Universidad de Cantabria

Facultad
de
Ciencias

Study of Dark Matter Production Models in the CMS Experiment at the LHC

Estudio de Modelos de Producción de Materia Oscura en el Experimento CMS del LHC.

Trabajo de Fin de Grado
para acceder al

GRADO EN FÍSICA

Autor: Unai Pérez Uribe

Director: Alicia Calderón Tazón

Codirector: Sergio Blanco Fernández

13-6-2025

Deseo expresar mi más sincero agradecimiento a mi tutora, Alicia, por su orientación, compromiso y apoyo durante todo el proceso de realización de este trabajo. Su acompañamiento ha sido fundamental. A mis amigos, por su compañía, comprensión y ánimo en los momentos necesarios, contribuyendo al equilibrio entre la exigencia académica y el bienestar personal. A mi pareja, por su paciencia, amor y confianza en mis capacidades. Su presencia ha sido, y será, imprescindible en mi vida. Y por último, a mi familia, por su constante respaldo, por inculcarme los valores del esfuerzo y la perseverancia, y por brindarme su apoyo incondicional en cada etapa de mi formación. Con especial mención a mi hermano, Ander, que siempre me saca una sonrisa cuando juego con él.

Resumen

En este trabajo se ha realizado un estudio de un modelo de producción de materia oscura basado en un mecanismo de Higgs oscuro. El modelo introduce tres nuevas partículas: un bosón de Higgs oscuro (s), un nuevo bosón mediador en el sector oscuro (Z') y una partícula de materia oscura fermiónica (χ).

Para estudiar el modelo en profundidad, se ha realizado generación Monte Carlo con MadGraph. Los datos generados han sido procesados y analizados con ROOT y PyROOT, lo que ha permitido el análisis de los canales principales de desintegración del Higgs oscuro, identificando el canal $s \to W^+W^- \to l^+\nu_l l^-\bar{\nu}_l$, como el más favorable.

Se ha analizado en profundidad el canal definiendo variables para diferenciar masas de las partículas introducidas y estudiando la fenomenología del modelo. Gracias a esto, se han identificado dos regímenes distintos en función de la masa del Higgs oscuro: bajo y alto. El fondo ha sido estimado mediante generaciones Monte Carlo de eventos con productos finales similares.

Por último, tras filtrar el fondo con cortes basados en estimadores de significancia, se ha estimado la eficiencia de selección de cada señal. A excepción de aquellas con $m_{Z'} < 500$ GeV en el régimen de m_s alto, debido a su baja sección eficaz. Sin embargo, para el resto de señales, se ha realizado un análisis estadístico mediante el método CL_s, obteniendo *upper limits* en las secciones eficaces con un 95% de nivel de confianza en función de $m_{Z'}$.

Como conclusión, se ha determinado que la cota superior de la sección eficaz de producción del proceso es del orden de cien veces menor que la sección eficaz de producción más pequeña observada en el CMS, lo que indica una necesidad de mayor sensibilidad para la comprobación experimental del modelo.

Abstract

In this work, a study of a dark matter (DM) production model based on a dark Higgs mechanism has been carried out. The model introduces three new particles: a dark Higgs boson (s), a new mediator boson in the dark sector (Z'), and a fermionic DM particle (χ).

To study the model in detail, Monte Carlo generations were performed using MadGraph. The generated data was processed and analyzed using ROOT and PyROOT, permitting the analysis of the main dark Higgs decay channels, identifying the channel $s \to W^+W^- \to l^+\nu_l, l^-\bar{\nu}_l$ as the most favorable.

The channel has been analyzed defining variables to discriminate between different masses of the dark sector and studying the model's phenomenology. As a result, two regimes were identified based on the mass of the dark Higgs: low and high. The background was estimated from Monte Carlo generations of Standard Model processes with similar final states.

Finally, after suppressing the background using significance based selection cuts, the signal selection efficiency was estimated for each signal, except for cases with $m_{Z'} < 500$ GeV in the high-m_s regime due to their low cross section. However, for the remaining signal points, a statistical analysis was carried out using the CL_s method, estimating *upper limits* on the cross section at a 95% confidence level as a function of $m_{Z'}$.

In conclusion, it has been determined that the upper bound on the production cross section of the process is about one hundred times smaller than the smallest production cross section ever recorded by CMS, indicating the need for greater sensitivity to experimentality test the model.

Index

1 Introduction

Currently, in particle physics, the most accurate model is the Standard Model (SM). The Standard Model is a theoretical framework that describes the fundamental particles and their interactions between them, via the fundamental forces, electromagnetic, strong, and weak. Among these elementary particles are fermions and bosons. The fermions, can be quarks and leptons, which constitute matter. Also, the model includes the bosons (W, Z, γ, and g), which mediate the fundamental interactions. Finally, it includes the Higgs boson, responsible for giving mass to particles through the electroweak symmetry breaking, also called as Higgs's mechanism.

Despite being a highly precise and experimentally validated model, it does not incorporate the gravitational interaction and it is not able to explain other phenomena, making it an incomplete theory. One of them is the existence of DM.

There are many indications of the existence of a completely unknown type of matter. There is evidence pointing to the need for more matter, beyond SM particles, for the rotational dynamics of galaxies to be consistent with our current understanding of gravity. Furthermore, observations also point to the need for additional matter to account for the curvature of light observed when passing through galaxy clusters, so that it matches theoretical predictions. On the other hand, thanks to cosmological observations, such as those from the Planck satellite on the cosmic microwave background (CMB) [1], it has been determined that the matter density of the universe is approximately the 32% , while the baryonic matter density is 5% , indicating a large amount of unknown matter beyond the Standard Model (SM).

Thanks to the cosmologycal observations, some characteristics of this new type of unknown matter are already known. It must not interact electromagnetically, otherwise it would be visible. This lack of electromagnetic interaction is actually the reason it's called "dark matter," and it is one of the main reasons why it is so difficult to detect, as most current instruments in astrophysics rely on electromagnetic interactions. On the other hand, because it is know, according to observations, that DM particles interact gravitationally, DM particles must be massive.

Also, it is known to be stable, to interact very little or not at all with itself, and it is possible that

DM particles interact via the weak force. So, in this work, DM is considered to be made of particles, but there are other models that don't assume this. It also will be assumed that DM interacts with SM particles via the weak force.

Currently, there are three types of experiments to detect DM, which are discussed in Section 1.1. With all this in mind, the aim of this work is to study DM through computational simulations of DM production in high energy particle colliders.

1.1 Dark Matter Detection

The search for DM can be divided into several detection strategies: direct detection, indirect detection, and production at particle colliders. The DM detection methods can be visualice in Figure 1.1.

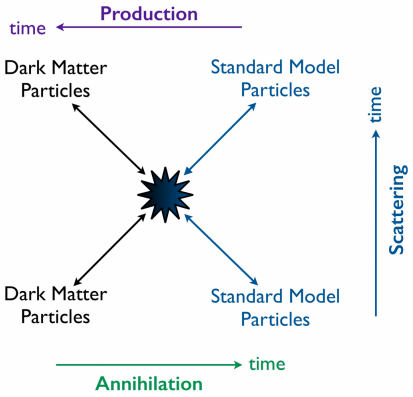


Figure 1.1: DM detection diagram [2].

Direct detection aims to observe an interaction between a DM particle and an atomic nucleus within a detector. These detectors are located in underground laboratories to shield them from other types of radiation, and it is expected that DM particles coming from outer space may interact with them. These experiments look for the nuclear recoil that would occur if such a collision took place via the weak interaction. This recoil could produce various detectable effects: ionization of atoms, scintillation light, or even vibrations in a crystal lattice, among others.The main problem is that these interactions are extremely rare, if they occur at all, and the energy transferred is expected to be very small. As a result, detectors must be large, highly sensitive, and well shielded to avoid mistaking known background interactions for potential DM signals. One example of this type of experiment is XENONnT [3], located in Italy, in which the detector is made of liquid Xenon. Other materials are also used such as: Germanium and silicon that are used in SuperCDMS [4], located in Canada, or NaI (thallium-doped sodium iodide) in the italian detector called DAMA/LIBRA [5].

Indirect detection focuses on identifying the products of dark sector processes, such as annihilation or decay of DM particles, that could result in conventional particles like gamma rays or neutrinos. These detection uses space telescopes, such as the Fermi-LAT [6], ground-based Cherenkov detectors, neutrino observatories, or antimatter detectors. In all of the cases, the goal is to find an excess of signals from regions of the universe where the cosmological model predicts high DM density, like the galactic center or nearby dwarf galaxies.

Finally, the high energy colliders try to produce DM particles in high energy collisions, a method that will be the focus of this work. If DMinteracts with SM particles, it might be produced in particle accelerators like the Large Hadron Collider (LHC) at CERN, and detected in experiments like the Compact Muon Solenoid (CMS). Even if it does, the interaction is expected to be extremely weak, resulting in a very small cross section that makes observation difficult, this will be discussed in more detail later.

The CMS experiment is one of the main detectors at the LHC, designed to study high-energy proton collisions. With its large cylindrical structure and powerful superconducting magnet, CMS can precisely detect a wide range of particles.

Also, the High Luminosity LHC (HL-LHC) will be an upgrade of the Large Hadron Collider, designed to significantly increase the accelerator's integrated luminosity $\mathcal{L}$, allowing the collection of 5 to 7 times more data than in its original configuration. Luminosity is a measure of the rate of potential collisions in a particle accelerator, typically expressed as the number of particles per unit area per unit time in the colliding beams. This boost in luminosity will improve precision measurements within the Standard Model, and increase the potential for discovering new physics. The HL-LHC is expected to begin operation around 2030 and continue running for at least a decade [7].

1.2 Dark Higgs Model

The model that will be studied throughout this work is a simplified model for the production of DM particle pairs at the LHC [8]. This model considers Weakly Interacting Particles (WIMPs), which are one of the most popular DM candidates, that interact only via the weak force and gravity.

The relic density is the currently predicted abundance of DM that "freeze out" of thermal equilibrium in the early universe as it expanded and cooled. One motivation for considering WIMPs as DM is the so-called "WIMP miracle", the remarkable coincidence that these theoretical particles replicate the relic density, which is ($\Omega_{\mathrm{DM}}h^2 \approx 0.12$)[8], if their masses are in the GeV-TeV range, so they are good candidates to be produced at the LHC energies.

There are many models that produce WIMP DM particles that try to explain the DM nature. In this work it is studied one DM model [8], in which a new complex Dark Higgs field S, that carries a new charge called q_S, is introduced. This field S is charged under a new $U(1)'$ gauge group symmetry. This model introduces three new particles beyond SM, the first one being the dark higgs, one dark sector mediator called Z' and one DM particle (χ).

The new DM particle (χ) is considered a Majorana fermion, which means it is its own antiparticle. The DM particle and Z' both obtain their mass thanks to the interaction with the dark Higgs field, in a process very similar to the SM Higgs mechanism. In short, as the early universe cooled, the $U(1)'$ gauge symmetry was spontaneously broken. This caused the vacuum expectation value (VEV) of S, the average value of the field in its minimum energy state, to acquire a constant, non-zero value (w). This is what causes the DM and Z' particles to acquire mass, in the same way it happends in the SM with the W and Z^0 bosons. Z' could be thought of as a new Z^0, but in the dark sector instead of SM, that works as a mediator between both sectors.

Also, as in SM, the symmetry breaking gives rise to a new physical Dark Higgs boson s, which is defined as $S = 1/\sqrt{2}(s + w)$. Also, in this particular model, the DM particle has a axial coupling to Z', which is not the only option as other models with vector couplings could been studied. In the Dark Higgs framework, the possible Feynman diagrams are shown in Figure 1.2.

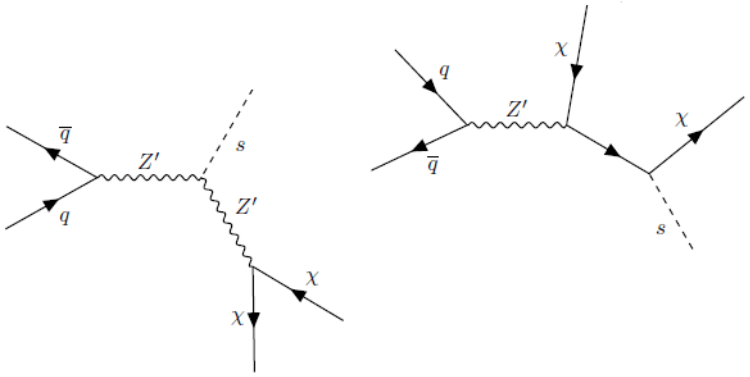


Figure 1.2: Feynamn's diagrams in the dark Higgs model.

The first diagram is a "Dark Higgs strahlung" in which Z' irradiates one dark Higgs, and later decays into $\chi\chi$. In the second one, Z' decays directly into $\chi\chi$, and one of the DM particles irradiates an s.

The Lagrangian of this model is given by the ec. (1.1).

$$\mathcal{L}_\chi = -\frac{1}{2} g' q_\chi Z'^\mu \bar{\chi}\gamma^5\gamma_\mu\chi - \frac{y_\chi}{2\sqrt{2}} s\,\bar{\chi}\chi + \frac{1}{2} g'^2 q_S^2 Z'^\mu Z'_\mu \left(s^2 + 2sw\right) - g_q Z'^\mu \bar{q}\gamma_\mu q \tag{1.1}$$

where g' is the $U(1)'$ gauge coupling and γ^μ are the Dirac gamma matrices, a set of four 4×4 matrices that are crucial in constructing Lorentz-covariant expressions involving Dirac spinors. Additionally, γ^5 is defined as $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$, and together with γ^μ it forms an axial-vector current in expressions like $\bar{\psi}\gamma^\mu\gamma^5\psi$. This appears in the first term of the Lagrangian, $\bar{\chi}\gamma^\mu\gamma^5\chi$, in which Z' couples to DM.

The second term in the Lagrangian is a mass term, in which χ acquires mass through its coupling with S. Being the DM mass $m_\chi \propto y_\chi w$, in which y_χ is the Yukawa coupling between s and χ.

In the third term, by expanding $(s+w)^2$, three terms are obteined, each multiplied by $\frac{1}{2}g'^2 q_S^2 Z'^\mu Z'\mu$. The term corresponding to the w^2 component is the one through which the Z' acquires mass by coupling to S, so $m_{Z'}$ depends on g', q_S, and w. The other two terms describe the couplings of one s with two Z' and two s with two Z', respectively.

The last term describes the interaction between the Z' and the SM quarks, serving as a bridge between the dark sector and the SM. The coupling is via a vector current of quarks, unlike the axial-vector current in DM, where g_q is the coupling between Z' and the quarks. As it is seen, in this model, Z' interacts with SM only via quarks, and Z' does not interact with leptons. There are some models when this occurs, but will not be studied.

Also, gauge invariance requires that $q_\chi = q_s/2$, where q_χ is the DM charge under the $U(1)'$ symmetry, and $g_\chi \equiv g' q_\chi$. With all this, the Lagrangian depends on six parameters: the masses of the three new particles ($m_{Z'}$, m_s, and m_χ) and the couplings: g_χ (the dark-sector coupling), g_q, and θ, the mixing angle between the SM Higgs boson h and its dark-sector counterpart s. Consequently, the decay width of s into SM particles is reduced by a factor of $\sin^2\theta$. With all this, the Lagrangian can be rewritten in terms of these parameters in Eq. 1.2.

$$\mathcal{L}_\chi = -\frac{1}{2} g_\chi Z'^\mu \bar{\chi}\gamma^5\gamma_\mu\chi - g_\chi \frac{m_\chi}{m_{Z'}} s\,\bar{\chi}\chi + 2 g_\chi Z'^\mu Z'_\mu \left(g_\chi s^2 + m_{Z'} s\right) - g_q Z'^\mu \bar{q}\gamma_\mu q \tag{1.2}$$

It is important to note that, among the six independent parameters, in the computational model, the couplings will be fixed and the masses will be taken as free parameters.

2 Computational Analysis

2.1 Monte Carlo Generation

In order to study the Dark Higgs model in detail, Monte Carlo generations have been used [9]. Monte Carlo methods are a class of algorithms based on random sampling, trying to obtain numerical results by repeating many generations. These algorithms model complex and probabilistic phenomena; in the context of particle physics, they are used to generate collision, production processes, and decays.

For this purpose, the program *MadGraph5_aMC@NLO*, also referred to as MadGraph, has been used in this work, to generate LHC collisions [10]. MadGraph is a software designed for the automated calculation of matrix elements in particle collision processes and Monte Carlo event generation. It was used in a Linux environment, offering the ability to select which particles participate in the collision, modify free parameters of the models, adjust beam energies, and more. By default, MadGraph operates at leading order (LO), corresponding to the simplest Feynman diagrams, which means the first term in the perturbative expansion describing the event will be the only one taken into account.

To test theories beyond Standard Model (BSM), as in this study, or to run simulations in SM, FeynRules is required. FeynRules is a Mathematica software package that allows one to define theoretical models from Lagrangians. It then derives all the associated Feynman rules for use in Monte Carlo generators [11].

Once FeynRules has processed the Lagrangian, the model is exported in the UFO (Universal FeynRules Output) format. The UFO contains all model information, such as couplings, particles and properties. MadGraph is designed to use these UFO files, enabling event generation in the new model. For this study, the UFO provided by the article *Hunting the Dark Higgs* [8], named *ZpHiggs_UFO*, was used to simulate events according to the model described in the previous section.

Monte Carlo event generation produces an output in the format LHE (*Les Houches Event Files*), which store all the generated information, such as the particles produced in each step and the 4-dimensional vector (E, pt) for each of them as well as the cross section the process . Generating

events and handling LHE files are computationally intensive tasks for a personal computer. Therefore, the CERN interactive computing service, *lxplus*, was used in this work [12]. *lxplus* is a remote access service to a Linux server cluster, offering many advantages. It provides access to computational resources far exceeding those of personal workstations and is integrated with CERN's large scale storage systems, enabling efficient management of heavy workloads. For all these reasons, *lxplus* was crucial in this work.

2.2 How to create events?

The model and the decay processes of the dark Higgs have been studied through MadGraph generations. To do so, several event samples were generated varying the free parameters of the model, with each sample containing 10.000 events. The couplings are fixed based on values suggested in the article: "*Dark Matter Benchmark Models for Early LHC Run-2 Searches: Report of the ATLAS/CMS Dark Matter Forum*" [13], being: $\theta = 0.01$, $g_q = 0.25$, $g_\chi = 1$ and the widths of Z' and s set to AUTO. So the free parameters will be m_χ, m_s, and $m_{Z'}$. Scripts were created in MadGraph, which are text files that contain a list of commands that MadGraph can run directly, simplifying the process. In order for MadGraph to read the script and start the simulation, the following command is used:

./bin/mg5_aMC "script_name"

As it is shown in the Listing 2.1, is to import the UFO model used in the simulation. Then, the particles involved in the interaction and the name of the output file must be specified. Additionally, parameters such as couplings, masses, and beam energies can be modified. An example of a script is:

Listing 2.1: Script example in MadGraph5_aMC@NLO

```
import model ZpHiggs_UFO
generate p p > zp > dm dm hs
output DM
launch
set wZp AUTO
set wHs AUTO
set MZp 1000
set MHs 300
set ebeam1 6800
set ebeam2 6800
set MDM scan:[200, 250, 300, 400, 500]
```

where in "`generate p p > zp > dm dm hs`", an event is generated where two protons collide, since the LHC produces proton–proton collisions. This leads to the production of a Z' boson, which decays into a final state with two DM particles and a dark Higgs. With this command, MadGraph generates all possible processes for all diagrams at LO.

