MUSiC: A Model Unspecific Search for New Physics Based on CMS Data at $\sqrt{s} = 8$ TeV

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften der RWTH Aachen University zur Erlangung des akademischen Grades einer Doktorin der Naturwissenschaften genehmigte Dissertation

vorgelegt von

Diplom-Physikerin

Deborah Duchardt

aus Leverkusen

Berichter:

Univ.-Prof. Dr. rer. nat. Thomas Hebbeker
Univ.-Prof. Dr. rer. nat. Christopher Wiebusch

Tag der mündlichen Prüfung: 16. August 2017

Abstract

Completing the first data-taking period of the Large Hadron Collider, around 20 fb^{-1} worth of proton-proton events produced in 2012 and recorded with the CMS experiment are considered in terms of a model independent analysis referred to as the "Model Unspecific Search in CMS" (MUSiC). The presented method aims at being unbiased, thus exempting dedicated theory model assumptions and kinematic selections made to uncover new effects beyond the established Standard Model of particle physics. Instead the search is carried out in several hundreds of final states, called event classes, containing electrons, muons, photons, jets as well as missing transverse energy. In doing so, the MUSiC analysis investigates regions of data even if no concrete signatures are anticipated there. In each event class three kinematic variables are examined by comparing the measured collision data with a Monte-Carlo simulated prediction of the Standard Model. A statistical measure of significance is determined for the strongest deviation in each event class distribution and a global overview of all results is given. After elaborating on prevailing aspects of the MUSiC analysis and new methods developed for this thesis, the algorithm is applied to the 2012 CMS data. The search results show a very good agreement between data and Standard Model Monte-Carlo expectation, revealing no obvious signs of new phenomena. All observed deviations can be explained within the systematic uncertainties of the Standard Model prediction. Moreover, the sensitivity of the analysis method is probed using simulated pseudo-data.

Zusammenfassung

Als Abschluss der ersten Datennahmephase des Large Hadron Collider Beschleunigers werden rund 20 fb^{-1} an Proton-Proton Ereignissen, welche mit dem CMS Experiment aufgezeichnet wurden, unter dem Gesichtspunkt der modellunabhängigen Analyse "Model Unspecific Search in CMS"(MUSiC) betrachtet. Die vorgestellte Methode zielt darauf ab unvoreingenommen zu sein, indem davon abgesehen wird dedizierte Annahmen zu theoretischen Modellen oder zur kinematischen Selektion zu machen, die dazu dienen bestimmte neue Effekte jenseits des etablierten Standardmodells der Teilchenphysik aufzudecken. Stattdessen wird die Suche in hunderten von Endzuständen, sogenannten Ereignisklassen, durchgeführt, welche Elektronen, Myonen, Photonen, Jets sowie fehlende transversale Energie enthalten können. Durch dieses Verfahren begutachtet die MUSiC Analyse auch Bereiche der Daten in denen keine konkreten Signaturen vorhergesagt sind. In jeder Ereignisklasse werden drei kinematische Variablen untersucht, indem die gemessenen Kollisionsdaten mit einer Monte-Carlo simulierten Vorhersage des Standardmodells verglichen werden. Für jede Verteilung einer Ereignisklasse wird ein statistisches Maß für die Stärke der größten Abweichung bestimmt, welche schlussendlich zu einem globalen Überblick aller Ergebnisse zusammengefügt werden. Nachdem zunächst die bestehenden Aspekte der MUSiC Analyse, sowie einige neue Entwicklungen erläutert werden, wird der Algorithmus auf die CMS Daten von 2012 angewandt. Die Ergebnisse der Suche zeigen eine sehr gute Übereinstimmung der Messung mit der Monte-Carlo Erwartung des Standardmodells, ohne klare Hinweise auf neue Phänomene aufzuweisen. Alle beobachteten Abweichungen können innerhalb der systematischen Unsicherheiten auf die Standardmodell-Vorhersage erklärt werden. Des weiteren wird die Sensitivität der Analyse Methode anhand von simulierten Pseudo-Daten untersucht.

Contents

Chapter 1
Introduction

As proven by the discovery of the long sought after Higgs boson by the CMS and ATLAS Collaborations, the LHC Run I, conduced from 2010 through 2012, presents an ideal experimental environment to examine signatures of the **S**tandard **M**odel (SM) of particle physics and search for phenomena beyond the established description of the sub-atomic domain. Analyses of collider data are often tailored to specific search channels, corresponding to a concrete set of final state particles, centered around dedicated models to describe the investigated **B**eyond **S**tandard **M**odel (BSM) scenarios. Though a great sensitivity to individual signatures can be reached with such an approach by optimizing the search strategy to a high signal-over-background ratio, it is rather time-consuming and ill-suited to cover a large number of final states. Furthermore, only regions are investigated that are perceived as promising for an existing BSM theory. In consequence, new effects might be overlooked simply because no analysis is in place for the specific part of phase-space.

To mitigate this data loss an alternative strategy is chosen. The **M**odel **U**nspecific **S**earch **i**n **C**MS (MUSiC) approach follows a BSM model independent concept where the measured proton-proton collision data is compared to Monte Carlo simulations of the SM. Hundreds of final states, referred to as event classes, are investigated taking three kinematic variables into consideration. The applied kinematic selection is kept at a minimum to avoid bias. A search algorithm automatically scans each distribution for the strongest upward or downward deviation of the data from the SM expectation, for which a statistical significance is calculated. The distribution of observed deviations is compared to expected deviations given the SM prediction and its uncertainties, offering a consistent global overview of the analyzed data.

Following a similar strategy as carried out at the LEP, Tevatron and HERA colliders, the MUSiC method is applied to proton-proton collision data taken with the CMS detector. Two separate analyses have already been conducted on data recorded in 2010 [1] and 2011 [2] at a center of mass energy of $\sqrt{s} = 7$ TeV, relying on methods established before data-taking [3, 4]. The work presented in this document can be understood as the continuation of these analyses in the context of data-taking during 2012 at $\sqrt{s} = 8$ TeV, which, yielding about 20 fb^{-1}, provides the largest consistent set of data collected for all analyses mentioned above. Within MUSiC a focus is placed on events containing electrons or muons, and further developments of the general methodology are included. A full account of the analysis steps and obtained results is given.

This thesis is built up into four parts. Initially, Part I summarizes the theoretical basis of the analysis introducing the SM and the necessity for theories beyond. The experimental setup is addressed in Part II with a concise account of the LHC and its CMS detector. An emphasis is set on algorithms used to reconstruct particle physics objects from the measured quantities in the detector. A full account of the search methodology is given in Part III, including information from previous works and new developments introduced for the analysis leading up to this thesis. The results of the MUSiC approach performed on the 2012 CMS data taken at $\sqrt{s} = 8$ TeV are presented in Part IV. This includes a full statistical analysis of the data with respect to its agreement with simulations of the SM. Furthermore, the sensitivity of the analysis is probed and an outlook concerning future developments of the MUSiC analysis is given.

As is customary in high energy physics, natural units are implied throughout this thesis, setting c and $\hbar$ to unity. Advantages of this convention are simplified formulas and a unit system reflecting physical boundaries relevant to elementary particles. Energies are given in multiples of electron

volts (eV), representing the amount of energy gained by an electron passing through an electrostatic potential difference of one volt. Thus, the dimensions of basic quantities can be expressed as:

$$[\text{energy}] = [\text{mass}] = [\text{momentum}] = [\text{length}]^{-1} = [\text{time}]^{-1} = \text{eV}. \tag{1.1}$$

Individual components of the analysis elaborated on in this document are published in form of a CMS Physics Analysis Summary CMS PAS EXO-14-016 [5] with accompanying CMS Analysis Note CMS AN 2014/098 [6], of which I am the leading author. Figures taken directly from the publication are labeled with "CMS Preliminary" and suitable attribution is given in the caption. Figures labeled with "CMS Private" were developed for this thesis and are, in this form, not part of the public analysis documentation.

Part I

Theoretical Framework

Chapter 2
The Standard Model of Particle Physics

A driving force of particle physics is the pursuit of the fundamental constituents of matter and interactions. The SM of elementary particle physics represents a renormalizable quantum field theory identified by the $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge group. Electromagnetic and weak interactions are unified and the strong interaction is embedded to provide a description of observed particles such as protons and neutrons. Within this mathematical framework, particles and forces are expressed as quantum fields complying with the laws of special relativity.

The following chapter summarizes characteristics of the SM such as its particle content and the fundamental forces. Although eminently important on the macroscopic scale, the gravitational force is negligible in terms of elementary particle interactions. Einstein's theory of relativity [7] extends Newton's classical description for high energies and velocities close to the speed of light. Devising a consistent quantum mechanical theory of gravitational effects is still one of the outstanding questions of modern day physics. Only recently the first observation of gravitational waves originating from a binary black hole merger was reported by the LIGO and Virgo Collaborations, setting an upper bound of $1.2 \cdot 10^{-22}$ eV on the mass of the hypothetical graviton [8]. In absence of a dedicated reference, information for this chapter is taken from [9–12] and references therein.

2.1 Particle Content of the Standard Model

Within the SM a possible categorization of all elementary particles is given by identifying fermions with half-integer spin and bosons with integer spin. An overview of all known elementary particles and their characteristic properties can be found in Tab. 2.1 and Tab. 2.2. Particles are called elementary if they do not show an internal structure and can be assumed as point-like. Present-day accelerators operate at center of mass energies of several TeV and, following Heisenberg's uncertainty principle, are thus able to resolve distances down to the order of 10^{-20} m [9, 10]. Searches at these colliders have not observed any indications of particle substructure [13–15].

Elementary fermions are the constituents of all visible matter and can be subdivided into those affected by the strong interaction (see Sec. 2.4), called quarks, and those that are not, called leptons. With respect to the weak interaction quarks and leptons can each be grouped into three left-handed doublets and one right-handed singlet. As indicated in Tab. 2.1, these groups of multiplets are referred to as generations [18]. Charged particle masses increase from one generation to the next, where only first generation particles are stable.

Neutrinos are uncharged fermions participating only in the weak interaction. Though the minimum SM assumes them to be massless, numerous experiments based on atmospheric [19], solar [20], reactor [21] and accelerator [22] neutrinos indicate oscillations between the three generations, necessitating non-zero masses. Upper bounds can be given for the individual neutrino masses from dedicated experiments. The electron neutrino mass is evaluated from the energy spectrum of tritium β-decays, giving an upper limit of $m_{\nu_e} \lesssim 2$ eV [23, 24]. Measurements of charged pion decays deliver a value of $m_{\nu_\mu} \lesssim 0.19$ MeV [25]. The tau neutrino mass is reconstructed from tau decays to $m_{\nu_\tau} \lesssim 18.2$ MeV [26].

	I	II	III	Spin	Charge	Isospin
quarks	$\boldsymbol{u}$ up quark $m \approx 2.3$ MeV	$\boldsymbol{c}$ charm quark $m \approx 95$ MeV	$\boldsymbol{t}$ top quark $m \approx 173$ GeV	$\frac{1}{2}$	$+\frac{2}{3}$	$+\frac{1}{2}$
	$\boldsymbol{d}$ down quark $m \approx 4.8$ MeV	$\boldsymbol{s}$ strange quark $m \approx 1.28$ GeV	$\boldsymbol{b}$ bottom quark $m \approx 4.18$ GeV	$\frac{1}{2}$	$-\frac{1}{3}$	$-\frac{1}{2}$
leptons	$\boldsymbol{\nu_e}$ electron neutrino $m \lesssim 2$ eV	$\boldsymbol{\nu_\mu}$ muon neutrino $m \lesssim 0.19$ MeV	$\boldsymbol{\nu_\tau}$ tau neutrino $m \lesssim 18.2$ MeV	$\frac{1}{2}$	0	$+\frac{1}{2}$
	e electron $m = 0.511$ MeV	$\boldsymbol{\mu}$ muon $m = 106$ MeV	$\boldsymbol{\tau}$ tau $m = 1777$ MeV	$\frac{1}{2}$	-1	$-\frac{1}{2}$

Table 2.1: Matter fields, overview of all fermions in the SM. Roman numerals indicate the quark and lepton generations. Details on the quark masses given in Sec. 2.4. The electric charge q is given in units of the elementary charge $e = 1.602 \cdot 10^{-19}$ C. Values given for the weak isospin T_3 refer to left-handed particles. All properties are taken from [16].

	Spin	Charge	Mass
$\boldsymbol{\gamma}$ photon	1	0	0
$\boldsymbol{W^\pm}$ W boson	1	± 1	80.4 GeV
$\boldsymbol{Z^0}$ Z boson	1	0	91.2 GeV
$\boldsymbol{g}$ gluon	1	0	0
$\boldsymbol{H^0}$ Higgs boson	0	0	125 GeV

Table 2.2: Gauge fields, overview of all bosons in the SM. The electric charge q is given in units of the elementary charge $e = 1.602 \cdot 10^{-19}$ C. The stated photon and gluon masses correspond to their SM prediction. The Higgs boson mass is given as reported in [17]. All other properties are taken from [16].

Derived by generalizing the non-relativistic quantum mechanical Schrödinger equation of motion for particles, the wave nature of a free spin-$\frac{1}{2}$ particle is described by the relativistic Dirac equation:

$$(i\gamma^\mu \partial_\mu - m)\,\psi = 0 \tag{2.1}$$

where γ^μ are the 4×4 gamma matrices, fulfilling the relation $\{\gamma^\mu, \gamma^\nu\} = g^{\mu\nu}\cdot\mathbb{1}$, with $g^{\mu\nu}$ representing the metric tensor. The derivative of all four space-time coordinates is denoted by ∂_μ and acts on

the four-element Dirac spinor $\psi = \psi(x)$ of a fermion with mass m. Assuming a plane-wave solution, the Dirac equations leads to the relativistic energy-momentum relation:

$$E^2 = \mathbf{p}^2 + m^2. \tag{2.2}$$

This equation exhibits two solutions when solved for the energy, a positive one describing matter states and a negative one describing antimatter states. Under the matter-antimatter transformation all additive quantum numbers, e.g. charge, spin, lepton/baryon number, are inverted, whereas the particle's mass remains unchanged.

Elementary bosons are the force carriers mediating interactions described in Sec. 2.2 and Sec. 2.4. Analogous to Eq. (2.1), the Schrödinger equation can be extended for relativistic free particles with integer spin. For the case of scalar spin-0 bosons one finds the Klein-Gordon equation to hold:

$$(\partial^\mu \partial_\mu + m^2)\, \phi = 0 \tag{2.3}$$

which is of second order in time. Here the wave function $\phi = \phi(x)$ represents a scalar field. Once again two solutions can be obtained, representing an incoming and outgoing wave function interpreted as matter and antimatter fields, respectively. Gauge or vector bosons with spin-1 follow the Proca equation:

$$\partial_\mu \left(\partial^\mu A^\nu - \partial^\nu A^\mu\right) + m^2 A^\nu = 0 \tag{2.4}$$

where $A_\mu = A_\mu(x)$ denotes a vector field. For a massless vector field ($m = 0$) the Proca equation reduces to Maxwell's equations of electrodynamics in vacuum.

2.2 Quantum Electrodynamics

As already indicated in this chapter's introduction, the SM is represented by a set of relativistic quantum gauge field equations. Conservation laws are addressed by requiring fields $\phi = \phi(x)$ to be invariant under a D dimensional group of local gauge transformations U:

$$\phi \to \phi' = U(\theta^a) \cdot \phi = e^{i\, g\, \theta^a\, T^a} \cdot \phi \qquad\qquad (a = 1, 2, ..., D) \tag{2.5}$$

with $\theta^a = \theta^a(x)$ representing numerical parameters of the local transformations. Moreover, g denotes the coupling constant and the T^a are the symmetry group's generators. In the SM these generators are realized as complex hermitian matrices, obeying the commutator relations:

$$[T^a, T^b] = i\, C_{ABC}\, T^c \tag{2.6}$$

where C_{ABC} are the symmetry group's structure constants. In the case of **q**uantum **e**lectro**d**ynamics (QED) the underlying symmetry group U of Eq. (2.5) can be expressed by a unitary group $U(1)_{\mathrm{QED}}$, i.e. an abelian group of 1×1 matrices, i.e., complex numbers.

Within this framework interactions are expressed by Lagrangian densities $\mathscr{L}$ determined through a given gauge group and the conditions of Lorentz invariance and renormalizablity. The Lagrangian of a free spin-$\frac{1}{2}$ particle can be constructed in accordance with the Dirac equation Eq. (2.1) as:

$$\mathscr{L} = \bar{\psi}\, (i\gamma^\mu \partial_\mu - m)\, \psi. \tag{2.7}$$

Although this particular $\mathscr{L}$ is invariant under global gauge transformations, where θ^a from Eq. (2.5) is space-time independent, it is not invariant under the local gauge transformation with $\theta^a = \theta^a(x)$. To ensure such an invariance an additional vector field A_μ must be included, transforming as:

$$A_\mu \to A'_\mu = A_\mu + \frac{1}{g}\, \partial_\mu\, \theta^a, \tag{2.8}$$

along with a covariant derivative:

$$D_\mu = \partial_\mu + i\, g\, A_\mu \tag{2.9}$$

introducing an interaction between the fermion and the field.

The vector field A_μ can be identified as the neutral photon field coupling with $\alpha = \frac{g^2}{\hbar c} = \frac{e^2}{\hbar c}$ to any charged particle, in this case an electron. A kinetic term for the gauge field is added to ensure its propagation, rendering the complete Lagrangian of quantum electrodynamics as:

$$\mathscr{L}_{\text{QED}} = \bar{\psi}\,(i\gamma^\mu D_\mu - m)\,\psi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \tag{2.10}$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the four dimensional electromagnetic field strength tensor. The photon does not couple to neutral particles without loops of charged particles, and consequently self-coupling does not occur, leading to a potential inversely proportional to the distance of the interacting particles. Photons are taken to be massless in the SM, an additional mass term for the photon would break the gauge symmetry. Astrophysical measurements prove to be consistent with this assumption, setting an upper limit of $m_\gamma < 10^{-18}$ eV [27].

2.3 Quantum Flavordynamics and Electroweak Unification

2.3.1 The Gauge Sector

The weak force differentiates between fermion chiralities, i.e. $W^\pm$ bosons only couple to left-handed fermions and right-handed antifermions. The special non-abelian unitary group $SU(2)_L$ is the simplest group that can account for this behavior. The label L implies that only left-handed fermions, combined in weak isospin doublets ψ_L, participate in the interaction. Right-handed fermions, i.e. quarks and charged leptons, are assigned to isospin singlets ψ_R. The $SU(2)_L$ is constructed by three generators $T^a (a = 1, 2, 3)$, which are linked to the gauge bosons and can be represented by the 2×2 Pauli matrices.

The Lagrangian of a weakly interacting fermion can be given as:

$$\mathscr{L}_{SU(2)_L} = i\,\bar{\psi}_L\,\gamma^\mu\,D_\mu\,\psi_L + i\,\bar{\psi}_R\,\gamma^\mu\,\partial_\mu\,\psi_R - \frac{1}{2}\,\mathrm{Tr}(W^a_{\mu\nu} W^{a\,\mu\nu}) \tag{2.11}$$

with the corresponding covariant derivative D_μ of the gauge fields $W^a_\mu (a = 1, 2, 3)$ defined analogous to Eq. (2.9). In comparison to the abelian case, an additional term is present in the field strength tensor:

$$W^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + i\,g\,[W^a_\mu, W^a_\nu]. \tag{2.12}$$

Though the $SU(2)_L$ Lagrangian reproduces the charged current processes of the weak interaction, i.e. the exchange of $W^\pm$ bosons, the neutral current processes included in the first term of Eq. (2.11) are not described in full. The third generator corresponds to a neutral boson coupling only to the left handed fermions, thus cannot be identified as the photon of QED or any other known particle.

To consistently include the electromagnetic interaction the symmetry group is extended by a unitary group to $SU(2)_L \times U(1)_Y$. The weak hypercharge $Y = 2(q - T_3)$, combines the electromagnetic charge q and the third component of the weak isospin T_3. The $U(1)_Y$ symmetry introduces a new field B_μ, gauge coupling g' and field strength tensor $B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$, resulting in the following Lagrangian:

$$\begin{aligned}\mathscr{L}_{SU(2)_L \times U(1)_Y} = {} & i\,\bar{\psi}_L\,\gamma^\mu\big(\partial_\mu + i\,g'\,\frac{Y}{2}\,B_\mu + i\,g\,W^a_\mu\big)\psi_L + i\,\bar{\psi}_R\,\gamma^\mu\big(\partial_\mu + i\,g'\,\frac{Y}{2}\,B_\mu\big)\psi_R \\ & - \frac{1}{2}\,\mathrm{Tr}(W^a_{\mu\nu} W^{a\,\mu\nu}) - \frac{1}{4}\,B_{\mu\nu} B^{\mu\nu}\end{aligned} \tag{2.13}$$

unifying the electromagnetic and weak interactions, as initially shown by Glashow [28], Weinberg [29] and Salam [30] (GWS). The first two terms of $\mathscr{L}_{SU(2)_L \times U(1)_Y}$ correspond to the fermion's kinetic energy and coupling to the electroweak fields in charged and neutral current processes. The last two terms represent a kinematic term and a self-coupling of the gauge fields W_μ^a and B_μ. Physical states are given by linear combinations constituting four vector bosons:

$$\begin{aligned} W^\pm &= \frac{1}{\sqrt{2}} \left(W_\mu^1 \mp i\, W_\mu^2 \right) \\ Z^0/\gamma &= \frac{1}{\sqrt{g^2 + g'^2}} \left(g\, W_\mu^3 \mp g'\, B_\mu \right). \end{aligned} \tag{2.14}$$

In accordance with the experimental findings, both the photon and Z^0 boson couple to left and right-handed fermions[1]. The $W^\pm$ bosons couple only to left-handed fermions and change their flavor. Within the GWS gauge theory all mediating vector bosons are expected to be massless, i.e. the mass term given in the Proca equation Eq. (2.4) is not invariant under local gauge transformations. However, as already stated in Tab. 2.2, observation shows that $W^\pm$ and Z^0 bosons actually have a rather large mass. A possible resolution of this disparity is summarized in Sec. 2.3.2.

The range of the weak force is inversely proportional to the relatively large mass of the $W^\pm$ bosons and consequently only around $2.5 \cdot 10^{-18}$ m [10]. Furthermore, this leads to rather long lifetimes of particles restricted to weak decays. At small distances, i.e. high energies, the strength of the weak interaction converges to that of the electromagnetic interaction. At low energies the electroweak interaction is completely dominated by QED effects.

2.3.2 The Higgs Sector

The presence of massive gauge bosons, such as the $W^\pm$ and Z^0 of the electroweak interaction, indicates a spontaneous breaking of the gauge symmetry. Such a spontaneous breaking can be realized by introducing a two-component complex scalar field ϕ, forcing the system into a specific non-vanishing ground state or vacuum expectation value v. This approach is referred to as the Brout-Englert-Higgs mechanism [31–33]. The four degrees of freedom introduced by ϕ give rise to masses of three weak gauge bosons and an additional massive spin-0 particle, the Higgs boson H^0.

The Lagrangian of electroweak theory Eq. (2.13) is extended by terms describing the coupling of the weak bosons to the Higgs field and also the Higgs field self-coupling. Furthermore, Yukawa couplings y are introduced, giving masses $m_f = y \cdot v$ to the fermion fields f. The bosonic couplings to the Higgs field are proportional to the square of the corresponding boson mass, whereas the fermionic couplings depend linearly on the fermion masses.

$$\begin{aligned} \mathscr{L}_{\text{Higgs}} =& \frac{1}{2}\, \partial_\mu H\, \partial^\mu H + \left(1 + \frac{1}{v}\frac{H}{\sqrt{2}}\right)^2 m_W^2\, W_\mu^+\, W^{\mu -} + \left(1 + \frac{1}{v}\frac{H}{\sqrt{2}}\right)^2 m_Z^2\, Z_\mu\, Z^\mu \\ & - \frac{\mu^2 v^2}{2} + \mu^2 \left(\frac{H}{\sqrt{2}}\right)^2 + 2\frac{\mu^2}{v}\left(\frac{H}{\sqrt{2}}\right)^3 + \frac{\mu^2}{2v^2}\left(\frac{H}{\sqrt{2}}\right)^4 \\ & - \left(1 + \frac{1}{v}\frac{H}{\sqrt{2}}\right) m_f\, \bar{f}\, f \end{aligned} \tag{2.15}$$

Here μ represents a constant of the Higgs potential $V(\phi)$, which can be related to the Higgs boson mass $m_{H^0} \equiv \sqrt{2\mu^2}$. Using parameters determined by the electroweak theory, the vacuum expectation value can be calculated to $v = \sqrt{\mu^2/(2\lambda)} = (\sqrt{2} \cdot G_\text{F})^{-\frac{1}{2}} = 246.22$ GeV with the Fermi constant G_F. The Higgs boson mass is left as a free parameter of the theory and must be ascertained experimentally, like all other masses with exception of the $m_{W^\pm}/m_{Z^0}$ ratio.

[1] Note that in the neutrino sector only couplings to left-handed particles and right-handed antiparticles are realized.

Following decades of collider searches, only increasing exclusion limits on the Higgs boson mass, the ATLAS and CMS Collaborations of the **L**arge **H**adron **C**ollider (LHC) were able to announce the discovery of a new boson in 2012 based on data taken at center of mass energies of 7 TeV and 8 TeV [34–36]. Detailed studies in various decay channels reveal the new boson's properties to correspond well with those of a sought for SM Higgs boson with a mass of $m_{H^0} = 125$ GeV [9, 17].

2.4 Quantum Chromodynamics

The discovery of the Δ^{++}-resonance, a particle composed of three up quarks, necessitated an additional quantum number for particles interacting under the strong force. All three up quarks of the Δ^{++} have spin $s = +\frac{1}{2}$, apparently violating the Pauli principle for fermions [37]. A consistent method of describing this and other composite quark states is by introducing a so-called *color* quantum number, which can be either red, green or blue. In accordance to this naming scheme, the field theory describing strong interactions is known as **q**uantum **c**hromo**d**ynamics (QCD). Each quark is assigned one color or anti-color, whereas gluons, the gauge bosons of color symmetry, hold both a color and and anti-color, enabling gluons to couple to quarks but also to each other.

As required by observation, colorless bound quark states, referred to as hadrons, are composed of either a quark and an antiquark or of three quarks and named mesons and baryons, respectively. Recently, evidence for additional bound states involving four quarks [38], called tetraquarks, and five quarks [39], the pentaquarks, was reported by the LHCb experiment.

Properties of QCD can be described in the language of gauge fields using a $SU(3)_C$ symmetry group, where a quark q_j is described by a Dirac spinor and a three-element vector representing its color charge. The eight color-changing generators represent the spin-1 gluons. Relevant terms for quarks and gluons of the QCD Lagrangian can be written as:

$$\mathscr{L}_{\mathrm{QCD}} = \sum_{j=1}^{n_f} \bar{q}_j (i\,\gamma^\mu\, D_\mu - m_j)\, q_j - \frac{1}{4} \sum_{A=1}^{8} F^{A\,\mu\nu} F^A_{\mu\nu} \tag{2.16}$$

where n_f is the number of quark flavors. The covariant derivative D_μ is once again given in analogy to Eq. (2.9), whereas the field strength tensor is given as:

$$F^A_{\mu\nu} = \partial_\mu\, G^A_\nu - \partial_\nu\, G^A_\mu - g_s\, C_{ABC}\, G^B_\mu\, G^C_\nu \tag{2.17}$$

with G^A_μ as the $A = 1, .., 8$ gluon fields and g_s representing the coupling constant of quarks to gluons. No mass term is included for the gauge bosons of the gluon field. This assumption is supported by arguments based on quark confinement revealing $m_g < 1.3$ MeV and cosmological reasoning providing an upper limit of $m_g < 2 \cdot 10^{-10}$ MeV [40].

The effective coupling measurable on experimental level $\alpha_s = \frac{1}{4\pi}\, g_s^2$ depends on the square of the momentum transferred during the interaction. At small distances r, i.e. large momenta, the QCD potential is proportional to $\alpha_s(r) \cdot \frac{1}{r}$. Consequently, on the scale of nuclei the strong force is large enough to over-compensate repulsive electromagnetic interactions of like-signed charged particles. At even smaller distances quarks and gluons behave as unbound particles in a state referred to as asymptotic freedom. Resulting from the gluon self-coupling, the potential grows linearly with r at large distances, increasing the energy stored in the gluon field. Once enough energy has been accumulated, secondary pairs of colored particles are created from the vacuum. This behavior is summarized by the term confinement and is the reason that quarks and gluons cannot be observed as isolated particles, but only in colorless bound states, the hadrons. Accordingly, the quark masses stated in Tab. 2.1 cannot be measured directly, but are estimated from mass-dependent hadronic effects relying on QCD calculations [16].

Chapter 3
Beyond the Standard Model

The SM, as described in Ch. 2, has proven to be a successful and predictive theory. Especially in the QED sector, theoretical calculations of, e.g., the anomalous magnetic moment of the electron and muon, corresponding to quantum mechanical loop corrections of their magnetic dipole moment, are given up to thirteen significant figures for the electron [41] and nine [42] for the muon anomaly. Moreover, these are among the most precisely measured quantities in physics, given at an experimental accuracy of twelve [43] and nine [44] significant figures.

Furthermore, precision tests of the SM electroweak sector over a wide range of observables including particles masses, coupling constants, mixing angles, branching ratios as well as various total and differential cross sections give a conclusive picture of the SM throughout different energy regimes. Results are obtained by comparing precise measurements from lepton-hadron and hadron-hadron colliders [45] as well as neutrino deep inelastic scattering [46] and atom spectra [47] with estimates from SM theory, revealing good agreement up to a scale of around 200 GeV.

Moreover, the SM theory has the invaluable trait of being able to predict the existence of specific particles and deliver estimates of their properties, e.g., mass ratios and branching ratios. A fourth quark, the charm quark, was hypothesized within the theory of weak interaction [48]. Four years later signatures of a charmed hadron, the J/Ψ, were observed [49, 50]. Similarly the gluon and a new strange baryon were predicted in describing the strong force [51]. The Ω^- baryon was found after two years, whereas the gluon was discovered 16 years later in three-jet events [52]. The list of achievements from SM predictions and subsequent discoveries continues with W and Z bosons [28–30] found at the **S**uper **P**roton **S**ynchrotron (SPS) collider [53, 54], the top quark [55] discovery at the Tevatron collider [56, 57], and finally the Higgs boson at the LHC, as recalled in Sec. 2.3.2.

Albeit these and other accomplishments, the SM alone cannot represent the complete description of physics on the sub-atomic level. The following paragraphs highlight specific shortcomings, indicating the necessity of extensions beyond the SM. A qualitative introduction and motivation for BSM physics is given in [10, 11, 58, 59] on which this chapter is based.

3.1 Experimental Indications

Given by the most recent interpretation of the cosmic microwave background[1], measured with the Planck satellite, the universe is composed of 4.8% baryonic matter, 25.9% dark matter and 69.3% dark energy [62]. The SM currently only describes the small amount of ordinary baryonic matter.

The majority of the universe's energy density is associated with an unknown source called dark energy, which is assumed to permeate all space uniformly. Supernova surveys performed in the late 1990s [63, 64] revealed the accelerated expansion of the universe, for which dark energy is found to be a consistent source. In theory, this phenomenon can be associated with a cosmological constant Λ, as introduced in field equations of general relativity [65].

A first hint for the existence of dark matter was observed by Zwicky [66] in clusters of galaxies and later confirmed by Rubin in the rotation curves of galaxies [67, 68]. According to classical gravitation, the velocity should decrease with the square root of the distance r between a star and the galaxy center. Instead the measurement shows a steady increase, which can be explained through the presence of additional gravitationally interacting matter, collected in a halo around

[1] Previous analyses of the cosmic microwave background were performed with data collected by the **W**ilkinson **M**icrowave **A**nisotropy **P**robe (WMAP) satellite [60, 61].

the luminescent parts of a galaxy. Further evidence of dark matter has since been observed in, e.g., gravitational lensing effects and structures of the cosmic microwave background. Both collider and astroparticle experiments are on the look-out for suitable weakly interacting massive particles to constitute dark matter.

Even within the 4.8% of the observable universe the SM cannot sufficiently describe all measured phenomena. Deviations between theory and experiment can be found in precision measurements of the SM. While the calculations and measurements of the electron anomalous magnetic moment agree up to an accuracy of nine significant figures, the muon anomalous magnetic moment shows a disagreement with the SM prediction at the sixth significant figure corresponding to 3.4 Gaussian standard deviations [42]. The muon anomaly is given at a lower precision, but it is $(m_\mu/m_e)^2 \approx 4 \cdot 10^4$ times more sensitive to contributions of heavy virtual particles [10, 69]. The new experiment Muon (g-2) is underway to deliver a more detailed measurement and determine if the observed discrepancy is indeed an indication of physics beyond the SM [69].

As described in Sec. 2.1, neutrinos in the SM are massless, an assumption contradicted by the observed neutrino oscillations. Every extension of the SM, aiming at including massive neutrinos, involves introducing new degrees of freedom.

Moreover, the universe observed today is filled with stable matter particles building the atoms of everyday life, but no antimatter. Within the SM no mechanism is known to cause such a large asymmetry. Although the charge parity violation observed in the quark sector of the weak interaction in decay of K^0 mesons delivers evidence for processes where matter and antimatter are treated differently [70], the effect is not large enough to account for the measured baryon asymmetry of the universe.

3.2 Theoretical Indications

At energies close to the Planck scale $\mathcal{O}(10^{19}$ GeV) loop corrections of the Higgs boson mass become very large, being proportional to the square of the scale [71]. An extreme *fine-tuning* is necessary to retain the measured Higgs boson mass of 125 GeV. In other words, this so-called *hierarchy problem*, shows that the SM is not likely to be valid up to the Planck scale.

Since a unification of the electromagnetic and weak force is already in place it appears compelling to also include the strong force. Such a gauge coupling unification is not given in the SM, but can be realized in **G**rand **U**nified **T**heory (GUT) or supersymmetry scenarios. As already mentioned in Ch. 2, the gravitational interaction is currently not consistent with the SM. A true *theory of everything* can only be achieved if general relativity is brought together with quantum field theory.

The list of experimental and theoretical motivations summarized here is not exhaustive and many BSM theories have been developed to tackle them. One example is supersymmetry, a new symmetry between fermions and bosons, which doubles the number of degrees of freedom and offers a dark matter candidate, mitigates the hierarchy problem, and unifies all three couplings well below the Planck scale [71]. The concept of quantum gravity is taken up in various model frameworks introducing extra dimensions [59] in which only gravity can propagate, also explaining the relative smallness of its coupling strength. In current understanding the SM represents an effective low energy manifestation of a underlying theory extending to the Planck scale. Until a conclusive description of all unexplained phenomena is found, it is worthwhile to search for new particles or effects that could lead to promising extensions of the SM.

Part II

Experimental Setup

Chapter 4
The Large Hadron Collider

Located at the CERN research center by Geneva, Switzerland, the LHC is a 26.7 km long storage ring in which protons or lead nuclei are brought to collision at relativistic energies. The following chapter gives a concise overview of the LHC machine, summarized predominantly from [72]. Specific values associated to LHC operations are based on 2012 run conditions as given by [73, 74].

Conceptualized to supply a sufficient amount of data to clarify the existence of the Higgs boson on a reasonable time scale, a rather high luminosity goal was set for the LHC, favoring a particle-particle over a particle-antiparticle collider. The confined space within the former **L**arge **E**lectron-**P**ositron (LEP) collider tunnel is addressed by means of a twin-bore magnet design. Two separate beam pipes are necessary to deflect counter-rotating particle beams and each, with its corresponding vacuum system, is enclosed in a set of NbTi superconducting dipole magnet coils, providing a field of up to 8.3 T, used to bend the particles onto the LHC path. Each of the 1232 dipoles contains a cooling system based on superfluid helium to reduce the magnets' temperature to below 2 K. Furthermore, 392 quadrupole magnets are used to focus the particles in each of the beam's bunches. Although far less pronounced than for lighter particles such as electrons, synchrotron radiation causes the proton beam to loose energy over time, accompanied by heating of the surrounding machine components. Conveniently, this effect can be exploited to focus the beam both transversally and longitudinally during acceleration of the beam [75].

A chain of accelerators, shown schematically in Fig. 4.1a, is employed to gradually increase the particle's energy. Protons are obtained by ionizing hydrogen atoms and groups of protons, referred to as *bunches*, are inserted into the **Lin**ear **Ac**celerator 2 (Linac2), accelerating them to 50 GeV. The bunches are passed on to the **P**roton **S**ynchrotron **B**ooster (PSB), the **P**roton **S**ynchrotron (PS) and the SPS, collimating the bunches to a length of 1.35 ns and raising their energy to 450 GeV. Finally, reaching a bunch spacing of 50 ns, they are injected into the LHC itself where they are accelerated to their final energy and brought to collision.

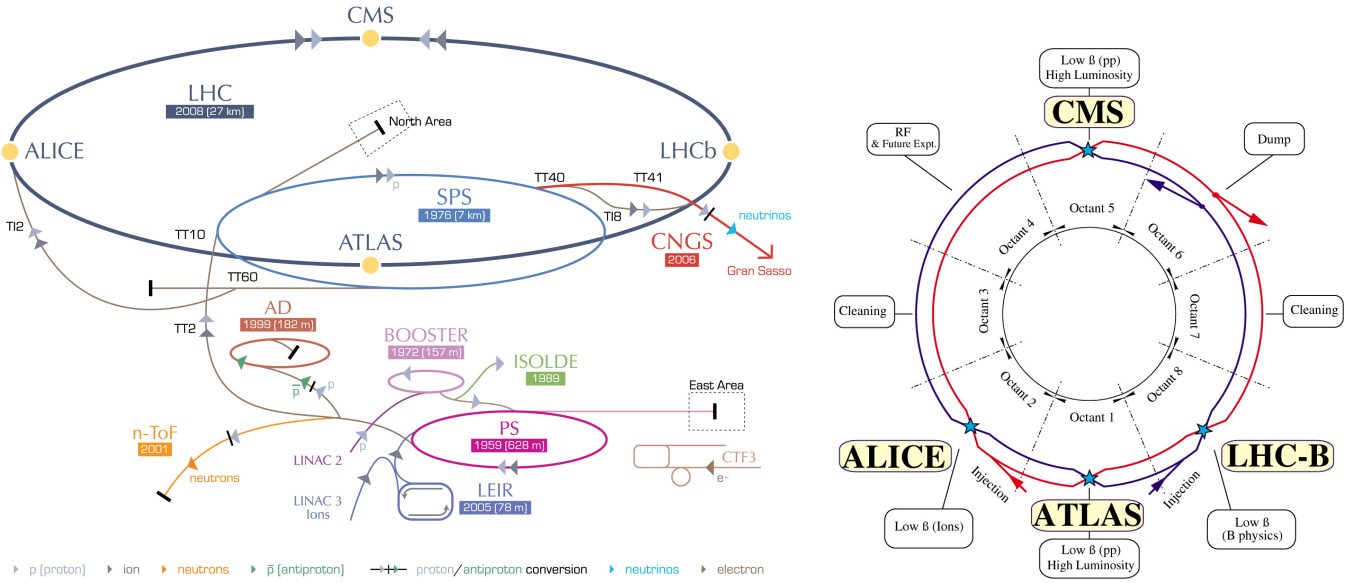
The LHC experimental layout is divided into octants. As shown in Fig. 4.1b the beams are interchanged between the inner and outer track at each of the four interaction points, ensuring that all bunches travel the same distance overall. The two large multi-purpose experiments **A** **T**oroidal **L**HC **A**pparatu**S** (ATLAS) [78] and **C**ompact **M**uon **S**olenoid (CMS) [79] designed to carry out a broad range of measurements with proton and heavy ion collisions are placed in opposing octants at Point 1 and 5. This secures that the same bunches collide in both detectors, nominally delivering the same amount of integrated luminosity. Adjacent to ATLAS, octants 2 and 8 each accommodate a low instantaneous luminosity experiment, the **A** **L**arge **I**on **C**ollider **E**xperiment (ALICE) [80] detector, especially designed to study heavy ion collisions, and **LHC** **b**eauty (LHCb) [81] dedicated to investigating, in particular, properties of b-hadrons in proton-proton collisions.

The rate of events $\dot{N}_{\text{ev}}$ of a given process produced in particle collisions per unit of time depends on the process's productions cross section σ_{proc} at the corresponding center of mass energy $\sqrt{s}$ and on the instantaneous luminosity:

$$\dot{N}_{\text{ev}} = \sigma_{\text{proc}} \cdot \mathcal{L}. \tag{4.1}$$

The instantaneous luminosity $\mathcal{L}$ per interaction region is given by parameters of the collider [82]:

$$\mathcal{L} = \frac{N_p^2 \, n_b \, f_{\text{rev}}}{4\pi \, \sigma_x^* \, \sigma_y^*} \, F, \tag{4.2}$$



(a) **(b)**

Figure 4.1: CERN accelerator complex including the chain leading up to the LHC is shown in Fig. 4.1a as given in 2012 (modified from [76]). Fig. 4.1b gives an overview of the LHC beam layout including the injection, dump and the four main experiments (modified from [77]).

where simplifications, such as a Gaussian beam profile and uncorrelated particle densities in all three dimensions, are assumed. Apart from the revolution frequency f_{rev}, determined by the circumference of the collider ring to about 11 kHz, the parameters of Eq. (4.2) can be chosen to maximize the delivered instantaneous luminosity. A geometrical reduction factor F arises from the beam overlap resulting from the angle at which the beams are crossed. For the 2012 run, the horizontal and vertical beam size at the interaction point σ_x^* and σ_y^* were at about 19 µm in Point 1 and 5 [83]. Each of the beam's $n_b = 1374$ bunches contained around $N_p = 1.7 \cdot 10^{11}$ protons. Aspects of luminosity measurements are discussed in Sec. 5.3 for the CMS detector.

Over time the instantaneous luminosity within one fill will decrease due to beam losses from the collisions. During operation in 2012 a peak instantaneous luminosity of $7.7 \cdot 10^{33}$ cm^{-2} s^{-1} [74] was reached and the luminosity lifetime amounted to around 10 h [84]. The integrated luminosity $L = \int \mathcal{L}\, dt$ is a measure for the total number of events produced within a fixed timespan.

Run I of the LHC was comprised of three data-taking periods, two at a center of mass energy of $\sqrt{s} = 7$ TeV delivering around 45.0 pb^{-1} (2010) and 6.1 fb^{-1} (2011) of integrated luminosity, as well as one at 8 TeV yielding approximately 23.3 fb^{-1} (2012) [85]. During a first long shutdown, the injectors and LHC machine were upgraded and data-taking commenced for Run II in 2015 at 13 TeV [86]. Following a further long shutdown around 2019, the LHC is planned to run at its design center of mass energy of 14 TeV [87]. The machine will be upgraded to the so-called High Luminosity LHC, aiming to deliver around 300 fb^{-1} annually over the following ten years [88].

Chapter 5
The Compact Muon Solenoid Detector

Situated at Point 5 of the LHC, the CMS experiment is a 28.7 m long and 15.0 m wide multi-purpose particle physics detector. Though conceptualized with a focus on uncovering the nature of electroweak symmetry breaking, CMS is well suited to investigate the multitude of final states created in the collisions. As a result of the high instantaneous luminosity delivered by the LHC, several interactions take place during a single bunch crossing. This effect, referred to as **p**ile**u**p (PU) (for details see Sec. 8.2.4), necessitates a high detector granularity, enabling individual particle trajectories to be resolved clearly. Furthermore, an excellent time resolution is vital to be equipped for the short timespan between consecutive collisions. Constructed in a cylindrical shape concentrically around the interaction point, CMS consists of many specialized detector subsystems. The interplay of all components makes it possible to identify a wide variety of objects and diverse event topologies. A five wheel structure was chosen to facilitate construction, maintenance and upgrades. This so-called *barrel* region is complemented by three planar *endcap* segments on each side delivering a nearly complete solid angle coverage, a key element for describing events containing neutrinos or undetectable BSM particles. An overview of the detector and its principle components is shown in Fig. 5.1. In the following a concise summary of the main subdetectors as well as the reconstruction methods used to measure and identify the encountered physics objects are given.

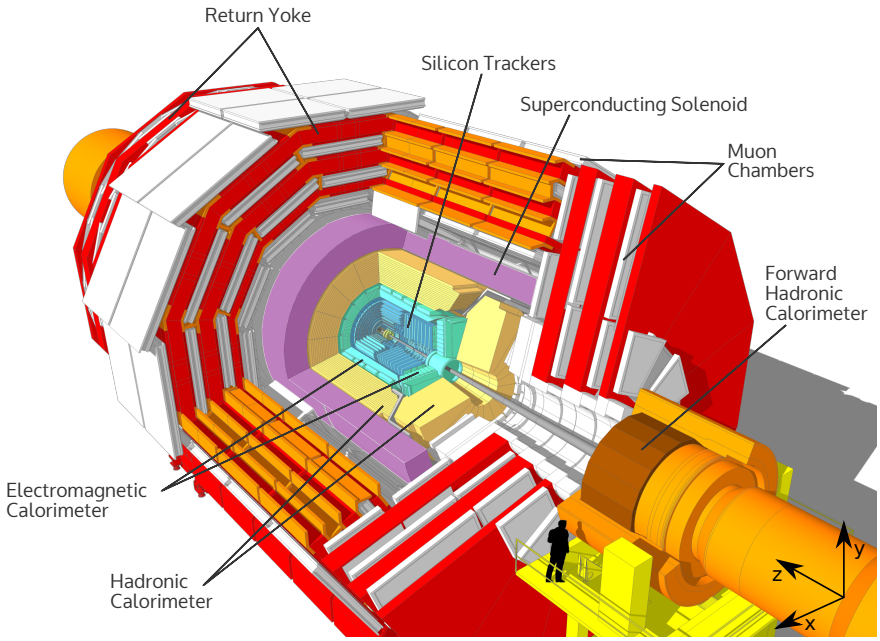


Figure 5.1: Overview of the CMS detector highlighting its main components, adapted from [89].

Centered at the nominal interaction point, the CMS coordinate system's Cartesian x-axis points inward towards the center of the LHC ring. The y-axis is oriented upwards and the z-axis is set along the beam line pointing from octant 5 to 4. Starting from the x-axis, the azimuthal angle ϕ is defined in the x-y plane, where the radial distance is denoted by r. Measured from the z-axis to a respective object, the polar angle θ strongly depends on an object's boost in beam direction. Therefore, the pseudorapidity is defined as $\eta = -\ln(\tan(\theta/2))$, in which differences $\Delta\eta$ for ultra-relativistic massless particles are Lorentz invariant under a z-boost. A value of $\eta = 0$ is given at

the y-axis and $|\eta| \to \infty$ in $\pm z$-direction. Lorentz invariant quantities are aspired, since the amount of energy given to each of the proton's constituents is not known a priori. Along these lines the transverse component of momentum vectors p_{T}, a projection onto the x-y plane, and the transverse energy E_{T} are commonly used to describe collision products.

5.1 Detector Components

Beginning from the innermost surrounding the LHC beam pipe, the subsequent sections give details on individual detector subsystems. Information is based mainly on [79], where additional references are given if applicable. At the heart of the CMS experiment, a finely segmented silicon tracking system is used to retrace the trajectories of charged particles. Also enclosed within the bore of the extraordinary 3.8 T solenoid magnet, calorimetry detectors are built to stop traversing particles and determine the amount of energy they deposit. Outermost, the four stations of the muon system are interlaced between the magnet return yoke, all within range of the magnet's large bending power.

5.1.1 Tracking System

Built in barrel and endcap layers around the beam pipe, the tracking system is the first active part of the instrumentation a particle passes through, and provides a robust and precise track and vertex assignment in the high particle flux environment given by the LHC instantaneous luminosity. Intense radiation and the strong magnetic field are set against requirements of high granularity and quick response time, making silicon based pn-junction detectors cooled to 4 °C [90] a viable solution. A schematic overview of the full CMS tracker is shown in Fig. 5.2.

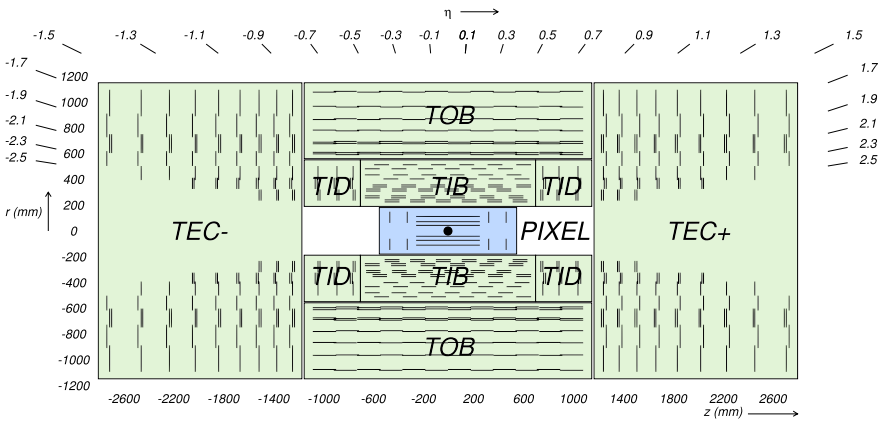


Figure 5.2: Vertical plan of the CMS tracking system in the r-z plane, adapted from [79]. The pixel tracker is underlayed in blue. Components of the strip tracker are shown in green and divided into four regions, **T**racker **I**nner **B**arrel (TIB) and **T**racker **I**nner **D**isk (TID), **T**racker **O**uter **B**arrel (TOB) and **T**racker **E**nd **C**ap (TEC).

Covering a pseudorapidity range of $|\eta| < 2.5$, the pixel detector consists of three cylindrical barrel modules accompanied by two endcap disks at both ends. The modules are equipped with around $66 \cdot 10^6$ individual 100 µm × 150 µm pixel cells, making up an active area of about 1 m^2. Forward detector elements are tilted to enhance the effect of interpolating over the Lorentz drift of free charge carriers produced by the traversing particle to increase the spacial resolution. Providing a fine segmentation in the r-ϕ and z directions, a precise three dimensional reconstruction of primary and secondary vertices is made possible.

The large number of read-out channels and the high cost associated with a pixel solution motivate a micro-strip layout for the consecutive tracker layers. Wafers with a thickness of 320 µm and strip

pitch of 80 µm to 141 µm are deployed in the four TIB and three TID layers, delivering a single point resolution of roughly 30 µm. The outer components of the tracker consists of six barrel layers TOB and nine endcap disks TEC on each side, as shown in Fig. 5.2. A reduced particle flux in this region allows for an increasingly wider strip pitch.

In all parts of the strip tracker, stereo modules, consisting of two sensors mounted together at an angle, are introduced to identify a particle's path along the strips. These are indicated by double lines in Fig. 5.2. In consequence, this setup secures that at least nine hits are given by each particle traversing the entire tracker within $|\eta| < 2.4$, of which more than three hits are measured in two dimensions simultaneously. The momentum resolution of isolated high p_T tracks reconstructed with combined information from pixel and strip detectors can be parameterized as [91]:

$$\frac{\Delta p_\mathrm{T}}{p_\mathrm{T}} = \left(c(\eta) \cdot \frac{p_\mathrm{T}}{\mathrm{TeV}}\right) \oplus 0.5\%, \tag{5.1}$$

where $c(\eta) \approx 15\%$ in the central and around 60% in the forward regions. In this notation $\oplus$ refers to the quadratic summation of uncorrelated properties in a factorized approach.

5.1.2 Electromagnetic Calorimeter

Situated hermetically around the outer tracker layer, the **E**lectromagnetic **Cal**orimeter (ECal) is made of over $68 \cdot 10^3$ lead tungstate $PbWO_4$ crystals deployed in barrel modules and endcap Dees, as shown in Fig. 5.3, reaching a coverage of $|\eta| < 3.0$. An advantage of this inorganic scintillator its small Molière radius, allowing around 25 radiation lengths within the confined space inside the solenoid's bore. Furthermore, its fast scintillation decay time and relatively strong radiation hardness make it the ideal material for propagating electromagnetic showers induced by the primary particle. Thermal effects of the crystals are minimized by keeping the entire ECal at a temperature of (18.00 ± 0.05) °C. A laser monitoring system is installed to continuously measure the optical transmission of the ECal crystals which grow less transparent over time because of radiation.

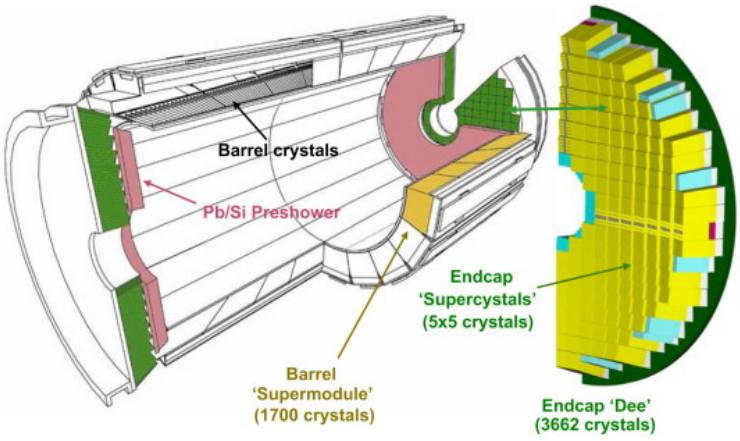


Figure 5.3: Sketch of the crystal layout in the CMS ECal, taken from [92].

The produced scintillation light is amplified and registered via two Avalance Photo Diodes mounted on each crystal in the barrel region and a single Vacuum Photo Triode for each endcap crystal. Both types of photodetector are well suited for operation under the strong magnetic field and ensure a rapid read-out matching the LHC bunch crossing time.

An additional preshower absorber is installed in front of both ECal endcaps. Consisting of a layer of lead radiators and subsequent silicon strip sensors, this sampling calorimeter was introduced to

improve the positional reconstruction of electrons as well as photons and aids in disambiguating neutral pions and genuine high p_T photons.

Resulting from test beam measurements with electrons up to 250 GeV, the energy resolution of in the barrel is determined to [79]:

$$\frac{\Delta E}{E} = \left(\frac{2.8\%}{\sqrt{E/\mathrm{GeV}}}\right) \oplus \left(\frac{0.12}{E/\mathrm{GeV}}\right) \oplus 0.30\%. \tag{5.2}$$

Statistical fluctuations of the photodetectors as well as stochastic effects of the lateral shower containment and the difference of energy deposited and measured in the preshower detector are summarized in the first term. Contributions of electronics, digitization as well as PU noise are given in the second term. A constant corrects for intercalibration errors as well as non-uniform light collection, which is already reduced by depolishing one lateral side of each barrel crystal, mitigating the effect of non-parallel crystal faces.

5.1.3 Hadronic Calorimeter

Although also depositing energy in the ECal, hadrons will often not be absorbed completely therein. The **H**adronic **Cal**orimeter (HCal), as shown in Fig. 5.4, is designed to measure the full extent of hadronic showers. Components placed between the ECal and the magnet coil, **H**adron **B**arrel (HB) and **H**adron **E**ndcap (HE), are completed by an **H**adron **O**uter (HO) detector segment outside of the solenoid. The combined material, including ECal, HB and HO as well as the iron return yoke, amounts to an absorber thickness of around 10 to 15 nuclear interaction lengths (depending on $|\eta|$) in the barrel region [93].

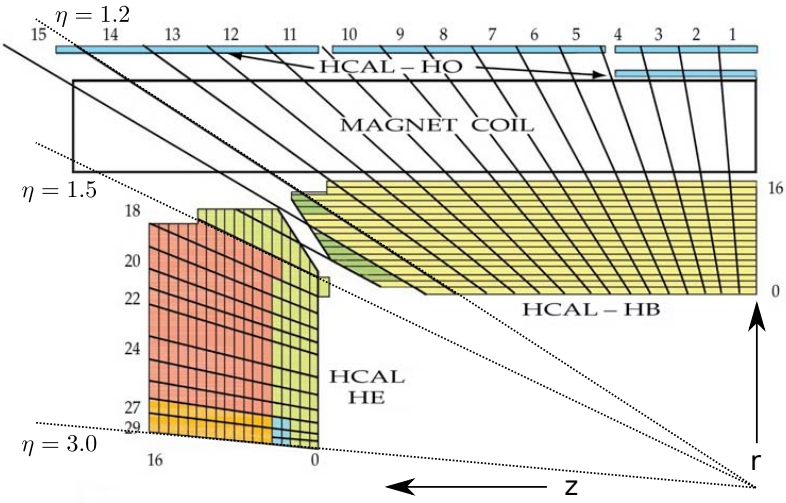


Figure 5.4: View of the CMS HCal with details of the tower segmentation, adapted from [79].

All three HCal subsystems rely on a sampling configuration alternating brass absorbers and plastic scintillators encased in an inner and outer housing of steel for stability. Matching the ECal and muon system granularity, a segmentation of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ is chosen in the barrel, resulting in 16 wedges in η direction, the so-called HCal towers. The over $70 \cdot 10^3$ scintillator tiles are grouped into trays along ϕ and installed between 14 brass absorber plates. Each tile is equipped with a wavelength shifting fiber and all fibers of one tray are fed to an **H**ybrid **P**hoto **D**iode (HPD) with 19 hexagonal pixels. Following the same basic detection principles, the HE seals off the HB almost hermetically, ranging from $1.3 < |\eta| < 3$, and is still inside the barrel of the solenoid magnet.

Placed outside of the last magnet return yoke segment on either side of the barrel, as can be seen in Fig. 5.1, the **H**adron **F**orward (HF) calorimeters extend the geometrical coverage to $|\eta| < 5.2$ for measuring the missing transverse energy of an event. Extreme particle fluxes in the forward region, on average seven times higher than found at smaller η ranges, necessitate a truly radiation hard technology. Based on Cherenkov light emission, over 1000 km of quartz fibers are embedded in a steel absorber structure. Read-out is performed with **P**hoto **M**ultiplier **T**ubes (PMTs) housed in separate boxes, shielded from the radiation by steel, lead and polyethlene.

Combining all calorimetry elements, test beam studies give an energy resolution of [94]:

$$\frac{\Delta E}{E} = \left(\frac{100\%}{\sqrt{E/\mathrm{GeV}}} \right) \oplus 4.5\% \tag{5.3}$$

for jets of 30 GeV to 1 TeV summarized over all measured η and ϕ values. The full calorimeter setup improves the resolution over values measured with pions exclusively for the joined ECal and HCal barrel region [95].

5.1.4 Magnet System

Bending of their trajectories within a magnetic field is used to determine the sign and momentum of charged particles. Reaching field strengths of 3.8 T in the central homogeneous region, the CMS solenoid provides an ideal setting for precise measurements and is the single most important element for the CMS muon reconstruction. The magnet's monolithic bore is constructed of four layers of reinforced NbTi coils cooled down to 4.5 K and spans 12.5 m in length with a diameter of 5.9 m [96]. A rather homogeneous field is given inside of the solenoid. Outside, the $10 \cdot 10^3$ t flux-return yoke helps mitigate inhomogeneities and values of around 2 T are reached [97]. Divided into five barrel wheels and six endcap disks, the iron yoke also doubles as an ideal support structure for the other detector components, which can be slid apart during periods of maintenance and upgrade.

5.1.5 Muon System

To ensure a good resolution within the full interval of $|\eta| < 2.4$ and over a wide muon momentum range, three types of detectors are installed in the shielding material of the flux-return yoke comprising the four stations of the CMS muon system. The barrel region is instrumented with **D**rift **T**ubes (DTs), while **C**athode **S**trip **C**hambers (CSCs) are better suited for magnet field inhomogeneities and increased radiation in the endcaps. In addition **R**esistive **P**late **C**hambers (RPCs) are interspersed throughout. All three chamber types rely on the principle of gas ionization and corresponding detectors come at a moderate cost, an important aspect to consider for a subsystem covering an area of around $25 \cdot 10^3$ m^2. Fig. 5.5 gives an overview of the CMS detector, highlighting the muon chamber placement.

A magnetic flux of only around 0.4 T is present throughout the barrel region [98], permitting the use of DT chambers comprised of drift cells with sophisticatedly shaped electrical fields. Formed between the four electrodes on the drift cell's walls and the anode wire in the center of the cell, the field guaranties a linear behavior over the drift volume, with a maximum path of 21 mm. A gas mixture of 85% Ar and 15% CO_2 warrants a drift velocity of about 54 µm/ns and good quenching properties resulting in a maximum drift time of 380 ns [99], appropriate for the relatively low muon rate. Individual cells are arranged in a staggered fashion to a superlayer, allowing for absolute time measurements and an improved spacial resolution. In term two or three superlayers are combined at a 90° angle to form one DT chamber, allowing measurements of the ϕ and z coordinate. Four

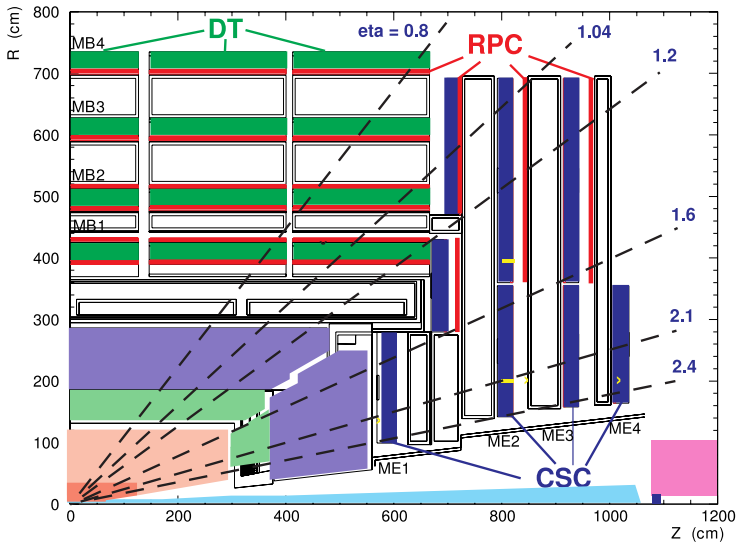


Figure 5.5: Layout of the CMS muon system in one quarter of the r-z plane, taken from [79].

layers of DTs are inserted in the barrel wheels, covering $|\eta| < 1.2$ with a spacial resolution in the r-ϕ plane of appropriately 80 µm to 120 µm per DT chamber [98].

Each endcap is instrumented with four layers of trapezoidal CSCs arranged in rings of 18 or 36 chambers perpendicular to the beam, covering $0.9 < |\eta| < 2.4$ and a contiguous range in ϕ. Consisting of six planes of anode wires and seven cathode panels, these multiwire proportional chambers are filled with a 40% Ar, 50% CO_2 and 10% CF_4 gas mixture. Cathodes are segmented into strips to determine the ϕ coordinate via charge interpolation over several adjacent strips. The short distance between individual wires yields a quick response time of 5 ns, enabling the association of a detected muon to the corresponding bunch crossing. Precision of the track measurement in the r-ϕ plane is given as 47 µm to 243 µm, depending on the specific ring [100].

Placed throughout both barrel and endcap regions up to $|\eta| < 1.6$[1], RPCs are included in the detector design to efficiently trigger events containing muons over a wide transverse momentum range with a rapid response time below 3 ns [98] required at collision rates up to 1 kHz/cm^2 [101]. Each RPC consists of a double gap structure of respectively two bakelite electrodes separated by a joined read-out strip and set in a gas tight volume flooded with a 96.2% $C_2H_2F_4$, 3.5% C_4H_{10} and 0.3% SF_6 mixture. The striped design allows a spacial measurement in ϕ direction.

As shown by simulation given in Fig. 5.6, both the tracker and muon station subsystem deliver independent measurements of the transverse momentum of tracks belonging to muons, enabling a redundancy. Especially for low p_T muons the tracker measurements give the better p_T resolution, as the muon system suffers from multiple scattering in the surrounding material. Combining the two measurements leads to an increased overall resolution which, considering muons above a transverse momentum of 100 GeV, can be written as [91]:

$$\frac{\Delta p_T}{p_T} = 4.5\% \cdot \frac{p_T}{\text{TeV}}. \tag{5.4}$$

Measurements based on meson [102] and Z boson [103] decays find the resolution to range from 1% to 6% over the full pseudorapidity range for muons below $p_T = 100$ GeV. Cosmic ray muons are used to determine a value of below 10% for muons up to 1 TeV, reconstructed in the central region of the detector [103].

[1] During LS1 the reach was extended by a fourth layer of endcap CSCs and RPCs to $|\eta| < 2.1$ [88].

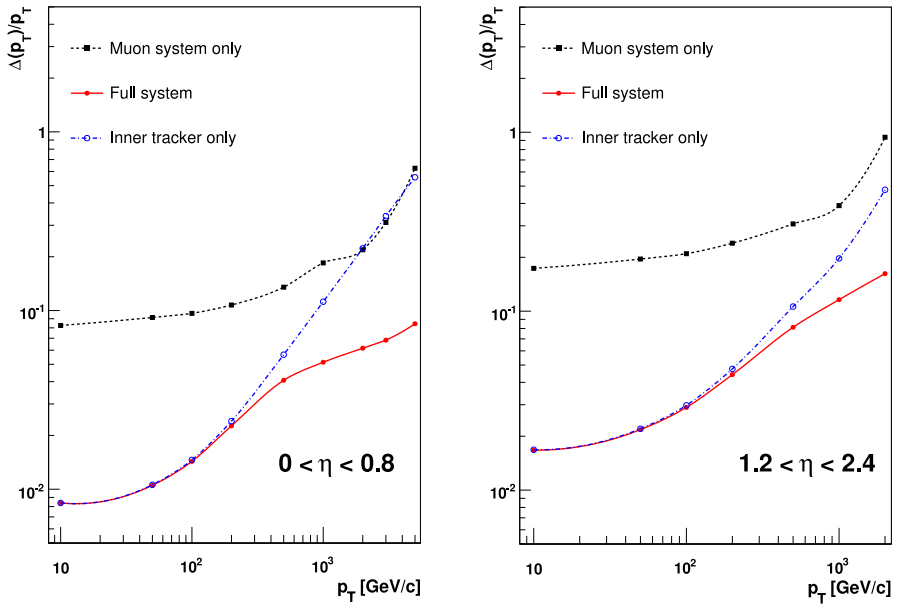


Figure 5.6: Simulation of muon transverse momentum resolution in two η ranges, taken from [79]. Individual curves are given for reconstruction using only the muon system or tracker, as well as their combination.

5.2 Trigger and Data Acquisition

The LHC collision rate of 16 MHz, on average during 2012, is several orders of magnitude higher than what can be processed and recorded and must be reduced to around 100 kHz [104]. A two stage setup is employed, consisting of a custom electronics based **L**evel-**1** (L1) trigger system and a software based **H**igh **L**evel **T**rigger (HLT) run on a commercial computing farm.

The L1 trigger [105] has the intricate task of supplying a rapid differentiation between the plentiful soft interactions and the much rarer hard interactions of heightened interest. Designed for the nominal bunch crossing rate of 25 ns, the L1 trigger relies on a reduced set of coarse detector information, originating solely from the calorimeters and muon system. Technological restraints given by the front-end buffers of tracker and ECal preshower dictate a maximum latency of 3.2 µs, corresponding to 128 bunch crossings. This tight time frame encapsulates reading out the data, forming an L1 trigger decision and propagating it back from an underground computing room, adjacent to the experimental cavern, back to the front-end electronics on the detector itself.

Once an L1 accept is issued, data fragments from $55 \cdot 10^6$ subsystem channels are bundled into 700 modules and assembled to form a complete event view by the CMS **D**ata **A**cquisition (DAQ) [106]. On average, a single bunch crossing results in 1 MB of data passed to one of the $13 \cdot 10^3$ **C**entral **P**rocessing **U**nits (CPUs) of the HLT at a rate of about one event per 10 µs. Since the computing farm is located on the surface at Point 5, sophisticated networking technology is needed, suited for a throughput of 800 Gbit/s and to ensure buffer time for the ∼1 s long HLT step.

Faster streamlined versions of the full reconstruction algorithms described in Sec. 5.4 are run to make the final decision of sending an event to mass storage. Individual HLT *paths* are implemented, each requiring different event compositions and topologies. Criteria of the HLT selection used in the presented analysis are covered in Sec. 9.1.3. To reduce the data output individual HLT paths can be assigned a *prescale* factor N, specifying that only every N^{th} event is recorded. During a fill of the LHC, the instantaneous luminosity will drop with time, allowing the prescale values to be reduced while exploiting the full capacity of the DAQ structure. In total, CMS writes roughly 460 events per second [107] to storage at the CERN **T**ier-**0** (T0) facility (see Sec. 5.5.1).

5.3 Luminosity Measurement

During online data-taking the HF calorimeter, which can be operated even if the LHC beams are not fully focused and stabilized, is used for a real-time measurement of the average luminosity per bunch which is proportional to the average transverse energy per HF tower [108]. However, the estimated value suffers from a non-linear detector response to PU conditions and systematic shifts in the used detection methods, and can only be seen as a preliminary result.

To increase the precision of the integrated luminosity value, an important ingredient for most physics analyses, a method based on counting clusters in the pixel tracker [109] is applied offline. Exploiting only information from pixel modules active during the entire data-taking, the high amount of pixels ensures a linear dependence between the number of pixel cluster hits and the actual number of interactions. Thus the instantaneous luminosity can be determined by:

$$\mathcal{L} = \frac{f_{\text{rev}} \cdot \langle n \rangle}{\sigma_{\text{vis}}} \tag{5.5}$$

with the given LHC revolution frequency f_{rev}, the average number of counted pixel clusters per event $\langle n \rangle$ and the so-called visible cross section $\sigma_{\text{vis}} = \sigma_{\text{tot}} \cdot n_1$, expressed via the total inelastic cross section and the average number of clusters per inelastic collision (compare also Eq. (4.2)).

The Van de Meer scan technique [110] is carried out in dedicated short data-taking periods to determine σ_{vis}. By displacing one proton beam against the other in the vertical and horizontal plane, profiles of the interaction rate are attained relative to the beam separation. Combining these with the accurate number of protons in each bunch and the average number of pixel clusters during the bunch crossing, given by the amplitude of the rate profile, results in the desired quantity.

5.4 Reconstruction

Signals from the CMS detector components are passed through a series of algorithms and combined to the objects used to interpret each collision event in terms of a particle physics final state. The following section outlines the offline reconstruction techniques relevant for objects considered in the analysis presented in this thesis. Further criteria are applied to the reconstructed objects to ensure their quality, and are discussed in Sec. 9.2 with relevance to this specific analysis.

A high absolute measurement accuracy can only be achieved if backed by rigorous calibration and alignment of each subdetector system, which is coordinated by the CMS **Al**ignment and **Ca**libration (AlCa) group [111]. Prominent objectives are the calibration of the ECal crystals using a dedicated laser monitoring system, calibration of the HCal as well as calibration and alignment of the tracker using events expected to show a symmetric distribution of tracks, alignment of the muon system, taking effects of the magnetic field into account.

5.4.1 Tracking

For charged particles the innermost part of the reconstruction originates from the silicon sensors of the tracker's pixel and strip detectors. Signals passing the internal zero-suppression are clustered into *hits* based on the position of measured free charges initiated by ionization processes in the active material. Applying several dedicated algorithms, adjacent pixel cells or neighboring strips are combined to account for the particle's Lorentz drift and result in a combined spacial resolution in the order of 10 µm in the barrel region of the pixel detector [112].

A *iterative tracking* sequence is performed on the collection of hits, to determine the particle trajectory's position and momentum. Four steps are carried out over several iterations, where initial parameters and selection criteria are adapted successively. After each iteration, hits associated with

identified tracks are removed from the collection of processed hits. Track *seeds* are selected from two or three hits in the tracker's inner region, taking advantage of the pixel's high spatial resolution and a low channel occupancy. Track finding is based on pattern recognition and fitting techniques of an adapted combinatorial **K**alman **F**ilter (KF) [113]. Layer by layer, the coarse seed tracks are extrapolated to outer tracker regions, taking energy losses and curvature into account. Once a given minimum number of valid hits is reached, potential additional hits are included through an inwards search. Final track candidates are given after a second KF and smoothing procedure and reevaluation with a χ^2 requirement. To reject fake tracks, not associated with an actual charged particle, selection criteria on the number of tracker layers with hits and on parameters of the track fit are applied. Three levels of selection purity are defined for which severity is increased with each iteration. Tracks passing the least restricting requirements are kept in the general collection of tracks. After completing all iterations, tracks passing the most stringent requirements from all iterations, are merged to form an ambiguity-free collection of *high-purity* tracks.

5.4.2 Primary Vertices

Reconstruction of a primary vertex is based on clustering of track information. An approach is chosen that performs well for events with many vertices, as expected under the high PU conditions, but refrains from splitting a single genuine vertex into several individual ones. For all tracks and vertices deterministic annealing clustering [114] is applied to assess the probability with which a specific track is associated to a specific vertex. In the algorithm's initial phase all reconstructed high-purity tracks are collapsed into a single vertex. Gradually, sub-clusters based on the z-coordinate of the track's closest point to the nominal beam line are formed until a predefined cut-off, motivated by the track resolution, is reached. The final position of each primary vertex is reached through a fitting procedure, where only vertices with at least two tracks are taken into account [112]. As given in [115] the efficiency of the primary vertex reconstruction reaches values close to 100% for vertices with at least two tracks, assuming an efficient track reconstruction. An event's leading primary vertex is determined by the sum of corresponding track momenta and is taken to be the hard interaction.

5.4.3 Muons

Initially, the muon trajectory is constructed separately in the tracker volume and muon system. Hits in the silicon detectors are used to reconstruct a set of *tracker tracks*. Similarly, track segments created from hits in DTs and CSCs are used as seeds in the inner chambers. Based on information from the full muon system, a *standalone-muon track* is determined by a KF technique. Finally, the nominal interaction point is used to extrapolate the muon's track into its place of origin [116].

As already discussed in Sec. 5.1.5 and especially Fig. 5.6, the muon transverse momentum resolution improves when combining the independent tracker and muon system measurements, profiting from the excellent tracker resolution for low momentum muons and long lever arm of the muon chambers under a strong magnetic field for higher momenta. Nevertheless, high energetic muons are challenging to reconstruct accurately. Their trajectory is rather straight and subject to notable energy loss through induced showers of secondary particles. Four track reconstruction algorithms were developed to address these issues and will be explained briefly in the following. The *Tracker* and *Global Muon* approaches are the most common and applicable for low and medium p_{T} muons. For increasing p_{T}, muons loose their properties as minimum ionizing particle and tend to emit bremsstrahlung photons. Specialized methods, the **T**racker **P**lus **F**irst **M**uon **S**tation (TPFMS) and *Picky Muon* approach, improve the momentum measurement of high p_{T} muons [103, 117, 118].

The Global Muon reconstruction is based on an outside-in matching in a defined η-ϕ area by selecting a subset of tracker tracks that roughly correspond to a given standalone-muon track. A global refit to each track pair is performed and the one with the best χ^2 value is chosen as the full muon trajectory. However, the refits do not necessarily always return a valid result for every standalone-muon track. The efficiency is optimized for higher p_{T} muons with track segments in at least two muon chambers. For a Tracker Muon, hits in the muon system are used as candidate muon seeds, but the final properties, resulting from a track fit considering all tracks with $p_{\mathrm{T}} > 0.5$ GeV and $p > 2.5$ GeV, are based solely on information from the silicon tracker. Since low p_{T} muons do not traverse the entire muon system[2] and are subject to multiple scattering, it is advantageous to only require at least one muon station hit in the final inside-out track-to-segment matching.

Although about 99% of muons initiating hits in the muon system are reconstructed as either a Global or a Tracker Muon, the estimate of their original momentum can differ greatly from the value retained from the dedicated high p_{T} algorithms. The TPFMS method refits a given Global Muon track considering all tracker hits, but only the innermost muon station containing hits. Thus, the influence of showering in subsequent chambers is reduced. Since some of the outer muon stations might in fact yield valid hits, the Picky Muon approach only removes hits from chambers apparently containing showers. If the occupancy in a chamber is above the chosen threshold, hits not compatible with a χ^2 comparison to the extrapolated Global Muon tracker track are ignored. The precise application of the muon reconstruction algorithms is given in Sec. 9.2.2, emphasizing methods relevant for the analysis documented in this thesis.

5.4.4 Electrons and Photons

Electrons and photons initiate electromagnetic showers measured in a number of adjacent ECal crystals, reconstructed by clustering algorithms [119, 120]. The cluster's shape is influenced by several effects, as will be discussed in the following. Being charged particles, electrons also leave hits in the silicon tracker layers. While crossing through the tracker material, electrons follow a curved trajectory in the magnetic field which is altered because of significant energy loss though bremsstrahlung. To incorporate the energy of the radiated photons spread out in ϕ along the electron's path but rather confined in η, the Hybrid clustering algorithm searches for a seed crystal above an energy threshold of 1 GeV and then allocates additional groups in $\eta \times \phi$ of 3×1 and 5×1 crystals centered around the seed crystal. The resulting *supercluster* is removed from the potential seed crystals and the procedure is re-run in descending cluster E_{T} until all crystals above a noise threshold of $E_{\mathrm{T}} = 0.1$ GeV have been assigned to a supercluster.

The crystal layout in the endcaps does not follow a fixed η-ϕ grid, so clustering is achieved with an alternative procedure. Beginning with the most energetic crystal, the Multi5x5 algorithm constructs a 5×5 matrix of crystals around the seed. All crystals with $E_{\mathrm{T}} > 0.18$ GeV can initiate a new cluster if not already assigned to a cluster and if they represent a local maximum of their four direct neighbors. To incorporate effects of bremsstrahlung, the outer 16 crystals of the matrix are not assigned exclusively to one cluster and can be used to seed overlapping clusters. Finally, a strip of $\Delta\eta \times \Delta\phi = 0.14 \times 0.6$ is moved along ϕ direction, combining all clusters with $E_{\mathrm{T}} > 1$ GeV to a supercluster, unless they have already been assigned to a supercluster with a higher E_{T} sum.

The high amount of bremsstrahlung has a strong influence on the electron's path through the tracker layers. 35% of 50 GeV electrons radiate over 70% of their initial energy before reaching the ECal [119]. Estimating the energy loss distribution by a single Gaussian, as assumed by the default KF technique (Sec. 5.4.1), cannot be applied to electrons. Instead, the **G**aussian **S**um **F**ilter (GSF) is based on a Gaussian mixture model for track extrapolation and subsequent smoothing

[2] Only muons with $p_{\mathrm{T}} > 5$ GeV reach the first muon station.

[121]. Each GSF track is seeded by the energy weighted position of an ECal supercluster, to which groups of two or three hits in the inner pixel and TEC layers are matched [122]. The rate of jets misidentified as electrons is reduced by applying selection criteria at seeding level. Moreover, secondary electrons resulting from photon conversions following the bremsstrahlung can lead to ambiguously assigning two electron candidates to a single supercluster. A cleaning is applied, retaining the candidate with the innermost tracker hit. If two candidate hits are found in equal tracker layers, the candidate with the more appropriate energy to momentum ratio is chosen.

ECal superclusters lacking associated tracks are taken as photon candidates. However, the material budget of the tracker of up to two radiation lengths causes over 40% of photons to convert into electron[3] pairs [123, 124]. Tracks from the produced electrons are displaced with respect to the primary vertex and are not identified efficiently by the seeding procedure in the pixel detector. A combination of two steps is used to supply satisfactory results. The inward tracking approach searches for hits in the three outer tracker layers compatible with a given ECal supercluster under the photon to electron pair hypothesis. For runtime reasons a KF is used to extrapolate the hits to the inner tracker. Two of the resulting oppositely charged tracks, those with the highest number of hits, are then used to initiate the second method which propagates outwards. Independently, both of these tracks are seeded by the innermost hit, assumed as the conversion vertex, and pairs of hits are identified in subsequent tracker layers.

5.4.5 Particle Flow

Representing a thorough interplay of all subdetectors, the **P**article-**F**low (PF) algorithm [125, 126] profits from the CMS detector's precise tracking and high granularity calorimeters. Improved values for muon, electron and photon momenta are determined, allowing excellent identification and measurement of composite objects at high efficiency. Initiated by reconstructed subdetector signatures, i.e., tracker tracks, muon system tracks, calorimeter clusters, the PF method employs a linking algorithm to connect the individual elements into a consistent description, referred to as *blocks*, where overlapping entries are resolved to reveal close-by particles. Based on methods summarized in the sections above, five types of particles are identified from the PF blocks in each event. In the order given below, motivated by the achievable detector resolution, particles are built successively and coinciding blocks are removed. Tracks not associated with muons or electrons are required to pass additional quality criteria.

- Muon: A PF muon is based on Global and Tracker Muons, adding subsidiary information from both calorimeters. Dedicated selection criteria allow the PF algorithm to also retain non-isolated muons originating in jets [103, 127].
- Electron: In addition to the electron reconstruction methods mentioned above, PF includes a clustering algorithm optimized to identify even low p_{T} electrons in jets [127, 128].
- Charged hadron: Remaining tracks are interpreted as charged hadrons if their tracker momentum is compatible with linked ECal and HCal deposits under a charged pion hypotheses. Calorimeter clusters can be split to better match p_{T} determined by the tracker. An additional track fit combining tracker and calorimeter information gives the final four-momentum.
- Photon: If no track is linked to a given ECal cluster a PF photon is reconstructed. Furthermore, a charged hadron candidate is assigned an additional photon if the energy determined from the ECal is much larger than its track momentum estimate.
- Neutral hadron: Similar to PF photons, a trackless HCal cluster initiates a PF neutral hadron.

[3]For sake of conciseness, the term electron is used to refer to electrons and positrons throughout this document.

These PF candidate particles are in term used to construct composite objects such as jets or the missing transverse energy of the event, described individually in the following sections.

5.4.6 Jets

Following the nature of the strong interaction, gluons and quarks cannot be observed directly as isolated particles, but give rise to colorless bound states. In a collider environment fragmentation and hadronization processes result in collimated bundles of stable hadrons [129]. During fragmentation new quark-antiquark pairs arise from gluons radiated off the initial partons. This process cannot be described by perturbative QCD calculations and is modeled with, e.g., PYTHIA6 [130], on which more details are given in Sec. 8.2. The hadronization then describes the transferal of primary and secondary partons to the measured hadrons. For initial particles with sufficiently high momenta, the rather light constituents are boosted in a common direction, relieving a roughly cone shaped topology referred to as a *jet*.

Several algorithms have been established to address the task of reverse-engineering the original proton partons from jets given in the event, which can be categorized as cone and sequential recombination algorithms [131]. The main task consists of clustering the individually reconstructed particles to well-defined and reasonably confined objects, given the challenge of a potentially very high number of jets per event. Moreover, desirable properties of such an algorithm are collinear[4] and infrared[5] safety for jets over a certain p_{T} threshold [132]. Both ensure that the determined jets are unambiguous representations of the intrinsic perturbative QCD processes. Furthermore, a quick computation time is advantageous and ample resolution of the jet's energy is essential.

The anti-k_{T} algorithm [133] is a sequential clustering algorithm and defines the distance measures:

$$d_{ij} = \min\{p_{\mathrm{T}}(i)^{-2}, p_{\mathrm{T}}(j)^{-2}\} \cdot \frac{\Delta \mathrm{R}^2}{R^2} \qquad \text{and} \qquad d_{iB} = p_{\mathrm{T}}(i)^{-2} \tag{5.6}$$

between four-vectors of two entities ij and between an entity i and the beam B. The parameter R relates to the radius of the jet's cone in the η-ϕ plane and $\Delta \mathrm{R} = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ determines their separation. In an iterative procedure, the smallest d_{ij} is identified and the corresponding entities are recombined. If at any step d_{iB} is smaller than all d_{ij}, object i is declared a jet and removed from subsequent iterations. This approach fulfills all criteria mentioned above and yields cone shaped jets, to which each particle is assigned unambiguously. Throughout analysis of CMS Run I data, a radius parameter of $R = 0.5$ is commonly chosen. The implementation relies on FastJet [134], which ensures rapid computational performance based on reducing the number of sequences needed for a fixed number of input particles.

PF candidates are used as input to the clustering and the resulting jets are referred to as PF jets. Since all candidates are interpreted as jets, a cleaning procedure is applied to remove jets which have already been identified as other particles by dedicated algorithms, such as isolated muons or electrons. The effect of PU is partially mitigated by the so-called **C**harged **H**adron **S**ubtraction (CHS) [135]. Any track not associated with the leading (in terms of track momenta sum) primary vertex is removed from the jet clustering procedure if it is reconstructed as a PF charged hadron and passes further quality criteria to insure it originates from PU. Around 50% of PU occurring in the tracker region is eliminated by the CHS [136]. The neutral components are corrected for PU later on, following procedures outlined in Sec. 9.2.1.

To ensure reconstructed jets represent a good description of underlying parton level processes, several event based calibrations are applied. These influence the amount and resolution of energy

[4] Results are insensitive to splitting a single jet into several parallel ones
[5] Results are insensitivity to soft gluon radiation

assigned to each jet, referred to as the **J**et **E**nergy **S**cale (JES) and **J**et **E**nergy **R**esolution (JER) [136]. Both issues are not explicitly addressed by AlCa operations, since jet reconstruction relies strongly on the complex interplay of all subsystems. In a factorized approach called **J**et **E**nergy scale **C**orrections (JEC), a calibration of the JES in measured data and in simulation is performed. Initially, a supplementary energy offset from remaining PU contributions is subtracted by estimating an average p_T density per unit jet area in the η-ϕ space of every event. The correction is given as a function of the jet η and p_T and determined from simulation. Additionally, residual differences between measured and simulated PU values are accounted for in data. In a second step, non-linearities of the calorimeter response are estimated from simulation by matching jets constructed from particles on generator level with those passed through the reconstruction. These calibrations are applied to simulation and data. Finally, two corrections applied only to measured data, adjust for residual η and p_T dependent differences to jets from simulation. The balanced topology of di-jet systems determined fully from data is exploited in events exhibiting one jet in the central and one in the forward regions to determine differences in the amount of reconstructed jet energy as a function of η. An absolute JES is probed with events containing precisely measurable objects and no neutrinos, i.e., a photon or Z^0 boson and jets at various p_T values. Since the JER is found to be sharper in simulation than in data, a smearing of the reconstructed jet p_T is applied to simulated events. An appropriate scale factor is estimated by methods similar to those used for the residual JES corrections, now placing emphasis on the width of the response distribution and not its mean value. The η dependent factors are propagated to the simulated jets on analysis level following a procedure from [137], as is discussed in Sec. 9.2.5.

5.4.7 Missing Transverse Energy

Projected onto the transverse plane, the initial partons' momenta cancel out. Following momentum conservation this must also hold true for all particles produced in the collision. An event's so-called *missing transverse energy* MET is calculated as the negative vector sum of the four-vector transverse momenta - $\sum p_T$ of a consistently reconstructed set of objects in the event. This quantity is a vital indicator for the amount of energy resulting from the momentum imbalance originating from undetectable objects such as neutrinos or specific novel BSM particles.

Superiority of the PF algorithm is harnessed to precisely determine the amount of MET based on the consistent set of all PF objects [138]. The intrinsic MET determined by PF suffers from three influences that are addressed as outlined here briefly. Although PU interactions lead to additional energy in the event, the extent of introduced MET is negligible. Yet, PU can influence the resolution of particles used to construct PF MET, leading to overestimation. Similar to the CHS, the so-called *Type-0* correction removes contributions from charged particles matched to a PU vertex, but also takes neutral PF particles into account. The *Type-I* correction incorporates all JEC in the clustered part of MET, where the effective sum of all corrections is applied. As a result, all PF jets with $p_T < 10$ GeV are summarized in a quantity referred to as unclustered energy. Furthermore, a modulation in ϕ is observed for transverse momenta reconstructed with PF, which is due to imperfections in the alignment of detector subsystems and further calibration factors. This asymmetry translates to a shift along the x and y components of MET and is mitigated on event-by-event basis via factors determined from clear signals of $Z^0 \to \mu\mu$ events, where no MET is expected. Further effects, stemming mostly from detector anomalies, can lead to large amounts of MET in single events and are overcome with dedicated measures summarized in Sec. 9.1.2.

5.5 Data Processing

During data-taking periods, over 5 PB are collected annually from the combined CMS trigger system. This tremendous amount of data must be stored and processed by offline reconstruction algorithms, whose output must in term be stored in a manner easily accessible to end-users. The following two sections give a survey of the computing hardware and software techniques engaged to cope with this demanding task. Many more details can be found in [139].

5.5.1 Data Storage

A globally distributed grid of interconnected computing centers, the **W**orldwide **L**HC **C**omputing **G**rid (WLCG) [140], provides the hardware infrastructure for analyzing and storing measured data as well as simulations (see Sec. 8.2). Based on a tiered hierarchy, requirements of the major LHC experiments are met. CMS data processed by the HLT is written to the T0 center at CERN, also used to run reconstruction algorithms and offer long term storage. Throughout Europe, North America and Asia, **T**ier-**1** (T1) centers provide resources for re-running the reconstruction and offer a permanent archive, creating redundancy. Located at universities and laboratories in almost 40 different countries, roughly 150 **T**ier-**2** (T2) centers are used to ensure a direct user access. Storage services for data acquired from T1s, as well as computing and storage means for simulated data sets are given here. Hence, the chosen setup offers longevity and flexible scalability.

5.5.2 CMS Software

Used for both offline reconstruction and physics analysis, the underlying software provides a highly modular framework for the thousands of globally dispersed collaborators. The centrally maintained software package **CMS S**oft**w**are (CMSSW) [95] is built around a `C++` class bundling all relevant properties of the collision in the so-called **E**vent **D**ata **M**odel (EDM) event. Users can reliably access all necessary provenance information and construct an analysis sequence from preexisting modules, or adding ones designed to meet their specific needs [141]. Input and output methods are based on `ROOT` [142] files. Additionally, a collection of tools performing common selection and cleaning procedures, required by the majority of CMS analyses, the **P**hysics **A**nalysis **T**oolkit (PAT) is available [143, 144]. This supplementary layer provides higher-level objects for easier access and comparability and helps reduce the amount of time invested on solving common tasks.

Reconstructed events are available in several data formats. These so-called data *tiers* differ with respect to their level of detail and are conceptualized for various applications such as detector alignment and calibration as well as physics analyses. Information returned by the HLT to the T0 constitutes the RAW format which contains the full recorded detector read-out and trigger decision. The RAW data are split into distinct *primary data sets*, depending on the fired triggers. Following the time-consuming pattern recognition and compression algorithms of the reconstruction, events are transfered into the **Reco**nstruction (RECO) format. Sufficient information is encompassed to permit a consistent recalibration or alternative reconstruction at this stage. To reduce storage space and runtime, the compact **A**nalysis **O**bject **D**ata (AOD) format offers the foundation for physics analyses, by removing much of the low-level detector information. Concerning events simulated with Monte Carlo techniques (see Sec. 8.2), output of the event generators is stored in the **Gen**eration (GEN) format, including the full process information. The CMS detector simulation is divided into two steps. First, energy deposits in the detector volumes are saved in the **Sim**ulation (SIM) format, followed by simulation of the digitized detector response retained in the **Digi**talization (DIGI) format. After running the L1 and HLT selection, the events are converted into RECO and subsequently AOD format.

Part III

Model Unspecific Analysis

Chapter 6
The Concept of Model Independent Searches

6.1 Motivation

The circumstances highlighted in Ch. 3 give indications that the SM as known today does not represent a complete description of particle physics and searching for new signatures at collider experiments provides promising possibilities of extending our knowledge. Typically, the search strategy for a collider-data analysis is developed around specific BSM model predictions. Investigated final states and kinematic selections are chosen and optimized to enhance sensitivity to a given model. The sizable number of proposed new physics models makes it worthwhile to consider more generalized approaches, helping to reduce the needed time and workforce. Furthermore, unforeseen phenomena not yet included in any theoretical model could be overlooked by a model-driven search, let alone phenomena resulting from detector or simulation performance.

A model independent search does not optimize its investigated final states or kinematic regime to a dedicated BSM scenario. Instead, only arguments based on detector performance and validity of the simulation are used to restrain the data. In accordance with this strategy, easily comprehensible, robust and unbiased selection criteria are applied to ensure well-measured events and objects. Yet, this comes at the price of a reduced sensitivity compared to analyses specialized to a designated signal. Moreover, a subsequent detailed investigation is needed to understand the underlying origin of potential deviations uncovered by such a model independent search.

6.2 Pre-LHC Model Independent Searches

Analysis groups at collider experiments preceding the LHC have already successfully carried out model independent search strategies. An initial study based on LEP data collected with the L3 experiment calculated the significance of deviations between data and SM simulation events based on a χ^2 test [145] and introduced many of the concepts used by the analysis presented in this thesis.

The SLEUTH quasi-model-independent search was applied to proton-antiproton collider data from Tevatron Run I taken with the D0 experiment. After carrying out variable transformations to ensure a uniform distribution of the SM expectation, the significance of excesses in high p_{T} data events containing at least one electron and at least one muon was calculated [146]. Later, the SLEUTH search regime was broadened to additional inclusive event topologies, including a single lepton and MET, like-sign same-flavor lepton pairs or all combinations of leptons and photons including at least three particles[1] [147, 148]. These results were reinterpreted by the QUAERO analysis in terms of generic 95% confidence level upper limits on the cross section of a provided BSM signal, thus offering the possibility to gain a quantitative outcome in case no signs of new physics were found in the analyzed data [149].

A general search in final states with at least two objects (electrons, muons, photons, jets, MET) was performed based on (anti)electron-proton collisions from HERA at the H1 experiment at up to $\sqrt{s} = 319$ GeV [150, 151]. An algorithm was developed to identify the most significant deviation between data and SM simulation in a given distribution. Inspired by the SLEUTH statistical methods for calculating significances, the algorithm was extended to locating excesses and deficits over the full observed kinematic range.

[1] Here leptons refers to only electrons and muons, not taus.

At the **C**ollider **E**xperiment at **F**ermilab (CDF) experiment during Tevatron Run II, SLEUTH was accompanied by the VISTA correction model, and the search regime was expanded to excesses and deficits in several hundred final states [152]. Based on the comparison of data and SM simulation, VISTA was designed to determine scale factors from the bulk of the distributions and deviations in high p_{T} tails of the distributions were uncovered by SLEUTH. The VISTA corrections were subsequently applied to the SM expectation, provided the dependencies could be proven to originate from inadequate modeling of the SM processes or detector response. Performing these steps in an iterative fashion simultaneously on all final states under investigation provides a way of determining theoretical and experimental correction factors comparable to systematic uncertainties at first hand, without withdrawing sensitivity to unexpected signals. The search was extended to a larger data set, also accompanied by a method to search for localized bumps in invariant mass distributions [153]. Later, the combination of VISTA and SLEUTH was also applied to data taken with the D0 experiment [154].

Of the analyses summarized above, none reported significant deviations from the SM expectation that could be attributed to new physics phenomena, but each contributed to a better understanding of the SM simulation in the context of collider experiments.

6.3 Model Independent Searches at the LHC

The **M**odel **U**nspecific **S**earch **i**n **C**MS (MUSiC) analysis is a generic search using CMS proton-proton collision data and is based on many of the techniques developed by the H1 general search, as will be elaborated on in Ch. 7. Introduced to the experiment before the start of the LHC, the initial algorithm was implemented and studies assessing the analysis performance and proofs of principle for the applied strategies were carried out with pseudo-data [3, 4, 155]. A focus was set on events containing at least one electron or muon and any amount of jets, photons and missing transverse energy MET. It could be shown that injected BSM signals or detector effects can be recovered by the algorithm.

Applying the analysis to $L = 36.1\ \mathrm{pb}^{-1}$ of collisions taken in 2010 at $\sqrt{s} = 7$ TeV revealed good agreement between the measured data and the fully **M**onte **C**arlo (MC) simulation based SM expectation [156]. This also held true when photon triggered events and event classes distinguished by the net lepton charge were included in the search [1]. At the same center of mass energy, $5.5\ \mathrm{fb}^{-1}$ of data taken in 2011 were analyzed, also including τ+MET triggered events. Although throughout good agreement between data and SM MC was observed, individual event classes showed issues of slight mismodeling or insufficient number of generated events in the available SM prediction for this increased amount of data [2]. The analysis presented in this thesis represents the continuation of the MUSiC analysis with data taken in 2012 at 8 TeV.

Beginning with the 2011 data-taking period at 7 TeV, a general search also following the H1 procedures was performed within the ATLAS Collaboration on $4.7\ \mathrm{fb}^{-1}$ of proton-proton collision data [157, 158]. Events triggered by leptons, photon, jets and MET were analyzed, also taking bottom quark tagging for jets into consideration. Results show the data to be in accordance with the SM expectation. The ATLAS general search was re-iterated with $20.3\ \mathrm{fb}^{-1}$ worth of data taken in 2012 at 8 TeV in which no significant deviation was observed [159, 160]. Moreover, collision at 13 TeV, amounting to $3.2\ \mathrm{fb}^{-1}$, were analyzed following the same methods and showed good agreement between data and SM expectation [161].

Chapter 7
Analysis Workflow

The MUSiC analysis is divided into three distinct steps, differentiated in concept by their purpose in the analysis but mainly by the disjunct types of computing resources used to perform each individual task. In short, these analysis steps are referred to as skimming, classification and scanning, which are outlined in the following sections. Many parts of the MUSiC workflow were established by analyses summarized in Sec. 6.3 and are the basis for further developments made in this thesis and in parts in close collaboration with S. Knutzen, as documented in [162].

7.1 Data Pre-processing

The CMS Collaboration supplies simulation and collision event data sets in several dedicated data-formats with varying degree of detail. Like the majority of CMS searches, MUSiC is based on the AOD format (see Sec. 5.5.2), which condenses the full recorded detector information into physics objects passed through specialized reconstruction algorithms. Consequently, the size per event is reduced from roughly 1.5 MB containing the full detector simulation (see Sec. 5.2) to 0.05 MB in the AOD format[1] [139]. Even in this reduced data-format the AOD data sets, stored remotely at different T2 sites of the WLCG (see Sec. 5.5.1), contain more information than needed by MUSiC. In a combined analysis step, referred to as *skimming*, superfluous information, i.e., low energy objects and in-depth details of the reconstruction, is disregarded and the desired data sets are retrieved to storage on the local T2 storage facility in Aachen. To cope with the large amount of processed data, tasks are run highly parallelized with the help of the **C**MS **R**emote **A**nalysis **B**uilder (CRAB) [164]. The slimmed output is written using **P**hysics e**X**tension **L**ibrary (PXL) [165], an object-oriented framework optimized for storing, accessing and manipulating event and particle information through suitable containers, reducing the event size by roughly a factor of five.

The skimming step is implemented as a dedicated analyzer module of the centrally distributed software introduced in Sec. 5.5.2. For the analysis of 2012 CMS data CMSSW version 5_3_14 [166] is used, where details of the data sets are given in Sec. 8.1 and Sec. 8.2. At this stage no events are discarded, but flags are set to identify ones failing criteria of proper detector performance, amounting to under 2% as described in Sec. 9.1.2. Furthermore, parts of the reconstruction algorithms are re-run with the help of PAT, to apply corrections on jets and MET as explained in Sec. 9.2.

7.2 Classification of Events

A key feature of MUSiC is the consistent analysis of hundreds of final states. The full procedure, referred to as classification, is described in the following paragraphs. Reading the information saved to the PXL files written by the skimmer, data and simulation events are analyzed within a custom framework developed in the Physics Institut IIIA, Aachen over the last years. Each final state is handled independently in an object-oriented manner and written to a `ROOT` [142] file. Details of the implementation are summarized in [2]. To speed up the classification process, tasks are run in parallel either on the local Aachen T2 with help of **C**omputing **R**esource **E**xecution **a**nd **M**anagement (CREAM) gLite middleware [167], or via HTCondor [168] on the Aachen desktop cluster, enabling the entire step to be completed in under 12 hours.

