RHEINISCH-WESTFÄLISCHE TECHNISCHE
HOCHSCHULE AACHEN

BACHELOR THESIS IN PHYSICS

Introducing the Jensen-Shannon distance as a new metric for the search algorithm to the Model Unspecific Search in CMS

Written by:
Lucas Karwatzki

Presented to:
The Faculty of Mathematics, Computer Science and Natural Sciences of RWTH Aachen University
III. Institute of Physics A

Supervised by:
Univ. Prof. Dr. Rer. nat. Thomas Hebbeker

As second reviewer:
Univ. Prof. Dr. Rer. nat. Alexander Schmidt

March 8, 2024

Abstract

The "Model Unspecific Search in CMS" (MUSiC) is a general approach to the search for new physics. CMS data and Monte-Carlo simulations of the Standard Model are first systematically categorised into classes and the deviations quantified. Model-independent searches are an essential extension to more significant dedicated searches, as new physics may manifest itself in ways that have not yet been predicted. MUSiC is based on a hybrid Bayesian-frequentist approach in which a sophisticated algorithm finds the Region of Interest (RoI). This must then be corrected for the "look elsewhere effect" (LEE).
This thesis presents an initial study of a global search algorithm that potentially offers many advantages. In particular, the complexity of the algorithm is reduced by the complexity of the ROI search $\mathcal{O}(n^2)$, making a multidimensional search possible. The Jensen-Shannon divergence (JSD) was first examined as a search algorithm. Comparability with MUSiC for simplified 1D histograms and an increase in sensitivity in a 2D analysis were found. However, a significant deterioration in sensitivity was observed when CMS Monte Carlo samples were investigated. This was explained by the lack of consideration of systematic uncertainty and regularisation was subsequently introduced. This significantly improved the sensitivity and the analysis showed comparable results, while specific cases where one algorithm performed better were evaluated. Finally, further possible improvements are discussed.

Zusammenfassung

Die "Model Unspecific Search in CMS" (MUSiC) ist eine allgemeine Herangehensweise an die Suche nach neuer Physik. Hierbei werden systematisch CMS-Daten und Monte-Carlo Simulationen des Standardmodells zuerst in Klassen eingeteilt um anschließend die Abweichungen vom Standardmodel zu quantifizieren. Model unabhängige Suchen sind eine essenzielle Erweiterung zu signifikanteren spezifischen Suchen, da neue Physik sich in Weisen manifestieren könnte, die bisher noch nicht vorhergesehen wurden. MUSiC basiert dabei auf einen hybriden Bayesianisch-frequentistischer Ansatz, in welchem ein Algorithmus die Region der stärksten Abweichung (ROI) findet. Diese muss anschließend um den "Look elsewhere effect" (LEE) korrigiert werden.
In dieser Arbeit wird eine erste Studie zu einem global Suchalgorithmus vorgestellt. Diese Herangehensweise bietet potenziell viele Vorteile. Insbesondere wird die Komplexität des Algorithmus um die Komplexität der ROI suche $\mathcal{O}(n^2)$ reduziert, was eine mehr dimensionale Suche möglich macht. In der Arbeit wurde zunächst die Jensen-Shannon Divergence (JSD) als Suchalgorithmus untersucht. Hierbei wurde eine Vergleichbarkeit mit MUSiC für vereinfachte 1D Histogramme und eine Steigerung der Sensitivität bei einer 2D Analyse festgestellt. Beim Übergang auf CMS MC-Histogramme wurde jedoch eine deutliche Verschlechterung der Sensitivität beobachtet. Dies wurde auf die mangelnde Berücksichtigung systematischer Fehler zurückgeführt und anschließend wurde eine Regularisierung eingeführt. Dies verbesserte die Sensitivität deutlich, die Analyse zeigte zu MUSiC vergleichbare Ergebnisse, wobei einzelne Klassen, welche abweichende Ergebnisse liefern, weiter untersucht wurden. Schlussendlich wurden Verbesserungsmöglichkeiten diskutiert.

Contents

Physical units

In this thesis the natural unit system will be used, as is convention in high energy particle physics. The speed of light and the reduced Planck constant are set to c = 1 and $\hbar = 1$. Therefore the mass, the momentum and the energy are expressed in units of energy, the electron-Volt (eV), which correspond to the kinetic energy gained by an electron in a one volt electric potential. The conversion to the SI energy unit Joule(J) is: 1 eV = 1.602×10^{-19} J. Additionally, the elementary e is defined as e = 1.602×10^{-19} C.

1 Standard Model

The Standard Model of particle physics (SM) in its current form was finalised in the 1970s and describes the fundamental particles and their interaction [1]. It consists of 17 elementary particles, which are divided into matter particles or fermions and into interaction particles or bosons, by their spin and its consequent statistical behaviour. All particles predicted by the standard model have been discovered, the last being the Higgs boson which was discovered 2012 in the LHC [2, 3]. The Standard Model also predicts an antiparticle for each of the elementary particles. It is assumed that the antiparticles have the same properties (e.g. Mass, Spin, lifetime), but inverted charges (electric charge and colour charge). [4]. It has to be noted that not all particles have an antiparticle, the Photon, Z-Boson, Higgs and the neutrinos are considered to be their own antiparticle [5].

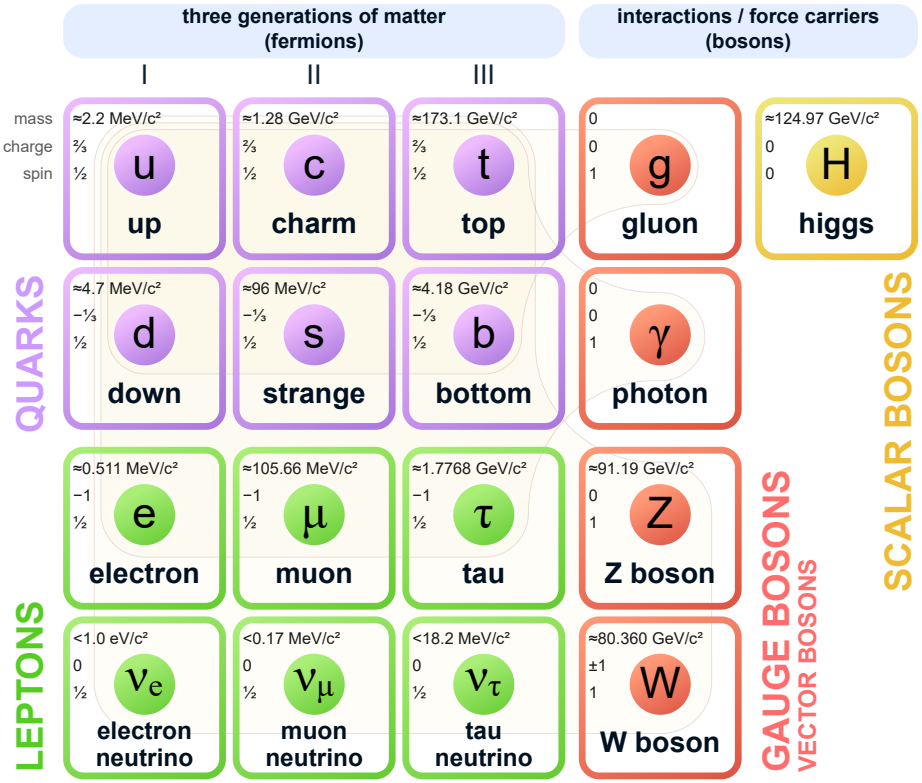


FIGURE 1.1: Illustration of the particle contents of the Standard Model of Particle Physics [6, modified]. The 12 fundamental matter particles are shown on the left and the 5 fundamental interaction particles are shown on the right. The matter particles are divided into quarks (purple) and leptons (green) and both appear in 3 generations. The interaction particles are divided into gauge bosons (red) and scalar bosons (yellow). The brown loops in the background indicate interactions between the particles.

The Standard Model includes three of the four known fundamental forces. It includes the electromagnetic, the weak and the strong force (Section 1.2), while gravity has not been successfully incorporated into the SM yet [7, 8, 9].

1.1 Matter Particles

In the Standard Model 12 of the 17 particles are matter particles. The matter particles all have spin 1/2 and are therefore considered fermions, whence follows, that they follow the Fermi-Dirac statistics and obey the Pauli exclusion principle. The group of matter particles consists of the quark (q) and lepton (ℓ) subgroups, which are themselves divided into three generations. The particles of each generation differ only in the mass that increases with the generation. Therefore, particles, except neutrinos, of the second and third generation, decay after a characteristic time into particles of the first generation. This explains why most of the visible matter in our Universe consists of particles of the first generation [5, 10].
The Standard Model includes six quarks (u, d, c, s, t, b) which all carry one of the three colour charges (red, green, blue). The quarks also carry an electrical charge of $2/3\ e$ (u, c, t) or $-1/3\ e$ (d, s, b), where e is the elementary charge. Quarks cannot exist isolated but form hadrons that consist of either two quarks, known as mesons, or three quarks, known as baryons[1] [12]. The Standard Model further includes six leptons, three are electrically charged leptons (e, μ,τ), with a charge of $1e$ and three corresponding electrically neutral leptons (ν_e, ν_μ, ν_τ), which are called neutrinos [9, 12].

1.2 Fundamental Interaction and exchange particles

The Standard Model describes three out of the four fundamental forces. The representative carriers of the forces are the interaction bosons. They are gauge bosons with a spin of one. Additionally, the scalar Higgs boson, with spin zero, was introduced. The Higgs boson is the carrier of the Higgs field, which gives the mass to elementary particles [8, 9]. Fundamental forces can be classified by their interaction range and relative strength which are measured in the coupling parameters, listed in Table 1.1 [13, 14].

Interaction	Strong	Weak	Electromagnetic
Exchange particle	8 gluons	Z, $W^{\pm}$	γ
Couples to	colour charge	weak hypercharge	electric charge
coupling strength	α_s	$\approx \alpha_s \cdot 10^{-6}$	$\approx \alpha_s \cdot 10^{-2}$
coupling range [m]	$\approx 10^{-15}$	$\approx 10^{-17}$	Indefinite

TABLE 1.1: Table of the most important properties at low energy of the 3 fundamental forces included in the Standard Model.

1.2.1 Electromagnetic interaction

The electromagnetic interaction describes the forces exchanged between electrically charged particles. These are the elementary quarks and the charged leptons (e, μ, τ). Further, composited particles like bosons, mesons and atoms can be charged. The carrier of this force is the massless and chargeless photon (γ), which couples only to electrically charge particles. Because of the null mass of the photon, the force has an unlimited range. The electromagnetic interaction is responsible for a wide variety of the physics we can observe in our common world. It defines the bound of atoms and molecules and characterises most of chemistry [8].

[1] Exotic unstable hadrons with four (tetraquark) or five (pentaquarks) quarks have been discovered[11].

1.2.2 Weak interaction

The weak interaction has the lowest coupling strength of the three fundamental forces. Since all particles carry the weak hypercharge, the weak force interacts with all particles of the Standard Model and is therefore the only force acting on neutrinos [15]. The carriers of the weak force are the Z-boson with a mass of $m_Z \approx 91$ Gev and the $W^{\pm}$-bosons with a mass of $m_W \approx 81$ GeV [9]. Due to the heavy mass, the Z and W bosons decay fast ($\approx 10^{-25}s$)[16], which limits their effective range to $\approx 10^{-17}$m [14]. A break in symmetry has been observed by the Wu experiment [17], which discovered that particles with a spin aligned with their momentum (right-handed particles) do not participate in the weak interaction [18].

1.2.3 Electroweak unification

The unification of the weak and the electromagnetic force has first been predicted in the 1960s [19, 20, 21]. It has been experimentally observed in two stages, by discovering the neutral Z boson and then the charged W boson. The unification describes the fact, that above the unification energy of 246 GeV the weak and the electromagnetic force became indistinguishable [14, 15].

1.2.4 Strong Interaction

The strong force describes the interaction between particles that carry a colour charge, which are quarks and gluons. There are three colours blue, green, and red, and their respective anticolors. If you combine a colour and an anticolour, you get one of the eight massless gluons, which are the carrier of the strong force [15]. All particles involved in the strong interaction are subject to confinement, which describes the fact that the particles can only be observed in bound states. This is due to the special form of the potential of the strong force, which contains a linear term [8]:

$$V(r) = \frac{4\alpha_s}{r} + b \cdot r + c \tag{1.1}$$

The form of the potential reduces the effective range of the strong force to $\approx 10^{-15}$ m. The strong force is the force that holds particles that are composites of quarks together, for example, the proton or neutron and the cores of atoms [12].

1.2.5 Higgs mechanism

The Higgs mechanism describes how fundamental particles obtain their mass. It was proposed in the 1960s and was experimentally proven with the LHC in 2012 by identifying the Higgs boson with a mass $m_H = 125$ GeV. Contrary to its gauge boson counterparts, the Higgs boson has a spin quantum number of S = 0, distinctively devoid of any force-mediating attributes [15].

1.3 Incompleteness and motivation

The Standard Model of Particle Physics, while highly successful, has limitations and fails to provide a comprehensive theory. In response to new discoveries, researchers have proposed extensions or modifications to the Standard Model to provide a more comprehensive understanding of particle physics. Gravity, one of the four fundamental forces is not yet described by the Standard Model. In the following chapter some of the phenomena which cause discrepancy between the Standard Model and our observations are described.

1.3.1 Dark matter

Dark matter is a hypothetical form of matter that does not participate in the electromagnetic interaction, but acts through gravitational effects. The proposed state of matter could explain a magnitude of observations, described by neither the general relativity theory nor the Standard Model. One of the observations is the speed of the rotation of the arms of galaxy's around its core. By Kepler's third law, we assume that the rotation velocities decrease with distance from the core, but this is not observed, as can be seen in Figure 1.2. Many further deviating observations could be justified by dark matter, such as the formation of galaxy clusters, the inhomogeneity of the cosmic microwave background or the effect of gravitational lensing, which is not in compliance with the visable mass [22].

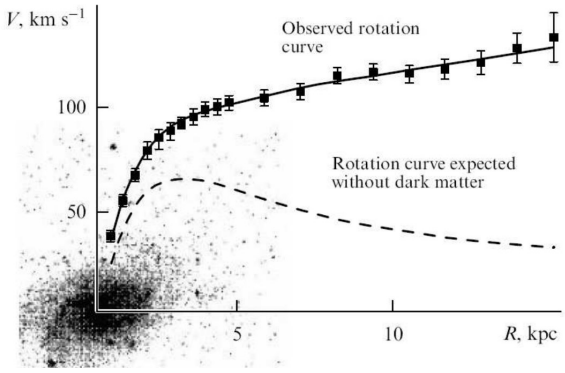


FIGURE 1.2: Illustration of the observed rotation curve and the rotation curve, which is expected by Kepler's third law [23].

1.3.2 Matter antimatter asymmetry

Matter and antimatter extinguish each other, when interacting. The Standard Model predicts that the Big Bang created an even amount of matter and antimatter. Since our world is made up of matter, there must be an asymmetry between antimatter and matter. Although the violation of the CP symmetry in the weak interaction was discovered, the effect is too small to explain the domination of matter [24].

1.3.3 Neutrino oscillation

Neutrino oscillation describes the phenomenon of neutrino changing its flavour. This has been first observed in the Homestake experiment [25] for solar neutrinos. This process is not in accordance with the Standard Model, because the probability of a neutrino changing its flavour is dependent on the second power of the mass difference $(\Delta m)^2$ of the two neutrino flavours. In the Standard Model neutrinos have no mass and therefore the probability of the flavour change would be zero [26].

1.3.4 Gravity

Gravity as one of the four fundamental forces is not included in the Standard Model. Because of its relatively low coupling constant $\alpha_G/\alpha_s \approx 10^{-39}$ it has a neglected role in the interactions typical studied in particle physics. The charge to which gravity couples is the mass, which is only positive. This explains why it plays a major role in interaction at larger scales, such as the Universe [14].

2 Beyond the Standard Model

The Standard Model of particle physics, while successful in describing a large variety of observed physics, has its limitation, which have been outlined in Section 1.3. Therefore, physicists have been searching for new theories to expand the Standard Model, which are known as Beyond the Standard Model(BSM) theories. Two of these theories will be briefly explained and will later be used as a benchmark for the new search algorithm [8].

2.1 W'

On small scales the weak interaction couples only to left-handed particles. To mathematically describe this symmetry break, a left-right symmetric (LRSM) extension of the standard model has been proposed [27]. This theory predicts a new gauge group at high energies, so that the symmetry is broken at lower energies [28].

$$SU(2)_L \otimes SU(2)_R \otimes U(1) \tag{2.1}$$

This approach predicts new bosons W′ and Z′, which are heavier copies of the Standard Model equivalents[29, 30]. A further advantage of this theory is that it explains the mass of neutrinos through the Seesaw mechanism and a new set of heavy neutrinos [31]. There are many more theories that predict a further charged gauge boson, such as grand unified theory models to unify all forces at higher energies, kaluza-Klein-theories to explain gravity or the Little Higgs model to stabilise the mass of the Higgs boson [28]. The properties of W′ are described in the "Reference Model", which predicts similar coupling and decay properties, besides a suppressed decay into Z and W, as the W boson in the Standard Model, but leaves only the mass of the boson as an open parameter [29, 32]. The decay channels of the W′ boson are deducted from the conservation of charge and only the most probable processes that are in a generation are displayed [28].

$$W'^{-} \rightarrow \bar{u}d, \bar{s}c, \bar{t}b, e^{-}\overline{\nu_e}, \mu^{-}\overline{\nu_\mu}, \tau^{-}\overline{\nu_\tau} \tag{2.2}$$

It has to bee noted, that through the higher mass of the $W^{-\prime}$ the decay into $\bar{t}b$ is possible, which is not the case for the W boson of the Standard Model [32]. The branching ratios of the W′ boson can be found in Table 2.1. Although the branching ratios are larger for the quark channels, than for the lepton channels, the most studied channels are leptonic, since they offer a better separation between background and signal [33, 34].
The W′ prime boson has been excluded for a mass below 6 TeV for Atlas [35] and 5.7 TeV for CMS [36].

Branching ratios of the Reference Model

$M_{W'}$	1Tev	2Tev	5 TeV	$M_{W'}$	1Tev	2Tev	5 TeV
$W'^{-} \rightarrow \bar{u}d$	24.2 %	24.0 %	23.9 %	$W'^{-} \rightarrow e^{-}\overline{\nu_e}$	8.2 %	8.2 %	8.2 %
$W'^{-} \rightarrow \bar{c}s$	24.2%	24.0 %	23.9 %	$W'^{-} \rightarrow \mu^{-}\overline{\nu_\mu}$	8.2 %	8.2 %	8.2 %
$W'^{-} \rightarrow \bar{t}d$	24.3 %	25.0 %	25.1 %	$W'^{-} \rightarrow \tau^{-}\overline{\nu_\tau}$	8.2 %	8.2 %	8.2 %

TABLE 2.1: Branching ratios of the Standard Model for relevant decay channels [32].

2.2 Sphaleron

The vacuum states of the electroweak theory (see section 1.2) are not absolute and the potential has a periodic structure of minima, which are characterised by the Chern-Simons numbers (N_{CS}) [13, 37]. The change of the system from one vacuum state to another is achieved with a process involving the sphaleron or with a tunnelling process involving an instanton, represented in Figure 2.1 [38].

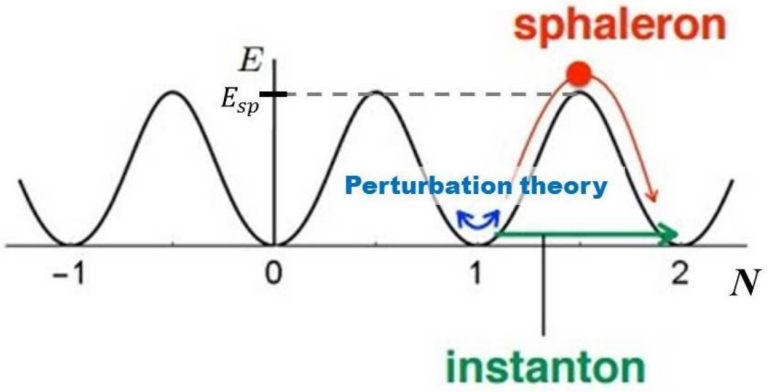


FIGURE 2.1: Illustration of the periodic structure of the electroweak energy density as a function of Chern-Simons number, illustrating the sphaleron and the instanton process [38].

The sphaleron is a static solution, decays immediately and the process can not be described by a perturbation theory [37, 39]. To change from one eigenstate to another, the sphaleron must pass the potential barrier E_{sph}. The threshold barrier can be calculated from the static energy solution of the electroweak field equations and is shown in equation 2.3, where m_W and m_H are the masses of the Higgs and W bosons, α_W is the electroweak constant and B is a tabulated function [8, 38] .

$$E_{sph} = \frac{2m_W}{\alpha_W} B\left(\frac{m_H}{m_W}\right) \approx 9.1\ TeV \tag{2.3}$$

Sphaleron induced processes always include one member of each electroweak doublet (ℓ) and (anti) quark (q), from all three generations and three colours. The leading order sphaleon donor does not involve electroweak bosons and could look like in equations 2.4 and 2.5, where X is a set of particles with B = L = 0 [40, 39, 41].

$$q\,q \rightarrow \bar{\ell}\,\bar{\ell}\,\bar{\ell}\,\bar{q}\,\bar{q}\,\bar{q}\,\bar{q}\,\bar{q}\,\bar{q}\,\bar{q}\,\bar{q} + X \qquad (\Delta N_{CS} = -1) \tag{2.4}$$

$$q\,q \rightarrow \ell\,\ell\,\ell\,q\,q\,q\,q\,q\,q\,q\,q\,q + X \qquad (\Delta N_{CS} = +1) \tag{2.5}$$

As can be seen in equations 2.4 and 2.5, the sphaleron transition changes the baryion number(B) and the lepton number(L) by $3 \cdot \Delta N_{CS}$, while preserving the difference between the baryon and the leptop number (B-L), but violates the sum of both (B+L) [39, 41]. The factor of 3 is due to the expected number generation, as it is assumed, that each generation causes a change of ΔN_{CS} in baryon and lepton number [41].
The change of the baryon and lepton number in the sphaleron process is unique, since they are conserved in any Standard Model process [13, 37]. This process could lead to further understanding of the matter antimatter asymmetry, which describes the fact that in the early universe more baryons than antibaryons were formed in a process called baryogenesis [13, 41]. Current exclusion limits depend on the expected threshold energy and a scaling factor of the cross section [42].

3 Experimental Setup

3.1 CERN and the Large Hadron Collider

CERN (Conseil Européen pour la Recherche Nucléaire), the European organisation for nuclear research, was established in 1954 in Meyrin, Switzerland. Today, it features a variety of experiments from many fields, but is most known for its particle accelerators and its contribution to particle physics [43].
The Large Hadron Collider (LHC), is the last step in a CERN's acceleration chain, it is a synchrotron collider in which charged particles are accelerated in opposing directions and collided in one of the four intersection points. The particles travel along two beam pipes in the tunnel of the former Large Electron Positron Collider(LEP) with a circumference of 26.7 kilometres and is approximately 100 meter below the surface [44]. To prevent particles from interacting with air molecules and slowing down, the pipes incorporate an ultra-high vacuum of up to 10^{-14} bars[45]. 1232 dipole magnets with field strength up to 8.3 Tesla are used to keep the particles on their circular path, while 858 quadrupoles and 6208 magnets of higher order are focusing and correcting the beam [46]. To produce such high field strength, magnet coils are produced from superconducting material and cooled to 1.9 Kelvin [5, 47, 46].

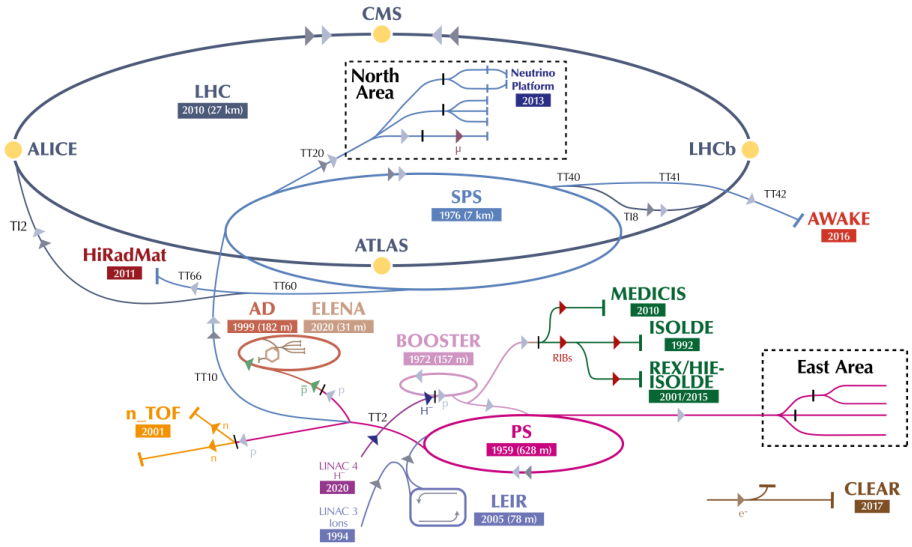


FIGURE 3.1: Illustration of the accelerator complex and some of the most important experiments of CERN [48, modified].

The particle accelerators are designed to accelerate and collide either protons and protons (pp) or heavy ions such as lead (PbPb). Currently, the LHC reaches centre of mass energies for the proton-proton collision of $\sqrt{s} = 13.6$ TeV, which is the highest energy ever achieved in a particle accelerator. It is designed for even higher energies of up to $\sqrt{s} = 14TeV$ and a maximum instantaneous luminosity of $\mathcal{L} = 10^{34}\, cm^{-2}s^{-1}$ [44].

3.1.1 Proton acceleration chain

The acceleration of the particles takes place in multiple stages, which are illustrated in Figure 3.1. The Proton acceleration starts by accelerating negatively ionised hydrogen in the linear accelerator 4 (LINAC 4). The LINAC uses multiple acceleration techniques to accelerate hydrogen ions to 160 MeV, while separating them into pulses of 400 ms [49]. The next step in the acceleration chain is the Proton Synchrotron Booster (PSB), which separates the electrons from the hydrogen ions and accelerates the protons, like all later accelerators, employing an alternating high voltage in radio frequency cavities [50, 51]. The PSB accelerates protons to 2 GeV and then injects them into the Proton Synchrotron (PS), which accelerates the proton up to 26 GeV [52]. The Super Proton Synchrotron is the next accelerator, which operates up to energies of 450 Gev [53], before reaching the LHC with a design energy of up to 7 TeV [54]. In the LHC the protons are split into 2808 bunches of approximately $1.2 \cdot 10^{11}$ protons [46].

3.1.2 Particle Detectors

The two particle beams intercept at four points around the LHC. To observe the collisions, nine experiments are conducted[55]. The four large experiments are:

- **ALICE** (A Large Ion Collider Experiment) is dedicated to heavy-Ion physics and is designed to study the strong interaction at extreme energy densities[56].
- **ATLAS** (A Tordoidal LHC Apparatus) and **CMS** (Compact Muon Solenoid) are general purpose detectors, which are designed to investigate a wide range of physics [57, 58].
- **LHCB** (Large Hadron Collider beauty) specialises in the study of the b quark to investigate small differences between matter and antimatter [59].

In addition to the four major experiments, five smaller experiments are conducted at the LHC:

- **Totem** (Total, elastic and diffractive cross-section measurement) makes precision measurements of protons in the forward direction [60].
- **LHCf** (Large Hardron Collider forward) uses the LHC to simulate and study cosmic rays in laboratory conditions [61].
- **MoMEDAL-MAPP** (Monopole and Eotics Detector at the LHC) searches for magnetic monopoles and other exotic particles [62].
- **FASER** (Forward Search Experiment) is designed to search for light and very weak interacting particles [63].
- **SND@LHC** (Scattering and Neutrino Detector at the LHC) specialises in the study of neutrinos [64].

3.2 The Compact Muon Solenoid Detector

The Compact Muon Solenoid has a length of 21.6 metres, a diameter of 15 meters and a total weight of 14.000 tonnes, making it one of the largest LHC detectors. CMS and Atlas are the two general purpose detectors of the LHC, which are designed

to detect a wide range of physics processed mainly in proton-proton collisions [50]. Both detectors have strengths in different subsystems, while the main difference is in the name giving superconducting magnet. The CMS detector is designed to have high geometrical acceptance around the interaction point and therefore the subdetectors are placed in layers around the centre, which can be seen in Figure 3.2. Each subsystem plays an essential role in the detection of the different interactions, and they are explained individually in the next sections.

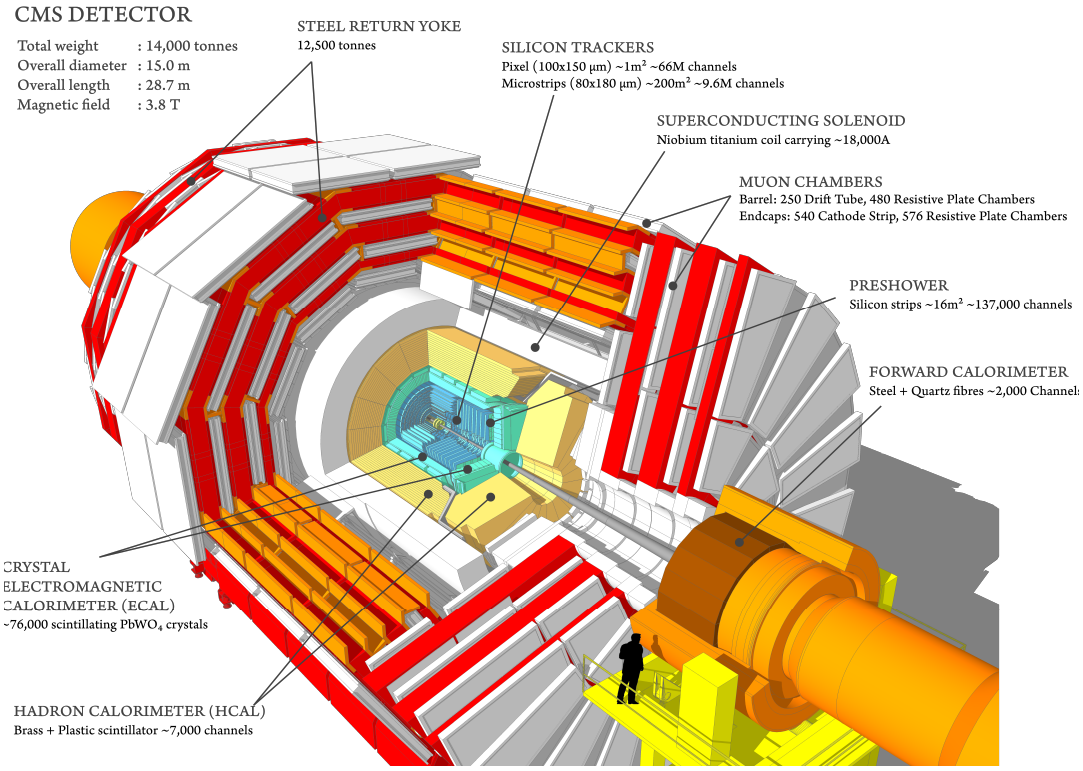


FIGURE 3.2: Illustration of the CMS detector with explanation and facts about the detector and its subsystems [65].

3.2.1 Coordinate system

The origin of the CMS coordinate system is in the nominal intersection point, where the two particle beams are expected to cross. In Cartesian coordinates, the x axis points towards the centre of the LHC, the y axis vertically upwards, and the z axis along the beam, in the direction of the Jura mountains. The spherical coordinates correspond to the classical definitions, ϕ is measured in the x/y plane. θ is measured between the z-axis and the x/y plane, while r corresponds to the distance from the intersection point [67].

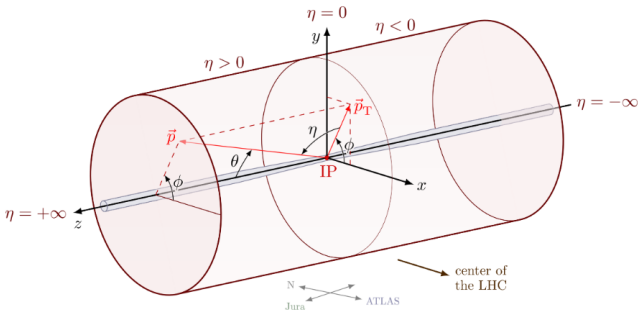


FIGURE 3.3: Illustration of the coordinate system used in CMS [66].

