Discovery Potential of a Model Independent Search for New Physics at the LHC

von
Jonas Lieb

Masterarbeit im Fach Physik

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

im
August 2017

angefertigt am
III. Physikalischen Institut A

bei
Prof. Dr. Thomas Hebbeker
(Zweitkorrektur: Prof. Dr. Martin Erdmann)

Ich versichere, dass ich die Arbeit einschließlich beigefügter Darstellungen und Tabellen selbständig angefertigt und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie Zitate kenntlich gemacht habe.

Jonas Lieb
Aachen, den 16. August 2017

Abstract

In 2015 and 2016, the CMS detector recorded proton-proton collisions at an unprecedented center of mass energy of $\sqrt{s} = 13\,\text{TeV}$. The Model Unspecific Search in CMS (MUSiC) provides an automated search for various possible signatures of new physics in these data.

In a three step process, MUSiC first classifies events according to the physics content of the final state, searches a set of kinematic distributions for the most significant deviations between Standard Model Monte Carlo simulations and observed data and finally applies a statistical hypothesis test to draw conclusions about indications of new physics in the observed dataset.

In this thesis, the discovery potential towards new physics is assessed. For this purpose, a quantification of the test power is defined. Subsequently, the framework is applied to simulated events of four benchmark models for new physics. Alongside the discovery potential towards these theories, the influence of several existing and newly introduced features and parameters on the sensitivity is measured.

Kurzdarstellung

In den Jahren 2015 und 2016 wurden vom CMS-Experiment Kollisionen von Protonenpaaren bei einer Schwerpunktsenergie von $\sqrt{s} = 13\,\text{TeV}$ beobachtet. Mithilfe der modellunabhängigen Suche in CMS (MUSiC) können die aufgezeichneten Daten automatisiert nach Anzeichen neuer Physik durchsucht werden.

Die MUSiC Analyse besteht aus einem mehrstufigen Verfahren, bei dem die aufgezeichneten Ereignisse zuerst ihrem Endzustand gemäß in verschiedene Klassen eingeteilt werden. In jeder Klasse werden daraufhin Histogramme einiger kinematischer Variablen aggregiert und die größten Abweichungen zwischen der beobachteten Verteilung und einer Erwartung aus Monte Carlo-Simulationen des Standardmodelles berechnet. Im Rahmen eines statistischen Hypothesentests wird schließlich festgestellt werden, ob beobachtete Daten Hinweise auf neue Physik beinhalten.

In dieser Arbeit wird das Entdeckungspotential der Analyse untersucht. Zu diesem Zweck wird zuerst eine geeignete quantitative Teststärke definiert. Diese wird anschließend für Simulationsergebnisse vier verschiedener Modelle neuer Physik ausgerechnet. Neben dem absoluten Wert in Bezug auf diese Modelle wird der Einfluss verschiedener Funktionalitäten und Parameter auf die Sensitivität untersucht.

Contents

1 Introduction

This thesis deals with the Model Unspecific Search in CMS (MUSiC), which is a particle physics analysis aiming to find new physics in data observed by the Compact Muon Solenoid (CMS) experiment.

In the first chapter, the Standard Model of particle physics and the CMS experiment are introduced. The introduction is extended in the second chapter, which provides a description of the MUSiC analysis in its current state. Different strategies and algorithms of the analysis framework are motivated and explained. Afterwards, the third chapter serves as a reference of simulations used in this thesis. The remaining chapters contain new contributions to the analysis: In the fourth chapter, the existing local test statistic θ is evaluated and a possible alternative are discussed. Additionally, a global measure for the discovery potential is introduced. It is used in the final chapter, where simulations of new physics processes are used to assess the discovery potential of the aforementioned methods.

1.1 Units and Abbreviations

Throughout this work, the natural unit system will be used. This means that the speed of light and the Planck constant are fixed to 1:

$$\hbar = c = 1$$

It follows that the only unit needed to express most other physical quantities is the unit of energy. As customary in high energy particle physics, it is notated in gigaelectronvolts: $1\,\mathrm{GeV} = 1 \times 10^{9}\,\mathrm{eV} = 1.60218 \times 10^{-10}\,\mathrm{J}$. Two exceptions to this are the cross section and luminosity, which are expressed in femtobarn and inverse femtobarn respectively: $1\,\mathrm{fb} = 10^{-43}\,\mathrm{m}^2$.

1.2 The Standard Model

Finding the substructure of matter has interested mankind for thousands of years. Ancient Greek philosophers already contemplated about the basic building blocks of matter [1]. They coined the term *atom* for what they thought to be indivisible. Since then, our understanding of the smallest parts and what holds them together has vastly improved. This field of science is nowadays called *Elementary Particle Physics*.

As of 2017, the most broadly accepted model in particle physics is the so-called *Standard Model of Particle Physics*. The Standard Model describes matter and its interactions through *particles* and *forces*. It is based on the mathematical foundation of quantum field theory, exhibiting several concepts such as gauge invariance, spontaneous symmetry breaking, perturbative and non-perturbative behavior.

Although it has provided accurate predictions, making it a successful theory, recent observations have yielded some inconsistencies, which motivate the scientific field to advance. The following sections will be a short introduction to the Standard Model, its historical development, fundamental ideas and open questions. Motivated by these questions, possible extensions to the Standard Model will be introduced afterwards.

1.2.1 Particles and Interactions

Elementary particles are described as point-like indivisible structureless objects, much like the atoms were to the ancient Greeks. These objects are characterized by physical observables such as mass, charge and spin.

Elementary particles can be subdivided into several groups, as shown in figure 1.1. At the first level, one distinguishes between two groups: *fermions* and *bosons*.

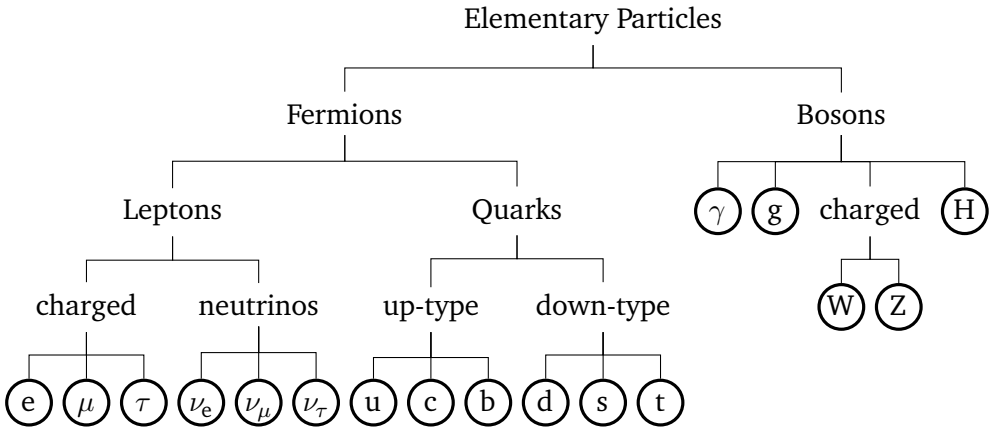


Figure 1.1: Nested groups of elementary particles. The actual particle symbols are indicated by circles.

Fermions are elementary particles with a spin of $1/2$. They follow Fermi-Dirac statistics and have to obey Pauli's exclusion principle which states that no more than one particle can occupy a certain state characterized by its quantum numbers. The group of fermions can be further divided into *quarks* and *leptons*. There are six quarks: up (u), down (d), charm (c), strange (s), bottom (b) and top (t). Quarks carry color charge and are the only fermions to take part in the strong interaction. The remaining fermions are called leptons. There are three charged leptons: The *electron* (e), the *muon* (μ) and the *tau* (τ). Each of the charged leptons has a massless, electrically neutral *neutrino* (ν_e, ν_μ, ν_τ), associated with it.

The second group of particles are called bosons. They mediate interactions (also called *forces*) and possess an integer spin. Bosons follow Bose-Einstein statistics and may thus occupy the same quantum state. Each kind of boson is responsible for one kind of elementary force: the electrodynamic force is mediated by the *photon* (γ), the strong force by the *gluon* (g) and the weak force by the Z and W bosons. In addition, there is the *Higgs boson* (H), which uses the Higgs mechanism to give mass to all mentioned massive elementary particles.[1]

Lepton	Mass		Quark	Mass		Boson	Mass	
e	511	keV	u	2	MeV	γ	0	
ν_e	< 2	eV	d	5	MeV	g	0	
μ	106	MeV	c	1.3	GeV	W	80.4	GeV
ν_μ	< 2	eV	s	96	MeV	Z	91.2	GeV
τ	1.78	GeV	b	4.18	GeV	H	125	GeV
ν_τ	< 2	eV	t	173	GeV			

Table 1.1: Known elementary particles and their masses [2].

1.2.2 Interactions

Quantum Electrodynamics

The theory of electromagnetism at extremely small or extremely high energetic scales is called *quantum electrodynamics* (QED). It is a field theory that unifies classical quantum mechanics, special relativity and gauge invariance.

Because the probability amplitude of a quantum mechanical wave function is independent of the wave's phase, two wave functions that only differ in phase are equivalent. Thus, physics processes should behave exactly the same, regardless of the phase term. This concept is called *invariance under phase transformations* and is a desirable property for quantum field theories.

In 1928, the British physicist Paul Dirac derived an relativistic extension of the classical Schrödinger equation [3] by substituting observables with corresponding operators in the relativistic energy-momentum relation $E^2 = p^2 + m^2$. The resulting so-called *Dirac equation* describes the behavior of free spin-½ particles and even predicts the existence of antiparticles. However, the Dirac equation is not invariant under phase transformations and thus Richard Feynman et al. developed a gauge

[1] Note that this does not apply to composite particles (like protons) which gain mask mostly through binding energy.

invariant extension of Dirac's theory [4]. This extension, quantum electrodynamics, was awarded the Nobel Prize in 1965 [5].

The theory of QED evolves around the photon coupling to the electric charge e: The photon interacts with charged particles with a strength of $\alpha \propto e^2$. All possible basic vertices are displayed in figure 1.2: Two charged particles may annihilate into one photon, a photon may be emitted or absorbed, or an oppositely charged particle-antiparticle pair may be created from the photon energy. Note that in any case, the electric charge as well as the fermion type are preserved.

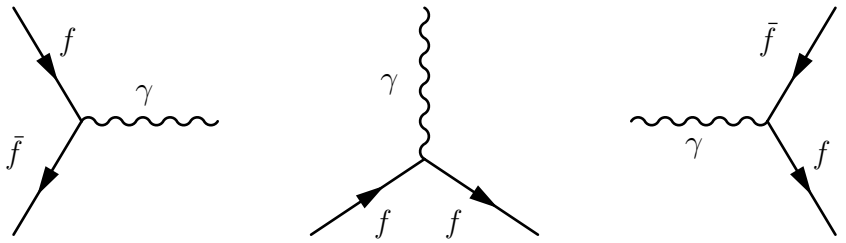


Figure 1.2: Basic interaction vertices of QED: Annihilation of a particle-antiparticle pair into one photon, emission or absorption of a photon, pair production from photon energy. The time direction is from left to right. Note that for kinetic reasons, these vertices cannot exist on their own, but are part of larger diagrams.

One challenge that arises during the calculation of interaction probabilities for a particular QED process is that one has to take into account all possible scenarios that lead to the same indistinguishable result. This includes diagrams containing fermion or photon loops, as shown in figure 1.3. In order to calculate the scattering amplitude for this particular example, one has to integrate over all possible momenta k in the loop. Unfortunately, this integral diverges quickly. A solution for this problem is *renormalization*: By making the coupling strength $\alpha(q^2)$ dependent on the momentum transfer, infinities can be avoided. $\alpha(q^2)$, which now "runs" with the momentum transfer is consequently called *running coupling constant*. Its value is $\alpha(0) \approx \frac{1}{137}$ in the low energy limit and increases towards higher energies [6].

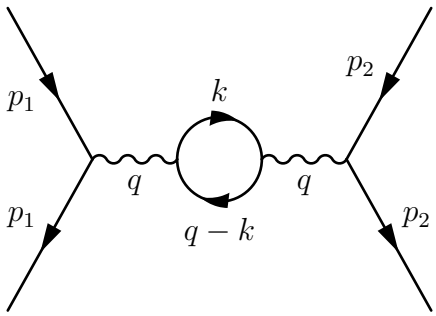


Figure 1.3: Higher order Feynman diagram of electron-positron annihilation. Renormalization is required because of divergence of the integral over k.

Quantum Chromodynamics

After QED, the theory of *quantum chromodynamics* (QCD) was proposed in the early 1960s, as bubble chamber experiments had unexpectedly discovered large numbers of hadrons with different masses and physical properties. Instead of assuming that these particles were fundamental, theorists Murray Gell-Mann et al. proposed a theory where they were made up of constituents, so-called quarks [7,8]. The theory of bound states of the three known quarks (u, d, s) and their excitations was able to reproduce many of the observations. However, the observed Δ^{++} baryon seemed to violate Pauli's exclusion principle: It consists of three u quarks with parallel spins. This dilemma was resolved in 1964, when Oscar W. Greenberg proposed a new quantum number, called *color charge* [9]. Alongside this quantum number, the existence of a gauge boson was predicted, the gluon, which was discovered at DESY in 1979 in three-jet events [10].

The basic interaction vertices of QCD are shown in figure 1.4. The charge of the strong interaction is called color-charge, its possible values are usually called *red*, *green*, *blue*, antired, antigreen and antiblue[2]. Quarks carry one unit of color charge, while gluons carry two units. Color charge is conserved in each interaction vertex. An important principle of QCD is that all naturally occurring particles are colorless: A composite particle carries either all colors (red + green + blue) or a color and its anti-color (e.g. red + antired). The former is found in baryons and the latter in mesons.

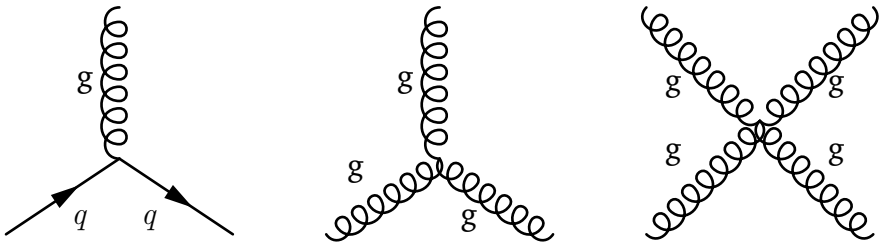


Figure 1.4: Basic interaction vertices of the strong force: Absorption or emission of a gluon from a quark, self interactions between gluons. Note that each diagram can be read in different directions, similarly to figure 1.2.

Similarly as QED, QCD also has to undergo renormalization. But unlike QED, where only fermionic loops are possible, gluons can also interact with each other, resulting in gluon loops. Therefore, renormalization in QCD is much more complicated. Its result is again a running coupling constant, however this time, the strong coupling constant $\alpha_s(q^2)$ decreases at higher energies. In fact, at infinite energies the quarks appear to be free, therefore this effect is called *asymptotic freedom*.

The contrary effect at low energies is also visible: As the distance between quarks is increased, which is equivalent to decreasing the probing energy, the coupling strength

[2] Note that this concept is completely unrelated to visual perception of color.

rises. The expended energy in turn allows new quark-antiquark pairs to form from the vacuum. Each of the created quarks binds with one of the separated quarks such that all quarks again end up in a bound, color-neutral state. This effect tis called *confinement*.

Confinement is also relevant in the experimental context: Quarks and gluons in the final state of a collision often possess excess kinetic energy and thus move away from each other. During that process, confinement causes new quarks and eventually hadrons to form. This effect is called *hadronization*. Repeated hadronization leads to the formation of so-called *jets*, cones of hadrons and subsequent decay products.

Electroweak Interaction

Although theoretical research on the *weak interaction* started in 1933 when Enrico Fermi proposed a contact interaction theory explaining the β-decay [11], it was only further explored in the second half of the 20th century.

In 1968, Sheldon Glashow [12], Abdus Salam [13] and Steven Weinberg [14] independently formulated a theory unifying QED with the weak interaction, called *electroweak theory*. They discovered that instead of classifying particles by their type and electrical charge, they could be regarded by quantum numbers called *weak isospin* and *weak hypercharge*. The theory introduces four gauge fields, W_1, W_2, W_3 and B which couple to these charges. It then predicts the existence of four bosons: The charged massive W^+ and W^-, which directly correspond to the fields W_1 and W_2 and the massive Z and massless photon which arise from mixing of the W_3 and B fields. While the photon had been known for a long time, the predicted vector bosons were experimentally confirmed by Carlo Rubia's group at CERN [15, 16] 15 years later. The mixing angle used for the Z and photon is called *Weinberg angle*, and in turn also relates the coupling strengths of the weak interaction and QED with the electroweak couplings g and g' [17].

Besides QED, the electroweak theory predicts the weak force, which is mediated by the heavy gauge bosons $W^\pm$ and Z. The name "weak force" originates from its finite range and coupling strength which is much smaller than α or α_s. Although the weak interaction would not be able to exist without electroweak unification, it is sometimes treated as a separate theory, called *quantum flavordynamics* (QFD).

The basic interaction vertices of QFD are displayed in figure 1.5. As indicated in the top row, interaction via the W boson allows changing particles' types (also called *flavor*). In the case of leptons, the interaction always contains one neutrino, one charged lepton and a W boson. For quarks, the W boson couples to a quark pair. Each quark pair has a predetermined probability interact via the W boson, as expressed by the CKM matrix [18]. This flavor changing mechanism allows heavy fermions to decay into lighter decay products, as long as there is a lighter particle to decay into.

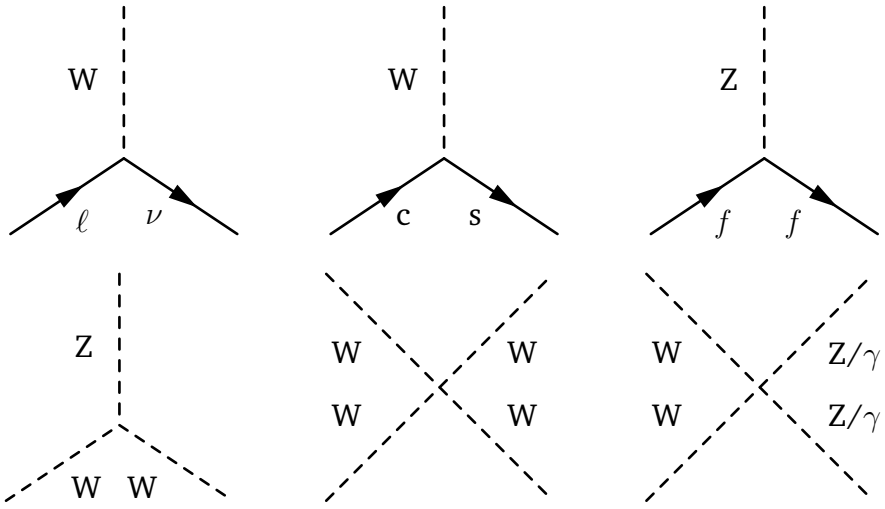


Figure 1.5: Basic vertices of the weak interaction: In the top row, interaction of a W boson with a lepton and neutrino and a two quarks is shown. The third diagram in the top row shows the flavor-conserving interaction of a fermion with a Z boson. The bottom row shows interactions between the bosons, including the photon. Note that, similarly to figure 1.2, all diagrams can be read in different directions.

Consequently, only the lightest fermions are stable: the electron, the u- and d-quarks and the neutrinos.

The Z boson interacts similarly to the photon. In fact, every interaction mediated by a photon can also be mediated by a Z boson. Additionally, the Z boson couples to the electrically neutral neutrinos.

1.2.3 Higgs Mechanism

The naive introduction of boson mass terms in a theory breaks its gauge invariance. While this does not pose a problem for QED and QCD, which are mediated by the massless photon and gluon, QFD, with its massive gauge bosons W and Z, would not be able to exist. Thus, another mechanism is required to give mass to these particles.

A solution was independently proposed by Peter Higgs [19], François Englert et. al. [20] in 1964: a scalar field, called *Higgs field*, and an additional potential in the SM Lagrangian. The introduced potential is symmetric, but not around its minimum. This has far-reaching consequences: *Spontaneous symmetry breaking* is necessary to transform into a non-symmetric ground state. This induces mass-like terms in the Lagrangian for the gauge bosons, as well as a field associated with a spin-0 particle, the Higgs boson. Furthermore, fermion mass terms can be replaced with interactions with the Higgs field. Therefore, the Higgs boson couples predominantly to heavy particles.

In 2012, the experimental signature of a particle compatible with the predicted Higgs boson was discovered at the Large Hadron Collider (LHC) [21,22], leading to a Nobel Prize for Higgs and Englert in 2013 [23].

1.2.4 Open Questions

With the discovery of the Higgs boson, all particles predicted by the Standard Model have been experimentally observed. The model as described in the previous sections has proven to be a successful theory. However, as mentioned earlier, in the last decade, several observations have been made that stand in conflict with its prediction. Additionally, the Standard Model is subject to some theoretical shortcomings. A selection of open questions from both areas will be discussed in the following sections.

Astrophysical Observations

One astrophysical research method is to probe whether observed gravitational effects can be completely accounted for by visible matter distributions. For this purpose, the rotation curves of galaxies have been measured and shown to be incompatible with simulations of visible matter only. Also, fluctuations in gravitational lensing without a visible origin has been observed [24,25]. Furthermore, several experiments have probed the cosmic microwave background for anisotropies. These data have subsequently been compared to the matter distributions obtained by simulations of the early universe according to the ΛCDM model. The results indicated that only about 5 % of the universe is made up of conventional (baryonic) matter [26].

All of these findings suggest that there must be a large amount of matter in the universe that only interacts gravitationally. Because of its electromagnetic invisibility, it is called *Dark Matter*. Since the Standard Model does not provide a fitting Dark Matter candidate, this is an active field of research also at particle collider experiments.

Neutrino Masses

Several experiments have been conducted to analyze the influence on particle's helicity on interactions [27,28]. The results indicate that only left-handed particles (and right-handed antiparticles) take part in the weak interaction. Neutrinos, which can only interact weakly, therefore are believed to only exist as left-handed particles. For that reason, the Standard Model was crafted in a way that only left-handed neutrinos exist. Recent observations of neutrino oscillations [29–32], however, indicate that neutrinos are massive. For the theory to remain renormalizable, neutrinos must thus also exist in their right-handed state [33], which demands for an extension or modification of the Standard Model.

Theoretical Considerations

From a theoretical point of view, the SM shows several deficiencies.

First, there is no consensus about unification of the gauge interactions. Unification of the electroweak and strong interactions would require the coupling strengths to converge at very high energy scales ($\sim 10^{15}$ GeV). However, the Standard Model does not predict a common point of convergence [34].

A second open question is how to combine gravity with the Standard Model. Although it is possible to derive a quantum field theory of gravity, which includes the prediction of a *graviton*, this theory is not renormalizable, i.e. infinities that appear during the calculation cannot be absorbed into a running coupling constant [35].

In any case, if there were a theory of quantum gravity, it would be expected to become relevant at the Planck scale $M_{\mathrm{Pl}} \approx 10^{18}$ GeV, where a particle's Compton wave length is comparable to its Schwarzschild radius. However, all processes that have been observed so far appear at the electroweak scale $m_{\mathrm{EW}} \approx 10^{3}$ GeV. If the Standard Model is believed to hold up to the Planck scale, the Higgs mass would require very precise cancelation terms to result in the observed effective mass of 125 GeV. This problem is called *mass hierarchy problem* [34] and will be addressed in the context of the black hole SM extension in the following chapter.

1.3 Extensions of the Standard Model

Some solutions have been proposed to these problems. In this work, four extensions beyond the Standard Model will be relevant and thus introduced in the following sections.

Extra Dimensions and Black Holes

One class of approaches addressing the mass hierarchy problem is the introduction of additional spatial dimensions. Within the Arkani-Hamed-Dimopoulos-Dvali (ADD) model [36], in addition to our three known spatial dimensions, n extra dimensions are assumed to exist. These dimensions are compactified on a length scale of R. On this scale and below, gravity is able to penetrate multiple dimensions, making it stronger than it appears in the classical four dimensions. Thus, the true Planck scale for n extra dimensions $M_{\mathrm{Pl}(4+n)}$ is much lower than the four-dimensional Planck scale M_{Pl}:

$$M_{\mathrm{Pl}}^2 \sim M_{\mathrm{Pl}(4+n)}^{2+n} R^n \tag{1.1}$$

An interesting consequence of these assumptions is the possibility to produce black holes at collider experiments [37]: As derived from a semiclassical perspective, the Schwarzschild radius of a black hole with $M_{\mathrm{BH}} \sim$ TeV is larger than the minimum separation distance of two partons at the collision. Thus, a microscopic black hole can form.

This thesis will deal with two different models of black holes at the LHC: semiclassical black holes and quantum black holes.

- A *semiclassical black hole* is presumed to behave similarly to a classical astronomical black hole, except for the fact that it can only exist through additional extra dimensions: After formation, it reaches thermal equilibrium (*thermalizes*) and subsequently evaporates via Hawking radiation. During this phase, a large number of SM particles is produced. The relative abundance of these particles is expected to follow the number of degrees of freedom per SM particle. Among the possible final states are many that violate the lepton number conservation and thus make excellent experimental signatures with low expected SM contribution [38].

- A *quantum black hole* is too short-lived to thermalize. Instead, it behaves like a quantum object and decays into few particles, as illustrated in figure 1.6. Similar to the semiclassical case, lepton flavor violating final states are accessible in this model. For this thesis, a benchmark model has been chosen that decays solely into a e + μ pair. As this final state violates conservation of lepton number, there is no irreducible SM contribution. However, through experimental reality, particles from an SM process such as the decay of a t $\bar{\mathrm{t}}$-pair can be misclassified and contribute to the event yield [39].

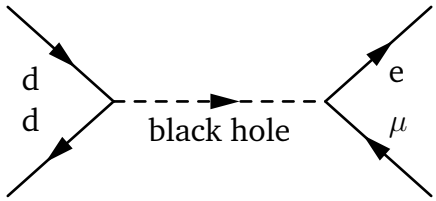


Figure 1.6: e μ signature of the quantum black hole (QBH) model. A black hole is created from the annihilation of two down quarks. The black hole subsequently decays into an e μ pair, violating the lepton number.

Seesaw Type III Model

The so-called Seesaw model aims to explain the low neutrino masses. In this model, the neutrino masses are considered to arise via the mediation of massive fermion partners. Two additional charged partners $\Sigma^{\pm}$ are predicted alongside a neutral Majorana particle Σ^{0}. Within the LHC, the particles are produced in pairs as electroweak decay products as shown in figure 1.7.

Each of the Σ particles is assumed to decay into a pair of SM particles. The most important decay channels are:

- $\Sigma^0 \to W^\pm \ell^\mp$
- $\Sigma^\pm \to W^\pm \nu$
- $\Sigma^0 \to Z\nu$
- $\Sigma^\pm \to Z\ell^\pm$
- $\Sigma^0 \to H\nu$
- $\Sigma^\pm \to H\ell^\pm$

Subsequent decays of the gauge bosons as well as the τ lepton allow for a plethora of accessible final states. The highest sensitivity is expected to arise from final states with at least three leptons, as the number of SM processes contributing to these final states is very low.

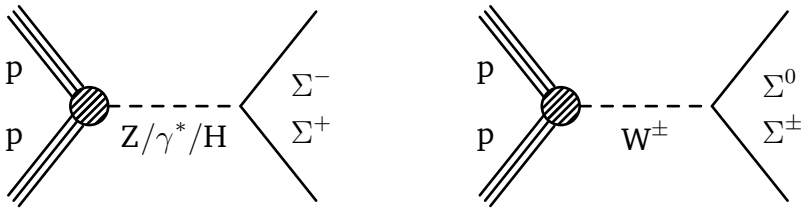


Figure 1.7: Production of heavy fermions $\Sigma^\pm, \Sigma^0$ from proton-proton collisions in the Seesaw Type-III model.

Sequential Standard Model and W′

Heavy gauge bosons are postulated by several extensions of the SM, most prominently by grand unification theories [40]. In order to provide a consistent simplified benchmark model, the *Sequential Standard Model* (SSM) [41] has been constructed, which introduces heavier copies of the SM W and Z bosons, consequently called W′ and Z′. The heavier copies interact with SM particles just like their SM counterparts, but because of the much larger mass, decays into the t and b-quarks are accessible. In some theories, this decay channel is even assumed to dominate as the W′ boson couples more strongly to fermions of the third generation [42, 43].

In this thesis, a variant of the SSM has been chosen where only the right-handed W′ boson (W'_R) exists. Possible decay channels are $W'_R \to \ell\nu_R$ and $W'_R \to t\bar{b}$. To further enhance the sensitivity, the hypothetical right-handed neutrino ν_R is assumed to be heavier than the W′ boson, therefore the former decay channel is kinematically forbidden and the branching ratio to $t + \bar{b}$ is 1. This corresponds to the scenario used in dedicated analyses performed by the CMS and ATLAS collaborations [44–46].

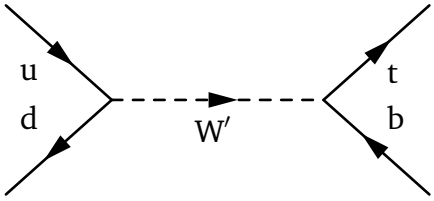


Figure 1.8: Experimental signature of the W′ model. The heavy boson is created from the annihilation of a quark pair and subsequently decays into a t + b pair.

1.4 Experiments

Beginning with Rutherford's experiments in the early 20th century [47], there have been an abundance of particle physics experiments using the approach of particle collisions. Today, most scattering experiments can be classified in two groups: Fixed target experiments and collider experiments. In the former case, accelerated particles collide with a fixed target, such as Rutherford's gold foils. In the latter case, two beams of particles are brought to collision at an interaction point. This allows for a higher center of mass energy and is thus favorable whenever the technical circumstances allow.

In both cases, outgoing scattered particles are registered. Because of the statistical nature of quantum mechanics, the Standard Model only describes scattering processes in terms of probabilities. These probabilities depend on particle properties before and after the collision. Controlling the incoming particles and counting outgoing particles with certain properties is thus a very efficient way of testing the theory.

1.5 The Large Hadron Collider

The to date largest particle collider experiment is located in a tunnel about 100 m underground between Geneva and the Jura mountains. It is operated by the *European Organization for Nuclear Research* (CERN) and is called *Large Hadron Collider* (LHC). The Large Hadron Collider [48] is a proton-proton collider with a circumference of 26.7 km. It has been built in the early 2000s in the tunnel of the former *Large Electron Positron Collider* (LEP). The tunnel contains the storage ring as well as accelerator structures and experiments. Within two beam pipes, in an ultra-high vacuum (about 10^{-11} mbar), groups of protons circulate in opposite directions. They are accelerated using radio-frequency cavities and kept on track by 1 232 superconducting dipole and 450 quadrupole magnets.

The protons are then brought to collision at one of the four *interaction points* along the ring (see figure 1.9). The energy of each accelerated proton before collision is 6.5 TeV in the rest system, resulting in a total center-of-mass energy of $\sqrt{s} = 13$ TeV.

Around each interaction point, a detector has been constructed. The task of two of the experiments is to perform very specialized measurements with lead ions (ALICE) and hadrons made of c or b quarks (LHCb). The other two experiments are aimed at more general measurements. These so-called *general purpose* experiments are the *Compact Muon Solenoid* (CMS) and the ATLAS experiment.

This work will focus on the CMS experiment, which will be discussed in the next sections.

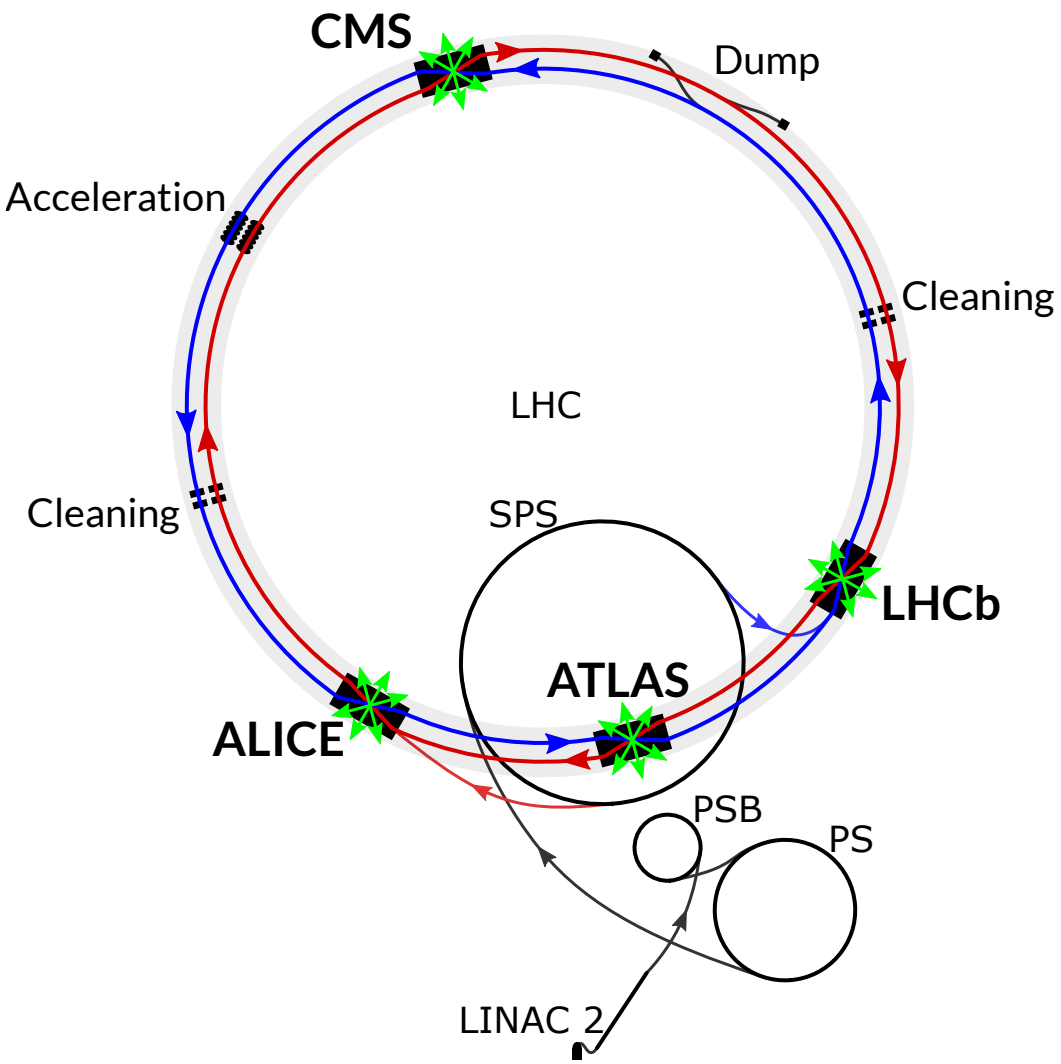


Figure 1.9: Schematic illustration of the CERN accelerator complex [49–51]. Protons first pass through several stages of preacceleration at the linear accelerator (LINAC) 2, Proton Synchrotron (PS), the Proton Synchrotron Booster (PSB) and the Super Proton Synchrotron (SPS). Eventually, they are injected into the LHC where they circulate in opposite directions. The two beams are further accelerated in radio-frequency cavities and their shape is restored within the cleaning sections of the ring. At each revolution, some of the protons are brought to collision at one of the four experiments CMS, ATLAS, LHCb and ALICE. Once they are not needed anymore, the remaining protons can be dumped into massive metal blocks at the dump site.

1.6 The Compact Muon Solenoid

The CMS detector consists of multiple particle detector subsystems surrounding the interaction point. Its goal is to measure outgoing particles created at proton-proton collisions. The observed properties include particle type, direction, momentum, energy and charge. Each of the detector subsystems is dedicated to measuring one or more of these characteristics. The subsystems are read out electronically and the data are later analyzed on a computing grid.

An overview of the detector can be seen in figure 1.10. The discussion of the detector subsystems in the following sections is based on [52] if not explicitly stated otherwise.

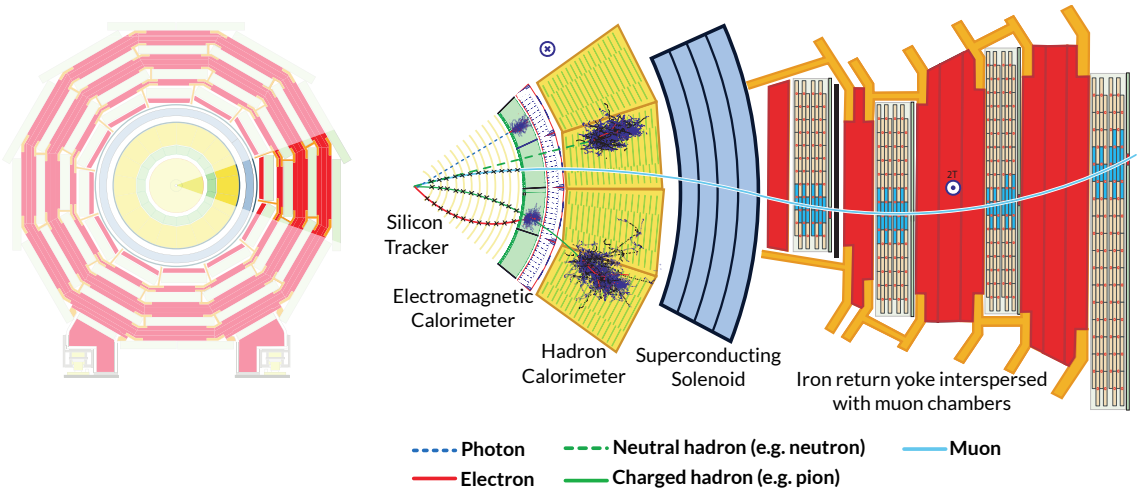


Figure 1.10: Slice through the CMS detector barrel. From left to right the following subsystems are drawn: the silicon tracker, the electromagnetic and hadron calorimeter, the superconducting solenoid coil, and the iron return yoke with the muon chambers. The solid lines represent charged particles which are bend due to the magnetic field [53, modified].

1.6.1 Detector Geometry and Coordinates

In a region starting at 23 m before the detector, both beam pipes are united [48]. The proton groups traveling in opposite directions share the same beam pipe, which defines the z-axis of the detector coordinate system. The collisions occur approximately at the interaction point at $z = 0$. The detector forms a barrel around the beam pipe. The barrel is subdivided into seven slices: five evenly sized *wheels* and one so-called *endcap* on each open side of the barrel. The central wheel is centered at $z = 0$. Detector layers in the wheels are mostly arranged in a cylindrical manner around the beam pipes, while the layers in the endcaps are mounted orthogonally to the beam pipe.

The direction of an outgoing particle originating in the interaction point can be characterized either using right-handed cartesian coordinates (x, y, z) or a spherical coordinate system (ϕ, θ). The x-axis connects the interaction point with the center of the LHC ring. The y-axis is perpendicular to the other axes, pointing upwards. Regarding spherical coordinates, the azimuthal angle ϕ is measured in the x-y plane with $\phi = 0$ along the x-axis. θ indicates the polar angle between the z-axis and the direction of interest. However, θ is rarely used because it is not invariant under Lorentz-boosts in the z direction. Instead, the Lorentz-invariant pseudo-rapidity $\eta := -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]$ is used to describe the angular separation between a particle and the beam pipe. The detector components in context of η ranges can be found in figure 1.11.

To indicate angular separation between particles, the quantity $R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, which is Lorentz-invariant, can be defined. η and R will be relevant in context of particle reconstruction.

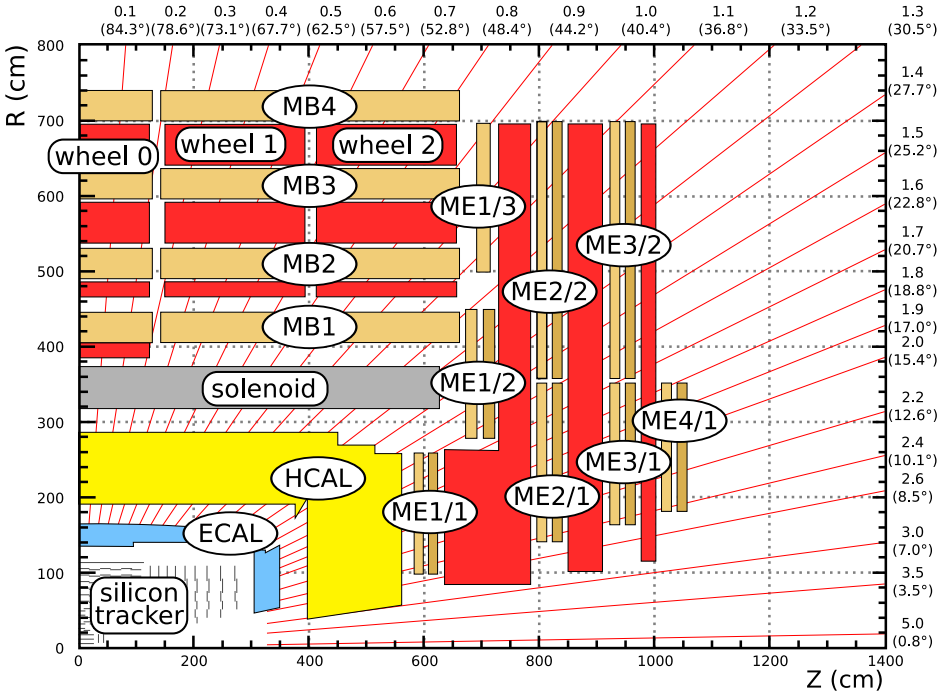


Figure 1.11: Schematic view of a CMS quadrant, taken from [54]. Besides the angle θ, the values for $|\eta|$ are shown around the frame. This illustration focuses on the muon system, which consists of barrel (MB) and endcap (ME) stations. It extends to $|\eta| = 2.4$.