[1] Due to the expected increase of collected data, an even conciser event data-format was developed for LHC Run 2. The `MiniAOD` format is about ten times smaller than AOD [163].

7.2.1 Event Classes

Individual units based on objects reconstructed in an event's final state, so-called *event classes*, are constructed to separate different types of physics processes, helping to locate and understand causes of potential unanticipated signatures. To determine an event's physics object content unambiguously, selection criteria optimized for the precise LHC and CMS run conditions (e.g., center of mass energy and reconstruction details) are applied to the data and SM simulation, returning a defined number of well-identified final objects state. The used criteria are given in Sec. 9.

Each event is considered separately under three event class definitions, as shown schematically in Fig. 7.1. *Exclusive* event classes represent orthogonal subsets of the analyzed events, in that they place an exact constraint on the multiplicity of all objects in the final state, implicitly applying a veto on objects not included in the event class's name. Therefore, every event is assigned to precisely one exclusive event class, facilitating the statistical treatment by minimizing correlations between event classes. Signatures of one new phenomenon can occur simultaneously in several final states due to, e.g., lepton universality. A deviation's significance may be increased by condensing these events into a single event class. Accordingly, *inclusive* event classes contain events with at least the number of objects stated in the event class name, where "+X" is used to denote potential additional objects. As a consequence, a single event can contribute to more than one inclusive event class. At a hadron collider the hard-scattering process is frequently accompanied by jets produced from initial/final state radiation and components of the underlying event [169]. In such cases the number of jets does not necessarily characterize the intrinsic physics process of interest in a specific final state. It can be profitable in terms of sensitivity to combine these events into a single *jet-inclusive* event class, which fixes the number of all objects with the exception of jets. A further merit is that this type of event class eases the comparison with other CMS analysis, which often confine their search to a specific number of leptons or photons but do not require a specific jet multiplicity. The suffix "+Njet" is added in the event class notation.

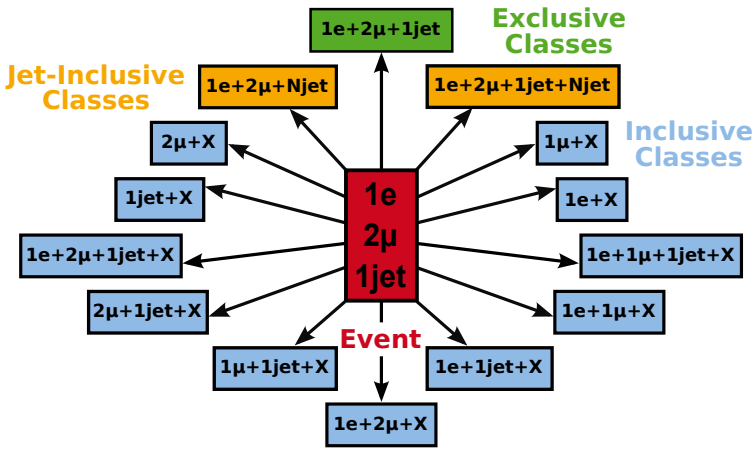


Figure 7.1: Schematic example for the MUSiC definition of event classes. The event, shown here in red, contains one electron, two muons and one jet and is assigned to exactly one exclusive event class (green). It also contributes to several inclusive (blue) and jet-inclusive (orange) event classes.

As described in Sec. 5.2, events in CMS are only recorded if they pass a specific trigger requirement, where specifics are given in Sec. 9.1.3. In approximately 10% of the time, an event is triggered by an object which does not pass criteria applied to insure a satisfactory reconstruction (see Sec. 9.2). Such events are excluded from the analysis.

A study of top-antitop quark pair events shows an increase of systematic uncertainties, given especially by the statistical uncertainty of the used simulation data sets, to over 20% for events with more than five jets [170]. Only relatively few such events occur at $\sqrt{s} = 8$ TeV so applying an upper threshold on the number of jets helps reduce issues of low number of generated glsmc simulation

events, since many of the simulated processes do not include more than four jets on matrix element level (see Sec. 8.2.1 for details). Consequently, all events contributing to exclusive event classes with six or more jets are combined into a single exclusive event class, e.g., 1e+2μ+≥6jets. Jet-inclusive event classes with at least as many jets as given in the threshold, e.g., 1e+2μ+6jets+Njet, are identical to the corresponding exclusive event class and not considered separately.

Lepton Charge Aware Analysis

Within the SM pairs of charged leptons produced in proton-proton collisions most frequently carry opposite sign. This is not necessarily the case for BSM signatures found in models of, e.g., supersymmetry [171] or vector-like quarks [172, 173]. Consequently, dividing the analyzed events into ones with same-sign and ones with opposite-sign di-leptons may enhance a search's sensitivity by reducing contributions from the SM background in the same-sign final states.

Similarly, MUSiC constructs event classes by considering an event's absolute total lepton-charge $Q = |\sum q_l|$. Despite a longer runtime caused by the increased number of event classes and additional uncertainties introduced to cover a possible charge misassignment for each lepton type, the analysis profits from separating different types of physics processes. Observed deviations can be associated more easily with potential causes, and sensitivity to many BSM models is enhanced.

Under consideration of the net lepton charge, the event discussed in Fig. 7.1 could induce two exclusive event classes 1e+2μ+1jet[1Q] and 1e+2μ+1jet[3Q], as well as many inclusive event classes such as 1e+1μ+X[0Q] and 1e+1μ+X[2Q] and jet-inclusive event classes.

7.2.2 Kinematic Distributions

The actual statistical analysis discussed in Sec. 7.3 is conducted on distributions of event properties constructed from object kinematics in the event class. Here, only the kinematic properties of objects mentioned explicitly in the event class name are considered. In the (jet-)inclusive case, the corresponding number of most energetic objects are used to represent the event's kinematics. For an event containing two muons, distributions of the 1μ+X event class only account for properties of the muon with the highest transverse momentum. If a specific amount of jets is explicitly given in a jet-inclusive event class, i.e., 1e+2jets+Njet, properties of the corresponding number of most energetic jets (in this case two) are used in the kinematic distributions, others are disregarded.

Three kinematic variables are chosen, which highlight different aspects of the collisions and are assumed to be especially sensitive to signs of unexpected effects. They are introduced briefly in the following, including shorthands used throughout this thesis.

Missing Transverse Energy

The amount of missing transverse energy MET is a valuable indicator for undetectable BSM particles traversing the detector. As introduced in Sec. 5.4.7, this quantity is calculated as the negative vector sum of four-vector transverse momenta of all N objects in the event:

$$\mathrm{MET} \equiv \left| -\sum_{i}^{N} \vec{p}_{\mathrm{T},i} \right| \tag{7.1}$$

and is a measure for the total momentum imbalance in the event. An overall increase in rate would appear in this variable through the production of novel non-interacting particles. For events with MET below the chosen analysis level threshold (see Sec. 9.2.6), this quantity is not considered, as the low MET regime is especially susceptible to mismeasurement and reconstruction inaccuracies. As a result the MET distribution is only evaluated for event classes containing events with MET above the threshold.

Scalar Sum of Transverse Momenta

A rather general overview of the collisions' kinematics is given by the scalar sum of transverse momenta of all N objects given explicitly in the event class name, calculated as:

$$\sum |\vec{p}_{\mathrm{T}}| \equiv \sum_{i}^{N} |\vec{p}_{\mathrm{T},i}| \,. \tag{7.2}$$

Here, $\vec{p}_{\mathrm{T},i}$ is the vectorial transverse momentum of an object i, where MET (above the analysis level threshold) is also taken as one such momentum. The $\sum |\vec{p}_{\mathrm{T}}|$ is calculated for all considered event classes. This measure is related to the total amount of energy involved in the hard interaction. Many BSM theories involve new heavy particles that would show up at high values of $\sum |\vec{p}_{\mathrm{T}}|$, where only few events are expected from SM physics. A similar definition of this quantity is used by, e.g., [174] and referred to there as S_{T}.

Combined Mass

The combined mass of all objects expands the sensitivity to angular distributions of objects in the collision and is calculated for every event with at least two objects. Event classes such as 1e(+X/+Njet) and 1μ(+X/+Njet) are disregarded in this variable. Particles produced resonantly appear as fairly localized peak in the combined mass distribution of their decay products and are the basis for several BSM scenarios. For events without a significant amount of MET, this variable is calculated via:

$$M \equiv \sqrt{\left(\sum_{i}^{N} E_i\right)^2 - \left(\sum_{i}^{N} \vec{p}_i\right)^2} \tag{7.3}$$

and referred to as the *invariant mass*. In this case $\vec{p}_i$ denotes the object's full four-vector momentum. Considering events were MET is accounted for, the *transverse mass*:

$$M_T \equiv \sqrt{\left(\sum_{i}^{N} E_{\mathrm{T},i}\right)^2 - \left(\sum_{i}^{N} \vec{p}_{\mathrm{T},i}\right)^2} \tag{7.4}$$

is evaluated, where the z-component does not contribute. The generic term *combined mass* is used when referring to both event classes containing MET below and above the analysis threshold and denoted by $M_{(\mathrm{T})}$ throughout this document.

Considering $\sum |\vec{p}_{\mathrm{T}}|$, the minimum possible value achievable in each event class can be calculated from momentum thresholds introduced by the trigger and object selection, as stated in Sec. 9.1.3 and Sec. 9.2. The MET distribution also has a sharp cut-off at the chosen analysis level restraint (see Sec. 9.2.6). Only the $M_{(\mathrm{T})}$ distribution allows for entries down to zero. The highest considered value in each distribution is given by the LHC center of mass energy, corresponding to $\sqrt{s} = 8$ TeV in the data analyzed for this thesis.

In principle, different or additional kinematic distributions can be considered, perhaps even ones more sensitive to specific aspects of BSM theories. For instance, the MUSiC analysis could profit from introducing an invariant mass distribution of opposite-sign same-flavor lepton pairs for event classes with at least two leptons. The leptons found in these events correspond to ones potentially produced by the decay of a Z boson. Measurements of the Z resonance can deliver valuable indications for physics beyond the SM. Nevertheless, it is important to keep in mind that the number of investigated distributions has a direct effect on the analysis' overall sensitivity due to an increase of the **l**ook-**e**lsewhere **e**ffect (LEE), which is explained in Sec. 7.3.4. Thus, the number of kinematic distributions should be kept small. Furthermore, if the chosen variables show strong correlations amongst each other the resulting significances cannot be seen as independent.

7.2.3 Resolution Motivated Bin Width

Further processing of each kinematic distribution is done in form of binned histograms, for which a common bin width must be determined for measured data and simulated MC events. The wide range of different final states under consideration does not allow motivating a single fixed binning for each kinematic variable, yet the large number of investigated event classes makes an automated procedure inevitable. Modifying the amount of bins in a distribution changes the number of regions investigated by the search algorithm (see Sec. 7.3) and can in general alter the overall significance determined for each event class. Thus, in order to avoid an additional bias, the binning is fixed before carrying out the actual statistical analysis.

Though a relatively large bin width is favorable in terms of computation time and also reduces the severity of the LEE (see Sec. 7.3.4), it decreases the analysis' sensitivity to narrow deviations. By choosing a small bin width, random fluctuations can gain in importance and possibly mask the actual deviation of interest. A tangible choice is a variable width, motivated by the expected total detector resolution of all objects in the specific event. A minimum bin width of 10 GeV is set as a starting point, of which all bin widths are integer multiples.

Several assumptions are made to determine an appropriate resolution for every event filled into the kinematic distributions, as the procedure is applied at the level of event variables ($\sum |\vec{p}_{\mathrm{T}}|$, $M_{(\mathrm{T})}$ and MET) and not the event's individual objects. The total detector resolution strongly depends on an event's topology and object content. Yet, estimates of the detector resolution are available only for individual physics objects as a function of their transverse momentum or energy. Consequently, a method is used to generalize these functions yielding variables accessible for each event without further knowledge of the individual object's properties. In the next paragraphs, p_{T} dependent resolution functions for all considered objects are summarized, followed by details of how these functions are translated to the bin width in each event class distribution.

For muons, a p_{T} dependent resolution function is determined from muon-gun simulations as summarized in [175]. The MUSiC analysis relies on fitting the results given in [175], yielding:

$$\sigma_{\mu}(p_{\mathrm{T}}) = \left(0.015 \cdot \frac{p_{\mathrm{T}}^2}{\mathrm{GeV}} + 1.6 \cdot p_{\mathrm{T}}\right) \cdot 10^{-2} \tag{7.5}$$

for the muon transverse momentum resolution σ_{μ} as a function of the muon's p_{T} (in GeV).

Along the lines discussed for the ECal resolution in Sec. 5.1.2, a function for the transverse energy resolution electrons and photons is derived from test beam measurements with electrons to [79]:

$$\sigma_{\mathrm{e}/\gamma}(E_{\mathrm{T}}) = \sqrt{(0.3 \cdot E_{\mathrm{T}})^2 + 2.8^2 \text{ GeV} \cdot E_{\mathrm{T}} + (0.12 \text{ GeV})^2} \cdot 10^{-2}. \tag{7.6}$$

Here, E_{T} refers to the transverse component of the electron or photon energy.

An estimate for the jet transverse momentum resolution is obtained for PF jets with $2 < \eta < 2.5$ from [136]. Assuming 25 pileup interactions in each event and a jet area of $A = \pi \cdot 0.5^2$ in $\eta - \phi$ space motivated by the anti-k_{T} algorithm's radius parameter, the jet p_{T} dependent function becomes:

$$\sigma_{\mathrm{jet}}(p_{\mathrm{T}}) = \sqrt{(0.043 \cdot p_{\mathrm{T}})^2 + 0.85^2 \text{ GeV} \cdot p_{\mathrm{T}} + (3 \text{ GeV})^2} \ . \tag{7.7}$$

The resolution of MET is shown to be proportional to the event's total transverse momentum sum. A parameterization is derived from measurements of γ+jets and $Z \to \mathrm{ee}/\mu\mu$ events [138]:

$$\sigma_{\mathrm{MET}}(p_{\mathrm{T}}) = 1.78 \text{ GeV} + 0.63 \cdot \sqrt{\sum |\vec{p}_{\mathrm{T}}| \ \text{ GeV}} \ , \tag{7.8}$$

where $\sum |\vec{p}_{\mathrm{T}}|$ is evaluated using Eq. (7.2).

For each object a resolution function of the object's transverse momentum or energy are now translated into a consistent binning of the three kinematic variables. The width of each bin in the $\sum |\vec{p}_{\mathrm{T}}|$ distribution is determined by assuming the event's total $\sum |\vec{p}_{\mathrm{T}}|$ is evenly distributed over all objects in the event. With $\langle p_{\mathrm{T},i} \rangle$ representing this mean transverse momentum assigned to each of the N objects i in an event, the total resolution $\sigma_{\sum |\vec{p}_{\mathrm{T}}|}$ of the $\sum |\vec{p}_{\mathrm{T}}|$ variable is calculated as:

$$\sigma_{\sum |\vec{p}_{\mathrm{T}}|} = \sqrt{\sum_{i}^{N} N_i \cdot \sigma_i^2(\langle p_{\mathrm{T},i} \rangle)}, \tag{7.9}$$

where $\sigma(p_{\mathrm{T}})$ is the resolution of the individual object i and N_i is the multiplicity of each object type (electron, muon, photon, jet, met) found in the event. The MET resolution σ_{MET} is given as a function of the events total $\sum |\vec{p}_{\mathrm{T}}|$ by Eq. (7.8).

Simplifying assumptions are made to estimate the total mass resolution as a function of the combined mass. All objects in the events are assumed to be distributed equidistantly in ϕ and assigned a pseudo-rapidity of $\eta = 0$. In this reduced phase-space region an event's $\sum |\vec{p}_{\mathrm{T}}|$ and combined mass are identical, and the $M_{(\mathrm{T})}$ resolution $\sigma_{M_{(\mathrm{T})}}$ is approximated by Eq. (7.9).

As mentioned above, an event's missing transverse energy is correlated with the total $\sum |\vec{p}_{\mathrm{T}}|$ in the event. For events with genuine MET the value of $\sum |\vec{p}_{\mathrm{T}}|$ and MET are typically in the same order of magnitude. Following this simplification, the MET resolution for each bin value is approximated by replacing $\sum |\vec{p}_{\mathrm{T}}|$ with MET in Eq. (7.8).

Conserving the overall shape of a distribution, while using a dynamic bin size, is achieved by normalizing the number of events in each bin to the bin width. Thus, a smooth continuous function is also represented by a smooth distribution, even when abrupt large changes in bin width occur.

7.3 Search Algorithm

A search algorithm centered around scanning all bins of a given event class distribution constitutes the final component of the MUSiC analysis. In each of the several hundred distributions, a statistical procedure is used to locate and quantify deviations of the measured collision data from the expectation given by MC simulations of the SM. Details of the analysis input, specific to the analysis of data taken in 2012 at $\sqrt{s} = 8$ TeV, are given in Ch. 8.

In the following paragraphs, the methodology of the search algorithm's individual steps is explained. Based on time-efficient `C++` modules, steered through `python` scripts, the MUSiC *Scanner* covers the full statistical analysis of the output obtained from the classification. More details of the implementation can be found in [2]. First, search regions are constructed from all connected bins in each event class distribution (see Sec. 7.3.1). A significance measure, the p-value (given in Sec. 7.3.2), is calculated in each of these regions. The strongest deviation of the measured data from the SM MC expectation is corrected for the bias of investigating many regions simultaneously (see Sec. 7.3.4) and then represents the final significance measure of the given distribution. A joined global overview of all event classes for each kinematic variable and event class type is finally provided in a graphical approach, as explained in Sec. 7.3.7. Furthermore, methods for dealing with the low generated event count of the simulation in individual regions (see Sec. 7.3.3) and runtime optimizations (see Sec. 7.3.6) are discussed.

7.3.1 Constructing Regions

In each event class distribution, the MUSiC search algorithm relies on a significance measure, the p-value (details given in Sec. 7.3.2), to pinpoint the strongest deviation found between measured data and SM MC expectation. Deviations may appear over broad areas of the distribution or

be localized around specific values of the kinematic variables. To cover both cases, bins of the kinematic distributions are grouped into *regions*, which are formed by combining a variable number of consecutive adjacent bins. In terms of significance calculation, the bin i of an event class distribution is described fully by three quantities:

- $N_{\mathrm{obs},i}$ the number of observed data events
- $N_{\mathrm{SM},i}$ the number of events expected from the SM simulation
- $\sigma_{\mathrm{SM},i}$ the systematic uncertainty on the number of events expected from the SM simulation.

Here, $\sigma_{\mathrm{SM},i}$ also includes the uncertainty arising from the number of generated simulation events. Quantities of the individual bins are transferred to the corresponding region by adding up contributions from all enclosed bins. The total number of observed data events N_{obs} in the region and the total event count given by the SM MC prediction N_{SM} in the region are simply given as the sum of all corresponding values in the contributing bins. The total systematic uncertainty in each bin $\sigma_{\mathrm{SM},i}$ is composed of different systematic effects, some of which are taken as fully correlated over all bins of the distribution, others as fully uncorrelated. To determine the total systematic uncertainty of a region σ_{SM}, the systematic effects are treated separately: Contributions from the individual systematic uncertainty sources in the bins constituting the region are summarized according to their nature (either fully correlated or fully uncorrelated), as specified in Ch. 10. Finally, individual systematic uncertainty sources are assumed to be independent of each other and combined by quadratic summation, yielding the total uncertainty σ_{SM} in the region.

Each bin can be assigned to several different overlapping regions. With N^{bins} representing the total number of bins in the distribution, the maximum number of regions constructible amounts to $N^{\mathrm{regions}} = \frac{N^{\mathrm{bins}}}{2} \cdot (N^{\mathrm{bins}} + 1)$. When dealing with $\sum |\vec{p}_{\mathrm{T}}|$ or MET distributions a minimum region width of at least three bins is chosen to reduce sensitivity to abrupt changes in the SM MC prediction (in particular so-called *spikes*) caused by a lack of simulated events. Deviations in these kinematic variables are expected to be rather broad, so the MUSiC discovery sensitivity is not diminished. This requirement is however dropped for $M_{(\mathrm{T})}$ distributions, where resonances of new particles could potentially appear in form of narrow peaks in the measured data.

The agreement of N_{obs} to $N_{\mathrm{SM}} \pm \sigma_{\mathrm{SM}}$ is evaluated in every considered region of the distribution. Finally, the region exhibiting the most significant deviation, determined by the smallest p-value, is identified and dubbed the **R**egion **o**f **I**nterest (RoI). The deviation observed in the RoI is the one least likely to be caused by fluctuations of the SM expectation. All further steps of the statistical evaluation are based on this deviation. A schematic of the procedure is shown in Fig. 7.2.

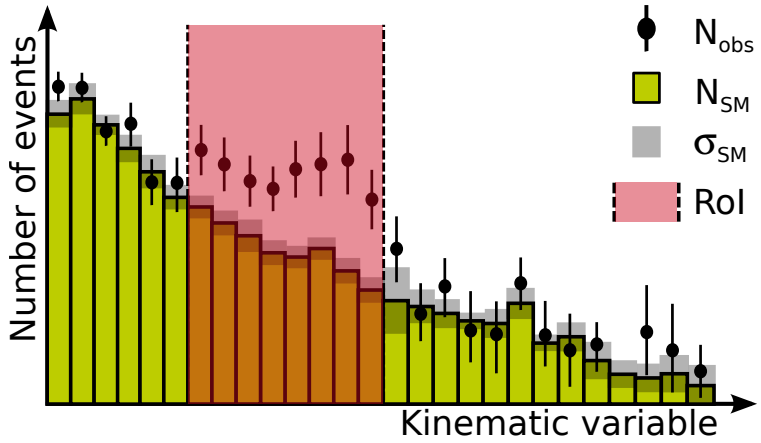


Figure 7.2: Schematic of a kinematic distribution including contributions from observed data N_{obs} as well as the SM expectation $N_{\mathrm{SM}} \pm \sigma_{\mathrm{SM}}$. The resulting RoI is shown in red and represents the region showing the strongest deviation between observation and expectation.

To quickly gain a general overview of the analyzed data, the entire kinematic distribution is condensed into one bin representing the total event yield in the corresponding event class. Since only a single region is investigated with this method, the necessary computing time is small (in total under one hour for all event class types in the data set considered for this thesis) and deviations stemming from overall normalization issues can be spotted early on.

7.3.2 Calculating the Significance of Deviations

In terms of statistical theory the MUSiC approach represents the comparison of measured data X_{obs} to a null hypothesis H_0. A p-value quantifies the probability $Pr()$ that an effect at least as strong as the one observed in the comparison could occur purely from statistical fluctuations of the null hypothesis itself [176, 177]:

$$p = Pr\Big(T(x = X_{H_0}) \geq T(x = X_{\text{obs}}) \;\Big|\; H_0\Big). \tag{7.10}$$

An appropriate test statistic $T(x)$ is chosen to evaluate the randomly distributed data x and assess the agreement of the measured data X_{obs} to the prediction given by the null hypothesis X_{H_0}. For the MUSiC analysis the test statistic is given by the number of events N in a region, for which the measured collision data N_{obs} is compared to MC simulations of the SM N_{SM}.

Processes involved in the particle collisions leading to a measured or simulated number of events N can be described by a Poisson probability density function [59]:

$$\mathcal{P}(N;\mu) \equiv Pr(N;\mu) = \frac{\mu^N \mathrm{e}^{-\mu}}{N!}. \tag{7.11}$$

Here, μ refers to the true mean of the distribution, which is not known exactly in the analysis at hand, but estimated through MC simulations. Systematic uncertainties arising from the simulation and measurement process are addressed by extending the probability description with a *prior* distribution $\pi(\theta)$, depending on a number of *nuisance parameters* θ of the model. The prior summarizes all knowledge of θ given before conducting the experiment, representing the degree of belief. As already successfully applied by the H1 general search [150], a Gaussian probability density distribution for an observed number of events x:

$$\mathcal{G}(x;\mu,\sigma^2) \equiv Pr(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \mathrm{e}^{\frac{-(x-\mu)^2}{2\sigma^2}}, \tag{7.12}$$

is used as the prior to model the influence of systematic uncertainties. The true mean μ is estimated by the number of events N_{SM} obtained from the SM simulation, whereas the distribution's width σ is given by the total systematic uncertainty σ_{SM} on the SM expectation. Nuisance parameters θ have an impact on μ and σ. By introducing a prior, the approach represents a Bayesian-frequentist hybrid method, also referred to as a *prior predictive* p-value [178, 179], for which an average of the nuisance parameters is taken according to weights introduced by the prior function. In conclusion, the p-value can be written as:

$$p = \begin{cases} \underbrace{\sum_{i=N_{\text{obs}}}^{\infty} C}_{\text{normalization}} \cdot \underbrace{\int_0^{\infty} \mathrm{d}x \, \exp\left(-\frac{(x-N_{\text{SM}})^2}{2\,\sigma_{\text{SM}}^2}\right)}_{\text{systematics}} \cdot \underbrace{\frac{x^i \, \mathrm{e}^{-x}}{i!}}_{\text{statistics}} & \text{if } N_{\text{obs}} \geq N_{\text{SM}} \\ \underbrace{\sum_{i=0}^{N_{\text{obs}}} C}_{\text{normalization}} \cdot \underbrace{\int_0^{\infty} \mathrm{d}x \, \exp\left(-\frac{(x-N_{\text{SM}})^2}{2\,\sigma_{\text{SM}}^2}\right)}_{\text{systematics}} \cdot \underbrace{\frac{x^i \, \mathrm{e}^{-x}}{i!}}_{\text{statistics}} & \text{if } N_{\text{obs}} < N_{\text{SM}} \end{cases} \tag{7.13}$$

where the cases of an excess of observed events ($N_{\text{obs}} \geq N_{\text{SM}}$) and a deficit ($N_{\text{obs}} < N_{\text{SM}}$) are treated separately in the sum arising from the Poisson distribution of a discrete number of events. The Gaussian distribution is truncated at zero, the lowest possible number of events. Therefore, a normalization C is introduced, given as:

$$C^{-1} = \sqrt{2\pi}\sigma_{\text{SM}} \cdot \int_0^\infty \mathrm{d}x\ \mathcal{G}(x; N_{\text{SM}}, \sigma_{\text{SM}}) \cdot \sum_{i=0}^{\infty} \mathcal{P}(i; x) = \sqrt{2\pi}\sigma_{\text{SM}} \cdot \int_0^\infty \mathrm{d}x\ \mathcal{G}(x; N_{\text{SM}}, \sigma_{\text{SM}}). \quad (7.14)$$

Resulting from the Gaussian component in Eq. (7.13), an analytical solution cannot be found and the integration must be carried out numerically, rendering the computation rather time consuming. Several methods of runtime optimization are applied, as discussed in Sec. 7.3.6, allowing for a sufficiently precise result within viable time-frame.

Overall, reasonable statistical properties are observed for the chosen p-value, such as monotony for increasing uncertainties and a sensible coverage probability, i.e., when drawing pseudo-data from a known distribution how often is the retained p-value at least as large as the one calculated with the true probability given from the input distribution [177, 180]. Furthermore, a study relying on a log-normal description of the systematic uncertainties proved to be a promising alternative, possibly improving coverage properties especially in the regime of few expected events accompanied by large systematic uncertainties [180]. A fully Bayesian approach, extending the p-value's credibility for very small event numbers expected from the SM, has also been considered [181]. Yet, for reasons of feasibility, numerical stability and computation time the p-value as given by Eq. (7.13) remains the MUSiC analysis' method of choice.

7.3.3 Low Generated Event Count Treatment

The SM and models beyond provide a rich topology of final states and it is the MUSiC goal to investigate as many of them as possible. SM MC simulation provides the basis for this search and the computing time needed to produce the simulation poses a hard restraint on the number of events providable within a reasonable time scale. Although, as described in Sec. 8.2.1, the simulation is subdivided into various physics processes, covering a wide range of phase space, not all final states or kinematic regions can be addressed adequately. As a result the SM description may show discontinuities, especially in the high energy tails of distributions with a large decline in differential cross section, or in event classes dominated by rare processes not simulated individually.

Regions containing neither an SM expectation nor a corresponding systematic uncertainty present a difficulty for the MUSiC algorithm. A sensible p-value cannot be calculated without knowledge of the expectation. Over the last years, two methods have been developed to address this issue and are introduced in the following paragraphs, including a qualitative comparison elaborating on each method's advantages and disadvantages. The quantitative assessment of their influence on the MUSiC sensitivity is given later on in Sec. 11.2.4 of this thesis.

SM Simulation Based Uncertainty Fill-Up

Originally developed for the MUSiC analysis of 2010 data at $\sqrt{s} = 7$ TeV [1, 156], this method introduces an additional uncertainty on the SM simulation to cover regions were the SM expectation is insufficient. The amount of uncertainty is determined individually for each simulated process j based on its contribution to previous bins. The total uncertainty fill-up $\sigma_{\text{fill-up}}$ for one bin of an event class distribution is determined by the sum of all individual $\sigma_{\text{fill-up},j}$ from data sets j. Only SM data sets actually contributing to the event class in question are taken into account.

Fig. 7.3 illustrates the procedure used to evaluate the amount and extent of uncertainty fill-up for a single simulation data set. To determine which bins of a distribution should be filled with $\sigma_{\text{fill-up},j}$, it is efficacious to consider the ranges where such an additional contribution is not applicable. For one, kinematic requisites on objects in the event class, i.e., p_{T} thresholds explained in Sec. 9, offer a constraint on the lowest possible filled bin b_{min} of a given event class distribution. It is not sensible to add any type of expectation below this value. The LHC center of mass energy presents a restriction on the highest filled bin. Additionally, since many of the considered distributions exhibit a steeply falling spectrum it is reasonable to truncate the uncertainty filling at a bin b^* where the considered process's contribution is expected to be negligible. Accordingly, twice the highest actually filled bin $b_{\text{max},j}$ with the value $c_{\text{max},j}$ was found to be feasible at the time, sufficiently mitigating the effect of low generated event counts in the observed regions of the considered simulation. Thus b^* is the bin corresponding to the value $2 \cdot c_{\text{max},j}$.

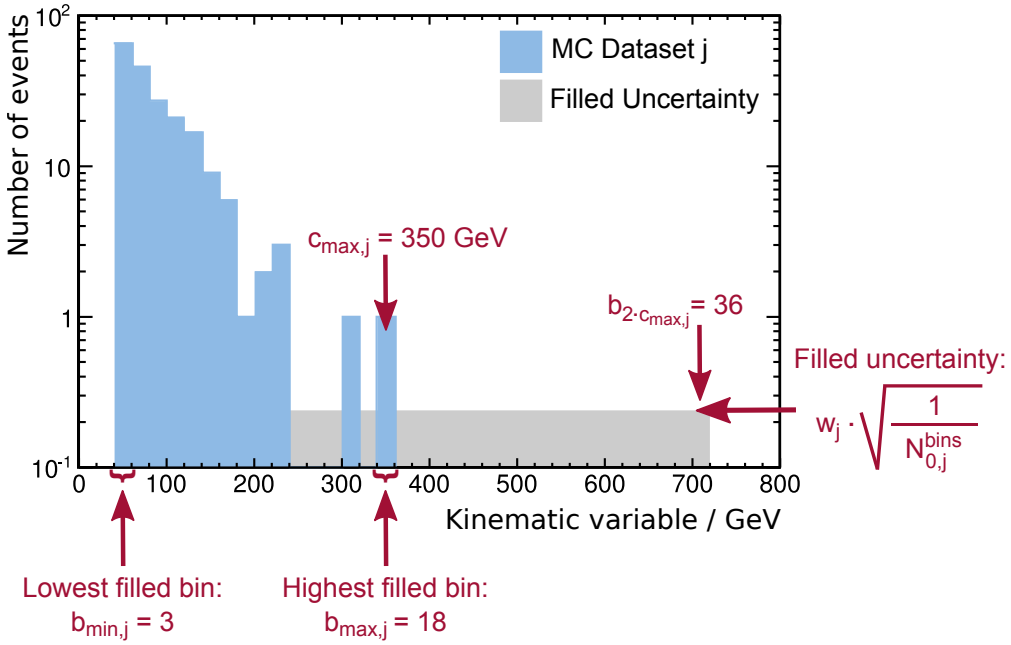


Figure 7.3: Schematic overview of the uncertainty fill-up method for a single MC simulation data set j. The gray area represents the resulting additional uncertainty calculated from the parameters given in red annotations. Here the lowest possible bin b_{min} is also the lowest actually filled bin $b_{\text{min},j} = 3$. At its highest filled bin $b_{\text{max},j} = 18$ the data set has a central value of $c_{\text{max},j} = 350$ GeV. The statistical uncertainty of one weighted event w_j is distributed over a all $N_{0,j}^{\text{bins}} = 21$ empty bins of the data set, up to the bin corresponding to $2 \cdot c_{\text{max},j} = 700$ GeV, in this case bin 36.

To estimate an uncertainty, one weighted event w_j is assumed to be spread out over all empty bins between $b_{\text{min},j}$ and $2 \cdot c_{\text{max},j}$. The corresponding statistical uncertainty is then used to determine the magnitude of $\sigma_{\text{fill-up},j}$ to:

$$\sigma_{\text{fill-up},j} = w_j \cdot \sqrt{\frac{1}{N_{0,j}^{\text{bins}}}}, \tag{7.15}$$

which is then actually filled into all $N_{0,j}^{\text{bins}}$ empty bins. For each data set j the weight w_j is chosen to be the mean event weight $\langle \alpha_j \rangle$ determined from the individual weights α used to scale each generated simulation event to the corresponding event yield, as explained in Sec. 8.2. Alternative choices for w_j are discussed in [2].

Processes produced in a binned fashion over several data sets, e.g., ones split into consecutive intervals of a kinematic generator variable (examples are given in Sec. 8.2.1), are treated as one data set.

SM Simulation Statistics Based Region Veto

An alternative approach, developed for the analysis presented in this thesis, is based on the premise that regions containing an insufficient amount of unweighted SM MC events do not provide enough knowledge of the underlying SM processes and effects to calculate an accurate p-value. Significance values calculated in these regions only represent a rough estimate of the actual p-value which would be given by a sufficient amount of generated SM MC event in the considered region. Consequently, the method described below identifies regions of low generated event count as such and refrains from calculating a p-value there.

The relative statistical uncertainty resulting from the sum of all processes contributing to each region is used to determine if the amount of SM MC events is sufficient. All valid regions must have a value of 60% or less:

$$\frac{\sigma_{\mathrm{SM}}^{\mathrm{stat}}}{N_{\mathrm{SM}}} \leq 0.6, \tag{7.16}$$

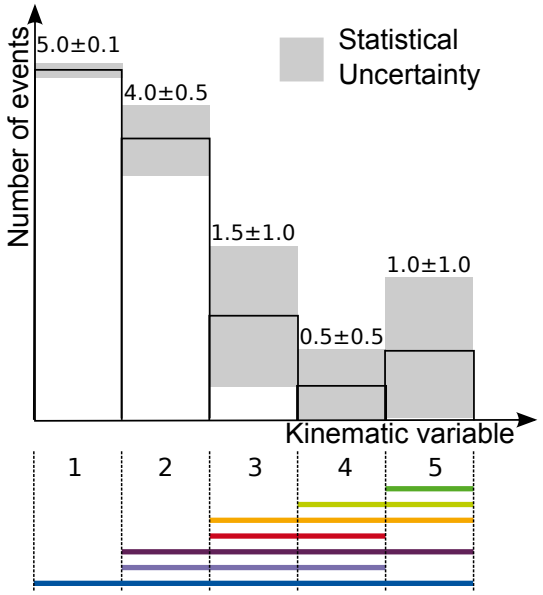
were the statistical uncertainty is given as:

$$\sigma_{\mathrm{SM}}^{\mathrm{stat}} = \sqrt{\sum_{j} w_j^2(\mathrm{event})}, \tag{7.17}$$

i.e., the square root taken from the sum of all weighted events squared, given by all MC processes j contributing to the region. For regions with $N_{\mathrm{SM}} = 0$ the relative statistical uncertainty is taken to be 100%. Applying a small threshold for the relative statistical uncertainty increases the number of disregarded regions, a high threshold value leads to p-values being calculated in more regions, but potentially also ones of insufficient modeling, making them unreliable. The chosen value of 60% represents a balance between both extremes. Studies performed to determine this value furthermore prove the algorithm to be stable against moderate changes of the threshold. A schematic example is shown in Fig. 7.4 to illustrate how the relative statistical uncertainty criterion is applied to the bins of each distribution.

The procedure summarized above identifies regions where contributing SM MC processes are not present in a sufficient amount, but it is not capable of determining regions in which an important SM process is missing all together, i.e., zero events of the process contribute to the region under investigation. A generalized approach to establish, a priori, which processes represent leading contributions in specific kinematic regions of all considered event classes is a difficult task. Each final state represents a separate case presenting varying relative contributions of several SM processes. In some cases a single SM process decisively dominates the entire kinematic rage, e.g., W+jets in an event class such as 1e+MET, in other cases two or three processes are equally relevant to the adequate modeling of the SM. Moreover, the dominant process may change from low to high values of the kinematic variable. Thus, the total number of events given by a specific process in the event class will not necessarily lead to a correct description of the distribution.

Instead, a method is used that compares processes contributing in the region under investigation to all processes contributing to bins surrounding the region. These surrounding bins are combined and referred to as **n**eighbor**h**ood (NH) regions, as illustrated in Fig. 7.5. A width of four bins above and below the region under consideration is chosen to define the NH. Even though a large NH region width ensures that small contributions from individual SM processes are evened out to reduce their influence, the NH region is more likely to extend into areas of the distribution that show a different SM composition than expected in the investigated region.



Example Regions:

Bins in Region	N_{SM} in Region	σ^{stat}_{SM}	$\sigma^{stat}_{SM}/N_{SM}$
5 - 5	1.0	1.00	100% ✗
4 - 5	1.5	1.12	75% ✗
3 - 5	3.0	1.50	50% ✓
3 - 4	2.0	1.12	56% ✓
2 - 5	7.0	1.58	23% ✓
2 - 4	6.0	1.22	20% ✓
1 - 5	12.0	1.87	16% ✓

...

Figure 7.4: Mock example of the relative statistical uncertainty criterion. A region is disregarded (no p-value is calculated) if the total relative statistical uncertainty $\sigma^{stat}_{SM}/N_{SM}$ on the total number of weighted SM events N_{SM} exceeds 60%, which is the case for the first two regions in this example, as listed in the table. Yet, the corresponding bins 4 and 5 are not excluded entirely from the significance evaluation, since they are included in other regions which do pass the criterion and for which a p-value may be calculated.

Contributions from individual SM MC processes are combined from both NH regions and ordered by their weighted event yield in the NH. The region under investigation is considered valid if it contains all leading NH processes. Setting the requirement to all NH processes, not only leading ones, increases the number of regions disregarded because of small contributions of negligible processes in the NH regions. Hence, the requirement is defined as all processes adding up to 95% of the SM yield in the NH regions must be present in the considered region. The process yielding the strongest contribution is considered first and subsequently, the next frequent processes are assessed until the threshold of 95% is reached. The chosen value can be motivated by the smallest total systematic uncertainty on the SM simulation, which is around 5%. Increasing the percentage above 95% would imply a better knowledge of the SM expectation than actually at hand.

A set of benchmark distributions, representing common cases investigated by the MUSiC analysis, shows that variations of both the NH region width and the required percentage of SM yield have only a small influence on the search's final outcome. Also, considering the lower and the upper NH region separately does not change the results.

Although the relative statistical uncertainty and NH region criteria are often not fulfilled in all of the distribution's bins, the prevailing majority of event class distributions provide at least one region that passes both requirements and in which a p-value is calculated. The small number of event class distributions containing only regions failing the criteria are considered separately in the analysis work flow and discussed individually.

Comparison of Uncertainty Fill-Up and Region Veto Method

The uncertainty fill-up method could be refined by determining a more advanced functional form for $\sigma_{\text{fill-up}}$, rather than assuming a constant value. Yet, finding a general parameterization for all final states investigated by MUSiC is a cumbersome task. While some event class distributions show a steep decline in the number of events, others are almost flat or contain only sporadic events. A way would have to be found to address all of these cases consistently and automatable. Nevertheless,

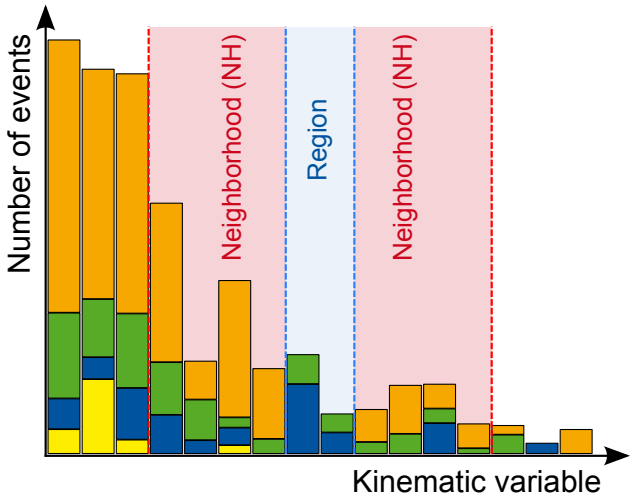


Figure 7.5: Schematic of the NH region definition. The region under consideration is shown in light blue and two NH regions with a width of four bins each are shown in red. Colored bars represent the MC based SM prediction. Individual processes (orange, green, blue, yellow bars) are sorted according to their overall contribution to the distribution. Since the orange process, dominant in the NH regions, is not present in the investigated region, no p-value is calculated.

assigning any type of additional uncertainty has a direct influence on the calculated significance. Varying the input parameters of the uncertainty fill-up method can lead to a change of significance in the region under consideration and in the most extreme cases alter the position of the RoI. The assumptions made to determine the parameter values are rather pragmatic and cannot ensure a viable description of the true SM processes the uncertainty is designed to mimic. Consequently, the calculated p-value can only be seen as a rough estimate of the true significance.

Even though a p-value is not calculated in all regions, when using the MC generated event count based region veto, the far majority of data events is typically enclosed in at least one region where a significance is determined. The uncertainty fill-up method principally allows for a p-value calculation in every filled bin, bearing in mind the limitations mentioned above, but it can introduce a very large systematic uncertainty, at times larger than the mean value in the bin itself. An extreme amount of additional uncertainty in a region blinds the MUSiC search algorithm to excesses.

On first glance it might seem that simply generating a larger amount of SM events would solve all issues directly which are addressed by both methods of low MC generated event count treatment. In reality this would unfortunately only shift the predicament to higher values of the kinematic variables or into event classes representing final states with even lower expected cross sections. Naturally, if a great deal (several orders of magnitude) more events were generated throughout all investigated regions the insufficiency could be handled. Yet, the necessary amount of computing time and manpower is much larger than what can be achieved reasonably. Thus, the MUSiC analysis will always have to handle regions with an insufficiently low number of generated SM events. After close consideration of both approaches presented above, the region veto method is adopted as the default way of handling issues given by a low number of generated events.

7.3.4 Look-Elsewhere Effect

The strongest deviation in each event class distribution is given by the region with the smallest p-value $p_{\text{min}}^{\text{data}}$, the RoI. To interpret the p-value as a true statistical significance, i.e., the probability to see such a deviation in the analyzed events purely from fluctuations of the null-hypothesis, the bias introduced by searching simultaneously in several different regions must be accounted for. The so-called **l**ook-**e**lsewhere **e**ffect (LEE) occurs if the exact search region for which the p-value

is calculated is not fixed a priori [182]. By considering many different regions it becomes more probable to encounter a significant deviation anywhere in the distribution.

In case of fully independent search regions a trial factor correcting the decrease in significance given by the LEE can be calculated from known quantities, leading to a post-trial p-value $\tilde{p}$ of:

$$\tilde{p} = 1 - (1 - p_{\text{min}}^{\text{data}})^{N^{\text{regions}}} \tag{7.18}$$

for a given uncorrected p-value $p_{\text{min}}^{\text{data}}$ and N^{regions} considered regions. However, in the case at hand, the simplified assumption of independent regions does not hold. The same bin may contribute to several regions. Moreover, some of the accounted for systematic uncertainties have an effect on all bins of the kinematic distribution and cannot be treated as uncorrelated. Since the precise relationships are too complex to model analytically, pseudo-experiments based on the SM MC prediction and its systematic uncertainty are performed in each event class distribution [59, 183]. Details of generating pseudo-experiments are given in Sec. 7.3.5.

For each event class distribution, SM-only pseudo-distributions are generated and individually compared to the original SM MC estimate. Following the same methods established to analyze the measured data, a RoI and corresponding p-value $p_{\text{min}}^{\text{pseudo}}$ are determined for every pseudo-distribution. The post-trial p-value is finally estimated as the fraction of the number of pseudo-experiments yielding a p-value at least as significant as $p_{\text{min}}^{\text{data}}$, the p-value retained from the original comparison of measured data with the SM MC prediction, to the total number of pseudo-experiments N^{pseudo}:

$$\tilde{p} = \frac{N^{\text{pseudo}} \text{ with } p_{\text{min}}^{\text{pseudo}} \leq p_{\text{min}}^{\text{data}}}{N^{\text{pseudo}}}. \tag{7.19}$$

This principle is illustrated by an example in Fig. 7.6. The red histogram shows the number of pseudo-experiments yielding a specific p-value, where the negative logarithm of the p-value is given on the x-axis to emphasize strong deviations. Small p-values, i.e., strong deviations appear on the right side of the distribution, insignificant p-values close to 1.0 are given in the left side.

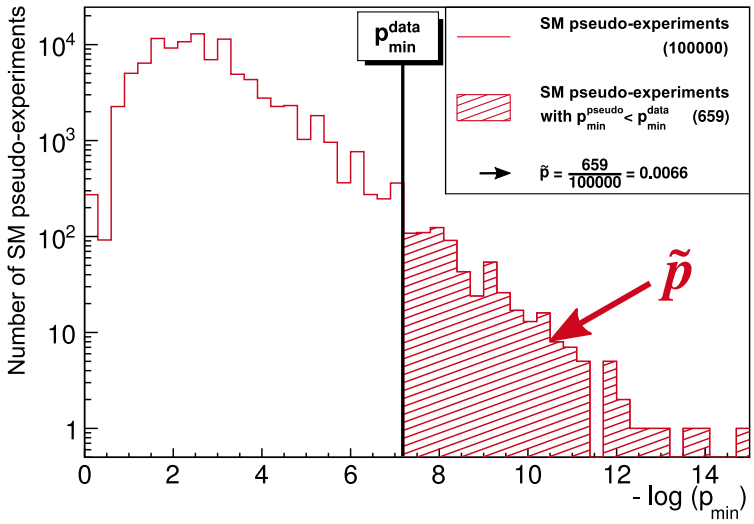


Figure 7.6: Mock example for determining the LEE corrected p-value $\tilde{p}$ via SM-only pseudo-experiments. As given in Eq. (7.19), $\tilde{p}$ is defined as the fraction of pseudo-experiments exhibiting a p-value $p_{\text{min}}^{\text{pseudo}}$ at least as significant as the one found in the data vs. SM expectation comparison ($p_{\text{min}}^{\text{data}}$) and the total number of trials. In this example a total of 10^5 pseudo-experiments are created, of which 659 (hatched area) show deviations at least as significant as $p_{\text{min}}^{\text{data}}$ (black vertical line), resulting in $\tilde{p} = \frac{659}{100000} = 0.0066$.

The number of SM-only pseudo-experiments is based on the desired precision to which a given p-value is corrected. About 10^6 pseudo-data rounds are necessary to expect at least one to contribute to the numerator of Eq. (7.19) for a p-value of $p_{\text{min}}^{\text{data}} = 2.9 \cdot 10^{-7}$, i.e., a $5\,\sigma$ deviation[2]. If the given p-value is larger, i.e., less significant, fewer trials are sufficient to reach the same precision.

Due to the large number of considered distributions, but primarily given by the approach invoked to determine the LEE correctio, the full procedure is rather computing intense with a long runtime. The time perceived by the user is reduced by running the scanning tasks in parallel on a dedicated 64-core server. Further methods of runtime reduction are highlighted in Sec. 7.3.6, which are based on specific details of the statistical analysis itself. Depending on the significance of observed deviations, a typical scan of one event class type and kinematic variable takes 5 to 10 hours.

The method described above covers the multitude of search regions within each individual distribution. In general, considering many different final states simultaneously also leads to a sensitivity reduction. So far, no consistent way has been found to treat this aspect within the MUSiC analysis.

7.3.5 Generating Pseudo-Data

Both the LEE correction of the significance in every event class distribution (see Sec. 7.3.4) and studies performed to assess the analysis' sensitivity (see Ch. 12) rely on pseudo-distributions generated from the given MC expectation including its statistical and systematic uncertainties. Following the procedure described below, pseudo-data is only generated in bins containing an MC expectation (mean value or systematic uncertainty). Pseudo-events are not migrated between bins.

In line with the Gaussian prior used in the p-value calculation (see Eq. (7.13)) a Gaussian distribution is the basis for generating a pseudo-data histogram. Bin by bin, the expectation given by the number of weighted MC events $N_{\text{SM},i}$ and its systematic uncertainty $\sigma_{\text{SM},i}$ per bin i are taken as input for a Gaussian distribution $\mathcal{G}(\theta; \mu, \sigma^2)$. The procedure is performed individually for each systematic uncertainty source $\sigma_{\text{SM},ij}$. Details on the individual contributions j are discussed in Ch. 10. Three cases occur differing in the parameters assumed for the Gaussian distribution:

- Systematic uncertainty is known as absolute number: Obtain random value from $\mathcal{G}(\theta; 0, 1)$ and multiply result with the uncertainty value $\sigma_{\text{SM},ij}$.
- Systematic uncertainty is known as relative number: Obtain random value from $\mathcal{G}(\theta; 0, \sigma_{\text{SM},ij})$ and multiply result with the number of expected events $N_{\text{SM},i}$.
- Systematic uncertainties on the assumed cross section for a MC process, given as a relative value, are treated individually for each process k: Obtain random value from $\mathcal{G}(\theta; 1, \sigma_{\text{SM},ik})$ and multiply result with the number of expected events from this MC process.

Between the bins of one event class distribution as well as between different event classes systematic uncertainties are either taken to be fully correlated or fully uncorrelated. For those which are assumed as correlated, the same generated pseudo-data value is used throughout each pseudo-experiment round. A systematic shift of a given quantity in data, e.g., the integrated luminosity measurement, has the same influence on all final states and bins. Therefore, the same random value is used in the pseudo-experiments. The results from all systematic uncertainty sources are added up, giving a pseudo-event count $N_{\text{SM},i}^{\text{pseudo}}$ in the bin i.

The statistical uncertainty on the number of events in the bin is then included by determining a random value from $\mathcal{G}(\theta; N_{\text{SM},i}^{\text{pseudo}}, \sigma_{\text{SM},i}^{\text{stat}})$, where the third parameter is calculated as given in Eq. (7.17). Finally, the Gaussian random value is taken as the mean μ of a Poisson distribution $\mathcal{P}(N; \mu)$, yielding the bin's total pseudo-event count.

[2] one-sided Gaussian standard deviation; See footnote 4 on p. 52 for details.

7.3.6 Runtime Optimization

In light of reducing the needed computational resources, especially runtime, it is favorable to limit the full search algorithm to regions which contain significant deviations. Evaluating the integral involved in the p-value calculation (Eq. (7.13)) is among the single most runtime consuming operations performed by the search algorithm, amounting to about 200 µs per integral with around $\mathcal{O}(10^9)$ integrals per scan [184]. The integral calculation itself is already implemented in a highly efficient manner, yet reducing the overall number of calculated integrals is a promising alternative.

As a first step, no p-value is calculated in regions that can a priori be determined as not significant. Along these lines, regions containing no observed data events and a SM expectation less than three standard deviations from zero are not regarded. With the observation and expectation being in accordance within the systematic uncertainties, they will never be of particular interest. Moreover, this is a regime where the chosen p-value tends to over-cover [177, 180]. Such regions are marked in dark green in the plot mapping regions to p-values or criteria of the low generated event count treatment discussed in Sec. 7.3.3, for which an example is given in Fig. 11.20b.

Additionally, a study summarized in [184] shows that a scan's total runtime can be reduced by approximately a factor nine if the costly p-value calculation in each region is preceded by a simplistic but more rapid statistical estimator χ:

$$\chi = \frac{|N_{\text{obs}} - N_{\text{SM}}|}{\sqrt{\sigma_{\text{SM}}^2 + N_{\text{SM}}}} \tag{7.20}$$

quantifying the deviation between the number of observed events N_{obs} and the number of events expected from the SM N_{SM} while accounting for the statistical as well as the systematic uncertainty σ_{SM}. A strong deviation leads to a large χ value. Applied as the final step in the scanning algorithm, this estimator is evaluated for each considered region and the full p-value (Eq. (7.13)) is only calculated for the 200 most significant regions chosen by χ. To ensure the actual RoI is among the 200 candidates an additional requirement is introduced if more regions than candidates are available. If a region yielding an excess is fully enclosed in a larger region also showing an excess and only the SM expectation (not the number of observed events) is increased by the additional bins, the larger region is not chosen as a candidate region. Otherwise, this can lead to an over-representation of a specific range of the distribution among the selected candidate regions due to an individual strong deviation (in terms of χ) and can prevent the algorithm from selecting other potentially significant regions (in terms of p-value) from unconnected ranges of the distribution and at worst missing the actual RoI. Such behavior was observed especially in high energy tails of distributions where the overall event count is close to one.

Approximately 10^6 pseudo-data rounds are needed to determine the LEE correction for a p-value corresponding to a deviation of $5\,\sigma$ significance[3], running the full scan algorithm for each round. Yet often only a small fraction of the analyzed distributions show such strong deviations. For less significant deviations it is runtime efficient to terminate the procedure once a sufficient precision has been reached for $\tilde{p}$. Therefore, the number of rounds is restricted to twice the number of trials necessary to achieve 100 (1000) p-values smaller than $p_{\text{min}}^{\text{data}}$ using 10^5 ($2{\cdot}10^6$) rounds.

No LEE correction is applied to event class distributions with a p-value greater than 0.3, corresponding to about $0.5\,\sigma$, as this would only further decrease the significance of deviations that do not warrant a detailed investigation. Moreover, the global result described in Sec. 7.3.7 summarizes all deviations under $1\,\sigma$ into a single bin.

[3] one-sided Gaussian standard deviation; See footnote 4 on p. 52 for details.

All runtime optimizations are introduced under the requisite that they do not significantly alter the analysis' outcome, i.e., p and $\tilde{p}$-values of individual event class distributions and the overall distribution of $\tilde{p}$-values. Consequently, the simplified calculation using the χ estimator is applied while scanning SM-only induced pseudo-data as used for the $\tilde{p}$ calculation (Sec. 7.3.4). The method is not used for determining the RoI in scans performed on SM expectation vs. observed data, as a runtime reduction is not necessary for only a single round.

A further concept for a runtime reduction, especially useful for sensitivity studies, is outlined in App. A. This procedure has not been included in the MUSiC framework yet. Moreover, it is worthwhile to investigate introducing a lookup table containing significance values for possible combinations of N_{data}, N_{SM} and σ_{SM}. Retrieving the p-values from the table, relying on interpolation between points in the defined grid, is more rapid than the numerical p-value calculation. A study of this possible extension of the MUSiC analysis is currently undergoing [185].

7.3.7 Global Overview of Results

Ultimately, the scan algorithm as described above provides results in the form of hundreds of individual event class distributions, each allocated with a corresponding $\tilde{p}$-value. For new BSM phenomena that cause strong deviations in only few final states it may suffice to focus on the most significant classes to understand its nature. Yet, other models predict signals which spread out over a multitude of final states, resulting in only a moderately increased significance in each individual event class. It would be beneficial to find a consistent way to condense all results into a single measure of significance, corrected for a global LEE over all final states. A numerical or analytical method for combining all $\tilde{p}$-values suffers from difficulties such as considering all correlations between final state topologies and uncertainties adequately. Alternatively, a graphical approach is used to represent the overall outcome of the search, as described in the following.

All $\tilde{p}$-values obtained from the data vs. SM MC comparison are inserted into a histogram, depicted by black markers as shown in the example in Fig. 7.7. The negative decadic logarithm is chosen for the axis of abscissas to emphasize strong deviations. To reduce correlations results from each event class type and kinematic variable are summarized in separate distributions of $\tilde{p}$ values. Event classes for which no $\tilde{p}$-value was calculated, because of the low generated event count treatment or runtime optimizations, are not included in this global overview.

A numerical estimate for the distribution of significances expected from the SM-only hypothesis can be obtained from the pseudo-experiments generated for determining the LEE correction in each event class (see Sec. 7.3.4). The p-value $p_{\text{min}}^{\text{pseudo}}$ from each round is set into relation to p-values from all other rounds analogous to Eq. (7.19), returning a $\tilde{p}$-value representing the strongest deviation found because of pure fluctuations of the SM MC expectation. In Fig. 7.7 this is shown by the light blue line labeled "MC vs. MC", which represents the histogram of mean $\tilde{p}$-values from all pseudo-experiments, i.e., normalized to the number of rounds.

The p-value as given in Eq. (7.13) represents a proper probability estimate and as such it and its derived quantity $\tilde{p}$ are distributed uniformly [176]. As a result, a constant value would occur if a linear x-axis is chosen in Fig. 7.7. Neglecting correlations between event classes, the uniformity of $\tilde{p}$ allows an analytical expression to be found for the expected amount of event class distributions contributing to each bin of the distribution of $\tilde{p}$-values based solely on the total number of event classes and the respective bin edge values of the distribution of $\tilde{p}$-values. The mean of this expectation is shown by a dashed green line in the global distribution of $\tilde{p}$-values (Fig. 7.7). Serving as a comparison to the observation, the median expectation is represented by a black dotted line accompanied by its 68% ($\pm 1\,\sigma$) and 95% ($\pm 2\,\sigma$) uncertainty bands given at the respective two-sided binomial confidence level accounting for statistical fluctuations. For bins at high abscissa values the

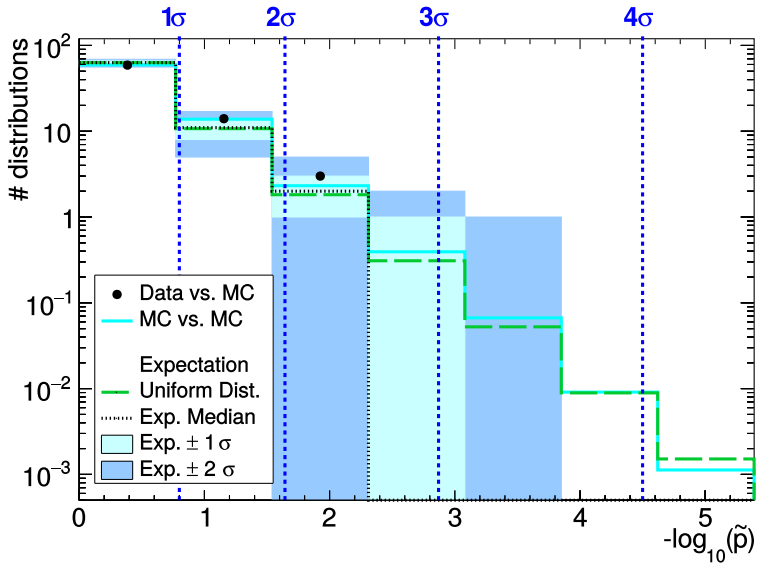


Figure 7.7: Mock example of the graphical global representation of MUSiC analysis results. Black markers refer to the distribution of $\tilde{p}$-values obtained from comparing data and SM expectation in each event class distribution, whereas $\tilde{p}$-values from SM-only pseudo-data vs. SM expectation are given by a solid light blue line. The mean analytical expectation is shown by a dashed green line and its median is represented by a black dotted line accompanied by uncertainty bands given at the respective two-sided confidence level.

binomial distribution tends towards a Poisson distribution. Accounting for correlations between event classes has no influence on the mean or median values, but only on the uncertainty bands.

The assumption of uniformity only holds if the corresponding estimator has exact coverage properties, which is not completely true for $\tilde{p}$, due to necessary runtime optimizations (see Sec. 7.3.6). A good measure for determining this influence can be achieved by comparing the mean analytical expectation with the numerical description, where the latter undergoes the same runtime optimizations as the observation. Nevertheless, the analytical form is beneficial as it is not influenced by statistical limitations of the pseudo-data procedure for small $\tilde{p}$-values.

Offering a more intuitive measure of significance, $\tilde{p}$-values are translated in terms of one-sided[4] Gaussian standard deviations σ which are indicated by blue dashed vertical lines in the distribution of $\tilde{p}$-values. In the high-energy physics community a deviation of at least $5\,\sigma$ is typically required to claim the discovery of a new effect, corresponding to a significance of $2.9{\cdot}10^{-7}$ [59, 183].

[4] A one-sided interpretation is chosen to quantify the observed deviation. The p-value as given in Eq. (7.13) differentiates between excesses and deficits. For a given observation a stronger deviation is defined by the one-tailed probability of the distributed number of events individually for each case. For an excess only more events and for a deficit only fewer events would yield a stronger deviation.

Chapter 8
Analysis Input

Analysis methods established in Ch. 7 are performed on measured data taken by the CMS experiment and simulated events modeling SM and BSM processes. Both types of input are explained in detail in the following chapter and are specific for the considered data-taking period in 2012, during which the LHC delivered collisions at a center of mass energy of $\sqrt{s} = 8$ TeV.

8.1 Collision Data

During 2012 the CMS detector recorded 21.8 fb^{-1} worth of proton-proton collision data at 8TeV [85]. According to the central certification of the Data Quality Monitoring group all CMS subsystems were fully functional and running for 19.7 fb^{-1} of integrated luminosity. Runs and luminosity sections appropriate for physics analysis are given in a dedicated file[1] corresponding to the so-called `22Jan2013ReReco` reconstruction campaign performed with CMSSW versions `5_3_X`. The precise detector and run conditions during data-taking are encoded in a global tag named `FT_53_V21_AN6::All` [188].

Information from the HLTs (see Sec. 5.2 for details) is used to group reconstructed events into so-called *data streams* during the individual data-taking periods (Run I A, B, C, D). Each data stream consists of events passing at least one of the triggers given in the respective trigger group. Tab. B.1 lists all data streams considered for the analysis presented in this thesis, focusing on electron and muon triggered events. Details of the applied trigger selection are given in Sec. 9.1.3. Integrated luminosity values mentioned in Tab. B.1 are determined with the pixel cluster counting method (see Sec. 5.3) as implemented in the official CMS tool `pixelLumiCalc` [189]. In total $7.1{\cdot}10^8$ data events are analyzed by the MUSiC data pre-processing step as decribed in Sec. 7.1.

Since the same event can be recovered by different triggers, MUSiC applies a cross-cleaning to remove duplicate events. The stream requiring a single muon in the event is taken as a starting point and all such events are removed from the other data streams. Furthermore, events found in the single electron stream are removed from those requiring two leptons and events from the double muon stream are vetoed in the double electron stream.

8.2 Simulation Data sets

Though a conclusive description of SM processes is given by quantum gauge field equations introduced in Ch. 2, this approach cannot provide a prediction of all phenomena relevant for collider experiments. Instead, the expectation of specific observables given from theory models is obtained by an elaborate simulation procedure consisting of consecutive steps, factorizing the complex task into components that can be handled individually with the necessary precision [190, 191]:

- **Parton distribution function:** Protons are composite objects made up of several strongly interacting particles called *partons*, which actually participate in the collision and are the basis for the hard scattering. Probability density functions can be derived to model the partons' momentum distributions, as explained in more detail in Sec. 8.2.3.
- **Hard-scattering process:** The primary process occurring during the collision in the *inelastic* or *hard* parton-parton scattering is evaluated using Feynman diagram rules to calculate the differential cross section. In Sec. 8.2.1 and Sec. 8.2.2 details are given of which SM and BSM processes are considered in the simulation used for this analysis.

[1] `Cert_190456-208686_8TeV_22Jan2013ReReco_Collisions12_JSON.txt` [186, 187]

Common **L**eading **O**rder (LO) generators like PYTHIA6 [130] and MADGRAPH [192] consider processes and derive cross sections at tree-level. In doing so, PYTHIA6 can account for a maximum of two incoming and four outgoing particles, whereas MADGRAPH is able to deliver $2 \to n$ processes. While PYTHIA6 considers the radiation of soft gluons and photons in a later step, the *parton shower*, the MADGRAPH generator can include them directly during calculation of the differential cross section. Generators like POWHEG [193–195] and SHERPA [196] are able to calculate contributions in higher order in QCD perturbation theory, introducing couplings with an additional factor α_s, resulting in cross sections at **N**ext-to-**L**eading **O**rder (NLO). Often a k-factor acquired through the ratio of the higher order to the LO cross section can be used to scale a process given by an LO generator to a higher order estimate.

Numerical **M**onte **C**arlo (MC) methods, following the procedure of random sampling, are applied to carry out an integration of the differential cross section yielding the process' total cross section. As a consequence, the final outcome of the full simulation procedure is often referred to as an MC data set. The result of this first step is a number of events that can be addressed individually by the following stages of the generation procedure.

- **Particle decays parton shower:** The decay of unstable particles is addressed through dedicated methods, where, e.g., TAUOLA [197] is used for decays of τ leptons.

  As mentioned above, soft (low p_{T}) gluons and photons can be radiated from initial and final state particles alike. Especially the color charged gluons give rise to cascades of additional particles, contributing to the event's final state. These *parton showers* are modeled with PYTHIA6, following an evolution from high transverse momenta, close to the scale of the hard scattering, down to low momenta where non-perturbative effects of confinement are dominant. The parton shower approach can be used to describe complex processes involving a larger number of **I**nitial **S**tate **R**adiation (ISR) and **F**inal **S**tate **R**adiation (FSR) particles.

  In addition, the so-called *underlying event* [198] includes the behavior of beam-beam remnants, i.e., partons of both protons left over after the hard interaction, and multiple particle interactions, i.e., additional semi-hard scattering of other partons during the same proton-proton interaction. All particles resulting from the underlying event are also put into the context of the full event description during the parton showering process.

  All particles initiated during this step must be matched to the outcome of the hard scattering. Thus momenta and color charges are adjusted to deliver a consistent picture throughout the entire event generation process.

- **Fragmentation and Hadronization:** Further development of radiated gluons and their secondary color charged particles is also treated with PYTHIA6. During the fragmentation process, the distance between two partons increases until sufficient potential energy is built up by the gluon field for new quark-antiquark pairs to emerge. Within the PYTHIA6 model, hadronization is pursued until only color neutral hadrons on their mass-shell are left. Finally, the hadrons are decayed into stable particles including leptons and photons.

- **Event pileup:** The large amount of protons in each bunch makes it likely for more than one proton-proton collision to take place during a single bunch crossing. Modeling of this event PU is discussed separately in Sec. 8.2.4.

- **Detector response:** Simulating the propagation of all final state objects through the CMS detector material is performed individually for each detector component. Furthermore, the component's response and read-out is modeled. Details on this are given in Sec. 8.2.5.

Not all aspects of the simulation can be derived from first principles. Especially for modeling of the parton shower, underlying event or hadronization, PYTHIA6 therefore resorts to so-called *tunes* [199]. The generator's parameters are optimized simultaneously to include correlations, where measured data from previous collider experiments and also the LHC is taken as input. Simulation data sets produced for the analyses of data taken in 2012 make use of the Z2* tune [200].

Fully data-driven techniques are generally not applied within the MUSiC strategy as all regions are treated as potential signal regions[2]. Thus, the SM expectation is derived foremost from MC simulation. Moreover, since event classes are constructed on-the-fly from the analyzed data streams and simulation data sets, a single consistent group of data sets describing the SM processes in all final states must be found. Different data sets for the same physics process, e.g., relying on different generators, cannot be used simultaneously.

8.2.1 Standard Model Simulation

The complex description of SM effects resulting from the proton-proton collisions is pieced together from simulation data sets of individual physics processes appearing at LHC energies. Interactions from pure soft scattering can be disregarded because of the analysis requirement of at least one isolated high p_{T} lepton in every event. An overview of considered SM simulation data sets is given in Tab. 8.1, with more details listed in App. B.3. To enhance the number of generated MC simulation events in the tails of specific kinematic variables, additional data sets binned in various event properties are often produced for the main SM processes. Duplicate events appearing in more than one data set must be removed when combining different data sets of the same SM process.

From the MUSiC analysis point of view, processes produced with MadGraph are often advantageous, as they frequently include additional jets on matrix element level. Regarding the MadGraph data sets listed in Tab. 8.1, one additional jet was generated for tri-boson processes, two additional jets are included in di-boson, $\mathrm{t\bar{t}}$+V[3] and WWW processes, whereas up to three jets are given in data sets of Drell-Yan and W+jets processes.

As far as possible, MC simulation data sets from official CMS production campaigns are used. Privately produced data sets are only included if no alternative from the official production is available. Thus, the MUSiC analysis serves as a validation for SM simulations widely relied on by the CMS Collaboration. More details on the individual SM MC processes are given in the following, beginning with those delivering the overall largest contributions in all considered event classes.

W Boson Production

At the LHC, the production of a W boson is one of the most frequent electro-weak processes. Gluons radiated from the initial state quarks lead to associated jets in the final state along side the isolated charged lepton and MET from the leptonic decay of the W boson itself. Hadronic decays of the W boson are not accounted for in this analysis as these pure jet events are negligible with respect to large contributions from QCD multi-jet processes.

Event generation is performed at LO with MadGraph and scaled to an **N**ext-to-**N**ext-to-**L**eading **O**rder (NNLO) cross section of 12 200 pb [201] calculated with FEWZ 3.1 [202]. Apart from the main inclusively generated W+jets data sets, two types of additional data sets are used to ensure a sufficiently high event count in the tail of kinematic distributions. The interplay of all three data set types is shown in Fig. 8.1.

The scalar sum of transverse energy of all jets H_{T} is chosen to bin a second set of data sets contributing mainly to the high energy tails. The minimum requirement of $H_{\mathrm{T}} = 150$ GeV is inverted to remove overlap in the main W+jets data sets. For additional W+jets simulations binned in the W boson mass the MUSiC analysis must resort to privately produced data sets[4]. To

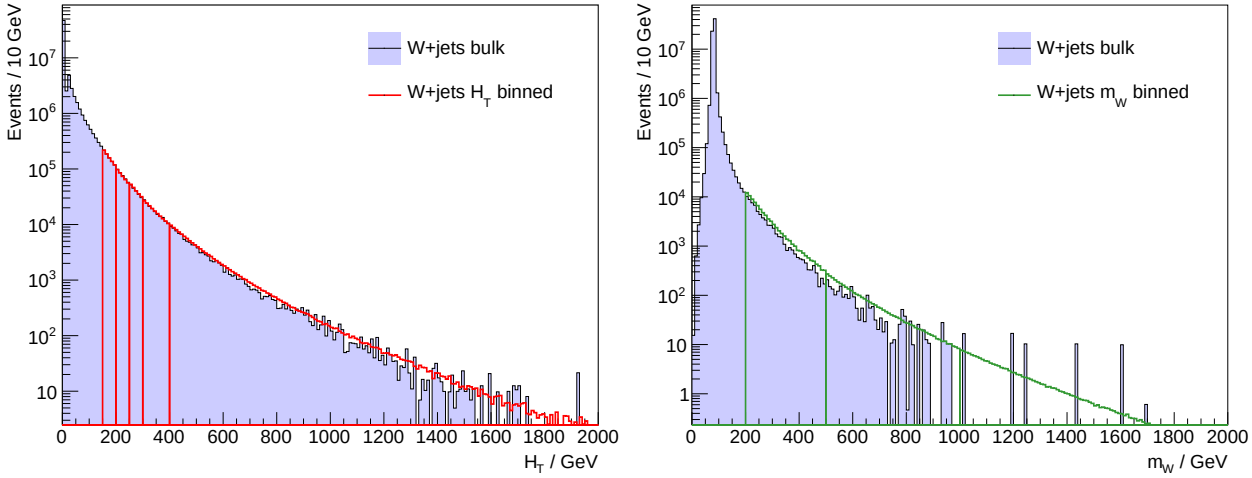
[2] First development of methods based on studies before data-taking are summarized in [4]. An approach is made by inverting lepton identification and isolation requirements. Nevertheless, a control region at low p_{T} values is introduced, which can then no longer serve as a search region.

[3] Here, V refers to a vector-boson γ, Z^0 or $W^{\pm}$.

[4] These data sets were produced by Klaas Padeken.

Process	Details	Generator	σ Order	N_{events}
W+Jets		MADGRAPH	NNLO	$6.0 \cdot 10^7$
	$M_{l^\pm\nu} > 200$ GeV	MADGRAPH	NNLO	$1.2 \cdot 10^6$
	$H_T > 150$ GeV	MADGRAPH	NNLO	$4.3 \cdot 10^7$
	Wbb $\to l^\pm\nu$bb	MADGRAPH	NLO	$2.0 \cdot 10^7$
Drell-Yan	$M_{l^\pm l^\pm} > 10$ GeV	MADGRAPH	NNLO	$6.0 \cdot 10^7$
	$p_T(Z) > 50$ GeV	MADGRAPH	LO	$3.8 \cdot 10^8$
	$M_{l^\pm l^\pm} > 200$ GeV	POWHEG	NNLO	$2.2 \cdot 10^6$
	Z+Jets $\to \nu\nu$	MADGRAPH	LO	$3.0 \cdot 10^7$
γ+Jets		PYTHIA6	NLO	$1.6 \cdot 10^6$
Multi-Boson	$\gamma\gamma$	PYTHIA6	LO	$7.7 \cdot 10^6$
	WW	PYTHIA6	NLO	$1.0 \cdot 10^7$
	WW via DPS	PYTHIA6	LO	$8.0 \cdot 10^5$
	$W^\pm W^\pm$	MADGRAPH	LO	$1.8 \cdot 10^5$
	gg $\to$ WW	gg2VV	NLO	$1.0 \cdot 10^5$
	ZZ $\to 4l^\pm$	POWHEG	NLO	$1.5 \cdot 10^7$
	ZZ $\to 2l^\pm$2q, $2l^\pm 2\nu$, 2q2ν	MADGRAPH	LO	$2.7 \cdot 10^6$
	gg $\to$ ZZ	gg2VV	NLO	$9.0 \cdot 10^5$
	WZ	MADGRAPH	LO	$7.9 \cdot 10^6$
	Vγ	MADGRAPH	NLO	$1.0 \cdot 10^7$
	WWW	MADGRAPH	NLO	$2.0 \cdot 10^5$
	ZZZ	MADGRAPH	NLO	$2.0 \cdot 10^5$
	WWZ	MADGRAPH	NLO	$2.0 \cdot 10^5$
	WZZ	MADGRAPH	NLO	$2.0 \cdot 10^5$
	WWγ	MADGRAPH	LO	$5.0 \cdot 10^5$
	W$\gamma\gamma$	MADGRAPH	LO	$2.0 \cdot 10^6$
QCD Multi-Jet		PYTHIA6	LO	$4.5 \cdot 10^8$
$t\bar{t}$		POWHEG	NNLO	$2.6 \cdot 10^7$
	$t\bar{t}$ ($M_{t\bar{t}} > 700$ GeV)	POWHEG	NNLO	$4.0 \cdot 10^6$
	$t\bar{t}\gamma$	MADGRAPH	LO	$1.0 \cdot 10^6$
	$t\bar{t}$W	MADGRAPH	NLO	$1.0 \cdot 10^5$
	$t\bar{t}$Z	MADGRAPH	NLO	$2.0 \cdot 10^5$
	$t\bar{t}$WW	MADGRAPH	LO	$2.0 \cdot 10^5$
$t\bar{t}t\bar{t}$		MADGRAPH	NLO	$9.0 \cdot 10^4$
Single Top		POWHEG	NNLL	$5.2 \cdot 10^6$
Higgs	gg $\to$ ZZ	MiNLO	NLO	$8.0 \cdot 10^5$
	gg $\to$ bb, $\gamma\gamma$, $\tau\tau$, Zγ, WW	POWHEG	NLO	$2.6 \cdot 10^6$
	VBF $\to$ bb, $\gamma\gamma$, $\tau\tau$, Zγ, VV	POWHEG	NLO	$2.5 \cdot 10^6$
	WH $\to \mu\mu$, bb	POWHEG	NLO	$1.1 \cdot 10^6$
	ZH $\to \mu\mu$, bb	POWHEG	NLO	$2.9 \cdot 10^6$
	TTH $\to$ bb, $\gamma\gamma$, Zγ	PYTHIA6	NLO	$1.2 \cdot 10^6$
	VH $\to \gamma\gamma$, Zγ	PYTHIA6	NLO	$1.1 \cdot 10^6$
	VH, TTH $\to \tau\tau$, VV	PYTHIA6	NLO	$9.0 \cdot 10^5$
Upsilon	Υ(1S)/Υ(2S)/Υ(3S) $\to \mu\mu$	EVTGEN	LO	$5.0 \cdot 10^6$

Table 8.1: Summary of SM MC simulation data sets used in this analysis. A detailed overview with unabridged data set names and corresponding cross section estimates can be found in Tab. B.3.



(a) **(b)**

Figure 8.1: Comparison of main W+jets MC data sets (bulk: blue histogram; index 13 and 14 in Tab. B.3) and additional data sets binned in kinematic variables. Fig. 8.1a includes data sets binned in H_{T} as red lines (index 18-22 in Tab. B.3), whereas Fig. 8.1b shows ones binned in the W boson mass in green (index 16-17 in Tab. B.3). Kinematic cuts are introduced to address the overlap of events seen above.

speed up the simulation process the FastSIM package [203] is used for modeling the CMS detector response (see Sec. 8.2.5 for details). An upper threshold of 200 GeV is applied on the invariant mass of the $l^{\pm}\nu$ system in the remaining W+jets data sets. A separate data set is included to simulate W boson production with two associated bottom quarks. The MadGraph shape is scaled to NLO following studies performed in [204] based on calculations from [205]. The corresponding cross section is about two orders of magnitude smaller than that of the inclusive W+jets production.

Drell-Yan Process

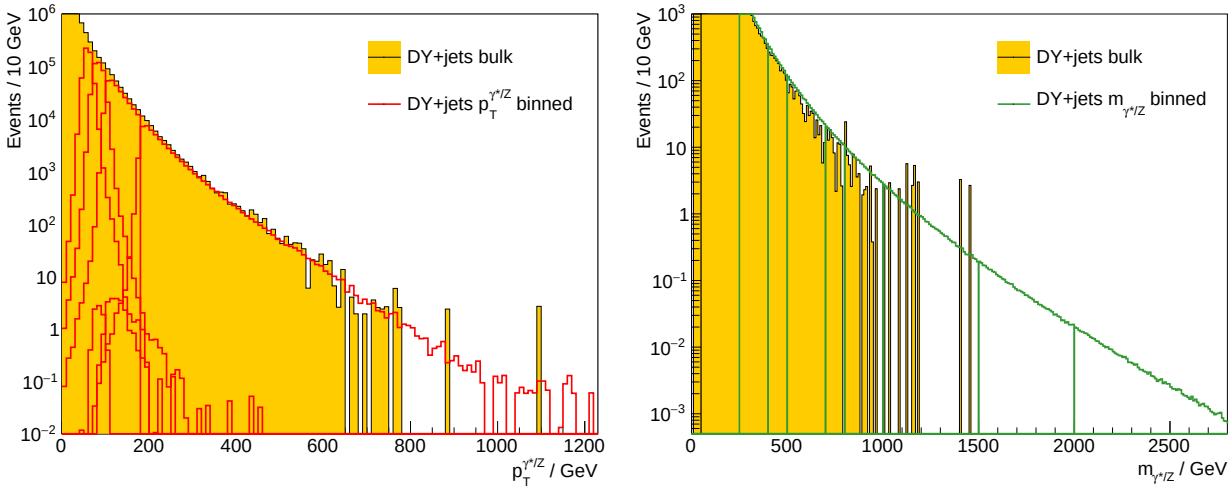
Similar to W+jets production, different data sets are combined to model the **D**rell-**Y**an (DY) process, the production of a virtual photon or Z boson decaying into two oppositely charged same flavor leptons [206]. The bulk of the distribution around the Z boson peak is addressed by events generated at LO with MadGraph and scaled to an NNLO cross section of 1177.3 pb [201] calculated for events with $M_{l^{\pm}l^{\pm}} > 50$ GeV with fewz 3.1 [202]. During generation, a lower threshold of 10 GeV is imposed on the γ^*/Z mass in these events, which is also the case for data sets of other processes as mentioned in the following section. Thus, an overall minimum threshold of 10 GeV is applied to the MUSiC $M_{(\mathrm{T})}$ variable on reconstruction level to measured data and simulated events alike. Generator level information, based on the particles' four vectors before the pythia6 showering, is not available so the exact requirement cannot be transfered to the analysis.

Data sets generated with MadGraph and binned in the γ^*/Z transverse momentum from 50 GeV upwards are included. To remove overlapping events from the main DY data sets a threshold of $p_{\mathrm{T}}(\gamma^*/Z) < 50$ GeV is applied. Since the $p_{\mathrm{T}}(\gamma^*/Z)$ information used during event generation is accessed before the pythia6 showering, jet matching and event balancing, the corresponding histograms shown with information after the pythia6 step in Fig. 8.2a do not have sharp thresholds at the corresponding $p_{\mathrm{T}}(\gamma^*/Z)$ values.

Moreover, to drastically increase the generated (unweighted) number of events found in the tails of di-lepton distributions, data sets binned in the γ^*/Z mass are included from 250 GeV upwards, as shown in Fig. 8.2b. A corresponding threshold is applied on the invariant mass of the lepton

pair in the remaining DY data sets. The high mass data sets are generated at NLO with POWHEG and scaled to an NNLO cross sections using a flat k-factor of 1.024 given in [207, 208].

The process of a virtual photon or Z boson decaying into two neutrinos is addressed with a separate data set generated with MADGRAPH for $p_{\mathrm{T}}(\gamma^*/Z) \geq 70$ GeV.



(a) **(b)**

Figure 8.2: Comparison of main DY+jets MC data sets (bulk, yellow histogram; index 24 and 25 in Tab. B.3) and additional data sets binned in kinematic variables. Fig. 8.2a shows data sets binned in the γ^*/Z transverse momentum in red lines (index 26-31 in Tab. B.3), whereas Fig. 8.2b includes data sets binned in the γ^*/Z mass in green (index 32-47 in Tab. B.3). The three types of data sets are combined to model the full DY+jets process, where kinematic cuts are introduced to address the overlap of events seen in the distributions above.

γ+Jets

The production of hard photons in association with jet is simulated with PYTHIA6. The process is divided into bins of the photon transverse momentum. Though this analysis requires each event to contain at least one isolated charged lepton (see Sec. 9.1) the γ+jets process may contribute to various event classes, predominantly ones in which the photon is misidentified as an electron. Especially events containing high energetic photons pass the analysis selection. Following the strategy documented in [207], a data-based k-factor of 1.3 is applied to scale all γ+jets data sets to NLO. The estimate is given from [209], where an NLO k-factor is calculated with DIPHOX [210] and compared to a γ+jets spectrum taken during 2011 at $\sqrt{s} = 7$ GeV.

Multi-Boson Production

A cornucopia of processes involving two or even three vector bosons is also accounted for by the MUSiC analysis. Exhibiting cross sections throughout the range of $\mathcal{O}(0.001 - 10\ \mathrm{pb})$, proper modeling of these processes is especially crucial for event classes with a high object count, specifically a high lepton multiplicity.

Pairs of photons are produced through the *born* level process of quark-antiquark annihilation at LO in QCD and additionally in the so called *box* process corresponding to NNLO in α_s, where a quark-loop is induced from two initial state gluons. Though the cross section of the box process is smaller, the large number of initial state gluons over quarks at the LHC makes it an important SM contribution for this search. Both processes are generated with PYTHIA6 and are binned in the photon transverse momentum. An alternative data set produced with the NLO SHERPA generator

[196] was also investigated, but did not prove advantageous over the $\gamma\gamma$ data sets generated with pythia6.

Several different generators are used to describe the production of two heavy vector bosons. The main part of the WW process is simulated with pythia6 at LO and scaled to NLO calculations from **M**onte **C**arlo for **F**e**M**tobarn processes (MCFM) [211] given on [201]. pythia6 is also used for the rare process of producing two W bosons via **D**ouble **P**arton **S**cattering (DPS), where two hard interactions take place during the same bunch crossing [212]. Moreover, the similarly rare production of two same-sign W bosons is modeled at LO with MadGraph and is especially important when considering event classes aware of the net lepton charge in the event. A dedicated generator, gg2vv [213], is invoked to cover the production of two W bosons via gluon-gluon fusion. Applying a k-factor of 1.4, as given in [214], the cross section is given to NLO in QCD.

A group of data sets produced at NLO with powheg are used to describe the full-leptonic decay of pair produced Z bosons, which are divided into the different charged lepton flavor combinations. All semi-leptonic ZZ decays and ones involving uncharged leptons (l^+l^-qq, $l^+l^-\nu\nu$, $qq\nu\nu$) are given at LO from MadGraph, where two additional jets are included on matrix element level. Once more, the gg2vv generator is relied on for NLO calculations of Z pair production via gluon-gluon fusion [215].

The production of a photon in association with a heavy boson, as well as all combinations of three electro-weak bosons are simulated with MadGraph at LO. One additional jet is simulated on matrix element level, with the exception of the WWW process, where two jets are considered. Following calculations from MCFM given in [216], a k-factor of about 1.2 is applied to data sets from both $W\gamma$ and $Z\gamma$ processes. For the cases involving three massive bosons, NLO cross section estimates are available from the aMC@NLO package [217] as given on [201].

QCD Multi-Jet Production

At the LHC the QCD multi-jet process, the production of quark or gluon jets, appears most frequently. Resulting from its large total cross section QCD multi-jet production represents a difficult process to simulate and a very high number of events must be generated to allow a statistically representative result. Furthermore, the process itself must be well modeled, including the sensitivity to higher order α_s corrections, since slight imprecisions on selection efficiencies will lead to a strong effect given through the large number of events.

This analysis relies on a composite prediction of QCD multi-jet production generated at LO with pythia6. All quark types apart from the top quark are considered in the simulation. To enhance the production of events containing high energetic leptons or bottom quark jets, an array of dedicated data sets enriched in specific topologies are used alongside the generic multi-jet data sets. The full list of data sets can be found in Tab. B.3 from index 126 to 184, with LO cross sections ranging from 0.0018 pb to $5 \cdot 10^{10}$ pb as given by [218]. A combination of data set relying on the following generator level filters is used in this analysis:

- `MuEnrichedPt5,15`: Every event contains at least one muon with $p_{\mathrm{T}} \geq 5, 15$ GeV;
- `EMEnriched`: Every event contains at least one isolated electron, photon, charged kaon or pion;
- `BCtoE`: Every event contains at least one electron originating from a jet produced by a bottom or charm quark;
- `bEnriched`: Every event contains at least one bottom quark jet.

The overlap of identical event types is addressed by applying the inverse of the generator level filters used for the enriching on data sets where the specific filter was not run.

Top Quark Production

Out of processes involving top quarks at the LHC, the top quark pair production $t\bar{t}$ via the strong interaction is the most frequent with an NNLO cross section of 245.8 pb at $\sqrt{s} = 8$ TeV [219]. Unlike all other quarks, the heavy and short-lived top quark decays before the hadronization process can set in, where the weak decay into a W boson and a bottom quark is given at a branching ratio of over 95% [59]. Since each of the two W bosons produced through $t\bar{t}$ can decay either hadronically or leptonically, a wide range of final states contain contributions from this process. The POWHEG NLO generator is used to model $t\bar{t}$, including a group of dedicated high top quark mass ($M_{t\bar{t}} > 700$ GeV) data sets as used in [207]. All data sets are scaled to NNLO cross section estimates and the overlap of events is removed between the main and high mass data sets.

The $t\bar{t}$ process can also appear in associated production with up to two vector bosons: $t\bar{t}+\gamma$, $t\bar{t}+W$, $t\bar{t}+Z$, $t\bar{t}+WW$, listed in descending order of the respective total cross section given at 1.4 pb (LO) and lower. All four processes are generated with MADGRAPH including two additional jets on matrix element level. Higher order QCD calculations are available for $t\bar{t}+W$ and $t\bar{t}+Z$ and k-factors to corresponding NLO cross sections are used [220, 221]. Moreover, radiation of a gluon can lead to the simultaneous production of two top quark pairs. This $t\bar{t}t\bar{t}$ process is rather rare with a cross section of only 0.7 fb, which is produced at LO with MADGRAPH and scaled to an NLO aMC@NLO estimate [201].

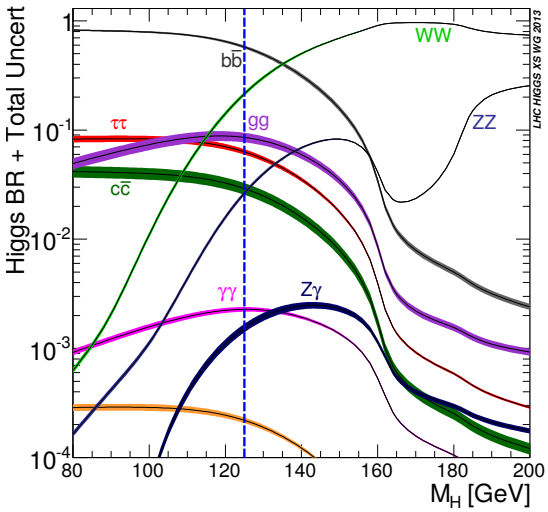
Produced via the strong interaction, signatures of single top quarks also appear in the LHC collisions. The data sets, generated at NLO with POWHEG and scaled to an approximate NNLO (also referred to as **N**ext-to-**N**ext-to-**L**eading-**L**ogarithmic (NNLL)) cross section calculation given by [222], are divided into the three single top quark production channels, the s, t and tW-channel for top quark and top antiquarks, respectively. With a total cross section of 87.2 pb at $\sqrt{s} = 8$ TeV the t-channel is the dominant of the trio, where a top quark is produced from a virtual W boson together with a further quark.

Higgs Boson Production

Following the discovery of an SM-like Higgs boson (see Sec. 2.3.2), a simulation of the SM is only complete if processes including the novel boson are also accounted for. Mirroring the manifold production and decay possibilities, a large variety of data sets modeling the SM Higgs ($m_H = 125$ GeV) are combined. All five production channels are taken into consideration, given in decreasing order of production cross section at the LHC ($\sqrt{s} = 8$ TeV) as, gluon-gluon fusion $pp \to H$, vector-boson fusion $pp \to qqH$, Higgs-strahlung $pp \to WH$ and $pp \to ZH$, as well as $t\bar{t}$ fusion $pp \to t\bar{t}H$.

An overview of the nine Higgs boson decay channels is given in Fig. 8.3. To economize computing resources and time, not all possibilities are considered in the provided MC simulation data sets. Although decays to gluons provide the third highest branching ratio for a 125 GeV Higgs boson, it is not possible to disambiguate its signature from the vast amount of jets produced in QCD multi-jet events or through PU. A similar argument is valid for the decay into charm quarks. Large, irreducible contributions from DY also deluge Higgs boson signatures from decays into opposite sign muon pairs from events produced by gluon-gluon and vector-boson fusion. Di-muon events produced by $t\bar{t}$ fusion have the smallest cross section times branching ratio of all Higgs boson signatures and are not accounted for in the simulation.

A dedicated procedure MiNLO [224] is used for $gg \to ZZ$ and interfaced in the POWHEG BOX framework. Matrix elements of all remaining decays of the Higgs boson from gluon-gluon fusion, as well as all decays from vector-boson fusion and decays to muons or bottom quarks from Higgs-strahlung are produced with POWHEG. For $t\bar{t}$ fusion calculations from PYTHIA6 are used, where decays to tau and vector boson pairs are produced in a combined data set with Higgs-strahlung.



	gg	VBF	WH	ZH	tt̄H
bb	✓	✓	✓	✓	✓
WW	✓	✓	✓	✓	✓
gg	✗	✗	✗	✗	✗
ττ	✓	✓	✓	✓	✓
cc	✗	✗	✗	✗	✗
ZZ	✓	✓	✓	✓	✓
γγ	✓	✓	✓	✓	✓
Zγ	✓	✓	✓	✓	✓
μμ	✗	✗	✓	✓	✗

Figure 8.3: The left plot illustrates all possible decay modes of the SM Higgs boson (adapted from [223]). For each channel the Higgs boson branching ratio is given as a function of the Higgs boson mass. A blue dashed line marks $m_H = 125$ GeV, as chosen for the MC data sets used in this analysis. On the right, a table summarizes which production and decay modes are taken into consideration by the CMS MC simulations. A green check mark indicates combinations accounted for in available data sets.

Furthermore, PYTHIA6 is chosen to simulate Higgs-strahlung with subsequent decays to di-photons and $Z\gamma$. Cross sections at NLO are evaluated for all processes, relying on information from the LHC Higgs Cross Section Working Group [223, 225].

Low Mass Resonances

A variety of low mass resonances are produced in the LHC proton-proton collisions. While the imposed minimum threshold on lepton transverse momenta introduced by the trigger strategy of this analysis (see Sec. 9.1.3) removes many of these mesons from the analysis reach, highly boosted lepton pairs from the Upsilon resonances $\Upsilon(1S)$, $\Upsilon(2S)$, $\Upsilon(3S)$ are energetic enough to pass the selection criteria. All three of these $b\bar{b}$ mesons have a mass of around 10 GeV. Simulation samples for the decay into a muon pair are available at LO from EVTGEN [226], a dedicated generator for processes involving bottom quarks, and are included in the SM description of the MUSiC analysis.

8.2.2 Beyond Standard Model Simulation

As elaborated on in Sec. 12.2, benchmark studies considering dedicated BSM signals are performed to evaluate the MUSiC discovery potential. For this installment of the analysis a hypothetical model is chosen that produces a heavy vector boson, the **S**equential **S**tandard **M**odel (SSM) W' [227]. Decay modes and branching fractions are identical to those of the SM $W^{\pm}$ boson apart from a correction for the additional $W' \to tb$ channel. Closely following the model developed for [228], an interference between the W' and $W^{\pm}$ boson is not taken into account and the mass of the postulated new particle is a free parameter of the theory. The signal's leptonic decay signature, given by an isolated high transverse momentum lepton (e, μ, τ) and substantial MET originating from the accompanying right-handed neutrino, leads to a broad Jacobian peak in the transverse mass distribution. The hadronic W' decay is not considered individually but is accounted for within the production for the leptonic channels and also mirrored in their cross section over the corresponding branching fraction. This regarded BSM physics scenario is one commonly used as a benchmark for collider experiments, helping to understand properties of the experiment at an early

stage of data-taking. Corresponding data sets considering different W' masses are summarized in Tab. 8.2, with more details given in App. B.2. The SSM processes are generated at LO in perturbation theory with PYTHIA6 and scaled to NNLO QCD cross sections obtained with FEWZ [229, 230] for each W' mass. Since high mass W' bosons are predominantly produced off mass-shell, effects resulting from the proton PDF occur in a similar form as given for lower masses. This leads to an increase of the k-factor for masses over 3 TeV.

Process	**Details**	**Generator**	**k-Factor (Order)**	$\mathbf{N_{events}}$
$W' \to l^{\pm}\nu$	$M_{W'} = 2000$ GeV	PYTHIA6	1.21 (NNLO)	$2.0 \cdot 10^4$
$W' \to l^{\pm}\nu$	$M_{W'} = 2500$ GeV	PYTHIA6	1.14 (NNLO)	$2.0 \cdot 10^4$
$W' \to l^{\pm}\nu$	$M_{W'} = 3000$ GeV	PYTHIA6	1.15 (NNLO)	$2.0 \cdot 10^4$
$W' \to l^{\pm}\nu$	$M_{W'} = 3500$ GeV	PYTHIA6	1.24 (NNLO)	$2.0 \cdot 10^4$

Table 8.2: Summary of BSM signal MC simulation data sets used in this analysis. Here $l^{\pm}$ stands for e, μ, τ and ν stands for the corresponding neutrino. A detailed overview can be found in Tab. B.2.

8.2.3 Parton Distribution Function Simulation

The formalism of QCD, as summarized in Sec. 2.4, can be used to precisely calculate interactions between individual quarks and gluons. Yet at a hadron collider the situation for the colliding protons is more complex. Protons are composite objects consisting of three distinct valence-quarks and many more quark anti-quark pairs, referred to as *sea-quarks*, as well as gluons. During a collision of two protons any two of these so-called *partons* may actually be the ones interacting. Individual probabilities of interaction for each parton type strongly depend on the momentum transfer Q^2 during the collision and are modeled by a so-called **P**arton **D**istribution **F**unction (PDF) [59, 191]. Given the partonic cross section $\sigma(p_i p_j \to X)$, the total hadronic cross section of two protons P colliding can be written as:

$$\sigma_{PP\to X} = \sum_{i,j} \int_1^0 \int_1^0 \mathrm{d}x_1 \mathrm{d}x_2 \; f_i(x_1, Q^2) \; f_j(x_2, Q^2) \cdot \sigma_{p_i p_j \to X}, \tag{8.1}$$

where $f_i(x_k, Q^2)$ is the **P**arton **D**istribution **F**unction (PDF) of parton k with momentum fraction x_k and the sum runs over all partons i, j potentially involved in the interaction. The **D**okshitzer **G**ribov **L**ipatov **A**ltarelli **P**arisi (DGLAP) equations [231–234] can be used to determine the Q^2 dependence of the PDFs. Specifically, once the PDF is known at a scale Q_0^2 the evolution to higher energy scales Q^2 is given through the DGLAP equations. Moreover, the x dependence of the PDFs cannot be calculated analytically in perturbation theory. Measurements of the PDFs at an initial scale Q_0^2 and a function of the x dependence are determined from global fits to experimental data from, e.g., deep-inelastic scattering or lepton-hadron colliders. Several different groups are devoted to determining these fit functions specialized on different combinations of experiments, covering a wide range in x and Q^2.

For analyses of LHC data, results from the **C**oordinated **T**heoretical-**E**xperimental Project on **Q**CD (CTEQ) [235, 236], **M**artin **S**tirling **T**horne **W**att (MSTW) [237] and **N**eural **N**etwork **PDF** (NNPDF) [238] groups are used predominantly. While CTEQ and MSTW retain fit results from a Hessian approach [239] using $\mathcal{O}(10)$ free parameters in a χ^2 approach at fixed α_s, the NNPDF group relies on neural-networks, allowing a simultaneous variation of α_s. More details of the different approaches and implications for resulting systematic uncertainties are discussed in Sec. 10.7.

The analysis described in this thesis relies on simulation data sets generating the hard scattering process with MadGraph and pythia6, which make use of the LO PDF set `CTEQ6L1` [235]. Data sets based on powheg calculations are given with the NLO PDF set `CT10` [236], with exception of those modeling single top quark processes, for which the NLO PDF set `CTEQ6M` [235] is chosen during event generation.