Instead of using θ, the pseudorapidity η is commonly used and defined by equation 3.1.

$$\eta = -ln\left(\frac{tan(\theta)}{2}\right) \tag{3.1}$$

This has the advantage, that in the limit $E \gg m$, the pseudorapidity is approximately the particles rapidity $y = artanh(v_z) \approx \eta$, which is a lorenz invariant quantity. With the pseudorapidity the angular distance ΔR of two objects can be calculated as [67]:

$$\Delta R = \sqrt{(\eta_2 - \eta_1)^2 + (\phi_2 - \phi_1)^2} \tag{3.2}$$

Further the detector is divided into two regions, the barrel which is the cylindric part and the endcaps which are the covers enclosing the of ends of the cylinder.

3.2.2 Tracking system

The subdetector closest to the interaction point is the inner tracking system, which measures the trajectories of charged particles and is used to reconstruct the vertices [67, 68]. To achieve this task with the highest precision, the particle flux must be considered and therefore three different systems are used in different radii[67].

- In the region closest to the detector (20cm < r) pixel detectors with a pixel size of $\approx 100\mu m \times 150\mu m$ are used [67].
- In the intermediate region (20cm < r < 55cm) silicon micro strip detectors, with a minimum cell size of $\approx 10cm \times 80\mu m$, are used [67].
- In the outermost region (r > 55cm) of the inner tracker, larger silicon micro strips, with a minimum cell size of $\approx 25cm \times 180\mu m$, are used [67].

3.2.3 Electromagnetic Calorimeter

The next layer is the electromagnetic calorimeter (ECAL), which performs measurements of the energy of photons and electrons with a precision of up to 1%. It utilises leadthungstate ($PbWO_4$) scintillator crystals, which combine absorber and detection material. 61,200 Crystals are places in the Barrel ($|\eta| < 1.479$) and 7324 crystals are placed in the endcap ($1.479 < \eta < 3.0$).
The particles shower, which describes a process of alternating pair production and bremsstrahlung, in the scintillator, which takes the deposited energy and emits photons with a wavelength of approximately 420 nm. These are then detected by avalanche photodiodes in the barrel and vacuum photodiodes in the endcaps [68].

3.2.4 Hadron Calorimeter

The hadronic Calorimeter is a second calorimeter and is specialised to detect the energy of hadron and jets produced by hadrons. Because of the higher "nuclear interaction length" of the hadron, more material is needed to stop them, but they still deposit some energy in the ECAL. To absorb as much of the remaining energy, a sandwich calorimeter is used, where interchanging layers of brass plates as absorber while Bicron and Kuraray as scintillators are used.
The particles shower through processes of the strong interaction, similar as in the ECAL. The photons produced by the scintillators are transported by waveguides to

hybrid photodiodes. The HCAL covers the same angular range as the ECAL ($\eta <$ 3.2), but in addition a forward hadron callorimeter (HF) extends the angular range to ($\eta <$ 5.2) [69].

3.2.5 Superconducting Magnet

To precisely measure the transverse momentum p_T of charged particles with high momentum, a strong magnetic field is required. This is provided by a super conducting solenoid magnet. The solenoid is composed of a four layers coil of superconducting niobium–titanium (NbTi), which is cooled to 4.5 K to achieve a magnetic field strength of 3.8 Tesla. The field lines are parallel to the beamline and therefore the charged particles are bent in the transverse plane. The magnetic flux is returned by a thick iron yokes, which is the outer layer of the detector [68].

3.2.6 Muon Chambers

The last subsystem is the muon chamber, designed to identify muons, measure their momentum and provide signals for the trigger. They are placed as the last system and are arranged in the steel yoke, to ensure a low probability for any particle besides the muon and weakly interacting neutrinos to enter the muon chambers [8]. Four different gaseous detector technologies are used in the muon chamber. Gaseous detectors work on the principle of a passing muon ionising the gas inside the detector. Gaseous detectors have been chosen to maximise the detector volume for a more precise momentum estimation, while keeping the cost low.
Drift Tubes are used in the barrel region ($|\eta| < 1.2$), while the endcap. ($0.9 < |\eta| <$ 2.4) features Cathode Strip Chambers. Additionally Resistive Plate Chamber ($|\eta| <$ 1.9) are added and for high pseudo rapidity ($1.55 < |\eta| < 2.18$) area Gas Electron Multipliers are used[1] [70].

3.2.7 Trigger and data acquisition

Proton proton collisions occur approximately 40 million times a second. To successfully extract useful information's out of these collisions, the amount of data must be reduced. This is achieved by employing a system of triggers, which acts as prefilters and decides, which data is stored.
The L1 trigger evaluates the signals of the muon detectors and both calorimeters to significantly reduce the data. The L1 trigger is hardware based and implemented through programmable electronics. The system rejects events with characteristics that have already been studied to reduce the data rate to a frequency of approximately 100 kHz. This frequency is the maximum, which can be accepted by the CMS readout system [71]. The High Level Trigger (HLT) is the second trigger and software based. It performs a streamlined version of the reconstruction algorithm and selects events of interest [70, 71].

3.2.8 Particle Flow

Particles in CMS can interact with multiple subdetectors as illustrated in Figure 3.4. To reconstruct and identify the different stable particles in CMS, the information measured by the different subsystems is used. To combine the information from all subdetectors, the particle flow algorithm (PF) is used [18]. Through the use of the

[1] The Gas Electron Multipliers where not part of CMS for RUN 2 to which the Monte-Carlo samples studied in this thesis correspond.

information obtained by all subdetectors, misidentification is reduced and energy resolution increases [12, 72].
The Particle Flow algorithm works in multiple steps. In the first step, each subdetector reconstructs the tracks and the clusters of deposited energy separately. Then the tracks in the inner tracking system are extrapolated to find matching clusters in the calorimeters [18]. The tracks between the inner tracker and the muon system are matched [18, 73].
In the last step, the PF builds the particle candidates from the linked tracks and clusters one by one, removing the identified particles. First muons are reconstructed from the tracks measured in the inner tracker and the muon system. Electrons are identified by the track in the inner tracker and the energy cluster in the ECAL. Because photons are not electrically charged, they can be identified by a cluster in the ECAL without a corresponding track in the inner tracker. In the same way, neutral and charged hadrons can be identified by clusters of energie in the HCAL and their interaction with the inner tracker. Higher observables, such as Missing Transverse Energy (MET), are then calculated from the properties of the reconstructed objects [5, 18, 72].

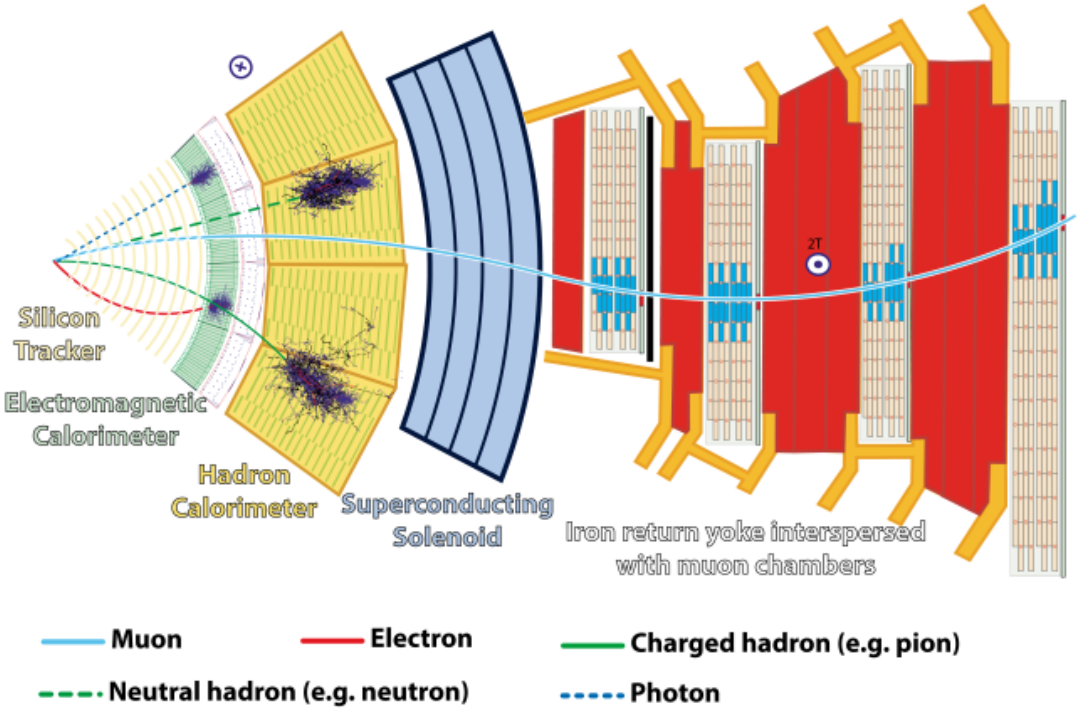


FIGURE 3.4: Illustration of the interaction of different particles in the CMS detector [74]. It shows how electrically charged particles (electron, muon, charged hadron) are bent by the magnetic field of the superconducting solenoid, while neutral particles (photon, neutral hadrons) have a straight path. It shows that photons and electrons are stopped in the ECAL, while charged and neutral hadrons are stopped in the HCAL. Only Muons and neutrinos (not indicated) reach the muon chambers.

4 Model Unspecific Search in CMS

As highlighted in section 1.3 and 2, it is likely that the Standard Model of particle physics in its current form is an incomplete description of the Universe. Particle colliders are designed to test the Standard Model and search for possible extensions. Typically these searches are developed to investigate a specific extension of the Standard Model and therefore study distinct final states with kinematic selections to optimise for a given signal. Because of the large number of theories for New Physics, the given limitation on computational and personnel resources and because New Physics may manifest itself in ways that have not been predicted yet model independent searches are a complementary alternative to dedicated investigations. A model independent search does not optimise for a specific BSM theory in the final states nor in the signal selections of the kinematic distribution. This strategy results in a lower sensitivity than dedicated searches, but covers a broader phase space and parameters. Should a model independent search find a strong enough deviation, a dedicated search can be triggered [8, 15, 75].
The "Model Unspecific Search in CMS" (MUSiC) is such a general approach to the search for new physics for CMS proton-proton collision data. MUSiC considers 3 kinematic distributions and hundreds of final states, which are systematically categorised into event classes. For each event class in each distribution, the data is compared to Monte-Carlo simulations of the Standard Model and the deviation is evaluated. There are no elaborate filters or strict kinematic limitations in place to optimise the examined phase space during the analysis. MUSiC was first introduced in 2008 and since then the development has been continuously improved to evaluate the most recent data, optimise the search algorithm and extend the analysed phase space [5, 8].
In this thesis a new search algorithm is investigated using the MUSiC framework is investigated. To evaluate the performance of the algorithm, a sensitivity study is conducted using Monte-Carlo simulations corresponding to a centre of mass energy of $\sqrt{s} = 13\ TeV$ and an integrated luminosity of 137.0 fb^{-1}.

4.1 Analysis procedure

4.1.1 Dataset and simulated samples

The MUSiC analysis requires a complete of Monte-Carlo simulation of the Standard Model including all possible interaction and information's about the detector and its subsystems (e.g. Geometry, materials and their properties, efficiencies, etc.). For the different interactions, specialised MC event generators are employed [1]. These simulations are typically generated by the CMS collaboration and made available through the Data Aggregation System (DAS), with the corresponding DAS key. The sample

[1]The MC event generators used are:PYTHIA 8.212 , MADGRAPH5 aMC@NLO version 2.2.2 , POWHEG v2 and SHERPA 2.1.1 . For more information, including all simulated processed and the corresponding generator please regard: [76]

files provided by CMS contain more information than needed for the analysis, therefore in a preprocessing step, called skimming, additional information, which is not needed, is stripped.

4.1.2 Preparation

4.1.2.1 Object and event selection

It is essential to correctly evaluating the physics object content for each event for the MUSiC analysis. This involves identifying each reconstructed object (see Section 3.2.8) and eliminating any overlap between them. Given that the analysis depends on simulation for background prediction, selection criteria are applied to the various objects, minimising the impact of misidentification while maintaining a reasonably high efficiency in object selection [76]. The objects considered in the MUSiC analysis are e, μ, γ, jets, b-tagged jets[2] and missing transverse momentum (p_T^{miss} / MET)[3]. The corresponding object selection criteria can be seen in Table 4.1

Object	p_T (GeV)	Pseudo rapidity
Muon (μ)	> 25	$\lvert\eta\rvert$ < 2.4
Electron (e)	> 25	0 < $\lvert\eta\rvert$ < 1.4442
		or 1.566 $\lvert\eta\rvert$ < 2.5
Photon (γ)	> 25	$\lvert\eta\rvert$ < 1.4442
Jet	> 50	$\lvert\eta\rvert$ < 2.4
b-tagged jet	> 50	$\lvert\eta\rvert$ < 2.4
Missing transverse momentum (p_T^{miss})	> 100	-

TABLE 4.1: Summary of the object selection criteria, taken from [76]. The p_T thresholds and the cuts in η are to ensure high quality objects for the analysis.

Every event must fulfil one of the single or double lepton triggers or the single photon trigger as listed in Table 4.2. This limits the MUSiC analysis to events with at least one lepton (e/μ) or one photon. Every event that was triggered has a higher threshold than the detector triggers listed in the middle column, but the analysis requirements are slightly higher. This is done to maximise efficiency, while ensuring a high enough statistics [12, 76]. To prevent counting a single event multiple times, the overlap between triggers is removed [76].

Trigger	Trigger level requirement	Analysis requirement
Single muon trigger	1 μ with p_T > 50 GeV	1μ with p_T > 53 GeV
Single electron trigger	1 e with p_T > 115 GeV	1e with p_T > 120 GeV
Double muon trigger	1 μ with p_T > 17 GeV	2μ with p_T > 20 GeV each
	and 2nd μ with p_T > 8 GeV	
Double electron trigger	2 e with p_T > 33 GeV each	2e with p_T > 40 GeV each
Single photon trigger	1 with p_T > 175 GeV	1 with p_T > 200 GeV

TABLE 4.2: Summary of online and offline trigger criteria, taken from[76]. The analysis requirements are higher than the offline triggers to ensure that the selections is above the turn on of the efficiency curves.

[2] A b-tagged jet is jet originating from a b quark [77].
[3] Missing transverse momentum is explained in section 4.1.3.1

4.1.2.2 Event cleaning

Physical objects in proximity may result in misidentification during analysis. Therefore, if two objects are too close to each other, they will be cleared. The threshold for muons and electrons is $\Delta R < 0.4$ and for jets and bjets $\Delta R < 0.5$. If two objects are too close, the particles are ordered by their expected purity, resulting in the following order: μ, e, γ and jets. Then the particles with the expected lower purity will be removed [76, 78].

4.1.2.3 MC weighting

The number of events generated from the Monte Carlo simulations is independent from the expected event rates, measured by the experiment. Therefore, the MC events must be reweighted. The expected event rate is proportional to the luminosity $\mathcal{L}$, which is determined by the experiment and the cross section σ of the physical process:

$$N = \sigma \cdot \int \mathcal{L}\, dt \tag{4.1}$$

The implementation of these information, as well as the k-factor, which corrects for higher order in the perturbation theory, if available, result in the first weight [5, 79]:

$$\omega_{lumi} = \frac{k \cdot \sigma \cdot \int \mathcal{L}\, dt}{N_{MC}} \tag{4.2}$$

where N_{MC} is the number of events generated by the simulation. Another effect which must be taken into account is the Pileup (PU), which describes the fact that with the high luminosity of the LHC multiple collisions happened during a bunch crossing. This results in additional interactions, which must be taken into account by the simulation, resulting in another weight ω_{PU}. Monte-Carlo processes may also introduce a negative contribution to the Monte-Carlo yield, and this can be incorporated by introducing a generator weight factor denoted ω_{gen}. The total weight factor ω_{MC}, which is applied to Monte Carlo events, can now be calculated as [5, 79]:

$$\omega_{MC} = \omega_{lumi} \cdot \omega_{PU} \cdot \omega_{gen} \tag{4.3}$$

4.1.2.4 Event classification

For the MUSiC analysis events in data and Monte Carlo samples are categorised into different classes, which correspond to their final states, based on the physics content of each object. This is done to preserve the model unspecific aspect of the MUSiC analysis and not dedicate towards a specific final state. MUSiC discriminates between three types of event classes, which will be briefly elaborated with the example event containing $\{1e, 2\mu, 1jet\}$, which is also shown in Figure 4.1 [76, 78]:

- **Exclusive** classes only contain the selected physical object. Therefore, each event is assigned only one exclusive class. In the case of $\{1e, 2\mu, 1jet\}$ the exclusive class is "$1e + 2\mu + 1jet$".
- **Inclusive** classes contain a core of the physical objects explicitly named in the event class, but may contain additional objects, specified by the suffix **"+X"**. The additional objects can be one or multiple of the following:e, μ, γ, jet, bjets, p_T^{miss}. Therefore, an event will belong to multiple inclusive event classes, for

example " $1e + 2\mu + 1jet$ " is in " $1e + X$ ", " $1e + 1\mu + X$ ", " $1e + 1\mu + 1jet + X$ ", and many more.

- **Jet-inclusive** classes are defined as inclusive classes, but restrict the additional objects to jets. Because higher jet multiplicities have a high expected disagreement between data and simulations and are therefore expected to not be accurate, all exclusive classes with five or more jets are instead assigned to the jet inclusive "X+5jets+Njets" class. The example class $\{1e, 2\mu, 1jet\}$ is present in two jet inclusive classes: " $1e + 2\mu + Njets$ " and " $1e + 2\mu + 1jet + Njets$ " [15]

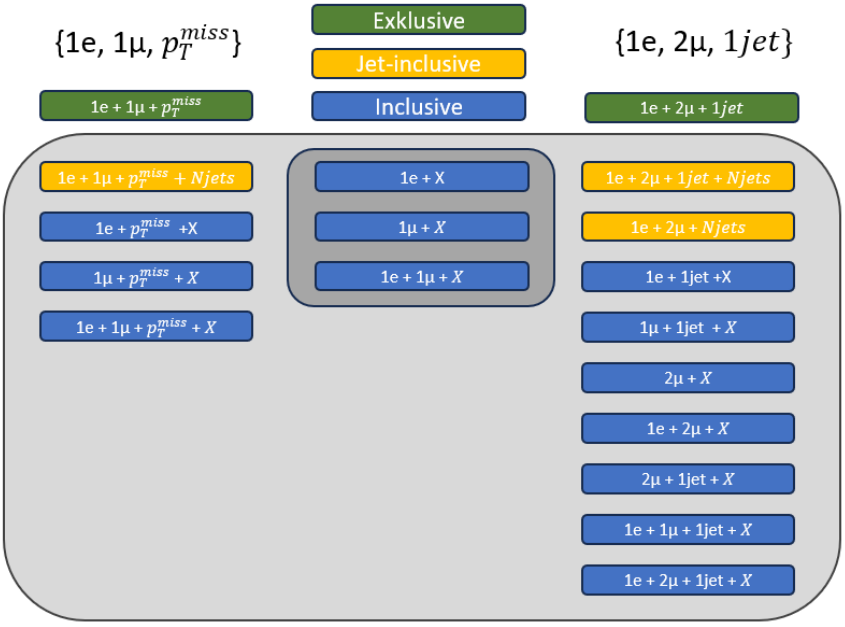


FIGURE 4.1: Illustration of the MUSiC classification for the two final states $\{1e, 1\mu, p_T^{miss}\}$ and $\{1e, 2\mu, 1jet\}$. The first final state results in one exclusive, one jet-inclusive and six inclusive states, which are shown on the left side. The second final state results in one Exclusive, two jet-inclusive and 10 Inclusive event classes, shown on the right side. The overlap between the two final states is marked in a darker grey in the middle.

An event class is only further considered for the MUSiC analysis if the MC event yield is greater than 0.1 and at least one data count is present in the event class.

4.1.3 The Scan

4.1.3.1 Kinematic distributions

MUSiC analyses focus on three kinematic distribution. The distributions have been chosen because they are expected to be the most sensitive to physics, at high transverse momenta, as predicted by many BSM theories [76]. The distributions are as follows:

1. **Sum of transverse momenta** S_T The Sum of transverse momenta S_T is computed by summing the absolute value of the transverse momenta $|\vec{p_T}|$ of all N physics objects in the event class:

$$S_T = \sum_i^N |\vec{p_{T,i}}| = \sum_i^N \sqrt{p_{T,x,i}^2 + p_{T,y,i}^2} \tag{4.4}$$

If missing transverse momenta (MET) is named in the event class, then the missing transverse momentum $|\vec{p_T^{miss}}|$ is also considered in the sum. This variable has been chosen, because many BSM theories predict new particles with heavy mass (see Section 2), which would be visible on the high energy tails of the S_T distribution [76].

2. **Invariant mass** m_{inv} The invariant mass m_{inv} is calculated from the energy E_i and the momentum $\vec{p}_i$ of the particles in the event class:

$$M_{inv} = \sqrt{\left(\sum_i^N E_i\right)^2 - \left(\sum_i^N p_{x,i}\right)^2 - \left(\sum_i^N p_{y,i}\right)^2 - \left(\sum_i^N p_{z,i}\right)^2} \tag{4.5}$$

In cases where there is a missing transverse momentum p_T^{miss} is present the invariant mass is not a good metric, because the longitudinal component of the missing momentum cannot be accurately evaluated. Instead, the transverse mass M_T is used[4]:

$$M_T = \sqrt{\left(\sum_i^N E_i\right)^2 - \left(\sum_i^N p_{x,i}\right)^2 - \left(\sum_i^N p_{y,i}\right)^2} \tag{4.6}$$

This distribution is of special interest for new physics, where a new heavy particle is expected. The resonance of the particle could be visible through the mass distribution of its decay products. The invariant mass is only calculated when at least two objects are present in the event class[76].

3. **Missing transverse momentum** p_T^{miss} When missing transverse momentum p_T^{miss} is present in the event class, a third distribution is taken into account. The missing transverse momentum is defined as the negative sum of the transverse momenta:

$$p_t^{miss} = |\vec{p_t^{miss}}| = |-\sum_i^N \vec{p_{T,i}}| \tag{4.7}$$

The missing transverse momentum is of special importance for non interacting particle such as neutrinos or hypothetical non interacting BSM particles [76].

For the exclusive event class, all included events share the same objects and the kinematic distribution can be evaluated based on all objects in the final state. There is an ambiguity for inclusive and jet inclusive event classes. Therefore, to evaluate all events on the same properties, only objects explicitly mentioned in the event class name are considered, when calculating the kinematic distributions. Additionally, if the multiplicity of an object in an event is higher than in the event class of interest, the particles with the highest transverse momentum p_T are considered. For example, in the event class $1e + 2\mu + p_T^{miss} + X$ only the four named objects contribute to the kinematic properties [76].
To evaluate kinematic distributions in a histogram, the data and Monte Carlo samples must be binned. The bin widths are chosen as a trade-off between a larger bin width, advantageous for computational efficiency but disadvantageous for sensitivity to potential narrow signals, and a smaller bin width, where random fluctuations

[4]When refereeing to this kinematic variable, generally the term invariant mass is used, without discriminating between the invariant mass or the transverse mass.

may become more prominent and potentially obscure the actual deviations of interest. The optimal bin width is automatically determined, on the basis of the expected detector resolution for the different kinematic regions. This leads to larger bins at higher energies. Generally, all bin widths are integer multiples of 10 GeV [76].

4.1.3.2 p-Value

To quantify the significance of the deviation between the Monte-Carlo background and the data, MUSiC follows a hybrid Bayesian frequentist approach and calculates a p value. The p-value quantifies the likelihood that an effect of at least a comparable strength to the observed one could arise exclusively from the statistical fluctuations inherent in the null hypothesis [15].
To evaluate the p-value, a test metric must be defined to compute the strength of the deviation. In the MUSiC analysis the test statistics is the number of events in a region, where the number of observed particles N_{obs} is compared to the expected number of events from the Monte Carlo simulations of the SM N_{MC}. The number of events in a region, can be expressed in a Poison distribution:

$$\mathcal{P}(N;\mu) = \frac{\mu^N e^{-\mu}}{N!} \tag{4.8}$$

where μ is the true mean of the distribution and N is the number of events. As Monte-Carlo simulation encompasses systematic uncertainties, it would be inaccurate to consider the MC count as the Standard Model expectation (N_{SM}). Instead, the systematic uncertainty in the MC simulation is incorporated using a truncated Gaussian distribution. Combing these assumptions, the p-value can be calculated as:

$$p_{data} = \begin{cases} \sum_{i=N_{data}}^{\infty} \cdot \alpha \int_0^{\infty} dx\, exp\left(-\frac{(x-N_{MC})^2}{2\cdot\sigma_{MC}}\right) \cdot \frac{x^i e^{-x}}{i!}, & \text{if } N_{data} \geq N_{MC} \\ \sum_{i=0}^{N_{data}} \cdot \alpha \int_0^{\infty} dx\, exp\left(-\frac{(x-N_{MC})^2}{2\cdot\sigma_{MC}}\right) \cdot \frac{x^i e^{-x}}{i!}, & \text{if } N_{data} < N_{MC} \end{cases} \tag{4.9}$$

with a normalisation factor α from the truncated Gaussian. In the calculation of the p-value, the case of exceeding data events ($N_{data} \geq N_{MC}$) and the case of less data than MC ($N_{data} < N_{MC}$) is treated separately [15, 76, 80].

4.1.3.3 Region of Interest scan

To find the region with the greatest deviation, the Region of Interest (RoI) scanner, goes through all possible regions, to find the lowest p-value $p_{data,min}$. A region is defined as any combination of adjacent bins. The total number of region is $N_{bins}(N_{bins}+1)/2$, where N_{bins} is the number of bins in the histogram. This process is illustrated in Figure 4.2.

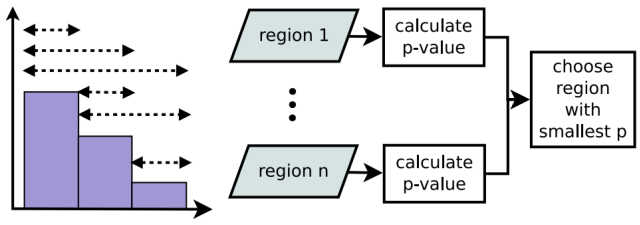


FIGURE 4.2: Illustration of the MUSiC ROI scanner [80]. In every possible region of adjacent bins the p-value is calculated and compared, to find the region with the lowest p-value.

The region with the lowest p-value in the distribution, which shows the deviation with the least likelihood to be compatible between the observed data and the simulated MC background, is defined as the Region of Interest. A schematic illustration can be seen in Figure 4.3.

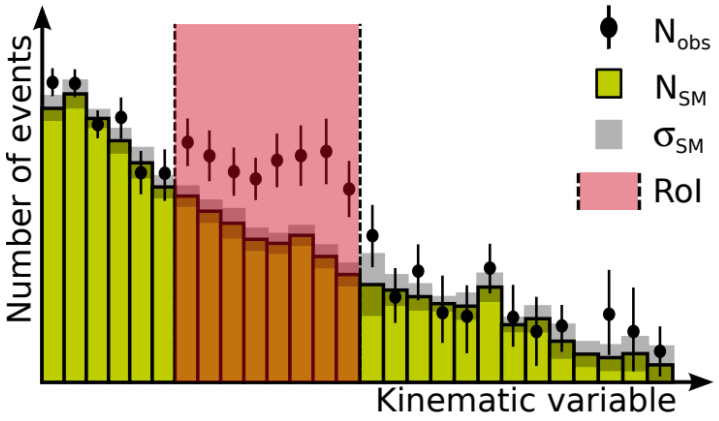


FIGURE 4.3: Schematic of a kinematic distribution [15]. The figure shows the SM expectation with uncertainty $N_{SM} \pm \sigma_{SM}$ as well as observed data N_{obs}. The Region of Interest is illustrated and marked in red.

To reduce statistical fluctuations, a threshold is set on the minimum number of bins of which a region can consist of. The threshold is three for the sum of transverse momentum and the missing transverse momentum and one for the invariant mass distribution. The threshold for the invariant mass is smaller because the deviation caused by a resonant particle might only influence one bin in the invariant mass distribution, but causes a broader deviation in the other classes. There is also a veto system that prohibits certain regions, to increase the robustness of the analysis[5].

4.1.3.4 Post trial probability ($\tilde{p}$) and the look elsewhere effect

The p-value quantifies only the likelihood of a local deviation. To evaluate if a given data distribution matches the expected MC distribution from the Standard Model, we need to take the whole distribution, with its uncertainties into account. This is done by correcting the so called "look-elsewhere effect" (LEE) which describes that it becomes increasingly probable to find a significant local deviation, just by a fluctuation, the more regions are probed [81]. The probability of finding a deviation with the significance α increases roughly as $p \approx 1 - (1 - \alpha)^n$, where n is the number of regions [10]. To correct this effect, the post-trial probability $\tilde{p}$, which corrects the look elsewhere effect, is calculated. The $\tilde{p}$ value, expresses the probability of finding a deviation as strong as the observed one, in one region, caused by statistical fluctuations [10]. Applying an analytical correction factor, based on the distribution, is not possible because the different regions and their uncertainties are correlated. Instead, up to 10^5 simulated pseudo experiments are used. The number of pseudo experiments is a compromise between sensitivity and computational expenses. These pseudo experiments vary the predicted SM event counts N_{SM} for each bin, corresponding to the associated uncertainty [15].
Each of pseudo experiment is then compared to the SM prediction, the RoI scan is performed and a smallest p-value p_{min} is determined and stored. The p-value from

[5]The details of the veto system can be read in the MUSiC analysis notes [80]

the data distribution $p_{data,min}$is then compared with the p-values of the pseudo experiments, to calculate $\tilde{p}$ [80]:

$$\tilde{p} = \begin{cases} \frac{N_{pseudo}(p_{min} < p_{data,min})}{N_{pseudo}} & , where\ N_{pseudo}(p_{min} < p_{data,min}) > 0 \\ \frac{1}{N_{pseudo}} & , where\ N_{pseudo}(p_{min} < p_{data,min}) = 0 \end{cases} \tag{4.10}$$

The Post trial probability $\tilde{p}$ now expresses the probability of a random deviation as strong as the observed one, in any region of the distribution. A lower limit for the $\tilde{p}$ is set, because the absence of a stronger deviation in pseudo experiments, can not be interpreted as zero probability for such a deviation. A distribution of p-values and the estimation $\tilde{p}$ is shown in Figure 4.4.

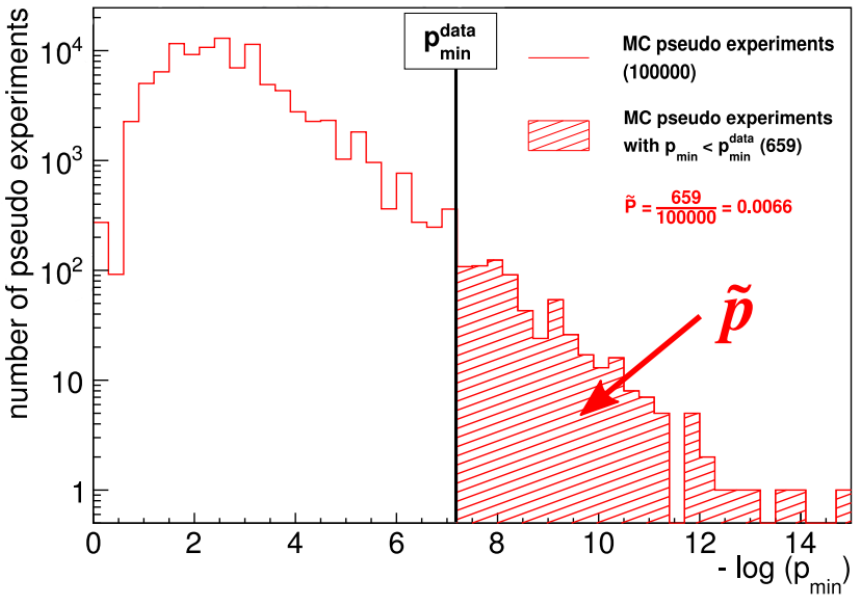


FIGURE 4.4: Illustration of p-value distribution and the calculation of the $\tilde{p}$ -value [80]

4.1.3.5 Pseudo Experiments

In every pseudo experiment, the Standard Model (SM) expectation in each bin is varied, based on both the expected value and the diced shifts that represent associated systematic uncertainties following the null hypothesis.
For each region n and systematic uncertainty i the width of a symmetrised 68% confidence interval $\Delta_{i,n}$ is calculated. In addition, to account for the correlation of the uncertainties, a random number k_i, is drawn from a standard normal distribution with expectation value zero and uncertainty of one. The expected Standard Model $N_{SM,n}$ event counts are shifted:

$$N_{SM,shifted} = N_{SM,n} + \sum_i k_i \cdot \frac{\Delta_{i,n}}{2} \tag{4.11}$$

If the shifted expectation $N_{SM,shifted}$ is negative, it is set to zero. The expectation $N_{SM,shifted}$ is then smeared with a Poisson distribution, to account for the limited sample size. No shifts are applied if a region has zero Monte Carlo expectation [15, 76, 78].

