Model Unspecific Search for New Physics with pp Collisions at $\sqrt{s} = 13$ TeV with the CMS Experiment

von

Jonas Till Roemer

Masterarbeit in Physik

vorgelegt der
Fakultät für Mathematik, Informatik und Naturwissenschaften
der RWTH Aachen University

im Januar 2017

erstellt im
III. Physikalischen Institut A
Prof. Dr. Thomas Hebbeker

Zweitgutachter
Prof. Dr. Martin Erdmann

Abstract

In 2015, the Large Hadron Collider resumed its physics program after more than two years of technical shutdown with an unprecedented center of mass energy of $\sqrt{s} = 13$ TeV. The Model Unspecific Search in CMS (MUSiC) performed in this thesis analyzes recorded data equivalent to 2.3 fb^{-1} of integrated luminosity in order to find physics beyond the Standard Model of particle physics (BSM).

A global search for deviations between the measured data and the standard model prediction is performed independently of BSM models. This approach is complementary to many direct searches which are generally optimized to maximize the sensitivity to one or a few specifically predicted signatures. In contrast, this analysis aims to be as model independent as possible in order to be able to find physics beyond the standard model which is not yet considered or even thought of by any theoretical model.

The analysis uses electron and muon triggered data and relies on Monte Carlo (MC) simulation for the standard model prediction. All data are sorted into event classes according to their final state and a p-value is used to quantify the agreement between data and MC simulation in the three kinematic distributions $\sum |p_T|$, M_{inv} and E_T^{miss} of every event class. The agreement is not only calculated for the entire distribution, but also for all regions which can be formed as subsets of the whole distribution. The local p-value is translated into the probability to find a deviation as large or larger as observed anywhere in the distribution using a look-elsewhere-effect correction.

Finally, the most significant deviations are discussed and the global agreement between data and MC simulation is investigated. No clear evidences of new physics were found.

Zusammenfassung

Der Large Hadron Collider begann im Jahr 2015 nach mehr als zwei Jahren technischer Pause mit Proton Kollisionen bei einer bis dahin unerreichten Schwerpunktsenergie von $\sqrt{s} = 13$ TeV. Diese Arbeit präsentiert die Modellunabhängige Suche in CMS (MUSiC) und analysiert Daten entsprechend einer Luminosität von 2.3 fb^{-1} zur Suche nach neuer Physik jenseits des Standardmodells der Teilchenphysik.

Dafür wird, unabhängig von theoretischen Modellen, global nach Abweichungen zwischen den gemessenen Daten und der theoretischen Vorhersage gesucht. Dieser Ansatz ist komplementär zu den meisten Analysen, welche ihre Sensitivität auf die Signaturen einzelner Modelle optimieren. MUSiC versucht so modellunabhängig wie möglich zu bleiben, um auch neue Physik finden zu können, die von keiner Theorie vorhergesagt wird.

Die Analyse benutzt Daten, die mindestens ein Elektron oder Myon enthalten, und basiert weitgehend auf Monte Carlo (MC) Simulationen für die Standardmodellvorhersage. Alle Daten werden abhängig vom Endzustand in Ereignisklassen sortiert und die Übereinstimmung zwischen Messung und Vorhersage wird mit Hilfe eines p-Wertes in den drei kinematischen Verteilungen $\sum |p_T|$, M_{inv} und E_T^{miss} quantifiziert. Es wird nicht nur nach Abweichungen in der ganzen Verteilung gesucht, sondern in allen zusammenhängenden Regionen die in der Verteilung gebildet werden können. Durch die Korrektur des "Look-elsewhere Effekts" wird ein globaler p-Wert berechnet, der die Wahrscheinlichkeit angibt, eine Abweichung zu finden, die mindestens so stark wie die gemessene ist.

Zusammenfassend werden die signifikantesten Abweichungen eingehender untersucht und die globale Übereinstimmung zwischen Messung und Vorhersage qualitativ diskutiert. Es wurden keine klaren Anzeichen neuer Physik gefunden.

Contents

1. Theory

The first section focuses on the theory of particle physics, specifically the Standard Model of particle physics (SM). The SM is an established and very precisely tested model [1,2] that explains interactions of all known particles. This chapter gives a summary about all known elementary particles and their interactions.

If not stated otherwise the description in this chapter is based on [3,4,5,6].

1.1. Units and Notation

This thesis uses mostly natural units in order to provide more accessible results compared to SI units since the energy regime is far below 1 J. The unit electron volt (eV) or for convenience GeV $= 10^9$ eV is used. In order to get rid of some constants the speed of light and $\hbar$ are set to one ($c = \hbar = 1$). This allows for example to provide masses or momenta in units of GeV as well. Charges will be provided, if not stated otherwise, in multiples of the elementary charge e.

Vectors are denoted with arrows (e.g. $\vec{v}$) and the implicit sum convention by Einstein is used implying the summation over terms with one upper and one lower index (e.g. $\vec{a} \cdot \vec{b} = a^i b_i$).

1.2. Particle Zoo

The Standard Model of particle physics describes the interactions[1] of twelve elementary half integer spin particles, called fermions, mediated by a set of gauge bosons with integer spin. Every particle is described by its mass and a set of quantum numbers containing the spin, weak isospin, electrical and color charge.

Fermions can be divided further in two groups: *Quarks* and *leptons*. The difference between them is the existence of a color charge for quarks. All fermions are sorted into three generations according to their mass as illustrated in figure 1.1. The up-type quarks have an electric charge of $+2/3$ while the down-type quarks have a charge of $-1/3$. The leptons are divided in electrically charged particles such as the electron and uncharged leptons called neutrinos. For each particle an antiparticle exists with the same features but flipped charges.

The four bosons with spin 1, namely the photon, the $W^\pm$ and Z bosons and eight gluons mediate the electromagnetic, weak and strong interaction respectively. The photon and the gluon are massless and electrically neutral but the latter possesses a color charge. The W and the Z boson are massive particles and the W is electrically charged.

---

[1] With the exception of gravity which is not contained in the SM.

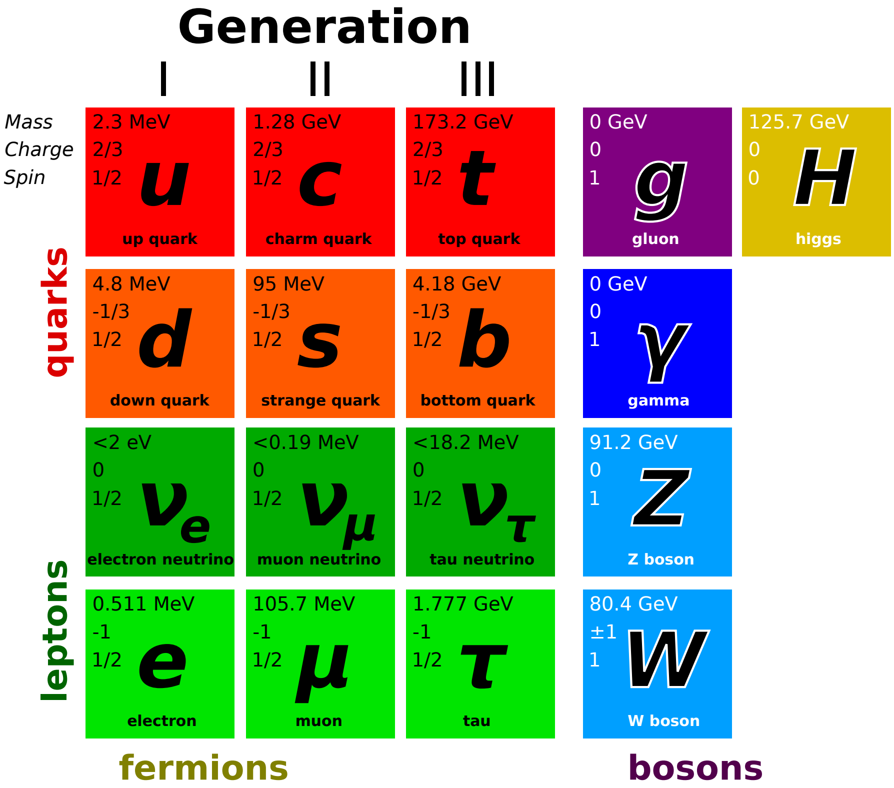


Figure 1.1.: Overview over all particles in the Standard Model. On the left hand side the fermions and on the right hand side the bosons are shown. The numbers in the upper left corner correspond to mass[2], electric charge and spin of the particle. Quarks are painted in red (orange) for the upper (lower) members of each generation. The leptons are divided in neutrinos painted in a dark green and charged leptons in a lighter green. The gluon as mediator particle of the strong interaction is painted in purple, the photon (Z- and W bosons) as mediator of the electroweak interaction in a dark (light) blue. In addition the Higgs is painted in yellow. All numbers taken from [6].

In 2012 a new heavy particle was discovered [7,8] which was identified as the Higgs boson with a mass of 125 GeV. It is a massive neutral scalar boson whose interactions are outlined later in section 1.5.1.

1.3. Interactions

The theoretical approach to explain the interactions between the previously introduced particles is the exploitation of features of the symmetry groups. The Standard Model can be represented in group theory as an $U(1)_Y \times SU_L(2) \times SU(3)$ group. The electromagnetic interaction is described by the $U(1)_Y$ symmetry group and requires its participants to carry an electromagnetic charge and is explained in detail in section 1.4. The $SU_L(2)$ symmetry group corresponds to the weak interaction (section 1.5) and the strong interaction is described by the $SU(3)$ symmetry group and is mediated by the gluon (section 1.6).

---

[2] For the u-, d- and s-quark the *current quark mass* is given and for the c- and b-quark the *running mass*.

In this section a closer look at the interactions between the particles is taken. In general the interaction of a non-relativistic free particle with a potential V can be described by solving the Schrödinger equation. This equation is not analytically solvable for the Standard Model potentials and perturbation theory is used to obtain an approximate solution for the transition of a particle from the initial state Φ_i to the final state Φ_f

$$W_{fi} = 2\pi \int \mathrm{d}^3x \Phi_f^*(\vec{x}) V(\vec{x}) \Phi_i(\vec{x}) \rho(E_i) \tag{1.1}$$

with the density of final states $\rho(E_f) = \rho(E_i)$ due to energy conservation.

This equation corresponds to the leading order approximation as the particle interacts only once with the potential and the precision can be improved by taking multiple interactions into account.

In order to describe the interaction of two particles we interpret the potential V as the potential produced by another particle. The potential itself varies for each type of interaction and can be derived from the corresponding Lagrangian. This procedure is outlined by an example in the next section.

1.4. Quantum Electro Dynamics

1.4.1. Example of Local Gauge Invariance

Quantum electro dynamics (QED) is the mathematically easiest interaction to describe. Starting point for the following calculation is the Lagrangian of a free massive fermion, the Dirac Lagrangian, which reproduces the Dirac equation by applying the Euler-Lagrangian-equations.

$$\mathcal{L} = i\bar{\Psi}\gamma^\mu \partial_\mu \Psi - m\bar{\Psi}\Psi, \tag{1.2}$$

with the wave function Ψ of the particle, its mass m and the *Gamma matrices* γ^μ.

In the next step, local gauge invariance is imposed by requiring that the addition of a local phase to the wave function $\Psi \rightarrow \Psi e^{i\Phi(x)}$ shall not change the Lagrangian. The phase may depend on the space coordinate x and does not commute with the derivative ∂_μ. Hence a term

$$\mathcal{L} = (i\bar{\Psi}\gamma^\mu \partial_\mu \Psi - m\bar{\Psi}\Psi) - (q\bar{\Psi}\gamma^\mu \Psi) A_\mu, \tag{1.3}$$

with the charge q and the newly added vector field A_μ has to be introduced which couples to the particle Ψ. However, a term describing the free vector particle has to be added as well due to the addition of the vector field A_μ. The Lagrangian of a free massive vector particle is given by the Proca-Lagrangian

$$\mathcal{L} = -\frac{1}{16\pi} F^{\mu\nu} F_{\mu\nu} + \frac{1}{8\pi} \left(m_A\right)^2 A^\nu A_\nu, \tag{1.4}$$

with the mass of the vector particle m_A and the electromagnetic field tensor $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$. The second term is not invariant under a local gauge transformation so the mass of the vector particle has to be equal to zero.

The resulting Lagrangian describes the propagation of free fermions and one massless boson corresponding to the photon and the interaction of the fermion with electric charge q to the photon:

$$\mathcal{L} = \underbrace{i\bar{\Psi}\gamma^\mu\partial_\mu\Psi - m\bar{\Psi}\Psi}_{\text{Propagation of leptons}} - \underbrace{\frac{1}{16\pi}F^{\mu\nu}F_{\mu\nu}}_{\text{Propagation of photon}} - \underbrace{(q\bar{\Psi}\gamma^\mu\Psi)A_\mu}_{\text{Interaction of lepton with photon}} \tag{1.5}$$

Similar descriptions can be found for the weak interaction (in conjunction with the electromagnetic interaction) and the strong interaction but are not further described here due to their increased complexity. Their features will be described in section 1.5 and 1.6.

1.4.2. Features of QED

Since the photon is massless, electrically neutral and stable, the electro magnetic interaction has infinite range and is therefore experienced in everyday life. Another consequence of the massless nature of the photon is the missing longitudinal spin orientation, $m_s = 0$. The photon can only couple to charged particles. Uncharged particles do not take part in electromagnetic interactions.

1.5. Weak Interaction

Contrary to the photon the bosons of the weak interaction, the $W^\pm$ and the Z bosons, are heavy with weights measured to 80.4 GeV and 91.2 GeV respectively [6]. They are unstable and decay quickly with a lifetime in the order of 10^{-25} s [6] which leads to a broad width.

The weak interaction is the only interaction that can change the flavor of particles. For example a top quark decays to a bottom quark by radiating a W boson[3].

In 1956 the Wu experiment [11] discovered that parity is not conserved in the weak interaction. The W boson couples only to fermions with left handed chirality (and anti-fermions with right handed chirality). The chirality is connected to the helicity which is defined as the direction of the spin orientation with respect to the direction of motion. In general chirality is not the same as helicity except for the case of massless particles or in the ultra-relativistic regime. Low energetic, massive particles can be described as superpositions of left-and right-handed helicity states since reference frames exist in which the direction of flight points to the opposite direction. However as neutrinos are massless, chirality is equal to helicity for them. Due to right-handed neutrinos not taking part in the weak interaction and the weak interaction being the only interaction neutrinos in general take part of, right handed neutrinos (and left-handed anti-neutrinos) do not interact with ordinary matter.[4]

---

[3] In case of the top quark all other decay chains (e.g. to strange or down quarks) are suppressed by the Cabbibo-Kobayashi-Maskawa (CKM) matrix [9, 10].

[4] The observation of neutrino oscillations [12, 13, 14] show that neutrinos are in fact not massless. However, in the SM neutrinos are considered massless. Even when considered massive the masses of neutrinos would be low compared to the charged leptons leaving the amount of right-handed neutrinos negligible compared to left-handed neutrinos.

Due to the high masses of the mediator bosons, the interaction is very short ranged and has typically lower cross sections than electromagnetic interactions at low energies compared to the boson masses. At ≈ 246 GeV the coupling strengths become comparably large because the coupling strengths depend on the energy scale [6][5].

The Glashow-Weinberg-Salam (GWS) model [15, 16, 17] unifies the electro-magnetic and the weak interaction to the electro-weak interaction. It predicts four massless fields (W^1, W^2, W^3, B) which mix to the physical bosons by

$$W^{\pm} = \sqrt{\frac{1}{2\pi}}\left(W^1 \mp iW^2\right) \tag{1.6}$$

$$Z = -B\sin\Theta_W + W^3\cos\Theta_W \tag{1.7}$$

$$A = B\cos\Theta_W + W^3\sin\Theta_W, \tag{1.8}$$

with the weak mixing angle Θ_W which is measured to $\sin^2\Theta_W = 0.23$ [6]. The masses of the bosons are not explained by the GWS model itself but in the Higgs mechanism.

The interaction of the bosons with the fermions are described by the Lagrangian

$$\mathcal{L} = \bar{L}\gamma^{\mu}\left(i\partial_{\mu} - g\frac{I_3^i}{2}W_{\mu}^i - g'\frac{Y}{2}B_{\mu}\right)L + \bar{R}\gamma^{\mu}\left(i\partial_{\mu} - g'\frac{Y}{2}B_{\mu}\right), \tag{1.9}$$

with the left-handed and right-handed fermions (L and R), the gamma-matrices γ^{μ}, the coupling constants g and g', the hypercharge (Y) and the third component of the weak isospin I_3 which are connected to the electrical charge Q via the Gell-Mann–Nishijima equation

$$Q = I_3 + \frac{1}{2}Y. \tag{1.10}$$

It can be seen in equation 1.9 that only the bosons mixing to the Z-boson or the photon couple to right-handed fermions while the bosons mixing to the $W^{\pm}$-boson do not couple to left-handed fermions.

1.5.1. Higgs Mechanism

The masses of $W^{\pm}$ and Z bosons cannot be explained by the classical Lagrangian of the electroweak interaction without breaking local gauge invariance.

This dilemma can be solved by introducing the Higgs field Φ [18, 19, 20, 21, 22, 23] with the Higgs potential of the form

$$V(\Phi) = +\frac{1}{2}\mu^2\Phi^2 + \frac{1}{4}\lambda\Phi^4, \tag{1.11}$$

with the real constant λ and the complex constant μ. With this additional field mass terms for the heavy gauge bosons can be introduced without breaking gauge symmetry. The potential has a maximum at the symmetry axis $\Phi = 0$ and a minimum at $\sqrt{-\mu^2/\lambda}$. This is the defining

---

[5] A short explanation why the coupling constants are energy dependent is given in section 1.7

difference compared to the other potentials in the Lagrangian. The potential is constructed to be not symmetric around the minimum and perturbation theory requires us to vary around a stable minimum. Therefore, the coordinate system has to be transformed to have its origin in the stable minimum. Performing a perturbative calculation in the transformed coordination system creates the mass terms for the massive bosons in the Lagrangian. This mechanism is called spontaneous symmetry breaking as the stable state is less symmetrical than the underlying potential. Combining this concept with local gauge invariance leads to mass terms for the heavy gauge bosons in the Lagrangian without breaking its invariance.

The fermions gain their mass through the Yukawa couplings to the Higgs boson. Measurements show good agreement to the prediction of a SM Higgs boson [24].

Another consequence of introducing the Higgs field is the existence of a Higgs boson which was detected in 2012 [7,8].

1.6. Quantum Chromo Dynamics

The strong interaction is described by the theory of Quantum Chromo Dynamics (QCD). Its mediators are eight massless gluons which couple to particles with color charge. The gluon also carries a color charge and therefore couples not only to quarks but also to itself in contrast to the photon in the QED. Due to multiple possible combinations of color charges, eight gluons carrying a color and an anti-color exist in total.

Each quark inherits one of the three different colors: red, blue and green[6]. A bound state of three quarks with different colors (baryon) or a quark and an anti-quark with matching color and anti-color (meson) are color neutral states. An externally color neutral particle does not participate in the strong interaction[7].

In contrast to the potential of QED an additional linear term is part of the potential of the strong interaction that leads to a constant force at high distances. If two colored particles (e.g. two quarks) with enough energy move away from each other they eventually form a new quark-anti-quark pair which will then bound to the initial quarks. This effect is known as *confinement*. When they get bound to each other they form two color neutral particles which are not bound anymore by the strong force and can move freely away from each other[8]. Especially at high energies this description is overly simplified thus a more realistic model is given in section 1.8.2.

Contrary, for small distances the coupling strength decreases leading to *asymptotic freedom* as the linear term does not dominate the potential.

---

[6] Each anti-quark has the color anti-red, anti-blue or anti-green.

[7] This only holds true if the energy with which the particle is probed is sufficiently low. If the particle is probed on a scale comparable to the size of the composite object the individual particles of the composite object, which are of course not color neutral, can be distinguished.

[8] They move only freely with respect to the strong interaction. They might be still effected by e.g. the electromagnetic force.

1.7. Feynman Graphs and Renormalization

An elegant way of displaying scattering processes are Feynman graphs. In figure 1.2a the Feynman graph of an interaction of an electron and a muon is shown as an example. With time ascending from left to right an incoming electron and an incoming muon interact via a photon which carries momentum from one particle to the other. The identity of the particles remains unchanged in the process but momentum is transfered.

This Feynman diagram corresponds to the first order approximation which is called the leading order or tree level. For a calculation of the total cross section one has to add additional vertices like in figure 1.2b or figure 1.2c as every possible interaction with the same final state has to be taken into account.

Technically each possible Feynman diagram represents one term in the series expansion around the minimum of the potential. Higher order corrections are suppressed as they include higher powers of the coupling strength (which needs to be small for perturbative calculations). Nevertheless, the convergence of this series is not given due to infinite contributions from loop corrections where arbitrary large momenta contribute. These effects can be avoided when the coupling strength is allowed to vary with the momentum scale of the interaction. For example in QED, the charge of the electron is screened by the polarization of the vacuum due to the loop corrections and the screening depends on the energy scale of the interaction. Hence the charge measured in experiments is not the bare charge of the electron but the screened charge. The bare charge of the electron is in fact not finite. This method is called *renormalization*. The coupling constant of the weak interaction behaves in a similar way and unifies with the electro-magnetic interaction above the electroweak scale.

As the series expansions is performed in the coupling constant the leading order approximation can only be justified if the coupling strength is much smaller than unity which is the case for electroweak interactions. In QCD the coupling strength differs with the momentum transfer as well but reaches values in the order of one for small momentum transfers. With higher order processes becoming even more important than the tree level diagram these processes in the low energy regime cannot be described by perturbation theory. For high momentum transfers the coupling strength decreases to values much smaller than one since not only virtual quark loops but also gluon loops are allowed due to their self coupling. The contribution of gluon loops counteract the contribution of fermion loops and dominate for QCD. This leads to asymptotic freedom as discussed in 1.6 allowing perturbative calculations in the high energy regime.

1.8. Monte Carlo Event Generation

In particle physics Monte Carlo event generators are often used to calculate the cross section of a process as well as the full event kinematics. They use pseudo random number generators to simulate events according to the theoretical predictions. The simulation of an event in a hadron collider can be split in several steps: The simulation of the hard scattering process, the parton shower of quarks and gluons, the hadronization of these and the detector simulation. An overview of the event generation can be found in figure 1.3.

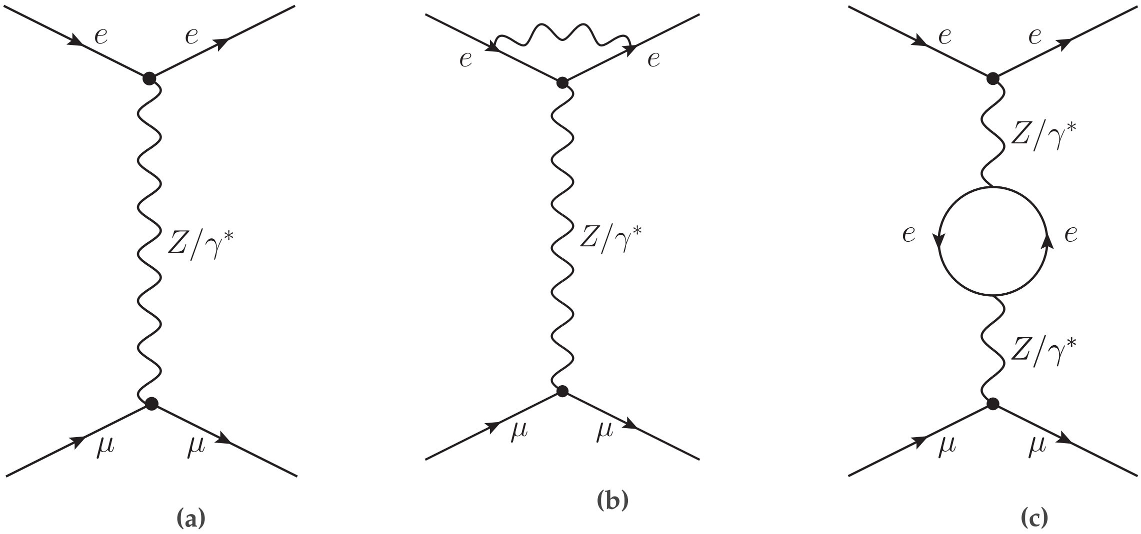


Figure 1.2.: Feynman graph of a scattering process of an electron and a muon via an exchange of a virtual photon or a Z boson. Time increases from left to right. In (a) the tree level diagram is shown, in (b) an example for a vertex correction, due to the emission of a photon and in (c) an example for a loop correction.

The event is calculated at leading order (LO) or next-to-leading order (NLO) precision taking additional vertices into account depending on the generator used. If the cross section of some samples is known by theoretical calculations with a higher precision than they are calculated by the MC generator, it can be corrected by a k-factor which is defined by

$$k = \frac{\sigma(\text{higher order})}{\sigma(\text{LO})}. \tag{1.12}$$

The same hard scattering process can lead to different jet multiplicities in the final state, e.g. due to initial state radiation. Some generators (e.g. MadGraph [26]) can simulate additional jets on matrix-element level while other generators (e.g. Pythia8 [27,28]) add additional jets posterior and rebalance the event in order to ensure energy and momentum conservation.

1.8.1. Parton Density Functions

Knowledge of the initial state of the collision is important to simulate the hard scattering process. Probing the structure of the proton e.g. with an electron[9] one can discover that a proton is a composite particle and consists of three valence quarks (two up quarks and one down quark). Secondly it was discovered that a significant amount of momentum is carried by electrical neutral particles, the gluons. With increasing momentum transfer a larger fraction of momentum is carried by sea quarks which can be produced in loops of the gluons (figure 1.4). In order to quantify the momentum each parton carries in the proton, structure functions $f(x, Q^2)$ are introduced which

---

[9] HERA measurements of electron-proton collisions combined with LHC data are the foundation of modern PDF sets like NNPDF [29].

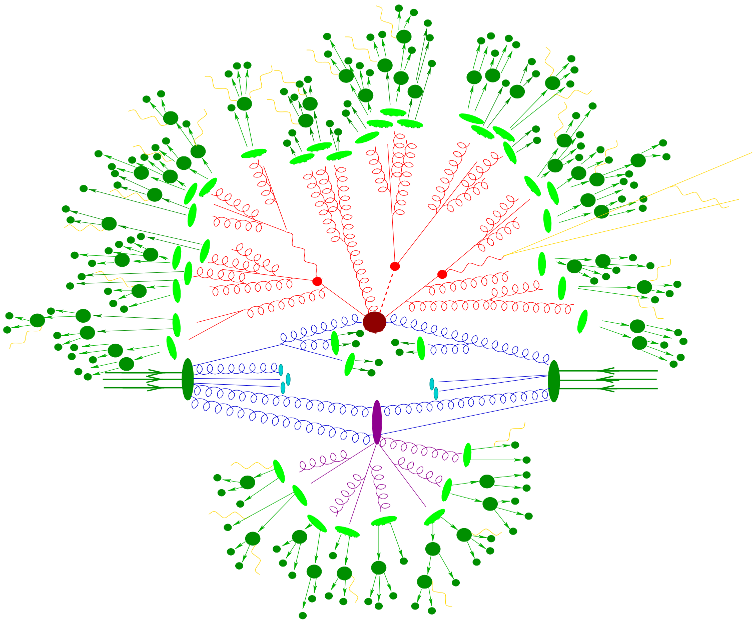


Figure 1.3.: Schematic view on the interaction of two protons as simulated by an event generator. The two incoming protons (large, dark green blobs on the left- and right hand side with three incoming lines denoting the valence quarks of a proton interact in the hard scattering process (large, red blob in the middle). The particles created by the hard scattering process further decay (small, red blobs) and the red lines show the evolution of the parton shower. At the hadronization scale the parton shower stops, the partons hadronize (light, green blobs) and decay in stable hadrons (small, dark green blobs). Photons may be radiated from any charged particle (yellow lines). This initial state parton shower is shown below the large, red blob. Other parts of the proton may interact as well (large, purple blob called underlying event). (Picture taken from [25].)

describe the probability to scatter of a quark or gluon inside the proton with the momentum fraction x carried by the parton at the energy scale Q^2.

The structure functions for all partons are summarized in the parton density functions (PDF). The precise knowledge of the PDFs are paramount for a reasonable Monte Carlo simulation of scattering processes. PDFs are measured mainly at lower energies than the LHC provides and have to be extrapolated using the DGLAP equations[10] [30, 31, 32], which describe the evolution of parton densities to higher energy scales.

With the PDF functions and the knowledge of the underlying Lagrangians of the process the cross section and kinematics of the hard scattering process can be simulated.

1.8.2. The Parton Shower

Quarks and gluons simulated in the hard scattering process do not transverse the detector as fundamental particles but shower (due to gluon radiations and pair production of quarks) and hadronize due to the nature of the strong interaction (see section 1.6). Gluons radiated from

---

[10] The equations were developed individually by Yuri Dokshitzer, Wladimir Naumowitsch Gribow and Lew Nikolajewitsch Lipatow as well as Guido Altarelli and Giorgio Parisi explaining the name of the equations.

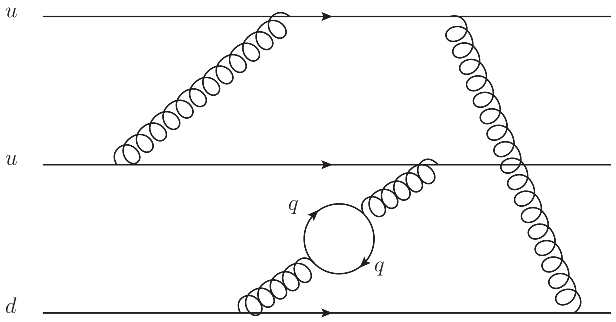


Figure 1.4.: Inner structure of a proton. One can see the three valence quarks u, u and d, which can exchange gluons and produce sea quarks denoted with q.

the initial state of the interaction also have to be showered. The evolution of a parton shower is described by the splitting functions which give the probability to split one parton in two partions (e.g. a quark radiating a gluon or a gluon producing a quark anti-quark pair) such as the structure functions described above [30,31,32].

Due to the boost of the initial particle created by the hard scattering process, the parton shower will be collimated in one direction.

If the hard scattering process and the parton shower are simulated by a different event generators jets produced by the parton showering are matched to additional generated jets on matrix-element level to avoid double counting [33].

1.8.3. Hadronization

The parton shower is stopped if the individual partons reach the hadronization scale. At this energy scale the partons form hadrons which is modeled by the string or the cluster model [34,35]. The created hadrons can subsequently decay to stable hadrons[11] which traverse the detector.

1.9. Beyond Standard Model

The Standard Model of particle physics works excellently and is measured up too a high precision by many experiments [1,2]. However, it fails to describe some phenomena:

- Dark matter is strongly motivated by astrophysical observations [36] but not included in the Standard Model.
- With the discovery of neutrino oscillations [12,13,14] it became clear that neutrinos are not massless as predicted by the Standard Model.
- The Standard Model yields no description of gravity although it is a clearly present interaction.

---

[11] Hadrons are considered as stable if their lifetime is long enough to traverse the detector without decaying further.

- It has no explanation for the large discrepancy between matter and antimatter in the universe [37, 38, 39].
- Fine-tuning of radiative corrections are necessary in order to explain the measured mass of the Higgs boson [40].

Many theoretical models exist which would solve some of these problems. As this analysis does not focus on a special theory no theory is discussed in detail here but the presence of physics beyond the Standard Model is strongly motivated.

2. Experimental Setup

2.1. Large Hadron Collider

The Large Hadron Collider (LHC) [42] is located near Geneva across the border between France and Switzerland. It is a circular collider stationed in the tunnel previously used for the Large Electron Positron Collider (LEP) [43] between 45 m and 170 m under ground with a circumference of 26.7 km. The LHC is accelerates protons or heavy ions. In case of proton acceleration the LHC is able to accelerate them from an energy of 450 GeV up to an energy of 7 TeV per proton. Pre-accelerators using smaller rings and linear accelerators at CERN accelerate the protons to the minimal energy threshold of the LHC. The main injector for the LHC is the Super Proton Synchrotron (SPS). The whole accelerator chain can be seen in figure 2.1.

Four major experiments are located at the LHC. A Large Ion Collider Experiment (ALICE) [44] concentrates, as its name suggests, on the analysis of heavy ion collision, which simulates the state of the universe a fraction of a second after the big bang. The Large Hadron Collider beauty (LHCb) [45] experiment is specialized to analyze decays of b-hadrons and analyzing the differences between matter and antimatter. The ATLAS [46] and Compact Muon Solenoid (CMS) [47] experiments are multi purpose particle detectors designed to be able to search for new physics at the highest energies.

2.2. Compact Muon Solenoid

The Compact Muon Solenoid (CMS) is a multi purpose particle detector located at the interaction point (IP) 5 of the LHC. It is designed to measure proton-proton collisions at a center of mass energy of $\sqrt{s} = 14$ TeV at a luminosity of $\mathcal{L} = 10^{34}\,\mathrm{cm}^{-1}\,\mathrm{s}^{-1}$ and lead lead collisions with an energy of 2.75 TeV per nucleon. Particles are not collided in a continuous stream but in discrete packages called *bunches*. To reach the design luminosity every 25 ns two bunches collide. Therefore, all detector components need to have a fast readout and a very short dead time. Due to the high rate of collisions, all detector components have to be resistant to radiation in order to not heavily decrease their performance over the operation of the LHC. Typically, more than two protons collide during the one bunch crossing with a triggered interaction and energy depositions in the calorimeters from secondary interactions or interactions in previous bunch crossings might be still be present in the readout of the detector. This leads to additional tracks and energy depositions called pile-up [48]. The description of the detector is based on [47] if not stated otherwise.

CMS is designed to achieve four major goals:

- A good muon reconstruction with a high momentum resolution up to muons with a momentum at the TeV scale.

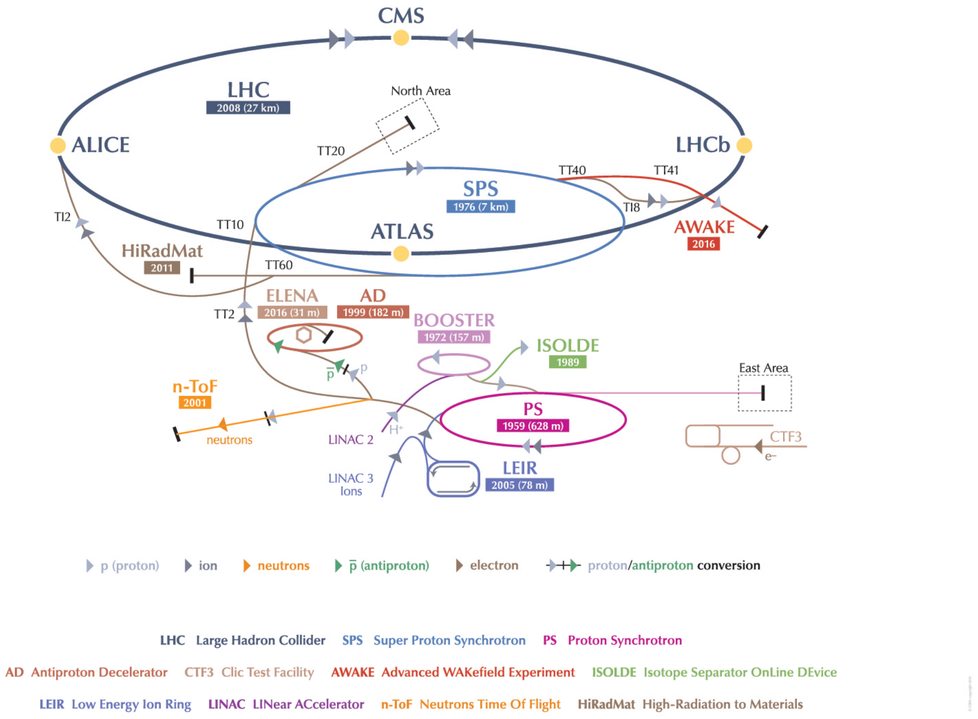


Figure 2.1.: Schematic view of the CERN accelerator complex with the LHC, its pre-accelerators (LINAC2, BOOSTER, PS and SPS) and other accelerators. Taken from [41]

- A high reconstruction efficiency in the tracker and the ability to tag b-jets and τ-leptons.
- A high and di-photon mass resolution and good lepton and photon isolation efficiencies.
- A precise measurement of the energy of jets and the missing transverse energy.

The detector can be divided into a cylindrical barrel region along the beam pipe with two endcaps. It consists of several layers with different purposes. An illustration with all subdetectors can be seen in figure 2.2.

2.2.1. Coordinate System and Kinematic Variables

In the coordinate system of the CMS detector (see figure) the z-axis points counterclockwise in direction of the beam pipe and the x-y-plane perpendicular to it with x pointing to the center of the LHC and y vertically upwards. Spherical coordinates are used to provide the position of particles. The CMS detector is symmetrical around the beam pipe and the azimuthal angle Φ is defined in the x-y-plane from the x-axis.

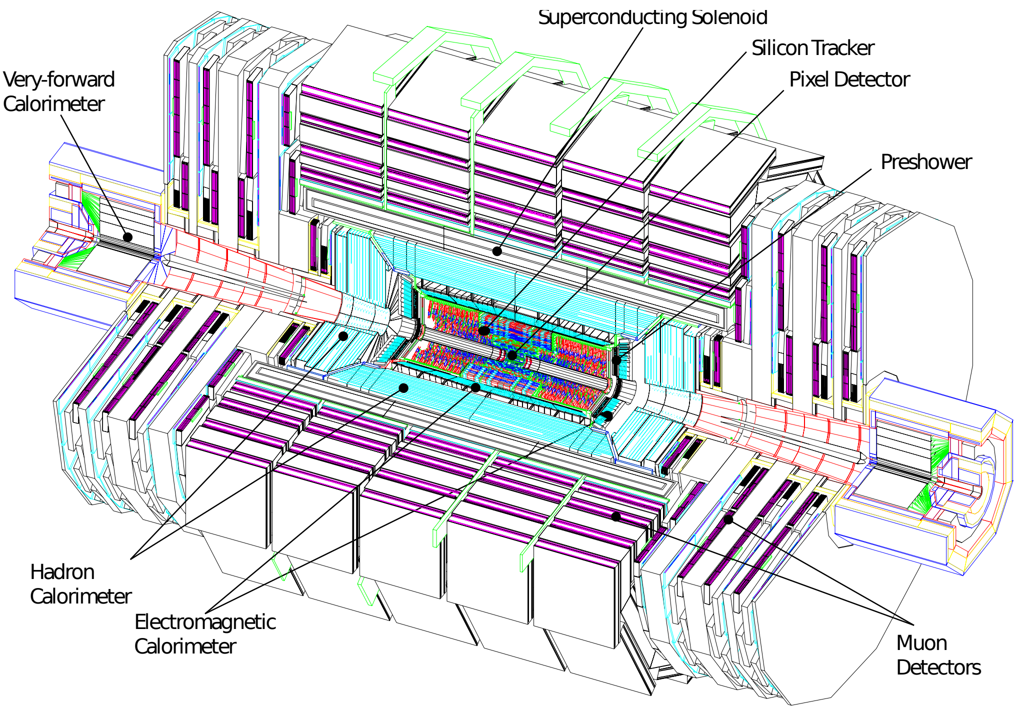


Figure 2.2.: Schematic view of the CMS detector with all subdetectors. Taken from [47].

