right) for 2017 data. Also shown are the contours corresponding to the observed and expected lower limits, including their uncertainties, on the gluino and neutralino masses.

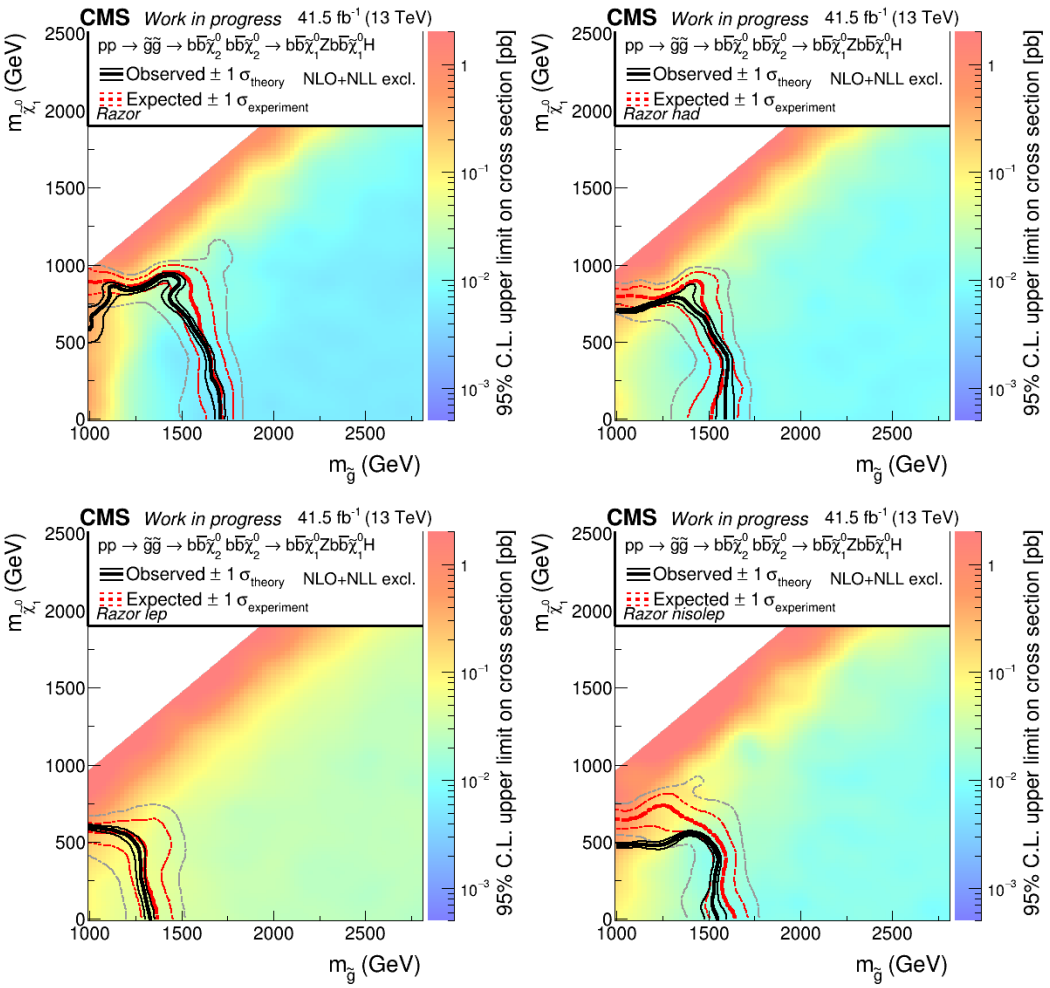


Figure 4.88: Observed 95% CL upper limits on the signal cross sections versus the gluino and neutralino mass for the T5bbbbZH model with All SRs (top left), hadronic SRs (top right), isolated leptonic SRs (top left), and nonisolated SRs (top right) for 2017 data. Also shown are the contours corresponding to the observed and expected lower limits, including their uncertainties, on the gluino and neutralino masses. The results are under investigation.

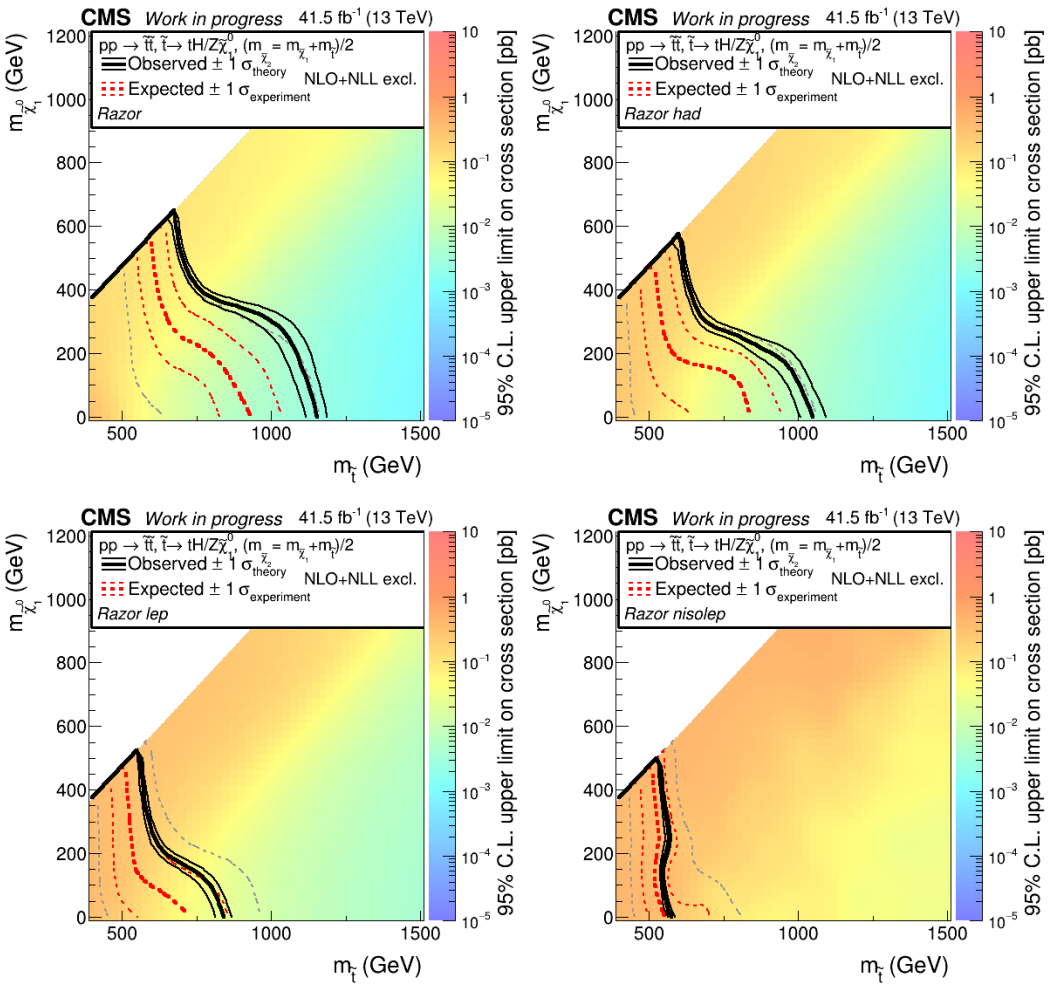


Figure 4.89: Observed 95% CL upper limits on the signal cross sections versus the gluino and neutralino mass for the T6ttZH model with All SRs (top left), hadronic SRs (top right), isolated leptonic SRs (top left), and nonisolated SRs (top right) for 2017 data. Also shown are the contours corresponding to the observed and expected lower limits, including their uncertainties, on the top squark and neutralino masses. The results are under investigation.

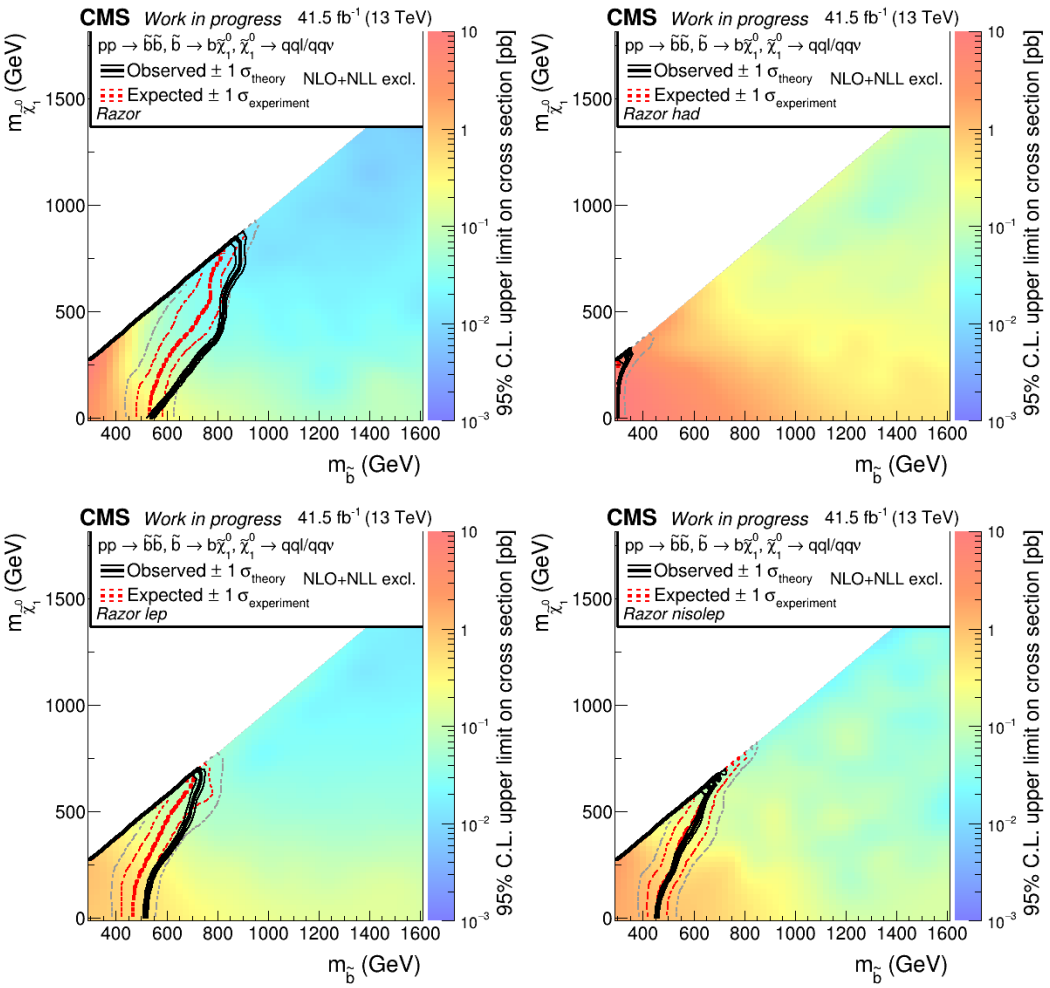


Figure 4.90: Observed 95% CL upper limits on the signal cross sections versus the gluino and neutralino mass for the T5bbbbZH model with All SRs (top left), hadronic SRs (top right), isolated leptonic SRs (top left), and nonisolated SRs (top right) for 2017 data. Also shown are the contours corresponding to the observed and expected lower limits, including their uncertainties, on the bottom squark and neutralino masses. The results are under investigation.

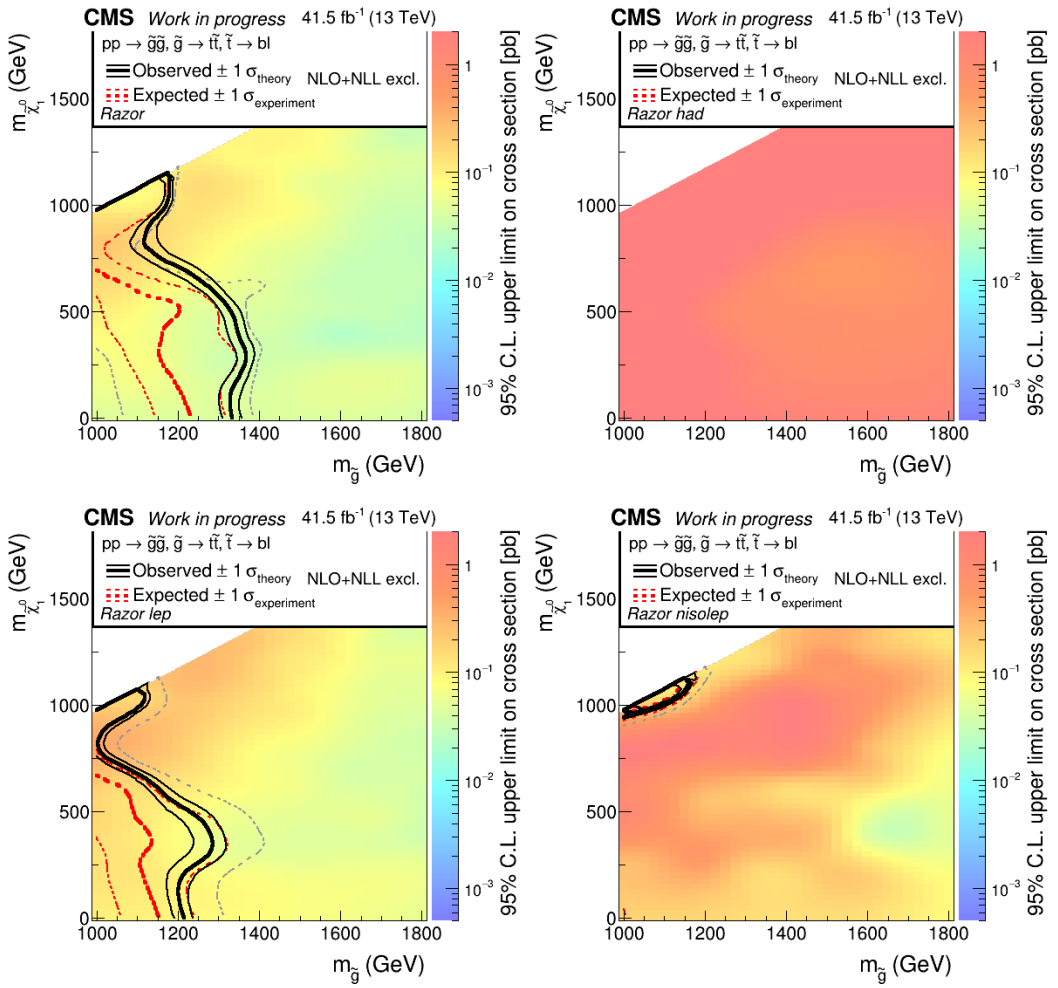


Figure 4.91: Observed 95% CL upper limits on the signal cross sections versus the gluino and neutralino mass for the T5bbbbZH model with All SRs (top left), hadronic SRs (top right), isolated leptonic SRs (top left), and nonisolated SRs (top right) for 2017 data. Also shown are the contours corresponding to the observed and expected lower limits, including their uncertainties, on the gluino and neutralino masses. The results are under investigation.

Chapter 5

Conclusion

We present a search for new physics in hadronic final states with at least one boosted W boson and a b-tagged jet or at least one boosted top quark using data binned at high values of the razor kinematic variables M_R and R^2. The analysis uses 138 fb^{-1} of 13 TeV proton-proton collision data collected by the CMS experiment. The SM backgrounds are estimated using control regions in data. Correction factors derived from simulations connect these control regions to the signal region. A deviation from the SM was searched in signal regions defined by jet, b-tagged jet, boosted W boson, and boosted top quark multiplicities. We have estimated the SM background expectation in the signal regions using data control regions and transfer factors derived from simulation connecting signal and control regions. The background estimation procedure was verified using multiple validation regions with kinematics resembling the signal regions. The run2(138 fb^{-1}) expected limits shows we could reject gluino mass up to 2 TeV and top squark mass up to 1 TeV. We did unblinding 2017 (41.5 fb^{-1}) and observed and expected limits are agreed within 2σ.

Appendix A

Appendix

A.1 Razor variables

Supersymmetry predicts new particles are pair-produced in the pp collisions at the collider. The supersymmetric particles have a very short lifetime except for LSP(Lightest Supersymmetric particle) and decayed to SM particles. LSP is a stable particle, and it can't be detected in the detector.

Razor variables are calculated from this feature. Two massive SUSY particles are produced from the pp collisions. That decays to an invisible massive particle like LSP and a massless visible particle, i.e., interacting with detector material, SM particles. This signature is illustrated in A.1.

Kinematical variables for new physics search in collider targeting this topology [32]- [33]. Most kinematical variables rely on the presence of the LSP. This causes the visible system to deviate from a dijet topology, resulting in possibly large missing transverse momentum, altered angular distributions, etc. All of this can be used to distinguish the SUSY signal from the known SM background processes. Mass reconstruction of the new particles cannot be possible. Because of the LSPs, there is not enough information available to fully constrain the problem. But we can approximate the mass scale of the particles in a collision, which can be used to check for evidence of new particles. The variables calculated from the mass scale exhibit a kinematic edge. The razor variables [34]- [35] are also have them. One

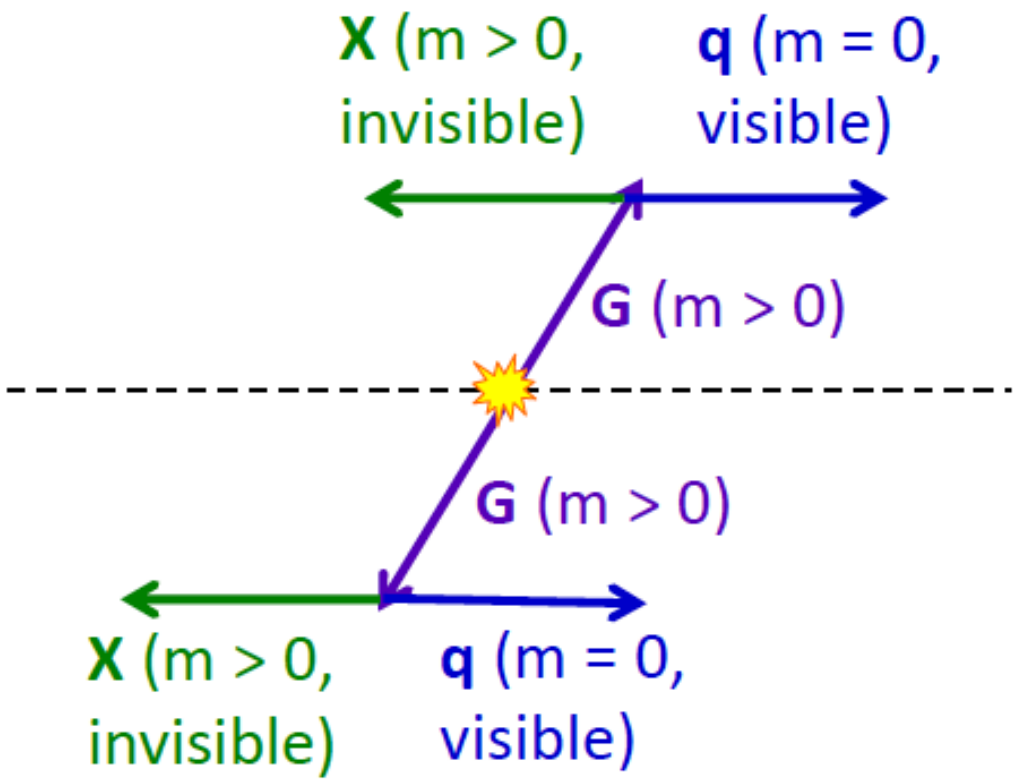


Figure A.1: Generic new physics signature. Two massive SUSY particles, G, are produced in pp collisions at the LHC, and, consequently, decay to a visible particle q and an invisible particle X.

advantage of the razor variables is they reconstruct the mass scale as a peak, in addition to an edge. The razor variables, denoted MR and R2, which use longitudinal and transverse information, estimate a characteristic mass scale associated with the new particles.

Kinematic Configuration

In A.1, we can consider Gs are two body products. Each G decay to the visible mass less SM particle q and invisible particle X. The four relevant reference frames determine a characteristic mass scale of the new physics process. The below paragraphs define the notations that will be used, as well as deriving relations between several variables.

Center of mass frame

In the center of mass(CM) frame of the considered from collision events of the particles G is produced with equal and opposite velocities $\vec{\beta}^{CM}$. The boost $\vec{\beta}^{CM}$ is how far above the threshold the X particles are produced, but unfortunately, this is invisible at hadron colliders.

To go from the rest frame of X to the CM frame, we need to boost the four momenta $p_1^{CM}(p_2^{CM})$ to the frame traveling at velocity $\vec{\beta}^{CM}(-\vec{\beta}^{CM})$ with respect to the G rest frame. The four vectors of particles G in the CM frame are also obtained by boosting according to $\vec{\beta}^{CM}$. They can be written as,

$$\begin{aligned} p_1^{CM} &= \left\{E_1^{CM}, \vec{p}_1^{CM}\right\} = M_G \gamma^{CM} \left\{1, \vec{\beta}^{CM}\right\} \\ p_2^{CM} &= \left\{E_2^{CM}, \vec{p}_2^{CM}\right\} = M_G \gamma^{CM} \left\{1, -\vec{\beta}^{CM}\right\} \end{aligned} \tag{A.1}$$

and satisfy the following

$$\sqrt{\hat{s}} = p_1^{CM} + p_2^{CM} = 2\gamma^{CM} M_G \tag{A.2}$$

$\sqrt{\hat{s}}$ is the CM energy of the collision.