8.2.4 Pileup Simulation

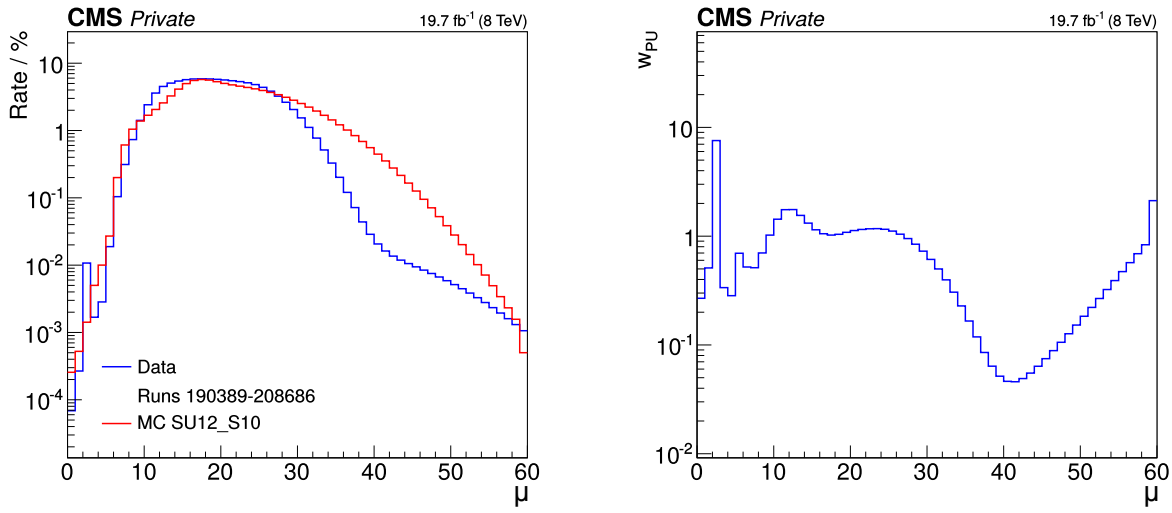
The high instantaneous luminosity at the LHC makes it rather probable for more than one proton-proton collision to occur in a single bunch crossing. The interaction of interest (hard interaction) is accompanied by energy from additional interactions, referred to as **p**ile**u**p (PU). In data taken during 2012, an average of 21 interactions take place per bunch crossing with peaks exceeding 40 interactions [85]. Although PU interactions cannot be separated from the primary interaction, the resulting energy offset can be estimated from measurable quantities of the event, e.g., the number of primary vertices, where the combined track and vertex reconstruction efficiency ($\sim$70%) must be taken into account. Alternatively, the instantaneous luminosity per bunch crossing multiplied with the total inelastic proton-proton cross section and divided by the protons' revolution frequency in the LHC gives the average number of interactions per bunch crossing. This method is preferable because of its smaller systematic uncertainties and since it does not introduce the bias of calibrating the data with itself. The blue curve in Fig. 8.4a shows how the mean number μ of interactions is distributed per bunch crossing in the analyzed data events. It is estimated from a minimum bias[5] inelastic proton-proton cross section of $\sigma_{\text{inel}} = 69.4$ mb at $\sqrt{s} = 8$ TeV [136] and instantaneous luminosity measured individually for each luminosity section[6] of the analyzed runs via the pixel cluster counting method (see Sec. 5.3).

During event generation an interim distribution of the mean number of PU events, the PU scenario, is used, as these are usually processed before data-taking has been completed. In addition to the hard interaction minimum bias events are passed through the detector simulation where the "true" number of simulated interactions is drawn from the input PU distribution. In addition to the dominant contribution occurring in-time within a single bunch crossing, energy referred to as out-of-time PU is also accumulated from previous and subsequent collisions, due to the calorimeter response time. Such contributions are accounted for during simulation by sampling additional interactions from a Poisson distribution with a mean driven by a value drawn from the PU scenario. The distribution measured in data corresponds to the sum of all these effects.

For simulation data sets from the `Summer12_DR53X` campaign PU scenario `PU_S10` [241] is chosen and its distribution is shown by the red curve in Fig. 8.4a. Adapting to the actual PU distribution during data-taking is performed through an event-by-event re-weighting applied to the simulated events using centrally supplied software documented in [242]. The distribution of the resulting event weight w_{PU} is given in Fig. 8.4b, along with information on how the scaling is applied.

[5] Minimum bias refers to a set of minimum (trigger) requirements made to account for the detector's geometric acceptance. A non-negligible dependence on the selected η range is observed for the proton-proton cross section [240], so a range is chosen that reflects one used by the majority of CMS analyses. In simulation this is emulated by inclusively generating events from all SM processes, which are dominated by ones of soft QCD interactions.

[6] A luminosity section (LS) is defined as the time it takes for the proton bunches to orbit the LHC 2^{18} times [109] and is used to divide a run into subsets of roughly the same amount of time, about 23 s. Within this time the instantaneous luminosity is expected to be roughly constant.



(a) (b)

Figure 8.4: The left plot shows the distribution of the average number of PU events μ per bunch crossing. The blue curves represents the number from data evaluated from the measured luminosity per bunch crossing, whereas the red curve is the PU distribution assumed during generation of MC simulation events. Both are normalized to unit area. The ratio of both histograms is given on the right and corresponds to the weight w_{PU} used to scale simulated events to the estimated data distribution on analysis level.

8.2.5 Detector Simulation

Once full modeling of the event on generator level is given by the steps described above, it is passed to a simulation of the CMS detector. A detailed database of the entire experiment including precise information on each component's position, size, shape and material is in place to ensure an accurate result [243]. The task of simulating the generated particles' reconstruction in the detector is interfaced with the GEANT4 package [244, 245]. A two-step procedure is carried out [246]. First, each event is considered under the aspect of all particles' passage through the detector, their interaction in the given matter and the sub-detectors response in form of simulated hits. The result is written in the SIM data format. Next, the response expected from the sub-detector front-end read-out electronics is addressed, leading to an event in DIGI format that can be passed to the reconstruction.

The procedure executed with GEANT4, referred to as **Full Sim**ulation (FullSIM), involves time-consuming algorithms especially within the SIM step. The **Fast Sim**ulation (FastSIM) package offers a substantially quicker alternative for event generation [203]. The method relies on modules using more simplistic models of the detector geometry, parameterizations of particle interactions in matter as well as electronics response and even the event PU. Per processed event, the computing time can be reduced by a factor of approximately 100. All modules are frequently tuned to results obtained with FullSIM and comparisons prove FastSIM to be sufficiently precise for many observables and over wide energy ranges.

All configurations applied during the detector simulation are stored in a so-called *global tag*. The simulation data sets analyzed for this thesis are processed with CMSSW using global tag `START53_V27::All` [247], as motivated by the recommended version of the JEC applied in each event. Finally, all MC simulation data sets are passed through the same reconstruction algorithms as the measured data.

8.2.6 Simulation Event Weighting

For each simulated process the generator returns a cross section σ, which may be adjusted to higher order calculations in QCD perturbation using a scale factor k. The final number of events, used to model the physics process in the considered distributions, is weighted to the rate expected to be observed in the measured data using the accumulated integrated luminosity L stated in Sec. 8.1:

$$w_{\text{lumi}} = \frac{k \cdot \sigma \cdot L}{N_{\text{SM}}}, \tag{8.2}$$

where N_{SM} is the number of events generated for the respective process.

Following the approach established in Sec. 8.2.4, a weight w_{PU} is determined to account for modeling of PU. Determined by the ratio of distribution of the average expected number of PU event per bunch crossing in data and in MC simulation, this weight is applied on event-by-event basis to all MC events.

Therefore, the total weight applied to every MC simulated event of a given process is given by:

$$w(\text{event, proc}) = w_{\text{lumi}}(\text{proc}) \cdot w_{\text{PU}}(\text{event}) \cdot w_{\text{gen}}(\text{event, proc}). \tag{8.3}$$

The additional weight w_{gen} accounts for potential event scaling during generation. Specific higher order MC generators, e.g., SHERPA or aMC@NLO, introduce weights on cross sections of different matrix elements used simultaneously in one process.

Chapter 9
Event and Object Selection

In contrast to model-driven analyses which often optimize their event selection to enhance a specific signal yield, the MUSiC strategy as a model independent search is to be as inclusive as possible. Nevertheless, criteria to ensure the quality of considered events and physics objects therein are applied. Recommendations tailored to the specific run conditions are based on dedicated studies from many individual groups within the CMS Collaboration, as summarized in the following chapter.

Known background or noise is rejected by methods highlighted in Sec. 9.1.1 and Sec. 9.1.2. Events passing these requirements are then assessed in terms of High-Level triggers for which details are given in Sec. 9.1.3. Subsequently, an unambiguous collection of physics objects is determined for each event applying the selection discussed in Sec. 9.2.

9.1 Event Selection

9.1.1 Primary Vertex

Background from non-hard-scattering events, i.e., beam-gas interactions occurring at a rate of appropriately 0.5 Hz per 10^{10} protons per bunch, is suppressed by requiring at least one primary vertex to be present in the vicinity of the nominal interaction point of each event [115]. To be considered a viable candidate a primary vertex must be closer to the beam spot than $|d_z| = 24$ cm along and $|d_0| = 2$ cm perpendicular to the beam axis. Moreover, the primary vertex fit (see Sec. 5.4.2) must show $N_{\text{dof}} \geq 5$ degrees of freedom, closely related to the number of tracks. Studies given in [114] report a misidentification rate well below 1% for this selection.

9.1.2 Event Filters

Throughout the first years of CMS data-taking several effects were observed, resulting in misreconstructed objects and events unsuitable for physics analysis. In most cases these anomalous effects lead to an over-estimated amount of MET. Recommendations for dedicated filters developed to identify and dismiss such events are provided by the MET group [138, 248]. A concise outline of filters used in the presented analysis is given in the following. In total only under 2% of the analyzed 2012 data events are affected by the mentioned filters. However, since the MUSiC analysis is especially sensitive to deviations in the tail of kinematic distributions, it is crucial to reject these events which lead to unphysical entries at high MET. Filters are applied to both measured data and simulated events.

Scraping Filter: A high number of reconstructed tracks can be an indication of a beam induced background, so-called *beam scraping*. In this case protons collide with the collimators deployed to remove protons scattered outside of the machine acceptance, which can potentially damage the superconducting magnets [249]. To reject these signatures, events with more than ten tracks are required to have a fraction of at least 25% high-purity tracks.

CSC Tight Beam Halo Filter: Resulting from interactions of the proton beam with residual gas in the beam-pipe, showers of secondary particles are produced, referred to as beam halo [250]. Especially halo muons traversing the entire detector along the z-axis, will lead to a significant increase of MET. Based on signals collected by the CSCs, halo muons are identified with a dedicated L1 trigger, as well as criteria ascertaining halo-like kinematics and timing characteristics.

Anomalous HCal Noise Filter: Several types of noise originating from the HCal HPDs instrumentation are observed, uncorrelated to the collisions. Filters based on HPD pulse shape, hit timing and multiplicity as well as a coincidence with the ECal are applied. The noise signals lead to HCal tower measurements of up to $\mathcal{O}$(TeV). Details on the noise characteristics are reported in [251] and a description of the filter can be found on [252].

HCal Laser Event Filter: Throughout Runs 2012 A to C laser pulses used to monitor the HCal HPDs and PMTs response during LHC inter-fill periods [253] were unintentionally fired during collision events. Details of a filter designed to discard these events are given on [254].

ECal Dead Cell Filter: Around 1% of ECal crystals show excessive noise and are disregarded in the reconstruction. Unphysical MET will arise if particles in forward direction encounter these masked ECal crystals. A combination of two filters [255] removes such events. L1 trigger information is used to identify saturated ECal cells and the sum of all energy collected in crystals bounding a masked 5×5 ECal supercrystal is used to asses large energy losses.

Bad EE Supercrystal Filter: Anomalously large signals are occasionally encountered from two specific 5x5 supercrystals of the ECal endcaps. The total supercrystal energy and number of hits failing reconstruction criteria in these supercrystals are used as indicators for such events.

ECal Laser Correction Filter: Over time ECal crystals lose transparency, due to the harsh radiation especially in the endcaps. Calibration lasers are used throughout data-taking to measure the current loss and determine a correction factor [256]. In rare cases the reconstruction delivers corrections several times higher than the usual values. Such events are removed from the analysis.

Tracking Failure Filter: Two effects lead to substantial calorimeter entries without corresponding tracks. An exceptionally high number of clusters can cause unintentional termination of the iterative tracking algorithm. Secondly, tracks displaced by more than 75 mm from the nominal interaction point may not be recognized by the tracking algorithm. Such events are identified and rejected over the ratio of the transverse momentum sum of all tracks belonging to well-reconstructed primary vertices and the total transverse energy of all particle-flow jets in the event.

Coherent Noise Filters: Noise originating from the strip tracker leads to an anomalous abundance of clusters incompatible with a much lower given number of observed pixel clusters. Resulting events can either show hundreds of tracks or even zero, due to an overload of the track reconstruction. Measures applied during the L1 trigger reject most of these events, but must be accompanied by dedicated filters on analysis level. An absolute and relative comparison of the number of clusters found in the pixel and strip trackers provides a good rejection method.

9.1.3 Trigger Selection

As explained in Sec. 5.2 not all events produced during the proton-proton collisions can be processed by the detector readout system and computing intense reconstruction algorithms. Instead they are passed through a sequence of increasingly complex trigger requirements. On analysis level specific HLT paths are chosen to select relevant event topologies. For this thesis, a focus is placed on events containing isolated electrons and muons. An overview of the corresponding triggers is shown in Tab. 9.1, consisting of single and double-lepton triggers that were unprescaled over all analyzed data runs. In an extension of the analysis, jet-triggered events are also taken into consideration [257]. Each data and simulation event is required to have fired at least one of the given triggers. Additionally, the event topology is constrained by requiring a minimum electron or muon multiplicity and transverse energy or momentum. Details of the respective particle identification on analysis level are given in Sec. 9.2.

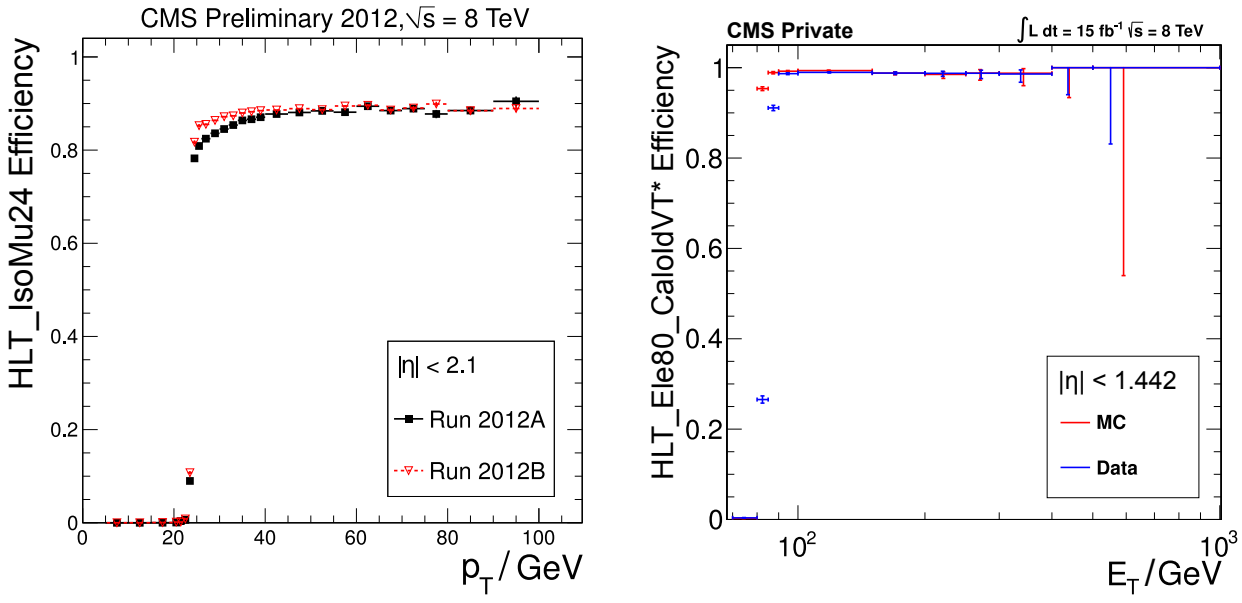
Single-Muon	`HLT_IsoMu24_eta2p1_v*`
⇒	$\geq 1\mu$ with $p_T > 30$ GeV
Double-Muon	`HLT_Mu17_Mu8_v*`
⇒	$\geq 2\mu$ with $p_T > 25$ GeV each
Single-Electron	`HLT_Ele80_CaloIdVT_TrkIdT_v*` or `HLT_Ele80_CaloIdVT_GsfTrkIdT_v*`
⇒	≥ 1e with $E_T > 100$ GeV
Double-Electron	`HLT_Ele17_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_v*`
⇒	≥ 2e with $E_T > 25$ GeV each

Table 9.1: Summary of criteria applied to events triggered by the stated single and double lepton triggers. In the HLT-path, `v*` serves as a placeholder for the trigger version number as given by the respective trigger menu used during data-taking or generation of simulation events. Motivated by the efficiency of each trigger, requirements on the leptons transverse momentum or energy are applied.

The `HLT_IsoMu24_eta2p1` trigger relies on tracker and muon system information to reconstruct muon candidates with a transverse momentum of at least 24 GeV within an acceptance of $|\eta| < 2.1$. In the final step a combined fit, returning so-called *Level-3* muons, is performed to exploit the benefits of both detector subsystems and improve the muon momentum resolution. An isolation criteria based on tracks and calorimeter deposits within a cone of $\Delta R = 0.3$ around the muon is applied, allowing for a lower momentum threshold. As can be seen in Fig. 9.1a, a pileup mitigation technique employed for the isolation requirement (see Sec. 9.2.1 for details) results in a steeper trigger efficiency response. The plateau of about 90% is reached slightly beyond the trigger's momentum threshold, justifying an analysis level requirement of $p_T > 30$ GeV for the most energetic muon in the event. As reported in [258], data and simulation show a similar turn-on behavior, compatible within under 2% of each other.

The double-muon trigger `HLT_Mu17_Mu8` requires two Level-3 muon candidates. Consequently the constraint on pseudorapidity can be lifted and the momentum thresholds are lowered to 17 GeV and 8 GeV. To ensure both muons originate from the same vertex their tracks must be closer than $d_z = 0.2$ cm [258], measured at the respective point of closest approach to the nominal beam axis. On analysis level at least two muons in the event are required to have $p_T > 25$ GeV, which is given by the quality of the muon identification summarized in Sec. 9.2.2.

Electron triggers are based on ECal signatures to identify Level-1 electron / photon candidates. In the HLT step a disambiguation is achieved through the presence or lack of a nearby track. During Run B of 2012 data-taking the tracking method for the single-electron trigger was changed from KF tracking (`HLT_Ele80_CaloIdVT_TrkIdT`) to GSF tracking (`HLT_Ele80_CaloIdVT_GsfTrkIdT`), leading to a trigger rate reduction of around 25% without loosing efficiency [260]. In all simulation events the GSF was used throughout. In addition to $E_T > 80$ GeV, selection criteria are imposed on cluster shape and isolation variables, reducing the rate of electron candidates to around 1 kHz [258]. The E_T dependent trigger efficiency for electrons reconstructed in the barrel region is shown in Fig. 9.1b for data and simulation events, as taken from [259]. A compatible behavior is found for the endcap regions. Compared to the single-muon trigger shown in Fig. 9.1a, the turn-on is less steep (note the logarithmic x-axis in Fig. 9.1b) and a plateau of nearly 100% is reached for electrons with about 100 GeV which is placed as the analysis level E_T requirement for at least one electron in the event.



(a) (b)

Figure 9.1: Overview of measured efficiencies for single-lepton triggers given in Tab. 9.1. Fig. 9.1a (adapted from [104]) shows the muon-momentum dependent curve for the isolated single-muon HLT with data from 2012 Run A (black squares) and Run B (red triangles). The steeper turn-on in Run B data results from additional pileup mitigation techniques. In Fig. 9.1b (adapted from [259]) the isolated single-electron HLT efficiency is given for electrons reconstructed in the barrel region for data (blue markers) and simulation (red markers) events. At around 100 GeV both curves have reached a plateau close to 100%.

A double-electron trigger is used to regain events lost because of the relatively high E_{T} cut applied in response to the single-electron trigger threshold. Especially the low energy region of $\sum |\vec{p}_{\mathrm{T}}|$ and $M_{(\mathrm{T})}$ distributions for event classes such as 2e and 1e+X are recovered. The corresponding HLT path selects at least two electrons with $E_{\mathrm{T}} > 17$ GeV and 8 GeV, respectively. The track, cluster shape and isolation criteria are looser than those applied for the single-electron case. ECal transparency corrections mentioned in Sec. 9.1.2 have a strong influence on the efficiency for electrons reconstructed in the endcap regions. Once these have been applied the E_{T} dependent efficiency in both barrel and endcap show a sufficiently steep turn-on behavior (distributions can be found in [258]) and allow lowering the analysis level requirement to $E_{\mathrm{T}} > 25$ GeV for at least two well-reconstructed electrons in the event, with electron selection criteria as given in Sec. 9.2.3.

9.2 Object Selection

As described in Sec. 7.2, the MUSiC classification algorithm divides analyzed events into classes defined by the final state's object content. Thus, a well-defined set of objects must be determined for every event. The following section gives an overview of the criteria used to identify objects taken into account by the analysis presented in this document, where a focus is set on events containing isolated leptons. Many of the criteria are based closely on parameters of the reconstruction algorithms explained in Sec. 5.4. For the first analysis of CMS data taken at 8 TeV, well understood objects with relatively small misidentification rates are chosen. In principle, additional criteria can be applied to distinguish heavy flavor quarks or tau leptons from light flavor quark jets. A study with b-jets in MUSiC was performed for the 8 TeV data set and summarized in [257]. Tagging of b-jets can enhance sensitivity to particular types of BSM effects, since the final states are grouped into

units of similar physics processes. On the other hand, issues of an insufficient number of generated events of the SM MC simulation can become more relevant if similar final states are spread out over several event classes.

A compromise must be found between rejecting correctly identified objects by applying tight cuts and accepting wrongly interpreted detector signals, due to rather loose criteria. For MUSiC a tighter selection is usually advantageous. In general, events will not be rejected completely but rather assigned to a different event class with a lower or different object-type multiplicity. It is important to ensure that a selected object's type is assigned correctly. Consequently, the criteria summarized in the following paragraphs have been chosen to minimize the misidentification rate. Objects are removed from the event if they do not fulfill any consistent set of criteria introduced in the following sections. In summary, the analysis presented in this thesis considers objects summarized in the following table.

Object	$\mathbf{p_T}$	$\lvert\eta\rvert$
μ	≥ 25 GeV	≤ 2.1
e	≥ 25 GeV	≤ 2.5
γ	≥ 25 GeV	≤ 1.442
jet	≥ 50 GeV	≤ 2.4
MET	≥ 50 GeV	–

Table 9.2: Summary of object selection acceptance criteria used for the analysis presented in this document. These requirements are applied to objects in events that fulfill at least one of the triggers summarized in Tab. 9.1, including the introduced momentum thresholds.

9.2.1 Pileup Mitigation

Techniques are applied to mitigate the effect of supplementary energy added by pileup interactions and the underlying event [261, 262]. For each simulation event an average energy density ρ is calculated which quantifies the amount of supplementary transverse momentum per unit area A. To account for inhomogeneities in the distribution of pileup throughout the event the median is chosen to assess this quantity. On average additional particles are distributed uniformly in η and ϕ throughout the event. Thus, a jet's spacial extent is a good measure for the amount of supplementary energy it accumulates. To determine the area A of each jet[1], low momentum *ghost* particles are superimposed on the event. Jet clustering is then reevaluated simultaneous to jet reconstruction (see Sec. 5.4.6) using FastJet [134] with the k_{T}-algorithm of $\Delta\mathrm{R} = 0.6$. The vicinity containing ghost particles assigned to a specific jet gives the jet area.

In each event ρ is used to correct isolation variables of electrons and photons. For each object an effective area A_{eff} is determined from the mean isolation and energy density as a function of the number of primary vertices in the event. The resulting PF based isolation is then given by:

$$I_{\mathrm{PF}}^{\mathrm{corr}} = \max\{I_{\mathrm{PF}} - A_{\mathrm{eff}} \cdot \rho, 0\}. \tag{9.1}$$

A different technique, based on the ratio of neutral to charged particles $\Delta\beta$, is used to correct muon isolation variables. Details on both methods are given in the following sections.

[1]Note that the jet area is not the same as the jets's cone. Jets are comprised of point-like particles to which no intrinsic area can be assigned.

9.2.2 Muons

Four algorithms are available to reconstruct muons in different energy regimes (see Sec. 5.4.3). In the final muon selection used in this analysis, properties from the Tracker Muon approach are used for muons with $p_{\mathrm{T}} \leq 200$ GeV, since low p_{T} muons are subject to multiple scattering and the tracker measurement has a superior resolution in this momentum range. Above this value, effects from showering increase and the so-called *Cocktail Muon* reconstruction using *TuneP* is chosen [103]. This method compares the relative transverse momentum uncertainty $\sigma_{p_{\mathrm{T}}}/p_{\mathrm{T}}$, and the track fits' goodness-of-fit values χ^2/N_{dof} of all four algorithms. As a result the majority of high p_{T} muons are reconstructed under the TPFMS or Picky Muon methods. Further selection criteria are applied to muon candidates to suppress muons originating from cosmic-rays or jets (mesons decaying in flight) as well as *punch-through* particles (mostly hadrons) traversing the muon system and misidentified as muons. The tight working point of the *Baseline* muon selection, which is still more efficient than 95%, is applied below $p_{\mathrm{T}} = 200$ GeV. Muons with a larger transverse momentum must fulfill requirements of the *High-p_T* muon selection. All criteria are summarized in Tab. 9.3 and will only be discussed here briefly. Details can be found in [103, 263].

Criteria used in both Baseline and High-p_{T} muon selection	
$\lvert\eta\rvert$	≤ 2.1
Global Muon	`True`
$N_{\text{pixel hits}}$	≥ 1
$N_{\text{tracker hits}}$	≥ 6
d_{xy}	≤ 0.2 cm
d_z	≤ 0.5 cm
$N_{\text{matched stations}}$	≥ 2
$N_{\text{muon chamber hits}}$	≥ 1
I_{PF}	≤ 0.12
Criteria used exclusively in Baseline muon selection	
p_{T}	≥ 25 GeV, ≤ 200 GeV
χ^2/N_{dof}	≤ 10
PF Muon	`True`
Criteria used exclusively in High-p_{T} muon selection	
p_{T}	> 200 GeV
$\sigma_{p_{\mathrm{T}}}/p_{\mathrm{T}}$	≤ 0.3

Table 9.3: Concise summary of criteria used to identify muons within the tight working point of the Baseline or with the High-p_{T} muon selection. The requirement on the muon isolation I_{PF} is also listed here. All values are taken from [264].

Muons are regarded within an acceptance of $|\eta| \leq 2.1$ (motivated by the muon trigger acceptance), with $p_{\mathrm{T}} \geq 25$ GeV, and must be identified as a Global Muon. To reject muons from meson decays, the muon track must show at least one hit in the pixel tracker and several in the full tracker volume, also ensuring sufficient measurements for an adequate muon p_{T} determination. Upper thresholds on the transverse d_{xy} and longitudinal d_z impact parameters of the track with respect to the primary vertex help to additionally reduce contributions from cosmic-ray muons. Matching the tracker track to the muon system warrants a much smaller amount of punch-through particles. Muon track segments must be found in at least two muon stations and at least one valid muon chamber hit must be included in the final Global Muon track fit. For muons with $p_{\mathrm{T}} \leq 200$ GeV, the Global

Muon fit must furthermore exhibit a reasonable χ^2/N_{dof} and the muon must be identified as such by the PF algorithm (see Sec. 5.4.5). These two criteria are dropped for high-p_{T} muons. Instead, quality of the muon track, evaluated from the specialized high-p_{T} reconstruction, is assessed by the relative transverse momentum uncertainty. In summary, the fraction of punch-through hadrons, a small quantity owing to the material budget of over 16 interaction lengths in front of the muon system [265], is further reduced to 0.5% by the tight working point of the muon selection [263]. Additionally, a maximum value is imposed on each muon's isolation, calculated via:

$$I_{\text{PF}} = \frac{1}{p_{\text{T}}} \cdot \left[\Sigma E_{\text{T}}(\text{chHad from PV}) + \max\{0,\ \Sigma E_{\text{T}}(\text{neutHad}) + \Sigma E_{\text{T}}(\gamma) - \Delta\beta \cdot \Sigma E_{\text{T}}(\text{chHad not PV})\} \right] \quad (9.2)$$

within a cone of $\Delta\text{R} < 0.4$ around the PF muon candidate [266]. Individual energy contributions from charged and neutral hadrons as well as photons, given by the PF algorithm, are set in relation to the muon transverse momentum. Contributions from pileup are subtracted from the neutral components using a factor $\Delta\beta = 0.5$ of the charged hadron transverse energy not allocated to the leading primary vertex, since charged hadrons, e.g., $\pi^{\pm}$ are twice probable as neutral ones π^0.

9.2.3 Electrons

The identification of prompt, isolated electrons must show robustness against misidentified jets and converted photons. The **H**igh **E**nergy **E**lectron **P**airs (HEEP) selection (version `v4.1`), as given in Tab. 9.4, exploits the superior momentum resolution of the ECal over the tracker for rather straight tracks of high-p_{T} electrons, while still efficient at moderate energies [267–269]. Selection criteria are optimized individually for barrel and endcap regions, resulting from differences in the geometrical layout of the ECal instrumentation. Electrons falling into the transition region ($1.442 \leq \eta \leq 1.56$) are excluded from the analysis, as reconstruction of showers extending in both barrel and endcap crystals leads to an increased misidentification rate and lower efficiency.

	Barrel	Endcap
$\lvert\eta\rvert$	< 1.442	$> 1.56, \leq 2.5$
E_{T}	> 25 GeV	> 25 GeV
ECal driven	`True`	`True`
$\lvert\Delta\eta_{\text{in}}\rvert$	< 0.005	< 0.007
$\lvert\Delta\phi_{\text{in}}\rvert$	< 0.06	< 0.06
$N_{\text{lost hits}}$	< 2	< 2
$\lvert d_{xy}\rvert$	< 0.02 cm	< 0.05 cm
H/E	< 0.05	< 0.05
$\lvert\sigma_{i\eta i\eta}\rvert$	–	< 0.03
Shower shape	$E^{1\text{x}5}/E^{5\text{x}5} > 0.83$, or $E^{2\text{x}5}/E^{5\text{x}5} > 0.94$	–
I_{track}	< 5 GeV	< 5 GeV
I_{calo}	< 2 GeV $+ 0.03 \cdot E_{\text{T}} + 0.28 \cdot \rho$	< 2.5 GeV $+ 0.28 \cdot \rho$, for $E_{\text{T}} < 50$ GeV, else: < 2.5 GeV $+ 0.03 \cdot (E_{\text{T}} - 50$ GeV$) + 0.28 \cdot \rho$

Table 9.4: Concise summary of criteria used to identify electrons within the HEEP selection. Values taken from [269].

Instead of relying on seeds from tracker information and the PF algorithm for an estimate of the electron trajectory, reconstruction is based on energy deposits in the ECal (*ECal driven*) once

calibration and corrections have been applied. Consequently, the transverse energy E_{T}, evaluated from ECal supercluster entries, is favored over the transverse momentum evaluated from the track curvature. A minimum value of 25 GeV is required in this analysis. Moreover, thresholds in η and ϕ are imposed on the difference between the position of the supercluster and the track extrapolated from the innermost tracker position to the interaction vertex and subsequently to the ECal, to ensure a good match between tracker and calorimeter measurements. By comparing the detected number of hits along the electron trajectory in the tracker with the expected number of hits, electrons originating from photon conversions or produced in jets can be rejected. Such electrons are not produced in the primary vertex, but further outwards in the tracker and lack hits in the innermost layers. Along the same lines, a maximum value is imposed on the transverse distance d_{xy} between the electron's track and the primary vertex.

Since hadronic showers are broader than electromagnetic showers, variables sensitive to the shower shape are exploited to reject jets misidentified as electrons. The ratio H/E of energy collected in HCal towers within a cone of $\Delta\mathrm{R} = 0.15$ around the electron's supercluster deposit and the energy associated to the supercluster itself is used in both barrel and endcap regions. The lateral spread of energy is quantified via $\sigma_{i\eta i\eta}$, which relates the energy of the seed crystal to deposits found in the surrounding 5×5 matrix in units of an averaged crystal width. In the barrel an alternative variable shows better performance. The energy sum within the 5×5 crystal grid around the seed crystal is set in relation to the energy found within $\eta \times \phi$ strips of 1×5 and 2×5. Well reconstructed electrons are expected to be narrow in η. The two different η-strip geometries are used to cover electrons hitting ECal crystals centrally and those close to the crystal's edge.

An upper threshold I_{track} is set on the sum of transverse momenta found within a cone of $\Delta\mathrm{R} = 0.3$ around the electron in the tracker, where the region of $\Delta\mathrm{R} < 0.04$ is disregarded to exclude the electron track. A strip in ϕ direction, three crystal wide in η, removes photons from bremsstrahlung [270]. As elaborated on in Sec. 9.2.1, pileup has a strong influence on electron properties such as shower shape, isolation and energy measured by the calorimeters. Thus, requirements on the calorimeter based isolation I_{calo} are parameterized by the median energy density ρ in each event and use an effective area fixed to $A_{\mathrm{eff}} = 0.28$. The isolation is calculated as the sum of energy E_{T} from ECal and HCal entries, excluding energy allocated to the electron itself.

9.2.4 Photons

Following approaches comparable to those invoked for electron identification, criteria given by the tight working point of the *Simple Cut Based* photon selection, see Tab. 9.5, address the rejection of neutral hadrons and electrons falsely taken to be photons [271]. To ensure misidentification rates are sufficiently low, only photons reconstructed in the barrel region are considered in the MUSiC analysis. Furthermore, E_{T} must exceed 25 GeV. Once more, the ratio of energy collected in the ECal supercluster and deposits in the HCal, as well as $\sigma_{i\eta i\eta}$ are used to identify electromagnetic showers induced by photons. Differing from the H/E variable used for electron selection, only the HCal tower closest to the ECal seed crystal is used in the energy sum [272], reducing effects from pileup. Isolation requirements are based on properties of PF candidates within a cone of $\Delta\mathrm{R} = 0.3$, where the photon itself is excluded from the isolation sum through additional geometrical restraints. Pileup mitigation techniques are applied when calculating photon and neutral hadron isolation variables, using effective areas evaluated from simulated γ+jets events [273].

In addition to jets, electrons with insufficient tracks may also be misidentified as photons. Two approaches are available to reject such electrons. On the other hand, up to 60% of photons convert into electron pairs before reaching the last three tracker layers [271]. Both presented electron rejection techniques undergo dedicated measures to correctly incorporate such photons. The *conversion-safe*

E_{T}	> 25 GeV
$\lvert\eta\rvert$	≤ 1.442
Single tower H/E	< 0.05
$\sigma_{i\eta i\eta}$	< 0.011
$I_{\text{Charged Hadron}}$	< 0.7 GeV
$I_{\text{Neutral Hadron}}$	< 0.4 GeV $+ 0.04 \cdot p_{\mathrm{T}}(\gamma)$
I_{Photon}	< 0.5 GeV $+ 0.005 \cdot p_{\mathrm{T}}(\gamma)$
Pixel track seed veto	`True`

Table 9.5: Concise summary of criteria used to identify photons based on the tight working point of the Simple Cut Based selection. Values taken from [273].

electron veto, developed for analyses with clean photon signatures like [274], rejects candidates whose ECal supercluster is matched to a track with no hits missing in the innermost pixel tracker layers. Yet the candidate is retained if it can be associated to a reconstructed conversion vertex. Almost exclusively photons converting before the first tracker layer remain unrecovered. Recommended for event topologies containing prompt electrons as well as photons, the *pixel track seed veto* offers a more restrictive photon selection. This method rejects all photon candidates whose ECal supercluster is in the vicinity of a track associated to at least two hits in the pixel detector. Converted photons can be recovered if they occur outside of the first two pixel layers. Compared to the conversion-safe electron veto, the probability of misidentifying electrons as photons is reduced by the pixel track seed veto, accompanied by only a slight loss in efficiency for photons reconstructed in the barrel region [271, 275]. For MUSiC, the latter approach offers the strong advantage of a diminished electron to photon misidentification rate.

9.2.5 Jets

Identification criteria based on the PF jet composition are applied on jet candidates of transverse momenta surpassing 50 GeV. The loose working point is chosen as recommended on [276], enabling rejection of calorimeter noise, non-prompt jets as well as electrons and photons misidentified as jets. For genuine jets the selection is still efficient to around 99% [277]. Reasonable noise rejection is only given for jets found within the tracker's geometrical acceptance of $|\eta| \leq 2.4$ [278]. As summarized in Tab. 9.6, each jet must be reconstructed from at least two constituents, of which at least one must be charged. In the barrel region, the five energy fractions defined through PF jet reconstruction add up to unity. Thresholds are applied individually to each of the energy fractions given by the cumulative energy of the respective PF candidate type.

As studied in [257], a tighter jet selection is advantageous if high p_{T} jets are also used to potentially trigger events, resulting in additional event classes containing jets but no leptons.

As previously introduced in Sec. 5.4.6, the energy resolution of jets in simulation is smeared to correct for differences encountered with respect to measured data [136]. Following a procedure given on [137], a matching is performed in each simulation event between reconstruction level (`RECO`) and generator level (`GEN`) jets. If a match within a given η and p_{T} range is found, simulation-to-data scale factors c, determined from di-jet asymmetry, are applied to the reconstructed jet's p_{T} via:

$$p_{\mathrm{T}}^{\texttt{RECO}} = \max\{p_{\mathrm{T}}^{\texttt{GEN}} + c \cdot (p_{\mathrm{T}}^{\texttt{RECO}} - p_{\mathrm{T}}^{\texttt{GEN}}), 0\}, \tag{9.3}$$

where c spans from 1.05 to 1.29 throughout ranges of η. A generator level match cannot be assigned to every `RECO` jet, especially those originating from pileup, and an alternative method must be used.

p_{T}	≥ 50 GeV
$\lvert\eta\rvert$	≤ 2.4
$N_{\text{Total Constituents}}$	> 1
$N_{\text{Charged Constituents}}$	> 0
Neutral Hadron EF	< 0.99
Neutral Electromagnetic EF	< 0.99
Charged Hadron EF	> 0
Charged Electromagnetic EF	< 0.99
Muon EF	< 0.8

Table 9.6: Concise summary of criteria used to identify jets within the loose working point of the Particle Flow jet selection. Here, energy fraction is abbreviated as EF. Values taken from [276].

A random smearing is applied following a Gaussian distribution centered at one. The distribution's width is set to $\sigma = \sqrt{c^2 - 1} \cdot \sigma_{\mathrm{MC}}$, where the relative difference of `RECO` and `GEN` level jet p_{T} is used to estimate σ_{MC}. Since this distribution depends strongly on the considered final state, no general value can be given for σ_{MC}. A dedicated study based on a selection as described in this chapter using the full set of SM data sets given in Tab. 8.1 was performed in [257], revealing a value of $\sigma_{\mathrm{MC}} = 0.1$ to be suitable over the full η range.

9.2.6 Missing Transverse Energy

Building on the PF algorithm and corrections discussed in Sec. 5.4.7, further effects are propagated to the MET used in this analysis. The muon transverse momentum considered during MET reconstruction is evaluated straightforward from the (corrected) PF candidates. Thus, for high-p_{T} muons the difference between the transverse momentum given by PF and the Cocktail Muon approach must be applied to MET, to avoid an artificially high amount of MET in the corresponding events. Moreover, the influence of JER smearing applied to jets in simulation is accounted for.

After including all corrections and shifts, a lower threshold of 50 GeV is imposed on MET. This implies that not all event classes contain MET as an object and the MET kinematic distribution is not investigated for those where it is not present. The low MET regime is especially prone to mismeasurement and reconstruction inaccuracies.

9.2.7 Resolving Object Ambiguities

As outlined previously, reconstruction and identification is performed separately, for the most part, for each object type. Imperfections in the algorithms and selection can lead to simultaneously identifying the same candidate as two different objects. Such incidents can lead to an incorrect total number of particles in the event, since candidates are counted repeatedly. Moreover, nearby objects may share individual subdetector entries, potentially altering the objects' kinematic properties. Both cases must be resolved before an event is classified by the MUSiC algorithm (see Sec. 7.2), to guarantee a largely correct and unambiguous description of the final state. A solution is found by removing objects in the vicinity of others, addressed in a hierarchic order given by the reliability of an object's identification. Only candidates fulfilling the appropriate selection criteria are considered. The distance between potentially overlapping objects is evaluated within a ΔR range corresponding to the isolation cone used during reconstruction.

- Initially, the overlap among electrons is removed. If two electrons are within $\Delta\mathrm{R} < 0.3$ of each other and are associated to the same ECal supercluster seed or the same tracker track, the one with the lower energy is removed from the list of electrons.
- In the following step, all electrons within a cone of $\Delta\mathrm{R} < 0.4$ of a muon are removed. Muons are detected with a higher purity, whereas bremsstrahlung photons emitted from a muon combined with the muon's track can be misidentified as electrons.
- Analogous to electrons, of two photons falling within $\Delta\mathrm{R} < 0.3$ of each other and assigned to the identical ECal supercluster seed, only the higher energetic one is retained.
- Moreover, photons are removed from the event if they are within $\Delta\mathrm{R} < 0.3$ of an electron, also ruling out that they are associated to the same ECal supercluster.
- Overlap between photons and muons within $\Delta\mathrm{R} < 0.4$ around the muon is removed to reduce effects from bremsstrahlung at high muon p_T.
- Finally, all jets closer than $\Delta\mathrm{R} < 0.5$ to any remaining particle are removed. Since the PF algorithm initiates a jet for every reconstructed PF candidate it is essential to eliminate these duplicates. The used anti-k_T jets are constructed without overlap, so a cleaning among jets is superfluous.

In principle overlap can also be found among muon tracks. However, high momentum Z^0 candidates and decays from Upsilon mesons produce genuine close by boosted muon pairs which would be removed by a muon-against-muon cleaning, although both muons can be reconstructed and separated sufficiently well. Consequently, the PF based muon isolation variable (see Sec. 9.2.2) is insensitive to the transverse momenta of additional muons within the isolation cone. For electrons on the other hand, only the E_T of the electron in the center of the isolation cone is removed. Thus, the majority of close by electrons will already be rejected by the isolation criteria. The dedicated overlap removal described here does not have a large impact on the remaining electrons.

All selection requirements discussed in this chapter are closely motivated by the reconstruction algorithms and optimized to the detector's response to each individual object type. Thus, criteria for muons and electrons, e.g., isolation requirements, are not identical and differences are expected to show up when comparing the resulting event classes.

Chapter 10
Systematic Uncertainties

Methods invoked to reconstruct and identify the different objects used in this analysis are often based on measurements with limited precision. Moreover, the SM MC simulation cannot be expected to fully describe all final states and energy regimes, since the theoretical predictions and assumptions made to model the complex nature of each collision event can only give an estimate of the true processes. On experimental level sophisticated quantities such as detector response and identification efficiencies must be accounted for adequately. Effects leading to a difference in the behavior observed in measured data and simulated events must be understood. Therefore, a key issue is determining reasonable systematic uncertainties to cover areas where the description through simulation is not ideal. For MUSiC, relevant sources of systematic uncertainties are energy and momentum scales of all considered objects, as well as their reconstruction efficiencies, misidentification rates and charge misidentification rate. Moreover, the normalization of each SM MC process is influenced by uncertainties on the assumed integrated luminosity, SM cross section, parton distribution function and the number of generated events.

All systematic uncertainties accounted for are taken to follow Gaussian prior functions, as required by the p-value definition given in Eq. (7.13). In each region of a distribution, uncertainty contributions are calculated individually and subsequently combined over all bins in the region. Bins are taken to be fully correlated within each uncertainty source, apart from the one attributed to the finite number of simulated events. Correlations between event classes are also considered for uncertainties on reconstruction efficiencies, the integrated luminosity, SM cross sections and parton distribution functions. Finally, the individual uncertainty sources are assumed to be uncorrelated among each other and summed in quadrature, resulting in a single value σ_{SM} per region.

As described below, a refined event based estimate of the uncertainty resulting from the proton substructure is applied. Uncertainties resulting from the applied JEC are given binned in jet η and p_{T}. Other uncertainties are also assessed in terms of the object's transverse momentum or given by, e.g., the number of objects of one type reconstructed in the event. Initially, some estimates may seem relatively large and potentially overestimated. Yet, one must bear in mind that effects occurring throughout all investigated final states must be incorporated. The uncertainties must not only cover event classes dominated by clear signatures such as single lepton or di-jet events, but also more complex ones exhibiting high jet multiplicities.

10.1 Energy and Momentum Scale

The energy or momentum of objects measured and reconstructed in the detector may be subject to a systematic shift. An uncertainty is introduced to reflect the level of precision between the measured data and the MC estimate for each object type separately. To incorporate the influence on the final event yield in each event class distribution, the energy or momenta of individual objects of one type are shifted upwards and downwards by one standard deviation of the assumed systematic for each SM MC event. Therefore, the entire analysis is performed three times for each considered object, once with the nominal energy/momentum values, once with the values shifted upwards and once with them shifted downwards. Since a change in energy/momentum can lead to a shifted object passing or failing the threshold applied during the selection step (see Sec. 9.2), events may migrate between different event classes. Assume a jet is reconstructed with its p_{T} close above the selection threshold of 50 GeV. Then a sufficient downwards shift will lead to this jet failing the

selection criteria and the event is sorted into the corresponding event class with one fewer jet. For a small number of event classes this approach can lead to cases in which distributions only contain contributions from energy/momentum scale uncertainties and no SM prediction or measured data.

An uncertainty on the number of events is evaluated from the three histograms obtained from the individual shifting procedure for muons, electrons, photons and jets. In each case the respective uncertainty contribution in each bin i of the event class distribution is given by the difference of the number of events in histogram associated to the upwards/downwards shift to the number of events given by the nominal histogram:

$$\sigma_{\mathrm{SM},i}^{\mathrm{up/down}} = N_{\mathrm{SM},i}^{\mathrm{up/down}} - N_{\mathrm{SM},i}. \tag{10.1}$$

Both values are combined to a symmeterized total uncertainty for each energy scale (ES):

$$\sigma_{\mathrm{SM},i}^{\mathrm{ES}} = \frac{\left|\sigma_{\mathrm{SM},i}^{\mathrm{up}}\right| + \left|\sigma_{\mathrm{SM},i}^{\mathrm{down}}\right|}{2}, \tag{10.2}$$

which can be handled by the Gaussian prior of the p-value (see Sec. 7.3.2). In the following, details are given for individual objects, including values assumed for each energy/momentum shift.

Muons: An offset in the muon momentum measurement can occur though a misalignment of the detector. The absolute extent of such shifts is studied with events containing signatures from cosmic-ray muons [102]. The so-called *cosmic-ray muons end-point* method exploits the fact that a single muon traverses the entire detector, leading to reconstructing two individual tracks for the same object. The muon track curvature can be used to determine a charge dependent offset in the momentum measurement to $\kappa = q/p_{\mathrm{T}}$, where no muon is expected to have infinite momentum, i.e., a curvature of $\kappa = 0$. An artificial momentum shift is determined which optimizes the agreement between the measured data and the MC simulation. For analyses considering high energetic muons ($p_{\mathrm{T}} > 200$ GeV) an uncertainty of 5% per 1 TeV of the muon transverse momentum is recommended on the shift which itself is found to be close to zero [279], grounded on an update of the end-point analysis performed with cosmic-ray muon events taken in 2011 [280]. This value is used to shift the inverse muon transverse momentum and to determine the global influence on the event yield in each event class containing muons.

Electrons and Photons: The energy of electrons and photons is determined by deposits in the ECal. Following calibrations of ECal crystals before commissioning and in-situ measurements with a laser monitoring system, the absolute ECal energy scale is determined from the Z mass distribution of events in the ee and $\mu\mu\gamma$[1] decay channels [281, 282]. Studies based on events taken in 2012 at 8 TeV show a total systematic variation of the electron and photon energy measurement of 0.6% in the barrel and 1.5% in the endcap regions [267, 271]. Considering each object type separately, the four vectors of electrons and photons are shifted by the corresponding value to estimate the overall influence on the number of events in the distribution.

Jets: As described in Sec. 5.4.6, the JEC are applied to every jet, calibrating measured and simulated jets in a factorized approach. Systematic uncertainties are determined for each step of the correction process individually and factors including a systematic shift of one standard deviation are provided binned in jet η and p_{T} [136]. All uncertainty sources are taken to be fully correlated and the total JEC uncertainty is in the range of 1% to 3% for the jets considered in this analysis.

Unclustered Energy: Energy deposits in the detector not associated to muons, electrons, photons or jets with $p_{\mathrm{T}} > 10$ GeV are summarized as unclustered energy. This quantity is not considered

[1] In $Z \to \mu\mu\gamma$ the photon is radiated from one of the muons.

separately in the analysis, but a systematic shift would directly influence the reconstructed MET. An uncertainty of 10% on the total unclustered energy in the event [283] is assumed and introduced on the event's MET.

Missing Transverse Energy: Applying a shift to the four-momenta of individual objects, as described above, will alter the vector sum of all objects in the event. Consequently, the respective influence is propagated to the event's MET. This is performed individually for each object type when the corresponding upwards/downwards shift is applied. Furthermore, an official CMS tool [284] is used to assess the influence on MET of energy and momentum shifts of all objects in the event, not only those fulfilling the full analysis requirements.

10.2 Reconstruction and Selection Efficiency

Although all reconstruction algorithms and object selection criteria undergo extensive verification, slight differences still occur between their performance on measured data and on simulated events. To account for such offsets, reasonable uncertainties are derived from dedicated studies performed to determine the efficiency of the used identification criteria. No systematic uncertainty on the reconstruction efficiency for jets or MET is considered since the observed difference is below 1% for jets reconstructed using the PF algorithm [285]. In each region of a given event class distribution the relative uncertainty on the difference between reconstruction and selection efficiencies given in data and MC is calculated from the quadratic sum over the number N_α of reconstructed objects of type α multiplied with the corresponding object dependent systematic uncertainty $\sigma_{\alpha,\mathrm{Eff}}$:

$$\sigma_{\mathrm{SM}}^{\mathrm{Eff}} = \sqrt{\sum_{\alpha} N_\alpha^2 \cdot \sigma_{\alpha,\mathrm{Eff}}^2} \,. \tag{10.3}$$

This value is multiplied with the total number of events expected from the MC simulation, yielding the total reconstruction and selection efficiency uncertainty in each region. The estimated value is also assumed to cover the small effects of the trigger efficiency given at each trigger's plateau.

To determine the identification efficiency for the individual object types reconstructed in the event and selected with the criteria given in Sec. 9.2, the so-called *tag-and-probe* method is used [286]. Based on the production of same type object pairs in the decay of a known mass resonance, e.g., the Z boson, one of the objects, the *tag*, is selected under the full identification criteria being studied. For the second object, the *probe*, the criteria under investigation are loosened. Moreover, the invariant mass constructed from the tag and probe object is required to reproduce the resonance mass. Finally, the desired efficiency measure is determined from the number of probe objects passing the loosened selection criteria divided by the number of probe objects selected by the invariant mass requirement.

Muons: Relying on the tag-and-probe method using muons with 45 GeV $\geq p_\mathrm{T} \geq$ 300 GeV from Z boson decays, a study of the muon identification efficiency shows the High-p_T muon selection to be very efficient at 93.6% to 94.7%, depending on the considered pseudorapidity range [287]. Moreover, the corresponding MC-to-data scale factors of 0.991 to 0.998 show the simulation to provide a very good estimate of the muon efficiency measured in data. To cover the small difference between data and simulation a systematic uncertainty of $\sigma_{\mu,\mathrm{Eff}} = 1\%$ is introduced in the MUSiC analysis.

Electrons: Considering Z boson decays into electrons, reconstructed with a double-electron trigger at $E_\mathrm{T} > 33$ GeV (analysis level requirement of $E_\mathrm{T} \geq 35$ GeV) and the GSF algorithm, the tag electron must additionally fulfill the HEEP criteria. For transverse energies above 100 GeV the efficiency curve reaches a plateau of over 90% [288]. Throughout the entire E_T range, MC-to-data

scale factors close to unity are determined, where the strongest deviation is given at 0.983 for probes in the barrel region with $E_{\mathrm{T}} \geq 100$ GeV. In consequence, a generous systematic uncertainty of $\sigma_{\mathrm{e,Eff}} = 2\%$ is assigned for this quantity.

Photons: The Z boson decay into di-electron pairs is also used to determine the selection efficiency of the Simple Cut Based photon identification criteria [289]. In addition, events from the Z boson decay to $\mu\mu\gamma$ are also investigated to obtain an efficiency scale factor for the conversion safe electron veto. In total MC-to-data scale factors are close to unity ranging from 0.978 to 1.000 for photons reconstructed in the barrel region. To also account for the different choice of electron veto, a conservative systematic uncertainty of $\sigma_{\gamma,\mathrm{Eff}} = 3\%$ is used for this analysis.

10.3 Misidentification Rate

Throughout this analysis, modeling of misidentified particles is taken from MC simulation and an uncertainty is estimated to cover the difference to measured data. For each considered particle type a matching to the generator level particle is performed within a cone of $\Delta\mathrm{R} = 0.2$ around the reconstructed particle. The number $N_{\alpha,\mathrm{Mis}}$ of particles on reconstruction level lacking a match is multiplied with a corresponding uncertainty $\sigma_{\alpha,\mathrm{Mis}}$. The event's total uncertainty estimate is:

$$\sigma_{\mathrm{SM}}^{\mathrm{Mis}} = \sqrt{\sum_{\alpha} N_{\alpha,\mathrm{Mis}}^2 \cdot \sigma_{\alpha,\mathrm{Mis}}^2} \,, \tag{10.4}$$

where contributions from an object type α are seen as fully correlated amongst each other and fully uncorrelated to contributions from other object types. Here, cases are considered for which objects are incorrectly identified as muons, electrons or photons. Specifically, it is irrelevant which other object is misidentified as the object under consideration, i.e., effects of jets misidentified as photons as well as electrons misidentified as photons are summarized in one value σ_α, obtained from detailed studies outlined below. Misidentification of jets or MET as other objects is not considered.

All event class distributions are reevaluated under an upwards and downwards shift, where a weight $w_{\mathrm{up/down}}^{\mathrm{Mis}} = 1 \pm \sigma_{\mathrm{SM}}^{\mathrm{Mis}}$ is evaluated using Eq. (10.4) and applied to every event. Thus, the fraction of events containing the calculated number of misidentified particles in simulation is varied according to the difference in misidentification rate observed between data and MC simulation. In each event class the histograms resulting from upwards and downwards shifts are used to calculate a symmeterized total uncertainty on the number of events as described previously in Sec. 10.1.

Muons: Non-prompt muons, not originating from the primary interaction vertex, can be produced in pion decays and other hadronic processes. Furthermore, so-called *punch-through* hadrons can reach the muon detectors and lead to a signal misidentified as a muon. In both cases the misidentification probability itself is rather small. A study of pions from K_s^0 decays and protons from Λ decays ($p_{\mathrm{T}} > 4$ GeV for the pion or proton track) quantities the rate of hadrons misidentified as muons within the applied tight working point of the Baseline muon selection (see Sec. 9.2.2) to around $\mathcal{O}(10^{-3})$. Good agreement is seen between data and MC simulation for both hadron types. Given in bins of the hadron p_{T} and η measurement and simulation show a maximum difference of around 50% of the simulation estimate [290]. Relying on a slightly looser muon selection, the number of misidentified muons in jet triggered events is determined for both data and MC simulation by [291]. The disparity between the two estimates is at up to 50%, depending on the considered SM processes. In summary, an uncertainty of $\sigma_{\mu,\mathrm{Mis}} = 50\%$ is assigned to the MUSiC analysis workflow for covering the difference in muon misidentification between data and simulation.

Electrons: Hadronic jets can also mimic signatures otherwise given by isolated electrons, e.g., through the conversion of a radiated photon into an electron pair. Though the misidentification rate

of electrons identified under the HEEP selection (see Sec. 9.2.3) is well below 10% for most electron transverse momentum ranges and regions of the detector [207, 267], the difference observed between data and simulation can become rather large. Motivated by a study of the electron misidentification rate in data and simulation given in [291] an uncertainty estimate of $\sigma_{\mathrm{e,Mis}} = 100\%$ is applied to account for the observed disparity.

Photons: As described in Sec. 9.2.4, the pixel track seed veto efficiently rejects photon candidates to which a charged track may be associated. Nevertheless, a small fraction of electrons misidentified as photons remains. For data events containing two photon candidates, of which one does not pass the electron veto, events in which the second photon is also given by a misidentified electron are observed at a rate of $(1.57 \pm 0.10)\%$. The corresponding MC simulation estimate is only half as large at $(0.80 \pm 0.16)\%$ [292]. A similar MC-to-data scale factor of 1.46 ± 0.20 is determined in [293]. Furthermore, hadronic jets of π^0 or η mesons can also be misinterpreted as isolated photons. Given in bins of photon p_{T} the misidentification rate of jets as photons ranges from about 0.1 at low p_{T} to under 0.05 at $p_{\mathrm{T}} > 100$ GeV [294]. Overall, differences are rather large when comparing the rate in data to simulation and well within 100% of the value given by simulation. For the MUSiC analysis a combined estimate of the difference between the photon misidentification in data and in MC simulation is assumed as $\sigma_{\gamma,\mathrm{Mis}} = 100\%$, covering contributions from electrons and jets.

10.4 Charge Misidentification Rate

The meticulous task of assigning the correct charge sign to leptons, even at high transverse momentum, becomes crucial when classifying events based on their net charge of electrons and muons. A systematic uncertainty is estimated individually for both lepton types to cover differences in the charge misidentification probability encountered between data and simulation. Both are applied to the analysis following the procedure established above for the object type misidentification.

Muons: Dedicated cosmic-ray muon events recorded in 2008 are used to determine the charge misidentification rate for muons in CMS. By combining tracker and muon system measurements the rate is reduced to around 1% for muons reconstructed in the barrel with $p_{\mathrm{T}} \approx 500$ GeV and below 0.1% for $p_{\mathrm{T}} < 100$ GeV [116]. A preliminary update of this study, based on cosmic-ray muon data taken in 2010, gives an estimate for the muon p_{T} dependent charge misidentification rate in data as well as simulation [295]. For global muons the rate itself is found to be 0.001% to 1%, where a difference of about 50% is found in the TeV range between data and simulation. This value is used as a conservative uncertainty estimate.

Electrons: As described in [267], three separate methods are available to measure an electron's charge based on the measured curvature of, respectively, the GSF track, the KF track matched to the GSF track, as well as the supercluster position given by the vectorial difference in ϕ of the vector spanning from the beam spot to the supercluster position and the vector between the beam spot and the first GSF track hit. The final estimate is based on the *majority method*, i.e., the assigned charge is taken from the two out of three methods yielding the same result. The rate of electrons that are assigned with an incorrect charge, especially as a result of photon conversions, is effectively reduced from 10% to 1.5%. A comparison of the rate in data and MC is found in [296] and is the basis for the systematic uncertainty of 20% assumed for the analysis presented here.

10.5 Integrated Luminosity

The *Pixel Cluster Counting* method, as summarized in Sec. 5.3, is used to determine the amount of integrated luminosity collected within the analyzed collision data. Following Eq. (8.2), this value is used to scale the amount of simulated SM events accordingly. Thus, the precision of the luminosity

estimate of 2.6% [109] is taken into consideration for all simulated processes. After combining all SM contributions in a region, the total expected number of events is multiplied with the relative luminosity uncertainty, retaining the effect on the total event yield.

10.6 Total Cross Section of Standard Model Processes

Though the total cross sections of physics processes are independent of characteristics of the experiment, they are susceptible to model assumptions made during the calculation. Specific simplifications can lead to non-conformance with respect to the actually observed process probabilities. Apart from POWHEG, all generators used for this analysis produce events at LO in perturbation theory. To provide a better modeling of the underlying physics processes calculations at higher order in α_s are used whenever accessible. A flat[2] k-factor is determined from the total yield of the process, which helps adjust the overall normalization but cannot account for differences in the shape of a distribution. Moreover, a k-factor does not reproduce the cross section to an exact precision, as it is limited by the considered orders in perturbation theory. In most cases the k-factor only varies slowly throughout considered kinematic variables, but for specific processes a strong dependence may be found. At times the higher order cross section is not determined with exactly the same phase-space assumptions or PDF sets as used during generation.

To cover the limited precision of the cross section estimates, a conservative set of systematic uncertainties is introduced into the analysis based on the order in α_s of QCD calculations:

- 50% for all processes known to LO,
- 10% for electroweak processes known to NLO and approximate NNLO (NNLL),
- 5% for electroweak processes known to NNLO.

Applied individually for each SM process, the uncertainties are chosen to reflect the level of knowledge given by the used k-factor. Within one generated process, the same (higher-order) cross section and uncertainty are assumed over all event classes. Three exceptions to the values stated above are given for specific SM processes, leading to uncertainties of:

- 30% for H_{T} binned W+jets, since the NNLO k-factor shows an H_{T} dependence, but is assumed to be flat (see Sec. 11.2.2 for details),
- 10% for low mass DY, since the flat NNLO k-factor is only calculated for $M_{l^{\pm}l^{\pm}} > 50$ GeV,
- 10% for $\mathrm{t\bar{t}}$, following the uncertainty of the cross section calculation given in [219].

10.7 Parton Distribution Functions and α_s

As summarized in Sec. 8.2.3 one specific PDF set is considered when generating events for each MC process. In general, the choice of PDF influences the number of events predicted to contribute to a specific final state. A very precise but extremely computing intense way of determining the systematic influence of the choice of PDF set is to rerun the generation step with different PDF sets, especially ones from other PDF groups. Instead, to save resources, an event-by-event reweighting approach is used to establish an estimate of this systematic uncertainty on the number of events in a given event class. A weight is determined for each event based on alternative PDF sets from different groups. The precise procedure depends on how each group estimates the PDFs and corresponding uncertainties on the fit values. Following recommendations of the PDF4LHC Working Group, as given in a practical summary [297] based on [298, 299], the methods of determining the analysis

[2] Here, "flat" refers to a constant factor throughout all kinematic variables.

susceptibility to variations of PDF sets and also α_s within one set is implemented with the help of modules supplied by the **L**es **H**ouches **A**ccord **PDF** (LHAPDF) library [300].

In the Hessian approach carried out by the CTEQ and MSTW groups, all free parameters of the resulting fit to experimental data are varied upwards and downwards to account for the fit's uncertainty in so-called PDF *members* or error sets. The *best fit* or central set is the one with the smallest χ^2 value. An uncorrelated set of parameters is obtained through diagonalizing the Hessian matrix of second derivatives of χ^2, resulting in N_{PDF} eigenvectors. For CTEQ, the $N_{\text{PDF}} = 52$ eigenvector sets from `CT10` [236] are used, whereas $N_{\text{PDF}} = 40$ sets are given by the `MSTW2008` [237] variations. Separately for each group, the uncertainty on an observable X is given as:

$$\sigma^{\pm}_{\text{Hess}}(\text{PDF}) = \frac{1}{C} \cdot \sqrt{\sum_{i=1}^{N_{\text{PDF}}} \left(\max\{\pm X_i^{+} \mp X_0,\ \pm X_i^{-} \mp X_0,\ 0\}\right)^2} \,. \tag{10.5}$$

Here, X_0 is the observable found under the central set and $X_i^{\pm}$ are obtained from the parameter variations. The "$+''$ component of Eq. (10.5) extracts the strongest increase of the observable given by either upwards or downwards shifts of the fit parameters. Analogously, the "$-''$ component sums up contributions that lead to the strongest decrease of X.

Furthermore, the influence of variations to α_s is evaluated for the central PDF set in which α_s at the Z mass is taken to be 0.118 for CTEQ and 0.120 for MSTW. A difference of $\Delta\alpha_s = 0.0012$ is defined as one Gaussian standard deviation of α_s. The resulting uncertainty is calculated via:

$$\sigma^{\pm}_{\text{Hess}}(\alpha_s) = \frac{1}{C'} \cdot \left(X^{\pm}_{\alpha_s} - X_0\right), \tag{10.6}$$

using dedicated PDF sets $X^{\pm}_{\alpha_s}$ including an upwards and downwards shift. Due to different choices of uncertainty intervals made by the different groups, coefficients C and C' must be introduced to rescale all uncertainties to 68% confidence level.

Estimates from both PDF fit parameter and α_s variations are then combined to a total uncertainty for each PDF group, assuming them to be uncorrelated:

$$\sigma^{\pm}_{\text{Hess}}(\text{PDF}, \alpha_s) = \sqrt{\sigma^{\pm}_{\text{Hess}}(\text{PDF})^2 + \sigma^{\pm}_{\text{Hess}}(\alpha_s)^2} \,. \tag{10.7}$$

The NNPDF group follows an MC approach in which α_s is not only varied for the central set, but simultaneously for all parameter estimates. For the central fit α_s is set to 0.119 at the Z boson mass and to values from 0.116 to 0.122 in increments of 0.001 for the so-called *replicas*, of which 100 each are available for every α_s value. Assuming a Gaussian probability distribution around the central α_s value different amounts of replicas are chosen depending on their distance from the central value. Specifically, a total of $N_{\text{PDF}} = 50$ PDF sets are chosen for the seven α_s values, distributed as: (1, 4, 12, 16, 12, 4, 1). Under this assumption the NNPDF systematic uncertainty is determined by:

$$\sigma_{\text{NNPDF}}(\text{PDF}, \alpha_s) = \sqrt{\frac{1}{N_{\text{PDF}} - 1} \sum_{i=1}^{N_{\alpha_s}} \sum_{j=1}^{N_{\text{PDF}}^{\alpha_s^i}} (X_{i,j} - X_0)^2} \,. \tag{10.8}$$

Finally, the total uncertainty on the number of events in each bin of an event class distribution is then evaluated as the combination of results from Eq. (10.7) and Eq. (10.8):

$$\begin{aligned} \sigma^{\text{PDF}} = \frac{1}{2} \cdot (&\max\{X_{0,\text{CTEQ}} + \sigma^{+}_{\text{CTEQ}},\ X_{0,\text{MSTW}} + \sigma^{+}_{\text{MSTW}},\ X_{0,\text{NNPDF}} + \sigma_{\text{NNPDF}}\} \\ &-\min\{X_{0,\text{CTEQ}} - \sigma^{-}_{\text{CTEQ}},\ X_{0,\text{MSTW}} - \sigma^{-}_{\text{MSTW}},\ X_{0,\text{NNPDF}} - \sigma_{\text{NNPDF}}\}), \end{aligned} \tag{10.9}$$

giving a conservative estimate by symmeterizing the strongest difference in the number of events.

10.8 Number of Simulated Events

In each region the SM event yield is estimated by N_{SM} weighted MC entries, which follow a Poisson distribution $\mathcal{P}(N_{\mathrm{SM}};\mu)$ around the unknown true expectation value μ, as given in Eq. (7.11). The total statistical uncertainty in each region, as expressed in Eq. (7.17), is calculated via the linear summation of events from one SM process, under consideration of the generator weight introduced in Sec. 8.2.6, and the uncorrelated sum over contributions from different SM processes. This uncertainty source is the only one taken to be uncorrelated throughout different bins of the distribution. Thus, a separate random value is drawn for each bin when generating pseudo-experiments (see Sec. 7.3.5). Following the region veto method to mitigate issues of an insufficient number of generated events (see Sec. 7.3.3), the relative statistical uncertainty $\sigma^{\mathrm{stat}}_{\mathrm{SM}}/N_{\mathrm{SM}}$ is limited to a maximum of 60%.

10.9 Summary of Systematic Uncertainties

To provide a concise overview of all systematic uncertainties considered by the MUSiC analysis, the individual contributions and assigned values are summarized in Tab. 10.1.

Systematic	**Value**	**Additional Information**
Energy and Momentum Scale		
μ	5% per 1 TeV	
e, γ	barrel: 0.6%, endcap: 1.5%	
jets	3% to 5%	jet η and p_{T} dependent
Unclustered Energy	10%	
MET	varies	given by objects in event
Reconstruction and Selection Efficiency		
μ	1%	
e	2%	
γ	3%	
Misidentification Rate		
μ	50%	
e	100%	
γ	100%	
Charge Misidentification Rate		
μ	50%	
e	20%	
Integrated Luminosity	2.6%	
SM Cross Sections	5% to 50%	process dependent
Parton Density Functions	varies	reweighting method
Number of MC events	varies	relative value $< 60\%$

Table 10.1: Summary of systematic uncertainties and assigned values taken into consideration in this analysis. Details of how these values are applied are given in corresponding sections of Ch. 10. The statistical uncertainty on the number of events in a given SM MC data set is treated as a systematic uncertainty.

Part IV

Analysis Results

Chapter 11
Search Results

The following chapter summarizes results of the MUSiC analysis of CMS data taken in 2012 at a center of mass energy of $\sqrt{s} = 8$ TeV, considering electron and muon triggered events. A total of $8.8 \cdot 10^7$ data events and $(9.1 \pm 0.7) \cdot 10^7$ weighted SM MC simulation events, with corresponding combined systematic uncertainty, fulfill all selection criteria addressed in Sec. 9. These are sorted into 337 exclusive, 341 inclusive and 321 jet-inclusive[1] event classes by the classification algorithm (see Sec. 7.2). About 45% of these final states contain measured data events, while only the simulated SM expectation contributes to the others. For each event class type approximately 70% of the event classes contain electrons and around 70% contain muons. Photons are found in about 40% of event classes and around 45% show MET $>$ 50 GeV.

Each event class type is analyzed separately by the search algorithm as described in Sec. 7.3. In a first step, the total event yield of each event class is used to calculate a single p-value for a data vs. SM MC comparison. Afterwards, the RoI is determined for three kinematic distributions in every event class, enabling a more detailed investigation of observed deviations.

Members of the CMS Collaboration can access the full set of MUSiC scan results on an internal website at:

https://cms-project-music.web.cern.ch/cms-project-music/

11.1 Scan of Total Event Yield

An initial impression of the data-MC comparison is gained by treating each event class as a single-bin counting experiment, considering only the total event yield and not the distribution of a kinematic property. Such an overview can help rapidly identify issues concerning the overall normalization of the simulated SM processes, where the runtime is short since every event class consists of only a single region. Fig. 11.1 shows the 50 exclusive event classes with the highest data yield. The observed data is shown by a black marker and individual processes from the SM prediction are represented by colored bars, where the total systematic uncertainty is given by a gray crosshatched band. Each bin is labeled with the corresponding event class name (bottom) and p-value (top). In this scanning method the LEE (Sec. 7.3.4) does not apply since the entire distribution is condensed into a single search region. Accordingly, the p-value represents the final measure of significance.

Five orders of magnitude are covered by the event classes shown in Fig. 11.1 and good agreement between the measured data and SM expection is found throughout. Resulting from the substantially higher p_{T}-threshold of the single-electron trigger (see Sec. 9.1.3) the 1μ exclusive event class (first bin of Fig. 11.1) contains around 250 times more data events than 1e (bin 15). Both double-lepton triggers have a comparable p_{T}-threshold, so the amount of events contributing to 2μ (bin 3) and 2e (bin 4) is about the same.

Event classes containing precisely two same flavor leptons and at least 50 GeV of missing transverse energy (2e+MET, 2μ+MET in bins 17 and 19) show a noticeable excess with respect to the SM prediction, which is however fully covered by the systematic uncertainty, in this case mainly the one stemming from unclustered energy contributions in the event. Associated event classes with one jet $2l$+1jet+MET (bins 28 and 29) as well as two jets $2l$+2jets+MET (bins 33 and 36)

[1] Events with six and more jets are summarized in a single event class (see Sec. 7.2.1) which is treated as an exclusive event class. The corresponding identical jet-inclusive class is no longer considered. This is reflected in the stated numbers: More exclusive than jet-inclusive event classes are considered.

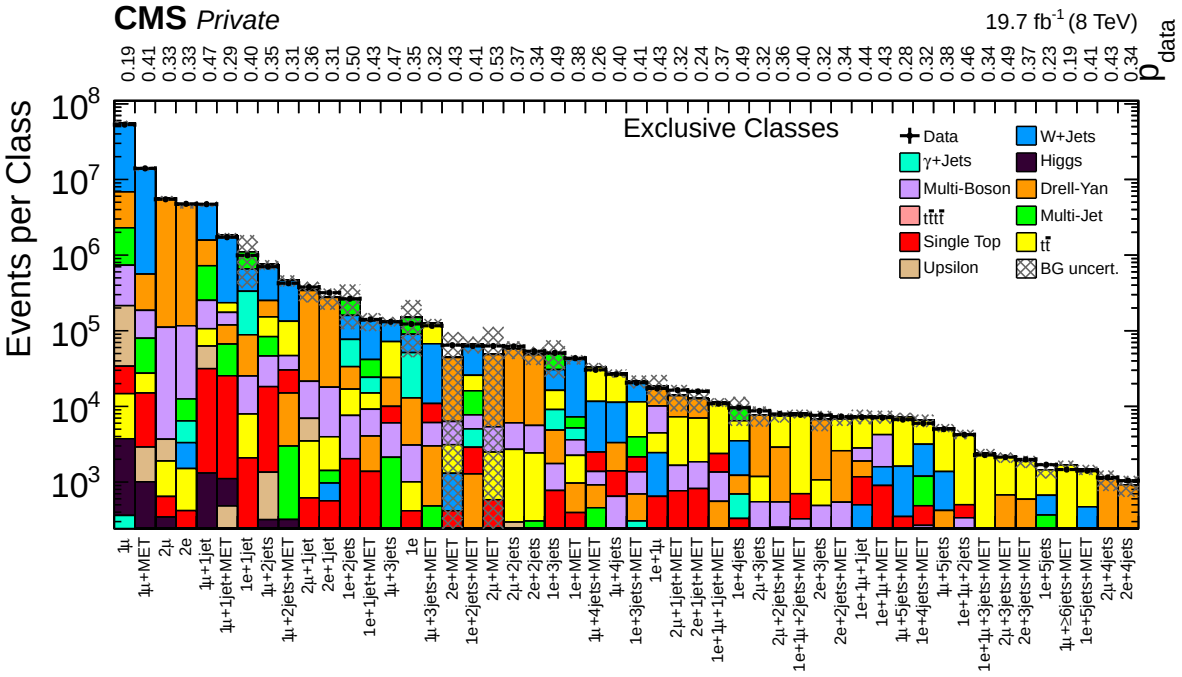


Figure 11.1: Overview of 50 exclusive event classes sorted by their total data event yield.

show a similar trend. This observation is discussed in Sec. 11.2.2, taking kinematic distributions and additional studies into consideration.

As explained in Sec. 7.3.3 a p-value is only calculated in regions with sufficient SM MC prediction. As a consequence, respectively about 170 exclusive, inclusive and jet-inclusive event classes are not considered during the further statistical analysis. Of these, only the seven event classes listed in Tab. 11.1 contain data. Due to overlap, altogether only five different data events contribute to these event classes.

Event Class	N_{data}	$N_{\text{SM}} \pm \sigma_{\text{SM}}$	Contributing SM Processes
2e+1γ+2jets+MET	3	0.89±0.61	Multi-Boson, $t\bar{t}$, DY, Higgs
2e+1γ+3jets	1	0.74±0.57	Multi-Boson, DY, $t\bar{t}$, Higgs
2e+1γ+3jets+X	2	0.84±0.59	Multi-Boson, DY, $t\bar{t}$, Higgs, Single Top, $t\bar{t}t\bar{t}$
2e+1γ+3jets+Njet	2	0.75±0.57	Multi-Boson, DY, $t\bar{t}$, Higgs, Single Top
2e+1γ+4jets	1	0.015±0.019	Multi-Boson, Higgs, Single Top
2e+1γ+4jets+X	1	0.025±0.022	Multi-Boson, Higgs, $t\bar{t}t\bar{t}$, Single Top
2e+1γ+4jets+Njet	1	0.015±0.019	Multi-Boson, Higgs, Single Top

Table 11.1: Details of all exclusive, inclusive and jet-inclusive event classes containing data but only insufficient SM MC prediction. Total data and SM event yields are given. Contributing processes from SM simulation are listed in order of decreasing event yield.

All event classes given in Tab. 11.1 show a similar combination of objects, two electrons, one photon and two or more jets. In event classes with more than two jets no substantial amount of MET is present. Several collision events contribute to more than one of the listed event classes. Corresponding visualizations and details of all five data events can be found in Fig. 11.2 through Fig. 11.6. Here, tracks reconstructed in the tracker are shown by green lines, where a minimum requirement of $p_{\text{T}} = 5$ GeV is imposed to allow a clearer view of the event. In all cases the invariant mass of either the two electrons or both electrons and the photon, possibly originating

from FSR, comes close to the on-shell Z boson hypothesis. Throughout this document the stated uncertainties on kinematic event variables are determined using the per object resolutions given in Sec. 7.2.3 which can been seen as a rough estimate.

As listed in Tab. 11.1, not only exclusive but also (jet-)inclusive event classes are disregarded for the three and four jet case but not for the two jet case. Corresponding event classes with five jets or more do not contain any data events and only a very small contribution of under 0.0005 events is expected from the SM. In consequence these additional events do not add a sufficient MC prediction to the already very low event count. On the other hand enough events are included in 2e+1γ+2jets+MET+X and 2e+1γ+2jets+MET+Njet to ensure a valid MC description. In both cases moderate p-values of 0.11 are achieved with three events observed in data and 0.97 ± 0.63 events expected from the SM simulation.

Event classes similar to those in Tab. 11.1, but without an identified photon, i.e., 2e+2jets+MET, 2e+3jets and 2e+4jets, show a much higher overall event count with over 1000 events observed in data. The agreement in these classes is good with p-values of around 0.33, where main SM contributions come from DY and $\mathrm{t\bar{t}}$. Requiring MET < 50 GeV in the two jet case, i.e., 2e+1γ+2jets, leads to an accumulation of multi-boson and DY events. A sufficient SM expectation is given and a moderate significance of $p_{\mathrm{data}} = 0.11$ is evaluated with one event observed in data and 4.5 ± 1.6 given by the glssm MC. In the three and four jet event classes with substantial MET, i.e., 2e+3jets+MET and 2e+4jets+MET, no events are found in data and the low SM expectation is compatible with zero within its systematic uncertainties.

In the scan of kinematic distributions discussed in Sec. 11.2 not all of event classes from Tab. 11.1 are disregarded. In some cases a valid region is found where both the neighborhood criterion (irrelevant when considering the total event yield) and the requirement of a relative statistical uncertainty of below 60% is fulfilled. These event classes are among those with the most significant deviations. Further investigations, centered around modeling the misidentification rate of photons in the SM MC simulation, are discussed in Sec. 11.2.5.

Turning now to event classes where a p-value was calculated, somewhat more excesses (exclusive: 88, inclusive: 98, jet-inclusive: 91) than deficits (79, 71, 70) are observed. This is not surprising, since especially event classes with no data events and a small SM prediction, often compatible with zero expected events, are removed from the statistical analysis by the runtime optimization, as explained in Sec. 7.3.6. Results for the 50 most significant exclusive event classes are shown in Fig. 11.8 and details for the 15 most significant event classes are summarized in Tab. 11.2. The observed data p-values are rather moderate, with the smallest one found in 1e+1γ+MET at $p_{\mathrm{data}} = 0.0015$ corresponding to around $3.0\,\sigma$[2]. This is also one of the most significant event classes in terms of the scan of kinematic distributions and results from detailed studies performed to understand this deviation are presented in Sec. 11.2.5. Overall, many of the most significant deviations appear in final states where only few events are expected and slight excesses are found.

As introduced in Sec. 7.3.7, a global overview in terms of significance for all event classes is given by the distribution of p-values. Per event class up to 10^5 rounds of SM-only pseudo-data are generated to gain a numerical estimate for the distribution of p-values and is shown alongside p-values from the data-MC comparison in Fig. 11.9. As every event class contains only one region, runtime optimizations discussed in Sec. 7.3.6 can be relaxed at this stage of the analysis. Pseudo-data is created for every event class, not only for ones with p-values smaller than 0.3. The observed p-values are in good accordance with the SM expectation. All bins show an agreement within $2\,\sigma$ of the uniform expectation. Overall, the systematic uncertainties assigned to various sources in this analysis seem to be chosen rather reasonably. A large global overestimation of systematic

[2] one-sided Gaussian standard deviation; See footnote 4 on p. 52 in Sec. 7.3.7 for details.

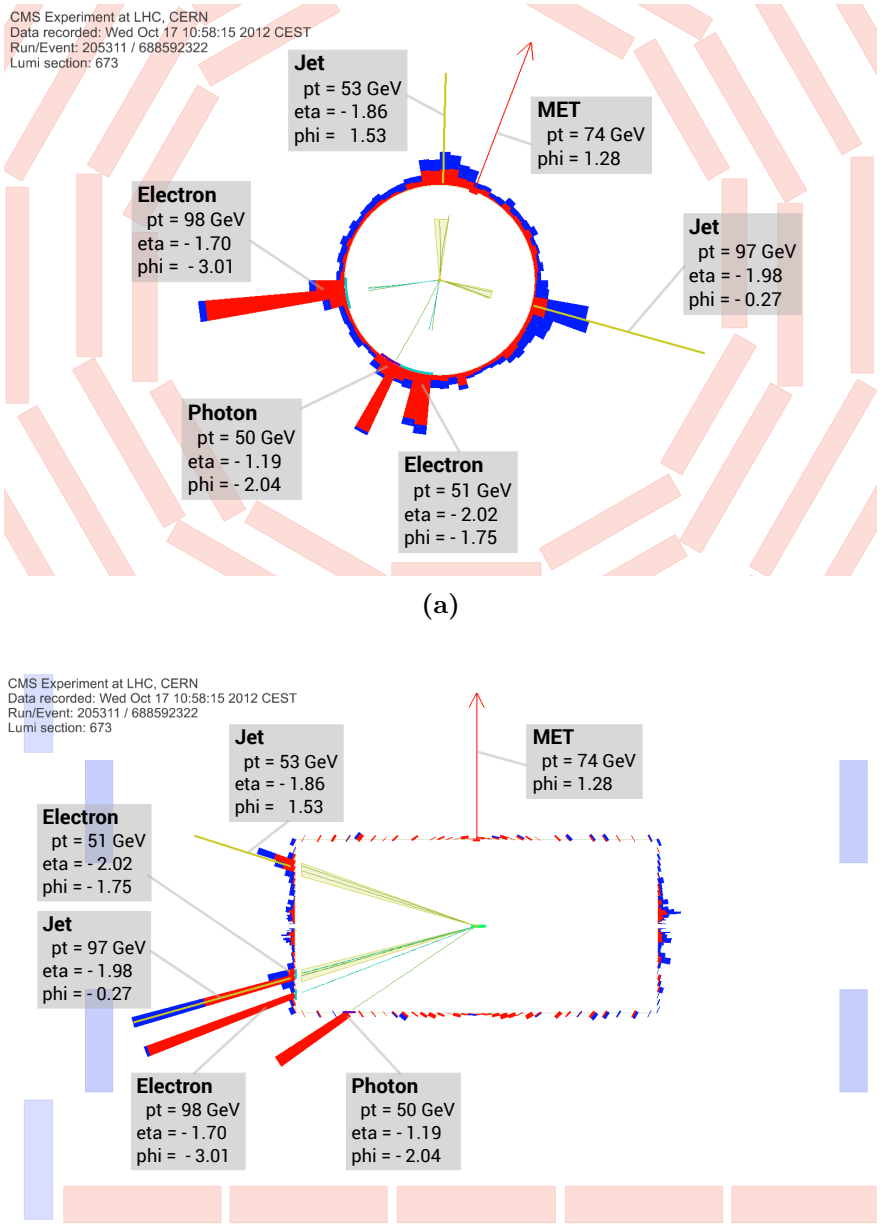
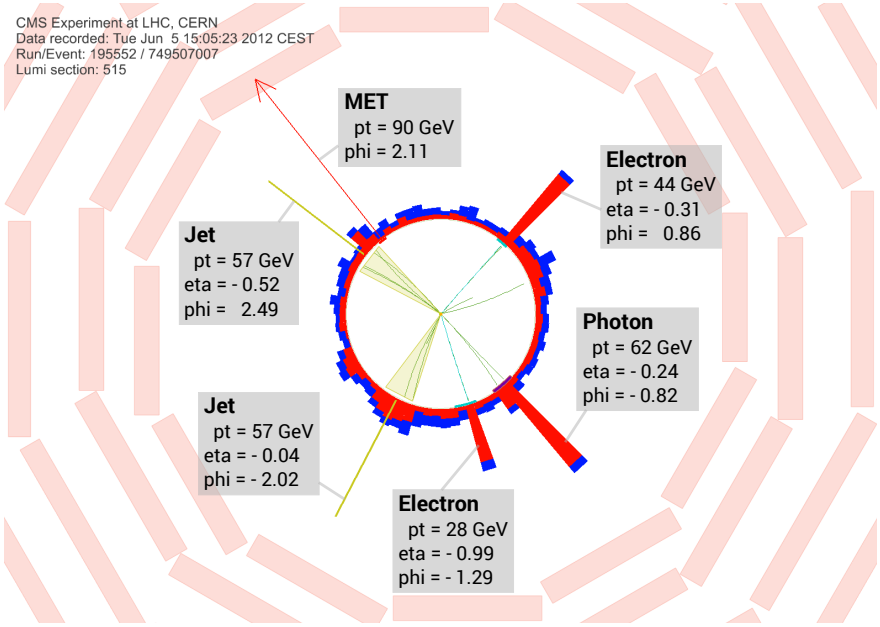
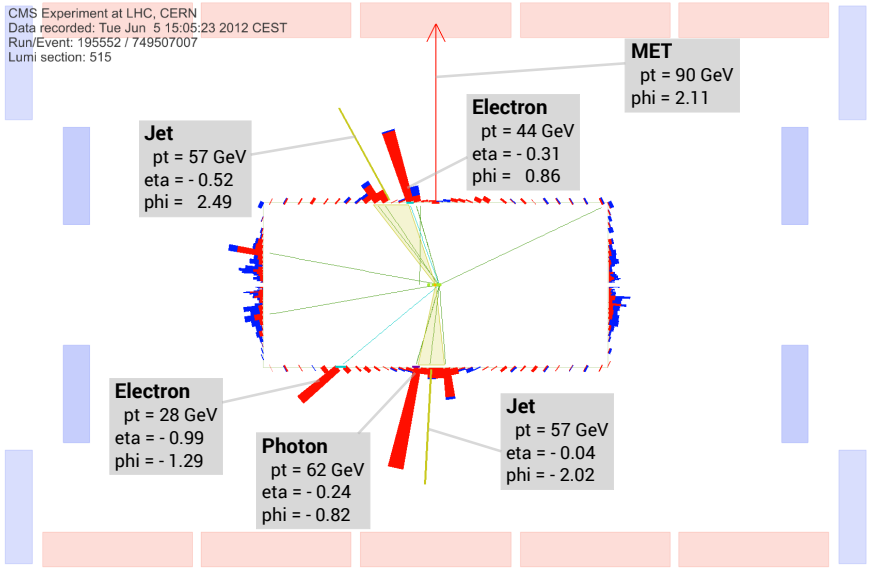


Figure 11.2: 205311:673:688592322 (run:lumi section:event) – double-electron trigger stream
Visualization of the data event contributing to event class 2e+1γ+2jets+MET in the ρ-ϕ (a) and ρ-z (b) view of the CMS detector. Red / blue towers indicate deposits in the ECal / HCal proportional to the collected energy. Both selected electrons (light blue tracks and bars) are not energetic enough to pass the single-electron trigger requirement, although one only falls a few GeV short of the threshold. With an invariant mass of $M_{\mathrm{inv}}(\mathrm{e},\mathrm{e}) = (86 \pm 1)$ GeV, the electrons might originate from the decay of a Z boson, including also the photon yields $M_{\mathrm{inv}}(\mathrm{e},\mathrm{e},\gamma) = (122 \pm 1)$ GeV. Shown by a purple bar, the photon is separated well from both electrons and the two reconstructed jets, which are indicated by yellow cones. Apart from these objects, the event is rather clean. All objects are boosted towards the forward region, both electrons and jets are measured in the endcap region and only little energy is deposited in the barrel. MUSiC kinematic variables of this event are: $\sum |\vec{p}_{\mathrm{T}}| = (423 \pm 19)$ GeV; $M_{(\mathrm{T})} = (423 \pm 19)$ GeV; MET $= (74 \pm 15)$ GeV;

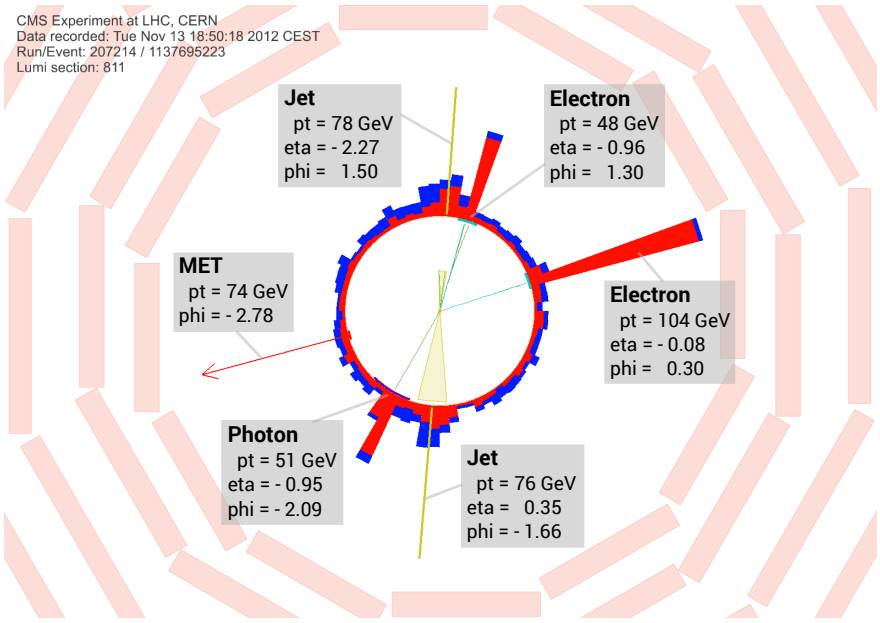


(a)

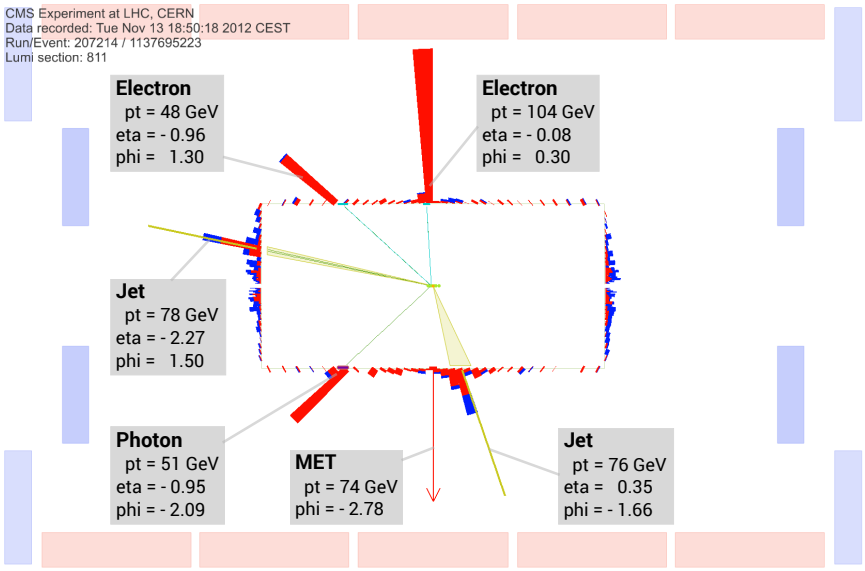


(b)

Figure 11.3: 195552:515:749507007 (run:lumi section:event) – double-electron trigger stream
Visualization of the data event contributing to event class 2e+1γ+2jets+MET in the ρ-ϕ (a) and ρ-z (b) view of the CMS detector. Red / blue towers indicate deposits in the ECal / HCal proportional to the collected energy. Both selected electrons (light blue tracks and bars) are by far not energetic enough to pass the single-electron trigger requirement. With an invariant mass of $M_{\mathrm{inv}}(\mathrm{e},\mathrm{e}) = (67 \pm 1)$ GeV, the electrons do not likely originate from the decay of a Z boson. Yet, the invariant mass of both electrons and the photon comes close to that of the on-shell case $M_{\mathrm{inv}}(\mathrm{e},\mathrm{e},\gamma) = (109 \pm 1)$ GeV. Shown by a purple bar, the photon is separated well from both electrons and the two reconstructed jets, which are indicated by yellow cones. All objects are reconstructed in the barrel and are sufficiently separated from each other. Some additional tracks and calorimetric energy are visible in the endcap regions but do not pass the analysis' object selection criteria. MUSiC kinematic variables of this event are: $\sum |\vec{p}_{\mathrm{T}}| = (338 \pm 17)$ GeV; $M_{(\mathrm{T})} = (338 \pm 17)$ GeV; MET $= (90 \pm 13)$ GeV;

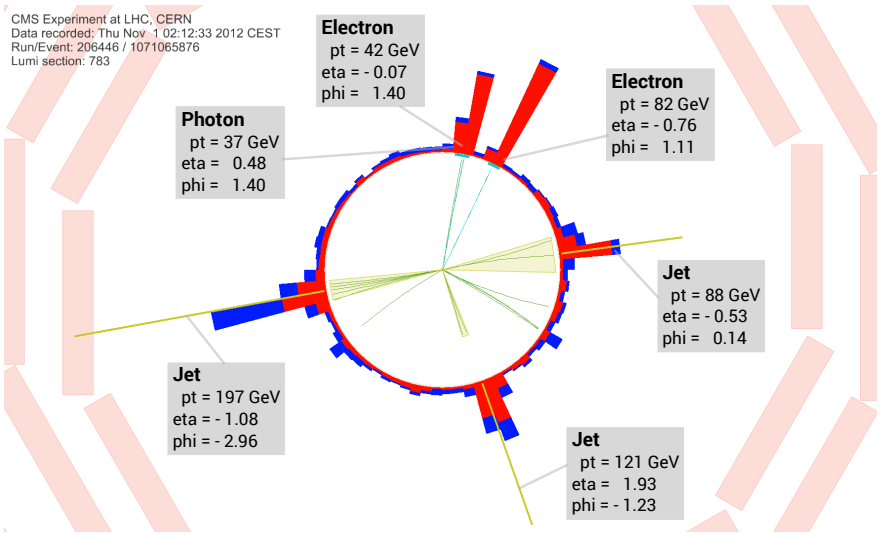


(a)

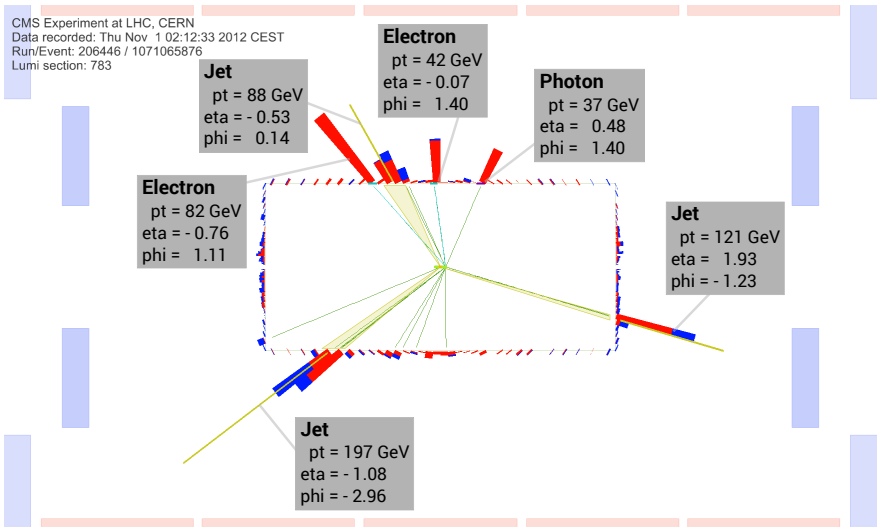


(b)

Figure 11.4: 207214:811:1137695223 (run:lumi section:event) – single-electron trigger stream
Visualization of the data event contributing to event class 2e+1γ+2jets+MET in the ρ-ϕ (a) and ρ-z (b) view of the CMS detector. Red / blue towers indicate deposits in the ECal / HCal proportional to the collected energy. One of the electrons (light blue tracks and bars) has enough p_{T} to be accepted by the single-electron trigger requirements, while the other one is far less energetic. With an invariant mass of $M_{\mathrm{inv}}(\mathrm{e},\mathrm{e}) = (93 \pm 1)$ GeV, the electrons might originate from the decay of a Z boson, including also the photon yields $M_{\mathrm{inv}}(\mathrm{e},\mathrm{e},\gamma) = (203 \pm 1)$ GeV. Shown by a purple bar, the photon is separated well from both electrons and the two reconstructed jets, which are indicated by yellow cones. One of the jets is reconstructed in the endcap region, while all other objects are measured in the barrel. Some additional calorimetric deposits are visible in the endcaps. MUSiC kinematic variables of this event are: $\sum |\vec{p}_{\mathrm{T}}| = (432 \pm 19)$ GeV; $M_{(\mathrm{T})} = (432 \pm 19)$ GeV; MET $= (74 \pm 14)$ GeV;

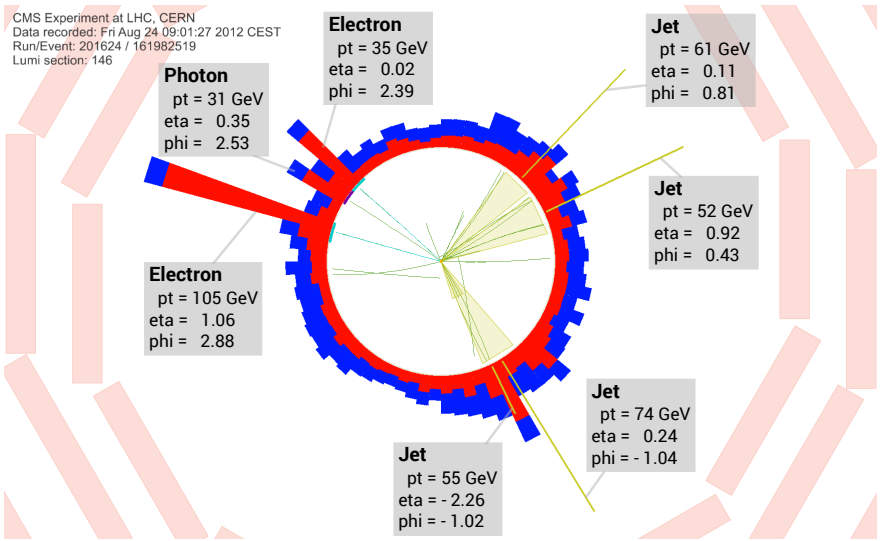


(a)

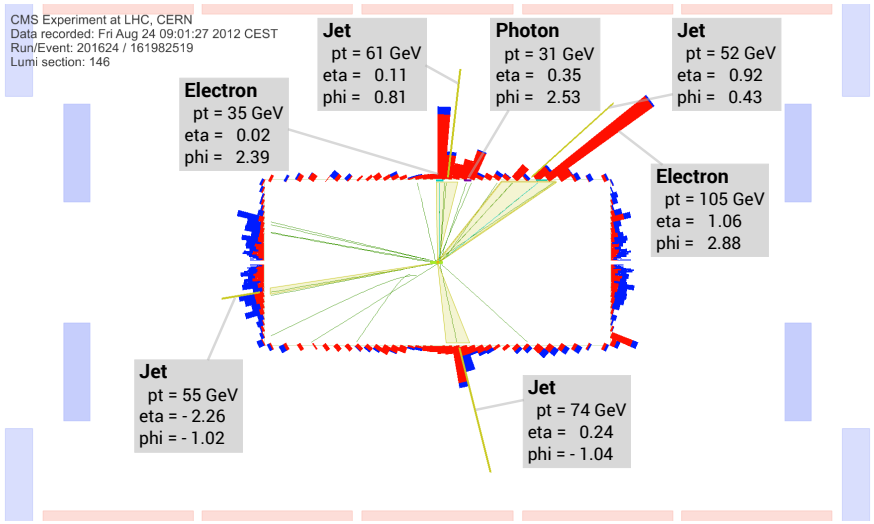


(b)

Figure 11.5: 206446:783:1071065876 (run:lumi section:event) – double-electron trigger stream
Visualization of the data event contributing to event classes 2e+1γ+3jets(+X/+Njet) in the ρ-ϕ (a) and ρ-z (b) view of the CMS detector. Red / blue towers indicate deposits in the ECal / HCal proportional to the collected energy. Both selected electrons (light blue tracks and bars) are not energetic enough to pass the single-electron trigger requirement. With an invariant mass of $M_{\mathrm{inv}}(\mathrm{e},\mathrm{e}) = (45 \pm 2)$ GeV, it is not likely that the electrons originate from the decay of a Z boson. Yet, the invariant mass of both electrons and the photon comes close to that of the on-shell case: $M_{\mathrm{inv}}(\mathrm{e},\mathrm{e},\gamma) = (89 \pm 2)$ GeV. Shown by a purple bar, the photon is hidden behind one of the electrons in the ρ-ϕ view, but as can be seen in ρ-z it is somewhat more forward than the other objects. Three jets, indicated by yellow cones, are reconstructed in different regions of the detector and lead to an overall very balanced event. Some additional tracks and calorimetric energy are visible in the endcap regions but do not pass the analysis' object selection criteria. MUSiC kinematic variables of this event are: $\sum |\vec{p}_{\mathrm{T}}| = (567 \pm 20)$ GeV; $M_{(\mathrm{T})} = (1044 \pm 24)$ GeV; MET $= (4 \pm 16)$ GeV (disregarded in analysis);



(a)



(b)

Figure 11.6: 201624:146:161982519 (run:lumi section:event) – single-electron trigger stream
Visualization of the data event contributing to event classes 2e+1γ+3jets+X(+Njet) and 2e+1γ+4jets(+X/+Njet) in the ρ-ϕ (a) and ρ-z (b) view of the CMS detector. Red / blue towers indicate deposits in the ECal / HCal proportional to the collected energy. One of the electrons (light blue tracks and bars) has enough p_{T} to be accepted by the single-electron trigger requirements, while the other one is far less energetic. With an invariant mass of $M_{\mathrm{inv}}(\mathrm{e},\mathrm{e}) = (71 \pm 2)$ GeV, it is not likely that the electrons originate from the decay of a Z boson. Yet, the invariant mass of both electrons and the photon comes close to that of the on-shell case $M_{\mathrm{inv}}(\mathrm{e},\mathrm{e},\gamma) = (85 \pm 1)$ GeV. Four jets, indicated by yellow cones, are reconstructed in different regions of the detector. Shown by a purple bar, the photon is separated well from all other objects in the event. A high amount of calorimetric energy and many primary vertices are present throughout this event. See Fig. 11.7 for a closer view of the tracker region and more details. Only a small amount of MET is found in the event. MUSiC kinematic variables of this event are: $\sum |\vec{p}_{\mathrm{T}}| = (362 \pm 14)$ GeV; $M_{(\mathrm{T})} = (634 \pm 18)$ GeV (for 2e+1γ+3jets+X(+Njet)); $\sum |\vec{p}_{\mathrm{T}}| = (414 \pm 16)$ GeV; $M_{(\mathrm{T})} = (718 \pm 20)$ GeV (for 2e+1γ+4jets(+X/+Njet)); MET $= (27 \pm 14)$ GeV (disregarded in analysis); Figure (a) is published in [5].