Instead of the polar angle Θ the pseudorapidity

$$\eta = -\ln\left(\tan\left(\frac{\Theta}{2}\right)\right) \tag{2.1}$$

is used which has the advantage that pseudorapidity differences are invariant under Lorentz boosts.

The spatial distance of two particles used e.g. for the measurement of isolation of particles is defined as the distance in the $\eta - \Phi$ space as

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\Phi)^2}. \tag{2.2}$$

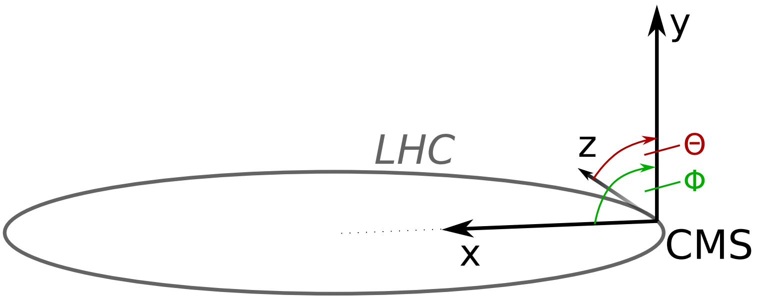


Figure 2.3.: Sketch explaining the coordinate system of the CMS detector. The x-axis points towards the center of the LHC ring and the y-axis vertically upwards.

The invariant mass of particles is given by

$$M_{\mathrm{inv}} = \sqrt{\left(\sum_{\text{particles}} E\right)^2 - \left|\sum_{\text{particles}} \vec{p}\right|^2} \,. \tag{2.3}$$

Due to the unknown energy of the interacting parton, only energy conservation in the transverse plane can be exploited under the assumption that the transverse momentum for the initial state particles is negligible. The properties of a particle perpendicular to the beam pipe are denoted by the subscript T, like the transverse momentum p_T. Some particles like neutrinos[1] can not be detected and might create missing transverse energy E_T^{miss}[2] in an event which is defined as

$$E_T^{\mathrm{miss}} = \left| - \sum_{\text{particles}} \vec{p_T} \right| . \tag{2.4}$$

If a decay involves undetected particles its invariant mass cannot be calculated since the longitudinal momentum of the missing energy is not known. In this cases the transverse mass is used

$$M_T = \sqrt{\left(\sum_{\text{particles}} E_T\right)^2 - \left|\sum_{\text{particles}} \vec{p_T}\right|^2} \,. \tag{2.5}$$

2.2.2. Magnet

The solenoid magnet creates a magnetic field of 3.8 T in the inner detector and is paramount for the transverse momentum measurement of tracks of charged particles. An iron return yoke surrounds the solenoid containing the magnetic field lines and serving as frame for the muon chambers.

2.2.3. Tracker

The innermost part of the detector consists of a silicon semiconductor based tracker. A schematic view of the tracker can be found in figure 2.4. Due to the large number of close by tracks a high spatial resolution is required to distinguish between individual particles. Therefore, the first three layers in the barrel region and two disks perpendicular to the beam pipe of the tracker are built out of silicon pixels (PIXEL in figure 2.4) to cover a pseudorapidity range of $|\eta| < 2.5$. This pixel detector reaches a spatial resolution of $15 - 20$ µm [47].

The rest of the tracker is built out of larger silicon strips. The inner part of the silicon strip detector consists four barrel layers (TIB) and three disks (TID) while the outer silicon strip tracker has

---

[1] Or some particles predicted by models beyond the Standard Model like dark matter particles

[2] The label *missing transverse momentum* would fit better to this definition. Historically the missing transverse energy was determined using the calorimeters which indeed measured the energy of particles and not the momentum. However, using the tracker information a higher precision can be reached.

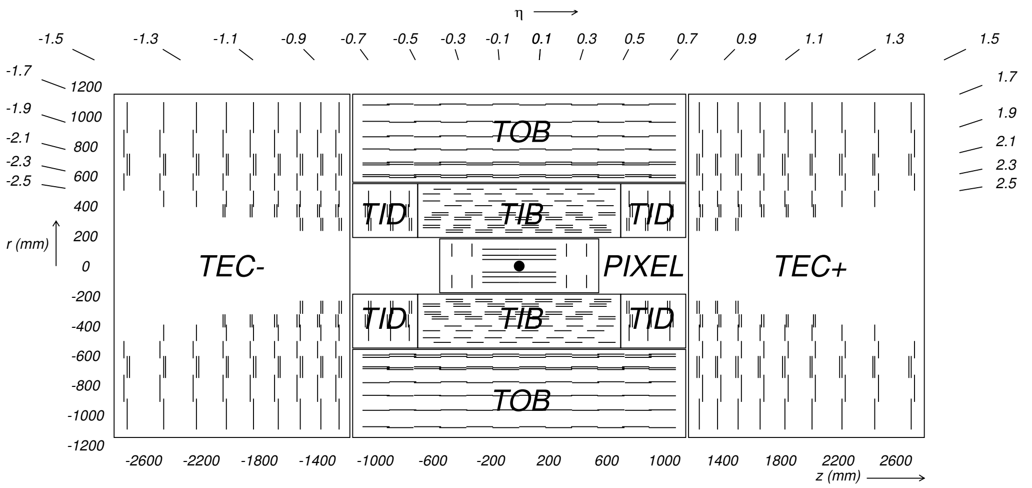


Figure 2.4.: Schematic view on the tracker in the $r - z$ plane. Taken from [47].

even larger strips with six layers in the barrel (TOB) and nine disks in the endcaps (TEC). The tracker covers a pseudorapidity range of $|\eta| < 2.4$ with all tracker layers and ultimately ends at a pseudorapidity of $|\eta| = 2.5$.

In conjunction with the magnetic field the tracker allows a precise determination of the transverse momentum of a charged particle with a resolution of about $1\,\% - 2\,\%$ in the central region $|\eta| < 1.6$ for track momenta around 100 GeV [47].

2.2.4. Electromagnetic Calorimeter

The tracker is surrounded by the electromagnetic calorimeter (ECAL). It is built out of lead tungstate crystals which have a short electromagnetic radiation length and scintillate themselves. In the barrel the length of the crystals corresponds to 25.9 electromagnetic radiation lengths while the crystals in the endcap are slightly shorter. The ECAL covers a pseudorapidity range of $|\eta| < 3.0$. However, it has less sensitive material in the overlap region of barrel and endcap, as can be seen in figure 2.5.

The ECAL is continuously tested by a laser monitoring system which measures the transparency of the crystals and corrects the energy calibration. The energy resolution of electromagnetic showers is energy dependent and can be parametrized as

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{2.8\,\%}{\sqrt{E/\,\mathrm{GeV}}}\right)^2 + \left(\frac{0.12}{E/\,\mathrm{GeV}}\right)^2 + (0.3\,\%)^2, \tag{2.6}$$

with the energy E [49,50]. The first term corresponds to stochastic uncertainties like fluctuations on event-to-event basis, the second to noise, e.g. electronic noise, and the constant term to calibration errors or leakage of energy if the shower is not completely contained inside the ECAL. Therefore, the energy resolution reaches $\sigma_E \approx 0.5$ GeV at the Higgs mass.

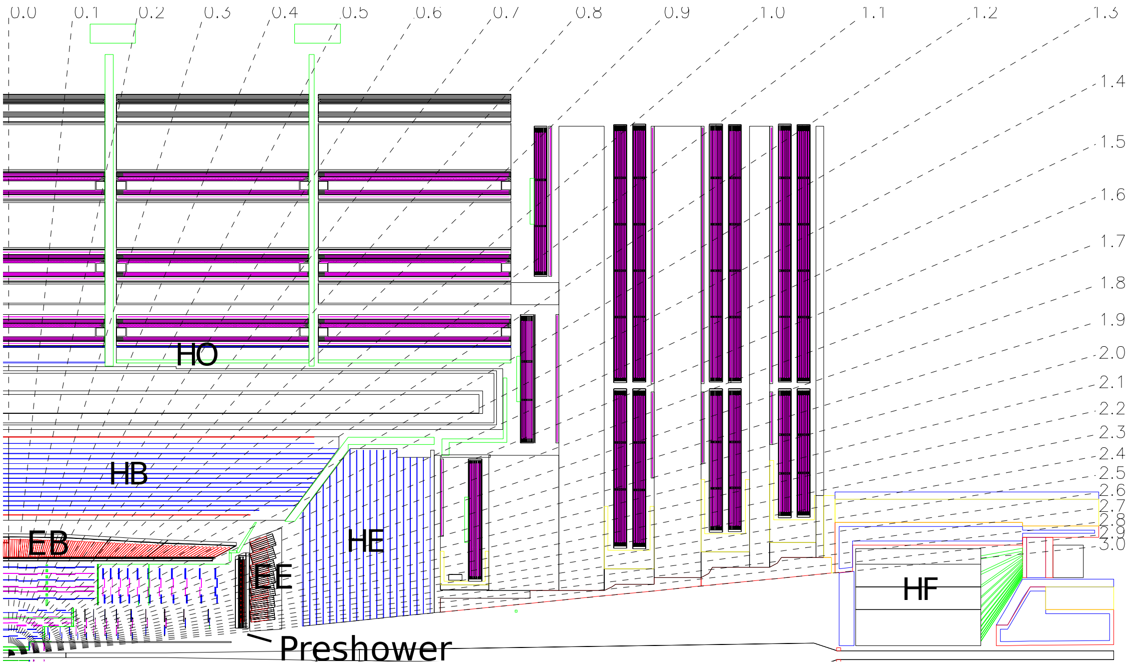


Figure 2.5.: Schematic view of a quarter of the CMS detector in the $r - z$ plane. The ECAL in the barrel (EB) and endcap (EE) can be seen in red. The parts of the HCAL are labeled with HB (HCAL Barrel), HE (endcap), HO (outer) and HF (forward). Taken from [47].

2.2.5. Hadronic Calorimeter

A lot of processes measured in a hadron collider have hadrons in their final state. The ECAL is too thin to contain the full shower of hadronic particles and a second calorimeter is required outside the ECAL. The hadronic calorimeter (HCAL) is also paramount for the determination of the energy of jets and the missing transverse energy. A schematic view can be found in figure 2.5.

The HCAL is a sampling calorimeter with brass as absorber and plastic scintillators as active material with embedded wavelength shifting fibers. Including the ECAL the total thickness of the endcaps is at about ten hadronic interaction lengths (λ_i). In order to increase the thickness of the HCAL and to catch the tails of particle cascades $1 - 2$ tiles are added outside the solenoid (hadron outer, HO) to increase the size to at least $11.8\lambda_i$. This results in an improved energy resolution for hadrons with low pseudorapidity.

To extend the pseudorapidity range of the HCAL beyond $|\eta| < 3.0$ the HCAL is supplemented by a forward calorimeter (HF). It is exposed to more radiation than the rest of the calorimeter and cannot use the same design. Therefore, a Cherenkov-light based approach is chosen. Quartz fibers are inserted into a steel block to collect Cherenkov light and guide it through wavelength shifting fibers to the readout. This forward HCAL extends the pseudorapidity range up to $|\eta| < 5.0$ which is important for the precise calculation of the E_T^{miss}.

The E_T^{miss} resolution was parametrized with the 2012 dataset at a center of mass energy of

$\sqrt{s} = 8$ TeV with

$$\sigma\left(E_T^{\text{miss}}\right) = 1.78 \text{ GeV} + 0.63 \text{ GeV} \cdot \sqrt{\sum p_T / \text{ GeV}} \tag{2.7}$$

with the sum of the transverse momentum of all reconstructed particles [51].

The resolution of jets reconstructed using the anti-k_t algorithm (see section 5.1.2) was calculated for the 2012 dataset as well. It is parametrized in the product of the jet area and the mean pile-up per event and can therefore be adapted to the 2015 data with a mean pile-up of 11.5 per event and a jet area of $\pi \cdot R^2$ with the distance parameter $R = 0.4$. The resolution [52] is given by

$$\sigma\left(p_T^{\text{jet}}\right) = \sqrt{(2.2 \text{ GeV})^2 + (0.045)^2 \text{ GeV} \cdot p_T + 0.88 \cdot p_T^2}. \tag{2.8}$$

More information about the reconstruction of jets and E_T^{miss} can be found in section 5.1. The reconstruction of both E_T^{miss} and jets is not only based on the HCAL but also on the ECAL and the tracker.

2.2.6. Muon System

The muon system is the outermost part of the detector and is placed in gaps inside the iron return yoke or outside of it. It consists out of three different detector parts: Drift tubes (DT), cathode strip chambers (CSC) and resistive plate chambers (RPC). This system is designed to measure high momentum muons with a good resolution and provide fast information for the trigger. In total it covers a pseudorapidity range up to $|\eta| < 2.4$.

Drift tubes are gaseous detectors and are used in the barrel due to the relatively low muon rate in the low pseudorapidity range and the homogeneous magnetic field. Due to the high muon rates in the endcap and the distorted magnetic field, it is impossible to use drift tubes since the drift time is much higher than the bunch spacing and multiple muons hitting one drift tube in a short time frame create unsolvable ambiguities. In addition the drift electrons would be deflected by the inhomogeneous magnet field in the endcap.

Therefore, CMS chose to use cathode strip chambers which are multi wire proportional chambers. The positions of the CSCs can be seen in figure 2.6. The principle of operation is the same as for drift tubes with the shortage that the drift time cannot be measured.

The resistive plate chambers are used in both barrel and endcaps. They are gaseous detectors as well and operate in the avalanche mode. When a particle traverses the RPC, gas gets ionized and the electrons create an avalanche of secondary electrons leading to a discharge which can be measured.

RPCs do not have the excellent spatial resolution of the DTs and CSCs but are much faster and help to assign tracks to a bunch crossing.

The momentum resolution was determined with the 2010 dataset at a center of mass energy of $\sqrt{s} = 7$ TeV [53] to

$$\sigma\left(p_T^{\mu}\right) = 0.016 \cdot p_T + 0.00015 \cdot p_T^2 / \text{ GeV}. \tag{2.9}$$

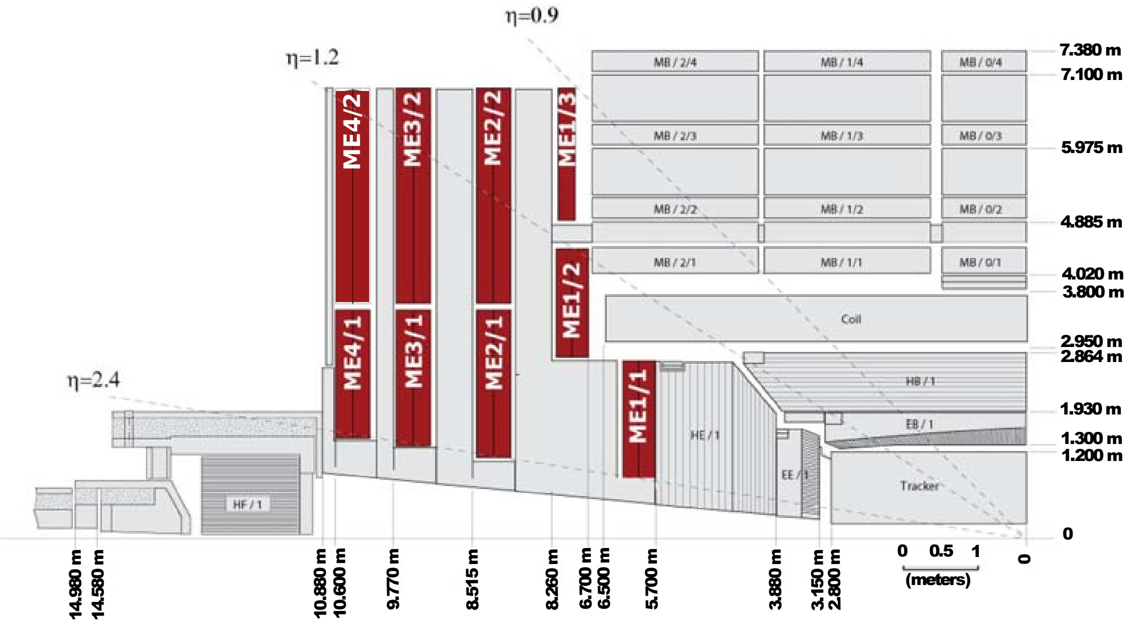


Figure 2.6.: Schematic view of a quarter of the CMS detector in the $r - z$ plane. The CSCs are highlighted in red.

This resolution is achieved by combining the reconstruction of the tracker and the muon chambers.

2.3. Trigger

During data taking CMS protons collide with a rate of 40 MHz and the amount of produced data exceeds the available resources for computing and storage. Triggers are used to limit the event rate from about 40 MHz to around 100 events per second. This is done by a two-step trigger system. The level one (L1) Trigger system operates on hardware basis and programmable electronics located on or near the detector and limits the event rate to a maximum of 100 kHz which is further reduced by the high level trigger (HLT) that runs on a dedicated computing farm of around one thousand commercial processors, has access to the full read-out data and runs an advanced event reconstruction close to the offline analysis.

The L1 Trigger uses input from the calorimeters and the muon system. It reconstructs e/γ candidates using deposited energy in the calorimeters and flags them with the measured transverse energy and a quality flag. The muon system reconstructs possible tracks and selects them based on the transverse momentum and the reconstruction quality requirements. The Global Muon trigger combines tracks from the different muon detectors and has also access to isolation in the calorimeter. In the end the global trigger can decide whether the event is discarded or transmitted to the HLT.

The exact requirements for all HLT triggers used in the thesis are described in section 5.2.1.

2.4. Computing

The CMS experiment uses a large network of computing sites called Worldwide LHC Computing GRID (WLCG) [54]. The computing system of the CMS experiment is distributed around the world and separated in different tiers. The Tier0 is located at CERN and Budapest and serves as the storage of the raw recorded data and the first reconstruction of it. Several Tier1 data centers are distributed around the world serving as backup storage, second-pass reconstruction, MC production and fast access to large data samples. In addition multiple in comparison small Tier2 data centers[3] exists which serve as a platform for second-pass reconstruction, Monte Carlo production but also to support local analyses.

The datasets are stored in different formats with different amounts of information. The RAW data tier for example contains the whole information of the detector response while the MiniAOD (Mini Analysis Object Data) format used in this analysis contains only high level physic objects.

---

[3] One Tier2 is located at the RWTH Aachen University.

3. Data and MC

3.1. Analyzed Data Sets

In this thesis the data taken at a center of mass energy of $\sqrt{s} = 13$ TeV in the data taking period `2015D` with 25 ns bunch spacing are analyzed. The data are split in several runs and further in subsections called luminosity sections. Each section is centrally certified to ensure full availability of the detector and well reconstructed objects. The validated dataset yields an integrated luminosity of 2.3 fb^{-1} which was determined with the pixel cluster counting method [55].

Only events triggered by at least one lepton are used in this analysis. To maximize the event yield not only single lepton triggers, but also double lepton triggers are used, as the double lepton triggers have a lower p_T threshold. A combination of single and double lepton triggers are used to maximize the total number of analyzed events and to lower the p_T thresholds for some final states. The triggers will be described in more detail in section 5.2.1. Detailed information about the used datasets and the validation of the data can be found in appendix D.1.

3.2. Standard Model Simulation

The MUSiC analysis aims to be as model independent as possible and considers all regions under investigation as potential signal regions. This approach restricts the use of data-driven description of SM expectations and the analysis fully relies on Monte Carlo simulations. The functionality of MC event generators is described in section 1.8.

This SM expectation is simulated by MadGraph [26] and Pythia8 [27,28] at LO precision and Sherpa [25], Powheg [56], MadGraph5_aMC@NLO [57] and MCFM [58] at NLO precision[1] and is corrected by k-factors if theoretical calculations have a higher precision than the cross section calculated by the MC generator.

MadGraph, Powheg, MadGraph5_aMC@NLO and MCFM do not apply a parton shower simulation themselves but use the parton shower provided by Pythia8.

As the cross section of most processes declines rapidly to very high energies, some samples have dedicated extensions to enrich the number of generated events at high energies. Overlaps between different samples are removed in order to avoid double counting of events.

The detector response to the generated particles is simulated with GEANT4 [59].

---

[1] Sherpa can produce LO and NLO samples. In the samples used in this analysis, the most important NLO contributions are taken into account and the samples are treated as NLO simulation.

The complete set of Standard Model simulations can be found in table D.2 in the appendix D.2. Additional remarks for some samples are given in the following.

3.2.1. W+Jets

The single W boson production is one of the most important Standard Model processes for the MUSiC analysis with a very high production cross section and prompt leptons in 30 % of the branching ratio of the W boson decay which are important to trigger the event. This process is modeled by MadGraph and Pythia8 at LO precision. Due to the importance of the process, constant k-factors are applied for all samples improving the precision to (N)NLO. In order to have a sufficiently smooth MC simulation for high jet multiplicities and energies, samples binned in the variable HT are used. This is the sum of all transverse momentum of non-leptonic particles (i.e. jets and photons) on generator level (i.e. before the detector simulation). A k-factor is only available for the fully inclusive sample. The same k-factor is used for the HT binned samples as the underlying process for these is the same as for the fully inclusive sample. However as the phase space is restricted for the HT binned samples, the uncertainty is increased compared to usual NLO calculations to 30 % from 10 %. A variable k-factor dependent on HT would neglect other dependencies of the k-factor like the dependence on the jet multiplicity which showed an even larger effect.

3.2.2. Drell-Yan

The Drell-Yan process is the production via quark annihilation of a Z/γ^* and its subsequent decay into two charged leptons and is an important Standard Model process as well for the same reasons as for the W boson production. It is simulated by MadGraph and Powheg. The Powheg samples have the advantage of a NLO cross section calculation, but lack the contribution of decays into taus. Therefore the MadGraph sample is used up to a mass of $M_{ll} < 120$ GeV in order to include the contribution of decays into taus in the peak region around the mass of the Z boson. At higher masses the contribution of decays into taus is missing, but since the hadronic decay of taus is dominant the effect is small.

3.2.3. QCD Multijet

The production of multiple jets by strong interactions is simulated with Pythia8. As only lepton triggers are used and the QCD process has no inherent isolated leptons in its final state, it can only contribute by misidentified jets or leptons from jets which are not filtered by the isolation criteria. As the misidentification rates are typically low, most events of a normal QCD sample would be discarded. To enhance the number of events passing the trigger and selection criteria special electromagnetic and muon enriched QCD samples are simulated. During generation, these events have to pass a filter to increase the fraction of muons for the muon enriched sample and the fraction of electrons or photons in the electromagnetic enriched sample. Using these filtered samples allows to have a sufficiently smooth description without having to simulate an enormous amount of events.[2]

The weight factor (see section 3.3) of simulated QCD events is typically very high (particular greater than one) as the number of simulated events is smaller than the number of expected events for this process according to its cross section and luminosity. The calculation of event weights is described in more detail in section 3.3.

3.2.4. Top production

The energy of the LHC is well above the mass of the top quark and the production of single top quarks and $t\bar{t}$ production is dominant in many final states, especially at high particle multiplicities. A top quark decays via W bosons which can decay to leptons. This is important for the MUSiC analysis since it uses only lepton triggers.

3.2.5. Multi-boson

Multi-boson production (e.g. WW, $W\gamma$ or ZZ) have inherent lepton decays, but typically a lower production cross section compared to the previously mentioned processes. It is still important enough to be dominant in some final states with a typically high lepton multiplicity but low jet multiplicity, as the $t\bar{t}$ production has typically a high jet multiplicity and therefore does not contribute primarily in those classes.

3.2.6. Rare processes

There are more simulated Standard Model processes, which are subdominant in most of the final states as either the production cross section is low, or they do not yield prompt leptons to trigger the event. These processes are not described in detail in this section but are listed in the table D.2.

3.3. Weights of Monte Carlo Simulation

Simulated MC events have to be weighted according to their cross section, the luminosity of the data and the pile-up distribution of data to simulate the real conditions in the experiment. The weights can be factorized in three different types: The luminosity weight, the pile-up weight and the generator weight.

---

[2] In principle the same amount of events have to be simulated using the filter compared to not using the filter as events without leptons are just discarded and not inherently included in the simulation in a way that only events with sufficient leptons are simulated. However, the most time consuming part (especially for LO generators) is the detector simulation and not the event simulation itself and this step is skipped for events which do not pass the filter.

The luminosity weight scales the MC events to the luminosity, the cross section of the process and, if present, additional k-factors for the cross section (see 3.2) and is calculated by

$$w_{\text{luminosity}} = \frac{N_{MC}}{L \cdot \sigma \cdot k}, \tag{3.1}$$

with the number of simulated events N_{MC}, the integrated luminosity L and the cross section σ.

The MC events are simulated with an approximate pile-up scenario and are reweighted to the actual pile-up distribution measured in data by

$$w_{\text{pile-up}} = \frac{w_{vtx}^{\text{data}}}{w_{vtx}^{\text{MC}}}, \tag{3.2}$$

with the weight of the pileup distribution w_{vtx} for the number of simulated primary vertices in data and MC simulation.

Generator weights w_{gen} are used by some generators to correctly model the cross section of the process e.g. negative contributions in NLO calculations of matrix elements [57].

The total weight is calculated by

$$w_{\text{tot}} = w_{\text{luminosity}} \cdot w_{\text{pile-up}} \cdot w_{\text{gen}}. \tag{3.3}$$

and are applied to the MC events.

4. Model Unspecific Search in CMS

4.1. Motivation and General Concept

Most searches for beyond the SM (BSM) physics study if theoretical predicted effects can be measured in data. The Model Unspecific Search in CMS (MUSiC) [60,61,62,63,64,65,66,67] follows a complementary approach and does not optimize the analysis to be sensitive to one specific model, but tries to be as general as possible and searches for deviations from the Standard Model expectation independent of theoretical predictions for BSM physics.

First, events are selected according to the used triggers (see section 5.2.1). Next in every selected event, reconstructed objects are identified as explained in section 5.3. This analysis considers electrons, muons, photons, jets and E_T^{miss} as objects. The general workflow starting after event and object selection is illustrated in figure 4.1 and is described in detail in the following.

4.2. Classification

We start selecting all events which pass the trigger criteria (see section 5.2.1) and special event filters used to veto badly reconstructed events (see section 5.2.2). Every selected event is sorted into *event classes* depending on the reconstructed objects in the event. There are three kinds of event classes:

- Exclusive event classes are defined by the exact number of objects in the event.
- Jet-inclusive event classes are defined by the exact number of electrons, muons, photons, jets and if present E_T^{miss} stated in the event classes names plus any number of additional jets.
- (Fully) inclusive event classes contain the exact number of objects stated in its name and may incorporate any number and kind of additional objects.

An example for the classification of a single event is shown in figure 4.2.

In every event class (independent of their kind) up to three binned distributions are filled: The invariant mass of the full event, based on all particles in the event class (M_{inv}), the sum of all transverse momenta ($\sum |p_T|$) and the missing transverse energy (E_T^{miss}) in the event if present in the event class[1]. In case missing transverse energy is present in the event class the transverse mass (M_T) is calculated instead of the invariant mass since the longitudinal component of the missing energy is unknown.

---

[1] Some missing transverse energy will be present in every event due to detector resolution effects, pile-up etc. Therefore it has to be greater than a threshold of $E_T^{\text{miss}} > 100$ GeV to be considered as described in section 5.3.

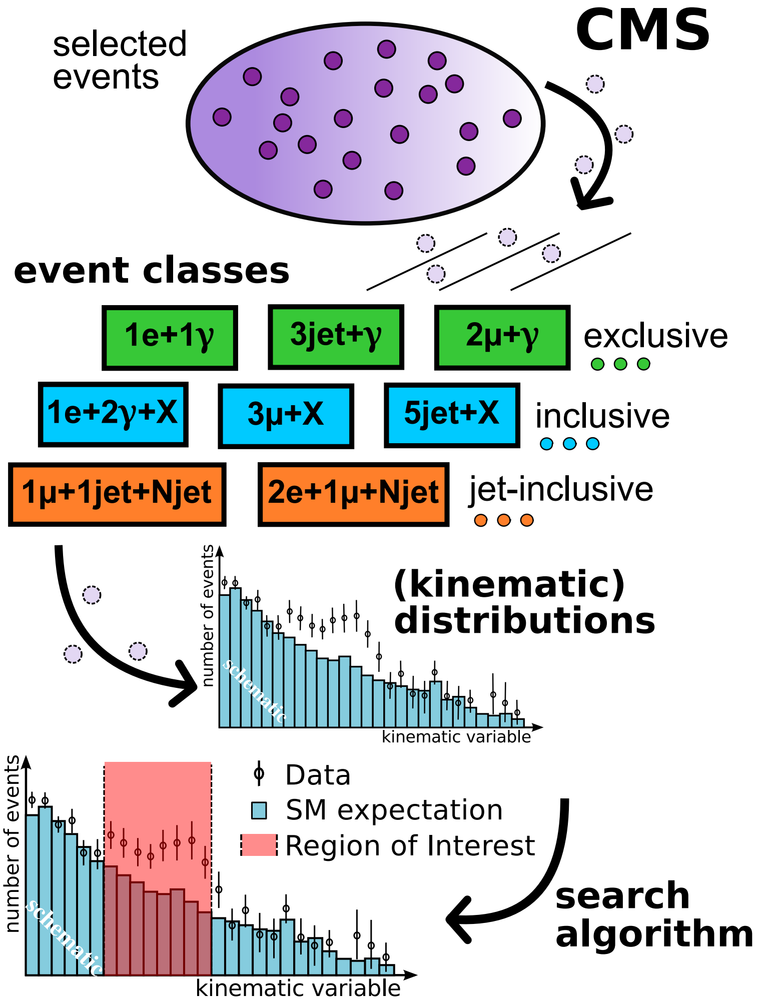


Figure 4.1.: Schematic view of the workflow of the MUSiC analysis. After the event and object selection, all events are classified in event classes depending on their final state. In each event class up to three distributions are filled and automatically scanned for deviations between Standard Model expectations and data. (Picture taken from [66].)

It is important to note that for any inclusive or jet-inclusive event class only the mass or $\sum |p_T|$ of the objects stated explicitly in the name of the event class are calculated and additional objects are ignored in the calculation. For example an event with one muon, two electrons and one jet is sorted amongst others in the $1e$ 1μ inclusive event class. The invariant mass (or $\sum |p_T|$) is then calculated only from the electron and the muon. If more objects of one kind are present in an inclusive event class than stated in its name the object with the highest p_T is used to calculate the kinematic properties. In this example, the muon and the electron with the highest p_T would be used.

Both data and simulated events are classified in order to scan the event classes for deviations as explained in the next section.

The binning of the distribution is designed to reflect the resolution of the objects in the event with a minimal bin width of 10 GeV. The energy dependent resolution of the event class is calculated

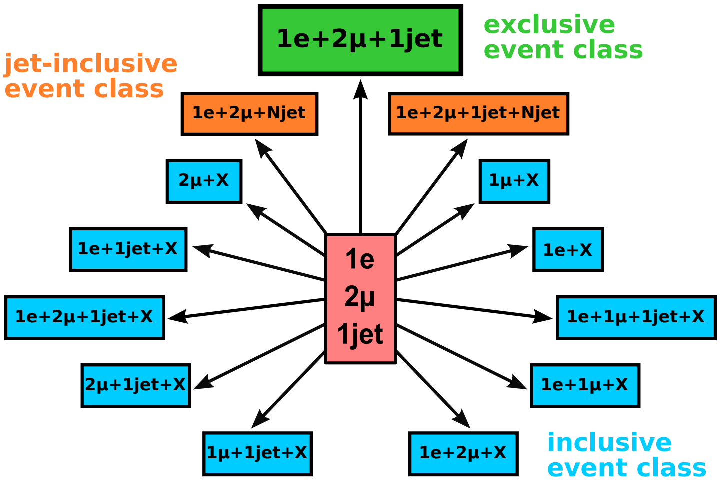


Figure 4.2.: Example for the classification of an event in exclusive, jet inclusive and fully inclusive event classes. The event is sorted in the exclusive event class defined by the exact final state of the event, inclusive event classes that contain at least one of the objects in the final state and jet-inclusive event classes which contain the exact number of leptons, photons and E_T^{miss}. (Picture taken from [66].)

under the assumption that the energy is distributed equally among all particles. The resolutions for each particle is stated in chapter 2.2. For electrons and photons the resolution of the ECAL is used.

4.3. Scanning

An automated scanning algorithm was developed to analyze all event classes and search systematically for deviations between measured data and the Standard Model expectation. As different theoretical models predict different signal shapes, e.g. narrow resonances[2] or broad excesses[3], every possible combination of connected bins is investigated. For the $\sum |p_T|$ and E_T^{miss} distributions a minimal region width of three bins is used as we do not expect narrow resonances in the $\sum |p_T|$ or E_T^{miss} spectrum and adding these regions would enlarge the look-elsewhere effect (see section 4.3.3) without providing sensitivity to new physics. As resonances in the invariant mass spectrum can be as narrow as the bin width of a single bin, the region may contain only one or two bins.

4.3.1. Handling of Regions with Insufficient Number of Monte Carlo Events

As the Standard Model expectation is simulated with a finite number of events it is almost inevitable that the number of simulated events of the MC simulation may not be sufficient to calculate a reasonable Standard Model expectation [66]. Examples for such regions are the high energy tails of distributions or event classes with a high number of objects.

---

[2] For example in case of a heavy partner of the Z boson [68].

[3] ADD models can predict such broad signatures [69].

Every region with a relative statistical uncertainty larger than 60 % is vetoed [66]. This ensures that processes with high weights (compared to other processes in the region) have enough simulated events in order to trust the MC simulation as an appropriate SM prediction. It is important to note that the region but not the bins contained in the region in general are vetoed. A wider region containing the bin with lacking MC statistics and additional bins with more simulated events might pass this criterion as can be seen in figure 4.3a.

Even if the region has enough simulated MC events to pass this criterion it may not describe the Standard Model appropriately, if it is missing contributions from dominating processes in that energy regime entirely. The dominant processes are determined using the neighborhood of a region. All processes which contribute to 95 %[4] of the total yield in the neighborhood of the region have to be present in the region under investigation. The neighborhood is defined as the four bins below and the four bins above the region.[5] The neighborhoods are not treated separately but added up together. A schematic example can be found in figure 4.3b.

The last requirement arises from the method some NLO generators simulate MC events. The NLO cross section can be calculated by generating events with positive and negative weights. Usually events with positive weights lead to an overall positive contribution per bin but in some regions especially in the high energy tails of the distribution, regions with enough simulated MC events to pass the first requirement might be found with a negative total event yield. As a negative expectation is clearly unphysical and the negative weights are produced only to balance the positive weighted events to simulate the correct NLO shape, these regions are vetoed as well.

4.3.2. The p-value Calculation

For every region passing the region veto a local p-value is calculated, which gives the probability to observe deviation at least as strong as examined. The p-value is given by

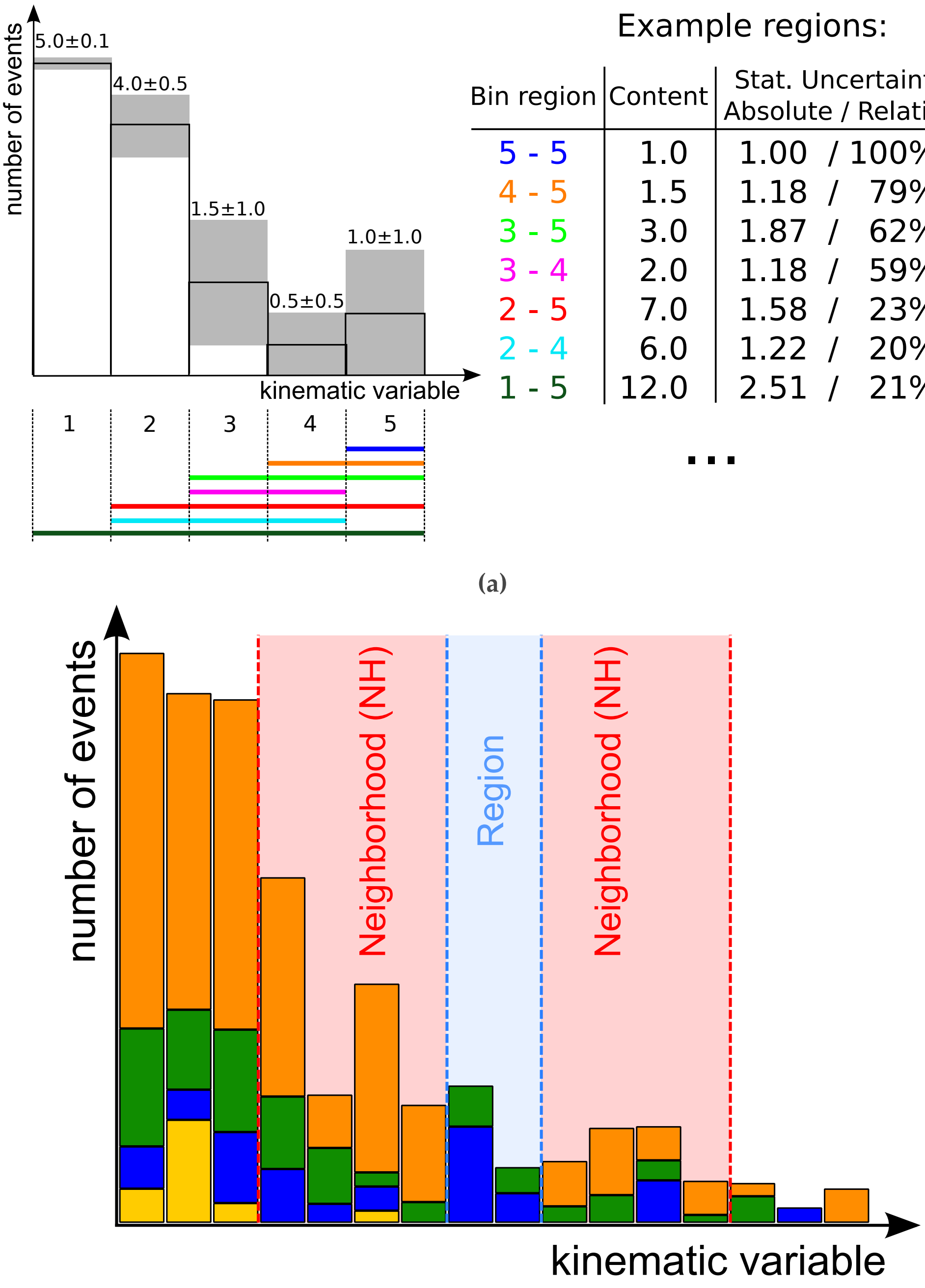
$$p = \begin{cases} \sum\limits_{i=N_{\text{data}}}^{\infty} C \cdot \int\limits_{0}^{\infty} \mathrm{d}\lambda \ \exp\left(-\frac{(\lambda - N_{\text{SM}})^2}{2\sigma_{\text{SM}}^2}\right) \cdot \frac{\mathrm{e}^{-\lambda}\lambda^i}{i!} & \text{if } N_{\text{data}} \geq N_{\text{SM}} \\ \underbrace{\sum\limits_{i=0}^{N_{\text{data}}} C \cdot \int\limits_{0}^{\infty}}_{\text{Normalization}} \mathrm{d}\lambda \ \underbrace{\exp\left(-\frac{(\lambda - N_{\text{SM}})^2}{2\sigma_{\text{SM}}^2}\right)}_{\text{Systematics}} \cdot \underbrace{\frac{\mathrm{e}^{-\lambda}\lambda^i}{i!}}_{\text{Statistics}} & \text{if } N_{\text{data}} < N_{\text{SM}}, \end{cases} \tag{4.1}$$

where N_{data} (N_{SM}) is the number of data (MC) events in the region and σ_{SM} the total uncertainty of the Standard Model simulation. The total uncertainty is calculated adding up uncertainties of each simulated Standard Model process taking correlation effects between different MC simulations into account. The uncertainties and their correlations are described in detail in 5.4. The statistical uncertainty of the MC prediction is taken into account as a systematic uncertainty for the p-value calculation as well.

