Alma Mater Studiorum · Università di Bologna

Scuola di Scienze
Dipartimento di Fisica e Astronomia
Corso di Laurea in Fisica

Measurement of the Z and W production cross section in pp collisions at LHC using a bayesian approach

Relatore:
Dott. Lorenzo Bellagamba

Presentata da:
Simone Ragoni

Anno Accademico 2015/2016

Contents

Abstract

The aim of all my work has been to compute the fiducial production cross sections of $W^{\pm}$ and Z^0 bosons in their leptonic (e, μ) decays using the data collected by the ATLAS detector at LHC with a center of mass energy of $\sqrt{s} = 13$ TeV during summer 2015. The selected events are exactly the same as the ones employed by the recently published article by the ATLAS Collaboration over the same topic, enabling us to compare the obtained results. Necessary comparison, if I may, for the results were obtained with two different procedures: baseline (classical) for the article, bayesian in this thesis. The bayesian approach allows for a natural combination among the many channels and a straightforward treatment of the systematic uncertainties. The obtained results are in excellent agreement with the Standard Model predictions and those published by ATLAS.

Abstract

L'obiettivo di tutto il mio lavoro è stato quello di misurare le sezioni d'urto di produzione dei bosoni deboli $W^{\pm}$ e Z^0 nei loro decadimenti leptonici (e, μ) coi dati raccolti dal rivelatore ATLAS a LHC con un'energia del centro di massa di $\sqrt{s} = 13$ TeV relativi all'estate 2015. Gli eventi selezionati sono gli stessi di quelli del recente articolo della Collaborazione ATLAS sullo stesso argomento, in modo anche da poter operare un confronto tra i risultati ottenuti. Confronto peraltro necessario, poichè i risultati sono stati ottenuti con due metodologie differenti: tradizionale (classica) per l'articolo, bayesiana in questa tesi. L'approccio bayesiano permette di combinare i vari canali e di trattare gli effetti sistematici in modo del tutto naturale. I risultati ottenuti sono in ottimo accordo con le predizioni dello Standard Model e con quelli pubblicati da ATLAS.

Introduction

I have always been interested in particle physics, so when my supervisor actually offered to me this thesis work I gladly accepted it.

The aim of this script of mine is to compute the production cross sections for the weak bosons using the first proton-proton collisions data provided by the Large Hadron Collider (LHC) at a center of mass energy $\sqrt{s} = 13$ TeV and collected by the ATLAS detector in summer 2015. The ATLAS Collaboration has recently published the W and Z production cross section based on the first $\sqrt{s} = 13$ TeV data provided by LHC after the energy upgrade. In this thesis I follow the same approach employed in a recent published ATLAS paper for the event selection and the background evaluation but use a complete different method for the extraction of the final cross-section results. In particular, instead of a classical statistical procedure, implemented in the *HERAverager* tool [2, 3], referenced in the following as *baseline* method, I exploit the features of the bayesian approach which allows a natural combination of the different channels and a straightforward inclusion of the systematic uncertainty effects.

It has been a long and hard work. At the very beginning, I had to study how bayesian inference worked: did not know beans about it. Thankfully there were online a few notes by prof. Giulio D'Agostini [1], they helped a lot.

The software my supervisor employed to perform bayesian analysis is a C++ code, BAT (Bayesian Analysis Toolkit), so the next step was to learn how to use it (how to write codes, learn its hierarchy system, a few touches here and there and we are good to go).

We wanted to make for BAT a new way to input data at runtime, without having to actually put our hands into its code, so a function to read data from the outside was necessary together with a way to present them inside the external file. My supervisor saw inside the XML format a good way to go

around the problem, using its markup system, and I must say the XML files that came from this work were actually really beautiful and user-friendly. Still, I had to first study the XML language (at the ultra-novice level) and think about a way to make XML and C++ interact. CMarkup, an XML parser [6], descended from the sky to help the man in its adventure.

Finally I was at the starting point. After many trials the analysis was made, but many questions arose from it, questions needing an answer. The solutions to the many problems were found little by little, and yet a few things were left in oblivion.

Again the way was sought into coding, specifically R [28] codes. My supervisor often uses it in order to make statistical analysis over great amount of data. After learning it I could actually make a few steps on my own.

One thing I noticed is that the analysis required to solve every problem was really long. Sometimes it could take several days, before even checking that everything done until that point was even right. So I told my supervisor that something needed to be done about it. After some discussions also involving the BAT authors we proposed a possible solution to improve the efficiency and the performances of the code. We are still working over it, together with the BAT management, but I guess that by the time of the discussion of this thesis work we still won't be done, still I wanted to convey to you all my feelings of hardships, but even satisfactory ones.

I guess that it is time to pass over to the real part of the thesis. Running over it you will eventually see that it represents everything I have done these last months.

In the first part you will see an introduction to bayesian inference, after that a description of LHC and the ATLAS detector, the fundamental tools which provide the data for the analysis. The following sections are about the selection of the data and the evaluation of the backgrounds for the different channels, and finally the presentation and discussion of the results. There are a few really technical parts, that were too heavy to be put inside the main body of this script, and you will find them inside the appendices, one on BAT and our codes, one on R, and another one over my work in summer 2015 at ATLAS about performance studies on the Insertable B-Layer (IBL), a new layer of the pixel tracking detector installed during the LHC shutdown for the energy upgrade.

Before leaving I want to express my great feelings of joy for this period of

hard but fulfilling work, under the aegis of a caring and patient supervisor, Dott. Lorenzo Bellagamba, all the ATLAS group (both Bologna and Cern groups), surrounded by a loving family, that supported me in these long three years away from home.

Part I

Bayesian approach

Chapter 1

Bayes' theorem

The basic concept of probability is measuring the degree of belief that an event will occur. But this doesn't tell us how to obtain the cherished numerical value.

1.1 Subjective probability?

Let us toss a coin. If we want to compute the probability of obtaining head or tails, we usually reason like:

A) the possible outcomes are head or tails (ideal coin with zero width), so we expect to obtain tails half of the times;

B) calling *head* H and *tails* T, tossing many times the coin we obtain HTHHHTTTHHTH... and if we go on like this we expect to obtain tails roughly half of the times.

These are seemingly "two sides of the same coin", but actually they represent two different ways of thinking about probability. The former is called *classical* definition of probability, where you compute the ratio of the wanted outcomes out of all the possible, while the latter is usually called *frequentist* definition, where you compute the actual frequency of an event happening and you suppose that the event will continue to happen with the same law even in the future.

Until now all fine and dandy, but let us delve further in their actual meaning. The former lacks the clause "if all the possible outcomes are equally probable". That is rather peculiar, at least. That is because it employs probability in the very definition of probability... So this approach actually only tell us a useful way of *qualitatively* evaluate probability, without telling us how to *quantitatively* to do.

The latter instead would need (to be picky) the condition of infinite repetitions, but we can just skip on it. The actual point is, as the reader would have already guessed, that no one actually tell us that the frequency of the outcome in the future will be the same as until now [1].

So we have to really define what is probability. At the start of these lucubrations, I have quoted the statement that probability is the measure of the degree of belief that an event will occur. This represents the subjective definition of probability, and the tool that allows us to develop computations is the *Bayes' theorem*.

1.2 Bayes' theorem: discrete case

It is much more natural to always start talking about probability in the discrete case, because it is the one we are most familiar with.

In the bayesian approach, it is of utmost importance to define the so called "conditional probability".

Let's consider the following example:

> In my university there are in all 1000 people. 400 of them are males, the others women of every age. Half of the males have to put on glasses, while only a hundred females have to do so. Suppose that people have at most one pair of glasses each, and none if they don't need them. Suppose you spot an abandoned pair of glasses far away and you ask yourself what is the gender of the unlucky owner.

Then you think that there are 600 females, 100 with glasses, 400 males, 200 of which shortsighted. So you say that you expect to find a female owner with a probability of $100/300 = \frac{1}{3}$.

What you have done is actually:

$$P(Female|g) = \frac{P(g|Female) \cdot P(Female)}{P(g|Female) \cdot P(Female) + P(g|Male) \cdot P(Male)}$$

Where with the notation $P(Female|g)$ we mean the conditional probability and we read "the probability that a person wearing glasses is a female". This is an instance of the Bayes' Theorem:

$$P(H_i|E) = \frac{P(E|H_i) \cdot P(H_i)}{\sum P(E|H_j) \cdot P(H_j)} \tag{1.1}$$

where E is an event and the H_i are a *complete* set of mutually excluding hypothesis. This is the start of everything. In the following all the functions in the form $f(E|H_i)$ will be called `likelihoods`, because they represent the way the event is influenced by the H_i, and the functions in the form $f(H_i)$ will be called `priors`, because they represent the probability distribution that would express one's beliefs about the quantity before some evidence is taken into account.

1.3 Monty Hall's problem

This is maybe the most famous example of bayesian reasoning. The first instance of solution for the Monty Hall problem came from Steve Selvin, solution that didn't even nearly achieve the popularity it had after being printed and solved by Marylin vos Savant in 1990.

> Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?[1]

Both Steve and Marylin stated it was for the best to always change, because you would achieve a probability of $\frac{2}{3}$ of winning the car. This statement was

[1] In the film "21" it is explained well enough. I highly recommend to watch it. The solution of the Monty Hall problem by the protagonist, constitutes the turning point of the film.

considered preposterous by a lot of scholars so that the magazine Marylin worked for was stormed by nearly 10^4 letters! Many Ph.D. wrote as well, but their comments were as to say offensive:

> Dear Marilyn, [...] Since you seem to enjoy coming straight to the point, I'll do the same. You blew it! Let me explain. If one door is shown to be a loser, that information changes the probability of either remaining choice, neither of which has any reason to be more likely, to 1/2. As a professional mathematician, I'm very concerned with the general public's lack of mathematical skills. Please help by confessing your error and in the future being more careful.
>
> Robert Sachs, Ph. D., Gorge Mason University [7]

The others were all along the lines of this one, we could even say that the previous one was even polite...

> I am sure you will receive many letters on this topic from high school and college students. Perhaps you should keep a few addresses for help with future columns.
>
> W. Robert Smith, Ph.D. Georgia State University

Needless to say that the solution stated by Steve and Marylin was the correct one. The reasoning is the following: if you had chosen the door with the car, the host will choose one of the other so that by switching you will have obtained the goat. Instead by choosing one of the doors with the goat you will force the host to choose the other with the goat. In the end, by switching you would achieve a probability of $\frac{2}{3}$ of winning the car.

I will try to write two equivalent ways of solving it (please note that in the following h = door opened by the *host*, c = door with the *car*, p = door chosen by the *player*).

a) You must switch and you have three possibilities: CAR⇒GOAT, GOAT1⇒CAR, GOAT2⇒CAR, so the probability of winning is
$0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = \frac{2}{3}$

b) Let's suppose a flat prior for the car:

$$P(c|h,p) = \frac{P(h|c,p) \cdot P(c|p)}{P(h|p)}$$

where $P(h|c,p)$ is the probability of the host opening a certain door, given that the car is behind door c and the player has chosen door p ($P(h|c,p) = 0(h = p, c), \frac{1}{2}(p = c \neq h), 1(p \neq c \neq h)$) , always $P(c|p) = \frac{1}{3}$ (position of the car given the choice of the player), and

$$P(h|p) = \sum_{c=1}^{3} P(h|c,p) \cdot P(c|p)$$

so that we can write, without losing general validity [8]:

$$P(c = 2|p = 1, h = 3) = \frac{1 \cdot \frac{1}{3}}{1 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3}} = \frac{2}{3}$$

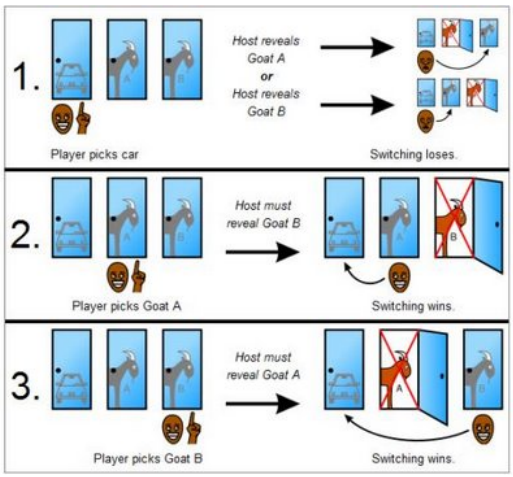

Figure 1.1: Monty Hall's problem.

You could even consider the version where you know that the host does instead not know where the car is. In this way you have:
$P(h|c,p) = 0(h = p, c), \frac{1}{3}(h = c), \frac{1}{2}(p = c \neq h), \frac{1}{2}(p \neq c \neq h)$, so that

$$P(c = 2|p = 1, h = 3) = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3}} = \frac{2}{3} = \frac{1}{2}.$$

And in the most general case where you don't know whether the host knows we just have to consider a new parameter $k = 0(hostknows), 1(doesnot)$ and giving to it a flat prior so that by Bayes':

$$P(c = 2|p = 1, h = 3) = \frac{\frac{1}{2} \cdot 1 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2} \cdot [1 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3}] + \frac{1}{2} \cdot [\frac{1}{2} \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3}]} = \frac{3}{5}.$$

So it is always for the best to switch to the other door.

1.4 Bayes' theorem: continuum case

It is fairly easy to generalize the theorem to the continuum case [1].

$$f(H|E) = \frac{f(E|H) \cdot f(H)}{\int f(E|H) \cdot f(H) dH} \tag{1.2}$$

We just have to assume the set of different hypothesis to become continuous values so that the sum may be replaced by an integral. This is the formula we will use to measure the cross-sections. What we will do is to consider $\vec{x} = \{x_1, x_2, ...\}$ the vector of observables, $\vec{\mu} = \{\mu_1, \mu_2, ...\}$ the vector of the physical quantities we want to measure, $\vec{h} = \{h_1, h_2, ...\}$ the vector of all the possible realizations of the influence factors (for instance the systematics and all the other possible variables which could influence the result), so that you may consider the $\vec{x}$ as the event in action and the $\vec{\mu}$ as the hypothesis you want to check.

The likelihood of the sample $\vec{x}$ being produced from $\vec{\mu}$ and $\vec{h}$ and the prior are respectively:

$$f(\vec{x}\,|\vec{\mu}\,,\vec{h}\,) \tag{1.3}$$

$$f_0(\vec{\mu}\,,\vec{h}\,). \tag{1.4}$$

In the end we obtain:

$$f(\vec{\mu}\,,\vec{h}\,|\vec{x}\,) = \frac{f(\vec{x}\,|\vec{\mu}\,,\vec{h}\,) \cdot f_0(\vec{\mu}\,,\vec{h}\,)}{\int f(\vec{x}\,|\vec{\mu}\,,\vec{h}\,) \cdot f_0(\vec{\mu}\,,\vec{h}\,) d\vec{\mu}\, d\vec{h}}. \tag{1.5}$$

Most of the times we want to eliminate the dependence of the posterior from $\vec{h}$, so what we need to do is to integrate the posterior with respect to $d\vec{h}$. This process is called *marginalization*.

1.5 Monte Carlo Methods to solve the Bayes' formula

In order to evaluate the *posterior probability density function* (p.d.f.) applying the Bayes' theorem, we will employ Markov Chain Monte Carlo (MCMC) technique, in particular Random Walk MCMC. A Markov Chain is a random process that undergoes transitions from one state to another on a state

space. It must possess a property that is usually characterized as "memorylessness": the probability distribution of the next state depends only on the current state and not on the sequence of events that preceded it.

When an MCMC method is used for approximating a multi-dimensional integral, an ensemble of walkers move around randomly. At each point where a walker steps, the integrand value at that point is counted towards the integral. The walker then may make a number of tentative steps around the area, looking for a place with a reasonably high contribution to the integral to move into next.

We will focus on the Metropolis algorithm (Wikipedia does a decent job on it [9]), an MCMC method for obtaining a sequence of random samples from a probability distribution for which direct sampling is difficult. It works by generating a sequence of sample values in such a way that, as more and more sample values are produced, the distribution of values more closely approximates the desired distribution, $P(x)$. These sample values are produced iteratively, with the distribution of the next sample being dependent only on the current sample value (thus making the sequence of samples a Markov chain). At each iteration, the algorithm picks a candidate for the next sample value based on the current sample value. Then, with some probability, the candidate is either accepted (in which case the candidate value is used in the next iteration) or rejected (in which case the candidate value is discarded, and current value is reused in the next iteration). Going a bit further the procedure is as follows:

a) let $f(x)$ be a function that is proportional to the desired probability distribution $P(x)$ (a.k.a. a target distribution);

b) *initialization*: choose an arbitrary point x_0 to be the first sample, and choose an arbitrary probability density $Q(x|y)$ which suggests a candidate for the next sample value x, given the previous sample value y. A powerful choice for the proposal function (Q) is the Breit-Wigner that makes it possible to explore even the tails.

c) *iteration t*: generate a candidate x_{t+1} for the next sample by picking from the distribution $Q(x_{t+1}|x_t)$. Calculate the acceptance ratio $\alpha = f(x_{t+1})/f(x_t)$, which will be used to decide whether to accept or reject the candidate. Because f is proportional to the density of P, we have that $\alpha = f(x_{t+1})/f(x_t) = P(x_{t+1})/P(x_t)$. If $\alpha \geq 1$, the candidate is more likely than x_t, and it is automatically accepted. Otherwise the

candidate is accepted with probability α. If the candidate is rejected, set $x_{t+1} = x_t$, instead.