G rest frame

Particles q and X are produced back-to-back with an equal magnitude of momentum in the rest frame of the G particle from two body decay kinematics. We can compute the magnitude of momentum in terms of the X and q masses. To do this, we start from the four vectors of the q and X particles in the G rest frame,

$$\begin{aligned} P[q] &= \nu_q^G = \left\{E_q^G, \vec{\nu}_q^G\right\} \\ P[X] &= \nu_X^G = \left\{E_X^G, \vec{\nu}_X^G\right\} \end{aligned} \tag{A.3}$$

Energy conservation in the G rest frame leads to

$$E_q^G + E_X^G = M_G \tag{A.4}$$

This can also be expressed as

$$M_G = \sqrt{M_q^2 + (\vec{\nu}_q^G)^2} + \sqrt{M_X^2 + (\vec{\nu}_X^G)^2} \tag{A.5}$$

q is massless, and the equal amount of momenta $|\vec{\nu}_q^G| = |\vec{\nu}_X^G| = |\vec{p}^G|$, the above equation can be simplified as,

$$\begin{aligned} M_G &= |\vec{p}^G| + \sqrt{M_X^2 + |\vec{p}^G|^2} \\ (\vec{p}^G)^2 &= M_G^2 - 2M_G\sqrt{M_X^2 + |\vec{p}^G|^2} + M_X^2 + |\vec{p}^G|^2 \\ M_G^2 + M_X^2 &= 2M_S\sqrt{M_X^2 + |\vec{p}^G|^2} \\ M_G^4 + 2M_G^2M_X^2 + M_X^4 &= 4M_G^2(M_X^2 + |\vec{p}^G|^2) \\ (M_G^2 - M_X^2)^2 &= 4M_G^2|\vec{p}^G|^2 \end{aligned} \tag{A.6}$$

We thus find the magnitude of the momentum of q and X in the G rest frame

$$|\vec{p}^G| = \frac{M_G^2 - M_X^2}{2M_G} \equiv \frac{M_\triangle}{2} \tag{A.7}$$

q is massless,

$$E_q^G = |\vec{p}^G| \equiv \frac{M_\triangle}{2} \tag{A.8}$$

in the rest of the frame,

$$\begin{aligned} E_{q_1}^R &= \gamma^R(E_{q_1}^{lab} - \beta^R \vec{p}_{1z}^{lab}) \\ \vec{p}_{1z}^R &= \gamma^R(\vec{p}_{1z}^{lab} - \beta^R E_{q_1}^{lab}) \end{aligned} \tag{A.9}$$

$E_{q_1}^R$ and $E_{q_2}^R$ is same so,

$$\begin{aligned} E_{q_1}^{lab} - \beta^R \vec{p}_{1z}^{lab} &= E_{q_2}^{lab} - \beta^R \vec{p}_{2z}^{lab} \\ \beta^R &= \frac{E_{q_1}^{lab} - E_{q_2}^{lab}}{\vec{p}_{1z}^{lab} - \vec{p}_{2z}^{lab}} \\ \gamma^R &= \frac{1}{\sqrt{1 - \left(\frac{E_{q_1}^{lab} - E_{q_2}^{lab}}{\vec{p}_{1z}^{lab} - \vec{p}_{2z}^{lab}}\right)^2}} \\ &= \frac{\vec{p}_{1z}^{lab} - \vec{p}_{2z}^{lab}}{\sqrt{\left(\vec{p}_{1z}^{lab} - \vec{p}_{2z}^{lab}\right)^2 - \left(E_{q_1}^{lab} - E_{q_2}^{lab}\right)^2}} \end{aligned} \tag{A.10}$$

$$\begin{aligned} M_R &= E_{q_1}^R + E_{q_2}^R \\ &= \gamma^R \left(E_{q_1}^{lab} - \beta^R \vec{p}_{1z}^{lab} + E_{q_2}^{lab} - \beta^R \vec{p}_{2z}^{lab}\right) \\ &= 2\sqrt{\frac{\left(E_{q_2}^{lab} \vec{p}_{1z}^{lab} - (E_{q_1}^{lab} \vec{p}_{2z}^{lab}\right)^2}{\left(\vec{p}_{1z}^{lab} - \vec{p}_{2z}^{lab}\right)^2 - \left(E_{q_1}^{lab} - E_{q_2}^{lab}\right)^2}} \end{aligned} \tag{A.11}$$

Let's consider longitudinal boost (β^{L^*}) and transverse boost ($\beta_T^{R^*}$) from the lab frame. momenta of q_1, q_2 is equal in R^* frame.

$$\begin{aligned} E_{q_1}^{L^*} &= \gamma^{L^*} \left(E_{q_1}^{lab} - \beta^{L^*} p_{1z}^{lab}\right) \\ E_{q_2}^{L^*} &= \gamma^{L^*} \left(E_{q_2}^{lab} - \beta^{L^*} p_{2z}^{lab}\right) \\ E_{q_1}^{R^*} &= \gamma_T^{R^*} \left(E_{q_1}^{L^*} - \vec{\beta}_T^{R^*} \vec{p}_{1T}^{lab}\right) \\ E_{q_2}^{R^*} &= \gamma_T^{R^*} \left(E_{q_2}^{L^*} - \vec{\beta}_T^{R^*} \vec{p}_{2T}^{lab}\right) \end{aligned} \tag{A.12}$$

$E_{q_1}^{R^*}$ and $E_{q_2}^{R^*}$ is same. So,

$$\begin{aligned} &\gamma_T^{R^*} \left(E_{q_1}^{L^*} - \vec{\beta}_T^{R^*} \vec{p}_{1T}^{lab}\right) - \gamma_T^{R^*} \left(E_{q_2}^{L^*} - \vec{\beta}_T^{R^*} \vec{p}_{2T}^{lab}\right) = 0 \\ &\gamma^{L^*} \left(E_{q_1}^{lab} - E_{q_2}^{lab}\right) - \gamma^{L^*} \beta^{L^*} \left(p_{1z}^{lab} - p_{2z}^{lab}\right) = \vec{\beta}_T^{R^*} \left(\vec{p}_{1T}^{lab} + \vec{p}_{2T}^{lab}\right) \end{aligned} \tag{A.13}$$

$\vec{\beta}_T^{R^*}$ is represent to $\beta_T^{R^*}\hat{\beta}_T^{R^*}$, now we can modify the equation as,

$$\beta^{R^*} = \frac{\gamma^{L^*}\left(E_{q_1}^{lab} - E_{q_2}^{lab}\right) - \gamma^{L^*}\beta^{L^*}\left(p_{1z}^{lab} - p_{2z}^{lab}\right)}{\hat{\beta}_T^{R^*}\left(\vec{p}_{1T}^{lab} + \vec{p}_{2T}^{lab}\right)} \tag{A.14}$$

$$\begin{aligned} M_{R^*} &= E_{q_1}^{R^*} + E_{q_2}^{R^*} \\ &= \gamma^{R^*}\left(\gamma^{L^*}\left(E_{q_1}^{lab} + E_{q_2}^{lab}\right) - \gamma^{L^*}\beta^{L^*}\left(p_{1z}^{lab} + p_{2z}^{lab}\right) - \beta_T^{R^*}\left(\vec{p}_{1T}^{lab} - \vec{p}_{2T}^{lab}\right)\right) \\ &= \frac{2\beta^{L^*}\hat{\beta}_T^{R^*}\left(\left(E_{q_1}^{lab}\vec{p}_{2T}^{lab} + E_{q_2}^{lab}\vec{p}_{1T}^{lab}\right) - \beta^{L^*}\left(p_{1z}^{lab}\vec{p}_{2T}^{lab} + p_{2z}^{lab}\vec{p}_{1T}^{lab}\right)\right)}{\sqrt{\left|\hat{\beta}_T^{R^*}\left(\vec{p}_{1T}^{lab} + \vec{p}_{2T}^{lab}\right)\right|^2 - \left(\gamma^{L^*}\right)^2\left(E_{q_1}^{lab} - E_{q_2}^{lab} - \beta^{L^*}\left(p_{1z}^{lab} - p_{2z}^{lab}\right)\right)^2}} \end{aligned} \tag{A.15}$$

We can consider transverse information only, so $\hat{\beta}_T^{R^*}$ represents to

$$\hat{\beta}_T^{R^*} = \frac{\vec{p}_{1T}^{lab} + \vec{p}_{2T}^{lab}}{\left|\vec{p}_{1T}^{lab} + \vec{p}_{2T}^{lab}\right|} \tag{A.16}$$

Then M_{R^*} is

$$M_{R^*} = \sqrt{\left(E_{q_1}^{lab} + E_{q_2}^{lab}\right)^2 + \left(p_{1z}^{lab} + p_{2z}^{lab}\right)^2 - \left(\frac{\left|\vec{p}_{1T}^{lab}\right|^2 - \left|\vec{p}_{2T}^{lab}\right|^2}{\left|\vec{p}_{1T}^{lab} + \vec{p}_{2T}^{lab}\right|}\right)^2} \tag{A.17}$$

M_{R^*} is invariant under longitudinal boost, so $\frac{\partial M_{R^*}}{\partial \beta_{L^*}} = 0$. From that equation, γ_{L^*} can be written by,

$$\beta_{\gamma_T^{R^*}} = \sqrt{\frac{\left(E_{q_1}^{lab} + E_{q_2}^{lab}\right)^2 - \left(p_{1z}^{lab} + p_{2z}^{lab}\right)^2}{\left(E_{q_1}^{lab} + E_{q_2}^{lab}\right)^2 + \left(p_{1z}^{lab} + p_{2z}^{lab}\right)^2 - \left(\frac{\left|\vec{p}_{1T}^{lab}\right|^2 - \left|\vec{p}_{2T}^{lab}\right|^2}{\left|\vec{p}_{1T}^{lab} + \vec{p}_{2T}^{lab}\right|}\right)^2}} \tag{A.18}$$

M_R is,

$$M_R = \gamma_T^{R^*} M_{R^*} = \sqrt{\left(E_{q_1}^{lab} + E_{q_2}^{lab}\right)^2 - \left(p_{1z}^{lab} + p_{2z}^{lab}\right)^2} \tag{A.19}$$

M_{2S} represent using 4-vectors, q_1, ν_1, q_2, ν_2

$$\begin{aligned} M_{2S} &= \sqrt{\frac{1}{2}\left[\left(q_1^{lab}+\nu_1^{lab}\right)^2+\left(q_2^{lab}+\nu_2^{lab}\right)^2\right]} \\ &= \sqrt{q_1^{lab}\nu_1^{lab}+q_2^{lab}\nu_2^{lab}+M_\chi^2} \end{aligned} \tag{A.20}$$

Let's assume $+M_\chi^2 = 0$,

$$M_{2S} = \sqrt{E_{q_1}^{lab}E_{\nu_1}^{lab}-\vec{q}_{1T}^{lab}\vec{\nu}_{1T}^{lab}-\vec{q}_{1z}^{lab}\vec{\nu}_{1z}^{lab}+E_{q_2}^{lab}E_{\nu_2}^{lab}-\vec{q}_{2T}^{lab}\vec{\nu}_{2T}^{lab}-\vec{q}_{2z}^{lab}\vec{\nu}_{2z}^{lab}} \tag{A.21}$$

$\vec{\nu}_{1z}^{lab} = \vec{\nu}_{2z}^{lab} = 0$, $E_{q_1}^{lab} = E_{q_2}^{lab} = 0$ in detector level. $\vec{\nu_{1T}^{lab}} = \vec{\nu_{2T}^{lab}}$, $\vec{E}_T^{miss} = \vec{\nu_{1T}^{lab}} + \vec{\nu_{2T}^{lab}}$

$$\begin{aligned} (M_{2S})_T &= \sqrt{\left|\vec{q}_{1T}^{lab}\right|\left|\vec{\nu}_{1T}^{lab}\right|-\vec{q}_{1T}^{lab}\vec{\nu}_{1T}^{lab}+\left|\vec{q}_{2T}^{lab}\right|\left|\vec{\nu}_{2T}^{lab}\right|-\vec{q}_{2T}^{lab}\vec{\nu}_{2T}^{lab}} \\ &= \sqrt{\frac{\left|\vec{E}_T^{miss}\right|}{2}\left(\left|\vec{q}_{1T}^{lab}\right|+\left|\vec{q}_{2T}^{lab}\right|\right)-\frac{\vec{E}_T^{miss}}{2}\left(\vec{q}_{1T}^{lab}+\vec{q}_{2T}^{lab}\right)} \end{aligned} \tag{A.22}$$

$$R = \frac{M_T^R}{M_R} \tag{A.23}$$

Appendix B

Appendix

B.1 PF Hadron calibration

In the particle flow algorithm, accurate calibration of the ECAL and HCAL energies is essential for reconstructing jets. This calibration is necessary because jets typically deposit energy in both the ECAL and HCAL, and the response of these calorimeters differs for different particles.

The calibration involves correcting the ECAL and HCAL energies separately to estimate the true hadronic energy deposits in each. Here's an overview of the calibration steps:

1. ECAL Calibration: The ECAL response to hadrons, electrons, and photons can vary. To account for these differences, the ECAL cluster energies are recalibrated. This calibration aims to estimate the true energy deposited by hadrons in the ECAL by considering the specific response characteristics of hadrons in this calorimeter.

2. HCAL Calibration: The response of the HCAL to hadrons is nonlinear. This means that the energy measured in the HCAL needs to be corrected to obtain an accurate estimate of the true hadronic energy. The calibration function used for the HCAL takes into account the energy measured in the HCAL and the pseudo-rapidity (η) of the HCAL cluster. The function is designed to correct for the nonlinear response and provide a more accurate measurement of the hadronic energy deposited

in the HCAL.

By calibrating the ECAL and HCAL cluster energies separately, the particle flow algorithm can better estimate the true hadronic energy deposits in each calorimeter. This allows for improved jet reconstruction, which is crucial for many physics analyses in particle physics experiments.

B.1.1 Run2

In the particle flow algorithm used during Run2, the calibration of calorimeter energy associated with charged hadrons is performed using a calibration function. This function accounts for the energy measured in the ECAL and HCAL, as well as the pseudo-rapidity (η) of the HCAL cluster. The calibrated energy ($E_{calibrated}$) is given by:

$$E_{calibrated} = a(E,\eta)E_{ECAL} + b(E,\eta)E_{HCAL} + c \tag{B.1}$$

Here, E_{ECAL} and E_{HCAL} represent the energies measured in the ECAL and HCAL, respectively. E is an estimate of the true energy, which is chosen to be either the total charged particle momentum or the total uncalibrated calorimetric energy, whichever is larger. The calibration function includes three terms: $a(E,\eta)$, $b(E,\eta)$, and c. These terms are derived separately for different categories of hadrons based on where their shower originates.

Specifically, the equation can be categorized into EH hadrons (hadrons with showers starting in ECAL) and H hadrons (hadrons with showers starting in HCAL), and it can be expressed as follows:

$$\begin{aligned} E_{corrected,EH\,hadron} &= a(E_{true})E_{ECAL} + b(E_{true})E_{HCAL} + o_{EH} \\ E_{corrected,H\,hadron} &= c(E_{true})E_{HCAL} + o_H \end{aligned} \tag{B.2}$$

In these equations, E_{true} represents the true energy of the hadron obtained from simulation. o_{EH} and o_H are the offset terms, with values of 3.5 GeV and 2.5 GeV, respectively.

To determine the calibration parameters $a(E_{true})$, $b(E_{true})$, and $c(E_{true})$, a χ^2

minimization is performed separately for the Barrel region ($|\eta| < 1.5$) and the Endcap region ($1.5 < |\eta| < 3.0$). The χ^2 is defined as:

$$\begin{aligned} \chi^2_{EH\,hadron} &= \sum_{hadrons} \left[\frac{E_{true} - E_{corrected}}{\sigma(E_{ECAL} + E_{HCAL})} \right]^2 \\ \chi^2_{H\,hadron} &= \sum_{hadrons} \left[\frac{E_{true} - E_{corrected}}{\sigma(E_{HCAL})} \right]^2 \end{aligned} \tag{B.3}$$

The calibration parameters are obtained by minimizing the χ^2 for each category of hadrons.

Figure B.1 shows the fitting results of the calibration functions $a(E_{true})$, $b(E_{true})$, and $c(E_{true})$ in both the Barrel and Endcap regions.

Figures B.2 and B.3 illustrate the response and resolution versus energy before and after the energy correction for EH hadrons and H hadrons. It can be observed that the energy correction improves both the response and resolution in all cases.

These calibration procedures and corrections are crucial for achieving accurate jet reconstruction and precise measurements in high-energy physics experiments.

After the energy correction, the response of the reconstructed jets exhibits a dependence on the pseudorapidity (η). To parametrize this η dependence, we introduce α and β as correction functions for EH hadrons and H hadrons. The energy-corrected values of E_{ECAL} and E_{HCAL} are denoted as E'_{ECAL} and E'_{HCAL}, respectively.

For EH hadrons, the η-dependent correction function is given by:

$$\begin{aligned} &|\eta| < 1.5 : E_{corrected} = (1 + \alpha(E_{true}) + \beta(E_{true}) \cdot |\eta|^2) \cdot E'_{ECAL} + E'_{HCAL} \\ &1.5 < |\eta| < 2.5 : E_{corrected} = (1 + \alpha(E_{true})) \cdot E'_{ECAL} + E'_{HCAL} \\ &2.5 < |\eta| : E_{corrected} = (1 + \alpha(E_{true}) + \beta(E_{true}) \cdot ((|\eta| - 1.5)^2 + 0.6)) \cdot E'_{ECAL} + E'_{HCAL} \end{aligned} \tag{B.4}$$

Similarly, for H hadrons, the η-dependent correction function is given by:

$$\begin{aligned} &|\eta| < 1.5 : E_{corrected} = (1 + \alpha(E_{true}) + \beta(E_{true}) \cdot |\eta|^2) \cdot E'_{HCAL} \\ &1.5 < |\eta| < 2.5 : E_{corrected} = (1 + \alpha(E_{true}) + \beta(E_{true}) \cdot 0.05) \cdot E'_{HCAL} \\ &2.5 < |\eta| : E_{corrected} = (1 + \alpha(E_{true}) + \beta(E_{true}) \cdot ((|\eta| - 1.5)^4 - 1.1)) \cdot E'_{HCAL} \end{aligned} \tag{B.5}$$

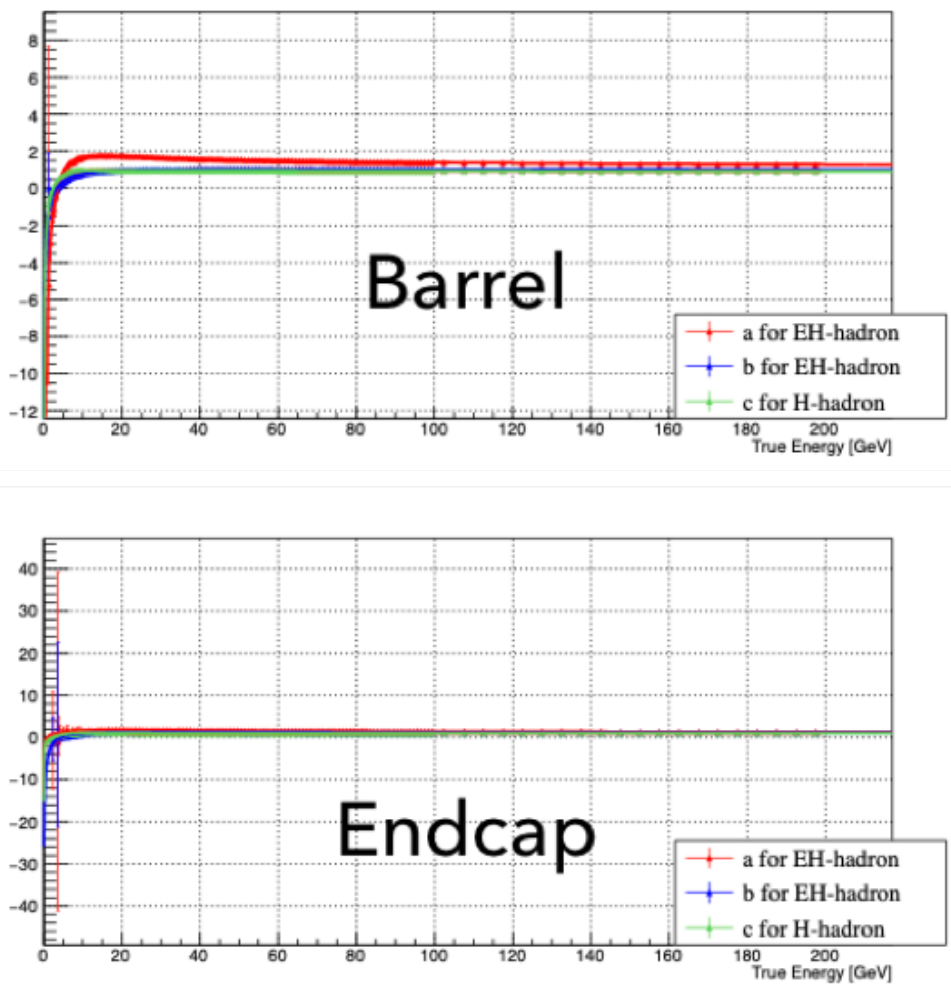


Figure B.1: Fitting results of calibration functions A, B, and C in the Barrel and Endcap regions.