---

[4] taking the weights of the simulated events into account

[5] If less than four bins neighboring the region are present the algorithm takes only these into account without adding the missing numbers of bins on the other side. This happens at the edges of the distribution.



Example regions:

Bin region	Content	Stat. Uncertainty Absolute / Relative
5 - 5	1.0	1.00 / 100% x
4 - 5	1.5	1.18 / 79% x
3 - 5	3.0	1.87 / 62% x
3 - 4	2.0	1.18 / 59% ✓
2 - 5	7.0	1.58 / 23% ✓
2 - 4	6.0	1.22 / 20% ✓
1 - 5	12.0	2.51 / 21% ✓

(a)

(b)

Figure 4.3.: Schematics explaining the region veto algorithm used to handle regions with insufficient MC simulation. (a) shows the threshold of maximal statistical uncertainty per region and (b) illustrates the building of neighborhood regions. (Pictures taken from [66].)

The p-value represents the probability of measuring as many or more (less) events than observed in data in case of a excess (deficit) with respect to the MC prediction. It is calculated by the sum over a Poissonian distribution with the expectation value of the MC prediction. To account for the systematic uncertainties of the Standard Model expectation, the Poissonian distribution is convoluted with a Gaussian with the width of the total systematical uncertainty in the region. As there cannot be a negative Standard Model expectation the Gaussian is truncated at zero, leading to a normalization factor C to correct for this truncation. Since the p-value describes the probability to find a deviation at least as strong as the observed (corresponding to N_{data}), the probability distributions are summed up from N_{data} to infinity in case of an excess or from zero to the number of data events in case of a deficit.

Boundaries of Validity of p-value

For some combination of data, MC simulation and total uncertainty the p-value calculation might overestimate the significance of deviations. For example in case of zero data events and a SM expectation close to zero (i.e. much smaller than one) and a total uncertainty which is order of magnitudes larger than the SM expectation value, the deviation is definitely not significant, as the SM expectation is compatible with the measured data events even with a fraction of 1σ but the p-value indicates a significant deficit. A parameter scan for various number of SM simulation events N_{SM} and the total absolute uncertainty σ_{abs} with zero observed data can be found in figure 4.4. All p-values above the $N_{\text{SM}} = \sigma_{\text{abs}}$ line should be insignificant as data and MC expectation are compatible within 1σ but for $\sigma_{\text{abs}} > 1$ the p-values indicate significant deviations.

This behaviour is not a malfunction of the p-value itself but the usual interpretation of the p-value is not valid in this region of the parameter space. In case of a deficit the p-value corresponds to the probability to find less or equal events than measured with the given MC expectation. If the uncertainty is much larger than the number of measured and expected events it is improbable to find less or equal events than measured in data. For example if no events were measured, close to zero events are expected and the uncertainty is 100 events the probability to find exactly zero events is relatively small. This, however, is not the desired interpretation of the p-value as all deficits covered by the uncertainty of the MC simulation show no signs of new physics.

Therefore all regions with a total uncertainty five times as large as the number of simulated MC events are vetoed.

4.3.3. Look-Elsewhere Correction

The p-value provides the probability to find a deviation as large or larger as observed *in the local region*. However, the probability to find such a deviation *in any region* of the whole distribution is of greater interest. This so called *look-elsewhere effect* is taken into account by dicing pseudo experiments as the correlations between regions prohibit an analytic calculation.

In the pseudo experiments data are diced according to the SM only hypothesis. First, a new expectation value is diced according to a Gaussian modeling the systematic uncertainties. The total uncertainty is calculated as described in section 4.3.2, and their sources and correlations are described in section 5.4. The shifts due to the systematic uncertainties are treated as fully

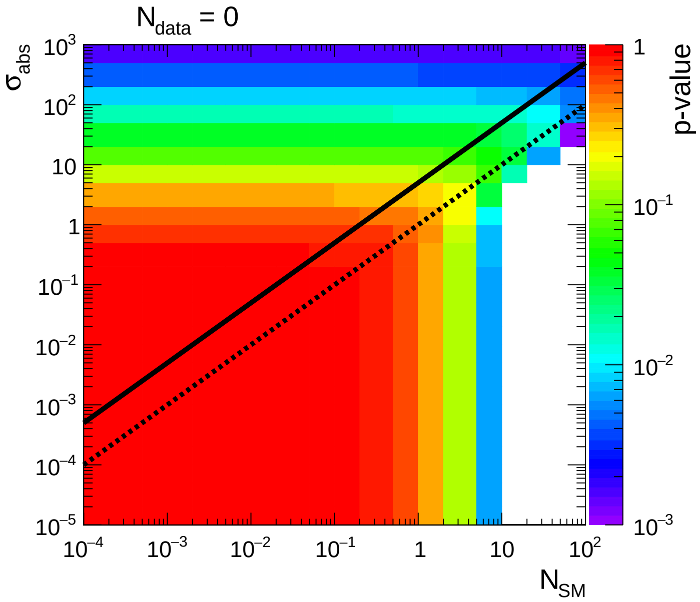


Figure 4.4.: Parameter scan for $N_{\text{data}} = 0$ and several values for the total absolute uncertainty σ_{abs} and number of MC events N_{MC}. The p-values are depicted by the color code with large (insignificant) p-values in red and small (significant) p-values in blue. The white areas have p-values below the minimum of the scale. The small p-values in the lower right corner of the histogram are expected as regions with no data events, a large MC expectation and a low uncertainty show a clear deviation. However, all p-values above the $N_{\text{SM}} = \sigma_{\text{abs}}$ line (black dashed line) should be insignificant as data and MC expectation are compatible within 1σ. This is not the case for uncertainties greater than one. The solid black line denotes the requirement of $\sigma_{\text{rel}} < 5$ which is chosen in order to not lose sensitivity.

correlated over different event classes per dicing round[6]. The pseudo data point is then calculated by dicing a Poissonian distribution with the new expectation value.

The $\tilde{p}$-value for a given distribution is calculated by the fraction of p-values calculated in the pseudo data which are lower (i.e. more significant) than the p-value calculated with the real data for this distribution and all diced p-values as visualized in figure 4.5.

Additional look-elsewhere effects have to be considered to interpret the final significance of a deviation: One introduced by the number of event classes (see section 4.3.4) and one due to the different kinematic variables and event class types (i.e. exclusive, jet-inclusive and inclusive) as they are considered separately.

4.3.4. Global $\tilde{p}$ Distribution

The need of calculating a $\tilde{p}$-value arises from the amount of different regions under investigation. The $\tilde{p}$-value is able to provide a measure of how significant a deviation is in the whole event class and not one region. However, it does not provide the probability to find a deviation as large as

---

[6] This means that all values for one kind of systematic are shifted in the same direction and by the same (relative) amount in every bin, distribution and event class. For example, if the energy scale of electrons (see section 5.4.6) would be systematically 5 % too low all electrons in all event classes would be affected in the same way. Another example would be an incorrect determination of the luminosity. For different dicing rounds the shifts may (and should) differ.

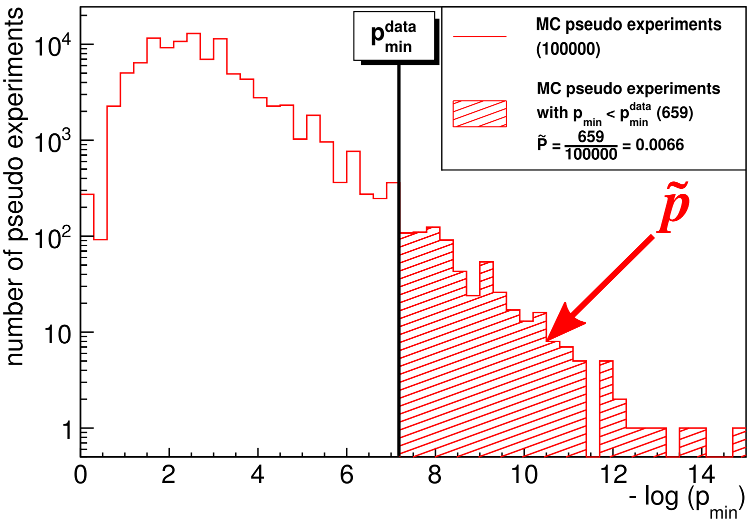


Figure 4.5.: Calculation of post-trial p-value. The $\tilde{p}$-value is defined as the fraction of p-values of pseudo experiments with a lower p-value than in the data distribution. (Picture taken from [66].)

observed in *any* event class as not only multiple regions per event class are investigated but also numerous event classes.

This effect is only handled qualitatively by comparing deviations measured in data to the deviations found in the pseudo experiments.

In a first step, the $\tilde{p}$-values are filled in a histogram as can be seen in figure 4.6. The observed deviations are denoted by the red lines. The x-axis of the histogram is chosen to be the negative logarithm of the $\tilde{p}$-value as new physics would result in small $\tilde{p}$-values which are more pronounced in such a logarithmic representation.

It is possible to evaluate the expected number of deviations as well by calculating $\tilde{p}$-values for each pseudo dicing round and filling the resulting $\tilde{p}$-values for each dicing round in a similar histogram. The $\tilde{p}$-values for this comparison are determined by comparing the p-value of the dicing round with the p-values of all other dicing rounds in the same way as for data. This yields one histogram for each dicing round with the $\tilde{p}$-values found by dicing pseudo data with the MC only hypothesis. The mean (denoted by the solid turquoise line), the median (dashed black line) and the 68 % and 95 % quantiles (purple and light turquoise) for each bin are calculated and filled in the same histogram as the observed deviations.

This procedure allows to evaluate if the observed deviations are expected using the MC only hypothesis. Another benefit is the ability to find not only single deviations with high significance but also deviations in multiple event classes with a lower significance per event class which would not be considered on their own. In this case, the global $\tilde{p}$ distribution would show an excess in multiple bins at lower significances.

The theoretically predicted mean $\tilde{p}$-value is shown as well since p-values are expected to be distributed uniformly between 0 and 1. This allows to check that the $\tilde{p}$-value indeed behaves like a p-value.

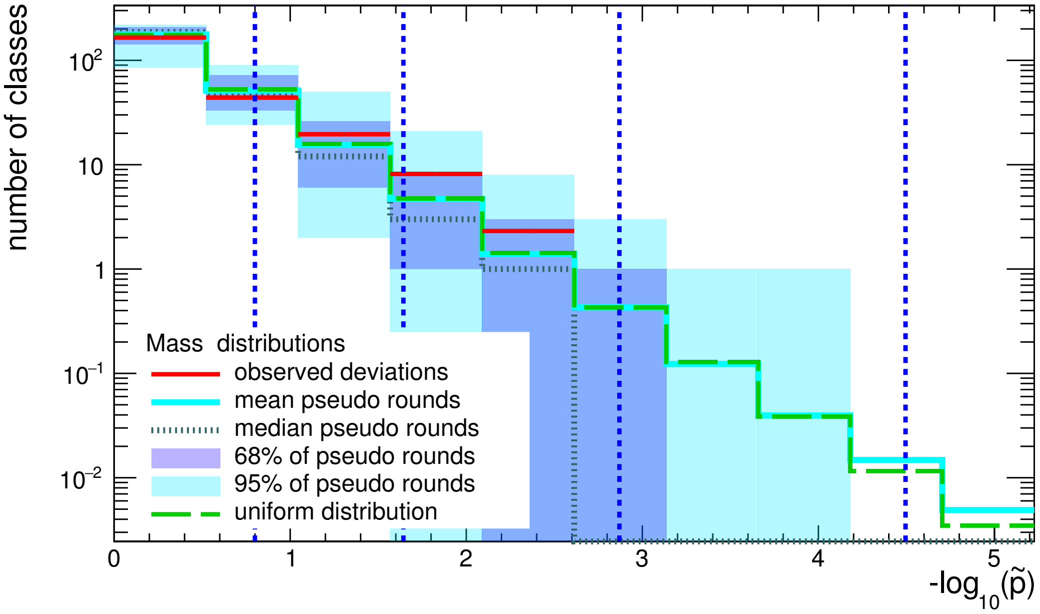


Figure 4.6.: A fictitious global ptilde distribution of $\tilde{p}$-values. The mean and median expected deviation calculated with the pseudo experiments are shown in the turquoise solid and the black dashed line. The $\pm 1\sigma$ and $\pm 2\sigma$ uncertainty bands are given in purple and light turquoise. The $\tilde{p}$-values measured in data are illustrated as red solid lines and the theoretical expectation of a uniform p distribution as the dashed green line.

However, the $\tilde{p}$-values are only expected to follow the theoretical prediction of being equally distributed if infinite pseudo experiments are performed. For event classes with a SM expectation close to zero the number of dicing rounds of 10000 used in this thesis is not sufficient. Classes without data events and a SM expectation below 0.3 events are therefore rejected in the global $\tilde{p}$ distribution.

4.4. Computing

As the analysis does not make use of low level objects, the CMS MiniAOD data format (see section 2.4) in its second version is used as input for the classification.

In a first step of the analysis, the centrally provided data and Monte Carlo sets are analyzed in a pre-selection step to reduce the number of information in order to provide small and easy-to-handle files. This step is performed on the WLCG [54] and the result is stored on the Tier2 data center located at the RWTH Aachen University in the Physics eXtension Library (PXL) format [70], which was developed at the RWTH Aachen University.

The classification and the scan is performed on the local Tier2 computing element using the PXL files and the classification output as input.

4.5. Previous Results of Model Unspecific Searches

Analyses with the same motivation as the MUSiC analysis have been done before the LHC era. The origin of the MUSiC analysis [71] can be found at the L3 experiment [72] at the LEP collider [43]. A global search for deviations between SM simulation and data [73,74,75,76] was performed at the D0 experiment [77] and by the CDF experiment [78] located at the Tevatron collider and by the H1 experiment at HERA [79].

At the LHC a model unspecific search for deviations between expectation and data is performed by both large multi purpose detectors ATLAS [80] and CMS. The foundation of the MUSiC analysis was laid in 2008 [60,61,62]. In the following years the data taken in 2010 at a center of mass energy of $\sqrt{s} = 7$ TeV [63,64] and subsequently at a center of mass energy of $\sqrt{s} = 8$ TeV in 2011 [65] and 2012 [66,67] were analyzed.

In this thesis the first analysis of the 2015 dataset at a center of mass energy of $\sqrt{s} = 13$ TeV and a first look at the data taken in 2016 with the same energy is presented.

5. Object and Event Selection

5.1. Object Reconstruction

In this section the reconstruction of all physics objects used in this analysis, namely muons, electrons, photons jets and missing transverse energy is described.

5.1.1. The Particle Flow Algorithm

The detector does not measure objects e.g. electrons or muons themselves but relies on reconstruction algorithms to reconstruct the object from the given detector response. The particle flow (PF) algorithm [81, 82] is used to reconstruct most particles in a consistent way and can be divided in three steps: First, tracks are fitted and clusters of energy depositions in the calorimeters are built, second, tracks, clusters and hits in muon chambers are linked to each other and third, the particles are reconstructed based on the linked tracks and clusters.

The prerequisite for this algorithm to work are high granularity detector components and an efficient tracking [82]. All of these are fulfilled by the CMS detector [82].

Track Fitting and Cluster Building An iterative tracking algorithm is used to achieve a high reconstruction efficiency of tracks while preserving a low rate of misidentified tracks. Track reconstruction is a computationally challenging task and is performed using combinatorial Kalman Filters [83, 84, 85]. For the first attempt of track reconstruction tight reconstruction criteria are set to suppress misidentified tracks. Tracks reconstructed with the tight criteria are removed from the collection of all tracks and the remaining tracks are reconstructed using looser criteria. This procedure is repeated five times in order to associate nearly all tracker hits to tracks [81].

Tracks in the muon chambers are reconstructed using the same techniques [86].

Secondly, calorimeter clusters are built in the ECAL and HCAL separately using the cells with the highest energy deposit as a seed and adding adjacent cells as long as the energy deposition exceeds a certain threshold of a few hundred MeV depending on the electronics noise [81].

Linking The next step is to link tracks and clusters to blocks and assign a quality measure for the link called link distance.

At first, tracks are extrapolated to the calorimeters and clusters near the extrapolated track are linked together with the distance in the η, ϕ-plane as the link distance. Secondly, clusters in the ECAL are linked together by extrapolating tangents from intersection points of tracks and tracker layers as they might originate from bremsstrahlung photons. The third step is to link close-by

ECAL and HCAL clusters together and lastly establish links between tracks and hits in muon chambers with the χ^2 of the combined track fit as link distance.

Particle Identification At the start, muons are identified from the collection of previously linked muon candidates if the tracker momentum and the momentum measured in the muon chambers coincide roughly. The track of the muons is then removed from the collection of all tracks. Electrons are identified using special fits (Gaussian Sum Filter), to take the energy loss due to bremsstrahlung into account, which alters the track and attaches clusters of bremsstrahlung photons in the ECAL to the electron and removes them from the collection [87].

The remaining tracks with linked clusters are reconstructed as charged hadrons taking the link distance into account to solve ambiguities and remaining clusters in the ECAL (HCAL) are reconstructed as photons (neutral hadrons).

These reconstructed objects are then used to reconstruct jets and calculate the missing transverse energy of the event.

5.1.2. Jet Reconstruction

As quarks and gluons produced in the interaction do not traverse the detector as single particles but hadronize as described in section 1.8.3 the momenta of all particles originating from the parton produced in the hard interaction should be added up in order to access the kinematic properties of the initial particle.

This can be achieved by an infrared and collinear splitting (IRC) safe jet clustering algorithm. Collinear splitting denotes the splitting of the energy of the initial quark or gluon produced in the hard interaction in two nearly equally energetic partons moving in almost the same direction. The jet clustering algorithm should not be sensitive to such splittings, e.g. not changing the number of jets as both partons could be reconstructed as separate jets, but rather combine both jets to one jet. The direction should also not depend on small deviations in the energy splitting between both partons. An algorithm centering a cone around the highest energetic particle and adding the energy of all neighboring particles would for example be sensitive to small energy deviations, as it would center the cone around one of the two partons depending on their energy even if they have nearly the same energy [88].

Infrared safe algorithms do not produce additional jets due to soft radiations, but combine the soft radiation with the rest of the jet [88].

The anti-k_t algorithm [89] used for the jet reconstruction in CMS is such an IRC safe algorithm and tends to provide jets with a conical shape with a distance parameter of $R = 0.4$.[1]

---

[1] If the jet is isolated from other hard particles the jet is indeed perfectly circular. If other hard particles are nearby the shape of the jet is more complex.

Jet Energy Scale and Resolution Corrections

The detector response to jets can differ between data and MC simulation and also shows a difference in the MC simulation between the reconstructed and the generated energy. The measurement of the jet energy scale (JES) is corrected to account for contamination of pile-up, non-linear detector response and residual differences between data and simulation [52]. These corrections are referred to as jet energy corrections (JEC).

In a first step the energy deposited by pile-up jets is removed. It also accounts for differences between the pile-up in data in its description in the Monte Carlo simulation. Next, Monte Carlo simulation is used to determine a correction between the p_T of the generated particle before and after the detector simulation. In the last step the JES is corrected for the difference between data and MC simulation. Of course, this step is only applied to data.

In addition it has been observed that the jet energy resolution (JER) is worse in data than in MC simulation. The resolution of simulated jets is smeared to reproduce the JER measured in data [90].

5.2. Event Selection

As described in section 2.3 not every collision event in the CMS detector is recorded, but triggers are used to select interesting events and reduce the event rate to an acceptable level which can be handled by the computing resources. In addition several event filters are used in order to reject events with degraded reconstruction efficiency.

5.2.1. Trigger Selection

This section describes the single and double lepton triggers used in the MUSiC analysis (see 3.1).

Every trigger has a threshold for the transverse momentum of the triggering particle and also might demand additional object requirements. The double lepton triggers are used to increase the event yield since the trigger thresholds and object requirements for the double lepton triggers are lower than for the single lepton triggers. The efficiency for the trigger does not increase sharply at the momentum threshold, but increases over a small momentum range. This behaviour is called turn-on and is not perfectly modeled in the Monte Carlo simulation which leads to large differences in the trigger efficiency in this region between data and simulation. To avoid huge scale factors between data and MC simulation a momentum threshold above the trigger threshold is required in the offline analysis for the triggering particle. This requirement is applied on the first (and second) object in a p_T ordered list for single (double) lepton triggers. The trigger names and the p_T threshold for each trigger can be found in table 5.1.

An event has to be triggered by at least one of these triggers. The trigger requirements are encrypted in the trigger name (see table 5.1). The numbers behind the shortcuts *Ele* or *Mu* correspond to the p_T threshold in GeV, e.g. $p_T > 105$ GeV in case of the single electron trigger or $p_T > 17$ GeV for the first muon and $p_T > 8$ GeV for the second muon in case of the double

Table 5.1.: Trigger used in the MUSiC analysis. An event has to be triggered by at least one of these triggers. The offline p_T threshold for the triggering particle is given in the last column. In case of double lepton trigger a p_T threshold for the two particles with the highest p_T are given.

Type	Trigger Name	Thresholds / GeV
Single Muon	`HLT_Mu50`	53
Single Electron	`HLT_Ele105_CaloIdVT_GsfTrkIdT`	120
Double Muon	`HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ`	35, 35
Double Muon	`HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ`	35, 35
Double Electron	`HLT_DoubleEle33_CaloIdL_GsfTrkIdVL`	50, 50

muon trigger[2]. The rest of the trigger names encrypts object requirements for the identification of objects and is explained in detail in appendix C.

5.2.2. Event Filter

Despite the event certification (see section 3.1) single events can be poorly reconstructed due to several reasons like anomalous noise in the HCAL, particles of the beam halo passing the detector[3] or high energy depositions near masked ECAL crystals [92].

The whole list of event filters can be found in appendix B. The filters are used on both data and MC simulation and do not aim to correct these events but rather veto them to avoid misidentified objects and a wrong E_T^{miss} calculation.

5.2.3. Primary Vertex Reconstruction

At least one primary vertex must be found in the event which is not displaced more than 24 cm in longitudinal direction and 2 cm in the transverse plane from the interaction point, or have less than 5 sufficiently well reconstructed tracks originating from the vertex.

This requirement is used to reduce the rate of events triggered by cosmic muons since for them ideally only two tracks are reconstructed and the vertex might be displaced in such events. For normal collision events in a hadron collider several tracks originating from the primary vertex are expected due to the hard scattering process and the underlying event.

5.3. Object Selection

The MUSiC analysis sorts every event in event classes defined by the final state of the event (see section 4). Reconstructed particles might be of low quality or are misidentified and a set of object requirements are used to reject such objects. The general reconstruction of each object is described

---

[2] The offline p_T requirement for the double muon trigger is set to a high value of 35 GeV to avoid the relatively long turn-on of the trigger [91].

[3] Particles of the beam might get deflected, for example by interactions with gas particles in the beam pipe, but still pass the CMS detector due to deflections in the beam optics [92].

in section 5.1. This section describes the identification requirements for each object, i.e. electrons, muons, photons, jets and E_T^{miss}. The threshold values differ for particles in the barrel and in the endcap region and are listed in appendix A (tables A.1 to A.6) and the acceptance requirements are summarized in table 5.2.

5.3.1. Electrons

Electrons are reconstructed using the Gaussian sum filter (Gsf) as described in section 5.1.1. The additional identification criteria differ depending on the transverse momentum of the electron candidate. For electrons with $p_T < 100$ GeV the tight working point of the cut based electron ID [93] is used. Most requirements are imposed to limit the misidentification rate of jets [94]. In the following each identification requirement and its purpose is described.

- The shower shape is measured in a variable classifying the spread of the shower in η direction called $\sigma_{i\eta i\eta}$. The shower shape for an electromagnetic shower differs from the shape of an hadronic shower as electromagnetic showers are typically spread less widely.
- The position in η and ϕ between the track measurement and the supercluster measurement in the calorimeter has to match within acceptable bounds to reject photons and neutral hadrons with a falsely linked track.
- An electron should deposit most of its energy in the ECAL due to its thickness of 25.9 electromagnetic radiation lengths (see section 2.2.4). Thus the fraction of energy deposited in the HCAL should be small for electrons.
- Prompt electrons are typically isolated compared to particles in jets. Pile-up can reduce the isolation of the electron, thus the isolation requirement for electrons depends on the mean deposited energy in the ECAL in the event [95]. The isolation is then calculated by

$$< F9 > \left(\sum p_T^{\text{ch. had. pv}} + \max\left\{0, \sum E_T^{\text{neu. had.}} + \sum E_T(\gamma) - \cdot \sum p_T^{\text{ch. had. pu}}\right\}\right) / p_T(e). \quad (5.1)$$

- The momentum measurement of the tracker should coincide with the energy measurement in the calorimeter. In jets energy can be carried by electrically neutral particles leading to disagreement between both measurements.
- The minimal difference between the position of the track extrapolated to the beam pipe and the primary vertex should be smaller than a few millimeters to reject non-prompt electrons. The values vary for the displacement in the xy-plane and the z-direction and for electrons in the barrel and the endcap. The displacement criteria on the xy-plane and the barrel are typically tighter than for the z-direction and the endcap. The values reach from 0.1 mm for the displacement in the xy-plane for electrons in the barrel region and 4.2 mm for the displacement in the z-direction for electrons in the endcap.
- The electron needs to have not more than one (two) missing hit in the innermost layers of the tracker for the barrel (endcap) to suppress electrons arising from photon conversion.
- A high energetic bremsstrahlung photon can convert in the detector to an electron positron pair. These electrons are rejected according to the χ^2 of a fit to the electron and the positron using the constraint that they have the same tangent at the conversion point [96].

For electrons with $p_T > 100$ GeV a dedicated high electron identifier (High Energy Electron Pairs ID - HEEP ID [97, 94]) is used. The variables used to identify electrons are very similar but the values are tuned to identify high energy electrons.

- For the endcap region a requirement on $\sigma_{i\eta i\eta}$ is used similar to the low energy ID. For the barrel region another measure of the energy spread is used which fulfills the same purpose.
- Similar requirements are demanded for the difference between track and supercluster position[4] and the fraction of energy deposited in the HCAL.
- The electron needs to be identified by the calorimeter and not only the tracker as the momentum measurement of the tracker gets imprecise for high energy particles.
- The track might be slightly more displaced (0.2 mm for barrel electrons) from the primary vertex in the xy-plane compared to the lower energy ID, since the track measurement for high energy particles is worse. The requirement on the displacement in the z-direction is dropped for the high energy ID.
- Only one hit in the innermost layers of the tracker might be missed to suppress conversions of photons.
- The electron needs to be isolated from the energy depositions in the first readout towers of the HCAL in a cone around the electron ($R = 0.3$, excluding the center of the electron shower ($R = 0.15$) to suppress misidentified jets.
- The sum of the transverse momentum of a track in a cone around the electron ($R = 0.3$) must be below 5 GeV in order to suppress misidentified jets as well.

Additionally for electrons identified by the HEEP ID, the energy and pseudorapidity measurement of the calorimeter is used instead of the track measurement.

5.3.2. Photons

The identification criteria of photons are very similar to the criteria of electrons, since both produce electromagnetic showers in the ECAL. The main difference to electrons is the missing track. In this analysis the tight working point of the cut based ID for 25 ns bunch spacing [98] is chosen.

The photon must fulfill requirements on the shower shape and the fraction of energy deposited in the HCAL for the same reasons as for electrons. Furthermore, it needs to be isolated from the energy depositions of charged and neutral hadrons and other photons [95], which were identified by the particle flow algorithm (see section 5.1.1).

It also needs to pass a pixel seed veto to suppress converted electrons. This veto is recommended for analyses sensitive to the misidentification rate of electrons to photons [99]. It vetoes every photon which has at least two hits in the pixel detector that suggesting a charged particle arriving at the same position in the ECAL as the photon.

---

[4] As explained in section 5.1.1 the calorimeter entries are clustered together and the final cluster is called supercluster.

5.3.3. Muons

The identification of muons is split into low and high energetic muons. Bremsstrahlung of the muon in the iron return yoke complicate the momentum measurement of the muon at high energies. While the momentum measurement of the tracker might be sufficiently accurate for low energy muons, CMS can use the dedicated muon system in order to reconstruct high energetic muons with high precision. The momentum of the muon is determined by the *TuneP* algorithm which uses a decision tree to choose the best fitting algorithm[5] on a muon-by-muon basis [53].

Below a p_T of 200 GeV the tight working point of the cut-based muon identification is used while for higher energetic muons slightly different requirements are set [100]. The goal of the identification requirements is to suppress hadronic punch-trough of jets and the identification of muons produced by decays of particles inside the detector. In the following the identification criteria for the tight working point of the cut based ID and the dedicated high-energy ID are given. Some criteria apply for both IDs.

[both] For both IDs the muon is required to be a global muon meaning it has hits in the tracker and in the muon system.

[tight] For the low energetic ID muons have to be identified by the particle flow algorithm. The particle flow algorithm is designed to identify muons with a medium p_T. Therefore the requirement is dropped for the high-p_T ID.

[tight] Low energetic muons need to have a good quality of the global fit with a $\chi^2/\mathrm{ndf} < 10$ to suppress hadronic punch-through and muons created by decays of particles in the detector.

[both] At least one muon chamber must be included in the global muon fit and at least hits in two muon stations need to be matched.

[both] The muon track is not allowed to be displaced from the primary vertex more than 2 mm in the xy-plane and 5 mm in the z-direction in order to suppress non-prompt muons.

[both] At least hits in one pixel and six tracker layers have to be matched to the muon to ensure a good p_T resolution and suppress non-prompt muons.

[high-p_T] Since the momentum measurement for high energetic muons is more difficult due to straight tracks, the relative p_T uncertainty has to be smaller than 30 %.

In addition the muon needs to be isolated against charged hadrons from the primary vertex (ch. had. pv), neutral hadrons (neu. had.) and photons (γ). The isolation requirement is relaxed with the amount of charged hadrons from pile-up (ch. had. pu) in the event as the muon should be isolated from particles originating from the same vertex and not from pile-up contamination, and is defined as

$$\left(\sum p_T^{\text{ch. had. pv}} + \max\left\{0, \sum E_T^{\text{neu. had.}} + \sum E_T(\gamma) - 0.5 \cdot \sum p_T^{\text{ch. had. pu}}\right\}\right) / p_T(\mu) < 0.15. \tag{5.2}$$

---

[5] Examples for different fitting algorithms are the *Tracker-Plus-First-Muon-Station Fit* which only takes the hit in the first layer of muon chambers into account or the *Picky Fit* which ignores hits caused by showers in the muon stations [53].

5.3.4. Jets

Jets are reconstructed using the anti-k_t algorithm and their energy is corrected for e.g. pile-up contribution as described in section 5.1.2.

All reconstructed jets must fulfill the requirements of the tight working point of the jet ID [101] and its values can be found in table A.6 in appendix A. Jets are required to have at least 2 constituents to suppress single particles reconstructed as jets and have additional requirements on the energy fraction of neutral hadrons, charged hadrons and charged and neutral electromagnetic energy depositions, as a typical high energetic jet is expected to consist of these various components.

5.3.5. Missing Transverse Energy

The missing transverse energy is reconstructed by the PF algorithm outlined in section 5.1.1. Its value is corrected for the JEC (Type-1 correction), the JER (see section 5.1.2) and the assignment of the muon momentum by the TuneP algorithm (see section 5.3.3) to keep the event balanced in the transverse plain [92, 102]. For electrons identified by the HEEP ID (see section 5.3.1) the energy measurement of the calorimeter cluster is used instead of the track measurement and the E_T^{miss} is adapted accordingly.

Missing transverse energy can be caused particles which do not interact inside the detector (e.g. neutrinos). Additionally E_T^{miss} can be created by two kinds of not perfectly measured objects: Clustered objects like muons or jets and unclustered objects. The unclustered part of the E_T^{miss} will remain unchanged by these corrections, but the clustered part will shift when applying the corrections. The JEC will introduce an additional uncertainty on the E_T^{miss} scale as well.

5.3.6. Event Cleaning

The PF algorithm (see section 5.1.1) can attribute the same track or calorimeter cluster to multiple particles. For example muons are often reconstructed as jet as well. In order to avoid double counting of objects a cleaning algorithm is applied. This is mandatory since the MUSiC analysis is sensible to the whole event content and not only single predefined particles.

The event cleaning is performed consecutively for muons, electrons, gammas and jets. In the first step low energetic electrons are removed if they have a higher energetic electron close-by and use the same supercluster seed or the same track. This kind of cleaning against other nearby particles of the same kind is not performed for muons in order to not lose sensitivity for boosted topologies, where two muons with small spatial distance are expected. Radiation of bremsstrahlung photons from high energetic muons can be reconstructed as electrons. Therefore electrons close to muons are removed.

The cleaning for photons is very similar to the cleaning of electrons with the difference that they obviously cannot share the same track. Photons close to electrons and muons are removed as they could be misidentified bremsstrahlung which are not matched to the electron or the muon.

In the last step all jets are removed which are very close to any other particle.

Table 5.2.: Acceptance requirements for every particle type on pseudorapidity and transverse momentum.

Particle	Pseudorapidity	Transverse Momentum
Muons	$\|\eta\| < 2.4$	$p_T > 25$ GeV
Electrons	$0 < \|\eta\| < 1.4442$ or $1.566 < \|\eta\| < 2.5$	$p_T > 25$ GeV
Photons	$\|\eta\| < 1.4442$	$p_T > 25$ GeV
Jets	$\|\eta\| < 2.4$	$p_T > 50$ GeV
E_T^{miss}	-	$E_T^{\text{miss}} > 100$ GeV

5.3.7. Reconstruction and Identification Efficiency

The reconstruction and identification efficiencies for each object and ID may differ between data and MC simulation. Scale factors have been determined to adjust the efficiency of the MC simulation to match the efficiency measured in data. The scale factors are determined depending on η and p_T by the respective physics object group.

Scale factors are applied for the efficiencies for the identification and isolation of muons [103], the identification of electrons and photons and the reconstruction of electrons [104]. The identification efficiency for jets agrees between data and MC simulation [105] and no scale factors are applied. In general the scale factors are in the order of magnitude of 1.

5.3.8. Acceptance Requirements

Every object needs to fulfill additional acceptance requirements on the pseudorapidity and the transverse momentum. The pseudorapidity requirements are mainly used to exclude parts of the detector with suboptimal reconstruction efficiency, e.g. the overlap region between barrel and endcap in the ECAL or the end of physical coverage of the detector, while the p_T thresholds are used to ensure a good identification efficiency and preserving a low misidentification rate. All requirements are summarized in table 5.2.

5.4. Systematic Uncertainties

The Standard Model expectation in the MUSiC analysis is calculated using Monte Carlo simulations, which do not give an exact representation of the Standard Model but rather an estimate describing the Standard Model within uncertainties. Those uncertainties arise for example due to imprecise calculations of perturbation theory (see section 1.7), an inaccurate description of hadronization of partons (see section 1.8.3), imprecise simulation of the geometry of the detector leading to a mis-modeled momentum resolution, or the finite number of simulated events.

The MUSiC analysis tries to be as inclusive as possible and to consider every final state, thus it is unavoidable that the description of the systematic uncertainties cannot be accurate for all final states. It is inevitable that the uncertainties are overestimated for some regions as conservative values are used for all systematic uncertainties.

Table 5.3.: Uncertainty values for each order of cross section.

Cross Section Order	Uncertainty
LO	50 %
NLO	10 %
NNLO	5 %

In the following every considered source of uncertainties is explained. The impact of each uncertainty cannot be stated globally as it varies vastly for each event class and region.

5.4.1. Luminosity

The number of counted events is directly proportional to the luminosity $\mathcal{L}$

$$N = \mathcal{L} \cdot \sigma \cdot A \cdot \epsilon, \tag{5.3}$$

with the cross section σ, the acceptance A and the total efficiency (trigger, reconstruction and identification) ϵ. The uncertainty on the luminosity translates directly on the number of events. The value of the luminosity itself is calculated by the Pixel Cluster Counting method and the uncertainty is estimated to be 2.7 % [55].