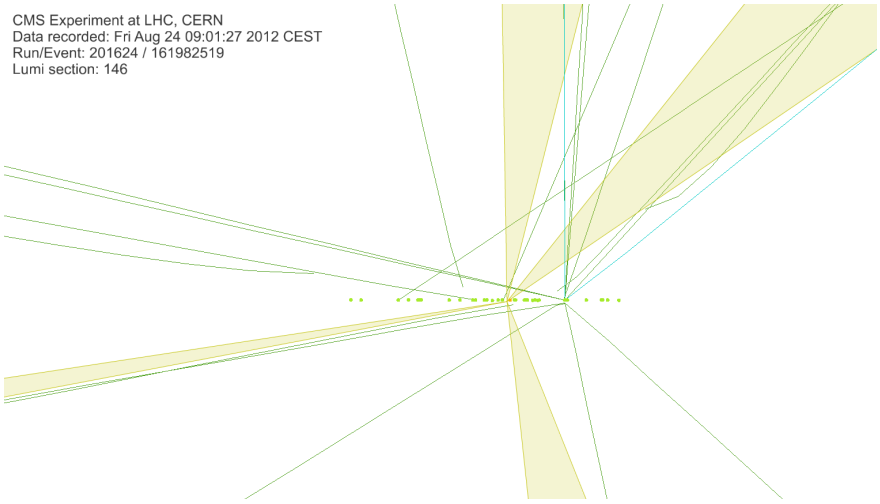


Figure 11.7: Close-up of the data visualization for the event contributing to event classes 2e+1γ+3jets+X(+Njet) and 2e+1γ+4jets(+X/+Njet). In this ρ-z view of the CMS detector around 30 reconstructed primary vertices are visible. Each vertex is shown by a green circle and the nominal beam-spot is given in orange. Such an amount of event pileup is rather typical for data taken at $\sqrt{s} = 8$ TeV (compare to the measured distribution in Fig. 8.4).

effects would distort the distributed p-values observed in the measured data with respect to the ones estimated in SM-only pseudo-experiments. Significantly more event classes with large p-values (first bin of the distribution) and fewer with small p-values would be found in the data than expected from the simulation. Yet, this is not the case here.

Corresponding results for (jet-)inclusive event classes are given in Fig. 11.10 through Fig. 11.13. Many of the most significant exclusive events classes are also significant in these interpretations of the data and have comparable p-values. Yet, the jet-inclusive event class 1e+1γ+MET+Njet associated to the most significant exclusive event class 1e+1γ+MET ($p_{\text{data}} = 0.0015$) has only a moderate significance of $p_{\text{data}} = 0.060$. In total, events from five exclusive event classes are summarized in the jet-inclusive case, 1e+1γ+MET ($p_{\text{data}} = 0.0015$), 1e+1γ+1jet+MET ($p_{\text{data}} = 0.19$), 1e+1γ+2jets+MET ($p_{\text{data}} = 0.073$), 1e+1γ+3jets+MET ($p_{\text{data}} = 0.38$) and 1e+1γ+4jets+MET ($p_{\text{data}} = 0.57$). Large p-values are found in the exclusive classes with one jet or more and diminish the significance found in the zero jet case, resulting in an overall good agreement in the jet-inclusive event class. Similar arguments are applicable for the inclusive event class 1e+1γ+MET+X ($p_{\text{data}} = 0.069$), where even more events are superimposed on the 1e+1γ+MET events since all object types are taken into consideration.

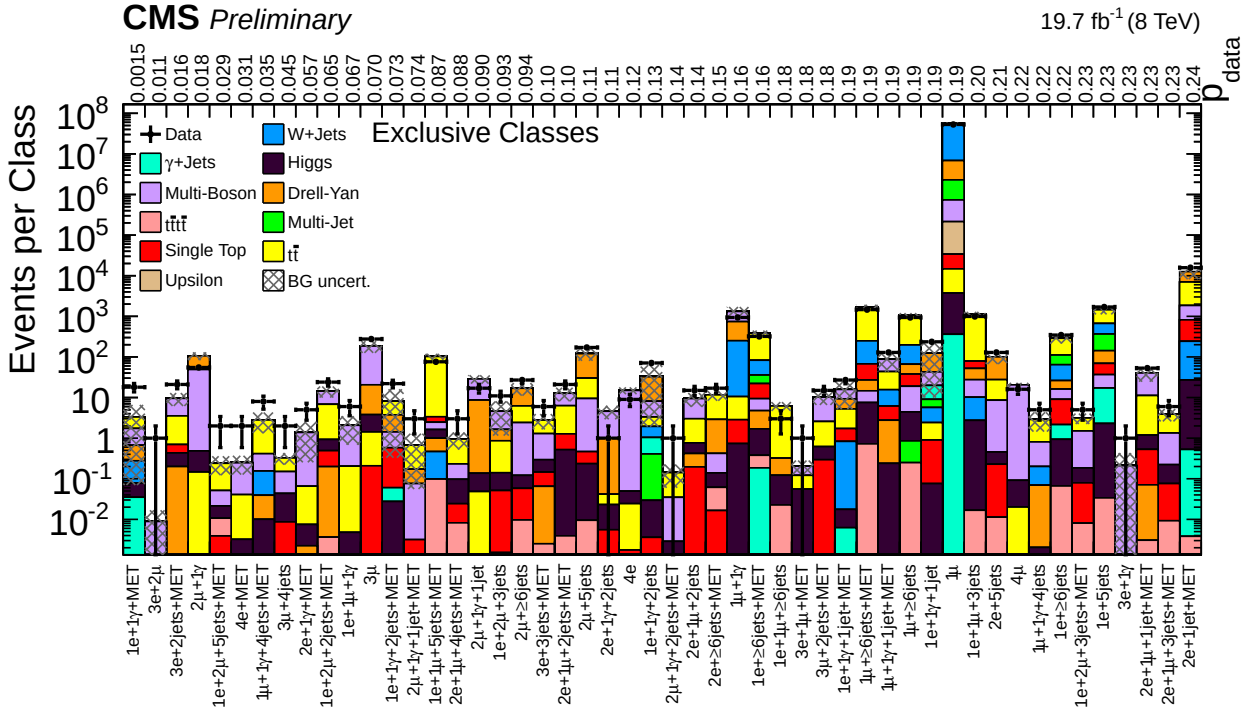


Figure 11.8: The 50 most significant exclusive event classes for the scan of total event yield sorted by increasing p-value, as stated at top of diagram. Details of the 15 most significant ones are given in Tab. 11.2. An identical plot is published in [5].

Event Class	p	N_{data}	$N_{SM} \pm \sigma_{SM}$
1e+1γ+MET	0.0015	18	3.3±3.3
3e+2μ	0.011	1	$(8.9 \pm 7.8) \cdot 10^{-3}$
3e+2jets+MET	0.016	21	9.6±3.3
2μ+1γ	0.018	55	106±23
1e+2μ+5jets+MET	0.029	2	0.24±0.12
4e+MET	0.031	2	0.25±0.12
1μ+1γ+4jets+MET	0.035	8	2.8±1.5
3μ+4jets	0.045	2	0.32±0.10
2e+1γ+MET	0.057	5	1.4±1.1
1e+2μ+2jets+MET	0.065	24	14.3±4.4
1e+1μ+1γ	0.067	6	2.1±1.4
3μ	0.070	276	185±59
1e+1γ+2jets+MET	0.073	22	8.1±7.8
2μ+1γ+1jet+MET	0.074	3	0.67±0.62
1e+1μ+5jets+MET	0.087	76	103±17

Table 11.2: Details of the 15 most significant exclusive event classes from the scan of total event yield as shown in Fig. 11.8.

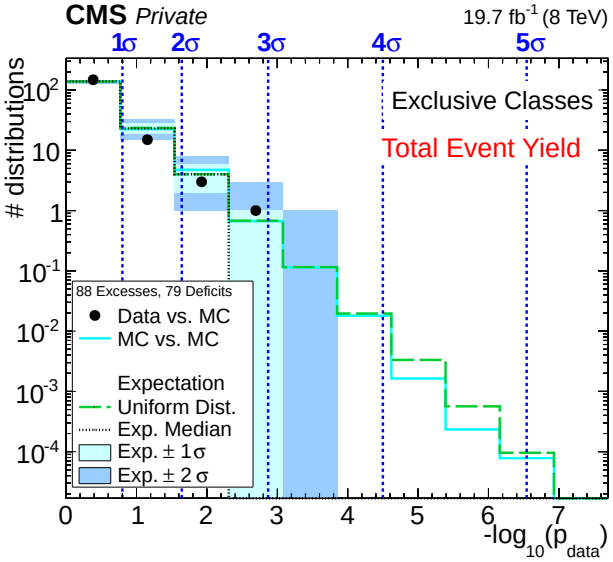


Figure 11.9: Distribution of p-values for exclusive event classes in the scan of total event yield. In the first bin 148 distributions are observed, with $139^{+4}_{-5}(1\,\sigma)^{+\,9}_{-10}(2\,\sigma)$ expected from the SM. A similar plot is published in [5].

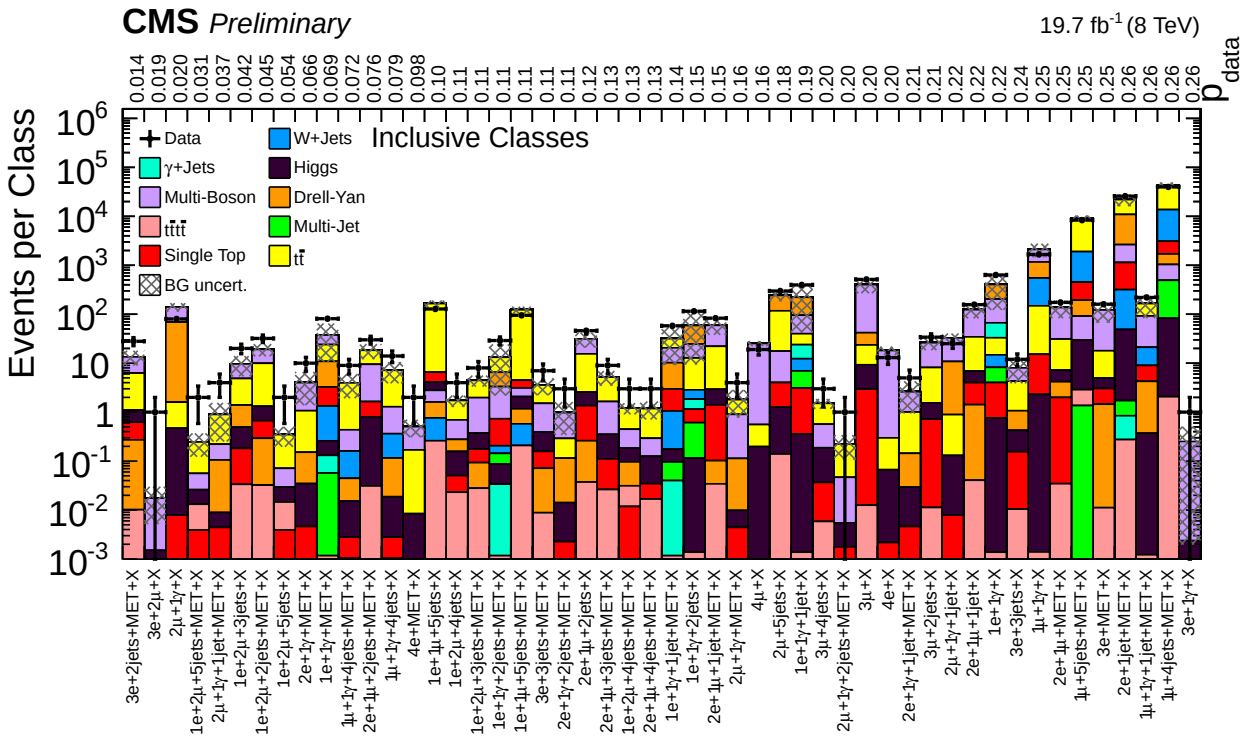


Figure 11.10: The 50 most significant inclusive event classes for the scan of total event yield sorted by increasing p-value, as stated at top of diagram. Details of the 15 most significant ones are given in Tab. 11.3. An identical plot is published in [5].

Event Class	p	N_{data}	$N_{SM} \pm \sigma_{SM}$
3e+2jets+MET+X	0.014	28	13.5±4.5
3e+2μ+X	0.019	1	0.018±0.012
2μ+1γ+X	0.02	80	141±27
1e+2μ+5jets+MET+X	0.031	2	0.25±0.12
2μ+1γ+1jet+MET+X	0.037	4	0.90±0.64
1e+2μ+3jets+X	0.042	20	9.6±4.3
1e+2μ+2jets+MET+X	0.045	32	19.1±5.3
1e+2μ+5jets+X	0.054	2	0.35±0.15
2e+1γ+MET+X	0.066	10	4.0±2.6
1e+1γ+MET+X	0.069	81	38±27
1μ+1γ+4jets+MET+X	0.072	9	4.0±2.0
2e+1μ+2jets+MET+X	0.076	30	18.3±6.2
1μ+1γ+4jets+X	0.079	14	7.2±3.3
4e+MET+X	0.098	2	0.50±0.20
1e+1μ+5jets+X	0.10	128	165±26

Table 11.3: Details of the 15 most significant inclusive event classes from the scan of total event yield as shown in Fig. 11.10.

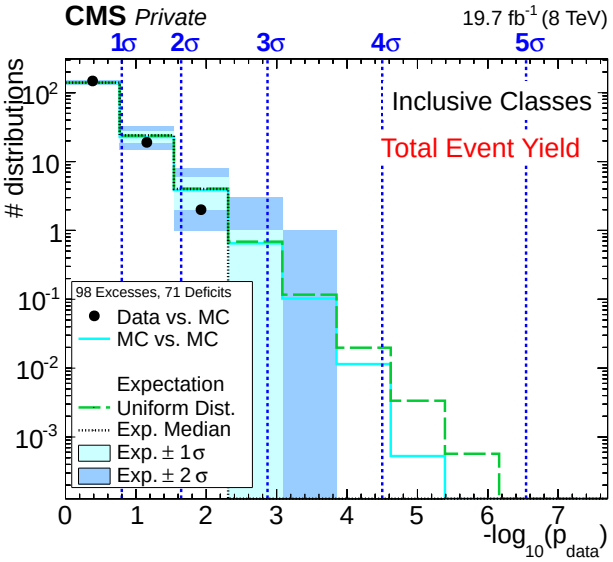


Figure 11.11: Distribution of p-values for inclusive event classes in the scan of total event yield. In the first bin 148 distributions are observed, with $140 \pm 5(1\,\sigma) \pm 10(2\,\sigma)$ expected from the SM.

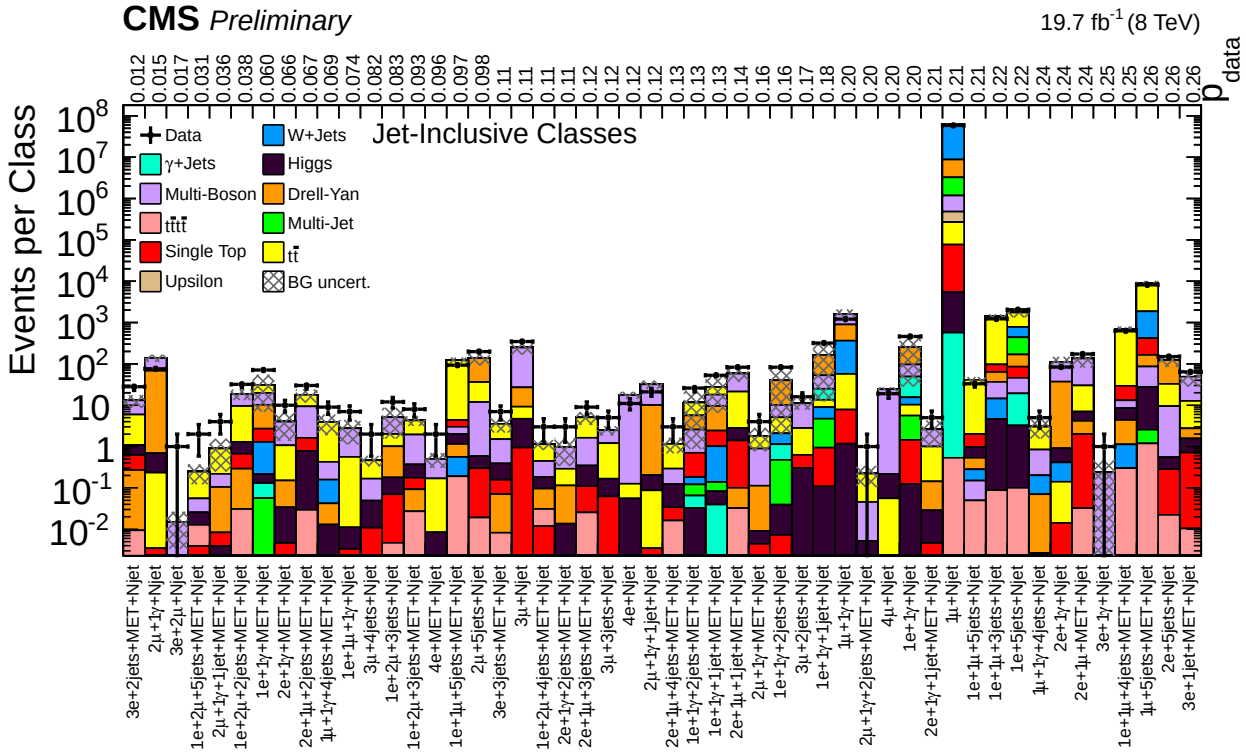


Figure 11.12: The 50 most significant jet-inclusive event classes for the scan of total event yield sorted by increasing p-value, as stated at top of diagram. Details of the 15 most significant ones are given in Tab. 11.4. An identical plot is published in [5].

Event Class	p	N_{data}	$N_{SM} \pm \sigma_{SM}$
3e+2jets+MET+Njet	0.012	28	13.1±4.4
2μ+1γ+Njet	0.015	76	138±27
3e+2μ+Njet	0.017	1	0.015±0.012
1e+2μ+5jets+MET+Njet	0.031	2	0.24±0.12
2μ+1γ+1jet+MET+Njet	0.036	4	0.89±0.64
1e+2μ+2jets+MET+Njet	0.038	32	18.6±5.3
1e+1γ+MET+Njet	0.060	71	31±24
2e+1γ+MET+Njet	0.066	10	4.0±2.6
2e+1μ+2jets+MET+Njet	0.067	30	17.8±6.2
1μ+1γ+4jets+MET+Njet	0.069	9	3.9±2.0
1e+1μ+1γ+Njet	0.074	7	2.8±1.6
3μ+4jets+Njet	0.082	2	0.46±0.13
1e+2μ+3jets+Njet	0.083	12	5.1±3.6
1e+2μ+3jets+MET+Njet	0.093	8	4.3±1.1
4e+MET+Njet	0.096	2	0.50±0.19

Table 11.4: Details of the 15 most significant jet-inclusive event classes from the scan of total event yield as shown in Fig. 11.12.

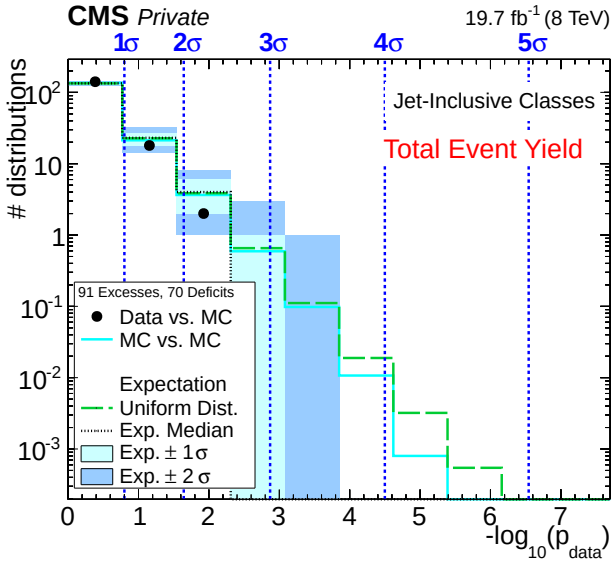


Figure 11.13: Distribution of p-values for jet-inclusive event classes in the scan of total event yield. In the first bin 141 distributions are observed, with $134^{+4}_{-5}(1\,\sigma)^{+\,9}_{-10}(2\,\sigma)$ expected from the SM.

11.2 Scan of Kinematic Distributions

The MUSiC analysis' sensitivity to small localized deviations or an overall difference in shape between the expectation and measurement is greatly increased by moving from a total event yield per event class to binned distributions in each case. Results obtained from the scan of electron and muon triggered events in kinematic distributions ($\sum |\vec{p}_{\mathrm{T}}|$, $M_{(\mathrm{T})}$, MET) of exclusive, inclusive and jet-inclusive event classes are discussed in the following sections.

11.2.1 Initial Validation

General methods of the MUSiC analysis are illustrated in the following, based on example results given in Fig. 11.14. Final states are chosen that allow an individual validation of dominantly contributing SM processes. Characteristic distributions for all event class types and kinematic variables are discussed. Throughout this analysis such plots of kinematic distributions for event classes contain the following information.

- The kinematic variable is plotted on the abscissa against the number of events per bin width (minimum of 10 GeV) on the ordinate.
- The event class name is shown at the top, left of the legend.
- Measured data is shown by black markers with bars representing the statistical uncertainty.
- The SM expectation given by MC simulation is depicted by stacked colored bars. For better visibility processes are combined into larger groups, as outlined in Tab. 8.1. In each distribution processes are given in ascending order of their integrated total number of events. In some cases the total number of events in each process is also stated in the legend. A gray hatched band gives the combined systematic uncertainty of all processes.
- The determined RoI is indicated by blue dashed lines, in some cases accentuated by a label.
- The p-value and $\tilde{p}$-value calculated in the RoI are written below the event class name.
- At the bottom of the plot, a ratio of the measurement and expectation is shown to highlight agreement or disagreement in individual regions of the distribution.

The $\sum |\vec{p}_{\mathrm{T}}|$ distribution of events containing precisely one electron and one muon passing the MUSiC selection, i.e., the exclusive 1e+1μ event class, is depicted in Fig. 11.14a. Due to thresholds set on each lepton's p_{T} during trigger and object selection the smallest possible $\sum |\vec{p}_{\mathrm{T}}|$ value is 55 GeV, representing events which pass the single-muon trigger requirements. In the region below 100 GeV DY is the dominant SM process. Towards higher $\sum |\vec{p}_{\mathrm{T}}|$ values contributions from multi-boson processes become more frequent. In this plot the right edge of the RoI extends beyond the depicted abscissa range to 1110 GeV. The thus included additional SM events make the observed deficit in data more significant. Nevertheless, the observed deviation of $N_{\mathrm{data}} = 15$ vs. $N_{\mathrm{SM}} = 30.3 \pm 9.9$ events with a $\tilde{p}$-value of 0.77 is not significant. Good agreement between measurement and SM expectation is found throughout the full kinematic range.

A final state dominated fully by DY is shown in Fig. 11.14b by the distribution of combined mass calculated between the two muons in the 2μ exclusive event class. The DY process is simulated well and with a sufficient number of generated events up to high combined mass values, as enabled by dedicated data sets binned in the di-muon mass produced with the POWHEG generator. Driven by the CMS muon resolution, this plot is a good example of a distribution whose bins become wider at higher values of the kinematic variable. The bin containing the data event with the highest combined mass value is 440 GeV wide. In $M_{(\mathrm{T})}$ the required minimum region width is relaxed from

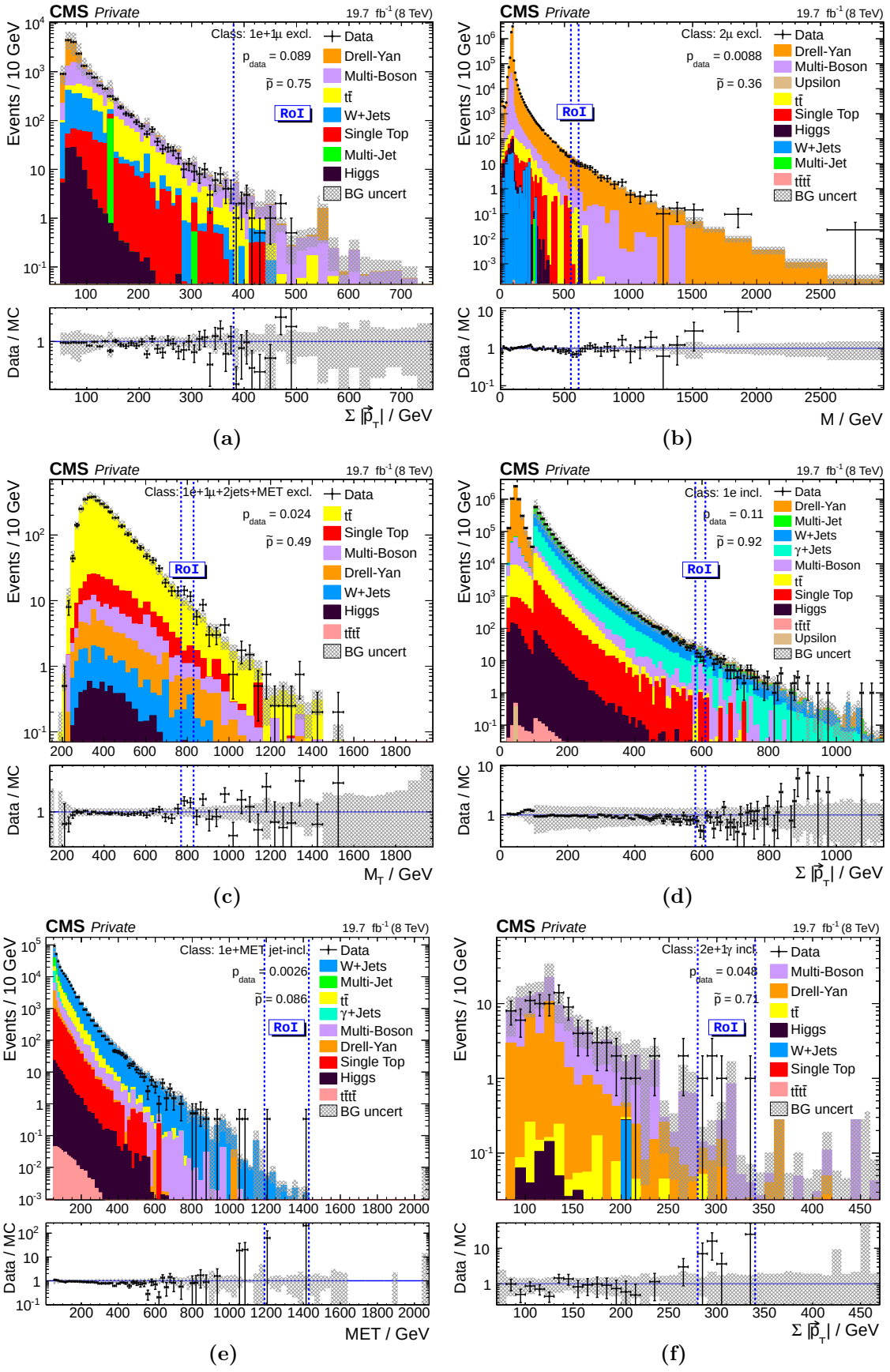


Figure 11.14: Examples of results in six different event class distributions. Measured data is given by black markers and contributions from SM processes are represented by colored bars. The event class name is shown left of the legend along with the p-value and $\tilde{p}$-value. In each case the RoI is indicated by blue dashed lines and a label. Figures (b), (c), (d) and (e) are published in similar form in [5].

three to one bin. The RoI in this case is only two bins wide, exhibiting an insignificant deficit of 63 data events against 98.3 ± 9.0 expected SM events. Due to the large number of bins in this distribution the LEE reduces the p-value by around two orders of magnitude to $\tilde{p} = 0.36$. This final state is under close investigation by a dedicated analysis searching for resonances of heavy new particles in the invariant mass distribution of opposite-sign lepton pairs [287]. The analysis relies on the single-muon data stream and uses a muon trigger without isolation criteria and consequently require both selected muons to have $p_{\mathrm{T}} > 45$ GeV, along with other reconstruction quality criteria following the High-p_{T} muon selection also used by MUSiC (see Sec. 9.2.2). Moreover, only opposite-sign muons pairs are chosen to contribute to the final di-muon spectrum, and their tracks are re-fit to incorporate a common vertex. An upper threshold is applied on the three-dimensional angle between both muons to reduce signatures stemming from cosmic-ray muons. Good agreement throughout the entire kinematic range is reported and exclusion limits are set on a series of BSM theories. The event with the highest di-muon mass is observed at 1.8 TeV and four events exceed 1.5 TeV. Following the MUSiC strategy, the event with the highest di-muon mass is found at around 2.6 TeV, and five events are given over 1.5 TeV, all of which contain an opposite-sign muon pair. The two highest events uncovered by [287] are also found by MUSiC at slightly altered invariant mass values, given by the common vertex re-fit of the dedicated analysis. The most energetic di-muon event observed by MUSiC is rejected in the dedicated analysis through the requirement on the muon pair three-dimensional angle. In this event both muons are very energetic and nearly back-to-back. In the electron channel three data events are observed above 1500 GeV [207], also found in the MUSiC 2e event class at corresponding di-electron invariant mass values.

An event class well suited to isolate $\mathrm{t\bar{t}}$ production from other SM processes is 1e+1μ+2jets+MET, whose M_T distribution is depicted in Fig. 11.14c. The two bottom quarks produced in the top quark decay are identified as jets. The associated neutrinos are reconstructed as MET, thus the transverse mass is calculated in these events. Contributions from DY are small given by the fully leptonic decay into two different flavor charged leptons. In this event class distribution the RoI finds an insignificant excess, where overall the SM simulation provides an excellent description of the data.

A feature of the applied trigger strategy becomes visible in the $\sum |\vec{p}_{\mathrm{T}}|$ distribution of the inclusive event class containing at least one electron 1e+X, shown in Fig. 11.14d. Events can contribute to this event class in different ways. Low values of $\sum |\vec{p}_{\mathrm{T}}|$ are populated by events recovered by the double-electron trigger (thresholds at 8 TeV and 17 TeV). The DY process is dominant in this regime, as all events must contain at least two electrons. Alternatively, events can pass the single-electron trigger if they are more energetic, resulting in a distinct edge at 100 GeV. In this part of the distribution three processes are dominant, being QCD multi-jet events containing an electron from a bottom quark decay or jets misidentified as electrons, events from the leptonic W boson decay, as well as γ+jets events in which a jet or photon is mistaken as an electron.

The W+jets process is clearly the dominating contribution in the MET distribution of the 1e+MET+Njet event class, which summarizes all events with precisely one electron, a substantial amount of MET and any number of jets (also zero). The corresponding plot is shown in Fig. 11.14e. Resulting from the quality selection applied on MET the distribution begins at 50 GeV. Good agreement is found throughout the distribution. Yet, a slight trend towards an excess of SM prediction is visible at MET over 200 GeV until around 600 GeV, within the steeply falling flank. More details of this observation will be given later on in Sec. 11.2.2. The strongest deviation is found as an excess in the tail of the distribution with two events observed in data against 0.071 ± 0.025 expected from the SM in a 200 GeV wide region. Though the calculated significance of $\tilde{p} = 0.087$ ($\sim 1.4\,\sigma$) is moderate and not statistically significant, this distribution is a good example for the MUSiC algorithm's sensitivity to signatures in the high MET regime, originating, e.g., from BSM processes containing new particles escaping detection or introducing

heavy W boson-like particles. The M_T distribution of this final state is scrutinized by a dedicated analysis searching for signatures in the high energy tails of events containing a single electron or muon and MET [228]. To reduce contributions other than the W+jets process, requirements on the ratio of lepton p_{T} and MET are set, as well as a lower threshold on the ϕ angle between lepton and MET of $\Delta\phi(l, \mathrm{MET}) > 2.5$. In the transverse mass of lepton and MET, the highest event in the muon channel is observed at 2.1 TeV and at 2.3 TeV in the electron channel. A single event can be found in the corresponding M_T distributions of the 1μ+MET+Njet and 1e+MET+Njet event classes[3] at the same values. Comparisons of the MUSiC results with event displays given in [259], reveals that these are precisely the events observed by the dedicated analysis. Furthermore, the M_T spectra observed by MUSiC are very similar to the ones observed by the dedicated search. Over 1.5 TeV both approaches observe three events in the muon and one in the electron case.

The final plot in Fig. 11.14f shows the $\sum |\vec{p}_{\mathrm{T}}|$ distribution of the 2e+1γ+X inclusive event class. Multi-boson processes show a large expected number of events in this final state, where $Z\gamma$ with its prompt photon, is by far the main contribution. A further prominent contribution is DY, where the photon can originate from a misidentified jet or from ISR and FSR. Accordingly, a large part of the total systematic uncertainty on the SM expectation is given by the uncertainty originating from misidentification. Overall, the data agree well with the SM simulation. A slight, insignificant excess is found around 300 GeV. As shown in this examples, the MUSiC analysis is well suited to investigate final states containing photons in addition to the triggering leptons.

11.2.2 Analysis Improvements

Three distinct features of the MUSiC analysis with 8 TeV data are discussed in this section, as they represent general aspects that were especially addressed and improved during the work leading up to this thesis. In all cases the highlighted event classes show only moderate $\tilde{p}$-values and are not among the most significant deviations.

Contributions from Higgs Boson Production

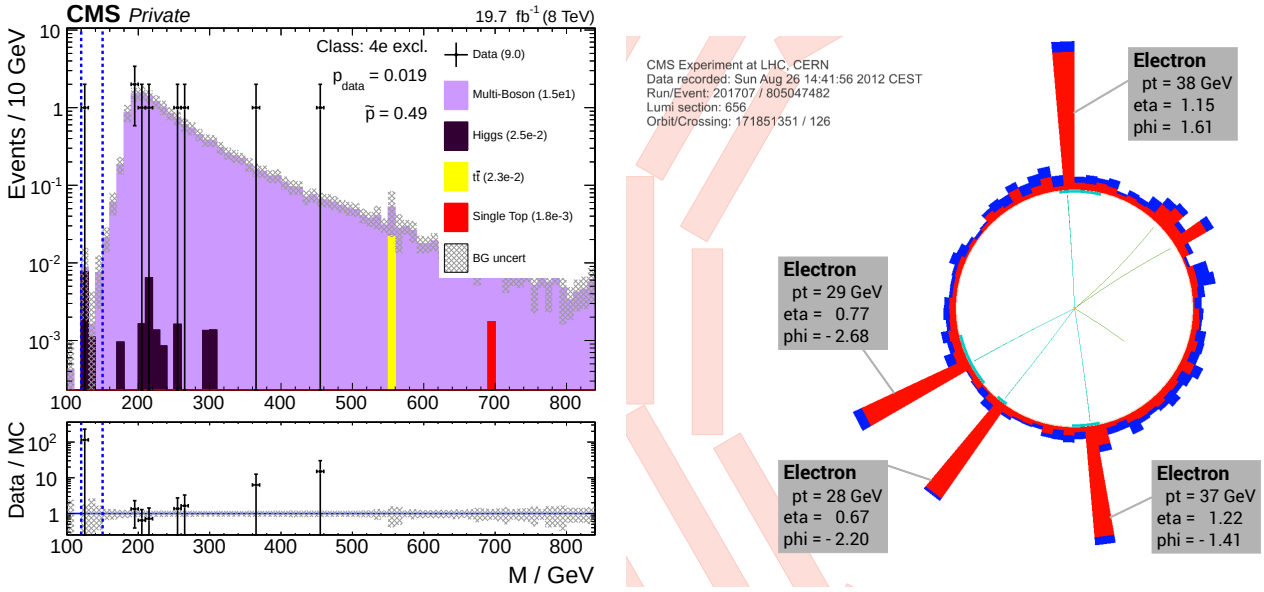
As mentioned in Sec. 2.3.2, a resonance was observed at a mass of around 125 GeV in data taken at $\sqrt{s} = 7$ TeV and 8 TeV, corresponding well with the SM Higgs boson. Consequently, to achieve a simulation of the SM that is as inclusive as possible, the MUSiC analysis also includes processes linked to the production and decay of the Higgs boson, as discussed in Sec. 8.2.1.

When including Higgs boson production in the SM simulation, the largest change in p-value[4], for exclusive event classes is observed in the mass distribution of 4e, as shown in Fig. 11.15a. It is overwhelmingly dominated by multi-boson production, most of which is given by Z boson pair production. Although the total number of events stemming from Higgs boson processes is three orders of magnitude smaller than the multi-boson event count (total number of events given in the legend), the contribution in the RoI itself is rather large. Neglecting the Higgs boson production leads to an increased significance of $p_{\mathrm{data}} = 0.0027$ in a slightly shifted RoI from 100 GeV to 140 GeV, yet still an overall insignificant deviation with $\tilde{p} = 0.12$. An event display of the Higgs boson candidate data event located in the RoI is shown in Fig. 11.15b, illustrating the clean nature of events given in the 4e event class.

For many collider experiments, the four lepton final state was investigated to uncover the existence of the Higgs boson and subsequently, at the LHC used to determine its properties. Naturally, the MUSiC analysis is far less sensitive to the observation of events containing Higgs boson signatures than the corresponding dedicated searches. In the combined analysis of decay channels

[3] The dedicated analysis [228] is performed without an explicit jet-veto

[4] By comparing p-values in the RoI, and not $\tilde{p}$-values, statistical effects of the LEE estimation are factored out.



(a) (b)

Figure 11.15: The left plot shows the mass distribution for the 4e exclusive event class, whose p-value is influenced most when including the Higgs boson process. The RoI is given from 120 GeV to 150 GeV, with a very moderate overall significance. An event display of the data event (run:lumi section:event - 201707:656:805047482) found in the RoI is shown on the right side of this figure. Four clearly separated electrons are identified in the barrel region of the detector, all of comparable momenta. Apart from these leptons, no large activity is seen in the event. MUSiC kinematic variables of this event are: $\sum |\vec{p}_{\mathrm{T}}| = (131.7 \pm 0.4)$ GeV; $M_{(\mathrm{T})} = (121.8 \pm 0.4)$ GeV; MET $= (12 \pm 9)$ GeV (disregard in analysis);

$H \to ZZ \to 4l/2l2\tau$ ($l = \mathrm{e}, \mu$) of CMS data taken at 7 and 8 TeV [301] a local significance of $6.7\,\sigma$ is reported. Of the 41 data events observed in electron and muon final states at 8 TeV in the range of m_{4l} =106 GeV to 141 GeV [302], 28 events pass the MUSiC analysis selection and appear in the resulting event classes. Details of events observed by both the dedicated and the MUSiC analysis are given in Tab. C.1, along with the corresponding final state identified by [302] and the MUSiC event class. Mismatches in the reconstructed event composition occur mainly because of unequal lepton p_{T} thresholds and isolation criteria. This leads to differences in misidentification rates and selection efficiency, causing individual electrons and muons to be identified as other objects or completely rejected from the event. The main advantage of the dedicated search is a more refined trigger and object selection. In addition to the double-lepton triggers also used by the MUSiC analysis, a triple electron trigger and cross-triggers selecting events containing an electron/muon with $p_{\mathrm{T}} > 8$ GeV and a muon/electron with $p_{\mathrm{T}} > 17$ GeV are considered. Minimum thresholds on electron and muon p_{T} are lower than those considered by MUSiC, since more complex selection criteria dedicated to the reconstruction of lower momentum leptons are used. Furthermore, a multistage scheme based on the invariant mass of opposite sign same flavor lepton is invoked to select events containing Z boson candidates.

Overall, processes from Higgs boson production appear in many MUSiC event classes. Apart from cases where far less than one event ($< \mathcal{O}(10^{-3})$ events) is expected from the SM simulation, contributions involving the Higgs boson are never dominant throughout the distribution and have only a small influence on the final significance of the event class.

Flat k-factor for W+jets Production

The W+jets process is one of the main SM components to many of the investigated final states. It represents the primary SM contribution in event classes with one charged lepton, large MET and few jets, and also appears along side similarly frequent processes in a variety of event classes with moderate MET, multiple leptons or a large number of jets. Therefore, it is important to provide a universally good modeling of this process. A sufficient number of generated events must be given, but also a reliable description of the shape in kinematic distributions. To achieve such a result, this analysis relies on a composition of three conceptually different MC simulation data sets, which each focus on describing the process in a specific part of the phase-space, as explained in Sec. 8.2.1. The validity of applying a flat k-factor to all of these cases must be scrutinized. All used data sets rely on W+jets processes generated at LO with MADGRAPH and are then scaled to an NNLO cross section calculation using the k-factor. As previously demonstrated in Fig. 11.14e, this procedure performs very well in single-lepton event classes with an inclusive number of jets. The main W+jets contribution in these final states is given by events with zero jets. The 1μ+MET event class, with its implicit jet veto, contains a total of $1.4{\cdot}10^7$ data events, whereas 1μ+MET+Njet, with any number of jets, contains only around 20% more events (total of $1.7{\cdot}10^7$ events).

As can be seen in Fig. 11.16a, exclusive event classes with one lepton and MET are also modeled appropriately by the chosen W+jets simulations. The data are described well over a large range in transverse mass, with event numbers dropping from $\sim 6{\cdot}10^6$ at the W boson peak to single events at over 2 TeV. From 500 GeV onwards the MC simulation systematic uncertainty is driven mainly by estimates of the proton PDFs (see individual uncertainty contributions in Fig. 11.16b). Unfortunately, exclusive W+jets dominated event classes with one or two identified jets are not represented well by the simulation, as shown by an example in Fig. 11.16c with 1μ+2jets+MET. The same behavior is observed for electrons. A clear trend towards an overshoot of MC simulation events is seen from 300 GeV onwards. The RoI, though not significant, is also located in this region of poor agreement. As illustrated by plotting the H_T binned data sets of W+jets process separately in Fig. 11.16c, the tail of the distribution is fully dominated by kinematics of W boson events with an H_T over of 400 GeV. In the transverse mass distribution of 1μ+MET, contributions from such events are located in the region below the W boson peak and play only a minor role. Overall, such an overshoot of W+jets MC events is only visible in a handful of distributions per event class type, i.e., ones containing a single lepton, substantial MET with few identified jets. For similar final states with three jets and more, contributions from $\mathrm{t\bar{t}}$, generated at NLO, become more dominant. Thus, for the overall analysis results, the impact of the W+jets k-factor is only marginal.

The flat NNLO k-factor, used to scale all W+jets contributions, is dominated by events at the W boson peak, those in the high H_T regime are not accounted for appropriately by this value. This behavior is also observed by a dedicated search based on CMS data-taking during 2015 at $\sqrt{s} = 13$ TeV [303] and is reproduced similarly in calculations of the W+1jet differential cross section at NNLO for 8 and 13 TeV [304] and W+0,1,2jets at 13 TeV [305]. All three analyses report a strong H_T dependence over the full kinematic range, e.g., ranging from $k = 1.0$ to 1.5 in [303]. Yet, because of differences in event composition, jet definition[5] and H_T bins, neither of these k-factor estimates can be transferred to the data sets used at $\sqrt{s} = 8$ TeV. A viable solution for the MUSiC analysis is to apply an increased systematic uncertainty on the H_T binned W+jets data sets. A conservative estimate of 30% is chosen, based on the difference between the total W+jets k-factor at 8 TeV and values given by the dedicated analysis [303]. The overall effect on the uncertainty can be seen in Fig. 11.16d. Contributions from cross section estimates are by far

[5] MC simulation data sets produced for CMS data-taking in 2015 use a newer software version PYTHIA8 for showering jets, including revised tunes. Also, the jet radius parameter is reduced to $R = 0.4$.

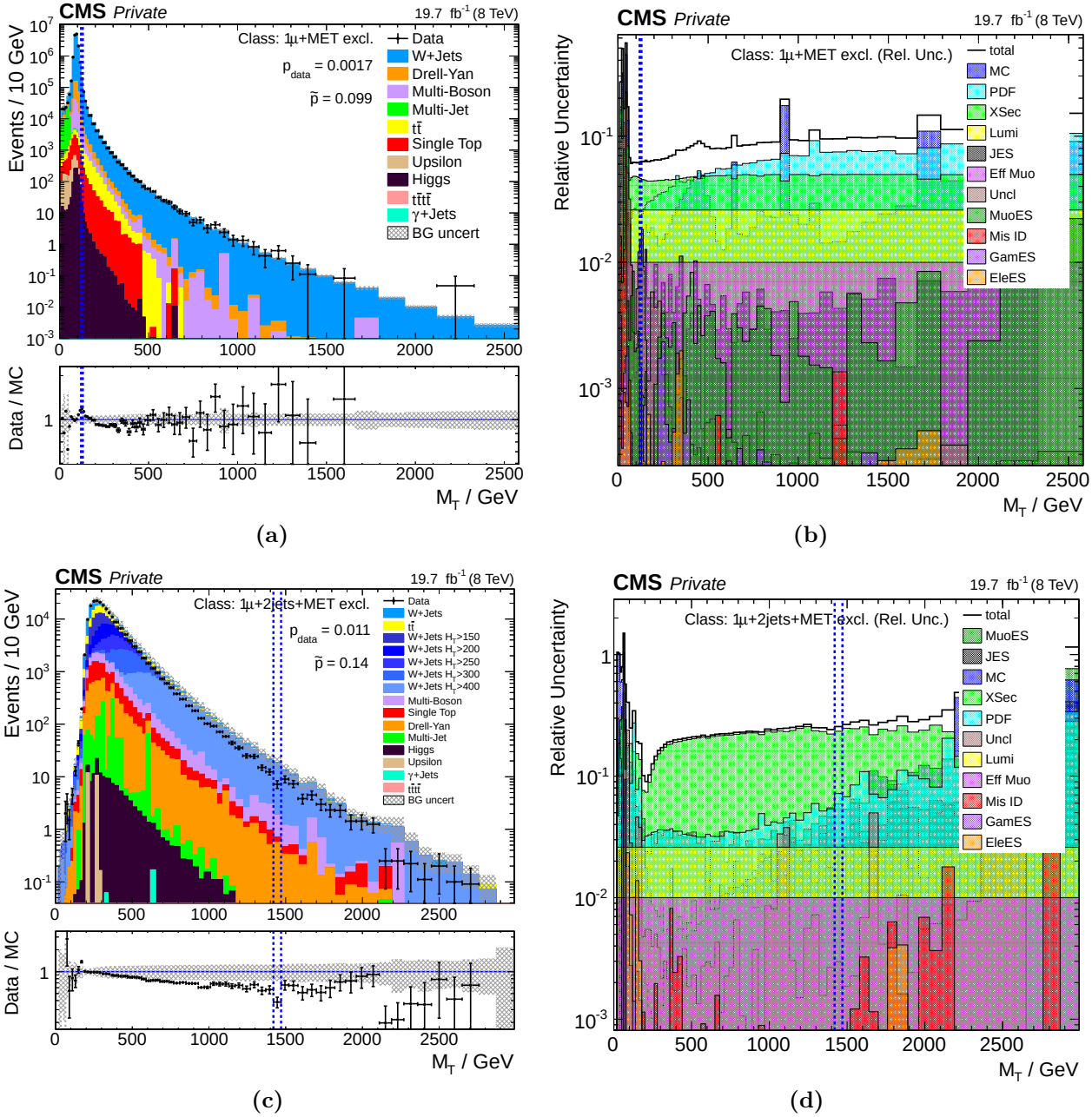


Figure 11.16: Distributions of the transverse mass for events in the 1μ+MET (top) and 1μ+2jets+MET event class (bottom). An insignificant deviation is found in both final states, which are clearly dominated by the W+jets process. The right side of this figure gives an overview of the respective individual systematic uncertainty contributions. Values are given relative to the total number of weighted events from the simulation. A black line gives the total relative systematic uncertainty. In Fig. 11.16a all W+jets data sets are combined and shown in blue. Below, in Fig. 11.16c, W+jets data sets binned in H_{T} on generator level are emphasized in different shades of blue. The remaining contribution labeled "W+jets" includes the main data sets as well as ones binned in the W boson mass and ones including dedicated bottom quark contributions.

the dominant systematic over an extended range of the distribution and much more pronounced than in the example of 1μ+MET. As always for a general search, a compromise must be found between underestimating uncertainties in specific final states and overestimating them in other channels, thus unnecessarily blinding the analysis.

A procedure that would go beyond simply mitigating the effect of a flat k-factor and would a priori lead to a better description of the W+jets process, is to produce data sets using higher order generators. Indeed, such approaches are pursued for CMS data-taking in Run-II. POWHEG and MADGRAPH aMC@NLO [306] are invoked to generate inclusive W+jets events at NLO. Regrettably, in both cases the generated number of events is too low to ensure a good description of tail regions and data sets binned in H_{T} or similar variables are not available at higher order.

Unclustered Energy Contributions

As mentioned previously for the scan of total event yield in Sec. 11.1, an excess of data is observed in event classes with two same flavor leptons and substantial MET. Dedicated studies performed for this thesis show that the deviation does not depend strongly on the amount of pileup in the event or on the charge of the two isolated leptons (same-sign or opposite-sign). The origin of the excess can be understood by considering the MET distribution of the corresponding exclusive event classes. These are shown in Fig. 11.17 alongside the distribution of their relative systematic uncertainties. For events below 100 GeV in MET, a fairly large contribution is given from the systematic uncertainty stemming from an estimate of the difference in unclustered energy between data and simulation. In respective $\sum |\vec{p}_{\mathrm{T}}|$ and $M_{(\mathrm{T})}$ distributions the excess events are spread out over the Z boson mass peak, also fully within the extent of the unclustered energy uncertainty. In both cases the RoI is located in the distribution's tail and exhibits an insignificant excess.

Especially at low MET the distribution of 2l+MET events shows a very steep slope. A slight systematic shift towards higher or lower MET values in each reconstructed event has a strong impact on the observed event rate. This region of the distribution is dominated by events from DY production, in which no produced particle is expected to remain undetected, yet MET may emerge because of imperfect measurements. In the tail of the distribution, dominated by events containing genuine MET from $\mathrm{t\bar{t}}$ production, the slope is considerably shallower and a shift along the x-axis has a negligible effect. The 2l+MET event classes include an implicit jet veto for jets with $p_{\mathrm{T}} > 50$ GeV. Thus, the main influence on misreconstructed MET in the corresponding DY events originates from imperfectly determined unclustered energy and not directly from mismeasured high energy jets. In event classes with additional jets, the effect is less pronounced. $\mathrm{t\bar{t}}$ is the overall dominant SM contribution for $N_{\mathrm{jet}} \geq 2$, delivering more genuine MET to the final state even at low MET values.

A study of the MET performance at 8 TeV [307] shows a similar impact of mismeasured MET on systematic uncertainties in the low MET regime of final states which are dominated by events lacking genuine MET. An increased systematic uncertainty is given in the region around 40 GeV to 100 GeV in MET, where the data show an upwards trend over the DY MC estimate. Contributions from unclustered energy are known to be poorly described, which is due to the lack of a detailed understanding of the underlying causes, e.g., detector noise. Improvements to the understanding are expected to directly translate into a correction of the MET itself.

For future instances of the MUSiC analysis, the unclustered energy uncertainty should be followed up upon. Moreover, it might be feasible to raise the minimum requirement on the event's MET to 100 GeV, since this regime is prone to a substantial amount of imperfectly measured MET and not expected to be especially sensitive to new BSM effects.

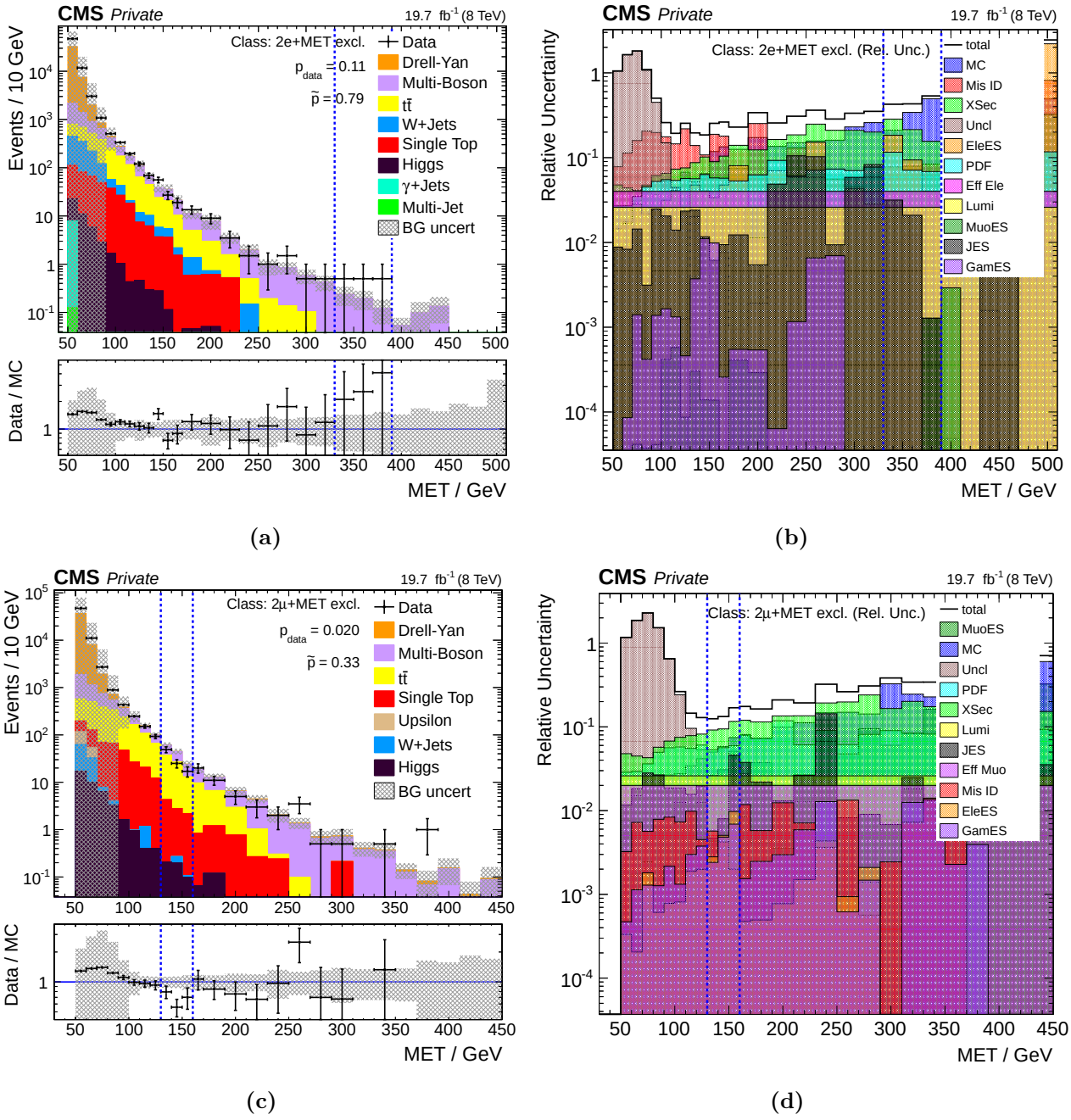


Figure 11.17: MET distribution of the 2e+MET (top) and 2μ+MET (bottom) event classes. Corresponding distributions of relative systematic uncertainties are shown in the right column. Large contributions to the regime below 100 GeV are given by the unclustered energy uncertainty.

11.2.3 Overview of the Statistical Analysis

An overall impression and statistical evaluation of the analyzed events is achieved by comparing $\tilde{p}$-values, obtained from the comparison of simulated SM expectation and measured data, with those expected from the SM itself (see Sec. 7.3.7 for details). Event classes failing criteria of the low generated event count treatment are not included in the distribution of $\tilde{p}$-values and handled separately in Sec. 11.2.4. Furthermore, to reduce runtime a $\tilde{p}$-value is not calculated for insignificant

distributions without measured data, as explained in Sec. 7.3.6. In total a $\tilde{p}$-value is estimated for the $\sum |\vec{p}_{\mathrm{T}}|$ distribution of 152 (out of 337) exclusive, 150 (out of 341) inclusive and 146 (out of 321) jet-inclusive event classes and accounted for in the global statistical overview. Fewer event classes are considered for the other two kinematic variables, as they cannot be calculated for every event class ($M_{(\mathrm{T})}$: 150 (out of 335) excl., 150 (out of 339) incl., 145 (out of 319) jet-incl.; MET: 69 (out of 161) excl., 68 (out of 150) incl., 68 (out of 150) jet-incl.). Up to $2{\cdot}10^6$ SM-only pseudo-experiments are used to estimate the LEE correction, allowing $\tilde{p}$-values of up to $5\,\sigma$ to be assessed.

All distributions of $\tilde{p}$-values for exclusive event classes are shown in Fig. 11.18, those from inclusive and jet-inclusive event classes summarized in Fig. 11.19. Compared to results from the scan of total event yield, a somewhat worse agreement is found. Thus, the observed effects are enhanced or appear because of a difference in shape and not solely from the total number of events in each kinematic distribution.

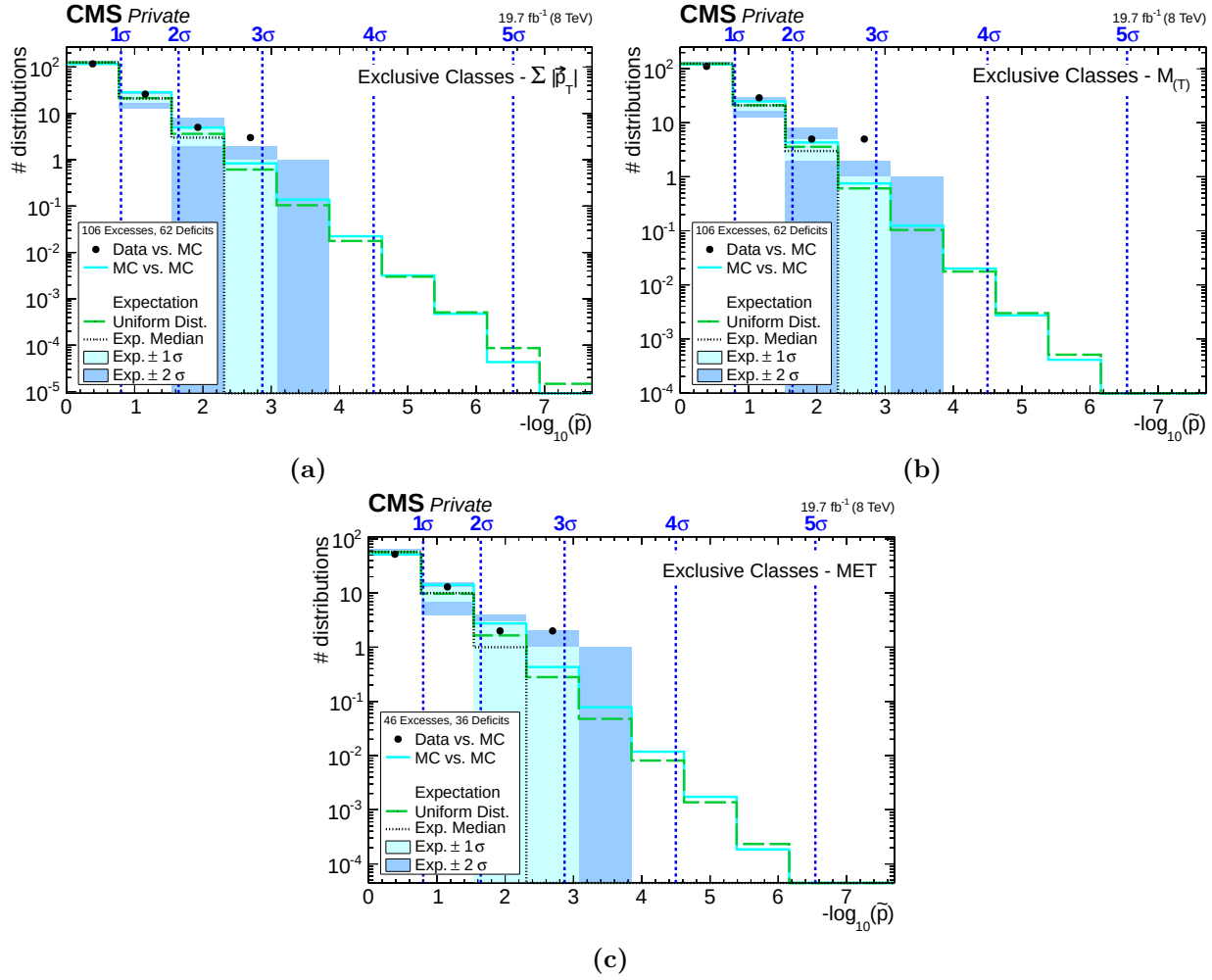


Figure 11.18: Distribution of $\tilde{p}$-values for $\sum |\vec{p}_{\mathrm{T}}|$, $M_{(\mathrm{T})}$ and MET distributions of exclusive event classes. Black markers represent the measured data compared to the SM MC expectation. In the first bin a: 117 [b: 111, c: 52] distributions are observed in data, with a: $125^{+5}_{-4}(1\,\sigma) \pm 9(2\,\sigma)$ [b: $125^{+4}_{-5}(1\,\sigma)^{+8}_{-10}(2\,\sigma)$, c: $57 \pm 3(1\,\sigma) \pm 6(2\,\sigma)$] expected from the SM. All figures are published in similar form in [5].

The total number of observed excesses and deficits is stated in the legend of the distribution of $\tilde{p}$-values for each event class type and kinematic variable. Note that the sum of both values is larger than the number of event class distributions a $\tilde{p}$-value was calculated for, as it also contains cases

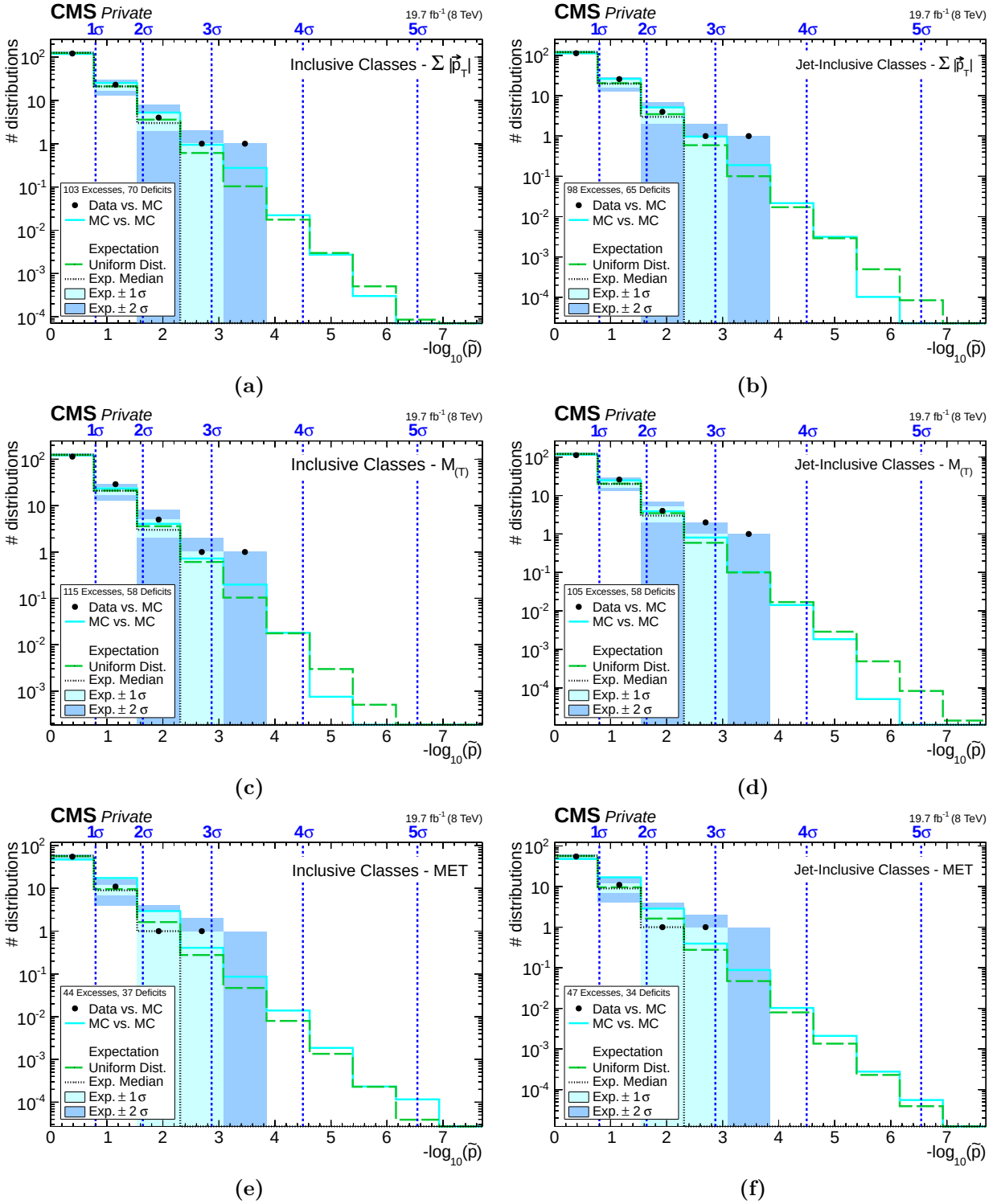


Figure 11.19: Distributions of $\tilde{p}$-values for kinematic distributions of inclusive (left side: a: $\sum |\vec{p}_{\mathrm{T}}|$, c: $M_{(\mathrm{T})}$, e: MET) and jet-inclusive (right side: b: $\sum |\vec{p}_{\mathrm{T}}|$, d: $M_{(\mathrm{T})}$, f: MET) event classes. Black markers represent the measured data compared to the SM MC expectation. In the first bin a: 121 [b: 114, c: 114, d: 112, e: 55, f: 55] distributions are observed in data, with a: $125^{+4}_{-5}(1\,\sigma)^{+8}_{-10}(2\,\sigma)$ [b: $121^{+5}_{-4}(1\,\sigma) \pm 9(2\,\sigma)$, c: $125^{+4}_{-5}(1\,\sigma)^{+8}_{-10}(2\,\sigma)$, d: $120^{+5}_{-4}(1\,\sigma) \pm 9(2\,\sigma)$, e: $57^{+3}_{-4}(1\,\sigma)^{+5}_{-7}(2\,\sigma)$, f: $57^{+3}_{-4}(1\,\sigma)^{+5}_{-7}(2\,\sigma)$] expected from the SM. All figures are published in similar form in [5].

for which only a p-value but no LEE correction was evaluated. Overall more excesses are found than deficits, a trend already observed in the results of the scan of total event yield in Sec. 11.1. Many cases with zero observed data events and only a small SM MC yield are removed from the p-value estimation by the low generated event count treatment.

In exclusive classes, a slight trend towards strongly deviating regions is found. Though the first three bins of the distribution are still within the $2\,\sigma$ band of the uniform distribution, the fourth bin shows a change in slope, departing from the uniform expectation. In all three variables, the median expectation (black dashed line) is zero from the fourth bin onwards. For $\sum |\vec{p}_{\mathrm{T}}|$ and MET the observed number of deviations is within the $2\,\sigma$ expectation of the uniform distribution, if the fifth is also taken into account. This is not the case for $M_{(\mathrm{T})}$ distributions. Only a maximum of three event classes are expected in the sum of bins four and five, with five event classes observed in data. Better agreement is found in inclusive and jet-inclusive event classes. Over all event class types, distributions of MET accord best with the SM prediction. All bins in the distribution of $\tilde{p}$-values show an agreement within $2\,\sigma$ of the uniform expectation. The largest deviation in MET is found at $\tilde{p} = 0.0013 \approx 3.0\,\sigma$ in a (jet-)inclusive event class.

The ten most significant deviations for each event class type and kinematic variable are summarized in App. C.4. Of these, the two most significant are discussed in Sec. 11.2.5. Often the same deviation appears in related event classes and throughout all kinematic variables. Many of the investigated effects can be explained through shortcomings of the SM simulation and help pin-point areas worthwhile improving in future CMS analysis. No obvious indications of new BSM effects are seen. The overall agreement with the SM prediction is excellent in many final states. Throughout all scan types satisfactory agreement is seen between the numerically estimated $\tilde{p}$ distribution, shown as a light blue line, and the analytic uniform distribution, shown by a green dashed line.

11.2.4 Evaluation of the Low Generated Event Count Treatment

Over the last years, two fully automatable algorithms were developed to address the issue of an insufficient number of simulated SM MC events through the event class distributions of the MUSiC analysis. Both are introduced in Sec. 7.3.3 and their results are compared on the basis of the 8 TeV analysis in the following section.

Discussion of Results from the Region Veto Method

Not all regions are suitable for calculating a p-value and thus disregarded in the RoI scan based on the number of generated SM MC events. In the most extreme cases entire distributions with a low generated MC event count are vetoed. Overall, distributions of five event classes containing at most three data events were completely disregarded and are given in Tab. 11.5. All of these event classes show a similar combination of objects, two electrons, one photon and two or more jets. In event classes with more than two jets no substantial amount of MET is present. Several collision events contribute to more than one of the listed event classes, in total five different data events are distributed over these classes. The $\sum |\vec{p}_{\mathrm{T}}|$ distribution and mapping of regions to veto criteria of the low generated event count treatment is shown for each event class in App. C.3. As given previously in Tab. 11.1, all of these event classes are also disregarded in the scan of total event yield, since only an insufficient SM MC prediction is provided considering the largest region, i.e., all bins of the distribution. In the scan of total event yield two additional event classes are disregarded, 2e+1γ+4jets+X and 2e+1γ+4jets+Njet. The mass distributions of 2e+1γ+3jets(+X/+Njet) are not disregarded. Depending on the precise position of the individual SM MC contributions in the distribution, valid regions can emerge. The NH criterion may be fulfilled if all contributions of one SM process are moved completely out of the NH. As given in Sec. 10.8, the relative statistical

uncertainty $\sigma_{\mathrm{SM}}^{\mathrm{stat}}/N_{\mathrm{SM}}$ is taken to be fully correlated over all weighted events of one process type, but fully uncorrelated over different simulated processes in the investigated region. Thus, the total statistical uncertainty value also depends on how event contributions to the same process are distributed through the kinematic range. The resulting six kinematic distributions, whose integral does not pass the low generated event count treatment, contain deviations that are among the most significant $\tilde{p}$-values found in the statistical analysis of event class distributions and are discussed in Sec. 11.2.5. In many valid regions the relative statistical uncertainty is close to the assumed threshold of 60%. As established in the further discussion of these deviations, they most likely do not originate from an unknown effect but can be categorized as a lack of precision in the SM MC simulation data sets. Consequently, future instances of the MUSiC analysis might consider tightening the requirements of the low generated event count treatment to rule out such distributions from the start. Nevertheless, a careful measure must be found to insure interesting deviations are not overlooked.

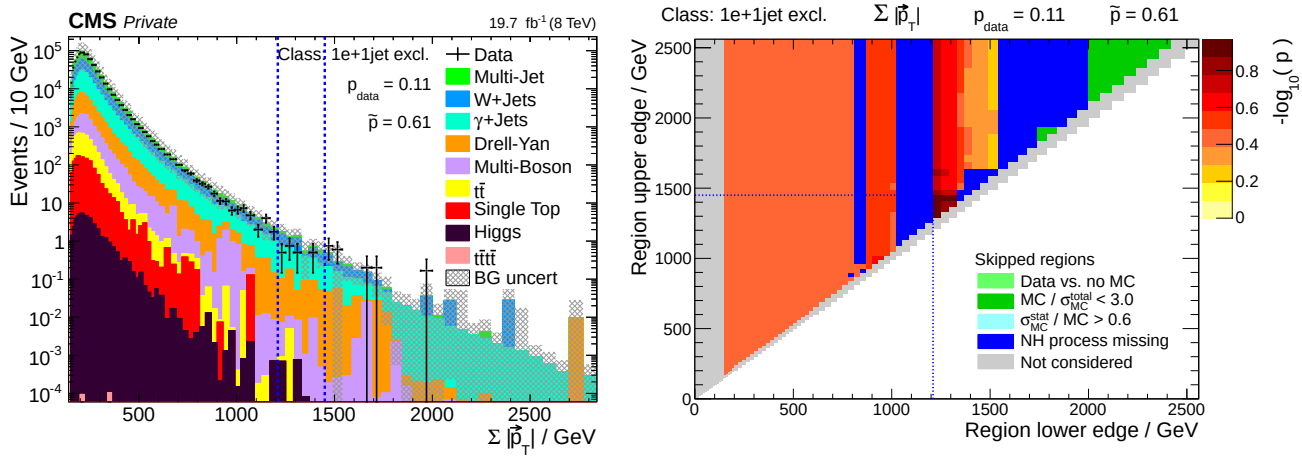
Event Class	N_{data}	$N_{\mathrm{SM}}\pm\sigma_{\mathrm{SM}}$	Distribution	Event Details
2e+1γ+2jets+MET	3	0.89±0.61	$\sum \lvert\vec{p}_{\mathrm{T}}\rvert$, $M_{(\mathrm{T})}$, MET	Fig. 11.2, Fig. 11.3, Fig. 11.4
2e+1γ+3jets	1	0.74±0.57	$\sum \lvert\vec{p}_{\mathrm{T}}\rvert$	Fig. 11.5
2e+1γ+3jets+X	2	0.84±0.59	$\sum \lvert\vec{p}_{\mathrm{T}}\rvert$	Fig. 11.5, Fig. 11.6
2e+1γ+3jets+Njet	2	0.75±0.57	$\sum \lvert\vec{p}_{\mathrm{T}}\rvert$	Fig. 11.5, Fig. 11.6
2e+1γ+4jets	1	0.015±0.019	$\sum \lvert\vec{p}_{\mathrm{T}}\rvert$, $M_{(\mathrm{T})}$	Fig. 11.6

Table 11.5: Details of all exclusive, inclusive and jet-inclusive event classes containing data but only insufficient SM MC prediction in all regions of the corresponding distribution. Total data and SM event yields are given. Respective data events are discussed in the figures referred to in the last column.

An overview of regions containing data but no SM MC prediction is given in App. C.2. A p-value cannot be calculated there, yet all of these data events are enclosed in larger regions with a sufficient expectation. The majority of these regions are located in tails of the distributions and contain only single data events. The 2μ+1jet event class presents a different case and is discussed in detail in the following section, as it contains one of the overall most significant deviations.

To retrace which regions of a distribution lead to a valid p-value and which are disregarded, a mapping of regions to requirements of the analysis or p-values is created for each event class distribution. An example is shown in Fig. 11.20, where the RoI is indicated by blue dashed lines. The lower and upper edge of the investigated regions are given on the axes of Fig. 11.20b. Regions in which a p-value was calculated are indicated on a yellow to red scale according to the negative decadic logarithm of the p-value. Light green regions contain measured data events but no SM prediction. No data is present in dark green regions which are disregarded in terms of runtime optimization, as explained in Sec. 7.3.6, since the SM MC prediction is compatible with zero.

Furthermore, regions constructed by adding a completely empty bin to an already investigated region are disregarded automatically by the algorithm to reduce computing time, such regions are given in gray. For $\sum \lvert\vec{p}_{\mathrm{T}}\rvert$ and MET distributions all regions containing less than three bins are not considered (for motivation see Sec. 7.3.1) and are also given in gray, as is the case for the two bins left of the diagonal in Fig. 11.20b. Regions marked in light blue fail the relative statistical uncertainty criterion of Eq. (7.16) and those in dark blue are disregarded under the NH region criterion. Moreover, Fig. 11.20a represents a good example for illustrating the necessity of invoking the NH criterion on only 95% and not on 100% of the most frequent SM processes. Contributions from γ+jets or $\mathrm{t\bar{t}t\bar{t}}$ processes in individual bins of the $\sum \lvert\vec{p}_{\mathrm{T}}\rvert$ distribution of 1μ+MET would



(a) (b)

Figure 11.20: The distribution of $\sum |\vec{p}_{\mathrm{T}}|$ in 1e+1jet is given on the left, showing good agreement between measured data and the SM prediction. On the right, a mapping of regions to p-values or failed analysis criteria is given for this distribution. Two bins containing very small contributions from $\mathrm{t\bar{t}t\bar{t}}$ are present below 400 GeV, which do not lead to a veto of the region since the NH criterion is applied on only the 95% most frequent processes. Not all forms of disregarded regions are present in this distribution (light green and light blue are not addressed).

prohibit calculating a p-value in the surrounding regions, even though a fully sufficient description of SM processes is present in the region under investigation.

The global impact of the region veto algorithm on all event class distributions is studied based on the number of regions which are omitted from the p-value calculation because of either the relative statistical uncertainty or NH criterion. The ratio of regions failing these two criteria to all regions given as input provides an impression of how many regions are actually affected by a low generated event count in the used simulation data. Overall, around 23% of the total considered regions are found to contain an insufficient number of generated MC events, i.e., $0.65{\cdot}10^6$ regions are disregarded, whereas a p-value is calculated in $2.2{\cdot}10^6$ regions. Moreover, this ratio is determined individually in each event class distribution and shown in Fig. 11.21. Event classes contributing to the far right of the distribution represent cases where no p-value is calculated and correspond to event class distributions discussed in Tab. 11.5 as well as 14 further ones without measured data and in which many regions are already disregarded by $\mathrm{MC}/\sigma_{\mathrm{MC}}^{\mathrm{total}} < 3.0$ and the remaining ones by the NH criterion. As can be seen in Fig. 11.21, the majority of event class distributions contains more regions in which a p-value was calculated, than regions that are disregarded.

Comparison to the Uncertainty Fill-Up Method

A direct comparison to selected results obtained using the uncertainty fill-up method, to deal with regions of low MC event count, is evaluated in the following. In a first step, the full analysis chain is re-processed substituting the region veto method with the uncertainty fill-up method for all $\sum |\vec{p}_{\mathrm{T}}|$ distributions of exclusive event classes. The corresponding distribution of $\tilde{p}$-values is given in Fig. 11.22 alongside the original result obtained using the region veto method, as already discussed in Sec. 11.2.3. Similarly, good agreement is found between $\tilde{p}$-values in measured data and in SM-only pseudo-data. In fact, fewer strong deviations are observed when relying on the uncertainty fill-up approach. The introduced additional systematic uncertainty taken into account during the p-value calculation apparently reduces the significance of several deviations. Overall, the uncertainty fill-up does not seem to introduce a large amount of regions containing a severely

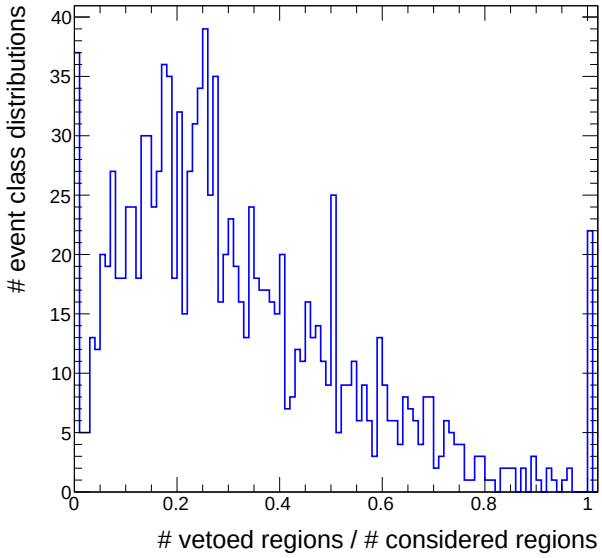
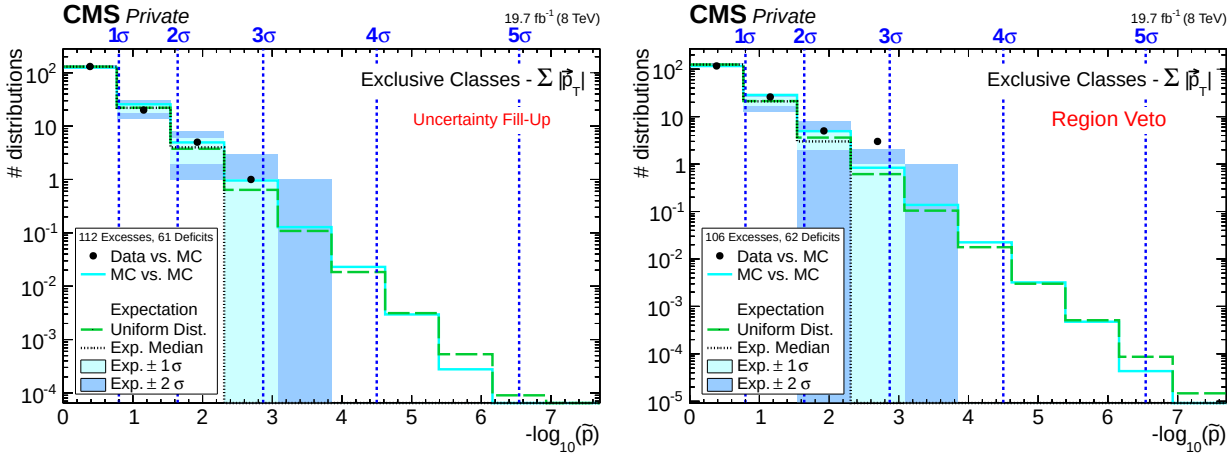


Figure 11.21: Global effect of the region veto method. For each event class and kinematic variable, the number of disregarded regions is set into relation with the total number of regions given as input to the region veto algorithm, i.e., ones that are ultimately vetoed and ones where a p-value is calculated.

overestimated systematic uncertainty, which would be visible in the comparison of data $\tilde{p}$-values to SM-only pseudo-data $\tilde{p}$-values. Overestimated uncertainties would lead to an excess of observed event classes in the first bins of the distribution.



(a) **(b)**

Figure 11.22: Distribution of $\tilde{p}$-values for $\sum |\vec{p}_{\mathrm{T}}|$ distributions of exclusive event classes. Black markers represent the measured data compared to the SM MC expectation. The result on the left is obtained using the uncertainty fill-up method, where as the result on the right (identical to Fig. 11.18a) is given using the region veto method. In the first bin a: 131 [b: 117] distributions are observed in data, with a: $130^{+5}_{-4}(1\,\sigma) \pm 9(2\,\sigma)$ [b: $125^{+5}_{-4}(1\,\sigma) \pm 9(2\,\sigma)$] expected from the SM.

Comparing the ten most significant $\sum |\vec{p}_{\mathrm{T}}|$ distributions of exclusive event classes using the uncertainty fill-up, given in Tab. C.3, to ones obtained using the region veto method (Tab. C.4) reveals that many of the same event classes are among those with the most significant deviations.

Seven out of the first ten most significant deviations using uncertainty fill-up are also among the ten most significant deviations considering the region veto method. The 2e+1γ+4jets event class, disregarded in the scan using region veto (see Tab. 11.5), is the sixth most significant $\sum |\vec{p}_{\mathrm{T}}|$ distribution when considering uncertainty fill-up. A p-value is calculated in a region containing one data event and an expectation given solely by $\sigma_{\mathrm{SM}} = 0.0035$ events of uncertainty fill-up. Three more cases of a RoI in regions containing data and only contributions from uncertainty fill-up are given among the ten most significant event classes. All of these are reproduced in the scan using the region veto, showing larger $\tilde{p}$-values for RoIs at a different position in the distribution, where a sufficient SM prediction is given.

The largest change in $\tilde{p}$-value for $\sum |\vec{p}_{\mathrm{T}}|$ distributions of exclusive event classes is found for 1μ+1γ, with $\tilde{p} = 0.17$ (fill-up) and $\tilde{p} = 0.013$ (region veto). As shown in Fig. 11.23, in both cases the RoI is located in the distribution's tail. A large amount of uncertainty fill-up is present at high $\sum |\vec{p}_{\mathrm{T}}|$ values and introduces a prediction in the two bins otherwise containing only a data event. The result in Fig. 11.23a represents a good example of how the uncertainty fill-up aims at mimicking the shape of the distribution's tail. Nevertheless, an artificial contribution on the SM expectation is added resulting in altered significance values. The issue of low SM MC generated event count is addressed differently in the region veto method. The RoI is broadened until a sufficient amount of SM MC prediction is given and a representative p-value can be calculated. The resulting p-value is more significant than the one in Fig. 11.23a.

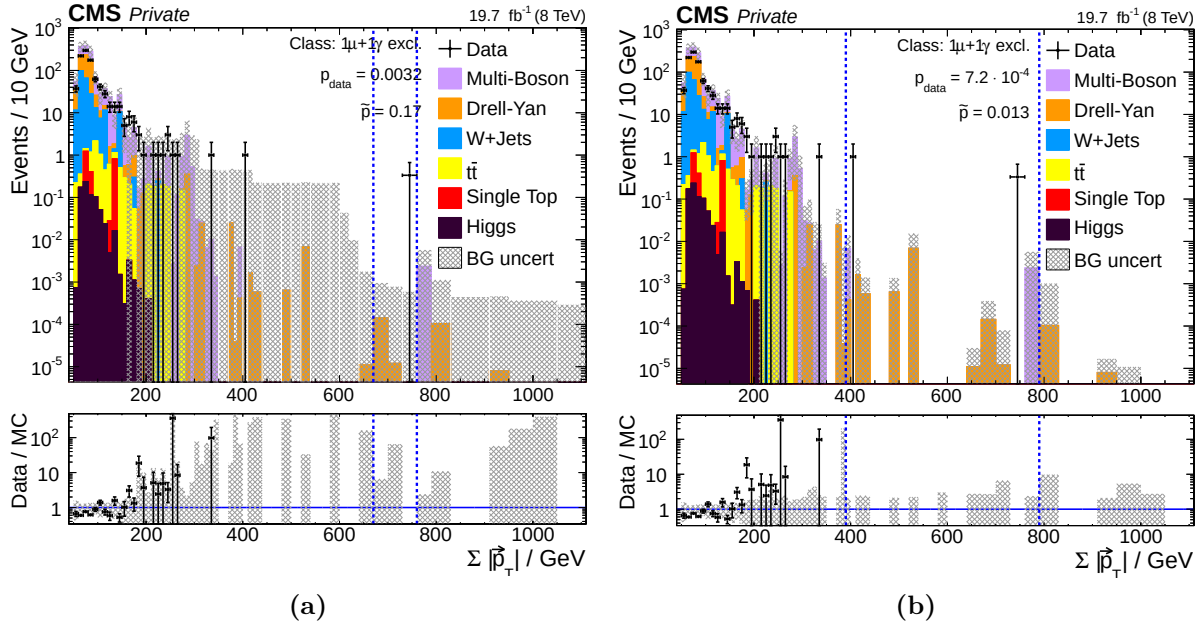


Figure 11.23: Distribution of $\sum |\vec{p}_{\mathrm{T}}|$ in 1μ+1γ. On the left the uncertainty fill-up method is used to address bins with a low number of generated MC events. The result on the right is made using the region veto method.

To understand how contributions from the uncertainty fill-up influence the considered regions, Fig. 11.24 sets $\sigma_{\text{fill-up}}$ into relation with the number of SM MC events N_{SM} in each region. Around 80% of regions are included in the first bin, where the amount of uncertainty fill-up is moderate and its influence on the p-value is small. Yet, as can be seen by the distribution's extensive tail, regions fully dominated by the additional systematic also exist. The majority of these regions are ones as shown from 440 GeV upwards in Fig. 11.23a. Only sporadic MC events are present and empty bins between these filled bins are described with the introduced uncertainty. Such regions

are recovered by the region veto algorithm and only disregarded if their total SM MC prediction is not sufficient.

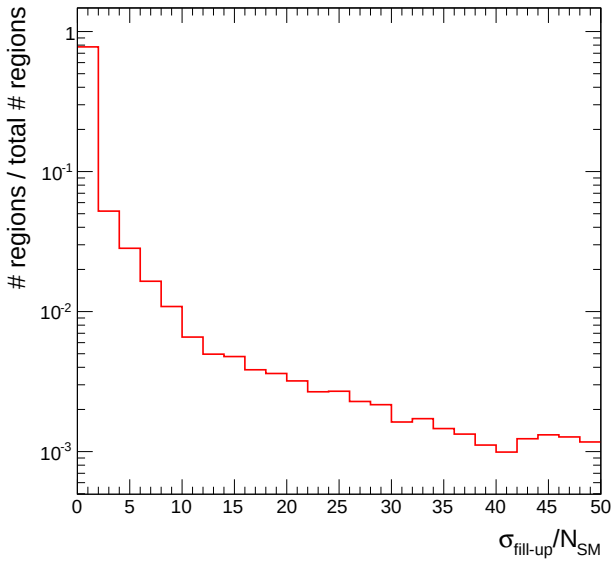


Figure 11.24: Distribution of regions in which the contribution from uncertainty fill-up $\sigma_{\text{fill-up}}$ is x-times larger than the total weighted amount of SM MC events N_{SM}. All event class types and kinematic variables are considered, resulting in $11.8{\cdot}10^6$ regions. The number of entries in each bin is normalized to this value. Only regions with $N_{\text{SM}} > 0$ are accounted for.

In extreme cases a large amount of uncertainty fill-up can lead to fully blinding valid regions of the distribution. Even strong deviations between data and the mean of the SM MC expectation are generously covered by the additional uncertainty and do not lead to significant $\tilde{p}$-values. The invariant mass distribution of 2e, discussed in Fig. 11.25a, is an ideal example for this undesirable behavior. Single bin contributions of QCD multi-jet MC below 350 GeV lead to an uncertainty of substantially over ten times the SM MC mean event yield in bins up to 700 GeV, resulting in distinct and unphysical steps in the total systematic uncertainty band. The individual systematics contributions are given relative to the weighted number of SM MC events in each bin in Fig. 11.25b. Here, uncertainty fill-up is shown in gray and clearly dominates all bins above 200 GeV. The three bin wide RoI is given around an excess of data in the tail of the distribution around 1000 GeV, resulting in a significance of $\tilde{p} = 1.0$. The result obtained in this event class distribution decisively changes when applying the region veto method, as shown in Fig. 11.25c. The one bin wide RoI with $\tilde{p} = 0.25$ is located in the regime overrun with fill-up uncertainty in Fig. 11.25a. As can be seen in Fig. 11.25d, a p-value can actually be calculated in many of the regions containing bins from 350 GeV to 700 GeV and only a small amount is disregarded because of missing SM processes found in the NH (dark blue bins). In this case, the region veto method is able to recover many significant regions that are missed by the uncertainty fill-up approach.

In consequence, the newly developed region veto algorithm provides an improved handling of regions with low generated event count over the formerly used fill-up approach. The region veto is now the method of choice for the MUSiC analysis.

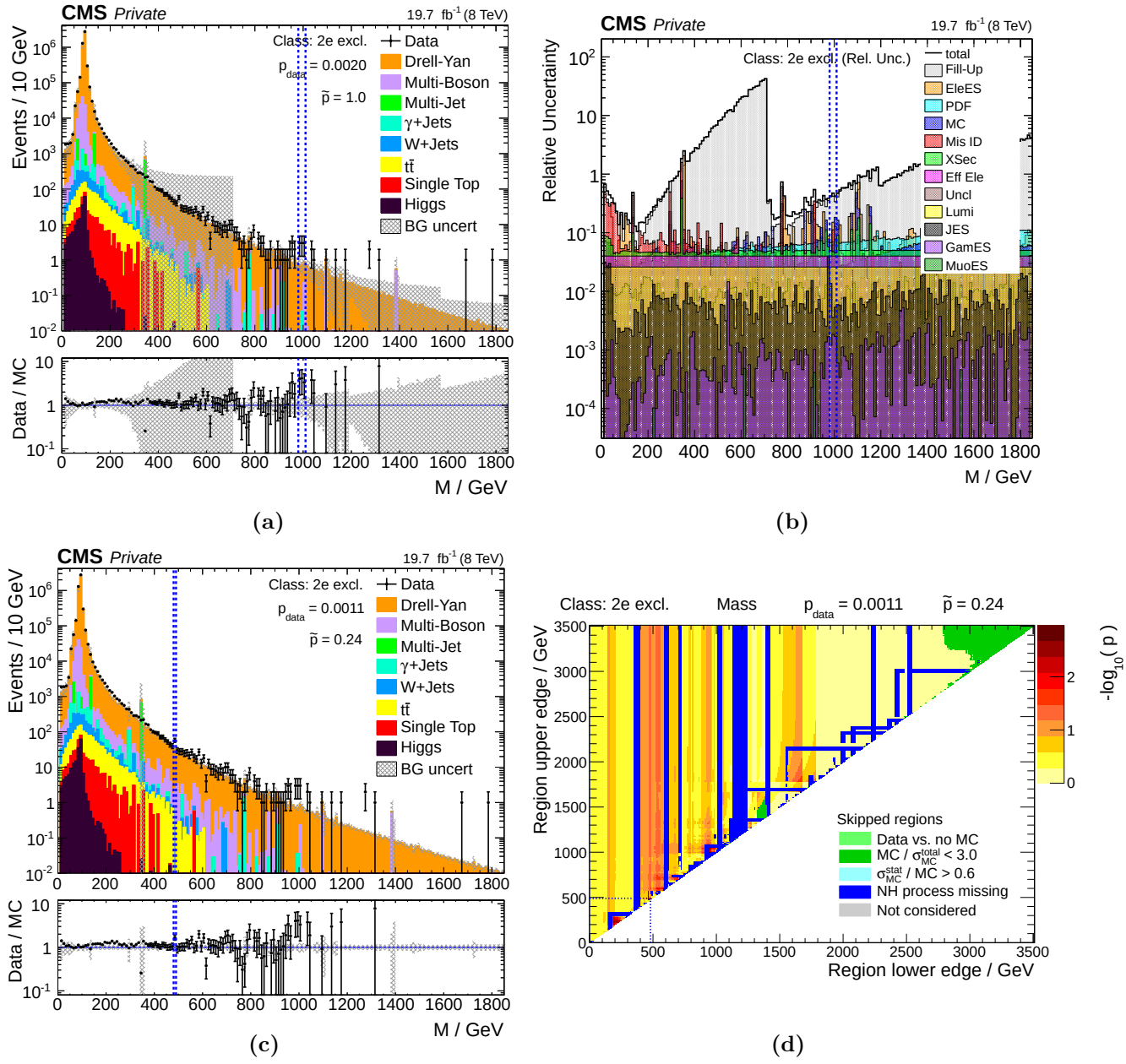


Figure 11.25: Distribution of the invariant di-electron mass. Results in the upper row use the uncertainty fill-up method to address bins with a low number of generated MC events. The distribution of individual systematic uncertainty contributions is given in Fig. 11.25b. Results in the lower row are made using the region veto method. The lower right side shows the mapping of regions to p-values or failed analysis criteria.

11.2.5 Discussion of the Most Significant Event Classes

In the following section detailed information is given on the respective two most significant deviations observed in all event class types and kinematic variables. An overview is shown in Tab. 11.6, in which groups of similar event class content are marked in the same color. Overall seven different topologies show small $\tilde{p}$-values ranging from $0.00017 \sim 3.6\,\sigma$ to $0.013 \sim 2.2\,\sigma$ within the two most significant deviations. All cases will be discussed hereafter in descending order of significance.

