Search for single top quark production with a Z boson and a quark using 137 fb^{-1} of data at $\sqrt{s} = 13$ TeV by the CMS detector at CERN

Kathryn Coldham, *Brunel University London*

Supervisors: Dr Jo Cole and Professor Akram Khan.
Research Development Advisor: Dr Paul Kyberd.

26 June 2023

Abstract

A search for tZq production in association with a Z boson in dilepton final states is conducted by analysing 137.1 fb^{-1} of data recorded by the CMS detector of the Large Hadron Collider at CERN. A hadronically-decaying W boson candidate from the decay of a top quark and a leptonically-decaying Z boson candidate are considered, where the dilepton is either a pair of electrons or muons. Across the entire dataset, the observed significances were found to be 2.85, 0.60 and 11.9σ for 2016, 2017 and 2018, respectively. In addition to being an important test for the predictions of the Standard Model. The tZq process forms an irreducible background to the tZq flavour changing neutral current (tZq-FCNC) production mechanism, which is predicted by Beyond the Standard Model Physics, the tZq dilepton cross section and signal strength measurements obtained could be useful for future Run 3 analyses.

Declaration of Authorship

The work contained in this thesis was solely conducted by the author. This is with the exception of where collaboration with others has occurred, which was primarily with other members of the High Energy Physics group at Brunel University London and members of the CMS collaboration. The primary contribution of the author was the analysis of 137 fb^{-1} of data recorded by the CMS detector for the tZq search (Sections 9 and 10), the creation of a RIVET plug-in for the tZq process in dilepton final states (Section 7) and a hit resolution measurement study (Section 6) and contributions to the migration from HLS Vivado to Vitis for the Level 1 Track Finding Chain of the Phase 2 CMS Tracker 8. No part of this thesis has been previously submitted to this or any other university as part of the requirement for a higher degree. When the published work of others has been consulted, it has been clearly attributed within the text.

Acknowledgements

Firstly, and most importantly, I would like to say a huge thank you to my husband *Jack*. You have been incredibly supportive over the past five years: from us first meeting in student accommodation at Brunel while I was (unsurprisingly) making myself a cup of tea, going long-distance during my periods of long-term attachment at CERN, living together throughout the lockdowns of the COVID-19 pandemic, and finally getting married towards the end of my PhD. You are my rock and words cannot express how much I love you.

Thank you to my ever-supportive parents, *Tony Coldham* and *Patricia Coldham.* You both have always encouraged me to pursue my dreams and have always been there for me, and I cannot thank you enough for absolutely everything you have both done for me. Thank you to my twin sister *Jennifer Coldham* and my little sister *Kiera Coldham* for your sisterly support.

I would like to say a huge thank you to *Dave Barney.* Your advice, support and mentorship over these past ten years have been invaluable. Thank you so much for taking the time to have our catch ups during my PhD; your encouragement has helped me to keep going through the ups and downs. You have been a complete role model. The bamboo has grown!

Thank you to all the friends I have made during my PhD, including *Katherine Rybacki*, *Naomi Cooke*, *Julia Dancu* and *Vilius Cepatis.* I had such a great time exploring different parts of Switzerland and France with you all, and our scenic hikes. I will cherish those memories forever. Thank you to *Liv Våge* for our chats during LTA while you kindly let me use your washing machine. I'll fondly remember our catch ups with pistacios, chocolate and 'door tea'!

Thank you to *Kin Jong Jamie Poon*, *Vishnu Varma*, *Sabrina Alam*, *Massi Hamadouche*, *Aggie Quesada* for being incredibly supportive friends and keeping in touch over the years, including during my PhD. I've enjoyed our Skype chats while you have all been in Europe, North America and Australia! I'm looking forward to see which countries we will all be in in the years to come.

A big thank you to my fellow members of the CMS Young Scientists Committee:*Ansar Iqbal*, *David Anthony*, *Patrick Asenov*, *Aneliya Karadzhinova-Ferrer*, *Feng Zhang*, *Adelina Lintuluoto*, *Bruno Alves*, *Melissa Quinnan*, *Giovanna Salvi*, *Andris Potrebko*, *Antra Gaile*, *Chris Brown*, *Charlotte Cooke* and *Jenny Lunde.* It has been an absolute pleasure working alongside you all to help improve the working lives of the untenured members of the CMS collaboration. You have also formed a part of my support network, which I am truly grateful for.

Thank you to *Ansar Iqbal* for our analysis chats, technical help and for proof-reading my thesis. Thank you to *Adelina Lintuluoto* for your technical help, teaching me all about HTCondor and Docker. Thank you to *Pieter David* for all of your guidance for the hit resolution study and thank you to *Denise Müller* for all of your guidance regarding NanoAOD and RIVET. Thank you to *Ian Tomalin*, *Alexx Perloff*, *Katie Walkingshaw-Pass* and *Mei-Li Holmberg* with your guidance on HLS Vivado and Vitis.

Thank you to *Ivan Reid* for your help when I ran into computing issues when using the Brunel machines. Thank you to *Catherine Mackay* for helping me at the start of my PhD. Thank you to *Corin Hoad* and *Alexander Morton* for taking the time to answer my analysis questions, and to *Shirin Golpayegani* for being supportive. Thank you to my Research Development Advisor *Paul Kyberd* and to my second supervisor *Akram Khan* for discussing my work during panel meetings. Thank you to my primary supervisor, *Joanne Cole*, for your thesis feedback.

Finally, I would like to thank *Peter Hobson* and the *Science and Technologies Facilities Council*, for awarding me the PhD studentship and making this PhD possible.

Contents

List of Figures

List of Tables

1 Introduction

Chapter 2 will give a brief overview of the Standard Model: the behaviours of the particles predicted by it, the interactions between them and phenomena that it does not account for. Chapter 3 will provide more detail about the top quark, the heaviest elementary particle of the Standard Model. Chapter 4 will explain the structures of the Large Hadron Collider and the CMS detector. Chapter 5 will give a brief overview on the algorithms used in high energy physics for event reconstruction. Chapter 6 will describe a study that was carried out to obtain hit resolution measurements for the CMS Tracker. Chapter 7 will briefly discuss the RIVET plug-in created for tZq production in dilepton final states. Chapter 8 will describe contributions to the migration from HLS Vivado to Vitis for the Level 1 Track Finding Chain of the Phase 2 CMS Tracker. Chapter 9 will describe the methodology steps for the search for tZq production in dilepton final states. Chapter 10 will summarise the results of the search for tZq production in dilepton final states. Chapter 11 will discuss the conclusions from the results of the tZq search and future work that could be carried out for the analysis.

2 The Standard Model

2.1 The Structure of the Atom

The ancient Greeks believed that atoms were the smallest of entities in existence and they could not be broken down any further. Hence, they were given the name “atoms”, which is from the Ancient Greek word “atomos”, meaning “indivisible”. It was later found that atoms contain an internal substructure: a densely-packed nucleus in the centre that contains protons and neutrons, which is surrounded by electrons in mostly empty space. Electrons have an electric charge of -1.602×10^{-19} C while protons have the opposite electric charge of 1.602×10^{-19} C. Neutrons are neutral in electric charge. Electrons are elementary particles, which means that they cannot be broken down further into smaller constituents. The electron is an example of a lepton; this classification of subatomic particle will be further discussed in Section 2.2.2. Meanwhile, protons and neutrons are not elementary particles and each comprise of a smaller classification of subatomic particle called a “quark”. Quarks are further discussed in Section 2.2.1.

The electron was discovered by J.J. Thompson in 1897 at the Cavendish Laboratory in Cambridge, after performing an experiment involving cathode rays. A cathode ray tube was used in the experimental set up; a diagram of this is shown in Figure 1. A glass tube was used, which was evacuated of air to form a vacuum. Inside the tube was a negatively-charged electrode (a cathode) and a positively-charged electrode (an anode). A high voltage was applied across the electrodes to create a potential difference, such that a cathode ray originating at the cathode could propagate from the cathode to the anode, passing through a slit present in each. Oppositely-charged plates were positioned at the middle of the tube with a magnet, and a deflection of the cathode ray was observed.

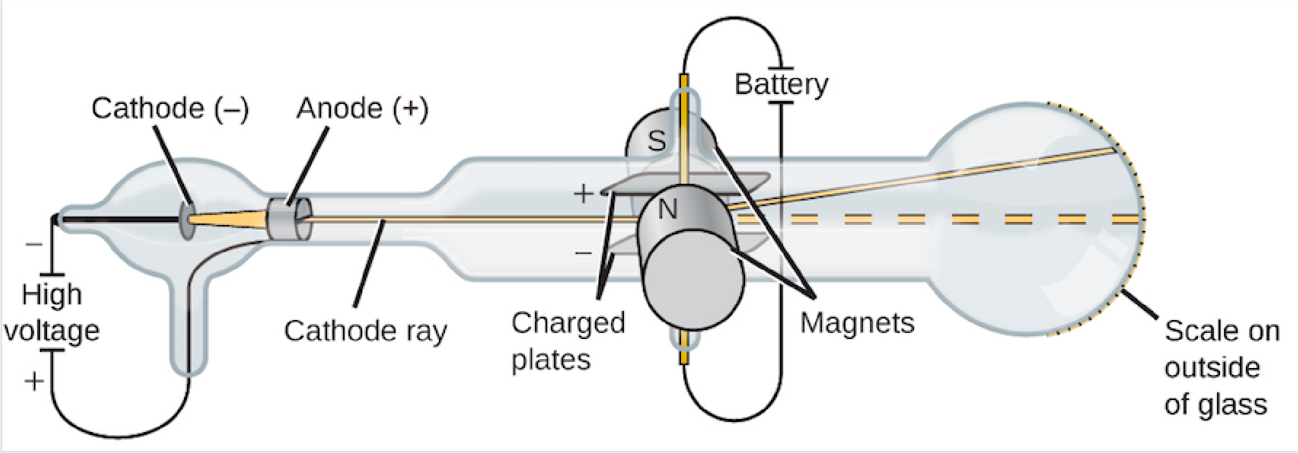


Figure 1: The experimental set up for Thompson’s cathode ray experiment [1].

From this, Thompson inferred that cathode rays actually comprise electrically-charged particles. Thompson dubbed the particles “corpuscles”. He could determine that the particles were negatively charged, based on the direction of curvature of the paths of the cathode rays. He passed the rays through electric and magnetic fields, gradually reducing the field strength until the net deflection of the cathode rays were zero. This enabled the calculation of the velocity of the “corpuscles”, as well as their charge-to-mass ratio. Previously, in 1891, Irish Physicist George Johnstone Stoney referred to the fundamental unit of charge as “electron”, but, later, the scientific community renamed the “corpuscles” to “electrons”. Thompson theorised that electrons are subatomic particles. However, atoms have a net neutral charge and are much heavier than electrons, which indicated that atoms must contain positively-charged, heavier subatomic particles to compensate for this. This led to Thompson proposing his "plum pudding model" to explain the atomic structure. In the model the negatively-charged electrons are dispersed in a sphere of positive charge, as shown in Figure 2, much like the plums in a plum pudding.

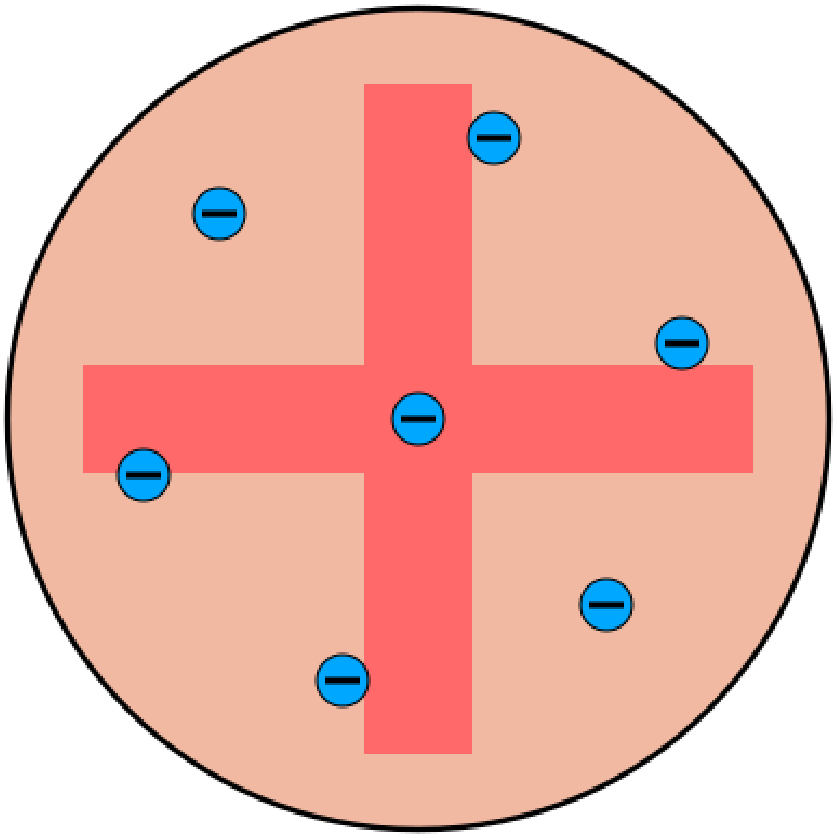


Figure 2: The plum pudding model of the atom, proposed by J.J. Thompson in 1909 [2].

To test this model, in 1909, Ernest Rutherford, Hans Geiger and Ernest Marsden performed the Geiger-Marsden experiment. In the experiment, a beam of alpha particles, where an alpha particle is composed of 2 protons and 2 neutrons, was fired towards a thin sheet of gold foil. A diagram showing the experimental set up is shown in Figure 3. Most of the alpha particles passed through the gold foil with their trajectories only slightly unchanged, while others bounced off the gold foil at large angles. This indicated that atoms contain positively-charged, dense nuclei at their centres. Additionally, the atomic nucleus occupies a tiny fraction of the total volume of the atom. Thus, the hydrogen atom was given the name "proton" by Rutherford, since hydrogen is the lightest atom [3].

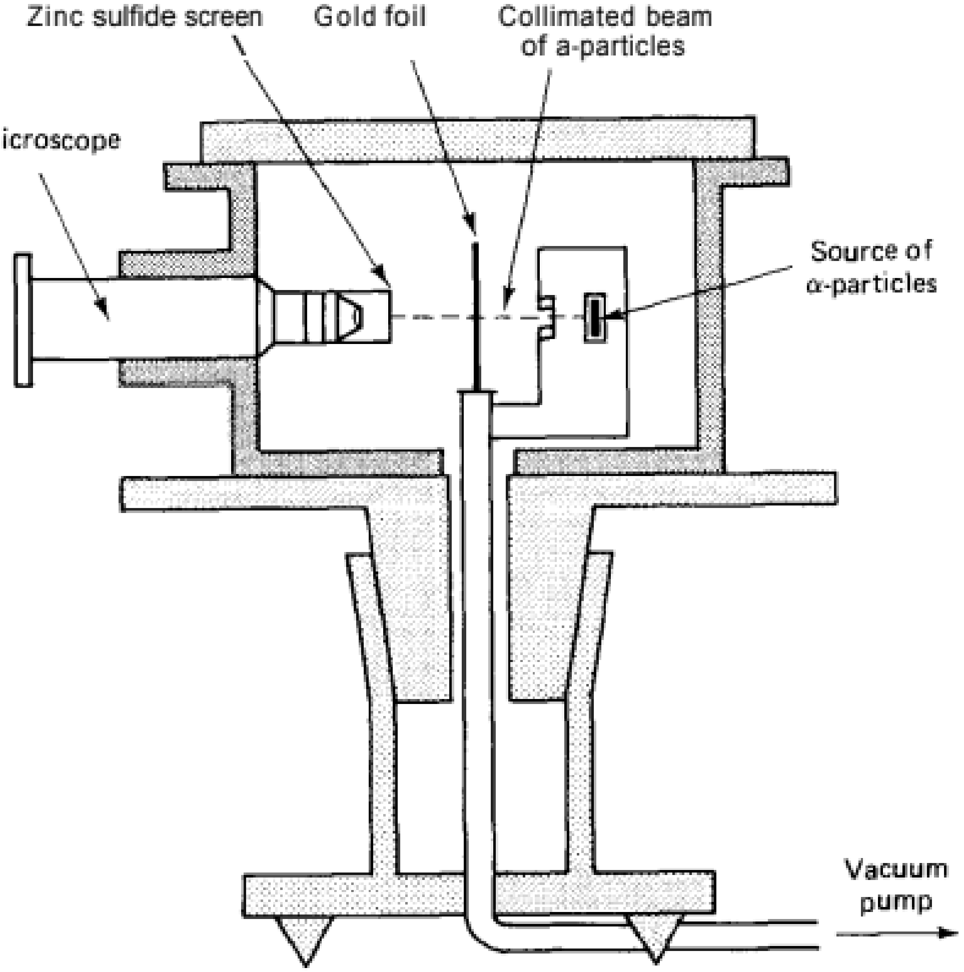


Figure 3: The experimental set up for the Geiger-Marsden experiment [3].

Four years later, the Bohr model for hydrogen was proposed by Niels Bohr, in which a hydrogen atom consists of an electron orbiting a proton at the atomic centre. The orbit is maintained by the attraction of opposite charges. In 1932, James Chadwick discovered that the atomic nucleus also contains a neutrally charged subatomic particle called the neutron. It was later determined that protons and neutrons are composed of constituents called quarks. Quarks will be further discussed in Section 2.2.1.

2.2 The Standard Model

The Standard Model (SM) of particle physics comprises all known elementary particles, and an explanation of the mechanism by which they interact. The elementary particles can be divided into three classes: quarks, leptons and gauge bosons. This is in addition to their anti-particles, which have the same mass as their matter counterparts but opposite charge. A table of the Standard Model particles is shown in Figure 4.

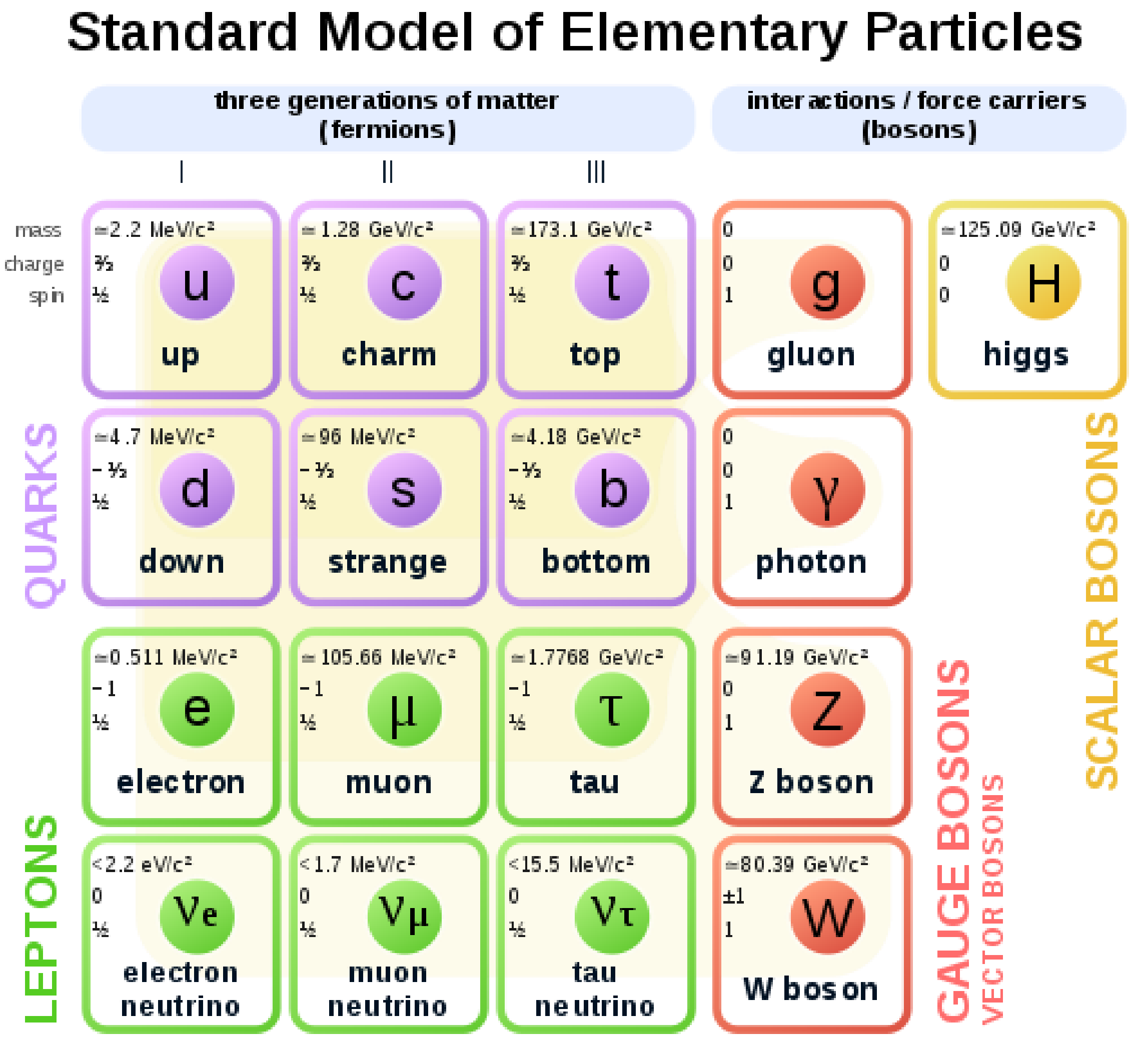


Figure 4: A table of the elementary particles contained in the Standard Model [4].

2.2.1 Quarks and Hadrons

Hadrons are another classification of subatomic particle and can be divided into two sub-classifications: baryons and mesons. Protons and neutrons are examples of baryons.

Baryons and mesons can be arranged into geometrical patterns based on their values of charm and strangeness. The patterns form the "Eightfold Way", which was created by Murray Gell-Mann in 1961. In the Eightfold Way, the three patterns are the baryon octet, the meson octet and the baryon decuplet. These are shown in Figure 5. In the baryon octet, baryons with the same charge Q are situated on the same diagonal line and baryons with the same strangeness are on the same horizontal line. Each subsequent horizontal line represents a strangeness difference of -1 and the strangeness values vary from $S = 0$ to $S = -2$. The meson octet follows a similar convention but, instead, the strangeness values vary from $S = 1$ to $S = -1$. The baryon decuplet consists of 10 heavier baryons than the baryon octet and is arranged in a triangular configuration, where the strangeness values vary between $S = 0$ to $S = -3$ and the charge values vary from $Q = -1$ to $Q = +2$. Anti-baryons and anti-mesons, which have opposite charge and strangeness values to baryons and mesons, can similarly be arranged in the anti-baryon octet, anti-meson octet and anti-baryon decuplet [5].

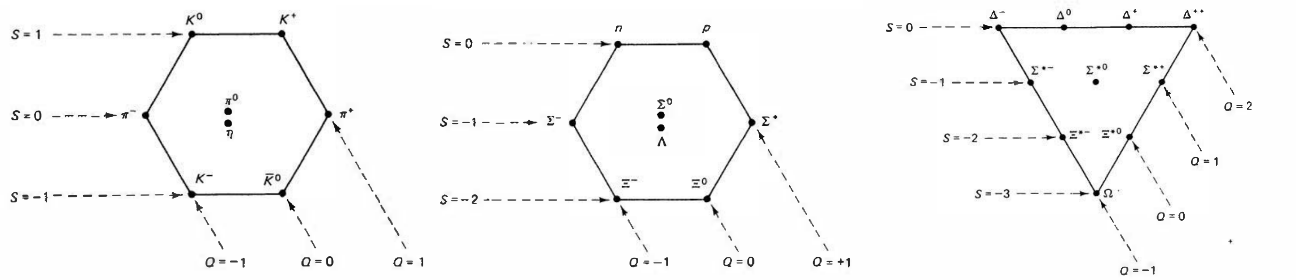


Figure 5: The meson octet (left), the baryon octet (centre) and the baryon decuplet (right) [5]

In the remainder of the 1960s and the 1970s, many more baryons and mesons were observed. Thus, supermultiplets were created that account for all of the possible baryons and mesons that arise from the four quark flavours. The supermultiplets are shown in Figure 6.

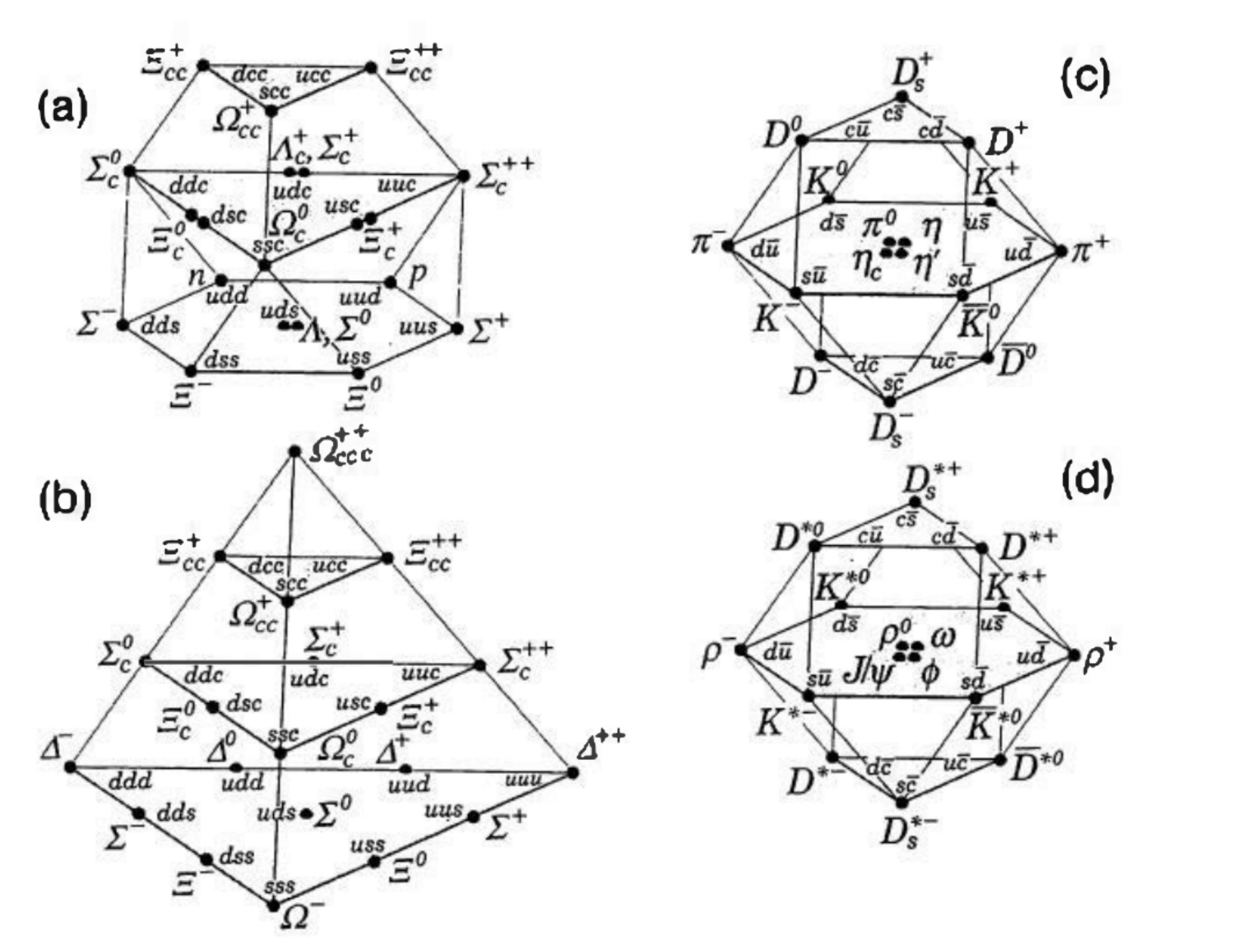


Figure 6: The baryon supermultiplets (a and b) and the meson supermultiplets (c and d) [5].

To explain why hadrons conform to the patterns displayed in supermultiplets, the existence of point-like particles that form hadrons was independently proposed by Gell-Mann and Zweig in 1964. Gell-Mann named these point-like particles “quarks” while Zweig called them “aces”; the former name has since been used. In the quark model, quarks are point-like structures that form baryons and mesons. Baryons consist of three quarks and mesons consist of a quark and an anti-quark. In the Standard Model, quarks are divided into three generations of increasing mass: the up and down quarks form the first generation, the charm and strange quarks form the second and the top and bottom quarks are in the third generation. For example, the proton is a baryon that consists of two ups and one down quark, while the J/ψ particle is a meson that comprises a charm quark and an anti-charm quark. The up, charm and top quarks have a charge of +2/3 and are referred to as "up-type" quarks while the down, strange and bottom quarks have a charge of -1/3 and are called "down-type" quarks. Quarks cannot be observed individually as they are confined inside baryons and mesons. This is known as quark confinement and arises due to the phenomenon of asymptotic freedom. In asymptotic freedom, the attractive potential between quarks increases as the distance between quarks increases, so they can never be unbound [5].

Oscar Wallace Greenberg proposed that quarks possess the property of "colour charge". This was to prevent quarks from violating Pauli's exclusion principle, which states that no two fermions can occupy the same state. The three colours are red, blue and green, so quarks of the same type would not occupy the same state if they are of different colour. The colours of quarks inside baryons and mesons must cancel to result in baryons and mesons being colourless.

In addition to baryons and mesons, the existence of pentaquarks (sub-atomic particles composed of four quarks and one anti-quark) and tetraquarks (comprised of two quarks and two anti-quarks) have been confirmed. Tetraquark production was first observed by the LHCb collaboration at CERN in 2014 [6], while the observation of pentaquark production was reported by the LHCb collaboration in 2015 [7].

The top quark will be the main focus of the research conducted as part of this thesis. Motivations behind researching the top quark and a detailed explanation of its properties are provided in Section 3 and Section 3.1, respectively.

2.2.2 Leptons

The Standard Model also contains leptons which, like quarks, are divided into three generations, called flavours. In the first generation is the electron, in the second is the muon and in the third generation is the tau particle.

The muon was discovered by Carl D. Anderson and Seth Neddermeyer in 1936 after studying cosmic radiation. It is approximately 200 times heavier than the electron and has a lifetime of 2.2 μs.

The muon acts like a spinning magnet and possess a property called the magnetic dipole moment. Also known as the "g-factor", the magnetic dipole moment indicates how strong the magnet is and the rate of its gyration. The equation for the magnetic dipole moment of a lepton (g_l) is shown in Equation 1. In this equation, q represents charge, m_l is the mass of the lepton and $\vec{s}$ represents spin. The concept of spin will be explained in 2.2.4.

$$\vec{\mu_l} = g_l \frac{q}{2m_l} \vec{s} \tag{1}$$

From the Dirac equation, g_l was predicted to have a value of 2. However, in experimental data, the value of g_μ has been shown to not be equal to exactly 2. The difference between the measured value and the value predicted by the Standard Model is known as the anomalous magnetic dipole moment and is a result of higher order contributions from quantum field theory. In 2021, the Muon g-2 collaboration at Fermilab announced a measurement of the anomalous magnetic dipole moment that challenged the value predicted by the Standard Model [8], as shown in Figure 7. The g-2 storage ring used to achieve the measurement is shown in Figure 8.

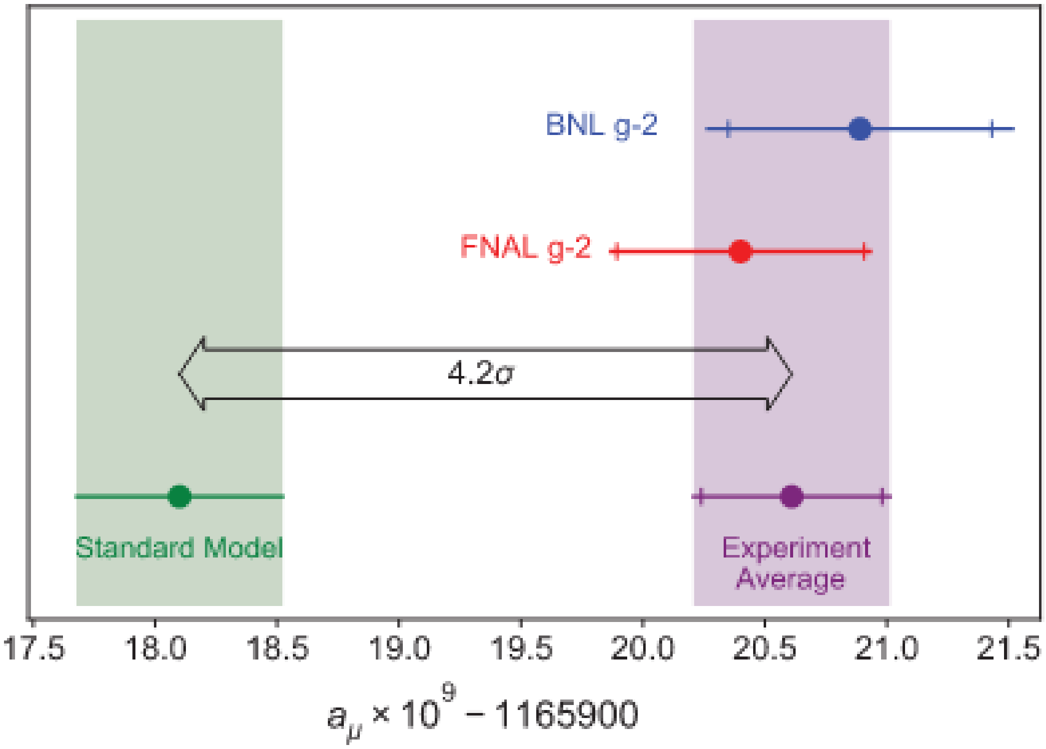


Figure 7: The value of the anomalous magnetic dipole moment predicted by the Standard Model, the 2021 result from the Muon g-2 experiment at Fermilab (FNAL g-2), the 2006 Brookhaven result (BNL g-2) and the experimental average [8].

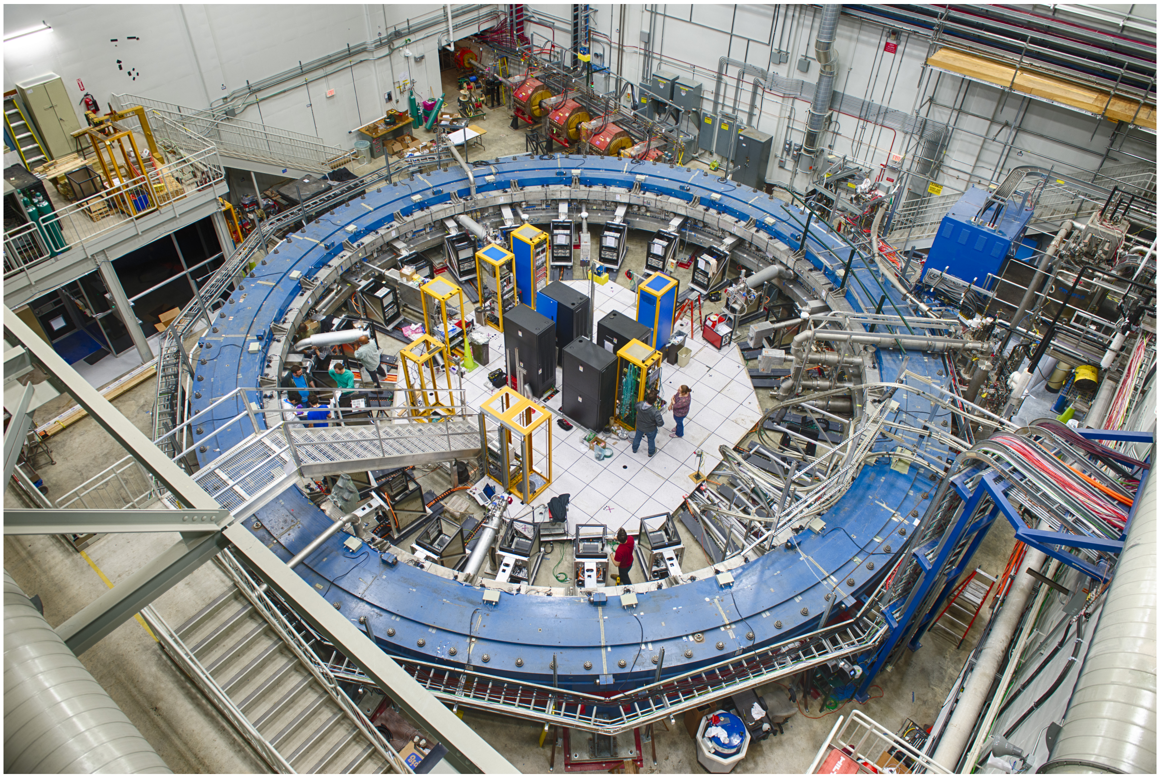


Figure 8: The g-2 storage ring magnet at Fermilab [9].

Like electrons and muons, the tau particle is also an example of a lepton. Its existence was discovered after a series of experiments conducted between 1974 and 1977 by a team led by Martin Lewis Perl at the Stanford Linear Accelerator Center (SLAC). An electron-positron collider called SPEAR was used to study electron-positron annihilation. It was anticipated for events to contain only electrons or positrons, or muons or anti-muons for the interaction products. However, anomalous events were observed in data, such that additional final state particles were detected. It was theorised that at least two additional, undetected particles formed in order for the conservation of energy and momentum to hold. It was determined that tau anti-tau pairs were produced as a result of the interaction, which would go on to decay and produce final state leptons [10].

The three flavours of lepton can be expressed in the doublets shown in Equation 2. The electron (e), muon (μ) and tau (τ) each share a flavour with their corresponding neutrino: the electron neutrino (ν_{e^-}), the muon neutrino (ν_{μ^-}) and the tau neutrino (ν_{τ^-}). Neutrinos are further discussed in Section 2.2.3.

$$\begin{pmatrix} e^- \\ \nu_e \end{pmatrix}, \begin{pmatrix} \mu^- \\ \nu_\mu \end{pmatrix}, \begin{pmatrix} \tau^- \\ \nu_\tau \end{pmatrix} \tag{2}$$

The antileptons can be expressed in the doublets shown in Equation 3.

$$\begin{pmatrix} e^+ \\ \bar{\nu}_{e^+} \end{pmatrix}, \begin{pmatrix} \mu^+ \\ \bar{\nu}_{\mu^+} \end{pmatrix}, \begin{pmatrix} \tau^+ \\ \bar{\nu}_{\tau^+} \end{pmatrix} \tag{3}$$

Leptons possess a quantum number called "lepton number" that must be conserved in interactions. For the electron, muon and tau particles, lepton number can be subdivided into electron number (L_e), muon number (L_μ) and tau number (L_τ), respectively. The formulae for these are shown in

Equations 4-6, where N represents the number of each type of particle participating in the interaction. For leptons $L = +1$, for antileptons $L = -1$ and $L = 0$ for particles that are neither leptons nor antileptons [5].

$$L_e \equiv N(e^-) + N(e^+) + N(\nu_e) - N(\bar{\nu_e}) \tag{4}$$

$$L_\mu \equiv N(\mu^-) + N(\mu^+) + N(\nu_\mu) - N(\bar{\nu_\mu}) \tag{5}$$

$$L_\tau \equiv N(\tau^-) + N(\tau^+) + N(\nu_\tau) - N(\bar{\nu_\tau}) \tag{6}$$

Charged leptons can interact via the electromagnetic and the weak interactions while the neutrinos, which are neutrally charged, can only interact via the weak force.

2.2.3 Neutrinos and Neutrino Oscillation

The three flavours of neutrinos are the electron neutrino, the muon neutrino and the tau neutrino. Neutrinos can be detected via their weak interaction but, like the neutron, the photon and the K_0, they leave no track in a detector. Neutrinos can change their flavour as they propagate, a phenomenon known as neutrino oscillation. In this, the three flavours of neutrino correspond to the neutrino mass eigenstates ν_1, ν_2 and ν_3. The basis of mass eigenstates is related to the weak eigenstates ν_e, ν_μ and ν_τ by a unitary matrix. The weak eigenstates are the different possible flavours of the lepton produced in a charged current weak interaction and the unitary matrix that relates them to the mass eigenstates is called the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix. The relationship between the mass and weak eigenstates is shown in Equation 7 [11].

$$\begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} = \begin{pmatrix} U_{e_1} & U_{e_2} & U_{e_3} \\ U_{\mu_1} & U_{\mu_2} & U_{\mu_3} \\ U_{\tau_1} & U_{\tau_2} & U_{\tau_3} \end{pmatrix} \begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} \tag{7}$$

2.2.4 The Dirac Equation and Spin

Quarks and leptons are fermions, particles with a half integer spin value. Spin is an intrinsic form of angular momentum and the quarks and leptons have a spin value of one half. Additionally, fermions obey the Dirac equation. This is shown in Equation 8 [11].

$$\hat{E}\psi = (\alpha \cdot \hat{\boldsymbol{p}} + \beta m)\psi \tag{8}$$

In Equation 8, $\hat{E}$ is the energy operator, ψ is the wavefunction, $\hat{\boldsymbol{p}}$ is the momentum operator, m is the particle mass. The term $(\alpha \cdot \hat{\boldsymbol{p}} + \beta m)$ is equal to the Dirac Hamiltonian operator, $\hat{H}_D$. The α and β terms are the matrices shown in Equations 10 and 11. In Equation 10, $\vec{\sigma_i}$ represents the Pauli matrices (where $i = 1, 2$ or 3). These are shown in Equation 12. Substituting the expressions for the energy and momentum operators into Equation 8 gives Equation 9

$$i\frac{\partial\psi}{\partial t} = (-i\alpha_x\frac{\partial}{\partial x} - i\alpha_y\frac{\partial}{\partial y} - i\alpha_z\frac{\partial}{\partial z} + \beta m)\psi \tag{9}$$

$$\alpha_i = \begin{pmatrix} 0 & \vec{\sigma_i} \\ \vec{\sigma_i} & 0 \end{pmatrix} \tag{10}$$

$$\beta = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{11}$$

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{12}$$

Squaring Equation 9 gives Equation 13; expanding this results in Equation 14.

$$-i\frac{\partial^2\psi}{\partial^2 t} = (i\alpha_x\frac{\partial}{\partial x} + i\alpha_y\frac{\partial}{\partial y} + i\alpha_z\frac{\partial}{\partial z} - \beta m)(i\alpha_x\frac{\partial}{\partial x} + i\alpha_y\frac{\partial}{\partial y} + i\alpha_z\frac{\partial}{\partial z} - \beta m)\psi \tag{13}$$

$$\frac{\partial^2\psi}{\partial t^2} = \alpha_x^2\frac{\partial^2\psi}{\partial x^2} + \alpha_y^2\frac{\partial^2\psi}{\partial y^2} + \alpha_z^2\frac{\partial^2\psi}{\partial z^2} - \beta^2 m^2\psi + (\alpha_x^2\alpha_y^2 + \alpha_y^2\alpha_z^2)\frac{\partial^2\psi}{\partial z\partial y} + (\alpha_y^2\alpha_z^2 + \alpha_z^2\alpha_y^2)\frac{\partial^2\psi}{\partial y\partial z} + (\alpha_z^2\alpha_x^2 + \alpha_x^2\alpha_z^2)\frac{\partial^2\psi}{\partial z\partial x} \tag{14}$$

Equation 14 reduces to the Klein-Gordon equation, shown in Equation 15, if the α and β matrices satisfy the relations shown in Equations 16-18 (where $j \neq k$). Relativistic fermions must satisfy Equation 8 and also Equation 15, since the Einstein energy-momentum relationship must be satisfied.

$$\frac{\partial^2\psi}{\partial t^2} = \frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2} - m^2\psi \tag{15}$$

$$\alpha_x^2 = \alpha_y^2 = \alpha_z^2 = \beta^2 = I \tag{16}$$

$$\alpha_j\beta + \beta\alpha_j = 0 \tag{17}$$

$$\alpha_j\alpha_k + \alpha_k\alpha_j = 0 \tag{18}$$

The Dirac equation describes spin-half particles. As previously stated, the Dirac Hamiltonian $\hat{H}_D$ is equal to $(\alpha \cdot \hat{\boldsymbol{p}} + \beta m)$. It does not commute with the angular momentum operator $\hat{L}$, defined as the cross product between the position operator $\hat{r}$ and the momentum operator $\hat{p}$. The Dirac Hamiltonian also does not commute with the 4×4 spin operator, $\hat{S}$, which is formed from the Pauli spin-matrices. The mathematical definition of $\hat{S}$ is shown in Equation 19.

$$\hat{S} \equiv \frac{1}{2}\hat{\Sigma} \equiv \frac{1}{2}\begin{pmatrix} \sigma & 0 \\ 0 & \sigma \end{pmatrix} \tag{19}$$

where:

$$\hat{\Sigma_x} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \hat{\Sigma_y} = \begin{pmatrix} 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \end{pmatrix}, \hat{\Sigma_z} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{20}$$

To find the total spin of a particle, s, the eigenvalue of the matrix shown in Equation 21 can be taken. This is because spin can be quantised in the same way as angular momentum.

$$\hat{S}^2 = \frac{1}{4}(\hat{\Sigma}_x^2 + \hat{\Sigma}_y^2 + \hat{\Sigma}_z^2) = \frac{3}{4}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{21}$$

From this the result $\hat{S}^2 \left|s, m_s\right\rangle = s(s+1)\left|s, m_s\right\rangle$ is obtained. Therefore, the square of the spin operator acting on any Dirac spinor, ψ, yields the result shown in Equation 22 only for spin-half particles 22.

$$\hat{S}^2\psi = s(s+1)\psi = \frac{3}{4}\psi \tag{22}$$

The commutation of $\hat{H}_D$ and $\hat{L}$, represented by $[\hat{H}_D, \hat{L}]$ gives a result of $-i\alpha \times \hat{p}$. This is shown in Equations 23- 26.

$$[\hat{H}_D, \hat{L}] = [\alpha \cdot \hat{p} + \beta m, \hat{L}] = [\alpha \cdot \hat{p}, \hat{r} \times \hat{p}] + [\beta m, \hat{L}] = [\alpha \cdot \hat{p}, \hat{L}] \tag{23}$$

The $\hat{L}$ is equal to $\hat{L}_x + \hat{L}_y + \hat{L}_z$ and $\alpha \cdot \hat{p}$ is equal to $\alpha_x\hat{p}_x + \alpha_y\hat{p}_y + \alpha_z\hat{p}_z$. By firstly considering the contribution of $\hat{L}_x$, which is equal to $y\hat{p}_z - z\hat{p}_y$, Equation 23 can be expressed as Equation 24.

$$[\alpha \cdot \hat{p}, \hat{L}] = [\alpha_x\hat{p}_x + \alpha_y\hat{p}_y + \alpha_z\hat{p}_z, y\hat{p}_z - z\hat{p}_y] = [\alpha_x\hat{p}_x + \alpha_y\hat{p}_y + \alpha_z\hat{p}_z, y\hat{p}_z] - [\alpha_x\hat{p}_x + \alpha_y\hat{p}_y + \alpha_z\hat{p}_z, z\hat{p}_y] \tag{24}$$

By using the Leibniz rules and the relations $[\hat{x}, \hat{p}_x] = [\hat{y}, \hat{p}_y] = [\hat{z}, \hat{p}_z] = i$, Equation 24 and be expressed as Equation 25

$$[\hat{H}_D, \hat{L}_x] = \alpha_y[\hat{p}_y, \hat{y}]\hat{p}_z - \alpha_z[\hat{p}_z, \hat{z}]\hat{p}_y = i(\alpha_y\hat{p}_z - \alpha_z\hat{p}_y) = -i(\boldsymbol{\alpha} \times \hat{\boldsymbol{p}})_x \tag{25}$$

Similarly, $[\hat{H}_D, \hat{L}_y] = -i(\boldsymbol{\alpha} \times \hat{\boldsymbol{p}})_y$ and $[\hat{H}_D, \hat{L}_z] = -i(\boldsymbol{\alpha} \times \hat{\boldsymbol{p}})_z$. From this, the result for $[\hat{H}_D, \hat{\boldsymbol{L}}]$ can be obtained. This is shown in Equation 26.

$$[\hat{H}_D, \hat{\boldsymbol{L}}] = -i\alpha \times \hat{\boldsymbol{p}} \tag{26}$$

The commutation of $\hat{H}_D$ and $\hat{S}$, represented by $[\hat{H}_D, \hat{S}]$ gives a result of $+i\alpha \times \hat{p}$. This is shown in Equations 27-33.

$$[\hat{H}_D, \hat{S}] = [\hat{\alpha} \cdot \hat{p} + \beta m, \frac{1}{2}\hat{\Sigma}] = [\hat{\alpha} \cdot \hat{p}, \frac{1}{2}\hat{\Sigma}] + [\beta m, \frac{1}{2}\hat{\Sigma}] = [\hat{\alpha} \cdot \hat{p}, \frac{1}{2}\hat{\Sigma}] = [\hat{\alpha}_x\hat{p}_x + \hat{\alpha}_y\hat{p}_y + \hat{\alpha}_z\hat{p}_z, \frac{1}{2}(\hat{\Sigma}_x + \hat{\Sigma}_y + \hat{\Sigma}_z)] \tag{27}$$

By firstly considering $\hat{S}_x = \frac{1}{2}\hat{\Sigma}_x$, Equation 28 is obtained.

$$[\hat{H}_D, \hat{S}_x] = [\hat{\alpha}_x\hat{p}_x + \hat{\alpha}_y\hat{p}_y + \hat{\alpha}_z\hat{p}_z, \frac{1}{2}\hat{\Sigma}_x] = [\hat{\alpha}_x\hat{p}_x, \frac{1}{2}\hat{\Sigma}_x] + [\hat{\alpha}_y\hat{p}_y, \frac{1}{2}\hat{\Sigma}_x] + [\hat{\alpha}_z\hat{p}_z, \frac{1}{2}\hat{\Sigma}_x] \tag{28}$$

This results in Equation 29.

$$[\hat{H}_D, \hat{S}_x] = \frac{1}{2}\hat{p}_x[\hat{\alpha}_x, \hat{\Sigma}_x] + \frac{1}{2}\hat{p}_y[\hat{\alpha}_y, \hat{\Sigma}_x] + \frac{1}{2}\hat{p}_z[\hat{\alpha}_z, \hat{\Sigma}_x] \tag{29}$$

The terms $[\hat{\alpha}_x, \hat{\Sigma}_x]$, $[\hat{\alpha}_y, \hat{\Sigma}_x]$ and $[\hat{\alpha}_z, \hat{\Sigma}_x]$ can be written in matrix form. This is shown in Equation 30, where $i = x$, y or z.

$$[\hat{\alpha}_i, \hat{\Sigma}_x] = \begin{pmatrix} 0 & \sigma_i \\ \sigma_i & 0 \end{pmatrix} \begin{pmatrix} \sigma_x & 0 \\ 0 & \sigma_x \end{pmatrix} - \begin{pmatrix} \sigma_x & 0 \\ 0 & \sigma_x \end{pmatrix} \begin{pmatrix} 0 & \sigma_i \\ \sigma_i & 0 \end{pmatrix} = \begin{pmatrix} 0 & [\sigma_i, \sigma_x] \\ [\sigma_i, \sigma_x] & 0 \end{pmatrix} \tag{30}$$

Substituting the cyclic relations $[\sigma_x, \sigma_x] = 0$, $[\sigma_y, \sigma_x] = -2i\sigma_z$ and $[\sigma_z, \sigma_x] = 2i\sigma_y$ into Equation 30 gives the expressions in Equation 31.

$$[\alpha_x, \Sigma_x] = 0, [\alpha_y, \Sigma_x] = -2i\alpha_z, [\alpha_z, \Sigma_x] = 2i\alpha_y \tag{31}$$

This simplifies Equation 31 to 32

$$[\hat{H}_D, \hat{S}_x] = \frac{1}{2}(-2i\hat{p_y}\alpha_z + 2i\hat{p_z}\alpha_y) = \frac{1}{2}[2i(\alpha \times \hat{p})_x] = (\alpha \times \hat{p}_x) \tag{32}$$

Similarly, $[\hat{H}_D, \hat{S}_y] = (\alpha \times \hat{p})_y$ and $[\hat{H}_D, \hat{S}_z] = (\alpha \times \hat{p})_z$. This leads to the general result shown in Equation 33.

$$[\hat{H}_D, \hat{S}] = (\alpha \times \hat{p}) \tag{33}$$

Since $\hat{H}_D$ does not commute with $\hat{L}$ or $\hat{S}$, the observables $\hat{L}$ or $\hat{S}$ do not correspond to observed quantities. However, finding the commutation of $\hat{H}_D$ and the total angular momentum $\hat{J}$, which is equal to $\hat{L} + \hat{S}$, gives a result of zero. This is shown in Equation 34. Therefore the spin observable corresponds to a conserved quantity.

$$[\hat{H}_D, \hat{J}] = [\hat{H}_D, \hat{L} + \hat{S}] = [\hat{H}_D, \hat{L}] + [\hat{H}_D, \hat{S}] = (-i\alpha \times \hat{p}) + (+i\alpha \times \hat{p}) = 0 \tag{34}$$

2.2.5 Bosons and the Fundamental Forces

Alternatively, particles with an integer spin are called bosons. These are responsible for transporting the fundamental forces between quarks and leptons, as described in Quantum Field Theory (QFT). In order of decreasing strength, the fundamental forces are: the strong nuclear force, the electromagnetic force, the weak nuclear force and the gravitational force. The gauge bosons in the Standard Model are the gluon, the photon and the W and Z bosons. In the theory of Quantum Chromodynamics (QCD), the massless gluon transports the strong nuclear force between quarks with colour charge, keeping quarks tightly bound in nuclei. In the theory of Quantum Electrodynamics (QED), the virtual photon, which is also massless, transports the electromagnetic force between charged particles. This results in the attraction and repulsion between the charged particles. The W and Z bosons transport the weak nuclear force between particles with weak isospin. There are two oppositely-charged W bosons, the W^+ and the W^-, that have a mass of 80.2 GeV. The W bosons mediate the weak charged-current interaction, which couples to quarks that differ by one unit of charge and is responsible for nuclear beta decay and nuclear fusion. The weak charged-current interaction coupling is greatest for quarks of the same flavour. For leptons, the weak interaction couples a charged lepton with its corresponding neutrino. The electrically-neutral Z boson has a mass of 91.2 GeV and mediates the weak neutral-current interaction.

Each fundamental force has an associated coupling strength, represented by g. For QED, $g_{QED} = +|e|$, where e is the charge of the electron (1.602×10^{-19} C). For QCD the strong coupling constant is represented by g_s and, for the weak interaction, the weak coupling constant is g_W. For each fundamental force, the coupling strength g is related to a dimensionless constant α, called the intrinsic strength, where $\alpha \propto g^2$. For the electromagnetic interaction, α_{QED} is the fine structure constant. The fine structure constant is shown in Equation 35, where e is the electron charge, ϵ_0 is the permittivity of free space ($8.85 \times 10^{-12} Fm^{-1}$), $\hbar$ is the reduced Planck's constant and c is the speed of light. For QCD $\alpha_s \sim 1$ and, for the weak interaction, $\alpha_W \sim \frac{1}{30}$.

$$\alpha_{QED} = \frac{e^2}{4\pi\epsilon_0\hbar c} = \frac{1}{137} \tag{35}$$

Gauge Theory

In quantum mechanics, wavefunctions that satisfy wave equations are used to describe single particles. In Quantum Field Theory, particles are instead described by excitations of a quantum field that must satisfy quantum mechanical field equations. The Lagrangian (L) can be used to obtain equations of motion and is a function dependent on generalised coordinates (q_i) and their time derivatives ($\dot{q}_i$). This is shown in Equation 36, where T and V represent the kinetic and potential energies of the system, respectively [12].

$$L = T - V \tag{36}$$

The equations of motion can then be determined by using the Euler-Lagrange equations, shown by equation 37

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} = 0 \tag{37}$$

However, for the dynamics of Quantum Field Theories, continuous systems rather than discrete systems of particles are considered. Therefore, the Lagrangian density is used, which is shown in Equation 38. The Lagrangian density is expressed as a function of fields ($\phi_i(x, y, z)$) and by their derivatives with respect to the four space-time coordinates, which is shown in Equation 39 [12].

$$L(q_i, \frac{dq_i}{dt}) \rightarrow \mathcal{L}(\phi_i, \partial_\mu \phi_i) \tag{38}$$

$$\partial_\mu \phi_i \equiv \frac{\partial \phi_i}{\partial x_\mu} \tag{39}$$

For such systems, the Lagrangian L can be expressed by Equation 40

$$L = \int \mathcal{L} d^3\mathbf{x} \tag{40}$$

The Euler-Lagrange equation for fields can be derived using the principle of least action. This is shown in Equation 41.

$$\partial_\mu \left(\frac{\partial L}{\partial(\partial_\mu \phi_i)} \right) - \frac{\partial L}{\partial \phi_i} = 0 \tag{41}$$

Quantum Chromodynamics and the gluons

The strong nuclear force is transported by the gluon. The Quantum Field Theory of the strong interaction is Quantum Chromodynamics (QCD), which requires invariance under SU(3) local phase transformations. This is represented by Equation 42 [12].

$$\psi(x) \rightarrow \psi'(x) = exp[ig_s \alpha(x).\hat{\mathbf{T}}]\psi(\mathbf{x}) \tag{42}$$

The eight generators of the SU(3) symmetry group are represented by $\hat{T}$. The generators do not commute, which is why QCD is a non-Abelian gauge theory. The generators' dependence on the Gell-Mann matrices (λ^a) is shown by Equation 43 [12], for $a = 1, 2, ..., 8$.

$$T^a = \frac{1}{2}\lambda^a \tag{43}$$

The Dirac equation under SU(3) local phase transformations is shown in Equation 44

$$i\gamma^\mu[\partial_\mu + ig_s(\partial_\mu \alpha.\hat{\mathbf{T}})]\psi = \mathbf{m}\psi \tag{44}$$

Eight new fields $G^a_\mu(x)$, where $a = 1, ..., 8$ and each correspond to a generator of SU(3) symmetry, are introduced to achieve the local gauge invariance. These eight fields correspond to gluons. If the new fields transform as shown in Equation 45, this leads to the Dirac Equation being invariant under local SU(3) phase transformations, as shown in Equation 46.

$$G^k_\mu \rightarrow G^{k'}_\mu = G^k_\mu - \partial_{k\mu} - g_S f^{ijk} \alpha_i G^j_\mu \tag{45}$$

$$i\gamma^\mu[\partial^\mu + ig_S G^a_\mu T^a]\psi - m\psi = 0 \tag{46}$$

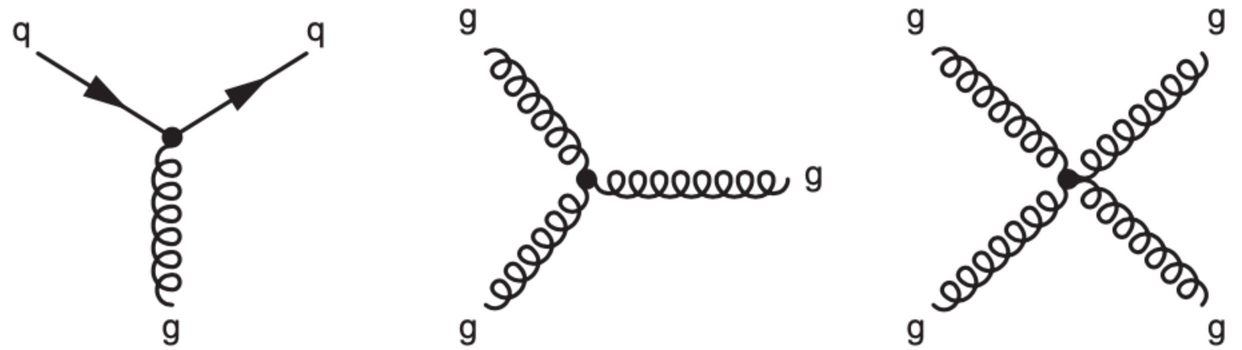


Figure 9: The gluon self-interactions that arise from SU(3) local gauge invariance [12].

In Equation 46, f_{ijk} represent the structure constants of SU(3) and are defined by the commutation relations $[\lambda_i, \lambda_j] = 2if_{ijk}\lambda_k$. The term $g_S f^{ijk}\alpha_i G^j_\mu$ arises from the non-Abelian nature of QCD and results in triple and quartic gluon self-interactions [12].

From Equation 46, the form of the qqg interaction vertex can be obtained, as shown in Equation 47.

$$g_S T^a \gamma^\mu G^a_\mu \psi = g_s \frac{1}{2} \lambda^a \gamma^\mu G^a_\mu \psi \tag{47}$$

A new degree of freedom, called "colour" charge, must be included in the wavefunction ψ, since 3 by 3 matrices represent the generators of SU(3). Colour charge is the conserved quantity in QCD. The terms "red" (r), "green" (g) and "blue" (b) are used to label the colour charges, which are represented by the wavefunctions shown in Equation 48.

$$r = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, g = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, b = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{48}$$

These states are rotated in colour space under the SU(3) local phase transformation. This rotation takes place about an axis that has a different direction at every point in space time. Particles with non-zero colour charge couple to gluons and the strength of the QCD interaction depends on the colour charge of the quark. The third component of colour isopsin I^c_3 and colour hypercharge Y^C are additive quantum numbers that can be used to label the colour states of quarks and antiquarks [12].

Quantum Electrodynamics and the photon

The electromagnetic interaction is transported by the photon. The photon is massless, which results in it having an infinite range. The Quantum Field Theory that describes this interaction is Quantum Electrodynamics (QED), which requires the Dirac equation to be invariant under U(1) local phase transformations. Fields under this transformation can be expressed by Equation 49

$$\psi(x) \rightarrow \psi'(x) = e^{iq\chi(x)}\psi(x) \tag{49}$$

The Lagrangian for a free spin-half particle is shown in Equation 50.

$$L = i\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - m\bar{\psi}\psi \tag{50}$$

Under the transformation, Equation 50 becomes L', where χ is the local phase. This is shown in Equation 51

$$L' = ie^{-iq\chi}\bar{\psi}\gamma^{\mu}[e^{iq\chi}\partial_{\mu}\psi + iq(\partial_{\mu}\psi) + iq(\partial_{\mu}\chi)e^{iq\chi}\psi]me^{-iq\chi}\bar{\psi}e^{iq\chi}\psi = L - q\bar{\psi}\gamma^{\mu}(\partial_{\mu}\chi)\gamma \tag{51}$$

For Equation 51 to be invariant under the U(1) local phase transformation, the derivative ∂_{μ} must be replaced by the covariant derivative D_{μ} shown in Equation 52.

$$\partial_{\mu} \rightarrow D_{\mu} = \partial_{\mu} + iqA_{\mu} \tag{52}$$

In this, A_{μ} represents a new field which transforms as shown in Equation 53

$$A_{\mu} \rightarrow A'_{\mu} = A_{\mu} - \partial_{\mu}\chi \tag{53}$$

Considering A_{μ}, the gauge-invariant Lagrangian for a half-spin fermion can be expressed as Equation 54

$$L = \bar{\psi}(i\gamma^{\mu}\partial_{\mu} - m)\psi - q\bar{\psi}\gamma^{\mu}A_{\mu}\psi \tag{54}$$

In QED, electric charge is conserved and the massless photon mediates the interaction between the electron (for which $q = -e$) and A_{μ}. Therefore, the Lagrangian of QED can be expressed as Equation 55, where $F_{\mu\nu}F_{\mu\nu}$ is the kinetic term for the massless spin-1 field and is itself invariant under U(1) local gauge transformations [12].

$$L_{QED} = \bar{\psi}(i\gamma^{\mu}\partial_{\mu} - m_e)\psi - e\bar{\psi}\gamma^{\mu}\psi A_{\mu} - \frac{1}{4}F_{\mu\nu}F_{\mu\nu} \tag{55}$$

The Weak Nuclear Interaction and the W and Z bosons

The weak nuclear interaction is transported by the heavy W and Z bosons; the W boson carries a charge of ±1 while the Z boson is neutral. The weak nuclear interaction is responsible for radioactive decay and couples to weak isospin (T). The electron, the muon, the tau, the down quark, the strange quark and the bottom quark each have weak isospin values of $-\frac{1}{2}$. Neutrinos and the up, charm and top quarks have weak isospin values of $+\frac{1}{2}$. Isospin is defined in terms of the Pauli spin-matrices, as shown in Equation 56 [12].

$$\hat{\mathbf{T}} = \frac{\mathbf{1}}{\mathbf{2}}\sigma \tag{56}$$

The mathematical structure of isospin is defined by the Abelian SU(2) symmetry group. The three generators of the SU(2) group can be represented by $\hat{T}_1$, $\hat{T}_2$ and $\hat{T}_3$ and satisfy the commutation relations shown in Equation 57.

$$[\hat{T}_1, \hat{T}_2] = i\hat{T}_3, [\hat{T}_2, \hat{T}_3] = i\hat{T}_1, [\hat{T}_3, \hat{T}_1] = i\hat{T}_2 \tag{57}$$

The quantity T_3, defined as the projection of T along the z-axis, is conserved in weak nuclear interactions. The W boson has a T_3 value of $\pm\frac{1}{2}$ whilst the Z has a T_3 value of 0 [12].

The relationship between electric charge Q, weak isospin T_3 and weak hypercharge Y_W is shown by Equation 58. Weak hypercharge is the conserved quantum number in the gauge symmetry U(1). It is associated with interactions with the Higgs field; the concept of the Higgs field will be explained in Section 2.2.6.

$$Q = T_3 + \frac{1}{2}Y_W \tag{58}$$

2.2.6 The Higgs Boson

The latest addition to the Standard Model is the Higgs boson, which has a mass of 125.10 ± 0.14 GeV [?]. It possesses a spin of zero, so it is a scalar particle, and it is responsible for providing mass to the other elementary particles via the Brout-Englert-Higgs mechanism. Its main production mechanisms are, in order of dominance at the LHC: gluon fusion, vector boson fusion, higgs-strahlung and top fusion. Feynman diagrams for these are shown in Figure 10. In gluon fusion, two gluons collide to form a virtual top or bottom quark loop that produces a Higgs boson. In vector boson fusion, virtual W or Z bosons are exchanged between a pair of fermions or anti-fermions and produce a Higgs. In higgs-strahlung, a fermion and an anti-fermion collide to form a virtual W or Z boson that radiates a Higgs. Finally in top fusion, a top quark and an anti-top quark are produced in top-antitop pairs via gluon fusion. The top and antitop go on to fuse to form a Higgs.

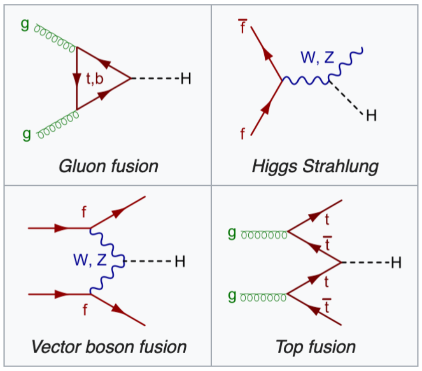


Figure 10: The production mechanisms for the Higgs boson [13].

The five main decay channels for a 125 GeV Higgs boson are: $H \rightarrow ZZ \rightarrow 4l$, $H \rightarrow WW \rightarrow l\nu l\nu$, $H \rightarrow b\bar{b}$, $H \rightarrow \tau\bar{\tau}$ and $H \rightarrow \gamma\gamma$. Observations of $H \rightarrow ZZ \rightarrow 4l$, $H \rightarrow WW \rightarrow l\nu l\nu$ and $H \rightarrow \gamma\gamma$ were reported by the ATLAS and CMS collaborations in 2012 [14] [15], while observations of the $H \rightarrow \tau\bar{\tau}$ and $H \rightarrow b\bar{b}$ decay channels were reported in 2017 and 2018, respectively [16] [17].

2.3 Beyond the Standard Model Physics

There are many physics phenomena that the Standard Model does not account for. The unexplained phenomena include an unobserved candidate for dark matter, the observed asymmetry between matter and antimatter in the universe and a force carrier for the gravitational force.

2.3.1 Dark Matter and Dark Energy

In the ΛCDM model of cosmology, baryonic matter only comprises 5% of the total-energy matter density of the universe ($\Omega_B = 0.05$). The rest is predicted to consist of cold dark matter ($\Omega_C \approx 0.23$) and dark energy ($\Omega_\Lambda = 0.72$). The concept of dark energy arises from the non-zero cosmological constant in Einstein's equations of general relativity and results in the acceleration of the universe. The existence of dark matter was proposed to provide an explanation for a number of astrophysical phenomena. These include conclusions drawn from rotation curves of spiral galaxies. The majority of the mass in a spiral galaxy is located at the bulge in the centre. However, for a star in the disc region, its mass m and tangential velocity v are related by Equation 59

$$\frac{mv^2}{r} \approx \frac{Gm}{r^2} M(r) \tag{59}$$

In Equation 59, r is the radial distance between the star and the centre of the galaxy and $M(r)$ is the total mass contained in r. Thus, as r increases v should decrease by $1/r^2$. However, observations showed a slow decrease of v with r, indicating that the mass distribution followed a $M(r) \propto r$ relationship. To explain the observations, the existence of dark matter was proposed as this would provide additional, non-luminous matter to the galaxies.

A number of cold dark matter candidates have been proposed. These include the hypothetical Weakly Interacting Massive Particles (WIMPS) that interact with the nuclei of baryonic matter via elastic scattering. Therefore, experiments attempt to measure the recoil of nuclei that have potentially been subjected to this scattering process to infer the existence of WIMPs. The recoiling nucleus can emit ionisation energy that leads to the production of scintillation light in sodium iodide crystals or liquid noble gas detectors. Alternatively, pure silicon or germanium crystals that are cooled to low temperatures can be used as a detection medium. As of May 30, 2024, there has been no observation of WIMPs [11].

2.3.2 Supersymmetry

In supersymmetry (SUSY) every SM particle corresponds to a sparticle, which differs from its SM partner by half of a unit of spin. The SM particles and their corresponding sparticles are listed in Table 2.3.2

SM Particle	Sparticle
Quark (q)	Squark
Lepton (l)	Slepton
Neutrino (ν)	Sneutrino
Gluon (g)	Gluino
Photon (γ), Z boson (Z)	Neutralino
Higgs boson (H)	Higgsino
W boson ($W^{\pm}$)	Chargino

Table 1: The SM particles and their corresponding sparticles.

The existence of sparticles would provide a solution to the hierarchy problem in physics. The presence of quantum loops in the Higgs boson propagator contribute to the Higgs mass; the size of the mass contributions are quadratic in the energy scale, Λ. Thus, the contributions become large at the Grand Unified Theory ($\Lambda_{GUT} \sim 10^{16}$ GeV) and Planck ($\Lambda_p \sim 10^{19}$ GeV) scales, which would result in the

Higgs mass no longer being at the electroweak scale of 10^2 GeV. The Minimal Supersymmetric Model (MSSM) could provide a solution to this. In the MSSM, a loop in the Higgs boson propagator would be cancelled out by a Higgsino loop that is opposite in sign. For an exact cancellation the Higgs and Higgsino would be of the same mass. Thus, the SM particles cannot have exactly the same mass as their corresponding sparticles, else they would have already been observed. The MSSM predicts that SUSY particles decay into final states that include the stable lightest supersymmetric particle (LSP). A LSP is neutrally-charged so cannot be directly detected but, instead, missing transverse momentum calculations are used in the searches.

Additionally, SUSY may provide a dark matter candidate. The lightest neutralino is a stable particle that decays via the weak interaction so is a potential WIMP. Searches are being conducted at the LHC for SUSY particles but, as of May 30, 2024, no candidates have been observed. However, squark and gluino masses below 1 TeV have been excluded [11].

2.3.3 Matter-Antimatter Asymmetry

It was once believed that the Big Bang created equal amounts of matter and antimatter in the universe. However, there is actually more matter than antimatter so an interaction must exist that violates both baryon number and lepton number conservation. For these to be violated, two conditions must have been met. Firstly, there must have been a period of time when thermal equilibrium temporarily ceased to result in a net change in baryon number. Secondly, charge-parity (CP) violation must exist.

The first evidence of charge-parity violation came in 1964 from a study of the decay of neutral kaons [18]. Neutral kaons are mesons that consist of a down quark and an anti-strange quark. Neutral anti-kaons consist of a strange quark and an anti-down quark. The K^0-short (K_S^0) and K^0-long (K_L^0) are superpositions of neutral kaons and anti-kaons, as shown in Equations 60 and 61. The K_S^0 and K_L^0 have masses around 498 MeV/c^2 but they have different lifetimes. The lifetime of K_S^0 is 9×10^{-11}s and the lifetime of K_L^0 is 5×10^{-8}s [5].

$$K_S^0 = \frac{K^0 - \bar{K^0}}{\sqrt{2}} \tag{60}$$

$$K_L^0 = \frac{K^0 + \bar{K^0}}{\sqrt{2}} \tag{61}$$

The K_S^0 can either decay into two neutral pions (BR = 31%) or a pion-antipion pair (BR = 69%). The K_L^0 can either decay into three neutrally-charged pions (BR = 20%), a pion, an anti-pion and a neutral pion (BR = 0.13%) or a pion/antipion, a lepton/antilepton and a neutrino/antineutrino (BR = 0.67%). The decays of K_S^0 and K_L^0 are shown in Equations 62 and 63.

$$\begin{aligned} K_S^0 &\rightarrow \pi^0 + \pi^0 \\ K_S^0 &\rightarrow \pi^+ + \pi^- \end{aligned} \tag{62}$$

$$\begin{aligned} K_L^0 &\rightarrow \pi^0 + \pi^0 + \pi^0 \\ K_L^0 &\rightarrow \pi^+ + \pi^0 + \pi^- \\ K_L^0 &\rightarrow \pi^{+/-} + l^{-/+} + \nu_l(\bar{\nu}_l) \end{aligned} \tag{63}$$

The kaon has a spin value of zero. Thus, in the rest frame of the kaon, the orbital momentum of the pion pair must be zero in order to conserve the total orbital angular momentum L. [5]. The intrinsic parity of the pion P_π is equal to -1, which results in the kaon having a parity of 1. This is shown in Equation 64.

$$P = P_\pi^2(-1)^L = (-1)^2(-1)^L = 1 \tag{64}$$

Additionally, the charge conjugation of the kaon, C, can be calculated using the charge conjugations of the produced pions, C_{π^0}.

$$C = (C_{\pi^0})^2 = 1 \tag{65}$$

Therefore, for final state of a neutral pion pair $CP = +1$ so CP is not conserved. Similarly, for the $\pi^+\pi^-$ final state, $CP = +1$ so CP is also not conserved. For the decay of K_L^0 into three pions, the total angular momentum in the rest frame of the kaon must be zero. This is shown in Equation 66, where L_{12} is the orbital angular momentum of a pair of the produced pions and L_3 is the orbital angular momentum of the third pion.

$$L = L_{12} + L_3 = 0 \tag{66}$$

For Equation 66 to hold true, $L_{12} = L_3$. The parity can then be calculated, as shown in Equation 67

$$P = P_\pi^3(-1)^{L_{12}}(-1)^{L_3} = (-1)^3(-1)^0(-1)^0 = -1 \tag{67}$$

For the three-neutral pion final state, the charge conjugation is calculated in Equation 68

$$C = (C_{\pi^0})^3 = (1)^3 = 1 \tag{68}$$

Overall, this gives a result of $CP = -1$ for the three-neutral pion state. For the $\pi^+\pi^0\pi^-$ state, $P = -1$ but the charge conjugation differs by a factor of $(-1)^{L_{12}}$. This is shown in Equation 69. The study of the pions' angular distribution has determined $(-1)^{L_{12}} = 0$. This results in $CP = -1$, so charge-parity is again violated.

$$C = C_{\pi^0}(-1)^{L_{12}} = (1)(-1)^{L_{12}} \tag{69}$$

The first experiment to provide evidence of CP violation was conducted by Christenson *et al.* in 1964; the experimental set up for which is shown in Figure 11. A 30 GeV proton beam was fired into a stationary beryllium target and a lead collimator was used to produce a secondary beam from the resultant particles. A bending magnet was used to remove charged particles from the secondary beam and the beam traversed through a block of lead to absorb any photons contained. At the block, the beam contained a mixture of K_S^0 and K_L^0. After this point, a helium bag was used. The K_S^0 and K_L^0 decays occurred inside the helium bag. At 18 metres away from the block, two spectrometers composed of spark chambers were used to detect the final products in the beam. Cherenkov and scintillation counter were situated immediately after the spectrometers for detection. It was found that all of the K_S^0 had decayed so only K_L^0 and neutrons were present in the beam. From angular distributions of the decay products, a peak at $\theta = 0$ degrees was observed. The peak corresponded to products with an invariant mass range of $494 < m < 504$ GeV/c^2. Thus, neutral kaons were observed in the system, providing evidence for CP violation [5].

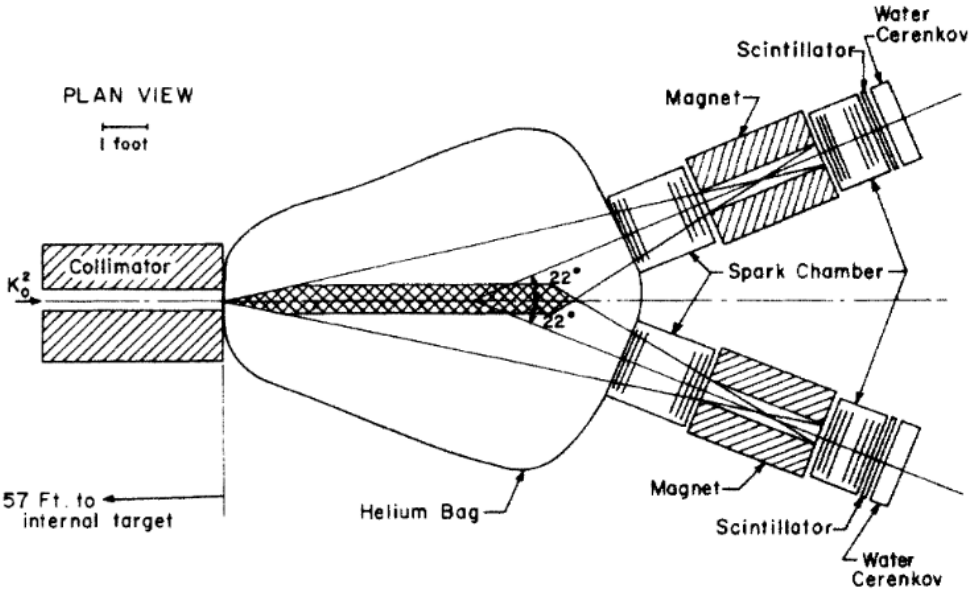


Figure 11: The experimental set-up for the kaon study conducted by Christenson *et al.* [18].

2.3.4 Force Carrier for the Gravitational Force

The SM does not contain an observed force carrier for the gravitational force. However, the existence of a proposed force carrier, called the Randall-Sundrum (RS) graviton [19] has been suggested. The RS graviton is a hypothetical spin-2 excitation of the gravitatonal field. In addition to mediating the force of gravity, it also provides a solution to the hierarchy problem. The hierarchy problem refers to how quantum loops in the Higgs boson propagator contribute to the Higgs mass. At the Grand Unified Theory ($\Lambda_{GUT} \approx 10^{16}$ GeV) and Plank energy scales ($\Lambda_P \approx 10^{19}$ GeV), this would result in the predicted Higgs mass being much larger than the electroweak scale of 10^2 GeV. A potential solution was provided in the RS model, in which the RS graviton exists on the five-dimensional "Planck" brane while the other SM particles are confined to the four-dimensional "TeV" or "SM" brane; a schematic representing the branes is shown in Figure 12. As the RS graviton travels from the Planck brane to the TeV brane, its mass exponentially decreases. This results in the Higgs mass being at the electroweak scale. It also explains why the strength of the gravitational force is much weaker than the other forces, since the gravitional force loses strength while it travels from the Planck brane to the TeV brane [11].

