W' and Z' bosons search at $\sqrt{s} = 13$ TeV with the ATLAS detector at the LHC

Student: Arturo Prieto Tirado
Advisor: Magnar Kopangen Bugge

University of Oslo

Spring 2020

Abstract

Invariant and transverse mass distributions of Beyond the Standard Model signals of Z' and W' bosons have been studied using 1 lepton and 2 lepton final states in ATLAS data, resulting in exclusion with more than 95% confidence for 2 and 3 TeV Z' and W', as well as setting exclusion limits of 3.3 TeV on the Z' and 3.7 TeV on the W'. Also, data-driven hadronic jet background estimation and high-mass background fitting have been studied and implemented.

1 Introduction

1.1 Aim

The Higgs boson, the last ingredient of the particle physics theory known as the Standard Model (SM), has been discovered by the ATLAS [1] and CMS [2] experiments at the Large Hadron Collider (LHC). The next step is to measure the Higgs boson properties and nature. Is it the long-awaited SM particle? Is it the first in a long series of scalar particles? With the advent of higher energies and higher collision rates the Large Hadron Collider (LHC) continues the exciting voyage towards new physics phenomena, allowing physicists all over the world to explore a previously unknown territory full of promise. The ambitious LHC physics programme may shed light on some of the greatest mysteries in physics today: (i) the nature of dark matter (DM), (ii) the behaviour of the gravitational force at the microscopic scale, and (iii) do new fundamental forces show up at the LHC the way the Z and Higgs bosons did? The project described here concentrates on item (iii). The focus of this work is to search for evidence of new weak interactions leading to new gauge bosons, using proton-proton (pp) collisions at the LHC from Run II at 13 TeV with leptonic final states, either with two leptons or one lepton in association with missing transverse momentum.

These data will be compared to the corresponding SM predictions as well as predictions obtained from beyond the Standard Model (BSM) theories. Such theories regularly introduce a number of additional gauge bosons and aim at resolving some of the limitations of the SM.

With the advent of modern particle physics, the wide range of forces occurring in nature has been discerned as simply scale-dependent manifestations of four fundamental forces. Out of these four, three (electromagnetic, weak and strong forces) are seen as exchanges of gauge bosons (γ, $W^{\pm}$, Z, and 8 gluons), between interacting particles, all of which are described in the Standard Model (SM). The fourth (the gravitational force) stands out by being far weaker than the other three, and is therefore routinely neglected in calculations. A way to search for signatures of new gauge bosons at the LHC, and thus reveal the existence of additional fundamental interactions, is to look for new resonances decaying, among other final states, into (i) a pair of oppositely charged leptons, directly measurable in particle detectors, or (ii) a charged lepton and an electrically neutral neutrino, in which case the neutrino must be indirectly measured by considering the momentum balance in the event. New gauge bosons are predicted by several models motivated by solutions to several short-comings of the SM, such as the hierarchy problem [3], the mystery of parity violation [4], and the need for a unification of all fundamental forces [5].

Models with dilepton resonances are predicted in many scenarios for new physics [6]. Among these are grand-unification models, which are motivated by gauge unification or a restoration of the left-right symmetry violated by the weak interaction. These models predict the existence of additional neutral, spin-1 vector gauge bosons called Z' bosons, due to the existence of larger symmetry groups that break to yield the SM gauge group and additional U(1) gauge groups. Specific examples typically considered include the Z' bosons of the E_6-motivated and minimal models [7]. Another Z' signal, the Z'$_{SSM}$, is considered due to its inherent simplicity and usefulness as a benchmark model [8]. The sequential Standard Model (SSM) includes a Z'$_{SSM}$ boson with couplings to fermions equivalent to those of the SM Z boson. In some theories, such as those motivated by restoration of left-right (parity) symmetry, the Z' boson naturally arises in association with a charged boson W'. Similar to the Z' case, a useful benchmark model is the SSM, where the W'$_{SSM}$ boson has couplings to fermions equivalent to those of the SM W boson.

1.2 The ATLAS detector

In order to get particles to interact with each other as described in the particle physics models, one needs to give them some energy so that this is more probable to happen. For processes involving the production of new, heavy particles, there are also energy thresholds (because we need to have at least enough energy to be able to generate the mass of the new particle at rest), so that it would really be impossible to happen for energies below the threshold. On the other hand, for us to maximize the number of events we get, it's interesting to have as many particles as possible and in a confined space or "beam" so that interactions occur easily. These are the reasons why particle accelerators are used to study particle physics.

In particular, the experiment whose data we are going to use in this analysis is the ATLAS experiment [9]. This experiment is based in Geneva, Switzerland, and uses the Large Hadron Collider (LHC) as its accelerator for proton-proton collisions. There are other experiments that use this same accelerator, like CMS or Alice, and another accelerator, LEP, that used the same tunnel to collide electron-positron instead of protons. This collection of experiments are all part of the European Association for Nuclear Research, also known as CERN for its name in French (Conseil Européen pour la Recherche Nucléaire).

The LHC, as shown in figure 1 is a 27 km-long ring built with superconducting magnets and accelerating structures so that the protons get accelerated while they are doing laps around the circuit. These magnets need to produce very strong magnetic fields because the particles are moving very fast, close to the speed of light, that's why usually, state of the art technology, like new superconducting magnets, need to be developed just for this purpose. Some of the magnets are used to bend the beam, these are called "dipoles". Quadrupoles focus the beam so that they don't go away. Remember that they are protons, so have the same electric charge and will tend to repel each other, so some counter action is needed. Other magnets, sextupoles, octupoles, etc. are used to squeeze the particles so that the probability of interaction is higher. The accelerating structures are called Radio Frequency Cavities or RF cavities. They work by releasing radio waves whose electric field impulses the protons like the waves of the sea can impulse us. Also, since the pulse is a wave, it helps to get the particles together since the fastest one will get less impulse due to being further, whilst the others that are on the crest of the wave will be given more.

Once the protons are ready to collide, i.e. they have the energy we want, two beams are made to collide going in opposite directions. The point where they collide, one sets up a detector to measure the different particles that result from the collision. In particular, the ATLAS detector is made of different layers, designed to target a specific particle/group of particles. A general overview is explained in figure 2.

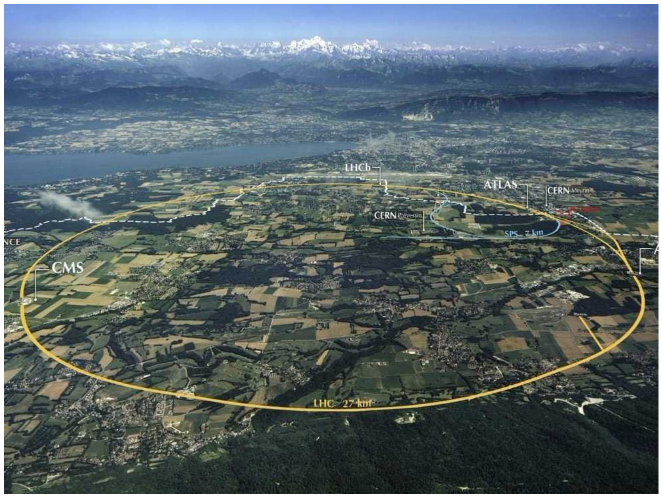

Figure 1: LHC ring air view

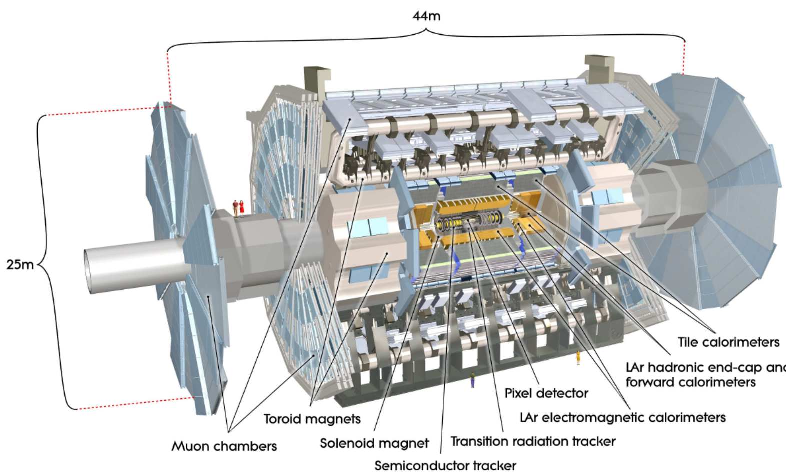

Figure 2: Picture of the ATLAS detector summarizing up its main parts. The inner part is composed of trackers that track the movement of the charged particles, so that one can know the charge and momentum of the particle by studying its bending due to the magnetic field of the present magnets. Curving left or right gives the sign of the charge while the radius of curvature gives an idea of the momentum of the particle. Afterwards, specific calorimeters are designed for each kind of particle to deposit all of its remaining energy so that we can measure it: those are electromagnetic calorimeters, for electrons and photons and hadronic, for hadrons. Muons and neutrinos are two elusive particles. The first are heavy so they interact little in the electromagnetic calorimeter and therefore specific muon chambers are built in the outer part to serve as additional tracking detectors for muons. Neutrinos, however, are almost impossible to detect so they go away without any measurement, one detects them by analyzing the missing transverse momentum.

After each particle triggers the detector, the events as a whole are saved or discarded depending if the data passes some series of restrictions or filters called "the trigger system". This is necessary because the amount of information generated is really large since there are billions of collisions per second. After a lot of filtering, a really tiny part of the total number of events is stored for further analysis.

1.3 Search for the bosons

Now, the way one searches or checks physical theories is by comparing the experimental results with simulated ones obtained assuming the theory one wants to test and the experimental conditions. These simulations are done following the Monte Carlo method, which generates a distribution of events following the SM rules which are then used to simulate the detection and the reconstruction of the information. Regarding this work, it consists in an analysis of ATLAS data, taken at the highest available energies, and a comparison to simulation data.

Since we only measure the final state of the process in the detector, we would need to infer from that the particles that were previously there. There are multiple possible origins for the final state particles that we measure. They could come from one of our "new bosons" Z' and W' or they could have just been produced by SM processes, this is called the background. Comparing the real data with the Monte Carlo samples of this SM background, as well as corresponding samples for the new physics signal, will help us on determining the feasibility of the new model(s) performing some statistical analysis. In our case, we will look at the invariant and transverse mass distribution and look for a "bump" in the data that is not compatible with the SM background but with the new boson signals. Here, we are going to look for final states with one lepton or two leptons, so the following processes will be used:

Search for Z' : The Z' would be produced in this form: pp $\rightarrow$ Z'+X $\rightarrow$ l^+ l^- + X. But other SM processes like the Drell-Yan process (Z$\rightarrow$ l^+ l^-), top quark pair production and decay or $W^+W^- \rightarrow l^+ \nu$ $l^- \bar{\nu}$ will also generate two oppositely charged leptons (l^+, l^-), so, as said before, it would be necessary to study the compatibility of the background and the signal to the data.