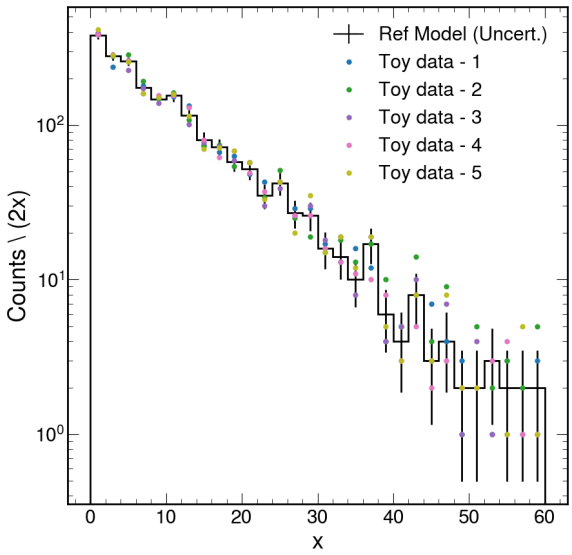


FIGURE 4.5: Illustration of the sampling with a simplified reference model. The Figure shows an exponential distribution with the corresponding uncertainty and five sampled toys.

The sampling example is illustrated in Figure 4.5. The kinematic distribution is approximated by exponential distribution, to show the sampling. It is shown that the uncertainty in the sampled toys corresponds to the uncertainty of the reference model.

4.1.4 Global overview

To get a broad overview and not focus on the event classes with the strongest deviation, a plot is created that contains the information from all event classes in a given distribution. Some BSM candidates, such as quantum black hole models [82], could show a small deviation, but in several different final states, which could be detected through a global illustration of all event classes [80].
To produce the global overview, a histogram from the $\tilde{p}$ values of all event classes for a given distribution is built first. This $\tilde{p}$-distribution can be compared with distributions obtained from pseudo experiments. For each round of toys, the $\tilde{p}$ values are calculated, comparing the expectation of SM with its shifted version. From these $\tilde{p}$ values, a new histogram can be built, which represents the null hypothesis. This process is repeated for each round of toys. To build the global plot, the median, as well as the 68% and 95% percentiles per bin, are calculated [80]. Additionally a numerical expectation can be calculated, following the assumption that all $\tilde{p}$ values are uniformly distributed. The analytically expected number of event classes in each bin can now be evaluated with the lower edge of the bin B_{low} and the upper edge B_{up} and the total number of event classes N_{Ec} [80]:

$$N_{bin} = (10^{-B_{low}} - 10^{-B_{up}}) \cdot N_{Ec} \tag{4.12}$$

The entire process and the resulting global plot can be seen in Figure 4.6. If a simulated signal is used instead of observed data, shifts for the signal are calculated because of the uncertainty inherent in the signal. Therefore, the signal is treated in the same way as the pseudo experiments before. This leads to the fact that the signal values in the global have uncertainties as well.

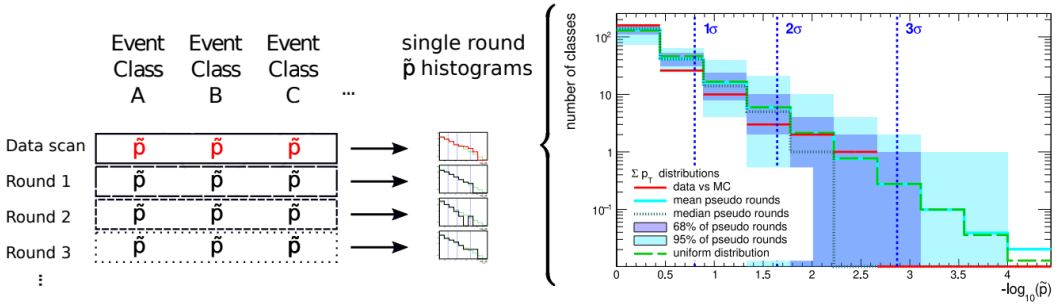


FIGURE 4.6: Illustration of the global $\tilde{p}$ distribution [80]. On the left side, the building of the global plot is illustrated, where the histogram from the data is first built, resulting in the red lines in the global plot. Then, for each round of pseudo experiments, a new $\tilde{p}$ histogram is calculated. Then, for each bin of the pseudo experiment histogram, the median and mean are calculated and inserted into the global plot as blue (mean) and dashed (median) lines. Then the one (light blue) and two σ (dark blue) intervals are calculated and inserted as shaded areas. Finally, the analytical expectation is inserted as a green dashed line. The last bin here displays the overflow bin. In this example no significant deviation from the SM is indicated.

4.2 Sensitivity study

To demonstrate, quantify and compare the capabilities of the MUSiC analysis sensitivity studies are an important tool. For sensitivity studies no data recorded by CMS is going to be used, but pseudo-data signals, which are Monte Carlo simulations of the Standard Model with the extension of a BSM theory. With these informations, it becomes clear, in which event classes the analysis is sensitive and in which it is not, potentially leading to improvements.
In this thesis a new search algorithm will be presented. To quantify the capabilities of the algorithm, a sensitivity study is conducted, using MUSiC as the benchmark on the same BSM signals. Therefore, the results for the MUSiC algorithm will be briefly introduced here. A complete analysis on this set of simulated samples has been done by MUSiC [76] at a luminosity of $35.9 fb^1$ and by Yannik Kaiser[8] at the same increased luminosity as this thesis $137.0 fb^{-1}$.

4.2.1 W' sensitivity study

The W' boson could decay into a lepton and neutrino or into a quark antiquark pair as discussed in chapter 2. The decay into quarks offer a worse separation of signal and background, due to the large QCD dijet and multijet crossection [34], therefore only leptonic decays were simulated in the Signal samples. A leading order Feynman diagram can be seen in Figure 4.7. In CMS such events are reconstructed into high energy lepton, a large amount of missing transverse energy. Additionally, an arbitrary number of jets and photons resulting from higher orders of the perturbation theory can be present in the Signal. These procresses are suppressed by the corresponding coupling factors, see Table 1.1.

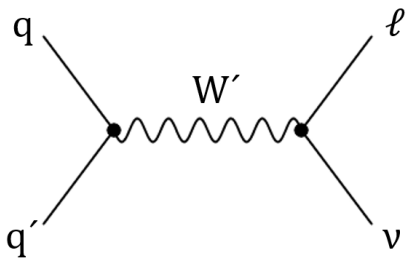


FIGURE 4.7: Feynman diagram of the leading order leptonic decay of the W′ boson.

A typical distribution for the W′ Signal, $1\mu + MET$, together with the SM background, is shown in Figure 4.8. The transverse mass distribution shows a clear excess of Signal over the SM-background. The highest deviation where the SM + signal exceeds the SM approximately 150 times is around the expected Breit-Wigner peak at a sum of transverse mass of $M_T = 2TeV$.

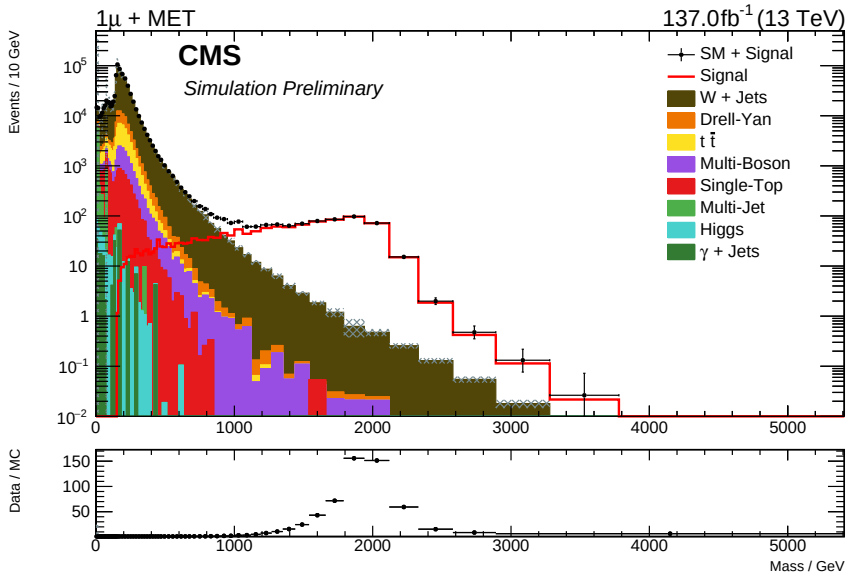


FIGURE 4.8: Transverse mass distribution of the $1\mu + MET$ distribution with a W′ ($M_T = 2TeV$). The different process groups of the SM background are indicated.

The global $\tilde{p}$ distributions for the invariant mass distribution, for all W′ masses and including all event classes, are illustrated in Figure 4.9. The analytical assumption does not fit the expected SM distribution. This is due to an excess in the first bin, which consequently translates into deficits in the later bins. This has also been previously observed and was connected to the minimum yield threshold [10] or an overestimation of the uncertainty [76]. Further explanations could be the Veto system or the problematic samples, but due to the time constraints, this is not further investigated in this Thesis.
In the invariant mass distributions, for all W′ masses, a significant excess of the event classes can be observed in the last bin. The presence of more low $\tilde{p}$ values than expected is a clear indication of a deviation from the standard model, which hints for new physics. The amount of the event classes observed in the last bin decreases as expected with increasing W′ mass, from the high relative uncertainty in the high tail of the distributions. For the highest investigated mass of 5 TeV, two event classes have been found, which is still outside of the expected event class count.

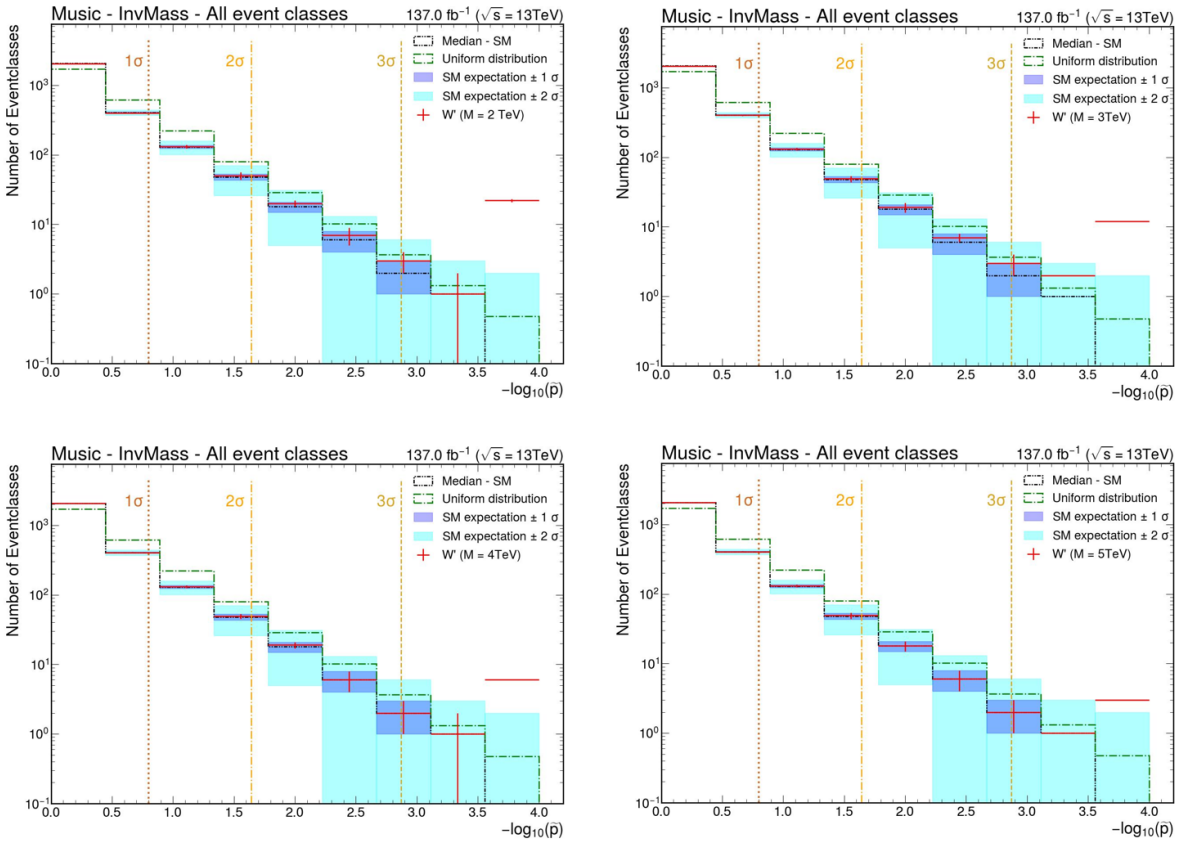


FIGURE 4.9: Histograms of the $\tilde{p}$-values for the invariant mass kinematic distribution. All W′ masses under investigation are shown.

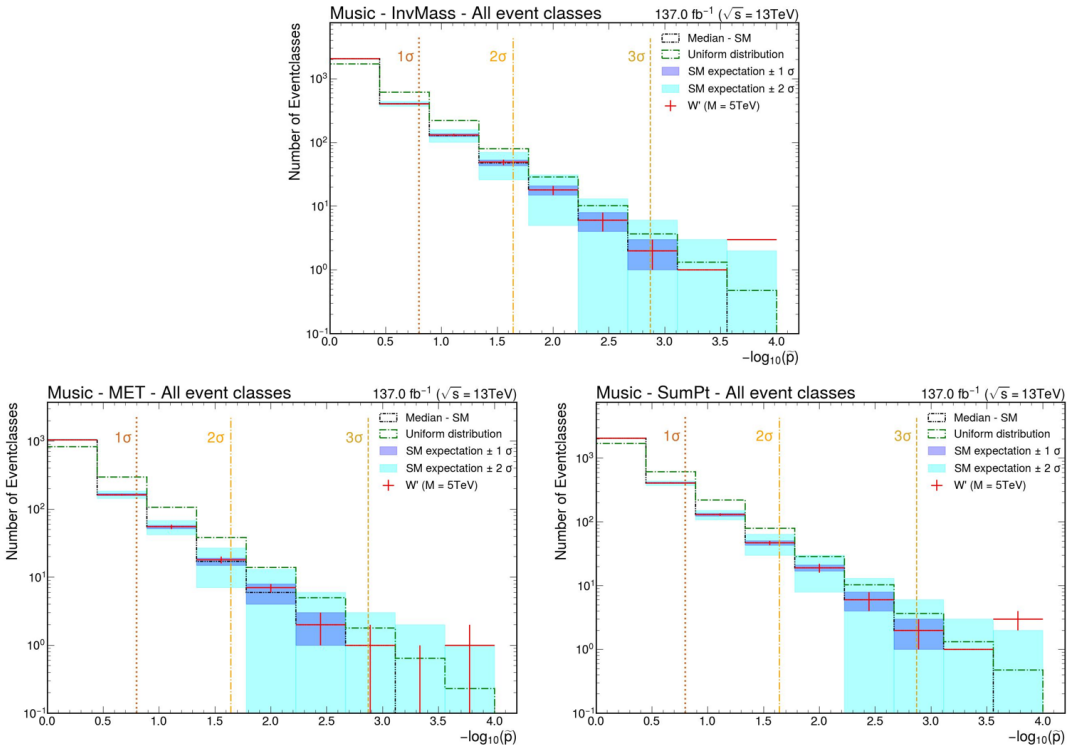


FIGURE 4.10: Histograms of the $\tilde{p}$-values for all kinematic distribution and a W′ mass of 5 TeV.

In Figure 4.10 the global $\tilde{p}$ distribution is shown for a W' mass of 5 TeV for all kinematic distributions. An excess of event class count for the lowest $\tilde{p}$ bin is observed in all distributions. The most significant deviation can be seen in the invariant mass distribution. More $\tilde{p}$-distributions are shown in the Appendix A.1.

A summary of all event classes that exceed a $\tilde{p}$ value less than 3 two-sided standard deviations is shown in Table 4.3. The significant deviations all occur in the expected event classes, which are described above. The results are compared with the one obtained by Yannik Kaiser [8]. In the invariant mass distribution, the detected classes are similar, with the exception of the "2e+MET" distribution for a mass of 2 TeV. This event class has been found in the analysis of Yannik Kaiser, but not in the analysis conducted during this thesis.

4.2.2 Sphaleron sensitivity study

The sphaleron signal investigated in this theses was generated at a sphaleron energy of E_{sph} = 8 TeV and thus the signal is expected in the high energy tails of the distributions. As discussed in chapter 2, we expect event classes with high multiplicity in leptons and jets, therefore the distributions of interest are mainly in the inclusive and jet-inclusive event classes. This thesis investigates the sphaleron signal at different pre-exponential-factors (PEF). The PEF is correlated to the cross sections σ of sphaleron processes through:

$$\sigma = \sigma_0 \cdot PEF \tag{4.13}$$

With $\sigma_0 = 121$ fb [40]. A distribution for the sphaleron signal, "$2\mu + 6jet + X$", together with the SM background, is shown in Figure 4.11, for a small pre-exponential-factor of 0.01. The transverse momentum distribution shows an excess of signal over the SM background. The deviation was found by the MUSiC algorithm, even though the signal is relatively small and in a high uncertainty region.

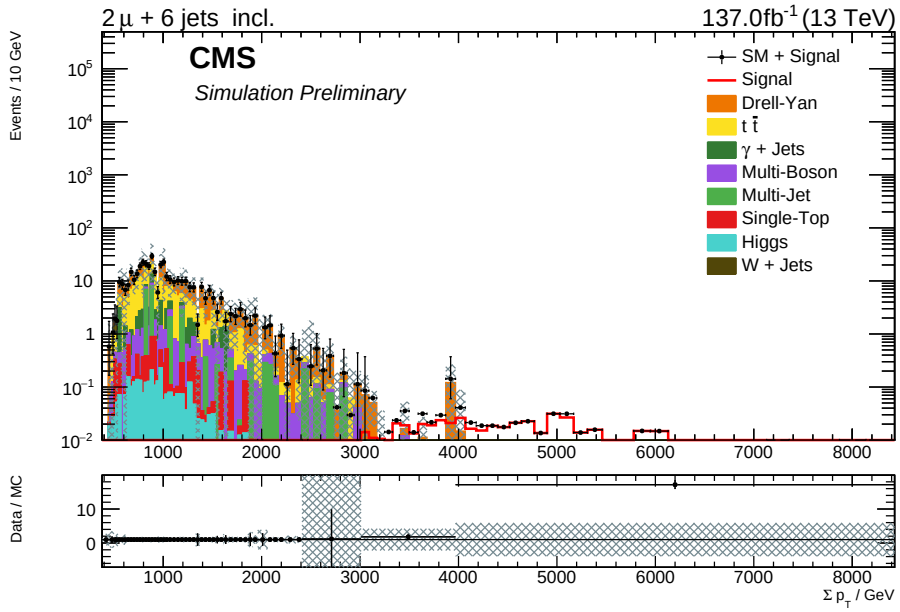


FIGURE 4.11: Transverse momentum distribution of the $2\mu + 6Jet + X$ distribution with a sphaleron ($PEF = 0.01$). The different process groups of the SM background are indicated.

The global $\tilde{p}$ distributions for the transverse momentum distribution for all sphaleron PEF's and including all event classes, are illustrated in Figure 4.12. The analytical assumption again does not fit the SM distribution perfectly.

In the transverse momentum distributions, for all sphaleron PEF's an excess of the event classes, stronger than the combined uncertainty ,can be observed in the last bin.The amount of the event classes observed in the last bin increases with increasing PEF as expected, to increase of the relative uncertainty in the high tail of the distributions. For the lowest investigated pre-exponential-factor of 0.01, four event classes have been found, which is outside of two standard deviations of the expected event class count.

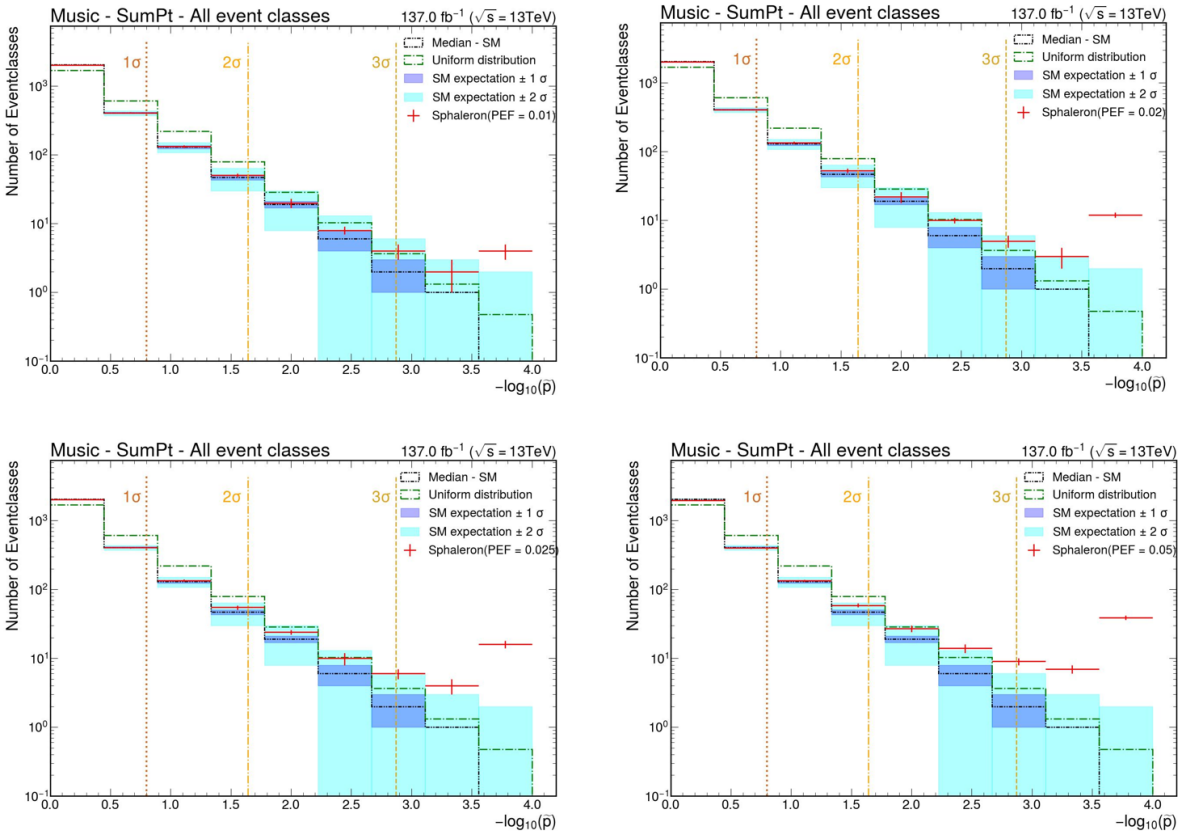


FIGURE 4.12: Histograms of the $\tilde{p}$-values for the transverse momentum kinematic distribution. All sphaleron pre-exponential-factors under investigation are shown.

In Figure 4.13 the global $\tilde{p}$ distribution for a PEF of 0.01 is shown for all kinematic distributions. An excess of event class count for the lowest $\tilde{p}$ bin is observed in all distributions. For the missing transverse momentum distribution, the excess is exactly inside two standard deviations. The most significant deviation can be seen in the transverse momentum distribution. More $\tilde{p}$ distributions are displayed in Appendix A.2.

A summary of all event classes exceeding a $\tilde{p}$-value lower than 3 two sided standard deviations, shown in Table 4.4. All significant deviations occur in the expected classes, which are described above. The results are compared to the one obtained by Yannik Kaiser [8]. In the transverse momentum distribution, the detected classes are similar with the exception of the "$1e + 1\mu + 1bJet + 1Jet + MET + X$" and "$1e + 1\mu + 3Jet + MET + X$" distributions for a PEF of 0.02. The values are likely only different due to random flucations inherent to the sampling process, leading to different classes passing the defined threshold. The exact values of the classes were evaluated and only a small difference from the threshold was detected. Furthermore, "$1e + 1bJet + 2Jet + MET + X$" and "$1e + 1\mu + 1bJet + 1Jet + MET + X$" for a PEF of 0.025 have been found in the analysis of Yannik Kaiser, but not, to the same extent, in the analysis conducted during this thesis.

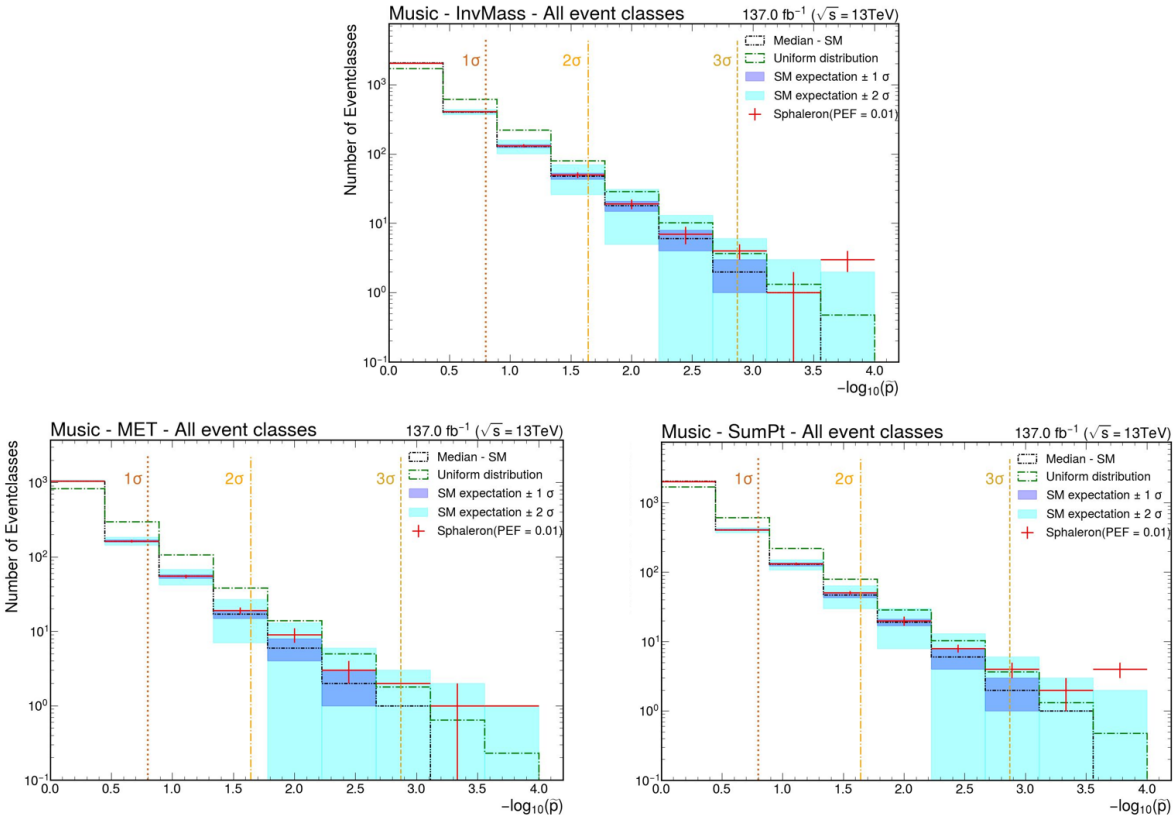


FIGURE 4.13: Histograms of the $\tilde{p}$-values all kinematic distribution and a PEF of 0.01.

Signal	M_{inv}		S_T		P_T^{miss}	
	Event class	$\tilde{p}$	Event class	$\tilde{p}$	Event class	$\tilde{p}$
W' 2TeV	1μ+1Jet+p_T^{miss}	< 0.0001	1μ+1Jet+p_T^{miss}	< 0.0001	1μ+1Jet+p_T^{miss}	< 0.0001
	1μ+p_T^{miss}	< 0.0001	+1μ+p_T^{miss}	< 0.0001	1μ+p_T^{miss}	< 0.0001
	1μ+1γ+p_T^{miss}	< 0.0001	1μ+1γ+p_T^{miss}	< 0.0001	1μ+1γ+p_T^{miss}	< 0.0001
	1e+p_T^{miss}	< 0.0001	1e+p_T^{miss}	< 0.0001	1e+p_T^{miss}	< 0.0001
	1e+1Jet+p_T^{miss}	< 0.0001	1e+1Jet+p_T^{miss}	< 0.0001	1e+1Jet+p_T^{miss}	< 0.0001
	1e+1γ+p_T^{miss}	< 0.0001	1e+1γ+p_T^{miss}	< 0.0001	1e+1γ+p_T^{miss}	< 0.0001
	1e+2γ+1Jet+p_T^{miss}	< 0.0001	1e+2γ+1Jet+p_T^{miss}	< 0.0001		
			2e+p_T^{miss}	< 0.0001		
					1μ+2Jet+p_T^{miss}	< 0.0001
					1μ+1bJet+p_T^{miss}	< 0.0001
					1e+1bJet+p_T^{miss}	< 0.0001
					1e+1bJet+1Jet+p_T^{miss}	< 0.0001
					1e+1γ+1Jet+p_T^{miss}	< 0.0001
					1e+2Jet+p_T^{miss}	< 0.0001
					1e+3Jet+p_T^{miss}	0.0003
W' 3TeV	1μ+1Jet+p_T^{miss}	< 0.0001	1μ+1Jet+p_T^{miss}	< 0.0001	1μ+1Jet+p_T^{miss}	< 0.0001
	1μ+p_T^{miss}	< 0.0001	1μ+p_T^{miss}	< 0.0001	1μ+p_T^{miss}	< 0.0001
	1e+1Jet+p_T^{miss}	< 0.0001	1e+1Jet+p_T^{miss}	< 0.0001	1e+1Jet+p_T^{miss}	< 0.0001
	1e+p_T^{miss}	< 0.0001	1e+p_T^{miss}	< 0.0001	1e+p_T^{miss}	< 0.0001
					1μ+1γ+p_T^{miss}	< 0.0001
W' 4TeV	1μ+p_T^{miss}	< 0.0001	1μ+p_T^{miss}	< 0.0001	1μ+p_T^{miss}	< 0.0001
	1e+p_T^{miss}	< 0.0001	1e+p_T^{miss}	< 0.0001	1e+p_T^{miss}	< 0.0001
W' 5TeV	1e+p_T^{miss}	< 0.0001	1e+p_T^{miss}	< 0.0001	1e+p_T^{miss}	< 0.0001
			1μ+p_T^{miss}	< 0.0001	1μ+p_T^{miss}	< 0.0001

TABLE 4.3: Table of all exclusive event classes for all kinematic distributions, which have a $\tilde{p}$-value which is smaller than a two sided uncertainty of 3σ.

