Università degli Studi di Trento
Dipartimento di Fisica

Doctoral Thesis
XXXII Cycle

Search for boosted Higgs bosons decaying to b-quarks at $\sqrt{s} = 13$ TeV with the ATLAS detector

Author:
Francesco Maria Follega

Supervisor:
Dr. Roberto Iuppa

Abstract

A search for the Higgs boson produced with large transverse momentum and decaying as H$\rightarrow b\bar{b}$ is presented, using the 80 fb^{-1} from the dataset collected by the ATLAS detector at $\sqrt{s}$=13 TeV. The Higgs boson is reconstructed as a large-R jet with two b-tagged variable radius track-jets. The work focuses on a peculiar event topology in which the Higgs is produced in association with another Large-R jet. Considering events with reconstructed p_T above 480 GeV and with a reconstructed mass from 70-230 GeV, a signal significance of 1.6σ for the Higgs and of 5σ for the V (Z boson + W boson) is observed. Furthermore, the new Full Run2 analysis is presented, using 136 fb^{-1} from the dataset collected by the ATLAS detector at $\sqrt{s}$=13 TeV. The strategy up to the computation of the expected sensitivity for the Higgs boson is presented in this thesis. The validation of the background modeling is described in details, including tests on data. A preliminary extraction of the Z boson signal is performed and the measured signal strength corresponds to $\mu_Z = 0.82 \pm 0.09$.

List of acronyms

ALICE - A Large Ion Collider Experiment

ASCOT - Apparatus with Super COnducting Toroids

ATLAS - A Toroidal LHC ApparatuS

BAT - Bayesian Analysis Toolkit

BDT - Boosted Decision Tree

BR - Branching Ratio

BSM - Beyond Standard Model

CL - Confidence Level

CCDF - Complementary Cumulative Distribution Function

CDM - Cold Dark Matter

CERN - Conseil Europeen pour la Recherche nucleaire

CKM - Cabibbo-Kobayashi-Maskawa

CLIC - Compact Linear Collider

CMB - Cosimic Microwave Background

CMS - Compact Muon Solenoid

CP - Charge Parity

CR - Control Region

CSC - Cathode Strip Chambers

DAQ - Data Acquisition

DM - Dark Matter

EAGLE - Experiment for Accurate Gamma, Lepton and Energy Measurements

EFT - Effective Field Theory

EM - ElectoMagnetic

EMCAL - ElectoMagnetic CALorimeter

EW - ElectroWeak

FCC - Future Circular Collider

FPGA - Field Programmable Gate Array

ggF - gloun-gloun Fusion

HADCAL - HADronic CALorimenter

HLT - High Level Trigger

IBL - Insertable B-Layer

ICEA- International Committee for Future Accelerators

ID - Inner Detector

ILC - International Linear Collider

IP - Impact Parameter

JMS - Jet Mass Scale

LEP - Large Electron-Positron Collider

LHC - Large Hadron Collider

LHCHXSWG - LHC Cross Section Working Group

LLR - Log Likelihood Ratio

LO - Leading Order

LSP - Lightest Supersymmetryc Particle

MC - Monte Carlo

MCMC - Markov Chain Monte Carlo

MDT - Monitored Drift Tubes

ME - Matrix Element

MSSM - Minimal SuperSymmetryc Model

NLL - Negative Log Likelihood

NLO - Next-to-Leading Order

NNLO - Next-to-Next-Leading Order

NP - Nuisance Parameter

PDF - Parton Density Functions

PDG - Particle Data Group

PE - Pseudo Experiment

PMT - Photo Multiplier Tube

PS - Parton Shower

QCD - Quantum ChromoDynamics

QED - Quantum ElectroDynamics

QFT - Quantum Field Theory

RMS - Root Mean Square

RPC - Resistive Plate Chambers

SCT - Semi Conductor Tracker

SLAC - Stanford Linear Accelerator Center

SM - Standard Model

SMEFT - Standard Model Effective Field Theory

SPS - Super Proton Syncrothron

SR - Signal Region

SSB - Spontaneous Symmetry Breaking

STXS - Simplified Template Cross Section

SUSY - SUper SYmmetry

TDAQ - Trigger Data Acquisition

TOTEM - TOTal Elastic and diffractive cross section Measurement

TRT - Transition Radiation Tracker

VBF - Vector Boson Fusion

WIMP - Weak Interacting Massive Particle

WP - Working Point

Contents

List of Figures

List of Tables

Introduction

Higgs boson searches dominated the particle physics scenario for decades [1–3]. This elusive particle has been deeply investigated in all its couplings with massive particles. Nowadays, some Higgs properties are known with high precision: mass, spin and parity. Despite that, many others are still largely unknown. Two of the most important aspects under investigation are the Higgs decay and production widths and the behavior of its coupling as a function of the transverse momentum. The outcome of these studies might confirm the Standard Model prediction, or, if statistically significant discrepancies are found, constitute a bridge to new physics. New measurements of these quantities can be used to constrain the contribution of beyond Standard Model theories to the Higgs production, such as the one given by Effective Field Theories (EFTs) [4, 5].

EFTs predict a large deviation from the Standard Model for the gluon-gluon fusion cross section manifesting at high p_T [6–8]. These effects can be probed by looking at Higgs boson produced with high transverse momentum decaying to b-quarks, which represents the dominant fraction of Higgs decays. Such a measurement presents many challenges: a poorer mass resolution, when compared to leptonic channels, a large background coming from QCD interactions and the presence of broad resonances nearby the Higgs peak.

This thesis reports about a novel and pioneering search for high momentum Higgs bosons, produced in association with a jet and decaying to a pair of b-quarks. It uses proton-proton collisions at center of mass energy of $\sqrt{s} =$ 13 TeV collected in 2015-2018 with the ATLAS detector at the Large Hadron Collider. The signature consists of two high p_T jets, recoiling against each other. One jet contains the $b\bar{b}$ system, very collimated because of the large Lorentz boost.

This work goes through two stages of the analysis: the first one focused on the measurement of the inclusive signal strength of low-mass mediators decaying to b-quarks and carried out with an integrated luminosity of 80 fb^{-1}; the second one focused on the Higgs boson, using the full Run2 dataset, which corresponds approximately to 140 fb^{-1}.

The first search was performed in the 70-230 GeV mass range and looking for objects with a p_T larger than 480 GeV. Besides the Higgs, a search for leptophobic Z' Dark Matter mediators in the 100-200 GeV mass range was performed.

Since only high p_T jets are considered, this work makes use of fixed radius jets with an R = 1, also called large-R jets, as signal candidates. The main issue of this study is the presence of large backgrounds, such as the QCD multijet production and the resonant ones, like the Z boson, the W boson and the $t\bar{t}$. QCD is by far the dominant background. To deal with it a parametric model has been developed and validated on a dataset logically orthogonal to the signal region, specifically designed for this purpose. After QCD, systematic uncertainties on the Z and W peaks most affect the sensitivity. For the Standard Model Higgs boson, the observed signal strength is $\mu_H = 5.8 \pm 3.1$ (stat.) $\pm$ 1.9 (syst.) $\pm$ 1.7 (th.), which is consistent with the background-only hypothesis within 1.6 standard deviations. No evidences of other significant excesses of events are found and limits on Z' production cross-section are set in the probed mass range. The Z+W (V) signal strength was fit to data simultaneously with the Higgs one. The observed signal strength is $\mu_V = 1.5 \pm 0.22$ (stat.) $^{+0.29}_{-0.25}$ (syst.) $\pm$ 0.18 (th.), corresponding to a significance of 5σ.

The second search is focused on the Higgs boson and targets two different results: (i) an inclusive measurement of the signal strength; (ii) a differential measurement of the signal strength in three p_T bins. A number of new solutions have been developed and implemented on top of the previous work, to improve the results more than what comes from the statistical factor. The first change was performed on the trigger strategy: new 2017 and 2018 unprescaled triggers with lower p_T thresholds, down to 450 GeV, have been introduced. This choice provides a large gain in signal acceptance and lowers the p_T threshold. The second improvement is the inclusion of subleading large-R jets as possible Higgs candidates. Looking at the subleading the global sensitivity of the analysis increases, especially in the lowest part of the p_T spectrum. Last, the introduction of a correction for jet mass and p_T, which takes into account the possibility that the b-quarks have a semileptonic decay. Furthermore, a new fitting framework able to fit at the same time a large number of signal and control regions was adopted, improving the signal extraction efficiency. Since the Full Run2 version of the analysis is still on the review process, no results on the Higgs boson will be given. The final Higgs unblinding has been delayed for reasons not in author's control. This thesis work contains a description up to the current state-of-the-art, i.e. the measurement of inclusive Z signal strength.

The *author's work* had an important impact on both searches. In particular, for the first one he worked on:

- studies of signal selection and event categorization;
- alternative signal selection studies including the D_2 substructure variables to select two-prong events;
- modeling studies on the V+jet MC templates and on the MC generator uncertainty, comparing Sherpa and Herwig++ generators.

Instead, within the second search he:

- performed studies of signal selection and event categorization;
- designed and validated the QCD background modeling with a parametric approach;
- performed spurious and signal injection test;
- calculated expectations for the inclusive and differential measurements using Asimov datasets;
- tested signal extraction on validation and signal regions, both for the V and Z processes.

Chapter 1

Standard Model and beyond

The Standard Model (SM) describes the nature of the elementary components of the matter and the forces which make them interact. It combines two theories: the Quantum Electrodynamics (QED) and the Quantum Chromodynamics (QCD). During the years it proved itself to be robust and very precise in the phenomenological predictions. Even though nowadays we take it for granted, using it as starting point for each new physics theory or new experiment, part of its beauty lies in the history of its foundation.
The Standard Model was the result of an extraordinary and extended joint effort of theoretical and experimental physicists all around the world. It was formulated in almost 50 years with many consecutive steps and not without errors, which played a big role in the final outcome. Four main ideas contributed the most to complete the theory:

1. **The idea of gauge symmetry on *SU(3)*.** It was originally proposed by Yang and Mills in 1954, by extending the previous idea of the simple U(1) gauge symmetry [9]. They built this theory on the group SU(3) of isotopic spin conservation, which became the theory of strong interaction. An interesting feature of their work was that the gauge group was non-Abelian and furthermore there were self-interactions among gauge bosons. The main drawback was that the gauge symmetry requires the bosons to be massless, therefore they should have been detected at that time. This work proved to be the foundation of the theory of strong interaction, namely the QCD.

2. **The quark model.** It was proposed independently by Gell Mann and Zweig in 1964 [10]. The idea is that hadrons are composed of elementary constituents called *quarks*, which shade light on the fast growing number of discovered resonances. In 1968 the experiment SLAC gave the first positive indication. It was found that electrons were often scattered with a large angle by nucleons. This phenomenum was interpreted with the fact that neutrons and protons are composed by point-like particles called *partons*, which will be identified afterwords as the aforementioned quarks.

3. **The spontaneous symmetry breaking (SSB)**. In 1964 Higgs [11], Englert and Brout [12] formulated this concept independently. The hypothesis is that there might be symmetries in the Lagrangian that break in the vacuum state. It was known, according to the Goldstone theorem [13] that for each spontaneous symmetry breaking there must be a spinless and massless boson, called the Goldstone boson. Unfortunately, none of these bosons had been observed in strong interaction physics. For that reason Higgs hypothesized that the Goldstone boson was the helicity-0 part of a gauge boson with a mass.

4. **The electroweak (EW) interaction.** This theory was developed by Weinberg [14] and Salam [15]. They built a gauge group $SU(2) \times U(1)$ in which the bosons are two massive particles called W and Z. The spontaneous symmetry breaking gives mass to these bosons and to the fermions and predicts the existence of the photon and the presence of a massive neutral scalar particle, namely the Higgs boson (still unobserved at that time). This work will allow the unification of electromagnetic and weak interactions, called afterwords EW.

The formulation of these ideas allowed the beginning of a great work on the experimental side. It took experimental physicists years to prove them. The most important experimental successes related to the topic are:

- the proof of parity violation in 1957 [16];
- the discovery of CP violation in 1964 [17].
- the discovery of the W and Z bosons and the measurement of their mass in 1983 [18, 19];
- the discovery of quarks and leptons of all the three generations, in years 1968-2000 [20–26];

1.1 Summary of particles and their interactions

The chief goal achieved by the SM is the correct prediction of the properties of the most elementary building blocks of the ordinary matter and the particle who mediates the forces between them. The former were identified as half-integer spin particles called fermions, the latter as integer spin particles called bosons. According to the relativistic quantum mechanics the existence of a fermion implies the existence of its antiparticle, with the same mass and opposite charge. Therefore, each fermion has its corresponding antifermion. These particles are further divided in two categories, leptons and quarks, further subdivided in three generations or families, fig 1.1. The reason for this division lies in the mass, which separates the leptons and roughly groups the quarks two-by-two. In addition to that, the CKM [27] coefficients strongly suppress decays of quarks belonging to a family to quarks belonging to another family,

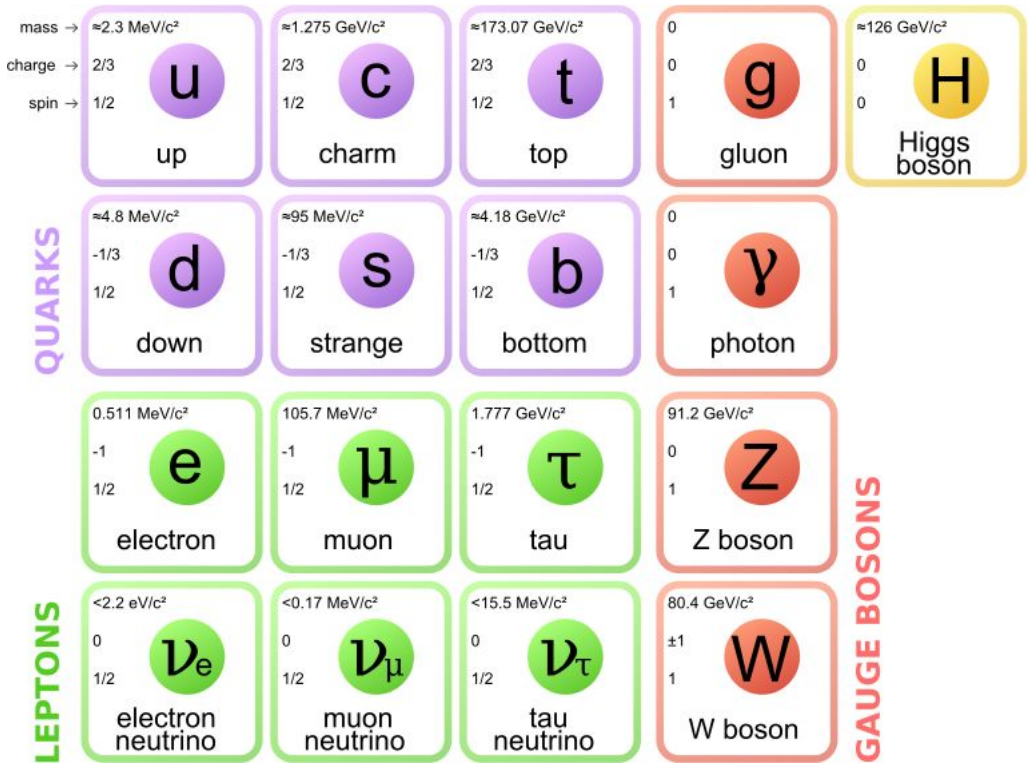


Figure 1.1: Scheme of fermions and boson predicted by the standard model. The first three columns are respectively the three fermions generations. The fourth column lists the interaction mediators and at the last place the Higgs boson responsible for the mass of all the aforelisted particles is added.

which creates a stronger division in generations. Leptons interact with other leptons and with quarks via the electromagnetic and weak interactions, while quarks interact with other quarks via the electromagnetic, weak and strong interactions. In nature, we only see first generation quarks and leptons, namely the u and d quark and the electron. Other generations leptons are unstable and quickly decay to more stable particles. The number of leptons is conserved for these decays, therefore an additive quantum number is assigned to these processes called leptonic numbers. As to the quarks, they can only be seen as bound state of a quark and an antiquark or as group of three quarks. The first state is defined as mesonic state while the second one as baryonic state. Mesons have integer spin and null baryonic number, whereas the baryons have half-integer spin and integer baryonic number. They are arranged into multiplets of spin and parity, forming a wide landscape of different resonances. Moreover, the baryonic number is conserved, an empiric evidence of that is the stability of the proton, $\tau_p > 10^{29}$ years. The reason why quarks cannot be isolated is that they carry another quantum number, called the color, introduced to ensure the Pauli principle in the mesonic and baryonic state. States with a color different from zero can not be observed in nature, this principle is called

Interaction	Boson	Range (m)	Gauge group
Strong	8 gluons	10^{-15}	$SU(3)_C$
Weak	Z^0,$W^{\pm}$	10^{-18}	$SU(2)_L$
Electromagnetic	γ	∞	$U(1)_Y$

Table 1.1: Table of all the interactions predicted by the Standard Model with their respective bosons [29].

the confinement rule [28]. As anticipated, the model describes all the interaction between particles, except for gravitation, and predicts the characteristics of the particles that carry the forces. These interactions differ in strength and range and are therefore mediated by different gauge bosons, tab. 1.1.

1.2 The Standard Model framework

The aim of the theory is to give a complete description of the fundamental particles and their interaction. It provides a unified model for the electromagnetic, weak and strong interactions. This model was realized to explain interactions at particle scale and therefore does not include gravitational interaction, which has a negligible contribution given such low masses. It makes use of a Quantum Field Theory (QFT) framework, requiring gauge symmetry. The Lagrangian formalism is used to describe particles and fields. After the definition of the Lagrangian density $\mathcal{L}$, which is a functional of the fields $\psi(x)$ and of their derivatives $\partial_x\psi$, the equations of motion are obtained by minimizing the action, defined as:

$$S = \int dx^4 \mathcal{L}(\phi, \partial\psi).$$

The form of the $\mathcal{L}$ is fixed by the requirement of gauge and relativistic invariance. The theory defined by this Lagrangian density is gauge invariant if the action S remains unchanged after the application of a continuous group of local transformation of the fields. Each symmetry of $\mathcal{L}$ under gauge transformation is associated to a conserved charge by the Neother's theorem [30]. As a consequence, conservation laws can be inferred by symmetries of the Lagrangian.

1.2.1 QED, EW and QCD

The original motivation of the gauge symmetry of the SM Lagrangian lies in the fact that it includes the QED. The electromagnetic potentials $\vec{A}$ and ϕ are not unique for a given $\vec{E}$ and $\vec{B}$ and the transformation of the potentials which preserves them are gauge transformations. A gauge transformation can be written as

$$A_\mu \rightarrow A'_\mu = A_\mu + 1/e\partial_\mu A_\mu$$

Left-handed quark	$\begin{pmatrix} u \\ d \end{pmatrix}_L \begin{pmatrix} c \\ s \end{pmatrix}_L \begin{pmatrix} c \\ s \end{pmatrix}_L$
Left-handed lepton	$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L$
Right-handed quark	u_R , c_R , t_R d_R , s_R , b_R
Right-handed lepton	e_R , μ_R , τ_R ν_R^e , ν_R^μ , ν_R^τ

Table 1.2: Table of all left-handed doubles and right-handed singles of weak isospin.

where $A_\mu = (\phi, -\vec{A})$ and $\partial_\mu = (\partial/\partial t, \vec{\Delta})$. The EM Lagrangian defined as

$$\mathcal{L}_{EM} = \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \quad , \quad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \,,$$

is unchanged under this transformation and it is therefore gauge invariant. This is only part of the full QED Lagrangian, which is defined as

$$\mathcal{L}_{QED} = \frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \bar{\psi}(x)[i\gamma^\mu D_\mu - m]\psi(x)$$

where $\psi(x)$ is the Dirac field, γ_μ the Dirac matrices and $D_\mu = \delta_\mu - ieA_\mu$ is the covariant derivative.

The other theory included in the SM is the electroweak theory (EW). It assumes that electromagnetic and weak interactions are different manifestation of the same force and thus unifies the $U(1)_Y$ with $SU(2)_L$. This theory predicts the non-conservation of parity in weak interaction, since the weak isospin current only couples to left-handed fermions. Therefore, two different spinors are introduced:

$$\psi_R = P_R = 1/2(1 + \gamma_5)\psi$$
$$\psi_L = P_L = 1/2(1 - \gamma_5)\psi,$$

where P_R and P_L are the chirality operators, respectively left- and right-handed. Consequently, the fermions are grouped in left-handed isospin doublets and right-handed singlets (tab. 1.2). For their single-handed nature they are invariant under isospin transformations. The gauge symmetry associated $SU(2)_L \times U(1)_Y$ translates into the conservation of the hypercharge defined as $Y = T_3 + Q$, where T_3 is the third component of the isospin vector. This theory also requires the presence of 4 bosons, two with charge and two neutral. Furthermore, the short range of the weak interaction requires that they are all massive, as $W^\pm$ and Z^0, while the infinite range of the electromagnetic requires its boson (the photon) to be massless.

The last building block of SM is the QCD. It describes the interaction between quarks and bosons of the strong interaction, the gluons, both having color. As a consequence, the theory is written in terms of colored states in the Lagrangian. The group describing the color symmetry is the $SU(3)_C$, where C

is the color carried by the particle. The theory predicts 8 bosons with 0-mass and with color. Since they carry color, they can interact with other gluons and with quarks. This gluon self-interaction causes the phenomenon known as asymptotic freedom. The coupling constant of strong interaction $\alpha_s = g_s^2/4\pi$, depends indeed on the momentum scale: it increases as the parton distance decreases and as a consequence there is color confinement in the hadrons.
Based on these properties the Lagrangian of the QCD can be written as the following

$$\mathcal{L}_{\mathcal{QCD}} = \frac{1}{4} G^a_{\mu\nu} (G^{\mu\nu})_a + \sum_k^{n_f} \bar{\psi}^k [i\gamma^\mu D_\mu - m_k] \psi^k \,,$$

where $D_\mu = \partial_\mu - i g_s T_a G^a_\mu$ and $G^{\mu\nu}$ is the strong field tensor (defined as $\partial_\mu G^a_\nu - \partial^a_\mu + g_s f^{abc} G^b_\mu G^c_\nu$, with f^{abc} to be the $SU(3)$ structure constant), T^a the generator of the group and n_f going from one to six (f = u,d,c,s,b,t). The aforementioned Lagrangian can be rewritten as the sum of three terms:

$$\mathcal{L}_{QCD} = \mathcal{L}_G + \mathcal{L}_q + \mathcal{L}_{Int}.$$

In this definition $\mathcal{L}_G$ is the self-interaction of the gluons, $\mathcal{L}_q$ is the mass part of the quarks and $\mathcal{L}_{Int}$ models the interaction within quarks and gluons.

1.2.2 The spontaneous symmetry breaking and the Higgs mechanism

The key for the electroweak unification and for the explanation of the mass of all fermions is a mechanism called Spontaneous Symmetry Breaking (SSB). Let us suppose to have an Abelian gauge invariant model with a local U(1) symmetry. The Lagrangian [11] will be the following

$$\mathcal{L} = \frac{1}{4} F_{\mu\nu} F^{\mu\nu} + (D_\mu \phi)^* D^\mu \phi - V(|\phi|^2),$$

where $\phi(x) = \phi_1 + \phi_2$ is a complex scalar field and $V(|\phi|^2) = -\mu^2 |\phi|^2 + \lambda |\phi|^4$ is a potential of the scalar field with $\lambda, \mu \in \mathcal{R}$, see fig. 1.2.
Given that, there are three possible solutions for the minimum of the potential:

1. $\lambda < 0$, then the potential has not boundaries and no stable minimum;
2. $\lambda > 0$ and $-\mu^2 > 0$, the function has a unique stable minimum at $|\phi_0| = \sqrt{(\phi\phi^\dagger)} = 0$;
3. $\lambda > 0$ and $-\mu^2 < 0$, the potential has two stable minima at $|\phi_0| = \mu/\sqrt{\lambda}$.

Since the Lagrangian is gauge invariant it means that there is a continuum set of fields that minimize the potential given by $(A^{vac}_\mu + 1/e\partial_\mu \alpha(x), \exp[i\alpha(x)\phi^{vac}])$. Then, choosing a particular $\alpha(x) = 0$ the configuration becomes

$$A^{vac}_\mu = 0 \quad , \quad \phi^{vac} = \phi_0/\sqrt{2}.$$

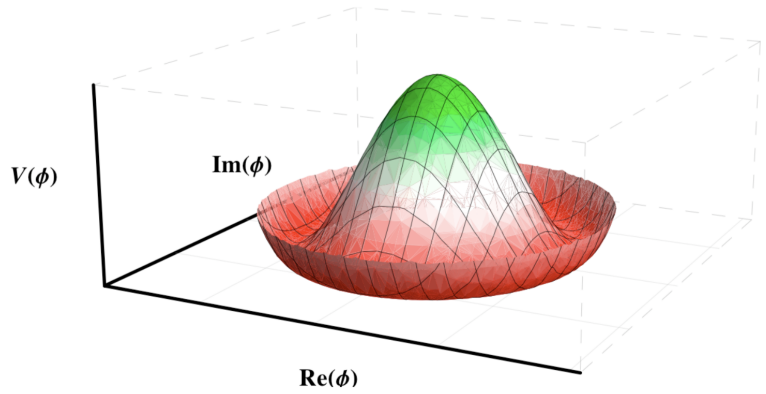


Figure 1.2: Illustration of the "Mexican hat" shape of the Higgs field potential as defined with μ^2,$\lambda > 0$.

The field can be rewritten as the sum of two real fields $\chi_1(x)$ and $\chi_2(x)$:

$$\phi(x) = \frac{1}{\sqrt{2}}(\phi_0 + \chi_1(x) + i\chi_2(x))$$

Substituting this field in the Lagrangian one gets a massive vector field B_μ with a mass equal to $m_V = e\phi_0 = e\mu\sqrt{\lambda}$ and a massive scalar field χ_1 with a mass equal to $\mu\sqrt{2}$, while the mass of χ_2 disappears. The scalar field is called Higgs field and the associated particle, the Higgs boson.

Masses of the gauge bosons

The SSB was used by S. Windberg [14] and A.Salam [15] to unify the electromagnetic and weak interactions. They theorized the SSB of the EW symmetry $SU(2)_L \times U(1)_Y$ with the Higgs mechanism. Before this passage, the Lagrangian can be written as

$$\mathcal{L} = -\frac{1}{4}F^a_{\mu\nu}F^{\mu\nu}_a - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} + (D_\mu\phi)^\dagger D^\mu\phi - V(\phi^\dagger, \phi)$$

where $F^a_{\mu\nu} = \partial_\mu W^a_\nu \partial_\nu W^a_\mu + g_2\epsilon^{abc}W_{\mu b}W_{\nu c}$, ϵ^{abc} is the structure constant of $SU(2)$ and $W^a_\mu(x)$ and B_μ are the gauge field of $SU(2)_L$ and $U(1)_Y$. The potential can be written as

$$V(\phi^\dagger, \phi) = \lambda(\phi\phi^\dagger - \frac{1}{2}v^2)^2.$$

In this case $\phi(x) = (\phi_1, \phi_2)$ is a weak hypercharge doublet and the covariant derivatives is defined as

$$D_\mu\phi = (\partial_\mu + g_2T_aW^a_\mu - ig_1YB_\mu)\phi$$

where T^a are the generators of the $SU(2)_L$ with coupling g_2 and Y is the generator of $U(1)_Y$ with coupling g_1. For the vacuum state we can take

$$W^a_\mu B_\mu = 0 \ , \ \phi^{vac} = (0v/\sqrt{2}),$$

and therefore $\phi(x)$ will be defined as,

$$\phi(x) = \big(0, (v + H(x))/\sqrt{2}\big)$$

Then we can introduce two complex and two real fields:

$$W^\pm_\mu = \frac{1}{\sqrt{2}}(W^1_\mu \mp iW^2_\mu)$$

$$Z_\mu = \frac{1}{\sqrt{g_1^2 + g_2^2}}(g_2 W^3_\mu - g_1 B_\mu)$$

$$A_\mu \frac{1}{\sqrt{g_1^2 + g_2^2}}(g_1 W^3_\mu + g_2 B_\mu)$$

such that $Z^2_\mu + (A_\mu)^2 = (W^3_\mu)^2 + B^2_\mu$.
After the redefinition the electroweak Lagrangian can be written as

$$\mathcal{L}^{(2)} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{2}W^{(+)}_{\mu\nu}W^{(-)\mu\nu} - \frac{1}{4}Z_{\mu\nu}Z_{\mu\nu} + m^2_W W^{(+)}_\mu W^{(-)\mu} + \frac{1}{2}m^2_Z Z_\mu Z^\mu + \frac{1}{2}(\partial_\mu H \partial^\mu H) - \frac{1}{2}m^2_H H^2.$$

It describes the following: a massless vector A_μ associated with the photon, a massive complex vector field $W^\pm_\mu$ with a mass $m_W = g_2 v/2$ associated with the W bosons, a massive real vector Z_μ with a mass $m_Z = v\sqrt{g_1^2 + g_2^2}/2$ and a real scalar field H with mass $m_H = v\sqrt{2\lambda}$. The parameter λ represents the Higgs self-coupling and, once fixed, defines the mass of the Higgs boson, while the g_2 and v determine the mass of the vector bosons. Furthermore, a weak mixing angle [27] can be introduced to connect the masses of the bosons

$$\cos\theta_W = \frac{g_2}{\sqrt{g_1^2 + g_2^2}}$$

this angle regulates the percentage of Z_μ and A_μ in the W^3_μ and B_μ, thus $m_Z = m_W/\cos\theta_W$.

Masses of the fermions

The fact that fermions are divided in left-handed doublets and right-handed singlets introduces a problem regarding the mass term of fermions, which can be written as

$$-m(\bar{\psi_L}\psi_R + \bar{\psi}_R\psi_L).$$

This term combine right- and left-handed terms and it is therefore not invariant under $SU(2)_L$. Hence, the way to give mass to fermions is using Yukawa

interaction coupled with the Higgs field. For instance, considering electrons and positrons we will have a Lagrangian as

$$\mathcal{L}^{e}_{Yukawa} = -g_e(\psi^e_L \phi e_R + \bar{e}_R \phi^\dagger \psi^e_L) = \frac{g_e v}{\sqrt{2}}(\bar{\psi^e_L}\psi^e_R + \bar{\psi^e_R}\psi^e_L) + \frac{g_e}{\sqrt{2}}(\bar{\psi^e_L}\psi^e_R + \bar{\psi^e_R}\psi^e_L)H = -m_{\bar{e}e} - g_{\bar{e}e}H$$

where g_f is the coupling constant between the Higgs field and the fermion spinors, f_R is the singlet of the fermionic field and ψ^L are the fermionic doublets. For quarks the Yukawa Lagrangian will appear as

$$\mathcal{L}^{q}_{Yukawa} = -\Gamma^u_{ij}\psi^{q,i}_L \phi_c \psi^{u,j}_R - \Gamma^d_{ij}\psi^{q,i}_L \phi \psi^{d,j}_R + [h.c.]$$

where Γ is the 3×3 Yukawa matrix and ϕ_c is the complex conjugate of the Higgs field.
Having excitation above the vacuum state there is a spontaneous symmetry breaking, and thus the quarks acquire mass. Their masses can be written as

$$M_f = V_{f,L}\Gamma_f (V_{f,r})^\dagger \frac{v}{\sqrt{2}}$$

where $f = u, d$ and v the vacuum expectation value. As a consequence, the fermions have an interaction with themselves and with the scalar field, the latter proportional to their mass $g_f = \sqrt{2}m_f/v$.

1.2.3 Higgs boson physics and characterization

The Higgs boson is one of the main actor of the SM, since it provides masses to bosons and fermions and creates a bridge between weak and electromagnetic interactions. Since it interacts with all massive particles, it is present in many terms of the Lagrangian density [29] :

$$\mathcal{L}_{H,Int} = -g_{Hf\bar{f}} + \delta_V V_\mu V^\mu (g_{HVV}H + g_{HVV}/2H^2) + \frac{g_{HHH}}{6}H^3 + \frac{g_{HHHH}}{24}H^4,$$

where V are the weak bosons Z and $W^\pm$, $\delta_W = 1$, and $\delta_Z = 1/2$. Since the coupling with the Higgs depends strongly on the mass, it is most likely that it couples with weak bosons and third generation fermions. The coupling with gluons and photons is also possible, but only via a fermionic loop.
In a hadron collider as the Large Hadron Collider (LHC), it is produced manly via four mechanisms [31], see fig. 1.3.

1. *gluon gluon fusion (ggF).* Two gluons fuse via a fermionic loop (t, b dominant). The cross section, calculated with perturbative QCD (N^3LOQCD), of this mode is 48.61 pb.

2. *vector boson fusion (VBF).* Two quarks irradiate bosons which fuse and produce the Higgs boson. The cross section, calculated with perturbative QCD (NNLO QCD and NLO EW), of this mode is 3.766 pb.

3. *Higgs strahlung (VH).* A W or Z boson emits a Higgs boson. The cross section of this mode, calculated with perturbative QCD (NNLO QCD and NLO EW), depends on the vector boson that is involved: for WH is 1.358 pb and for ZH is 0.880 pb.

4. $t\bar{t}H$. Two third generation quarks fuse and generate a Higgs boson. The cross section of this mode, calculated with perturbative QCD (NLO QCD and NLO EW), is 0.613 pb.

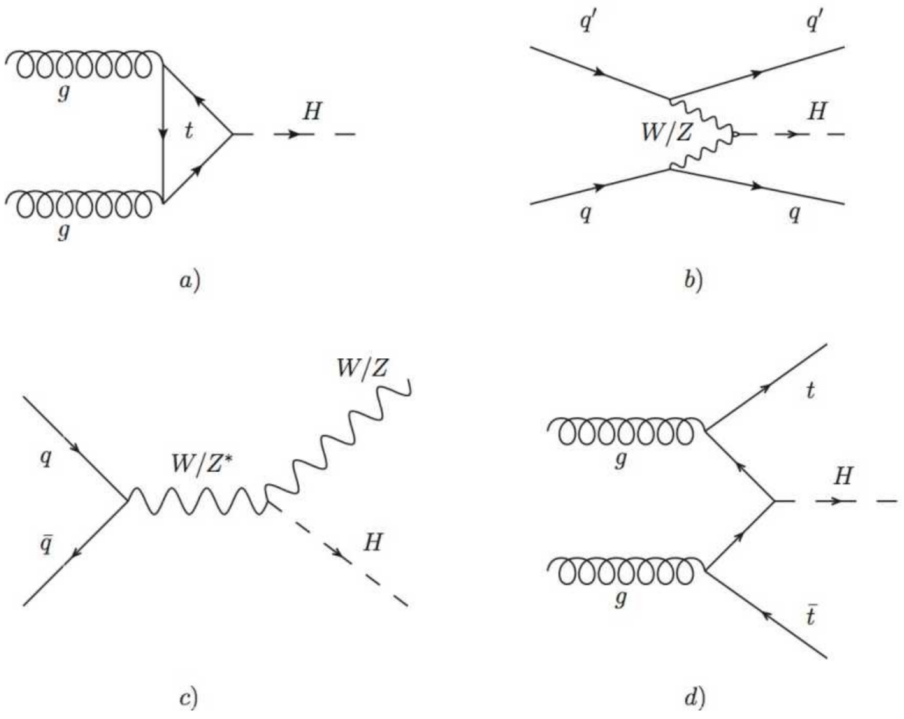


Figure 1.3: The four main Feynman diagrams for the Higgs production: (a) ggF, (b) VBF, (c) VH, and (d) $t\bar{t}$H.

The main mode is the gluon-gluon fusion, which is more than an order of magnitude above the others. After its production, the Higgs boson suddenly decays. Since the top quark it too heavy, the H has the largest branching ratio for decays to $b-$pairs, tab.1.3.

In July 2012, one year after the official start of the Large Hadron Collider (LHC), a new particle was discovered by ATLAS [32] and CMS [33] with mass of approximately 125 GeV. At that time the analysis was carried out with a dataset of 5 $pb^{-}1$ at center of mass energy of 8 TeV. The decay channels used for the measurement were the $H \to \gamma\gamma$, $H \to W^{+}W^{-}$, $H \to ZZ$, as they had the best signatures. An excess of signal was searched over the background mass

Decay Mode	BR(%)
$b\bar{b}$	58.24%
W^+W^-	21.37 %
gg	8.19%
$\tau^+\tau^-$	6.27%
$c\bar{c}$	2.89%
ZZ	2.62%
$\gamma\gamma$	0.23 %
$Z\gamma$	0.15 %
$\mu^+\mu^-$	0.02 %

Table 1.3: Table of main decaying modes of the Higgs boson, sorted by decreasing branching ratio [31].

spectrum of two back-to-back photons. Its significance was assessed by computing the p-value of a background fluctuation able to produce such a bump in the mass spectrum, as shown in fig. 1.4. The found signal strength μ, the ratio between observed and expected cross-section times branching ratio, was in good agreement with the SM, $\mu = 1.4 \pm 0.3$.

Higgs boson has been further investigated using the channels with the best

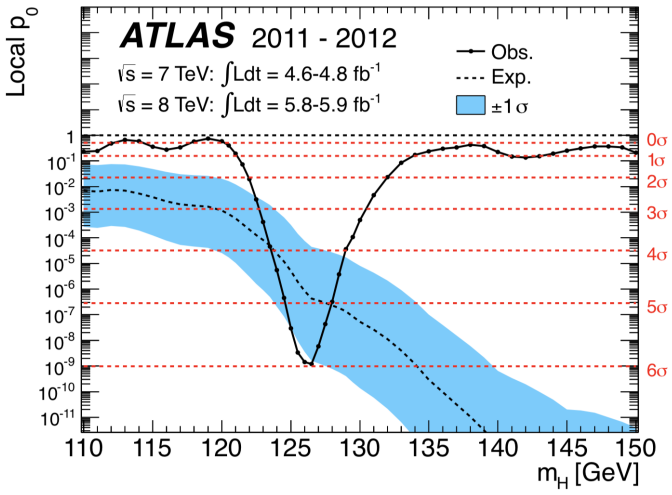


Figure 1.4: The p-value obtained summing 2011 and 2012 ATLAS dataset (10 pb^-1) and combining all the decaying channels $\gamma\gamma$, W^+W^-, ZZ [32].

mass resolution, namely $H \to \gamma\gamma$ and $H \to ZZ$ (mass resolution of 0.2%). By using both the channels with the ATLAS dataset of all Run1 (36 fb^-1) the measured mass was $m_H = 124.97 \pm 0.24$ GeV. The width of the Higgs boson is predicted to be of the order of 4 MeV. For that reason, it is too small to be directly measured by ATLAS or CMS, because they have a mass resolution of the order of 1 GeV. The SM predicts the Higgs boson to be a spinor with

$J^P = 0^+$ and in order to test SM hypotheses and alternative configurations the $\gamma\gamma$, $ZZ \to 4l$ and $WW \to l\nu l\nu$ have been used. All the alternative non-SM hypotheses have been excluded at 99.9% confidence level [34].

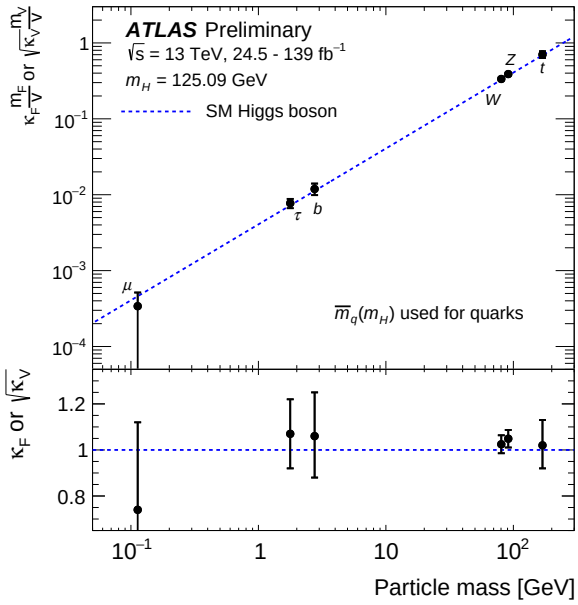


Figure 1.5: The reduced coupling strength modifiers $k_F \frac{m_F}{v}$ for fermions F and $\sqrt{k_v}\frac{m_v}{v}$ for weak gauge bosons V, for a vacuum expectation value of the Higgs $v = 246$ GeV, are shown in the figure. The dotted line corresponds to the SM predictions and the points to the measurements at the ATLAS detector.

All the measured Higgs production cross sections and branching ratios are in good agreement with the SM, see fig. 1.5. The global signal strength measured by ATLAS with 13 TeV is $\mu = 1.13^{+0.09}_{-0.08}$. Moreover, each observed decay branching ratio and production mode has a signal strength consistent with 1. Even tough each measurement concerning the Higgs boson turned out to be consistent with the SM hypothesis, this sector offers some hints for new physics searches and shows some problems of the current model:

- the coupling and Higgs self-coupling goes to infinite for finite energies [35]. This is difficult to prove because it happens at the plank scale.
- there is a really large mismatch between the mass of the boson and its bare mass [36]. Therefore, the bare mass and the radiative correction have to cancel up to plank scale, which requires an unnatural fine-tuning. The other explanation for this relatively low observed Higgs mass is the presence of new physics at a higher energy scale. These new objects might have radiative corrections which counterbalance those from the SM particles.

1.3 Theories beyond standard model

The SM was confirmed as extremely accurate and its predictions cover the large majority of physics phenomena at particle scale, as shown in fig. 1.6. Nevertheless, it leaves a few fundamental questions unanswered: what is the nature of the Dark Matter and of the dark energy; what is the reason for the unnaturally low vacuum expectation value; why matter is a factor 10^9 more abundant than anti-matter. Furthermore, some of the model predictions revealed to be different or at least far from the model results: the observation of neutrino flavor oscillations [37] demonstrates that they are massive particles and recent measurement shows some tension in lepton flavor universality [38] and in the muon magnetic momentum [39]. These shortcomings have moti-

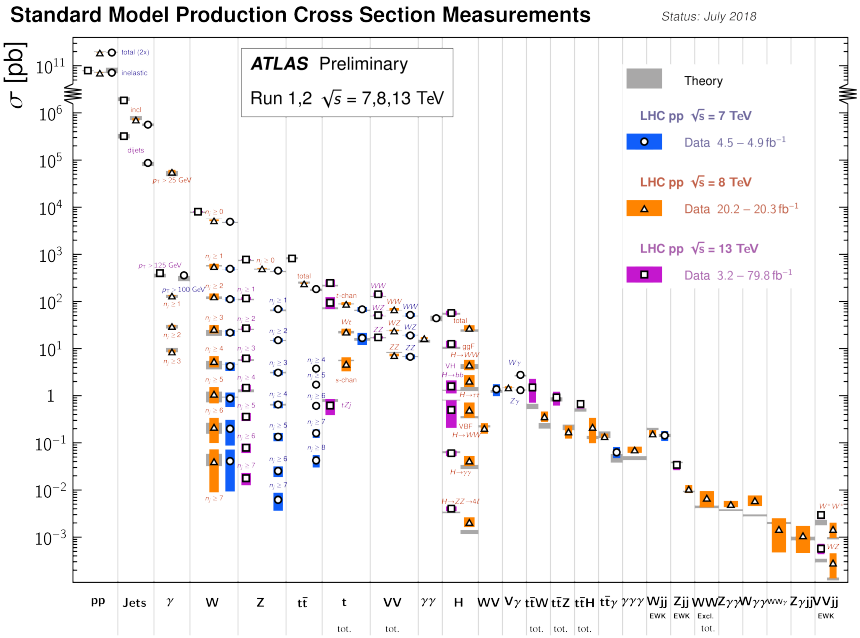


Figure 1.6: Summary of several Standard Model total and fiducial production cross section measurements, corrected for leptonic branching fractions, compared to the corresponding theoretical expectations. All theoretical expectations were calculated at NLO or higher.

vated a wide number of searches and measurements for new physics Beyond the Standard Model (BSM). The most relevant BSM models will be briefly described in the following paragraph.

1.3.1 Dark matter

The evidence of Dark Matter derives from astrophysical observation at different scales. All these measurements are indirect and a particle process including Dark Matter is still unobserved.

- **Galactic scales**. The amount of visible matter from stars and interstellar gas is not sufficient to explain the motions of the bodies going perpendicularly to the galactic disk. Since luminous matter is not enough to explain this behavior, an additional matter in the galaxy must be accounted for.

- **Inter-galactic scales**. In the study of the rotation curves of galaxies the luminous objects move faster than predictions, if they were only attracted by the gravitational force of other visible objects. If one considers just Newtonian gravity the velocity as a function of the distance from the galactic center is given by:

$$v(r) = \sqrt{\frac{G_N M(r)}{r}} \propto 1/\sqrt{r}$$

  where $M(r) = 4\pi \int dr\, \rho(r) r^2$ is the mass as function of the distance. This fact is contradicted by the observation, which supports a velocity which is constant with r, see fig. 1.7. Therefore, it was hypothesized a halo of matter surrounding the galaxy, identified as with the Dark Matter.

- **Cosmic Microwave Background (CMB)**. The Dark Matter is theorized by the Λ Cold Dark Matter (ΛCDM model). The analysis of the CMB temperature fluctuation spectrum allows a precise determination of the parameters of this model. The temperature anisotropies can be expressed as sum of harmonics:

$$\frac{\Delta T}{T} = \sum_{l,m} a_m Y_{lm}(\theta, \phi)$$

  where θ and ϕ are inclination and declination angles and $l = 0, 1, ...$ and $-l \geq m \leq l$ are the multipole moments. The position of the first, second and third peak of the power spectrum (fig. 1.8) are sensible to the density energy of the universe and the total, baryon and Dark Matter energy, can be extracted. The result of the fit divides the universe in three portions: 68.3 made of % Dark Energy, 26.8 made of % Dark Matter and made of 4.9 % ordinary matter.

One of the favored Dark Matter candidates is a particle called Weakly Interacting Massive Particle (WIMP). Moreover, some simplified supersymmetric models imply the presence of the Lightest Supersymmetic Particle (LSP). The latter is a stable particle with characteristics really close to the WIMP example. Therefore, direct searches for Dark Matter with collider experiments look for WIMPs in a restricted phase space. These searches can be performed in two different ways: indirectly, by high precision measurements of electroweak observables as new particles can contribute through loop corrections, or directly, by producing Dark Matter particles from the collision of two SM particles. It is assumed that Dark Matter is produced only via pair production and it traverses the detector without interaction. For that reason its only signature is

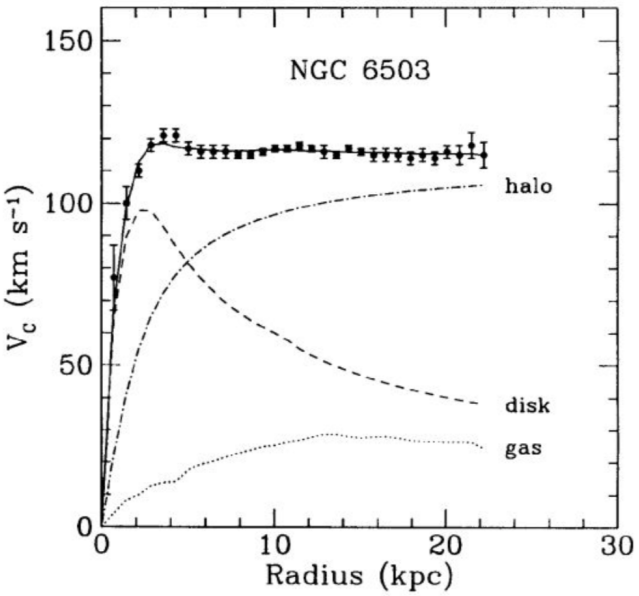


Figure 1.7: The rotation curve of the galaxy NGC 6503 compared with the predictions from visible and dark matter.