5.4.2. Pile-up Calculation

The pile-up in data is calculated using the total inelastic proton-proton cross section (minimum bias cross section) and the instantaneous luminosity [55]. The minimum bias cross section is measured to 71.3 mb with an uncertainty of 5 %. This uncertainty is propagated to calculate an up- and downwards shifted pile-up weight (see section 3.3) whose difference to the nominal weight gives an estimate of the impact of the uncertainty on the pile-up calculation [106].

5.4.3. Cross Section of Standard Model Processes

The cross sections of the simulated Standard Model processes are calculated by perturbation theory. Depending on how many higher-oder corrections are taken into account (see 1.7) the precision of the cross section calculation improves. The assumed uncertainties for each perturbation order can be found in table 5.3 and the order of the cross section calculation for every sample can be found in table D.2.

For some samples the integral cross section is analytically known to a higher order than available for the full event generation and a constant correction factor is used to match the cross section calculated at higher orders (see 3.2). In this case the uncertainty corresponds to the corrected cross section. The cross section order corresponds only to the order of QCD corrections and not the electro-weak corrections.

The values used for the uncertainties (see table 5.3) are rough estimates and chosen to be conservative in most final states in order to cover the uncertainty in different final states.

5.4.4. Parton Distributions Functions

In order to calculate the cross section Parton Distribution Functions (PDF) are used, since the proton is a composite particle (see section 1.8.1). The uncertainty is calculated following the `PDF4LHC` recipe [107].

The PDF uncertainty can be estimated by re-generating each event with different PDF sets which has the disadvantage of being very time consuming. The `PDF4LHC` recipe follows a computational much faster re-weighting method instead which uses a set of variations around the central value. For the samples generated at NLO precision the weights are stored as they cannot be re-computed using LHAPDF. The `PDF4LHC` set combines multiple PDF sets into a single set of variations. However, the `PDF4LHC` sets were not available by the time the production of the RunII MC begun [108] and weights of other PDF sets (mostly `NNPDF30`) are used. All PDF sets used for this analysis are listed in table F.1 in appendix F and are accessed via LHAPDF [109] if the weights are not stored in the event.

5.4.5. Factorization and Renormalization Scale Uncertainty

The choice of the factorization (μ_F) and renormalization (μ_R) scale, which is related to missing higher orders, influence the cross section calculation of NLO calculations as well [56, 57]. The uncertainty of the chosen value is estimated by varying the scales by a factor one half and two around the central value. The maximal difference is taken as the uncertainty [110].

This uncertainty is negligible in most final states except for $t\bar{t}$ dominated event classes with high jet multiplicities. These show a large systematic deficit in data which is covered by this uncertainty as the jet matching between Powheg and Pythia8 is not perfectly tuned [111].

5.4.6. Object Energy Scale Related Uncertainties

The energy of a reconstructed particle cannot be measured without an uncertainty as for example the detector alignment and calibration is not known exactly. The classification is repeated with energy values which are shifted by $\pm 1\sigma$ of the corresponding energy scale uncertainty and the difference to the central value is taken as the uncertainty to account for the impact of these energy scale uncertainties. The energy scale of each type of particles (e.g. electrons or jets) is shifted separately as the energy scale of different particle types is assumed to be uncorrelated but the energy scale of particles of the same type is fully correlated.

The shift of the energy can lead to migration of events to other event classes as the particle which normally passes the p_T threshold might not pass it if the energy scale is shifted downwards (and vice versa for shifts upwards). Therefore events that do not enter an event class can still contribute to the uncertainty of this class.

Jet Energy Scale

As described in section 5.1.2, the energy of jets is corrected to map the measured deposited energy to the particle-level energy. The uncertainty of these corrections does not only change the energy of jets, but also the missing transverse energy, which is adapted to the JEC (see section 5.3.5).

The uncertainties on the JEC are treated by shifting the energy of jets by $\pm\sigma$ of the JEC uncertainties and adapt the E_T^{miss} accordingly. Since both, jets and E_T^{miss}, require a minimal threshold energy to be reconstructed, these shifts can lead to migration of events to other event classes.

Lepton and Photon Energy Scale

The uncertainty of the energy scale of leptons and jets are treated analogously to the jet energy scale. For electrons the relative uncertainty is estimated to be 0.2 % in the barrel and 0.3 % in the endcap for low energetic electrons and 2 % for HEEP electrons [96] using the differences between data and simulation of masses of well known resonances (e.g. the Z-boson) [96]. The energy scale uncertainty of photons is evaluated using the same method but discarding the tracker information of the electrons. It is estimated to 0.15 % in the barrel and 0.3 % in the endcap [99].

The momentum scale uncertainty of the muon was determined dependent on η, ϕ and p_T with the generalized endpoint method [112] which exploits the fact that the curvature of two muons from a Z-boson decay should be equal to each other [113]. The p_T dependence is estimated by scaling the uncertainty σ_0 linearly with the p_T of the muon in TeV

$$\sigma = \sigma_0 \cdot \frac{p_T}{1\,\text{TeV}}. \tag{5.4}$$

For muons the shift is not applied on p_T but rather $1/p_T$ as the momentum is measured by the bending of the track which is proportional to $1/p_T$.

Unclustered Energy Scale

After reconstructing all particles in the detector some energy depositions will not be matched to reconstructed particles, leaving unclustered energy in the event. This energy is not linked to any physical object but is taken into account for the E_T^{miss} calculation. The uncertainty is treated similar to other energy scale uncertainties described above. Its value varies on an event to event basis and is calculated by varying each particle type within their own resolution [114, 115].

5.4.7. Energy Resolution

Not only the energy scale of objects is known within given uncertainties but also their energy resolution. This uncertainty can lead to migration to other event classes if particles are shifted above or below the momentum threshold.

Lepton Momentum Resolution

The momentum resolution uncertainty is applied by randomly smearing the momentum of the particle according to a Gaussian prior. This uncertainty is only applied to muons with an uncertainty of 3.2 % of the p_T of the muon and is determined by comparing the momentum measurement of cosmic muons in the upper and lower detector halves [53].

Jet Energy Resolution

The energy resolution for jets is corrected in simulation to reproduce the resolution measured in data as described in section 5.1.2. The uncertainty of this measurement is determined using a centrally provided tool to assess the uncertainty based on jet p_T and η [90].

5.4.8. Reconstruction and Identification Efficiency

As mentioned in section 5.3.7 the reconstruction and identification efficiency is adjusted for the MC simulation to match the efficiency measured in data. The uncertainty of these scale factors are taken into account by shifting the scale factors by one standard deviation and compare the resulting histogram with the histograms obtained with the unshifted scale factors. Within MUSiC, events cannot migrate to other event classes due to this uncertainty and the uncertainty is small compared to other uncertainties in most distributions. The uncertainties for all scale factors can be found in [103, 104].

5.4.9. Misidentification Probability

Despite the identification criteria for each object, some objects might be misidentified, e.g. a small fraction of jets is identified as electrons. The misidentification rate of objects might differ between data and MC simulation. Several dedicated searches for supersymmetry [116] and resonances in the di-electron spectrum [117] indicate that an uncertainty of 100 % on the scale factor between data and MC for misidentified electrons, muons and photons is an acceptable choice. The scale factor itself is in the order of magnitude of 1 and the misidentification rate for example of jets to electrons in the order of magnitude of 10^{-2} [96]. Objects are tagged as misidentified objects if no generated particle of the same type is found in a cone of $R = 0.2$ around the reconstructed particle.

It is important to note that this uncertainty is not used to correct for misidentified objects but to describe the difference in the misidentification rate between data and MC simulation.

5.4.10. Simulated Number of Events

The Standard Model expectation is estimated by MC simulation. Since the number of generated events is finite, the MC simulation is only an estimate of the true value of Standard Model events contributing to one bin. A Poissonian uncertainty for every simulated sample of $\sqrt{N} \cdot \alpha$ is introduced with the number of generated MC events N and the total weight α.

5.4.11. Correlation of Uncertainties

All uncertainties apart from the uncertainty on the total number of simulated events and the cross section calculation are considered as fully correlated among each other and are added linearly across all bins of a region. The different kinds of uncertainties are then added in quadrature, as they are treated as totally uncorrelated.

The total numbers of simulated events per region of different datasets are not correlated to each other and are their uncertainties added in quadrature. The uncertainty on the cross section calculation is considered fully correlated across the same physical process (e.g. single W production) and fully uncorrelated across different processes (e.g. W production and Drell-Yan).

6. Results of the 2015 Dataset

This chapter presents the results of the full analysis of the 2015 dataset. In a first step, the general agreement between data and MC simulation is checked using a set of examples for different spectra (also visualizing special features of the MUSiC analysis) and the p-value for the integral distribution of the event classes with the most data events. After that, the results of a full scan for deviations are presented and the most significant deviations are discussed in detail.

6.1. Normalization Check and Examples

6.1.1. M_{inv} Distribution of the 2μ Inclusive Event Class

The first example is the invariant mass distribution of the 2μ inclusive event class and can be found in figure 6.1. All contributing MC simulations are shown as a colored, stacked histogram[1] with the data points in black on top. The combined statistical and systematical uncertainty of the MC simulation is shown as a gray, hatched area. The ratio of data and MC simulation is shown below the spectrum with a binning to contain at least one MC event per bin. This helps judging if single events in high energy tails are significant or if the amount of measured events in this region is compatible with the SM expectation. A direct ratio between data and MC simulation would not account for neighboring bins without data events and would make it difficult to judge the significance of this excess. The legend provides the integrated number of weighted events per process.

The red dashed lines show the region of interest found by the scanning algorithm described in section 4. It shows a deficit in one bin which is not statistically significant according to the $\tilde{p}$-value of 0.96 shown on top.

The spectrum shows a distinct peak at a mass of $90 - 100$ GeV which corresponds to the resonant production of a Z-boson decaying into two oppositely charged muons. The position and height of the peak matches in data and MC simulation and validates the NNLO k-factor. The MC simulation verifies this assumption as the Drell-Yan process is the most dominant SM process in this event class especially in the peak region. The spectrum falls continuously to higher masses and is well described by the MC simulation whose shape is mostly calculated at NLO precision. The highest mass event has an invariant mass of ≈ 2.3 TeV. The agreement between data and MC simulations indicates a proper absolute normalization of the MC simulation.

---

[1] The assignment which processes are associated with each color can be found in table E.1 in appendix E.

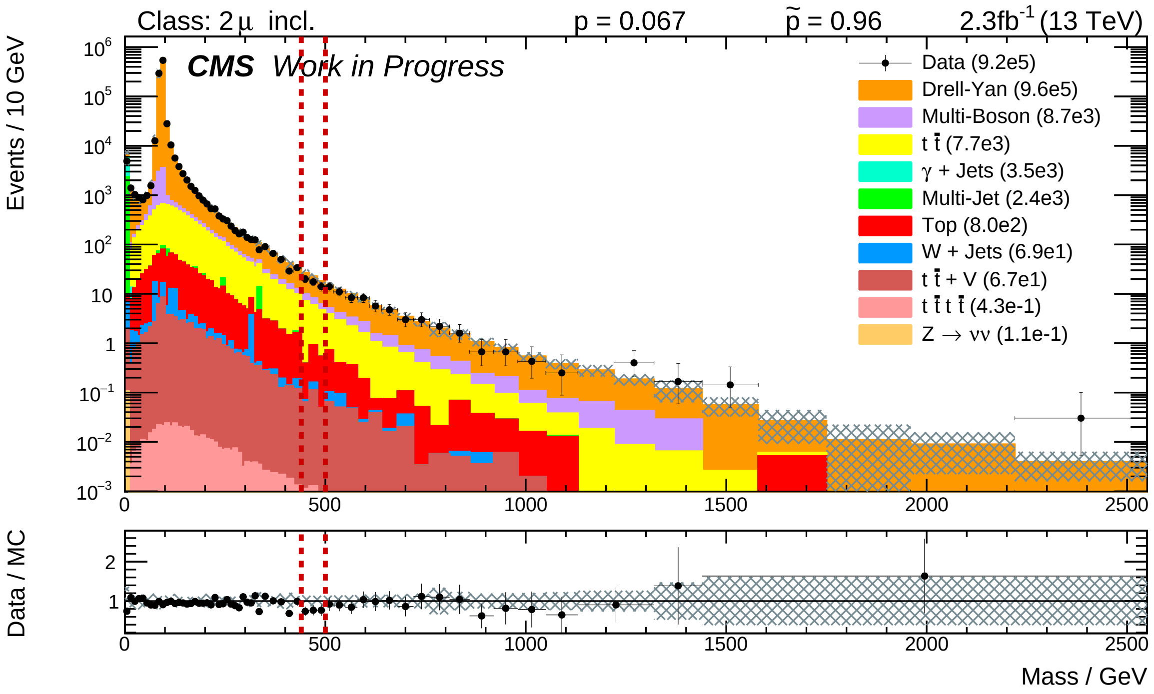


Figure 6.1.: Invariant mass spectrum of the di-muon inclusive event class.

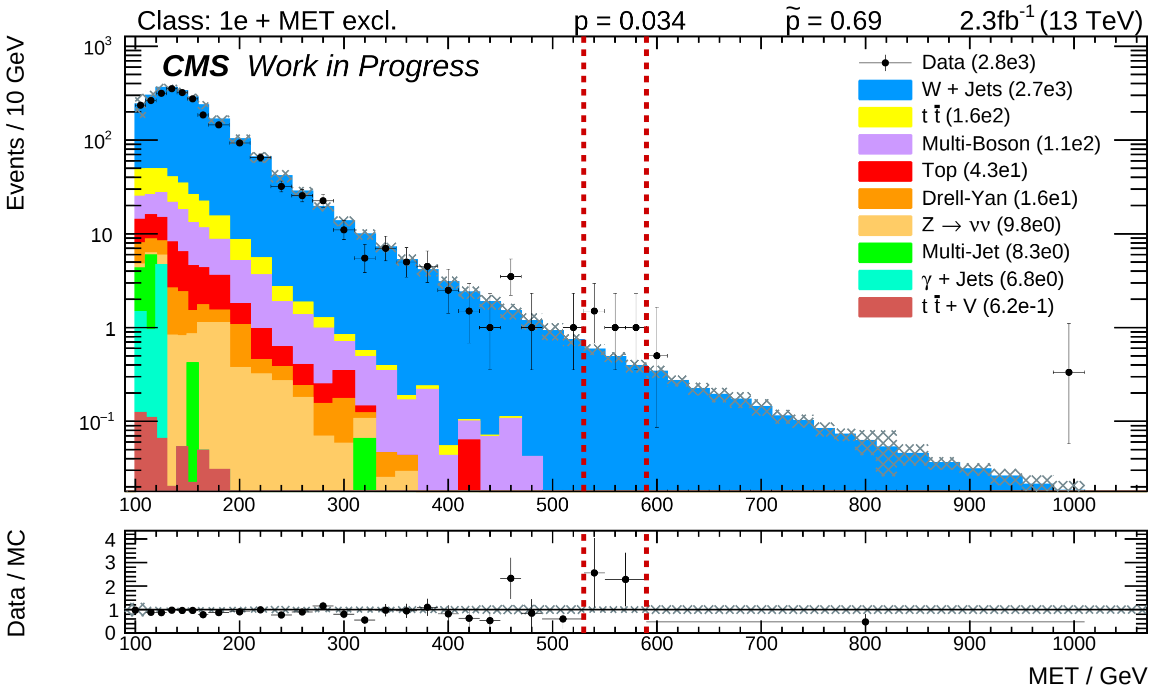


Figure 6.2.: Missing transverse energy spectrum of the $1e + E_T^{\text{miss}}$ exclusive event class.

6.1.2. E_T^{miss} Distribution of the $1e + E_T^{\text{miss}}$ Exclusive Event Class

The E_T^{miss} distribution of the $1e + E_T^{\text{miss}}$ exclusive event class is dominated by the single W-boson production as expected as this event class corresponds directly to the final state of the W-boson decay into an electron and a neutrino which is reconstructed as E_T^{miss} as it travels through the detector undetected. The spectrum can be found in figure 6.2.

The narrow turn-on region visible at low energies can be explained by the trigger requirement for the single electron trigger. The trigger threshold for the single electron trigger is at $p_T = 120\,\text{GeV}$ (see section 5.2.1) while the lower edge of the spectrum is at 100 GeV. If a single W boson is produced without any other particles the p_T of the electron and the E_T^{miss} have to be perfectly balanced due to momentum conservation in the transverse plane of the detector. In order to create E_T^{miss} below the electron trigger threshold additional soft particles have to be created which do not pass the selection criteria as no additional objects are allowed in events entering exclusive event classes.

Above the peak the spectrum decreases as the cross section falls off to higher energies as either the W-boson has to be produced off its mass shell or a boosted topology is required to produce high energy neutrinos with an on-shell W-boson.

The region of interest shows a small deficit of data which is not significant with a $\tilde{p}$-value of 0.69.

This spectrum shows good agreement for one of the most important SM processes in the MUSiC analysis. It further validates the correct matching of the different samples used to simulate the W+Jets process and checks the applied k-factors.

6.1.3. M_T Distribution of the $1e + 1\mu + 1jet + E_T^{\text{miss}}$ Jet Inclusive Event Class

The third example shows the transverse mass distribution of the $1e + 1\mu + 1jet + E_T^{\text{miss}}$ jet-inclusive event class (see figure 6.3) which is one of many $t\bar{t}$ dominated event classes. The (anti) top quark decays to a W boson and a bottom quark and the W bosons further decay to leptons. This yields up to two prompt leptons of possibly different flavour. Only the transverse mass is calculated for this event class as missing transverse energy is present.

The deviation found at around 900 GeV is not significant with $\tilde{p}$ = 0.4 and shows the good agreement for the $t\bar{t}$ MC simulation. The $t\bar{t}$ MC dominates over 40 % of event classes with different lepton flavours or with high multiplicities due to the multiple possible final states due to the decay chains of the top quark.

6.1.4. $\sum|p_T|$ Distribution of the $1e$ Inclusive Event Class

The last example shows the $\sum|p_T|$ spectrum of the $1e$ inclusive event class (figure 6.4). It contains contributions of the different trigger streams and verifies the cleaning between them. The multiple steps in the distribution correspond to the trigger thresholds of different triggers and each region and each step is explained in the following starting at the lowest momentum.

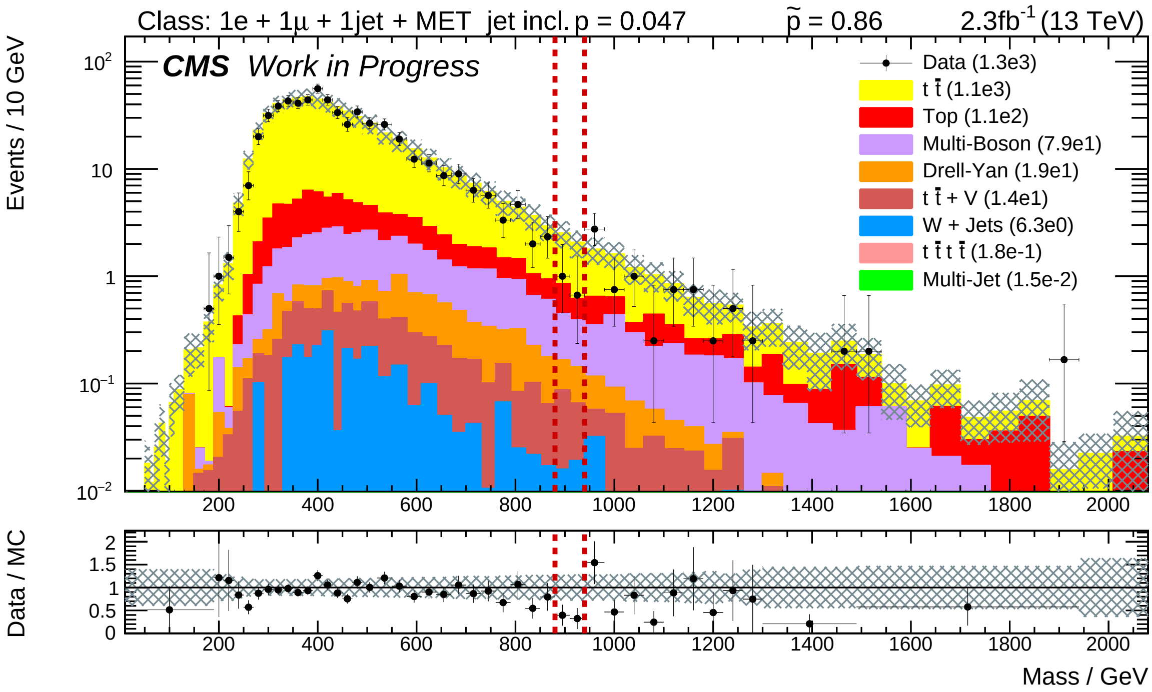


Figure 6.3.: Transverse mass spectrum of the $1e + 1\mu + 1jet + E_T^{\text{miss}}$ jet-inclusive event class.

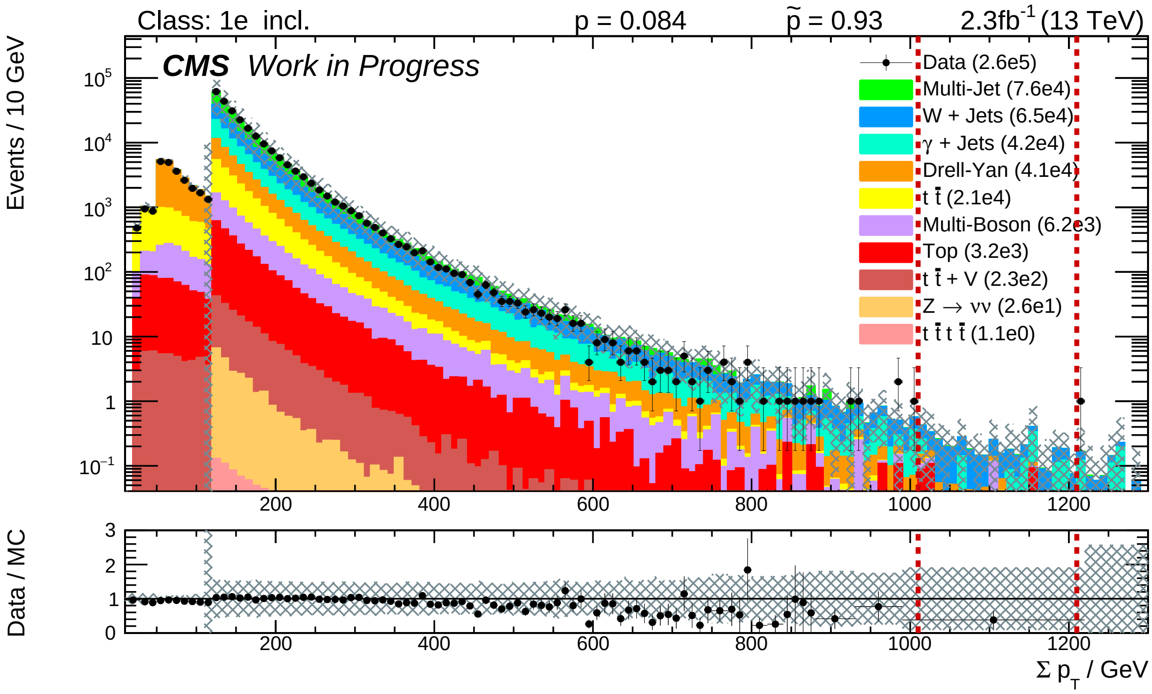


Figure 6.4.: $\sum |p_T|$ spectrum of the one electron inclusive event class.

The lowest trigger threshold for electron triggers is 50 GeV for the double electron trigger. If multiple particles of the same kind are present in an event which is sorted into an event class with a lower multiplicity of this particle the particle with the higher p_T is always taken. Therefore events below 50 GeV must be triggered by muon triggers since the electron has then only to pass the object identification requirement of $p_T > 25$ GeV. The dominant SM process which yields prompt, isolated electrons and muons is the $t\bar{t}$ process[2] which dominates the region between 20 GeV and 50 GeV. The first bin has less events than the following bins as the p_T threshold of 25 GeV is at the bin center and not the lower edge.

The threshold of the double electron trigger is at $p_T > 50$ GeV. This explains the first step in the distribution as above this threshold electron triggered events can enter the distribution. The dominant SM process yielding two leptons is the Drell-Yan process which dominates the region between the two steps at 50 GeV and 120 GeV.

The last step in the distribution at 120 GeV corresponds to the trigger threshold of the single electron trigger. It enables significant contribution of the $W +$ Jets process which yields only one prompt lepton and the multijet and $\gamma +$ Jets process which both do not yield prompt electrons but have a large cross section compared to processes with prompt leptons. The electrons identified in these processes are misidentified photons or jets and are dominant in this event class as no other leptons or photons are required which would require a second misidentified object and suppress these SM processes. The uncertainty in the region above 120 GeV is much greater than below this threshold due to the dominance of the cross section uncertainty above 120 GeV. The cross section of the multi-jet and γ+Jets process is only simulated at LO precision in contrast to the Drell-Yan and $t\bar{t}$ MC simulation with precisions of (N)NLO. The large uncertainty in the bin at 110 GeV just below the step is due to the electron energy scale which shifts electrons from the bin at 120 GeV (which has a lot more events) downwards.

The deviation found in the high energy tail of the distribution is not significant with $\tilde{p} = 0.93$.

6.1.5. Scans of Integral Distributions

A global overview over the normalization and general data and MC agreement can be achieved with a statistical analysis of the integral distributions. Here, only one region containing the complete spectrum is considered. The result of the analysis of the integral distributions of the 15 exclusive, jet-inclusive and inclusive event classes with the highest number of observed events can be found in figures 6.5 to 6.7. One can see that the agreement between data and MC simulation fits within the systematic uncertainties resulting in insignificant p-values.

This validates the general agreement between data and MC simulation and enables a further analysis of the event classes. The most significant deviations of the integral distributions are not shown here because they are correlated to the deviations found by the full scan of all possible regions.

The presented tests and examples validate that the MC simulation describes the known and dominant SM processes sufficiently well. Thus it is possible to proceed with a full scan of all possible regions in all distributions to look for deviations caused by new physics in a small subset of all regions.

---

[2] The event class of one electron, one muon and one additional jet and E_T^{miss} is shown in section 6.1.3.

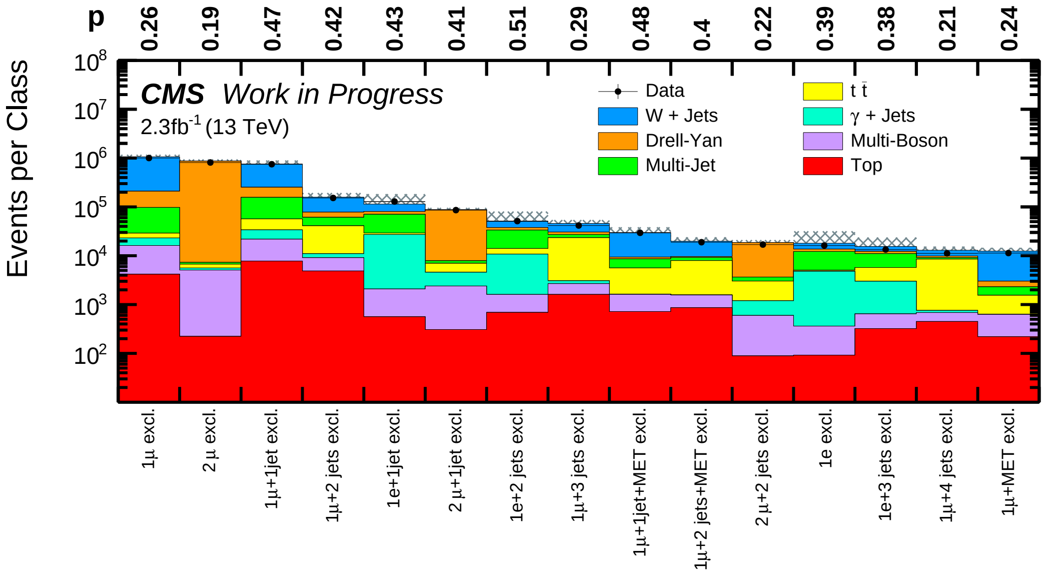


Figure 6.5.: Result of the statistical analysis of the integral $\sum |p_T|$ distribution of the 15 exclusive event classes with the highest event yield. Each bin represents the integral of data, MC simulation and systematic uncertainty for one event class. The corresponding p-value can be found on top.

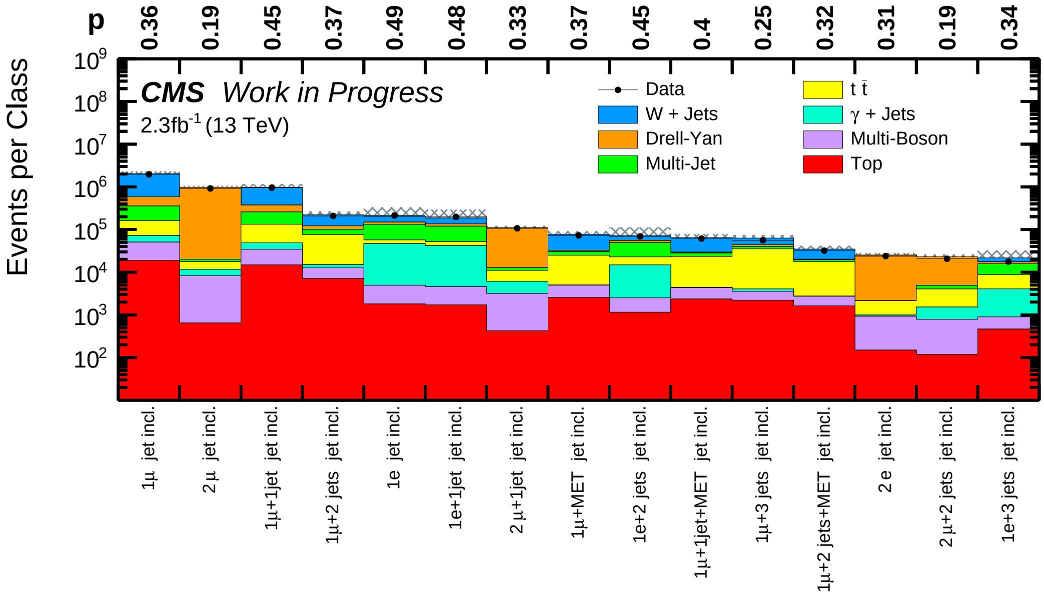


Figure 6.6.: Result of the statistical analysis of the integral $\sum |p_T|$ distribution of the 15 jet-inclusive event classes with the highest event yield. Each bin represents the integral of data, MC simulation and systematic uncertainty for one event class. The corresponding p-value can be found on top.

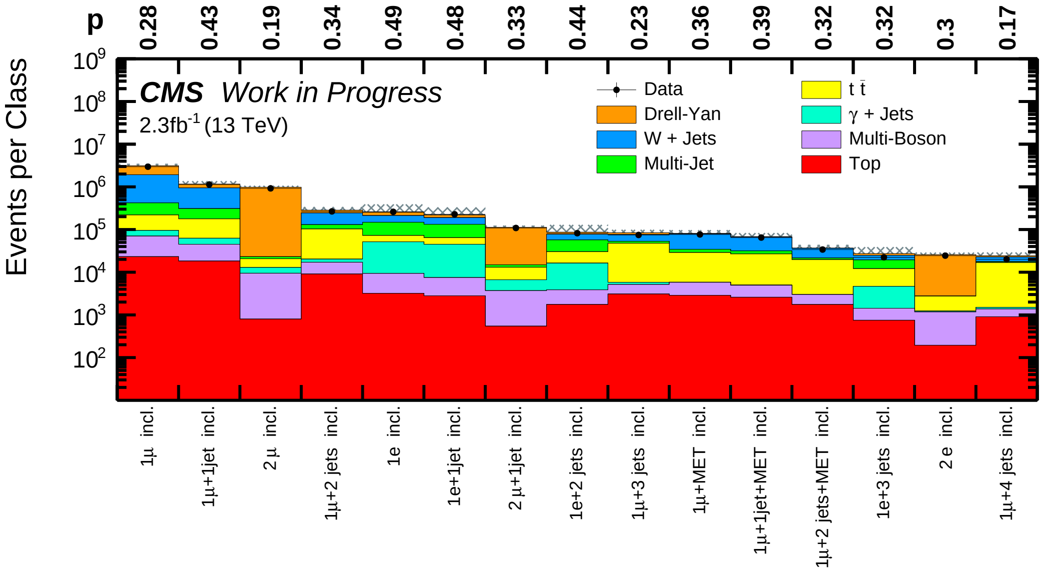


Figure 6.7.: Result of the statistical analysis of the integral $\sum |p_T|$ distribution of the 15 inclusive event classes with the highest event yield. Each bin represents the integral of data, MC simulation and systematic uncertainty for one event class. The corresponding p-value can be found on top.

6.2. Result of the Region Scan

In this section, the result of the full MUSiC analysis chain introduced in chapter 4 is presented. The $\tilde{p}$-distributions, introduced in section 4, and the most significant deviations for all types of distributions and the exclusive, jet-inclusive and fully inclusive event classes are shown and discussed. Finally, the most significant deviations across all distributions and event class types will be discussed in detail.

The number of jets in an event with large jet multiplicities is primarily determined by the soft gluon splitting in the parton showering (see 1.8.2) and is known to be described insufficiently in simulation from other analyses [118]. A threshold of six or more jets is therefore applied to all event classes. Events associated to exclusive classes with more than six jets are combined in the corresponding 6jets jet-inclusive class. These jet inclusive classes are also considered as exclusive in the following[3].

6.2.1. $\tilde{p}$ Distributions

In figures 6.8 to 6.16 and tables 6.1 to 6.9 the $\tilde{p}$ distributions and the most significant classes in all three kinematic distributions for exclusive, jet-inclusive and fully inclusive event classes are provided.

---

[3] 6jets jet-inclusive event classes are indeed exclusive since every event in these event classes is not contained in any other considered event class using this definition.

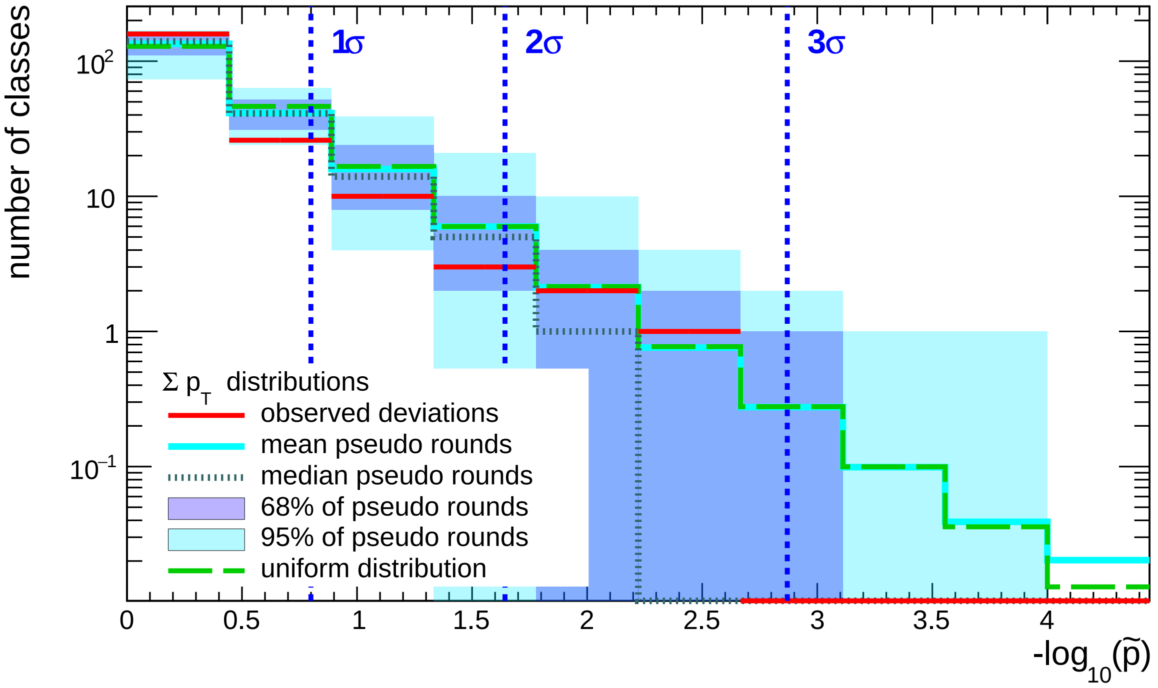


Figure 6.8.: $\tilde{p}$ distribution of the $\sum |p_T|$ kinematic variable for exclusive event classes.

Table 6.1.: Overview of the 15 most significant classes in the $\sum |p_T|$ distribution of exclusive event classes.