Search for W' : pp $\rightarrow$ W'+X $\rightarrow$ l ν+X. In analogy to the Z' decay, we'll expect a lepton final state, but in this case together with a neutrino. Because the detector doesn't detect the neutrino, its momentum can only be indirectly measured as the "missing transverse momentum". With "transverse", we refer to the plane orthogonal to the beam line, the longitudinal direction. As a result of this, the invariant mass cannot be reconstructed, and the transverse mass, defined as $m_T = \sqrt{2p_T E_T^{miss}(1 - \cos\phi_{l\nu})}$, is used instead as the discriminating variable in the search, where $\phi_{l\nu}$ is the azimuthal angle between the lepton transverse momentum p_T and the missing transverse energy E_T^{miss} directions in the transverse plane. That is, the direction of motion of the lepton and the neutrino, respectively. In this final state, one expects a larger contribution from background processes where a hadronic jet ("QCD jets") is mistakenly reconstructed as a lepton, for which a data-driven background estimation may be necessary. The feasibility of performing such a background estimation within the ATLAS Open Data context will be evaluated.

1.4 Statistics introduction

The resulting invariant mass distribution we will get will look something like figure 3. Just from the plot we can visually judge whether the background and signal fit the experimental data.

However, it would be necessary to have some quantitative measure to either confirm the signal, which will mean a discovery, or to reject it, therefore excluding the model from reality. The parameter that quantifies the discovery is called "significance" Z. For the exclusion, it is called "confidence level" CL.

In order to do this, we have to extract quantitative information from the histogram. To do that, we integrate (i.e. sum the content of its bins) the histogram from a given mass cut until the end. To avoid biases in this cut, one chooses the cut that maximizes the expected significance, which we will introduce in a moment. After the integration, we have summed up the histogram information into three numbers: the number of observed experimental events, n_{obs}; the expected number of SM background events, b and the expected number of signal events, s. At this stage, there are two approaches one could use, frequentist or Bayesian. We will use both of them in the analysis.

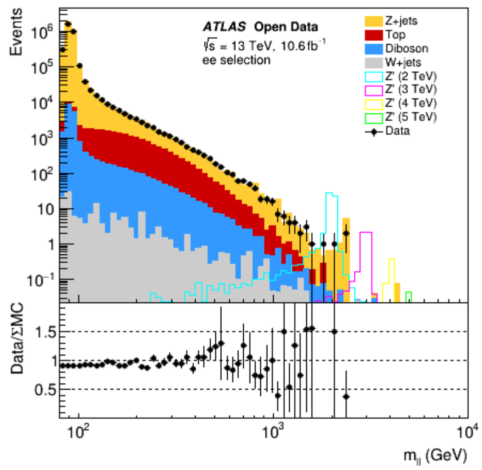

Figure 3: Example of an invariant mass distribution plot. We can see the experimental data as well as the different SM backgrounds that contribute to generate, in this case, electron-positron pairs. We could argue in favour of the signal hypothesis if we saw a "bump" or an increase of the data relative to the SM background. On the other hand , we could speak about excluding the signal model if the experimental data is clearly below the signal prediction (or even zero, as it's the case for high mass Z's).

Assuming a Poisson distribution for the events, the frequentist analysis gives the following expressions for the expected and observed significances (not well-defined for $b = 0$):

$$Z_{exp} = \sqrt{2[(s+b)\ln(1+\frac{s}{b}) - s]} \tag{1}$$

$$Z_{obs} = \sqrt{2[n_{obs}\ln(\frac{n_{obs}}{b}) - n_{obs} + b]} \tag{2}$$

These don't take into account uncertainties on s, b. Not taking into account b's uncertainty means they are optimistic. On the other hand, when $n_{obs} = 0$, the frequentist analysis yields the simple expression

$$CL_{s+b} = e^{-(s+b)} \tag{3}$$

Therefore, to exclude with 95% confidence will result in having less than 3 signal events maximum, but also depending on the background we could even need negative s, thus being susceptible to background fluctuations.

$$s = -\ln(0.05) - b = 3 - b \tag{4}$$

Since $s > 0$ obviously, this problem has been solved and the typical solution (shared also in the bayesian analysis, but without this problem from the beginning) is

$$CL_s = e^{-s} \tag{5}$$

So that the expression we will use for the Confidence limit is the following one. Which means we can exclude a signal if $s > 3$ when $n_{obs} = 0$

$$CL = 1 - e^{-s} \tag{6}$$

This frequentist analysis is only valid when we have no events in the experimental data, $n_{obs} = 0$. It could also be generalised when we have $n_{obs} \neq 0$. However, we will use the Bayesian formulation to achieve that in order to benefit from the numerical tools available in the FYS5555 course. Also, the Bayesian formulation directly solves the exclusion dependence on the background shown in equation 4.

The main idea of the Bayesian analysis is using the Bayes theorem.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \rightarrow P(Theory|Data) = \frac{P(Data|Theory)P(Theory)}{P(Data)} \tag{7}$$

This way, we have the probability of the theory given the data, $P(Theory|Data)$. That is, it quantifies our degree of belief in the new theory after we see the experimental data, which is really what we want when speaking about accepting or discarding theories. On the other hand, $P(Data|Theory)$ represents the likelihood of the events we see, i.e., the chances we have on actually measuring a specific number of events if assuming the theory as true so that, if we repeated the measurement a very big number of times, N, the number of occurrences of that specific number of events would approach $N \times P(Data|Theory)$. This last probability is what can be obtained with the frequentist analysis whilst $P(Theory|Data)$ is not possible. It's also interesting to introduce the concept of "prior probabilities", that makes the Bayesian formulation special. $P(Theory)$ is the prior probability of the theory, that is, our degree of belief on the theory before seeing the data, an assumption we have to make in order to work with the Bayes theorem.

Now we need to specify the different probabilities. Defining $\lambda = s + b$:

$$P(Data|Theory) = L(s) = P(n_{obs}|s) = \frac{\lambda^{n_{obs}} e^{-\lambda}}{n_{obs}!} \tag{8}$$

We get to

$$P(s|n_{obs}) = \frac{P(n_{obs}|s)P(s)}{P(n_{obs})} \tag{9}$$

The probability of the observed data, $P(n_{obs})$, has no dependence on s. Also, if we assume that there is no prior knowledge on the signal, $P(s)$ is also independent of s, this is referred to as a flat prior. This way, we can just englobe them into a normalization constant, N.

$$P(s|n_{obs}) = NP(n_{obs}|s) = N(s+b)^{n_{obs}} e^{-(s+b)} \tag{10}$$

$$N \quad \text{determined} \quad \text{by} \int_0^\infty P(s|n_{obs})ds = 1 \tag{11}$$

The final step in this discussion would be to take into account the fact that the background has an uncertainty δb. Assuming variations δb on b around its measured value $\bar{b}$. Then, the likelihood becomes a function of both s and b: $L(s, b)$ thus implying:

$$P(s, b|n_{obs}) = \frac{P(n_{obs}|s, b)P(s)P(b)}{P(n_{obs})} = N(s+b)^{n_{obs}} e^{-(s+b)} e^{-\frac{(b-\bar{b})^2}{2(\delta b)^2}} \tag{12}$$

This way we can calculate the posterior probability for the signal by getting rid of the background dependence:

$$P(s|n_{obs}) = \int_0^\infty P(s, b|n_{obs})db \tag{13}$$

A numerical tool provided in the FYS5555 course has been used to determine the signal exclusion limit (i.e. the maximum value of the signal count before it's considered excluded, the 95% quantile of the posterior probability) based on these Bayesian concepts once we had previously selected the best way of integrating the histogram (maximum expected significance) from the frequentist analysis, getting s, b, δb, n_{obs}. The results of these calculations will be explained in an exclusion limit plot like the one shown in figure 4.

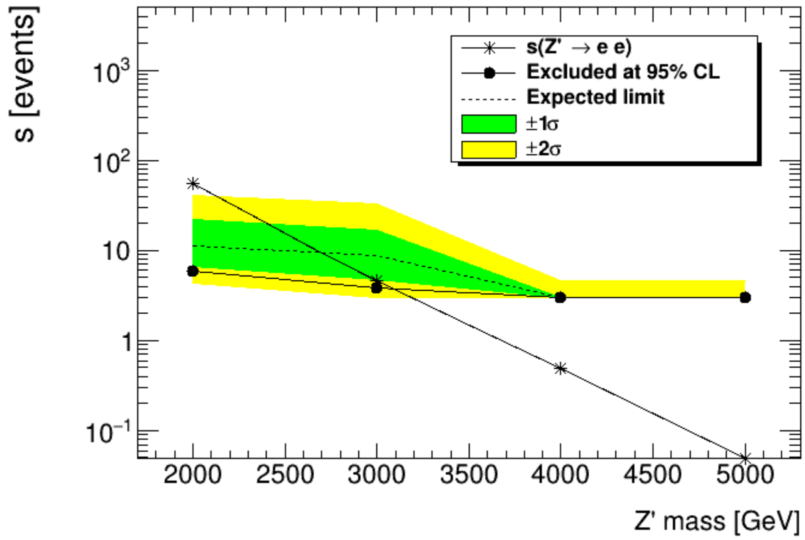

Figure 4: Example of a limit plot, showing the signal event count s for a given Z' signal model characterized by the boson's mass. The crossed points are the ones that represent the actual signal count we got from the histogram. The dotted points represent the value they should have so that we could exclude them at least with 95% confidence, the limit value. Therefore, any value (crossed point) that is higher than these points indicates us that the signal can be excluded. In this case, we see that 2 and 3 TeV signals could be excluded. The higher mass signals can't be excluded but we do set an upper limit for their exclusion by looking at the intersection between the observed data line and the exclusion limit line. In this case, that point is around 3100 GeV, so we would exclude all Z' whose mass is below 3.1 TeV. Also, the expected limit and green/yellow bands represent an ensemble of limits calculated with pseudo-data drawn from the background distribution, to take into account the background uncertainty and the possible effects of statistical fluctuations.