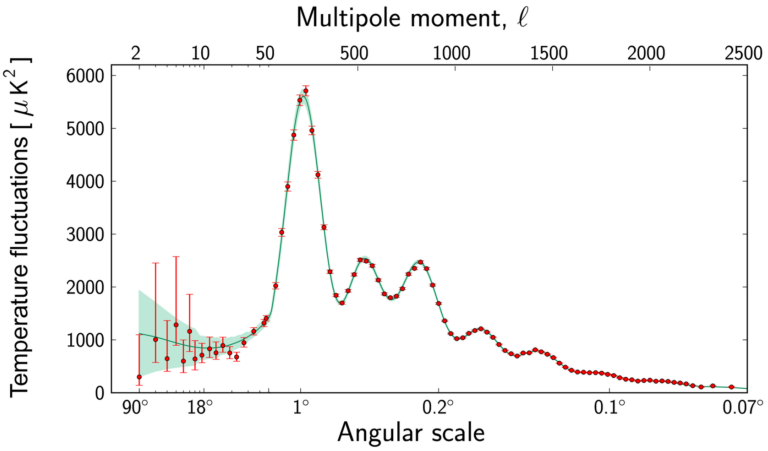


Figure 1.8: The Cosmic Microwave Background power spectrum measured by the Planck satellite, fitted with the prediction of the ΛCDM model [40].

missing energy E_T or large momentum imbalance. Such processes are referred as mono-X or E_{miss} + X signature. At the LHC there are many programs for Dark Matter detection led by ATLAS. For these studies the interaction between SM and Dark Matter particle is described by contact interaction either in Effective Field Theory (EFT) framework or by simplified models, see fig. 1.9.
In the EFT it is made the assumption that the new physics is suppressed at

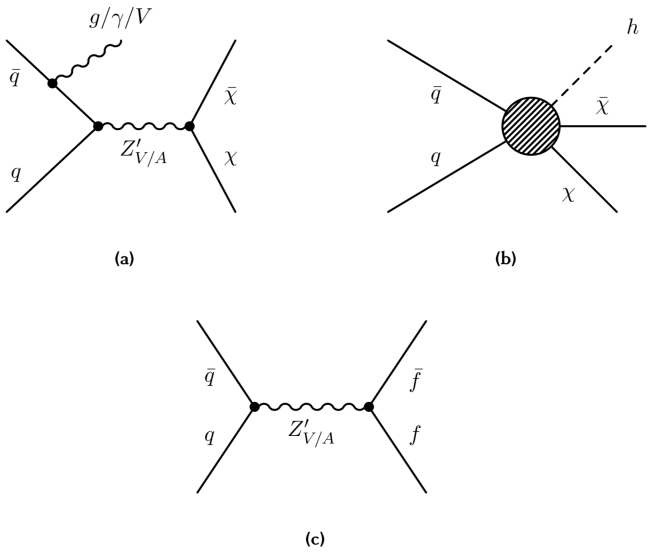


Figure 1.9: Feynman diagrams showing (a) the pair production of Dark Matter particles in association with the boson, from radiation of a quark or (b) a Higgs boson and (c) the production of fermion pairs via a vector or axial-vector spin-1 mediator.

the collision energy. For simplified models instead, new heavy mediators responsible for the interaction between Dark Matter and SM particles with a certain coupling structure are introduced. The Higgs sector is tied up with Dark Matter searches and can be used as starting point. One can look for invisible Higgs boson decays into Dark Matter. If the Higgs boson decays to Dark Matter there would be a missing decay width, under the hypothesis that the Higgs boson is the only connection between Dark Matter and SM particles. The only process contributing to invisible Higgs decay is $H \to ZZ' \to \nu\nu\nu\nu$, which is a contribution of 0.1 % to the total width of the Higgs boson. Currently, the best ATLAS observed (expected) upper limits on B($H \to inv.$) are 0.23 (0.24) [41]. Another channel important channel for Dark Matter searches linked with Higgs is the mono-Higgs searches [42].

1.3.2 Supersymmetry

Initially supersymmetry was introduced to explain why all particle interactions are ruled by the Poincare symmetries (semi-direct product between translation and Lorentz transformation), but only a fraction of them are manifest in nature [43]. Therefore, it was supposed that there is a larger symmetry group that states what interaction is allowed to show in nature and what is not. The most natural way to extend the Poincare group is with the spin symmetry, as it is the only degree of freedom which is not included. Therefore, this larger symmetry group was called supersymmetry and its generators can transform the spin of a particle, by summing or subtracting a fundamental spin unit $\frac{\hbar}{2}$, and thus it

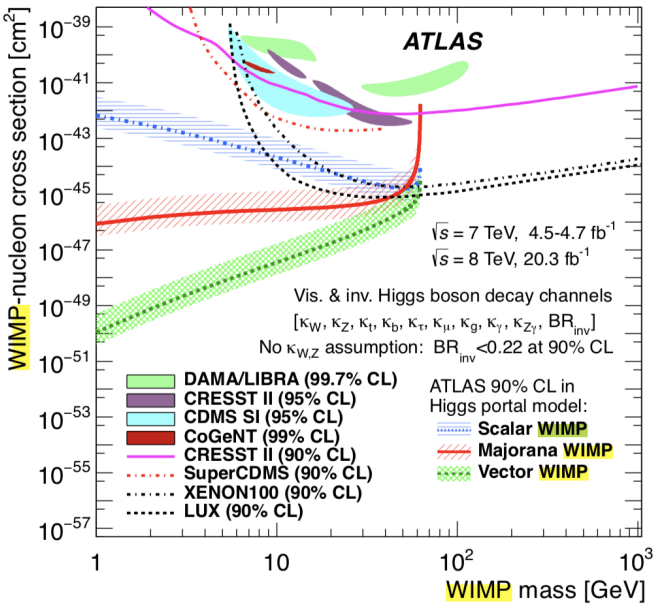


Figure 1.10: ATLAS upper limit at the 90% CL on the WIMP. Nucleon scattering cross section in a Higgs portal model as a function of the mass of the dark-matter particle, shown separately for a scalar, Majorana fermion, or vector-boson WIMP. It is determined using the limit at the 90% CL of BRinv < 0.22 derived using both the visible and invisible Higgs boson decay channel. Excluded and allowed regions from direct detection experiments at the confidence levels indicated are also shown.

changes fermions in bosons and vice versa. The operator responsible for this can be named as $\hat{Q}$ and its action defined as:

$$\hat{Q}|fermion\rangle = |boson\rangle \quad , \quad \hat{Q}|boson\rangle = |fermion\rangle \, .$$

Since the spin contributes to angular momentum and it must be conserved by this transformation, the operator itself needs to carry a spin of $\frac{\hbar}{2}$. The pair of particles linked by this operator are called supermultiplets and the corresponding fermions and bosons are called superfields. This theory introduces another quantum number, and therefore there are particles which are identical, except for this number and for the spin. They are called superpartners: squarks for the quarks and sleptons for the leptons. Two superpartners have the same mass and the same charge.

From its first formulation this theory was seriously considered by the scientific community, because it could naturally solve the hierarchy problem. The presence of supersymmetric partners contributes with opposite sign to the vacuum expectation value, providing the cancellation of the quadratic correction. However, this model showed two problems, one at the beginning and one only recently: the supersymmetric terms could make the proton unstable; no supersymmetry have been observed, even at the energy scale at which physicists

Supermultiplets	Spin 0	Spin 1/2	spin 1	number of fields
Higgs, Higgsino	$\begin{pmatrix} H_u \\ H_d \end{pmatrix}$	$\begin{pmatrix} \tilde{H}_u \\ \tilde{H}_d \end{pmatrix}$		$\begin{pmatrix} 2 \\ 2 \end{pmatrix}$
quark, squark	$\begin{pmatrix} \tilde{Q} \\ \tilde{u}_R^* \\ \tilde{d}_R^* \end{pmatrix}$	$\begin{pmatrix} Q_L \\ u_R^\dagger \\ d_R^\dagger \end{pmatrix}$		$\begin{pmatrix} 6 \\ 3 \\ 3 \end{pmatrix}$
lepton, slepton	$\begin{pmatrix} \tilde{L}_L \\ \tilde{e}_R^* \end{pmatrix}$	$\begin{pmatrix} L_L \\ e_R^\dagger \end{pmatrix}$		$\begin{pmatrix} 6 \\ 3 \end{pmatrix}$
B , bino		$\tilde{B}$	B	1
W, wino		$\tilde{W}$	W	1
gluon, gluino		$\tilde{g}$	g	1

Table 1.4: Summary of the field in a MSSM.

expected them. The first problem can be quickly avoided introducing an additional symmetry called R-parity [44], which requires the superpartners to be produced or annihilated in couples. The second caveat is more problematic, since supersymmetry should have been seen at the same energy scale of the corresponding SM particles. It was suggested that the supersymmetric particles are much heavier than their partners. From a theoretical point of view this can be achieved stating that the symmetry is only broken softly and that other terms giving extra mass to the supersymmetric particles, can be added.

The minimal supersymmetric model (MSSM)

The supersymmetry can be introduced in multiple ways into the SM. One of the simplest ones is the MSSM [45], which predicts the existence of other 41 fields, 33 are those that superfields matching the SM ones and the remaining 8 arising from a second Higgs doublet. In tab. 1.4, there is a summary of the field included in the minimal supersymmetric model. The supersymmetric Lagrangian can be written as:

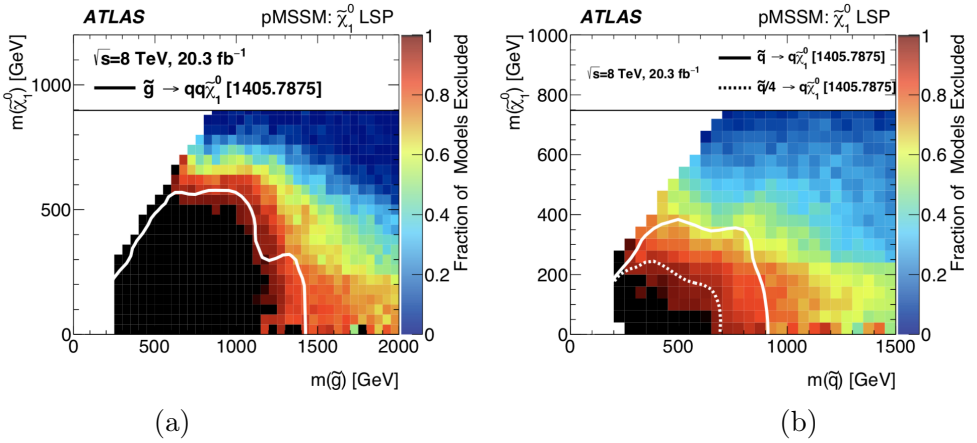
$$\begin{aligned}\mathcal{L}_{SOFT,MSSM} = -1/2(M_1\tilde{B}\tilde{B} + M_2\tilde{W}\tilde{W} + M_3\tilde{g}\tilde{g}) - \tilde{\tilde{u}}\hat{a}_u\tilde{Q}H_u - \tilde{\tilde{d}}\hat{a}_d\tilde{Q}H_d \\ -\tilde{\tilde{e}}\hat{a}_e\tilde{Q}H_e - \tilde{Q}^\dagger\hat{m}_Q^2\tilde{Q} - \tilde{L}^\dagger\hat{m}_L^2\tilde{L} - \tilde{\tilde{u}}\hat{m}_u^2\tilde{\tilde{u}}^\dagger - \tilde{\tilde{d}}\hat{m}_d^2\tilde{\tilde{d}}^\dagger - \tilde{\tilde{e}}\hat{m}_e^2\tilde{\tilde{e}}^\dagger \\ -bH_uH_d - m_{H_u}^2H_u^*H_u - m_{H_d}^2H_d^*H_d\end{aligned}$$

where a and m are complex 3x3 matrices. It generates a theory with 107 unknown parameters. This large number of parameters needs to be explored in order to prove MSSM as a model which describes nature. Of course, that is impossible to do exhaustively. Nevertheless, there are a lot of observations already done, that can be used to constrain the model parameters. The two most important constrained version of the MSSM are the cMSSM (constrained MSSM) and the pMSSM (phenomenological MSSM).

1. the cMSSM [46] makes many assumptions, in order to reduce the number of free parameters from 107 up to 5. The most important one states

that there is an energy (10^{16} GeV) at which the superpartners become mass degenerate. The remaining free parameters are $m_{1/2}$,m_0,A_0,B_0 and $sign(\mu)$. Because of the low model dimensionality it can be tested very easily with the observations already done. It did not turn out to play a crucial role in the supersymmetric physics at this stage.

2. the pMSSM [47] is based on fewer assumption and, in particular, it does not make any statement on the degeneracy of the superpartners at high energy. This model makes 5 assumptions: there are no new flavor changing currents, there are no new sources of CP violation at tree level, the first and second generation fermions masses are degenerate and that the lightest possible particle is the neutralino. This leaves the model with 19 free parameters: three gaugino masses, two higgsino mass parameter, ten sfermion masses, three trilinear couplings and the ratio between the two Higgs expectation value. This model captures most of the phenomenology of the MSSM and it has been already used by ATLAS in results interpretation [48], as shown in fig. 1.11.



(a) (b)

Figure 1.11: Fraction of pMSSM points excluded by the combination of 8 TeV ATLAS searches in the (a) $\tilde{g} - \chi_1^0$ and (b) the $\tilde{q} - \chi_1^0$ mass planes. The color scale indicates the fraction of pMSSM points excluded in each mass bin.

1.3.3 Standard Model Effective Field Theories (SMEFT)

One of the simplest ways to introduce deviation from the SM Higgs sector is the SMEFT framework [8]. In this approach we add higher dimensional terms to the SM Lagrangian, in first approximation dimension six operators are used. These terms are defined in such a way to be suppressed by powers of Λ, the New Physics (NP) scale energy, therefore small deviations from SM can be interpreted as NP phenomena. For what concerns the Higgs boson, this new operators modify the ggF production mode and the coupling with the gluon with the third generation quarks, see fig. 1.12. An example of effective

Lagrangian is given by,

$$\mathcal{L} = \mathcal{L}_{SM} + \sum_i \frac{c_i}{\Lambda^2}\mathcal{O}_i$$

where $\mathcal{O}_i$ are the dimension six operators, c_i are the coefficients which defines the magnitude of the effect brought by the operator and Λ is the new physics scale. The dimension-six operators may be expressed as,

$$\frac{c_1}{\Lambda^2}\mathcal{O}_1 \rightarrow \frac{\alpha_s}{\pi v} c_g h G^a_{\mu\nu} G^{a,\mu\nu}$$

$$\frac{c_2}{\Lambda^2}\mathcal{O}_2 \rightarrow \frac{m_t}{v} c_t h \bar{t}t$$

$$\frac{c_3}{\Lambda^2}\mathcal{O}_3 \rightarrow \frac{m_b}{v} c_b h \bar{b}b$$

$$\frac{c_4}{\Lambda^2}\mathcal{O}_4 \rightarrow c_{tg}\frac{g_s m_t}{2v^3}(v+h)G^a_{\mu\nu}(\bar{t}_L \sigma\mu\nu T^a t_R + h.c.)$$

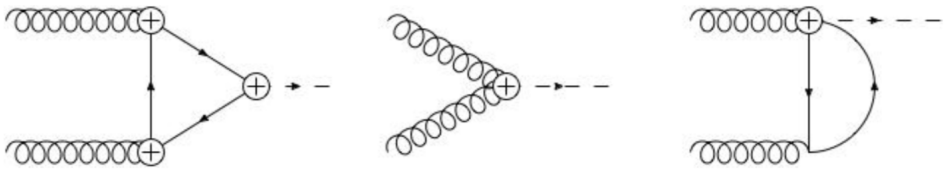


Figure 1.12: Feynman diagrams contributing to ggH production at LO. The possible insertions of dimension-six operators are marked by a cross in a circle.

$\mathcal{O}_1$ is a point-like operator that models the interaction between the Higgs boson and the gluon. $\mathcal{O}_2$ and $\mathcal{O}_3$ are modifications of the interaction between the Higgs with top and bottom quarks. $\mathcal{O}_4$ is the chromomagnetic dipole-momentum operator, which corrects the coupling between the top and the gluon. The parameters c_t, c_b, c_g and c_{tg} are called Wilson coefficients and can be tested studying processes involving the Higgs boson. The first two can be probed in processes such as ttH and bbH, while the c_g through the ggH production mode. The last one is constrained by the cross section of the top pair production.

By omitting the bottom operator contribution, the amplitude of the corresponding Feynman diagrams, see fig. 1.12, can be expressed as,

$$\mathcal{M}(g(p_1) + g(p_2) \rightarrow H) = i\frac{\alpha_S}{3\pi v}\epsilon_{1\mu}\epsilon_{2\nu}[p_1^\nu p_2^\mu + (p_1 p_2)g^{\mu\nu}]F(\tau)$$

, where $\tau = \frac{4m_t^2}{m_H^2}$ and ϵ_i are the polarization vectors of the gluons. The complete expression of $F(\tau)$ in the HLT limit, where $m_t^2 \gg m_H^2$, is

$$F(\tau) = \Gamma(1+\epsilon)\Big(\frac{4\pi\mu^2}{m_t^2}\Big)^\epsilon \Big(c_t + 12c_g(\mu_R) + 6Re(c_{tg})\frac{m_t^2}{v^2}(\ln(\frac{\mu_R^2}{m_t^2}) - 1)\Big)$$

The impact of these operators on the total Higgs production cross section is 20%. The contribution of c_t and c_g can be summarized as

$$\sigma \approx |12c_g + c_t|^2 \sigma_{SM},$$

therefore there is not a way to constrain one of them from the total cross section.

A possibility to constrain these parameters is given by the p_T differential cross section of the Higgs boson. The presence of the $\mathcal{O}_i$ operators strongly affects the computation of $\frac{d\sigma_H}{dp_T}$. The contribution of each operator can be varied through the corresponding coefficient as much as possible, while the other are fixed to SM values. The variation is made under the requirement that the total cross section does not deviate by more than 20% from SM prediction, fig 1.13. As expected c_b is the dominant contribution for the deviations from the SM

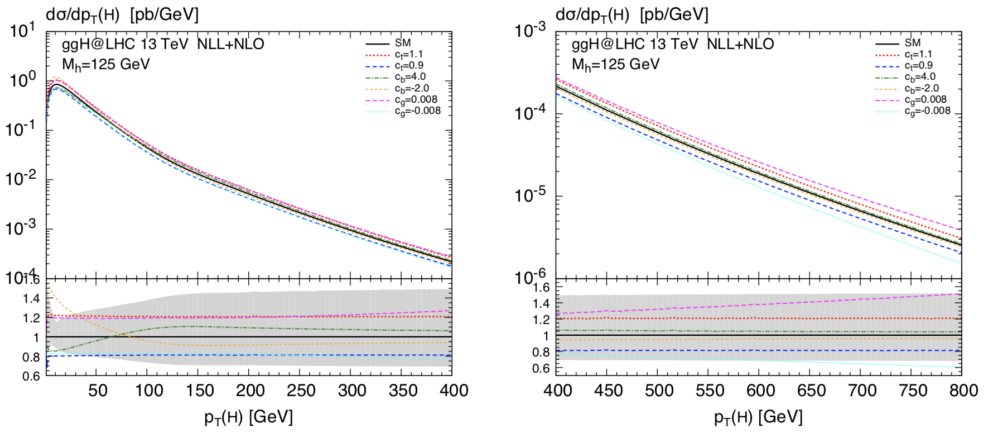
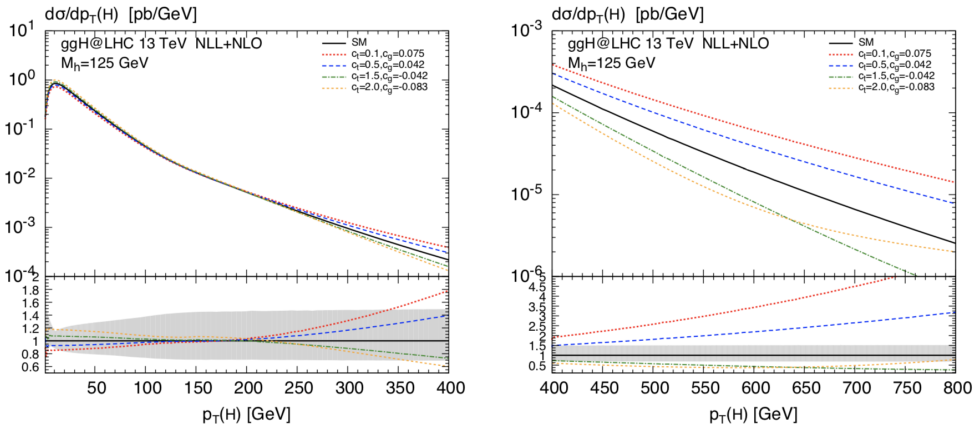


Figure 1.13: Higgs p_T spectrum in the SM (black) compared to separate variations of the dimension-six operators (dashed and colored). The shaded band in the ratio indicates the uncertainty due to scale variations.

at low p_T, the point-like Higgs-gluon coupling c_g becomes important at high p_T, while c_t just gives us a rescaling of the p_T spectrum. However, the best deviation from the SM p_T spectrum is given by the simultaneous variation of more coefficients, as shown in fig. 1.14. These large deviations can be detected at high p_T by studying the Higgs production via gluon-gluon fusion mode. Therefore, achieving a precise measurement of the Higgs production modes, especially ttH and ggH, would allow to asses the presence of SMEFT and of new physics BSM.



(a) (b)

Figure 1.14: Higgs p_T spectrum in the SM (black) compared to the ones obtained with simultaneous variations of c_t and c_g (dashed and colored). The lower frame shows the ratio with respect to the SM prediction. The shaded band in the ratio indicates the uncertainty due to scale variations.

Chapter 2

The Large Hadron Collider and the ATLAS detector

With the project of the Large Hadron Collider (LHC), CERN laboratories acquired a leading role in experimental particle physics, which nowadays is still undisputed. At the time of its first proposal there were many facilities spread around Europe and USA, see fig. 2.1.

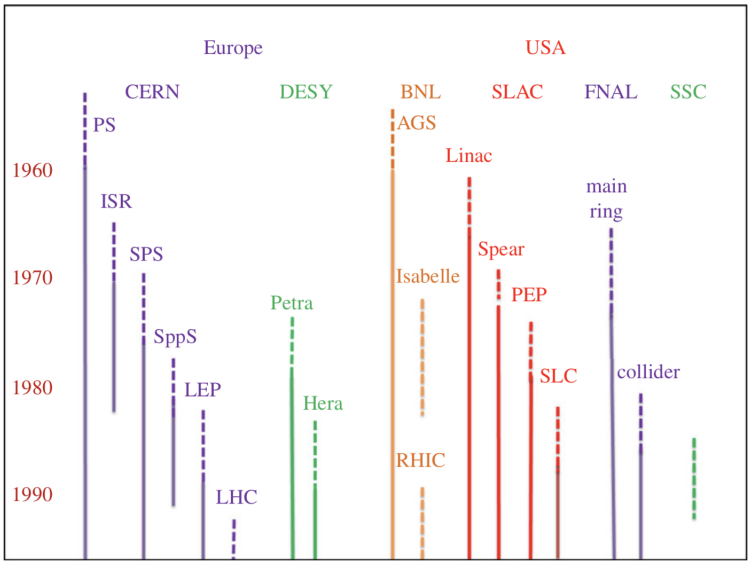


Figure 2.1: Major particle physics facilities in Europe and USA from the 1950s to the late 1990s. Construction and operational phases are indicated by dotted and solid lines, respectively.

Despite the uniqueness of the machine, the robustness of the scientific case and the unanimous support of the world's particle physics community, the approval and the construction of such a huge machine was not straightforward. Two key factors favored this project: it was the natural development of LEP (approved in 1981) and the well recognized CERN previous successes. For this reason the

importance of the scientific outcomes which made the LHC fame, is equaled by the conceptual and economic effort put in its construction by the member and not-member states, especially in the years before its first power-on (10th September 2008). Besides the economic and technological commitments, another difficulty was the competition with the Superconducting Super Collider, the American collider which should have reached 20 TeV center-of-mass energy, proposed in 1983, but abandoned in 1993 for founding problems.
The first public discussion on the overall theoretical case for the LHC took place in 1984 at the International Committee for Future Accelerators workshop (ICEA). It was presented as a machine able to provide "an unparalleled reach in the search for fundamental particles" and a "unique insights into the structure of matter and the nature of the Universe". The approval of the detectors progressed in parallel with the LHC's one. In July 1993, the Large Hadron Collider Committee evaluated the letters of intent to build general purpose detectors. Two of them were selected: ATLAS (A Toroidal LHC ApparatuS), resulted by the merging of two previous projects: ASCOT (Apparatus with Super COnducting Toroids) and EAGLE (Experiment for Accurate Gamma, Lepton and Energy Measurements), and CMS (Compact Muon Solenoid) to proceed to write technical proposals, which were approved in 1996. ALICE (A Large Ion Collider Experiment) and LHCb followed few years later.
The principal objectives of the particle physics at that time were:

- prove the existence of the Higgs boson or bosons, to answer the question of why some particle are massive and some other not;
- find supersymmetric particles, if they exist, revealing a deep connection between those who compose the matter and those who mediate forces between them.

The first goal was reached on the 4th July 2012, when in a joined conference ATLAS and CMS announced the discovery of a particle with a mass of 125 GeV, compatible with the one theorized by Peter Higgs, Robert Brout and Francois Englert. The second one it is still an open question, since no evidence of the presence of supersymmetric particles has been found, even though the energy range and the statistics collected are large.

2.1 The Large Hadron Collider

The LHC is the most energetic collider of the CERN accelerator complex, see fig. 2.2. It is located at CERN near Geneva and it is designed to collide proton beams with a center-of-mass energy of 14 TeV at an unprecedented luminosity of $2 \times 10^{34} cm^2 s^{-1}$. It receives bunches of protons by the Super Proton Synchrotron (SPS) and then accelerates them from 450 GeV up to 7 TeV in approximately 20 minutes. The particles are accelerated by superconducting radio-frequency cavities that operate at 400 MHz and they are bent in circular trajectories by a 8.34 T magnetic field, produced by dipole magnets. Such a

large magnetic field is provided by 1232 superconducting dipole magnets using niobium-titanium (NiTi) technology, which, in order to achieve superconductivity properties, are cooled down to 1.9 K.

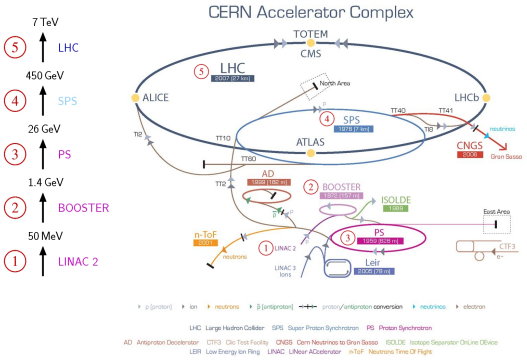


Figure 2.2: CERN Accelerator Complex and its main experiments [49].

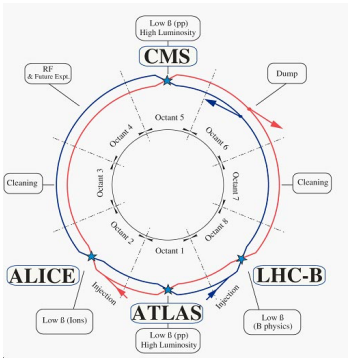


Figure 2.3: Schematic layout of the LHC. [49]

The superconducting NbTi Rutherford cables are coiled in two layers and distributed in six blocks. The two layers are surrounded by a unique stainless steel collar with two apertures. This component is necessary to ensure structural stability, by opposing the electromagnetic forces acting at the interfaces between the collared coils and the magnetic yoke with the forces set up by the shrinking cylinder. The basic layout of the LHC follows the LEP tunnel geometry with eight arcs and eight straight sections, as shown in fig. 2.3. Each straight section is 528 meters long and it can host an experimental or utility insertion. The arc sections of LHC lattice are composed by cells each 107 m long. The overall components length adds up to 27 km, located in an underground tunnel.

2.1.1 Luminosity

In a collider experiment one of the key parameters is the number of events per second generated by the collisions for a physics process under study. Its definition is:

$$N_p^{ev} = L\sigma_p.$$

where σ_p is the cross section of the given process and L is the luminosity of the machine. The LHC beams are not a continuum stream of protons going clockwise and anticlockwise within the cavities. The protons are indeed accelerated in several bunches, each containing 1.15×10^{11} protons. When the ring is completely full these bunches are spaced approximately by 7.48 m, which corresponds to a spacing time of 25 ns. This placement allows to store up to 2808 bunches in the LHC ring. When two bunches cross, the protons collide at an energy of approximately 13 TeV in the center of mass.
The transversal profile of a proton bunch can be approximated to a gaussian

distribution and therefore the luminosity $\mathcal{L}$ is:

$$\mathcal{L} = \frac{N^2 n_b f \gamma_r}{4\pi\sigma_x\sigma_y}$$

where N is the number of particles in the bunch, n_b the number of bunches stored in the beam, f is the revolution frequency, γ_r is the relativistic factor, $\sigma_{x,y}$ is the size of the beam along the x,y axis according to a gaussian distribution. However, since the two proton bunches are crossed with a certain angle and the transversal profile is not constant over all the bunch length, the formula is slightly different:

$$\mathcal{L} = \frac{N^2 n_b f \gamma_r}{4\pi\epsilon_n\beta^*} F$$

where ϵ_n is the normalized emittance, β^* is the beta function and F is a geometric factor accounting for the crossing angle. The LHC is designed to provide p-p collisions to the five main experiments located in the interaction points. They are:

- ATLAS and CMS. They are twin experiments working at $\mathcal{L} \sim 10^{34} cm^{-2} s^{-1}$, see fig. 2.4. They both investigate high energetic p-p collisions and explore physics at the Terascale;
- LHCb focused on b-physics. It reaches its efficiency at a luminosity of $\mathcal{L} \sim 10^{32} cm^{-2} s^{-1}$, lower than the nominal one, in order to have mainly just one collision per event;
- TOTEM (TOTal Elastic and diffractive cross section Measurement) built for high precision measurements of p-p interaction cross sections and designed for a deep study of the proton inner structure. It works at a luminosity of $\mathcal{L} \sim 10^{29} cm^{-2} s^{-1}$;
- ALICE, working at a luminosity of $\mathcal{L} \sim 10^{27} cm^{-2} s^{-1}$. This experiment differs from the previous p-p ones because it makes use of ion beams. It is optimized for heavy ion collisions, such as lead-lead collisions, in order to study the physics of strong interacting matter in very dense environments.

During the 9 years of operation, these experiments recorded a huge number of event from LHC collisions. This amount is usually quantified in terms of integrated luminosity, defined as

$$L_{int} = \int \mathcal{L}(t) dt.$$

The L_{int} is expressed in inverse barn (b^{-1}), which give us an intuitive way to evaluate the statistical power available with such amount of collected collisions, since it is the data required to observe one event of a process with $\sigma_p = L_{int}^{-1}$.

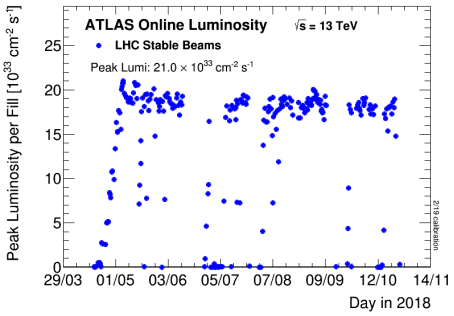


Figure 2.4: Peak of instantaneous luminosity delivered by the LHC to the ATLAS experiment in 2018

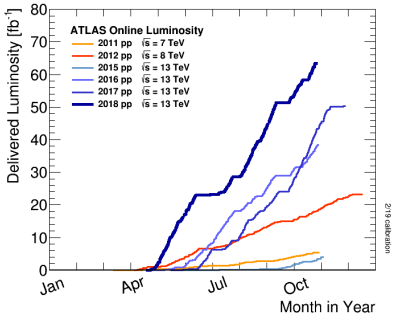


Figure 2.5: Integrated luminosity collected by the ATLAS experiment from 2011-2018.

In order to take into account the machine stops and the beam lifetime, the integrated luminosity is parametrized as:

$$L_{int} = L_0 \tau_L [1 - e^{-T_{run}/\tau_L}]$$

where T_{run} is the total duration of the run, and τ_L is the characteristic luminosity lifetime. Hence, the luminosity L_{int} strongly depends on the ratio between the duration of the run and the characteristic luminosity lifetime. This parameter quantifies the number of collisions available for physics analysis. At the end of 2018, ATLAS experiment recorded an integrated luminosity of approximately 140 fb^{-1}, see fig. 2.5.

2.1.2 Pile-up

When two bunches cross in one of the interaction points, multiple p-p collisions can occur due two the high beam densities reached by the LHC. This phenomenon is known as pile-up (μ) and with $< \mu >$ is indicated the average number of collisions for each bunch crossing.
There are two different categories of pile-up: in-time and out-of-time. The first one refers to interactions taking place in the same bunch crossing, the second one to those happening in subsequent bunch crossings. The pile-up makes more difficult the reconstruction of the physics objects necessary for the analysis. It also generates ambiguity in the primary vertex location and mistracking. The LHC increased the number of protons contained in each bunch to get more and more collisions per second, hence the importance has grown effect year-by-year (fig. 2.6). For that reason all the main experiments started to develop algorithms which are robust against pile-up and which disentangle from soft-scattering vertices.

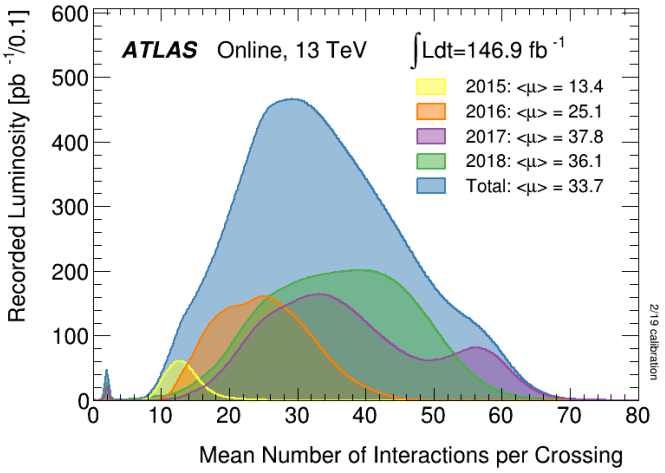


Figure 2.6: Luminosity-weighted distribution of the mean number of interactions per crossing for the 2015-2018 pp collision data at 13 TeV center-of-mass energy.

2.2 The ATLAS detector

The ATLAS collaboration was born in 1992, when the members of EAGLE and ASCOT collaborations joined to design and build a modern general-purpose detector for the LHC. In 1993 an official letter of intent was submitted to the LHC Committee and in 1996 the technical proposal was approved. The ATLAS detector was built in five years (from 2003 to 2008) thanks to the collaborations and the studies of a large number of physicists and technicians. It was designed to be the largest and complete particle detector of that time and therefore it was required to provide:

- a high space-time resolution, for tracking in dense environments;
- a good energy resolution and high granularity in the calorimeters;
- a fast electronics and information processing in such a way to pick only a small fraction of initial events and store just the interesting one;
- a strong magnetic field to measure the momentum of charged particles;
- good radiation-hardness especially for the innermost components.

The ATLAS detector is formed by several subsystems: the solenoidal and toroidal magnets; the tracking system, used to track particles and find vertices; the electromagnetic and hadronic calorimeters, employed to measure the energy of the collision products; the muon system, which detects muons [50].

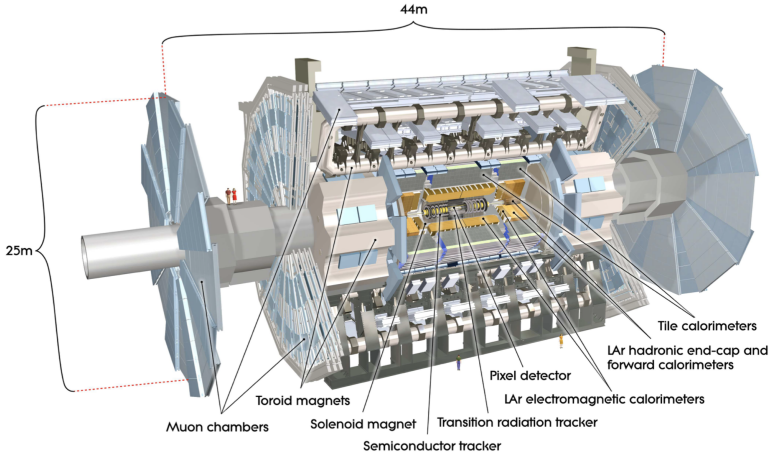


Figure 2.7: Picture of the ATLAS detector and its subsystems [50].

2.2.1 Geometry and coordinate system

A schematic view of the ATLAS detector is shown in figure 2.7. It is symmetrical around the beam axis with a diameter of 25 meters, a length of 45 meters and a total weight of 7000 tons. It uses a right-handed coordinate system where the z-axis is aligned to the beam pipe, the x-axis points toward the center of the ring and the y-axis is directed upwards. The angular coordinates are θ and ϕ: the former is the polar angle with respect to the z-axis and the latter is radial angle around the z-axis. In collider experiment instead of θ the rapidity y is used, defined as

$$y = \ln(\frac{E + p_L}{E - p_L})$$

where E is the energy of the particle, p is the momentum and p_L is longitudinal component of the momentum. This variable is preferred to θ because it is a Lorenz invariant and at high energies it can by approximated by the pseudorapidity η,

$$\eta = -\ln(\tan(\theta/2)) = \ln(\frac{p + p_L}{p - p_L})\,.$$

In this system a new distance between tracks or particles is defined exploiting the new coordinates. Its definition is the following,

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}.$$

2.2.2 Magnet System

ATLAS is provided with a magnetic system [50] to generate high fields able to bend extremely energetic particles produced by LHC. In particular, ATLAS detector is equipped with a hybrid system of two superconducting Al stabilized NiTi magnets (fig. 2.8): a solenoidal magnet and a toroidal magnet. The whole

system has an overall dimension of 22 m in diameter times 26 m in length and it provides a high magnetic field in a volume of 12000 m^3 ($0 < |\eta| < 2.7$). Since it makes use of superconducting magnets, it needs to be cooled down at approximately 4.5 K. The cooling is provided by a dedicated cryogenic system which supplies the whole apparatus.

The solenoid is the innermost and it has a cylindrical symmetry around the beam axis. It was designed to generate a magnetic field in the Inner Detector and at the same time to minimize the amount of material in front of the calorimeter (0.66 radiation length). It provides a magnetic field of 2T at an operational current of 7730 kA.

The outermost is the toroid, divided in a barrel and two end-caps and it produces a toroidal field of 0.5-1 T. They supply the calorimeter and the muon system, and they consist in eight coils inserted in stainless-steel vacuum containers, with a length of 25.3 m each and a diameter of 9.4 m for the inner and 20.1 m for the outer.

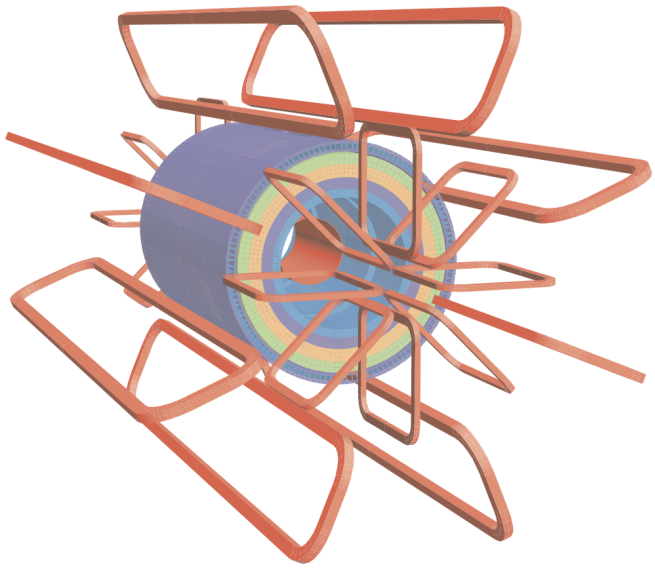


Figure 2.8: View of the magnet windings and tile calorimeter steel. The eight barrel toroid coils, with the end-cap coils are visible, while the solenoid winding lies inside the calorimeter volume. [50].

2.2.3 Inner Detector

The inner detector (ID), shown in fig. 2.9, is used to track particles and reconstruct vertices [50]. It contributes to the identification of charged particles like electrons, quarks and muons. The tracking procedure and the measurements of the transverse momentum and energy are possible thanks to the strong

magnetic field, which bends particle trajectories. The momentum p can be extracted by the bending radius ρ as follows [50]

$$p = \frac{qB\rho}{c}$$

where c is the speed of light and q the charge of the particle.

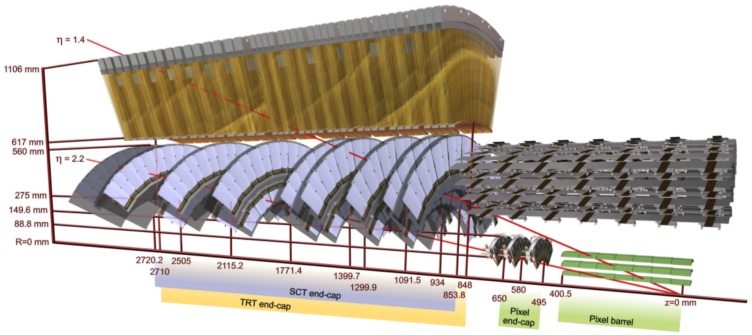


Figure 2.9: Drawing showing the sensors and structural elements of the ATLAS ID, traversed by two charged tracks of 10 GeV p_T [50].

The ID extends between $-2.5 < \eta < 2.5$ and it has a symmetric structure in η and ϕ. It has a spacial resolution on the impact parameters < 15 μm and on vertices < 1 mm, which enables it to identify the flavor composition of the particle jets (flavor tagging). It is composed of many subsystems, starting from the center they are: the pixel detector, the semiconductor tracker and the transition radiation tracker. Therefore, at inner radii, silicon pixels and pairs of silicon microstrip layers provide high-resolution pattern recognition capabilities, which give discrete space points. At larger radii, the transition radiation tracker gives the hit positions with several layers of gaseous straw tube elements interleaved with transition radiation material. All the system produces on average 36 hits per track, if the particle transverses all the ID material.

The pixel detector is the closest module to the interaction point located at radii 50.5 mm to 194.6 mm (fig. 2.10) [50]. It is composed by 1744 identical silicon pixel sensors (external dimensions 19x63 mm^2 each) providing high granularity while measuring charged particles traversing at least three layers for trajectories with $|\eta| < 2.5$. Each sensor is 250 μm thick and uses oxygenated n-type wafers with readout pixels on the $n^{\pm}$ implanted side of the detector to ensure a good charge collection with different voltages and a good radiation tolerance. The nominal pixel size is 50×400 μm^2 (about 90% of the pixels) due to the readout pitch of the front-end electronics. There are 47232 pixels on each sensor for a total of 46080 readout channels. The pixels are located in a central barrel and in two end-caps, each of them arranged in three subsequent layers. This geometry allows the pixel detector to guarantee at least three space points over the full tracking pseudo-rapidity range.

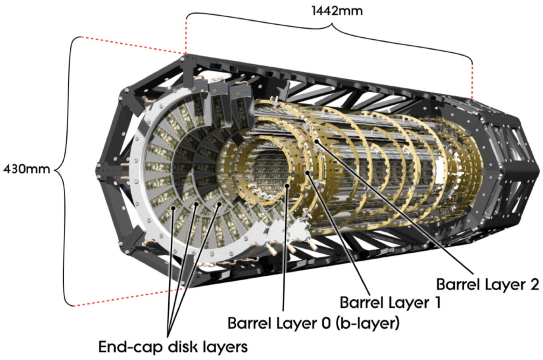


Figure 2.10: Cut away of the Pixel Detector: barrels and end-caps [50].

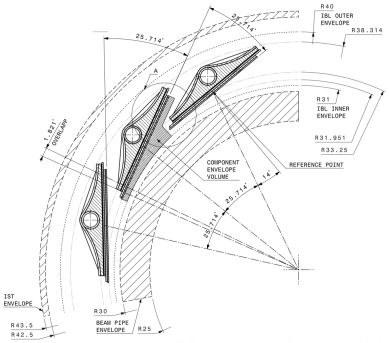


Figure 2.11: IBLpixel layout: $r\phi$ view [50].

In May 2014 the insertion of another pixel layer, called Insertable B-Layer (IBL) was performed. This layer targeted early decaying particle identification and tracking and brought large improvements in track and vertex reconstruction as well as in b-tagging performance [51]. As shown in fig.2.11, it is the innermost detector and it consists of 14 carbon fiber staves each 2 cm wide and 64 cm long, and tilted by 14° in ϕ surrounding the beam-pipe at a mean radius of 33 mm. Since this module is the most exposed to the LHC radiation it was built to be substituted with another module before its performances degrade too much. Nevertheless, it has been qualified to absorb a total ionizing dose of 300 Mrad, which means it can collect 300 fb^{-}1.
The semiconductor tracker (SCT) uses a classic single-sided p-in-n technology with AC-coupled readout strips. It is formed by four barrel layers ($1.1 < |\eta| < 1.4$) and nine disks in each of the two end-caps, extending from 255 mm to 560 mm starting from the center. Each sensor is $285 \pm 15\,\mu m$ thick and is composed by 768 active strips of 12 cm length operating at 250-350 V (plus 2 strips at bias potential at the edges), with a pitch of 80 μm. This detector allows to reconstruct the positions of charged particles with an accuracy of 17 μm per layer, in the direction transverse to the strips. The total number of readout channels in the SCT is approximately 6.3 million.
The transition radiation tracker (TRT) is divided in barrel and end-caps, formed by Polyimide drift straw of 4 mm diameter. It extends radially from 563 to 1066 mm and axially in z-direction until 712 mm (barrel) and 2710 mm (end-cap) and covers a volume of $12m^3$ (η<2). In the barrel there are 50000 straws, each 144 cm long, while in the end-caps there are 250000 straws, each 39 cm long. The detector consists of drift tubes with a diameter of 4 mm in which a gas mixture of 70% Xe,27% CO2 and 3% O2 is flushed with 5 to 10 mbar. For both the barrel and end-cap straws, the anodes are 31 μm diameter tungsten (99.95%) wires plated with 0.5-0.7 μm gold. The presence of Xenon is essential to combine tracking information of the detected hit position with the particle nature ones. The Xenon gas can indeed absorb the low-energy tran-

sition radiation photons emitted by passing particles (i.e. electrons, pions), yielding to a much higher signal amplitude than minimum-ionizing particles. The TRT distinguishes between tracking signals and transition radiation signal on a straw-by-straw basis, by using separate low and high thresholds in the front-end electronic.

2.2.4 ID Alignment

The Inner Detector modules are subject to large thermal excursion (especially the ones close to the beam pipe, i.e. IBL) and mechanical stress induced by the high magnetic fields. These agents may cause small shifts from the nominal module position or even changes to the shape of the detector modules, producing misalignments at the micrometer scale. Since the intrinsic resolution of the pixels is $\sim 10\ \mu m$ along the r-ϕ direction and $\sim 115\ \mu m$ along z direction, those misalignments have a not negligible impact on the tracking performance. The effect produced are well represented by the sketch in figure 2.12.