This algorithm proceeds by randomly attempting to move about the sample space, sometimes accepting the moves and sometimes remaining in place. Note that the acceptance ratio α indicates how probable the new proposed sample is with respect to the current sample, according to the distribution $P(x)$. If we attempt to move to a point that is more probable than the existing point (i.e. a point in a higher-density region of $P(x)$), we will always accept the move. However, if we attempt to move to a less probable point, we will sometimes reject the move, and the more the relative drop in probability, the more likely we are to reject the new point. Thus, we will tend to stay in (and return large numbers of samples from) high-density regions of $P(x)$, while only occasionally visiting low-density regions. Intuitively, this is why this algorithm works, and returns samples that follow the desired distribution $P(x)$.

Part II

LHC and the ATLAS detector

Chapter 2

LHC and the ATLAS detector

In this chapter we want to describe the LHC facility, and the ATLAS detector which provides and analyzes the proton-proton collisions used in our analysis.

2.1 LHC

If we go from Geneva (CH, Switzerland) towards the northern French border, we come across the CERN (European Organization for Nuclear Research), the largest research centre in the world, with the greatest and most powerful collider of all times, LHC (figure 2.1). Most people tend to think that LHC is the whole accelerating apparatus. It is actually only its final component. By "only", we mean a 27 km ring, 100 m underground, kept at the coldest temperature achieved at a so large scale of something like 2 K. This is already amazing. But you have to think that its actual planned center of mass energy ($\sqrt{s}$) for the colliding beams is 14 TeV, where until now we have just hit 13 TeV with an instant peak luminosity of 10^{34} cm^{-2}s^{-1}. We define as *luminosity* L the quantity that multiplied by the cross-section σ gives the event rate as in:

$$\frac{dN}{dt} = \sigma \cdot L.$$

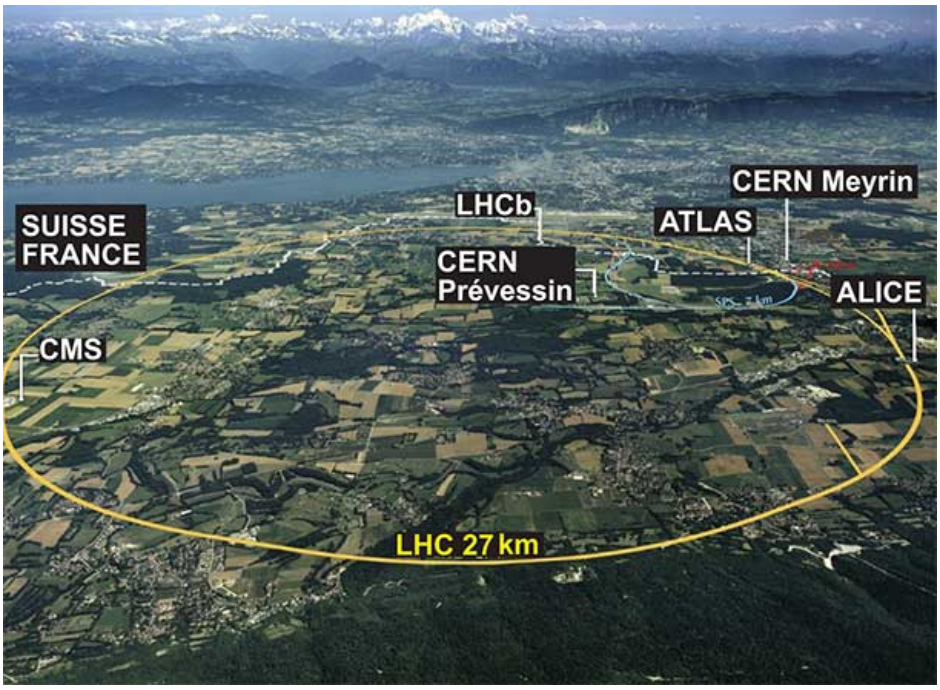

Figure 2.1: Aerial view of the area where the LHC ring is located.

We can even define the *integrated luminosity* L_{int} as[1]:

$$L_{int} = \int L dt.$$

If we walk alongside LHC, we find four caverns, containing the four detectors operating at LHC: LHCb (its main aim is to study the problem of CP violation), CMS (Compact Muon Solenoid, multipurpose detector with same goals of ATLAS but different technologies), ALICE (A Large Ion Collider Experiment, whose focus lies in the study of quark-gluon plasma formed in heavy nucleons collisions), and finally ATLAS (A Toroidal LHC Apparatus), whose data will be later used to study the W and Z boson production in pp interactions.

[1] All collider experiments aim to maximize their integrated luminosities, as the higher the integrated luminosity, the more data is available to analyze. It is useful to note that L is explicitly given by geometrical factors of the proton bunches.

2.2 The ATLAS detector

ATLAS is a multipurpose detector [10] which works mainly at high luminosity to discover signature of new physics and provides precise measurements of Standard Model particles' features. With its total length of 42 m, 11 m radius and weighing 7000 tons, it is the largest LHC experiment (figure 2.2). Its main organization is, from inside out:

- Inner Detector;
- Calorimetric system;
- Muon Spectrometer.

The particle trajectory and the momentum of the charged particles are evaluated by the inner detector, through the bending given by the magnetic field of roughly 2 T generated by a superconducting solenoid, while the energy and the particle identification are provided by the calorimeters and the momentum of the muons is measured in the muon spectrometer.

The ATLAS standard coordinate system is right-handed, the beam direction corresponds to the z axis while the x-y plane is on the plane transverse to the beam pipe with x pointing towards the center of LHC and the origin is located at the nominal interaction point. It is useful to employ polar coordinates: the azimuth angle (ϕ) is measured around the beam axis while the polar angle (θ) is the angle a particle forms with respect to the beam axis. The θ coordinate is usually substituted with the *pseudorapidity* defined as:

$$\eta = -\ln(\tan\frac{\theta}{2}).$$

2.2.1 The inner detector

The inner detector (figure 2.3) is embedded within a 2 T thin superconducting solenoid magnet of 2.5 m diameter and it is composed by a silicon pixel, a micro-strip detector (SCT) and a straw-tube transition radiation tracker (TRT). The precision tracking detector, formed by the pixel detector and SCT, is arranged on concentric cylinders around the beam axis while in the end-cap regions consists of disks perpendicular to the beam axis. The inner

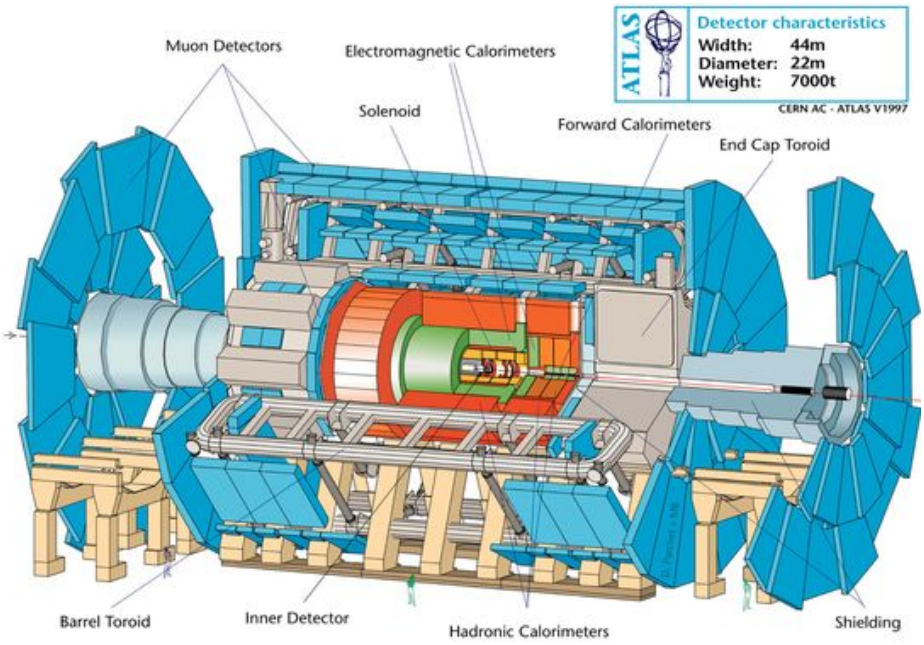

Figure 2.2: The ATLAS detector with its main components.

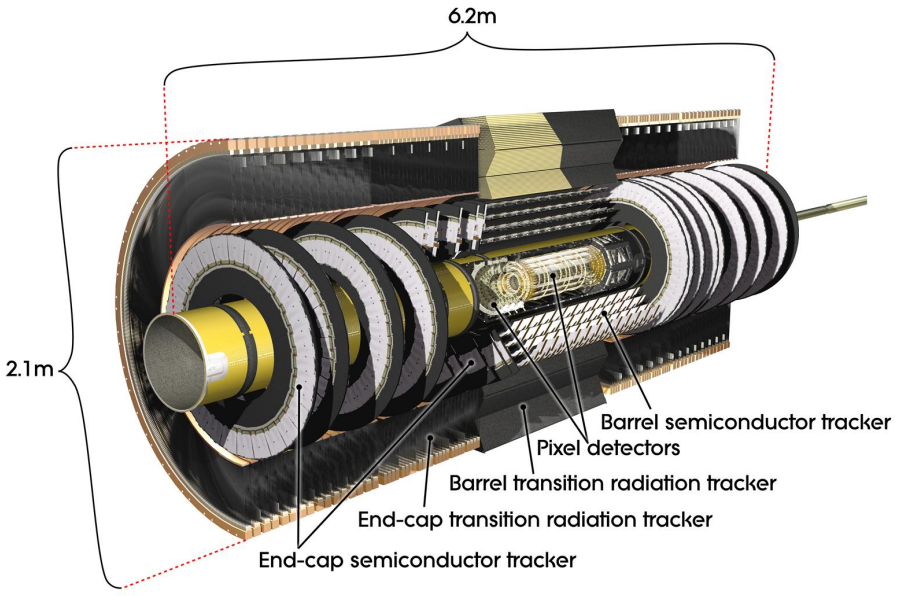

Figure 2.3: ATLAS inner detector.

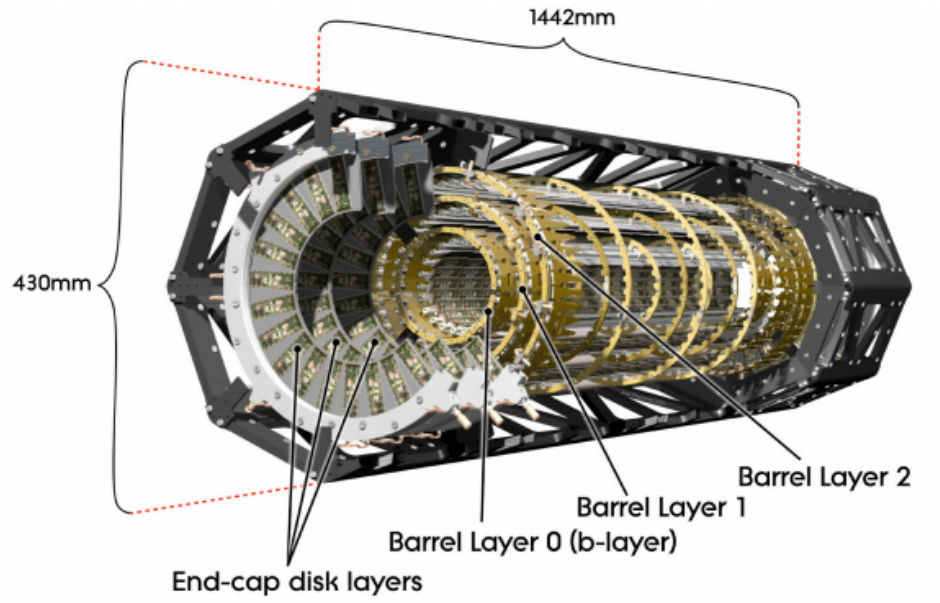

Figure 2.4: ATLAS old pixel detector. The recently installed IBL is missing in this layout.

detector provides tracking and vertexing capabilities within $|\eta| < 2.5$, as well as electron identification thanks to the TRT within $|\eta| < 2.0$.

The pixel detector (figure 2.4) is the nearest to the collision point and measures the particle impact parameters and the decay vertexes of short living particles. In the old days it consisted of merely three layers (b-layer, layer 1, layer 2), but it has recently been updated with the new IBL[2] (Insertable B-layer, figure 2.5), comprised of two different type of sensors: Planar and 3D (see fig. 2.6).

During summer 2015, I have spent a period of roughly 2 months working in ATLAS for the pixel detector group, doing performance studies over the Lorentz Angle (for further informations Appendix B).

The SCT (Semi Conductor Tracker) system is designed to provide track precision measurements and contribute to the measurement of momentum, impact parameter and vertex position in the intermediate radial range.

The TRT combines drift tube chamber tracking capabilities with transition

[2]The performance of the b-layer is in constant decline because of radiation damage. The IBL was designed and installed in order to prevent a decrease in efficiency and performances of the inner detector with the increase of the integrated luminosity.

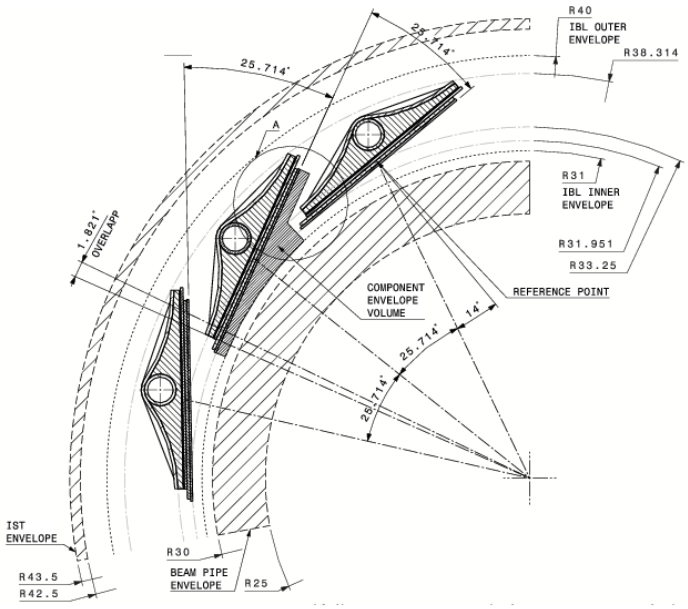

Figure 2.5: IBL design.

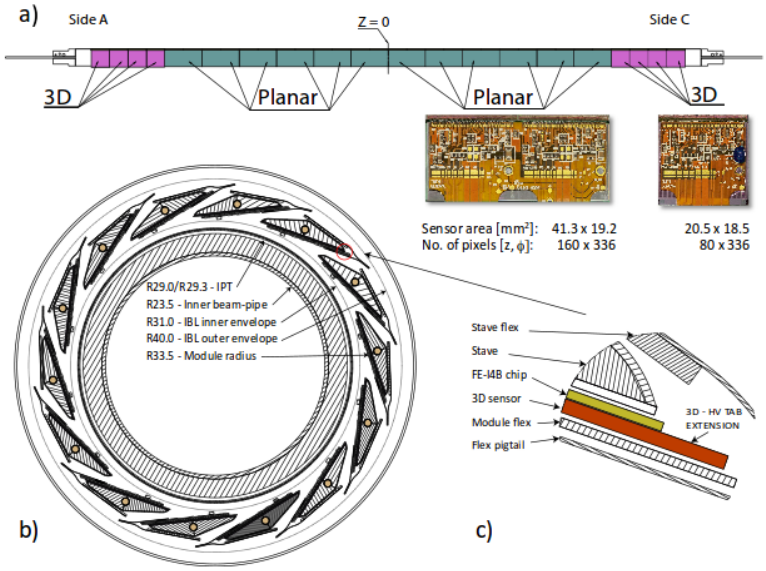

Figure 2.6: (a) Stave layout with the organization of planar and 3D sensor modules. (b) Layout of the IBL detector with the 14 staves around the IBL positioning tube (IPT) and (c) zoom of one stave side where a 3D sensor module is visibile.

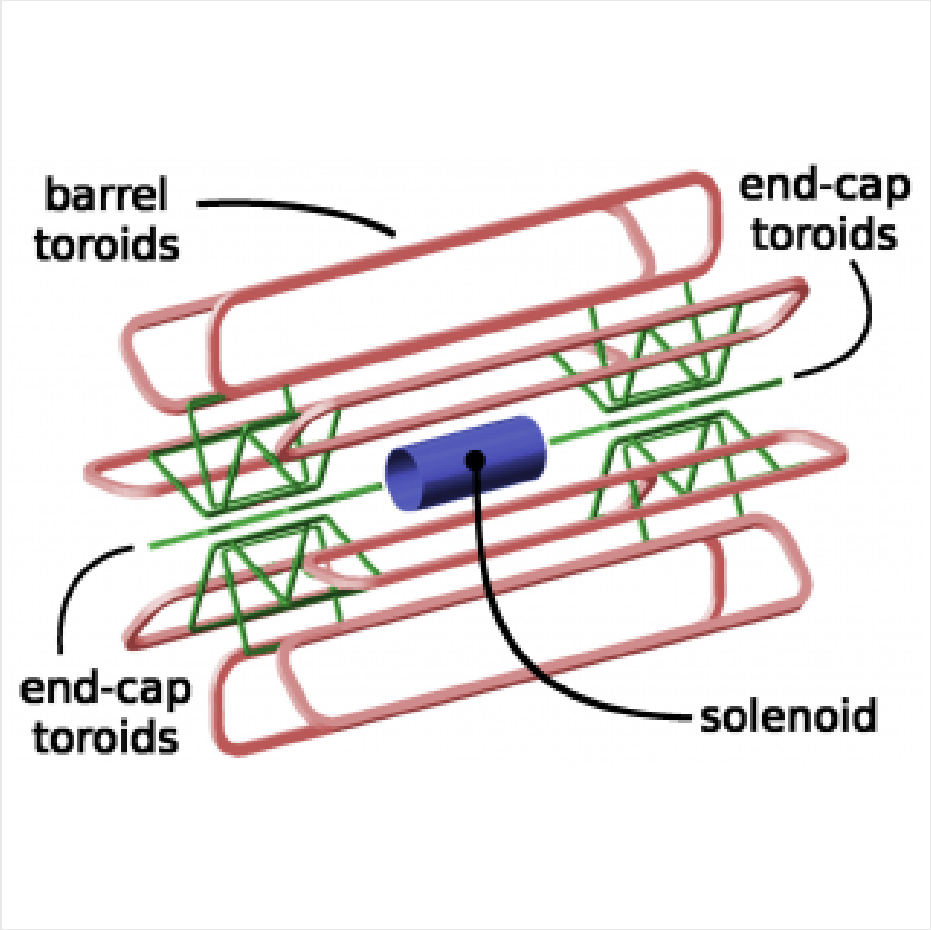

Figure 2.7: Magnetic system design.

radiation detector features in electron/pion discrimination.