Scan Type		Most Significant Event Class	$p_{\text{data}}/\ \tilde{p}$	Second Most Significant Event Class	$p_{\text{data}}/\ \tilde{p}$
Total Event Yield	excl.	**1e+1γ+MET**	0.0015	**3e+2μ**	0.011
	incl.	**3e+2jets+MET+X**	0.014	**3e+2μ+X**	0.019
	jet-incl.	**3e+2jets+MET+Njet**	0.012	**2μ+1γ+Njet**	0.015
$\sum \lvert\vec{p}_{\text{T}}\rvert$	excl.	**1e+1γ+MET**	0.00097	**3e+2jets+MET**	0.0027
	incl.	**2e+1γ+4jets+X**	0.00069	**3e+2μ+X**	0.0041
	jet-incl.	**2e+1γ+4jets+Njet**	0.00015	**3e+2μ+Njet**	0.0040
$M_{(\text{T})}$	excl.	**1e+1γ+MET**	0.0020	**1μ+1γ**	0.0021
	incl.	**2e+1γ+4jets+X**	0.00071	**2μ+1jet+X**	0.0016
	jet-incl.	**2e+1γ+4jets+Njet**	0.00017	**2μ+1jet+Njet**	0.0014
MET	excl.	**1e+1γ+MET**	0.0038	**2μ+1γ+1jet+MET**	0.0039
	incl.	**2μ+1γ+1jet+MET+X**	0.0013	**3e+2jets+MET+X**	0.013
	jet-incl.	**2μ+1γ+1jet+MET+Njet**	0.0013	**3e+2jets+MET+Njet**	0.0095

Table 11.6: Overview of the two most significant event classes from each scan type. Similar event class compositions are shown in the same color. Corresponding significance values are given in each case, where the p-value represents the full significance in the scan of total event yield. The $\tilde{p}$-value is stated for scans of kinematic distributions.

A summary of the ten most significant deviations per event class type and kinematic variable is given in App. C.4, the full overview, including plots of all distributions, can be accessed on the website mentioned at the beginning of this chapter. Many of the most significant event class distributions not discussed in this thesis are dominated by SM processes with low cross sections and show only a few observed data events. To help definitively rule out contributions from BSM theories in such cases, these event classes should be closely monitored in future instances of the MUSiC analysis at higher center of mass energies and with more integrated luminosity. Due to the large amount of final states investigated by this analysis, seeing occasional deviations in the order of $3\,\sigma$ is expected purely from fluctuations of the SM only hypothesis. It is important to keep the overall good agreement in mind, as shown by the individual distributions of $\tilde{p}$-values. Significances given by the $\tilde{p}$-value represent deviations corrected for the LEE within the corresponding kinematic distribution of investigated event classes. They however do not account for a LEE between different event classes or kinematic variables.

2e+1γ+4jets+X(/+Njet)

The most significant deviation is found in the 2e+1γ+4jets+Njet event class, exhibiting a $\tilde{p}$-value of $1.5 \cdot 10^{-4} \sim 3.6\,\sigma$. Related event classes also show similarly strong significance values in both $\sum \lvert\vec{p}_{\text{T}}\rvert$ and $M_{(\text{T})}$ distributions. Tab. 11.7 summarizes these findings. As discussed previously in Sec. 11.1, all of these event classes are disregarded in the scan of total event yield because of an overall insufficient amount of SM MC simulation events. When analyzing the corresponding kinematic

distributions, a small number of regions pass all veto criteria and some of the RoIs represent the most significant ones found throughout all event class types and kinematic distributions. As already discussed in Sec. 11.2.4, the number of valid regions and also the final significance in such sparingly populated event classes depends on the position of individual SM MC events. It might be feasible to disregard all distributions of event classes failing criteria of the low generated event count treatment in the scan of total event yield. Yet, such assumptions must be introduced with care, as not to spoil the MUSiC analysis' aspiration of being inclusive and unbiased. For the particular data set analyzed in this thesis consequences of such a requirement would be small, additionally disregarding only the four distributions discussed in Tab. 11.7.

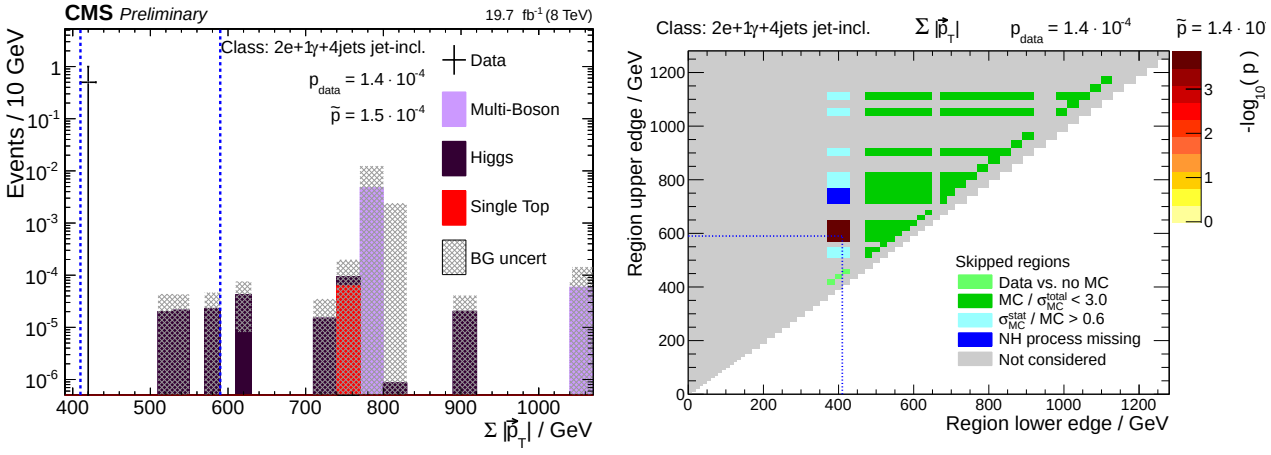
Event Class	Variable	p_{data}	$\tilde{p}$	RoI / GeV	N_{data}	$N_{\text{SM}} \pm \sigma_{\text{SM}}$
2e+1γ+4jets+Njet	$\sum \lvert\vec{p}_{\text{T}}\rvert$	$1.5 \cdot 10^{-4}$	$1.4 \cdot 10^{-4}$	410 - 590	1	$(12.9 \pm 8.1) \cdot 10^{-5}$
2e+1γ+4jets+Njet	$M_{(\text{T})}$	$1.7 \cdot 10^{-4}$	$1.5 \cdot 10^{-4}$	590 - 830	1	$(12.7 \pm 8.6) \cdot 10^{-5}$
2e+1γ+4jets+X	$\sum \lvert\vec{p}_{\text{T}}\rvert$	$6.9 \cdot 10^{-4}$	$6.9 \cdot 10^{-4}$	410 - 610	1	$(5.2 \pm 3.3) \cdot 10^{-4}$
2e+1γ+4jets+X	$M_{(\text{T})}$	$7.1 \cdot 10^{-4}$	$7.0 \cdot 10^{-4}$	630 - 740	1	$(3.9 \pm 3.1) \cdot 10^{-4}$

Table 11.7: Details of deviations in 2e+1γ+4jets+X(/+Njet) event classes sorted by increasing $\tilde{p}$-value. Stated number of data events N_{data} and SM events N_{SM} refer to the RoI.

Details of the data event contributing to these distributions can be found in Fig. 11.6. The event shows an increased amount of calorimetric energy, especially in the endcap regions. Around 30 primary vertices are reconstructed, indicating a high amount of PU. Of the seven identified objects, many are close to the required p_{T} lower threshold (100 GeV for the leading electron and 25 GeV for the sub-leading; 25 GeV for the photon; 50 GeV for the jets), indicating that the event is close to the minimum possible $\sum \lvert\vec{p}_{\text{T}}\rvert$ for this event class. The invariant mass of both electrons and the photon $M_{\text{inv}}(\text{e}, \text{e}, \gamma) = (85 \pm 1)$ GeV comes close to the Z boson on-shell mass.

The overall most significant distribution is shown in Fig. 11.26. As can be seen from the mapping of p-values or veto criteria to all investigated regions, only a very small number of valid regions are given in this case. Furthermore, the overall low number of filled bins, and thus regions, leads to only a small LEE correction, p-value and $\tilde{p}$-value are nearly identical. Distributions of the other event classes summarized in Tab. 11.7 are very similar to the one shown below in Fig. 11.26a.

The number of data and SM simulation events in 2e+1γ+4jets+Njet is influenced by the criterion used to reject electrons misidentified as photons. As explained in Sec. 9.2.4 two different approaches are available for this task, the conversion-safe electron veto and the pixel track seed veto. The MUSiC analysis relies on the latter, motivated by its smaller misidentification rate and comparable efficiency. In this final state using the conversion-safe electron veto leads to an increased total event count, $N_{\text{data}} = 2$ and $N_{\text{SM}} = 0.096 \pm 0.072$, where the same is observed for the other event classes given in Tab. 11.7. In simulation, a matching of reconstruction to generator level objects shows that the fraction of misidentified objects in these topologies is high ($\sim 80\%$), considering both the misidentification of electrons and jets as photons. As done for all misidentification rates, a fully MC simulation based approach is chosen for modeling these in the MUSiC analysis. Many dedicated analyses undergo additional effort to assess data-driven estimates for their specific final state. Determining such values is no trivial task and the respective misidentification rate depends strongly on the event composition. Although an uncertainty on the number of misidentified objects is estimated within each event class, the MUSiC analysis currently neglects migration of events from one class to another if individual object types are interchanged. Future instances of the analysis could profit from introducing such a systematic uncertainty and achieve a better modeling of event classes dominated by misidentified objects by extrapolating events into adjacent classes. However,



(a) (b)

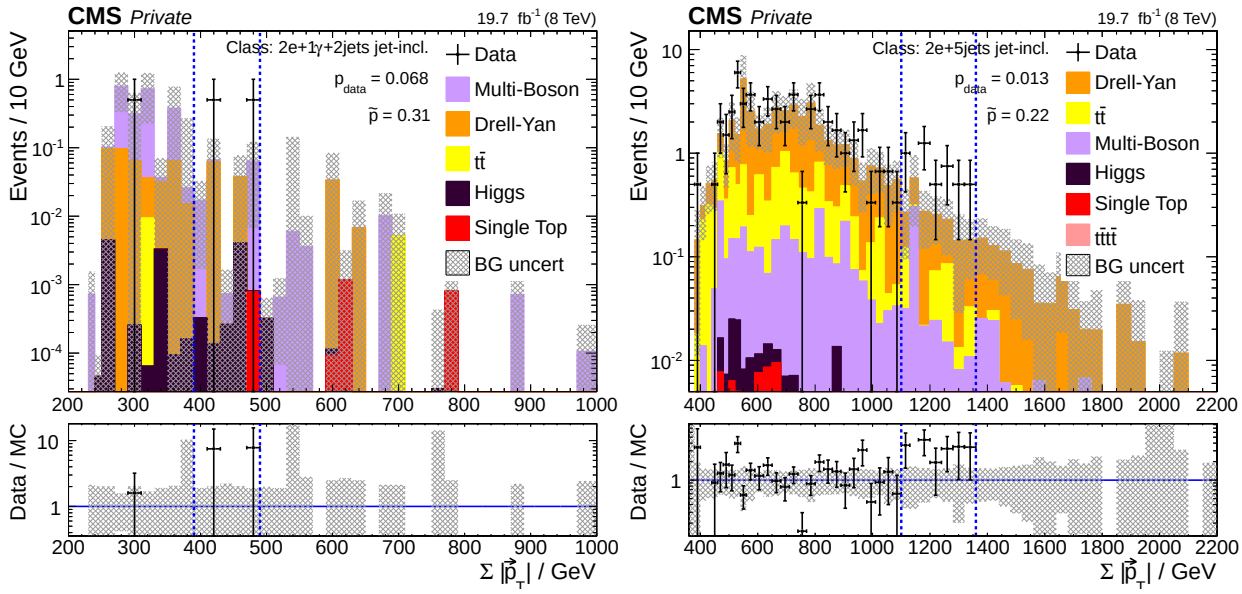
Figure 11.26: The distribution of $\sum |\vec{p}_{\mathrm{T}}|$ in events with two electrons, one photon and at least four jets is shown on the left side of this figure. The mapping of investigated regions to p-values or veto criteria is given on the left. This event class distribution contains the most significant deviation found in the analyzed data. Figure (a) is published in [5].

such a method would need to be very time-efficient, as the combinatorics involved with propagating every possible object type into all neighboring event classes are immense.

A better understanding of the observation can be gained by considering event classes with slightly altered event composition. Fig. 11.27a shows the $\sum |\vec{p}_{\mathrm{T}}|$ distribution of 2e+1γ+2jets+Njet, i.e., the event class with two instead of four explicitly required jets. Additional events are observed in data and simulation, yielding a total of $N_{\mathrm{data}} = 3$ and $N_{\mathrm{SM}} = 5.3 \pm 1.8$. Strikingly, contributions from the DY process are now found and represent the second most dominant process in this event class, individual DY events are also found in the three jet case 2e+1γ+3jets+Njet. This is an indication for an insufficient modeling of DY events with many jets and a photon in the used SM MC data sets. Only up to three jets are generated on matrix element level in the corresponding MadGraph data sets, all subsequent jets are introduced by the showering performed with pythia6. The same holds for photons from ISR and FSR. Simulation events actually contributing to 2e+1γ+4jets+Njet are ones produced with a relatively high number of events for processes with rather low cross sections.

An abundance of events is added if the required photon is replaced by a jet, as shown in Fig. 11.27b. Main contributions come from DY and $\mathrm{t\bar{t}}$ processes, which provide an overall good description of the measured data. In total $N_{\mathrm{data}} = 150$ and $N_{\mathrm{SM}} = 123 \pm 99$ events contribute to this event class. As studied in [292] the rate of jets misidentified as photons is around 0.1%. Applying this value to the number of simulation events given in 2e+5jets+Njet yields more SM events than predicted in 2e+1γ+4jets+Njet. Such estimates could be the basis for developing a systematic uncertainty on the modeling of misidentified objects in the simulation, by estimating the spillover from one event class into neighboring ones. The same effects are observed in the corresponding event classes with two muons substituted for the two electrons. In data and in simulation, somewhat more events are given in final states with two muons, which is not unexpected given the different identification schemes.

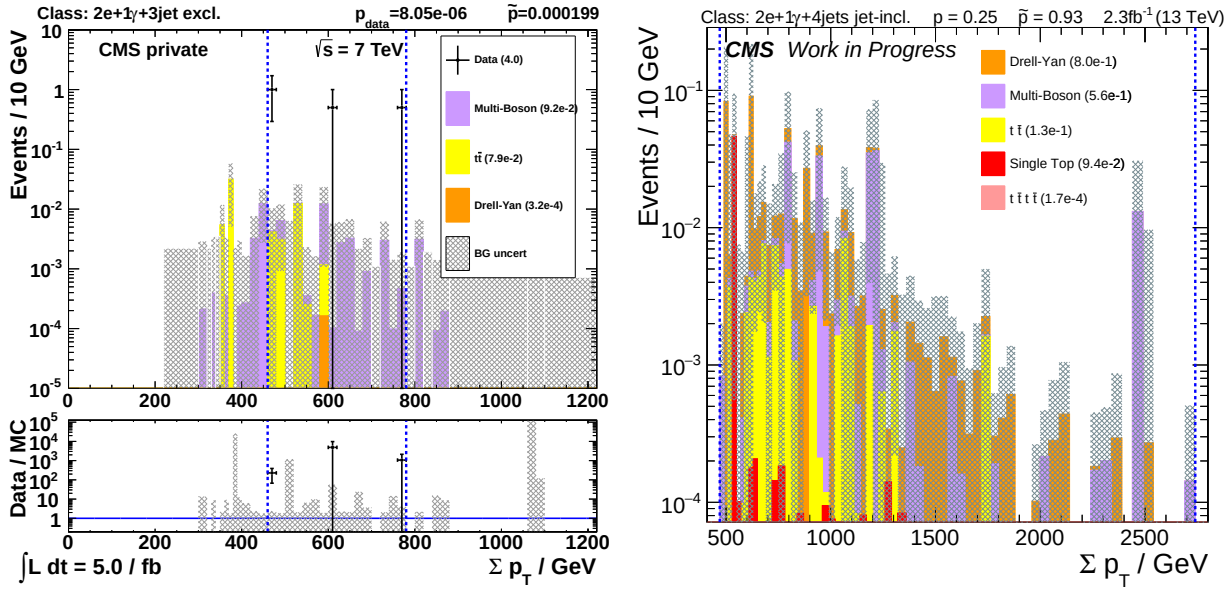
Taking a glance back at the unpublished MUSiC analysis of data taken during 2011 at $\sqrt{s} = 7$ TeV [2], shows the $\sum |\vec{p}_{\mathrm{T}}|$ distribution of 2e+1γ+3jets as the most significant deviation observed in lepton triggered events, with $\tilde{p} = 1.99 \cdot 10^{-4}$, as depicted in Fig. 11.28a. Similar conclusions concerning an insufficient DY simulation are drawn in [2]. As given in Tab. 11.7, this event class' significance is reduced to $\tilde{p} = 0.001$ in the 8 TeV MUSiC analysis. In 2011 data, the corresponding



(a) (b)

Figure 11.27: The distribution of $\sum |\vec{p}_{\mathrm{T}}|$ in 2e+1γ+2jets+Njet and 2e+5jets+Njet, which represent neighboring event classes of 2e+1γ+4jets+Njet.

final state with four jets does not contain any data events, which is not surprising considering only roughly a fourth (5.0 fb^{-1}) of the integrated luminosity, accumulated for the 2012 analysis, was available in the 2011 data set. Jet-inclusive event classes were not considered in previous analyses, so an exact comparison with the most significant event class presented here cannot be made.



(a) (b)

Figure 11.28: The distribution of $\sum |\vec{p}_{\mathrm{T}}|$ in 2e+1γ+3jets as found in the MUSiC analysis of data taken in 2011 at $\sqrt{s} = 7$ TeV (from [2]) is shown on the left, which is the most significant deviation in the corresponding analysis of lepton triggered events. Given on the right, preliminary results of 2e+1γ+4jets+Njet from the analysis of data taken during 2015 at $\sqrt{s} = 13$ TeV are shown (provided by J. Roemer).

A glance in the opposite direction, given by preliminary results of the MUSiC analysis with data taken in 2015 at $\sqrt{s} = 13$ TeV [308], yields the result shown in Fig. 11.28b. No data event is observed in 2e+1γ+4jets+Njet, considering about an eighth (2.3 fb^{-1}) of the integrated luminosity given at 8 TeV. A much higher amount of (unweighted) simulation events is found in this final state, in particular from DY. The majority of these events belong to high H_{T} DY processes. Similar events generated with MadGraph and binned in H_{T} are also available for analysis of 8 TeV, but do not deliver any events into the event classes discussed here. To sustain a simple description of the DY process only the three data set types explained in Sec. 8.2.1 are used in this analysis.

In summary, all investigations performed to understand the deviation observed in 2e+1γ+4jets event classes, point towards an insufficient description of final states with many jets and a photon in the used DY data sets. Nevertheless, it is an interesting final state and should be closely monitored in future instances of MUSiC analysis with more accumulated data.

2μ+1γ+1jet+MET (+X/+Njet)

The most significant deviations found in distributions of MET are listed in Tab. 11.8 and incorporate up to four data events in a region with under one expected event. In total 0.89 ± 0.64 are expected from the SM throughout the entire distribution of 2μ+1γ+1jet+MET+Njet and a medium size p-value of 0.036 is calculated in the scan of total event yield, where it is the fifth most significant jet-inclusive event class. Information on the four data events is summarized in Tab. 11.9.

Event Class	Variable	p_{data}	$\tilde{p}$	RoI / GeV	N_{data}	$N_{\mathrm{SM}} \pm \sigma_{\mathrm{SM}}$
2μ+1γ+1jet+MET+Njet	MET	$3.55 \cdot 10^{-4}$	0.0013	60 - 110	4	0.22±0.16
2μ+1γ+1jet+MET+X	MET	$3.57 \cdot 10^{-4}$	0.0013	60 - 100	4	0.22±0.16
2μ+1γ+1jet+MET	MET	0.0019	0.0039	60 - 100	3	0.14±0.15

Table 11.8: Details of deviations in 2μ+1γ+1jet+MET(+X/+Njet) event classes, sorted by increasing $\tilde{p}$-value. Stated number of data events N_{data} and SM events N_{SM} refer to the RoI.

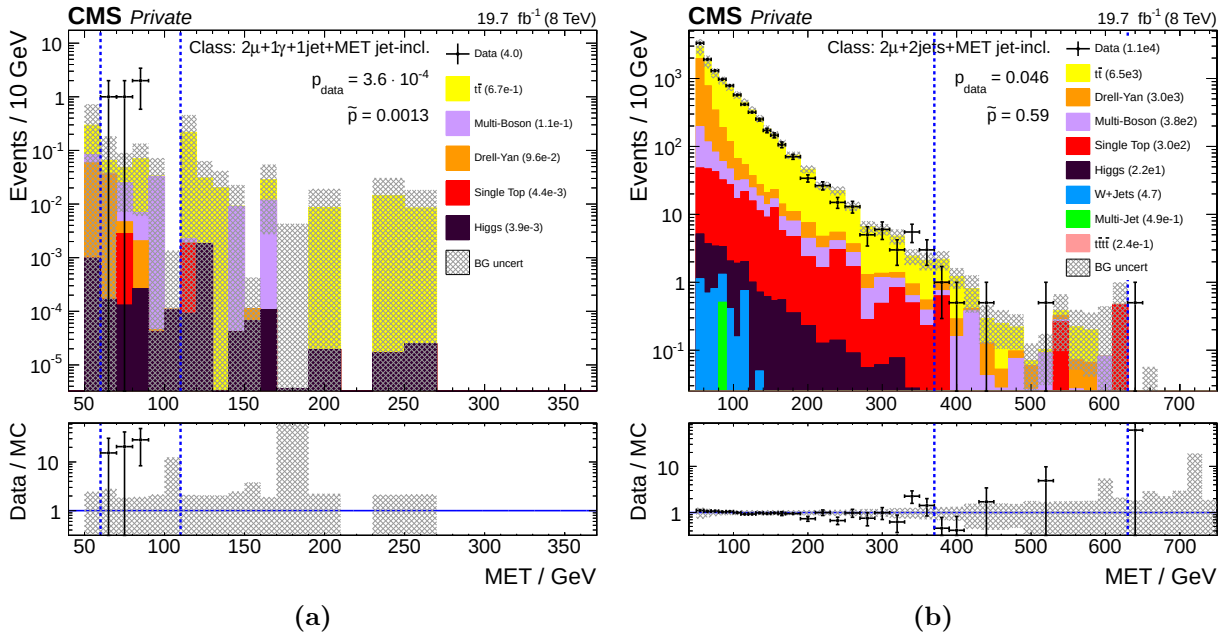
All events are triggered by a single-muon trigger and show little additional activity apart from the well separated, reconstructed objects given in the event class name. In all cases the two muons are oppositely charged, and in one event the invariant mass of the two is near the Z boson mass.

Run:LumiSec:Event $\sum \lvert\vec{p}_{\mathrm{T}}\rvert$	$M_{(\mathrm{T})}$	MET	$M_{\mu\mu}$	$M_{\mu\mu\gamma}$	$p_{\mathrm{T}}(\mu)$	$p_{\mathrm{T}}(\gamma)$	p_{T}(jet)
196218:536:789930803							
792±30	7851±25	63±20	311±42	507±20	162±7, 222±11	75±1	95±10, 270±18
200991:564:714662585							
426±21	421±18	71±15	70±14	89±7	27±1, 130±5	31±1	166±13
203002:395:535204974							
334±16	333±16	81±13	35±10	94±5	46±1, 57±1	68±1	81±9
207487:414:668892522							
326±17	326±17	86±13	92±9	221±5	43±1, 45±1	45±1	108±10

Table 11.9: Details of data events observed in 2μ+1γ+1jet+MET(+X/+Njet) event classes. All kinematic values are given in GeV. Stated uncertainties are determined using resolutions given in Sec. 7.2.3.

The MET distribution of events in 2μ+1γ+1jet+MET+Njet is given in Fig. 11.29, revealing $\mathrm{t\bar{t}}$ to be the main contribution but events from multi-boson and DY processes also play an important role. In comparison to the exclusive case, a slightly larger $\mathrm{t\bar{t}}$ and multi-boson contribution is found

in the inclusive and jet-inclusive event class, while the amount of DY events remains the same. Overall, the systematic uncertainty in this event class is driven by the low number of generated simulation events.



(a) (b)

Figure 11.29: The distribution of MET in events of 2μ+1γ+1jet+MET+Njet is shown on the left side of this figure. This is the most significant MET distribution. The right side shows the MET distribution in the neighboring 2μ+2jets+MET+Njet event class.

Similar to the discussion presented above for 2e+1γ+4jets (+X/+Njet) event classes, the rate of jets misidentified as photons in simulated events seems to be different than the one given by the observation. Considering the neighboring event class 2μ+2jets+MET+Njet, obtained by replacing the photon with a jet, show a larger relative contribution of DY events. As can be seen in Fig. 11.29b, DY represents the second largest overall SM contribution at only about 50% fewer events then expected from $t\bar{t}$. Here the relative multi-boson event yield is substantially smaller than what is found in 2μ+1γ+1jet+MET+Njet. The observed deviation is far less significant with a large $\tilde{p}$-value of 0.59. Apparently not all simulation data sets deal photons from ISR/FSR or with jets misidentified as photons in an equally good and sufficient manner.

The same trend is found in past and future instances of this analysis. Both the MUSiC analysis of data taken during 2011 at $\sqrt{s} = 7$ TeV [2] and during 2015 at $\sqrt{s} = 13$ TeV [308] observe a different relative contribution of the individual SM processes and more significant deviation in the 2μ+1γ+1jet+MET than in the 2μ+2jets+MET event class.

1e+1γ+MET

Overall, 1e+1γ+MET is the most significant exclusive event class in all three kinematic variables, for which details are given in Tab. 11.10. Moreover, with a p-value of 0.0015 ($N_{\text{data}} = 18$, $N_{\text{SM}} = 3.3 \pm 3.3$), it is also the most significant exclusive class the scan of total event yield. In this event class the final p-value observed in the total event yield can become more significant than the $\tilde{p}$-value determined from a kinematic distribution, since the large number of bins in the distributions introduces a substantial LEE correction. Although the distribution of events in 1e+1γ+MET is very similar in $\sum |\vec{p}_{\text{T}}|$ and MET, a stronger LEE is found for $M_{(\text{T})}$. Since a minimum of one and not three bins is imposed on the potential RoI for $M_{(\text{T})}$, a higher amount of regions is considered.

Event Class	Variable	p_{data}	$\tilde{p}$	RoI / GeV	N_{data}	$N_{\text{SM}} \pm \sigma_{\text{SM}}$
1e+1γ+MET	$\sum \lvert\vec{p}_{\text{T}}\rvert$	$1.05 \cdot 10^{-4}$	$9.7 \cdot 10^{-4}$	270 - 330	7	0.47 ± 0.51
1e+1γ+MET	$M_{(\text{T})}$	$1.98 \cdot 10^{-4}$	0.0020	270 - 330	7	0.49 ± 0.58
1e+1γ+MET	MET	$5.36 \cdot 10^{-4}$	0.0038	50 - 90	14	1.8 ± 2.1

Table 11.10: Details of deviations in the 1e+1γ+MET event class, sorted by increasing $\tilde{p}$-value. Stated number of data events N_{data} and SM events N_{SM} refer to the RoI.

Fig. 11.30 showa $\sum \lvert\vec{p}_{\text{T}}\rvert$ and MET distributions of 1e+1γ+MET. In both cases the RoI is three bins wide and contains an excess of data located in the distribution's bulk. A telltale sign of an insufficient SM description is found here once again. Systematic uncertainties in this event class are driven by misidentified objects and a low number of generated MC events. In $\sum \lvert\vec{p}_{\text{T}}\rvert$ and $M_{(\text{T})}$ the lower end of the distribution shows bins containing measured data but no SM prediction. In this regime the search region is broadened by the low generated event count treatment.

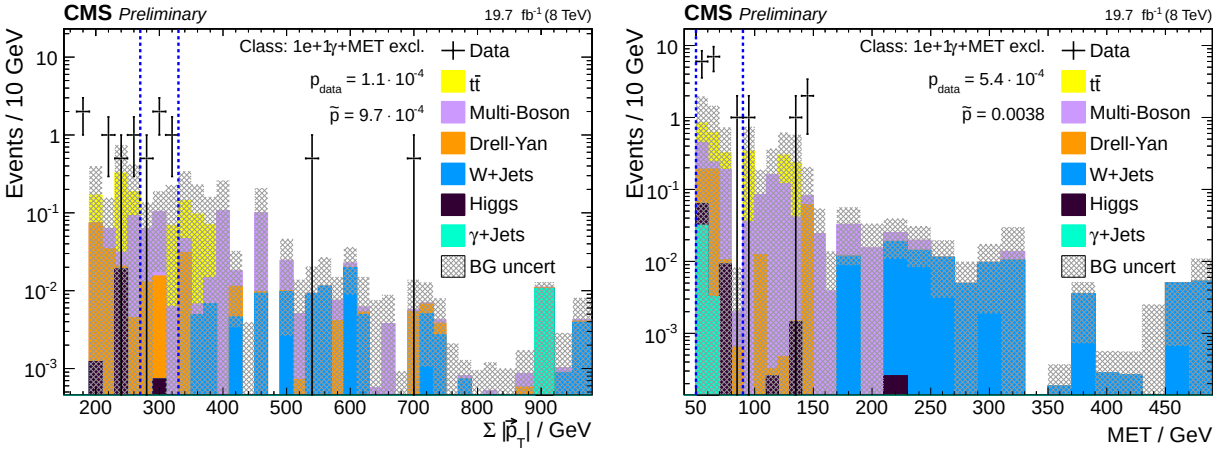


Figure 11.30: The distribution of $\sum \lvert\vec{p}_{\text{T}}\rvert$ and MET in events of 1e+1γ+MET. This is the most significant event class in the scan of total event yield. Both figures are published in [5].

Strikingly, W+jets is not the main contribution in 1e+1γ+MET and even completely missing in the RoI, although is is by far the dominant process in event classes with a single-lepton, MET and potentially jets. Due to the precise way the individual W+jets contributions are spread out over the bins, the NH criterion of the low generated event count treatment does not consider it to be a dominant process for the determined RoI.

Furthermore, exchanging the electron veto, used in photon selection from the pixel track seed veto to the conversion-safe electron veto, increases the number of events observed in this final state to $N_{\text{data}} = 70$ vs. $N_{\text{SM}} = 70 \pm 33$, yielding an insignificant $\tilde{p}$-value of 0.96. Since the conversion-safe electron veto is prone to a higher rate of electrons misidentified as photons, obtaining these additional events by such misidentified objects is a probable scenario. A glance into the neighboring 2e+MET, as given in Fig. 11.17, shows a much higher relative contribution of DY events.

In conclusion, this final state suffers from an insufficient modeling of especially the W+jets and DY processes in the used SM data sets. As already seen before, the presence of a photon obtained either from ISR/FSR or by misidentification of a jet or electron may be simulated at a different rate than observed in data.

1μ+1γ

A similar event topology as the one described in the previous paragraph is given by 1μ+1γ, which shows up as the second most significant $M_{(\mathrm{T})}$ and sixth most significant $\sum|\vec{p}_{\mathrm{T}}|$ distribution. As illustrated in Fig. 11.31a, an excess of three data events over $N_{\mathrm{SM}} = 0.032 \pm 0.022$ events expected from the SM is found in the tail of the mass distribution from 460 GeV to 830 GeV, yielding a significance of $\tilde{p} = 0.0021 \sim 2.9\,\sigma$. Two of the data events lie in bins were no SM expectation is present and the corresponding regions are disregarded by the low generated event count treatment. Throughout the majority of bins, systematic uncertainties are dominated by a low number of generated MC events and a high fraction of misidentified objects, given in blue and red in Fig. 11.31b. Considering the total event yield in this class a deficit of data is found with $N_{\mathrm{data}} = 937$ and $N_{\mathrm{SM}} = 1320 \pm 380$ leading to an insignificant p-value of 0.16.

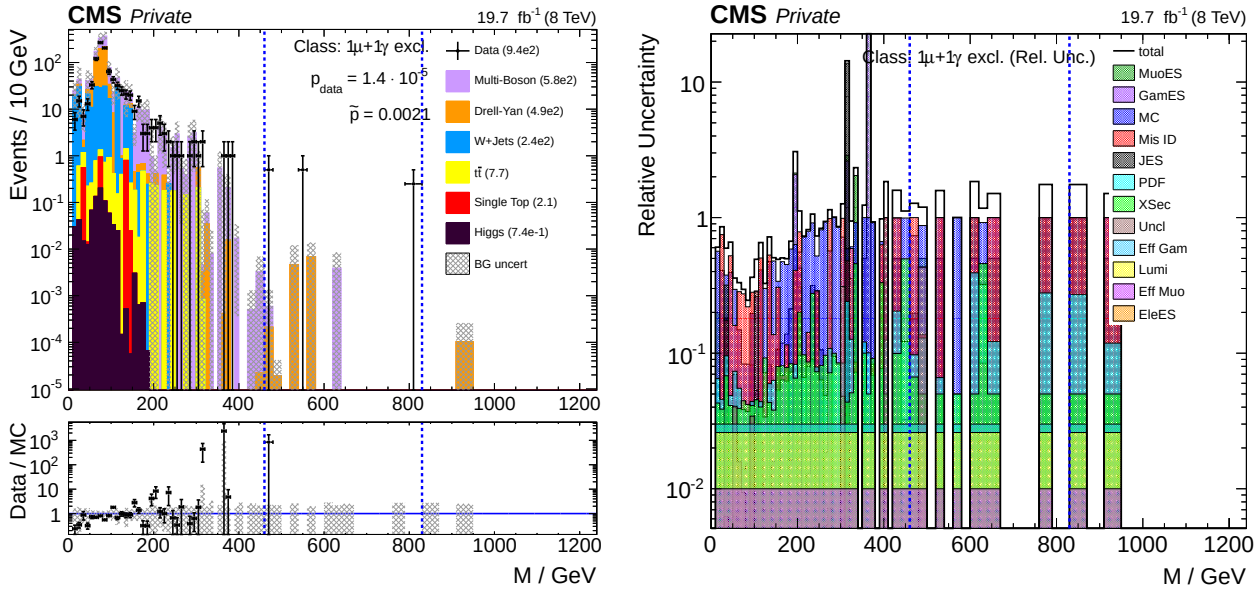


Figure 11.31: The invariant mass distribution of events in 1μ+1γ. On the right side individual systematic uncertainty contributions in each bin, relative to the total number of expected SM events, are given.

Details of the three data events observed in the RoI are listed in Tab. 11.11. In all cases the muon and photon have comparable transverse momenta and little other activity is recorded in the event. The two particles occur back-to-back in the $\rho - \phi$ plane, where those found in the first two events have a boost in approximately z direction.

Run:LumiSec:Event	$\sum\|\vec{p}_{\mathrm{T}}\|$/GeV	$M_{(\mathrm{T})}$/GeV	MET/GeV	$p_{\mathrm{T}}(\mu)$/GeV	$p_{\mathrm{T}}(\gamma)$/GeV
196452:745:1017011550	402±9	544±16	6±14	193±9	210±1
200190:342:417596149	749±24	812±31	44±19	352±24	397±1
203002:1537:1736105409	338±7	460±12	3±13	162±7	176±1

Table 11.11: Details of the three data events found in the RoI of 1μ+1γ. In all cases MET is below the analysis level threshold of 50 GeV and is disregarded. The stated rough uncertainties on kinematic event variables are determined using resolutions given in Sec. 7.2.3.

As discussed previously, a glance into the neighboring event class by exchanging the photon with a jet can help understand the deviation found in 1μ+1γ. In 1μ+1jet, the $4.7{\cdot}10^6$ measured

data events are reproduced well by the SM simulation, W+jets being the main contribution with $3.2{\cdot}10^6$ events, followed by DY ($8.5{\cdot}10^5$ events), QCD multi-jet ($4.7{\cdot}10^5$ events) and multi-boson ($1.5{\cdot}10^5$ events) processes. The relative SM composition found in 1μ+1γ is quite different, especially contributions from W+jets are completely lacking in the tail of the distribution.

Exchanging the pixel track seed veto for the conversion-safe electron veto approximately doubles the amount of events in this final state, with $N_{\text{data}} = 2700$ and $N_{\text{SM}} = 3230 \pm 900$ in total. Moreover, the 1e+1μ shown at the beginning of this chapter in Fig. 11.14a exhibits a very good agreement and similar relative SM contribution to the event class discussed here.

Apparently the excess observed in 1μ+1γ is caused by the same effect as uncovered for previous most significant event class distributions. A difference is found between data and simulation in the modeling of photons originating from ISR/FSR or by misidentification of a jet or electron.

2μ+1jet+X(/+Njet)

Two individual effects become visible in the mass distribution of 2μ+1jet+X(/+Njet), which are among the four most significant $M_{(\text{T})}$ distributions in all event class types. As shown in Fig. 11.32a, a $\tilde{p}$-value of 0.0016 is evaluated for 2μ+1jet+X in the two bin wide RoI from 100 GeV to 120 GeV given by an excess in data of $N_{\text{data}} = 2269$ vs. $\sigma_{\text{SM}} = 1530{\pm}170$. DY is by far the most contributing process and a good normalization is given around 200 GeV at the peak of the distribution. At higher mass values the systematic uncertainties become larger, mainly because of the uncertainty assumed on the LO k-factor used for $p_{\text{T}}(\gamma^*/Z)$ binned DY data sets which contribute dominantly in this regime. The three data set types combined to model the DY process are shown individually in Fig. 11.32b. To account for high jet multiplicities the bulk of the DY events are generated binned in the invariant di-lepton mass with MadGraph, labeled as "Drell-Yan M (MG)" in the plot. These data sets are most dominant in the region before the distribution's peak. The NLO shape for the tail of the distribution is addressed by using a group of di-lepton mass binned data sets produced with powheg and labeled as "Drell-Yan M (PH)", having only a very small influence in the tail of the 2μ+1jet+X mass distribution. Furthermore, data sets binned in $p_{\text{T}}(\gamma^*/Z)$ generated with MadGraph are used to include events with a boosted topology, given as "Drell-Yan p_{T} (MG)". For more details on the data sets see Sec. 8.2.1. The combined uncertainties on assumed SM cross sections are shown in green in Fig. 11.32c and are driven by the uncertainty assumed for the LO cross section of the $p_{\text{T}}(\gamma^*/Z)$ binned DY data sets. The agreement in terms of event yield between data and SM simulation is only moderate in the distribution's tail, yet fully covered by the systematic uncertainty band.

The very low mass regime of the distribution below the actual RoI, as shown by a closeup in Fig. 11.32b, suffers from an effect of the DY MC production. As explained in Sec. 8.2.1, events for this group of data sets were generated only for an γ^*/Z mass above 10 GeV and an analysis level requirement of $M_{(\text{T})} \geq 10$ GeV is imposed on all event classes. The exact value of the threshold is a compromise chosen to treat all event classes consistently. For the distribution discussed here it could be tightened, since the invariant mass also includes the jet four vector. For other event compositions a higher threshold would reject well modeled events. This effect is not visible in $\sum |\vec{p}_{\text{T}}|$, since a minimum threshold calculated from the object p_{T} selection is applied. Fortunately, the MUSiC low MC generated event count treatment takes care of the few critical bins at low mass in 2μ+1jet+X, where no DY events are present. The first seven bins of the distribution, up to 100 GeV, are disregarded. As can be seen in Fig. 11.32d the majority of regions do not contain all necessary processes from the NH regions.

A second effect becomes visible in the intermediate mass regime around 100 GeV to 150 GeV, where an undershoot of SM simulation events is given below the distribution's peak (see Fig. 11.32b). An analogous observation is made in $\sum |\vec{p}_{\text{T}}|$ and a similar trend is found in the exclusive and jet-

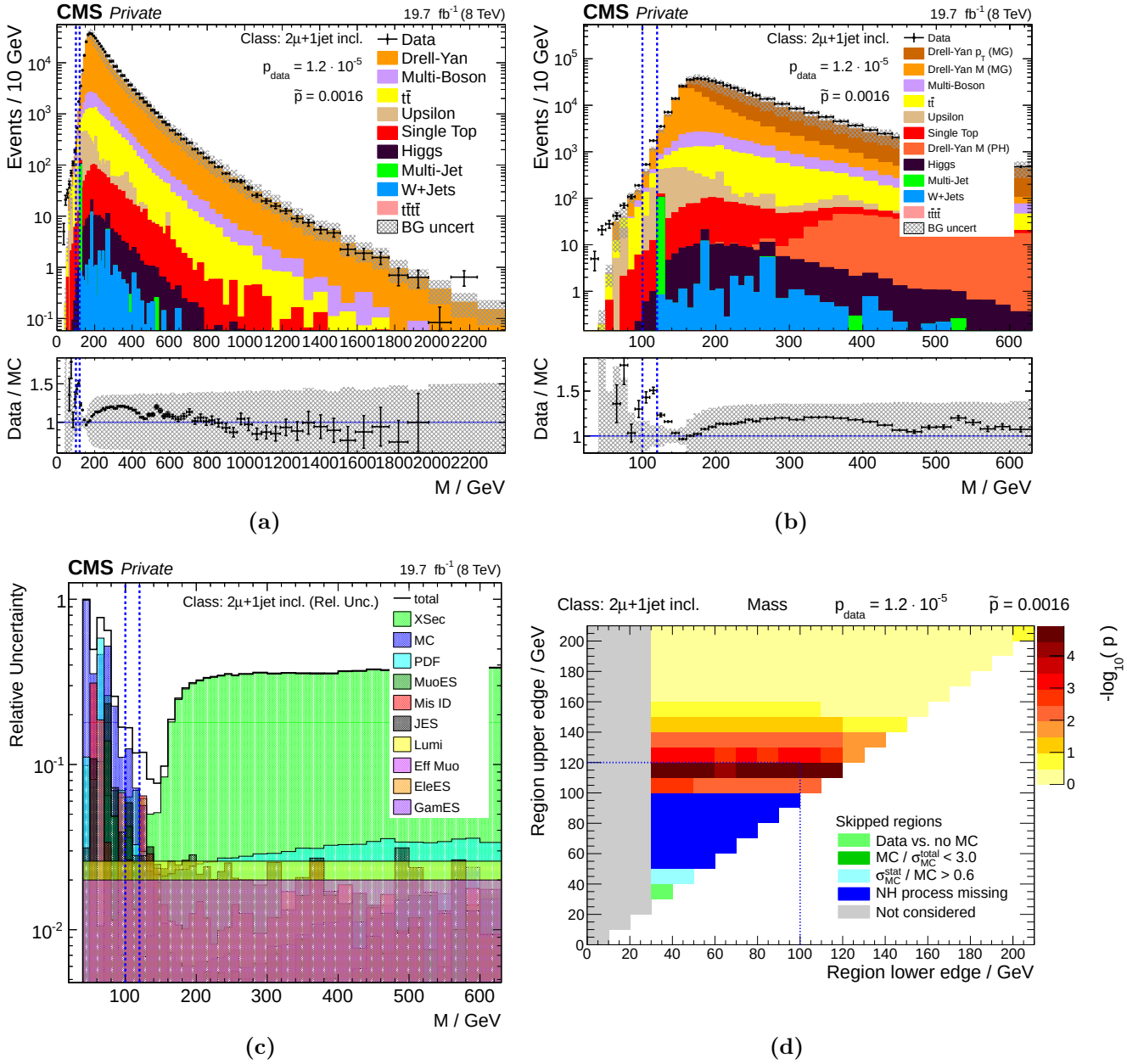
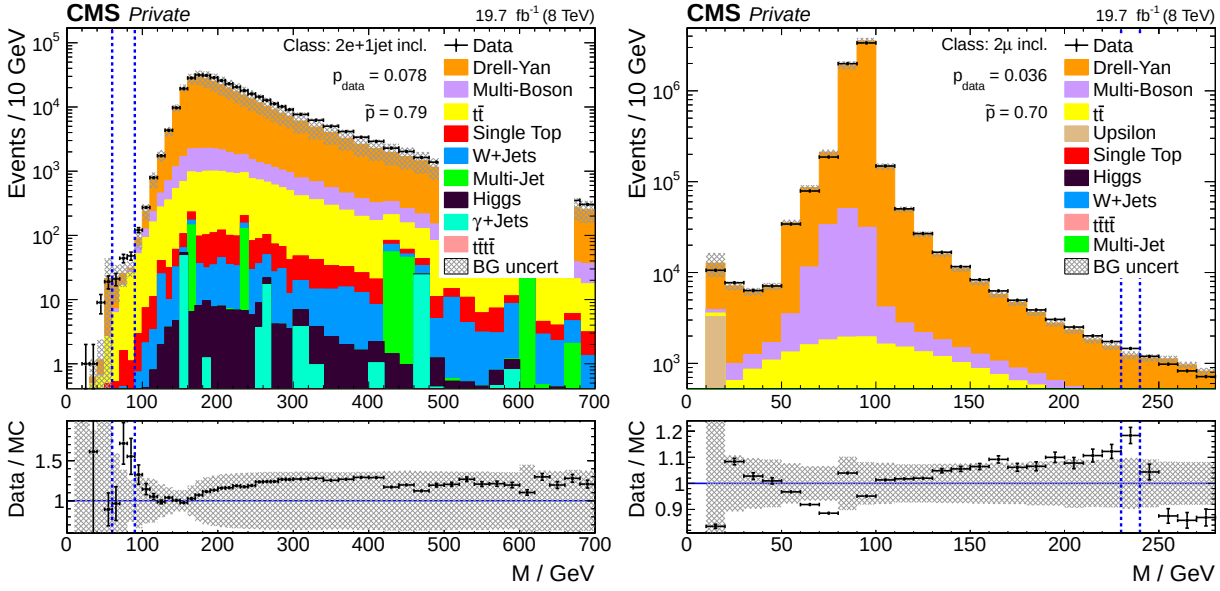


Figure 11.32: Fig. 11.32a shows the mass distribution in events of 2μ+1jet+X in its entirety. This is one of the most significant $M_{(\mathrm{T})}$ distributions. On the right side, in Fig. 11.32b, a closeup of the low mass range emphasizes what is observed in the RoI. Here contributions from DY are split up into the different contributing data set types. Fig. 11.32c shows the corresponding relative systematic uncertainties. For bins below 200 GeV in mass, the mapping of regions to p-values or criteria of the low generated event count treatment is shown in Fig. 11.32d. Ratio plots are given with a linear abscissa.

inclusive event classes. Moreover, the same feature is also found in the corresponding event class with electrons 2e+1jet+X, as shown in Fig. 11.33a. Here the deviation is far less significant, which is mainly due to larger systematic uncertainties. Contributions from misidentified objects are larger than in the muon case, which are given in red in Fig. 11.32c. However, the effect does not appear in events containing a di-lepton pair without the additional jet requirement, as given for example by 2μ+X in Fig. 11.33b. Here, the intermediate mass regime shows better agreement and no strong

deviation is found. The RoI is located at the transition region between DY data sets generated with MadGraph and high mass data sets generated with powheg, but the $\tilde{p}$-value of $0.70 \sim 0.52\,\sigma$ is absolutely insignificant.



(a) **(b)**

Figure 11.33: Mass distribution in events of 2e+1jet+X and 2μ+X. In both cases the ordinate range is chosen to emphasize the low mass region. More data events are located at higher values not shown in the plots. Ratio plots are given with a linear abscissa.

These observations lead to the conclusion that the required jets have an influence which shows up in the intermediate mass region of 2μ+1jet+X and similar event classes. Throughout the analysis two procedures are adopted to improve the description of jets in the SM simulation. As explained in Sec. 10.1, the applied jet energy scale corrections introduce a systematic uncertainty which accounts for shifts within the kinematic distribution but also migration of events to neighboring event classes. Especially the low energy range is susceptible to low momentum jets being shifted below or above the selection threshold of $p_{\mathrm{T}} \geq 50$ GeV and the resulting JEC uncertainty is given by a black histogram in Fig. 11.32c. Nevertheless, the uncertainty in mass range of the RoI is dominated by the number of generated MC events. Furthermore, a smearing of the jet p_{T} resolution is performed to account for differences observed between data and simulation (see Sec. 9.2.5). The uncertainties on this procedure are relatively small (2-5% of the correction factor), but may become large in individual kinematic regions, such as the one observed in 2μ+1jet+X. The effect does not appear in 2μ+X or 2μ+Njet, since the jets themselves are not considered in the kinematic variable. Future instances of the MUSiC analysis could profit from including the effect of the JER uncertainty in their SM prediction to achieve a better modeling of final states such as the one discussed here. The effect of this additional systematic on the work presented in this thesis would be small and only noticeably contribute to results in few event classes.

Moreover, the low mass region of 2μ+1jet+X represents events with extreme kinematics. Considering the minimum p_{T} requirements applied during the selection an event must have at least $\sum |\vec{p}_{\mathrm{T}}| > 100$ GeV (assume double-muon trigger, then 25 GeV for each muon and 50 GeV for the jet). In the mass distribution events below this threshold contain strongly boosted objects, produced less frequently.

3e+2jets+MET (+X/+Njet)

Event classes containing 3e+2jets+MET(+X/+Njet) are among the most significant deviations in all kinematic variables and also in the scan of total event yield. In all cases an excess of data over the SM expectation is observed. The smallest $\tilde{p}$-values are given in the three event class distributions summarized in Tab. 11.12. Overall the significance values in this group of event classes ranges from $\tilde{p} = 0.0027 \sim 2.8\,\sigma$ in $\sum |\vec{p}_{\mathrm{T}}|$ for the exclusive to $\tilde{p} = 0.017 \sim 2.1\,\sigma$ in $\sum |\vec{p}_{\mathrm{T}}|$ for the inclusive case.

Event Class	Variable	p_{data}	$\tilde{p}$	RoI / GeV	N_{data}	$N_{\mathrm{SM}} \pm \sigma_{\mathrm{SM}}$
3e+2jets+MET	$\sum \lvert\vec{p}_{\mathrm{T}}\rvert$	$10.0 \cdot 10^{-5}$	0.0027	950 - 1230	6	0.55±0.24
3e+2jets+MET+Njet	MET	$9.6 \cdot 10^{-4}$	0.0095	90 - 190	15	4.4±1.5
3e+2jets+MET+X	MET	0.0012	0.013	90 - 190	15	4.6±1.5

Table 11.12: Details of deviations in 3e+2jets+MET(+X/+Njet) event classes, sorted by increasing $\tilde{p}$-value. Stated number of data events N_{data} and SM events N_{SM} refer to the RoI.

The distribution of events for the two most significant cases is shown in Fig. 11.34. Throughout nearly the entire distribution more events are measured in data than expected from the SM MC prediction. Multi-boson and $t\bar{t}$ production are the dominant processes and appear rather smooth in the respective bulk of the distribution. The systematic uncertainties are dominated by impreciseness of the SM cross section estimates, the uncertainty stemming from the number of generated MC events plays only a secondary role.

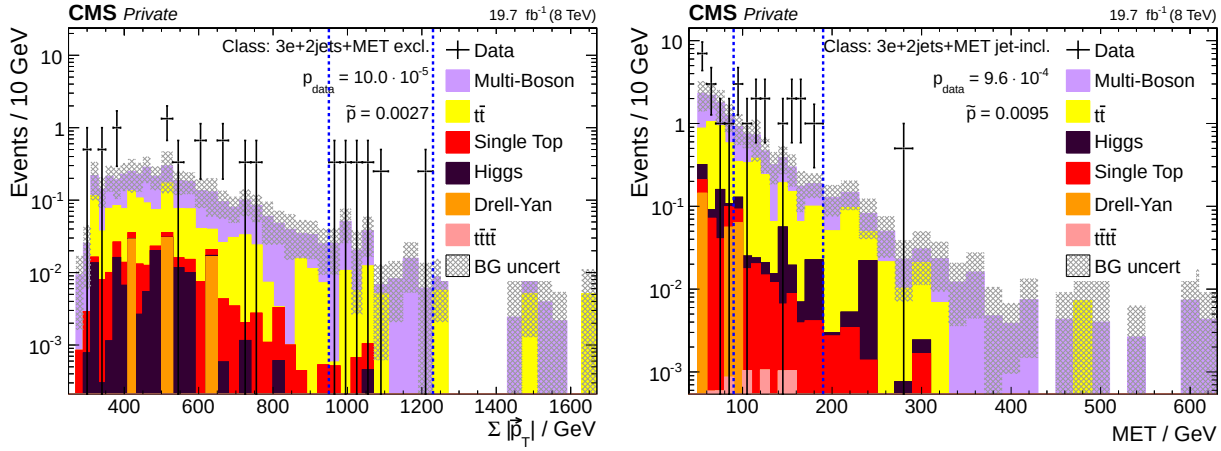


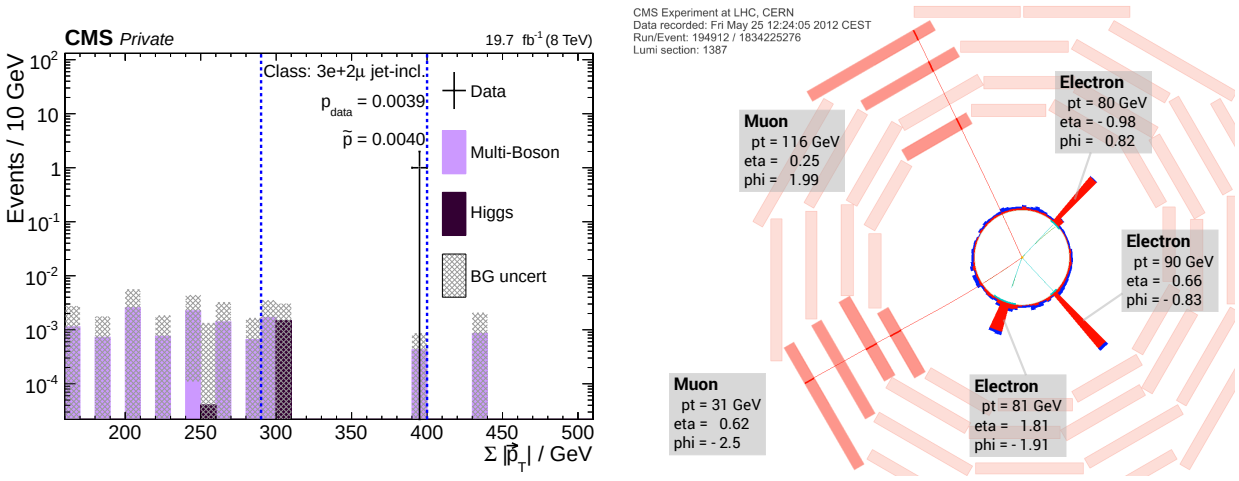
Figure 11.34: The distributions of $\sum |\vec{p}_{\mathrm{T}}|$ in events of 3e+2jets+MET and MET in 3e+2jets+MET+Njet.

Interestingly, neighboring event classes with one or three jets do not show such strong deviations. In $\sum |\vec{p}_{\mathrm{T}}|$ of 3e+1jet+MET a $\tilde{p}$-value of 0.80 is found. The RoI includes a single data event in the tail of the distribution ($N_{\mathrm{SM}} = 0.115 \pm 0.079$). Overall 36 data events are observed with 36±17 expected from the SM. The three jet case 3e+3jets+MET is more significant than the one jet event class, but still an order of magnitude less significant than 3e+2jets+MET. The RoI encloses all six observed data events against an SM expectation of 1.46 ± 0.58 events, leading to $\tilde{p} = 0.059$. Similarly, the (jet-)inclusive and three jet event classes are less significant than their two jet counterpart. Since neighboring event classes with more and fewer jets are noticeably less significant, it is unlikely that the deviation seen in 3e+2jets+MET(+X/+Njet) originates from an unknown physics process. Yet, no definitive conclusion can be made from the current amount of data.

Only one data event is observed in 3e+2jets+MET for both the MUSiC analysis of data taken during 2011 at $\sqrt{s} = 7$ TeV [2] and of data taken during 2015 at $\sqrt{s} = 13$ TeV [308]. A small total SM yield is expected in both cases, and the deviation proves to be insignificant, with $N_{\mathrm{SM}} = 3.1 \pm 1.0$, $\tilde{p} = 0.99$ (2011) and $N_{\mathrm{SM}} = 0.52 \pm 0.23$, $\tilde{p} = 0.35$ (2015). In the inclusive case (jet-inclusive event classes are not investigated with 2011 data) the findings are similarly insignificant, with $N_{\mathrm{data}} = 3$, $N_{\mathrm{SM}} = 3.9 \pm 1.2$ in total, $\tilde{p} = 0.83$ (2011) and $N_{\mathrm{data}} = 1$, $N_{\mathrm{SM}} = 0.79 \pm 0.37$ in total, $\tilde{p} = 0.36$ (2015). Additional data-taking, resulting in a higher integrated luminosity is needed to gain a better insight of these final states.

3e+2μ+X/+Njet

Finally, an excess of a single data event over a small SM expectation of $N_{\mathrm{SM}} = 0.0036 \pm 0.0024$ events in 3e+2μ(+X/+Njet) event classes is discussed. Significances of $\tilde{p} \sim 2.6\,\sigma$ are observed, with details given in Tab. 11.13. These event classes are also among the two most significant in the scan of total event yield, where their p-values are roughly three times larger than the $\tilde{p}$-values given below. The $\tilde{p}$-value of the inclusive event class is slightly larger than that of the jet-inclusive case, shown in Fig. 11.35a, since an additional multi-boson contribution at around 750 GeV induces more regions, increasing the LEE.



(a) **(b)**

Figure 11.35: The left plot shows the distribution of $\sum |\vec{p}_{\mathrm{T}}|$ in events of 3e+2μ+Njet. A visualization of the data event (run:lumi section:event - 194912:1387:1834225276) found in the RoI is shown on the right side of this figure. Apart from the five leptons given in the event class name the event shows little activity. All objects are separated well from each other. The three electrons are of comparable p_{T}, whereas a large difference in p_{T} is observed between the two muons. MUSiC kinematic variables of this event are: $\sum |\vec{p}_{\mathrm{T}}| = (399 \pm 4)$ GeV; $M_{(\mathrm{T})} = (589 \pm 6)$ GeV; MET $= (30 \pm 14)$ GeV (disregarded in analysis); The stated rough uncertainties on kinematic event variables are determined using resolutions given in Sec. 7.2.3.

Event Class	Variable	p_{data}	$\tilde{p}$	RoI / GeV	N_{data}	$N_{\mathrm{SM}} \pm \sigma_{\mathrm{SM}}$
3e+2μ+Njet	$\sum \lvert\vec{p}_{\mathrm{T}}\rvert$	0.0039	0.0040	290 - 400	1	0.0036±0.0024
3e+2μ+X	$\sum \lvert\vec{p}_{\mathrm{T}}\rvert$	0.0039	0.0041	290 - 400	1	0.0036±0.0024

Table 11.13: Details of deviations from in 3e+2μ event classes, sorted by increasing $\tilde{p}$-value. Stated number of data events N_{data} and SM events N_{SM} refer to the RoI.

A visualization of the observed data event is shown in Fig. 11.35b. The event contains little activity apart from the five clearly separated leptons. The Z boson hypothesis is applied to three combinations of same flavor opposite sign leptons with invariant masses of: $M(\mathrm{e}_1, \mathrm{e}_3) = 200$ GeV, $M(\mathrm{e}_2, \mathrm{e}_3) = 345$ GeV and $M(\mu_1, \mu_2) = 96$ GeV. In consequence, only the two muons are likely to originate from a Z boson decay.

No events are observed in this final state in the MUSiC analysis of data taken during 2011 at $\sqrt{s} = 7$ TeV [2], nor in the analysis of data taken during 2015 at $\sqrt{s} = 13$ TeV [308]. Thus, it is worth keeping an eye on this event class with increasing integrated luminosity values during future runs of the LHC. Until then, no definitive answer can be given on the nature of this observation. Although an occasional observation of such significance is within the expectation from statistical fluctuations of the SM, it could also represent the signature of a Z boson decaying into muons produced alongside a heavy resonance of an unknown particle decaying into three electrons. The invariant mass of all three electrons amounts to $M(\mathrm{e}_1, \mathrm{e}_2, \mathrm{e}_3) = 422$ GeV. A close follow-up on this event class should be made with additional data.

11.2.6 Conclusion of the Search Results

Overall, both the scans performed on the total event yield in each event class and those pursued for kinematic distributions show good agreement between data collected with single and double-lepton triggers and the MC simulation based SM prediction. Refined systematic uncertainties, especially considering the effect of unclustered energy, and a new approach of dealing with the occurrence of an insufficient amount of generated events improve the portrayal of properties observed with the CMS detector at 8 TeV. A small number of distributions show strong deviations from the modeled description, at most $3.6\,\sigma$ for the $\sum |\vec{p}_{\mathrm{T}}|$ distribution of the 2e+1γ+4jets+Njet event class. Many of the observed discrepancies are traced back to the limited correctness of the simulation based approach for determining object misidentification rates, especially photons. Interesting excesses are observed in final states with 3e+2jets+MET as well as 3e+2μ and should be monitored closely in future data-taking periods. In total, the number of observed deviations is in accordance with expected fluctuations of the SM within the assumed uncertainties.

11.3 Lepton-Charge Aware Analysis

Analogous to the procedure established in Sec. 11.2 the complete analysis is reevaluated based on event classes taking the net lepton charge of the event into consideration. The majority of SM processes lead to final states with opposite-sign lepton pairs. However, even such decays can result in signatures of non-zero net charge if a produced lepton does not pass the analysis selection or if its charge is not assigned correctly. Nevertheless, the resulting event classes present a valuable asset, since signs of specific BSM processes, e.g., given by scenarios of supersymmetry, would appear precisely in such final states of uncompensated electromagnetic charge. Difficulties already encountered in the scan disregarding lepton charges are increased when considering this additional event characteristic. The rate of leptons assigned with an incorrect charge, as well as the misidentification rate of other objects as leptons become more pronounced in this approach.

The classification separates the events into 484 exclusive, 501 inclusive and 473 jet-inclusive event classes, corresponding to an increase of about 45% with respect to the amount of event classes constructed without lepton charge awareness. An overview of the global distributions of $\tilde{p}$-values is given in Fig. 11.36 for exclusive and in Fig. 11.37 for (jet-)inclusive event classes. Once again more excesses than deficits are observed in all cases, and the ratio is even more extreme here than found in scan results already discussed in Sec. 11.2.3.

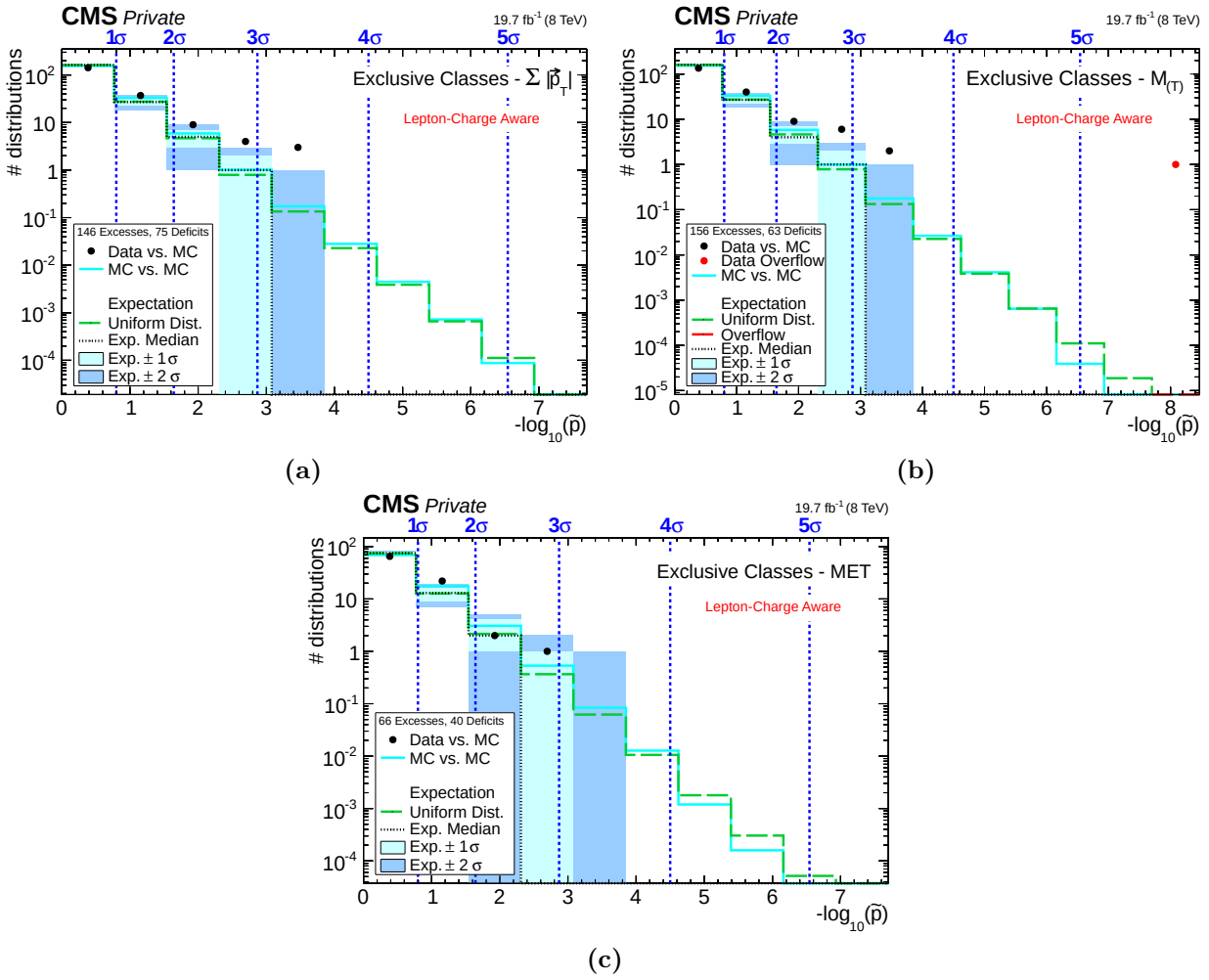


Figure 11.36: Distribution of $\tilde{p}$-values for $\sum |\vec{p}_{\mathrm{T}}|$, $M_{(\mathrm{T})}$ and MET distributions of exclusive event classes taking lepton charge into account. Black markers represent the measured data compared to the SM MC expectation. In the first bin a: 142 [b: 135, c: 65] distributions are observed in data, with a: $162 \pm 5(1\,\sigma)^{+10}_{-11}(2\,\sigma)$ [b: $160 \pm 5(1\,\sigma) \pm 10(2\,\sigma)$, c: $75^{+3}_{-4}(1\,\sigma)^{+6}_{-8}(2\,\sigma)$] expected from the SM.

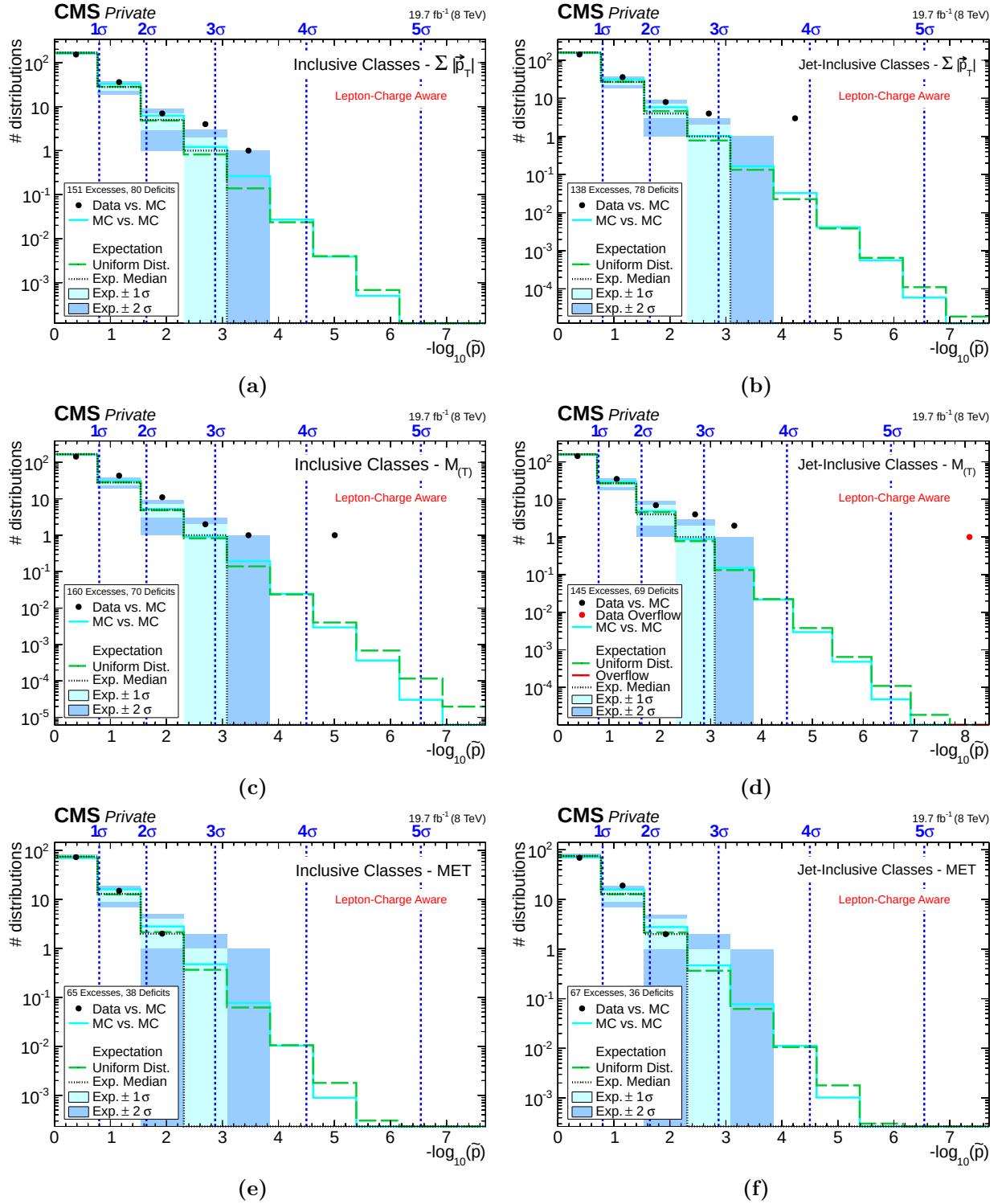


Figure 11.37: Distributions of $\tilde{p}$-values for kinematic distributions of inclusive (left side: a: $\sum|\vec{p}_{\mathrm{T}}|$, c: $M_{(\mathrm{T})}$, e: MET) and jet-inclusive (right side: b: $\sum|\vec{p}_{\mathrm{T}}|$, d: $M_{(\mathrm{T})}$, f: MET) event classes taking lepton charge into account. Black markers represent the measured data compared to the SM MC expectation. In the first bin a: 153 [b: 142, c: 144, d: 142, e: 73, f: 69] distributions are observed in data, with a: $167 \pm 5(1\,\sigma)^{+10}_{-11}(2\,\sigma)$ [b: $160 \pm 5(1\,\sigma) \pm 10(2\,\sigma)$, c: $168^{+5}_{-6}(1\,\sigma)^{+10}_{-11}(2\,\sigma)$, d: $159^{+5}_{-6}(1\,\sigma)^{+9}_{-11}(2\,\sigma)$, e: $75^{+3}_{-4}(1\,\sigma)^{+6}_{-8}(2\,\sigma)$, f: $75^{+3}_{-4}(1\,\sigma)^{+6}_{-8}(2\,\sigma)$] expected from the SM.

In kinematic distributions of $\sum|\vec{p}_{\mathrm{T}}|$ and $M_{(\mathrm{T})}$ stronger deviations are found than in the scan disregarding the lepton charge (see Sec. 11.2.3). In two cases the disagreement between data and SM MC prediction is so large that no LEE pseudo experiment within $2{\cdot}10^6$ rounds yields a smaller p-value than the one found when comparing data and SM expectation. For each event class type, the ten most significant deviations are listed in Tab. C.7, Tab. C.8 and Tab. C.9, respectively. The two most significant deviations are summarized in Tab. 11.14, of which many are also observed to be among the most significant in the scan disregarding lepton charge and are given in blue.

	Most Significant		Second Most Significant	
	Event Class	$\tilde{p}$	Event Class	$\tilde{p}$
$\sum\|\vec{p}_{\mathrm{T}}\|$	**2e+≥6jets[2Q]**	$2.2{\cdot}10^{-4}$	**2μ [2Q]**	$3.8{\cdot}10^{-4}$
	2e+1γ+4jets+X[0Q]	$7.3{\cdot}10^{-4}$	**2e+1μ+2jets+X[3Q]**	0.0034
	2μ+1jet+Njet[2Q]	$2.7{\cdot}10^{-5}$	**2μ+Njet[2Q]**	$1.1{\cdot}10^{-4}$
$M_{(\mathrm{T})}$	**2μ [2Q]**	$< 5{\cdot}10^{-7}$	**2e+1γ+4jets[0Q]**	$1.9{\cdot}10^{-4}$
	2μ+X[2Q]	$7.0{\cdot}10^{-6}$	**2e+1γ+4jets+X[0Q]**	$7.5{\cdot}10^{-4}$
	2μ+Njet[2Q]	$< 5{\cdot}10^{-7}$	**2e+1γ+4jets+Njet[0Q]**	$1.5{\cdot}10^{-4}$
MET	**1e+1γ+MET**	0.0039	**3e+2jets+MET [1Q]**	0.0082
	3e+2jets+MET+X[1Q]	0.011	**2μ+1γ+1jet+MET+X[0Q]**	0.024
	3e+2jets+MET+Njet[1Q]	0.0083	**2μ+1γ+1jet+MET+Njet[0Q]**	0.024

Table 11.14: Overview of the two most significant event classes from each scan type taking the net charge of all leptons in the event into consideration. Deviations also among the two most significant event class distributions without lepton charge awareness, as discussed in Sec. 11.2.5, are typeset in blue. Those shown in other colors are only so significant in the scan discussed here.

The overall most significant deviations are found in event classes containing two muons of the same charge. Here, the observed disagreement between data and SM MC is so large, that more than $2{\cdot}10^6$ pseudo-experiments would have to be generated to determine the LEE correction. As shown in Fig. 11.38a, a broad excess is found around the peak of the mass distribution of 2μ[2Q] which is dominated by multi-boson processes. In the RoI, 355 events are observed in data, where only 94 ± 30 events are expected from the SM simulation. To get to the bottom of this striking deviation, the corresponding event class with opposite-sign muons is shown in Fig. 11.38b, where a clear peak from the DY process is visible at 90 GeV. No large deviation is observed in this highly populated event class. In the same-sign case the peak is situated below 90 GeV, indicating that the majority of events contributing to this final state do not originate from DY events with a charge misidentified muon. The dominant contribution in simulation is given by the WZ process, where one of the muons from the Z boson decay does not pass the full set of analysis requirements. Moreover, the W+jets process with MET under 50 GeV and a muon given by a jet is found in this final state, but not nearly as frequently as the mentioned di-boson process due to the restriction on MET and the low rate of muons originating from jets. Clearly visible by numerous empty intermediate bins, the amount of generated events given by the W+jets data sets is low in 2μ[2Q], in which dedicated data sets representing processes of W bosons accompanied by two bottom quark tagged jets are also included. A noticeably larger contribution of W+jets is present in the 2μ+MET[2Q] event class. This final state shows a better agreement between data and simulation yielding a significance of $\tilde{p} = 0.0099$ given around the peak region of the M_T distribution.

At the DY mass resonance a muon charge misidentification rate of around 0.001% is found in cosmic-ray muon data [295]. Based on this value the roughly $3{\cdot}10^6$ events populating the peak region of 2μ[0Q] are expected to deliver around 30 events in the same-sign scenario, yet in total

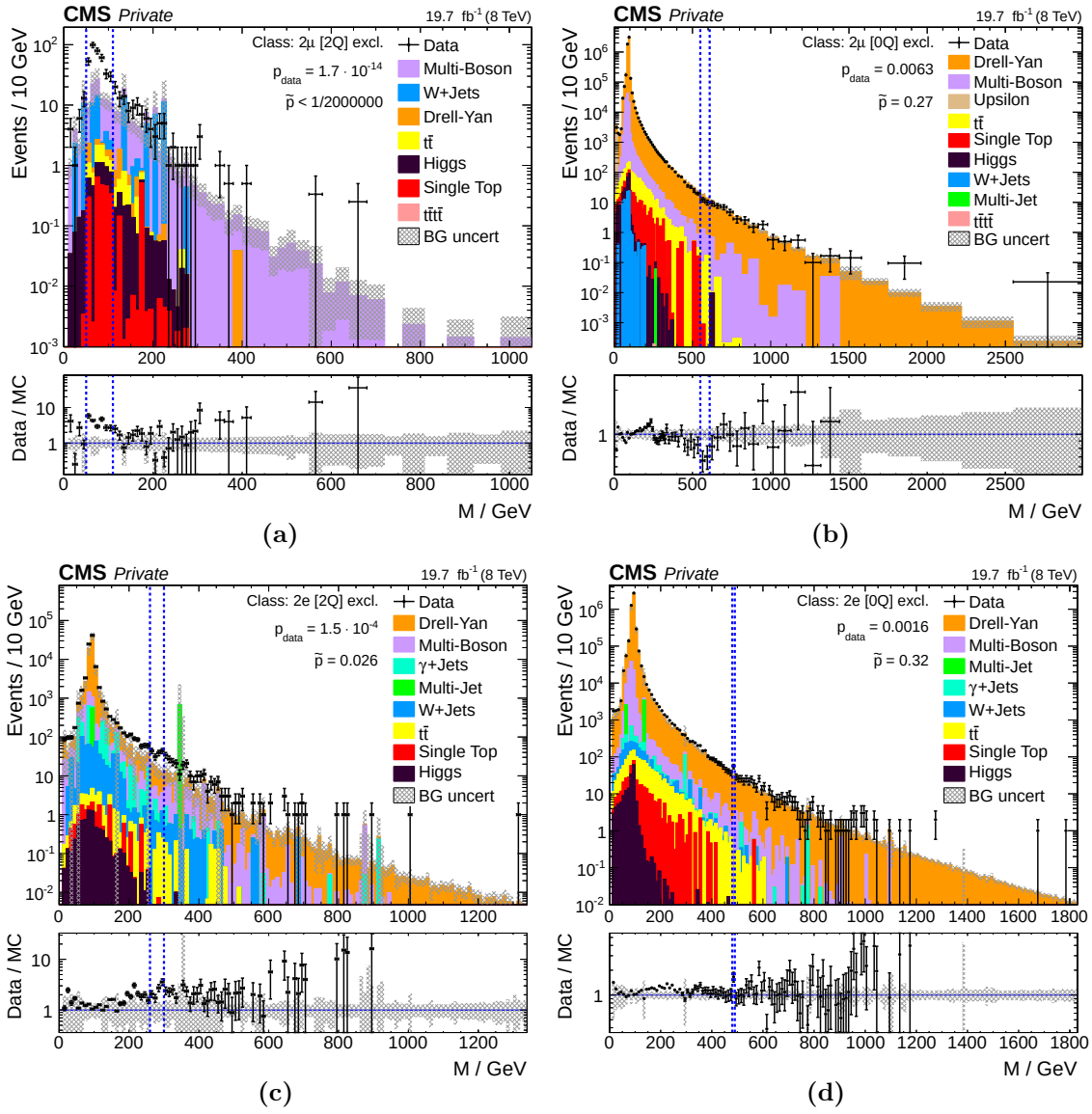


Figure 11.38: The most significant deviation observed when considering an event's net lepton charge is represented by 2μ[2Q], shown in Fig. 11.38a. Moreover, the 2μ[0Q] event class with two oppositely charged muons is given on the right. In the lower row, the di-electron case 2e[2Q] and 2e[0Q] is presented.

only around six DY events are found. A possible explanation is given by the method of determining the muon transverse momentum, which based on the curvature of the muon track throughout the detector. If the charge is not assigned correctly, the curvature and thus the muon trajectory is also likely to be measured very imprecisely. Such muons are prone to fail track selection criteria (see Sec. 9.2.2) and are removed from the analysis, resulting in less DY events than initially expected.

Furthermore, a glance at the corresponding event classes with electrons is given in Fig. 11.38c for the same-sign and in Fig. 11.38d for the opposite-sign electron charge case. A strong difference in rate, especially for DY, is seen between the 2μ[2Q] (Fig. 11.38a) and 2e[2Q] (Fig. 11.38c) event classes. The muon charge misidentification rate is much smaller than that of electrons (1.5% [267]), yet the main effect is that the reconstructed electron four-vector is not influenced so strongly by its curvature. Even though, as given in Sec. 10.4, the systematic assigned to account for the difference of charge misidentification in data and modeling in MC is smaller for electrons (20%) than for

muons (50%), the actual rate of charge misidentification leads to larger uncertainty contributions in distributions containing electrons. In total, the deviation observed for two electrons of the same electromagnetic charge is less severe and located more towards the tail of the distribution. The peak region is modeled well. Contributions from γ+jets and QCD multi-boson processes are also involved, given by a photon misidentified as an electron and electrons originating from jets.

A search for supersymmetry in events of same-sign di-leptons and at least two jets [309, 310] reports a number of observed events comparable to 2μ+2jets+Njet[2Q] and 2e+2jets+Njet[2Q]. In the dedicated analysis a data-driven method is used to estimate the contribution of muons originating from jets, representing the dominant source of events. In the final selection the data-driven approach yields 47 events, providing a good description of the observed data, whereas the pure MC simulation yields only 33 events. The MC based approach, relying on object misidentification rates purely from simulation, as chosen by MUSiC, underestimates contributions of muons given by jets in this final state. A similar effect was also observed in the MUSiC analysis of 2011 data [2].

Two more event classes are given in Tab. 11.14, 2e+≥6jets[2Q] and 2e+1μ+2jets+X[3Q], both exhibiting only a few data events with an SM MC expectation below one event. The observed deviations yield $\tilde{p}$-values of $2.2 \cdot 10^{-4} \approx 3.5\,\sigma$ and $0.0034 \approx 2.7\,\sigma$. Corresponding event classes with the alternative lepton charge combination show substantially more events with overall good agreement and only rather insignificant deviations. All four distributions are shown in Fig. 11.39.

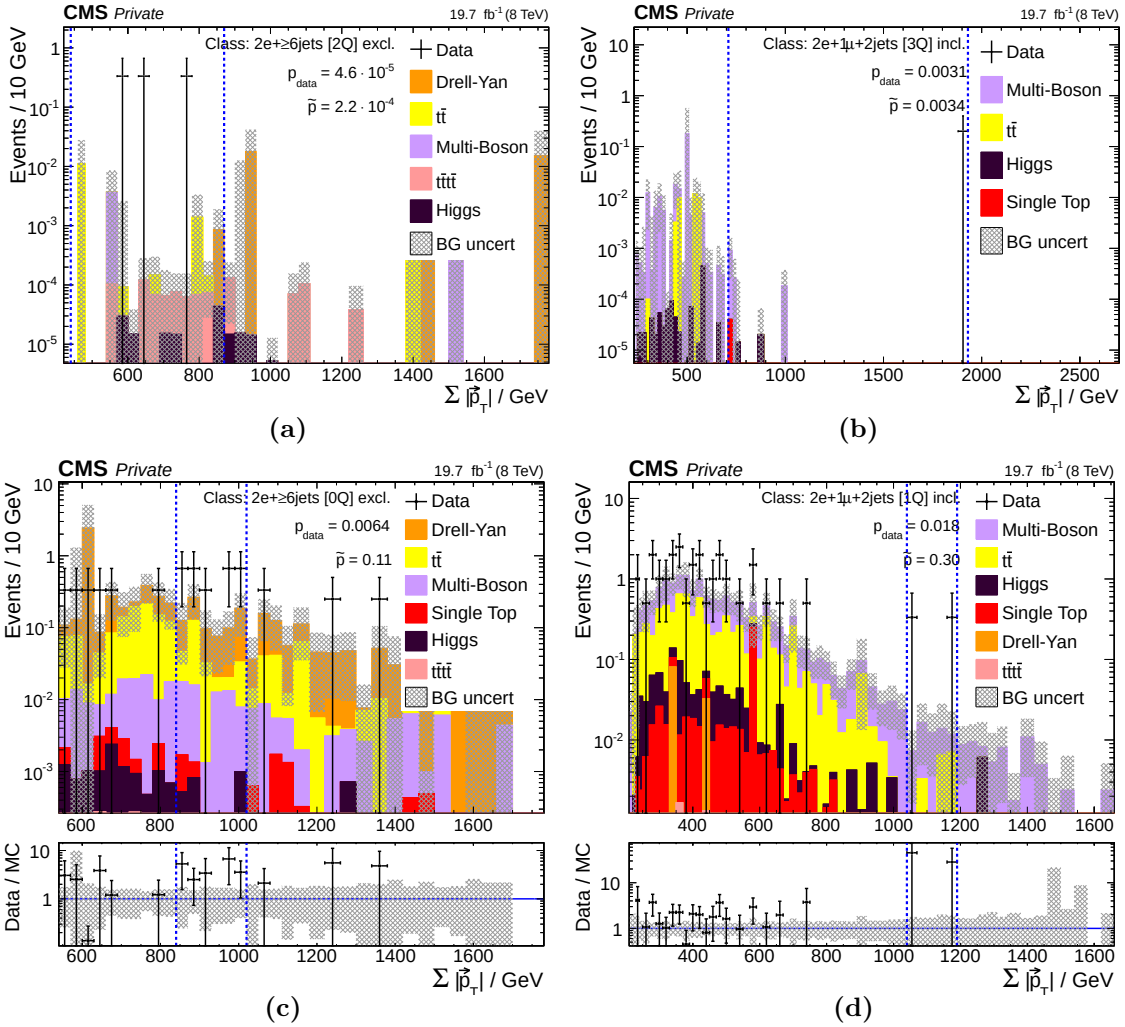


Figure 11.39: Two of the most significant event classes considering the net lepton charge of the event, 2e+≥6jets[2Q] in Fig. 11.39a and 2e+1μ+2jets+X[3Q] in Fig. 11.39b. For both cases the event class distribution with the alternative lepton charge combination is given in the row below.

A similar composition of SM processes can be found in both the same-sign as in the opposite-sign cases. Within the scan disregarding lepton charges, both event classes are insignificant, 2e+≥6jets is the 20$^{\text{th}}$ most significant $\sum|\vec{p}_{\text{T}}|$ distribution of an exclusive event class and 2e+1μ+2jets+X is the 37$^{\text{th}}$ most significant $\sum|\vec{p}_{\text{T}}|$ distribution within the inclusive event classes. In both cases the RoI is in the same position as in Fig. 11.39c and Fig. 11.39d, resulting in very similar significance values. The single data event observed in 2e+1μ+2jets+X[3Q] is visualized in Fig. 11.40, showing a clean reconstruction of all five identified objects which are all well above the respective analysis level transverse momentum thresholds. No obvious explanation can be given for the strong significances found in the same-sign case. It remains to be seen if these excesses persist in a larger amount of data expected to be delivered by the 2016 LHC run.

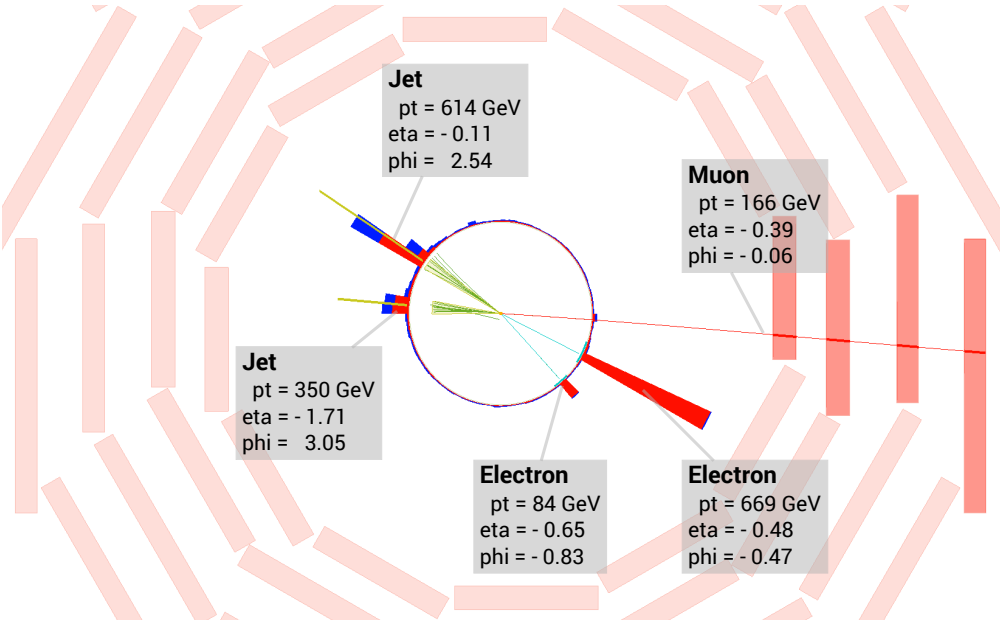


Figure 11.40: 204563:261:337432057 (run:lumi section:event) – single-muon trigger stream
Visualization of the data event contributing to the 2e+1μ+2jets+X[3Q] event class in the ρ-ϕ view of the CMS detector. Red / blue towers indicate deposits in the ECal / HCal proportional to the collected energy. All objects are separated well from each other. The three leptons are identified with positive electromagnetic charge. MUSiC kinematic variables of this event are: $\sum|\vec{p}_{\text{T}}| = (1883 \pm 41)$ GeV; $M_{(\text{T})} = (2206 \pm 52)$ GeV; MET $= (24 \pm 29)$ GeV (disregarded in analysis). The stated rough uncertainties on kinematic event variables are determined using resolutions given in Sec. 7.2.3.

Chapter 12
Sensitivity Studies

In order to warrant credibility to the MUSiC analysis claims concerning the comparison of measured data and SM simulation, it must be shown that the algorithm will not find a significant deviation if no new phenomena are actually present. Furthermore, it must be validated that MUSiC can indeed locate signal contributions within its search regime. Three separate studies are performed to address these prerequisites. Initially, as summarized in Sec. 12.1, pseudo-data is generated under the SM-only hypothesis and compared with the SM MC expectation itself to confirm that the MUSiC search method does not reveal very small p-values if only fluctuations of the SM are observed in data. In the second study, Sec. 12.2, a BSM signal is injected on top of the SM, indicating how evidence of new phenomena would manifest itself in the analysis. Finally, two known SM processes are omitted from the SM simulation and a comparison to the measured data is performed to evaluate the potential of rediscovering them with MUSiC.

For sensitivity studies addressed in Sec. 12.1 and 12.2, statistical fluctuations and the systematic uncertainties of the simulation must be accounted for. A total of $N_{\text{pseudo}} = 100$ pseudo-data sets, whose respective nature will be explained for each individual case, are determined for each event class and kinematic variable to insure the outcome is statistically stable. For each of the N_{pseudo} experiments the most significant p-value is calculated between signal induced pseudo-data and SM MC expectation. A median p-value $p_{1/2}^{\text{pseudo}}$ is determined from all pseudo-experiment p-values and the round corresponding to the median p-value is chosen for representative purposes. To account for the LEE (see Sec. 7.3.4), the SM-only hypothesis is used to create up to $1{\cdot}10^5$ pseudo-experiments in each event class distribution. Following the runtime reduction procedure established in Sec. 7.3.6, generation of pseudo-data for the LEE correction is terminated effectively based on the significance of the p-value. The cutoff is set on twice the number of rounds needed to generate 100 SM-only pseudo-experiments yielding a more significant p-value than the smallest p-value $p_{\text{min}}^{\text{pseudo}}$ found in the N_{pseudo} signal induced pseudo-data rounds. Here it is vital to base the number of LEE correction pseudo-experiments on $p_{\text{min}}^{\text{pseudo}}$, the most extreme signal-induced pseudo-data set, and not on $p_{1/2}^{\text{pseudo}}$ to ensure a sufficiently stable result. Furthermore, statistical outliers in the signal pseudo-data are removed. If the smallest pseudo-data p-value is at least three orders of magnitude smaller than the next smallest p-value, the corresponding pseudo-experiment is removed. Although such extreme cases are statistically valid, run-time limitations do not allow increasing the number of (SM-only hypothesis) rounds needed to incorporate these small p-values. Such outliers are very seldom at under 0.1% per scan of event class type and kinematic variable.

The studies given in Sec. 12.3 do not require additional signal induced pseudo-data, but rely on the measured data already analyzed in Ch. 11. Nevertheless, the LEE correction is taken into consideration with up to $2{\cdot}10^6$ pseudo-experiments carried out under the SM-only hypothesis.

12.1 False Discovery Rate

Given that the observed data behave as expected from the SM prediction, the MUSiC analysis should not encounter extreme deviations of high significance. Along these lines, the algorithm is verified by generating $N_{\text{pseudo}} = 100$ sets of pseudo-data according to the MC based SM-only hypothesis, referred to here as signal. These are compared to the SM MC prediction itself, calculating a separate $\tilde{p}$-value for each pseudo-experiment.

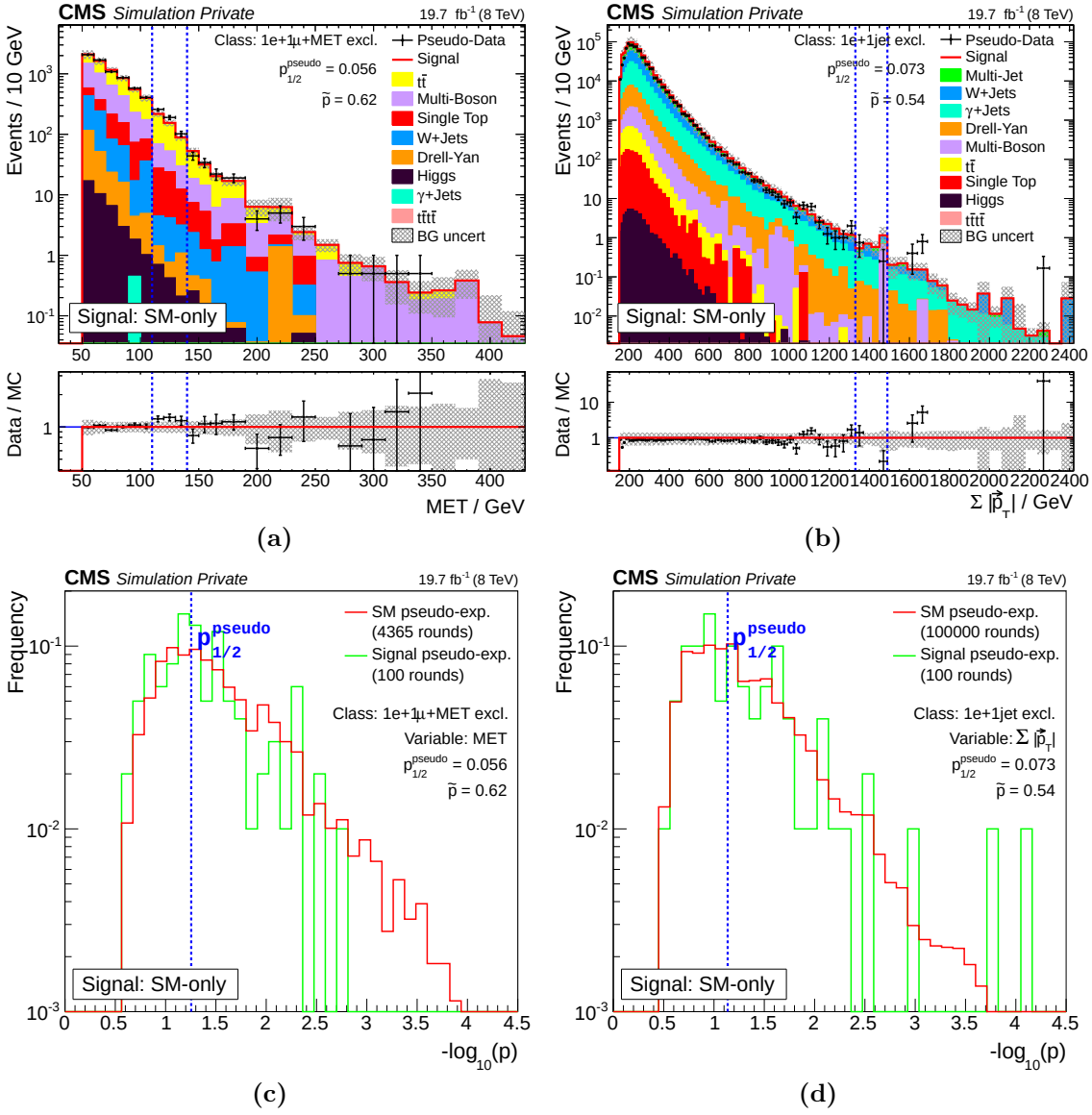


Figure 12.1: The upper figures show kinematic distributions for two event classes with the median result of 100 SM-only signal rounds. The signal is given by a red line and black markers represent pseudo-data. The lower figures show the respective distribution of p-values from the SM-only signal rounds (green line) and up to 10^5 pseudo-experiments created for the LEE correction (red line). Histograms are normalized to the corresponding number of pseudo-experiments (given in legend). In both cases the p-value histograms overlap fully, proving the signal events to be in accordance with the SM MC prediction.

Two example event class distributions are shown in Fig. 12.1, along with the corresponding distribution of p-values from signal rounds and from the LEE calculation. In Fig. 12.1a and

Fig. 12.1b the signal is shown by a red line whereas black markers represent pseudo-data generated under the SM-only hypothesis modeled with MC simulation. For each event class the chosen distribution corresponds to the signal round yielding the median p-value $p_{1/2}^{\text{pseudo}}$ over all rounds. As shown in Fig. 12.1c and Fig. 12.1d the median p-value is calculated from the values shown in green. The number of pseudo-experiments conducted for the LEE correction, shown in red, is based on the signal round with the strongest deviation. Since a smaller $p_{\text{min}}^{\text{pseudo}}$ was obtained for the SM-only signal in Fig. 12.1d more LEE correction pseudo-experiments are generated than for the case in Fig. 12.1c (10^5 vs. 4365 rounds). Both examples illustrate that the SM-only signal pseudo-experiments show the same behavior in terms of significance as the LEE correction pseudo-experiments. Consequently, $\tilde{p}$-values calculated for each event class distribution do not indicate a significant deviation.

The global overview of results is shown in Fig. 12.2 for all three kinematic distributions of exclusive event classes. Corresponding distributions of $\tilde{p}$-value for (jet-)inclusive event classes can be found in App. D.1. In all cases the pseudo-data do not deviate more strongly from the SM MC description than expected from statistical fluctuations within the SM. On average, far less than one event class distribution shows a deviation of more than $3\,\sigma$, fully in accordance with predictions from the SM modeling. As can be concluded from this study, the MUSiC analysis will not claim significant deviations if the data behave as expected from the SM simulation.