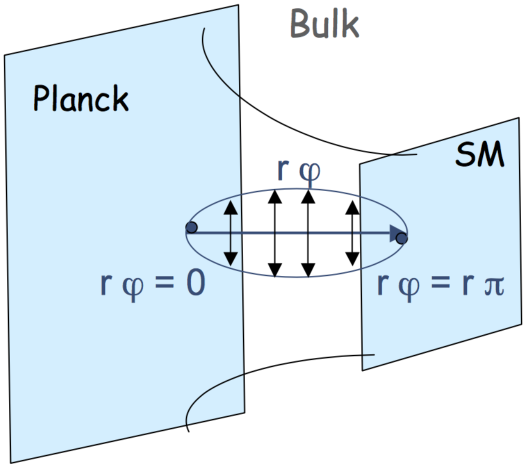


Figure 12: The Planck and SM branes that the RS Graviton travels between [20].

3 The Top Quark and the Discovery of the Top Quark

The top quark was discovered in 1995 at the Tevatron at Fermilab. Its existence was anticipated to complete the three generations of quark. With a mass of 173.1 ± 0.6 GeV [21], it is the heaviest of the elementary particles. It has a lifetime of 0.5×10^{-24} s [22], so decays before hadronisation can occur. The top quark is the only quark to have a mass greater than that of the W boson. It is therefore the only quark that is able to decay to produce a W boson and a b-jet. The CKM matrix is shown in Equations 71 and 70 and provides the relationship between the weak eigenstates, denoted by d', s' and b', and the mass eigenstates, denoted by d, s and b. This relation is shown in Equation 70 [12]. Therefore, top quark studies can be used to test the values of the $|V_{tb}|$ CKM matrix element.

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{70}$$

$$\begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} = \begin{pmatrix} 0.97428 & 0.2253 & 0.00347 \\ 0.2252 & 0.97345 & 0.0410 \\ 0.00862 & 0.0403 & 0.99915 \end{pmatrix} \tag{71}$$

Top quarks can either be produced in $t\bar{t}$ pairs or can be produced singly. They have been studied using the ATLAS and CMS detectors at the Large Hadron Collider (LHC), and the CDF and D0 experiments at the Tevatron. The structure of the LHC will be discussed in Section 4. The Tevatron was a synchrotron at the Fermi National Accelerator Laboratory (Fermilab), which was in operation between 1983 and 2011. It had a circumference of 6.28 km and collided protons with anti-protons ($p\bar{p}$ collisions) at centre of mass energies ($\sqrt{s}$) up to 1 TeV. The Tevatron contained two particle detectors: the Collider Detector at Fermilab (CDF) and D0.

At the LHC, for 90% of the top quark production processes at $\sqrt{s} = 14$ TeV, top quarks are produced in $t\bar{t}$ pairs via gluon-gluon scattering. Alternatively, 85% of its production at the Tevatron was via quark-antiquark scattering [21]. The production of top quark pairs and single top production will be further discussed in sub-sections 3.1 and 3.4, respectively.

3.1 Top Quark Properties

3.1.1 Mass measurements

The latest average of the top quark measurements recorded from LHC and Tevatron runs is 172.69 ± 0.30 GeV [23].

3.1.2 Top Quark Pair Cross Section Measurements

The production of top quark pairs is the dominant top quark production mode at the LHC. Cross section measurements for the final states of $t\bar{t}$ production (dilepton, lepton+jets and all jets) have been conducted by the CDF, D0, ATLAS and CMS collaborations.

The latest measurement of the $t\bar{t}$ cross section published by the CMS collaboration was in 2018 [24]. It was recorded to be 803 ± 2 (stat.) ± 25 (syst.) ± 20 (lumi.) pb, for events in dileptonic final states. Events corresponded to a centre of mass energy of 13 TeV for data at an integrated luminosity of $35.9 fb^{-1}$.

The latest measurement of the $t\bar{t}$ inclusive cross section published by the ATLAS collaboration was in 2020 [25]. It was recorded to be 830 ± 0.4 (stat.) ± 36 (syst.) ± 14 (lumi.) pb, for events in lepton+jets final states. Events corresponded to a centre of mass energy of 13 TeV for data at an integrated luminosity of $139 fb^{-1}$.

3.1.3 Measurements of Top Quark Pair Charge Asymmetry

The measurements of the angular distributions between top quarks and antitop quarks, referred to as the charge asymmetry, have been conducted by the CDF, D0, ATLAS and CMS collaborations. The latest combined measurement of the charge asymmetry conducted by the ATLAS and CMS collaborations was recorded to be $0.55 \pm 0.23 \pm 0.25$ for collisions at a centre of mass energy of 8 TeV [23] .

3.1.4 Measurements of the Top Quark Pair Mass Difference

In the SM, the CPT transformation predicts a mass difference of zero between particles and their antiparticles. As of May 30, 2024, there have been two measurements of the top quark pair mass difference. The first was published in 2012 [26], for pp collisions at $\sqrt{s} = 7$ TeV. Events were selected with a muon or electron and at least 4 jets in the final state. The data corresponded to an integrated luminosity of $4.96 \pm 0.11\ fb^{-1}$ and a mass difference value of $\Delta m = -0.44 \pm 0.46$ (stat.) ± 0.27 (syst.) GeV was measured.

The second analysis was published in 2016 [27] for pp collisions at $\sqrt{s} = 8$ TeV. Events were selected with a muon or electron and at least 4 jets in the final state. The data corresponded to an integrated luminosity of $19.6\ fb^{-1}$ and a mass difference value of $\Delta m = -0.15 \pm 0.19$ (stat.) ± 0.09 (syst.) GeV was measured.

3.1.5 W Helicity Measurements

Helicity is defined as the normalised component of spin projected along the direction of motion of the particle. The mathematical definition is shown in Equation 72, where h is the helicity, $\boldsymbol{S}$ represents spin and $\boldsymbol{p}$ represents the momentum of the particle.

$$h \equiv \frac{\boldsymbol{S} \cdot \boldsymbol{p}}{p} \tag{72}$$

Helicity measurements of the W boson are of interest to research by the particle physics community because they are related to the V-A structure of the weak charged current. The measurements are also sensitive to BSM physics. Helicity measurements can be obtained through the study of angular distributions of $t\bar{t}$ decay products.

The latest combination result for measurements of the W boson helicity conducted by the ATLAS and CMS collaborations was published in 2020 and corresponded to proton-proton collisions at a centre-of-mass energy of 8 TeV. Results were quoted as fractions of W bosons with longitudinal (F_0), left-handed (F_L), or right-handed (F_R) polarisations, which were found to be $F_0 = 0.693 \pm 0.014$ and $F_L = 0.315 \pm 0.011$ and $F_R = 0.008 \pm 0.007$, respectively [28].

3.1.6 Spin Correlations and Top Quark Polarisation Measurements

The lifetime of the top quark is $3.29^{+0.90}_{-0.63} \times 10^{-25}$ s, which is lower than the timescale required for spin decorrelation $m_t/\Lambda^2_{QCD} \approx 10^{-21}$ s. Spin decorrelation is when there is no longer a correlation between the spin of the parent particle and the spins of its decay products. Therefore, the top quark can pass on its spin information to its daughter particles in decays. This information is accessed by analysing angular distributions of the decay products.

The top quark polarisation P is calculated using $P = 2A_P$, where A_P is the asymmetry variable shown in Equation 73. In this, θ^*_l is the angle of a charged lepton in the rest frame of its parent top quark or antiquark, measured in the helicity frame. The number of events counted is represented by N.

$$A_P = \frac{N[cos\theta_l^* > 0] - N[cos\theta_l^* < 0]}{N[cos\theta_l^* > 0] + N[cos\theta_l^* < 0]} \tag{73}$$

To discriminate between correlated and uncorrelated spins, the variable $A_{\Delta\phi}$ can be obtained by using Equation 74.

$$A_{\Delta\phi} = \frac{N[\Delta\phi_{l^+l^-} > \pi/2] - N[\Delta\phi_{l^+l^-} < \pi/2]}{N[\Delta\phi_{l^+l^-} > \pi/2] + N[\Delta\phi_{l^+l^-} < \pi/2]} \tag{74}$$

The spin correlation coefficient C_{hel} can be determined by calculating $C_{hel} = -4A_{c_1}c_2$. The mathematical definition of the $A_{c_1}c_2$ variable is shown in Equation 75, in which $c_1 = cos(\theta_{l^+}^*)$ and $c_2 = cos(\theta_{l^-}^*)$.

$$A_{c_1}c_2 = \frac{N[c_1c_2 > 0] - N[c_1c_2 < 0]}{N[c_1c_2 > 0] + N[c_1c_2 < 0]} \tag{75}$$

Latest measurements of the top quark polarization and top quark pair spin correlations were published by the CMS collaboration in 2019. Events were recorded in 2016 and corresponded to an integrated luminosity of 35.9 fb^1, contained two oppositely-charged leptons produced in proton-proton collisions at a centre-of-mass energy of 13 TeV. The normalized differential cross sections were used to constrain the anomalous chromomagnetic and chromoelectric dipole moments of the top quark to $-0.24 < C_{tG}/\Lambda^2 < 0.07$ TeV^2 and $0.33 < C_{ItG}/\Lambda^2 < 0.20$TeV^2, respectively, at the 95% confidence level [29].

3.1.7 Top Quark Decay Width And Ratio Of Branching Fractions

The ratio of the top quark branching fractions is represented by $R = \frac{t \to Wb}{t \to Wq}$, where q is a down-type quark (d, s or b) and the denominator is summed over all q. The latest average result was recorded as 0.957 ± 0.034 by the Particle Data Group [23].

The top quark decay width is represented by Γ and the latest average was recorded as $1.42\ ^{+0.19}_{-0.15}$ by the Particle Data Group [23].

3.2 Four Top Production

The top quark can be produced via $t\bar{t}t\bar{t}$ production. The $t\bar{t}t\bar{t}$ production Feynman diagrams are shown in Figure 13. This is a rare top quark production mechanism that, at LO, has a SM-predicted cross section of ≈ 1 fb at $\sqrt{s} = 8$ TeV. Its dominant background is $t\bar{t}$ production.

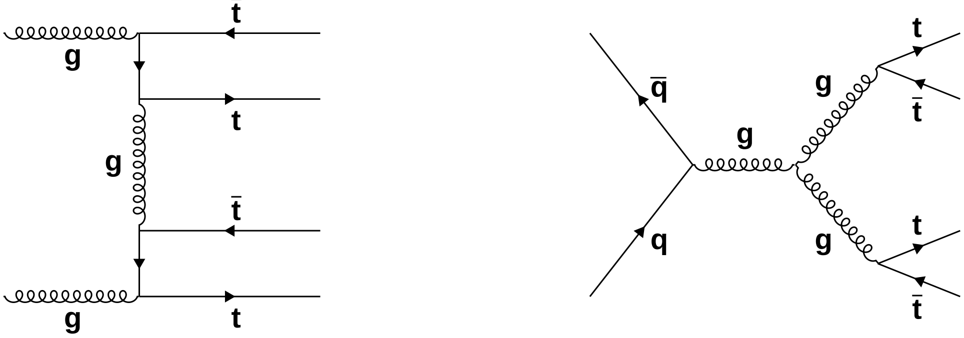


Figure 13: The leading-order Feynman diagrams for $t\bar{t}t\bar{t}$ production via gluon fusion (left) and quark-antiquark annihilation (right) [30].

Results of the latest search for four top production by the CMS collaboration were published in 2019. Same sign or multi-lepton final states were considered for proton-proton collisions at a centre-of-mass energy of 13 TeV, corresponding to an integrated luminosity of 137 fb^{-1}. The observed and expected significances were found to be 2.6 and 2.7 standard deviations, respectively, and the $t\bar{t}t\bar{t}$ cross section was measured to be $12.6^{+5.8}_{-5.2}$ fb [31].

3.3 Charge-Parity Violation in the Top Quark Sector

There has been one reported analysis by the CMS collaboration of charge-parity violation (CPV) in the top quark sector [32]. Events were analysed from the decay of $t\bar{t}$ pairs from pp collisions at $\sqrt{s} =$ 8 TeV. The data corresponded to an integrated luminosity of 19.7 fb^{-1}. Events were required to have at least one electron or muon and four jets in the final state.

3.4 Single Top Quark Production

There are three general single-top production mechanisms. These are s-channel, t-channel and associated tW production. Feynman diagrams for these are shown in Figure 14.

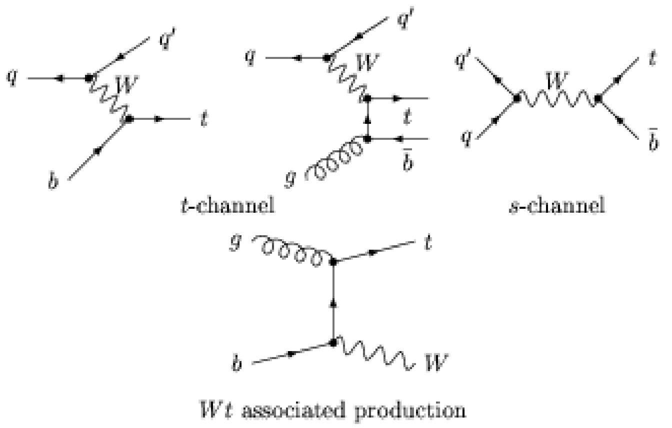


Figure 14: The single top production channels [33].

The first observations for electroweak single top production in the t-channel were reported in 2009 by the D0 and CDF collaborations [34] [35]. The D0 collaboration reported an observed signal significance of 5.0 standard deviations after analysing data from $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV, corresponding to an integrated luminosity of 2.3 fb^{-1}. The cross section was measured to be $\sigma(p\bar{p} \to tb + X, tqb + X) = 3.94 \pm 0.88$ pb. The CDF collaboration also reported an observed signal significance of 5.0 standard deviations, after analysing data from $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. The data corresponded to an integrated luminosity of 3.2 fb^{-1} and the measured cross section was determined to be $2.3^{+0.6}_{-0.5}$ (stat+syst) pb. A value of 0.91 ± 0.11 (stat+syst) 0.07 (theory) was obtained for the $|V_{tb}|$ CKM matrix element.

3.4.1 Single Top Production in the s-channel

In 2013, the D0 collaboration reported evidence for single top production in the s-channel [36]. For $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV that corresponded to an integrated luminosity of $9.7 fb^{-1}$, the collaboration reported a measured cross section of $\sigma(p\bar{p} \to tb{+}X) = 1.10^{+0.33}_{-0.31}$ pb. This corresponded to a significance of 3.7 standard deviations.

The first observation of the s-channel was announced in 2014 after results from the D0 and CDF collaborations were combined [37]. Data from $p\bar{p}$ collisions was analysed at a centre of mass energy of 1.96 TeV, corresponding to integrated luminosities of $9.7 fb^{-1}$ per experiment. The cross section was measured as $\sigma = 1.29^{+0.26}_{-0.24}$ pb and the observed signal significance was found to be 6.3 standard deviations.

As of the time of writing, the most recent cross section measurement for the s-channel was reported in a combination of ATLAS and CMS measurements for pp collisions at $\sqrt{s} = 8$ TeV. The measured cross section was 4.9 ± 1.4 pb [39].

3.4.2 Single Top Production in the tW channel

No searches for single top production in the tW-channel were conducted by the CDF and DZero collaborations since its cross section is negligible at a centre-of-mass energy of 1.96 TeV. In 2014, the CMS collaboration reported an observation for the associated production of a single top quark and W boson in pp collisions at $\sqrt{s} = 7$ TeV [40]. The data corresponded to an integrated luminosity of 12.2 fb^{-1}. The production cross section was measured to be 23.4 ± 5.4 pb and the observed signal significance was 6.1 standard deviations. In 2018, a measured cross section of $63.1 \pm 1.8(stat) \pm 6.3(sys) \pm 2.1(lumi)$ pb was obtained from analysing data corresponding to pp collisions at $\sqrt{s} = 13$ TeV and an integrated luminosity of 35.9 fb^{-1} [41].

Finally, in 2019, the combinations of s-channel measurements by ATLAS and CMS at $\sqrt{s} = 7$ and 8 TeV were reported [42]. The combined measurements for the cross secion were 16.3 ± 4.1 pb at $\sqrt{s} = 7$ TeV and 23.1 ± 3.6 pb at $\sqrt{s} = 8$ TeV.

3.4.3 Single Top Production in the t-channel

An example of a single top quark production process for the t-channel is tZq SM production. The Feynman diagrams for leading order tZq production are shown in the five diagrams of Figure 15 [43]. In this, a quark scatters off of a bottom quark, exchanging a W boson. This leads to either the scattered quark radiating a Z boson (shown in the first two diagrams of Figure 15), the Z boson being radiated from the incoming bottom quark or outgoing top quark (shown in the third and fourth diagrams, respectively) or the incoming quarks each radiating a W boson which fuse to form a Z (shown in the fifth diagram). This rare SM process is of interest to study since it is related to WZ production, so cross section measurements provide a test for the Standard Model [43].

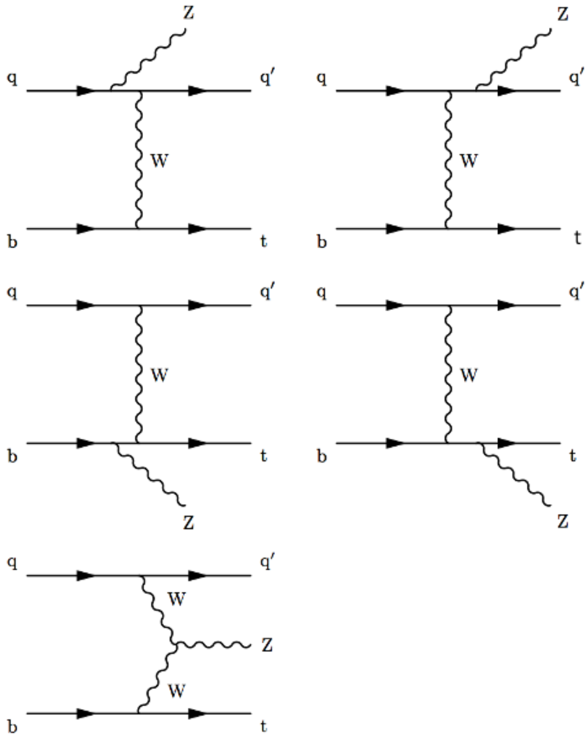


Figure 15: The Feynman diagrams for tZq production [43].

There are various final states for tZq production, whose Feynman diagrams are shown in Figure 16. The outgoing quark from the scatter can be either an up quark, a down quark, a charm quark or a strange quark. The top quark in the process can decay to a b-jet and a W boson. In the case that both the W and Z bosons decay to form quark-antiquark pairs, this is called the hadronic final state. If the W decays into a lepton and neutrino while the Z boson decays into a quark and an antiquark, this is called the single lepton final state. However, if the Z decays instead into a lepton-antilepton pair, this is called the trilepton final state. Alternatively, if the W instead decays into a quark-antiquark pair and the Z decays into a lepton-antilepton pair, this is called the dilepton final state. The quark-antiquark pairs in all final states consist of an up-type quark and a down-type quark. A search for production of tZq in dilepton final states is the main topic of this thesis. The methodology for the search will will be discussed in Section 9. The tZq production mechanism in dilepton final states is of particular interest since it is a rare SM process and also forms a background for other processes (SM tH production and BSM FCNCs).

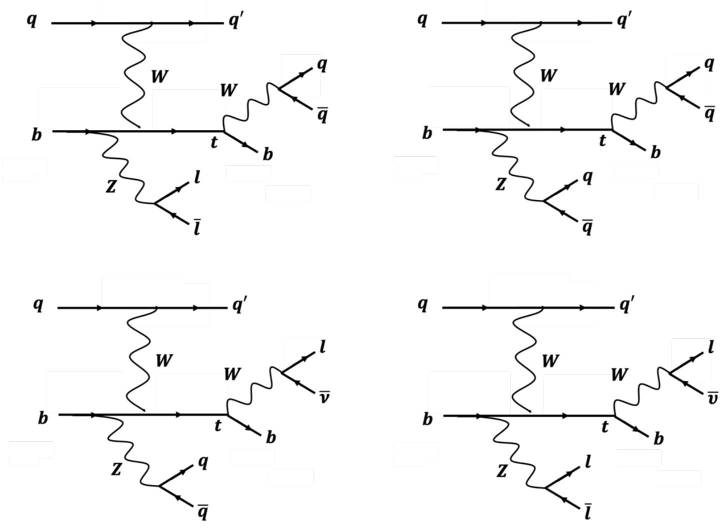


Figure 16: Feynman diagrams for all tZq final states. These are: the dilepton final state (top-left), the hadronic final state (top-right), the single lepton final state (bottom-left) and the trilepton final state (bottom-right).

3.5 FCNC production

Additionally, tZq production can occur via flavour changing neutral current (FCNC) channels. These are non-SM processes and the diagrams for FCNC single top production are shown in Figure 17. In the SM, the charged current interaction is the only flavour changing process. The SM forbids FCNC processes at tree level as these require a loop for virtual W boson exchange. The GIM mechanism suppresses higher orders. Therefore, evidence for FCNCs would indicate the presence of physics beyond the Standard Model.

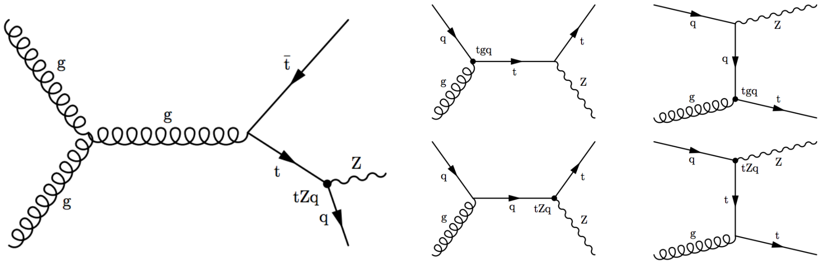


Figure 17: Left: The tZq-FCNC channel. Right: The tZ-FCNC channel [43].

In nearly 100% of events, the top quark is predicted to decay to a W boson and a b quark ($t \rightarrow Wb$). However, it may also decay into a neutral Z boson and an up or charm quark ($t \rightarrow Zq$, where q = u or c). This decay channel has a predicted branching ratio $\sim 10^{-14}$. It is forbidden by the SM at tree level as it requires the presence of quantum loop corrections. It is suppressed at higher orders by the Glashow–Iliopoulos–Maiani (GIM) mechanism. Beyond the SM (BSM) models (e.g. R parity-violating supersymmetric models) predict a branching ratio $\sim 10^{-4}$, which has a required sensitivity beyond the scope of current detection technology at the LHC. The $t \rightarrow Zq$ channel is predicted in FCNCs so an observation of FCNCs at the LHC would therefore be indicative of BSM physics.

The most recent results for upper limits for FCNC branching fractions were published by the CMS collaboration in 2021. A search was conducted for flavor-changing neutral current interactions of the top quark and the Higgs boson. Events in lepton+jets final states were considered, corresponding to proton-proton collisions at 13 TeV at an integrated luminosity of 137 fb^{-1}. The observed (expected) upper limits at 95% confidence level were reported to be $B(t \rightarrow Hu) < 0.079(0.11)\%$ and $B(t \rightarrow Hc) < 0.094(0.086)\%$ [44].

3.6 Backgrounds to tZq Production Processes

The backgrounds associated with tZq production in the dilepton final state include top pair production ($t\bar{t}$, $t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}H$), boson+jets ($Z/\gamma * +jets$, $W + jets$), dibosonic processes (WW, WZ, ZZ, tWZ) and tribosonic processes (WWW, WWZ, WZZ, ZZZ) [45]. However, the dominant backgrounds are $t\bar{t}$ and $Z + jets$ [46]. The backgrounds to tZq production in the trilepton final state contain at least one top quark or more than two gauge bosons, either at leading order or next to leading order. Backgrounds include the production of tHW, tHq, tWZ, ttV, ttH, ttVV and VVV (where V = W or Z, and H is the Higgs boson) [43].

Additionally, non-prompt lepton (NPL) background processes also contribute to the above three processes. A non-prompt lepton is a lepton produced from an event containing either a jet that has been incorrectly reconstructed as a lepton, or it contains a lepton originating from the decay of heavy quarks. Such events mainly arise from semi-leptonic $t\bar{t}$ decays and $W + jets$ production. Single top and QCD production provide smaller contributions to the NPL background.

4 The Large Hadron Collider

4.1 An overview of the Large Hadron Collider

The Large Hadron Collider (LHC) is a particle accelerator at CERN, which is near Geneva, Switzerland. It was built using the tunnels of its predecessor, the Large Electron-Position Collider (LEP), which collided electrons and positrons at a centre of mass energy of 209 GeV. Unlike LEP, the LHC is a proton-proton collider. The decision was made to collide protons rather than electrons and positrons for many reasons. For example, accelerating protons leads to considerably less synchrotron radiation being emitted. Synchrotron radiation is emitted when a charged particle is accelerated and is represented by Equation 76. In this equation, the amount of energy lost in the form of synchrotron radiation is represented by ΔE. The particle charge is given by q and β and γ are Lorentz indices ($\gamma = 1/\sqrt{1-\beta^2}$ and $\beta = v/c$, where v is the particle velocity and c is the speed of light). The permittivity of free space is represented by ϵ_0 and ρ is the radius of curvature of the orbit [5].

$$\Delta E = \frac{4\pi q^2 \beta^3 \gamma^4}{3\epsilon_0 \rho} \tag{76}$$

For relativistic particles, $\beta \ \sim 1)$ and $\gamma = \frac{E}{mc^2}$. Therefore $\Delta E \propto 1/m^4$. Therefore the energy loss is inversely proportional to m^4 and means that an electron circulating the LHC would lose 10^{13} times more energy in the form of synchrotron radiation than a proton.

Another benefit to using proton beams is protons are much heavier than the electron so much higher energies can be achieved using a proton-proton collider. However, disadvantages with proton-proton colliders include the collisions not being as clean. This is because protons consist of quarks and gluons, which collide with other quarks and gluons in collisions, forming jets.

Before entering the LHC, the protons pass through an injector chain. They begin as a bottle of hydrogen gas and then pass through Linac2. The protons then travel to the Proton Synchrotron Booster (PSB). In both Linac2 and the PSB, radiofrequency cavities are used to accelerate the protons. They then enter the Proton Synchrotron (PS) and finally the Super Proton Synchrotron (SPS) [47].

With a circumference of 27 kilometres, the LHC is the world's largest particle accelerator and produces the most energetic collisions. Proton bunches are made to collide, where 10^{11} protons are contained in a single bunch and 2808 bunches circulate the LHC. These bunches have a 25 ns spacing. Collisions occur at four beam-crossing points, at each a detector is situated. The detectors are: A Toroidal LHC Apparatus (ATLAS) [48], A Large Ion Collider Experiment (ALICE) [49], Compact Muon Solenoid (CMS) [50] and LHC-beauty (LHCb) [51]. The TOTal cross section, Elastic scattering and diffraction dissociation Measurement at the LHC (TOTEM) detector [53] and the LHC forward (LHCf) detector [54] are also situated on the LHC ring. TOTEM is used to obtain precise measurements of the proton-proton cross section while LHCf uses forward particles to simulate cosmic rays.

Data recorded by the CMS Experiment will be analysed for the research conducted for this thesis. It is located 100 metres below the village of Cessy in France and is situated at point 5 of the Large Hadron Collider. This is shown in Figure 18.

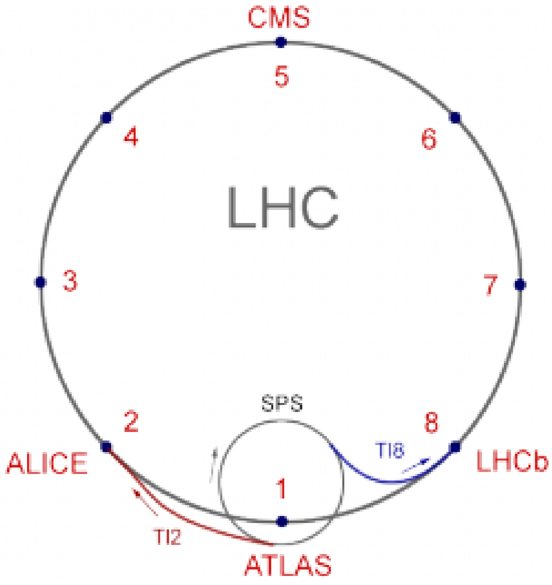


Figure 18: Points on the LHC ring [55].

4.2 The Compact Muon Solenoid Experiment

The CMS detector is a general-purpose detector, meaning it is used for a range of different studies. Such studies include searches for dark matter candidates and the study of the Higgs boson. Studies are conducted to test the Standard Model, which was discussed in Section 2.2. The name Compact Muon Solenoid arises firstly from the detector being compact. It contains 14,000 tonnes of material while having a length of 21.6 metres and a diameter of 15 metres [56]. The CMS detector comprises layers and its outermost layer gives a signal when a muon has passed through. Finally, the detector makes use of a solenoid magnet to bend charged particles under the Lorentz force. The structure of the CMS detector and its subsystems will be explained in Sections 4.2.1-4.2.6.

4.2.1 The CMS Coordinate System

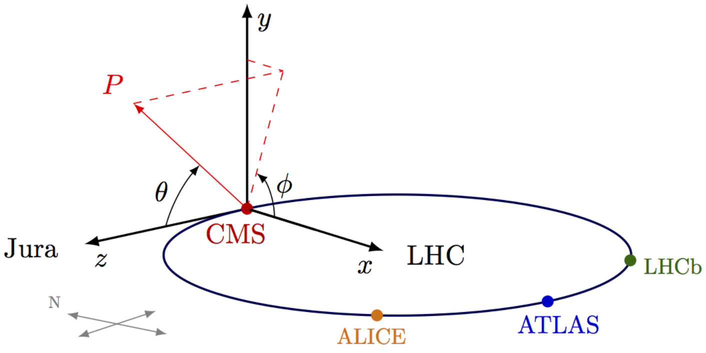


Figure 19: A schematic of the CMS coordinate system.

A diagram of cylindrical coordinate system used in the CMS detector is shown in Figure 19. In this coordinate system, the z-axis points in the direction of the beam pipe, which is towards the Jura mountains from Point 5. The x-axis points from the CMS experiment in the direction of the centre of the LHC. Finally, the y-axis is perpendicular to both the x and z axes and points vertically upward.

Two angles arise from these axes: the polar angle, θ, which is measured from the z-axis, and the azimuthal angle, ϕ, which is measured from the x-axis and towards a physics object in the $x-y$ plane. From this coordinate system, the rapidity variable can be calculated. The definition for rapidity is shown in Equation 77.

$$y = \frac{1}{2} ln(\frac{E + p_Z}{E - p_Z}) \tag{77}$$

A rapidity value of zero corresponds to a particle produced at the centre of the z-axis. For a highly-relativistic particle with $E >> m$, $p_Z \simeq Ecos(\theta)$. Substituting this into Equation 77, using the trigonometric relation $\frac{1-cos(\theta)}{(1+cos(\theta))} = tan^2(\frac{\theta}{2})$ and letting $y \sim \eta$ obtains the definition of pseudorapidity, η. This definition is shown in Equation 78.

$$\eta = -ln(\tan\left(\frac{\theta}{2}\right)) \tag{78}$$

The pseudorapidity is zero for particles that are traveling perpendicular to the beam ($\theta = \pi/2$), and it approaches infinity for angles closer to the beam. In CMS, the sign of η is chosen to be positive for the side of the detector closer to the Jura. Each subdetector of the CMS detector spans a range of pseudorapidity values and the number of particles in each unit of η is constant. These values are shown in Table 2.

Region	**Pseudorapidity**
Tracking System	$\|\eta\| < 2.4$
ECAL Barrel (EB)	$0 < \|\eta\| < 1.479$
ECAL Endcaps (EE)	$1.479 < \|\eta\| < 3.0$
Hadron Barrel (HB)	$\|\eta\| < 1.4$
Hadron Outer (HO)	$\|\eta\| < 1.26$
Hadron Endcap (HE)	$1.3 < \|\eta\| < 3.0$
Hadron Forward (HF)	$3.0 < \|\eta\| < 5.0$
Muon System (barrel region)	$\|\eta\| < 1.2$
Muon System (endcaps)	$\|\eta\| < 2.4$

Table 2: The variation of η values for the different CMS detector regions.

A schematic showing a tranverse slice of the CMS detector is shown in Figure 20. The structures and functions of its sub-detectors will be further explained in Sections 4.2.2-4.2.6.

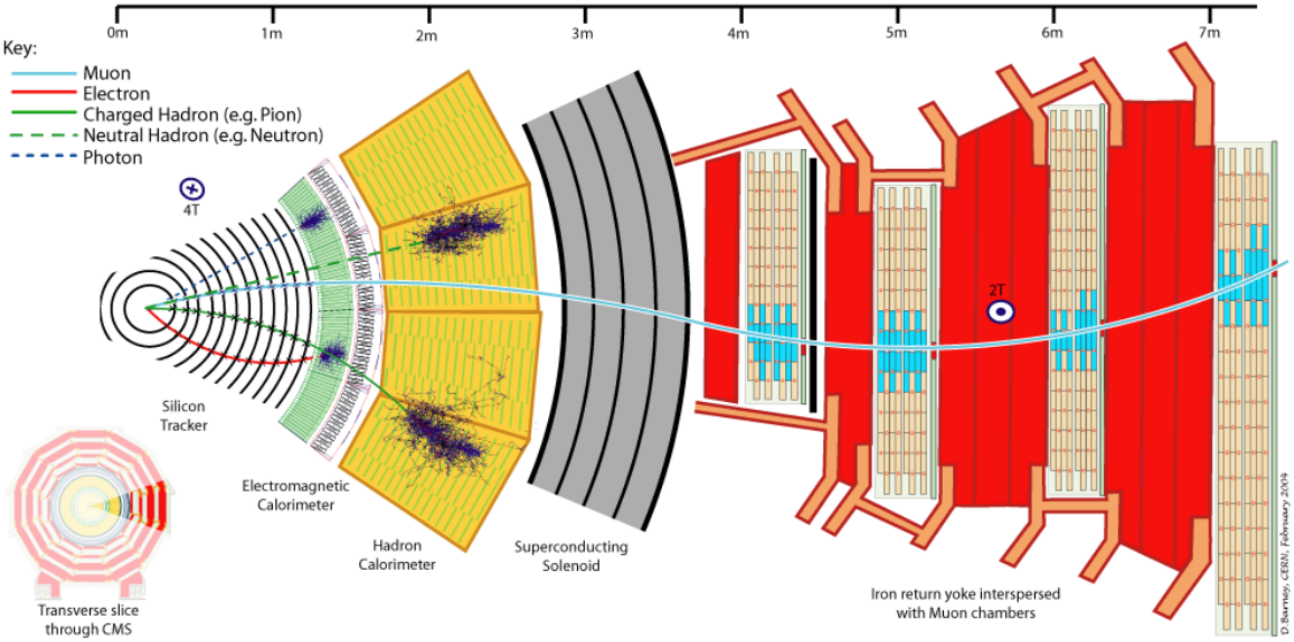


Figure 20: A schematic showing the different CMS sub-detectors [57].

4.2.2 Tracking System

Figure 20 [57] shows a transverse slice of the CMS detector, with its subdetectors. The far-left of the image is the beam pipe, inside which the collisions occur.

The region of the CMS detector that is closest to the beam pipe is the inner tracking system. Charged particles, such as electrons, leave hits in the tracker while neutral particles, such as photons, do not. It is essential to measure and detect the momenta of charged particles for identification. Ionisation occurs when a charged particle traverses a medium, producing a trail of ionised atoms and electrons that have been liberated from the atoms. The detection of the ionisation enables the reconstruction of the charged particle trajectory, from which a momentum measurement is obtained. Two main tracking detector methods are used in the CMS. The first is the use of a large gaseous tracking volume, inside which a traversing charged particle ionises the gaseous atoms and the liberated electrons drift towards an anode wire. A signal is recorded at the anode wire and the track is reconstructed. Alternatively, semiconductor strips or pixels can be used for tracking. This is implemented in the ATLAS and CMS detectors. Ionisation occurs when a charged particle traverses a doped silicon wafer, which produces $\sim$ 10,000 electron-hole pairs. The potential difference that is applied across the silicon results in the electrons and holes drifting along the electric field lines until they are collected at p-n junctions. Amplification electronics are then used to obtain a clear signal, called a 'hit'. The hits from each strip or pixel are used to reconstruct the charged particle's trajectory. The tracking system is located inside the superconducting solenoid magnet, which will be discussed in detail in Section 4.2.5. The magnet generates a uniform magnetic field in the direction of the beam axis, so particles in the tracker are subjected to the Lorentz force, F. An expression for the Lorentz force is shown in Equation 79, where q represents the charge of the particle, $\vec{v}$ represents the velocity and B represents the magnetic field strength.

$$\vec{F} = q(\vec{v} \times B) \tag{79}$$

A charged particle in an axial magnetic field will follow a helical trajectory, from which measurements of the radius R and the pitch angle λ can be obtained. The momentum, p, can then be determined by using the relation shown in Equation 80.

$$p = \frac{0.3BR}{cos(\lambda)} \tag{80}$$

In Equation 80, B represents the magnetic flux density (in Tesla), p is measured in units of GeV/c, R is in metres and λ is in degrees. The momentum is proportional to the radius so, for charged particles with high momenta, the tracks appear to be almost straight [11].

The CMS tracking system is cylindrical in shape, with a length of 5.8 metres and a diameter of 2.6 metres. It is divided into two subdetectors: the strip tracker, comprising of silicon microstrip detectors, and the pixel tracker, comprising of silicon hybrid pixel detectors.

The barrel region of the strip tracker is divided into two regions: the Tracker Inner Barrel (TIB) and the Tracker Outer Barrel (TIB). The endcap regions of the strip tracker are also divided into two regions: the Tracker End Cap (TEC) and the Tracker Inner Disks (TID).

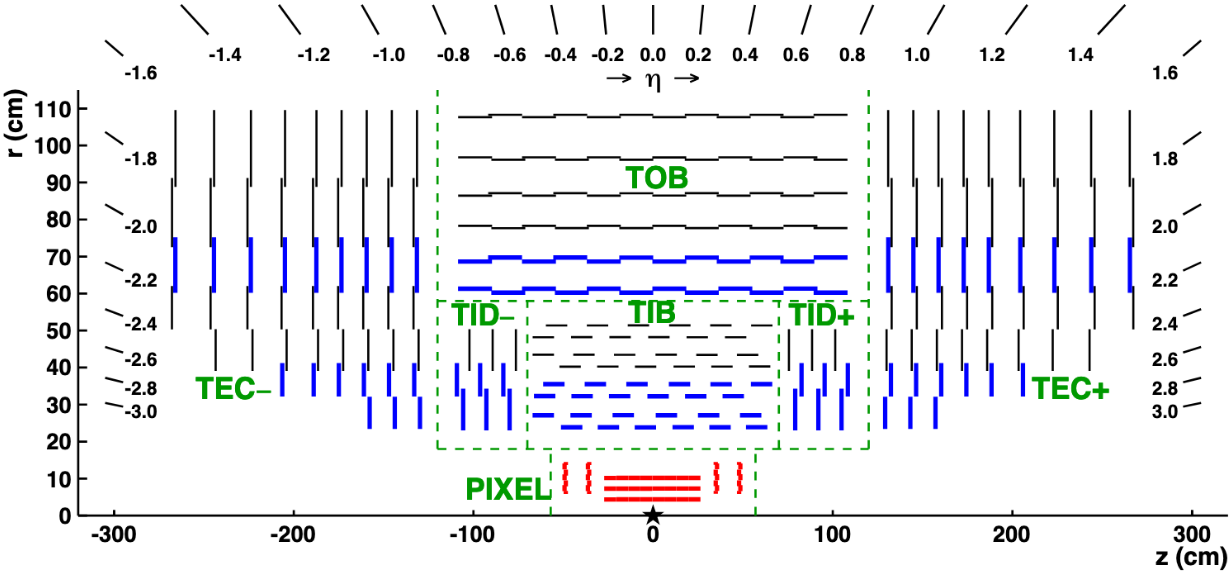


Figure 21: A schematic showing the layout of the CMS tracking system [58].

The pixel tracker also consists of a barrel region and endcap regions. In the barrel region, there were originally three layers of silicon pixel detectors but this has been increased to four layers during the Extended Year End Technical Stop (EYETS) of 2016/2017. The layers are 53 cm long and are situated close to the interaction region, at radii of 4.4cm, 7.3 cm and 10.2 cm. The close proximity improves the measurement of the impact parameter of the tracks, defined as the distance between the track and the primary vertex at closest approach. The close proximity also improves the measurement of the position of secondary vertices. Outside of the pixel tracker barrel region are two endcap disks. B-tagging algorithms are used to identify jets that originate from bottom quarks in the tracker. Such jets are referred to as ‘b-jets’. The hadrons that decay into bottom quarks have lifetimes lower than that of light quark hadrons, resulting in them decaying inside the silicon tracker. The algorithms are discussed in Section 5.

4.2.3 The Electromagnetic Calorimeter

Particles that undergo electromagnetic interactions such as electrons and photons leave energy deposits in the Electromagnetic Calorimeter (ECAL). Here X_0 represents the radiation length, which is defined as the average distance taken for the electron energy to decrease by 1/e, via the bremsstrahlung energy loss mechanism. The relationship between X_0 and the atomic number of the absorbing material Z is shown in Equation 81.

$$X_0 \approx \frac{1}{4\pi\alpha n Z^2 r_e^2 ln(287/Z^{1/2})} \tag{81}$$

In Equation 81, r_e represents the critical radius of the electron ($r_e = e^2/4\pi\epsilon_0 m_e c^2 = 2.85 \times 10^{-15}$ m), n represents the density of the atomic nuclei and α represents the EM coupling strength. Therefore, materials with a high Z have a short X_0.

The CMS ECAL can detect the products of electromagnetic showers up to 25 X_0. The EM showers are initiated when an electron interacts with a medium and loses energy via bremsstrahlung radiation. A bremsstrahlung photon is produced which goes on to produce an electron-positron pair via pair production. This process continues such that the number of particles contained in the electromagnetic shower doubles with every radiation length travelled. An EM shower is depicted in Figure 22 [11].

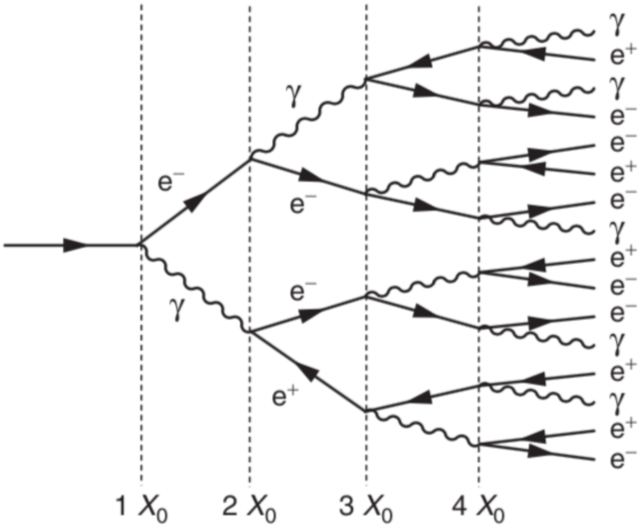


Figure 22: A diagram showing how the number of particles contained in an EM shower varies with radiation length [11].

The EM shower progression leads to the relationship between the average energy of the particles in the EM shower $< E >$ and the number of radiation lengths traversed x, shown in Equation 82.

$$< E > \approx \frac{E}{2^x} \tag{82}$$

Therefore, the high the number of radiation lengths traversed, the lower the average energy of the EM shower particles. After a maximum value of x, which is represented by x_{max}, the $< E >$ will become less than the critical energy E_c. At the critical energy, ionisation becomes the dominant energy loss mechanism over bremsstrahlung. The value of x_{max} can be calculated by using Equation 83 [11].

$$x_{max} = \frac{ln(E/E_c)}{ln(2)} \tag{83}$$

Detection is made by the use of lead tungstate crystals ($PbWO_4$), which is an example of an inorganic scintillator. When an incident particle is absorbed, the electrons of the scintillator's atoms enter an excited state. They return to their ground state by emitting photons in the UV region. Silicon avalanche photodiodes (APDs) are used to detect scintillation light in the barrel region of the detector, while vacuum phototriodes (VPTs) detect scintillation light in the endcaps.

In front of the ECAL endcaps there is a preshower system. This is used to distinguish between the photons produced from pions and the photons produced from other particles, such as the Higgs boson.

4.2.4 The Hadron Calorimeter

The products of hadronic showers are deposited in the subdetector that surrounds the ECAL, which is called the Hadron Calorimeter (HCAL). A hadronic shower is produced when a relativistic hadron strongly interacts with the nuclei of the medium it traverses, which causes the hadron to lose energy via ionisation. After the primary interaction, the hadron undergoes a secondary interaction, producing a shower of particles. The mean distance travelled between hadronic interactions is defined as the radiation length λ_1.

The hadronic shower also includes an electromagnetic component, due to pion decay into two photons. These photons go on to produce an electron-positron pair via pair production, which go on to emit bremsstrahlung photons. This was discussed in Section 4.2.3. The HCAL is used to detect the products of hadronic showers. It is a sampling calorimeter, where a sampling calorimeter is a type of calorimeter design that comprises layers of passive absorber material (e.g. lead or copper) alternating with layers of active detector material (e.g. silicon, liquid argon or plastic scintillator). The HCAL is composed of brass and plastic scintillator tiles. It comprises four different calorimeter regions: the Hadron Barrel (HB), Hadron Endcap (HE), Hadron Outer (HO) and Hadron Forward (HF) [56].

The ECAL or HCAL energy deposits, which are called clusters, are matched to the corresponding hits in the tracker using the Particle Flow (PF) algorithm [59]. This algorithm will be explained in Section 5.

4.2.5 The Magnet

Surrounding the calorimeters solenoid magnet, which runs at 3.8T and is used to bend the paths of the charged particles using the Lorentz force ($F = q\bar{v} \times \bar{B}$). The magnet is situated after the calorimeters because, in order to maximise the resolution of the particles' momenta measurements, either a larger magnetic field or a larger coil length must be used. This can be seen from Equation 84 [60]. In this equation, p represents momentum, $\frac{dp}{p}$ represents the resolution in a momentum measurement, B represents the magnetic field strength and l represents the coil length.

$$\frac{dp}{p} \propto \frac{p}{Bl^2} \tag{84}$$

4.2.6 The Muon System and the Return Yoke

Muons are highly-penetrating particles so are detected in the muon system, the outermost subdetector of the CMS detector. Ionisation is the dominant energy-loss process for muons with energies less than 100 GeV, so they can travel distances on the order of metres in dense materials. The rate of ionisation energy loss per unit length travelled by a singly charged particle, $\frac{dE}{dx}$, is given by the Bethe-Bloch equation. This is shown in Equation 85.

$$\frac{dE}{dx} \approx -4\pi\hbar c^2\alpha^2\frac{nZ}{m_e v^2}(ln[\frac{2\beta^2\gamma^2c^2m_e}{I_e}] - \beta^2) \tag{85}$$

In this equation, n is the number density of material being traversed, Z is the atomic number, v is the velocity where $v = \beta c$ and I_e is the effective ionisation potential of the material averaged over all atomic electrons ($I_e \sim 10ZeV$). If the number density is expressed instead as $n = \frac{\rho}{Am_u}$, where A is the atomic mass number and m_u is the unified atomic unit ($m_u = 1.66 \times 10^{-27}$ kg), Equation 85 can be re-written as Equation 86.

$$\frac{1}{\rho}\frac{dE}{dx} \approx \frac{-4\pi\hbar c^2\alpha^2}{m_e v^2 m_u}\frac{Z}{A}(ln[\frac{2\beta^2\gamma^2c^2m_e}{I_e}] - \beta^2) \tag{86}$$

In Equation 86, $\frac{Z}{A}$ is a constant, since an atomic nucleus has approximately the same numbers of protons and neutrons. In addition, the rate of ionisation energy loss per unit length traversed is dependent on the density of the material. For relativistic particles, v ≈ c. This leads to $\frac{dE}{dx}$ becoming logarithmically dependent on $(\beta\gamma)^2$, where $\beta\gamma = \frac{p}{Mc}$. The logarithmic dependence results in a slow increase in $\frac{dE}{dx}$ as $\beta\gamma$ increases [11]. This is shown in Figure 23.

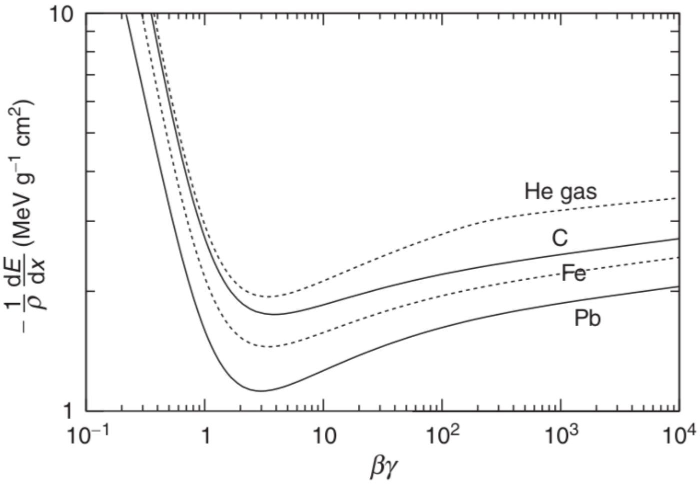


Figure 23: The rate of ionisation energy loss per unit length travelled by a singly charged particle as a function of $\beta\gamma$, expressed in units of GeV/c^2. The curves are for helium gas, carbon, iron and lead. The minima of each curve is at $\beta\gamma \approx 3$, which corresponds to Mimimum Ionising Particles (MIPs) [11].

The CMS Muon System comprises four muon stations containing Aluminum Drift Tubes (DTs), Cathode Strip Chambers (CSCs) and Resistive Plate Chambers (RPCs). The barrel region of the muon system comprises DTs and RPCs, whereas the endcaps of the muon system contain CSCs and RPCs.

Neutrinos pass through the CMS detector undetected. They are neutral in charge so do not leave hits in the tracker, are not deposited in the calorimetry and pass through the muon spectrometer. Instead, their presence is inferred using missing transverse energy and momenta calculations [56].

4.2.7 The Trigger System

The trigger system is used to select data corresponding to events that show potential signs of interesting physics. Otherwise, data is discarded. The system consists of two tiers: the Level 1 (L1) trigger and the High Level Trigger (HLT) [61].

The L1 trigger is a hardware system used to accept or reject events containing candidate objects by using trigger primitives from the calorimeters and muon detectors. The decision to accept or reject events takes place within 4 μs after a collision.

The L1 calorimeter trigger consists of two subsystems: the regional calorimeter trigger (RCT) system and the global calorimeter trigger (GT) system. The RCT receives and processes transverse energy and quality flag information from ECAL and HCAL towers. This information is then outputted as $e\mu$ candidates and regional E_T sums. This output is then sent to the GCT, which further sorts $e\gamma$ candidates and uses E_T sums to identify central, forward and tau jets and calculates global quantities, such as the missing transverse energy. Four $e\gamma$ candidates, which are either isolated or non-isolated, are then produced as output [61].

The L1 Muon Trigger comprises the three muon detector sub-detectors. Front-end trigger electronics in the DT and CSC systems are used to identify track segments using hit information. Optical fibres are then used to collect and submit the track segments to regional track finders in the electronics service cavern. Pattern recognition algorithms are then applied to identify and measure the momenta of muon candidates. This information is then shared between the DT and CSC track finders. Pattern Comparator Trigger logic boards are used to identify muon candidates, after they receive hits recorded by RPCs from the front-end electronics. The identified muon candidates are sorted using three regional track finders, and up to 4 or 8 candidates are then transmitted to the Global Muon Trigger (GMT) for every bunch crossing. Muon candidates identified by multiple systems are merged by the GMT, which is also used to discard low-quality candidates [61].

The HLT is implemented in software and is used to further refine the purity of physics objects. On average, it selects a rate of 400 Hz for offline storage [61].

4.3 The High-Luminosity LHC

The LHC is currently being upgraded, under the High-Luminosity LHC (HL-LHC) project. A diagram that provides a timeline of the HL-LHC project is shown in Figure 24.

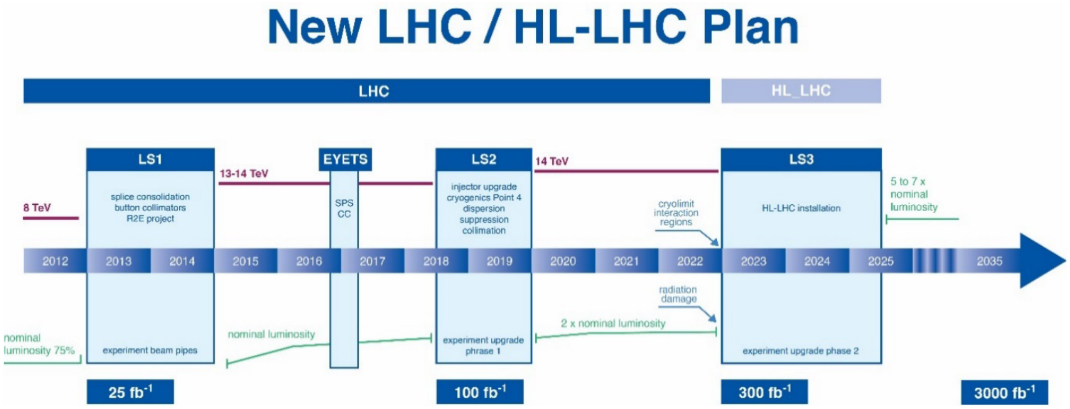


Figure 24: A timetable giving an overview of the planned increase in luminosity hoped to be provided by the HL-LHC project [62].

The upgrades will result in the instantaneous luminosity increasing by a factor of five and the integrated luminosity increasing by a factor of ten. Instantaneous luminosity, $\mathcal{L}$, is the rate at which collisions take place. The mathematical definition is shown in Equation 87. In this, γ is the Lorentz factor, f represents the revolution frequency, k_B represents the number of bunches, N_P represents the number of protons per bunch, ϵ_n represents the normalized transverse emittance, β^* represents the betatron function at the interaction point and F is the reduction factor due to the crossing angle.

$$\mathcal{L} = \frac{\gamma f k_B N_P^2}{4\pi\epsilon_n\beta^*}F \tag{87}$$

When $\mathcal{L}$ is multiplied by the cross section, σ, the observed event rate R can be calculated. The observed event rate is the number of interactions per unit time, where an interaction is defined as an inelastic proton-proton collision that leads to a significant energy deposit in the CMS calorimetry. The relationship between the R, $\mathcal{L}$ and σ is shown in equation 88. Taking the integral of $\mathcal{L}$ gives the integrated luminosity, $\int \mathcal{L}$dt. The integrated luminosity is the integral of the delivered luminosity over a period of time and can be used to measured the total number of collisions that occurs over a period of time. By multiplying the $\int \mathcal{L}$dt by the cross section, σ, the number of interactions, N, is obtained. This is shown in equation 89.

$$R = \mathcal{L}\sigma \tag{88}$$

$$N = \sigma \int \mathcal{L} dt \tag{89}$$

The LHC has a design luminosity of 10^{34} cm^{-2} s^{-1}, which was reached in June 2016. At this value, there are 25 interactions for every bunch crossing. Therefore, increasing the luminosity further would increase the number of interactions that would occur per bunch in the LHC. The HL-LHC upgrades are hoped to achieve a peak luminosity of 5×10^{34} cm^{-2} s^{-1}, an integrated luminosity of 250 fb^{-1} per year and 3000 fb^{-1} by 2038 [63].

4.3.1 CMS Phase-2 Upgrades

Barrel Calorimeters

The ECAL and HCAL calorimeters were designed to operate up to an integrated luminosity of 500 fb^{-1} over 10 years of LHC data taking at an instantaneous luminosity of 1×10^{34} $cm^2 s^{-1}$. The HCAL Phase 1 upgrades increased the detector's capability to operate at an instantaneous luminosity of 1×10^{34} $cm^2 s^{-1}$. However, in order to maintain this performance for increased pile up and at integrated luminosities of up to an 4500 fb^{-1}, the front-end electronics of the EB and the off detector electronics of the EB and HB must be replaced.

Specifically, the EB electronics must be able to operate under an increased trigger latency from $4\mu s$ to $12\mu s$, and an increased L1 Trigger rate from 100 kHz to 750 kHz [64].

Endcap Calorimeter

The HL-LHC will pose challenges for radiation tolerance and event pileup on the CMS detectors. These will particularly affect the existing forward calorimeters, which were designed to operate up to 500 fb^{-1}; beyond this integrated luminosity, a significant loss in physics performance is anticipated. Therefore, the current endcap calorimeters will be replaced by a high granularity calorimeter (HGCAL).

The HGCAL will have fine longitudinal and lateral granularity, which will result in a high sensor cell count. A fine longitudinal granularity is required for fine sampling of the longitudinal development of showers. This fine sampling will enable a good electromagnetic energy resolution, pattern recognition and pile up discrimination. A fine lateral granularity is required for a low energy equivalent of electronics noise. This will yield sufficient signal-to-noise ratio for Minimum Ionising Particles (MIPs).

This ratio must be good after $3000fb^{-1}$ to achieve a clean detection in each cell and to best achieve a suitable (on the level of $\approx 3\%$) inter-cell calibration accuracy. A fine lateral granularity will also aid in the observation of narrow jets and two shower separation. It will also lead to a reduction in the region used to measure energies, which will minimise the inclusion of energy measurements of particles from pile up interactions.

Additionally, the HGCAL will be dense in order to preserve the lateral compactness of the particle showers. It will also need time precision measurements of high energy showers from each silicon cell. This will contribute to pile up rejection and triggering interaction vertex identification. From simulations, it is predicted that the HGCAL must be able to withstand a fluence of 1.5×10^{16} n_{eq}/cm^2, where n_{eq}/cm^2 represents the number of 1 MeV equivalent neutrons per square centimetre, and a dose of $2MGy$.

The HGCAL will comprise an electromagnetic compartment (CE-E) and a hadronic compartment (CE-H), with a 120mm thick polythene neutron moderator layer in front of the calorimeter to reduce neutron flux in the tracker. Perpendicular to its layers, the HGCAL and the neutron moderator layer has a total calorimeter thickness of 10.7λ. All HGCAL layers read out information for energy measurements. The L1 trigger primitives are made using all CE-H layers and alternate layers in the CE-E.

The front sections of HGCAL, which covers the entire electromagnetic part and the majority of the hadronic part, will comprise silicon sensors, with each sensor being 300 μm, 200 μm or 120 μm thick, in regions of increasing fluence, respectively. In studies conducted by groups in CMS, it was determined that silicon would be able to tolerate such high radiation fluence and dosage levels. To achieve the required signal-to-noise ratio for Minimum Ionising Particles (MIPs), the silicon sensors will contain small ($\approx 0.5 - 1cm^2$) low-capacitance silicon cells.

The rear sections will comprise highly-segmented plastic scintillator tiles ($\approx 4 - 30cm^2$) and silicon photomultipliers (SiPMs) to read out scintillation light. The SiPMs must be operated at -30 degrees celsius in order to keep the radiation-induced energy equivalent of electronics noise low. Therefore, the HGCAL as a whole will be cooled to and operate at -30 degrees celsius by a $C0_2$ system.

The CE-E will consist of 28 sampling layers, with each layer having a thickness of 34cm. Each layer will be approximately 26 X_0 and 1.7 λ deep. The sampling layers are formed by 14 layers of cassettes, where a cassette is a 60 degree unit of the plane of the CE-E structure. A cassette of the CE-E will contain silicon modules. A diagram of an individual silicon module is shown in Figure 25 and comprises a silicon wafer sandwiched between a 1.4mm thick WCu baseplate covered by a 105 μm thick kapton sheet, covered by a layer of gold foil, for electrical insulation and a printed circuit board, known as a hexaboard, which contains the front-end electronics. A photograph of a hexaboard is shown in Figure 26. The kapton sheet is wirebonded to the hexaboard through a hole at each of the six corners and the wirebonds are covered by a clear, radiation-tolerant silicon elastomer. The Silicon sensors in the silicon wafer will be used as the active detector element. Each sensor will be 163mm wide, forming an 8 inch hexagonal wafer. A hexagonal shape was chosen to maximise the amount of material used while preventing gaps in detector coverage. Modules are positioned on either side of a 6mm thick Cu cooling baseplate. The Cu cooling baseplate and the two WCu baseplates form an absorber layer. Either side of this sandwich are two, 2.1mm thick lead plates clad with 0.3mm stainless steel sheets.

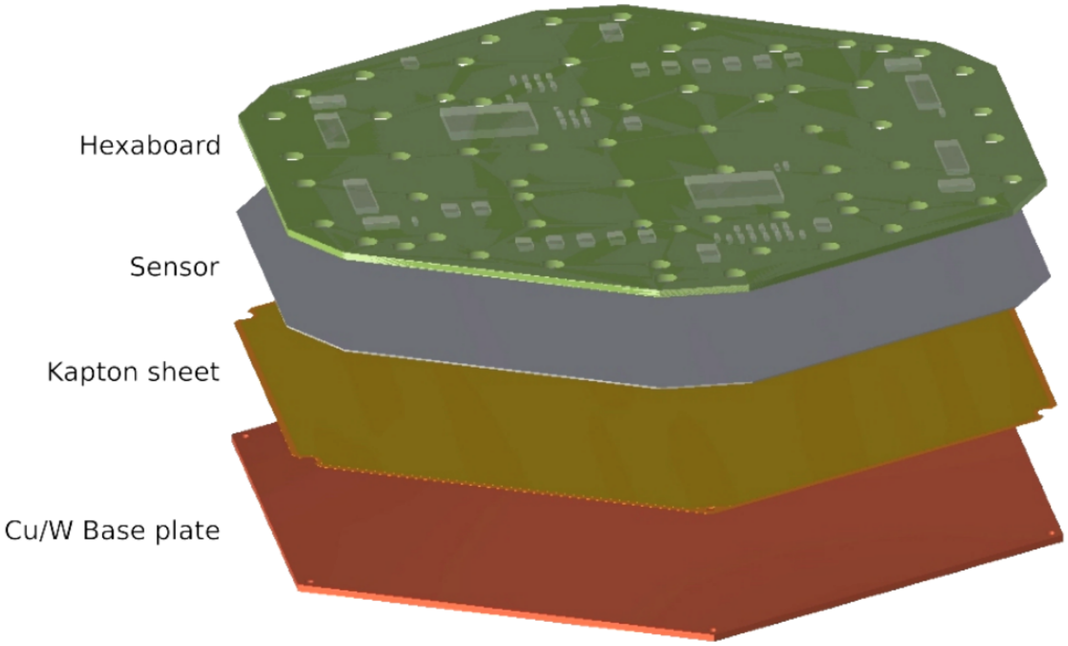


Figure 25: The layers of a silicon module in the CE-E [52].

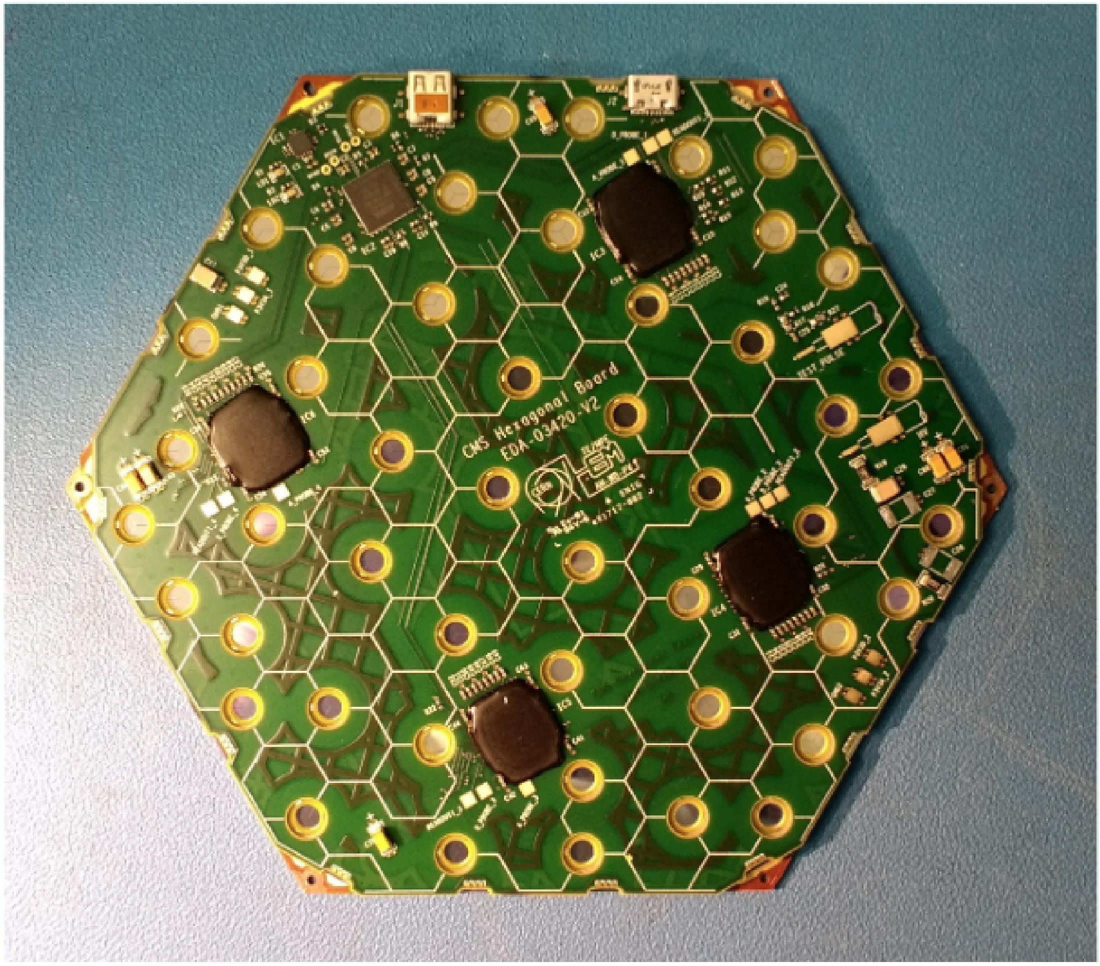


Figure 26: A hexaboard containing the SKIROC2-CMS front end readout chips and an FPGA [52].

Unlike the CE-E, the hadronic compartment of HGCAL (CE-H) will consist of 24 samplings subdivided into 30 degree wide cassettes. Each cassette contains silicon modules and scintillator tileboards; the radii at which silicon modules and scintillator tileboard are used in the CE-H are dependent on the radiation levels. The radiation dose induced light-loss in the scintillators must not exceed 50% and the neutron fluence must not exceed 8×10^{13} n_{eq}/cm^2. This will keep the energy equivalent of the electronics noise low to enable a good measurement of the MIP response. The energy equivalent of electronics noise arises from SiPM leakage current and light loss. The absorber material is treated as its own structure and comprises 12 planes of 35mm thick stainless steel plates and a further 12 planes of 68 mm thick stainless steel plates. In between the plates are the silicon modules and scintillator tileboards. Unlike the cassettes in the CE-E, the sensors in the CE-H silicon modules and scintillator tileboards are located on one side of the cooling plate only.

For the HGCAL, two types of scintillator material have been proposed: polyvinyltoluene-based (PVT) or polystyrene-based (PS). For the SiPM photodetectors, the proposed design was based on a prototype design for the Analog Hadron Calorimeter (AHCAL) of the Calorimeter for Linear Collider Experiment (CALICE). The CALICE is calorimetry for future International Linear Collider (ILC) experiments. The design was used in the Phase-1 upgrade of the CMS HCAL, and yielded a high gain of 10^5 and a photodetection efficiency greater than 30%.

L1 Trigger

Like other CMS subsystems, the increased pile up in the HL-LHC will pose challenges for the L1 trigger. This increased pile up will result in a higher number of particles produced per bunch collision which, consequently, will result in an increased ambient energy in calorimeter measurements. There is also expected to be an increase in low-momentum muons being misidentified as high-momentum muons. This motivates the need for the trigger system to be upgraded in order to improve particle identification in the trigger.

Minimum Ionising Particles Timing Detector

The Minimum Ionising Particles Timing Detector (MTD) will be a new subdetector of the CMS experiment and is expected to enable the precise measurement of the production time for Mimimum Ionising Particles (MIPs), charged hadron identification and searches for long-lived particles. It is anticipated that the MTD will give MIP timing information with a resolution of 30-40 μs by 2026 and 50-60 μs by the end of HL-LHC operation. This time degradation is expected as a result of radiation damage.

Muon System

In the 2016/17 winter shutdown, ’demonstrator’ Gas Electron Multiplier (GEM) chambers of the type GE1/1 were installed. This was to improve the performance of the muon trigger and tracking under high luminosity conditions. The GE1/1 system covers a pseudorapidity range of $1.6 < |\eta| < 2.4$ and utilises GEM technology. A GEM chamber comprises two layers of thin polyimide foil with holes etched in a hexagonal pattern, clad by thin conductive copper layers. A voltage is applied across the two polyimide layers, resulting in a 60-100kV/cm electric field produced across the two layers. Traversing particles ionise the Ar/C0$_2$ gas contained, which leads to an avalanche of electrons that induces a readout signal. The GE1/1 system comprises a ‘Triple-GEM detector’ (i.e. three cascaded GEM foils) to achieve a charge amplification factor of up to 10^5. A diagram of the GE1/1 system is shown in Figure 27. A pair of Triple-GEM detectors form a “super-chamber”, which help to maximise detector efficiency [65].

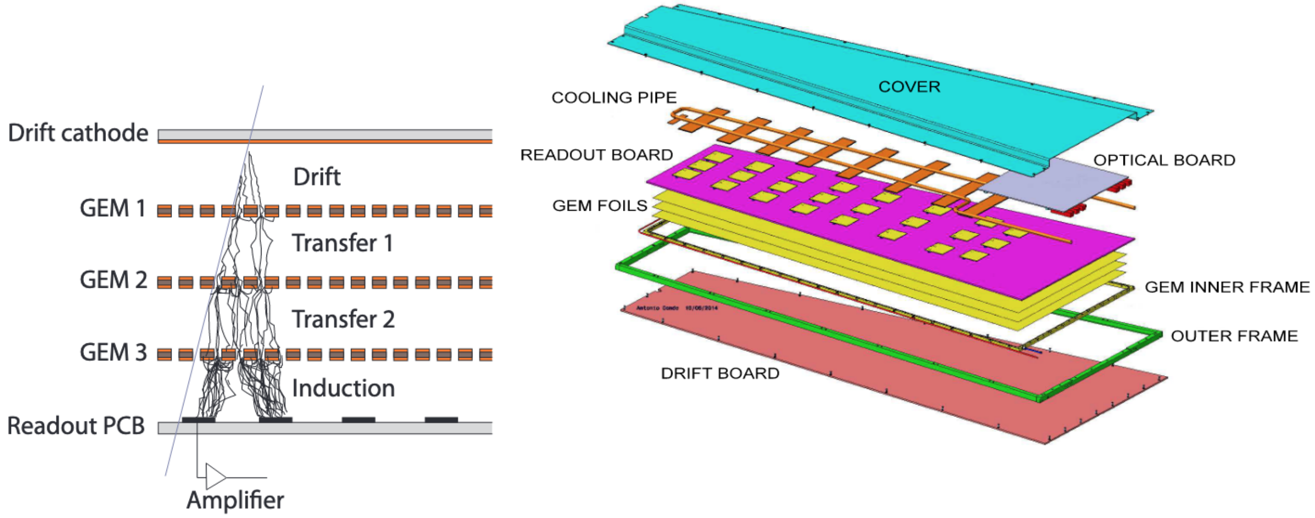


Figure 27: Left: A Triple-GEM detector. Right: An exploded view of a Triple-GEM detector [65].

Tracker Upgrades

Upgrades to the CMS Tracking System will be made, as part of the HL-LHC project. The upgrade project is known as the Phase-2 upgrade for the CMS tracker. The upgrades will be made to the L1 trigger for the tracker, the Inner Tracker (IT) and the Outer Tracker (OT). The IT will consist of four layers of silicon pixel modules whereas the OT will consist of six layers of silicon modules that will be composed of strip and macro-pixel detectors [66].

5 Event reconstruction

5.1 Particle Flow

The Particle Flow (PF) algorithm uses information from all CMS sub-detectors to reconstruct stable particles in each event [59]. The PF algorithm enables a highly efficient event reconstruction while giving a low fake rate. The information from each sub-detector used is referred to as an "element" in the PF algorithm. The elements are hits left by charged particle tracks, calorimeter clusters and muon tracks. Tracker information is used to reconstruct charged particle tracks by implementing an iterative-tracking strategy. In the first iteration, tracks are seeded and reconstructed. The reconstruction is carried out with very tight seeding criteria applied. In subsequent iterations, unambiguous hits are removed to reduce combinatorics and achieve a low fake rate ($\sim 1\%$). Looser seeding criteria are also applied in later iterations to improve tracking efficiency.

A clustering algorithm is used to identify calorimeter clusters in each sub-region of the ECAL and HCAL, with the exception of HF (EB, EE, HB, HE, PS1 and PS2). The clustering algorithm is not applied to the HF sub-region so a cluster will form from each cell instead. The algorithm comprises three steps. In the first, cells are identified with maxima larger than a certain energy threshold. The value of the threshold is equal to the value two standard deviations above the mean value of the electronics noise in a calorimeter. This corresponds to 80 MeV ECAL barrel region, 300 MeV in the ECAL endcap regions and 800 MeV in the HCAL. The cells that meet this criterion are defined as "cluster seeds". A topological cluster is then formed from surrounding cells with at least one side in contact with the cluster seed. The surrounding cells must also be above the same threshold energy. A PF cluster is formed from multiple topological clusters and the contribution of the energy of each topological cluster to the total energy of the PF cluster is dependent on the distance between the cell and cluster. An iterative method is used to determine the energies and distances [59].

An individual particle gives rise to multiple PF elements and the link algorithm connects PF elements to fully reconstruct particles. One, two or three elements in an event are identified and directly or indirectly linked in a single, small "block". Blocks form an input to the particle reconstruction and identification algorithm.

A charged particle leaves a hit in the tracker and energy is deposited in either the ECAL (e.g. electrons or bremsstrahlung photons radiated from electrons), HCAL (e.g. charged hadrons) or the muon chambers (for muons). The link algorithm is used to link a charged particle track from the tracking system to a calorimeter cluster. Firstly, an extrapolation is made from the last hit measured in the tracker to a calorimeter region. The distance over which the extrapolation takes place is defined as the "link distance". For the PS, the region the extrapolation is made to is its last two layers. For the ECAL, this corresponds to a depth of the calorimeter material at which a maximum would be reached for the typical energy of an EM shower. For the HCAL, the region extrapolated to is a depth of one radiation length in the calorimeter material. The calorimeter region must be within the set boundary of a single cell size in each direction. A single cell size was chosen for the boundary as this sufficiently accounts for any gaps in the detector coverage, the uncertainty in the EM shower maximum measurement and the effect of multiple scattering. The tracker and the extrapolated position are then linked using the algorithm.

Electron reconstruction and identification is achieved by firstly identifying short tracks. This is because energy loss via bremsstrahlung gives rise to electrons with short tracks. A Gaussian-Sum filter is used to re-fit electron tracks and to enable the full electron trajectory to be identified. Numerous variables used in tracking and calorimetry are then considered to determine the particle identification. The resulting electron is referred to as a PF electron. To prevent the misidentication of a hadronic cluster in future iterations, the relative uncertainty on the measured p_T must be less than the expected relative calorimetric energy resolution for a charged hadron. The identified track and ECAL cluster elements are no longer considered in the block for further iterations of the link algorithm. The remaining elements are used for the reconstruction and identification of charged or neutral hadrons, photons and muons.

The same method is applied for the reconstruction of charged hadrons. However, neutrally-charged particles do not leave hits in the tracker and are deposited in either the ECAL (for particles governed by the EM interaction e.g. photons) or the HCAL (for strongly interacting particles e.g. neutral hadrons such as π^0). Neutrally-charged particles are reconstructed from cluster candidates with no corresponding track in the tracker system, or by comparing the track four momentum with the total calibrated calorimetric energy. The calibrated calorimetric energy for HCAL clusters is calculated using the calibration function, shown in Equation 90. In Equation 90, a, b and c are calibration coefficients. The ECAL and HCAL energies are represented by E_{ECAL} and E_{HCAL}, respectively. The pseudorapidty of the HCAL cluster is given by η and E is the true energy; this is taken to either be the largest value out of the total uncalibrated calorimetric energy or the total charged-particle momentum. The comparison between the track four momentum and the total calibrated calorimetric energy is further explained in the penultimate paragraph of this section.

$$E_{calib} = a + b(E, \eta)E_{ECAL} + c(E, \eta)E_{HCAL} \tag{90}$$

The calibration coefficients b and c are determined from a χ^2 distribution for a given value of a. The coefficient a is a correction factor used for the thresholds applied to calorimetric energies. The summation equation for this is shown in Equation 91, which is applied over all events and represented in bins of E. In Equation 91, the true energy and the expected calorimetric energy resolution of the ith hadron are represented by E^i and σ_i, respectively. The χ^2 for each bin of the distribution is minimised.

$$\chi^2 = \sum_{i=1}^{N} \frac{(E^i_{calib} - E^i)^2}{\sigma_i^2(E^i_{calib})} \tag{91}$$

Muon tracks in the tracker are linked to tracks in the muon chambers by using the link algorithm. A global fit is carried out between the tracks and the corresponding value of χ^2 is calculated. Any muon that gives rise to an accepted χ^2 value is referred to as a global muon. After exclusion criteria are applied, the muon is referred to as a PF muon and the track is no longer considered as an element in the block.

Exclusion criteria are applied to link a hit left in the tracking system with the optimal match out of multiple cluster candidates. In the scenario where multiple tracks could be linked to a calorimeter cluster, the sum of the tracks' four momenta is calculated and is considered if the sum is smaller than the total calibrated calorimetric energy. However, if the sum is significantly larger than the total calibrated calorimetric energy, the candidate is considered to be either a PF photon or a PF neutral hadron. If the sum is significantly larger than the total calibrated calorimetric energy calculated for the ECAL, the candidate is considered to be a PF photon while the excess energy difference is considered as the energy of a PF neutral hadron. This is due to the photon energy providing a higher contribution to the total energy of jets compared to neutral-hadrons. Alternatively, if an individual track could be linked to multiple calorimeter cluster candidates, only the track that corresponds to the shortest linked distance is used. In the case where a track in the tracker system could be linked to multiple muon chamber track candidates, the muon that results in lowest value of χ^2 is chosen. Additionally, the four momentum of the muon must be within three standard deviations of the corresponding track in the tracking system [59].

Finally, missing transverse energy calculations are used to infer the presence of neutrinos and other weakly interacting particles.

5.2 b-Tagging Algorithms

In the tracker, jets are produced from the hadronisation of quarks. Among the constituents of the jet are b-hadrons, which have a long lifetime of 1.6 ps. The long lifetime means that, after production at the primary vertex (PV), the b-hadron travels an average finite distance of 500 μm before decaying at the secondary vertex (SV). The b-hadron can produce approximately five charged particles in an individual decay. The distance between the PV and the point of closest approach to the track is defined as the impact parameter (IP), as shown in Figure 28.