Signal	M_{inv}		S_T		P_T^{miss}	
	Event class	$\tilde{p}$	Event class	$\tilde{p}$	Event class	$\tilde{p}$
PEF = 0.01	1e+6Jet+p_T^{miss}	< 0.0001	1e+6Jet+p_T^{miss}	< 0.0001		
			2μ+6Jet	0.0002		
PEF = 0.02	1e+6Jet+p_T^{miss}	<0.0001	1e+6Jet+p_T^{miss}	<0.0001		
	1μ+3Jet+p_T^{miss}	0.0005				
	1e+5Jet+p_T^{miss}	<0.0001				
	1μ+1bJet+2Jet+p_T^{miss}	0.0011	1μ+1bJet+2Jet+p_T^{miss}	0.0006		
			1μ+1bJet+3Jet	0.0004		
			1e+1μ+3Jet+p_T^{miss}	0.0011		
			1e+2bJet+1Jet+p_T^{miss}	<0.0001		
			1e+1μ+1bJet+3Jet	0.0004		
			1e+1μ+5Jet	<0.0001		
			2μ+6Jet	0.0002		
					1μ+5Jet+p_T^{miss}	<0.0001
PEF = 0.025	1e+6Jet+p_T^{miss}	<0.0001	1e+6Jet+p_T^{miss}	<0.0001		
	1μ+3Jet+p_T^{miss}	<0.0001				
	1μ+1bJet+2Jet+p_T^{miss}	<0.0001	1μ+1bJet+2Jet+p_T^{miss}	<0.0001		
	1e+5Jet+p_T^{miss}	<0.0001				
	1e+1μ+2bJet+4Jet+p_T^{miss}	0.0012				
			1μ+1bJet+2Jet	0.0006		
			1μ+1bJet+3Jet	<0.0001		
			1e+1μ+2Jet	0.0006		
			1e+1μ+3Jet+p_T^{miss}	<0.0001		
			1e+2bJet+1Jet+p_T^{miss}	<0.0001		
			1e+2bJet+3Jet	0.0004		
			1e+1μ+1bJet+3Jet	0.0002		
			1e+1μ+5Jet	<0.0001		
			2μ+6Jet	<0.0001		
					1μ+5Jet+p_T^{miss}	<0.0001
					1e+1bJet+3Jet+p_T^{miss}	0.0004

TABLE 4.4: Table of all inclusive event classes for all kinematic distributions, which have a $\tilde{p}$-value which is smaller than a two sided uncertainty of 3σ. The "+X" has not been written in the Event class.

5 Jensen Shannon Divergence

To evaluate the deviation from CMS data and Monte Carlo Standard Model simulations, a search algorithm which takes the whole distribution into account could be interesting. Because of the global nature, no Look elsewhere effect is present and no RoI scan needs to be performed. Therefore, it is expected that the global method has a shorter calculation time, which could be comparable to the MUSiC integral scan.

5.1 Theory

5.1.1 Probability density function

To evaluate if two kinematic distributions are globally compatible, we approximate the underlying probability density function (PDF). To achieve this, first the counts N in each bin $N(x)$ are divided by the width of the bin $BW(X)$. To get to a probability, we have to normalise the counts through dividing by the sum of the counts. The probability for each bin $p(x)$ is evaluated as:

$$p(x) = \frac{N(x)}{BW(X)} \cdot \left(\sum_{x \in \chi} N(x)/BW(x) \right)^{-1} \tag{5.1}$$

where χ is the sample space, in our case our histogram bins.

5.1.2 Kullback–Leibler divergence

The Neyman–Pearson lemma, states that the most powerful method to correctly evaluate if a test sample Y is drawn from distribution P or Q is the ratio of their log likelihoods [83]:

$$H = log(P) - log(Q) = log\left(\frac{P(x)}{Q(x)}\right) \tag{5.2}$$

The Kullback–Leibler divergence $D_{KL}(P||Q)$ is defined by the expected value of the method provided by the Neyman–Pearson lemma, given that the test sample is drawn from P [84]:

$$D_{KL} = \sum_{x \in \chi} P(x) \cdot log\left(\frac{log(P(x))}{log(Q(x))}\right) \tag{5.3}$$

The KL-divergence is known in statistics as a measure of the dissimilarity of a probability distribution Q from a reference distribution P. The Kullback-Leibler divergence is always positive and no upper bounds exist. The smaller the Kullback-Leibler divergence, the more similar the two distributions are[83, 84].

5.1.3 Jensen Shannon divergence

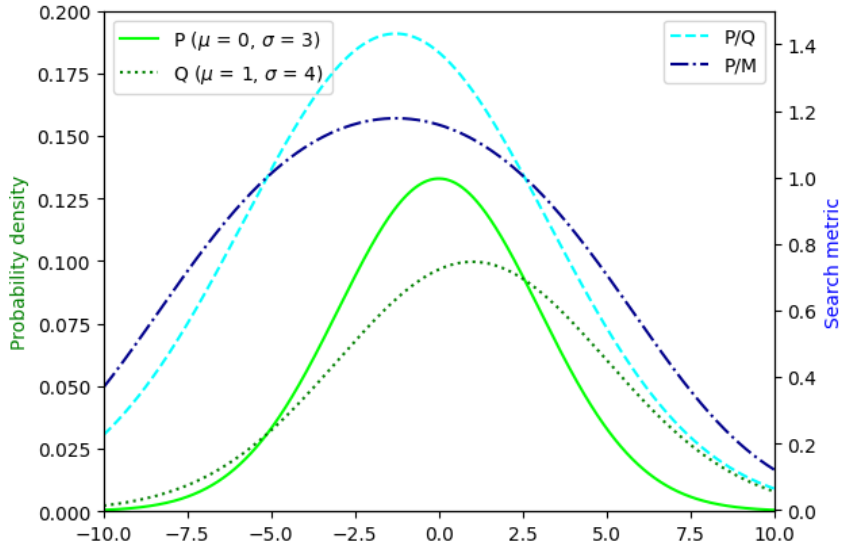
The Kullback–Leibler divergence has two properties, which are not desired in a model independent search. First, the Kullback-Leiber divergence is not symmetric and discriminates between the reference distribution P and the tested distribution Q. In addition, for the comparison between a non-zero probability of the reference distribution $p(x) \geq 0$ and a zero probability of the test distribution the Kullback-Leiber divergence diverges.
This is solved by the Jensen–Shannon divergence (JSd) [85]. The Jensen-Shannon divergence measures the dissimilarity of two PDFs and is defined as [86]:

$$JSd(P||Q) = \frac{1}{2}D_{KL}(P||M) + \frac{1}{2}D_{KL}(Q||M) \tag{5.4}$$

where $M = (P+Q)/2$ is a mixture distribution of P and Q. As defined in equation 5.4, the Jensen–Shannon divergence is symmetric and non divergent. To gain the interpretation of distance between two PDF's the Jensen–Shannon Distance (JSD) is calculated:

$$JSD = \sqrt{JSd} = \sqrt{\frac{1}{2}D_{KL}(P||M) + \frac{1}{2}D_{KL}(Q||M)} \tag{5.5}$$

The JSD is non zero and bounded by $0 \geq JSD(P||Q) \geq 1$. The main idea, why we would prefer JSD over the Kullback–Leibler divergence, is that we profit off the mentioned properties, as well as that if a deviation is strong enough, the deviation from the mean M should present a strong deviation as well, as illustrated in figure (5.1).



0.9

FIGURE 5.1: Two normal distributions, with different means(μ) and different uncertainty(σ) are displayed and the difference between the quotient of the two distributions and quotient between one distribution and the mean are illustrated.

Another idea why JSD might fit an model unspecific analysis is, because the Kullback–Leibler divergence will be positive and negative for different evaluations of P and Q in each bin. Therefore, the up and down variation caused by the uncertainty in the measurement cancel out. This results in a small contribution for similar bins between p and q, but highlights the difference, which itself presents a behaviour similar to a RoI scan.

5.1.4 JSD P-Value

In order to calculate a p value, which describes the likelihood of a deviation at least as likely as the observed one, we need to transform the search metric. To include all uncertainties, we follow the sample sampling procedure as MUSiC for the post trial probability (described in Sections 4.1.3.4 and 4.1.3.5).
For this thesis, toys for the SM background and samples for the signals are produced, and their corresponding JSD value is calculated. With this, the p-value can be calculated:

$$p_{JSD} = \begin{cases} \frac{N_{pseudo}(JSD_{pseudo} \geq JSD_{data})}{N_{pseudo}} & , N_{pseudo}(JSD_{pseudo} \geq JSD_{data}) > 0 \\ \frac{1}{N_{pseudo}} & , N_{pseudo}(JSD_{pseudo} \geq JSD_{data}) = 0 \end{cases} \tag{5.6}$$

In figure 5.2 a distribution of JSD values is shown for an exponential distribution. A roughly Gaussian like shape is recognisable, which is different from the MUSiC example shown in figure 4.4.

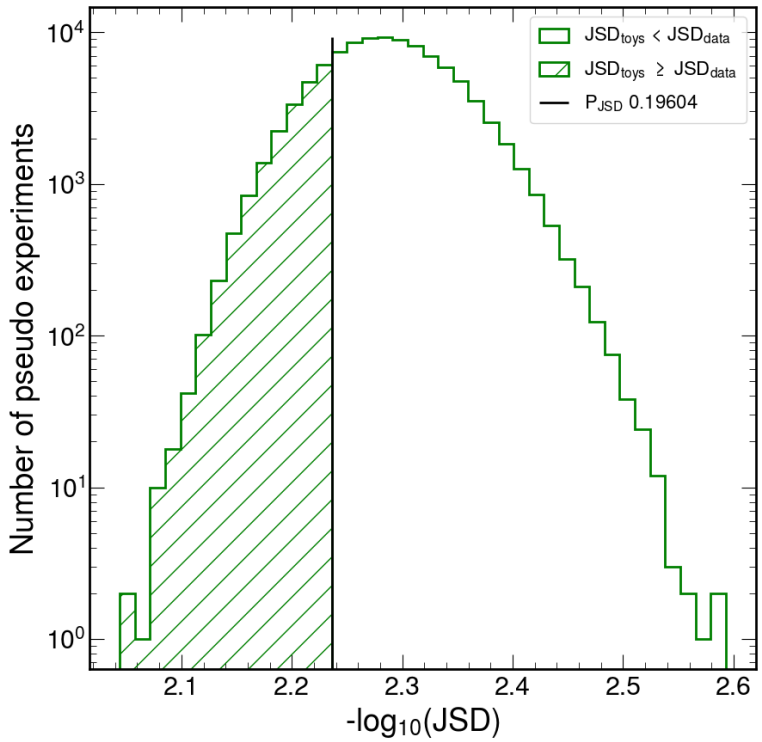


FIGURE 5.2: Illustration of the distribution of the sampled JSD-values of the pseudo experiment. The JSD value for the observed deviation is marked by the black line, and pseudo experiments with a stronger deviation are marked with a hacked area.

5.2 Evaluation with pseudo distribution

To better understand the JSD algorithm and confirm the desired behaviour, a first study with controlled pseudo samples is conducted.

5.2.1 Distributions

For the evaluation in a controlled environment we build our histograms based on a exponential distribution $f_{exponential}(x, \lambda)$:

$$f_{exponential}(x, \lambda) = \begin{cases} \frac{1}{\lambda} \cdot e^{-\frac{x}{\lambda}} & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{5.7}$$

Here λ is called scale parameter and is the mean value. This distribution has been chosen, because it approximates large parts of actual kinematic distributions relatively well (see 4.8) .Then a positive or negative Gaussian signal is added as illustrated in figure 5.3.

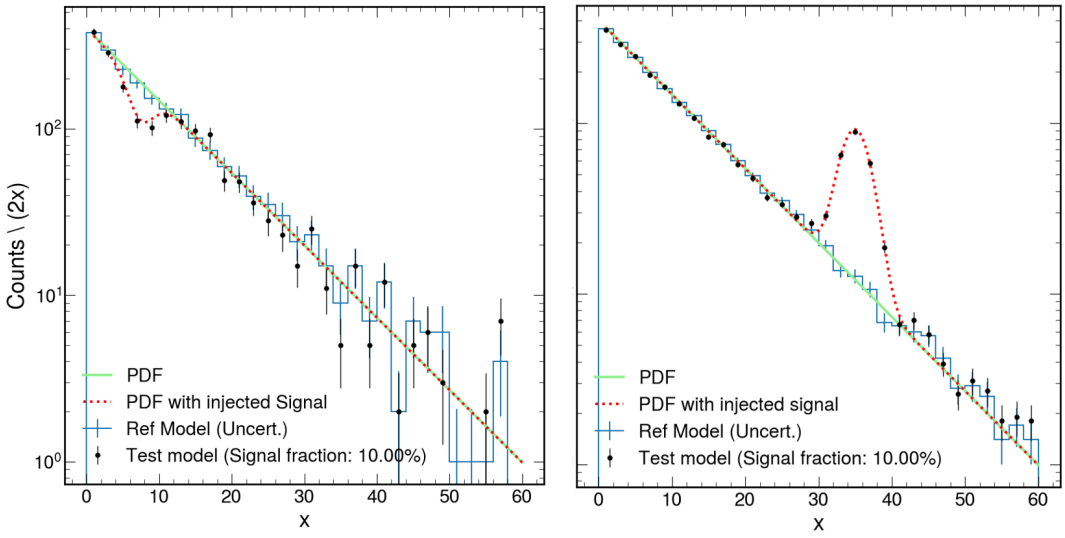


FIGURE 5.3: Examples of the exponential distributions, with corresponding uncertainty and an injected signals. 2000 Events where drawn from an exponential distribution and with a signal fraction of 10% 200 Events where drawn from a Gaussian distribution. For the left(right) figure the Gaussian was centered around $\mu = 7$($\mu = 35$) and an uncertainty of $\sigma = 2$($\sigma = 2$). The resulting observation with the uncertainty is shown in black. The Exponential PDF is illustrated in green and the PDF injected signal with the exponential PDF in red.

The uncertainty must be defined to use the full MUSiC framework. Statistical uncertainty is defined as the uncertainty expected from the Poison distribution $\sigma_{stat} = \sqrt{N}$, where N is the number of counts in a bin. In addition, a set of systematic uncertainty's must be defined in order to use the MUSiC sampling. We therefore define three arbitrary pseudo systematic uncertainties as:

$$\sigma_{syst,i} = \frac{\sigma_{stat}}{i+4} \tag{5.8}$$

Here $\sigma_{syst,i}$ is the i-th systematic uncertainty. The uncertainty is then smeared with a normal distribution with an expected value of $\mu = 1$ and an uncertainty $\sigma = 1$.

5.2.2 1D Evaluation

To evaluate the sensitivity of MUSiC and JSD we compare the MUSiC $\tilde{p}$-value and the JSD p-value for varying sample sizes, which models a decrease in uncertainty comparable to the increase in the integrated luminosity. For this we sample 15,000 toys and 3,000 signal samples, from an exponential distribution, modelled with 30 bins and a scale parameter of $\lambda = 10$. The sample size describes the number of events drawn from the exponential distribution. Additionally, a percentage of the sample size, the signal fraction, is added as the signal.
In Figure 5.4 a comparison can be seen between the MUSiC $\tilde{p}$ and the JSD p value for different signal fractions. It is shown that the $p/\tilde{p}$ decreases with increasing sample size. This is expected, since the uncertainty corresponds to the sample size and therefore a deviation with given relative strength becomes more significant. In addition, it can be seen that the slope at which the MUSiC $\tilde{p}$ and the JSD p value decreases, increases with higher signal fraction, as expected.
For this example, it can be observed that the JSD p value is slightly but consistently lower than the value of MUSiC $\tilde{p}$. Given that the uncertainty is large and the behaviour was not always consistent between different signals, no significant deviation is observed and both calculations share a large overlap.

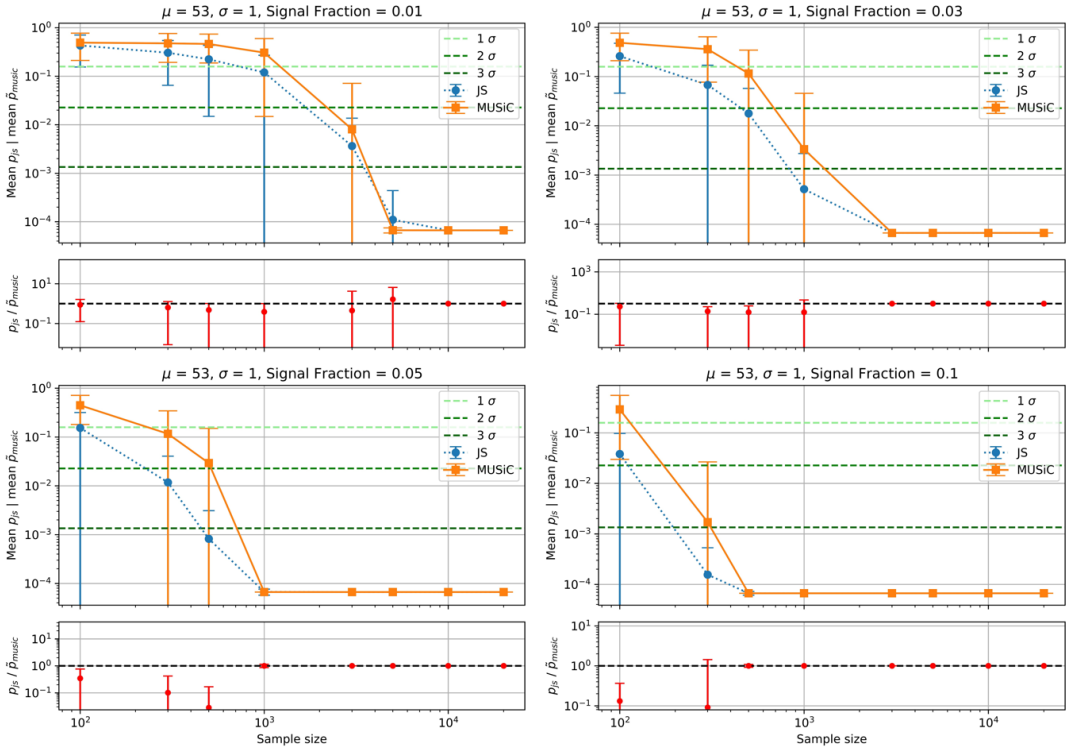


FIGURE 5.4: The calculated $\tilde{p}$ for MUSiC and p for JSD is plotted against the sample size. Additionally a plot is shown below each figure, which shows the quotient between the two metrics. Four plots are shown at the same expectation value and same uncertainty for the injected normal distribution, but with different signal fraction. The different signal fractions and the parameters of the injected signal are shown in the title

In Figure 5.5 a comparison between MUSiC $\tilde{p}$ and JSD p is illustrated for different signal standard deviations. The broader signal causes a worse separation between background and signal and therefore leads to a reduced sensitivity. This is observed for both MUSiC and JSD.

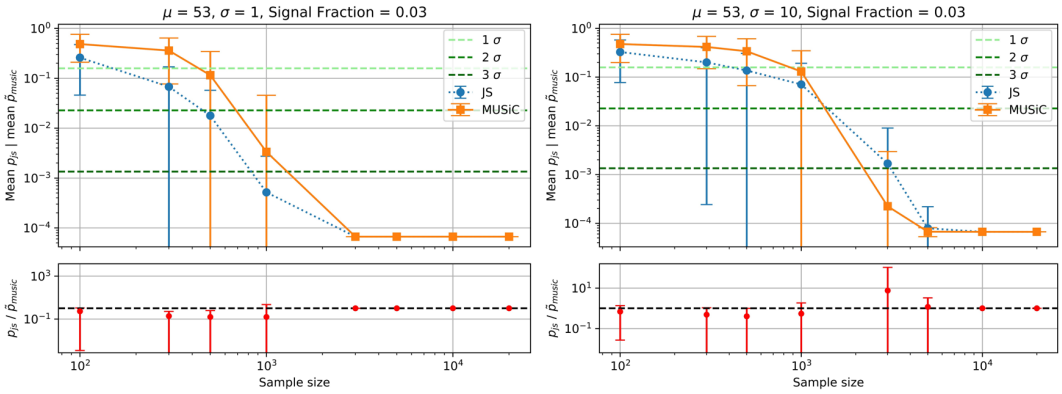


FIGURE 5.5: The calculated $\tilde{p}$ for MUSiC and p for JSD is plotted against the sample size. Additionally a plot is shown below each Figure, which shows the quotient between the two metrics. Two plots are shown at the same expectation value and same signal fraction for the injected normal distribution, but with different uncertainty. The signal fractions and the parameters of the injected signal are shown in the title

In Figure 5.6 a comparison between the MUSiC $\tilde{p}$ and the JSD p value is shown for two different signal positions. Because the ratio between the signal and the background increases with the expectation value, for a fixed signal fraction, we expect the deviation with higher expectation value to be more significant. This is also observed for both MUSiC and JSD.

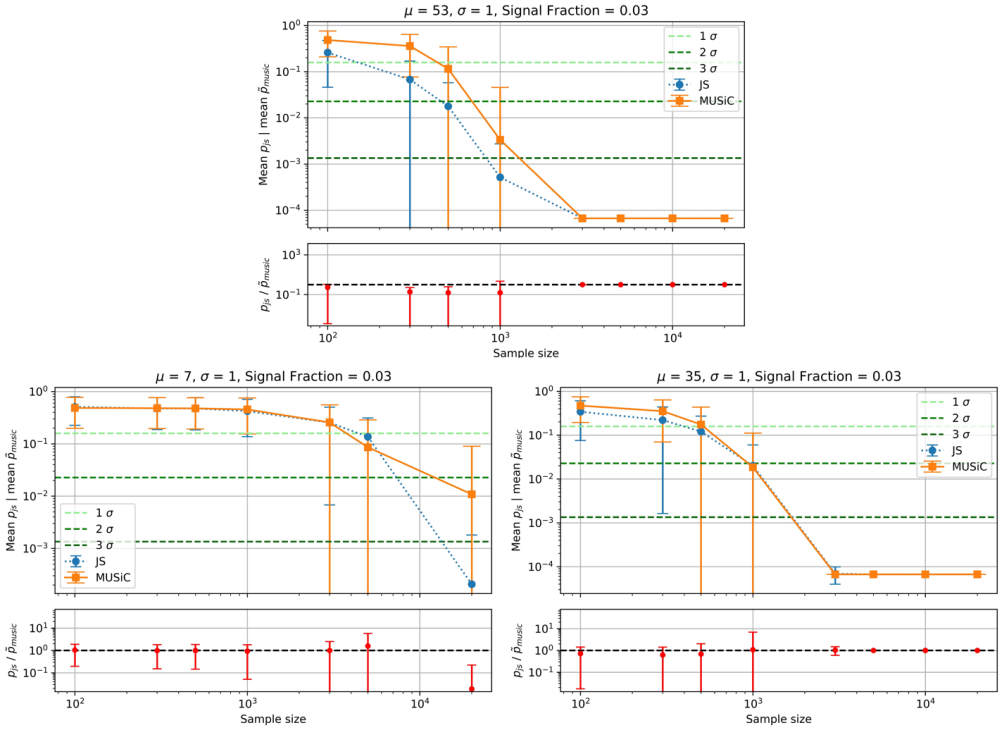


FIGURE 5.6: The calculated $\tilde{p}$ for MUSiC and p for JSD is plotted against the sample size. Additionally a plot is shown below each figure, which shows the quotient between the two metrics. Three plots are shown at three different expectation value and the same uncertainty and signal fraction for the injected normal distribution . The signal fractions and the parameters of the injected signal are shown in the title

In Figure 5.7 a comparison between the MUSiC $\tilde{p}$ and the JSD p value for a constructive and destructive signal is illustrated. We would generally expect a similar behaviour for positive and a negative signal, which can be observed for both MUSiC and JSD. It can be seen that the last $\tilde{p}$-value in both distributions is higher than the JSD p-value.

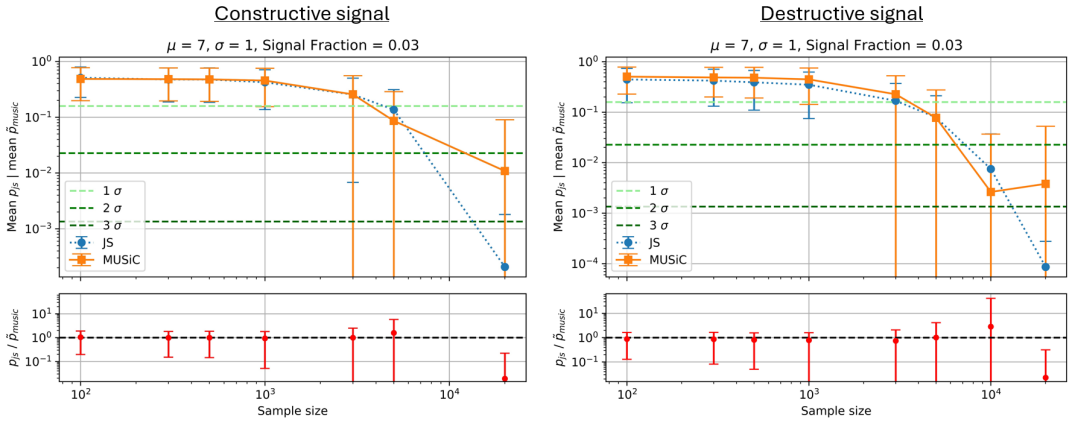


FIGURE 5.7: The calculated $\tilde{p}$ for MUSiC and p for JSD is plotted against the sample size. Additionally a plot is shown below each Figure, which shows the quotient between the two metrics. Two plots are shown at the same expectation value, same uncertainty and signal fraction for the injected normal distribution. On the left a constructive Signal is injected, while on the right a destructive signal is added. The signal fractions and the parameters of the injected signal are shown in the title

5.2.3 2D Evaluation

To evaluate the potential of a multidimensional scan, a comparison of the result of a one-dimensional scan to a two-dimensional scan, based on the Jensen-Shannon divergence, was made. A similar procedure as previously is conducted, sampling pseudo events, for which the X coordinate and the Y coordinate are assumed to be uncorrelated, from an exponential distribution with a scale parameter $\lambda = 10$. In the same way, we sample a fraction of the events with uncorrelated Gaussians for X and Y and add them to the background. We then perform three p-value calculations, one for each dimension and one taking both dimensions into account. In this process, we grouped the data into five bins per dimension.
This is illustrated in Figure 5.8, where we can see two samples of 2D distributions and their corresponding evaluation. For a symmetric signal, a stronger sensitivity and steeper slope, exceeding the combined uncertainty can be noticed. For the signal with the unsymmetrical expectation value of $\mu = [53, 35]$ a lower p-value is estimated by the one dimensional JSD where the deviation is at the tail.

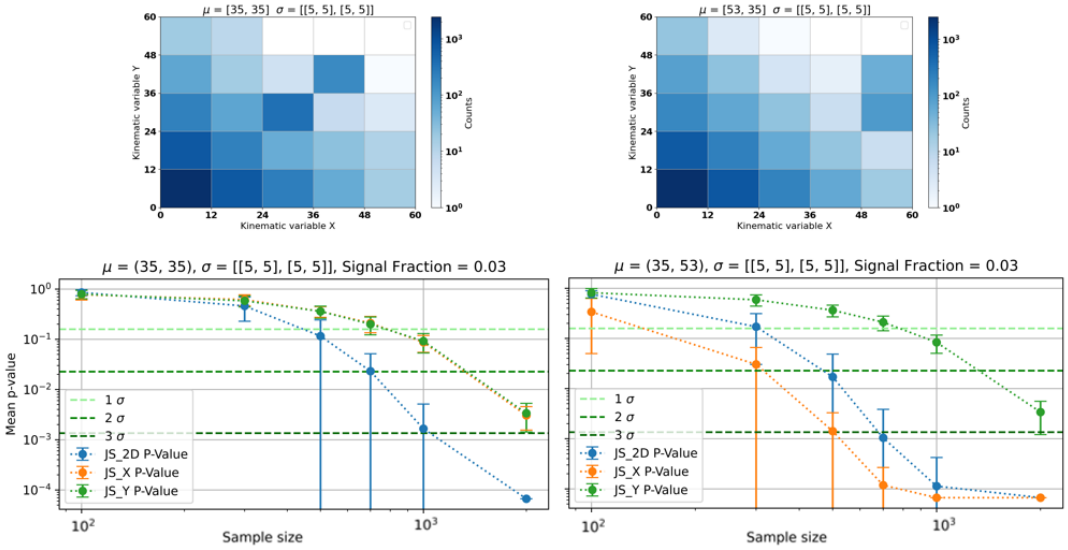


FIGURE 5.8: A sample distribution for 5.000 events sampled from an exponential distribution and injected signal with signal fraction of 3% is shown in the upper plots. Shown below is their corresponding evaluation. Two one dimensional evaluation and a two dimensional evaluation is shown.

5.3 Evaluation on simulated data

To evaluate and compare the JSD algorithm, a sensitivity study is performed. The same procedure and samples as discussed in Chapter 4 is used.

5.3.1 Merging and Veto

The higher complexity of real event classes can lead to unwanted effects. For example, classes without injected signal can be misdetected as deviating from the Standard Model through the sampling process. If the relative uncertainty is high or only a few bins are occupied, the distribution will change strongly through the sampling, which can cause the JSD or MUSiC algorithm to detect an event class, which is in accordance with the Standard Model, to appear as deviating. An example of the resulting p-value distribution is shown in Figure 5.10. Some classes that caused problematic behaviour are illustrated in Figure 5.9. On the left side an example with a small MC yield is shown. The second example shows a very high uncertainty. This is due to an erroneous calculation of the muon resolution, which can cause a relative uncertainty of more than 100.
To solve these issues, a merging algorithm has been developed to ensure that only bins of high quality, defined by the event yield and the relative uncertainty, are evaluated. The merging algorithm starts from the right side merging towards the left side. Therefore, single bins with high relative uncertainty can still be evaluated. The merging occurs in multiple steps:

1. We merge until we reach a Monte Carlo yield of 0.1. This is intended to guarantee that enough information is obtained in the simulation process. This is not the complete truth, since we work with scaled Monte-Called and the unweighted Monte-Carlo event yield preferable, but was not implemented due to time constraints.

2. Then we examine the relative systematic uncertainty. If the relative uncertainty is below one, we add the bin to our final distribution. If the relative systematic uncertainty is between one and five, we continue merging. If the relative systematic uncertainty is above five, we have to discard the bin. This has been done, because of the faulty calculated uncertainty's, which would lead all later bins having a too high uncertainty.

3. At last, it is checked if the resulting distribution has at least five bins, in order to analyse a full distribution.

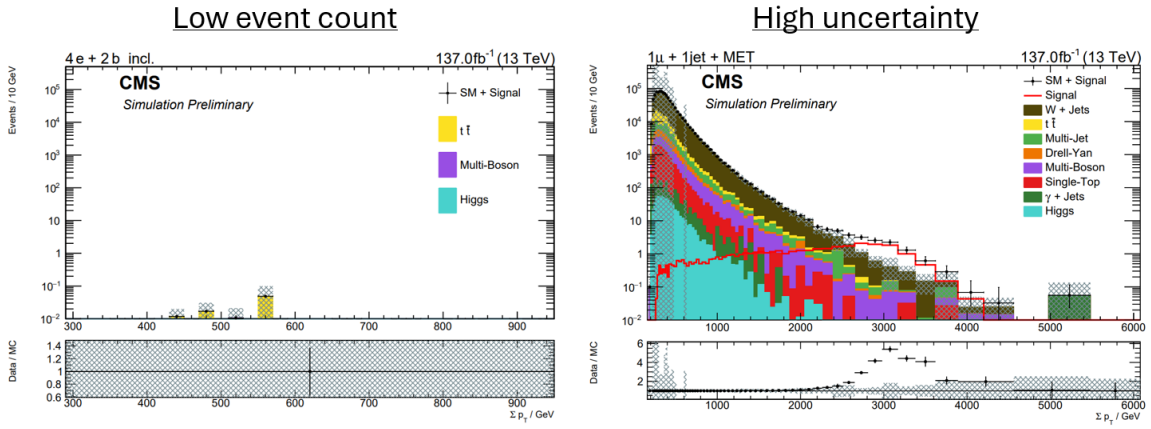


FIGURE 5.9: Kinematic distributions to illustrate problematic conditions, which can lead to faulty calculations. In the left plot a very low occupied event class is shown, which is not robust against sampling. On the right a plot is shown, where in the low sum of transverse momentum region high uncertainties are observed.