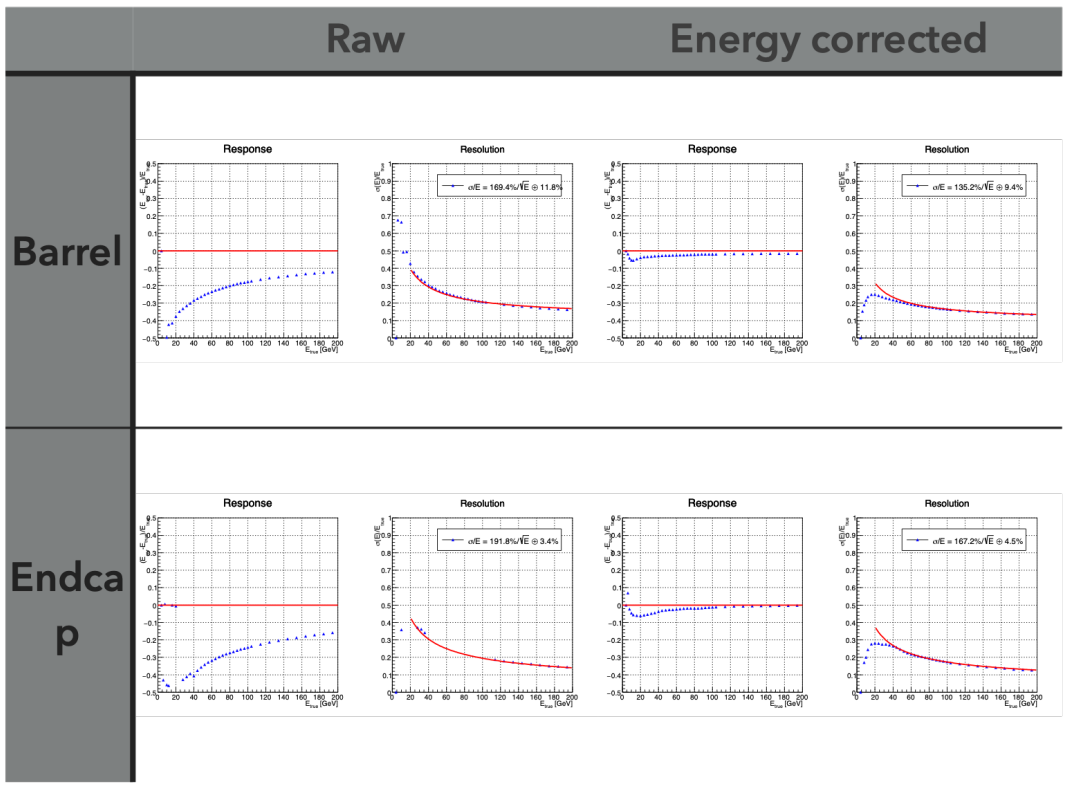


Figure B.2: Response and resolution versus energy before and after energy correction for EH hadrons.

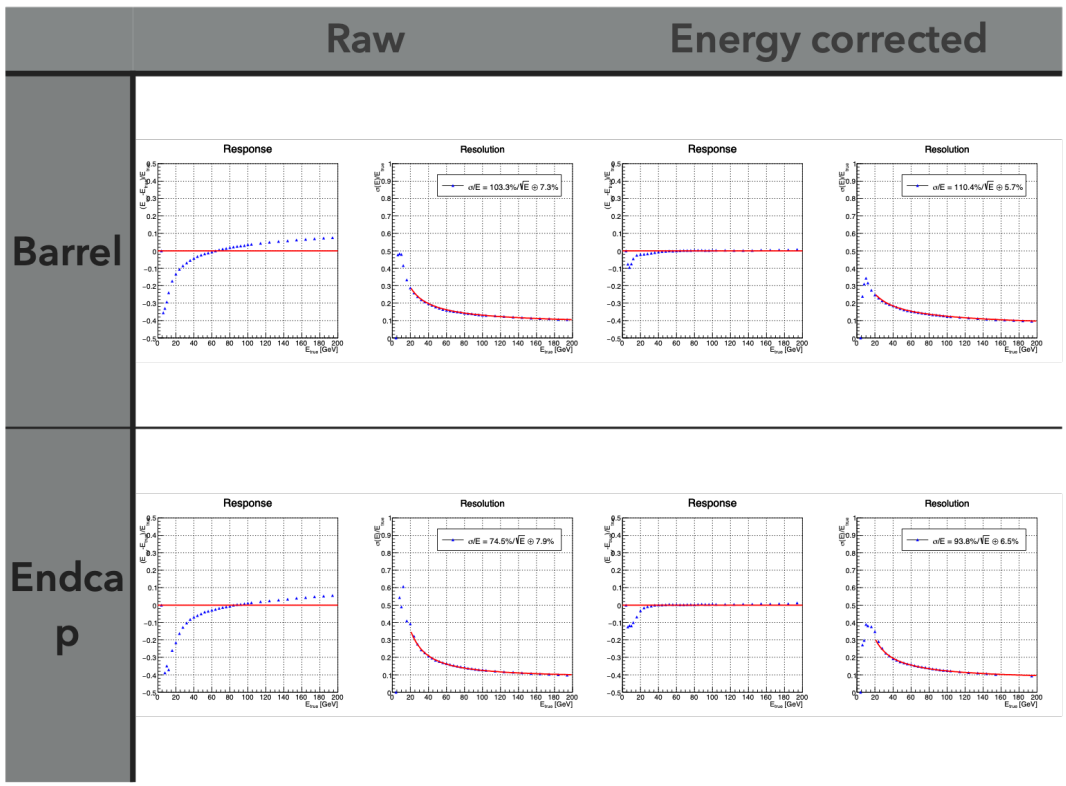


Figure B.3: Response and resolution versus energy before and after energy correction for H hadrons.

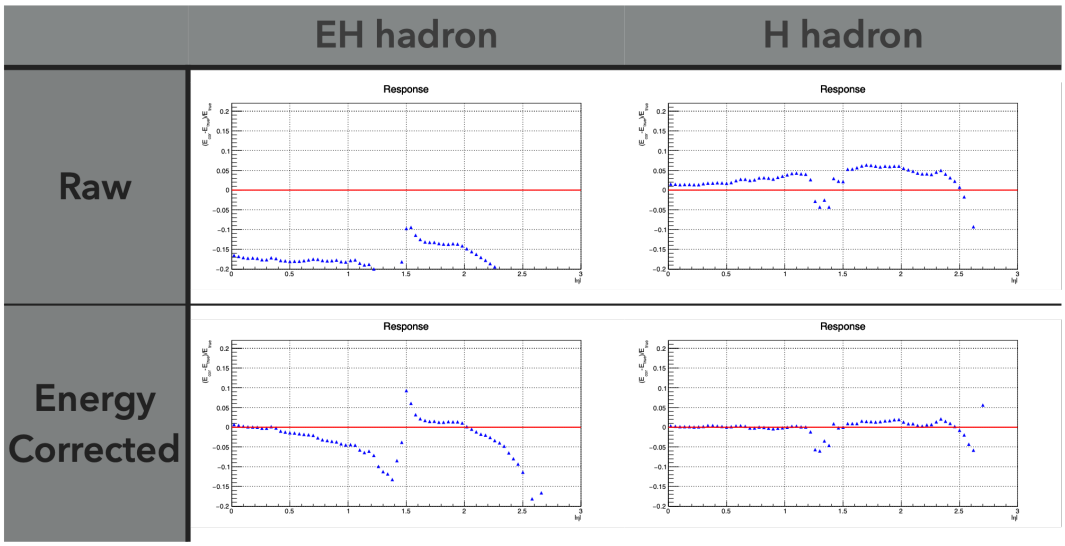


Figure B.4: Response vs $|\eta|$ with EH hadron and H hadron

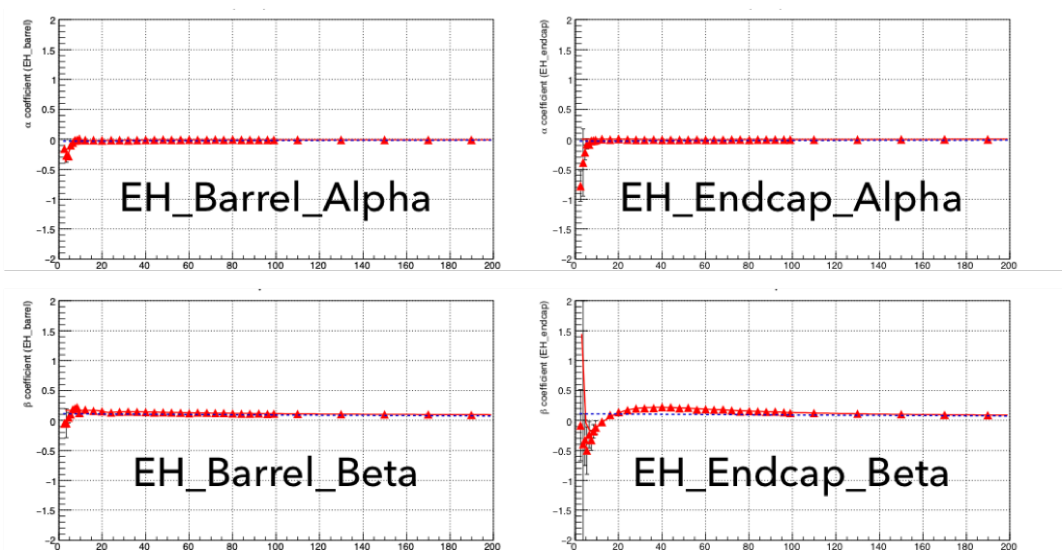


Figure B.5: Fitting results of the EH hadron η correction functions α and β in the Barrel and Endcap regions.

Figure B.5 and B.6 show the fits of the η-dependent correction functions $\alpha(E_{true})$ and $\beta(E_{true})$ for EH hadrons and H hadrons in the Barrel and Endcap regions.

Figures B.7 and B.8 depict the response and resolution versus actual energy with

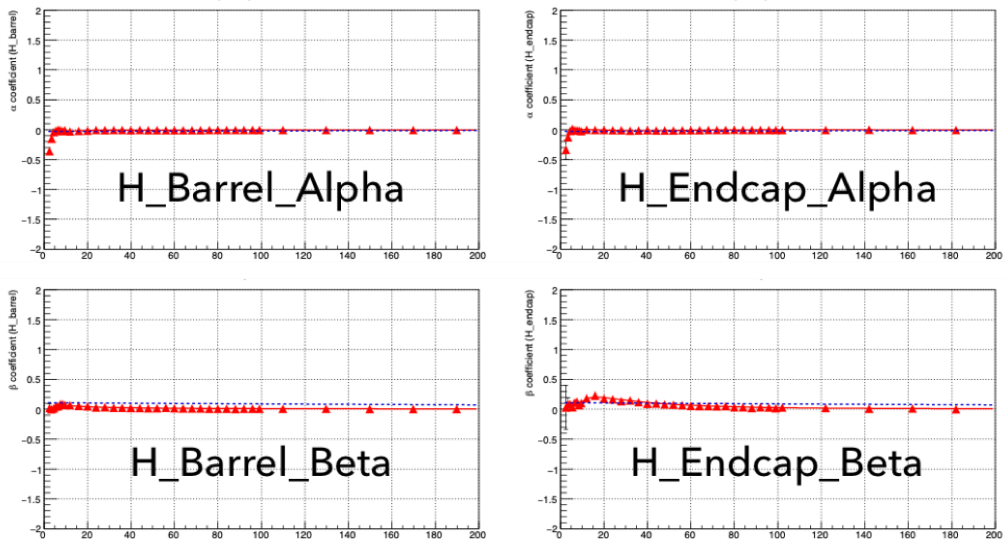


Figure B.6: Fitting results of the H hadron η correction functions α and β in the Barrel and Endcap regions.

EH hadrons and H hadrons after applying the η correction. The response shows significant improvement, particularly for energies exceeding 30 GeV.

Furthermore, Figures B.9 and B.10 show the response versus $|\eta|$ and η for EH hadrons and H hadrons after the η correction. The responses are significantly improved after applying the η correction. The response is worse near $|\eta| = 1.5$, corresponding to the ECAL gap region. Although we corrected the energy using the absolute value of η, the response versus η appears okay.

These calibration functions are implemented in the data acquisition process of the CMS experiment to correct the jet energy and ensure accurate measurements.

B.1.2 Run3

The algorithm of PF hadron calibration is the same as Run2. PF hadron calibration energy range is increased to 500 GeV for run3. Above 500 GeV, we extrapolate the calibration coefficient function. Figure B.11 shows the function of energy and eta calibration coefficient.

Figure B.12 shows the response and resolution of EH and H hadrons in each

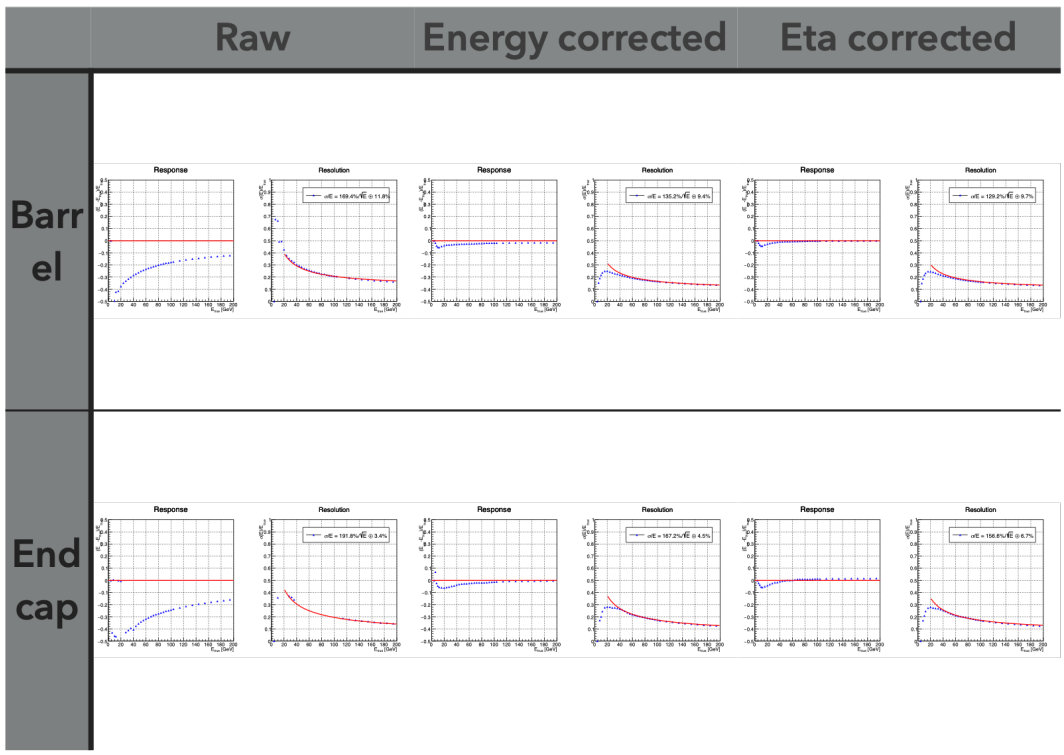


Figure B.7: Response and resolution versus true energy for EH hadrons after η correction.

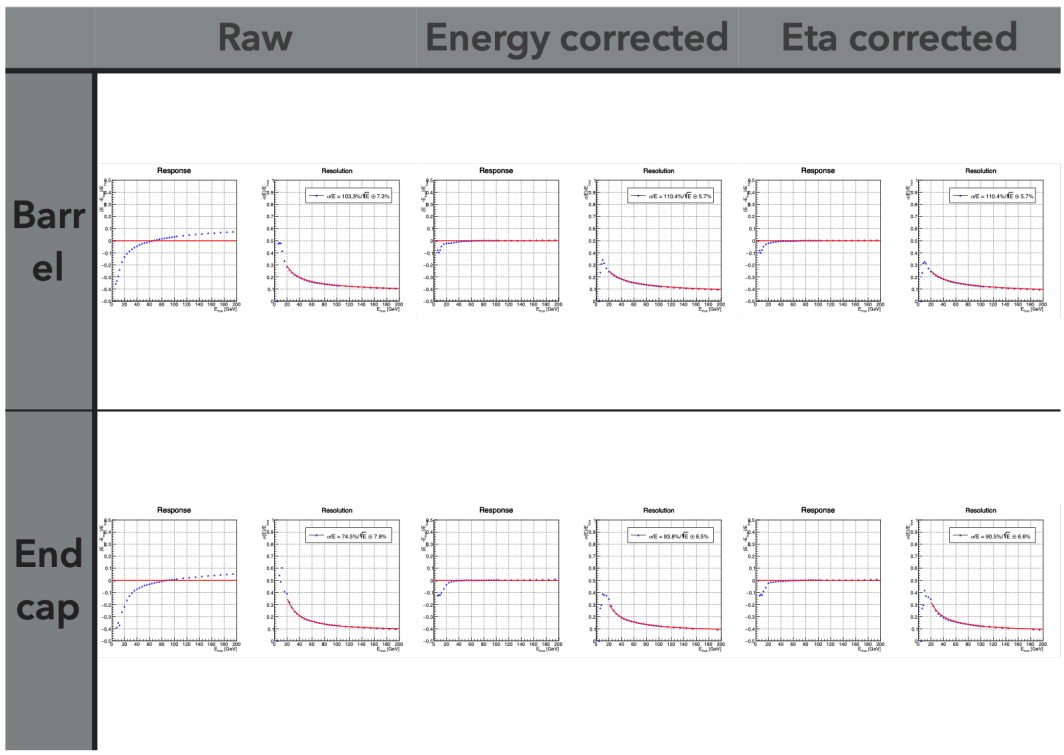


Figure B.8: Response and resolution versus true energy for H hadrons after η correction.

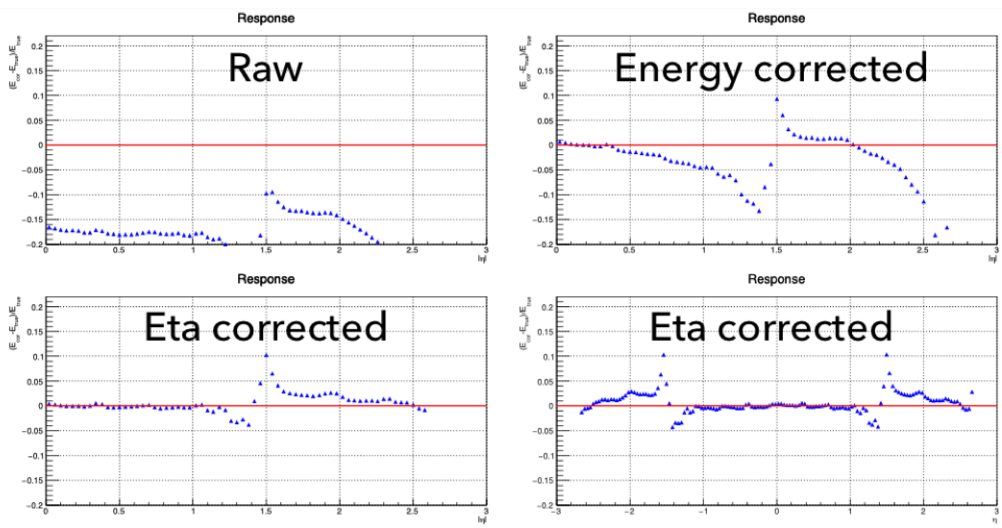


Figure B.9: Response versus $|\eta|$ and η for EH hadrons after η correction.

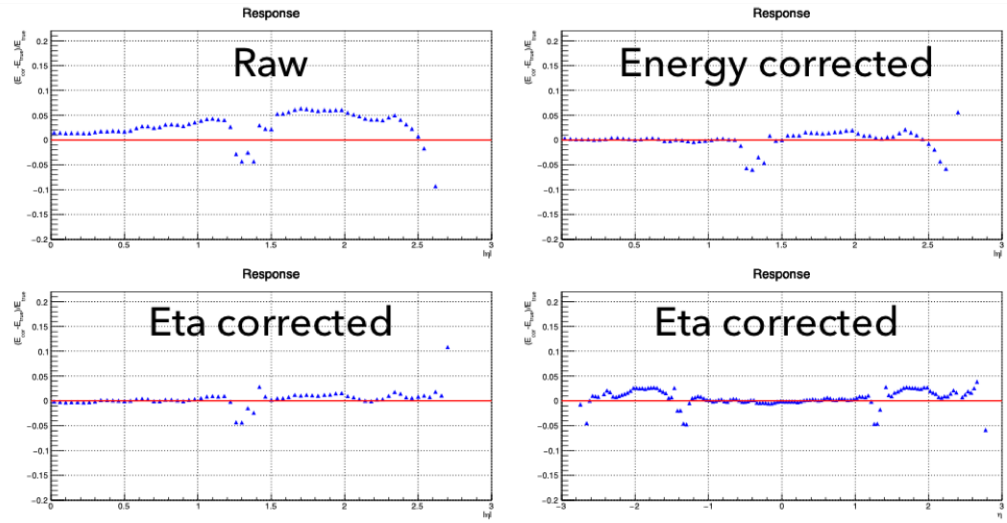


Figure B.10: Response versus $|\eta|$ and η for H hadrons after η correction.

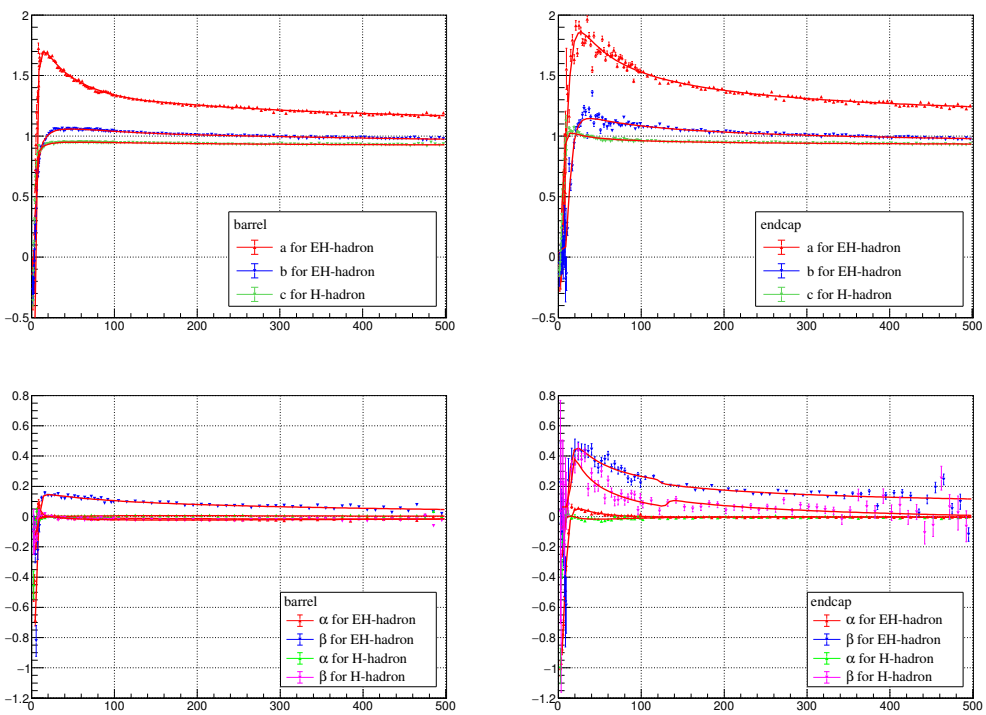


Figure B.11: Fluxes of calibration coefficients in each barrel and endcap case

barrel and endcap as an energy function. From the results, after the correction, the responses are flatter and closer to 1 in whole cases.