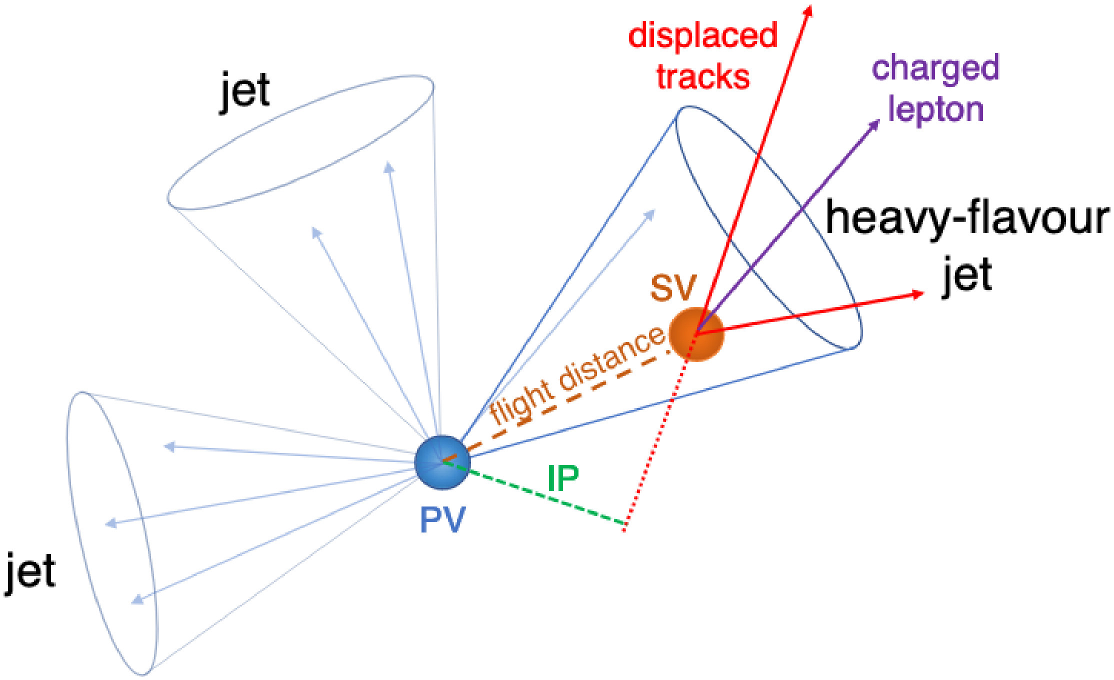


Figure 28: A schematic of the formation of a jet [67].

Thus, the b-hadron can be exploited by the use of b-tagging algorithms to identify bottom quarks produced in the tracker. The b-tagging algorithms used by the CMS collaboration include the Combined Secondary Vertex version 2 algorithm (CSVv2) and the DeepCSV algorithms. Both algorithms make use of a deep neural network (DNN), which is analogous to neurons in the brain. A DNN comprises an input layer, an output layer and multiple "hidden" layers in between. Firstly, a transformation maps the input obtained from the input layer $\boldsymbol{x}$ to the first hidden layer h_0. The general form of a transformation is shown in Equation 92, where h_{i+1} is the transformation of the information received from the previous layer, g_i is the activation function, W is a matrix whose elements are weights used in the function and $\boldsymbol{b}$ is a vector whose elements are biases [68].

$$h_{i+1} = g_i(W_i \boldsymbol{h_i} + \boldsymbol{b_i}) \tag{92}$$

An iterative process, called training, is carried out to reduce the size of the cost function C. The cost function indicates how well the classifier performs and is minimised in training and is mathematically defined as the sum of the loss term L and the regularisation term Ω ($C = L + \Omega$). The weights and biases are adjusted in order to minimise C. The loss function indicates how well the classifier classifies b-jet and non-b-jet candidates and is the sum of the binary cross-entropies for all binary events. The binary cross-entropy is shown in Equation 93, where $y = 1$ for b-tagged candidates and $y = 0$ otherwise. The greater the difference between y and $\hat{y}$, the larger the binary cross-entropy, which leads to a larger loss function.

$$l(y, \hat{y}) = (y - 1)log(1 - \hat{y}) - ylog(\hat{y}) \tag{93}$$

The second term in the cost function is the regularisation term Ω, which provides an indication of the complexity of the model under consideration. A higher value of Ω indicates a more complex model, which could be the result of unwanted noise being classified in the training data. This would lead to overtraining. The mathematical definition of Ω is shown in Equation 94. In this, w represents the weight, n represents the number of weights in the network and α and λ are coefficients whose values are chosen for each layer of the NN [69].

$$\Omega = \frac{\alpha}{n}\sum_{n}|w| + \frac{\lambda}{2n}\sum_{n}w^2 \tag{94}$$

The CSVv2 algorithm is based upon the CSV algorithm, which was used during Run 1. There are two versions of the algorithm: the inclusive jet finding (IVF) algorithm and the adaptive vertex reconstruction (AVR) algorithm. The IVF algorithm uses all reconstructed tracks in the event with transverse momenta greater than 30 GeV, whereas the AVR algorithm only uses selected tracks in reconstructed jets. For the CSVv2 algorithm, the rate of misidentifying light jets as b-jets at the loose, medium and tight working points are 10%, 1% and 0.1%, respectively. The DeepCSV algorithm also uses a deep neural network and the same input variables as the CSVv2 algorithm, but instead uses a larger number of particle tracks. It contains six layers in total and four of the layers are hidden [68].

5.3 Jet-Clustering Algorithms

The inclusive k_t, anti-k_t and the Cambridge/Aachen jet-clustering algorithms [71] are among the algorithms used to identify jets that arise from hadronic collisions. A jet clustering algorithm must be infrared and collinear safe. Infrared refers to when soft particles are added to the jet while collinear refers to the splitting of particles. The algorithms use Equation 95.

$$d_{ij} = min(k_{ti}^{2p}, k_{tj}^{2p})\frac{\Delta_{ij}^2}{R^2} \tag{95}$$

In Equation 95, R is the radius parameter, $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$. The transverse momentum is represented by k_{ti}, y_i is the rapidity of the ith particle and ϕ_i is the azimuth of the ith particle. The case where $p = 1$ corresponds to the inclusive k_t algorithm, $p = 0$ corresponds to the inclusive Cambridge/Aachen algorithm and $p = -1$ corresponds to the anti-k_t algorithm.

6 Hit Resolution Measurements Study for the CMS Tracker

Hit resolution measurements were conducted for various CMS Tracker regions by using the 'pair method [72] to help assess the performance of the CMS Tracker. As part of this, tracks were considered that satisfied the following criteria:

- The charged particle must have traversed two sufficiently-overlapping detector modules.
- Tracks must have transverse momenta greater than 3 GeV.
- Track momentum is > 15 GeV for strips and > 5 GeV for pixels.
- More than six hits must be left in the SST layers and the pixel detector.
- No more than four of these hits can have a large cluster width.
- The track fit must yield a good χ^2 probability of $\chi^2 \geqslant 10^{-2}$.
- The predicted path (the distance of propagation of the charged particle from one layer to the next) must be $<$ 7cm in the Tracker Inner Barrel (TIB) and Tracker Outer Barrel (TOB).
- The predicted path must be $<$ 2cm for pixels.
- The error in the distance between two hit resolution measurements must be < 0.0025.
- Clusters in the silicon module under consideration must have the same width.
- Clusters positioned on the edge of modules are not considered.

The hit resolution measurements (σ_{hit}) were then obtained using Equation 96, where $\sigma^2_{HitDX-TrackDX}$ is the standard deviation of a Gaussian-fitted distribution of the double differences between the measured and predicted hit positions between two sensors. The squared mean of the distribution of the predicted positions is represented by $\sigma^2_{TrackDXE}$.

$$\sigma_{hit} = \sqrt{\frac{\sigma^2_{HitDX-TrackDX} - \sigma^2_{TrackDXE}}{2}} \tag{96}$$

The results are shown in Figure 29. Hit resolution measurements obtained for the TIB layers 1 and 2 and TOB layer 5 are in agreement with previous results, which are shown in Figure 30. Results for the TIB layers 3 and 4 and TOB layers 1-4 were observed to be lower than expected; further investigation is needed to identify the cause, which was not possible due to time constraints. The result for the TOB layer 6 was found to be 0; this was possibly due to a bug in the code since no events passed the filter containing the definition for TOB layer 6 but this was not resolved due to time constraints so would require further investigation.

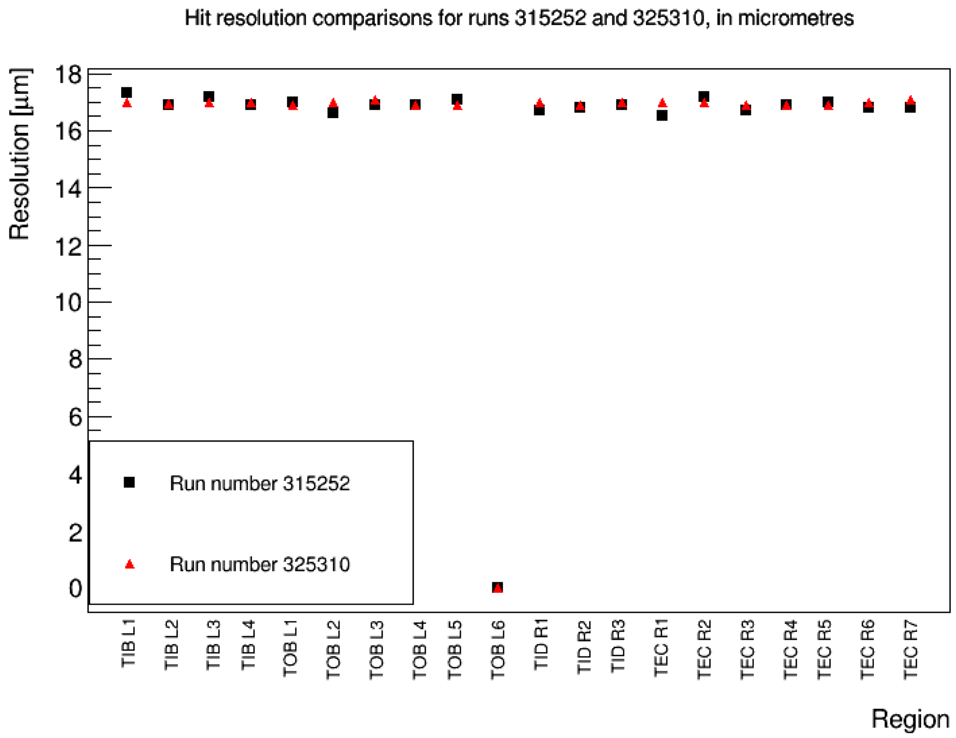


Figure 29: Hit resolutions for various tracker regions, for runs 315252 and 325310. The runs took place on the 26/04/2018 and the 26/10/2018, respectively. The resolutions are quoted in micrometres. From left to right, the regions are: Tracker Inner Barrel (TIB) Layers 1-4, Tracker Outer Barrel (TOB) Layers 1-6, Tracker Inner Disk (R1-3), Tracker Endcap (R1-7).

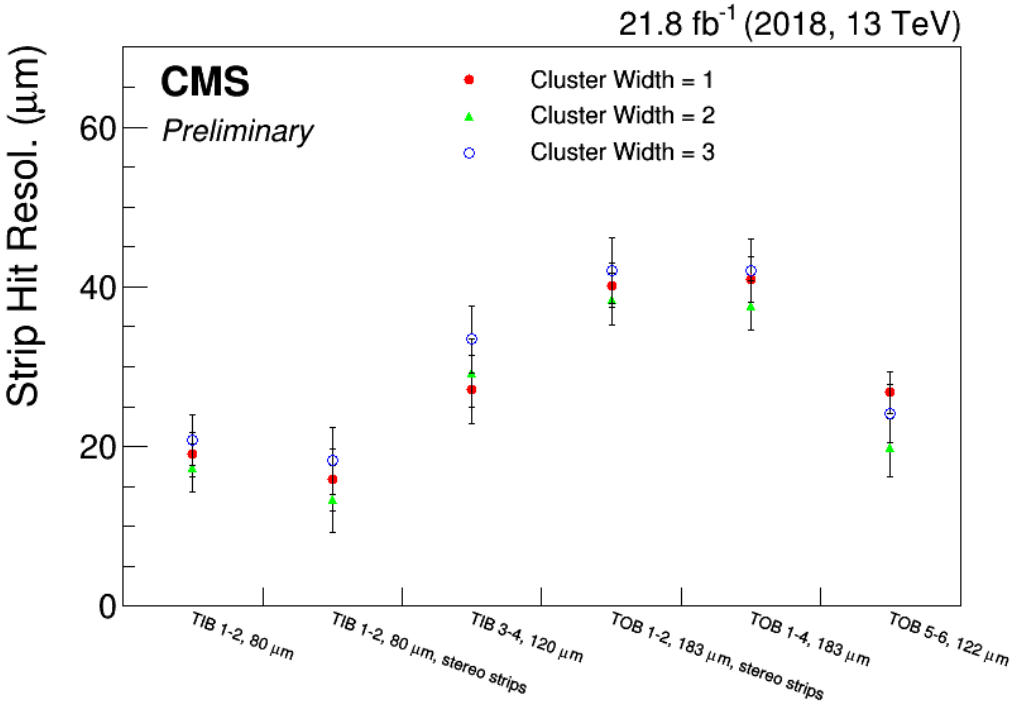


Figure 30: Previous results for the strip hit resolution measurements for 2018 [73].

7 RIVET Plug-in Development for tZq Production in Dilepton Final States

A RIVET [74] (Robust Independent Validation of Experiment and Theory) plug-in was developed for tZq production in the dilepton final state. The RIVET system can be used for the development, validation and tuning of Monte Carlo event generators. The plug-in provides a test of the performance of Monte Carlo event generators to determine if their outputs agree with SM predictions. The plug-in was used to test MC event generators for generating tZq signal events. The samples simulate data taken from the CMS detector in 2016, 2017 and 2018. The plug-in was run over approximately 14.5 million tZq MC signal events for 2016 and approximately 13 million tZq MC signal events for 2017 and 2018. Events were generated using the Pythia8 Monte Carlo generator.

The combined distributions for the Run 2 tZq MC signal samples are shown in Figures 31- 43. The pseudorapidity and transverse momentum distributions for all final-state jets are shown in Figure 31-left and Figure 31-right, respectively. The pseudorapidity and transverse momentum distributions for the leading final state jet are shown in Figure 32-left and Figure 32-right, respectively. The pseudorapidity and transverse momentum distributions for the sub-leading final state jet are shown in Figure 33-left and Figure 33-right, respectively. The jet H_T distribution and jet multiplicity distribution for all final state jets are shown in Figure 34-left and Figure 34-right, respectively. The pseudorapidity and transverse momentum distributions for the leading b-jet are shown in Figure 35-left and Figure 35-right, respectively. The mass distribution for the leading b-jet is shown in Figure 36-left while the distribution for the difference in pseudorapidity between the leading b-jet and the W boson is shown in Figure 36-right. The distribution for the $\Delta\phi$ value between the leading b-jet and the W boson is shown in Figure 37-left and the distribution for the ΔR value between the leading b-jet and the W boson is shown in Figure 37-right. For the leading lepton, the distributions for the transverse momentum and the pseudorapidity are shown in Figure 38-left and Figure 38-right, respectively. For the sub-leading lepton, the distributions for the transverse momentum and the pseudorapidity are shown in Figure 39-left and Figure 39-right, respectively. The pseudorapidity and transverse momentum distributions for the leading jet from the W pair are shown in Figure 40-left and Figure 40-right, respectively. The pseudorapidity and transverse momentum distributions for the sub-leading jet from the W pair are shown in Figure 41-left and Figure 41-right, respectively. Finally, the W boson and Z boson mass distributions are shown in Figure 43-left and Figure 43-right, respectively.

Results were consistent between 2016, 2017 and 2018 for all variables except for the jet multiplicity, shown in Figure 34 right. This is expected since the lepton selection criteria were increased between 2016 and 2017 due to increased trigger thresholds, and the same selection criteria were used for 2017 and 2018. The selection cuts are given in Section 9.2.

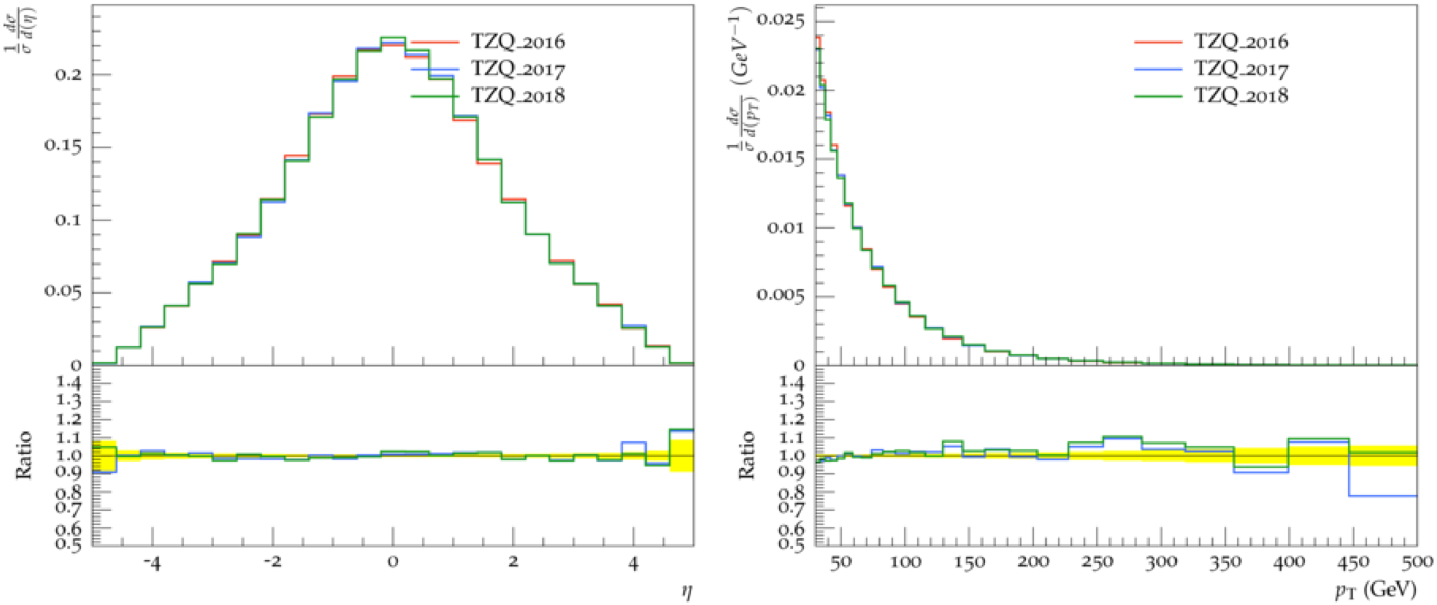


Figure 31: The pseudorapidity (left) and transverse momenta (right) values of all final state jets.

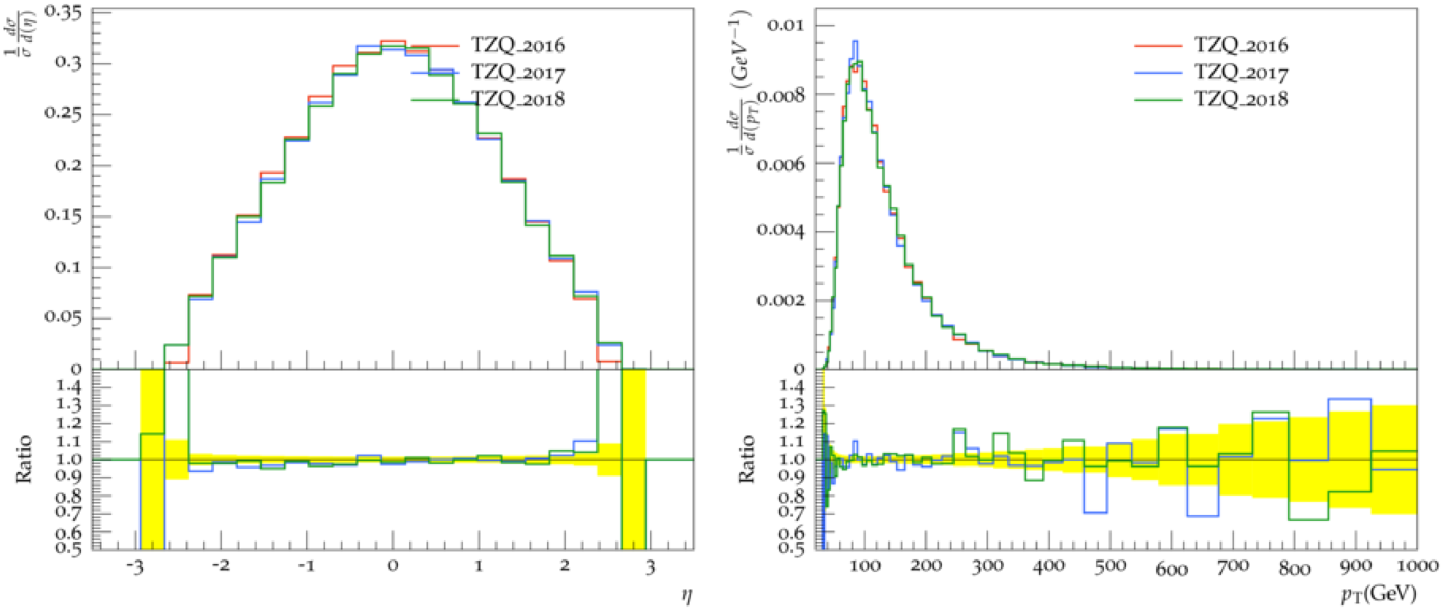


Figure 32: The pseudorapidity (left) and transverse momenta (right) values for the leading final state jet.

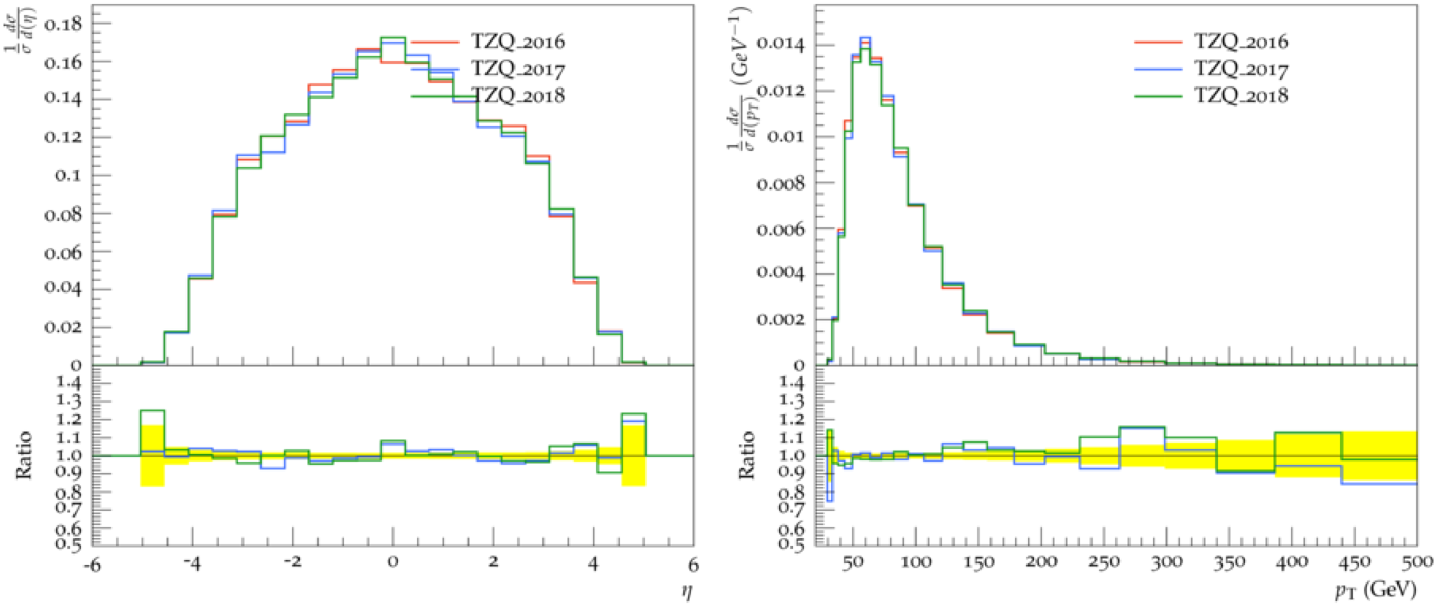


Figure 33: The pseudorapidity (left) and transverse momenta (right) values for the sub-leading final state jet.

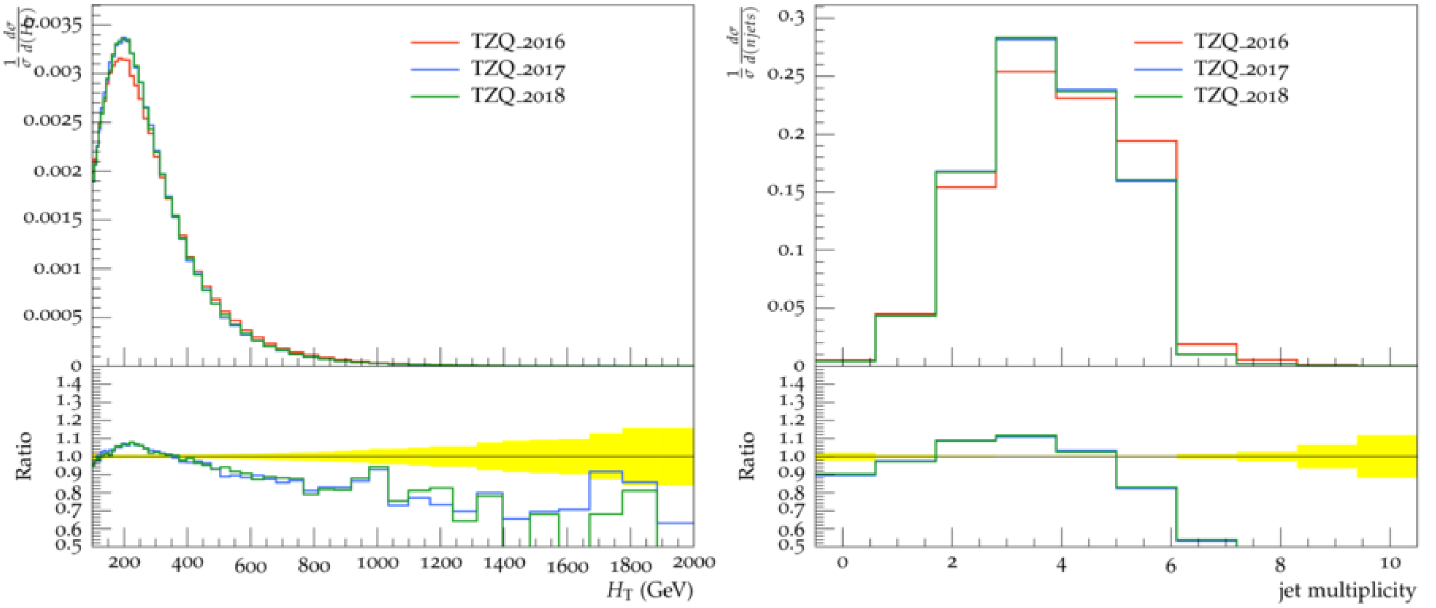


Figure 34: The jet H_T (left) and the jet multiplicities (right) for all final state jets.

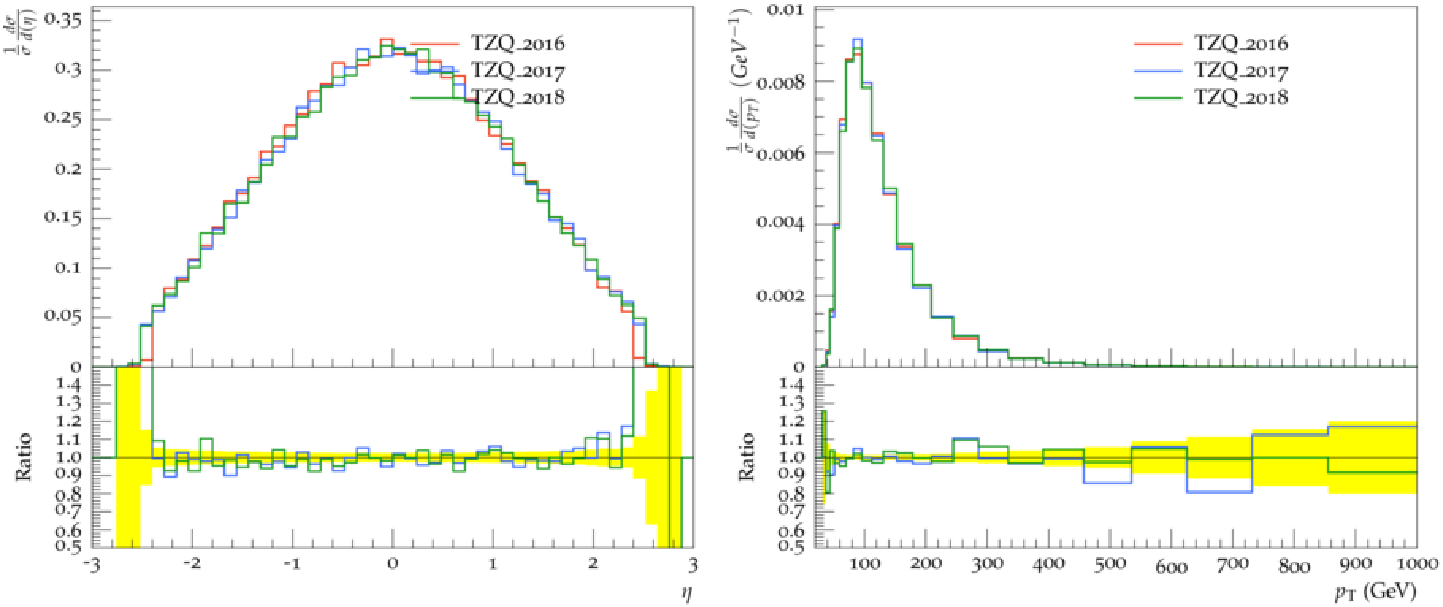


Figure 35: The eta (left) and transverse momenta (right) values for the leading b-jet.

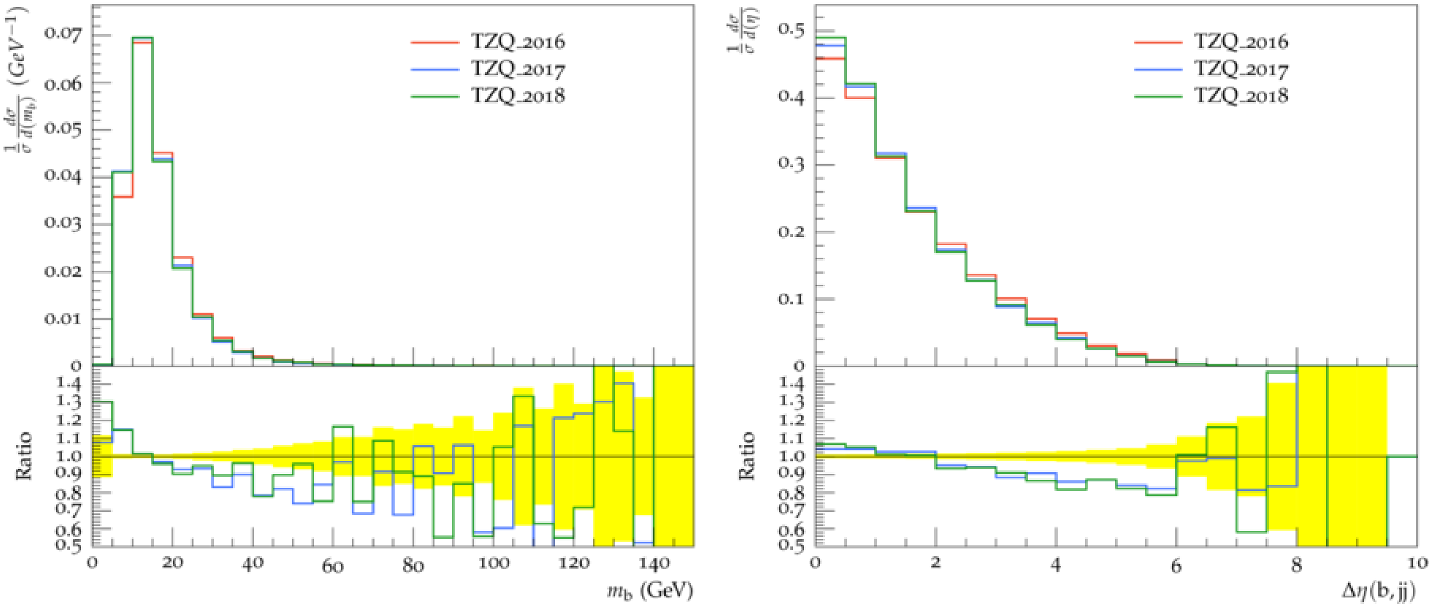


Figure 36: The mass of the leading b jet (left) and the delta eta value between the leading b jet and W boson.

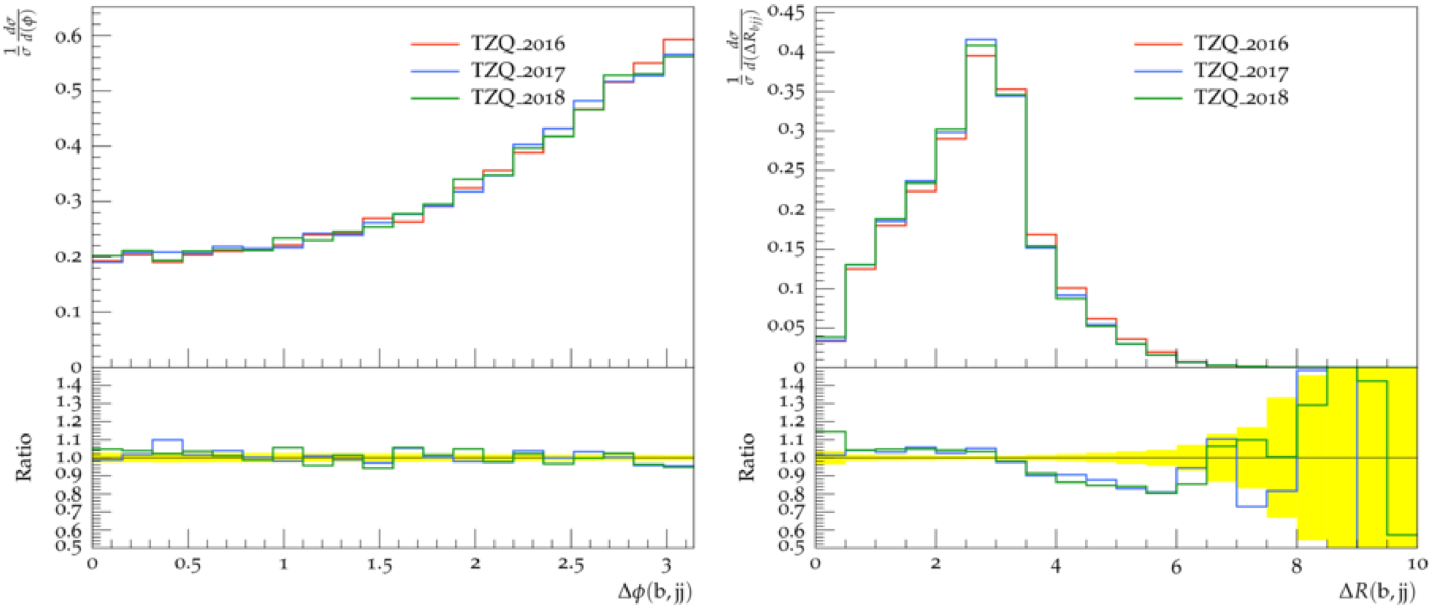


Figure 37: The delta phi value between the of the leading b jet and the W boson (left). The delta R value between the leading b jet and W boson (right).

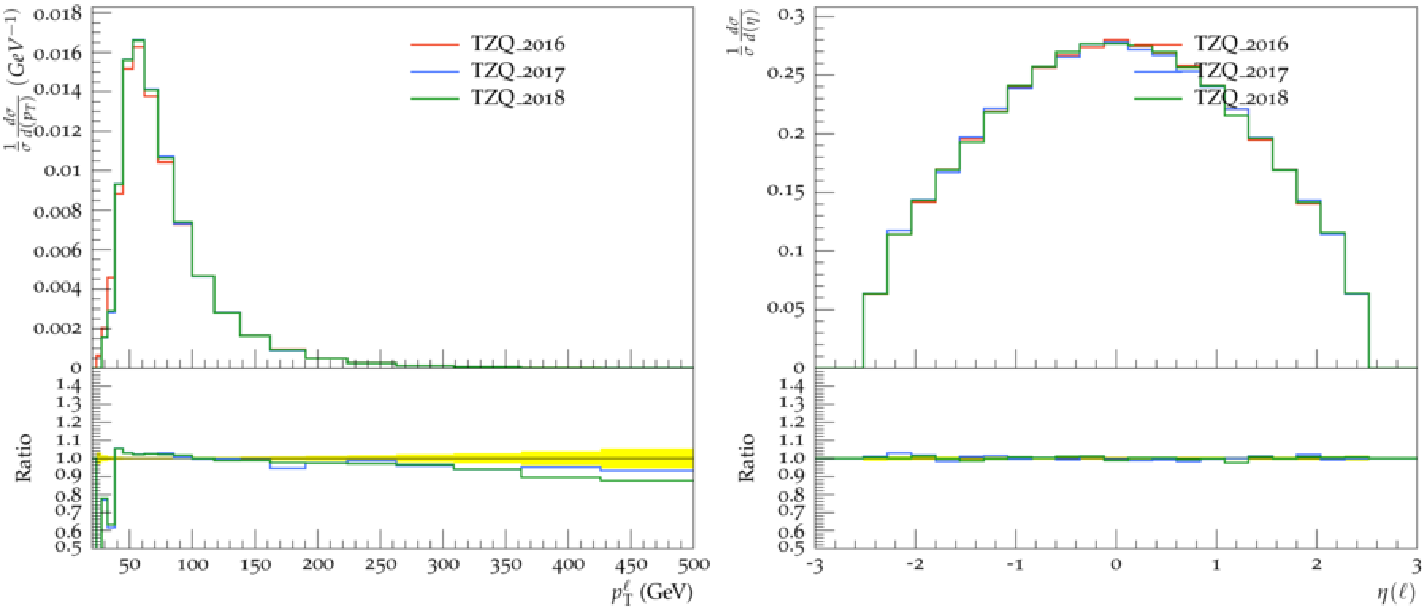


Figure 38: The transverse momenta (left) and pseudorapidity (right) values of the leading lepton.

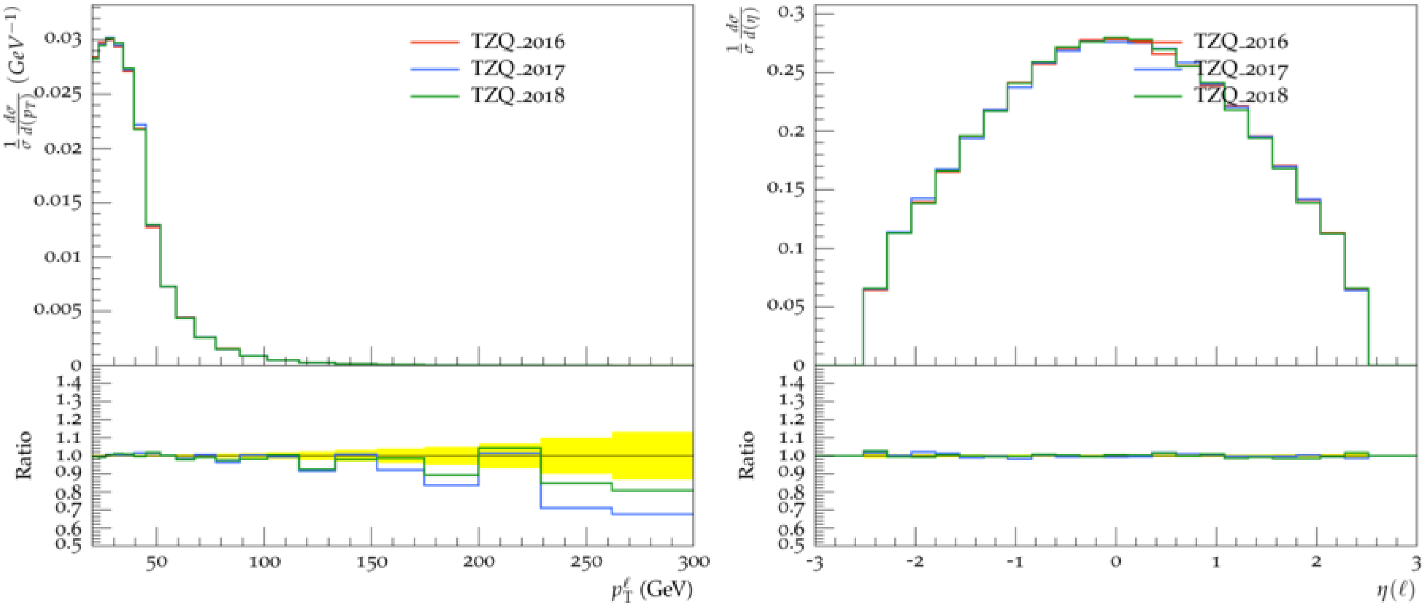


Figure 39: The transverse momenta (left) and pseudorapidity (right) values of the sub-leading lepton.

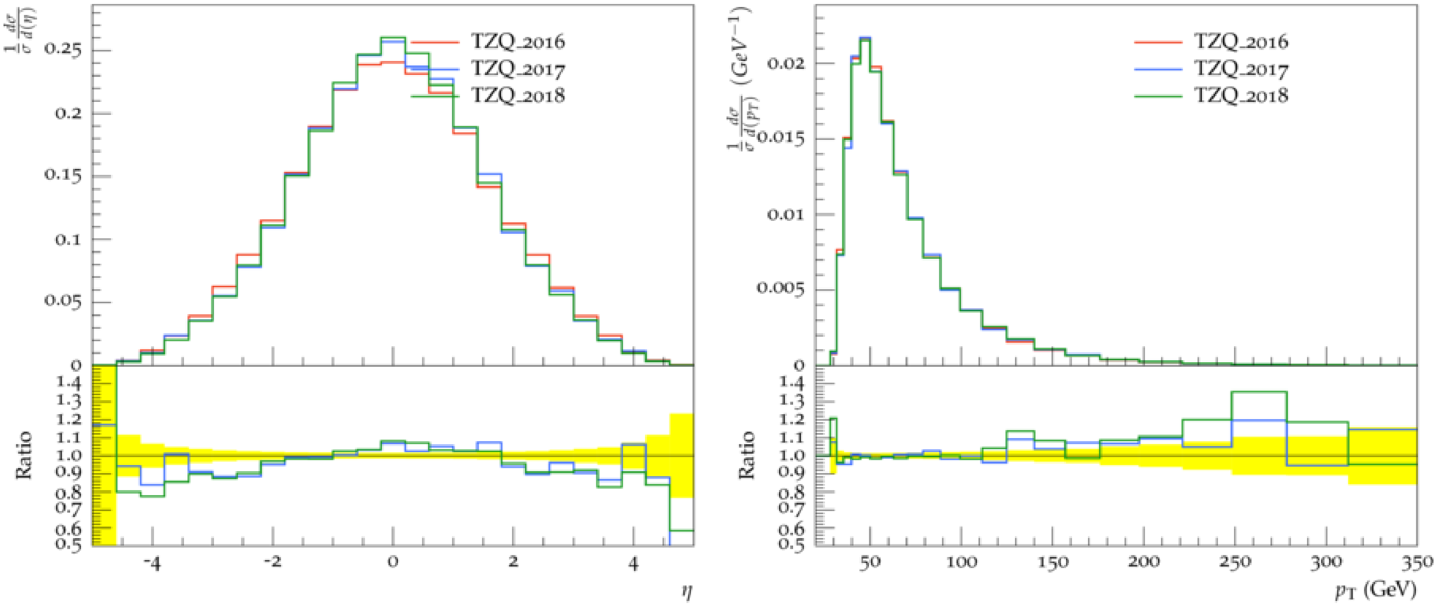


Figure 40: The pseudorapidity (left) and transverse momenta (right) values of the leading jet from the W pair.

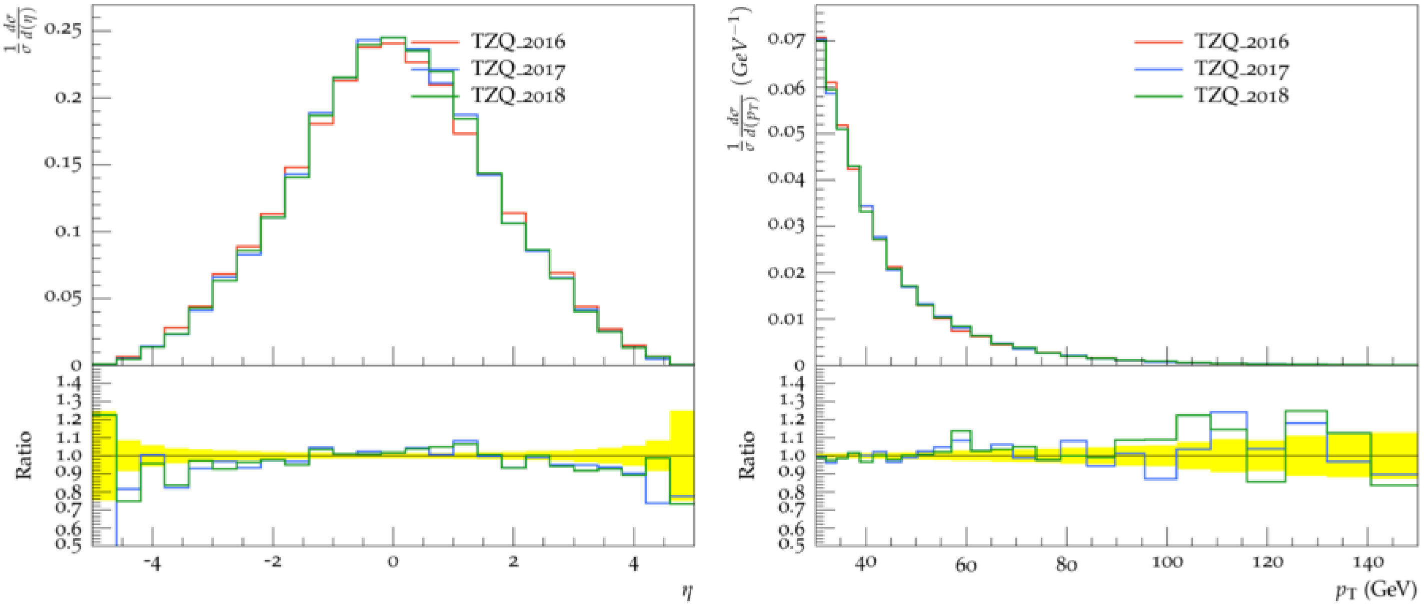


Figure 41: The pseudorapidity (left) and transverse momenta (right) values of the sub-leading jet from the W pair.

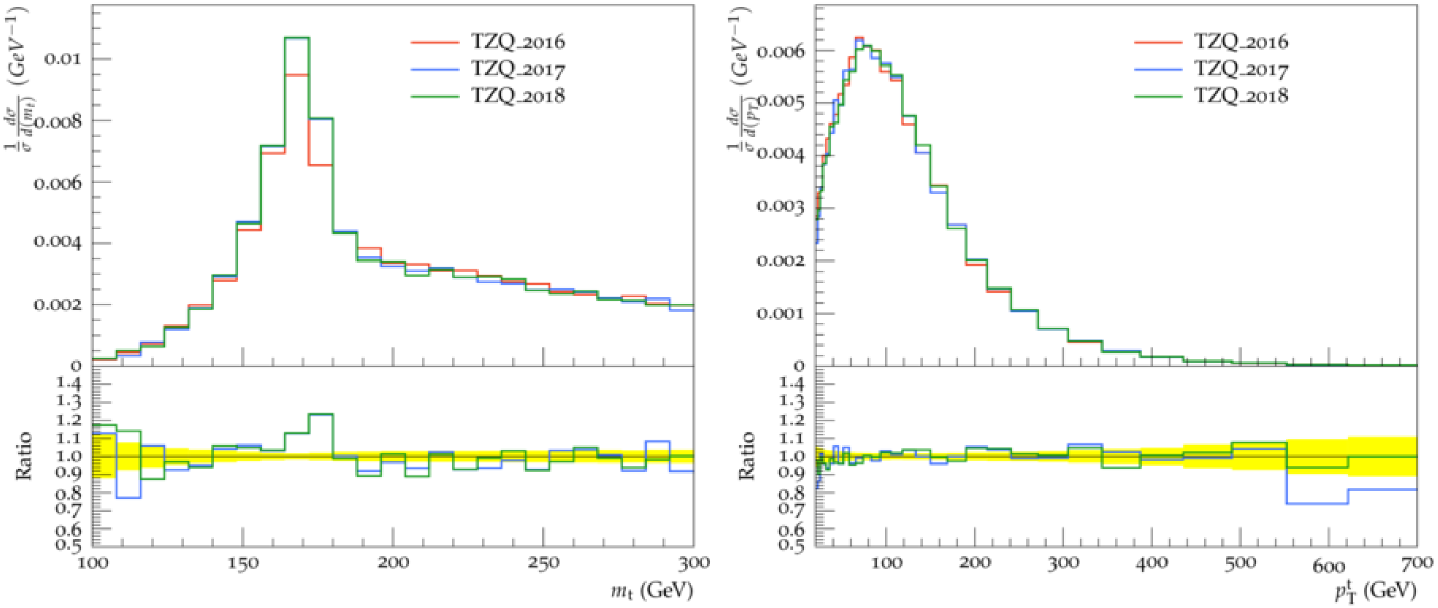


Figure 42: The mass (left) and transverse momenta (right) values of the top quark.

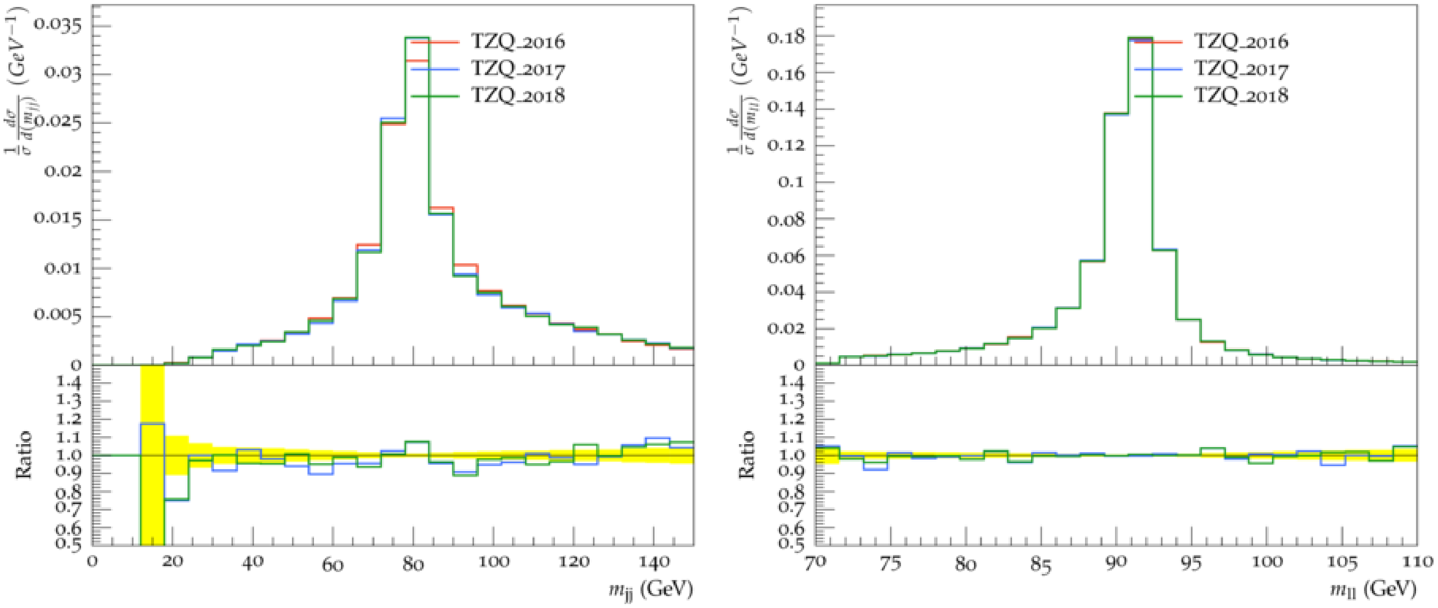


Figure 43: The masses of the W boson (left) and of the Z boson (right).

8 Vitis migration for the Phase 2 Tracker Level 1 Track Finding Algorithm

A track finding algorithm, called the Hybrid algorithm [75], is currently being developed in CMS. Steps of the algorithm are implemented in firmware by using a chain of "virtual modules". These are divided temporally, such that a new event is received in each 150 ns time slice (i.e. when a new event is received by a virtual module, the previous event is sent to the next one). A sector processor is used for each time slice, which is a hardware board containing FPGAs. Additionally for the modules, the CMS detector is subdivided spatially into 28 ϕ regions. Both the spatial and temporal subdivisions are applied to help achieve a suitable resource usage and latency.

The algorithm begins with a tracklet road search algorithm, in which "seeding" is performed. In this, pairs of stubs are identified in adjacent disks or layers by making projections to other disks and layers. A pair of stubs forms a "seed", which can also be referred to as a "tracklet".

The virtual modules operate as follows:

- The Input Router module sorts input stubs by layer.
- The VM router module then sorts the stubs into smaller units with respect to z and ϕ.
- The Tracklet Engine module selects possible pairs of stubs.
- The Tracklet Calculator calculates the tracklet parameters.
- The Projection Transceiver transmits projections that point to neighbouring sectors.
- The Projection Router then routes the projections based on their virtual module.
- The Match Engine matches projections to stubs.
- The Match Calculator calculates the tracklet stub difference in positions.
- The Match Receiver transmits matches back to the original sector.
- The Track Fit module fits final tracks.
- The Duplicate Removal module removals duplicate tracks.

The track finding chain firmware was previously written using HLS Vivado. The Match Engine module was optimised, with the aim of achieving a sufficient latency and resource usage under Vitis. In future HLS versions, Vivado will be replaced by Vitis. Like Vivado, Vitis can used as a standalone hardware design tool but it additionally has software capabilities to allow integration for more dynamic use. These reasons motivate the migration of the track finding chain to Vitis.

Comparisons of the performances of the algorithm under Vivado and Vitis for the Match Engine (ME) and Projection Router (PR) modules is shown in Figures 44- 47. The metrics compared are: the total number of clock cycles it takes to fully process one bunch crossing (ap_clk), the number of clock cycles to process one bunch crossing (latency), the total LUT (look-up table) estimate (where a LUT is an element that performs logic operations) and the total FF (flip flop) estimate (where a FF is register element that stores the result of an LUT).

For both the ME and PR virtual modules, the minimum (maximum) latency cycles were found to be 112 (113) for HLS Vivado and 324 (325) for Vitis. Additionally the interval, defined as the number of clock cycles between when two bunch crossings start, was found to be 108 and 324 under HLS Vivado and Vitis, respectively, for both the Match Engine and the Projection Router virtual modules. Since had interval of 300+ was observed under Vitis, further investigation is required as the interval is required to be 108 for the chain to function properly.

For the Match Engine module, *ap_clk* summaries above the 4ns target were observed under both HLS Vivado and Vitis, whereas *ap_clk* values for all solutions under Vitis were higher than that when under HLS Vivado. Additionally, latency summary values for all solutions under Vitis were higher than that when under HLS Vivado. The total LUT and FF estimates under Vitis were lower than that when under HLS Vivado.

For the Projection Router module, *ap_clk* summaries below the 4ns target were observed under both HLS Vivado and Vitis, whereas *ap_clk* values for all solutions under Vitis were higher than that when under HLS Vivado. Additionally, latency summary values for all solutions under Vitis were higher than that when under HLS Vivado. The total LUT and FF estimates under Vitis were lower than that when under HLS Vivado.

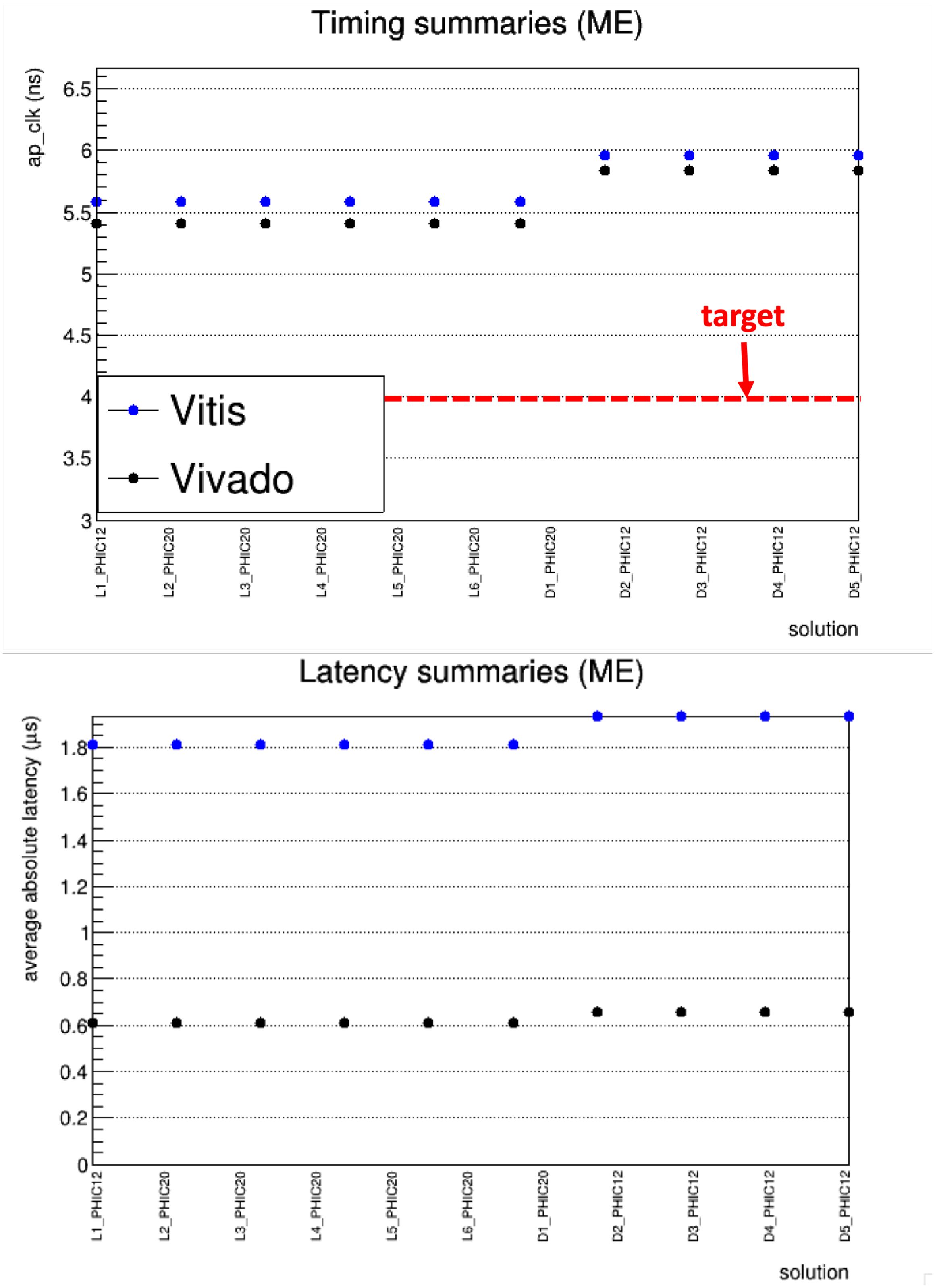


Figure 44: A comparison of the timing (top) and latency (bottom) summaries for various solutions for the Match Engine module. Here, results under Vitis are shown in blue and results under HLS Vivado are dhown in black.

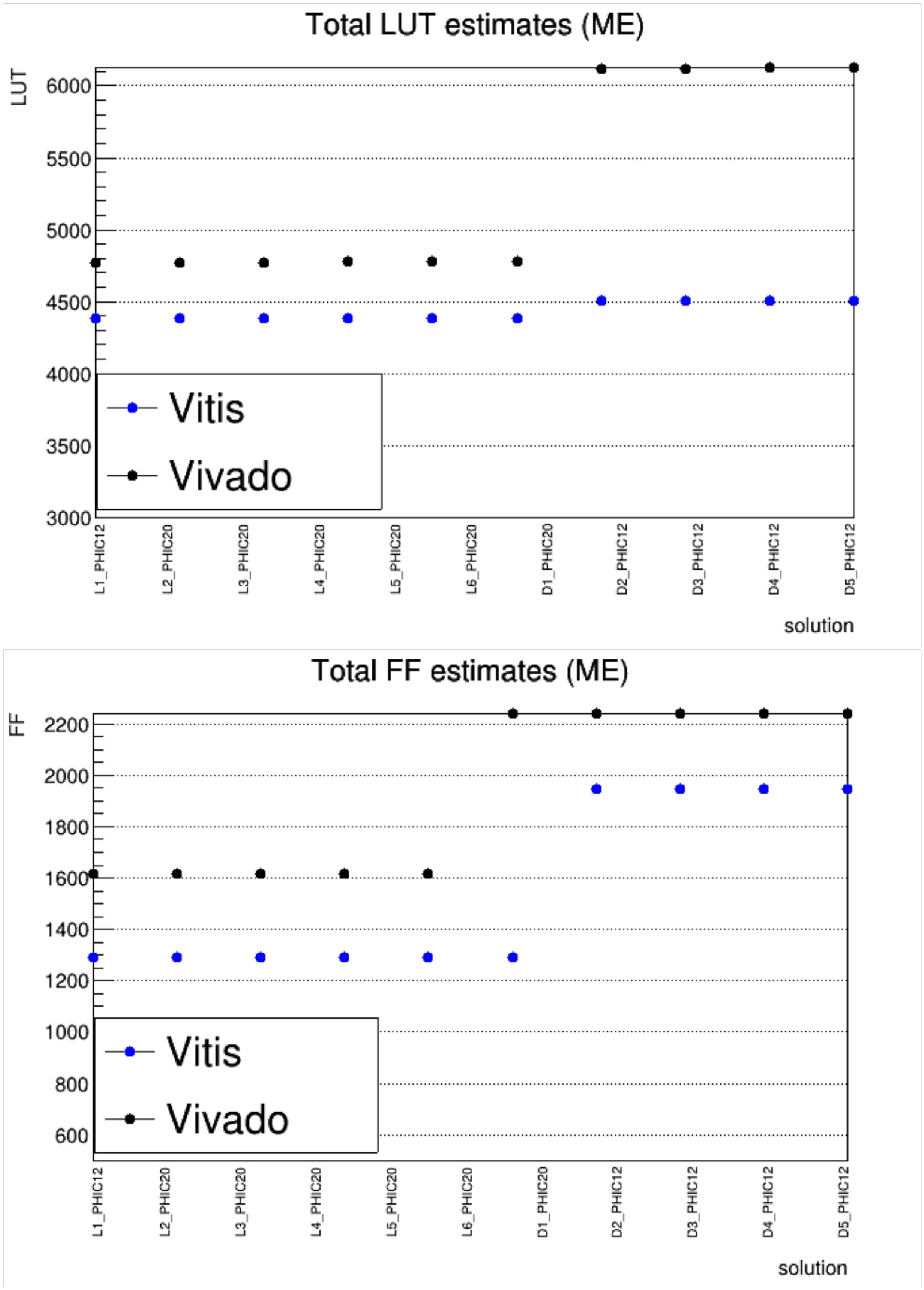
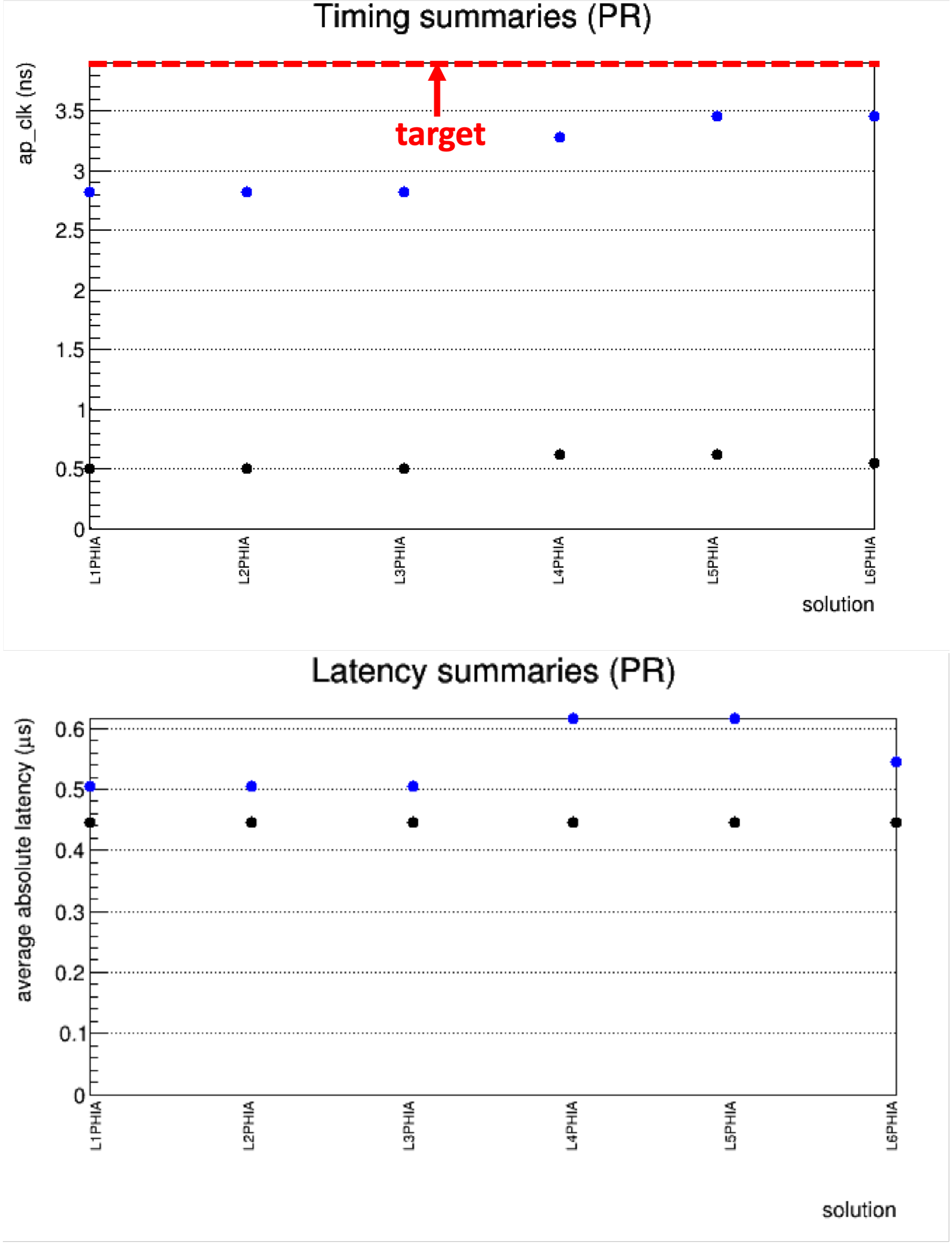


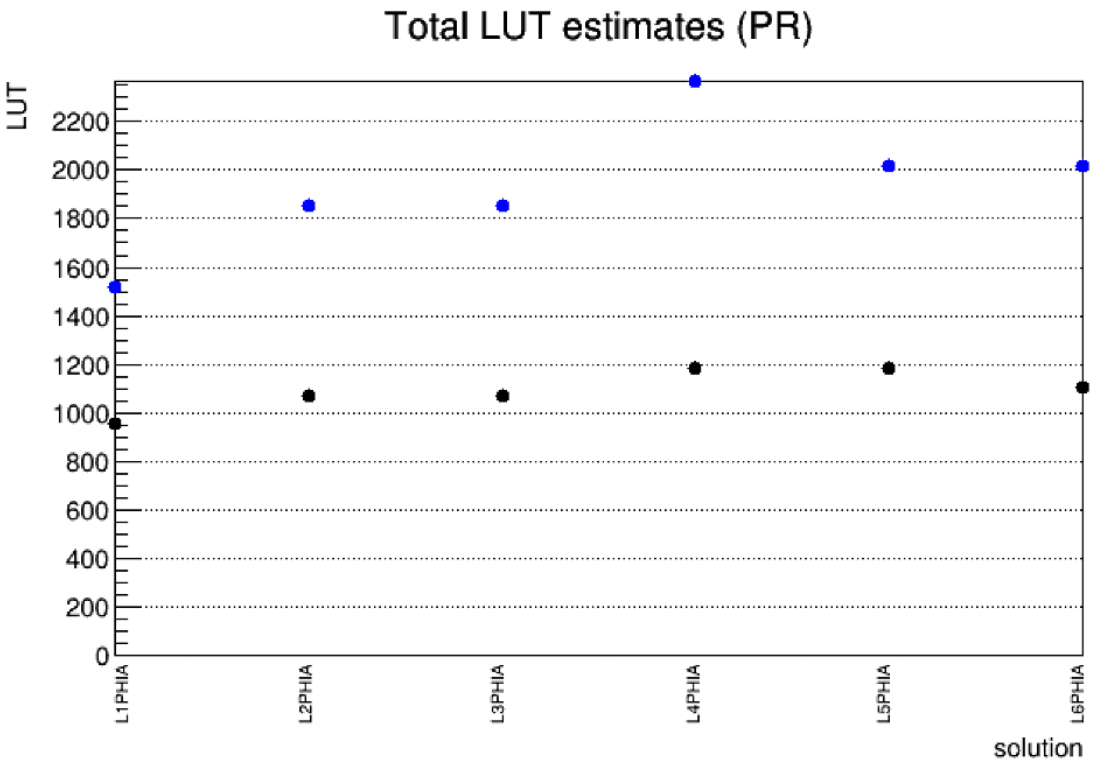
Figure 45: A comparison of the total LUT estimates (top) and the total FF estimates (bottom) for various solutions for the Match Engine module. Here, results under Vitis are shown in blue and results under HLS Vivado are dhown in black.



Timing summaries (PR)

Latency summaries (PR)

Figure 46: A comparison of the timing (top) and latency (bottom) summaries for various solutions for the Projection Router module. Here, results under Vitis are shown in blue and results under HLS Vivado are shown in black.



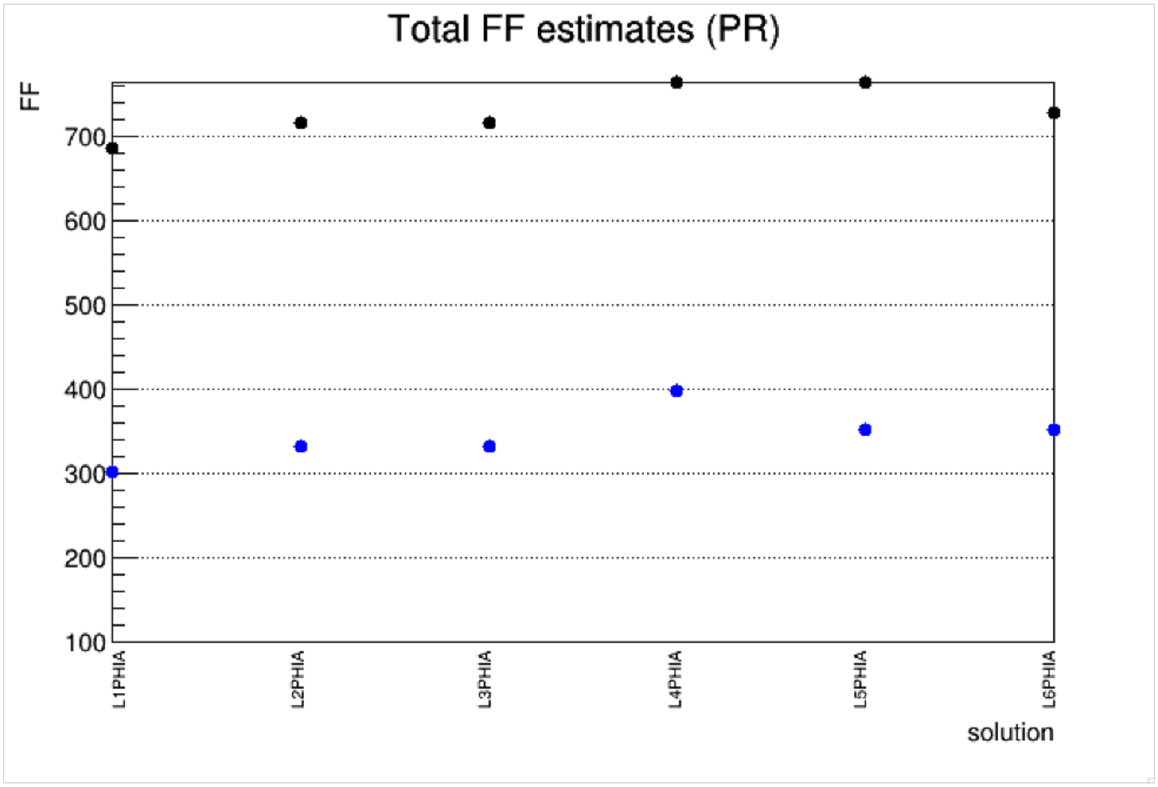


Figure 47: A comparison of the total LUT estimates (top) and the total FF estimates (bottom) for various solutions for the Projection Router module. Here, results under Vitis are shown in blue and results under HLS Vivado are shown in black.

9 Methodology: Search for tZq Production in Dilepton Final States

9.1 Datasets and Simulation Samples

Monte Carlo simulation samples were chosen based on the cross sections of the processes (how likely it is for a process to occur) and how similar the final state particles of the processes are to those in the signal sample. The samples following a naming convention, which is of the general format:

$$\boldsymbol{PROCESS\_RANGETYPE - RANGELOWToRANGEHIGH\_FILTER\_TUNE}$$
$$\boldsymbol{\_COMMENT\_COMENERGY - GENERATOR}$$

In this "PROCESS" is the name of the physics process that the sample corresponds to. The "RANGETYPE" refers to the variable that the sample is binned in, such as transverse momentum or mass, and "RANGELOW" and "RANGEHIGH" are the lower and upper bounds of this range, respectively. The "FILTER" corresponds to any additional filters that have been applied. The "TUNE" is the underlying event tune. In the sample name, additional comments may be provided in the "COMMENT" section and "COMENERGY" is the centre of mass energy. Finally "GENERATOR" is the generator used. The list of possible generators is shown in Table 3 [76].

Pythia6	pythia6
Pythia8	pythia8
Herwig6	herwig6
Herwig++	herwigpp
Herwig7	herwig7
Sherpa	sherpa
MadGraph/MG5$_aMC@NLO(LO)$	madgraph
MadGraph/MG5$_aMC@NLO(LO)$ $e.g.$ $showered$ $with$ $Pythia8$	madgraph-pythia8
MadGraph/MG5$_aMC@NLO(NLO)$	amcatnlo
Alpgen	alpgen
MC@NLO	mcatnlo
POWHEG	powheg
HARDCOL	hardcol
BCVEGPY 2	bcvegpy2

Table 3: The different types of generators used to generate monte carlo events and their corresponding keywords used to represent them in sample names [76].

Taking the example of the tZq signal sample used in this analysis for 2016:

$$\boldsymbol{/tZq\_ll\_4f\_ckm\_NLO\_TuneCP5\_PSweights\_13TeV - amcatnlo - pythia8/}$$

The PROCESS here is $\boldsymbol{tZq\_ll\_4f\_ckm\_NLO}$. The ll indicates dileptonic final states and the $4f$ represents that this sample corresponds to the four flavour scheme. The NLO represents that the process is at next-to-leading order.

The TUNE is TuneCP5; a tune defines a set of parameters in simulation that can be adjusted in order to better describe data [77]. A COMMENT has been included to indicate that the sample contains parton shower weights (PSWeights). The COMENERGY is 13 TeV and events were generated using the pythia8 generator.

The remainder of the sample name is called the chain. The $RunIISummer16NanoAODv7$ part indicates that the sample simulates events recorded in Summer 2016, which is during Run II, and the sample corresponds to the seventh version of the NanoAOD format, where improvements and fixes are made from one version to the next. In order of decreasing size, the ROOT file formats used for both data and simulation in CMS are RECO (3 MB per event), AOD (400-500 kB per event), MiniAOD [78] (35-60 kB per event) and NanoAOD (1-2 kB per event). For this analysis, simulation and data samples in the NanoAOD format were analysed. This is in contrast to previous analyses for the tZq dilepton decay channel [45] [46], for which samples in the MiniAOD format were used. The NanoAOD format is more than 20 times smaller than that of MiniAOD and only contains top-level information. Using this more compact format speeds up runtime and also negates the need for skimming samples, which is the process by which events are removed based on trigger bits or high-level reconstructed quantities [79]. Additionally, unlike in MiniAOD, in NanoAOD certain jet energy corrections are already applied; these are outlined in Section 9.3.4.

The naming convention for data samples is slightly different from that of simulation samples. Taking the example of the single electron dataset used in the tZq dilepton analysis:

$$\boldsymbol{/SingleElectron/Run2016C - Nano25Oct2019 - v1/NANOAOD}$$

Here, the final state contains a single electron and events were recorded in the data taking era of Run C in 2016. The version of the dataset is Nano25Oct2019-v1, and the dataset is in the NanoAOD format.

The full list of data and simulation samples analysed for the tZq dilepton analysis is shown in Tables 4-12

Process	Sample
tZq (signal)	tZq_ll_4f_13TeV-amcatnlo-pythia8
tZq (scale up)	tZq_ll_4f_scaleup_13TeV-amcatnlo-pythia8
tZq (scale down)	tZq_ll_4f_scaleup_13TeV-amcatnlo-pythia8
Z+jets (MC@NLO)	DYJetsToLL_M-50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8
	DYJetsToLL_M-10to50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8
	DYJetsToLL_M-10to50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8
Z+jets (madgraph)	DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8
	DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8
	DYJetsToLL_M-10to50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8
Z+jets (p_T-binned)	DYJetsToLL_Zpt-0To50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8
	DYJetsToLL_Pt-50To100_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8
	DYJetsToLL_Pt-100To250_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8
	DYJetsToLL_Pt-250To400_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8
	DYJetsToLL_Pt-400To650_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8
	DYJetsToLL_Pt-650ToInf_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8

Table 4: The tZq signal and Z+jets MC samples used for 2016.

Process	Sample
Single top (t-channel top)	ST_t-channel_top_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1
Single top (t-channel top, scale up)	ST_t-channel_top_4f_scaleup_inclusiveDecays_13TeV-powhegV2-madspin-pythia8
Single top (t-channel top, scale down)	ST_t-channel_top_4f_scaledown_inclusiveDecays_13TeV-powhegV2-madspin-pythia8
Single top (t-channel antitop)	ST_t-channel_antitop_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1
Single top (t-channel antitop, scale up)	ST_t-channel_antitop_4f_scaleup_inclusiveDecays_13TeV-powhegV2-madspin-pythia8
Single top (t-channel antitop, scale down)	ST_t-channel_antitop_4f_scaledown_inclusiveDecays_13TeV-powhegV2-madspin-pythia8
Single top (t-channel top, hdamp up)	ST_t-channel_top_4f_hdampup_InclusiveDecays$_{T}uneCP5\_PSweights\_13TeV - powheg - pythia8$
Single top (t-channel top, hdamp down)	ST_t-channel_top_4f_hdampdown_InclusiveDecays$_{T}uneCP5\_PSweights\_13TeV - powheg - pythia8$
Single top (t-channel antitop, hdamp up)	ST_t-channel_antitop_4f_hdampup_InclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8
Single top (t-channel antitop, hdamp down)	ST_t-channel_antitop_4f_hdampdown_InclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8
Single top (s-channel)	ST_s-channel_4f_InclusiveDecays_13TeV-amcatnlo-pythia8
Single top (tW)	ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1
Single top (tW, scale up)	ST_tW_top_5f_scaleup_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1
Single top (tW, scale down)	ST_tW_top_5f_scaledown_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1
Single top ($\bar{t}W$)	ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1
Single top ($\bar{t}W$, scale up)	ST_tW_antitop_5f_scaleup_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1
Single top ($\bar{t}W$, scale down)	ST_tW_antitop_5f_scaledown_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1
Single top (tHq)	THQ_Hincl_13TeV-madgraph-pythia8_TuneCUETP8M1
Single top (tWZ tll)	ST_tWll_5f_LO_13TeV-MadGraph-pythia8
Triboson (WWW)	WWW_4F_TuneCUETP8M1_13TeV-amcatnlo-pythia8
Triboson (WWZ)	WWZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8
Triboson (WZZ)	WZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8
Triboson (ZZZ)	ZZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8
Diboson ($ZZ \to 4l$)	ZZTo4L_13TeV-amcatnloFXFX-pythia8
Diboson ($ZZ \to 2l2\nu$)	ZZTo2L2Nu_13TeV_powheg_pythia8
Diboson ($ZZ \to 2l2Q$)	ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8
Diboson ($WW \to 2l2\nu$)	WWTo2L2Nu_13TeV-powheg
Diboson ($WZ \to 1l1\nu 2Q$)	WZTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin_pythia8
Diboson ($WW \to l\nu QQ$)	WWToLNuQQ_13TeV-powheg
	WWToLNuQQ_13TeV-powheg
Diboson ($WZ \to 3l\nu$)	WZJToLLLNu_TuneCUETP8M1_13TeV-amcnlo-pythia8
Diboson ($WZ \to 2l2Q$)	WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8
W+jets (W+jets$\to$ $l\nu$)	WJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8
	WJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/

Table 5: The single top, triboson, diboson and W+jets MC samples used for 2016.

