Final Report

Search for Z'-pair production decaying into Dark Matter and boosted Jets at CMS

Claudio Andrea Manzari

University of Bari

claudioandrea.manzari@gmail.com

Fermilab Summer School

07-31-17/09-29-17

Supervisor: Matteo Cremonesi

Abstract

This study aims to develop a search for dark matter in a signature of missing transverse energy and jets in the CMS detector at the CERN LHC. I generated and validated a Z'-pair production model and I set up the analysis strategy for the background estimation, defining control regions and preparing the main background extraction from these.

1. Introduction

The existence of Dark Matter would imply strong evidence for physics beyond standard model (SM) [1, 2]. Cosmological observations demonstrate that around the 85% of the mass of the Universe comprised of DM, however there is no experimental evidence of its non-gravitational interactions with SM particles. These observations along with further constraints make it highly likely that DM is composed primarily of weakly interacting massive particles (WIMPs). If non-gravitational interactions exist between DM and SM particles, DM could be produced by colliding SM particles at high energy. In many theories, the pair production of DM particles in hadron collisions proceeds through a spin-0 or spin-1 bosonic mediator [3], with the DM particles leaving the detector without a measurable signature. One way to observe them is when they recoil with large transverse momentum (p_T)

against additional jets radiated from the initial state, resulting in an overall transverse momentum imbalance in the collision event. Such jets from the initial-state radiation (ISR) can then be used to tag the interesting events. These are called "mono-jet" events. Figure 1 shows the mono-jet production diagram assuming a spin-1 mediator.

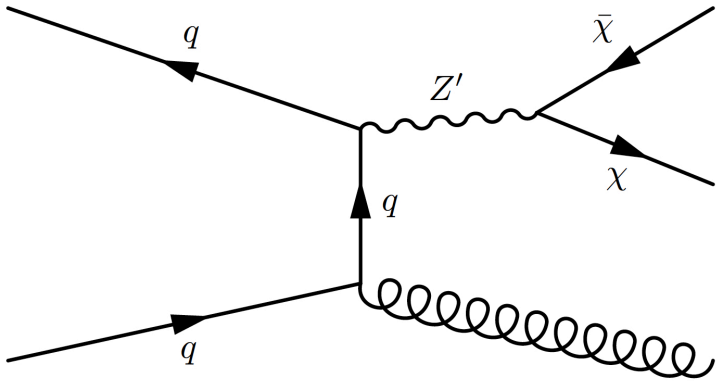

Figure 1: Mono-jet production diagram with a spin-1 mediator.

With the same mechanism, the associated production of DM and a single photon (mono-photon) or a vector boson (mono-V , V =W,Z) can occur. If the spin-1 mediator has flavor-changing couplings to top and light quarks, a single top quark can be produced in association with DM particles (mono-top) [4]. Recent phenomenological studies [5] are proposing a class of DM models with a GeV-scale dark Z'. At hadron colliders, the production of DM particles naturally leads to associated production with a Z', which can appear as a large-cone jet after it decays hadronically. Contrary to the usual mono-jet signal from ISR, a Z'-pair production can generate the signature of a visible Z'-jet plus missing transverse energy, arising from the second Z' decaying to DM. Performing a jet-substructure analysis to tag the visible Z' jet, a dedicated search for mono-Z' events can lead to over an order of magnitude stronger bounds on the interpreted DM-nucleon scattering cross sections when compared to mono-jet. My first task was the generation of the mono-Z' process using Madgraph, a software package that provides a matrix element generator for parton-level processes giving a specific Lagrangian for the theoretical model. Mono-Z' events are generated in LHE format. Their

kinematic properties are analyzed and validated before starting a massive production of files in the most suitable format for the search in CMS data. My final task was to set up the analysis strategy, studying the main background for this process.

2. The CMS detector

The CMS detector, described in detail in Ref. [6], is a multipurpose apparatus designed to study a wide gamut of physics processes in pp and heavy-ion collisions. A superconducting solenoid occupies its central region, providing a magnetic field of 3.8 T parallel to the beam direction. Charged-particle trajectories are measured by the silicon pixel and strip trackers, which cover a pseudorapidity region of $|\eta| < 2.5$. A lead tungstate ($PbWO_4$) crystal electromagnetic calorimeter (ECAL), and a brass and scintillator hadron calorimeter (HCAL) surround the tracking volume and cover $|\eta| < 3$. The steel and quartz-fiber Ĉerenkov hadron forward (HF) calorimeter extends the coverage to $|\eta| < 5$. The muon system consists of gas-ionization detectors embedded in the steel flux-return yoke of the solenoid, and covers $|\eta| < 2.4$. The first level of the CMS trigger system is designed to select events in less than 4 μs, using information from the calorimeters and muon detectors. The high-level trigger processor farm then further reduces the event rate to several hundred Hz that are recorded for further analysis.