On the other hand, "`wZp`" and "`wHs`" are the widths of the Z' boson and the dark Higgs, respectively. Additionally, in this case, the dark Higgs mass is set to 300 GeV with the command set MHs 300, and

a scan over DM masses is performed with the command "`set MDM scan:[200, 250, 300, 400, 500]`".

Also, it is important to emphasize that the LHC began its third data-taking period (Run 3) in 2022, and proton-proton collisions with a center of mass energy of 13.6 TeV [14]. Therefore, throughout this work, generations are performed at this energy. Because of this, the energy of each beam is set at 6800 GeV to recreate the actual energy of the LHC collisions.

Several scenarios will be analyzed separately by varying the masses of χ, s, and Z'. In each case, two of the masses are fixed while a scan is performed over a range of values for the third. The ranges used are shown in Table 2.1.

Setup	m_χ / GeV	m_s / GeV	$m_{Z'}$ / GeV
1	[100, 150, 200]	160	700
2	[200, 500]	300	1000
3	100	[160, 180, 200, 300]	700
4	100	160	[195, 200, 295, 300, 500, 1000, 1500]
5	100	300	[195, 200, 295, 300, 500, 1000, 1500]

Table 2.1: Mass configuration for each setup of generation.

The mass ranges used in this work has been defined obased on the minimal produccion cross section for which CMS could be sensitive.

2.3 How to analyse events?

Once the generation is completed, it creates a compressses `.gz` file of the LHE , which must be unpacked using the *gunzip* command, resulting in a `.lhe` file. In this work, however, the files are analyzed in SWAN (Service for Web-based ANalysis), which has access to EOS [15], where the files are transferred. EOS is a disk-based, scalable storage system developed at CERN to handle large-scale data, and SWAN is a web platform developed at CERN that uses Jupyter Notebooks [16] for data analysis. SWAN combines the storage and computing capabilities of EOS with capacity to create notebooks to analyze the LHE.

The analysis in SWAN is mainly performed with ROOT [17]. ROOT is a data analysis framework developed at CERN in C++, which provides powerful tools for handling large datasets, including statistical tools and TTrees [17]. TTrees are data structures that store large collections of real/simulated events. Each row or *entry* in a TTree corresponds to a single event (in this case a collision), while each column represents a variable or set of variables, as the transverse momemntum of each particle, the energy, the invariant masses and other new variables defined. To transform the information of the LHE files into ROOT files in order to store the information of the events into TTrees, a C++ script was used within ROOT. This script was extended to include additional variables needed for this work, which are further discussed in Section 3.3.

For each sample generated one TTree is created. The analysis of this ROOT files is performed in SWAN using PyROOT. PyROOT provides Python bindings for ROOT, enabling interaction between Python scripts and ROOT objects, such as TTrees. This approach facilitates efficient event analysis in Python using ROOT functionalities. In addition, the analysis makes use of Numpy [18],

Uproot [19], and standard Python libraries such as `subprocess` and `random` [20].

All the code developed throughout this project is publicly accessible in a GitHub repository I have created [21].

3 Results

3.1 Study of the Dark Higgs Decay Channels

First, starting from the Feynman diagrams in Figure 1.2, it is necessary to analyze the possible decay channels of the s. The dark Higgs has the same decay channels as the SM Higgs, so it can decay into two W bosons, a pair of bottom quarks ($b\bar{b}$), top quarks ($t\bar{t}$), two Z bosons, or two SM Higgs bosons (h), among other that will not be considered. To perform a more precise analysis, al background should be studied.

The Branching Ratio (BR) of each decay channel has been estimated for a range of s masses, with $m_s \in [0, 350]$ GeV. This was done via a parameter scan similar to the one described in Section 2.2, using MadGraph generation. Additionally, a script named `getXsec.sh` was used on *lxplus* to store the BR from each simulation in a text file, facilitating data collection. Using these values, the BR of each channel as a function of m_s has been calculated.

The BR represents the probability that a particle, in this case, the dark Higgs, decays into a specific channel. The relationship between the BR and the width (Γ) of each channel (σ_i) is given by Equation 3.1.

$$BR_i = \frac{\Gamma_i}{\Gamma_T} \tag{3.1}$$

where Γ_T is the sum of the widths of all decay channels.

Based on this, the BR for each decay chnannel is shown in Figure 3.1 as a funcition of the dark higgs mass. As can be observed, the dominant decay channel if $m_s < 160$ GeV is the $b\bar{b}$ channel. This is because the bottom quark mass is $m_b \approx 4.18$ GeV [22], while the final state particles of the other channels are more massive. In this mass range, the dark Higgs is heavy enough to decay on-shell into $b\bar{b}$, but a higher mass is required to produce the particles of the other channels. Decays into photons and gluons are less probable because these processes are not LO. Decays into lighter fermions are also less likely because the Higgs boson, responsible for the decay through the mixing angle, couples to fermions with a strength proportional to the fermion mass. In this range

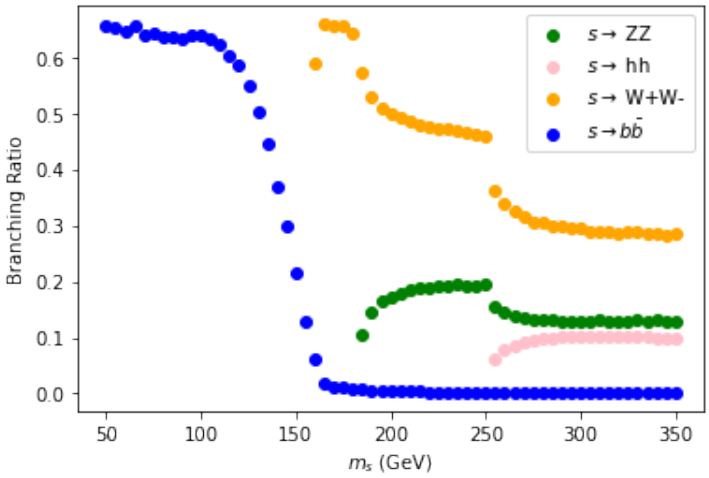


Figure 3.1: BR for each decay channel of the dark Higgs boson (s) as a function of its (m_s), obtained in MadGraph.

($m_s < 160$ GeV), the b quark is the heaviest accessible fermion, making the $b\bar{b}$ channel the most probable.

Around 160 GeV, the dominant decay channel is W^+W^-. This is due to the fact that the mass of the W boson is approximately 81 GeV [22]. This explains the high BR values around 160 GeV, which corresponds to twice the mass of the W boson.

On the other hand, at slightly higher masses, the ZZ channel appears, around 180 GeV, corresponding to twice the mass of the Z boson, being $m_Z \approx 91$ GeV [22]. There are four factors that lead to a factor of 2 difference between the BR of W^+W^- and ZZ, in favor of W^+W^-. The first factor is in favor of a larger BR for the ZZ channel, while the other three are in favor of the W^+W^- one.

The first one, the one favouring the ZZ channel, is that the Higgs couplings to weak bosons are proportional to the square of their masses: $g_{h,WW} \propto M_W^2$ and $g_{h,ZZ} \propto M_Z^2$. This results in a larger BR in favor of ZZ due to its higher mass. This is a direct consequence of electroweak symmetry breaking in the Higgs mechanism, a topic that will not be discussed in detail in this study [23].

The second one, is a reduction in phase space, due to the larger mass of the Z boson, although the difference is small, a phenomenon that will be explained in more detail in section 3.2.
The third one, is because, the decay into two identical Z bosons requires a symmetry factor of $1/2$ to avoid double counting of final states. This factor does not appear in the W^+W^- since both are distinguisable particles.

The fourth, appears because the W and Z bosons appear from the breaking of the $SU(2)_L \times U(1)_Y$ symmetry. The precise numerical coefficients in the interaction terms, which determine the decay widths, appear from the gauge structure and the way the Goldstone bosons (the degrees of freedom of the Higgs field) become W and Z. This phenomenon occurs in favor of a higher BR in the W^+W^- channel. In summary, this explanation is far from trivial and, as it is not a central objective of this study, it will not be explored in depth [23].

Finally, the hh channel appears around 250 GeV, which corresponds to twice the Higgs boson mass [22]. It is also observed that the appearance of this new channel reduces the bBR of both the W^+W^- and ZZ channels. The BR of this process is smaller compared to the others, since the coupling between the SM Higgs and the dark Higgs bosons is proporcional to sen$^2\theta$ [24], while θ is the mixing angle, set in $\theta = 0.01$.

With all this information, and thanks to the BR data from Figure 3.1, it is necessary to choose which decay channel will be studied in detail throughout the work. A good candidate is the $b\bar{b}$ channel, but the $b - jets$ produced in this decay are complex to analyze in MadGraph. Therefore, the W^+W^- channel will be chosen, as it has the highest BR after the $b\bar{b}$ channel.

In the W^+W^- channel, once both bosons are produced, they decay into a lepton-neutrino pair (l and ν_l), specifically: $W^+ \to l^+\nu_l$ and $W^- \to l^-\bar{\nu}_l$. They can also decay into quarks, but it will not be study for the same reason as the $b\bar{b}$ channel. These leptons and neutrinos, along with the DM particles, are the final particles that could be leave signal in the detector, and hence they are the only ones used in the current work. Also, during the simulation, τ and ν_τ are not considered. It is only considered properly electronic and mounic leptons because tauonic decays are not considered in MadGraph.

3.2 Cross Section of the Events

In this section, the effect of the mass of each dark sector particle to the cross section of the process will be analyzed. The command needed to generate the process is:

```
generate p p > zp > dm dm hs, (hs > w+ w- > l+ vl l- vl )
```

where the first part is the same as the one in the Section 2.3. Then the dark Higgs decays into two W^+W^-, that later decay into two leptons with opposite charge, one neutrino and one anti-neutrino. So, once all generations of the W^+W^- channel have been completed, the corresponding cross sections are collected in Table 3.1.

First, the variation of the total cross section as a function of m_s will be analyzed. In agreement with Figure 3.1, it can be observed that in all scenarios with $m_s = 300$ GeV, the BR is significantly reduced, affecting the total cross section. This makes any statistical analysis in these cases really complicated, with the current luminosity at the LHC. This issue will be discussed in detail in Section 3.6.

To better understand how the cross section depends on the masses m_χ and $m_{Z'}$, one can observe the two diagrams in Figure 1.2. Also, in Figure 3.2, the generated invariant mass of the two DM particles, is shown, for three different DM masses of 100, 150 y 200 GeV.
This variable is not experimentally accessible, since DM particles are invisible to the detectors and their invariant mass cannot be reconstructed. Nevertheless, the two peaks observed in the plot allow us to distinguish between the two diagrams contributing to the process. The biggest peak corresponds to the left diagram in Figure 1.2, where the DM particles are produced directly from the Z' decay. By cuadrimomentum conservation, their invariant show a resonance at the mass of the Z', which in this case is 700 GeV.

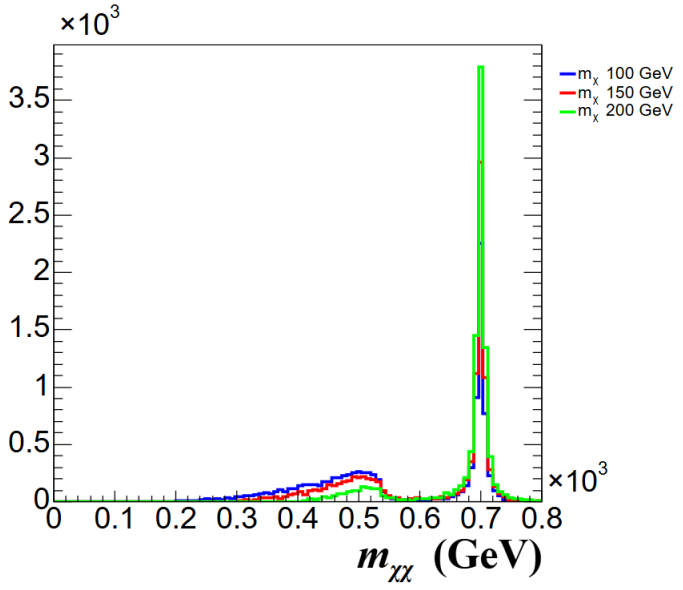


Figure 3.2: Number of events as a function of the invariant mass of both DM particles ($m_{\chi\chi}$), with $m_s = 160$ GeV and $m_{Z'} = 700$ GeV.

On the other hand, the smaller peak around 500 GeV corresponds to the right diagram in Figure 1.2, where the invariant mass of the DM particles is not at $m_{Z'}$, but rather $m^2_{Z'} = (p_{\chi_1} + p_{\chi_2} + p_s)^2$. Qualitatively, the peak appears around $m_{\chi\chi} \approx m_{Z'} - m_s = 540$ GeV.
Also, it is evident that most of the events are produced by an on-shell Z' boson.

Taking all this into account, three kinematic limits can be identified depending on m_χ and $m_{Z'}$. If $2m_\chi > m_{Z'}$, which occurs in several scenarios, the cross section drops drastically. For example, in the case $m_s = 160$ GeV and $m_\chi = 100$ GeV, Equation 3.2 applies:

$$\sigma(m_{Z'} = 295 \text{ GeV}) \approx 53 \cdot \sigma(m_{Z'} = 195 \text{ GeV}) \tag{3.2}$$

In this example, the configuration with $m_{Z'} = 195$ GeV satisfies the condition $2m_\chi > m_{Z'}$, and the 53 factor appears from dividing both cross sections using the data located in Table 3.1.

This occurs because, the Z' is required to decay into a pair of χ particles. For this to occur kinematically, the condition $m_{Z'} > 2m_\chi$ must be satisfied due to cuadrimomentum conservation. Therefore, when this condition is not fulfilled, neither of both diagrams contribute to the cross section.

The smaller peak in Figure 3.2 corresponds to a decay of Z' such that $Z' \to \chi\chi s$. This implies that, in order for the process to be kinematically allowed, the condition $2m_\chi + m_s > m_{Z'}$ must be satisfied due to cuadrimomentum conservation. This introduces a second kinematic limit, which removes the lower weight diagram if the condition is fulfilled, reducing the cross section.

On the other hand, when $m_{Z'}$ is sufficiently large to avoid these kinematic issues, a different phenomenon appears. Rather than the cross section continuing to increase with $m_{Z'}$, a slight decrease is observed. When the mass of Z' increases, due to cuadrimomentum conservation, a larger portion of the initial energy is used in the Z' mass, leaving less kinetic energy available for the products. This reduces the phase space volume and, consequently, the total cross section.

m_χ / GeV	m_s / GeV	$m_{Z'}$ / GeV	σ / pb	$\Delta\sigma$ / pb
100	160	700	0.01391	$2 \cdot 10^{-5}$
150	160	700	0.008927	$2 \cdot 10^{-6}$
200	160	700	0.005107	$4 \cdot 10^{-6}$
200	300	1000	0.001063	$7 \cdot 10^{-6}$
250	300	1000	$7.555 \cdot 10^{-4}$	$6 \cdot 10^{-7}$
300	300	1000	$5.311 \cdot 10^{-4}$	$4 \cdot 10^{-7}$
400	300	1000	$2.362 \cdot 10^{-4}$	$4 \cdot 10^{-7}$
500	300	1000	$1.048 \cdot 10^{-5}$	$4 \cdot 10^{-7}$
100	160	700	0.01391	$4 \cdot 10^{-5}$
100	180	700	0.06424	$2 \cdot 10^{-5}$
100	200	700	0.09136	$9 \cdot 10^{-5}$
100	300	700	$1.602 \cdot 10^{-5}$	$7 \cdot 10^{-8}$
100	160	195	0.01739	$4 \cdot 10^{-5}$
100	160	200	0.02722	$3 \cdot 10^{-5}$
100	160	295	0.9739	$6 \cdot 10^{-4}$
100	160	300	0.9921	$5 \cdot 10^{-4}$
100	160	500	0.8925	$5 \cdot 10^{-4}$
100	160	1000	0.3264	$9 \cdot 10^{-4}$
100	160	1500	0.0938	$7 \cdot 10^{-5}$
100	300	195	$1.187 \cdot 10^{-6}$	$2 \cdot 10^{-9}$
100	300	200	$1.681 \cdot 10^{-6}$	$3 \cdot 10^{-9}$
100	300	295	$7.397 \cdot 10^{-5}$	$7 \cdot 10^{-8}$
100	300	300	$7.819 \cdot 10^{-5}$	$9 \cdot 10^{-8}$
100	300	500	$1.833 \cdot 10^{-4}$	$2 \cdot 10^{-7}$
100	300	1000	$4.625 \cdot 10^{-4}$	$5 \cdot 10^{-7}$
100	300	1500	$4.014 \cdot 10^{-4}$	$6 \cdot 10^{-7}$

Table 3.1: Cross section for each generated samples with MadGraph. The cross section uncertainty is estimated by MadGraph.

The third kinematic limit occurs when $m_s > 2m_\chi$. When this condition is fulfilled, a new decay channel s appears: $s \to \chi\chi$. This channel has a larger BR than the W^+W^- channel, since $BR(s \to \chi\chi) \propto g_\chi^2$ with $g_\chi = 1$, while $BR(s \to W^+W^-) \propto \sin^2\theta$, and $\sin^2\theta \approx 10^{-5}$. As a result, the presence of this channel suppresses the decay into W^+W^-, no leptons are produced, and the scenario becomes unobservable for this analysis.

These are the effects of varying masses on the total cross section. The different orders of magnitude estimated in Table 3.1 are the result of a combination of several factors, some of which tend to increase the total cross section, while others reduce it.

3.3 Event Analysis

As previously mentioned in Section 2.3, the starting point was a C++ script, to which new variables were added. These variables and their definitions are shown in Appendix B. All the variables were defined under the CMS coordinate system, shown in Figure 3.3 and are based only in detectable observables at CMS.