Process	Sample
$t\bar{t}$ (madgraph)	TTJets_DiLept_TuneCUETP8M1_13TeV-madgraphMLM-pythia8
$t\bar{t}$ (MC@NLO)	TTJets_TuneCUETP8M2T4_13TeV-amcatnloFXFX-pythia8
$t\bar{t} > 2l2\nu$	TTTo2L2Nu_TuneCP5_PSweights_13TeV-powheg-pythia8
$t\bar{t}$ (hdamp up)	TT_hdampUP_TuneCUETP8M2T4_13TeV-powheg-pythia8
$t\bar{t}$ (hdamp down)	TT_hdampDOWN_TuneCUETP8M2T4_13TeV-powheg-pythia8
$t\bar{t}$ (isr up)	TT_TuneCUETP8M2T4_13TeV-powheg-isrup-pythia8
$t\bar{t}$ (isr down)	TT_TuneCUETP8M2T4_13TeV-powheg-isrdown-pythia8
$t\bar{t}$ (fsr up)	TT_TuneCUETP8M2T4_13TeV-powheg-fsrup-pythia8
$t\bar{t}$ (fsr down)	TT_TuneCUETP8M2T4_13TeV-powheg-fsrdown-pythia8
$t\bar{t}V$ ($ttZ \rightarrow QQ$)	TTZToQQ_TuneCUETP8M1_13TeV-amcatnlo-pythia8
$t\bar{t}V$ ($ttZ \rightarrow ll$)	TTZToLLNuNu_M-10_TuneCUETP8M1_13TeV-amcatnlo-pythia8
$t\bar{t}V$ ($ttZ \rightarrow ll\nu\nu$)	TTZToLLNuNu_M-10_TuneCUETP8M1_13TeV-amcatnlo-pythia8
$t\bar{t}V$ (ttW jets $\rightarrow l\nu$)	TTWJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8
$t\bar{t}V$ (ttW jets $\rightarrow QQ$)	TTWJetsToQQ_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8
$t\bar{t}V$ ($ttH \rightarrow bb$)	ttHTobb_M125_TuneCUETP8M2_ttHtranche3_13TeV-powheg-pythia8
$t\bar{t}V$ ($ttH \rightarrow$ non-bb)	ttHToNonbb_M125_TuneCUETP8M2_ttHtranche3_13TeV-powheg-pythia8

Table 6: The $t\bar{t}$ and $t\bar{t}V$ MC samples used for 2016.

Process	Sample
Single electron	/SingleElectron/Run2016B_ver2-Nano25Oct2019_ver2-v1/NANOAOD
	/SingleElectron/Run2016C-Nano25Oct2019-v1/NANOAOD
	/SingleElectron/Run2016D-Nano25Oct2019-v1/NANOAOD
	/SingleElectron/Run2016E-Nano25Oct2019-v1/NANOAOD
	/SingleElectron/Run2016F-Nano25Oct2019-v1/NANOAOD
	/SingleElectron/Run2016G-Nano25Oct2019-v1/NANOAOD
	/SingleElectron/Run2016H-Nano25Oct2019-v1/NANOAOD
Double electron	/DoubleEG/Run2016B-22Aug2018_ver2-v1/NANOAOD
	/DoubleEG/Run2016C-Nano25Oct2019-v1/NANOAOD
	/DoubleEG/Run2016D-Nano25Oct2019-v1/NANOAOD
	/DoubleEG/Run2016E-Nano25Oct2019-v1/NANOAOD
	/DoubleEG/Run2016F-Nano25Oct2019-v1/NANOAOD
	/DoubleEG/Run2016G-Nano25Oct2019-v1/NANOAOD
	/DoubleEG/Run2016H-Nano25Oct2019-v1/NANOAOD
Single muon	/SingleMuon/Run2016B_ver2-Nano25Oct2019_ver2-v1/NANOAOD
	/SingleMuon/Run2016C-Nano25Oct2019-v1/NANOAOD
	/SingleMuon/Run2016D-Nano25Oct2019-v1/NANOAOD
	/SingleMuon/Run2016E-Nano25Oct2019-v1/NANOAOD
	/SingleMuon/Run2016F-Nano25Oct2019-v1/NANOAOD
	/SingleMuon/Run2016G-Nano25Oct2019-v1/NANOAOD
	/SingleMuon/Run2016H-Nano25Oct2019-v1/NANOAOD
Double muon	/DoubleMuon/Run2016B_ver2-Nano14Dec2018_ver2-v1/NANOAOD
	/DoubleMuon/Run2016C-Nano25Oct2019-v1/NANOAOD
	/DoubleMuon/Run2016D-Nano25Oct2019-v1/NANOAOD
	/DoubleMuon/Run2016E-Nano25Oct2019-v1/NANOAOD
	/DoubleMuon/Run2016F-Nano25Oct2019-v1/NANOAOD
	/DoubleMuon/Run2016G-Nano25Oct2019-v1/NANOAOD
	/DoubleMuon/Run2016H-Nano25Oct2019-v1/NANOAOD
Muon-electron	/MuonEG/Run2016B_ver2-Nano14Dec2018_ver2-v1/NANOAOD
	/MuonEG/Run2016C-Nano14Dec2018-v1/NANOAOD
	/MuonEG/Run2016D-Nano14Dec2018-v1/NANOAOD
	/MuonEG/Run2016E-Nano14Dec2018-v1/NANOAOD
	/MuonEG/Run2016F-Nano14Dec2018-v1/NANOAOD
	/MuonEG/Run2016G-Nano14Dec2018-v1/NANOAOD
	/MuonEG/Run2016H-Nano14Dec2018-v1/NANOAOD

Table 7: The datasets used for 2016.

Process	Sample
tZq (signal)	tZq_ll_4f_ckm_NLO_TuneCP5_PSweights_13TeV-amcatnlo-pythia8
Z+jets (MC@NLO)	DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8
	DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8
Single top (t-channel, top)	ST_t-channel_top_4f_InclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8
Single top (t-channel, antitop)	ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8
Single top (s-channel)	ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-amcatnlo-pythia8
Single top (tW)	ST_tW_top_5f_inclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8
Single top ($\bar{t}W$)	ST_tW_antitop_5f_inclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8
Single top (tZq, hadronic Z and leptonic W)	tZq_W_lept_Z_hadron_4f_ckm_NLO_13TeV_amcatnlo_pythia8
Single top (tHq)	THQ_4f_Hincl_13TeV_madgraph_pythia8
Single top ($tWZtll$)	ST_tWll_5f_LO_TuneCP5_PSweights_13TeV-madgraph-pythia8
Tribosonic (WWW)	WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8
Tribosonic (WWZ)	WWZ_4F_TuneCP5_13TeV-amcatnlo-pythia8
Tribosonic (WZZ)	WZZ_TuneCP5_13TeV-amcatnlo-pythia8
Tribosonic (ZZZ)	ZZZ_TuneCP5_13TeV-amcatnlo-pythia8
Dibosonic ($ZZ \rightarrow 4l$)	ZZTo4L_13TeV_powheg_pythia8
Dibosonic ($ZZ \rightarrow 2l2\nu$)	ZZTo2L2Nu_13TeV_powheg_pythia8
Dibosonic ($ZZ \rightarrow 2l2Q$)	ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8
Dibosonic ($WW \rightarrow 2l2\nu$)	WWTo2L2Nu_NNPDF31_TuneCP5_PSweights_13TeV-powheg-pythia8
Dibosonic ($WW \rightarrow 1l1\nu2Q$)	WWTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin_pythia8
Dibosonic ($WW \rightarrow l\nu QQ$)	WWToLNuQQ_NNPDF31_TuneCP5_PSweights_13TeV-powheg-pythia8
Dibosonic ($WZ \rightarrow 3l\nu$)	WZTo3LNu_TuneCP5_13TeV-amcatnloFXFX-pythia8
Dibosonic ($WZ \rightarrow 2l2Q$)	WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8
Dibosonic ($WZ \rightarrow 1l1\nu2Q$)	WZTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin_pythia8

Table 8: The tZq, Z+jets, single top, tribosonic and dibosonic samples used for 2017.

Process	Sample
$t\bar{t}$ (madgraph)	TTJets_TuneCP5_13TeV-amcatnloFXFX-pythia8
$t\bar{t}$ ($t\bar{t} \rightarrow$ hadronic)	TTToHadronic_TuneCP5_PSweights_13TeV-powheg-pythia8
$t\bar{t}$ ($t\bar{t} \rightarrow$ semileptonic)	TTToSemiLeptonic_TuneCP5_PSweights_13TeV-powheg-pythia8
$t\bar{t}$ (MC@NLO)	TTJets_TuneCP5_13TeV-amcatnloFXFX-pythia8
$t\bar{t}$ ($t\bar{t} \rightarrow> 2l2\nu$)	TTTo2L2Nu_TuneCP5_PSweights_13TeV-powheg-pythia8
$t\bar{t}V$ ($tt\gamma$)	TTGamma_Dilept_TuneCP5_PSweights_13TeV_madgraph_pythia8
$t\bar{t}V$ ($ttZ \rightarrow QQ$)	TTZToQQ_TuneCP5_13TeV-amcatnlo-pythia8
$t\bar{t}V$ ($ttZ \rightarrow ll$)	TTZToLL_M-1to10_TuneCP5_13TeV-amcatnlo-pythia8
$t\bar{t}V$ ($ttZ \rightarrow ll\nu\nu$)	TTZToLLNuNu_M-10_TuneCP5_13TeV-amcatnlo-pythia8
$t\bar{t}V$ (ttW jets $\rightarrow l\nu$)	TTWJetsToLNu_TuneCP5_PSweights_13TeV-amcatnloFXFX-madspin-pythia8
$t\bar{t}V$ (ttW jets $\rightarrow QQ$)	TTWJetsToQQ_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8
$t\bar{t}V$ ($ttH \rightarrow bb$)	ttHTobb_M125_TuneCP5_13TeV-powheg-pythia8
$t\bar{t}V$ ($ttH \rightarrow$ non-bb)	ttHToNonbb_M125_TuneCP5_13TeV-powheg-pythia8
W+jets	WJetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8

Table 9: The $t\bar{t}$, $t\bar{t}V$ and W+jets MC samples used for 2017.

Process	Sample
Single electron	/SingleElectron/Run2017B-Nano25Oct2019-v1/NANOAOD
	/SingleElectron/Run2017C-Nano25Oct2019-v1/NANOAOD
	/SingleElectron/Run2017D-Nano25Oct2019-v1/NANOAOD
	/SingleElectron/Run2017E-Nano25Oct2019-v1/NANOAOD
	/SingleElectron/Run2017F-Nano25Oct2019-v1/NANOAOD
Double electron	/DoubleEG/Run2017B-Nano14Dec2018-v1/NANOAOD
	/DoubleEG/Run2017C-Nano14Dec2018-v1/NANOAOD
	/DoubleEG/Run2017D-Nano14Dec2018-v1/NANOAOD
	/DoubleEG/Run2017E-Nano14Dec2018-v1/NANOAOD
	/DoubleEG/Run2017F-Nano14Dec2018-v1/NANOAOD
Single muon	/SingleMuon/Run2017B-Nano25Oct2019-v1/NANOAOD
	/SingleMuon/Run2017C-Nano25Oct2019-v1/NANOAOD
	/SingleMuon/Run2017D-Nano25Oct2019-v1/NANOAOD
	/SingleMuon/Run2017E-Nano25Oct2019-v1/NANOAOD
	/SingleMuon/Run2017F-Nano25Oct2019-v1/NANOAOD
Double muon	/DoubleMuon/Run2017B-Nano14Dec2018-v1/NANOAOD
	/DoubleMuon/Run2017C-Nano14Dec2018-v1/NANOAOD
	/DoubleMuon/Run2017D-Nano14Dec2018-v1/NANOAOD
	/DoubleMuon/Run2017E-Nano14Dec2018-v1/NANOAOD
	/DoubleMuon/Run2017F-Nano14Dec2018-v1/NANOAOD
Muon-electron	/MuonEG/Run2017B-Nano14Dec2018-v1/NANOAOD
	/MuonEG/Run2017C-Nano14Dec2018-v1/NANOAOD
	/MuonEG/Run2017D-Nano14Dec2018-v1/NANOAOD
	/MuonEG/Run2017E-Nano14Dec2018-v1/NANOAOD
	/MuonEG/Run2017F-Nano14Dec2018-v1/NANOAOD

Table 10: The datasets used for 2017.

Process	Sample
tZq (signal)	tZq_ll_4f_ckm_NLO_TuneCP5_13TeV-madgraph-pythia8
Z+jets	DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8
	DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8
Single top (t-channel, t)	t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8
Single top (t-channel, $\bar{t}$)	t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8
Single top s-channel	ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-madgraph-pythia8
Single top (tW)	ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8
Single top ($\bar{t}W$)	ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8
Single top (tZq)	tZq_Zhad_Wlept_4f_ckm_NLO_TuneCP5_PSweights_13TeV-amcatnlo-pythia8
Single top (tHq)	THQ_4f_Hincl_13TeV_madgraph_pythia8
Single top ($tWZtll$)	ST_tWll_5f_LO_TuneCP5_PSweights_13TeV-madgraph-pythia8
Tribosonic (WWW)	WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8
Tribosonic (WWZ)	WWZ_TuneCP5_13TeV-amcatnlo-pythia8
Tribosonic (WZZ)	WZZ_TuneCP5_13TeV-amcatnlo-pythia8
Tribosonic (ZZZ)	ZZZ_TuneCP5_13TeV-amcatnlo-pythia8
Dibosonic ($ZZ \rightarrow 4l$)	ZZTo4L_TuneCP5_13TeV_powheg_pythia8
Dibosonic ($ZZ \rightarrow 2l2\nu$)	ZZTo2L2Nu_TuneCP5_13TeV_powheg_pythia8
Dibosonic ($ZZ \rightarrow 2l2Q$)	ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8
Dibosonic ($WW \rightarrow 2l2\nu$)	WWTo2L2Nu_NNPDF31_TuneCP5_13TeV-powheg-pythia8
Dibosonic ($WW \rightarrow 1l1\nu2Q$)	WWToLNuQQ_NNPDF31_TuneCP5_13TeV-powheg-pythia8
Dibosonic ($WZ \rightarrow 3l\nu$)	WZTo3LNu_TuneCP5_13TeV-amcatnloFXFX-pythia8
Dibosonic ($WZ \rightarrow 2l2Q$)	WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8
$t\bar{t}$ (madgraph)	TTJets_DiLept_TuneCP5_13TeV-madgraphMLM-pythia8
$t\bar{t}$ ($t\bar{t} \rightarrow$ hadronic)	TTToHadronic_TuneCP5_13TeV-powheg-pythia8
$t\bar{t}$ ($t\bar{t} \rightarrow$ semileptonic)	TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8
$t\bar{t}$ (MC@NLO)	TTJets_TuneCP5_13TeV-amcatnloFXFX-pythia8
$t\bar{t}$ ($t\bar{t} > 2l2\nu$)	TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/
$t\bar{t}V$ ($tt\gamma$)	TTGamma_Dilept_TuneCP5_13TeV-madgraph-pythia8
$t\bar{t}V$ ($ttZ \rightarrow QQ$)	TTZToQQ_TuneCP5_13TeV-amcatnlo-pythia8
$t\bar{t}V$ ($ttZ \rightarrow ll$)	TTZToLL_M-1to10_TuneCP5_13TeV-amcatnlo-pythia8
$t\bar{t}V$ ($ttZ \rightarrow ll\nu\nu$)	TTZToLLNuNu_M-10_TuneCP5_13TeV-amcatnlo-pythia8
$t\bar{t}V$ (ttW jets $\rightarrow l\nu$)	TTWJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8
$t\bar{t}V$ (ttW jets $\rightarrow QQ$)	TTWJetsToQQ_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8
$t\bar{t}V$ ($ttH \rightarrow bb$)	ttHTobb_M125_TuneCP5_13TeV-powheg-pythia8
$t\bar{t}V$ ($ttH \rightarrow$ non-bb)	ttHToNonbb_M125_TuneCP5_13TeV-powheg-pythia8
W+jets	WJetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8

Table 11: The tZq, Z+jets, single top, tribosonic, dibosonic, $t\bar{t}$, $t\bar{t}V$ and W+jets MC samples used for 2018.

Process	Sample
Single electron and double electron	/EGamma/Run2018B-Nano25Oct2019-v1/NANOAOD /EGamma/Run2018C-Nano25Oct2019-v1/NANOAOD /EGamma/Run2018D-Nano25Oct2019-v1/NANOAOD
Single muon	/SingleMuon/Run2018B-Nano25Oct2019-v1/NANOAOD /SingleMuon/Run2018C-Nano25Oct2019-v1/NANOAOD /SingleMuon/Run2018D-Nano25Oct2019-v1/NANOAOD
Double muon	/DoubleMuon/Run2018B-Nano14Dec2018-v1/NANOAOD /DoubleMuon/Run2018C-Nano14Dec2018-v1/NANOAOD /DoubleMuon/Run2018D-Nano25Oct2019_ver2-v1/NANOAOD

Table 12: The datasets used for 2018.

9.2 Event Selection and Experimental Blinding

Selection criteria were applied to filter events such that only events that contain a pair of final-state lepton candidates and a pair of final-state quark candidates remain. The lepton candidate pair was used to reconstruct the Z boson candidate, while a pair of quark candidates was used to reconstruct the W boson candidate. The W boson and b-jet candidates were then used to reconstruct the top quark candidate.

The analysis was conducted in four different regions: the signal region (SR), the $t\bar{t}$ and Z+jets control regions ($t\bar{t}$ CR and Z+jets CR, respectively) and the side-band region (SBR). The definitions for all four regions are explained in this section. The SR is enriched in tZq signal events and, for this region, the following selection criteria were applied to the samples of 2016:

Lepton Selection Criteria

For 2016 samples, only events containing a total of two oppositely-charged, same flavour final state particle flow leptons are selected. Events that do not satisfy this condition are vetoed. Lepton cut-based identification (ID) criteria are applied, which are defined based on their average selection efficiency in simulation. In order of increasing purity and decreasing mis-identification rate, the four electron cut-based ID criteria are: veto, loose, medium and tight, as defined by the CMS EGamma and Muon Physics Object Groups [80][81]. Events must contain leptons that satisfy the tight electron ID criterion, which has a 70% efficiency. The veto electron ID criterion has an efficiency of 90%.

For the 2016 analysis, final state leading (subleading) electrons were selected with $p_T > 35$ (15) GeV and $|\eta| < 2.4$. Final state leading (subleading) muons were selected with with $p_T > 26$ (20) GeV and $|\eta| < 2.5$.

To ensure that electrons originate from the detector interaction point, criteria are imposed on the transverse impact parameter, dxy, and the longitudinal impact parameter, dz. Tight electrons in the barrel region ($|\eta| < 1.479$) must have a $dz < 0.1$ cm and a $dxy < 0.05$ cm, while tight electrons in the endcap regions ($1.479 < |(\eta)| < 3.0$) must have a $dz < 0.2$ cm and a $dxy < 0.1$ cm.

For the 2017 and 2018 samples the same selection criteria were applied as above, with the exception of the leading electron requiring a $p_T > 38$ GeV and the leading muon requiring $p_T > 29$ GeV. This is due to increased trigger thresholds between 2016 and 2017.

Jet Selection Criteria

For the 2016, 2017 and 2018 samples, final state jets were selected with $p_T > 30$ GeV, within the range of $|\eta| < 4.7$ and with an anti-k_T parameter of 0.4. Thus, the selected jets are AK4 jets. Only events containing a total number of jets between 4 and 6 were selected. Events with a total outside of this range were vetoed. Jets must satisfy the tight ID criterion; those that satisfy the loose ID criterion are vetoed.

Selection Criteria for B-Tagged Jets

For the 2016, 2017 and 2018 samples, only events containing 1 or 2 b-tagged jets are selected, using the Combined Secondary Vertex Version 2 (CSVv2) b-tagging algorithm [82]. The CSVv2 algorithm was chosen for use in this analysis as it was the recommended algorithm for use by the CMS B-Tagging and Vertexing group. The algorithm uses a multivariate analysis technique to combine displaced track information with information about secondary vertices that are associated with the jet. The loose, medium and tight working points correspond to values of 58%, 88% and 97%, respectively [83]. The b-tagged jet candidates were required to pass the medium working point in this analysis and events that did not satisfy this condition were vetoed. For 2016, b-tagged jets were selected within the full detectable range of the CMS tracker, $|\eta| < 2.4$. For 2017 and 2018, b-tagged jets were selected within $|\eta| < 2.5$ due to the increased tracker acceptance.

Reconstruction of the Z Boson Candidate

For the 2016, 2017 and 2018 samples, the Z boson was reconstructed from the leading and subleading leptons. The combined leading and subleading lepton were required to have a reconstructed mass within 20 GeV of the known Z boson mass m_Z, where $m_Z = 91.2$ GeV.

Reconstruction of the W Boson Candidate

For the 2016, 2017 and 2018 samples, the W boson was reconstructed from the pair of final state jets with a mass closest to the known W boson mass m_W, where $m_W = 80.4$ GeV. The final state jets used to reconstruct the W excludes the leading b jet. A mass window of 60 GeV $< m_W <$ 100 GeV was applied.

Reconstruction of the Top Quark Candidate

For the 2016, 2017 and 2018 samples, the top quark candidate was reconstructed from the reconstructed W boson candidate and the leading b-tagged jet with a reconstructed mass closest to the known top quark mass of 173.3 GeV.

Additional Criteria

The additional selection criteria applied to the 2016, 2017 and 2018 samples are as follows:

- Jets and leptons were required to have a ΔR separation > 0.4.
- The detector region of $1.4442 \leq |\eta| \leq 1.566$ is a transition region between the ECAL barrel and endcap regions. Therefore, electron candidates from this region are not considered since this region contains material that cannot provide accurate detection of physics objects.

The Z+jets and $t\bar{t}$ Control Regions

The $t\bar{t}$ and Z+jets backgrounds form irreducible backgrounds to the signal sample and are predicted to be the dominant sources of background due to their large cross sections. This motivates the need to ensure that the $t\bar{t}$ and $Z + jets$ background simulation samples are correctly modelled. The control regions are enriched with $t\bar{t}$ and $Z + jets$ background events, and are topologically similar to the signal region but orthogonal to it. The $t\bar{t}$ control region contains $t\bar{t}$ events while the Z+jets control region contains events with a Z boson and one or more jets. To form the two control regions, the selection criteria applied are the same as above but with the following exceptions:

- $t\bar{t}$ CR – Lepton selection changed: exactly one tight electron and one tight muon are required with $p_T > 25$ GeV. Additional loose leptons are vetoed.
- Z+jets CR – The W boson mass requirement is inverted: no pair of jets may have an invariant mass within 20 GeV of the nominal W boson mass, excluding the leading b jet. The event must have less than 50 GeV of missing transverse energy.

Experimental Blinding

The side-band region is used to blind the analysis. Experimental blinding is carried out to prevent bias when choosing selection cuts and the methodology applied in this analysis was based on that applied in $HH \rightarrow b\bar{b}$ analyses. The methodology is implemented by using the χ^2 variable shown in equation 97. In this equation, m_W^{rec} represents mass of reconstructed W boson candidate, m_{Top}^{rec} represents the mass of the reconstructed top candidate, m_W is the accepted value of the W mass (80.3 GeV), m_{Top} is the accepted value of the top quark mass (173.5 GeV), σ_W is the resolution of the weighted mass distribution of the reconstructed W boson candidate after a Gaussian fit has been applied and σ_{Top} is the resolution of the weighted mass distribution of the reconstructed top quark candidate after a Gaussian fit has been applied.

$$\chi^2 = \left(\frac{m_W^{rec} - m_W^2}{\sigma_W}\right)^2 + \left(\frac{m_{Top}^{rec} - m_{Top}^2}{\sigma_{Top}}\right)^2 \tag{97}$$

A χ^2 distribution was obtained using nominal tZq MC events. Sub-ranges of χ^2 values were then obtained to define the signal and side band regions. Their definitions are as follows:

- A sub-range of χ^2 values was identified, for which the reconstructed W mass used to calculate them was within $5\sigma_W$ of the known W mass of 80.3 GeV.
- The maximum χ^2 in this sub-range, χ^2_{Max} was identified and the side-band region was defined as events with $\chi^2 < \chi^2_{Max}$
- The first sub-range of χ^2 was used to obtain a second sub-range of χ^2 values that was used to define the signal region. Distributions of χ^2 values were iterated over for all channels and a cumulative total for the number of events was obtained. The χ^2 value that corresponded to where the total number of events was 68% of the total number of χ^2 events in the first sub-range was identified. This χ^2 value is represented by χ^2_{Min} and the signal region is defined as events with $\chi^2_{Min} < \chi^2 < \chi^2_{Max}$.

The signal and side-band regions were defined to be $\chi^2 < 5$ and $5 < \chi^2 < 30$ respectively, as was calculated in previous work for the tZq dilepton analysis [45] [46]. For this thesis, the analysis was conducted in the side-band region and blinded results were obtained. After permission is granted by CMS, unblinded results can be obtained in the signal region.

9.3 Simulation Corrections

9.3.1 Lepton Efficiency

The total lepton efficiency is shown in Equation 98, where $\epsilon_{reconstruction}$, $\epsilon_{identification}$, $\epsilon_{trigger}$ and $\epsilon_{isolation}$ represent the reconstruction, identification and trigger efficiencies, respectively.

$$\epsilon_{total} = \epsilon_{reconstruction} \times \epsilon_{identification} \times \epsilon_{trigger} \times \epsilon_{isolation} \tag{98}$$

The lepton reconstruction, identification and isolation efficiencies and scale factors are dependent on p_T and η and are provided by the CMS collaboration [84]. The lepton trigger efficiencies and scale factors are calculated for each of the possible final states of the tZq dilepton channel: ee, $\mu\mu$ and $e\mu$. To achieve this, the 'cross trigger' method was applied, which is based on the method outlined in a previous $t\bar{t}$ analysis [85]. Cross triggers are defined as triggers that are weakly correlated to the high-level lepton triggers used in this analysis. The high-level lepton triggers chosen are shown in Tables 13-15. Triggers for the missing transverse energy (MET) were selected as the cross triggers; these are shown in Table 16.

The high-level lepton triggers follow the general format of:

HLT_ObjectMinPt_AdditionalCriteria

HLT stands for high-level trigger; the high-level trigger system was previously described in Section 4.2.7. The Object refers to the objects that the trigger paths select (e.g. Ele for electrons, Mu for global muons or TkMu for tracker muons). The MinPt is the minimum transverse momentum threshold, expressed in GeV. Additional criteria can then be given, such as particle identification (Id) or isolation (Iso) requirements. These requirements rely on data from the tracker (Trk) or calorimeters (Calo), and can be loose (L), very loose (VL) or very very loose (VVL), in order of increasing tightness. The term DZ may be given at the end of the path name to indicate that the HLT rejects objects that do not originate from the interation point.

In the cross trigger method, the number of MC events that pass the lepton selection criteria was calculated for each channel. Then, the number of MC events that passed the lepton triggers (shown in Tables 13, 14 and 15 for 2016, 2017 and 2018, respectively) were calculated. After this, the number of MC events that passed the MET triggers (shown in Table 16) was calculated. Once these values were obtained for each channel, the number of MC events that passed both the event selection and the lepton triggers were calculated. Then, the number of MC events that passed both the lepton selection and the MET triggers were calculated. Finally, the number of MC events that passed the lepton selection, the lepton triggers and the MET triggers were calculated. Data events with bad lumisections were filtered. A lumisection is a set of events in data, where an event corresponds to one beam crossing [86]. A lumisection corresponds to approximately 23 seconds of data taking. After filtering bad lumisections, the above steps are repeated.

Year	Trigger paths
2016	HLT_Ele25_eta2p1_WPTight_Gsf (e)
	HLT_Ele32_eta2p1_WPTight_Gsf (e)
	HLT_Ele27_WPTight_Gsf (e)
	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ (ee)
	HLT_IsoMu24 (μ)
	HLT_IsoMu24_eta2p1 (μ)
	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ ($\mu\mu$)
	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ ($e\mu$)
	HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ ($e\mu$)
	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ ($e\mu$)
	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL ($e\mu$)
	HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL ($e\mu$)
	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL ($e\mu$)

Table 13: The single electron (e), double electron (ee), single muon (μ), double muon ($\mu\mu$) and electron-muon ($e\mu$) triggers used in the cross trigger method for 2016.

Year	Triggers
2017	HLT_Ele35_WPTight_Gsf (e)
	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL (ee)
	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ (ee)
	HLT_IsoMu24 (μ)
	HLT_IsoMu27 (μ)
	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ ($\mu\mu$)
	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8 ($\mu\mu$)
	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL$_DZ$(eμ)
	HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ ($e\mu$)
	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ ($e\mu$)

Table 14: The single electron (e), double electron (ee), single muon (μ), double muon ($\mu\mu$) and electron-muon ($e\mu$) triggers used in the cross trigger method for 2017.

Year	Triggers
2018	HLT_Ele32_WPTight_Gsf_L1DoubleEG (e) HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL (ee) HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ (ee) HLT_IsoMu24 (μ) HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8 ($\mu\mu$) HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8 ($\mu\mu$) HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ ($e\mu$) HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ ($e\mu$) HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ ($e\mu$)

Table 15: The single electron (e), double electron (ee), single muon (μ), double muon ($\mu\mu$) and electron-muon ($e\mu$) triggers used in the cross trigger method for 2018.

Year	MET Trigger
2016	HLT_MET200
	HLT_MET250
	HLT_PFMET120_PFMHT120_IDTight
	HLT_PFMET170_HBHECleaned
	HLT_PFHT300_PFMET100
2017	HLT_PFMET120_PFMHT120_IDTight
	HLT_PFMET130_PFMHT130_IDTight
	HLT_PFMET140_PFMHT140_IDTight
	HLT_PFMETNoMu120_PFMHTNoMu120_IDTight
	HLT_PFHT1050
	HLT_PFHT180
	HLT_PFHT500_PFMET100_PFMHT100_IDTight
	HLT_PFHT500_PFMET110_PFMHT110_IDTight
	HLT_PFHT700_PFMET85_PFMHT85_IDTight
	HLT_PFHT700_PFMET95_PFMHT95_IDTight
	HLT_PFHT800_PFMET75_PFMHT75_IDTight
	HLT_PFHT800_PFMET85_PFMHT85_IDTight
2018	HLT_PFMET120_PFMHT120_IDTight
	HLT_PFMET130_PFMHT130_IDTight
	HLT_PFMET140_PFMHT140_IDTight
	HLT_PFMETNoMu120_PFMHTNoMu120_IDTight
	HLT_PFHT1050
	HLT_PFHT180
	HLT_PFHT500_PFMET100_PFMHT100_IDTight
	HLT_PFHT500_PFMET110_PFMHT110_IDTight
	HLT_PFHT700_PFMET85_PFMHT85_IDTight
	HLT_PFHT700_PFMET95_PFMHT95_IDTight
	HLT_PFHT800_PFMET75_PFMHT75_IDTight
	HLT_PFHT800_PFMET85_PFMHT85_IDTight

Table 16: The MET triggers used for 2016, 2017 and 2018 data and MC for the cross trigger method.

Equation 99 was used to determine if the chosen triggers show a weak correlation to “cross triggers”. The strength of the correlation is represented by α, ϵ_{trgLL}^{MC} represents the efficiency for passing the lepton trigger, $\epsilon_{trigMissE_T}^{MC}$ is the efficiency of passing the missing E_T trigger and $\epsilon_{trgLL,trgMissE_T}^{MC}$ represents the efficiency of passing both the lepton and MET triggers. A value of $\alpha = 1$ indicates that there is no correlation between the lepton triggers and the cross triggers because the triggers being independent would yield the expression shown in Equation 100.

$$\alpha = \frac{\epsilon_{trgLL}^{MC} \times \epsilon_{trigMissE_T}^{MC}}{\epsilon_{trgLL,trgMissE_T}^{MC}} \tag{99}$$

$$\epsilon_{trgLL,trgMissE_T}^{MC} = \epsilon_{trgLL}^{MC} \times \epsilon_{trigMissE_T}^{MC} \tag{100}$$

MET triggers are used as the cross triggers since they should have no dependence on lepton production and should show a weak correlation with the dilepton triggers. A weak correlation must be displayed because, to determine the lepton trigger efficiencies in data, the data sample must show no bias to the trigger efficiency. Additionally, MET triggers provide sufficient statistics to give a low statistical uncertainty. The MET datasets used in the cross trigger method are shown in Table 17.

Year	MET Dataset Names
2016	/MET/Run2016B_ver2-Nano25Oct2019_ver2-v1/NANOAOD /MET/Run2016C-Nano25Oct2019-v1/NANOAOD /MET/Run2016D-Nano25Oct2019-v1/NANOAOD /MET/Run2016E-Nano25Oct2019-v1/NANOAOD /MET/Run2016F-Nano25Oct2019-v1/NANOAOD /MET/Run2016G-Nano25Oct2019-v1/NANOAOD /MET/Run2016H-Nano25Oct2019-v1/NANOAOD
2017	/MET/Run2017B-Nano14Dec2018-v1/NANOAOD /MET/Run2017C-Nano14Dec2018-v1/NANOAOD /MET/Run2017D-Nano14Dec2018-v1/NANOAOD /MET/Run2017E-Nano14Dec2018-v1/NANOAOD /MET/Run2017F-Nano14Dec2018-v1/NANOAOD
2018	/MET/Run2018B-Nano14Dec2018-v1/NANOAOD /MET/Run2018C-Nano14Dec2018-v1/NANOAOD /MET/Run2018D-Nano14Dec2018_ver2-v1/NANOAOD

Table 17: The MET datasets used in the cross trigger method.

These were chosen since they have constant MET trigger thresholds and also have a similar event topology to signal events tZq in the dilepton channel. For MC, the $t\bar{t}$ samples used are shown in Table 18.

Year	$t\bar{t}$ Sample Names
2016	ttbar (inclusive): TT_TuneCUETP8M2T4_13TeV-powheg-pythia8
2017	$t\bar{t} > 2l2\nu$: TTTo2L2Nu_TuneCP5_PSweights_13TeV-powheg-pythia8
	$t\bar{t}$ (madgraph): TTJets_TuneCP5_13TeV-amcatnloFXFX-pythia8
	$t\bar{t}$ (hadronic): TTToHadronic_TuneCP5_PSweights_13TeV-powheg-pythia8
	$t\bar{t}$ (semileptonic): TTToSemiLeptonic_TuneCP5_PSweights_13TeV-powheg-pythia8
	$t\bar{t}$ (MCatNLO): TTJets_TuneCP5_13TeV-amcatnloFXFX-pythia8
2018	$t\bar{t} > 2l2\nu$: TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8
	$t\bar{t}$ (madgraph): TTJets_TuneCP5_13TeV-amcatnloFXFX-pythia8
	$t\bar{t}$ (hadronic): TTToHadronic_TuneCP5_13TeV-powheg-pythia8
	$t\bar{t}$ (semileptonic): TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8
	$t\bar{t}$ (MCatNLO): TTJets_TuneCP5_13TeV-amcatnloFXFX-pythia8

Table 18: The $t\bar{t}$ samples used in the cross trigger method.

The number of events that pass the MET triggers, the dilepton triggers and the event selection is represented by $N_{trigME_T+trgLL}$. The number of events that pass only the MET triggers and the event selection is represented by N_{trigME_T}. The dilepton trigger efficiency is defined as the ratio of $N_{trigME_T+trgLL}$ and N_{trigME_T}. This is shown in Equation 101.

$$\epsilon_{trigger} = \frac{N_{trigME_T+trgLL}}{N_{trigME_T}} \tag{101}$$

Pile up reweighting is applied to the MC samples on a per event basis, which will be explained in Section 9.3.6. Equation 101 is applied for both data and MC. The ratio between the MC trigger efficiency ϵ_{MC} and the data efficiency ϵ_{data} is taken, which is the definition for the trigger scale factor (SF). The definition for the trigger SF is shown in Equation 102.

$$SF_{trigger} = \frac{\epsilon_{MC}}{\epsilon_{data}} \tag{102}$$

The Particle Data Group recommends the use of frequentist Clopper-Pearson intervals to calculate the errors associated with efficiency and scale factor values [87]. A Clopper-Pearson interval level of 0.60 is used. An error function, called a trigger turn on, is used to better describe trigger behaviour. The turn on curves that result from this are shown in Section 10.1.

The alpha, efficiency and scale factor values calculated using 2016, 2017 and 2018 data and MC are shown in Section 10.1.

9.3.2 Electron Reconstruction and Selection Efficiency Scale Factors

The energy regression corrects the electron energy recorded in order to improve the resolution of the reconstructed particle by using detector information. Energy scale and smearing corrections were applied to correct the energy scale and resolution in simulation. Electron regression, electron energy scale and electron smearing corrections are already applied in nanoAOD samples, but this is not the case for the electron reconstruction and selection efficiency scale factors. The latter two are centrally provided by the CMS Electron Gamma (CMS EGM) group and are dependent on the p_T and $|\eta|$ of electron candidates.

9.3.3 Rochester Corrections

Rochester Corrections are applied to muon momentum measurements to correct for the effects of bias. This can be a result of detector misalignment, bias in software reconstruction or uncertainties that arise from the magnetic field [88]. The mean value of the reciprocal of the muon momentum, $\langle 1/p_T^{\mu} \rangle$, is calculated in bins of charge, η and ϕ for muons produced from the decay $Z \rightarrow \mu\mu$. The Rochester Correction is given by $\langle p_T^{\mu} \rangle$. The average invariant mass of the reconstructed Z boson in the decay, $\langle M_{\mu\mu}^{Z} \rangle$, is calculated to tune the Rochester Corrections. The Rochester Corrections are provided by the CMS collaboration [89].

9.3.4 Jet Energy Corrections

Jet energy corrections (JECs) comprise jet energy scale and jet smearing corrections. The jet energy scale corrections have already been applied in the centrally produced nanoAOD samples. Descriptions of these jet energy scale corrections are as follows:

- Pile-up corrections: Applied to both data and simulation, these corrections remove excess energy that arises from pile up and from the jet energy.

- Relative and absolute corrections: Applied to both data and simulation, these corrections account for the non-uniform response of the detector for p_T and η.
- Residual corrections: Applied only to simulated jets, these account for any remaining discrepancies between data and simulation.

Jet smearing corrections are applied to the jets' four momenta. They comprise of a scale factor for the jet p_T resolution (σ^2_{JER}) and a data-to-simulation core resolution factor (S_{JER}); both of which are centrally-provided [90].

The hybrid method is used to obtain the jet smearing correction factor, C_{JER}. This method is a mixture of two methods: the scaling method and the stochastic method. If Equation 104 holds true, in which the ΔR between the jet transverse momentum (p_T) and the transverse momentum of a matching particle-level jet (p_T^{ptcl}) is less than the jet cone size parameter (0.4 for anti-k_T jets, which was previously defined in Section 5.3), and if Equation 105 also holds true, the scaling method is used. In the scaling method, C_{JER} is calculated by using Equation 103. However, if neither of these conditions do hold, the stochastic method is used. For the stochastic method, Equation 106 is used. In this, $N(0, \sigma_{JER})$ represents a random number sampled from a normal distribution with a mean of zero and a standard deviation of σ_{JER}. The square root of $\sigma^2_{JER} - 1$ is taken if the value of $\sigma^2_{JER} - 1$ is greater than 0, otherwise C_JER is taken to be equal to 1.

$$C_{JER} = 1 + (S_{JER} + 1)\frac{p_T - p_T^{ptcl}}{p_T} \tag{103}$$

$$\Delta R < R_{cone}/2 \tag{104}$$

$$\left|p_T - p_T^{ptcl}\right| > 3 \times \sigma_{JER} \times p_T \tag{105}$$

$$C_{JER} = 1 + N(0, \sigma_{JER})\sqrt{max(\sigma^2_{JER} - 1, 0)} \tag{106}$$

9.3.5 b-Tagging Scale Factors

A b-Tagging scale factor, ω_b, is calculated for each event to address any misidentfication made by the CSVv2 b-tagging algorithm. To calculate ω_b, firstly the ratio between the number of true bjets in MC events that have been correctly identified by CSVv2 and the number of bjets in MC is determined. This ratio is represented by ϵ_i. The same ratio is calculated for non-bjets, which is represented by ϵ_j. A centrally-provided scale factor is determined for bjets (SF_i) and non-bjets (SF_j) that is dependent on the CSVv2 operating point, the measurement type, the systematic type, the jet flavour, the η range in which the jet η lies, the p_T range in which the jet p_T lies and the CSVv2 discriminant value. The values of ϵ_i, ϵ_j, SF_i and SF_j are used to calculate the probabilities of a given configuration of jets in MC and data, which are represented by $P(MC)$ and $P(DATA)$, respectively. The equations used to calculate $P(MC)$ and $P(DATA)$ are shown in Equations 107 and 108, respectively. The ratio between $P(DATA)$ and $P(MC)$ is calculated to obtain the b-tagging scale factor, as shown in Equation 109.

$$P(MC) = \prod_{i=tagged} \epsilon_i \prod_{j=not\ tagged} (1 - \epsilon_j) \tag{107}$$

$$P(DATA) = \prod_{i=tagged} SF_i\epsilon_i \prod_{j=not\ tagged} (1 - SF_j\epsilon_j) \tag{108}$$

$$\omega_b = \frac{P(DATA)}{P(MC)} \tag{109}$$

9.3.6 Pile Up Modelling

Monte Carlo samples are weighted to correctly model the pile up (PU) shown in data. Pile up is defined as the number of interactions that occur away from the primary vertex. Pile up can vary year on year due to a change in certain conditions, like instantaneous luminosity. The PU reweighting was achieved by applying a scale factor determined from a distribution for the number of primary vertices in minimum bias data.

9.3.7 Missing Transverse Energy

The missing transverse energy of particle flow objects is corrected to take the uncertainties associated with the jet smearing corrections into account. The p_x and p_y values of the smeared jet four momentum is subtracted from the four momentum of the MET.

9.3.8 Top p_T Re-Weighting

A correction factor is applied to correct a previously-observed discrepancy in the top transverse momentum spectra [91] between simulation and data. Events in simulation are re-weighted by the weight ω_T, which is shown in Equation 110. This weight is dependent on two scale factors: one calculated using the transverse momenta of generator-level top quarks and a second calculated using the transverse momenta of generator-level antitop quarks. These scale factors are shown in Equation 111.

$$\omega_T = \sqrt{SF(t)SF(\bar{t})} \tag{110}$$

$$SF(p_T) = e^{-0.0615-0.0005p_T} \tag{111}$$

9.4 Shape Uncertainties

Shape uncertainties are a type of systematic uncertainty that affect the shapes of the kinematic distributions, which can affect the overall normalisation. The shape uncertainties that were varied in this analysis are explained in this section.

9.4.1 Jet Energy Corrections

The up and down variations in the jet energy scale and resolution scale factors were provided by CMS. For these, the central values were varied by $\pm 1\sigma$.

9.4.2 Pile-up Reweighting

The up and down variations for the pile-up re-weighting were implemented by varying the cross section $\pm 4.6\%$.

9.4.3 B-Tagging Scale Factors

The up and down variations for the b-tagging scale factors were provided by CMS, for which the central value was varied by $\pm 1\sigma$.

9.4.4 Missing Transverse Energy

For the up variation, the Δx of the MET unclustered energy is added to the p_x and p_y of the MET four momentum. For the down variation, the the Δx of the MET unclustered energy is subtracted from the p_x and p_y of the MET four momentum.

9.4.5 Parton Density Functions

Weights for the nominal and systematic variations for PDF sets are stored in NanoAOD samples. For a PDF set that corresponds to a Hessian set, Equation 112 was used to calculate the PDF uncertainty, δ^{PDF}. In this, N_{mem} is the number of error sets of the specific combined set. For example, for the PDF sets $PDF4LHC15\_30$ and $PDF4LHC15\_100$, N_{mem} has values of 30 and 100, respectively [94].

$$\delta^{PDF}\sigma = \sqrt{\sum_{k=1}^{N_{mem}} (\sigma^{(k)} - \sigma^{(0)})^2} \tag{112}$$

However, for a PDF set that corresponds to an MC set, such as $PDF4LHC15\_mc$, the standard deviation of the distribution given by Equation 113 was used. In this, the mean value of the cross section $\langle\sigma\rangle$ is given by Equation 114.

$$\delta^{PDF}\sigma = \sqrt{\frac{1}{N_{mem}-1} \sum_{k=1}^{N_{mem}} (\sigma^{(k)} - \sigma^{\langle\sigma\rangle})^2} \tag{113}$$

$$\langle\sigma\rangle = \frac{1}{N_{mem}} \sum_{k=1}^{N_{mem}} \sigma^{(k)} \tag{114}$$

9.4.6 Perturbative and Non-Perturbative Factorisation and Renormalisation Scales

The uncertainties associated with varying the factorisation μ_F and normalisation scales μ_R in perturbative and non-perturbative QCD were addressed at the matrix element and parton shower levels. The factorisation scale refers to the energy at which the PDF is used, while the renormalisation scale is associated with the strong coupling constant, α_s, of the matrix element.

Matrix Element Level

Dedicated MC samples were used to consider the uncertainties associated with μ_F and μ_R at the Matrix Element (ME) level. For this, "scale up" and "scale down" samples shown in Table 4 were used. However, these simulation samples make use of the MADGRAPH5 aMC@NLO 2.2.2 MC event generator. This event generator considers the tree level and one loop Feynman diagrams for any given process. For events where the leading order and higher order terms destructively interfere, the event generator assigns a negative event weight. This can impact the total numbers of positively-weighted and negatively-weighted events. To correct for this, a scale factor was applied to the overall weight for each event. Additionally, a generator weight was applied to correct the cross section normalisation.

Centrally-provided samples were also used to determine the impact of independently varying μ_F and μ_R on the initial state radiation (ISR) and the final state radiation (FSR) of $t\bar{t}$ events. These are also shown in Table 6, where the names of the samples include "isr up", "isr down", "fsr up" and "fsr down".

Parton Shower Level

For 2017 and 2018, Parton Shower (PS) weights were used to vary μ_F and μ_R at the PS level. The variations correspond to six cases: 1. When only μ_F has been doubled; 2. When only μ_R has been doubled; 3. When only μ_F has been halved; 4. When only μ_R has been halved; 5. When both μ_F and μ_R have been doubled and 6. When both μ_F and μ_R have been halved.

Single top tW samples

Since the PS weights were not available in the single top tW nominal sample for 2016, centrally-produced samples were used to consider the uncertainty associated with independently varying μ_F and μ_R at both the ME and PS levels. The sample names contain the phrases "scale up" and "scale down" and are shown in Table 5.

9.4.7 Matching Threshold Energy

The damping function shown in Equation 115 is used in POWHEGv2 samples to scale the cross section in ME-PS matching. The function is dependent on the transverse momentum of the leading additional jet for $t\bar{t}$ dilepton final state, p_T, and the hdamp variable. The hdamp variable is equal to $1.58 \times m_{top}$, where m_{top} is the mass of the top quark. Centrally-produced samples have been provided that consider the impact of varying hdamp up and down by one standard deviation. These samples are shown in Table 5.

$$\frac{hdamp^2}{hdamp^2 + p_T^2} \tag{115}$$

Samples for the up and down variations for hdamp have been provided by CMS for the single top t-channel, single top tW, single antitop t-channel and single antitop tW channel MC background samples for 2016, 2017 and 2018.

9.5 Rate Uncertainties

Rate uncertainties are a type of systematic uncertainty that affect the overall normalisation of distributions, but do not affect their shapes. The rate uncertainties that were varied for the analysis are explained in this section.

9.5.1 Integrated luminosity

The uncertainty in the total integrated luminosity collected using the CMS detector was estimated to be 2.5% in 2016 [92], 2.3% in 2017 [93] and 2.5% in 2018 [95].

9.5.2 Cross Section Normalisation

The value of the cross section used to calculate the normalisation weight was varied by ±30%. This was to address the uncertainty in the cross sections of simulated samples and a value of 30% was chosen based in accordance with a previous measurement for tZq in trilepton final states [96].

9.5.3 Lepton Efficiencies

The lepton efficiency scale factors provided by CMS were varied by $\pm 1\sigma$, using centrally-provided scale factors provided by the CMS collaboration.

9.6 Multivariate Analysis

The boosted decision tree (BDT) [97] multivariate analysis technique was used to separate the signal and dominant backgrounds for the tZq in dilepton final states. For tZq production, backgrounds include top pair production ($t\bar{t}$, $t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}H$), boson+jets ($Z/\gamma{*}{+}jets$, $W{+}jets$), dibosonic processes (WW, WZ, ZZ, tWZ) and tribosonic processes (WWW, WWZ, WZZ, ZZZ) [45]. However, the dominant backgrounds are $t\bar{t}$ and $Z + jets$ [46]. Boosted Decision Trees were used since they are well understood within the group studying Single Top production and were found to be the optimal machine learning technique for the analysis [45]. Boosted Decision Trees comprise a series of nodes, at each of which a decision is made as to how signal-like or background-like an event is. Decisions at each node are dependent on of previous nodes, until a final decision is made at the last node. Boosting can be carried out using algorithms, such as Adaptive Boosting (AdaBoost) [99] and Gradient Boosting, to apply a weight to each classifier to improve efficiency [100].

9.7 Extraction of the Signal Strength and Significance

The signal strength and significance were extracted by performing a binned maximum likelihood fit on the classifier response; the fit was performed simultaneously for the ee and $\mu\mu$ channels for 2016, 2017 and 2018. This was achieved using the Higgs Analysis Combined Limit Tool (combine) [101], which makes use of the Likelihood Model. Binned distributions of the BDT discriminator is used and, in this model, the expected yield in a given bin (λ_i) is shown in Equation 116. In this, the expected number of signal events is represented by s_i, the expected number of background events in bin i for background j is b_{ij} and the number of background processes is $nbkg$

$$\lambda_i = rs_i + \sum_{j}^{nbkg} b_{ij} \tag{116}$$

The signal strength, r, is defined as the ratio of the observed cross section σ_{obs} and the expected cross section σ_s; this is shown in Equation 117.

$$r = \frac{\sigma_{obs}}{\sigma_s} \tag{117}$$

A set of nuisance parameters, represented by θ, can be used to introduce the statistical and systematic uncertainties. These parameters adjust the signal and background contributions of each bin.

In Poisson statistics, for a given bin, the probability of observing n_i events, P_i, is shown in Equation 118.

$$P(n_i) = \frac{\lambda_i^{n_i}}{n_i!} e^{-\lambda_i} = \frac{(rs_i(\theta) + b_i(\theta))^{n_i}}{n_i!} e^{-rs_i(\theta) - b_i(\theta)} \tag{118}$$

The probability of observing a given dataset is therefore obtained by the product across all n bins, which is shown by Equation 119.

$$P(n|r,\theta) = \prod_{i=1}^{N} P(n_i) \tag{119}$$

A probability density function can be used to describe the sources of systematic uncertainty, since these are treated as being 100% correlated. A log-normal distribution is used to model the rate-based uncertainties whilst the nominal shapes are varied by $\pm 1\sigma$ in order to model the shape-based systematic uncertainties. The Barlow-Beeston-lite approach [102] was used to model the statistical uncertainties. In this approach, to reduce fitting time and improve the stability of the fit, one nuisance parameter per bin is created instead of one bin per sample.

The probability density of nuisance parameters, $\rho(\theta|\tilde{\theta})$, can be multiplied by the probability of observing a given dataset to obtain the likelihood for the entire dataset L. This is shown in Equation 120.

$$L = \prod_{i=1}^{N} P(n_i)\dot{\rho}(\theta|\tilde{\theta}) \tag{120}$$

To extract the significance, the log-likelihood ratio is used as the test statistic. This ratio can be represented by q_r, which is shown in Equation 121. The maximum likelihood estimators of θ for a given r is represented by $\hat{\theta}_r$. The global maxima of the likelihood are represented by $\hat{r}$ and $\hat{\theta}$ [103].

$$q_r = \begin{cases} -2ln(\frac{L(data(r,\hat{\theta_r}))}{L(data(\hat{r},\hat{r}))}) & \hat{\mu} \geq 0 \\ 0 & \hat{r} \leq 0 \end{cases}$$

10 Results: Search for tZq Production in Dilepton Final States

10.1 Trigger Efficiency Measurements

Tables 19-21 show the efficiency, trigger SF and α values obtained using the cross trigger method, which was previously described in Section 9.3. Figures 48- 51 show comparisons between NanoAOD and MiniAOD for the simulation efficiencies, data efficiencies, trigger scale factor and α values obtained for 2016 and 2017. The values correspond to the ee and $\mu\mu$ channels for the side band region, and the $e\mu$ channel for the $t\bar{t}$ control region. NanoAOD results for 2016 and 2017 were found to be in agreement with results previously obtained using MiniAOD samples, for the ee and $\mu\mu$ channels. While scale factor results were as expected for the ee and $\mu\mu$ channels, results for the $e\mu$ channel were observed to be lower than expected across all three years. Due to time constraints this was not resolved in time, but there was possibly a bug in the code for the lepton selection.

channel	ϵ_{data}	ϵ_{MC}	SF	α
ee	$0.972^{0.001}_{-0.003}$	0.97 ± 0.01	1.002 ± 0.003	0.999
$\mu\mu$	0.960 ± 0.001	0.976 ± 0.004	0.983 ± 0.001	0.999
$e\mu$	$0.972^{0.001}_{-0.003}$	0.510 ± 0.001	0.403 ± 0.001	0.999

Table 19: The alpha, efficiencies and scale factor values obtained using the cross trigger method for 2016 MC samples and datasets.

channel	ϵ_{data}	ϵ_{MC}	SF	α
ee	$0.972^{+0.001}_{-0.003}$	$0.974^{+0.002}_{-0.003}$	0.991 ± 0.005	0.999
$\mu\mu$	0.956 ± 0.001	$0.973^{+0.001}_{-0.002}$	0.982 ± 0.001	0.999
$e\mu$	$0.335^{+0.0001}_{-0.0008}$	0.822 ± 0.001	0.407 ± 0.001	0.999

Table 20: The alpha, efficiencies and scale factor values obtained using the cross trigger method for 2017 MC samples and datasets.

channel	ϵ_{data}	ϵ_{MC}	SF	α
ee	$0.969\pm^{+0.001}_{-0.007}$	0.974 ± 0.002	0.994 ± 0.01	0.999
$\mu\mu$	0.970 ± 0.004	$0.979\pm^{+0.002}_{-0.001}$	0.990 ± 0.004	0.999
$e\mu$	0.294 ± 0.001	$0.821^{+0.001}_{-0.0003}$	0.358 ± 0.002	0.999

Table 21: The alpha, efficiencies and scale factor values obtained using the cross trigger method for 2018 MC samples and datasets.

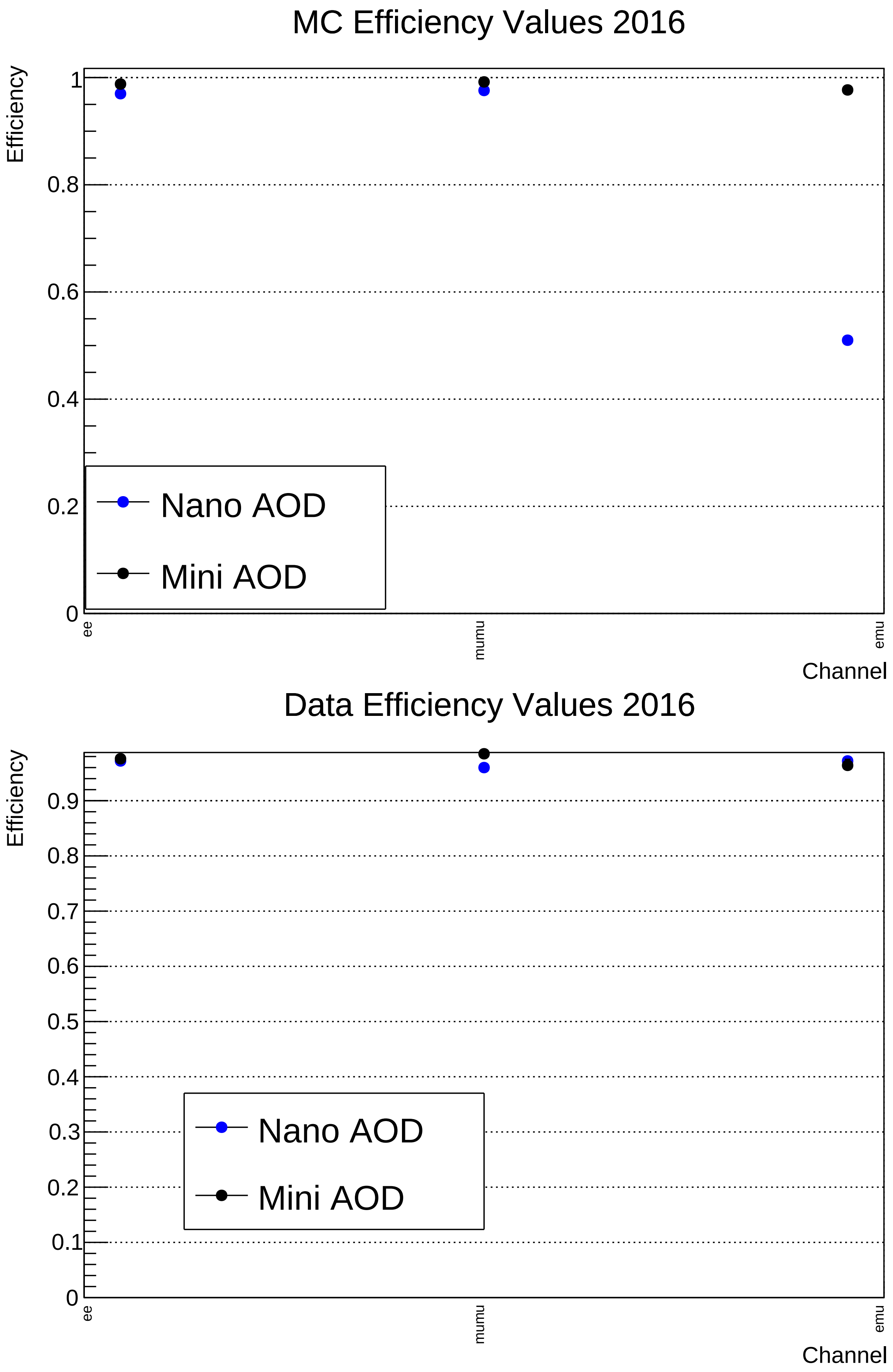


Figure 48: Comparisons between the trigger efficiency values obtained using MiniAOD and NanoAOD samples for $t\bar{t}$ MC (top) and MET data (bottom) for 2016.

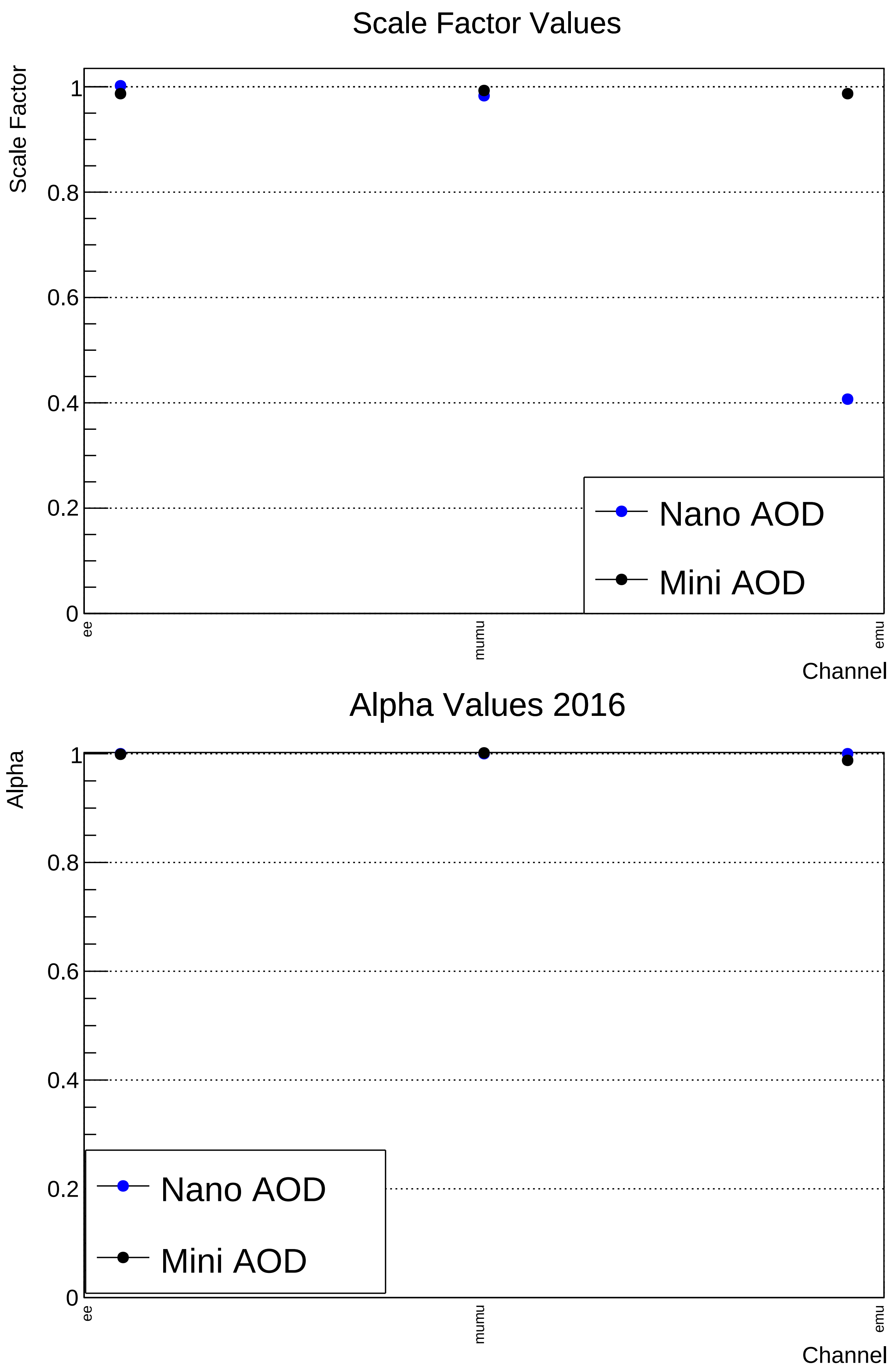


Figure 49: Comparisons the trigger scale factor (top) and alpha values (bottom) obtained using MiniAOD and NanoAOD samples for 2016.

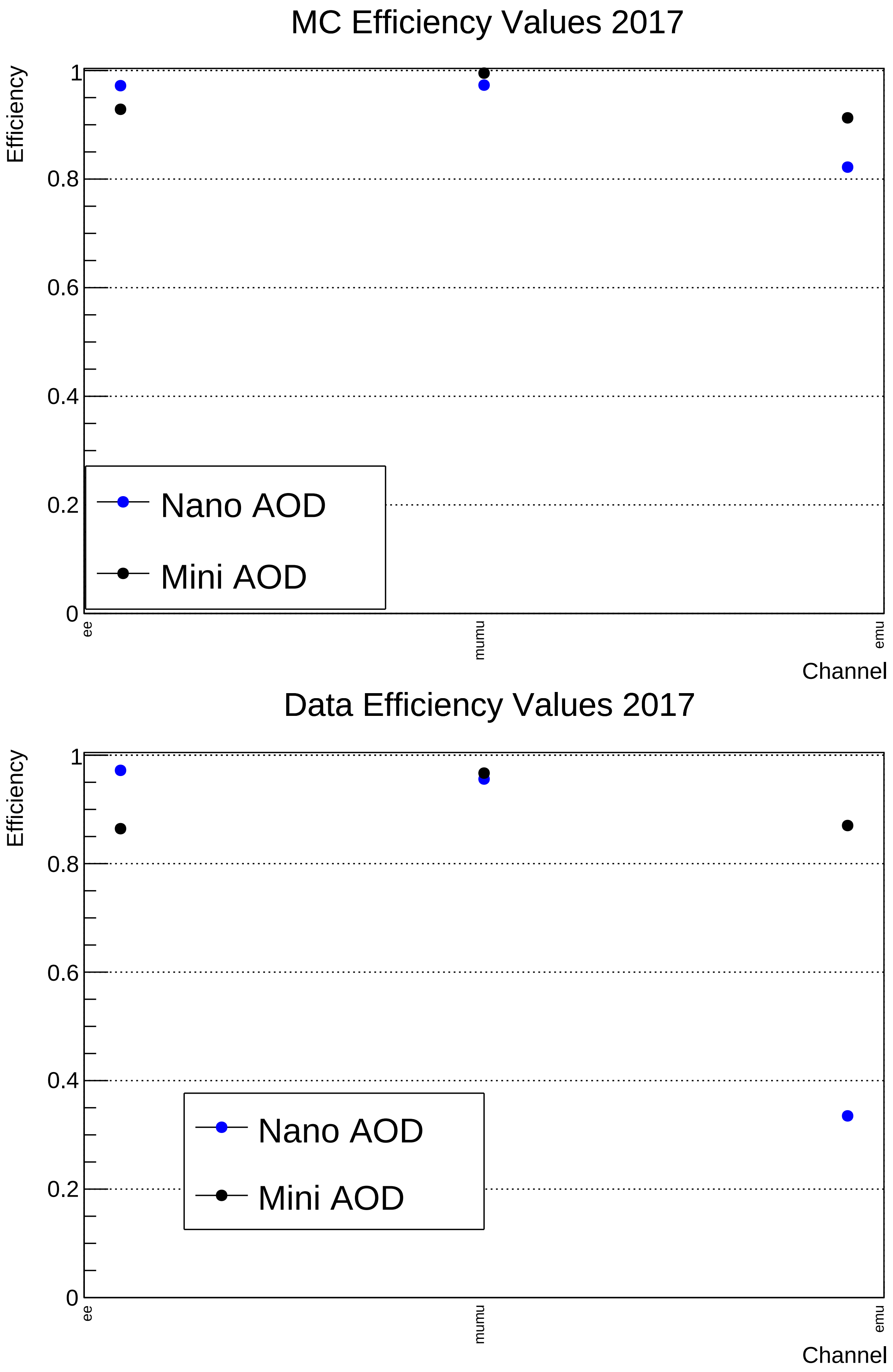


Figure 50: Comparisons between the trigger efficiency values obtained using MiniAOD and NanoAOD samples for $t\bar{t}$ MC (top) and MET data (bottom) for 2017.

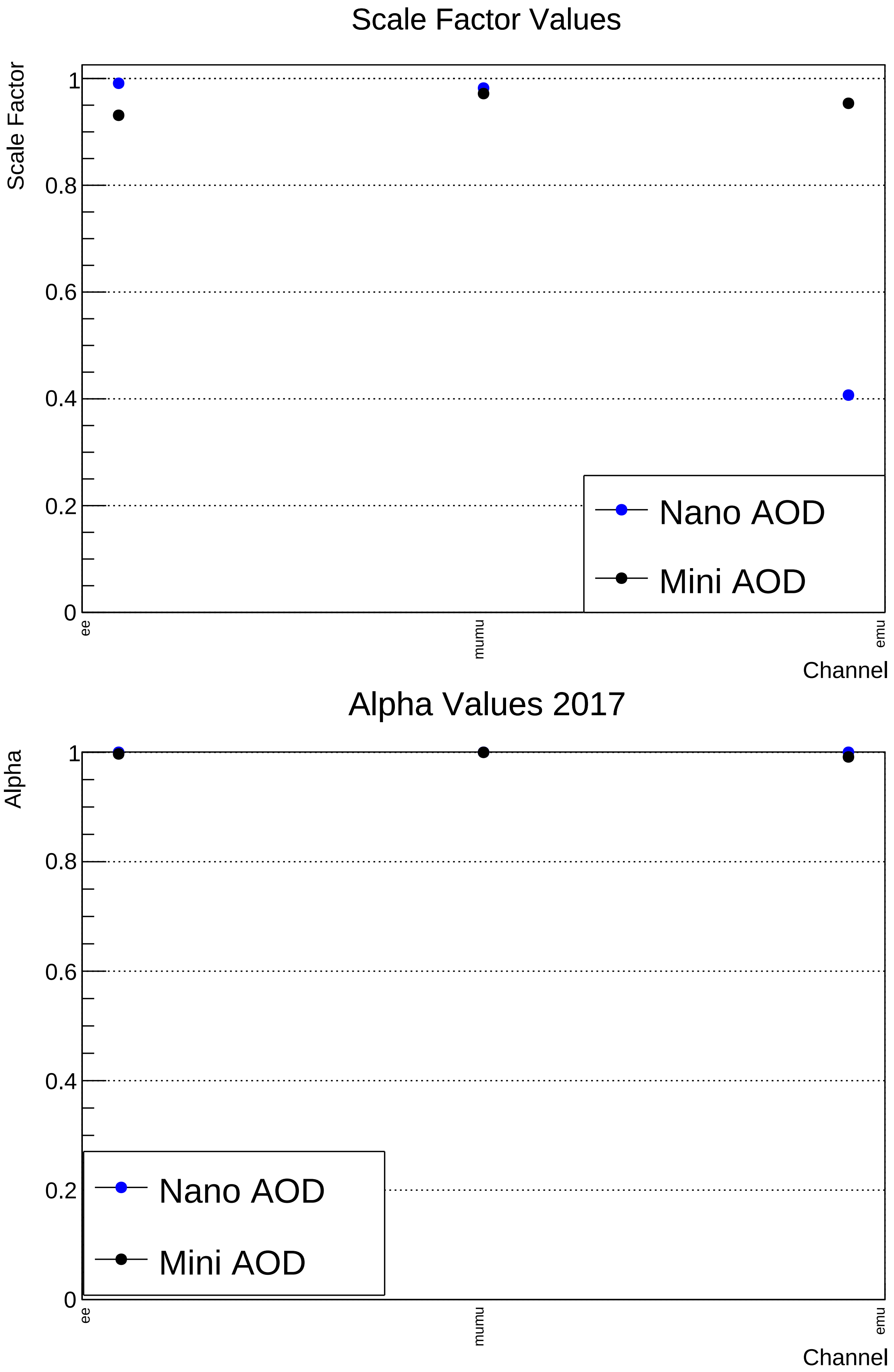


Figure 51: Comparisons the trigger scale factor (top) and alpha values (bottom) obtained using MiniAOD and NanoAOD samples for 2017.

10.2 Distributions of selected features

10.2.1 2016: ee channel, side-band region

Figures 52- 55 show the distributions of selected features that were used as inputs for the multivariate analysis. These correspond to 2016 data and simulation in the *ee* channel, and in the side-band region. The selected features are: b-tagged discriminator of the leading b-tagged jet, the transverse momentum of the leading jet candidate, ΔR between the leading and subleading jet candidates, ΔR between the second and third jet candidates, the transverse momentum of the fourth jet candidate, the sum of the masses of the four jet candidates, the missing transverse energy, the invariant top mass, the pseudorapidity of the subleading jet from the reconstructed W boson candidate, the mass of the reconstructed Z boson candidate, the transverse momentum of the reconstructed Z boson candidate, ΔR between the reconstructed Z boson and the leading b-jet candidates and the maximum ΔR between the Z boson candidate and any jet candidate in the system. These features were chosen as they were previously found to have the highest discriminating power[45] [46].

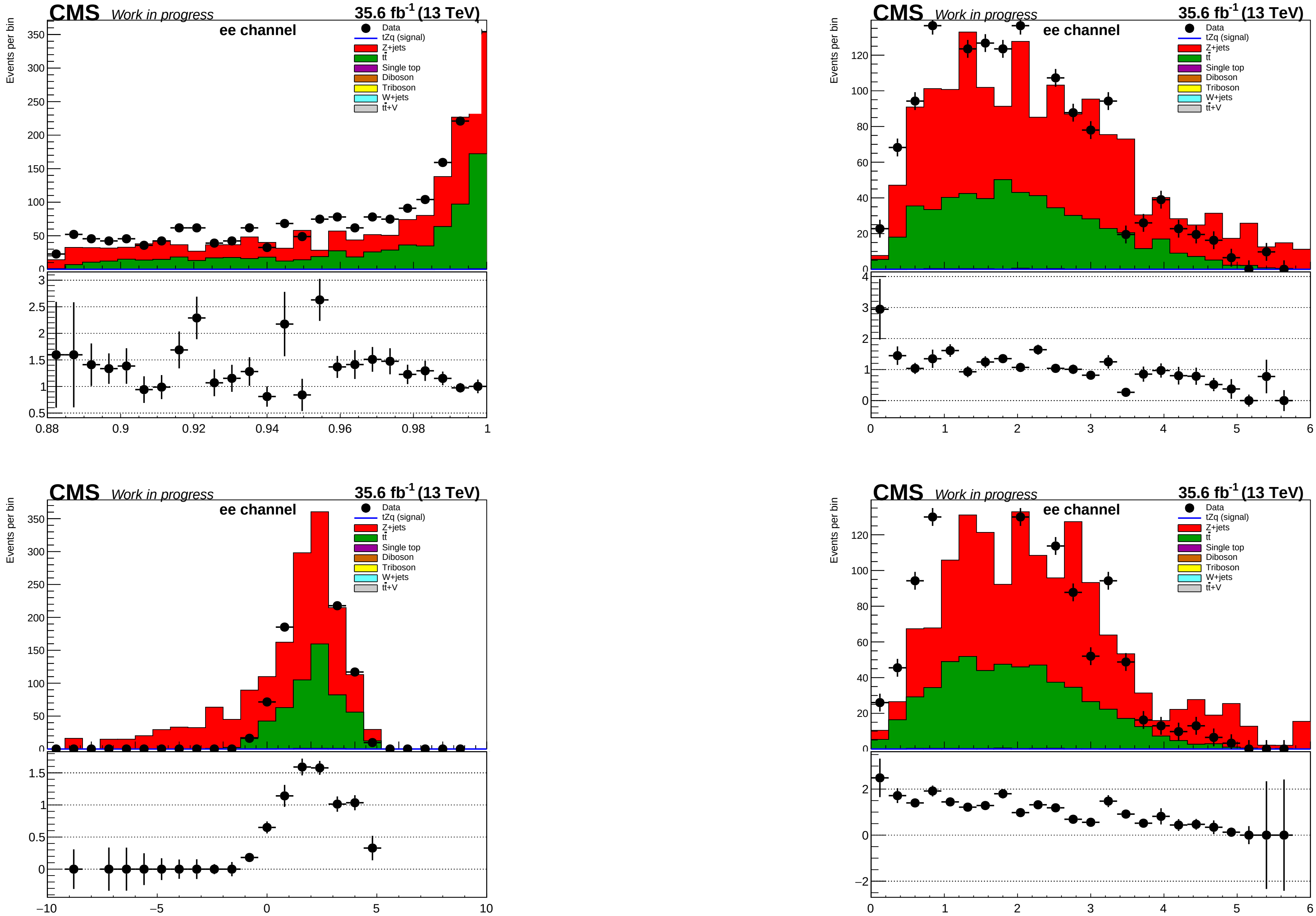


Figure 52: The distributions of the features: b-tagged discriminator of the leading b-tagged jet, the transverse momentum of the leading jet candidate, ΔR between the leading and subleading jet and ΔR between the second and third jet candidates. The features in this figure correspond to the ee channel in the side-band region, for simulation and data recorded in 2016.

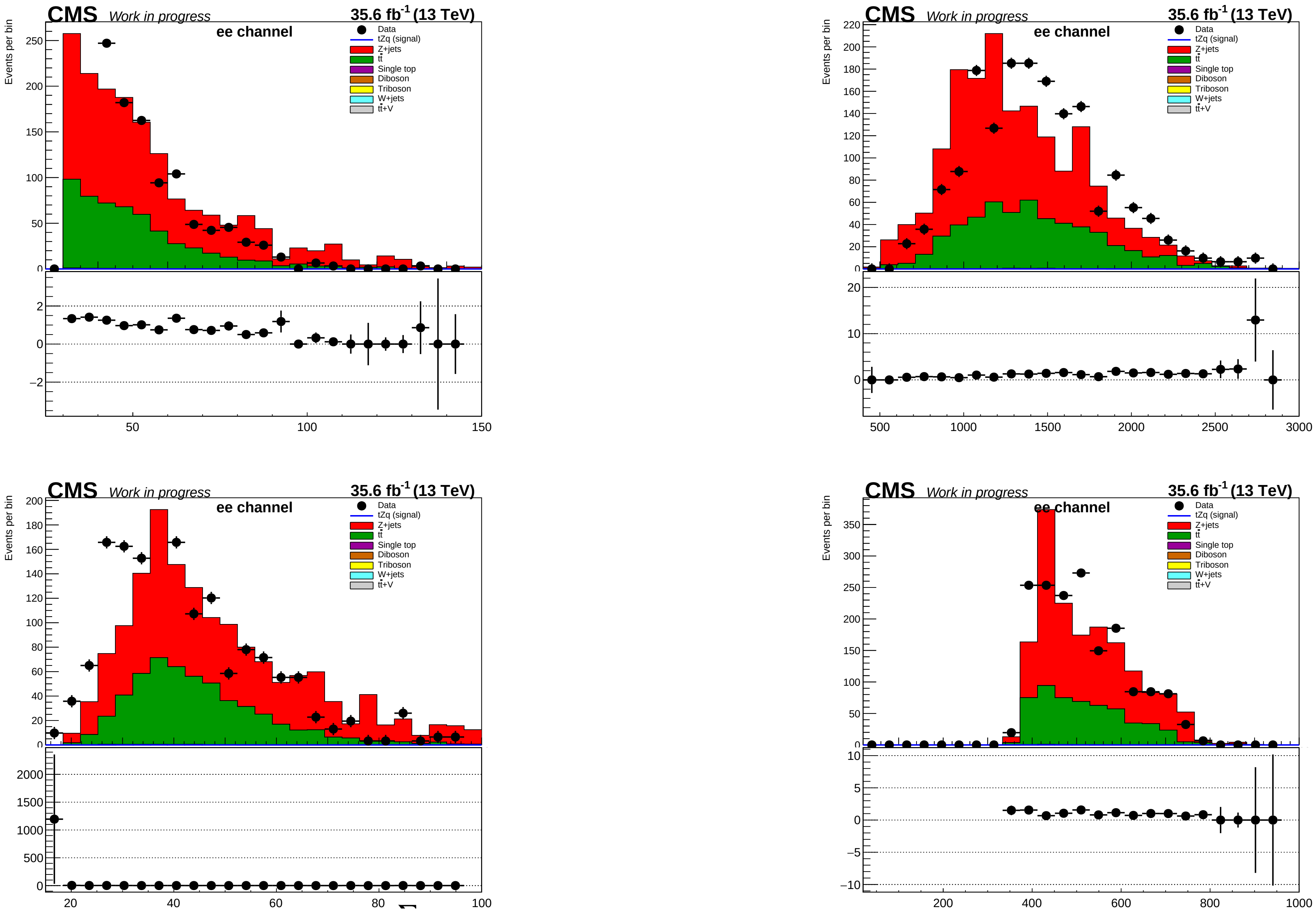


Figure 53: The distributions of the features: the transverse momentum of the fourth jet candidate, the sum of the masses of the four jet candidates, the missing transverse energy and the invariant mass of the top quark candidate. The features in this figure correspond to the *ee* channel in the side-band region, for simulation and data recorded in 2016.