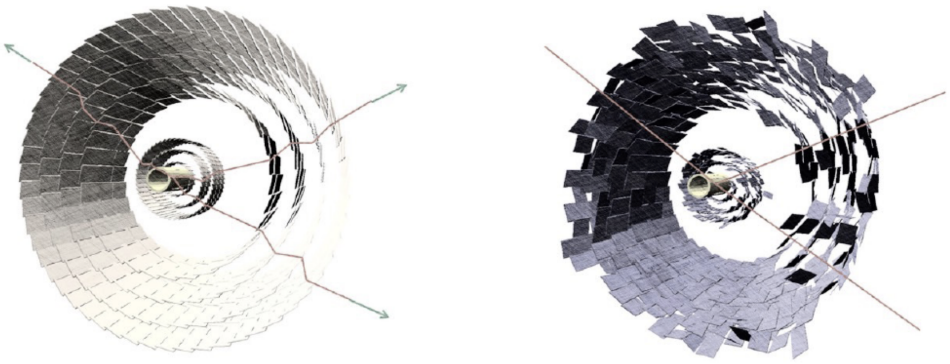


Figure 2.12: Sketch of the nominal ID geometry (a) and of the actual ID geometry after several deformations (b). On the left, tracks appear distorted because the nominal position of the modules is considered, while on the right, the trajectories have the expected behavior since the actual geometry is taken into account.

Hence, a dedicated alignment procedure has been developed to correct for these distortions and to determine the actual detector geometry [52]. This method is based on the minimization of the residuals between the reconstructed track and the hits used to build it. Therefore, if $\mathbf{r}(\mathbf{a}, \tau)$ is the track-hit residuals vector of a given track, with a as the alignment constant and τ as the track parameters, the χ^2 can be defined as

$$\chi^2 = \sum_{tracks} \mathbf{r}(a,\tau)^T V^{-1} \mathbf{r}(a,\tau)$$

where V is the covariance matrix of the hits. This minimization will be performed with respect to the alignment constant, enforcing the condition $\frac{d\chi^2}{da} = 0$.

With an iterative procedure the constants which minimize the χ^2 are determined. Thanks to the high number of tracks and the continuous stream of particles, the inner detector can be aligned every 20 minutes, calculating δa, allowing for a stability of the tracking performance even in presence of rapid detector distortions.

Weak Modes analysis: sagitta distortion

A more subtle class of detector distortions are those which leave the global χ^2 almost unchanged, since they preserve the helical shape of the track. They are called weak modes [53] and lead to large biases on the particle momentum and on the track impact parameters. For that reason, the ID is constantly monitored and detailed time-consuming analyses are performed to identify these peculiar misalignments. I contributed to develop the methods used to monitor and constrain them, namely:

- using standard candle in particle physics, such as Z, J/Ψ and K_s resonances to measure if their mass/p_T is reconstructed correctly by the tracker;
- use external detector constrains as E/p, where E is provided by the calorimeter and p by the tracker;
- using cosmic rays trajectories with no magnetic field surrounding the detector, in order to have straight tracks in the ID.

The results that come from these analyses are converted into new alignment constants, which are used to update the prompt alignment and reprocess data to correct for the misalignment. One of the most dangerous weak mode affecting the p_T measurement is the curl. It is caused by a movement of the inner detector modules, that can be modeled as a curling of the layers with magnitude increasing with the radius of the detector. This phenomenon produces a $\Delta\phi$ for the ID modules and therefore a bias in the measurement of the track sagitta in an anti-symmetric way depending on the charge of the particle, fig. 2.13.
It then directly affects the calculation of the transverse momentum, depending on the charge of the particle. Hence, the reconstructed momenta will become:

$$p_{reco} = p/(1 + qp_T\delta_{sagitta})$$

where $\delta_{sagitta}$ quantifies the magnitude of the bias and q is the charge of the particle.
Since this bias propagates on the reconstructed masses, one of the most efficient ways to monitor and detect it is the use of the mass distribution of the $Z \to \mu\mu$ decay. Markers depending on the deviation from the MC mass and from the MC width of the Z as a function of the (η, ϕ) coordinate have been constructed in the following way [54]:

$$\frac{M^Z_{reco} - M^Z_{MC}}{M^Z_{MC}} \quad , \quad \frac{\sigma^Z_{reco} - \sigma^Z_{MC}}{\sigma^Z_{MC}}$$

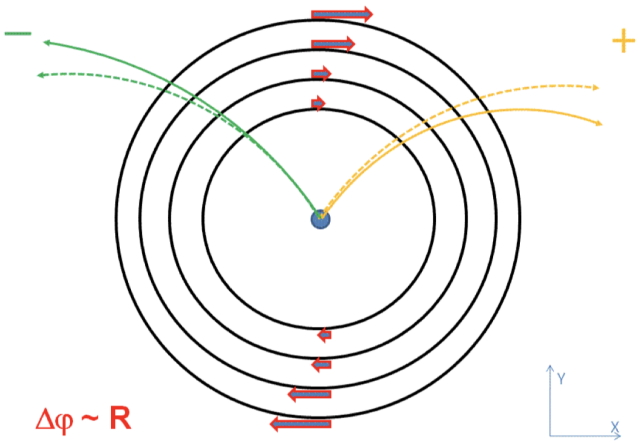


Figure 2.13: Sketch of curl distortion of the ID. The sagitta of the particle bent in the positive direction will be decreased (same direction of the curl), while the sagitta of the particle bent in the negative direction will be increased (opposite direction of the curl).

The behaviour of these markers is shown in fig 2.14 using 2016 data. From these plots the misaligned regions are highlighted by values of the mean and of the width of the mass distribution that deviate from the MC values. Since the curl is anti-symmetric distortion with respect to the charge of the particle, the p_T of positive and negative muons is biased in the opposite way. The mean of the mass distribution is therefore shifted from its nominal value and, for the same reason, the width of the mass distribution increases when using the leading p_T, where the two muon populations (positive and negative) are mixed.

After having identified the biased region it can be corrected using the following expression:

$$\frac{m^2_{d\mu\mu} - m^2_{0\mu\mu}}{m^2_{d\mu\mu}} = (p^+_{Td}\delta^+ - p^-_{Td}\delta^-)$$

where m_d is the distorted mass, m_0 is the nominal mass, $p^{\pm}_{Td}$ is the distorted momentum for the positive or negative muon and $\delta^{\pm}$ is the corresponding sagitta distortion. Such relation can be inverted and used to compute δ_s iteratively for each (η, ϕ) bin. In fig. 2.15 an input $\delta_{sagitta}$ distortion map is introduced at reconstruction level to bias a sample of simulated $Z \to \mu\mu$. The convergence efficiency and the goodness of the bias correction are improved by using as nominal mass the MC mass, which depends on the (η,ϕ) coordinate, instead of a flat one coming from Particle Data Group, since all the detector effects induced by the geometry are taken into account.

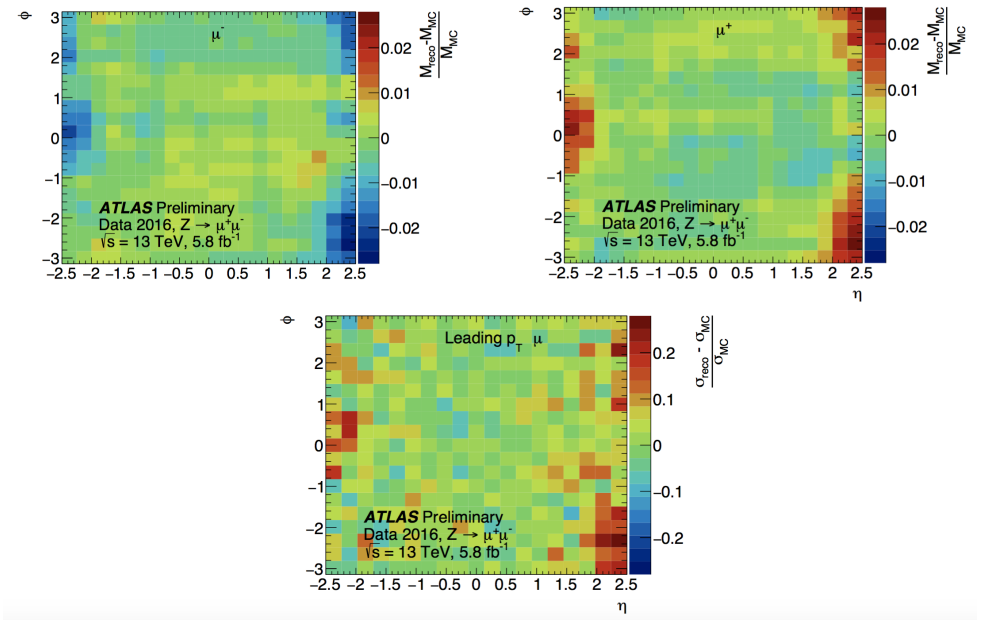


Figure 2.14: In the plots reconstructed quantities are calculated using 2016 data not reprocessed with the optimal alignment and the MC ones are calculated from simulated events with perfectly aligned detector geometry. The markers are plotted as a function of the η and ϕ of cells which is hit by the positive, negative or pT leading muons.

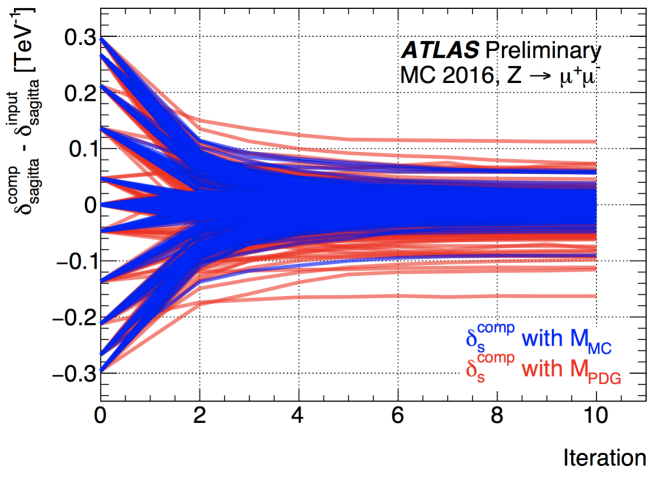


Figure 2.15: The convergence of $\delta^{comp}_{sagitta}$ to $\delta^{input}_{sagitta}$, as a function of the iteration is shown. The red lines correspond to the case in which $\delta^{comp}_{sagitta}$ is computed using the PDG mass of the Z in each (η, ϕ) bin. The blue lines correspond to the case in which $\delta^{comp}_{sagitta}$ is calculated using MC mass as reference. Each line refers to a single (η, ϕ) bin.

2.2.5 Calorimeter

There are two calorimeters in the ATLAS (fig.2.16): the Electromagnetic Calorimeter (EMCAL) and the Hadronic Calorimeter (HADCAL) [50]. Both of them are sampling calorimeters. The first one measures the energy of particles such as e, e^+, γ. It makes use of Liquid Argon technology, with lead as absorbing material. The second one detects hadrons, jets and taus energy deposits. It uses scintillating tiles as active material and steel as absorber.
The calorimeters form two concentric cylindrical layers covering a region between $-4.9 < \eta < 4.9$. They are designed to have a good containment of the particle showers, in order to keep all the information of the interaction without losing energy resolution. Furthermore, in order to focus the detector capabilities where they are more needed, they have different granularity depending on the η region as shown in tab. 2.1.

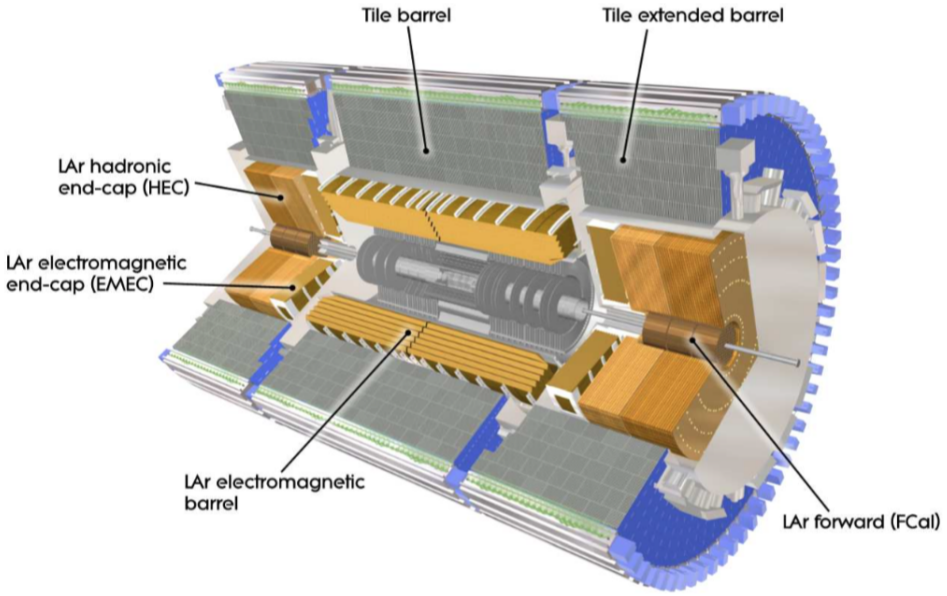


Figure 2.16: Schematic view of the electromagnetic and hadronic ATLAS calorimeters surrounding the inner detector [50].

The Electromagnetic Calorimeter

The electromagnetic calorimeter is a LAr-lead calorimeter with fine granularity and it completely surrounds the inner detector. When a particle interact with the calorimeter material, it produces a shower which develop in the lead, ionizes

Calorimeter	η	$\Delta\eta$ x $\Delta\phi$ granularity
Electromagnetic Barrel	$\lvert\eta\rvert < 1.475$	0.3 x 0.1
Electromagnetic End-Cap	$1.375 < \lvert\eta\rvert < 3.2$	0.1 x 0.1
Hadronic Barrel	$\eta\rvert < 1.0$	0.1 x 0.1
Hadronic End-Cap	$1.5 < \lvert\eta\rvert < 2.5$	0.1 x 0.1
	$2.5 < \lvert\eta\rvert < 3.2$	0.2 x 0.2
Forward Calorimeter	$3.1 < \lvert\eta\rvert < 4.9$	0.2 x 0.2

Table 2.1: The granularity in the different regions of the ATLAS calorimeters [50].

the LAr between two lead absorbers and the signal is collected on the kapton electrodes. This EMCAL has an inner structure: the Electromagnetic Barrel and the Electromagnetic End-Cap Calorimeter.

The barrel is composed of 16 identical modules, each of them formed by three layers and a pre-sampler, see fig. 2.17. These layers divided are the front layer, the middle layer and the back layer. The granularity decreases from the front layer to the back layer. The Electromagnetic End-Cap Calorimeter is composed of two disks arranged in eight modules instead. Unlike the barrel, the end-caps absorbers thickness increases with η coordinate, and is symmetrical in ϕ symmetry. The resolution of the EM calorimeter of the ATLAS detector is designed to be [55]

$$\frac{\sigma(E)}{E} = \frac{10\%}{\sqrt{E}} \oplus 0.7\%$$

where the first term stands for the sampling term and the second one describes the detector imperfections.

The Hadronic Calorimeter

The hadronic calorimeter surrounds the EMCAL and it covers the same angular region. It is divided in three main blocks: the Tile Calorimeter, the End-Cap Calorimeter and the Forward Calorimeter. The first one is divided in 64 modules (fig. 2.18), each of them with three layers. The active part are the scintillating tiles, 3 mm thick with a trapezoidal shape, placed in a steel container perpendicularly to the beam axis. When a particle passes through the tiles a scintillation light is produced. Then, the wavelength shifting fibers shift to the blue light and a PMT system collects it.

The End-Cap Calorimeters are located in $1.4 < |\eta| < 3.2$. They are a system of four disks, two for each end-cap. They are all composed of 32 modules with 2 layers and use of liquid Argon as active material, placed between two copper plates.

The LAr Forward Calorimeter (FCal) works at high $|\eta|$. Each of the end-caps is divided in three modules (FCAL1,FCAL2,FCAL3): FCAL1 is made of copper,

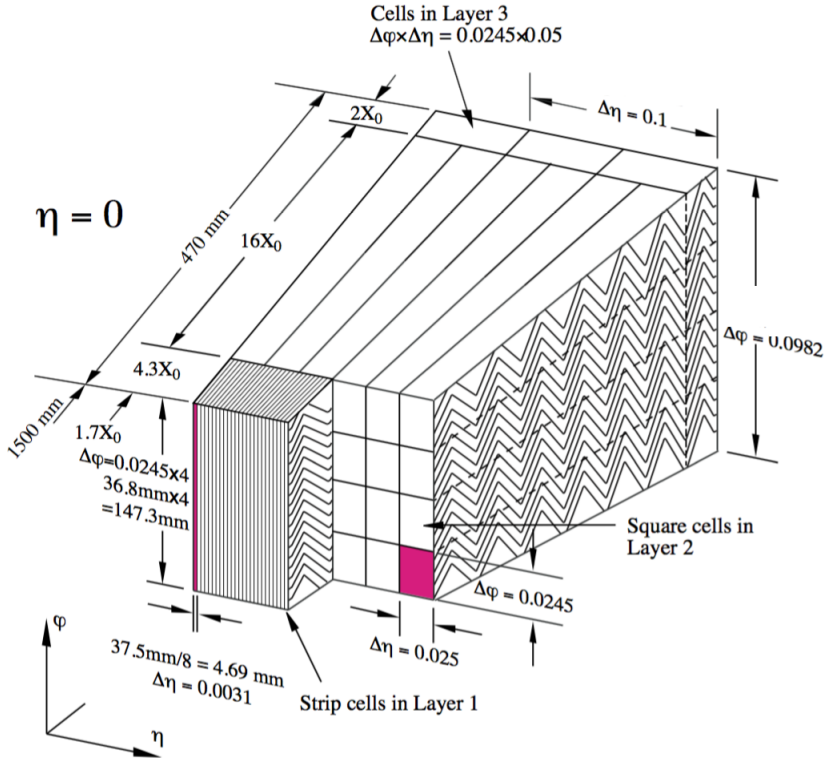


Figure 2.17: Picture of the modules which composes the barrel of the electromagnetic calorimeter [50].

FCAL2 and FCAL3 are made of tungsten. The first layer is for electromagnetic measurements, the other two are placed at the very high η in order to provide shower containment.
The energy resolution estimated for the hadronic calorimeter is [55]

$$\frac{\sigma(E)}{E} = \frac{50\%}{\sqrt{E}} \oplus 3\% \quad , \quad |\eta| < 3$$

$$\frac{\sigma(E)}{E} = \frac{100\%}{\sqrt{E}} \oplus 10\% \quad , \quad 3 < |\eta| < 5 \ .$$

2.2.6 Muon Spectrometer

The Muon Spectrometer is the system that ATLAS uses in order to detect muons. It exploits the magnetic field generated by the magnets and measures the curvature of the muon trajectories, which is inversely proportional to the transverse momentum. The whole structure is symmetrical in η and so we will describe just the positive η regions.
The system makes use of two different kind of detectors: the precision-tracking chambers and the trigger chambers (see fig. 2.19).

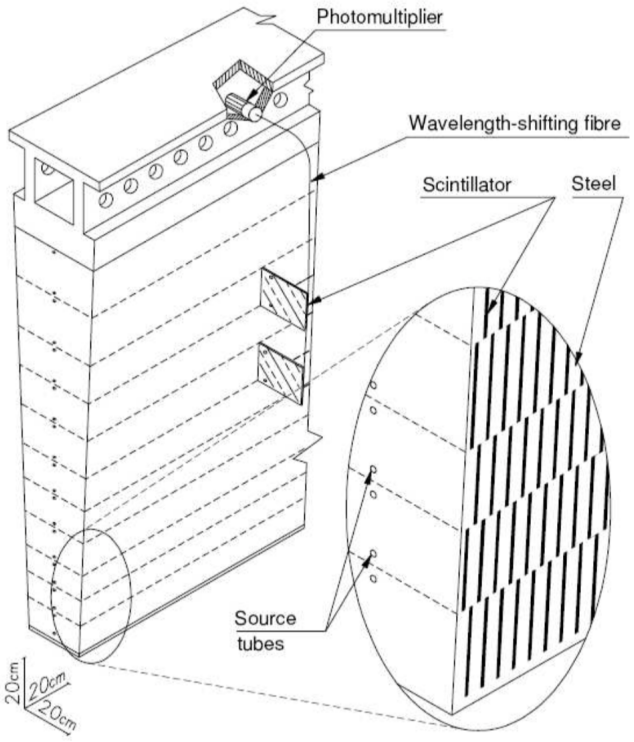


Figure 2.18: Picture of a Tile module [50].

The precision-tracking chambers are the Monitored Drift Tubes (MDTs) and Cathode Strips Chambers (CSCs). They are used to track muons with high precision. The MDTs are placed at the border of the η range. They are arranged in three barrels parallel to the beam axis and four wheels composed by eight layers of drift tubes. The dimension of the tubes is 30 mm in diameter. When a muon passes through the MDT, ionization charges are produced and collected on the electrodes, achieving the optimal spacial resolution.
The CSCs are used at $2 < \eta < 2.7$. They are multi-wire proportional chambers, with higher granularity than MDTs. They are used also because they collect charges very fast.

The trigger chambers are the Resistive Plate Chambers (RPCs) and the Thin Gap Chambers (TGCs). They are employed to trigger on muons. The former are used at low $\eta < 1$, while the latter are more distant from the beam axis, placed at $1 < \eta < 2.4$. The RPCs are gas detectors: a gas is located in a narrow gap between two plastic plates. When a particle passes through it, it triggers a localized avalanche. The charges are rapidly collected by the metal strips of the plates. The optimal configuration for the RPC is to place them in series in order to have more trajectory points and extrapolate a track.

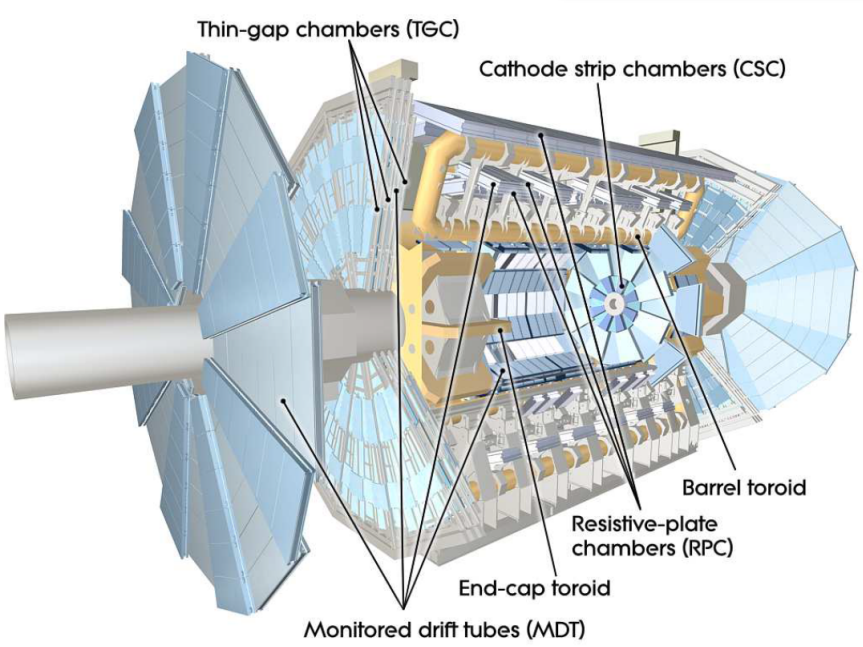


Figure 2.19: Picture of the ATLAS muon spectrometer [50].

The TGCs are multi-wire proportional chambers as the CSCs. They have a modified geometry that increases the time resolution. Their tasks are to contribute to the triggering on the detected particle and to estimate the azimuthal angle [50].

2.3 The Trigger System

The trigger system has two distinct levels: Level-1 (L1) and the High Level Trigger (HLT) (see fig. 2.20). Each trigger level refines the decision made by the previous one by applying stricter selections. The first level uses a reduced amount of the information coming from the detectors (Calorimeters and Muon System) and reduces the initial rate (40 MHz) to 100 kHz. After the L1 Trigger decision, the High Level Trigger adds more information and performs a stronger selection reducing the rate to 1 kHz. At the end of the trigger and storing procedure, each event has a size of approximately 1.3 Mbyte.
The Level-1 trigger is used to look for events with high transverse energy carried by charged particles, jets or photon. Its decision is based on the reduced information coming from the hadronic or electromagnetic calorimeters and from the muon spectrometer. The Level-1 Muon and Level-1 Calorimeter Trigger (L1Calo) information is processed by the central trigger processor. Furthermore, in each event, the L1 trigger defines one or more Regions-of-Interest, i.e. the η and ϕ coordinates where the trigger saw interesting physics. The High Level Trigger uses the full calorimeter granularity information. It

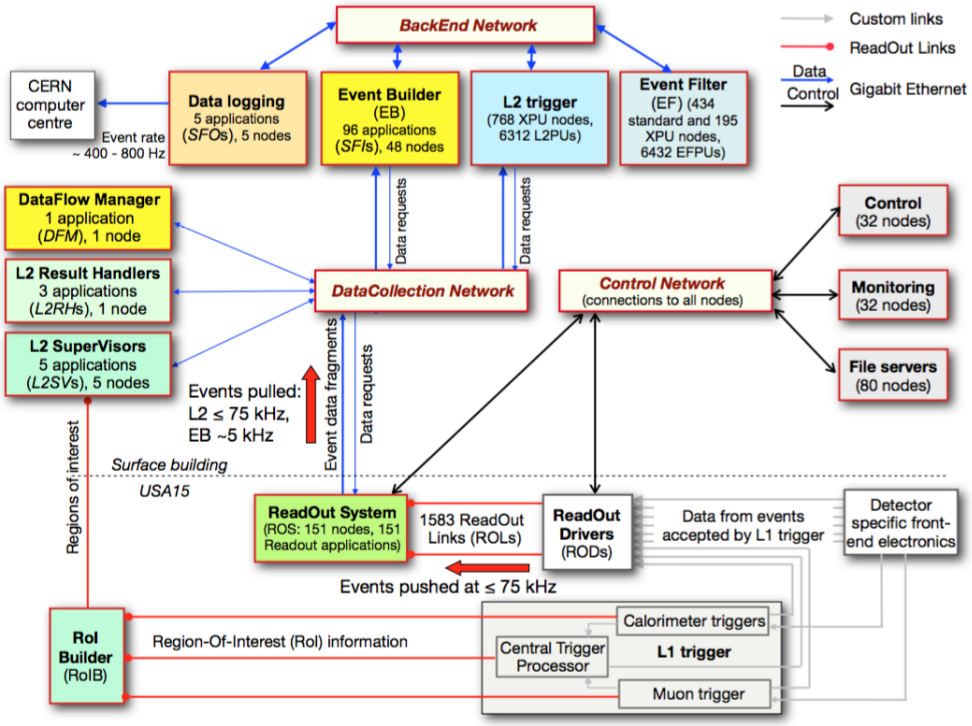


Figure 2.20: Block scheme of the Trigger and DAQ system [56].

unites the previous L2 and the Event Filter systems and reduces the Level-1 trigger rate.
After the decision of HLT the event is discarded or built. Then, offline algorithms perform finer selections when the event data is passed to the Event Filter of the HLT. This trigger level can access further information from the calorimeters and spectrometers and perform a combination of them with the data from the ID to further refine the selection. For instance, it performs the reconstruction of decays as is required for detecting b decays. The Event Filter is formed of approximately 3200 quad-core CPUs. If the Event Filter accepts the event, it is passed to the farm and then stored on disk [56].

2.4 Access to new precision and new physics

The LHC facility gained a great success allowing the discovery of the Higgs boson and the confirmation of the extraordinary predictive power of the Standard Model. Despite that, there are still a few experimental facts that requires explanations or the extension of the current paradigm, such as:

- the Hierarchy Problem, which can be explored by a better understanding of SUSY models and EFTs.

- the reason for the big mass differences between fermions with different flavor;
- the abundance of matter with respect to antimatter;
- the indirect evidence of Dark Matter as an important constituent of the universe mass, not predicted by the SM;
- the evidence of a non-zero neutrino mass, which contradict the SM predictions;

Further investigations have been carried out towards super symmetric particles, which are beyond the reach of LHC (exclusion limit up to 1 TeV).
Most of these questions can be addressed by a next generation of experiments and therefore new facilities which overcomes the current limitation in center-of-mass energy and intensity have to be designed. The attempts are two:

- increase precision of measurements related to the Higgs boson;
- increase center of mass energy to access more massive first states;

2.4.1 High Lumi LHC

The more short term project is a major upgrade of the LHC which will take place in the 2020s. It will increase the luminosity reached by the machine by a factor 5 and to extend its operations for another decade. The general thought is that they will allow to extend the investigation of the present LHC range for mass and coupling by 20-50%. They will also contribute in constraining, and potentially discovering, new physics that is presently unexplored.

The new average instantaneous luminosity will be around $5 \times 10^{34} cm^{-2} s^{-1}$ (peak luminosity at $7.5 \times 10^{34} cm^{-2} s^{-1}$). This will bring to an integrated luminosity per year of 250 fb^{-1}, as shown in fig. 2.21), which is almost twice the luminosity collected from 2011 [57]. The High Luminosity LHC (HL-LHC), will rely on a number of key innovations, representing exceptional technological challenges:

- **Cutting-edge 11–12 T superconducting magnets**. The change of the magnets in the high luminosity insertions is the cornerstone of the LHC upgrade. New Nb_3Sn quadruples with a 12 T peak field on the coils will be installed, as well as new Nb_3Sn dipoles with 11 T peak field and a reduced magnetic length of 11.2m.
- **Very compact superconducting cavities** for beams with ultra-precise phase control. The new superconducting cavities designed for the HL-LHC upgrade are called Crab Cavities (CC). They allow very small values of β^*, but requires a robust control of the phase and of the voltage. The cavities per beam will be installed in IP1-IP5, where they will provide a transverse kick voltage of 3.4 MV.

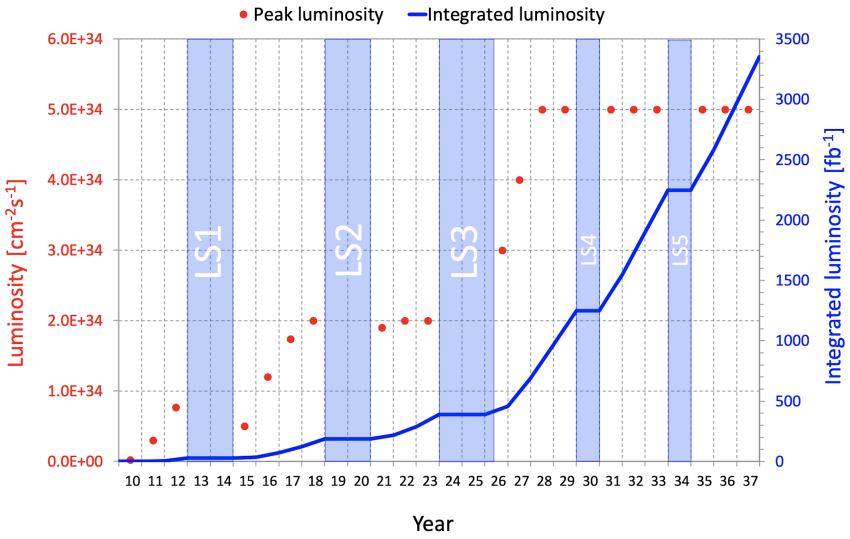


Figure 2.21: Plans for the delivered instantaneous and integrated luminosity with HL-LHC [57].

- **new technology for beam collimation**. Low impedance collimators equipped with molybdenum-coated jaws made out of Mo-Graphite composite will replace old ones. Moreover, to reduce the dispersion of off-momentum particle, a new possibility will come by the use of the new dipoles. The space gained by the reduced length of the magnets will allow the installation of an additional collimator;
- **high-power superconducting links** made out of high-temperature superconductors (YBCO or Bi-2223) and mainly of MgB2 superconductors, which shows almost zero energy dissipation, will be added.

With this machine new scenarios and the next era of precision will begin. Both SM and BSM measurements will be performed to shade light on the main open points of modern physics.
Several of these measurements performed at LHC are limited by the uncertainties on the knowledge of the partonic inner structure of the proton, i.e. Parton Density Functions (PDFs). This limitation will be reduced by the new precision measurement of global PDF and of Standard Model parameters, such as the weak mixing angle and the W mass. The other frontier, for which this upgraded have been demanded, is the possibility to study BSM physics. SUSY or EFT with extra dimensions share the prediction of new degrees of freedom close to the EW scale (1-1000 GeV) and therefore this new machine will be an occasion to test these theories still unconstrained.

For ATLAS and CMS the triggering system together with the inner detectors will be replaced and the triggering and tracking abilities will be assessed. Also LHCb will be upgraded to run at a luminosity of 2 x $10^{-34} cm^{-2} s^{-1}$.

2.4.2 ATLAS upgrade

The R&D and the upgrades of the detectors and of the new HL-LHC ones proceed at the same time. For ATLAS, the trigger and the tracking systems will be upgraded to face the new machine performances:

- center-of-mass energy at $\sqrt{s} = 14$ TeV;
- instantaneous luminosity of $7.5 \times 10^{34} cm^{-1} s^{-1}$ and integrated luminosity of $4000 fb^{-1}$;
- average number of interactions per bunch crossing $\mu = 200$.

From a technical point of view these items represent challenges for many detector subsystems:

- for the trigger system: higher particle fluxes, larger event sizes and higher trigger rate;
- for the tracking system: higher occupancy and larger radiation damages for the modules (especially the closer ones to the beam pipe);

The major upgrades planned by the collaboration involve the muon spectrometer, the inner tracking detector and the trigger and DAQ systems [58].
For the inner tracking detector one of the most important changes studied for this upgrade concerns the layout: it was rearranged from four pixel layers and five strip layers (4P 5S) to a layout with five pixel layers and four strip layers (5P 4S), to achieve better resolution on the close-by tracks. Furthermore, a η-coverage up to 4 (previously 2.5) will be provided by a forward pixel system (Pixel End-cap Ring System) comprising 12 disks. Two different technologies have been studied for the pixel layers, tailored on the radiation environment: 3D pixel modules will cover the two innermost layers and the traditional planar pixel sensors will cover the other ones.
For the muon spectrometer most of the detector readout and trigger electronics for the RPCs, the TGCs, and the MDTs chambers will be replaced to withstand the higher trigger rates and longer latencies necessary for the new level-0 trigger. Additionally, the new RPCs will be added to increase the acceptance and robustness of the detector. A consistent fraction of the MDT chambers will be substituted with the new small-diameter MDTs. The doublets TGCs will be replaced with the triplet chambers, to suppress random coincidences which increase the trigger rate in that region. Finally, the CSCs and the MDTs chambers of the innermost end-cap wheels will be replaced by micromegas and small-strip TGCs.
The enhancement of the ATLAS Liquid Argon (LAr) calorimeter with electronics and optical-fiber cabling is also foreseen. This will greatly improve the resolution of the detector at trigger level, providing four-times higher granularity to allow particles to be better identified and events to be better filtered. The on- and off-detector Tile calorimeter will be replaced as well during the

shutdown of 2024-2025. That will contribute to the reliability of the system. For the TDAQ the approved upgrades comprises: a hardware-based low-latency real-time trigger operating at 40 MHz based on FPGA, data acquisition which combines custom readout with commodity hardware and networking to deal with 5.2 TB/s input, and an event filter running at 1 MHz.

2.4.3 Future colliders

There is an extensive discussion about which is the most useful new collider to develop and therefore about which technology we have to invest on now. The discussion resulted in several proposal supported by different countries and different collaborations, some of them are more straightforward than others [59]. There are projects to build new hadron colliders, exploiting the LHC, as well as ideas of electron-positron collider, electron-hadron colliders (fig. 2.22) or muon colliders. The main proposals are:

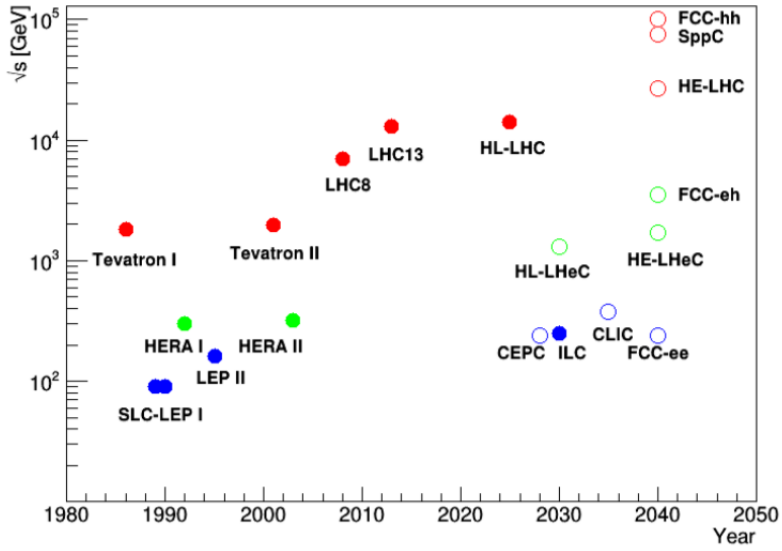


Figure 2.22: Center of mass energy reached by the facilities built and foreseen in the next years [59].

- **The International Linear Collider (ILC)** . The linear colliders have the main advantage to not require dipole magnets, but just focusing magnets. Furthermore, since the trajectories are not bent in a magnetic field, there is not energy loss due to synchrotron radiation. In particular the ILC is based on a technology using superconducting accelerating cavities, which turns out to be the best suited to meet the goals of a machine that operates between the Z boson mass and 1 TeV. In the technical design report presented in 2013 the ILC collaboration assumes 500 GeV center-of-mass energy as the baseline. A similar project is the Compact Linear Collider (CLIC).

- **Future Circular Collider for hadrons (FCC-hh)**. It was proposed for the first time in 2013 by the European Union in the context of the European Strategy for Particle Physics. The FCC-hh project aims to a machine with a center-of-mass energy of $\sqrt{s} \approx 100$ TeV. It would reside in a circular tunnel of about 100 km in circumference, using the LHC as an injector. The biggest challenge for the construction of this machine is the design and the realization of dipole magnets which provides 16 T of magnetic field over a length of 13,5 m. These developments are part of a worldwide program for innovative magnets technologies, which considers the application of Nb_3Sn as conductor material.

- **Future Circular Collider for electrons (FCC-ee)**. The proposal of FCC-hh is also motivated by the idea to install, prior to the hadron machines, circular e^+e^- colliders in the tunnel. The plan is to run with several center-of-mass energies to test specific processes: 91.2GeV (Z mass),160 GeV (WW-threshold), 240 GeV(ZH-threshold) and 365 GeV (slightly above the $t\bar{t}$ threshold). The main advantage with respect to the linear technology is the higher luminosity that can be achieved. The main limitation for a circular electron-positron collider are: beam to beam electric field interactions, beam-strahlung, beam instabilities in case of asymmetric charge.

- **Future Circular Collider for electron-hadron (FCC-eh)**. The electron machine in this case would be an Electron Recovery Linac, while the hadron machine will be the LHC. This will allow for measurements of relative couplings of the Higgs boson and to study single top production, as well as the probing of the structure functions.

- **Muon colliders**. This machine has great potential for high-energy physics, since it can produce lepton collisions at very high energies, without limitation from synchrotron radiation (as in the electron colliders). The main challenge for this interesting development in experimental physics is the short muon lifetime and therefore the difficulty to produce large numbers of muons in bunches with small emittance.

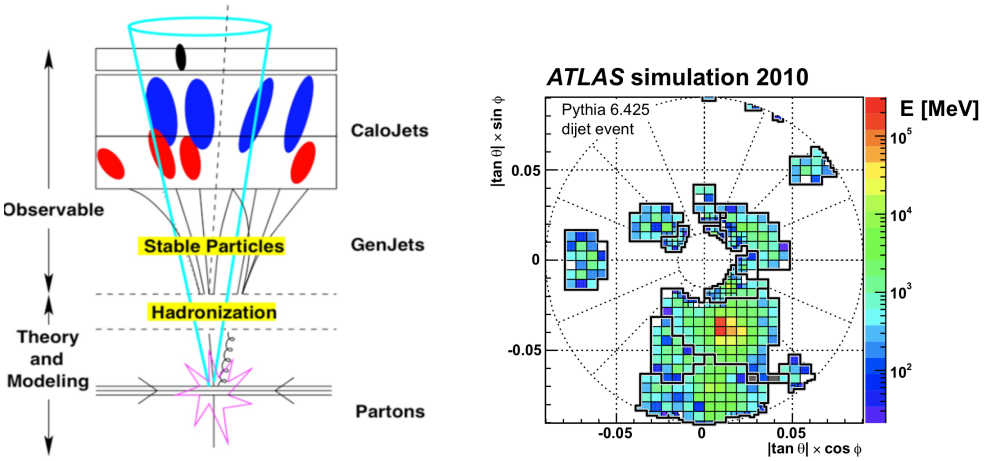
Chapter 3

Object reconstruction for the H(bb) analysis with the ATLAS detector

In this chapter we will describe the objects, tools and techniques used in a hadronic particle physics analysis at the ATLAS experiment, focusing on the $H \to b\bar{b}$ case.

3.1 Particle jets

The dominant products of the high-energy p-p collision at LHC are highly collimated sprays of energetic particles, named particle jets [60]. They arise from the fragmentation of the protons or from the decay of an unstable particle, see fig. 3.1(a). Jets provide a link between the observed colourless stable particles and the underlying physics at the partonic level. The reconstruction of these objects is performed combining the calorimetry and tracking information and several steps are required before computing the common jet variables: p_T, mass and radius. In ATLAS, the jets are reconstructed starting from energy deposits in the calorimeter cells. A particle releases its energy passing trough the detector material and its 3D energy deposit in the calorimeter is called topological cluster or topo-cluster [61]. This cluster is formed starting from a cell, called seed, and adding iteratively neighboring cells, according to the energy recorded and the overall noise level expected, see 3.1(b).
This cluster is classified in hadronic or electromagnetic, depending on its position, shape and energy. The energy is computed assuming it was produced by EM interaction and then it is calibrated to hadronic scale with a cluster reweighting. Besides clusters, also tracks can be used as jet seed, these objects are named track-jets. In this case a track is used as pointer towards energy deposit in the calorimeter, which are possibly re-clustered to form a jet [62].



(a) (b)

Figure 3.1: On the right, a sketch of the collision between two protons and the hadronization of the decaying products. On the left, the calorimeter cells clustered to form several ATLAS topo-clusters.

3.1.1 Jet reconstruction

During the operation of the modern particle physics experiment, physicist developed several jet reconstruction algorithms, with structure and performance depending on the condition and requirements [63]. The two main classes of reconstruction algorithms are:

1. **the cone algorithms**. They work under the assumption that particles in a jet are distributed within a cone. All constituents within a radius R are considered part of a jet. The jet-axis can be defined as the direction of the hardest p_T constituent. In the Iterative Cone (IC) algorithms, it is recalculated using all constituents every time a new one is added to the jet. The other option to form a jet is to consider the cones as proto-jets, then constituents are removed from the cones and the procedure is repeated until no more cones are found. Finally, all the overlapping proto-jet are merged in the final jet.

2. **the sequential algorithms**, among which the major exponents are the k_t, anti-k_t and the Cambridge/Aachen algorithms. They are preferred in particle physics and used in several forms in all the main LHC experiments. They assume that the particles belonging to the same jet have small differences in p_T and therefore particles are grouped in cluster of transverse momenta. The result is a jet distributed in a defined $\eta - \phi$ area.

The jet reconstruction algorithm used in this analysis is the Anti-K_t [64]. It has been chosen because it shows higher performance in very dense environments with the presence of soft-radiation. It is divided in several steps:

- the distance between two particles is calculated as,

$$d_{ij} = min(1/p_{Ti}^2, 1/p_{Tj}^2)\frac{\Delta R_{ij}}{R}$$

where $\Delta R_{ij} = \sqrt{(\Delta y)_{ij}^2 + (\Delta\phi)_{ij}^2}$ is the distance in the $\eta - \phi$ space and R is the radius of the jet, usually between 0.4-0.7.

- a second variable is computed as the momentum space distance between the beam axis and the considered jet constituent, defined as:

$$d_{iB} = \frac{1}{p_{Ti}^2}.$$

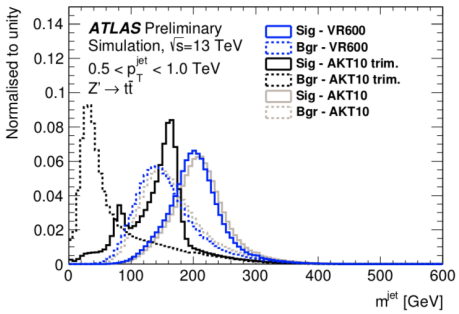
The algorithm searches in the entire particle set the minimum between d_{ij}, d_{iB}. If for two particles is found that the minimum is d_{ij}, these two particles 4-vectors are summed and removed from the list of particles. If d_{iB} is found to be the minimum of the set, the i is labeled as final jet and removed from the jet list

- when all particles are part of a jet with distance between jet axis larger than R the procedure stops. The clustering can be stopped also when the defined amount of jet is found. With this procedure we have that jet are clustered mainly around hard particles, which attracts all the soft ones within a distance R.

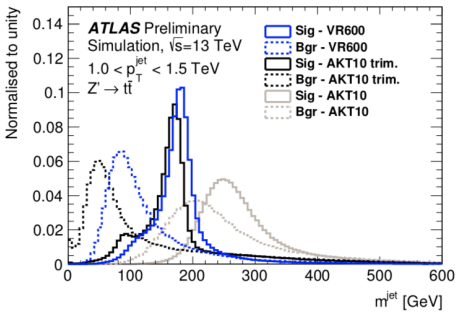
The most used jet topology in ATLAS are anti-k_t jets with R = 0.4, that are also called small jets. But when the decaying particle has a large transverse momentum its decay products will be boosted along its flight direction and therefore they will merge in a single anti-k_t jet with $R \leq 1.0$, called large-R jet. Large-R jets are the basic objects in many ATLAS boosted analysis with hadronic final states. This jet topology allows grouping in a single object all the products of the decay and to compute quite easily kinematic quantities as the jet mass and the jet p_T. In very boosted regimes, to resolve the internal structure of a large-R jet, small jets with variable radius are used, they are called variable-R jets (VR). Conversely, with the traditional small-jets or large-R jets, which have fixed radius, these objects have a p_T dependent amplitude given by:

$$R_{VR} = \frac{\rho}{p_T}$$

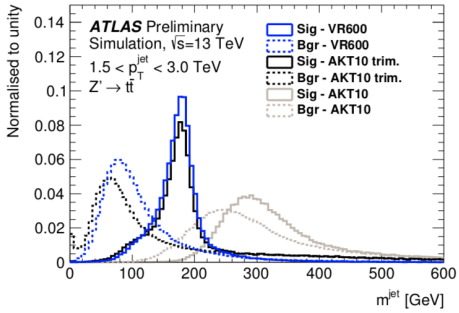
where ρ is a parameter which determines how fast the effective jet size decreases with the transverse momentum of the jet, between the fixed values R_{min} and R_{max}. Therefore, VR jets are more dynamic and suitable to reconstruct boosted objects. The performance of mass reconstruction for large-R jets and for VR jets are shown in fig. 3.2.



(a) 0.5 TeV < p_T^{jet} < 1.0 TeV



(b) 1.0 TeV < p_T^{jet} < 1.5 TeV



(c) 1.5 TeV < p_T^{jet} < 3.0 TeV

Figure 3.2: Jet mass distributions on Z' signal (solid lines) and multijet background (dotted lines) events for ungroomed VR jets (blue), trimmed anti-kt R = 1.0 jets (black) and ungroomed anti-kt R = 1.0 jets (grey) in three different p_T regimes.