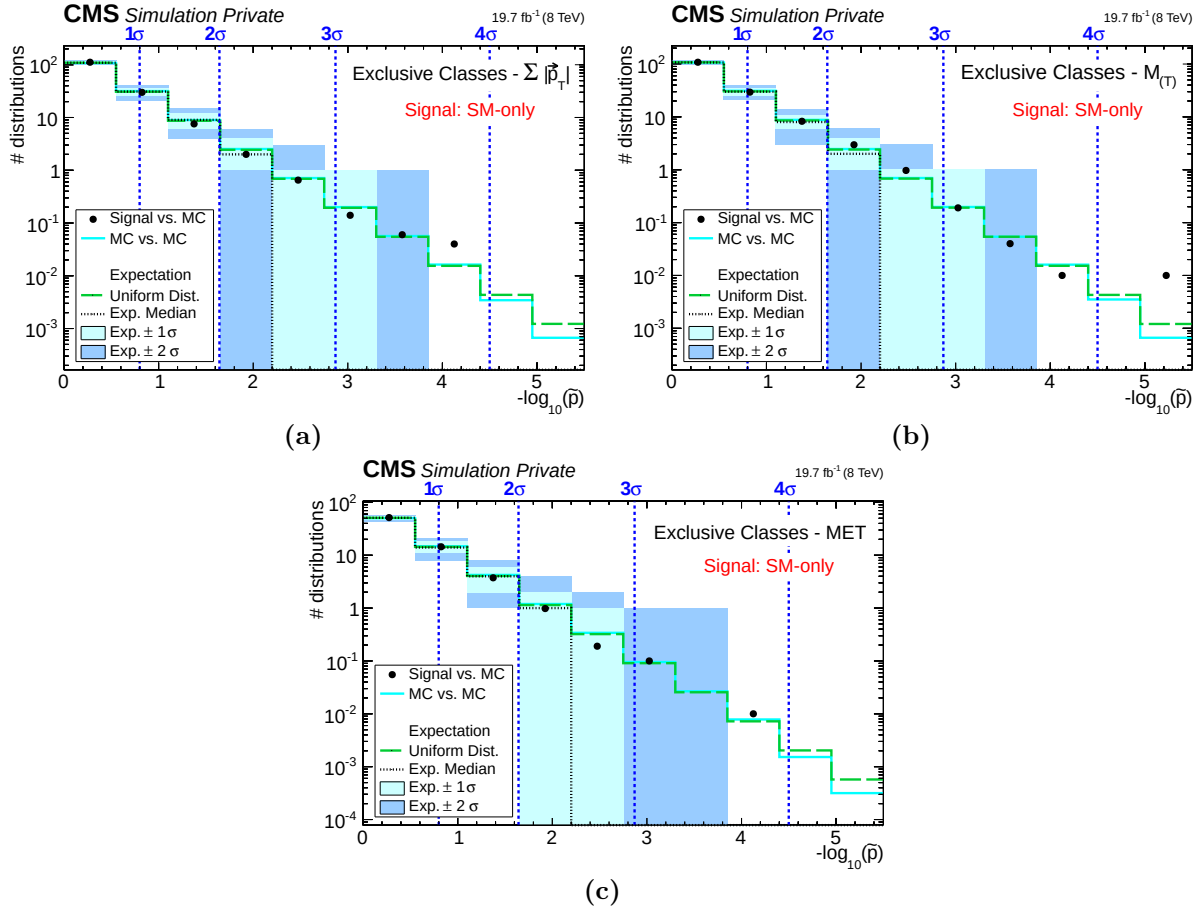


Figure 12.2: Distributions of $\tilde{p}$-values for kinematic distributions (a: $\sum |\vec{p}_{\text{T}}|$, b: $M_{(\text{T})}$, c: MET) of exclusive event classes. Black markers represent the averaged results of pseudo-data from 100 SM-only signal rounds in each event class. In the first bin a: 111.6 [b: 108.2, c: 51.5] distributions are observed in pseudo-data ("Signal vs. MC"), with a: $109^{+6}_{-5}(1\,\sigma) \pm 11(2\,\sigma)$ [b: $107 \pm 5(1\,\sigma) \pm 11(2\,\sigma)$, c: $51 \pm 4(1\,\sigma)^{+7}_{-8}(2\,\sigma)$] expected from the SM. Corresponding results for (jet-)inclusive event classes are given in Fig. D.1.

12.2 Sensitivity to Beyond the Standard Model Signals

In a second study, the MUSiC algorithm is validated by introducing a dedicated BSM signal, from which pseudo-data are generated. Once again, 100 pseudo-experiments per event class distribution are compared to the SM MC expectation. Such a study provides a proof of principle that the MUSiC analysis would be sensitive to a given BSM scenario. The chosen signal is expected to introduce a strong and localized excess of events in individual final states. Under assumptions of the SSM, a heavy vector boson W' is produced and its leptonic decays are reconstructed as a single isolated and high-energetic muon or electron along with substantial MET. Additionally, the events my contain a number of jets resulting from ISR as well as PU. Details of the underlying theoretical model and the used data sets are given in Sec. 8.2.2. The M_T distributions of 1e+MET+Njet and 1μ+MET+Njet are comparable to the search region of a dedicated CMS analysis investigating W' like signals [228]. Since strong deviations found in several event classes lead to a long processing time, and similar results are expected from all event class types and kinematic variables, a focus is set on the mass distribution of jet-inclusive event classes.

Four different W' boson mass scenarios, $M_{W'} = 2.0,\ 2.5,\ 3.0$ and 3.5 TeV, are evaluated separately and resulting $\tilde{p}$ distributions are shown in Fig. 12.3. In each case few event classes exhibit a p-value so small that no pseudo-experiment from the LEE correction obtains an even stronger significance. These event class distributions contribute to the overflow-bin, given in red, and on average two are given for both $M_{W'} = 2.0$ and 2.5 TeV. Within the SM, under 10^{-3} event classes are expected to deviate so strongly. Additionally, all W' scenarios induce numerous deviations above $3\,\sigma$. Large W' boson masses lead to an increasingly broader and less pronounced signal with smaller cross section (see Tab. B.2 for details), becoming more compatible with the SM prediction. Correspondingly, fewer strongly deviating event classes are found for high masses, declining below one observed distribution for 3.5 TeV, and are less obvious for the MUSiC analysis to uncover.

As listed in Tab. 12.1, giving an overview of the ten most significant event class distributions, the 1e+1jet+Njet and 1μ+1jet+Njet event classes show the strongest deviations for all considered W' boson masses. In the two scenarios with the lightest masses all signal induced pseudo-experiments in single lepton and MET event classes yield a p-value so small, that no LEE corrected value could be determined within the precision of pseudo-experiment generation. Such rounds are summarized by the **O**verflow **F**raction (OF) which decreases with increasing W' boson masses. Events classes containing jets in addition to a single lepton and MET also result in small $\tilde{p}$-values, but not nearly as significant as in the cases without jets. Other event topologies are barely influenced by the injected signal. Moderate $\tilde{p}$-values are observed because of fluctuations within the uncertainties of the SM simulation and are comparable to those obtained from the study to determine the false discovery rate, as summarized in Sec. 12.1. It should be noted, that some of the p-values stated here represent extreme cases reaching values below $\mathcal{O}(10^{-100})$ and should not be taken by face value. Since the p-value calculation is performed numerically its associated precision prevents such small values to be evaluated reliably. One can only take these deviations to be exceedingly significant.

Respectively, the two most significant event classes are discussed for each W' boson mass in Fig. 12.4 through Fig. 12.7. The Jacobian peak around the associated mass of the signal, shown in red, becomes clearly visible in the transverse mass distribution of 1e+1jet+Njet and 1μ+1jet+Njet. Especially in the distribution's tail, bins of the event class distribution for the muon final state are wider than in the electron case, reflecting the corresponding lepton momentum resolution in the CMS detector. As a result, the event classes containing electron are more sensitive to the introduced BSM signal. For $M_{W'} = 2.0$ and 2.5 TeV pseudo-data are distributed well throughout the signal's peak, leading to event rates around 100 times higher than expected from SM processes. Correspondingly, p-values of pseudo-data generated under consideration of the W' scenario (green

	Event Class	$p_{1/2}^{\text{pseudo}}$	$\tilde{p}$	RoI / GeV	N_{Pseudo}	$N_{\text{SM}} \pm \sigma_{\text{SM}}$	OF / %
$M_{W'} = 2.0$ TeV	1e+MET+Njet	$2.30 \cdot 10^{-214}$	$<10^{-5}$	1400 - 2170	210	4.07±0.52	100
	1μ+MET+Njet	$3.23 \cdot 10^{-192}$	$<10^{-5}$	1350 - 2890	210	6.15±0.67	100
	1e+1jet+MET+Njet	$7.38 \cdot 10^{-5}$	0.0051	1970 - 2350	34	10.4±3.8	12
	1μ+1jet+MET+Njet	0.0048	0.063	1850 - 3140	83	43±13	1
	2μ+5jets+MET+Njet	0.013	0.42	870 - 1460	69	43.8±8.1	0
	1e+1γ+MET+Njet	0.056	0.42	770 - 890	1.0	0.041±0.049	0
	1e+1μ+5jets+Njet	0.0099	0.42	980 - 1250	27	14.4±2.7	0
	2e+MET+Njet	0.016	0.42	1670 - 1790	1.0	0.0164±0.0030	0
	2e+5jets+Njet	0.019	0.43	1360 - 1480	1.0	11.0±4.3	0
	1e+2μ+2jets+Njet	0.042	0.43	540 - 1000	1.0	12.6±8.4	0
$M_{W'} = 2.5$ TeV	1e+MET+Njet	$3.66 \cdot 10^{-43}$	$<10^{-5}$	1970 - 2770	29	0.303±0.058	100
	1μ+MET+Njet	$1.85 \cdot 10^{-28}$	$<10^{-5}$	1660 - 3780	34	1.82±0.24	100
	1e+1jet+MET+Njet	0.011	0.21	2650 - 2970	3.0	0.34±0.23	0
	2e+MET+Njet	0.0089	0.30	670 - 770	18	7.7±2.3	0
	2μ+3jets+Njet	0.032	0.37	2140 - 2370	33	16.6±7.0	0
	2e+1μ+3jets+MET+Njet	0.048	0.37	790 - 1140	4.0	1.28±0.33	0
	2μ+1jet+MET+Njet	0.036	0.40	470 - 580	790	1340±300	0
	2μ+MET+Njet	0.016	0.40	10 - 40	960	490±220	0
	1μ+4jets+Njet	0.033	0.41	1510 - 1710	600	453±77	0
	1μ+1jet+MET+Njet	0.050	0.41	1770 - 2490	87	56±16	0
$M_{W'} = 3.0$ TeV	1e+MET+Njet	$6.12 \cdot 10^{-10}$	$<10^{-5}$	2530 - 3370	5.0	0.026±0.015	65
	1μ+MET+Njet	$7.73 \cdot 10^{-5}$	0.0034	2580 - 3780	3.0	0.075±0.018	23
	1e+1jet+MET+Njet	0.012	0.23	80 - 140	180	343±71	0
	3μ+2jets+Njet	0.026	0.40	860 - 1130	4.0	1.01±0.33	0
	2μ+3jets+Njet	0.036	0.41	2820 - 3270	8.0	2.8±1.5	0
	1e+2μ+2jets+MET+Njet	0.026	0.41	370 - 800	4.0	13.5±3.6	0
	2e+2jets+Njet	0.031	0.42	110 - 200	33	79±23	0
	1μ+1γ+2jets+MET+Njet	0.047	0.42	280 - 340	2.0	15.5±9.2	0
	3e+4jets+Njet	0.13	0.42	1130 - 1520	1.0	0.140±0.063	0
	2e+1μ+3jets+Njet	0.073	0.42	370 - 450	1.0	0.063±0.056	0
$M_{W'} = 3.5$ TeV	1e+MET+Njet	0.0017	0.12	690 - 770	53	86.7±7.8	4
	1μ+MET+Njet	0.01	0.23	1660 - 2580	6.0	1.75±0.23	0
	1e+1γ+Njet	0.014	0.31	20 - 140	5.0	130±130	0
	2μ+2jets+MET+Njet	0.028	0.40	1470 - 1680	4.0	19.0±7.2	0
	2μ+MET+Njet	0.016	0.41	340 - 400	310	441±60	0
	1e+1jet+MET+Njet	0.028	0.41	100 - 160	1200	640±300	0
	3μ+3jets+MET+Njet	0.057	0.41	880 - 990	2.0	0.37±0.13	0
	2μ+1γ+2jets+Njet	0.18	0.41	420 - 1330	5.0	2.7±1.1	0
	3e+3jets+Njet	0.099	0.41	370 - 470	1.0	0.101±0.054	0
	1e+1μ+1γ+Njet	0.24	0.42	70 - 220	4.0	2.2±1.2	0

Table 12.1: Overview of the ten most significant jet-inclusive event classes for M_T distributions of SM MC compared to pseudo-data created under four different W' signal scenarios, considering masses of $M_{W'} = 2.0,\ 2.5,\ 3.0$ and 3.5 TeV. Here, $p_{1/2}^{\text{pseudo}}$ is the median p-value of 100 signal pseudo-experiments and $\tilde{p}$ is the corresponding LEE corrected significance. The bin region of the RoI is given, as well as the number of pseudo-data events N_{Pseudo} and SM MC events with a $\tilde{p}$-value $< 10^{-5}$ is given by the OF.

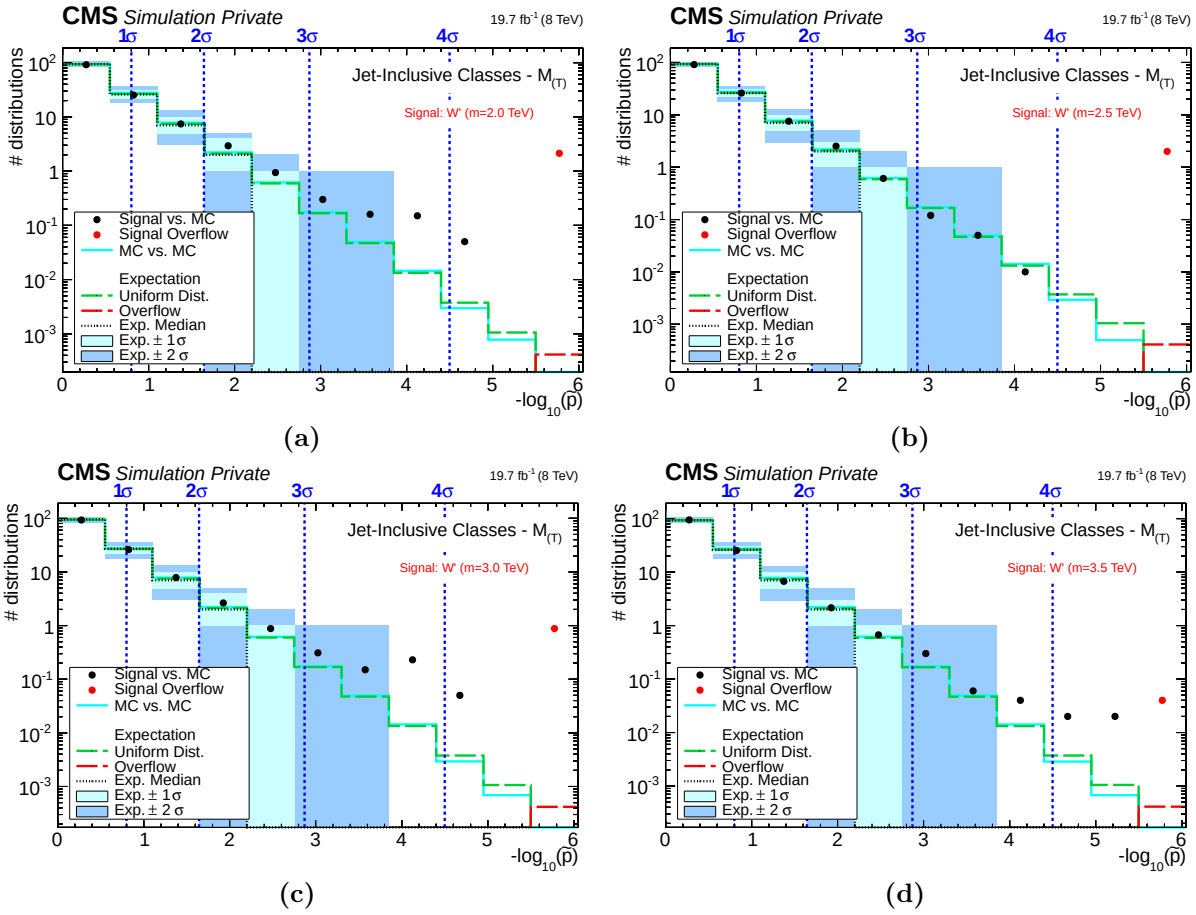


Figure 12.3: Distributions of $\tilde{p}$-values for the M_T distribution of jet-inclusive event classes. Black or red markers represent the averaged results of pseudo-data from 100 signal rounds in each event class. Four different W' boson masses ($M_{W'} = 2.0, 2.5, 3.0, 3.5$ TeV) are considered. In the first bin a: 91.7 [b: 91.1, c: 93.0, d: 94.6] distributions are observed in pseudo-data ("Signal vs. MC"), with a: $94 \pm 5(1\,\sigma) \pm 10(2\,\sigma)$ [b: $93 \pm 5(1\,\sigma) \pm 10(2\,\sigma)$, c: $95 \pm 5(1\,\sigma)^{+10}_{-11}(2\,\sigma)$, d: $93 \pm 5(1\,\sigma) \pm 10(2\,\sigma)$] expected from the SM. All figures are published in similar form in [5].

line in Fig. 12.4c and Fig. 12.4d, or similar figures) are several orders of magnitude larger than ones from pseudo-data under the SM-only hypothesis. With increasing W' boson masses the distribution of signal induced p-values approaches that of the SM-only induced p-values, finally showing a strong overlap at $M_{W'} = 3.5$ TeV as depicted in Fig. 12.7c and Fig. 12.7d. The calculated median p-value $p_{1/2}^{\text{pseudo}}$ is compatible with those obtained from the pure SM pseudo-experiments.

An initially unsettling behavior can be observed in the pseudo-data given in especially Fig. 12.4a and Fig. 12.6a at low values of M_T. As can be seen most directly in the ratio plot, the pseudo-data deviates very strongly from the signal around which it is generated. The reason for these extreme pseudo-data values is given by the large systematic uncertainties on the SM simulation in the afflicted region of the distribution. For every round, the pseudo-data are generated from the SM expectation within its uncertainties, for which random values are determined by which each systematic is shifted. Statistically, some cases are expected to appear where the pseudo-data lie in the outskirts of the systematic uncertainty band. The round shown in the distribution is the one corresponding to the median p-value. In the given event class one specific systematic uncertainty, stemming from misidentified objects, is rather large in the region around the single electron trigger turn on. A strong shift of the uncertainty on the electron misidentification rate data-MC scale factor

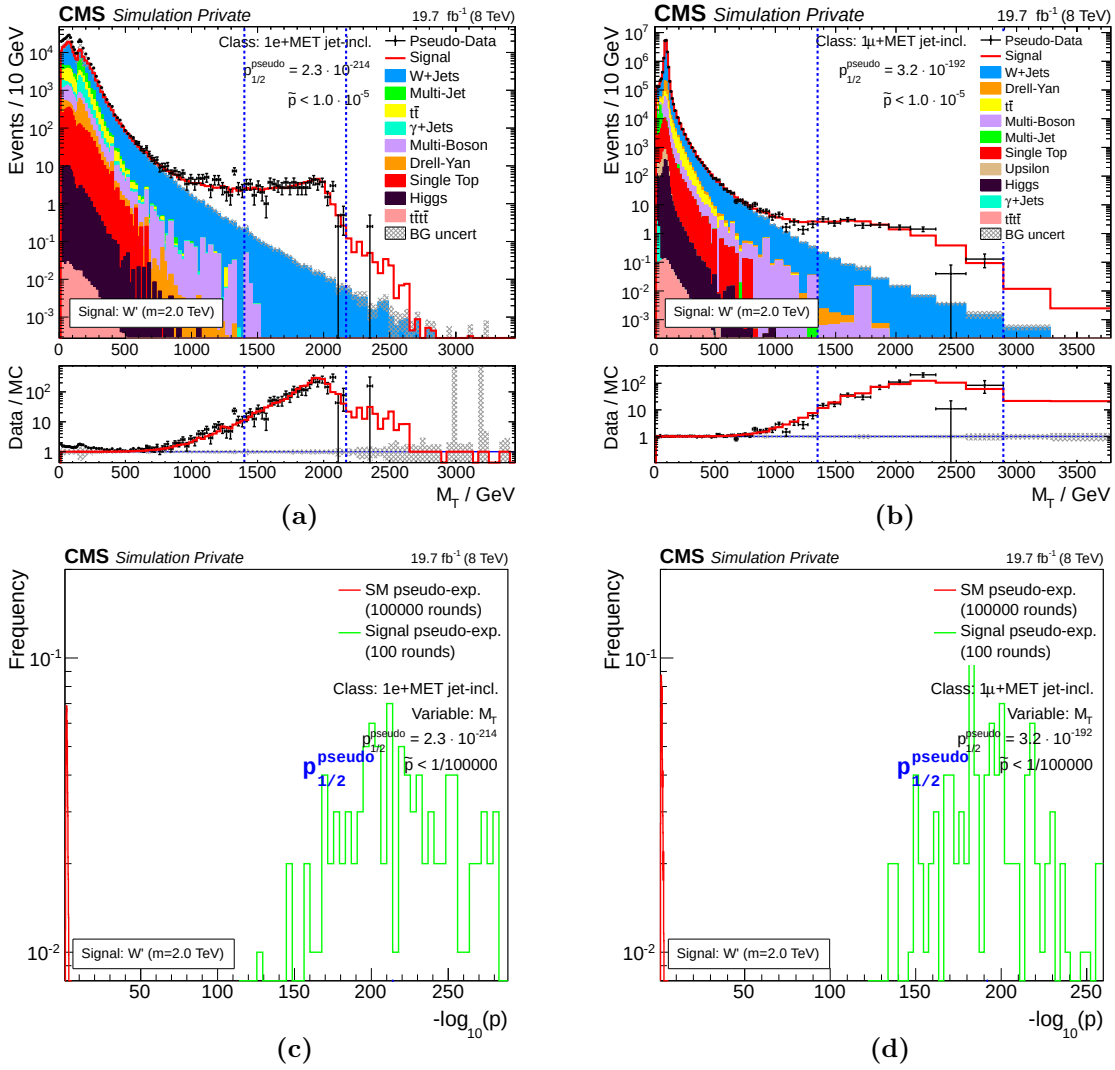


Figure 12.4: M_T distribution (a, b) and corresponding p-value distribution (c, d) for the 1e+1jet+Njet and 1μ+1jet+Njet jet-inclusive event classes. In the upper two figures, black markers represent the median result of pseudo-data from 100 signal rounds assuming a W' scenario with $M_{W'} = 2.0$ TeV. The signal itself is shown by a red line. In the lower two figures, pseudo-experiments from the LEE calculation are given in red, those generated under the signal hypothesis are shown in green and the median p-value is given by a blue dashed line. Histograms are normalized to the number of pseudo-experiments (given in legend) and a separate binning is used to accent features of each histogram individually.

can have a large influence on the pseudo-data generated in this region. Below 150 GeV this effect becomes less prominent, as it is less likely to obtain both electrons required by the double electron trigger from misidentification. Nevertheless, in both cases the RoI is located in the tail of the distribution and not in the those of extreme pseudo-data, since the p-value also accounts for the total uncertainty assigned to the region. For example, the bin from 150 GeV to 160 GeV in Fig. 12.6a, which shows the strongest difference between pseudo-data and SM expectation in the regime below 500 GeV, yields a p-value of only $p_{\text{pseudo}} = 0.0019 \sim 2.9\,\sigma$ ($N_{\text{data}} = 237$, $N_{\text{SM}} = 11994 \pm 5693$), much less significant than the one given in the RoI around 3.0 TeV.

Finally, a brief comparison is made to results obtained by the dedicated analysis covering precisely these signatures. As given in [228], the most energetic event in 19.7 fb^{-1} of measured CMS data at

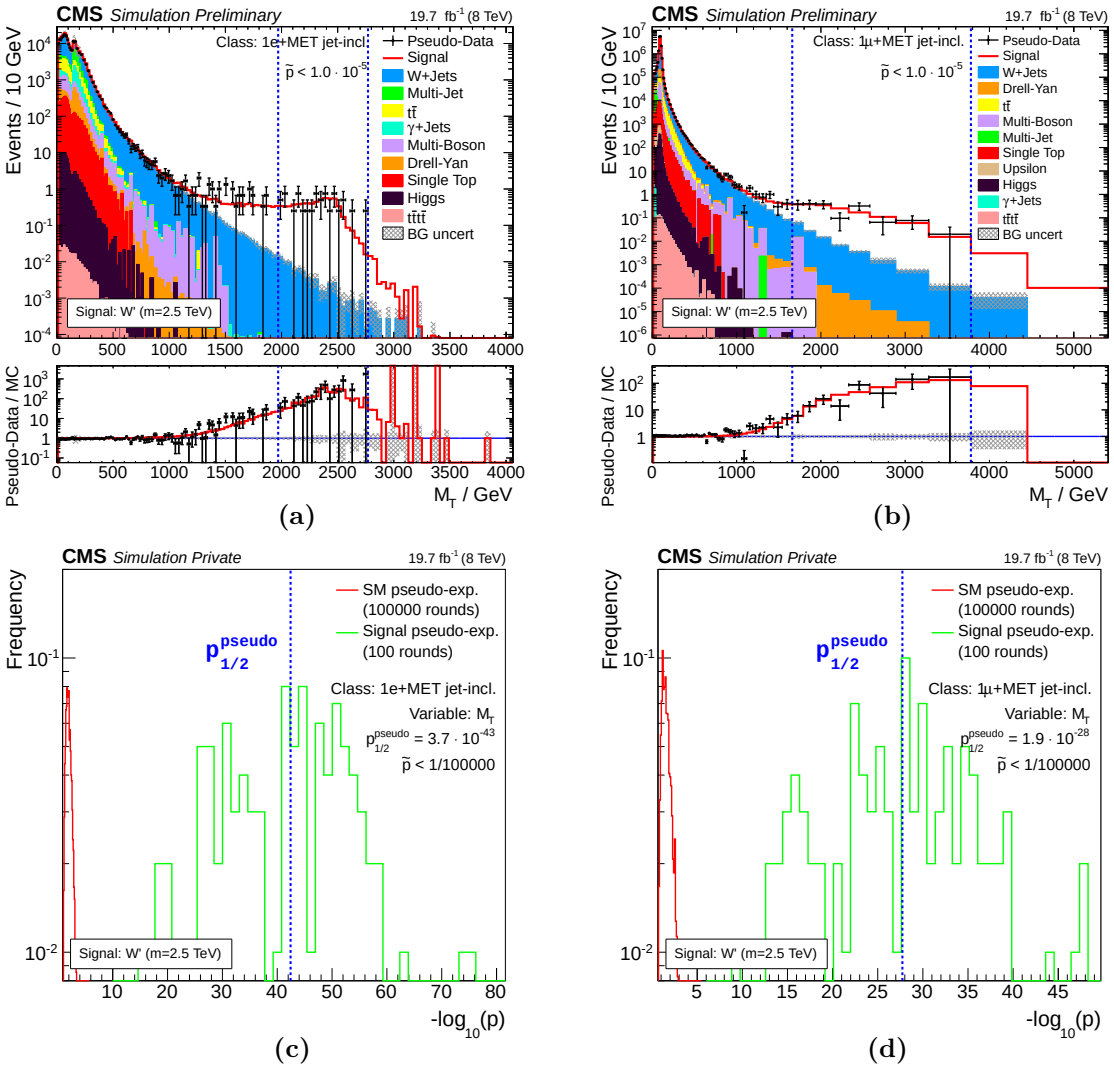


Figure 12.5: M_T distribution (a, b) and corresponding p-value distribution (c, d) for the 1e+1jet+Njet and 1μ+1jet+Njet jet-inclusive event classes. In the upper two figures, black markers represent the median result of pseudo-data from 100 signal rounds assuming a W' scenario with $M_{W'} = 2.5$ TeV. The signal itself is shown by a red line. In the lower two figures, pseudo-experiments from the LEE calculation are given in red, those created under the signal hypothesis are shown in green and the median p-value is given by a blue dashed line. Histograms are normalized to the number of pseudo-experiments (given in legend) and a separate binning is used to accent features of each histogram individually. Figures in the top row are published in identical form in [5].

$\sqrt{s} = 8$ TeV yields $M_T = 2.3$ TeV[1] in the electron and 2.1 TeV in the muon channel. An additional kinematic selection is applied to enhance the signal yield with respect to the SM. Throughout the considered search region measured data agree well with the SM expectation, showing no indications of a W' signal. In consequence the analysis proceeds to set upper exclusion limits on the mass of the investigated hypothetical boson. An observed limit at 95% confidence-level of 3.28 TeV is determined for the combination of electron and muon channel. Similar conclusions can be reached from the results of the MUSiC analysis as represented by the global overview in Fig. 12.3

[1] In the analysis documented in [228] M_T is calculated under an approximation assuming massless particles. For high transverse momentum particles this M_T definition converges towards the one used by MUSiC, i.e., Eq. (7.4).

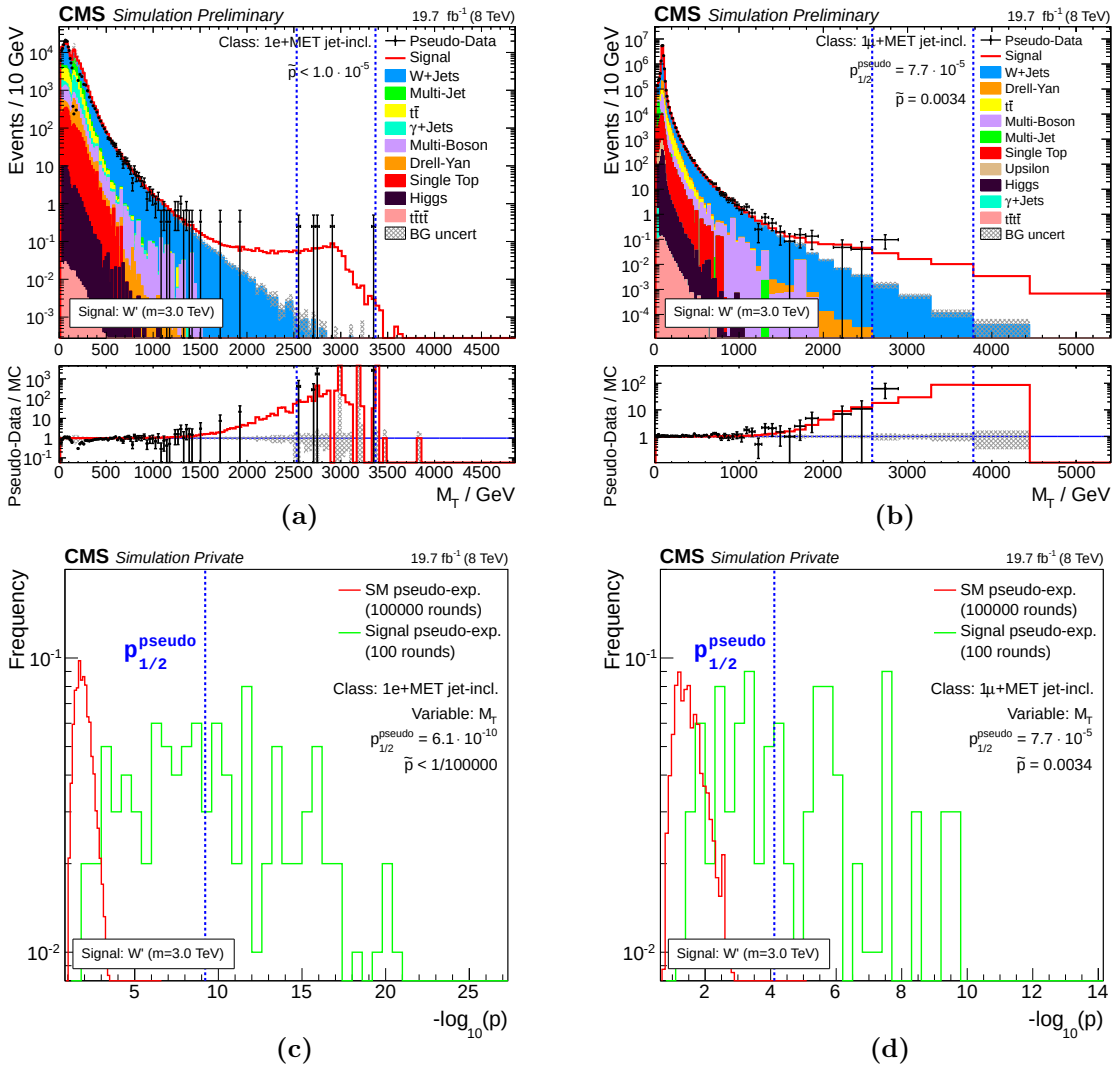


Figure 12.6: M_T distribution (a, b) and corresponding p-value distribution (c, d) for the 1e+1jet+Njet and 1μ+1jet+Njet jet-inclusive event classes. In the upper two figures, black markers represent the median result of pseudo-data from 100 signal rounds assuming a W' scenario with $M_{W'} = 3.0$ TeV. The signal itself is shown by a red line. In the lower two figures, pseudo-experiments from the LEE calculation are given in red, those created under the signal hypothesis are shown in green and the median p-value is given by a blue dashed line. Histograms are normalized to the number of pseudo-experiments (given in legend) and a separate binning is used to accent features of each histogram individually. Figures in the top row are published in identical form in [5].

and details of the most significant event classes in Tab. 12.1. A considerable number of event class distributions deviate strongly from the SM expectation for $M_{W'} \leq 3.0$ TeV, considering the **O**verflow **F**raction (OF) of up to 100%. For $M_{W'} = 3.5$ TeV only four of the 100 signal induced pseudo-data distributions become ultimately significant and on average far less than one distribution deviates more strongly than expected from the SM (see Fig. 12.3d).

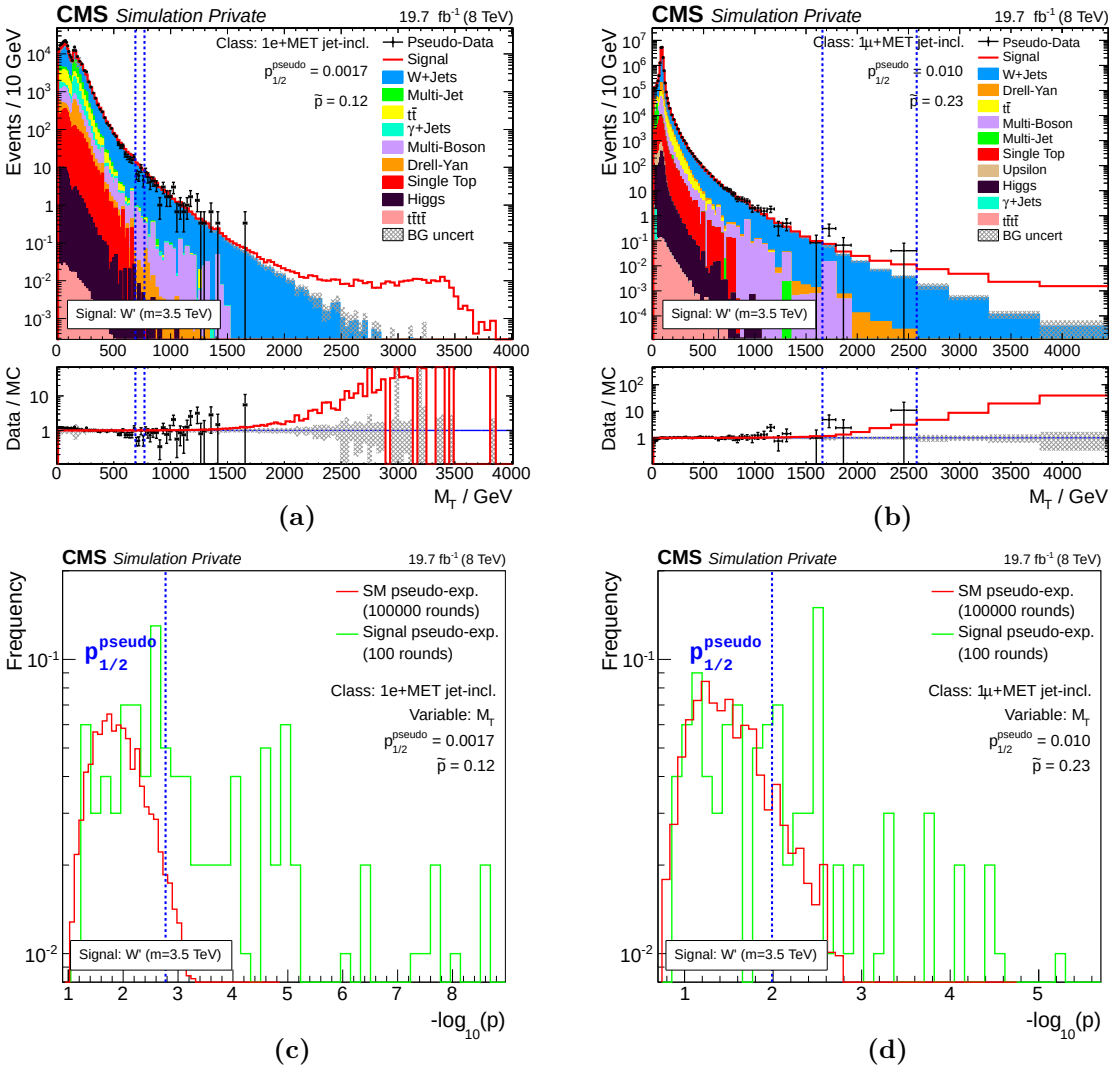


Figure 12.7: M_T distribution (a, b) and corresponding p-value distribution (c, d) for the 1e+1jet+Njet and 1μ+1jet+Njet jet-inclusive event classes. In the upper two figures, black markers represent the median result of pseudo-data from 100 signal rounds assuming a W' scenario with $M_{W'} = 3.5$ TeV. The signal itself is shown by a red line. In the lower two figures, pseudo-experiments from the LEE calculation are given in red, those created under the signal hypothesis are shown in green and the median p-value is given by a blue dashed line. Histograms are normalized to the number of pseudo-experiments (given in legend) and a separate binning is used to accent features of each histogram individually.

12.3 Rediscovering Standard Model Processes

A third approach is chosen to evaluate the sensitivity of the MUSiC analysis, for which a known SM process is removed from the MC simulation. Relying on this reduced SM description, all event classes are reevaluated using the full set of measured CMS data considered for the main part of this analysis (see Sec. 8.1). Analogous to the procedure established in Ch. 11, up to $2{\cdot}10^6$ SM-only pseudo-experiments, based on the incomplete SM simulation, are generated for every event class distribution to determine a LEE corrected significance for the deviation in the identified RoI.

Rediscovering the ZZ Process

Initially, a study is performed to uncover the production of two Z bosons from the measured data. Given at $\sqrt{s} = 8$ TeV with a next-to-leading order cross section of around 3 pb (based on values from simulation data sets explained in Sec. 8.2.1), a very clear signature is expected from the decay into four charged leptons. An overview of the complete results is given in Fig. 12.8, showing distributions of $\tilde{p}$-values for all three kinematic variables of exclusive event classes. In $\sum |\vec{p}_{\mathrm{T}}|$ and $M_{(\mathrm{T})}$ two event classes become so significant that the LEE correction cannot be evaluated within the maximum number of rounds. Moreover three (two) event classes show a deviation larger than $3\,\sigma$, of which only up to one is given within 95% of the uniform expectation's median. Unsurprisingly,

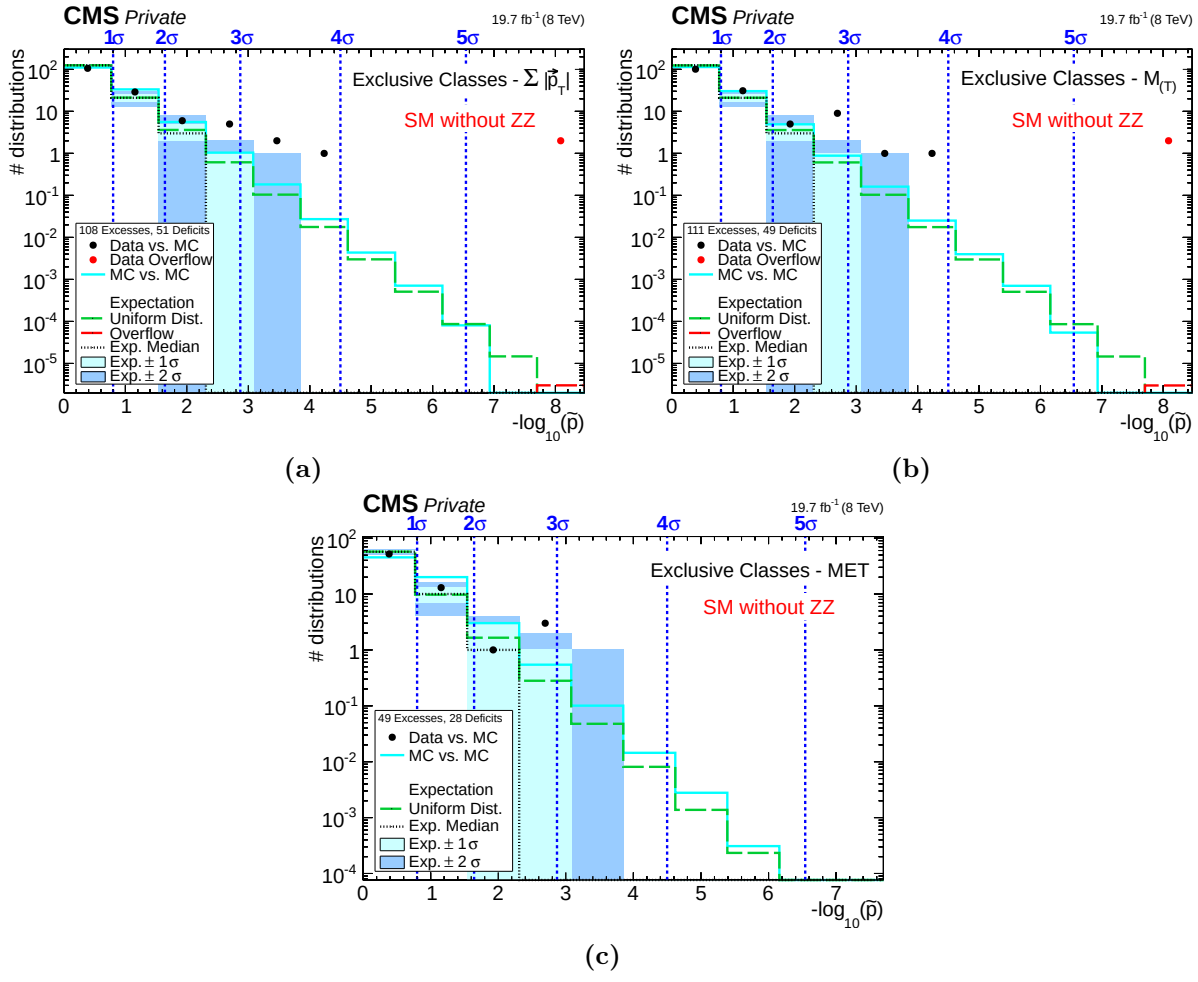


Figure 12.8: Distributions of $\tilde{p}$-values for kinematic distributions (a: $\sum |\vec{p}_{\mathrm{T}}|$, b: $M_{(\mathrm{T})}$, c: MET) of exclusive event classes. Black or red markers represent measured data compared to the SM MC expectation from which modeling of ZZ production is removed. In the first bin a: 106 [b: 101, c: 52] distributions are observed in data, with a: $125^{+5}_{-4}(1\,\sigma) \pm 9(2\,\sigma)$ [b: $125^{+4}_{-5}(1\,\sigma)^{+8}_{-10}(2\,\sigma)$, c: $57 \pm 3(1\,\sigma) \pm 6(2\,\sigma)$] expected. All figures are published in similar form in [5].

the MET distribution does not provide good sensitivity to the ZZ process, as only a rather small amount of MET is anticipated. Genuine MET can be initiated by Z bosons decaying into neutrinos. However, the majority of such events will be produced from non-boosted Z bosons, resulting in back-to-back neutrinos and a balanced event with MET below the analysis selection threshold.

Looking more closely at the ten most significant event classes Tab. 12.2 reveals precisely final states with four charged leptons to deviate strongly, 2e+2μ, 4μ, 4e and corresponding (jet-)inclusive event classes. A comparison of the corresponding $\sum|\vec{p}_{\mathrm{T}}|$ distributions with those obtained using the complete SM description is given in Fig. 12.9. Obvious excesses of data events are visible throughout a large portion of the kinematic range in each case. Similar observations can be made in the $M_{(\mathrm{T})}$ distributions of exclusive event classes (see Fig. D.3). No striking differences become apparent when considering the event's MET (see Fig. D.4). The RoI remains in the same position for all shown distributions.

	Event Class	p_{data}	$\tilde{p}$	RoI / GeV	N_{data}	$N_{\mathrm{SM}}\pm\sigma_{\mathrm{SM}}$
$\sum\|\vec{p}_{\mathrm{T}}\|$ distributions	2e+2μ	$7.42\cdot10^{-29}$	$< 5\cdot10^{-7}$	140 - 340	38	2.23±0.37
	4μ	$4.33\cdot10^{-13}$	$< 5\cdot10^{-7}$	150 - 690	16	1.11±0.15
	4e	$1.22\cdot10^{-6}$	$3.00\cdot10^{-5}$	130 - 430	9	0.89±0.19
	2e+1γ+4jets	$4.58\cdot10^{-4}$	$4.70\cdot10^{-4}$	410 - 920	1	$(3.8\pm3.3)\cdot10^{-4}$
	1e+1γ+MET	$7.70\cdot10^{-5}$	$7.79\cdot10^{-4}$	270 - 330	7	0.43±0.48
	2e+2μ+1jet	$7.39\cdot10^{-5}$	0.001	250 - 740	5	0.383±0.090
	1e+1μ+1γ	$6.58\cdot10^{-4}$	0.0011	230 - 370	2	0.022±0.026
	4e+MET	0.001	0.0021	210 - 330	2	0.042±0.019
	3e+2jets+MET	$9.58\cdot10^{-5}$	0.0028	950 - 1230	6	0.54±0.24
	4μ+1jet	$8.24\cdot10^{-4}$	0.0041	250 - 620	3	0.170±0.040
$M_{(\mathrm{T})}$ distributions	2e+2μ	$1.59\cdot10^{-28}$	$< 5\cdot10^{-7}$	170 - 700	39	2.46±0.40
	4μ	$3.68\cdot10^{-13}$	$< 5\cdot10^{-7}$	170 - 220	12	0.466±0.063
	4e	$7.89\cdot10^{-7}$	$3.80\cdot10^{-5}$	110 - 460	9	0.85±0.18
	2e+1γ+4jets	$4.58\cdot10^{-4}$	$4.72\cdot10^{-4}$	650 - 1100	1	$(3.8\pm3.3)\cdot10^{-4}$
	1e+1γ+MET	$1.53\cdot10^{-4}$	0.0015	250 - 330	9	0.80±0.87
	4e+MET	0.0011	0.0017	210 - 330	2	0.044±0.019
	1μ+1γ	$1.25\cdot10^{-5}$	0.0018	460 - 830	3	0.031±0.022
	4μ+1jet	$2.47\cdot10^{-4}$	0.0018	380 - 700	3	0.110±0.031
	2μ+1γ+1jet+MET	$4.10\cdot10^{-4}$	0.0023	320 - 440	3	0.073±0.090
	2e+2μ+1jet	$1.42\cdot10^{-4}$	0.0023	300 - 880	5	0.450±0.090
MET distributions	4e+MET	0.0011	0.0016	50 - 80	2	0.045±0.018
	2μ+1γ+1jet+MET	0.0017	0.0025	60 - 100	3	0.13±0.14
	1e+1γ+MET	$5.09\cdot10^{-4}$	0.0037	50 - 90	14	1.8±2.1
	3e+2jets+MET	$8.27\cdot10^{-4}$	0.0089	140 - 190	6	0.89±0.32
	1e+2μ+5jets+MET	0.024	0.033	50 - 140	2	0.21±0.11
	3μ+MET	0.0067	0.054	230 - 290	3	0.31±0.16
	1μ+1γ+4jets+MET	0.034	0.055	50 - 210	8	2.8±1.5
	2e+1μ+4jets+MET	0.022	0.060	100 - 190	3	0.47±0.29
	1e+3μ+MET	0.04	0.067	170 - 330	1	0.039±0.021
	2e+1γ+MET	0.042	0.073	50 - 140	5	1.2±1.0

Table 12.2: Overview of the ten most significant $\sum|\vec{p}_{\mathrm{T}}|$, $M_{(\mathrm{T})}$ and MET distributions for exclusive event classes of measured data compared to an SM MC expectation from which modeling of ZZ production is removed. p-value and $\tilde{p}$ value are stated in each case. The bin region of the RoI is given, as well as the number of data events N_{data} and SM MC events $N_{\mathrm{SM}}\pm\sigma_{\mathrm{SM}}$ in the RoI. Corresponding results for (jet-)inclusive event classes are given in Tab. D.1 and Tab. D.2.

Furthermore, many of the most strongly deviating event classes from the analysis of the full SM description (compare Tab. C.4) become even more significant once the ZZ process is removed. Additionally, the 2e+1γ+4jets event class, disregarded in the main analysis as discussed in length

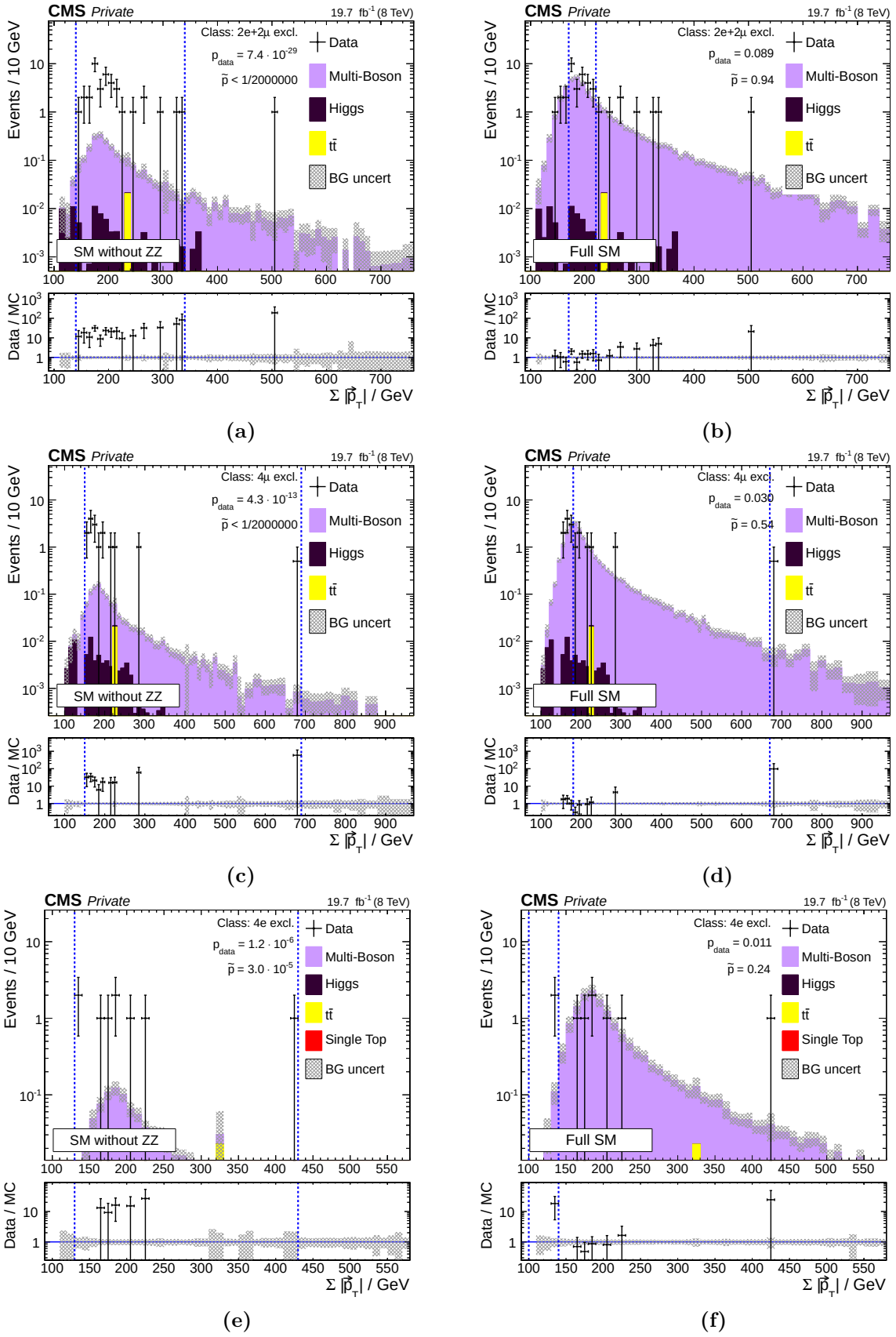


Figure 12.9: The left column (a, c, e) shows the three most significant $\sum |\vec{p}_{\mathrm{T}}|$ distributions for exclusive event classes of measured data compared to an SM MC expectation from which modeling of ZZ production is removed. Strong deviations are observed. For comparison, results of the same distributions using the full SM description (including ZZ production) are shown in the right column. Figures in the top row are published in similar form in [5].

in Sec. 11.2.5, is no longer rejected by the low generated event count treatment once the ZZ process is removed. It becomes the fourth most significant $\sum|\vec{p}_{\mathrm{T}}|$ distribution of exclusive event classes. Corresponding distributions and the mapping of investigated regions to p-values or failed criteria are summarized in Fig. 12.10. Removing data sets associated to the ZZ process sufficiently reduces the relative statistical uncertainty in a small number of regions and a p-value can be calculated. However, this event class distribution represents an extreme case and only a small number of regions (six) pass all criteria of the low generated event count treatment. All of these regions show a relative statistical uncertainty of around $\sigma^{\mathrm{stat}}_{\mathrm{SM}}/N_{\mathrm{SM}} = 0.55$, just below the threshold of 0.6 used to veto regions of insufficient number of generated MC events. Consequently, the calculated p-value should be taken with a grain of salt.

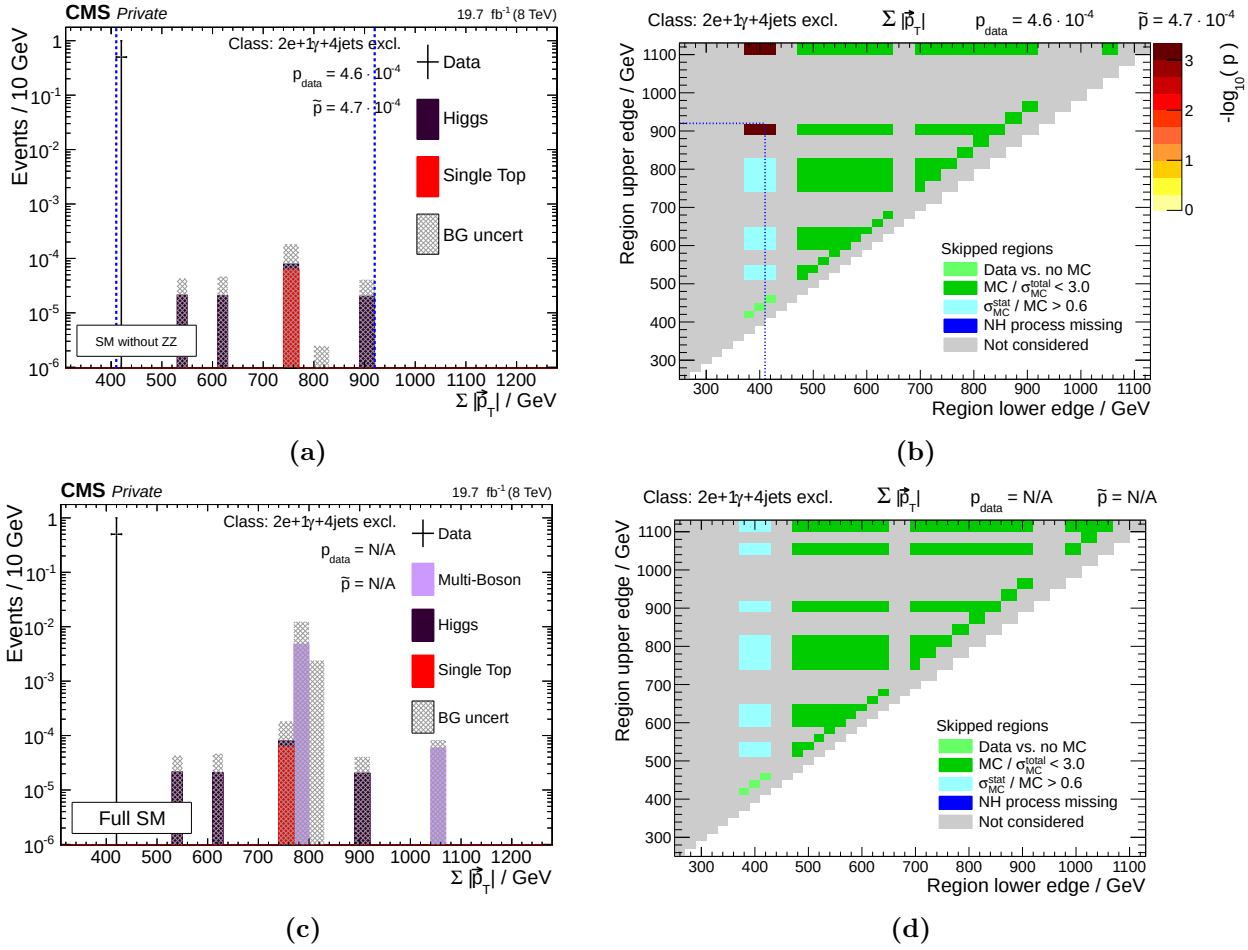


Figure 12.10: $\sum|\vec{p}_{\mathrm{T}}|$ distribution and mapping of considered regions to p-values for 2e+1γ+4jets. The upper row shows results obtained from the study of removing ZZ production from the SM MC. Results in the lower row are obtained using the full SM MC modeling. Only very few events are expected overall in this final state. Contributions of ZZ production in two bins of the distribution introduce large statistical uncertainties, causing the corresponding regions to be excluded by the MUSiC low generated event count treatment. Without ZZ production a small number of regions pass all criteria and a RoI (blue lines) can be determined.

In addition to the event class distributions listed in Tab. 11.5 the $\sum|\vec{p}_{\mathrm{T}}|$ and $M_{(\mathrm{T})}$ distributions of 3e+2μ are disregarded because of an insufficient amount of SM MC events if the ZZ process is removed. As can be seen in Fig. 12.11, this event class contains only one observed event and a small number of expected (0.0089 ± 0.0078) events, even when the ZZ process is accounted for. Only

few regions allow a feasible p-value calculation and removing ZZ production eliminates even these. Careful investigation of the event class in Fig. 12.11a could lead to first hints of an overlooked SM process, but the clear excesses in event classes with four charged leptons offer a much more comprehensive insight.

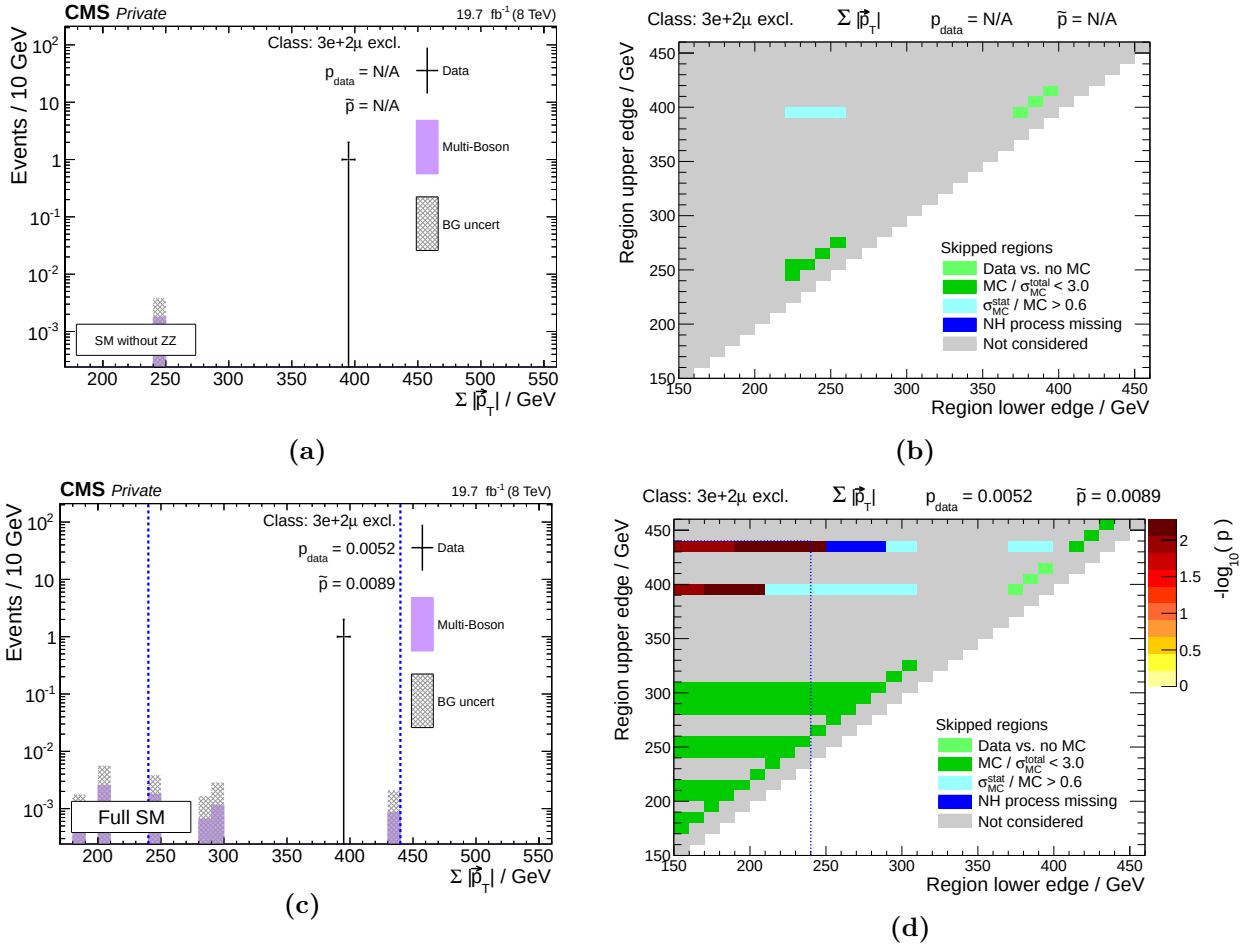


Figure 12.11: $\sum |\vec{p}_{\mathrm{T}}|$ distribution and corresponding mapping of considered regions to p-values for 3e+2μ. The upper row represents results obtained by removing ZZ production from the SM MC. In this case the event class does not contain a sufficient amount of SM MC events and a p-value is not calculated. Results in the lower row are obtained using the full SM MC modeling. A few regions pass all criteria of the low generated event count treatment and a RoI (blue lines) with corresponding p-value is determined.

Corresponding results for (jet-)inclusive event classes are given in App. D.2. Overall similar features are found in these event class definitions, leading to comparable conclusions. Absence of the ZZ process is clearly visible, especially in final states with four leptons. In some cases the inclusiveness enhances deviations by combining events from several significant exclusive event classes, e.g., 4e ($p_{\mathrm{data}} = 1.22{\cdot}10^{-6}$), 4e+MET ($p_{\mathrm{data}} = 0.001$) and others containing four electrons are summarized in 4e+X ($p_{\mathrm{data}} = 1.47 \cdot 10^{-9}$). In other cases a combination diminishes the significance, since significant deviations are concealed under additional events.

Rediscovering the WW Process

Results of a similar study, excluding the production of WW events from the SM MC description, can be found in Fig. 12.12, for all three kinematic variables of exclusive event classes. Based on simulations, this process is produced with a next-to-leading order cross section of around 55 pb in proton-proton collisions at $\sqrt{s} = 8$ TeV. An overview of the respective eight most significant exclusive event classes for all three kinematic distributions is given in Tab. 12.3. Comparing these

results to those obtained with the full SM expectation (see Tab. C.4) reveals that a very similar outcome is observed. Removing modeling of the WW process causes individual event classes to become more significant within the stated precision of the p-value, these are accentuated by a green check mark in Tab. 12.3. Some event class distributions become slightly less significant for cases in which a deficit becomes less pronounced. These are accentuated by a red cross in Tab. 12.3. The strongest change in significance over all kinematic variables of exclusive event classes is found in the $M_{(\mathrm{T})}$ distribution of 1e+1μ+MET, with $\tilde{p} = 0.63$ reduced to $\tilde{p} = 4.3 \cdot 10^{-3}$ once WW production is removed. In this final state events from DY are suppressed by considering two leptons of different flavor. Also, the implicit jet veto applied in exclusive event classes reduces $\mathrm{t\bar{t}}$ contributions. Yet overall no strong influence is seen. The distributions of $\tilde{p}$-values agree well with the expectation and do not indicate that a process is missing from the modeling of the SM.

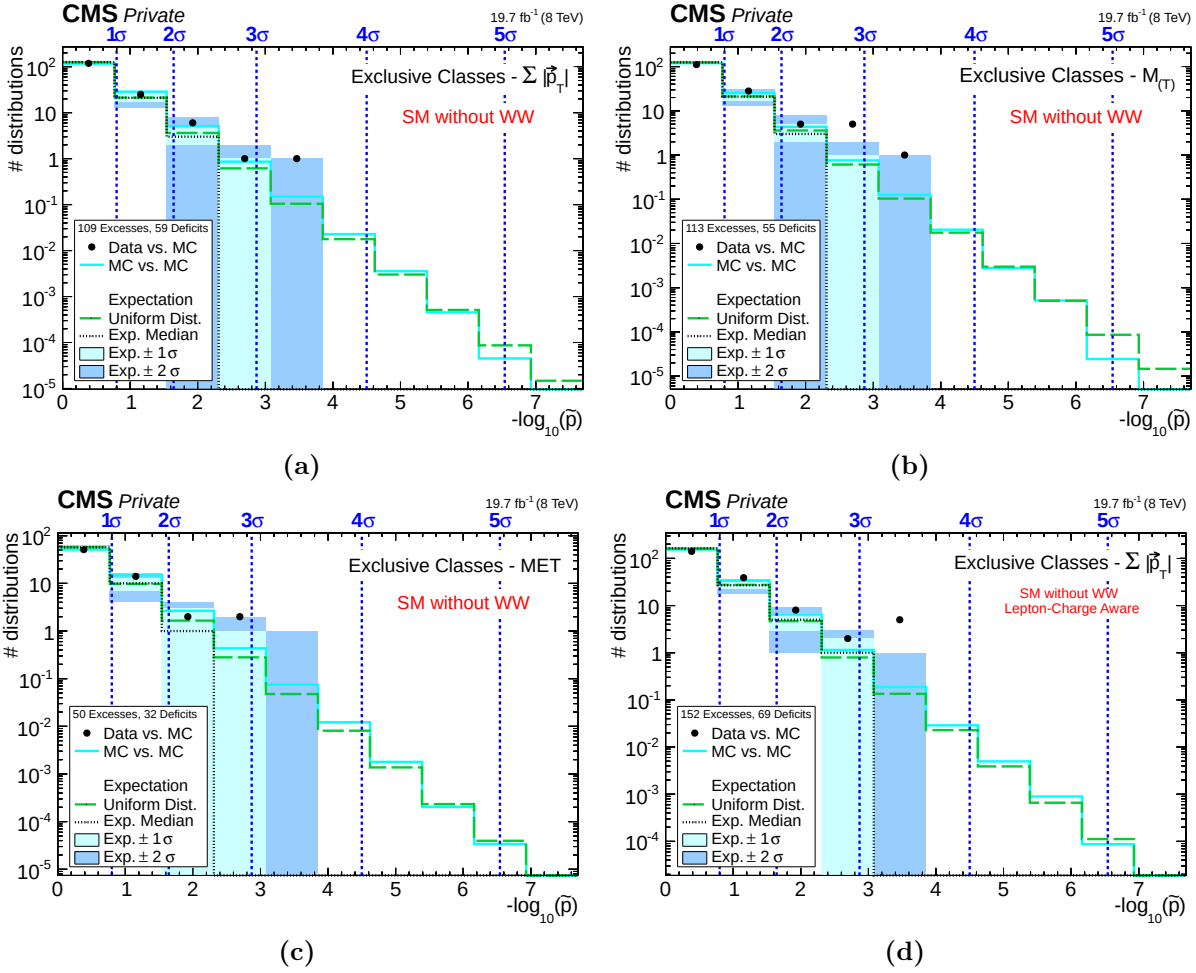


Figure 12.12: Distributions of $\tilde{p}$-values for kinematic distributions (a: $\sum |\vec{p}_{\mathrm{T}}|$, b: $M_{(\mathrm{T})}$, c: MET) of exclusive event classes. Black markers represent measured data compared to the SM MC expectation from which modeling of WW production is removed. The distribution given in (d) shows results for $\sum |\vec{p}_{\mathrm{T}}|$ distributions of exclusive event classes taking lepton charge into account. In the first bin a: 118 [b: 111, c: 51, d: 141] distributions are observed in data, with a: $125^{+5}_{-4}(1\,\sigma) \pm 9(2\,\sigma)$ [b: $125^{+4}_{-5}(1\,\sigma)^{+8}_{-10}(2\,\sigma)$, c: $57 \pm 3(1\,\sigma) \pm 6(2\,\sigma)$, d: $162 \pm 5(1\,\sigma)^{+10}_{-11}(2\,\sigma)$] expected from the SM without WW production.

Since the WW cross section is over ten times larger than that of the ZZ process, this result is initially surprising, but can be understood when contemplating the applied event selection criteria. As explained in Sec. 9.1.3, an event is only taken into consideration by the analysis if it contains

	Event Class	p	$\tilde{p}$	RoI / GeV	N_{data}	$N_{\text{SM}} \pm \sigma_{\text{SM}}$	Comment
$\sum \lvert\vec{p}_{\text{T}}\rvert$ distributions	1e+1γ+MET	$6.04 \cdot 10^{-5}$	$5.81 \cdot 10^{-4}$	270 - 330	7	0.35±0.49	✓
	3e+2jets+MET	$9.99 \cdot 10^{-5}$	0.0027	950 - 1230	6	0.55±0.24	–
	1e+1μ+1γ	0.0019	0.0050	230 - 370	2	0.041±0.043	–
	3e+2μ	0.0052	0.0091	240 - 440	1	0.0044±0.0037	–
	1e+2μ+5jets+MET	0.0073	0.011	670 - 1050	2	0.115±0.052	–
	1μ+1γ	$7.25 \cdot 10^{-4}$	0.014	390 - 790	2	0.032±0.020	–
	4e+MET	0.012	0.022	210 - 330	2	0.157±0.041	–
	1μ+≥6jets	$2.69 \cdot 10^{-4}$	0.028	1060 - 1260	23	63.0±9.9	✗
$M_{(\text{T})}$ distributions	1e+1γ+MET	$5.92 \cdot 10^{-5}$	$6.58 \cdot 10^{-4}$	250 - 330	9	0.60±0.79	✓
	1μ+1γ	$1.41 \cdot 10^{-5}$	0.0022	460 - 830	3	0.032±0.022	–
	3e+2jets+MET	$9.08 \cdot 10^{-5}$	0.0025	950 - 1230	6	0.54±0.24	–
	2μ+1jet	$3.25 \cdot 10^{-5}$	0.0026	90 - 120	1830	1190±150	✓
	1e+1μ+1γ	$8.63 \cdot 10^{-4}$	0.0035	310 - 540	2	0.027±0.029	–
	1e+1μ+MET	$9.44 \cdot 10^{-5}$	0.0043	150 - 160	662	378±72	✓
	1e+1γ+2jets	$3.07 \cdot 10^{-4}$	0.0053	2110 - 2260	1	$(1.1 \pm 3.3) \cdot 10^{-4}$	–
	2μ+5jets	$5.40 \cdot 10^{-5}$	0.0063	1990 - 2220	13	2.1±1.1	–
MET distributions	1e+1γ+MET	$1.65 \cdot 10^{-4}$	0.001	50 - 90	14	1.5±1.9	✓
	2μ+1γ+1jet+MET	0.0019	0.0041	60 - 100	3	0.14±0.15	–
	3e+2jets+MET	$8.47 \cdot 10^{-4}$	0.0084	140 - 190	6	0.89±0.32	–
	4e+MET	0.017	0.025	50 - 80	2	0.190±0.045	–
	1e+2μ+5jets+MET	0.024	0.033	50 - 140	2	0.21±0.11	–
	1μ+1γ+4jets+MET	0.024	0.042	50 - 210	8	2.6±1.4	✓
	1e+3μ+MET	0.018	0.051	170 - 230	1	0.0173±0.0096	–
	2e+1μ+4jets+MET	0.022	0.060	100 - 190	3	0.47±0.29	–

Table 12.3: Overview of the eight most significant $\sum \lvert\vec{p}_{\text{T}}\rvert$, $M_{(\text{T})}$ and MET distributions for exclusive event classes of measured data compared to an SM MC expectation from which modeling of WW production is removed. p-value and $\tilde{p}$ value are stated in each case. The bin region of the RoI is given with corresponding number of data N_{data} and SM MC $N_{\text{SM}} \pm \sigma_{\text{SM}}$ events. Event classes with a p-value smaller than the one obtained with the full SM simulation are marked with a green check, ones with a larger p-value are marked with a red cross (compare to Tab. C.4).

at least one rather high energetic isolated lepton or two isolated same flavor leptons with moderate transverse momenta. In ZZ events only one Z boson must decay leptonically for the event to pass the trigger criteria, where low p_{T} leptons are included over the double lepton triggers. For events given by WW production either both W bosons must decay leptonically to pass the double lepton trigger, or if only one W boson decays leptonically, the corresponding charged lepton must have a rather high p_{T} (especially for electrons). These restrictions greatly reduce the total number of WW events taken into account by the analysis.

Dedicated analyses performed to measure the ZZ and WW cross sections [311, 312] apply a trigger selection similar to the one used in this thesis, but are able to reduce contributions from other processes. Also, the number of search channels is extended to also include events with reconstructed τ leptons decaying hadronically. Isolation criteria and a Z boson candidate selection via a di-lepton invariant mass restraint around the Z boson mass are applied to enrich the number of resonantly produced ZZ pairs. For WW production the criteria are more sophisticated. The two identified leptons must have opposite charge. Contributions from $\text{t}\bar{\text{t}}$ are reduced by using top quark tagging methods and rejecting events with well reconstructed jets. DY events are removed by applying a minimum threshold on the events' MET. Combining all these criteria allows the dedicated analysis to efficiently measure WW production.

The MUSiC analysis' generalized approach is not tailored to target such a specific process, especially when other SM processes deliver high contributions in the same final state. A method to gain more sensitivity to the WW process is given by the lepton-charge aware event classes, as discussed in Sec. 11.3. Events from $W^+ W^-$ are separated more clearly from events with two like-sign leptons, originating from other processes. A reevaluation of the $\sum \lvert\vec{p}_{\text{T}}\rvert$ distributions of exclusive

event classes considering the net lepton charge of each event shows a slight increase of deviations over $3\,\sigma$ (compare Fig. 11.36b and Fig. 12.12d). Although the most significant deviations are still given by those discussed in Sec. 11.3, 1e+1μ+MET[0Q] becomes noticeably more significant, with $\tilde{p}$ =0.53 reduced to $\tilde{p}$ =0.031 when WW production is removed from the SM simulation. The decrease in $\tilde{p}$ is less pronounced if lepton charge is not accounted for. A comparison of both cases is given in Fig. 12.13. In conclusion, contributions from WW are not clear enough for the MUSiC algorithm to deliver a definite result. Sensitivity to this process could be increased by applying quark flavor tagging techniques. As can be seen in Fig. 12.13, processes involving top quarks are dominant in final states where the absence of WW simulation is most obvious.

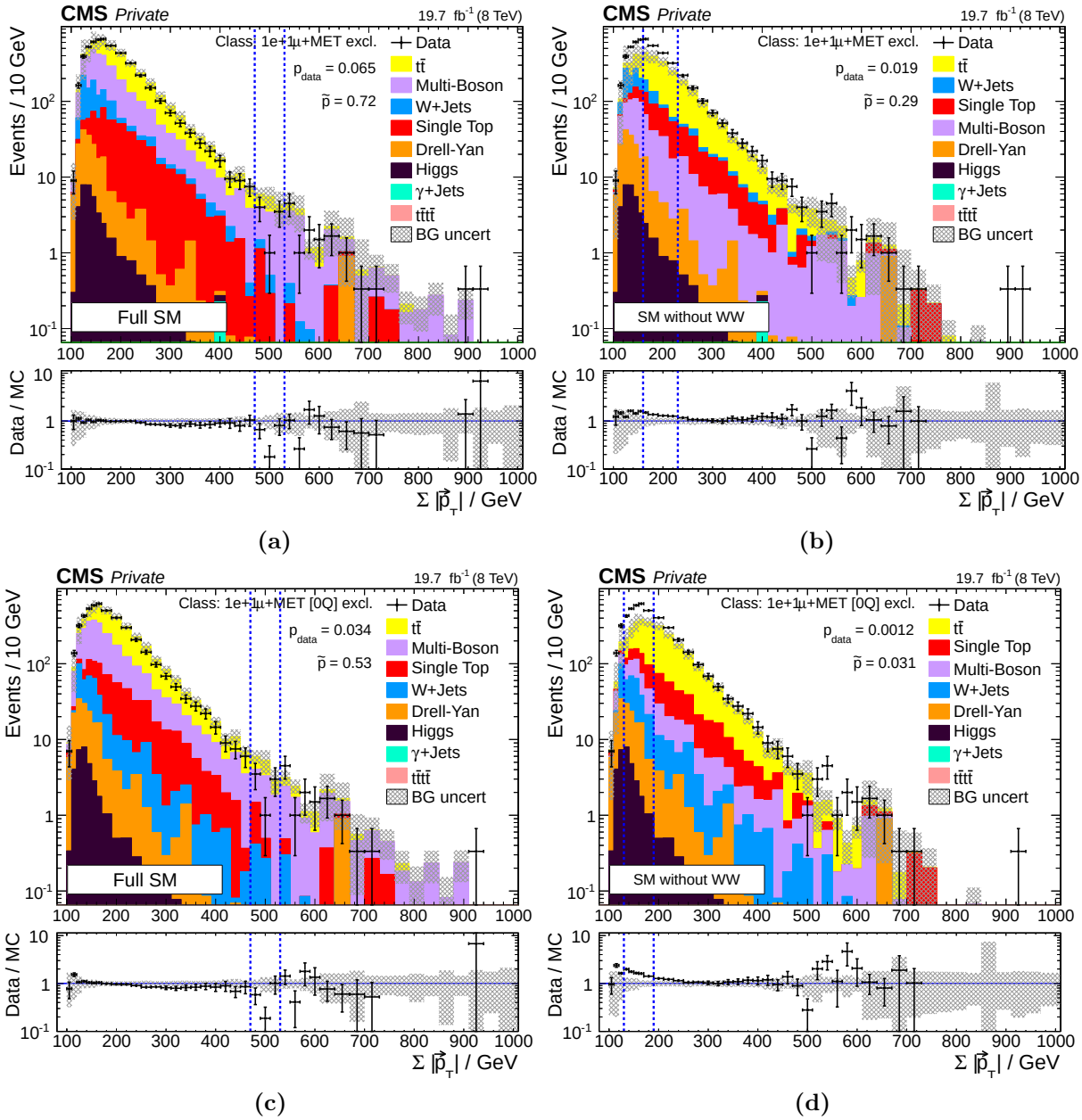


Figure 12.13: $\sum |\vec{p}_{\mathrm{T}}|$ distribution of events containing precisely 1e+1μ+MET. In the upper row events are summarized in one event class, in the lower row only events with oppositely charged leptons are considered. The left column includes the full SM simulation, whereas in the right column WW production is removed.

Chapter 13
Outlook

Closely associated with the analysis performed for this thesis, parallel and partially ongoing developments of the MUSiC workflow are presented here in a short summary. Many more details are given in the referenced master and doctoral theses.

Jet-Triggered Events and Bottom Quark Jet Tagging

Based on methods and intermediate results obtained during the work leading up to this thesis, an extension of the MUSiC search strategy is carried out on events passing a single jet trigger [257]. In consequence, the total amount of analyzed data is increased by roughly 10%. Furthermore, a tagging of jets initiated by a bottom quark is included in the analysis. Adding these additional event topologies increases the total number of exclusive and inclusive event classes that do not contain an electron or muon by over 200 event classes. Considering all additional event classes, the majority is introduced by dividing all identified jets into ones originating from bottom quarks and ones associated to other quark types or gluons, yielding almost 1000 new exclusive event classes. Jet-inclusive event classes were not yet implemented at the time and no upper threshold on the number of jets resolved in an event class was set.

Many of the issues already reported on in [257] have now been better understood and mitigated by methods developed for the analysis presented in this thesis, such as insufficient modeling of event classes and kinematic regions with a very low number of generated MC simulation events, unreliable description of events with a very high jet multiplicity, systematic uncertainty contributions from unclustered energy and modeling of W+jets using a flat NLO k-factor. Apart from these observations several interesting deviations are uncovered in event classes containing bottom quark tagged jets and also in ones containing di-jets. Some deviations are attributed to characteristics of the used simulation data sets, including the unsatisfactory LO description of QCD multi-jet events and the relatively small number of QCD multi-jet events containing bottom quark jets. The latter issue was addressed by increasing the generated number of bottom quark jet enriched QCD multi-jet events. Moreover, the used LO data sets generated purely with PYTHIA6 are compared to data sets generated at LO with MadGraph and at NLO in QCD with SHERPA. The choice of data set has an influence especially on modeling of the low energy region in the di-jet spectrum. Furthermore, an interesting excess is found in the tail of the mass distribution of the 1e+1μ+1b(+X) event classes, which should be monitored in future instances of the analysis with more data. In summary, good agreement was found in the majority of investigated event class distributions.

Model Independent Exclusion Limits and Reinterpretation

The analysis strategy documented in [162] offers an alternative statistical interpretation of the MUSiC classification results presented in this thesis. Instead of running the search algorithm to calculate the significance of observed deviations, each event class distribution is used to determine exclusion limits on the number of additional events above the SM prediction, potentially originating from BSM signal. The motivation behind this approach is to offer a quantitative result, especially for the theory model-building community. Limits are provided if the MUSiC analysis proves reasonable agreement of the measured data to the SM simulation. A frequentist exclusion limit is calculated in every valid kinematic region of all MUSiC event classes. Restraints from the low MC generated event count treatment are taken into consideration as well as a dedicated preselection of regions to reduce the computing time costly limit calculation by restraining the method

to fewer regions. Each upper exclusion limit represents a confidence interval stating the amount of additional events above the SM expectation that would still be consistent with the observed data at 95% confidence level. In a standalone second step, these signal model independent results can be reinterpreted in terms of given BSM signal models, yielding an exclusion limit in terms of signal cross section. Knowledge of the BSM signal acceptance and efficiency in all kinematic regions and search channels is necessary for translating the limit on the number of events into one on a specific signal cross section. To enable non-CMS members access of these efficiencies, the widespread Delphes framework [313] is used to simulate detector response and event reconstruction in the CMS detector.

Details of the limit calculation procedure and an optimization of Delphes parameterizations for the object energy reconstruction are discussed in [162]. After the full strategy is established, model independent exclusion limits are calculated for all regions given by the 2012 CMS data. To validate the method, two BSM scenarios introducing heavy vector bosons W' and Z' are passed through the reinterpretation method. In both cases limits based on the MUSiC results are comparable with ones obtained by the corresponding dedicated analysis.

Analysis of Data Taken at $\sqrt{s} = 13$ TeV

As already referred to in the previous chapter, a continuation of the MUSiC analysis is performed for LHC Run II data at $\sqrt{s} = 13$ TeV as summarized in [308]. The full analysis chain was carried out on events containing at least one electron or muon for 2.3 fb^{-1} worth of proton-proton collision data collected in 2015. Good agreement between the measurement and the SM MC simulation is found throughout all event class types and kinematic variables. The strongest deviation is observed in a multi-boson dominated final state 3e+3jets at $3.0\,\sigma$, which is fully within the expected significance range given by fluctuations of the SM. Corresponding inclusive and jet-inclusive event classes as well as 3μ+3jets are less significant.

Furthermore, an initial glance at 26.4 fb^{-1} of data taken during 2016 (13 TeV) is discussed. Simulation data sets accounting for the changed run conditions were not available yet, so this first survey of the data is carried out with the SM simulation generated for the 2015 data-taking period. Thus, the full analysis including the distribution of $\tilde{p}$-values is not processed yet. Nevertheless, the most significant deviations observed in the 2015 data set are revisited and found not to be enhanced by the increased amount of data. Furthermore, details of reconstruction issues were investigated and discussed with the CMS Collaboration based on a survey of early 2016 experimental data.

Chapter 14
Conclusion

Following an approach unbiased by BSM signal models, the MUSiC analysis was applied to 19.7 fb^{-1} of proton-proton $\sqrt{s} = 8$ TeV collision data recored with the CMS detector in 2012. Concepts, methods and dedicated algorithms of the model independent search were presented, including extensions devised during the work leading up to this thesis. A new strategy for dealing with kinematic regions suffering from the influence of an insufficient amount of generated MC events was developed and compared to the previous method. Satisfactory results with improved statistical properties are achieved with the revised approach, yielding meaningful p-values in all investigated regions. Only very few distributions do not contain any valid region and are excluded from the global statistical analysis but are investigated individually. Moreover, an additional method to reduce the total runtime of the scanning step was deployed successfully.

Placing a focus on events containing electrons or muons and also considering photons, jets and MET, over 300 final states were respectively analyzed for exclusive, inclusive and jet-inclusive event classes, for which three kinematic variables were treated independently. A detailed description of the SM was constructed using an aligned set of MC data sets modeling individual processes, for the first time also considering signatures of the SM Higgs boson. Also, extended systematic uncertainties originating from various experimental and theoretical sources were included. All selection criteria were revised in the context of the 2012 data-taking and dedicated requirements for dealing with high p_{T} muons were added.

Results obtained from the MUSiC workflow on data taken at $\sqrt{s} = 8$ TeV show no obvious signs of physics beyond the SM. Good agreement between the data and MC-based SM expectation is found throughout a wide range of final states covering a multitude of different event compositions. The small number of strong deviations originating from individual event classes is still compatible with the global prediction from the SM. A detailed discussion of the most significant deviations was presented. These lead to conclusions about imperfectly modeled misidentification rates of photons in particular, but also leave room for possible first indications of new phenomena which must be followed up on in future analyses. Comparisons to previous MUSiC analyses and preliminary results from Run II of the LHC exhibit fewer events in the corresponding event classes and cannot be consulted for a definitive answer on the most significant deviations. It remains to be seen if further data, acquired during CMS operations in 2016 and later, can give a clearer picture. When considering the net-charge of leptons in the event class discrepancies are found which can be attributed to the simulation based modeling of the muon charge misidentification rate and the rate of muons produced by jets. Similar behavior was observed by previous instances of the MUSiC analysis and also by a dedicated CMS analysis, which resorts to a data-driven estimate of muons stemming from jets.

Through probing the SM expectation not only against pseudo-experiments generated from BSM signals, but also against ones created under an SM description lacking individual processes and the full SM-only hypothesis, the MUSiC analysis' sensitivity and false discovery rate could be evaluated. As anticipated, no strong deviations were observed for the SM-only case. Sensitivity to specific mass points of a W' boson scenario and the SM ZZ process were found, while the signature of the WW process it not strong enough for MUSiC to uncover, yet.

Part V

Appendix

Appendix A
Runtime Reduction via Bonferroni Correction

Often the majority of event classes considered in a scan exhibits only moderate p-values and the corresponding LEE correction only takes a few minutes of runtime per distribution, since the generation of pseudo-data is terminated once a sufficient precision is reached. Only a small number of strongly deviating distributions, for which the full amount of SM pseudo-experiments must be carried out, make up the majority of the runtime. Especially when considering sensitivity studies, as explained in Sec. 12, a handful of distributions become so significant that no $\tilde{p}$-value can be calculated, even within the full amount of rounds. A large portion of runtime can be saved by evaluating an estimate of the LEE correction beforehand and setting event class distributions to ultimately significant if the $\tilde{p}$-value estimate is sufficiently small without performing the SM pseudo-experiment generation.

A crude estimate of the LEE corrected significance can be achieved by the Bonferroni correction [314], in the MUSiC case given as:

$$\tilde{p} \leq p_{\text{data}} \cdot N_{\text{regions}} \leq p_{\text{data}} \cdot \frac{N_{\text{bins}}}{2} \cdot (N_{\text{bins}} - 1) \equiv \tilde{p}', \tag{A.1}$$

where the number of regions N_{regions} depends on the number of bins N_{bins} in the distribution, but will often be reduced by the MUSiC low MC generated event count treatment (Sec. 7.3.3) and minimum region width requirement (Sec. 7.3.1).

An example of the considerable runtime reduction based on results from an actual scan performed for the sensitivity study explained in Sec. 12.3 is shown in the following table.

Event Class	Runtime	p	$\tilde{p}$	N_{bins}	$\tilde{p}'$
2e+2μ+X	10 h	$1.16 \cdot 10^{-27}$	$<5 \cdot 10^{-7}$	90	$4.65 \cdot 10^{-24}$
4μ+X	11 h	$5.50 \cdot 10^{-13}$	$<5 \cdot 10^{-7}$	88	$1.93 \cdot 10^{-10}$
4e+X	15 h	$1.47 \cdot 10^{-9}$	$<5 \cdot 10^{-7}$	129	$1.21 \cdot 10^{-5}$
2e+1γ+4jets+X	1 h	$5.58 \cdot 10^{-4}$	$7.07 \cdot 10^{-4}$	60	0.988
3e+2μ+X	0.5 h	$4.51 \cdot 10^{-3}$	$5.01 \cdot 10^{-3}$	40	0.271

The five most significant inclusive event classes are given for a scan of the $\sum |\vec{p}_{\text{T}}|$ distribution of data and SM MC from which the ZZ process has been removed. The first three event class distributions are so significant that none of the $2 \cdot 10^6$ SM pseudo-experiments results in a smaller p-value. The scan total runtime of around 46.5 h is dominated by these three event classes. An approximation of the $\tilde{p}$-value following Eq. (A.1) shows that the first two event classes result in a $\tilde{p}'$ value considerably smaller than the smallest $\tilde{p}$-value calculable with the given number of rounds. Here, the number of bins N_{bins} in each distribution is estimated by the highest filled bin, as the MUSiC scan algorithm refrains from adding empty bins to a potential RoI, thus reducing the considered number of regions. In this example the total runtime of the scan can be halved by assessing the approximate $\tilde{p}$-value beforehand. The third most significant event class 4e+X would need to be processed in the usual manner to exclude claiming a higher significance than actually on-hand.

Refinements of the proposed method could aid in returning $\tilde{p}'$ estimates closer to the actual $\tilde{p}$-value. Determining the actual number of considered regions, i.e., including all criteria used to

exclude bins from the region building and accounting for the minimum region width, would further reduce $\tilde{p}'$. A better estimate could help remove cases like that of 4e+X from the LEE correction procedure.

A disadvantage of omitting pseudo-data generation for ultimately significant event class distributions is that the corresponding SM pseudo-experiments are not available for the global overview given by the distribution of $\tilde{p}$-value (Sec. 7.3.7). Nevertheless, since a good agreement with the analytical description of the uniform distribution is given, the comparison of SM expectation and (signal induced pseudo-) data could be based on the analytical estimate alone.

Appendix B
Data Streams and Monte Carlo Simulation Data Sets

B.1 Overview of Collision Data Streams

Run Range	Data Stream	L in pb^{-1}	N_{events}
190456 - 193621	/SingleMu/Run2012A-22Jan2013-v1	880	$1.4 \cdot 10^7$
193834 - 196531	/SingleMu/Run2012B-22Jan2013-v1	4400	$5.3 \cdot 10^7$
198022 - 203742	/SingleMu/Run2012C-22Jan2013-v1	7100	$8.5 \cdot 10^7$
203777 - 208686	/SingleMu/Run2012D-22Jan2013-v1	7400	$9.0 \cdot 10^7$
190456 - 193621	/SingleElectron/Run2012A-22Jan2013-v1	880	$1.1 \cdot 10^7$
193834 - 196531	/SingleElectron/Run2012B-22Jan2013-v1	4400	$6.2 \cdot 10^7$
198022 - 203742	/SingleElectron/Run2012C-22Jan2013-v1	7000	$9.8 \cdot 10^7$
203777 - 208686	/SingleElectron/Run2012D-22Jan2013-v1	7400	$1.0 \cdot 10^8$
190456 - 193621	/DoubleMu/Run2012A-22Jan2013-v1	880	$4.8 \cdot 10^6$
193834 - 196531	/DoubleMuParked/Run2012B-22Jan2013-v1	4400	$2.6 \cdot 10^7$
198022 - 203742	/DoubleMuParked/Run2012C-22Jan2013-v1	7000	$3.5 \cdot 10^7$
203777 - 208686	/DoubleMuParked/Run2012D-22Jan2013-v1	7400	$3.6 \cdot 10^7$
190456 - 193621	/DoubleElectron/Run2012A-22Jan2013-v1	880	$9.5 \cdot 10^6$
193834 - 196531	/DoubleElectron/Run2012B-22Jan2013-v1	4400	$2.1 \cdot 10^7$
198022 - 203742	/DoubleElectron/Run2012C-22Jan2013-v1	7100	$3.3 \cdot 10^7$
203777 - 208686	/DoubleElectron/Run2012D-22Jan2013-v1	7400	$3.3 \cdot 10^7$
Total		19700	$7.1 \cdot 10^8$

Table B.1: Summary of CMS data streams used in this analysis in AOD format. Stated integrated luminosity values L are determined with the pixel cluster counting method. The given number of events N_{events} corresponds to the analysis input.

B.2 Overview of Beyond Standard Model Monte Carlo Data Sets

Table B.2: Detailed overview of BSM MC simulation data sets used in this analysis. For cross section and k-factor references see Sec. 8.2.2. All data sets were produced under PU scenario PU_S10 [241] and conditions START53_V7A-v1. The filter efficiency ϵ_{filter} at generator level is 1.0 for all data sets. The stated k-factors are given at NNLO.

Data Set Name	σ / pb	**k-Factor**	$\mathbf{N_{events}}$	$\alpha_{\mathbf{MC}}$
WprimeToENu_M-2000_TuneZ2star_8TeV-pythia6	0.021	1.214	$2 \cdot 10^4$	0.025
WprimeToMuNu_M-2000_TuneZ2star_8TeV-pythia6	0.021	1.214	$2 \cdot 10^4$	0.025
WprimeToTauNu_M-2000_TuneZ2star_8TeV-pythia6-tauola	0.021	1.214	$2 \cdot 10^4$	0.025
WprimeToENu_M-2500_TuneZ2star_8TeV-pythia6	0.0047	1.14	$2 \cdot 10^4$	0.0053
WprimeToMuNu_M-2500_TuneZ2star_8TeV-pythia6	0.0047	1.14	$2 \cdot 10^4$	0.0053
WprimeToTauNu_M-2500_TuneZ2star_8TeV-pythia6-tauola	0.0047	1.14	$2 \cdot 10^4$	0.0053
WprimeToENu_M-3000_TuneZ2star_8TeV-pythia6	0.0014	1.151	$2 \cdot 10^4$	0.0016
WprimeToMuNu_M-3000_TuneZ2star_8TeV-pythia6	0.0013	1.151	$2 \cdot 10^4$	0.0015
WprimeToTauNu_M-3000_TuneZ2star_8TeV-pythia6-tauola	0.0013	1.151	$2 \cdot 10^4$	0.0015
WprimeToENu_M-3500_TuneZ2star_8TeV-pythia6	0.00051	1.242	$2 \cdot 10^4$	0.00062
WprimeToMuNu_M-3500_TuneZ2star_8TeV-pythia6	0.00051	1.242	$2 \cdot 10^4$	0.00062
WprimeToTauNu_M-3500_TuneZ2star_8TeV-pythia6-tauola	0.00051	1.242	$2 \cdot 10^4$	0.00062

B.3 Overview of Standard Model Monte Carlo Data Sets

Please turn to the next page for Tab. B.3.

Table B.3: Detailed overview of SM MC simulation data sets used in this analysis, where * in the column labeled "Conditions" stands for START53. For cross section and k-factor references see Sec. 8.2.1. All data sets were produced under PU scenario PU_S10 [241]. Data sets listed in italic do not pass selection criteria given in Sec. 9.

#	Data Set Name	σ / pb	ϵ_{filter}	k-Factor	N_{events}	α_{MC}	Conditions
1	*G_Pt-0to15_TuneZ2star_8TeV_pythia6*	$9.1 \cdot 10^7$	1.0	1.3 (NLO)	$2.0 \cdot 10^6$	$1.2 \cdot 10^6$	*_V7A-v1
2	*G_Pt-15to30_TuneZ2star_8TeV_pythia6*	$2.0 \cdot 10^5$	1.0	1.3 (NLO)	$1.0 \cdot 10^6$	2600	*_V7A-v1
3	G_Pt-30to50_TuneZ2star_8TeV_pythia6	$2.0 \cdot 10^4$	1.0	1.3 (NLO)	$1.0 \cdot 10^6$	260	*_V7A-v1
4	G_Pt-50to80_TuneZ2star_8TeV_pythia6	3300	1.0	1.3 (NLO)	$1.0 \cdot 10^6$	43	*_V7A-v1
5	G_Pt-80to120_TuneZ2star_8TeV_pythia6	560	1.0	1.3 (NLO)	$1.0 \cdot 10^6$	7.2	*_V7A-v1
6	G_Pt-120to170_TuneZ2star_8TeV_pythia6	110	1.0	1.3 (NLO)	$2.0 \cdot 10^6$	1.4	*_V7A-v1
7	G_Pt-170to300_TuneZ2star_8TeV_pythia6	30	1.0	1.3 (NLO)	$2.0 \cdot 10^6$	0.39	*_V7A-v1
8	G_Pt-300to470_TuneZ2star_8TeV_pythia6	2.1	1.0	1.3 (NLO)	$2.0 \cdot 10^6$	0.027	*_V7A-v1
9	G_Pt-470to800_TuneZ2star_8TeV_pythia6	0.21	1.0	1.3 (NLO)	$1.0 \cdot 10^6$	0.0027	*_V7A-v1
10	G_Pt-800to1400_TuneZ2star_8TeV_pythia6	0.0071	1.0	1.3 (NLO)	$1.0 \cdot 10^6$	$9.2 \cdot 10^{-5}$	*_V7A-v1
11	G_Pt-1400to1800_TuneZ2star_8TeV_pythia6	$4.5 \cdot 10^{-5}$	1.0	1.3 (NLO)	$1.0 \cdot 10^6$	$5.8 \cdot 10^{-7}$	*_V7A-v1
12	G_Pt-1800_TuneZ2star_8TeV_pythia6	$1.9 \cdot 10^{-6}$	1.0	1.3 (NLO)	$1.0 \cdot 10^6$	$2.5 \cdot 10^{-8}$	*_V7A-v1
13	WJetsToLNu_TuneZ2Star_8TeV-madgraph-tarball	$3.0 \cdot 10^4$	1.0	1.21 (NNLO)	$1.0 \cdot 10^7$	39	*_V7A-v1
14	WJetsToLNu_TuneZ2Star_8TeV-madgraph-tarball	$3.0 \cdot 10^4$	1.0	1.21 (NNLO)	$5.0 \cdot 10^7$	13	*_V7A-v2
15	WJetsToLNu_M-200_TuneZ2star_8TeV-madgraph (private FSIM)	35	0.30	1.21 (NNLO)	$3.0 \cdot 10^5$	0.69	*_V19_FSIM
16	WJetsToLNu_M-500_TuneZ2star_8TeV-madgraph (private FSIM)	0.44	0.61	1.21 (NNLO)	$4.0 \cdot 10^5$	0.013	*_V19_FSIM
17	WJetsToLNu_M-1000_TuneZ2star_8TeV-madgraph (private FSIM)	1.0	0.011	1.21 (NNLO)	$5.0 \cdot 10^5$	$5.1 \cdot 10^{-4}$	*_V19_FSIM
18	WJetsToLNu_HT-150To200_8TeV-madgraph	240	1.0	1.21 (NNLO)	$2.0 \cdot 10^7$	0.26	*_V7C-v1
19	WJetsToLNu_HT-200To250_8TeV-madgraph	90	1.0	1.21 (NNLO)	$1.0 \cdot 10^7$	0.21	*_V7C-v1
20	WJetsToLNu_HT-250To300_8TeV-madgraph_v2	48	1.0	1.21 (NNLO)	$4.0 \cdot 10^6$	0.23	*_V7A-v1
21	WJetsToLNu_HT-300To400_8TeV-madgraph_v2	38	1.0	1.21 (NNLO)	$5.0 \cdot 10^6$	0.18	*_V7A-v1
22	WJetsToLNu_HT-400ToInf_8TeV-madgraph_v2	25	1.0	1.21 (NNLO)	$4.0 \cdot 10^6$	0.12	*_V7A-v1
23	WbbJetsToLNu_Massive_TuneZ2star_8TeV-madgraph-pythia6_tauola	210	1.0	1.79 (NLO)	$2.0 \cdot 10^7$	0.20	*_V7A-v1
24	DYJetsToLL_M-10To50_TuneZ2Star_8TeV-madgraph	$1.1 \cdot 10^4$	1.0	1.2 (NNLO)	$3.0 \cdot 10^7$	6.9	*_V7A-v1
25	DYJetsToLL_M-50_TuneZ2Star_8TeV-madgraph-tarball	3000	1.0	1.2 (NNLO)	$3.0 \cdot 10^7$	2.3	*_V7A-v1
26	DYJetsToLL_PtZ-50To70_TuneZ2star_8TeV_ext-madgraph-tarball	89	1.0	1.0 (LO)	$1.0 \cdot 10^7$	0.11	*_V7C-v1
27	DYJetsToLL_PtZ-70To100_TuneZ2star_8TeV_ext-madgraph-tarball	53	1.0	1.0 (LO)	$1.0 \cdot 10^7$	0.089	*_V7C-v1
28	DYJetsToLL_PtZ-100_TuneZ2star_8TeV_ext-madgraph-tarball	33	1.0	1.0 (LO)	$1.0 \cdot 10^7$	0.052	*_V7C-v1
29	DYJetsToLL_PtZ-180_TuneZ2star_8TeV-madgraph-tarball	4.6	1.0	1.0 (LO)	$1.0 \cdot 10^6$	0.058	*_V7C-v1
30	ZJetsToNuNu_PtZ-70To100_8TeV	33	1.0	1.0 (LO)	$2.0 \cdot 10^7$	0.030	*_V7C-v1
31	ZJetsToNuNu_PtZ-100_8TeV-madgraph	21	1.0	1.0 (LO)	$1.0 \cdot 10^7$	0.030	*_V7C-v1

Continued on next page...