1.6.2 Bunch Crossings and Pile-Up

For acceleration reasons, protons in the LHC are grouped into packets called *bunches*. At the interaction point, proton bunches from both directions collide every 25 ns, defining the so-called *bunch crossing*. During each bunch crossing, multiple proton pairs may interact with each other. This multitude of collisions is called *pile-up effect*.

Similarly, the collision of a single proton pair may result in multiple partons interacting with each other. This ambiguity is resolved by only analyzing the interaction with the highest momentum transfer, called *hard interaction*. Other interactions contribute to the so-called *underlying event*.

1.6.3 Subsystems

Inner Tracking System

The *inner tracking system* [55] is used for precise measurements of the direction and curvature of charged particles. It is composed of silicon semiconductor pixel layers surrounded by strip modules.

Each pixel cell consists of a silicon semiconductor in reverse bias direction. Charged particles passing through the semiconductor induce ionization and thus allow small currents to flow. These currents are amplified and read out by dedicated electronics. There are about 66 million pixel cells, each with a size of 100 µm $\times$ 150 µm, allowing for a high spatial resolution.

The strip modules function similarly to the pixel cells. For reduced production costs, typical cell sizes are 10 cm $\times$ 80 µm. There are 24 244 silicon strips in the tracker.

A very high spatial resolution around the interaction point is necessary for finding the origin of decay products. This is achieved by tracing the particle tracks back to the location of their parent particles. The information is later used to discard particle tracks originating in pile-up events or the underlying event (see section 1.6.2), as well as for identifying decays of B-mesons (see section 2.6.8).

Electromagnetic Calorimeter

The goal of the *electromagnetic calorimeter* (ECAL) [56] is to measure the energy of outgoing electrons, positrons and photons.

It consists of multiple layers of dense, transparent lead tungstate ($PbWO_4$) crystals with a total thickness of 25 radiation lengths. Within these crystals, high energetic

electrons and photons cause electromagnetic cascades: Electrons emit photons in the process of bremsstrahlung, while photons convert to electron-positron pairs in the process of pair-production. This leads to large numbers of lower-energy electrons and photons, until the electron energy loss is dominated by ionization losses and the photon energy is below the pair-production threshold of $2m_{\mathrm{e}} = 1\,022\,\mathrm{keV}$. Low energetic photons and those emitted by ionization or excitation are eventually registered using photodiodes. The initial particle energy is calculated from the light yield [2].

The energy resolution of the ECAL has been studied on $\mathrm{Z} \rightarrow \mathrm{e} + \mathrm{e}$ and $\mathrm{H} \rightarrow \gamma\gamma$ simulations. The electron energy resolution (at $E_{\mathrm{T}} \approx 45\,\mathrm{GeV}$) is below 2 % in the center of the barrel and 2 % to 5 % elsewhere. The photon resolution (at $E_{\mathrm{t}} \approx 60\,\mathrm{GeV}$) has been determined to be below 3 % in the barrel and 2 % to 5 % in the endcaps [57].

Hadron Calorimeter

In combination with the ECAL, the energy of hadrons produced during the collision is measured by the *hadron calorimeter* (HCAL) [58].

It is designed as a sampling calorimeter with alternating layers of brass absorbers and plastic scintillator tiles. Hadrons passing through the calorimeter scatter on nuclei of the denser brass absorbers and cause hadronic showers [2].

The shower products are detected via scintillation. In the plastic scintillators, the passage of charged particles excites electrons of the scintillator. During the subsequent deexcitation, optical photons are emitted, which are guided through wavelength-shifting fibers to photodiodes.

Because of the short nuclear interaction lengths in the absorber material, sampling calorimeters are more space efficient than homogeneous calorimeters. However, as some of the energy is deposited within the absorber material, the total shower energy can only be estimated from the scintillation light. Overall, the HCAL has a thickness of 7 to 11 interaction lengths, depending on η.

Similarly to the electromagnetic calorimeter, a high resolution can be achieved only as long as the showers do not leak out of the calorimeter.

Solenoid Magnet

The CMS solenoid magnet [59] is designed to deliver a magnetic field strength of 4 T. This is achieved by cooling the 220 t cold mass to 4.6 K, such that the NbTi wires become superconducting and the current can reach up to 19.5 kA. The magnet surrounds the calorimeters and has a diameter of 6 m. Within the coil, the magnetic

field is approximately uniform with the field lines being parallel to the beam pipe. Outside, the magnetic flux is returned using an iron yoke, which is interspersed with the muon system.

Particles passing the field are subject to the Lorentz force which subsequently bends their tracks into helices. This effect allows to distinguish between charged and neutral particles and enables measurement of particle momenta.

Muon System

The goal of the muon system [60] is to detect muons and measure their tracks, from which the muon momenta are calculated.

Because of their high mass compared to electrons, muons are less affected by the electromagnetic fields within matter, reducing their radiative losses. This makes them the only detectable particles to pass through all inner detector subsystems as well as the solenoid coil. For this reason, the muon system is the outermost part of the detector. It surrounds the solenoid coil and consists of *resistive plate chambers* (RPCs), *drift tubes* (DTs) in the barrel section and *cathode strip chambers* (CSCs) in the endcaps. The muon barrel section extends in $|\eta|$ up to 1.2, the endcap region lies between $|\eta|$ of 0.9 and 2.4. In the range between 0.9 and 1.2, the muon systems overlap.

Drift tubes have a length of about 2 m and a cross section of 13 mm $\times$ 42 mm. In the center of each tube, there is an anode wire. Between the wire and the tube walls exists a high electric potential. The tubes are filled with a gas mixture. As the charged muons pass through the gas, they ionize a few of the gas atoms. The released electrons drift along the electric field lines towards the anode wires. Eventually, the accumulated charges are deposited in the wires and are measurable as current spike [2]. By measuring the drift time and combining results from several layers, a spatial resolution of less than 100 µm can be achieved. However, determining the location in the wire direction is not possible with this arrangement.

In the endcap region, the muon system consists of cathode strip chambers instead of drift tubes. CSCs are gaseous detectors that contain several wires and cathode strips that are aligned orthogonally and read out separately. CSCs do not make use of drift effects, therefore are able operate under non-uniform magnetic fields and at high rates of muons, which makes them suited for use in the endcap region.

In both barrel and endcap, the DTs and CSCs are complemented with RPCs. Similarly to DTs and CSCs, RPCs are gaseous detectors, but are operated in avalanche mode. They possess a low spatial resolution but an excellent timing resolution, which helps to assign tracks to a bunch crossing.

The overall muon momentum resolution has been determined from measurements of cosmic muons and is approximately 7 % for a p_{T} range of 350 GeV to 2 000 GeV [61].

1.6.4 Trigger System

As mentioned in section 1.6.2, the LHC delivers collisions about every 25 ns, resulting in around 40 million events per second. Even with a raw event size of about 500 kB [62], storing all events would require a bandwidth of 20 TB s^{-1}. Neither the read-out chips of the submodules nor the computer network between the detector and storage can currently handle data traffic of this magnitude.

Therefore, the rate of events accepted for storage and analysis has to be drastically reduced. This is the task of the *triggering system*. Based on the physics content, statistics and bandwidth restrictions, triggers decide whether or not to store an event to disk.

Triggering is implemented in a two step process: First, the *level-1* (L1) trigger is consulted. It is implemented in dedicated hardware and uses information from the calorimeters and the muon system to reject low-energetic events. Events passing L1 will be read out, temporarily stored and further evaluated by the *high level trigger* (HLT). The HLT is implemented in software and runs on computers next to the detector cavern. It has access to the full detector information to decide whether an event is eventually to be stored or rejected [62].

1.6.5 Computing Grid

Reconstruction and analyses are ran on the stored events. For this purpose, the *Worldwide LHC Computing Grid* (WLCG) project was founded, which consists of several computing facilities operated by the CERN member states.

The facilities are organized in layers called *tiers*. There are two tier 0 computing centers, one at the CERN data center in Geneva and one at the Wigner Research Centre for Physics in Budapest, which store and reconstruct raw data. From there, the data are distributed to one of the 13 tier 1 data centers worldwide, where they are further processed, stored for safe-keeping and distributed to tier 2 facilities.

There are about 160 tier 2 computing centers, most of which are operated by universities and other scientific institutes. Their task is to provide computing power for event generation and analyses [63, 64]. One of these tier 2 centers is operated at the physics department of the RWTH Aachen University, currently providing around 5 200 processing cores and 3 PB of storage [65].

Model Unspecific Search

2

2.1 Motivation

A collision at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ can create a large amount of different particles. Solely by combinatorics, it follows that a lot of different final states are accessible.

Most dedicated searches tend to focus on one or a few final states that represent the signature of the theory under investigation. This leaves many final states not examined, either because they are not covered by any new physics model or because there is no analysis group currently working on the corresponding theory.

One of the goals of the model unspecific search is to gain knowledge from these additional final states. Furthermore, the model unspecific search aims to obtain a global interpretation of the agreement between simulation and observed data across a broad range of final states.

2.2 Previous Works

The approach of a model unspecific search is not a new concept: In 1998, a note about an unspecific search has been written at the L3 experiment (LEP) [66], and in 2004, a similar approach has been applied to data of the D0-experiment at the Tevatron proton-antiproton collider at Fermilab [67].

At CMS, the analysis has been developed and regularly applied to observed data since 2009 [68–79]. This thesis is based on the most advanced implementation of the MUSiC analysis, which has recently been applied to analyze the dataset of the 2015 LHC run [76] containing events with an integrated luminosity of $2.3\,\text{fb}^{-1}$ at $\sqrt{s} = 13\,\text{TeV}$.

A similar approach is also being pursued by the ATLAS collaboration, which published analyses using a model unspecific search at the LHC at $\sqrt{s} = 7\,\text{TeV}$ [80], 8 TeV [81] and 13 TeV [82].

2.3 Procedure

In the following sections, the existing analysis will be outlined.

The analysis starts out with reconstructed events of observed as well as Monte Carlo (MC) simulated data which have been centrally preprocessed by the CMS collaboration. As first step, requirements on events and physics objects are applied, discarding unwanted events and extracting particles for the MUSiC analysis (sections 2.5 and 2.6).

Afterwards, each event is sorted into so-called *event classes*, sets of events that share the same final state (section 2.7). For each event class, a set of kinematic variables of each event is calculated and *kinematic distributions* are aggregated (section 2.8).

Subsequently, an automated search algorithm finds the largest deviation between data and simulation in each kinematic distribution of each event class (section 2.10). The significance of each deviation is corrected for the look-elsewhere-effect using pseudo-experiments from SM-only simulations (section 2.10.4).

Finally, the observed deviations are aggregated and the distribution of deviations is compared to the expected distribution given the SM-only hypothesis as a measure for the global agreement between simulation and data (section 2.10.6).

2.4 Transverse Momentum and Energy

As prerequisite for the following sections it is helpful to introduce the so-called transverse quantities *transverse momentum* and *transverse energy*. This section provides a motivation and definition for both.

Interaction cross sections generally depend on the amount of momentum transferred, which can be obtained by calculating the combined invariant mass of the decay products. In order to achieve this goal, one must know all four components of the momentum of all created particles. If the final state contains neutrinos, which are not detectable by the detector, this is not achievable.

In other experiments like LEP, one could use conservation of momentum and the known momentum of colliding particles to reconstruct the momentum transfer for processes involving one neutrino. At the LHC this is impossible: The total longitudinal momentum of the proton is distributed unevenly between the constituent quarks and gluons and thus the momentum involved in a collision is only known up to statistical probability.

A mitigation is to ignore the longitudinal component altogether. This approach is pursued by many analyses at CMS by only regarding the kinematics in the transverse plane, orthogonally to the beam pipe. The projection of the kinematic (three-)momentum onto the transverse plane gives the transverse momentum $\vec{p}_{\mathrm{T}}$.

There are other properties of the transverse momentum that make it an interesting quantity: It can be straightforwardly measured from the tracks since the magnetic field lines are approximately parallel to the beam direction and the Lorentz force only acts in the transverse plane. Additionally, the transverse momentum is not only invariant to the actual distribution of momenta in the proton, but also to Lorentz boosts along the z-axis.

Assuming p_x, p_y, E and m can be observed directly (e.g. via track curvature) or indirectly (e.g. via particle identification), one can define the transverse momentum and transverse energy as follows:

$$\begin{aligned}
\vec{p}_{\mathrm{T}} &:= \left(p_x, p_y, 0\right)^T \\
p_{\mathrm{T}} &:= |\vec{p}_{\mathrm{T}}| \\
\vec{E}_{\mathrm{T}} &:= E \cdot \frac{\vec{p}_{\mathrm{T}}}{|\vec{p}|} = \frac{\vec{p}_{\mathrm{T}}}{\sqrt{1 - (m/E)^2}} \\
E_{\mathrm{T}} &:= \left|\vec{E}_{\mathrm{T}}\right|
\end{aligned}$$

Note that in the approximation of massless ($m \approx 0$) particles: $\vec{p}_T = \vec{E}_T$.

An additional quantity that is commonly defined in this context is *missing transverse energy*, denoted by $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$:

$$\begin{aligned}
\vec{E}_{\mathrm{T}}^{\mathrm{miss}} &:= -\sum_i \vec{E}_{T,i} \\
E_{\mathrm{T}}^{\mathrm{miss}} &:= \left|\vec{E}_{\mathrm{T}}^{\mathrm{miss}}\right|
\end{aligned}$$

This quantity denotes the imbalance of transverse momentum. If all particles in the final state can be correctly reconstructed, the imbalance is $E_{\mathrm{T}}^{\mathrm{miss}} = 0$. But there are several possible causes for this quantity to become non-zero: Incorrect measurements, such as noisy, dead or asymmetric modules, and non-detectable particles like neutrinos. If one or more neutrinos are produced during the interaction, one can observe only their combined four momentum as $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$. The same applies to hypothetical new physics particles which do not interact with the detector, making $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ an interesting quantity to work with.

2.5 Event Selection

The following section will discuss how events are selected to be used for the MUSiC analysis. Since this thesis aims to be a general evaluation of the MUSiC approach, it will not go deeply into detail about the algorithms used, nor the exact parameter values. For the sensitivity studies in chapter 5, however, specific values have to be chosen. If not explicitly stated otherwise, values of the 2015 analysis run are used and can be found in [76].

2.5.1 Triggers

In section 1.6.4 the triggering system has been introduced, a dedicated system that decides whether observed events are recorded to disk. In the vocabulary of an analyst, the term "trigger" does not only refer to the system itself, but also to the set of rules that are applied. In practice, triggering hardware is configured to run multiple sets of rules in a so-called *trigger menu*, in order to maximize the usage of readout bandwidth without overloading the available resources. The triggering decision is also simulated on MC generated events: Using the same information available to the triggering system and a trigger emulator, each event is assigned tags about which decision lead to it being stored.

Because multiple trigger configurations are running at the same time, the same event might end up in two separate trigger streams. Such duplicate events are removed in the analysis. An additional problem arises within the first few GeV beyond the p_{T} threshold: In this so-called *turn-on* region, disagreement between observed data and simulation has been found which is attributed to insufficiencies in the simulation. Because of this, an additional minimum p_{T} threshold close above the trigger threshold is applied.

Stream Name	trigger $p_{\mathrm{T,\,min}}$	custom $p_{\mathrm{T,\,min}}$
`HLT_Ele115_CaloIdVT_GsfTrkIdT`	115 GeV	120 GeV
`HLT_DoubleEle33_CaloIdL_GsfTrkIdVL_MW`	33 GeV, 33 GeV	50 GeV, 50 GeV
`HLT_Mu50`	50 GeV	53 GeV
`HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ` `HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ`	17 GeV, 8 GeV	35 GeV, 35 GeV

Table 2.1: Triggers used in the MUSiC analysis. Listed are the HLT names. The second, fourth and fifth triggers require two particles, where the second minimal threshold lies much lower than the first. Additional identification requirements are applied as indicated by the name [76]. Note that the single electron trigger p_{T} threshold is higher than the one used in [76] to be comparable to the 2016 scenario.

Table 2.1 provides a reference of triggers used in the analysis. In comparison to the 2015 analysis, the p_{T} threshold of the single electron trigger has been raised from 105 GeV to 115 GeV. This choice ensures consistency between the 2015 and 2016 analysis for the study of the discovery potential.

2.5.2 Event Filters

Event filters are used to mask off events that contain observations of unwanted detector behavior rather than physics content from a collision. Undesirable experimental effects include defective modules, anomalous electronics noise, saturation of subsystems by calibration lasers or particles scattered on residual gas in the beam pipe. They sometimes mimic real particles and events and thus cause the trigger to store the event. In the offline analysis, these events are recognizable by a large momentum imbalance of the reconstructed particles, resulting in large values for $E_{\mathrm{T}}^{\mathrm{miss}}$. Using a set of rules (so-called $E_{\mathrm{T}}^{\mathrm{miss}}$-filters) established at CMS, these events are excluded from the analysis. The list of applied filters can be found in [76].

2.6 Reconstruction and Selection of Objects

Now that a set of events to analyze has been established, this section will focus on the reconstruction and selection of the particles from an event for the analysis.

In practice, the reconstruction is performed centrally by CMS working groups. For each particle, several reconstructed candidates are stored, from which the analyst can select. Besides selecting only the particles relevant for the analysis, one often chooses a *working point* for particle identification criteria. The possible options are "tight", "medium" and "loose". Tight criteria are more strict, limiting the rate of misclassified particles at the cost of discarding some legitimate particles. On the other hand, loose criteria allow for more events, reducing the statistical uncertainty, with a higher probability of misclassification. As the name indicates, "medium" aims to provide a compromise between the two extremes.

For certain final states, it has been found that the probability of events being misclassified differs between MC simulation and observation. However, the scaling factor to correct this effect has not been determined in all final states regarded by MUSiC. Thus, to evade deviations that are caused by different misclassification rates, the particle selection generally requires tight identification criteria, except for light jets, which are selected with loose criteria to ensure a complete event description.

The following paragraphs aim to give a short overview over the reconstruction methods and object selection criteria used in the analysis.

2.6.1 Particle Flow

MUSiC uses the so-called *Particle-Flow* (PF) algorithm [83], which aims to combine measurements from separate detector subsystems into a consistent picture of all stable particles in the event. This is achieved by allowing the separate particle reconstruction algorithms to use shared definitions and communicate, avoiding duplicate interpretation of the same data.

The PF algorithm can be divided into three steps: Fitting of tracks and clustering, linking of tracks and clusters, reconstruction of particles.

In the first step, clusters of tracker cells that observed a current spike during the event are linked between the layers using a Kalman-filter, yielding reconstructed tracks. The same algorithm is also applied to the tracks of muons in the muon chambers. In the same step, clusters are found in the energy deposits of the calorimeters using a greedy clustering algorithm.

In the second step, tracks and clusters from the first step are linked to each other. Tracks from the inner tracking system are extrapolated to find matching clusters in the calorimeters. Similarly, clusters observed in the ECAL as well as the HCAL are related. Finally, tracks are matched between the inner tracker and the muon system.

The final step builds the actual particle candidates from the available information. First, muons are identified if the momentum of linked tracks is compatible. Tracks that have been successfully associated to a muon are then removed from the track collection. Electrons are next: The tracks are refit using a Gaussian-Sum Filter and matched to energy deposits in the ECAL. Again, information that has been related to an electron is removed from the track and cluster collections. The remaining tracks are afterwards associated to charged hadrons. Finally, the remaining energy clusters in ECAL and HCAL give rise to reconstructed photons and neutral hadrons respectively.

2.6.2 Jets

Jets form during the hadronization of quarks and gluons from the final state. The goal of jet reconstruction is to obtain the mass, energy and momentum of the originating parton or gluon.

In the detector, jets are measured as tracks and clusters of stable particles. Starting with the PF objects, in the first step, charged contributions from pile-up are removed. The undesired tracks are selected by displacement from the primary vertex and then removed from the set of candidate particles. After the correction, a jet clustering algorithm arranges the remaining particles into clusters. Within the MUSiC analysis,

the anti-k_t algorithm [84] with the characteristic radius $R = 0.4$ is deployed. It uses pairwise recombination to form cone-like clusters around the hardest particles.

As output, one obtains a set of clusters, each containing multiple PF objects. The combined four-vector of all particles within each cluster is calculated and assigned to the jet.

Measurements have shown that the energy resolution of jets is worse in observed data than in the MC simulation. As a mitigation, the transverse momenta of simulated jets are smeared with a p_{T} and η-dependent factor [85]. After the jet energy correction, the jets used in the MUSiC analysis are selected using established jet-id criteria at the tight working point [86]. It ensures high data quality and low misclassification rates by requiring a minimum p_{T} threshold of 50 GeV and poses certain additional requirements on the composition of the jet.

2.6.3 Muons

Muons are reconstructed from tracks in the inner tracker and in the muon system. For low energetic muons ($p_{\mathrm{T}} < 200$ GeV), the momentum resolution of the inner tracker is superior to the muon chambers. Therefore, the momentum information is extracted from the tracker measurement, while the muon chambers are only used for selecting the tracks used. At high momenta ($p_{\mathrm{T}} \geq 200$ GeV), the resolution can be increased by adding curvature measurement from the muon chambers and is therefore reconstructed using a combined methods.

The algorithms used for identification also depend on the muon momentum. Low energetic muons are identified using a series of cut-based criteria with a tight working point. High energetic muons are identified by a dedicated high-p_{T} algorithm which attempts to combine information from multiple muon chambers to achieve the best resolution [87].

Both identification criteria require the muon to be fully contained in the detectable range ($|\eta| < 2.4$). The second main goal of the identification is to ensure that the muon candidate originates from the hard interaction. Secondary muons from jets or from pile-up collisions are vetoed based on displacements from the primary vertex, isolation criteria and a p_{T} threshold.

After a muon has been successfully selected for the analysis, close-by low-energetic electrons, photons and jets are removed from the event description. This avoids double-counting of physics content and thus strives to restore a consistent event description.

2.6.4 Electrons

Electrons are recognized by their curved track in the tracker and energy deposit in the ECAL. For the reconstruction of low energetic electrons (below 100 GeV), the momentum information from the tracker and the ECAL are combined using the estimated uncertainty as a weight. At higher energies, using only the ECAL measurement has been shown to improve performance, thus the tracker momentum information is discarded.

The selection of electrons for the analysis is also applied in a piecewise manner: In the low energy regime, the tight cut-based electron identification criteria are required, while for high energies, the *high energy electron pairs* (HEEP) requirements [88] are imposed.

Both criteria aim to differentiate between electrons from the hard interaction and electrons from jets or pile-up. Again, a maximum displacement from the primary vertex is required, as well as a minimal isolation and a minimum p_{T} value. Jets are also suppressed by vetoing energy contributions in the HCAL.

In addition, converted electrons and photons are vetoed by settings tight quality requirements on the electron track.

2.6.5 Photons

Photons do not leave tracks in the inner tracker. Besides that, their experimental signature resembles the signature of electrons. Therefore, photons are selected and identified in a similar matter. The MUSiC analysis uses the cut-based photon identification criteria [89] over the entire energy range.

In order to veto photons emitted as bremsstrahlung, photons with more than one signal in the tracker are discarded.

2.6.6 Missing Transverse Energy

As mentioned earlier, neutrinos and possible new physics particles that do not leave a signature in the detector can be grasped via the momentum imbalance in the detector. This momentum imbalance is characterized in the missing transverse energy $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$, which can be treated just like other physics objects in this section.

All corrections that change the p_{T} value of PF candidates (e.g. jet energy corrections) are passed onto the $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ calculation to preserve a consistent event description. Using similar arguments as for the other particles, only $E_{\mathrm{T}}^{\mathrm{miss}}$ with $p_{\mathrm{T}} \geq 100$ GeV is identified as object in the analysis.

2.6.7 Possible Extensions

The existing analysis can be extended to identify more particle types by their decay products, called *tagging*. Possible tagging choices are b-tagging, τ-tagging or Z-tagging. This thesis will investigate the increase of sensitivity using b-tagging, as discussed below.

2.6.8 b-Tagging

The goal of b-tagging is to distinguish jets from b-quarks ("b-flavored jets") from those originating from light quarks. One way of achieving this goal is to exploit the long lifetime of hadrons formed during the hadronization of b-quarks. B-mesons for example possess a comparably long lifetime of $\tau \approx 1.5 \times 10^{-12}$ s [2]. This allows B-mesons to travel approximately 1 to 10 cm for energies between 100 and 1 000 GeV. With CMS' tracking resolution of up to 20 µm in the transverse and 30 µm in the longitudinal direction [90], a displacement of such magnitude can be resolved efficiently and is therefore an appropriate discriminator.

To introduce b-tagging in the MUSiC analysis, the *Combined Secondary Vertex version 2* (CSVv2) algorithm [91] is used in addition to the jet requirements described in section 2.6.2. The algorithm combines information of displaced tracks with information about displaced vertices in a multivariate analysis. The discriminator output is then compared to a threshold defined by the working point. For the sensitivity studies, the "tight" working point with a discriminator threshold of 0.935 is used. Studies of the 2015 data and simulation have found that with the tight working point, approximately 50 % of jets originating from b-quarks are correctly identified as such, while less than 0.5 % of the b-tagged jets do not actually originate from b decays [92].

2.7 Event Classes

An event class is a set of events sharing the same final state. The final state is indicated by the name of the event class: All events in the class named $2e + 1\mu$, for example, contain two electrons and one muon in the final state.

There are three types of event classes: *exclusive*, *inclusive* and *jet inclusive*.

Events in the exclusive event classes contain exactly the indicated (and no additional) particles in their final state. Each event thus belongs to exactly one exclusive event class.

Inclusive event classes are denoted with the suffix "+ X" in the name (for example 2e + 1μ + X). Their final state contains the explicitly stated particles plus any number (including zero) of additional ones. Each event can be assigned any number of inclusive event classes. One advantage of this procedure is a larger number of events per class which lower the statistical uncertainty. However, the correct combination of statistical results across multiple inclusive event classes is not trivial.

Jet inclusive event classes are denoted with the suffix "+ Njets" in the event class name (e.g. 2 e + 1μ + Njets). Events contained in a jet inclusive event class may contain any number of jets in addition to the explicitly stated objects. This increases the number of events per class, as effects like initial or final state radiation are ignored, leading to reduced statistical uncertainties.

In any case, the kinematic variables are only calculated from the objects explicitly stated in the class name, in the example of two electrons and one muon. Event classes are dynamically generated based on the classified event content. In the next sections, the total number of generated event classes will be denoted with n.

2.8 Kinematic Variables and Distributions

For each event, the MUSiC classification calculates three scalar variables from the particles defining the event class: the sum of transverse momenta, the invariant mass and the missing transverse energy.

As discussed in section 2.4, the sum of transverse momenta is the most directly observable variable and is invariant to Lorentz-boosts in the z direction. In the analysis it is denoted by Σp_{T} and calculated from the scalar sum of the magnitudes of the transverse momenta of all particles under consideration:

$$\Sigma p_{\mathrm{T}} := \sum_{i} \left|\vec{p}_{\mathrm{T},i}\right| \tag{2.1}$$

The invariant mass is, as mentioned before, very sensitive to resonances in the t-channel. If the event class definition contains missing transverse energy passing the selection criteria, the invariant mass will not be computed, instead the transverse

mass is used. The variables are denoted with M_{inv} and M_{T} respectively and defined as:

$$M_{\text{inv}} := \sqrt{\left(\sum_i E_i\right)^2 - \left(\sum_i \vec{p}_i\right)^2} \tag{2.2}$$

$$M_{\text{T}} := \sqrt{\left(\sum_i E_{T,i}\right)^2 - \left(\sum_i \vec{p}_{T,i}\right)^2} \tag{2.3}$$

The third kinematic variable is missing transverse energy. This variable is only calculated if the event class definition explicitly contains $\vec{E}_{\text{T}}^{\text{miss}}$ (above the threshold of 100 GeV) and describes the magnitude of the missing transverse momentum, denoted by $E_{\text{T}}^{\text{miss}}$ (as defined in section 2.4).

From each of the kinematic variables, *kinematic distributions*, histograms of the corresponding calculated variable, are aggregated. The bin size of these histograms is not constant, but instead chosen to match the estimated detector resolution in the variable under investigation. However, the bins are guaranteed to be at least 10 GeV wide.

2.9 Systematic Uncertainties

When comparing observed data to simulation, *systematic uncertainties* have to be considered. For various reasons, one will never be able to exactly predict how many events will pass the analysis. Possible causes include inaccuracies in the theoretical prediction, uncertainty on how a particle affects the detector and statistical uncertainty from performing a counting experiment on independent events. Therefore, it is only possible to estimate the observable event yield up to a certain interval. In this analysis, it is assumed that most of the arising uncertainties can be modeled with a normal distribution and thus they are denoted with a mean value and a 68 % confidence interval within $\pm 1\sigma$ around the mean. The uncertainties are propagated onto the event yield of each kinematic bin and used in the subsequent automated search as well as the generation of pseudo-experiments.

There are multiple quantities involved in the calculation of the predicted number of events N. Here, I will categorize these effects into three groups:

$$N = \underbrace{\sigma}_{\text{theory}} \cdot \underbrace{\mathcal{L}}_{\substack{\text{independent} \\ \text{measurements}}} \cdot \underbrace{A \cdot \epsilon}_{\substack{\text{experimental} \\ \text{response}}} \tag{2.4}$$

This categorization will be used in the following sections, which will give an overview about the currently implemented systematic uncertainties. As with section 2.6, the concepts are only roughly motivated, more comprehensive information and actual parameter values can be found in [76].

Theoretical Uncertainties

The dominating theoretical uncertainty stems from the perturbative calculation of the cross section. In quantum field theory, the interaction potential is expanded in discrete orders. The computation of higher order terms is analytically impossible and can often only be estimated up to an uncertainty. Therefore, the *cross section uncertainty* is chosen depending on the highest calculated order for a given sample, and lies between 5 and 50 % of the cross section.

The second considered theoretical uncertainty deals with the parton density function. Its task is to model the distribution of momentum between quarks and gluons in a proton. As it is not possible to analytically calculate this distribution, combined results from various deep inelastic scattering experiments are used and numerically extrapolated to high momenta. The uncertainty of these predictions is propagated onto the predicted yield by varying between multiple available prediction sets.

Independent Measurements

The systematic uncertainty on the luminosity directly affects the expected number of events. The current instantaneous luminosity is estimated by the activity in the pixel tracker. The measurement is complemented by a calibration from periodical precise van-der-Meer scans. This procedure is sufficiently accurate, as its uncertainty is estimated to be only 2.7 % [93].

The number of pile-up interactions in an event can be estimated from the total proton-proton cross section, combined with the instantaneous luminosity. The uncertainty of these independent measurements is propagated onto the number of pile-up collisions using a reweighting method to study the effect on the expected yield.

Event-Dependent Uncertainties

As discussed before, the energy of a reconstructed particle is estimated from the measured data of various detector components. The uncertainty on this mapping is called *energy scale uncertainty* and is considered for all particles: jets, electrons, muons, photons and $E_{\mathrm{T}}^{\mathrm{miss}}$. It is determined by studying differences between simulation and observed data of well-known physical processes. The uncertainty is then propagated onto the event by shifting the object momenta by $\pm\sigma(p_{\mathrm{T}})$ and repeating the classification. Note that the objects whose p_{T} values have been shifted may

subsequently pass certain selection criteria and thus contribute to different bins or even different event classes than the original event.

Similarly to the energy scale uncertainty, the simulated detector resolution is subject to uncertainty. This uncertainty is determined from independent observations, such as cosmic muons and Z-decays, and is also propagated by shifting object momenta. The energy resolution uncertainty is only significant for muons and jets, depends on η and p_{T} and is 3 to 5 % of the objects' momentum.

In addition to uncertainties on the measured momenta, the analysis also considers uncertainties on the identification of particles. Two effects have to be taken into account: Particles may not be recognized with the correct type (*identification efficiency*) or contributions that do not originate from particles might be incorrectly classified as particles (*misidentification rate*). Ideally, the probability of these effects occurring would be the same in simulation as in observed data, therefore there are several techniques that aim to correct the MC expectation.

To reproduce the identification probability of data in simulation, the expected yield is multiplied by so-called scale factors. The uncertainties of these scale factors are propagated by shifting the scale factors and observing the effects on the event yield.

The treatment of incorrectly identified objects is more difficult: For each identified particle, there is a nonzero chance that it may not actually correspond to the reconstructed particle type. To account for differences of this behavior between observation and prediction, incorrectly classified objects are identified by matching simulated particles to reconstructed candidates. Subsequently, for each incorrectly classified object, an uncertainty of 1 event is assigned to the corresponding bin (for electrons, muons and photons) [76]. For b-tagged jets, the mismatch of the misclassification rate between data and prediction has been studied in 2016 data and found to be up to 30 % [94, Fig. 73], so 0.3 events of uncertainty are assigned for each misclassified event.

The actual misclassification rates depend on the final state but are in the order of 1 % for most event classes.

Statistical Uncertainty

The amount of simulated MC events is limited, mostly by computational costs. When comparing the scaled number of events to observed data, one has to take into account statistical effects of the discrete event production. The scaled number of events is calculated by multiplying the generated number of events N with a scalar s. The

statistical uncertainty on the generated number of events is derived from the standard deviation of the Poisson distribution:

$$n = s \cdot N, \quad \sigma_{\text{stat},N} = \sqrt{N} \tag{2.5}$$

$$\Rightarrow \sigma_{\text{stat},n} = s\sqrt{N} \tag{2.6}$$

$$\Rightarrow \text{relative uncertainty} = \frac{\sigma_{\text{stat},n}}{n} = \frac{s\sqrt{N}}{sN} = \frac{1}{\sqrt{N}} \tag{2.7}$$

One can see that an increase in the number of simulated events directly decreases the statistical uncertainty on the prediction.

2.10 Search for Deviations

After each event has been classified, the kinematic distributions have been aggregated and the systematic uncertainties have been propagated onto the event yield follows the automated search for deviations. The search is performed on each distribution of each event class separately as follows: First, multiple connected bins are combined to a *region*, according to the rules in sections 2.10.1 and 2.10.2. Then, a test statistic θ for each region is calculated (section 2.10.3). Of all possible connected bin combinations, the region with the lowest value of θ, called θ_{min}, is selected. This region will be called *region of interest* (RoI). If two or more regions with the same value of θ_{min} are found, the smallest one is classified as RoI.

Subsequently, a global $\tilde{p}$-value for the region is calculated. This $\tilde{p}$-value expresses the probability of finding a deviation with $\theta \leq \theta_{\text{min}}$ anywhere in the distribution purely by chance. Details on the procedure can be found in section 2.10.4.

2.10.1 Search Space

The search is performed on so-called regions. A region is a set of adjacent bins and is represented by the combined number of expected and observed events as well as a combined uncertainty.

As new physics phenomena are not expected to appear resonantly in the Σp_{T} and $E_{\text{T}}^{\text{miss}}$ kinematic distributions, each region in these distributions must consist of at least 3 adjacent bins. Within the M_{inv} and M_{T} distribution, narrow resonant deviations are possible, thus the minimal number of bins in these distributions is set to 1. In addition, the MUSiC analysis employs a veto on regions where the simulation is incomplete, which is described in the next section.

The combined number of events and the combined uncertainty for each region is calculated as

$$N_{\text{total,exp}} = \sum_i N_{i,\text{exp}} \tag{2.8}$$

$$N_{\text{total,obs}} = \sum_i N_{i,\text{obs}} \tag{2.9}$$

$$\sigma_{\text{total}} = \sqrt{\sum_i \sigma_i^2 + 2\sum_{i,j} \rho_{i,j}\sigma_i\sigma_j} \tag{2.10}$$

where $\rho_{i,j}$ is the correlation coefficient for the uncertainty σ between the bins i and j.

Some uncertainties, such as the statistical uncertainty on the simulated number of events, are completely uncorrelated between the bins. The correlation coefficient thus is $\rho_{i,j} = 0$ and the total uncertainty is the quadratic sum of the bin uncertainties.

Other uncertainties are correlated between bins. An example for this is the uncertainty on the luminosity measurement, which affects all bins of all distributions the same way. Any uncertainty that is assumed to be correlated to some extend is treated as fully correlated ($\rho_{i,j} = 1$) and the combined uncertainty is calculated as linear sum. This method possibly overestimates the total uncertainty, in case the correlation coefficient is somewhere between 0 and 1, leading to a lower sensitivity.

2.10.2 Region Veto

Although the analysis aims to cover as much phase space as possible by combining large amounts of MC events, there are regions where the MC description is incomplete. Because the null-hypothesis (the Standard Model) is not correctly represented in some areas, one must not probe for new physics in these regions, i.e. any statistical inference from these regions would be invalid.

Therefore, several rules have been identified and implemented in order to skip calculation of θ values for incorrectly modeled regions. These rules may only depend on the SM values, not on the observed data per region. In addition, note that any region is usually embedded in a larger region with a higher confidence in simulation. If this is not the case, i.e. the largest region, which is the integral of the entire distribution, is vetoed, the event class is excluded from the analysis. A similar criterion will be introduced in section 2.10.5.

Some of the rules do not only depend on the region itself, but also on a *neighborhood*. The neighborhood is defined as combination of adjacent bins below and above the region boundaries, 4 in each direction. In the neighborhood, a set of *dominant*

neighborhood processes is defined as the set of processes with the largest contributions, which in total make up more than 95 % of the total yield in the neighborhood.

The following rules are tested; if any of them matches, the region is vetoed, no θ value is calculated and thus not considered for RoI.

- No or negative total MC yield: $N_{\mathrm{SM}} \leq 0$.
- Any process contributes more than 2 % of the event yield in the negative direction.
- Any process which has been identified as dominant in the neighborhood has no or negative total MC yield.
- Region has large total statistical uncertainty: $\frac{\sigma_{\mathrm{stat}}}{N_{\mathrm{SM}}} > 60\,\%$.

Negative total MC yield can arise from few events with negative weights. In order to account for possible double counting between leading order matrix element calculations and parton showering, which is for some samples performed by a different application, MC generators may assign negative weights to a small number of events [95]. Therefore, especially in phase space areas with few simulated events, situations can occur where only negative event weights contribute to a region. In these regions the original assumption, having enough events for a regression to the mean, is not fulfilled, thus the SM is not correctly represented and the regions must be vetoed.

2.10.3 Test Statistic

For each region, a test statistic is calculated. Its goal is to express "extremity" of a deviation based on the number of expected events N_{SM} and their uncertainty σ_{SM}, and the number of observed (or pseudo-observed) events N_{obs}.

The choice of such a measure is ambiguous: All test statistics have strengths and weaknesses and are more or less sensitive in various areas of their three-dimensional input domain. For the MUSiC analysis, especially for the subsequent calculation of $\tilde{p}$, it is desirable to use a test statistic based on a null-hypothesis for which pseudo-experiments can be reliably generated.

Ideally, the test statistic would express a local p-value directly, in order to be comparable to other CMS results. However, as shown in the following section, only the ordering of θ-values is relevant for the $\tilde{p}$ calculation and thus any test statistic $\theta(p)$ that is monotonous in p will suffice.

The following sections motivate the test statistic as used in the analysis.