3.1.2 Jets substructure

A key role to understand the properties of the particles which originate the jet, is to study the inner structure, called jet substructure [65]. There are several observables that characterize jet substructure and distinguish between boosted hard scattering objects and soft ones:

- **Jet Mass:** it is defined as the module of the four-momentum sum among all the jet constituents;
- k_t **splitting scales:** they are calculated during the re-clustering the constituent of a jet with the k_t algorithm. The two harder constituents of the jets, called at this stage proto-jets, are used to compute the distance d_{ij}, called splitting scale variable:

  $$\sqrt{d_{ij}} = \min(p_{Ti}, p_{Tj}) \times \Delta R_{ij}$$

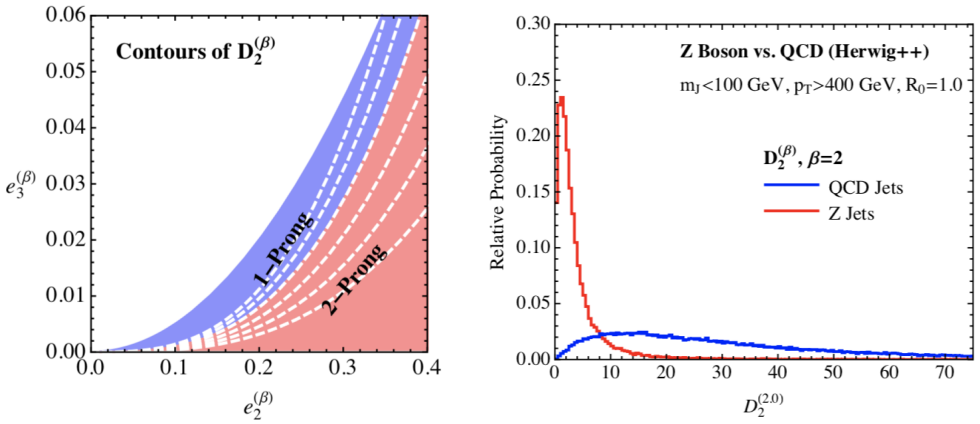
  For a two-body heavy-particle decay the $\sqrt{d_{12}} \approx \frac{m_{particle}}{2}$, while for light quarks and gluons the values are lower;
- N**-subjettiness:** this variable, τ_N, rely on the sub-jets multiplicity. It is calculated during the re-clustering k_t algorithm requiring that exactly N jets are found in the event. These N jets define an axis within the jet. It is calculated with the following formula,

  $$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \times \min(\delta R_{1k}, \delta R_{2k}, ..., \delta R_{Nk}) \ , \ d_0 = \sum_k p_{T,k} \times R$$

  where R is the jet radius parameter and δR_{ik} is the distance between the sub-jet constituents i and k. It basically tells how well the jet is described by N or less sub-jets. If $\tau_{21} = \frac{\tau_2}{\tau_1} \approx 1$ it means that the jet is well described by one constituent, instead if the value is smaller the constituents are likely more than one.
- **D2**. It is one of the most powerful discriminator for one- or two- pronged structures, as in fig. 3.3(a). It is defined as:

  $$D2^\beta = \frac{e_3^\beta}{(e_2^\beta)^3}$$

  where e_i is the i-points energy correlation function and β is the angular exponent. It can be applied to distinguish signal decaying in jet pairs from multi-jet background, as in fig. 3.3.



(a) (b)

Figure 3.3: On the left the correlation between e_2 and e_3 for two- and three-pronged jets, for different β values is shown. The white lines correspond to different β values [65]. The plot on the right shows the difference in the D2 values for a sample of Z boson decaying to two b-pairs and for a sample of QCD multi-jet background [65].

3.1.3 Jet grooming

An important problem to face in jet reconstruction is the presence of pile-up and of objects not associated to the primary vertex. In order to deal with it, a wide variety of techniques and algorithms have been developed to improve the correct identification and the resolution on the jet observables. These algorithms are designed to clean-up jets by subtracting the contributions of unassociated radiation. This process is called jet grooming and the main algorithms used are: mass-drop filtering, trimming and pruning [60].

Mass-drop filtering: this technique is optimized for C/A jets. It aims to find symmetric and well separated sub-jets with mass significantly smaller than the original jet. It operates in two steps:

1. Mass-drop and symmetry: the jet is split in two, ordered in mass and then the following relation are checked:

$$\frac{m^{j1}}{m^{jet}} > \mu_{frac} \ , \ \frac{\min[(p_T^{j_1})^2, (p_T^{j_2})^2]}{(m^{jet})^2} \times \Delta R^2_{j_1,j_2} > y_{cut}$$

   if these relations are not satisfied, the jet is discarded. Where μ_{frac} and y_{cut} values are optimized for each analysis.

2. *Filtering*: j_1 and j_2 are re-clustered using C/A with $R_{filt} = \min[0.3, \Delta R_{j_1,j_2}/2]$, where $R_{filt} < \Delta R_{j_1,j_2}$. All the constituents outside the three hardest sub-jet are discarded, to exclude additional radiation.

Pruning: The procedure is based on three steps:

1. C/A or k_t reconstruction algorithm is run;

2. at each recombination step of j_1 and j_2 the following relation must be satisfied

$$p_T^{j_2}/p_T^{j_1+j_2} > z_{cut} \ , \Delta R_{j_1,j_2} < R_{cut} \times (2m^{jet}/p_T^{jet})$$

   if not j_2 is discarded and the algorithm continues;

3. if previous condition is satisfied j_1 and j_2 are merged.

Trimming: this technique uses the fact that the p_T of the softer components from pile-up, radiation and multiple parton interaction, is much smaller than the partons associated with harder interactions. The other advantage is that with trimming there are more chances to retain the radiation from parton emission. This tends to be distributed in clusters, while unassociated radiation is more uniformly distributed. The trimming procedure has three steps:

1. The k_t-algorithm is used to create sub-jet with R_{sub}.

2. Any constituent with $\frac{p_T^i}{p_T^{jet}} < f_{cut}$ (a few percent, i.e 0.03) is removed, where $< f_{cut}$ is the cut-off parameter. With this requirement the low-mass jets from soft products loose 30-50% of their energy and instead the jets containing boosted objects lose less.

3. the final jet is the sum of the remaining constituents.

3.2 Muons

In ATLAS, muons are reconstructed in several ways and different reconstruction qualities are assigned to them [66]. Muon tracks are reconstructed independently in ID and in MS and then the information is combined. The muon spectrometer builds tracks starting from hit patterns inside the muon chambers. In MDT hits are grouped in segments by fitting a straight line to the hits laying on each layer, while in the CSC detectors they are reconstructed using a combinatorial search in the $\eta - \phi$ space. After the segment finding, the tracks are formed from segments by fitting together the hits from different layers. A pattern recognition algorithm uses as seed the middle layer segments. The minimum number of segment that have to be connected to build a track is set to three, except for the region where there is the transition from barrel to end-cap, in which is required just one segment. A fit is performed to the matched segments. If the fit χ^2 of this fit satisfies the selection criteria, the track is accepted and a procedure of hit recovering is performed to add more hit close to the selected track. After that, the combined reconstruction is performed including the information coming from the inner detector and the calorimeters. The reconstructed muons are divided in four categories:

1. **Combined (CB)**: muon tracks are reconstructed independently with MS and ID and then a global fit is performed using hits of the matched track segments;

2. **Segment-tagged (ST)**: muon tracks are reconstructed in the ID and then matched with segments of the MDT or CSC chambers.

3. **Calorimeter-tagged (CT)**: muon tracks are reconstructed in ID and then they are matched with a energy deposit compatible with a minimum ionizing particle.

4. **Extrapolated Muons (EM)**: muon tracks are reconstructed using the MS segment and a loose compatibility with the interaction point from which the particle is required.

There are several reconstruction procedures depending on the different muon identification qualities. The lowest quality category are the loose muons, which are muons of all type. Then there are the medium muons which are required either to be CB with at least three hits in two MDT layers (except for η <0.1) or EM muons with at least three hits MDT/CSC layers (applied only for $2.5 < \eta < 2.7$). The best quality objects are the tight muons, which are required to fulfill the medium criteria and to be CB with hits in at least two stations of the MS. There is another category, based on p_T, the high-p_T muons, which are required to be CB muons with p_T above 100 GeV, passing the medium selection and having at least three hits in three MS stations.
Furthermore, muons may be required to be isolated. This isolation helps to reduce the mis-identification rates. The isolation is in first approximation a cut on the activity surrounding the trajectory of a particle in the detectors. There are two types of isolation:

1. **Track isolation**: activity is measured by looking at the sum of the transverse momentum of the tracks surrounding the direction of the particle in the ID. The track-based isolation variable is the $p_T^{varcone30}$ defined as the sum of p_T of the tracks surrounding the muon in a $\Delta R = \min(10\,GeV/p_T^{\mu}, 0.3)$.

2. **Calorimetric isolation**: activity is measured summing up the energy deposits surrounding the trajectory of the particle in the calorimeters, removing the energy deposited by the particle itself. The calo-based isolation variable is the $E_T^{toppcone20}$ defined as the sum of E_T of the particles surrounding the muon in a $\Delta R = 0.2$.

Therefore, isolation working points can be defined. The most common is the loose isolation working point, defined cutting on $p_T^{varcone30}/p_T^{\mu}$ and $E_T^{topocone20}/E_T^{\mu}$ in a way that as the identification efficiency is 99% across the $\eta - p_T$ space.

3.3 b-tagging algorithms

The correct identification of b-hadrons is a crucial requirement for the ATLAS physics analyses, especially for those related to the Higgs and to the top quarks. b-quarks are the lightest third generation quarks and therefore they are allowed to decay just in first and second generation quarks. However, this decay is strongly suppressed since its probability depends on a CKM off-diagonal coefficient. This fact provides bottom-particles with a relatively long lifetime $\tau_b \approx 1.5$ ps, which for a b produced at 50 GeV corresponds to a traveled distance in the transverse plane of approximately 3 mm (in the lab reference frame). This long path allows these particles to be directly observed in the ATLAS inner detector. Therefore, all b-tagging algorithms exploits this key feature by looking at the displacement between primary and secondary vertex, as well as the high mass and decay multiplicity of b hadrons, the presence of a muon coming from a semileptonic decay and the hard b-quark fragmentation function [67].
Multivariate techniques are widely used in this field and they provide strong discriminating observables. They are called b-tagging algorithms. Their performance is characterized with physics events which include the presence of pile-up. The most important value to measure and calibrate is the b-tagging efficiency, ϵ_b, defined as the fraction of b-quarks correctly identified by a specific algorithm. Its calibration is crucial since mismodelling of the input variables used for the identification of b-quarks would translate in a different output and therefore in a different tagging rate between data and MC. These effects are taken into account by comparing the performance obtained in the simulation with those obtained in the data, using dedicated calibration analyses. The outcome of the calibration are the so-called scale factors, defined as:

$$SF_b = \frac{\epsilon_b^{DATA}}{\epsilon_b^{MC}}$$

where ϵ_b^{DATA} is the b-efficiency on data and ϵ_b^{MC} is the one obtained on MC samples. The b-efficiency is usually measured on samples of $t\bar{t}$ events with one or two leptons in the final state, since the top quarks almost always decay b-quarks. The identification efficiency for light and charm quarks is calibrated similarly, using specific analyses for the comparison with data. The main b-tagging algorithms have calibrated working points at fixed efficiency: 60%, 70%, 77% and 85%. This variable is calibrated and for each b-tagging algorithm and WP, different scale factors are derived. The b-tagging algorithms in ATLAS exploits different strategies/variables. The most important ones are: the impact parameter-based algorithm, the secondary vertex reconstruction algorithm and a decay chain multi-vertex reconstruction algorithm.

IP2D and IP3D algorithms [68]: as mentioned above, the most distinctive characteristic of the hadrons containing b-quark is the long lifetime. Therefore, a b-hadron topology is characterized by a secondary vertex displaced from the

interaction point. Therefore, the following selection are applied on the track, to select tracks originating from b-quarks:

- track p_T is above 1 GeV;
- $|d_0| < 1$ mm and $|z_0 \sin\theta| < 1.5mm$;
- seven or more silicon hits, no more than two holes and at least one hit in the pixel detector.

Then these algorithms perform a per track decision to discriminate between b and light jets. The IP2D uses the transverse impact parameter significance, d_0/σ_{d_0}, while IP3D the longitudinal and the transverse parameter significance, $z_0 \sin\theta/\sigma_{z_0}$. Several refinements are introduced afterwords:

- number of pixel hits relaxed to one;
- tracks from conversion, Λ and K_S decays, are ignored;
- no hits in the two innermost layers.

Secondary vertex reconstruction algorithm (SV) [68]: it reconstructs a displaced secondary vertex within the jet. Therefore, all tracks associated with a jet are tested to see if they come out from a secondary vertex. Any two track vertices is rejected if it originates from a long-lived particle, a photon or material conversion. The selected tracks are required to have at least seven hits and at most one shared hit on the silicon detector. These tracks are then fitted to each available two-tracks vertex. Then the primary and secondary vertex are required to be significantly displaced. Several selections are performed on the IP significances (d_0/σ_{d_0},z_0/σ_{z_0}) to require that each track of the secondary vertex is far enough from the primary vertex. Additional criteria are applied to reduce fake rates:

- tracks hits cannot be recorded at a radius smaller than the one of the secondary vertex;
- cases with invariant mass of the two-vertices tracks above 6 GeV are removed;
- tracks from Λ and K_s are removed.

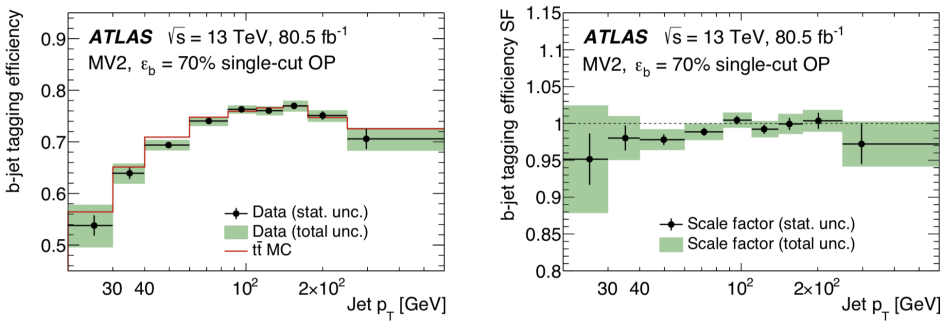
JetFitter [68]: it exploits the structure of weak b-hadron decays and tries to reconstruct the b-hadron decay chain. It can find a line on which the primary and the secondary vertices lie. Therefore, a single track is enough for the identification b-vertices.

Multivariate Algorithm: MV2

The input variables of all these low-level taggers are then combined together with other kinematic variables in a more complex algorithm, called MV2c10 [68, 69], which is based on a Boosted Decision Tree (BDT) [70]. The input variables used in the decision tree are:

- transverse and longitudinal impact parameter significance;
- properties of secondary vertex;
- flight direction of the b-hadron connecting primary and secondary vertex;
- p_T and η of the jet;

The training of the algorithm has been carried out by optimizing the BDT performance in a large hyper-parameter space, tab. 3.1. In this way we achieve the best separation power between the signal and the background. The dataset used is a subset of events from a simulated $t\bar{t}$ sample + Z', considering b-jet as signal and c-jet and light-jet as background. To enhance the background rejection, the fractions of c-jet and light-jet are optimized to be respectively 7% and 93%.



(a) (b)

Figure 3.4: The (a) b-jet tagging efficiency and (b) b-jet tagging efficiency simulation-to-data scale factors for the $\epsilon_b = 70\%$ single-cut WP of the MV2 tagger as a function of jet pT [69]. The efficiency measurement is shown together with the efficiency derived from $t\bar{t}$ simulated events passing the selection. Vertical error bars include data statistical uncertainties only while the green bands correspond to the sum in quadrature of statistical and systematic uncertainties.

In figure 3.4 the b-tagging efficiency and the scale factors as a function of the jet p_T for a 70% working point are shown. The values of the efficiency for MC and Data are compatible within the uncertainties. This reflects in the corresponding scale factors, which are always consistent with one. In this case, it can be asserted that either no significant mis-modeling is present in the MC variables used by MV2c10 or, if present, it does not play an important role.

Hyper-parameter	Values
Number of trees	1000
Depth	30
Minimum node size	0.05%
Cuts	200
Boosting type	Gradient boost
Shrinkage	0.1
Bagged sample fraction	0.5

Table 3.1: List of optimized hyper-parameters used in the MV2 tagging algorithm [69].

Deep Learning Algorithm: DL1

DL1 is an Artificial Deep Neural Network designed for flavor tagging, which combine inputs of the lower-level taggers [68, 69]. It was designed and trained in order to be as independent as possible to η and p_T. The distributions of these kinematic observables for b-jets have been reweighted to match the c-jet and l-jet ones (which is a common practice in a tagger training). The input variables belong to different categories: kinematic variables, as p_T and $|\eta|$, and input of low level taggers, as IP2D, IP3D, SV and JetFitter. Moreover, different version of DL1 have been produced using additional set of variables. The training dataset of DL1 is the same of MV2c10. The performance have been optimized tuning the algorithm parameters by scanning a large hyper-parameters space (tab. 3.2 for DL1).

Hyper-parameters	Values
Number of hidden layers	8
Number of nodes [per layer]	[78, 66, 57, 48, 36, 24, 12, 6]
Number of Maxout layers [position]	3 [1, 2, 6]
Number of parallel layers per Maxout layer	25
Number of training epochs	240
Learning rate	0.0005
Training minibatch size	500

Table 3.2: List of optimized hyper-parameters used in the DL1 tagging algorithm [69].

The outputs of the tagger are the probabilities that the considered jet is a b-jet (p_b), c-jet (p_c) and a light-jet (p_l). They are then combined in a log-likelihood ratio, which is used to discriminate between b-jets and c- or l- flavor jet. The final DL1 b-tagging discriminant is defined as:

$$DL1 = \ln\left(\frac{p_b}{f_{c-jets}p_c + (1 - f_{c-jets})p_l}\right)$$

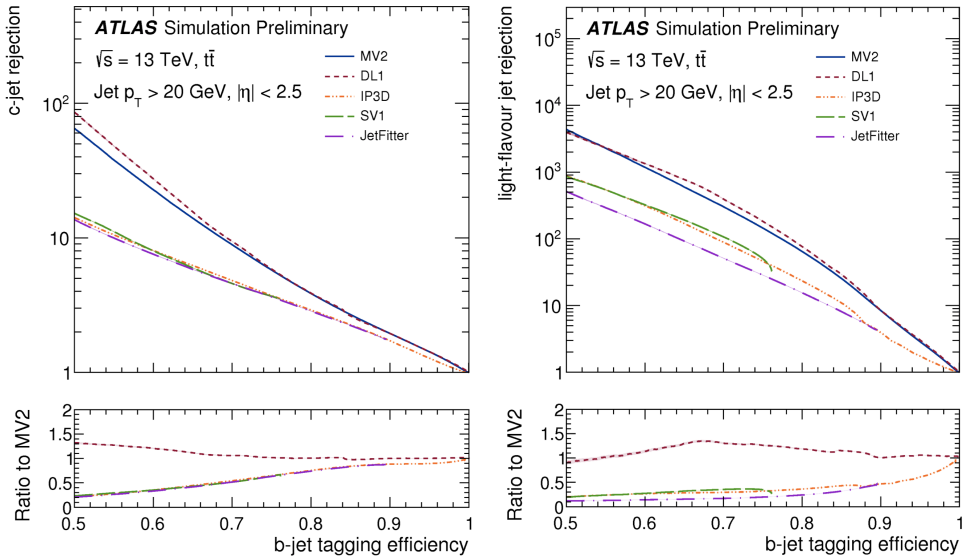


Figure 3.5: On the left the c-jet rejection as a function of the b-tagging efficiency is shown comparing all the current b-tagging algorithms. On the right the light-jet rejection as a function of the b-tagging efficiency is shown, comparing all the current b-tagging algorithms [69].

where f_{c-jet} is the effective c-jet fraction in the background training sample. In this framework, the value for the c-jet fraction can be chosen a posteriori in order to optimize the performance of the algorithm. Currently, a c-jet fraction of 8% is used. The most recent studies show that it improves discrimination of b-jets against light flavor-jets and c-jets, as in fig. 3.5(a) and fig. 3.5(b). These plots show that over all the b-tagging efficiency range the rejection power of DL1 is always equal or larger than the other algorithms, for both c-quarks and l-quarks. For that reason, analysis teams within the ATLAS collaboration are starting a migration from MV2c10 to DL1, if available and calibrated on the physics objects utilized.

Chapter 4

Search for boosted H$\rightarrow b\bar{b}$ produced in association with a jet

This chapter reports the first version of the H(bb)+jet analysis that the author contributed to, published in [71]. Although conducted on a limited amount of data, it constitutes the baseline of the new version of the analysis, still ongoing and discussed in chapter 5.

4.1 Recent observation of $H \rightarrow b\bar{b}$

In 2018 the ATLAS collaboration announced the observation of the $H \rightarrow b\bar{b}$ decay [72]. Since the b-quark is the heaviest fermion the Higgs can decay to, this measurement was very important to further constrain the Higgs decaying width and its coupling to massive particles. In that analysis, most sensitivity comes from the VH(bb) channel, which has the best signature. Events are required to have a couple of b-jets and isolated leptons, E_T^{miss}, or both of them. Therefore, they were classified in three categories, depending on the number of leptons produced by the V decay: 0-lepton ($VH \rightarrow \nu\nu bb$), 1-lepton ($VH \rightarrow l\nu bb$), 2-lepton ($VH \rightarrow \nu\nu bb$). Once all lepton channels are combined, the probability p_0 to have those results from a background only hypothesis is 5.5×10^{-7}, corresponding to an excess of 4.9 σ; the measured signal strength is:

$$\mu_{VH}^{bb} = 0.98 \pm 0.14(stat.)^{+0.17}_{-0.16}(syst.).$$

By combining the results coming from VH with the ones coming from ggH and ttH an excess compatible with 5.4 standard deviation is observed, to be compared with an expected one of 5.5 standard deviations. The measured signal strength is:

$$\mu_{Hbb} = 1.01 \pm 0.12(stat.)^{+0.16}_{0.15}(syst.).$$

From the results in fig. 4.1, the measurements are in agreement with the SM prediction.

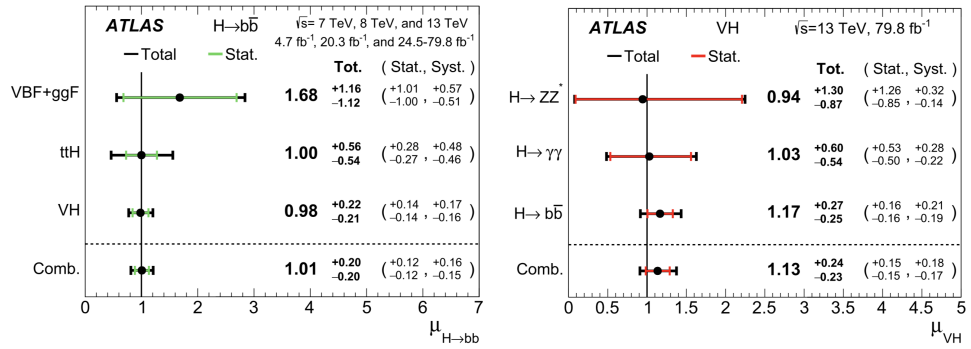


Figure 4.1: On the left, the fitted values of the Higgs boson signal strength H→ $b\bar{b}$ for a Higgs mass of 125 GeV for the VH, ttH and VBF+ggF processes. On the right, the fitted values of the Higgs boson signal strength μ_{VH} for a Higgs mass of 125 GeV for the $H \to b\bar{b}$, $H \to \gamma\gamma$ and $H \to ZZ' \to 4$ decay modes. The signal strength values are obtained from a simultaneous fit with the signal strengths for each process floating independently.

4.2 Analysis Overview

The process in which a boosted Higgs boson decays to a pair of b-quarks has been studied by the ATLAS collaboration using a dataset of 80.5 fb^{-1} with collision at a center-of-mass energy of $\sqrt{s} = 13$ TeV [71]. The study has been carried out in a broader context, i.e. searching for low mass resonances decaying to b-pairs. This study was motivated by exotic models predicting new light particles coupled with SM ones; an example is the leptophobic axial-vector Z'_A [73]. Since in ATLAS physicists mostly focused to searches for particle with masses above the TeV, the low mass multijet phase space was left partially unexplored, fig 4.2 [74, 75]. The b-tagging requirement on the multijet improves analysis sensitivity. Indeed, if we have a leptophobic axial-vector with Higgs-like coupling it will prefer the decay to b-pairs. Moreover, searches for b-pairs reduce the multijet background, which is highly dominated by light flavor jet.
Besides new physics searches, the other target of this analysis is to provide a measurement of the boosted H(bb) cross section focusing on the ggH production mode. At high-p_T its production is sensitive to possible contributions from new resonances and to the presence of anomalous couplings which can enhance the observed cross section significantly.
Even using the b-tagging, b-jets coming from QCD multijet background still represent the dominant source of background for this analysis. An additional difficulty of this search is that in the low mass range (from 70 to 230 GeV) there are broad resonances approximately every 30 GeV. This fact lowers the

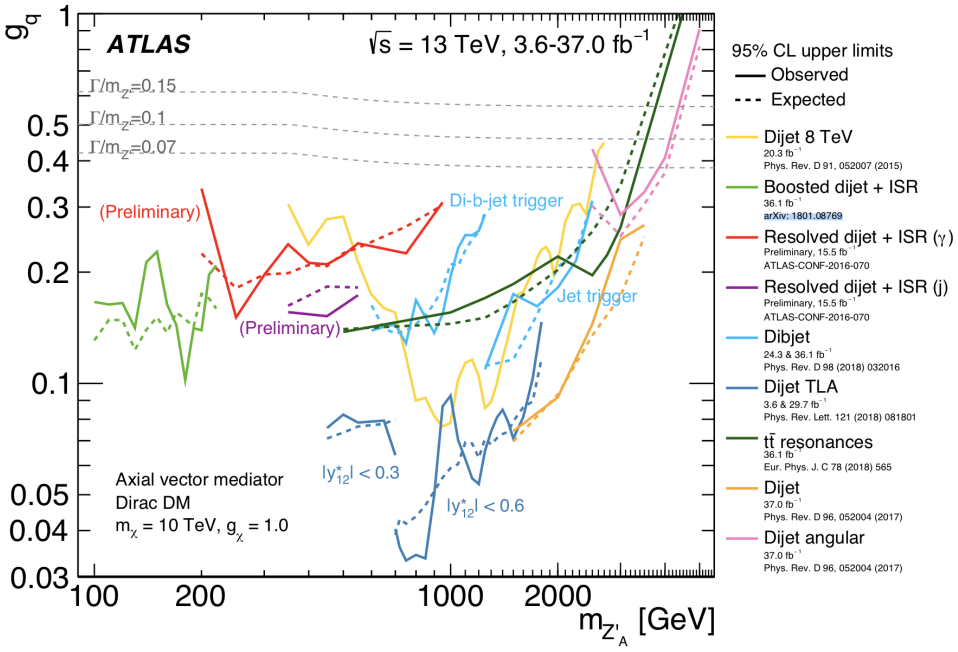


Figure 4.2: Dijet search contours for 95% CL upper limits on the coupling g_q as a function of the resonance mass $m_{Z'_A}$ for the leptophobic axial-vector $m_{Z'_A}$ model are shown. The expected limits from each search are indicated by dotted lines. Coupling values above the solid lines are excluded, as long as the signals are narrow enough to be detected using these searches. The dijet angular analysis is sensitive up to $\Gamma/m_{Z'} = 50\%$. No limitation in sensitivity arises from large width resonances in the $t\bar{t}$ resonance analysis. Benchmark width lines are indicated in the canvas. $\Gamma/m_{Z'} = 50\%$ lies beyond the canvas borders [76].

sensitivity to a new mass peak, due to the big modeling uncertainty. The strategy elaborated to mitigate the QCD background is to trigger on a high-p_T jet with an associated recoiling jet. High-p_T events have mainly two large-R jets, one for the signal and one for the recoil, since the system tends to be balanced. In fig. 4.3 there are the leading Feynman diagrams contributing to this signature.

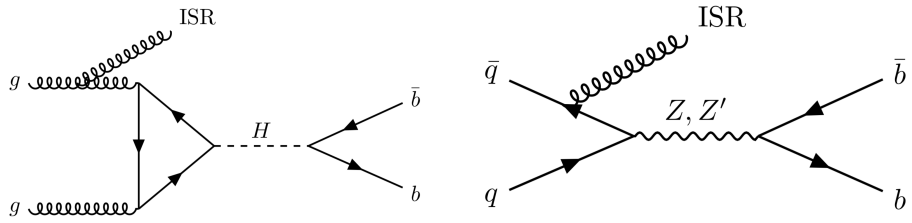


Figure 4.3: Leading order Feynman diagram for H+jet and Z'+jet produced at LHC.

In the boosted regime the dijet system is very collimated along the flight direction of the decaying particle and therefore it is reconstructed as a single large-R jets with $R = 1$ (see page 52). To reduce the complexity of the analysis, the recoiling jet is also reconstructed as a large-R jet. Furthermore, the large-R jets are required to have b-tagged sub-jets. Variable R jets have been chosen as sub-jets, because of their flexibility especially at high p_T.

4.3 Data and Simulated samples

The analysis has been carried out using 2015 data (3.2 fb^{-1}), 2016 data (33.0 fb^{-1}) and 2017 data (44.3 fb^{-1}) at $\sqrt{s} = 13$ TeV. The acquisition is triggered by at least one large-R jet with online p_T from 360 to 460 GeV. This threshold was gradually increased along the years, according to L1 trigger requirements to follow the increase of the instantaneous luminosity.

For what concerns MC samples, they have been used to develop analysis strategies and to model resonant and non-resonant backgrounds and signals. Two signal samples have been simulated: one for a SM $H \to b\bar{b}$ and one for the leptophobic Z' with democratic coupling with all quarks up, top excluded.
The Higgs is considered the main signal and therefore all analysis is optimized for it. Its simulated samples are produced considering all the main production modes which contribute to the signal region: ggF, VBF and VH (associated W/Z production). Different generators and procedures have been used to simulate them:

- ggF plus jet events are generated with the HJ+MiNLO prescription with finite top mass using Powheg-Box 2 [77, 78] and the NNPDF30 NNLO parton distribution function set [79]. The showering is performed using

Pythia 8.212 [80] with the AZNLO tune [81] and the CTEQ6L1 parton distribution function set [82]. The decay of b-hadrons is performed using EvtGen [83]. Since the analysis basic objects are high p_T large-R jet, the events are propagated by Powheg only if they have a Born k_T above 200 GeV.

- VBF events are generated in the same way of ggF samples but without the HJ+MiNLO prescription.

- VH, are generated using Pythia 8.212 with the AZNLO tune and the CTEQ6L1 parton distribution function set, with the decay of b-hadrons done with EvtGen. In this case, to account for higher order effects, the cross-section is corrected with k-factors 1.25 and 1.47, respectively for WH and ZH [84]. Only events with a boson with true $p_T > 350$ GeV are run with the full simulation chain.

The other simulated signal is the Z' decaying in two b-quarks. It has been simulated with Simplified Dark Matter (dma) $MadGraph5\_aMC@NLO$ model [85, 86] and the showering is done using Pythia 8 with the A14 tune and the NNPDF23 LO parton distribution function set. All events are required to have a jet with a generated $p_T > 350$ GeV and the jet must have a parton with $p_T > 100$ GeV. More samples were produced at different masses $m_{Z'}$: 100 GeV, 125 GeV, 150 GeV, 175 GeV, 200 GeV and 250 GeV. For Z' the total width is defined as:

$$\Gamma_{Z'} = \Gamma_{SM} + \Gamma_{DM}.$$

The definition of these two terms is the following,

$$\Gamma_{SM} = \sum_{q(m_q<m'_Z/2)} \frac{3m_{Z'}g_q^2}{12\pi} \times \left(1 - \frac{(2m_q)^2}{m_{Z'}^2}\right)^{3/2}$$

$$\Gamma_{DM} = \begin{cases} \frac{m_{Z'}g_{DM}^2}{12\pi} \times \left(1 - \frac{(2m_{DM})^2}{m_{Z'}^2}\right)^{3/2}, & \text{if } m_{DM} < m_{Z'}/2. \\ 0, & \text{otherwise.} \end{cases}$$

The calculation is implemented using the DarkMatterWidthCalculator [87,88]. For this simulation $m_{DM} = 10$ GeV and g_{DM} is set to 1, therefore the presence of Dark Matter is negligible and the total width is reduced to the first term. For what concerns g_q it was set to 0.25 for this study.

The other simulated samples are the background ones: QCD multijet is the non-resonant background while $t\bar{t}$,V+jet are the resonant ones.

- The QCD background is the dominant one and it was computed using Pythia 8.186 [80] with the A14 tune and the NNPDF23 LO PDF and the decay of b-hadrons is done with EvtGen [80]. The p_T spectrum of this process falls really quickly and therefore the production has been carried out splitting the simulated events in several slices of large-R jet p_T truth, to obtain samples with equivalent statistical power.

- The W+jet and Z+jet samples decaying to b-quarks have been produced using Sherpa 2.1.1 [89] with CT10 parton distribution function set. An alternative sample has been produced for both of them using Herwig++ 2.7.1 [90] to model account for possible generator uncertainty.
- The $t\bar{t}$ were generated at tree-level using Powheg-Box 2 [78] and the NNPDF30 NLO parton distribution function set. The quark hadronization is performed using Pythia 8.230 with the A14 tune and the NNPDF23 LO parton distribution function set. An alternative sample was simulated with Sherpa 2.2.1 using the NNPDF30 NNLO parton density function.

4.4 Event reconstruction

In this analysis large-R jets are widely used. Quality criteria are used to clean beam-induced background, coming from cosmic rays or from calorimeter noise [91]. They are reconstructed from calibrated topological clusters and trimmed using $R_{sub} = 0.2$ and $f_{cut} = 0.05$ to mitigate pile-up. The jet energy is corrected using a calibration (p_T, η) dependent [92]. In order to improve the mass resolution, the mass assigned to the jet is the track-assisted jet mass, built combining tracker and calorimeter information [93]. After reconstruction, only large-R jets with a $p_T > 250$ GeV and $\eta < 2.5$ are selected to be used in the subsequent steps of the analysis.

This analysis makes use of variable radius (VR) track-jets to access the large-R jets substructure [62]. Their variable radius is reconstructed using $\rho = 30$ GeV, $R_{min} = 0.02$ and $R_{max} = 0.4$. The tracks around which the VR jets are constructed must have $p_T > 0.5$ GeV and $|\eta| < 2.5$. Furthermore, some track quality requirements are applied to the tracks: at least seven hits in the pixel and SCT detector are required, no more than one hit shared between more tracks and no more than one missing hit in the pixels and two in the SCT. Final VR track jets are required to have $p_T > 10$ GeV and not to have complete overlap with each other. Tracks not originated from the primary vertex are rejected by requiring $|z_0 sin\theta| < 3$ mm. B-tagging is applied on selected objects, using the MV2c10 algorithm. Two different working points are used, corresponding to 77% (tight) and 85% (loose) b-tagging efficiency.

Finally, combined muons are reconstructed in order to use them in a $t\bar{t}$ control region. Similarly to VR track jets, they are required to have a p_T above 10 GeV and $|\eta|$ <2.5 . They are then selected requiring medium quality muons with loose isolation working point.

4.5 Event selection and classification

4.5.1 Trigger

Three large-R jets triggers, per year, have been used in a logic OR to exploit all the available luminosity from 2015 to 2017 data. All triggers have a threshold on the large-R jet p_T:

- 2015 trigger, HLT_j360_a10_lcw_sub_L1J100 with p_T offline threshold of 410 GeV;
- 2016 trigger, HLT_420_a10_lcw_L1J100 with p_T offline threshold of 450 GeV;
- 2017 trigger, HLT_j460_a10_lcw_jes_L1J100 with p_T offline threshold of 480 GeV;

The efficiency curves have been studied in MC samples and compared with data, as shown in fig. 4.4. Since the online threshold was raised year by year,

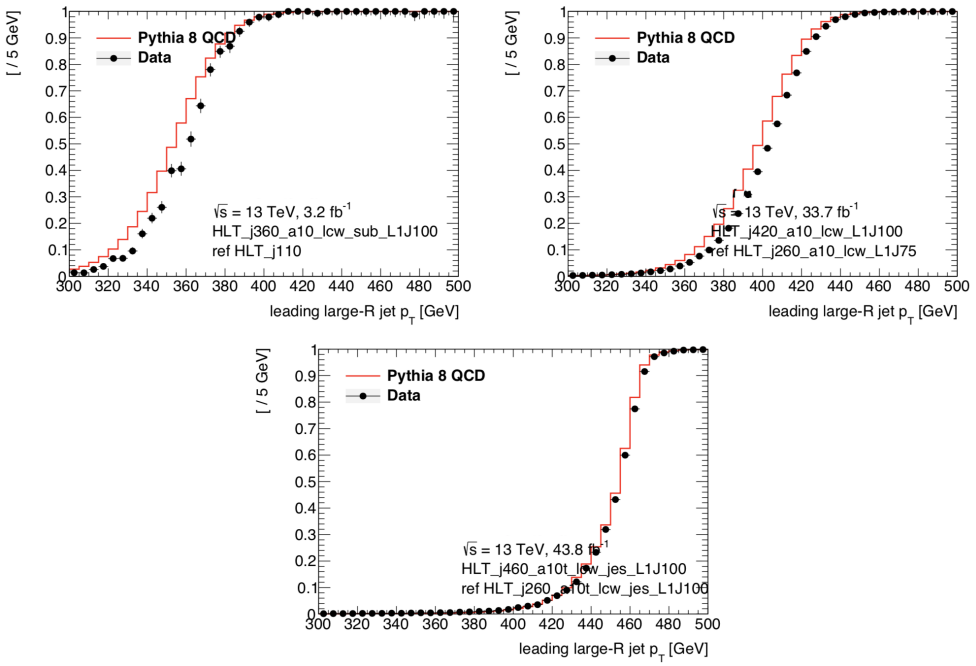


Figure 4.4: Efficiency curves for chosen trigger for the 2015 (up-left corner), 2016 (up-right corner) and 2017 (bottom). The efficiency is given as a function of the leading p_T large-R jet, which is always the one who fires the trigger in this case.

the highest p_T threshold have been used as an offline threshold for large-R jets from all data. This is done to avoid the possibility that a discontinuity in the p_T spectrum propagates to the mass spectrum.

4.5.2 Event selection

The selected events must have at least two trimmed large-R jets. The leading large-R jet must have $p_T > 480$ GeV and the subleading large-R jet must have $p_T > 250$ GeV. The leading jet p_T cut selects event kinematics for which all three triggers are fully efficient. The subleading jet p_T cut is used to reduce the size of MC samples.

Since the large-R jet is assumed to contain the decay products of the Z' or Higgs, a list of large-R jets is created with the ones who survive this pre-selection:

- $2m_J/p_{T,J}<1$ (boosted selection cut);
- at least 2 VR track jets with $p_T>10$ GeV;

The first step requires that the large-R jet is boosted and that all the decay products are collimated and contained in the cone size. The second step is performed to ensure ourselves to have at least two VR track jets to inspect for b-tagging, corresponding to the b-quark hadronizations. The highest p_T large-R jet of the remaining list is taken as the Higgs candidate, with an additional requirement of $p_T > 480$ GeV. This p_T cut is performed to ensure that the signal candidate large-R jet is the triggering one.

4.5.3 Event categorization

The event is further classified using b-tagging. The two p_T leading VR track jets contained in the signal candidate large-R jet are inspected for b-tagging. Two different working point have been used for this classification: 77 % and 85% b-tagging efficiency. The first one (tight WP) is used to ensure a low rate of wrong identified b-jet. The second one (loose WP) to be sure to reject a VR track jet only if the looser WP b-tagging fails. Then, two regions have been constructed: the signal region (SR), filled with signal events, and the CR_{QCD}, mostly filled by QCD background events. The CR_{QCD} is used to perform background modeling studies on the multijet background. These two regions are built in the following way:

- CR_{QCD} is composed by events where the signal candidate has exactly 0 loose b-tagged VR track jets;
- SR is composed by events where the signal candidate has exactly 2 tight b-tagged VR track jets;

The mass range considered for both regions is between 70 GeV and 230 GeV. This mass window has been chosen to ensure the equivalence between the shapes of the two regions on multijet background, see fig 4.5. Indeed, in the SR distribution below 70 GeV a b-tagging inefficiency in dense environments shows-up. In fact, low-mass objects with high p_T are very collimated and the

resolution on the jet-axis is poorer, degrading the b-tagging performance. The upper bound ensures that the VR track jets are still boosted along the large-R jet axis. This mass window is wide enough to contain the Higgs signal and all the resonant backgrounds, see fig. 4.6. Also the leptophobic Z' shapes have been checked in the SR for several mass hypothesis, see fig. 4.7. The flavor composition of the SR, as in fig. 4.8, is strongly dominated by events with 2 b-quarks, but there is a not negligible amount of events with one c-quark or light flavor quark instead of a b-quark. Events with no b-quark are just 5% of the total. This contamination is present because of the mis-tag rate of the MV2c10 tool [68].

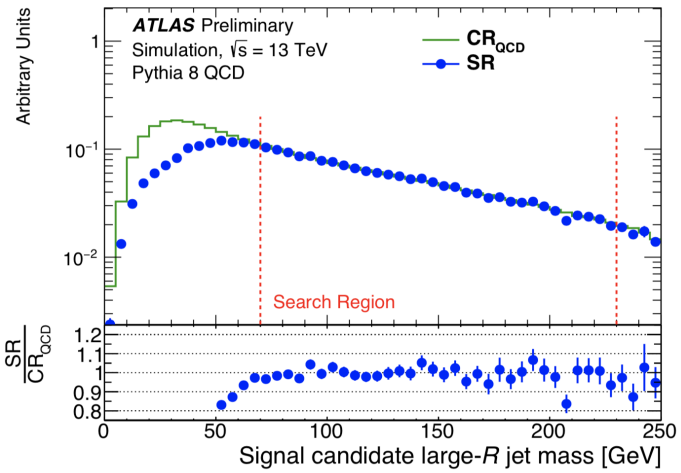


Figure 4.5: The shape comparison between the multijet background in the SR and in CR_{QCD} normalized to one in 70 GeV $< m_J <$ 230 GeV.

The fraction of remaining events all along the analysis cutflow for collision and Monte Carlo data, are reported in tabs. 4.1, tab. 4.2 and tab 4.3. As expected the main source of background over all the mass spectrum are multijet events, while for the resonance the main contribution is given by the $t\bar{t}$. For what concerns the V+jet the main component in the signal region is given the Z boson, since it can decay in a pair of b-jet, while in a decay of the W boson only one b-jet is produced.

4.6 Additional control region for the $t\bar{t}$ process

An additional control region for the $t\bar{t}$ process has been defined to derive the scale factor to correct the yield predicted with $t\bar{t}$ MC. This scale factor is obtained by fitting the normalization of an enriched $t\bar{t}$ MC template in the $CR_{t\bar{t}}$. It will be used to correct the yield of the $t\bar{t}$ template used to fit the data SR. The signal candidate for this CR is selected as explained before in pag. 71, then it is required that:

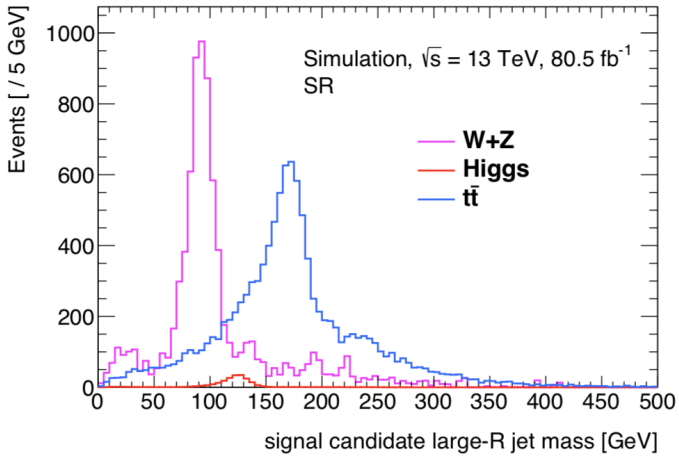


Figure 4.6: The shapes of the Higgs signal and of the resonant backgrounds (W, Z and $t\bar{t}$) in the SR, normalized to their respective expected yields, are shown in the plot.

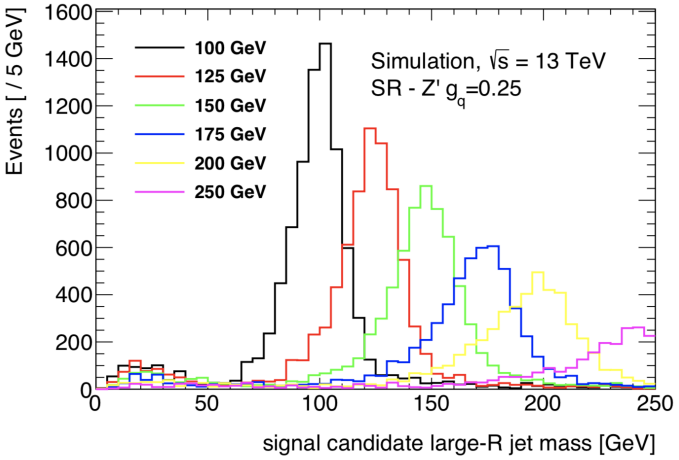


Figure 4.7: The shapes of the several simulated Z' with different masses, normalized to their respective expected yields.

	Cut	QCD	W+jets	Z+jets	$t\bar{t}$	Data
Preselection	Trigger	1.000	1.000	1.000	1.000	1.000
	Jet Cleaning	1.000	0.999	0.999	0.999	0.999
	Lead large-R jet p_T >480 GeV	0.200	0.301	0.290	0.263	0.484
	Sublead large-R jet p_T>250 GeV	0.181	0.284	0.275	0.218	0.446
	At least one signal candidate	0.160	0.272	0.263	0.210	0.402
	Signal candidate p_T > 480 GeV	0.135	0.246	0.236	0.177	0. 338
	Signal candidate mass> 40 GeV	0.099	0.208	0.198	0.159	0.248
CRqcd	0 b-tagged VR sub-jets (85% WP)	0.076	0.154	0.134	0.068	0.187
SR	2 b-tagged VR sub-jets (77% WP)	0.001	0.001	0.010	0.006	0.003

Table 4.1: The cutflow efficiency of the different regions using simulated background events and data is shown.

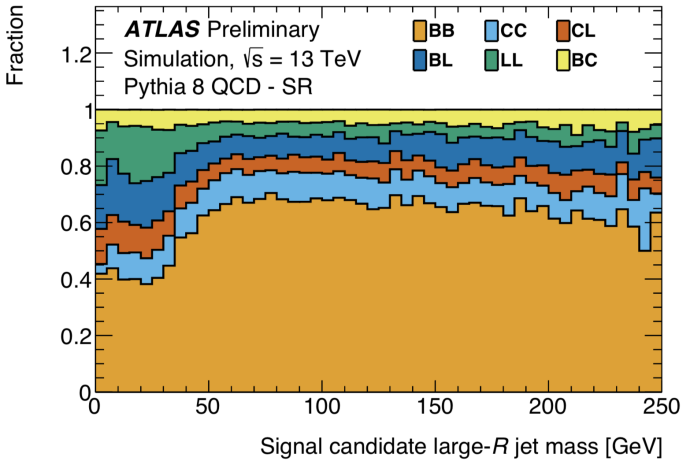


Figure 4.8: Predicted flavor composition of the multijet background in the SR based on the truth-matched hadron content of the two leading-p_T track-jets associated to the signal candidate large-R jet. B/C labels indicates the presence of a b/c-quark and L stays for the light quarks or gluons.