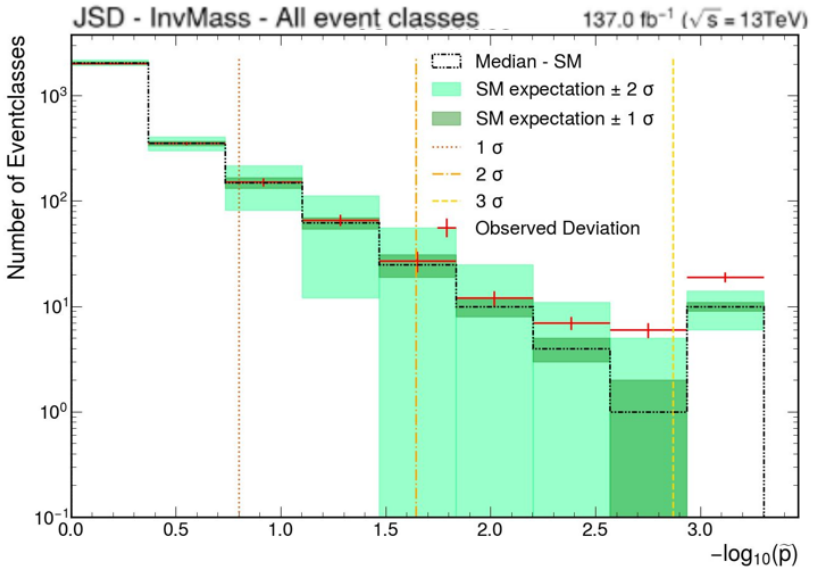


FIGURE 5.10: P-value distribution for event classes without a veto system. A high amount of strongly deviating event classes can be observed for the SM expectation.

5.3.2 Evaluation

The Jensen-Shannon Distance is evaluated on the same CMS simulation samples as previously MUSiC. First, the two-dimensional comparison graph of the JSD p values and the MUSiC $\widetilde{p}$ values in Figure 5.11 must be investigated to evaluate and understand the distribution of the corresponding metric and to see the correlation between JSD and MUSiC. In the histograms, the first bin of MUSiC and Jensen-Shannon Distance has a high amount of event classes due to the vetoing. For $\widetilde{p}$ values between 0.6 and 0.9 MUSiC shows a relatively constant number of event classes. Around 0.5 a peak is observed for MUSiC. The last bins are again more occupied because there is a signal present. Jensen Shannon generally shows a similar behaviour, but the second bin is empty, and it shows significantly less events in the last bin, with the highest deviation and a more dominant peak around 0.5.
In the lower left quadrant, where the p-value is low, indicating a deviation, little correlation between MUSiC and Jensen Shannon is visible. The amount of event classes in the last bin is significantly greater for MUSIC than for JSD. This can also be observed in Figure 5.12, where the global distributions of p / $\widetilde{p}$ distributions are shown. While for MUSiC a clear deviation can be observed, JSD does not show a significant difference between SM expectation and signal, except for the missing transverse momentum distribution, where a small excess is visible.
The sensitivity of the JSD search algorithm for the CMS simulations is significantly lower compared to MUSiC. Since in the controlled environment MUSiC and JSD showed similar sensitivity, the difference between the simulations and the controlled environment must be evaluated. In addition to the more complex shape of the distributions and the signal, the main difference between the simulations and the controlled environment is the existence of a systematic uncertainty.

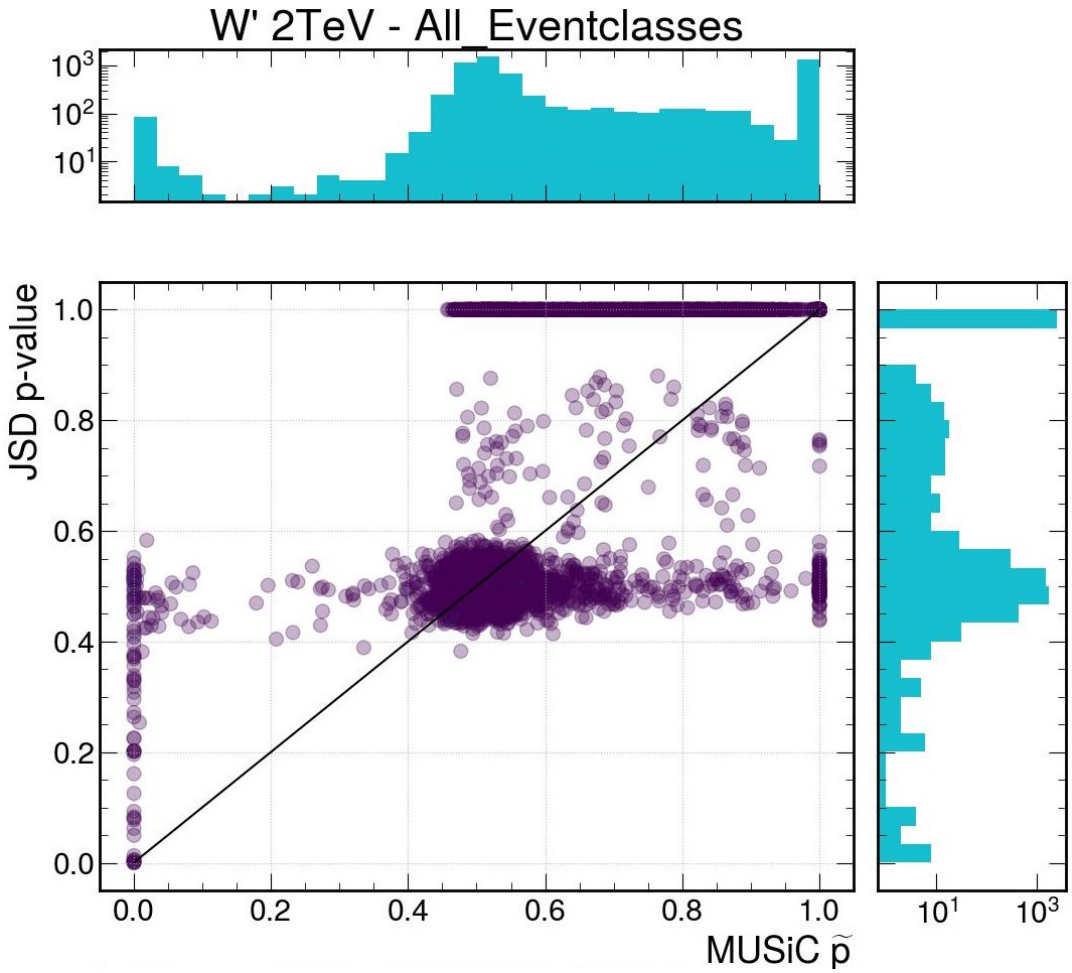


FIGURE 5.11: Illustration of the distribution of the median JSD p-values and MUSiC $\widetilde{p}$-values. In the middle Figure each data point represents an event class in one of the three kinematic distribution. In both axes a histogram of the values is shown, representing the $p/\widetilde{p}$ distribution. This example is for all event classes, all kinematic distributions and for the W′ ($M_{W'}$ = 2 TeV) example

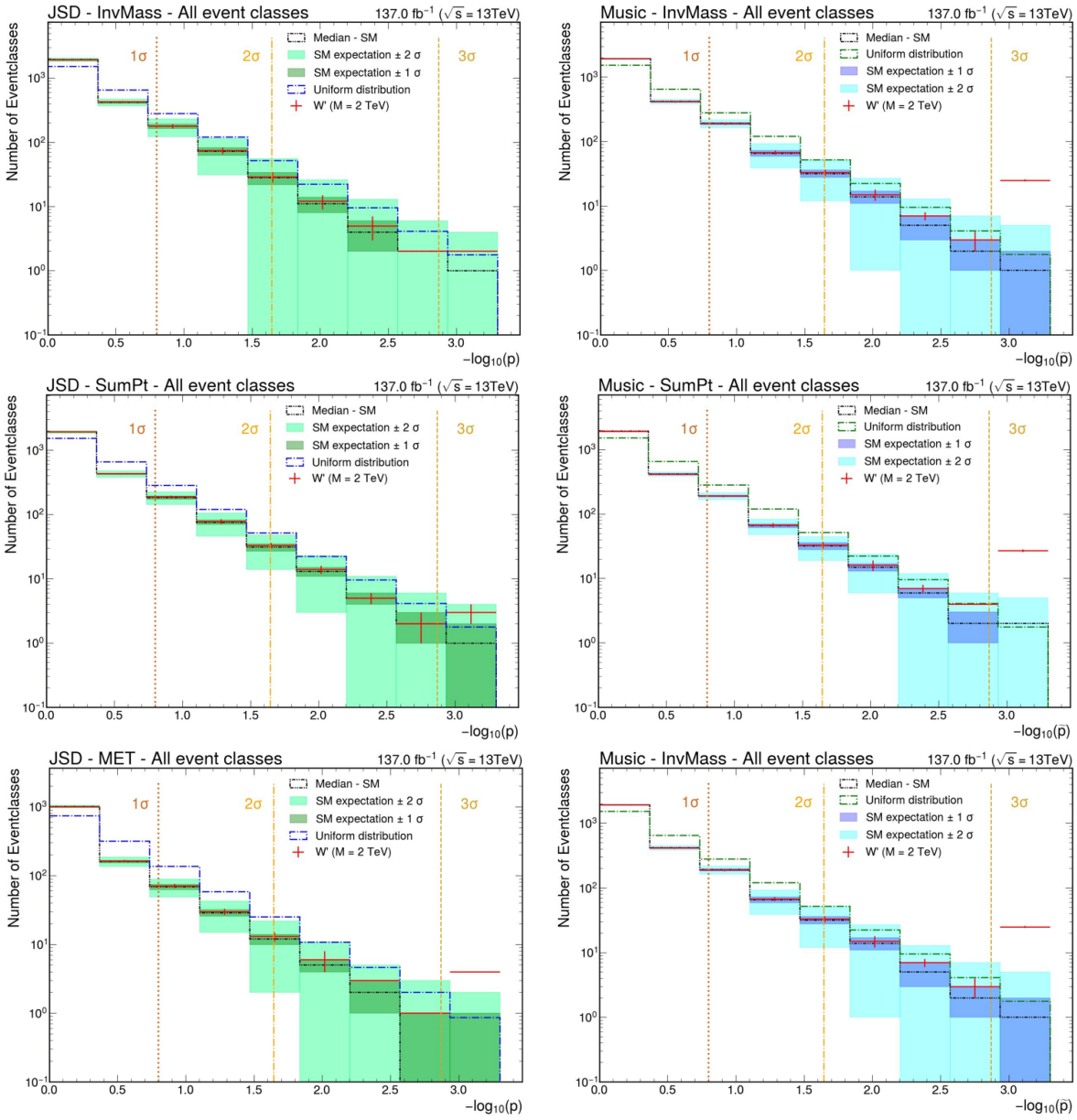


FIGURE 5.12: Comparison of the Histograms of the global $\tilde{p}$ and p distributions for all kinematic distributions and the W′ ($M_{W'}$ = 2 TeV) samples. The green distributions correspond to the regularised JSD evaluation and the blue distributions to the MUSiC evaluation.

6 Adjusted Jensen Shannon Divergence

To incorporate the systematic uncertainty to the Jensen-Shannon Distance, we add a weight factor proportional to the total relative uncertainty, which we will refer to as regularisation. In order to keep the symmetry of JSD, the regularisation is performed at the level of the Kullback-Leibler-divergence. We use the total uncertainty of the Standard Model σ_{SM} weighted by the counts in the corresponding bin of the P sample N_p. The idea is to decrease the importance of the bin according to the total uncertainty weighted by the events in the bin. Therefore, bins with a strong deviation relative to the uncertainty become more important.

$$D_{KL}^{\mathcal{R}}(P||Q) = \sum_{x \in \chi} e^{-\alpha \cdot \frac{\sigma_{SM}(x)}{N_P(x)}} \cdot P(x) \log\left(\frac{P(x)}{Q(X)}\right) \tag{6.1}$$

Here α is a hyperparameter that scales the strength of the regularisation and that must be optimised. This implementation has the desirable property, that if the uncertainty would vanish, we would return to the original Jensen-Shannon Distance:

$$\lim_{\sigma_{SM} \to 0} D_{KL}^{\mathcal{R}}(P||Q) = D_{KL}(P||Q) \tag{6.2}$$

With the adjusted Kullback-Leibler-divergence, the regularised Jensen Shannon Distance is defined as:

$$JSD^{\mathcal{R}} = \sqrt{\frac{1}{2} D_{KL}^{\mathcal{R}}(P||M) + \frac{1}{2} D_{KL}^{\mathcal{R}}(Q||M)} \tag{6.3}$$

6.1 Evaluation on simulated Data

6.1.1 Hyperparameter optimisation

To roughly estimate the hyperparameter in order to conduct further studies, α is varied and we compare to the amount of event classes, deviating more than a 2 sided 3σ. This is shown in Figure 6.1, for all studied variations of W' and Sphaleron. The number of JSD event classes show a steep slope from low hyper parameter, reaching a relatively constant area around $\alpha \approx 500$ - 5000 and then decreasing steeply again. The position of the plateau region is roughly similar for all signal samples and the large covered α interval by the plateau show the robustness of the method.
A better estimation of the hyperparameter needs a broader range of signal samples, to test robustness against different signal shapes, as well as repeated calculations per α to estimate the uncertainty on the metric.
Given that this way to determine the hyperparameter is still to some degree model dependent, future iterations of this work could profit from the development of a model independent metric to evaluate α.

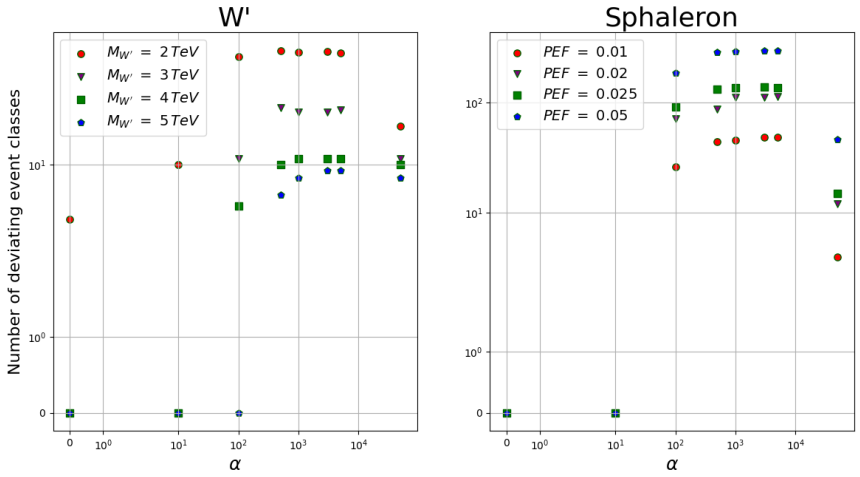


FIGURE 6.1: Comparison of the hyper parameter and the amount of event classes deviating more than three two sided standard deviations. The number of JSD event classes is shown for all W' and sphaleron samples investigated in this thesis.

6.1.2 Overview

Regularised JSD is evaluated on the CMS simulation samples. First, the two-dimensional comparison plot of the JSD p values and the MUSiC $\widetilde{p}$ in Figure 6.2 is investigated. The plot shows that both methods have a large amount of vetoed event classes, as can be observed in the first bins of the histogram. A strong difference between MUSiC and the adjusted JSD is observed for high p values in the area of approximately 0.5 - 1. While MUSiC shows a relatively constant number of event classes as expected, the regularised JSD shows nearly no event classes in that area, deviating from the new hypotheses expectation. From 0.5 - 0, the histograms look rather similar, while MUSiC has fewer events between the lowest p-value and 0.5. Generally, a stronger correlation can be observed, especially in the region of low p-values.

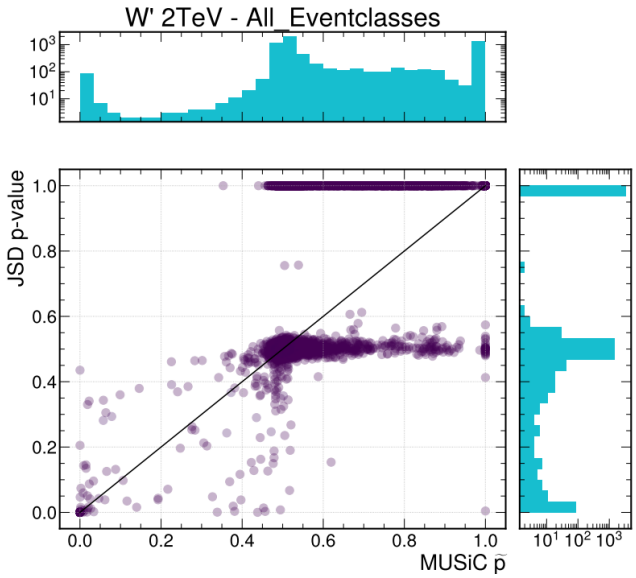


FIGURE 6.2: Illustration of the distribution of the median JSD p-values and MUSiC $\widetilde{p}$-values. In the middle figure, each data point represents a kinematic distribution of an event class. On both axes, a histogram of the values is shown, representing the $p/\widetilde{p}$ distribution. This example is for all event classes, all kinematic distributions and for W' ($M_{W'}$ = 2 TeV)

JS / MUSiC / Overlap	Invariant mass	SumPT	MET	Total
W' 2 TeV	19 / 18 / 18	20 / 20 / 20	28 / 30 / 25	67 / 68 / 63
W' 3 TeV	9 / 13 / 9	7 / 11 / 7	8 / 11 / 7	24 / 35 / 23
W' 4 TeV	4 / 5 / 4	4 / 4 / 4	3 / 2 / 2	11 / 11 / 10
W' 5 TeV	4 / 3 / 3	3 / 3 / 2	2 / 2 / 2	9 / 8 / 7
Sphaleron 1.0 %	16 / 1 / 1	32 / 1 / 0	0 / 0 / 0	48 / 2 / 1
Sphaleron 2.0 %	42 / 4 / 3	64 / 6 / 3	5 / 0 / 0	111 / 10 / 6
Sphaleron 2.5 %	52 / 7 / 6	78 / 11 / 8	8 / 1 / 1	138 / 19 / 15
Sphaleron 5.0 %	115 / 23 / 20	155 / 32 / 29	24 / 6 / 3	294 / 61 / 52

TABLE 6.1: Number of event classes exceeding a $p \mid \widetilde{p}$ of more than a two sided 3 σ. This table is classified per kinematic distribution and displays the results for all examined signal samples. The overlap describes the case where a class deviates more than the threshold for MUSiC and JSD.

JS / MUSiC / Overlap	Exclusive	Inclusive	Jet inclusive	Total
W' 2 TeV	22 / 25 / 21	23 / 22 / 21	22 / 21 / 21	67 / 68 / 63
W' 3 TeV	9 / 13 / 9	8 / 11 / 7	7 / 11 / 7	24 / 35 / 23
W' 4 TeV	6 / 6 / 6	2 / 2 / 2	3 / 3 / 2	11 / 11 / 10
W' 5 TeV	6 / 4 / 4	1 / 2 / 1	2 / 2 / 2	9 / 8 / 7
Sphaleron 1%	0 / 0 / 0	39 / 2 / 0	9 / 0 / 0	48 / 2 / 1
Sphaleron 2%	0 / 0 / 0	112 / 19 / 15	26 / 0 / 0	138 / 19 / 15
Sphaleron 5%	0 / 0 / 0	221 / 50 / 42	73 / 11 / 10	294 / 61 / 52

TABLE 6.2: Number of event classes exceeding a $p \mid \widetilde{p}$ of more than a two sided 3 σ. This table is classified per event class type and displays the results for all examined signal samples.Overlap describes the case where a class deviates more than the threshold for MUSiC and JSD.

The amount of event classes with a deviation of more than 3 σ are displayed in Table 6.1 and in Table 6.2. In Table 6.1 the event classes are categorised by the kinematic distribution. Generally, most of the classes found by MUSiC, were found as well by JSD. The total overlap per signal sample ranges from 50%, where one of two deviating classes was not found, to 96 %, with a mean coverage of 76%. It can be observed that the difference between the overlap and the event classes found by MUSiC is significantly higher for the Sphaleron samples than for the W' sample. It can be further noted that the number of event classes for W' is relatively similar for both MUSiC and JSD, while JSD finds more event classes for sphaleron samples. This will be evaluated in more detail in Section 6.1.4.
In total, 819 exclusive, 952 inclusive and 901 Jet inclusive classes are present in the analysis. In Table 6.2 the deviating classes are separated by their event class type. While for the W' samples a relatively even split between the event class type is observed, for sphaleron, no exclusive event classes with a deviation are observed as expected (see chapter 2). Most classes with an observed deviation are as expected in the inclusive class, while some deviations are observed for the jet-inclusive class.

6.1.3 W' sensitivity studies

For the W' sample, the most interesting sample, which corresponds to the mass of $M_{W'}$ = 5Tev will be discussed. It is the most difficult to detect and closest to the current lower mass limit [36]. Illustrations of other samples corresponding to other masses can be found in the Appendix C.

The comparison of the global JSD p and MUSiC $\widetilde{p}$ distribution is shown in Figure 6.3. The distributions for both JSD and MUSiC are similar in the median SM expectation and in signal. Both distributions show more event classes within the SM expectation in the first bin, as discussed previously, than the uniform expectation and a deficit in the following bins.

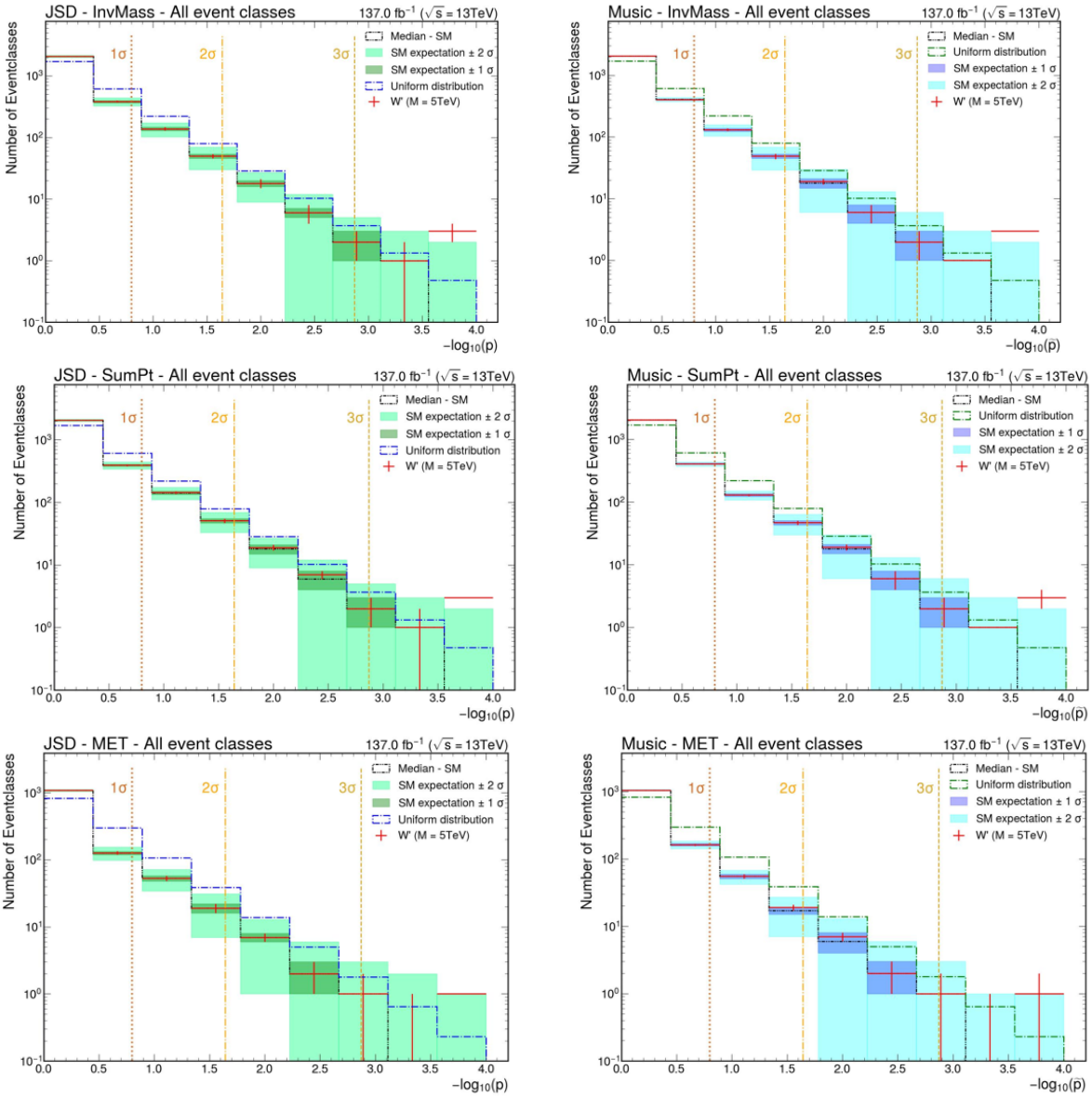


FIGURE 6.3: Comparison of the Histograms of the global $\widetilde{p}$ and p distributions for all kinematic distributions and the W' ($M_{W'}$ = 5TeV) samples. The green distributions correspond to the regularised JSD evaluation and the blue distributions to the MUSiC evaluation.

As can be seen in Table 6.1, three classes deviate from the threshold of $3\,\sigma$ between JSD and MUSiC. The first event class "$1\mu + 1MET$" shows a difference between the two algorithm in the Invariant Mass and the transverse momentum distribution, where in both variables JSD detects a more significant deviation. This is one of the main channels into which W' decays, which has been detected for lower W' masses by both algorithms. Furthermore, this event class is found in the missing transverse momentum kinematic distribution for both MUSiC and JSD. The distributions are

shown in Figure 6.4. Both distributions show a clear excess of signal over background, primarily in a region where very little SM background is detected, here the signal exceeds the background up to 200 times. For the invariant mass distribution, no region from (1540-5410) to (5410 - 6880) is vetoed. For the transverse momentum distribution, the only relevant region vetoed is (5410 - 13000). Because of the low statistics, in some cases, no integer bin count was sampled in the relevant region, leading to a nonsignificant deviation in the MUSiC analysis.

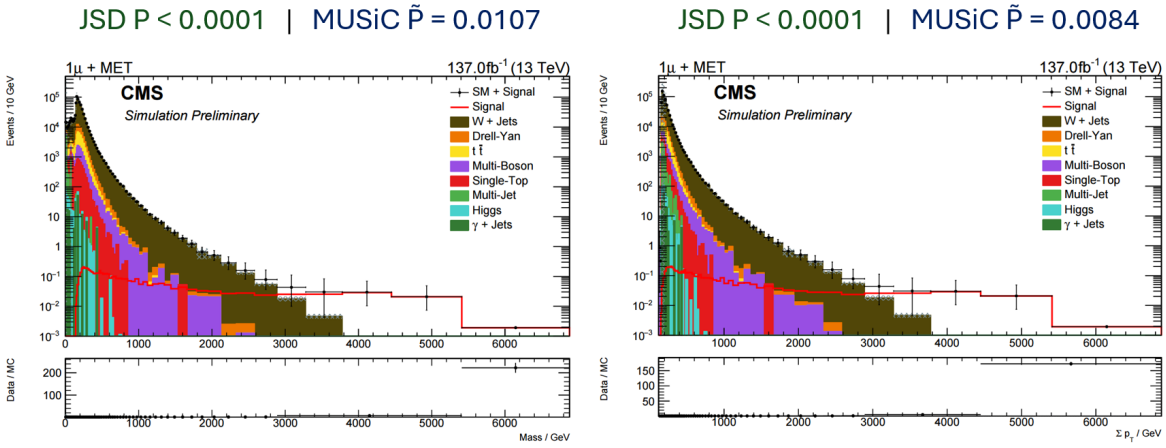


FIGURE 6.4: Invariant mass and transverse momentum distributions of the $1\mu + 1MET$ event class for $M_{W'} = 5TeV$. The calculated JSD p-value and MUSiC $\tilde{p}$ are stated above the figures.

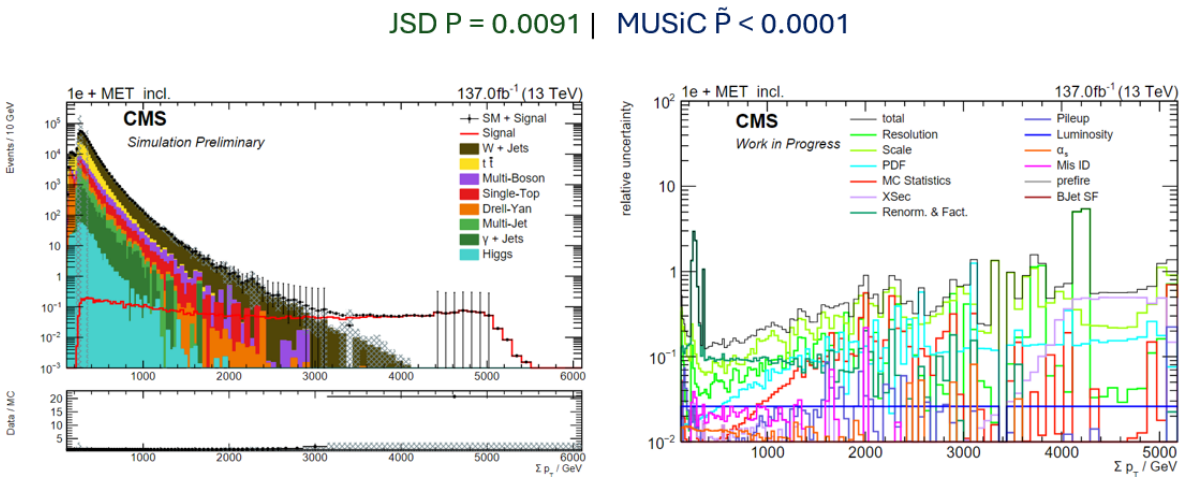


FIGURE 6.5: Transverse momentum distributions and uncertainty of the $1e + 1MET + X$ event class for $M_{W'} = 5TeV$. The calculated JSD p-value and MUSiC $\tilde{p}$ are stated above the figure.

The distribution in which MUSiC detects a more significant deviation is $1e + 1MET + X$ for the missing transverse momentum variable. The distribution and its uncertainties are shown in figure 6.5. Here, an excess of signal of up to 20 times the background is visible. Given the lower difference between SM expectation and the higher uncertainty compared to Figure 6.5, a smaller deviation is expected. Through merging, the majority signal is reduced to the singular bin, which makes the whole distribution not different enough for the JSD to recognise the deviation as definite as MUSiC.

6.1.4 Sphaleron sensitivity studies

For the sphaleron samples, the lowest pre-exponential-factor (PEF = 0.01) is examined, since it is the most difficult to detect. Illustrations of further samples corresponding to other pre-exponential-factors can be found in the Appendix C.2.