2 Data analysis and results

First, the data from the CERN Open data portal has been downloaded [10] [11], including measured data values and background MC samples. The MC samples also include simulations of the particles beyond the Standard Model that we're looking for: the Z' and W' bosons.

The models used for the Z' are the so-called Z'$_\chi$ of the E_6-motivated models. In the case of the W' it's a bit more elaborated, because we had to "reproduce" some W' samples based on the Z' ones. Due to the unavailability of W' signal samples in the open data, we created samples mimicking the W' process by using the events of the Z' samples and treating one of the leptons as invisible (i.e. contributing to the missing transverse momentum, as if it was the neutrino). The cross sections used for normalisation of the W' samples correspond to the SSM model, although with the couplings scaled down to match the width of the Z'$_\chi$ samples.

In order to work with the data from the ATLAS open data portal we first have to filter the information we want. The events have been sorted and classified using ROOT and C++, gathering the information in ROOT histogram files according to the desired variable: Invariant mass for dilepton final states and transverse mass for one lepton final states in order to look for the "bump" we mentioned above. Also, other variables like the momenta of the particles or the missing transverse energy (met) have been studied for more completeness.

2.1 Data validation and event selection

First of all, it's good to check that the MC describes well the data so that we really know we can trust it. See figure 5.

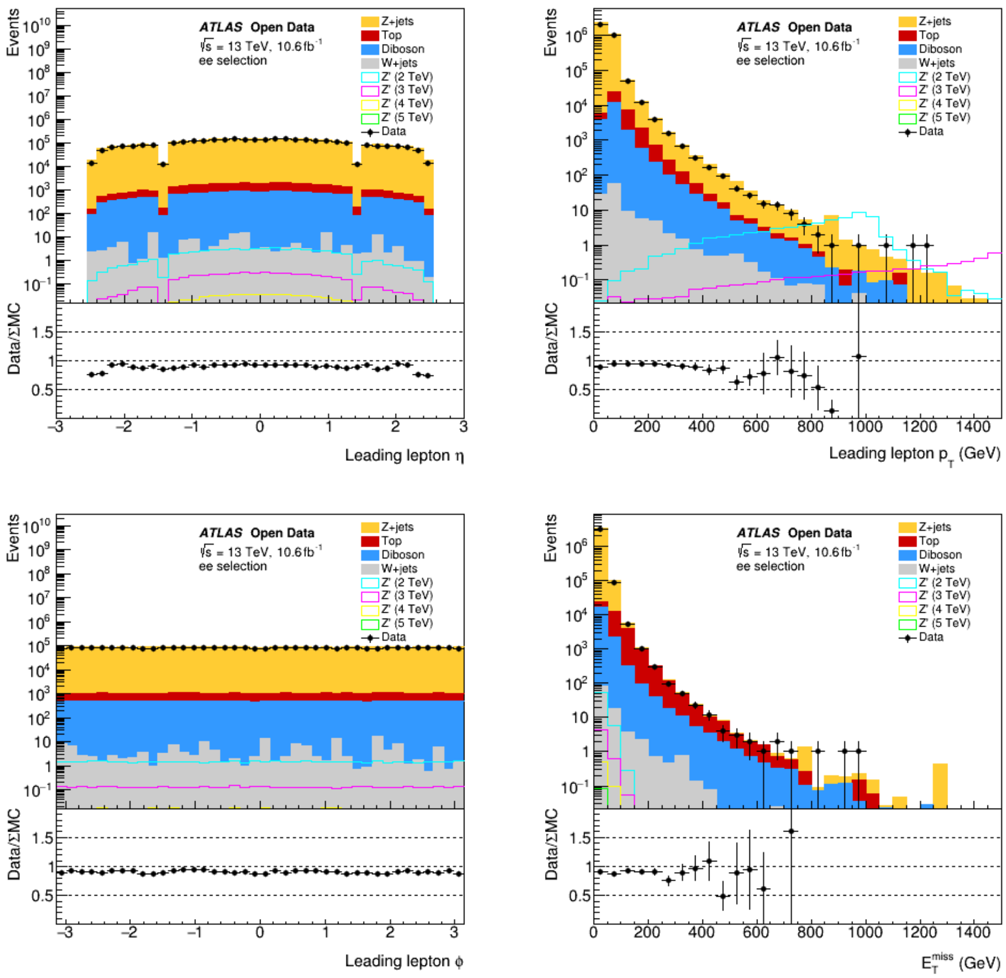

Figure 5: Example distribution for the 2 lepton study of: pseudorapidity, η; transverse momentum, p_T; azimuthal angle ϕ and missing transverse energy, E_T^{miss}

There is good agreement so now we can motivate the different cuts we are going to apply in order to search for the events we want. In the 2 lepton analysis, we select all the events with only two leptons, with opposite charge and same flavour (electron-positron, muon-antimuon). In the 1 lepton analysis we only need to require 1 lepton. Then, we perform some additional filters on the kinematics shown in table 1. Due to an issue discovered in the open data ntuples, the recommended cut on the d_0 (transverse impact parameter) significance could not be applied, so it's not included.

2.2 Dilepton final states: Z'

2.2.1 Invariant mass distribution

The first analysis is the one of the invariant mass distribution for final states of electron-positron and muon-antimuon. These dilepton final states, as mentioned earlier, allow us to study the Z' boson produc-

	ee channel 2lep	e ν channel 1lep	$\mu\mu$ channel 2lep	$\mu\nu$ channel 1lep
ID level	Tight			
$z_0 \sin(\theta)$/mm	< 0.5			
$etcone20$/lepton p_T	> -0.1 and < 0.2			
lepton p_T/GeV	> 25	> 65	> 25	> 55
lepton $\lvert\eta\rvert$	< 2.47		< 2.4	
lepton trigger	Electron trigger		Muon trigger	
M/GeV	> 80	-	> 80	-

Table 1: Required filters applied to the data

tion, so different signals for different masses of the Z' have been included in the analysis shown in figure 6 and 7.

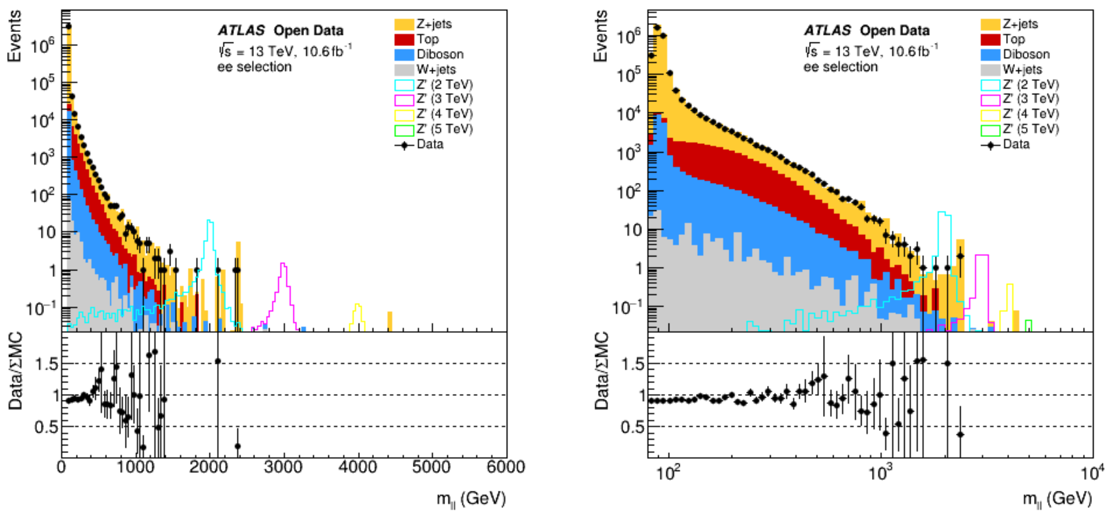

Figure 6: Invariant mass m_{ll} distribution on the electron channel in the 2-lepton study. Left, normal binning; right, logarithmic binning. Experimental data are shown as black dots; different Standard Model backgrounds contributing to the process, as solid color. New bosons' signals are also introduced, having a resonance peak around their mass energy value.

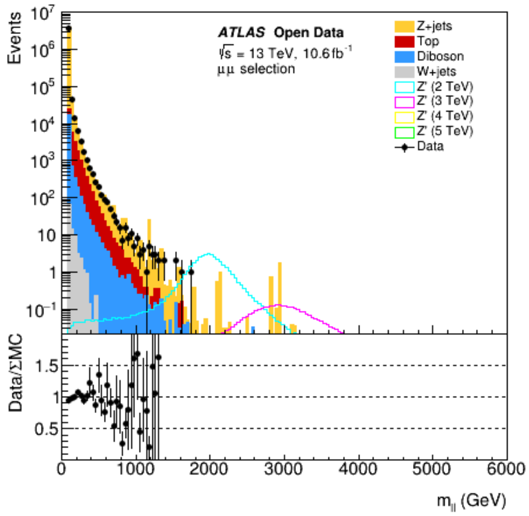

Figure 7: Invariant mass m_{ll} distribution on the muon channel in the 2-lepton study. Left, normal binning; right, logarithmic binning. Experimental data are shown as black dots; different Standard Model backgrounds contributing to the process, as solid color. New bosons' signals are also introduced, having a resonance peak around their mass energy value.

2.2.2 Histogram integration and Statistics

Now, we would like to know the confidence limit that we can attribute to each of the Z' signals given how they adjust to the present background and data. In order to do that, as explained before, we'll perform the integration of the histogram and find b, δb (from MC statistical uncertainty), n_{obs} and s.

However, there are multiple cuts from where we could start integrating. We would like to know the best cut to make in order to get the maximum expected significance. To that end, figure 8 shows the distribution of the expected significance depending on where we perform the cut.

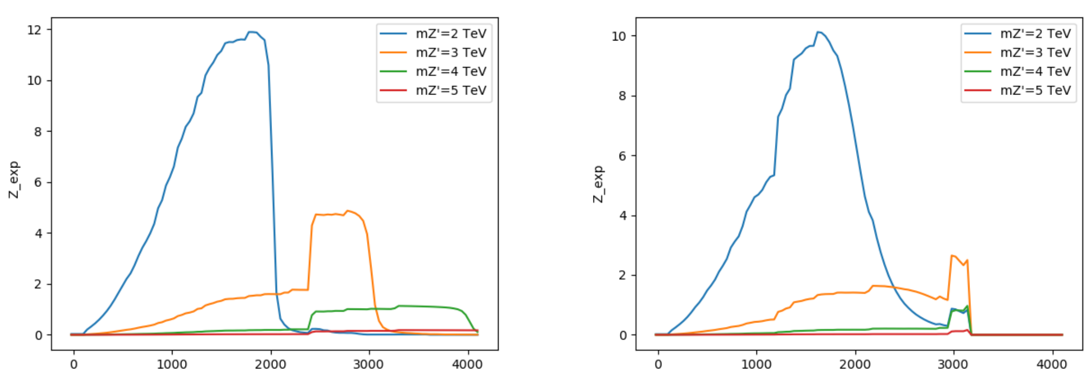

Figure 8: Expected significance optimization for the 2 lepton analysis: The expected significance for a given bin cut of the histogram integration is analysed for each of the respective signals. Left-hand side corresponds to electron channel; right-hand side, to the muon channel.