2.2.2 The magnet

ATLAS has a system of four large superconducting magnets that provide the magnetic field required for the momentum reconstruction of the many particles that are produced in the collision process. The system (figure 2.7) is made up by:

- the solenoid, aligned with the beam axis, providing a 2T axial magnetic field for the inner detector;

- the large super-conducting air-core toroid system, constituted by eight Barrel Toroids 25m long, and two End-Cap Toroids (5m long).

Such system provides the toroidal magnetic field for the muon spectrometer.

2.2.3 Calorimetric System

The calorimeter is designed to trigger and to provide precision measurements of electrons, photons, hadron jets, and missing transverse energy. It is comprised of an electromagnetic (EM) and a hadronic calorimeter. The system covers the range $|\eta| < 4.9$ using different techniques, suited to the widely varying requirements of the physics processes of interest and of the radiation environment. The finer granularity is reached in the EM, over the region matched to the inner detector, and is ideally suited for precision measurements of electrons and photons. The jet measurements are performed in the hadronic calorimeter with less granularity.

The foremost design feature of the calorimeters is their depth (they must both provide good containment for particles showers and limit punch-through in the muon system).

2.2.4 Muon Spectrometer

The outer part of the ATLAS detector constitutes the muon spectrometer, designed to detect muons going out of the calorimeters and to measure their momentum in the pseudorapidity range $|\eta| < 2.7$. It even triggers on said particles in the range $|\eta| < 2.4$. The momentum is measured thanks to the deflection provided by the barrel and end-cap magnet toroid. In the barrel region,tracks are measured in chambers arranged in three cylindrical layers around the beam axis; in the transition and end-cap regions, the chambers are installed in planes perpendicular to the beam, in three layers yet again.

Part III

Analysis

Chapter 3

Analysis setup, statistical procedure

Our aim is to measure the production cross-sections for $W^{\pm}$ and Z^0 in pp collisions at $\sqrt{s} = 13$ TeV. These results represent the first measurements of SM processes performed by the ATLAS Collaboration at 13 TeV.

3.1 Channels and fiducial phase spaces

Measurements of the $W^{\pm} \to l^{\pm}\nu$ and $Z \to l^{+}l^{-}$ production cross-sections (where $l^{\pm} = e^{\pm}, \mu^{\pm}$) in pp collisions at $\sqrt{s} = 13$ TeV are presented using data recorded by ATLAS corresponding to a total integrated luminosity of 81 pb^{-1}. Measurements of electroweak bosons' production provide a benchmark for the understanding of quantum chromodynamics (QCD) and electroweak (EW) processes. All the data were collected during the period of June 13 to July 16, 2015. The results have been recently published [15] using a classical statistical approach to evaluate the production cross-sections. In this chapter, after summarizing the event selection, the main sources of backgrounds and the systematic uncertainties, we'll describe the alternative statistical analysis we used to process the data.

The total cross-section for the $W^{\pm}$ boson times the branching ratio for the decays into a single-lepton may be expressed as a ratio of the numbers of background subtracted data events N to the product of the integrated luminosity of the data L_{int}, an acceptance factor A and a correction factor

C:

$$\sigma^{tot}_{W^{\pm}} = \frac{N}{L \cdot A \cdot C}. \tag{3.1}$$

The cross-sections are defined similarly for the Z boson in the dilepton invariant mass range $66 < m_{ll} < 116$ GeV (σ^{tot}_{Z}). The production cross-section can be measured in the whole phase space or in a fiducial part of the total phase space defined in such a way to be as similar as possible to the detector acceptance. The correction factor C is the ratio of the total number of generated events which pass the final selection requirements after reconstruction to the total number of generated events within the fiducial acceptance and is used to define the fiducial cross-section. This factor, defined before the decay leptons emit photons via final-state radiation, includes the efficiencies for triggering on, reconstructing, and identifying the $W^{\pm}$ and Z decay products within the acceptance, and also accounts for the slight difference between the fiducial and reconstructed phase spaces. The acceptance factor A is expressed as the fraction of events satisfying the fiducial acceptance (geometric and kinematic requirements) at the Monte Carlo generator level and it is used to extrapolate the fiducial cross section to the full phase-space. The production cross-sections defined without the acceptance factors ($\sigma^{tot} \cdot A$) are referred to as the fiducial cross-sections ($\sigma^{fid}_{W^{\pm}}$ and σ^{fid}_{Z}). For the $W^{\pm}$ measurements, the fiducial phase space is defined by the lepton transverse momentum $p^{l}_{T} > 25$ GeV, the lepton pseudorapidity $|\eta_{l}| < 2.5$, the neutrino transverse momentum $p^{\nu}_{T} > 25$ GeV and the $W^{\pm}$ transverse mass[1] $m_{T} > 50$ GeV. For the Z production the definition is based on $p^{l}_{T} > 25$ GeV, $|\eta_{l}| < 2.5$, and $66 < m_{ll} < 116$ GeV [15].

3.2 Event selection

Electron and muon candidate events are selected using triggers which require at least one electron or muon with transverse momentum thresholds $p_{T} = 24$ GeV or 20 GeV, respectively, with loose isolation requirements [15]. To recover possible efficiency losses at high momenta, additional electron and muon triggers which do not make any isolation requirements are included with thresholds of $p_{T} = 60$ GeV and 50 GeV, respectively.

Electron candidates are required to have $p_{T} < 25$ GeV and to pass *medium* likelihood-based identification requirements [20, 21] optimized for the 2015

[1] We define $m_{T} = \sqrt{2p^{l}_{T}p^{\nu}_{T}[1 - cos(\phi_{l} - \phi_{\nu})]}$ with azimuthal angle of the charged lepton ϕ_{l} and azimuthal angle of the neutrino ϕ_{ν}.

operating conditions, within the fiducial region of $|\eta| < 2.47$, excluding candidates in the transition region between the barrel and endcap electromagnetic calorimeters so with $1.37 < |\eta| < 1.52$. Muons are reconstructed for $|\eta| < 2.4$ with $p_T > 25$ GeV and must pass *medium* requirements again optimized for 2015 operating conditions [22]. At least one of the lepton candidates is required to match the lepton that triggered the event. The electrons and muons must also satisfy p_T-dependent cone-based isolation requirements, using both tracking detector and calorimeter informations. The isolation requirements are tuned so that the lepton isolation efficiency is at least 90% for all $p_T > 25$ GeV, increasing to 99% at 60 GeV.

Jets are reconstructed from energy deposits in the calorimeter using the anti-k_t algorithm [23] with radius parameter $R = 0.4$. All jets with energies calibrated at the electromagnetic scale, must have $p_T > 20$ GeV and $|\eta| < 4.5$. The missing transverse momentum, which in the the $W^{\pm}$ analysis acts as a proxy for the transverse momentum of the neutrino, is defined as the negative of the global sum of all identified physics objects (electrons, muons, jets) as well as *soft terms* accounting for soft tracks and calorimeter energy clusters not matched to any object.

The event selection for the $W^{\pm}$ signature requires exactly one identified electron or muon. The event is required to have $E_T^{miss} > 25$ GeV, and the transverse mass of the $W^{\pm}$ calculated using the missing transverse momentum is required to satisfy $m_T > 50$ GeV. In order for the $W^{\pm}$ selection to be consistent with the missing transverse momentum reconstruction methodology, an overlap removal algorithm is applied to the selection for the events with jets and leptons found at a distance of $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.4$ of each other, removing either one or the other object. After the full $W \to l\nu$ selection, a total of 463063 $W^{\pm}$ candidates (256923 W^+ and 206140 W^-) pass all requirements in the electron channel, and 487090 candidates (272841 W^+ and 214249 W^-) in the muon channel.

Events containing a Z candidate are selected by requiring exactly two selected leptons of the same flavour but of opposite charge with invariant mass of $66 < m_{ll} < 116$ GeV. No overlap removal is applied in the Z analysis as missing transverse momentum is not required in the selection. A total of 34955 candidates pass all requirements in the electron channel and 44899 candidates in the muon channel.

The numbers of W and Z candidate events surviving the selection for the different channels, denoted as N_obs, are also listed in the first column of table 3.2.

3.3 Simulation samples, background evaluation

Monte Carlo simulations are used to evaluate the selection efficiency for signal events and the contribution of several background processes to the analyzed dataset (see [15] for details on the employed Monte Carlo generators and their settings). All of the simulated samples are processed with the Geant4-based simulation [16] of the ATLAS detector [17].

The background contributions from the electroweak (single-boson and diboson) and top-quark (single-top and top-quark pair) production processes have been estimated via simulated samples. The $W\to\tau\nu$ and $Z\to\tau\tau$ processes with the subsequent leptonic decay of the τ are treated as background. The dominant contributions, given as percentages of the total number of simulated events passing the signal selection in each analysis, are as follows: the $W\to\tau\nu$ and top-quark production contribute approximately 2% and 1%, respectively, in the $W^{\pm}$ analysis, the $Z\to e^+e^-$ and $Z\to\mu^+\mu^-$ processes contribute 1% and 5% respectively in $W\to e\nu$ and $W\to\mu\nu$, while the total background in $Z\to l^+l^-$ is approximately 0.5%, dominated by $t\bar{t}$ production (the sum of all electroweak background is 0.2%). The expected backgrounds estimations from EW and $t\bar{t}$ processes in the different channels, denoted as N_{bkg}^{EW}, are listed in the second column of table 3.2 .

Events involving semileptonic decays of heavy quarks, hadrons misidentified as leptons, and in the case of the electron channel, electrons from photons conversions (all referred to collectively as *multijets events*) are a fairly large source of background in the $W^{\pm}$ analysis. The multijet background in the Z analysis is estimated from simulation to be less than 0.1% and is therefore neglected.

The multijet contribution to the electron and muon channels of the $W^{\pm}$ analysis is estimated with a data driven approach, performing maximum likelihood fits on the data with template distributions to exploit the discriminating power between signal and background in certain kinematic distributions. The discriminant variables used in the multijet evaluations are m_T, E_T^{miss}, p_T^l and $\Delta\phi$ between the lepton and transverse missing momentum. Two fit regions are used to extract the multijet normalization. The first fit region is defined as the full event selection but removing the m_T requirement, and the second one is defined as the full event selection but removing the E_T^{miss} requirement. Several multijet-enriched data samples (multijet templates) are built

from events passing all selection requirements in each fit region except lepton isolation. Mutually exclusive requirements (*intervals*) in either tracking (or calorimeter) based isolation variables are chosen to create statistically independent multijet templates. These samples are designed to be progressively closer to the signal-candidate selection by fixing one of the isolation criteria to that of the signal region and varying the other one; four such samples are built for each isolation type in the electron channel and four (for tracking based isolation) or six (for calorimeter based isolation) in the muon channel. Templates are similarly constructed from simulation for $W^{\pm}$ signal and electroweak and top backgrounds, in order to account for potential contaminations in the multijet template. For each isolation interval the normalization of the multijet template is extracted with a maximum likelihood fit to the data in the two fit regions and separately for each one of the discriminant variables and charged lepton samples. In each fit region, the normalization of the signal template derived from simulation is left free to float while the remaining background luminosity and the predicted cross sections are fixed (but are permitted to vary within the integrated luminosity uncertainty of 2.1%). It was verified that the value of the signal normalization extracted from this fit has no significant impact on the multijet estimate. The multijet background event yield in each region is then estimated from this normalization together with the signal region requirement of either $m_T > 50$ GeV, or $E_T^{miss} > 25$ GeV. For each discriminant variable and separately for calorimeter and track based isolation and for each fit region, the estimates obtained in the isolation intervals are used to build a linear extrapolation to the isolation selection used in the signal region. The extrapolation is performed assuming that the individual estimates are uncorrelated.

Separately for the calorimeter based isolation variables and for each fit region, an estimate of the background yield is obtained from a weighted average of the extrapolations obtained with the individual discriminant variables and their uncertainties. The average of the four multijet background estimated fractions found from the track and calorimeter isolation requirements in each fit region is then taken as the nominal multijet background yield in each channel. The uncertainties derived from the linear extrapolations are propagated as systematic uncertainties of the method. A systematic uncertainty for the choice of isolation variable is obtained from half the difference between the averages of the calorimeter based isolation estimated fractions in the two fit regions, and the track based averages. Similarly, a systematic uncertainty due to the use of different fit regions is evaluated as half the difference of the averages obtained from the different types of isolation in the two separate fit regions.

The estimated multijet background fractions with respect to the total number of observed candidate events are 8% and 10% for the electron selection and 4% and 5% for the muon selection in the W^+ and W^- channels, respectively. The corresponding relative uncertainties range from approximately 20% to 30% for the muon and electron channels, and are similar for both the positively and negatively charged samples.

The expected background estimations for the multijet events for the different channels, denoted as N_{bkg}^{QCD}, are reported in the third column of table 3.2 .

3.4 Evaluation of systematic uncertainties

The experimental systematic uncertainties in the measurements of the fiducial cross sections, that we are going to estimate in the next chapter, enter via the evaluation of the correction factor C, the acceptance A (only for the total cross-sections), the integrated luminosity L in the denominator of eq. (3.1), and through the estimation of the background subtracted from the candidate events in its numerator.

Table 3.1: Relative systematic uncertainties (%) for the correction factors C in the different channels.

$\delta C/C$ [%]	$Z\to e^+e^-$	$W^+\to e^+\nu$	$W^-\to e^-\bar{\nu}$	$Z\to \mu^+\mu^-$	$W^+\to \mu^+\nu$	$W^-\to \mu^-\bar{\nu}$
Lepton trigger	0.1	0.3	0.3	0.2	0.6	0.6
Lepton recon, ident	0.9	0.5	0.6	0.9	0.4	0.4
Lepton isolation	0.3	0.1	0.1	0.5	0.3	0.3
Lepton scale & resol	0.2	0.4	0.4	0.1	0.1	0.1
Charge identification	0.1	0.1	0.1	-	-	-
JES and JER	-	1.7	1.7	-	1.6	1.7
E_T^{miss}	-	0.1	0.1	-	0.1	0.1
Pileup modeling	<0.1	0.4	0.3	<0.1	0.2	0.2
PDF	0.1	0.1	0.1	<0.1	0.1	0.1
Total	1.0	1.9	1.9	1.1	1.8	1.8

The sources of systematic uncertainties for the correction factors C are summarized in table 3.1, and listed below:

- *Lepton trigger*: the lepton trigger efficiency is estimated in simulation,

with a dedicated data driven analysis performed to obtain the simulation to data trigger correction factors and the corresponding uncertainties;

- *Lepton reconstruction, identification and isolation*: the lepton selection efficiencies as determined from simulations are corrected with the simulation to data correction factors and their associated uncertainties [18, 19];

- *Lepton energy* and *momentum scale* and *resolution*: uncertainties in the lepton calibrations are applied as they can cause a change of acceptance because of migration of events across the p_T threshold and m_{ll} boundaries;

- *Charge identification*: electron charge misidentification may occur when electrons radiate early in the detector and the resulting photons subsequently convert and are reconstructed as high p_T tracks. A particle with reconstructed charge opposite to the parent electron may then accidentally be associated with the calorimeter cluster. The effect of electrons having their charge reconstructed wrongly is studied using a control sample of $Z \to ee$ events in which both electrons are reconstructed with the same charge and is found to be well described by the Monte Carlo simulations, within the statistical uncertainty of the control sample. An uncertainty is assessed to cover any small residual differences between data and simulation. the probability of charge misidentification is negligible in the muon channel;

- *Jet Energy Scale/Resolution (JES & JER)*: the corresponding uncertainties [24] are propagated to the calculation of the missing transverse momentum;

- E_T^{miss}: uncertainties in the soft component of the E_T^{miss} scale and resolution (evaluated as described in [25]) are included;

- *pileup*: incorrect modeling of pileup effects can lead to acceptance changes and is accounted for with dedicated studies [15];

- *PDF*: the impact of the parton density functions (PDF) eigenvector variations is propagated to the correction factor and to the acceptance (see [15] for details).

The many sources of systematic uncertainties for the factor C have been categorized in common items. A detailed lists will be presented in the next

section where the statistical method used is described. Indeed in the statistical analysis every single source has been treated individually in order to minimize possible correlation effects.