Figure B.13 shows the response results in η space. Also, in this response distribution, the response is close to 1 and flatter after correction except for the transition region, and there are no particular signatures in the sign of η.

This PF hadron calibration results are implemented when CMS data is taken for run3 to correct online jet energy.

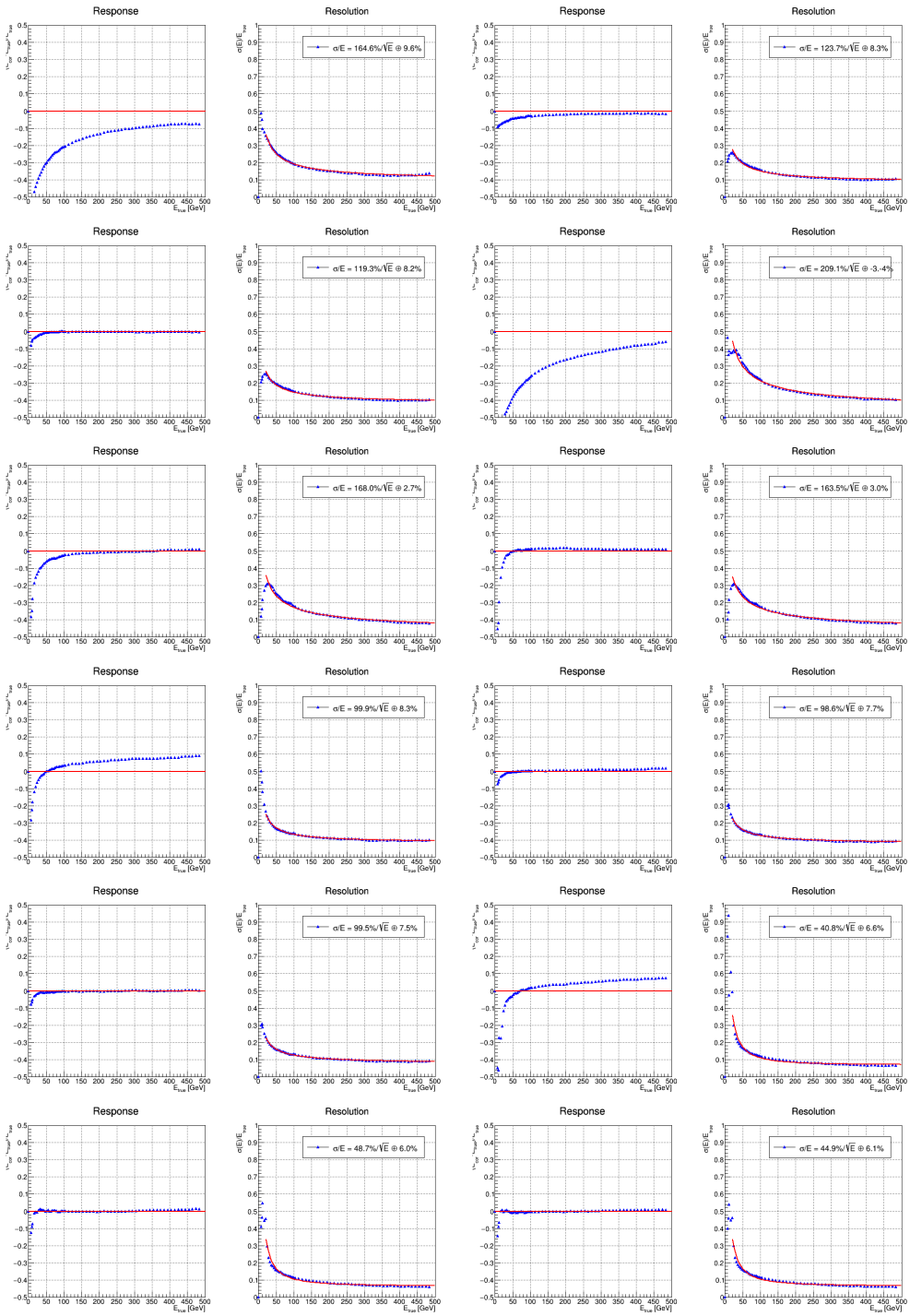


Figure B.12: Response and resolution of EH and H hadron in each barrel and endcap

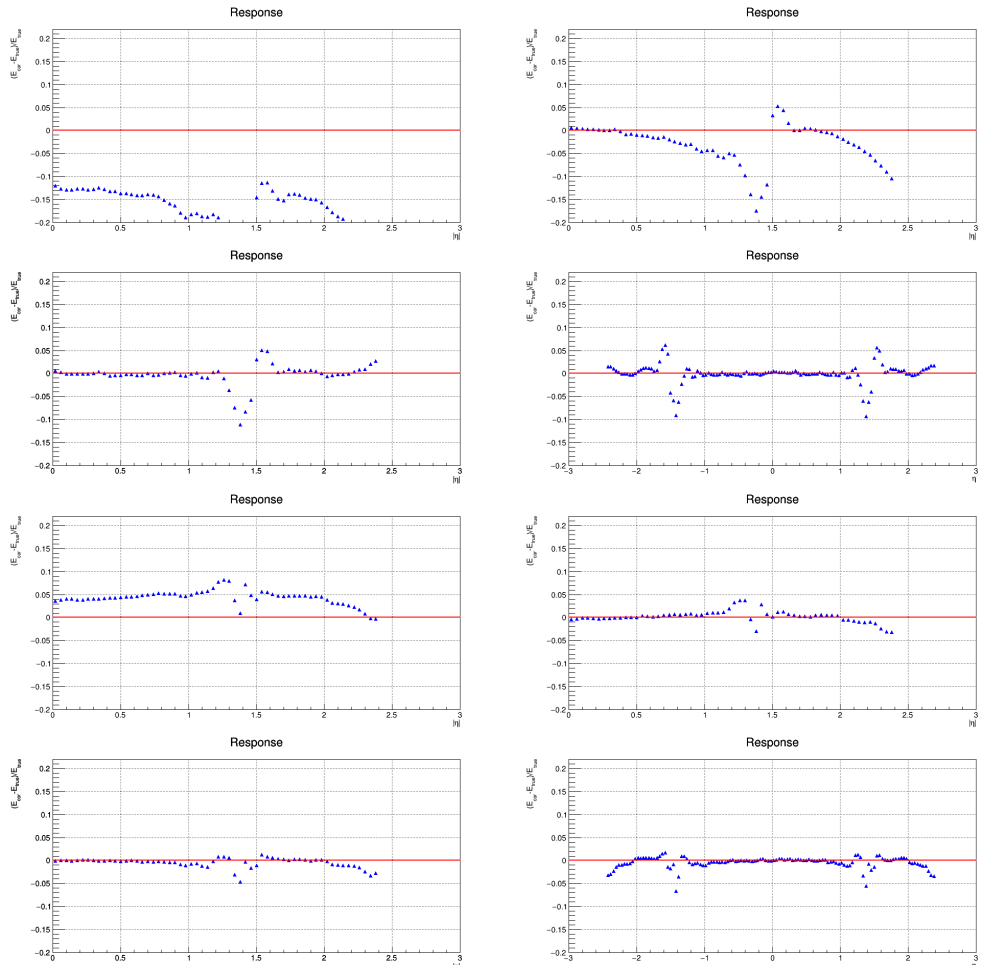


Figure B.13: Response of EH and H hadron in η

B.2 Jet Energy Correction at HLT

The purpose of Jet Energy Corrections (JECs) is to accurately reconstruct the energy of jets in the CMS detector. Jets are reconstructed from the particle flow algorithm, but the energy measurements from the ECAL and HCAL need to be corrected due to various factors. The composition of a jet is random, depending on the energy deposits of different particles, including electromagnetic (EM) components such as photons originating from decaying π^0 mesons and other signatures in the calorimeters. Neutrons inside a jet are chargeless, making measuring their energy difficult. As a result, jet energy measurements typically have worse resolution compared to electrons or photons.

The procedure for Jet Energy Corrections involves two steps: L1 pileup offset corrections and L2L3 response corrections. These corrections are applied at the Monte Carlo (MC) and the High-Level Trigger (HLT) levels.

The L1 pileup offset corrections aim to remove the pileup contribution within the jet cone. This is done by calculating the average difference in p_T between matched jets in simulated samples with and without pileup overlay. By subtracting the pileup contribution, the energy measurement is corrected for the effects of additional interactions in the same bunch crossing.

The L2L3 response corrections align the reconstructed jet p_T with the particle-level jet p_T. This is achieved by matching reconstructed (rec) and generated (gen) jets and inverting the median response, defined as $R = p_T^{\text{rec}}/p_T^{\text{gen}}$. The corrections are determined by comparing the response distributions of rec and gen jets and applying the necessary scaling factors. These corrections ensure that the reconstructed jets have a response consistent with the particle-level jets.

Residual corrections are then applied to account for differences between data and simulation. These corrections correct for any remaining discrepancies and ensure that the jet energy calibration is accurate in the data-taking environment.

For HLT JECs, two jet categories are used: AKPFJet and AKCaloJet, with jet cone sizes of 0.4 and 0.8, respectively. AKPFJet represents reconstructed jets using the Particle Flow (PF) algorithm, which combines information from all CMS subdetectors. AKCaloJet, on the other hand, is reconstructed using only calorimeter

information.

The offset value is defined as Offset = $p_T^{\text{rec}}(\text{PU}) - p_T^{\text{rec}}(\text{noPU})$, where PU refers to pileup and noPU represents the absence of pileup. The offset mainly accounts for neutral hadron contamination. The mean offset before the L1 correction is shown in Figure B.14. To correct for the pileup effect, a two-dimensional fit is performed, fitting ¡Offset/jet area¿ as a function of $< \rho >$ (pileup energy density) and $< p_T^{\text{rec}} >$ in each η bin. The fit function used is a "semi-simple" function:

$$f(\rho, p_T) = p_1(\rho - p_0)\left(1 + p_2 \ln\left(\frac{p_T}{15\,\text{GeV}}\right)\right) \tag{B.6}$$

After the L1 correction is applied, as shown in Figure B.15, the pileup effect becomes very small in all cases.

We calculate the median response, denoted as $R = p_T^{rec}/p_T^{ptcl}$, in fine bins of η^{rec} and p_T^{ptcl}. Similarly, we calculate the $< p_T^{rec} >$ in the same fine bins of η^{rec} and p_T^{ptcl}. To determine the L2L3 correction factors, we fit the inverse of the median response as a function of $< p_T^{rec} >$ using the "Standard+Gaussian" function:

$$f(p_T^{reco}) = p_0 + \frac{p_1}{\log_{10}^2(p_T^{reco}) + p_2} + p_3 e^{-p_4(\log_{10}(p_T^{reco}) - p_5)^2} + p_6 e^{-p_7(\log_{10}(p_T^{reco}) - p_8)^2} \tag{B.7}$$

The L2L3 correction factors are calculated after applying the L1 correction and are shown in Figure B.16.

The response results after applying the L1 and L2L3 corrections are shown in Figure B.17.

While Figure B.17 provides some simple validations, full validation is required for both jets and MET. MET validation is particularly important since MET is derived from jet information. Physics event Monte Carlo (MC) samples with VBFHToInvisible are used for MET validation.

The jet validation results for response and resolution are shown in Figure B.18 and Figure B.19. The response is close to 1 in all p_T and the detector regions examined.

Figure B.20 displays the MET validation results. MET validation is performed

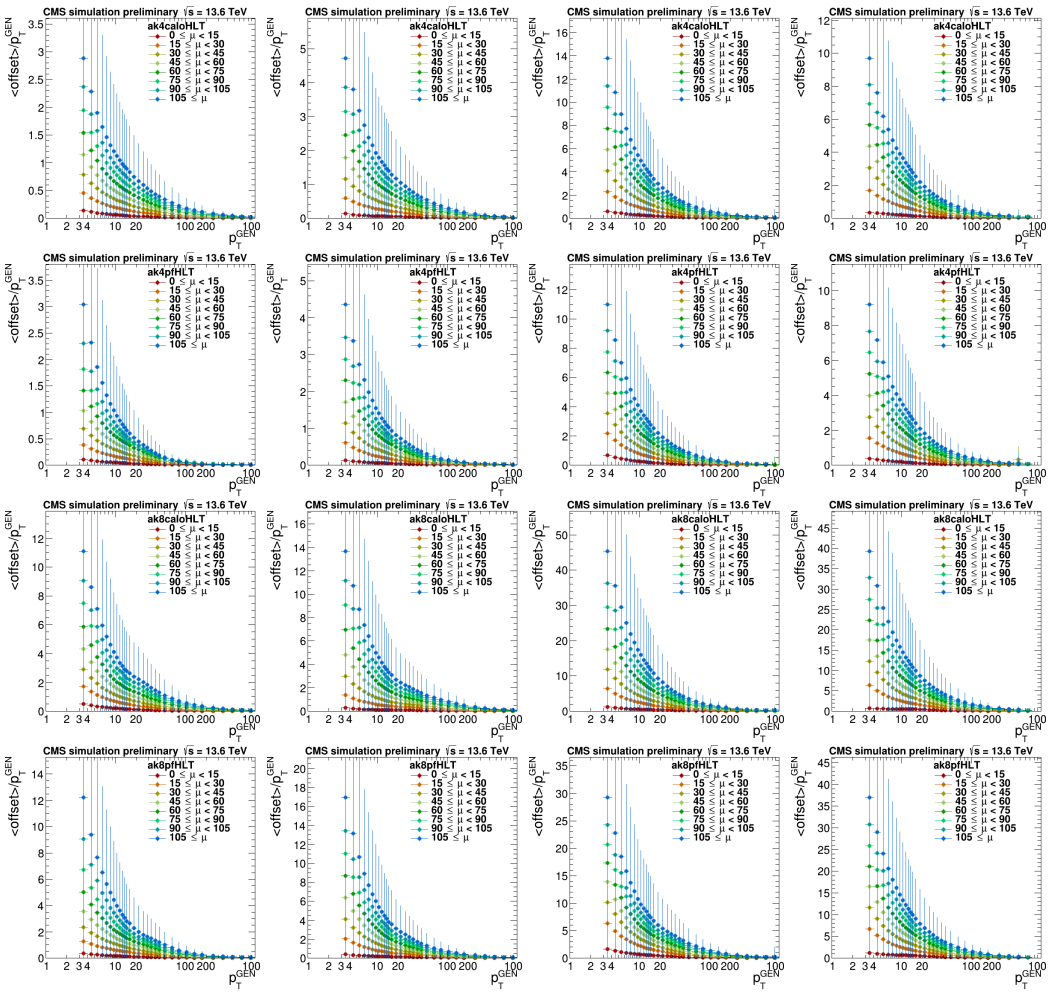


Figure B.14: Offset divided by p_T distribution vs. p_T of each jet category, cone size with detection region.

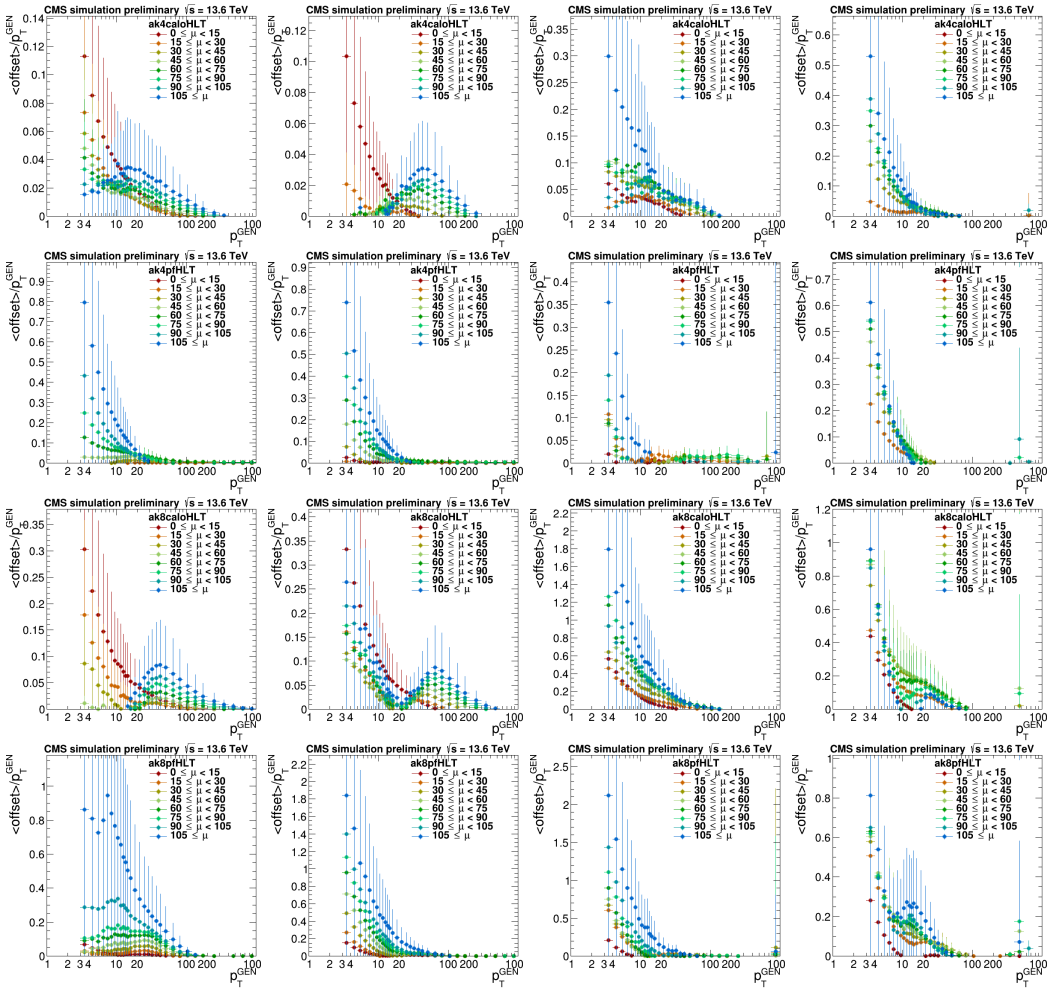


Figure B.15: Offset divided by p_T distribution vs. p_T of each jet category, cone size with detection region.

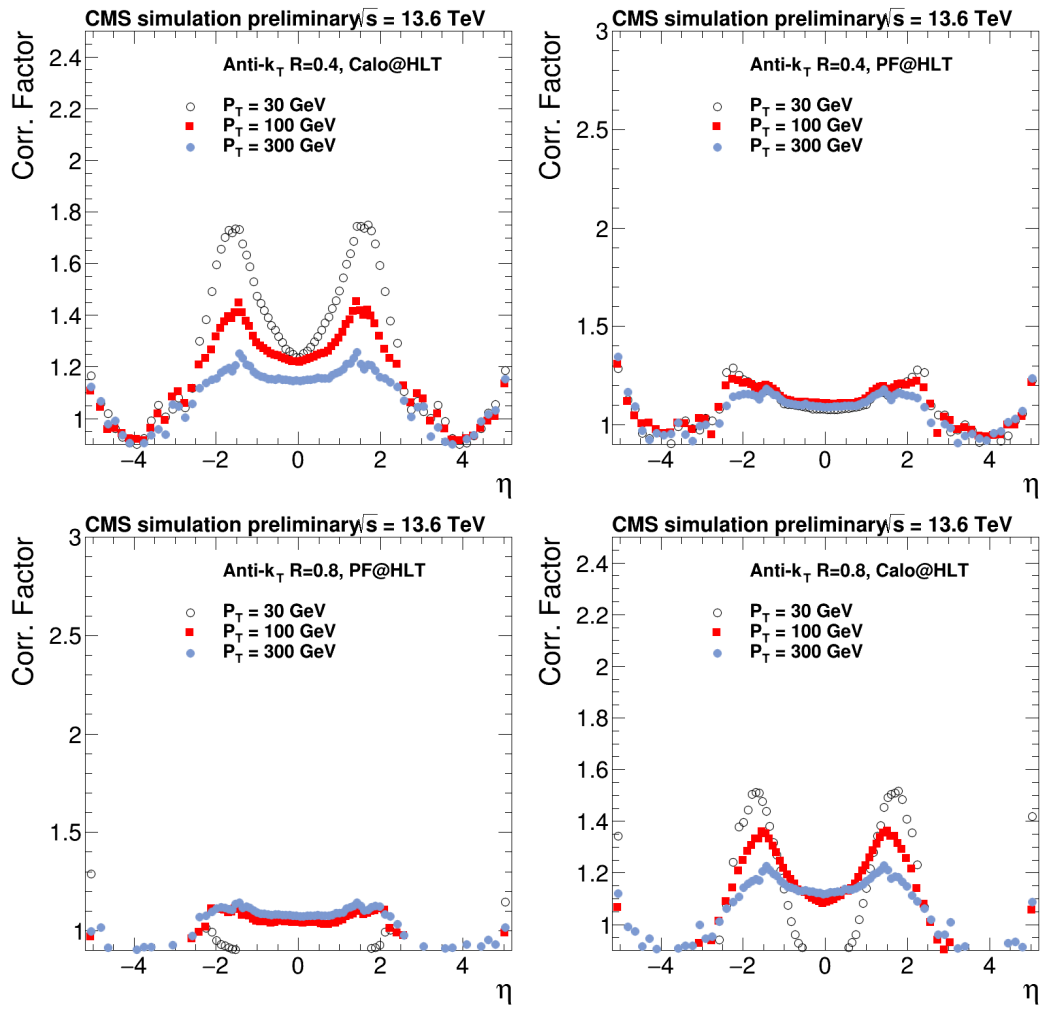


Figure B.16: L2L3 response correction scale factor vs. η in each jet category and cone size.