Results in figure 8 are not so smooth due to the emptiness of some of the bins at high masses that generate discontinuities when searching for the optimal cut. However, based mostly on the 2 TeV curve, which has no big discontinuity, we can establish a general criterion of around 80% of the mass of the boson. After we perform the cut and the integration, we arrive at the following results in table 2 and 3.

$Zmass$/GeV	Cut/GeV	b/counts	δb/counts	s/counts	n_{obs}/counts	Z_{exp}	Z_{obs}
2000	1600	10	7	55.0	4	11.88	-
3000	2400	6	5	4.64	1	4.87	-
4000	3200	0.12	0.12	0.49	0	1.13	-
5000	4000	0.08	0.08	0.05	0	0.18	-

Table 2: Results of the histogram integration for the electron channel in the 2 lepton analysis. For each of the different signals representing a Z' boson of mass Zmass the integration has been performed from a bin cut of Cut. Results of the background b, signal s and data n_{obs} are shown. From these results, the expected significance Z_{exp} has been estimated by using the results shown in section 1.4. The observed one Z_{obs} is not applicable because $n_{obs} < b$.

$Zmass$/GeV	Cut/GeV	b/counts	δb/counts	s/counts	n_{obs}/counts	Z_{exp}	Z_{obs}
2000	1600	7	5	37.83	2	10.12	-
3000	2400	3	2	3.05	0	2.65	-
4000	3200	0.0	0.0	0.3	0	0.97	-
5000	4000	0.0	0.0	0.03	0	0.16	-

Table 3: Results of the histogram integration for the muon channel in the 2 lepton analysis. For each of the different signals representing a Z' boson of mass Zmass the integration has been performed from a bin cut of Cut. Results of the background b, signal s and data n_{obs} are shown. From these results, the expected significance Z_{exp} has been estimated by using the results shown in section 1.4. The observed one Z_{obs} is not applicable because $n_{obs} < b$.

Now that we have the statistics condensed into the parameters we need to perform the Bayesian calculation, we can set the exclusion limits on the signal, shown in figure 9.

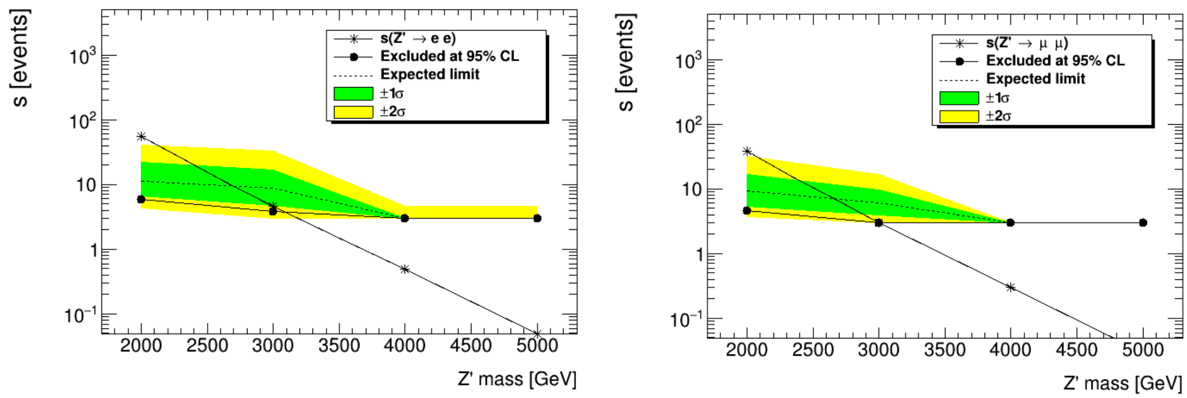

Figure 9: Limit plots for the electron channel (left) and muon channel (right) analysis. The crossed points are the ones we got from table 2 and 3. The dotted points represent the value they should have so that we could exclude them at 95% confidence. Expected limit and green/yellow bands represent an ensemble of limits calculated with pseudo-data drawn from the background distribution. We can clearly see how 2 and 3 TeV Z's are excluded. In fact, if we look at the intersection with the observed limit and the data lines, we see that any signal below 3 TeV would be excluded.

2.2.3 Background fitting

We can see how the events we get decrease with the invariant mass. Therefore, the total number of MC events, the "statistics" we have, is important to have some precision on high masses. As we can see in previous plots, it gets to a point where there are basically just a couple of background events, which causes that statistical fluctuations or limitations in the MC set production can give us an actual background level of zero. We can see that this creates some problems in a significance optimization and the subsequent process by generating some discontinuities. In order to avoid this inconvenience, one could think of fitting and interpolating the background so that we get a "smoother" shape trying to alter the results as little as possible. The next analysis will consist on fitting the most relevant background, that is, the Z+ jets.

In order to do that, we use three different functions:

$$f_1(m_{ll}) = e^{-a} m_{ll}^{b} m_{ll}^{c\log(m_{ll})}$$

$$f_2(m_{ll}) = \frac{a}{(m_{ll}+b)^c}$$

$$f_3(m_{ll}) = f_{BW,Z}(m_{ll})(1-x^c)^b x^{\sum_{i=0}^{3} p_i log(x)^i}$$

Where a, b, c, p_i are the fitting parameters, m_{ll} is the invariant mass, $x = \frac{m_{ll}}{\sqrt{s}}$ and $f_{BW,Z}$ is a non-relativistic Breit-Wigner function with m_Z=91.1876 GeV and Γ_Z=2.4952 GeV. In particular, f_3 is the function used to fit the invariant mass distribution of the data in the official ATLAS Z' analysis [12]. The different fits are shown on figure 10.

From here, we can take the average and add the difference as uncertainty. The resulting invariant mass plot is shown in figure 11 from where we can analyse the optimal cut, now with smoother curves in figure 12. The plot is indicates something slightly higher, but not too far from the 80% "rule". However, in any case, for exclusion one would typically benefit from having a bit lower threshold. Therefore, we still use the 80% "rule" and perform the histogram integration shown in table 4 and 5.

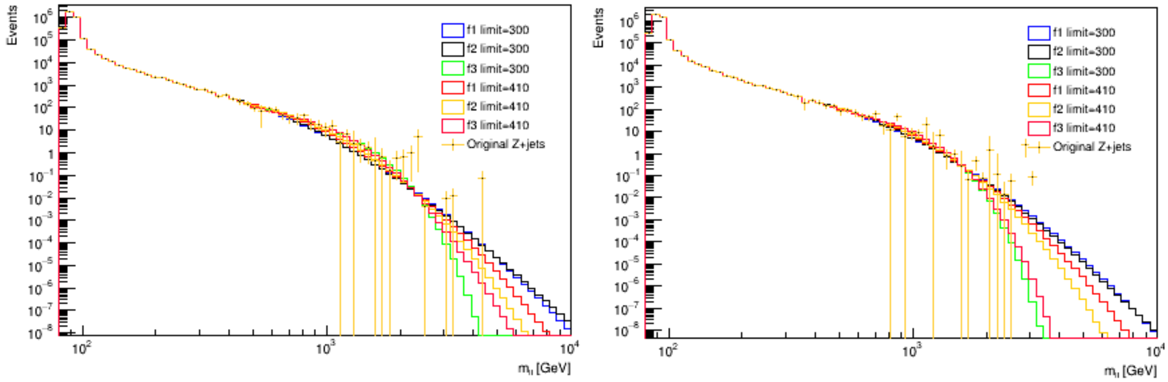

Figure 10: Different fits for the invariant mass distribution of the Z+ jets background in the 2 lepton analysis. Left hand side corresponds to electron channel; right hand side, to the muon channel. The fitting region was from a lower limit of limit in the legend, either 300 or 410 GeV, to 3500 GeV. Two different lower limits were chosen so that we had several different fits to average. The original background is also shown together with its uncertainty.

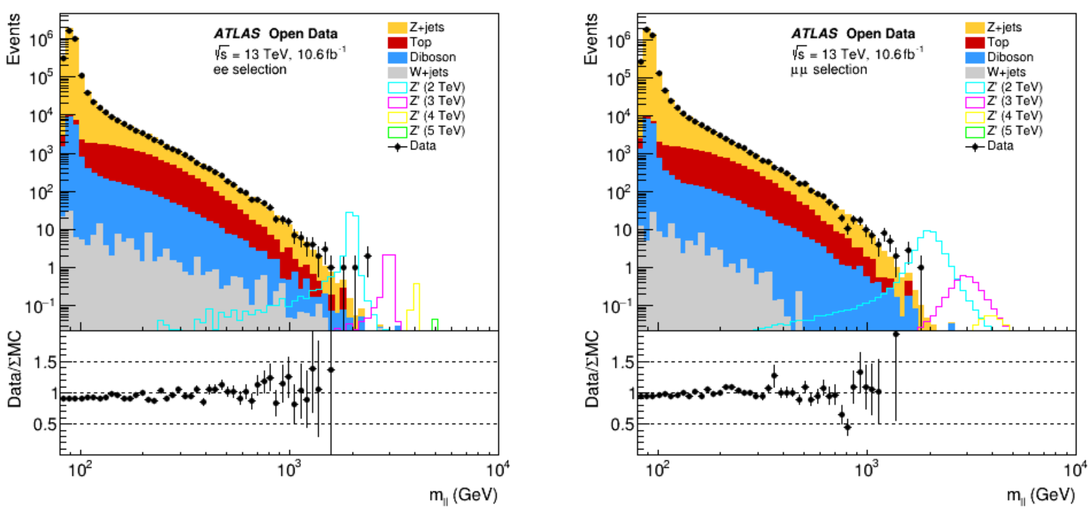

Figure 11: Invariant mass m_{ll} distribution on the electron channel (left) and muon channel(right) in the 2 lepton study after performing the Z+ jets background fitting. Experimental data are shown as black dots; different Standard Model backgrounds contributing to the process, as solid color. New bosons' signals are also introduced, having a resonance peak around their mass energy value.