The first (reasonable) assumption in the derivation is that the probability of observing a certain number of events given an exactly known prediction N_{true} follows a Poisson distribution:

$$\Pr(N_{\text{obs}}|N_{\text{true}}) = \frac{{N_{\text{true}}}^{N_{\text{obs}}}}{N_{\text{obs}}!} e^{-N_{\text{true}}} \tag{2.11}$$

In context of a p-value which expresses the probability of obtaining a deviation equally or more extreme than the observed one just by chance, one has to sum over all "more extreme" cases. In case of an excess ($N_{\text{obs}} \geq N_{\text{true}}$), this is

$$p(N_{\text{obs}}|N_{\text{true}}) = \sum_{n=N_{\text{obs}}}^{\infty} \frac{{N_{\text{true}}}^{n}}{n!} e^{-N_{\text{true}}} \tag{2.12}$$

For a deficit, the sum is finite and corresponds to:

$$p(N_{\text{obs}}|N_{\text{true}}) = \sum_{n=0}^{N_{\text{obs}}} \frac{{N_{\text{true}}}^{n}}{n!} e^{-N_{\text{true}}} \tag{2.13}$$

In reality, N_{true} can never be known with perfect precision. Instead, the expectation can be expressed by a probability density $\rho(x|N_{\text{SM}}, \sigma_{\text{SM}})$ which is then folded into the p value:

$$p(N_{\text{obs}}|N_{\text{SM}}, \sigma_{\text{SM}}) = \int_{-\infty}^{\infty} p(N_{\text{obs}}|x) \cdot \rho(x|N_{\text{SM}}, \sigma_{\text{SM}})\, \mathrm{d}x \tag{2.14}$$

Here, one can imagine that x takes the role of all possible N_{true} values and ρ assigns different weights to contributions from equations 2.12 and 2.13.

In the case of MUSiC the smearing procedure is applied after deciding whether a region contains an excess or a deficit, which allows one to move the piecewise definition out of the integral:

$$p(N_{\text{obs}}|N_{\text{SM}}, \sigma_{\text{SM}}) = \begin{cases} \displaystyle\int_{-\infty}^{\infty} \rho(x|N_{\text{SM}}, \sigma_{\text{SM}}) \sum_{n=0}^{N_{\text{obs}}} \frac{x^n}{n!} e^{-x} \mathrm{d}x & N_{\text{obs}} < N_{\text{SM}} \\ \displaystyle\int_{-\infty}^{\infty} \rho(x|N_{\text{SM}}, \sigma_{\text{SM}}) \sum_{n=N_{\text{obs}}}^{\infty} \frac{x^n}{n!} e^{-x} \mathrm{d}x & N_{\text{obs}} \geq N_{\text{SM}} \end{cases} \tag{2.15}$$

This is turn also implies that contributions of the integral where e.g. $N_{\text{SM}} < N_{\text{obs}} < x$ are treated as if they belong to an excess and thus sum up from N_{obs} to infinity. This conservative choice of procedure is not expected to have a high impact to the results because it is also corrected during the calculation of $\tilde{p}$.

The choice of the prior $\rho(x|N_{\mathrm{SM}}, \sigma_{\mathrm{SM}})$ is not unambiguous. For previous MUSiC publications and the signal study in this thesis, a truncated Gaussian prior is used:

$$\rho_N(x|N_{\mathrm{SM}}, \sigma_{\mathrm{SM}}) = C(N_{\mathrm{SM}}, \sigma_{\mathrm{SM}}) \cdot \Theta(x) \cdot \exp\left(\frac{(x - N_{\mathrm{SM}})^2}{2{\sigma_{\mathrm{SM}}}^2}\right) \tag{2.16}$$

Here, $\Theta(x)$ denotes the step-function and $C(N_{\mathrm{SM}}, \sigma_{\mathrm{SM}})$ is a normalization constant to correct for the truncated ($x < 0$) part of the normal distribution.

However, it has been under discussion in earlier theses [68] to use a log-normal prior instead:

$$\rho_{LN}(x|N_{\mathrm{SM}}, \sigma_{\mathrm{SM}}) = \frac{\Theta(x)}{\sqrt{2\pi}|x| \ln\left(1 + \frac{\sigma_{\mathrm{SM}}}{N_{\mathrm{SM}}}\right)} \exp\left(-\frac{1}{2}\left(\frac{\ln(x/N_{\mathrm{SM}})}{\ln\left(1 + \frac{\sigma_{\mathrm{SM}}}{N_{\mathrm{SM}}}\right)}\right)^2\right) \tag{2.17}$$

The log-normal prior represents the multiplicative nature of uncertainties and features a larger coverage range, but has disadvantages concerning the generation of pseudo-experiments. In section 4.2, the alternative of a log-normal prior will be discussed in further detail. For a rough derivation of this formula, see appendix A.3.

The p-value in equation (2.15) in combination with the Gaussian prior will be denoted as $\theta(N_{\mathrm{obs}}, N_{\mathrm{SM}}, \sigma_{\mathrm{SM}}) = p(N_{\mathrm{obs}}|N_{\mathrm{SM}}, \sigma_{\mathrm{SM}})$ in this thesis. Section 4.1 will discuss to what extends the aforementioned assumptions match the definition of a p-value. Furthermore, whenever comparisons to the log-normal prior are drawn, it will be denoted with θ_{LN}.

2.10.4 Consideration of the Look-Elsewhere-Effect

So far, only the local magnitude of a deviation has been estimated. In order to quantify and compare it, it has to be put in context: As the number of regions n grows, the probability of finding a deviation with a significance of α increases, roughly $p \sim 1 - (1 - \alpha)^n$. The analysis accounts for this so-called *look-elsewhere effect* by calculating a corrected post-trial p-value called $\tilde{p}$. It expresses the probability of finding a deviation equally large or larger than the observed one somewhere in the distribution, solely by chance.

The simplest way of adapting the local magnitude would be to apply an analytical correction factor to it. A suitable correction factor could be estimated using combinatorics if the possible values of the local p-value were uncorrelated. This method is also called *Bonferroni correction* [96]. In the case of the automated search, a purely analytically calculation is not possible because the connected bin regions overlap and thus the local significances are correlated.

Instead a simulation-based approach is used: Possible experimental outcomes are simulated based on the Standard Model, and refer to each simulation round as *pseudo-experiment*. During each pseudo-experiment, a substitute observed distribution is randomly drawn from the expected number of events. The entire automated search is repeated on each generated distribution and its result $\theta_{\text{pseudo, min}}$ is stored.

Finally, the global $\tilde{p}$-value is calculated by determining the rate of pseudo-experiments where a more significant deviation has been found: $\theta_{\text{pseudo, min}} < \theta_{\text{obs, min}}$. Thus, the $\tilde{p}$-value automatically fulfills the definition of a p-value ("probability to find a deviation equally large or larger as the observed one by chance, given the null hypothesis") if the pseudo-experiment generation corresponds to the null hypothesis. For this thesis, the number of pseudo-experiments has been chosen to be $k = 10\,000$, therefore the minimal meaningful result is $\tilde{p} = \frac{1}{10000} = 1 \times 10^{-4}$, which corresponds to a two-sided Z-score of 3.7σ (see appendix A.4).

Figure 2.1 illustrates the procedure for generating a pseudo-experiment. The pseudo-experiment generation is performed separately for each bin of the distribution. First, a possible true value N_{true} is generated from the known SM mean and systematic uncertainties, then the counting experiment is simulated by drawing from a Poisson distribution $\mathcal{P}$.

In order to simulate reality as close as possible, during each pseudo-experiment, every uncertainty assumes a certain value, fully correlated between all event classes and bins. This is done by drawing a "bias" x_i from a standard normal distribution $\mathcal{N}$ and subsequently multiplying it with the magnitude of each uncertainty of each bin σ_i to determine its effect on the event yield:

$$N_{\text{obs}}{}' \sim \mathcal{P}(\lambda = N_{\text{true}}{}') \tag{2.18}$$

$$N_{\text{true}}{}' = N_{\text{SM}} \cdot \sum_i x_i \sigma_i \tag{2.19}$$

$$x_i \sim \mathcal{N}(\mu = 0, \sigma = 1) \tag{2.20}$$

2.10.5 Minimum Yield

For further statistical inference it is important to note that any non-zero event yield in a certain final state will cause the creation of an event class corresponding to that final state. Therefore, by pure combinatorics, the number of event classes grows as more simulated events are added to the analysis. These almost empty event classes cause problems during the pseudo-experiment procedure: As the event yield is very low, the pseudo-data generation will generate $N_{\text{obs}} = 0$ in most cases and $N_{\text{obs}} = 1$

Input: $\mathsf{N}_{i=1..n}$: number of expected events in each bin i
Input: $\mathsf{corr}_{i=1..n,j=0..m}$: correlated uncertainty j in each bin i
Input: $\mathsf{stat}_{i=1..n}$: statistical uncertainty in each bin i
Output: $\mathsf{n}_{i=1..n}$: pseudo-data event yield in bin i

```
// Once for all distributions:  choose systematic uncertainty bias
```
foreach *correlated uncertainty* j *in distribution* **do**
    $\mathsf{x}_j \leftarrow \text{RandNormal}(\mu = 0, \sigma = 1)$
end

foreach *bin* i *in distribution* **do**
    // Start with the expected yield
    $\mathsf{n}_i \leftarrow \mathsf{N}_i$

    // Simulate correlated uncertainties using the pre-chosen bias
       multiplied by an individual magnitude
    foreach *correlated uncertainty* j *in bin* i **do**
        $\mathsf{n}_i \leftarrow \mathsf{n}_i + \mathsf{x}_j \cdot \mathsf{corr}_{i,j}$
    end

    // Simulate (uncorrelated) statistical uncertainty by drawing a
       number from a normal distribution
    $\mathsf{n}_i \leftarrow \text{RandNormal}(\mu = \mathsf{n}_i, \sigma = \mathsf{stat}_i)$

    // Simulate independent counting experiment
    $\mathsf{n}_i \leftarrow \text{RandPoisson}(\mathsf{n}_i)$
end

Figure 2.1: Pseudo-event generation algorithm. The algorithm takes absolute values for uncertainties and number of events as input and sequentially stacks pseudo-true values for uncertainties on top of the expected yield. Finally, a random number is drawn from a Poisson distribution, simulating the physical process of a counting experiment.

in very few cases. The 0-case will be insignificant and the few 1-cases will yield a very significant θ_{pseudo} and $\tilde{p}$-value. Because of this discretization, the $\tilde{p}$-value will not be uniformly distributed, which is one fundamental assumption for the following statistical inference. For this reason, event classes with a total integrated yield of $N_{\text{SM}} < 0.1$ are excluded from statistical inference, but are checked by hand if they contain observed data events.

A similar choice is also applied in a model-independent search in LHC data by the ATLAS collaboration [81, 82].

2.10.6 Distribution of $\tilde{p}$-Values

Similarly to the look-elsewhere effect in a single distribution, the probability of finding a significant deviation in any distribution increases as the search is extended to multiple event classes.

Since $\tilde{p}$ fulfills the definition of a p-value, the probability of observing an event class with a $\tilde{p}$-value of P in the absence of new physics is P, thus $\tilde{p}$-values should be uniformly distributed. Note, that this also implies that, in the absence of correlations, among n event classes one expects the most significant class to have a $\tilde{p}$-value of $\frac{1}{n+1}$ (see appendix A.2), e.g. $\tilde{p} = 0.002, Z = 2.9$ with 500 event classes.

This presumption is subsequently used to set $\tilde{p}$-values into context. For this purpose, the $\theta_{\text{pseudo, min}}$ values of all pseudo-experiments can be reused: In the correction step, one $\theta_{\text{pseudo, min}}$ value for each pseudo-experiment round in each event class was calculated. Now, for each event class, each pseudo-experiment result can be corrected using all other $k - 1$ pseudo-experiments from the same event class. This procedure is applied to all n event classes, yielding $n \cdot k$ $\tilde{p}$-values. Then, a histogram of all $\tilde{p}$-values belonging to one pseudo-round are aggregated (n values). All k histograms are combined to calculate the mean, median, 68 % and 95 % intervals on each bin content of the $\tilde{p}$ distribution. Similarly, the distribution of $\tilde{p}$-values of all event classes from observed data is calculated. Up to this thesis, both distributions are only qualitatively compared (by eye). In section 4.4, a quantitative solution to this approach will be discussed.

2.11 Signal Study

The goal of the signal study is to investigate how signatures of new physics propagate to the result of the analysis. For this purpose, simulated events of new physics processes are classified into event classes and merged with the classification results from the Standard Model. Experimental uncertainties between the SM classification results and the signal results are assumed to be fully correlated and accordingly

combined. However, no uncertainty on the signal cross section is assumed, because the signal study investigates one specific cross section scenario.

Subsequently, pseudo-experiments are generated from the combined signal and background and compared to the SM simulation by the automated search for deviations, using the pseudo-experiments instead of observed data. Subsequently, the resulting θ_{min} values are corrected for the look-elsewhere effect using a precomputed set of SM θ values, resulting one distribution of $\tilde{p}$-values for each signal pseudo-experiment round.

The interpretation of the $\tilde{p}$-distributions is performed qualitatively, analogous to the interpretation of real observed data. If the resulting $\tilde{p}$-values are not distributed uniformly anymore and there is a deviation from SM-only simulation or the uniform distribution, it is interpreted as sensitivity of the MUSiC analysis towards that specific model. Because the qualitative inference is very vague, I will discuss an alternative in section 4.4, where multiple pseudo-experiment rounds are used to calculate the quantified discovery potential towards the model under investigation.

2.12 Implementation

The analysis combines various tools from different authors (CMS, local institute collaboration, open source), which are used on a heterogeneous computing environment (WLCG, local computing cluster, local computer). From raw data to statistical inference there are many intermediate states which have to be managed. Doing this manually does not only require manpower, but is also more prone to human error, such as confusing files or missing steps. Therefore, several levels of automation and validation are critical for a rapid development of the analysis.

The following sections illustrate the practical steps required to reproduce the analysis and point out various starting points for automation. In addition, this section intends to be a reference about the exact software packages and versions used for the signal study.

Events triggered by the detector are stored on the WLCG. They are centrally reconstructed by CMS and stored on the grid in the MINIAOD format [97], accessible only to CMS members. In the first step, these events are preprocessed by the Aachen Institute IIIA working group and stored by sample on the local Tier-2 storage element in the *Pxl I/O* file format [98], which has also been developed in Aachen. The programs used for preprocessing make use of a software framework for CMS analyses, called CMSSW. Preprocessing of 2015 SM and signal samples was performed with CMSSW version 7.6.3. In addition to the Pxl I/O files, a working-group wide shared database

is used for meta-information about samples (such as cross sections and correction factors).

The further analysis is performed with version 8.0.26 of the CMSSW framework, which includes ROOT version 6.06/01 [99], C++11 with GCC 5.3.0 and PYTHON 2.7.11 [100]. The MUSiC analysis starts with the classification of events on the local computing element. The classification is implemented as plugin in the PXLANALYZER, a framework written in C++11 collaboratively by the Aachen Institute IIIA working group. The PXLANALYZER reads Pxl I/O files from the previous step, classifies the event and stores the resulting event classes in the ROOT file format. Subsequently, event samples that have been classified separately are merged together. The entire process is managed by the LUIGI framework [101](version 2.4.0), which allows automated execution of dependent task such as classification and sample merging.

As next step follows the automated search for deviations. The automated search is implemented in PYTHON as well as C++11. A custom script converts the classified ROOT files into JSON files, configuring the automated search. It also distributes the workload into multiple *jobs* which are executed in parallel on the local computing cluster. Each job then operates on one or more JSON files, searching for deviations between simulation and data or optionally generating and evaluating pseudo-experiments. After retrieval from the computing grid, its output is parsed into a local SQLITE database for statistical analysis.

Eventually, several tools written in PYTHON perform statistical inference (such as $\tilde{p}$ and $\hat{p}$) and creation of figures from the search results. These tools make use of the libraries NUMPY 1.11.1 [102], which provides fast vectorized math, SCIPY 0.16.1 [103] for statistical analysis and ROOT as well as MATPLOTLIB 1.5.2 [104] for visualization.

The dependency of intermediate states, from the preprocessed ("skimmed") samples to results published in this thesis, is illustrated in appendix A.5, which served as orientation for the employment of automation.

2.12.1 Lookup-Table

During the automated search algorithm, most of the computation time is spent calculating θ (see appendix A.6), which involves integration and summation over a large number of terms.

To decrease the amount of time spent in this step, I implemented a three dimensional *lookup table* (LUT), as suggested by T. Hebbeker (personal communication, 2015). The table consists of θ values for the most frequently used input parameters, such that

the expensive runtime computation is replaced by retrieving a precomputed value from a static block of memory.

The LUT is generated as a separate file during compilation and is about 80 MB in size.

Implementation

The precise lookup procedure for $\theta_{\text{LUT}}(N_{\text{obs}}, N_{\text{SM}}, \sigma_{\text{SM}})$ is as follows:

1. Determine real-valued indices i, j, k from the parameter values $N_{\text{obs}}, N_{\text{SM}}, \sigma_{\text{SM}}$
2. Check whether the calculated indices are valid, e.g. within table bounds
3. Fetch four adjacent values from memory: N_{obs} entry at $\lfloor i \rfloor$, N_{SM} and σ_{SM} entries at $\lfloor j \rfloor, \lceil j \rceil, \lfloor k \rfloor, \lceil k \rceil$[1].
4. Use two dimensional linear interpolation in the j, k plane to get a better estimate for θ:

$$\begin{aligned}\theta_{\text{LUT}}(i,j,k) \approx\ & \theta(\lfloor i \rfloor, \lfloor j \rfloor, \lfloor k \rfloor)(\lceil j \rceil - j)(\lceil k \rceil - k) \\ & + \theta(\lfloor i \rfloor, \lceil j \rceil, \lfloor k \rfloor)(j - \lfloor j \rfloor)(\lceil k \rceil - k) \\ & + \theta(\lfloor i \rfloor, \lfloor j \rfloor, \lceil k \rceil)(\lceil j \rceil - j)(k - \lfloor k \rfloor) \\ & + \theta(\lfloor i \rfloor, \lceil j \rceil, \lceil k \rceil)(j - \lfloor j \rfloor)(k - \lfloor k \rfloor)\end{aligned}$$

   Note that $j - \lfloor j \rfloor = 1 - (\lceil j \rceil - j)$.

5. Discard the result if it is too small ($\theta_{\text{LUT}} < 0.005$)

If the value can not be found in the LUT, either because of the table bounds or an otherwise untrusted value, the implementation falls back to the full expensive calculation of θ. Thus, the efficiency of the LUT is determined by the probability that a requested value is contained in the table.

In order to improve accuracy around the most commonly used parameters ($N_{\text{obs}} = N_{\text{SM}}$, N_{obs} small), the LUT is organized in three regions, as illustrated in figure 2.2. In the so-called "grid region", the point spacing for the expected number of events is independent if the observed number of events. This results in a rectangular grid up to an observed yield of 10 and an expected yield of 20. The second, so-called "linear region" stretches up to an observed yield of 100. In this region, every integer observed value corresponds to exactly one index, $i = N_{\text{obs}}$. Lastly, there is an exponential region, in which the spacing of observed points gradually becomes less dense, keeping the relative spacing constant. In the dimension of the expected number of events, the

[1] $\lfloor x \rfloor$ denotes the floor, $\lceil x \rceil$ the ceiling function, always rounding down and up respectively.

LUT is more densely spaced around $N_{\text{SM}} = N_{\text{obs}}$, because this is statistically the most common scenario. There is an exponential falloff in density which in theory expands from $N_{\text{SM}} = 0.1 \cdot N_{\text{obs}}$ up to $5.0 N_{\text{obs}}$. Very distant values, however, are vetoed by the $\theta_{\text{LUT}} > 0.005$ threshold. In the third dimension, uncertainty, the grid is exponentially spaced in terms of relative uncertainty, from $\sigma_{\text{SM}} = 0.01 N_{\text{SM}}$ up to $2.0 N_{\text{SM}}$.

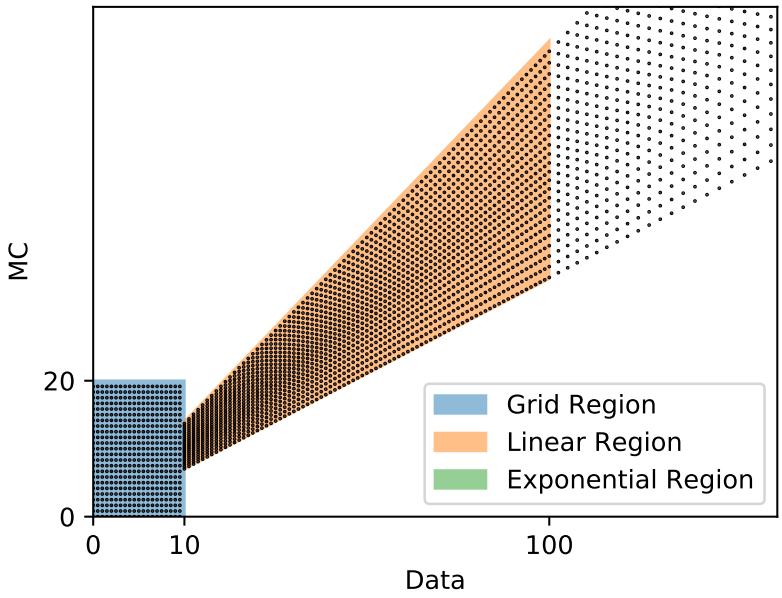


Figure 2.2: Illustration of the point-density for various areas in the LUT. Only the exact θ values for the values indicated by the black dots are stored, interpolation is done in-between. The "grid region" is defined with equidistant spacing of exact N_{obs} and N_{SM} points. The "linear region" is equidistant in N_{obs}, but two-sided exponential in N_{SM}. In the "exponential region", the spacing between exact N_{obs} points also grows exponentially.

Validation

The LUT can be easily validated by comparing θ values from the complete calculation with θ values obtained from the LUT algorithm. Because the input domain for both algorithms is three-dimensional, the validation is performed on one-dimensional slices of the input domain and discuss its generality afterwards. The parameter values have been chosen to ensure the relative deviation $\frac{|\theta_{\text{LUT}} - \theta|}{\theta}$ between the algorithms is less than 1 %.

First, the relative uncertainty is varied while keeping N_{obs} and N_{SM} fixed. The results are shown in figure 2.3. On the left hand side, N_{obs} is fixed to 0, while $N_{\text{SM}} = 1$. One can see that over the entire range of covered relative uncertainty (from 0 % to 130 %), the relative deviation between LUT and full calculation is below 0.1 %.

Similarly, figure 2.4 shows results from varying N_{SM} over a fixed value of N_{obs} and relative uncertainty. The dashed black line indicates the requirement of 1 %.

In both cases the relative differences show a repetitive pattern mostly below 0.1 %. Except for a non-differentiable feature at $N_{\mathrm{SM}} = N_{\mathrm{obs}}$ and extreme values of N_{SM} and low uncertainties, the maximum of the deviation lies below 1 %, fulfilling the requirement. In the case of the extreme values, the low θ value (< 0.005) causes the LUT value to be vetoed and the full calculation to take place. The repetitive pattern is also expected as it shows the high accuracy close to exactly calculated points and lower, but sufficient, accuracy in the linearly interpolated regions in between.

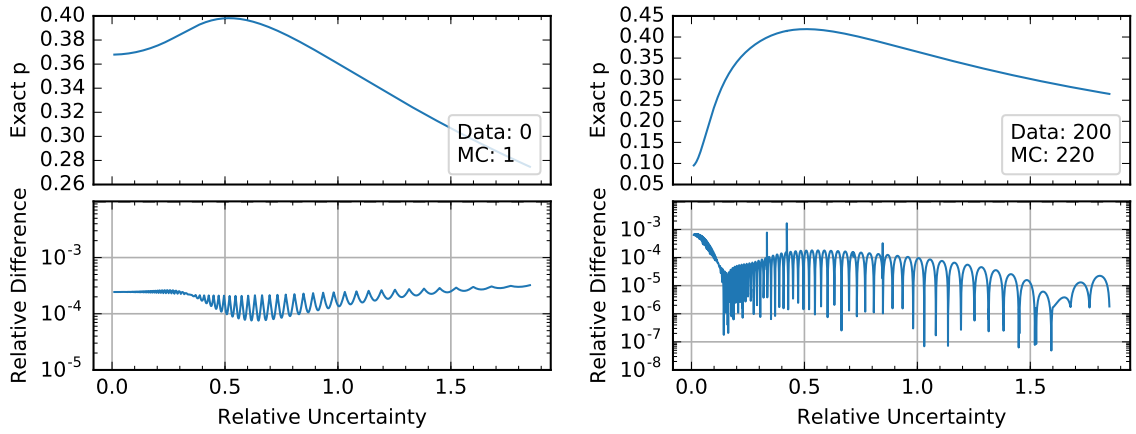


Figure 2.3: Relative difference between the complete calculation of θ and results from the LUT, varying the relative uncertainty while keeping N_{obs} and N_{SM} fixed. The relative difference is below 0.1 % over the entire covered range.

Discussion

Within the validated regions, the LUT has only shown small deviations from the results of a full computation, orders of magnitude below the previously set threshold of 1 %. Under these conditions it is therefore believed to provide a suitable alternative to the full computation of θ values.

Performance measurements (see appendix A.6) indicate that the time spent on calculating θ values, which was previously the dominating computational expense, has been reduced by a factor of ≈ 8. However, additional expenses arise, such as transferring the file to worker nodes and subsequently reading the file from the hard drive. The total gain will thus be less.

Overall, the LUT contributes great value to the automated search, as a reduction in computing time allows the analyst to increase the number of pseudo-experiments, enhancing the resolution of $\tilde{p}$-values. In future developments of the analysis, I would suggest further optimization of the LUT parameters, as they have only been roughly chosen to match the given requirements.

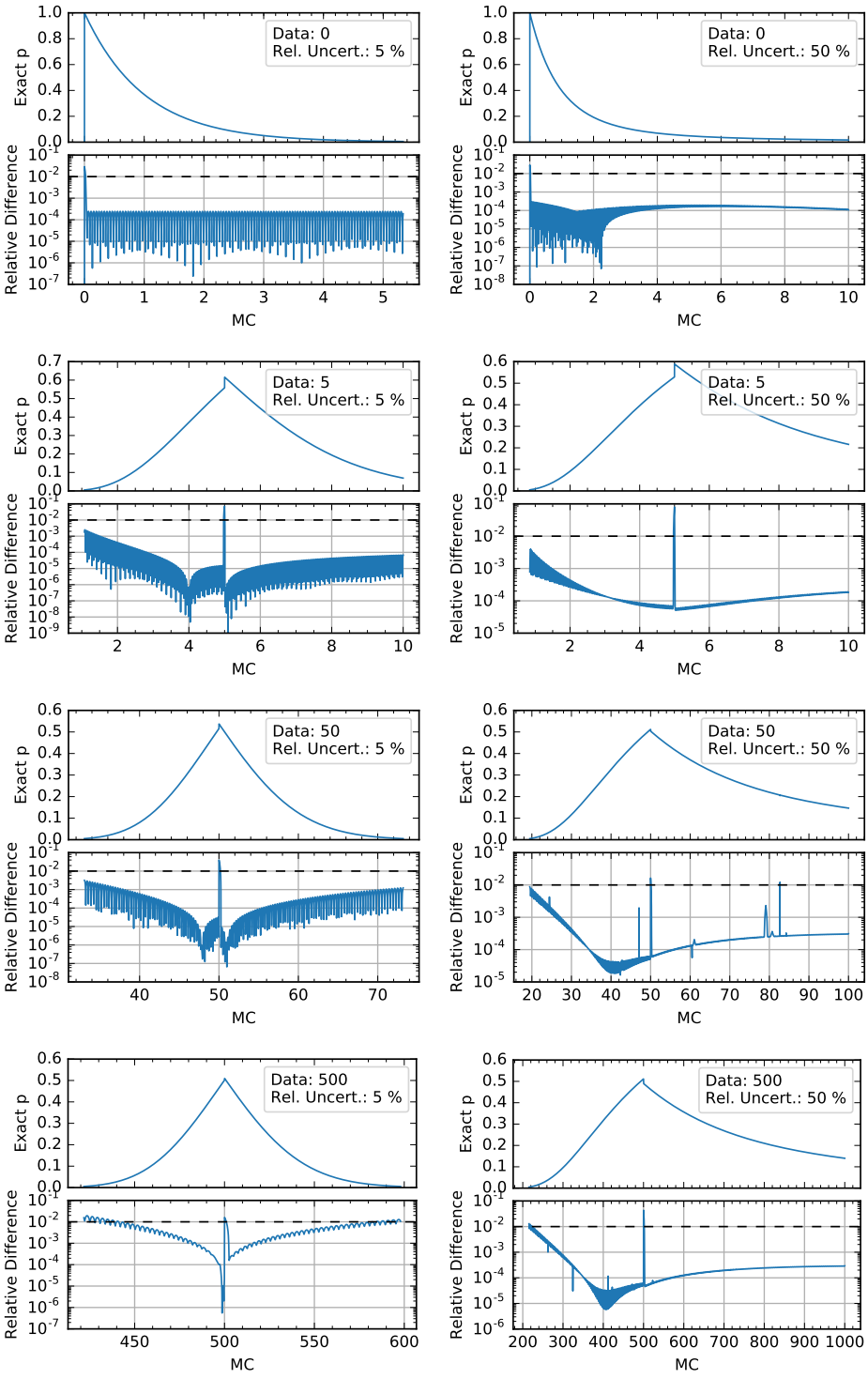


Figure 2.4: Relative differences between the complete calculation of θ and results from the LUT, varying N_{SM} while keeping N_{obs} and the relative uncertainty fixed. The relative difference between the algorithms is below 1 % (dashed line) in almost all regions.

Datasets

3

3.1 Monte-Carlo Simulations

The MUSiC analysis aims to search for deviations in a large number of final states and a wide kinematic phase space. Deviations are expected in any area under investigation and therefore there is no signal-free region that could be used for validation or fine tuning. Thus, MUSiC avoids data-driven methods as much as possible and rely on theoretical predictions in form of Monte Carlo simulations.

The Monte Carlo simulation is performed on an event-by-event basis. Each event is generated in multiple steps [105]: First, the sharing of proton momentum between partons is simulated. This is done sampling momenta from the empirically determined *parton density functions*. Afterwards, the hard scattering process is simulated. Similarly as with the momentum distribution, different scenarios are possible, each with a probability corresponding to the predicted scattering probability. In the next step, parton showering and hadronization is applied: Since the required higher order terms of QCD cannot be computed exactly, several parametrized models are used to simulate the formation of hadrons from quarks and gluons. In the last step, the detector response is simulated, also using parametrized interactions of the final state particles with detector components.

In practice, events of different physics processes are simulated in separate sets, so-called *samples*. Additionally, some samples only contain events within a certain p_{T} or mass range. This is done in order to provide a low statistical uncertainty in the tails of steeply falling distributions. In order to have consistent and validated samples for all CMS analyses, the MC samples are generated centrally by CMS on the WLCG [106].

In addition to weight assigned by generators, each event is assigned a further weight during the analysis. It is computed from the expected cross section, an expected higher order correction factor (k-factor) and the luminosity. Additionally, events are reweighted to compensate for difference pileup simulation between expectation and data. The expected event yield with certain properties can now be obtained by summing the weights of the desired events.

Within this thesis, various luminosity scenarios are analyzed. The MC events have not been regenerated for each scenario, instead the events are scaled to the described luminosity with the method mentioned above. This allows for a direct comparison ignoring hardware changes made to the detector between the data taking periods in 2015 and 2016.

3.1.1 Standard Model

The SM samples used by MUSiC are produced with the following generator applications: MADGRAPH 5 [107], SHERPA [108], POWHEG [109, 110], MADGRAPH_aMC@NLO [111], MCFM [112] and PYTHIA 8 [113]. For the generators MADGRAPH 5, POWHEG, MADGRAPH_aMC@NLO and MCFM, the subsequent hadronization is applied separately using PYTHIA 8. The detector response is simulated with Geant 4 [114].

The full list of used MC samples can be found in appendix A.1.

3.1.2 Signal Samples

The signal study uses the same set of signal samples as the corresponding dedicated analyses [38, 39, 45, 115]. This makes results comparable and also allows reusing central production and reconstruction by CMS. The mass points are chosen to cover a broad range including the resulting limits set by the dedicated analyses. However, since the automated search has to be re-run for each signal sample, computation time limits the amount of samples analyzed.

- For the QBH model, the Arkani-Hamed-Dimopoulos-Dvali (ADD) model with $n = 4$ extra dimensions is analyzed. The chosen mass points M are 1 TeV, 2 TeV, 3 TeV, 4 TeV and 5 TeV. The events are simulated by the QBH 2.0 [116] generator.

- In the black hole study, the case of a non-rotating black hole is considered. The fundamental Planck mass is set to $M_{\mathrm{Pl}(4+n)} = 4\,\mathrm{TeV}$ with $n = 6$ extra dimensions. This choice corresponds to the benchmark result published in [38]. The black hole mass is varied between 6 TeV, 7 TeV, 8 TeV, 9 TeV, 10 TeV. The MC generator used for the model is BLACKMAX [117].

- For the Seesaw-TypeIII samples, only the mass points $M_\Sigma = 380\,\mathrm{GeV}$ and $M_\Sigma = 500\,\mathrm{GeV}$ are tested. The samples are generated with MADGRAPH_aMC@NLO [111].

- The W′ model is tested for W′ masses of 3 TeV, 4 TeV, 5 TeV. The branching ratio of $\mathrm{W}' \to \mathrm{tb}$ is set to 1. This sample uses the CALCHEP [118] generator.

Again, where necessary, hadronization is applied using PYTHIA 8 and the detector response is simulated with Geant 4.

Statistical Evaluation

4

4.1 Coverage Analysis

This chapter investigates to what extend the test statistic θ introduced in section 2.10.3 can be directly interpreted as p-value, as intended: $p = \theta$. For that purpose, it will refer to the definition and requirements of a p-value and test whether θ fulfills those requirements. Additionally, the analysis will be repeated with θ_{LN}, using the log-normal prior (as suggested in section 2.10.3) instead of the Gaussian term in θ and compare the results.

A p-value is an indicator for the significance of a deviation, where a smaller p-value corresponds to a higher significance. The p-value of a result is compared to a significance threshold α, which is fixed before the statistical analysis. If the observed p-value is smaller than α, the null hypothesis is rejected [119]. In this case, rejection of the null hypothesis corresponds to falsification of the Standard Model and thus the discovery of new physics.

Both values are constructed in a way that the null hypothesis is (incorrectly) rejected by chance with a probability of α:

$$\Pr(H_0 \text{ rejected} | H_0) = \alpha \tag{4.1}$$

$$\Rightarrow \Pr(p < \alpha | H_0) = \alpha \tag{4.2}$$

This equation is tested during the *coverage analysis*.

In the case of θ, a deviation from this equality is expected. One reason for that is the truncation of the normal distribution. Usually, the probability of including the true value in a certain interval around the expected value is the same as including an expected value in an interval with the same width around the true value:

$$\Pr(N_{\text{SM}} - x \cdot \sigma \leq N_{\text{true}} < N_{\text{SM}} + x \cdot \sigma) \tag{4.3}$$

$$= \Pr(N_{\text{true}} - x \cdot \sigma \leq N_{\text{SM}} < N_{\text{true}} + x \cdot \sigma) \tag{4.4}$$

This property holds for all x if the probability density is a (non-truncated) Gaussian distribution. Truncation however breaks this property. The effect becomes more visible as the truncated probability increases and therefore should be most dominant in areas of very large uncertainty.

4.1.1 Procedure

To test equation (4.2), one can construct pseudo-experiments based on the null hypothesis H_0 and calculate the corresponding p-value, in this case θ or θ_{LN}. After generating and examining sufficiently many pseudo experiments, the rate of significant findings can be determined and compared to the claimed significance threshold:

$$p_{\text{true}} = \frac{\text{number of pseudo-experiments where } p < \alpha}{n_{\text{toys}}} = \Pr(p < \alpha | H_0) \tag{4.5}$$

The pseudo-experiments are based on MUSiC's null hypothesis: It is assumed that there is a constant probability that events end up in a particular region, and thus a constant true mean value N_{true}. Additionally, it is assumed that this true mean value is only known up to an expected event yield, N_{SM}. The way that N_{SM} is derived from N_{true} differs between θ and θ_{LN}: For θ, N_{SM} is drawn from a normal probability density, θ_{LN} will use a log-normal distribution instead.

Additionally, the physics process of performing a counting experiment has to be simulated: Here it is assumed that the event yield is caused by independent statistical processes with a fixed probability, thus the observed event yield follows a Poisson distribution around N_{true}.

The way that the assumed uncertainty enters the pseudo experiment also differs between θ and θ_{LN}: For θ, the absolute uncertainty is kept constant: $\sigma_{\text{SM}} = \sigma_{\text{true}}$. For θ_{LN}, the uncertainty is recalculated after drawing N_{SM} in order to keep the relative uncertainty constant: $\sigma_{\text{SM}} = \frac{N_{\text{SM}}}{N_{\text{true}}}\sigma_{\text{true}}$. This has been discussed by [68], a more rigorous discussion can be found in appendix A.7.

In the implementation, each pseudo-experiment begins with drawing N_{SM} from the probability density around N_{true}. At the same time, N_{obs} is drawn from a (discrete) Poisson distribution with a mean of N_{true}. From these two values, as well as the uncertainty σ_{SM}, θ (or θ_{LN}) is calculated.

This process is repeated for n_{toys} pseudo-experiments. An estimation for the true p value is afterwards calculated using equation (4.5), which yields the left hand side of equation (4.2).

In order to state the so called "coverage value" for the tuple (N_{true}, σ_{true}), both sides of equation (4.2) are translated to Z-scores (see appendix A.4), resulting in $Z_{\text{true}} = Z(p_{\text{true}})$ representing the observed and $Z_{\text{claim}} = Z(\alpha)$ representing the claimed rate ob significant results. The coverage is finally reported as

$$\Delta Z := Z_{\text{true}} - Z_{\text{claim}}. \tag{4.6}$$

4.1.2 Possible Outcomes

Three different result cases can arise, depending on the coverage value in equation (4.6):

1. $\Delta Z = 0 \Leftrightarrow \Pr(H_0 \text{ rejected} | H_0) = \alpha$: This is the ideal case. It indicates the p-value performs according to its definition.

2. $\Delta Z < 0 \Leftrightarrow \Pr(H_0 \text{ rejected} | H_0) > \alpha$: The background hypothesis is rejected more often than with a probability of α. This case is called "undercoverage" and corresponds to "liberal" behavior. The p-value overestimates the significance of deviations.

3. $\Delta Z > 0 \Leftrightarrow \Pr(H_0 \text{ rejected} | H_0) < \alpha$: The background hypothesis is not rejected in some cases where it should have been rejected. This effect is called "overcoverage" and corresponds to "conservative" behavior where the p-value underestimates the significance of deviations.

4.1.3 Results

Since the MUSiC p-value is required to cover a large range of possible N_{true} and relative uncertainty $\sigma_{\text{true}}/N_{\text{true}}$ values, the coverage is evaluated on a two dimensional grid in this parameter space. The computed results for θ can be found in figure 4.1, for θ_{LN} in figure 4.2.

Areas tinted with dark blue color were determined to be overcovered (conservative behavior), while the bright yellow shows areas of undercoverage. Within the blank areas, the coverage value could not be determined since no significant result could be observed. This is expected, especially for deficits below $N_{\text{true}} = 1$, because at $N_{\text{SM}} < 1$, the only possible deficit is $N_{\text{obs}} = 0$, which is not significant purely by the Poisson contribution in θ ($p \geq e^{-1} = 0.37$ with $\sigma_{\text{SM}} \to 0$).

Further illustrations in non-logarithmic view can be found in appendix A.9.

Results for θ

For the Gaussian prior in θ, one can see almost ideal coverage for $N_{\text{true}} > 1$ and $\sigma_{\text{SM}}/N_{\text{true}} < 50\,\%$. This is compatible with the results from earlier studies [68], which have been reproduced partially in appendix A.8.