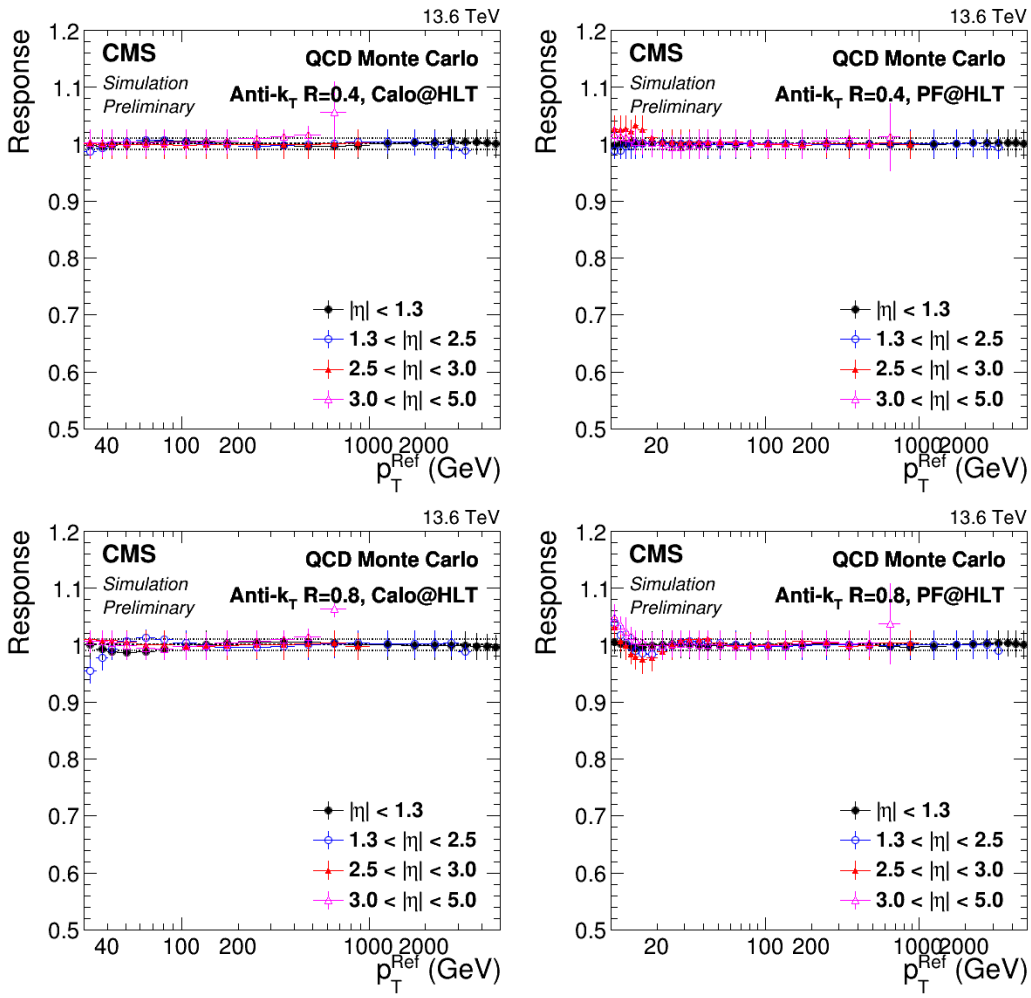


Figure B.17: response vs. η in each jet category and cone size.

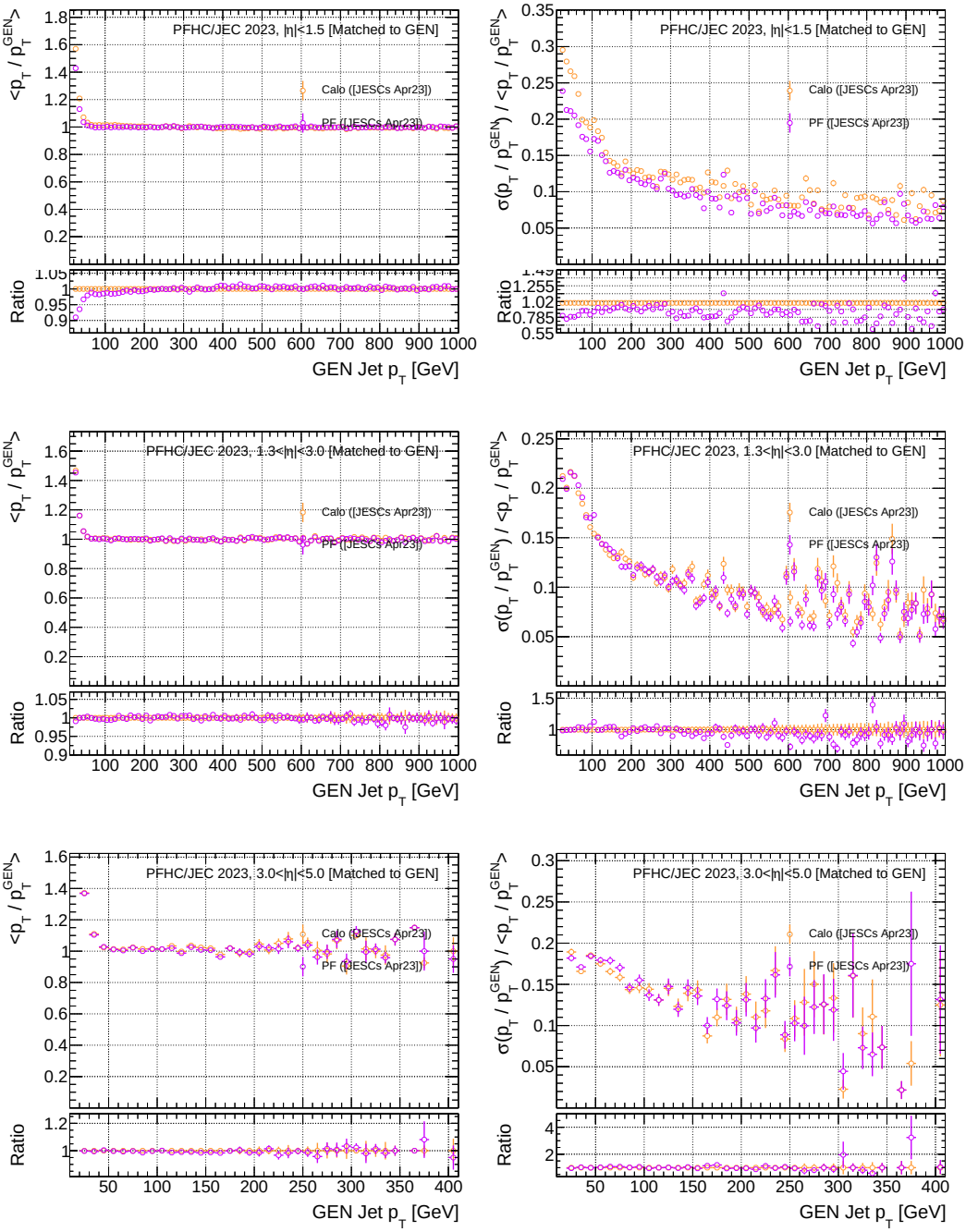


Figure B.18: Response and resolution vs. p_T for AK4 jets.

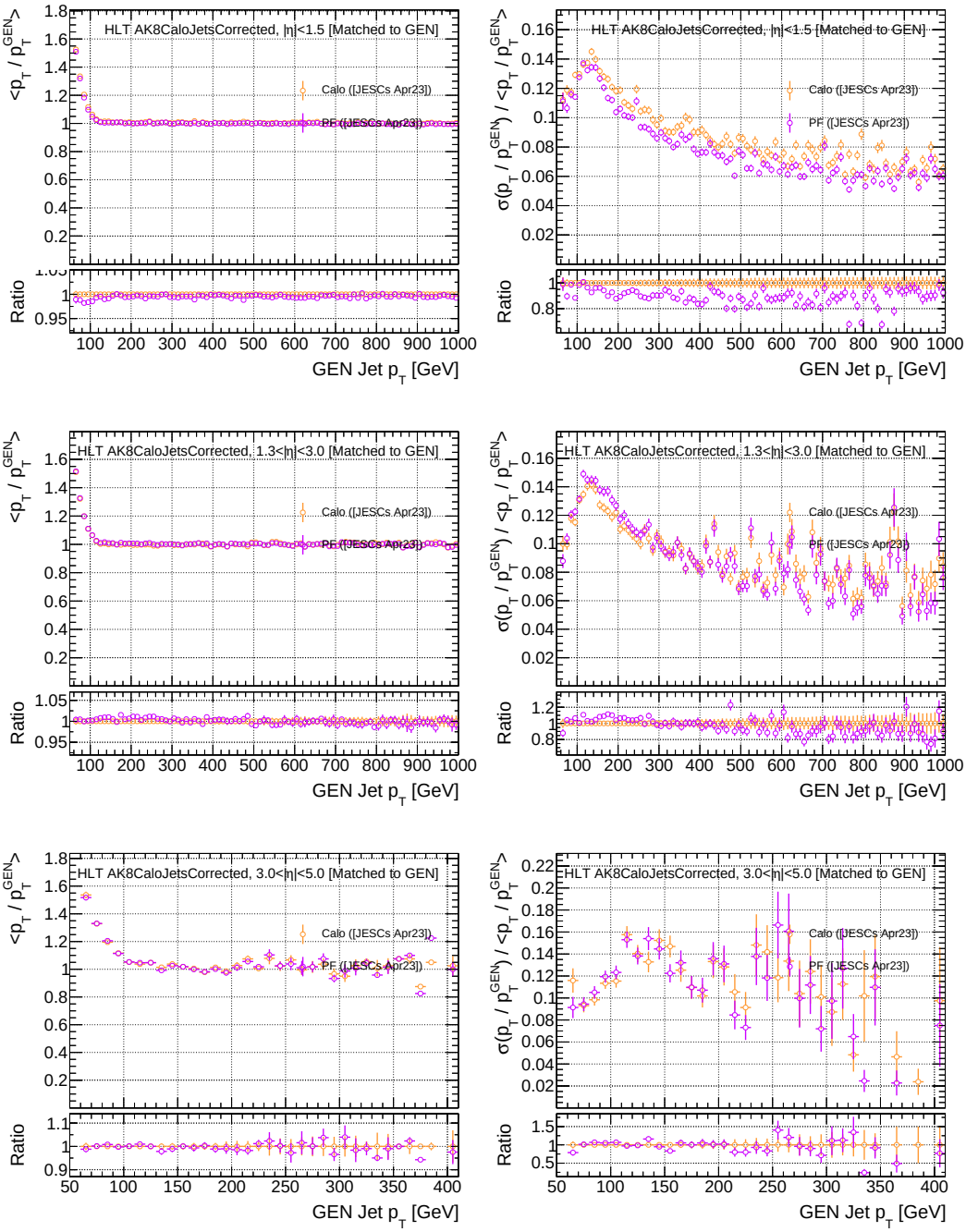


Figure B.19: Response and resolution vs. p_T for AK8 jets.

using physics simulation, and the MET response appears satisfactory.

These validation studies ensure the reliability and accuracy of the Jet Energy Corrections at the HLT level, providing confidence in the subsequent jet and MET measurements in the CMS experiment.

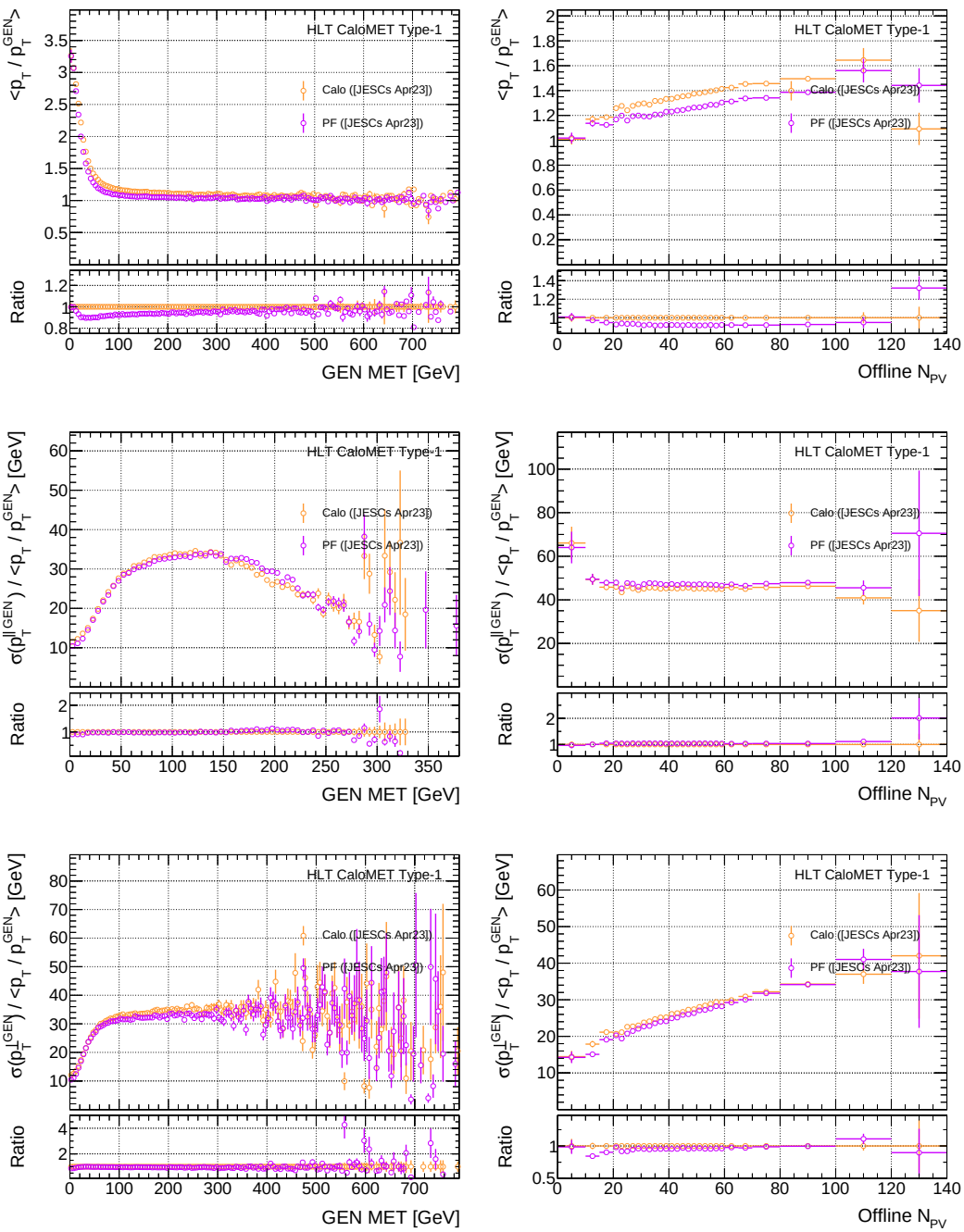


Figure B.20: The response and resolution vs. p_T for MET

B.3 Background study for Korea Experiments on Magnetic Monopole

Classical electromagnetism is a very beautiful theory and we seek its completion. A magnetic monopole (m), analogous and similar to the electric monopole (e), is a natural complement in Maxwell's equations of classical electromagnetism that would restore the complete symmetry of electric and magnetic fields [36, 37]. The simplicity and symmetry of Maxwell's equations with magnetic charge are seen directly in natural units ($\varepsilon_0 = \mu_0 = c = 1$) as:

$$\nabla \cdot \mathbf{E} = \rho_e \qquad -\nabla \times \mathbf{E} = \partial \mathbf{B}/\partial t + \mathbf{j_m}$$

$$\nabla \cdot \mathbf{B} = \rho_m \qquad \nabla \times \mathbf{B} = \partial \mathbf{E}/\partial t + \mathbf{j_e}$$

The electromagnetic force law for electric charge, e, and magnetic charge, g, at velocity $\mathbf{v}$ is

$$\mathbf{F} = e(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \; + \; g(\mathbf{B} - (\mathbf{v}/c^2) \times \mathbf{E}),$$

and this completes the description of classical electrodynamics. We are motivated by this symmetry to search in the mass region below the electron mass, and recognize that a charged particle at such low masses may radiate vigorously by bremsstrahlung. We have estimated that the effective radiation length of a magnetic charge (X_m) will scale relative to that of an electric charge as

$$X_m \approx \frac{(m/m_e)^2}{(g/e)^4} \, X_0,$$

so that a charged particle at very low mass will *hyper*-EM radiate and be unobservable in most particle physics experiments. This feature will dominate the design of our experiments.

Maxwell did not include magnetic charges due, apparently, to the absence of any evidence of radial magnetic fields. Pierre Curie [36] argued for the existence of "courants magnétiques" from symmetry and, in a footnote, describes an experiment he did to sense a magnetic current, commenting that it may not be "possible to ap-

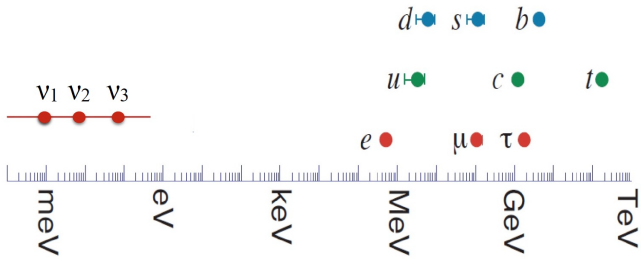


Figure B.21: Masses of known fermions from meV to 0.2 TeV [**?**].

preciate a very low magnetic conductivity." In 1931 Paul Dirac [38] discussed the quantum mechanics of a two-particle system, an electric monopole (e) and a magnetic monopole (g), concluding that quantization of angular momentum requires that these charges are related by

$$g_D = \frac{e}{2\alpha} \approx 68.5e,$$

in their lowest angular momentum state. Such a highly charged particle would be observed in nearly all particle physics experiments. Recent reviews of experiments and theoretical expectations are by K.Milton [39] and N.E.Mavromatos and V.A.Mitsou [40].

This Dirac monopole has been sought in cosmic rays [41–44], including in water and ice detectors [45–47] using Cenrenkov light [48], in materials exposed to fluxes of accelerator produced energetic particles [49], in geologic materials [50–52] including Moon rocks [53], in hadron collider experiments [54–56], in e^+e^- collider experiments [57–59], and in current LHC pp and ion-ion collider experiments [60–62]. There has been zero evidence of such a highly charged particle [63].

Two experiments have reported single Dirac monopoles, one at high altitude in a lexan plastic detector [64], and one in a superconducting loop [65] in a physics laboratory, the former with a simpler explanation as a heavy nucleus [66] and the latter without reproducible confirmation [67]. All of these searches have assumed a Dirac charged particle, $g_D = 68.5e$, have been sensitive to this charge or higher and, notably, restricted to high masses from GeV/c^2 to multi-TeV/c^2.

The masses of the elementary fermions are shown in Fig. B.21, in which the neutrino masses are estimated from the measured mass-squared differences. There are seven orders of magnitude between the neutrino and charged lepton masses where we will search. There are many speculations of unseen particles: supersymmetric particles, axions, skyrmions, dyons, dilatons, dark matter particles, etc., and more recently milli-charged particles derived from kinetic mixing ideas with a wide range of physical properties, and many experiments are underway to search for these particles [40].

We speculate that the symmetry of **E** and **B** in Maxwell's equations should be reflected in the properties and masses of the fundamental electric and magnetic particles. Therefore, we find it unreasonable that a magnetic monopole would have a huge mass, from $m > 10^3$ GeV/c^2 up to the GUT scale at 10^{16} GeV/c^2, that is physically incompatible with the electron mass. We suggest that magnetic charges may be *hiding in plain sight* with properties that would prevent their discovery by all previous and present experiments. Therefore, we are designing and developing new techniques and modifications and augmentations of current detectors for this experiment. This includes techniques to allow sensitivity to a low-mass Dirac charged magnetic particle.

We have chosen to search for magnetic charged (g) particles with masses less than the electron mass in e^+e^- annihilation at rest through a virtual photon which couples to a magnetic charge-anticharge pair m^+m^- is illustrated in Fig. B.22 with a coupling proportional to g^2,

$$\alpha_{m^+m^-} = \frac{g^2}{4\pi\varepsilon_0\hbar c}.$$

^{22}Na is a positron source that can be embedded in the annihilation target and which yields a low energy positron (e^+) at($E_{e^+} < 0.545$ MeV that precedes by 3.7 ps a photon (γ) at $E_\gamma = 1.275$ MeV. The positron annihilates on an atomic electron near rest in a metallic target within a few picoseconds [68] yielding two back-to-back 0.511-MeV γs. The 1.275-MeV γ is used as a pre-trigger or start-of-event signal. The dominant processes in annihilation at rest are $e^+e^- \to \gamma\gamma$ and $e^+e^- \to \gamma\gamma\gamma$.

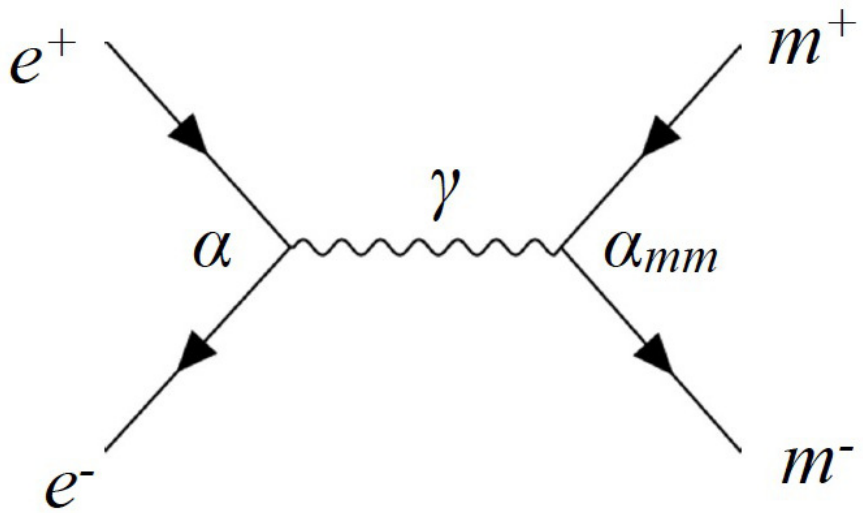


Figure B.22: Feynman diagram for $e^+e^- \to \gamma^* \to m^+m^-$.