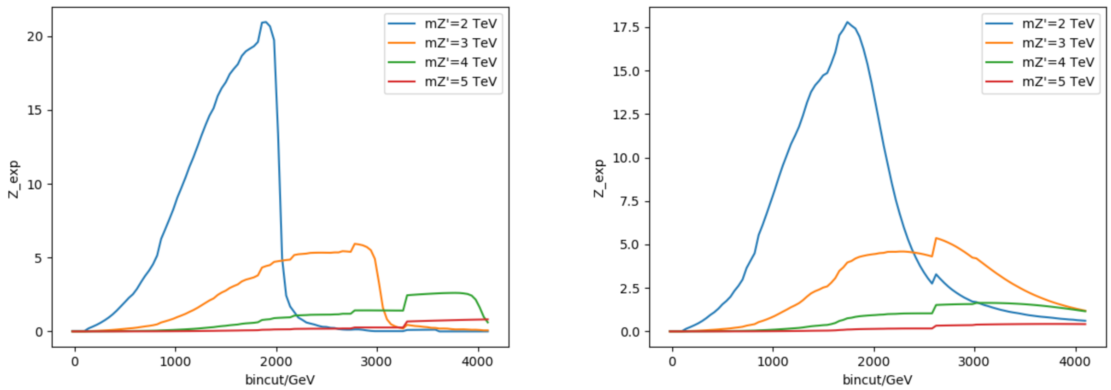

Figure 12: Expected significance optimization for the 2 lepton analysis with Z+jets background fitting: The expected significance for a given bin cut of the histogram integration is analysed for each of the respective signals. Left-hand side corresponds to electron channel; right-hand side, to the muon channel.

$Zmass$/GeV	Cut/GeV	b/counts	δb/counts	s/counts	n_{obs}/counts	Z_{exp}	Z_{obs}
2000	1600	1.3	0.2	55.0	4	20.96	1.92
3000	2400	0.09	0.05	4.64	1	5.93	1.74
4000	3200	0.03	0.03	0.49	0	2.61	-
5000	4000	$5.1\ 10^{-5}$	$1.3\ 10^{-5}$	0.05	0	0.27	-

Table 4: Results of the histogram integration for the electron channel in the 2 lepton analysis after the Z+ jets background fitting. For each of the different signals representing a Z' boson of mass $Zmass$ the integration has been performed from a bin cut of Cut. Results of the background b, signal s and data n_{obs} are shown. From these results, the expected and observed significances Z_{exp} and Z_{obs} have been estimated by using the results shown in section 1.4.

$Zmass$/GeV	Cut/GeV	b/counts	δb/counts	s/counts	n_{obs}/counts	Z_{exp}	Z_{obs}
2000	1600	0.7	0.2	37.83	2	17.79	1.33
3000	2400	0.04	0.03	3.05	0	5.37	-
4000	3200	0.0014	0.0012	0.3	0	1.64	-
5000	4000	0.0004	0.0004	0.03	0	0.43	-

Table 5: Results of the histogram integration for the muon channel in the 2 lepton analysis after the Z+ jets background fitting. For each of the different signals representing a Z' boson of mass $Zmass$ the integration has been performed from a bin cut of Cut. Results of the background b, signal s and data n_{obs} are shown. From these results, the expected and observed significances Z_{exp} and Z_{obs} have been estimated by using the results shown in section 1.4.

Finally, we can use the integration parameters to calculate the signal exclusion limits, shown in figure 13.

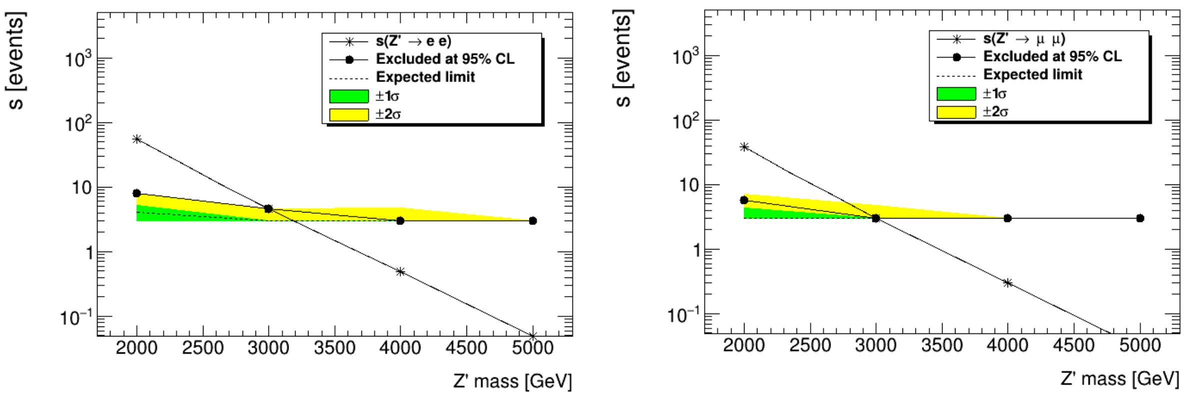

Figure 13: Limit plot for the electron and muon channel of the 2-lepton selection with Z+ jets fitting implemented. Expected limit and green/yellow bands represent an ensemble of limits calculated with pseudo-data drawn from the background distribution. We see that 2 and 3 TeV are above the limit, so we can exclude those signals. The general exclusion limit is set on around 3 TeV.

2.3 One lepton final state: W'

In this final state we don't know all the invariant mass because we cannot detect the neutrino, so we work with transverse mass distributions. The same procedure of extracting the data, MC and signals and storing them in histograms is also applied here, leading to the results shown in figure 14 and 15.

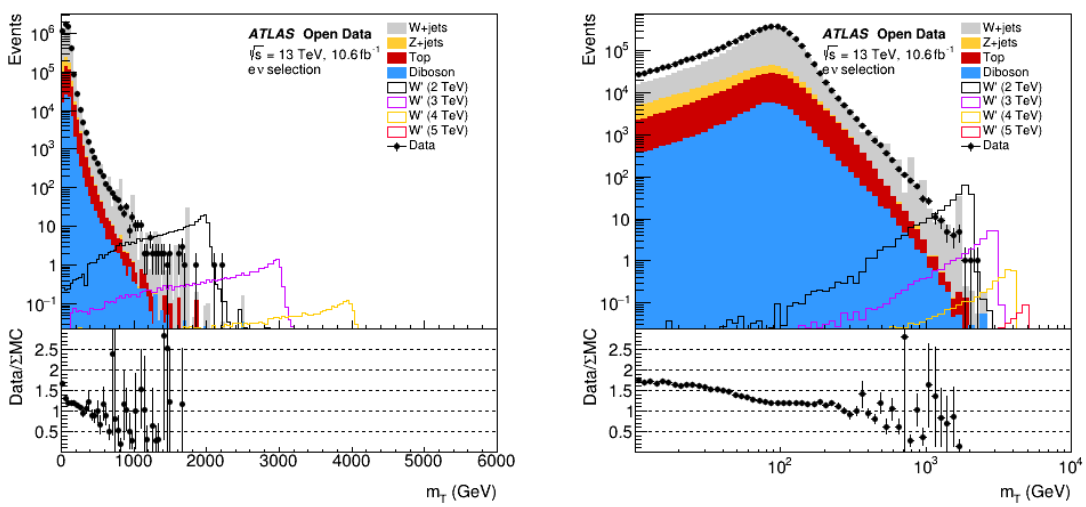

Figure 14: Transverse mass m_T distribution on the electron channel in the 1-lepton study. Left, normal binning; right, logarithmic binning. Experimental data are shown as black dots; different Standard Model backgrounds contributing to the process, as solid color. New bosons' signals are also introduced, with a long tail that extends from the left to their energy mass value from where it decreases rapidly. The so-called Jacobian peak.

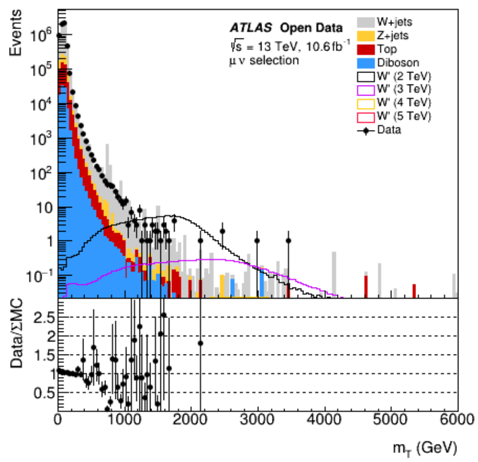

Figure 15: Transverse mass m_T distribution on the muon channel in the 1-lepton study. Left, normal binning; right, logarithmic binning. Experimental data are shown as black dots; different Standard Model backgrounds contributing to the process, as solid color. New bosons' signals are also introduced, with a long tail that extends from the left to their energy mass value from where it decreases rapidly. The so-called Jacobian peak.

2.3.1 QCD analysis

We can see in the previous electron channel m_T distribution, figure 14, that points in the low mass region seem to have some systematic deviation from the background events. This effect suggests that we could be missing some other backgrounds not taken into account, for example, leptons that could have been generated in QCD jets, as described in the introduction.

The main ideas of the QCD background estimation procedure are: first, defining a QCD control region from where we can extract the "Pure QCD" shape of the QCD background by subtracting the other MC backgrounds. Then, assuming this shape holds for the signal region. This is, there is no correlation between the control variable, for example the isolation variable "etcone20", and the variables of interest: transverse mass, missing transverse energy, etc. Finally, scale the pure QCD so that it agrees with the shape in the low mass region of the m_T distribution, where it's assumed to be QCD dominated.

In particular, we could obtain an estimation of the QCD jets by making a requirement that works with the so called isolation variable, "etcone20" which is the sum of energy measurements in the calorimeters in a narrow region around the electron: $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} < 0.2$ (the 0.2 corresponds to the "20" in the variable name). The lepton being isolated refers to the fact that adjacent parts of the detector have not been triggered by other particles. Non-isolated particles come together with others that do trigger the surrounding detectors. This is the case mainly for the particles coming from the QCD jets but is also applicable for the other backgrounds as well. Therefore, if we would like to obtain a "Pure QCD" distribution, we should subtract the other MC backgrounds. However, in order for the pure QCD to not be biased by this MC subtraction, it would also be desirable that this subtraction doesn't account for a big percentage of the total quantity.