As expected, the coverage equality breaks down at large uncertainties: In the deficit case undercoverage becomes visible, while the excess case shows overcoverage. For several regions, the θ-value is too conservative, leading to the effect that not a single

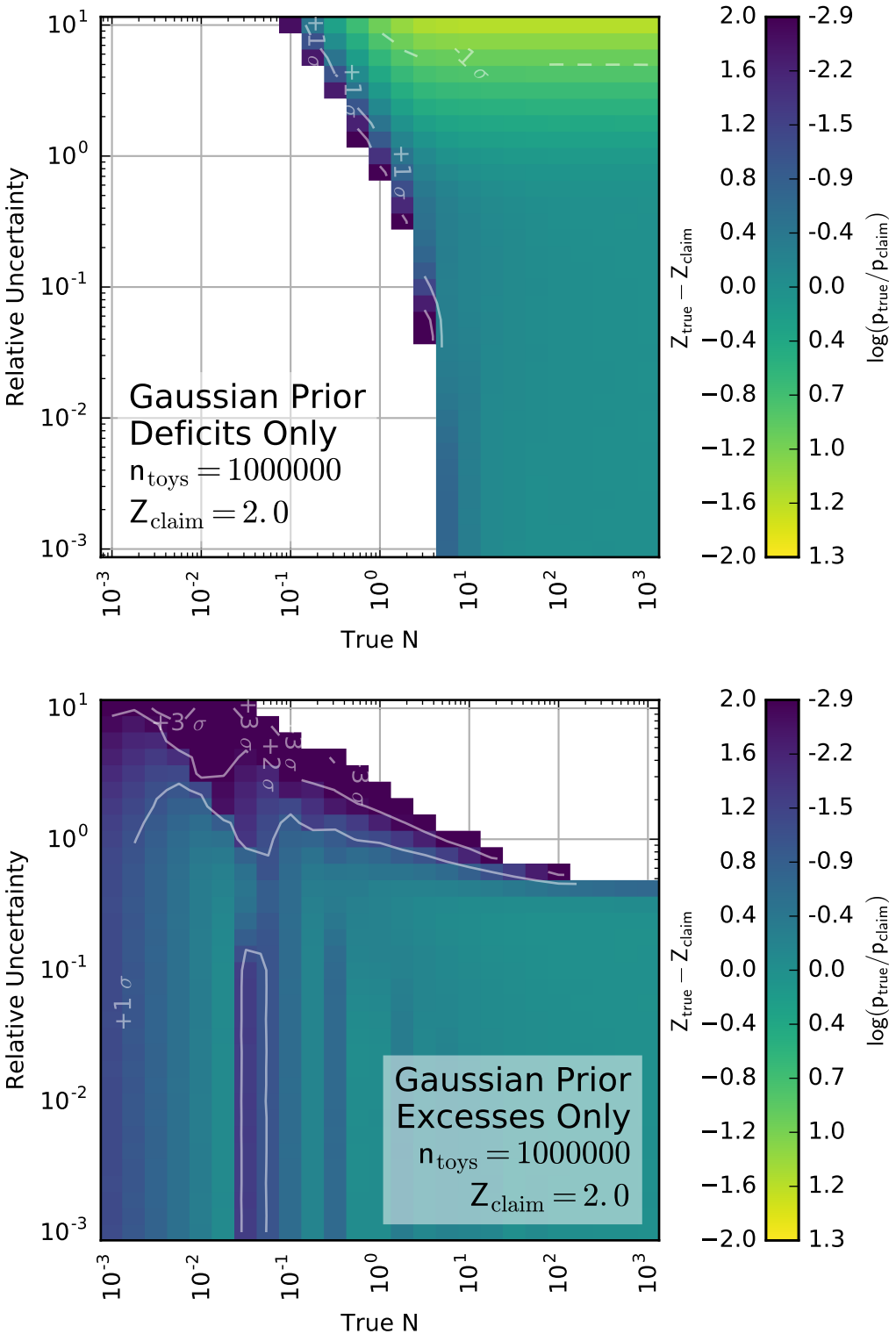


Figure 4.1: Results of the coverage analysis for θ, using the Gaussian prior. The figure on top shows that coverage for deficits, the bottom figure indicates coverage behavior for excesses. Within the blank areas, the coverage value could not be determined since no significant result could be observed.

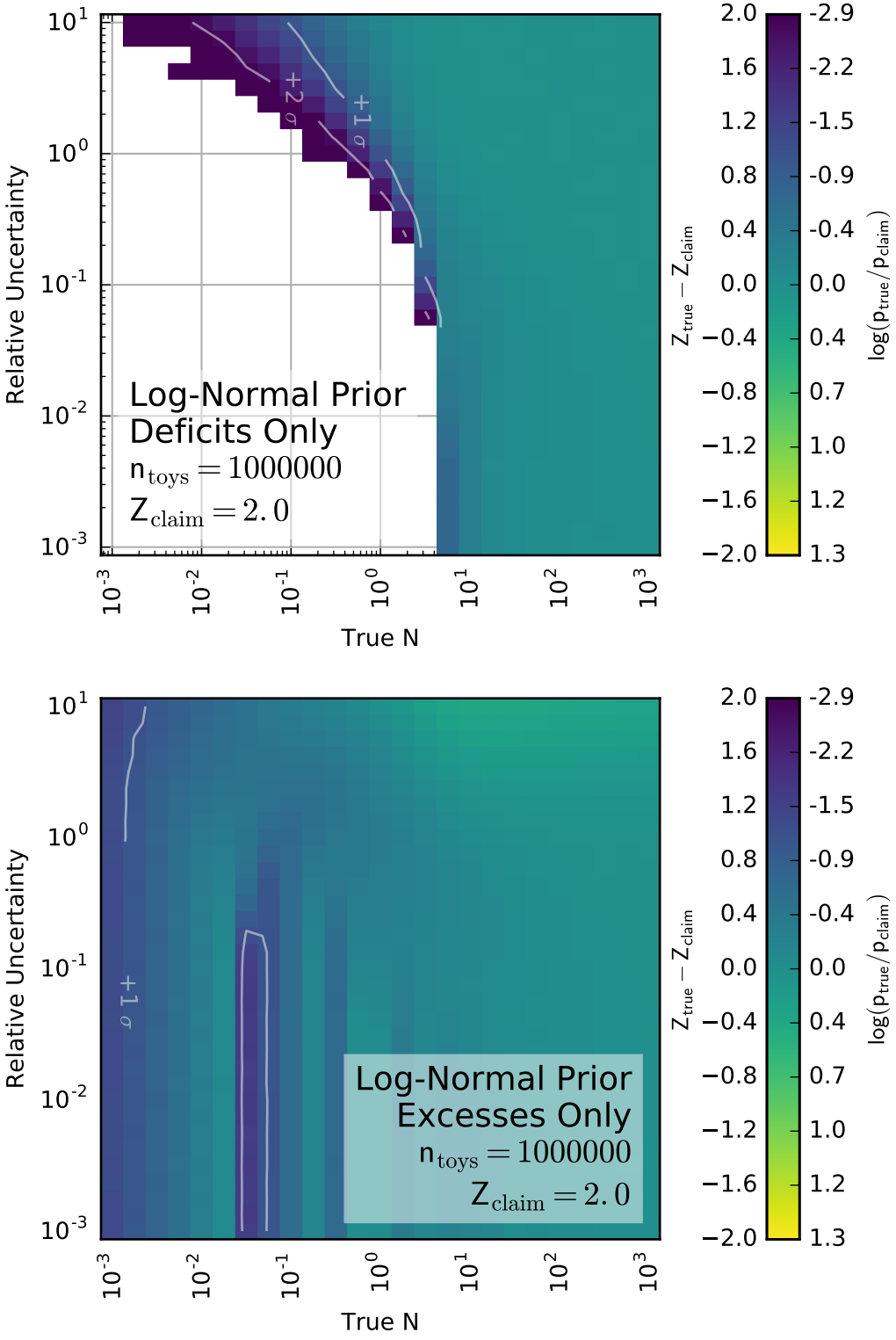


Figure 4.2: Results of the coverage analysis for θ_{LN}, using the log-normal prior. The figure on top shows that coverage for deficits, the bottom figure indicates coverage behavior for excesses.

significant round can be obtained during the coverage study. In these regions, the coverage value cannot be computed and therefore the area is displayed blank.

Results for θ_{LN}

The log-normal prior shows much better coverage values across the probed range. There are areas of slight overcoverage, but especially in the excess case the overcoverage does not exceed 1σ.

4.1.4 Discussion

A coverage analysis of the local test statistic θ with a Gaussian and a log-normal prior has been conducted. As expected, the Gaussian prior shows deficiencies especially at high values of relative uncertainty. One apparent solution would be to correct the θ-value with a parametrized correction factor to reestablish perfect coverage. In practice, it is not obvious how to parametrize the coverage difference in all three $(N_{\text{obs}}, N_{\text{SM}}, \sigma_{\text{SM}})$ dimensions. Therefore, this approach is not further pursued in this thesis.

In comparison, the log-normal prior performs well over the entire tested range, with a coverage of $|\Delta Z| \leq 1$. Therefore, if the goal of the local test statistic is to correspond to the definition of a p-value, the log-normal prior is to prefer over the Gaussian prior.

4.2 Log-Normal p-Value

As observed in the previous section, using the truncated Gaussian distribution as prior in the p-value has limited validity at high relative uncertainties. At the same time, it has been observed that using a log-normal prior improves test coverage in the problematic regions.

In addition, a log-normal prior is more suited for modeling uncertainties in event numbers: Analogously to the normal distribution, which is generated from a sum of independent random variables, the log-normal distribution results from the multiplication of independent random variables (see appendix A.3). The latter corresponds to the situation of event numbers, as most uncertainties on the event yield originate from shifting the event yield by a multiplicative factor. Also note that in the limit of $N_{\text{SM}} \to \infty$ and $\sigma_{\text{SM}} \to 0$, the log-normal distribution recovers the shape of a normal distribution, which explains similar coverage properties in that region.

Changing the prior in the test statistic also implies a new null-hypothesis, namely that all uncertainties are distributed in a log-normal fashion. Therefore, the pseudo-

experiment generation, which is based on the null-hypothesis, must also be adapted. Instead of choosing a new pseudo-mean value using a sum of uncertainties multiplied by a normally distributed bias (see equation (2.19)), one now has to use a product of relative uncertainties:

$$N_{\text{true}}{}' = N_{\text{SM}} \cdot \prod_i \left(1 + \frac{\sigma_i}{N_{\text{SM}}}\right)^{x_i} \tag{4.7}$$

However, this approach also comes with its disadvantages: When combining multiple bins into a region, usually, individual bin contents are added. Assuming normally distributed bin contents, this approach is valid because the sum of two normally distributed random variables is also normally distributed. Yet, this principle does not apply to log-normal distributions. The sum of log-normally distributed random variables is not log-normally distributed.

In the following, some solutions to this challenge are discussed, some of which have also been considered in previous works on the log-normal prior [68].

- Use Gaussian error propagation in any case: While this solution may approximately hold for two bins with the same number of events and relative uncertainty, it breaks down with more unequal numbers as the combined distribution differs significantly from the distribution obtained by summation of two (or more) random numbers drawn separately.

- Use an approximation: The challenge of approximating a sum of log-normally distributed random variables is of interest for various applications and has therefore been thoroughly researched. The oldest, most popular approach is the Fenton-Wilkinson approximation which is obtained by matching mean and variance of the sum to a log-normal distribution. Unfortunately, this approach breaks down when combining more than two bins [120]. Other possible alternatives are the Schwartz-Yeh approximation [121], which matches the first two moments of the logarithms with additional weight functions, and a recent procedure by Mehta et. al. [122] which uses a Gauss-Hermite approximation and poses the most accurate solution. The downside of the latter approaches is that they are computationally expensive, involving the calculation of integrals. Since a region can consist out of up to ~ 100 bins, the number of regions per event class can reach up to $\sim 10\,000$. Thus using a computationally cheap algorithm is absolutely necessary. Because of the high number of parameters (two for each bin) using a LUT is also not feasible.

- Combine bins to regions first, generate pseudo-experiments on the complete regions instead of bins: This approach completely ignores the direct correla-

tion between overlapping regions, and therefore does not represent the null-hypothesis.

- Generate pseudo-experiments from a normal distribution, evaluate the θ-value with the log-normal prior: This approach has been pursued in an earlier signal study [68]. As this signal study was performed before the start of data taking at CMS, uncertainties were underestimated. With larger uncertainties, this procedure is no longer feasible. The problem can be illustrated using an example from the turn-on region of the distribution. The uncertainty in these areas mostly originates from distribution shifts and therefore can be very large. Using a Gaussian prior, the scenario $N_{\mathrm{SM}} = 1\,000 \pm 2\,000, N_{\mathrm{obs}} = 0$ yields a θ-value of 0.00025, while a log-normal prior would assign 2.5×10^{-8}. This tension of 1.9σ causes very significant pseudo-experiments in these regions of large uncertainties and large event numbers.

Although the problem of combining bins is the only problem arising from a log-normal prior, it is so grave that the log-normal prior will not be used for the signal study in this thesis. Instead, an alternative solution to the overcoverage of the normal θ-value will be presented in the next section.

4.3 Region Veto Against Overcoverage

In section 2.10.2, several rules have been introduced to prevent inference for regions with incomplete simulation. In this section, add another rule will be added, which however does not intend to limit the search space because of invalid input for the test statistic, but to avoid shortcomings of the Gaussian test statistic itself.

In section 4.1, it was shown that the test statistic θ with the Gaussian prior has severe overcoverage for regions of high relative systematic uncertainty. This section now introduces a simple rule excluding these problematic regions from the search. In the bottom plot of figure 4.2, one can see that the contour line coverage $= +2\sigma$ follows a straight path in double logarithmic presentation. This line can be approximately parametrized with

$$\frac{\sigma_{\mathrm{SM}}}{N_{\mathrm{true}}} = 1.2 \cdot {N_{\mathrm{true}}}^{-0.2} \tag{4.8}$$

Regions with a larger systematic uncertainty (above the line) should be excluded from the search due to insufficient coverage of the test statistic.

As N_{SM} is the best estimator for N_{true}, N_{true} is replaced with N_{SM}. Additionally, in order not to unnecessarily veto regions with large N_{true}, and not to diverge at $N_{\mathrm{SM}} \to 0$, this value is restricted to the range between 0.5 and 5.0:

$$\frac{\sigma_{\mathrm{SM}}}{N_{\mathrm{SM}}} = \max(0.5, \min(1.2 \cdot {N_{\mathrm{SM}}}^{-0.2}, 5.0)) \tag{4.9}$$

The excluded area is illustrated in figure 4.3.

4.4 A Global p-Value

The $\tilde{p}$ distribution provides the ability to notice an increase of not very significant deviations in many classes by inspecting the central part of the distribution. However, this manual method is only qualitative. I will now introduce a possible extension to the analysis which provides a quantitative method of combining all $\tilde{p}$-values of a search in one kinematic variable for all classes.

This allows the analyst not only to draw conclusions about the presence of new physics in the observed data set, but also to quantify MUSiC's sensitivity for simulations of known benchmark models. Furthermore, an automated analysis can be used as a regression test to decide whether future modifications to the analysis improve its sensitivity.

As input, the algorithm takes the look-elsewhere-corrected $\tilde{p}$-values of all classes. In the case of MC simulation, either for SM or signal studies, such a set of $\tilde{p}$-values can be provided for each pseudo experiment round, for observed data one only obtains one list (one value for each class).

The algorithm for the SM distribution performs as follows: For each pseudo experiment round, the distribution of $\tilde{p}$-values (of all classes) is compared to a reference distribution using a statistical test. The reference distribution is the distribution of all $\tilde{p}$-values of SM-only pseudo-experiments, of all n event classes and k pseudo-experiment rounds, as presented in figure 4.4. Note that the distribution features a spike at $\tilde{p} \approx 1$, which originates from the remaining discretization of $\tilde{p}$ discussed in section 2.10.5.

The statistical test outputs a scalar result t for each of the k rounds, where a larger value indicates a larger deviation from the reference distribution. The set of k test statistic values is stored. Given a single observed $\tilde{p}$ distribution, one can analogously compute the test statistic t_{obs} and a p-value expressing to the significance of the observation:

$$\hat{p} = \Pr(t > t_{\text{obs}} | H_0), \tag{4.10}$$

where H_0 indicates the null hypothesis given by the Standard Model. For a decision whether or not the null hypothesis should be rejected, $\hat{p}$ has to be compared to a significance threshold α. α is equivalent to the type-I error rate, the probability of incorrectly rejecting the null hypothesis if it were true [119]:

$$\Pr(\hat{p} < \alpha | H_0) = \alpha \tag{4.11}$$

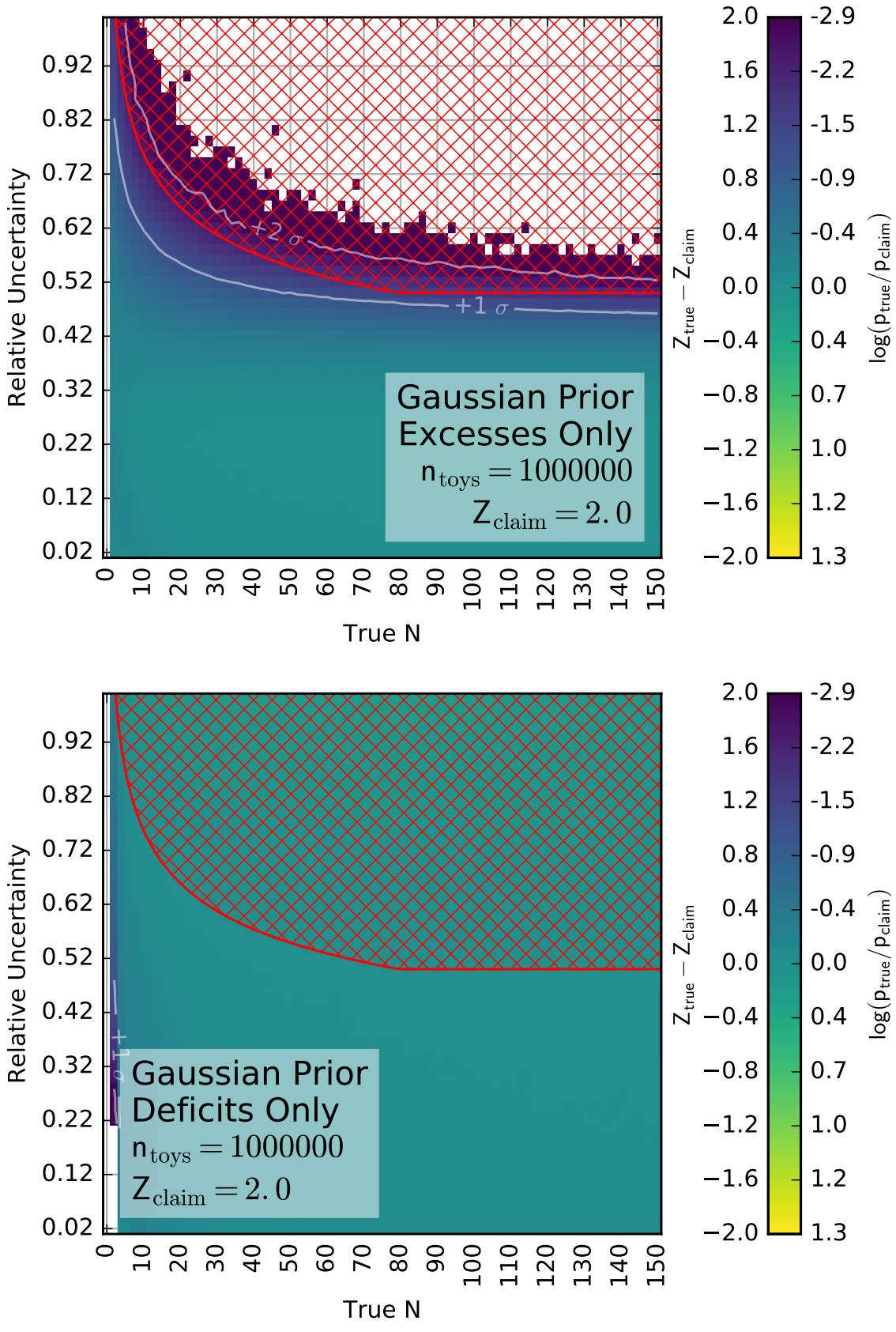


Figure 4.3: Illustration of the regions vetoed due to insufficient coverage (areas with red hatching pattern).

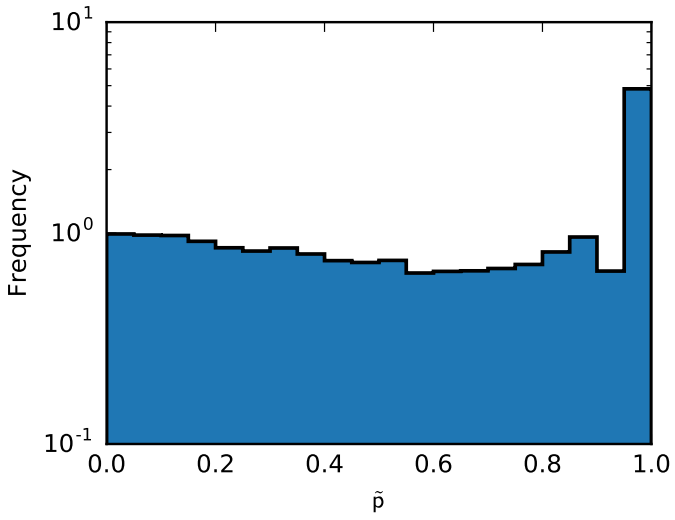


Figure 4.4: Reference distribution $g(\tilde{p})$, serving as input to the analysis. In this instance, the Σp_{T} distribution has been analyzed, shown are all $\tilde{p}$-values of exclusive event classes.

Equivalently, a critical value t_{crit} of the test statistic can be defined, such that

$$\Pr(t > t_{\mathrm{crit}}|H_0) = \alpha \tag{4.12}$$

Similarly, the type-II error rate can be identified, which denotes the probability of incorrectly accepting H_0 if it is false [119]:

$$\Pr(\hat{p} > \alpha|H_1) = \Pr(t < t_{\mathrm{crit}}|H_1) = \beta \tag{4.13}$$

The quantity $1 - \beta$ is called *test power* [119] and in here indicates the probability of correctly rejecting the SM if the dataset under investigation contains new physics. The test power of a test therefore directly describes the sensitivity towards a particular new physics scenario.

The test power can be calculated from simulations of new physics, such as the combined signal and background pseudo-experiments as described in section 2.11. The test power does not only depend on the chosen test statistic and signal model, but also on α. A larger value of α is less restrictive and will correctly accept the alternative hypothesis more often, at the cost of incorrectly rejecting the SM.

The choice of α is subjective [123]. Rather than following a strict convention, it is important to state which significance threshold has been used and, if the test statistic indicates that the null hypothesis should be discarded, to act accordingly. In the case of the MUSiC analysis, finding a deviation from the SM would certainly initiate a dedicated analysis at CMS. Therefore, in this case, an aggressively high value

of $\alpha = 5\,\%$[1] can be used, accepting the fact that in 5 % of the cases MUSiC would incorrectly reject the SM and initiate a dedicated analysis.

So far, the choice of the test statistic t has not been specified. In the following, multiple options are discussed. The distribution of $\tilde{p}$-values under investigation is denoted as $f(\tilde{p})$, having $N_f = k$ values, the reference distribution is $g(\tilde{p})$, with N_g values. The cumulative distribution functions are $F(\tilde{p})$ and $G(\tilde{p})$ respectively.

- χ^2-Test [124, 125]: This test is used to compare two sets of values, binned into distributions. For each bin i, the difference between bin contents g_i' for the reference and the sample bin content f_i' is evaluated.

$$t = \frac{\chi^2}{\text{ndof}} = \frac{1}{\text{number of bins}} \sum_i^{\text{bins}} \frac{\left(\sqrt{N_g/N_f} \cdot f_i' - \sqrt{N_f/N_g} \cdot g_i'\right)^2}{N_f + N_g} \tag{4.14}$$

  A disadvantage of this method is that binning discards information. Additionally, the method works best for large number of event classes.

- Kolmogorov-Smirnov [125,126]: This test does not require binning. It compares the empirical cumulative distribution functions of reference and comparison, finding the maximal distance between F and G.

$$t = \left(\sqrt{N_e} + 0.12 + 0.11/\sqrt{N_e}\right) \cdot \sup_i |F(\tilde{p}_i) - G(\tilde{p}_i)| \tag{4.15}$$

  where

$$N_e = \frac{N_f \cdot N_g}{N_f + N_g} \tag{4.16}$$

  is an effective number of values.

  A disadvantage of this test statistic is that it is most sensitive to deviations in the center of the distribution. Additionally, only the largest difference is used for the result, instead of accumulating the difference.

- Anderson-Darling [125–127]: This test belongs to the group of quadratic EDF (empirical distribution function) statistics. This group consists of test statistics that integrate over the quadratic difference between the reference and sample cumulative functions:

$$t = k \int_{-\infty}^{\infty} (F(x) - G(x))^2 \, w(x) \, \mathrm{d}F(x) \tag{4.17}$$

  Here, the weighting term $w(x)$ can be used to focus on specific parts of the distribution. The simplest possible weight is $w(x) = 1$, which is used by the

[1] For comparison, a discovery in particle physics is usually claimed at $\alpha \sim 10^{-7}$ [119].

Cramér-von-Mises test statistic. The Anderson-Darling metric uses a more sophisticated weight $w(x) = [G(x)(1 - G(x))]^{-1}$ instead, giving more weight to both tails of the distribution.

For this thesis, the discrete two-sample version is used, which is described in [128].

- Simple: t is the ratio of $\tilde{p}$-values below a critical value, in this case 0.01.

A thorough comparison of the afore-mentioned test statistics can be found in [127].

Alternatively, the reference distribution can be replaced by a uniform distribution. This also simplifies the formulas of most test statistics. Additionally, this variant has the advantage of enabling the analyst to identify whether the SM-only distribution agrees with the assumption of being uniformly distributed. However, as will be shown in chapter 5, the comparison to a uniform distribution can only be done with a higher minimum yield threshold of e.g. $N_{\mathrm{SM}} \geq 1.0$, which reduces the number of event classes with $\tilde{p} \approx 1$ and therefore brings each distribution closer to a uniform distribution.

The MUSiC analysis is not the only model-independent search for the LHC striving at a global measure for sensitivity: A corresponding analysis performed by the ATLAS collaboration [82] employs a statistical interpretation similar to the newly introduced "simple test". The approach is also to count the number of pseudo-experiments which yield a certain fraction of classes beyond a threshold p. However, instead of determining the fraction of classes t_{crit} from SM-only simulations, ATLAS uses $t_{\mathrm{crit}} \sim$ one class (alternatively also two or three classes). The analysis then uses the p-threshold as a variable. For any given observed dataset, the p-threshold is determined at which one (or two or three) event class(es) pass.

Discovery Potential

5

The study of the discovery potential, also called *sensitivity study*, has multiple purposes: On one hand, its goal is to assess the absolute sensitivity of the analysis towards certain benchmark models for new physics. On the other hand, it can also be used to evaluate the sensitivity for a single model using different analysis variants. This enables the analyst to test which features, such as algorithms or parameters, have the largest impact on the discovery potential.

The chapter will start with the latter goal: After defining a set of features and key questions to be answered in this chapter, the corresponding results will be presented in form of tables and graphics, and the implications for the MUSiC analysis will be discussed. In the second part, the sensitivity towards the benchmark models introduced earlier will be explicitly assessed and compared to dedicated analyses performed on a similar dataset by CMS.

5.1 Evaluation of Features

The goal of the first part of the chapter is to evaluate various features of the analysis, including some that have been newly introduced with this thesis. The evaluation will be guided by the following set of key questions that arise in this context:

- Validation: Using only the SM as new physics input, do the results agree with the SM-only search?
- How does the $\tilde{p}$-distribution of the SM-only search scale with luminosity?
- How does the minimum-yield-per-event-class threshold (section 2.10.5) affect the SM-only distribution of $\tilde{p}$-values?
- How do the region vetoes (including the overcoverage veto introduced in section 4.3) reduce the available search space?
- Which test statistic t shows the best sensitivity for the global p-value?
- Can a uniform distribution be used as test statistic reference?
- Is MUSiC sensitive to new physics that is only visible in one or few final states?
- Is MUSiC sensitive to new physics with a small impact on several final states?
- How does an increase in luminosity affect the sensitivity? For which kind of models does the analysis benefit from a higher luminosity?

5.1.1 Interpretation of the Result Displays

The result figures presented in the following sections contain a large amount of information that might be difficult to interpret for readers not accustomed to MUSiC. Therefore, this section aims to quickly introduce the material used to present the results and explain how they can be interpreted. An overview of the information presented is schematically shown in figure 5.1.

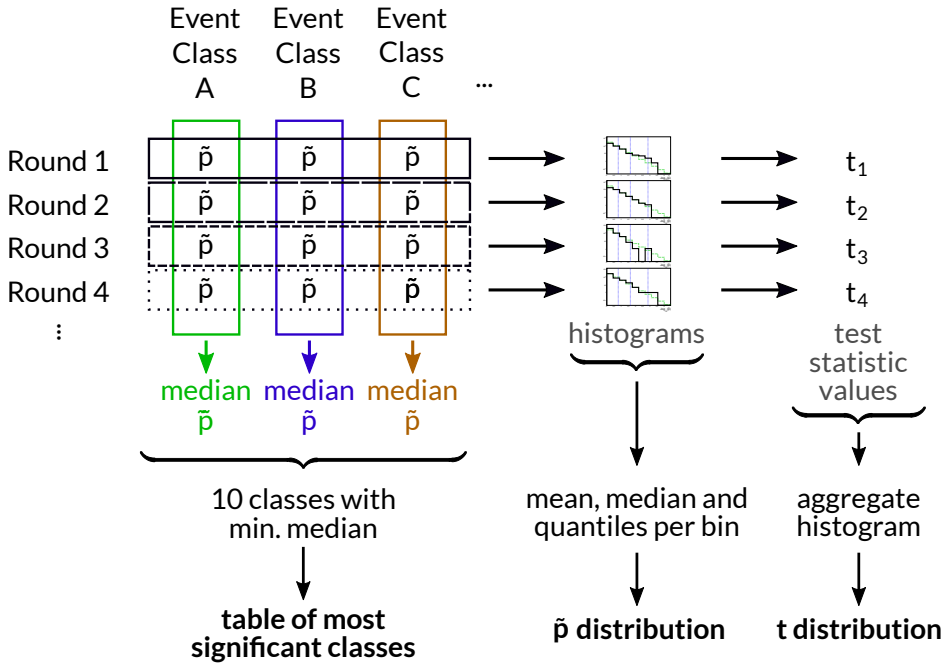


Figure 5.1: Flow of information from the sets of $\tilde{p}$-values to the results presented in this chapter. The schematic shows the vertical aggregation of $\tilde{p}$-values from $k = 4$ pseudo-experiments into median $\tilde{p}$-values as well as the horizontal aggregation of distributions of $\tilde{p}$-values for each round. Furthermore, the test statistic t is also computed horizontally and aggregated into a distribution.

$\tilde{p}$-Distribution

This illustration has conceptually been introduced in section 2.10.6. One example can be seen in figure 5.2. It is produced by aggregating $-\log_{10}\tilde{p}$ values: First, for each of the k rounds, the n results from n classes are sorted into a histogram. This results in k histograms, therefore k entries per bin. In each bin, the median bin content, the mean bin content and 68 %- and 95 % quantiles around the median are calculated. The highest bin serves as overflow bin, as one cannot obtain a precise value for $\tilde{p} < 1 \times 10^{-4}$ from 10 000 SM-only pseudo-experiments.

The aggregation procedure is done for the (corrected) $\tilde{p}$-values originating from searches for deviations between the SM and pseudo-experiments based on the SM and its result is depicted in the turquoise and blue bands, as well as the turquoise line and the black dotted line. Additionally, the illustrations show a dashed green

line which represents a uniform distribution. Because of the logarithmic binning, the number of uniform events is not equal for each bin. The uniform distribution can be used to tell whether $\tilde{p}$ fulfills the requirement of a p-value of being uniformly distributed.

Results of the pseudo-experiments originating from a signal study (see section 2.11) are depicted in the same figure as red data points. The vertical error bar around these points indicate the 68 %-quantiles around the median. This thesis uses 100 signal pseudo-experiments for each event class, thus 100 bin contents have been used to calculate the magnitude of the error bar.

Table of Most Significant Classes

The second way of presenting results is the table of median most significant classes. It is aggregated by regarding the corrected $\tilde{p}$-values of each event class for all rounds. For each event class, the median $\tilde{p}$-value is computed. Subsequently, the event classes with the 10 smallest median values are presented, alongside with the Z-score expressing the deviation in terms of standard deviations of a normal distribution (for the conversion formula see appendix A.4).

Distributions of t

The third way of presenting the results is to aggregate distributions of t-values. Several examples are shown in figure 5.3. There are always multiple distributions depicted in one figure: The histogram filled in gray shows the distribution of t-values from SM-pseudo-experiments. In addition there are one or more distributions originating from signal studies, drawn as colored lines.

A vertical dashed line indicates the critical value t_{crit}. It is defined as the value of t which separates $\alpha = 5\,\%$ of the area under the gray SM-only distribution on its right side. The test power can also be read from the result figure as the area to the right of t_{crit} under each signal line. Its numerical value is indicated next to each entry in the legend below the figure.

Choice of Distributions

The statistical inference, i.e. the aggregation of distributions of $\tilde{p}$-values as distribution, table or test statistic, is performed separately on each type of event class (exclusive, jet-inclusive and inclusive, see section 2.7) and on each kinematic variable (M_{inv}, Σp_{T}, $E_{\text{T}}^{\text{miss}}$) separately. Therefore, the result set for each signal study consists of nine different channels.

All results presented in this chapter have been obtained in regard to the Σp_{T} kinematic variable and the inclusive event classes. The former choice was ambiguous, in

most signal models, the results of all three kinematic variables have shown similar sensitivity.

The latter choice is motivated by sensitivity: For the majority of investigated models, the highest sensitivity could be obtained from inclusive event classes. Events in inclusive event classes can contain any number of objects in addition to the objects explicitly stated in the event class name. This in turn also means that events usually are sorted into multiple inclusive event classes, introducing correlations between the event classes. Due to the way that MUSiC treats correlations in SM-pseudo-experiments, this is not expected to pose a problem.

Note that in any case, the calculation of the kinematic quantity (here Σp_{T}) only considers the objects that are explicitly stated in the event class name.

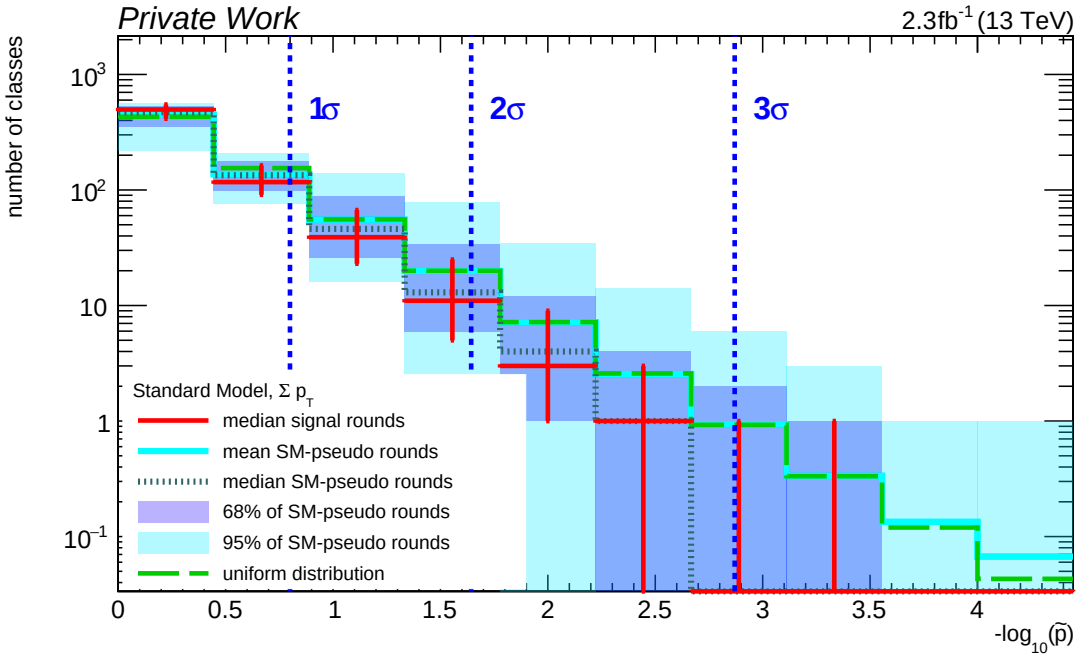
5.1.2 Validation Using the SM

The automated search takes two sets of event classes as input when performing a sensitivity study: A set of event classes from classified events of the Standard Model, and a set of event classes where simulated new physics events have been added to the Standard Model expectation. In order to validate the analysis, the latter input set can be replaced by another set of SM-only event classes.

In this case, one expects not to find any significant event classes. Furthermore, there should be no significant deviation in the distribution of $\tilde{p}$-values between the median signal rounds and the mean SM-pseudo rounds. The test power of the global test should be $1 - \beta \approx \alpha = 5\,\%$.

The results of such a validation run are depicted in figure 5.2. As expected, the distribution of $\tilde{p}$-values shows no significant deviation between the median bin contents for the signal study, indicated by the red data points and the median bin content of the SM-only pseudo-experiments, drawn as black dotted line. The table on the bottom shows the 10 median most significant inclusive event classes. The median most significant event class deviates by 0.2σ, thus being in complete agreement with the expectation.

Figure 5.3 shows results of the newly introduced global p-value. As expected, the gray SM-only distribution and the blue validation distribution overlap for all test statistics. Furthermore, the expected test power is $1 - \beta \approx \alpha = 5\,\%$, which confirms that the automated search would find a significant deviation in the validation data set as often as in the SM.



Event Class	Median $\tilde{p}$	Z
1 e + 1 μ + 1 jet + $E_{\mathrm{T}}^{\mathrm{miss}}$ + 3 b-jets + X	4.23×10^{-1}	0.2
1 μ + 2 γ + 1 b-jet + X	4.28×10^{-1}	0.2
2 e + 1 μ + 1 b-jet + X	4.29×10^{-1}	0.2
2 μ + 1 jet + $E_{\mathrm{T}}^{\mathrm{miss}}$ + 3 b-jets + X	4.32×10^{-1}	0.2
1 e + 4 b-jets + X	4.36×10^{-1}	0.2
2 γ + 2 jets + $E_{\mathrm{T}}^{\mathrm{miss}}$ + X	4.38×10^{-1}	0.2
3 jets + 1 b-jet + X	4.42×10^{-1}	0.1
2 μ + 2 jets + $E_{\mathrm{T}}^{\mathrm{miss}}$ + X	4.42×10^{-1}	0.1
3 μ + X	4.44×10^{-1}	0.1
1 e + 1 μ + 6 jets + 1 b-jet + X	4.46×10^{-1}	0.1

Figure 5.2: Distribution of $\tilde{p}$-values and table of most significant inclusive event classes for the SM-only validation. For a detailed explanation on the displayed information, see section 5.1.1. As expected, the distribution does not show any significant deviation between the validation (red data points) and the SM-only pseudo-experiments (dotted line). The table below shows the inclusive event classes with the smallest median of $\tilde{p}$ between the signal simulation rounds. The most significant class shows a deviation of 0.2σ, which is not significant.

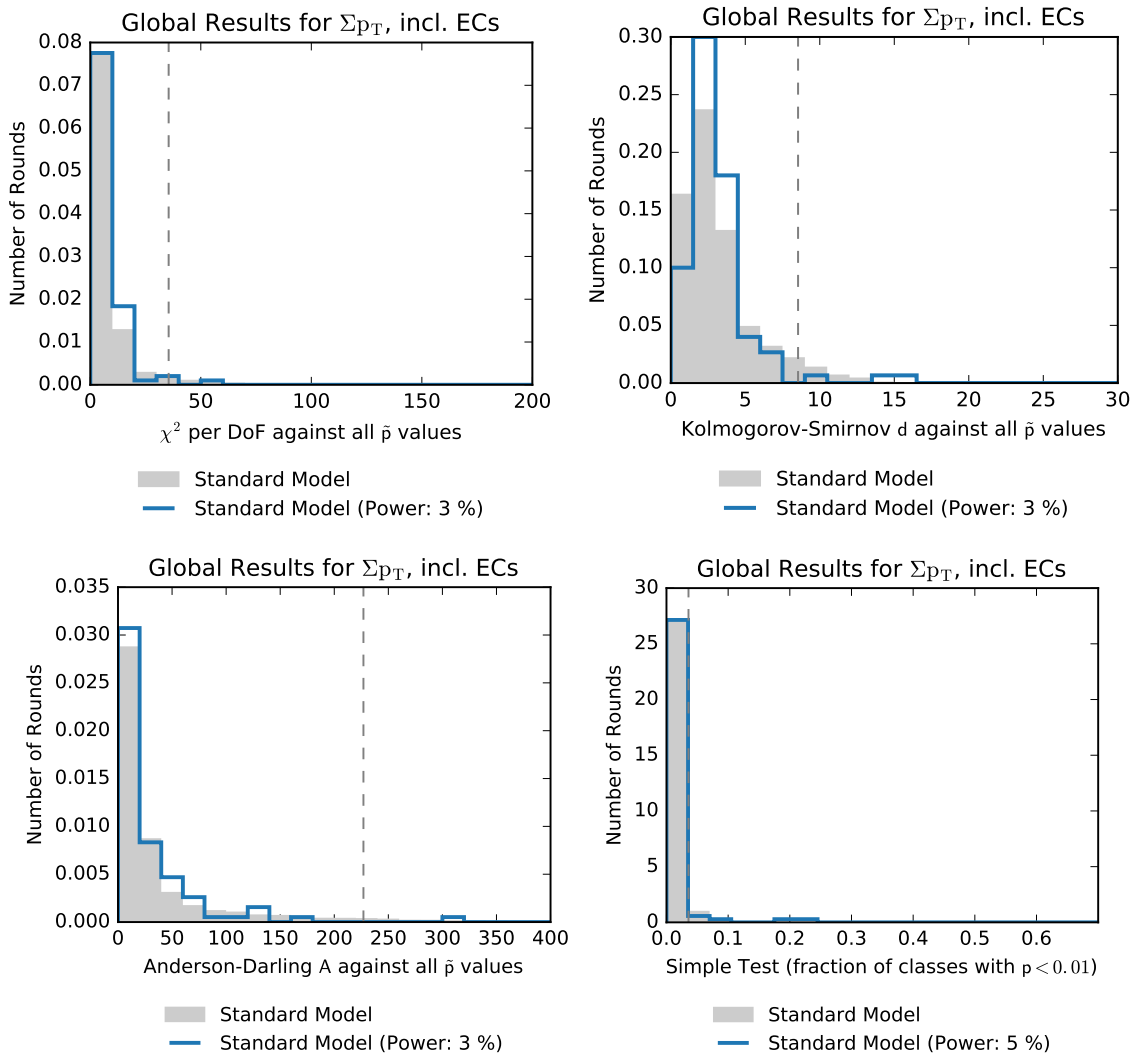


Figure 5.3: Distribution of t-values of the validation. As expected, the blue line, corresponding to results of the signal study, is in complete agreement with the gray histogram, as the same SM-only event classes have been specified as signal input. Consequently, the test power, which is defined as area under the blue curve to the right of the dashed vertical line at t_{crit}, is approximately 5 %. Note that the blue line has been generated from only 100 entries, therefore the test power is subject to statistical fluctuations.