	Cut	ggF	VBF	WH	ZH	Total
Preselection	Trigger	1.000	1.000	1.000	1.000	1.000
	Jet Cleaning	1.000	0.999	1.000	1.000	1.000
	Lead large-R jet p_T >480 GeV	0.216	0.243	0.329	0.323	0.0.246
	Sublead large-R jet p_T>250 GeV	0.199	0.216	0.278	0.282	0.220
	At least one signal candidate	0.193	0.209	0.272	0.275	0.214
	Signal candidate p_T > 480 GeV	0.171	0.178	0.250	0.250	0. 190
	Signal candidate mass> 40 GeV	0.146	0.144	0.231	0.230	0.164
CRqcd	0 b-tagged VR sub-jets (85% WP)	0.051	0.038	0.107	0.098	0.060
SR	2 b-tagged VR sub-jets (77% WP)	0.042	0.050	0.039	0.051	0.044

Table 4.2: The cutflow of the different regions using simulated signal events is shown.

Process	CR_{QCD} Yield 80.5 fb^{-1}	SR Yield 80.5 fb^{-1}
QCD	29435344	400020
$Z \to q\bar{q}$ + jet	84389	6173
$W \to q\bar{q}$ + jet	219353	1506
$t\bar{t}$	110905	10553
$H \to b\bar{b}$ (ggF)	140	115
$H \to b\bar{b}$ (VBF)	41	53
$H \to b\bar{b}$ (WH)	71	26
$H \to b\bar{b}$ (ZH)	40	21
$H \to b\bar{b}$ (Total)	292	216
Z' (m = 100 GeV)	87739	8160
Z' (m = 125 GeV)	82796	7251
Z' (m = 150 GeV)	78257	6699
Z' (m = 175 GeV)	71144	5553
Z' (m = 200 GeV)	64816	4912
Z' (m = 250 GeV)	52008	3275
Data	29883000	484600

Table 4.3: The yields in the CR_{QCD} and SR for all the backgrounds, the Higgs boson and Z' boson signals and data.

- the signal candidate large-R jet has 1 b-tagged VR track-jet. This requirement ensures that the signal candidate is a top quark, which had a semileptonic decay;
- there is one loose isolated muon, with $p_T > 40$ GeV and $\Delta\phi(\mu, \text{large-R jet}) > 2\pi/3$. This requirement ensures that there is another top quark in the opposite hemisphere, which had a semileptonic decay;
- there is one extra large-R jet with one b-tagged VR track-jet with a $\Delta R(\mu, \text{large-R jet}) < 1.5$

The final estimation for the scale factor is $SF_{t\bar{t}} = 0.83 \pm 0.11$, which has been used in the final fit.

4.7 Signal and resonant backgrounds modeling

The signal and the resonant backgrounds (Z+jet, W+jet and $t\bar{t}$) are modeled using dedicated MC templates. The W and Z boson templates were merged in a single V template to facilitate the fit. To mitigate the effect of statistical fluctuations, a functional form was used to smooth the template distributions. All the templates have been fit in the 70-230 GeV mass range, except for the V+jet one, which was fit starting from 55 GeV in order to properly model the W tail, see fig. 4.9. Ad-hoc parametric forms have been chosen to fit the different templates:

- for the fit of the Higgs and Z' templates, a sum of three gaussians plus a constant term was used;

- for the fit of the V+jet template, a sum of three gaussians plus a constant term was used;
- for the fit of the $t\bar{t}$ template, a double-sided crystal ball was used.

All systematic variation histograms have also been smoothed with the same functional choices.

4.8 QCD modeling

An accurate estimation of the QCD background is fundamental to subtract it correctly and fit small resonances, such as the Higgs one. Two different procedures can be used to provide a robust estimation:

- estimation with MC templates: the templates are constructed running on simulated background events. This procedure is widely used when the event generator is reliable and the MC statistics is sufficiently high.
- data-driven estimation with a parametric function: the parametric function is built by finding the best agreement with data CR, which must be kinematically equivalent to the SR. This procedure is used when the MC statistics is not sufficient to build a reliable template.

The second approach was employed for this analysis because the MC accuracy was not enough to model the QCD background properly.
Therefore, the multijet background estimation was validated using the CR_{QCD} on data, while the V+jet and $t\bar{t}$ are taken from MC, with normalization fitted to data. Since the data CR_{QCD} contains more events than the SR, 60 independent data slices, each with statistics equivalent to the SR, have been obtained slicing the CR_{QCD} dataset. These slices have been used to provide independent samples to validate the background fit.

4.8.1 Parametric function selection

To fit the CR_{QCD} an exponentially falling function have been chosen, which can be defined as:

$$f_n\left(x|\vec{\theta}\right) = \theta_0 \exp\left(\sum_{i=1}^{N} x^i \theta_i\right)$$

where x = $\frac{m_J - 150GeV}{80GeV}$ maps the fit range from [70,230] GeV to [-1,1] and θ_i are the fit parameters. This function have been preferred to a simple polynomial since for the same amount of free parameters it is more flexible and suitable to describe a falling distribution: its derivatives maintain the same number of free parameters at every order. Good flexibility is required in this case, since there is no a priori reason why the QCD spectrum follows a functional behavior.
To select the minimum number of model parameters to describe the shape of the CR_{QCD} two different statistical test have been carried out:

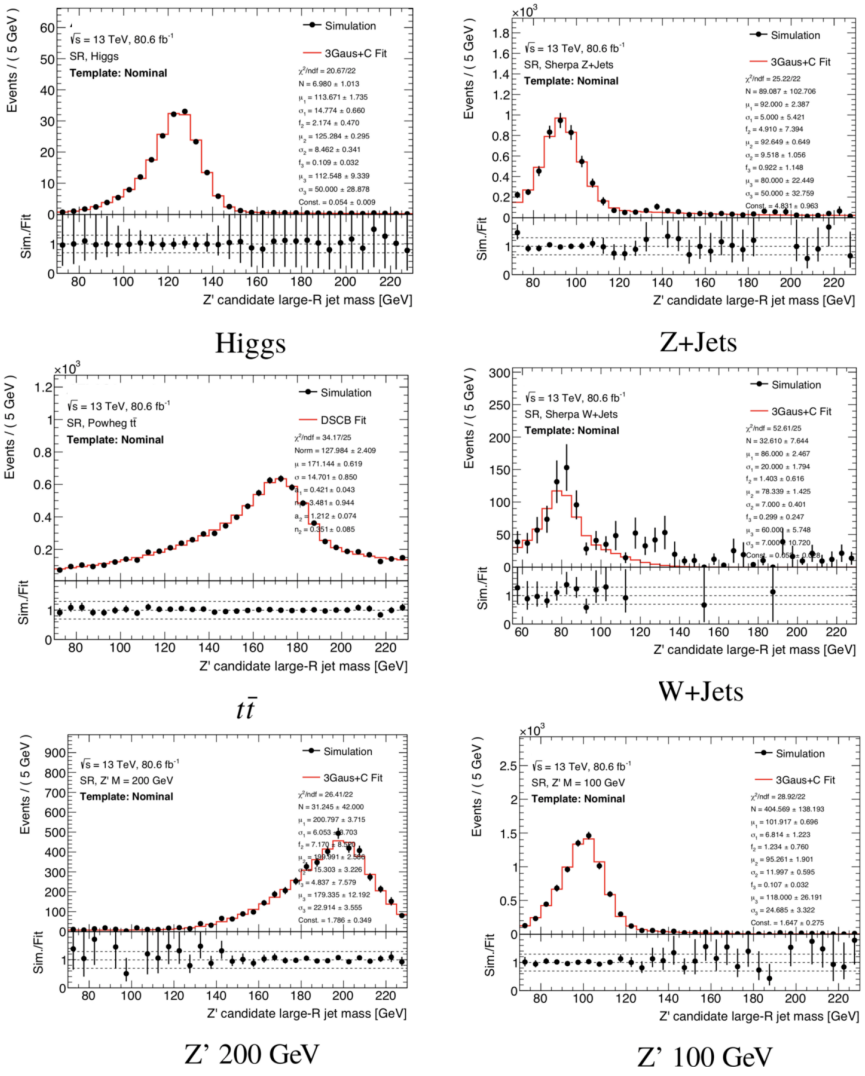


Figure 4.9: The results of the different parametric fits to the signal template are shown.

- the F-test [94]. The F-test statistics can be defined as:

$$F = \frac{\chi^2_2/\nu_2}{\chi^2_1/\nu_1}$$

where i=1,2 refers to the two compared models, while ν_i represent the model's degrees of freedom. The F-statistics is distributed according to the F-distribution [94], so the one-side p-value of this test statistics is evaluated as the complementary cumulative distribution function at the observed F.

- the log-likelihood ratio test [95]. The Wilk's theorem states that for two functions in a nested family of functions $f(x, \vec{\theta_a})$ and $f(x, \vec{\theta_b})$, where $a < b$ are the number of parameters, the log-likelihood ratio

$$t_\theta = -2 \log \frac{L(\vec{\theta_a})}{L(\vec{\theta_b})}$$

is asymptotically distributed as the χ^2 of a random variable with a-b degrees of freedom:

$$f(t_{\vec{\theta}}|\vec{\theta}) \rightarrow \chi^2_d$$

The one-sided p-value is the complementary cumulative distribution function (CCDF) of the χ^2_d distribution evaluated at the observed $t_{\vec{\theta}}$.

In both cases if the p-value is below a given threshold, $\alpha = (0.05)$, the difference in the fits of the two models is considered statistically significant. If this is the case, a model with a highest number of parameters is favored. The tests for the polynomial exponential family of functions were repeated for all CR_{QCD} data slices to construct the cumulative distribution function. The results obtained on a representative slice are shown in the table 4.4

Number of parameters compared	log likelihood test p-value	F-test p-value
4 vs. 5	0.0001	0.0785
5 vs. 6	0.3355	0.4867

Table 4.4: The observed p-values for the log-likelihood ratio test and F-test for comparing the polynomial exponential models with different number of parameters.

Since the log-likelihood ratio test favors the five parameters model, while the F-test the four parameters one, a conservative choice has been done by selecting the five parameters model.
The final step of the background modeling consists in the validation of the parametric function by fitting the data slices. Therefore, the slices with the parametric model plus templates for the Z+Jets, W+Jets, $t\bar{t}$ contributions have been fitted. This fitting model has a good agreement with data and χ^2/N_{dof} follows the expected behavior for the given number of data slices and degree of freedom, as shown in fig. 4.10

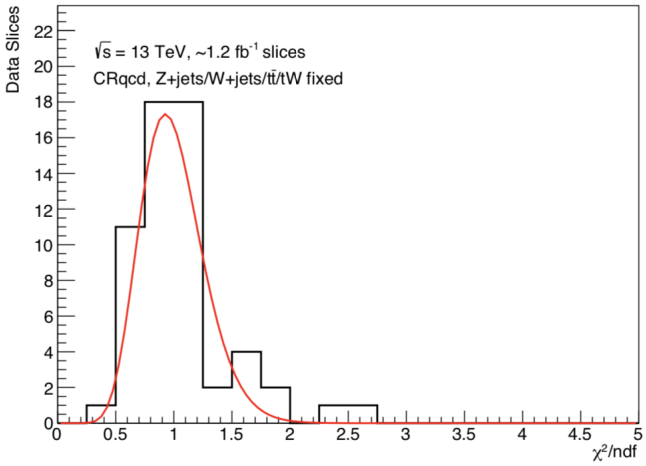


Figure 4.10: The distribution of the χ^2/N_{dof} obtained fitting all the slices, with the MC templates with fixed normalization. The red line is the expected behavior.

4.8.2 Spurious signal test

This procedure is run to see if the chosen parametric function is sufficiently robust against statistical fluctuations and does not extract a significant signal from a data slice with no signal injected. To build such a data sample, the CR_{QCD} slices are taken and all the resonances are subtracted by using the corresponding templates for V+jet, H+jet and $t\bar{t}$. If the fit extract a signal from the slices, then a systematic uncertainty needs to be added to cover this effect, if not, no further uncertainty needs to be added. Therefore, the test is performed as follows:

- a data slice of CR_{QCD} with resonances subtracted is build;
- a signal assumption is made and a template coherent with this assumption is built;
- the fit is run trying to extract this template from data and the statistical significance of this signal is checked, as μ_{fit}/σ_{fit}.

Different signal assumptions are tested. No significant signal was extracted for the Z', V and Higgs signal assumption. Given these results, no extra systematic uncertainty was added.

4.8.3 Signal injection test

This test is made to prove that parametric shape is smooth enough to allow to extract the same amount of injected signal and it was tested for the V+jet and the Z' signals, over a wide range of injected signal strengths. To do that, a proxy for the SR distribution was built summing up the H+jet, V+jet

and $t\bar{t}$ templates and the QCD parametric function with post-fit parameters, obtained with a fit of the CR_{QCD} with resonances subtracted. After that, pseudo-experiments (PE) were generated from this distribution, sampling each bin from a Poisson distribution with

$$\mu = bin_i^{Asimov} \text{ and } \sigma = \sqrt{bin_i^{Asimov}}.$$

Each of this pseudo-experiments is then fitted with the full-fit model and the relative difference between extracted and injected signal strength is computed,

$$pull = \frac{\mu_{fit} - \mu_{inj}}{\sigma_{fit}}.$$

For a reasonable high number of fitted PE the pull distribution becomes a Gauss distribution with mean 0 and standard deviation 1. Also in this case no particular trend was observed and the pull distribution is consistent with the expectations.

4.9 Systematic uncertainties

Systematic uncertainties are related to the simulation, to the detector, to the calibration procedure of the physics observables, and to the modeling of signal and background. These systematic uncertainties can be divided in two main categories: uncertainties on the QCD multi-jet background (data-related) and uncertainties on the resonant background and signals (MC-related). The first category is composed by:

- uncertainty on the fitting function: to estimate an uncertainty on the choice of the fitting function for QCD modeling, another model has been designed. The alternative model is a Laurent series defined as:

$$f_n^{alt} = f\left(x|\vec{\theta}\right) = a\sum_{i=0}^{n}\frac{\theta_i}{x^{x+1}}, \quad a = 10^5, \; x = \frac{m_J + 90GeV}{160GeV}$$

  Therefore, 100 toys were generated sampling from the nominal functional choice fit to a CR_{QCD} slice. These toys are then re-fit with the nominal and the alternate functional form. The mean and the standard deviation of the difference between the two functions creates an uncertainty band. A nuisance parameter is then added to interpolate inside the uncertainty band.

- statistical uncertainty on fit: this uncertainty represents the combination of the error on the fit coefficients. A single nuisance parameter has been defined for each parameter. The variation for a given parameter is obtained by varying it by $\pm$ 1 σ, $\pm$ 2 σ , etc... while keeping the others constant. They were then used with a uniform prior in the fitting procedure.

The uncertainties on the resonant background and signal are a much larger set:

- the uncertainty on the Large-R jet energy and mass scale and resolution. It accounts for 10 additional nuisance parameters;
- the uncertainties on the b-tagging. They are parametrized with separate nuisance parameters for b, c and light track-jets.
- the uncertainty on the measurement of the total integrated luminosity, which is used to normalize the MC yields to the measured data. It is quantified as 2.1% [96] and it is applied to the yields of all MC processes.
- the uncertainty on the normalization of the $t\bar{t}$ is a $\pm$ 0.11 (page 72), used as the width of the gaussian prior to constrain the normalization of the ttbar MC template in the SR fit.
- the uncertainty for the $t\bar{t}$ modeling is considered by allowing for an extra degree of freedom in the shape of the mass peak. A nuisance parameter was added to account for differences seen in the mass shapes between a Powheg (nominal) and Sherpa (alternative) sample, fig. 4.11. It has been implemented as a Gaussian prior, where 0 refers to a template equal to Powheg, and 1 equal to Sherpa (no values below 0 or above 1 are allowed).
- the uncertainty on the V+jets is accounted for by adding an extra nuisance parameter for differences seen in the mass shapes between a V+Jets Sherpa (nominal) and Herwig (alternative) sample, see fig. 4.11. It has been implemented as a Gaussian prior, where 0 refers to a template equal to Sherpa, and 1 equal to Herwig (no values below 0 or above 1 are allowed). More details on the study of the V+jet alternative generator can be found in the appendix A.
- the theoretical uncertainty for the Higgs signal is obtained summing in quadrature different sources of uncertainty: 30% coming from $ggH \to b\bar{b}$ prediction, 1.2 % coming from QCD scale, 1.8 % coming from PDF uncertainty and 0.7% from α_S.

The impact of each systematic on the signal significance is resumed in tab. 4.5. The dominant systematics are those concerning the jet mass and energy reconstruction and the theoretical uncertainty on signal and backgrounds. The impact of systematic, as large-R jet ones, which affects normalization and shape, is assessed by varying mass and energy within their uncertainty and propagating the effect in the signal extraction chain. The effect of b-tagging affects only the normalization and therefore it is assessed by varying yields within uncertainty and checking the effect on the signal. The impact of a single systematic on μ is computed performing fits with all the others nuisance parameters fixed.

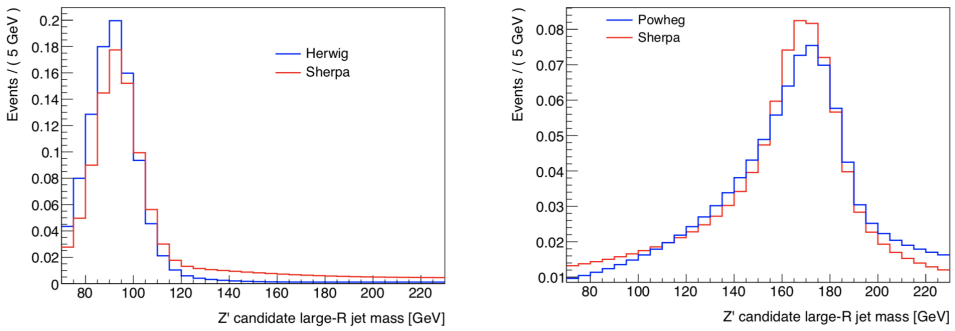


Figure 4.11: The difference between the nominal and the alternative generator mass distributions for the Z+jets (left) and for the $t\bar{t}$ is shown.

		Impact on signal($\sqrt{\Delta\sigma^2}/\mu$)			
Source	Type	V+jets	Higgs	Z' (100 GeV)	Z' (175 GeV)
Jet energy and mass scale	Norm. & Shape	15%	14%	23%	18%
Jet mass resolution	Norm. & Shape	20%	17%	30%	20%
V+jets modeling	Shape	9%	4%	4%	$< 1\%$
$t\bar{t}$ modeling	Shape	$< 1\%$	1%	$< 1\%$	11%
b-tagging (b)	Normalization	11%	12%	11%	15%
b-tagging (c)	Normalization	3%	1%	3%	5%
b-tagging (l)	Normalization	4%	1%	4%	7%
$t\bar{t}$ scale factor	Normalization	2%	3%	2%	58%
Luminosity	Normalization	2%	2%	2%	3%
Alternative QCD function	Norm. & Shape	4%	4%	3%	17%
W/Z and QCD (Theory)	Normalization	14%	–	–	–
Higgs (Theory)	Normalization	–	30%	–	–

Table 4.5: Summary of the impact of the main systematic uncertainties on the uncertainty σ on the measurement of the signal strength μ for the V+jets, Higgs boson and Z' signals.

4.10 Fitting Strategy

In this section details on the statistical procedure used to extract the results in this analysis are provided. A description of the use of the Bayesian Analysis Toolkit (BAT) framework [97] used to extract the observed signal significances and the limits on the Z' signals and of the search using BumpHunter [98] is provided.

4.10.1 The Bayesian Analysis Toolkit

The BAT is an analysis package based on Bayes Theorem [97]. In the estimation of the of H+jets and V+jets significance and in the limit setting for Z' models, this tool was used to calculate a full posterior probability distribution as a function of number of signal events, ν, for the signal under consideration. The expression of Bayes Theorem used in this toolkit to obtain the final marginalized (integrated) posterior, p(ν|Data), is:

$$p(\nu|Data) \propto \int \mathcal{L}(\nu|\theta)\pi(\nu)\prod_i \pi(\theta_i)d\theta$$

where θ_i are the nuisance parameters, ν is the parameter of interest (in our case the number of events for a given process), in this case the normalization of the resonance peaks, and $\pi(\nu)$ and $\pi(\theta)$ are the prior knowledge about the parameters. The prior knowledge is therefore encoded in the priors π, and is updated by the outcome of the experiment, represented by the likelihood $L(\nu, \theta|Data)$, in order to obtain the marginalized posterior p(ν|Data). In the marginalization the ν corresponds to the normalization of the V+jet and H+jet which are extracted simultaneously from a single fit. For each systematic, described in the previous section, a nuisance parameter is introduced, and they are treated correlated among signals. The marginalization over the nuisance parameters is performed by a Markov Chain Monte Carlo (MCMC) engine.
In the fit the signal or background process are represented by a separate template with a nominal shape and normalization, μ, which has an associated nuisance parameter which defines it in the calculation. For all resonant signals and backgrounds the nominal templates are normalized to 1, in order to have this value if the observed number of events for that process is equal to the expected one. In the combined fit to extract H+jet and V+jet normalization uniform priors are used for the signals, set to the yield of the MC templates in the signal regions, while a gaussian prior is used for $t\bar{t}$, with mean and width set to the values obtained by the dedicated CR.

The fitting procedure is therefore as follows:

- for each iteration of the MCMC, a set of parameters is drawn and the effects of each systematic are calculated and applied to the signal and background templates;

- These parameters are used to compute the expected mass distribution and are then compared to data.
- A log likelihood is calculated using the expected and the observed mass distribution.

In that way, the MCMC uses that likelihood and the priors for the parameters to determine the final posterior on number of signal events.
The expected limits on the dark matter search are found by fixing the number of expected events to the SM prediction (0 for exotic signals). The obtained distributions after marginalization were then used to generate pseudo-experiments. A distribution of re-fitted values was produced, where the mean of that distribution represents the expected value of the parameter of interest and the deviations from the mean define the uncertainty bands of the exclusion limits. The 95 % percentile of that distribution gives the observed exclusion limit at 95 % CL. These limits are therefore expressed as 95% Credibility Level (CL) limit on the cross section times acceptance times branching ratio, $\sigma\times$ A $\times$ BR.

4.10.2 Search using BumpHunter

To properly take into account the Look Elsewhere Effect [99, 100], the BumpHunter algorithm [98] is used. In this case, it was used to establish the presence or absence of the H + jets process or of new physics phenomena. It operates on the binned spectrum, comparing the background estimate (post fit distribution of V + jets plus $t\bar{t}$ plus QCD) with the data in mass intervals of varying contiguous bin multiplicities. It looks for signal in an increasing mass window: from a two bin window, up to range equal to the half of the mass spectrum. For each point in the scan, it computes the significance of the difference between the data and the background. In that way the algorithms performs a series of pseudo-experiments sampling from the background estimate to determine the probability random fluctuations in the background-only hypothesis would create an excess anywhere in the spectrum at least as significant as the one observed. The region of the bump is defined as the one with the smallest probability to be produced by a statistical fluctuation (background Poisson distributed).

4.11 Alternative offline event selection studies

I proposed alternative selection using the D_2 substructure variable (see page 55) to select the signal candidate Large-R jet. This substructure variable is one of the most powerful identifier of large-R jet with a two prong structure. Since it is the ratio between the two- and three-point energy correlation function, it is ≈ 1 for two prong jets and larger for other topologies. The proposal was to sort the large-R jet list, built after the selection, in increasing D_2 instead of decreasing p_T. The signal candidate would be chosen always as the first element of the list, which, in this case, is the large-R jet with the lowest D_2.

This sorting will reinforce the requirement of at least two VR jets in the large-R jet during pre-selection, narrowing the event topologies to the signal one. Figure 4.12 shows the mass distribution for the SR using the official and the proposed definition of signal candidate. The event population off the 125 GeV resonance peak is reduced using the D2 ordering, indicating that the large-R jet containing the Higgs boson is correctly picked more often. This translates in an improvement in sensitivity summarized $S/\sqrt{B}$, summarized in table 4.6.

Selection	$S/\sqrt{B}$ for V	$S/\sqrt{B}$ for H
p_T, cand $p_T > 250$ GeV	8.79	0.52
D_2, cand $p_T > 250$ GeV	11.11	0.54
p_T, cand $p_T > 480$ GeV	9.47	0.49
D_2, cand $p_T > 480$ GeV	9.42	0.51

Table 4.6: The $S/\sqrt{B}$ on Higgs and V bosons for different selection using either p_T or D_2 ordering and the effect of using the cut on the p_T of the signal candidate are shown.

The improvement of using D_2 is of the order of $\approx 5\%$ on Higgs signal. It is much more significant on V+jet without p_T cut on candidates $\approx 25\%$. The drawback of D_2 ordering can be seen in the QCD spectrum. The choice of the most two-pronged-like large-R jet pushes the shoulder to higher masses. This reduces the lever-arm below the Z peak for fitting a smooth function and makes estimation of the QCD background harder. Since the QCD modeling was one of the most problematic issue of the whole analysis, D_2 ordering was not used.

4.12 Results

The bayesian analysis toolkit is used to calculate the expected signal strengths and the posterior likelihood for the different signal hypotheses.
Two different results have been therefore obtained in this analysis. The first result is the combined fit of the V+jets and H+jets components and the second is the limit setting on the presence of low-mass resonances in the explored range. The post-fit distributions are shown in fig. 4.13.
The observed signal strength for the V+jets process is

$$\boxed{\mu_V = 1.5 \pm 0.22(stat.)^{+0.29}_{-0.25}(syst.) \pm 0.18(th.)},$$

corresponding to a significance of 5 standard deviations. This is the first evidence at the ATLAS experiment of the presence of a vector boson in final states decaying to bottom quarks in the boosted regime. For the H+jets process, the observed signal strength is

$$\boxed{\mu_H = 5.8 \pm 3.1(stat.) \pm 1.9(syst.) \pm 1.7(th.)},$$

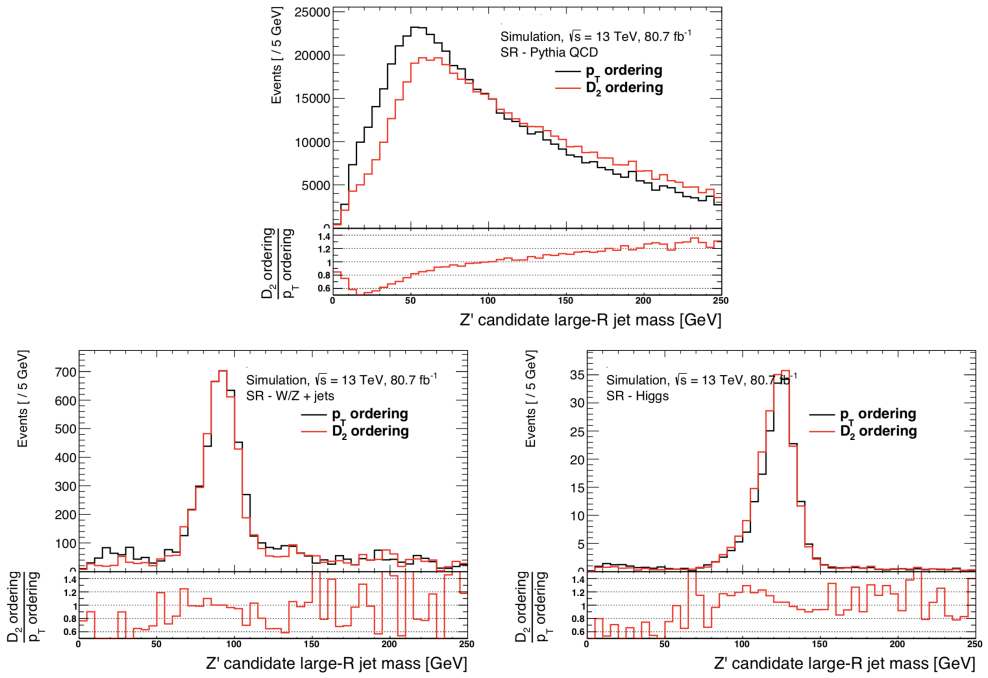


Figure 4.12: Comparison of the signal candidate mass spectra for the expected QCD background (top), V+jets (bottom-left) and Higgs signal (bottom-right) when p_T ordering or D_2 ordering is used. The signal region selection is applied.

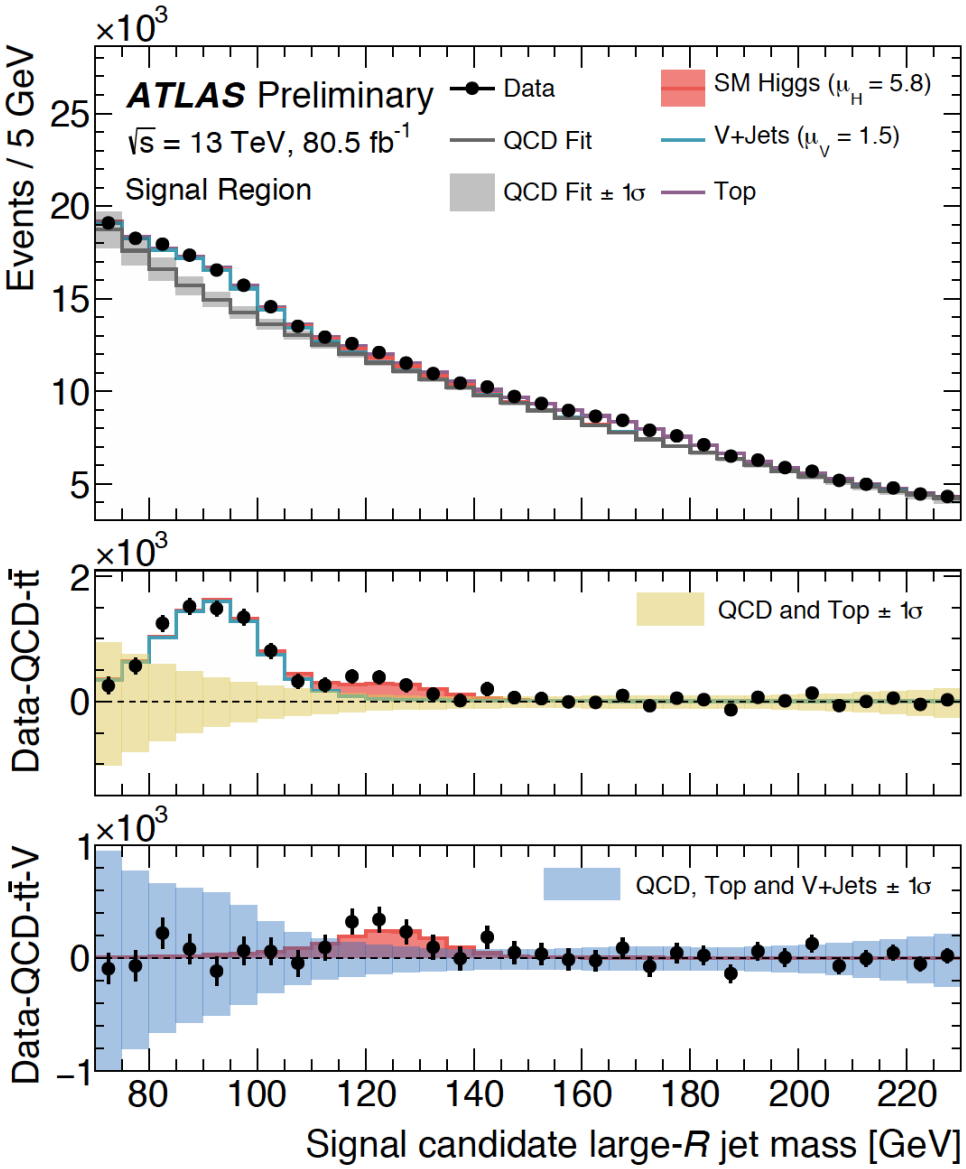


Figure 4.13: The SR with the different components (QCD,$t\bar{t}$,V+jet,H) fit to the distribution (top plot), the SR QCD subtracted with the signal templates fit to the distribution (middle plot) and the SR background and V+jet subtracted with the Higgs component from the fit in red (bottom plot) are shown in the plot.

consistent with the background-only hypothesis at 1.6σ. The observed significances represent the values of the excesses over the background, independently of the model used to interpret them. The breakdown of the error on the signal strengths in statistical and systematic is realized as follows: the statistical uncertainty is the one coming from a stat-only fit (turning-off all the systematics), while the systematic one is the sum in quadrature of the uncertainties introduced by each systematic independently (turning-off all the nuisance parameters except that one). It is therefore an approximation of the real uncertainty value, since it does not account for the correlation among uncertainties. The combined likelihood of the simultaneous fit of μ_V and μ_H is shown in fig. 4.14.

While for the V+j process the statistical uncertainty is of the same order of the systematic one, the extraction of the H+j yield is statistically limited. The impact of the systematic uncertainties is summarized in table 4.5. As expected, the main source of uncertainties is the jet mass/energy scale and the jet mass resolutions, since the two measured peaks are relatively close. Also the uncertainty on the b-tagging efficiency plays a not negligible role, since our SR has a double b-tagging requirement.
After the extraction of the signal strengths the BumpHunter searching procedure was performed. It assumes the post-fit shapes coming from the fit values for the $t\bar{t}$ and V+jets and searches for extra signals in the mass range, which are represented by deviation from the background model, see fig. 4.15.
The only significant deviation (p-value = 0.54) was found at a mass of 125 GeV, corresponding to the position of the Higgs mass peak. Finally, 95% confidence level intervals have been set on signals from Dark Matter mediators that decay to quarks, with masses between 100 and 200 GeV. These limits can be re-expressed as a function of g_q, the coupling parameter that controls the coupling of the DM mediator to quarks (see page 67). They are computed by multiplying the cross-section by g_q^2. These results are shown in figure 4.16.

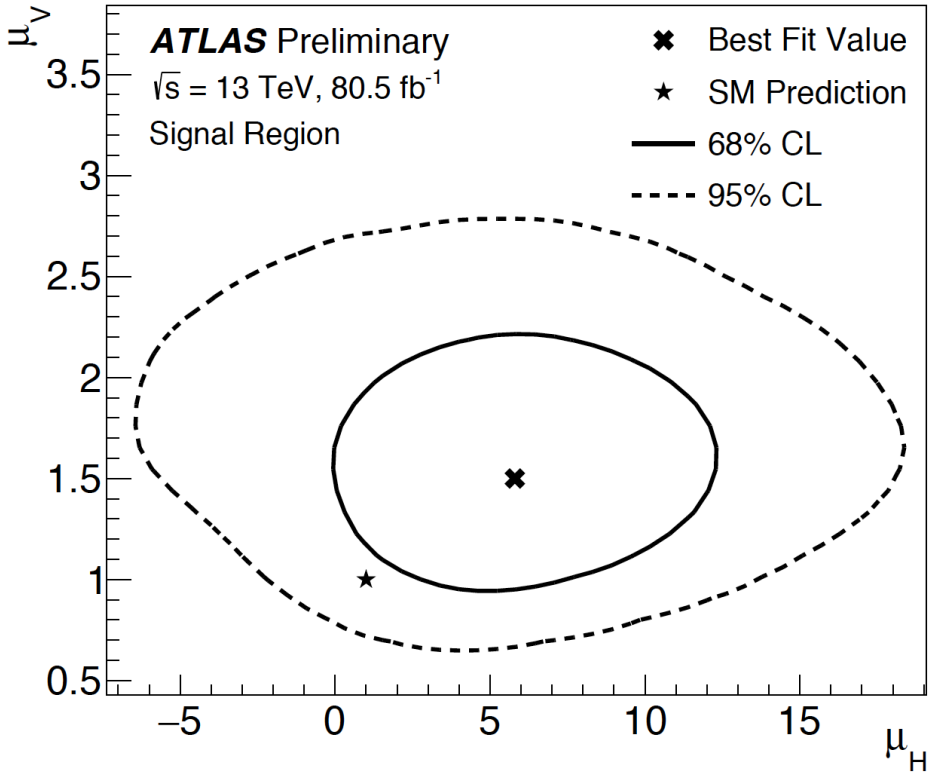
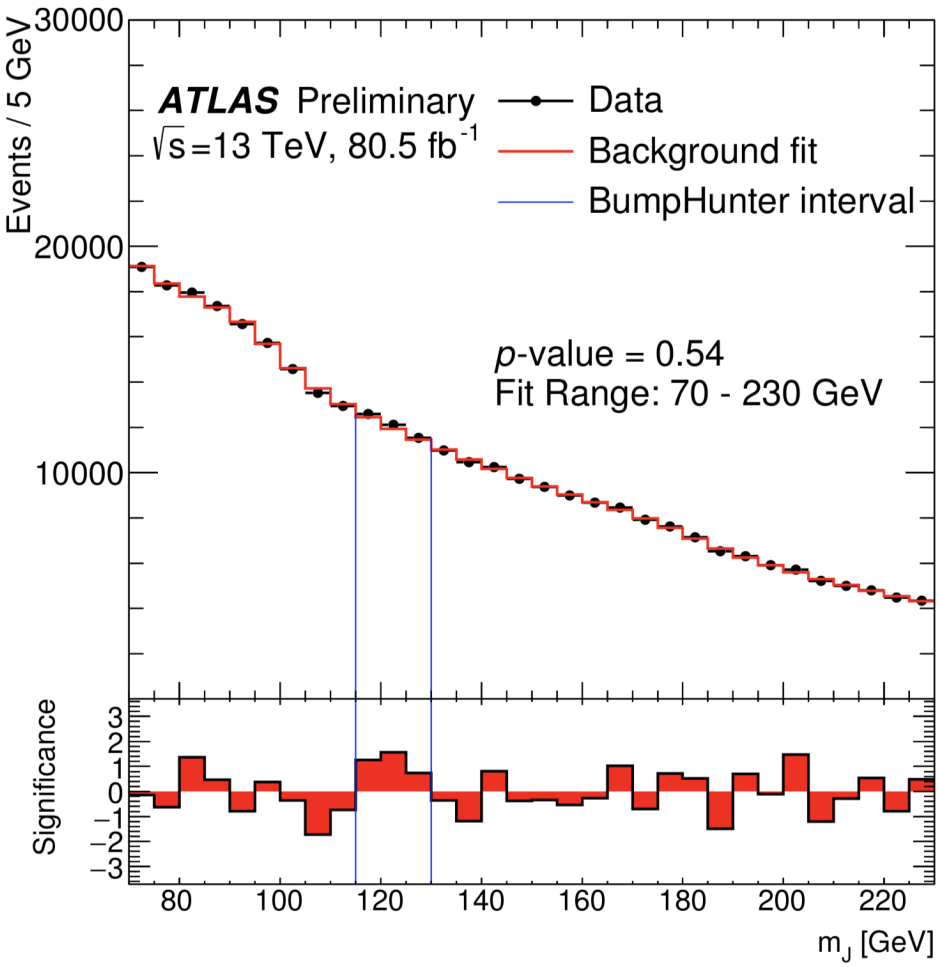
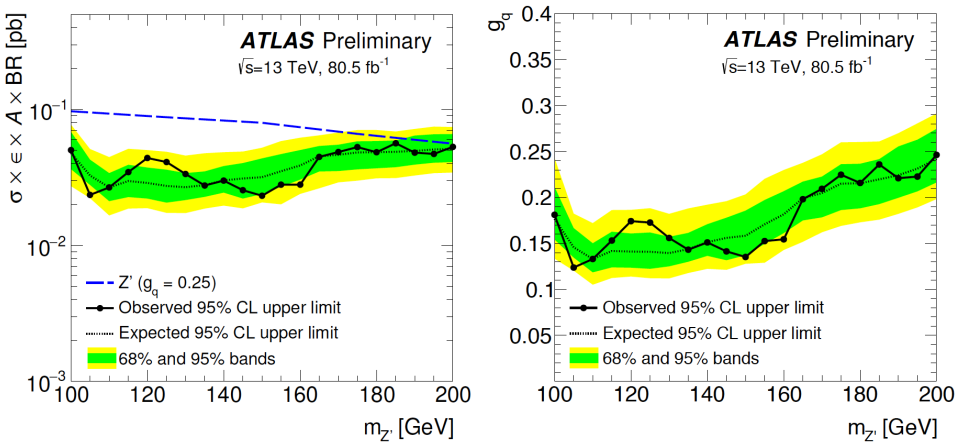


Figure 4.14: Combined probability distribution of μ_H and μ_V from the SR fit.



(a) (b)

Figure 4.15: The SR mass distribution is shown, the solid red line is the background prediction(QCD + V+jets and $t\bar{t}$). The blue lines point the region where BumpHunter [98] found the most discrepant bins. The low panel shows the local bin-by-bin difference between data and background divided by the statistical uncertainty.



(a) (b)

Figure 4.16: The 95% confidence-level upper limits obtained from the invariant mass distribution on the cross-section times acceptance times branching ratio times efficiency for the Z' model described in the text (a) and on the g_q parameter that controls the decay width of the DM mediator into SM particles (b).

Chapter 5

Full Run2 Analysis

During 2018 ATLAS collected 60.3 fb^{-1} from the collisions produced at LHC. The great majority of this integrated luminosity, 58.5 fb^{-1}, fulfills the data quality criteria and is good for physics analysis [101], fig. 5.1. This portion completes the Run 2 statistics and represents the bulk the total amount of data. Hence, the usage of data from 2015-2018 allows to perform a physics analysis on the largest high-energy physics dataset ever existed, approximately 140 fb^{-1}. The improved statistics is a good reason to rerun a large number

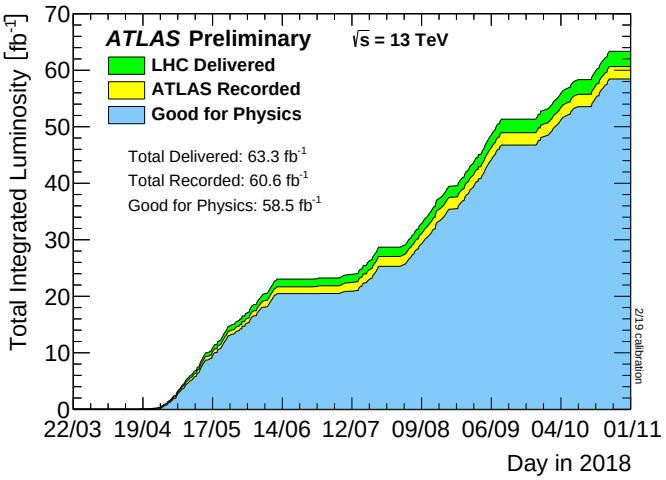


Figure 5.1: Integrated luminosity profile growing during the 2018. The plot shows the delivered luminosity by LHC, the recorded luminosity by ATLAS, which accounts for the DAQ inefficiency, and the luminosity good for physics [102].

of the ATLAS physics analysis. For most of them the statistical error is the dominant source of uncertainty and it would be reduced approximately by a factor $\sqrt{2}$. This is the case of the boosted Higgs to $b\bar{b}$ analysis as well, which is severely limited by the statistical error.

The full Run2 analysis has two planned deliverables:

1. an inclusive Higgs cross-section measurement;

2. a differential Higgs cross-section measurement, in three Higgs p_T bins;

For that purpose, some upgrades have been developed and implemented on top of the previous work, to improve the results more than what comes from the statistical factor. The focus of the measurement is on the differential result (p_T binned), which has never been shown for this regime of p_T. The p_T binned measurement is performed in three bins of the large-R jet p_T: the first with 250 GeV $< p_T <$ 450 GeV, the second with 450 GeV $< p_T <$650 GeV and the third one where 650 GeV $< p_T <$ 1000 GeV. This three bins of large-R jet p_T will be mapped in three Higgs Boson p_T bins [300 GeV,450 GeV), [450 GeV,650 GeV), [650 GeV, ∞).

5.1 Improvements

The improvements studied for this version of the analysis are mainly aimed at increasing the signal acceptance. Another objective is widening the p_T range of the reconstructed large-R jets and improving mass and p_T resolution.

5.1.1 Trigger strategy

A new set of triggers have been exploited in the full Run2 analysis. The main novelty is the presence of a mass cut on the large-R jet. This additional requirement on the jet substructure allows to relax the cut on the reconstructed large-R jet p_T. This has a strong impact on the analysis sensitivity, allowing to gain signal events at low p_T. Even if QCD contribution increases as well, the sensitivity grows because it behaves as $S/\sqrt{B}$ .The new trigger proposed to be included in the trigger strategy are:

- HLT_j360_a10_lcw_sub_L1J100, this trigger has online (offline) threshold on large-R jet p_T>360(410) GeV. It was applied on 2015 data;

- HLT_j420_a10_lcw_jes_L1J100, this trigger has online (offline) threshold on large-R jet p_T>420(450) GeV. It was applied on 2016 data;

- HLT_j390_a10t_lcw_jes_30smcINF_L1J100 with a logic OR with HLT_j440_a10t_lcw_jes_L1J100, the combination of these triggers has online (offline) threshold on large-R jet p_T>390(420) GeV and on mass 30(50) GeV. They were applied on 2017 data;

- HLT_j420_a10t_lcw_jes_35smcINF_L1J100 with a logic OR with HLT_j460_a10t_lcw_jes_L1J100, the combination of these triggers have online (offline) threshold on large-R jet p_T>420(450) GeV and on mass 35(60) GeV. They were applied on 2018 data;

The new triggers provide a large gain in signal acceptance on a large part of Run2 statistics. To give an order of magnitude of the gain using the new triggers, the signal efficiency is evaluated for ggF signal in a loose phase space and the results are listed in tab. 5.1. An additional advantage of using these new set of triggers comes from the mass cut, which helps reducing the QCD background a low masses. The efficiency curves have been studied to set offline threshold to reach the point at which the trigger is 99% efficient in data and MC (fig. 5.2, fig. 5.3, fig. 5.4, fig. 5.5).

Trigger	$\mathcal{L}$ $[fb^{-1}]$	$\epsilon_S(\%)$	$\epsilon_S \times \mathcal{L}$
HLT_j460_a10_lcw_jes_L1J100	**44.3**	**4.0**	**1.77**
HLT_j420_a10_lcw_jes_L1J100	33.0	9.0	2.97
HLT_j390_a10t_lcw_jes_30smcINF_L1J100	41.8	7.8	3.26
HLT_j420_a10t_lcw_jes_35smcINF_L1J100	58.6	5.5	3.22

Table 5.1: List of proposed trigger to cover mostly of the Run2 data compared to the trigger used in the past version of the analysis (in bold). The luminosity is the one at which offline large-R jets are 99%, while ϵ_S is the signal acceptance for that trigger in an inclusive phase space.

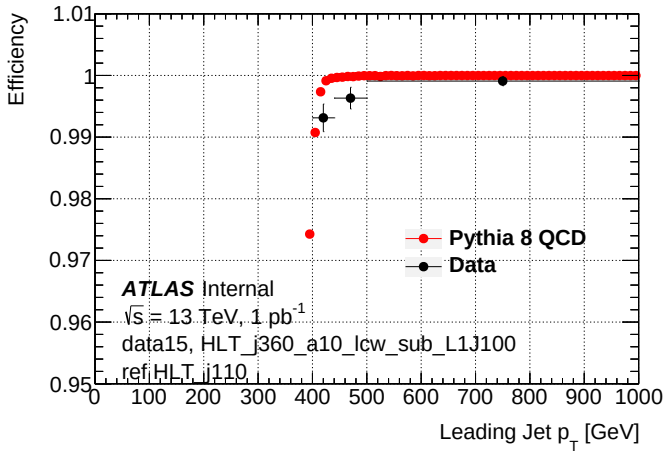


Figure 5.2: The efficiency of the trigger configuration active in 2015 data-taking period, as a function of the leading large-R jet p_T. The efficiency in MC is calculated as the fraction of event firing the trigger over the number of generated events at a given p_T/mass. The data efficiency is defined as the number of events passing the OR between the reference trigger and the trigger in question over the number of events firing the reference trigger at a given p_T/mass.