In the Z channel, the systematic uncertainties from the background evaluation contribute negligibly to the experimental cross section uncertainty. This is not the case for the $W^{\pm}$ channel: the multijet background evaluation results in uncertainties of up to 3.4% on the cross section measurements in the electron channel and up to 1.4% in the muon channel.

For clarity's sake we may want to describe the uncertainties on the electroweak backgrounds and on the acceptance factor A. They may be found inside table 3.3: for the electroweak background we have a $\pm 2\%$ from the experimental uncertainties' propagation and a $\pm 5\%$ from the theoretical uncertainty on the cross-sections' computation (both $W^{\pm}$ and Z). For the acceptance A the uncertainty comes from the PDFs and is $^{+1.5\%}_{-1.1\%}$ for the Z and $\pm 1.8\%$ for the $W^{\pm}$.

The cross sections have a further 2.1% uncertainty in the measurement of the integrated luminosity, which is derived, following a methodology similar to that detailed in [26], from a calibration of the luminosity scale using a pair of $x - y$ beam separation scans performed in August 2015. Apart from the determination of the luminosity, the dominant sources of experimental systematic uncertainties in the cross section evaluations are the jet energy scale and resolution and the multijet background for the $W^{\pm}$ measurements while they are lepton reconstruction and identification efficiencies for the Z measurements.

3.5 Cross section estimation

The production cross-sections have been estimated using a bayesian approach to process the experimental informations presented in the previous sections. The standard Bayes formula:

$$p(x|data) = \frac{p(data|x) \cdot p_0(x)}{\int p(data|x') \cdot p_0(x') dx'} \tag{3.2}$$

where x is the variable of interest and $data$ a generic set of experimental informations, can be expressed in terms of the relevant parameters and systematic

uncertainties as:

$$p(\sigma^i|N^i_{obs}, N^i_{bkg}, C^i, A^i, L) = \\ = \frac{\int p(N^i_{obs}|\sigma^i, N^i_{bkg}(\chi_k), C^i(\chi_k), A^i(\chi_k), L(\chi_k))\cdot p_0(\sigma^i)\cdot\prod_k p_0(\chi_k)\cdot d\chi_k}{\int\int p(N^i_{obs}|\sigma'^i, N^i_{bkg}(\chi_k), C^i(\chi_k), A^i(\chi_k), L(\chi_k))\cdot p_0(\sigma'^i)\cdot\prod_k p_0(\chi_k)\cdot d\chi_k\cdot d\sigma'^i} \tag{3.3}$$

where i is an index running over the different production and decay channels, σ^i is the cross section to be estimated, N^i_{obs} the number of observed events, N^i_{bkg} the number of predicted background events, C^i the selection efficiency defining the fiducial phase space, A^i the geometrical acceptance used for the extrapolation to the full phase space, L the integrated luminosity, χ_k the sources of systematic uncertainties treated as nuisance parameters, $p_0(\sigma^i)$ and $p_0(\chi_k)$ the prior density probabilities for the cross sections and systematic uncertainties, respectively. The denominator of the formula is the factor required by the normalization condition. The different cross sections are evaluated all together at the same time and the global likelihood is the product of the individual likelihoods. The likelihoods $p(N^i_{obs}|\sigma^i, N^i_{bkg}(\chi_k), C^i(\chi_k), A^i(\chi_k), L(\chi_k))$ are poissonian with mean value $\mu = \sigma^i\cdot L(\chi_k)\cdot C^i(\chi_k)\cdot A^i(\chi_k) + N^i_{bkg}(\chi_k)$ so that we have:

$$p(N^i_{obs}|\sigma^i, N^i_{bkg}(\chi_k), C^i(\chi_k), A^i(\chi_k), L(\chi_k)) = \frac{e^{-\mu}\cdot\mu^{N^i_{obs}}}{N^i_{obs}!} \tag{3.4}$$

A flat prior has been employed for the cross section while normal priors $N(0,1)$ have been assumed for the sources of systematic uncertainties:

$$x = x_0\cdot\left(1 + \sum_{k=1}^{N_{syst}}\epsilon_k\cdot\frac{1}{\sqrt{2\pi}}exp\left(-\frac{\chi_k^2}{2}\right)\right)$$

where x is a generic parameter, x_0 its nominal value, N_{syst} the number of the sources of systematic uncertainties, ϵ_k the relative variation of the parameter when the source k is shifted by one standard deviation and χ_k the shift of the source k.

The input parameters used in the Bayes formula are summarized in table 3.2 for the different channels while the list of systematic uncertainties are presented in table 3.3 for the predicted background events, the acceptance and the luminosity and in tables 3.4, 3.5, 3.6 for the correction factor C.

Table 3.2: Input parameters for the different channels.

channel [%]	N_{obs}	N_{bkg}^{EW}	N_{bkg}^{QCD}	C	A	Lumi(pb^{-1})
$W^+ \to \mu^+ \nu$	272841	19310±760	12200±3500	0.653±0.012	0.383±0.007	81.0±1.7
$W^+ \to e^+ \nu$	256923	9620±520	15220±6700	0.602±0.012	0.398±0.007	81.0±1.7
$W^- \to \mu^- \nu$	214249	17340±680	11500±3100	0.650±0.012	0.383±0.007	81.0±1.7
$W^- \to e^- \nu$	206140	8670±470	15640±6900	0.614±0.012	0.398±0.007	81.0±1.7
$Z \to \mu^+ \mu^-$	44899	191±10	-	0.711±0.008	0.393±0.007	81.0±1.7
$Z \to e^+ e^-$	34955	144±8	-	0.552±0.006	0.393±0.007	81.0±1.7

Table 3.3: Relative systematic uncertainties (%) for the predicted background events, acceptance and luminosity.

par/syst [%]	$EW\mu$	EWe	$EWtheoretical$	$QCD\mu$	$QCDe$	$PDFe/\mu$	Lumi
N_{EW}^{Z}	±2	±2	±5	-	-	-	±2.1
A_Z	-	-	-	-	-	$^{+1.5}_{-1.1}/^{+1.5}_{-1.1}$	-
N_{EW}^{W+}	±2	±2	±5	-	-	-	±2.1
N_{QCD}^{W+}	-	-	-	±22	±26	-	-
A_{W+}	-	-	-	-	-	±1.8/±1.8	-
N_{EW}^{W-}	±2	±2	±5	-	-	-	±2.1
N_{QCD}^{W-}	-	-	-	±30	±31	-	-
A_{W-}	-	-	-	-	-	±1.8/±1.8	-
Lumi	-	-	-	-	-	-	±2.1

Table 3.4: Summary of the different sources contributing to the uncertainty on C_W for electron and muon final states. The decomposition has been made such that correlations between the various contributions are negligible. The systematic uncertainties have been grouped according to the objects they refer to: electrons (e), muons (μ), jets (JET), missing E_T (E_T^{miss}).

	Process	$W^+ \to$	$e\nu$	$W^- \to$	$e\nu$	$W^+ \to$	$\mu\nu$	$W^- \to$	$\mu\nu$
	Uncertainties on $\delta C_W/C_W$ [%]	Up	Down	Up	Down	Up	Down	Up	Down
e	EG_RESOLUTION_ALL	-0.04	0.06	-0.04	0.05	-	-	-	-
	EG_SCALE_ALL	0.5	-0.4	0.4	-0.4	-	-	-	-
	EL_EFF_Reco_TotalCorrUncertainty	0.5	-0.5	0.5	-0.4	-	-	-	-
	EL_EFF_ID_COMBMCTOY	0.3	-0.3	0.3	-0.3	-	-	-	-
	EL_EFF_Iso_COMBMCTOY	0.14	-0.14	0.13	-0.13	-	-	-	-
	EL_EFF_Trig_COMBMCTOY	0.3	-0.3	0.3	-0.3	-	-	-	-
μ	MUONS_MS	-	-	-	-	0.004	0.002	-0.01	0.001
	MUONS_ID	-	-	-	-	-0.02	-0.005	-0.02	0.007
	MUONS_SCALE	-	-	-	-	-0.14	0.15	-0.13	0.13
	MUON_EFF_STAT	-	-	-	-	0.3	-0.3	0.3	-0.3
	MUON_EFF_SYS	-	-	-	-	0.3	-0.3	0.3	-0.3
	MUON_ISO_STAT	-	-	-	-	0.3	-0.3	0.3	-0.3
	MUON_ISO_SYS	-	-	-	-	0.12	-0.12	0.12	-0.12
	MUON_EFF_TrigStatTOYUncertainty	-	-	-	-	0.2	-0.2	0.2	-0.2
	MUON_EFF_TrigSystUncertainty (w/o q)	-	-	-	-	0.5	-0.5	0.5	-0.5
	MUON_EFF_TrigSyst q dep.	-	-	-	-	0.3	-0.3	-0.3	0.3
JET	JET_GroupedNP_1	-1.4	1.4	-1.4	1.4	-1.3	1.3	-1.3	1.3
	JET_GroupedNP_2	-0.2	0.2	-0.2	0.2	-0.2	0.2	-0.2	0.2
	JET_GroupedNP_3	-1	1	-0.9	0.9	-1	1	-0.9	0.9
	JET_JER_SINGLE_NP	0.3	-0.3	0.3	-0.3	0.3	-0.3	0.3	-0.3
E_T^{miss}	MET_SoftTrk_Scale	-0.09	0.06	-0.09	0.03	-0.13	0.09	-0.1	0.08
	MET_SoftTrk_ResoPerp	0	-0.16	0	-0.15	0	-0.2	0	-0.18
	MET_SoftTrk_ResoPara	0	-0.19	0	-0.13	0	-0.2	0	-0.17
	Pile-up	-0.3	0.3	-0.4	0.4	-0.2	0.2	-0.2	0.2
	Charge misidentification	0.1	0	0	-0.1	-	-	-	-
	PDF Uncertainty	0.1	-0.1	0.12	-0.12	0.09	-0.09	0.12	-0.12
	Total uncertainty [%]	1.9	-1.9	1.9	-1.9	1.9	-1.8	1.8	-1.8

Table 3.5: Summary of the different terms contributing to the uncertainty on C_Z for electron final state. We note that the systematic uncertainties related to the electron object have been assumed to be not correlated between W and Z channels. This is denoted by the presence of the _1 in the corresponding labels.

Parameter	Up[%]	Down[%]
Statistics	0.05	-0.05
EG_RESOLUTION_ALL_1	-0.02	0.01
EG_SCALE_ALL_1	0.22	-0.23
EL_EFF_ID_COMBMCTOY_1	0.48	-0.48
EL_EFF_Iso_COMBMCTOY_1	0.29	-0.29
EL_EFF_Reco_TotalCorrUcertainty_1	0.77	-0.76
EL_EFF_Trig_COMBMCTOY_1	0.05	-0.05
Pileup	0.01	-0.01
Opposite charge requirement	-0.15	0.15
PDF	0.14	-0.08
Total	1.00	-0.99

Table 3.6: Summary of the different terms contributing to the uncertainty on C_Z for muon final state. We note that the systematic uncertainties related to the muon object have been assumed to be not correlated between W and Z channels. This is denoted by the presence of the _1 in the corresponding labels.

Parameter	Up[%]	Down[%]
Statistics	0.05	-0.05
MUONS_ID_1	-0.05	-0.01
MUONS_MS_1	-0.01	0.00
MUONS_SCALE_1	-0.07	0.06
MUON_EFF_STAT_1	0.61	-0.61
MUON_EFF_SYS_1	0.64	-0.64
MUON_EFF_TrigSystUncertainty_1	0.17	-0.17
MUON_EFF_TrigStatTOYUncertainty_1	0.10	-0.10
MUON_ISO_STAT_1	0.49	-0.48
MUON_ISO_SYS_1	0.22	-0.21
Pileup	-0.01	-0.03
Opposite charge requirement	0.00	0.00
PDF	0.02	-0.01
Total	1.05	-1.05

Chapter 4

Results

We employed a dedicated code, BAT (Bayesian Analysis Toolkit [5], for further details Appendix A), entirely written in C++, finely tuned for our purpose. Thanks to BAT our analysis went almost smoothly. And yet, a few things must be pointed out. For instance luminosity. This is a troublesome systematic, since it is the dominant source of systematic uncertainty, fully correlated among all the channels and with a large impact on the measurement. The convergence of the Markov Chain is then in peril, because by using Gelman-Rubin diagnostics (see Appendix C) we observe difficulties for the chains to reach a well defined convergence. The only thing we can do is to run the chain for as long as we can.

It has been noticed the convergence got sped up by reducing the uncertainty on the luminosity. Recently the ATLAS collaboration, thanks to detailed studies performed with the fundamental contribution of the Bologna group, was able to considerably reduce the uncertainty on the luminosity measurement from 5% to 2.1% thus helping our work a great deal.

All the results presented here have been obtained using this new reduced uncertainty for the integrated luminosity.

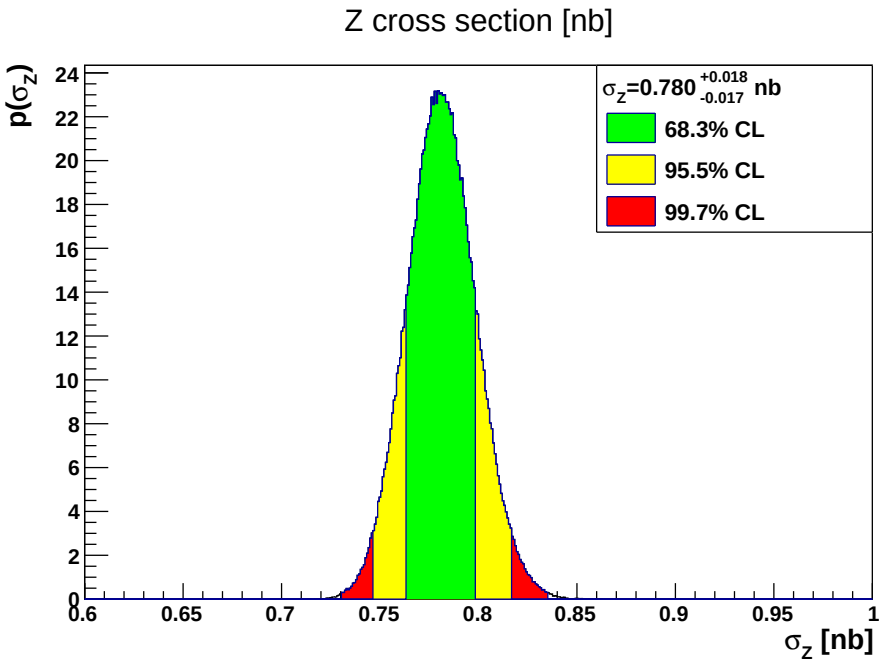

Figure 4.1: Z cross-section, the y-axis is normalized.

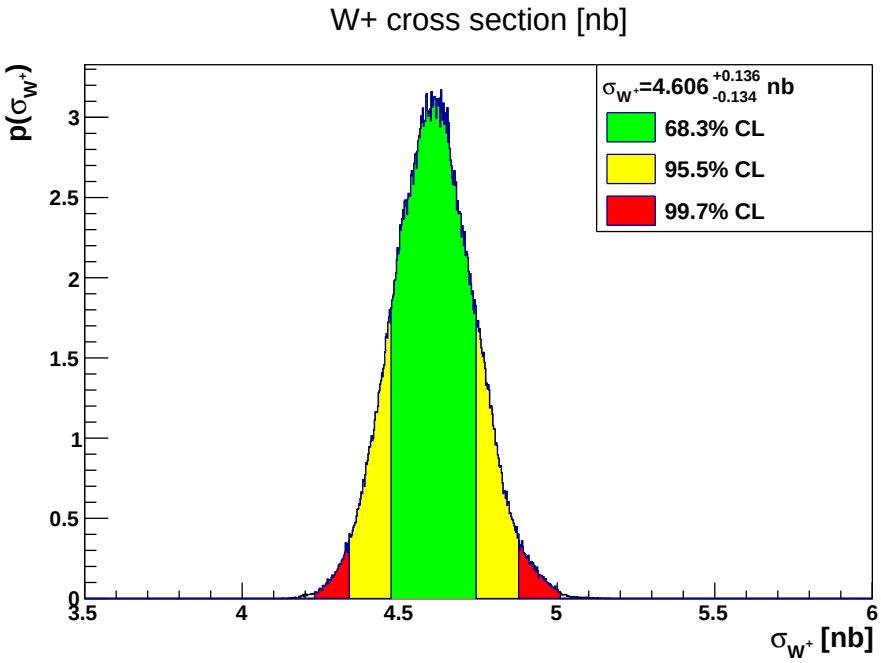

Figure 4.2: W^+ cross-section, the y-axis is normalized.

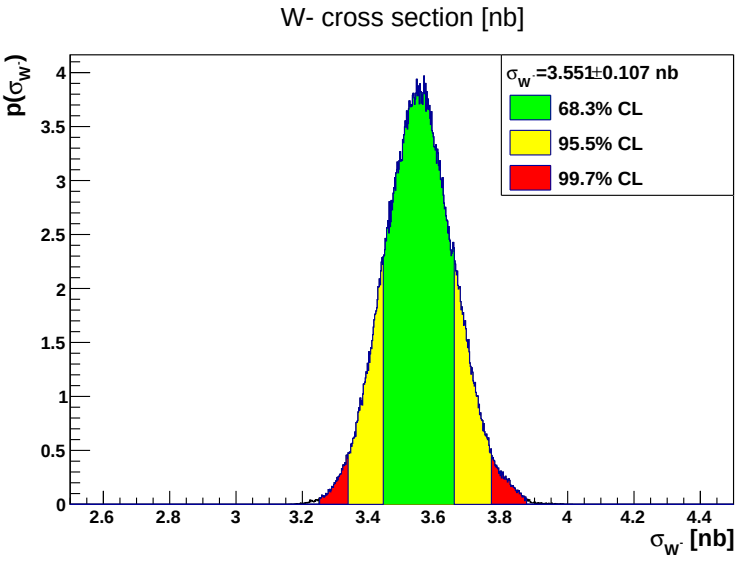

Figure 4.3: W^- cross-section, the y-axis is normalized.