So, to create the QCD control region we can require absolute isolation, as we said, or relative isolation, which is the absolute divided by the particle momentum. Different configurations have been studied:

First attempt was done requiring relative isolation, etcone20/p_T >0.04, as well as absolute isolation: etcone20>6GeV. The QCD control regions obtained are shown in figure 16. As said before, it's interesting not to have a big MC backgrounds relevance not to introduce a bias. We see that the MC subtraction becomes sizeable above 100 GeV. Ideally one could make the control region more pure by cutting harder on the relative isolation, but this is not possible due to the preselection in the open data, which already requires etcone20/p_T <0.1.

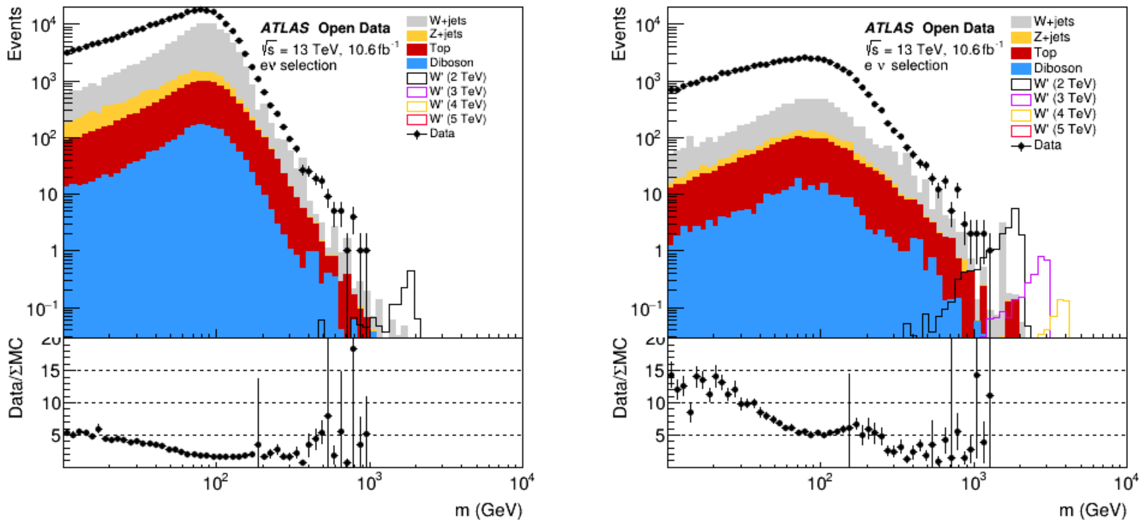

Figure 16: QCD region invariant mass distribution obtained with relative isolation (left) and absolute isolation (right) for the 1 lepton analysis.

From here we can produce the "pure QCD" by subtracting the MC background and scaling the resulting distribution, so that we have a data/MC ratio as close to 1 as possible for low mass in the complete transverse mass distribution, shown in figure 17.

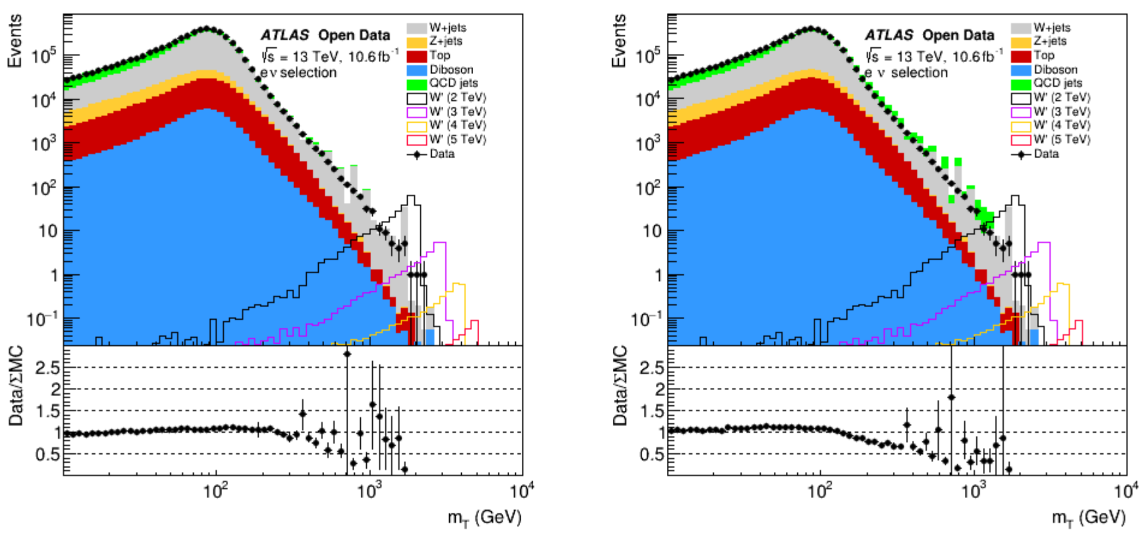

Figure 17: Transverse mass distribution in the 1 lepton analysis including QCD background estimation for relative (left) and absolute (right) isolation approach.

As we can see, the QCD estimate obtained with the relative isolation control region fits the missing background to the data in a very precise way. Both methods agree at low mass, which is the region we want to fit in as accurately as possible. We cannot judge whether or not any of the fits are good on the signal region (high m_T) because that's the area of the discovery, so we should not force data-MC agreement there. In order to discriminate between them, we are going to recreate the above plots but enhancing the QCD contribution. This way, we will see which one of the absolute or relative approaches works better. In order to do this, we are going to work with the "ID level" variable. The tight filter serves to reduce the QCD so, if we look for not tight on all the process there would be more presence of the hadronic jets in general and we could get QCD-enhanced results and a higher purity of the QCD control region and, as a result, less dependence on the MC background due to the subtraction.

One interesting thing to mention is that some pre-selection cuts are usually applied in the early stages of the data processing. This means that the files we're working with have an implicit momentum cut on each of the "ID level" cases. There are three specifications for this variable: "Tight", "Medium" or "Loose". If, as we say, require "not Tight", we will be working with the electrons present in "Medium" and "Loose". However, each one of those has a different momentum cut at the trigger level: around 60 GeV for the "Medium" and around 140 GeV for the "Loose". Therefore, there is a discontinuity and bias around 140 GeV due to the sudden allowance of the "Loose" electrons. We can see it visually in the p_T plots on figures 18 and 19. These figures also show the differences between the relative and absolute isolation approaches in the "not tight" QCD-enhanced environment. We analyse three different variables and not only m_T to have more information available to argue in favour of one particular approach.

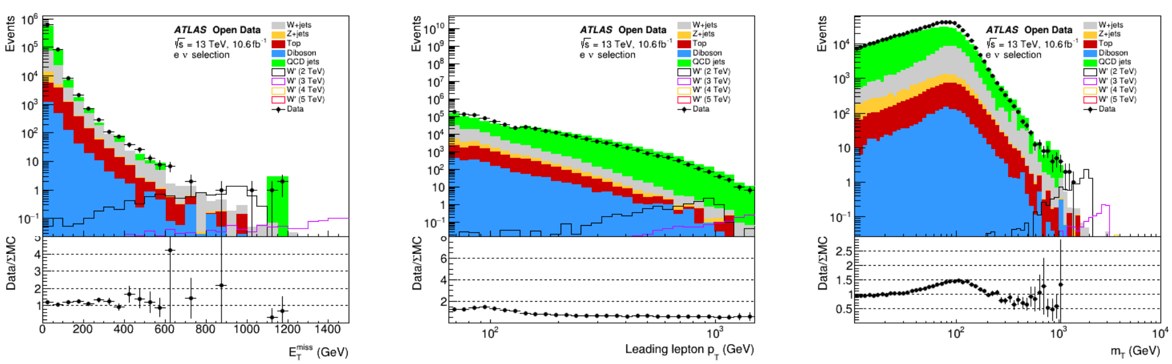

Figure 18: Transverse mass, lepton transverse momentum and missing transverse energy distributions for the 1 lepton analysis with QCD background estimation using relative isolation>0.04 on the QCD region and not tight on the QCD region but also on the normal distribution. Note that the same scale factor obtained for m_T was used for MET and p_T.

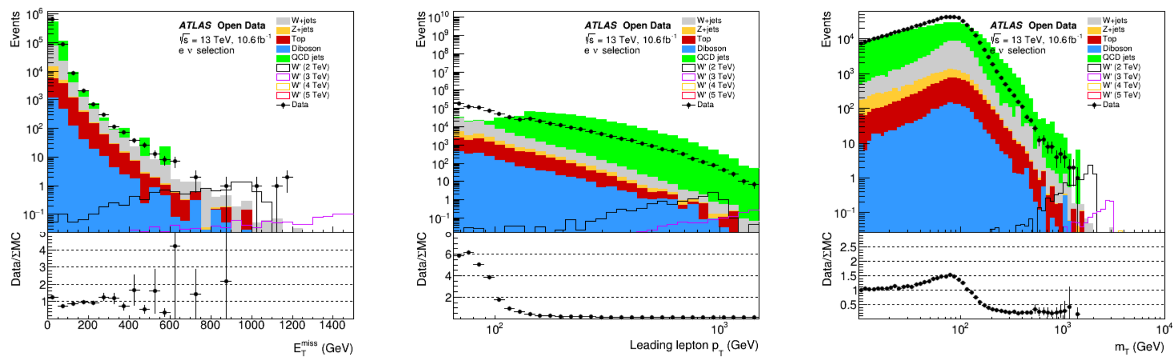

Figure 19: Transverse mass, lepton transverse momentum and missing transverse energy distributions for the 1 lepton analysis with QCD background estimation using absolute isolation>6 GeV on the QCD region and not tight on the QCD region but also on the normal distribution. Note that the same scale factor obtained for m_T was used for MET and p_T.

We can see that the m_T distribution is not perfect in neither of the approaches, but much better at high m_T, which is our primary concern for the search, on the relative isolation method. However, there is a discrepancy around 100 GeV shared by both of them. The other distributions, p_T and MET clearly show, though, that the relative approach describes way better the distribution in the QCD-enhanced regime, with a very decent agreement between data and background, in particular in the transverse momentum distribution. So, the relative isolation approach is reinforced by the results from these other configurations and therefore the final m_T distribution for the 1 lepton analysis is the one with the QCD estimate obtained using relative isolation>0.04, shown in figure 21.