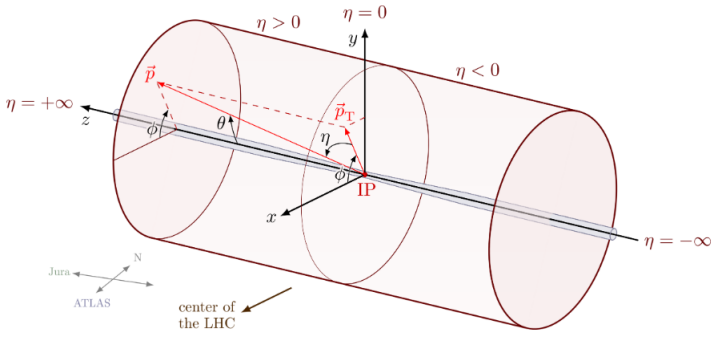


Figure 3.3: CMS coordinate system [25].

In Figure 3.3, ϕ is the azimuthal angle, which measures the angle in the x-y plane from the x-axis and ranges in $\phi \in [-\pi, \pi]$. Meanwhile, η is the pseudorapidity, which describes the angle relative to the z-axis and is defined in Eq. 3.3, with θ being the polar angle.

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) \tag{3.3}$$

Also, it is worth taking a closer look at some of the most relevant variables added. Additionally, since a ROOT environment was used, variables must be properly defined. For this, the variable `ptl1`, corresponding to the transverse momentum (x-y plane) of the leading lepton, is used as an example:

Listing 3.1: How to correctly define a variable in ROOT.

```
float ptl1_; // Create a variable to store ptl1
tree->Branch("ptl1", &ptl1_, "ptl1/F"); // Create a branch in the ROOT tree to store it
ptl1_ = -99; // Initialize with a default value
ptl1_ = v_tlv_all_leptons.at(0).GetP4().Pt(); // Assign pT of the first lepton in the
    list
```

In a real experiment, there are, on one hand, particles that appear in the final state: electrons, muons, neutrinos, and the lack of cuadrimomentum produced by DM particles. On the other hand, there are particles that decay at the collision point and can therefore only be studied through their decay products in a real experiment. So, it will be the same in the analysis.

Moreover, since DM and neutrinos do not interact electromagnetically or via the strong force, it would not leave any direct trace in the detectors. However, it could be seen via the energy conservation in the transversal plane x-y. It is defined as the vector sum of the negative transverse momentum (in the x–y plane) of each reconstructed particle in each collision, and then the magnitude of this sum is used.

This type of signal is known as missing transverse energy (MET), and while it could suggest the presence of DM, neutrinos produce a similar effect, complicating the analysis. At generator level, it can be defined as the sum of the transverse momentum of neutrinos and DM particles, since these are the undetectable particles at the LHC. The MET is, in principle, the most distinctive one in DM production, although later analysis will show that its relevance may not be as significant as initially expected.

Another variable defined is the transverse mass (m_T) [26], which is defined in the transverse plane x-y. It is defined for one or more visible particles, whose transverse energy is E_T, together with the MET and the angle in the transverse plane between the visible object's direction and the MET, $\Delta\phi$. The expression is given in Eq. 3.4.

$$m_T = \sqrt{2E_T \cdot MET \cdot (1 - \cos(\Delta\phi))} \tag{3.4}$$

In addition to these, it will be seen that, the most important variables in the analysis are m_{ll} and ΔR. The m_{ll} corresponds to the invariant mass of the two leptons produced in the collision, defined in Eq. 3.5, and played a key role in Section 3.3.2.

$$m_{ll} = \sqrt{(p_{l_1} + p_{l_2})^2} \tag{3.5}$$

being p_l the cuadrimomentum of the final two leptons.

Finally, the angular separation (ΔR) is defined in Eq. 3.6, and quantifies how far apart two particles or particle groups are after the collision.

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{3.6}$$

One thing to consider when defining ΔR; which was used between various objects such as ΔR_{ll} (angular separation between the two leptons), $\Delta R_{ll,MET}$ (angular separation between the two leptons and the MET), $\Delta R_{l_1,MET}$ (angular separation between the most energetic lepton and the met), and $\Delta R_{l_2,MET}$ (angular separation between the least energetic lepton and the met); is to handle the definition of $\Delta\phi$ carefully. In order to keep $\Delta\phi \in [-\pi, \pi]$, the function `DeltaPhi` has been defined:

Listing 3.2: Function that ensures DeltaPhi stays within the correct range.

```
double DeltaPhi(double phi1, double phi2) {
double dphi = phi1 - phi2;
while (dphi > M_PI) dphi -= 2 * M_PI;
while (dphi < -M_PI) dphi += 2 * M_PI;
return dphi;
}
```

With all these variables defined, a detailed analysis of each scenario in Table 2.1 can be performed. This analysis tries to identify variables that can distinguish events with different masses, so that a statistical analysis can later be performed in Section 3.6. The following variables related to the energy of the particles are: MET, m_{ll}, the transverse mass of the two leptons and the MET ($mt_{ll,MET}$) and the transverse mass of each lepton and the MET ($mt_{l_1,MET}$ and $mt_{l_2,MET}$).

Also, variables related to angular distance have also been studied, being difference of ϕ of the two leptons ($\Delta\phi_{ll}$), the two leptons and the MET ($\Delta\phi_{ll,MET}$) and one of each leptons with the MET

($\Delta\phi_{l_1,MET}$ and $\Delta\phi_{l_2,MET}$). The same variables are defined but with η instead of ϕ ($\Delta\eta_{ll}$, $\Delta\eta_{ll,MET}$, $\Delta\eta_{l_1,MET}$ and $\Delta\eta_{l_2,MET}$) and the ones related with ΔR defined earlier.

3.3.1 DM Analysis

In this section, Setups 1 and 2 from Table 2.1 are analyzed. The defined variables will be studied to evaluate their ability to distinguish between various DM masses. The first setup corresponds to $m_\chi = [100, 150, 200]$ GeV with $m_s = 160$ GeV and $m_{Z'} = 700$ GeV, while the second one corresponds to $m_\chi = [200, 500]$ GeV with $m_s = 300$ GeV and $m_{Z'} = 1000$ GeV. As explained in Section 3.2, in order to study higher values of m_χ, it is needed to increase m_s to avoid the kinematic limit $s \to \chi\chi$.

Based on this, the analyzed variables are plotted in Figures 3.4, 3.5, and 3.6, where the distributions are normalized. The plots on the left correspond to the analysis of the first setup, while the second setup is analyzed on the right.

Although the MET is in principle a promising observable for distinguishing DM masses (since larger m_χ is expected to lead to higher MET), it is observed that in the first setup (located on the left plot of the first row in Figure 3.4) a dependency with m_χ is not seen. In the second setup (located on the right plot of the first row in Figure 3.4), although $m_\chi = 500$ GeV shows slightly higher MET, this difference is still not significant. This is because MET takes into account not only the DM particles, but also the neutrinos. The neutrinos are produced in the opposite direction to the DM particles, and since MET is defined as the negative vector sum of transverse momentum, the transverse momentum of the neutrinos tends to partially cancel that of the DM particles. This partial cancellation effect hides the real contribution of the DM particles to the MET, preventing the total value from increasing significantly with m_χ. As a result, MET loses much of its power as a discriminating variable, since it is diluted by the presence of neutrinos.

This effect propagates to other MET dependent variables such as the transverse masses $mt_{ll,MET}$, $mt_{l_1,MET}$, and $mt_{l_2,MET}$, located on the other four plots of Figure 3.4 and the first row of Figure 3.5. No significant differences between DM masses can be observed in these distributions.

The same happens with the angular variables, located on the last four plots of Figure 3.5 and all of Figure 3.6, including those related to angular separation (ΔR), $\Delta\phi$, and $\Delta\eta$, where no differentiation by m_χ is observed either. It is seen that the vector sum of the cuadrimomentum of the DM particles and neutrinos lies at a $\Delta\phi$ of approximately π with respect to the leading lepton, although neutrinos are produced in the same direction as the leptons, because $p(\chi\chi) \gg p(\nu\nu)$. This variable, located in the second row of Figure 3.6, is shown with respect to the leading lepton instead of both leptons, because in that case $|\Delta\phi_{ll,MET}|$ is practically π, and the tails are not visible. This explains the threshold around three observed in $\Delta R_{ll,MET}$, since, according to Eq. 3.6, and assuming $\Delta\eta_{ll,MET} = 0$, $\Delta R_{ll,MET} \approx |\Delta\phi_{ll,MET}| \approx \pi$.

With all this, it is concluded that m_χ is not a suitable parameter for the analysis, as neither of the variables studied are able to differentiate DM masses.

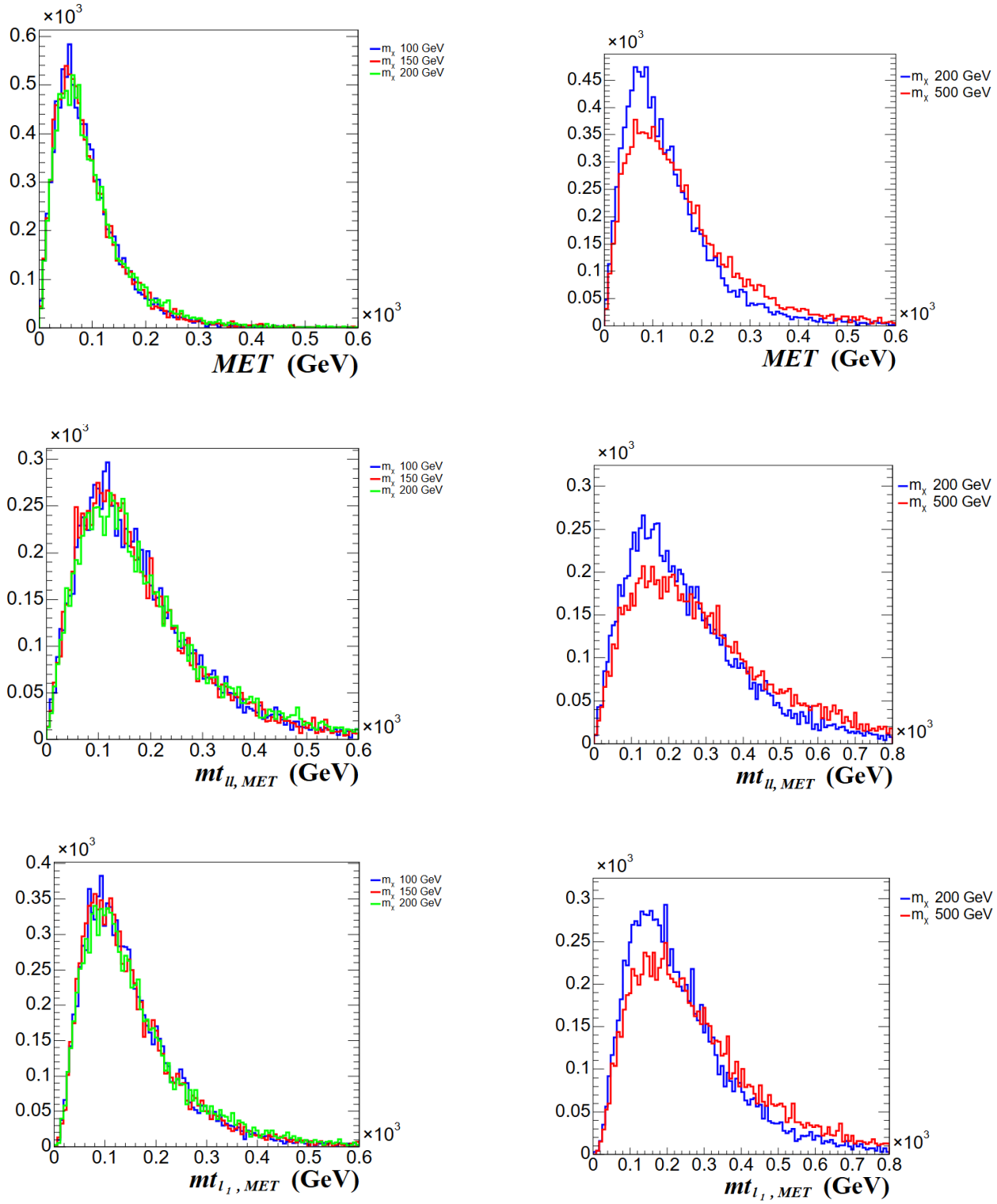


Figure 3.4: Number of events as a function of the MET, $mt_{ll,MET}$ and $mt_{l_1,MET}$ with $m_\chi = [100, 150, 200]$ GeV, $m_s = 160$ GeV and $m_{Z'} = 700$ GeV on the left. While, $m_\chi = [200, 500]$ GeV, $m_s = 300$ GeV and $m_{Z'} = 1000$ GeV on the right. Neither of the distributions are weighted by the cross section.

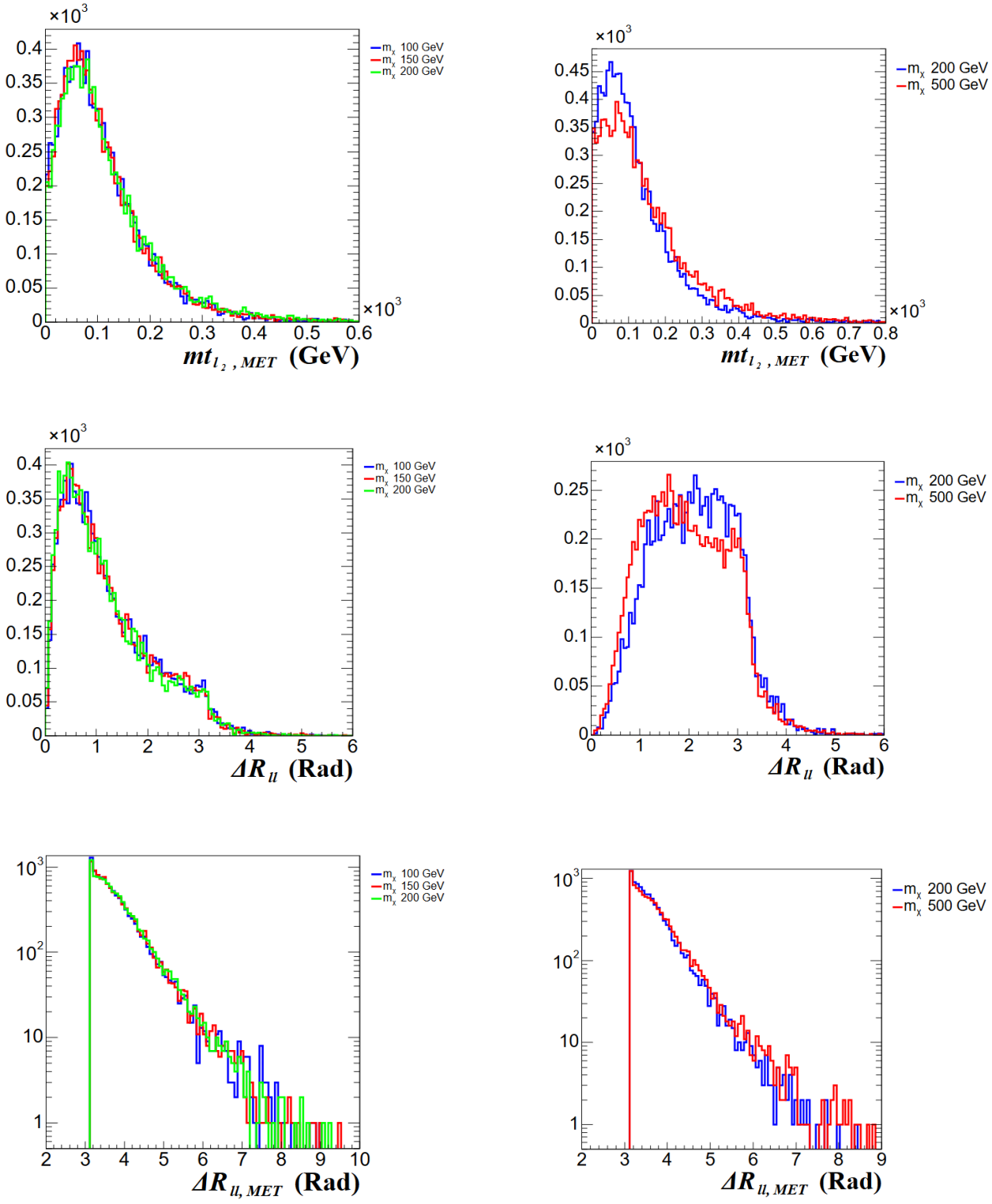


Figure 3.5: Number of events as a function of the $mt_{l_2,MET}$, ΔR_{ll} and $\Delta R_{ll,MET}$ with $m_\chi = [100, 150, 200]$ GeV, $m_s = 160$ GeV and $m_{Z'} = 700$ GeV on the left. While, $m_\chi = [200, 500]$ GeV, $m_s = 300$ GeV and $m_{Z'} = 1000$ GeV on the right. Neither of the distributions are weighted by the cross section.

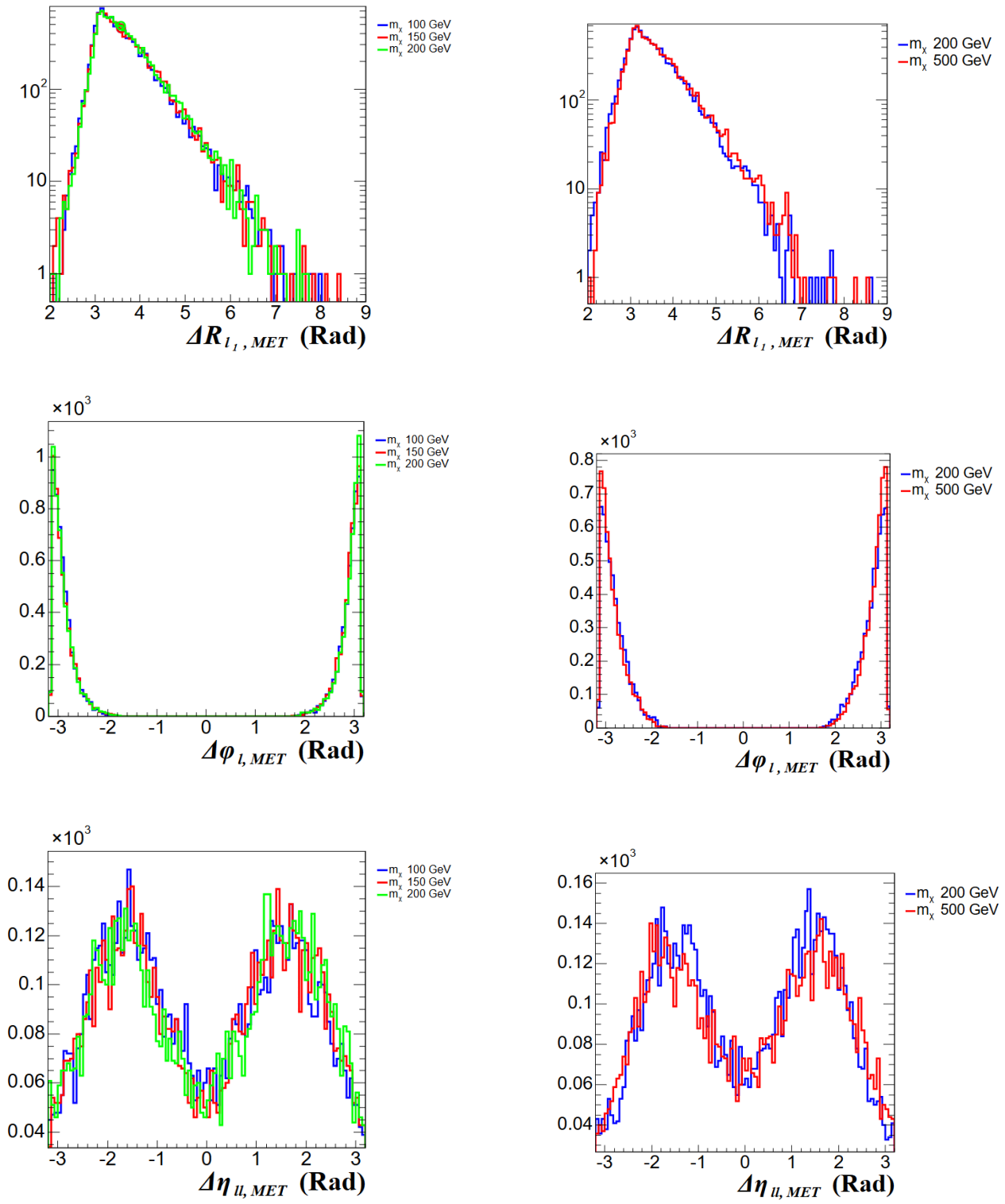


Figure 3.6: Number of events as a function of the $\Delta R_{l_1,MET}$, $\Delta\phi_{l,MET}$and $\Delta\eta_{ll,MET}$ with $m_\chi = [100, 150, 200]$ GeV, $m_s = 160$ GeV and $m_{Z'} = 700$ GeV on the left. While, $m_\chi = [200, 500]$ GeV, $m_s = 300$ GeV and $m_{Z'} = 1000$ GeV on the right. Neither of the distributions are weighted by the cross section.