The particle-flow (PF) event algorithm [7, 8] reconstructs and identifies each individual particle with an optimized combination of information from the various elements of the CMS detector. The energy of photons is directly obtained from the ECAL measurement, corrected for zero-suppression effects. The energy of electrons is determined from a combination of the electron momentum at the primary interaction vertex as determined by the tracker, the energy of the corresponding ECAL cluster, and the energy sum of all bremsstrahlung photons spatially compatible with originating from the electron track. The energy of muons is obtained from the curvature of the corresponding track. The energy of charged hadrons is determined from a combination of their momentum measured in the tracker and the matching ECAL and HCAL energy deposits, corrected for zero-suppression effects and for the response function of the calorimeters to hadronic showers. Finally, the energy of neutral hadrons is obtained from the corresponding corrected ECAL and HCAL energy.

The missing transverse momentum vector ($\vec{p}_T^{miss}$) is computed as the negative vector sum of the transverse momenta ($\vec{p}_T$) of all the PF candidates in an event, and its magnitude is denoted as E_T^{miss}. Jets are reconstructed by clustering PF candidates using the anti-k_t algorithm [9]. Jets clustered with distance parameters of 0.4 and 0.8 are referred to as AK4 and AK8 jets, respectively. The primary vertex with the largest sum of p_T^2 of the associated tracks is chosen as the vertex corresponding to the hard interaction in an event. All charged PF candidates originating from any other vertex are ignored during the jet reconstruction. Jet momentum is determined as the vectorial sum of all particle momenta in the jet, and is found from simulation to be within 5% to 10% of the true momentum over the whole p_T spectrum and detector acceptance. An offset correction is applied to jet energies to take into account the contribution from additional proton-proton interactions within the same or nearby bunch crossings (pileup). Jet energy corrections are derived from simulation, and are confirmed with in situ measurements of the energy balance in dijet, multijet, γ+jet, and leptonic Z+jet events [10]. These are also propagated to the E_T^{miss} calculation [11].

3. Model Generation

In order to design a dedicated mono-Z' search and to study its sensitivity to the associated production of dark matter with a Z' boson, signal samples are generated using a multipurpose Monte Carlo generator called Madgraph [12]. The decay of the mediator in SM particles has been generated with Madspin [13]. The mono-Z' model is obtained following the prescription provided by the LHC Dark Matter Working Group [3], which devised a set of recommendations for the generation of simplified models for dark matter production at the LHC. The Madgraph study consists of generating signal events and scanning the relevant parameters, like the couplings to SM and DM particles that in the case of a mono-Z' model are assumed to be vector and axial-vector. Following the official prescription, the chosen value for the coupling to the SM (g_{sm}) is 0.25, while the coupling to DM particles (g_{DM}) is chosen as 1. The Z' particle is generated assuming minimal decay width, making the boson coupling only with DM particles and SM quarks. It should be noted that if there were other particles coupling to the Z', the width would increase reducing the overall sensitivity of the analysis [14]. The

mono-Z' signal samples have been generated with different hypotheses for the mass of the boson and for the mass of the Dark Matter particle. The mass grid used is reported in Table 1.

M_χ (GeV)	M_ϕ (GeV)										
1	10	100	300	500	750	1000	1500	1750	2000	2250	2500
10								1750	2000	2250	
50		100									
100			300					1750	2000	2250	
150				500				1750	2000	2250	
200		100		500				1750	2000	2250	
300			300	500	750	1000	1500	1750	2000	2250	2500
400			300			1000		1750	2000	2250	
500				500				1750	2000	2250	
600					750	1000	1500				2500
700											2500

Table 1: This Table reports the grid of DM and Mediator masses for model generation.