To have a unique trigger strategy for all the data years we use the logic OR of the aforementioned triggers with fixed thresholds in mass and p_T, which guarantee an efficiency of 99% for all the triggers across the years. Therefore, we retain only events in which at least one of the triggers fires and in which

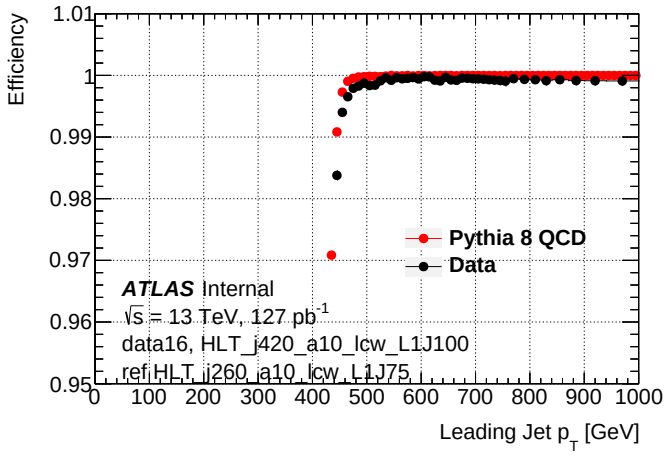
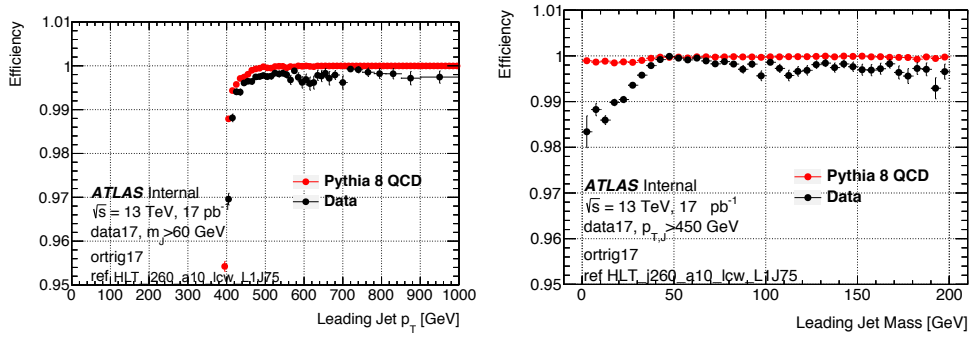
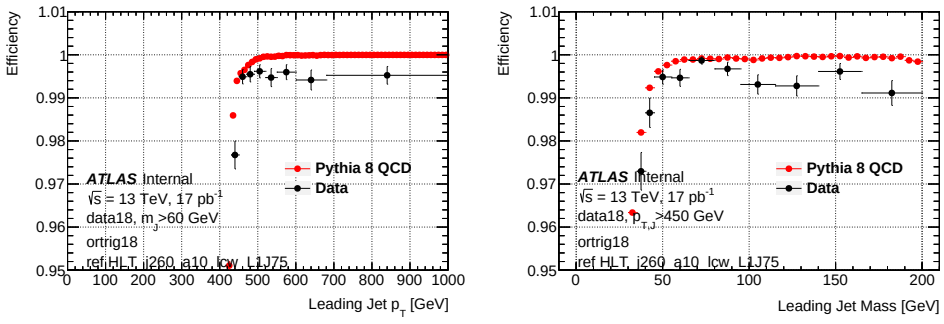


Figure 5.3: The efficiency of the trigger configuration active in 2016 data-taking period, as a function of the leading large-R jet p_T.



(a) (b)

Figure 5.4: The efficiency of the trigger configuration active in 2017 data-taking period, as a function of the leading large-R jet p_T (a) and of the large-R jet mass (b).



(a) (b)

Figure 5.5: The efficiency of the trigger configuration active in 2018 data-taking period, as a function of the leading large-R jet p_T (a) and of the large-R jet mass (b).

there is a large-R jet with $m > 60$ GeV and $p_T > 450$ GeV.

5.1.2 Subleading large-R jet

Another important improvement developed for this analysis is the use of the large-R jet with the second largest p_T in the event, referred to as subleading jet. This addition is motivated by the fact that, in the Higgs plus jet topology, events are dominated by two back-to-back large-R jets. Since the system is balanced in p_T the two objects must be close in p_T. Therefore, it is expected half of the Higgs bosons to be the leading jet and the other half the subleading jet in the event, see fig. 5.6. For that reason instead of limiting the analysis

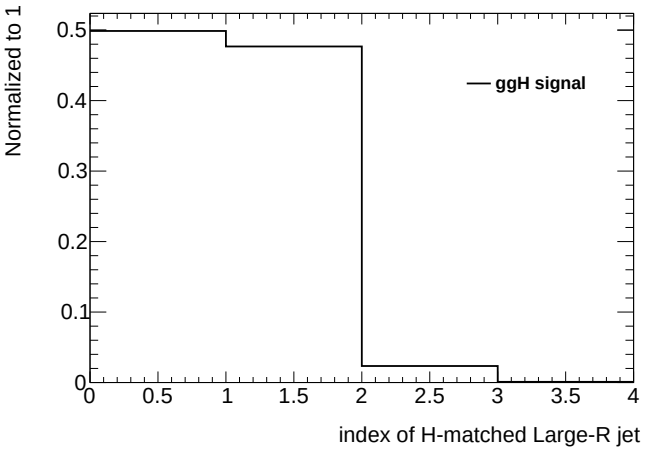


Figure 5.6: The index of the Higgs matched large-R jet in p_T order is shown for ggH signal sample. The index is p_T ordered so it is equal to 0 and 1 respectively for the leading and the subleading large-R jets.

to the leading large-R jet also the subleading jet was considered as a possible signal candidate. The gain from this change reflects both in sensitivity and in low large-R jet p_T events. For the leading and the subleading Higgs, two separate signal regions have been designed. More details will be given in the next sections.

5.1.3 Muon-in-jet correction

When a b-hadron decays semileptonically, the muon (electron) and the neutrino take away a fraction of the p_T from the objects, resulting in a p_T^{reco} lower than p_T^{true}. For this fact, the Higgs decaying semileptonically is often reconstructed as the subleading large-R jet of the event. This effect is evident by looking at the difference between the large-R jet matched with the Higgs and the recoil one, as shown in fig. 5.7, which presents a large tail due to semileptonic decays for the subleading Higgs bosons.

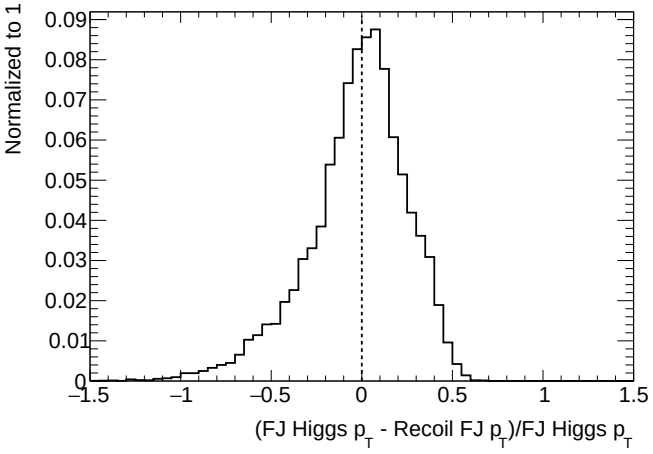


Figure 5.7: The relative difference between the p_T of the large-R jet matched with the Higgs and the recoil one is shown. Leading p_T large-R jet Higgs candidates are at the right hand of the dashed line, whereas subleading ones are at the left hand. The asymmetric tail is caused by the energy carried outside the jet by the muon and the neutrino produced in the semileptonic decays.

B-hadrons has a not negligible semileptonic branching ratio and dedicated energy corrections are being considered for this analysis. This correction takes into account the energy taken away by the muon (electron), but cannot consider the neutrino, which escapes detection. Standard jet energy calibrations do not account for these events, leading to a worse mass resolution in the case of semileptonic b-hadron decays for the signal processes considered here. Due to the difficulty of disentangling electron energy depositions in the calorimeter from the one of the jets, only muon-decaying b-hadrons are corrected for.

The general approach is to identify muons within the signal candidate large-R jet and add their four-vectors to the energy-momentum four-vector of the large-R jet. This correction is therefore called muon-in-jet correction. The correction used in this analysis had been already implemented for the $H \to b\bar{b}$ tagger [103]. In this case muons are required to have $p_T > 10$ GeV, $|\eta| < 2.4$, and be within $\Delta R < \min(0.4, 0.04 + 10GeV/p_T^{muon})$ of the b-tagged track jets. In case more than one muon is found within a given track jet, only the muon closer to the jet axis is retained. This correction has a large impact on the mass resolution, mostly affecting the subleading Higgs bosons, as shown in fig. 5.8. The percentage of the leading Higgs bosons corrected for the missing muon is much lower than the subleading ones and therefore the effect in the last one is enhanced, see tab. 5.2 .
An improvement of the mass resolution for the Higgs peak is fundamental in this analysis, since in the mass spectrum between 60 and 200 GeV there are several known resonant background contributions. The muon-in-jet correction

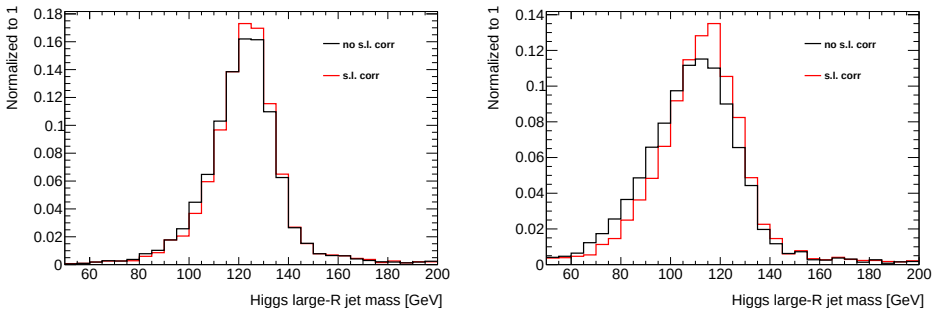


Figure 5.8: The mass distribution of leading and of subleading large-R jets matched with the Higgs boson, with and without the muon-in-jet correction.

Higgs matched Large-R jet	correction frac.	ΔM_{peak} [GeV]	$\Delta\sigma/\sigma$
leading large-R jet	0.33	0.45	-0.046
subleading large-R jet	0.13	1.52	-0.116

Table 5.2: The impact of the muon-in-jet correction on leading jet and subleading jet matched with the Higgs, in terms of fraction of corrected, peak shift and shrinkage of the distribution.

can indeed help to better constrain their width and improve separation between the vector bosons, $t\bar{t}$ and the Higgs, since there is not negligible amount of semileptonic decays for all these processes at this p_T regime.

5.2 Updates in Data and Monte Carlo Samples

The analysis uses of the ATLAS collisions dataset recorded from 2015 to 2018 (last version of the analysis up to 2017 data). Some events coming from period with bad detector performance have been removed and for that reason the integrated luminosity good for physics measurements amounts to 36.2 fb^{-1}, 41.0 fb^{-1} and 58.5 fb^{-1} respectively for 2015+2016, 2017 and 2018 data-taking periods. Summing up all years a total integrated luminosity of $136 fb^{-1}$ is obtained.

MC samples were largely inherited from the previous analysis, described in chapter 4. Therefore, in this section we will highlight what is different and what was updated because a sample with better calculation precision was available.
For what concerns the background samples:

- simulated QCD multijet events are generated using with the same strategy but using Pythia 8.235 [80] (Pythia 8.186 used previously);
- The reference V+jets sample is generated with Sherpa 2.2.8 [89] (Sherpa 2.2.5 used previously) with improved accuracy (QCD and EW NLO) and

it provides: PDF uncertainties (via internal weights), scale uncertainties in ME+PS , alternative fragmentation models.

Regarding the signal samples:

- The Higgs boson events produced in association with a W or Z boson are generated at tree-level using Powheg-Box 2 [78] at NLO (previously Sherpa was used) in QCD and the NNPDF30 NLO parton distribution function set. They include the $gg \to ZH$ contribution at leading order. The events are showered using Pythia 8.240 for hadronic and 8.212 for leptonic decays of the vector boson with the AZNLO tune and the CTEQ6L1 parton distribution function set. The decay of b-hadrons is performed using EvtGen [83].

- The ttH Higgs production mode was added to the analysis (not considered previously). The Higgs boson events produced in association with two top quarks are generated at tree-level using Powheg-Box 2 [78] at NLO in QCD and the NNPDF30 NLO parton distribution function set [79]. Separate samples are generated for each decay mode of the two top quarks: all-hadronic, semi-leptonic and di-leptonic. The events are then showered using Pythia 8.230 with the AZNLO tune and the CTEQ6L1 [82] parton distribution function set. The decay of b-hadrons is performed using EvtGen.

5.3 Event selection and categorization

In order to include the improvements and the new features of this version of the analysis, some items of the event selection and categorization have been updated. The most important change is that the categorization looks at large-R jets possibly having leading or subleading p_T in the event. This choice is motivated by the fact that we are looking for a system of two high-p_T large-R jets, and therefore the Higgs can be the subleading. With this new definition third ranked large-R jets of the event are discarded, but this does not impact sensitivity, as previously shown in fig. 5.6. The list of candidates is constructed looking at the large-R jet in the event and applying the selection listed in table 5.3.
Finally, new event selection is summarized in table 5.4.
After the event selection, we have a list of possible candidates to enter the signal or validation regions. For the inclusive measurement, we categorize the events in 4 regions: two separate signal regions for leading and subleading large-R jets and two related validation regions. The categorization, described in the diagram in fig. 5.9, is performed looking at the b-tagging of the two leading variable R jets in the candidate in this precise sequence:

1. if the leading large-R jet has exactly two b-tagged variable R jets it enters the signal region of the leading jets (SRL);

Signal candidate large-R jet list selection
$2m/p_T<1$, to have only boosted large-R jets in the list
at least two VR jets with $p_T > 10$ GeV inside the large-R jet
at least two VR jet with $p_T > 10$ GeV
no fully overlap with any other VR jet with $p_T > 5$ GeV
$p_T > 250$ GeV and m> 60 GeV

Table 5.3: Requirements on the large-R jet reconstructed in the event to enter the signal candidate large-R jet list

Event selection
one of the selected triggers is fired
at least two large-R jet with $p_T > 200$ and $\|\eta\| < 2$ in the event
$p_T>450$ GeV and m>60 GeV for at least one large-R jet in the event
at least one candidate large-R in the event

Table 5.4: Description of the selection applied to the events in the analysis.

2. if the subleading large-R (if any) jet has exactly two b-tagged variable R jets it enters the signal region of the subleading jets (SRS);

3. if the leading large-R jet has exactly zero b-tagged variable R jets and it is the only large-R jet in the list of candidates, it enters the validation region of the leading jet (VRL);

4. if the leading and the subleading large-R jets have exactly zero b-tagged variable R jets they enter the validation region of the leading of the subleading jet (VRS) respectively.

When the event is categorized in one of the four regions, the muon-in-jet correction is applied to the candidate large-R jet. Events are further categorized in p_T bins, based on the of the large-R jet after eventual muon-in-jet correction. These categories are: 250-450 GeV (SRx0/VRx0), 450-650 GeV (SRx1/VRx1), 650-1000 GeV (SRx2/VRx2) with separate regions for leading (x=L) and subleading (x=S) jets. The lowest analysis p_T bin is only populated by the subleading candidates due to the trigger cuts requiring a jet with p_T above 450 GeV. The signal regions are used in the final fit to extract the contribution of each process to the data mass spectrum, while the validation regions are utilized to perform background modeling studies and validation.
In fig. 5.10 the event fraction after each cut is shown for resonant backgrounds and signal, while in the tables 5.5 and 5.6 the number of events surviving each cut is shown. To render the impact of the selection on signals and backgrounds, we report that for ggF and QCD. For ggF and QCD multi-jet production events survival rate is of the order 10% and of 0.1% respectively.

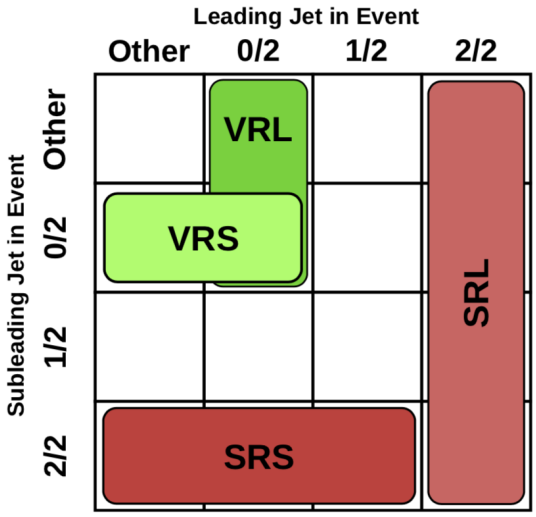


Figure 5.9: A diagram showing the event categorization criteria. The columns (rows) are divided into 4 categories: "other" when the leading (sub-leading) jet is not a candidate jet , "0/2" when neither of the first two p_T-ordered track jet is b-tagged, "1/2" when one of the track jets is b-tagged and "2/2" when both track jets are b-tagged

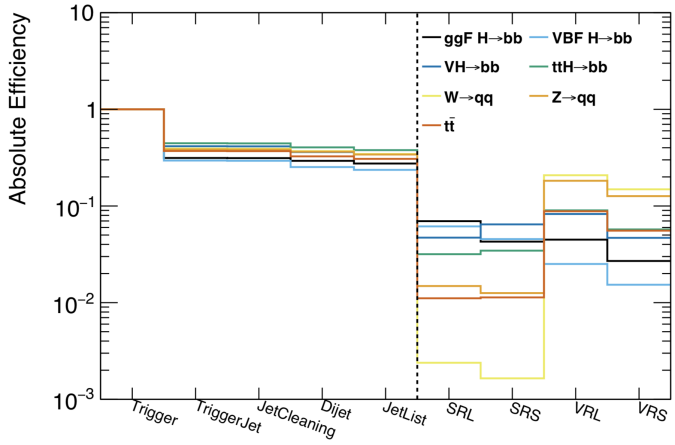


Figure 5.10: The fraction of events surviving each analysis cut for the resonant samples. The dashed line marks the separation of the selection and the categorization phase.

Cut	**QCD**	**W+jets**	**Z+jets**	$t\bar{t}$
Trigger	573216064	2189060	897215	2226760
≥1 large-R jet, p_T >450 GeV, M>60 GeV	116799232	859557	345338	825079
Jet Cleaning	116404304	856312	344112	822212
At least one signal candidate	96269984	754917	305728	684132
Signal Region Leading	811931	5226	13328	24702
Signal Region Subleading	637357	3605	11268	25223
Validation Region Leading	59733908	455415	163603	196199
Validation Region Subleading	40467792	325614	113495	123659

Table 5.5: The MC prediction for the expected number of events surviving each analysis cut for all background processes

Cut	**ggF**	VBF	**VH**	**ttH**
Trigger	3257	1062	1504	6583
≥1 large-R jet, p_T >450 GeV, M>60 GeV	1022	313	624	2934
Jet Cleaning	1019	311	622	2924
At least one signal candidate	895	251	515	2491
Signal Region Leading	226	65	70	208
Signal Region Subleading	139	48	97	227
Validation Region Leading	146	26	124	593
Validation Region Subleading	87	16	70	376

Table 5.6: The MC prediction for the expected number of events surviving each analysis cut for all the Higgs signal processes

5.4 Signal Modeling

The inclusive SR and VR mass distribution for the different decay modes are shown in fig 5.11. The peaks in the validation regions represents the mass of the resonant recoiling jets: the W and Z bosons, the Higgs and the top quark. In signal regions, the excess above the top mass in the ttH production channel represents events in which we pick as signal candidate one of the two top quarks. Figure 5.12 shows the contribution of different production modes as a function of the signal candidate large-R jet p_T in the inclusive signal region Higgs mass window, defined as 105 GeV $< m_J <$ 140 GeV, while tab. 5.7 shows the fractional contribution of each production mode in the inclusive and differential signal regions.
The calculation of the Higgs cross sections at high p_T have been recently improved. The ggF cross section computation is now available at NNLO order in the heavy top quark effective limit and at NLO in full QCD. Electroweak corrections to the VH, VBF and ttH processes at high p_T are implemented in the samples used in this analysis. The LHC Higgs cross section working group has recently summarized the status of these calculations and issued a set of recommendations [104]. In the final fit procedure only the mass cumulative templates, including all the production modes, will be used to extract the

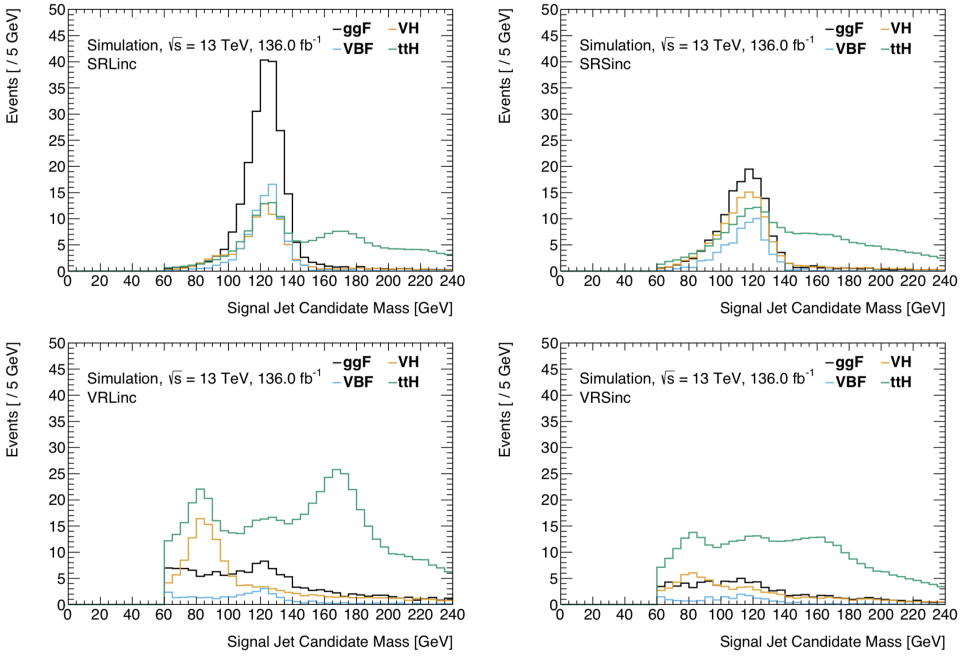


Figure 5.11: The mass distribution for the four main production modes entering the SRs and the VRs: the inclusive leading (left) and subleading (right) signal (top) and validation (bottom) regions are shown.

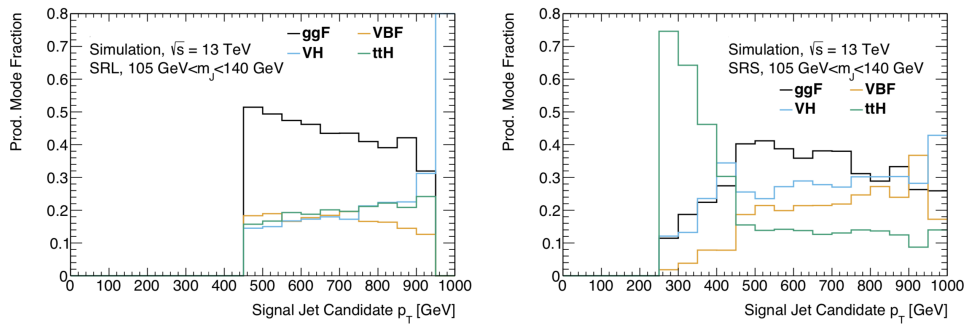


Figure 5.12: p_T distribution of the signal candidate large-R jet p_T for the four main production modes entering the SRs: the inclusive leading (left) and subleading (right) signal regions are shown.

Mode	Inc.	$250 < p_T^{jet} <$450 GeV	$450 < p_T^{jet} <$650 GeV	$650< p_T^{jet} <$ 1000 GeV
	SRL			
ggF	0.53	-	0.54	0.49
VBF	0.16	-	0.16	0.16
ttH	0.18	-	0.17	0.22
VH	0.14	-	0.13	0.14
	SRS			
ggF	0.36	0.24	0.44	0.41
ttH	0.27	0.47	0.15	0.14
VH	0.23	0.23	0.23	0.25
VBF	0.14	0.06	0.18	0.21

Table 5.7: The fractional contribution of each production mode to a given analysis bin in the Higgs peak, defined by 105 $GeV < m_J <$ 140 GeV. The fraction is given with respect to the total yield in the p_T bin

Higgs signal strength, see fig. 5.13. While in the SRs only the Higgs peak is present, in the validation region there are three structures in both the leading and the subleading templates. These bumps represent events in which also the recoil jet is selected as signal candidate and therefore the V+jet, the Higgs (when it does not have 2 b-tagged VR track-jets) and the $t\bar{t}$ peak are present.

5.4.1 Signal region mass range

The VRs are used to model data without looking at SRs. VRs are constructed to be nearly indistinguishable from SRs kinematics. As shown in fig. 5.14, the main difference between them comes from the lowest end of the mass spectrum. At low masses, for a given p_T, the jet boost increases and the b-tagging performance degrades. In particular, the efficiency decreases and any difference in the b-jet content between SRs and VRs immediately translates into visible discrepancies in the mass spectrum. Since in this low-mass region the SR and VR are not equivalent, we start the fit at masses around 70 GeV, where the turn-on is not present. The MC distributions and the ratio plots are shown in fig. 5.14 for the mass and the p_T of the candidate large-R jets. Differential results are shown in fig. 5.15. The starting point of turn-on slightly depends on the jet p_T of the sample: it begins at lower masses for lower p_T bins and begins at higher masses for the largest p_T bin. The higher bound instead is limited by boosting requirement ($2M/p_T>1$) and it was set to 210 GeV. Slightly different mass windows have been optimized for the SRs in which the fit is performed:

1. SRL mass window 65-210 GeV;
2. SRS mass window 70-210 GeV;
3. SRS0 mass window 70-210 GeV;
4. SRS1 mass window 70-210 GeV;

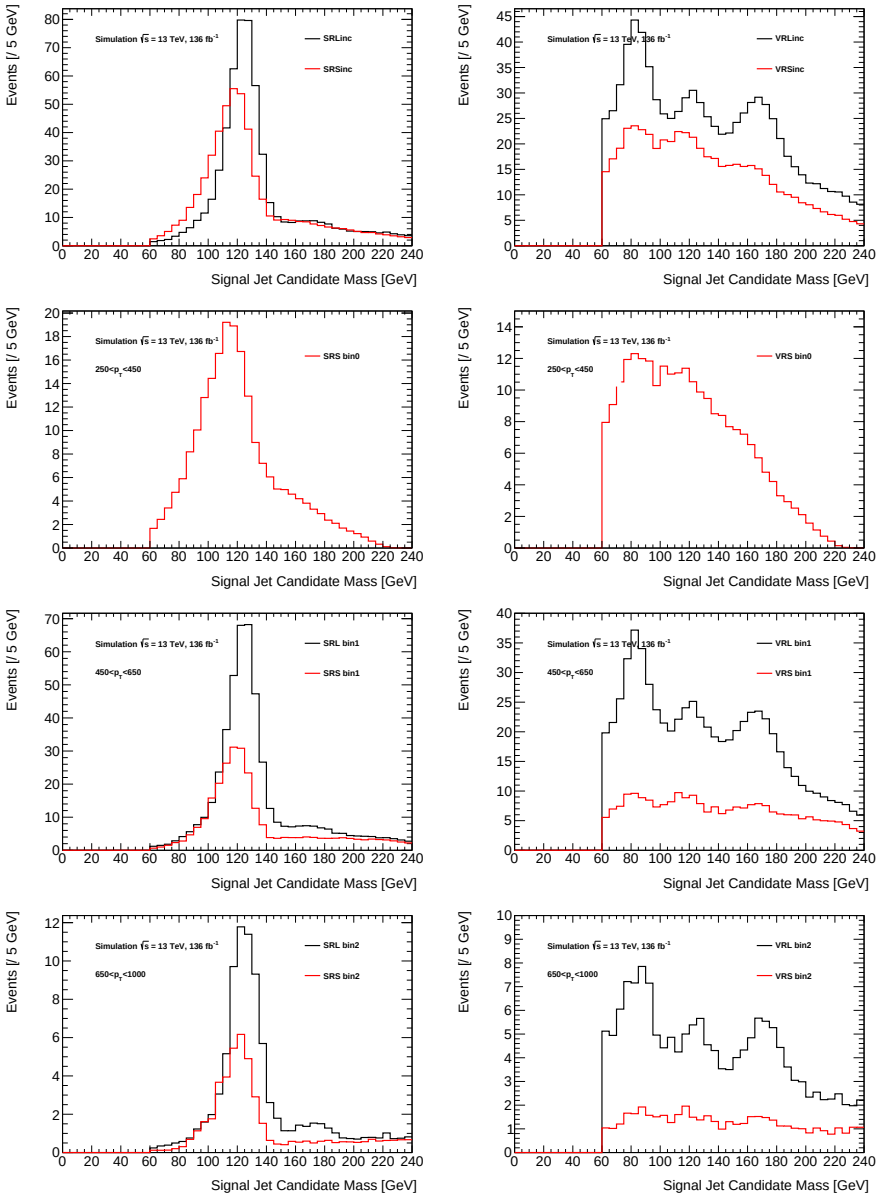


Figure 5.13: The mass templates for the leading and the subleading signal (left) and validation (right) regions are shown for the inclusive and the p_T binned ones. In the first p_T bins only the subleading distributions are present, since, in that region, the leading large-R jets would be below the trigger p_T threshold.

5. SRS2 mass window 70-210 GeV;
6. SRL1 mass window 65-210 GeV;
7. SRL2 mass window 70-210 GeV;

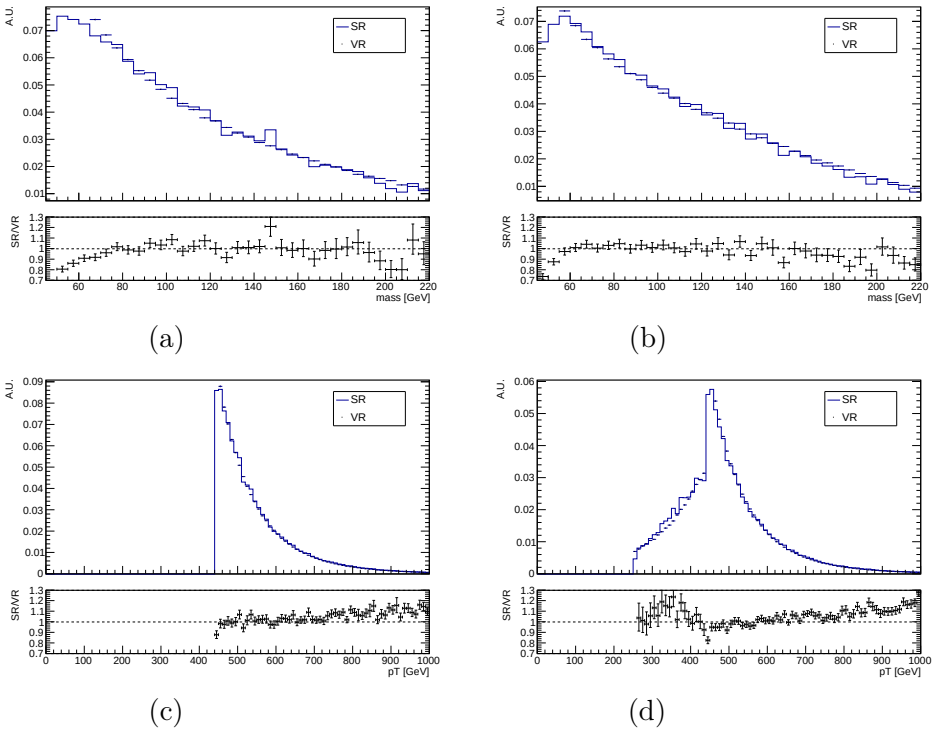


Figure 5.14: Comparison of signal and validation region for leading (a and c) and subleading (b and d) large-R jets. Monte Carlo samples are used, both for mass and p_T spectra. The low-mass cut has been removed for these plots to show the SR/VR discrepancy at low mass values. The area of distributions between 70 and 210 GeV is normalized to 1. The error bars only represent the statistical error. The mis-modeling seen the p_T as well, as others discrepancies in MC between VRs and SRs, is the reasons why a data-driven approach was preferred.

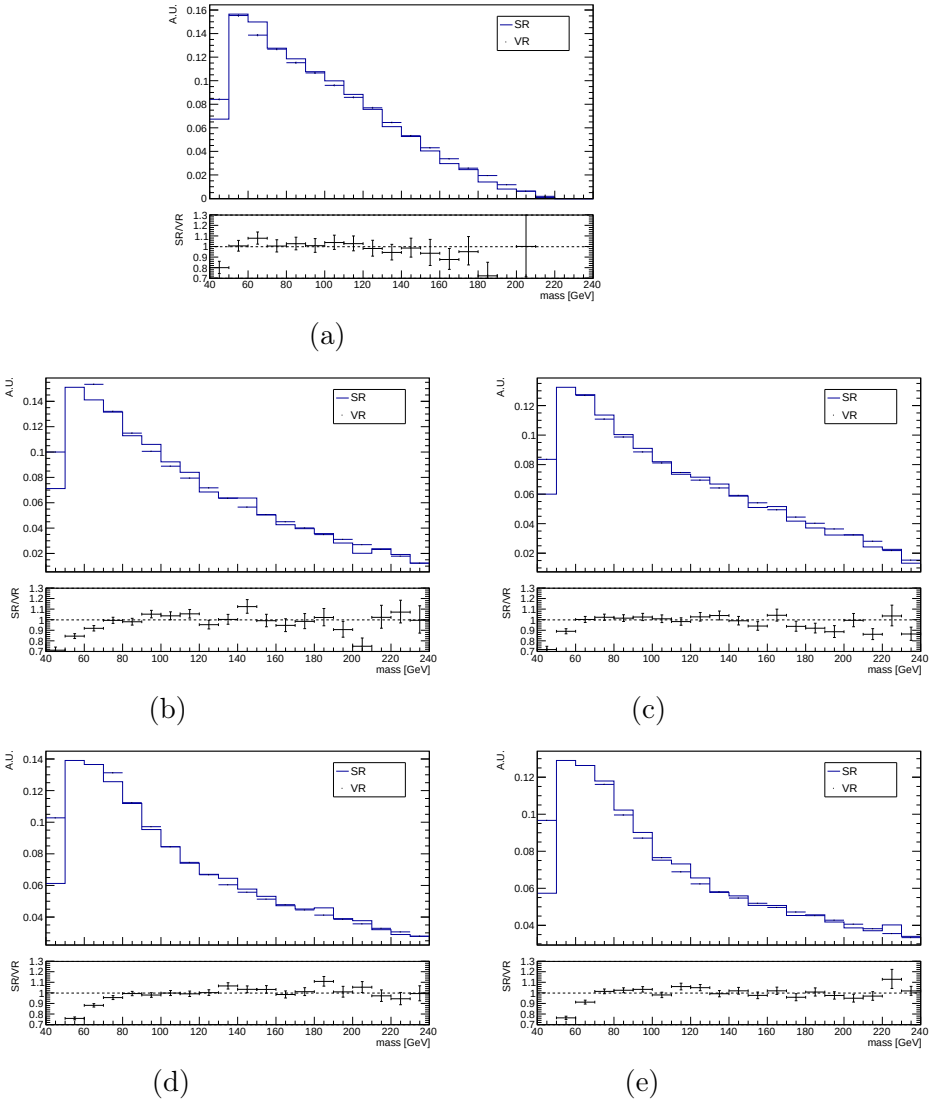


Figure 5.15: Comparison of signal and validation region for leading (a and c) and subleading (b and d) large-R jets. Monte Carlo samples are used for mass spectra. Figure (a),(b) and (d) show respectively the first, second and third p_T bin of the subleading SR. Figure (c) and (e) show respectively the first and second p_T bin of the leading SR. The low-mass cut has been removed for these plots to show the SR/VR discrepancy at low mass values. The area of distributions between 70 and 210 GeV is normalized to 1. The error bars only represent statistical error.

5.5 Background modeling

Since the dominant background of this analysis is still given by QCD multi-jet events, the strategy for the background modeling and estimation remains unchanged with respect to the version of the analysis: QCD background modeling will be data-driven and the resonant background modeling will be done with templates derived from the MC samples. Therefore, QCD background will be modeled with a dedicated validation region constructed with data not containing signal events, but optimized to be equivalent from the kinematic point of view.

5.5.1 QCD modeling

The QCD background is modeled with a parametric fit on the data validation region. The functional form chosen for this modeling is an exponentially falling function with N parameters:

$$f(x,\theta) = \theta_0\, exp\Big(\sum_{i=1}^{N} \theta_i x^i\Big)$$

where $x = (m_J - 140GeV)/70$ GeV and θ_i are the parameters of the fit. The background modeling validation strategy follows what already explained at page 76.

Determination of the optimal degree of QCD-modeling functions

The optimal number of parameters in each signal region is the result of the combination of the LLR test and of the spurious signal test results for the Higgs signal.

In general, the number of parameters needed for a functional form to be able to describe a given dataset depends on the statistics available, i.e. on the integrated luminosity [105]. Therefore, they need to be estimated in each region. With too few parameters, the function does not properly model data, while too many parameters cause the function to follow resonances and fluctuations. As the VR has many more events compared to the SR, it can be sliced in different subsets containing as many events as SRs. Table 5.8 contains the number of data events in each region and the number of VR slices created in each analysis jet p_T bin.
Therefore, the number of parameters needed in each signal region is determined by performing the log-likelihood ratio (LLR) test on all the VR slices. As described in detail at page 78, this test compares the log-likelihood values of two nested function models, looking at the p-value of their ratio. If two models, the first with n and the second with $n+1$ parameters, are proved to be equally accurate in the modeling of data, the one with the lowest number of parameters is favored. To help visualize the fit quality difference depending

	Inclusive		250-450 GeV		450-650 GeV		650-1000	
Leading	VR:	51149504	–		VR:	40712891	VR:	6239597
	SR:	700654	–		SR:	551808	SR:	92242
	Slices:	73	–		Slices:	74	Slices:	68
Subleading	VR:	32853230	VR:	9272185	VR:	20472758	VR:	2902422
	SR:	524426	SR:	153544	SR:	313929	SR:	52322
	Slices:	63	Slices:	60	Slices:	65	Slices:	55

Table 5.8: Summary of VR slices.

on the number of parameters used, the residuals are shown in fig. 5.16 and in fig.5.17. These clearly show that the fit is biased when a too low number of parameters is used. Also, the fit quality is not improved when using more parameters than the adequate one.

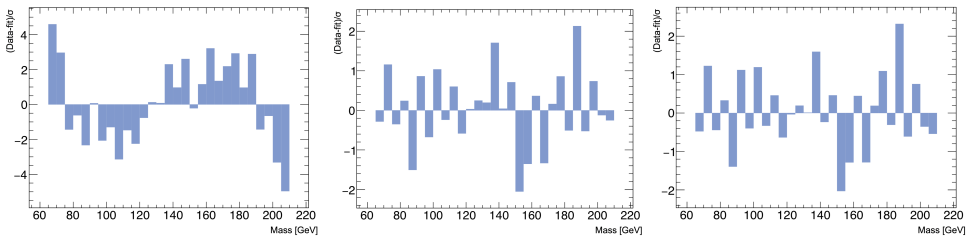


Figure 5.16: Residuals from the QCD fit with 2, 3, and 4 parameters (left to right) to a random slice of the inclusive VRL data are shown. Too low a degree of the polynomial fit clearly introduce biases in the background estimation.

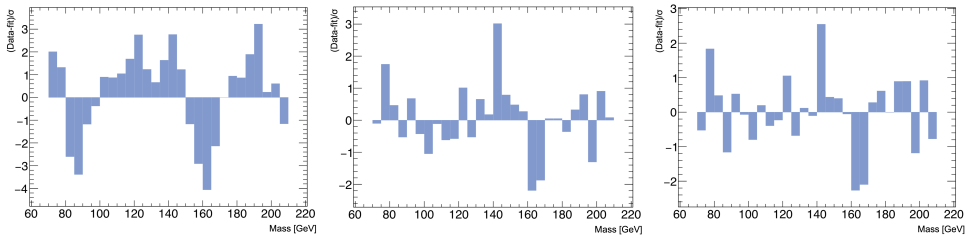


Figure 5.17: Residuals from the QCD fit with 2, 3, and 4 parameters (left to right) to a random slice of the inclusive VRS data are shown. Too low a degree of the polynomial fit clearly introduce biases in the background estimation.

To make the results on the choice of number of parameters robust against statistical fluctuations, the LLR test have been performed on the ensemble of VR slices. The smallest number of parameters giving a flat probability distribution of the LLR tests has been used for each jet category and p_T bin. As figure of merit we choose the cumulative distribution function of the LLR p-

value, realized over all the VR slices. In this way, a flat probability distribution is mapped into a linear form and that is what it is expected when we compare two models with equivalent power to describe data. Figure 5.18 shows the tests for the leading inclusive and subleading inclusive validation region slices, the other regions are shown in the appendix C. In the leading inclusive VR, the line 3 vs 4 shows an acceptable linear behaviour, therefore 3 parameters are enough to describe this region according to the LLR test. The same can be said for the line representing 4 vs 5 for the LLR test on the subleading inclusive VR, therefore 4 parameters are enough to describe this region according to the LLR test. The same criteria is applied to extract the number of parameters for the p_T binned validation regions.

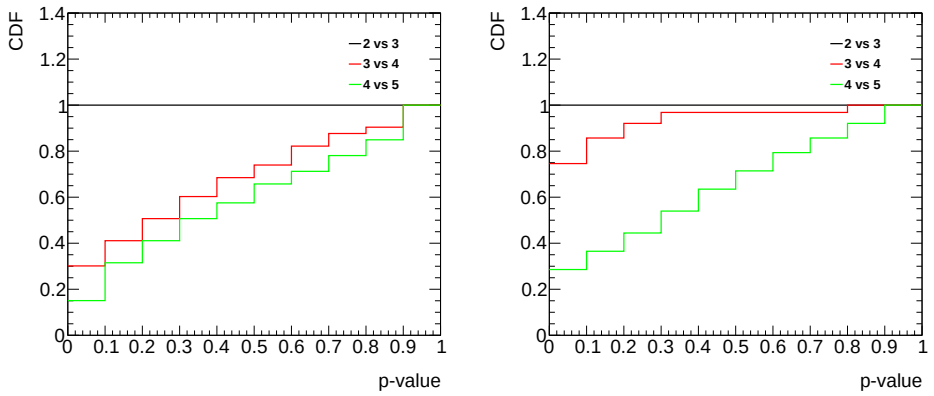


Figure 5.18: Cumulative distribution of the probability for the difference of the log-likelihood obtained for exponential polynomial functions of order n and n + 1. The results from the VRL and VRS are shown on the left-hand side and on the right-hand side respectively.

The log-likelihood ratio tests gives as an indication about the goodness of the global description of the data. However, it does not exclude that our model has some localized biases, i.e. close to the Higgs peak. This can be investigated performing the spurious signal tests in the Higgs mass window.

Spurious signal tests

In order to assess that our functional form does not create an artificial signal in mass the region window of the Higgs, a spurious signal test is performed on the validation region data slices. The contributions of V+jets, Higgs and $t\bar{t}$ are subtracted using MC templates computed for the validation region at the SM-predicted rates. Then, the subtracted data slices are fitted with the QCD function and the SR Higgs templates leaving free the QCD normalization, the polynomial parameters and the signal rate. The QCD shape and normalization and signal strength of the Higgs boson μ_H are extracted with a fit. The presence of spurious signal is evaluated by looking at the occurrence of fitted

signal strength values in excess of 2σ with respect to 0, as shown in fig. 5.19, where σ is the fit statistical uncertainty.

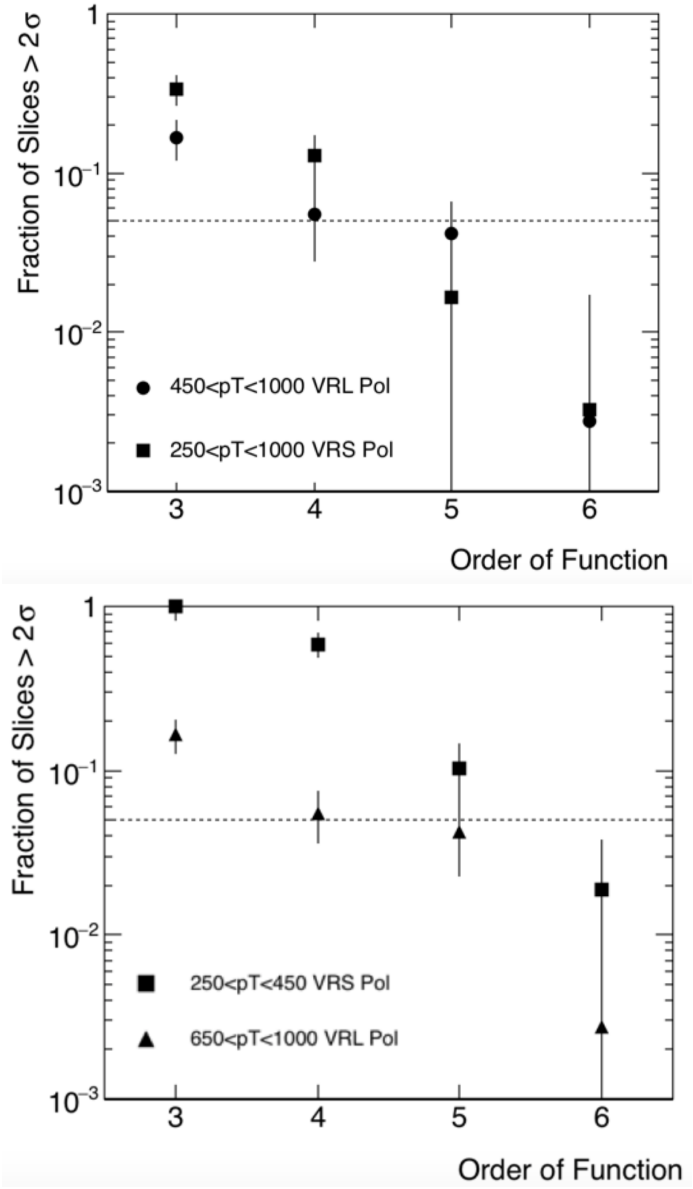


Figure 5.19: On the top the fraction of fitted signal values in excess of 2σ for all the VR slices of the leading and on the bottom the same for the subleading inclusive validation regions in the two most extreme p_T bins. The dashed line at 0.05 represents the 2 σ threshold for the fraction of outliers.

From fig. 5.19 it is clear that with a too low number of parameters in the function we have some not negligible signal artificially created by the background modeling.

Final Parametric choice

The optimal number of parameters in each signal region is the result of the combination of the LLR test and of the spurious signal test results for the Higgs signal. All these results are balanced with the fact that the statistical uncertainty on the signal increases with the number of parameters, as shown in 5.20. The final parametric choice is summarized in tab. 5.9 and it is the result of a compromise between minimizing the statistical uncertainty and the spurious signal excess. The optimal number of parameters is the one which makes the fraction of outliers compatible with 5% and it is has been studied in this way for each region. Then the absolute value of μ/σ_μ is computed for each region. This excess will be treated as an additional systematic on the μ/σ_μ of the Higgs signal and therefore the final error on will be inflated by this amount. A table of the excess of spurious signals, if any, in all VRs is shown in tab. 5.10.