3.3.2 s Analysis

In this section, the third setup from Table 2.1 is analyzed, where $m_s = [160, 180, 200, 300]$ GeV, $m_\chi = 100$ GeV and $m_{Z'} = 700$ GeV. So, the dependency of the variables defined with m_s is going to be studied. The corresponding distributions are shown in Figures 3.7 and 3.8.

As in the previous section, the MET and $mt_{ll,MET}$ are not able to clearly differentiate m_s, this can be seen in Figure 3.7, which is why $mt_{l_1,MET}$ and $mt_{l_2,MET}$ are not displayed, since they do not provide additional information.

However, unlike the earlier case, the remaining variables are able to discriminate m_s. The least effect is seen in $\Delta\eta_{ll}$ (located in the first plot on the last row of Figure 3.8), where as m_s increases, the dark Higgs is produced with less kinetic energy, since more energy is spent on generating its mass. As a result, when it decays into W^+W^- and later into leptons, the leptons are emitted less boosted, leading to a higher $\Delta\eta_{ll}$. This behavior is more visible in the $\Delta\phi_{ll}$ distribution, in the last plot of Figure 3.8. When $m_s = 300$ GeV, where s is produced nearly at rest, the leptons are produced almost back-to-back, rather than highly aligned, where $\Delta\phi_{ll} \approx 0$, as observed in the other cases. A similar pattern appears in ΔR_{ll}, where the angular separation is larger for $m_s = 300$ GeV compared to the other mass values, because of the same phenomenon, which is shown in the first plot of Figure 3.8.
Additionally, since both leptons originate from the decay $s \to W^+W^- \to l^+l^-$, the invariant mass of both leptons (m_{ll}) increases with m_s, this effect is shown in the second plot of the first row of Figure 3.8.

Overall, in contrast to the case of m_χ, the variable m_s does provide a clear way to differentiate between mass values. Two mass regimes can be identified based on m_s: one for $m_s = [160, 180, 200]$ GeV and another for $m_s = 300$ GeV. This distinction motivates the analysis in the next section, where different Z' mass ranges will be explored within these two regimes: low and high values of m_s. This will allow Section 3.5 to treat both scenarios separately.

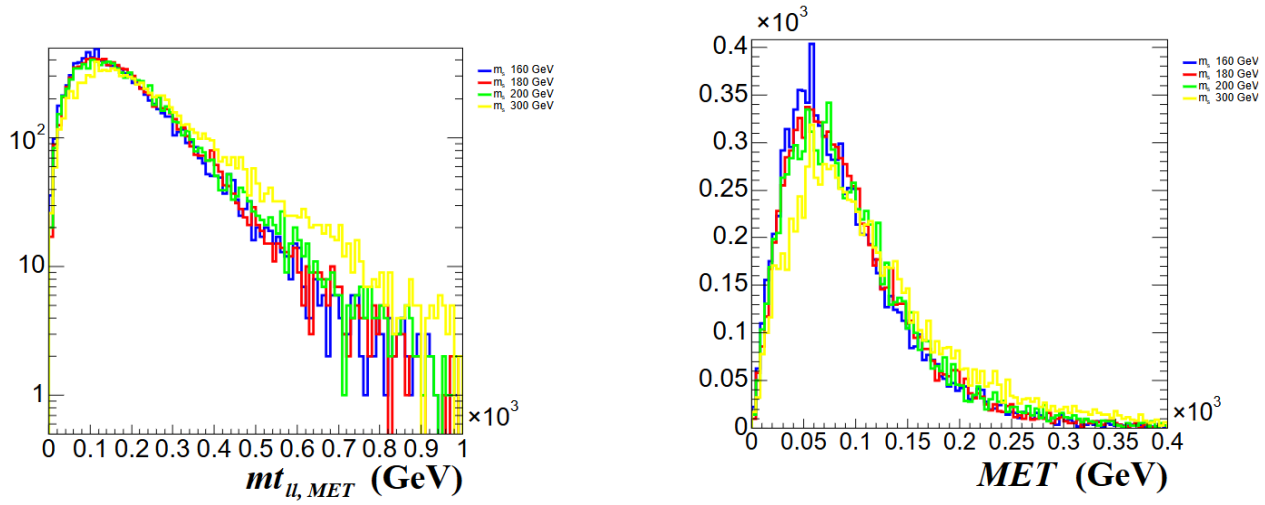


Figure 3.7: Number of events as a function of the $mt_{ll,MET}$ and MET with $m_s = [160, 180, 200, 300]$ GeV, $m_\chi = 100$ GeV and $m_{Z'} = 700$ GeV. Neither of the distributions are weighted by the cross section.

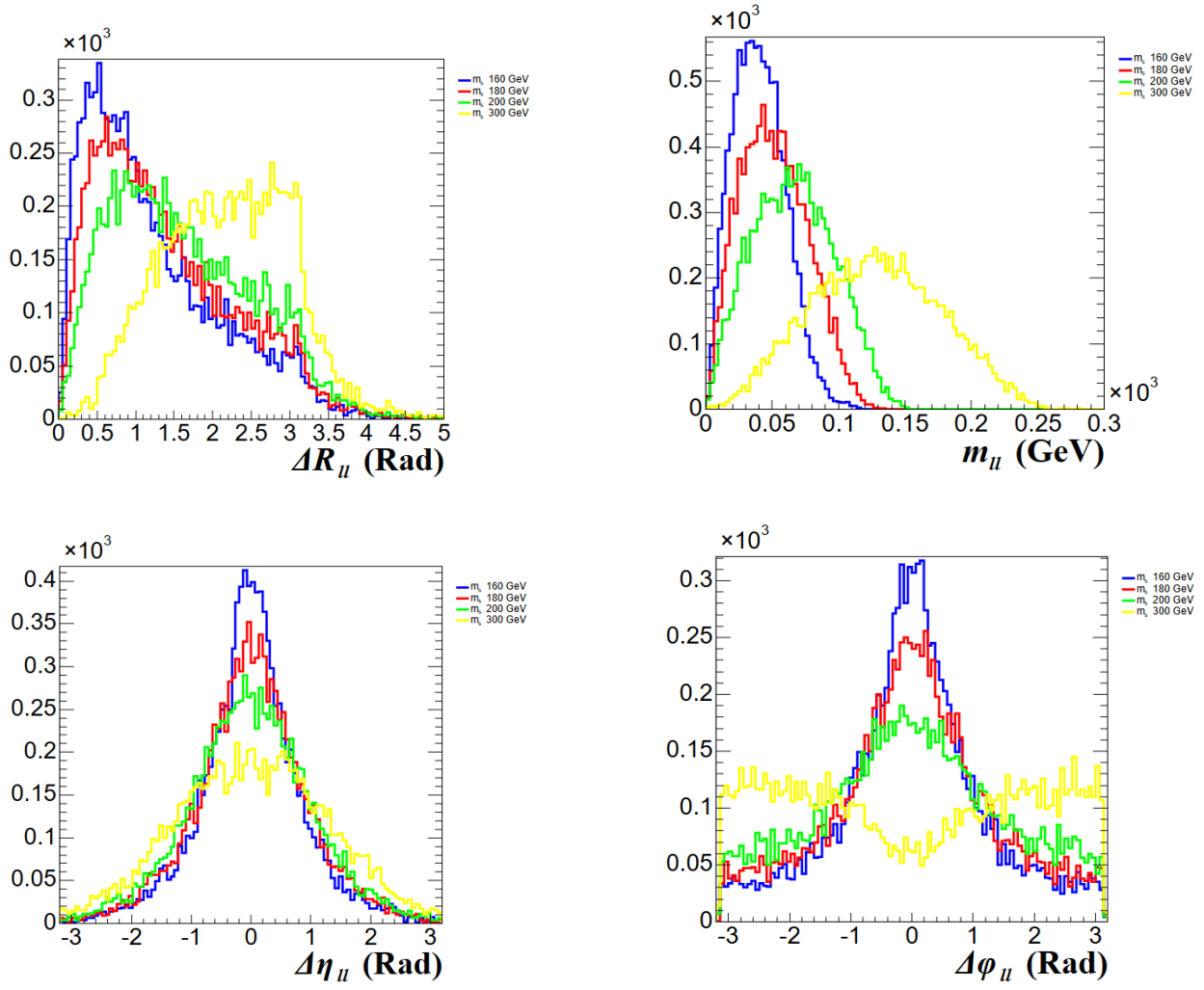


Figure 3.8: Number of events as a function of ΔR_{ll}, m_{ll}, $\Delta\eta_{ll}$ and $\Delta\phi_{ll}$ with $m_s = [160, 180, 200, 300]$ GeV, $m_\chi = 100$ GeV and $m_{Z'} = 700$ GeV. Neither of the distributions are weighted by the cross section.

3.3.3 Z' Analysis

In this section, the last two setups from Table 2.1 are analyzed, corresponding to a scan over $m_{Z'} = [195, 200, 1000, 1500]$ GeV with $m_\chi = 100$ GeV in two different regimes: $m_s = 160$ GeV and $m_s = 300$ GeV. So, the dependency of the variables defined with $m_{Z'}$ is going to be studied in each regime. The results are shown in Figures 3.9, 3.10 and 3.11.

Firstly, for the same reasons discussed at the beginning of Section 3.3.1, it is observed on Figure 3.9 that the MET $mt_{ll,MET}$ distributions, can not discriminate $m_{Z'}$. With the exception of a small increase in the higher values of $m_{Z'}$ because the DM particles are produced from the Z' decay, so higher $m_{Z'}$ results in more energetic DM. However, as already mentioned in Section 3.3.1, the contribution of neutrinos counteracts this effect.

On the other hand, the strong differentiation observed in m_{ll} and ΔR_{ll} in Section 3.3.2, is still clear across the two regimes in first and last rows of Figure 3.10. In the case of ΔR_{ll}, it is seen that as $m_{Z'}$ increases, its decay products become more energetic, resulting in leptons emitted more boosted, which reduces ΔR_{ll}. However, this effect is not particularly pronounced. Moreover, the m_{ll} distribution shows little variations in $m_{Z'}$, since m_{ll} mainly depends on m_s.

A small separation between different $m_{Z'}$ values is observable in $\Delta R_{ll,MET}$ and $\Delta R_{l_1,MET}$ on the Figure 3.11. This due to the fact that a heavier Z' is produced with less kinetic energy and closer to being at rest, leading to a more back-to-back configuration of its decay products and a larger angular separation between the leptons and the MET. Again, this distinction is diluted by the presence of neutrinos. The same trend can be seen in $\Delta\eta_{ll,MET}$, where a heavier Z' results in larger angular separation on the middle row of Figure 3.10.

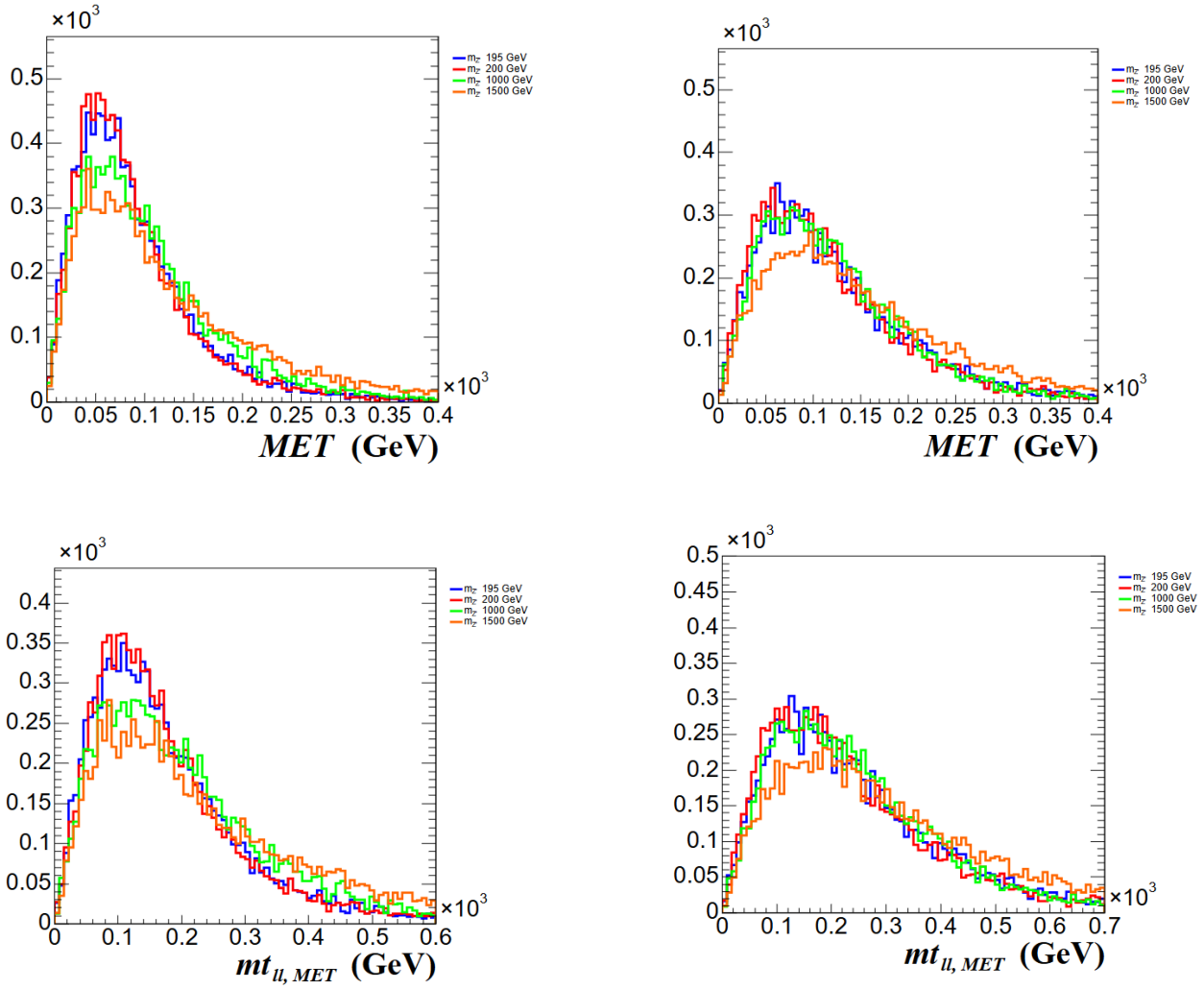


Figure 3.9: Number of events as a function of the MET and $mt_{ll,MET}$ with $m_{Z'} = [195, 200, 1000, 1500]$ GeV, $m_s = 160$ GeV and $m_\chi = 100$ GeV on the left. While, $m_{Z'} = [195, 200, 1000, 1500]$ GeV, $m_s = 300$ GeV and $m_\chi = 100$ GeV on the right. Neither of the distributions are weighted by the cross section.