We can also now quickly check that the muon channel didn't have a very important QCD contribution. We can perform the relative isolation>0.04 approach the same way to estimate the possible QCD background contribution. The QCD region obtained is shown in figure 20 and we can see how the data are almost perfectly described by the already present backgrounds, except for the very low mass part. This way, the QCD correction is very small, at the same level as MC and data statistical fluctuations. Accounting for this little contribution, the definitive version of the muon channel transverse mass distribution is also included in figure 21.

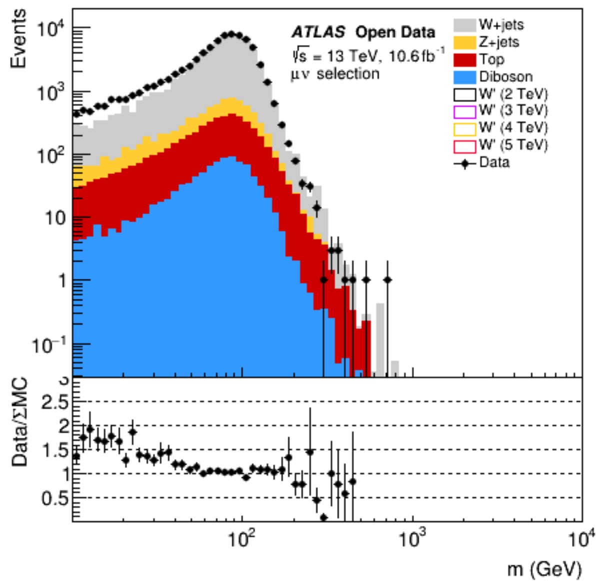

Figure 20: QCD control region for the muon channel. We see how there are very few events that actually correspond to QCD jets since we can describe almost everything with the previous MC backgrounds.

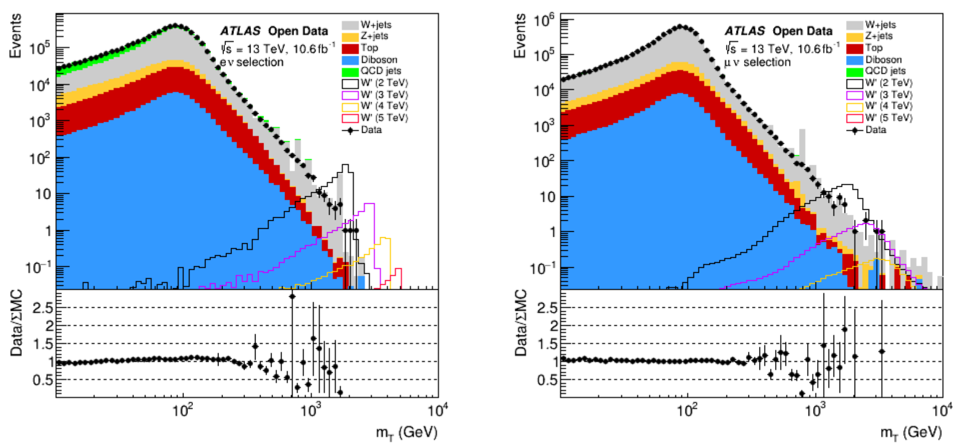

Figure 21: Definitive transverse mass m_T distribution on the electron channel (left) and muon channel (right) in the 1-lepton study, including QCD background estimation with relative isolation > 0.04. Experimental data are shown as black dots; different Standard Model backgrounds contributing to the process, as solid color. New bosons' signals are also introduced, with a long tail that extends from the left to their energy mass value from where it decreases rapidly. The so-called Jacobian peak.

2.3.2 Histogram integration and Statistics

With the final electron channel distribution and the very slightly corrected muon one, we're now ready to follow the same line of work followed for the 2 lepton final state analysis. First, we need to know the optimal cut, maximizing the expected significance. The results of the optimization are shown in figure 22.

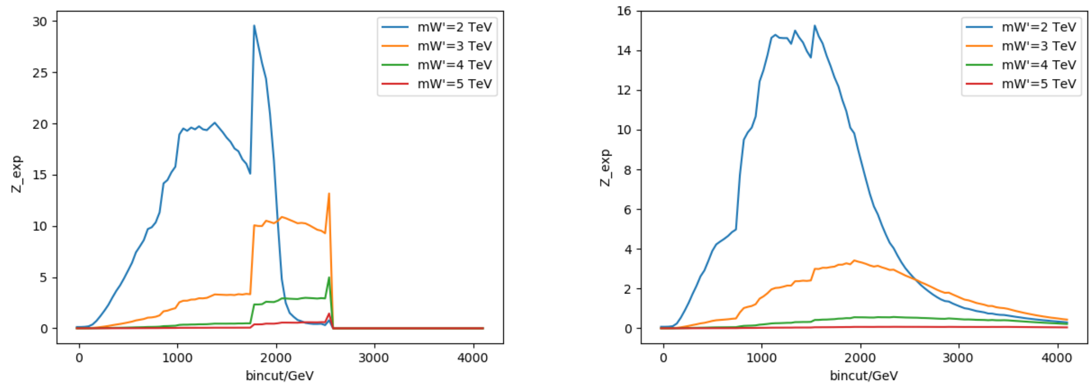

Figure 22: Expected significance optimization for the 1 lepton analysis: The expected significance for a given bin cut of the histogram integration is analysed for each of the respective W' signals. Left-hand side corresponds to electron channel; right-hand side, to the muon channel.

Due to the lack of events at higher masses, there is no information beyond 2600 GeV. Also, the electron plot is not very clear. However, we can see how a sensible approximation of the maximum is around 75% of the mass value in the muon channel for 2 TeV and 3 TeV signals. Then, we assume it would behave the same way for the electron and the other two high mass signals at the muon channel and assume a general criterion of 75% of the mass. A bit lower in the W' case than in the Z' case due to the wider signal distributions for the W' case. Now we can perform the optimal cut, with results from the integration shown in table 6 and 7.

$Wmass$/GeV	Cut/GeV	b/counts	δb/counts	s/counts	n_{obs}/counts	Z_{exp}	Z_{obs}
2000	1500	40	30	165.04	11	29.56	-
3000	2250	0.24	0.14	15.96	1	13.17	1.17
4000	3000	0.0	0.0	1.84	0	-	-
5000	3750	0.0	0.0	0.2	0	-	-

Table 6: Results of the histogram integration for the electron channel in the 1 lepton analysis. For each of the different signals representing a W' boson of mass $Wmass$ the integration has been performed from a bin cut of Cut. Results of the background b, signal s and data n_{obs} are shown. From these results, the expected and observed significances Z_{exp} and Z_{obs} have been estimated by using the results shown in section 1.4.

$Wmass$/GeV	Cut/GeV	b/counts	δb/counts	s/counts	n_{obs}/counts	Z_{exp}	Z_{obs}
2000	1500	23	8	92.5	18	15.24	-
3000	2250	4.4	1.0	7.86	4	3.42	-
4000	3000	2.5	1.0	0.77	2	0.57	-
5000	3750	1.3	1.0	0.07	0	0.07	-

Table 7: Results of the histogram integration for the muon channel in the 1 lepton analysis. For each of the different signals representing a W' boson of mass $Wmass$ the integration has been performed from a bin cut of Cut. Results of the background b, signal s and data n_{obs} are shown. From these results, the expected and observed significances Z_{exp} and Z_{obs} have been estimated by using the results shown in section 1.4.

From these results one can construct the limit plot shown in figure 23 from where to extract the exclusion limits.

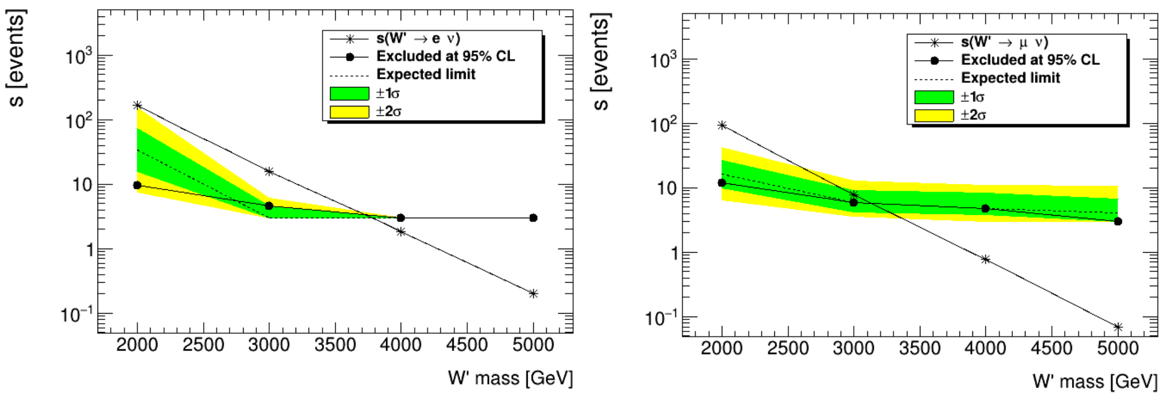

Figure 23: Limit plots for the electron channel (left) and muon channel (right) in the 1 lepton analysis. The crossed points are the ones we got from table 6 and 7. The dotted points represent the value they should have so that we could exclude them at 95% confidence. Expected limit and green/yellow bands represent an ensemble of limits calculated with pseudo-data drawn from the background distribution. We see that we can exclude the 2 and 3 TeV W' signals and that the exclusion limit is set on 3700 GeV for the electron channel and 3100 GeV for the muon one.

2.4 Merging results from the electron and muon channels

As we just said, the electron and muon analysis establish confidence limits on the signals in an independent way. However, one could think of combining the results together such that one had a more powerful argument. That's what we do next.

In order to do that, it's not possible to work with s anymore. We will work with the cross section, which should be the same due to lepton universality. The relation between the number of events and the cross section is $s = L_{int}\epsilon\sigma$, where s is the number of signal events; σ, the cross section for that signal process; ϵ, the signal efficiency and L_{int}, the integrated luminosity.

In this combination, the likelihood becomes a product of the individual Poisson probabilities for the channels, given by

$$L(L_{int}, \sigma, \boldsymbol{\epsilon}_{sig}, \boldsymbol{\sigma}^{eff}_{bkg}) = \prod_{i=1}^{N} \frac{\lambda_i^{n_{obs,i}} e^{-\lambda_i}}{n_{obs,i}!} \tag{14}$$

Where $N = 2$ is the number of channels and $\lambda_i = L_{int}(\sigma\epsilon_i + \sigma^{eff}_{bkg,i})$: expected number of events in channel number i. $\boldsymbol{\epsilon}_{sig}$ is a vector whose components are the signal efficiencies for each of the channels. $\boldsymbol{\sigma}^{eff}_{bkg}$, is a vector whose components are the total background effective cross sections for each of the channels.