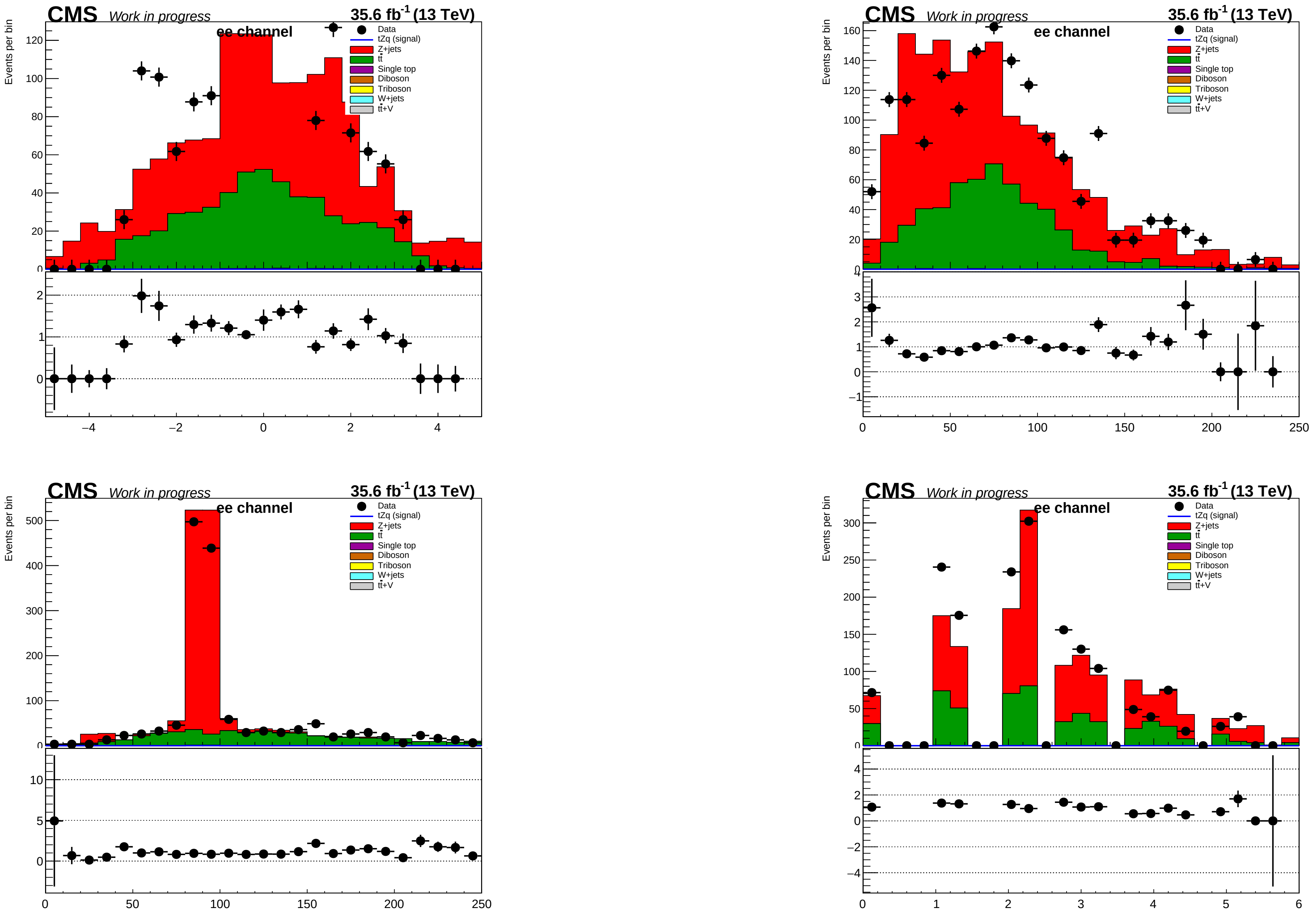


Figure 54: The distributions of the features: the pseudorapidity of the subleading jet from the reconstructed W boson candidate, the mass of the reconstructed Z boson candidate, the transverse momentum of the reconstructed Z boson candidate and ΔR between the reconstructed Z boson and the leading b-jet candidate. The features in this figure correspond to the *ee* channel in the side-band region, for simulation and data recorded in 2016.

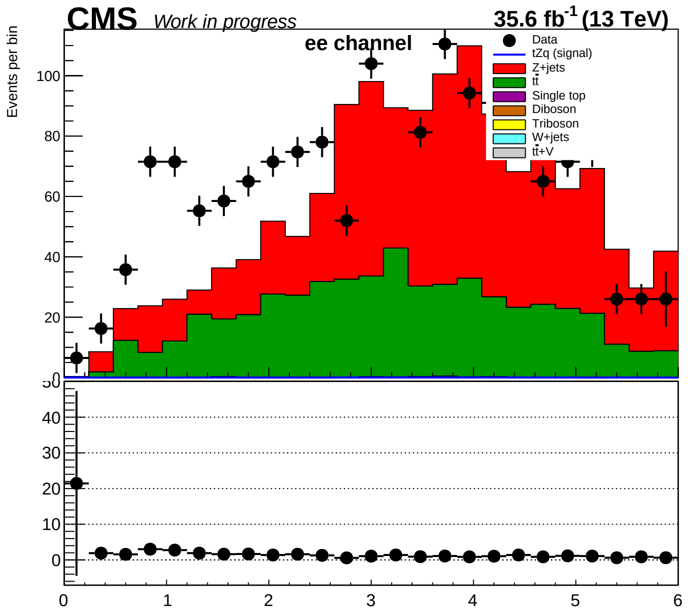


Figure 55: The distribution of the feature the maximum ΔR between the Z boson candidate and any jet candidate in the system. The feature in this figure corresponds to the *ee* channel in the side-band region, for simulation and data recorded in 2016.

10.2.2 2016: $\mu\mu$ channel, side-band region

Figures 56- 59 show the distributions of selected features that were used as inputs for multivariate analysis. These correspond to 2016 data and simulation in the $\mu\mu$ channel, and in the side-band region. The selected features were: b-tagged discriminator of the leading b-tagged jet, the pseudorapidity of the leading jet candidate, ΔR between the leading and subleading jet candidates, the transverse momentum of the subleading jet candidate, the sum of the masses of the four jet candidates, the transverse momentum of the leading lepton, the missing transverse energy, the invariant top mass, the pseudorapidity of the reconstructed W boson candidate, the pseudorapidity the reconstructed Z boson candidate, the invariant mass of the reconstructed Z boson candidate, ΔR between the reconstructed Z boson and the leading b-jet candidates and the maximum ΔR between the Z boson candidate and any jet candidate in the system. These features were chosen as they were previously found to have the highest discriminating power[45] [46].

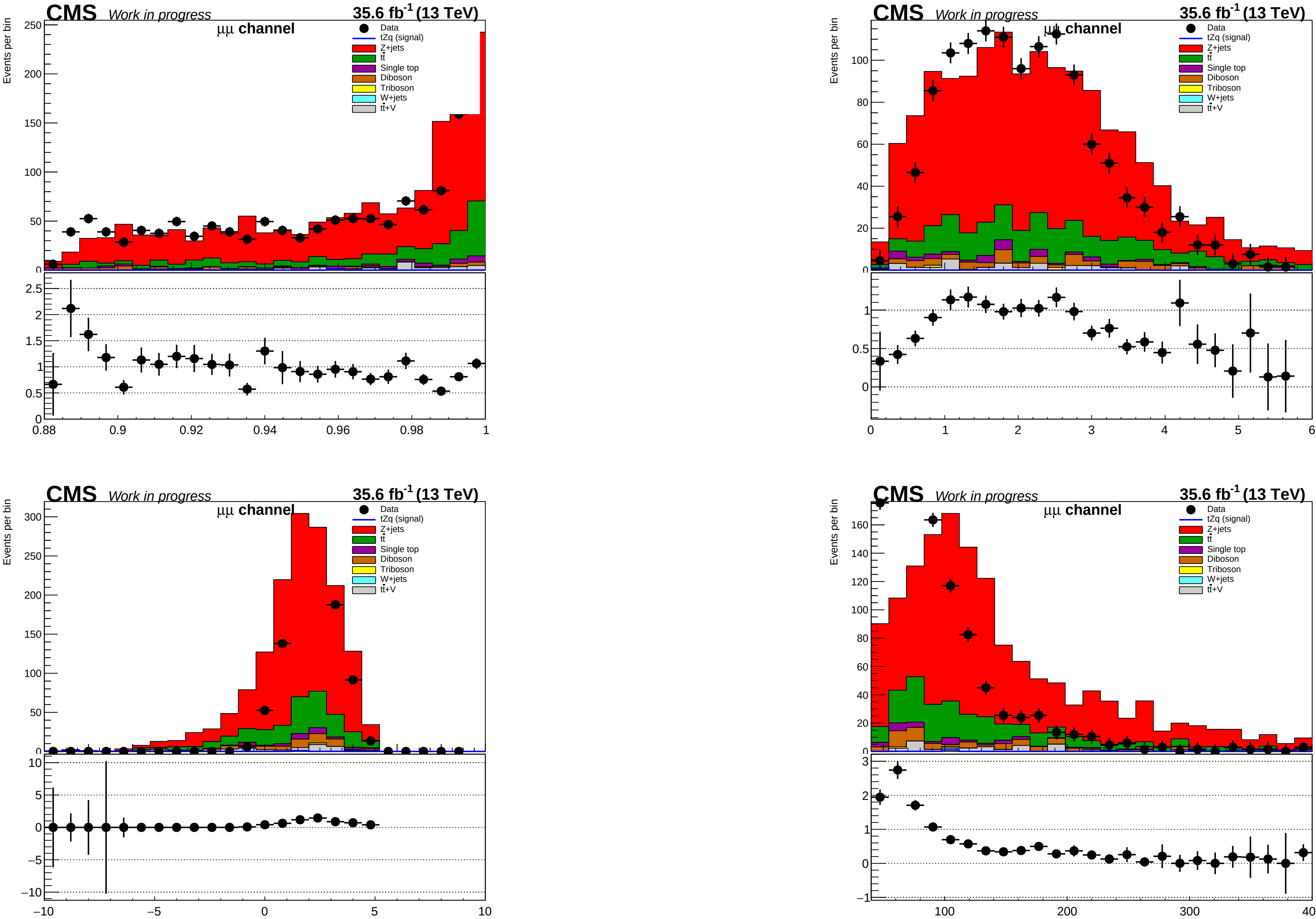


Figure 56: The distributions of the features: b-tagged discriminator of the leading b-tagged jet, the pseudorapidity of the leading jet candidate, ΔR between the leading and subleading jet candidates and the transverse momentum of the subleading jet candidate. The features in this figure correspond to the $\mu\mu$ channel in the side-band region, for simulation and data recorded in 2016.

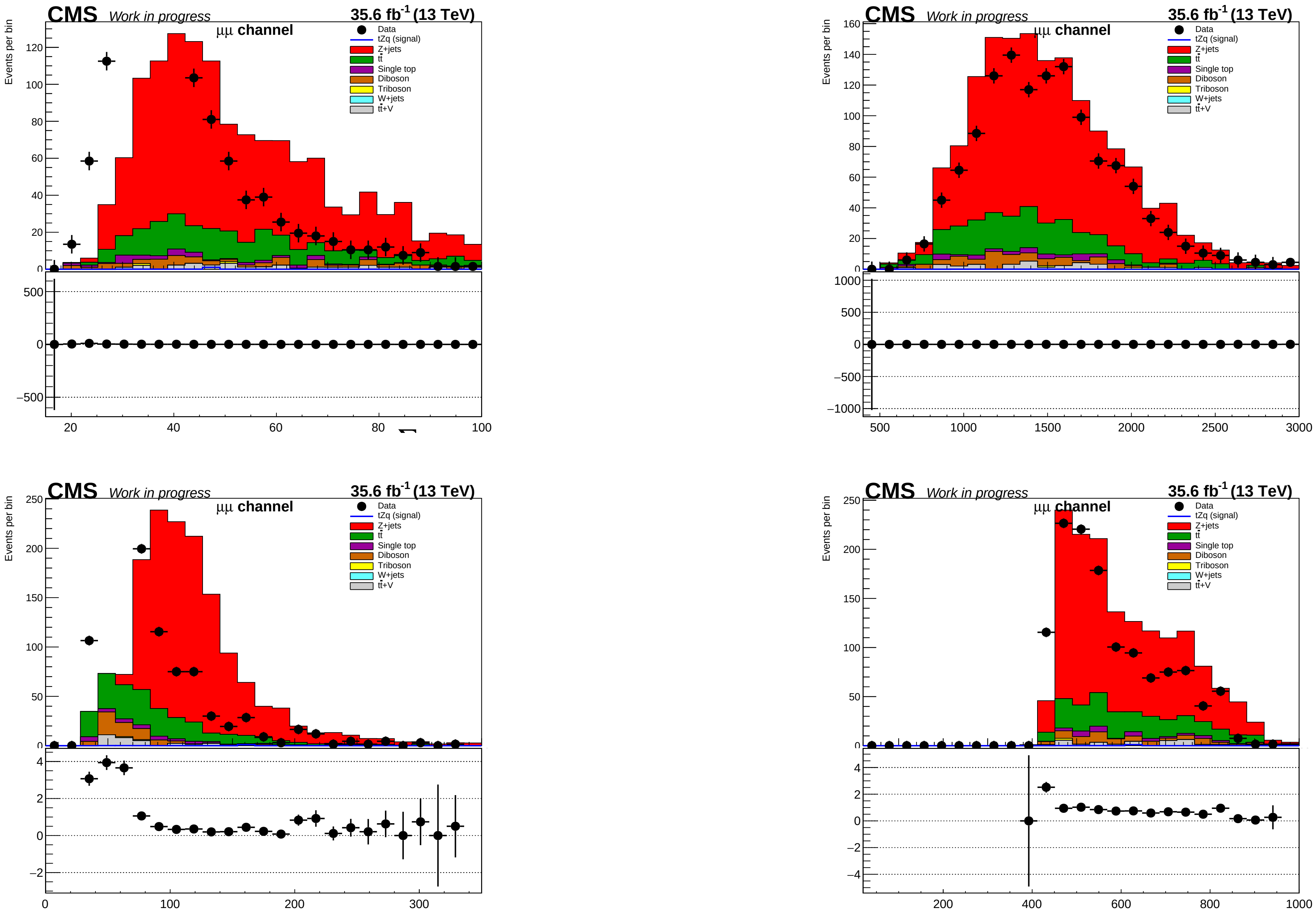


Figure 57: The distributions of the features: the sum of the masses of the four jet candidates, the transverse momentum of the leading lepton, the missing transverse energy and the invariant mass of the reconstructed top quark candidate. The features in this figure correspond to the $\mu\mu$ channel in the side-band region, for simulation and data recorded in 2016.

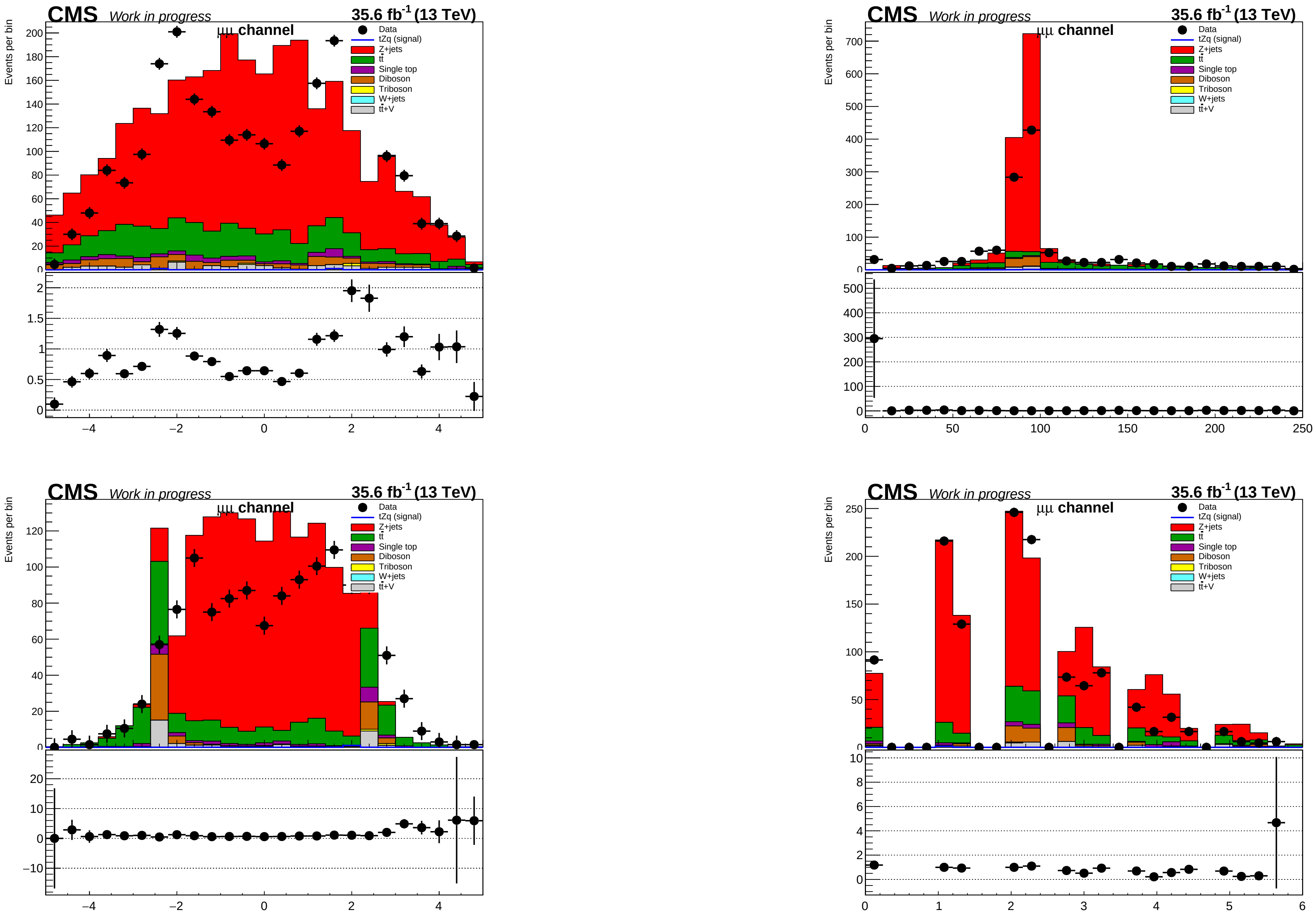


Figure 58: The distributions of the features: he pseudorapidity of the reconstructed W boson candidate, the pseudorapidity the reconstructed Z boson candidate, the invariant mass of the reconstructed Z boson candidate, ΔR between the reconstructed Z boson and the leading b-jet candidates. The features in this figure correspond to the $\mu\mu$ channel in the side-band region, for simulation and data recorded in 2016.

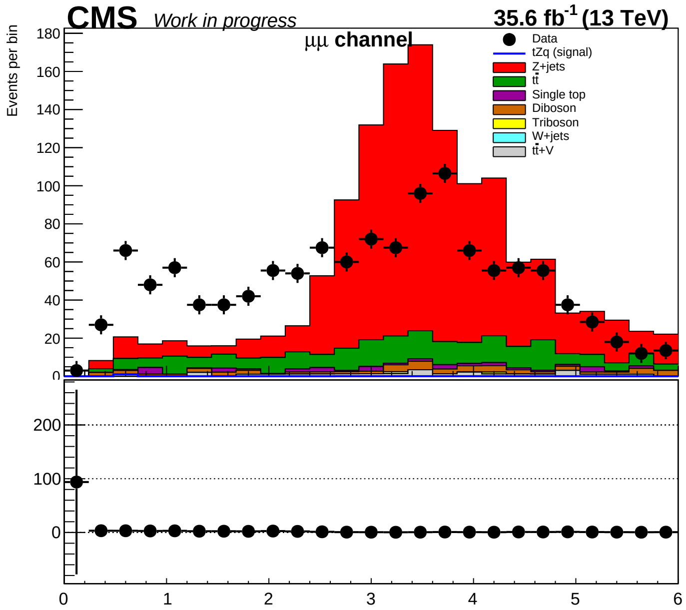


Figure 59: The distribution of the feature the maximum ΔR between the Z boson candidate and any jet candidate in the system. The feature in this figure corresponds to the $\mu\mu$ channel in the side-band region, for simulation and data recorded in 2016.

10.2.3 2017: *ee* channel, side-band region

Figures 60- 63 show the distributions of selected features that were used as inputs for multivariate analysis. These correspond to 2016 data and simulation in the *ee* channel, and in the side-band region. The selected features were: b-tagged discriminator of the leading b-tagged jet, the χ^2 variable used for experimental blinding, ΔR between the leading and subleading jet candidates, the transverse momentum of the subleading jet candidate, the transverse momentum of the fourth jet candidate, the sum of the masses of the four jet candidates, the transverse momentum of the leading lepton, the missing transverse energy, the transverse momentum of the leading jet from the reconstructed W boson candidate, the invariant mass of the reconstructed Z boson candidate and the maximum ΔR between the Z boson candidate and any jet candidate in the system. These features were chosen as they were previously found to have the highest discriminating power [45] [46].

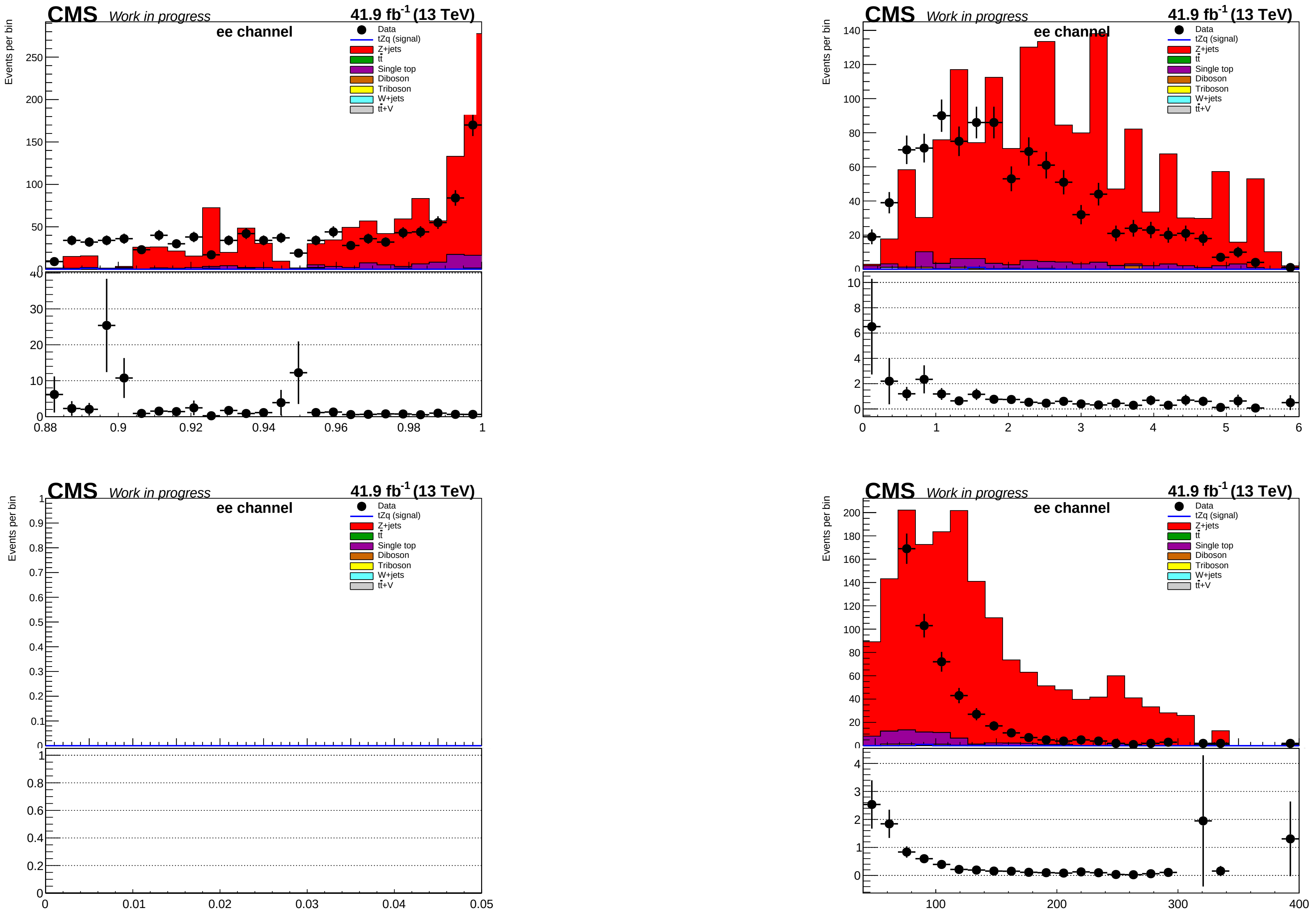


Figure 60: The distributions of the features: the b-tagged discriminator of the leading b-tagged jet, the χ^2 variable used for experimental blinding, ΔR between the leading and subleading jet candidates and the transverse momentum of the subleading jet candidate. The features in this figure correspond to the ee channel in the side-band region, for simulation and data recorded in 2017.

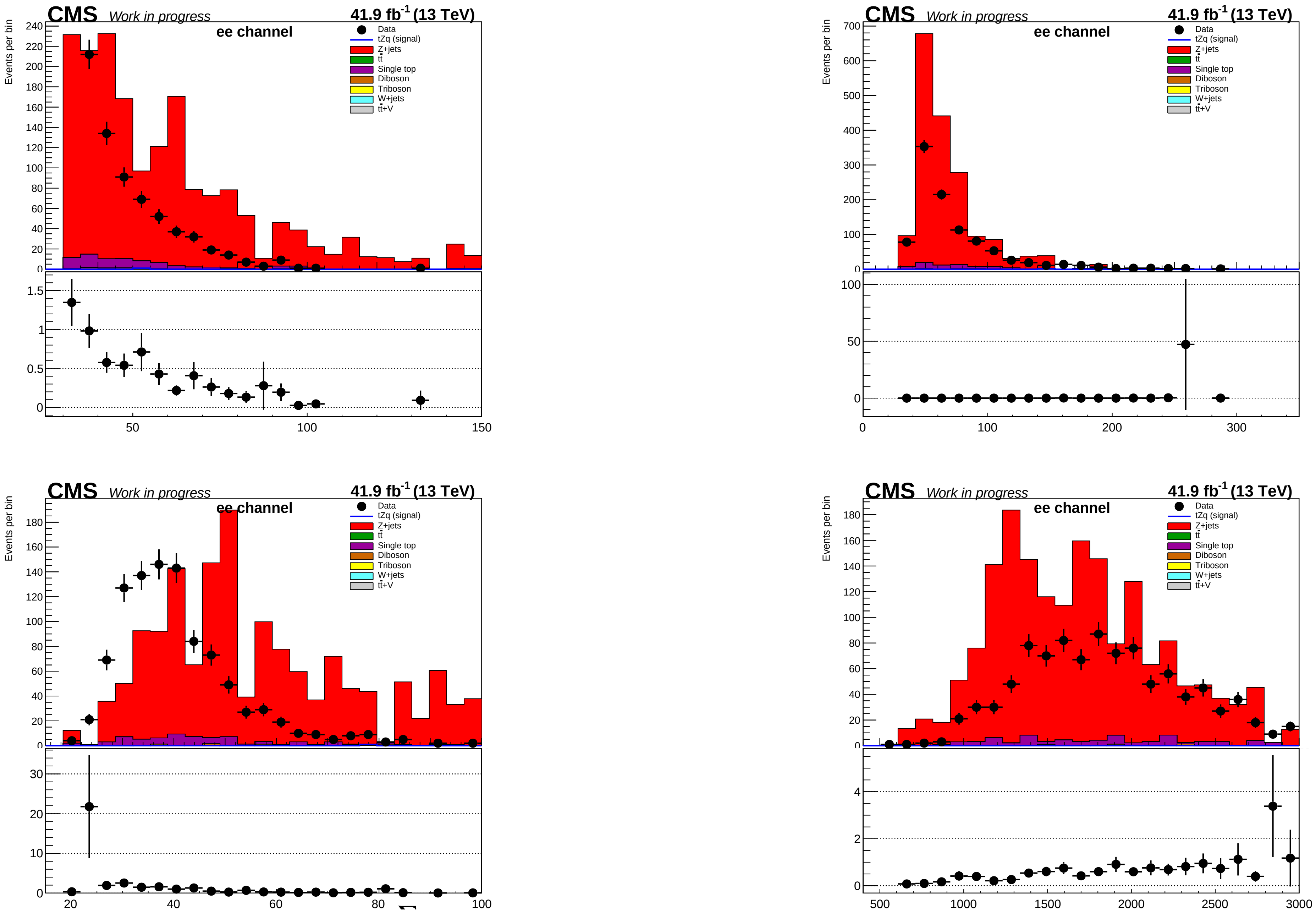


Figure 61: The distributions of the features: the transverse momentum of the fourth jet candidate, the sum of the masses of the four jet candidates, the transverse momentum of the leading lepton and the missing transverse energy. The features in this figure correspond to the *ee* channel in the side-band region, for simulation and data recorded in 2017.

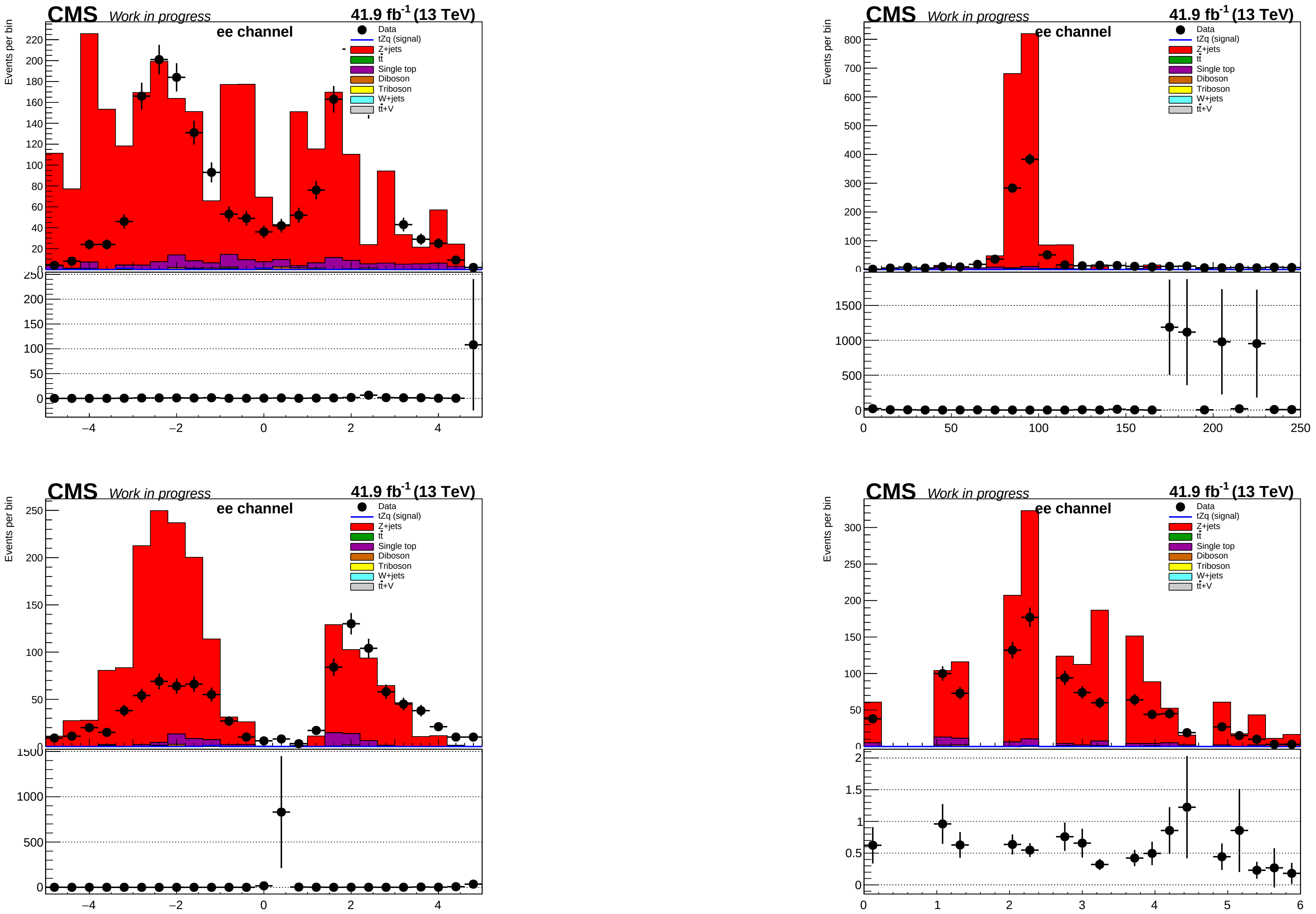


Figure 62: The distributions of the features: he pseudorapidity of the reconstructed W boson candidate, the pseudorapidity the reconstructed Z boson candidate, the invariant mass of the reconstructed Z boson candidate and the ΔR between the reconstructed Z boson and the leading b-jet candidates. The features in this figure correspond to the *ee* channel in the side-band region, for simulation and data recorded in 2017.

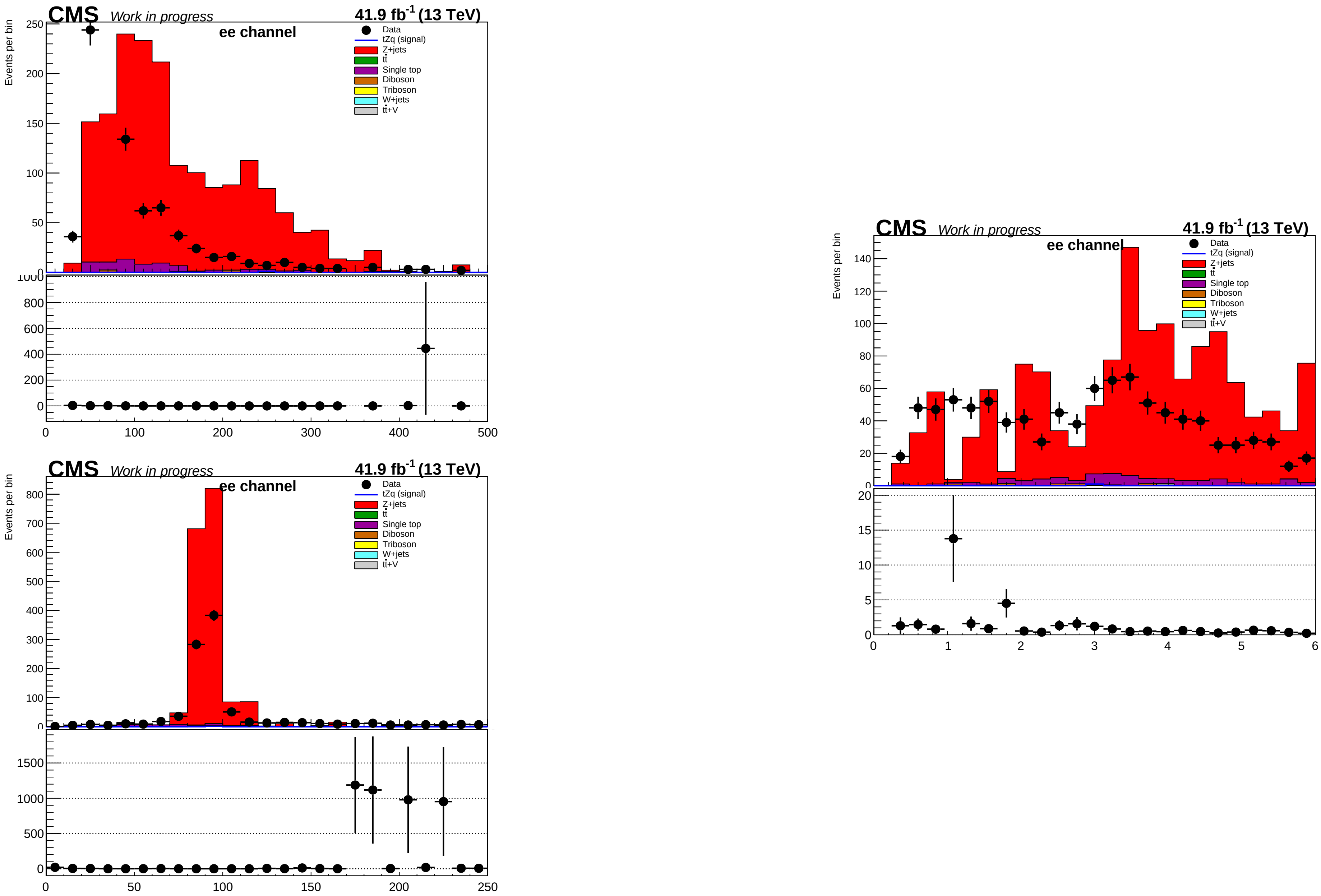


Figure 63: The distribution of the features: the transverse momentum of the leading jet from the reconstructed W boson candidate, the invariant mass of the reconstructed Z boson candidate and the maximum ΔR between the Z boson candidate and any jet candidate in the system. The features in this figure corresponds to the *ee* channel in the side-band region, for simulation and data recorded in 2017.

10.2.4 2017: $\mu\mu$ channel, side-band region

Figures 64- 66 show the distributions of selected features that were used as inputs for multivariate analysis. These correspond to 2016 data and simulation in the *ee* channel, and in the side-band region. The selected features were: b-tagged discriminator of the leading b-tagged jet, the χ^2 variable used for experimental blinding, the pseudorapidity of of the leading jet candidate, the transverse momentum of the subleading jet candidate, the transverse momentum of the fourth jet candidate, the sum of the masses of the four jet candidates, the missing transverse energy, the transverse momentum of the subleading jet from the reconstructed W boson candidate, the invariant mass of the reconstructed Z boson candidate, the transverse momentum of the reconstructed Z boson candidate, the ΔR between the reconstructed Z boson candidate and the leading bjet and the maximum ΔR between the Z boson candidate and any jet candidate in the system. These features were chosen as they were previously found to have the highest discriminating power [45] [46].

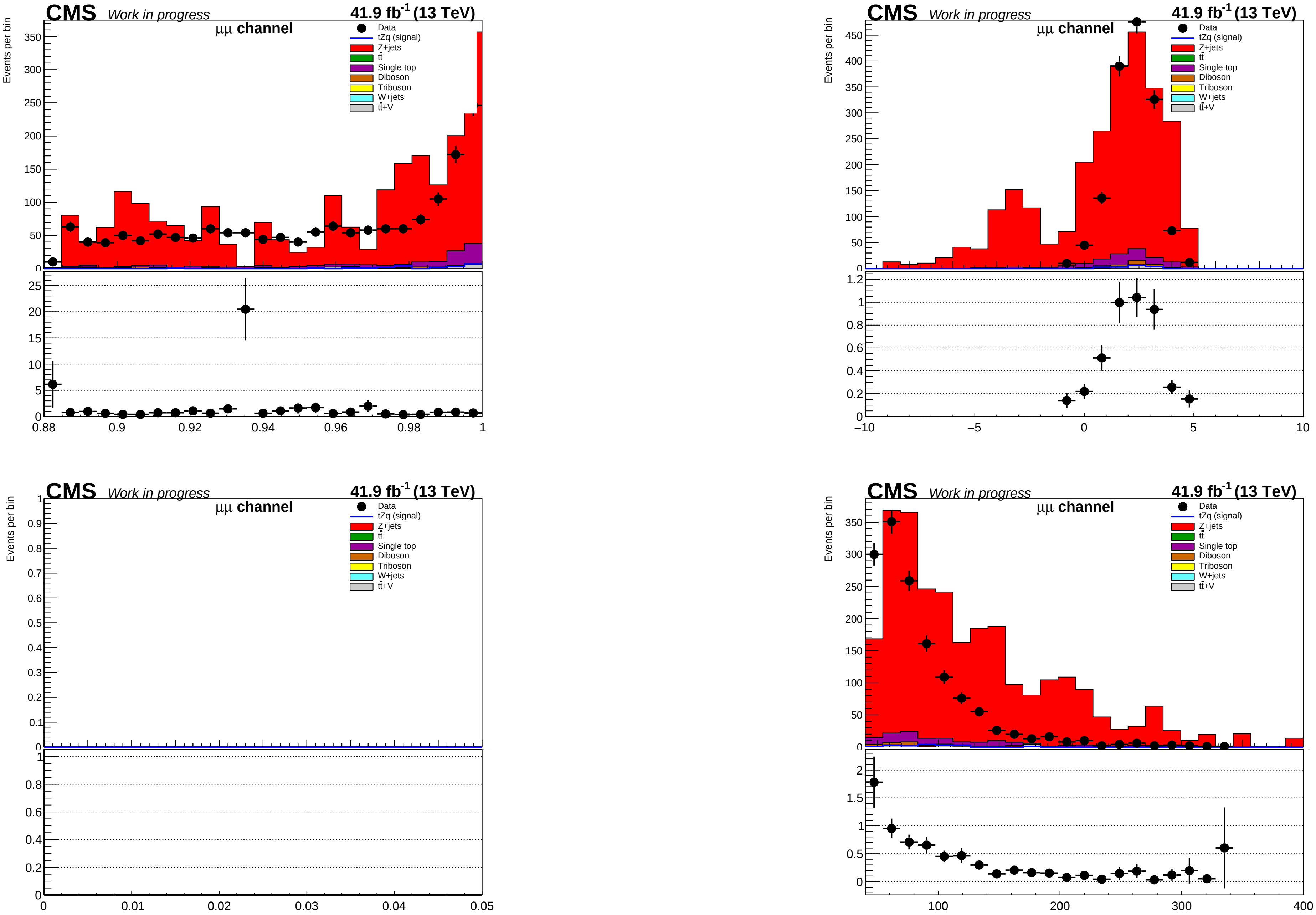


Figure 64: The distributions of the features: b-tagged discriminator of the leading b-tagged jet, the χ^2 variable used for experimental blinding, the pseudorapidity of of the leading jet candidate and the transverse momentum of the subleading jet candidate. The features in this figure correspond to the $\mu\mu$ channel in the side-band region, for simulation and data recorded in 2017.

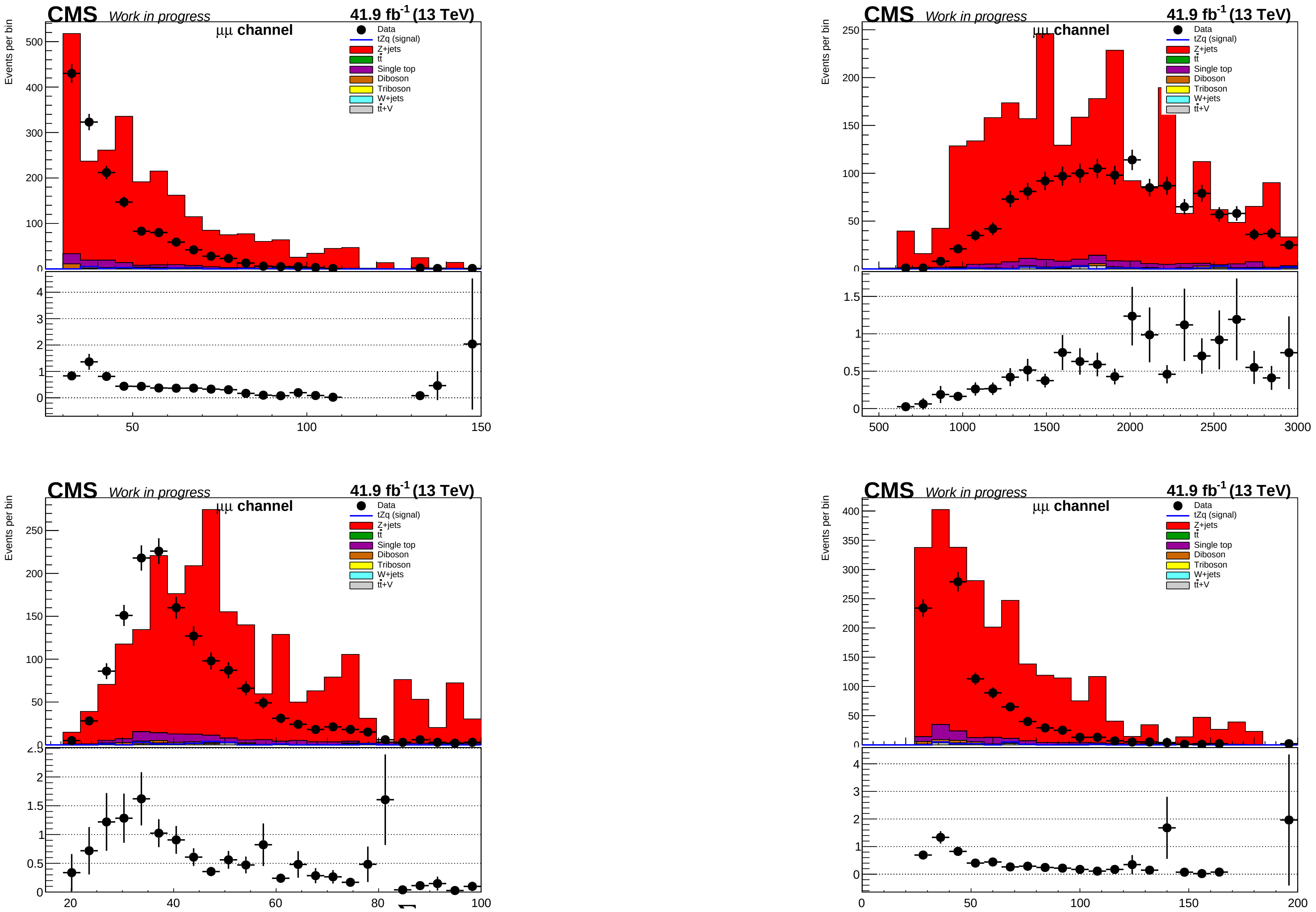


Figure 65: The distributions of the features: the transverse momentum of the fourth jet candidate, the sum of the masses of the four jet candidates, the missing transverse energy and the transverse momentum of the subleading jet from the reconstructed W boson candidate. The features in this figure correspond to the $\mu\mu$ channel in the side-band region, for simulation and data recorded in 2017.

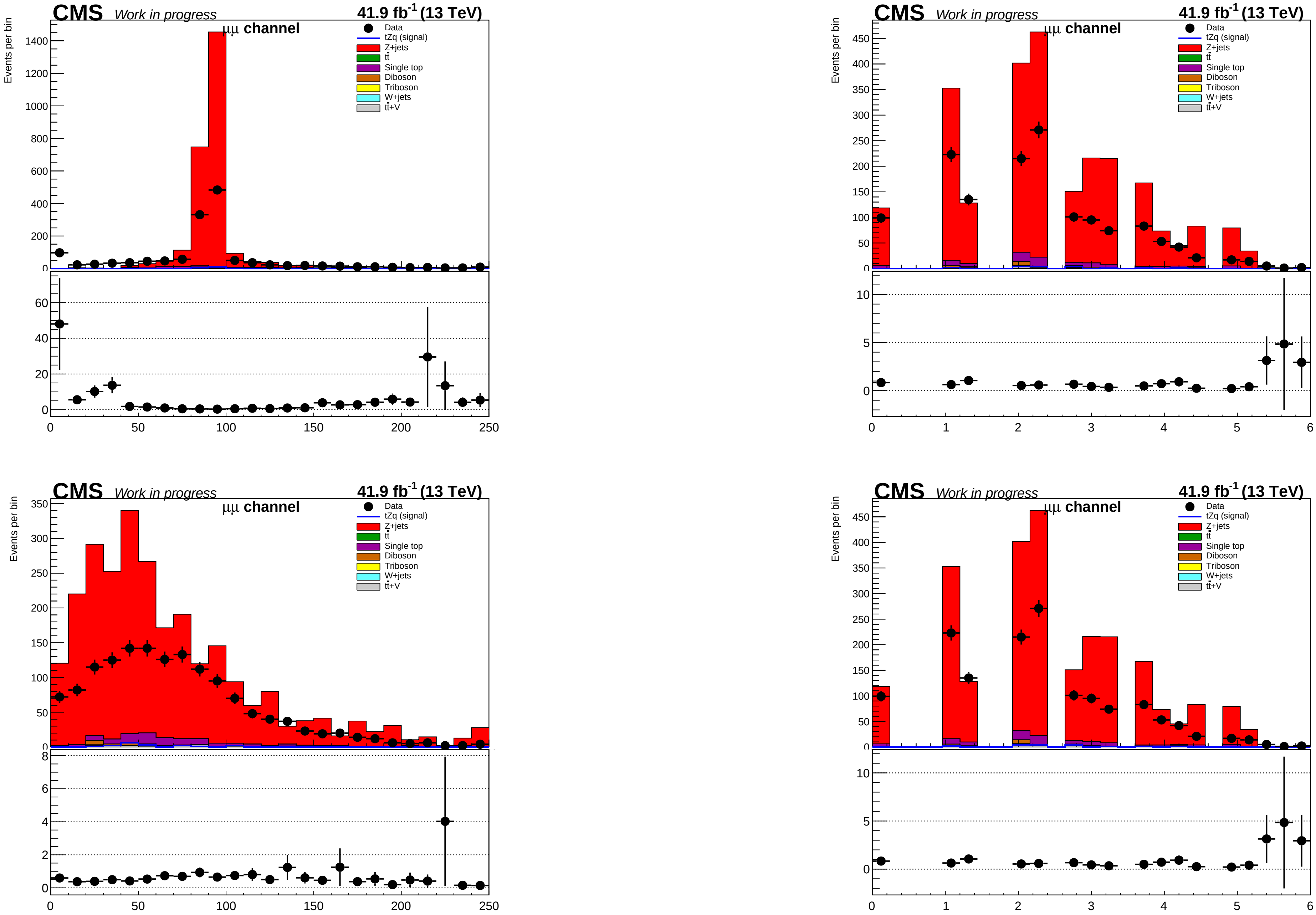


Figure 66: The distributions of the features: the invariant mass of the reconstructed Z boson candidate, the transverse momentum of the reconstructed Z boson candidate, the ΔR between the reconstructed Z boson candidate and the leading bjet and the maximum ΔR between the Z boson candidate and any jet candidate in the system. The features in this figure correspond to the $\mu\mu$ channel in the side-band region, for simulation and data recorded in 2017.

10.2.5 2018: *ee* channel, side-band region

Figures 67- 70 show the distributions of selected features that were used as inputs for multivariate analysis. These correspond to 2018 data and simulation in the *ee* channel, and in the side-band region. The selected features were: b-tagged discriminator of the leading b-tagged jet, the χ^2 variable used for experimental blinding, ΔR between the leading and subleading jet candidates, the transverse momentum of the subleading jet candidate, the transverse momentum of the fourth jet candidate, the sum of the masses of the four jet candidates, the transverse momentum of the leading lepton, the missing transverse energy, the transverse momentum of the leading jet from the reconstructed W boson candidate, the invariant mass of the reconstructed Z boson candidate and the maximum ΔR between the Z boson candidate and any jet candidate in the system.

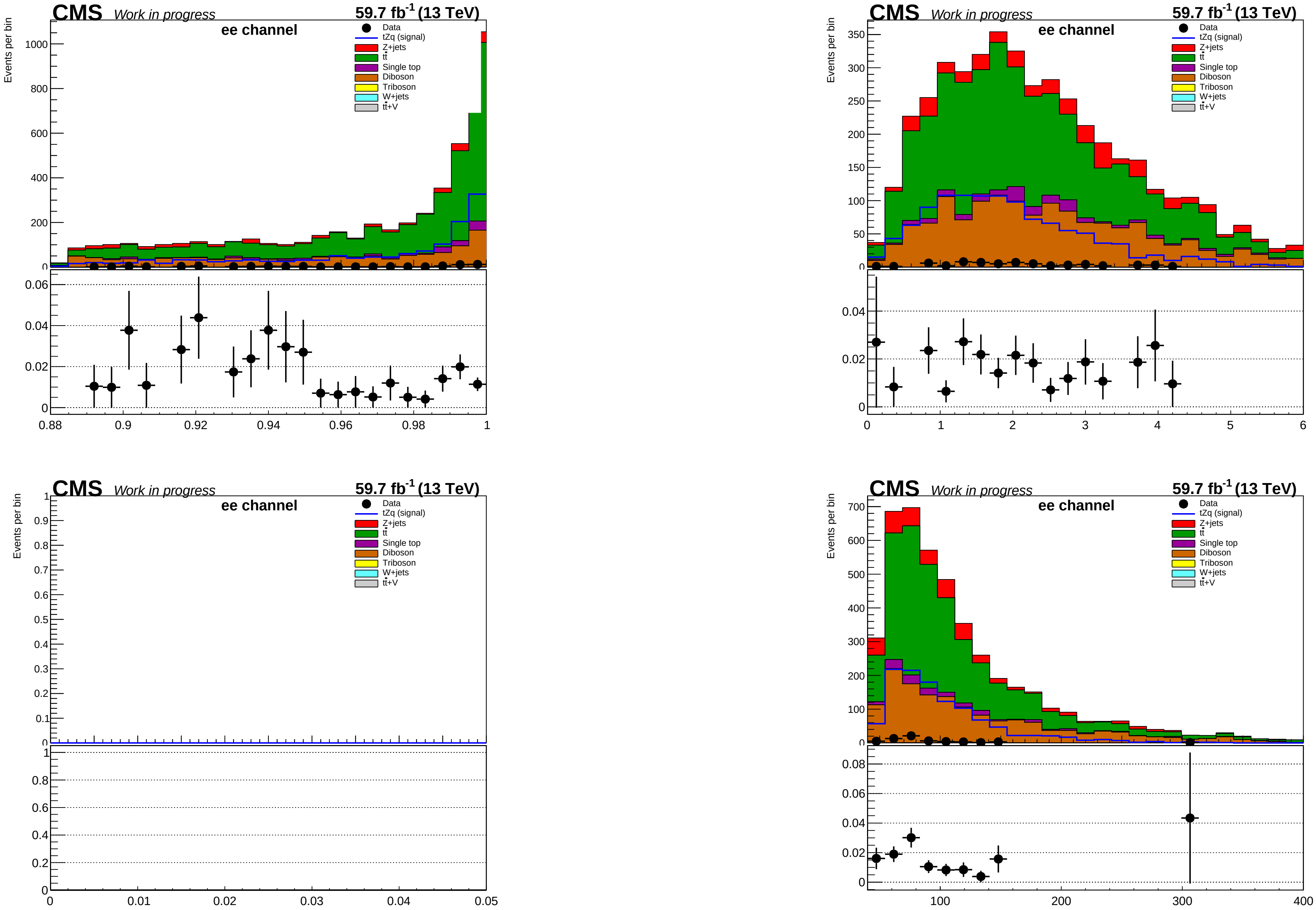


Figure 67: The distributions of the features: the b-tagged discriminator of the leading b-tagged jet, the χ^2 variable used for experimental blinding, ΔR between the leading and subleading jet candidates and the transverse momentum of the subleading jet candidate. The features in this figure correspond to the *ee* channel in the side-band region, for simulation and data recorded in 2018.

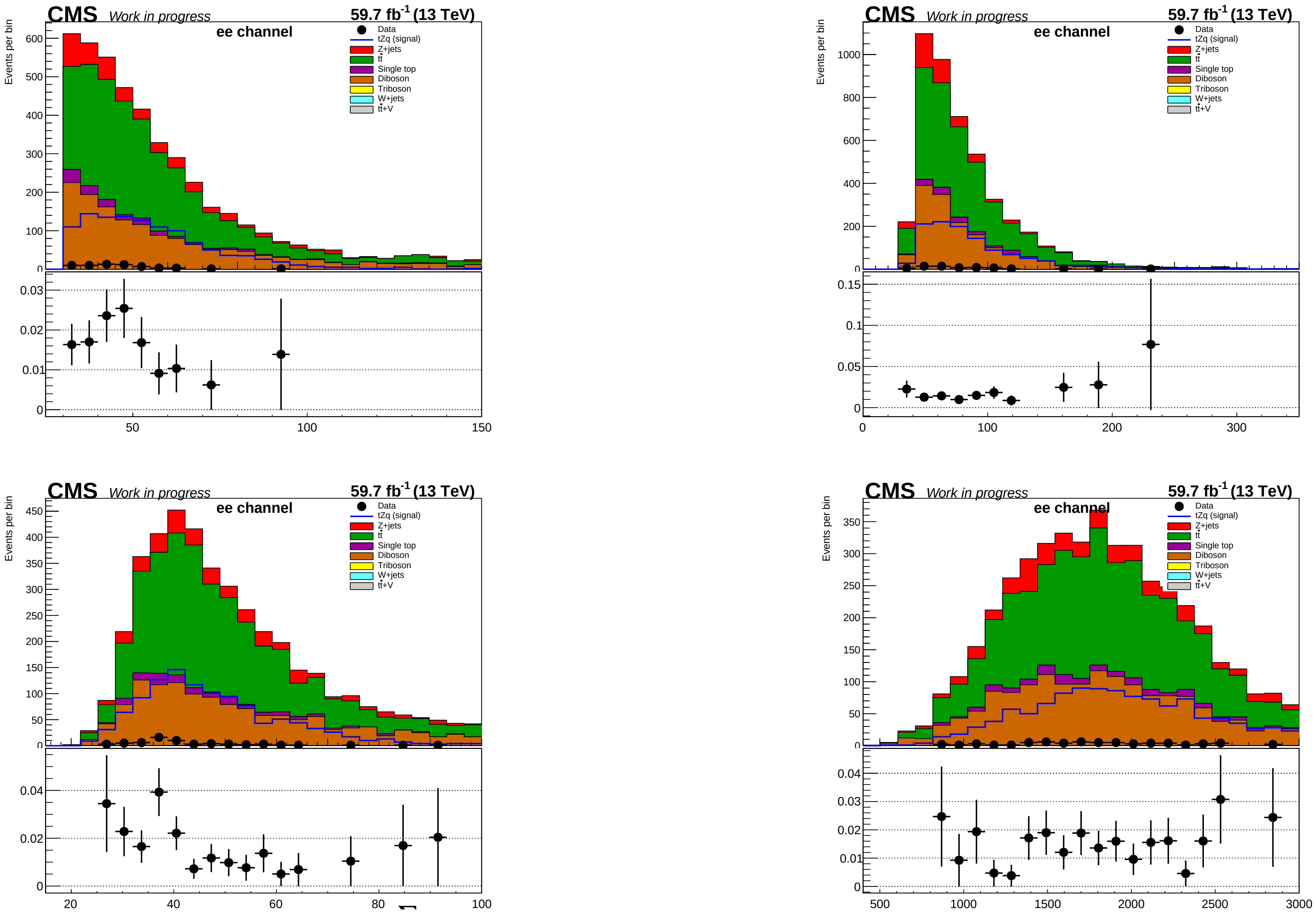


Figure 68: The distributions of the features: the transverse momentum of the fourth jet candidate, the sum of the masses of the four jet candidates, the transverse momentum of the leading lepton and the missing transverse energy. The features in this figure correspond to the *ee* channel in the side-band region, for simulation and data recorded in 2018.

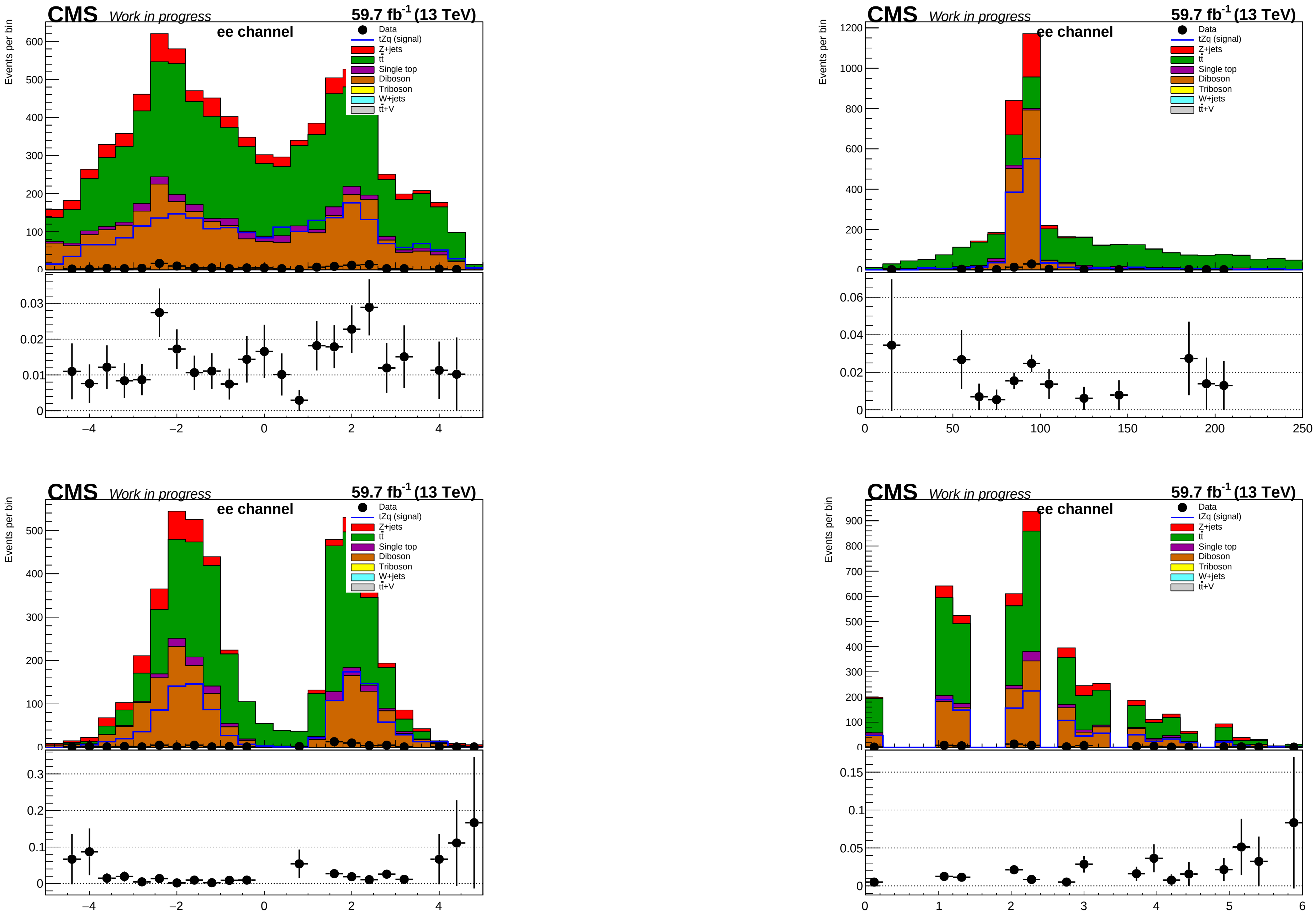


Figure 69: The distributions of the features: the pseudorapidity of the reconstructed W boson candidate, the pseudorapidity the reconstructed Z boson candidate, the invariant mass of the reconstructed Z boson candidate and the ΔR between the reconstructed Z boson and the leading b-jet candidates. The features in this figure correspond to the *ee* channel in the side-band region, for simulation and data recorded in 2018.

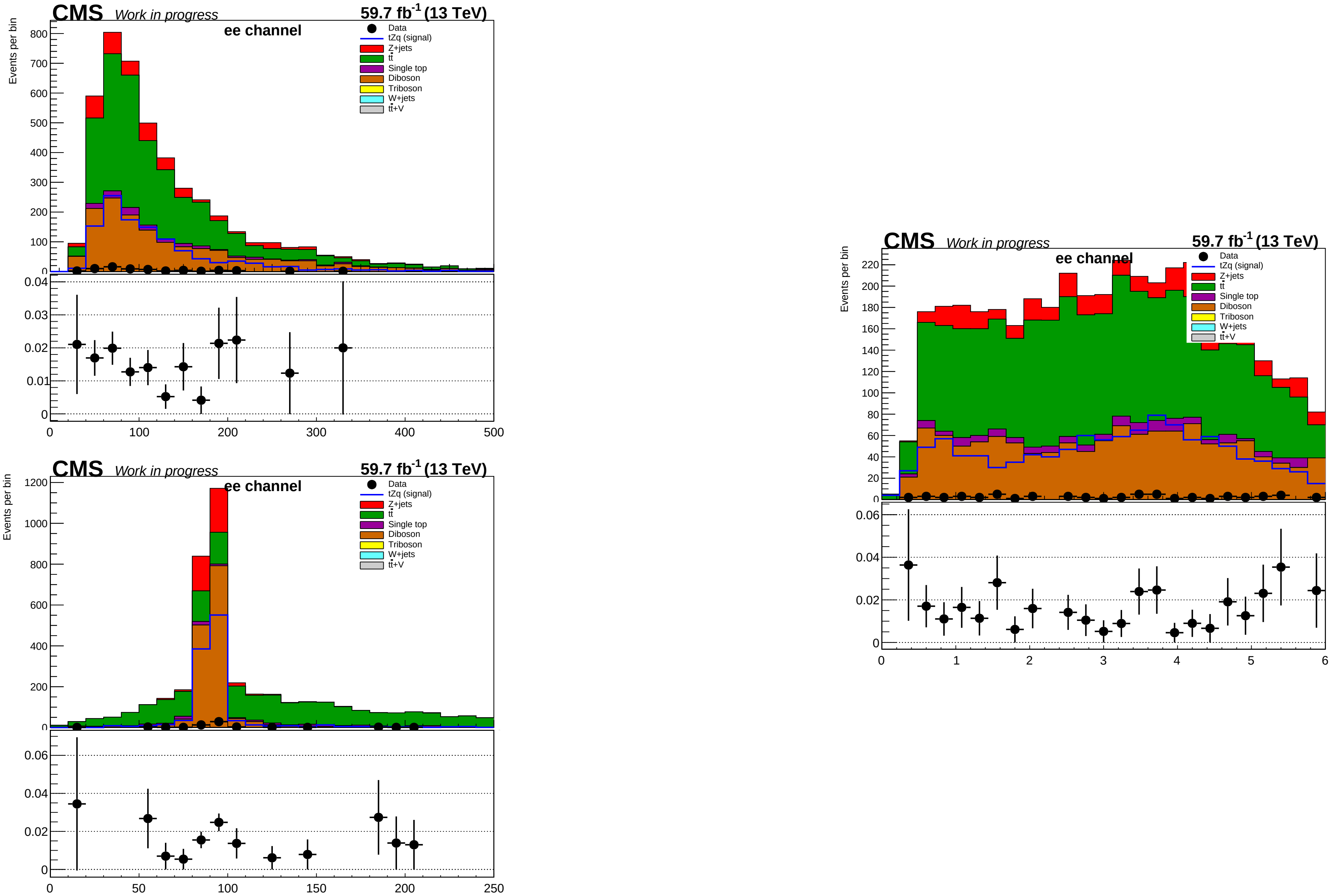


Figure 70: The distribution of the features: the transverse momentum of the leading jet from the reconstructed W boson candidate, the invariant mass of the reconstructed Z boson candidate and the maximum ΔR between the Z boson candidate and any jet candidate in the system. The features in this figure corresponds to the *ee* channel in the side-band region, for simulation and data recorded in 2018.

10.2.6 2018: $\mu\mu$ channel, side-band region

Figures 71- 73 show the distributions of selected features that were used as inputs for the multivariate analysis. These correspond to 2018 data and simulation in the *ee* channel, and in the side-band region. The selected features are: b-tagged discriminator of the leading b-tagged jet, the χ^2 variable used for experimental blinding, the pseudorapidity of of the leading jet candidate, the transverse momentum of the subleading jet candidate, the transverse momentum of the fourth jet candidate, the sum of the masses of the four jet candidates, the missing transverse energy, the transverse momentum of the subleading jet from the reconstructed W boson candidate, the invariant mass of the reconstructed Z boson candidate, the transverse momentum of the reconstructed Z boson candidate, the ΔR between the reconstructed Z boson candidate and the leading bjet and the maximum ΔR between the Z boson candidate and any jet candidate in the system.

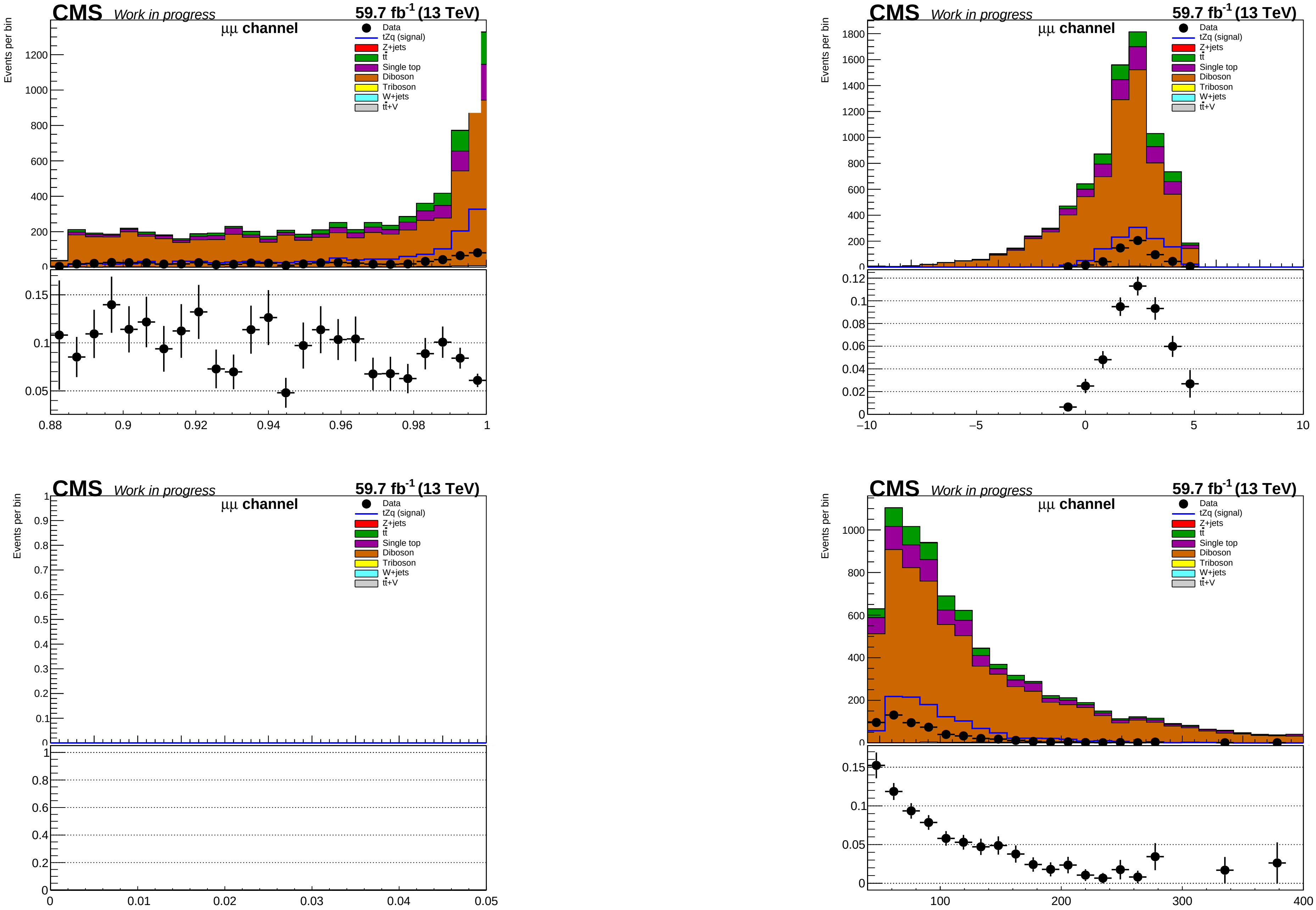


Figure 71: The distributions of the features: b-tagged discriminator of the leading b-tagged jet, the χ^2 variable used for experimental blinding, the pseudorapidity of of the leading jet candidate and the transverse momentum of the subleading jet candidate. The features in this figure correspond to the $\mu\mu$ channel in the side-band region, for simulation and data recorded in 2018.

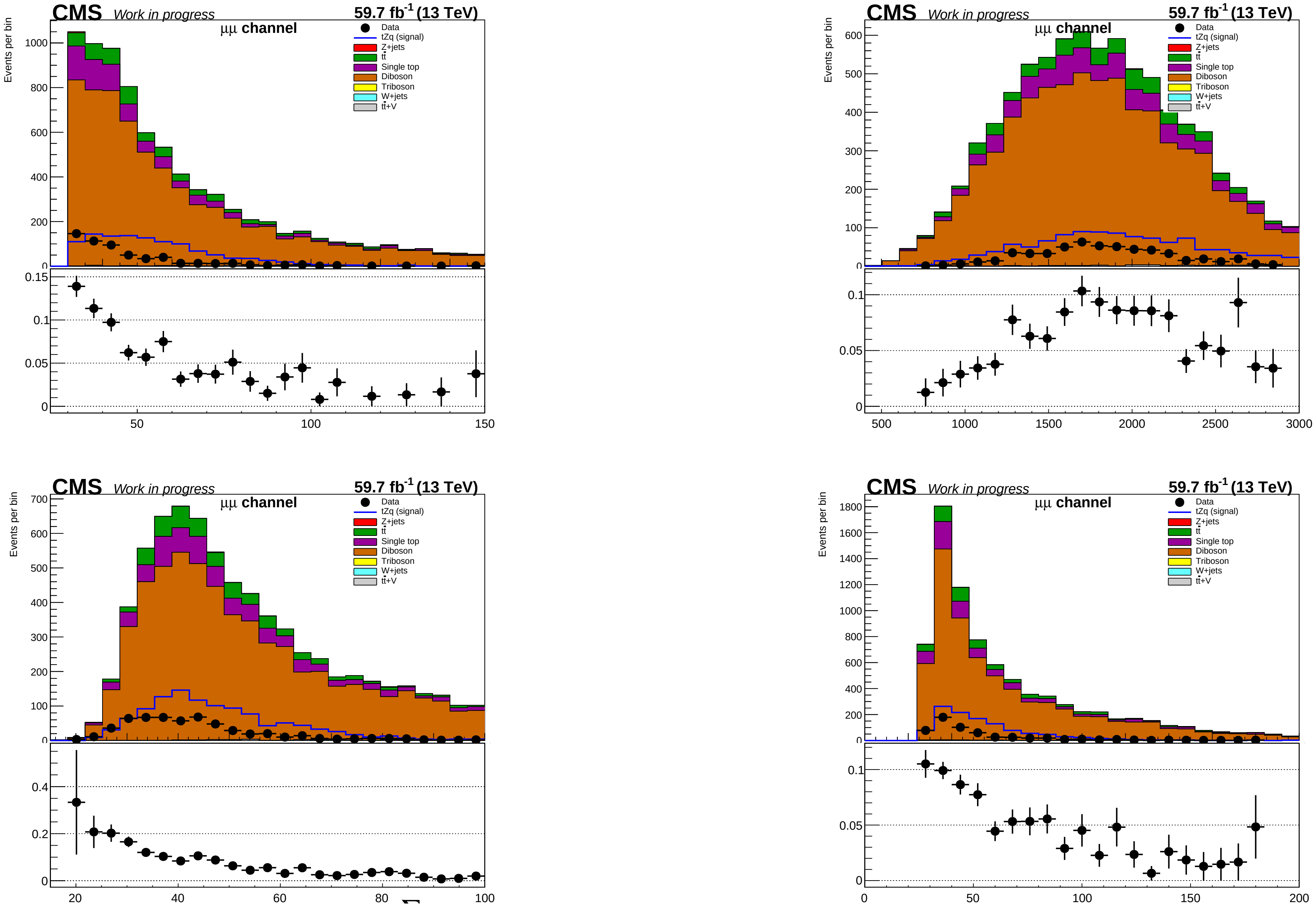


Figure 72: The distributions of the features: the transverse momentum of the fourth jet candidate, the sum of the masses of the four jet candidates, the missing transverse energy and the transverse momentum of the subleading jet from the reconstructed W boson candidate. The features in this figure correspond to the $\mu\mu$ channel in the side-band region, for simulation and data recorded in 2018.

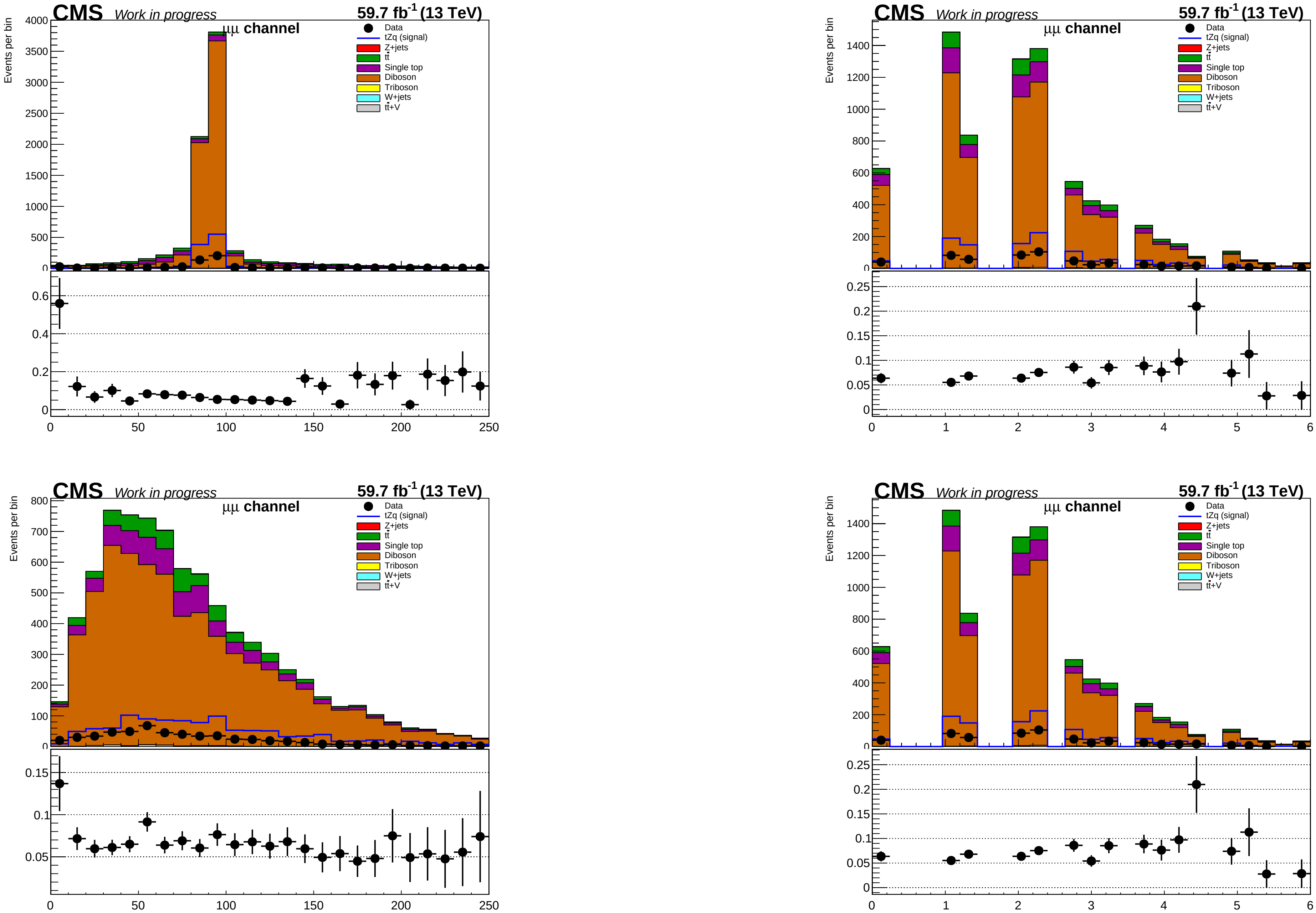


Figure 73: The distributions of the features: the invariant mass of the reconstructed Z boson candidate, the transverse momentum of the reconstructed Z boson candidate, the ΔR between the reconstructed Z boson candidate and the leading bjet and the maximum ΔR between the Z boson candidate and any jet candidate in the system. The features in this figure correspond to the $\mu\mu$ channel in the side-band region, for simulation and data recorded in 2018.