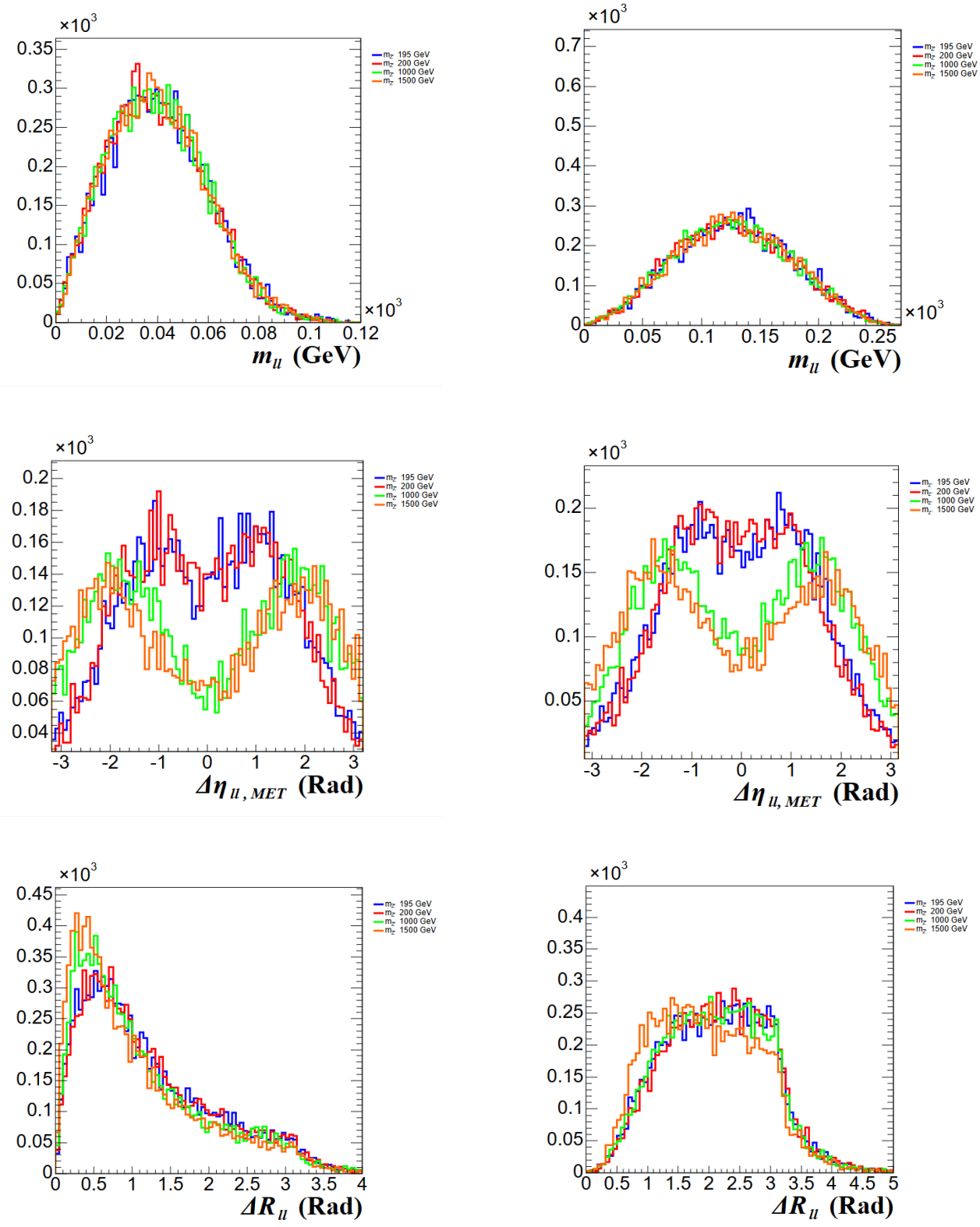


Figure 3.10: Number of events as a function of the m_{ll}, $\Delta\eta_{ll,MET}$ and ΔR_{ll} with $m_{Z'} = [195, 200, 1000, 1500]$ GeV, $m_s = 160$ GeV and $m_\chi = 100$ GeV on the left. While, $m_{Z'} = [195, 200, 1000, 1500]$ GeV, $m_s = 300$ GeV and $m_\chi = 100$ GeV on the right. Neither of the distributions are weighted by the cross section.

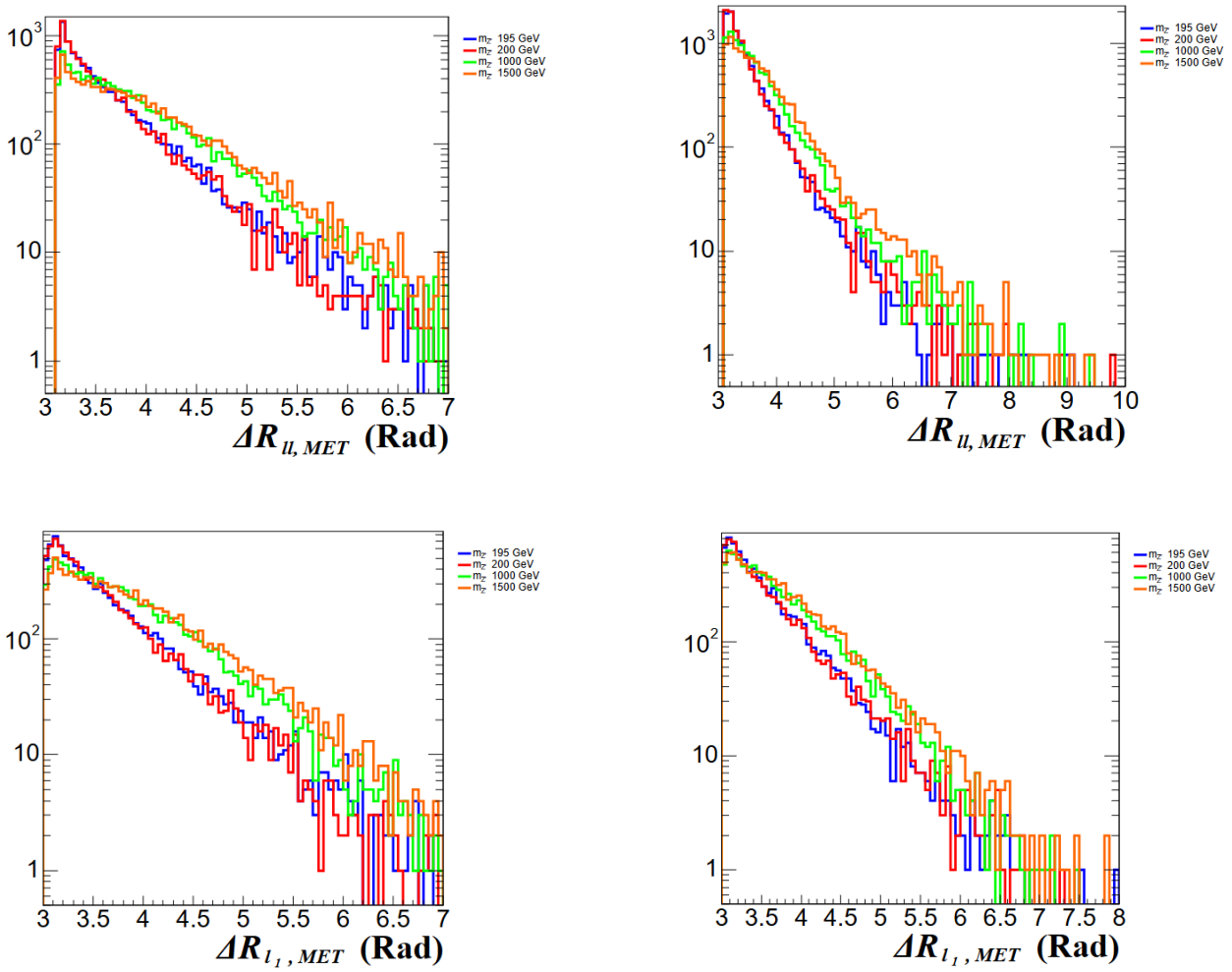


Figure 3.11: Number of events as a function of the $\Delta R_{ll,MET}$ and $\Delta R_{l_1,MET}$ with $m_{Z'} = [195, 200, 1000, 1500]$ GeV, $m_s = 160$ GeV and $m_\chi = 100$ GeV on the left. While, $m_{Z'} = [195, 200, 1000, 1500]$ GeV, $m_s = 300$ GeV and $m_\chi = 100$ GeV on the right. Neither of the distributions are weighted by the cross section.

In conclusion, some variables differentiate $m_{Z'}$, but is not sufficient for clear mass discrimination. So, the two m_s regimes: high and low, will be studied as a function of the Z' mass, for the event selection and statistical analysis.

3.4 Backgrounds

In particle physics, the background refers to any physical process that mimics or resembles the signal being searched for, but is not the physical process of interest. The possible backgrounds that can hinder tagging the signal must be considered. To do this, SM processes with two leptons as final products are evaluated. Many backgrounds could be relevant, but in this particular case, the three with the most cross section will be studied. Ideally, all possible backgrounds should be considered.

First, it is considered the production of two W bosons decaying into leptons and neutrinos. This process, like the signal, results in two oppositely charged leptons and MET, which in this case originates only from the neutrinos. This process has a significant cross section in the energy range studied. There are three posibles diagrams that represent this process, that are shown in Figure 3.12.

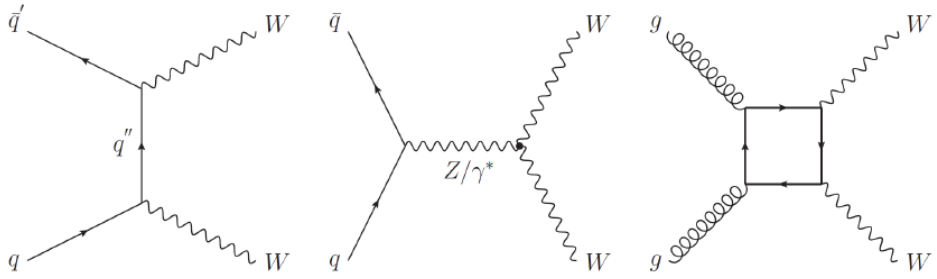


Figure 3.12: Three main Feynman diagrams of the W^+W^- at the LHC [26].

The second background to consider is the production of a SM Higgs boson decaying into W^+W^-. In this case, the final state products are the same as in the previous background. The cross section is smaller than that of the previous background, but it should be noted that near the Higgs resonance, around 125 GeV, this background becomes more problematic. Additionally, this diagram does not occur at LO, so next-to-leading order (NLO) calculations are required to simulate it. In NLO calculations, the first and second terms in the perturbative expansion is taken into account. This includes more diagrams with more vertex, such as loops. The diagram that represents this process is shown in Figure 3.13.

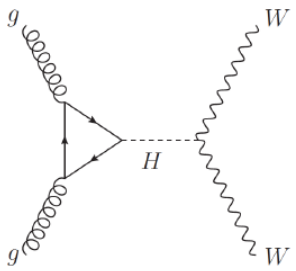


Figure 3.13: Feynman diagram of the $h \to W^+W^-$ at the LHC [26].

The last background considered is the production of a pair of quarks top ($t\bar{t}$), where both top quarks decay into a bottom quark and a W boson, which is the same final state as before plus two bottom quarks. The diagram that represents this collision is shown in Figure 3.15. This background has the highest cross section among all those considered. Being $\sigma_{t\bar{t}} \approx 500$ pb, while $\sigma_{WW} \approx 60$ pb and $\sigma_{h\to WW} \approx 2$ pb. Although it has two b-jets in the final state, the algorithms used to identify them (b-tagging) are not perfectly efficient and may classify events with two real b-jets as having none. For tracking this, it is assumed a b-tagging efficiency of 90% [26]. Therefore, the probability of detecting both $b-jets$ is $0.9 \cdot 0.9 = 0.81$, the probability of detecting none is $0.1 \cdot 0.1 = 0.01$, and the probability of detecting only one is 0.18. Thus, it is only necessary to multiply the cross section of this background by 0.01 to account for events with zero b-jets, as shown in Figure 3.14.

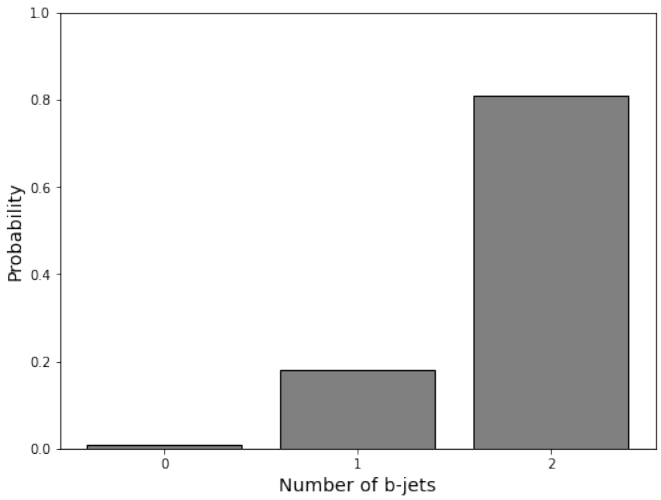


Figure 3.14: Probability of measuring cero, one or two b-jets on a $t\bar{t}$ background event.

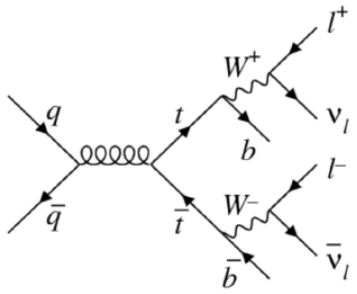


Figure 3.15: Feynman diagram of $t\bar{t} \to b\bar{b}\ W^+W^-$ at the LHC [26].

Lastly, a common issue in MadGraph generations, that occurs in any Monte Carlo generation, is the lack of events in the tails of the distributions when a huge number of events is not generated. This issue affects the analysis even after generating 10.000 events. This occurs because the cross sections of typical SM processes, such as the studied backgrounds, fall rapidly with energy, which creates the false impression in simulations that there are no events in these regions. However, when compared to signals with small cross sections, as in this study, there might even be more background events than signal events in the tails.

To solve this, the same background has been simulated again, but imposing a minimum transverse momentum to electrons or muons (pt_l): 100 GeV in the case of the $h \to W^+W^-$ background, and 150 GeV for the W^-W^+ and $t\bar{t}$ backgrounds. Then, both background have been combined into a single TTree, weighted by its corresponding `EvtWeight`, to obtain a new TTree that accounts for the tail behavior. An example of how the tails are recovered is shown in Figure 3.16, where one can see in logarithmic scale that the background including tails better reproduces the values of pt_{l_1} at higher energies for the WW background.

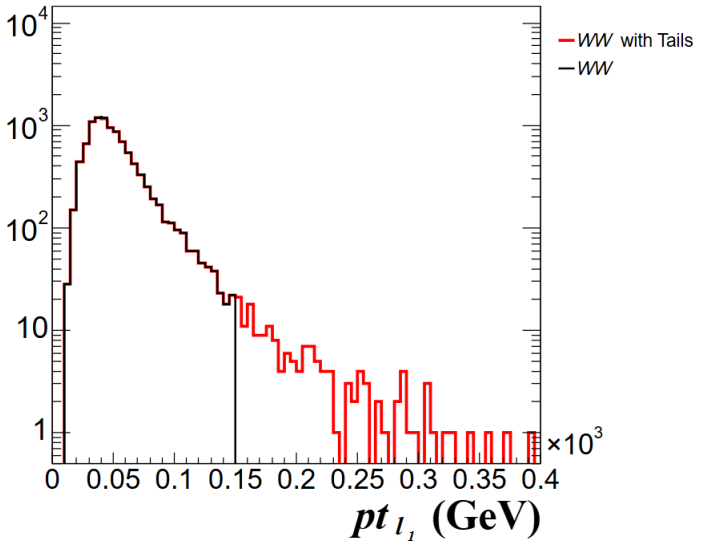


Figure 3.16: Number of events for the transverse momentum of the leading electron or muon in the WW background, comparing the distribution with and without the added tails.

3.5 Event Selection

Thanks to the analysis carried out in Section 3.3, it was concluded that one low m_s regime ($m_s = 160$ GeV) and one high m_s regime ($m_s = 300$ GeV) will be studied, for a range of Z' masses, specifically $m_{Z'} = [195, 200, 295, 300, 500, 1000, 1500]$ GeV.

To perform this analysis, one must find a way to isolate the background processes from the signal, in order to identify a signal region free from background contamination. To do this, the significance is used. The statistical significance is a parameter that helps determine the selection point at which the signal significance is maximized with respect to the background [27]. For this purpose, three different definitions will be used, each useful depending on the relation between the signal and the background.

The first is the so-called *Naive Significance*, which measures the statistical fluctuation of both background and signal. It is suitable when both the signal and the background contribute to the noise, and is defined in Eq. 3.7.

$$\textit{Naive Significance} = \frac{S}{\sqrt{S+B}} \tag{3.7}$$

The second to be considered is the *Poisson Significance*, which only considers the background fluctuation as noise. It assumes that the signal is negligible compared to the noise, as it equals the *Naive Significance* when $S \to 0$. Therefore, it is used when the signal is very small, and is defined in Eq. 3.8.

$$\textit{Poisson Significance} = \frac{S}{\sqrt{B}} \tag{3.8}$$

The last one is the *Asymptotic Significance* ($\sqrt{q_{0,A}}$) [28], used in modern analyses such as those at the LHC. It is accurate even with low background and signal. It is based on an asymptotic

approximation, and is defined in Eq. 3.9.

$$\sqrt{q_{0,A}} = \sqrt{2\left[(S+B)\ln\left(1+\frac{S}{B}\right) - S\right]} \tag{3.9}$$

To filter the signal, all three significances have been taken into account. Also, it must be noted that not all backgrounds have been considered, only three. Therefore, a loose cut has been applied. Because of this, the cut was not made exactly at the optimal point suggested by the significance values but was instead more permissive. With this in mind, the event selection was applied to both the high and low m_s regimes in the Z' distributions.

3.5.1 Event Selection in the Low-m_s Regime as a Function of $m_{Z'}$

As previously said, the chosen method for the event selection on the variables has been the use of significances. As an example, Figure 3.17 shows the selection applied to the variable ΔR_{ll}. As already explained, the event selectione+ is applied loosely due to the lack of background. Therefore, in the plot on the right, it can be seen that instead of choosing the maximum significance value at $\Delta R_{ll} \approx 0.5$ Rad, a more permissive selection is applied at $\Delta R_{ll} = 0.9$ Rad. In the left plot, it can be observed that this event selection is reasonable, as it allows a significant portion of the background to be removed without losing too much signal.

With all of this, the significances, indicating the type among the three studied in each case, are shown in Figures 3.18, 3.19, 3.20, 3.21, and 3.22. Also, the selection of each variable is shown in the caption of each figure.

It is worth noting that, depending on the case, the cuts were clearer using the *Naive Significance* or the *Asymptotic Significance*, but the information provided by the *Poisson Significance* was less helpful for cutting the signal.

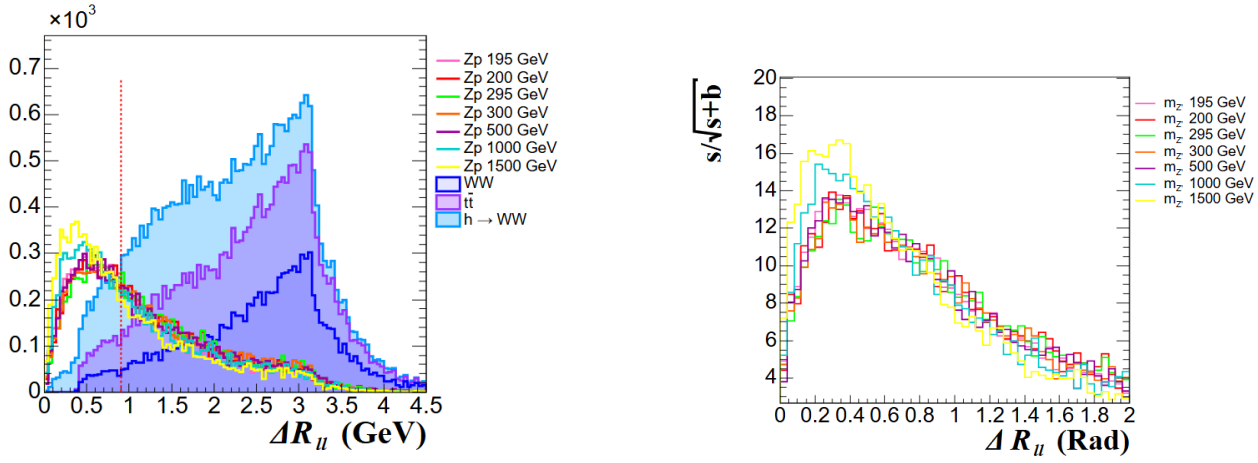


Figure 3.17: *Naive Significance* as a function of ΔR_{ll}, resulting in a loose selection in $\Delta R_{ll} < 0.9$ Rad, in the right. While, the number of events of every $m_{Z'}$ distribution as a function of ΔR_{ll} with the background, in the left. The red vertical line in the left figure represents the selection suggested from the *Naive Significance* in the right. This distributions correspond to the low-m_s regime.