5.1.3 Differences between 2.3 fb^{-1} and 35.9 fb^{-1} of Events

To assess the differences between the luminosities of the data taking periods in 2015 (with a luminosity of 2.3 fb^{-1}) and 2016 (35.9 fb^{-1}), one can compare the results from the $\tilde{p}$-distribution in figure 5.2 to similarly obtained results in figure 5.4. For this distribution, all event yields have been scaled to the new luminosity and the automated search was repeated. Because of the increase in the scaled event yield, more event classes will pass the threshold of $N_{\text{SM}} \geq 0.1$, therefore more event classes will be searched for deviations in the 2016 dataset. This also shows e.g. in the number of exclusive event classes, which rises from 460 to 676 (table 5.2).

The validation procedure presented in the previous chapter was also applied on the dataset with the luminosity of 35.9 fb^{-1}, as depicted in red in figure 5.4. Again, no deviation between the validation dataset and the SM-only pseudo-experiments is visible.

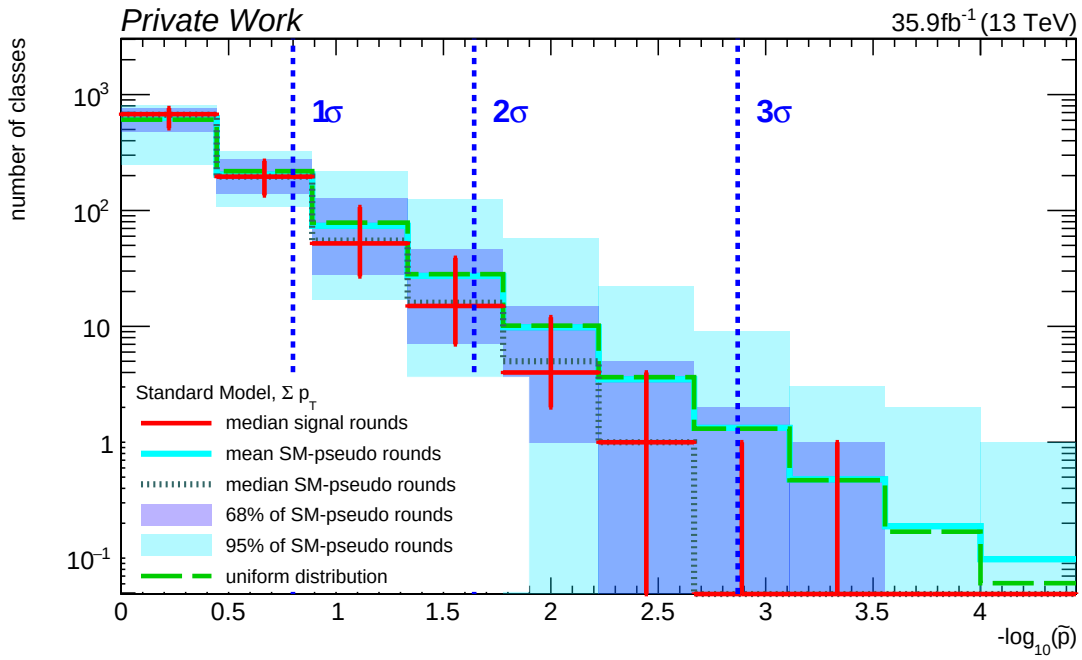


Figure 5.4: Distribution of $\tilde{p}$-values for the SM-only validation using the 35.9 fb^{-1}. This illustration should be compared with the distribution in figure 5.2 to illustrate the increase in inclusive event classes.

5.1.4 Effect of the Minimum Yield Threshold

The minimum yield threshold was introduced in section 2.10.5 in order to suppress the creation of almost-empty event classes, as the discretization of the θ- and $\tilde{p}$-values would invalidate the interpretation of the $\tilde{p}$-value as a probability. This effect has been analyzed by performing automated searches and aggregating distributions of $\tilde{p}$-values with several different values of the minimum yield threshold. The results are presented in figure 5.5 and table 5.2.

The first $\tilde{p}$-distribution has been generated with $N_{\mathrm{SM}} \geq 0$, only filtering out event classes if the region veto (section 2.10.2) applies to the entire distribution. The SM-only distribution clearly shows deviations from the uniform distribution, especially in the bins containing less significant event classes. This is not the case in the second figure, showing a distribution generated with a threshold of $N_{\mathrm{SM}} \geq 0.1$.

Alternatively, one can regard the reduction in the number of event classes between several threshold values, as shown in table 5.2. Overall, the minimum yield threshold reduces the number of event classes by more than a factor of two for the 2015 dataset. Because the same MC simulated events were used for the 2015 and 2016 scenarios, the number of initial event classes remains the same between the studies. The increase in luminosity by a factor of 16 however causes more event classes to pass the threshold.

For this thesis, a threshold of $N_{\mathrm{SM}} \geq 0.1$ was chosen, as it brings the $\tilde{p}$-distribution of the SM-only pseudo-experiments and the uniform distribution into agreement. A larger value would be impractical as the analyst manually has to review classes that do not pass the threshold but contain observed data.

5.1.5 Impact of Vetoes

In section 2.10.2, several rules have been introduced that aim to exclude regions from the automated search where the SM simulation is incomplete and therefore no inference can be made. The set of rules was expanded in section 4.3 in order to avoid statistical inference on regions where the test statistic is known to be incorrect because of overcoverage.

This section assesses the impact of the region vetoes on the number of regions that are searched for deviations. The numbers in table 5.3 have been obtained by recording the reason for each veto during an automated search of the Σp_{T} distribution of SM-only pseudo-experiments of all event classes. Because the vetoes are evaluated in a given order, aborting after the first matching rule, the numbers on the latter vetoes are only an approximation and a lower bound. The total number of vetoes, however, can be accurately determined using this feature.

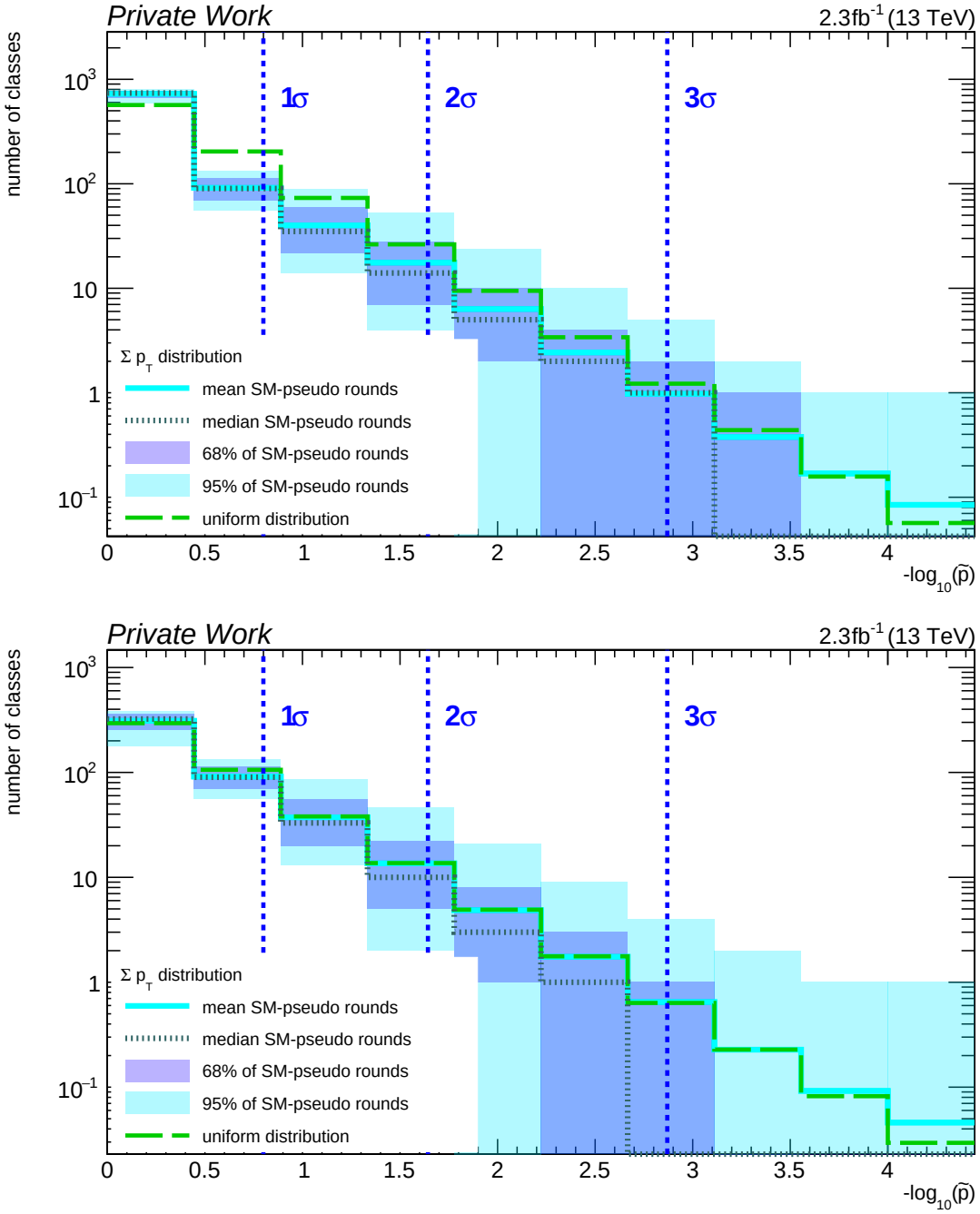


Figure 5.5: Distribution of $\tilde{p}$-values of inclusive event-classes for different minimum yield thresholds at 2.3 fb^{-1}. The distribution above was created with a threshold of 0, the second one with a threshold of 0.1. Although the standard model and its validation agree in the upper illustration, neither distribution is in accordance with a uniform distribution. Therefore, the validity of $\tilde{p}$ without such a threshold is questionable.

Event classes created with a luminosity of $2.3\,\text{fb}^{-1}$:

	without req.	vetoes only	$N_{\text{SM}} \geq 0$	$N_{\text{SM}} \geq 0.01$	$N_{\text{SM}} \geq 0.1$	$N_{\text{SM}} \geq 1$
exclusive	1 249	902	885	647	460	239
jet-inclusive	1 359	988	968	718	508	343
inclusive	1 686	1 236	1 212	915	672	465

Event classes created with a luminosity of $35.9\,\text{fb}^{-1}$:

	without req.	vetoes only	$N_{\text{SM}} \geq 0$	$N_{\text{SM}} \geq 0.01$	$N_{\text{SM}} \geq 0.1$	$N_{\text{SM}} \geq 1$
exclusive	1 249	902	885	773	676	492
jet-inclusive	1 359	988	968	863	747	549
inclusive	1 686	1 236	1 212	1 085	951	710

Table 5.2: Number of event classes created with several veto combinations at a luminosity of $2.3\,\text{fb}^{-1}$ (upper table) and $35.9\,\text{fb}^{-1}$ (lower table). As mentioned earlier in section 2.10.2, in some cases, entire event classes may be skipped if the largest region is vetoed. This case is illustrated in the second column. The other columns additionally require the event class to have a minimum total yield, as described in section 2.10.5. As this threshold is increased, less event classes pass.

veto reason	vetoed regions
empty bin added	38.8 %
overcoverage threshold (section 4.3)	4.9 %
negative total yield	1.0 %
large negative contribution of any process	2.5 %
negative/low leading contribution	5.3 %
large statistical uncertainty	6.2 %
total vetoed	58.7 %

Table 5.3: Impact of region vetoes. The vetoes are applied in the order listed here. The first veto, "empty bin added", which is not explained in section 2.10.2, originates from an optimization where the test statistic θ is not recomputed after an empty bin has been added to the region.

In addition to the aforementioned rules, the entry "empty bin added" is listed. This is due to an optimization in the automated search, where regions are not reassessed after an empty bin has been added. The motivation behind this optimization is that the value of θ would be the same (as neither N_{SM}, σ_{SM} or N_{obs} change) and the region would be larger, therefore not a candidate to become RoI.

Overall, more than half of the regions (58.7 %) are skipped. However, the optimization accounts for 38.8 %, leaving only about 20 % to the rules that prevent an invalid inference. The newly introduced veto amounts for about 5 % of vetoed regions, which is an acceptable reduction in the number of regions given that it restores coverage of the θ-value.

5.1.6 Comparison of Test Statistics t

In this section, the results of using several test statistics on the semiclassical black hole model are explored. The distributions of the four aforementioned test statistics can be found in figure 5.6. The test power towards each model is indicated in the legend and in this case varies between 53 % and 100 % for the given test statistics.

The largest test power of 100 % is given by the "simple" test statistic, which just measures the fraction of classes with $\tilde{p} < 0.01$. While this may appear surprising at first, considering that the other test statistics are much more sophisticated, there exists a possible explanation: In the past, the MUSiC analysis has been developed while assessing the sensitivity by eye from the $\tilde{p}$-distribution. The $\tilde{p}$-distribution, which is binned in logarithmic bin sizes, strongly favors deviations in bins with $\tilde{p} \to 0$. Therefore, the analysis might have been optimized for this area of sensitivity.

Only now, with the introduction of a quantitative measure for deviations in the bulk of the distribution, the analysis can be optimized for sensitivity in this region. Therefore, the test power of these statistical test should be considered alongside the "simple" test during future design decisions of the analysis.

5.1.7 Uniform Reference Distribution for t

So far, all results from the global p-tilde value have been obtained by comparing the empirical distribution to a reference distribution given by all SM-only values of $\tilde{p}$. As mentioned in section 4.4, a uniform distribution can be used as alternative reference. In this section the challenges are shown that arise when comparing to a uniform distribution with $N_{\mathrm{SM}} \geq 0.1$ and how the alternative performs if one requires $N_{\mathrm{SM}} \geq 1.0$.

Figure 5.7 shows results of a SM-only validation similar to the results obtained earlier (figure 5.3), but using a uniform distribution as reference. However, unlike earlier

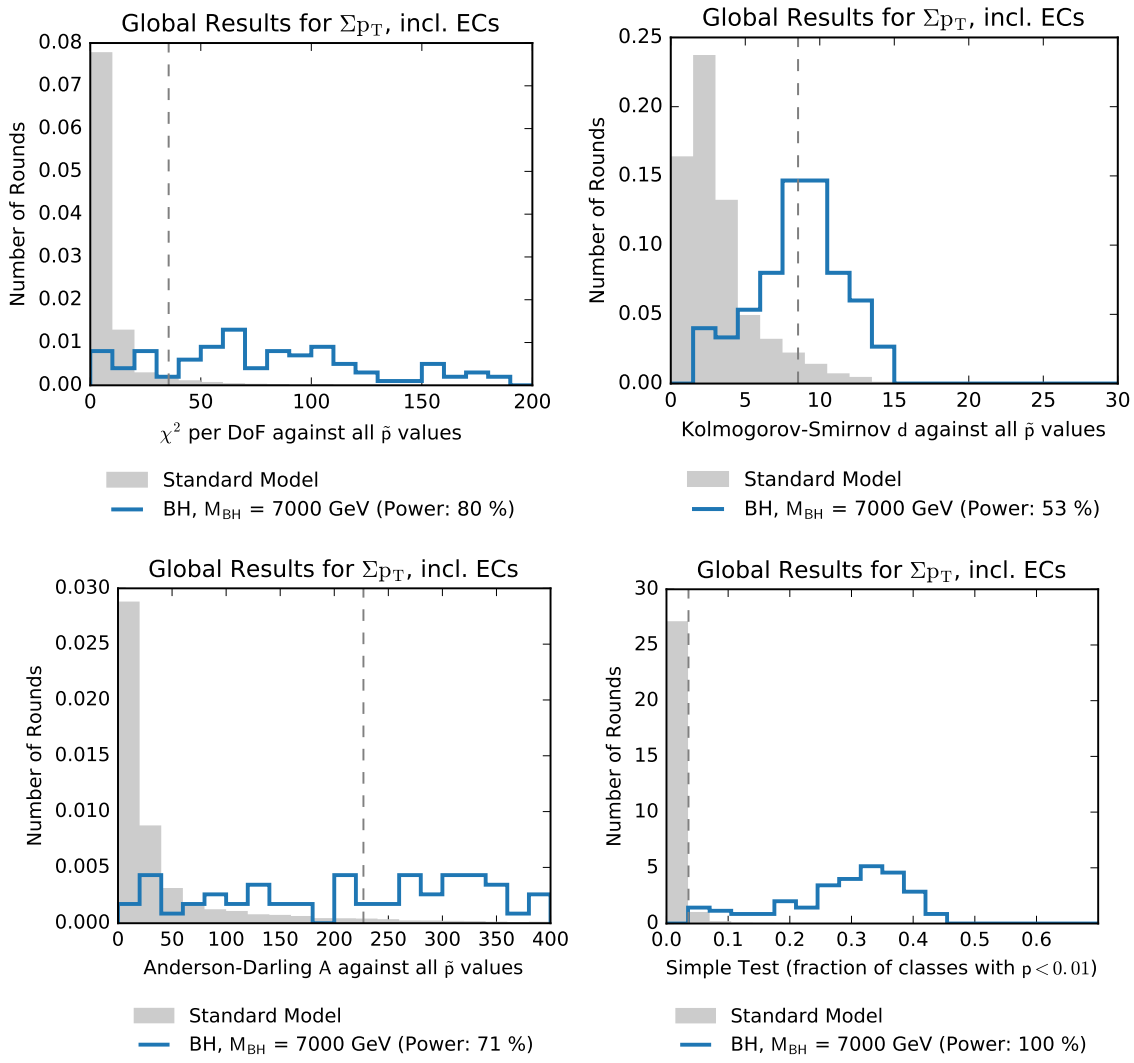


Figure 5.6: Comparison of test statistics for the BH model. For demonstration purposes, the test statistic values of the $M_{\mathrm{BH}} = 7\,000\,\mathrm{GeV}$ signal study are drawn. The deviations between the two distributions in each figure are clearly visible. In this case, the simple test shows the largest power of 100 %, followed by the χ^2 test with 80 %.

results, the test power on the validation is not $\approx 5\,\%$, as it would be expected, but more than 15 %, indicating that using the test statistics in combination with the uniform distribution yields invalid results under these circumstances. Figure 5.8 presents the situation with $N_{\mathrm{SM}} \geq 1.0$. This time, the validation results (upper two figures) indicate a test power of approximately 5 %, as desired. Therefore, one can continue to perform inferences from these results. The two figures on the bottom show distributions obtained from pseudo-experiments of the semiclassical black hole model, similar to figure 5.6 in the previous section. The obtained test power is comparable to the case of a sampled reference, e.g. 81 % instead of 80 % for the χ^2-test, 75 % instead of 53 % for the Kolmogorov-Smirnov test.

Overall one can conclude that this approach is also feasible, it even seems to be more sensitive in some cases. However, as the minimal yield has to be increased up to $N_{\mathrm{SM}} \geq 1.0$, in practice there will be event classes that do not pass the threshold but contain data. These event classes will have to be reviewed manually by the analyst. To reduce the amount of required manual involvement, the uniform approach with a larger threshold is not further used in this thesis.

5.1.8 Sensitivity in Few Final States (QBH model)

The QBH model introduced in section 1.3 serves as benchmark for new physics appearing in few final states. Since the simulated black holes are assumed to decay completely into a e + μ pair, this final state is expected to dominate the list of significant classes. To illustrate this, the black hole mass of $M = 4\,000\,\mathrm{GeV}$ is considered. It is the highest mass point for this model which generates a significant deviation from the SM. The results of the analysis are presented as a distribution of $\tilde{p}$-values and table of most significant classes in figure 5.9 and the distribution of t-values in figure 5.10.

As expected, the set of most significant classes is dominated by final states containing an e + μ pair. The most significant event class 1e + 1μ + X has $\tilde{p} < 1 \times 10^{-4}$. The inequality symbol indicates that in none of the 10 000 SM-only pseudo-experiments, a more significant deviation has been found.

The second most significant event class is the final state 1e + 1μ + $E_{\mathrm{T}}^{\mathrm{miss}}$ + X. Its significance in the signal study is 3.5σ, followed by an insignificant event class with $Z = 2.1 < 3$.

The $\tilde{p}$-distribution indicates a similar result: There is no apparent deviation within the bulk of the distribution, but the overflow bin contains two event classes in the median, occasionally also none or more than two. However, overall the distribution is compatible with the prediction from SM-only pseudo-experiments. The same applies to the distribution of t-values in figure 5.10. The only test statistic that

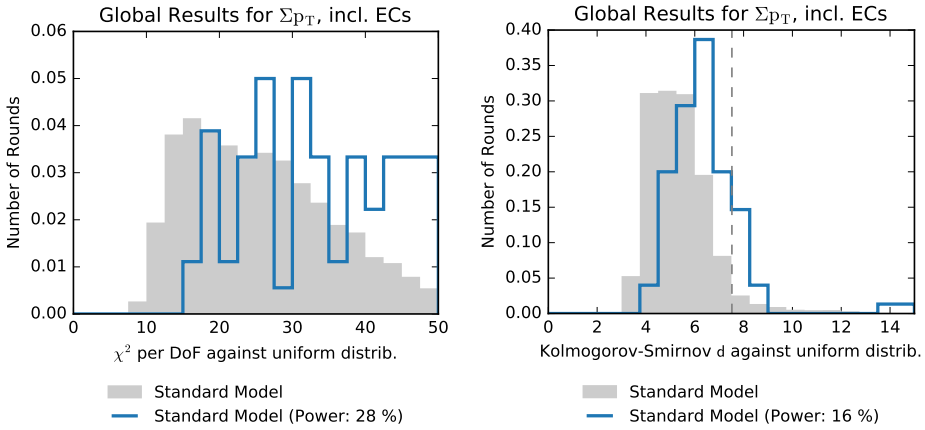


Figure 5.7: Distributions of t-values, obtained by comparing to a uniform distribution. This case, calculated with $N_{\mathrm{SM}} \geq 0.1$, shows a failed validation attempt: Instead of 5 %, the test power is 16 % to 28 %, indicating that the method cannot be used with these parameters.

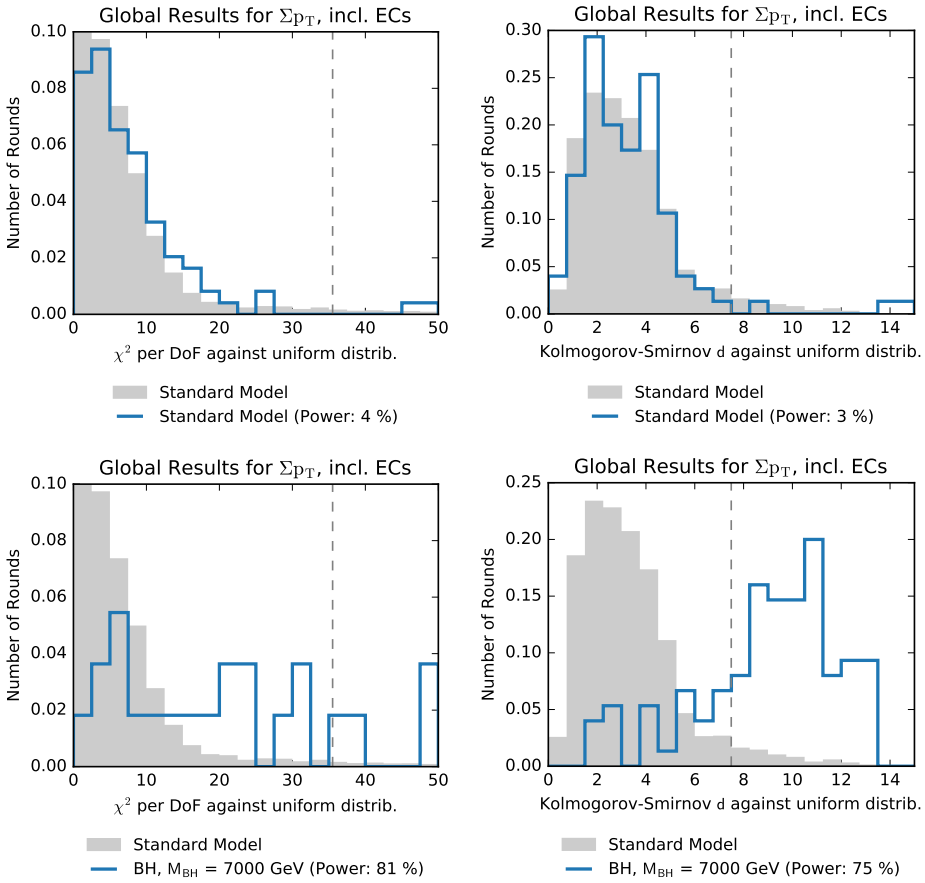
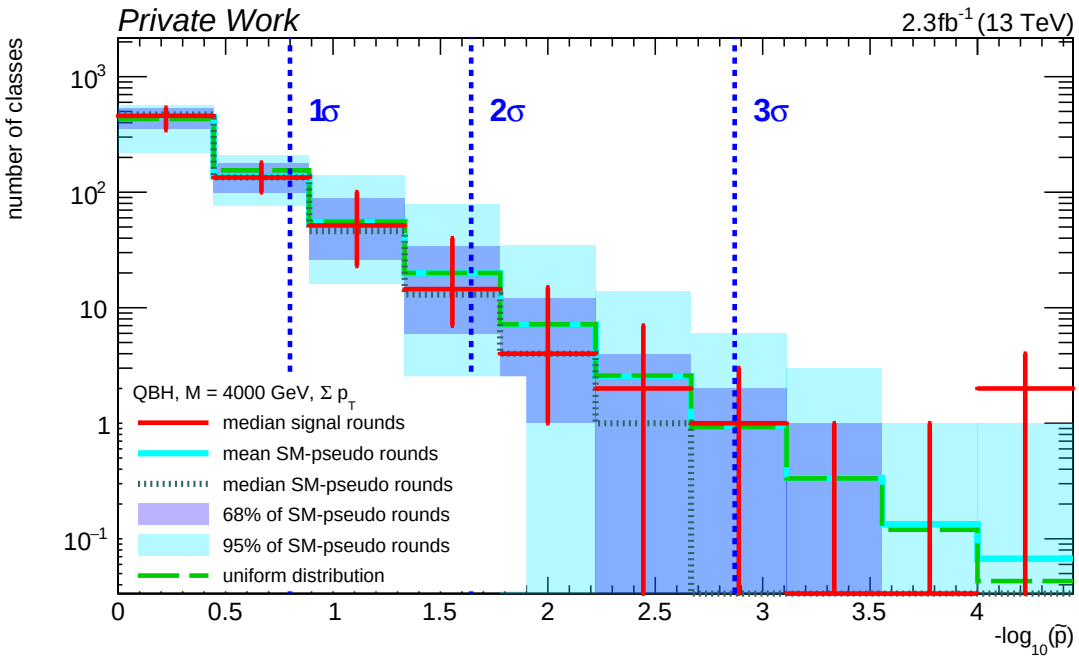


Figure 5.8: Distributions of t-values, obtained by comparing to a uniform distribution, with an increased minimal yield requirement of $N_{\mathrm{SM}} \geq 1.0$. The two upper illustrations show the same validation as in figure 5.7. As desired, the test power is $1 - \beta \approx \alpha = 5\,\%$. The illustrations below show the results of the black hole model, suggesting that the test power is comparable to figure 5.6.



Event Class	Median $\tilde{p}$	Z
1 e + 1 μ + X	$<1.00 \times 10^{-4}$	>3.7
1 e + 1 μ + $E_{\mathrm{T}}^{\mathrm{miss}}$ + X	2.00×10^{-4}	3.5
1 e + X	1.96×10^{-2}	2.1
1 e + 1 μ + 1 jet + X	3.00×10^{-2}	1.9
1 e + 1 μ + 1 jet + $E_{\mathrm{T}}^{\mathrm{miss}}$ + X	1.49×10^{-1}	1.0
1 μ + X	1.66×10^{-1}	1.0
1 e + $E_{\mathrm{T}}^{\mathrm{miss}}$ + X	2.17×10^{-1}	0.8
1 μ + $E_{\mathrm{T}}^{\mathrm{miss}}$ + X	2.50×10^{-1}	0.7
1 e + 1 μ + 3 jets + X	2.98×10^{-1}	0.5
1 e + 1 μ + 2 jets + X	3.06×10^{-1}	0.5

Figure 5.9: Distribution of $\tilde{p}$-values and most significant inclusive event classes for the quantum black hole decaying to e + μ, $M = 4\,000$ GeV. This model represents new physics appearing in few final-states. As expected, only one highly significant event class has been found.

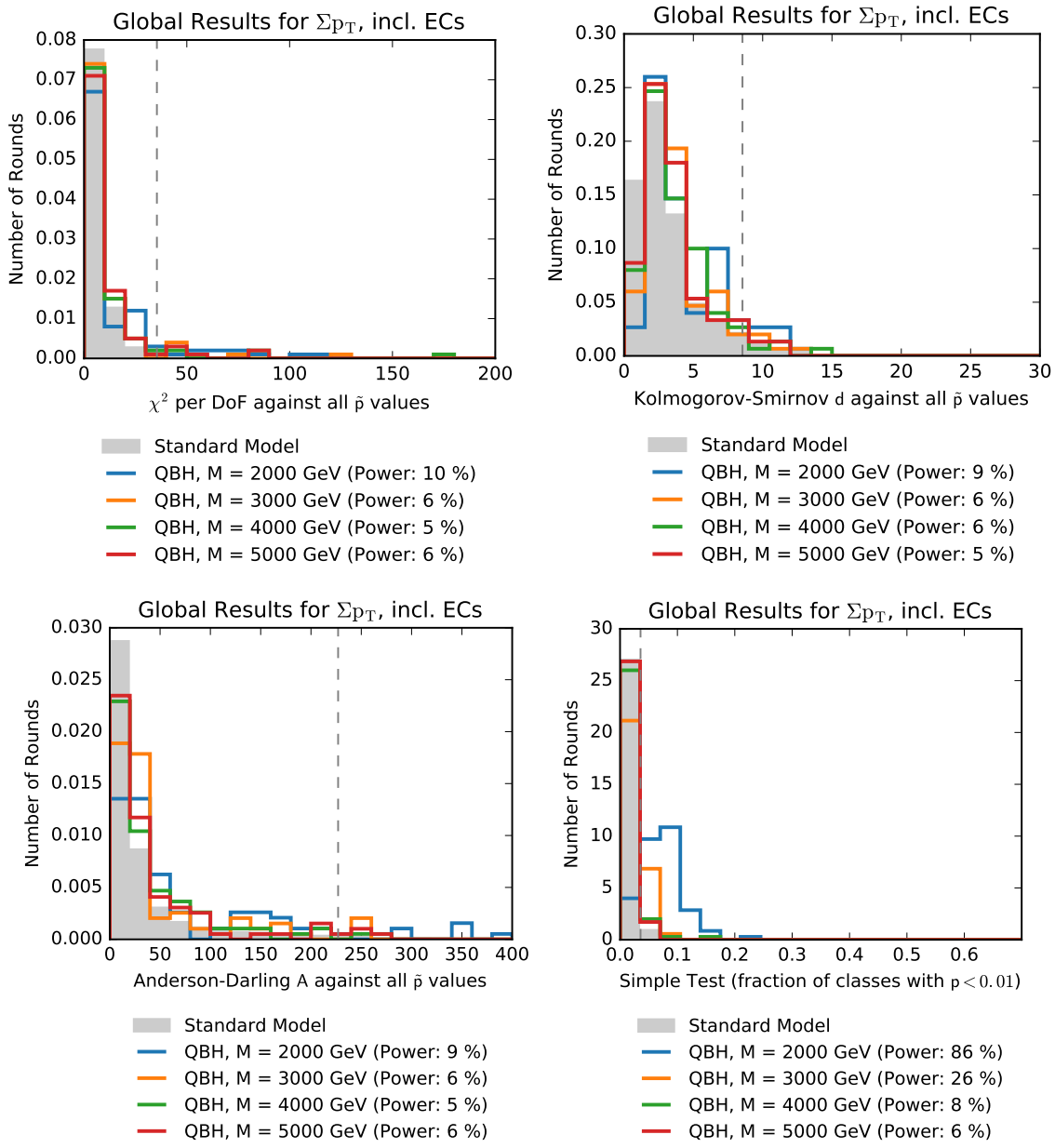


Figure 5.10: Distribution of t-values for the global p-value. The shown signal model is the QBH model, which represents a model appearing in few very significant event classes. None of the test statistics feature a sensitivity comparable to the table of most significant event classes.

shows sensitivity to this model is the simple test, with a test power of 86 % for $M = 2\,000$ GeV. However, this mass is much lower than the most sensitive mass that could be discovered using the table of most significant classes.

Overall, this study shows that the sensitivity towards a model that appears in one or a few final states is mostly given by the table of median most significant classes. Neither the distribution of $\tilde{p}$-values nor the distributions of t-values can provide a comparable sensitivity in this case.

5.1.9 Sensitivity in Multiple Final States (BH model)

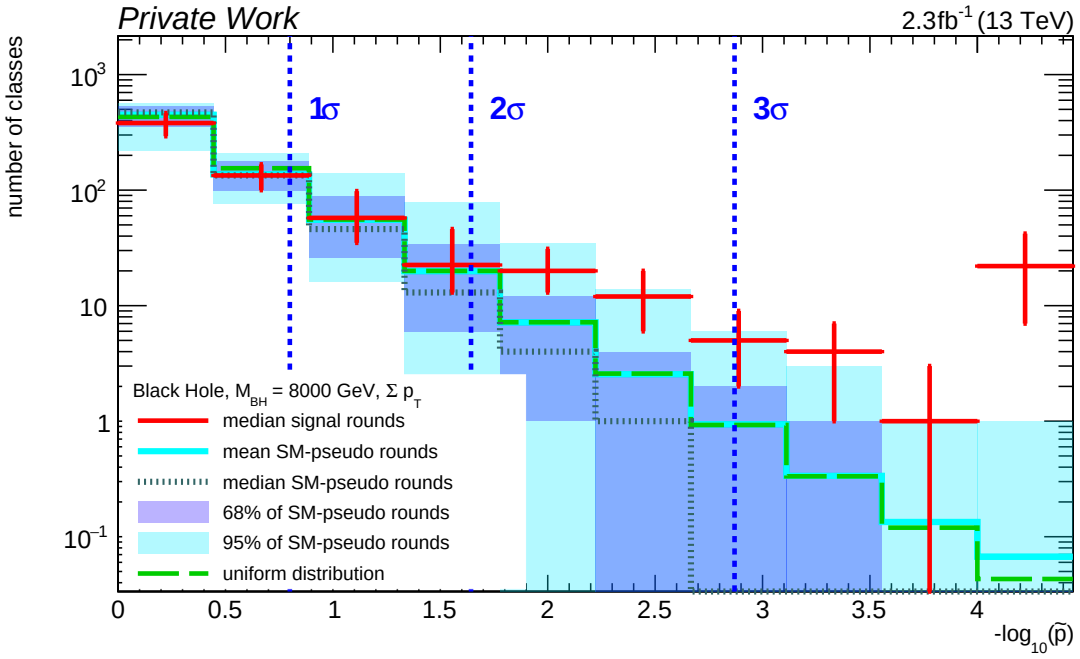
As opposed to the previous section, the goal of this section is to assess sensitivity of the MUSiC analysis towards a new physic that appear in multiple final states. The benchmark model for this analysis is the semiclassical black hole model introduced in section 1.3, with a mass of $M_{\mathrm{BH}} = 8\,000$ GeV. Similarly to the previous section, the results are first presented as distribution of $\tilde{p}$-values and as a table of most significant classes in figure 5.11.

The table of most significant classes shows, in contrast to the previous case, that there are multiple event classes that are sensitive to the model. The three most significant event classes have a Z-score of more than 3.7, followed by five more event classes above 3σ.

It is noticeable that the table of median most significant inclusive event classes is dominated by event classes containing several jets and some containing missing transverse energy. As mentioned in the introductory chapter, the branching ratio of the semiclassical black hole is proportional to the number of degrees of freedom of the decay products. Thus, quarks and gluons, which carry additional degrees of freedom through color charge, are heavily favored, resulting in multiple jets in the final state. A significant amount of $E_{\mathrm{T}}^{\mathrm{miss}}$ is also expected to appear during the evaporation of the black hole [38].

For this model, the distribution of $\tilde{p}$-values is very sensitive. During each round, there were about eleven event classes with $\tilde{p} < 1 \times 10^{-4}$, making the deviation from the SM-only distribution highly significant. Note that the event classes which are sorted into the last bin vary between pseudo-experiment rounds and therefore do not appear with the same significance in the results table. Again, a similar result is apparent in the distribution of t-values in figure 5.12. All test statistics show a median sensitivity up to a black hole mass of $M_{\mathrm{BH}} = 7\,000$ GeV, the simple test is even sensitive up to $M_{\mathrm{BH}} = 8\,000$ GeV, with a test power of 88 %.

The study in this section has shown different behavior than the previous study: For new physics appearing in many final states, the distribution of $\tilde{p}$-values as well as the



Event Class	Median $\tilde{p}$	Z
5 jets + $E_{\mathrm{T}}^{\mathrm{miss}}$ + X	$<1.00 \times 10^{-4}$	>3.7
1 μ + 5 jets + X	$<1.00 \times 10^{-4}$	>3.7
4 jets + X	$<1.00 \times 10^{-4}$	>3.7
5 jets + X	2.00×10^{-4}	3.5
4 jets + $E_{\mathrm{T}}^{\mathrm{miss}}$ + X	2.00×10^{-4}	3.5
3 jets + $E_{\mathrm{T}}^{\mathrm{miss}}$ + X	7.00×10^{-4}	3.2
6 jets + $E_{\mathrm{T}}^{\mathrm{miss}}$ + X	8.00×10^{-4}	3.2
1 μ + 3 jets + $E_{\mathrm{T}}^{\mathrm{miss}}$ + X	8.00×10^{-4}	3.2
6 jets + X	1.20×10^{-3}	3.0
1 μ + 4 jets + $E_{\mathrm{T}}^{\mathrm{miss}}$ + X	1.20×10^{-3}	3.0

Figure 5.11: Distribution of $\tilde{p}$-values and most significant inclusive event classes for the black hole model at the mass of $M_{\mathrm{BH}} = 8\,000$ GeV. This model has an effect on multiple final states. Therefore the distribution of $\tilde{p}$-values shows a deviation in the bulk of the distribution as well as the highest bin and there are several event classes with $Z > 3$.