Event Class	Global		Region of Interest				
(Exclusive)	n_{Data}	n_{MC}	n_{Data}	n_{MC}	σ_{MC}	Region / GeV	$\tilde{p}$
4μ+1jet	4	1.2	4	0.19	0.11	290 - 360	0.0033
3e+3jets	3	0.58	3	0.17	0.1	360 - 520	0.013
3e+1γ	1	0.032	1	0.0026	0.0024	240 - 270	0.015
1e+2γ+2jets+E_T^{miss}	1	0.034	1	0.015	0.013	570 - 840	0.021
1e+1μ+1γ+2jets	6	3.2	4	0.45	0.25	450 - 510	0.028
2e+1γ+E_T^{miss}	1	0.13	1	0.016	0.021	650 - 920	0.039
4μ+2jets	1	0.3	1	0.0088	0.0069	860 - 980	0.048
3e+E_T^{miss}	4	0.93	2	0.1	0.077	610 - 810	0.049
2e+1μ+2jets+E_T^{miss}	2	0.85	2	0.15	0.064	690 - 840	0.084
2e+1μ+4jets+E_T^{miss}	1	0.078	1	0.025	0.03	660 - 870	0.089
2e+1μ+E_T^{miss}	3	1.5	2	0.15	0.056	490 - 550	0.091
1e+1μ+1γ+1jet+E_T^{miss}	3	1.5	3	0.46	0.24	360 - 460	0.095
2e+1γ+2jets+E_T^{miss}	2	0.48	2	0.2	0.14	440 - 720	0.1
1e+1μ+1γ+1jet	7	4.7	2	0.14	0.06	160 - 190	0.11
1e+2μ	61	91	2	0.064	0.072	670 - 750	0.12

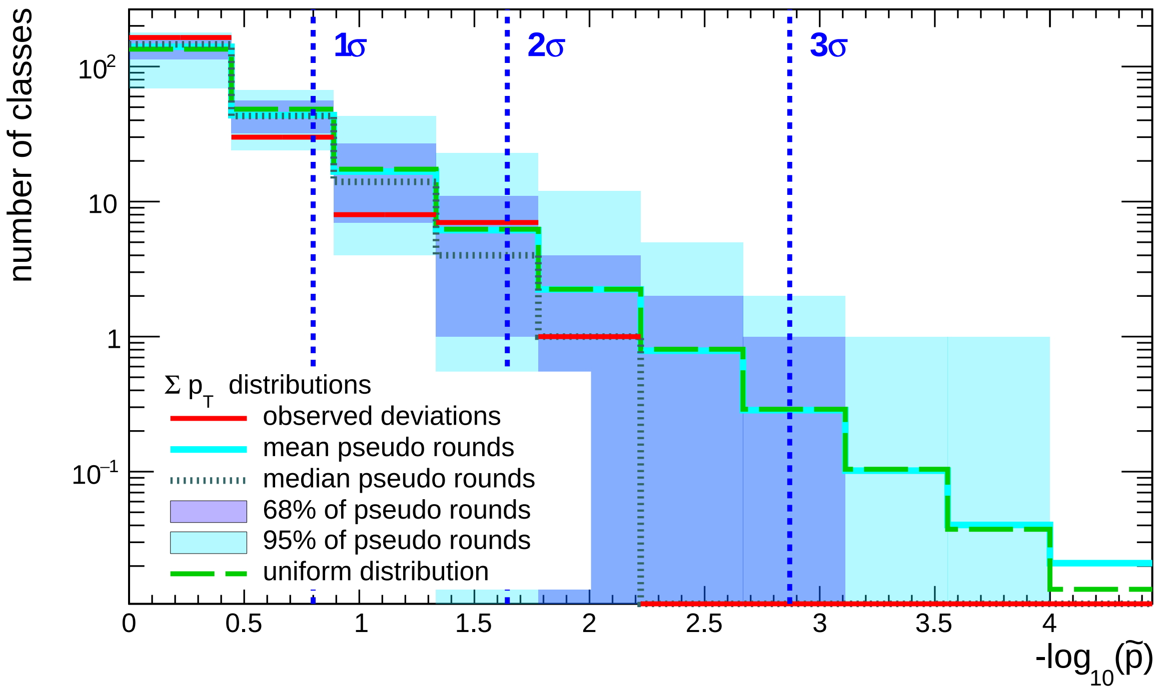


Figure 6.9.: $\tilde{p}$ distribution of the $\sum |p_T|$ kinematic variable for jet inclusive event classes.

Table 6.2.: Overview of the 15 most significant classes in the $\sum |p_T|$ distribution of jet inclusive event classes.

Event Class (Jet Inclusive)	Global		Region of Interest				
	n_{Data}	n_{MC}	n_{Data}	n_{MC}	σ_{MC}	Region / GeV	$\tilde{p}$
4μ+1jet	5	1.6	4	0.24	0.13	290 - 360	0.0066
3e+3jets	3	0.71	3	0.19	0.11	360 - 520	0.017
1e+1μ+1γ+1jet	13	10	3	0.15	0.077	160 - 190	0.021
1e+2γ+2jets+E_T^{miss}	1	0.095	1	0.015	0.011	600 - 720	0.025
3e+E_T^{miss}	8	2.8	4	0.52	0.12	350 - 410	0.035
3μ+1γ+E_T^{miss}	1	0.043	1	0.015	0.02	430 - 550	0.037
1e+2γ+E_T^{miss}	1	0.12	1	0.011	0.015	490 - 590	0.038
2e+1μ+2jets+E_T^{miss}	3	1.3	3	0.26	0.13	690 - 840	0.039
3e+1γ	1	0.07	1	0.016	0.011	220 - 270	0.074
4μ+2jets	1	0.37	1	0.015	0.0094	860 - 980	0.074
3e+2jets	5	3.3	3	0.21	0.19	280 - 360	0.075
1μ+4jets+E_T^{miss}	4516	5.6×10^3	0	8.9	2.6	2500 - 2740	0.089
1μ+5jets+E_T^{miss}	1284	1.8×10^3	0	11	3.6	2550 - 4400	0.097
3e+1jet	16	12	7	1.4	0.82	230 - 270	0.12
1e+2γ+1jet+E_T^{miss}	1	0.11	1	0.053	0.075	590 - 770	0.12

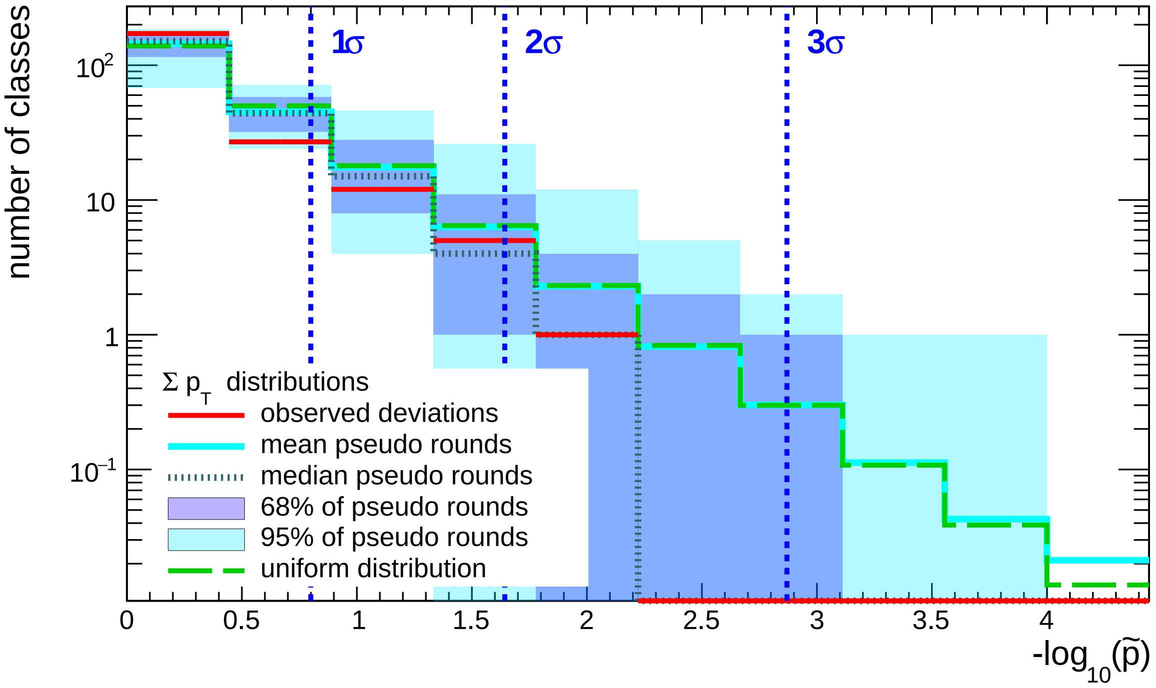


Figure 6.10.: $\tilde{p}$ distribution of the $\sum|p_T|$ kinematic variable for inclusive event classes.

Table 6.3.: Overview of the 15 most significant classes in the $\sum|p_T|$ distribution of inclusive event classes.

Event Class (Inclusive)	Global n_{Data}	Global n_{MC}	Region of Interest n_{Data}	Region of Interest n_{MC}	Region of Interest σ_{MC}	Region / GeV	$\tilde{p}$
4μ+1jet	5	1.6	4	0.25	0.13	290 - 360	0.0076
3e+3jets	3	0.97	3	0.2	0.11	360 - 520	0.023
3μ+1γ+E_T^{miss}	1	0.043	1	0.015	0.02	430 - 550	0.038
1e+1μ+1γ+1jet	18	15	3	0.17	0.083	160 - 190	0.038
3e+E_T^{miss}	8	2.8	4	0.52	0.12	350 - 410	0.039
1e+2γ+E_T^{miss}	1	0.14	1	0.011	0.014	490 - 570	0.046
2e+1μ+2jets+E_T^{miss}	3	1.3	3	0.27	0.13	690 - 840	0.05
1μ+4jets+E_T^{miss}	4716	5.9×10^3	0	9.8	2.7	2500 - 2740	0.059
1μ+3jets+E_T^{miss}	13837	1.6×10^4	6	30	7.6	2290 - 2700	0.077
3e+1γ	1	0.087	1	0.016	0.012	230 - 290	0.08
1μ+5jets+E_T^{miss}	1332	1.9×10^3	0	11	3.7	2550 - 4400	0.083
2e+2jets	3785	4.3×10^3	1	11	2.4	1540 - 2360	0.083
2e+1μ+2jets	6	4.7	5	0.93	0.37	530 - 650	0.097
1e+5jets+E_T^{miss}	379	5.3×10^2	1	14	4.3	1940 - 2180	0.1
4μ+2jets	1	0.39	1	0.021	0.011	860 - 1010	0.11

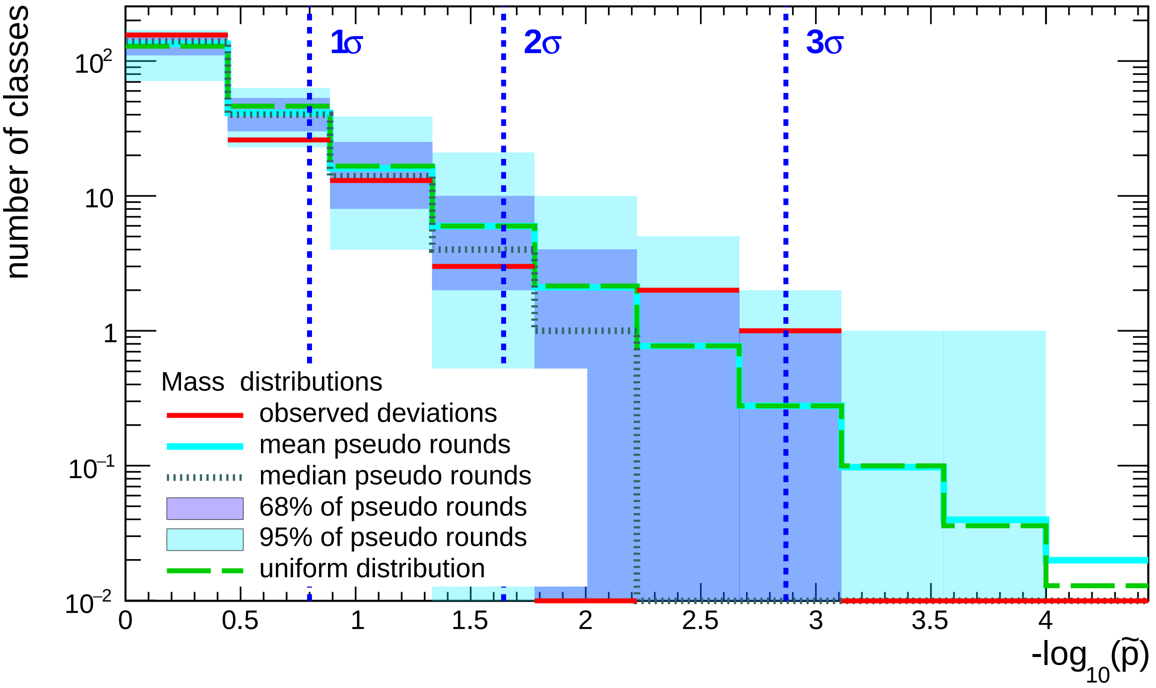


Figure 6.11.: $\tilde{p}$ distribution of the M_{inv} kinematic variable for exclusive event classes.

Table 6.4.: Overview of the 15 most significant classes in the M_{inv} distribution of exclusive event classes.

Event Class	Global		Region of Interest				
(Exclusive)	n_{Data}	n_{MC}	n_{Data}	n_{MC}	σ_{MC}	Region / GeV	$\tilde{p}$
3e+3jets	3	0.58	3	0.07	0.058	540 - 740	0.0013
3e+1γ	1	0.032	1	0.0016	0.0011	270 - 300	0.0047
4μ+1jet	4	1.2	3	0.11	0.02	380 - 420	0.0056
4μ+2jets	1	0.3	1	0.0024	0.0022	2090 - 2600	0.021
1e+1γ+E_T^{miss}	12	11	5	0.39	0.28	270 - 290	0.022
1e+2γ+2jets+E_T^{miss}	1	0.034	1	0.021	0.019	660 - 780	0.03
1e+2μ+4jets	1	0.44	1	0.013	0.01	1450 - 1530	0.057
2e+1γ+E_T^{miss}	1	0.13	1	0.016	0.022	650 - 1700	0.058
3e+E_T^{miss}	4	0.93	2	0.13	0.083	570 - 750	0.065
2e+1μ+4jets+E_T^{miss}	1	0.078	1	0.025	0.029	630 - 870	0.081
2e+1μ+E_T^{miss}	3	1.5	2	0.12	0.042	510 - 550	0.081
2e+4jets	114	1.7×10^2	1	12	2.5	1450 - 1650	0.082
2e+1μ+2jets+E_T^{miss}	2	0.85	2	0.15	0.06	690 - 840	0.086
1e+1jet	128596	1.1×10^5	1	4.6×10^{-5}	5.3×10^{-5}	5240 - 6090	0.11
1e+1μ+1γ+1jet+E_T^{miss}	3	1.5	3	0.44	0.24	360 - 460	0.11

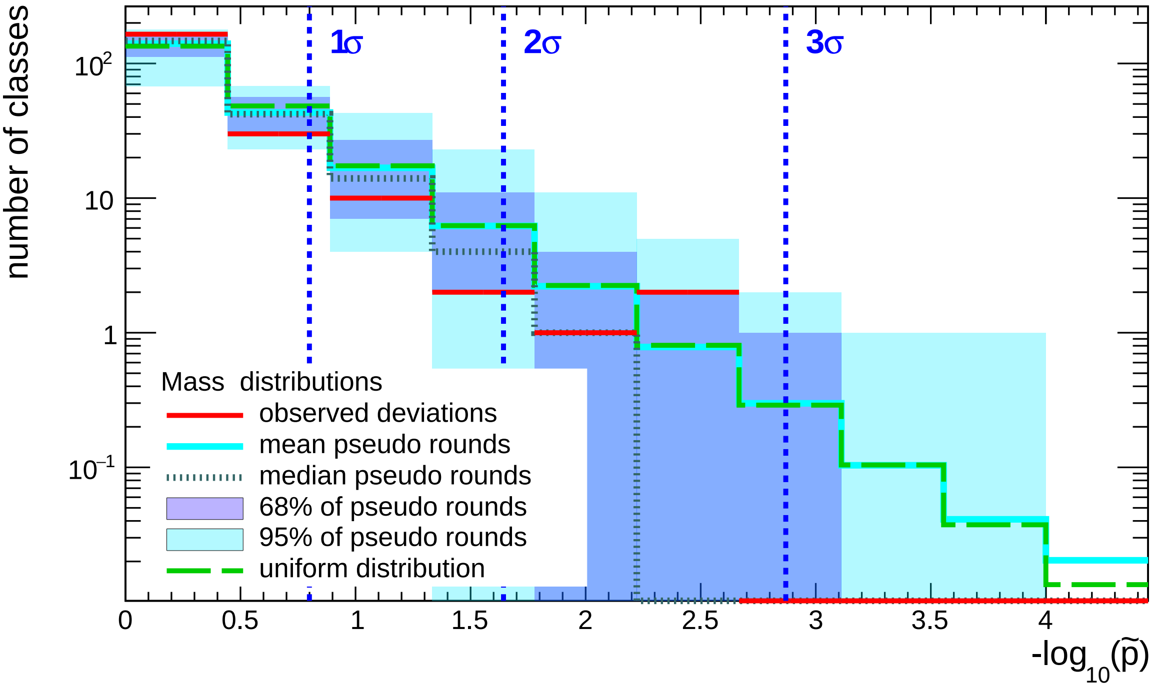


Figure 6.12.: $\tilde{p}$ distribution of the M_{inv} kinematic variable for jet inclusive event classes.

Table 6.5.: Overview of the 15 most significant classes in the M_{inv} distribution of jet inclusive event classes.

Event Class	Global		Region of Interest				
(Jet Inclusive)	n_{Data}	n_{MC}	n_{Data}	n_{MC}	σ_{MC}	Region / GeV	$\tilde{p}$
3e+3jets	3	0.71	3	0.082	0.062	540 - 740	0.0046
3e+1γ	1	0.07	1	0.0019	0.0018	260 - 280	0.0056
4μ+1jet	5	1.6	3	0.13	0.033	380 - 420	0.0095
4μ+2jets	1	0.37	1	0.0019	0.0014	2030 - 2220	0.019
2e+1μ+2jets+E_T^{miss}	3	1.3	3	0.25	0.12	690 - 840	0.041
1e+2γ+E_T^{miss}	1	0.12	1	0.019	0.02	430 - 570	0.048
3e+E_T^{miss}	8	2.8	3	0.2	0.12	610 - 750	0.05
3μ+1γ+E_T^{miss}	1	0.043	1	0.015	0.028	430 - 530	0.05
2μ+2jets	20664	2.3×10^4	0	7.9	1.4	2930 - 7170	0.061
1e+2μ+4jets	1	0.52	1	0.013	0.011	1450 - 1530	0.066
1e+2γ+2jets+E_T^{miss}	1	0.095	1	0.041	0.067	690 - 840	0.1
1μ+4jets+E_T^{miss}	4516	5.6×10^3	0	9	2.6	2500 - 2740	0.12
1e+2γ+1jet+E_T^{miss}	1	0.11	1	0.05	0.067	590 - 740	0.12
2e+1μ+4jets+E_T^{miss}	1	0.11	1	0.025	0.03	840 - 1060	0.13
1e+5jets+E_T^{miss}	358	4.9×10^2	2	21	6.3	1880 - 2240	0.13

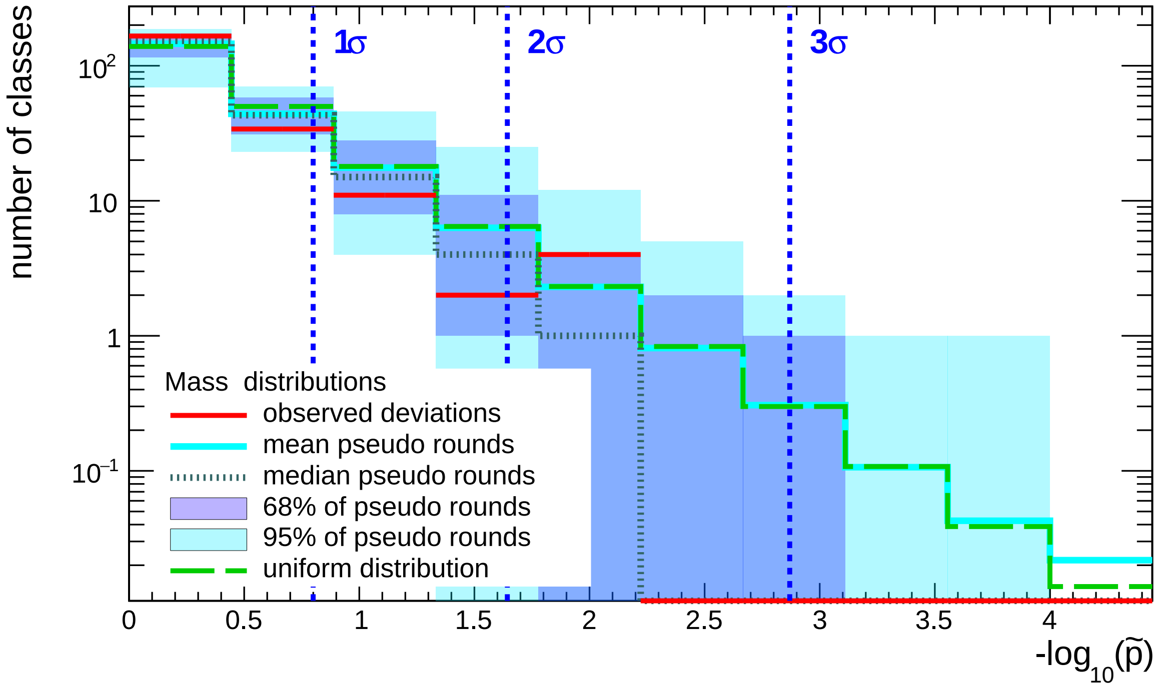


Figure 6.13.: $\tilde{p}$ distribution of the M_{inv} kinematic variable for inclusive event classes.

Table 6.6.: Overview of the 15 most significant classes in the M_{inv} distribution of inclusive event classes.

Event Class (Inclusive)	Global n_{Data}	n_{MC}	Region of Interest n_{Data}	n_{MC}	σ_{MC}	Region / GeV	$\tilde{p}$
3e+1γ	1	0.087	1	0.0019	0.0018	260 - 280	0.0096
4μ+1jet	5	1.6	3	0.13	0.035	380 - 420	0.0097
3e+3jets	3	0.97	3	0.11	0.083	540 - 740	0.012
2μ+2jets	21776	2.4×10^4	0	10	1.8	2930 - 7170	0.014
1e+2γ+E_T^{miss}	1	0.14	1	0.011	0.021	450 - 750	0.045
2e+1μ+2jets+E_T^{miss}	3	1.3	3	0.26	0.13	690 - 840	0.045
3μ+1γ+E_T^{miss}	1	0.043	1	0.015	0.028	430 - 530	0.048
3e+E_T^{miss}	8	2.8	3	0.21	0.13	610 - 750	0.052
1μ+4jets+E_T^{miss}	4716	5.9×10^3	0	9.9	2.8	2500 - 2740	0.087
1μ+4jets	20167	2.4×10^4	3	22	6	3590 - 4300	0.087
3μ+1γ	1	0.63	1	0.014	0.021	490 - 530	0.099
2e+1μ+4jets+E_T^{miss}	1	0.12	1	0.023	0.029	840 - 1020	0.1
1e+5jets+E_T^{miss}	379	5.3×10^2	2	22	6.8	1880 - 2240	0.11
2e+1μ+3jets+E_T^{miss}	1	0.45	1	0.031	0.029	780 - 810	0.12
2e+1γ+2jets+E_T^{miss}	2	1	2	0.17	0.14	460 - 630	0.12

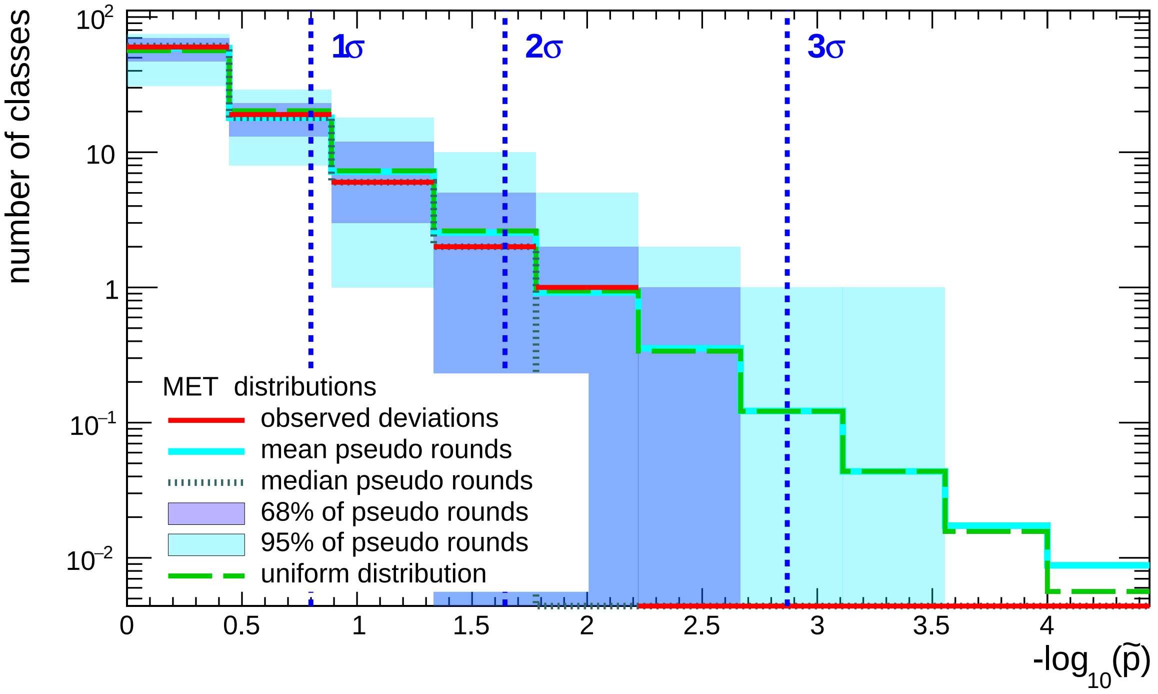


Figure 6.14.: $\tilde{p}$ distribution of the E_T^{miss} kinematic variable for exclusive event classes.

Table 6.7.: Overview of the 15 most significant classes in the E_T^{miss} distribution of exclusive event classes.

Event Class	Global		Region of Interest				
(Exclusive)	n_{Data}	n_{MC}	n_{Data}	n_{MC}	σ_{MC}	Region / GeV	$\tilde{p}$
3e+E_T^{miss}	4	0.93	3	0.17	0.058	170 - 290	0.006
3μ+1jet+E_T^{miss}	4	5	2	0.058	0.029	350 - 410	0.019
1e+2γ+2jets+E_T^{miss}	1	0.034	1	0.033	0.038	100 - 290	0.045
2e+1γ+E_T^{miss}	1	0.13	1	0.022	0.021	160 - 210	0.049
2e+1μ+4jets+E_T^{miss}	1	0.078	1	0.033	0.027	130 - 170	0.054
3μ+2jets+E_T^{miss}	3	1.9	3	0.4	0.2	130 - 160	0.081
1μ+1γ+2jets+E_T^{miss}	71	79	1	0.0051	0.0035	890 - 980	0.087
2e+1μ+E_T^{miss}	3	1.5	2	0.18	0.073	210 - 270	0.089
3μ+E_T^{miss}	3	3.7	1	0.013	0.024	390 - 470	0.12
2μ+6jets+E_T^{miss}	2	9.6	0	8.4	3.5	110 - 490	0.14
3e+1jet+E_T^{miss}	3	1.1	2	0.26	0.086	150 - 230	0.15
1e+1μ+5jets+E_T^{miss}	11	23	0	9.1	3.6	170 - 810	0.15
2e+1γ+2jets+E_T^{miss}	2	0.48	2	0.31	0.18	120 - 210	0.15
1e+2μ+2jets+E_T^{miss}	3	1.4	2	0.21	0.12	170 - 230	0.15
2e+1μ+2jets+E_T^{miss}	2	0.85	1	0.036	0.033	270 - 410	0.16

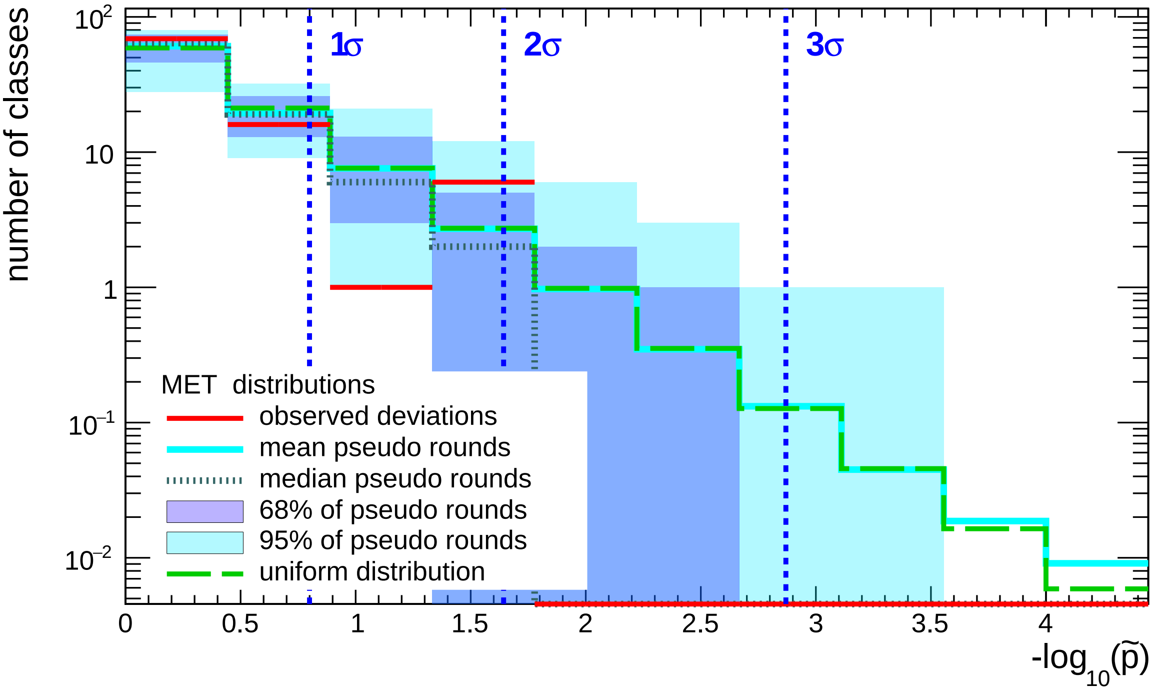


Figure 6.15.: $\tilde{p}$ distribution of the E_T^{miss} kinematic variable for jet inclusive event classes.

Table 6.8.: Overview of the 15 most significant classes in the E_T^{miss} distribution of jet inclusive event classes.

Event Class (Jet Inclusive)	Global n_{Data}	n_{MC}	Region of Interest n_{Data}	n_{MC}	σ_{MC}	Region / GeV	$\tilde{p}$
1e+2γ+2jets+E_T^{miss}	1	0.095	1	0.012	0.013	210 - 290	0.025
3μ+E_T^{miss}	11	12	3	0.18	0.14	350 - 410	0.028
1e+2γ+E_T^{miss}	1	0.12	1	0.017	0.019	210 - 290	0.031
1e+2γ+1jet+E_T^{miss}	1	0.11	1	0.017	0.019	210 - 290	0.031
3e+E_T^{miss}	8	2.8	3	0.27	0.095	210 - 290	0.036
3μ+1γ+E_T^{miss}	1	0.043	1	0.016	0.029	90 - 120	0.038
2e+1μ+4jets+E_T^{miss}	1	0.11	1	0.036	0.02	110 - 140	0.11
1e+1μ+5jets+E_T^{miss}	13	31	0	10	4.1	190 - 830	0.13
2μ+6jets+E_T^{miss}	2	9.6	0	8.4	3.5	110 - 490	0.14
2e+1γ+E_T^{miss}	3	1.7	2	0.32	0.16	160 - 210	0.15
3μ+1jet+E_T^{miss}	8	8.1	2	0.17	0.14	350 - 410	0.16
2e+1μ+E_T^{miss}	6	4.5	3	0.59	0.16	210 - 290	0.2
2e+1μ+2jets+E_T^{miss}	3	1.3	1	0.065	0.049	270 - 330	0.27
1e+1jet+E_T^{miss}	10552	1.1×10^4	9	28	7.1	530 - 590	0.28
3μ+2jets+E_T^{miss}	4	3.1	3	0.83	0.33	120 - 150	0.29

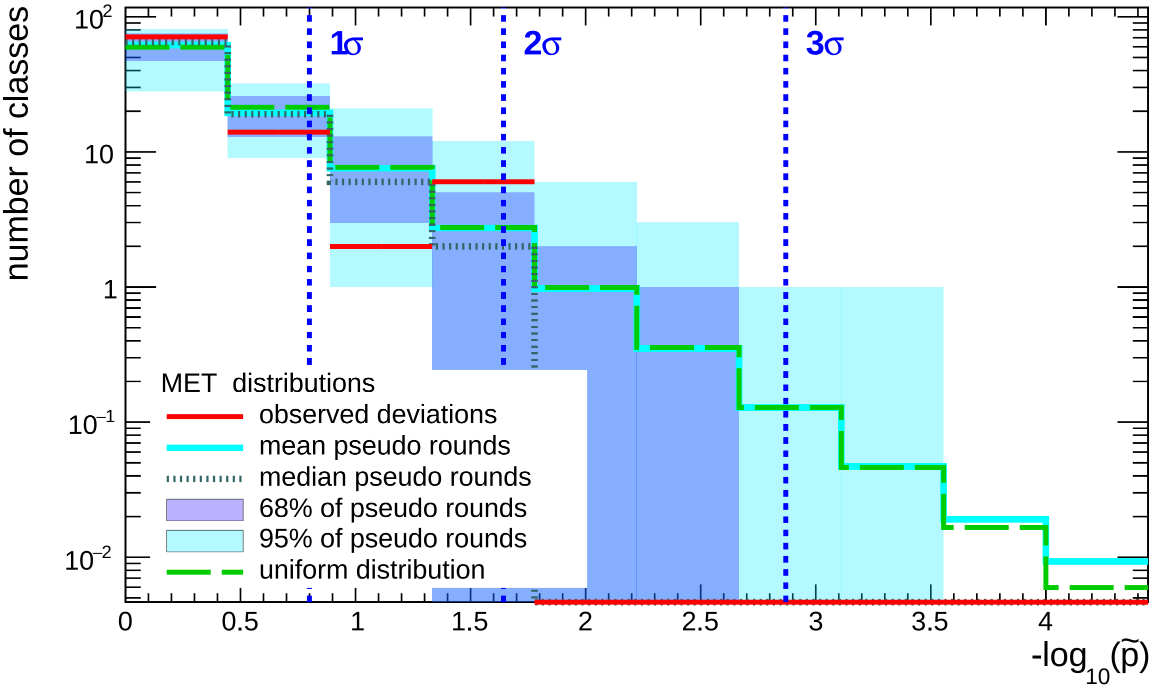


Figure 6.16.: $\tilde{p}$ distribution of the E_T^{miss} kinematic variable for inclusive event classes.

Table 6.9.: Overview of the 15 most significant classes in the E_T^{miss} distribution of inclusive event classes.

Event Class (Inclusive)	Global		Region of Interest				
	n_{Data}	n_{MC}	n_{Data}	n_{MC}	σ_{MC}	Region / GeV	$\tilde{p}$
$3\mu+E_T^{\text{miss}}$	12	12	3	0.18	0.14	350 - 410	0.029
$1e+2\gamma+E_T^{\text{miss}}$	1	0.14	1	0.017	0.019	210 - 290	0.035
$1e+2\gamma+1\text{jet}+E_T^{\text{miss}}$	1	0.13	1	0.017	0.019	210 - 290	0.037
$3e+E_T^{\text{miss}}$	8	2.8	5	0.99	0.26	150 - 290	0.037
$1e+2\gamma+2\text{jets}+E_T^{\text{miss}}$	1	0.11	1	0.02	0.033	170 - 290	0.039
$3\mu+1\gamma+E_T^{\text{miss}}$	1	0.043	1	0.016	0.029	90 - 120	0.045
$2e+1\mu+4\text{jets}+E_T^{\text{miss}}$	1	0.12	1	0.038	0.02	110 - 140	0.11
$1e+1\mu+5\text{jets}+E_T^{\text{miss}}$	13	31	0	11	4.1	190 - 830	0.12
$2\mu+6\text{jets}+E_T^{\text{miss}}$	2	9.7	0	8.5	3.5	110 - 490	0.13
$2e+1\gamma+E_T^{\text{miss}}$	3	1.8	2	0.33	0.16	160 - 210	0.16
$3\mu+1\text{jet}+E_T^{\text{miss}}$	8	8.3	2	0.18	0.14	350 - 410	0.17
$2e+1\mu+E_T^{\text{miss}}$	6	4.6	3	0.61	0.16	210 - 290	0.22
$2e+1\mu+2\text{jets}+E_T^{\text{miss}}$	3	1.3	1	0.066	0.049	270 - 330	0.28
$2e+1\gamma+2\text{jets}+E_T^{\text{miss}}$	2	1	1	0.12	0.071	190 - 250	0.29
$3\mu+2\text{jets}+E_T^{\text{miss}}$	4	3.2	3	0.85	0.32	120 - 150	0.29

The $\tilde{p}$ distributions of all kinematic distributions of all event classes show good agreement between the expected deviation of the MC simulation and the uniform distribution and reasonable 68 % and 95 % quantiles. This validates the complex concept with dicing of pseudo data of the $\tilde{p}$ calculation, the accurate description of systematic uncertainties, and allows to indeed interpret the $\tilde{p}$-value as a statistical p-value. It also shows good agreement between the $\tilde{p}$-values observed in data with the $\tilde{p}$-values calculated using MC simulation. All deviations are contained in the 95 % quantile of the MC simulation. The most significant deviations across all distributions and event classes has a $\tilde{p}$-value of 0.0013 which corresponds to a significance of 3.0σ and is well contained in the 95 % quantile of the MC expectation. It is important to note, that the significance of 3σ includes only the look-elsewhere correction for this distribution and no corrections for the look-elsewhere effects introduced by the amount of event classes and different distributions (see section 4.3.3). In summary, it can be stated that the MUSiC algorithm is working and reproduces the expected behaviour of a p-value with the pseudo data dicing. No evidence of new physics is found in the measured data.

The most significant classes are dominated by multi-boson processes and are discussed in detail in section 6.2.2. One event class is entirely skipped due to the region veto algorithm as no region can be found which passes the criteria described in section 4.3.1. It is discussed in section 6.2.6.

Among the less significant deviations are a lot of deficits in $t\bar{t}$ dominated final states like the $\sum |p_T|$ spectrum of the $1\mu + 5\text{jets} + E_T^{\text{miss}}$ jet inclusive event class with zero observed events compared to an expectation of 10 ± 3.6 events. The p-value is relatively insignificant due to the high renormalization and factorization scale uncertainty (see section 5.4.5). Deficits are observed in most $t\bar{t}$ dominated final states at high jet multiplicities but are mostly covered by this uncertainty.

6.2.2. Significant Multi-Boson Dominated Event Classes

At first sight it might seem improbable that the most significant deviations are found in event classes dominated by the same physics processes but considering the correlations of the processes across one process group this behaviour is not unexpected. It is also important to remark that the deviations are in the range of expected deviations. However, a closer look is taken at the $4\mu + 1\text{jet}$, $3e + 3\text{jets}$ and the $3e + 1\gamma$ event class as these three event classes contain the most significant deviations with $\log_{10} \tilde{p} > 2.3$.

In 2012, several 2σ deviations in di-boson searches were observed with the $\sqrt{s} = 8$ TeV dataset at masses around 2 TeV [119] but were not confirmed in recent analyses [120]. Nevertheless, excesses in di-boson dominated final states are also predicted by many theoretical models [121, 122] and are interesting final states to search for new physics.