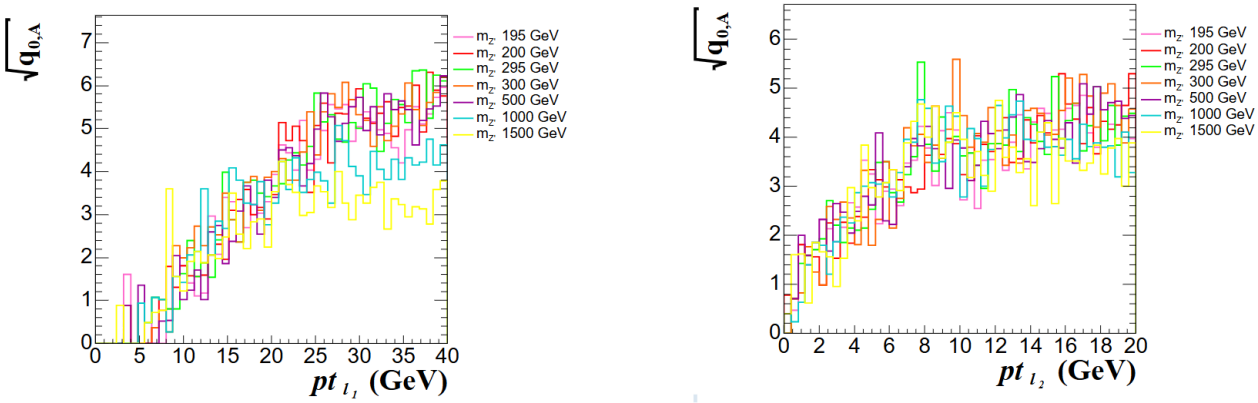


Figure 3.18: *Asymptotic Significance* as a function of pt_{l_1} that suggest a loose selection in $pt_{l_1} > 25$ GeV, in the left. While, the *Asymptotic Significance* as a function of pt_{l_2} in the right suggest a loose selection in $pt_{l_2} > 12$ GeV. This distributions correspond to the low-m_s regime.

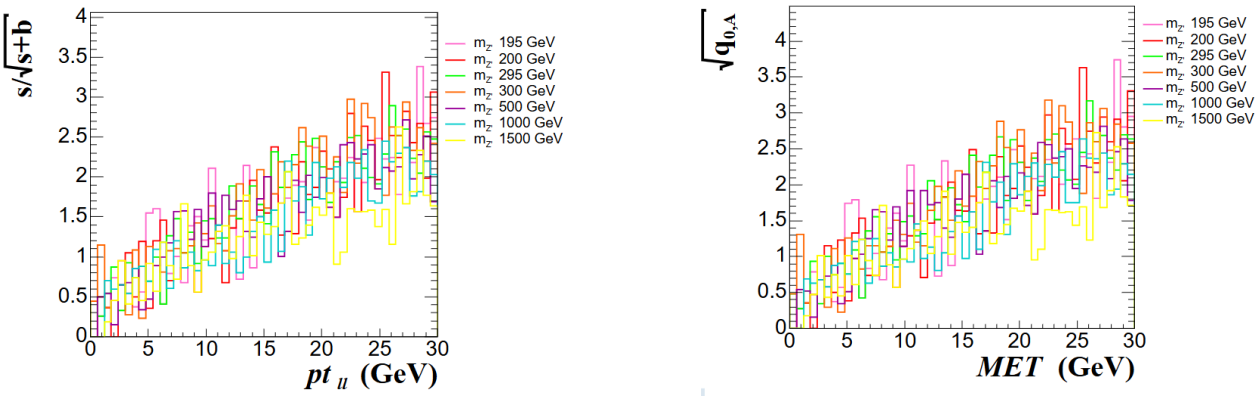


Figure 3.19: *Naive Significance* as a function of pt_{ll} that suggest a loose selection in $pt_{ll} > 20$ GeV, in the left. While, the *Asymptotic Significance* as a function of the MET in the right suggest a loose selection in $MET > 20$ GeV. This distributions correspond to the low-m_s regime.

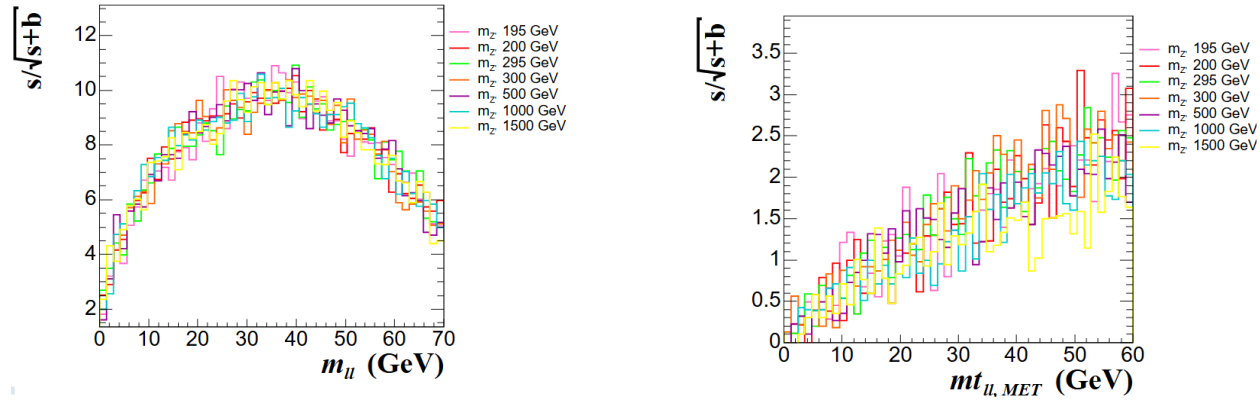


Figure 3.20: *Naive Significance* as a function of m_{ll} that suggest a loose selection in 20 GeV $< m_{ll} <$ 48 GeV, in the left. While, the *Naive Significance* as a function of $mt_{ll,MET}$ in the right suggest a loose selection in $mt_{ll,MET} > 40$ GeV. This distributions correspond to the low-m_s regime.

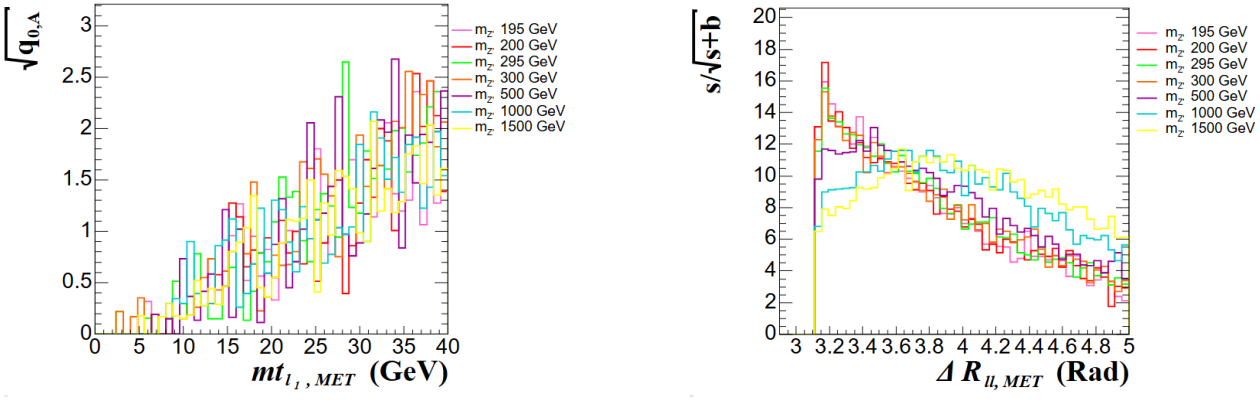


Figure 3.21: *Asymptotic Significance* as a function of $mt_{l_1,MET}$ that suggest a loose selection in $mt_{l_1,MET} > 30$ GeV, in the left. While, the *Naive Significance* as a function of $\Delta R_{ll,MET}$ in the right suggest a loose selection in $\Delta R_{ll.MET} > 3$ Rad. This distributions correspond to the low-m_s regime.

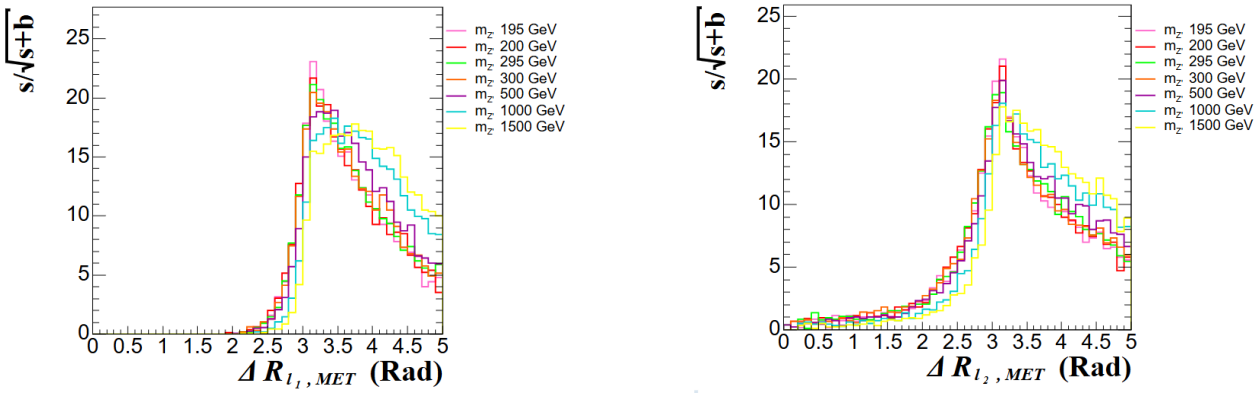


Figure 3.22: *Naive Significance* as a function of $\Delta R_{l_1,MET}$ that suggest a loose selection in $\Delta R_{l_1,MET} > 2.8$ Rad, in the left. While, the *Naive Significance* as a function of $\Delta R_{l_2,MET}$ in the right suggest a loose selection in $\Delta R_{l_2,MET} > 2.8$ Rad. This distributions correspond to the low-m_s regime.

3.5.2 Event Selection in the High-m_s Regime as a Function of $m_{Z'}$

Similarly to the previous section, the event selection is performed based on the significance values. The significances are shown in Figures 3.23, 3.24, 3.25, 3.26, and 3.27.

Unlike the low m_s regime, in this case, as shown in Table 3.1, the cross sections when $m_s = 300$ GeV drops drastically. These cross sections are approximately between 10^{-6} pb and 10^{-4} pb. This has led to a significant reduction in the number of events.

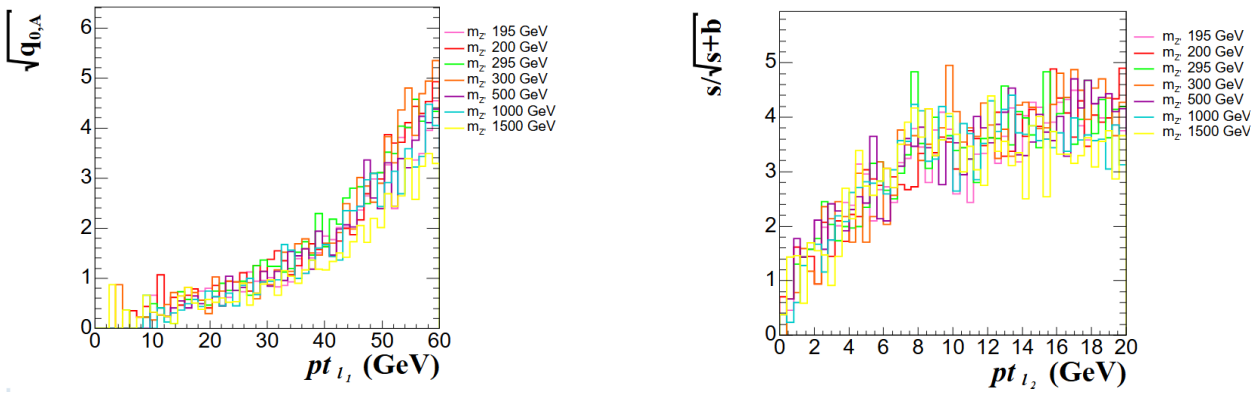


Figure 3.23: *Asymptotic Significance* as a function of pt_{l_1} that suggest a loose selection in $pt_{l_1} > 45$ GeV, in the left. While, the *Naive Significance* as a function of pt_{l_2} in the right suggest a loose selection in $pt_{l_2} > 10$ GeV. This distributions correspond to the high-m_s regime.

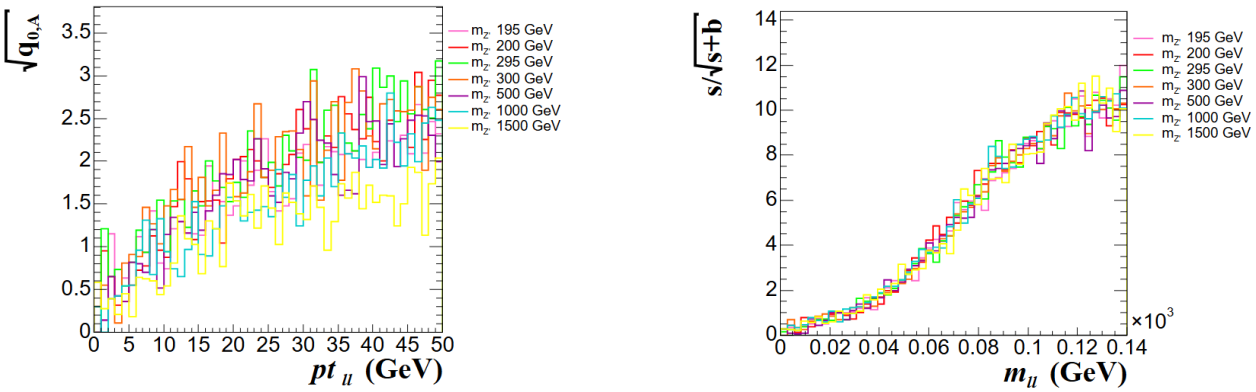


Figure 3.24: *Asymptotic Significance* as a function of pt_{ll} that suggest a loose selection in $pt_{ll} > 35$ GeV, in the left. While, the *Naive Significance* as a function of m_{ll} in the right suggest a loose selection in $m_{ll} > 110$ GeV. This distributions correspond to the high-m_s regime.

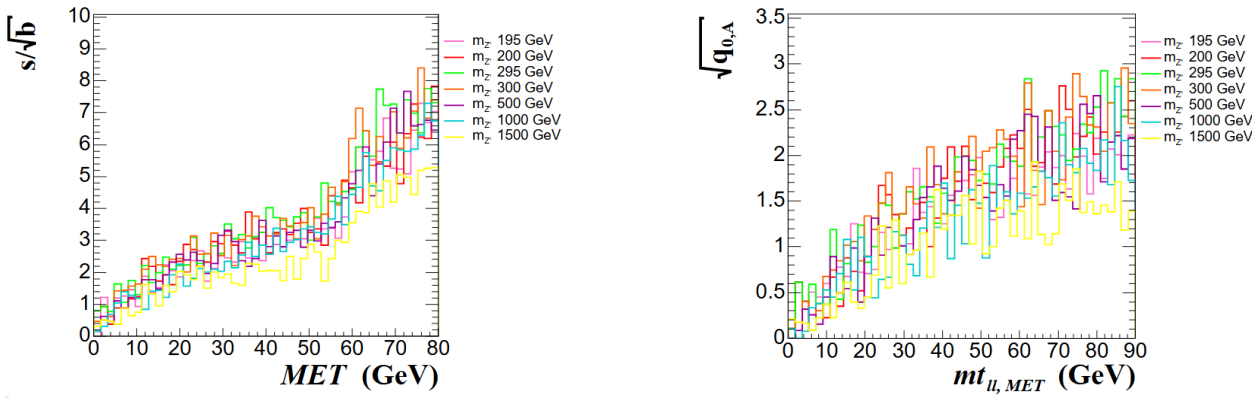


Figure 3.25: *Poisson Significance* as a function of the MET that suggest a loose selection in the $MET > 60$ GeV, in the left. While, the *Asymptotic Significance* as a function of the $mt_{ll,MET}$ in the right suggest a loose selection in $mt_{ll,MET} > 75$ GeV. This distributions correspond to the high-m_s regime.

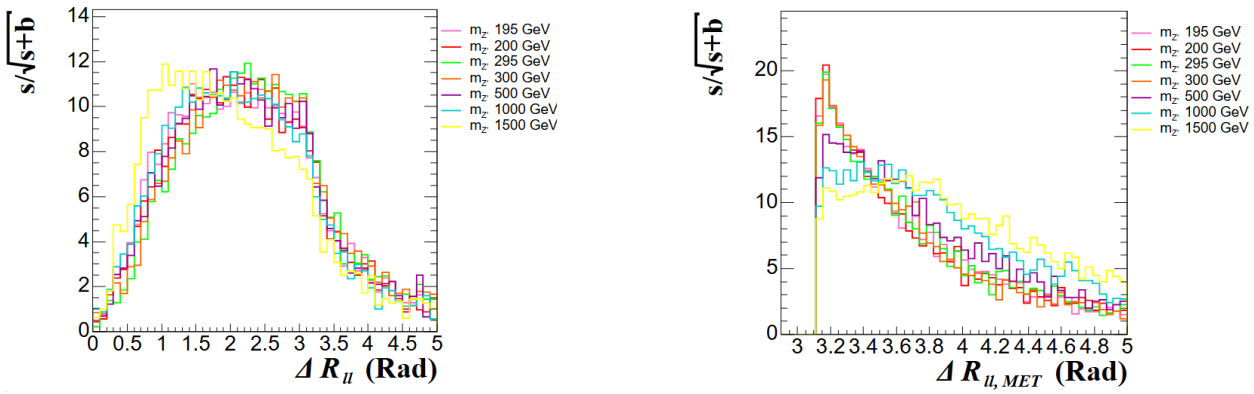


Figure 3.26: *Naive Significance* as a function of ΔR_{ll} that suggest a loose selection in $\Delta R_{ll} < 2.7$ Rad, in the left. While, the *Naive Significance* as a function of $\Delta R_{ll,MET}$ in the right suggest a loose selection in $\Delta R_{ll,MET} > 3.1$ Rad.This distributions correspond to the high-m_s regime.