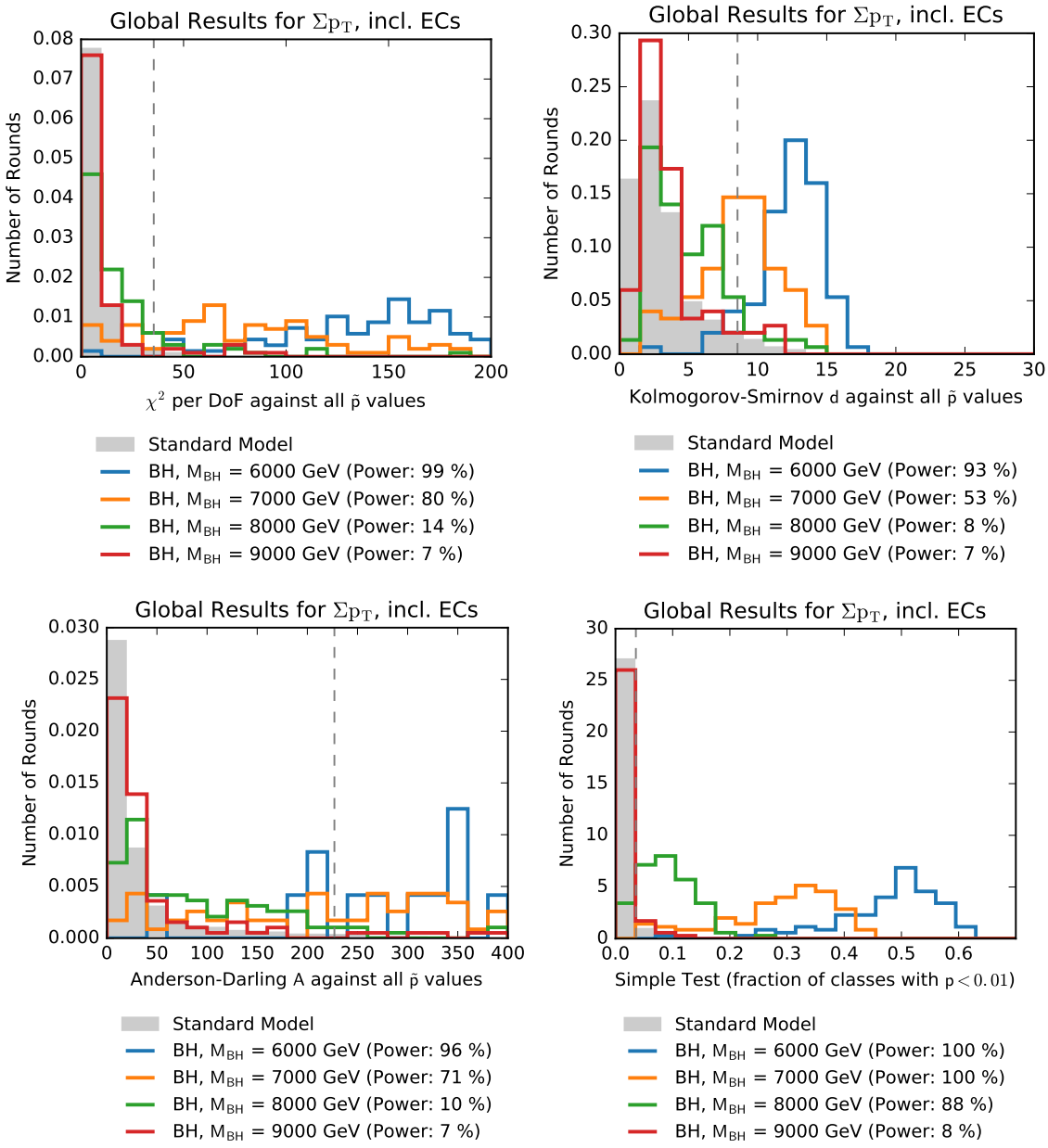


Figure 5.12: Distributions of t-values corresponding to figure 5.11. The semiclassical black hole model represents new physics that appear in multiple event classes. All test statistics are sensitive to the new physics model, the highest sensitivity is again obtained by the "simple" test, with a power of 88 % at $M_{\mathrm{BH}} = 8\,000\,\mathrm{GeV}$.

global p-value display enhanced sensitivity and therefore complement the table of median most significant classes.

5.1.10 Dependence on the Luminosity

In the previous section, the highest discoverable mass points of the QBH and BH models have been presented, with respect to a luminosity of 2.3 fb^{-1}. This section revisits the results, this time with the luminosity scenario of 2016 (35.9 fb^{-1}). The corresponding distributions are depicted in figure 5.13.

The figure on top shows the result for the QBH model with a black hole mass of $M = 4\,000$ GeV. At 2.3 fb^{-1}, there were two classes belonging to the overflow bin on average. One can see that this number has increased at 35.9 fb^{-1}, showing nine event classes in the same bin on average. The bulk of the distribution, however, remains unchanged: No significant deviations are apparent in all but the highest histogram bin.

The second figure shows the distribution of $\tilde{p}$ models from pseudo-experiments based on the BH model. This model highly benefits from the increase in luminosity. On average, there are over 100 classes in the overflow bin. This indicates that the analysis is sensitive to even higher black hole masses, as will be presented in section 5.2.1.

5.2 Results by Model

The following sections aim to present absolute sensitivity results towards the tested benchmark models. As the cross section of each model decreases with the mass of the new physics object (black hole, Σ or W′), the results will be shown for the highest mass that the analysis would be able to discover.

For this purpose, a model will be claimed to be discoverable if any of the following conditions are met: The distribution of $\tilde{p}$-values shows a considerable deviation, the most significant event class has a median Z-score larger than 3 or the test power of t is larger than 50 %, i.e. the median t is larger than t_{crit}, for any test statistic.

5.2.1 Semiclassical Black Hole

The results for the semiclassical black hole theory have already been presented and discussed in the previous sections. At a luminosity of 2.3 fb^{-1}, the largest black hole mass that the MUSiC analysis is very sensitive is about $M_{\text{BH}} = 8\,000$ GeV, with eleven event classes in the overflow bin.

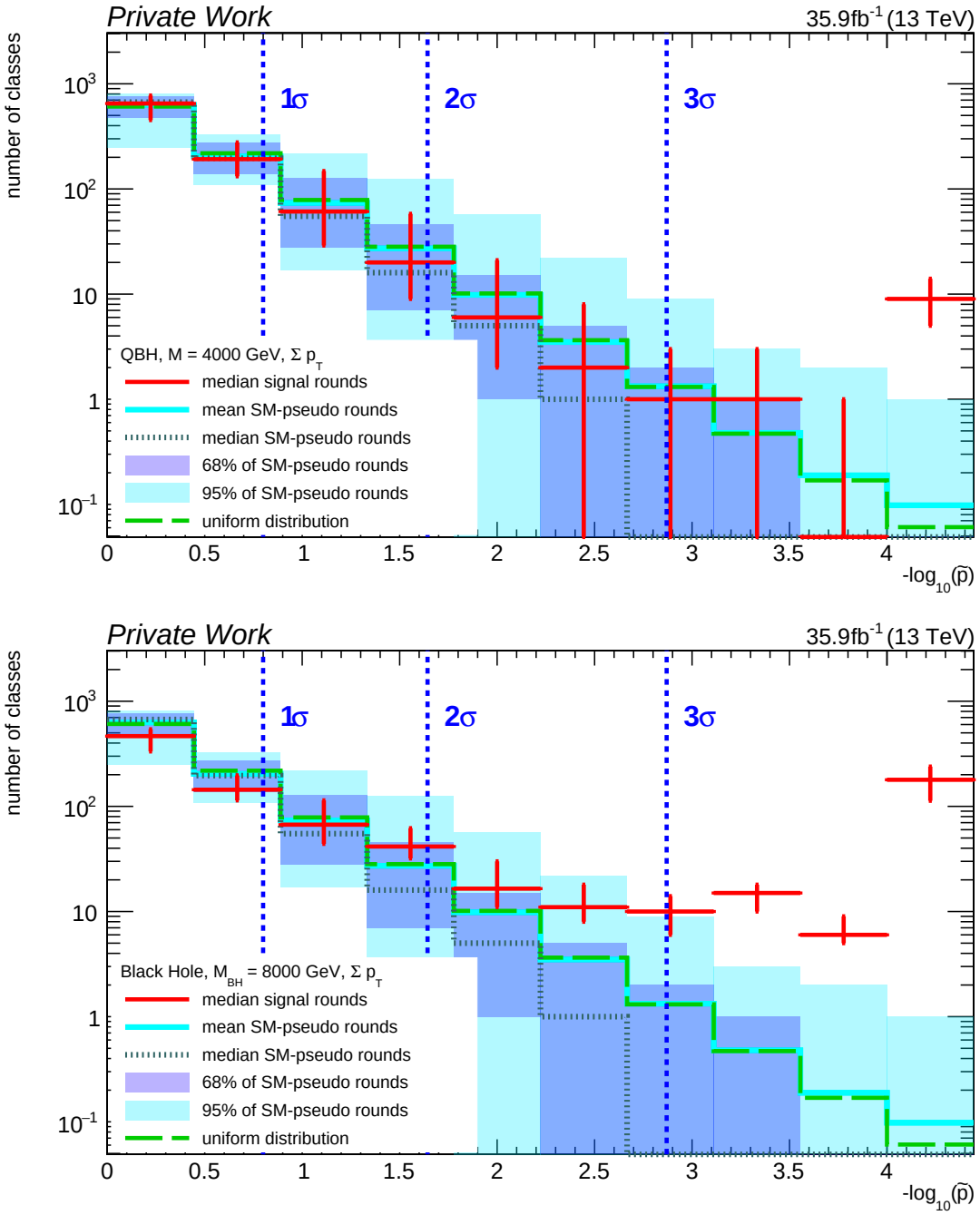


Figure 5.13: Distributions of figure 5.9 and figure 5.11 for the 2016 scenario with a luminosity of 35.9 fb^{-1}, using the same models and mass points. On top: quantum black hole model with $M = 4\,000\,\text{GeV}$, below: black hole model with $M_{\text{BH}} = 8\,000\,\text{GeV}$.

As suggested in the previous section, increasing the luminosity from 2.3 fb^{-1} to 35.9 fb^{-1} additionally enables sensitivity up to a black hole mass of $M_{\text{BH}} = 9\,000$ GeV, as shown in figure 5.14. At 35.9 fb^{-1}, there are three inclusive event classes with a median Z-score larger than 3 and a clearly visible deviation in the $\tilde{p}$-distribution.

5.2.2 Quantum Black Hole

Similarly to the semiclassical black hole, the results regarding the QBH model have also been presented earlier in this chapter. In figure 5.9, sensitivity to the QBH model up to a black hole mass of $M = 4\,000$ GeV is demonstrated.

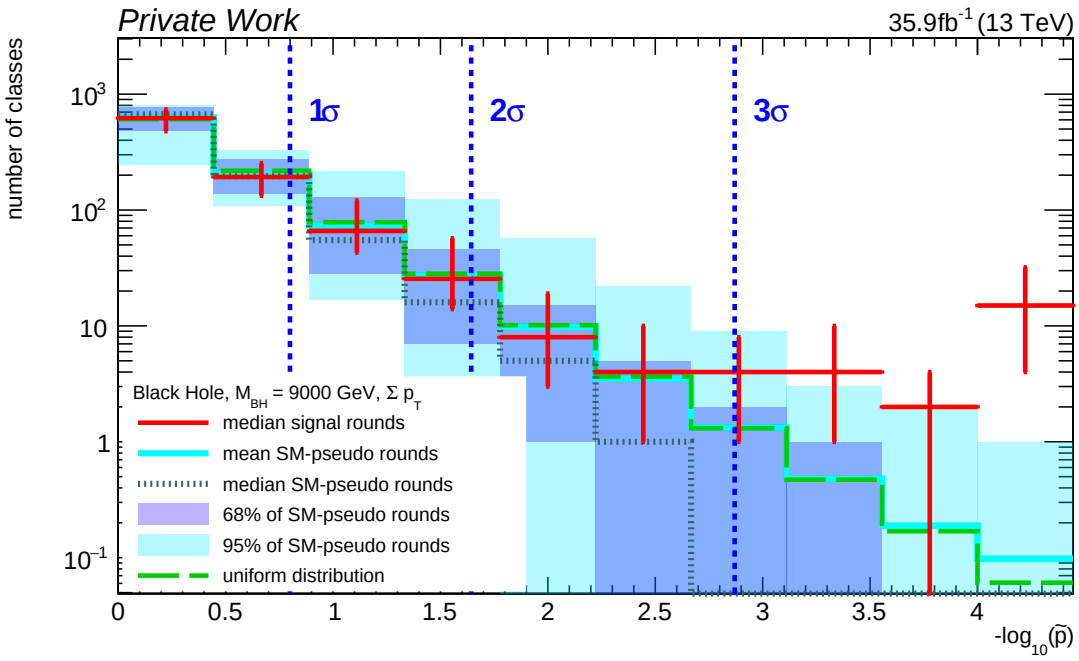
With an increased luminosity of 35.9 fb^{-1}, the analysis becomes sensitive up to a QBH mass of $M = 5\,000$ GeV (figure 5.15), with 1e + 1μ + $E_{\text{T}}^{\text{miss}}$ + X being the most significant event class ($Z = 3.5$).

5.2.3 Seesaw Type-III

The results for the Seesaw Type-III signal model are presented in figure 5.16. Because no sensitivity can be observed in the dataset of 2.3 fb^{-1}, the illustration directly contains the results corresponding to a luminosity of 35.9 fb^{-1}. At 35.9 fb^{-1}, the distribution of $\tilde{p}$-values shows a slight deviation from the SM-only distribution in the overflow bin. The table of median most significant classes shows that the event class 1e + 3μ + X stands out with a significance of $Z = 3.4$.

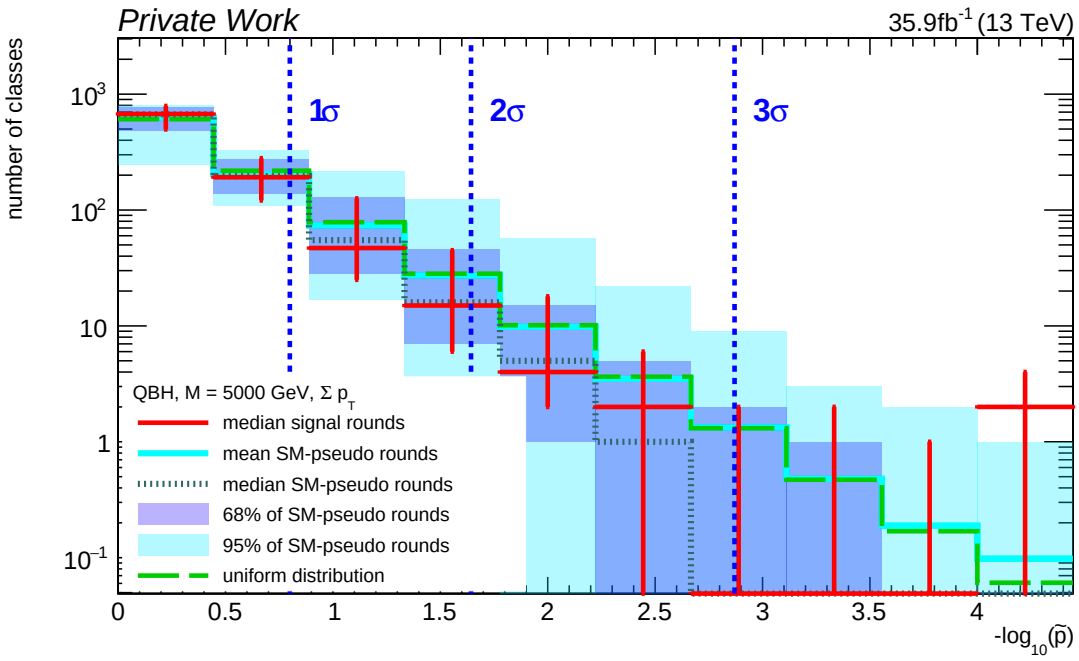
To understand the comparably low sensitivity towards this model, it is helpful to compare the selection and procedure of the MUSiC analysis to the dedicated analysis which was able to exclude the same model up to a mass of 440 GeV in 2015 [115] and 790 GeV in 2016 [129].

The dedicated analysis applies several selection criteria to maximize signal efficiency and suppress the SM contribution in the final states under investigation. First, only events with three or more leptons are considered. The leptons are expected to pass comparably low p_{T} thresholds of 25 GeV and less. Subsequently, events are classified into six statistically independent search channels which correspond to the decay channels of the Σ fermions. Intermediate Z bosons are reconstructed by matching pairs of leptons with the same flavor and opposite electrical charges. The pairs are then additionally binned by the invariant mass. For search channels including intermediate W bosons, the kinematic variable $\Sigma p_{\text{T}} + E_{\text{T}}^{\text{miss}}$ is considered in order to combine the momenta of visible leptons and neutrinos. Overall, the search strategy of the dedicated search is to start with as many events as possible and quickly narrow them down using precise selection criteria. Therefore, the signal efficiency is larger than in this analysis while sufficiently suppressing the SM contribution.



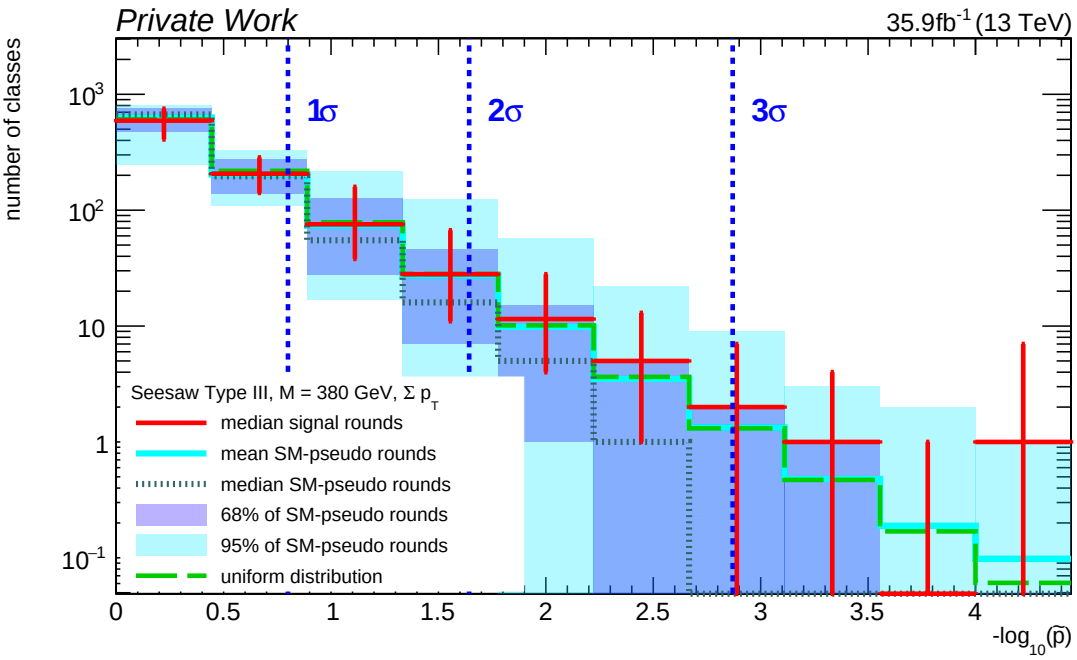
Event Class	Median $\tilde{p}$	Z
1 e + 6 jets + $E_{\mathrm{T}}^{\mathrm{miss}}$ + X	$<1.00 \times 10^{-4}$	>3.7
6 jets + $E_{\mathrm{T}}^{\mathrm{miss}}$ + X	2.00×10^{-4}	3.5
1 μ + 6 jets + $E_{\mathrm{T}}^{\mathrm{miss}}$ + X	1.00×10^{-3}	3.1
5 jets + X	1.20×10^{-3}	3.0
5 jets + $E_{\mathrm{T}}^{\mathrm{miss}}$ + X	1.30×10^{-3}	3.0
1 μ + 4 jets + $E_{\mathrm{T}}^{\mathrm{miss}}$ + X	1.40×10^{-3}	3.0
4 jets + $E_{\mathrm{T}}^{\mathrm{miss}}$ + X	1.70×10^{-3}	2.9
1 μ + 5 jets + $E_{\mathrm{T}}^{\mathrm{miss}}$ + X	1.90×10^{-3}	2.9
1 μ + 5 jets + X	2.10×10^{-3}	2.9
1 e + 5 jets + $E_{\mathrm{T}}^{\mathrm{miss}}$ + X	3.00×10^{-3}	2.7

Figure 5.14: Distribution of $\tilde{p}$-values and most significant inclusive event classes for the black hole (BH) model at the mass of $M_{\mathrm{BH}} = 9\,000\,\mathrm{GeV}$ and the 2016 luminosity scenario of $35.9\,\mathrm{fb}^{-1}$. For comparison, at $2.3\,\mathrm{fb}^{-1}$, the highest discoverable mass point was at 8 000 GeV.



Event Class	Median $\tilde{p}$	Z
1 e + 1 μ + $E_{\mathrm{T}}^{\mathrm{miss}}$ + X	2.00×10^{-4}	3.5
1 e + 1 μ + X	5.00×10^{-4}	3.3
1 e + X	1.82×10^{-2}	2.1
1 e + 1 μ + 1 jet + $E_{\mathrm{T}}^{\mathrm{miss}}$ + X	4.39×10^{-2}	1.7
1 e + 1 μ + 1 jet + X	1.47×10^{-1}	1.1
1 e + $E_{\mathrm{T}}^{\mathrm{miss}}$ + X	1.50×10^{-1}	1.0
1 e + 1 μ + 2 jets + $E_{\mathrm{T}}^{\mathrm{miss}}$ + X	3.48×10^{-1}	0.4
1 e + 1 γ + X	3.74×10^{-1}	0.3
2 e + 1 μ + 1 γ + $E_{\mathrm{T}}^{\mathrm{miss}}$ + X	3.74×10^{-1}	0.3
1 e + 1 γ + 3 jets + $E_{\mathrm{T}}^{\mathrm{miss}}$ + 2 b-jets + X	3.93×10^{-1}	0.3

Figure 5.15: Distribution of $\tilde{p}$-values and most significant inclusive event classes for the quantum black hole model at the mass of $M = 5\,000$ GeV and the 2016 luminosity scenario of 35.9 fb^{-1}. For comparison, at 2.3 fb^{-1}, the highest discoverable mass point was at 4 000 GeV.



Event Class	Median $\tilde{p}$	Z
1 e + 3 μ + X	3.00×10^{-4}	3.4
1 e + 3 μ + 1 jet + X	5.00×10^{-3}	2.6
1 e + 3 μ + $E_{\mathrm{T}}^{\mathrm{miss}}$ + X	1.31×10^{-2}	2.2
3 e + 1 μ + X	1.53×10^{-2}	2.2
2 e + 2 μ + $E_{\mathrm{T}}^{\mathrm{miss}}$ + X	3.25×10^{-2}	1.8
3 e + 1 μ + 1 jet + X	4.16×10^{-2}	1.7
2 e + 1 μ + $E_{\mathrm{T}}^{\mathrm{miss}}$ + X	4.71×10^{-2}	1.7
1 e + 2 μ + $E_{\mathrm{T}}^{\mathrm{miss}}$ + X	5.27×10^{-2}	1.6
1 e + 3 μ + 2 jets + X	6.34×10^{-2}	1.5
4 μ + $E_{\mathrm{T}}^{\mathrm{miss}}$ + X	8.84×10^{-2}	1.4

Figure 5.16: Distribution of $\tilde{p}$-values and most significant inclusive event classes for the Seesaw Type-III model at the mass of $M = 380\,\mathrm{GeV}$ and a luminosity of $35.9\,\mathrm{fb}^{-1}$.

5.2.4 $W' \to tb$

The W' model was originally included in this thesis in order to demonstrate the increase of sensitivity due to enabling b-tagged jets as analysis objects. However, as one can see in figure 5.17, even with b-tagged jets and the luminosity of 35.9 fb^{-1}, the MUSiC analysis is not sensitive to deviations caused by the new physics model: The $\tilde{p}$-distribution shows no significant deviation between the distribution of $\tilde{p}$-values from SM-pseudo-experiments and pseudo-experiments involving contributions of the W' boson. The median most significant class is insignificant with a Z-score of 0.3.

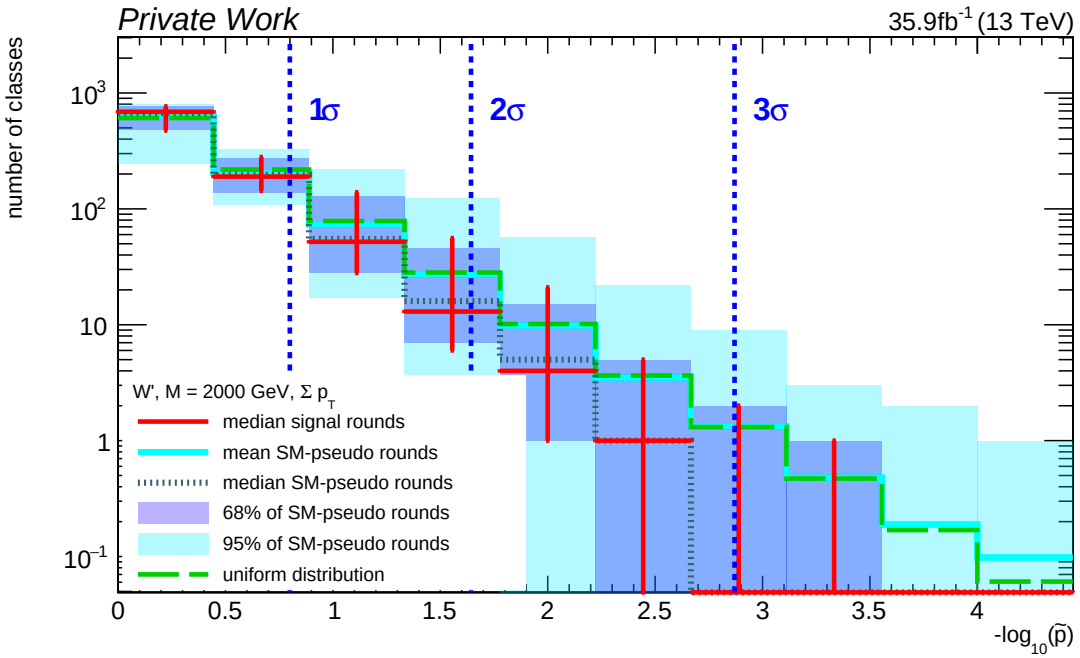


Figure 5.17: Distribution of $\tilde{p}$-values and most significant inclusive event classes for the W' model at the mass of $M = 2\,000$ GeV and a luminosity of 35.9 fb^{-1}.

In order to learn about improvement opportunities, again the search strategy of the dedicated search [46] should be discussed.

The dedicated analysis focuses on the decay cascade $W' \to tb \to Wbb \to \ell\nu bb$. The trigger thresholds used are comparable to the ones used in this analysis. Subsequently, events are required to contain exactly one lepton with $p_T \geq 180$ GeV, a significant amount of E_T^{miss} and two jets.

The search strategy consists of reconstructing the complete four momentum of the W' boson as follows: As first step, the four momentum of the W boson is reconstructed by combining the four momenta of the lepton and E_T^{miss}, where the the longitudinal component of $\vec{E}_T^{miss}$ is calculated from assuming the W invariant mass to be exactly 80.4 GeV. In a similar fashion, the W boson and one of the jets are combined to form an object with the invariant mass close to the nominal t-quark mass. Finally,

the top quark candidate is combined with the most energetic remaining jet in order to reconstruct the W′ boson. Standard Model contributions are further suppressed by requiring the top quark to have $p_{\mathrm{T}} > 650\,\mathrm{GeV}$ and both jets to have a combined momentum of $p_{\mathrm{T}} > 700\,\mathrm{GeV}$. This increases the signal to background ratio, and therefore the sensitivity.

5.2.5 Comparison to Limits of Dedicated Analyses

Several groups at CMS have optimized and applied dedicated analyses for the new physics models considered in this thesis. In those instances where these analyses do not find a significant deviation of the observed data from the SM expectation, *limits* are calculated.

The term "limit" is commonly used to describe two distinct quantities: In the first meaning, the term refers to a *cross section limit*. Roughly speaking, the cross section limit denotes the maximal cross section that a new physics process could exhibit while being in agreement with the observed data. The numerical value is usually stated to the 95 % confidence level, meaning that the new physics model is incorrectly rejected in only 5 % of the cases. Note that this is not directly comparable to the discovery threshold α used in this thesis, where one would incorrectly reject the Standard Model with a probability of 5 %.

As most theoretical cross sections fall steeply with the mass of the predicted particle, setting an upper bound on the possible cross section also sets a lower bound on the particle's mass. This is expressed in the *mass limit*, which is derived from the cross section limit.

The mass limits of the aforementioned dedicated analyses working on the same signal samples are listed in table 5.9, alongside with the highest sensitive mass points of the MUSiC analysis as determined in this chapter.

	$\mathcal{L} \approx 2.3\,\mathrm{fb}^{-1}$		$\mathcal{L} \approx 35.9\,\mathrm{fb}^{-1}$	
Model	CMS Limit	MUSiC	CMS Limit	MUSiC
QBH $n = 4$	4.2 TeV [39]	4 TeV	5.4 TeV [130]	5 TeV
Black Hole	8.6 TeV [38]	8 TeV	-	9 TeV
Seesaw	440 GeV [115]	< 380 GeV	790 GeV [129]	380 GeV
$\mathrm{W}' \to \mathrm{tb}$	2.4 TeV [45]	< 2 TeV	3.4 TeV [46]	< 2 TeV

Table 5.9: Comparison of the discovery thresholds determined in this chapter to the expected mass limits of comparable dedicated analyses published by the CMS collaboration.

The results indicate that the discovery thresholds for the semiclassical black hole and QBH correspond approximately to the mass limits obtained by optimized analyses. For the Seesaw model, the sensitivity by MUSiC is significantly lower than the exclusion limit by the dedicated analysis. In case of the W′ model, the comparison cannot be directly performed as the MUSiC discovery threshold is unknown, thus only an upper limit is given by the lowest object mass under investigation (2 TeV). Possible reasons for a greater sensitivity of the dedicated analyses have been discussed in section 5.2.3 and section 5.2.4.

5.3 General Validity for New Physics

In this thesis, the sensitivity of the MUSiC analysis towards four simulated models of new physics has been assessed. For both black hole models it was deduced that the analysis would discover the presence of this model in observed data for the luminosity scenarios 2015 and 2016. For one model, the Seesaw theory, only the enhanced luminosity of 35.9 fb^{-1} in 2016 enables MUSiC to observe the phenomena, the discovery threshold is however well below the exclusion limit by dedicated analyses. For the last model (W′), the analysis fails to reject the Standard Model even in the presence of new physics.

However, one goal of a *model unspecific* search is to also be sensitive to new physics that are not represented by a known theory at the time of the analysis. These theories in turn cannot be simulated and the ultimate sensitivity cannot be determined. Therefore it is important to discuss how far the knowledge gained on the tested models can be transferred to unknown new physics.

There are many possibilities for new physics to be discovered at CMS: If new physics appears as a new object, its mass could be anywhere between a few GeV up to several TeV to be produced at the LHC. In some scenarios, it might not even have a well defined mass and thus not produce a resonance. Subsequently, the object can possibly decay either into a few SM decay products with high momenta or alternatively into many low-energetic decay products. A decay might obey conservation laws of SM quantum numbers, or maybe it violates them. If the new particle only interacts weakly, the new physics object might even occur as displaced track or leave the detector unnoticed, leaving behind any amount of $E_{\text{T}}^{\text{miss}}$.

Unfortunately, the MUSiC analysis cannot cover all possibilities, neither can a study of the discovery potential. Nevertheless, in this thesis a large range of particle masses, from $\sim$ 100 GeV up to $\sim$ 9 TeV, has been probed. Additionally, several options for the multiplicity of decay products have been explored, ranging from two decay products as in the QBH model up to several jets in the BH model.

Overall, the analysis has been found to be especially sensitive towards new physics appearing in regions with low SM-contribution, including large values of the kinematic variables and final states that violate conservation laws of the Standard Model. In areas with a large SM contribution, dedicated analyses tend to perform better than MUSiC because of more sophisticated selection strategies.

5.4 Towards a Higher Sensitivity

The MUSiC analysis has many parameters and inputs that can be optimized to improve the sensitivity towards certain models. This section aims to provide a few suggestions based on the conclusions drawn from the study of the discovery potential in this chapter and with regard to the dedicated analyses of the benchmark models.

In case of the Seesaw Type-III model, it has been discussed that the signal efficiency of the MUSiC selection is significantly lower than the efficiency determined by the dedicated analysis [115, 129]. One possible explanation is the trigger threshold: The highest p_{T} requirement imposed by the dedicated analysis is $p_{\mathrm{T}} > 25\,\mathrm{GeV}$. As shown earlier in table 2.1, the MUSiC analysis imposes a higher transverse momentum threshold, possibly discarding signal events with multiple low energetic leptons. Therefore, in order to increase the sensitivity to these types of new physics models, I would suggest to use a trigger stream with a lower p_{T} threshold.

The second suggestion would likely increase the sensitivity in both the Seesaw Type-III as well as the W′ model: Both dedicated analyses make use of combining four momenta of decay products in order to reconstruct intermediate particles. This concept, tagging, could also be applied to the MUSiC analysis by extending the set of analysis objects to t-jets, τ-leptons or Z-bosons.

A final suggestion applies to all signal models, but has especially been presented on the example of the semiclassical black hole model: An increase in luminosity directly causes an increase in sensitivity towards new physics. Not only does a larger number of data events cause lower statistical uncertainties, but it also enables more event classes to pass the minimum yield threshold and thus participate in the statistical inference. After all, statistically combining several event classes is one of the largest advantages that the MUSiC analysis has compared to dedicated analyses. In practice, the amount of data events to analyze can usually not be influenced by the analyst. However, this suggestion shall motivate to further pursue a model independent approach, as the LHC is expected to deliver a total integrated luminosity of about $100\,\mathrm{fb}^{-1}$ of collisions up to 2018 [131].

Conclusion

6

The Model Unspecific Search in CMS (MUSiC) is a complex analysis with the ultimate goal to discover new physics in LHC data. It consists of a multitude of algorithms and parameters which have to be implemented and optimized by the analyst. This process is usually guided by physical intuition and manual inspection of analysis results. However, it is important to regularly perform a systematic reevaluation of the state of the analysis and study its discovery potential. In this thesis, the framework for such a study has been refined and subsequently applied to existing features as well as newly introduced extensions of the analysis.

Among the evaluated features was the inclusion of b-tagged jets as dedicated physics objects, which has not been part of the analysis since the increase in collision energy to $\sqrt{s} = 13\,\text{TeV}$ in 2015. Although the increase in sensitivity could not be explicitly shown on the provided benchmark models, the feature is predicted to increase sensitivity towards new physics and therefore should be further investigated in the future.

On the technical side of the analysis, the complexity of the tool chain has grown. This has motivated employing further automation, through which it was possible to cope with the additional workload posed by the exploration of the multidimensional parameter space. Additionally, a lookup table has been implemented and evaluated in order to optimize the performance of the automated search for deviations.

Taking up discussions published in a thesis eight years ago [68], the potential of a log-normal prior within the local test statistic has been evaluated. For this purpose, a study of the coverage behavior of both the Gaussian- and log-normal prior has been performed, indicating that the log-normal option features superior coverage properties. However, several difficulties regarding pseudo-experiments with a log-normal prior arise. Possible mitigations have been discussed, and it was decided to maintain using the Gaussian prior with additional constraints on the search space.

Another new feature within the analysis is the computation of a global p-value. It allows to quantify deviations within the distribution of $\tilde{p}$-values, instead of judging deviations by eye. The feature enables future analysts to observe whether a certain change of a parameter value or the implementation of a new feature increases the sensitivity towards certain benchmark models.

Finally, the sensitivity of the analysis towards four models of new physics has been assessed. These benchmark models have been chosen to cover a wide range of possible signatures of new physics. By combining simulated events of these models with Standard Model processes, generating pseudo-experiments and analyzing the

resulting distributions with the automated search, the impact of each model on the statistical inference has been evaluated. In two instances, the W′ and the Seesaw Type-III model, the sensitivity did not (or barely) suffice to discover the presence of the simulated events. In these cases, a comparison to the dedicated analyses was drawn, yielding several suggestions for features that may increase the sensitivity. The discovery potential for the two other models, semiclassical and quantum black holes, could successfully be shown. The highest discoverable object masses are in agreement with mass limits from corresponding dedicated analyses.

In the future, I would like to encourage analysts who pursue a model independent analysis approach to regularly reevaluate the sensitivity of their analysis to a wide range of new physics simulations. Furthermore, I would suggest to refine the idea of a global p-value and a measure of sensitivity, as informed decisions will help the analysis to remain lean and efficient for discovering new physics.

Bibliography

B

[1] A. G. Van Melsen, "From atomos to atom: The history of the concept atom". Courier Corporation, 2004.

[2] Particle Data Group, "Review of Particle Physics", *Chin. Phys.* **C40** (2016), no. 10, 100001. doi:10.1088/1674-1137/40/10/100001.

[3] P. A. M. Dirac, "The quantum theory of the electron", *Proc. Roy. Soc. Lond.* **A117** (1928) 610–624. doi:10.1098/rspa.1928.0023.

[4] R. P. Feynman, "Space - time approach to quantum electrodynamics", *Phys. Rev.* **76** (1949) 769–789. doi:10.1103/PhysRev.76.769.

[5] Nobel Media, "The Nobel Prize in Physics 1965". Nobelprize.org, 2014. Web, 06.02.2017.

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

[7] M. Gell-Mann, "The Eightfold Way: A Theory of strong interaction symmetry", technical report, California Inst. of Tech., Pasadena. Synchrotron Lab., 1961.

[8] M. Gell-Mann, "A Schematic Model of Baryons and Mesons", *Phys. Lett.* **8** (1964) 214–215. doi:10.1016/S0031-9163(64)92001-3.

[9] O. W. Greenberg, "Spin and Unitary Spin Independence in a Paraquark Model of Baryons and Mesons", *Phys. Rev. Lett.* **13** (1964) 598–602. doi:10.1103/PhysRevLett.13.598.

[10] D. P. Barber et al., "Discovery of Three Jet Events and a Test of Quantum Chromodynamics at PETRA Energies", *Phys. Rev. Lett.* **43** (1979) 830. doi:10.1103/PhysRevLett.43.830.

[11] E. Fermi, "Tentativo di una teoria dei raggi β", *Il Nuovo Cimento (1924-1942)* **11** (1934), no. 1, 1.

[12] S. L. Glashow, "Partial Symmetries of Weak Interactions", *Nucl. Phys.* **22** (1961) 579–588. doi:10.1016/0029-5582(61)90469-2.

[13] A. Salam, "Weak and Electromagnetic Interactions", *Conf. Proc.* **C680519** (1968) 367–377.

[14] S. Weinberg, "A model of leptons", *Physical review letters* **19** (1967), no. 21, 1264.

[15] UA1 Collaboration, "Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV", *Phys. Lett.* **122B** (1983) 103–116. [,611(1983)]. doi:10.1016/0370-2693(83)91177-2.

[16] G. Arnison, A. Astbury, B. Aubert et al., "Experimental Observation of Lepton Pairs of Invariant Mass Around 95 GeV/c^2 at the CERN SPS Collider", *Physics Letters B* **126** (1983), no. 5, 398–410.

[17] P. Schleper, "Teilchenphysik für Fortgeschrittene, Notizen zur Vorlesung im Wintersemester 2010-11". 03, 2011.

[18] M. Kobayashi and T. Maskawa, "CP Violation in the Renormalizable Theory of Weak Interaction", *Prog. Theor. Phys.* **49** (1973) 652–657. doi:10.1143/PTP.49.652.

[19] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (1964) 508–509. doi:10.1103/PhysRevLett.13.508.

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

[21] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", arXiv:1207.7235v2. doi:10.1016/j.physletb.2012.08.021.

[22] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", arXiv:1207.7214v2. doi:10.1016/j.physletb.2012.08.020.

[23] Nobel Media, "The Nobel Prize in Physics 2013". Nobelprize.org, 2014. Web, 07.07.2017.

[24] G. Bertone, D. Hooper, and J. Silk, "Particle dark matter: Evidence, candidates and constraints", *Phys. Rept.* **405** (2005) 279–390, arXiv:hep-ph/0404175. doi:10.1016/j.physrep.2004.08.031.

[25] P. J. E. Peebles and B. Ratra, "The Cosmological constant and dark energy", *Rev. Mod. Phys.* **75** (2003) 559–606, arXiv:astro-ph/0207347. doi:10.1103/RevModPhys.75.559.

[26] Planck Collaboration, "Planck 2015 results. XIII. Cosmological parameters", *Astron. Astrophys.* **594** (2016) A13, arXiv:1502.01589. doi:10.1051/0004-6361/201525830.

[27] C.-S. Wu, E. Ambler, R. Hayward et al., "Experimental test of parity conservation in beta decay", *Physical review* **105** (1957), no. 4, 1413.

[28] M. Goldhaber, L. Grodzins, and A. Sunyar, "Helicity of neutrinos", *Physical Review* **109** (1958), no. 3, 1015.

[29] KamLAND Collaboration, "Reactor On-Off Antineutrino Measurement with KamLAND", *Phys. Rev.* **D88** (2013), no. 3, 033001, arXiv:1303.4667. doi:10.1103/PhysRevD.88.033001.

[30] Double Chooz Collaboration, "Improved measurements of the neutrino mixing angle θ_{13} with the Double Chooz detector", *JHEP* **10** (2014) 086, arXiv:1406.7763. [Erratum: JHEP02,074(2015)]. doi:10.1007/JHEP02(2015)074, 10.1007/JHEP10(2014)086.