Purported magnetic charges m^+ and m^- will be accelerated in opposite directions in a solenoidal magnetic field (B) and gain energy $\Delta E = gB\ell$ in the distance ℓ. The energy gain of a Dirac monopole, $g_D = 68.5e$, in a $B = 1$ T field over 1 meter (converting Tesla to V/m by the factor of c) is

$$\Delta E = 68.5e \times (1T \times c) \times 1 \text{ meter} \approx 20.5 \text{ GeV}.$$

This is a huge energy and two e^+e^- collider experiments [57, 58] have used this for their magnetic monopole searches. For a magnetic charge, g, in our uniform solenoidal field of $\int B_z dz \approx 1$T·m the energy gain is

$$\Delta E = (300 \text{ MeV}) \times g/e,$$

Therefore, we are sensitive down to $g/e \approx 10^{-2}$ where the total magnetic particle energy is ~ 6 MeV compared to the ^{22}Na positron source energies of 1.27 and 0.511 MeV.

This two-sided vacuum chamber (diameter ~ 20 cm and length ~ 1.2 m) inside the solenoids is shown in Fig. B.23 with flanges on both ends for mounting the

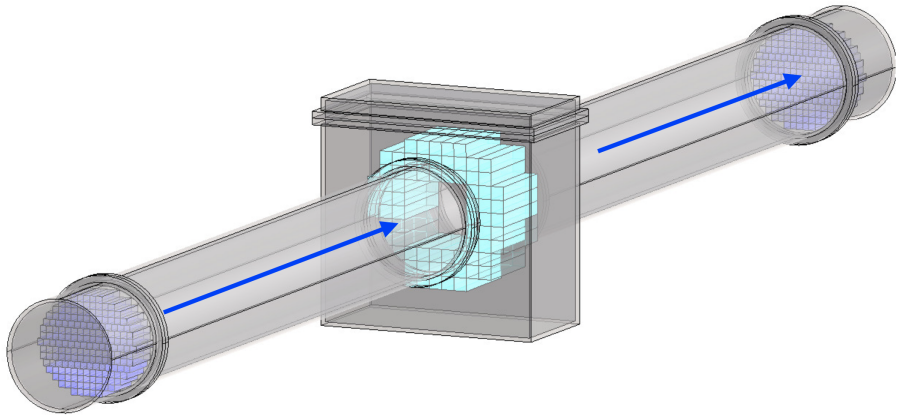


Figure B.23: Two-sided vacuum chamber showing EM crystal calorimeters at ends and central trigger-veto detectors in the middle, with solenoidal field shown as the blue arrows.

LYSO crystals of the EM calorimeters and a port in the middle for inserting the ^{22}Na source and annihilation target. Details of the crystals arrangements are shown in Fig. B.24. The locations of the trigger-veto detector and end-cap EM calorimeter crystals are noted in the photo on Fig. B.25. Near the center will be silicon-photomultipliers (SiPMs) photo-converters to record optical activity in the volume, including electromagnetic albedo from the EM crystals. We will use SiPM and DRS electronics to read out all EM crystals.

The solid angular acceptance of a 20-cm bore 1-meter from the source is about $d\Omega/\Omega \approx 0.0025$, so the $e^+e^- \to \gamma\gamma$ coincidence rate in the two EM calorimeters for a 10 μCi ^{22}Na source is about 100 Hz.

Hyper-EM radiation: A low-mass or low-charge magnetic particle will behave differently in a detector from, for example, an electron or muon. There are two considerations: (1) hyper-EM radiation from a low-mass charged particle (radiation rate goes like $1/m^2$), and (2) small momentum impulse to atomic electrons from a low-charge electromagnetic particle.

A charged particle with mass below the electron mass will radiate in the materials of the detector. The radiative energy loss depends strongly on the mass of

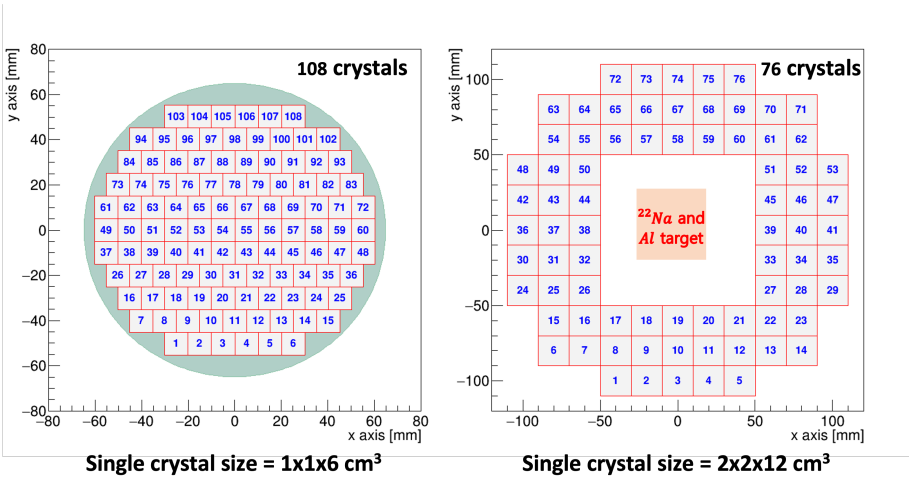


Figure B.24: Calorimeter crystals: trigger-veto detector(76 crystals) surrounding the positron annihilation target, and end-cap calorimeters (108 crystals) at the end of each acceleration channel.

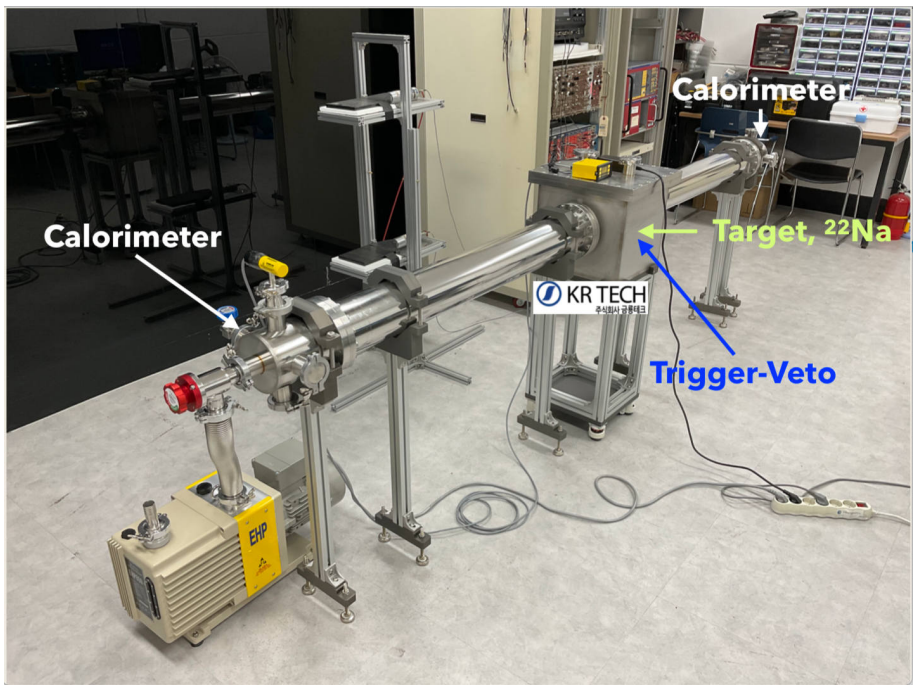


Figure B.25: Vacuum chamber with EM crystal calorimeters at ends and central trigger-veto detectors in the middle.

the charged particle and also on its charge. A simple semi-classical derivation for radiative energy loss leads to

$$\frac{dE_{\rm rad}}{dx} \approx \frac{4}{3\pi}\frac{\rho}{A}Z^2\alpha K\Big[\frac{(g/e)^2}{m/m_e}\Big]^2 \ln\Big(\frac{233\gamma(m/m_e)}{Z^{1/3}}\Big)\frac{E}{mc^2},$$

where E is the total energy of the magnetic charge, $\gamma = E/mc^2$, and $K = 0.3072$ MeV/g·cm^{-2}.

The effective radiation length of a monopole, X_m, in a material of $\rho Z/A$ is gotten from rearranging this expression as

$$\frac{dE_{\rm rad}}{E} = \frac{dx}{X_m},$$

so that

$$X_m = \Big[\frac{4}{3\pi}\frac{\rho}{A}Z^2\frac{\alpha K}{mc^2}\Big[\frac{(g/e)^2}{m/m_e}\Big]^2 \ln\Big(\frac{233\gamma(m/m_e)}{Z^{1/3}}\Big)\Big]^{-1}.$$

Comparing this to the radiation rate of an electric charge, the magnetic radiation length, X_m, scales relative to the $e^{\pm}, \gamma$ radiation length, X_0, as

$$X_m \propto \frac{(m/m_e)^2}{(g/e)^4}\, X_0.$$

This has consequences for detectors. The scale of an EM calorimeter is driven by its radiation length. For the electron (mass m_e and charge e) the radiation length of copper is about $X_0 \approx 1.4$ cm. For a magnetic particle of mass $m \approx 10^{-2} m_e$ and charge $g \approx e$, the radiation length is a few microns and a monopole may not escape the target or penetrate the wrapping of a detector. This radiation consists of photons from a dk/k spectrum with a normal electromagnetic radiation length of X_0 in the material, but the response of a particular detector configuration depends on Compton and photoelectric cross sections, coverings of detector volumes, and intervening materials. We have designed thin targets, low pressure gas chambers, and near-vacuum detector volumes to minimize the material traversed by a low-mass charged particle.

On the other hand, if $m \sim m_e$ and $g \ll e$, then the magnetic radiation length is very long and this magnetic particle becomes essentially undetectable in the present

$Lu_{2(1-x)}Y_{2x}SiO_5$	Value
Density, ρ(g/cm^3)	7.25
Yield (Photons/MeV)	33200
Decay time, τ(ns)	42
Refractive index, n	1.82
Radiation length, X_0(cm)	1.14

Table B.1: LYSO properties

experiment and all other experiments.

A particle with low magnetic charge $g < e$ will impart a small momentum impulse to atomic electrons and therefore lower ionization and lower excitation in scintillation materials. An estimate of the mean ionization rate for a relativistic magnetic particle of mass m and charge g is taken from Uehling [69, 70]. Making the replacement $e \to g\beta$ [71], the ionization energy loss by a low-mass magnetic charge is related to the ionization rate of the positron and is approximately given by

$$\frac{dE_m}{dx} \approx \Big(\frac{g\beta}{e}\Big)^2 \frac{dE_{e^+}}{dx} \approx \Big(\frac{g}{e}\Big)^2 K\frac{Z}{A}\Big[\ln\Big(\frac{\sqrt{2}mc^2\beta\gamma^{3/2}}{I^2}) - \beta^2)\Big].$$

This expression is mainly constrained by the magnetic charge squared (g^2) in the numerator, and the monopole mass (m) inside the logarithm.

The simulation was performed with GEANT4 (release 10.4.0) with FTFP-BERT physics list to estimate the physical background.
^{22}Na positron source and the 10 μm aluminum target are located in the center of the chamber, and a trigger-veto detector surrounds them. 1 m of vacuum cylinder is located, end-cap calorimeters are located in both ends as shown in B.26. Trigger-veto detector and end-cap calorimeters consist of LYSO crystals and SiPMs and LYSO crystal properties are shown in Table B.1.

Calibration We calibrate the particle's energy from the number of measured photoelectrons in the SiPM. The 662 keV gamma rays are used to calibrate detectors. We shoot the gamma rays to two different sizes of single crystals, used in trigger-

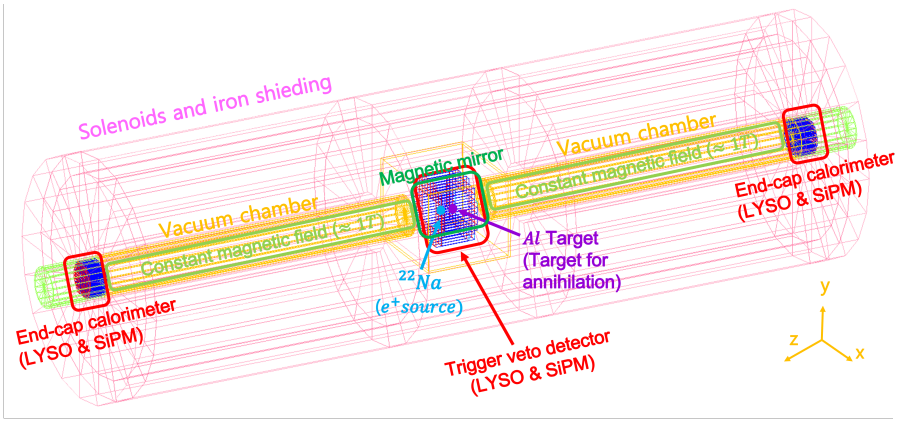


Figure B.26: Geometrical aspects of the simulations. They are the relevant feature of the KAEM experiments presented in FIG. B.23

veto detector and end-cap calorimeter. Then, measure the number of photoelectrons and fit the full peak with Gaussian distribution. The ratio from the mean of Gaussian distribution and energy of gamma rays is the calibration coefficient as shown in FIG. B.27. The calibration constant of end-cap calorimeter is 0.359, and the trigger-veto detector is 0.360.

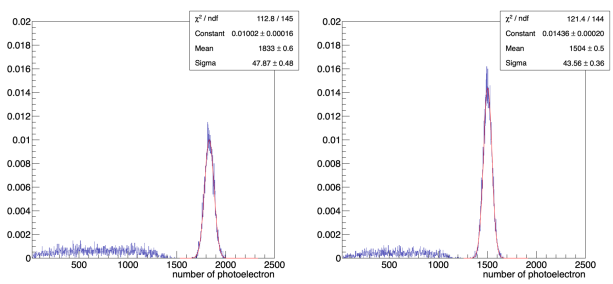


Figure B.27: Distribution of the number of photoelectrons in $2 \times 2 \times 12$ cm^3 and $1 \times 1 \times 6$ cm^3 single LYSO crystals.

Validation To validate the calibration coefficient, we first validate using trigger-veto detector and end-cap calorimeter using 662 keV gamma rays that the single crystal calibration constant is well performed in full detectors. Second, we validate using a full trigger-veto detector and end-cap calorimeter using 1 MeV gamma rays

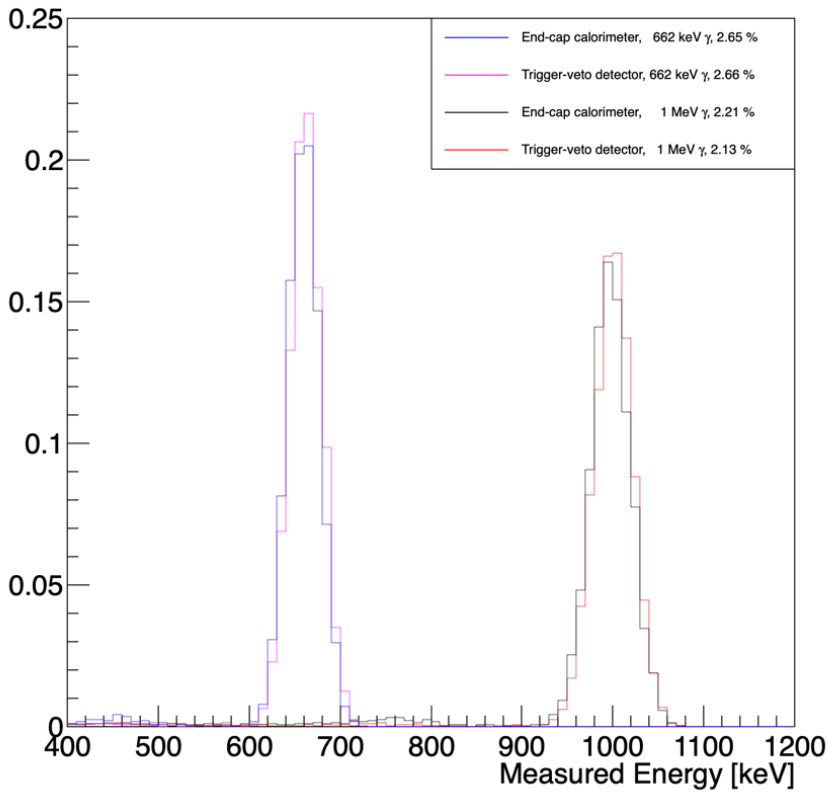


Figure B.28: Validation with 662 keV, 1 MeV gamma ray in trigger-veto detector and end-cap calorimeter

that the calibration constant is well performed at another energy. The validation result is shown in Fig. B.28. The resolution at 662 keV is 2.66%, 2.65%.

Backgrounds to a signal of a magnetic charge-anticharge pair can be separated into (*i*) ^{22}Na positron source related, (*ii*) intrinsic radioactivity from LYSO related, (*iii*) cosmic muons, both correlated and uncorrelated, (*iv*) cosmic electromagnetic particles, both correlated and uncorrelated, and (*v*) instrumental mismeasurements or miscalibrations.

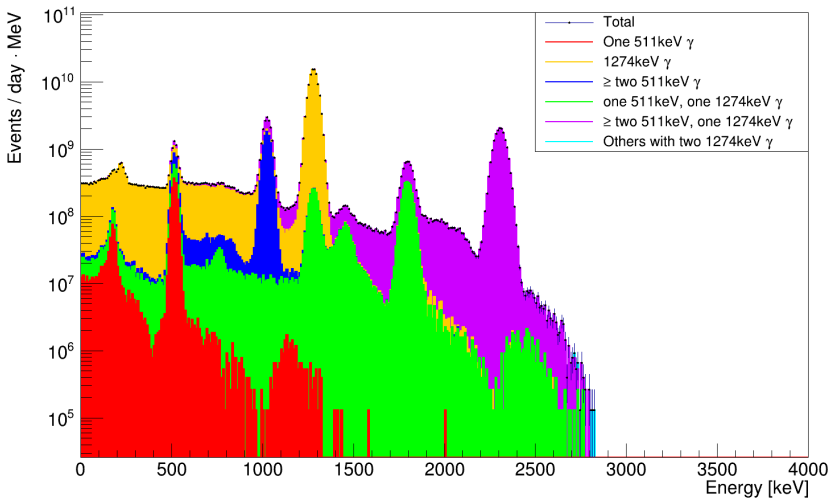


Figure B.29: Spectrum of trigger-veto detector from ^{22}Na positron source

Source related backgrounds The cascade decay

$$^{22}\mathrm{Na} \rightarrow {}^{22}\mathrm{Ne}^{*} + e^{+} \rightarrow {}^{22}\mathrm{Ne} + \gamma + e^{+},$$

is followed by positron annihilation

$$e^{+}e^{-} \rightarrow \gamma\gamma,$$

So the final state is three nearly simultaneous γs originating inside the thin positron target which is at the center of the trigger-veto detector. This distribution of trigger-veto detector and end-cap calorimeters energies for all ^{22}Na decays shows the collection of 0.511-MeV and 1.274-MeV γs in Fig. B.29 and B.30.

This distribution includes all processes taking place in the geometry of Fig. B.26: Compton backward scattering from the target material and from the inner walls of the vacuum chamber, grazing incidence of γ at the edges of the crystals, and overlaps of the 1.275-MeV γ with one or both of the 0.511-MeV γ.

Two simultaneous ^{22}Na decays within the acquisition time window, although infrequent, are shown in Fig. B.31.

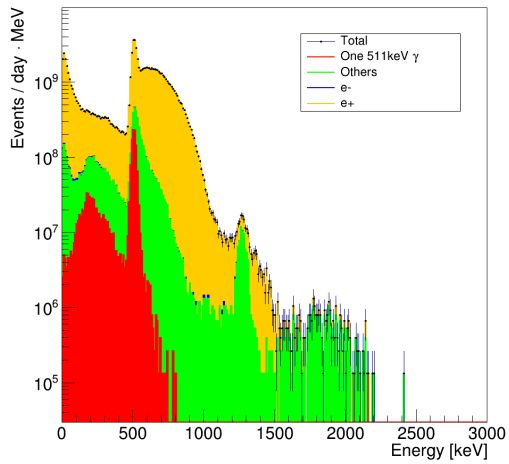


Figure B.30: Spectrum of end-cap calorimeter from ^{22}Na positron source

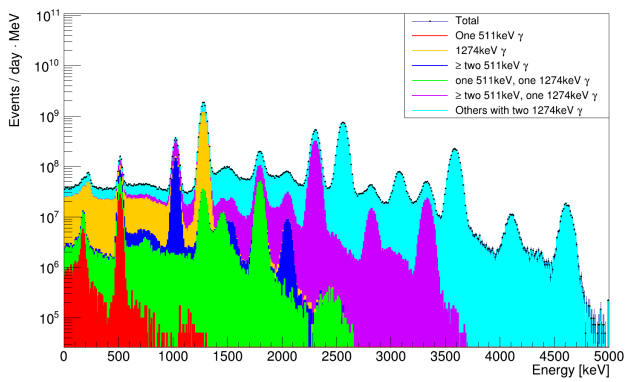


Figure B.31: The trigger-veto detector energy distribution for all near-simultaneous ^{22}Na decays.

Pattern recognition consists of clustering repeated Compton scatters in the crystal array. The requirement of one γ per calorimeter and energy deposition in the trigger-veto detector is consistent with a 1.274-MeV γ, reducing almost the source-related backgrounds. The pattern recognition, D_{xy}, is

$$D_{xy} = \sqrt{\frac{\sum d_i^2 E_i}{\sum E_i}}$$

where $d_i = \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2}$ is distance between crystal and hit position, x_i, y_i, and E_i is position and energy of trigger-veto detector and end-cap calorimeter. $x_0 = \sum x_i E_i / \sum E_i$, $y_0 = \sum y_i E_i / \sum E_i$, is position that calculate from center of mass method using the nearest crytals from hottest crystal in trigger-veto detector and end-cap calorimeter. D_{xy} distributions of trigger-veto detector and end-cap calorimeter with 1 or 2 γs are shown in FIG. B.32.