6.2.3. $3e + 3\text{jets}$ Exclusive Event Class

The most significant deviation is found in the M_{inv} spectrum of the $3e + 3\text{jets}$ exclusive event class (see figure 6.17) with a $\tilde{p}$-value of 0.0013 corresponding to a significance of 3σ. It is dominated by the WZ process. Three data events are found nearby to each other in the mass range of $540 - 740$ GeV with only 0.07 ± 0.06 expected MC events but no resonant structure is observed.

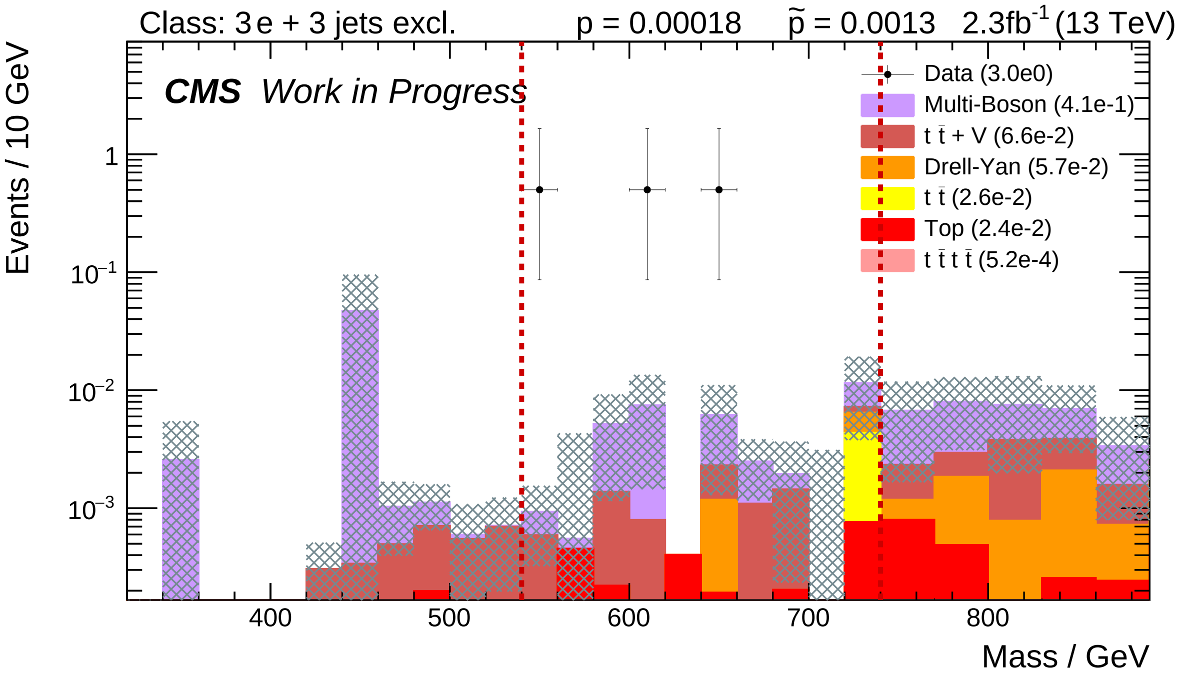


Figure 6.17.: Invariant mass spectrum of the $3e + 3$jets exclusive event class.

In the integral there are also three data events with 0.6 ± 0.4 MC events which is still an observed excess but with a much higher p-value of 0.037.

It is interesting to see if the excess persists when additional objects are allowed. Taking a look at the inclusive event class, the excess becomes less significant as well with a $\tilde{p}$-value of 0.012 as no additional data enter the event class but the MC prediction increases to 0.11 ± 0.08 and 1.0 ± 0.6 overall. As expected the jet inclusive event class positions itself between the exclusive and inclusive event class with a $\tilde{p}$-value of 0.0046 as it contains no additional data events and more MC than the exclusive event class and less than the inclusive event class.

In addition to the inclusive event classes it is interesting to take a look at the event classes with the other lepton flavour since deviations are expected in both cases if the leptons are produced in flavour universal processes like the SM di-boson processes. The $3\mu + 3$jets exclusive event class shows an excess of 5 events in data and an expectation of 1.5 ± 0.6 events between 460 GeV and 980 GeV but is less significant with $\tilde{p} = 0.224$. The combined significance of the integral distribution of the $3e + 3$jets and the $3\mu + 3$jets exclusive event class is approximately 2.0σ.

The 8 TeV dataset taken in 2012 shows good agreement between data and MC simulation in the $3e + 3$jets class with $\tilde{p} = 0.32$ as well as in the $3\mu + 3$jets class with $\tilde{p} = 0.12$ [66].

6.2.4. $4\mu + 1$jet Exclusive Event Class

The second most significant deviation with a $\log_{10} \tilde{p}$ value of 2.5 ($\tilde{p} = 0.0033$) is found in the $\sum |p_T|$ distribution of the $4\mu + 1$jet exclusive event class (see figure 6.18). It shows an excess in data with four observed events in the energy range of $290 - 360$ GeV with an expectation of 0.2 ± 0.1

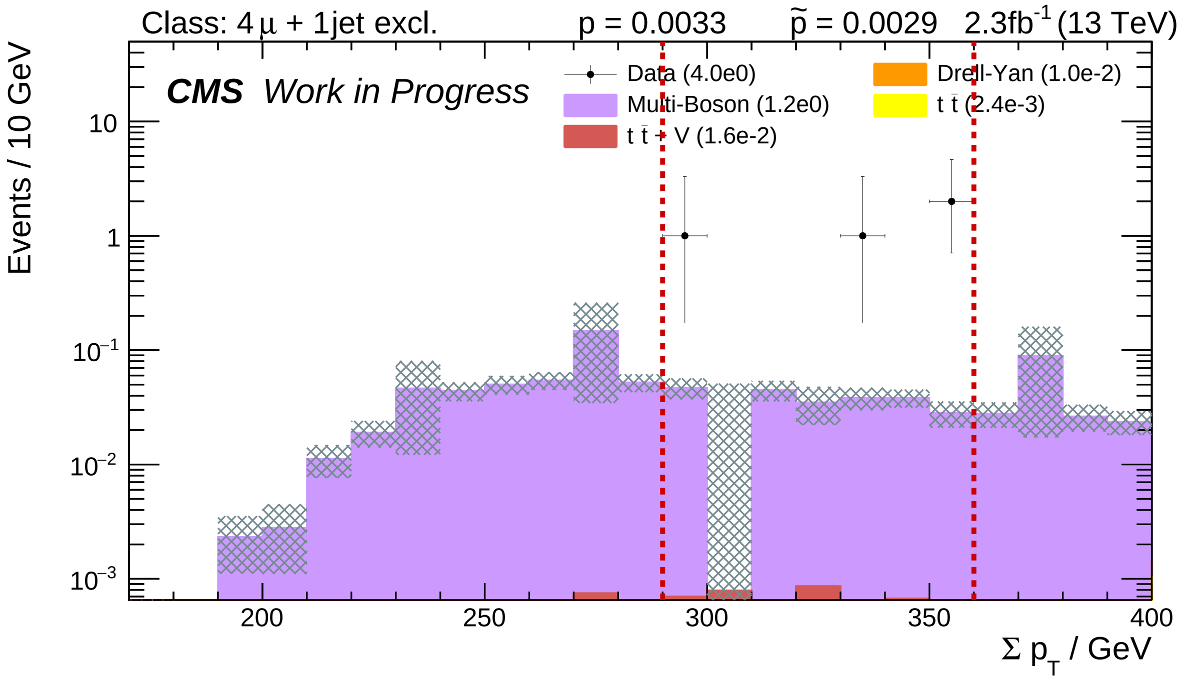


Figure 6.18.: $\sum |p_T|$ spectrum of the $4\mu + 1$jet exclusive event class.

events. The dominant process in the event class is the ZZ process. In the integral distribution no additional data event is found but the expectation increases to 1.2 ± 0.4 events yielding a p-value of 0.05.

The inclusive event class shows an excess in data as well with a $\tilde{p}$-value of 0.0069 which is not much less significant than the excess in the exclusive event class.

The corresponding classes with electrons instead of muons in the decay chain show no excess in data as no events are observed in the $4e + 1$jet exclusive event class as well as in the $2e + 2\mu + 1$jet exclusive event class which fits the MC expectation in these classes.

In 2012 the data showed good agreement with the MC simulation with $\tilde{p} = 0.66$ [66].

6.2.5. $3e + 1\gamma$ Exclusive Event Class

The third most significant excess is found in the invariant mass distribution of the $3e + 1\gamma$ exclusive event class (see figure 6.19) with a $\tilde{p}$-value of 0.0047 which is dominated by the ZZ process. It has a single data event at 270 GeV and a SM expectation of 0.002 ± 0.001 in the region of interest and 0.03 ± 0.03 in total. The inclusive event class does not yield any more data events but the MC prediction is increased leading to a less significant $\tilde{p}$-value of 0.099. In the muon channel no event is observed, which fits the SM expectation, but one event with additional E_T^{miss} is observed and discussed in section 6.2.6.

In 2012 one event was found at 200 GeV leading to an insignificant $\tilde{p}$-value of 0.20 [66].

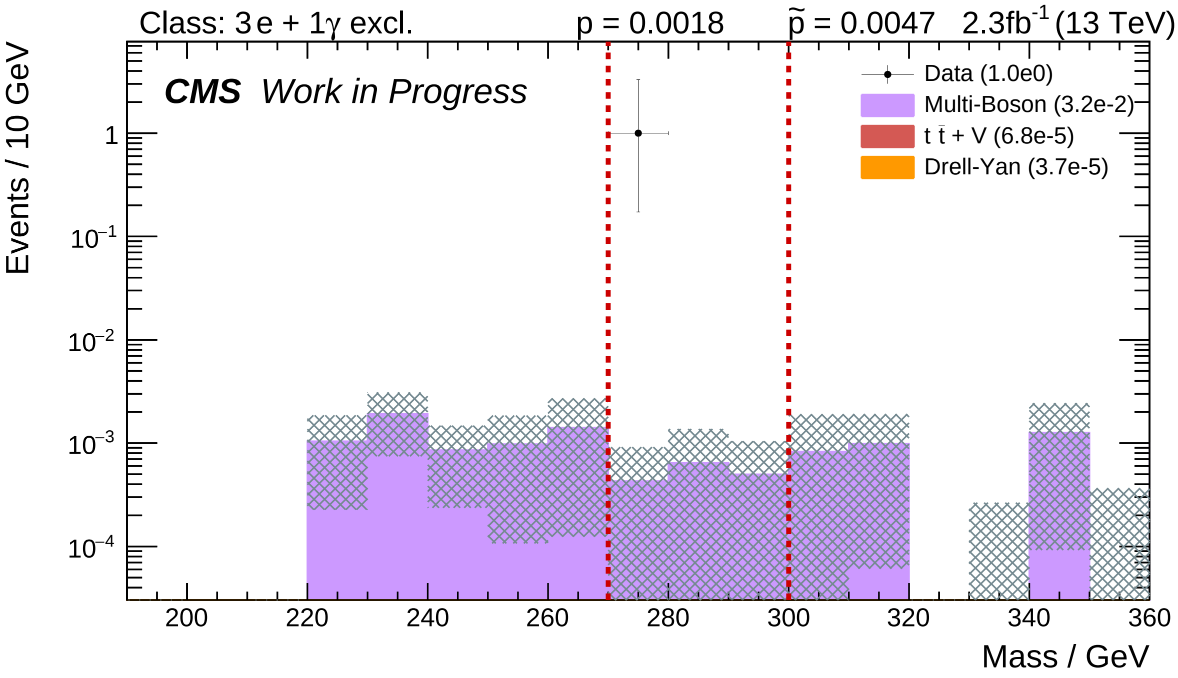


Figure 6.19.: Invariant mass spectrum of the $3e + 1\gamma$ exclusive event class.

6.2.6. Skipped Event Classes

Due to the region veto algorithm described in section 4.3.1, all possible regions and even the integral region of the $3\mu + 1\gamma + E_T^{\text{miss}}$ exclusive event class are discarded. This is the only event class which is skipped entirely due to the region veto algorithm, contains data events, and is skipped in all three distributions. The $\sum |p_T|$ spectrum can be found in figure 6.20. The reasons why each region was skipped can be found in figure 6.21.

In total the event class contains one data event and a SM expectation of 0.01 ± 0.02 which would yield a p-value of 0.02 corresponding to a significance of 2σ. However, the SM description is not sufficient to trust this value. The fact that an excess is found in both the electron and the muon channel (see section 6.2.5) is still interesting and will be monitored in the future to see if the excesses persist.

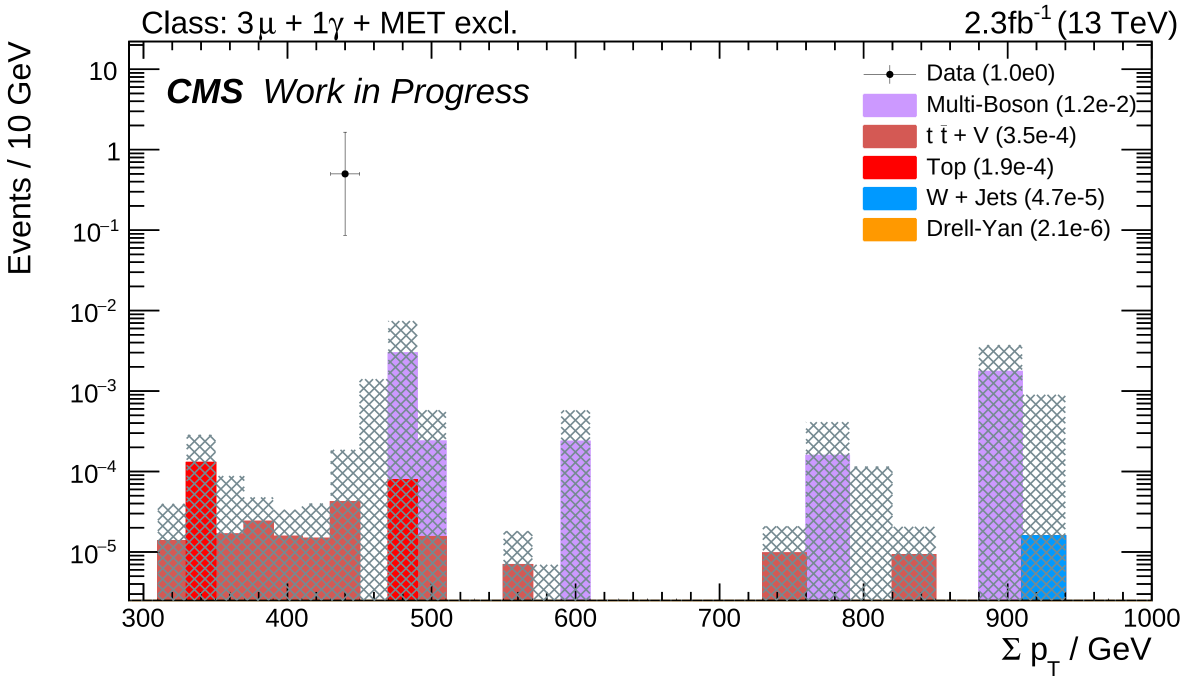


Figure 6.20.: $\sum |p_T|$ distribution of the $3\mu + 1\gamma + E_T^{\text{miss}}$ exclusive event class which is the only event class which is skipped entirely by the region veto algorithm.

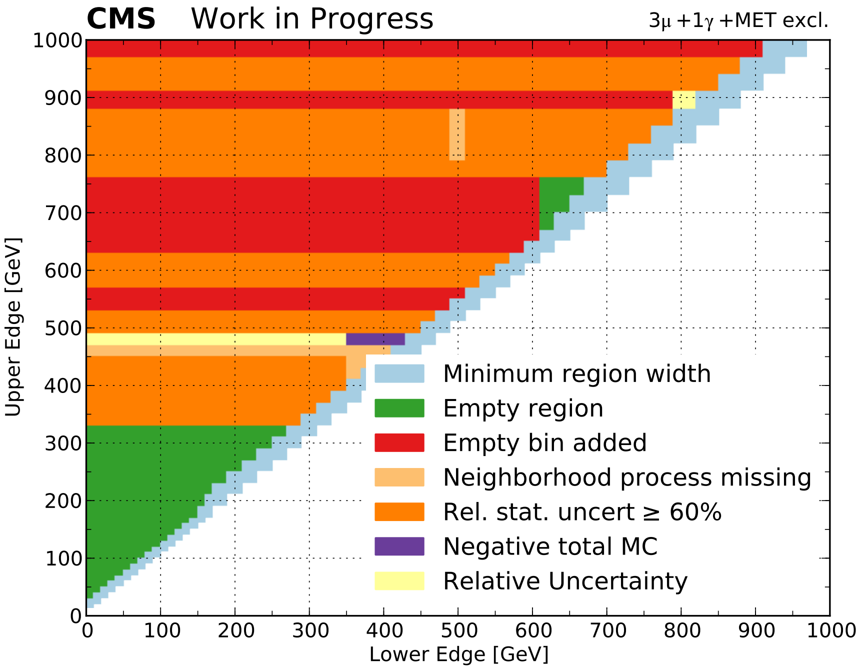


Figure 6.21.: Skip reasons for each possible region in the $3\mu + 1\gamma + E_T^{\text{miss}}$ exclusive event class. The x-axis denotes the lower edge of the region and the y-axis the upper edge.

7. First Look at the 2016 Dataset

The analysis of the 2016 dataset offers with a validated, integrated luminosity of $36.46\,\mathrm{fb}^{-1}$ a much greater dataset than the data taken in 2015. However the analysis suffers from some issues which were not fixable within a reasonable time frame for this thesis. For a future analysis of the 2016 dataset all of these issues can be solved. The most severe is a lacking MC simulation for the 2016 data for the following reasons:

- A new version of the reconstruction algorithms and L1 trigger between 2015 and 2016 requires a new version of MC simulation.
- The trigger menu for the 2016 MC simulation production was not finished before the first MC samples were submitted and include no trigger information due to the new assembled L1 trigger in the year end technical shutdown between 2015 and 2016. There is no feasible way to simulate the trigger efficiency for the MUSiC analysis as a logical OR of five different triggers with different efficiencies is used. It would have been possible to reduce the number of trigger to only one trigger, but that would drastically limit the analyzed dataset. In the future all MC samples will be re-produced with trigger information.
- The pile-up distributions between the 2015 and 2016 data are vastly different due to the increased intensity per bunch and the reduction of the emittance due to BCMS (beam compression merge splitting) beams first used in the `Run2016E`. The pile-up distributions of both years and the 2015 MC samples can be found in figure 7.1. One can see, that the mean pile-up matches nicely for the 2015 data and MC simulation but also that in 2016 the mean is shifted to much higher pile-up. As now MC events are simulated with such pile-up values, pile-up weights can get enormously high or it is even not possible to reweight the MC simulation anymore.
- The dynamic strip inefficiency was observed in 2015 but was a much more severe issue at higher pile-up such as in the 2016 data. Highly ionizing particles lead to a saturation of read-out boards in the tracker which were then blinded for the next bunch crossings. This lead to a decrease in the tracking efficiency which was not simulated in the MC simulation. Meanwhile this issue is fixed by changing the hardware settings [123] and mitigation issues for data taken previous to the fix will be applied in a re-reconstruction of the 2016 data.
- The new L1 trigger showed some inefficiencies for both high energetic electrons and muons for several reasons. These inefficiencies are not simulated in the MC simulation as well but are meanwhile also fixed and a future MC reprocessing will simulate this inefficiency correctly.

Another issue of the 2016 data is the inhomogeneous dataset since not all datasets taken in 2016 have the same reconstruction settings as e.g. the dynamic strip inefficiency was solved as described above since `Run2016G` [124] and the alignment of the detector was measured with

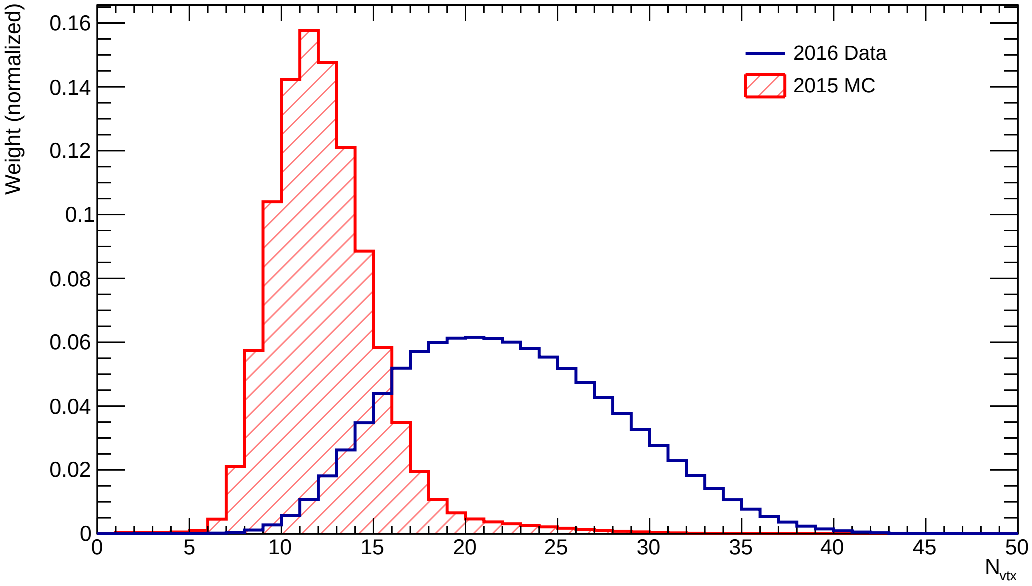


Figure 7.1.: Comparison of the pileup distributions of the 2015 MC simulation in red and the 2016 data of the eras `Run2016B-G` in blue. Both distributions are normalized to unity. The overlap between both distributions is relatively small.

improved precision and taken into account for the reconstruction since `Run2016E` [124]. This issue will also be solved with a re-reconstruction of all data which is expected by the end of 2016.

Due to the higher instantaneous luminosity, parts of the 2015 trigger set cannot be used for the analysis of the 2016 data. This is especially a problem for the `Run2016H` dataset which uses new double electron triggers which were not included in the 2015 MC simulation. Therefore, only the `Run2016B-G` datasets with an integrated luminosity of 26.4 fb^{-1} are used. The `Run2016A` dataset was taken without the magnetic field of the solenoid. In addition the single electron trigger uses a higher trigger threshold of 115 GeV instead of 105 GeV.

7.1. Search for Mis-reconstructed Events

Despite these problems all data were analyzed with the MUSiC classification in order to detect mis-reconstructed events as no MC simulation is required e.g. to pin down single high energetic events. As the event has to be balanced in the transverse plane all reconstruction defects have to show up in the missing transverse energy as well. Therefore a close look is taken to the inclusive $1e + E_T^{\text{miss}}$ and $1\mu + E_T^{\text{miss}}$ event classes which can be found in figure 7.2 and 7.3. The MC simulation is the same as for the 2015 dataset scaled to the luminosity measured in 2016. In the muon channel a lot of high energetic events can be found and were reported to the Physics Object Groups in order to improve the reconstruction.

As an example the event display in figure 7.4 shows an issue that a short track with hits only in the three pixel layers is reconstructed as a multi- TeV muon and a second muon with a $p_T = 60$ GeV

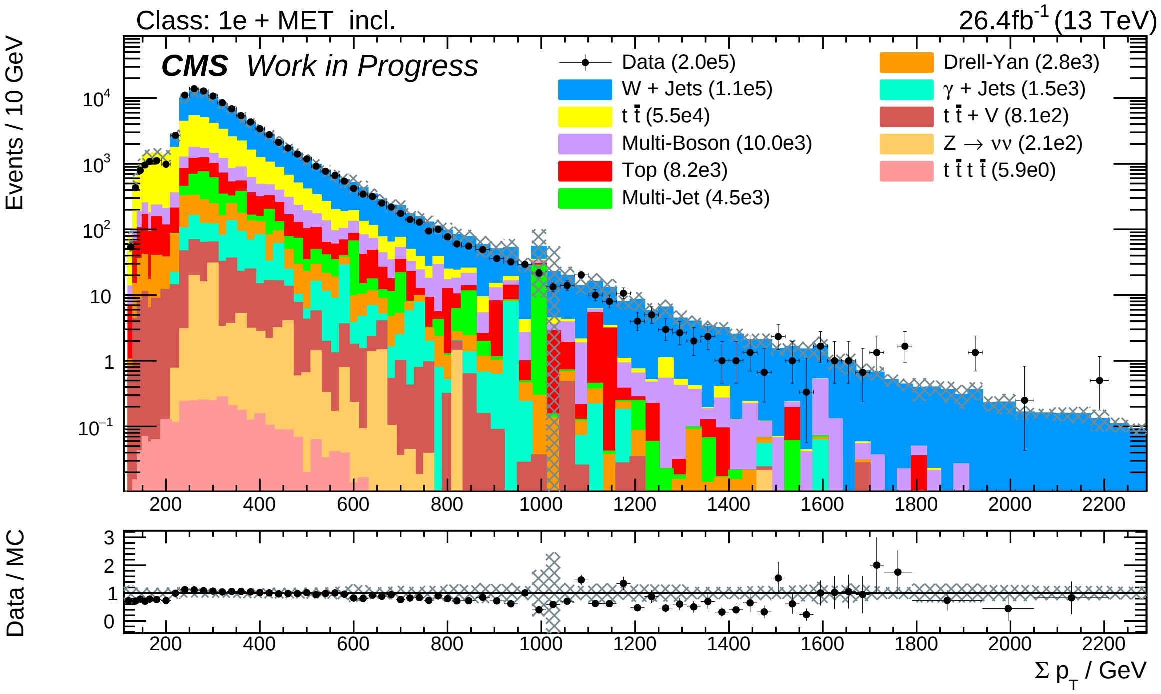


Figure 7.2.: $\sum |p_T|$ spectrum of the $1e + E_T^{\text{miss}}$ inclusive event class of the 2016 dataset.

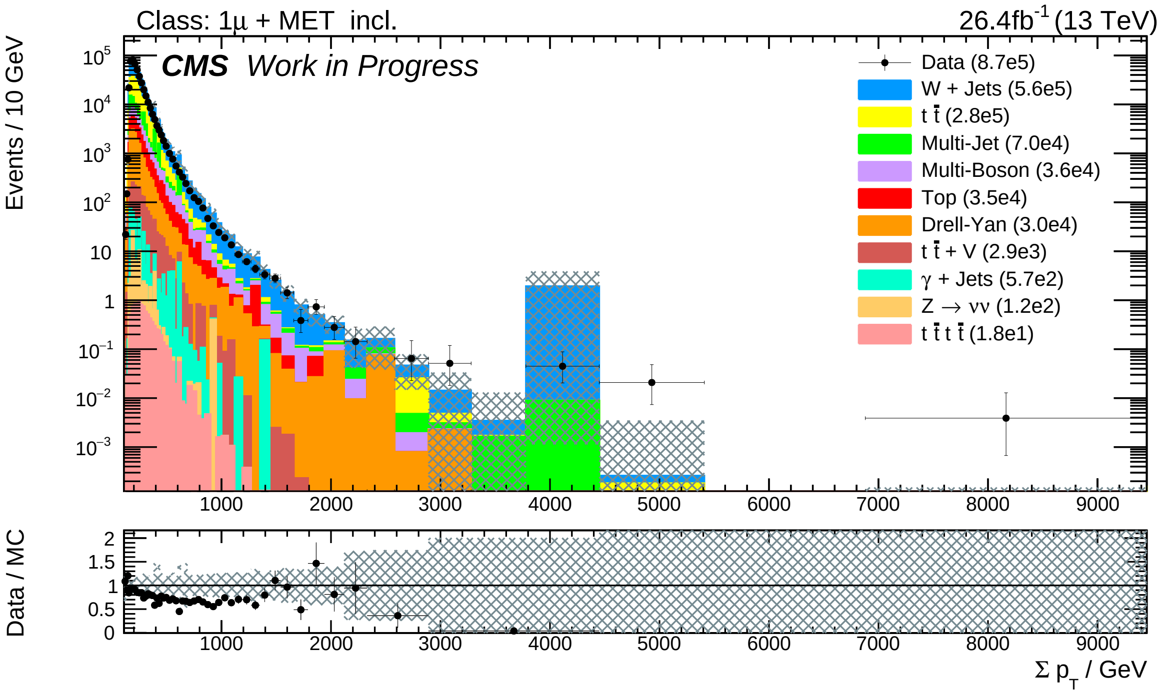


Figure 7.3.: $\sum |p_T|$ spectrum of the $1\mu + E_T^{\text{miss}}$ inclusive event class of the 2016 dataset.

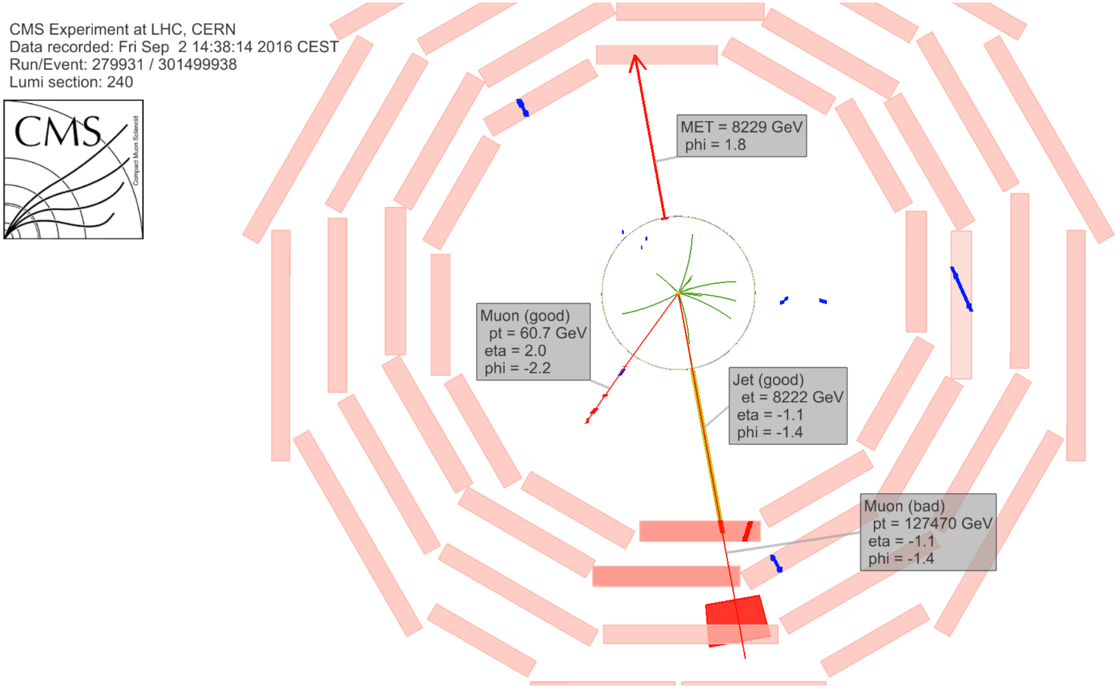


Figure 7.4.: Event display of a mis-reconstructed event showing the $\rho - \phi$ plane of the detector. The event features two muons. One of these passes the identification requirements and the other multi-TeV muon which fails the identification requirements. However, a jet with 8 TeV is reconstructed out of the inner track of the muon and the E_T^{miss} is balanced accordingly.

which triggers the event. The multi-TeV muon does not pass the reconstruction criteria and is also not accepted as a particle flow muon. However, the inner track of the muon with a momentum of over 8 TeV leads to the reconstruction of a jet with an energy of over 8 TeV as well. The jet is a valid PF candidate and the E_T^{miss} is adapted accordingly to keep the event balanced. It has three additional low energetic constituents and passes the relative energy fraction required by the tight ID (see section 5.3.4). Therefore this event shows up in the E_T^{miss} spectrum at high values as can be seen in figure 7.3.

7.2. Most Significant Classes of 2015

With the greater dataset of 2016 another look at the most significant classes in 2015 is taken. Since the same MC simulation is used deviations that appear in 2015 and 2016 are not automatically signs for new physics but can point to mis-modeling in the MC simulation. However, it might be possible to exclude statistical fluctuations.

The spectra of the most significant and completely skipped event classes in 2015 with the 2016 dataset can be found in figures 7.5 - 7.8. Due to the different single electron trigger and a different pile-up distribution of data, the MC prediction does not exactly reproduce the distribution with the 2015 dataset.

The excess in the $3e + 3$jets exclusive event class is still present in the 2016 dataset but not as

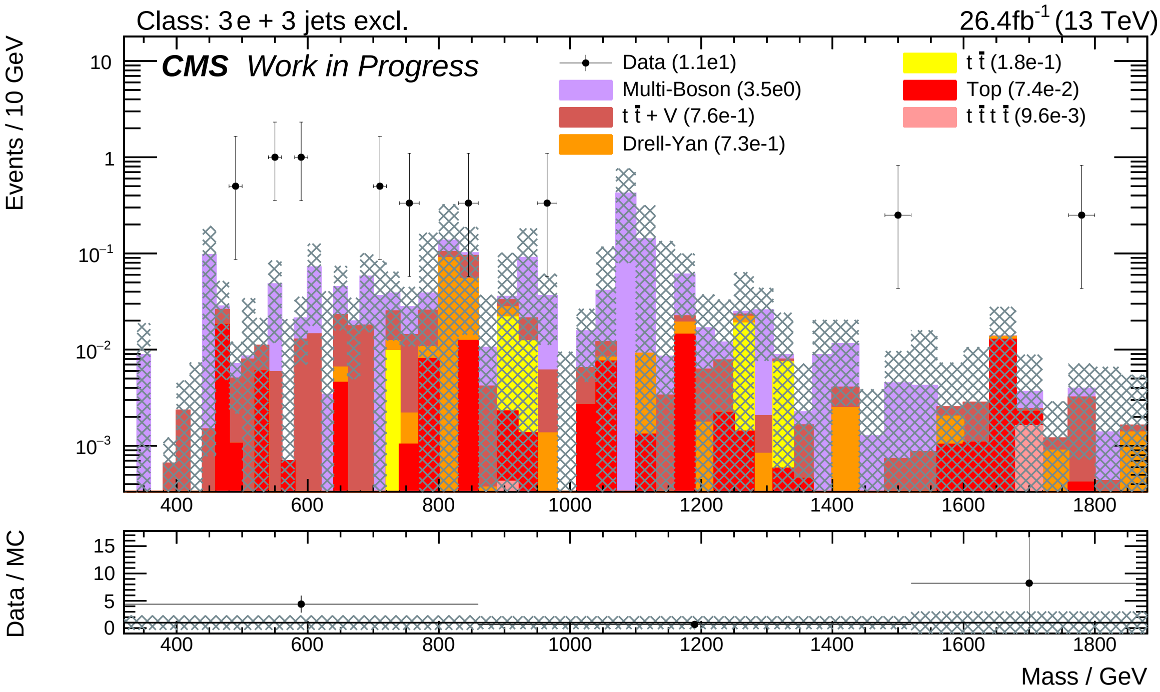


Figure 7.5.: Invariant mass spectrum of the $3e + 3$jets exclusive event class.

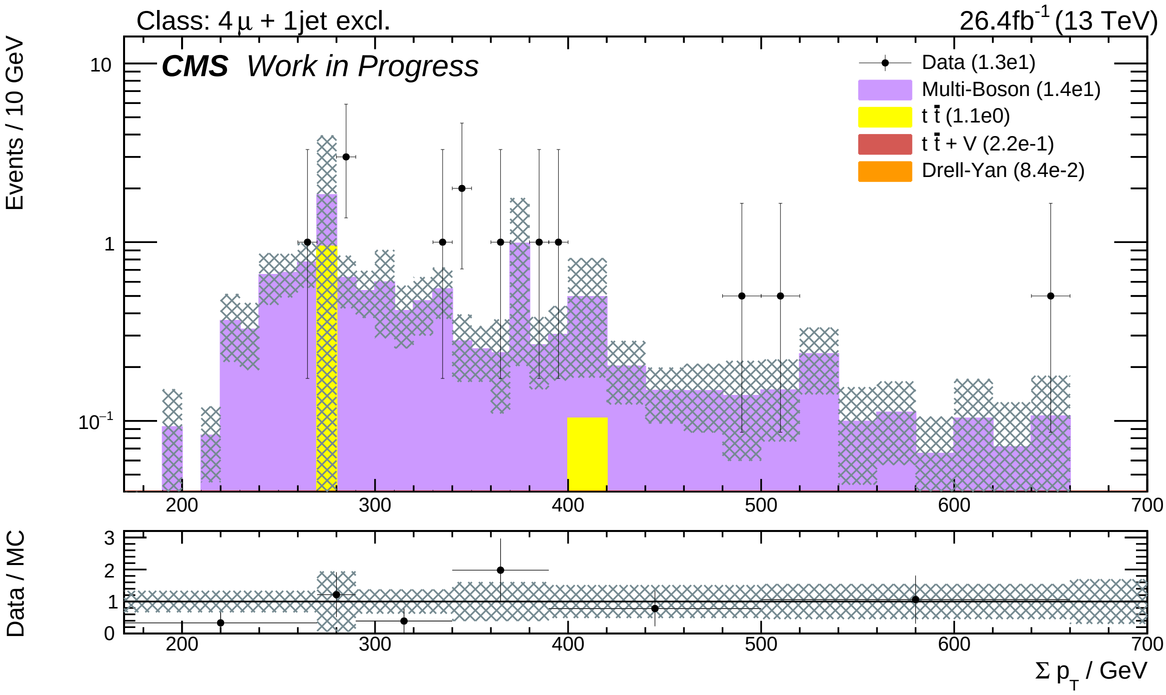


Figure 7.6.: $\sum |p_T|$ spectrum of the $4\mu + 1$jet exclusive event class.