[31] IceCube Collaboration, "Determining neutrino oscillation parameters from atmospheric muon neutrino disappearance with three years of IceCube DeepCore data", *Phys. Rev.* **D91** (2015), no. 7, 072004, arXiv:1410.7227. doi:10.1103/PhysRevD.91.072004.

[32] Daya Bay Collaboration, "New Measurement of Antineutrino Oscillation with the Full Detector Configuration at Daya Bay", *Phys. Rev. Lett.* **115** (2015), no. 11, 111802, arXiv:1505.03456. doi:10.1103/PhysRevLett.115.111802.

[33] F. R. Klinkhamer, "Neutrino mass and the Standard Model", *Mod. Phys. Lett.* **A28** (2013) 1350010, arXiv:1112.2669. doi:10.1142/S0217732313500107.

[34] U. Amaldi, W. de Boer, and H. Furstenau, "Comparison of grand unified theories with electroweak and strong coupling constants measured at LEP", *Phys. Lett.* **B260** (1991) 447–455. doi:10.1016/0370-2693(91)91641-8.

[35] J. F. Donoghue, "Introduction to the Effective Field Theory Description of Gravity", arXiv:gr-qc/9512024v1.

[36] N. Arkani-Hamed, S. Dimopoulos, and G. R. Dvali, "The Hierarchy problem and new dimensions at a millimeter", *Phys. Lett.* **B429** (1998) 263–272, arXiv:hep-ph/9803315. doi:10.1016/S0370-2693(98)00466-3.

[37] S. Dimopoulos and G. Landsberg, "Black Holes at the LHC", arXiv:hep-ph/0106295v1. doi:10.1103/PhysRevLett.87.161602.

[38] CMS Collaboration, "Search for Black Holes with Early Run 2 Data", Technical Report CMS-PAS-EXO-15-007, CERN, Geneva, 2015.

[39] CMS Collaboration, "Search for high-mass resonances and quantum black holes in the $e\mu$ final state in proton-proton collisions at $\sqrt{s} = 13$ TeV", Technical Report CMS-PAS-EXO-16-001, CERN, Geneva, 2016.

[40] P. Langacker, R. W. Robinett, and J. L. Rosner, "New Heavy Gauge Bosons in p p and p anti-p Collisions", *Phys. Rev.* **D30** (1984) 1470. doi:10.1103/PhysRevD.30.1470.

[41] G. Altarelli, B. Mele, and M. Ruiz-Altaba, "Searching for New Heavy Vector Bosons in $p\bar{p}$ Colliders", *Z. Phys.* **C45** (1989) 109. [Erratum: Z. Phys.C47,676(1990)]. doi:10.1007/BF01552335, 10.1007/BF01556677.

[42] D. Muller and S. Nandi, "A separate SU(2) for the third family: Topflavor", arXiv:hep-ph/9607328v1. doi:10.1016/S0920-5632(96)00560-9.

[43] E. Malkawi, T. Tait, and C. P. Yuan, "A Model of Strong Flavor Dynamics for the Top Quark", arXiv:hep-ph/9603349v1. doi:10.1016/0370-2693(96)00859-3.

[44] ATLAS Collaboration, "Search for $W' \to t\bar{b}$ in the lepton plus jets final state in proton-proton collisions at a centre-of-mass energy of $\sqrt{s} = 8$ TeV with the ATLAS detector", arXiv:1410.4103v4. doi:10.1016/j.physletb.2015.02.051.

[45] CMS Collaboration, "Searches for W' bosons decaying to a top quark and a bottom quark in proton-proton collisions at 13 TeV", arXiv:1706.04260v1.

[46] CMS Collaboration, "Search for W' boson resonances decaying into a top quark and a bottom quark in lepton + jet final states at $\sqrt{s} = 13$ TeV", Technical Report CMS-PAS-B2G-17-010, CERN, Geneva, 2017.

[47] E. Rutherford, "The scattering of alpha and beta particles by matter and the structure of the atom", *Phil. Mag. Ser.6* **21** (1911) 669–688. doi:10.1080/14786440508637080.

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

[49] R. Ley, "CERN Accelerators". http://ps-div.web.cern.ch/ps-div/PSComplex/accelerators.pdf, 2001.

[50] J.-L. Caron, "LHC Layout". https://cds.cern.ch/record/841573, 1997.

[51] C. De Melis, "The CERN accelerator complex". https://cds.cern.ch/record/2197559, 2016.

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

[53] CMS Collaboration, S. R. Davis, "CMS Slice". https://cds.cern.ch/record/2204899, 2016.

[54] CMS Collaboration, "Aligning the CMS Muon Chambers with the Muon Alignment System during an Extended Cosmic Ray Run", arXiv:0911.4770v2. doi:10.1088/1748-0221/5/03/T03019.

[55] CMS Collaboration, V. Karimäki, M. Mannelli, P. Siegrist et al., "The CMS tracker system project: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 1997.

[56] CMS Collaboration, "The CMS electromagnetic calorimeter project: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 1997.

[57] CMS Collaboration, "Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at sqrt(s) = 7 TeV", arXiv:1306.2016v2. doi:10.1088/1748-0221/8/09/P09009.

[58] CMS Collaboration, "The CMS hadron calorimeter project: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 1997.

[59] CMS Collaboration, "The CMS magnet project: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 1997.

[60] CMS Collaboration, "The CMS muon project: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 1997.

[61] CMS Collaboration, "Performance of CMS muon reconstruction in pp collision events at sqrt(s) = 7 TeV", arXiv:1206.4071v2. doi:10.1088/1748-0221/7/10/P10002.

[62] CMS Collaboration, "The CMS trigger system", arXiv:1609.02366v2. doi:10.1088/1748-0221/12/01/P01020.

[63] CERN, "The Grid: A system of tiers". https://cds.cern.ch/record/1997396, 08, 2012.

[64] WLCG, "Tier centres". http://wlcg-public.web.cern.ch/tier-centres.

[65] III. Physikalisches Institut B, RWTH Aachen University, “Grid-Computing”. http://www.institut3b.physik.rwth-aachen.de/cms/ParticlePhysics3B/Forschung/~gbvf/GRID-Computing/, 01, 2016. Accessed: 18.06.2017.

[66] T. Hebbeker, “A Global Comparison between L3 Data and Standard Model Monte Carlo - a first attempt”. L3 note 2305, 07, 1998. https://web.physik.rwth-aachen.de/~hebbeker/l3note_2305.pdf.

[67] P. Biallass, “Model Independent Search for Deviations from the Standard Model at the Tevatron”, Diploma Thesis, III. Physikalisches Institut A, RWTH Aachen University, 12, 2004.

[68] S. A. Schmitz, “Model Unspecific Search for New Physics with High pT Photons in CMS”, Diploma Thesis, III. Physikalisches Institut A, RWTH Aachen University, 10, 2009.

[69] C. Hof, “Implementation of a Model-Independent Search for New Physics with the CMS Detector exploiting the World-Wide LHC Computing Grid”. PhD thesis, III. Physikalisches Institut A, RWTH Aachen University, 10, 2009.

[70] E. Dietz-Laursonn, “Model Unspecific Search for New Physics with b-Hadrons in CMS”, Diploma Thesis, III. Physikalisches Institut A, RWTH Aachen University, 10, 2010.

[71] M. Olschewski, “Study of Alternative Statistical Methods for a Model Unspecific Search in CMS”, Diploma Thesis, III. Physikalisches Institut A, RWTH Aachen University, 07, 2011.

[72] M. Brodski, “Model Unspecific Search in CMS with tau Leptons”, Master’s thesis, III. Physikalisches Institut A, RWTH Aachen University, 09, 2012.

[73] H. Pieta, “MUSiC - A Model Unspecific Search in CMS based on 2010 LHC data”. PhD thesis, III. Physikalisches Institut A, RWTH Aachen University, 2012.

[74] P. Papacz, “Model Unspecific Search for new Physics in CMS Based on 2011 Data”. PhD thesis, III. Physikalisches Institut A, RWTH Aachen University, 05, 2014.

[75] A. A. E. Albert, “Extension of the Model Unspecific Search in CMS to Final States with Jets using 2012 Data”, Master’s thesis, III. Physikalisches Institut A, RWTH Aachen University, 11, 2015.

[76] J. Roemer, "Model Unspecific Search for New Physics with pp Collisions at $\sqrt{s} = 13$ TeV with the CMS Experiment", Master's thesis, III. Physikalisches Institut A, RWTH Aachen University, 02, 2017.

[77] CMS Collaboration, "MUSiC, a Model Unspecific Search for New Physics, in pp Collisions at $\sqrt{s} = 8\,\mathrm{TeV}$", Technical Report CMS-PAS-EXO-14-016, CERN, Geneva, 2017.

[78] S. Knutzen, "A software for the reinterpretation of model independent search results and constraining theories beyond the Standard Model". PhD thesis, III. Physikalisches Institut A, RWTH Aachen University, 2017.

[79] D. Durchardt, "MUSiC: A Model Unspecific Search for New Physics Based on CMS Data at $\sqrt{s} = 8$ TeV". PhD thesis, III. Physikalisches Institut A, RWTH Aachen University, 2017.

[80] ATLAS Collaboration, "A general search for new phenomena with the ATLAS detector in pp collisions at sort(s)=7 TeV.", Technical Report ATLAS-CONF-2012-107, CERN, Geneva, Aug, 2012.

[81] ATLAS Collaboration, "A general search for new phenomena with the ATLAS detector in pp collisions at $\sqrt{s} = 8$ TeV", Technical Report ATLAS-CONF-2014-006, CERN, Geneva, Mar, 2014.

[82] ATLAS Collaboration, "A model independent general search for new phenomena with the ATLAS detector at $\sqrt{s} = 13$ TeV", Technical Report ATLAS-CONF-2017-001, CERN, Geneva, Jan, 2017.

[83] CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET", Technical Report CMS-PAS-PFT-09-001, CERN, Geneva, Apr, 2009.

[84] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm", arXiv:0802.1189v2. doi:10.1088/1126-6708/2008/04/063.

[85] CMS Collaboration, "CMS TWiki: Jet Resolution". https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution?rev=57#Smearing_procedures, 05, 2017.

[86] CMS Collaboration, "CMS TWiki: Jet Identification". https://twiki.cern.ch/twiki/bin/view/CMS/JetID?rev=98#Recommendations_for_13_TeV_2015, 12, 2016.

[87] CMS Collaboration, "CMS TWiki: Muon Identification". https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonIdRun2?rev=28#HighPt_Muon, 01,

2017.

[88] CMS Collaboration, "CMS TWiki: HEEP Electron ID and isolation". https://twiki.cern.ch/twiki/bin/view/CMS/HEEPElectronID?rev=68, 06, 2015.

[89] CMS Collaboration, "CMS TWiki: Cut Based Photon ID for Run 2". https://twiki.cern.ch/twiki/bin/view/CMS/CutBasedPhotonIdentificationRun2?rev=35#SPRING15_selections_25_ns, 12, 2016.

[90] CMS Collaboration, "Identification of b quark jets at the CMS Experiment in the LHC Run 2", Technical Report CMS-PAS-BTV-15-001, CERN, Geneva, 2016.

[91] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", arXiv:1211.4462v2. doi:10.1088/1748-0221/8/04/P04013.

[92] CMS Collaboration, "Performance of b-tagging algorithms at 13 TeV", Technical Report CMS-AN-2016-036, CERN.

[93] CMS Collaboration, "CMS luminosity measurement for the 2015 data-taking period", Technical Report CMS-PAS-LUM-15-001, CERN, Geneva, 2017.

[94] CMS Collaboration, "Performance of flavour tagging algorithms in 2016 data at 13 TeV", Technical Report CMS-AN-2017-018, CERN.

[95] S. Frixione and B. R. Webber, "Matching NLO QCD computations and parton shower simulations", arXiv:hep-ph/0204244v2. doi:10.1088/1126-6708/2002/06/029.

[96] H. Abdi, "Bonferroni and Šidák corrections for multiple comparisons", *Encyclopedia of measurement and statistics* **3** (2007) 103–107.

[97] G. Petrucciani, A. Rizzi, C. Vuosalo et al., "Mini-AOD: A New Analysis Data Format for CMS", arXiv:1702.04685v2. doi:10.1088/1742-6596/664/7/072052.

[98] H. P. Bretz, M. Brodski, M. Erdmann et al., "A Development Environment for Visual Physics Analysis", arXiv:1205.4912v2. doi:10.1088/1748-0221/7/08/T08005.

[99] R. Brun and F. Rademakers, "ROOT: An object oriented data analysis framework", *Nucl. Instrum. Meth.* **A389** (1997) 81–86. doi:10.1016/S0168-9002(97)00048-X.

[100] T. E. Oliphant, "Python for Scientific Computing", *Computing in Science Engineering* **9** (May, 2007) 10–20. doi:10.1109/MCSE.2007.58.

[101] E. Bernhardsson, E. Freider, and A. Rouhani, "Github: Luigi". https://github.com/spotify/luigi/tree/2.4.0, 12, 2016.

[102] S. V. D. Walt, S. C. Colbert, and G. Varoquaux, "The NumPy array: a structure for efficient numerical computation", arXiv:1102.1523v1. doi:10.1109/MCSE.2011.37.

[103] E. Jones, T. Oliphant, P. Peterson et al., "SciPy: Open source scientific tools for Python". http://www.scipy.org/, 2001–.

[104] J. D. Hunter, "Matplotlib: A 2D graphics environment", *Computing In Science & Engineering* **9** (2007), no. 3, 90–95. doi:10.1109/MCSE.2007.55.

[105] B. Webber, "Parton shower Monte Carlo event generators", *Scholarpedia* **6** (2011), no. 12, 10662. revision #128236. doi:10.4249/scholarpedia.10662.

[106] CMS Collaboration, "Generator Validation". https://twiki.cern.ch/twiki/bin/view/CMS/GeneratorValidation?rev=9, 04, 2017.

[107] J. Alwall, M. Herquet, F. Maltoni et al., "MadGraph 5 : Going Beyond", arXiv:1106.0522v1. doi:10.1007/JHEP06(2011)128.

[108] T. Gleisberg, S. Hoeche, F. Krauss et al., "Event generation with SHERPA 1.1", arXiv:0811.4622v1. doi:10.1088/1126-6708/2009/02/007.

[109] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method", arXiv:0709.2092v1. doi:10.1088/1126-6708/2007/11/070.

[110] S. Alioli, P. Nason, C. Oleari et al., "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", arXiv:1002.2581v1. doi:10.1007/JHEP06(2010)043.

[111] J. Alwall, R. Frederix, S. Frixione et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", arXiv:1405.0301v2. doi:10.1007/JHEP07(2014)079.

[112] J. M. Campbell, R. K. Ellis, and C. Williams, "Vector boson pair production at the LHC", arXiv:1105.0020v1. doi:10.1007/JHEP07(2011)018.

[113] T. Sjöstrand, S. Mrenna, and P. Skands, "A Brief Introduction to PYTHIA 8.1", arXiv:0710.3820v1. doi:10.1016/j.cpc.2008.01.036.

[114] S. Agostinelli, J. Allison, K. a. Amako et al., "GEANT4—a simulation toolkit", *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** (2003), no. 3, 250–303.

[115] CMS Collaboration, "Search for Type-III Seesaw Heavy Fermions with Multilepton Final States using 2.3/fb of 13 TeV proton-proton Collision Data", Technical Report CMS-PAS-EXO-16-002, CERN, Geneva, 2016.

[116] D. M. Gingrich, "Monte Carlo event generator for black hole production and decay in proton-proton collisions", arXiv:0911.5370v4. doi:10.1016/j.cpc.2010.07.027.

[117] D.-C. Dai, G. Starkman, D. Stojkovic et al., "BlackMax: A black-hole event generator with rotation, recoil, split branes and brane tension", arXiv:0711.3012v4. doi:10.1103/PhysRevD.77.076007.

[118] A. Belyaev, N. D. Christensen, and A. Pukhov, "CalcHEP 3.4 for collider physics within and beyond the Standard Model", arXiv:1207.6082v2. doi:10.1016/j.cpc.2013.01.014.

[119] G. Cowan, "Statistics for Searches at the LHC", arXiv:1307.2487v1.

[120] P. Pirinen, "Statistical power sum analysis for nonidentically distributed correlated lognormal signals", in *The 2003 Finnish Signal Processing Symposium (FINSIG'03), Tampere, Finland*, pp. 254–258. 2003.

[121] S. C. Schwartz and Y.-S. Yeh, "On the Distribution Function and Moments of Power Sums With Log-Normal Components", *Bell Labs Technical Journal* **61** (1982), no. 7, 1441–1462.

[122] N. Mehta, A. Molisch, J. Wu et al., "Approximating the Sum of Correlated Lognormal or Lognormal-Rice Random Variables", in *IEEE International Conference on Communications (ICC)*, volume 4, pp. 1605–1610. June, 2006.

[123] G. Cowan, "Statistical Methods for Discovery and Limits in HEP Experiments". 06, 2011. https://www.pp.rhul.ac.uk/~cowan/stat/freiburg/cowan_freiburg_2011_1.pdf.

[124] G. Cowan, "Statistical data analysis". Oxford University Press, 1998.

[125] B. P. Flannery, W. Press, S. Teukolsky et al., "Numerical recipes in FORTRAN: The art of scientific computing", *Cambridge University Press* (1992) 704–716.

[126] D. A. Darling, "The kolmogorov-smirnov, cramer-von mises tests", *The Annals of Mathematical Statistics* **28** (1957), no. 4, 823–838.

[127] M. A. Stephens, "EDF statistics for goodness of fit and some comparisons", *Journal of the American Statistical Association* **69** (1974), no. 347, 730–737.

[128] F. W. Scholz and M. A. Stephens, "K-sample Anderson–Darling tests", *Journal of the American Statistical Association* **82** (1987), no. 399, 918–924.

[129] CMS Collaboration, "Search for evidence of Type-III seesaw mechanism in multilepton final states in pp collisions at $\sqrt{s} = 13$ TeV", Technical Report CMS-PAS-EXO-17-006, CERN, Geneva, 2017.

[130] CMS Collaboration, "Search for lepton flavour violating decays of heavy resonances and quantum black holes to an eμ pair in proton-proton collisions at $\sqrt{s} = 13$ TeV", Technical Report CMS-PAS-EXO-16-058, CERN, Geneva. in preparation for publication.

[131] M. Lamont, "LHC Commissioning - Longer term LHC schedule". `https://lhc-commissioning.web.cern.ch/lhc-commissioning/schedule/LHC-long-term.htm`, 01, 2016.

Glossary of Terms

G

ADD	Arkani-Hamed-Dimopoulos-Dvali model
ALICE	A Lead Ion Collider Experiment
ATLAS	ATLAS experiment (orig. A Toroidal LHC Apparatus)
BH	black hole
CERN	European Organization for Nuclear Research (orig. Conseil Européen pour la Recherche Nucléaire)
CKM	Cabibbo-Maskawa-Kobayashi matrix
CMS	Compact Muon Solenoid experiment
CSC	cathode strip chamber
CSVv2	Combined Secondary Vertex version 2 algorithm
DESY	Deutsches Elektronen-Synchrotron
DT	drift tube
ECAL	electromagnetic calorimeter
HCAL	hadron calorimeter
HEEP	high energy electron pairs selection criteria
HLT	high level trigger
L1	level-1 trigger
LEP	Large Electron Positron Collider
LHC	Large Hadron Collider
LHCb	Large Hadron Collider beauty experiment
LINAC	linear accelerator
LUT	lookup table
MC	Monte Carlo simulation
MUSiC	Model Unspecific Search in CMS
PF	Particle-Flow algorithm
PS	Proton Synchrotron
PSB	Proton Synchrotron Booster
QBH	quantum black hole
QCD	quantum chromodynamics
QED	quantum electrodynamics
QFD	quantum flavordynamics
RoI	region of interest
RPC	resistive plate chamber
RS	Randall-Sundrum model
SM	Standard Model
SPS	Super Proton Synchrotron
SSM	Sequential Standard Model
WLCG	Worldwide LHC Computing Grid project

Appendix

A

A.1 Monte Carlo Datasets

The following pages list the MC samples used within the analysis. The column "Group" indicates the process group. All cross section uncertainties within a process group are assumed to be correlated. The second column, "Dataset Name" lists the CMS internal dataset name, which also gives information about the simulated physics process. The third column shows the cross section in pb. If a higher order correction factor has been applied, it is listed in the column denoted by k-factor. This is also visible in the "Order" column, which shows the perturbative calculation order before and, where applicable, after application of the k-factor. Possible values are "LO" (leading order), "NLO" (next to leading order), "NNLO" (next to next to leading order). Another possible value is present only for processes containing the W boson and additional jets, "NLO_W", which corresponds to the next to leading order but is treated with a slightly larger uncertainty due to QCD effects. Finally, the filter efficiency indicates that the samples have already undergone some filtering. The numbers have to be combined into a total effective cross section for each sample by multiplying the indicated cross section, k-factor and filter efficiency.

A.1.1 Standard Model Samples

Process	Dataset Name	Cross Section (pb)	k-factor	Order	Filter Eff.	Events
DrellYan	DYJetsToLL_M-5to50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	71 310.000	1.23	LO → NLO	1.00	8 771 481
	DYJetsToLL_M-5to50_HT-100to200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	224.200	1.23	LO → NLO	1.00	1 018 459
	DYJetsToLL_M-5to50_HT-100to200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 (ext1)	224.200	1.23	LO → NLO	1.00	8 145 474
	DYJetsToLL_M-5to50_HT-200to400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	37.190	1.23	LO → NLO	1.00	1 025 157
	DYJetsToLL_M-5to50_HT-200to400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 (ext1)	37.190	1.23	LO → NLO	1.00	2 020 586
	DYJetsToLL_M-5to50_HT-400to600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	3.581	1.23	LO → NLO	1.00	1 012 963
	DYJetsToLL_M-5to50_HT-600toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1.124	1.23	LO → NLO	1.00	1 019 394
	DYJetsToLL_M-5to50_HT-600toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 (ext1)	1.124	1.23	LO → NLO	1.00	1 924 359
	DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	4 895.000	1.23	LO → NNLO	1.00	8 877 130
	DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 (ext1)	4 895.000	1.18	LO → NNLO	1.00	240 499 997
	DYJetsToLL_M-50_HT-100to200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	147.400	1.23	LO → NLO	1.00	2 564 547
	DYJetsToLL_M-50_HT-100to200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 (ext1)	147.400	1.23	LO → NLO	1.00	7 829 684
	DYJetsToLL_M-50_HT-200to400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	42.750	1.44	LO → NLO	1.00	962 195
	DYJetsToLL_M-50_HT-400to600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	5.678	1.23	LO → NLO	1.00	1 069 003
	DYJetsToLL_M-50_HT-400to600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 (ext1)	5.678	1.23	LO → NLO	1.00	3 431 833
	DYJetsToLL_M-50_HT-600toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	2.210	1.14	LO → NLO	1.00	1 031 103
	DYJetsToLL_M-50_HT-600toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 (ext1)	2.210	1.14	LO → NLO	1.00	4 058 472
	ZToEE_NNPDF30_13TeV-powheg_M_120_200	19.320	-	NLO	1.00	100 000
	ZToEE_NNPDF30_13TeV-powheg_M_200_400	2.731	-	NLO	1.00	100 000
	ZToEE_NNPDF30_13TeV-powheg_M_400_800	0.241	-	NLO	1.00	100 000
	ZToEE_NNPDF30_13TeV-powheg_M_800_1400	0.017	-	NLO	1.00	100 000
	ZToEE_NNPDF30_13TeV-powheg_M_1400_2300	0.001	-	NLO	1.00	100 000
	ZToEE_NNPDF30_13TeV-powheg_M_2300_3500	0.000	-	NLO	1.00	100 000
	ZToEE_NNPDF30_13TeV-powheg_M_3500_4500	0.000	-	NLO	1.00	100 000
	ZToEE_NNPDF30_13TeV-powheg_M_4500_6000	0.000	-	NLO	1.00	99 000
	ZToEE_NNPDF30_13TeV-powheg_M_6000_Inf	0.000	-	NLO	1.00	97 600
	ZToMuMu_NNPDF30_13TeV-powheg_M_120_200	19.320	-	NLO	1.00	100 000
	ZToMuMu_NNPDF30_13TeV-powheg_M_200_400	2.731	-	NLO	1.00	100 000
	ZToMuMu_NNPDF30_13TeV-powheg_M_400_800	0.241	-	NLO	1.00	99 600
	ZToMuMu_NNPDF30_13TeV-powheg_M_800_1400	0.017	-	NLO	1.00	97 600
	ZToMuMu_NNPDF30_13TeV-powheg_M_1400_2300	0.001	-	NLO	1.00	99 200
	ZToMuMu_NNPDF30_13TeV-powheg_M_2300_3500	0.000	-	NLO	1.00	100 000
	ZToMuMu_NNPDF30_13TeV-powheg_M_3500_4500	0.000	-	NLO	1.00	100 000
	ZToMuMu_NNPDF30_13TeV-powheg_M_4500_6000	0.000	-	NLO	1.00	100 000
	ZToMuMu_NNPDF30_13TeV-powheg_M_6000_Inf	0.000	-	NLO	1.00	99 200
Gamma	GJets_HT-40To100_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	20 790.000	-	LO	1.00	4 424 830

Process	Dataset Name	Cross Section (pb)	k-factor	Order	Filter Eff.	Events
	GJets_HT-100To200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	9 238.000	-	LO	1.00	4 877 095
	GJets_HT-200To400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	2 305.000	-	LO	1.00	10 467 654
	GJets_HT-400To600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	274.400	-	LO	1.00	2 406 285
	GJets_HT-600ToInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	93.460	-	LO	1.00	2 456 253
QCD	QCD_Pt-50to80_EMEnriched_TuneCUETP8M1_13TeV_pythia8	19 800 000.000	-	LO	0.15	22 463 814
	QCD_Pt-50to80_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8	19 222 500.000	-	LO	0.02	20 378 392
	QCD_Pt-80to120_EMEnriched_TuneCUETP8M1_13TeV_pythia8	2 800 000.000	-	LO	0.13	36 029 628
	QCD_Pt-80to120_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8	2 758 420.000	-	LO	0.04	13 749 024
	QCD_Pt-120to170_EMEnriched_TuneCUETP8M1_13TeV_pythia8	477 000.000	-	LO	0.13	36 202 409
	QCD_Pt-120to170_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8	469 797.000	-	LO	0.05	7 971 018
	QCD_Pt-170to300_EMEnriched_TuneCUETP8M1_13TeV_pythia8	114 000.000	-	LO	0.17	11 518 871
	QCD_Pt-170to300_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8	117 989.000	-	LO	0.07	7 910 182
	QCD_Pt-300to470_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8	7 820.250	-	LO	0.10	7 845 620
	QCD_Pt-300toInf_EMEnriched_TuneCUETP8M1_13TeV_pythia8	9 000.000	-	LO	0.15	7 212 550
	QCD_Pt-470to600_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8	645.528	-	LO	0.12	3 841 262
	QCD_Pt-600to800_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8	187.109	-	LO	0.13	3 984 898
	QCD_Pt-800to1000_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8	32.349	-	LO	0.15	3 666 110
	QCD_Pt-1000toInf_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8	10.431	-	LO	0.16	3 938 782
TG	TGJets_TuneCUETP8M1_13TeV_amcatnlo_madspin_pythia8	2.967	-	NLO	1.00	280 100
	TGJets_TuneCUETP8M1_13TeV_amcatnlo_madspin_pythia8 (ext1)	2.967	-	NLO	1.00	1 556 996
TTBar	TT_TuneCUETP8M1_13TeV-powheg-pythia8 (ext3)	730.000	1.14	NLO → NNLO	1.00	97 994 442
	TT_TuneCUETP8M1_13TeV-powheg-pythia8 (ext4)	730.000	1.14	NLO → NNLO	1.00	187 626 200
	TT_Mtt-700to1000_TuneCUETP8M1_13TeV-powheg-pythia8	730.000	1.14	NLO → NNLO	0.09	37 285 071
	TT_Mtt-700to1000_TuneCUETP8M1_13TeV-powheg-pythia8 (ext1)	730.000	1.14	NLO → NNLO	0.09	3 086 786
	TT_Mtt-1000toInf_TuneCUETP8M1_13TeV-powheg-pythia8 (ext1)	730.000	1.14	NLO → NNLO	0.02	1 782 160
	TT_Mtt-1000toInf_TuneCUETP8M1_13TeV-powheg-pythia8 (ext2)	730.000	1.14	NLO → NNLO	0.02	25 182 267
TTG	TTGJets_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8	3.697	-	NLO	1.00	4 874 116
TTGG	TTGG_0Jets_TuneCUETP8M1_13TeV_amcatnlo_madspin_pythia8	0.017	-	NLO	1.00	1 466 516
TTW	ttWJets_13TeV_madgraphMLM	0.243	-	LO	1.00	12 889 768
TTZ	ttZJets_13TeV_madgraphMLM	0.259	-	LO	1.00	22 157 947
TZQ	tZq_ll_4f_13TeV-amcatnlo-pythia8_TuneCUETP8M1	0.076	-	NLO	1.00	2 996 000
	tZq_nunu_4f_13TeV-amcatnlo-pythia8_TuneCUETP8M1	0.138	-	NLO	1.00	984 600
Top	ST_s-channel_4f_leptonDecays_13TeV-amcatnlo-pythia8_TuneCUETP8M1	3.360	-	NLO	1.00	998 400
	ST_t-channel_antitop_4f_leptonDecays_13TeV-powheg-pythia8_TuneCUETP8M1	25.300	-	NLO	1.00	1 630 900
	ST_t-channel_top_4f_leptonDecays_13TeV-powheg-pythia8_TuneCUETP8M1	41.900	-	NLO	1.00	3 299 200

Process	Dataset Name	Cross Section (pb)	k-factor	Order	Filter Eff.	Events
W	WJetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	50 690.000	1.21	LO → NNLO	1.00	47 161 328
	WJetsToLNu_HT-100To200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1 345.000	1.21	LO → NLO_W	1.00	10 205 377
	WJetsToLNu_HT-100To200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 (ext1)	1 345.000	1.21	LO → NLO_W	1.00	28 484 673
	WJetsToLNu_HT-200To400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	359.700	1.21	LO → NLO_W	1.00	4 692 696
	WJetsToLNu_HT-200To400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 (ext1)	359.700	1.21	LO → NLO_W	1.00	14 881 357
	WJetsToLNu_HT-400To600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	48.910	1.21	LO → NLO_W	1.00	1 943 664
	WJetsToLNu_HT-600To800_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	12.050	1.21	LO → NLO_W	1.00	3 767 766
	WJetsToLNu_HT-800To1200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	5.501	1.21	LO → NLO_W	1.00	1 568 277
	WJetsToLNu_HT-1200To2500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1.329	1.21	LO → NLO_W	1.00	246 239
	WJetsToLNu_HT-1200To2500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 (ext1)	1.329	1.21	LO → NLO_W	1.00	6 709 861
	WJetsToLNu_HT-2500ToInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	0.032	1.21	LO → NLO_W	1.00	251 982
	WJetsToLNu_HT-2500ToInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 (ext1)	0.032	1.21	LO → NLO_W	1.00	2 300 696
	WJetsToQQ_HT180_13TeV-madgraphMLM-pythia8	2 788.000	-	LO	1.00	22 689 606
	WJetsToQQ_HT-600ToInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	95.140	-	LO	1.00	1 025 100
	WToENu_M-200_TuneCUETP8M1_13TeV-pythia8	6.236	1.34	LO → NNLO	1.00	998 887
	WToENu_M-500_TuneCUETP8M1_13TeV-pythia8	0.214	1.33	LO → NNLO	1.00	993 370
	WToENu_M-1000_TuneCUETP8M1_13TeV-pythia8	0.013	1.33	LO → NNLO	1.00	998 728
	WToENu_M-2000_TuneCUETP8M1_13TeV-pythia8	0.001	1.26	LO → NNLO	1.00	998 096
	WToENu_M-3000_TuneCUETP8M1_13TeV-pythia8	0.000	1.14	LO → NNLO	1.00	995 350
	WToENu_M-4000_TuneCUETP8M1_13TeV-pythia8	0.000	0.45	LO → NNLO	1.00	993 166
	WToMuNu_M-200_TuneCUETP8M1_13TeV-pythia8	6.236	1.29	LO → NNLO	1.00	993 140
	WToMuNu_M-500_TuneCUETP8M1_13TeV-pythia8	0.214	1.27	LO → NNLO	1.00	997 511
	WToMuNu_M-1000_TuneCUETP8M1_13TeV-pythia8	0.013	1.26	LO → NNLO	1.00	992 260
	WToMuNu_M-2000_TuneCUETP8M1_13TeV-pythia8	0.001	1.17	LO → NNLO	1.00	994 764
	WToMuNu_M-3000_TuneCUETP8M1_13TeV-pythia8	0.000	1.04	LO → NNLO	1.00	996 545
	WToMuNu_M-4000_TuneCUETP8M1_13TeV-pythia8	0.000	0.41	LO → NNLO	1.00	992 708
	WToTauNu_M-200_TuneCUETP8M1_13TeV-pythia8	6.236	1.35	LO → NNLO	1.00	991 772
	WToTauNu_M-500_TuneCUETP8M1_13TeV-pythia8-tauola	0.224	-	LO	1.00	977 384
	WToTauNu_M-1000_TuneCUETP8M1_13TeV-pythia8	0.013	1.33	LO → NNLO	1.00	999 316
	WToTauNu_M-2000_TuneCUETP8M1_13TeV-pythia8-tauola	0.001	-	LO	1.00	977 400
WG	WGToLNuG_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	405.271	1.21	LO → NLO	1.00	6 102 260
	WGToLNuG_PtG-500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	0.012	-	LO	1.00	1 390 140
WW	WWTo2L2Nu_13TeV-powheg	10.480	1.16	NLO → NNLO	1.00	1 979 988
	WWTo2L2Nu_Mll_200To600_13TeV-powheg	10.481	1.16	NLO → NNLO	0.11	200 000
	WWTo2L2Nu_Mll_600To1200_13TeV-powheg	10.481	1.16	NLO → NNLO	< 0.01	200 000
	WWTo2L2Nu_Mll_1200To2500_13TeV-powheg	10.481	1.16	NLO → NNLO	< 0.01	198 400
	WWTo2L2Nu_Mll_2500ToInf_13TeV-powheg	10.481	1.16	NLO → NNLO	< 0.01	38 969
	WWTo4Q_13TeV-powheg	45.200	1.14	NLO → NNLO	1.00	2 000 000

Process	Dataset Name	Cross Section (pb)	k-factor	Order	Filter Eff.	Events
	WWToLNuQQ_13TeV-powheg	43.530	1.15	NLO → NNLO	1.00	1 924 400
	WWToLNuQQ_13TeV-powheg (ext1)	43.530	1.15	NLO → NNLO	1.00	6 996 000
	WW_DoubleScattering_13TeV-pythia8	1.640	-	LO	1.00	844 954
WWG	WWG_TuneCUETP8M1_13TeV-amcatnlo-pythia8 (ext1)	0.215	-	NLO	1.00	995 400
WWW	WWW_4F_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.209	-	NLO	1.00	240 000
WWZ	WWZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.165	-	NLO	1.00	248 400
WZ	WZTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin_pythia8	10.710	-	NLO	1.00	19 742 520
	WZTo1L3Nu_13TeV_amcatnloFXFX_madspin_pythia8	3.033	-	NLO	1.00	1 703 772
	WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	5.595	-	NLO	1.00	25 704 656
	WZTo3LNu_TuneCUETP8M1_13TeV-powheg-pythia8	4.430	-	NLO	1.00	2 000 000
WZG	WZG_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.041	-	NLO	1.00	997 400
WZZ	WZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.056	-	NLO	1.00	249 800
ZToInvisible	ZJetsToNuNu_HT-100To200_13TeV-madgraph	280.470	1.63	LO → NLO	1.00	5 240 199
	ZJetsToNuNu_HT-200To400_13TeV-madgraph	78.360	1.62	LO → NLO	1.00	2 170 249
	ZJetsToNuNu_HT-200To400_13TeV-madgraph (ext1)	78.360	1.62	LO → NLO	1.00	19 007 186
	ZJetsToNuNu_HT-400To600_13TeV-madgraph	10.944	1.46	LO → NLO	1.00	954 435
	ZJetsToNuNu_HT-600ToInf_13TeV-madgraph	4.203	1.39	LO → NLO	1.00	1 033 818
ZZ	ZZTo2L2Nu_13TeV_powheg_pythia8	0.564	-	NLO	1.00	8 785 050
	ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	3.220	-	NLO	1.00	15 301 695
	ZZTo2Q2Nu_13TeV_amcatnloFXFX_madspin_pythia8	4.040	-	NLO	1.00	30 806 034
	ZZTo4L_13TeV_powheg_pythia8	1.256	-	NLO	1.00	6 669 188
	ZZTo4Q_13TeV_amcatnloFXFX_madspin_pythia8	6.842	-	NLO	1.00	30 707 868
ZZZ	ZZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.014	-	NLO	1.00	250 000

A.1.2 Signal Samples

Dataset Name	Cross Section (pb)	k-factor	Order	Filter Eff.	Events
BlackHole_BH1_MD-4000_MBH-5000_n-6_TuneCUETP8M1_13TeV-blackmax	6.700	-	LO	1.00	10 000
BlackHole_BH1_MD-4000_MBH-6000_n-6_TuneCUETP8M1_13TeV-blackmax	1.230	-	LO	1.00	10 000
BlackHole_BH1_MD-4000_MBH-7000_n-6_TuneCUETP8M1_13TeV-blackmax	0.185	-	LO	1.00	10 000
BlackHole_BH1_MD-4000_MBH-8000_n-6_TuneCUETP8M1_13TeV-blackmax	0.021	-	LO	1.00	10 000
BlackHole_BH1_MD-4000_MBH-9000_n-6_TuneCUETP8M1_13TeV-blackmax	0.002	-	LO	1.00	10 000
BlackHole_BH1_MD-4000_MBH-10000_n-6_TuneCUETP8M1_13TeV-blackmax	0.000	-	LO	1.00	10 000
QBHToEMu_M-2000_n4_ADD_TuneCUETP8M1_13TeV-QBH-pythia8	0.744	-	LO	1.00	11 706
QBHToEMu_M-3000_n4_ADD_TuneCUETP8M1_13TeV-QBH-pythia8	0.042	-	LO	1.00	10 240
QBHToEMu_M-4000_n4_ADD_TuneCUETP8M1_13TeV-QBH-pythia8	0.003	-	LO	1.00	9 437
QBHToEMu_M-5000_n4_ADD_TuneCUETP8M1_13TeV-QBH-pythia8	0.000	-	LO	1.00	8 898
SeesawTypeIII_SIGMA-SIGMA0HH_M-380_13TeV-madgraph	0.001	1.44	LO → NLO	0.49	102 022
SeesawTypeIII_SIGMA-SIGMA0HW_M-380_13TeV-madgraph	0.003	1.44	LO → NLO	0.77	100 340
SeesawTypeIII_SIGMA-SIGMA0HZ_M-380_13TeV-madgraph	0.002	1.44	LO → NLO	0.36	102 481
SeesawTypeIII_SIGMA-SIGMA0WH_M-380_13TeV-madgraph	0.003	1.44	LO → NLO	0.12	100 575
SeesawTypeIII_SIGMA-SIGMA0WW_M-380_13TeV-madgraph	0.009	1.44	LO → NLO	0.25	202 275
SeesawTypeIII_SIGMA-SIGMA0WZ_M-380_13TeV-madgraph	0.004	1.44	LO → NLO	0.07	99 804
SeesawTypeIII_SIGMA-SIGMA0ZH_M-380_13TeV-madgraph	0.002	1.44	LO → NLO	0.37	101 795
SeesawTypeIII_SIGMA-SIGMA0ZW_M-380_13TeV-madgraph	0.004	1.44	LO → NLO	0.63	100 214
SeesawTypeIII_SIGMA-SIGMA0ZZ_M-380_13TeV-madgraph	0.002	1.44	LO → NLO	0.27	100 897
SeesawTypeIII_SIGMAplusSIGMA0HH_M-380_13TeV-madgraph	0.003	1.44	LO → NLO	0.50	101 922
SeesawTypeIII_SIGMAplusSIGMA0HW_M-380_13TeV-madgraph	0.008	1.44	LO → NLO	0.76	99 477
SeesawTypeIII_SIGMAplusSIGMA0HZ_M-380_13TeV-madgraph	0.004	1.44	LO → NLO	0.36	98 616
SeesawTypeIII_SIGMAplusSIGMA0WH_M-380_13TeV-madgraph	0.008	1.44	LO → NLO	0.12	98 681
SeesawTypeIII_SIGMAplusSIGMA0WW_M-380_13TeV-madgraph	0.019	1.44	LO → NLO	0.25	198 335
SeesawTypeIII_SIGMAplusSIGMA0WZ_M-380_13TeV-madgraph	0.010	1.44	LO → NLO	0.07	138 466
SeesawTypeIII_SIGMAplusSIGMA0ZH_M-380_13TeV-madgraph	0.004	1.44	LO → NLO	0.36	99 315
SeesawTypeIII_SIGMAplusSIGMA0ZW_M-380_13TeV-madgraph	0.010	1.44	LO → NLO	0.63	98 330
SeesawTypeIII_SIGMAplusSIGMA0ZZ_M-380_13TeV-madgraph	0.005	1.44	LO → NLO	0.27	101 633
SeesawTypeIII_SIGMAplusSIGMA-HH_M-380_13TeV-madgraph	0.002	1.44	LO → NLO	0.82	101 030
SeesawTypeIII_SIGMAplusSIGMA-HW_M-380_13TeV-madgraph	0.005	1.44	LO → NLO	0.38	99 829
SeesawTypeIII_SIGMAplusSIGMA-HZ_M-380_13TeV-madgraph	0.003	1.44	LO → NLO	0.71	97 230