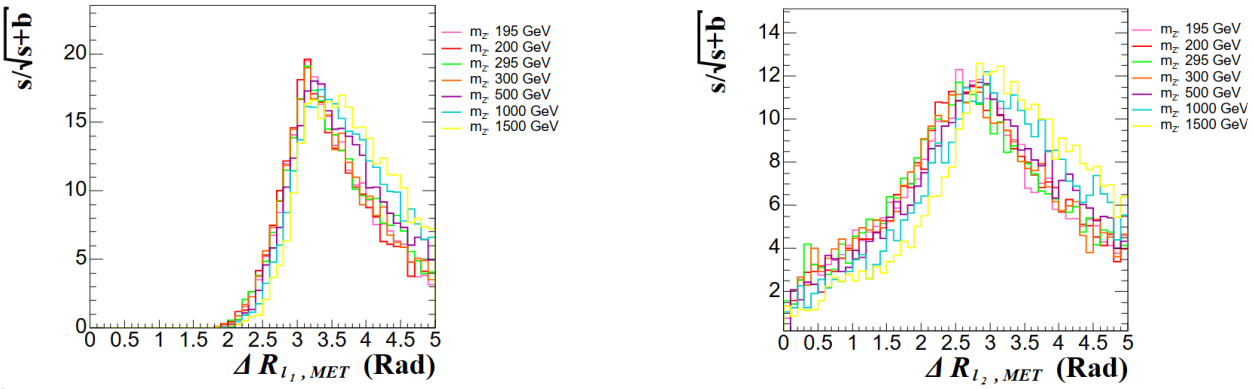


Figure 3.27: *Naive Significance* as a function of $\Delta R_{l_1,MET}$ that suggest a loose selection in $\Delta R_{l_1,MET} > 2.8$ Rad, in the left. While, the *Naive Significance* as a function of $\Delta R_{l_2,MET}$ in the right suggest a loose selection in $\Delta R_{l_2.MET} > 2.2$ Rad. This distributions correspond to the high-m_s regime.

3.5.3 Event Selection Analysis

The event selection is used to isolate the signal from the background, count events, and later perform a statistical analysis, which will be detailed in Section 3.6. For this purpose, a function was implemented in PyROOT that contains all the required selection filters. This function is applied in a loop that iterates over all entries in each TTree, retaining only those events that satisfy all the selected conditions. An example of this function is shown in the Listing 3.3.

It is also essential to account for the weight of each event. To properly scale the results, an event weight (`evtWeight`) is defined in Ec. 3.10. However, in the significance plots, in order to filter the signal from the background, the weight will not be used in favor of normalized plots, which are very useful when there is way more background than signal. On the other hand, it is important to weigh the events in order to count the number of signal and background events once the selection is finished.

$$\texttt{evtWeight} = \frac{\sigma\mathcal{L}}{N} \tag{3.10}$$

where σ is the cross section of each simulation, N is the number of generated events (in this case, 10000), and $\mathcal{L}$ represents the luminosity, which, replicating Run3 conditions, is set to 196400 pb^{-1} [29].

Listing 3.3: Definition of an event selection PyROOT function

```
def event_selection(entry):
    return(entry.ptl1 > 40 and entry.ptl2 > 30 and entry.mll > 29 and entry.mll < 48
           and entry.met > 50 and entry.ptll > 35
           and entry.mtllmet > 40 and entry.mtl1met > 30 and entry.dRl1l2 < 1.5
          and entry.dRllmet > 3.1 and entry.dRl1met > 3 and entry.dRl2met > 2.5)
```

So, with the selection of the previus subsections, it is possible to represent all the event selection of the defined variables, which is shown in Figure 3.2.

Variables	low-m_s regime	high-m_s regime
pt_{l_1} / GeV	> 25	> 45
pt_{l_2} / GeV	> 12	> 10
pt_{ll} / GeV	> 20	> 35
MET / GeV	> 20	> 60
m_{ll} / GeV	[20, 48]	> 110
$mt_{ll,MET}$ / GeV	> 40	> 75
$mt_{l_1,MET}$ / GeV	> 30	n/a
ΔR_{ll} / Rad	< 0.9	[1, 2.7]
$\Delta R_{ll,MET}$ / Rad	> 3	> 3.1
$\Delta R_{l_1,MET}$ / Rad	> 2.8	> 2.8
$\Delta R_{l_2,MET}$ / Rad	> 2.8	> 2.2

Table 3.2: Event selection of each defined variables for both regimes.

After applying the event selection, it can be used on the background and signal distributions, resulting in Tables 3.3 and 3.4.. The efficiency (ϵ) for each signal is defined in Eq. 3.11, where N_{events} is the number of signal events after the selection and N_T is the total number of generated signal events before the selection.

$$\epsilon = \frac{N_{events}}{N_T} \tag{3.11}$$

Also, the relative error for the number of events of the signal, is defined as $\sqrt{N_{events}}/N$. Where N_{events} is the number of events of each signal, and N is total number of events of the signal and background.

Additionally, the efficiencies for each $m_{Z'}$ in both regimes can be plotted, as shown in Figure 3.28.

$m_{Z'}$ / GeV	N	N_{events}	$\sqrt{N_{events}}/N$	Efficiency
195	173.9	43.1	0.04	0.247
200	272.3	64.6	0.03	0.237
295	9739.2	2188.4	0.004	0.224
300	9920.4	2283.7	0.004	0.225
500	8925.9	2193.1	0.005	0.245
1000	3264.2	828.8	0.009	0.254
1500	938.5	261.3	0.02	0.278

Table 3.3: Effect of the event selection on the number of events for the low-m_s regime, with N as the number of events before the event selection, and N_{events} as the number of events after the event selection. The number of background events after the selection have been: $N_{W^+W^-} = 6638.9$, $N_{h\to W^+W^-} = 532.1$ and $N_{t\bar{t}} = 217.5$. The number of background events before the selection was 402049, so a 98.16% reduction was made.

$m_{Z'}$ / GeV	N	N_{events}	$\sqrt{N_{events}}/N$	Efficiency
195	0.2	0	0	0
200	0.1	0	0	0
295	0.7	0.2	0.6	0.28
300	0.8	0.2	0.6	0.25
500	1.8	0.5	0.4	0.27
1000	78.1	24.1	0.06	0.301
1500	74.3	24.5	0.07	0.326

Table 3.4: Effect of the event selection on the number of events for the high-m_s regime, with N as the number of events before the event selection, and N_{events} as the number of events after the event selection. The number of background events after the selection have been: $N_{W^+W^-} = 2806$, $N_{h\to W^+W^-} = 55.3$ and $N_{t\bar{t}} = 169.2$. The number of background events before the selection was 402049, so a 99.25% reduction was made.

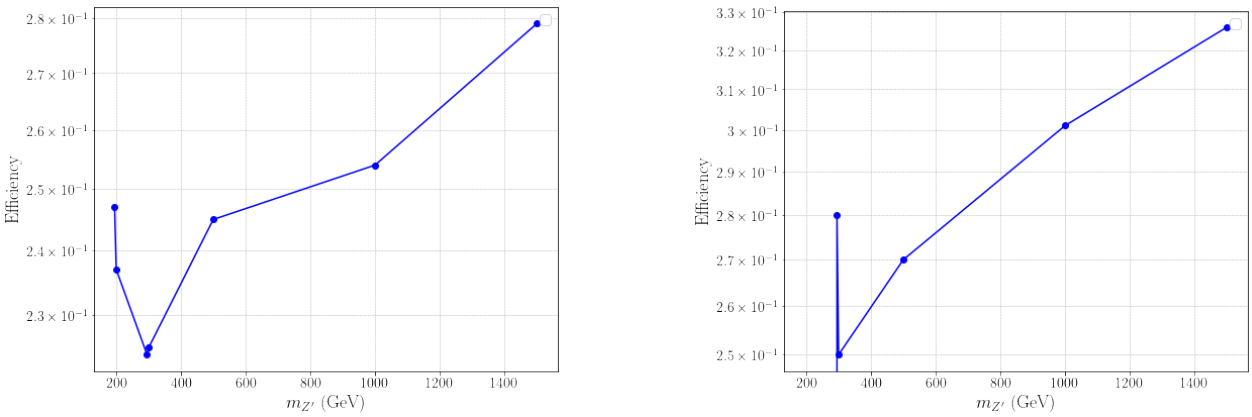


Figure 3.28: Efficiency for each $m_{Z'}$ in low-m_s regime on the left and high-m_s regime on the right

It has been observed, as expected, that the processes with the highest cross sections from Table 3.1 have resulted in a significantly larger number of events. Furthermore, it has been seen in both cases that the efficiencies increase with $m_{Z'}$. Also, it should be noted that the cuts applied were loose and not optimized for specific masses.

On the other hand, as previously observed in Table 3.1, the processes with $m_s = 300$ GeV have such a small cross section that in many cases almost no events are observed. For this reason, the statistical analysis for this regime will only be performed for the processes with the highest cross sections, $m_{Z'} = 1000$ GeV and $m_{Z'} = 1500$ GeV.

Finally, the distributions of the variables that best differentiate the signal from the background can be shown in Figures 3.29 and 3.31, which all come from the low-m_s regime, due to the small cross section of the high-m_s one. As observed, especially in the tails, after performing the event selection, processes from the model could be seen. Despite this, a detailed analysis of the cross section of the processes is necessary, which will be done in Section 3.6.

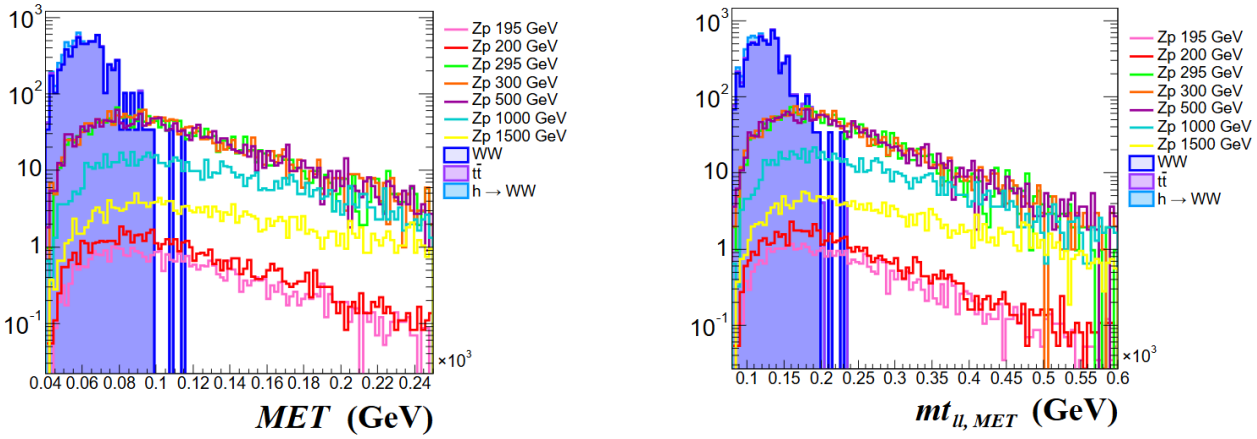
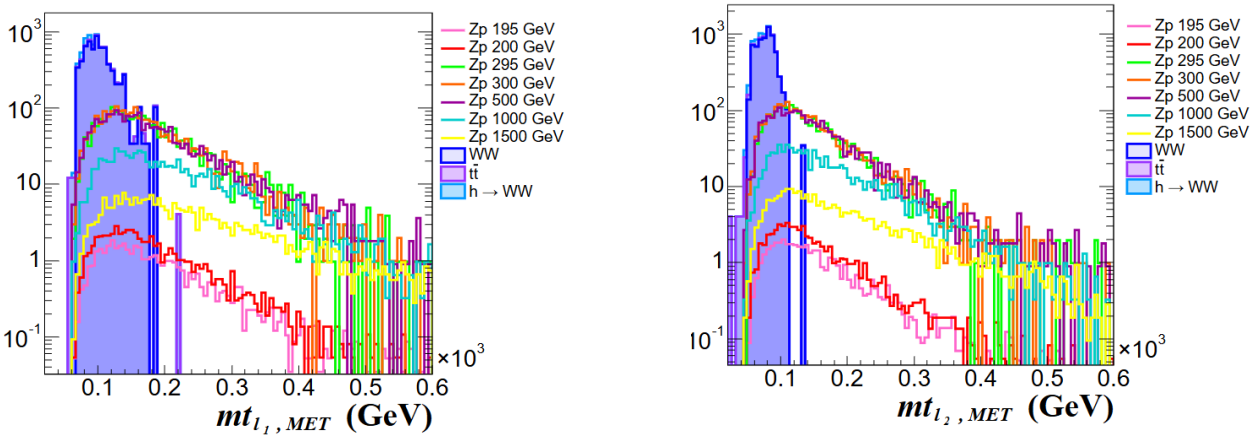


Figure 3.29: Distribution weighted by `EvtWeight` of the number of events as a function of the MET and $mt_{ll,MET}$, after applying the full event selection in the low-m_s regime. The three main background processes are also shown.



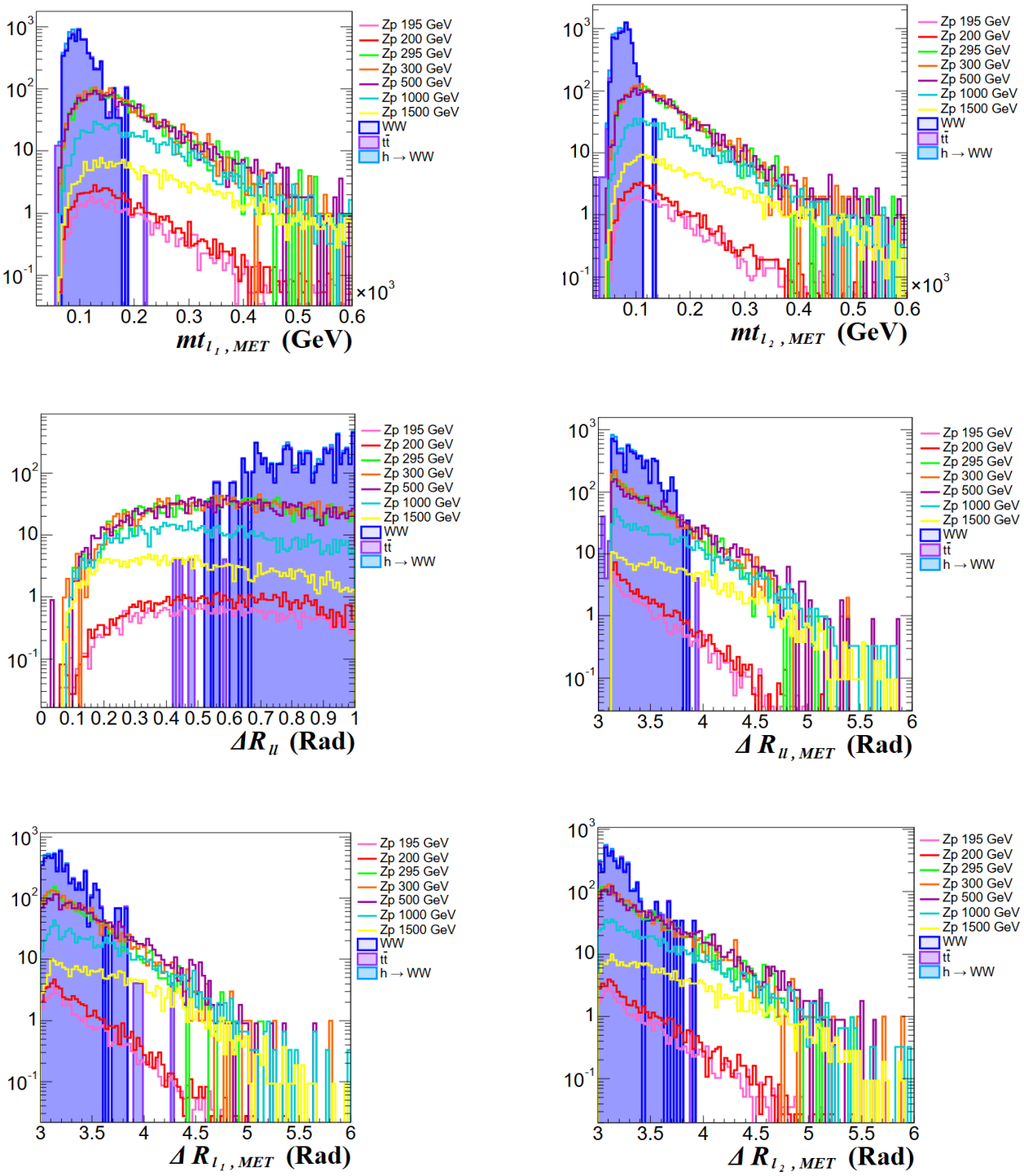


Figure 3.31: Distribution weighted by `EvtWeight` of the number of events as a function of $mt_{l_1,MET}$, $mt_{l_2,MET}$, ΔR_{ll}, $\Delta R_{ll,MET}$, $\Delta R_{l_1,MET}$ and $\Delta R_{l_2,MET}$, after applying the full event selection in the low-m_s regime. The three main background processes are also shown.

3.6 Statistical Analysis

Collisions in particle accelerators are quantum, random, and independent processes. The number of events observed in a particle physics experiment or generation follows a Poisson distribution. This is because the events are discrete, no two events can occur at the same exact time, and the expected value (μ) is constant and is defined as $\mu = \sigma \cdot \mathcal{L}$. The probability of observing exactly n events, given an expected value μ, is defined in Eq. 3.12.

$$P(n;\mu) = \frac{\mu^n e^{-\mu}}{n!} \tag{3.12}$$

Given that the distribution of the number of events follows a Poisson distribution, it is possible to estimate *upper limits* on the cross section (σ_{UL}). σ_{UL} represents the maximum value of the cross section for which one can claim with 95% confidence that the true cross section is below this limit. To estimate σ_{UL}, the Confidence Levels (CL_s) method is used.

This method compares two hypotheses: the background only hypothesis (b), and the signal plus background hypothesis ($s + b$). Also, the number of observed events (N_{obs}) is considered, and the goal is to determine whether this observation could be a background fluctuation or indicates the presence of signal. Since this work does not use real data, N_{obs} will be taken as the sum of the generated signal and background events.

Furthermore, two key probabilities are defined in this analysis. p_b is the probability of observing N_{obs} or more events assuming only background is present. It is given in Eq. 3.13.

$$p_b = P(N \geq N_{obs}|b) = \sum_{n=N_{obs}}^{\infty} P(n|b) \tag{3.13}$$

This is known as the Poisson tail, and it can be estimated using a Python script with the command: `p_b = stats.poisson.sf(N_obs - 1, b)`.