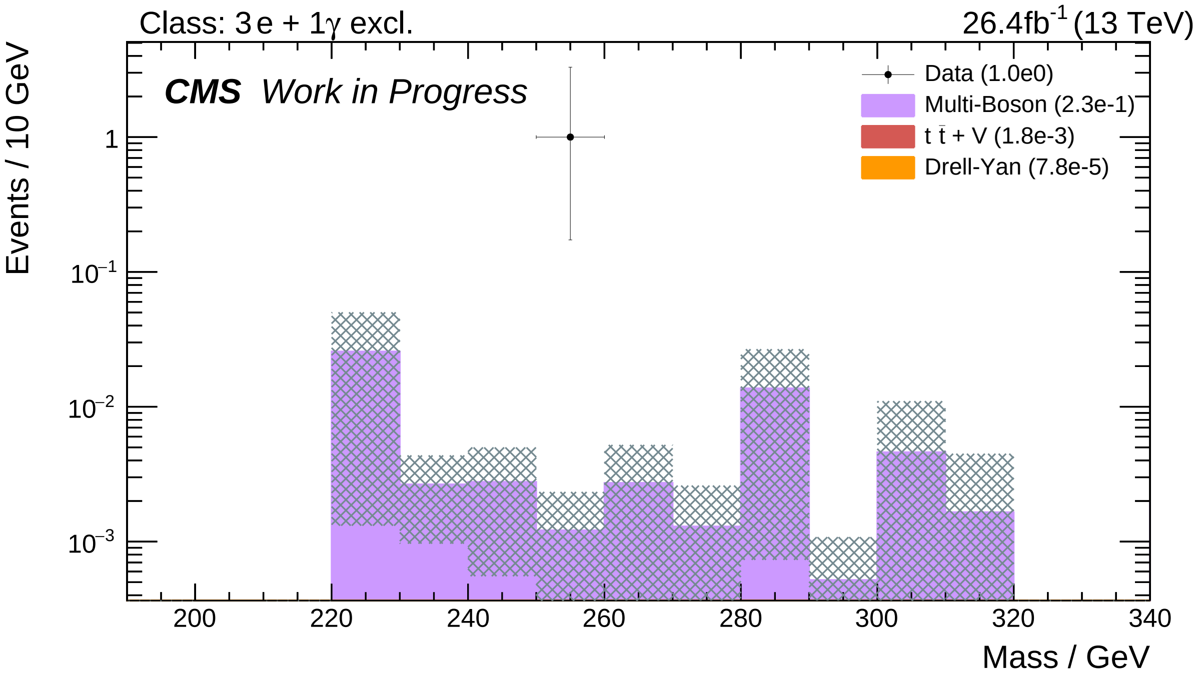


Figure 7.7.: Invariant mass spectrum of the $3e + 1\gamma$ exclusive event class.

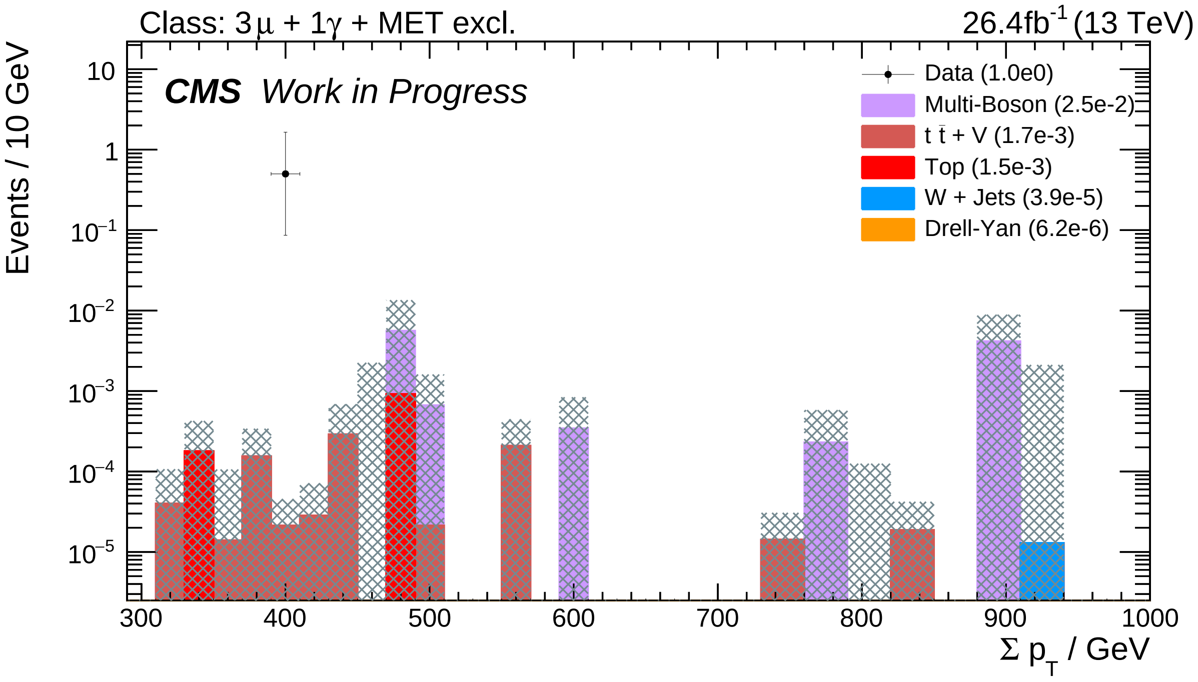


Figure 7.8.: $\sum |p_T|$ distribution of the $3\mu + 1\gamma + E_T^{\text{miss}}$ exclusive event class which is the only event class which is skipped entirely by the region veto algorithm.

significant in the integral distribution. In the $3e + 1\gamma$ exclusive event class another single event is found which is less significant due to the higher integrated luminosity as well. The excess in the $4\mu + 1$jet exclusive event class has vanished completely and the SM prediction now agrees with the measured data.

The $3\mu + 1\gamma + E_T^{\text{miss}}$ exclusive event class is the only event class which was skipped entirely by the region veto algorithm and has data events. In 2016 a single event is found in this event class as well at a similar position as in 2015. However, due to the increased luminosity the significance of this excess is further reduced compared to the 2015 dataset.

8. Conclusion and Outlook

8.1. Conclusion

In this thesis the Model Unspecific Search in CMS (MUSiC) for new physics beyond the Standard Model, its concepts and workflow have been presented. The used trigger, the event cleaning between different trigger streams, the reconstruction algorithms, identification criteria, and the event classification according to their final state have been discussed. Further, the statistical analysis of the agreement between measured data and Standard Model prediction was reviewed. A coherent set of MC simulation samples is found to represent the full SM prediction and a description of the systematic uncertainties of the SM prediction is presented.

The full analysis of single and double lepton triggered data taken by CMS in the year 2015 during the RunII of the LHC at a center of mass energy of $\sqrt{s} = 13$ TeV and an integrated luminosity of 2.3 fb^{-1} was performed. A preview of the 2016 dataset was provided.

No evidences for new physics beyond the Standard Model were found. The most significant deviation reaches the 3σ level which is in the scope of expected deviations using diced pseudo data. The three most significant deviations are excesses in di-boson dominated event classes and were discussed in detail. Overall, the data show good agreement with the Standard Model prediction and all deviations are within the scope of the expectation.

Finally an outlook to the dataset taken in 2016 was provided. Due to several issues like missing MC simulation with a matching trigger menu and pileup distribution no scan was performed but event classes containing the most significant deviations observed in the 2015 dataset were reviewed. The significance of the excesses were found to be lower in the 2016 dataset but no final conclusion can be drawn if the deviations observed in the 2015 dataset are caused by statistical fluctuations due to the missing MC simulation for the 2016 dataset.

The MUSiC analysis is performed by multiple people in Aachen and relies on the work of the whole CMS collaboration. The topics I worked on are listed in the appendix G.

8.2. Outlook

As soon as the MC production for the 2016 dataset is completed, a full scan of the data can be performed. The combination of the increase in the center of mass energy from 8 TeV to 13 TeV and the luminosity from 2.3 fb^{-1} (19.7 fb^{-1} in 2012) to 36.46 fb^{-1} will provide unprecedented sensitivity to all theoretical models in comparison to previous MUSiC analyses. Also, the increase in luminosity from 2015 to 2016 will be the largest in the following years. Therefore the full analysis of the 2016 dataset is very interesting and might reveal signs of new physics.

The analysis performed in this thesis will provide a solid base for the future analysis of RunII data and is extendible, e.g. *b*-tagged jets or other trigger streams can be included and might improve the sensitivity to many more theories.

Appendix A.

Identification Criteria

A.1. Muon Identification Criteria

Table A.1.: Identification criteria for the tight working point of the cut based ID for muons [100].

Criterion	Value
Transverse Momentum	< 200 GeV
Is Global Muon	yes
Is Particle Flow Muon	yes
Fit Quality	$\chi^2/\text{ndf} < 10$
Muon Chambers in Global Fit	> 0
Matched Muon Stations	> 1
Transverse Impact Parameter	< 2 mm
Longitudinal Impact Parameter	< 5 mm
Pixel Hits	> 1
Number Missing Tracker Hits	< 5

Table A.2.: Identification criteria for the high-p_T ID for muons [100].

Criterion	Value
Transverse Momentum	> 200 GeV
Is Global Muon	yes
Relative Uncertainty on p_T	< 0.3
Muon Chambers in Global Fit	> 0
Matched Muon Stations	> 1
Transverse Impact Parameter	< 2 mm
Longitudinal Impact Parameter	< 5 mm
Pixel Hits	> 1
Number Missing Tracker Hits	< 5

A.2. Electron Identification Criteria

Table A.3.: Identification criteria for the tight working point of the cut based ID for electrons [93].

Criterion	Value Barrel	Value Endcap
Transverse Energy	< 100 GeV	< 100 GeV
Shower Shape	$\sigma_{i\eta i\eta} < 0.0101$	$\sigma_{i\eta i\eta} < 0.0279$
Difference Track and Supercluster Position (η)	< 0.00926	< 0.00724
Difference Track and Supercluster Position (ϕ)	< 0.0336	< 0.0918
Fraction in HCAL over ECAL	< 0.0597	< 0.0615
Relative Isolation	< 0.0354	< 0.0646
Energy ECAL / Momentum Tracker $\left\lvert\frac{1}{E} - \frac{1}{p}\right\rvert$	$< 0.012/$ GeV	$< 0.00999/$ GeV
Transverse Impact Parameter	< 0.111 mm	< 0.351 mm
Longitudinal Impact Parameter	< 0.466 mm	< 4.17 mm
Number Missing Inner Hits	< 3	< 2
Pass Conversion Veto	yes	yes

Table A.4.: Identification criteria for the HEEP ID version 6.0 for electrons [97].

Criterion	Value Barrel	Value Endcap
Transverse Energy	> 100 GeV	> 100 GeV
ECAL Driven	yes	yes
Shower Shape	$E^{2x5}/E^{5x5} > 0.94$ or $E^{1x5}/E^{5x5} > 0.83$	$\sigma_{i\eta i\eta} < 0.03$
Difference Track and Supercluster Position (η)	< 0.004	< 0.006
Difference Track and Supercluster Position (ϕ)	< 0.06	< 0.06
Fraction in HCAL over ECAL	$<$ GeV$/E + 0.05$	< 5 GeV$/E + 0.05$
Calorimeter Isolation	$< 2\text{ GeV} + 0.03E_T + 0.28\rho$	$\begin{cases} < 2.5\text{ GeV} + 0.28\rho & E_T < 50\text{ GeV} \\ < 2.5\text{ GeV} + 0.28\rho & E_T > 50\text{ GeV} \\ \quad +0.03(E_T - 50\text{ GeV}) & \end{cases}$
Tracker Isolation	< 5 GeV	< 5 GeV
Transverse Impact Parameter	< 0.2 mm	< 0.5 mm
Number Missing Inner Hits	< 2	< 2

A.3. Photon Identification Criteria

Table A.5.: Identification criteria for the tight working point of the cut based ID for electrons. ρ is the mean energy density in the detector and corresponds to the amount of pileup in an event and EA are the effective areas of the corresponding object [98].

Criterion	Value Barrel	Value Endcap
Transverse Energy	< 25 GeV	< 25 GeV
Shower Shape	$\sigma_{i\eta i\eta} < 0.01$	$\sigma_{i\eta i\eta} < 0.0268$
Fraction in HCAL over ECAL	< 0.05	< 0.05
Charged Hadron Isolation	< 0.76 GeV	< 0.56 GeV
Neutral Hadron Isolation	$< 0.97\ \mathrm{GeV} + 0.014 p_T + 0.000019 p_T^2/\ \mathrm{GeV} + \rho \cdot EA$	$< 2.09\ \mathrm{GeV} + 0.0139 p_T + 0.000025 p_T^2/\ \mathrm{GeV} + \rho \cdot EA$
Photon Isolation	$< 0.08\ \mathrm{GeV} + 0.0053 p_T + \rho \cdot EA$	$< 0.16\ \mathrm{GeV} + 0.0034 p_T + 0.000025 p_T^2/\ \mathrm{GeV} + \rho \cdot EA$
Pixel Seed Veto	yes	yes

A.4. Jet Identification Criteria

Table A.6.: Identification criteria for the tight working point of the cut based ID for jets [101].

Criterion	Value Barrel
Transverse Energy	< 50 GeV
Neutral Hadron Fraction	< 0.9
Neutral EM Fraction	< 0.9
Number of Constituents	> 1
Charged Hadron Fraction	> 0
Charged Multiplicity	> 0
Charged EM Fraction	< 0.99

Appendix B.

E_T^{miss} Filter

The following E_T^{miss} filters are used in the 2015 analysis [125]:

CSCTightHalo2015Filter
Filters events with beam halo particles passing the detector [126].

eeBadScFilter
Some crystal regions of the ECAL measure sometimes anomalous energies. Events in which this problem occurs are rejected [127].

HBHENoiseFilter
Filters events with anomalous noise in the HCAL [128].

HBHENoiseIsoFilter
Filters events with anomalous noise in the HCAL not caught by algorithm of HBHENoiseFilter [128].

goodVertices
Events with much less tracks than clusters in the calorimeter are rejected. This happens for example if the hard interaction did not take place in the center of the interaction [127].

EcalDeadCellTriggerPrimitiveFilter
If energy is deposited in masked ECAL clusters the event is discarded [129].

In 2016 the beam halo filter is updated to the `globalTightHalo2016Filter`. In addition the `badMuon` and `badChargedHadron` filter are used which are supposed to filter events in which mis-reconstructed muons lead to anomalous E_T^{miss}.

Appendix C.

Trigger

In this appendix the encryption of the trigger names is explained. The numbers behind the shortcuts *Ele* or *Mu* correspond to the p_T threshold in GeV. The rest of the trigger names encrypts object requirements for the identification (*Id*) or isolation (*Iso*) in the tracker (*Trk*) and the calorimeters (*Calo*). Each of these requirements have different working points: Loose (*L*), tight (*T*), very tight (*VT*) or very very loose (*VVL*). The double muon trigger has an additional requirement on the longitudinal displacement of the track (*DZ*) with respect to the primary vertex.

Appendix D.

Datasets

D.1. Data

The table D.1 provides the datasets used for the analysis of 2015 data. The data were analyzed under the `76X_dataRun2_16Dec2015_v0` conditions. The certification of all runs and lumisections is provided by the golden JSON `Cert_13TeV_16Dec2015ReReco_Collisions15_25ns_JSON_v2.txt`. The normtag used to calculate the luminosity is `moriond16_normtag.json`.

Table D.1.: Name of the datasets used in the 2015 analysis.

Run	Dataset Name	Integrated Luminosity
Run2015D	/SingleMuon/Run2015D-16Dec2015-v1/MINIAOD	2.3 fb^{-1}
	/SingleElectron/Run2015D-08Jun2016-v1/MINIAOD	2.3 fb^{-1}
	/DoubleMuon/Run2015D-16Dec2015-v1/MINIAOD	2.3 fb^{-1}
	/DoubleEG/Run2015D-08Jun2016-v1/MINIAOD	2.3 fb^{-1}

D.2. MC

The following table D.2 shows the complete, detailed list of MC samples used in the MUSiC analysis. All samples have been analyzed under the `76X_mcRun2_asymptotic_RunIIFall15DR76_v1` conditions.

Table D.2.: Table with all MC samples used in the MUSiC analysis. The weight is the absolute value of the MC weight and is calculated by $\alpha_{MC} = \epsilon_{\text{filter}} * L * \sigma / N_{MC}$. The k-factor order *NLO_W* corresponds to a NLO k-factor with increased uncertainty as described in 3.2.1.

Process	**Dataset Name**	**σ / pb (Order)**	**Filter Efficiency**	**k-Factor (Order)**	**Number Events**	**Weight α_{MC}**	**Reference**
Drell-Yan	DYJetsToLL_M-50_HT-100to200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	140 (LO)	1.0	1.6 (NLO)	$2.7 \cdot 10^6$	0.19	[130]
	DYJetsToLL_M-50_HT-100to200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 (ext1)	140 (LO)	1.0	1.6 (NLO)	$8.4 \cdot 10^6$	0.061	[130]
	DYJetsToLL_M-50_HT-200to400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	43 (LO)	1.0	1.4 (NLO)	$9.6 \cdot 10^5$	0.15	[130]
	DYJetsToLL_M-50_HT-400to600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	5.5 (LO)	1.0	1.5 (NLO)	$1.1 \cdot 10^6$	0.018	[130]
	DYJetsToLL_M-50_HT-400to600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 (ext1)	5.5 (LO)	1.0	1.5 (NLO)	$9.0 \cdot 10^6$	0.0021	[130]
	DYJetsToLL_M-50_HT-600toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	2.2 (LO)	1.0	1.1 (NLO)	$1.0 \cdot 10^6$	0.0056	[130]
	DYJetsToLL_M-50_HT-600toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 (ext1)	2.2 (LO)	1.0	1.1 (NLO)	$4.1 \cdot 10^6$	0.0014	[130]
	DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	4900 (LO)	1.0	1.2 (NNLO)	$9.0 \cdot 10^6$	1.5	[131]
	DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 (ext1)	4900 (LO)	1.0	1.2 (NNLO)	$2.5 \cdot 10^8$	0.056	[131]
	DYJetsToLL_M-5to50_HT-100to200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	220 (LO)	1.0	1.2 (NLO)	$1.0 \cdot 10^6$	0.62	[131]
	DYJetsToLL_M-5to50_HT-100to200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 (ext1)	220 (LO)	1.0	1.2 (NLO)	$8.4 \cdot 10^6$	0.075	[131]
	DYJetsToLL_M-5to50_HT-200to400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	37 (LO)	1.0	1.2 (NLO)	$1.0 \cdot 10^6$	0.10	[131]
	DYJetsToLL_M-5to50_HT-200to400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 (ext1)	37 (LO)	1.0	1.2 (NLO)	$2.0 \cdot 10^6$	0.052	[131]
	DYJetsToLL_M-5to50_HT-400to600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	3.6 (LO)	1.0	1.2 (NLO)	$1.0 \cdot 10^6$	0.010	[131]
	DYJetsToLL_M-5to50_HT-600toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1.1 (LO)	1.0	1.2 (NLO)	$1.0 \cdot 10^6$	0.0031	[131]
	DYJetsToLL_M-5to50_HT-600toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 (ext1)	1.1 (LO)	1.0	1.2 (NLO)	$1.9 \cdot 10^6$	0.0017	[131]
	DYJetsToLL_M-5to50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	$7.1 \cdot 10^4$ (LO)	1.0	1.2 (NLO)	$8.8 \cdot 10^6$	23	[131]
	ZToEE_NNPDF30_13TeV-powheg_M_120_200	19 (NLO)	1.0	-	$1.0 \cdot 10^5$	0.44	[132]
	ZToEE_NNPDF30_13TeV-powheg_M_1400_2300	0.0014 (NLO)	1.0	-	$1.0 \cdot 10^5$	$3.2 \cdot 10^{-5}$	[132]
	ZToEE_NNPDF30_13TeV-powheg_M_200_400	2.7 (NLO)	1.0	-	$1.0 \cdot 10^5$	0.063	[132]
	ZToEE_NNPDF30_13TeV-powheg_M_2300_3500	$8.9 \cdot 10^{-5}$ (NLO)	1.0	-	$1.0 \cdot 10^5$	$2.1 \cdot 10^{-6}$	[132]
	ZToEE_NNPDF30_13TeV-powheg_M_3500_4500	$4.1 \cdot 10^{-6}$ (NLO)	1.0	-	$1.0 \cdot 10^5$	$9.5 \cdot 10^{-8}$	[132]

Table D.2.: Table with all MC samples used in the MUSiC analysis. The weight is the absolute value of the MC weight and is calculated by $\alpha_{MC} = \epsilon_{\text{filter}} * L * \sigma / N_{MC}$. The k-factor order NLO_W corresponds to a NLO k-factor with increased uncertainty as described in 3.2.1.

Process	Dataset Name	σ / pb (Order)	Filter Efficiency	k-Factor (Order)	Number Events	Weight α_{MC}	Reference
	ZToEE_NNPDF30_13TeV-powheg_M_400_800	0.24 (NLO)	1.0	-	$1.0 \cdot 10^5$	0.0055	[132]
	ZToEE_NNPDF30_13TeV-powheg_M_4500_6000	$4.6 \cdot 10^{-7}$ (NLO)	1.0	-	$9.9 \cdot 10^4$	$1.1 \cdot 10^{-8}$	[132]
	ZToEE_NNPDF30_13TeV-powheg_M_50_120	2000 (NLO)	1.0	-	$3.0 \cdot 10^6$	1.5	[132]
	ZToEE_NNPDF30_13TeV-powheg_M_6000_Inf	$2.1 \cdot 10^{-8}$ (NLO)	1.0	-	$9.8 \cdot 10^4$	$4.9 \cdot 10^{-10}$	[132]
	ZToEE_NNPDF30_13TeV-powheg_M_800_1400	0.017 (NLO)	1.0	-	$1.0 \cdot 10^5$	$3.9 \cdot 10^{-4}$	[132]
	ZToMuMu_NNPDF30_13TeV-powheg_M_120_200	19 (NLO)	1.0	-	$1.0 \cdot 10^5$	0.44	[132]
	ZToMuMu_NNPDF30_13TeV-powheg_M_1400_2300	0.0014 (NLO)	1.0	-	$9.9 \cdot 10^4$	$3.2 \cdot 10^{-5}$	[132]
	ZToMuMu_NNPDF30_13TeV-powheg_M_200_400	2.7 (NLO)	1.0	-	$1.0 \cdot 10^5$	0.063	[132]
	ZToMuMu_NNPDF30_13TeV-powheg_M_2300_3500	$8.9 \cdot 10^{-5}$ (NLO)	1.0	-	$1.0 \cdot 10^5$	$2.1 \cdot 10^{-6}$	[132]
	ZToMuMu_NNPDF30_13TeV-powheg_M_3500_4500	$4.1 \cdot 10^{-6}$ (NLO)	1.0	-	$1.0 \cdot 10^5$	$9.5 \cdot 10^{-8}$	[132]
	ZToMuMu_NNPDF30_13TeV-powheg_M_400_800	0.24 (NLO)	1.0	-	$10.0 \cdot 10^4$	0.0056	[132]
	ZToMuMu_NNPDF30_13TeV-powheg_M_4500_6000	$4.6 \cdot 10^{-7}$ (NLO)	1.0	-	$1.0 \cdot 10^5$	$1.0 \cdot 10^{-8}$	[132]
	ZToMuMu_NNPDF30_13TeV-powheg_M_50_120	2000 (NLO)	1.0	-	$3.0 \cdot 10^6$	1.5	[132]
	ZToMuMu_NNPDF30_13TeV-powheg_M_6000_Inf	$2.1 \cdot 10^{-8}$ (NLO)	1.0	-	$9.9 \cdot 10^4$	$4.8 \cdot 10^{-10}$	[132]
	ZToMuMu_NNPDF30_13TeV-powheg_M_800_1400	0.017 (NLO)	1.0	-	$9.8 \cdot 10^4$	$4.0 \cdot 10^{-4}$	[132]
W	WJetsToLNu_HT-100To200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1300 (LO)	1.0	1.2 (NLO_W)	$1.0 \cdot 10^7$	0.35	[131]
	WJetsToLNu_HT-100To200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 (ext1)	1300 (LO)	1.0	1.2 (NLO_W)	$2.8 \cdot 10^7$	0.13	[131]
	WJetsToLNu_HT-1200To2500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1.3 (LO)	1.0	1.2 (NLO_W)	$2.5 \cdot 10^5$	0.015	[131]
	WJetsToLNu_HT-1200To2500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 (ext1)	1.3 (LO)	1.0	1.2 (NLO_W)	$6.7 \cdot 10^6$	$5.5 \cdot 10^{-4}$	[131]
	WJetsToLNu_HT-200To400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	390 (LO)	1.0	1.2 (NLO_W)	$4.9 \cdot 10^6$	0.22	[131]
	WJetsToLNu_HT-200To400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 (ext1)	390 (LO)	1.0	1.2 (NLO_W)	$1.5 \cdot 10^7$	0.072	[131]
	WJetsToLNu_HT-2500ToInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	0.031 (LO)	1.0	1.2 (NLO_W)	$2.5 \cdot 10^5$	$3.4 \cdot 10^{-4}$	[131]
	WJetsToLNu_HT-2500ToInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 (ext1)	0.031 (LO)	1.0	1.2 (NLO_W)	$2.3 \cdot 10^6$	$3.7 \cdot 10^{-5}$	[131]
	WJetsToLNu_HT-400To600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	48 (LO)	1.0	1.2 (NLO_W)	$1.9 \cdot 10^6$	0.069	[131]
	WJetsToLNu_HT-600To800_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	13 (LO)	1.0	1.2 (NLO_W)	$3.8 \cdot 10^6$	0.0095	[131]

Table D.2.: Table with all MC samples used in the MUSiC analysis. The weight is the absolute value of the MC weight and is calculated by $\alpha_{MC} = \epsilon_{\text{filter}} * L * \sigma / N_{MC}$. The *k*-factor order *NLO_W* corresponds to a NLO *k*-factor with increased uncertainty as described in 3.2.1.

Process	Dataset Name	σ / pb (Order)	Filter Efficiency	*k*-Factor (Order)	Number Events	Weight α_{MC}	Reference
	WJetsToLNu_HT-800To1200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	5.3 (LO)	1.0	1.2 (NLO_W)	$1.6 \cdot 10^6$	0.0094	[131]
	WJetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	$5.1 \cdot 10^4$ (LO)	1.0	1.2 (NNLO)	$4.7 \cdot 10^7$	3.0	[131]
	WJetsToQQ_HT-600ToInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	95 (LO)	1.0	-	$1.0 \cdot 10^6$	0.21	[132]
	WJetsToQQ_HT180_13TeV-madgraphMLM-pythia8	2800 (LO)	1.0	-	$2.3 \cdot 10^7$	0.28	[132]
	WToENu_M-1000_TuneCUETP8M1_13TeV-pythia8	0.013 (LO)	1.0	1.3 (NNLO)	$10.0 \cdot 10^5$	$3.9 \cdot 10^{-5}$	[133]
	WToENu_M-2000_TuneCUETP8M1_13TeV-pythia8	$5.6 \cdot 10^{-4}$ (LO)	1.0	1.3 (NNLO)	$10.0 \cdot 10^5$	$1.6 \cdot 10^{-6}$	[133]
	WToENu_M-200_TuneCUETP8M1_13TeV-pythia8	6.2 (LO)	1.0	1.3 (NNLO)	$10.0 \cdot 10^5$	0.019	[133]
	WToENu_M-3000_TuneCUETP8M1_13TeV-pythia8	$2.9 \cdot 10^{-5}$ (LO)	1.0	1.1 (NNLO)	$10.0 \cdot 10^5$	$7.6 \cdot 10^{-8}$	[133]
	WToENu_M-4000_TuneCUETP8M1_13TeV-pythia8	$3.0 \cdot 10^{-6}$ (LO)	0.45	-	$9.9 \cdot 10^5$	$3.1 \cdot 10^{-9}$	[133]
	WToENu_M-500_TuneCUETP8M1_13TeV-pythia8	0.21 (LO)	1.0	1.3 (NNLO)	$9.9 \cdot 10^5$	$6.6 \cdot 10^{-4}$	[133]
	WToMuNu_M-1000_TuneCUETP8M1_13TeV-pythia8	0.013 (LO)	1.0	1.3 (NNLO)	$9.9 \cdot 10^5$	$3.7 \cdot 10^{-5}$	[133]
	WToMuNu_M-2000_TuneCUETP8M1_13TeV-pythia8	$5.6 \cdot 10^{-4}$ (LO)	1.0	1.2 (NNLO)	$9.9 \cdot 10^5$	$1.5 \cdot 10^{-6}$	[133]
	WToMuNu_M-200_TuneCUETP8M1_13TeV-pythia8	6.2 (LO)	1.0	1.3 (NNLO)	$9.9 \cdot 10^5$	0.019	[133]
	WToMuNu_M-3000_TuneCUETP8M1_13TeV-pythia8	$2.9 \cdot 10^{-5}$ (LO)	1.0	1.0 (NNLO)	$10.0 \cdot 10^5$	$7.0 \cdot 10^{-8}$	[133]
	WToMuNu_M-4000_TuneCUETP8M1_13TeV-pythia8	$3.0 \cdot 10^{-6}$ (LO)	1.0	0.41 (NNLO)	$9.9 \cdot 10^5$	$2.9 \cdot 10^{-9}$	[133]
	WToMuNu_M-500_TuneCUETP8M1_13TeV-pythia8	0.21 (LO)	1.0	1.3 (NNLO)	$10.0 \cdot 10^5$	$6.3 \cdot 10^{-4}$	[133]
	WToTauNu_M-1000_TuneCUETP8M1_13TeV-pythia8	0.013 (LO)	1.0	1.3 (NNLO)	$10.0 \cdot 10^5$	$3.9 \cdot 10^{-5}$	[133]
	WToTauNu_M-2000_TuneCUETP8M1_13TeV-pythia8-tauola	$5.6 \cdot 10^{-4}$ (LO)	1.0	1.2 (NNLO)	$9.8 \cdot 10^5$	$1.6 \cdot 10^{-6}$	[133]
	WToTauNu_M-200_TuneCUETP8M1_13TeV-pythia8	6.2 (LO)	1.0	1.3 (NNLO)	$9.9 \cdot 10^5$	0.019	[133]
	WToTauNu_M-3000_TuneCUETP8M1_13TeV-pythia8-tauola	$2.9 \cdot 10^{-5}$ (LO)	1.0	1.1 (NNLO)	$9.9 \cdot 10^5$	$7.5 \cdot 10^{-8}$	[133]
	WToTauNu_M-4000_TuneCUETP8M1_13TeV-pythia8-tauola	$3.0 \cdot 10^{-6}$ (LO)	1.0	0.44 (NNLO)	$10.0 \cdot 10^5$	$3.1 \cdot 10^{-9}$	[133]
	WToTauNu_M-500_TuneCUETP8M1_13TeV-pythia8-tauola	0.21 (LO)	1.0	1.3 (NNLO)	$9.8 \cdot 10^5$	$6.7 \cdot 10^{-4}$	[133]
γ	GJets_HT-100To200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	9200 (LO)	1.0	-	$5.1 \cdot 10^6$	4.2	[131]
	GJets_HT-200To400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	2300 (LO)	1.0	-	$1.0 \cdot 10^7$	0.51	[131]
	GJets_HT-400To600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	270 (LO)	1.0	-	$2.4 \cdot 10^6$	0.26	[131]
	GJets_HT-40To100_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	$2.1 \cdot 10^4$ (LO)	1.0	-	$4.4 \cdot 10^6$	11	[131]
	GJets_HT-600ToInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	93 (LO)	1.0	-	$2.5 \cdot 10^6$	0.088	[131]
QCD	QCD_Pt-1000toInf_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8	10 (LO)	0.16	-	$3.8 \cdot 10^6$	$9.7 \cdot 10^{-4}$	[132]
	QCD_Pt-120to170_EMEnriched_TuneCUETP8M1_13TeV_pythia8	$4.8 \cdot 10^5$ (LO)	0.13	-	$3.6 \cdot 10^7$	4.0	[132]
	QCD_Pt-120to170_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8	$4.7 \cdot 10^5$ (LO)	0.054	-	$8.0 \cdot 10^6$	7.3	[132]
	QCD_Pt-15to20_EMEnriched_TuneCUETP8M1_13TeV_pythia8	$1.3 \cdot 10^9$ (LO)	0.0018	-	$5.5 \cdot 10^6$	970	[132]
	QCD_Pt-15to20_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8	$1.3 \cdot 10^9$ (LO)	0.0030	-	$4.7 \cdot 10^6$	1900	[132]
	QCD_Pt-170to300_EMEnriched_TuneCUETP8M1_13TeV_pythia8	$1.1 \cdot 10^5$ (LO)	0.17	-	$1.2 \cdot 10^7$	3.8	[132]
	QCD_Pt-170to300_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8	$1.2 \cdot 10^5$ (LO)	0.073	-	$7.9 \cdot 10^6$	2.5	[132]
	QCD_Pt-20to30_EMEnriched_TuneCUETP8M1_13TeV_pythia8	$5.6 \cdot 10^8$ (LO)	0.0096	-	$9.3 \cdot 10^6$	1300	[132]

Table D.2.: Table with all MC samples used in the MUSiC analysis. The weight is the absolute value of the MC weight and is calculated by $\alpha_{MC} = \epsilon_{\text{filter}} * L * \sigma / N_{MC}$. The k-factor order NLO_W corresponds to a NLO k-factor with increased uncertainty as described in 3.2.1.

Process	Dataset Name	σ / pb (Order)	Filter Efficiency	k-Factor (Order)	Number Events	Weight α_{MC}	Reference
	QCD_Pt-20to30_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8	$5.6 \cdot 10^8$ (LO)	0.0053	-	$3.2 \cdot 10^7$	210	[132]
	QCD_Pt-300to470_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8	7800 (LO)	0.10	-	$7.8 \cdot 10^6$	0.23	[132]
	QCD_Pt-300toInf_EMEnriched_TuneCUETP8M1_13TeV_pythia8	9000 (LO)	0.15	-	$7.3 \cdot 10^6$	0.42	[132]
	QCD_Pt-30to50_EMEnriched_TuneCUETP8M1_13TeV_pythia8	$1.4 \cdot 10^8$ (LO)	0.073	-	$4.7 \cdot 10^6$	4900	[132]
	QCD_Pt-30to50_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8	$1.4 \cdot 10^8$ (LO)	0.012	-	$3.0 \cdot 10^7$	130	[132]
	QCD_Pt-470to600_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8	650 (LO)	0.12	-	$3.8 \cdot 10^6$	0.047	[132]
	QCD_Pt-50to80_EMEnriched_TuneCUETP8M1_13TeV_pythia8	$2.0 \cdot 10^7$ (LO)	0.15	-	$2.2 \cdot 10^7$	300	[132]
	QCD_Pt-50to80_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8	$1.9 \cdot 10^7$ (LO)	0.023	-	$2.0 \cdot 10^7$	49	[132]
	QCD_Pt-600to800_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8	190 (LO)	0.13	-	$4.0 \cdot 10^6$	0.014	[132]
	QCD_Pt-800to1000_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8	32 (LO)	0.15	-	$3.7 \cdot 10^6$	0.0030	[132]
	QCD_Pt-80to120_EMEnriched_TuneCUETP8M1_13TeV_pythia8	$2.8 \cdot 10^6$ (LO)	0.12	-	$3.6 \cdot 10^7$	22	[132]
	QCD_Pt-80to120_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8	$2.8 \cdot 10^6$ (LO)	0.038	-	$1.4 \cdot 10^7$	18	[132]
$t\bar{t}\gamma$	TTGJets_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8	3.7 (NLO)	1.0	-	$4.9 \cdot 10^6$	0.0017	[131]
$t\bar{t}\gamma\gamma$	TTGG_0Jets_TuneCUETP8M1_13TeV_amcatnlo_madspin_pythia8	0.017 (NLO)	1.0	-	$1.4 \cdot 10^6$	$2.9 \cdot 10^{-5}$	[131]
$t\bar{t}W$	TTWJetsToQQ_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8	0.41 (NLO)	1.0	-	$8.3 \cdot 10^5$	0.0011	[132]
	ttWJets_13TeV_madgraphMLM	0.24 (LO)	1.0	-	$1.3 \cdot 10^7$	$4.3 \cdot 10^{-5}$	[132]
$t\bar{t}Z$	TTZToQQ_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.53 (NLO)	1.0	-	$7.5 \cdot 10^5$	0.0016	[132]
	ttZJets_13TeV_madgraphMLM	0.26 (LO)	1.0	-	$2.2 \cdot 10^7$	$2.7 \cdot 10^{-5}$	[132]
$t\bar{t}$	TT_Mtt-1000toInf_TuneCUETP8M1_13TeV-powheg-pythia8 (ext1)	730 (NLO)	0.025	1.1 (NNLO)	$1.8 \cdot 10^6$	0.027	[131]
	TT_Mtt-1000toInf_TuneCUETP8M1_13TeV-powheg-pythia8 (ext2)	730 (NLO)	0.025	1.1 (NNLO)	$2.5 \cdot 10^7$	0.0019	[131]
	TT_Mtt-700to1000_TuneCUETP8M1_13TeV-powheg-pythia8	730 (NLO)	0.092	1.1 (NNLO)	$3.7 \cdot 10^7$	0.0047	[131]
	TT_Mtt-700to1000_TuneCUETP8M1_13TeV-powheg-pythia8 (ext1)	730 (NLO)	0.092	1.1 (NNLO)	$3.1 \cdot 10^6$	0.057	[131]
	TT_TuneCUETP8M1_13TeV-powheg-pythia8 (ext3)	730 (NLO)	1.0	1.1 (NNLO)	$9.8 \cdot 10^7$	0.020	[131]
	TT_TuneCUETP8M1_13TeV-powheg-pythia8 (ext4)	730 (NLO)	1.0	1.1 (NNLO)	$1.9 \cdot 10^8$	0.010	[131]
$t\bar{t}t\bar{t}$	TTTT_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.0091 (NLO)	1.0	-	$2.5 \cdot 10^5$	$8.4 \cdot 10^{-5}$	[132]
	TTTT_TuneCUETP8M1_13TeV-amcatnlo-pythia8 (ext1)	0.0091 (NLO)	1.0	-	$9.6 \cdot 10^5$	$2.2 \cdot 10^{-5}$	[132]
Single Top (s-channel)	ST_s-channel_4f_leptonDecays_13TeV-amcatnlo-pythia8_TuneCUETP8M1	3.4 (NLO)	1.0	-	$10.0 \cdot 10^5$	0.0077	[131]
Single Top (t-channel)	ST_t-channel_antitop_4f_leptonDecays_13TeV-powheg-pythia8_TuneCUETP8M1	25 (NLO)	1.0	-	$1.6 \cdot 10^6$	0.036	[132]
	ST_t-channel_top_4f_leptonDecays_13TeV-powheg-pythia8_TuneCUETP8M1	42 (NLO)	1.0	-	$3.3 \cdot 10^6$	0.029	[132]
tW	ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1	38 (NLO)	1.0	0.93 (NNLO)	$10.0 \cdot 10^5$	0.082	[131]

Table D.2.: Table with all MC samples used in the MUSiC analysis. The weight is the absolute value of the MC weight and is calculated by $\alpha_{MC} = \epsilon_{\text{filter}} * L * \sigma / N_{MC}$. The k-factor order NLO_W corresponds to a NLO k-factor with increased uncertainty as described in 3.2.1.