Dataset Name	Cross Section (pb)	k-factor	Order	Filter Eff.	Events
SeesawTypeIII_SIGMAplusSIGMA-WH_M-380_13TeV-madgraph	0.005	1.44	LO → NLO	0.38	101 206
SeesawTypeIII_SIGMAplusSIGMA-WW_M-380_13TeV-madgraph	0.013	1.44	LO → NLO	0.03	163 501
SeesawTypeIII_SIGMAplusSIGMA-WZ_M-380_13TeV-madgraph	0.007	1.44	LO → NLO	0.24	101 459
SeesawTypeIII_SIGMAplusSIGMA-ZH_M-380_13TeV-madgraph	0.003	1.44	LO → NLO	0.70	100 239
SeesawTypeIII_SIGMAplusSIGMA-ZW_M-380_13TeV-madgraph	0.007	1.44	LO → NLO	0.24	100 827
SeesawTypeIII_SIGMAplusSIGMA-ZZ_M-380_13TeV-madgraph	0.003	1.44	LO → NLO	0.53	120 855
SeesawTypeIII_SIGMA-SIGMA0HH_M-500_13TeV-madgraph	0.000	1.45	LO → NLO	0.51	105 795
SeesawTypeIII_SIGMA-SIGMA0HW_M-500_13TeV-madgraph	0.001	1.45	LO → NLO	0.78	101 865
SeesawTypeIII_SIGMA-SIGMA0HZ_M-500_13TeV-madgraph	0.000	1.45	LO → NLO	0.38	106 417
SeesawTypeIII_SIGMA-SIGMA0WH_M-500_13TeV-madgraph	0.000	1.45	LO → NLO	0.13	108 201
SeesawTypeIII_SIGMA-SIGMA0WW_M-500_13TeV-madgraph	0.001	1.45	LO → NLO	0.26	208 409
SeesawTypeIII_SIGMA-SIGMA0WZ_M-500_13TeV-madgraph	0.000	1.45	LO → NLO	0.07	233 526
SeesawTypeIII_SIGMA-SIGMA0ZH_M-500_13TeV-madgraph	0.000	1.45	LO → NLO	0.38	106 428
SeesawTypeIII_SIGMA-SIGMA0ZW_M-500_13TeV-madgraph	0.001	1.45	LO → NLO	0.64	100 090
SeesawTypeIII_SIGMA-SIGMA0ZZ_M-500_13TeV-madgraph	0.000	1.45	LO → NLO	0.28	103 906
SeesawTypeIII_SIGMAplusSIGMA0HH_M-500_13TeV-madgraph	0.001	1.45	LO → NLO	0.51	105 856
SeesawTypeIII_SIGMAplusSIGMA0HW_M-500_13TeV-madgraph	0.002	1.45	LO → NLO	0.78	99 020
SeesawTypeIII_SIGMAplusSIGMA0HZ_M-500_13TeV-madgraph	0.001	1.45	LO → NLO	0.38	105 350
SeesawTypeIII_SIGMAplusSIGMA0WH_M-500_13TeV-madgraph	0.001	1.45	LO → NLO	0.13	106 154
SeesawTypeIII_SIGMAplusSIGMA0WW_M-500_13TeV-madgraph	0.002	1.45	LO → NLO	0.26	207 893
SeesawTypeIII_SIGMAplusSIGMA0WZ_M-500_13TeV-madgraph	0.000	1.45	LO → NLO	0.07	108 745
SeesawTypeIII_SIGMAplusSIGMA0ZH_M-500_13TeV-madgraph	0.000	1.45	LO → NLO	0.38	106 633
SeesawTypeIII_SIGMAplusSIGMA0ZW_M-500_13TeV-madgraph	0.002	1.45	LO → NLO	0.64	101 611
SeesawTypeIII_SIGMAplusSIGMA0ZZ_M-500_13TeV-madgraph	0.000	1.45	LO → NLO	0.28	102 991
SeesawTypeIII_SIGMAplusSIGMA-HH_M-500_13TeV-madgraph	0.001	1.45	LO → NLO	0.83	102 422
SeesawTypeIII_SIGMAplusSIGMA-HW_M-500_13TeV-madgraph	0.001	1.45	LO → NLO	0.40	104 469
SeesawTypeIII_SIGMAplusSIGMA-HZ_M-500_13TeV-madgraph	0.001	1.45	LO → NLO	0.72	100 637
SeesawTypeIII_SIGMAplusSIGMA-WH_M-500_13TeV-madgraph	0.001	1.45	LO → NLO	0.40	131 607
SeesawTypeIII_SIGMAplusSIGMA-WW_M-500_13TeV-madgraph	0.000	1.45	LO → NLO	0.04	112 882
SeesawTypeIII_SIGMAplusSIGMA-WZ_M-500_13TeV-madgraph	0.001	1.45	LO → NLO	0.25	105 665
SeesawTypeIII_SIGMAplusSIGMA-ZH_M-500_13TeV-madgraph	0.001	1.45	LO → NLO	0.72	103 053
SeesawTypeIII_SIGMAplusSIGMA-ZW_M-500_13TeV-madgraph	0.001	1.45	LO → NLO	0.25	104 324
SeesawTypeIII_SIGMAplusSIGMA-ZZ_M-500_13TeV-madgraph	0.001	1.45	LO → NLO	0.55	101 385
WprimeToTB_TToHad_M-2000_RH_TuneCUETP8M1_13TeV-comphep-pythia8	0.076	1.25	LO → NLO	1.00	199 200
WprimeToTB_TToLep_M-2000_RH_TuneCUETP8M1_13TeV-comphep-pythia8	0.031	1.25	LO → NLO	1.00	199 400
WprimeToTB_TToHad_M-3000_RH_TuneCUETP8M1_13TeV-comphep-pythia8	0.007	1.25	LO → NLO	1.00	196 800
WprimeToTB_TToLep_M-3000_RH_TuneCUETP8M1_13TeV-comphep-pythia8	0.003	1.25	LO → NLO	1.00	197 600

Dataset Name	Cross Section (pb)	k-factor	Order	Filter Eff.	Events
WprimeToTB_TToHad_M-3500_RH_TuneCUETP8M1_13TeV-comphep-pythia8	0.002	1.25	LO → NLO	1.00	199 800
WprimeToTB_TToLep_M-3500_RH_TuneCUETP8M1_13TeV-comphep-pythia8	0.001	1.25	LO → NLO	1.00	199 600
WprimeToTB_TToHad_M-4000_RH_TuneCUETP8M1_13TeV-comphep-pythia8	0.001	1.25	LO → NLO	1.00	197 600
WprimeToTB_TToLep_M-4000_RH_TuneCUETP8M1_13TeV-comphep-pythia8	0.000	1.25	LO → NLO	1.00	196 000

A.2 Expected Minimal p-Values

The probability of observing a p-value of p (or more extreme) is p, thus the distribution of p-values is uniform.

I now want to derive the distribution of minimal p-values, depending on the sample size N: Taking N p-values, the probability that the smallest one has the value p_{min} can be calculated by requiring that the other $N-1$ p-values are in the range $p = [0, p_{\text{min}})$:

$$\Pr(p_{\text{min}} = p) = C \cdot (1-p)^{N-1} \tag{A.1}$$

The proportionality factor C can be calculated from normalization: $C = N$.

The entire probability density for minimal p-values as therefore:

$$\Pr(p_{\text{min}} = p) = N \cdot (1-p)^{N-1} \tag{A.2}$$

This derivation can be verified using pseudo-experiments: In each pseudo-experiment, N numbers are drawn from a uniform distribution $\mathcal{U}([0,1])$ and the minimal value is stored. This is repeated 10 000 times and the empirical distribution function is compared to the predicted distribution. This is shown in figure A.1.

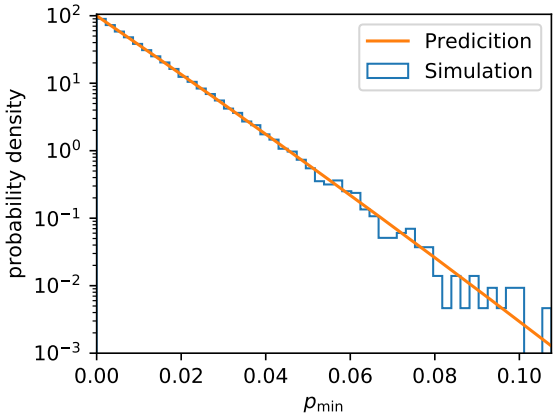


Figure A.1: Minimal p-values from 100 000 pseudo-experiments using $N = 100$.

One additional study on the probability density is to calculate the expected minimal value of N uniform p-values:

$$\langle p \rangle = \int_0^1 N \cdot (1-p)^{N-1} \cdot p \mathrm{d}p = \frac{1}{N+1} \tag{A.3}$$

A.3 Derivation of the Log-Normal p-value

Through the systematic uncertainties, the total number of events is scaled by an unknown factor X, which is a random variable that emerges as product of multiple other random variables x_i. In the following section illustrates a derivation on how X is distributed.

First, the product can be rewritten in terms of a sum within an exponential distribution:

$$X = \prod_i x_i = \prod_i \exp(\ln(x_i)) = \exp\left(\sum_i \ln(x_i)\right) \tag{A.4}$$

$\ln(x_i)$ is a random variable, and the central limit theorem implies that the sum is distributed according to the normal distribution.

$$Y := \sum_i \ln(x_i) \Rightarrow X = \exp(Y) \text{ and } f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right) \tag{A.5}$$

After defining Y as the normally distributed variable with the probability density $f_Y(y)$, one introduces a new unknown probability density function $g_X(x)$:

$$f_Y(y)\,\mathrm{d}y =: g_X(x)\,\mathrm{d}x \tag{A.6}$$

Finally, one can insert all definitions, calculate the derivative and write the distribution function dependent of x:

$$g_X(x) = f_Y(y)\,\frac{\mathrm{d}y}{\mathrm{d}x} \tag{A.7}$$

$$= f_Y(\ln x)\,\frac{\mathrm{d}\ln x}{\mathrm{d}x} \tag{A.8}$$

$$= f_Y(\ln x)\,\frac{1}{|x|} \tag{A.9}$$

$$= \frac{1}{\sqrt{2\pi}\sigma|x|} \exp\left(-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right) \tag{A.10}$$

The result is known as *log-normal distribution* and is the probability density function of a product of random variables.

By performing two substitutions, a parametrization with a more meaningful interpretation is obtained:

$$\sigma \to \ln k = \ln(1 + \sigma/x_0) \text{ and } \mu \to \ln x_0 \tag{A.11}$$

$$g_X(x) = \frac{1}{\sqrt{2\pi}|x| \ln k} \exp\left(-\frac{1}{2}\left(\frac{\ln(x/x_0)}{\ln k}\right)^2\right) \tag{A.12}$$

Here, x_0 is the expected value of x and $k = 1 + \sigma/\mu$ is the relative uncertainty on x_0.

A.4 Z-score and p-value

The probability p associated with a Z-score is the cumulative tail probability of a normal distribution at Z standard deviations from the mean:

$$p = \frac{1}{\sqrt{2\pi}} \int_{Z}^{\infty} \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x \tag{A.13}$$

This yields the conversion formulas:

$$p = \frac{1}{2}\left(1 - \operatorname{erf}\left(\frac{Z}{\sqrt{2}}\right)\right) \Leftrightarrow Z = \sqrt{2} \cdot \operatorname{erf}^{-1}(1 - 2p) \tag{A.14}$$

A table of values computed according to this formula can be found in table A.3.

Z	p	Z	p
0.0	0.500000	2.5	0.006210
0.1	0.460172	2.6	0.004661
0.2	0.420740	2.7	0.003467
0.3	0.382089	2.8	0.002555
0.4	0.344578	2.9	0.001866
0.5	0.308538	3.0	0.001350
0.6	0.274253	3.1	0.000968
0.7	0.241964	3.2	0.000687
0.8	0.211855	3.3	0.000483
0.9	0.184060	3.4	0.000337
1.0	0.158655	3.5	0.000233
1.1	0.135666	3.6	0.000159
1.2	0.115070	3.7	0.000108
1.3	0.096800	3.8	0.000072
1.4	0.080757	3.9	0.000048
1.5	0.066807	4.0	0.000032
1.6	0.054799	4.1	0.000021
1.7	0.044565	4.2	0.000013
1.8	0.035930	4.3	0.000009
1.9	0.028717	4.4	0.000005
2.0	0.022750	4.5	0.000003
2.1	0.017864	4.6	0.000002
2.2	0.013903	4.7	0.000001
2.3	0.010724	4.8	0.000001
2.4	0.008198	4.9	0.000000

Table A.3: Conversion table for Z-score to a one tailed p-value.

A.5 MUSiC Workflow

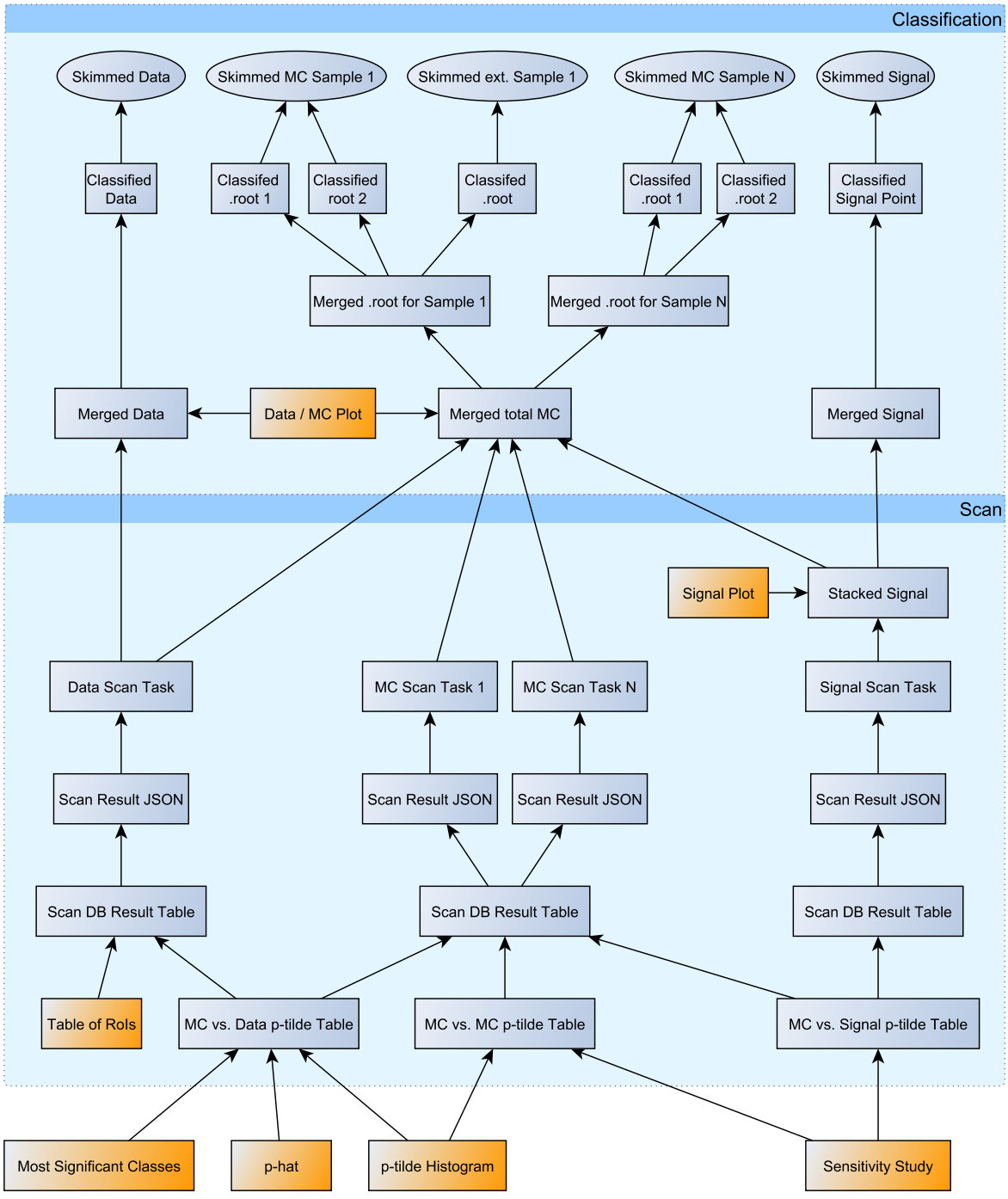


Figure A.2: Illustration of the dependencies of intermediate products during the analysis. Chronologically, the analysis starts at the top by classifying preprocessed ("skimmed") data and simulation samples. In the following steps, the classified samples are merged into a consistent set and the search for deviations is executed. Eventually, the results are stored in a database file and statistical inference is applied.

A.6 Performance of the Lookup-Table

The performance of the lookup table (LUT) is evaluated by applying the automated search on SM pseudo-experiments with and without a LUT. The classification output contains 4 297 event classes, including event classes with b-tagged jets. The pseudo-experiments are generated on the Σp_{T} kinematic distribution.

The LUT under evaluation is 80 MB in size. The performance is measured by wall-time, therefore is subject to small deviations ($\pm 3\,\%$), e.g. due to other processes using the same core. The test is performed twice, on 10 (table A.4) and 100 (table A.5) pseudo-experiments, to indicate scaling behavior.

	without LUT	with LUT
pseudo exp. generation	5 s	5 s
region building / veto	1 373 s	1 337 s
number of regions	223 877 750	
number of θ values	92 503 930	
LUT hits	-	86 681 242 (93.7 %)
time spent calculating θ values	5 220 s (56 µs per θ)	561 s (6 µs per θ)
other	910 s	1 397 s
total time	7 517 s	3 300 s

Table A.4: Performance results on the automated search on 4 297 event classes, 10 pseudo-experiments on the Σp_{T} distribution. All listed durations express wall-time.

	without LUT	with LUT
pseudo exp. generation	33 s	37 s
region building / veto	13 520 s	13 365 s
number of regions	2 238 777 500	
number of θ values	925 039 300	
LUT hits	-	862 902 004 (93.3 %)
time spent calculating θ values	53 746 s (58 µs per θ)	6 165 s (7 µs per θ)
other	1 022 s	1 615 s
total time	68 321 s	21 182 s

Table A.5: Performance results on the automated search on 4 297 event classes, 100 pseudo-experiments on the Σp_{T} distribution. All listed durations express wall-time.

A.7 Adapted Distributions for Coverage Tests

A probability distribution parameterized by μ and σ will in general not assign the same p-value to the value x as the same distribution would do if it were parametrized by x and σ for the value μ. This quite obvious fact becomes important during coverage testing: The pseudo-experiments generate a value N_{SM} from a distribution usually parametrized by N_{true}, but then in the following a distribution around N_{SM} is used in the evaluation of the significance.

In this section, I want to regard the special cases of a normal probability distribution and a log-normal distribution and derive a formula for adapting the distribution width parameter σ such that the p-value to include N_{true} in a (log-)normal distribution parametrized around N_{SM} will remain the same as for N_{SM} parametrized by N_{true}. However, some of the ideas mentioned here can be adapted to other distributions too.

We start by rephrasing the problem as an equation:

$$\Pr(x \leq N_{\mathrm{SM}} | N_{\mathrm{true}}, \sigma) = \Pr\left(x \geq N_{\mathrm{true}} | N_{\mathrm{SM}}, \sigma'\right) \tag{A.15}$$

This equation can also be graphically illustrated as in figure A.3. The blue dashed distribution indicates the initial probability distribution which is used to generate the pseudo-experiments. The orange distribution is the resulting distribution used to calculate the p-value, which has been modified according to the recipe from this chapter, in order to keep the integrals (circled and cross-hatched areas) equal.

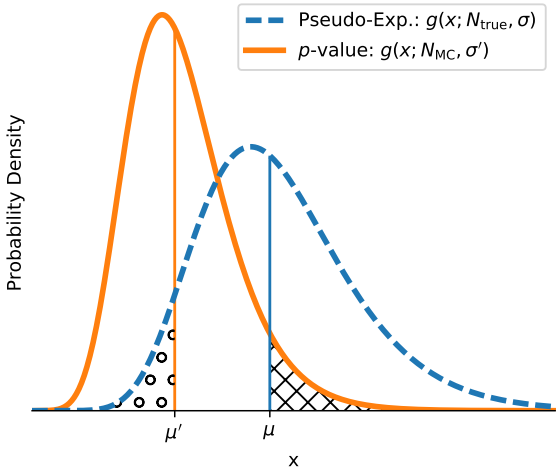


Figure A.3: Difference in widths of distributions while keeping the p-values constant.

Now, let $g(x|N_{\text{true}}, \sigma)$ be the distribution from which the pseudo-experiments are drawn and $G(x|N_{\text{true}}; \sigma)$ its cumulative distribution function. The left-hand-side now spells out:

$$\Pr(x \leq N_{\text{SM}}|N_{\text{true}}, \sigma) = \int_0^{N_{\text{SM}}} g(x|N_{\text{true}}, \sigma)\,\mathrm{d}x = G(N_{\text{SM}}|N_{\text{true}}, \sigma) - G(0|N_{\text{true}}, \sigma) \tag{A.16}$$

The right hand side consists of the distribution used for calculating the p-value:

$$\Pr(x \geq N_{\text{true}}|N_{\text{SM}}, \sigma) = \int_{N_{\text{true}}}^{\infty} g(x|N_{\text{SM}}, \sigma')\,\mathrm{d}x = G(\infty|N_{\text{SM}}, \sigma') - G(N_{\text{true}}|N_{\text{SM}}, \sigma') \tag{A.17}$$

Since G is a cumulative distribution function, one can substitute $G(0) = 0$ and $G(\infty) = 1$ and set both sides equal:

$$G(N_{\text{SM}}|N_{\text{true}}, \sigma) = 1 - G(N_{\text{true}}|N_{\text{SM}}, \sigma') \tag{A.18}$$

The width adaption can now be specified by solving this equation for $\sigma'(N_{\text{true}}, N_{\text{SM}}, \sigma)$.

For this purpose, first a general solution for $\Phi(x)$, which is the cumulative distribution function of a standard normal distribution $\mathcal{N}(\mu = 0, \sigma = 1)$, is derived:

$$\Phi(x) = 1 - \Phi(x') \tag{A.19}$$

$$\Rightarrow x' = \Phi^{-1}(1 - \Phi(x)) \tag{A.20}$$

$$= \sqrt{2}\,\mathrm{erf}^{-1}\left(2 \cdot \left[1 - \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)\right)\right] - 1\right) \tag{A.21}$$

$$= \sqrt{2}\,\mathrm{erf}^{-1}\left(-\,\mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)\right) \tag{A.22}$$

$$= \sqrt{2}\,\mathrm{erf}^{-1}\left(\mathrm{erf}\left(\frac{-x}{\sqrt{2}}\right)\right) \tag{A.23}$$

$$= -x \tag{A.24}$$

Here, it has been used that Φ can be expressed in terms of the error function, which in turn can be further expressed by an integral.

$$\Phi(x) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)\right) \tag{A.25}$$

$$\Phi^{-1}(x) = \sqrt{2}\;\mathrm{erf}^{-1}(2x - 1) \tag{A.26}$$

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\,\mathrm{d}t \tag{A.27}$$

The error function is odd, i.e. $\mathrm{erf}(-x) = -\mathrm{erf}(x)$.

Normal Distribution

A general normal distribution $\mathcal{N}(\mu, \sigma)$ has the following cumulative distribution function:

$$G(\mu'|\mu,\sigma) = \Phi\left(\frac{\mu'-\mu}{\sigma}\right); \quad G(\mu|\mu',\sigma') = \Phi\left(\frac{\mu - N_{\mathrm{SM}}}{\sigma'}\right) \tag{A.28}$$

One can now use the expression for Φ derived earlier to solve the initial equation:

$$\Phi\left(\frac{N_{\mathrm{SM}} - N_{\mathrm{true}}}{\sigma}\right) = 1 - \Phi\left(\frac{N_{\mathrm{true}} - N_{\mathrm{SM}}}{\sigma'}\right) \tag{A.29}$$

$$\Rightarrow \frac{N_{\mathrm{SM}} - N_{\mathrm{true}}}{\sigma} = -\frac{N_{\mathrm{true}} - N_{\mathrm{SM}}}{\sigma'} \tag{A.30}$$

$$\Rightarrow \sigma' = \sigma \tag{A.31}$$

The final result is that the width parameter does not have to be adapted for a normal distribution, which was expected (as it is symmetrical).

Log-Normal Distribution

The argumentation for the log-normal distribution is similar. The cumulative distribution of the log-normal distribution is

$$G(x|N_{\mathrm{true}},\sigma) = \Phi\left(\frac{\ln x - N_{\mathrm{true}}}{\sigma}\right) \tag{A.32}$$

Following the same argumentation, one obtains that again σ must be conserved. However, in contrast to the case of the normal distribution, σ now only depends on the *relative* uncertainty (see equation (A.11)).

Conclusion

Overall, one can conclude that the distributions used for pseudo-experiment generation and evaluation of the p-value are not necessarily the same. In order to keep the probabilities to obtain N_{true} from N_{SM} after drawing N_{SM} from N_{true} the same, in case of the log-normal distribution, the absolute error has to be recomputed with the newly drawn N_{SM}, such that the relative error remains the same. The normal property does not need such adaptation as it has the decent property of being symmetrical.

A.8 Comparison of Coverage Results with Thesis by Stefan Schmitz

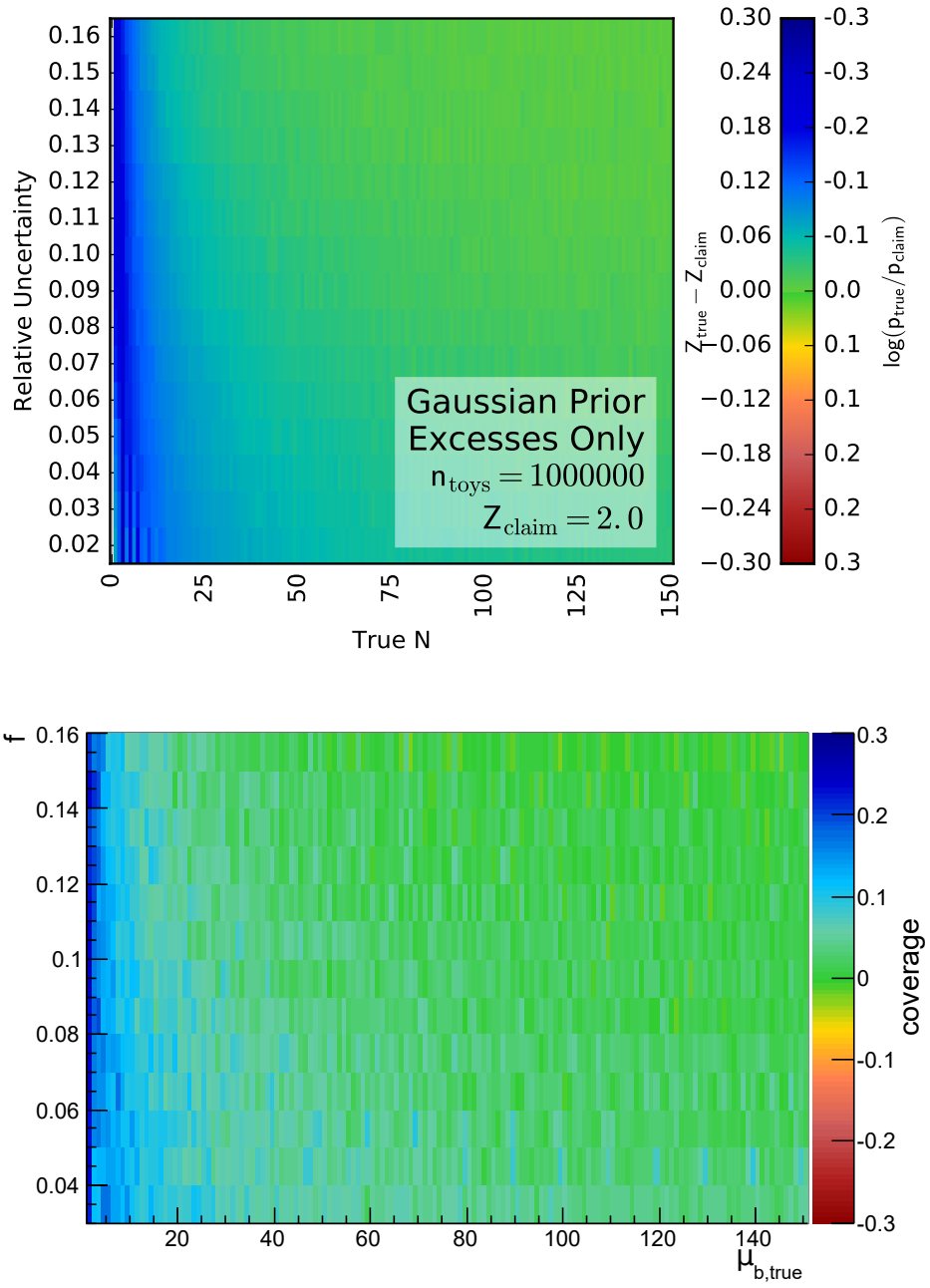


Figure A.4: Comparison of my results (top) with [68] (bottom). The test settings as well as plotting options have been chosen to match. From visual inspection I deduce that the results are very similar and therefore conclude that I can directly compare results from my implementation with the earlier thesis.

A.9 Additional Coverage Results for θ and θ_{LN}

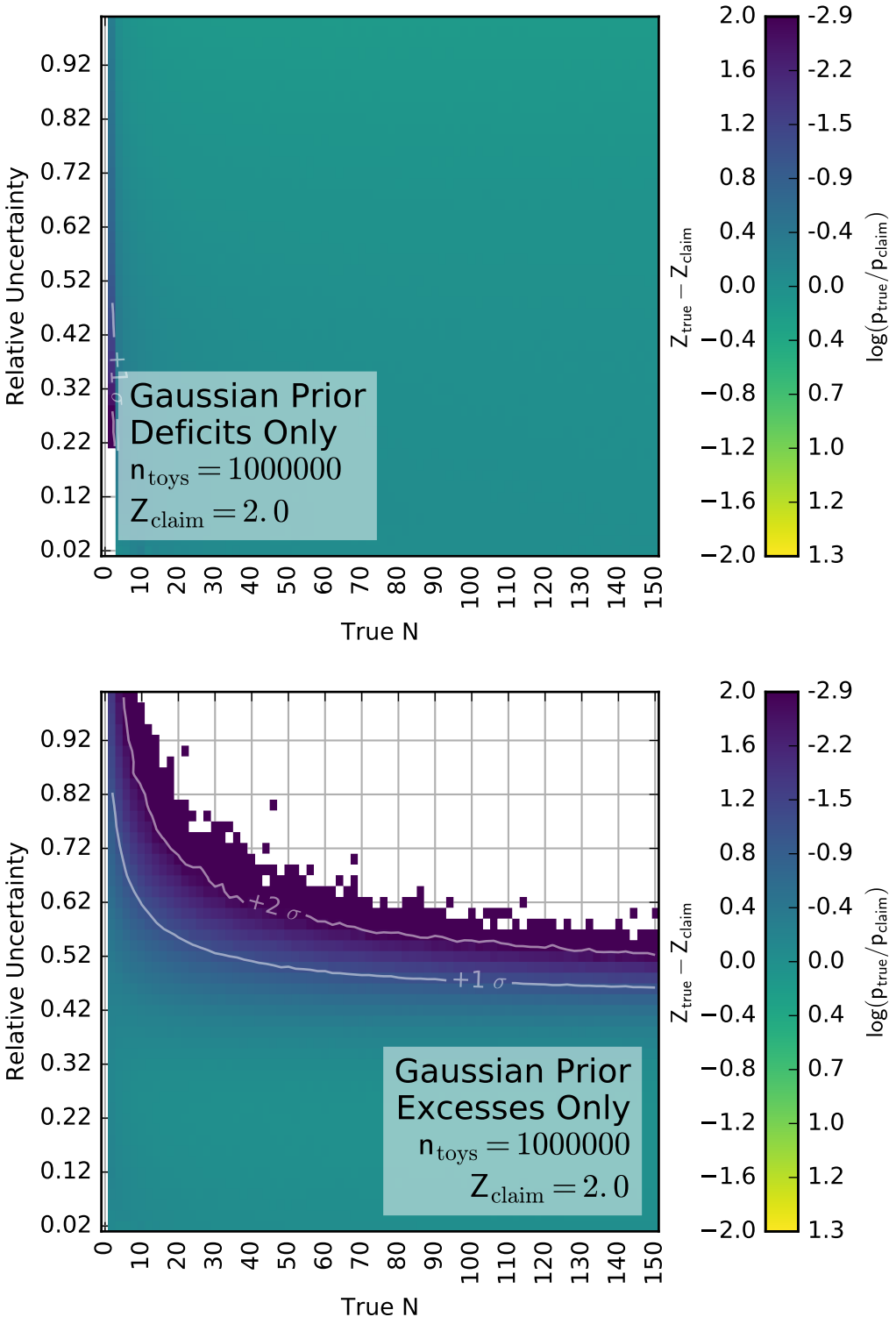


Figure A.5: Results of the coverage study of θ, similar to figure 4.1, but evaluated on a non-logarithmic grid.

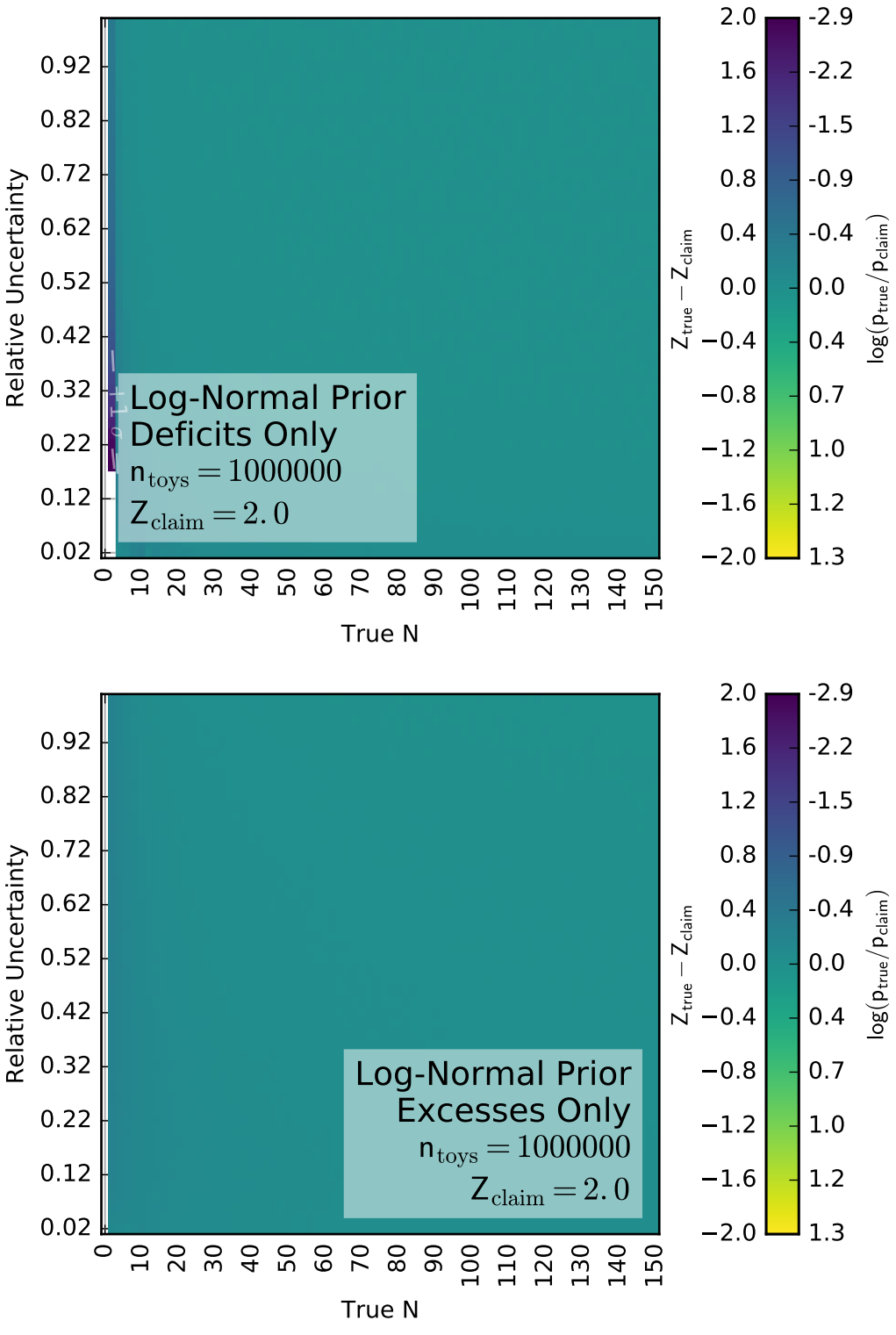


Figure A.6: Results of the coverage study of θ_{LN}, similar to figure 4.2, but evaluated on a non-logarithmic grid.

Danksagung

Diese Masterarbeit wäre nicht möglich gewesen ohne viele Menschen, die mich während des Studiums und der Arbeit unterstützt haben.

Für die Chance, dieses Thema im Rahmen der MUSiC-Analyse zu bearbeiten und damit direkt in der Forschung mitzuwirken, für die Unterstützung während der Entwicklung und für die Korrektur der Arbeit danke ich ganz besonders Herrn Professor Thomas Hebbeker. Des weiteren danke ich Dr. Arnd Meyer dafür, dass er über viele Jahre einen Überblick über die MUSiC-Analyse behalten hat, und mit seinem Expertenwissen bei Fragen zur Seite stand.

Ein ganz besonderer Dank gilt Tobias Pook, als Betreuer, MUSiC-Kollege und für das Korrekturlesen der Arbeit. Er hat die Analyse (MUSiC bei 13 TeV) mit hervorragendem physikalischen Gespür und Erfahrung in Softwareentwicklung geleitet. Mit seinem Blick für das Essentielle hat er mir stets geholfen, der Arbeit einen Fokus zu geben.

Für die lockere, professionelle Arbeitsatmosphäre danke ich auch der restlichen MUSiC-Gruppe, bestehend aus Jonas Roemer, der die MUSiC-Klassifikation maßgeblich weiterentwickelt hat, Simon Knutzen, mit dem ich stets komplizierte Statistikfragen diskutieren konnte und Debbie Duchardt, die mit der Analyse des CMS-Datensatzes von 2012 eine wichtige Grundlage gesetzt hat.

Ich danke außerdem den anderen Wissenschaftlern der CMS-Analysegruppe des III. Physikalischen Instituts A, mit denen ich eng zusammengearbeitet habe. Hierbei sind nicht nur fruchtbare Forschungsergebnisse, sondern auch Freundschaften entstanden.

Professor Martin Erdmann möchte ich nicht nur als Zweitkorrektor dieser Arbeit Dank aussprechen, sondern auch für seinen Einsatz für die Erasmus-Physikstudenten der RWTH. Er hat es mir ermöglicht, zwei Semester meines Masterstudiums an der KTH Royal Institute of Technology in Stockholm zu absolvieren.

Weiterhin möchte ich Dr. Markus Merschmeyer für die Stelle als studentische Hilfskraft danken, durch welche ich meine Kenntnisse der Webentwicklung ausbauen und mein Masterstudium mitfinanzieren konnte.

Zu guter Letzt danke ich meinen Eltern, meinem Bruder und meiner Oma, die mich während meines gesamten Studiums herzlich unterstützt haben, meiner Freundin, die während der stressigeren letzten Monate zu mir stand und meinen Freunden, aus dem Studium, meiner Laufgruppe und dem Aachener Studentenorchester.