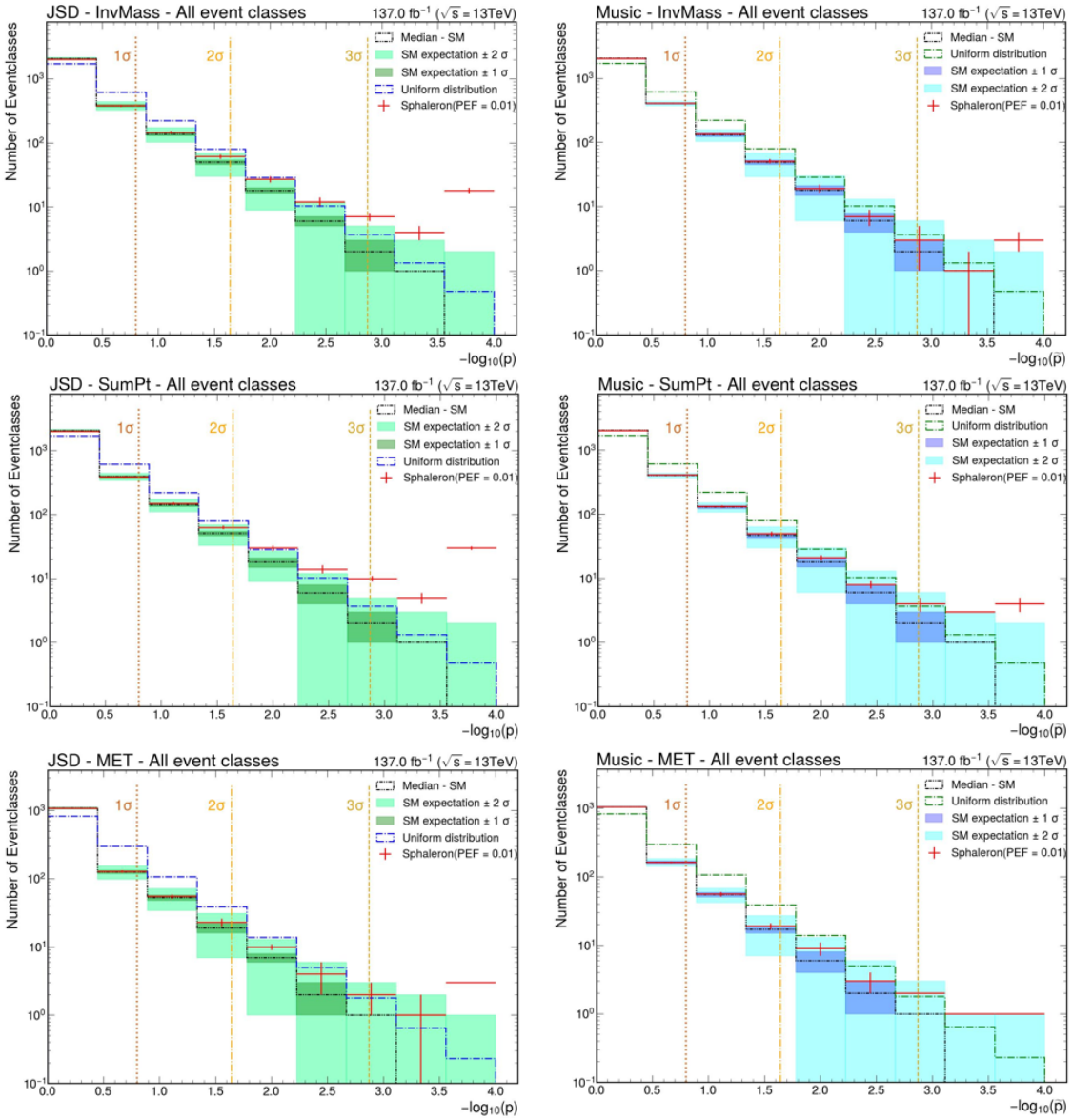


FIGURE 6.6: Comparison of the Histograms of the global $\widetilde{p}$ and p distributions for all kinematic distributions and the sphaleron (PEF = 0.01) samples. The green distributions correspond to the regularised JSD evaluation and the blue distributions to the MUSiC evaluation.

The comparison of the global JSD p and MUSiC $\widetilde{p}$ distribution is shown in Figure 6.6. The distributions for both JSD and MUSiC appear similar, both in the median SM expectation and in the signal. For the evaluated signal, while the shape is similar, a clear difference between MUSiC and JSD is visible. The number of event classes with a low p value found by the regularised JSD far exceeds the classes found by MUSiC. JSD even shows bins prior to the last bin, which exceed the combined uncertainty of the background and the signal, hinting that with a stronger PEF more classes will reach the highest quantified deviation, in the last bin.
The class where MUSiC shows a stronger deviation for a PEF of 0.01 is $1e + 6Jet +$

$1MET + X$ in the missing transverse momentum distribution. This is a class where due to the final state, a deviation is expected. The class is detected in the invariant mass distribution by MUSiC ($\widetilde{p}$ = 0.0014) and JSD (p $\leq$ 0.0001) and is also detected by JSD for every higher PEF. Furthermore, the inclusive class $1e + 6Jet + X$, which is a subset of $1e + 6Jet + 1MET + X$, is found only by JSD but not by MUSiC. The distribution is shown in Figure 6.7, where there is a deviation of up to five times the signal compared to the background. Due to the high observed uncertainty, the JSD merge procedure leads to failure. Many bins with a relative uncertainty of more than 10 are observed, and in the region of interest, the uncertainty is always above one, which leads to continuous merging.

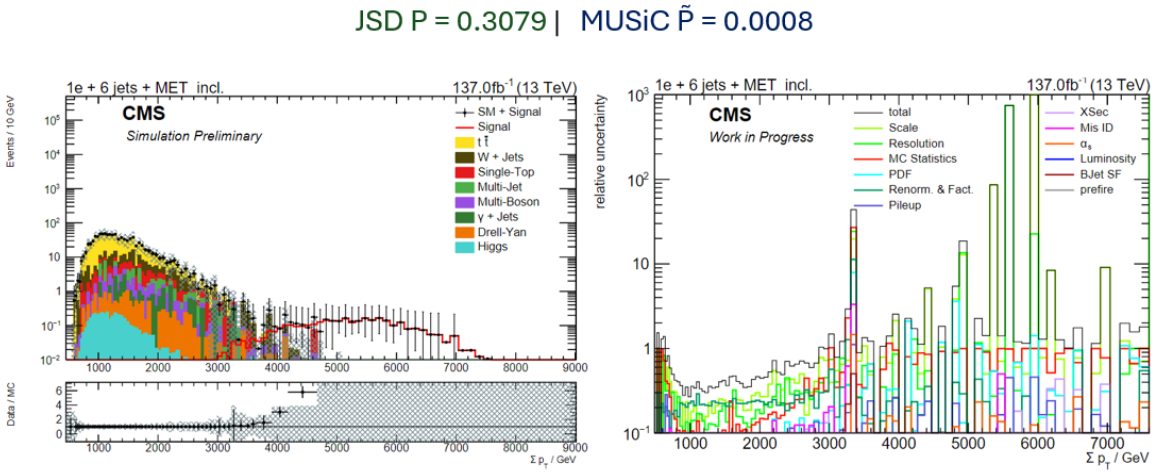


FIGURE 6.7: Invariant mass and transverse momentum distributions of the $1e + 6Jet + 1MET + X$ eventclass for PEF = 0.01. The calculated JSD p-value and MUSiC $\widetilde{p}$ are stated above the figures.

In Figure 6.8 two classes are shown in which Jensen-Shannon shows a stronger deviation than MUSiC. Both classes show a high relative excess but a low event count. Since no significant regions are vetoed, it is not clear why JSD is more sensitive in these cases.
In Figure 6.9 one extreme example is shown, where MUSiC does not show a significant deviation, even for a pre-exponential-factor of 0.05, in an event class, contrary JSD already detects it at a PEF of 0.01. Here, every region starting from 2990 GeV or higher is vetoed because of the adaptive coverage threshold, making it impossible for MUSiC to detect the deviation.
The adaptive coverage threshold was introduced to reduce the effect of high uncertainty and low event count for the calculation of the post trial p-value [73, 80] and is defined as:

$$\frac{\sigma_{SM}}{N_{True}} < max(0.5, min(5.0, 1.2 \cdot N_{SM}^{-0.2})) \tag{6.4}$$

Contrarily in the current MUSiC analysis code it is defined as:

$$\frac{\sigma_{SM}}{N_{True}} < min(1.0, max(1.2 \cdot N_{SM}^{-0.2}, 0.5)) \tag{6.5}$$

A change back to the original adaptive coverage threshold, as defined in the MUSiC analysis notes, does not change the results for the $1e + 1bJet + 5Jet + MET + X$ event class. For many more classes like $1\mu + 5Jet + X$, $1e + 1\mu + 4Jet + 1MET + X$ and $2\mu + 6Jet + 1MET + X$, all region with strong signal background separation are vetoed

by the adaptive coverage threshold. The mentioned event classes can be found in Appendix C.3.

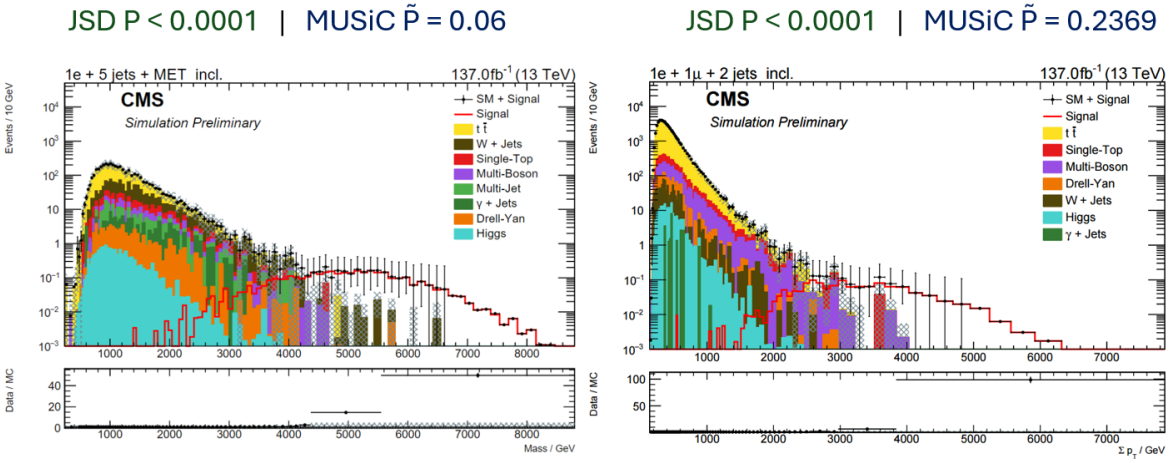


FIGURE 6.8: Invariant mass distribution of $1e + 5Jet + 1MET + X$ and transverse momentum distributions of the $1e + 1\mu + 2Jet + 1MET + X$ event class for PEF = 0.01. The calculated JSD p-value and MUSiC $\tilde{p}$ are stated above the figures.

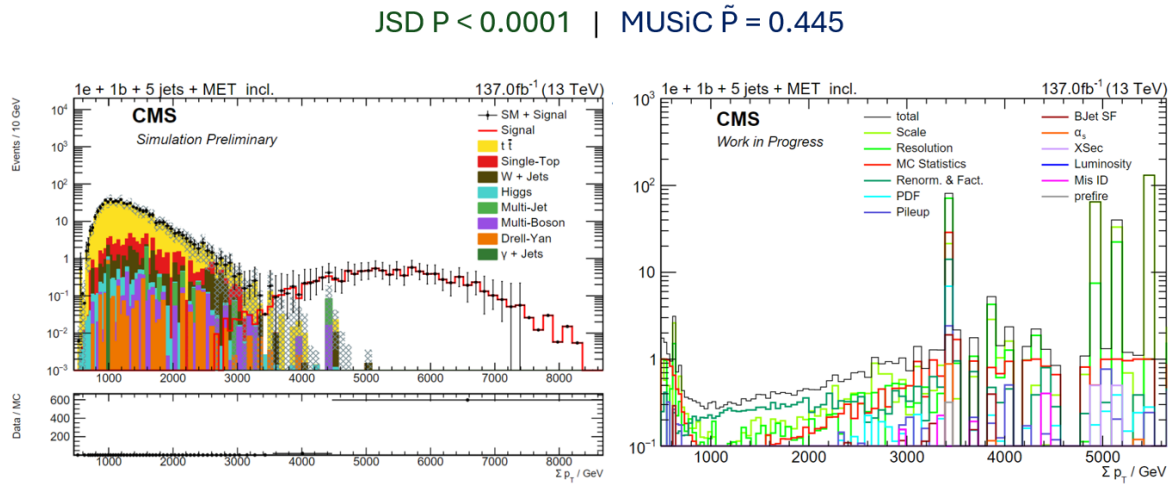


FIGURE 6.9: Transverse momentum distributions and uncertainty of the $1e + 1bJet + 5Jet + 1Met + X$ event class, for a sphaleron sample with PEF = 0.05. The calculated JSD p-value and MUSiC $\tilde{p}$ are stated above the Figure.

7 Conclusion and Outlook

In this thesis, a new method to evaluate a deviation within the MUSiC framework has been studied and compared to the MUSiC search algorithm. The new method showed promising results with simplified pseudo distributions, while possible optimisation through a multivariate approach is discussed briefly. The general advantage of the new method is the reduced computation time, compromised by the lack of the Region of Interest to be determined, still JSD shows at least similar performance as MUSiC, in the signal samples and conditions studied in this thesis. This is a minor deficit, since a detected deviation in a certain event class would trigger a dedicated analysis.
With simulated samples corresponding to the standard model expectation, low sensitivity and low correlation to MUSiC was observed. This has been corrected with the introduction of a regularisation. To compensate for problematic distributions and only evaluate bins with at most a given relative uncertainty to obtain more robustness in the sampling procedure, a merging algorithm was implemented. The final results of the W' sensitivity studies, between MUSiC and JSD, are comparable. For Sphaleron samples, the new algorithm showed good performance in finding most of the classes that MUSiC does find and many more in addition. The classes found by MUSiC and not by the regularised Jensen-Shannon algorithm, related to the merging algorithm. The reason why MUSiC did not detect many classes is due to the adaptive coverage threshold veto. This could give a hint of how to adjust this veto in future MUSiC analyses. If Jensen-Shannon could become a complementary test, the following improvements could be evaluated:

- Rework and optimise the merging algorithm. Many classes were not detected due to the merging procedure, which in areas of high uncertainty reduces the separation between signal and background. The total amount of unweighted MC events should be taken into account, as well as a study on the relative uncertainty cuts should be performed.
- An upgrade from Jensen-Shannon to H-Divergence [87] could be made. H-divergences are broader classes of divergences where Jensen Shannon is a limit for a specific set of hyperparameters. This has been studied and has a higher sensitivity in the examples evaluated by the seminal paper [87].
- Improved sensitivity has been shown for certain examples in a multivariate approach, therefore a study on CMS data, taking multiple kinematic distributions as input, could be promising.

To conclude, a promising new algorithm to quantify deviations between distributions has been studied within the MUSiC framework.There are possibilities to further enhance the sensitivity are possible, which leaves the Jensen-Shannon Distance as an interesting candidate for complementary approach to the Model Unspecific Search in CMS. This could be used to determine weaknesses in the MUSiC framework, and a study on real data could reveal interesting event classes.

Acknowledgements

First I want to express my gratitude toward Prof. Dr. Thomas Hebbeker for enabling me the great opportunity to write a bachelorthesis and to take my first steps into science in his research group and for valuable inputs in discussions leading the thesis in the right direction. I want to thank Prof. Dr. Alexander Schmidt for the discussions about physics and my thesis as well as for performing the second examination. I would like to thank Dr. Arnd Meyer for his guiding comments during presentations and the important input and counseling during discussions.
I want to express my deepest gratitude to my supervisor Dr. Felipe Torres da Silva de Araújo, who supported my ideas, taught me countless things and created an environment in which it was pleasure to do research and learn.
I want to thank all the "MUSiCians", firstly Msc. Yannik Kaiser, whose bachelorthesis was my first contact with MUSiC even before my thesis started. He provided me with the simulation samples, taught me how to use the plotting tools of MUSiC and helped with many technical questions. Next, I want to thank Msc. Chinmay Seth for helping me with an implementation to access the Linux cluster, the proof reading of my thesis and for the countless talks about physics and my results. I also want to thank Bsc. Niels Esper and Bsc. Patrik Kersten for the great time I had working with you and all the discussions we had.
I would also like to thank many members of the research group for the warm welcome, many interesting talks and the good times we had not only on our food Fridays.

I want to thank my friends for their support and the fun moments we had. I want to thank my parents and siblings for always being there for me and showing support in any way they can and especially my mother for showing special interest and curious questions. I want to express my gratitude to my girlfriend, who emotionally supported me throughout my studies even though I am far away.

Bibliography

[1] F. J. Sciulli M. K. Gaillard P. D. Grannis. The Standard Model of Particle Physics, 1998. URL: https://arxiv.org/abs/hep-ph/9812285.

[2] The ATLAS Collaboration. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc, 2012. URL: https://arxiv.org/abs/1207.7214.

[3] The CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC, 2012. URL: https://arxiv.org/abs/1207.7235.

[4] The ALPHA Collaboration. Observation of the effect of gravity on the motion of antimatter. *Nature*, vol 621, 2023. URL: https://www.nature.com/articles/s41586-023-06527-1.

[5] N. J. Esper. Exploring jet-only final states in the Model Unspecific Search in CMS (MUSiC) using 2018 data, 2023. URL: https://web.physik.rwth-aachen.de/user/hebbeker/theses/esper_bachelor.pdf.

[6] PBS NOVA, Office of Science Fermilab, United States Department of Energy, and Particle Data Group. Standard model of elementary physics. (accessed: 04.12.2023). URL: https://commons.wikimedia.org/w/index.php?curid=4286964.

[7] P. Langacker. The Standard Model and Beyond, 2017. URL: https://library.oapen.org/handle/20.500.12657/50890.

[8] Y. Kaiser. Sensitivity study of the MUSIC algorithm for LHC run 2, 2021. URL: https://web.physik.rwth-aachen.de/user/hebbeker/theses/kaiser_bachelor.pdf.

[9] R. L. Workman et al. (Particle Data Group). Review of Particle Physics , Prog. Theor. Exp. Phys. 2022, 083C01 (2022). URL: https://doi.org/10.1093/ptep/ptac097.

[10] J. Lieb. Development of a Fast Search Algorithm for the MUSiC Framework, 2015. URL: https://web.physik.rwth-aachen.de/user/hebbeker/theses/lieb_bachelor.pdf.

[11] Ahmed Ali, Jens Sören Lange, and Sheldon Stone. Exotics: heavy pentaquarks and tetraquarks. *Progress in Particle and Nuclear Physics*, 97:123–198, 2017. ISSN: 0146-6410. DOI: https://doi.org/10.1016/j.ppnp.2017.08.003. URL: https://www.sciencedirect.com/science/article/pii/S0146641017300716.

[12] P. H. H. Kersten. A study of hadronic tau lepton decays in model unspecific search in cms, 2023. URL: https://web.physik.rwth-aachen.de/user/hebbeker/theses/kersten_bachelor.pdf.

[13] F. Rehbein. Search for sphalerons in the eμ final state at s = 13 TeV with the CMS experiment, 2018. URL: https://web.physik.rwth-aachen.de/user/hebbeker/theses/rehbein_bachelor_corrected.pdf.

[14] P. D. Grannis M. K. Gaillard and F. J. Sciulli. The standard model of particle physics, 1999. DOI: https://doi.org/10.1103/RevModPhys.71.S96. URL: https://journals.aps.org/rmp/abstract/10.1103/RevModPhys.71.S96.

[15] D. Duchardt. Music: a model unspecific search for new physics based on cms data at s = 8 tev, 2017. URL: https://web.physik.rwth-aachen.de/user/hebbeker/theses/duchardt_phd.pdf.

[16] S. P. Martin and J. D. Wells. Elementary Particles and Their Interactions, 2022. ISSN: ISSN 1868-45. DOI: https://doi.org/10.1103/RevModPhys.71.S96. URL: https://link.springer.com/book/10.1007/978-3-031-14368-7.

[17] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson. Experimental test of parity conservation in beta decay. *Phys. Rev.*, 105:1413–1415, 4, February 1957. DOI: 10.1103/PhysRev.105.1413. URL: https://link.aps.org/doi/10.1103/PhysRev.105.1413.

[18] J. Lieb. Discovery Potential of a Model Independent Search for New Physics at the LHC, 2017. URL: https://web.physik.rwth-aachen.de/user/hebbeker/theses/lieb_master.pdf.

[19] Sheldon L. Glashow. Partial-symmetries of weak interactions. *Nuclear Physics*, 22(4):579–588, 1961. ISSN: 0029-5582. DOI: https://doi.org/10.1016/0029-5582(61)90469-2. URL: https://www.sciencedirect.com/science/article/pii/0029558261904692.

[20] Abdus Salam and J. C. Ward. Weak and electromagnetic interactions. *Il Nuovo Cimento (1955-1965)*, 11(4), February 1959. ISSN: 1827-6121. DOI: 10.1007/BF02726525. URL: https://doi.org/10.1007/BF02726525.

[21] Steven Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266, 21, November 1967. DOI: 10.1103/PhysRevLett.19.1264. URL: https://link.aps.org/doi/10.1103/PhysRevLett.19.1264.

[22] P. M. Agulló. Search for dark matter in proton-proton collision events with a muon and missing transverse energy in the CMS detector with s = 13TeV, 2016. URL: https://web.physik.rwth-aachen.de/user/hebbeker/theses/martinez_bachelor.pdf.

[23] A.V. Zasov, A.S. Saburova, A.V. Khoperskov, and S.A. Khoperskov. Dark matter in galaxies, 2017. DOI: 10.3367/UFNe.2016.03.037751. URL: https://arxiv.org/pdf/1710.10630.pdf.

[24] M. Dine and A. Kusenko. Origin of the matter-antimatter asymmetry. *Rev. Mod. Phys.*, 76, 1, 2003. DOI: 10.1103/RevModPhys.76.1. URL: https://link.aps.org/doi/10.1103/RevModPhys.76.1.

[25] Raymond Davis. A review of the homestake solar neutrino experiment. *Progress in Particle and Nuclear Physics*, 32:13–32, 1994. ISSN: 0146-6410. DOI: https://doi.org/10.1016/0146-6410(94)90004-3. URL: https://www.sciencedirect.com/science/article/pii/0146641094900043.

[26] S.M. Bilenky, C. Giunti, and W. Grimus. Phenomenology of neutrino oscillations. *Progress in Particle and Nuclear Physics*, 43:1–86, 1999. ISSN: 0146-6410. DOI: https://doi.org/10.1016/S0146-6410(99)00092-7. URL: https://www.sciencedirect.com/science/article/pii/S0146641099000927.

[27] W. Grimus. Introduction to left-right symmetric models. Technical report, Austria, Mar 1993, page 13. URL: http://inis.iaea.org/search/search.aspx?orig_q=RN:25012246. UWThPh–1993-10.

[28] W. Bender. Studie zum Nachweis schwerer, geladener Eichbosonen in leptonischen Zerfallskan¨alen mit dem CMS-Detektor, 2008. URL: https://web.physik.rwth-aachen.de/user/hebbeker/theses/bender_diploma.pdf.

[29] G. Altarelli and B. Mele ad M. Ruiz-Altaba. Searching for new heavy vector bosons in $p\bar{p}$ colliders. *Particles and Fields*, vol 45, 1989. DOI: https://doi.org/10.1007/BF01556677. URL: hhttps://link.springer.com/article/10.1007/BF01556677#citeas.

[30] L. O. Padeken. Search for new physics in the tau plus missing energy final states at CMS, 2017. URL: https://web.physik.rwth-aachen.de/user/hebbeker/theses/padeken_phd.pdf.

[31] M. Lindner, T. Ohlsson, and G. Seidl. See-saw Mechanisms for Dirac and Majorana Neutrino Masses. *Phys. Rev.*, vol 65, 2001. DOI: https://doi.org/10.1103/PhysRevD.65.053014. URL: https://journals.aps.org/prd/abstract/10.1103/PhysRevD.65.053014.

[32] C. Hof. Dete tionofNewHeavyChargedGaugeBosonswiththeFutureCMSDete tor, 2005. URL: https://web.physik.rwth-aachen.de/user/hebbeker/theses/hof_diploma.pdf.

[33] DØ Collaboration. Search for w boson production in the w $\rightarrow$ tb^{-} decay channel. *Physics Letters*, vol 641, 2006. DOI: https://doi.org/10.1016/j.physletb.2006.09.021. URL: https://www.sciencedirect.com/science/article/pii/S0370269306011713?via%3Dihub.

[34] The CMS collaboration. Searches for w bosons decaying to a top quark and a bottom quark in proton-proton collisions at 13 tev. *Journal of High Energy Physics*, 2017(8), 2017. ISSN: 1029-8479. DOI: 10.1007/jhep08(2017)029. URL: http://dx.doi.org/10.1007/JHEP08(2017)029.

[35] ATLAS Collaboration. Search for a heavy charged boson in events with a charged lepton and missing transverse momentum from pp collisions at s=13 tev with the atlas detector. *Phys. Rev. D*, vol 100, 2019. DOI: https://doi.org/10.1103/PhysRevD.100.052013. URL: https://journals.aps.org/prd/abstract/10.1103/PhysRevD.100.052013.

[36] CMS Collaboration. Search for new physics in the lepton plus missing transverse momentum final state in proton-proton collisions at s= 13 tev. *Journal of High Energy Physics*, vol 67, 2022. DOI: https://doi.org/10.1007/JHEP07(2022)0671. URL: https://arxiv.org/pdf/2202.06075.pdf.

[37] N. S. Manton. The inevitability of sphaleronsin field theory. *Phil.Trans. R. Soc*, vol 377, 2019. DOI: http://dx.doi.org/10.1098/rsta.2018.0327. URL: https://royalsocietypublishing.org/doi/epdf/10.1098/rsta.2018.0327.

[38] S. Daria. Search for sphalerons in proton-proton collisions, 2017. URL: https://cds.cern.ch/record/2278326/files/report_Satco.pdf?version=1.

[39] B. Burkic. Towards a search for sphalerons in leptonic channels with cms experiment, 2021. URL: https://cds.cern.ch/record/2278326/files/report_Satco.pdf?version=1.

[40] J. Ellis and K. Sakurai. Search for sphalerons in proton-proton collisions, 2016. URL: https://arxiv.org/pdf/1601.03654.pdf.

[41] C. Bravo1 and J. Hauser. Baryogen, a monte carlo generator for sphaleron-like transitions in proton-proton collisions. *J. High Energ. Phys*, vol 41, 2018. DOI: https://doi.org/10.1007/JHEP11(2018)041. URL: https://arxiv.org/pdf/1805.02786.pdf.

[42] The CMS collaboration. Search for black holes and sphalerons in high-multiplicity final states in proton-proton collisions at $$ \sqrt{s} = 13$$*tev*. *Journal of High Energy Physics*, 2018(11):42, November 2018. ISSN: 1029-8479. DOI: 10.1007/JHEP11(2018)042. URL: https://doi.org/10.1007/JHEP11(2018)042.

[43] A brief historyof cern. *CERN COURIER*, vol 19, 1979. URL: https://cds.cern.ch/record/1730465/files/vol19-issue6.pdf.

[44] L. Evans. The large hadron collider. *New Journal of Physics*, vol 9, 2007. URL: https://iopscience.iop.org/article/10.1088/1367-2630/9/9/335/pdf.

[45] CERN. Taking a closer look at lhc, accessed 20.12.2023. URL: https://www.lhc-closer.es/taking_a_closer_look_at_lhc/1.lhc.

[46] CERN. Lhc parameters, accessed 20.12.2023. URL: https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.lhc_parameters.

[47] O. Brüning, H. Burkhardt, and S. Myers. The large hadron collider. *Progress in Particle and Nuclear Physisc*, vol 67, 2007. URL: https://www.sciencedirect.com/science/article/pii/S0146641012000695?via%3Dihub.

[48] CERN. Facts and figures about the lhc, Accessed 20.12.2023. URL: https://home.cern/resources/faqs/facts-and-figures-about-lhc.

[49] CERN. Linear accelerator 4, accessed 20.12.2023. URL: https://home.cern/science/accelerators/linear-accelerator-4.

[50] S. ERDWEG. Search for charged lepton flavour violation with cms, 2018. URL: https://web.physik.rwth-aachen.de/user/hebbeker/theses/erdweg_phd.pdf.

[51] CERN. The proton synchrotron booster, accessed 20.12.2023. URL: https://home.cern/science/accelerators/proton-synchrotron-booster.

[52] CERN. The proton synchrotron, accessed 20.12.2023. URL: https://home.cern/science/accelerators/proton-synchrotron.

[53] CERN. The super proton synchrotron, accessed 20.12.2023. URL: https://home.cern/science/accelerators/super-proton-synchrotron.

[54] CERN. The large hadron collider, accessed 20.12.2023. URL: https://home.cern/science/accelerators/large-hadron-collider.

[55] CERN. Experiments, accessed 22.12.2023. URL: https://home.cern/science/experiments.

[56] The ALICE Collaboration. The alice experiment at the cern lhc. *Journal of Instrumentation*, 3, 2008. DOI: 10.1088/1748-0221/3/08/S08002. URL: https://dx.doi.org/10.1088/1748-0221/3/08/S08002.

[57] The ATLAS Collaboration. The atlas experiment at the cern large hadron collider. *Journal of Instrumentation*, 3, 2008. DOI: http://dx.doi.org/10.1088/1748-0221/3/08/S08003. URL: https://ri.conicet.gov.ar/handle/11336/64823.

[58] The CMS Collaboration. The cms experiment at the cern lhc. *Journal of Instrumentation*, 3(08):S08004, August 2008. DOI: 10.1088/1748-0221/3/08/S08004. URL: https://dx.doi.org/10.1088/1748-0221/3/08/S08004.

[59] The LHCb Collaboration. The lhcb detector at the lhc. *Journal of Instrumentation*, 3(08):S08005, August 2008. DOI: 10.1088/1748-0221/3/08/S08005. URL: https://dx.doi.org/10.1088/1748-0221/3/08/S08005.

[60] CERN. Totem, accessed 22.12.2023. URL: https://home.cern/science/experiments/totem.

[61] CERN. Lhcf, accessed 22.12.2023. URL: https://home.cern/science/experiments/lhcf.

[62] CERN. Moedal-mapp, accessed 22.12.2023. URL: https://home.cern/science/experiments/moedal-mapp.

[63] CERN. Faser, accessed 22.12.2023. URL: https://home.cern/science/experiments/faser.

[64] CERN. Snd@lhc, accessed 22.12.2023. URL: https://home.cern/science/experiments/sndlhc.

[65] CERN. Cms collaboration, accessed 22.12.2023. URL: https://cms.cern/index.php/detector.

[66] Izaak Neutelings. Cms coordinate system, accessed 23.12.2023. URL: https://tikz.net/axis3d_cms/.

[67] CMS Collaboration. Technical design report volume 1: detector performance and software, 2006. URL: https://inspirehep.net/literature/1614070.

[68] S. Wiedenbeck. Search for new physics in the e + met final state in the 2016 cms dataset, 2018. URL: https://web.physik.rwth-aachen.de/user/hebbeker/theses/wiedenbeck_master.pdf.

[69] G. Mocellin. Performance of the ge1/1 detectors for the upgrade of the cms muon forward system, 2021. URL: https://web.physik.rwth-aachen.de/user/hebbeker/theses/mocellin_phd.pdf.

[70] CMS Collaboration. Development of the cms detector for the cern lhc run 3, 2023. DOI: https://doi.org/10.48550/arXiv.2309.05466. URL: https://arxiv.org/abs/2309.05466.

[71] Daniele Trocino. The cms high level trigger. *Journal of Physics: Conference Series*, vol 513, 2014. DOI: 10.1088/1742-6596/513/1/012036. URL: https://dx.doi.org/10.1088/1742-6596/513/1/012036.

[72] Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. Technical report, CERN, Geneva, 2009. URL: https://cds.cern.ch/record/1194487.

[73] T. Pook. A model unspecific search in cms (music) using 13 tev data, 2022. URL: https://web.physik.rwth-aachen.de/user/hebbeker/theses/pook_phd.pdf.

[74] CERN. Cms detector slice, 2016. URL: https://cds.cern.ch/record/2120661.

[75] S.S. Ghosh, T. Hebbeker, A. Meyer, and T. Pook. *General Model Independent Searches for Physics Beyond the Standard Model*. 2020. ISBN: 9783030537821. URL: https://books.google.de/books?id=X16VzQEACAAJ.

[76] MUSiC, a model unspecific search for new physics, in pp collisions at sqrt(s)=13 TeV. Technical report, CERN, Geneva, 2020. URL: https://arxiv.org/abs/2010.02984.

[77] Luca Scodellaro. B tagging in atlas and cms, 2017. URL: https://arxiv.org/abs/1709.01290.

[78] MUSiC, a Model Unspecific Search for New Physics, in pp Collisions at $\sqrt{s} =$ 8 TeV. Technical report, CERN, Geneva, 2017. URL: https://cds.cern.ch/record/2256653.

[79] A. Andrade. Machine learning for a model unspecific search in cms, 2023. URL: hhttps://web.physik.rwth-aachen.de/user/hebbeker/theses/andrade_master.pdf#page=31&zoom=100,496,149.

[80] S. Ghosh, T. Hebbeker, A. Meyer, T. Pook, and J. Roemer. The region of interest (roi) scanner for the music analysis at 13 tev, 2020. URL: https://cms.cern.ch/iCMS/analysisadmin/cadilines?line=EXO-19-008&tp=an&id=2231&ancode=EXO-19-008.