10.2.7 2016: $e\mu$ channel, $t\bar{t}$ control region

Figures 74- 76 show some of the distributions obtained using 2016 data and simulation in the $e\mu$ channel, and in the $t\bar{t}$ control region. The selected features were: b-tagged discriminator of the leading b-tagged jet, the χ^2 variable used for experimental blinding, the pseudorapidity of of the leading jet candidate, the transverse momentum of the subleading jet candidate, the transverse momentum of the fourth jet candidate, the sum of the masses of the four jet candidates, the missing transverse energy, the transverse momentum of the subleading jet from the reconstructed W boson candidate, the invariant mass of the reconstructed Z boson candidate, the transverse momentum of the reconstructed Z boson candidate, the ΔR between the reconstructed Z boson candidate and the leading bjet and the maximum ΔR between the Z boson candidate and any jet candidate in the system.

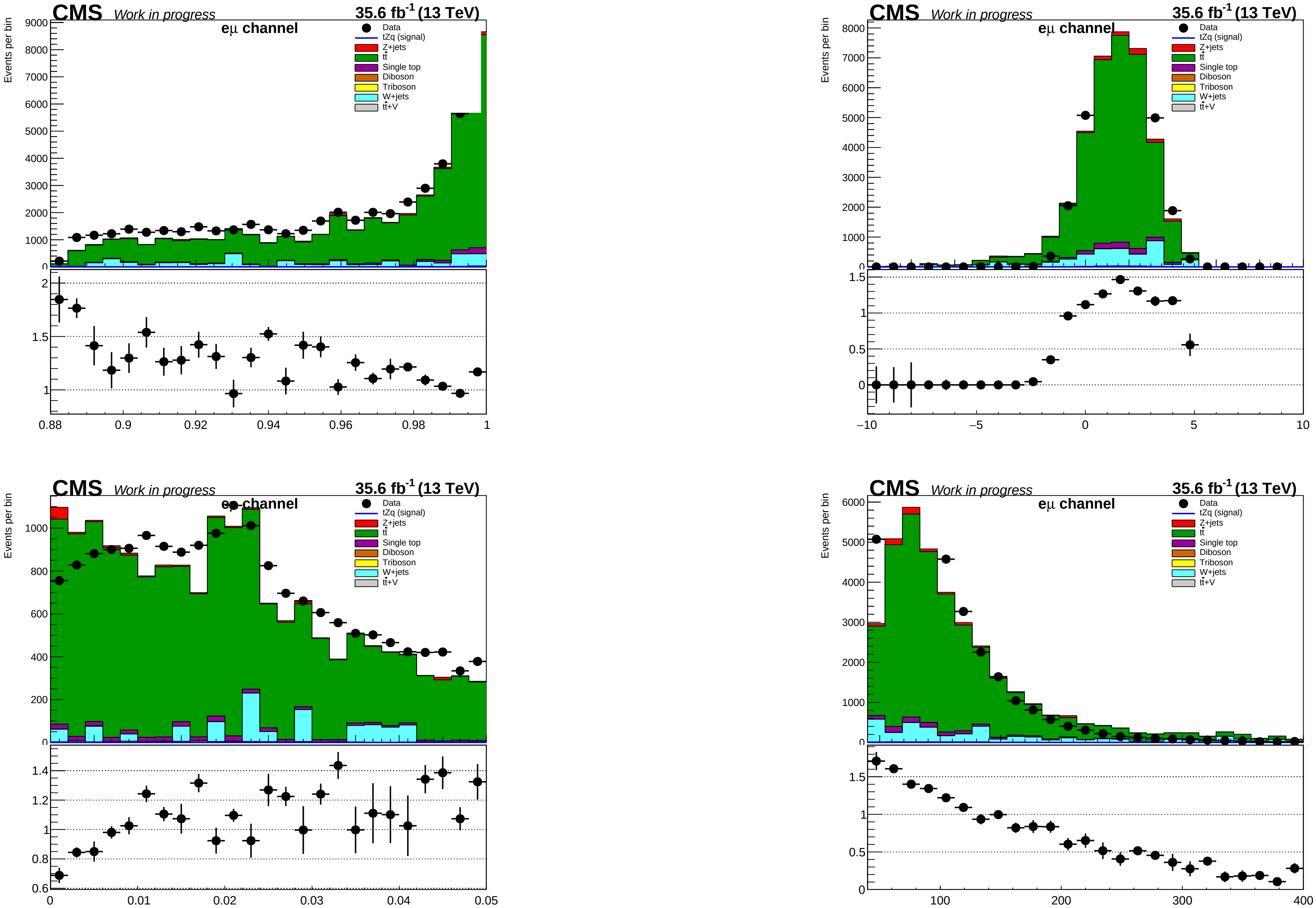


Figure 74: The distributions of the features: b-tagged discriminator of the leading b-tagged jet, the χ^2 variable used for experimental blinding, the pseudorapidity of of the leading jet candidate and the transverse momentum of the subleading jet candidate. The features in this figure correspond to the $e\mu$ channel in the $t\bar{t}$ control region, for simulation and data recorded in 2016.

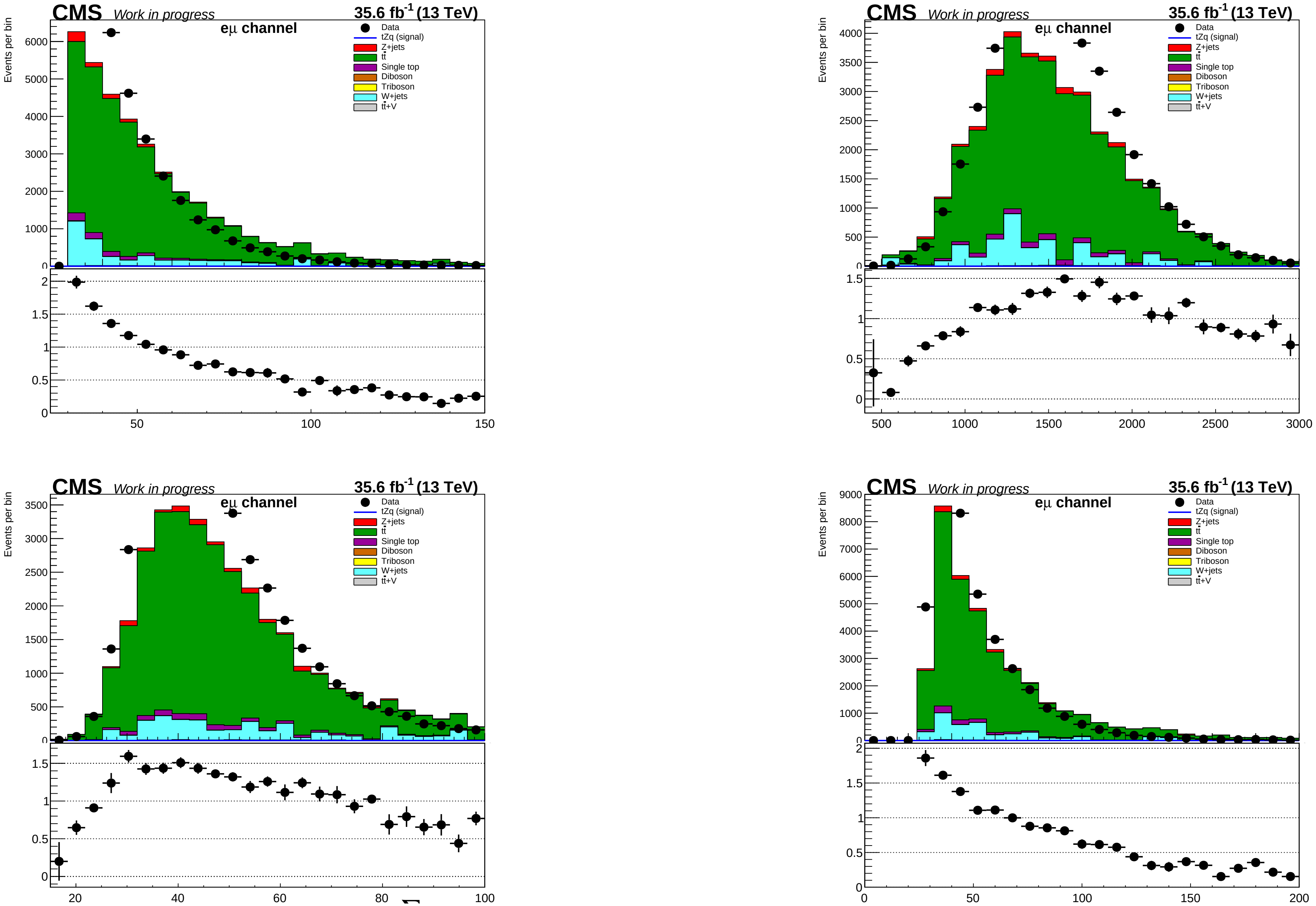


Figure 75: The distributions of the features: the transverse momentum of the fourth jet candidate, the sum of the masses of the four jet candidates, the missing transverse energy and the transverse momentum of the subleading jet from the reconstructed W boson candidate. The features in this figure correspond to the $e\mu$ channel in the $t\bar{t}$ control region, for simulation and data recorded in 2016.

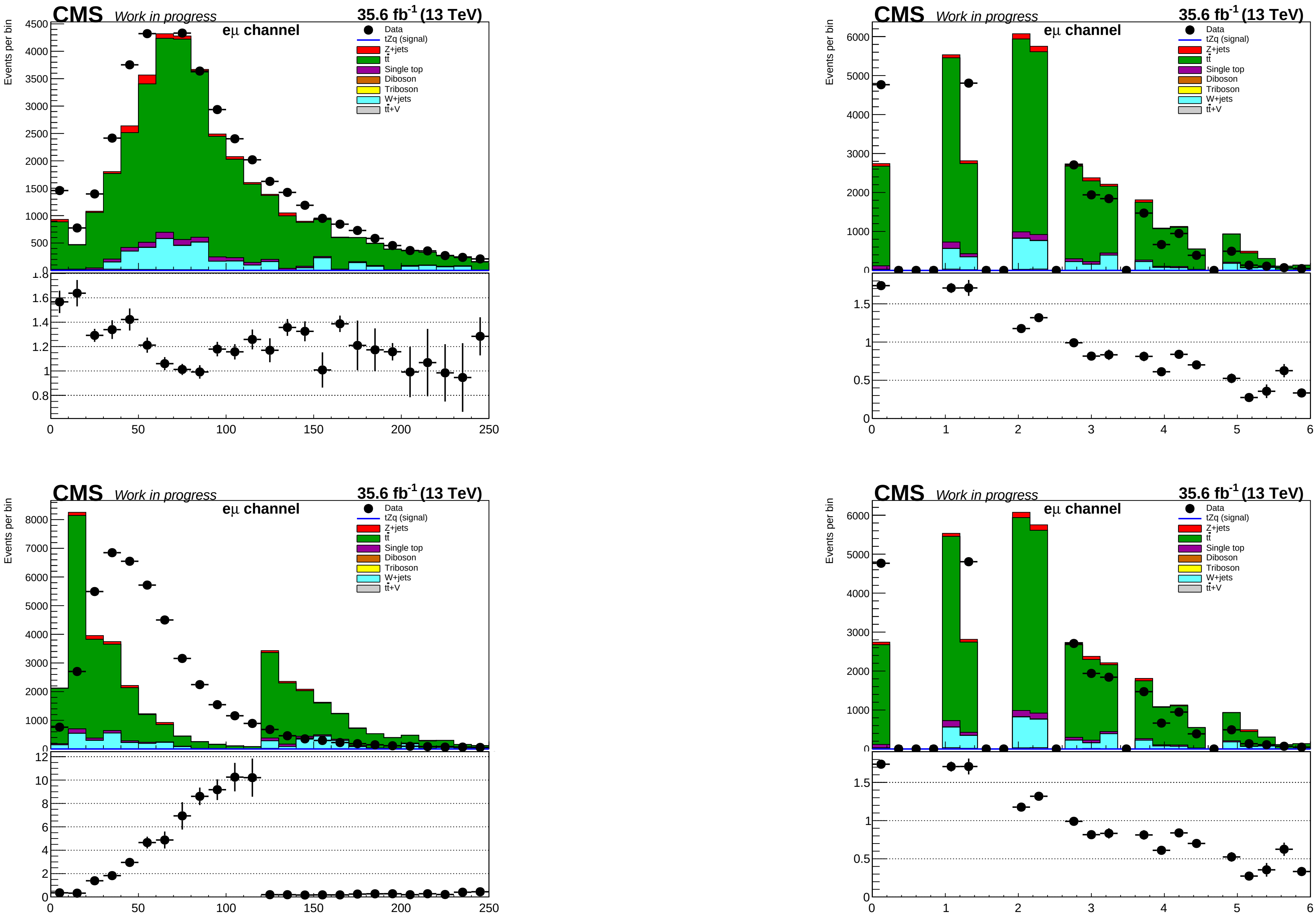


Figure 76: The distributions of the features: the invariant mass of the reconstructed Z boson candidate, the transverse momentum of the reconstructed Z boson candidate, the ΔR between the reconstructed Z boson candidate and the leading bjet and the maximum ΔR between the Z boson candidate and any jet candidate in the system. The features in this figure correspond to the $e\mu$ channel in the $t\bar{t}$ region, for simulation and data recorded in 2016.

10.2.8 2017: $e\mu$ channel, $t\bar{t}$ control region

Figures 77- 79 show some of the distributions obtained using 2017 data and simulation in the $e\mu$ channel, and in the $t\bar{t}$ control region. The selected features were: b-tagged discriminator of the leading b-tagged jet, the χ^2 variable used for experimental blinding, the pseudorapidity of of the leading jet candidate, the transverse momentum of the subleading jet candidate, the transverse momentum of the fourth jet candidate, the sum of the masses of the four jet candidates, the missing transverse energy, the transverse momentum of the subleading jet from the reconstructed W boson candidate, the invariant mass of the reconstructed Z boson candidate, the transverse momentum of the reconstructed Z boson candidate, the ΔR between the reconstructed Z boson candidate and the leading bjet and the maximum ΔR between the Z boson candidate and any jet candidate in the system.

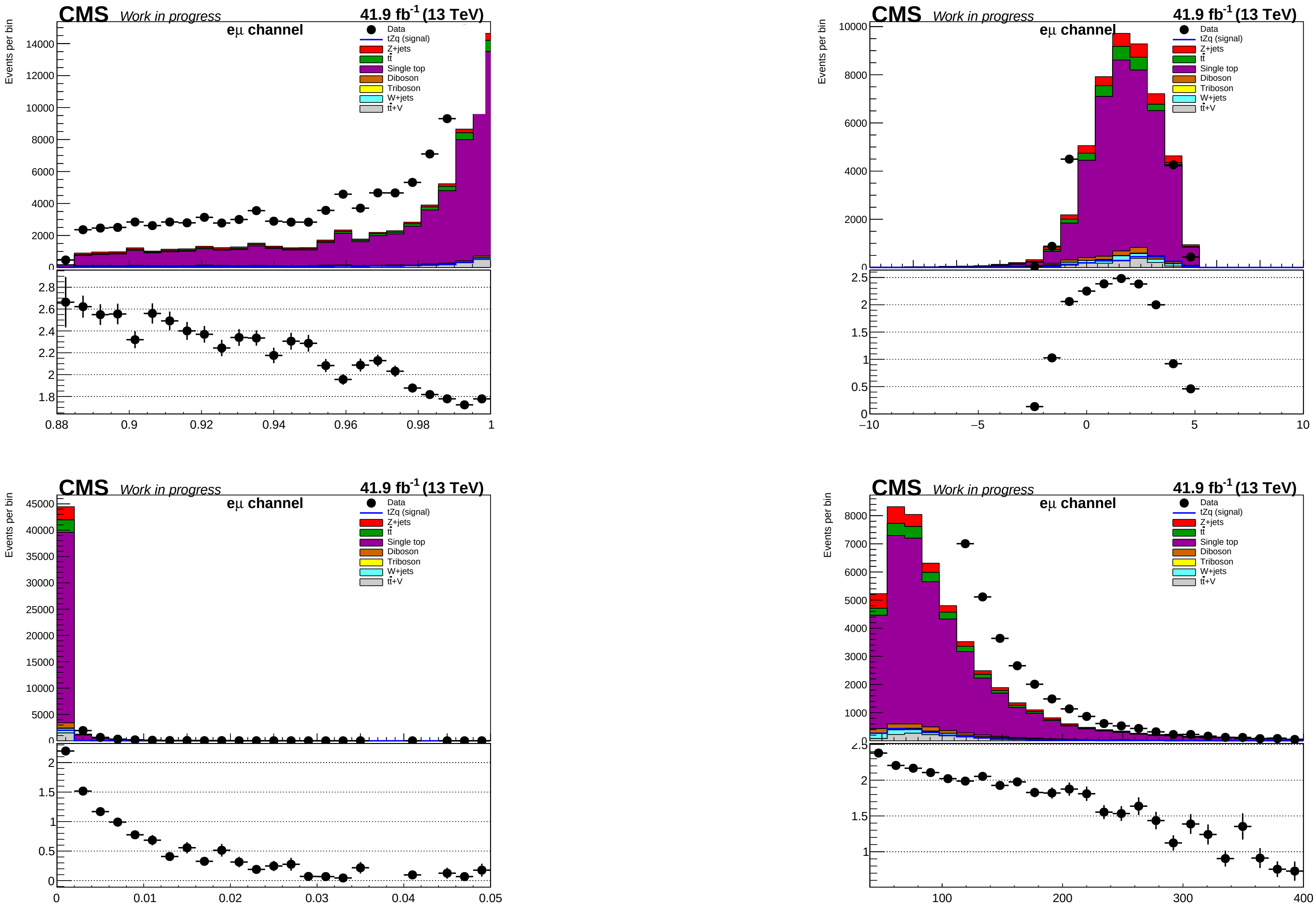


Figure 77: The distributions of the features: b-tagged discriminator of the leading b-tagged jet, the χ^2 variable used for experimental blinding, the pseudorapidity of of the leading jet candidate and the transverse momentum of the subleading jet candidate. The features in this figure correspond to the $e\mu$ channel in the $t\bar{t}$ control region, for simulation and data recorded in 2017.

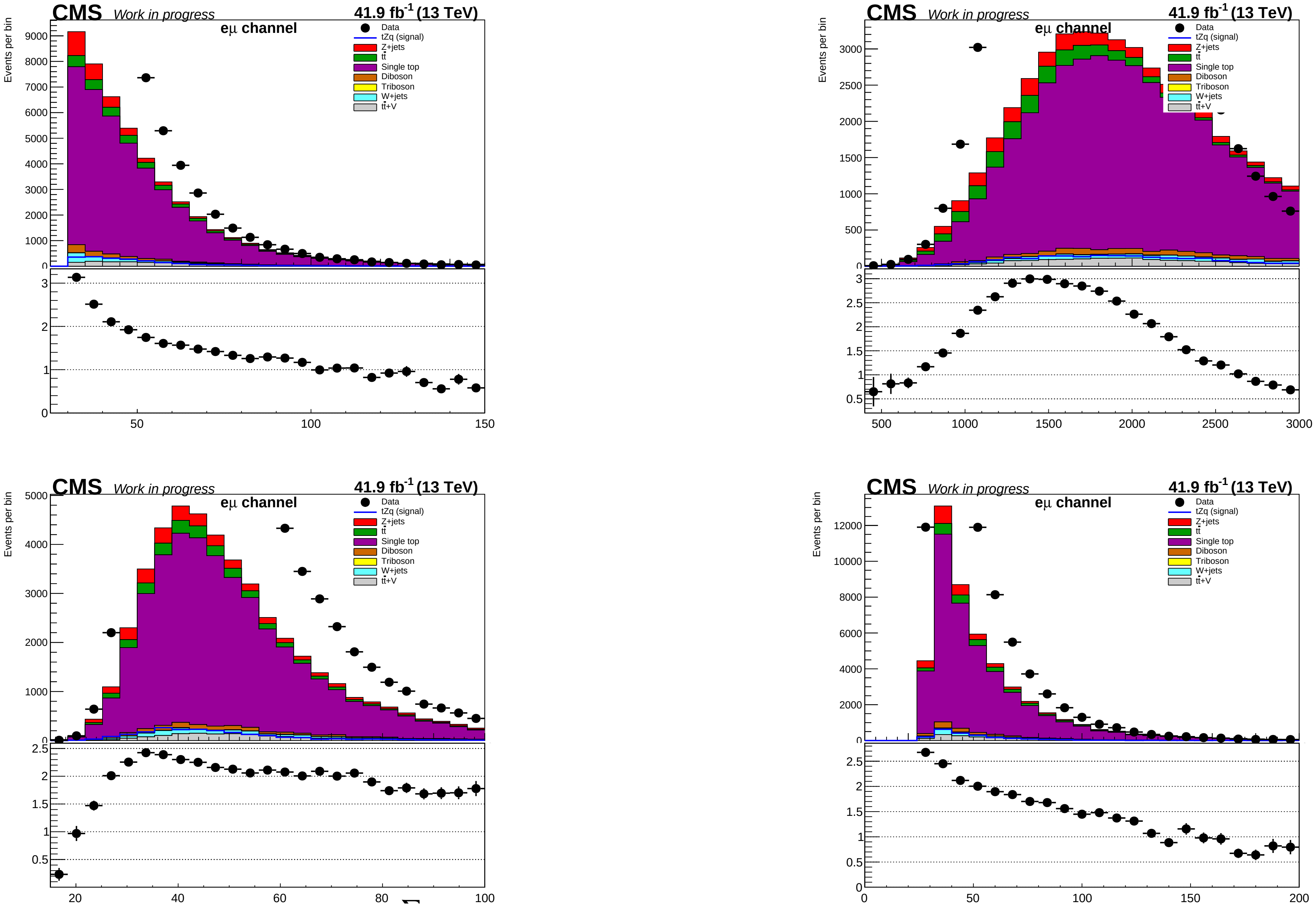


Figure 78: The distributions of the features: the transverse momentum of the fourth jet candidate, the sum of the masses of the four jet candidates, the missing transverse energy and the transverse momentum of the subleading jet from the reconstructed W boson candidate. The features in this figure correspond to the $e\mu$ channel in the $t\bar{t}$ region, for simulation and data recorded in 2017.

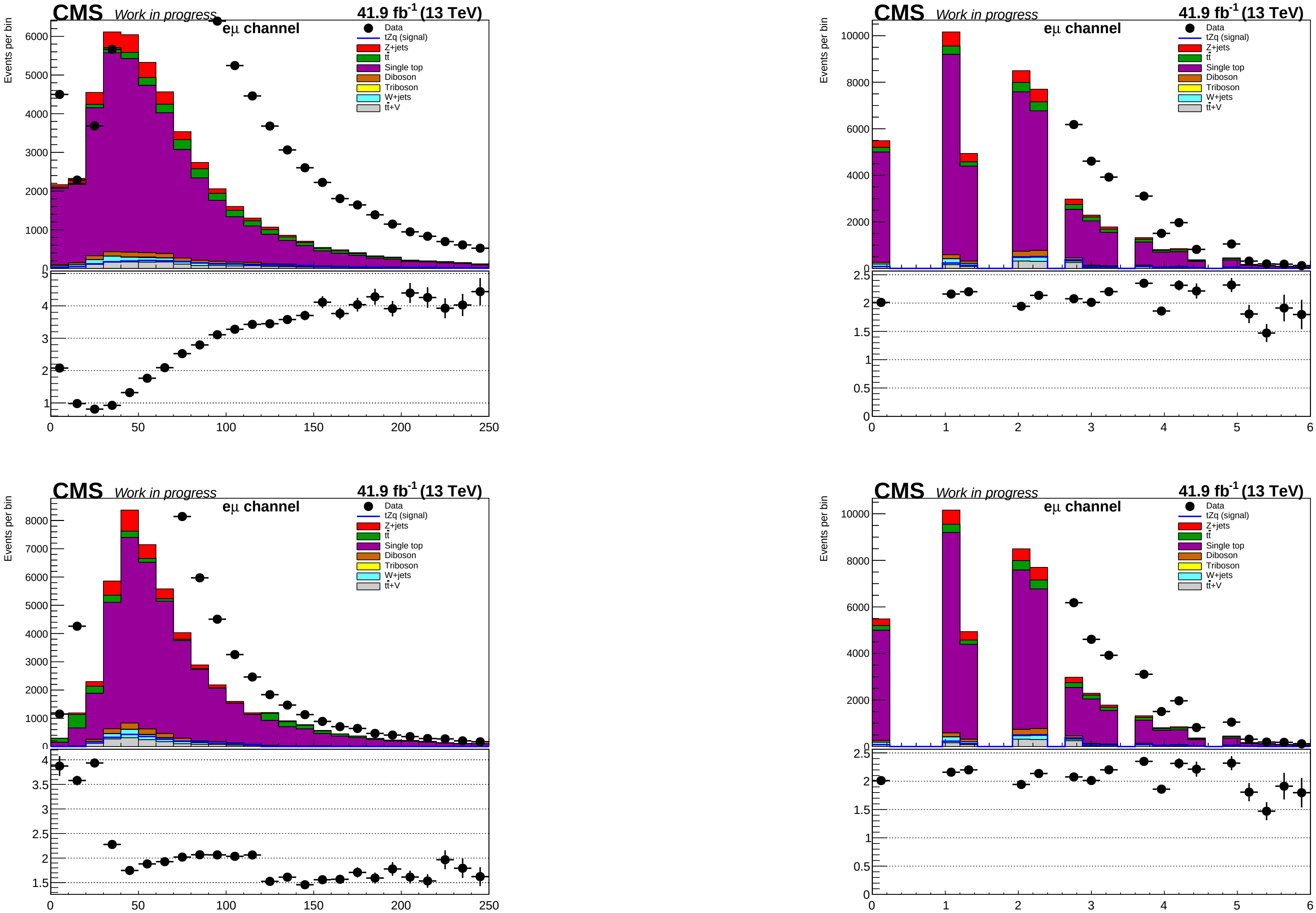


Figure 79: The distributions of the features: the invariant mass of the reconstructed Z boson candidate, the transverse momentum of the reconstructed Z boson candidate, the ΔR between the reconstructed Z boson candidate and the leading bjet and the maximum ΔR between the Z boson candidate and any jet candidate in the system. The features in this figure correspond to the $e\mu$ channel in the $t\bar{t}$ region, for simulation and data recorded in 2017.

10.2.9 2018: $e\mu$ channel, $t\bar{t}$ control region

Figures 80- 82 show some of the distributions obtained using 2018 data and simulation in the $e\mu$ channel, and in the $t\bar{t}$ control region. The selected features are: b-tagged discriminator of the leading b-tagged jet, the χ^2 variable used for experimental blinding, the pseudorapidity of of the leading jet candidate, the transverse momentum of the subleading jet candidate, the transverse momentum of the fourth jet candidate, the sum of the masses of the four jet candidates, the missing transverse energy, the transverse momentum of the subleading jet from the reconstructed W boson candidate, the invariant mass of the reconstructed Z boson candidate, the transverse momentum of the reconstructed Z boson candidate, the ΔR between the reconstructed Z boson candidate and the leading bjet and the maximum ΔR between the Z boson candidate and any jet candidate in the system. Due to time constraints an additional study to identify the features with the highest discriminating power was not undertaken, so the same features that were chosen for the 2016 and 2017 analyses were used for 2018.

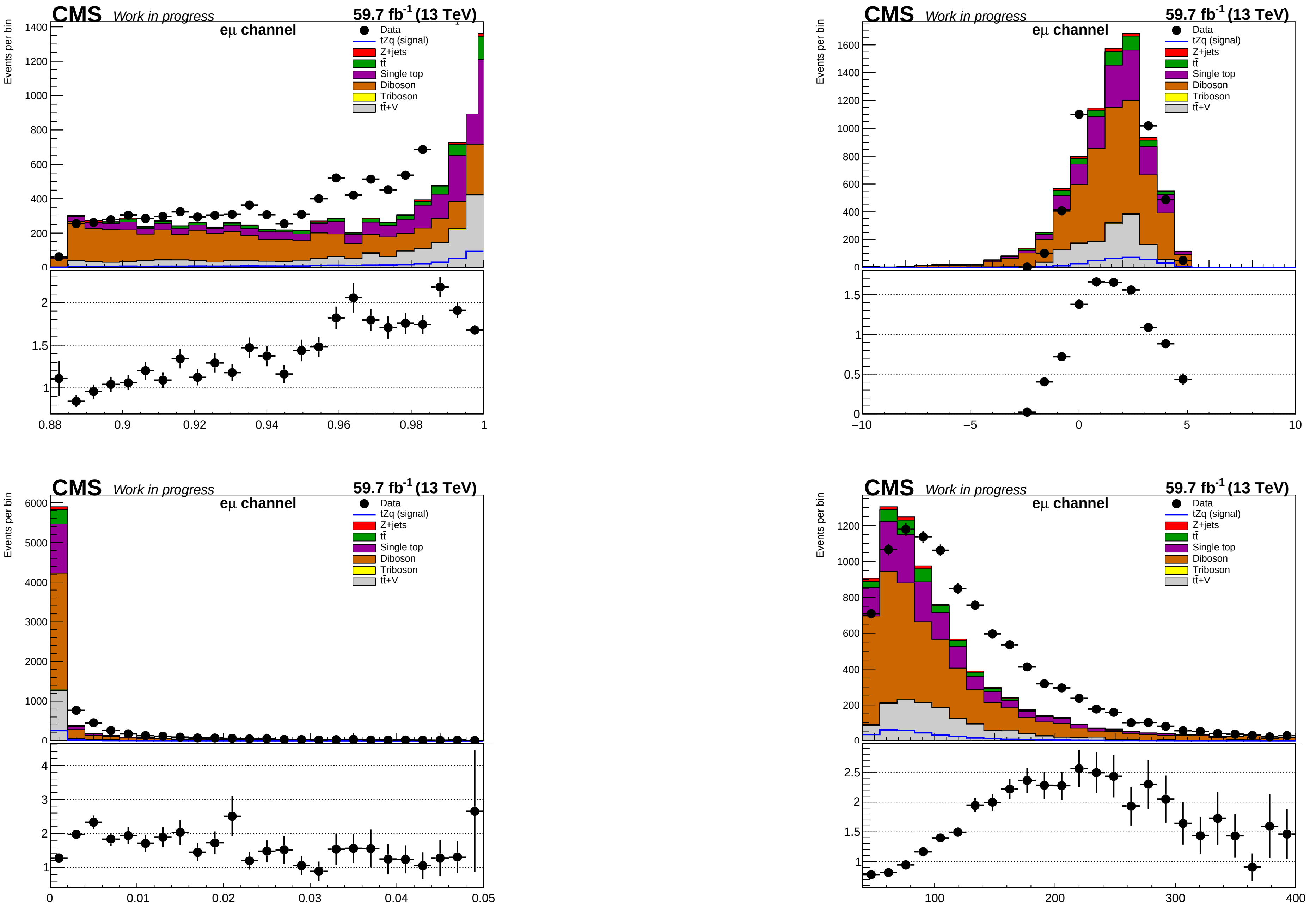


Figure 80: The distributions of the features: b-tagged discriminator of the leading b-tagged jet, the χ^2 variable used for experimental blinding, the pseudorapidity of of the leading jet candidate and the transverse momentum of the subleading jet candidate. The features in this figure correspond to the $e\mu$ channel in the $t\bar{t}$ control region, for simulation and data recorded in 2018.

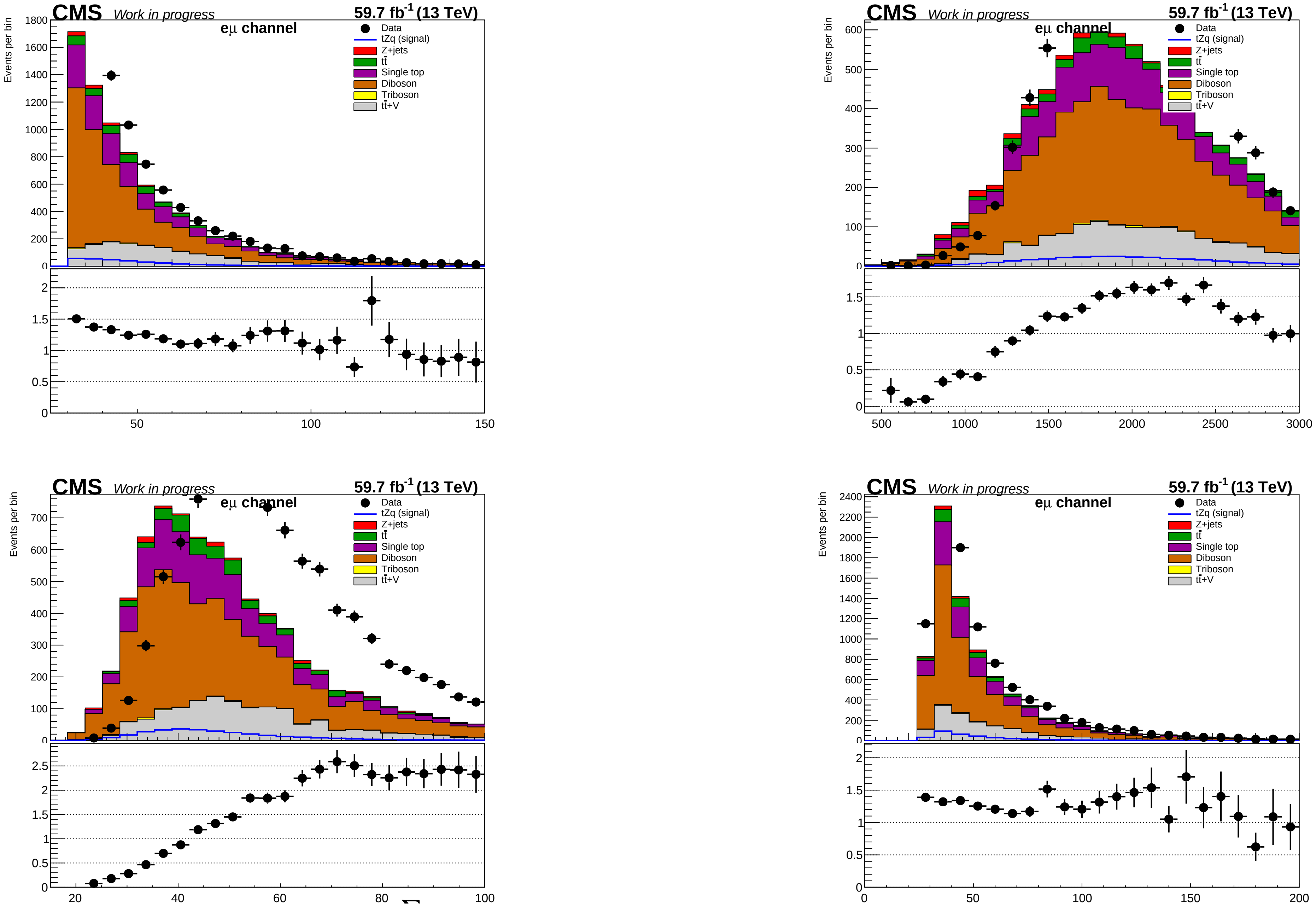


Figure 81: The distributions of the features: the transverse momentum of the fourth jet candidate, the sum of the masses of the four jet candidates, the missing transverse energy and the transverse momentum of the subleading jet from the reconstructed W boson candidate. The features in this figure correspond to the $e\mu$ channel in the $t\bar{t}$ control region, for simulation and data recorded in 2018.

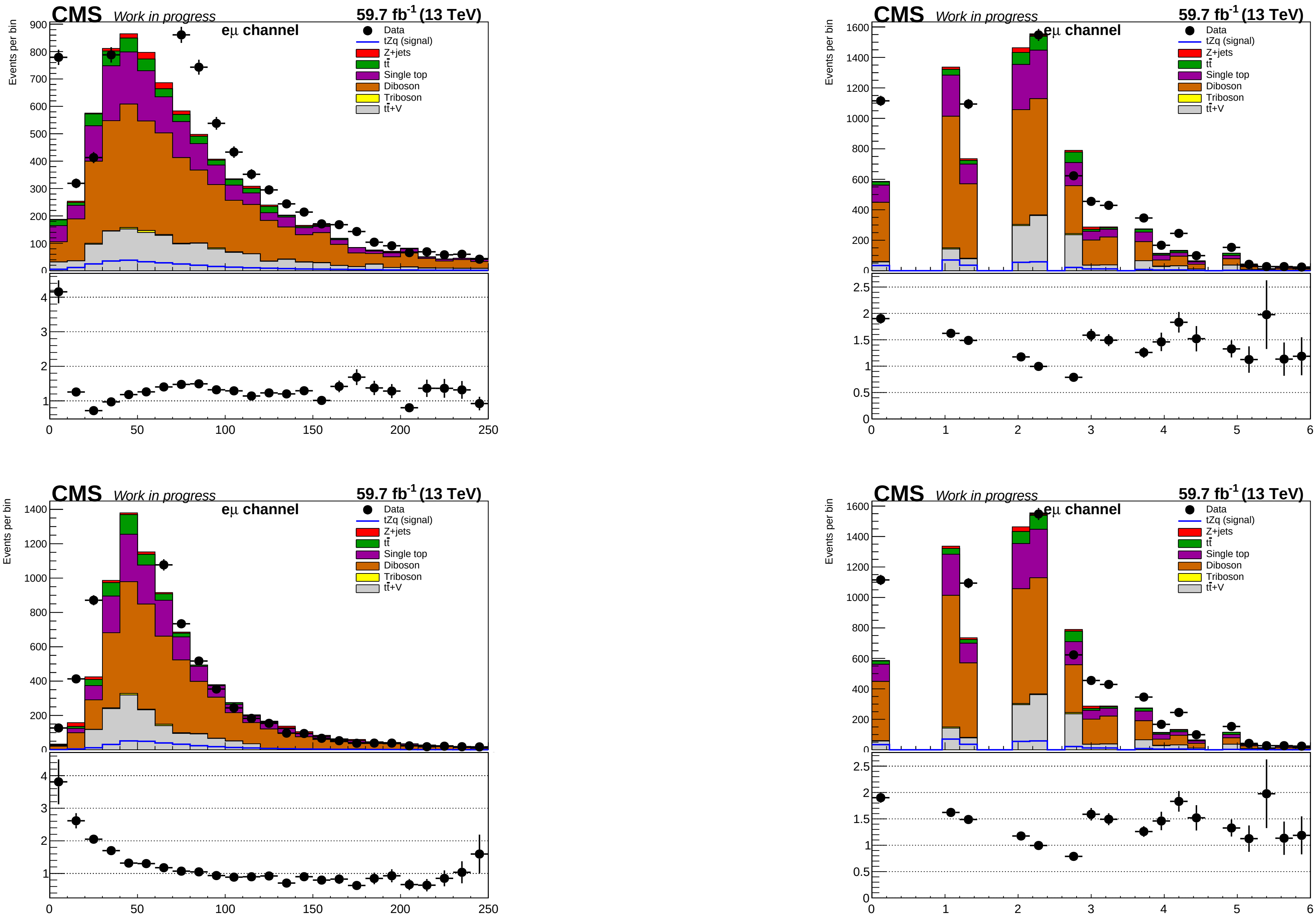


Figure 82: The distributions of the features: the invariant mass of the reconstructed Z boson candidate, the transverse momentum of the reconstructed Z boson candidate, the ΔR between the reconstructed Z boson candidate and the leading bjet and the maximum ΔR between the Z boson candidate and any jet candidate in the system. The features in this figure correspond to the $e\mu$ channel in the $t\bar{t}$ control region, for simulation and data recorded in 2018.

10.2.10 2016: ee channel, $Z + jets$ control region

Figures 83- 86 show the distributions of selected features that were used as inputs for the multivariate analysis. These correspond to 2016 data and simulation in the ee channel, and in the $Z + jets$ control region. The selected features are: b-tagged discriminator of the leading b-tagged jet, the transverse momentum of the leading jet candidate, ΔR between the leading and subleading jet candidates, ΔR between the second and third jet candidates, the transverse momentum of the fourth jet candidate, the sum of the masses of the four jet candidates, the missing transverse energy, the invariant top mass, the pseudorapidity of the subleading jet from the reconstructed W boson candidate, the mass of the reconstructed Z boson candidate, the transverse momentum of the reconstructed Z boson candidate, ΔR between the reconstructed Z boson and the leading b-jet candidates and the maximum ΔR between the Z boson candidate and any jet candidate in the system. These features were chosen as they were previously found to have the highest discriminating power [46] [45].

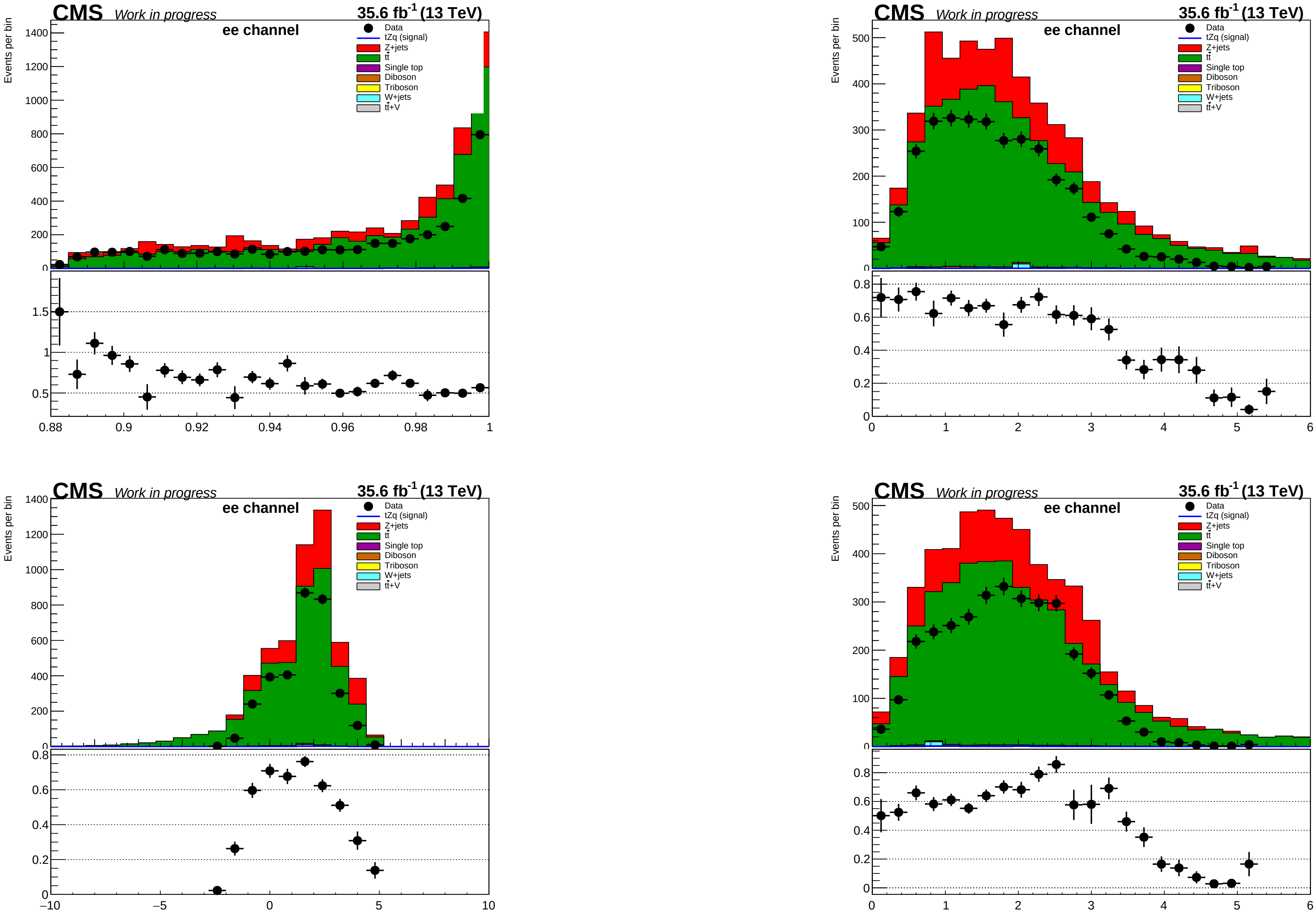


Figure 83: The distributions of the features: b-tagged discriminator of the leading b-tagged jet, the transverse momentum of the leading jet candidate, ΔR between the leading and subleading jet and ΔR between the second and third jet candidates. The features in this figure correspond to the ee channel in the $Z + jets$ control region, for simulation and data recorded in 2016.

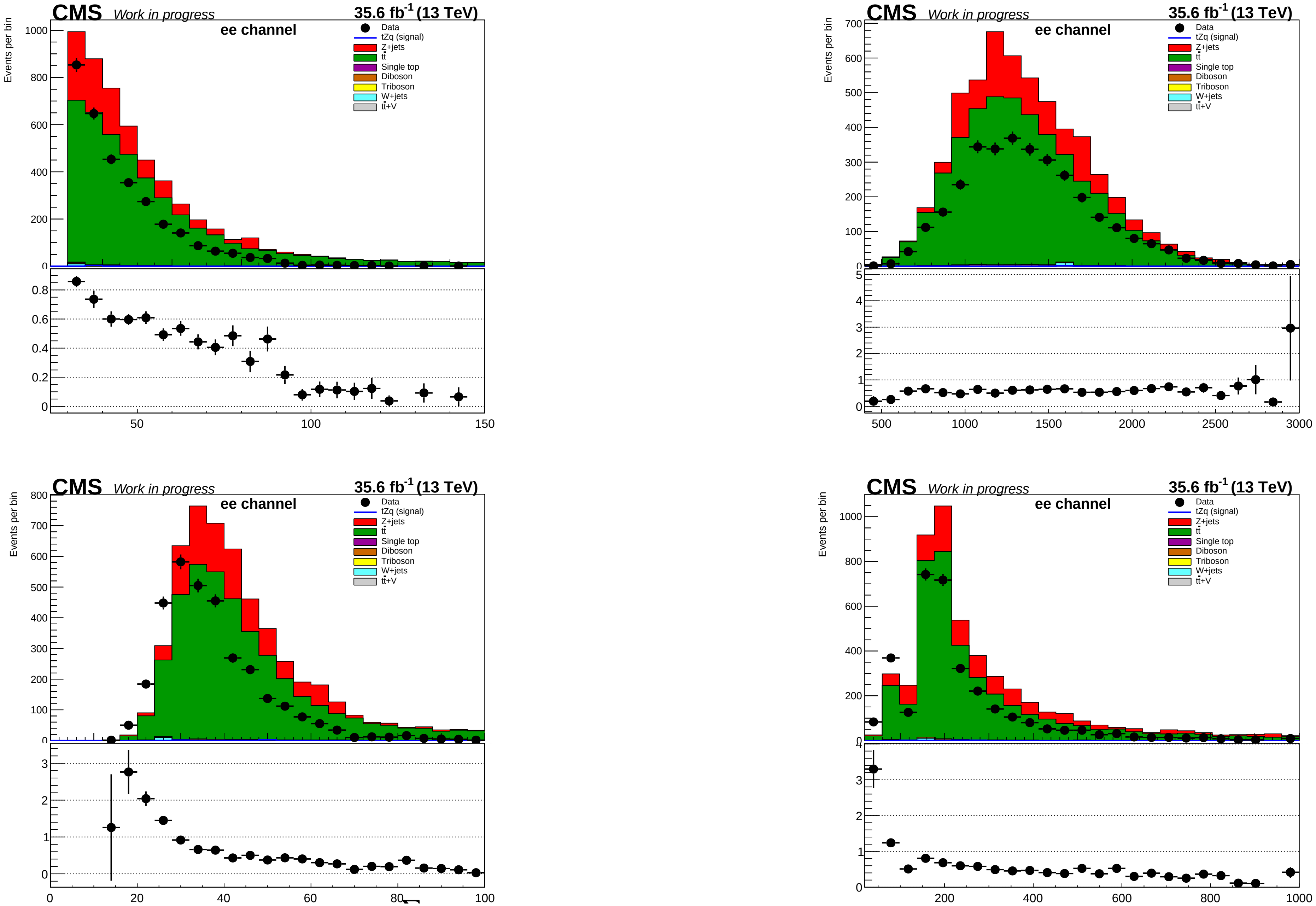


Figure 84: The distributions of the features: the transverse momentum of the fourth jet candidate, the sum of the masses of the four jet candidates, the missing transverse energy and the invariant mass of the top quark candidate. The features in this figure correspond to the *ee* channel in the $Z + jets$ control region, for simulation and data recorded in 2016.

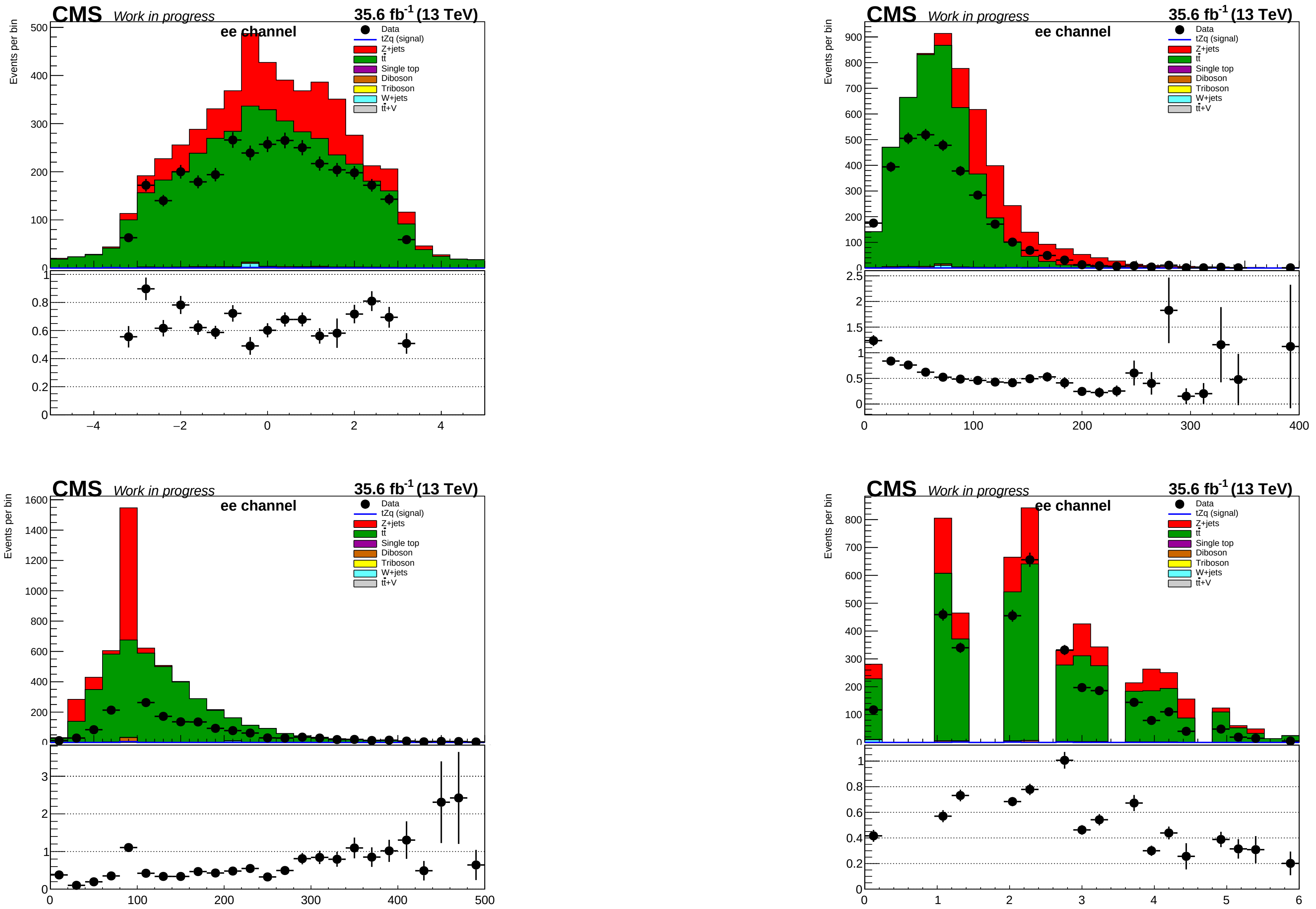


Figure 85: The distributions of the features: the pseudorapidity of the subleading jet from the reconstructed W boson candidate, the mass of the reconstructed Z boson candidate, the transverse momentum of the reconstructed Z boson candidate and ΔR between the reconstructed Z boson and the leading b-jet candidate. The features in this figure correspond to the *ee* channel in the $Z + jets$ region, for simulation and data recorded in 2016.

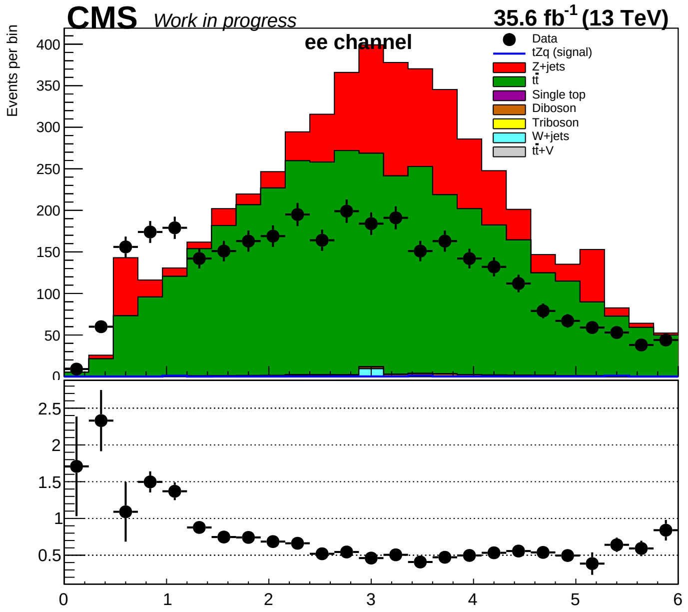


Figure 86: The distribution of the feature the maximum ΔR between the Z boson candidate and any jet candidate in the system. The feature in this figure corresponds to the ee channel in the $Z + jets$ control region, for simulation and data recorded in 2016.

10.2.11 2016: $\mu\mu$ channel, $Z + jets$ control region

Figures 87- 90 show the distributions of selected features that were used as inputs for the multivariate analysis. These correspond to 2016 data and simulation in the $\mu\mu$ channel, and in the $Z + jets$ control region. The selected features were: b-tagged discriminator of the leading b-tagged jet, the pseudorapidity of the leading jet candidate, between the leading and subleading jet candidates, the transverse momentum of the subleading jet candidate, the sum of the masses of the four jet candidates, the transverse momentum of the leading lepton, the missing transverse energy, the invariant top mass, the pseudorapidity of the reconstructed W boson candidate, the pseudorapidity the reconstructed Z boson candidate, the invariant mass of the reconstructed Z boson candidate, ΔR between the reconstructed Z boson and the leading b-jet candidates and the maximum ΔR between the Z boson candidate and any jet candidate in the system. These features are chosen as they were previously found to have the highest discriminating power [45] [46].

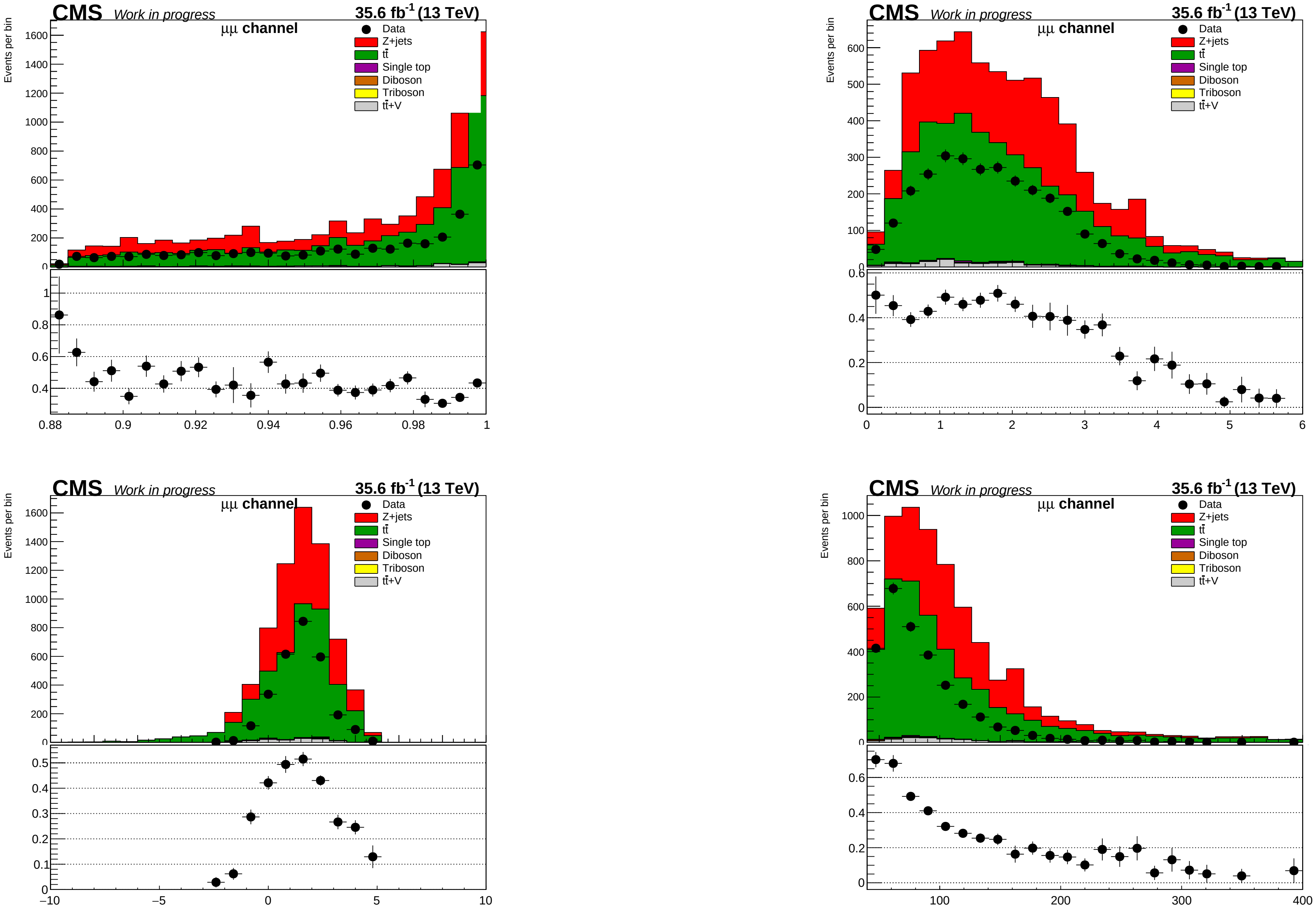


Figure 87: The distributions of the features: b-tagged discriminator of the leading b-tagged jet, the pseudorapidity of the leading jet candidate, ΔR between the leading and subleading jet candidates and the transverse momentum of the subleading jet candidate. The features in this figure correspond to the $\mu\mu$ channel in the $z + jets$ control region, for simulation and data recorded in 2016.

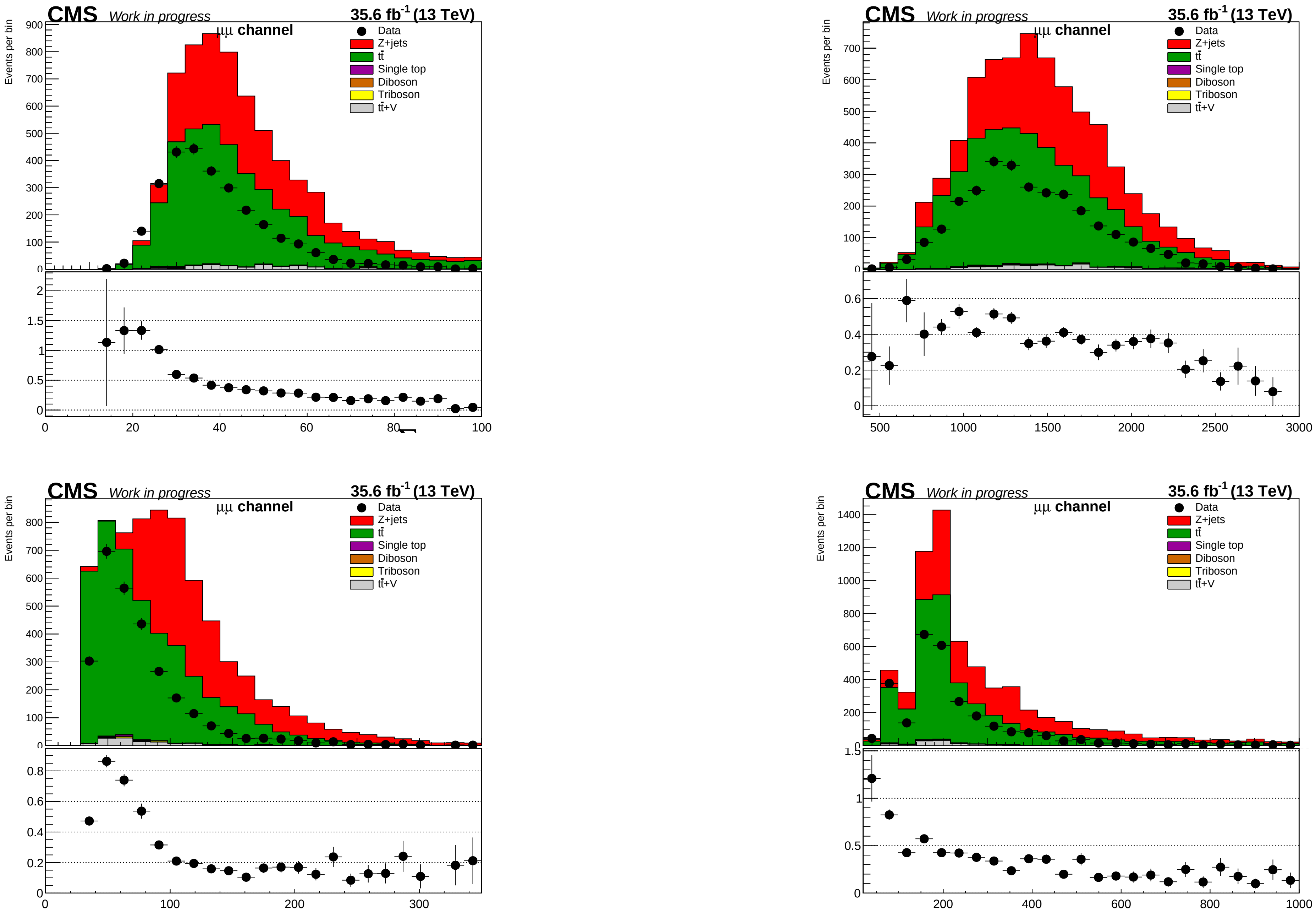


Figure 88: The distributions of the features: the sum of the masses of the four jet candidates, the transverse momentum of the leading lepton, the missing transverse energy and the invariant mass of the reconstructed top quark candidate. The features in this figure correspond to the $\mu\mu$ channel in the $z + jets$ control region, for simulation and data recorded in 2016.

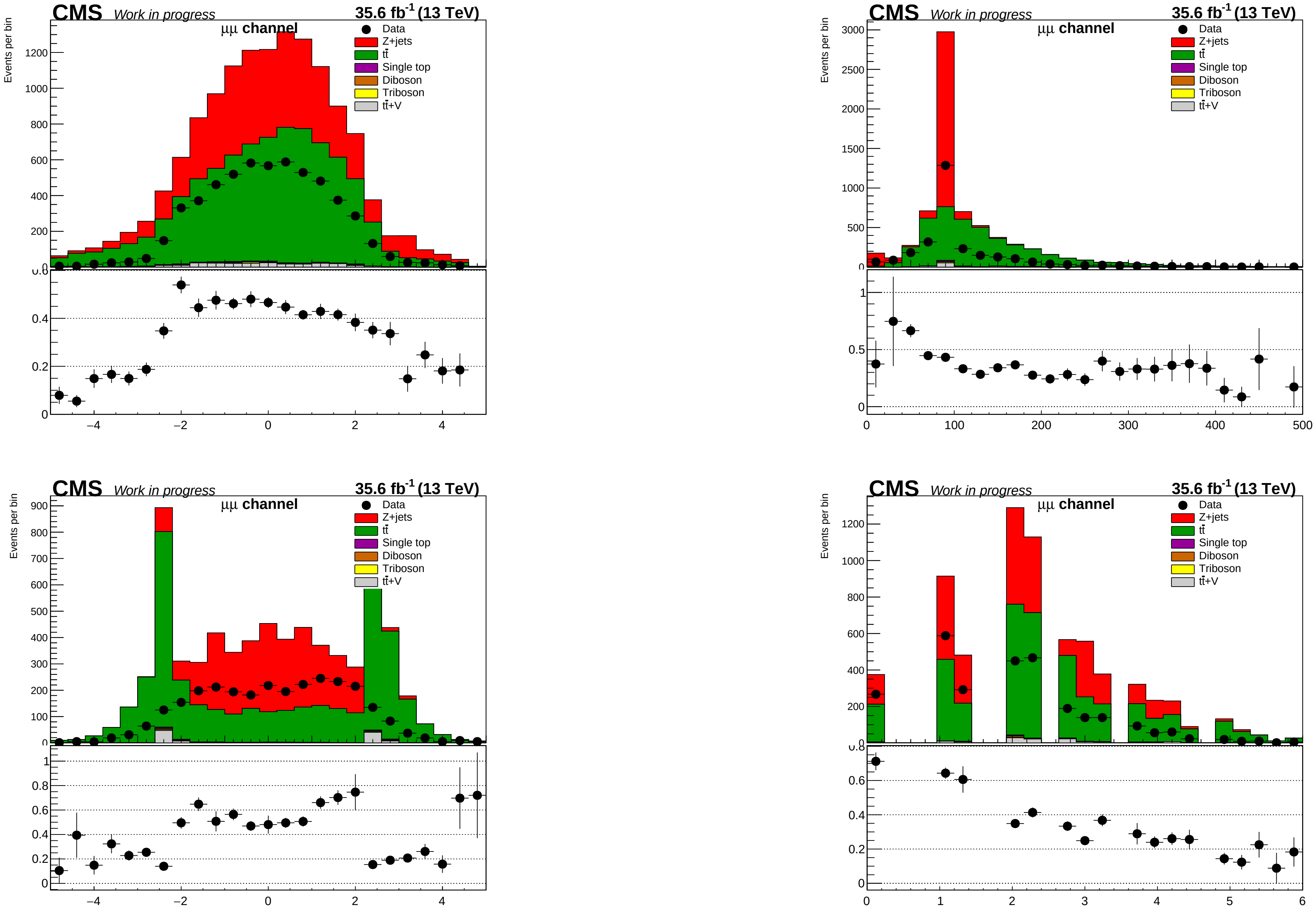


Figure 89: The distributions of the features: he pseudorapidity of the reconstructed W boson candidate, the pseudorapidity the reconstructed Z boson candidate, the invariant mass of the reconstructed Z boson candidate, ΔR between the reconstructed Z boson and the leading b-jet candidates. The features in this figure correspond to the $\mu\mu$ channel in the $z + jets$ control region, for simulation and data recorded in 2016.

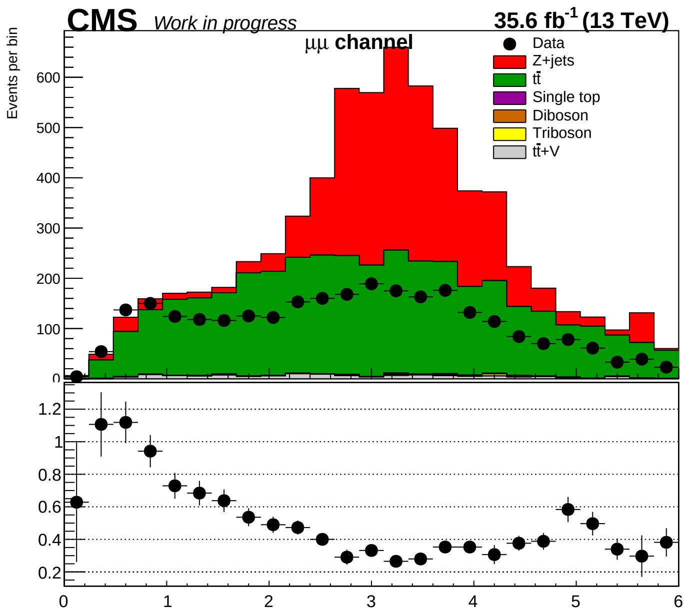


Figure 90: The distribution of the feature the maximum ΔR between the Z boson candidate and any jet candidate in the system. The feature in this figure corresponds to the $\mu\mu$ channel in the $z + jets$ control region, for simulation and data recorded in 2016.

10.2.12 2017: ee channel, $Z + jets$ control region

Figures 91- 94 show the distributions of selected features that are used as inputs for the multivariate analysis. These correspond to 2017 data and simulation in the ee channel, and in the $Z + jets$ control region. The selected features were: b-tagged discriminator of the leading b-tagged jet, the χ^2 variable used for experimental blinding, ΔR between the leading and subleading jet candidates, the transverse momentum of the subleading jet candidate, the transverse momentum of the fourth jet candidate, the sum of the masses of the four jet candidates, the transverse momentum of the leading lepton, the missing transverse energy, the transverse momentum of the leading jet from the reconstructed W boson candidate, the invariant mass of the reconstructed Z boson candidate and the maximum ΔR between the Z boson candidate and any jet candidate in the system. These features were chosen as they were previously found to have the highest discriminating power [45] [46].

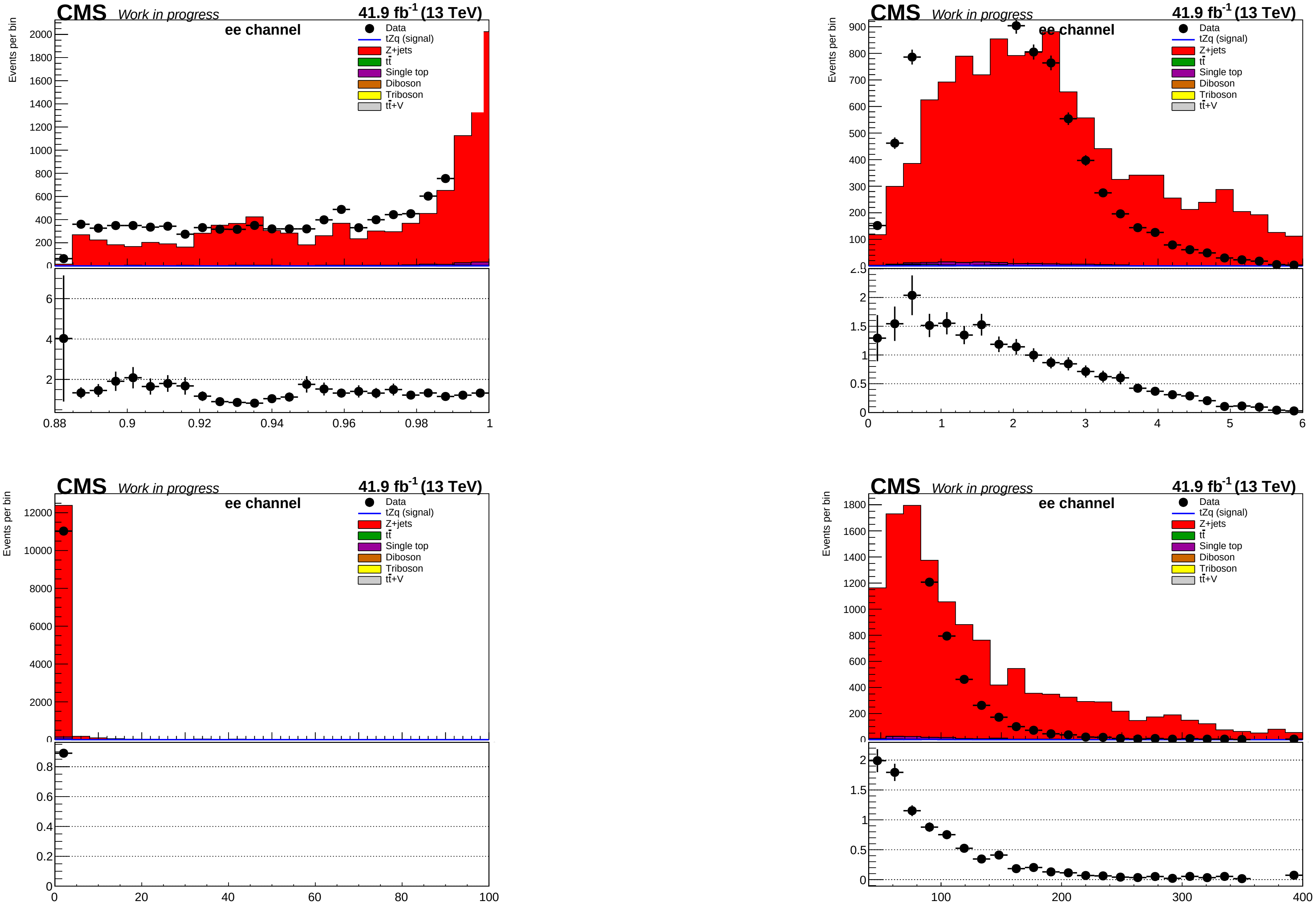


Figure 91: The distributions of the features: the b-tagged discriminator of the leading b-tagged jet, the χ^2 variable used for experimental blinding, ΔR between the leading and subleading jet candidates and the transverse momentum of the subleading jet candidate. The features in this figure correspond to the ee channel in the $Z + jets$ control region, for simulation and data recorded in 2017.

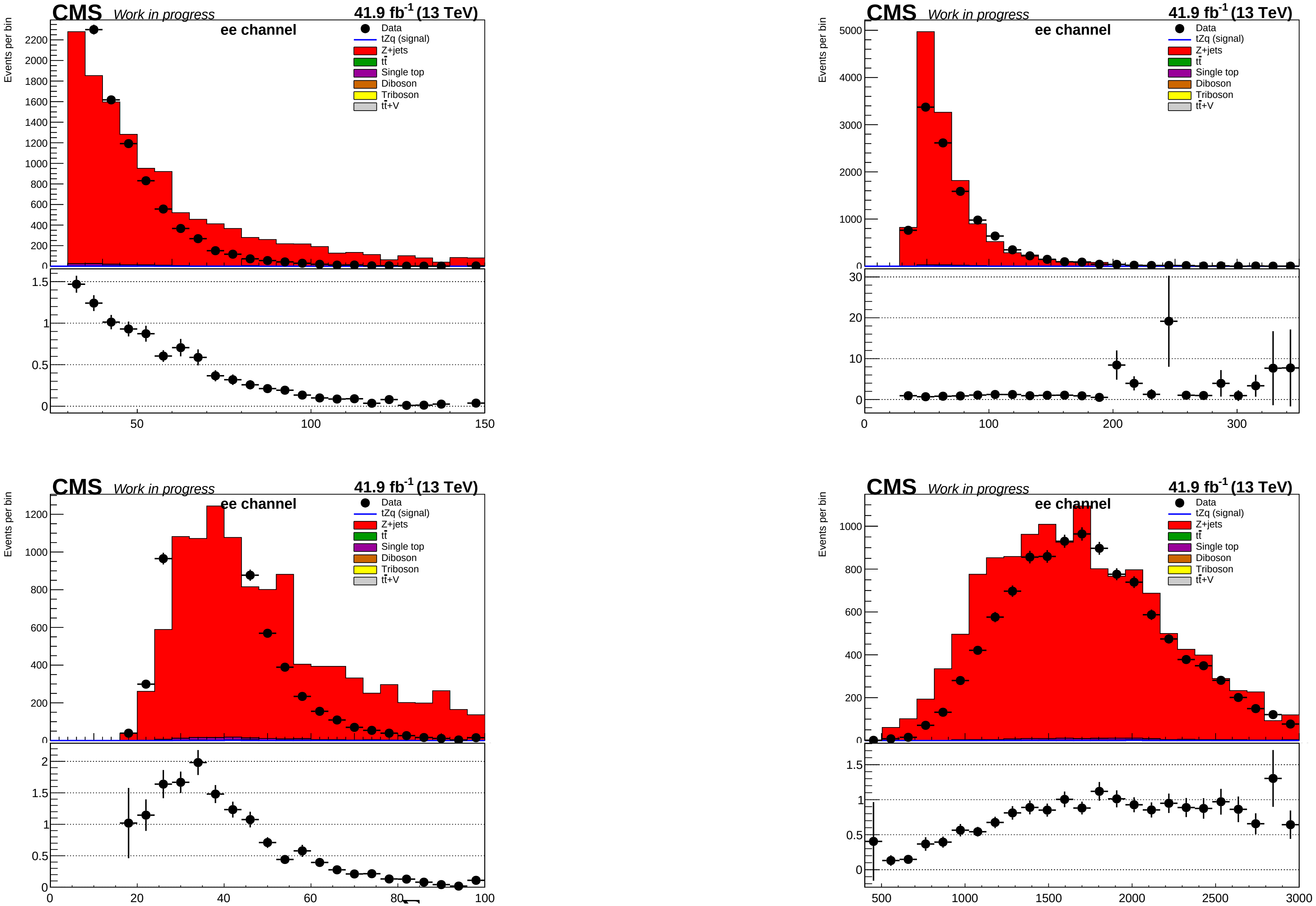


Figure 92: The distributions of the features: the transverse momentum of the fourth jet candidate, the sum of the masses of the four jet candidates, the transverse momentum of the leading lepton and the missing transverse energy. The features in this figure correspond to the ee channel in the $Z + jets$ control region, for simulation and data recorded in 2017.

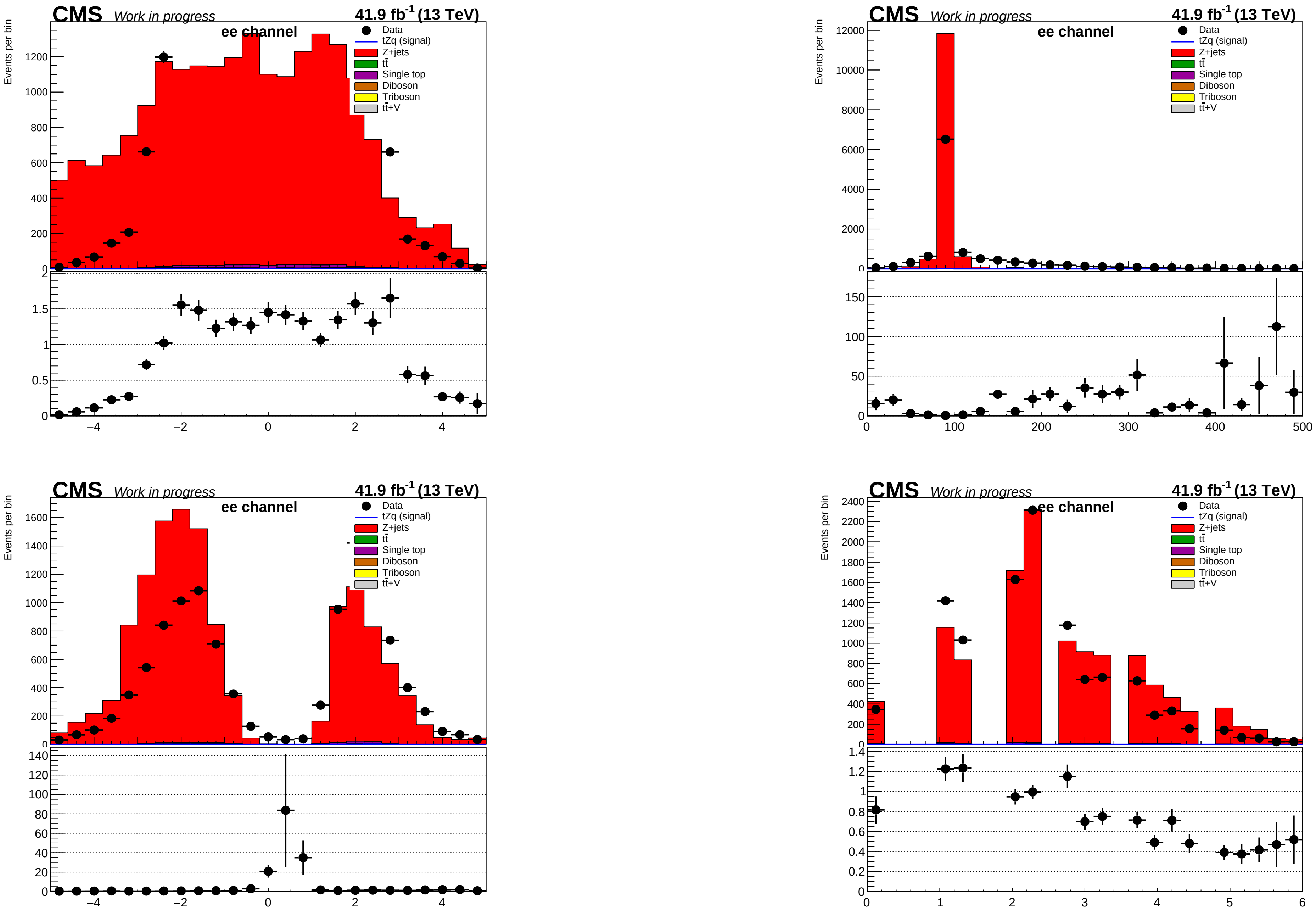


Figure 93: The distributions of the features: he pseudorapidity of the reconstructed W boson candidate, the pseudorapidity the reconstructed Z boson candidate, the invariant mass of the reconstructed Z boson candidate and the ΔR between the reconstructed Z boson and the leading b-jet candidates. The features in this figure correspond to the *ee* channel in the $Z + jets$ control region, for simulation and data recorded in 2017.