Table B.3: *...continued from previous page.*

#	Data Set Name	σ / pb	ϵ_{filter}	k-Factor	N_{events}	α_{MC}	Conditions
32	DYToEE_M-200_CT10_TuneZ2star_8TeV-powheg-pythia6	1.5	1.0	1.02 (NNLO)	$9.0 \cdot 10^4$	0.30	*_V7A-v1
33	DYToEE_M-400_CT10_TuneZ2star_8TeV-powheg-pythia6	0.11	1.0	1.02 (NNLO)	$9.0 \cdot 10^4$	0.022	*_V7A-v1
34	DYToEE_M-500_CT10_TuneZ2star_8TeV-powheg-pythia6	0.044	1.0	1.02 (NNLO)	$9.0 \cdot 10^4$	0.0089	*_V7A-v1
35	DYToEE_M-700_CT10_TuneZ2star_8TeV-powheg-pythia6	0.010	1.0	1.02 (NNLO)	$9.0 \cdot 10^4$	0.0021	*_V7A-v1
36	DYToEE_M-800_CT10_TuneZ2star_8TeV-powheg-pythia6	0.0055	1.0	1.02 (NNLO)	$9.0 \cdot 10^4$	0.0011	*_V7A-v1
37	DYToEE_M-1000_CT10_TuneZ2star_8TeV-powheg-pythia6	0.0018	1.0	1.02 (NNLO)	$9.0 \cdot 10^4$	$3.6 \cdot 10^{-4}$	*_V7A-v1
38	DYToEE_M-1500_CT10_TuneZ2star_8TeV-powheg-pythia6	$1.7 \cdot 10^{-4}$	1.0	1.02 (NNLO)	$9.0 \cdot 10^4$	$3.4 \cdot 10^{-5}$	*_V7A-v1
39	DYToEE_M-2000_CT10_TuneZ2star_8TeV-powheg-pythia6	$2.2 \cdot 10^{-5}$	1.0	1.02 (NNLO)	$9.0 \cdot 10^4$	$4.5 \cdot 10^{-6}$	*_V7A-v1
40	DYToMuMu_M-200_CT10_TuneZ2star_8TeV-powheg-pythia6	1.5	1.0	1.02 (NNLO)	$9.0 \cdot 10^4$	0.30	*_V7C1-v1
41	DYToMuMu_M-400_CT10_TuneZ2star_8TeV-powheg-pythia6	0.11	1.0	1.02 (NNLO)	$9.0 \cdot 10^4$	0.022	*_V7C1-v1
42	DYToMuMu_M-500_CT10_TuneZ2star_8TeV-powheg-pythia6	0.044	1.0	1.02 (NNLO)	$9.0 \cdot 10^4$	0.0089	*_V7C1-v1
43	DYToMuMu_M-700_CT10_TuneZ2star_8TeV-powheg-pythia6	0.010	1.0	1.02 (NNLO)	$9.0 \cdot 10^4$	0.0023	*_V7C1-v1
44	DYToMuMu_M-800_CT10_TuneZ2star_8TeV-powheg-pythia6	0.0055	1.0	1.02 (NNLO)	$9.0 \cdot 10^4$	0.0011	*_V7C1-v1
45	DYToMuMu_M-1000_CT10_TuneZ2star_8TeV-powheg-pythia6	0.0018	1.0	1.02 (NNLO)	$9.0 \cdot 10^4$	$3.6 \cdot 10^{-4}$	*_V7C1-v1
46	DYToMuMu_M-1500_CT10_TuneZ2star_8TeV-powheg-pythia6	$1.7 \cdot 10^{-4}$	1.0	1.02 (NNLO)	$9.0 \cdot 10^4$	$3.4 \cdot 10^{-5}$	*_V7C1-v1
47	DYToMuMu_M-2000_CT10_TuneZ2star_8TeV-powheg-pythia6	$2.2 \cdot 10^{-5}$	1.0	1.02 (NNLO)	$9.0 \cdot 10^4$	$4.5 \cdot 10^{-6}$	*_V7C1-v1
48	GluGluToHToBB_M-125_8TeV-powheg-pythia6	11	1.0	1.0 (NLO)	$9.0 \cdot 10^5$	0.22	*_V7A-v1
49	GluGluToHToGG_M-125_8TeV-powheg-pythia6	0.044	1.0	1.0 (NLO)	$9.0 \cdot 10^4$	0.0090	*_V7A-v1
50	GluGluToHToTauTau_M-125_8TeV-powheg-pythia6	1.2	1.0	1.0 (NLO)	$9.0 \cdot 10^5$	0.025	*_V7A-v1
51	GluGluToHToWWTo2LAndTau2Nu_M-125_8TeV-powheg-pythia6	0.43	1.0	1.0 (NLO)	$2.0 \cdot 10^5$	0.029	*_V7A-v1
52	GluGluToHToWWToLAndTauNuQQ_M-125_8TeV-powheg-pythia6	1.8	1.0	1.0 (NLO)	$4.0 \cdot 10^5$	0.075	*_V19-v1
53	GluGluToHToZG_M-125_8TeV-powheg-pythia6	0.030	1.0	1.0 (NLO)	$9.0 \cdot 10^4$	0.0058	*_V7A-v1
54	GluGluToHToZZTo2L2Nu_M-125_8TeV-minloHJJ-pythia6-tauola	0.010	1.0	1.0 (NLO)	$3.0 \cdot 10^5$	$5.1 \cdot 10^{-4}$	*_V19-v1
55	GluGluToHToZZTo2L2Q_M-125_8TeV-minloHJJ-pythia6-tauola	0.0036	1.0	1.0 (NLO)	$2.0 \cdot 10^5$	$2.5 \cdot 10^{-4}$	*_V19-v1
56	GluGluToHToZZTo4L_M-125_8TeV-minloHJJ-pythia6-tauola	0.010	1.0	1.0 (NLO)	$3.0 \cdot 10^5$	$5.7 \cdot 10^{-4}$	*_V19-v1
57	VBFToHToWWToLAndTauNuQQ_M-125_8TeV-powheg-pythia6	0.15	1.0	1.0 (NLO)	$2.0 \cdot 10^5$	0.0098	*_V19-v1
58	VBF_HToBB_M-125_8TeV-powheg-pythia8	0.91	1.0	1.0 (NLO)	$9.0 \cdot 10^5$	0.018	*_V7C-v1
59	VBF_HToGG_M-125_8TeV-powheg-pythia6	0.0036	1.0	1.0 (NLO)	$9.0 \cdot 10^4$	$7.1 \cdot 10^{-4}$	*_V7A-v1
60	VBF_HToTauTau_M-125_8TeV-powheg-pythia6	0.100	1.0	1.0 (NLO)	$9.0 \cdot 10^5$	0.0020	*_V7A-v1
61	VBF_HToWWTo2LAndTau2Nu_M-125_8TeV-powheg-pythia6	0.036	1.0	1.0 (NLO)	$2.0 \cdot 10^5$	0.0023	*_V7A-v1
62	VBF_HToZG_M-125_8TeV-powheg-pythia6	0.0024	1.0	1.0 (NLO)	$9.0 \cdot 10^4$	$4.8 \cdot 10^{-4}$	*_V7A-v1
63	VBF_HToZZTo2L2Q_M-125_8TeV-powheg-pythia6	$2.9 \cdot 10^{-4}$	1.0	1.0 (NLO)	$4.0 \cdot 10^4$	$1.2 \cdot 10^{-4}$	*_V7A-v1
64	VBF_HToZZTo4L_M-125_8TeV-powheg-pythia6	$8.5 \cdot 10^{-4}$	1.0	1.0 (NLO)	$4.0 \cdot 10^4$	$3.4 \cdot 10^{-4}$	*_V7A-v1
65	WH_HToMuMu_M-125_8TeV-powheg-herwigpp	$1.5 \cdot 10^{-4}$	1.0	1.0 (NLO)	$1.0 \cdot 10^5$	$3.0 \cdot 10^{-5}$	*_V19-v1
66	WH_WToLNu_HToBB_M-125_8TeV-powheg-herwigpp	0.13	1.0	1.0 (NLO)	$1.0 \cdot 10^6$	0.0026	*_V7A-v1

Continued on next page...

Table B.3: *...continued from previous page.*

#	Data Set Name	σ / pb	ϵ_{filter}	k-Factor	N_{events}	α_{MC}	Conditions
67	ZH_HToMuMu_M-125_8TeV-powheg-herwigpp	$9.1 \cdot 10^{-5}$	1.0	1.0 (NLO)	$1.0 \cdot 10^{5}$	$1.8 \cdot 10^{-5}$	*_V19-v1
68	ZH_ZToBB_HToBB_M-125_8TeV-powheg-herwigpp	0.036	1.0	1.0 (NLO)	$9.0 \cdot 10^{5}$	$7.1 \cdot 10^{-4}$	*_V7A-v1
69	ZH_ZToLL_HToBB_M-125_8TeV-powheg-herwigpp	0.024	1.0	1.0 (NLO)	$9.0 \cdot 10^{5}$	$4.7 \cdot 10^{-4}$	*_V7A-v1
70	ZH_ZToNuNu_HToBB_M-125_8TeV-powheg-herwigpp	0.048	1.0	1.0 (NLO)	$1.0 \cdot 10^{6}$	$9.4 \cdot 10^{-4}$	*_V7A-v1
71	TTH_HToBB_M-125_8TeV-pythia6	0.075	1.0	1.0 (NLO)	$1.0 \cdot 10^{6}$	0.0015	*_V7A-v1
72	TTH_HToGG_M-125_8TeV-pythia6	$2.9 \cdot 10^{-4}$	1.0	1.0 (NLO)	$1.0 \cdot 10^{5}$	$5.8 \cdot 10^{-5}$	*_V7A-v1
73	TTH_HToZG_M-125_8TeV-pythia6	$2.0 \cdot 10^{-4}$	1.0	1.0 (NLO)	$1.0 \cdot 10^{5}$	$3.9 \cdot 10^{-5}$	*_V7A-v1
74	WH_ZH_HToGG_M-125_8TeV-pythia6	0.0026	1.0	1.0 (NLO)	$1.0 \cdot 10^{5}$	$5.0 \cdot 10^{-4}$	*_V7A-v1
75	WH_ZH_HToZG_M-125_8TeV-pythia6	0.0017	1.0	1.0 (NLO)	$1.0 \cdot 10^{6}$	$3.4 \cdot 10^{-5}$	*_V7A-v1
76	WH_ZH_TTH_HToTauTau_M-125_8TeV-pythia6-tauola	0.079	1.0	1.0 (NLO)	$2.0 \cdot 10^{5}$	0.0078	*_V7A-v1
77	WH_ZH_TTH_HToWW_M-125_8TeV-pythia6	0.27	1.0	1.0 (NLO)	$2.0 \cdot 10^{5}$	0.026	*_V7A-v1
78	WH_ZH_TTH_HToZZ_M-125_8TeV-pythia6	0.033	1.0	1.0 (NLO)	$5.0 \cdot 10^{5}$	0.0013	*_V7A-v1
79	DiPhotonBorn_Pt-10To25_8TeV-pythia6	270	1.0	1.0 (LO)	$5.0 \cdot 10^{5}$	11	*_V7A-v1
80	DiPhotonBorn_Pt-10To25_8TeV_ext-pythia6	270	1.0	1.0 (LO)	$1.0 \cdot 10^{6}$	5.3	*_V7A-v1
81	DiPhotonBorn_Pt-25To250_8TeV-pythia6	25	1.0	1.0 (LO)	$5.0 \cdot 10^{5}$	1.0	*_V7A-v1
82	DiPhotonBorn_Pt-25To250_8TeV_ext-pythia6	25	1.0	1.0 (LO)	$1.0 \cdot 10^{6}$	0.50	*_V7A-v1
83	DiPhotonBorn_Pt-250ToInf_8TeV-pythia6	0.011	1.0	1.0 (LO)	$5.0 \cdot 10^{5}$	$4.3 \cdot 10^{-4}$	*_V7A-v1
84	DiPhotonBorn_Pt-250ToInf_8TeV_ext-pythia6	0.011	1.0	1.0 (LO)	$1.0 \cdot 10^{6}$	$2.1 \cdot 10^{-4}$	*_V7A-v1
85	DiPhotonBox_Pt-10To25_8TeV-pythia6	430	1.0	1.0 (LO)	$5.0 \cdot 10^{5}$	17	*_V7A-v1
86	DiPhotonBox_Pt-25To250_8TeV-pythia6	16	1.0	1.0 (LO)	$5.0 \cdot 10^{5}$	0.61	*_V7A-v1
87	DiPhotonBox_Pt-25To250_8TeV_ext-pythia6	16	1.0	1.0 (LO)	$8.0 \cdot 10^{5}$	0.37	*_V7A-v1
88	DiPhotonBox_Pt-250ToInf_8TeV-pythia6	$3.2 \cdot 10^{-4}$	1.0	1.0 (LO)	$5.0 \cdot 10^{5}$	$1.3 \cdot 10^{-5}$	*_V7A-v1
89	DiPhotonBox_Pt-250ToInf_8TeV_ext-pythia6	$3.2 \cdot 10^{-4}$	1.0	1.0 (LO)	$9.0 \cdot 10^{5}$	$6.5 \cdot 10^{-6}$	*_V7A-v1
90	WW_TuneZ2star_8TeV_pythia6_tauola	34	1.0	1.63 (NLO)	$1.0 \cdot 10^{7}$	0.11	*_V7A-v1
91	WpWpqq_8TeV-madgraph	0.25	1.0	1.0 (LO)	$9.0 \cdot 10^{4}$	0.049	*_V7A-v1
92	WmWmqq_8TeV-madgraph	0.089	1.0	1.0 (LO)	$9.0 \cdot 10^{4}$	0.018	*_V7A-v1
93	GluGluToWWTo4L_TuneZ2star_8TeV-gg2ww-pythia6	0.18	1.0	1.4 (NLO)	$1.0 \cdot 10^{5}$	0.045	*_V7A-v1
94	ZZTo2e2mu_8TeV-powheg-pythia6	0.18	1.0	1.0 (NLO)	$1.0 \cdot 10^{6}$	0.0023	*_V7A-v1
95	ZZTo2e2mu_8TeV_ext-powheg-pythia6	0.18	1.0	1.0 (NLO)	$3.0 \cdot 10^{6}$	$8.7 \cdot 10^{-4}$	*_V7C-v1
96	ZZTo2mu2tau_8TeV-powheg-pythia6	0.18	1.0	1.0 (NLO)	$8.0 \cdot 10^{5}$	0.0042	*_V7A-v1
97	ZZTo2mu2tau_8TeV_ext-powheg-pythia6	0.18	1.0	1.0 (NLO)	$3.0 \cdot 10^{6}$	$8.7 \cdot 10^{-4}$	*_V7C-v1
98	ZZTo2e2tau_8TeV-powheg-pythia6	0.18	1.0	1.0 (NLO)	$8.0 \cdot 10^{5}$	0.0042	*_V7A-v1
99	ZZTo2e2tau_8TeV_ext-powheg-pythia6	0.18	1.0	1.0 (NLO)	$1.0 \cdot 10^{6}$	0.0017	*_V7C-v1
100	ZZTo4e_8TeV-powheg-pythia6	0.077	1.0	1.0 (NLO)	$1.0 \cdot 10^{6}$	0.0010	*_V7A-v1
101	ZZTo4e_8TeV_ext-powheg-pythia6	0.077	1.0	1.0 (NLO)	$1.0 \cdot 10^{6}$	$7.6 \cdot 10^{-4}$	*_V7C-v1

Continued on next page...

Table B.3: *...continued from previous page.*

#	Data Set Name	σ / pb	ϵ_{filter}	k-Factor	N_{events}	α_{MC}	Conditions
102	ZZTo4mu_8TeV-powheg-pythia6	0.077	1.0	1.0 (NLO)	$1.0 \cdot 10^6$	0.0010	*_V7A-v1
103	ZZTo4mu_8TeV_ext-powheg-pythia6	0.077	1.0	1.0 (NLO)	$1.0 \cdot 10^6$	$7.6 \cdot 10^{-4}$	*_V7C-v1
104	ZZTo4tau_8TeV-powheg-pythia6	0.077	1.0	1.0 (NLO)	$8.0 \cdot 10^5$	0.0018	*_V7A-v1
105	ZZTo4tau_8TeV_ext-powheg-pythia6	0.077	1.0	1.0 (NLO)	$1.0 \cdot 10^6$	$7.6 \cdot 10^{-4}$	*_V7C-v1
106	ZZJetsTo2L2Nu_TuneZ2star_8TeV-madgraph-tauola	0.28	1.0	1.0 (LO)	$9.0 \cdot 10^5$	0.0058	*_V7A-v3
107	ZZJetsTo2L2Q_TuneZ2star_8TeV-madgraph-tauola	0.91	1.0	1.0 (LO)	$1.0 \cdot 10^6$	0.0093	*_V7A-v1
108	ZZJetsTo2Q2Nu_TuneZ2star_8TeV-madgraph-tauloa	0.49	1.0	1.0 (LO)	$8.0 \cdot 10^5$	0.012	*_V7A-v1
109	GluGluToZZTo2L2L_TuneZ2star_8TeV-gg2zz-pythia6	0.012	1.0	1.0 (NLO)	$4.0 \cdot 10^5$	$5.9 \cdot 10^{-4}$	*_V7A-v1
110	GluGluToZZTo4L_8TeV-gg2zz-pythia6	0.0048	1.0	1.0 (NLO)	$5.0 \cdot 10^5$	$1.7 \cdot 10^{-4}$	*_V7A-v1
111	WZJetsTo2L2Q_TuneZ2star_8TeV-madgraph-tauola	1.8	1.0	1.0 (LO)	$3.0 \cdot 10^6$	0.011	*_V7A-v1
112	WZJetsTo2Q2Nu_TuneZ2star_8TeV-madgraph-tauloa	0.70	1.0	1.0 (LO)	$9.0 \cdot 10^5$	0.015	*_V7A-v1
113	WZJetsTo2QLNu_8TeV-madgraph	3.1	1.0	1.0 (LO)	$2.0 \cdot 10^6$	0.021	*_V7C-v1
114	WZJetsTo3LNu_TuneZ2_8TeV-madgraph-tauola	0.87	1.0	1.0 (LO)	$2.0 \cdot 10^6$	0.0085	*_V7A-v1
115	WGToLNuG_TuneZ2star_8TeV-madgraph-tauola	460	1.0	1.2 (NLO)	$4.0 \cdot 10^6$	2.3	*_V7A-v1
116	ZGToLLG_8TeV-madgraph	130	1.0	1.2 (NLO)	$6.0 \cdot 10^6$	0.48	*_V7A-v1
117	WWWJets_8TeV-madgraph	0.082	1.0	0.98 (NLO)	$2.0 \cdot 10^5$	0.0072	*_V7A-v1
118	WWZNoGstarJets_8TeV-madgraph	0.063	1.0	0.92 (NLO)	$2.0 \cdot 10^5$	0.0052	*_V7A-v1
119	WZZNoGstarJets_8TeV-madgraph	0.019	1.0	1.02 (NLO)	$2.0 \cdot 10^5$	0.0018	*_V7A-v1
120	ZZZNoGstarJets_8TeV-madgraph	0.0046	1.0	1.2 (NLO)	$2.0 \cdot 10^5$	$5.0 \cdot 10^{-4}$	*_V7A-v1
121	WWGJets_8TeV-madgraph	0.53	1.0	1.0 (LO)	$2.0 \cdot 10^5$	0.048	*_V7A-v1
122	WWGJets_8TeV-madgraph_v2	0.53	1.0	1.0 (LO)	$3.0 \cdot 10^5$	0.034	*_V7A-v1
123	LNuGG_enhanced_FSR_8TeV_madgraph	1.8	1.0	1.0 (LO)	$1.0 \cdot 10^6$	0.036	*_V19-v1
124	LNuGG_enhanced_ISR_8TeV_madgraph	0.32	1.0	1.0 (LO)	$1.0 \cdot 10^6$	0.0063	*_V19-v1
125	WW_DoubleScattering_8TeV-pythia8	0.59	1.0	1.0 (LO)	$8.0 \cdot 10^5$	0.014	*_V7A-v1
126	*QCD_Pt-0to5_TuneZ2star_8TeV_pythia6*	$5.0 \cdot 10^{10}$	0.98	1.0 (LO)	$9.0 \cdot 10^5$	$9.6 \cdot 10^8$	*_V7A-v1
127	*QCD_Pt-5to15_TuneZ2star_8TeV_pythia6*	$4.3 \cdot 10^{10}$	1.0	1.0 (LO)	$1.0 \cdot 10^6$	$5.6 \cdot 10^8$	*_V7A-v1
128	*QCD_Pt-15to30_TuneZ2star_8TeV_pythia6*	$9.9 \cdot 10^8$	1.0	1.0 (LO)	$9.0 \cdot 10^6$	$1.9 \cdot 10^6$	*_V7A-v1
129	*QCD_Pt-15to30_TuneZ2star_8TeV_pythia6*	$9.9 \cdot 10^8$	1.0	1.0 (LO)	$3.0 \cdot 10^5$	$5.4 \cdot 10^7$	*_V7A-v2
130	*QCD_Pt-30to50_TuneZ2star_8TeV_pythia6*	$6.6 \cdot 10^7$	1.0	1.0 (LO)	$4.0 \cdot 10^5$	$3.2 \cdot 10^6$	*_V7A-v1
131	*QCD_Pt-30to50_TuneZ2star_8TeV_pythia6*	$6.6 \cdot 10^7$	1.0	1.0 (LO)	$6.0 \cdot 10^6$	$2.2 \cdot 10^5$	*_V7A-v2
132	*QCD_Pt-50to80_TuneZ2star_8TeV_pythia6*	$8.1 \cdot 10^6$	1.0	1.0 (LO)	$3.0 \cdot 10^5$	$4.2 \cdot 10^5$	*_V7A-v1
133	*QCD_Pt-50to80_TuneZ2star_8TeV_pythia6*	$8.1 \cdot 10^6$	1.0	1.0 (LO)	$5.0 \cdot 10^6$	$2.7 \cdot 10^4$	*_V7A-v2
134	QCD_Pt-80to120_TuneZ2star_8TeV_pythia6	$1.0 \cdot 10^6$	1.0	1.0 (LO)	$8.0 \cdot 10^5$	$2.4 \cdot 10^4$	*_V7A-v2
135	QCD_Pt-80to120_TuneZ2star_8TeV_pythia6	$1.0 \cdot 10^6$	1.0	1.0 (LO)	$5.0 \cdot 10^6$	3400	*_V7A-v3
136	QCD_Pt-120to170_TuneZ2star_8TeV_pythia6	$1.6 \cdot 10^5$	1.0	1.0 (LO)	$8.0 \cdot 10^5$	3500	*_V7A-v2

Continued on next page...

Table B.3: *...continued from previous page.*

#	Data Set Name	σ / pb	ϵ_{filter}	k-Factor	N_{events}	α_{MC}	Conditions
137	QCD_Pt-120to170_TuneZ2star_8TeV_pythia6	$1.6 \cdot 10^5$	1.0	1.0 (LO)	$5.0 \cdot 10^6$	510	*_V7A-v3
138	QCD_Pt-170to300_TuneZ2star_8TeV_pythia6	$3.4 \cdot 10^4$	1.0	1.0 (LO)	$2.0 \cdot 10^5$	2300	*_V7A-v1
140	QCD_Pt-170to300_TuneZ2star_8TeV_pythia6	$3.4 \cdot 10^4$	1.0	1.0 (LO)	$5.0 \cdot 10^6$	120	*_V7A-v2
141	QCD_Pt-170to300_TuneZ2star_8TeV_pythia6_v2	$3.4 \cdot 10^4$	1.0	1.0 (LO)	$1.0 \cdot 10^7$	34	*_V7A-v1
142	QCD_Pt-300to470_TuneZ2star_8TeV_pythia6	1800	1.0	1.0 (LO)	$3.0 \cdot 10^5$	110	*_V7A-v1
143	QCD_Pt-300to470_TuneZ2star_8TeV_pythia6	1800	1.0	1.0 (LO)	$5.0 \cdot 10^6$	5.8	*_V7A-v2
144	QCD_Pt-300to470_TuneZ2star_8TeV_pythia6_v2	1800	1.0	1.0 (LO)	$3.0 \cdot 10^6$	9.9	*_V7A-v1
145	QCD_Pt-300to470_TuneZ2star_8TeV_pythia6_v3	1800	1.0	1.0 (LO)	$1.0 \cdot 10^7$	1.7	*_V7A-v1
146	QCD_Pt-470to600_TuneZ2star_8TeV_pythia6	110	1.0	1.0 (LO)	$3.0 \cdot 10^5$	6.8	*_V7A-v1
147	QCD_Pt-470to600_TuneZ2star_8TeV_pythia6	110	1.0	1.0 (LO)	$3.0 \cdot 10^6$	0.56	*_V7A-v2
148	QCD_Pt-600to800_TuneZ2star_8TeV_pythia6	27	1.0	1.0 (LO)	$2.0 \cdot 10^5$	2.0	*_V7A-v1
149	QCD_Pt-600to800_TuneZ2star_8TeV_pythia6	27	1.0	1.0 (LO)	$3.0 \cdot 10^6$	0.13	*_V7A-v2
150	QCD_Pt-800to1000_TuneZ2star_8TeV_pythia6	3.6	1.0	1.0 (LO)	$3.0 \cdot 10^5$	0.22	*_V7A-v1
151	QCD_Pt-800to1000_TuneZ2star_8TeV_pythia6	3.6	1.0	1.0 (LO)	$3.0 \cdot 10^6$	0.017	*_V7A-v2
152	QCD_Pt-1000to1400_TuneZ2star_8TeV_pythia6	0.74	1.0	1.0 (LO)	$1.0 \cdot 10^6$	0.0074	*_V7A-v1
153	QCD_Pt-1400to1800_TuneZ2star_8TeV_pythia6	0.034	1.0	1.0 (LO)	$2.0 \cdot 10^6$	$3.3 \cdot 10^{-4}$	*_V7A-v1
154	QCD_Pt-1800_TuneZ2star_8TeV_pythia6	0.0018	1.0	1.0 (LO)	$9.0 \cdot 10^5$	$3.7 \cdot 10^{-5}$	*_V7A-v1
155	*QCD_Pt-15to20_MuEnrichedPt5_TuneZ2star_8TeV_pythia6*	$7.0 \cdot 10^8$	0.0039	1.0 (LO)	$1.0 \cdot 10^6$	$3.1 \cdot 10^4$	*_V7A-v2
156	*QCD_Pt-20to30_MuEnrichedPt5_TuneZ2star_8TeV_pythia6*	$2.9 \cdot 10^8$	0.0065	1.0 (LO)	$8.0 \cdot 10^6$	4300	*_V7A-v1
157	*QCD_Pt-30to50_MuEnrichedPt5_TuneZ2star_8TeV_pythia6*	$6.6 \cdot 10^7$	0.012	1.0 (LO)	$9.0 \cdot 10^6$	1700	*_V7A-v1
158	*QCD_Pt-50to80_MuEnrichedPt5_TuneZ2star_8TeV_pythia6*	$8.1 \cdot 10^6$	0.022	1.0 (LO)	$1.0 \cdot 10^7$	330	*_V7A-v1
159	QCD_Pt-80to120_MuEnrichedPt5_TuneZ2star_8TeV_pythia6	$1.0 \cdot 10^6$	0.040	1.0 (LO)	$9.0 \cdot 10^6$	86	*_V7A-v1
160	QCD_Pt-120to170_MuEnrichedPt5_TuneZ2star_8TeV_pythia6	$1.6 \cdot 10^5$	0.047	1.0 (LO)	$8.0 \cdot 10^6$	17	*_V7A-v1
161	QCD_Pt-170to300_MuEnrichedPt5_TuneZ2star_8TeV_pythia6	$3.4 \cdot 10^4$	0.068	1.0 (LO)	$7.0 \cdot 10^6$	5.9	*_V7A-v1
162	QCD_Pt-300to470_MuEnrichedPt5_TuneZ2star_8TeV_pythia6	1800	0.086	1.0 (LO)	$7.0 \cdot 10^6$	0.38	*_V7A-v1
163	QCD_Pt-470to600_MuEnrichedPt5_TuneZ2star_8TeV_pythia6	120	0.10	1.0 (LO)	$3.0 \cdot 10^6$	0.061	*_V7A-v1
164	QCD_Pt-600to800_MuEnrichedPt5_TuneZ2star_8TeV_pythia6	27	0.100	1.0 (LO)	$4.0 \cdot 10^6$	0.013	*_V7A-v1
165	QCD_Pt-800to1000_MuEnrichedPt5_TuneZ2star_8TeV_pythia6	3.6	0.10	1.0 (LO)	$4.0 \cdot 10^6$	0.0018	*_V7A-v1
166	QCD_Pt-1000_MuEnrichedPt5_TuneZ2star_8TeV_pythia6	0.77	0.11	1.0 (LO)	$3.0 \cdot 10^6$	$4.3 \cdot 10^{-4}$	*_V7A-v1
167	QCD_Pt_20_MuEnrichedPt_15_TuneZ2star_8TeV_pythia6	$3.6 \cdot 10^8$	$3.7 \cdot 10^{-4}$	1.0 (LO)	$7.0 \cdot 10^6$	350	*_V7A-v1
168	QCD_Pt_20_MuEnrichedPt_15_TuneZ2star_8TeV_pythia6	$3.6 \cdot 10^8$	$3.7 \cdot 10^{-4}$	1.0 (LO)	$2.0 \cdot 10^7$	120	*_V7A-v3
169	*QCD_Pt-15To30_bEnriched_TuneZ2star_8TeV-pythia6-evtgen*	$10.0 \cdot 10^8$	0.064	1.0 (LO)	$4.0 \cdot 10^6$	$2.7 \cdot 10^5$	*_V7A-v1
170	QCD_Pt-30To50_bEnriched_TuneZ2star_8TeV-pythia6-evtgen	$6.7 \cdot 10^7$	0.081	1.0 (LO)	$4.0 \cdot 10^6$	$2.2 \cdot 10^4$	*_V7A-v1
171	QCD_Pt-50To150_bEnriched_TuneZ2star_8TeV-pythia6-evtgen	$9.4 \cdot 10^6$	0.096	1.0 (LO)	$2.0 \cdot 10^6$	7100	*_V7A-v1
172	QCD_Pt-150_bEnriched_TuneZ2star_8TeV-pythia6-evtgen	$6.7 \cdot 10^4$	0.13	1.0 (LO)	$4.0 \cdot 10^5$	350	*_V7A-v1

Continued on next page...

Table B.3: *...continued from previous page.*

#	Data Set Name	σ / pb	ϵ_{filter}	k-Factor	N_{events}	α_{MC}	Conditions
173	QCD_Pt_20_30_EMEnriched_TuneZ2star_8TeV_pythia6	$2.9 \cdot 10^{8}$	0.010	1.0 (LO)	$3.0 \cdot 10^{7}$	1600	*_V7A-v1
174	QCD_Pt_30_80_EMEnriched_TuneZ2star_8TeV_pythia6	$7.4 \cdot 10^{7}$	0.062	1.0 (LO)	$3.0 \cdot 10^{7}$	2700	*_V7A-v1
175	QCD_Pt_80_170_EMEnriched_TuneZ2star_8TeV_pythia6	$1.2 \cdot 10^{6}$	0.15	1.0 (LO)	$3.0 \cdot 10^{7}$	100	*_V7A-v1
176	QCD_Pt_170_250_EMEnriched_TuneZ2star_8TeV_pythia6	$3.1 \cdot 10^{4}$	0.15	1.0 (LO)	$3.0 \cdot 10^{7}$	2.9	*_V7A-v1
177	QCD_Pt_250_350_EMEnriched_TuneZ2star_8TeV_pythia6	4300	0.13	1.0 (LO)	$3.0 \cdot 10^{7}$	0.32	*_V7A-v1
178	QCD_Pt_350_EMEnriched_TuneZ2star_8TeV_pythia6	810	0.11	1.0 (LO)	$3.0 \cdot 10^{7}$	0.052	*_V7A-v1
179	*QCD_Pt_20_30_BCtoE_TuneZ2star_8TeV_pythia6*	$2.9 \cdot 10^{8}$	$5.8 \cdot 10^{-4}$	1.0 (LO)	$1.0 \cdot 10^{6}$	1900	*_V7A-v1
180	QCD_Pt_30_80_BCtoE_TuneZ2star_8TeV_pythia6	$7.4 \cdot 10^{7}$	0.0022	1.0 (LO)	$2.0 \cdot 10^{6}$	1600	*_V7A-v1
181	QCD_Pt_80_170_BCtoE_TuneZ2star_8TeV_pythia6	$1.2 \cdot 10^{6}$	0.011	1.0 (LO)	$1.0 \cdot 10^{6}$	130	*_V7A-v1
182	QCD_Pt_170_250_BCtoE_TuneZ2star_8TeV_pythia6	$3.1 \cdot 10^{4}$	0.020	1.0 (LO)	$1.0 \cdot 10^{6}$	6.4	*_V7A-v1
183	QCD_Pt_250_350_BCtoE_TuneZ2star_8TeV_pythia6	4300	0.024	1.0 (LO)	$2.0 \cdot 10^{6}$	1.0	*_V7A-v1
184	QCD_Pt_350_BCtoE_TuneZ2star_8TeV_pythia6	810	0.029	1.0 (LO)	$1.0 \cdot 10^{6}$	0.24	*_V7A-v2
185	T_s-channel_TuneZ2star_8TeV-powheg-tauola	2.8	1.0	1.34 (NNLL)	$2.0 \cdot 10^{5}$	0.29	*_V7A-v1
186	T_t-channel_TuneZ2star_8TeV-powheg-tauola	47	0.98	1.22 (NNLL)	$3.0 \cdot 10^{6}$	0.29	*_V7A-v1
187	T_t-channel_TuneZ2star_8TeV-powheg-tauola	47	0.98	1.22 (NNLL)	$2.0 \cdot 10^{4}$	47	*_V7A-v2
188	T_t-channel_TuneZ2star_8TeV-powheg-tauola	47	0.98	1.22 (NNLL)	$9.0 \cdot 10^{4}$	11	*_V7A-v3
189	T_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola	11	1.00	1.04 (NNLL)	$4.0 \cdot 10^{5}$	0.44	*_V7A-v1
190	Tbar_s-channel_TuneZ2star_8TeV-powheg-tauola	1.6	1.0	1.12 (NNLL)	$1.0 \cdot 10^{5}$	0.25	*_V7A-v1
191	Tbar_t-channel_TuneZ2star_8TeV-powheg-tauola	25	0.98	1.25 (NNLL)	$1.0 \cdot 10^{6}$	0.31	*_V7A-v1
192	Tbar_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola	11	1.00	1.04 (NNLL)	$4.0 \cdot 10^{5}$	0.44	*_V7A-v1
193	TBZToLL_4F_TuneZ2star_8TeV-madgraph-tauola	0.022	1.0	0.53 (NLO)	$1.0 \cdot 10^{5}$	0.0015	*_V7C-v1
194	TT_CT10_TuneZ2star_8TeV-powheg-tauola	250	1.0	1.0 (NNLO)	$6.0 \cdot 10^{6}$	0.75	*_V7A-v1
195	TT_CT10_TuneZ2star_8TeV-powheg-tauola	250	1.0	1.0 (NNLO)	$2.0 \cdot 10^{7}$	0.22	*_V7A-v2
196	TT_Mtt-700to1000_CT10_TuneZ2star_8TeV-powheg-tauola	210	0.074	1.11 (NNLO)	$3.0 \cdot 10^{6}$	0.11	*_V7A-v1
197	TT_Mtt-1000toInf_CT10_TuneZ2star_8TeV-powheg-tauola	3.0	0.014	1.11 (NNLO)	$1.0 \cdot 10^{6}$	$7.2 \cdot 10^{-4}$	*_V7A-v1
198	TTGJets_8TeV-madgraph	1.4	1.0	1.0 (LO)	$1.0 \cdot 10^{6}$	0.017	*_V19-v1
199	TTWJets_8TeV-madgraph	0.21	1.0	1.08 (NLO)	$1.0 \cdot 10^{5}$	0.023	*_V7A-v1
200	TTZJets_8TeV-madgraph_v2	0.17	1.0	1.2 (NLO)	$2.0 \cdot 10^{5}$	0.019	*_V7A-v1
201	TTWWJets_8TeV-madgraph	0.0020	1.0	1.0 (LO)	$2.0 \cdot 10^{5}$	$1.8 \cdot 10^{-4}$	*_V7A-v1
202	TTTT_TuneZ2star_8TeV-madgraph-tauola	$7.2 \cdot 10^{-4}$	1.0	1.28 (NLO)	$9.0 \cdot 10^{4}$	$1.8 \cdot 10^{-4}$	*_V7A-v1
203	Upsilon1SToMuMu_2MuPtEtaFilter_tuneD6T_8TeV-pythia6-evtgen	$9.9 \cdot 10^{4}$	0.022	1.0 (LO)	$2.0 \cdot 10^{6}$	22	*_V7A-v1
204	Upsilon2SToMuMu_2MuPtEtaFilter_tuneD6T_8TeV-pythia6-evtgen	$1.3 \cdot 10^{5}$	0.011	1.0 (LO)	$2.0 \cdot 10^{6}$	14	*_V7A-v1
205	Upsilon3SToMuMu_2MuPtEtaFilter_tuneD6T_8TeV-pythia6	3800	0.15	1.0 (LO)	$1.0 \cdot 10^{6}$	5.7	*_V7A-v1

Appendix C
Search Results

C.1 Data Events Related to Higgs Boson Searches

$H \to 4l$	m_{4l}/GeV	Run	LS	Event	MUSiC	$\sum \lvert\vec{p}_{\mathrm{T}}\rvert$/GeV	$M_{(\mathrm{T})}$/GeV
4μ	107.3	202237	90	112735227	2μ	98.3	90.2
4μ	107.8	206446	1062	1374805900	2μ	74.1	66.5
4e	108.9	196531	226	305747400			
2e2μ	112.7	207269	140	150711847			
4μ	114.2	206243	537	807811295			
4e	116.7	193575	523	400912970	2e	54.4	52.8
4e	117.7	200188	139	187920671	2e	93.1	93.8
4μ	118.4	202973	713	681363400	2μ	54.0	19.0
2e2μ	118.8	206478	142	188791841			
4e	119.5	196027	79	123030498	2e	61.0	82.2
4μ	120.5	198271	27	36542571	1μ+1jet	95.5	95.7
2e2μ	122.2	195915	690	990388348			
4μ	122.2	202178	1286	1430970868	1μ	63.0	0.1
4μ	122.2	201707	503	635670564	2μ	93.3	88.1
4e	122.4	202299	304	421267699	2e	78.8	92.5
2e2μ	122.6	207454	359	554587442			
4μ	124.6	204577	1012	1164017963	3μ	103.3	109.3
4μ	124.8	206446	368	536901756	1μ+1jet	91.3	115.1
4e	125.3	205620	161	123983433	2e	92.9	91.0
4μ	125.5	199319	970	1203594102	2μ	68.0	73.5
4μ	125.7	191856	64	53791282	2μ	90.7	83.0
4e	126.2	201707	656	805047482	4e	131.7	121.8
4e	126.4	205718	403	674409735			
2e2μ	126.6	205238	521	544499395			
2e2μ	126.7	208686	425	656049185	1μ+2jet	169.0	185.8
2e2μ	126.7	201174	335	216745941	1e+1μ	78.8	51.7
2e2μ	127.0	200466	226	153791279	1μ+1jet	177.1	230.5
2e2μ	127.5	195099	115	137440354	1μ	43.4	0.1
2e2μ	128.1	198271	540	631637861	2e	88.9	90.1
4e	129.0	199876	75	92622577			
2e2μ	129.3	206466	260	463399210			
2e2μ	132.3	206859	391	594138345	2μ	89.3	88.3
4μ	132.4	191271	352	385661498			
4μ	132.7	204113	526	784493856			
2e2μ	132.8	200091	1481	1605749984	1μ	47.0	0.1
4μ	133.3	198941	217	262583111	1μ	74.3	0.1
4e	133.7	206513	300	278141880	2e	95.4	97.7
4e	135.4	198969	808	964868227	2e	88.5	91.8
2e2μ	136.5	200473	163	124745478	2e	82.6	100.3
4μ	139.6	195655	140	167570931	2μ+2jet	384.6	511.0
2e2μ	139.8	205781	435	592726525			

Table C.1: Details of events observed by the dedicated Higgs boson search $H \to ZZ \to 4l$ ($l = \mathrm{e}, \mu$) [302] and how these events are observed in the MUSiC analysis. The first two columns give the final state as identified by [302] and the reconstructed four-lepton invariant mass. Information on the run number, luminosity section and event number are given in the following three columns. The last three columns summarize how these events are observed in the MUSiC analysis, including the corresponding kinematic variables.

C.2 Regions with Data but No SM Mont-Carlo Prediction

Event Class	Distribution	N_{data}	Region without MC / GeV
1e+1γ+2jets+MET (+X/+Njet)	$M_{(\text{T})}$	1	1500 - 1540
1e+1γ+MET	$M_{(\text{T})}$	1	140 - 160
1e+1μ+1γ (+X/+Njet)	$M_{(\text{T})}$	1	50 - 540
1e+1μ+1jet+X (+Njet)	$M_{(\text{T})}$	1	30 - 40
1e+1μ+3jets (+Njet)	$M_{(\text{T})}$	1	2670 - 2750
1e+1μ+4jets	$M_{(\text{T})}$	1	2170 - 2230
1e+1μ+6jets+MET+Njet	$M_{(\text{T})}$	1	2180 - 2240
1e+4jets+MET (+X/+Njet)	MET	1	1010 - 1160
1e+5jets+MET (+X/+Njet)	MET	1	790 - 920
1e+6jets+Njet	$\sum \lvert\vec{p}_{\text{T}}\rvert$	1	390 - 470
1μ+1γ+1jet (+Njet)	$M_{(\text{T})}$	1	70 - 80
1μ+1γ+1jet+MET	$M_{(\text{T})}$	1	930 - 960
1μ+1γ+4jets+X	$M_{(\text{T})}$	1	1070 - 1100
1μ+1γ+4jets+MET	$M_{(\text{T})}$	1	1060 - 1100
1μ+2jets	$M_{(\text{T})}$	1	4390 - 4780
1μ+4jets+MET (+X/+Njet)	$M_{(\text{T})}$	1	190 - 210
1μ+5jets (+X/+Njet)	$M_{(\text{T})}$	1	290 - 310
2e+1γ+4jets (+X/+Njet)	$\sum \lvert\vec{p}_{\text{T}}\rvert$	1	370 - 470
2e+1γ+4jets (+Njet)	$M_{(\text{T})}$	1	710 - 740
2e+1γ+Njet	$M_{(\text{T})}$	1	50 - 60
2e+1jet+MET (+Njet)	MET	1	860 - 950
2e+1μ+1jet+Njet	$\sum \lvert\vec{p}_{\text{T}}\rvert$	1	1500 - 1620
2e+1μ+2jets (+Njet)	$\sum \lvert\vec{p}_{\text{T}}\rvert$	1	1780 - 2030
2e+1μ+4jets+X	$M_{(\text{T})}$	1	2130 - 2180
2e+3jets+MET	MET	1	590 - 670
2μ+1γ+2jets+MET (+X/+Njet)	$\sum \lvert\vec{p}_{\text{T}}\rvert$	1	800 - 950
	$M_{(\text{T})}$	1	860 - 890
2μ+1jet (+X/+Njet)	$M_{(\text{T})}$	5	30 - 40
3e+2μ	$\sum \lvert\vec{p}_{\text{T}}\rvert$	1	370 - 420
3e+2μ (+X/+Njet)	$M_{(\text{T})}$	1	580 - 590
3e+3jets+MET+X (+Njet)	$\sum \lvert\vec{p}_{\text{T}}\rvert$	1	1610 - 1750
	$M_{(\text{T})}$	1	1610 - 1650
3e+4jets (+X/+Njet)	$\sum \lvert\vec{p}_{\text{T}}\rvert$	1	1480 - 1680
	$M_{(\text{T})}$	1	1760 - 1800
3e+4jets+MET	$\sum \lvert\vec{p}_{\text{T}}\rvert$	1	1550 - 1800
3e+4jets+MET+X (+Njet)	$\sum \lvert\vec{p}_{\text{T}}\rvert$	1	1600 - 1800
3e+4jets+MET (+X/+Njet)	$M_{(\text{T})}$	1	1650 - 1700

Table C.2: Details for regions in distributions of exclusive, inclusive (+X) and jet-inclusive (+Njet) event classes with data but without any SM MC prediction. All of these data events are enclosed in larger regions that contain a sufficient SM expectation, so that a p-value can be calculated.

C.3 Event Classes Failing the Low Statistics Treatment

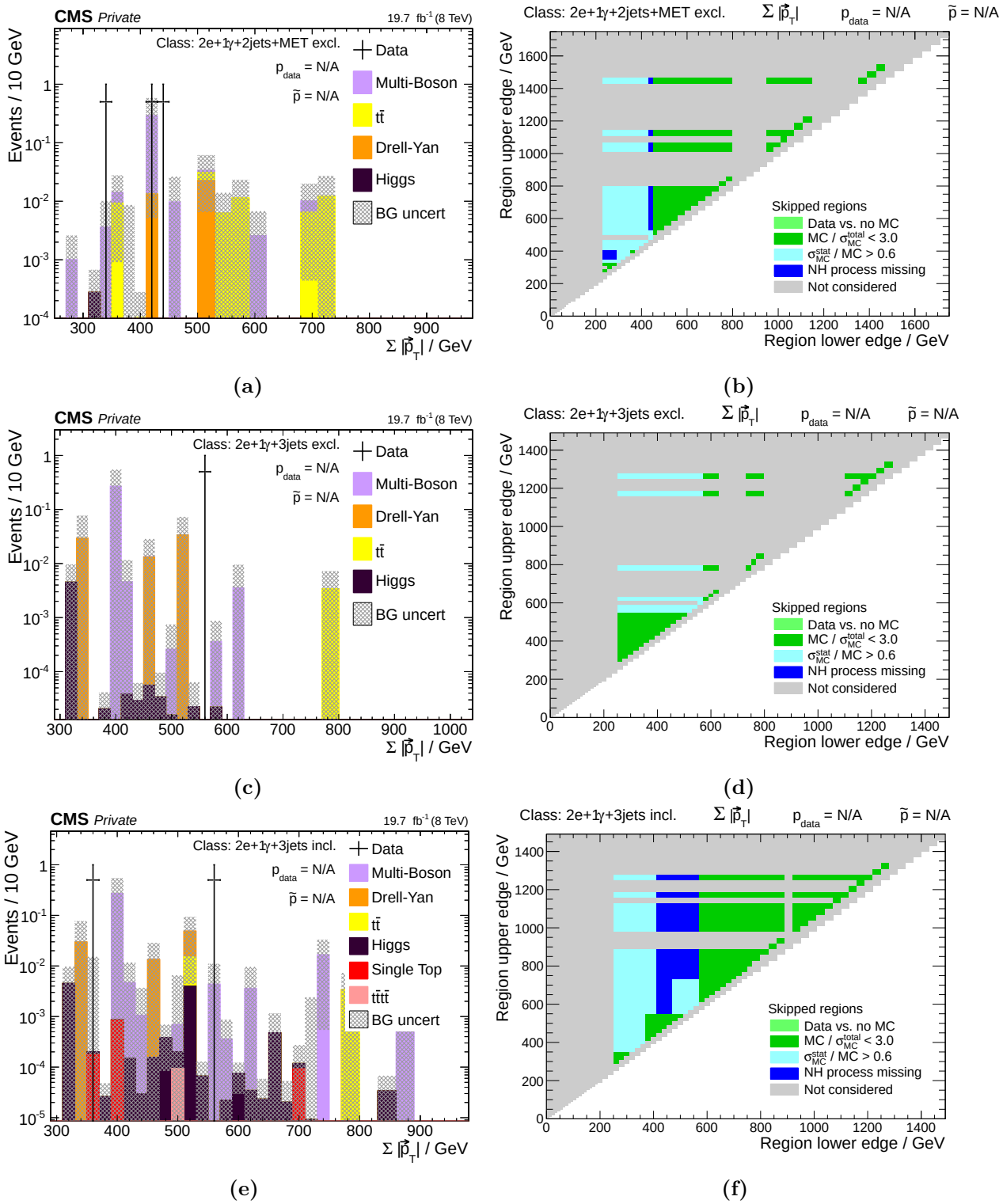


Figure C.1: Event classes in which no regions of the $\sum |\vec{p}_{\mathrm{T}}|$ distribution pass all criteria of the low generated event count treatment. (Continued in Fig. C.2)

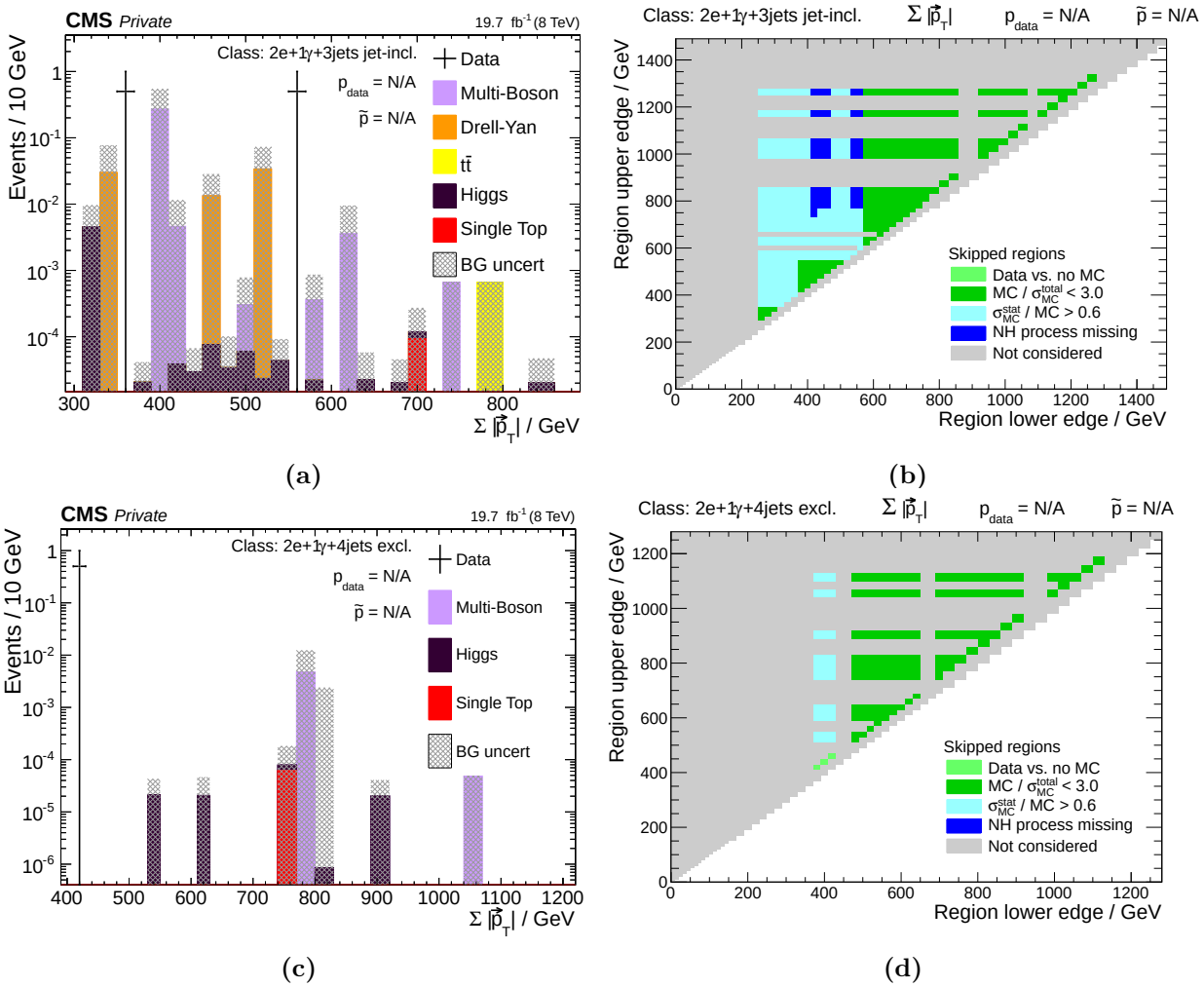


Figure C.2: Event classes in which no regions of the $\sum |\vec{p}_{\mathrm{T}}|$ distribution pass all criteria of the low generated event count treatment. (Continued from Fig. C.1)

C.4 Most Significant Event Class Distributions

	Event Class	p_{data}	$\tilde{p}$	RoI / GeV	N_{data}	$N_{\text{SM}} \pm \sigma_{\text{SM}}$
$\sum \lvert\vec{p}_{\text{T}}\rvert$	3e+2μ	$4.28 \cdot 10^{-4}$	0.0043	390 - 420	1	$(0.0 \pm 5.4) \cdot 10^{-4}$
	1e+1γ+MET	$1.12 \cdot 10^{-4}$	0.0052	170 - 330	16	2.0±2.1
	3e+2jets+MET	$1.09 \cdot 10^{-4}$	0.0059	950 - 1230	6	0.55±0.25
	2e+≥6jets+MET	$7.59 \cdot 10^{-5}$	0.0097	630 - 720	4	0.086±0.135
	1e+1μ+1γ	0.0018	0.024	290 - 370	2	0.0050±0.0600
	2e+1γ+4jets	0.0028	0.028	410 - 470	1	0.0000±0.0035
	1e+2μ+5jets+MET	0.0079	0.047	670 - 1050	2	0.115±0.062
	2μ+1γ+2jets+MET	0.0071	0.048	860 - 950	1	0.0000±0.0090
	1μ+≥6jets	$2.56 \cdot 10^{-4}$	0.06	1060 - 1260	23	63.3±9.9
	3e+4jets	0.0043	0.062	1560 - 1680	1	0.0000±0.0054

Table C.3: Overview of the most ten significant $\sum |\vec{p}_{\text{T}}|$ for exclusive event classes using the uncertainty fill-up method. p-value and $\tilde{p}$ value are stated in each case. The bin region of the RoI is given, as well as the number of data events N_{data} and SM MC events $N_{\text{SM}} \pm \sigma_{\text{SM}}$ in the RoI.

	Event Class	p_{data}	$\tilde{p}$	RoI / GeV	N_{data}	$N_{\text{SM}} \pm \sigma_{\text{SM}}$
$\sum \lvert\vec{p}_{\text{T}}\rvert$	1e+1γ+MET	$1.05 \cdot 10^{-4}$	$9.7 \cdot 10^{-4}$	270 - 330	7	0.47±0.51
	3e+2jets+MET	$9.99 \cdot 10^{-5}$	0.0027	950 - 1230	6	0.55±0.24
	1e+1μ+1γ	0.0019	0.0040	230 - 370	2	0.041±0.043
	3e+2μ	0.0052	0.0089	240 - 440	1	0.0044±0.0037
	1e+2μ+5jets+MET	0.0073	0.010	670 - 1050	2	0.115±0.052
	1μ+1γ	$7.25 \cdot 10^{-4}$	0.013	390 - 790	2	0.032±0.020
	4e+MET	0.012	0.023	210 - 330	2	0.157±0.041
	1μ+≥6jets	$2.56 \cdot 10^{-4}$	0.028	1060 - 1260	23	63.3±9.9
	3μ+4jets	0.012	0.030	550 - 790	2	0.153±0.058
	2e+≥6jets+MET	0.0014	0.049	600 - 780	8	1.57±0.69
$M_{(\text{T})}$	1e+1γ+MET	$1.98 \cdot 10^{-4}$	0.0020	270 - 330	7	0.49±0.58
	1μ+1γ	$1.41 \cdot 10^{-5}$	0.0021	460 - 830	3	0.032±0.022
	3e+2jets+MET	$9.08 \cdot 10^{-5}$	0.0023	950 - 1230	6	0.54±0.24
	2μ+1jet	$3.38 \cdot 10^{-5}$	0.0026	90 - 120	1830	1190±150
	1e+1μ+1γ	$8.63 \cdot 10^{-4}$	0.0030	310 - 540	2	0.027±0.029
	1e+1γ+2jets	$3.07 \cdot 10^{-4}$	0.0055	2110 - 2260	1	$(1.1 \pm 3.3) \cdot 10^{-4}$
	2μ+5jets	$5.40 \cdot 10^{-5}$	0.0061	1990 - 2220	13	2.1±1.1
	3e+2μ	0.0052	0.010	410 - 590	1	0.0044±0.0036
	3μ+4jets	0.0028	0.013	880 - 970	2	0.064±0.040
	4e+MET	0.011	0.023	210 - 330	2	0.154±0.040
MET	1e+1γ+MET	$5.36 \cdot 10^{-4}$	0.0038	50 - 90	14	1.8±2.1
	2μ+1γ+1jet+MET	0.0019	0.0039	60 - 100	3	0.14±0.15
	3e+2jets+MET	$8.47 \cdot 10^{-4}$	0.0085	140 - 190	6	0.89±0.32
	4e+MET	0.017	0.025	50 - 80	2	0.190±0.045
	1e+2μ+5jets+MET	0.024	0.035	50 - 140	2	0.21±0.11
	1e+3μ+MET	0.018	0.051	170 - 230	1	0.0173±0.0096
	1μ+1γ+4jets+MET	0.034	0.056	50 - 210	8	2.8±1.5
	2e+1μ+4jets+MET	0.022	0.061	100 - 190	3	0.47±0.29
	2e+1γ+MET	0.045	0.066	50 - 140	5	1.3±1.0
	3μ+MET	0.0077	0.069	230 - 290	3	0.34±0.16

Table C.4: Overview of the most ten significant $\sum |\vec{p}_{\text{T}}|$, $M_{(\text{T})}$ and MET distributions for exclusive event classes. p-value and $\tilde{p}$ value are stated in each case. The bin region of the RoI is given, as well as the number of data events N_{data} and SM MC events $N_{\text{SM}} \pm \sigma_{\text{SM}}$ in the RoI.

	Event Class	p_{data}	$\tilde{p}$	RoI / GeV	N_{data}	$N_{\text{SM}} \pm \sigma_{\text{SM}}$
$\sum \lvert\vec{p}_{\text{T}}\rvert$	2e+1γ+4jets+X	$5.58 \cdot 10^{-4}$	$6.92 \cdot 10^{-4}$	410 - 610	1	$(5.2 \pm 3.3) \cdot 10^{-4}$
	3e+2μ+X	0.0039	0.0041	290 - 400	1	0.0036±0.0024
	2e+1γ+MET+X	$8.31 \cdot 10^{-4}$	0.0055	210 - 310	8	1.10±0.82
	1e+2μ+5jets+MET+X	0.0078	0.011	670 - 1050	2	0.119±0.055
	3e+3jets+X	0.001	0.016	1340 - 1530	2	0.036±0.026
	3e+2jets+MET+X	$8.36 \cdot 10^{-4}$	0.017	920 - 1550	7	1.16±0.46
	3e+2jets+X	0.0016	0.033	910 - 1420	6	0.97±0.40
	3e+1jet+X	0.0013	0.034	820 - 1080	7	1.19±0.56
	1e+2μ+5jets+X	0.024	0.046	580 - 1060	2	0.224±0.081
	3e+1jet+MET+X	0.0034	0.060	940 - 1060	4	0.51±0.22
$M_{(\text{T})}$	2e+1γ+4jets+X	$4.49 \cdot 10^{-4}$	$7.09 \cdot 10^{-4}$	630 - 740	1	$(3.9 \pm 3.1) \cdot 10^{-4}$
	2μ+1jet+X	$1.16 \cdot 10^{-5}$	0.0016	100 - 120	2269	1530±170
	2e+1γ+3jets+X	0.0091	0.012	630 - 1250	2	0.111±0.082
	3e+2jets+MET+X	$8.08 \cdot 10^{-4}$	0.016	920 - 1550	7	1.16±0.46
	3e+2μ+X	0.005	0.019	460 - 1250	1	0.0045±0.0033
	1e+2μ+4jets+X	0.0025	0.027	1090 - 1280	3	0.235±0.090
	1e+2μ+5jets+X	0.015	0.027	1060 - 1400	2	0.168±0.072
	2μ+1γ+1jet+X	$6.94 \cdot 10^{-4}$	0.031	840 - 930	2	0.030±0.021
	1e+2μ+5jets+MET+X	0.027	0.034	670 - 1330	2	0.23±0.12
	2e+1γ+MET+X	0.012	0.034	210 - 390	8	1.7±1.4
MET	2μ+1γ+1jet+MET+X	$3.57 \cdot 10^{-4}$	0.0013	60 - 100	4	0.22±0.16
	3e+2jets+MET+X	0.0012	0.013	90 - 190	15	4.6±1.5
	1e+2μ+5jets+MET+X	0.024	0.035	50 - 140	2	0.21±0.11
	1μ+5jets+MET+X	$9.26 \cdot 10^{-4}$	0.039	430 - 510	1	14.9±3.7
	2μ+1γ+MET+X	0.024	0.049	60 - 90	4	0.86±0.48
	2e+1μ+3jets+MET+X	0.0062	0.053	130 - 190	5	0.95±0.39
	2e+1γ+1jet+MET+X	0.038	0.061	60 - 100	5	1.24±0.96
	2e+1μ+4jets+MET+X	0.028	0.072	100 - 190	3	0.54±0.29
	4e+MET+X	0.051	0.080	50 - 80	2	0.35±0.10
	1e+MET+X	0.0027	0.091	1190 - 1430	2	0.071±0.025

Table C.5: Overview of the ten most significant $\sum \lvert\vec{p}_{\text{T}}\rvert$, $M_{(\text{T})}$ and MET distributions for inclusive event classes. p-value and $\tilde{p}$ value are stated in each case. The bin region of the RoI is given, as well as the number of data events N_{data} and SM MC events $N_{\text{SM}} \pm \sigma_{\text{SM}}$ in the RoI.

	Event Class	p_{data}	$\tilde{p}$	RoI / GeV	N_{data}	$N_{\text{SM}} \pm \sigma_{\text{SM}}$
$\sum \lvert\vec{p}_{\text{T}}\rvert$	2e+1γ+4jets+Njet	$1.38 \cdot 10^{-4}$	$1.48 \cdot 10^{-4}$	410 - 590	1	$(12.9 \pm 8.1) \cdot 10^{-5}$
	3e+2μ+Njet	0.0039	0.0040	290 - 400	1	0.0036±0.0024
	2e+1γ+MET+Njet	$8.24 \cdot 10^{-4}$	0.0054	210 - 310	8	1.10±0.82
	1e+1μ+1γ+Njet	0.0028	0.0093	220 - 370	3	0.19±0.15
	1e+2μ+5jets+MET+Njet	0.0078	0.011	670 - 1050	2	0.118±0.055
	3e+2jets+MET+Njet	$6.33 \cdot 10^{-4}$	0.013	920 - 1550	7	1.11±0.44
	1e+5jets+Njet	0.0027	0.058	390 - 580	517	360±52
	3μ+3jets+Njet	0.0062	0.063	420 - 480	3	0.34±0.11
	2μ+1γ+1jet+MET+Njet	0.034	0.071	240 - 810	4	0.86±0.63
	2μ+1γ+Njet	0.0014	0.073	80 - 120	20	67±15
$M_{(\text{T})}$	2e+1γ+4jets+Njet	$1.40 \cdot 10^{-4}$	$1.66 \cdot 10^{-4}$	590 - 830	1	$(12.7 \pm 8.6) \cdot 10^{-5}$
	2μ+1jet+Njet	$1.02 \cdot 10^{-5}$	0.0014	30 - 120	2298	1480±180
	2e+1γ+3jets+Njet	0.0011	0.0017	610 - 1100	2	0.033±0.031
	3e+2jets+MET+Njet	$6.09 \cdot 10^{-4}$	0.012	920 - 1550	7	1.10±0.44
	3μ+4jets+Njet	0.0028	0.013	880 - 970	2	0.064±0.040
	3e+2μ+Njet	0.006	0.017	410 - 590	1	0.0054±0.0037
	1e+1μ+1γ+Njet	0.0064	0.027	310 - 540	2	0.083±0.077
	2e+1γ+MET+Njet	0.012	0.031	210 - 390	8	1.7±1.4
	3e+Njet	$3.87 \cdot 10^{-4}$	0.032	850 - 990	3	0.122±0.046
	1e+2μ+5jets+MET+Njet	0.027	0.035	670 - 1330	2	0.22±0.12
MET	2μ+1γ+1jet+MET+Njet	$3.55 \cdot 10^{-4}$	0.0013	60 - 110	4	0.22±0.16
	3e+2jets+MET+Njet	$9.56 \cdot 10^{-4}$	0.0095	90 - 190	15	4.4±1.5
	1e+2μ+5jets+MET+Njet	0.024	0.035	50 - 140	2	0.21±0.11
	1μ+5jets+MET+Njet	0.0011	0.045	430 - 510	1	14.4±3.6
	2μ+1γ+MET+Njet	0.024	0.047	60 - 90	4	0.86±0.48
	2e+1μ+3jets+MET+Njet	0.0058	0.052	130 - 190	5	0.93±0.39
	2e+1γ+1jet+MET+Njet	0.038	0.062	60 - 100	5	1.24±0.96
	2e+1μ+4jets+MET+Njet	0.028	0.067	100 - 190	3	0.54±0.29
	4e+MET+Njet	0.051	0.080	50 - 80	2	0.35±0.10
	1e+MET+Njet	0.0026	0.086	1190 - 1430	2	0.071±0.025

Table C.6: Overview of the ten most significant $\sum \lvert\vec{p}_{\text{T}}\rvert$, $M_{(\text{T})}$ and MET distributions for jet-inclusive event classes. p-value and $\tilde{p}$ value are stated in each case. The bin region of the RoI is given, as well as the number of data events N_{data} and SM MC events $N_{\text{SM}} \pm \sigma_{\text{SM}}$ in the RoI.

	Event Class	p_{data}	$\tilde{p}$	RoI / GeV	N_{data}	$N_{\text{SM}} \pm \sigma_{\text{SM}}$
$\sum \lvert\vec{p}_{\text{T}}\rvert$	2e+≥6jets [2Q]	$4.56 \cdot 10^{-5}$	$2.22 \cdot 10^{-4}$	440 - 870	3	0.042±0.037
	2μ [2Q]	$1.91 \cdot 10^{-5}$	$3.80 \cdot 10^{-4}$	50 - 90	369	140±52
	2μ+1jet [2Q]	$2.73 \cdot 10^{-5}$	$6.09 \cdot 10^{-4}$	160 - 200	43	15.2±4.0
	1e+1γ+MET	$1.07 \cdot 10^{-4}$	$9.67 \cdot 10^{-4}$	270 - 330	7	0.47±0.51
	3e+2jets+MET [1Q]	$9.95 \cdot 10^{-5}$	0.0027	950 - 1230	6	0.55±0.24
	2e+1μ+2jets [3Q]	0.0024	0.0028	650 - 1930	1	0.0023±0.0014
	2e+1jet+MET [2Q]	$2.14 \cdot 10^{-4}$	0.0032	1890 - 2040	1	$(2.0 \pm 1.2) \cdot 10^{-4}$
	1e+1μ+1γ[2Q]	0.0037	0.0079	110 - 370	4	0.34±0.36
	3e+2μ [1Q]	0.0052	0.0091	240 - 440	1	0.0044±0.0036
	1e+2μ+5jets+MET [1Q]	0.0099	0.014	670 - 1050	2	0.135±0.061
$M_{(\text{T})}$	2μ [2Q]	$1.70 \cdot 10^{-14}$	$< 5 \cdot 10^{-7}$	50 - 110	355	94±30
	2e+1γ+4jets [0Q]	$1.59 \cdot 10^{-4}$	$1.92 \cdot 10^{-4}$	650 - 1010	1	$(14.4 \pm 9.8) \cdot 10^{-5}$
	2e+≥6jets [2Q]	$4.51 \cdot 10^{-5}$	$2.10 \cdot 10^{-4}$	750 - 1380	3	0.042±0.037
	1e+1γ+MET	$2.00 \cdot 10^{-4}$	0.0016	270 - 330	7	0.49±0.58
	2e+1jet+MET [2Q]	$1.07 \cdot 10^{-4}$	0.0020	1940 - 2040	1	$(9.7 \pm 6.8) \cdot 10^{-5}$
	1μ+1γ	$1.41 \cdot 10^{-5}$	0.0022	460 - 830	3	0.032±0.022
	3e+2jets+MET [1Q]	$9.16 \cdot 10^{-5}$	0.0025	950 - 1230	6	0.54±0.24
	2μ+1jet [2Q]	$7.10 \cdot 10^{-5}$	0.0025	170 - 280	71	27.8±8.4
	2μ+1jet [0Q]	$3.51 \cdot 10^{-5}$	0.0027	90 - 120	1827	1200±150
	2μ+5jets [0Q]	$5.25 \cdot 10^{-5}$	0.0059	1990 - 2220	13	2.1±1.1
MET	1e+1γ+MET	$5.26 \cdot 10^{-4}$	0.0039	50 - 90	14	1.8±2.1
	3e+2jets+MET [1Q]	$8.43 \cdot 10^{-4}$	0.0082	140 - 190	6	0.89±0.32
	4e+MET [0Q]	0.017	0.018	50 - 110	2	0.194±0.046
	1e+2μ+5jets+MET [1Q]	0.028	0.037	50 - 140	2	0.23±0.12
	2μ+1γ+1jet+MET [0Q]	0.038	0.039	50 - 110	3	0.41±0.50
	3e+MET [3Q]	0.037	0.044	50 - 130	3	0.57±0.37
	1e+3μ+MET [0Q]	0.018	0.050	170 - 230	1	0.0173±0.0096
	1μ+1γ+4jets+MET	0.035	0.053	50 - 210	8	2.8±1.5
	2e+1μ+4jets+MET [1Q]	0.023	0.056	100 - 190	3	0.48±0.29
	2e+1γ+MET [0Q]	0.051	0.060	50 - 140	4	1.01±0.73

Table C.7: Overview of the ten most significant $\sum \lvert\vec{p}_{\text{T}}\rvert$, $M_{(\text{T})}$ and MET distributions for exclusive event classes taking lepton charge into account. p-value and $\tilde{p}$ value are stated in each case. The bin region of the RoI is given, as well as the number of data events N_{data} and SM MC events $N_{\text{SM}} \pm \sigma_{\text{SM}}$ in the RoI.

	Event Class	p_{data}	$\tilde{p}$	RoI / GeV	N_{data}	$N_{\text{SM}} \pm \sigma_{\text{SM}}$
$\sum \lvert\vec{p}_{\text{T}}\rvert$	2e+1γ+4jets+X[0Q]	$5.02 \cdot 10^{-4}$	$7.29 \cdot 10^{-4}$	410 - 610	1	$(4.5 \pm 3.2) \cdot 10^{-4}$
	2e+1μ+2jets+X[3Q]	0.0031	0.0034	710 - 1930	1	0.0030±0.0016
	2e+1γ+MET+X[0Q]	$7.57 \cdot 10^{-4}$	0.0040	210 - 310	7	0.87±0.63
	3e+2μ+X[1Q]	0.0039	0.0041	290 - 400	1	0.0036±0.0024
	2e+MET+X[2Q]	$2.24 \cdot 10^{-4}$	0.0042	1880 - 1970	1	$(2.0 \pm 1.5) \cdot 10^{-4}$
	2μ+X[2Q]	$2.33 \cdot 10^{-4}$	0.0054	50 - 80	464	205±71
	2μ+1jet+X[2Q]	$2.01 \cdot 10^{-4}$	0.0072	160 - 200	110	52±13
	1e+2μ+5jets+MET+X[1Q]	0.011	0.016	670 - 1050	2	0.139±0.066
	3e+2jets+MET+X[1Q]	$8.37 \cdot 10^{-4}$	0.017	920 - 1550	7	1.16±0.47
	3e+3jets+X[1Q]	0.0012	0.017	1340 - 1850	2	0.040±0.028
$M_{(\text{T})}$	2μ+X[2Q]	$1.10 \cdot 10^{-7}$	$7.00 \cdot 10^{-6}$	60 - 70	148	45±16
	2e+1γ+4jets+X[0Q]	$4.50 \cdot 10^{-4}$	$7.50 \cdot 10^{-4}$	630 - 740	1	$(3.9 \pm 3.1) \cdot 10^{-4}$
	2μ+1jet+X[0Q]	$1.33 \cdot 10^{-5}$	0.0017	90 - 120	2452	1670±180
	2e+MET+X[2Q]	$2.31 \cdot 10^{-4}$	0.0033	1850 - 1910	1	$(2.1 \pm 1.5) \cdot 10^{-4}$
	3e+X[3Q]	$3.62 \cdot 10^{-4}$	0.0056	440 - 470	2	0.021±0.016
	1μ+1γ+4jets+X	$2.45 \cdot 10^{-4}$	0.0070	860 - 1250	9	1.46±0.64
	2μ+1jet+X[2Q]	$1.43 \cdot 10^{-4}$	0.0070	150 - 190	78	33.6±9.2
	2e+1γ+3jets+X[0Q]	0.0062	0.0072	630 - 1400	2	0.092±0.065
	2μ+2jets+X[2Q]	$1.71 \cdot 10^{-4}$	0.011	280 - 360	39	16.3±3.4
	2μ+3jets+X[2Q]	$1.25 \cdot 10^{-4}$	0.012	1200 - 1550	9	1.60±0.38
MET	3e+2jets+MET+X[1Q]	0.0012	0.011	90 - 190	15	4.6±1.5
	2μ+1γ+1jet+MET+X[0Q]	0.012	0.024	50 - 100	4	0.50±0.51
	1μ+5jets+MET+X	$9.55 \cdot 10^{-4}$	0.039	430 - 510	1	15.0±3.8
	1e+2μ+5jets+MET+X[1Q]	0.028	0.040	50 - 140	2	0.23±0.12
	3e+MET+X[3Q]	0.038	0.050	50 - 160	4	1.00±0.58
	2e+1μ+3jets+MET+X[1Q]	0.010	0.074	120 - 190	5	1.11±0.41
	2e+1μ+4jets+MET+X[1Q]	0.028	0.084	100 - 190	3	0.55±0.29
	1μ+1γ+4jets+MET+X	0.053	0.086	50 - 210	9	3.7±1.8
	1e+MET+X	0.0026	0.097	1190 - 1430	2	0.071±0.025
	4e+MET+X[0Q]	0.048	0.12	50 - 80	2	0.34±0.10

Table C.8: Overview of the ten most significant $\sum \lvert\vec{p}_{\text{T}}\rvert$, $M_{(\text{T})}$ and MET distributions for inclusive event classes taking lepton charge into account. p-value and $\tilde{p}$ value are stated in each case. The bin region of the RoI is given, as well as the number of data events N_{data} and SM MC events $N_{\text{SM}} \pm \sigma_{\text{SM}}$ in the RoI.

	Event Class	p_{data}	$\tilde{p}$	RoI / GeV	N_{data}	$N_{\text{SM}} \pm \sigma_{\text{SM}}$
$\sum \lvert\vec{p}_{\text{T}}\rvert$	2μ+1jet+Njet [2Q]	$5.68 \cdot 10^{-7}$	$2.65 \cdot 10^{-5}$	160 - 200	60	19.8±5.0
	2μ+Njet [2Q]	$5.65 \cdot 10^{-6}$	$1.10 \cdot 10^{-4}$	50 - 110	520	209±67
	2e+1γ+4jets+Njet [0Q]	$1.38 \cdot 10^{-4}$	$1.14 \cdot 10^{-4}$	410 - 590	1	$(12.9 \pm 8.1) \cdot 10^{-5}$
	2e+1μ+2jets+Njet [3Q]	0.0024	0.0028	650 - 1930	1	0.0023±0.0014
	2e+1γ+MET+Njet [0Q]	$7.51 \cdot 10^{-4}$	0.0041	210 - 310	7	0.87±0.63
	3e+2μ+Njet [1Q]	0.0039	0.0041	290 - 400	1	0.0036±0.0024
	2e+MET+Njet [2Q]	$2.24 \cdot 10^{-4}$	0.0043	1880 - 1970	1	$(2.0 \pm 1.5) \cdot 10^{-4}$
	2e+1μ+1jet+Njet [3Q]	0.0022	0.0066	590 - 1540	1	0.0020±0.0013
	1e+1μ+1γ+Njet [0Q]	0.0032	0.0075	260 - 470	2	0.060±0.051
	1e+2μ+Njet [3Q]	0.0095	0.013	140 - 270	1	0.0083±0.0065
$M_{(\text{T})}$	2μ+Njet [2Q]	$1.38 \cdot 10^{-12}$	$< 5 \cdot 10^{-7}$	50 - 90	335	95±30
	2e+1γ+4jets+Njet [0Q]	$1.35 \cdot 10^{-4}$	$1.52 \cdot 10^{-4}$	590 - 950	1	$(12.4 \pm 8.1) \cdot 10^{-5}$
	2μ+1jet+Njet [2Q]	$9.87 \cdot 10^{-6}$	$4.19 \cdot 10^{-4}$	170 - 280	97	38±10
	2μ+1jet+Njet [0Q]	$1.10 \cdot 10^{-5}$	0.0019	30 - 120	2294	1490±180
	2e+1γ+3jets+Njet [0Q]	0.0017	0.0022	610 - 1400	2	0.047±0.034
	2e+MET+Njet [2Q]	$2.31 \cdot 10^{-4}$	0.0031	1850 - 1910	1	$(2.1 \pm 1.5) \cdot 10^{-4}$
	2e+Njet [2Q]	$1.01 \cdot 10^{-5}$	0.0048	280 - 310	161	57±21
	3e+2jets+MET+Njet [1Q]	$6.12 \cdot 10^{-4}$	0.011	920 - 1550	7	1.10±0.44
	3μ+4jets+Njet [1Q]	0.0028	0.014	880 - 970	2	0.065±0.040
	2e+5jets+Njet [2Q]	0.0018	0.016	1200 - 1240	2	0.048±0.035
MET	3e+2jets+MET+Njet [1Q]	$9.31 \cdot 10^{-4}$	0.0083	90 - 190	15	4.4±1.5
	2μ+1γ+1jet+MET+Njet [0Q]	0.012	0.024	50 - 100	4	0.50±0.51
	1e+2μ+5jets+MET+Njet [1Q]	0.028	0.041	50 - 140	2	0.23±0.12
	1μ+5jets+MET+Njet	0.0011	0.045	430 - 510	1	14.5±3.6
	3e+MET+Njet [3Q]	0.038	0.050	50 - 160	4	0.99±0.58
	2e+1μ+3jets+MET+Njet [1Q]	0.0094	0.074	120 - 190	5	1.08±0.41
	2e+1μ+4jets+MET+Njet [1Q]	0.028	0.078	100 - 190	3	0.55±0.29
	1e+MET+Njet	0.0026	0.086	1190 - 1430	2	0.071±0.025
	4e+MET+Njet [0Q]	0.048	0.096	50 - 80	2	0.34±0.10
	1μ+1γ+4jets+MET+Njet	0.041	0.11	150 - 210	2	0.27±0.17

Table C.9: Overview of the ten most significant $\sum \lvert\vec{p}_{\text{T}}\rvert$, $M_{(\text{T})}$ and MET distributions for jet-inclusive event classes taking lepton charge into account. p-value and $\tilde{p}$ value are stated in each case. The bin region of the RoI is given, as well as the number of data events N_{data} and SM MC events $N_{\text{SM}} \pm \sigma_{\text{SM}}$ in the RoI.

Appendix D
Sensitivity Studies

D.1 False Discovery Rate

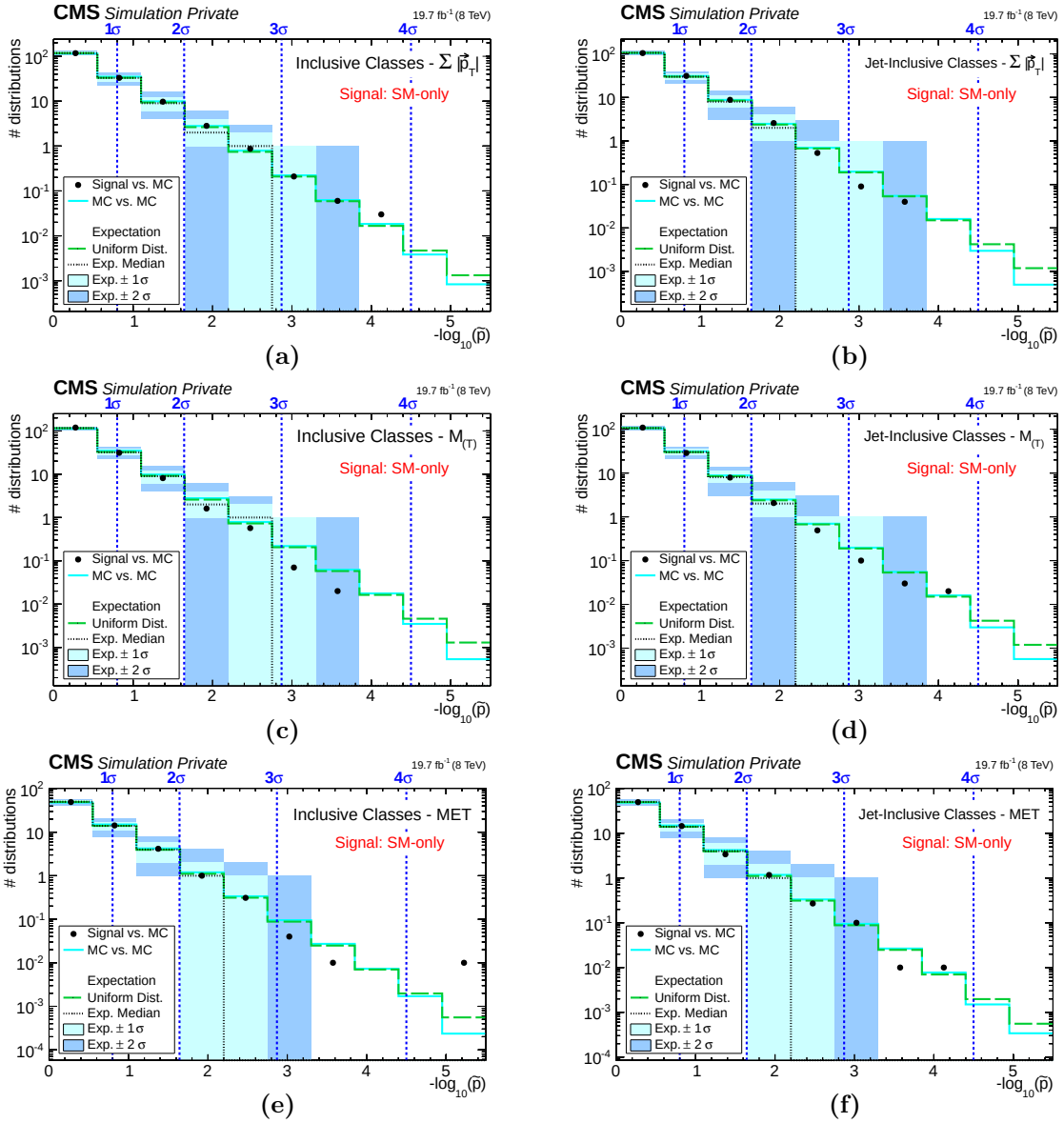


Figure D.1: Distributions of $\tilde{p}$-values for kinematic distributions of inclusive (left side: a: $\sum |\vec{p}_{\mathrm{T}}|$, c: $M_{(\mathrm{T})}$, e: MET) and jet-inclusive (right side: b: $\sum |\vec{p}_{\mathrm{T}}|$, d: $M_{(\mathrm{T})}$, f: MET) event classes. Black markers represent the averaged results of pseudo-data from 100 SM-only signal rounds compared with the SM MC expectation for each event class In the first bin a: 117.3 [b: 104.7, c: 119.5, d: 108.8, e: 49.2, f: 49.6] distributions are observed in pseudo-data ("Signal vs. MC"), with a: $118 \pm 6(1\,\sigma)^{+11}_{-12}(2\,\sigma)$ [b: $106^{+6}_{-5}(1\,\sigma) \pm 11(2\,\sigma)$, c: $116^{+5}_{-6}(1\,\sigma)^{+11}_{-12}(2\,\sigma)$, d: $106^{+6}_{-5}(1\,\sigma) \pm 11(2\,\sigma)$, e: $50^{+3}_{-4}(1\,\sigma)^{+7}_{-8}(2\,\sigma)$, f: $50^{+3}_{-4}(1\,\sigma)^{+7}_{-8}(2\,\sigma)$] expected from the SM.

D.2 Rediscovering Standard Model Processes

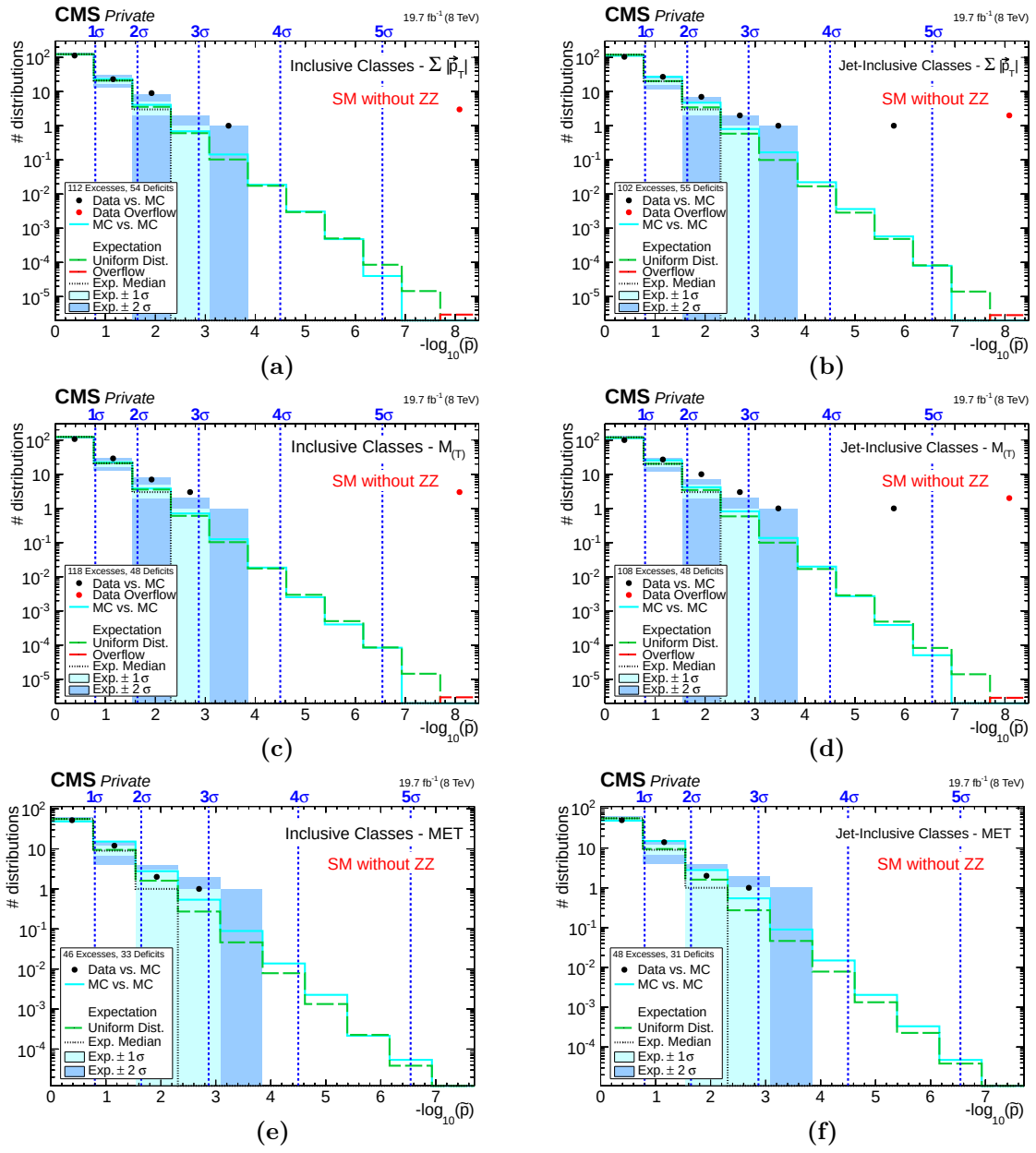


Figure D.2: Distributions of $\tilde{p}$-values for kinematic distributions of inclusive (left side: a: $\sum |\vec{p}_{\mathrm{T}}|$, c: $M_{(\mathrm{T})}$, e: MET) and jet-inclusive (right side: b: $\sum |\vec{p}_{\mathrm{T}}|$, d: $M_{(\mathrm{T})}$, f: MET) event classes. Black markers represent the measured data compared to the SM MC expectation from which modeling of ZZ production is removed. In the first bin a: 113 [b: 104, c: 107, d: 100, e: 52, f: 50] distributions are observed in data, with a: $124^{+4}_{-5}(1\,\sigma)^{+8}_{-10}(2\,\sigma)$ [b: $120^{+4}_{-5}(1\,\sigma)^{+8}_{-10}(2\,\sigma)$, c: $124^{+4}_{-5}(1\,\sigma)^{+8}_{-10}(2\,\sigma)$, d: $120^{+4}_{-5}(1\,\sigma)^{+8}_{-10}(2\,\sigma)$, e: $56 \pm 3(1\,\sigma)^{+5}_{-7}(2\,\sigma)$, f: $56 \pm 3(1\,\sigma)^{+5}_{-7}(2\,\sigma)$] expected from the SM. Corresponding information on the ten most significant event classes is given in Tab. D.1 and Tab. D.2.

	Event Class	p_{data}	$\tilde{p}$	RoI / GeV	N_{data}	$N_{\text{SM}} \pm \sigma_{\text{SM}}$
$\sum \lvert\vec{p}_{\text{T}}\rvert$ distributions	2e+2μ+X	$1.16 \cdot 10^{-27}$	$< 5 \cdot 10^{-7}$	140 - 340	43	3.24±0.56
	4μ+X	$5.50 \cdot 10^{-13}$	$< 5 \cdot 10^{-7}$	140 - 230	16	1.13±0.16
	4e+X	$1.47 \cdot 10^{-9}$	$< 5 \cdot 10^{-7}$	130 - 240	12	0.87±0.21
	2e+1γ+4jets+X	$5.58 \cdot 10^{-4}$	$7.06 \cdot 10^{-4}$	410 - 610	1	$(5.2 \pm 3.3) \cdot 10^{-4}$
	3e+2μ+X	0.0045	0.0050	240 - 400	1	0.0043±0.0025
	2e+2μ+1jet+X	$3.91 \cdot 10^{-4}$	0.0055	260 - 340	4	0.313±0.070
	2e+1γ+MET+X	0.0013	0.0077	210 - 330	8	1.18±0.89
	4e+MET+X	0.0021	0.0090	210 - 330	2	0.059±0.029
	3e+1γ+X	0.0047	0.011	170 - 200	1	0.0038±0.0036
	1e+2μ+5jets+MET+X	0.0078	0.011	670 - 1050	2	0.119±0.055
$M_{(\text{T})}$ distributions	2e+2μ+X	$1.18 \cdot 10^{-27}$	$< 5 \cdot 10^{-7}$	170 - 700	45	3.72±0.59
	4μ+X	$2.83 \cdot 10^{-14}$	$< 5 \cdot 10^{-7}$	130 - 220	14	0.611±0.090
	4e+X	$1.22 \cdot 10^{-8}$	$< 5 \cdot 10^{-7}$	90 - 460	13	1.25±0.32
	2e+1γ+4jets+X	$4.49 \cdot 10^{-4}$	0.0011	630 - 740	1	$(3.9 \pm 3.1) \cdot 10^{-4}$
	2μ+1jet+X	$1.08 \cdot 10^{-5}$	0.0015	100 - 120	2269	1520±170
	3e+2μ+X	0.0045	0.0046	300 - 590	1	0.0042±0.0025
	2e+1γ+3jets+X	0.0075	0.0081	610 - 1250	2	0.095±0.079
	2μ+1γ+1jet+MET+X	0.0029	0.011	320 - 870	4	0.44±0.25
	3e+2jets+MET+X	$7.62 \cdot 10^{-4}$	0.015	920 - 1550	7	1.14±0.46
	4e+MET+X	0.0038	0.016	210 - 330	2	0.082±0.036
MET distributions	2μ+1γ+1jet+MET+X	$3.17 \cdot 10^{-4}$	0.0012	60 - 100	4	0.20±0.16
	3e+2jets+MET+X	0.0012	0.012	90 - 190	15	4.6±1.5
	4e+MET+X	0.010	0.023	50 - 80	2	0.135±0.064
	1e+2μ+5jets+MET+X	0.024	0.036	50 - 140	2	0.21±0.11
	1μ+5jets+MET+X	$9.26 \cdot 10^{-4}$	0.039	430 - 510	1	14.9±3.7
	2μ+1γ+MET+X	0.023	0.047	60 - 90	4	0.84±0.48
	2e+1μ+3jets+MET+X	0.0062	0.053	130 - 190	5	0.95±0.39
	2e+1γ+1jet+MET+X	0.038	0.065	60 - 100	5	1.23±0.96
	2e+1μ+4jets+MET+X	0.028	0.070	100 - 190	3	0.54±0.29
	1e+MET+X	0.0027	0.094	1190 - 1430	2	0.071±0.025

Table D.1: Overview of the ten most significant $\sum \lvert\vec{p}_{\text{T}}\rvert$, $M_{(\text{T})}$ and MET distributions for inclusive event classes of measured data compared to an SM MC expectation from which modeling of ZZ production is removed. p-value and $\tilde{p}$ value are stated in each case. The bin region of the RoI is given, as well as the number of data events N_{data} and SM MC events $N_{\text{SM}} \pm \sigma_{\text{SM}}$ in the RoI.

	Event Class	p_{data}	$\tilde{p}$	RoI / GeV	N_{data}	$N_{\text{SM}} \pm \sigma_{\text{SM}}$
$\sum \lvert\vec{p}_{\text{T}}\rvert$ distributions	2e+2μ+Njet	$3.99 \cdot 10^{-30}$	$< 5 \cdot 10^{-7}$	140 - 510	44	3.04±0.49
	4μ+Njet	$2.52 \cdot 10^{-14}$	$< 5 \cdot 10^{-7}$	140 - 690	19	1.47±0.20
	4e+Njet	$9.27 \cdot 10^{-8}$	$1.50 \cdot 10^{-6}$	130 - 250	10	0.84±0.19
	2e+1γ+4jets+Njet	$1.38 \cdot 10^{-4}$	$1.46 \cdot 10^{-4}$	410 - 590	1	$(12.9 \pm 8.1) \cdot 10^{-5}$
	2e+2μ+1jet+Njet	$1.03 \cdot 10^{-4}$	0.0014	250 - 340	4	0.221±0.047
	1e+1μ+1γ+Njet	0.0015	0.0036	220 - 370	3	0.15±0.12
	3e+2μ+Njet	0.0045	0.0049	240 - 400	1	0.0043±0.0025
	2e+1γ+MET+Njet	0.0013	0.0078	210 - 330	8	1.18±0.89
	4e+MET+Njet	0.0021	0.008	210 - 330	2	0.059±0.029
	1e+2μ+5jets+MET+Njet	0.0078	0.011	670 - 1050	2	0.118±0.055
$M_{(\text{T})}$ distributions	2e+2μ+Njet	$2.74 \cdot 10^{-30}$	$< 5 \cdot 10^{-7}$	170 - 700	44	3.05±0.47
	4μ+Njet	$3.39 \cdot 10^{-15}$	$< 5 \cdot 10^{-7}$	130 - 280	17	0.96±0.13
	4e+Njet	$8.88 \cdot 10^{-8}$	$3.50 \cdot 10^{-6}$	90 - 460	11	1.06±0.24
	2e+1γ+4jets+Njet	$1.40 \cdot 10^{-4}$	$1.46 \cdot 10^{-4}$	590 - 830	1	$(12.7 \pm 8.6) \cdot 10^{-5}$
	2e+1γ+3jets+Njet	$7.62 \cdot 10^{-4}$	$8.88 \cdot 10^{-4}$	610 - 1250	2	0.027±0.026
	2μ+1jet+Njet	$9.61 \cdot 10^{-6}$	0.0014	30 - 120	2298	1480±180
	3e+2μ+Njet	0.0045	0.0048	300 - 590	1	0.0042±0.0025
	2e+2μ+1jet+Njet	$4.73 \cdot 10^{-4}$	0.0052	300 - 880	5	0.59±0.11
	4μ+1jet+Njet	$5.43 \cdot 10^{-4}$	0.0057	380 - 700	3	0.146±0.036
	2μ+1γ+1jet+MET+Njet	0.0029	0.011	320 - 870	4	0.43±0.25
MET distributions	2μ+1γ+1jet+MET+Njet	$3.17 \cdot 10^{-4}$	$9.41 \cdot 10^{-4}$	60 - 110	4	0.20±0.16
	3e+2jets+MET+Njet	$9.16 \cdot 10^{-4}$	0.010	90 - 190	15	4.4±1.5
	4e+MET+Njet	0.01	0.023	50 - 80	2	0.134±0.064
	1e+2μ+5jets+MET+Njet	0.024	0.035	50 - 140	2	0.21±0.11
	2μ+1γ+MET+Njet	0.023	0.047	60 - 90	4	0.84±0.48
	1μ+5jets+MET+Njet	0.0011	0.048	430 - 510	1	14.4±3.6
	2e+1μ+3jets+MET+Njet	0.0058	0.050	130 - 190	5	0.93±0.39
	2e+1γ+1jet+MET+Njet	0.038	0.065	60 - 100	5	1.23±0.96
	2e+1μ+4jets+MET+Njet	0.028	0.070	100 - 190	3	0.54±0.29
	1e+MET+Njet	0.0026	0.085	1190 - 1430	2	0.071±0.025

Table D.2: Overview of the ten most significant $\sum \lvert\vec{p}_{\text{T}}\rvert$, $M_{(\text{T})}$ and MET distributions for jet-inclusive event classes of measured data compared to an SM MC expectation from which modeling of ZZ production is removed. p-value and $\tilde{p}$ value are stated in each case. The bin region of the RoI is given, as well as the number of data events N_{data} and SM MC events $N_{\text{SM}} \pm \sigma_{\text{SM}}$ in the RoI.