[81] Adrian E. Bayer and Uroš Seljak. The look-elsewhere effect from a unified bayesian and frequentist perspective. *Journal of Cosmology and Astroparticle Physics*, 2020(10):009, October 2020. DOI: 10.1088/1475-7516/2020/10/009. URL: https://dx.doi.org/10.1088/1475-7516/2020/10/009.

[82] CMS Collaboration. Search for black holes and sphalerons in high-multiplicity final states in proton-proton collisions at $\sqrt{s} = 13$ tev. *Journal of High Energy Physics*, vol 11, 2018. ISSN: 1029-8479. DOI: 10.1007/jhep11(2018)042. URL: http://dx.doi.org/10.1007/JHEP11(2018)042.

[83] S. Eguchi and J. Copas. Interpreting kullback–leibler divergence with the neyman–pearson lemma. *Journal of Multivariate Analysis*, 97(9):2034–2040, 2006. ISSN: 0047-259X. DOI: https://doi.org/10.1016/j.jmva.2006.03.007. URL: https://www.sciencedirect.com/science/article/pii/S0047259X06000868. Special Issue dedicated to Prof. Fujikoshi.

[84] Pengfei Yang and Biao Chen. Robust kullback-leibler divergence and universal hypothesis testing for continuous distributions. *IEEE Transactions on Information Theory*, 65(4):2360–2373, 2019. DOI: 10.1109/TIT.2018.2879057.

[85] J. Lin. Divergence measures based on the shannon entropy. *IEEE Transactions on Information Theory*, 37(1):145–151, 1991. DOI: 10.1109/18.61115.

[86] Jop Briët and Peter Harremoës. Properties of classical and quantum jensen-shannon divergence. *Phys. Rev. A*, 79:052311, 5, 2009. DOI: 10.1103/PhysRevA.79.052311. URL: https://link.aps.org/doi/10.1103/PhysRevA.79.052311.

[87] Shengjia Zhao, Abhishek Sinha, Yutong He, Aidan Perreault, Jiaming Song, and Stefano Ermon. H-divergence: a decision-theoretic probability discrepancy measure, 2021. URL: https://openreview.net/forum?id=uBHs6zpY4in.

A Model Unspecific Search in CMS

A.1 Sensitivity study - W′

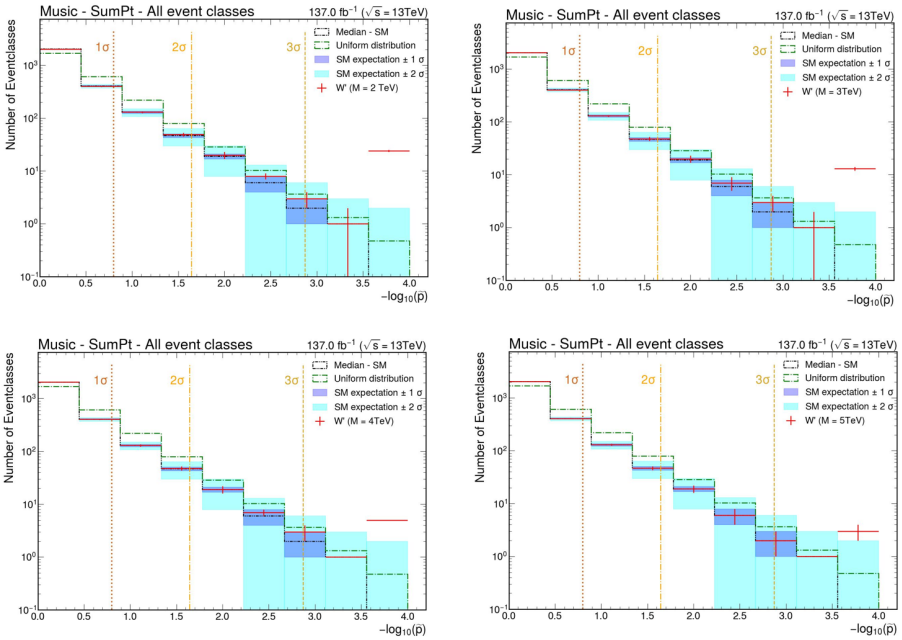


FIGURE A.1: Histograms of the $\tilde{p}$-values for the sum of transverse momentum kinematic distribution. All W′ masses under investigation are shown.

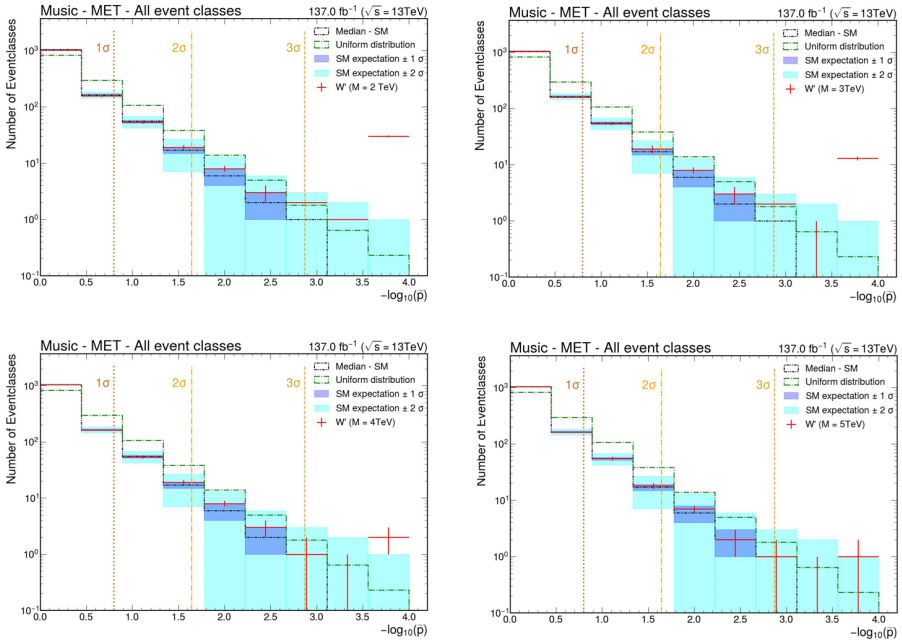


FIGURE A.2: Histograms of the $\tilde{p}$-values for the missing transverse momentum kinematic distribution. All W′ masses under investigation are shown.

A.2 Sensitivity study - Sphaleron

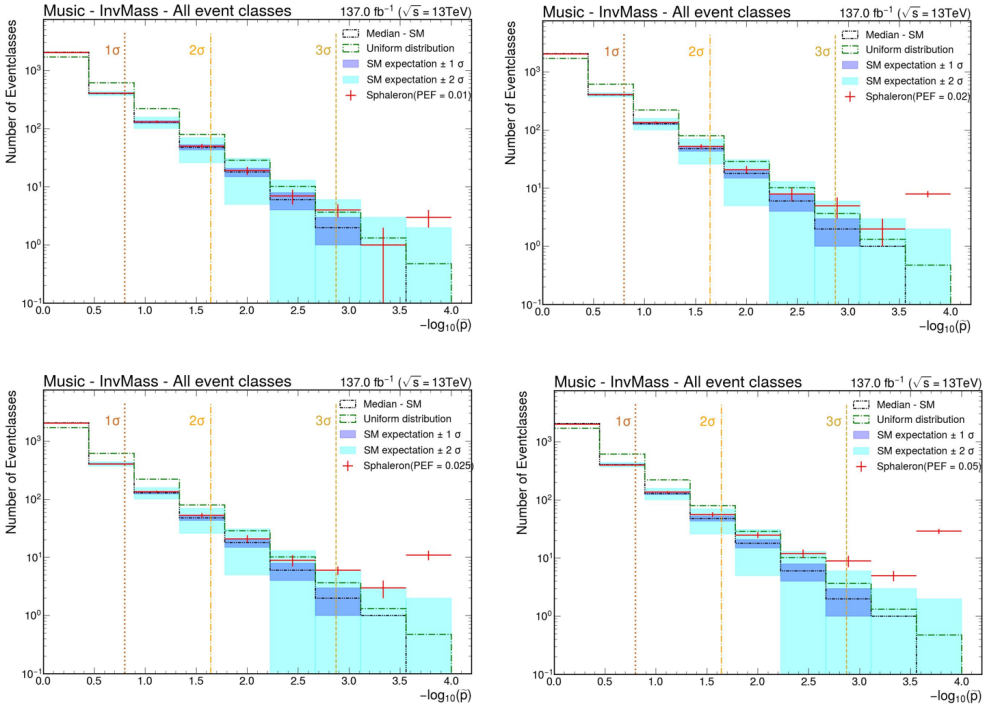


FIGURE A.3: Histograms of the $\tilde{p}$-values for the transverse momentum kinematic distribution. All sphaleron pre exponential factors under investigation are shown.

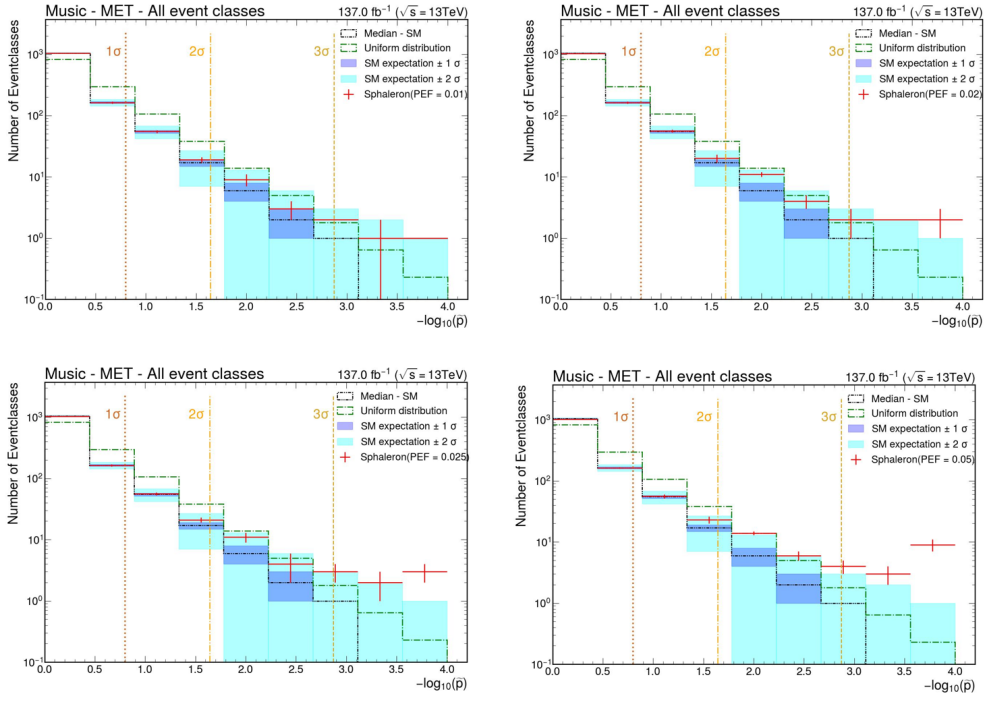


FIGURE A.4: Histograms of the $\tilde{p}$-values for the transverse momentum kinematic distribution. All sphaleron pre exponential factors under investigation are shown.

B Jensen Shannon Divergence

B.1 Evaluation with pseudo samples

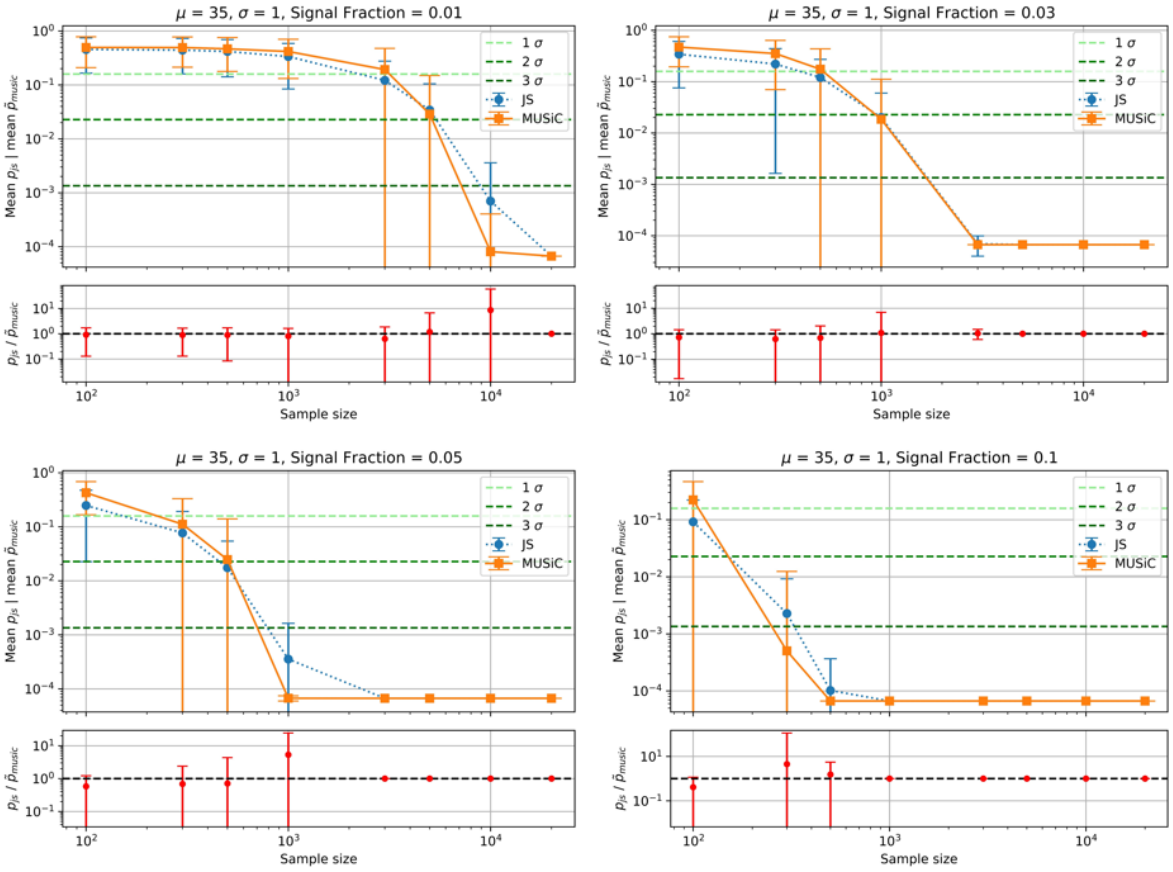


FIGURE B.1: The calculated $\tilde{p}$ for MUSiC and p for JSD is plotted against the sample size. Additionally, a plot is shown below each figure, which shows the quotient between the two metrics. Four plots are shown with the same expectation value and uncertainty for the injected normal distribution, but with different signal fractions. The different signal fractions and the parameters of the injected signal are shown in the title.

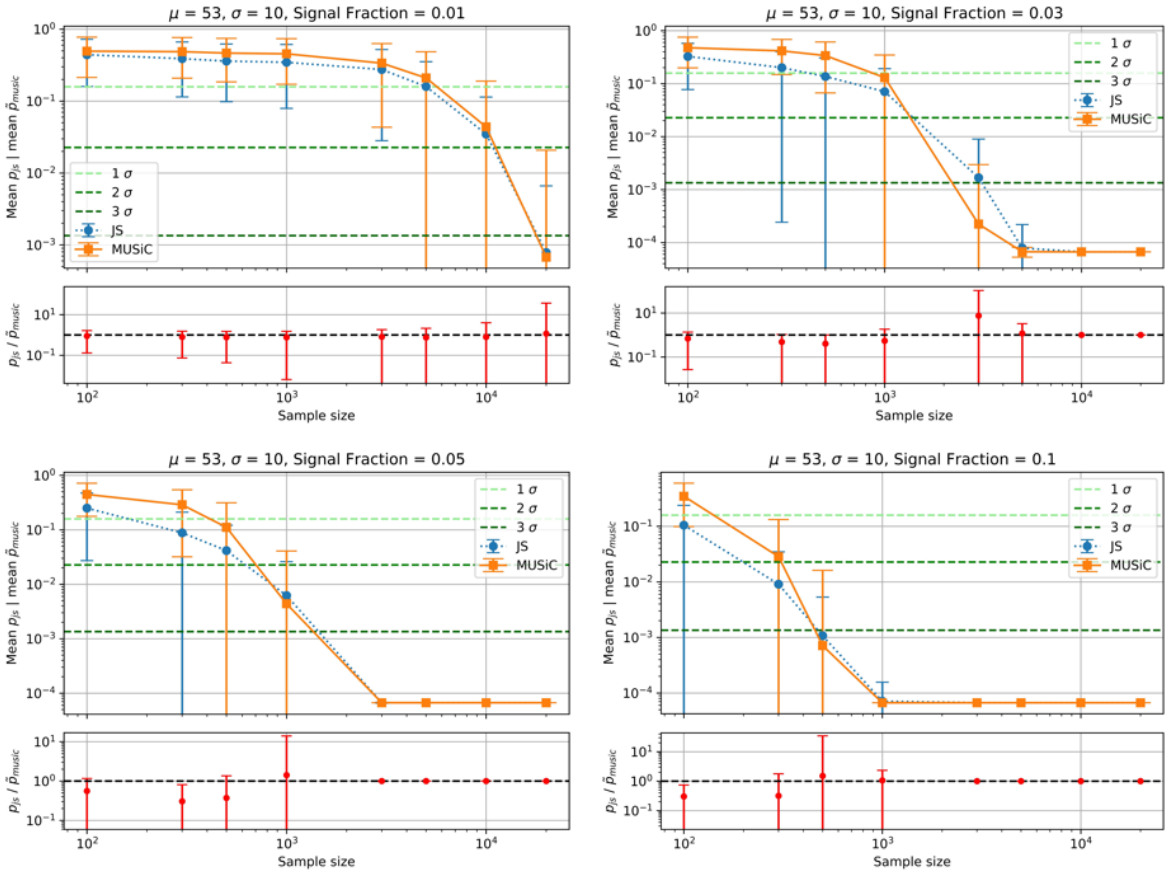


FIGURE B.2: The calculated $\tilde{p}$ for MUSiC and p for JSD is plotted against the sample size. Additionally, a plot is shown below each figure, which shows the quotient between the two metrics. Four plots are shown with the same expectation value and uncertainty for the injected normal distribution, but with different signal fractions. The different signal fractions and the parameters of the injected signal are shown in the title.

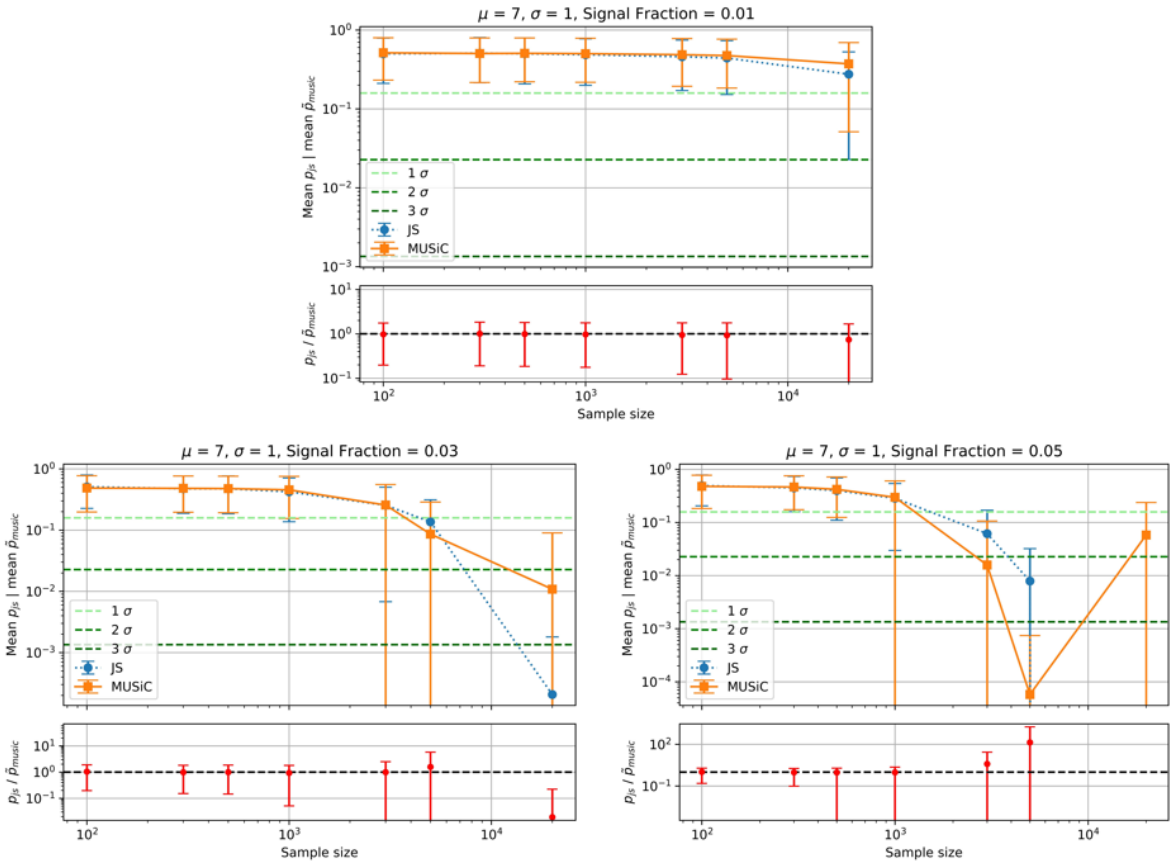


FIGURE B.3: The calculated $\tilde{p}$ for MUSiC and p for JSD is plotted against the sample size. Additionally, a plot is shown below each figure, which shows the quotient between the two metrics. Four plots are shown with the same expectation value and uncertainty for the injected normal distribution, but with different signal fractions. The different signal fractions and the parameters of the injected signal are shown in the title.

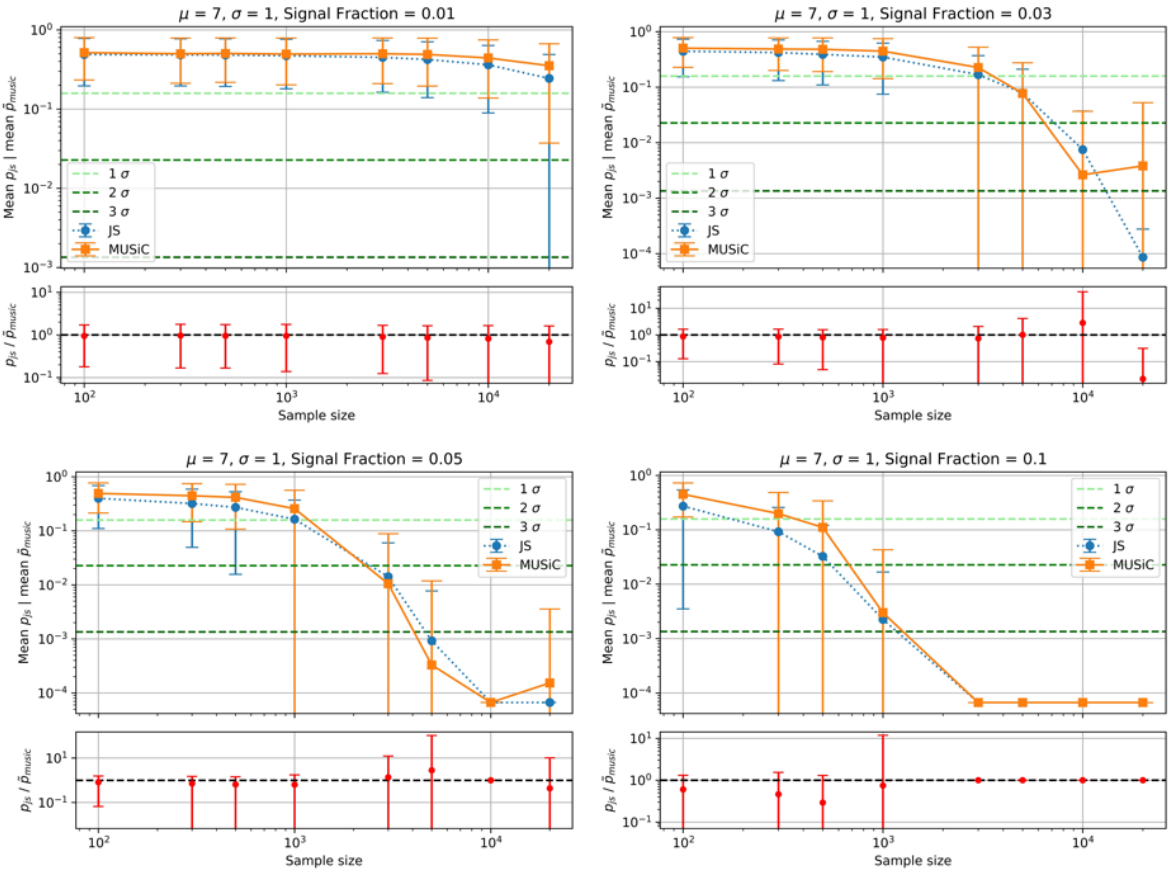


FIGURE B.4: The calculated $\tilde{p}$ for MUSiC and p for JSD is plotted against the sample size. Additionally, a plot is shown below each figure, which shows the quotient between the two metrics. Four plots are shown with the same expectation value and uncertainty for the injected negative normal distribution, but with a different signal fraction. The different signal fractions and the parameters of the injected signal are shown in the title.

C Adjusted Jensen Shannon Divergence

C.1 Sensitivity study - W′

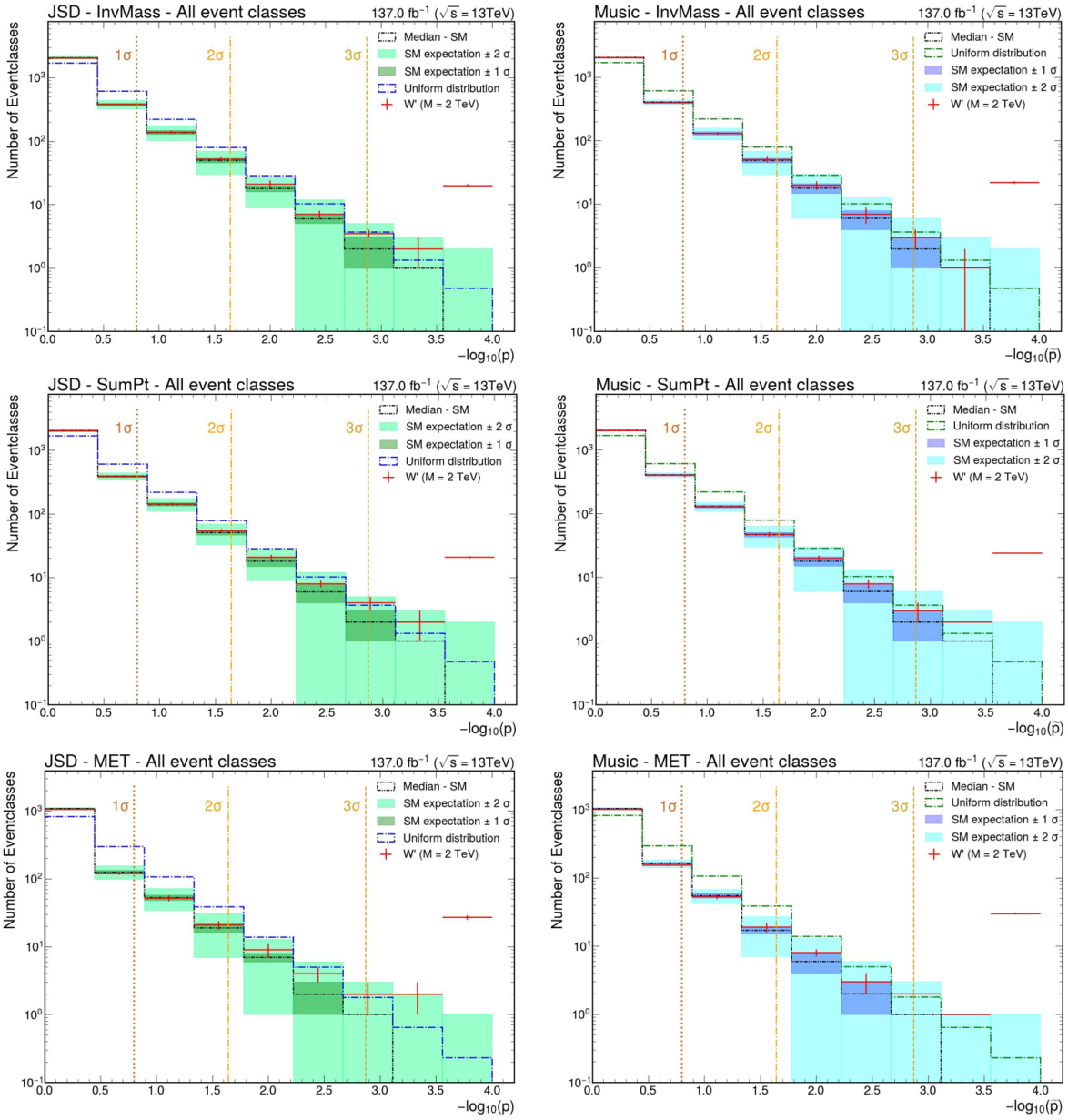


FIGURE C.1: Comparison of the Histograms of the global $\tilde{p}$ and p distributions for all kinematic distributions and the W′ ($M_{W'}$ = 2TeV) samples. The green distributions correspond to the regularised JSD evaluation and the blue distributions to the MUSiC evaluation.

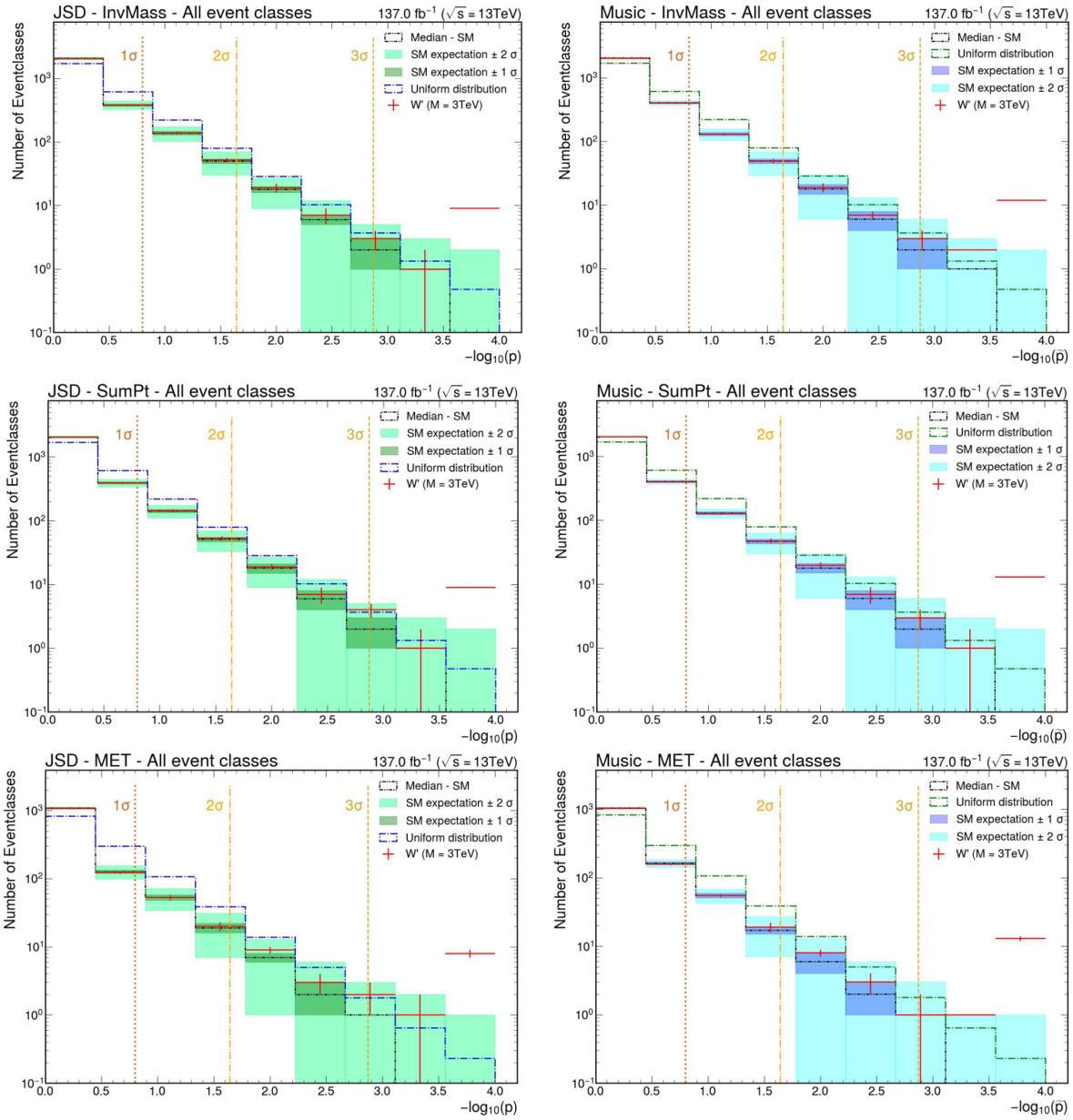


FIGURE C.2: Comparison of the Histograms of the global $\widetilde{p}$ and p distributions for all kinematic distributions and the W′ ($M_{W'}$ = 3TeV) samples. The distributions in green correspond to the regularised JSD evaluation and the distributions in blue to the MUSiC evaluation.