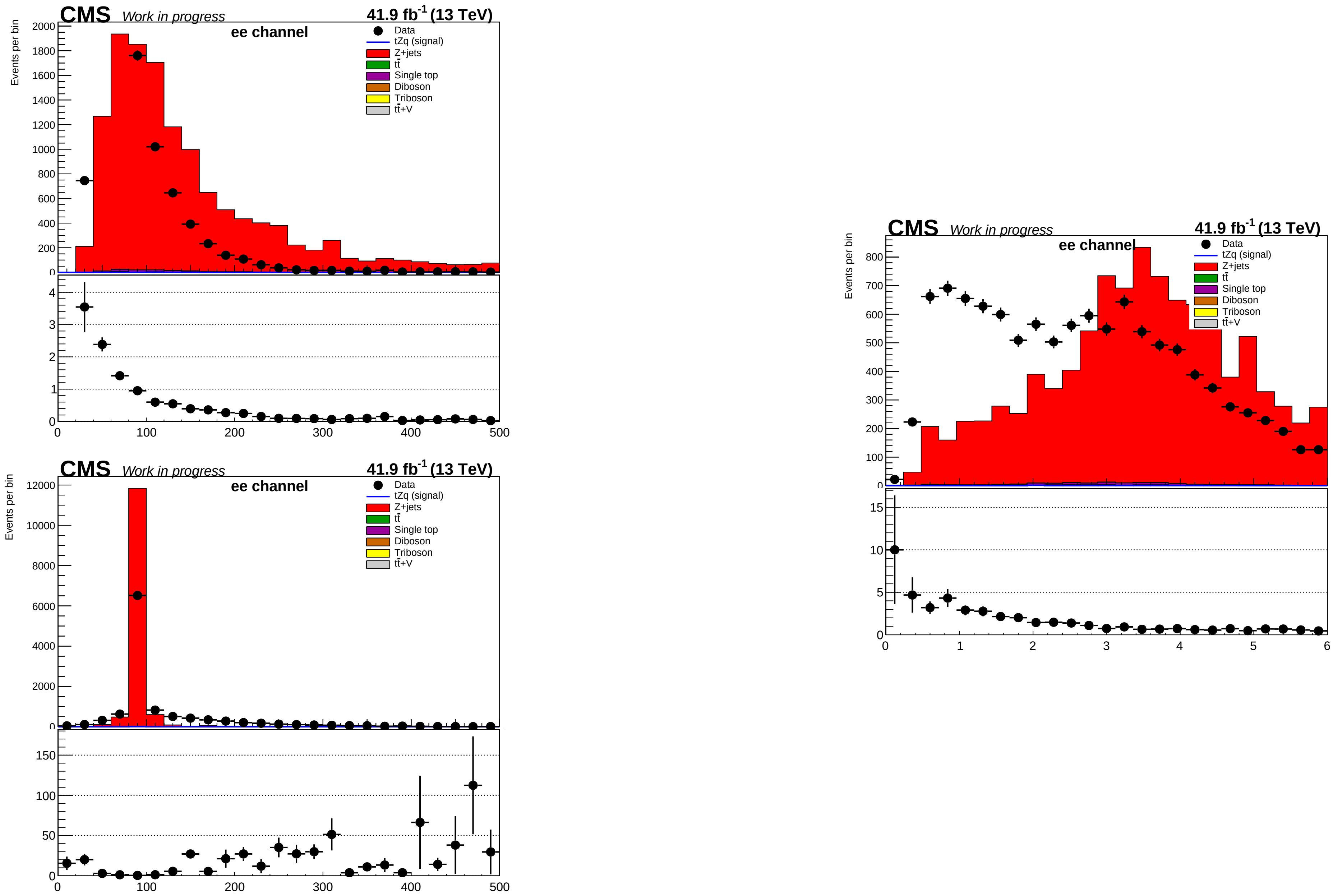


Figure 94: The distribution of the features: the transverse momentum of the leading jet from the reconstructed W boson candidate, the invariant mass of the reconstructed Z boson candidate and the maximum ΔR between the Z boson candidate and any jet candidate in the system. The features in this figure corresponds to the ee channel in the $Z + jets$ control region, for simulation and data recorded in 2017.

10.2.13 2017: $\mu\mu$ channel, $Z + jets$ control region

Figures 64- 66 show the distributions of selected features that were used as inputs for multivariate analysis. These correspond to 2016 data and simulation in the *ee* channel, and in the side-band region. The selected features were: b-tagged discriminator of the leading b-tagged jet, the χ^2 variable used for experimental blinding, the pseudorapidity of of the leading jet candidate, the transverse momentum of the subleading jet candidate, the transverse momentum of the fourth jet candidate, the sum of the masses of the four jet candidates, the missing transverse energy, the transverse momentum of the subleading jet from the reconstructed W boson candidate, the invariant mass of the reconstructed Z boson candidate, the transverse momentum of the reconstructed Z boson candidate, the ΔR between the reconstructed Z boson candidate and the leading bjet and the maximum ΔR between the Z boson candidate and any jet candidate in the system. These features were chosen as they were previously found to have the highest discriminating power [45] [46].

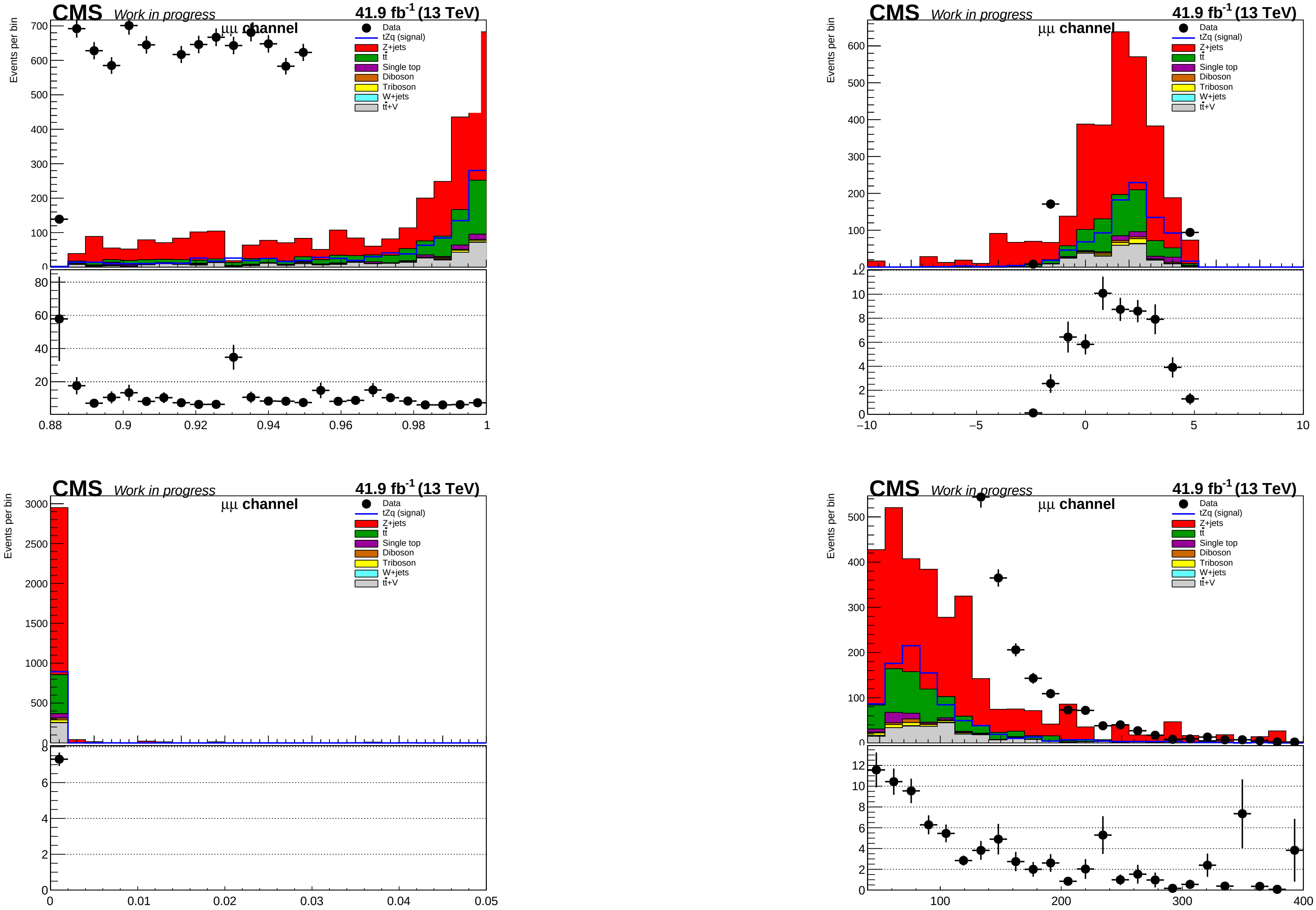


Figure 95: The distributions of the features: b-tagged discriminator of the leading b-tagged jet, the χ^2 variable used for experimental blinding, the pseudorapidity of of the leading jet candidate and the transverse momentum of the subleading jet candidate. The features in this figure correspond to the $\mu\mu$ channel in the $z + jets$ control region, for simulation and data recorded in 2017.

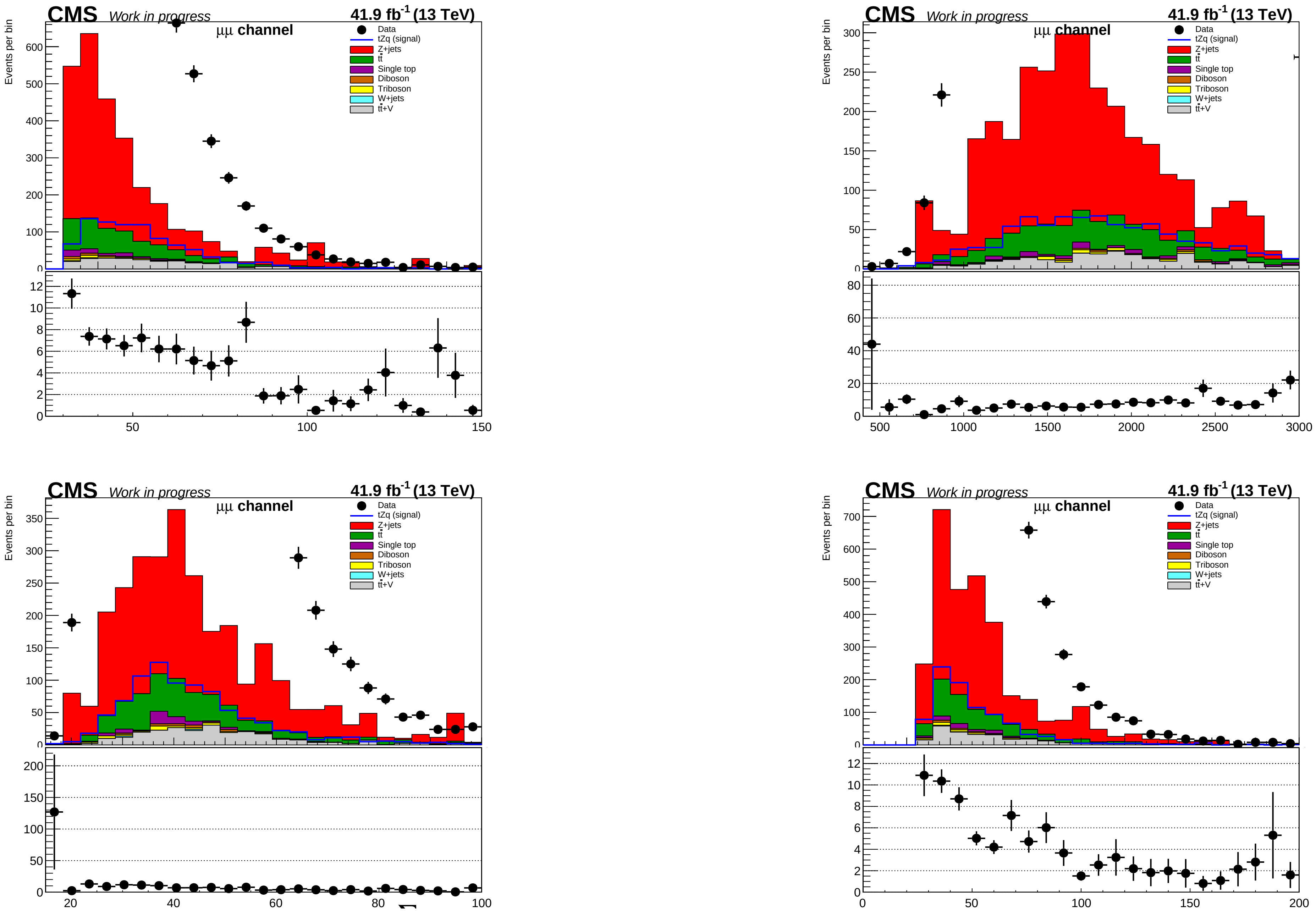


Figure 96: The distributions of the features: the transverse momentum of the fourth jet candidate, the sum of the masses of the four jet candidates, the missing transverse energy and the transverse momentum of the subleading jet from the reconstructed W boson candidate. The features in this figure correspond to the $\mu\mu$ channel in the $z + jets$ control region, for simulation and data recorded in 2017.

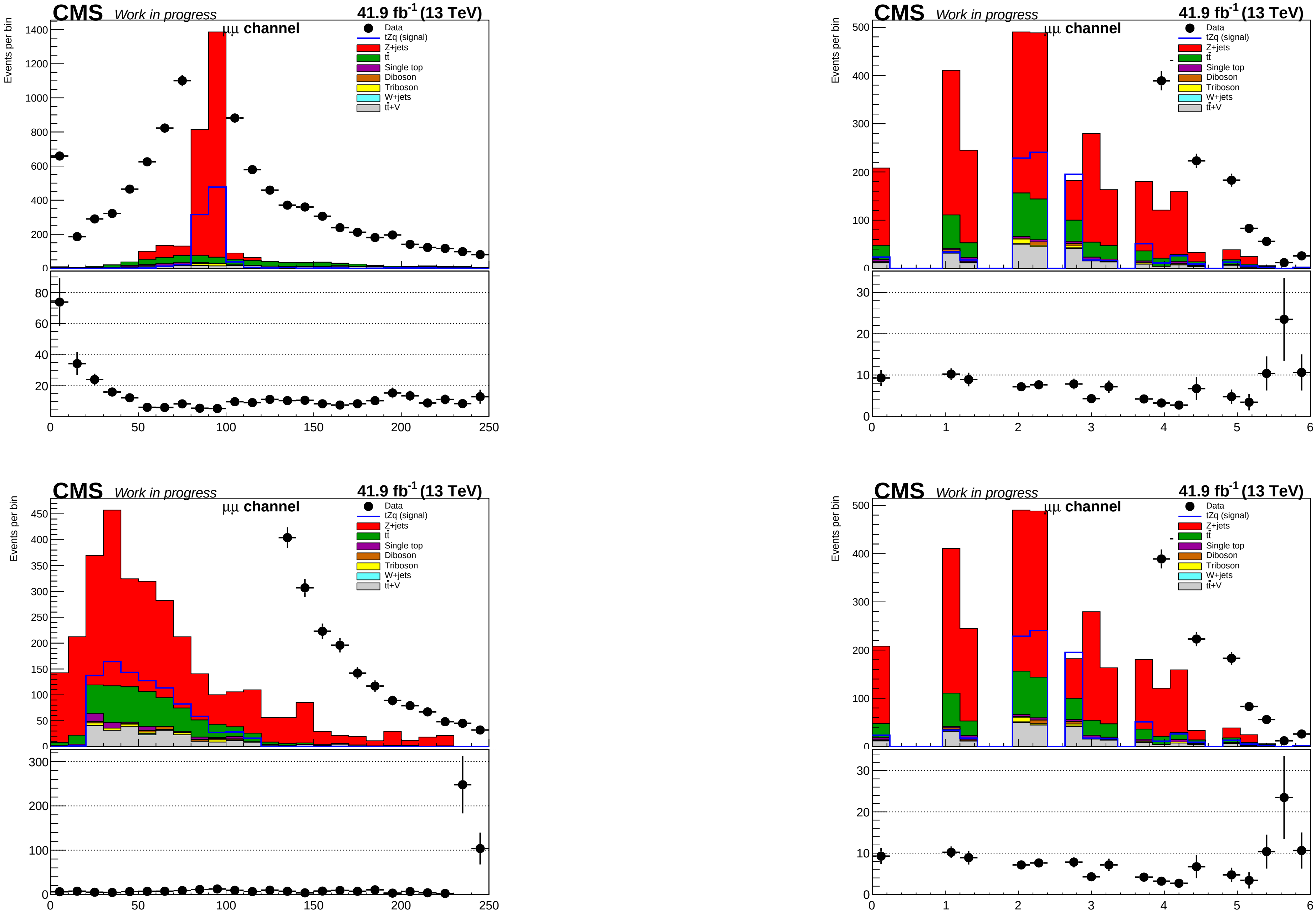


Figure 97: The distributions of the features: the invariant mass of the reconstructed Z boson candidate, the transverse momentum of the reconstructed Z boson candidate, the ΔR between the reconstructed Z boson candidate and the leading bjet and the maximum ΔR between the Z boson candidate and any jet candidate in the system. The features in this figure correspond to the $\mu\mu$ channel in the $z + jets$ control region, for simulation and data recorded in 2017.

10.2.14 2018: ee channel, $Z + jets$ control region

Figures 98- 101 show the distributions of selected features that were used as inputs for multivariate analysis. These correspond to 2018 data and simulation in the ee channel, and in the $z + jets$ control region. The selected features were: b-tagged discriminator of the leading b-tagged jet, the χ^2 variable used for experimental blinding, ΔR between the leading and subleading jet candidates, the transverse momentum of the subleading jet candidate, the transverse momentum of the fourth jet candidate, the sum of the masses of the four jet candidates, the transverse momentum of the leading lepton, the missing transverse energy, the transverse momentum of the leading jet from the reconstructed W boson candidate, the invariant mass of the reconstructed Z boson candidate and the maximum ΔR between the Z boson candidate and any jet candidate in the system. Due to time constraints an additional study to identify the features with the highest discriminating power was not undertaken, so the same features that were chosen for the 2016 and 2017 analyses were used for 2018.

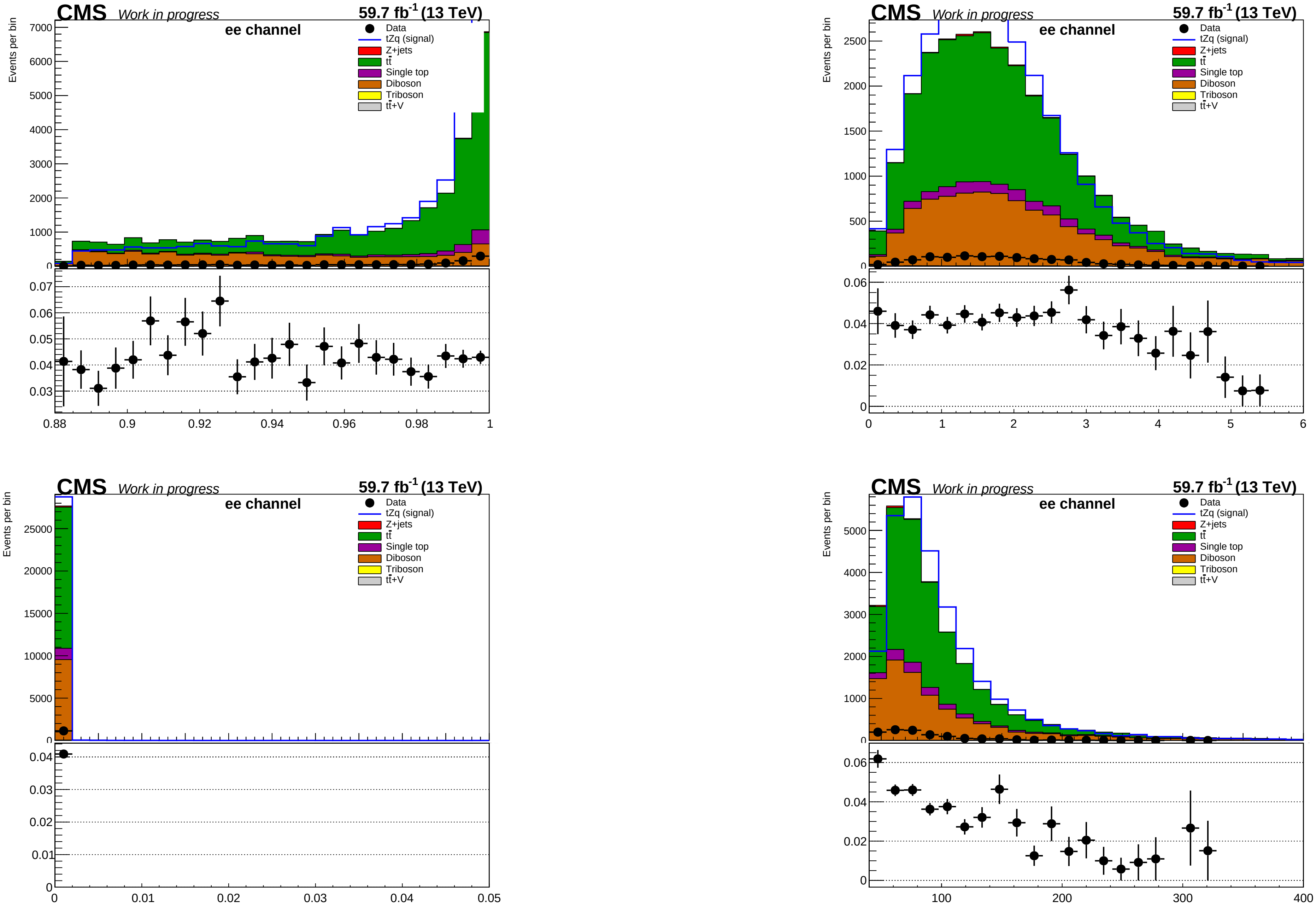


Figure 98: The distributions of the features: the b-tagged discriminator of the leading b-tagged jet, the χ^2 variable used for experimental blinding, ΔR between the leading and subleading jet candidates and the transverse momentum of the subleading jet candidate. The features in this figure correspond to the ee channel in the $z + jets$ control region, for simulation and data recorded in 2018.

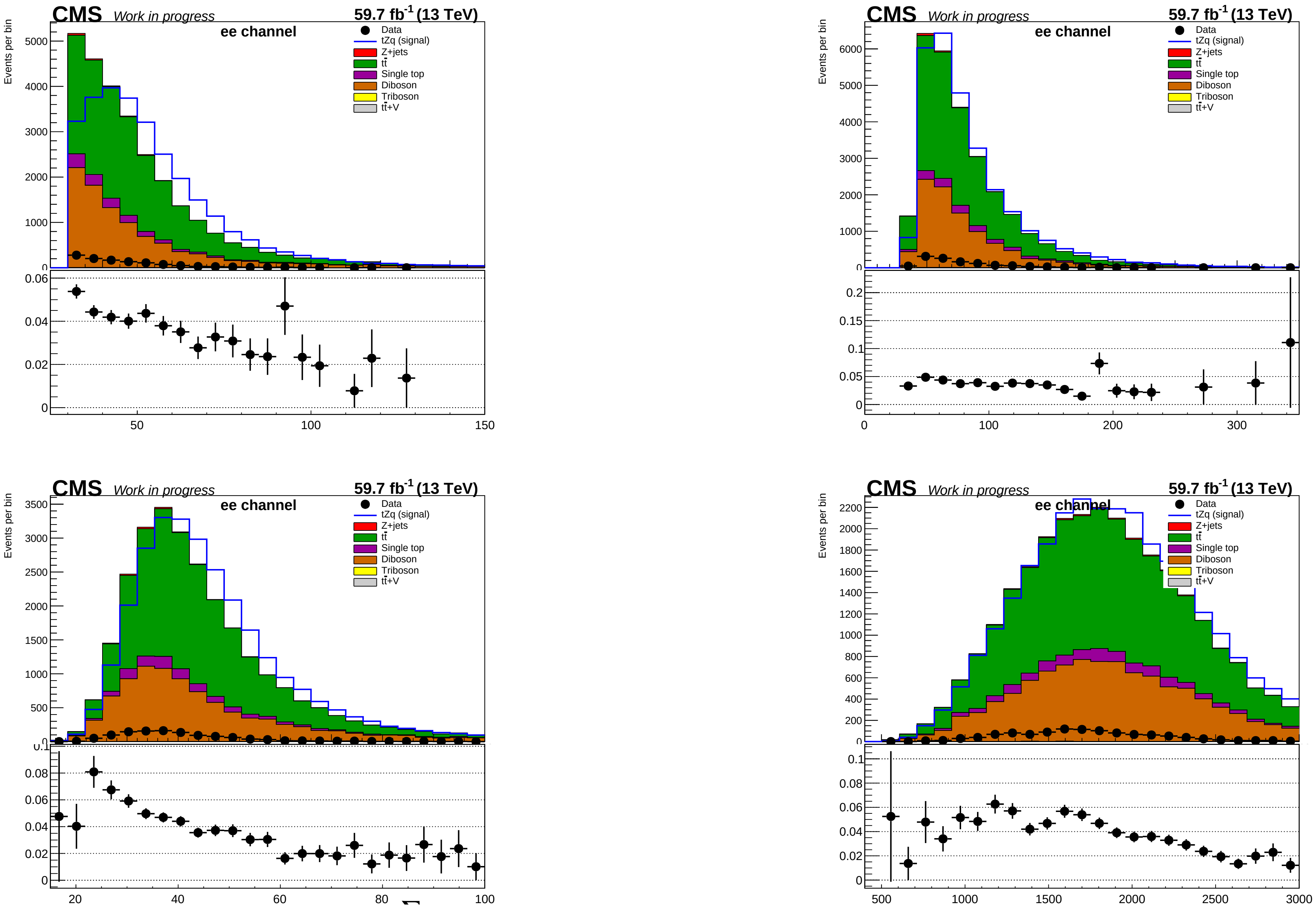


Figure 99: The distributions of the features: the transverse momentum of the fourth jet candidate, the sum of the masses of the four jet candidates, the transverse momentum of the leading lepton and the missing transverse energy. The features in this figure correspond to the ee channel in the $z + jets$ control region, for simulation and data recorded in 2018.

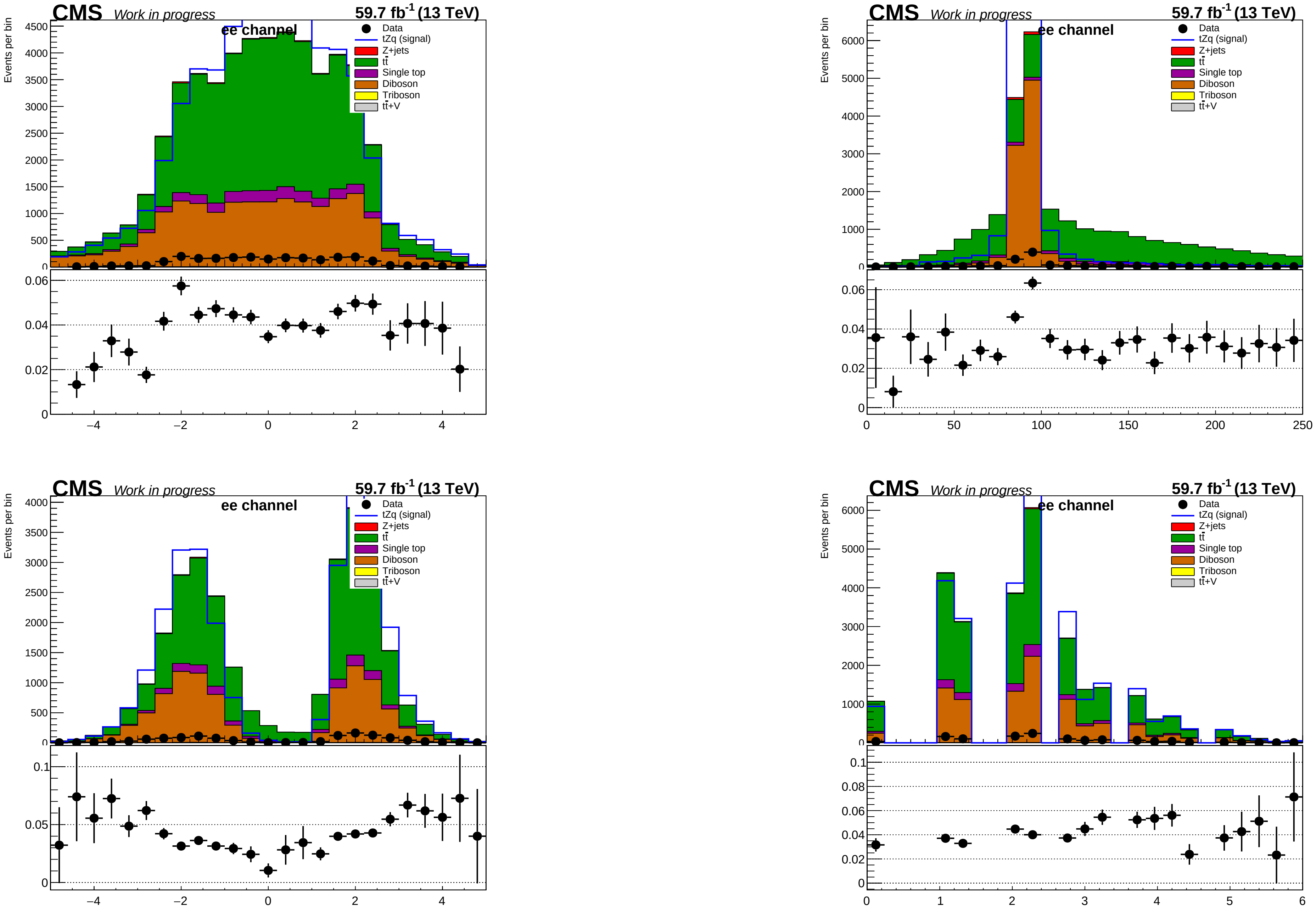


Figure 100: The distributions of the features: the pseudorapidity of the reconstructed W boson candidate, the pseudorapidity the reconstructed Z boson candidate, the invariant mass of the reconstructed Z boson candidate and the ΔR between the reconstructed Z boson and the leading b-jet candidates. The features in this figure correspond to the *ee* channel in the $z + jets$ control region, for simulation and data recorded in 2018.

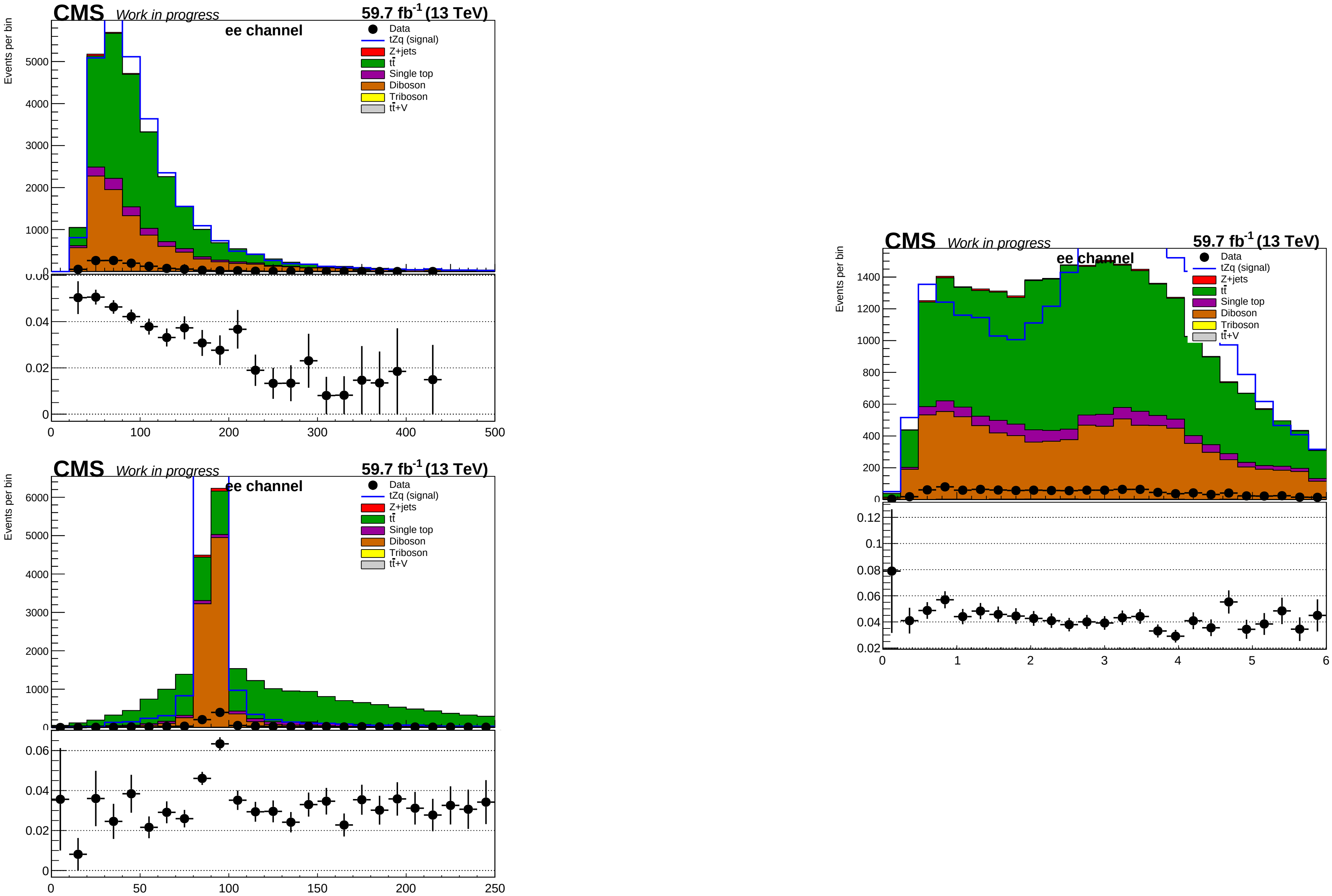


Figure 101: The distribution of the features: the transverse momentum of the leading jet from the reconstructed W boson candidate, the invariant mass of the reconstructed Z boson candidate and the maximum ΔR between the Z boson candidate and any jet candidate in the system. The features in this figure corresponds to the ee channel in the $z + jets$ region, for simulation and data recorded in 2018.

10.2.15 2018: $\mu\mu$ channel, $z + jets$ control region

Figures 102- 104 show the distributions of selected features that were used as inputs for multivariate analysis. These correspond to 2018 data and simulation in the ee channel, and in the $z + jets$ control region. The selected features were: b-tagged discriminator of the leading b-tagged jet, the χ^2 variable used for experimental blinding, the pseudorapidity of of the leading jet candidate, the transverse momentum of the subleading jet candidate, the transverse momentum of the fourth jet candidate, the sum of the masses of the four jet candidates, the missing transverse energy, the transverse momentum of the subleading jet from the reconstructed W boson candidate, the invariant mass of the reconstructed Z boson candidate, the transverse momentum of the reconstructed Z boson candidate, the ΔR between the reconstructed Z boson candidate and the leading bjet and the maximum ΔR between the Z boson candidate and any jet candidate in the system. Due to time constraints an additional study to identify the features with the highest discriminating power was not undertaken, so the same features that were chosen for the 2016 and 2017 analyses were used for 2018.

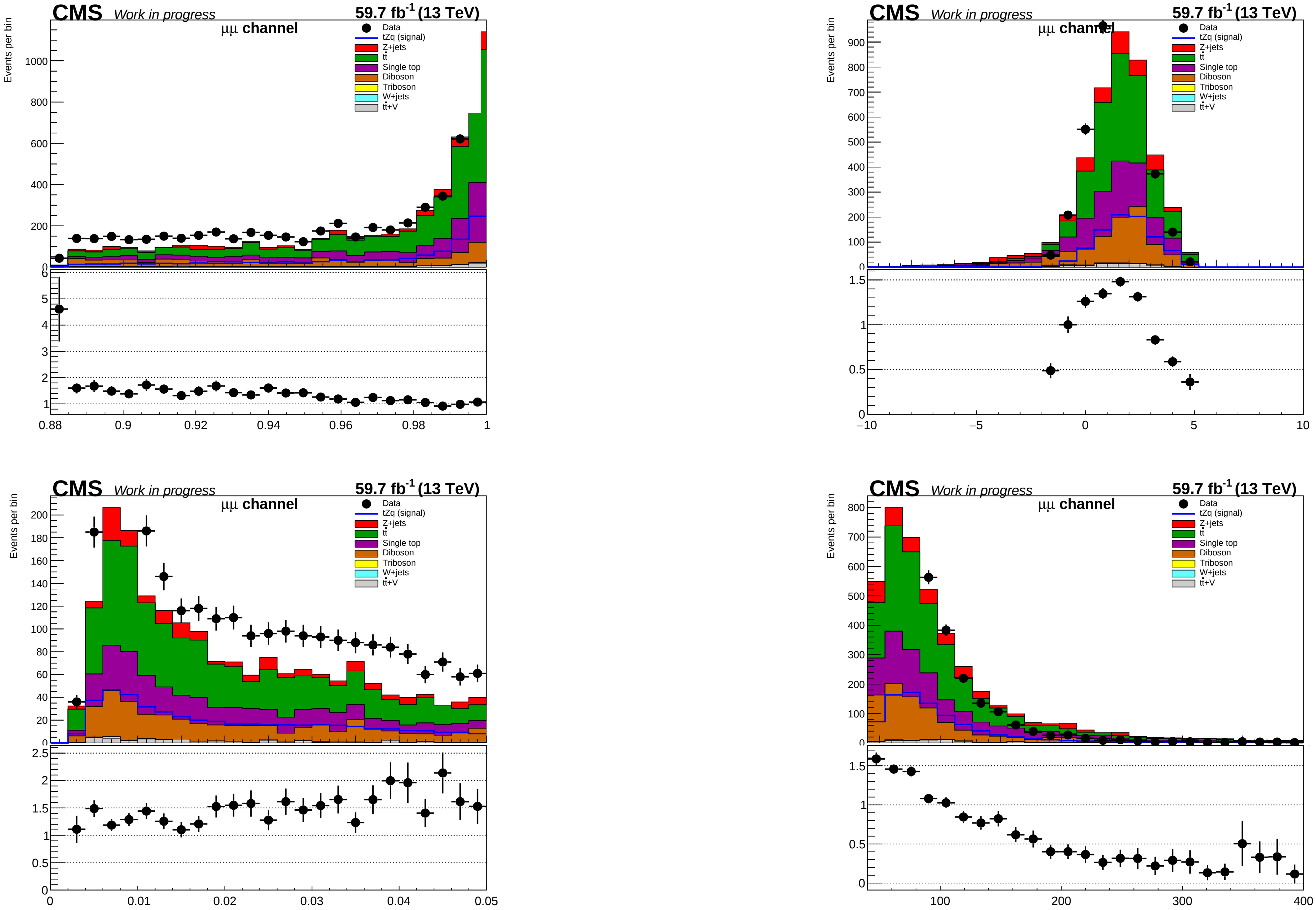


Figure 102: The distributions of the features: b-tagged discriminator of the leading b-tagged jet, the χ^2 variable used for experimental blinding, the pseudorapidity of of the leading jet candidate and the transverse momentum of the subleading jet candidate. The features in this figure correspond to the $\mu\mu$ channel in the $z + jets$ control region, for simulation and data recorded in 2018.

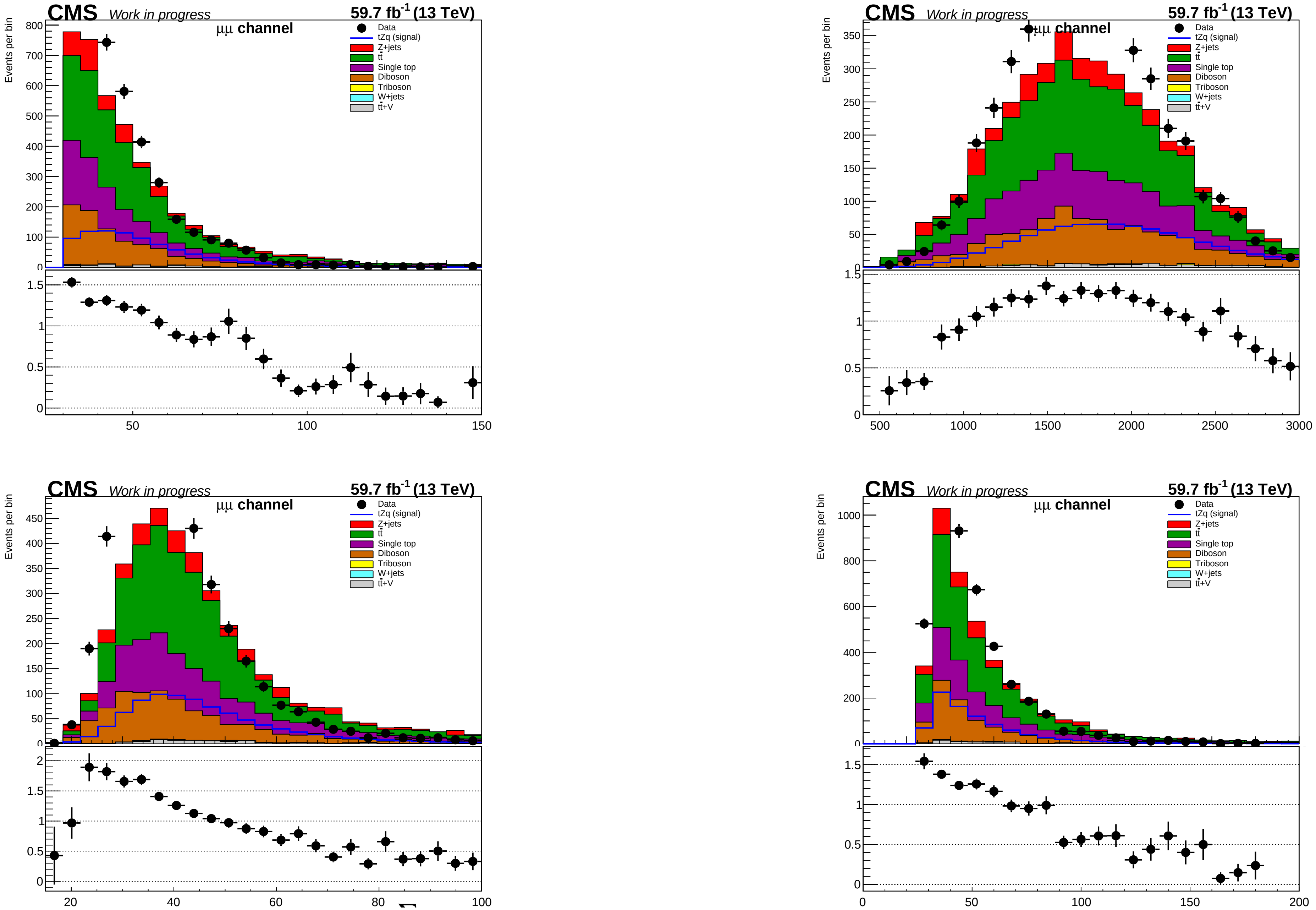


Figure 103: The distributions of the features: the transverse momentum of the fourth jet candidate, the sum of the masses of the four jet candidates, the missing transverse energy and the transverse momentum of the subleading jet from the reconstructed W boson candidate. The features in this figure correspond to the $\mu\mu$ channel in the $z + jets$ control region, for simulation and data recorded in 2018.

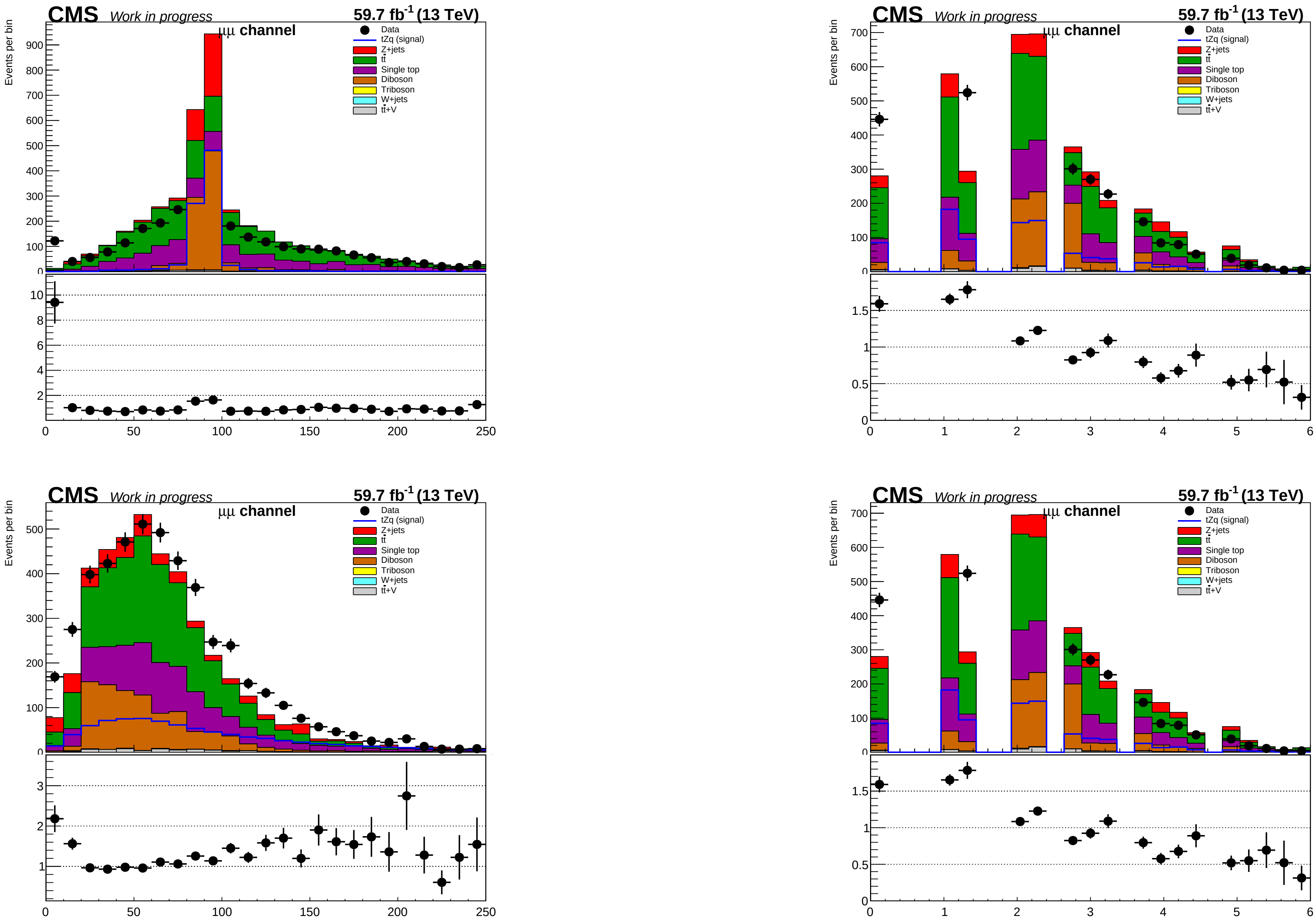
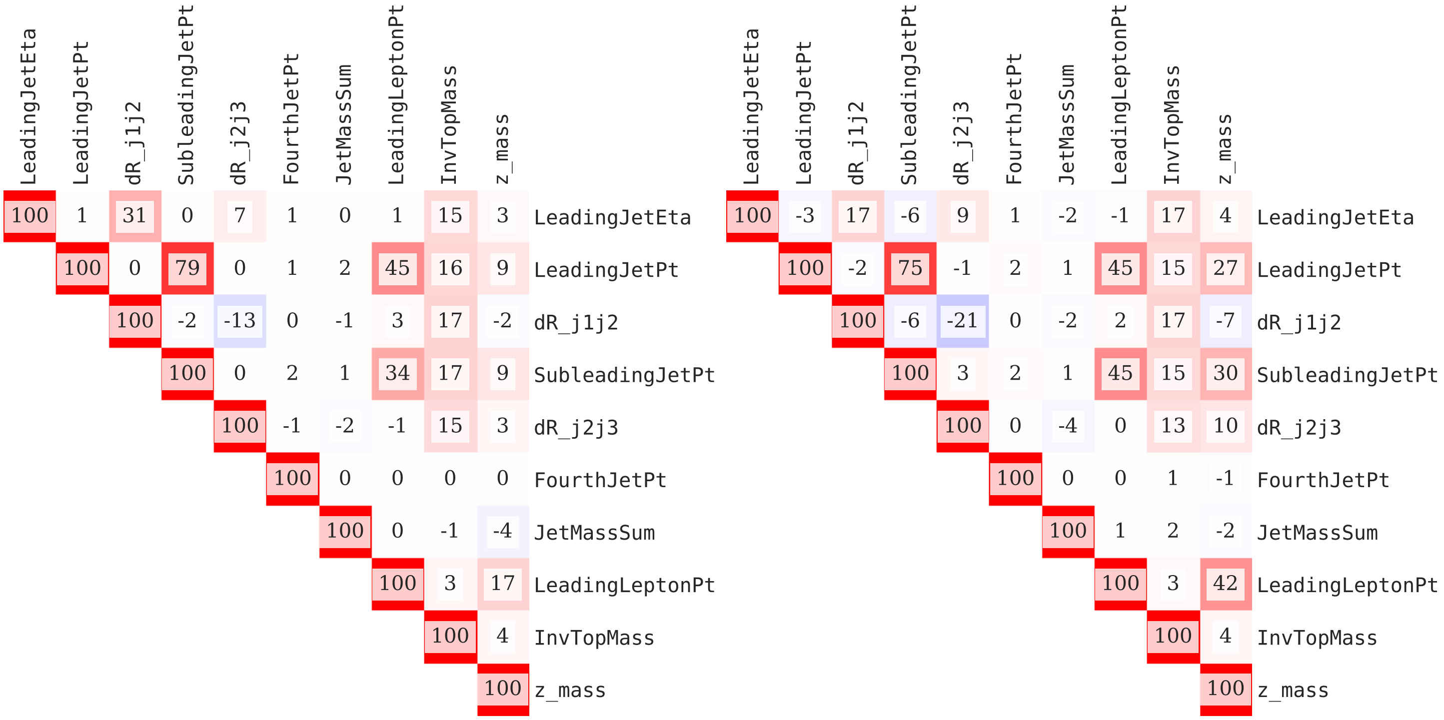


Figure 104: The distributions of the features: the invariant mass of the reconstructed Z boson candidate, the transverse momentum of the reconstructed Z boson candidate, the ΔR between the reconstructed Z boson candidate and the leading bjet and the maximum ΔR between the Z boson candidate and any jet candidate in the system. The features in this figure correspond to the $\mu\mu$ channel in the $z + jets$ control region, for simulation and data recorded in 2018.

10.3 Results: Multivariate Analysis

10.3.1 Correlations between the BDT Input Variables

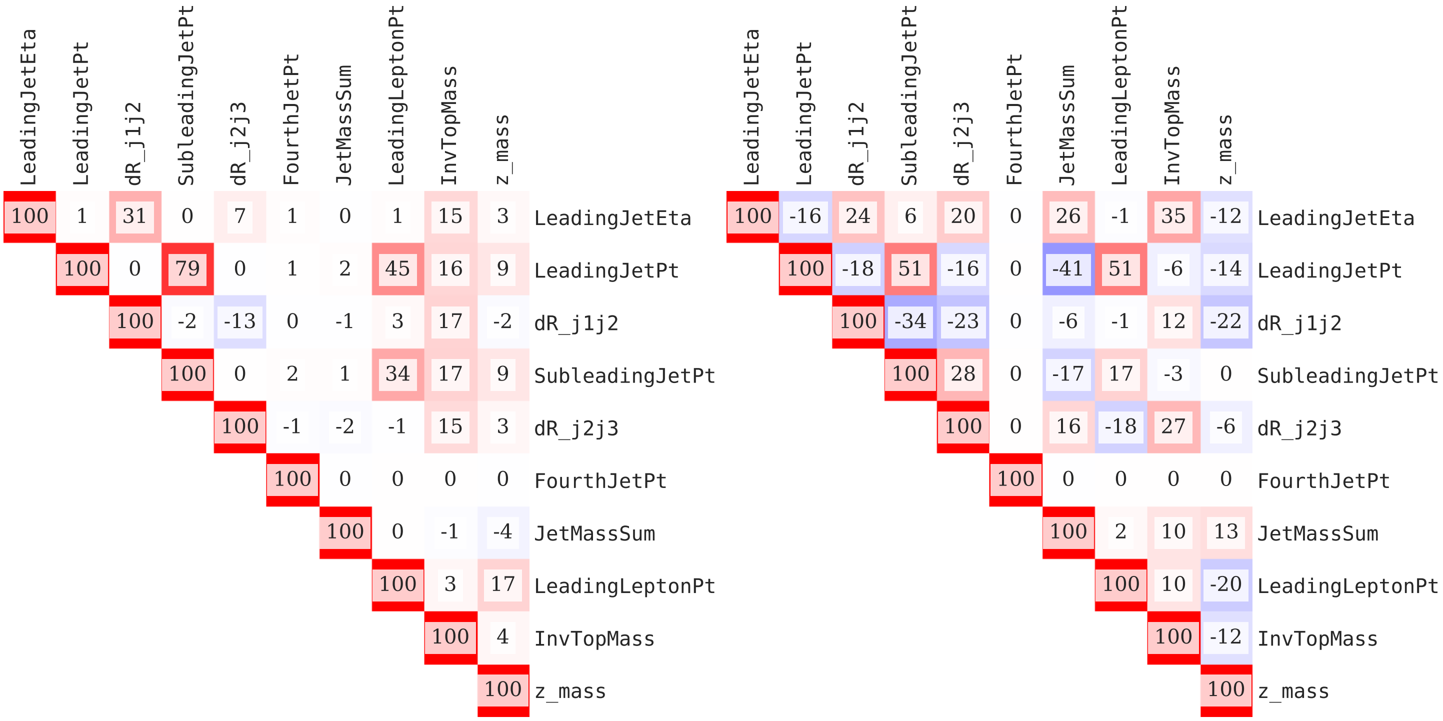
Figures 105-110 show the correlations between the BDT input variables for the signal and background samples, where values closer to 100 indicate a strong correlation. In the Figures, *LeadingJetEta* represents the pseudorapidity of the leading jet, *LeadingJetPt* represents the transverse momentum of the leading jet, *dR_j1j2* represents the ΔR between the leading and subleading final state jets, *SubleadingJetPt* represents the transverse momentum of the subleading jet, *dR_j2j3* represents the ΔR between the third and fourth final state jets, *FourthJetPt* represents the transverse momentum of the fourth most leading jet of the system, *JetMassSum* represents the sum of all jets in the system, *LeadingLeptonPt* represents the transverse momentum of the leading lepton, *InvTopMass* represents the invariant mass of the reconstructed top quark candidate and *z_mass* represents the invariant mass of the Z boson candidate. Results are discussed in Section 11.



	LeadingJetEta	LeadingJetPt	dR_j1j2	SubleadingJetPt	dR_j2j3	FourthJetPt	JetMassSum	LeadingLeptonPt	InvTopMass	z_mass
LeadingJetEta	100	1	31	0	7	1	0	1	15	3
LeadingJetPt		100	0	79	0	1	2	45	16	9
dR_j1j2			100	-2	-13	0	-1	3	17	-2
SubleadingJetPt				100	0	2	1	34	17	9
dR_j2j3					100	-1	-2	-1	15	3
FourthJetPt						100	0	0	0	0
JetMassSum							100	0	-1	-4
LeadingLeptonPt								100	3	17
InvTopMass									100	4
z_mass										100

	LeadingJetEta	LeadingJetPt	dR_j1j2	SubleadingJetPt	dR_j2j3	FourthJetPt	JetMassSum	LeadingLeptonPt	InvTopMass	z_mass
LeadingJetEta	100	-3	17	-6	9	1	-2	-1	17	4
LeadingJetPt		100	-2	75	-1	2	1	45	15	27
dR_j1j2			100	-6	-21	0	-2	2	17	-7
SubleadingJetPt				100	3	2	1	45	15	30
dR_j2j3					100	0	-4	0	13	10
FourthJetPt						100	0	0	1	-1
JetMassSum							100	1	2	-2
LeadingLeptonPt								100	3	42
InvTopMass									100	4
z_mass										100

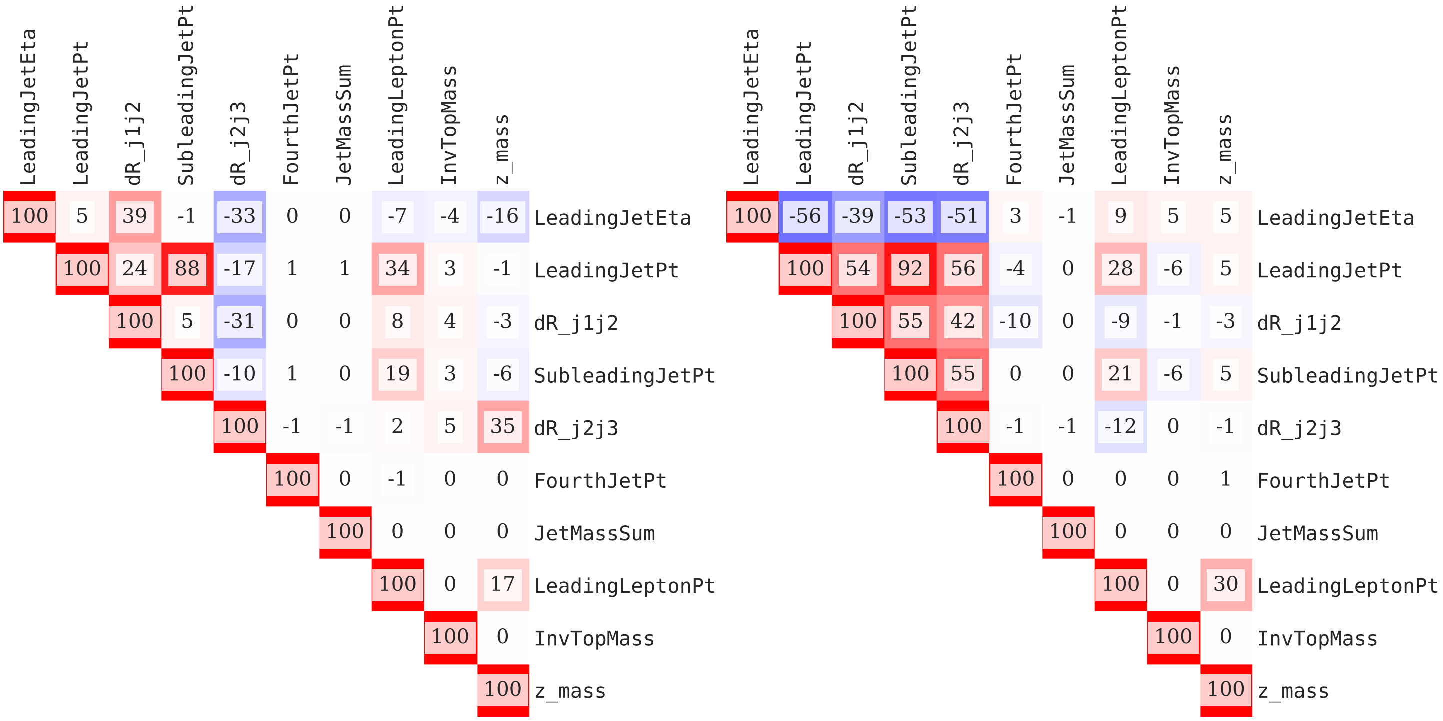
Figure 105: The correlations between BDT input variables for the signal and backgrounds samples of 2016, for the *ee* channel.



	LeadingJetEta	LeadingJetPt	dR_j1j2	SubleadingJetPt	dR_j2j3	FourthJetPt	JetMassSum	LeadingLeptonPt	InvTopMass	z_mass
LeadingJetEta	100	1	31	0	7	1	0	1	15	3
LeadingJetPt		100	0	79	0	1	2	45	16	9
dR_j1j2			100	-2	-13	0	-1	3	17	-2
SubleadingJetPt				100	0	2	1	34	17	9
dR_j2j3					100	-1	-2	-1	15	3
FourthJetPt						100	0	0	0	0
JetMassSum							100	0	-1	-4
LeadingLeptonPt								100	3	17
InvTopMass									100	4
z_mass										100

	LeadingJetEta	LeadingJetPt	dR_j1j2	SubleadingJetPt	dR_j2j3	FourthJetPt	JetMassSum	LeadingLeptonPt	InvTopMass	z_mass
LeadingJetEta	100	-16	24	6	20	0	26	-1	35	-12
LeadingJetPt		100	-18	51	-16	0	-41	51	-6	-14
dR_j1j2			100	-34	-23	0	-6	-1	12	-22
SubleadingJetPt				100	28	0	-17	17	-3	0
dR_j2j3					100	0	16	-18	27	-6
FourthJetPt						100	0	0	0	0
JetMassSum							100	2	10	13
LeadingLeptonPt								100	10	-20
InvTopMass									100	-12
z_mass										100

Figure 106: The correlations between BDT input variables for the signal (left) and background samples (right) of 2016, for the $\mu\mu$ channel.



	LeadingJetEta	LeadingJetPt	dR_j1j2	SubleadingJetPt	dR_j2j3	FourthJetPt	JetMassSum	LeadingLeptonPt	InvTopMass	z_mass
LeadingJetEta	100	5	39	-1	-33	0	0	-7	-4	-16
LeadingJetPt		100	24	88	-17	1	1	34	3	-1
dR_j1j2			100	5	-31	0	0	8	4	-3
SubleadingJetPt				100	-10	1	0	19	3	-6
dR_j2j3					100	-1	-1	2	5	35
FourthJetPt						100	0	-1	0	0
JetMassSum							100	0	0	0
LeadingLeptonPt								100	0	17
InvTopMass									100	0
z_mass										100

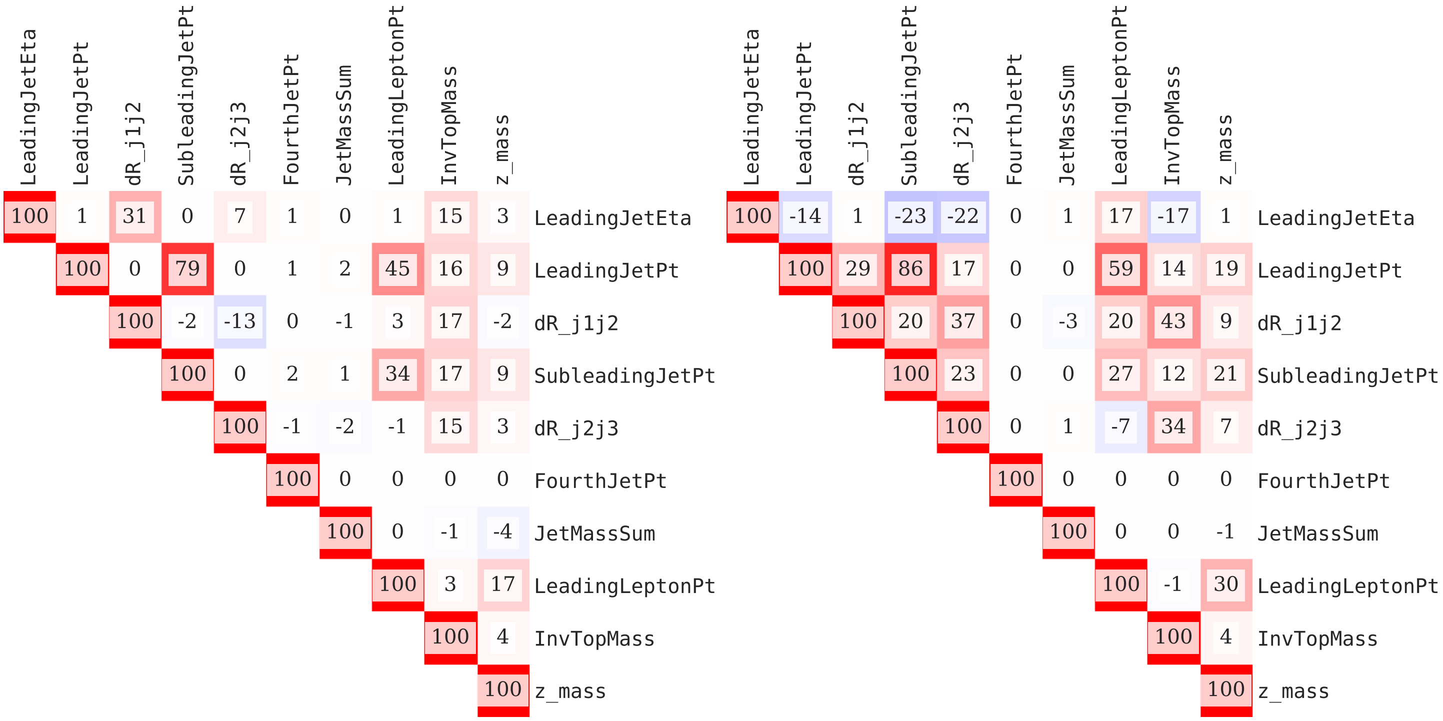
	LeadingJetEta	LeadingJetPt	dR_j1j2	SubleadingJetPt	dR_j2j3	FourthJetPt	JetMassSum	LeadingLeptonPt	InvTopMass	z_mass
LeadingJetEta	100	-56	-39	-53	-51	3	-1	9	5	5
LeadingJetPt		100	54	92	56	-4	0	28	-6	5
dR_j1j2			100	55	42	-10	0	-9	-1	-3
SubleadingJetPt				100	55	0	0	21	-6	5
dR_j2j3					100	-1	-1	-12	0	-1
FourthJetPt						100	0	0	0	1
JetMassSum							100	0	0	0
LeadingLeptonPt								100	0	30
InvTopMass									100	0
z_mass										100

Figure 107: The correlations between BDT input variables for the signal (left) and background (right) samples of 2017, for the ee channel.

	LeadingJetEta	LeadingJetPt	dR_j1j2	SubleadingJetPt	dR_j2j3	FourthJetPt	JetMassSum	LeadingLeptonPt	InvTopMass	z_mass
LeadingJetEta	100	-56	-39	-53	-51	3	-1	9	5	5
LeadingJetPt		100	54	92	56	-4	0	28	-6	5
dR_j1j2			100	55	42	-10	0	-9	-1	-3
SubleadingJetPt				100	55	0	0	21	-6	5
dR_j2j3					100	-1	-1	-12	0	-1
FourthJetPt						100	0	0	0	1
JetMassSum							100	0	0	0
LeadingLeptonPt								100	0	30
InvTopMass									100	0
z_mass										100

	LeadingJetEta	LeadingJetPt	dR_j1j2	SubleadingJetPt	dR_j2j3	FourthJetPt	JetMassSum	LeadingLeptonPt	InvTopMass	z_mass
LeadingJetEta	100	-55	-38	-56	-43	-1	0	0	1	0
LeadingJetPt		100	62	93	57	3	1	26	3	8
dR_j1j2			100	62	43	1	2	-6	12	2
SubleadingJetPt				100	60	3	1	17	3	8
dR_j2j3					100	2	0	-8	3	3
FourthJetPt						100	0	-3	0	0
JetMassSum							100	-1	0	0
LeadingLeptonPt								100	-3	24
InvTopMass									100	-5
z_mass										100

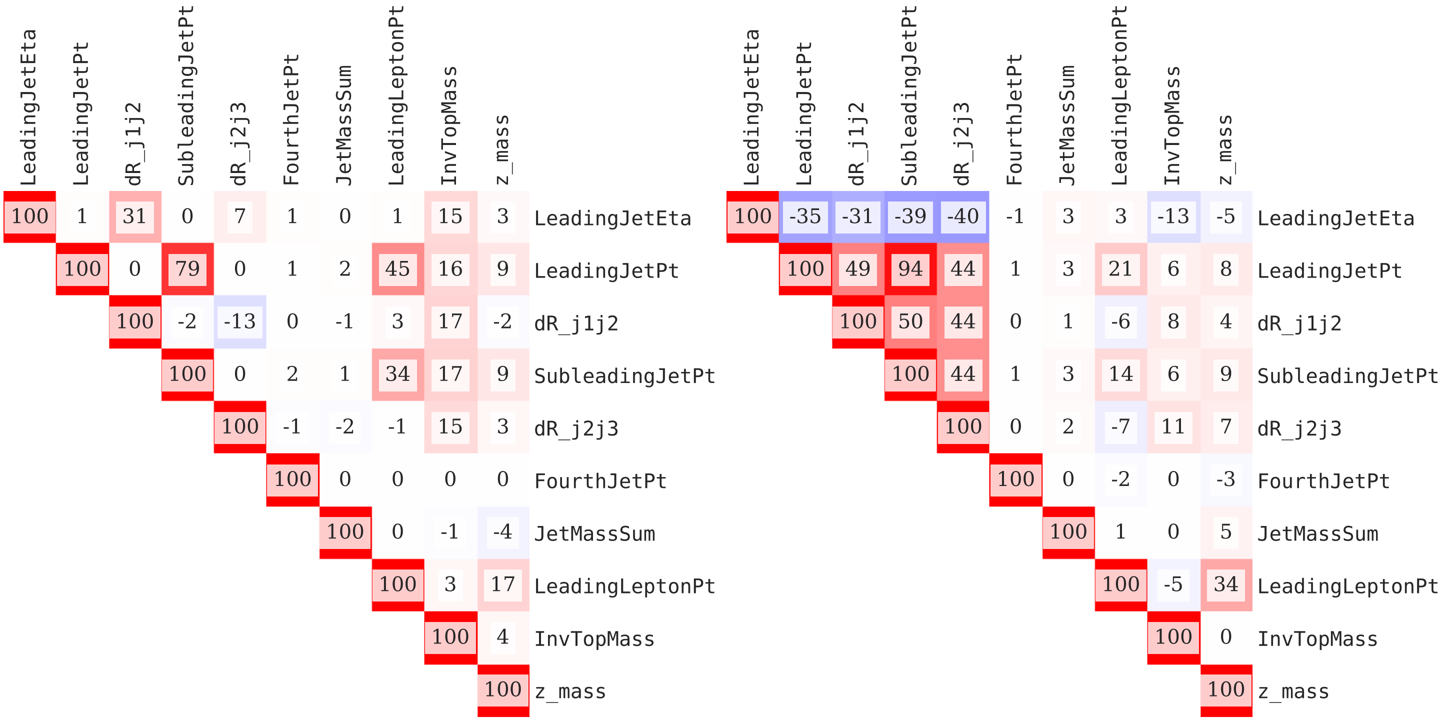
Figure 108: The correlations between BDT input variables for the signal (left) and background (right) samples of 2017, for the $\mu\mu$ channel.



LeadingJetEta	LeadingJetPt	dR_j1j2	SubleadingJetPt	dR_j2j3	FourthJetPt	JetMassSum	LeadingLeptonPt	InvTopMass	z_mass	
100	1	31	0	7	1	0	1	15	3	LeadingJetEta
	100	0	79	0	1	2	45	16	9	LeadingJetPt
		100	-2	-13	0	-1	3	17	-2	dR_j1j2
			100	0	2	1	34	17	9	SubleadingJetPt
				100	-1	-2	-1	15	3	dR_j2j3
					100	0	0	0	0	FourthJetPt
						100	0	-1	-4	JetMassSum
							100	3	17	LeadingLeptonPt
								100	4	InvTopMass
									100	z_mass

LeadingJetEta	LeadingJetPt	dR_j1j2	SubleadingJetPt	dR_j2j3	FourthJetPt	JetMassSum	LeadingLeptonPt	InvTopMass	z_mass	
100	-14	1	-23	-22	0	1	17	-17	1	LeadingJetEta
	100	29	86	17	0	0	59	14	19	LeadingJetPt
		100	20	37	0	-3	20	43	9	dR_j1j2
			100	23	0	0	27	12	21	SubleadingJetPt
				100	0	1	-7	34	7	dR_j2j3
					100	0	0	0	0	FourthJetPt
						100	0	0	-1	JetMassSum
							100	-1	30	LeadingLeptonPt
								100	4	InvTopMass
									100	z_mass

Figure 109: The correlations between BDT input variables for the signal (left) and background (right) samples of 2018, for the ee channel.



LeadingJetEta	LeadingJetPt	dR_j1j2	SubleadingJetPt	dR_j2j3	FourthJetPt	JetMassSum	LeadingLeptonPt	InvTopMass	z_mass	
100	1	31	0	7	1	0	1	15	3	LeadingJetEta
	100	0	79	0	1	2	45	16	9	LeadingJetPt
		100	-2	-13	0	-1	3	17	-2	dR_j1j2
			100	0	2	1	34	17	9	SubleadingJetPt
				100	-1	-2	-1	15	3	dR_j2j3
					100	0	0	0	0	FourthJetPt
						100	0	-1	-4	JetMassSum
							100	3	17	LeadingLeptonPt
								100	4	InvTopMass
									100	z_mass

LeadingJetEta	LeadingJetPt	dR_j1j2	SubleadingJetPt	dR_j2j3	FourthJetPt	JetMassSum	LeadingLeptonPt	InvTopMass	z_mass	
100	-35	-31	-39	-40	-1	3	3	-13	-5	LeadingJetEta
	100	49	94	44	1	3	21	6	8	LeadingJetPt
		100	50	44	0	1	-6	8	4	dR_j1j2
			100	44	1	3	14	6	9	SubleadingJetPt
				100	0	2	-7	11	7	dR_j2j3
					100	0	-2	0	-3	FourthJetPt
						100	1	0	5	JetMassSum
							100	-5	34	LeadingLeptonPt
								100	0	InvTopMass
									100	z_mass

Figure 110: The correlations between BDT input variables for the signal (left) and background (right) samples of 2018, for the $\mu\mu$ channel.

10.3.2 BDT Classifier Response

Figures 111-113 shows the BDT classifier response for 2016, 2017 and 2018, for the ee and $\mu\mu$ channels of the side-band region. A response closer to 1.0 corresponds to a signal-like event, whereas a response closer to 0 corresponds to a background-like event. Any response in the middle reflects an uncertainty from the model. Results shown are for the test and training sets and are discussed in Section 11.

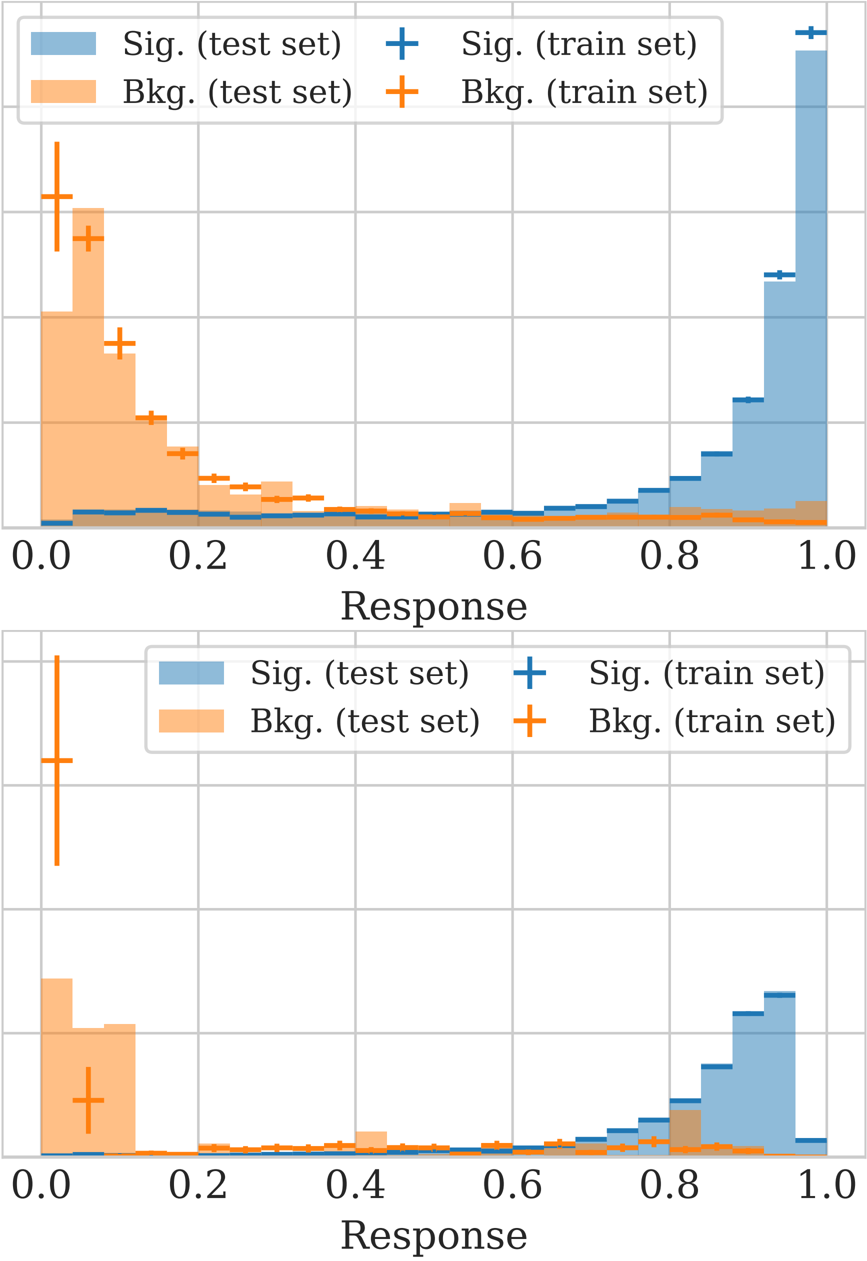


Figure 111: The BDT classifier response for the signal and background samples of 2016, for the ee (top) and $\mu\mu$ channels (bottom).