D_{xy} to determine a single particle is determined by the receiver operating characteristic curve(ROC curve) as shown in FIG. B.33. The single particle tagging efficiencies of D_{xy} at trigger-veto detector and end-cap calorimeter are 90% and 95% at 95% background rejection.

From the ROC curve, the pattern recognition value, D_{xy}, is 12 mm for end-cap calorimeter and 15 mm for trigger-veto detector. After applying pattern recognition result is shown in FIG. B.34.

intrinsic radioactivity from LYSO **related backgrounds** LYSO is a Lutetium-based scintillator that contains a naturally occurring radioactive isotope ^{176}Lu. The cascade decay

$$^{176}\mathrm{Lu} \rightarrow ^{176}\mathrm{Hf}^{*} + e^{-} \rightarrow ^{176}\mathrm{Hf} + \gamma + \gamma + \gamma + e^{-},$$

So the final state is three or fewer nearly simultaneous γs originating and leaving energies inside the calorimeters. The decay results in a 3 γs cascade of 306.78, 201.82, and 88.35 keV. The spectrum for all ^{176}Lu decays is shown in Fig. B.35.

Cosmic muon backgrounds Single cosmic $\mu^{\pm}$ deposit by ionization about $dE/dx \sim$ 1.3 MeV/g-cm^{-2}, which is about 9.6 MeV/cm for a density of ρ = 7.25 g/cm^3 for a

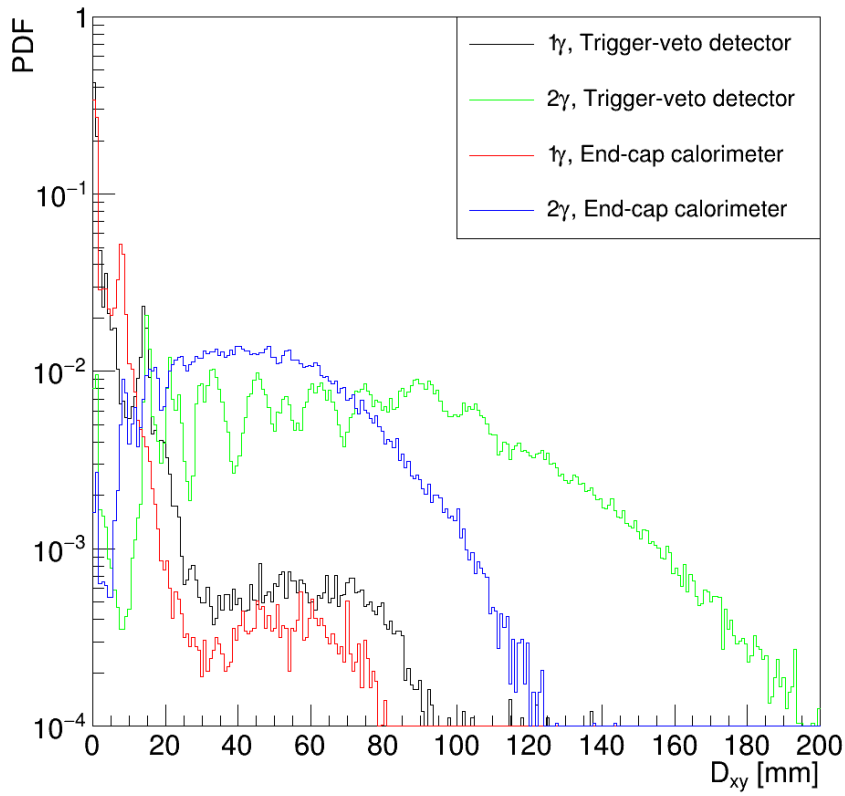


Figure B.32: Pattern recognition distribution in the trigger-veto detector and end-cap calorimeter.

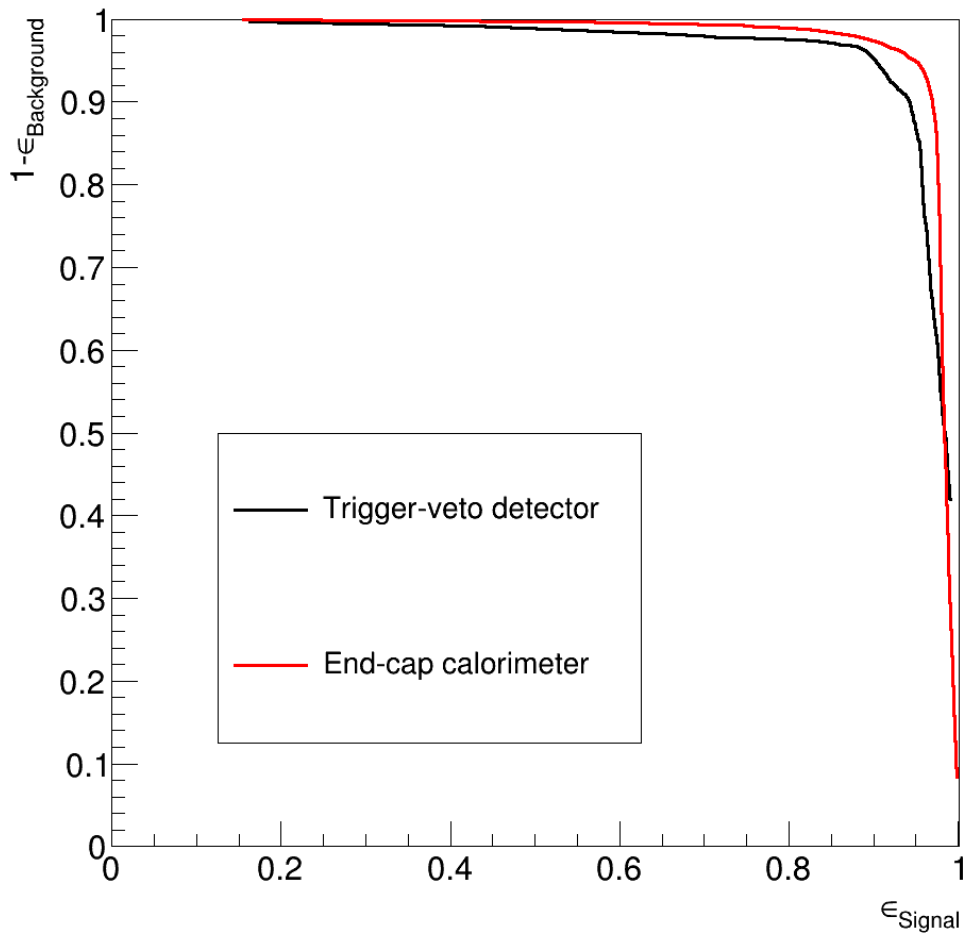


Figure B.33: Receiver operating characteristic curve in the trigger-veto detector and end-cap calorimeter.

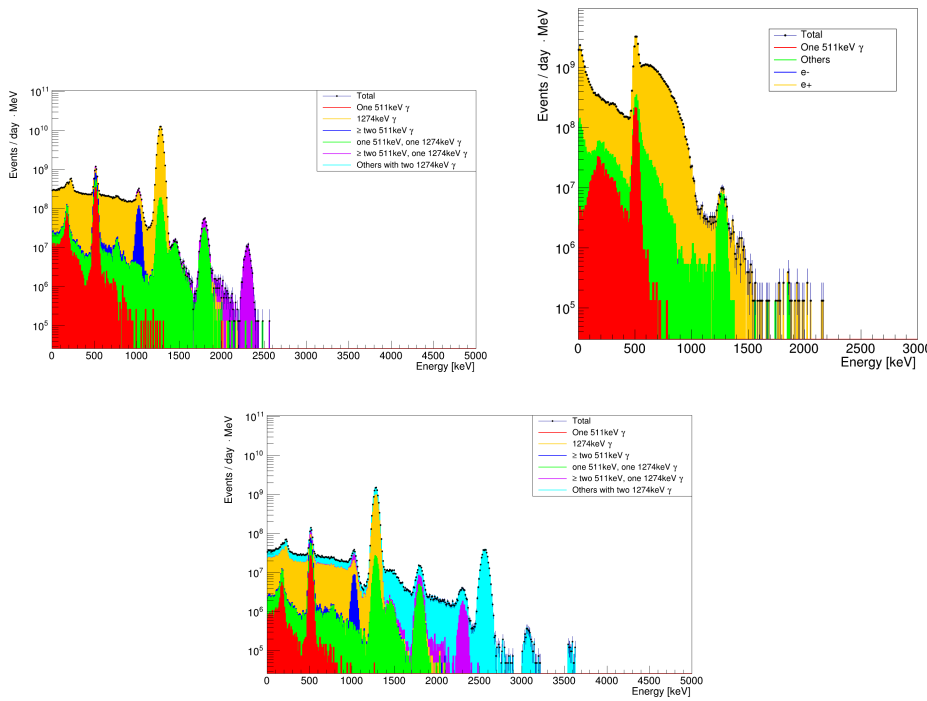


Figure B.34: Spectrum of trigger-veto detector(top) and end-cap calorimeter(middle) from ^{22}Na decay, and trigger-veto detector(bottom) from two simultaneous ^{22}Na decay after simple pattern recognition.

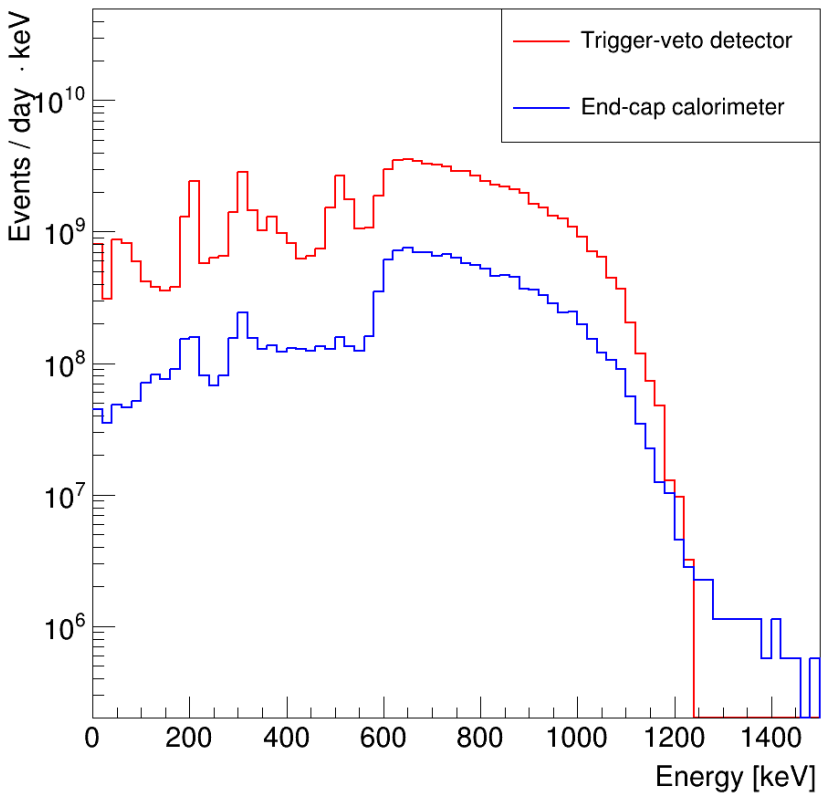


Figure B.35: The trigger-veto detector and end-cap calorimeters spectrum for ^{176}Lu decays

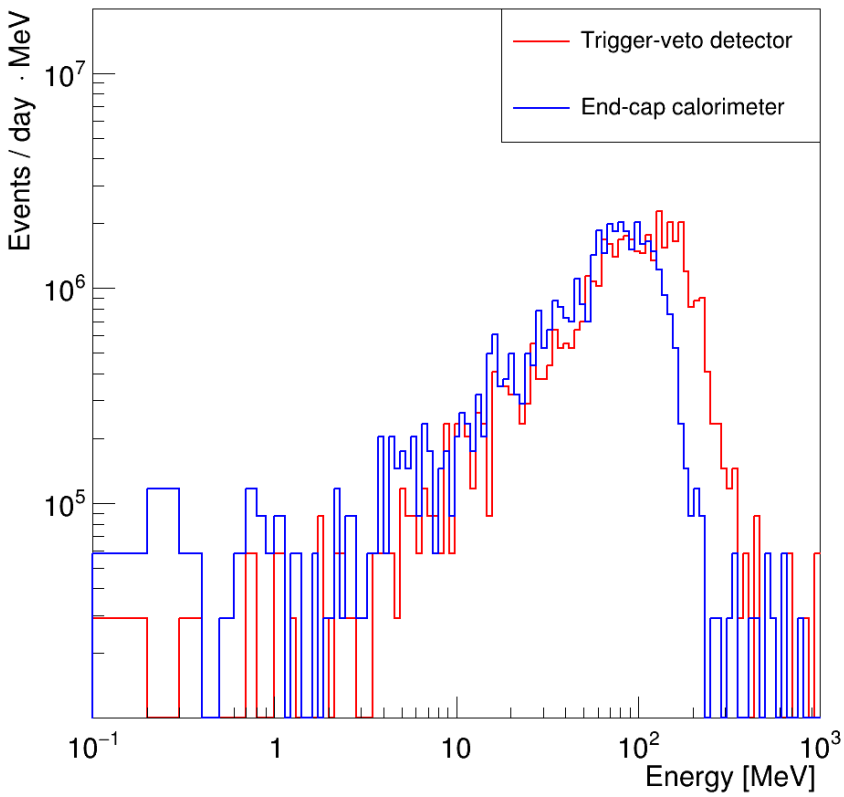


Figure B.36: The trigger-veto detector and end-cap calorimeters energy distribution for Cosmic muons.

LYSO crystal. The trigger-veto detector and end-cap calorimeter will have at least a few centimeters of muon path length in any direction, so a $\mu^{\pm}$ energy deposit will be far above the energy acceptance window of trigger-veto detector and will be the energy acceptance window of end-cap calorimeter. However, grazing passage through the edges of crystals will give a wide range of smaller signals.

Single cosmic muon spectra from the BESS experiments [72]. The distributions of trigger-veto detector and end-cap calorimeters for a single cosmic muon shows in Fig B.36. The deposited energy in the trigger-veto detector is far above the energy acceptance window. However, deposited energy in the end-cap calorimeter is overlapped with the magnetic monopole search region. Therefore, we set the muon

veto detector in the end-cap calorimeter to reject the cosmic muon background.

Cosmic electromagnetic backgrounds The rate of $e^{\pm}$ is approximately 1% of the $\mu^{\pm}$ rate, but any high energy EM particle that begins to shower will generate γs after shower maximum with energies near $E_{\gamma} \approx 1$ MeV. These γs have a long mean-free-path due to the minimum in the γN cross section just before the opening of the pair production channel. This is right near the energies of our signals and therefore is a serious background.

Cosmic EM particle spectra from the results of the real experiments [73]. The cosmic EM particle passage through the $6 \times 6 \times 1.25m^3$ concrete roof. The distributions of trigger-veto detector and end-cap calorimeters for a single cosmic EM particle are shown in Fig B.37. The total number of events from cosmic EM particle per month is 526.85 in the trigger-veto detector and 60.15 in the end-cap calorimeter. The number of events in the trigger energy range of the trigger-veto detector is 1.60 per year. The number of events in the magnetic monopole search region of the end-cap calorimeter is 5.33 per year.

Trigger and DAQ

The trigger for reading out all crystals is a clustered, isolated and identified energy deposit in the trig-veto that is consistent with a single 1.274 MeV photon. In addition, the time-coincidence of energy deposits in both end-cap calorimeters down to an energy just above the noise level is required. In addition, a possible geometrical requirement that the end-cap deposits be back-to-back in a final analysis.

The time resolution of the LYSO crystal calorimeters will allow a time-of-flight discrimination for masses near the electron mass, m_e, but only a time acceptance of a few nanoseconds relative to the trig-veto signal.

Waveform digitization All basic measurements are derived from waveform digitization of SiPM signals [74, 75] of the LYSO crystals comprising the trig-veto and end-cap calorimeters. Additional SiPMs monitor the single-photon light levels within the detector volume, including the optical albedo from any electromagnetic particle hitting the front face of the LYSO crystal crystals.

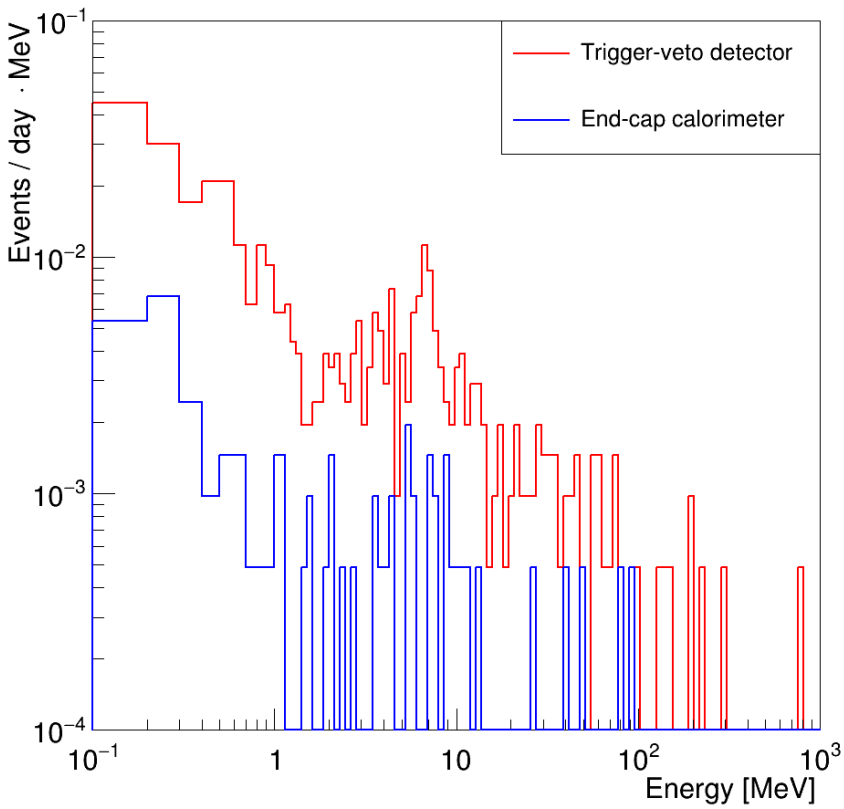


Figure B.37: The trigger-veto detector and end-cap calorimeters energy distribution for cosmic electromagnetic particles.

(a) Primitive	Description
t_0	Start-of-event time from barrel trigger-veto
t_1	Signal arrival time at $-z$ end-of-solenoid
t_2	Signal arrival time at $+z$ end-of-solenoid
E_0	Single-γ energy in barrel trigger-veto
E_1	Single-γ energy in $-z$ end-of-solenoid
E_2	Single-γ energy in $+z$ end-of-solenoid
$(\bar{x},\bar{y})_1$	Centroid in $-z$, $\phi_1 \sim \vec{r}_1/\ell$
$(\bar{x},\bar{y})_2$	Centroid in $+z$, $\phi_2 \sim \vec{r}_2/\ell$
$pe(t)$	Single photo-electron volume time history

(b) Object	Criterion	#σ	Condition for
ΔE_0	$= \lvert E_0 - 1.275\rvert < 0.08$ MeV	3σ	Start-of-event
ΔE_{12}	$= \lvert E_1 - E_2\rvert < 0.18$ MeV	3σ	$E_1 \approx E_2$
Δt_1	$= \lvert t_1 - t_0\rvert - \ell/c < 0.07$ ns	3σ	Near-c transit
Δt_2	$= \lvert t_2 - t_0\rvert - \ell/c < 0.07$ ns	3σ	Near-c transit
Δt_{12}	$= \lvert t_2 - t_1\rvert < 0.10$ ns	3σ	Equal time
$\Delta\phi_{12}$	$= \lvert\phi_1 - \phi_2\rvert < 0.014$ rad	3σ	Back-to-back

Table B.2: Trigger primitives (a) and trigger objects formed from them (b), showing trigger acceptance selections and the number of standard deviations for each selection.

Trigger primitives are shown in Table B.2(a) derived from the analyzed waveforms of the LYSO calorimeters, and the trigger objects and their selection criteria are shown in the Table B.2(b).

The study of magnetic monopole experiment simulation backgrounds with GEANT4 shows that backgrounds leave how much energy in calorimeters. Those provide to design rejecting backgrounds events from magnetic monopoles.

As a part of this work, simulation setup with GEANT4 (10.4.0) and FTFP-BERT physics list. We build LYSO crystal and SiPM to estimate particles' energy. Calibration and validation are processed, and it performed well in full calorimeters.

The four different types of physical backgrounds are considered. From the ^{22}Na simulation, we can precisely measure the trigger signal in the trigger-veto detector, and ^{22}Na positron sources do not leave energy in the magnetic monopole search region in end-cap calorimeters. From the pattern recognition, the efficiencies of

single and multiple particle discrimination from trigger-veto detector and end-cap calorimeter are 90% and 95% at 95% background rejection. The intrinsic radioactivity of LYSO does not affect both trigger signal and monopole search region. The cosmic muons and cosmic electromagnetic particles can leave energy in calorimeters up to 1 GeV, directly affecting the monopole search region. However, it can reject from the coincidence of another muon veto detector that surrounds the trigger-veto detector and end-cap calorimeters.

B.4 Cosmic Cloud API development

Cosmic Cloud is a cluster at the University of Maryland in the USA. This cloud's purpose of the store and analyze the cosmic ray data from ISS-CREAM or CREAM. Nowadays, most young developers move to the interpreter type of language, like Python. I designed and developed the CAI to access the datasets using Python to satisfy this request.

I choose MySQL connector to develop the CAI. The CAI has to support several functions. Log in to the database(DB), check user authority in DB, open and close DB and read, write and update data in the dataset. To access several datasets simultaneously, use Class to develop the CAI. When a Class is constructed, several pieces of information are inputted by the user and defined as private types to prohibit access to this information. The type of input and output is the list, and this type is easy to write and read variables.