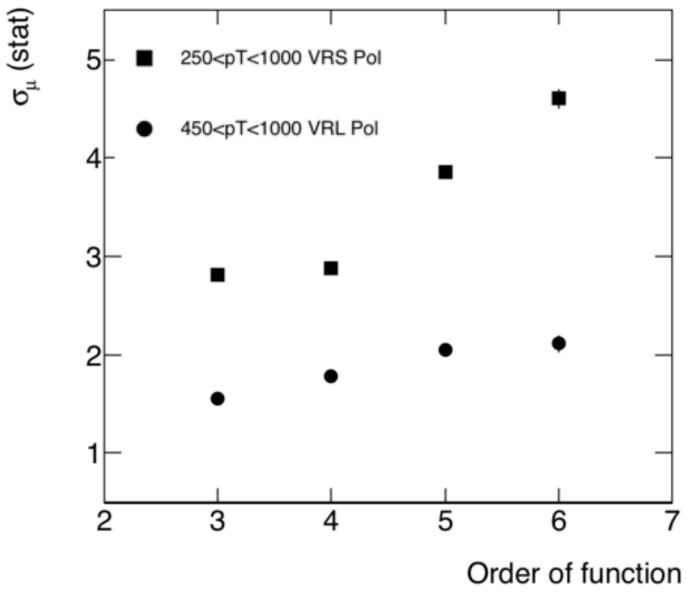


Figure 5.20: The behavior of the statistical uncertainty on μ_H determined by the fit increasing the number of parameters is shown for the SRL and SRS.

Region	Inclusive	250-450 GeV	450-650 GeV	650-1000 GeV
SR Leading	4	-	4	5
SR Subleading	5	5	4	4

Table 5.9: The number of parameters used in the QCD fit for each region determined from the likelihood ratio test and the spurious signal test results.

Example fits of a VR slice for each region are shown in figure 5.21 5.22 and 5.23. Fits are performed with an exponential polynomial function with the

Region	Inclusive	250-450 GeV	450-650 GeV	650-1000 GeV
VR Leading	0.14	-	0.30	0.10
VR Subleading	0.35	0.30	0.10	0.25

Table 5.10: The spurious signal excess evaluated in each in the validation regions, resulting from the fit with the final number of parameters .

number parameters previously determined.

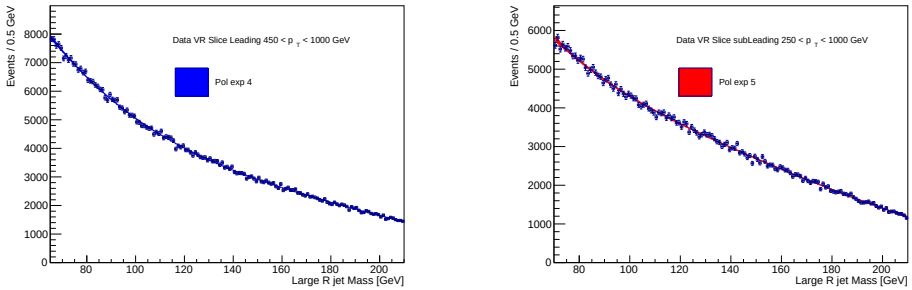


Figure 5.21: Leading large-R and subleading large-R jet invariant mass distribution in a data validation region slice. The shaded areas show the 90% C.L. around the fit on the VR slice. Exp 4 and exp 5 indicate the degree of the polynomial function at the exponent used in the fit.

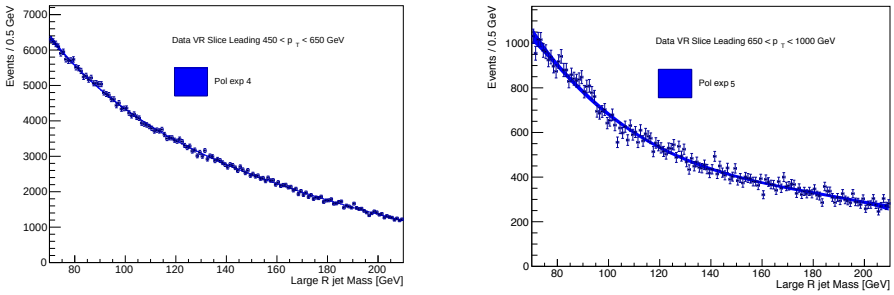


Figure 5.22: Leading large-R jet invariant mass distribution in a data validation region slice with $450 < p_T < 650$ GeV (left panel) and $650 < p_T < 1000$ GeV (right panel). The shaded areas show the 90% C.L. around the fit on the VR slice. Exp 4 and exp 5 indicate the degree of the polynomial function at the exponent used in the fit.

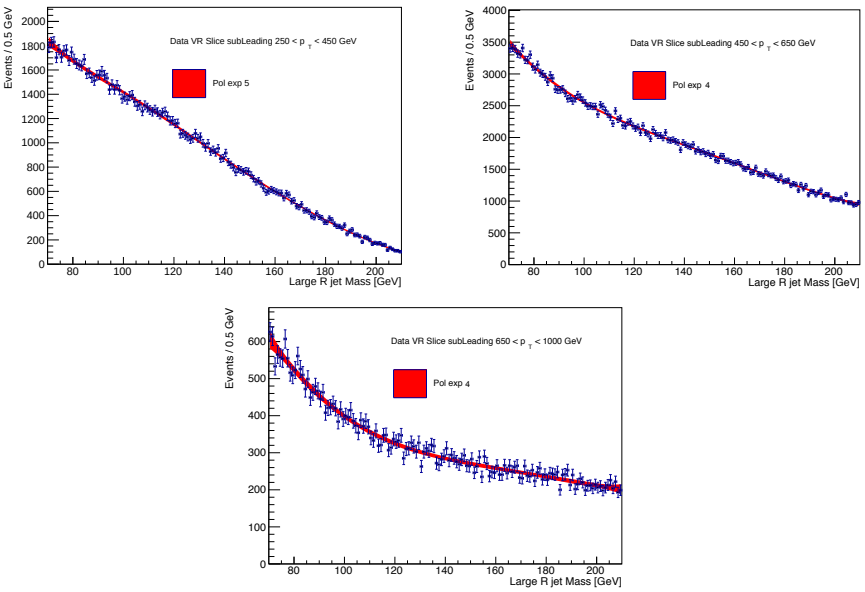


Figure 5.23: Subleading large-R jet invariant mass distribution in a data validation region slice for $250 < p_T < 450$ GeV (top left panel), $450 < p_T < 650$ GeV (top right panel) and $650 < p_T < 1000$ GeV (bottom panel). The shaded area shows the 90% C.L. around the fit on the VR slice. Exp 4 and exp 5 indicate the degree of the polynomial function at the exponent used in the fit.

5.5.2 V+jet modeling

The modeling of the V+jet (Z+jet and W+jet) background plays an important role in this analysis because of its closeness to the Higgs peak and its right tail extends in the Higgs mass window with a consistent contribution. Figure 5.24 shows the relative contribution of W+jets and Z+jets in the inclusive signal region and validation regions. In the signal region the dominant contribution to the V+jet background is given by the Z boson, since it can decay in a b-pairs. In the validation region the W boson is the main one, given that only 0 b-tag events are allowed to enter it.

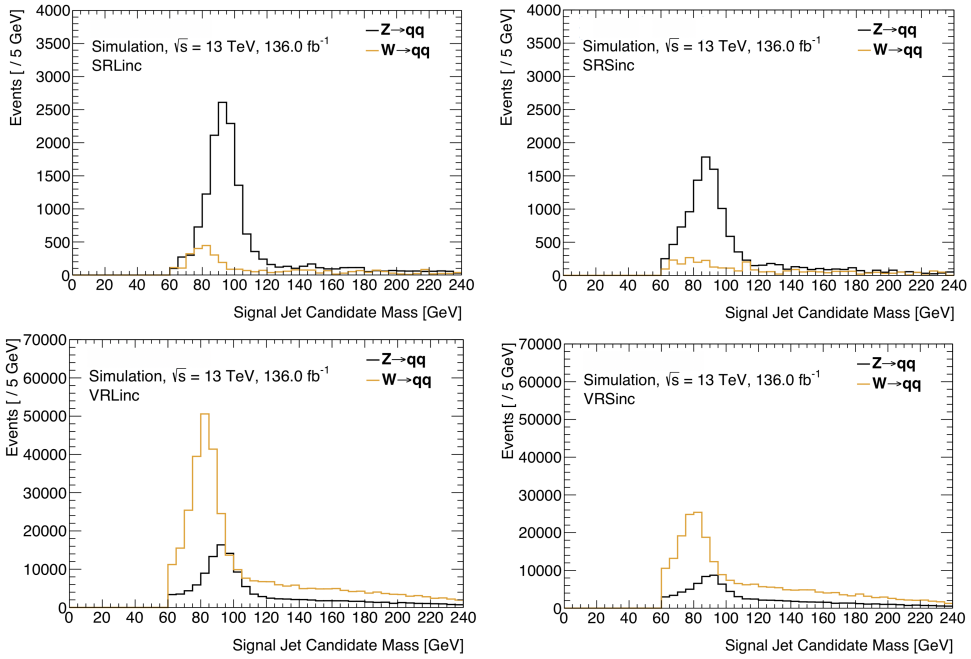


Figure 5.24: The V+jets contribution to the signal jet candidate mass (W+jets and Z+jets) in the inclusive leading (left) and subleading (right) signal (top) and validation (bottom) regions is shown.

5.5.3 $t\bar{t}$ modeling and $CR_{t\bar{t}}$

Since t-quarks decay into b-quarks, boosted $t\bar{t}$ events represent a non-negligible background contribution in the signal region. Figure 5.25 shows the breakdown of $t\bar{t}$ contributions in the inclusive signal regions and validation regions. In the signal region the top peak is well resolved while in the validation region there are two clear peaks, depending on whether only the W-jet or the whole top-decay chain are contained in the large-R jet candidate.
Although the MC prediction of the $t\bar{t}$ cross section is generally quite accurate [106], this analysis is sensitive to high-p_T effects, potentially creating discrepancies between MC and Data distribution. For this reason, as in page 72,

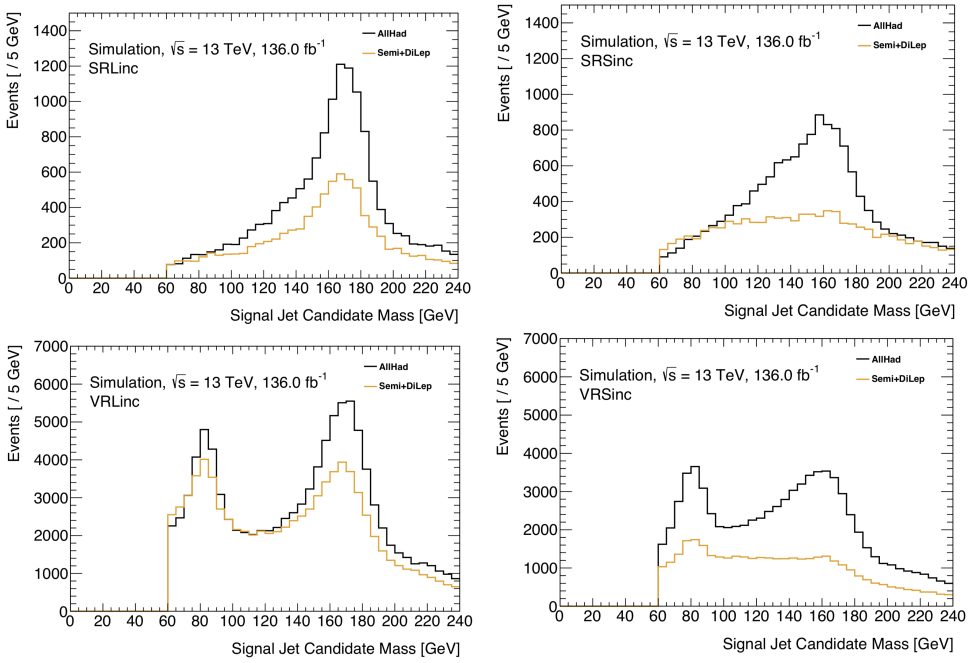


Figure 5.25: $t\bar{t}$ contributions to the signal jet candidate mass spectrum, divided into hadronic and leptonic. In the inclusive leading (left) and subleading (right) signal (top) and validation (bottom) regions are shown.

a control region ($CR_{t\bar{t}}$) enriched in semi-leptonic decays is used to help control the $t\bar{t}$ contributions. $CR_{t\bar{t}}$ is defined as having a boosted leptonically decaying top quark in one hemisphere (tag jet) and a boosted hadronically decaying top quark in the other (probe jet). To enter this region, events are required to fulfill these requirements:
Tag jet:

- isolated muon with $p_T > 40$ GeV ;
- $\Delta(\mu, \text{jet}) < 1.5$;
- leading VR track jet in tag-jet is b-tagged .

Probe Jet

- exactly 1 b-tagged track jet ;
- $\Delta\phi$ (μ, probe-jet) $> 2\pi/3$;
- $\Delta\phi$ (tag - jet, probe - jet) $> 2\pi/3$.

The ΔR cut reduces V+ jets and VV contamination and $\Delta\phi$ cut reduces QCD and V+jet contamination. The resulting $CR_{t\bar{t}}$ is binned in $p_T^{probe-jet}$, following the same binning used in the signal region. Finally, the shape of the mass spectrum is taken from MC, while the normalization is extracted from data. Fig.

5.26 reports Data/MC comparison in the control region, as outlined before. The normalization discrepancy is expected and the corresponding Data/MC scale factor (shown in tab 5.11) agree with ATLAS measurements reported in literature [107].

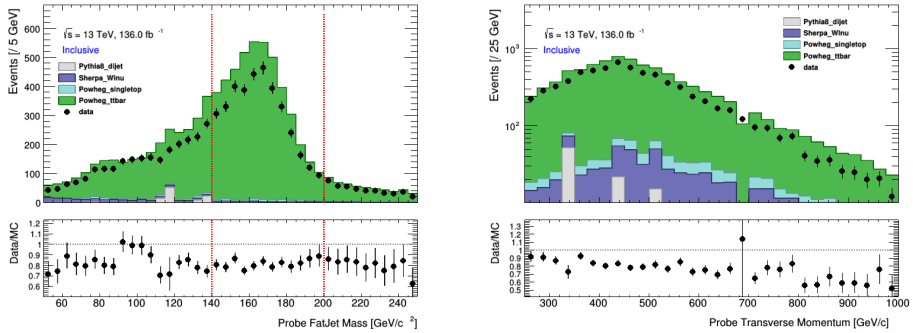


Figure 5.26: Data/MC of the inclusive $CR_{t\bar{t}}$ mass (left) and p_T distributions (right) are shown

Range [GeV]	Scale Factor	Stat Error
$[250, 450]$	0.86	0.02
$[450, 650]$	0.77	0.03
$[650, 1000]$	0.77	0.05

Table 5.11: The scale factors for the $t\bar{t}$ process is determined in the $CR_{t\bar{t}}$ in the 140 to 200 GeV mass range. The uncertainties are statistical only.

5.6 Fitting Strategy

This section describes the fitting strategy adopted in the analysis and the statistical framework used to derive results for inclusive and differential measurements.

5.6.1 Framework

The fitting framework used for this analysis is based on a tool called XML Analytic Workspace Builder (xmlAnaWSBuilder) [108]. A workspace is a container for the model and dataset of the analysis, accompanied by a standardized interface and its structure.
XmlAnaWSBuilder creates RooFit workspaces using one-dimensional observables and its workflow is summarized in fig. 5.27.

It is widely used in the ATLAS Higgs analysis because it provides interesting features, as the following ones:

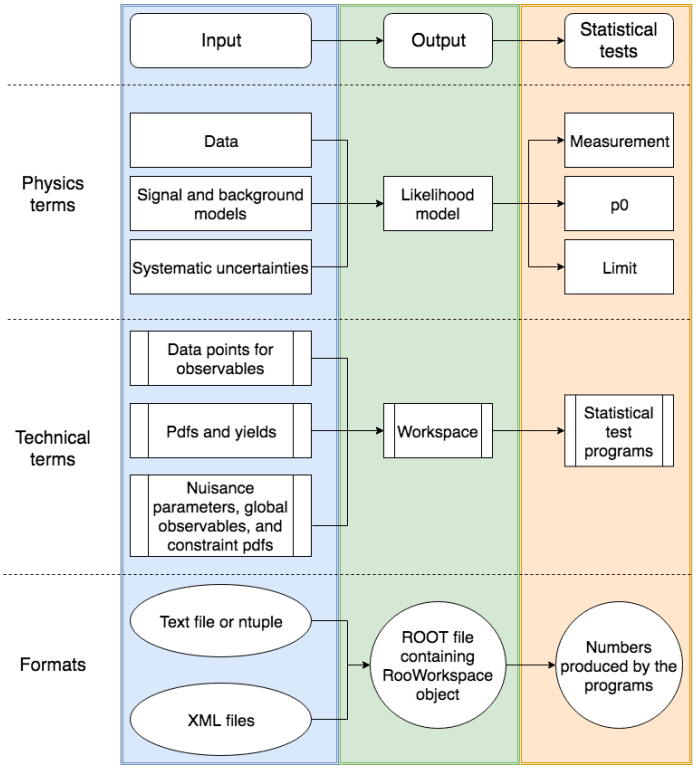


Figure 5.27: The xmlAnaWSBuilder workflow is shown in the figure [108].

- it is based on RooFit/RooStats [109] and it is therefore specialized in constructing analytic models and counting experiments;
- it can access external tools (like HistFactory [110]), it can also deal with histograms;
- it has a user-friendly interface, which allows easily to create a variety of likelihood models using simple syntax saved in XML cards. With this structure one may fit simultaneously multiple signal/control regions and get a combined result;
- the saved XML cards ensure reproducibility of the results, and allow easy adjustment of the model;
- it has a simple fitting functionality already integrated, which is activated at the same time of the workspace construction;
- the software is light-weighted and only depends on ROOT [111].

After the workspace creation, the fit itself is run with a separate tool, widely used in the ATLAS collaboration, called QuickFit [112]. The fit is based on a negative log-likelihood minimization and offers full compatibility with xmlAnaWSBuilder and good stability in complex hyperparameter spaces.

5.6.2 Statistical Model

The statistical model contains 8 regions, which we can list as follows:

- SRL[1-2] - The two p_T bins of the leading jet;
- SRS[0-2] - The three p_T bins of the sub-leading jet;
- TTbar[0-2] - The three p_T bins of the $CR_{t\bar{t}}$,

It includes also 4 floating parameters representing three normalizations modifiers (μ), normalized to SM values, and a QCD-background normalization fit to data:

- μ_H,
- μ_V,
- $\mu_{t\bar{t}}$,
- $\text{yield}_{\text{QCD}}$,

and several Gaussian or Log-Normal constrained nuisance parameters. For the inclusive measurement, the SRs are realized adding all the p_T bins contributions in one, while for the differential all jet p_T bins are fit once separately, to know the contribution in each bin, and once simultaneously for the final result. Among the normalization modifiers, μ_H and μ_V are defined as parameters of

interest (POIs) of the fit. MC statistical uncertainties are considered for bins with a statistical error $> 5\%$ with a jet mass distribution in 0.5 GeV/ bins. In order to account for statistical uncertainty for all the aforementioned bins, an independent process-by-process bin-by-bin nuisance parameter is defined with Gamma-function prior probability density functions and σ prior equal to the bin relative statistical uncertainty.
For the inclusive and differential measurement we fit at the same time one (or more)SR (SRs) and the corresponding $CR_{t\bar{t}}$ ($CR_{t\bar{t}}$s):

- Inclusive: SRL+SRS + $CR_{t\bar{t}}$
- Differential:
  1. first p_T bin: SRS0 + $CR_{t\bar{t}}0$
  2. second p_T bin: SRS1 + SRL1 + $CR_{t\bar{t}}1$
  3. third p_T bin: SRS2 + SRL2 + $CR_{t\bar{t}}2$

5.6.3 Signal injection test

Once the adequacy and robustness of the QCD parametric shape is assessed, with the of LLR test and the spurious signal test on the VR slices, another important check is the signal injection test. This test proves that the chosen parametric shape is smooth enough to allow the extraction of the same amount of signal that is injected.
As already explained in chapter 4, an Asimov dataset [113] is built summing up the H+jet, V+jet and $t\bar{t}$ templates and the QCD parametric function with post-fit parameters, obtained from a fit to the VRs, after the subtraction of resonances. After that, pseudo-experiments (PE) were generated from the Asimov data, sampling each bin of the histogram from a Poisson distribution with $\mu = bin_i^{Asimov}$ and $\sigma = \sqrt{bin_i^{Asimov}}$, where bin_i^{Asimov} are the bin contents of the Asimov dataset. Each of these pseudo-experiments is then fitted and the relative difference between extracted and injected signal strength is computed:

$$pull = \frac{\mu_{fit} - \mu_{injected}}{\sigma_{fit}}.$$

A grid $(0, 1, 2)\times(0, 1, 2)$ for (μ_H, μ_V) was explored using these pseudo-experiments. This test is repeated on 400 independent pseudo-experiments. When the gradient minimizer does not converge during the fit, the trial is discarded. A Normal distribution is fit to the distribution of pulls. As far as the fit result is unbiased, then one should expect the pulls to be normally distributed with a mean of 0 and width of 1,

$$\text{pulls} \sim \mathcal{N}\left(\mu = 0, \sigma = 1\right).$$

The deviation of the fit Normal distribution's mean from 0 is then an indicator of the bias of the pull given the choice of QCD model. The number of trials

was chosen to be 400 to estimate the mean distribution at 5% level

$$\sigma_{\hat{\mu}} = \left.\frac{\sigma}{\sqrt{\text{N trials}}}\right|_{\sigma=1,\text{N trials}=400} = 0.05.$$

Results of the signal injection tests are shown in figures 5.28-5.31 and in tables 5.12-5.15. The full set of results in different p_T bins is shown in appendix B. The fitting model and the framework used performed well in injection tests and therefore no systematic uncertainties associated to them have been included in the analysis.

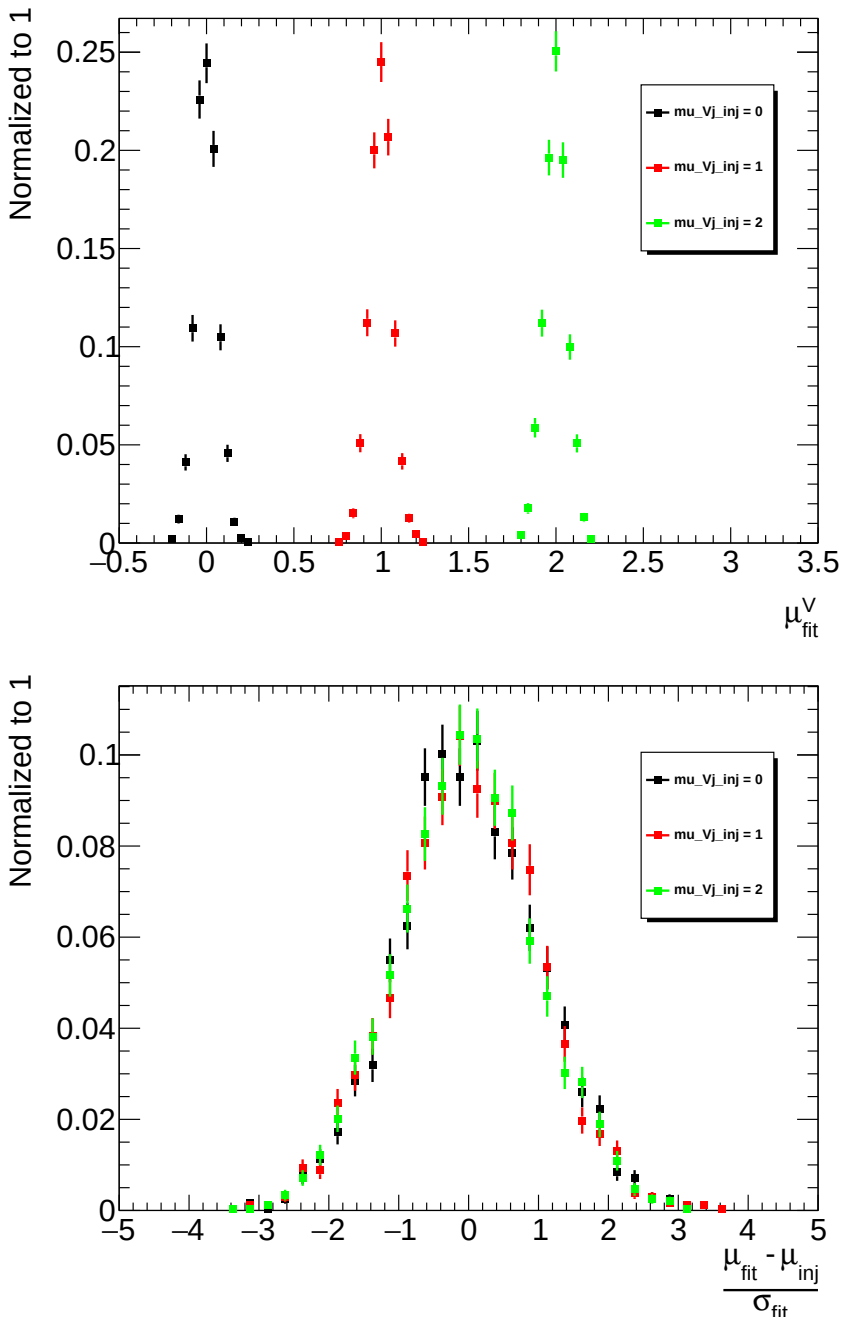


Figure 5.28: The extracted signal strength distributions for a given injected signal strength in the inclusive SRL are shown for V+jet. μ_V distributions are on the top and pull distributions on the bottom. The pulls are centered at 0 with a width of 1.

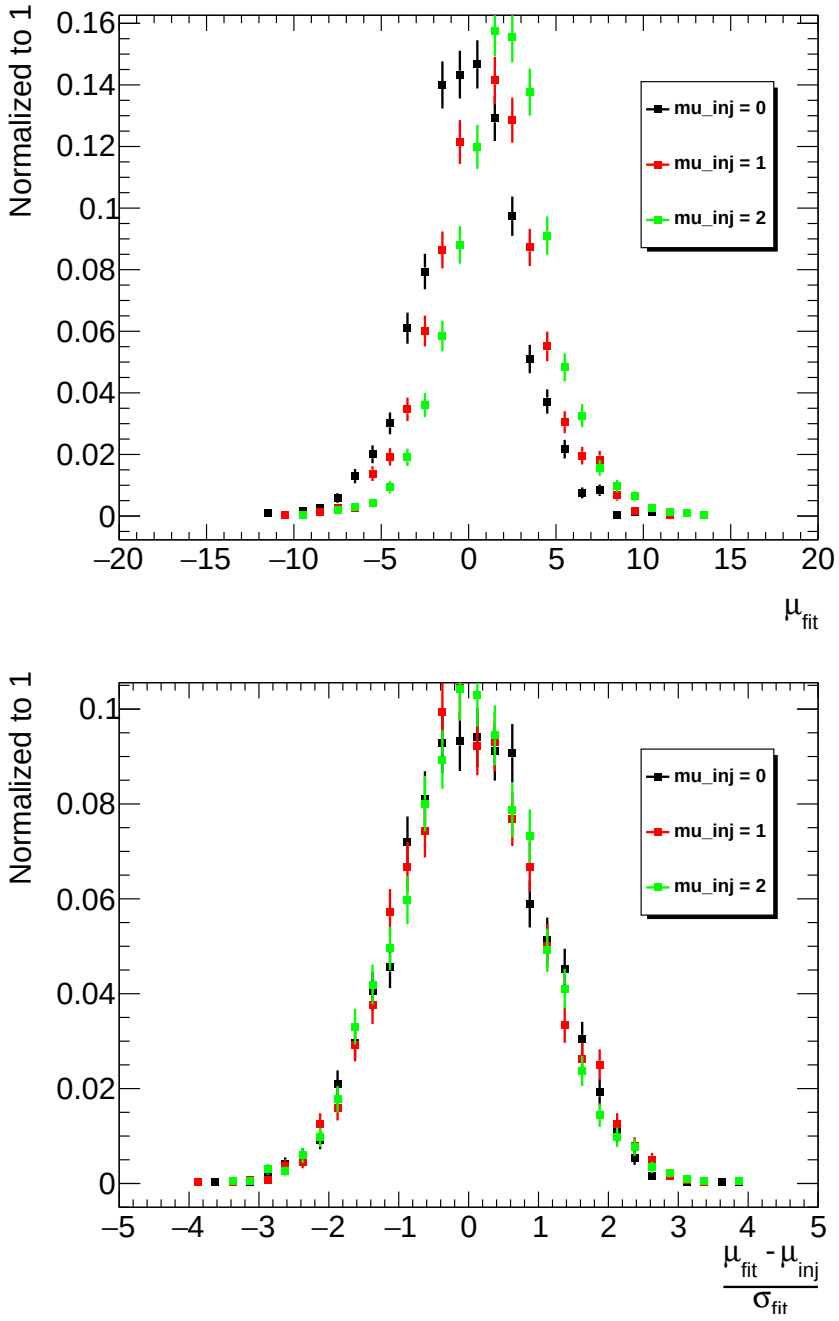


Figure 5.29: The extracted signal strength distributions for a given injected signal strength in the inclusive SRL are shown for the Higgs. μ distributions are on the top and pull distributions on the bottom. The pulls are centered at 0 with a width of 1.

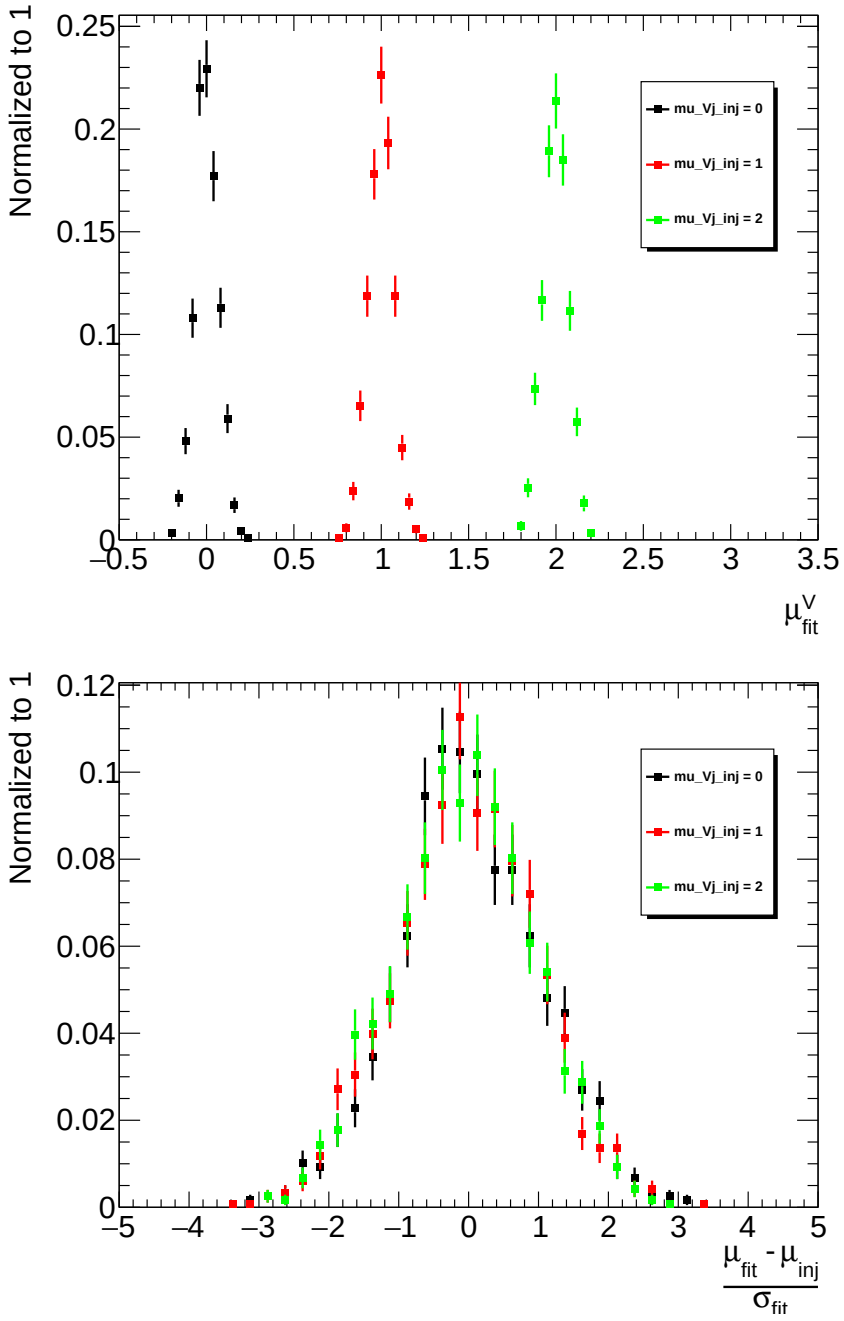


Figure 5.30: The extracted signal strength distributions for a given injected signal strength in the inclusive SRS are shown for V+jet. μ_V distributions are on the top and pull distributions on the bottom. The pulls are centered at 0 with a width of 1.

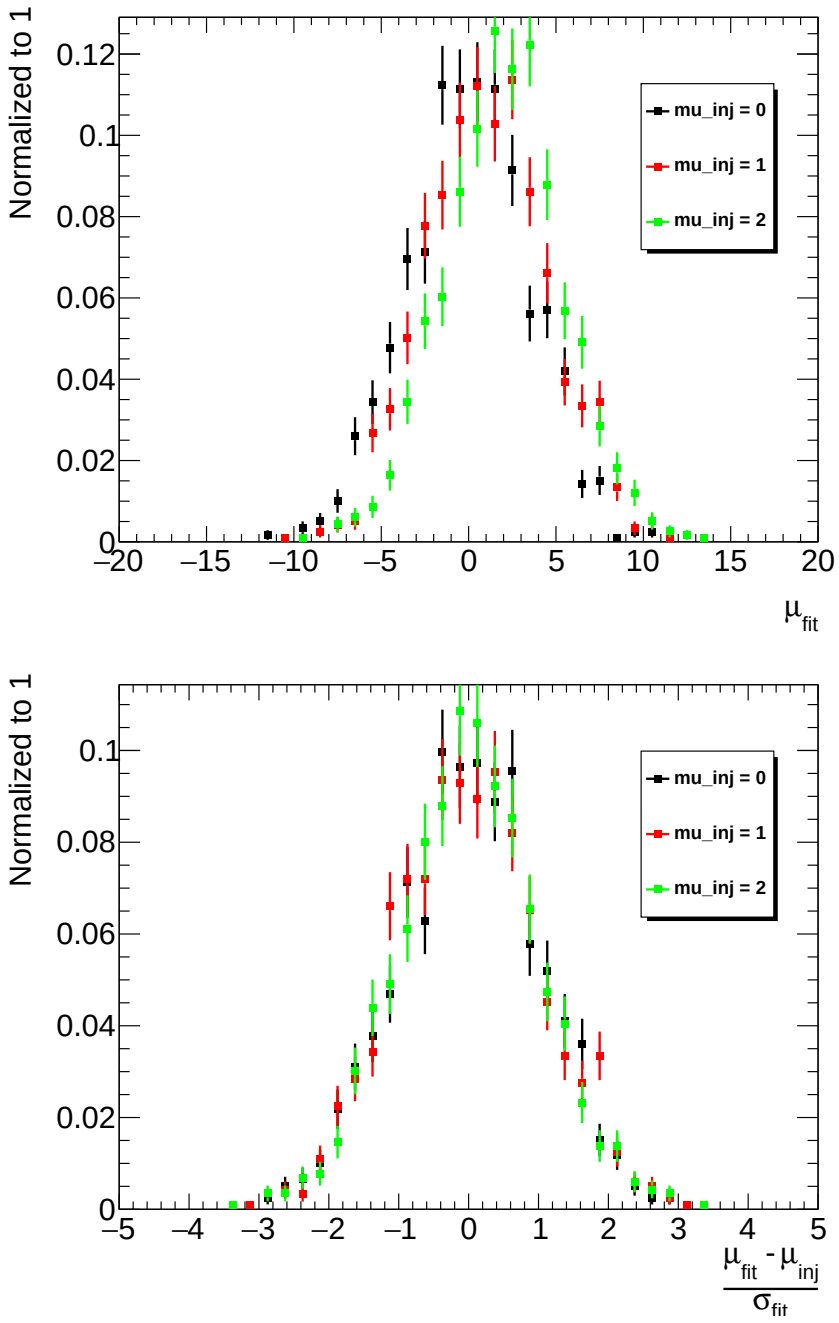


Figure 5.31: The extracted signal strength distributions for a given injected signal strength in the inclusive SRS are shown for the Higgs. μ distributions are on the top and pull distributions on the bottom. The pulls are centered at 0 with a width of 1.

Leading inclusive SR

μ_H^{inj}	μ_V^{inj}	μ_H^{fit}	$\sigma_{\mu_H}^{fit}$	μ_V^{fit}	$\sigma_{\mu_V}^{fit}$
0	0	0.006 ± 0.112	2.087 ± 0.090	0.002 ± 0.003	0.054 ± 0.002
0	1	-0.043 ± 0.114	2.123 ± 0.089	1.001 ± 0.003	0.061 ± 0.002
0	2	-0.015 ± 0.108	2.000 ± 0.088	1.999 ± 0.003	0.056 ± 0.002
1	0	1.054 ± 0.102	1.935 ± 0.072	0.000 ± 0.003	0.057 ± 0.002
1	1	0.906 ± 0.102	1.949 ± 0.092	1.002 ± 0.003	0.059 ± 0.002
1	2	0.964 ± 0.111	2.114 ± 0.089	1.998 ± 0.003	0.062 ± 0.002
2	0	1.975 ± 0.099	1.902 ± 0.073	-0.004 ± 0.003	0.057 ± 0.002
2	1	2.012 ± 0.105	2.053 ± 0.077	0.994 ± 0.003	0.057 ± 0.002
2	2	1.911 ± 0.111	2.058 ± 0.097	1.996 ± 0.003	0.062 ± 0.003

Table 5.12: μ and σ values are extracted with a gaussian fit to the $\mu_{H/V}$ distribution from the SRL.

Leading inclusive SR

μ_H^{inj}	μ_V^{inj}	$Pull_H^{fit}$	$\sigma(pull)_{\mu_H}^{fit}$	$Pull_V^{fit}$	$\sigma(pull)_{\mu_V}^{fit}$
0	0	-0.022 ± 0.053	1.014 ± 0.040	-0.033 ± 0.052	0.998 ± 0.038
0	1	-0.002 ± 0.053	1.014 ± 0.037	0.043 ± 0.053	1.025 ± 0.038
0	2	0.009 ± 0.050	0.941 ± 0.040	0.012 ± 0.050	0.956 ± 0.043
1	0	0.040 ± 0.052	0.974 ± 0.037	-0.030 ± 0.051	0.985 ± 0.038
1	1	-0.055 ± 0.051	0.962 ± 0.048	0.015 ± 0.052	0.991 ± 0.038
1	2	-0.011 ± 0.054	1.010 ± 0.044	-0.035 ± 0.052	1.006 ± 0.040
2	0	-0.028 ± 0.050	0.944 ± 0.037	-0.088 ± 0.051	1.011 ± 0.041
2	1	-0.017 ± 0.052	0.972 ± 0.041	-0.093 ± 0.050	0.968 ± 0.038
2	2	-0.050 ± 0.054	1.012 ± 0.047	-0.059 ± 0.053	1.028 ± 0.042

Table 5.13: μ and σ values are extracted with a gaussian fit to the $pull_{H/V}$ distribution from the SRL.

Subleading inclusive SR

μ_H^{inj}	μ_V^{inj}	μ_H^{fit}	$\sigma_{\mu_H}^{fit}$	μ_V^{fit}	$\sigma_{\mu_V}^{fit}$
0	0	0.023 ± 0.179	3.387 ± 0.144	0.004 ± 0.004	0.070 ± 0.003
0	1	0.085 ± 0.190	3.494 ± 0.172	0.994 ± 0.004	0.073 ± 0.003
0	2	-0.096 ± 0.191	3.479 ± 0.142	1.995 ± 0.004	0.073 ± 0.003
1	0	1.463 ± 0.189	3.529 ± 0.151	0.003 ± 0.004	0.071 ± 0.003
1	1	0.802 ± 0.195	3.542 ± 0.162	0.993 ± 0.004	0.069 ± 0.003
1	2	0.781 ± 0.187	3.441 ± 0.147	1.997 ± 0.004	0.075 ± 0.003
2	0	2.030 ± 0.184	3.364 ± 0.154	-0.005 ± 0.004	0.069 ± 0.003
2	1	1.994 ± 0.179	3.320 ± 0.158	1.006 ± 0.004	0.071 ± 0.003
2	2	2.044 ± 0.179	3.370 ± 0.158	1.998 ± 0.004	0.075 ± 0.003

Table 5.14: μ and σ values are extracted with a gaussian fit to the $\mu_{H/V}$ distribution from the SRS.

Subleading inclusive SR

μ_H^{inj}	μ_V^{inj}	$Pull_H^{fit}$	$\sigma(pull)_{\mu_H}^{fit}$	$Pull_V^{fit}$	$\sigma(pull)_{\mu_V}^{fit}$
0	0	0.017 ± 0.055	1.039 ± 0.047	0.029 ± 0.051	0.983 ± 0.036
0	1	-0.014 ± 0.055	0.988 ± 0.045	-0.080 ± 0.052	0.999 ± 0.046
0	2	0.013 ± 0.054	1.031 ± 0.038	-0.109 ± 0.057	1.024 ± 0.046
1	0	0.117 ± 0.059	1.085 ± 0.045	0.076 ± 0.054	1.032 ± 0.045
1	1	-0.084 ± 0.053	1.016 ± 0.042	-0.081 ± 0.050	0.986 ± 0.040
1	2	0.006 ± 0.053	0.995 ± 0.037	-0.063 ± 0.052	1.015 ± 0.038
2	0	0.035 ± 0.052	0.990 ± 0.041	-0.081 ± 0.051	0.978 ± 0.042
2	1	0.026 ± 0.051	0.948 ± 0.040	0.057 ± 0.053	1.009 ± 0.038
2	2	-0.014 ± 0.052	0.990 ± 0.041	-0.021 ± 0.052	0.997 ± 0.038

Table 5.15: μ and σ values are extracted with a gaussian fit to the $pull_{H/V}$ distribution from the SRS.

5.7 Systematics

The determination of the Higgs signal strength is statistically limited in all configurations, therefore, systematics mostly play a second-order role in the analysis sensitivity. They can affect shape, normalization or both and are divided in three main blocks, affecting shape, normalization or both of them: theoretical systematics, flavor tagging systematics and jet systematics.

1. **Theoretical Systematics**. They are evaluated with a change in selection which can introduce statistical noise.

   - Generator uncertainty: it accounts for shape-only variation coming from the comparison of the nominal sample to an alternative one. For V+jet we used Herwig as alternative sample and for $t\bar{t}$ we use Sherpa as alternative sample;
   - PDF uncertainties: evaluated via internal weights;
   - Scale uncertainties in the matrix element and in parton showering;
   - Alternative fragmentation models: evaluated by comparing cluster vs Lund fragmentation models.

   The theoretical systematic associated to V+jet is named alpha _wMUR0p5 _MUF0p5 and it accounts for matrix element uncertainty and parton showering scale variation. For $t\bar{t}$ we have alpha_ttbar_PowHer, which accounts for fragmentation/hadronization uncertainty using the comparison for Herwig, and alpha_ttbar _aM-cAtNloPy8, which accounts for generator uncertainty.

2. **Flavor Tagging Systematics**: the uncertainty on the calibration of the b-tagging algorithm is calculated using the eigenvector scheme developed within the collaboration [69]. The flavor composition versus mass is non-uniform since we have ttH in SRs. For this reason the area below the Higgs peak is dominated by b-quarks while the area around the top mass is dominated by c- and light quarks. This can cause different scale factor variations with p_T and flavor dependence, which translates in a mass dependence. Nuisance parameters are introduced in the fit to take into account this dependence. They represent the impact on scale and shape of the mass distribution produced by variation of the b-jet efficiency and c-jet and light-jet mis-tag rate. In the fit model these scale factors variation for b-jet, c-jet and l-jet are respectively Eigen_B_*, Eigen_C_* and Eigen_Light_*.

3. **Jet Systematics**. For this category nuisance parameters are introduced [114–116] to take into account for:

   - the Jet Energy Scale;
   - the Jet Mass Scale;

- the Jet Mass Resolution;
- experimental error on jet observables;
- statistical limitation on the determination of jet observables;
- peculiar topology of top-jet and V-jet;
- difference in the calorimeter response to quark or gluon;
- quark-gluon fraction.

In the fit model they are introduced including the nuisance parameters JET_* .

A pruning procedure is applied to all systematics. It acts on those producing a normalization change of the nominal sample of $< 3\%$ for a given background process, restricting them to be shape-only.

5.8 Expected results

The expected SR distributions are realized adding the MC templates of the V+jet, $t\bar{t}$ and H+jet to the QCD parametric form, with parameter fixed to the post-fit values extracted from the corresponding VR. With this procedure an Asimov dataset is built, which is the best representation of the SR and the expected sensitivity to the Higgs signal can be assessed. In fig. 5.32, the post-fit shapes and normalization for the inclusive regions are shown.
It is clear by looking at these shapes that this analysis has a large sensitivity to the V peak. Therefore, in the next future a differential measurement in p_T bins for the V will be performed, in addition to the inclusive one. Moreover, since the V is largely dominated by the Z boson in our signal region, a pure Z measurement could be performed, fixing the W template to its SM rate. For what concerns the Higgs boson, its yield is well within the data statistical fluctuations, as suggested by the large data error bars after the background subtractions.
For the differential cross section measurement a signal strength per truth Higgs p_T bin is defined and the correlation between the Higgs p_T and the large-R jet p_T is modeled with the use of the Monte Carlo samples. This is well represented by the migration matrix in fig. 5.33, which show for the resolution effects of the detector.
The fit model for the differential measurement has free-floating normalization parameters correlated across regions corresponding to the same pT bin (SRLX + SRSX + $CR_{t\bar{t}}$X). Conversely, the background normalizations are not correlated across different pT bins. Systematics variations are correlated across the analysis regions.

5.8.1 Summary of expected results

In tab. 5.16 there is a summary of the results in all the SRs. These values are estimated with Asimov datasets built with the QCD shape resulting from a fit

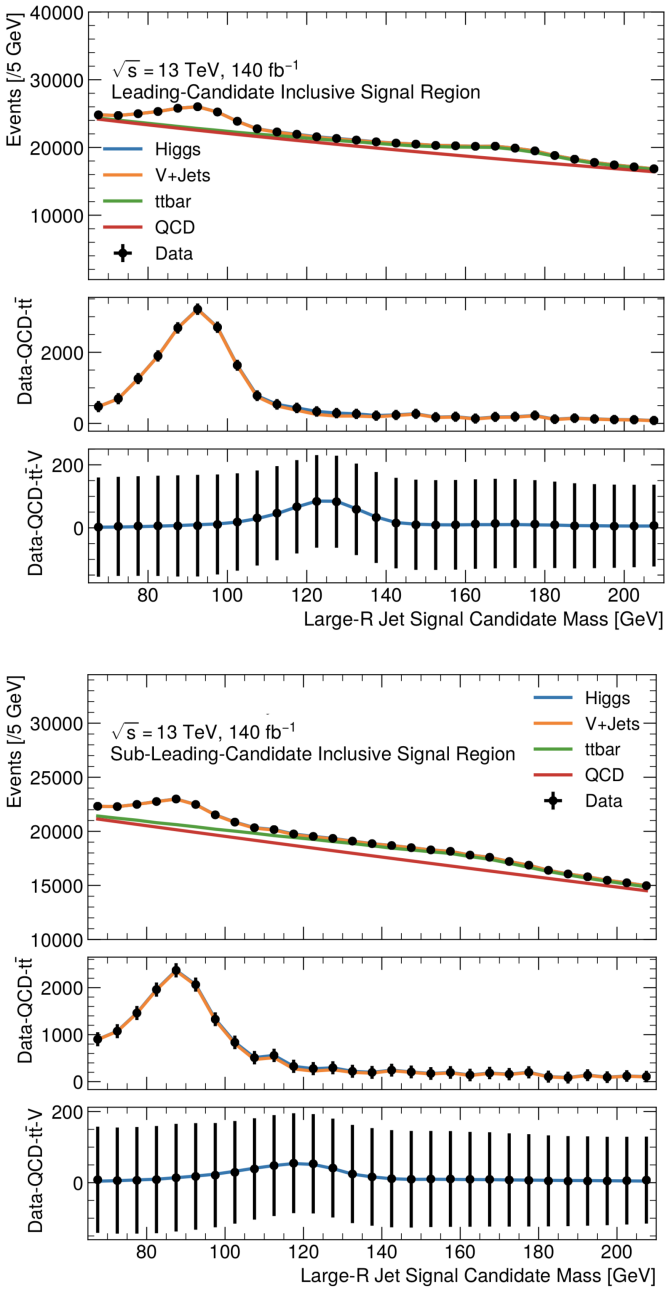


Figure 5.32: Post-fit stack plots for the SRL (left) and SRS (right) Asimov fit. In the top box all the background and signal contributions are shown, in the middle one the QCD function is subtracted and in the bottom one all the backgrounds are removed and only the Higgs signal is left.