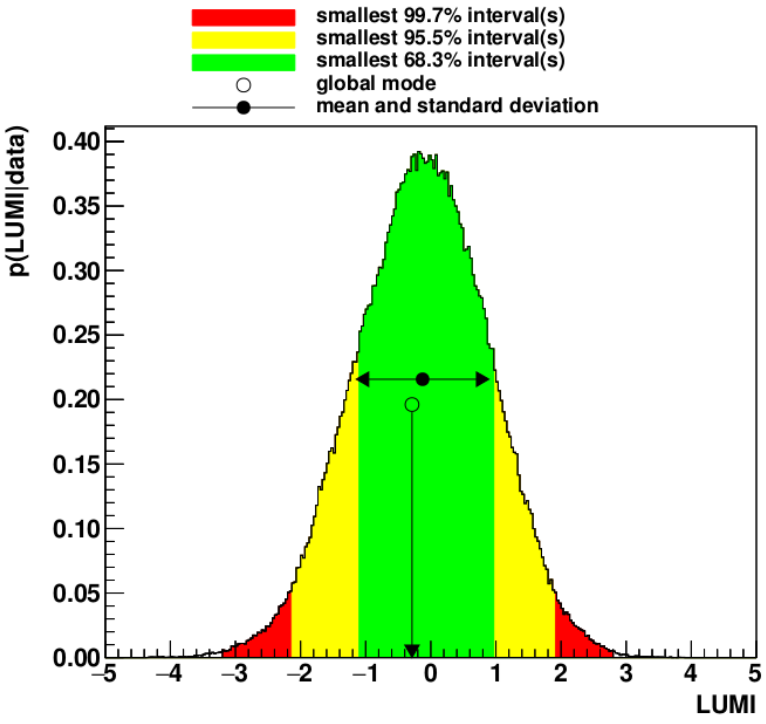

Figure 4.4: Marginalized integrated luminosity, the y-axis is normalized. We note that even the problematic luminosity is in the posterior bound by $\pm 1\sigma$.

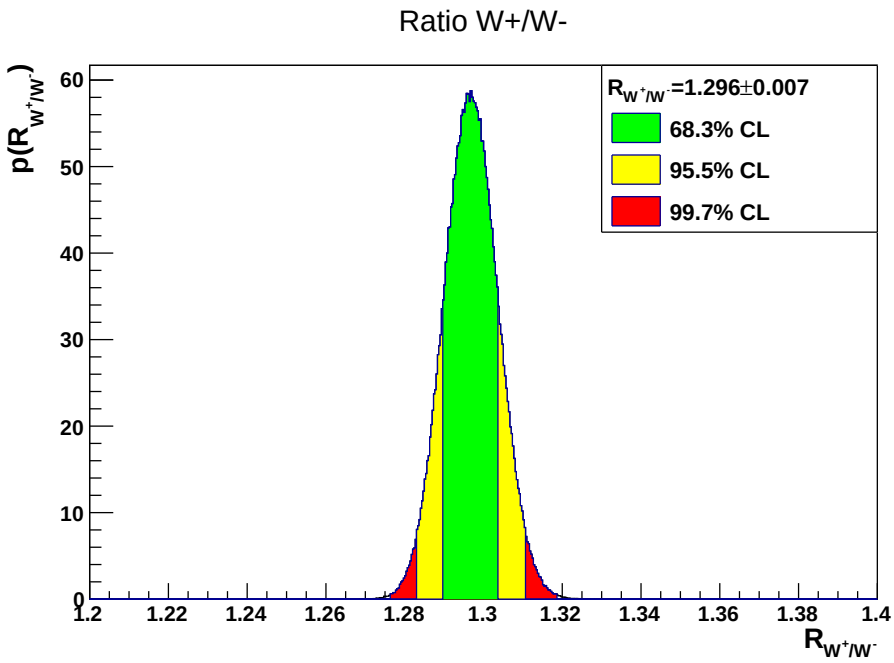

Figure 4.5: Ratio W^+ and W^- production cross section.

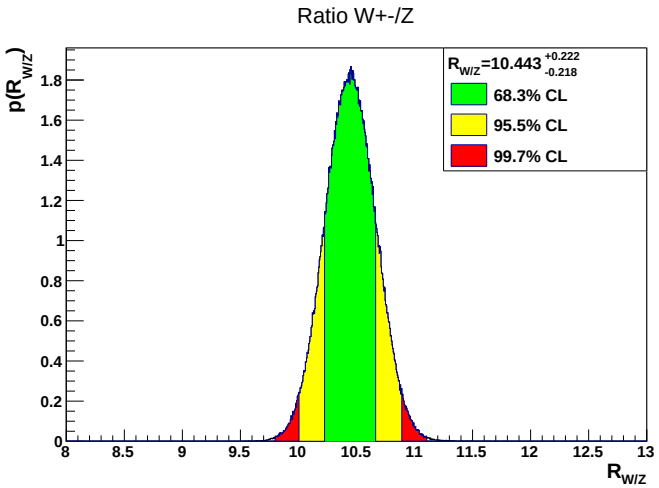

Figure 4.6: Ratio $W^\pm$ and Z production cross section.

Table 4.1: Correlation coefficients among W^- , W^+ , and Z boson production combined fiducial cross-section measurements excluding the common normalization uncertainty due to the luminosity calibration.

	W^+	W^-	Z
W^+	1	0.98	0.71
W^-		1	0.68
Z			1

Figure 4.7: Running mean plots on xsecZ, xsecW$^+$, xsecW$^-$ against the iteration number.

The MCMC technique allows us to sample the parameter space and to evaluate the probability density function for each parameter using a marginalization with respect to the others so that any function of the parameter can

Table 4.2: Ratios of the W^+ to W^- and $W^\pm$ to Z boson fiducial cross sections for the electron, muon and combined measurements. We even compared them to the baseline method's results actually published a few weeks ago.

	$\sigma^{fid}_{W^+}/\sigma^{fid}_{W^-}$	$\sigma^{fid}_{W^\pm}/\sigma^{fid}_{Z}$
our (bayesian approach)	1.296±0.007	10.44±0.22
article's baseline method	1.295±0.013	10.31±0.24

also be easily determined during the sampling. The marginalized probability density functions (PDF) for the three cross sections (Z, W^+ and W^-) are shown in figures 4.1, 4.2 and 4.3 . We note that the distributions do not present strange features and are approximately gaussian with a slight larger tail towards high cross section values. Since the distributions are not perfectly gaussian we chose to characterize them using the median and the smallest interval containing the 68.3% area of the probability density distribution (68.3% credibility level, CL, interval). The 68.3% (1σ), 95.5% (2σs) and 99.7% (3σs) credibility level intervals are also shown in the plots in green, yellow and red respectively. The results for all the channels are summarized in table 4.3 together with a comparison with the baseline method of the ATLAS paper [15]. Figure 4.4 shows the pdf of the luminosity which represents the dominant uncertainty, fully correlated among all the channels. Also in this case no strange features are observed, the posterior distribution is approximately gaussian with mean close to zero and σ close to 1. Two other very important observables are the cross section ratios $\sigma_{W^+}/\sigma_{W^-}$ and $\sigma_{W^\pm}/\sigma_Z$ which are presented in figs.4.5 and 4.6 respectively. these observables allow for stringent tests of the SM predictions since all the correlated uncertainties, including the luminosity related one, cancel in the ratio. The results are also summarized in table 4.2 along with a comparison with the results obtained with the baseline method [15].

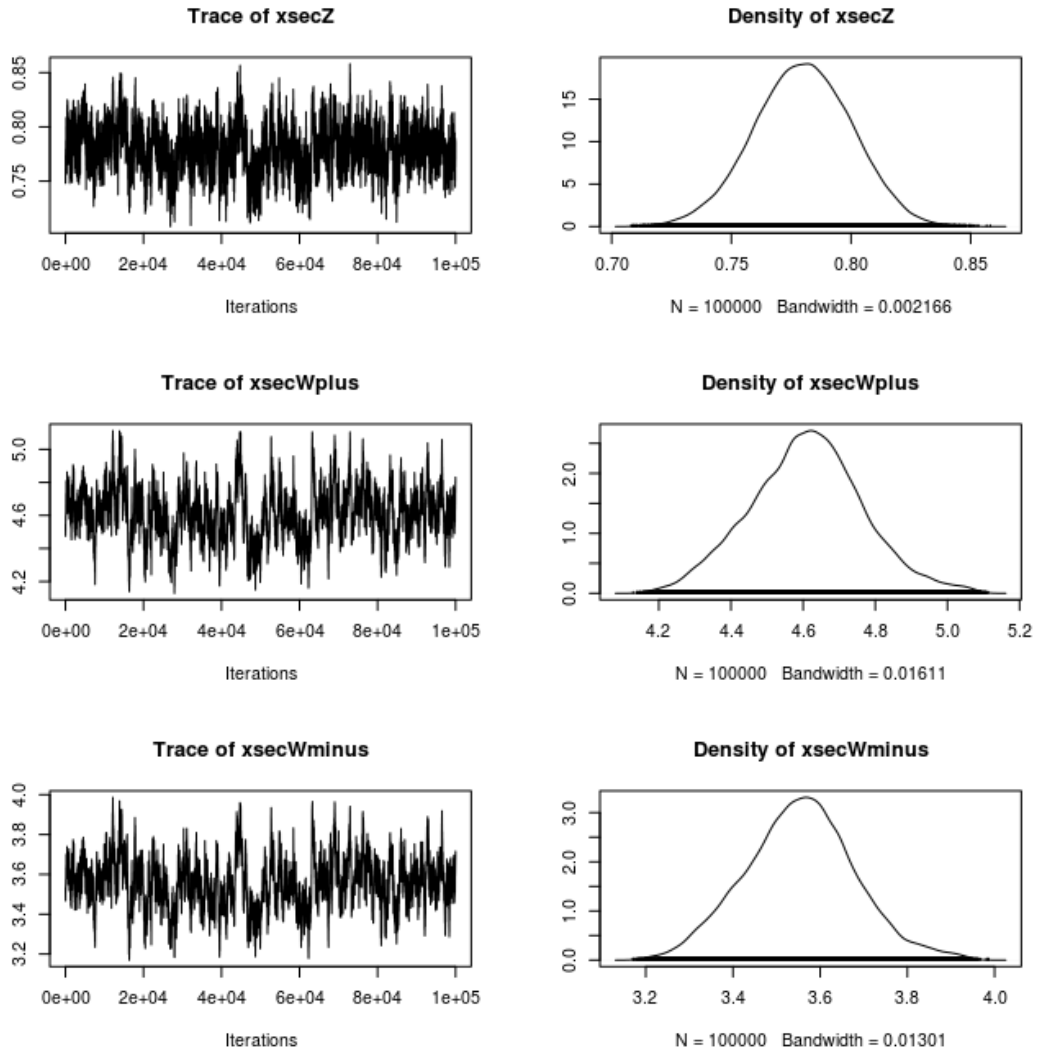

Figure 4.8: Parameter space sampling for σ_Z, σ_{W^+}, σ_{W^-} against the iteration number for one of the five chains used in the sampling (left). The corresponding density plots are shown on the right side.

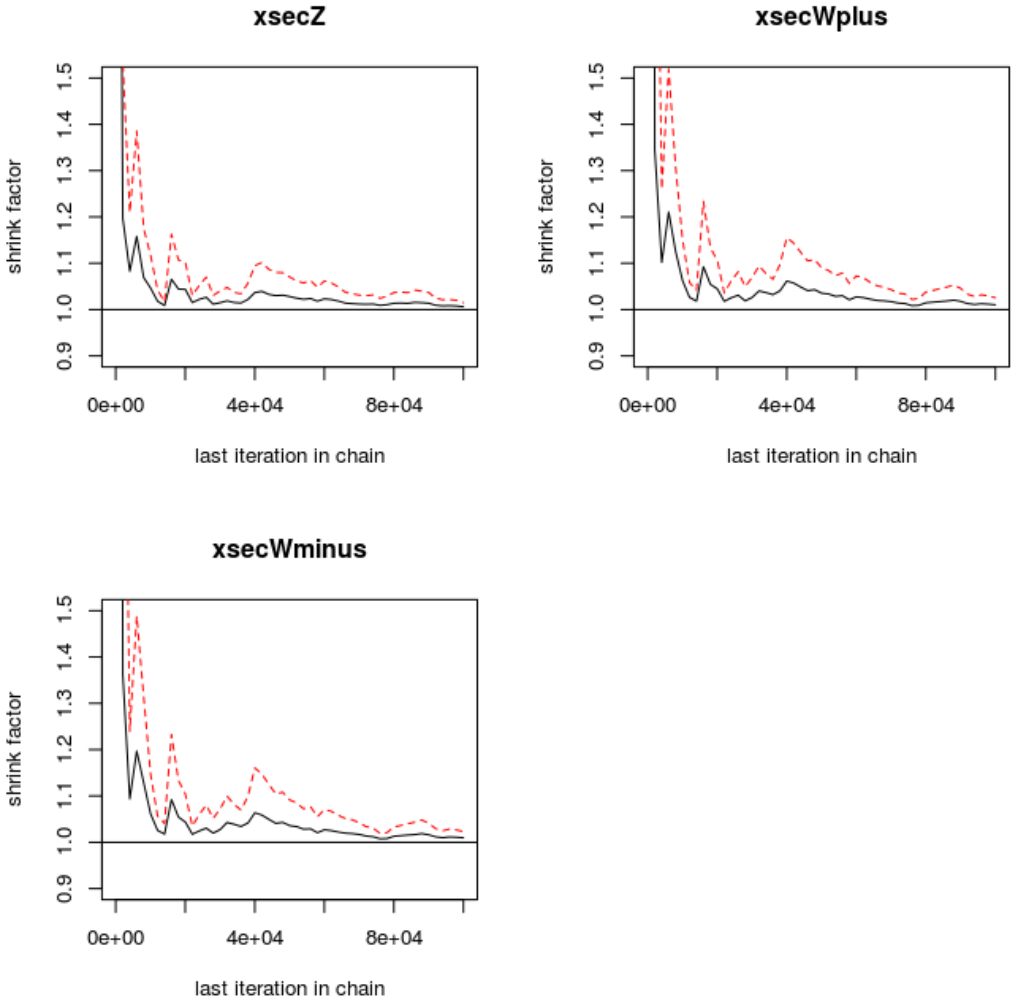

Figure 4.9: Shrink scale factor analysis for the parameters describing the Z, W^+ and W^- production cross sections against the iteration number.

The last set of plots present the checks we made in order to assess the convergence of the set of Markov Chains used to sample the parameter space. The results presented above have been obtained using 5 independent chains with a pre-run of maximum 10^6 iterations per chain to assess the convergence of the chains using the Gelman-Rubin convergence diagnostic (see Appendix C) and a sampling run of 10^6 iterations per chain.

By looking at fig. 4.8 we can ascertain the good mixing of the chains. This is clearly visible in the left side plots showing the path of one of the Markov Chains along the iteration number. The chain explores the whole space

and does not get stuck in any part of the parameter space. The shrink scale factors are presented in fig. 4.9, indicate the convergence status of the chains according to the Gelman-Rubin criterion and BAT checks its value for each parameter during the pre-run assuming that the convergence has been reached when all the factors for every parameter is below 1.1 . We can observe that all the parameters reach a shrink scale factor of almost 1.0 in the long run (thankfully the only thing that matters).

Table 4.3: Summary of the fiducial production cross section results for the Z, W^+, W^- and $W^\pm$ processes. The results of this analysis are presented in the first column and are compared to the public ATLAS results in the second column.

	bayesian [nb]	baseline [nb]
σ_Z^{prod}	0.781±0.017	0.780±0.017
$\sigma_{W^+}^{prod}$	4.61±0.13	4.53±0.13
$\sigma_{W^-}^{prod}$	3.55±0.10	3.50±0.10
$\sigma_{W^\pm}^{prod}$	8.17±0.23	8.03±0.23

Finally fig. 4.10 shows a summary of the mean and the root mean square of every systematic uncertainties, the so called Nuisance Parameters (NPs) which are integrated out to obtain the posterior pdf of the parameters of interest (the Z, W^+ and W^- production cross sections). As may be noted from the plot no strange features are observed, all the NPs do not show large departures from the prior which is a normal distribution with mean 0 and sigma 1. The only exception is the source of the systematic uncertainty n.46, the uncertainty on the multijet electron background, which exhibits a root mean square significantly smaller than 1. This deviation is probably due to an overestimation of this peculiar uncertainty.