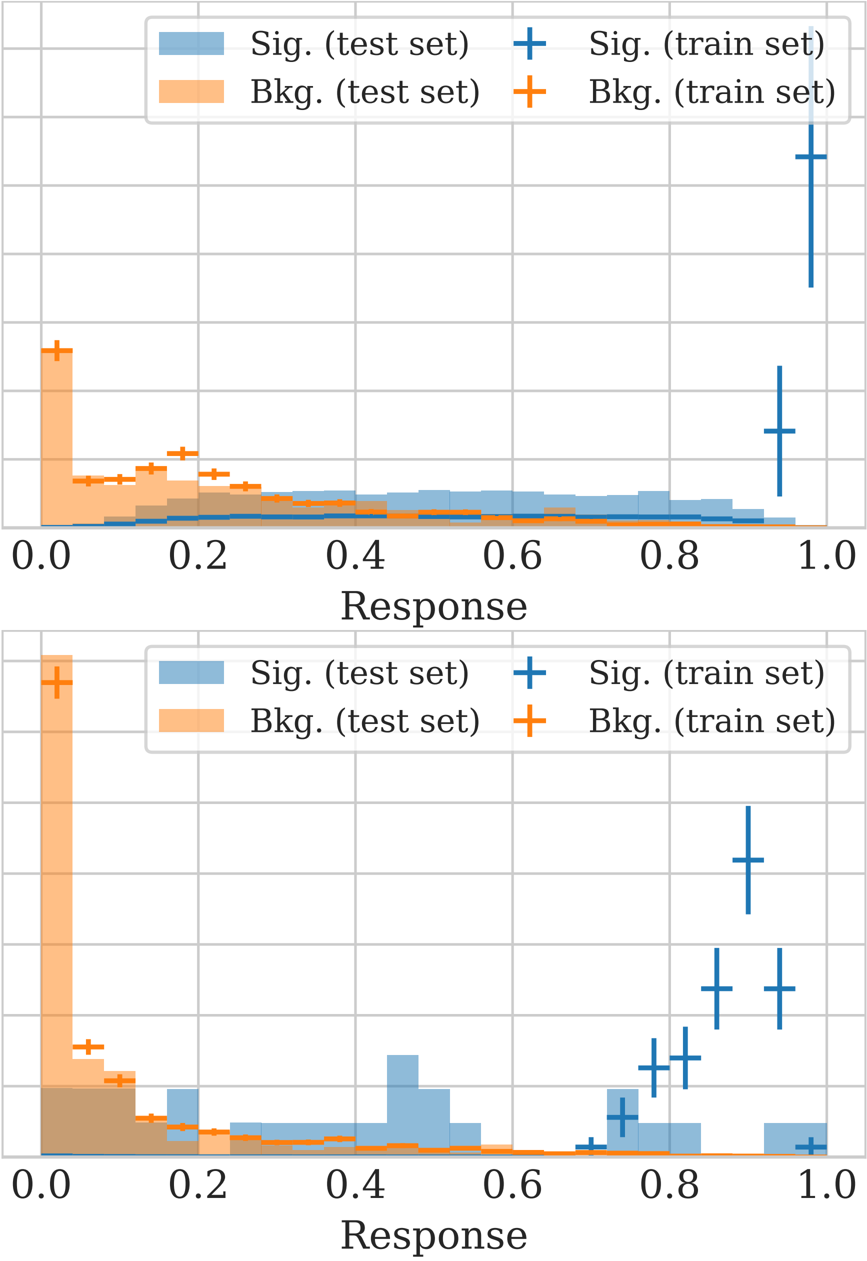


Figure 112: The BDT classifier response for the signal and background samples of 2017, for the *ee* (top) and $\mu\mu$ channels (bottom).

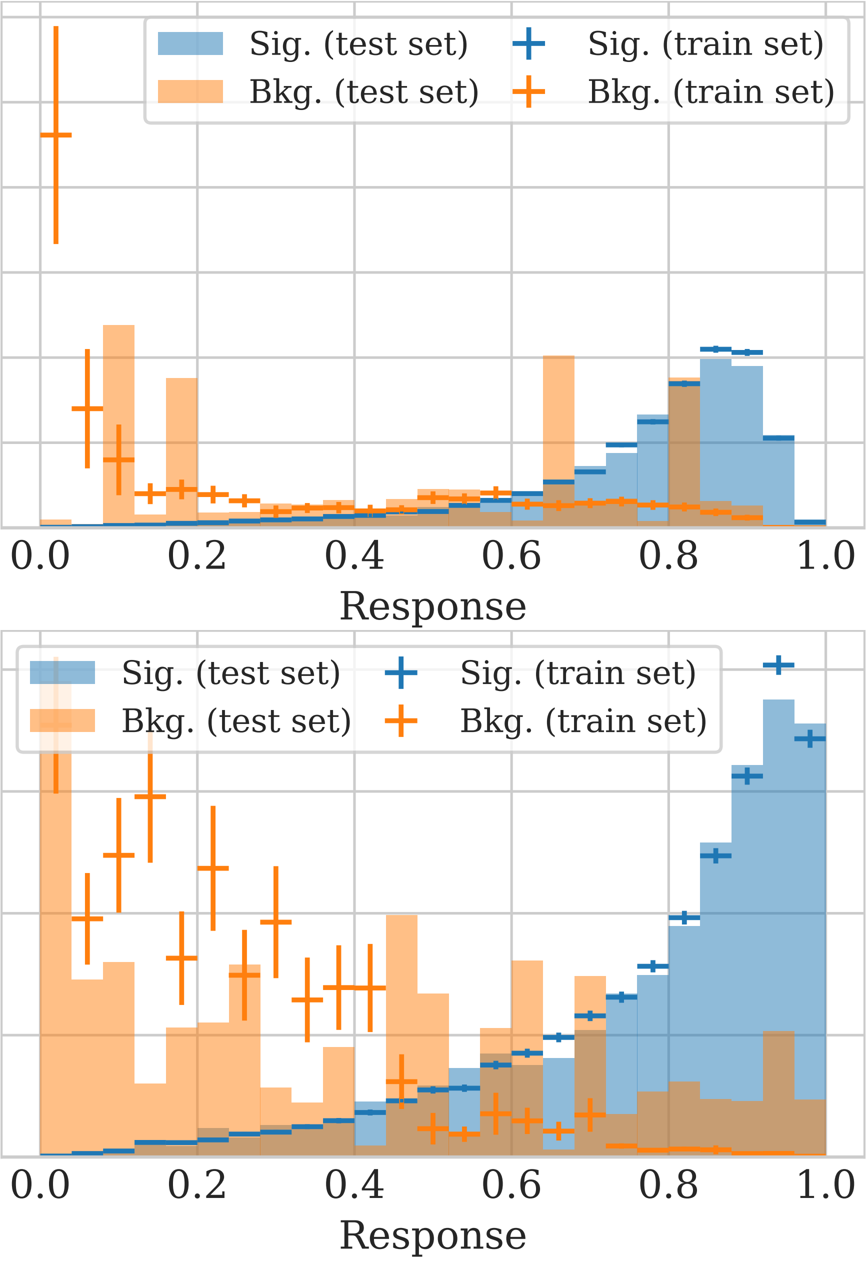


Figure 113: The BDT classifier response for the signal and background samples of 2018, for the *ee* (top) and $\mu\mu$ channels (bottom).

10.3.3 Receiver Operating Characteristic Curves

Figures 114-116 show the receiver operating characteristic (ROC) curves for 2016, 2017 and 2018, for the *ee* and $\mu\mu$ channels of the side band region. The curves shown are for the test and train sets. A receiver operating characteristic curve shows the true positive rate against the false positive rate. The true positive rate, which can also be referred to as the probability of detection, is the probability of detecting a signal-like or background-like event. The false positive rate, which can also be referred to as the probability of false alarm, is the probability of incorrectly classifying a signal-like or background-like event. Therefore, the top-left hand corner of an individual plot represents a false positive rate of zero and a true positive rate of one, which is the ideal scenario. The area under a ROC curve (AUROC) is of interest because the closer the curve is to the top-left corner, the greater the AUROC, which indicates a better performance.

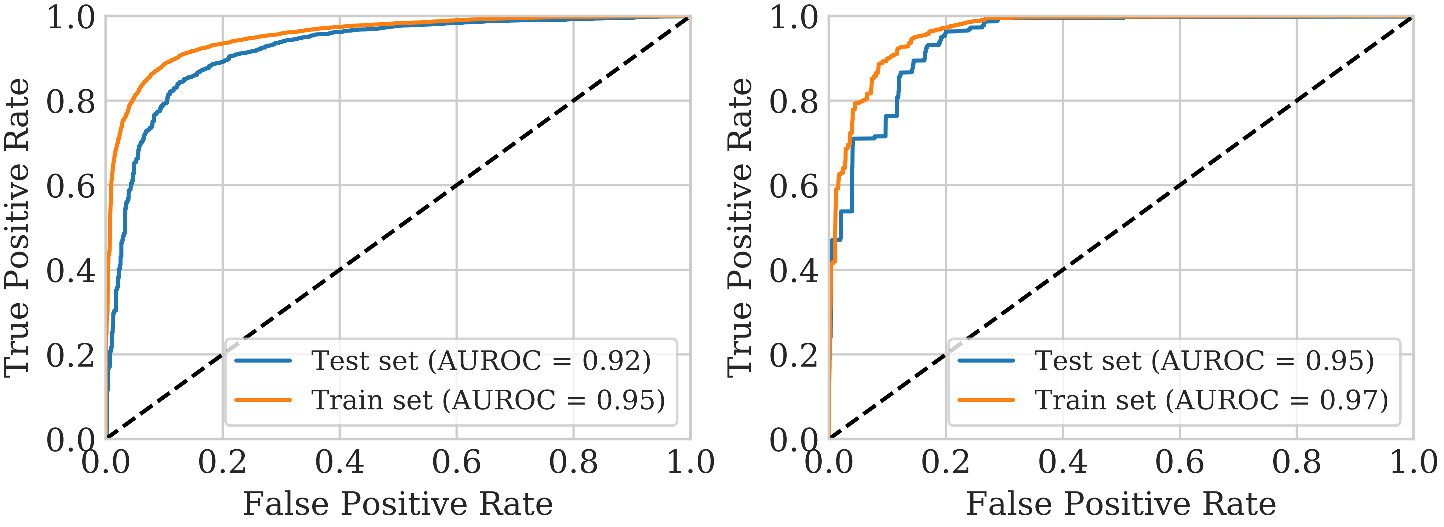


Figure 114: The ROC curves for the signal and background samples of 2016, for the ee (left) and $\mu\mu$ channels (right).

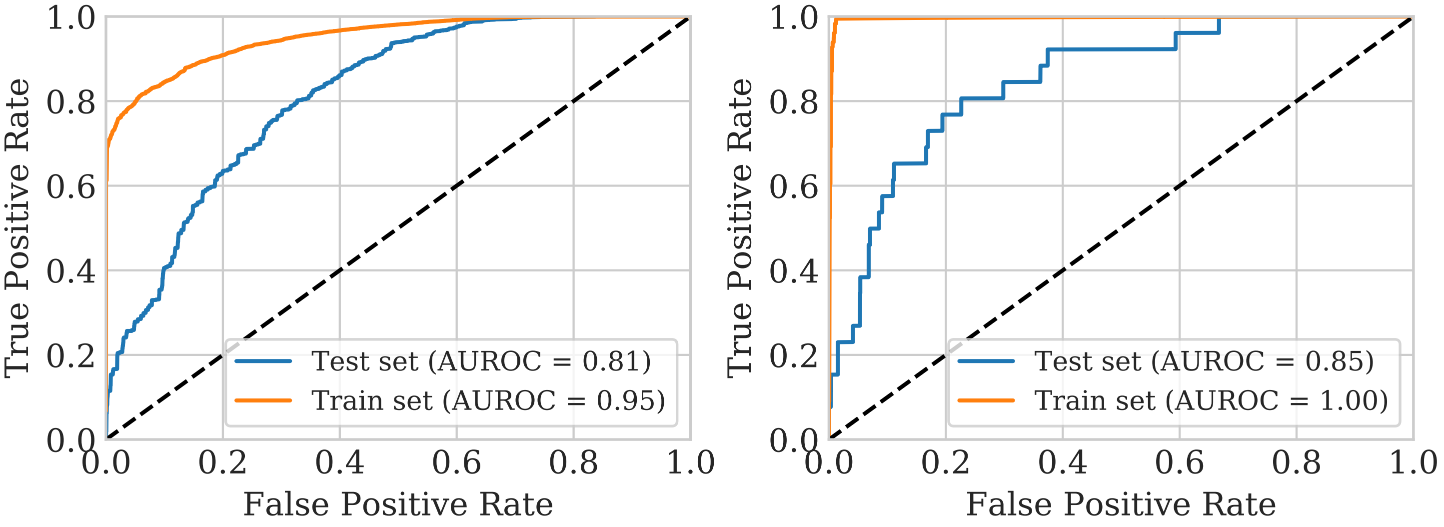


Figure 115: The ROC curves for the signal and background samples of 2017, for the ee (left) and $\mu\mu$ channels (right).

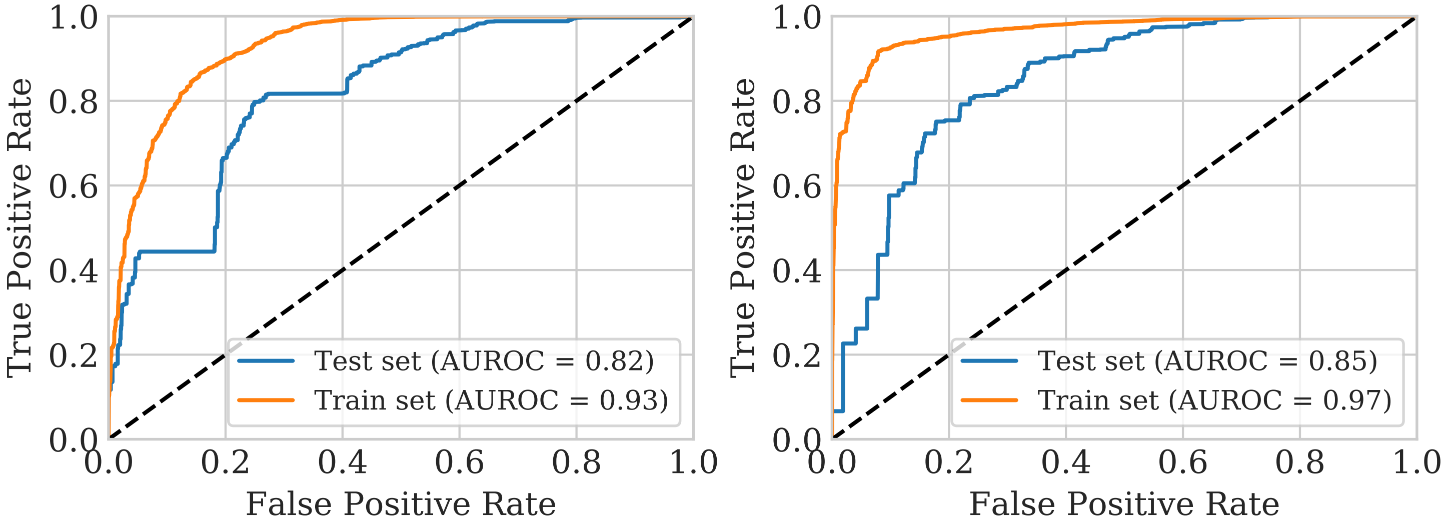


Figure 116: The ROC curves for the signal and background samples of 2018, for the *ee* (left) and $\mu\mu$ channels (right).

10.4 Pre-fit Impact of the Systematic Uncertainties

Figures 117-124 show the prefit impacts of the systematic uncertainties. Prefit impacts show the influence that each systematic uncertainty has on the overall yield of the simulated samples. The shape uncertainties that were considered were in: the pile up (denoted by *pileup_*), the b-tagging scale factors (denoted by *bTag_*), the lepton efficiencies (denoted by *trig_*), the jet energy resolution (denoted by *jer_*), the jet energy scale (denoted by *JEC_*), the parton distribution functions (denoted by *pdf_*), the matrix element (denoted by *ME_*), the missing transverse energy (denoted by *MET_*), the initial state radiation (denoted by *isr_*) and the final state radiation (denoted by *fsr_*). The rate uncertainties considered were on the luminosity (denoted by *lumi_*) and the rate uncertainties applied to specific processes, where results were available from the processes recommended by members of the CMS TOP group. These were: *ZPlusJets_rate* for the Z+jets process, *ZPlusJets_M50_aMCatNLO_rate* for the Z+jets aMCatNLO, *ttWJetsToLNu_rate* for $ttWJets \to L\nu$, *WWZTo4F_rate* for $WWZ \to 4F$, *ttWJetsToQQ_rate* for $ttWJets \to QQ$, *ttbarV_ttZToLL_rate* for $ttZ \to LL$ and *SingleTop_tchannel_top_rate* for single top t-channel.

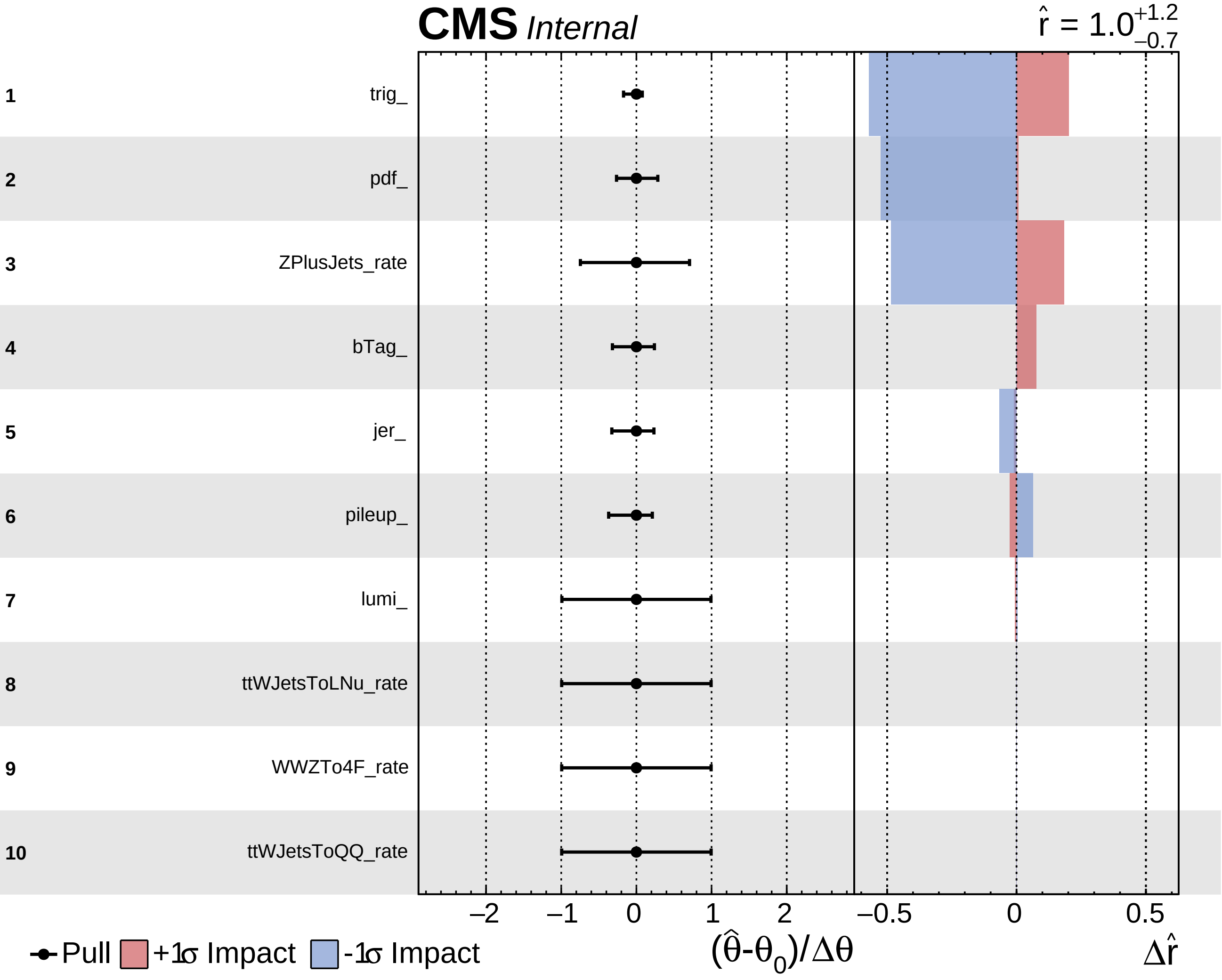


Figure 117: The pre-fit impacts of the systematic uncertainties for the *ee* channel, 2016.

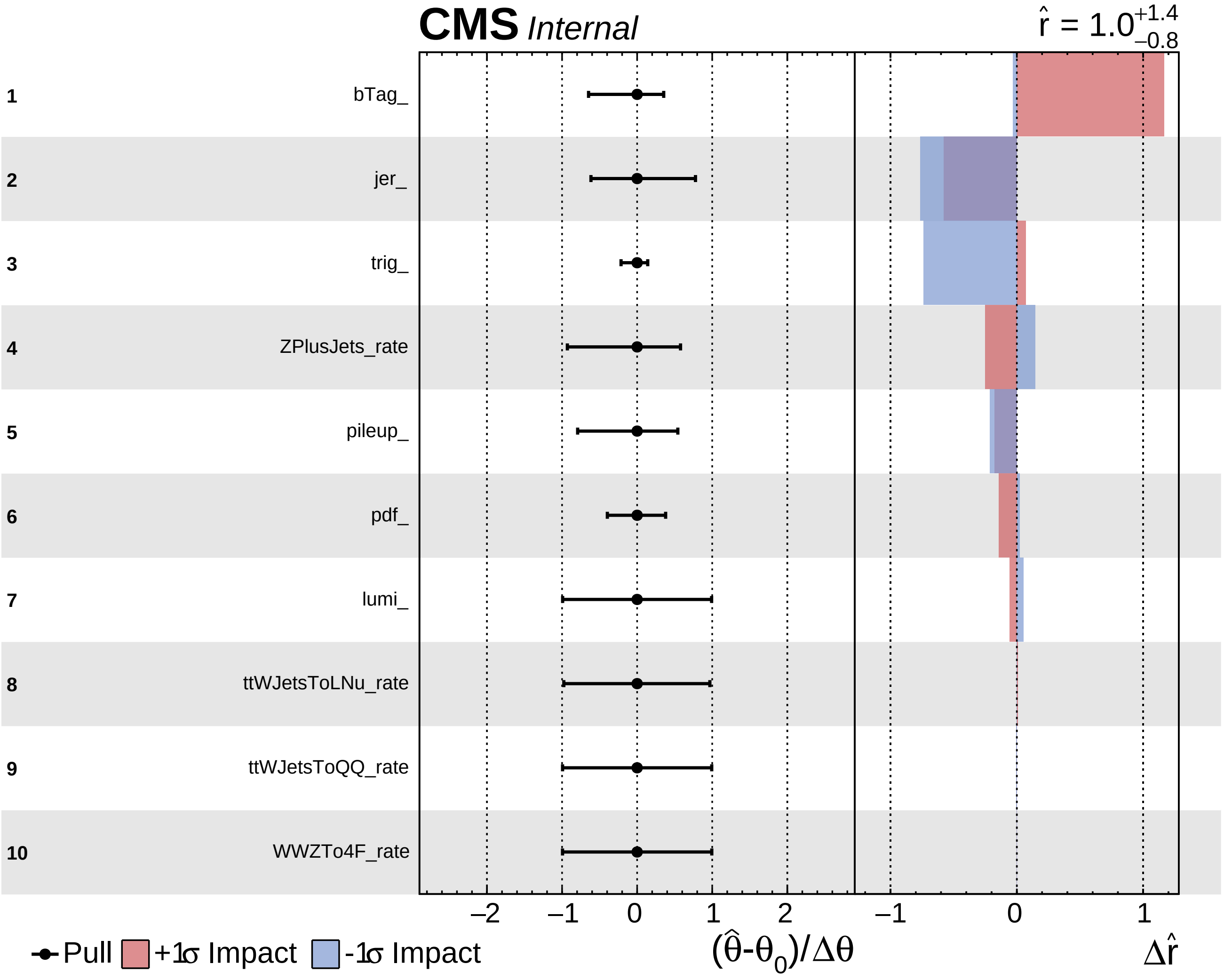


Figure 118: The pre-fit impacts of the systematic uncertainties for the $\mu\mu$ channel, 2016.

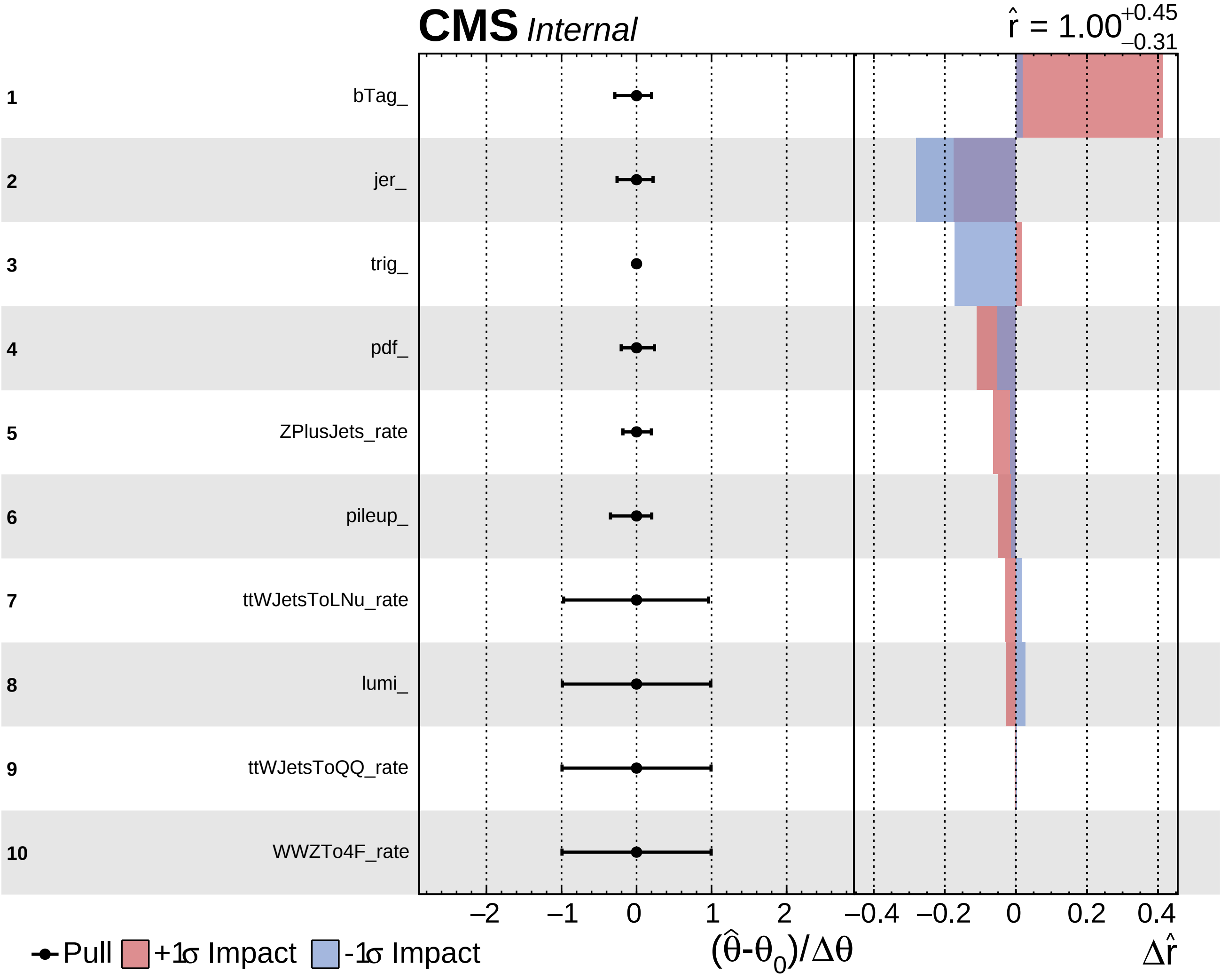


Figure 119: The pre-fit impacts of the systematic uncertainties, combined for the ee and $\mu\mu$ channels, 2016.

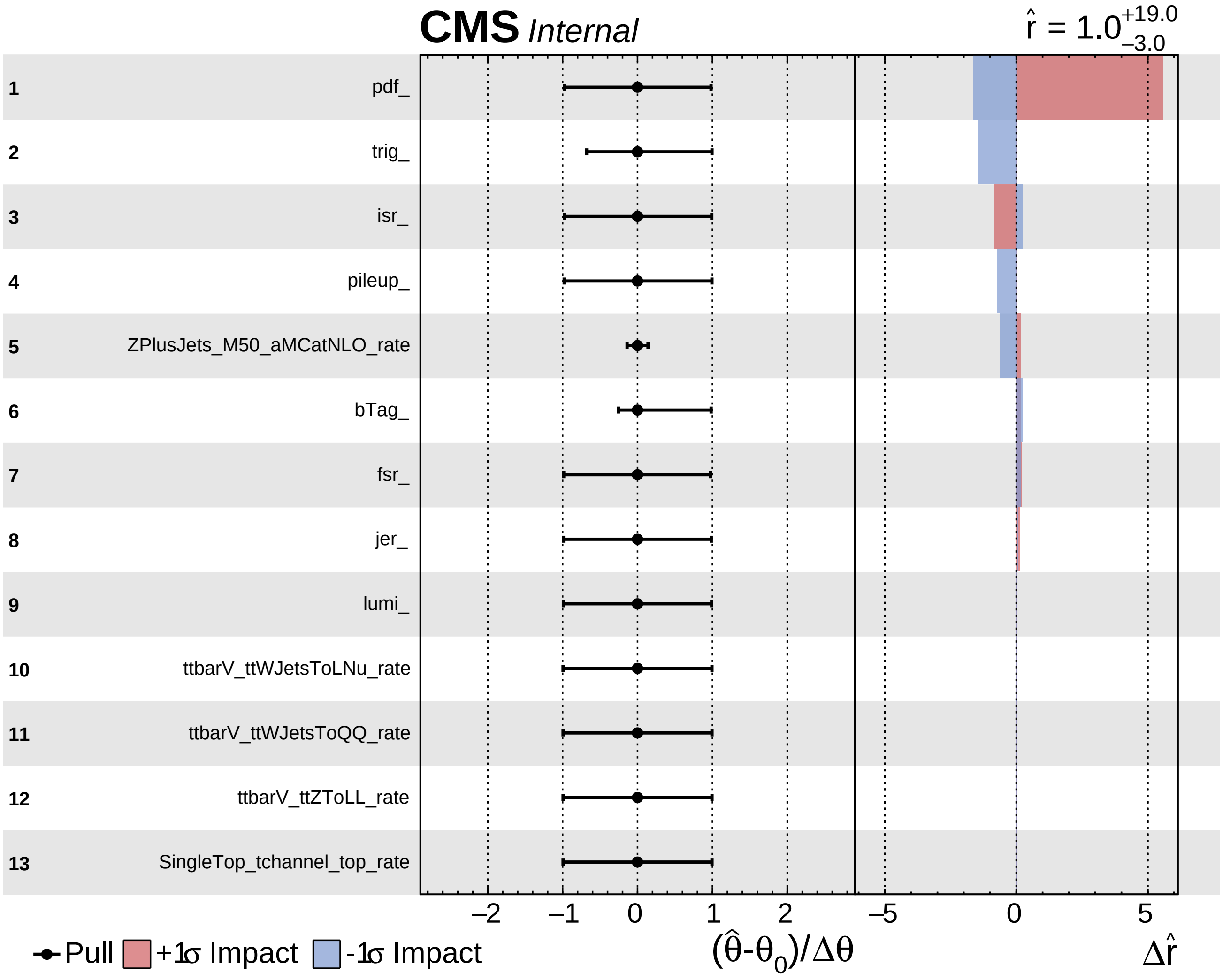


Figure 120: The pre-fit impacts of the systematic uncertainties for the *ee* channel, 2017.

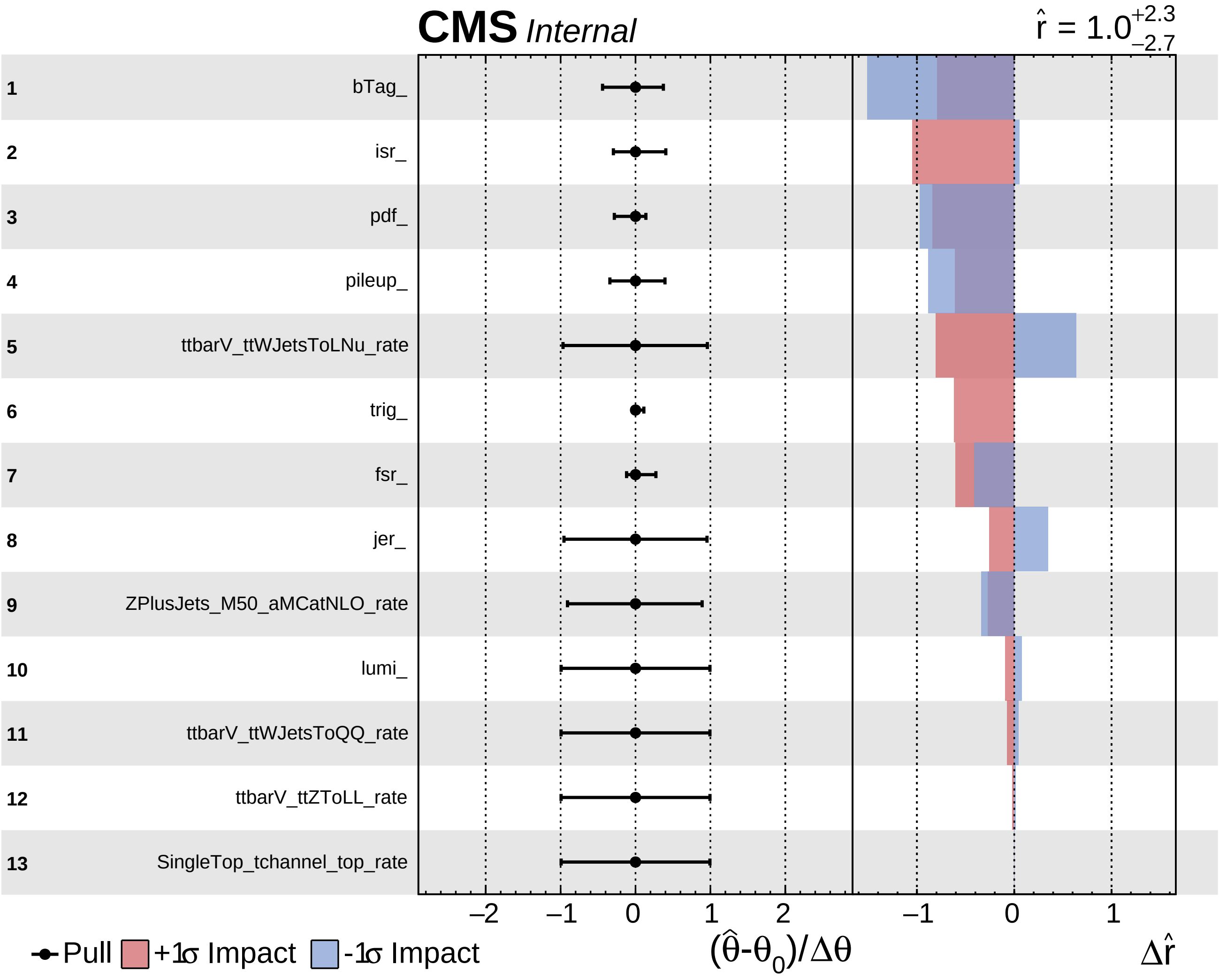


Figure 121: The pre-fit impacts of the systematic uncertainties for the $\mu\mu$ channel, 2017.

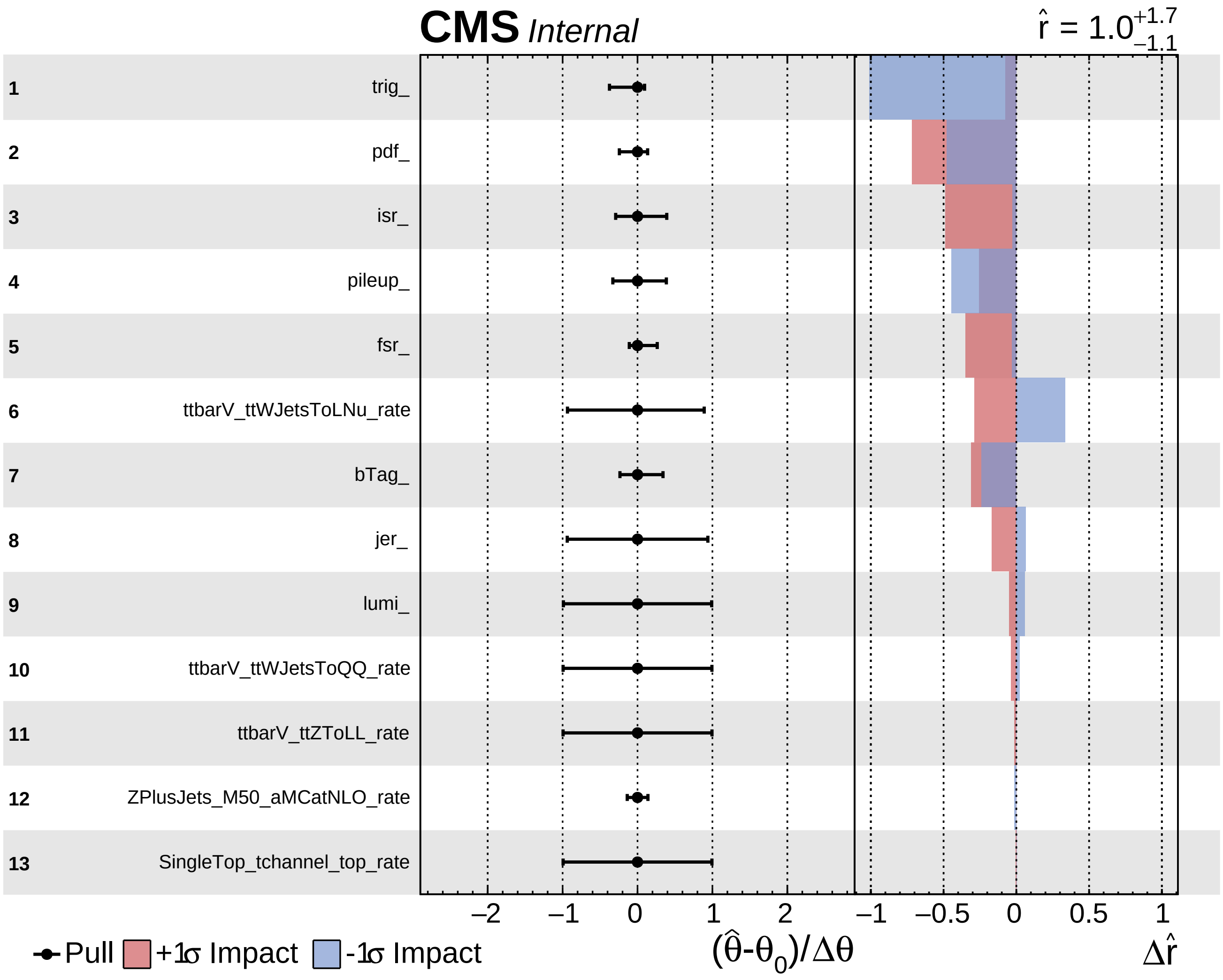


Figure 122: The pre-fit impacts of the systematic uncertainties, combined for the ee and $\mu\mu$ channels, 2017.

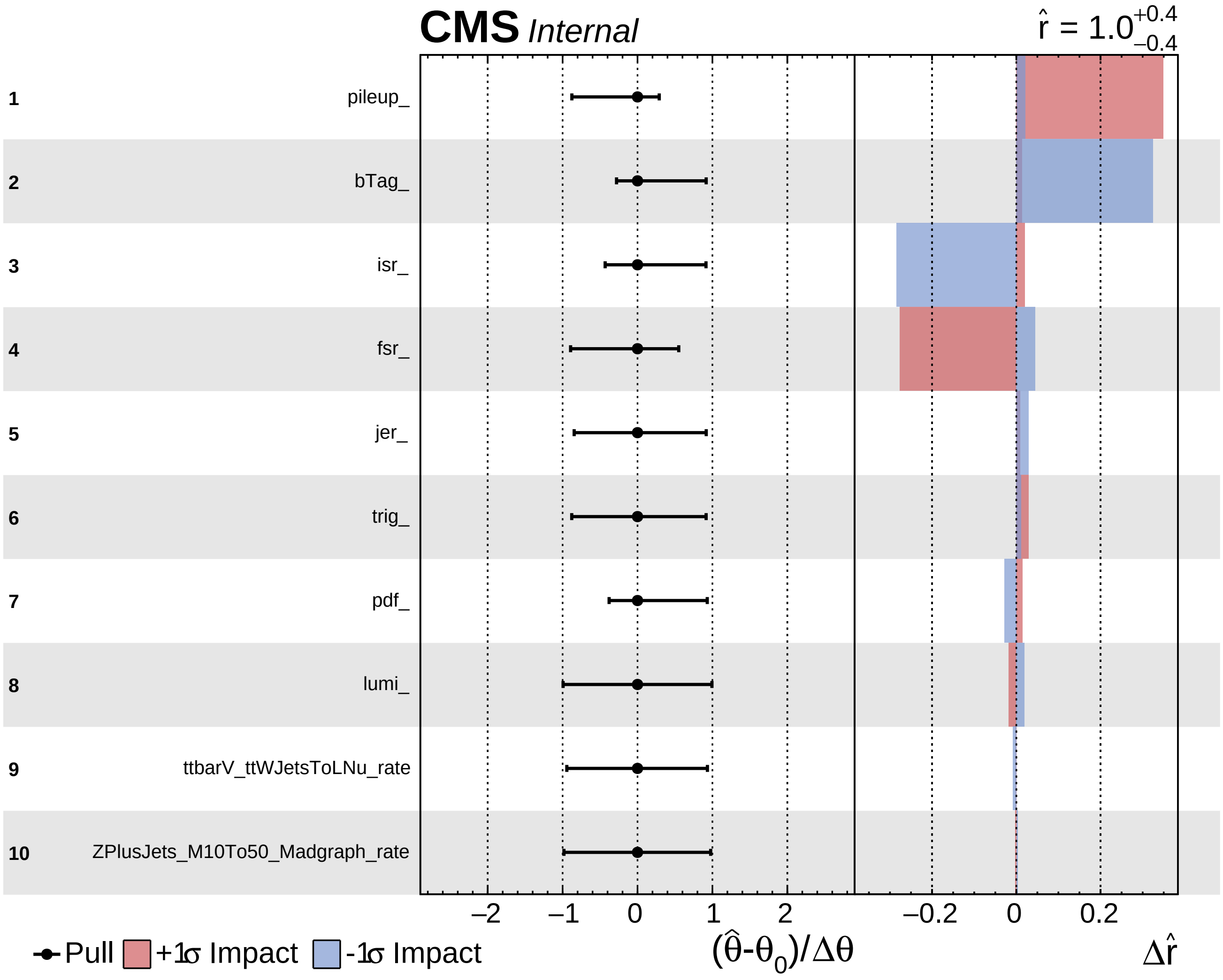


Figure 123: The pre-fit impacts of the systematic uncertainties for the *ee* channel, 2018.

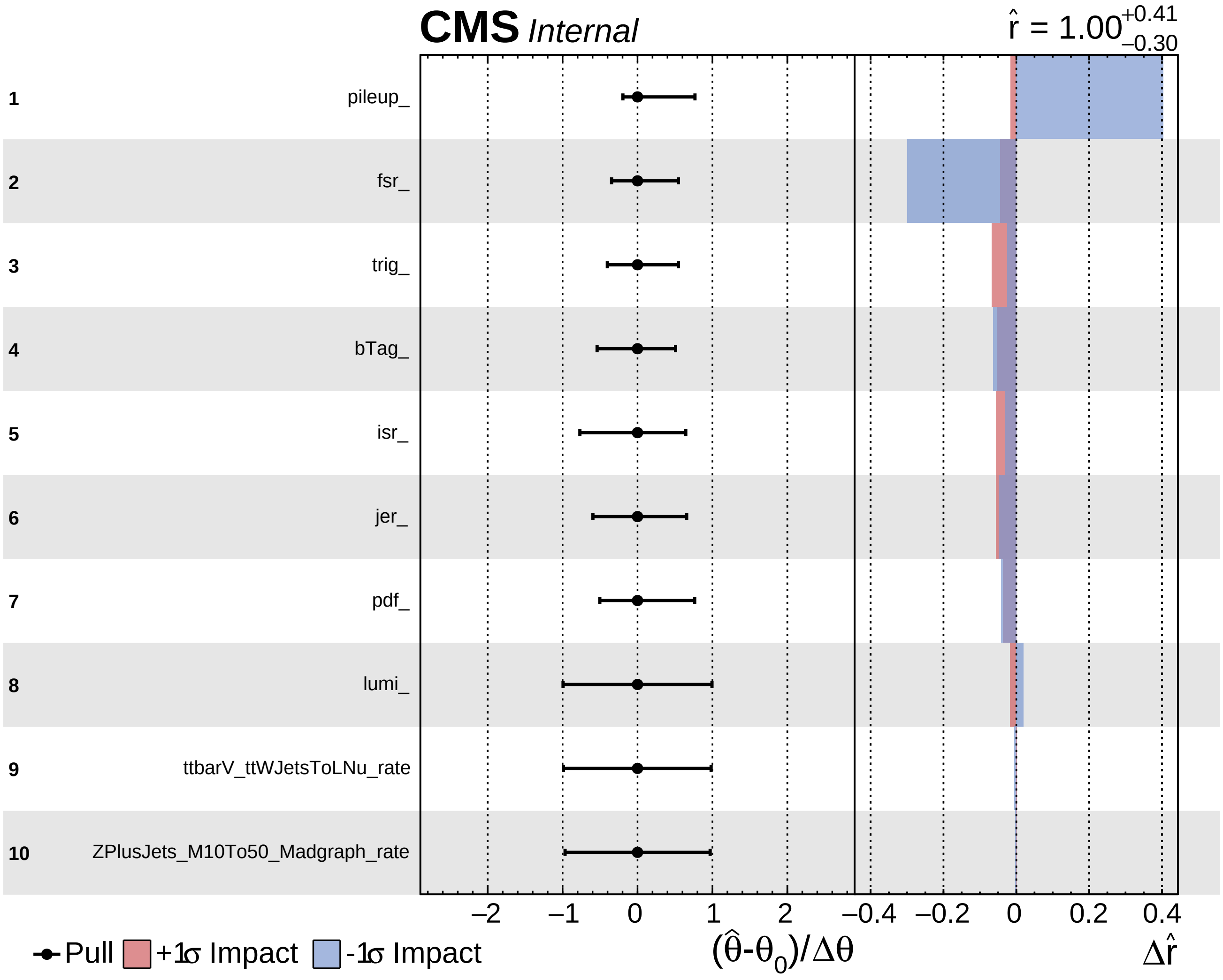


Figure 124: The pre-fit impacts of the systematic uncertainties for the $\mu\mu$ channel, 2018.

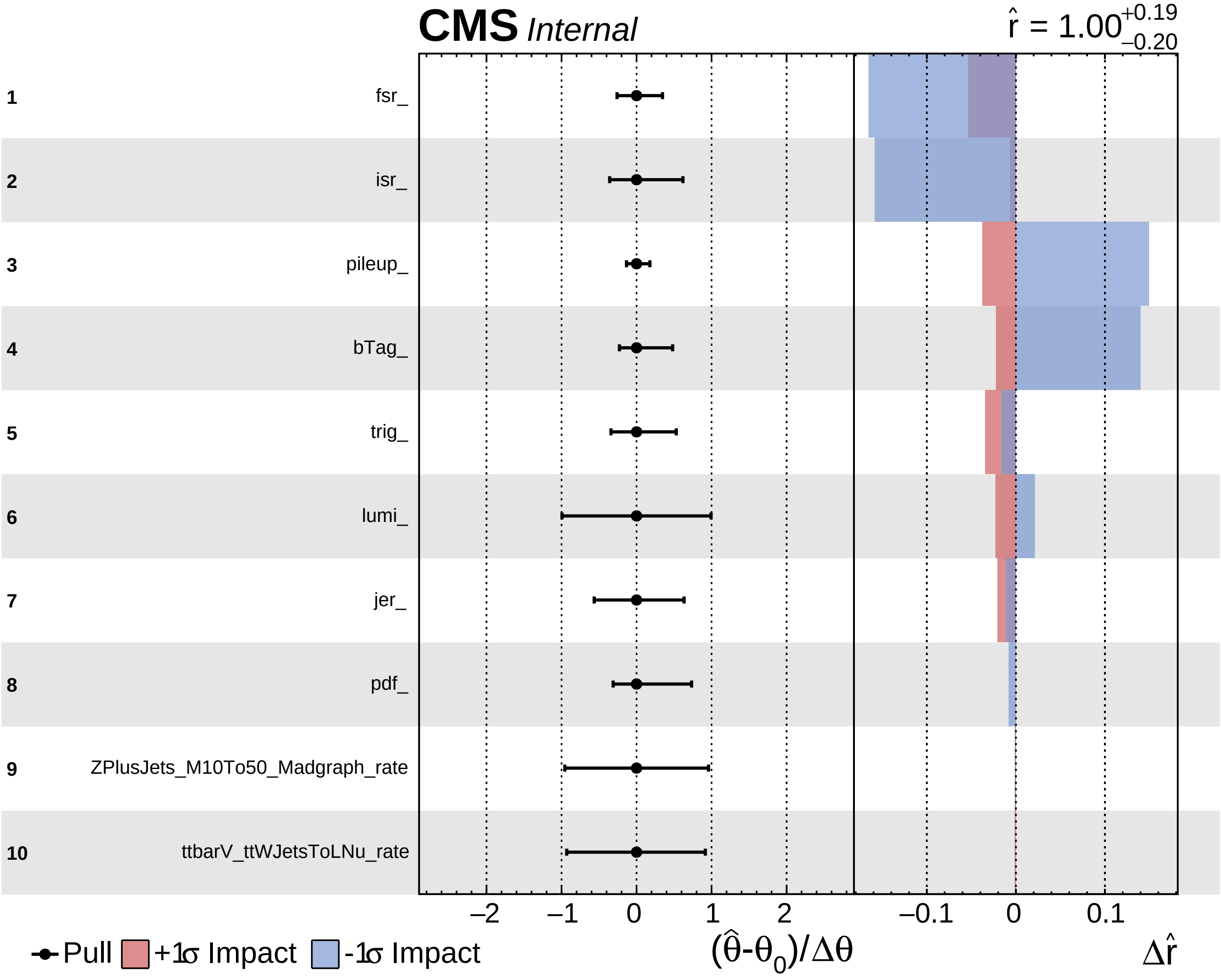


Figure 125: The pre-fit impacts of the systematic uncertainties, combined for the ee and $\mu\mu$ channels, 2018.

11 Conclusions and Future Work

A search was conducted for tZq production in dilepton final states by analysing 137 fb^{-1} of data recorded by the CMS detector at CERN in the data-taking period of 2016-2018. Previous tZq searches conducted by the ATLAS and CMS collaborations were for tZq production in trilepton final states, where the Z boson candidate decays leptonically into a pair of leptons while the W boson candidate decays semi-leptonically into a lepton and a neutrino. For the dilepton search, events containing two same-flavour opposite-sign final state particle flow leptons and 4-6 final jets, 1-2 of which being bjets, were selected. Two channels, ee and $\mu\mu$, were considered. The reconstruction of the Z boson candidate was achieved using the four momenta of the leading and subleading lepton candidates with an invariant mass within ± 20 GeV of the known Z boson mass. Similarly, the reconstruction of the W boson candidate was achieved using the four momenta of the leading and subleading jet candidates with an invariant mass within ± 20 GeV of the known W boson mass. The top quark candidate was fully reconstructed using the four momenta of the W boson and leading b-jet candidates. The analysis was conducted in the side-band region. Two control regions, the Z+jets and $t\bar{t}$ control regions, were assessed to ensure that the Z+jets and $t\bar{t}$ backgrounds were correctly modelled in simulation, as these were the dominant background contributions. Across the entire dataset, the observed significances were found to be 2.85, 0.60 and 11.9σ for 2016, 2017 and 2018, respectively.

Multivariate analysis was conducted on the selected events to better distinguish between signal-like and background-like events. Boosted Decision Trees were chosen as they had been found to be the optimal machine learning technique for the analysis when compared with other machine learning techniques, such as multilayer perceptrons [45]. An optimisation of the hyperparameters was conducted.

Due to technical issues encountered with HTCondor jobs unexpectedly failing, results for certain processes were not outputted. Due to time constraints this issue unfortunately could not be addressed, so having the missing output results would be of benefit in improving the accuracy of the results. The observed significances for 2016 and 2017 differ from those found previously (1.67 and 2.97, respectively [45]), which is possibly a knock-on effect from the technical issues experienced.

Apart from the difference between MC and data, it is a bit surprising that there is less $t\bar{t}$ in the $t\bar{t}$ control region. The diboson contribution on the other hand is rather large. This may indicate that there is something not correct with the b tagging selection [105].

For the transverse momentum distribution of the reconstructed Z boson candidate ($e\mu$ channel, 2016, $t\bar{t}$ CR), data looks reasonable but MC looks completely different. It is also surprising there is an edge at around 120 GeV. It looks as if the Z boson is reconstructed completely different for data and MC, which should not be the case, especially since there is at least the same information available in MC as in data. This may suggest a bug in the Z boson reconstruction. There may be an issue with the implementation of the lepton selection for the $t\bar{t}$ CR that might be causing this problem [105].

It has since been realised that there was a mistake in the lepton selection for 2017. The muon detector coverage is $|\eta| < 2.4$ for the full Run 2, thus, the muons have to be required within $|\eta| < 2.4$ also for 2017 and 2018, rather than $|\eta| < 2.5$ [106].

In 111, differences are observed in the shapes of the training and tests. For example, the test background is a lot lower and shows a higher uncertainty. This is quite likely to be the result of a lower statistic.

In 112, a large discrepancy in the signal region is observed. The discrepancies between signal and background in can be the result of overfitting or failed jobs. However, it is more likely to be the result of the failed HTCondor jobs as this largely impacted the analysis as a whole. The lower statistic could have resulted in an uneven distribution between signal and background.

In 113, the observed discprepancy for the background region could be as a result of fewer samples in the category, which likely would have been a result of the failed HTCondor jobs. Alternatively, this might have been the result of overfitting.

Correlation plots were also obtained. For example, in 105, the subleading and leading jet p_T are highly correlated as there are only minor differences in the selection cuts used to select them. This may also be due to very energetic events, in which the two leading jets would likely have high momenta. There is slightly less correlation with the leading lepton p_T, possibly due to further differences in the selection cuts applied.

Additionally, in 108 many high correlations are shown. Correlation plots are used to eliminate variables and there is often little reason to use variables that have most of the same information in a model, otherwise it could confuse the model. Therefore, a potential improvement would be to remove some of the highly-correlated variables. The correlations between variables for the signal is seen to be stronger in 108 than in 105. This is potentially due to the change in selection cuts used between 2016 and 2017. This would have affected the event yield potentially impacted the correlations observed.

A further improvement that could be made to the analysis is the use of the Deep Combined Secondary Vertex (DeepCSV) b-tagging algorithm, as opposed to the CSVv2 algorithm. The DeepCSV algorithm [104], is the leading b-tagging algorithm used in CMS. It is based on deep neural network training and uses more particle tracks than in the CSVv2 algorithm. A comparison of the misidentification probabilities of the DeepCSV and CSVv2 algorithms can be seen in Figure 126 and Figure 127 shows the comparison for the b-jet efficiency as a function of the jet transverse momentum.

Future work that could be of interest would be to conduct the search by analysing Run 3 data. Run 3 is the LHC data taking era taking place between 2022 and 2026, where proton-proton collisions will occur at 13.6 TeV (6.8 TeV per beam). This is expected to yield 280 fb^{-1} of data; this increase in statistics could help to reduce the $t\bar{t}$ and Z+jets backgrounds, which were found to be the two most dominant background contributions in the analysis. Additionally, achieving a combination result for the results of the Run 2 search and a potential Run 3 search could also be of interest.

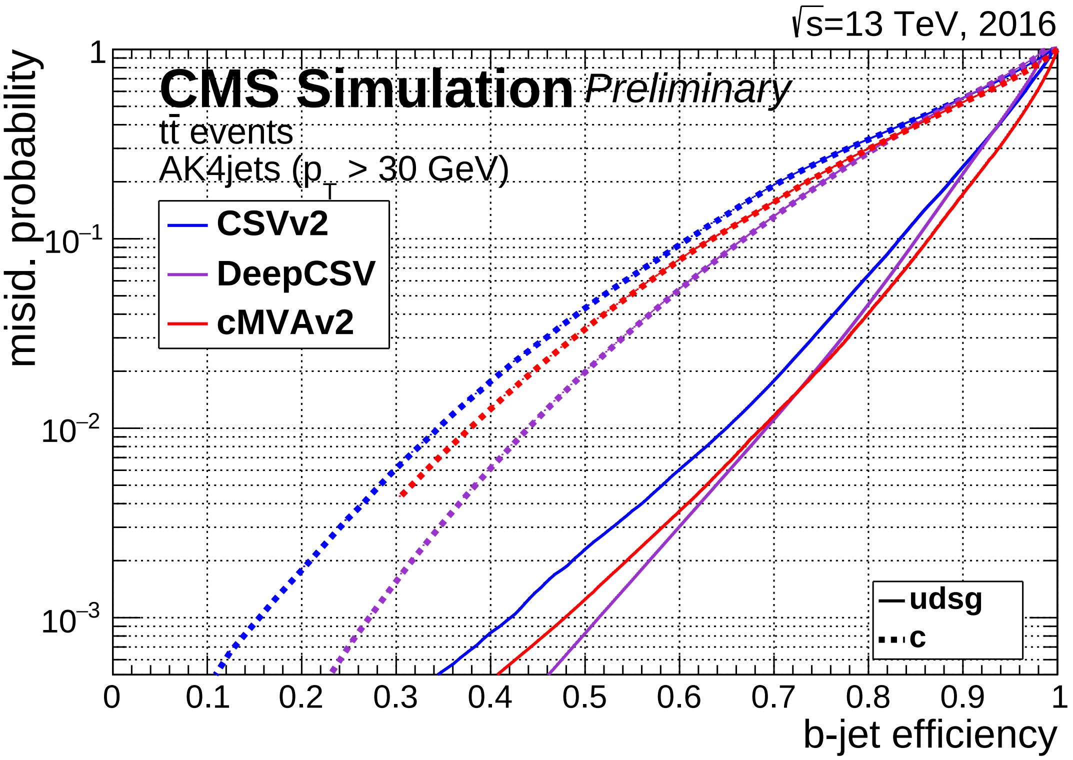


Figure 126: A comparison of the performances of the DeepCSV and CSVv2 b-tagging algorithms. The probability for non-b jets to be misidentified as b jet is shown as a function of the efficiency to correctly identify b jets [104].

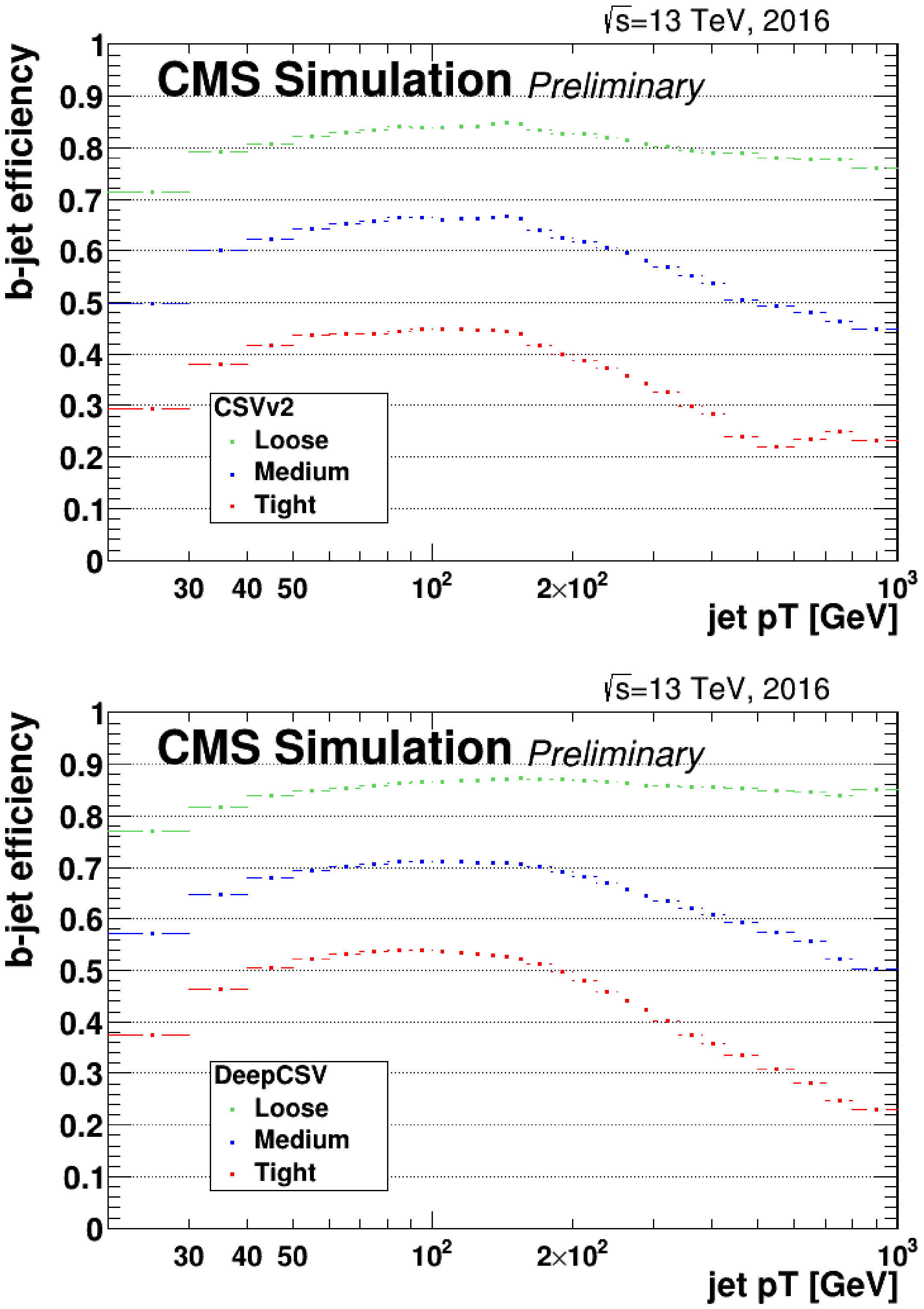


Figure 127: The b-jet efficiency as a function of the jet transverse momentum for the CSVv2 and DeepCSV algorithms [104].

References

[1] Discovery of the electron and nucleus. Khan Academy. Available: `https://www.khanacademy.org/science/chemistry/electronic-structure-of-atoms/history-of-atomic-structure/a/discovery-of-the-electron-and-nucleus`

[2] Kurzon. (2014). The plum pudding model of the atom, as proposed by JJ Thomson. Available: `https://commons.wikimedia.org/wiki/File:Plum_pudding_model.svg`

[3] Griffiths, D (2008). Introduction to Elementary Particles. 2nd ed. Weinhelm: Wiley-VCH. p13-15.

[4] The Standard Model of Elementary Particles. (2006). Available: `https://en.wikipedia.org/wiki/Standard_Model#/media/File:Standard_Model_of_Elementary_Particles.svg`. Last accessed: 20/10/2018.

[5] Martin B.R. and Shaw G. Particle Physics.

[6] LHCb Collaboration. (2014). Observation of the resonant character of the $Z(4430)^-$ state. Available: `https://arxiv.org/abs/1404.1903`

[7] LHCb Collaboration. (2015). Observation of $J/\psi p$ resonances consistent with pentaquark states in $\Lambda_b^0 \rightarrow J/\psi K^- p$ decays. Available: `https://arxiv.org/abs/1507.03414`

[8] Abi, B, et al. (2021). Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm. Available: `https://journals.aps.org/prl/pdf/10.1103/PhysRevLett.126.141801`

[9] Available: `https://commons.wikimedia.org/w/index.php?curid=58326280`

[10] Perl, M.L. et al. (1975). Evidence for anomalous lepton production in $e^+ - e^-$ annilhilation: `https://inspirehep.net/files/84721aabdc0df67f8f9e1c3b218d6de3`

[11] Thomson, M. (2013). Modern Particle Physics.

[12] Thomson, M. (2013). CP violation and weak hadronic interactions. In: Modern Particle Physics. Cambridge: Cambridge University Press. p370.

[13] Higgs boson. Available: `https://en.wikipedia.org/wiki/Higgs_boson`

[14] ATLAS Collaboration. (2012). Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. Available: `https://arxiv.org/abs/1207.7214`

[15] CMS Collaboration. (2012). Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. Available: `https://arxiv.org/abs/1207.7235`

[16] The CMS Collaboration. (2018). Observation of Higgs boson decay to bottom quarks. Available: `https://arxiv.org/pdf/1808.08242.pdf`

[17] The CMS Collaboration. (2017). Observation of the Higgs boson decay to a pair of tau leptons with the CMS detector. Available: `https://arxiv.org/abs/1708.00373`

[18] CMS Collaboration. (2012). Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. Available: `https://journals.aps.org/prl/pdf/10.1103/PhysRevLett.13.138`

[19] Randall, L and Sundrum, R. (1999). A Large Mass Hierarchy from a Small Extra Dimension. `https://arxiv.org/abs/hep-ph/9905221`

[20] Collard, C. (2014). RS Graviton Searches in CMS. Ecole Polytechnique Paris.

[21] Patrignani *et al* (Particle Data Group). (2017). Chin. Phys. C, 40 (100001). Available: `http://pdg.lbl.gov/2017/tables/rpp2017-sum-quarks.pdf`

[22] Particle Data Group. (2017). Top quark. Available: `http://pdg.lbl.gov/2018/reviews/rpp2018-rev-top-quark.pdf`

[23] R.L. Workman et al. (Particle Data Group), Prog. Theor. Exp. Phys. 2022, 083C01 (2022). Available: `https://pdg.lbl.gov/2022/listings/contents_listings.html`

[24] CMS Collaboration. (2018). Measurement of the $t\bar{t}$ production cross section, the top quark mass, and the strong coupling constant using dilepton events in pp collisions at $\sqrt{s}$= 13 TeV. Available: `http://cms-results.web.cern.ch/cms-results/public-results/publications/TOP-17-001/index.html`

[25] ATLAS Collaboration. (2020). Measurement of the $t\bar{t}$ production cross-section in the lepton+jets channel at $\sqrt{s}$=13 TeV with the ATLAS experiment. Available: `https://www.sciencedirect.com/science/article/pii/S0370269320306006?via=ihub`

[26] CMS Collaboration. (2012). Measurement of the mass difference between top and antitop quarks. Available: `http://cms-results.web.cern.ch/cms-results/public-results/publications/TOP-11-019/index.html`

[27] CMS Collaboration. (2016). Measurement of the mass difference between top quark and antiquark in pp collisions at $\sqrt{s}$= 8 TeV. Available: `http://cms-results.web.cern.ch/cms-results/public-results/publications/TOP-12-031/index.html`

[28] ATLAS and CMS Collaborations. (2020). Combination of the W boson polarization measurements in top quark decays using ATLAS and CMS data at $\sqrt{s}$= 8 TeV. Available: `https://cms-results.web.cern.ch/cms-results/public-results/publications/TOP-19-004/index.html`

[29] CMS Collaboration. (2019). Measurement of the top quark polarization and $t\bar{t}$ spin correlations using dilepton final states in proton-proton collisions at $\sqrt{s}$= 13 TeV. Available: `https://journals.aps.org/prd/abstract/10.1103/PhysRevD.100.072002`

[30] CMS Collaboration. (2014). Search for standard model production of four top quarks in the lepton + jets channel in pp collisions at $\sqrt{s}$ = 8 TeV. Available: `http://cms-results.web.cern.ch/cms-results/public-results/publications/TOP-13-012/index.html`

[31] CMS Collaboration. (2020). Search for production of four top quarks in final states with same-sign or multiple leptons in proton–proton collisions at $\sqrt{s}$ = 13 TeV. Available: `https://link.springer.com/article/10.1140/epjc/s10052-019-7593-7`

[32] CMS Collaboration. (2016). Search for CP violation in tt^{-} production and decay in proton-proton collisions at $\sqrt{s}$= 8 TeV. Available: `http://cms-results.web.cern.ch/cms-results/public-results/publications/TOP-16-001/index.html`

[33] Baskakov, A.V. et al. (2015). Single top quark production in heavy ion collisions at energies available at the CERN Large Hadron Collider. Phys.Rev. 4 (044901). Available: `http://inspirehep.net/record/1345175/plots`

[34] D0 Collaboration. (2009). Observation of Single Top-Quark Production. Available: `http://dx.doi.org/10.1103/PhysRevLett.103.092001`

[35] CDF Collaboration. (2009). Observation of Electroweak Single Top-Quark Production. Available: `http://dx.doi.org/10.1103/PhysRevLett.103.092002`

[36] D0 Collaboration. (2013). Evidence for s-channel single top quark production in pp collisions at $sqrts$ = 1.96 TeV. Available: `http://dx.doi.org/10.1016/j.physletb.2013.09.048`

[37] CDF and D0 Collaborations. (2014). Observation of s-Channel Production of Single Top Quarks at the Tevatron. Available: `http://dx.doi.org/10.1103/PhysRevLett.112.231803`

[38] ATLAS Collaboration. (2015). Evidence for single top-quark production in the s-channel in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector using the Matrix Element Method. Available: `http://www.arxiv.org/abs/1511.05980`

[39] ATLAS and CMS Collaborations. (2019). Combinations of single-top-quark production cross-section measurements and $|f_{LV}V_{tb}|$ determinations at $\sqrt{s}$ = 7 and 8 TeV with the ATLAS and CMS experiments. Available: `https://arxiv.org/pdf/1902.07158.pdf`

[40] CMS Collaboration. (2014). Observation of the associated production of a single top quark and a W boson in pp collisions at $\sqrt{s}$ = 8 TeV .Available: `http://cms-results.web.cern.ch/cms-results/public-results/publications/TOP-12-040/index.html`

[41] CMS Collaboration. (2018). Measurement of the production cross section for single top quarks in association with W bosons in proton-proton collisions at $\sqrt{s}$ 13 TeV . Available: `http://cms-results.web.cern.ch/cms-results/public-results/publications/TOP-17-018/index.html`

[42] ATLAS and CMS Collaborations. (2019). Combinations of single-top-quark production cross-section measurements and $|f_{LV}V_{tb}|$ determinations at $\sqrt{s}$= 7 and 8 TeV with the ATLAS and CMS experiments Available: `http://cms-results.web.cern.ch/cms-results/public-results/publications/TOP-17-006/index.html`

[43] CMS Collaboration. (2017). Search for associated production of a Z boson with a single top quark and for tZ flavour-changing interactions in pp collisions at $\sqrt{s} = 8$ TeV. JHEP. 7 (003). Available: `http://cms-results.web.cern.ch/cms-results/public-results/publications/TOP-12-039/index.html`

[44] CMS Collaboration. (2021). Search for flavor-changing neutral current interactions of the top quark and the Higgs boson decaying to a bottom quark-antiquark pair at $\sqrt{s}$= 13 TeV. Available: `https://cms-results.web.cern.ch/cms-results/public-results/publications/TOP-19-002/index.html`

[45] Cole, J *et al.* (2018). Search for single top production in association with a Z boson for the dilepton final state in pp collisions at $\sqrt{s} = 13$ TeV in the CMS detector. Available: `https://indico.cern.ch/event/764673/contributions/3174231/attachments/1738575/2812766/main.pdf`

[46] Morton, A. (2017). A search for the Standard Model production of tZq in the dilepton final state using data from CMS at 13 TeV. Available: `https://indico.shef.ac.uk/event/1/contributions/271/attachments/227/231/IOP_admorton.pdf`

[47] European Organisation for Nuclear Research. (2004). LHC Design Report, Volume III: The Injector Chain. Available: `https://cds.cern.ch/record/823808/files/CERN-2004-003-V3.pdf`

[48] The ATLAS Collaboration et al 2008 JINST 3 S08003. Available: `https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08003`

[49] The ALICE Collaboration et al 2008 JINST 3 S08002. Available: `https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08002`

[50] CMS Collaboration. (2016). CMS Physics Technical Design Report Volume I: Detector Performance and Software. Available: `https://cdsweb.cern.ch/record/922757/files/lhcc-2006-001.pdf`

[51] The LHCb Collaboration et al 2008 JINST 3 S08005. Available: `https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08005`

[52] CMS Collaboration. CERN LHC Experiments Committee. (2017). The Phase-2 Upgrade of the CMS Endcap Calorimeter. Available: `https://cds.cern.ch/record/2293646?ln=en`

[53] The TOTEM Collaboration et al 2008 JINST 3 S08007. Available: `https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08007`

[54] The LHCf Collaboration et al 2008 JINST 3 S08006. Available: `https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08006`

[55] ALICE Data Acquisition (DAQ). Available: `http://alice-daq.web.cern.ch/`. Last accessed 17/10/2018.

[56] The CMS Collaboration. (2008). The CMS experiment at the CERN LHC. JINST. 3 (S08004) `http://dx.doi.org/10.1088/1748-0221/3/08/S08004`

[57] Barney, D. (2004). Interactive Slice of the CMS detector. Available: `https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/ShowDocument?docid=4172`. Last accessed 17/10/2018.

[58] CMS Collaboration. (2014). Description and performance of track and primary-vertex reconstruction with the CMS tracker. Available: `https://arxiv.org/pdf/1405.6569.pdf`

[59] CMS Collaboration. (2009). Particle–Flow Event Reconstruction in CMS and Performance for Jets, Taus, and E_T^{miss}. Available: `https://cds.cern.ch/record/1194487/files/PFT-09-001-pas.pdf`

[60] Hill, C. (2010). Lecture 2 - The CMS and ATLAS detectors. Available: `https://warwick.ac.uk/fac/sci/physics/staff/academic/gershon/gradteaching/warwickweek/material/lhcphysics/chill_warwick_lhc_2010_2.pdf`. Last accessed 20/10/2018.

[61] CMS Collaboration. (2017). The CMS trigger system. Available: `https://arxiv.org/pdf/1609.02366.pdf`

[62] CERN. (2015). High-Luminosity Large Hadron Collider (HL-LHC)Preliminary Design Report. `https://cds.cern.ch/record/2116337/files/CERN-2015-005.pdf`

[63] CERN. (2017). High-Luminosity Large Hadron Collider (HL-LHC) Technical Design Report V. 0.1. Available: `http://inspirehep.net/record/1635816/files/11.pdf`

[64] CMS Collaboration. (2017). The Phase-2 upgrade of the Barrel Calorimeters Technical Design Report. Available. `https://cds.cern.ch/record/2283187/files/CMS-TDR-015.pdf`

[65] CERN. Geneva. The LHC experiments Committee. (2015). CMS Technical Design Report for the Muon Endcap GEM Upgrade. Available: `https://cds.cern.ch/record/2021453?ln=en`

[66] CMS Collaboration. (2018). The Phase-2 Upgrade of the CMS Tracker Technical Design Report. Available: `https://cds.cern.ch/record/2272264/files/CMS-TDR-014.pdf`

[67] CMS Collaboration. (2018). Identification of heavy-flavour jets with the CMS detectorin pp collisions at 13 TeV. Available: `https://arxiv.org/pdf/1712.07158.pdf`

[68] Guest D., Cranmer K. and Whiteson D. (2018). Deep Learning and Its Application to LHC Physics. Available: `https://arxiv.org/pdf/1806.11484.pdf`

[69] Hoad, C. (2019). Search for single top production in association with a Zboson for the dilepton final state. Available:`https://twiki.cern.ch/twiki/pub/CMS/SingleTopTZq2015/2017_analysis_slides.pdf`.

[70] Combine. Available: `https://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/`

[71] Salam G. and Caccari M. (2008). The anti-ktjet clustering algorithm. Available: `https://arxiv.org/pdf/0802.1189.pdf`

[72] CMS Collaboration. (2017). Operation and Performance of the CMS Outer Tracker. Available: `https://cds.cern.ch/record/2304928/files/Vertex%202017_013.pdf`

[73] Available: `https://twiki.cern.ch/twiki/bin/viewauth/CMS/StripsOfflinePlots2018`.

[74] Rivet — the particle-physics MC analysis toolkit. Available: `https://rivet.hepforge.org/`

[75] Bartz, E, et al. (2017). FPGA-Based Tracklet Approach to Level-1 Track Finding at CMS for the HL-LHC. Available: `https://arxiv.org/pdf/1706.09225.pdf`

[76] Rules for dataset names. Available: `https://cms-pdmv.gitbook.io/project/mccontact/rules-for-dataset-names`

[77] Sorrentino, G. (2021). Monte Carlo tuning in CMS.

[78] Perucianni, G, et al. (2017). Mini-AOD: A New Analysis Data Format for CMS. Available: `https://arxiv.org/pdf/1702.04685.pdf`

[79] Rizzi, A, et al. (2019). A further reduction in CMS event data for analysis: the NANOAOD format. Available: `https://www.epj-conferences.org/articles/epjconf/pdf/2019/19/epjconf_chep2018_06021.pdf`

[80] CMS Collaboration (2015). Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV. Available: `https://arxiv.org/abs/1502.02701`

[81] CMS Collaboration (2013). Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV. Available: `https://arxiv.org/abs/1206.4071`

[82] CMS Collaboration. (2017). Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. Available: `https://arxiv.org/abs/1712.07158`

[83] Heavy flavour tagging for 13 TeV data in 2017 and 94X MC. Available: `https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation94X`

[84] TOP Systematic Uncertainties (Run2). Available: `https://twiki.cern.ch/twiki/bin/viewauth/CMS/TopSystematics`

[85] CMS Collaboration. (2016). Trigger and Lepton Efficiencies for Top analyses in RunII. Available: `http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2016/025`

[86] M Borisyak et al 2017 J. Phys.: Conf. Ser. 898 092041. Available: `https://iopscience.iop.org/article/10.1088/1742-6596/898/9/092041/pdf`

[87] ROOT reference guide. Available: `https://root.cern.ch/doc/master/classTEfficiency.html#ae80c3189bac22b7ad15f57a1476ef75b`

[88] Rochester corrections. Available: `https://arxiv.org/abs/1208.3710`

[89] `https://twiki.cern.ch/twiki/bin/view/CMS/RochcorMuon#Rochester_Correction`

[90] Jet Energy Resolution. Available: `https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution`

[91] Czakon, M, Heymes, D and Mitov, A. (2015). High-precision differential predictions for top-quark pairs at the LHC. Available: `https://arxiv.org/abs/1511.00549`

[92] CMS Collaboration. (2017). CMS luminosity measurements for the 2016 data-taking period. Available: `https://cds.cern.ch/record/2257069/files/LUM-17-001-pas.pdf`

[93] CMS Collaboration. (2018). CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s}$=13 TeV. `https://cds.cern.ch/record/2621960/files/LUM-17-004-pas.pdf`

[94] Butterworth, J, et al. (2015). PDF4LHC recommendations for LHC Run II. Available: `https://arxiv.org/pdf/1510.03865.pdf`

[95] CMS Collaboration. (2019). CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV. Available: `http://cms-results.web.cern.ch/cms-results/public-results/preliminary-results/LUM-18-002/index.html`

[96] CMS Collaboration. (2017). Measurement of the associated production of a single top quark and a Z boson in pp collisions at $sqrts = 13$ TeV. Available: `https://arxiv.org/abs/1712.02825`.

[97] TMVA 4 Toolkit for Multivariate Data Analysis with ROOT Users Guide. (2018). Available: `https://arxiv.org/pdf/physics/0703039.pdf`

[98] ATLAS and CMS Collaborations. (2019). Combinations of single-top-quark production cross-section measurements and $|f_{LV}V_{tb}|$ determinations at $\sqrt{s}$= 7 and 8 TeV with the ATLAS and CMS experiments. Available: `http://cms-results.web.cern.ch/cms-results/public-results/publications/TOP-17-006/index.html`

[99] Y. Freund and R.E. Schapire, J. of Computer and System Science 55, 119 (1997).

[100] Hoecker, A, et al. (2007). TMVA - Toolkit for Multivariate Data Analysis. Available: `https://arxiv.org/abs/physics/0703039`.

[101] Combine. Available: `https://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/`

[102] Barlow, R, Beeston, C. (1993). Fitting using finite Monte Carlo samples. Available: `https://www.sciencedirect.com/science/article/abs/pii/001046559390005W?via%3Dihub`

[103] Conway, J.S. (2011). Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra. Available: `https://arxiv.org/abs/1103.0354`.

[104] Heavy flavor identification at CMS with deep neural networks. Available: `https://twiki.cern.ch/twiki/bin/view/CMSPublic/BTV13TeVDPDeepCSV`.

[105] In conversation with Nils Faltermann, former CMS t+X subgroup convenor and member of the CMS TOP Group, on Friday 26th May 2023.

[106] In conversation with David Walter, CMS t+X subgroup convenor, on Wednesday 31st May 2023.