I developed CAI based on the design. In addition, details like password hashing, split read function, and determined Python version are added. The purpose of password hashing is the protect the user's password from cracking. The read function is split into two categories. The first one is based on the fetch_all function in MySQL. It reads all rows of the dataset at once. Another uses the fetch_one function in MySQL. It read one row of data in the dataset. Python has two major versions, Python2 and Python3. Python2 officially does not support soon. So Python3 is determined.

The latest version of CAI that I developed is attached.

```
1 import numpy as np
2 import mysql.connector
3 import zlib
4 import bcrypt
5
6 class ccloud:
7   def __init__(
8     self, email=None, password=None, project="isscream",
    dataset=None, variablename=None, condition=None,
9   ):
```

```
10      self.__email = email
11      self.__password = password
12      self.__project = project
13      self.__dataset = dataset
14      self.__variablename = variablename
15      self.__condition = condition
16      self.__login_connection = mysql.connector.connect(
17                user='',
18                password='',
19                host='',
20                database='')
21      self.__login_cursor = self.__login_connection.cursor()
22
23      self.__open_connection = mysql.connector.connect(
24             user='',
25             password='',
26             host='',
27             database=self.__project)
28      self.__open_cursor = self.__open_connection.cursor()
29
30      self.__write_connection = mysql.connector.connect(
31             user='',
32             password='',
33             host='',
34             database=self.__project)
35      self.__write_cursor = self.__write_connection.cursor()
36
37      sql_query = self.__login_connection._cmysql.escape_string(
    """select userid, pwd from Userinfo where mail = %s""")
38      result = self.__login_cursor.execute(sql_query, (self.
    __email,))
39      record = self.__login_cursor.fetchone()
40      self.__name = record[0]
41      data = record[1]
42      if not (self.__name):
43        self.__login_connection.close()
44        self.__login_cursor.close()
45        self.__open_connection.close()
```

```
46      self.__open_cursor.close()
47      self.__write_connection.close()
48      self.__write_cursor.close()
49      print("wrong email address")
50    if not (bcrypt.checkpw(self.__password.encode('utf-8'),
    data.encode('utf-8'))):
51      self.__login_connection.close()
52      self.__login_cursor.close()
53      self.__open_connection.close()
54      self.__open_cursor.close()
55      self.__write_connection.close()
56      self.__write_cursor.close()
57      print("wrong password")
58
59    sql_query = 'select admin from Userinfo where mail="'+str(
    self.__email)+'"'
60    result = self.__login_cursor.execute(sql_query)
61    admin = self.__login_cursor.fetchone()
62    sql_query = 'select datasetid from DataSetInfo where name="
    '+self.__dataset+'"'
63    result = self.__login_cursor.execute(sql_query)
64    record = self.__login_cursor.fetchone()
65    if not record == None:
66      self.__dataview=0
67      sql_query = "select permit from DataAccess WHERE userid="
     + str(self.__name) + " AND datasetid=" + str(record[0])
68      result = self.__login_cursor.execute(sql_query)
69      record = self.__login_cursor.fetchone()
70    else:
71      self.__dataview=1
72      sql_query = 'select viewid from Viewinfo where name="'+
    self.__dataset+'"'
73      result = self.__login_cursor.execute(sql_query)
74      record = self.__login_cursor.fetchone()
75      sql_query = "select permit from ViewAccess WHERE userid="
     + str(self.__name) + " AND viewid=" + str(record[0])
76      result = self.__login_cursor.execute(sql_query)
77      record = self.__login_cursor.fetchone()
```

```
78      self.__permission=-1
79      self.__login_connection.close()
80      self.__login_cursor.close()
81      if(record[0] == "O" or admin[0] == "admin"):
82        self.__permission=2
83      elif(record[0] == "WR"):
84        self.__permission=1
85      elif(record[0] == "R"):
86        self.__permission=0
87      else:
88        self.__permission=-1
89        self.__open_connection.close()
90        self.__open_cursor.close()
91        self.__write_connection.close()
92        self.__write_cursor.close()
93        print("You don't have permission to access data")
94      if(self.__condition == None):
95        self.sql_query = "SELECT "+self.__variablename+" FROM "+
    self.__dataset
96        result = self.__open_cursor.execute(self.sql_query)
97      else:
98        self.sql_query = "SELECT "+self.__variablename+" FROM "+
    self.__dataset+" WHERE "+self.__condition
99        result = self.__open_cursor.execute(self.sql_query)
100
101   def __del__(self):
102     self.__open_connection.close()
103     self.__open_cursor.close()
104     self.__write_connection.close()
105     self.__write_cursor.close()
106
107   def Read(self):
108     record = self.__open_cursor.fetchone()
109     if record == None:
110       values = None
111       return values
112     values = list()
113     for output in record:
```

```
114       if(str(type(output)) == "<class 'bytes'>"):
115         temp = zlib.decompress(output)
116       else:
117         temp = output
118       values.append(temp)
119     return values
120 
121   def ReadAll(self):
122     sql_query = "SELECT "+self.__variablename+" FROM "+self.
    __dataset
123     result = self.__open_cursor.execute(sql_query)
124     i = 0
125     values = list()
126     for (data) in self.__open_cursor:
127       value_temp = list()
128       for output in data:
129         if(str(type(output)) == "<class 'bytes'>"):
130           temp = zlib.decompress(output)
131         else:
132           temp = output
133         value_temp.append(temp)
134       values.append(value_temp)
135     return values
136 
137   def Insert(self, values):
138     if(self.__permission>=1 and self.__dataview==0):
139       name = self.__variablename.split(',')
140       if(range(len(name)) == range(len(values))):
141         Input = ""
142         for temp in values:
143           if(str(type(temp)) == "<class 'bytes'>"):
144             print("NUMPY ARRAY DOES NOT SUPPORT YET")
145             return False
146           elif(str(np.shape(temp)) == "()"):
147             if(str(type(temp)) == "<class 'str'>"):
148               Input += '"'
149             Input += str(temp)
150             if(str(type(temp)) == "<class 'str'>"):
```

```
151                 Input += '"'
152             Input += ', '
153         Input = Input[:-2]
154 
155         sql_query = """INSERT INTO %s (%s) VALUES (%s)""" %(
    self.__dataset, self.__variablename, Input)
156         print(sql_query)
157         self.__write_cursor.execute(sql_query)
158         self.__write_connection.commit()
159       else:
160         print("Number of values is different of number of
    variables")
161     else:
162       print("You don't have permission to write dataset")
163 
164 
165   def Update(self, values, variables=None, condition=None,):
166     if(self.__permission>=1 and self.__dataview==0):
167       name = self.__variablename.split(',')
168       if not (varialbes == None):
169         Input = ""
170         for(temp1, temp2) in zip(variables, values):
171           Input += temp1
172           Input += ' = "'
173           if(str(np.shape(temp2)) == "()"):
174             Input += str(temp2)
175           else:
176             print("NUMPY ARRAY DOES NOT SUPPORT YET")
177             return False
178           Input += '", '
179         Input = Input[:-2]
180 
181         sql_query = "UPDATE "+self.__dataset+" SET " + Input +
    " WHERE " +condition
182         self.__write_cursor.execute(sql_query)
183         self.__write_connection.commit()
184       elif(range(len(name)) == range(len(values))):
185         Input = ""
```

```
186         for(temp1, temp2) in zip(name, values):
187           Input += temp1
188           Input += ' = "'
189           if(str(np.shape(temp2)) == "()"):
190             Input += str(temp2)
191           else:
192             print("NUMPY ARRAY DOES NOT SUPPORT YET")
193             return False
194           Input += '", '
195         Input = Input[:-2]
196 
197         sql_query = "UPDATE "+self.__dataset+" SET " + Input +
    " WHERE " +condition
198         self.__write_cursor.execute(sql_query)
199         self.__write_connection.commit()
200       else:
201         print("Number of values is different of number of
    variables")
202     else:
203       print("You don't have permission to write dataset")
204 
205   def Delete(self, condition=None,):
206     if(self.__permission>=2):
207       sql_query = "DELETE FROM " + self.__dataset + " WHERE " +
    condition
208       self.__write_cursor.execute(sql_query)
209       self.__write_connection.commit()
210     else:
211       print("You don't have permission to remove dataset")
```

The user and developer manuals are written as released. The user manual contains the basic concepts and how to use the CAI, and the developer manual includes detailed information about the CAI.

Bibliography

[1] S. P. Martin and A. Supersymmetry primer. Advanced series on directions in high energy physics. 18:1, 1998.

[2] Wikipedia. Standard Model — Wikipedia, the free encyclopedia. `http://en.wikipedia.org/w/index.php?title=Standard%20Model&oldid=1179748656`, 2023. [Online; accessed 13-November-2023].

[3] Steven Weinberg. Notes for a History of Quantum Field Theory, The Search for Unity, 1997.

[4] Michael E. Peskin and Daniel V. Schroeder. An Introduction to quantum field theory, 1995.

[5] L. Alvarez-Gaume and J. Ellis. Eyes on a prize particle", nature phys. 7 (2011). *no*, 1:2–3.

[6] N. Aghanim et al. Planck 2018 results. VI. Cosmological parameters. *Astron. Astrophys.*, 641:A6, 2020.

[7] P. Salucci. E. Corbelli. The extended rotation curve and the dark matter halo of M33. *Monthly Notices of the Royal Astronomical Society*, 311:441, 2000.

[8] A. G. Riess al. and Astrophysical Journal. 876. 85, 2019.

[9] J. Ellis et al. Supersymmetric relics from the big bang. *Nuclear Physics B*, 238(2):453–476, 1984.

[10] Wikipedia. Large Hadron Collider — Wikipedia, the free encyclopedia. `http://en.wikipedia.org/w/index.php?title=Large%20Hadron%20Collider&oldid=1183519026`, 2023. [Online; accessed 13-November-2023].

[11] O. S. Brning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock. *LHC Design Report*. Monographs, CERN Yellow Reports, 2004.

[12] S. Chatrchyan al. Commissioning of the cms experiment and the cosmic run at four tesla. *JINST*, 5, 2010.

[13] S. Chatrchyan al. Precise mapping of the magnetic field in the cms barrel yoke using cosmic rays. *JINST*, 5, 2010.

[14] L.Viliani et al. Cms tracker performance and readiness for lhc run ii, 2016.

[15] M. D'Alfonso al. Validation tests of the cms tib/tid structures. 2009.

[16] A. Benaglia al. The cms ecal performance with examples. *JINST*, 9, 2014.

[17] A. M. Sirunyan al. Calibration of the cms hadron calorimeters using proton-proton collision data at $\sqrt{s}$ = 13 tev. *JINST*, 15, 2020.

[18] S. Chatrchyan al. Performance of cms muon reconstruction in pp collision events at $\sqrt{s}$ = 7 tev. *JINST*, 7, 2012.

[19] Albert M Sirunyan et al. Inclusive search for supersymmetry in pp collisions at $\sqrt{s}$ = 13 TeV using razor variables and boosted object identification in zero and one lepton final states. *JHEP*, 03:031, 2019.

[20] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_{T} jet clustering algorithm. *JHEP*, 04:063, 2008.

[21] CMS JetMET POG. Recommended jet energy corrections and uncertainties for data and mc. 2018. https://twiki.cern.ch/twiki/bin/view/CMS/JECDataMC#2017_Data.

[22] CMS B-tag POG. b-tag pg twiki. https://twiki.cern.ch/twiki/bin/view/CMS/BtagPOG, 2015.

[23] CMS B-tag POG. Methods to apply b-tagging efficiency scale factors. https://twiki.cern.ch/twiki/bin/viewauth/CMS/BTagSFMethods, 2015.

[24] CMS Muon POG. Reference muon id, isolation and trigger efficiencies for 2016 legacy re-reco data. https://twiki.cern.ch/twiki/bin/view/CMS/MuonReferenceEffs2016LegacyRereco, 2018.

[25] CMS Muon POG. Reference muon id, isolation and trigger efficiencies for 2017 data. https://twiki.cern.ch/twiki/bin/view/CMS/MuonReferenceEffs2017, 2018.

[26] CMS JetMET POG. Particlenet sfs recommendations for run 2. 2016. https://twiki.cern.ch/twiki/bin/view/CMS/ParticleNetSFs#Top_Tagger_Nominal.

[27] CMS JetMET POG. Met filter recommendations for run 2. 2016. https://twiki.cern.ch/twiki/bin/viewauth/CMS/MissingETOptionalFiltersRun2.

[28] CMS Combine Tool. Likelihood for a counting experiment. https://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/part5/roofit/#a-likelihood-for-a-counting-experiment.

[29] CMS JetMET POG. Utilities for accessing pileup information in data. https://twiki.cern.ch/twiki/bin/viewauth/CMS/L1PrefiringWeightRecipe.

[30] CMS JetMET POG. Reweighting recipe to emulate l1 prefiring. https://twiki.cern.ch/twiki/bin/viewauth/CMS/L1PrefiringWeightRecipe.

[31] CMS SUS PAG. Isr systematics for fastsim. https://twiki.cern.ch/twiki/bin/view/CMS/SUSRecommendationsRun2Legacy#ISR_systematics_for_Signal_Monte.

[32] C. Lester and D. Summers. Measuring masses of semiinvisibly decaying particles pair produced at hadron colliders. *Phys.Lett. B*, 463:99–103, 1999.

[33] Y. Bai, H.-C. Cheng, J. Gallicchio, and J. Gu. Stop the top background of the stop search. *JHEP*, 1207:110, 2012.

[34] C. Rogan. *Kinematical variables towards new dynamics at the LHC*. [hep-ph].

[35] Cms Collaboration and V. Khachatryan al. Inclusive search for supersymmetry using razor variables in pp collisions at sqrt(s) = 13 tev. *Phys. Rev. D*, 95, 2017.

[36] P. Curie. Sur la possibilité d'existence de la conductibilité magnétique et du magnétisme libre. *Séances de la Société française de physique, Paris pp*, pages 76–77, 1894.

[37] Arttu Rajantie. Introduction to magnetic monopoles. *Contemporary Physics*, 53:3:195–211, 2012.

[38] P. A. M. Dirac. *Quantized Singularities in the Electromagnetic field*. Proc. Roy. Soc. London **A 133**, 60-72, 1931.

[39] K. A. Milton. Theoretical and experimental status of magnetic monopoles. *Rep. Prog. Phys*, 69:1637–1712, 2006.

[40] N. E. Mavromatos and V. A. Mitsou. Magnetic monopoles revisited: Models and searches at colliders and in the cosmos. *Int. J. of Modern Physics*, A35, August 2020.

[41] M. Ambrosio. Macro detector at gran sasso. *Nucl. Instrs. Meths.*, A 486, 2002.

[42] Pierre Auger Collaboration and A.et al. Aab. Search for ultra-relativistic magnetic monopoles with the pierre auger observatory. *Phys. Rev*, D 94, 2016.

[43] D. P. Hogan. et al., "relativistic magnetic monopole flux constraints from rice,". *Phys. Rev*, D 78, 2008.

[44] Anita ii Collaboration and M. Detrixhe. et al., search for ultra-relativistic magnetic monopoles with the pierre auger observatory," phys. *Rev. D*, 83, 2011.

[45] V. Aynutdinov. et al., "the BAIKAL neutrino experiment: Physics results and perspectives,". *Nucl. Instrs. Meth*, A 602:14–20, 2009.

[46] IceCube Collaboration and M. G. Aartsen. et al. *Eur. Phys. J*, C 74, 2019.

[47] Antares Collaboration and A. Albert. et al.,"search for relativistic magnetic monopoles with five years of the antares detector data,". *JHEP*, 07, 2017.

[48] W. Peter Trower. Magnetic monopoles and c light. *Nucl. Instr. Meths*, 248:130–135, 1986.

[49] G. R. Kalbfleisch. et al. *"Limits on production of magnetic monopoles utilizing samples from the D0 and CDF detectors at the Fermilab Tevatron,"*, D69, 2004.

[50] H. Jeon and M. J. Longo. Search for magnetic monopoles trapped in matter. *Phys. Rev. Lett*, 75, 1996.

[51] T. Ebisu and T. Watanabe. Search for magnetic monopoles trapped in old iron ores using a superconducting detector. *Phys. Rev*, D36:3359–3366, 1987.

[52] J. M. Kovalik and J. L. Kirschvink. New superconducting-quantum-interference-device-based constraints on the abundance of magnetic monopoles trapped in matter: An investigation of deeply buried rocks. *Phys. Rev*, A 33:1183, 1986.

[53] R. R. Ross. et al. *"Search for Magnetic Monopoles in Lunar Material Using an Electromagnetic Detector,"*, D8, 1973.

[54] H. Hoffmann. et al., "a new search for magnetic monopoles at the cern-isr with plastic detectors,". *Lett. Nuovo Cim*, 23:357, 1978.

[55] B. Aubert, P. Musset, M. Price, and J. P. Vialle. Search for magnetic monopoles in proton-antiproton interactions at 540 gev cm energy. *Phys. Lett*, 120B:465, 1983.

[56] Cdf Collaboration and A. Abulencia. et al., "direct search for dirac magnetic monopoles in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ tev,". *Phys. Rev. Lett*, 96, 2006.

[57] Cleo Collaboration and T. Gentile. et al. *"Search for Magnetically Charged Particles Produced in e^+e^- Annihilations at $\sqrt{s} = 10.6$ GeV*, D35, 1987.

[58] Tasso Collaboration and W. Braunschweig. et al. *"A search for particles with magnetic charge produced in e^+e^- annihilations at $\sqrt{s} = 35$ GeV*, C 38:543–550, 1988.

[59] G. Abbiendi. et al., "search for dirac magnetic monopoles in e^+e^- collisions with the opal detector at lep2,". *Phys. Lett*, B 663:2007–019, 2008.

[60] et al. B. Acharya. The physics programme of the moedal experiment at the lhc. *International Journal of Modern Physics A*, 29(23):1430050, 2014.

[61] et al. Acharya, B. Search for magnetic monopoles produced via the schwinger mechanism. *Nature*, 602, August 2022.

[62] Ana Lopes. Atlas homes in on magnetic monopoles, " `home.cern`, 14 june 2019.

[63] D. Milstead and E. J. Weinberg. Magnetic monopole searches. *(August*, 2015, 2016.

[64] P. B. Price al. Evidence for detection of a moving magnetic monopole. *Phys. Rev. Lett*, 35:487, 1975.

[65] B. Cabrera. First results from a superconductive detector for moving magnetic monopoles. *Phys. Rev. Lett*, 48:1378, 1982.

[66] L. W. Alvarez. Analysis of a reported magnetic monopole. In reprinted in Goldhaber, A. S., and W. P. Trower, editors, *Magnetic Monopoles*. American Association of Physics Teachers, Berwyn Park, Maryland, 1990.

[67] A. D. Caplin. et al. *"Observation of an unexplained event from a magnetic monopole detector,"*, 321, 1986.

[68] C. Hugenschmidt. Positrons in surface physics. 14, November 2016.

[69] E. A. Euhling. Penetration of heavy charged particles in matter. *Ann. Rev. Nucl. Sci*, 4, 1954.

[70] P. A. Zyla. et al., (particle data group). *Prog. Theor. Exp. Phys*, 2020, 2020.

[71] G. Giacomelli. et al., nucl. *Instrum. Meth*, 411:441–445, 1998.

[72] K. Tanizaki. et al., "geomagnetic cutoff effect on atmospheric muon spectra at ground level,". *28th ICRC*, pages 1163–1166, 2003.

[73] R. R Daniel and S. A. Stephens. "cosmic-ray-produced electrons and gamma rays in the atmosphere". *Reviews of geophysics and space physics*, 2:233, 1974.

[74] H. Kim, C.-T. Chen, N. Eclov, A. Ronzhin, P. Murat, E. Ramberg, S. Los, and C.-M. Kao. A silicon photo-multiplier signal readout using strip-line and waveform sampling for positron emission tomography. *Nucl. Instrum. Methods Phys Res*, A830:119–129, 2016.

[75] Stefan Gundacker. et al., "experimental time resolution limits of modern sipms and tof-pet detectors exploring different scintillators and cherenkov emission,". In *Phys. Med. Biol*. **65 025001, 2020.**

Inclusive search for supersymmetry in pp collisions at $\sqrt{s}$ = 13 TeV using boosted object identification and razor kinematic variables in zero and one lepton final states

Changgi Huh

Department of Physics, Kyungpook National University, Daegu 702-701, Korea

초 록

본 연구는 초대칭성(supersymmetry) 입자의 탐색을 Top 쿼크 또는 W/Z/H 보존 입자가 하드론과 렙톤으로 붕괴한 상태에 대하여 수행되었다. 질량 중심 에너지 13 TeV의 양성자-양성자 충돌 데이터는 통합 광도가 138 fb^{-1}인 LHC의 CMS 실험에서 수집한 데이터를 검색에 활용하였다. 제트 구조 기법을 통해 하드론 또는 렙톤 붕괴를 하는 운동량이 큰 Top 쿼크와 W/Z/H 보존의 이벤트를 식별한 다음 레이저 변수 M^R 및 R^2를 사용하여 분석하여 모든 잠재적 신호를 식별합니다. 신호 영역에서 관찰된 이벤트 수율은 데이터의 제어 샘플을 사용해 만든 표준 모델 프로세스에서 예측된 기여도와 일치합니다. 본 연구는 R-parity를 보존하는 모델과 위반하는 비대칭 모델을 모두 포함합니다. 연구 결과는 gluino 쌍의 생성과 관련하여 분석되며, 여기서 gluino는 중성미자와 함께 Top squark 및/또는 Top 쿼크 또는 제트 및 벡터 보손으로 분해됩니다. 또한, 이 연구는 Top squark가 Top 쿼크와 neutralino로 붕괴하는경우, bottom squark가 bottom 쿼크와 neutralino로 붕괴하는경우, neutralino와 Chargino가 벡터 보손과 다른 neutralino로 붕괴하는 경우들의 직접 생성을 탐구합니다. 138 fb`1 expected limits에서 gluino 질량이 최대 2.1 TeV 그리고 top squark의 질량이 1.2 TeV까지 통계적으로 존재하지 않는다는것을 확인하였다. 또한 2017년에 받은 광도 41.5 fb`1 결과의 observed limit과 expected limits이 2σ 이내로 일치하는 것을 확인하였다.