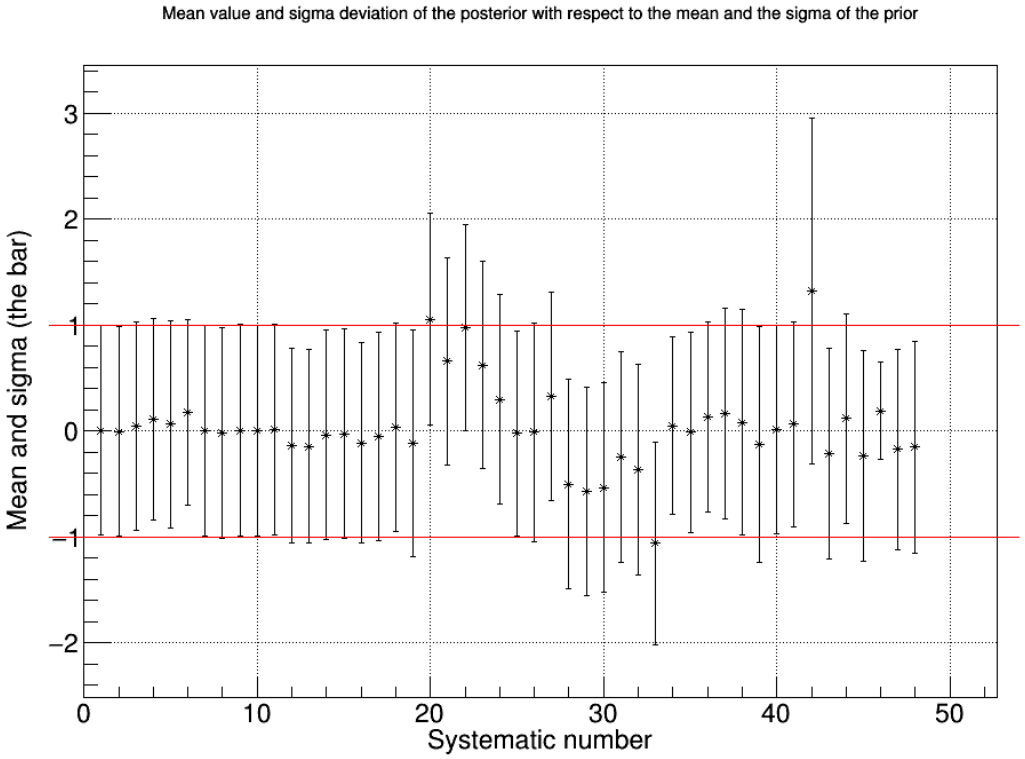

Figure 4.10: Summary for the systematics.

In order to make a complete comparison between our results and the article's the total production cross sections were computed as well using the same method and including in the likelihood also the acceptance A and its uncertainty in order to extrapolate the cross-sections to the full-phase space. These last results are shown in table 4.4 .

The results for the W^+, W^- and Z cross sections obtained using a bayesian approach which accounts for the correlations among the different channels introduced by the common source of systematic uncertainties, are in very good agreement with the SM predictions and the public ATLAS results [15] obtained using a classical approach. Figs. 4.11 and 4.12 show the level of the agreement for the fiducial and the full phase space cross-sections respectively, while fig. 4.13 the comparisons made for the ratios.

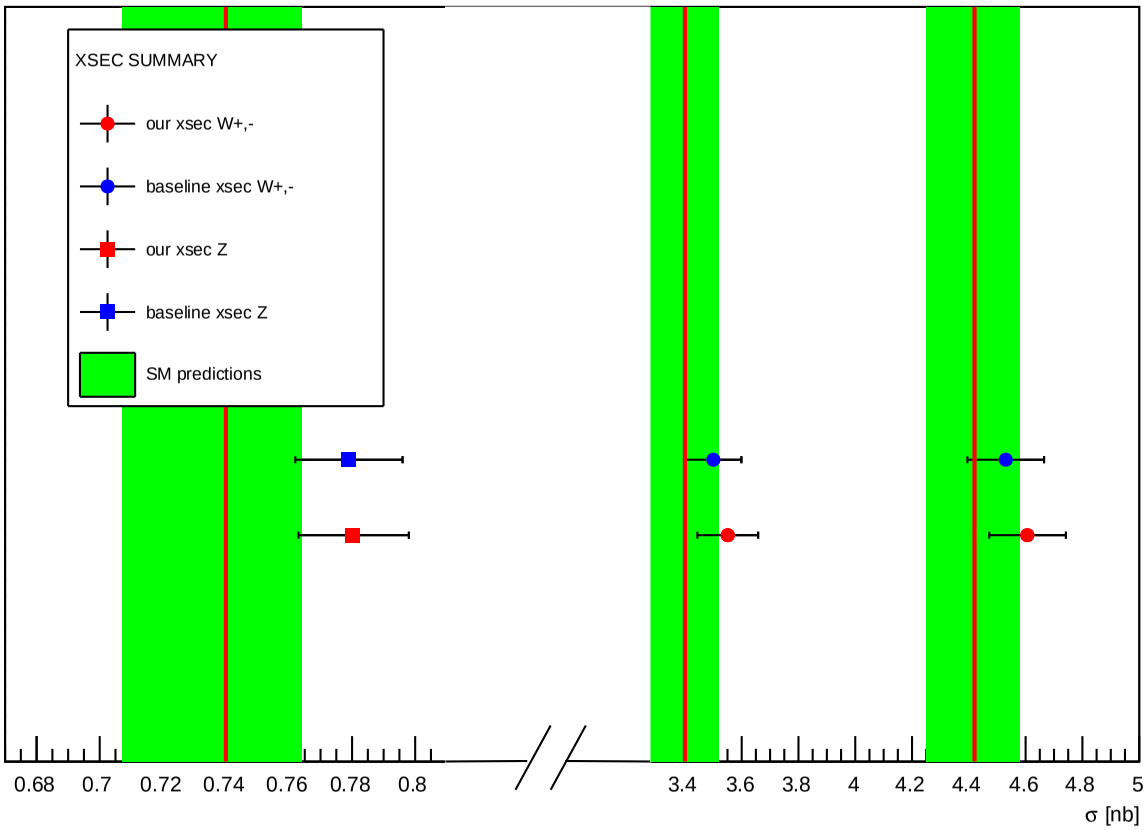

Figure 4.11: Summary of the results for the fiducial production cross sections including the Standard Model predictions.

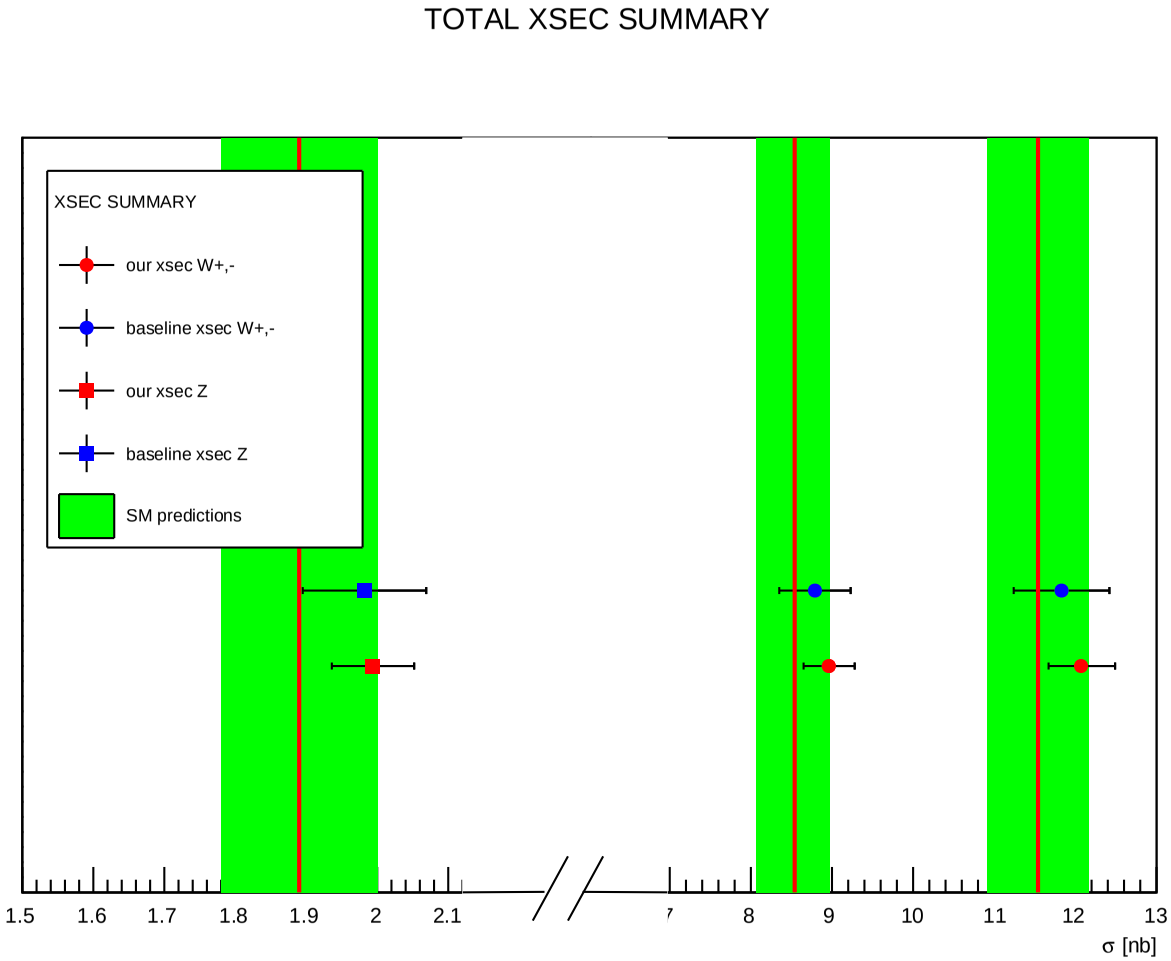

Figure 4.12: Summary of the results for the total production cross sections including the Standard Model predictions.

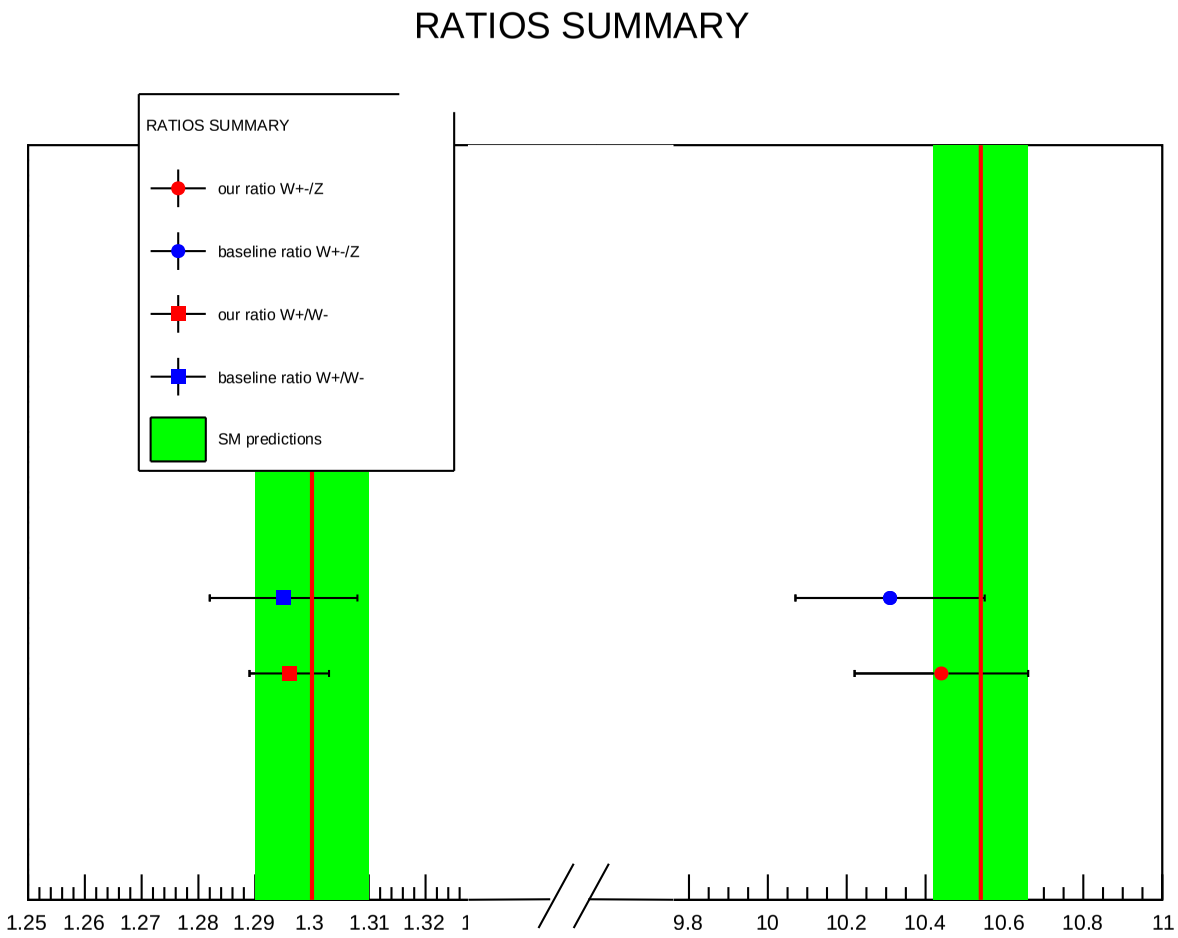

Figure 4.13: Summary of the results for the ratios including the Standard Model predictions.

Table 4.4: Summary of the total production cross section results for the Z, W^+, W^- and $W^\pm$ processes. To obtain them it was necessary to consider the acceptance factor A. The results of this analysis are presented in the first column and are compared to the public data results in the second column.

	bayesian [nb]	baseline [nb]
σ_Z^{TOT}	$1.993^{+0.059}_{-0.057}$	1.981±0.087
$\sigma_{W^+}^{TOT}$	$12.07^{+0.42}_{-0.40}$	11.83±0.59
$\sigma_{W^-}^{TOT}$	$8.96^{+0.32}_{-0.31}$	8.79±0.44
$\sigma_{W^\pm}^{TOT}$	$21.03^{+0.74}_{-0.71}$	20.64±1.00

Conclusions

In this thesis we have computed the production cross sections for $W^{\pm}$ and Z bosons using their leptonic decays. The analysis was performed using pp collision data provided by LHC at a center of mass energy of $\sqrt{s} = 13$ TeV and recorded by the ATLAS detector on summer 2015 for an integrated luminosity of $L = 81$ pb^{-1}. The analysis have been carried on using the same event selections reported in the recent ATLAS paper [15] but exploiting the features of the bayesian approach for the statistical treatment of the data and the evaluation of the probability density function for the the cross-sections of interest. The results are in very good agreement with the SM predictions and the ATLAS results, as witnessed by the summary of figs. 4.11 and 4.12 for the fiducial and the full phase space cross sections, respectively. The uncertainty of the measurements is dominated by the luminosity uncertainty which is fully correlated among the different channels. The ratios of the different cross-sections can hence be determined with a much better precision then the cross-sections themselves and constitute a powerful test of the SM predictions. Also the ratios $\sigma_{W^+}/\sigma_{W^-}$ and $\sigma_{W^{\pm}}/\sigma_Z$ have been evaluated and found again in good agreement with the ATLAS published results and SM predictions as shown in fig. 4.13 .

I think that the bayesian method is a powerful tool which allows to treat data correlations and systematic uncertainties in a very natural and straightforward way (meaning free of unclear recipes sometimes present in the classic statistical approach). It is hence of interest to go on with both approaches as one could be a check to the other.

Appendix A

Why BAT and how to use it

BAT is a package which employs everything explained until now. It makes easy performing bayesian analysis, and it is interfaced with ROOT. You can find it at [4]. The release considered in all the following is the 0.9.4.1, one of the latest.

A.1 How it works

By activating the bat-project script you first create a skeleton, comprised of a few components.

The core of the skeleton is the part listed below which allows the evaluation of the likelihood:

```
double TEST::LogLikelihood(const std::vector<double> & parameters) {
// This methods returns the logarithm of the conditional probability
// p(data|parameters). This is where you have to define your model.
// access parameters from vector by remembering their positions, e.g.
// double mu = parameters[0];
// or by looking up their indicies, e.g.
// double mu = parameters[fParameters.Index("mu")];
// Calculate your likelihood according to your model. You may find
// the built in functions such as BCMath::LogPoisson helpful.
// Return the logarithm of this likelood
```

```
      sig=parameters[0];           #not in the skeleton
      Nsig=sig*eff*lumi;           #same
      mu=Nsig+Npredicted;          #same
      logprob +=  BCMath::LogPoisson(Nobs,mu);   #same
  }

return logprob;       #you have to put it this way
//return -1;          #if you want your code to run
}
```

In this last case we are interested in the actual posterior:

$$P(\sigma|N_{pred}, N_{obs}, lumi, eff)$$

with the poissonian likelihood:

$$P(N_{obs}|\sigma, N_{pred}, lumi, eff) = \frac{\mu^{N_{obs}} \cdot e^{-\mu}}{N_{obs}!},$$

where N_{pred} is the predicted number of events, N_{obs} is the observed number of events, $lumi$ the luminosity and eff the efficiency.

This code will be explained better in the section A.5 when the example about a signal beyond the Standard Model is analyzed, because it is the code employed there.

A.2 XML: powerful labelling system

Extensible Markup Language (XML) is a markup language that defines a set of rules for encoding documents in a format which is both human-readable and machine-readable.

This labelling system proved to be very useful in providing easy to understand data to BAT. What we wanted to do was to be able to give data from outside the code. In other words, by using peculiar text files a.k.a. "control cards", we controlled the analysis. This is interesting for many reasons:

- if the core program is correct, small changes may be applied only to the control cards;

- using control cards, continuous acts of compiling for the program are not required anymore, because we can perform different analysis by modifying the input parameters in the control cards feeding it to our program ⇒less bugs to worry about;

- if we want to change the function for the likelihood from outside we just have to change the actual flag in the control cards, so we only have to consider more types of likelihood once for all in the core program coding phase;

- the labelling system allows us to produce a set of control cards easy to understand;

Thankfully the workload was eased by employing XML syntax analyzers, a.k.a. *parsers* , which translate our XML file in C++ code.