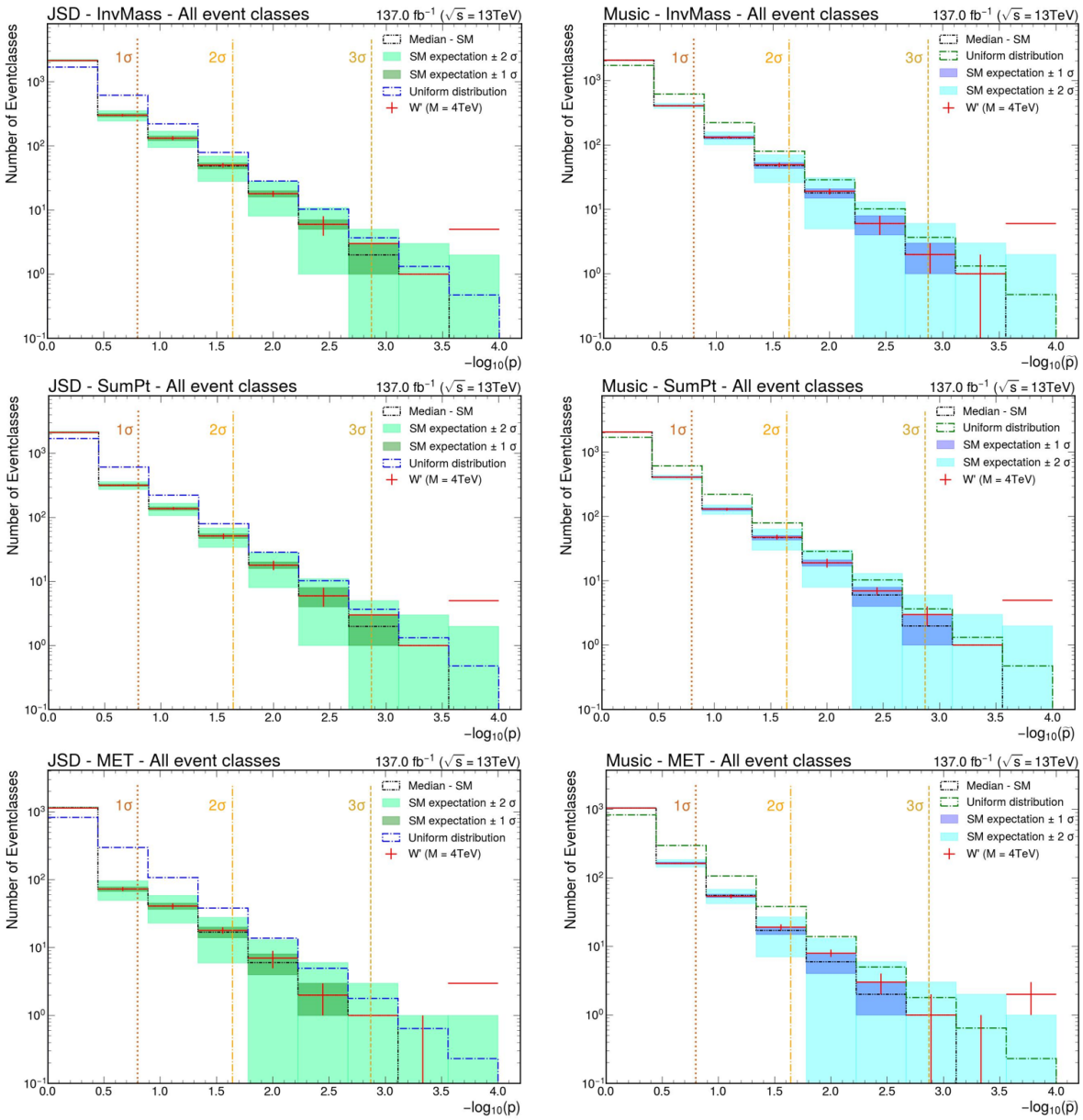


FIGURE C.3: Comparison of the Histograms of the global $\widetilde{p}$ and p distributions for all kinematic distributions and the W′ ($M_{W'}$ = 4TeV) samples. The distributions in green correspond to the regularised JSD evaluation and the distributions in blue to the MUSiC evaluation.

C.2 Sensitivity study - Sphaleron

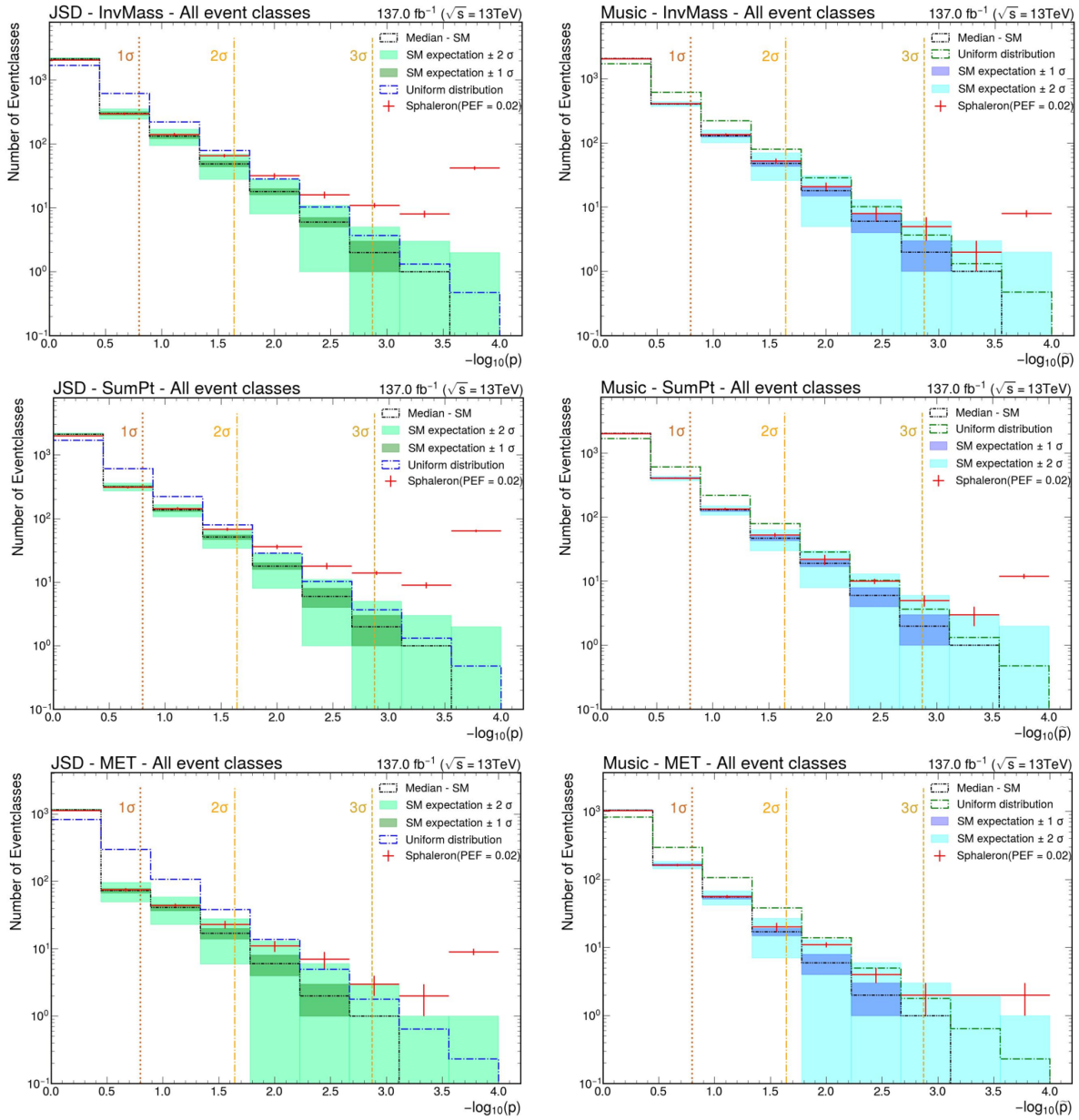


FIGURE C.4: Comparison of the Histograms of the global $\tilde{p}$ and p distributions for all kinematic distributions and the sphaleron (PEF = 0.02) samples. The distributions in green correspond to the regularised JSD evaluation and the distributions in blue to the MUSiC evaluation.

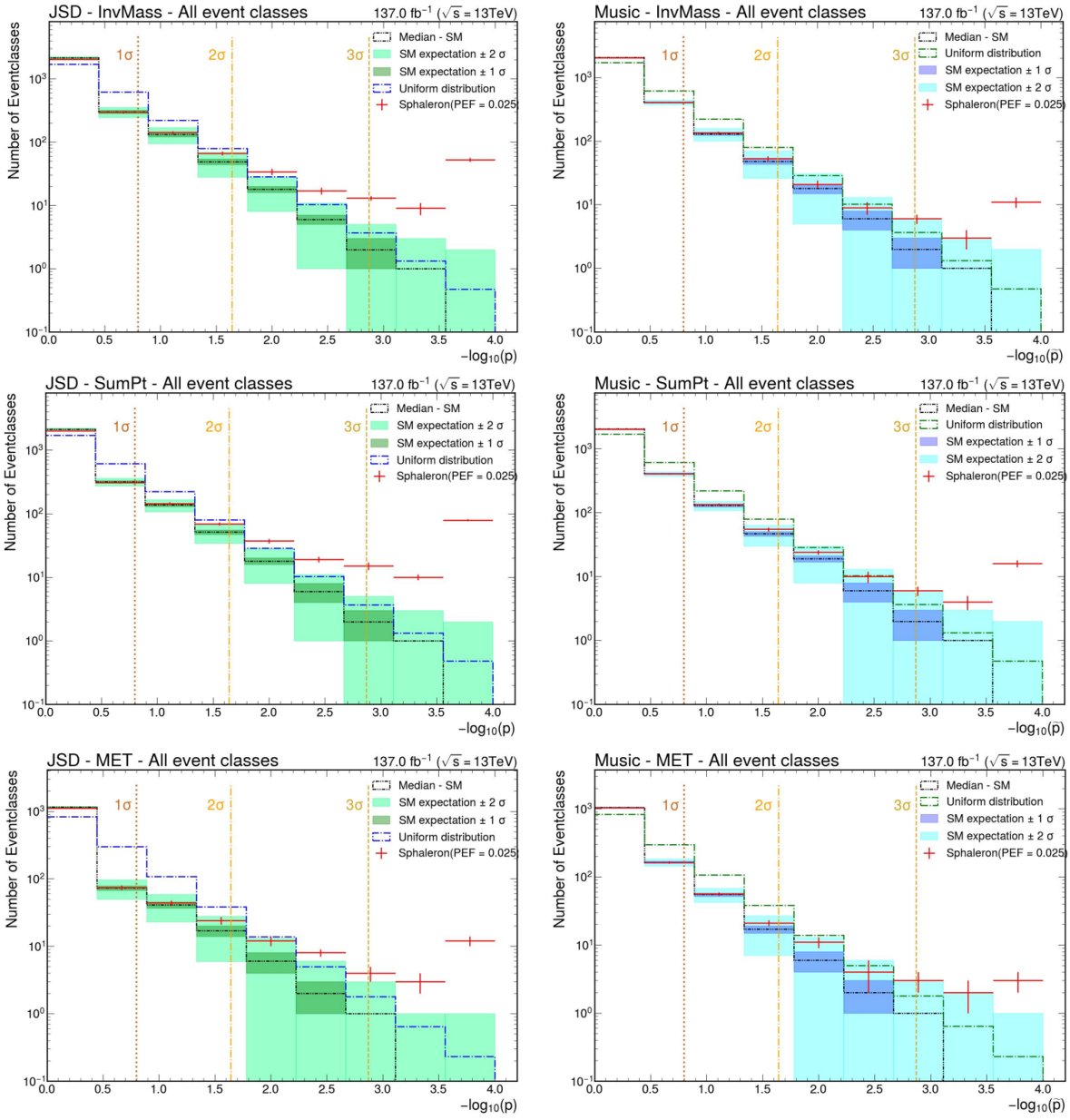


FIGURE C.5: Comparison of the Histograms of the global $\tilde{p}$ and p distributions for all kinematic distributions and the sphaleron (PEF = 0.025) samples. The distributions in green correspond to the regularised JSD evaluation and the distributions in blue to the MUSiC evaluation.

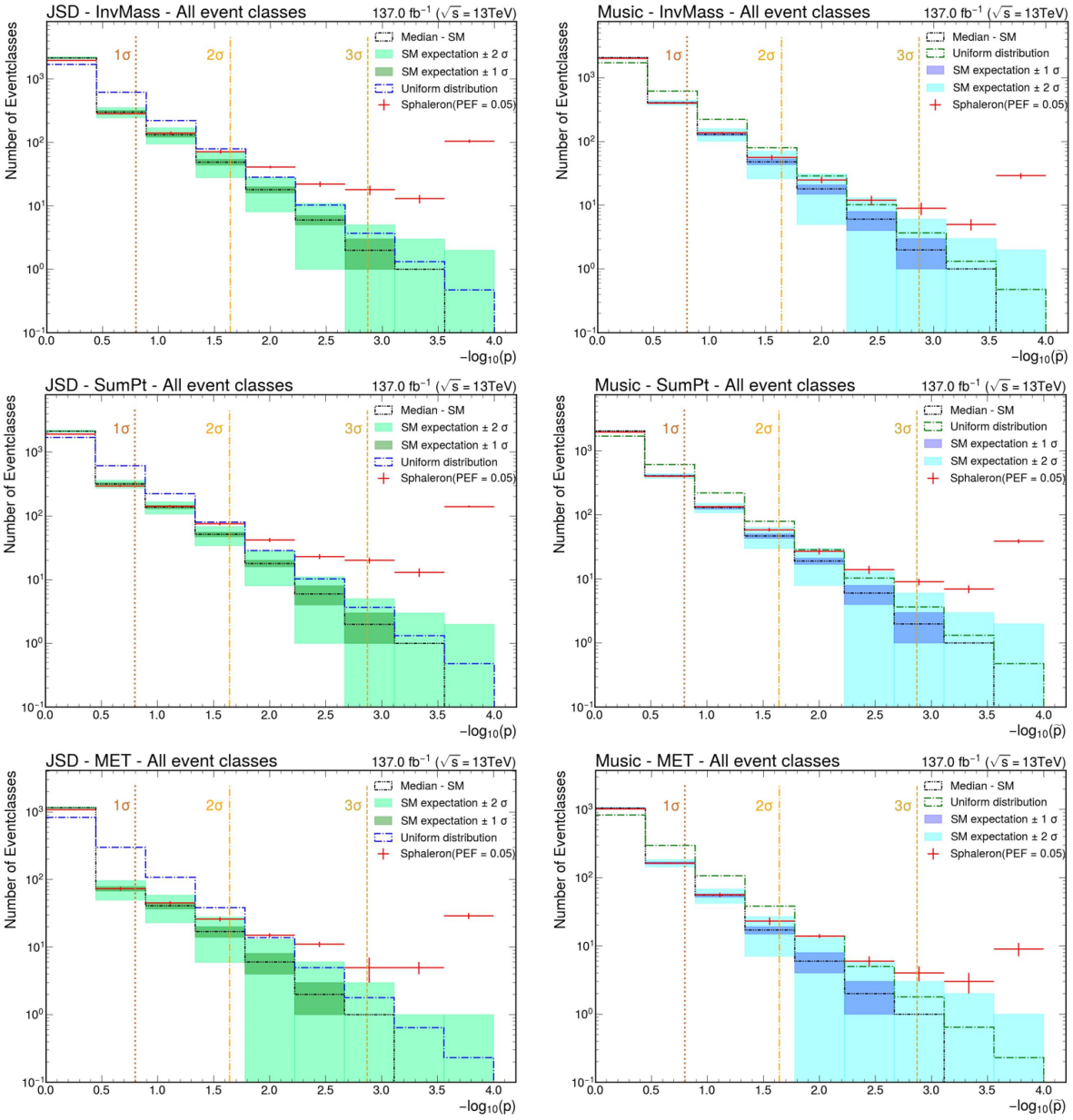


FIGURE C.6: Comparison of the Histograms of the global $\widetilde{p}$ and p distributions for all kinematic distributions and the sphaleron (PEF = 0.05) samples. The distributions in green correspond to the regularised JSD evaluation and the distributions in blue to the MUSiC evaluation.

C.3 Sphaleron event classes

In this section, more event classes are presented, which are not detected by MUSiC($\widetilde{>}0.1$) but by JSD($p < 0.001$) are presented.

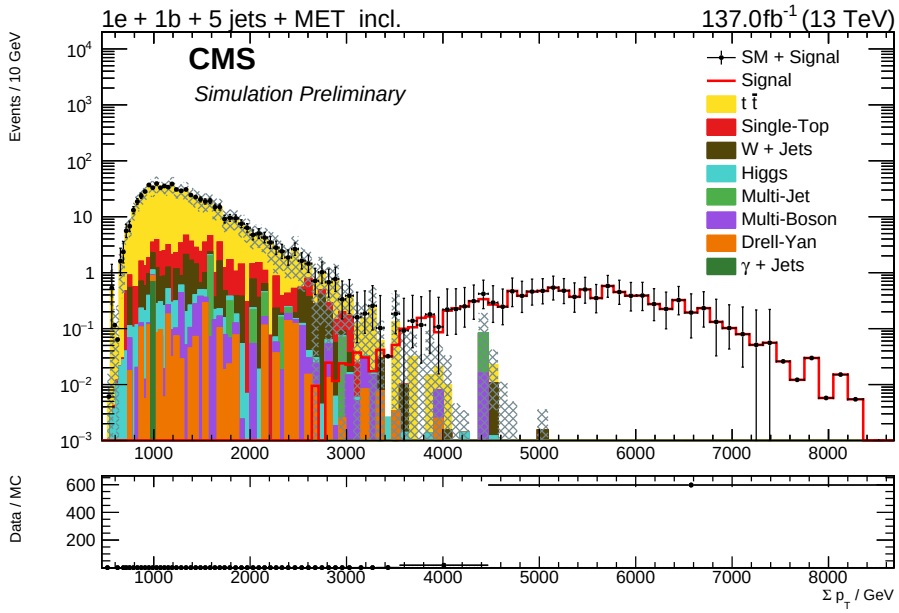


FIGURE C.7: Sum of transverse momentum distribution of $1e + 1bJet + 5Jet + MET + X$ for PEF = 0.05.

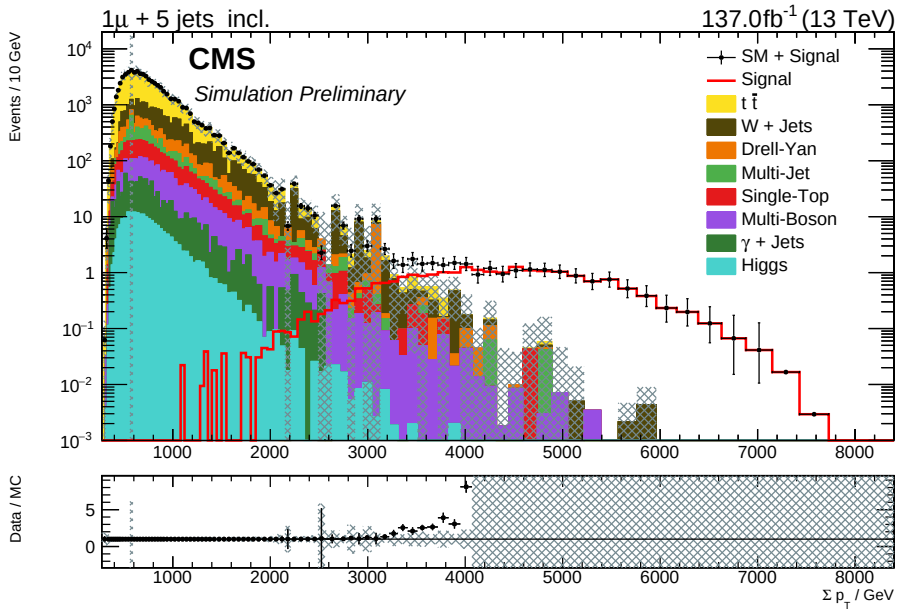


FIGURE C.8: Sum of transverse momentum distribution of $1\mu + 5Jet + X$ for PEF = 0.05.

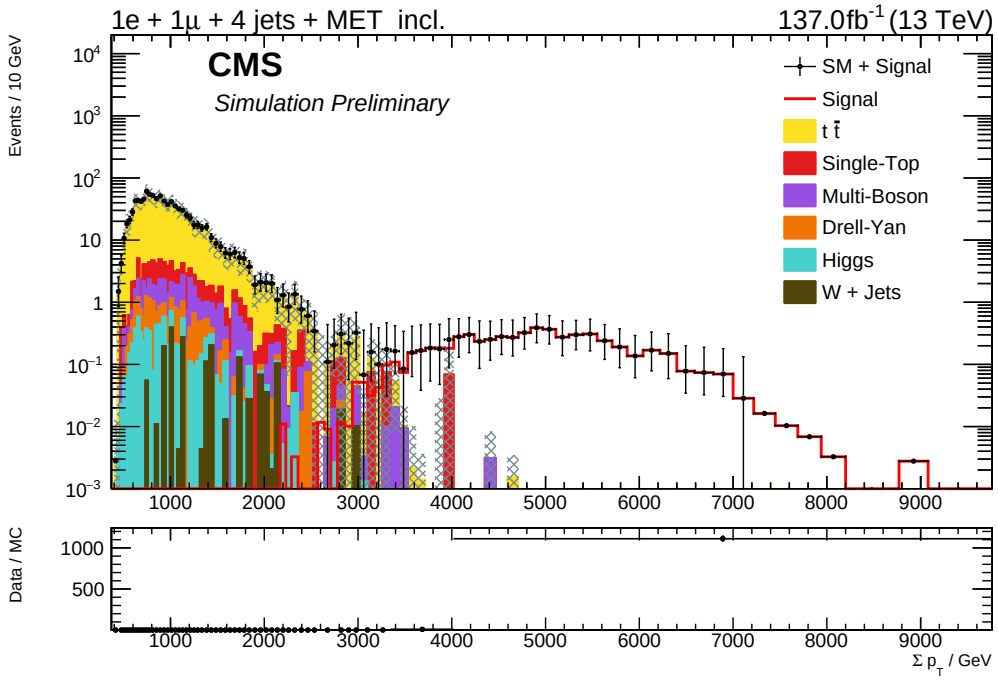


FIGURE C.9: Sum of transverse momentum distribution of $1e + 1\mu + 4Jet + MET + X$ for PEF = 0.05.

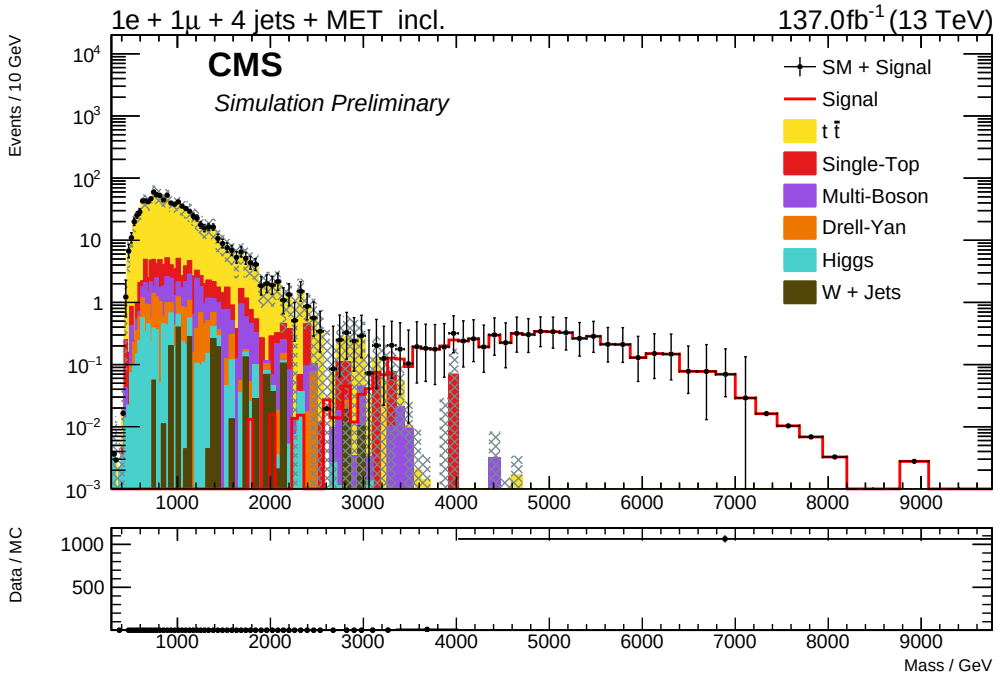


FIGURE C.10: Invariant mass distribution of $1e + 1\mu + 4Jet + MET + X$ for PEF = 0.05.

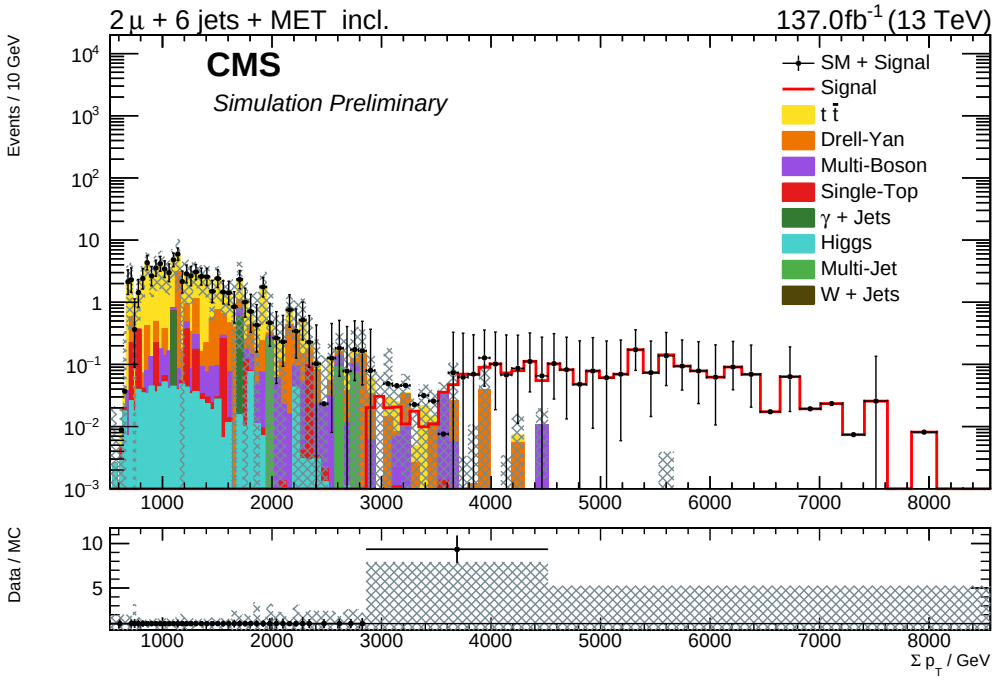


FIGURE C.11: Sum of transverse momentum distribution of $2\mu + 6Jet + MET + X$ for PEF = 0.05.

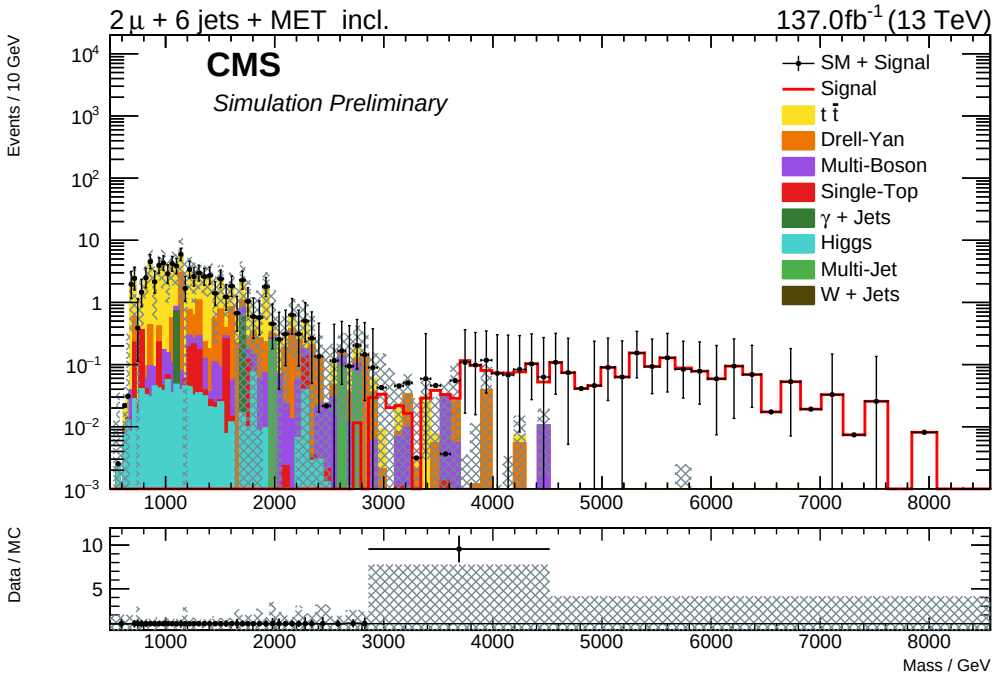


FIGURE C.12: Invariant mass distribution of $2\mu + 6Jet + MET + X$ for PEF = 0.05.

List of Figures

List of Tables

List of Abbreviations

ALICE	A Large Ion Collider Experiment
AOD	Analysis Object Data
ATLAS	A Toroidal LHC Apparatus
BOOSTER	Proton Synchrotron Booster
BSM	Beyond the Standard Model
CERN	European Organization for Nuclear Research
CMS	Compact Muon Solenoid
DAS	Data Aggregation System
DT	Drift Tubes
ECAL	Electromagnetic Calorimeter
FASER	Forward Search Experiment
HCAL	Hadronic Calorimeter
HLT	High-Level Trigger
JSd	Jensen-Shannon divergence
JSD	Jensen-Shannon Distance
KLD	Kullbacl-Leibler divergence
L1 Trigger	Level 1 Trigger
LEE	Look-Elsewhere Effect
LHC	Large Hadron Collider
LHCb	Large Hadron Collider beauty
LHCf	Large Hadron Collider forward
LINAC 2	Linear accelerator 2
LINAC 3	Linear accelerator 3
LINAC 4	Linear accelerator 4
LO	Leading Order
MC	Monte Carlo
MET	Missing Transverse Energy
MoEDAL	Monopole and Exotics Detector at the LHC
MUSiC	Model Unspecific Search in CMS
PDF	Probability density function
PF	Particle Flow
PS	Proton Synchrotron
RoI	Region of Interest
SM	Standard Model
SND@LHC	Scattering and Neutrino Detector
SPS	Super Proton Synchrotron
TOTEM	Total, elastic and diffractive cross-section measurement