Similarly, p_{s+b} is defined as the probability of observing N_{obs} or more events assuming both signal and background are present. It is given in Eq. 3.14, and in Python it is calculated as `p_bs = stats.poisson.sf(N_obs - 1, s+b)`.

$$p_{s+b} = P(N \geq N_{obs}|s+b) = \sum_{n=N_{obs}}^{\infty} P(n|s+b) \tag{3.14}$$

With all of this, an iterative method is implemented to determine σ_{UL} for each $m_{Z'}$. A statistical indicator (CL_s) is calculated, which must be smaller than a certain threshold α. In this case, $\alpha = 0.05$ ensures a 95% confidence level. This is shown in Eq. 3.15 [30].

$$CL_s \equiv \frac{p_{s+b}}{1 - p_b} < \alpha \tag{3.15}$$

Using this and the number of signal events and efficiencies estimated in the previous section in Tables 3.3 and 3.4, a range of hypothetical cross sections (σ_s) is scanned, and for each one, the expected number of signal events is calculated using Eq. 3.16. In this work, the scan is performed over the range $\sigma_s \in [0, 0.1]$.

$$s = \sigma_s \cdot \mathcal{L} \cdot \epsilon \tag{3.16}$$

Subsequently, CL_s is calculated for each value of s, and it is checked if $CL_s < \alpha$. The smallest value of σ_s that satisfies this condition is taken as the *upper limit*. To do this, CL_s is plotted as a function of σ_s. The *upper limit* corresponds to the value of σ_s where the curve intersects $CL_s = 0.05$. An example for $m_{Z'} = 300$ GeV in the low m_s regime is shown in Figure 3.32.

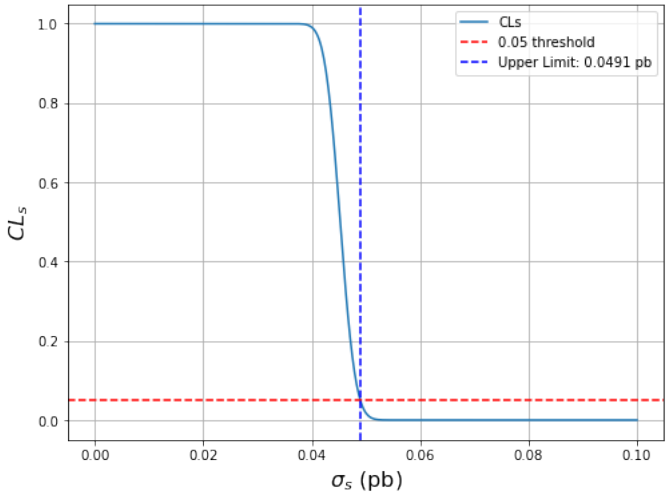


Figure 3.32: CL_s as a function of σ_s. The red line represents $CL_s = 0.05$, and the point in σ_s where CL_s intersects this line defines the *upper limit*. This example corresponds to $m_{Z'} = 300$ GeV in the low m_s regime.

Now, the *upper limits* can be computed for all signal points in both the low and high m_s regimes. As previously explained in Section 3.5, due to the lack of signal events for low $m_{Z'}$ in the high-m_s regime, only the cases when $m_{Z'} = 1000$ GeV and $m_{Z'} = 1500$ GeV are analyzed.

Additionally, *upper limits* are computed for different luminosity scenarios: corresponding to Run 3 (2022/23/24), the expected value for 2025 (Run 3 2022/23/24/25), and the projected value for the HL-LHC. The results for the low-m_s regime are shown in Table 3.5, and those for the high-m_s regime in Table 3.6.

$m_{Z'}$ / GeV	σ^1_{UL} / pb	σ^2_{UL} / pb	σ^3_{UL}
195	0.0041	0.0027	0.0003
200	0.0047	0.0031	0.0003
295	0.0535	0.0351	0.0036
300	0.0555	0.0363	0.0038
500	0.0491	0.0321	0.0032
1000	0.0196	0.013	0.0014
1500	0.0076	0.005	0.0006

Table 3.5: Table of the *upper limit* of the cross section as a function of $m_{Z'}$, for $m_\chi = 100$ GeV and $m_s = 160$ GeV. σ^1_{UL} corresponds to the total cross section limit considering the Run3 luminosity from the years 2022/23/24, with $\mathcal{L} = 196400$ pb^{-1}. σ^2_{UL} corresponds to the expected value in 2025 with $\mathcal{L} \approx 3 \cdot 10^6$ pb^{-1} [31], and σ^3_{UL} corresponds to the expected value at the HL-LHC, with $\mathcal{L} \approx 3 \cdot 10^7$ pb^{-1} [7].

$m_{Z'}$ / GeV	σ^1_{UL} / pb	σ^2_{UL} / pb	σ^3_{UL}
1000	0.0022	0.0014	0.0002
1500	0.0020	0.0013	0.0001

Table 3.6: Table of the *upper limit* of the cross section as a function of $m_{Z'}$, for $m_\chi = 100$ GeV and $m_s = 300$ GeV. σ^1_{UL} corresponds to the total cross section limit considering the Run3 luminosity from the years 2022/23/24, with $\mathcal{L} = 196400$ pb^{-1} [29]. σ^2_{UL} corresponds to the expected value in 2025 with $\mathcal{L} \approx 3 \cdot 10^6$ pb^{-1} [31], and σ^3_{UL} corresponds to the expected value at the HL-LHC, with $\mathcal{L} \approx 3 \cdot 10^7$ pb^{-1} [7].

With all of this, the *upper limit* on the cross section as a function of $m_{Z'}$ can be represented for the low-m_s regime in Figure 3.33.

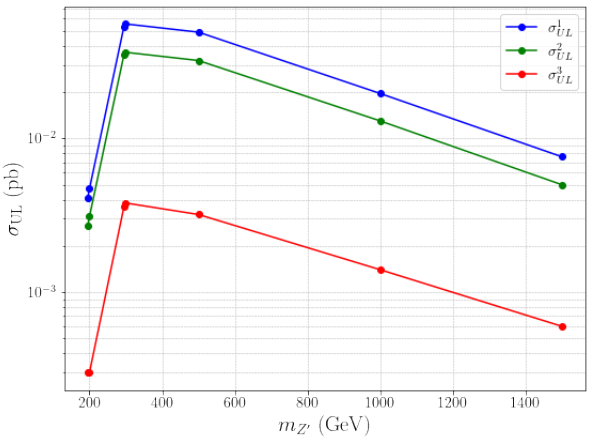


Figure 3.33: Representation of the *upper limit* on the cross section as a function of $m_{Z'}$, for $m_\chi = 100$ GeV and $m_s = 160$ GeV. σ^1_{UL} corresponds to the cross section limit obtained using Run 3 luminosity from the years 2022/23/24, with $\mathcal{L} = 196400$ pb^{-1} [29]. σ^2_{UL} corresponds to the projected value for 2025 with $\mathcal{L} \approx 3 \cdot 10^6$ pb^{-1}[31], while σ^3_{UL} refers to the expectation for HL-LHC, with $\mathcal{L} \approx 3 \cdot 10^7$ pb^{-1}[7]. The number of background events are: $N_{W^+W^-} = 6638.9$, $N_{h\to W^+W^-} = 532.1$, and $N_{t\bar{t}} = 217.5$.

The uncertainties in the *upper limits* have not been considered, since when plotting the three curves with very similar values, they would overlap visually. Moreover, given that the cuts applied were loose, additional background processes should be included and a more detailed statistical analysis performed in order to obtain more precise values with their proper uncertainties.

It is important to note that the *upper limits* shown correspond to the full cross section of the studied process, as expressed in Eq. 3.17, where σ_P is the production cross section, where 4 $BR(WW \to l\nu_l)^2$ accounts for the two lepton flavors.

$$\sigma_{UL} = \sigma_P \times BR(s \to W^+W^-) \times 4\ BR(WW \to l\nu_l)^2 \tag{3.17}$$

It is important to observe that the obtained *upper limits* are extremely small when compared to SM cross sections. This was expected, since Beyond-Standar-Model (BSM) processes are rare, and if their cross sections were of the same order as common SM processes, they would have been observed

at the LHC, if they actually exist.

To see how small the estimated *upper limits* are, it is worth consider that the smallest production cross section ever measured by CMS is $\sigma(\text{EW } qqZZ) = 330$ pb [32], corresponding to the electroweak component of double Z production from two quarks.

In order to be able to estimate a qualitative value of the production cross section *upper limit*, Eq. 3.17 can be used to isolate σ_P, using the highest estimated value at the Run 3 luminisity of σ_{UL}, which is $\sigma_{UL} = 0.0555$ pb. Also, approximating $BR(s \to W^+W^-)$ to $BR(s \to W^+W^-) \approx 0.65$, which is the highest value observed in Figure 3.1. And lastly, 4 $BR(WW \to l\nu_l)^2 \approx (2 \cdot 0.11)^2$ [22]. Using these, an upper bound for the production *upper limit* ($\sigma_P^{\max}$) is estimated in Eq.3.18.

$$\sigma_P^{\max} \approx \frac{0.055 \text{ pb}}{0.65 \cdot (2 \cdot 0.11)^2} \approx 3.5 \text{ pb} \tag{3.18}$$

Given that the smallest production cross section ever observed is $\sigma(\text{EW } qqZZ) = 330$ pb. It can be concluded that the upper bound for the production *upper limit* of the process analyzed is nearly one hundred times smaller than the lowest value ever measured by CMS.

4 Conclusions

This work presents a study of a dark matter production model based on a dark Higgs mechanism, which introduces a dark Higgs boson, a dark sector mediator Z', and a fermionic dark matter particle (χ). Using Monte Carlo generations, the main decay channels of the dark Higgs have been analyzed, identifying the most favorable one as $s \to W^+W^- \to l^+\nu_l, l^-\bar{\nu}_l$. For this, BR have been estimated and each decay mode studied in detail.

In addition, various mass ranges have been explored, and several variables have been defined to differentiate between the masses of the new dark sector particles. This has allowed a detailed analysis of the model's phenomenology and the influence of the mass spectrum on physical observables, identifying two regimes depending on the dark Higgs mass: low and high.

The background has been studied through Monte Carlo generation of Standard Model events with similar final states. After applying selection cuts based on statistical significance, the efficiency of each signal has been estimated, except for the signals with $m_{Z'} < 500$ GeV in the low-m_s regime, due to their suppressed cross section. For the remaining signals, the CL_s method was used to estimate 95% confidence level *upper limits* on the cross section for each value of $m_{Z'}$ in both regimes, as shown in Figures 3.5 and 3.6.

It was found that the upper bound on the production cross section of the process is approximately two orders of magnitude below the smallest cross section observed by CMS. This suggests that current sensitivity is still far from being able to probe the model exmperimentaly. However, future particle colliders such as the HL-LHC may reach the necessary precision to test such processes.

To complete the results of this study, additional background processes could be considered. This would allow the application of less-loose selection cuts, improve the precision of the *upper limit* estimates, and enable a more rigorous treatment of uncertainties. Also, the variation of the three couplings are free parameters of the model could be analyzed in a similar way as the variation of the masses.

Alternative decay channels of the dark Higgs could also be studied. The most promising one is the

$b\bar{b}$ channel, due to its high cross section for $m_s < 160$ GeV. This would require incorporating b-jets into the generation, either within MadGraph or through another event generator. In this case, the MET originates exclusively from dark matter particles, meaning neutrinos would not distort the variable.

The ZZ channel, despite its theoretical interest, since it produces final states without neutrinos or b-jets, has an even smaller cross section than W^+W^- when $m_s = 300$ GeV. Given the statistical challenges faced when analyzing the high-m_s regime, this suggests that the ZZ channel is likely inaccessible without an improved sensitivity.

Other decay modes beyond those explored in Section 3.1 could also be considered. For example, the decay $s \to \chi\chi$, already mentioned in Section 3.2, deserves further investigation. Also, one could evaluate the possibility of introducing a direct coupling between Z' and leptons, which would open new phenomenological channels. However, this interaction is not included in the model and no similar implementation was found, meaning it would have to be added manually in FeynRules, which is a non trivial task.

Finally, it would be interesting to apply a similar analysis to other dark matter production models, or even to real data from the CMS experiment, incorporating signal and background modeling with more advanced statistical tools for the search of dark matter.

Beyond the results obtained, this project has represented an important learning experience and has allowed me to develop a wide range of skills. These include the use of simulation tools like MadGraph, ROOT, and PyROOT; working with professional environments such as SWAN and *lxplus* under Linux; and programming in Python and C++ for big data analysis.

I, also, have improved my scientific writing, bibliografy management, critical analysis, and understanding of statistical tools such as the CL_s method and significance estimators. Furthermore, I have deepened my knowledge of the Standard Model, the Higgs mechanism (essential to understand the dark Higgs mechanism), and the theory behind the dark Higgs model, which I have explored in depth.

This project has allowed me to get a first look at the kind of analysis carried out by researchers in particle physics, and not only has it been an enriching experience, it has also reaffirmed my motivation to continue training in this field.

Referencias

[1] Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, et al. Planck 2018 results. VI. Cosmological Parameters. *Astronomy & Astrophysics*, 641:A6, 2020.

[2] Brown University Center for Particle Astrophysics. Dark matter — brown university. `https://particleastro.brown.edu/dark-matter/`. Accessed: 2025-06-03.

[3] E. Aprile and others [XENON Collaboration]. Constraining the spin-dependent WIMP-nucleon cross sections with XENONnT. *Phys. Rev. Lett.*, 131:041003, 2023.

[4] SuperCDMS Collaboration. Projected sensitivity of the supercdms snolab experiment. *Phys. Rev. D*, 102(5):052002, 2020.

[5] R. Bernabei et al. First model independent results from DAMA/LIBRA–phase2. *Nucl. Phys. At. Energy*, 19(4):307–325, 2018.

[6] W. B. Atwood and others [Fermi-LAT Collaboration]. The large area telescope on the fermi gamma-ray space telescope mission. *The Astrophysical Journal*, 697(2):1071–1102, 2009.

[7] G. Apollinari, I. Béjar Alonso, O. Brüning, M. Lamont, and L. Rossi. High-luminosity large hadron collider (hl-lhc): Technical design report v. 0.1. Technical Report CERN-2017-007-M, CERN, Geneva, 2017. CERN Yellow Reports: Monographs.

[8] Michael Duerr, Felix Kahlhoefer, Bjoern Penning, Kai Schmidt-Hoberg, Alexander Grohsjean, and Christian Schwanenberger. Hunting the Dark Higgs. *Journal of High Energy Physics*, 2017(7):1–26, 2017.

[9] IBM. ¿qué es la simulación de montecarlo? `https://www.ibm.com/es-es/topics/monte-carlo-simulation`. Accessed: 2025-06-13.

[10] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014.

[11] FeynRules Collaboration. Feynrules – a mathematica package for computing feynman rules. `https://feynrules.irmp.ucl.ac.be/`. Accessed: 2025-06-13.

[12] CERN Document Service. Lxplus documentation. `https://lxplusdoc.web.cern.ch/`. Accessed: 2025-06-13.

[13] Daniel Abercrombie, Nural Akchurin, Ece Akilli, Juan Alcaraz Maestre, Brandon Allen, Barbara Alvarez Gonzalez, Jeremy Andrea, Alexandre Arbey, Georges Azuelos, Patrizia Azzi, and others. Dark matter benchmark models for early lhc run-2 searches: Report of the atlas/cms dark matter forum. *Physics of the Dark Universe*, 27, 2020.

[14] Instituto de Física de Cantabria (IFCA). Hoy comienza run 3 del lhc y física a energías récord. `https://ifca.unican.es/es-es/news/Paginas/Hoy-comienza-Run-3-LHC-y-fisica-energias-record.aspx`. Accessed: 2025-06-13.

[15] A. Peters, J. Blomer, M. Bessone, C. Aguado-Sánchez, G. Govi, and E. Cano. EOS – CERN's disk storage system. *Journal of Physics: Conference Series*, 898(6):062032, 2017.

[16] Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, Brian Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica Hamrick, Jason Grout, Sylvain Corlay, Paul Ivanov, Damián Avila, Safia Abdalla, and Carol Willing. Jupyter notebooks – a publishing format for reproducible computational workflows, 2016. In F. Loizides and B. Schmidt (Eds.), *Positioning and Power in Academic Publishing: Players, Agents and Agendas*. IOS Press.

[17] ROOT Team. Root - data analysis framework. `https://root.cern/`. Accessed: 2025-06-13.

[18] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, et al. Array programming with numpy. *Nature*, 585(7825):357–362, 2020.

[19] Jim Pivarski, Eduardo Rodrigues, et al. Uproot: Scikit-hep's root i/o in pure python and numpy. *Comput. Softw. Big Sci.*, 3:1–9, 2019.

[20] Python Software Foundation. Python language reference, version 3.11. `https://www.python.org/`. Accessed: 2025-06-13.

[21] Unai Pérez Uribe. Study of dark matter production models in the cms experiment at the lhc. `https://github.com/UnaiPerezUribe/Study-of-Dark-Matter-Production-Models-in-the-CMS-Experiment-at-the-LHC`, 2025. GitHub repository.

[22] R.L. Workman et al. Review of Particle Physics. *Prog. Theor. Exp. Phys.*, 2022(083C01), 2022.

[23] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to Quantum Field Theory*. Addison-Wesley Publishing Company (now Westview Press), Reading, MA, 1995.

[24] Torben Ferber, Alexander Grohsjean, and Felix Kahlhoefer. Dark higgs bosons at colliders. *Progress in Particle and Nuclear Physics*, 136:104105, 2024.

[25] TikZ.net. Cms 3d axis in tikz. `https://tikz.net/axis3d_cms/`. Accessed: 2025-06-13.

[26] Pedro José Fernández Manteca. *Production of Two W^+W^- Bosons in Proton-Proton Collisions at a Center-of-Mass Energy of $\sqrt{s} = 13$ TeV with the CMS Detector at the LHC*. PhD thesis, University of Cantabria, Santander, 2021. Supervised by Alicia Calderón Tazón and Rocío Vilar Cortabitarte.

[27] Lucio Lista. Practical statistics for particle physicists. In M. Mulders and G. Zanderighi, editors, *CERN Yellow Reports: School Proceedings*, volume 5 of *Skeikampen 2016*, pages 213–285. CERN, Geneva, 2017.

[28] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 71:1554, 2011.

[29] CMS Collaboration. Integrated luminosity for pp collisions at different energies. `https://cmslumi.web.cern.ch/publicplots/multiYear/int_lumi_cumulative_pp_2.png`. Accessed: 2025-06-13.

[30] ATLAS Statistics Forum. The CLs method: information for conference speakers. Technical report, ATLAS Collaboration, July 2011. Internal Note, 5 July, 2011.

[31] CERN. And they're off! the 2025 lhc physics season gets underway, 2025. Accessed: 2025-06-05.

[32] CMS Collaboration. Summary of CMS Cross Section Measurements. `https://twiki.cern.ch/twiki/pub/CMSPublic/PhysicsResultsCombined/CMSCrossSectionSummaryBarChart.pdf`, 2024. Accessed: 2025-06-13.