Process	Dataset Name	σ / pb (Order)	Filter Efficiency	k-Factor (Order)	Number Events	Weight α_{MC}	Reference
	ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1	38 (NLO)	1.0	0.93 (NNLO)	$1.0 \cdot 10^6$	0.082	[131]
tZq	tZq_ll_4f_13TeV-amcatnlo-pythia8_TuneCUETP8M1	0.076 (NLO)	1.0	-	$3.0 \cdot 10^6$	$5.8 \cdot 10^{-5}$	[132]
	tZq_nunu_4f_13TeV-amcatnlo-pythia8_TuneCUETP8M1	0.14 (NLO)	1.0	-	$9.8 \cdot 10^5$	$3.2 \cdot 10^{-4}$	[132]
$t\gamma$	TGJets_TuneCUETP8M1_13TeV_amcatnlo_madspin_pythia8	3.0 (NLO)	1.0	-	$2.8 \cdot 10^5$	0.024	[131]
	TGJets_TuneCUETP8M1_13TeV_amcatnlo_madspin_pythia8 (ext1)	3.0 (NLO)	1.0	-	$1.6 \cdot 10^6$	0.0044	[131]
$\gamma\gamma$	DiPhotonJetsBox_M40_80-Sherpa	300 (NLO)	1.0	-	$5.0 \cdot 10^6$	0.14	[132]
	DiPhotonJetsBox_MGG-80toInf_13TeV-Sherpa	83 (NLO)	1.0	-	$5.9 \cdot 10^7$	0.0032	[132]
$W\gamma$	WGToLNuG_PtG-500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	0.012 (LO)	1.0	-	$1.4 \cdot 10^6$	$2.0 \cdot 10^{-5}$	[132]
	WGToLNuG_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	410 (LO)	1.0	1.2 (NLO)	$6.1 \cdot 10^6$	0.18	[132]
WW	GluGluWWTo2L2Nu_MCFM_13TeV	0.39 (NLO)	1.0	2.2 (NLO)	$5.0 \cdot 10^5$	0.0039	[131]
	WWTo2L2Nu_13TeV-powheg	10 (NLO)	1.0	1.2 (NNLO)	$2.0 \cdot 10^6$	0.014	[131]
	WWTo2L2Nu_Mll_1200To2500_13TeV-powheg	10 (NLO)	$2.9 \cdot 10^{-4}$	1.2 (NNLO)	$2.0 \cdot 10^5$	$4.1 \cdot 10^{-5}$	[131]
	WWTo2L2Nu_Mll_200To600_13TeV-powheg	10 (NLO)	0.11	1.2 (NNLO)	$2.0 \cdot 10^5$	0.016	[131]
	WWTo2L2Nu_Mll_2500ToInf_13TeV-powheg	10 (NLO)	$4.4 \cdot 10^{-6}$	1.2 (NNLO)	$3.9 \cdot 10^4$	$3.2 \cdot 10^{-6}$	[131]
	WWTo2L2Nu_Mll_600To1200_13TeV-powheg	10 (NLO)	0.0047	1.2 (NNLO)	$2.0 \cdot 10^5$	$6.5 \cdot 10^{-4}$	[131]
	WWTo4Q_13TeV-powheg	45 (NLO)	1.0	1.1 (NNLO)	$2.0 \cdot 10^6$	0.059	[131]
	WWToLNuQQ_13TeV-powheg	44 (NLO)	1.0	1.1 (NNLO)	$1.9 \cdot 10^6$	0.060	[131]
	WWToLNuQQ_13TeV-powheg (ext1)	44 (NLO)	1.0	1.1 (NNLO)	$7.0 \cdot 10^6$	0.016	[131]
	WW_DoubleScattering_13TeV-pythia8	1.6 (LO)	1.0	-	$8.4 \cdot 10^5$	0.0045	[132]
$WW\gamma$	WWG_TuneCUETP8M1_13TeV-amcatnlo-pythia8 (ext1)	0.21 (NLO)	1.0	-	$10.0 \cdot 10^5$	$5.0 \cdot 10^{-4}$	[132]
WWW	WWW_4F_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.21 (NLO)	1.0	-	$2.4 \cdot 10^5$	0.0020	[132]
WWZ	WWZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.17 (NLO)	1.0	-	$2.5 \cdot 10^5$	0.0015	[132]
WZ	WZTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin_pythia8	11 (NLO)	1.0	-	$2.0 \cdot 10^7$	0.0012	[131]
	WZTo1L3Nu_13TeV_amcatnloFXFX_madspin_pythia8	3.0 (NLO)	1.0	-	$1.7 \cdot 10^6$	0.0041	[131]
	WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	5.6 (NLO)	1.0	-	$2.6 \cdot 10^7$	$5.0 \cdot 10^{-4}$	[131]
	WZTo3LNu_TuneCUETP8M1_13TeV-powheg-pythia8	4.4 (NLO)	1.0	-	$2.0 \cdot 10^6$	0.0051	[132]
$WZ\gamma$	WZG_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.041 (NLO)	1.0	-	$10.0 \cdot 10^5$	$9.5 \cdot 10^{-5}$	[132]
WZZ	WZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.056 (NLO)	1.0	-	$2.5 \cdot 10^5$	$5.1 \cdot 10^{-4}$	[132]
$Z\gamma$	ZGTo2LG_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	120 (NLO)	1.0	-	$4.4 \cdot 10^6$	0.061	[132]
	ZLLGJets_MonoPhoton_PtG-130_TuneCUETP8M1_13TeV-madgraph	0.14 (LO)	1.0	-	$4.9 \cdot 10^5$	$6.7 \cdot 10^{-4}$	[132]
$Z \to \nu\nu$	ZJetsToNuNu_HT-100To200_13TeV-madgraph	280 (LO)	1.0	1.6 (NLO)	$5.2 \cdot 10^6$	0.20	[130]
	ZJetsToNuNu_HT-200To400_13TeV-madgraph	78 (LO)	1.0	1.6 (NLO)	$5.1 \cdot 10^6$	0.057	[130]
	ZJetsToNuNu_HT-200To400_13TeV-madgraph (ext1)	78 (LO)	1.0	1.6 (NLO)	$1.9 \cdot 10^7$	0.015	[130]
	ZJetsToNuNu_HT-400To600_13TeV-madgraph	11 (LO)	1.0	1.5 (NLO)	$9.5 \cdot 10^5$	0.039	[130]

Table D.2.: Table with all MC samples used in the MUSiC analysis. The weight is the absolute value of the MC weight and is calculated by $\alpha_{MC} = \epsilon_{\text{filter}} * L * \sigma / N_{MC}$. The k-factor order NLO_W corresponds to a NLO k-factor with increased uncertainty as described in 3.2.1.

Process	**Dataset Name**	**σ / pb (Order)**	**Filter Efficiency**	**k-Factor (Order)**	**Number Events**	**Weight α_{MC}**	**Reference**
	ZJetsToNuNu_HT-600ToInf_13TeV-madgraph	4.2 (LO)	1.0	1.4 (NLO)	$1.0 \cdot 10^6$	0.013	[130]
$Z \rightarrow qq$	ZJetsToQQ_HT600toInf_13TeV-madgraph	5.7 (LO)	1.0	-	$9.5 \cdot 10^5$	0.014	[132]
ZZ	ZZTo2L2Nu_13TeV_powheg_pythia8	0.56 (NLO)	1.0	-	$8.8 \cdot 10^6$	$1.5 \cdot 10^{-4}$	[132]
	ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	3.2 (NLO)	1.0	-	$1.5 \cdot 10^7$	$4.8 \cdot 10^{-4}$	[131]
	ZZTo2Q2Nu_13TeV_amcatnloFXFX_madspin_pythia8	4.0 (NLO)	1.0	-	$3.1 \cdot 10^7$	$3.0 \cdot 10^{-4}$	[131]
	ZZTo4L_13TeV_powheg_pythia8	1.3 (NLO)	1.0	-	$6.7 \cdot 10^6$	$4.3 \cdot 10^{-4}$	[132]
	ZZTo4Q_13TeV_amcatnloFXFX_madspin_pythia8	6.8 (NLO)	1.0	-	$3.1 \cdot 10^7$	$5.1 \cdot 10^{-4}$	[132]
ZZZ	ZZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.014 (NLO)	1.0	-	$2.5 \cdot 10^5$	$1.3 \cdot 10^{-4}$	[132]

Appendix E.

Process Groups and Plotting Groups

Plots showing kinematic distribution sometimes aggregate contributions from several process groups into one plotting group to make the plots easier readable. The table E.1 shows the relation between them and the process groups defined in table D.2.

Table E.1.: Assignment between plotting group (shown in plots) and process group (used as input for handling of regions with low MC simulation statistics and correlation of cross section uncertainty).

Plotting Group	Process Group
Drell-Yan	Drell-Yan
	$Z \rightarrow qq$
$Z \rightarrow \nu\nu$	$Z \rightarrow \nu\nu$
$W +$ Jets	W
$\gamma +$ Jets	γ
Multi-Jet	QCD
$t\bar{t}$	$t\bar{t}$
$t\bar{t} + V$	$t\bar{t}W$
	$t\bar{t}Z$
	$t\bar{t}\gamma$
	$t\bar{t}\gamma\gamma$
$t\bar{t}t\bar{t}$	$t\bar{t}t\bar{t}$
Top	Single Top (s-channel)
	Single Top (t-channel)
	tW
	tZq
	$t\gamma$
Multi-Boson	WW
	WZ
	ZZ
	$W\gamma$
	$\gamma\gamma$
	$Z\gamma$
	WWW
	WWZ
	WZZ
	ZZZ
	$WW\gamma$
	$WZ\gamma$

Appendix F.

PDF Sets

All samples produced by the same generator have the same PDF set. The PDF sets used by each generator can be found in table F.1.

Table F.1.: PDF sets for each generator.

Generator	PDF Set	Reference
PYTHIA8	`NNPDF2.3LO`	[134]
MADGRAPH	`NNPDF3.0LO`	[29]
SHERPA	`CT10`	[135]
POWHEG	`NNPDF3.0NLO`	[29]
MADGRAPH5_aMC@NLO	`NNPDF3.0NLO`	[29]

Appendix G.

My Contribution to the MUSiC Analysis

The MUSiC analysis is a project which is performed by many contributers and relies on the work of the whole CMS collaboration. I contributed myself mainly to the following parts of the analysis:

- Assembly of the MC set.
- Studies and validation of the trigger and object selection.
- Implementation of scale factors.
- Implementation and validation of several systematic uncertainties.
- Execution of the full analysis, understanding of the results and visualisation.
- Event scouting with the 2016 dataset.

Bibliography

[1] The ALEPH Collaboration, The DELPHI Collaboration, The L3 Collaboration et al., "Precision Electroweak Measurements on the Z Resonance", *Phys. Rept.* **427** (2006) 257.

[2] The ALEPH Collaboration, The DELPHI Collaboration, The L3 Collaboration et al., "Electroweak Measurements in Electron-Positron Collisions at W-Boson-Pair Energies at LEP", *Phys. Rept.* **532** (2013) 119.

[3] F. Halzen and A. D. Martin, "Quarks and leptons: an introductory course in modern particle physics". Wiley, New York, 1984.

[4] D. J. Griffiths, "Introduction to elementary particles". Physics textbook. Wiley-VCH, Weinheim, 2., rev. ed., 5. reprint edition, 2011. OCLC: 711866653.

[5] M. Erdmann, "Lecture Particle Physics I & II (RWTH Aachen University)", 2014.

[6] K. A. Olive and others, "Review of Particle Physics", *Chin. Phys.* **C38** (2014) 090001. `doi:10.1088/1674-1137/38/9/090001`.

[7] A. Adair, C. Boulahouache, V. Cuplov et al., "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Physics Letters B* **716** (2012), no. 1, 30.

[8] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the {ATLAS} detector at the {LHC}", *Physics Letters B* **716** (2012), no. 1, 1. `doi:http://dx.doi.org/10.1016/j.physletb.2012.08.020`.

[9] N. Cabibbo, "Unitary Symmetry and Leptonic Decays", *Phys. Rev. Lett.* **10** (June, 1963) 531. `doi:10.1103/PhysRevLett.10.531`.

[10] M. Kobayashi and T. Maskawa, "CP-Violation in the Renormalizable Theory of Weak Interaction", *Progress of Theoretical Physics* **49** (February, 1973) 652. `doi:10.1143/PTP.49.652`.

[11] C. S. Wu, Ambler, E., Hayward, R. W. et al., "Experimental Test of Parity Conservation in Beta Decay", *Phys. Rev.* **105** (February, 1957) 1413.

[12] Y. Fukuda, T. Hayakawa, E. Ichihara et al., "Evidence for oscillation of atmospheric neutrinos", *Physical Review Letters* **81** (1998), no. 8, 1562.

[13] Q. R. Ahmad, R. C. Allen, T. C. Andersen et al., "Measurement of the Rate of nu e + d -> p + p + e - Interactions Produced by B 8 Solar Neutrinos at the Sudbury Neutrino Observatory", *Physical Review Letters* **87** (July, 2001) 071301. `doi:10.1103/PhysRevLett.87.071301`.

[14] Q. R. Ahmad, R. C. Allen, T. C. Andersen et al., "Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory", *Physical Review Letters* **89** (June, 2002) 011301. `doi:10.1103/PhysRevLett.89.011301`.

[15] S. L. Glashow, "Partial-symmetries of weak interactions", *Nuclear Physics* **22** (February, 1961) 579. `doi:10.1016/0029-5582(61)90469-2`.

[16] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (November, 1967) 1264. `doi:10.1103/PhysRevLett.19.1264`.

[17] A. Salam, "Weak and electromagnetic interactions",. Proc. of the 8th Nobel Symposium on 'Elementary particle theory, relativistic groups and analyticity'.

[18] P. W. Higgs, "Broken symmetries, massless particles and gauge fields", *Physics Letters* **12** (1964), no. 2, 132.

[19] F. Englert and R. Brout, "Broken symmetry and the mass of gauge vector mesons", *Physical Review Letters* **13** (1964), no. 9, 321. `doi:10.1103/PhysRev.105.1413`.

[20] P. W. Higgs, "Broken symmetries and the masses of gauge bosons", *Physical Review Letters* **13** (1964), no. 16, 508.

[21] G. S. Guralnik, C. R. Hagen, and T. W. Kibble, "Global conservation laws and massless particles", *Physical Review Letters* **13** (1964), no. 20, 585.

[22] P. W. Higgs, "Spontaneous symmetry breakdown without massless bosons", *Physical Review* **145** (1966), no. 4, 1156.

[23] T. W. B. Kibble, "Symmetry breaking in non-Abelian gauge theories", *Physical Review* **155** (1967), no. 5, 1554.

[24] T. C. Collaboration, "Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV", *The European Physical Journal C* **75** (2015), no. 5, 212. `doi:10.1140/epjc/s10052-015-3351-7`.

[25] T. Gleisberg, S. Höche, F. Krauss et al., "Event generation with SHERPA 1.1", *Journal of High Energy Physics* **2009** (2009), no. 02, 007.

[26] J. Alwall, M. Herquet, F. Maltoni et al., "MadGraph 5: going beyond", *Journal of High Energy Physics* **2011** (2011), no. 6, 1. `doi:10.1007/JHEP06(2011)128`.

[27] T. Sjöstrand, S. Ask, J. R. Christiansen et al., "An introduction to PYTHIA 8.2", *Computer Physics Communications* **191** (2015) 159. `doi:http://dx.doi.org/10.1016/j.cpc.2015.01.024`.

[28] V. Khachatryan, A. M. Sirunyan, A. Tumasyan et al., "Event generator tunes obtained from underlying event and multiparton scattering measurements", *The European Physical Journal C* **76** (2016), no. 3, 1. `doi:10.1140/epjc/s10052-016-3988-x`.

[29] R. D. Ball, V. Bertone, S. Carrazza et al., "Parton distributions for the LHC run II", *Journal of High Energy Physics* **2015** (2015), no. 4, 40. `doi:10.1007/JHEP04(2015)040`.

[30] V. N. Gribov and L. N. Lipatov, "Deep inelastic electron scattering in perturbation theory", *Physics Letters B* **37** (1971), no. 1, 78.

[31] Y. L. Dokshitzer, "Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics.", *Sov. Phys. JETP* **46** (1977) 641. [Zh. Eksp. Teor. Fiz.73,1216(1977)].

[32] G. Altarelli and G. Parisi, "Asymptotic freedom in parton language", *Nuclear Physics B* **126** (1977), no. 2, 298.

[33] J. Alwall and others, "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions", *Eur. Phys. J.* **C53** (2008) 473. `doi:10.1140/epjc/s10052-007-0490-5`.

[34] B. Andersson, G. Gustafson, G. Ingelman et al., "Parton fragmentation and string dynamics", *Physics Reports* **97** (1983), no. 2, 31. `doi:http://dx.doi.org/10.1016/0370-1573(83)90080-7`.

[35] G. Marchesini, B. R. Webber, G. Abbiendi et al., "HERWIG 5.1 - a Monte Carlo event generator for simulating hadron emission reactions with interfering gluons", *Computer Physics Communications* **67** (1992), no. 3, 465. `doi:http://dx.doi.org/10.1016/0010-4655(92)90055-4`.

[36] R. J. Gaitskell, "Direct detection of dark matter", *Annual Review of Nuclear and Particle Science* **54** (December, 2004) 315. `doi:10.1146/annurev.nucl.54.070103.181244`.

[37] A. G. Cohen, D. B. Kaplan, and a. A. E. Nelson, "Progress in Electroweak Baryogenesis", *Annual Review of Nuclear and Particle Science* **43** (1993), no. 1, 27. `doi:10.1146/annurev.ns.43.120193.000331`.

[38] A. Riotto and M. Trodden, "RECENT PROGRESS IN BARYOGENESIS", *Annual Review of Nuclear and Particle Science* **49** (1999), no. 1, 35. `doi:10.1146/annurev.nucl.49.1.35`.

[39] W.-S. Hou, "Source of CP Violation for the Baryon Asymmetry of the Universe", *Chinese Journal of Physics* **47** (2009), no. 2, 134.

[40] E. Gildener, "Gauge-symmetry hierarchies", *Physical Review D* **14** (1976), no. 6, 1667.

[41] D. M. Cinzia, "CERN Accelerator Complex.jpg (JPEG Image)", August, 2016.

[42] L. Evans and P. Bryant, "LHC Machine", *Journal of Instrumentation* **3** (2008), no. 08, S08001.

[43] "LEP design report". CERN, Geneva, 1984. Copies shelved as reports in LEP, PS and SPS libraries.

[44] The ALICE Collaboration, "The ALICE experiment at the CERN LHC", *Journal of Instrumentation* **3** (2008), no. 08, S08002.

[45] The LHCb Collaboration, "The LHCb Detector at the LHC", *Journal of Instrumentation* **3** (2008), no. 08, S08005.

[46] The ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *Journal of Instrumentation* **3** (2008), no. 08, S08003.

[47] The CMS Collaboration, "The CMS experiment at the CERN LHC", *Journal of Instrumentation* **3** (2008), no. 08, S08004.

[48] T. C. Collaboration), "Pileup measurement and mitigation techniques in CMS", *Journal of Physics: Conference Series* **404** (2012), no. 1, 012045.

[49] CMS Collaboration, "Energy resolution of the barrel of the CMS Electromagnetic Calorimeter", *Journal of Instrumentation* **2** (2007), no. 04, P04004.

[50] CMS Collaboration, "Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in pp Collisions at sqrt(s) = 7 TeV", *JINST* **8** (2013) P09009. [JINST8,9009(2013)]. `doi:10.1088/1748-0221/8/09/P09009`.

[51] The CMS collaboration, "Performance of the CMS missing transverse momentum reconstruction in pp data at sqrt(s) = 8 TeV", *Journal of Instrumentation* **10** (2015), no. 02, P02006.

[52] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", Technical Report arXiv:1607.03663. CERN-PH-EP-2015-305. CMS-JME-13-004, CERN, Geneva, July, 2016.

[53] The CMS collaboration, "Performance of CMS muon reconstruction in pp collision events at sqrt(s) = 7 TeV", *Journal of Instrumentation* **7** (2012), no. 10, P10002.

[54] U. Schwickerath, R. Jones, J. Shiers et al., "LHC computing Grid", technical report, CERN, 2005.

[55] The CMS Collaboration, "CMS Luminosity Measurement for the 2015 Data Taking Period", Technical Report CMS-PAS-LUM-15-001, CERN, Geneva, 2016.

[56] J. Blümlein, S. Moch, T. Riemann et al., "Loops and Legs in Quantum Field Theory The POWHEG BOX", *Nuclear Physics B - Proceedings Supplements* **205** (2010) 36. `doi:http://dx.doi.org/10.1016/j.nuclphysbps.2010.08.016`.

[57] J. Alwall, R. Frederix, S. Frixione et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *Journal of High Energy Physics* **2014** (2014), no. 7, 1. `doi:10.1007/JHEP07(2014)079`.

[58] J. Blümlein, S. Moch, T. Riemann et al., "Loops and Legs in Quantum Field Theory MCFM for the Tevatron and the LHC", *Nuclear Physics B - Proceedings Supplements* **205** (2010) 10. `doi:http://dx.doi.org/10.1016/j.nuclphysbps.2010.08.011`.

[59] S. Agostinelli, J. Allison, K. Amako et al., "Geant4—a simulation toolkit", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** (2003), no. 3, 250. `doi:http://dx.doi.org/10.1016/S0168-9002(03)01368-8`.

[60] CMS Collaboration, "MUSIC – An Automated Scan for Deviations between Data and Monte Carlo Simulation", Technical Report CMS-PAS-EXO-08-005, CERN, 2008. Geneva, October, 2008.

[61] P. A. Biallass, "Commissioning of the CMS muon detector and development of generic search strategies for new physics". PhD thesis, Publikationsserver der RWTH Aachen University, Aachen, 2009. Aachen, Techn. Hochsch., Diss., 2009.

[62] C. Hof, "Implementation of a model independent search for new physics with the CMS detector exploiting the world-wide LHC computing grid". PhD thesis, Publikationsserver der RWTH Aachen University, Aachen, 2009. Aachen, Techn. Hochsch., Diss., 2009.

[63] H. Pieta, "MUSiC-A Model Unspecific Search in CMS based on 2010 LHC data". PhD thesis, Rheinisch-Westfälischen Technischen Hochschule Aachen, 2012.

[64] CMS Collaboration, "Model Unspecific Search for New Physics in pp Collisions at sqrt(s) = 7 TeV", Technical Report CMS-PAS-EXO-10-021, CERN, Geneva, 2011.

[65] P. Papacz, "Model Unspecific Search for new Physics in CMS Based on 2011 Data". PhD thesis, Rheinisch-Westfälischen Technischen Hochschule Aachen, 2014.

[66] D. Duchardt, "MUSiC: A Model Unspecific Search for New Physics Based on CMS Data at sqrt(s) = 8 TeV". PhD thesis, RWTH Aachen University, 2017. In preparation.

[67] CMS Collaboration, "MUSiC - A Model Unspecific Search for New Physics in pp Collisions at sqrt(s) = 8 TeV", Technical Report CMS-PAS-EXO-14-016, CERN, Geneva, 2016.

[68] CMS Collaboration, "Search for a high-mass resonance decaying into a dilepton final state in 13 fb-1 of pp collisions at sqrt(s) =13 TeV", Technical Report CMS-PAS-EXO-16-031, CERN, Geneva, 2016.

[69] CMS Collaboration, "Search for Extra Dimensions in Dimuon Events in pp Collisions at sqrt(s) = 8 TeV", Technical Report CMS-PAS-EXO-12-027, CERN, Geneva, 2013.

[70] H.-P. Bretz, M. Brodski, M. Erdmann et al., "A development environment for visual physics analysis", *Journal of Instrumentation* **7** (2012), no. 08, T08005.

[71] T. Hebbeker, "A Global Comparison between L3 Data and Standard Model Monte Carlo-a first attempt", technical report, L3 note 2305. 1998. url: http://web. physik. rwth-aachen. de/ hebbeker/l3note_2305. pdf.(Cit. on p. 51), 1998.

[72] The L3 Collaboration, "The construction of the L3 experiment", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **289** (1990), no. 1, 35. doi:http://dx.doi.org/10.1016/0168-9002(90)90250-A.

[73] DØ Collaboration, "Search for new physics in emuX data at DØ using SLEUTH: A quasi-model-independent search strategy for new physics", *Phys. Rev. D* **62** (October, 2000) 092004. doi:10.1103/PhysRevD.62.092004.

[74] DØ Collaboration, "Quasi-model-independent search for new physics at large transverse momentum", *Phys. Rev. D* **64** (June, 2001) 012004. doi:10.1103/PhysRevD.64.012004.

[75] CDF Collaboration, "Model-Independent and Quasi-Model-Independent Search for New Physics at CDF", *Phys. Rev.* **D78** (2008) 012002. doi:10.1103/PhysRevD.78.012002.

[76] CDF Collaboration, "Global search for new physics with 2.0/fb at CDF", *Phys. Rev. D* **79** (January, 2009) 011101. doi:10.1103/PhysRevD.79.011101.

[77] DØ Collaboration, "The D0 detector", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **338** (1994), no. 2, 185. doi:http://dx.doi.org/10.1016/0168-9002(94)91312-9.

[78] CDF Collaboration, "The CDF detector: an overview", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **271** (1988), no. 3, 387. doi:http://dx.doi.org/10.1016/0168-9002(88)90298-7.

[79] H1 Collaboration, "A General search for new phenomena in ep scattering at HERA", *Phys. Lett.* **B602** (2004) 14. doi:10.1016/j.physletb.2004.09.057.

[80] R. M. Bianchi and G. Herten, “A model-independent "General Search" for new physics with the ATLAS detector at LHC”. PhD thesis, Freiburg U., Freiburg, June, 2014. Presented 18 Mar 2014.

[81] CMS Collaboration, “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET”, Technical Report CMS-PAS-PFT-09-001, CERN, 2009. Geneva, April, 2009.

[82] F. Beaudette, “The CMS Particle Flow Algorithm”,.

[83] P. Billoir, “Progressive track recognition with a Kalman-like fitting procedure”, *Computer Physics Communications* **57** (1989), no. 1, 390. `doi:http://dx.doi.org/10.1016/0010-4655(89)90249-X`.

[84] P. Billoir and S. Qian, “Simultaneous pattern recognition and track fitting by the Kalman filtering method”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **294** (1990), no. 1, 219. `doi:http://dx.doi.org/10.1016/0168-9002(90)91835-Y`.

[85] R. Mankel, “A concurrent track evolution algorithm for pattern recognition in the HERA-B main tracking system”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **395** (1997), no. 2, 169. `doi:http://dx.doi.org/10.1016/S0168-9002(97)00705-5`.

[86] G. Abbiendi, N. Adam, J. Alcaraz et al., “Muon Reconstruction in the CMS Detector”, July, 2009.

[87] W. Adam, R. Frühwirth, A. Strandlie et al., “Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC”, *Journal of Physics G: Nuclear and Particle Physics* **31** (2005), no. 9, N9.

[88] G. P. Salam and G. Soyez, “A practical seedless infrared-safe cone jet algorithm”, *Journal of High Energy Physics* **2007** (2007), no. 05, 086. `doi:10.1088/1126-6708/2007/05/086`.

[89] M. Cacciari, G. P. Salam, and G. Soyez, “The anti-k_t jet clustering algorithm”, *Journal of High Energy Physics* **2008** (April, 2008) 063. `doi:10.1088/1126-6708/2008/04/063`.

[90] CMS TWiki, “Jet Energy Resolution”. https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution?rev=54#JER_Scaling_factors_and_Uncertai, July, 2016.

[91] A. Grebenyuk and H. Brun, “Di-muon trigger measurement with the reference trigger method: systematic uncertainties”, January, 2017.

[92] The CMS Collaboration, “Performance of missing energy reconstruction in 13 TeV pp data using the CMS detector”, Technical Report CMS PAS JME-16-004, CERN, Geneva, July, 2016.

[93] CMS TWiki, “Cut Based Electron ID for Run 2”. https://twiki.cern.ch/twiki/bin/view/CMS/CutBasedElectronIdentificationRun2?rev=34#Spring15_selection_25ns, July, 2016.

[94] CMS TWiki, “HEEP Electron ID and isolation”. https://twiki.cern.ch/twiki/bin/view/CMS/HEEPElectronID?rev=68, June, 2015.

[95] CMS TWiki, "Electron Gamma PF Based Isolation Run-II". https://twiki.cern.ch/twiki/bin/view/CMS/EgammaPFBasedIsolationRun2?rev=16, May, 2016.

[96] The CMS Collaboration, "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at sqrt(s) = 8 TeV", *Journal of Instrumentation* **10** (2015), no. 06, P06005.

[97] CMS TWiki, "HEEP Electron ID and isolation". https://twiki.cern.ch/twiki/bin/view/CMS/HEEPElectronIdentificationRun2?rev=17#Selection_Cuts_HEEP_V6_0_Recomme, April, 2016.

[98] CMS TWiki, "Cut Based Photon ID for Run 2". https://twiki.cern.ch/twiki/bin/view/CMS/CutBasedPhotonIdentificationRun2?rev=28#SPRING15_selections_25_ns, February, 2016.

[99] CMS Collaboration, "Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at sqrt(s) = 8 TeV", *JINST* **10** (2015), no. 08, P08010. `doi:10.1088/1748-0221/10/08/P08010`.

[100] CMS TWiki, "Baseline muon selections for Run-II". https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonIdRun2?rev=26, July, 2016.

[101] CMS TWiki, "Jet Identification". https://twiki.cern.ch/twiki/bin/view/CMS/JetID?rev=94#Recommendations_for_13_TeV_data, March, 2016.

[102] CMS TWiki, "MiniAOD Analysis Documentation". https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookMiniAOD2015?rev=96#ETmiss, April, 2016.

[103] CMS TWiki, "Reference muon id, isolation and trigger efficiencies for Run-II". https://twiki.cern.ch/twiki/bin/viewauth/CMS/MuonReferenceEffsRun2?rev=18#Results_for_CMSSW_7_6_X_dataset, July, 2016.

[104] CMS TWiki, "Instructions for applying electron and photon ID". https://twiki.cern.ch/twiki/bin/view/CMS/EgammaIDRecipesRun2?rev=32#Efficiencies_and_scale_factors, July, 2016.

[105] M. Diamantopoulou, N. Saoulidou, and E. Tziaferi, "Performance of the Particle-Flow jet identification criteria using proton-proton collisions at 13 TeV", Technical Report CMS AN-15-269, CERN, Geneva, March, 2016.

[106] The CMS Collaboration, "Measurement of the inelastic proton-proton cross section at sqrt(s)=13 TeV", Technical Report CMS-PAS-FSQ-15-005, CERN, Geneva, 2016.

[107] J. Butterworth and others, "PDF4LHC recommendations for LHC Run II", *J. Phys.* **G43** (2016) 023001. `doi:10.1088/0954-3899/43/2/023001`.

[108] J. Bendavid, "New MC Features for Analysis in 74x Plus more details on scale/PDF uncertainties", November, 2015.

[109] A. Buckley, J. Ferrando, S. Lloyd et al., "LHAPDF6: parton density access in the LHC precision era", *Eur. Phys. J.* **C75** (2015) 132. `doi:10.1140/epjc/s10052-015-3318-8`.

[110] CMS TWiki, “Theory systematics”.
https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TheorySystematics?rev=13, October, 2015.

[111] E. Yazgan and M. Seidel, “ttbar Jet Multiplicity Discrepancy - TOP PAG Report”, February, 2016.

[112] A. E. d. Valle and R. Castello, “Muon momentum scale at HighPt: Update in Generalized Endpoint”, April, 2016.

[113] R. Castello, “Measurement of the high pT muon momentum scale using a generalized endpoint method”, Technical Report CMS AN-2015/228, CERN, Geneva, October, 2015.

[114] CMS Twiki, “Official Prescription for calculating uncertainties on Missing Transverse Energy (MET)”.
https://twiki.cern.ch/twiki/bin/view/CMS/MissingETUncertaintyPrescription?rev=51, August, 2016.

[115] M. D’Alfonso, R. Schoefbeck, and and others, “Scale and resolution of ETmiss in early 13 TeV data”, Technical Report CMS AN-15-157, CERN, Geneva, July, 2016.

[116] The CMS Collaboration, “Validation of key observables used in SUSY searches with first data at sqrt(s) = 13 TeV”, Technical Report CMS-PAS-SUS-15-001, CERN, Geneva, August, 2015.

[117] B. Clerbaux, D. Cockerill, G. Daskalakis et al., “Dielectron resonance search in Run 2 at sqrt(s) = 13 TeV pp collisions”, Technical Report CMS AN-2015/222, CERN, Geneva, May, 2016.

[118] F. Blekman, D. Burns, E. Clement et al., “Top Pair Differential Cross Section Measurements with respect to Global Event Variables using 25ns Data at 13TeV”, Technical Report CMS-AN-2016/092, CERN, Geneva, July, 2016.

[119] The CMS Collaboration, “Search for a narrow spin-2 resonance decaying to Z bosons in the semileptonic final state”, Technical Report CMS-PAS-EXO-12-022, CERN, Geneva, 2013.

[120] The CMS Collaboration, “Search for massive resonances decaying into pairs of boosted W and Z bosons at sqrt(s) = 13 TeV”, Technical Report CMS-PAS-EXO-15-002, CERN, Geneva, 2015.

[121] K. Agashe, H. Davoudiasl, G. Perez et al., “Warped gravitons at the CERN LHC and beyond”, *Phys. Rev. D* **76** (August, 2007) 036006. `doi:10.1103/PhysRevD.76.036006`.

[122] D. Pappadopulo, A. Thamm, R. Torre et al., “Heavy vector triplets: bridging theory and data”, *Journal of High Energy Physics* **2014** (2014), no. 9, 60.
`doi:10.1007/JHEP09(2014)060`.

[123] Erik Butz, “Mitigation of the Strip Tracker Dynamic Inefficiency (previously known as HIP)”, August, 2016.

[124] G. Cerminara and M. Klute, “News”, September, 2016.

[125] CMS TWiki, “MET Filter Recommendations for Run II”. https://twiki.cern.ch/twiki/bin/viewauth/CMS/MissingETOptionalFiltersRun2?rev=97#76X_Recommendations_and_Recipes, July, 2016.

[126] CMS TWiki, “Beam Halo Id in CMS”. https://twiki.cern.ch/twiki/bin/view/CMS/BeamHaloId?rev=18, August, 2011.

[127] CMS TWiki, “MET Optional Filters”. https://twiki.cern.ch/twiki/bin/view/CMS/MissingETOptionalFilters?rev=66, May, 2015.

[128] CMS TWiki, “HCAL Noise Library”. https://twiki.cern.ch/twiki/bin/view/CMS/HcalNoiseInfoLibrary?rev=47, May, 2012.

[129] CMS TWiki, “ECAL Maksed Cell Issue Summary for SUSY Analyses”. https://twiki.cern.ch/twiki/bin/view/CMS/SusyEcalMaskedCellSummary?rev=32, October, 2011.

[130] L. Benato, Y.-H. Chang, C.-W. Chen et al., “Search for heavy resonances decaying into a vector boson and a Higgs boson in the (nu nu, l nu, l l) bb final state”, Technical Report CMS AN-2015/186, CERN, Geneva, May, 2016.

[131] CMS TWiki, “Summary table of samples produced for the 1 Billion campaign, with 25ns bunch-crossing”. https://twiki.cern.ch/twiki/bin/viewauth/CMS/SummaryTable1G25ns?rev=146, November, 2016.

[132] The CMS Collaboration, “McM Monte-Carlo Request Management”. https://cms-pdmv.cern.ch/mcm/.

[133] K. Padeken, M. Olschewski, and K. Hoepfner, “Cross section calculation for leptonicaly decaying W at NNLO QCD and NLO electroweak”, Technical Report CMS AN AN-14-263, CERN, Geneva, October, 2015.

[134] R. D. Ball, V. Bertone, S. Carrazza et al., “Parton distributions with LHC data”, *Nuclear Physics B* **867** (2013), no. 2, 244. `doi:http://dx.doi.org/10.1016/j.nuclphysb.2012.10.003`.

[135] H.-L. Lai, M. Guzzi, J. Huston et al., “New parton distributions for collider physics”, *Phys. Rev. D* **82** (October, 2010) 074024. `doi:10.1103/PhysRevD.82.074024`.

Danksagung

Abschließend möchte ich allen, die mir geholfen und es möglich gemacht haben diese Arbeit zu schreiben, meinen Dank aussprechen.

Zunächst danke ich Herrn Professor Hebbeker für die Möglichkeit, im III. Physikalischen Institut A in einer guten, inspirierender Umgebung an aktuellen Daten forschen zu können und damit diese Arbeit zu schreiben.

Des Weiteren möchte ich Arnd Meyer danken, der mir sowohl mit Ratschlägen zur Analyse als auch mit Korrekturen der Masterarbeit zu Hilfe stand.

Besonders möchte ich meinem Betreuer Tobias Pook danken, der nicht nur die Analyse gut organisiert hat und jederzeit für Fragen bereitstand, sondern auch meine Arbeit Korrektur gelesen hat. Debbie, Simon und Jonas, die an der gleichen Analyse arbeiten, danke ich für ihre hilfreichen Tipps. Herzlichen Dank auch an Simon für das Korrekturlesen der Arbeit.

Auch der gesamten Arbeitsgruppe, der ich jederzeit Fragen stellen konnte und die mir häufig Ratschläge für die Analyse gegeben hat, möchte ich ein besonderes Dankeschön aussprechen. Die lockere und freundliche Arbeitsatmosphäre sowohl in meinem Büro als auch in der gesamten Arbeitsgruppe hat sehr dazu beigetragen, dass die Arbeit mir viel Spaß gemacht hat.

Schließlich danke ich ganz besonders meinen Freunden und meiner Familie für die Unterstützung abseits der Arbeit auch in stressigen Zeiten.

Selbstständigkeitserklärung

Ich versichere, dass ich die Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie Zitate kenntlich gemacht habe.

Aachen, den 16.01.2017

Jonas Roemer