Our target parser is CMarkup [6], because we deemed it easiest to control and understand.

A.3 CMarkup in BAT

The few tags we introduced are the ones needed to compute cross-sections, so if need arises a few more would be no trouble at all to add.

First of all, XML needs a parent element (or root element) and the first element we usually call it <ROOT>. It must be closed at the end (that is true for every element) with the syntax < /ROOT> , but truth to be told, the name of the parent element is up to everyone's fantasy.

We added a channel system, so that an analysis may be made for many decay ways (like $Z \to e^+e^-$ or $Z \to \mu^+\mu^-$). How to name every channel is yet again up to your will.

The system will find the following labels:

- **Measure**: by the syntax $< Measure > thing < /Measure >$ and in succession all the others, we add the quantities that we want to obtain by the end of the analysis;

- **Measure-prior**: by the syntax $< Measure - prior > number <$

$/Measure-prior>$ in succession for the many Measure-s, we specify their prior, $0 = flat$, $1 = gauss$;

- **Measure-lim**: by the syntax $<Measure{-}lim>number1</Measure{-}lim>$, $<Measure-lim>number2</Measure-lim>$, we specify the ranges of the Measure-s, again one after the other;

- **Measure-par**: by the syntax $<Measure{-}par>number1</Measure{-}par>$, $<Measure-par>number2</Measure-par>$, we tell the program the mean and the sigma for the gaussian, and it will only be used in that instance;

- **Likelihood**: it is needed to determine the likelihood type $1 = poisson$, $2 = gauss$;

- **Parameters**: the parameters of the likelihood; again they must have the syntax $<Parameters>thing</Parameters>$;

- **Parameters-values**: the value of the parameters;

- **Data**: mainly employed for the N_{obs};

- **Data-value**: its value:

- **Syst**: nuisance parameters (NP), same name for NP on different data set means full correlation;

- **Syst-lim**: by the syntax $<Syst-lim>number1</Syst-lim>$, $<Syst-lim>number2</Syst-lim>$, we specify the ranges of the Syst-s, again one after the other;

- **Syst-prior**: it is the prior type for the systematic source and it needs to be specified for all the Syst-s, $0 = cost$, $1 = gauss$, $-1 = notconsidered$;

- **End-dataset**: it needs to be present as well, before the channel element everything belong to is closed;

- **Syst-whateveryouwant**: the system will check if whateveryouwant is a label corresponding to anyone of the many Parameter-s. If a match is found, it will enter the element and find the tag belonging to the Syst it has dependence from (concatenated dependance must be employed). It sounds tricky but an example in the following will certainly clear the way.

A.4 Example: the XML input data

This is how the input data, let's say example.xml, should be written:

```
<ROOT>
   <Channel>
      <Measure>xsecWplus</Measure>
      <Measure-prior>0</Measure-prior>
      <Measure-lim>0.6</Measure-lim>
      <Measure-lim>0.9</Measure-lim>
      <Likelihood>1</Likelihood>
      <Parameters>NbgEW_Wplus</Parameters>
      <Parameters-values>9625</Parameters-values>
      <Parameters>Lumi</Parameters>
      <Parameters-values>85000</Parameters-values>
      <Data>Nobs</Data>
      <Data-value>256858</Data-value>
      <Syst>BKG_EW_EL</Syst>
      <Syst>BKG_EW_TH</Syst>
      <Syst>LUMI</Syst>
      <Syst-lim>-5</Syst-lim>
      <Syst-lim>5</Syst-lim>
      <Syst-lim>-5</Syst-lim>
      <Syst-lim>5</Syst-lim>
      <Syst-lim>-5</Syst-lim>
      <Syst-lim>5</Syst-lim>
      <Syst-NbgEW_Wplus>
         <BKG_EW_EL>0.02</BKG_EW_EL>     #if the second value
         <BKG_EW_EL>0</BKG_EW_EL>        #is 0, the first
         <BKG_EW_TH>0.05</BKG_EW_TH>     #value is assumed
         <BKG_EW_TH>0</BKG_EW_TH>        #for the systematic
         <LUMI>0.09</LUMI>               #uncertainty
         <LUMI>0</LUMI>
      </Syst-NbgEW_Wplus>
      <Syst-Prior>1</Syst-Prior>
      <Syst-Prior>1</Syst-Prior>
      <Syst-Prior>1</Syst-Prior>
      <EndDataSet>0</EndDataSet>
   </Channel>
</ROOT>
```

And in the runTESTproject.cxx created by the BAT tools it is to be called by:

```
TESTproject* m = new TESTproject("example.cards");
```

A.5 Example: BSM limit

Suppose you want to study the possibility of a signal beyond the Standard Model (BSM). Let's consider the following data:

- $N_{obs} = 5$
- $N_{pred} = 4$
- $lumi = 100 pb^{-1}$
- $eff = 0.5$

where N_{pred} is the predicted number of events, N_{obs} is the observed number of events, $lumi$ the integrated luminosity and eff the efficiency.

Let's consider the following likelihood:

$$P(N_{obs}|\sigma_{BSM}, N_{pred}, lumi, eff) = Poisson(N_{obs}, \mu)$$

with μ computed as:

$$\mu = N_{pred} + \sigma_{BSM} \cdot lumi \cdot eff$$

We next impose a flat prior for σ_{BSM}, the cross section of a possible BSM process.

By putting together all this informations, our control cards should look like:

```
<ROOT>
<Channel>
```

```
<Measure>sigmaBSM</Measure>
<Measure-prior>0</Measure-prior>
<Measure-lim>0.</Measure-lim>
<Measure-lim>0.5</Measure-lim>
<Measure-par>0.</Measure-par>
<Measure-par>0.</Measure-par>
<Likelihood>1</Likelihood>
<Parameters>Eff</Parameters>
<Parameters>Npred</Parameters>
<Parameters>Lumi</Parameters>
<Data>Nobs</Data>
<Data-value>5</Data-value>
<Parameters-values>0.5</Parameters-values>
<Parameters-values>4</Parameters-values>
<Parameters-values>100</Parameters-values>
<Syst>LUMI</Syst>
<Syst-lim>-5</Syst-lim>
<Syst-lim>5</Syst-lim>
<Syst-Lumi>
    <LUMI>-0.05</LUMI>
    <LUMI>0.05</LUMI>
</Syst-Lumi>
<Syst-Prior>1</Syst-Prior>
<EndDataSet>0</EndDataSet>
</Channel>
</ROOT>
```

where a single systematic uncertainty was added for clarity's sake. Of course we are not done yet. We have to modify the *LogLikelihood*. As explained in the short introduction to BAT found on its website, we have to start editing the TestProject.cxx file. In our skeleton this is done in the following way:

```
double TestProject::LogLikelihood(const std::vector<double> & parameters) {

  const double pig=3.14159;
  double logprob = 0.;
  double par0, systdev;
  double Nsig, mu, Nobs;
  vector<double> parx;
  map<string,map<string,vector<double> > >::iterator it1;
  map<string,vector<double> >::iterator it2, itnobs;
  map<string,double>::iterator it3;
  map<string,vector<int> >::iterator it4;
  double a,b,sig;

  // loop over different datasets, ids represemts the number
```

```
  //of the channel you are working on
  for (int ids=0; ids<m_parsyst.size(); ids++){
    // loop on parameters
    for(int ipar=0; ipar<likelihood_par.size(); ipar++){
      if(debug>2)cout << " DEBUG ===> <=== " << ipar << " ";
      cout << likelihood_par.size();
      cout << " " << likelihood_par.at(ipar) << " " << endl;
      it3=m_parcv.at(ids).find(likelihood_par.at(ipar));
      par0=it3->second;
      it1=m_parsyst.at(ids).find(it3->first);
      systdev=0;
      // loop on systematics
      for(it2=it1->second.begin(); it2!=it1->second.end(); it2++){
  if(it2->second.at(0) != 0){
    it4=m_systtyp.find(it2->first);
    if(it4 != m_systtyp.end()){
      if(it2->second.at(1) == 0){
                systdev += it2->second.at(0)*parameters[it4->second.at(0)+idm-1];
      }
      else{
parabola(it2->second,&a,&b);
systdev += a*pow(parameters[it4->second.at(0)+idm-1],2)+b*parameters[it4->second.at(0)+idm-1];
      }
    }
  }
      }
      parx.push_back(par0*(1+systdev));
      systdev=0;
    }

    itnobs=m_data.at(ids).begin();
```

Now starts your coding:

```
    if(ids == 0){
      sig=parameters[0];
      Nsig=sig*parx.at(0)*parx.at(2);
      mu=Nsig+parx.at(1);
      parx.at(k) << endl;
    }
    Nobs=itnobs->second.at(0);
    if(likelihood.at(0) == 1)
      //IT IS:
      logprob +=  BCMath::LogPoisson(Nobs,mu);
    parx.clear();
  }
  return logprob;
}
```

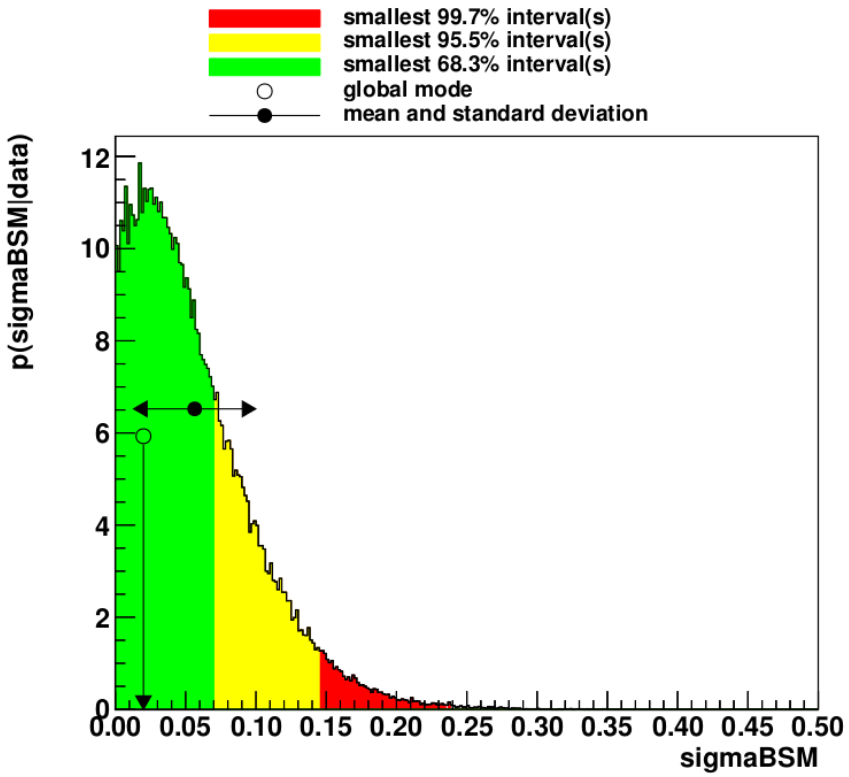

Figure A.1: Results for the BSM cross-section.

Fig. A.1 shows the BAT output for the posterior probability distribution function of σ_{BSM}. Using the above probability distribution, a limit at 95% credibility level may be evaluated as:

$$\int_0^{\sigma_{BSM}^{95}} P(\sigma_{BSM}|DATA)d\sigma_{BSM} = 0.95 \Rightarrow \sigma_{BSM}^{95} = 0.145\ pb.$$

Appendix B

Lorentz angle inside the Pixel Detector

The ATLAS pixel tracking detector is located close to the interaction region and exploits the drift of the charge carriers produced by the charged particles formed in the *pp* collisions to reconstruct the tracks of these particles. Furthermore it is immersed in a strong, solenoidal magnetic field which is used to bend the particle trajectories in order to estimate their momentum.

In the presence of an electric field E and a magnetic field B, with a component orthogonal to E, the charge carriers move along a direction that forms an angle (Lorentz angle) with the electric field. This angle affects the area of collection of the charge carriers. Resolution and efficiency of the detector depend on the track incidence angle and on the charge drift angle: the determination of this angle is therefore very important to define the mechanical design and optimize the detector performances.

The distribution of charge produced by ionizing particles while drifting to the read-out pixels can spread over more than one pixel. The spread depends on the particle incidence angle and is minimum for an angle equal to the Lorentz angle [13].

In order to evaluate the Lorentz angle inside the Pixel Detector I performed a study on the mean cluster sizes in the transverse component with respect to the azimuthal incidence angle.

Thankfully ROOT [27] comes in our aid, providing us with the TProfile class. The peculiarity of this class is that is will give us the mean and the spread

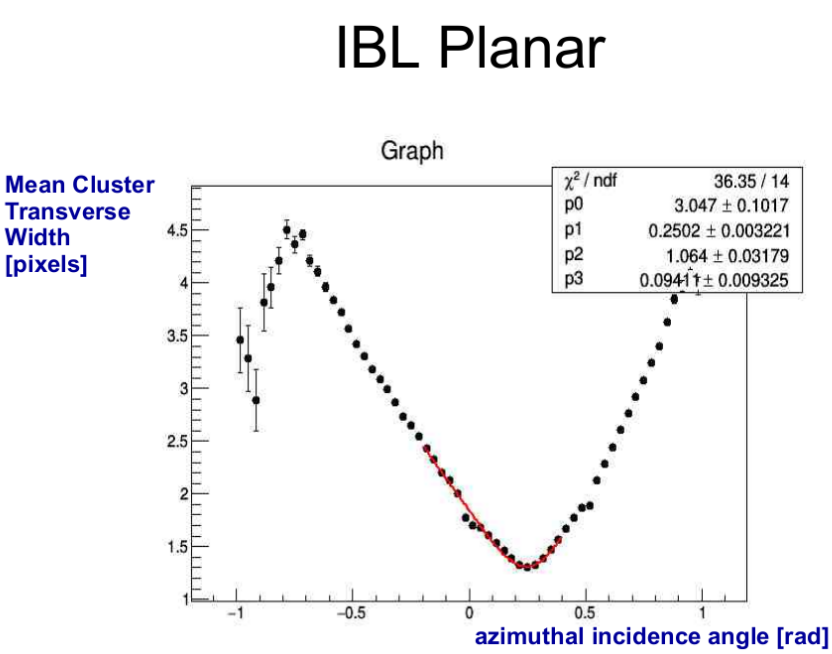

Figure B.1: IBL Planar, Lorentz angle. Incidence angle vs mean cluster size.

of the y-axis with respect to the same bin in the x-axis. So if we consider a two-dimensional plot with the cluster sizes in the y-axis and the azimuthal angle in the x-axis, it will give us exactly what we need.

The next step is to consider the quality cuts made in our analysis. Track selection required at least 2 pixel hits, 6 SCT hits, 15 TRT hits. Given how closely packed they are (we are reminded that IBL was exactly designed this way), we expect the IBL 3D Lorentz angle, which is the $p1$ parameter of the fit (all the fit parameters are listed in the plot legends), to be close to zero.

The results of this study were included in the pixel detector calibration.

Table B.1: Lorentz Angle Values, summary.

	Lorentz Angle (rad)
IBL Planar	$2.502e^{-1} \pm 3.2e^{-3}$
IBL 3D	$2.338e^{-2} \pm 5.7e^{-3}$
Layer 1	$1.725e^{-1} \pm 2.5e^{-3}$
Layer 2	$2.051e^{-1} \pm 3.4e^{-3}$
Layer 3	$2.025e^{-1} \pm 2.8e^{-3}$

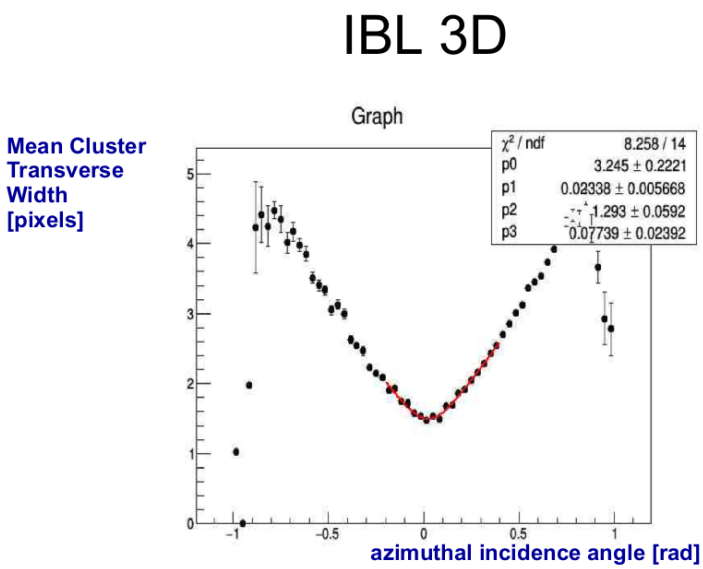

Figure B.2: IBL 3D, Lorentz angle. Incidence angle vs mean cluster size.