The transverse momentum of the Z' particle is expected to vary with the mass. In order to check our expectations and validate the model, Les Houches Event (LHE) files are generated and later converted in ROOT [15] files, allowing a visual check by simply plotting the relevant kinematic variables. The validation of the model has been done on three mass benchmarks reflecting three different possible physical scenarios:

	M_ϕ (GeV)	M_χ (GeV)
Resonant Case	300	100
Heavy Mediator	1000	1
Heavy DM	300	400

Table 2: Mass benchmarks used for model validation.

Figures 2, 3, 4 and 5, show the agreement between the mass distribution and the transverse momentum distribution of the generated Z' and the Z' reconstructed as the sum of the two DM particles from its decay.

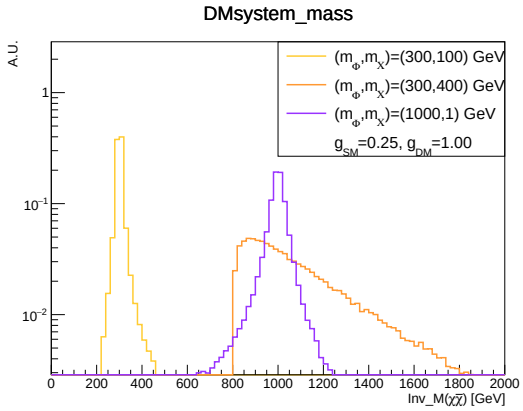

Figure 2: Invariant mass distribution of the DM system.

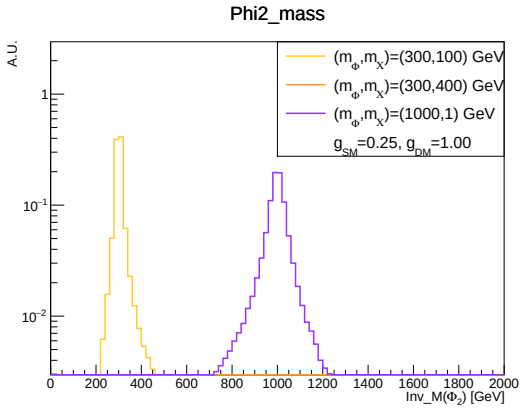

Figure 3: Mass distribution of the Z' boson decaying in DM.

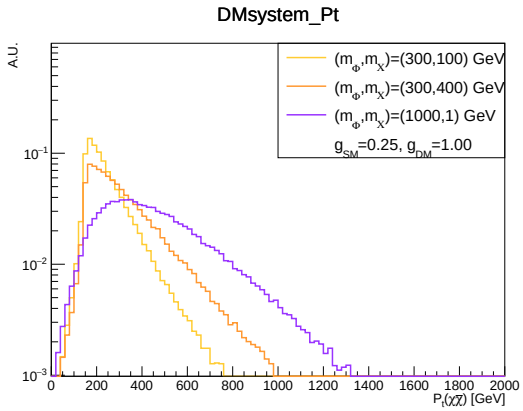

Figure 4: Transverse momentum distribution of the DM system.

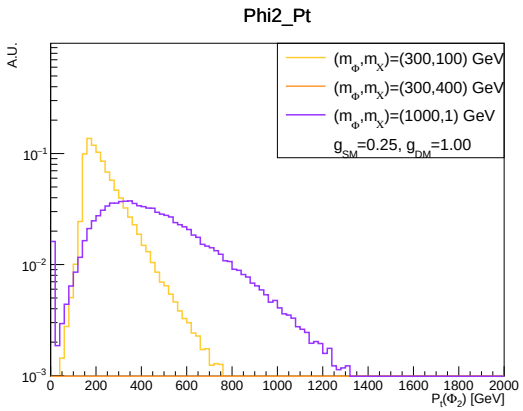

Figure 5: Transverse momentum distribution of the Z' boson decaying in DM.

Note that the orange distribution in Figure 2 does not show any mass peak since it corresponds to a model where the two dark matter particles are produced off-shell. The distributions showed are relative to the Vector case. The distributions obtained in the Axial case show the same features, and the same considerations are valid for them. To make the new generated samples usable in a real physics analysis, the LHE files have been hadronized with Pythia 8 [16] and are then propagated through a full CMS detector simulation using the CMSSW framework [17]. At the hadronization step, the

simulation has been validated again checking the distributions of the relevant kinematic variables.