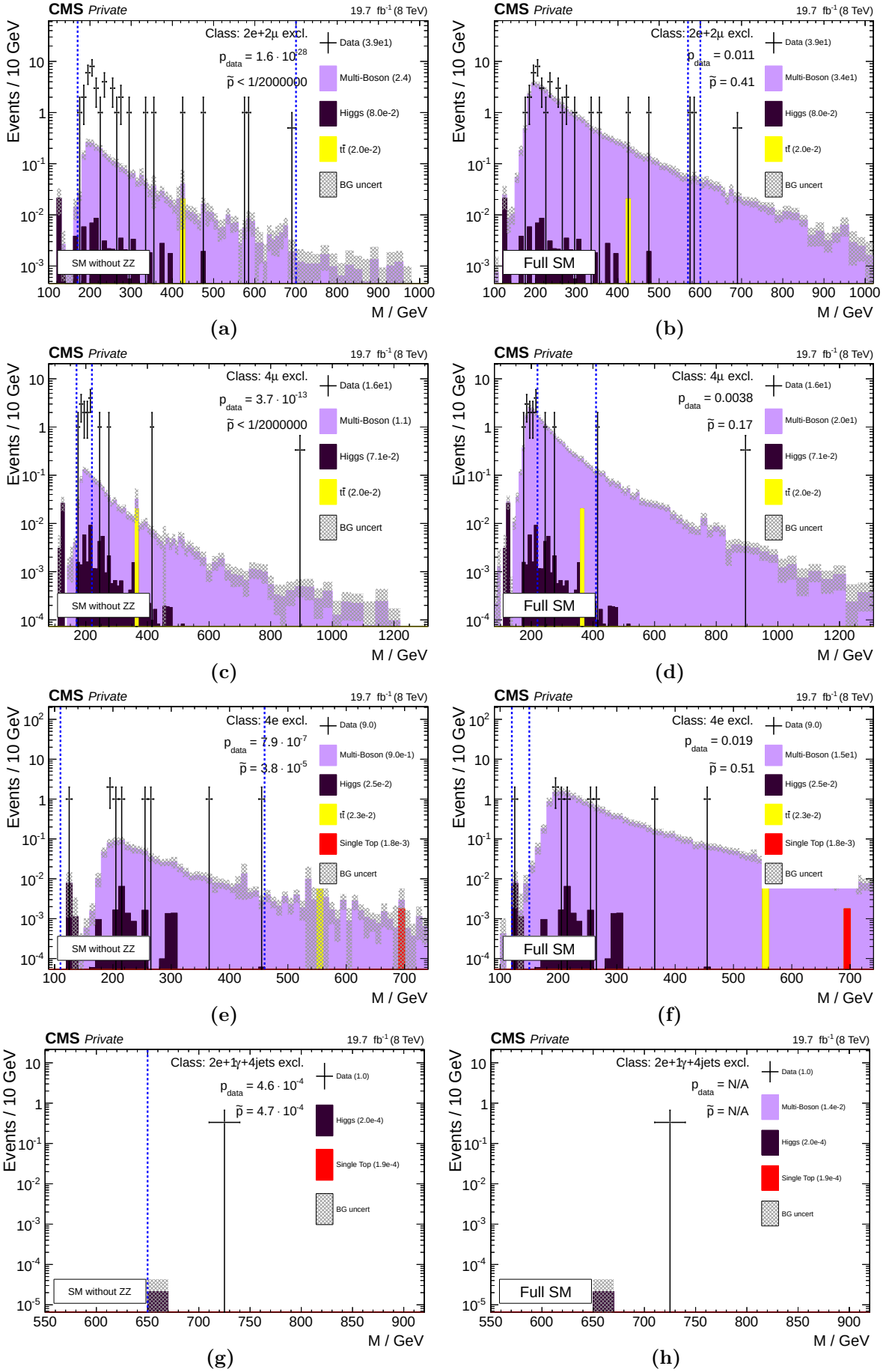


Figure D.3: The left column (a, c, e, g) shows the four most significant $M_{(\mathrm{T})}$ distributions for exclusive event classes of measured data compared to an SM MC expectation from which modeling of ZZ production is removed. Strong deviations are observed. For comparison, results of the same distributions using the full SM description (including ZZ production) are shown in the right column.

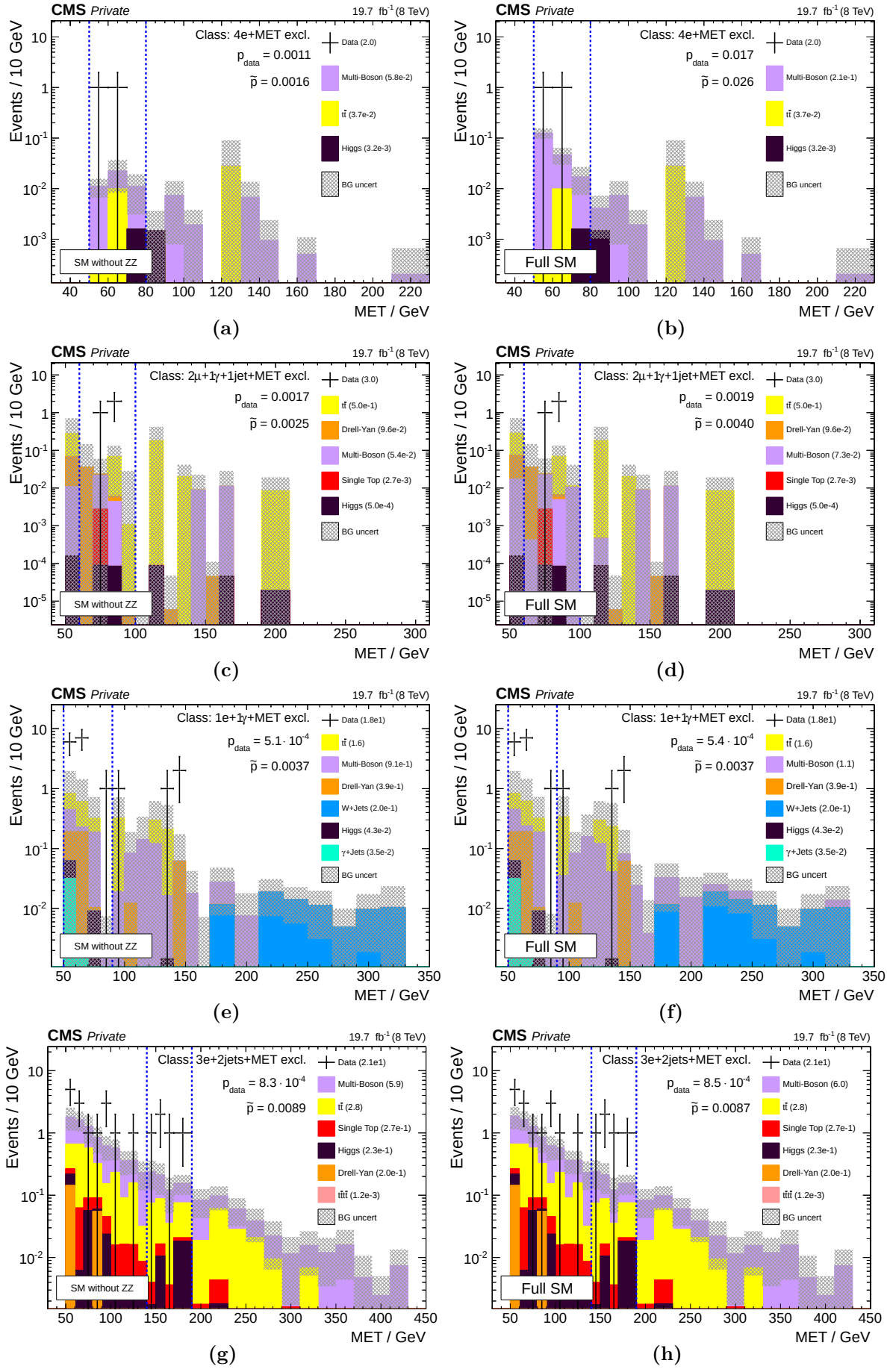


Figure D.4: The left column (a, c, e, g) shows the four most significant MET distributions for exclusive event classes of measured data compared to an SM MC expectation from which modeling of ZZ production is removed. Strong deviations are observed. For comparison, results of the same distributions using the full SM description (including ZZ production) are shown in the right column.

Acronyms

AlCa **Al**ignment and **Ca**libration. 24, 29
ALICE **A** **L**arge **I**on **C**ollider **E**xperiment. 15
AOD **A**nalysis **O**bject **D**ata. 30, 35, 165
ATLAS **A** **T**oroidal **L**HC **A**pparatu**S**. 1, 10, 15, 34

BSM **B**eyond **S**tandard **M**odel. 1, 11, 12, 17, 29, 33, 34, 37, 38, 51, 53, 61, 62, 70, 103, 108, 112, 119, 133, 139, 142, 157–159, 166

CDF **C**ollider **E**xperiment at **F**ermilab. 34
CERN Originally: **C**onseil **E**uropéen pour la **R**echerche **N**ucléaire; now: European Organization for Nuclear Research. 15, 16, 30
CHS **C**harged **H**adron **S**ubtraction. 28, 29
CMS **C**ompact **M**uon **S**olenoid. 1, 10, 15–24, 27, 28, 30, 34–36, 53–55, 57, 61, 63, 64, 67, 70, 89, 92–97, 101, 105, 106, 108, 112, 132, 138, 142, 145, 149, 158, 159, 165
CMSSW **CMS** **S**oft**w**are. 30, 35, 53, 64
CPU **C**entral **P**rocessing **U**nit. 23
CRAB **C**MS **R**emote **A**nalysis **B**uilder. 35
CREAM **C**omputing **R**esource **E**xecution **a**nd **M**anagement. 35
CSC **C**athode **S**trip **C**hamber. 21, 22, 25, 67
CTEQ **C**oordinated **T**heoretical-**E**xperimental Project on **Q**CD. 62, 85

DAQ **D**ata **A**cquisition. 23
DGLAP **D**okshitzer **G**ribov **L**ipatov **A**ltarelli **P**arisi. 62
DIGI **Digi**talization. 30, 64
DPS **D**ouble **P**arton **S**cattering. 59
DT **D**rift **T**ube. 21, 22, 25
DY **D**rell-**Y**an. 57, 58, 60, 84, 90, 91, 101, 103, 104, 108, 121, 123–125, 127–129, 135, 136, 154, 155

ECal **E**lectromagnetic **Cal**orimeter. 19–21, 23, 24, 26, 27, 39, 68–70, 73–75, 77, 80, 92–96, 138
EDM **E**vent **D**ata **M**odel. 30

FastSIM **Fast** **Sim**ulation. 64
FSR **F**inal **S**tate **R**adiation. 54, 91, 104, 121, 124, 125, 127
FullSIM **Full** **Sim**ulation. 64

GEN **Gen**eration. 30
GSF **G**aussian **S**um **F**ilter. 26, 27, 69, 81, 83
GUT **G**rand **U**nified **T**heory. 12
GWS **G**lashow **W**einberg **S**alam. 9

HB **H**adron **B**arrel. 20
HCal **H**adronic **Cal**orimeter. 20, 21, 24, 27, 68, 74, 92–96, 138
HE **H**adron **E**ndcap. 20
HEEP **H**igh **E**nergy **E**lectron **P**airs. 73, 81, 83
HERA English: **H**adron-**E**lectron **R**ing **A**ccelerator; German: **H**adron-**E**lektron-**R**inga**n**lage. 1, 33
HF **H**adron **F**orward. 21, 24
HLT **H**igh **L**evel **T**rigger. 23, 30, 53, 68–70
HO **H**adron **O**uter. 20
HPD **H**ybrid **P**hoto **D**iode. 20, 68

ISR **I**nitial **S**tate **R**adiation. 54, 104, 121, 124, 125, 127, 142

JEC **J**et **E**nergy scale **C**orrections. 29, 64, 79, 80, 129
JER **J**et **E**nergy **R**esolution. 29, 76, 129
JES **J**et **E**nergy **S**cale. 29

KF **K**alman **F**ilter. 25–27, 69, 83

L1 **L**evel-**1**. 23, 30, 67, 68
LEE **l**ook-**e**lsewhere **e**ffect. 38, 39, 47–51, 89, 103, 104, 110, 112, 119, 120, 124, 131, 135, 139–143, 145–149, 163, 164
LEP **L**arge **E**lectron-**P**ositron. 1, 15, 33
LHAPDF **L**es **H**ouches **A**ccord **PDF**. 85
LHC **L**arge **H**adron **C**ollider. 1, 10, 11, 15–19, 23, 24, 30, 33–36, 38, 44, 53, 55, 58–61, 63, 68, 104, 132, 138, 158, 159
LHCb **LHC** **b**eauty. 10, 15
LIGO **L**aser **I**nterferometer **G**ravitation **W**ave **O**bservatory. 5
Linac2 **Lin**ear **Ac**celerator 2. 15
LO **L**eading **O**rder. 54, 55, 57–63, 84, 106, 127, 157

MC **M**onte **C**arlo. 34, 39–49, 51, 54–58, 60–62, 64, 65, 71, 79–86, 89–91, 101, 106, 108, 110–121, 125–127, 129, 130, 132–137, 139–144, 149–155, 157–159, 163, 166, 167, 174, 177–188
MCFM **M**onte **C**arlo for **F**e**M**tobarn processes. 59
MSTW **M**artin **S**tirling **T**horne **W**att. 62, 85
MUSiC **M**odel **U**nspecific **S**earch **i**n **C**MS. 1, 34–43, 46, 47, 49, 51–53, 55, 57, 58, 61, 67, 70, 71, 74–76, 79, 81–83, 86, 89, 92–96, 101, 103–106, 108, 112, 113, 117, 119–124, 127, 129, 131, 132, 137–142, 146, 149, 152, 155–159, 163, 173

NH **n**eighbor**h**ood. 45–47, 112–114, 117, 125, 127

NLO **N**ext-to-**L**eading **O**rder. 54, 57–61, 63, 84, 106, 108, 127, 157
NNLL **N**ext-to-**N**ext-to-**L**eading-**L**ogarithmic. 60, 84
NNLO **N**ext-to-**N**ext-to-**L**eading **O**rder. 55, 57, 58, 60, 62, 84, 106
NNPDF **N**eural **N**etwork **PDF**. 62, 85

OF **O**verflow **F**raction. 142, 143, 147

PAT **P**hysics **A**nalysis **T**oolkit. 30, 35
PDF **P**arton **D**istribution **F**unction. 62, 63, 84, 85, 106
PF **P**article-**F**low. 27–29, 39, 71, 73–77, 81
PMT **P**hoto **M**ultiplier **T**ube. 21, 68
PS **P**roton **S**ynchrotron. 15
PSB **P**roton **S**ynchrotron **B**ooster. 15
PU **p**ile**u**p. 17, 20, 24, 25, 28, 29, 54, 60, 63–65, 120, 142, 166, 167
PXL **P**hysics e**X**tension **L**ibrary. 35

QCD **q**uantum **c**hromo**d**ynamics. 10, 28, 54, 55, 58–60, 62, 63, 65, 84, 103, 117, 127, 137, 157
QED **q**uantum **e**lectro**d**ynamics. 7–9, 11

RECO **Reco**nstruction. 30
RoI **R**egion **o**f **I**nterest. 41, 47, 48, 50, 51, 89, 101–106, 108, 112, 113, 116, 117, 120, 123–131, 135, 138, 143, 145, 149, 150, 152, 153, 155, 163, 177–182, 185, 186
RPC **R**esistive **P**late **C**hamber. 21, 22

SIM **Sim**ulation. 30, 64
SM **S**tandard **M**odel. 1, 5–8, 10–12, 33, 34, 36–38, 40–43, 45–53, 55, 56, 58, 60, 61, 63, 71, 76, 79, 80, 82–84, 86, 89–91, 97–106, 108–117, 119–121, 123–127, 129–135, 137–147, 149–159, 163, 164, 167, 174, 177–188
SPS **S**uper **P**roton **S**ynchrotron. 11, 15
SSM **S**equential **S**tandard **M**odel. 61, 62, 142

T0 **T**ier-**0**. 23, 30
T1 **T**ier-**1**. 30
T2 **T**ier-**2**. 30, 35
TEC **T**racker **E**nd **C**ap. 18, 19, 27
TIB **T**racker **I**nner **B**arrel. 18, 19
TID **T**racker **I**nner **D**isk. 18, 19
TOB **T**racker **O**uter **B**arrel. 18, 19
TPFMS **T**racker **P**lus **F**irst **M**uon **S**tation. 25, 26, 72

WLCG **W**orldwide **L**HC **C**omputing **G**rid. 30, 35
WMAP **W**ilkinson **M**icrowave **A**nisotropy **P**robe. 11

Bibliography

[1] H. Pieta. "MUSiC - a Model Unspecific Search in CMS based on 2010 LHC data". PhD thesis. RWTH Aachen University, June 2012. URL: http://publications.rwth-aachen.de/record/82887 (cit. on pp. 1, 34, 43).

[2] P. Papacz. "Model Unspecific Search for New Physics in CMS Based on 2011 Data". PhD thesis. RWTH Aachen University, Oct. 2014. URL: http://publications.rwth-aachen.de/record/465391 (cit. on pp. 1, 34, 35, 40, 44, 121, 122, 124, 131, 132, 137).

[3] P. Biallass. "Commissioning of the CMS muon detector and development of generic search strategies for new physics". PhD thesis. RWTH Aachen University, Mar. 2009. URL: http://publications.rwth-aachen.de/record/50694 (cit. on pp. 1, 34).

[4] C. Hof. "Implementation of a model independent search for new physics with the CMS detector exploiting the world-wide LHC computing grid". PhD thesis. RWTH Aachen University, Dec. 2009. URL: http://publications.rwth-aachen.de/record/51431 (cit. on pp. 1, 34, 55).

[5] CMS Collaboration. "MUSiC, a Model Unspecific Search for New Physics, in pp Collisions at $\sqrt{s} = 8$ TeV". CMS PAS EXO-14-016. Mar. 2017. URL: http://cds.cern.ch/record/2256653 (cit. on pp. 2, 96, 98–100, 102, 110, 111, 121, 125, 144, 146, 147, 149, 151).

[6] A. Albert et al. "MUSiC - A Model Unspecific Search for New Physics in pp Collisions at $\sqrt{s} = 8$TeV". CMS AN 2014/098. Jan. 2016. URL: http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2014/098 (cit. on p. 2).

[7] A. Einstein. "On the Relativity Priniciple and the Conclusions Drawn from It (Über das Relativitätsprinzip und die aus demselben gezogene Folgerungen)". In: *Jahrbuch der Radioaktivität und Elektronik* 4 (Dec. 1907), p. 252 (cit. on p. 5).

[8] B. P. Abbott et al. "Observation of Gravitational Waves from a Binary Black Hole Merger". In: *Physical Review Letters* 116.6 (Feb. 2016). DOI: 10.1103/physrevlett.116.061102 (cit. on p. 5).

[9] R. Wolf. "The Higgs Boson discovery at the Large Hadron Collider". 1st edition. Vol. 264. Springer tracts in modern physics. Springer, 2015. ISBN: 978-3-319-18512-5. URL: http://www.springer.com/us/book/9783319185118 (cit. on pp. 5, 10).

[10] G. Altarelli. "Collider Physics within the Standard Model: a Primer". arXiv:1303.2842, CERN-PH-TH-2013-020. Mar. 2013. URL: http://cds.cern.ch/record/1527372 (cit. on pp. 5, 9, 11, 12).

[11] D. Griffiths. "Introduction to Elementary Particles". 2nd edition. Wiley, 2008. ISBN: 978-3-527-40601-2. URL: http://eu.wiley.com/WileyCDA/WileyTitle/productCd-3527406018,subjectCd-PH20.html (cit. on pp. 5, 11).

[12] P. Schmüser. "Feynman-Graphen und Eichtheorien für Experimentalphysiker". 2nd edition. Springer, 1995. ISBN: 978-3-540-58486-5. DOI: 10.1007/978-3-642-57766-6 (cit. on p. 5).

[13] CMS Collaboration. "Search for excited quarks in the photon+jet final state in proton-proton collisions at $\sqrt{s} = 8$ TeV". In: *Physics Letters B* 738 (June 2014). CMS PAS EXO-13-003, 274–293. DOI: 10.1016/j.physletb.2014.09.048 (cit. on p. 5).

[14] ATLAS Collaboration. "Search for excited electrons and muons in $\sqrt{s} = 8$ TeV proton–proton collisions with the ATLAS detector". In: *New Journal of Physics* 15.9 (Sept. 2013). CERN-PH-EP-2013-131, p. 093011. DOI: 10.1088/1367-2630/15/9/093011 (cit. on p. 5).

[15] CMS Collaboration. "Search for excited leptons in proton proton collisions at $\sqrt{s} = 8$ TeV". CMS PAS EXO-14-015. Mar. 2015. URL: http://cds.cern.ch/record/2001655 (cit. on p. 5).

[16] K.A. Olive et al. (Particle Data Group). "Review of Particle Physics". In: *Chinese Physics C* 38.9 (Aug. 2014), p. 090001. DOI: 10.1088/1674-1137/38/9/090001 (cit. on pp. 6, 10).

[17] ATLAS and CMS Collaborations. "Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments". In: *Physical Review Letters* 114.19 (May 2015). DOI: 10.1103/physrevlett.114.191803 (cit. on pp. 6, 10).

[18] H. Harari. "Three Generations of Quarks and Leptons". In: *Proceedings of the V International Conference on Experimental Meson Spectroscopy: Boston, Mass., April 29-30, 1977*. July 1977, p. 0170. URL: http://www-public.slac.stanford.edu/sciDoc/docMeta.aspx?slacPubNumber=SLAC-PUB-1974 (cit. on p. 5).

[19] Super-Kamiokande Collaboration. "Evidence for an Oscillatory Signature in Atmospheric Neutrino Oscillations". In: *Physical Review Letters* 93.10 (Sept. 2004). DOI: 10.1103/physrevlett.93.101801 (cit. on p. 5).

[20] GNO Collaboration. "Complete results for five years of GNO solar neutrino observations". In: *Physics Letters B* 616.3-4 (June 2005), 174–190. DOI: 10.1016/j.physletb.2005.04.068 (cit. on p. 5).

[21] KamLAND Collaboration. "Measurement of Neutrino Oscillation with KamLAND: Evidence of Spectral Distortion". In: *Physical Review Letters* 94.8 (Mar. 2005). DOI: 10.1103/physrevlett.94.081801 (cit. on p. 5).

[22] M. H. Ahn et al. "Indications of Neutrino Oscillation in a 250 km Long-Baseline Experiment". In: *Physical Review Letters* 90.4 (Jan. 2003). DOI: 10.1103/physrevlett.90.041801 (cit. on p. 5).

[23] V. N. Aseev et al. "Upper limit on the electron antineutrino mass from the Troitsk experiment". In: *Physical Review D* 84.11 (Dec. 2011). DOI: 10.1103/physrevd.84.112003 (cit. on p. 5).

[24] C. Kraus et al. "Final Results from phase II of the Mainz Neutrino Mass Search in Tritium β Decay". In: *European Physical Journal C* C40 (Apr. 2005), pp. 447–468. DOI: 10.1140/epjc/s2005-02139-7 (cit. on p. 5).

[25] K. Assamagan et al. "Upper limit of the muon-neutrino mass and charged-pion mass from momentum analysis of a surface muon beam". In: *Physical Review D* 53.11 (June 1996), 6065–6077. DOI: 10.1103/physrevd.53.6065 (cit. on p. 5).

[26] ALEPH Collaboration. "An upper limit on the tau-neutrino mass from three-prongand five-prong tau decays". In: *The European Physical Journal C* 2.3 (Apr. 1998), p. 395. DOI: 10.1007/s100520050149 (cit. on p. 5).

[27] D. D. Ryutov. "Using plasma physics to weigh the photon". In: *Plasma Physics and Controlled Fusion* 49.12B (Nov. 2007), B429–B438. DOI: 10.1088/0741-3335/49/12b/s40 (cit. on p. 8).

[28] S. L. Glashow. "Partial-symmetries of weak interactions". In: *Nuclear Physics* 22.4 (Feb. 1961), 579–588. DOI: 10.1016/0029-5582(61)90469-2. URL: http://dx.doi.org/10.1016/0029-5582(61)90469-2 (cit. on pp. 9, 11).

[29] S. Weinberg. "A Model of Leptons". In: *Physical Review Letters* 19.21 (Nov. 1967), 1264–1266. DOI: 10.1103/physrevlett.19.1264 (cit. on pp. 9, 11).

[30] A. Salam. "Weak and Electromagnetic Interactions". In: *Proceedings, 8th Nobel Symposium Lerum, Sweden, May 19-25, 1968*. May 1968, pp. 367–377. URL: http://inspirehep.net/record/53083 (cit. on pp. 9, 11).

[31] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Physical Review Letters* 13.9 (Aug. 1964), 321–323. DOI: 10.1103/physrevlett.13.321 (cit. on p. 9).

[32] P. Higgs. "Broken symmetries, massless particles and gauge fields". In: *Physics Letters* 12.2 (Sept. 1964), 132–133. DOI: 10.1016/0031-9163(64)91136-9 (cit. on p. 9).

[33] P. W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Physical Review Letters* 13.16 (Oct. 1964), 508–509. DOI: 10.1103/physrevlett.13.508. URL: http://dx.doi.org/10.1103/PhysRevLett.13.508 (cit. on p. 9).

[34] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (Sept. 2012), 1–29. DOI: 10.1016/j.physletb.2012.08.020 (cit. on p. 10).

[35] CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716.1 (Sept. 2012), 30–61. DOI: 10.1016/j.physletb.2012.08.021 (cit. on p. 10).

[36] CMS Collaboration. "Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV ". In: *Journal of High Energy Physics* 2013.6 (June 2013). DOI: 10.1007/jhep06(2013)081 (cit. on p. 10).

[37] D. E. Nagle. "The Delta: The First Pion Nucleon Resonance, Its Discovery and Applications". LALP-84-27. Los Alamos National Laboratory. July 1984. URL: http://www.osti.gov/accomplishments/documents/fullText/ACC0011.pdf (cit. on p. 10).

[38] LHCb Collaboration. "Observation of the Resonant Character of the Z(4430)-State". In: *Physical Review Letters* 112.22 (June 2014), p. 222002. DOI: 10.1103/physrevlett.112.222002 (cit. on p. 10).

[39] LHCb Collaboration. "Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays". In: *Physical Review Letters* 115.CERN-PH-EP-2015-153 (July 2015), p. 072001. DOI: 10.1103/PhysRevLett.115.072001 (cit. on p. 10).

[40] F. Ynduráin. "Limits on the mass of the gluon". In: *Physics Letters B* 345.4 (Feb. 1995), 524–526. DOI: 10.1016/0370-2693(94)01677-5 (cit. on p. 10).

[41] T. Aoyama et al. "Tenth-order electron anomalous magnetic moment: Contribution of diagrams without closed lepton loops". In: *Physical Review D* 91.3 (Feb. 2015). DOI: 10.1103/physrevd.91.033006 (cit. on p. 11).

[42] K. Hagiwara et al. "Improved predictions for of the muon and $\alpha_{\mathrm{QED}}(M_Z^2)$". In: *Physics Letters B* 649.2-3 (May 2007), 173–179. DOI: 10.1016/j.physletb.2007.04.012 (cit. on pp. 11, 12).

[43] D. Hanneke, S. Fogwell Hoogerheide, and G. Gabrielse. "Cavity control of a single-electron quantum cyclotron: Measuring the electron magnetic moment". In: *Physical Review A* 83.5 (May 2011). DOI: 10.1103/physreva.83.052122 (cit. on p. 11).

[44] Muon g-2 Collaboration. "Final report of the E821 muon anomalous magnetic moment measurement at BNL". In: *Physical Review D* 73.7 (Apr. 2006). DOI: 10.1103/physrevd.73.072003. URL: http://dx.doi.org/10.1103/PhysRevD.73.072003 (cit. on p. 11).

[45] LEP Electroweak Working Group. "Electroweak measurements in electron-positron collisions at W-boson-pair energies at LEP". In: *Physics Reports* 532.4 (Nov. 2013), 119–244. DOI: 10.1016/j.physrep.2013.07.004 (cit. on p. 11).

[46] NuTeV Collaboration. "Precise Determination of Electroweak Parameters in Neutrino-Nucleon Scattering". In: *Physical Review Letters* 88.9 (Feb. 2002). DOI: 10.1103/physrevlett.88.091802 (cit. on p. 11).

[47] M. Y. Kuchiev and V. V. Flambaum. "Radiative corrections to parity-non-conservation in atoms". arXiv:hep-ph/0305053. May 2003. URL: http://arxiv.org/abs/hep-ph/0305053 (cit. on p. 11).

[48] S. L. Glashow, J. Iliopoulos, and L. Maiani. "Weak Interactions with Lepton-Hadron Symmetry". In: *Physical Review D* 2.7 (Oct. 1970), 1285–1292. DOI: 10.1103/physrevd.2.1285 (cit. on p. 11).

[49] J. E. Augustin et al. "Discovery of a Narrow Resonance in e+e- Annihilation". In: *Physical Review Letters* 33.23 (Dec. 1974), 1406–1408. DOI: 10.1103/physrevlett.33.1406 (cit. on p. 11).

[50] J. J. Aubert et al. "Experimental Observation of a Heavy Particle J". In: *Physical Review Letters* 33.23 (Dec. 1974), 1404–1406. DOI: 10.1103/physrevlett.33.1404 (cit. on p. 11).

[51] M. Gell-Mann. "Symmetries of Baryons and Mesons". In: *Physical Review* 125.3 (Feb. 1962), 1067–1084. DOI: 10.1103/physrev.125.1067 (cit. on p. 11).

[52] B. R. Stella and H.-J. Meyer. "Υ(9.46 GeV) and the gluon discovery (a critical recollection of PLUTO results)". In: *The European Physical Journal H* 36.2 (Sept. 2011), 203–243. DOI: 10.1140/epjh/e2011-10029-3 (cit. on p. 11).

[53] UA1 Collaboration. "Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s}$=540 GeV". In: *Physics Letters B* 122.1 (Feb. 1983), 103–116. DOI: 10.1016/0370-2693(83)91177-2 (cit. on p. 11).

[54] UA2 Collaboration. "Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider". In: *Physics Letters B* 122.5-6 (Mar. 1983), 476–485. DOI: 10.1016/0370-2693(83)91605-2 (cit. on p. 11).

[55] M. Kobayashi and T. Maskawa. "CP-Violation in the Renormalizable Theory of Weak Interaction". In: *Progress of Theoretical Physics* 49.2 (Feb. 1973), 652–657. DOI: 10.1143/ptp.49.652 (cit. on p. 11).

[56] CDF Collaboration. "Observation of Top Quark Production in $p\bar{p}$ Collisions with the Collider Detector at Fermilab". In: *Physical Review Letters* 74.14 (Apr. 1995), 2626–2631. DOI: 10.1103/physrevlett.74.2626 (cit. on p. 11).

[57] D0 Collaboration. "Search for High Mass Top Quark Production in $p\bar{p}$ Collisions at $\sqrt{s} = 1.8$ TeV". In: *Physical Review Letters* 74.13 (Mar. 1995), 2422–2426. DOI: 10.1103/physrevlett.74.2422 (cit. on p. 11).

[58] G. Altarelli. "The Higgs and the Excessive Success of the Standard Model". arXiv:1407.2122, CERN-PH-TH/2014-127. July 2014. URL: http://arxiv.org/abs/1407.2122 (cit. on p. 11).

[59] C. Patrignani et al. (Particle Data Group). "Review of Particle Physics". In: *Chinese Physics C* 40.10 (Oct. 2016), p. 100001. DOI: 10.1088/1674-1137/40/10/100001 (cit. on pp. 11, 12, 42, 48, 52, 60, 62).

[60] G. Hinshaw et al. "Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results". In: *The Astrophysical Journal Supplement Series* 208.2 (Sept. 2013), p. 19. DOI: 10.1088/0067-0049/208/2/19 (cit. on p. 11).

[61] C. L. Bennett et al. "Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results". In: *The Astrophysical Journal Supplement Series* 208.2 (Sept. 2013), p. 20. DOI: 10.1088/0067-0049/208/2/20 (cit. on p. 11).

[62] Planck Collaboration. "Planck 2015 results". In: *Astronomy & Astrophysics* 594 (Sept. 2016), A13. DOI: 10.1051/0004-6361/201525830 (cit. on p. 11).

[63] Supernova Search Team Collaboration. "Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant". In: *The Astronomical Journal* 116.3 (Sept. 1998), 1009–1038. DOI: 10.1086/300499 (cit. on p. 11).

[64] Supernova Cosmology Project Collaboration. "Measurements of Ω and Λ from 42 High-Redshift Supernovae". In: *The Astrophysical Journal* 517.2 (June 1999), 565–586. DOI: 10.1086/307221 (cit. on p. 11).

[65] A. Einstein. "Cosmological considerations on the general theory of relativity (Kosmologische Betrachtungen zur allgemeinen Relativitätsprinziptstheorie)". In: *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften* (Feb. 1917), pp. 142–152. URL: http://adsabs.harvard.edu/abs/1917SPAW.....142E (cit. on p. 11).

[66] F. Zwicky. "On the Masses of Nebulae and of Clusters of Nebulae". In: *The Astrophysical Journal* 86 (Oct. 1937), p. 217. DOI: 10.1086/143864 (cit. on p. 11).

[67] V. C. Rubin and W. K. Ford Jr. "Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions". In: *Astrophysical Journal* 159 (Feb. 1970), p. 379. DOI: 10.1086/150317 (cit. on p. 11).

[68] V. C. Rubin, W. K. Ford Jr., and N. Thonnard. "Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 /R = 4kpc/ to UGC 2885 /R = 122 kpc/". In: *Astrophysical Journal* 238 (June 1980), pp. 471–487. DOI: 10.1086/158003 (cit. on p. 11).

[69] Muon g-2 Collaboration. "Muon (g-2) Technical Design Report". FERMILAB-FN-0992-E. Jan. 2015. URL: http://arxiv.org/abs/1501.06858 (cit. on p. 12).

[70] J. H. Christenson et al. "Evidence for the 2π Decay of the K_2^0 Meson". In: *Physical Review Letters* 13.4 (July 1964), 138–140. DOI: 10.1103/physrevlett.13.138 (cit. on p. 12).

[71] S. P. Martin. "A Supersymmetry Primer". In: *Advanced Series on Directions in High Energy Physics* (July 1998), 1–98. DOI: 10.1142/9789812839657_0001 (cit. on p. 12).

[72] L. Evans and P. Bryant. "LHC Machine". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08001–S08001. DOI: 10.1088/1748-0221/3/08/s08001 (cit. on p. 15).

[73] M. Lamont. "Status of the LHC". In: *Journal of Physics: Conference Series* 455 (Aug. 2013), p. 012001. DOI: 10.1088/1742-6596/455/1/012001 (cit. on p. 15).

[74] G. Arduini. "Performance Reach in the LHC for 2012". In: *Proceedings of the 2012 Workshop on LHC Performance: Chamonix, France, February 6-10, 2012*. CERN-2012-006. Feb. 2012. URL: http://cds.cern.ch/record/1424362 (cit. on pp. 15, 16).

[75] S. Myers. "The Large Hadron Collider 2008–2013". In: *International Journal of Modern Physics A* 28.25 (Oct. 2013), p. 1330035. DOI: 10.1142/s0217751x13300354 (cit. on p. 15).

[76] C. Lefèvre. "The CERN accelerator complex". CERN-DI-0812015. Dec. 2008. URL: https://cds.cern.ch/record/1260465 (cit. on p. 16).

[77] J.-L. Caron. "LHC Layout". LHC-PHO-1997-060. Sept. 1997. URL: http://cds.cern.ch/record/841573 (cit. on p. 16).

[78] ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08003–S08003. DOI: 10.1088/1748-0221/3/08/s08003 (cit. on p. 15).

[79] CMS Collaboration. "The CMS experiment at the CERN LHC". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004 (cit. on pp. 15, 18, 20, 22, 23, 39).

[80] ALICE Collaboration. "The ALICE experiment at the CERN LHC". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08002–S08002. DOI: 10.1088/1748-0221/3/08/s08002 (cit. on p. 15).

[81] LHCb Collaboration. "The LHCb Detector at the LHC". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08005–S08005. DOI: 10.1088/1748-0221/3/08/s08005 (cit. on p. 15).

[82] W. Herr and B. Muratori. "Concept of luminosity". In: *Proceedings, CERN Accelerator School: Intermediate Course on Accelerator Physics, Zeuthen, Germany, September 15-26, 2003*. CERN-2006-002. Jan. 2006. DOI: 10.5170/CERN-2006-002.361 (cit. on p. 15).

[83] Talk by F. Zimmermann at the SLAC Summer Institute, July 23nd - August 3, 2012. "LHC: The Machine". URL: https://www-conf.slac.stanford.edu/ssi/2012/Presentations/Zimmermann.pdf (visited on 09/25/2016) (cit. on p. 16).

[84] M. Hostettler and G. Papotti. "Luminosity Lifetime at the LHC in 2012 Proton Physics Operation". In: *Proceedings of the 4th International Particle Accelerator Conference (IPAC 2013): Shanghai, China, May 12-13, 2013*. May 2013. URL: http://JACoW.org/IPAC2013/papers/tupfi029.pdf (cit. on p. 16).

[85] CMS Public Twiki. "Public CMS Luminosity Information". URL: http://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults (visited on 07/08/2016) (cit. on pp. 16, 53, 63).

[86] F. Bordry and K. Foraz. "Summary of Session 5 and 6: "Long Shutdown 1" (LS1) 2013-2014". In: *Proceedings of the 2012 Workshop on LHC Performance: Chamonix, France, February 6-10, 2012*. CERN-2012-006. Feb. 2012. URL: http://cds.cern.ch/record/1424362 (cit. on p. 16).

[87] Talk by R. Lindner at the LHC Performance Workshop (Chamonix 2016), January 28, 2016. "LHC Experiments Activities during LS2". URL: http://indico.cern.ch/event/448109/contributions/1942025/attachments/1219237/1781524/Chamonix_LHCexp_2016-RL.pdf (visited on 07/08/2016) (cit. on p. 16).

[88] CMS Collaboration. "Technical proposal for the upgrade of the CMS detector through 2020". CERN-LHCC-2011-006; LHCC-P-004. June 2011. URL: https://cds.cern.ch/record/1355706 (cit. on pp. 16, 22).

[89] T. Sakuma and T. McCauley. "Detector and Event Visualization with SketchUp at the CMS Experiment". In: *Journal of Physics: Conference Series* 513.2 (June 2014), p. 022032. DOI: 10.1088/1742-6596/513/2/022032 (cit. on p. 17).

[90] B. R. Vormwald. "The CMS inner tracker – transition from LHC Run I to Run II and first experience of Run II". CMS CR-2015/213. Oct. 2015. URL: https://cds.cern.ch/record/2059826 (cit. on p. 18).

[91] CMS Collaboration. "The CMS tracker system project: Technical Design Report". CERN-LHCC-98-006; CMS-TDR-5. Apr. 1997. URL: http://cds.cern.ch/record/368412 (cit. on pp. 19, 22).

[92] R. Brown and D. Cockerill. "Electromagnetic calorimetry". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 666 (Feb. 2012), 47–79. DOI: 10.1016/j.nima.2011.03.017 (cit. on p. 19).

[93] CMS HCAL Collaborations. "Design, performance, and calibration of the CMS hadron-outer calorimeter". In: *The European Physical Journal C* 57.3 (Oct. 2008). CERN-CMS-NOTE-2008-020, 653–663. DOI: 10.1140/epjc/s10052-008-0756-6 (cit. on p. 20).

[94] CMS Collaboration. "The CMS hadron calorimeter project: Technical Design Report". CERN-LHCC-97-03; CMS-TDR-2. June 1997. URL: http://cds.cern.ch/record/357153 (cit. on p. 21).

[95] CMS Collaboration. "CMS Physics: Technical Design Report Volume 1: Detector Performance and Software". CERN-LHCC-2006-001; CMS-TDR-8-1. Feb. 2006. URL: http://cds.cern.ch/record/922757 (cit. on pp. 21, 30).

[96] CMS Collaboration. "The CMS magnet project: Technical Design Report". CERN-LHCC-97-010; CMS-TDR-1. May 1997. URL: http://cds.cern.ch/record/331056 (cit. on p. 21).

[97] CMS Collaboration. "Precise mapping of the magnetic field in the CMS barrel yoke using cosmic rays". In: *Journal of Instrumentation* 5.03 (Mar. 2010). CMS PAS CFT-09-015, T03021–T03021. DOI: 10.1088/1748-0221/5/03/t03021 (cit. on p. 21).

[98] CMS Collaboration. "The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s} = 7$ TeV at the LHC". In: *Journal of Instrumentation* 8.11 (Nov. 2013). CMS PAS MUO-11-001, P11002–P11002. DOI: 10.1088/1748-0221/8/11/p11002 (cit. on pp. 21, 22).

[99] CMS Collaboration. "Performance of the CMS drift tube chambers with cosmic rays". In: *Journal of Instrumentation* 5.03 (Mar. 2010). CMS PAS CFT-09-012, T03015–T03015. DOI: 10.1088/1748-0221/5/03/t03015 (cit. on p. 21).

[100] CMS Collaboration. "Performance of the CMS cathode strip chambers with cosmic rays". In: *Journal of Instrumentation* 5.03 (Mar. 2010). CMS PAS CFT-09-011, T03018–T03018. DOI: 10.1088/1748-0221/5/03/t03018 (cit. on p. 22).

[101] CMS Collaboration. "Performance study of the CMS barrel resistive plate chambers with cosmic rays". In: *Journal of Instrumentation* 5.03 (Mar. 2010). CMS PAS CFT-09-010, T03017–T03017. DOI: 10.1088/1748-0221/5/03/t03017 (cit. on p. 22).

[102] CMS Collaboration. "Measurement of Momentum Scale and Resolution of the CMS Detector using Low-mass Resonances and Cosmic Ray Muons". CMS PAS TRK-10-004. July 2010. URL: http://cds.cern.ch/record/1279137 (cit. on pp. 22, 80).

[103] CMS Collaboration. "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV". In: *Journal of Instrumentation* 7.10 (Oct. 2012). CMS PAS MUO-10-004, P10002–P10002. DOI: 10.1088/1748-0221/7/10/p10002 (cit. on pp. 22, 25, 27, 72).

[104] D. Trocino. "The CMS High Level Trigger". In: *Journal of Physics: Conference Series* 513.1 (June 2014), p. 012036. DOI: 10.1088/1742-6596/513/1/012036 (cit. on pp. 23, 70).

[105] CMS Collaboration. "CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems". CERN-LHCC-2000-038; CMS-TDR-6.1. Dec. 2000. URL: http://cds.cern.ch/record/706847 (cit. on p. 23).

[106] CMS Collaboration. "CMS TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger". CERN-LHCC-2002-026; CMS-TDR-6.2. Dec. 2002. URL: http://cds.cern.ch/record/578006 (cit. on p. 23).

[107] I. Bird et al. "Update of the Computing Models of the WLCG and the LHC Experiments". CERN-LHCC-2014-014; LCG-TDR-002. Apr. 2014. URL: http://cds.cern.ch/record/1695401 (cit. on p. 23).

[108] CMS Collaboration. "Absolute Calibration of Luminosity Measurement at CMS: Summer 2011 Update". CMS PAS EWK-11-001. Aug. 2011. URL: http://cds.cern.ch/record/1376102 (cit. on p. 24).

[109] CMS Collaboration. "CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update". CMS PAS LUM-13-001. Sept. 2013. URL: http://cds.cern.ch/record/1598864 (cit. on pp. 24, 63, 84).

[110] S. van der Meer. "Calibration of the effective beam height in the ISR". CERN-ISR-PO-68-31. June 1968. URL: http://cds.cern.ch/record/296752 (cit. on p. 24).

[111] CMS Twiki. "Calibration and Alignment". URL: http://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideCalAli?rev=160 (visited on 09/27/2016) (cit. on p. 24).

[112] CMS Collaboration. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". In: *Journal of Instrumentation* 9.10 (Oct. 2014). CMS PAS TRK-11-001, P10009. DOI: 10.1088/1748-0221/9/10/P10009 (cit. on pp. 24, 25).

[113] R. Frühwirth. "Application of Kalman filtering to track and vertex fitting". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 262.2-3 (Dec. 1987), 444–450. DOI: 10.1016/0168-9002(87)90887-4 (cit. on p. 25).

[114] W. Erdmann. "Offline Primary Vertex Reconstruction with Deterministic Annealing Clustering". CMS IN-2011/014. June 2011. URL: http://cms.cern.ch/iCMS/user/noteinfo?cmsnoteid=CMS%20IN-2011/014 (cit. on pp. 25, 67).

[115] CMS Collaboration. "Tracking and Primary Vertex Results in First 7 TeV Collisions". CMS PAS TRK-10-005. July 2010. URL: http://cds.cern.ch/record/1279383 (cit. on pp. 25, 67).

[116] CMS Collaboration. "Performance of CMS muon reconstruction in cosmic-ray events". In: *Journal of Instrumentation* 5.03 (Mar. 2010). CMS PAS CFT-09-014, T03022–T03022. DOI: 10.1088/1748-0221/5/03/t03022 (cit. on pp. 25, 83).

[117] M. Chen et al. "Search for New High-Mass Resonances Decaying to Muon Pairs in the CMS Experiment". CMS AN-2007/038. Mar. 2008. URL: http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2007/038 (cit. on p. 25).

[118] G. Abbiendi et al. "Muon Reconstruction in the CMS Detector". CMS AN-2008/097. July 2009. URL: http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2008/097 (cit. on p. 25).

[119] S. Baffioni et al. "Electron reconstruction in CMS". In: *The European Physical Journal C* 49.4 (Jan. 2007). CMS NOTE 2006/40, 1099–1116. DOI: 10.1140/epjc/s10052-006-0175-5 (cit. on p. 26).

[120] M. Anderson et al. "Review of clustering algorithms and energy corrections in ECAL". CMS IN-2010/008. Mar. 2010. URL: http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20IN-2010/008 (cit. on p. 26).

[121] W. Adam et al. "Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC". CMS-NOTE-2005-001. Jan. 2005. URL: http://cds.cern.ch/record/815410 (cit. on p. 27).

[122] W. Adam et al. "Electron Reconstruction in CMS". CMS AN-2009/164. Jan. 2010. URL: http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2009/164 (cit. on p. 27).

[123] N. Marinelli. "Track finding and identification of converted photons". CMS-NOTE-2006-005. Jan. 2006. URL: http://cds.cern.ch/record/927374 (cit. on p. 27).

[124] N. Marinelli et al. "Track finding and identification of converted photons with the CMS Tracker and ECAL". CMS AN-2008/102. Nov. 2008. URL: http://cms.cern.ch/iCMS/user/noteinfo?cmsnoteid=CMS%20AN-2008/102 (cit. on p. 27).

[125] CMS Collaboration. "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET". CMS PAS PFT-09-001. Apr. 2009. URL: http://cds.cern.ch/record/1194487 (cit. on p. 27).

[126] F. Beaudette. "The CMS Particle Flow Algorithm". In: *Proceedings of the International Conference on Calorimetry for the High Energy Frontier (CHEF 2013): Paris, France, April 22-25, 2013*. Jan. 2014. URL: http://inspirehep.net/record/1279774 (cit. on p. 27).

[127] CMS Collaboration. "Particle-flow commissioning with muons and electrons from J/Psi and W events at 7 TeV". CMS PAS PFT-10-003. July 2010. URL: http://cds.cern.ch/record/1279347 (cit. on p. 27).

[128] F. Beaudette et al. "Electron Reconstruction within the Particle Flow Algorithm". CMS AN-2010/034. Feb. 2010. URL: http://cms.cern.ch/iCMS/user/noteinfo?cmsnoteid=CMS%20AN-2010/034 (cit. on p. 27).

[129] B. Webber. "Fragmentation and Hadronization". In: *International Journal of Modern Physics A* 15.supp01b (July 2000), 577–606. DOI: 10.1142/s0217751x00005334 (cit. on p. 28).

[130] T. Sjöstrand, S. Mrenna, and P. Skands. "PYTHIA 6.4 physics and manual". In: *Journal of High Energy Physics* 2006.05 (May 2006), pp. 026–026. DOI: 10.1088/1126-6708/2006/05/026 (cit. on pp. 28, 54).

[131] CMS Collaboration. "Performance of Jet Algorithms in CMS". CMS PAS JME-07-003. Aug. 2009. URL: http://cds.cern.ch/record/1198227 (cit. on p. 28).

[132] A. Banfi. "Hadronic Jets: An introduction". 1st edition. Morgan & Claypool Publishers, 2016. ISBN: http://id.crossref.org/isbn/978-1-6817-4073-7. DOI: 10.1088/978-1-6817-4073-7 (cit. on p. 28).

[133] M. Cacciari, G. P. Salam, and G. Soyez. "The anti-k_T jet clustering algorithm". In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), 063–063. DOI: 10.1088/1126-6708/2008/04/063 (cit. on p. 28).

[134] M. Cacciari, G. P. Salam, and G. Soyez. "FastJet user manual". In: *The European Physical Journal C* 72.3 (Mar. 2012). DOI: 10.1140/epjc/s10052-012-1896-2 (cit. on pp. 28, 71).

[135] CMS Collaboration. "Pileup Removal Algorithms". CMS PAS JME-14-001. Aug. 2014. URL: http://cds.cern.ch/record/1751454 (cit. on p. 28).

[136] CMS Collaboration. "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV ". CMS PAS JME-13-004. June 2016. URL: http://cds.cern.ch/record/2198719 (cit. on pp. 28, 29, 39, 63, 75, 80).

[137] CMS Twiki. "Jet Energy Resolution". URL: http://twiki.cern.ch/twiki/bin/viewauth/CMS/JetResolution?rev=32 (visited on 11/24/2014) (cit. on pp. 29, 75).

[138] CMS Collaboration. "Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ TeV". In: *Journal of Instrumentation* 10.02 (Feb. 2015). CMS PAS-JME-13-003, P02006–P02006. DOI: 10.1088/1748-0221/10/02/p02006 (cit. on pp. 29, 39, 67).

[139] G. L. Bayatyan et al. "CMS computing: Technical Design Report". CERN-LHCC-2005-023; CMS-TDR-7. May 2005. URL: http://cds.cern.ch/record/838359 (cit. on pp. 30, 35).

[140] CMS Collaboration. "LHC computing Grid: Technical Design Report". CERN-LHCC-2005-024. June 2005. URL: http://cds.cern.ch/record/840543 (cit. on p. 30).

[141] F. Fabozzi et al. "Physics Analysis Tools for the CMS Experiment at LHC". In: *IEEE Transactions on Nuclear Science* 55.6 (Dec. 2008), 3539–3543. DOI: 10.1109/tns.2008.2006979 (cit. on p. 30).

[142] R. Brun and F. Rademakers. "ROOT - An object oriented data analysis framework". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 389.1-2 (Apr. 1997), 81–86. DOI: 10.1016/s0168-9002(97)00048-x (cit. on pp. 30, 35).

[143] W Adam et al. "PAT: The CMS Physics Analysis Toolkit". In: *Journal of Physics: Conference Series* 219.3 (Apr. 2010), p. 032017. DOI: 10.1088/1742-6596/219/3/032017 (cit. on p. 30).

[144] CMS Twiki. "Physics Analysis Toolkit (PAT)". URL: http://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuidePAT?rev=90 (visited on 08/28/2014) (cit. on p. 30).

[145] T. Hebbeker. "A Global Comparison between L3 Data and Standard Model Monte Carlo - a first attempt". July 1998. URL: http://web.physik.rwth-aachen.de/~hebbeker/l3note_2305.pdf (cit. on p. 33).

[146] D0 Collaboration. "Search for new physics in $e\mu X$ data at DØ using SLEUTH: A quasi-model-independent search strategy for new physics". In: *Physical Review D* 62.9 (Oct. 2000). DOI: 10.1103/physrevd.62.092004 (cit. on p. 33).

[147] D0 Collaboration. "Quasi-model-independent search for new physics at large transverse momentum". In: *Physical Review D* 64.1 (June 2001). DOI: 10.1103/physrevd.64.012004 (cit. on p. 33).

[148] D0 Collaboration. "Quasi-Model-Independent Search for New High p_T Physics at D0". In: *Physical Review Letters* 86 (Apr. 2001), pp. 3712–3717. DOI: 10.1103/PhysRevLett.86.3712 (cit. on p. 33).

[149] D0 Collaboration. "Search for New Physics Using QUAERO: A General Interface to D0 Event Data". In: *Physical Review Letters* 87.23 (Nov. 2001). DOI: 10.1103/physrevlett.87.231801 (cit. on p. 33).

[150] H1 Collaboration. "A general search for new phenomena in ep scattering at HERA". In: *Physics Letters B* 602.1-2 (Nov. 2004), 14–30. DOI: 10.1016/j.physletb.2004.09.057 (cit. on pp. 33, 42).

[151] H1 Collaboration. "A general search for new phenomena at HERA". In: *Physics Letters B* 674.4-5 (Apr. 2009), 257–268. DOI: 10.1016/j.physletb.2009.03.034 (cit. on p. 33).

[152] CDF Collaboration. "Model-independent and quasi-model-independent search for new physics at CDF". In: *Physical Review D* 78.1 (June 2008). DOI: 10.1103/physrevd.78.012002 (cit. on p. 34).

[153] CDF Collaboration. "Global Search for New Physics with 2.0 fb^{-1} at CDF". In: *Physical Review D* 79 (Jan. 2009), p. 011101. DOI: 10.1103/PhysRevD.79.011101 (cit. on p. 34).

[154] D0 Collaboration. "Model independent search for new phenomena in $p\bar{p}$ collisions at $\sqrt{s} =$ 1.96 TeV ". In: *Physical Review D* 85.9 (May 2012). DOI: 10.1103/physrevd.85.092015 (cit. on p. 34).

[155] CMS Collaboration. "MUSIC – An Automated Scan for Deviations between Data and Monte Carlo Simulation". CMS-PAS-EXO-08-005. Oct. 2008. URL: http://cds.cern.ch/record/1152572 (cit. on p. 34).

[156] CMS Collaboration. "Model Unspecific Search for New Physics in pp Collisions at $\sqrt{s} = 7$ TeV". CMS PAS EXO-10-021. June 2011. URL: http://cds.cern.ch/record/1360173 (cit. on pp. 34, 43).

[157] ATLAS Collaboration. "A general search for new phenomena with the ATLAS detector in pp collisions at $\sqrt{s} = 7$ TeV". ATLAS-CONF-2012-107. Aug. 2012. URL: http://cds.cern.ch/record/1472686 (cit. on p. 34).

[158] R. M. Bianchi. "A model-independent "General Search" for new physics with the ATLAS detector at LHC". PhD thesis. Albert-Ludwigs-Universität Freiburg, Mar. 2014. URL: http://cds.cern.ch/record/1757501 (cit. on p. 34).

[159] ATLAS Collaboration. "A general search for new phenomena with the ATLAS detector in pp collisions at $\sqrt{s} = 8$ TeV". ATLAS-CONF-2014-006. Mar. 2014. URL: http://cds.cern.ch/record/1666536 (cit. on p. 34).

[160] S. Amoroso. "Tuning of event generators to measurements of $t\bar{t}$ production and a general search for new physics with the ATLAS experiment". PhD thesis. Albert-Ludwigs-Universität Freiburg, July 2015. URL: http://cds.cern.ch/record/2056368 (cit. on p. 34).

[161] ATLAS Collaboration. "A model independent general search for new phenomena with the ATLAS detector at $\sqrt{s} = 13$ TeV". ATLAS-CONF-2017-001. Jan. 2017. URL: http://cds.cern.ch/record/2243494 (cit. on p. 34).

[162] S. Knutzen. "A software for the reinterpretation of model independent search results and constraining theories beyond the Standard Model". In preparation. PhD thesis. RWTH Aachen University, 2017 (cit. on pp. 35, 157, 158).

[163] G Petrucciani, A Rizzi, and C Vuosalo. "Mini-AOD: A New Analysis Data Format for CMS". In: *Journal of Physics: Conference Series* 664.7 (Dec. 2015), p. 072052. DOI: 10.1088/1742-6596/664/7/072052 (cit. on p. 35).

[164] D. Spiga et al. "The CMS Remote Analysis Builder (CRAB)". In: *High Performance Computing – HiPC 2007* 4873 (Dec. 2007), pp. 580–586. DOI: 10.1007/978-3-540-77220-0_52 (cit. on p. 35).

[165] H.-P. Bretz et al. "A Development Environment for Visual Physics Analysis". In: *Journal of Instrumentation* 7.08 (Aug. 2012), T08005. DOI: 10.1088/1748-0221/7/08/t08005 (cit. on p. 35).

[166] CMS Twiki. "PAT Release Notes (compatible with CMSSW_5_2_X / CMSSW_5_3_X)". URL: http://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuidePATReleaseNotes52X?rev=189 (visited on 10/29/2014) (cit. on p. 35).

[167] C. Aiftimiei et al. "Using CREAM and CEMonitor for job submission and management in the gLite middleware". In: *Journal of Physics: Conference Series* 219.6 (Apr. 2010), p. 062001. DOI: 10.1088/1742-6596/219/6/062001 (cit. on p. 35).

[168] D. Thain, T. Tannenbaum, and M. Livny. "Distributed computing in practice: the Condor experience". In: *Concurrency - Practice and Experience* 17.2-4 (Feb. 2005), 323–356. DOI: 10.1002/cpe.938 (cit. on p. 35).

[169] T. Aaltonen et al. "Studying the underlying event in Drell-Yan and high transverse momentum jet production at the Tevatron". In: *Physical Review D* 82.3 (Aug. 2010). DOI: 10.1103/physrevd.82.034001 (cit. on p. 36).

[170] CMS Collaboration. "Measurement of the Jet Multiplicity in dileptonic Top Quark Pair Events at 8 TeV". CMS PAS TOP-12-041. May 2013. URL: http://cds.cern.ch/record/1547532 (cit. on p. 36).

[171] R. Michael Barnett, J. F. Gunion, and H. E. Haber. "Discovering supersymmetry with like-sign dileptons". In: *Physics Letters B* 315.3-4 (Oct. 1993), 349–354. DOI: 10.1016/0370-2693(93)91623-u (cit. on p. 37).

[172] J. Mrazek and A. Wulzer. "A strong sector at the LHC: Top partners in same-sign dileptons". In: *Physical Review D* 81.7 (Apr. 2010). DOI: 10.1103/physrevd.81.075006 (cit. on p. 37).

[173] CMS Collaboration. "Search for Top-Quark Partners with Charge 5 / 3 in the Same-Sign Dilepton Final State". In: *Physical Review Letters* 112.17 (Apr. 2014). CMS PAS B2G-12-012. DOI: 10.1103/physrevlett.112.171801 (cit. on p. 37).

[174] CMS Collaboration. "Search for microscopic black holes in pp collisions at $\sqrt{s} = 8$ TeV". In: *Journal of High Energy Physics* 2013.7 (July 2013). CMS PAS EXO-12-009. DOI: 10.1007/jhep07(2013)178 (cit. on p. 38).

[175] S. Bolognesi et al. "Calibration of track momentum using dimuon resonances in CMS". CMS AN 2010/059. Mar. 2010. URL: http://cms.cern.ch/iCMS/user/noteinfo?cmsnoteid=CMS%20AN-2010/059 (cit. on p. 39).

[176] J. Rice. "Mathematical Statistics and Data Analysis". 3rd edition. Thomson Learning, 2007. ISBN: 978-0-534-39942-9 (cit. on pp. 42, 51).

[177] R. D. Cousins, J. T. Linnemann, and J. Tucker. "Evaluation of three methods for calculating statistical significance when incorporating a systematic uncertainty into a test of the background-only hypothesis for a Poisson process". In: *Nuclear Instruments and Methods in Physics Research Section A* 595.2 (Oct. 2008), 480–501. DOI: 10.1016/j.nima.2008.07.086 (cit. on pp. 42, 43, 50).

[178] A. C. Micheas and D. K. Dey. "Prior and posterior predictive p-values in the one-sided location parameter testing problem". In: *Sankhya: The Indian Journal of Statistics* 65.1 (Feb. 2003), pp. 158–178. URL: http://www.jstor.org/stable/25053252 (cit. on p. 42).

[179] L. Lyons. "Open statistical issues in Particle Physics". In: *The Annals of Applied Statistics* 2.3 (Sept. 2008), 887–915. DOI: 10.1214/08-aoas163 (cit. on p. 42).

[180] S. A. Schmitz. "Model Unspecific Search for New Physics with High p_T Photons in CMS, Diploma thesis". Diploma thesis. RWTH Aachen University, Oct. 2009. URL: http://web.physik.rwth-aachen.de/~hebbeker/theses/schmitz_diploma.pdf (cit. on pp. 43, 50).

[181] M. Olschewski. "Study of Alternative Statistical Methods for a Model Unspecific Search in CMS, Diploma thesis". Diploma thesis. RWTH Aachen University, June 2011. URL: http://web.physik.rwth-aachen.de/~hebbeker/theses/olschewski_diploma.pdf (cit. on p. 43).

[182] E. Gross and O. Vitells. "Trial factors for the look elsewhere effect in high energy physics". In: *The European Physical Journal C* 70.1-2 (Oct. 2010), 525–530. DOI: 10.1140/epjc/s10052-010-1470-8 (cit. on p. 48).

[183] CMS Collaboration. "CMS Physics: Technical Design Report Volume 2: Physics Performance". CERN-LHCC-2006-021; CMS-TDR-8-2. June 2007, 995–1579. 669 p. URL: http://cds.cern.ch/record/942733 (cit. on pp. 48, 52).

[184] J. Lieb. "Development of a Fast Search Algorithm for the MUSiC Framework". Bachelor thesis. RWTH Aachen University, Sept. 2015. URL: http://web.physik.rwth-aachen.de/~hebbeker/theses/lieb_bachelor.pdf (cit. on p. 50).

[185] J. Lieb. "Discovery Potential of a Model Independent Search for New Physics at the LHC". In preparation. Master thesis. RWTH Aachen University, 2017. URL: http://web.physik.rwth-aachen.de/~hebbeker/theses/lieb_master.pdf (cit. on p. 51).

[186] CMS Twiki. "How to work with files for Good Luminosity Sections in JSON format". URL: http://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideGoodLumiSectionsJSONFile (visited on 07/29/2015) (cit. on p. 53).

[187] CMS website. "CMS Service DQM". URL: http://cms-service-dqm.web.cern.ch/cms-service-dqm/CAF/certification/Collisions12/8TeV/DCSOnly/ (visited on 08/29/2014) (cit. on p. 53).

[188] CMS Twiki. "Global Tags for Conditions Data". URL: http://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideFrontierConditions?rev=449 (visited on 08/28/2014) (cit. on p. 53).

[189] CMS Twiki. "The official CMS Luminosity Calculation Tools". URL: http://twiki.cern.ch/twiki/bin/viewauth/CMS/LumiCalc?rev=186 (visited on 11/21/2014) (cit. on p. 53).

[190] A. Buckley et al. "General-purpose event generators for LHC physics". In: *Physics Reports* 504.5 (July 2011), 145–233. DOI: 10.1016/j.physrep.2011.03.005 (cit. on p. 53).

[191] J. M. Campbell, J. W. Huston, and W. J. Stirling. "Hard interactions of quarks and gluons: a primer for LHC physics". In: *Reports on Progress in Physics* 70.1 (Dec. 2006), 89–193. DOI: 10.1088/0034-4885/70/1/r02 (cit. on pp. 53, 62).

[192] J. Alwall et al. "MadGraph 5: going beyond". In: *Journal of High Energy Physics* 2011.6 (June 2011). ISSN: 1029-8479. DOI: 10.1007/jhep06(2011)128 (cit. on p. 54).

[193] P. Nason. "A New Method for Combining NLO QCD with Shower Monte Carlo Algorithms". In: *Journal of High Energy Physics* 2004.11 (Nov. 2004), 040–040. DOI: 10.1088/1126-6708/2004/11/040 (cit. on p. 54).

[194] S. Frixione, P. Nason, and C. Oleari. "Matching NLO QCD computations with parton shower simulations: the POWHEG method". In: *Journal of High Energy Physics* 2007.11 (Nov. 2007), 070–070. DOI: 10.1088/1126-6708/2007/11/070 (cit. on p. 54).

[195] S. Alioli et al. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX". In: *Journal of High Energy Physics* 2010.6 (June 2010). DOI: 10.1007/jhep06(2010)043 (cit. on p. 54).

[196] T. Gleisberg et al. "Event generation with SHERPA 1.1". In: *Journal of High Energy Physics* 2009.02 (Feb. 2009), 007–007. DOI: 10.1088/1126-6708/2009/02/007 (cit. on pp. 54, 59).

[197] Z. Was. "TAUOLA the library for τ lepton decay, and KKMC/KORALB/KORALZ/... status report". In: *Nuclear Physics B - Proceedings Supplements* 98.1-3 (Apr. 2001), 96–102. DOI: 10.1016/s0920-5632(01)01200-2 (cit. on p. 54).

[198] CMS Collaboration. "Measurement of the Inclusive Jet Cross Section in pp Collisions at 7 TeV". CMS PAS QCD-10-010. July 2010. URL: https://cds.cern.ch/record/1280682 (cit. on p. 54).

[199] P. Z. Skands. "Tuning Monte Carlo generators: The Perugia tunes". In: *Physical Review D* 82.7 (Oct. 2010). DOI: 10.1103/physrevd.82.074018 (cit. on p. 55).

[200] CMS Collaboration. "Underlying Event Tunes and Double Parton Scattering". CMS PAS GEN-14-001. Apr. 2014. URL: https://cds.cern.ch/record/1697700 (cit. on p. 55).

[201] CMS Twiki. "Standard Model Cross Sections for CMS at 8 TeV". URL: http://twiki.cern.ch/twiki/bin/viewauth/CMS/StandardModelCrossSectionsat8TeV?rev=30 (visited on 03/05/2014) (cit. on pp. 55, 57, 59, 60).

[202] Y. Li and F. Petriello. "Combining QCD and electroweak corrections to dilepton production in the framework of the FEWZ simulation code". In: *Physical Review D* 86.9 (Nov. 2012). DOI: 10.1103/physrevd.86.094034 (cit. on pp. 55, 57).

[203] A. Giammanco. "The Fast Simulation of the CMS Experiment". In: *Journal of Physics: Conference Series* 513.2 (June 2014), p. 022012. DOI: 10.1088/1742-6596/513/2/022012 (cit. on pp. 57, 64).

[204] T. Perry et al. "Measurement of the production cross section for a W boson and two b jets in pp collisions at $\sqrt{s} = 8\ TeV$". CMS AN-2014/105. Sept. 2015. URL: http://cms.cern.ch/iCMS/user/noteinfo?cmsnoteid=CMS%20AN-2014/105 (cit. on p. 57).

[205] F. F. Cordero, L. Reina, and D. Wackeroth. "W- and Z-boson production with a massive bottom-quark pair at the Large Hadron Collider". In: *Physical Review D* 80.3 (Sept. 2009). DOI: 10.1103/physrevd.80.034015 (cit. on p. 57).

[206] S. D. Drell and T.-M. Yan. "Massive Lepton-Pair Production in Hadron-Hadron Collisions at High Energies". In: *Physical Review Letters* 25.5 (Aug. 1970), 316–320. DOI: 10.1103/physrevlett.25.316 (cit. on p. 57).

[207] B. Clerbaux et al. "Search for High Mass Resonances Decaying to Electron Pairs at 8 TeV with the Full 2012 dataset". CMS AN 2012/415. Oct. 2012. URL: http://cms.cern.ch/iCMS/user/noteinfo?cmsnoteid=CMS%20AN-2012/415 (cit. on pp. 58, 60, 83, 103).

[208] G. Alverson et al. "Search for High-Mass Resonances Decaying to Muon Pairs in pp Collisions at $\sqrt{s} = 8$ TeV". CMS AN 2012/422. Apr. 2014. URL: http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2012/422 (cit. on p. 58).

[209] A. Benaglia et al. "Search for a Higgs boson decaying into two photons in proton-proton collisions recorded by the CMS detector at the LHC". CMS AN 2011/129. Aug. 2012. URL: http://cms.cern.ch/iCMS/user/noteinfo?cmsnoteid=CMS%20AN-2011/129 (cit. on p. 58).

[210] T. Binoth et al. "A full next-to-leading order study of direct photon pair production in hadronic collisions". In: *The European Physical Journal C - Particles and Fields* 16.2 (Aug. 2000), 311–330. DOI: 10.1007/s100520050024 (cit. on p. 58).

[211] J. M. Campbell, R. K. Ellis, and C. Williams. "Vector boson pair production at the LHC". In: *Journal of High Energy Physics* 2011.7 (July 2011). DOI: 10.1007/jhep07(2011)018 (cit. on p. 59).

[212] CMS Collaboration. "Double Parton Scattering cross section limit from same-sign W bosons pair production in di-muon final state at LHC". CMS PAS FSQ-13-001. Nov. 2015. URL: https://cds.cern.ch/record/2103756 (cit. on p. 59).

[213] T. Binoth et al. "Gluon-induced W-boson pair production at the LHC". In: *Journal of High Energy Physics* 2006.12 (Dec. 2006), 046–046. DOI: 10.1088/1126-6708/2006/12/046 (cit. on p. 59).

[214] F. Caola et al. "QCD corrections to W+W- production through gluon fusion". In: *Physics Letters B* 754 (Mar. 2016), 275–280. DOI: 10.1016/j.physletb.2016.01.046 (cit. on p. 59).

[215] T. Binoth, N. Kauer, and P. Mertsch. "Gluon-induced QCD corrections to $pp \to ZZ \to l\bar{l}l'\bar{l}'$". In: *Proceedings of the XVI International Workshop on Deep-Inelastic Scattering and Related Topics* (July 2008). DOI: 10.3360/dis.2008.142 (cit. on p. 59).

[216] E. Avdeeva et al. "Measurement of the total and differential Wgamma production cross section at 8 TeV". CMS AN-2015/055. Apr. 2016. URL: http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2015/055 (cit. on p. 59).

[217] S. Frixione and B. R. Webber. "Matching NLO QCD computations and parton shower simulations". In: *Journal of High Energy Physics* 2002.06 (June 2002), 029–029. DOI: 10.1088/1126-6708/2002/06/029 (cit. on p. 59).

[218] CMS website. "PREP Admin Control Panel". URL: http://cms.cern.ch/iCMS/prep/campaignmanagement?id=Summer12_DR53X (visited on 11/20/2014) (cit. on p. 59).

[219] M. Czakon, P. Fiedler, and A. Mitov. "Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $\mathcal{O}(\alpha_S^4)$". In: *Physical Review Letters* 110 (June 2013), p. 252004. DOI: 10.1103/PhysRevLett.110.252004 (cit. on pp. 60, 84).

[220] J. M. Campbell and R. K. Ellis. "$t\bar{t}W^{\pm}$ production and decay at NLO". In: *Journal of High Energy Physics* 2012.7 (July 2012). DOI: 10.1007/jhep07(2012)052 (cit. on p. 60).

[221] M. V. Garzelli et al. "$t\bar{t}W^{\pm}$ + $t\bar{t}Z$ hadroproduction at NLO accuracy in QCD with Parton Shower and Hadronization effect". In: *Journal of High Energy Physics* 2012.11 (Nov. 2012). DOI: 10.1007/jhep11(2012)056 (cit. on p. 60).

[222] N. Kidonakis. "Differential and total cross sections for top pair and single top production". In: *Proceedings of the 20th International Workshop on Deep-Inelastic Scattering and Related Subjects (DIS 2012): Bonn, Germany, March 26-30, 2012*. May 2012, pp. 831–834. DOI: 10.3204/DESY-PROC-2012-02/251 (cit. on p. 60).

[223] LHC Higgs Cross Section Working Group. "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties". CERN-2013-004. June 2013. DOI: 10.5170/CERN-2013-004 (cit. on p. 61).

[224] K. Hamilton, P. Nason, and G. Zanderighi. "MINLO: multi-scale improved NLO". In: *Journal of High Energy Physics* 2012.10 (Oct. 2012). DOI: 10.1007/jhep10(2012)155 (cit. on p. 60).

[225] CMS Twiki. "SM Higgs production cross sections at $\sqrt{s} = 8\ TeV$ (update in CERN Report3)". URL: http://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt8TeV?rev=59 (visited on 09/28/2015) (cit. on p. 61).

[226] D. J. Lange. "The EvtGen particle decay simulation package". In: *Nuclear Instruments and Methods in Physics Research Section A* 462.1-2 (Apr. 2001), 152–155. DOI: 10.1016/s0168-9002(01)00089-4 (cit. on p. 61).

[227] G. Altarelli, B. Mele, and M. Ruiz-Altaba. "Searching for new heavy vector bosons in p $\bar{\text{p}}$ colliders". In: *Zeitschrift für Physik C Particles and Fields* 45.1 (1989), 109–121. DOI: 10.1007/bf01556677 (cit. on p. 61).

[228] CMS Collaboration. "Search for physics beyond the standard model in final states with a lepton and missing transverse energy in proton-proton collisions at $\sqrt{s} = 8$ TeV". In: *Phys. Rev. D* 91.9 (May 2015). CMS PAS EXO-12-060. DOI: 10.1103/physrevd.91.092005 (cit. on pp. 61, 104, 142, 145, 146).

[229] R. Gavin et al. "FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order". In: *Computer Physics Communications* 182.11 (Nov. 2011), 2388–2403. DOI: 10.1016/j.cpc.2011.06.008 (cit. on p. 62).

[230] S. Quackenbush et al. "W physics at the LHC with FEWZ 2.1". In: *Computer Physics Communications* 184.1 (Jan. 2013), 209–214. DOI: 10.1016/j.cpc.2012.09.005 (cit. on p. 62).

[231] Y. L. Dokshitser. "Calculation of structure functions of deep-inelastic scattering and e^+e^- annihilation by perturbation theory in quantum chromodynamics". In: *Journal of Experimental and Theoretical Physics* (Oct. 1977). URL: http://jetp.ac.ru/cgi-bin/e/index/e/46/4/p641 (cit. on p. 62).

[232] V. N. Gribov and L. N. Lipatov. "Deep inelastic e p scattering in perturbation theory". In: *Soviet Journal of Nuclear Physics* 15 (1972), pp. 438–450. URL: http://inspirehep.net/record/73449 (cit. on p. 62).

[233] L. N. Lipatov. "The parton model and perturbation theory". In: *Soviet Journal of Nuclear Physics* 20 (1975), pp. 94–102. URL: http://inspirehep.net/record/91556 (cit. on p. 62).

[234] G. Altarelli and G. Parisi. "Asymptotic freedom in parton language". In: *Nuclear Physics B* 126.2 (Aug. 1977), 298–318. DOI: 10.1016/0550-3213(77)90384-4 (cit. on p. 62).

[235] J. Pumplin et al. "New Generation of Parton Distributions with Uncertainties from Global QCD Analysis". In: *Journal of High Energy Physics* 2002.07 (July 2002), 012–012. DOI: 10.1088/1126-6708/2002/07/012 (cit. on pp. 62, 63).

[236] H. Lai et al. "New parton distributions for collider physics". In: *Physical Review D* 82.7 (Oct. 2010), p. 074024. DOI: 10.1103/physrevd.82.074024 (cit. on pp. 62, 63, 85).

[237] A. D. Martin et al. "Parton distributions for the LHC". In: *European Physical Journal C* 63.2 (July 2009), 189–285. DOI: 10.1140/epjc/s10052-009-1072-5 (cit. on pp. 62, 85).

[238] R. D. Ball et al. "A determination of parton distributions with faithful uncertainty estimation". In: *Nuclear Physics B* 809.1-2 (Mar. 2009), 1–63. DOI: 10.1016/j.nuclphysb.2008.09.037 (cit. on p. 62).

[239] J. Pumplin et al. "Uncertainties of predictions from parton distribution functions. II. The Hessian method". In: *Physical Review D* 65.1 (Dec. 2001), p. 014013. DOI: 10.1103/physrevd.65.014013 (cit. on p. 62).

[240] CMS Collaboration. "Measurement of the inelastic pp cross section at $\sqrt{s} = 7\ TeV$". CMS PAS QCD-11-002. Mar. 2012. URL: http://cds.cern.ch/record/1433413 (cit. on p. 63).

[241] CMS Twiki. "2012 Pileup Information". URL: http://twiki.cern.ch/twiki/bin/viewauth/CMS/PdmVPileUpDescription?rev=22 (visited on 08/28/2014) (cit. on pp. 63, 166, 167).

[242] CMS Twiki. "Utilities for Accessing Pileup Information for Data". URL: http://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupJSONFileforData?rev=16 (visited on 08/28/2014) (cit. on p. 63).

[243] CMS Twiki. "Detector Description and Geometry Offline Guide". URL: http://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideDetectorDescription?rev=66 (visited on 02/09/2017) (cit. on p. 64).

[244] S. Agostinelli et al. "Geant4 - a simulation toolkit". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (July 2003), 250–303. DOI: 10.1016/s0168-9002(03)01368-8 (cit. on p. 64).

[245] J. Allison et al. "Geant4 developments and applications". In: *IEEE Transactions on Nuclear Science* 53.1 (Feb. 2006), 270–278. DOI: 10.1109/tns.2006.869826 (cit. on p. 64).

[246] CMS Twiki. "How to Configure and Run Detector Simulation and Digitization". URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookSimDigi?rev=87 (visited on 02/09/2017) (cit. on p. 64).

[247] CMS Twiki. "Recommended Jet Energy Corrections and Uncertainties For Data and MC". URL: http://twiki.cern.ch/twiki/bin/viewauth/CMS/JECDataMC?rev=59 (visited on 08/28/2014) (cit. on p. 64).

[248] CMS Twiki. "MET Optional Filters". URL: http://twiki.cern.ch/twiki/bin/viewauth/CMS/MissingETOptionalFilters?rev=62 (visited on 11/03/2014) (cit. on p. 67).

[249] R. Bruce et al. "Simulations and measurements of beam loss patterns at the CERN Large Hadron Collider". In: *Physical Review Special Topics - Accelerators and Beams* 17.8 (Aug. 2014). DOI: 10.1103/physrevstab.17.081004 (cit. on p. 67).

[250] D. Acosta et al. "Beam Halo Event Identification in CMS Using the CSCs, ECAL, and HCAL". CMS AN-2010/111. Mar. 2011. URL: http://cms.cern.ch/iCMS/user/noteinfo?cmsnoteid=CMS%20AN-2010/111 (cit. on p. 67).

[251] J. P. Chou et al. "Anomalous HB/HE Noise at Startup: Characteristics and Rejection Algorithms". CMS IN-2010/006. Feb. 2010. URL: http://cms.cern.ch/iCMS/user/noteinfo?cmsnoteid=CMS%20IN-2010/006 (cit. on p. 68).

[252] CMS Twiki. "HCAL Noise Library". URL: http://twiki.cern.ch/twiki/bin/view/CMS/HcalNoiseInfoLibrary?rev=47 (visited on 08/12/2016) (cit. on p. 68).

[253] B. Dahmes et al. "Orbit Gap Operations for CMS HCAL". CMS IN-2010/020. July 2010. URL: http://cms.cern.ch/iCMS/user/noteinfo?cmsnoteid=CMS%20IN-2010/020 (cit. on p. 68).

[254] CMS Twiki. "Physics Data And Monte-Carlo Validation Known Features". URL: http://twiki.cern.ch/twiki/bin/viewauth/CMS/PdmVKnowFeatures?rev=19 (visited on 08/28/2014) (cit. on p. 68).

[255] CMS Twiki. "ECAL Maksed Cell Issue Summary for SUSY Analyses". URL: http://twiki.cern.ch/twiki/bin/view/CMS/SusyEcalMaskedCellSummary?rev=32 (visited on 08/15/2016) (cit. on p. 68).

[256] M. Anfreville et al. "Laser monitoring system for the CMS lead tungstate crystal calorimeter". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 594.2 (Sept. 2008), 292–320. DOI: 10.1016/j.nima.2008.01.104 (cit. on p. 68).

[257] A. Albert. "Extension of the Model Unspecific Search in CMS to Final States with Jets using 2012 Data". Master thesis. RWTH Aachen University, Nov. 2015. URL: http://web.physik.rwth-aachen.de/~hebbeker/theses/albert_master.pdf (cit. on pp. 68, 70, 75, 76, 157).

[258] CMS Collaboration. "The CMS Trigger System". CMS PAS TRG-12-001. Sept. 2016. URL: https://arxiv.org/abs/1609.02366 (cit. on pp. 69, 70).

[259] S. Chang et al. "Search for new physics in the single lepton+MET final states with the full 2012 dataset at $\sqrt{s} = 8$ TeV". CMS AN-2012/423. Aug. 2013. URL: http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2012/423 (cit. on pp. 69, 70, 104).

[260] CMS Twiki. "High Level Trigger Approved Results in Run I". URL: http://twiki.cern.ch/twiki/bin/view/CMSPublic/HighLevelTriggerRunIResults (visited on 08/18/2016) (cit. on p. 69).

[261] M. Cacciari and G. P. Salam. "Pileup subtraction using jet areas". In: *Physics Letters B* 659.1-2 (Jan. 2008), 119–126. DOI: 10.1016/j.physletb.2007.09.077 (cit. on p. 71).

[262] M. Cacciari, G. P. Salam, and G. Soyez. "The catchment area of jets". In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), 005–005. DOI: 10.1088/1126-6708/2008/04/005 (cit. on p. 71).

[263] CMS Collaboration. "Performance of muon identification in pp collisions at $\sqrt{s} = 7\ TeV$". CMS PAS MUO-10-002. July 2010. URL: http://cds.cern.ch/record/1279140 (cit. on pp. 72, 73).

[264] CMS Twiki. "Baseline muon selections". URL: http://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideMuonId?rev=48 (visited on 11/14/2014) (cit. on p. 72).

[265] CMS Collaboration. "The CMS muon project: Technical Design Report". CERN-LHCC-97-032; CMS-TDR-3. Dec. 1997. URL: http://cds.cern.ch/record/343814 (cit. on p. 73).

[266] CMC Collaboration. "Search for neutral Higgs bosons decaying to tau pairs in pp collisions at $\sqrt{s} = 7$ TeV". In: *Physics Letters B* 713.2 (June 2012). CMS PAS HIG-11-029, 68–90. DOI: 10.1016/j.physletb.2012.05.028 (cit. on p. 73).

[267] CMS Collaboration. "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV". In: *Journal of Instrumentation* 10.06 (June 2015). CMS PAS EGM-13-001, P06005–P06005. DOI: 10.1088/1748-0221/10/06/p06005 (cit. on pp. 73, 80, 83, 136).

[268] B. C. L. Thomas. "High Et electron selection presentation and optimisation in case of high pile-up conditions". CMS AN 2012/258. May 2012. URL: http://cms.cern.ch/iCMS/user/noteinfo?cmsnoteid=CMS%20AN-2012/258 (cit. on p. 73).

[269] CMS Twiki. "HEEP Electron ID and isolation". URL: http://twiki.cern.ch/twiki/bin/viewauth/CMS/HEEPElectronID?rev=65 (visited on 11/14/2014) (cit. on p. 73).

[270] L. Thomas. "Search for new heavy narrow resonances decaying into a dielectron pair with the CMS detector". PhD thesis. Brussels University, Sept. 2015. URL: http://cds.cern.ch/record/2048187 (cit. on p. 74).

[271] CMS Collaboration. "Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV". In: *Journal of Instrumentation* 10.08 (Aug. 2015). CMS PAS EGM-14-001, P08010–P08010. DOI: 10.1088/1748-0221/10/08/p08010 (cit. on pp. 74, 75, 80).

[272] CMS Twiki. "H/E for electron and photon ID in 2012". URL: http://twiki.cern.ch/twiki/bin/view/CMS/HoverE2012?rev=12 (visited on 12/10/2016) (cit. on p. 74).

[273] CMS Twiki. "Simple Cut Based Photon ID 2012". URL: http://twiki.cern.ch/twiki/bin/viewauth/CMS/CutBasedPhotonID2012?rev=23 (visited on 11/14/2014) (cit. on pp. 74, 75).

[274] CMS Collaboration. "Observation of the diphoton decay of the Higgs boson and measurement of its properties". In: *The European Physical Journal C* 74.10 (Oct. 2014). CMS PAS HIG-13-001. DOI: 10.1140/epjc/s10052-014-3076-z (cit. on p. 75).

[275] Talk by J. Bendavid at the CMS Egamma Identification Meeting, Febuary 9, 2012. "Tools for conversion ID and rejection". URL: http://indico.cern.ch/event/177272/contributions/1443588/attachments/231387/323681/egammaFeb9.pdf (visited on 07/18/2016) (cit. on p. 75).

[276] CMS Twiki. "Jet Identification". URL: http://twiki.cern.ch/twiki/bin/view/CMS/JetID?rev=82 (visited on 05/03/2015) (cit. on pp. 75, 76).

[277] N. S. Eirini Tziaferi. "Performance of the Particle-Flow jet identification criteria using proton-proton collisions at $\sqrt{s} = 8$ TeV". CMS AN-2014/227. Oct. 2014. URL: http://cms.cern.ch/iCMS/user/noteinfo?cmsnoteid=CMS%20AN-2014/227 (cit. on p. 75).

[278] N. Saoulidou. "Particle Flow Jet Identification Criteria". CMS AN-2010/003. Feb. 2010. URL: http://cms.cern.ch/iCMS/user/noteinfo?cmsnoteid=CMS%20AN-2010/003 (cit. on p. 75).

[279] CMS Twiki. "Reference muon momentum scale and resolution". URL: http://twiki.cern.ch/twiki/bin/viewauth/CMS/MuonReferenceResolution?rev=28 (visited on 02/06/2017) (cit. on p. 80).

[280] Talk by T. N. Kypreos at the CMS Muon DPG-PH Meeting, July 23, 2012. "New results on momentum scale with endpoint methods". URL: http://indico.cern.ch/event/188598/contributions/339056/attachments/264923/371000/mpogEndpointKypreos.pdf (visited on 02/06/2017) (cit. on p. 80).

[281] CMS Collaboration. "ECAL 2010 performance results". CMS DP-2011-008. June 2011. URL: http://cds.cern.ch/record/1373389 (cit. on p. 80).

[282] CMS Collaboration. "Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at $\sqrt{s} = 7$ TeV". In: *Journal of Instrumentation* 8.09 (Sept. 2013). CMS PAS EGM-11-001, P09009–P09009. DOI: 10.1088/1748-0221/8/09/p09009 (cit. on p. 80).

[283] CMS Twiki. "Official Prescription for calculating uncertainties on Missing Transverse Energy (MET)". URL: http://twiki.cern.ch/twiki/bin/viewauth/CMS/MissingETUncertaintyPrescription?rev=21 (visited on 02/01/2016) (cit. on p. 81).

[284] CMS Twiki. "PAT Tools - MET Systematics Tools". URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuidePATTools?rev=60#MET_Systematics_Tools (visited on 02/01/2016) (cit. on p. 81).

[285] P. Harris et al. "Pileup Jet Identification". CMS AN 2013/186. Aug. 2013. URL: http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2013/186 (cit. on p. 81).

[286] N. Adam et al. "Generic Tag and Probe Tool for Measuring Efficiency at CMS with Early Data". CMS AN-2009/111. July 2009. URL: http://cms.cern.ch/iCMS/user/noteinfo?cmsnoteid=CMS%20AN-2009/111 (cit. on p. 81).

[287] CMS Collaboration. "Search for Resonances in the Dilepton Mass Distribution in pp Collisions at $\sqrt{s} = 8$ TeV". CMS PAS EXO-12-061. Feb. 2013. URL: http://cds.cern.ch/record/1519132 (cit. on pp. 81, 103).

[288] L. Thomas and B. Clerbaux. "Measurement of the high energy electron selection efficiency using Drell-Yan events collected in 2012". CMS AN 2013/174. May 2013. URL: http://cms.cern.ch/iCMS/user/noteinfo?cmsnoteid=CMS%20AN-2013/174 (cit. on p. 81).

[289] Talk by S.-W. Li at the CMS Regular EGM Working Group Meeting, January 30, 2013. "Scale Factors for photon cut based ID with the full 2012 dataset". URL: http://indico.cern.ch/event/230885/contributions/1539278/attachments/382794/532499/Egamma_20130130.pdf (visited on 02/06/2017) (cit. on p. 82).

[290] CMS Collaboration. "Muon Identification performance: hadron mis-Id measurements and RPC Muon selections". CMS DP-2014/018. May 2014. URL: http://cms.cern.ch/iCMS/user/noteinfo?cmsnoteid=CMS%20DP-2014/018 (cit. on p. 82).

[291] L. Benato et al. "Search for heavy lepton partners of neutrinos in pp collisions at 8 TeV, in the context of Type III seesaw mechanism". CMS AN 2013/135. Mar. 2015. URL: http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2013/135 (cit. on pp. 82, 83).

[292] A. Askew et al. "Search for Dark Matter and Large Extra Dimensions in the gamma+MET final state in pp collisions at $\sqrt{s} = 8$ TeV in the CMS experiment with 19.6 fb^{-1} of data". CMS AN-2012/439. Apr. 2014. URL: http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2012/439 (cit. on pp. 83, 121).

[293] A. Ivanov et al. "Measurement of the top-quark pair + photon production cross section in the electron+jets and muon+jets channels in pp collisions at 8 TeV ". CMS AN 2015/165. Mar. 2016. URL: http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2015/165 (cit. on p. 83).

[294] S. Banerjee et al. "Search for excited leptons in the $l^+l^-\gamma$ final state at $\sqrt{s} = 8$ TeV". CMS AN-2012/418. Nov. 2014. URL: http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2012/418 (cit. on p. 83).

[295] Talk by J. Tucker at the CMS Muon POG Working Meeting, March 22, 2012. "Muon momentum resolution using cosmic muon data and MC". URL: http://indico.cern.ch/event/183007/contributions/316223/attachments/248827/348006/tucker.pdf (visited on 06/07/2016) (cit. on pp. 83, 135).

[296] S. Baffioni et al. "Electron Charge Identification using 8 TeV data". CMS AN 2014/164. July 2014. URL: http://cms.cern.ch/iCMS/user/noteinfo?cmsnoteid=CMS%20AN-2014/164 (cit. on p. 83).

[297] A. Vicinii. "Practical implementation of the PDF4LHC recipe". URL: http://www.hep.ucl.ac.uk/pdf4lhc/PDF4LHC_practical_guide.pdf (visited on 11/27/2014) (cit. on p. 84).

[298] S. Alekhin et al. "The PDF4LHC Working Group Interim Report". arXiv:1101.0536. Jan. 2011. URL: http://arxiv.org/abs/1101.0536 (cit. on p. 84).

[299] M. Botje et al. "The PDF4LHC Working Group Interim Recommendations". arXiv:1101.0538. Jan. 2011. URL: http://arxiv.org/abs/1101.0538 (cit. on p. 84).

[300] LHAPDF. "LHAPDF the Les Houches Accord PDF Interface". URL: http://projects.hepforge.org/lhapdf/ (visited on 11/24/2014) (cit. on p. 85).

[301] CMS Collaboration. "Properties of the Higgs-like boson in the decay H to ZZ to 4l in pp collisions at $\sqrt{s} = 7\ TeV$ and 8 TeV". CMS PAS HIG-13-002. Mar. 2013. URL: http://cds.cern.ch/record/1523767 (cit. on p. 105).

[302] N. Amapane et al. "Properties of the Higgs-like boson in the decay H→ZZ→4l in pp collisions at $\sqrt{s} = 7\ TeV$ and 8 TeV". CMS AN-2013/003. Mar. 2013. URL: http://cms.cern.ch/iCMS/user/noteinfo?cmsnoteid=CMS%20AN-2013-003 (cit. on pp. 105, 173).

[303] L. Benato et al. "Search for heavy resonances decaying into a vector boson and a Higgs boson in the ($\nu\nu$, $l\mu$, ll) $b\bar{b}$ final state". CMS AN-2015/186. May 2016. URL: http://cms.cern.ch/iCMS/user/noteinfo?cmsnoteid=CMS%20AN-2015/186 (cit. on p. 106).

[304] R. Boughezal, X. Liu, and F. Petriello. "W-boson plus jet differential distributions at NNLO in QCD". arXiv:1602.06965. Feb. 2016. URL: http://arxiv.org/abs/1602.06965 (cit. on p. 106).

[305] S. Kallweit et al. "NLO QCD+EW predictions for V+jets including off-shell vector-boson decays and multijet merging". In: *Journal of High Energy Physics* 2016.4 (Apr. 2016). DOI: 10.1007/jhep04(2016)021 (cit. on p. 106).

[306] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *Journal of High Energy Physics* 2014.7 (June 2014). DOI: 10.1007/jhep07(2014)079 (cit. on p. 108).

[307] CMS Collaboration. "MET performance in 8 TeV data". CMS PAS JME-12-002. Apr. 2013. URL: http://cds.cern.ch/record/1543527 (cit. on p. 108).

[308] J. Roemer. "Model Unspecific Search for New Physics with *pp* Collisions at $\sqrt{s} = 13$ TeV with the CMS Experiment". Master thesis. RWTH Aachen University, Jan. 2017. URL: http://web.physik.rwth-aachen.de/~hebbeker/theses/roemer_master.pdf (cit. on pp. 123, 124, 131, 132, 158).

[309] P. Millet et al. "Search for RPV SUSY resonant second generation slepton production in same-sign dimuon events at $\sqrt{s} = 8$ TeV". CMS AN-2013/225. June 2015. URL: http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2013/225 (cit. on p. 137).

[310] S. Thuer et al. "Search for RPV SUSY resonant first generation slepton production in same-sign dielectron events at $\sqrt{s} = 8$ TeV". CMS AN-2014/220. Apr. 2015. URL: http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2014/220 (cit. on p. 137).

[311] CMS Collaboration. "Measurement of W^+W^- and ZZ production cross sections in pp collisions at $\sqrt{s} = 8$ TeV". In: *Physics Letters B* 721.4-5 (Apr. 2013). CMS PAS SMP-12-024, 190–211. DOI: 10.1016/j.physletb.2013.03.027 (cit. on p. 155).

[312] CMS Collaboration. "Measurements of the ZZ production cross sections in the $2l2\nu$ channel in proton–proton collisions at $\sqrt{s} = 7$ TeV and 8 TeV and combined constraints on triple gauge couplings". In: *The European Physical Journal C* 75.10 (Oct. 2015). CMS PAS SMP-12-016. DOI: 10.1140/epjc/s10052-015-3706-0 (cit. on p. 155).

[313] J. de Favereau et al. "DELPHES 3: a modular framework for fast simulation of a generic collider experiment". In: *Journal of High Energy Physics* 2014.2 (Feb. 2014). DOI: 10.1007/jhep02(2014)057 (cit. on p. 158).

[314] O. J. Dunn. "Multiple Comparisons among Means". In: *Journal of the American Statistical Association* 56.293 (Mar. 1961), 52–64. DOI: 10.1080/01621459.1961.10482090 (cit. on p. 163).

Acknowledgements

My sincere gratitude goes to Prof. Thomas Hebbeker for the great opportunity of preforming a promising search such as the MUSiC analysis in the exciting times at the LHC following the Higgs boson discovery. I am very grateful for numerous visits to conferences, workshops and an extended stay at CERN as well as encouragement and much appreciated liberties in developing the MUSiC analysis at 8 TeV. Thank you Prof. Chistopher Wiebusch for being the second referee of this thesis and valuable comments on dark matter observations.

The MUSiC analysis was and is performed by a team of enthusiastic and inspirational individuals, some of which I would like to thank especially: Holger Pieta for the initial introduction to the analysis laced with your enthusiasm for HEP and IT topics; Paul Papacz for countless fruitful discussions and long hours of debating and tackling quirks of the MUSiC framework, also for your lasting technical and moral support; Simon Knutzen for an open ear on analysis details and coding topics (especially the scanner), as well as your help on the MUSiC 8 TeV PAS; Tobias Pook for new enthusiasm during the later part of my PhD research and creative discussions especially for the region veto method; Jonas Lieb for de-bugging help and a faster scanner; Jonas Roemer for a first glance into the MUSiC results in 13 TeV data; Arnd Meyer for being there from beginning to end with an open ear and decades of experience and knowledge in times of need;

I would also like to offer my sincere thanks to all members of the CMS Exotica working group, especially the Non-Hadronic subgroup. From the first presentation of the 8 TeV analysis to the final approval of the PAS, the MUSiC analysis has received support and highly constructive criticism from numerous generations of conveners, and particularly from the wonderful ARC of EXO-14-016. I sincerely appreciate the time you invested and your endless motivation. Thank you all very much for helping to scrutinize our hundreds of final states and bearing with us over the years.

The tedious task of proof-reading several versions of this thesis was performed with much devotion by Arnd Meyer, Leila Ali Cavasonza, Sören Erdweg, Klaas Padeken, Simon Knutzen, Tobias Pook and Eva Maria Duchardt-Hellbarth. Thank you dearly for sparing the reader typos, unclear explanations or missing information.

The Physics Institute III is lucky to have an excellent support in IT matters, which I have come to cherish over the past years. Many thanks to especially Thomas Kress, Andreas Nowack, Michael Bontenackels and Achim Burdziak for restoring deleted files, helping with the dCache and numerous other (unnoticed) tasks.

An outstanding working environment is made possible by the many people in the Physics Institute III, joining forces to collectively tackle task given to us by high energy physics. I greatly enjoyed the open-door-mentality at our institute, making it easy to ask questions, brainstorm new analysis methods and also have lots of fun on our way to the finish line. My years in particle physics and in Aachen would have not been the same without (given office-wise): Simon Knutzen, Jonas Lieb, Paul Papacz, Tobias Pook, Jonas Roemer, Matthias Endres, Sören Erdweg, Andreas Güth, Kerstin Hoepfner, Arnd Meyer, Markus Merschmeyer, Daniel Teyssier, Melanie Roder, Iris Rosewick, Adriana Del Piero, Andreas Albert, Michael Brodski, Swagata Mukherjee, Klaas Padeken, Marcel Materok, Mark Olschweski, Thomas Esch, Philipp Millet, Sebastian Thür, Markus Radziej, Leila Ali Cavasonza, Sarah Beranek and Ronja Hetzel. Thank you all so much for making this an experience to remember for a lifetime.