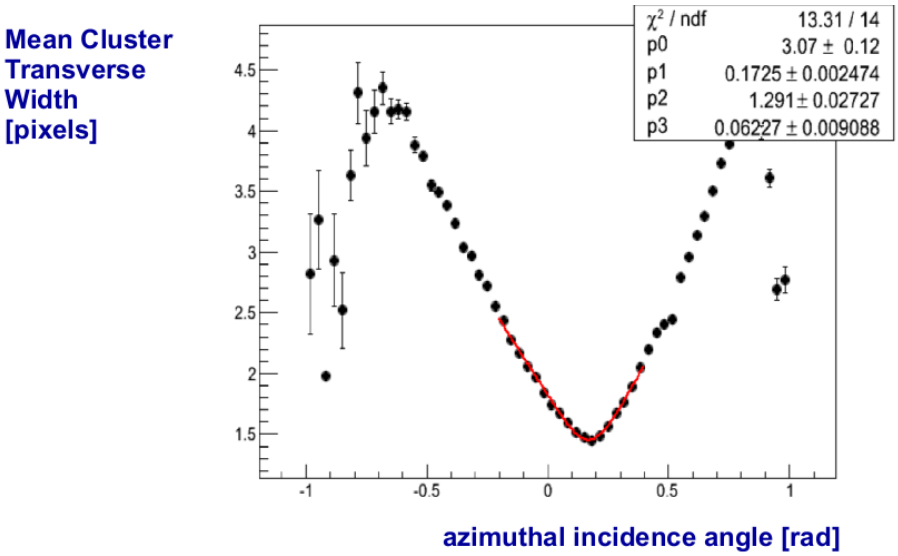

Figure B.3: Layer 1 (previously called B-layer), Lorentz angle. Incidence angle vs mean cluster size.

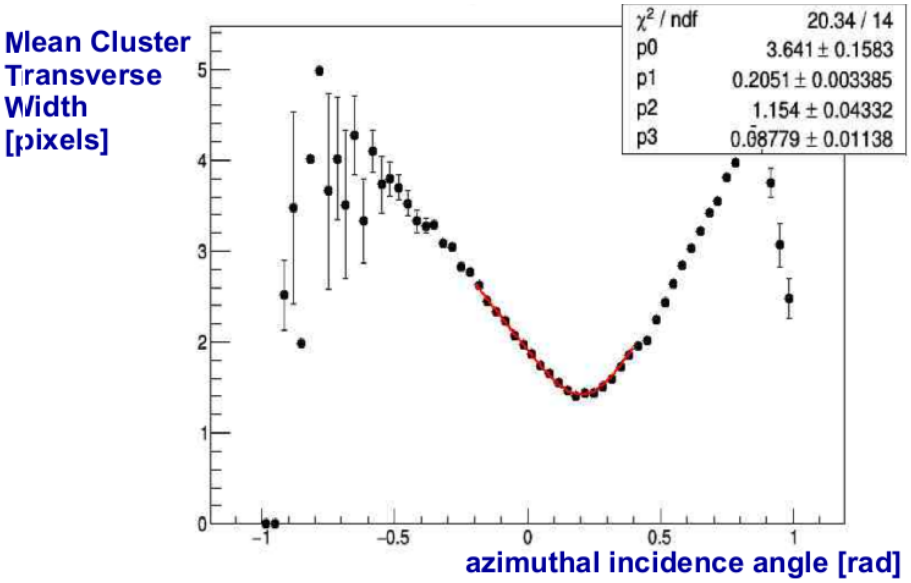

Figure B.4: Layer 2, Lorentz angle. Incidence angle vs mean cluster size.

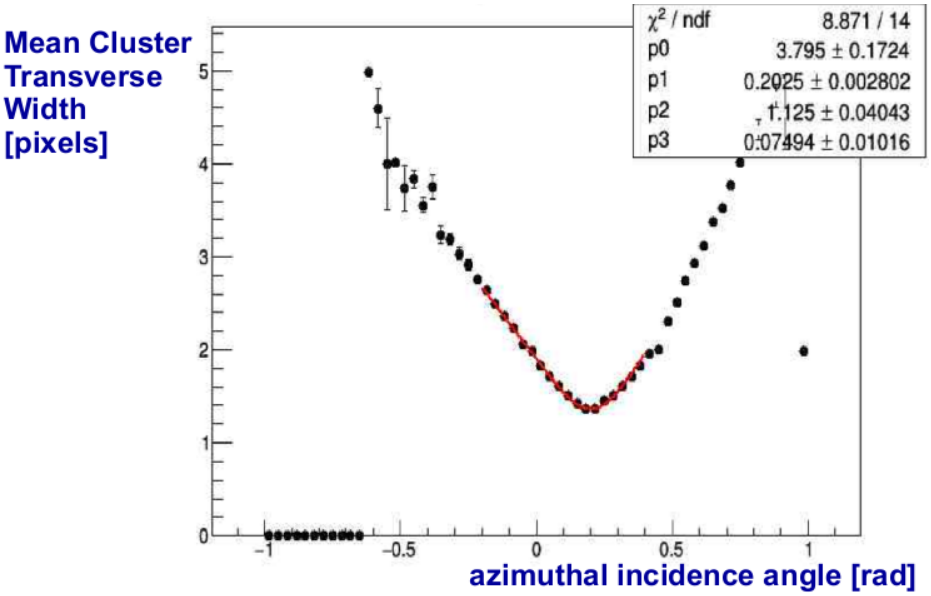

Figure B.5: Layer 3, Lorentz angle. Incidence angle vs mean cluster size.

Appendix C

R: useful MCMC analyzer

It could be of interest to study the Markov Chain provided by BAT in an another environment. Thanks to the CRAN [11] repository, R [28] proves to be a good choice. So much already made open source code!!

C.1 Convergence diagnostics

We expect our Markov Chains to converge to the target distribution, but we cannot be sure that they will converge after a certain number of draws. Just to keep it safe we can run a series of diagnostics tests. Thankfully the actual coding component of the problem is already no more. To put it bluntly, in order to go on with our analysis inside the R environment we just have to download from CRAN (Comprehensive R Archive Network, [11]) the CODA package [12] and importing it inside our R script, together with the package *mcmcplots* [29].

To see if a Markov Chain worked like we want it to do, it is useful to start from a visual way. What I mean is to check if the chain *mixed* well, so as to say, if it moved well inside the parameter space. It makes sense that if our chain is inside an area longer than expected, than it will take much longer to converge. This is possible to check through visual inspection thanks to *traceplots*, where you plot the iteration number against the actual value of the parameter you are checking. This needs to be done for every parameter. Let's consider fig.C.1, in the left side we see a good mixing, where the chain worked efficiently, while on the right side we can see that it got stuck somewhere.

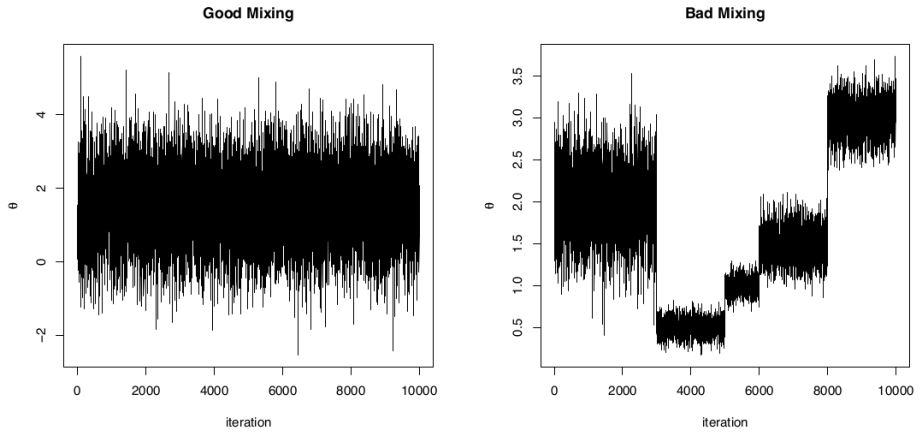

Figure C.1: Mixing, example.

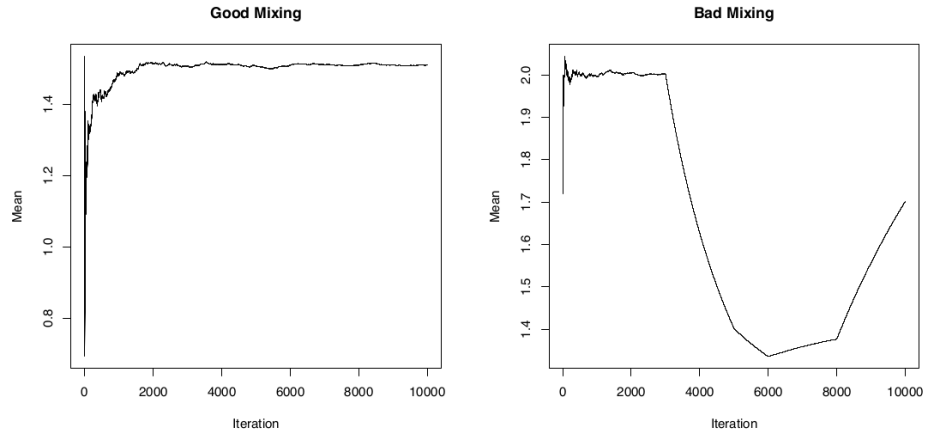

Figure C.2: Mixing through traceplots, example.

This is no good.

We can also use *running mean plots*, where we plot the iteration number against the mean up to each iteration of the parameter, in order to check how well did it go (figure C.2).

C.2 Gelman-Rubin convergence test

This is a more quantitative type of test. It is comprised of a few steps:

a) start running a number $m \geq 2$ of chains, $2n$ steps long, starting from m different and dispersed values;

b) discard the first n steps;

c) compute the within chain variance W;

d) compute the between chain variance B;

e) compute the estimated variance of the parameter as a weighted sum of the within-chain and between-chain variance;

f) compute the potential scale reduction factor.

We then define the variables involved:

$$W = \frac{1}{m} \cdot \sum_{j=1}^{m} s_j^2,$$

$$s_j^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (\vartheta_{ij} - \vartheta_{j,mean})^2,$$

where ϑ_{ij} is the i-ish value of the j-ish chain and $\vartheta_{j,mean}$ its mean value, so that s_j^2 is the variance!

$$B = \frac{n}{m-1} \cdot \sum_{j=1}^{m} (\vartheta_{j,mean} - \vartheta_{ALL,mean})^2,$$

where

$$\vartheta_{ALL,mean} = \frac{1}{m} \cdot \sum_{j=1}^{m} \vartheta_{j,mean}$$

but this means that B is the variance of the chain means times n. We can now define the estimated variance $V(\vartheta)$ of the chosen parameter as a weighted sum of W and B with respect to the number of considered iterations:

$$V(\vartheta) = (1 - \frac{1}{n}) \cdot W + \frac{1}{n} \cdot B.$$

The *shrink scale factor* R is defined as:

$$R = \sqrt{\frac{V(\vartheta)}{W}}. \tag{C.1}$$

When R is high enough, perhaps greater than 1.1 or 1.2 (BAT check if R is greater than 1.1), we should then run our chains longer to improve the convergence.

C.3 How to do the same in R

It should be fairly easy to do the same inside R. We want to have a script looking like:

```
#install.packages("coda")
#library(coda)
#install.packages("mcmcplots")
#library(mcmcplots)
```

[...] Let's suppose that here you have generated your Markov Chains or you have fetched them from file, and you have named them $mcNUMBER$:

```
mc.list <- mcmc.list(list(mc1, mc2, mc3, mc4, mc5))
```

Now we employ the function for the Rubin-Gelman diagnostics:

```
gelman.diag(mc.list)
```

We obtain then a summary of what has happened inside. We may be interested in a visual check. The function *gelman.plot()* spells out how the potential scale reduction factor changes through the iterations.

```
gelman.plot(mc.list)
```

That should give something like Fig. C.3:

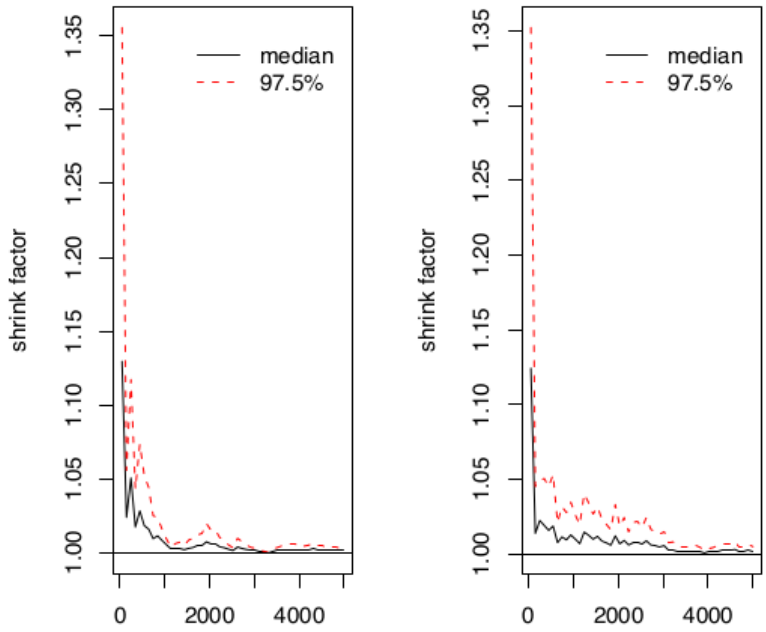

Figure C.3: R scale factor against the number of iterations, example.

Bibliography

[1] Giulio D'Agostini, Yellow Report, CERN, http://www.roma1.infn.it/ dagos/yr2.pdf

[2] A. Glazov, Averaging of DIS cross section data , AIP Conf. Proc. 792 (2005) 237 .

[3] H1 Collaboration, Measurement of the Inclusive ep Scattering Cross Section at Low Q^2 and x at HERA , Eur. Phys. J. C 63 (2009) 625 , arXiv: 0904.0929 [hep-ex] .

[4] BAT website, https://www.mppmu.mpg.de/bat/

[5] A. Caldwell, D. Kollar, K. Krninger, BAT - The Bayesian Analysis Toolkit, Comput. Phys. Commun. 180 (2009) 2197-2209 (ScienceDirect) [arXiv:0808.2552]

[6] CMarkup website, http://www.firstobject.com/dn_markup.htm

[7] Letters to vos Savant, http://marilynvossavant.com/game-show-problem/

[8] Lorenzo Bellagamba, Lectures slides, http://www.bo.infn.it/ bellag/didattica/2015/statistica.pdf

[9] Wikipedia, Metropolis Algorithm, https://en.wikipedia.org/wiki/Metropolis %E2%80%93Hastings_algorithm

[10] The ATLAS Collaboration, The ATLAS Experiment at the CERN Large Hadron Collider, JINST, 3, S08003, 2008

[11] CRAN, https://cran.r-project.org/

[12] CODA, MCMC analyzer, https://cran.r-project.org/web/packages/coda/coda.pdf

[13] ATLAS, http://arxiv.org/pdf/physics/0012050.pdf

[14] ATLAS, https://cds.cern.ch/record/1971961/files/ATL-INDET-PROC-2014-017.pdf

[15] ATLAS, Measurement of $W\pm$ and Z-boson production cross sections in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, arXiv:1603.09222

[16] S. Agostinelli et al., GEANT4: A Simulation toolkit, Nucl. Instrum. Meth. A 506 (2003) 250

[17] ATLAS Collaboration, The ATLAS Simulation Infrastructure, Eur. Phys. J. C 70 (2010) 823, arXiv: 1005.4568 [hep-ex]

[18] ATLAS Collaboration, Electron identification measurements in ATLAS using $\sqrt{s} = 13$ TeV data with 50 ns bunch spacing, ATL-PHYS-PUB-2015-041, 2015, http://cdsweb.cern.ch/record/2048202

[19] ATLAS Collaboration, Muon reconstruction performance of the ATLAS detector in protonproton collision data at $\sqrt{s} = 13$ TeV , Eur. Phys. J. C 76 (2016) 292, arXiv: 1603.05598 [hep-ex]

[20] ATLAS Collaboration, Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton-proton collision data , ATLAS-CONF-2014-032, 2014, url : http://cdsweb.cern.ch/record/1706245 .

[21] ATLAS Collaboration, Electron identification measurements in ATLAS using $\sqrt{s} = 13$ TeV data with 50 ns bunch spacing , ATL-PHYS-PUB-2015-041, 2015, url : http://cdsweb.cern.ch/record/2048202 .

[22] ATLAS Collaboration, Muon reconstruction performance of the ATLAS detector in protonproton collision data at $\sqrt{s} = 13$ TeV , Eur. Phys. J. C 76 (2016) 292 , arXiv: 1603.05598 [hep-ex] .

[23] M. Cacciari, G. P. Salam and G. Soyez, The anti-k_t jet clustering algorithm , JHEP 0804 (2008) 063 , arXiv: 0802.1189 [hep-ph] .

[24] ATLAS Collaboration, Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at $\sqrt{s} = 13$ TeV, ATL-PHYS-PUB-2015-015, 2015,http://cdsweb.cern.ch/record/2037613

[25] ATLAS Collaboration, Expected performance of missing transverse momentum reconstruction for the ATLAS detector at $\sqrt{s} = 13$ TeV, ATL-PHYS-PUB-2015-023, 2015, http://cdsweb.cern.ch/record/2037700

[26] ATLAS Collaboration, Improved luminosity determination in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector at the LHC, Eur. Phys. J. C 73 (2013) 2518, arXiv: 1302.4393 [hep-ex]

[27] ROOT, Data Analysis Framework, https://root.cern.ch/

[28] R, programming language, https://www.r-project.org/

[29] *mcmcplots* package, https://cran.r-project.org/web/ packages/mcmcplots/mcmcplots.pdf

[30] Central Limit Theorem, https://en.wikipedia.org/wiki/Central_limit_theorem