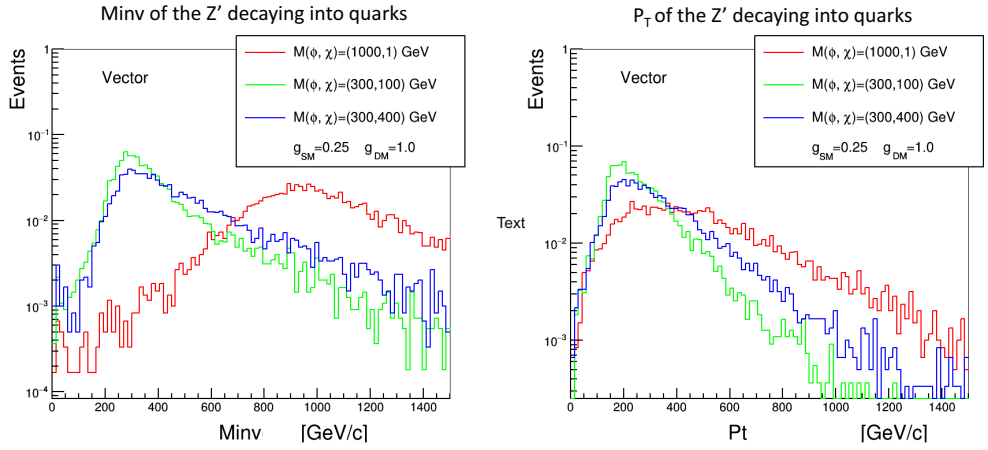

Figure 6: Invariant mass and transverse momentum distributions of the Z' decaying into quarks.

In Figure 6 are shown the invariant mass and transverse momentum distributions of the Z' decaying into quarks at the hadronization level. The invariant mass distributions peak at the expected boson masses. The transverse momentum distributions have the expected behaviour and show the momentum cut imposed at the generation level (to be larger than 150 GeV).

4. Event selection and simulated samples

Mono-Z' events are characterized by a large imbalance in the transverse momentum. In order to identify them, events are selected online requiring triggers with thresholds of 90, 100, 110, or 120 GeV on $E^{miss}_{T,trig}$. The $E^{miss}_{T,trig}$ corresponds to the magnitude of the vectorial sum of the p_T of all the candidates reconstructed at the trigger level. The trigger efficiency measured to be 97% for events passing the analysis selection for E^{miss}_T around 250 GeV and becomes fully efficient for events with $E^{miss}_T > 350$ GeV.

Referring to the model in Figure 1 the cut on E_T^{miss} is expected also on the momentum of the object recoiling against MET. Since from phenomenological studies we are interested in a light mass region for the Z' boson, around $0-150$ GeV, it is possible to estimate the angular distribution (dR) between the objects produced in the Z' decay from the empirical relation in Equation 1.

$$dR = \frac{2 * M^{Z'}}{P_t^{Z'}} \tag{1}$$

The dR expected is between $0 - 1.5$ therefore the $dR = 1.5$ CMS jet collection is used.

An event is categorized into mono-Z' final state, if $E_T^{miss} > 250$ GeV and the leading CA15 jet [18] of $p_T > 250$ GeV and $|\eta| < 2.4$ is identified as a jet arising from hadronic decays of Lorentz-boosted Z' boson. The mass of the leading CA15 jet is computed after pruning based on the technique [19] involving reclustering the constituents of the jet using Cambridge-Aachen algorithm and removing the soft constituents in every recombination step. This technique yields improved jet mass resolution due to reduced effects coming from the underlying event and pileup.

The leading jet is also required to have at least 10% of its energy coming from charged particles and less than 80% of its energy attributed to neutral hadrons. Such selection helps to reduce events originating from beam induced backgrounds. The energy fraction attributed to neutral hadrons in these jets is required to be less than 0.9. This requirement suppresses anomalous events with jets originating from detector noise. In addition, the analysis employs various event filters to reduce events with large misreconstructed E_T^{miss} [11] originating from non collision backgrounds.