Now, we can follow a similar procedure like for the individual channels and estimate the exclusion limits for the signal given a background with a certain uncertainty. The difference is that now, in addition, we need the individual efficiency (and its uncertainty, that we can get from $\delta\epsilon = \frac{\delta s}{L_{int}\sigma}$) of each channel in order to combine them together. The numerical tool present at [13] has also been used in this case to perform this estimation. The efficiencies have been estimated in table 8 and together with the results in tables 2, 3, 6, 7, the exclusion limits shown in figure 24 have been determined. With these new data, we can perform the cross section exclusion algorithm and arrive at the limit plot shown in figure 24.

Channel	Boson m/GeV	s/events	δs/events	ϵ	$\delta\epsilon$
Z'→ee	2000	55.0	0.5	0.587	0.006
	3000	4.64	0.05	0.543	0.006
	4000	0.490	0.005	0.446	0.005
	5000	0.0533	0.0007	0.275	0.004
Z'→ $\mu\mu$	2000	37.83	0.06	0.4019	0.0007
	3000	3.052	0.005	0.3586	0.0006
	4000	0.3046	0.0006	0.2781	0.0005
	5000	0.02803	$7\ 10^{-5}$	0.1142	0.0004
W'→eν	2000	165.0	1.9	0.433	0.005
	3000	15.96	0.18	0.423	0.005
	4000	1.84	0.02	0.371	0.005
	5000	0.204	0.003	0.244	0.004
W'→ $\mu\nu$	2000	92.5	0.2	0.2428	0.0005
	3000	7.863	0.017	0.2112	0.0005
	4000	0.774	0.002	0.1561	0.0004
	5000	0.0685	0.0002	0.0816	0.0003

Table 8: Signal efficiencies, ϵ, and their uncertainties, $\delta\epsilon$, determined from the number of signal events, s with uncertainty δs, for the different Z' and W' boson processes in the respective channels of the 1 lepton and 2 lepton analysis.

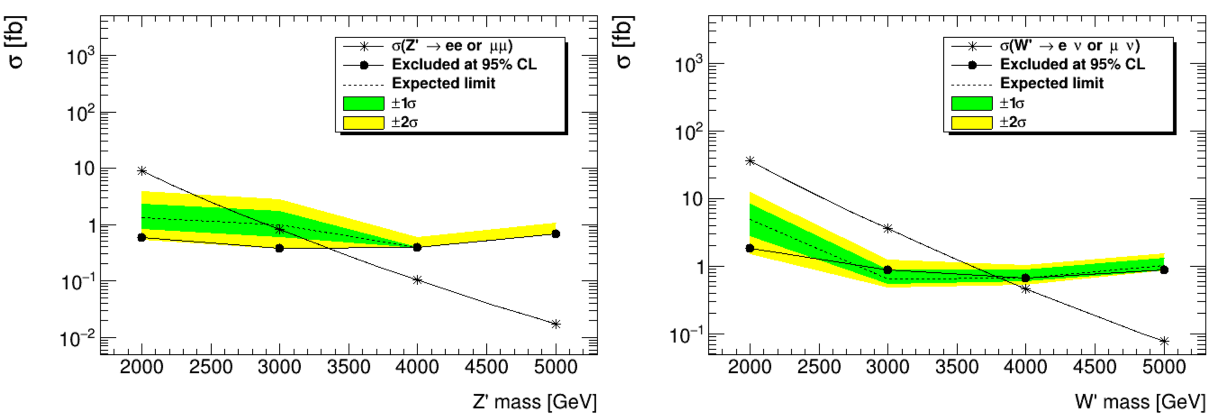

Figure 24: Cross section exclusion limits for the combination of channels on the 2 lepton study (left) and 1 lepton study (right). The line with star shaped points represents the cross section value for the signals of this analysis. The line with the black dots represents the (maximum) limit they can have, so that anything above is excluded. The green and yellow bands represent an ensemble of limits calculated with pseudo-data drawn from the background distribution that represent the effect statistical fluctuations would have on the result. As we can see, 2 and 3 TeV Z' and W' are excluded with more than 95% confidence. Furthermore, a exclusion limit is set on the Z' for masses of 3.3 TeV and, on the W', for masses of 3.7 TeV.

3 Discussion and conclusions

In conclusion, based on the 1 lepton and 2 lepton data study from CERN open data, the 2 and 3 TeV Z' from the E_6-motivated Z'$_\chi$ model and the 2 and 3 TeV narrow-width SSM W' are excluded with more than 95% confidence. The other signals can't be excluded but this procedure has allowed to determine a general exclusion limit for them: 3.3 TeV on the Z' and 3.7 TeV on the W'. This limit could be pushed farther by future experiments like the HL-LHC. No signature of Beyond the Standard Model physics has been observed, in agreement with recent results from [14], [15].

Finally, in order to effectuate a more precise analysis, background fitting techniques have been applied to the 2 lepton case and QCD background data driven estimation has been applied to the 1 lepton study, both successfully. An attempt was done for the background fitting on the 1 lepton analysis but fitting accurately in this case turned out to be more difficult. Regarding the QCD estimation method, a relative isolation approach turned out to be the best one to generate a QCD control region from where to extract the background estimation.

It is also interesting to mention some of the limitations encountered when working with the ATLAS open data:

First, the available MC statistics at high mass is a very important limitation when it comes to producing smooth results. Although this limitation was overcome to some extent in the 2-lepton study (and the same principle could have been applied in the 1-lepton study), it would be much better if the open data had dedicated MC samples covering the high-mass region, as used in the official ATLAS analyses [14] [15].

On the other hand, the application of an isolation cut as preselection for the open data significantly impacts the freedom that is left for performing the QCD estimate. If there was more room to, for example, cut tighter on the relative isolation, we could have reduced the MC subtraction and improved the quality of the pure QCD.

Finally, the unavailability of the medium ID decision greatly reduces the possibility of using "reverse-ID" methods for the QCD estimate because it means that the "not-tight" will always have a discontinuity around $p_T = 140$ GeV due to the selection applied at the trigger level. If we had the medium ID, we could require "isMediumID and not isTightID" which would be a reversed ID definition without a discontinuity.

References

[1] G. Aad, T. Abajyan, B. Abbott, J. Abdallah, S. Abdel Khalek, A. Abdelalim, O. Abdinov, R. Aben, B. Abi, M. Abolins, and et al., "Observation of a new particle in the search for the standard model higgs boson with the ATLAS detector at the LHC," *Physics Letters B*, vol. 716, no. 1, p. 1–29, Sep 2012. [Online]. Available: http://dx.doi.org/10.1016/j.physletb.2012.08.020

[2] S. Chatrchyan, V. Khachatryan, A. Sirunyan, A. Tumasyan, W. Adam, E. Aguilo, T. Bergauer, M. Dragicevic, J. Erö, C. Fabjan, and et al., "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," *Physics Letters B*, vol. 716, no. 1, p. 30–61, Sep 2012. [Online]. Available: http://dx.doi.org/10.1016/j.physletb.2012.08.021

[3] M. Chizhov and G. Dvali, "Origin and phenomenology of weak-doublet spin-1 bosons," *Physics Letters B*, vol. 703, no. 5, p. 593–598, Sep 2011. [Online]. Available: http://dx.doi.org/10.1016/j.physletb.2011.08.056

[4] G. Senjanovic and R. N. Mohapatra, "Exact left-right symmetry and spontaneous violation of parity," *Phys. Rev. D*, vol. 12, pp. 1502–1505, Sep 1975. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.12.1502

[5] D. London and J. L. Rosner, "Extra gauge bosons in E_6," *Phys. Rev. D*, vol. 34, pp. 1530–1546, Sep 1986. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.34.1530

[6] P. Langacker, "The physics of heavy Z ' gauge bosons," *Reviews of Modern Physics*, vol. 81, no. 3, p. 1199–1228, Aug 2009. [Online]. Available: http://dx.doi.org/10.1103/RevModPhys.81.1199

[7] E. Salvioni, G. Villadoro, and F. Zwirner, "Minimal Z' models: present bounds and early LHC reach," *Journal of High Energy Physics*, vol. 2009, no. 11, p. 068–068, Nov 2009. [Online]. Available: http://dx.doi.org/10.1088/1126-6708/2009/11/068

[8] G. Altarelli, B. Mele, and M. Ruiz-Altaba, "Searching for new heavy vector bosons in $p\bar{p}$ colliders," *Zeitschrift für Physik C Particles and Fields*, vol. 45, no. 1, pp. 109–121, Mar 1989. [Online]. Available: https://doi.org/10.1007/BF01556677

[9] "ATLAS experiment website," available at https://atlas.cern/.

[10] "ATLAS open data portal 2 lepton sample," available at opendata.cern.ch/15003.

[11] "ATLAS open data portal 1 lepton sample," available at opendata.cern.ch/15001.

[12] G. Aad, B. Abbott, D. Abbott, O. Abdinov, A. Abed Abud, K. Abeling, D. Abhayasinghe, S. Abidi, O. AbouZeid, N. Abraham, and et al., "Search for high-mass dilepton resonances using 139 fb^{-1} of pp collision data collected at $\sqrt{s}$=13 TeV with the ATLAS detector," *Physics Letters B*, vol. 796, p. 68–87, Sep 2019. [Online]. Available: http://dx.doi.org/10.1016/j.physletb.2019.07.016

[13] "FYS5555 GitHub repository," available at https://github.com/Etienne357/FYS5555/tree/master/Statistics.

[14] M. Aaboud, G. Aad, B. Abbott, O. Abdinov, B. Abeloos, S. H. Abidi, O. S. AbouZeid, N. L. Abraham, H. Abramowicz, and et al., "Search for new high-mass phenomena in the dilepton final state using 36 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV with the ATLAS detector," *Journal of High Energy Physics*, vol. 2017, no. 10, Oct 2017. [Online]. Available: http://dx.doi.org/10.1007/JHEP10(2017)182

[15] M. Aaboud, G. Aad, B. Abbott, O. Abdinov, B. Abeloos, S. H. Abidi, O. S. AbouZeid, N. L. Abraham, and H. Abramowicz, "Search for a new heavy gauge-boson resonance decaying into a lepton and missing transverse momentum in 36 fb^{-1} of pp collisions at $\sqrt{s} = 13$TeV with the ATLAS experiment," *The European Physical Journal C*, vol. 78, no. 5, May 2018. [Online]. Available: http://dx.doi.org/10.1140/epjc/s10052-018-5877-y