5. Background estimation

The main backgrounds to this search arise from the $Z(\nu\nu)$ + jets and $W(l\nu)$ + jets with a missed lepton processes. The $Z(\nu\nu)$+jets process is an

irreducible background and therefore constitutes the largest background in the search. The background from $W(l\nu)$ + jets is suppressed by imposing a veto on events containing one or more electrons or muons with $p_T > 10GeV$ or τ leptons with $p_T > 18$ GeV. Electron candidates are required to satisfy identification criteria based on the shower shape of its energy deposit in the ECAL and the matching of the electron track to the ECAL energy cluster. Muon candidates are required to be identified by the PF algorithm. Both electron and muon candidates are required to be isolated and must be consistent with originating from the primary vertex. The τ lepton candidates are required to pass identification criteria [45] which require a jet with an identified subset of particles whose mass is consistent with the hadronic decay products of τ leptons. Events that contain an isolated photon with $p_T > 15$ GeV that satisfies certain identification criteria based on its ECAL shower shape are also vetoed. This helps to suppress electroweak backgrounds in which a photon is radiated from the initial state. To reduce contamination from top quark backgrounds, events are rejected if they contain a b jet with $p_T > 20$ GeV identified using the loose working point of the combined secondary vertex algorithm, non overlapping with the Z'-candidate CA15 jet. [20, 21].

The largest background contributions from $Z(\nu\nu)$ + jets and $W(l\nu)$ + jets processes are estimated using data from mutually exclusive control samples composed by dimuon, dielectron, single-muon, single-electron events. The hadronic recoil p_T is used as a proxy for E^T_{miss} in these control regions and is defined by excluding the leptons from the E^T_{miss} calculation.

Dimuon and single-muon control region events are selected using full signal region criteria without the muon veto, and with the E_T^{miss} replaced by the p_T of the hadronic recoil system. Instead, in the dimuon control region, events with exactly two isolated muons with opposite charge, with $p_T^{\mu 1}$, $p_T^{\mu 2} > 20, 10$ GeV and an invariant mass between 60 and120 GeV are required. In the single-muon control region, exactly one tightly identified isolated muon with $p_T > 20$ GeV is required. The transverse mass of the muon-$\mathbf{p}_T^{miss}$ system is required to be less than 160 GeV. The transverse mass (MT) is computed as $MT = \sqrt{2E_T^{miss}p_T^{\mu}(1 - cos\Delta\phi)}$, where p_T^{μ} is the p_T of the muon, and $\Delta\phi$ is the angle between $\vec{p}_T^{\,\mu}$ and $\vec{p}_T^{\,miss}$.

Dielectron and single-electron control region events are selected using an

isolated single-electron trigger with a p_T threshold of 27 GeV combined with a non isolated single-electron trigger with a p_T threshold of 105 GeV. The non isolated trigger is required to recover some inefficiency caused by the boosted electrons getting included in each others isolation cones. The dielectron events are required to contain exactly two oppositely charged electrons with $p_T^{e1}, p_T^{e2} > 40$, 10 GeV. The invariant mass of the dielectron system is required to be between 60 and120 GeV, in order to be consistent with a Z boson decay. The single-electron events are required to contain exactly one tightly identified and isolated electron with $p_T > 40$ GeV. The QCD background in the single-electron control sample is suppressed by requiring $E_T^{miss} > 50$ GeV, and $MT < 160$ GeV.

All the control sample events are further required to satisfy all other selection requirements imposed on the signal events with the E_T^{miss} replaced by the p_T of the hadronic recoil system.

6. Signal Extraction

A binned likelihood fit to the data in all the hadronic recoil p_T bins of the control regions will be used to determine the E_T^{miss} and the jet mass spectra of the $Z(\nu\nu)$ + jets and $W(l\nu)$ + jets backgrounds. Transfer factors will be used to account for the extrapolation of specific backgrounds from the control regions to signal region. These transfer factors also rely on an accurate prediction of the ratio of V+jets cross sections. Therefore, LO simulations for the Z + jets, W + jets processes are corrected using p_T-dependent NLO QCD K-factors derived using MADGRAPH5 aMC@NLO. They are also corrected using p_T-dependent higher-order electroweak (EW) corrections extracted from theoretical calculations [22, 23, 24, 25, 26, 27].

Systematic uncertainties will be modeled as constrained nuisance parameters that allow variation of these transfer factors. These include both experimental and theoretical uncertainties.

For each control region, Z + jets, W + jets and QCD processes have been simulated at leading order (LO) using MADGRAPH5 aMC@NLO 2.2.3 [12], $t\bar{t}$, and single top processes have been produced using POWHEG 2.0 [28],

and a set of diboson samples with PYTHIA 8 [16].

In Figures 7 and 8 are reported the data and Monte Carlo simulations for the jet mass and the E_T^{miss} in $Z(e^+e^-)$ + jets and $W(e\nu)$ + jets control regions. As expected in the Figure 7, the dominant contribution comes from the $Z(ll)$ + jets process, while in the Figure 7 comes from the $W(l\nu)$ + jets process. The $Z(\mu^+\mu^-)$ + jets and $W(\mu\nu)$ + jets control regions show the same distributions.

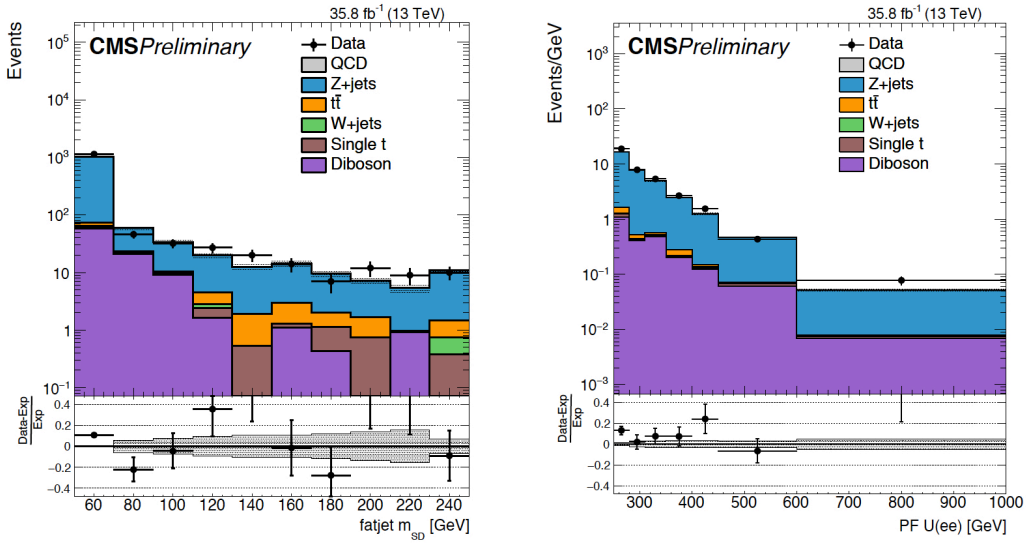

Figure 7: Data and Monte Carlo simulations in the Z(ll) + jets control region. The grey band in the bottom of the figure represent the theoretical uncertainty.

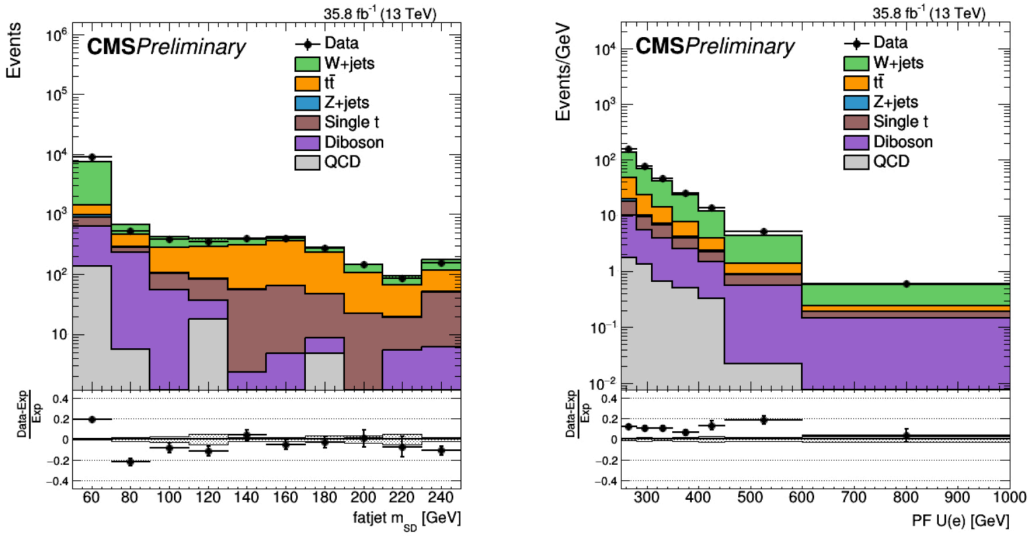

Figure 8: Data and Monte Carlo simulations in the W(lν) + jets control region. The grey band in the bottom of the figure represent the theoretical uncertainty.

7. Conclusion

A search for dark matter particles, invisible decays of a new 1-spin boson Z' has been presented. The model has been generated for different mass points, following LHC Dark Matter Forum recommendations, and validated. The analysis strategy has been set up: control regions have been defined and simulated in order to model the main background. The next step will consist in a simultaneous fit of the control regions and the signal region with CMS data and the model previously generated.

Bibliografia

[1] Giorgio Busoni, Andrea De Simone, Enrico Morgante, and Antonio Riotto. On the validity of the effective field theory for dark matter searches at the lhc. *Physics Letters B*, 728(Supplement C):412 – 421, 2014.

[2] O. Buchmueller, Matthew J. Dolan, and Christopher McCabe. Beyond effective field theory for dark matter searches at the lhc. *Journal of High Energy Physics*, 2014(1):25, Jan 2014.

[3] Daniel Abercrombie et al. Dark Matter Benchmark Models for Early LHC Run-2 Searches: Report of the ATLAS/CMS Dark Matter Forum. 2015.

[4] CMS Collaboration. Search for new physics in a boosted hadronic monotop final state using 12.9 fb^{-1} of $\sqrt{s} = 13$ TeV data. 2016.

[5] Yang Bai, James Bourbeau, and Tongyan Lin. Dark matter searches with a mono-$Z^{'}$ jet. *JHEP*, 06:205, 2015.

[6] S. Chatrchyan et al. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

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

[8] Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector. Technical Report CMS-PAS-PFT-10-001, 2010.

[9] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The Anti-k(t) jet clustering algorithm. *JHEP*, 04:063, 2008.

[10] Serguei Chatrchyan et al. Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS. *JINST*, 6:P11002, 2011.

[11] Performance of missing energy reconstruction in 13 TeV pp collision data using the CMS detector. Technical Report CMS-PAS-JME-16-004, CERN, Geneva, 2016.

[12] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. MadGraph 5 : Going Beyond. *JHEP*, 06:128, 2011.

[13] Pierre Artoisenet, Rikkert Frederix, Olivier Mattelaer, and Robbert Rietkerk. Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations. *JHEP*, 03:015, 2013.

[14] Philip Harris, Valentin V. Khoze, Michael Spannowsky, and Ciaran Williams. Constraining Dark Sectors at Colliders: Beyond the Effective Theory Approach. *Phys. Rev.*, D91:055009, 2015.

[15] R. Brun and F. Rademakers. ROOT: An object oriented data analysis framework. *Nucl. Instrum. Meth.*, A389:81–86, 1997.

[16] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008.

[17] https://github.com/cms-sw/cmssw.

[18] Yuri L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber. Better jet clustering algorithms. *JHEP*, 08:001, 1997.

[19] Stephen D. Ellis, Christopher K. Vermilion, and Jonathan R. Walsh. Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches. *Phys. Rev.*, D81:094023, 2010.

[20] Serguei Chatrchyan et al. Identification of b-quark jets with the CMS experiment. *JINST*, 8:P04013, 2013.

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

[22] Ansgar Denner, Stefan Dittmaier, Tobias Kasprzik, and Alexander Muck. Electroweak corrections to W + jet hadroproduction including leptonic W-boson decays. *JHEP*, 08:075, 2009.

[23] Ansgar Denner, Stefan Dittmaier, Tobias Kasprzik, and Alexander Muck. Electroweak corrections to dilepton + jet production at hadron colliders. *JHEP*, 06:069, 2011.

[24] Ansgar Denner, Stefan Dittmaier, Tobias Kasprzik, and Alexander Mück. Electroweak corrections to monojet production at the tevatron and the lhc. *The European Physical Journal C*, 73(2):2297, Feb 2013.

[25] Johann H. Kuhn, A. Kulesza, S. Pozzorini, and M. Schulze. Electroweak corrections to hadronic photon production at large transverse momenta. *JHEP*, 03:059, 2006.

[26] Stefan Kallweit, Jonas M. Lindert, Philipp Maierhfer, Stefano Pozzorini, and Marek Schnherr. NLO electroweak automation and precise predictions for W+multijet production at the LHC. *JHEP*, 04:012, 2015.

[27] Stefan Kallweit, Jonas M. Lindert, Philipp Maierhofer, Stefano Pozzorini, and Marek Schnherr. NLO QCD+EW predictions for V + jets including off-shell vector-boson decays and multijet merging. *JHEP*, 04:021, 2016.

[28] Carlo Oleari. The POWHEG-BOX. *Nucl. Phys. Proc. Suppl.*, 205-206:36–41, 2010.