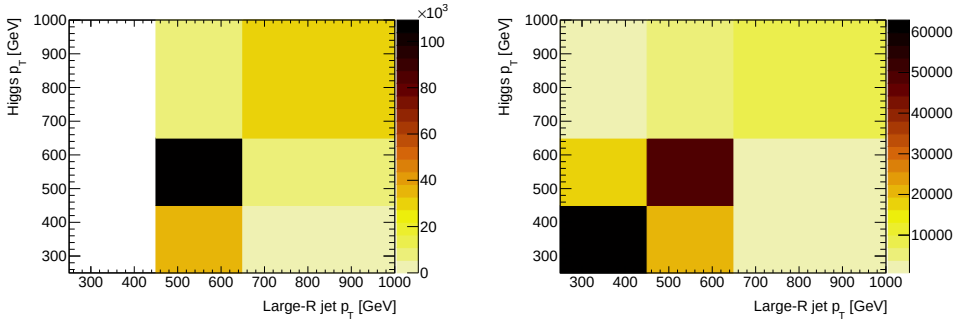


Figure 5.33: The migration matrix showing the generated p_T of the Higgs bosons as a function of that of the reconstructed large-R jets for the SR leading (left panel) and subleading (right panel) jet categories. On the z-axis the number of entries.

to a VR region slice and the other processes are included using templates fixed at SM rate.

jet p_T range [GeV]	Categories	$\mu_H/\sigma(\mu_H)$
250/450<p_T^{jet}<1000	SRS, SRL	0.57
250<p_T^{jet}< 450	SRS0,	0.14
450<p_T^{jet}< 650	SRS1, SRL1	0.54
650<p_T^{jet}<1000	SRS2, SRL2	0.30
250<p_T^{jet}< 450 450<p_T^{jet}< 650 650<p_T^{jet}<1000	SRS0, SRS1, SRL1 SRS2, SRL2	0.63

Table 5.16: Expected signal significance in various configurations: inclusive, in single p_T bins and combining all the p_T bins.

The first row of the table represents the inclusive expected significance and the following ones the expected values in the three p_T bins. In the second part of the table (rows 2-4) the expected significances are listed for each large-R jet p_T bin. The last row of the table is obtained by fitting the three p_T bins at the same time. The value is larger in this case than for the inclusive results because we use different templates with different p_T and mass resolution for the three different p_T bins.
The differential results can be translated in terms of Higgs p_T using the MC information. The result is shown in terms of μ_H/σ_H and expected asymptotic limits at 95% CL [117] in tables, 5.17 and 5.18.

With this analysis configuration we expect a sensitivity $\sim 0.6\ \sigma$ to the Higgs boson, mostly driven by the p_T bin right after the trigger threshold, 450<p_T<650. This analysis has been designed to measure the Higgs boson cross section at

Higgs p_T range [GeV]	$\mu_H/\sigma(\mu_H)$
$300<p_T^H<450$	0.14
$450<p_T^H<650$	0.45
$p_T^H>650$	0.26

Table 5.17: Expected signal significance in the three p_T^H bins.

p_T^H range [GeV]	limits on μ_H
$300<p_T^H<450$	$\mu_H^1<16$
$450<p_T^H<650$	$\mu_H^2<4.6$
$p_T^H>650$	$\mu_H^3<7.4$

Table 5.18: 95% CL limits in the three p_T^H bins.

high p_T and to test the existence of SMEFT which may largely enhance it. The main outcomes are therefore two: the possibility to see if there are large deviations from SM in the Higgs sector at boosted regime and the construction of a pioneering and solid strategy to measure the Higgs cross section, which can be re-run to as soon as we have more statistics (i.e. HL-LHC).

5.9 Results from data

In this section the results from data are presented. At the time of writing, the latest version of the analysis still undergoes the internal review process of the ATLAS collaboration. The review process is conducted by members of the collaboration who are not part of the analysis team. It is carried out on results based on data samples coming from Monte Carlo simulations and collisions belonging to Validation Regions. Such a procedure, often referred to as "blind", do not consider collisions from the signal region until the full analysis chain is validated against possible systematics. This approach protects science from biased results, avoiding procedures being tuned on data to be blindly analyzed.
Reviewers are currently evaluating the performance of the analysis on the Z boson in the signal region, to obtain a full estimation of the spurious signal induced at the Higgs mass. This thesis contains results up to there and does not report the measurement of the Higgs signal strength.

5.9.1 Tests on VR

One of the last tests to be performed before looking at the signal region, is to probe the full procedure of signal extraction in the validation region. Indeed, validation regions contain significant fractions of W, Z, and $t\bar{t}$ events and they constitute an ideal benchmark to assess the reliability of the signal extraction process.

The VR signal extraction test assumes the $t\bar{t}$ contribution to be well reproduced by Monte Carlo simulations, so that it can be subtracted from VR data to leave the test running on QCD and V+jet events only. The choice of subtracting $t\bar{t}$ is well motivated, as Monte Carlo simulations in use for this analysis were confirmed by previous ATLAS measurement [107].
The ratio of Z to W bosons in signal regions (75:25) is inverted in validation regions. Instead of introducing an *ad-hoc* unfolding procedure, the VR signal extraction test is performed on the inclusive V+jet signal, therefore taking the inclusive coupling strength μ_V as benchmark.
In order to avoid statistical biases, recalling that VRs contain much more events than SRs, the test is performed on the VR slices as defined at page 109. At the end of the procedure, μ_V results from all slices are averaged out.
Differently than signal regions, validation regions contain more W bosons than Z bosons (75:25 in VRs and 25:75 in SRs).
The test is performed independently on the VRL and VRS slices. For the VRL the resulted $\mu_V^{VRL} = 1.06 \pm 0.08$ and for the VRS is $\mu_V^{VRS} = 0.89 \pm 0.12$, which are both compatible with 1 and therefore with SM. Post-fit distributions and pulls from two representative slices are shown respectively in fig. 5.34, fig. 5.37. In the post-fit plots the W peak is pronounced around 81 GeV, since we have a good sensitivity to it and pulls are well-behaved.

From fig. 5.35 and 5.35, a high degree of correlation among QCD parameters and the QCD yield is seen, as expected. There is also an important correlation between QCD and the H and V+jet normalizations since they sit one on top each other.
μ_V/σ_V and μ_H/σ_H were tested as well, the results are shown in fig. 5.38. This was done to check the behavior of these quantities as a function of the number of parameters used in the fit. No systematic effects are induced from the fit model and the results of μ_H/σ_H are compatible with those from the QCD modeling studies previously shown.

5.9.2 Z boson signal extraction

In order to prove the stability and reliability of the fit model, the extraction of the Z boson signal strength has been performed. The full fit model was used and all three inclusive regions, SRL + SRS and $CR_{t\bar{t}}$, were fitted at the same time. The W boson was subtracted at its SM rate with a 10% uncertainty on the subtraction [118]. This uncertainty has been introduced in the fit as a Gaussian prior with sigma of 0.1. The Standard Model Higgs boson signal has been subtracted.
The post-fit distributions and the correlation matrix of the combined fit can be found in fig. 5.39 and 5.40. The fit results for the Z and the $t\bar{t}$ are listed in tab. 5.19 .
The results for $\mu_{t\bar{t}}$ are consistent with the previous version of the analysis (page 72) and with the dedicated measurement [107]. Looking at the figure

Process	μ values
μ_Z	0.82 ± 0.09
$\mu_{t\bar{t}}$	0.81 ± 0.04

Table 5.19: Fitted μ values for Z and $t\bar{t}$ from a combined fit of the inclusive SR.

54 of this reference, reported in figure 5.42 for the sake of convenience, in the boosted regime and for pT>450 GeV, the $t\bar{t}$ production cross section measured by ATLAS turns out to be ∼20% less than SM predictions.
The impact of such recent result could not be evaluated when this version of the analysis was devised. As we will immediately see, $t\bar{t}$ modeling hugely impacts on the full fit model.

In fact, the Z process is measured to be two sigmas below expectations. The result refers to a precise Standard Model measurement and the discrepancy with theoretical predictions is larger than expected. If the correlation matrix and the pull distributions are considered (fig. 5.40 and fig. 5.41 respectively), it can be seen that μ_Z is strongly correlated with $\mu_{t\bar{t}}$, something that was not observed in the VR signal extraction test where ttbar contribution was kept fixed at the SM value. Both μ_Z and $\mu_{t\bar{t}}$ anti-correlate with QCD yields, as expected, but these correlations are affected by the Z/$t\bar{t}$ correlation. This unexpected behavior triggered new studies on the impact of $t\bar{t}$ modeling on the analysis, as well as on the quality of the QCD background parametrization.

At the time of writing, the possibility of looking at the Higgs mass window in an unbiased way critically depends on full understanding of the impact of $t\bar{t}$ modeling and QCD parametrization on the analysis procedure.
From the post-fit shapes a Z peak is clearly visible and that is confirmed by the fit results from which we estimate a significance of $\approx 9\sigma$. This result is a step forward with respect to the V+jet measurement performed previously (page 85), since now we are measuring a pure process with high precision. The next step will be a differential measurement of the Z in p_T bins of the Z, since with the high inclusive sensitivity shown in this measurement, a good precision in each bin is expected.

Finally, an event for a boosted $Z \to b\bar{b}$ event have been selected from 2018 data and shown in fig 5.43. In this figure the back-to-back topology is shown by the two reconstructed large-R jets. The muon coming out from one them most likely represents a semileptonic decay of a b-quark and we can therefore identify this jet as the signal one.

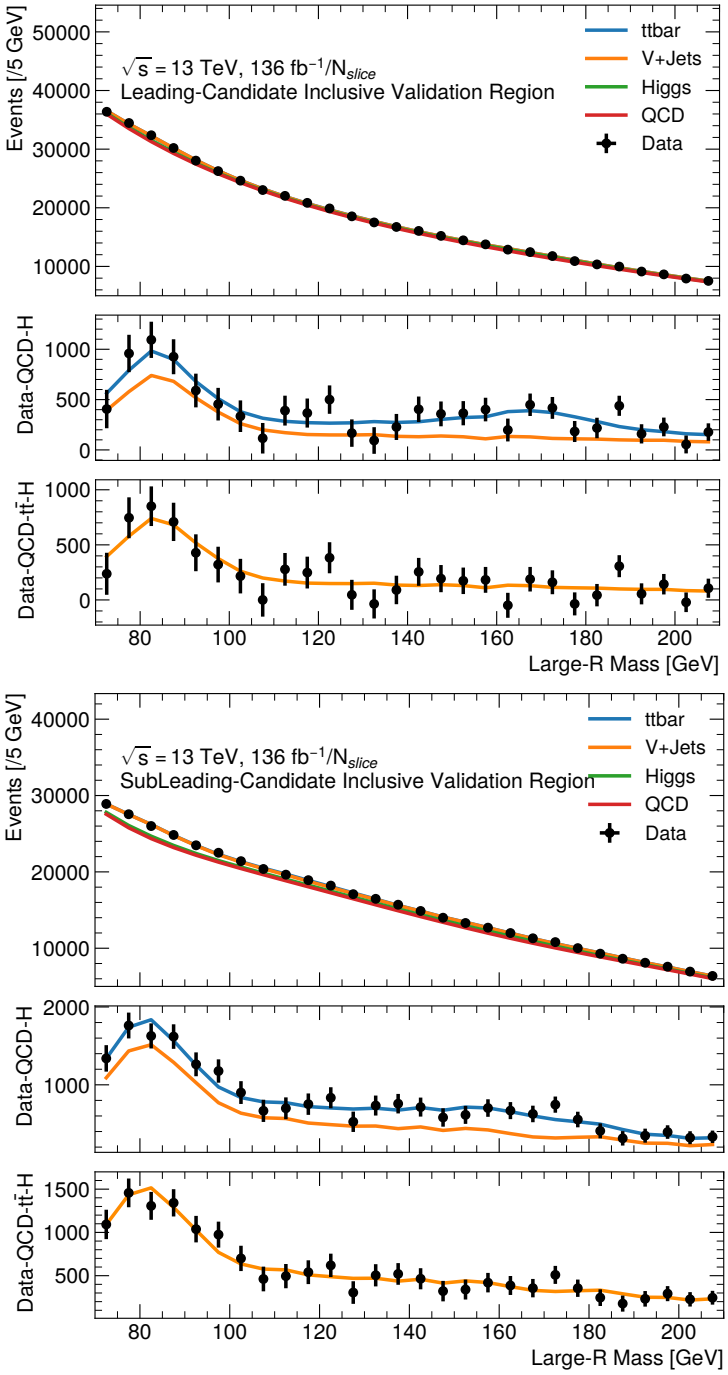


Figure 5.34: Post-fit plot for a VRL (top) and VRS (bottom) data slice analyzed with a complete fit model. In the top box all the background and signal contributions are shown, in the middle one the QCD function and the Higgs template are subtracted and in the bottom one only the V+jet process is left.

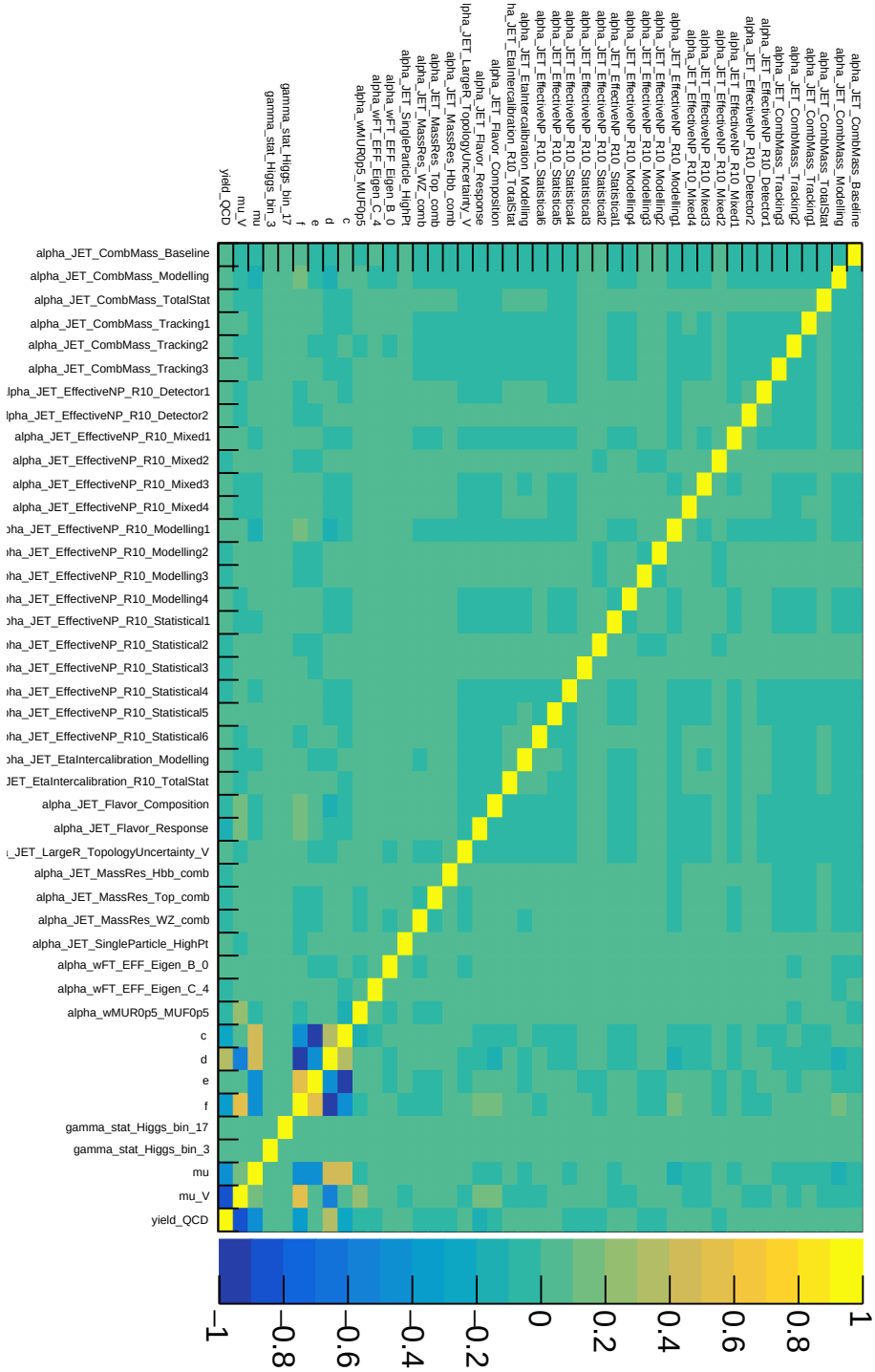


Figure 5.35: Correlation matrix plot for the VRL fit to data with a complete fit model.

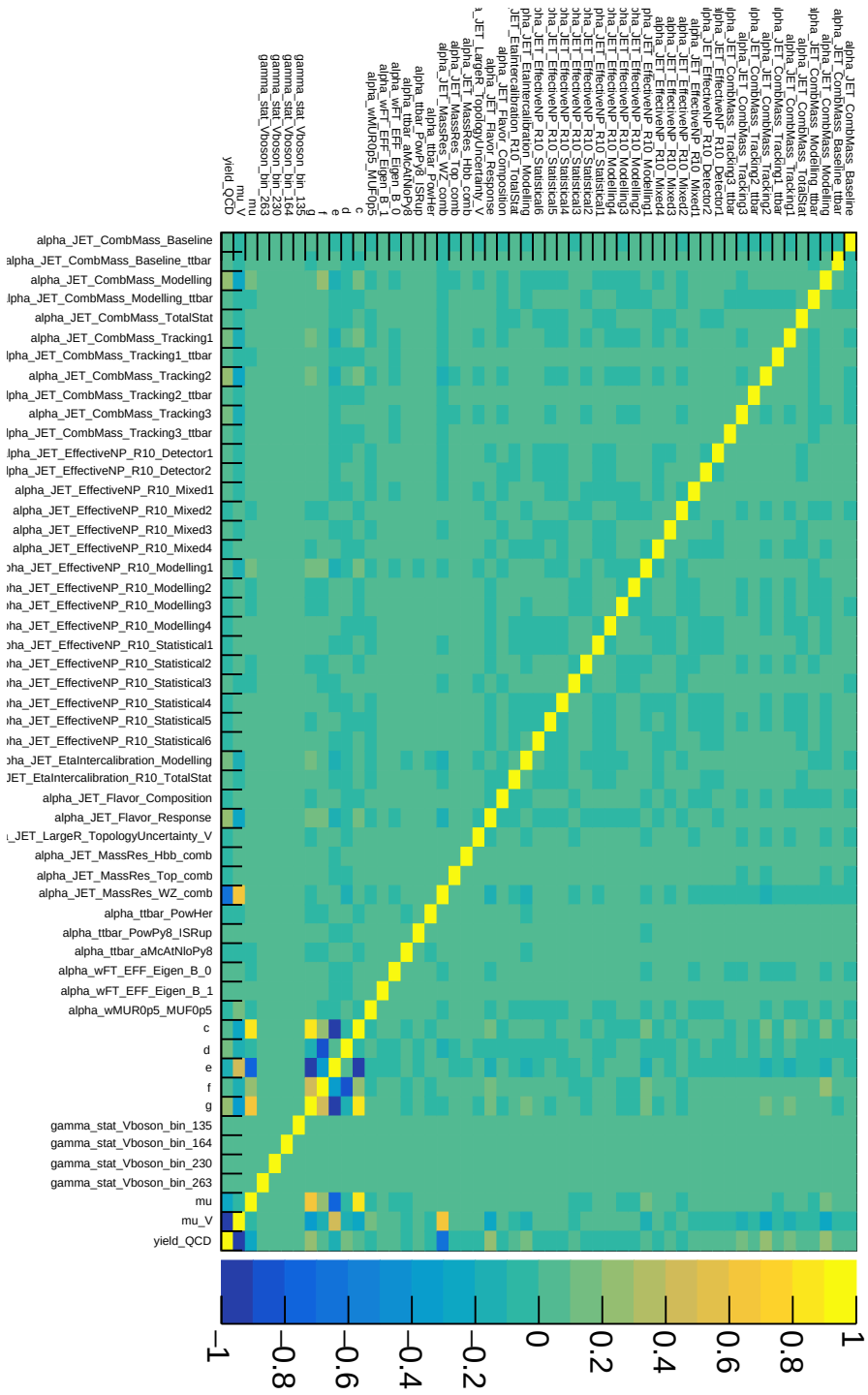


Figure 5.36: Correlation matrix plot for the VRS fit to data with a complete fit model

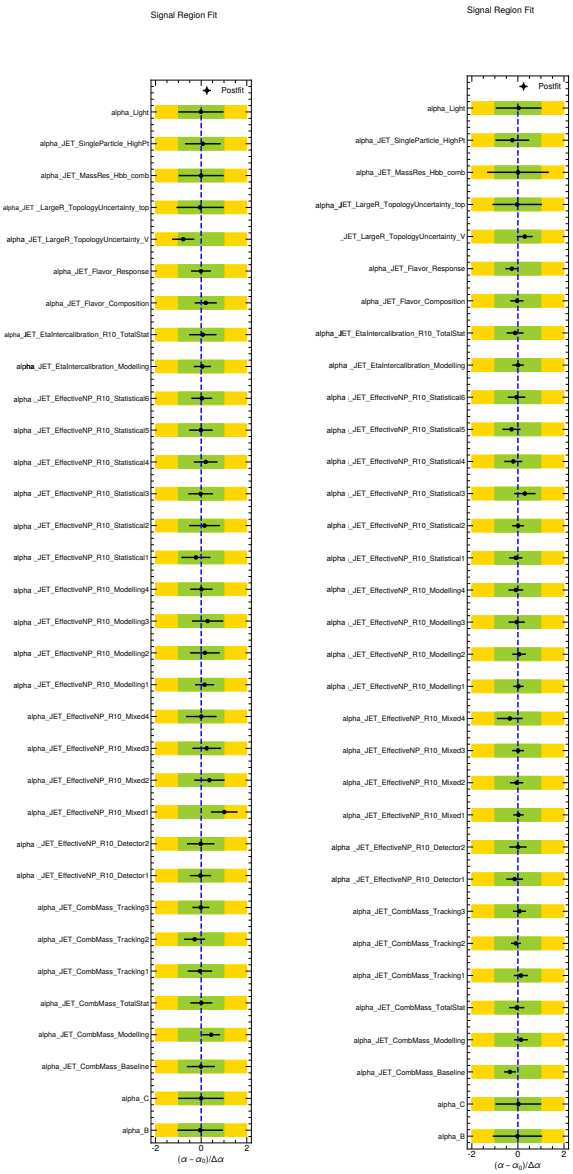


Figure 5.37: Representation of pulls (see definition at page 5.6.3)for the VRL (left) and VRS (right). Each VR slice was fit separately. Gamma parameters are not shown.

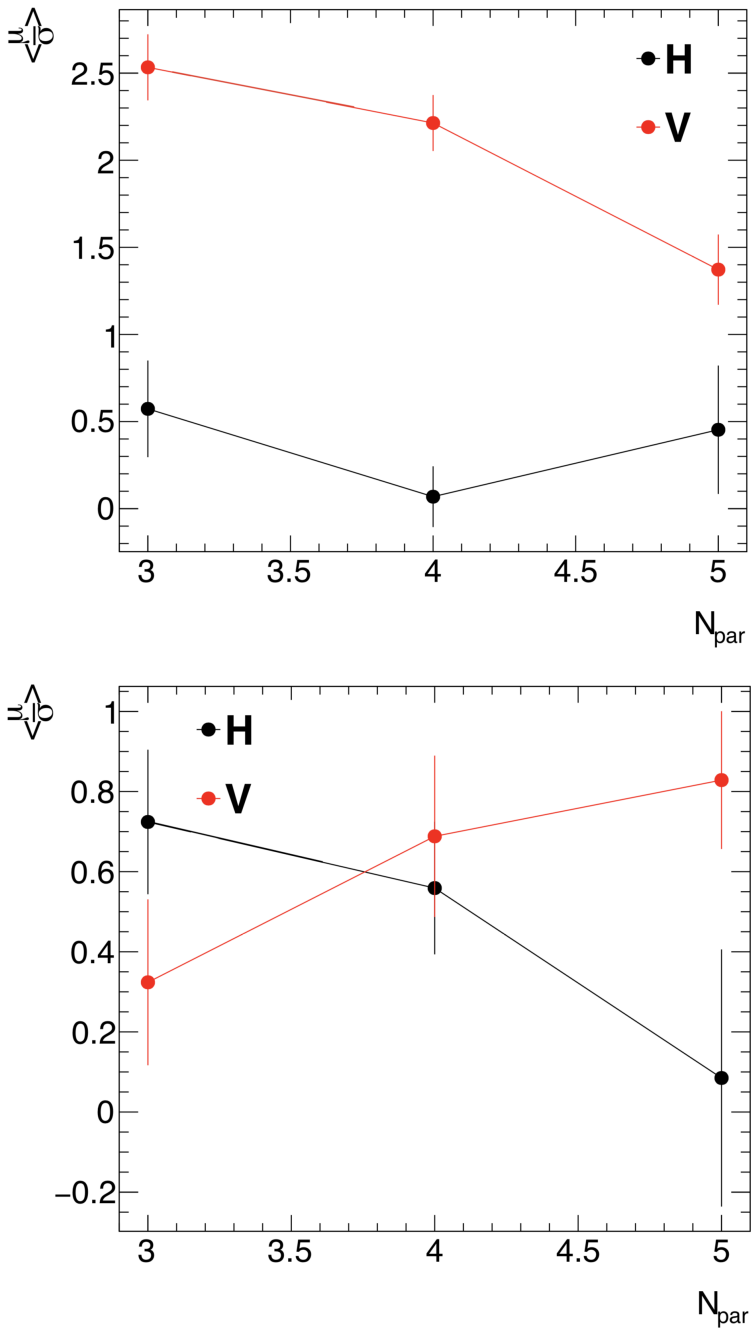


Figure 5.38: The average μ_V/σ_V (red) and μ_H/σ_H (right) versus N, for all the fits to the VRL (top) and VRS (bottom), where N is the number of parameters in the QCD function. For VRL, we use $N = 4$ and $N = 5$ for VRS. In the subleading plot opposite trends are seen for the Higgs and the V+j μ/σ when the number parameter increases. The correlation among these processes is under investigation.

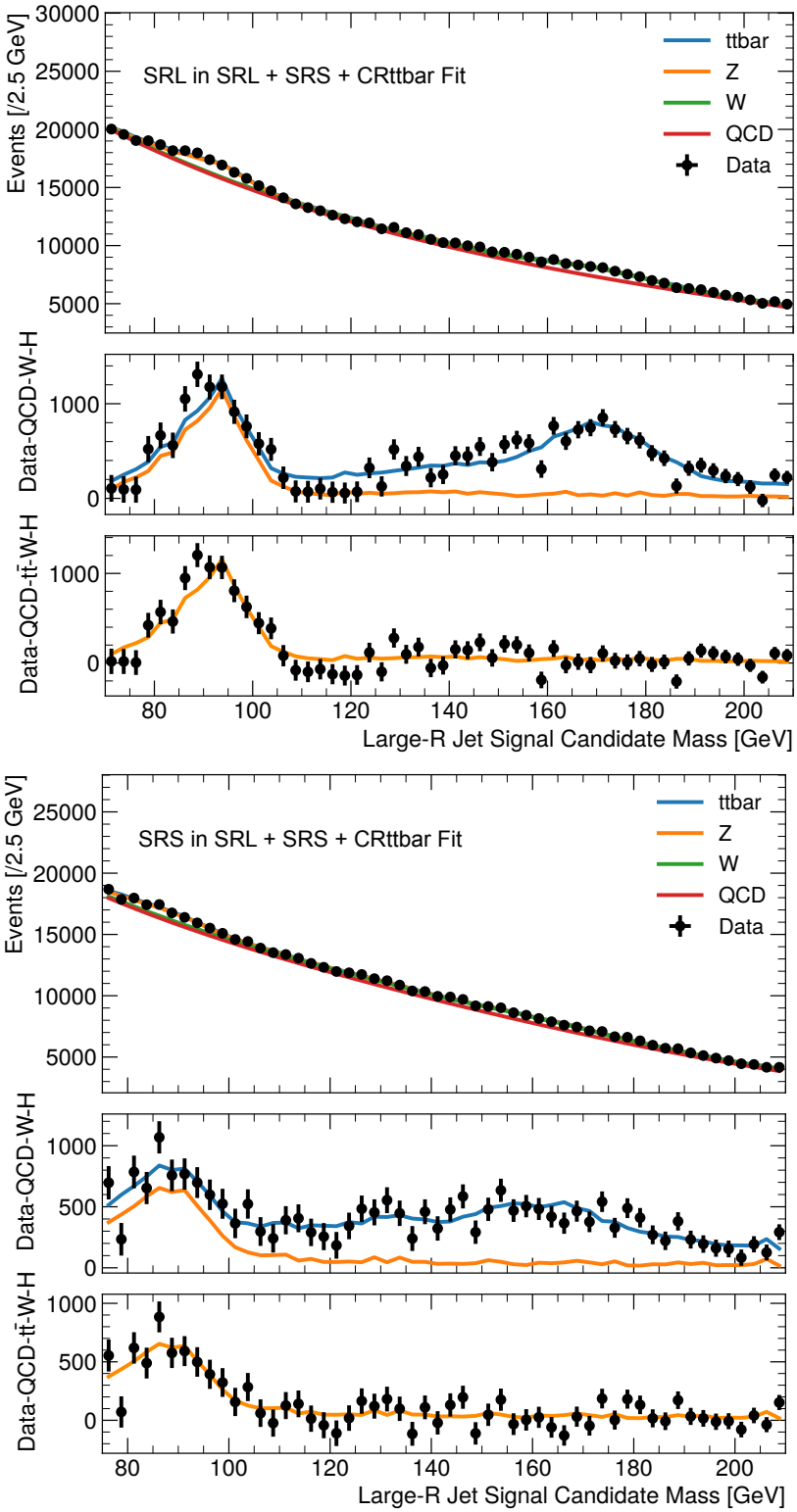


Figure 5.39: Post-fit plots from the combined fit on data SR regions. The plots from top to bottom show SRL and SRS, the Higgs template is subtracted at its SM rate.

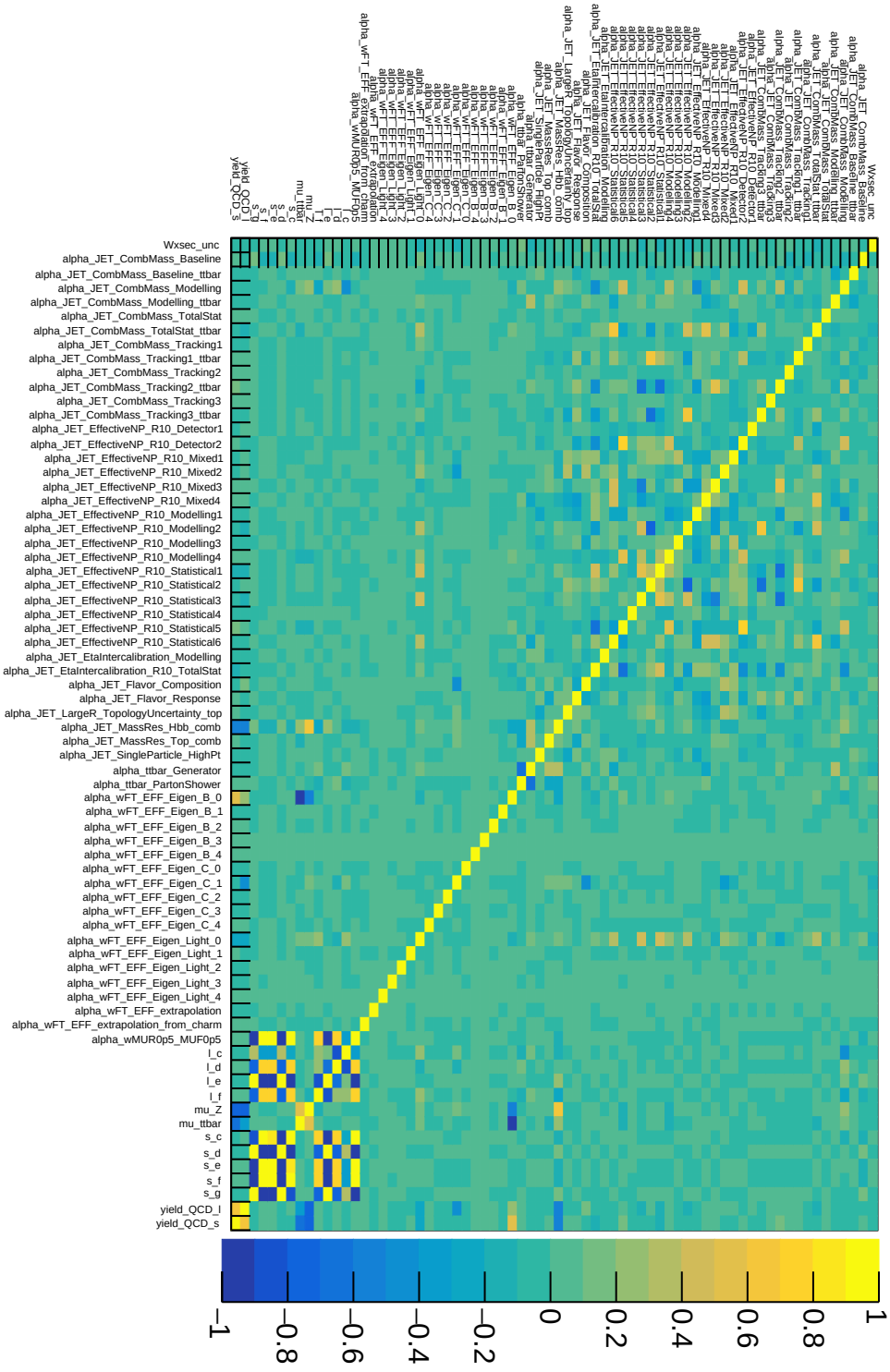


Figure 5.40: Correlation matrix of the combined fit. The parameters "$l\_$" and "$s\_$" are respectively the QCD parameters of the leading and the sub-leading SR.

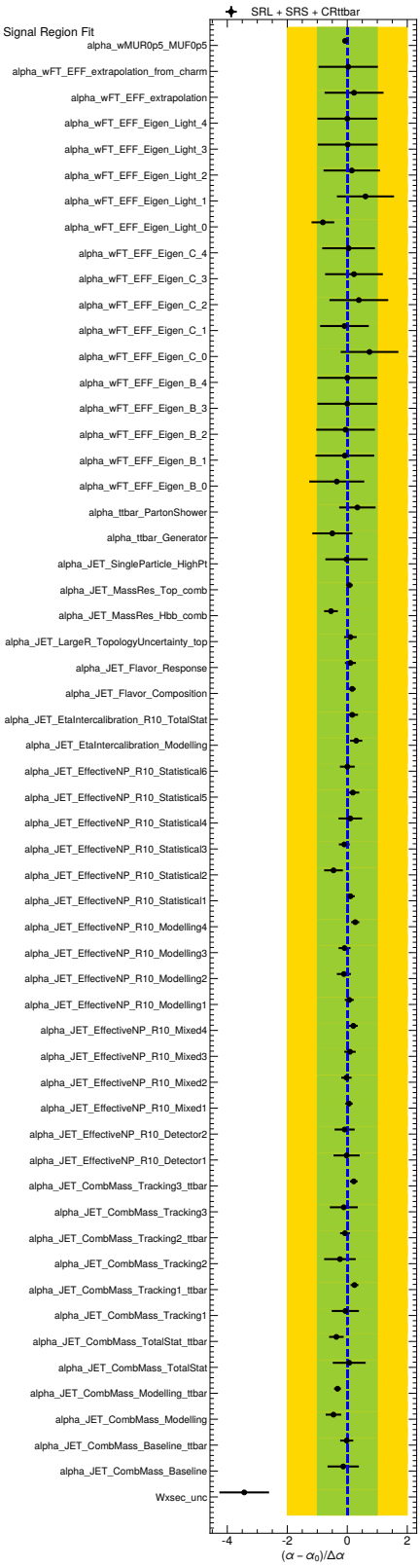


Figure 5.41: The pull plot of the combined fit is shown.

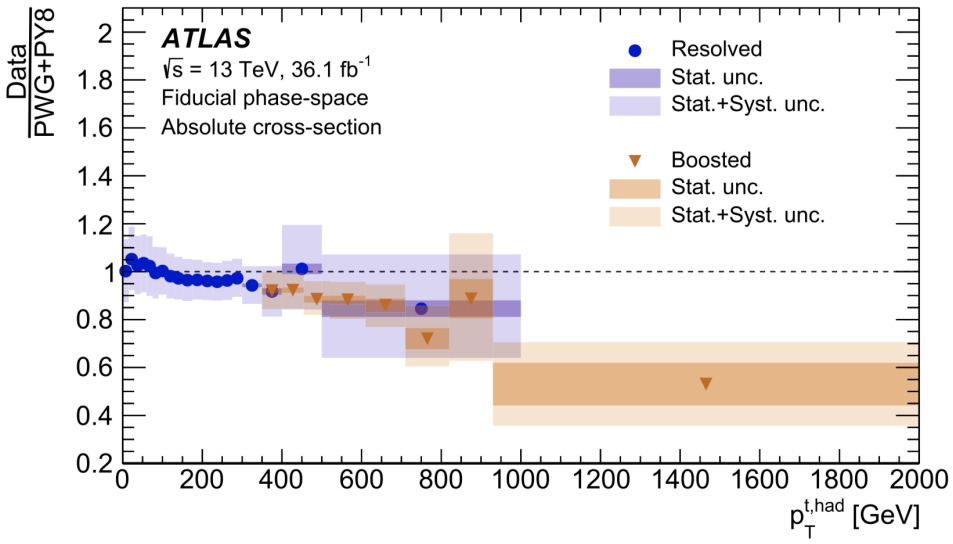


Figure 5.42: Figure from [107]. The ratios of the measured fiducial phase-space absolute differential cross-sections to the predictions obtained with the Powheg+Pythia8 MC generator in the resolved and boosted topologies as a function of the transverse momentum of the hadronic top quark. The bands indicate the statistical and total uncertainties of the data in each bin.

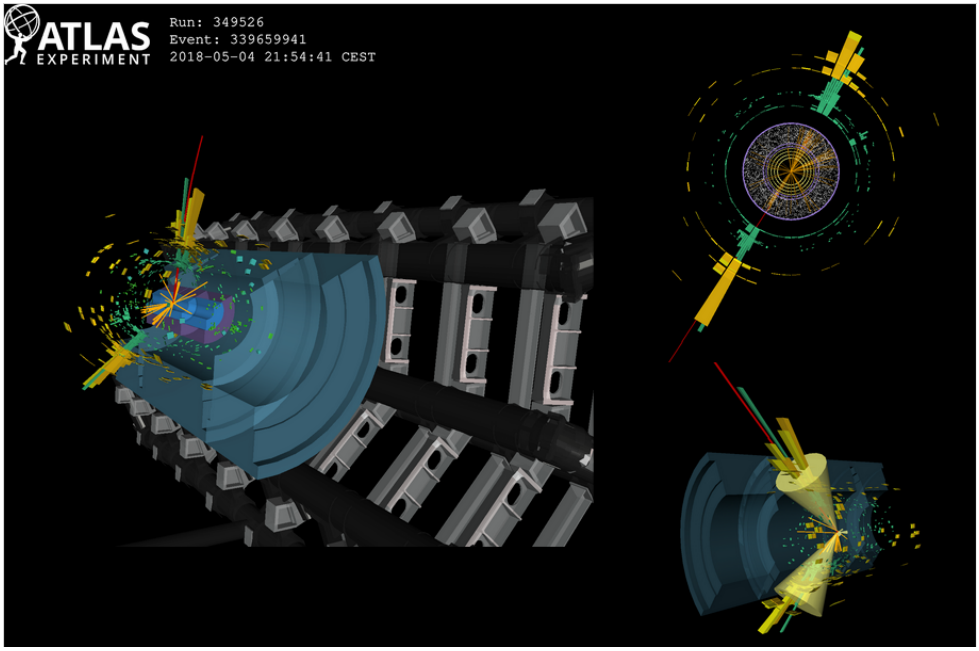


Figure 5.43: Display of a Z+jet event for the 2018 data taking. On the right the radial (top) and the longitudinal (bottom) views are shown. The green/yellow rectangles represent the energy deposits in the electromagnetic/hadronic calorimeter, while the red line stands for the flight direction of a muon.

Conclusions

In this thesis work, a search for the Higgs boson at high p_T and in association with a jet has been described in detail. Two versions of this analysis have been discussed: a first version with 80 fb^{-1} at $\sqrt{s} = 13$ TeV and a second one which makes use of the full Run2 statistics (140 fb^{-1}). In both cases the main limitation on the final results comes from the statistics and the main challenge is represented by QCD multijet background modeling.
The first version of the analysis was focused on the search for low-mass exotic resonances decaying into b-quark pairs, therefore not optimized towards the Higgs boson. This work describes all steps made to perform this first measurement.
Since this analysis is sensitive to vector bosons W and Z, the μ_V has been measured as well. The observed signal strength for the SM V+jets process is

$$\mu_V = 1.5 \pm 0.22(stat.)^{+0.29}_{-0.25}(syst.) \pm 0.18(th.).$$

Above the null hypothesis, once correlations are accounted for, it corresponds to a significance of 5 standard deviations. For the H+jets process, the observed signal strength is

$$\mu_H = 5.8 \pm 3.1(stat.) \pm 1.9(syst.) \pm 1.7(th.),$$

consistent with the background-only hypothesis at 1.6σ. No exotic resonances have been measured and upper limits in the analysis mass range have been set. This first set of results has shown the need for increasing the signal acceptance. For this reason, in the second version of the analysis I performed a comprehensive set of studies to increase the sensitivity to the Higgs boson and to effective field theories at high transverse momentum. The main achievements have been:

- introduction of new triggers with lower p_T cut and with a threshold on the large-R jet mass. This solution extends the analysis reach to events with lower p_T and enriches the selected dateset;
- inclusion of Higgs subleading large-R jet event topology;
- implementation of the muon-in-jet correction, improving the mass resolution all over the explored range.

These studies clearly showed the real analysis potential. In fact, the extension of the p_T explored range allowed to go for a differential measurement, a level of detail not addressed in the first version of the analysis. The complexity of the analysis grew accordingly, passing from one to seven signal regions to be considered. To properly model backgrounds became more problematic and each signal region had to be matched to a validation region, constructed *ad-hoc* to constrain the QCD parametrization.
All results currently available are shown, including the extraction of the Z boson signal strength in the signal region. The measurement of the Higgs signal strength is not reported, because the internal ATLAS review process has scheduled it beyond the scope of this PhD thesis.
This work poses the basis for future high-statistics/high-sensitivity analyses of H(bb)+jet in the boosted regime, as well as for next searches for beyond Standard Model physics in the J(bb)+J signature.

Appendix A

Alternative generator studies for V+jet background

The motivation behind this study is to check the size of the generator uncertainty on the $Z/W^{\pm}$+jets sample. This section compares important distributions of large-R jets for two different MC generators: Sherpa (nominal) and Herwig (alternative).
The comparison was performed in two steps:

- first the truth-level samples have been checked. Since it was a preliminary step, it was performed only for Z samples;
- second the agreement between the two MC generators using a full-simulation was checked. This test was performed on both Z and W samples.

In order to make these studies, truth-level and fully simulated versions of the V+jet samples have been produced. The focus of the comparison is the effect on the $Z/W^{\pm}$ mass distribution after the event selection.
The Sherpa sample contains $Z/W^{\pm}$ bosons generated with up to 4 jets. The Herwig sample only contains a single extra parton. This difference will be controlled by a cut on the number of large-R jets in the event.

A.0.1 Truth-Level Comparison

The full offline selection cannot be applied due to lack of b-tagging at truth level. Therefore, the event selection was applied up to that cut as follows:

1. p_T of the leading large-R jet above 480 GeV;
2. p_T of the subleading large-R jet above 250 GeV;
3. (extra) number of large-R jet in the event equal to two;

This selection was used to compare the two MC generators. The cut on the large-R jet number is used to bring the Sherpa and Herwig samples to the same

number of extra partons. It allows to disentangle the effects of the generator algorithms and order of the calculation.
Figures A.1 and A.2 show the p_T and mass spectra of the leading large-R jet respectively. The mass distributions are normalized to unity between 70 and 100 GeV. This allows to check the impact of the modeling uncertainty in the region around the peak. At the peak, the two mass distributions are very close. For what concerns the tails, Sherpa is always above Herwig, but this effect is expected as the NLO advantage of Sherpa affects the "non-Z" large-R jets forming the tails.

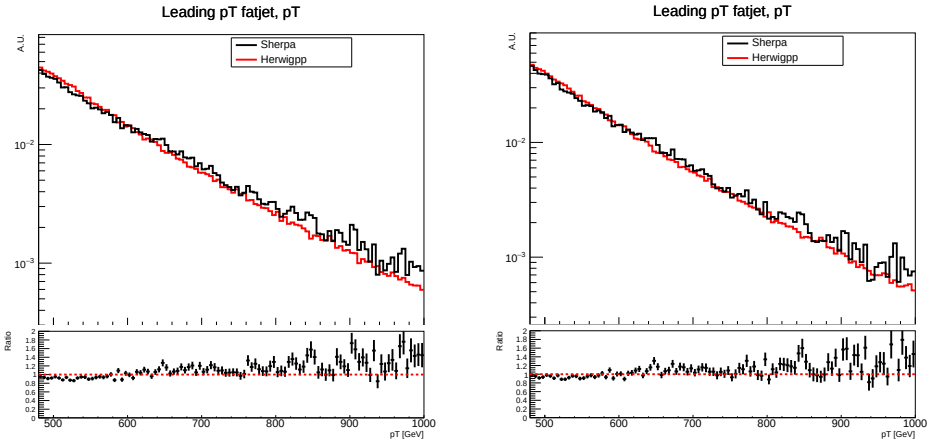


Figure A.1: Comparison of the truth leading large-R jet p_T distributions for Z+jet samples generated using Sherpa (black) and Herwig (red). The $N_{large-R} = 2$ cut is used for the right plot. The distributions are normalized to the number of events between 480 and 1000 GeV

A.0.2 Full Simulation Comparison

The full simulation samples have all the variables necessary to reproduce the complete event selection and therefore provide a better estimate of the modeling uncertainty in the analysis. The same selection as described for the signal region in the 80 fb^{-1}, with the following variations on the b-tagging requirement:

- the W+jet sample without any requirement on the number of b-tags in the signal candidate,
- the Z+jet sample without any requirement on the number of b-tags in the signal candidate,
- the Z+jet sample with the same double b-tagging as used in the analysis,
- and a second selection on the Z+jet sample uses truth b-tagging to increase the chance that the signal candidate contains the Z boson.

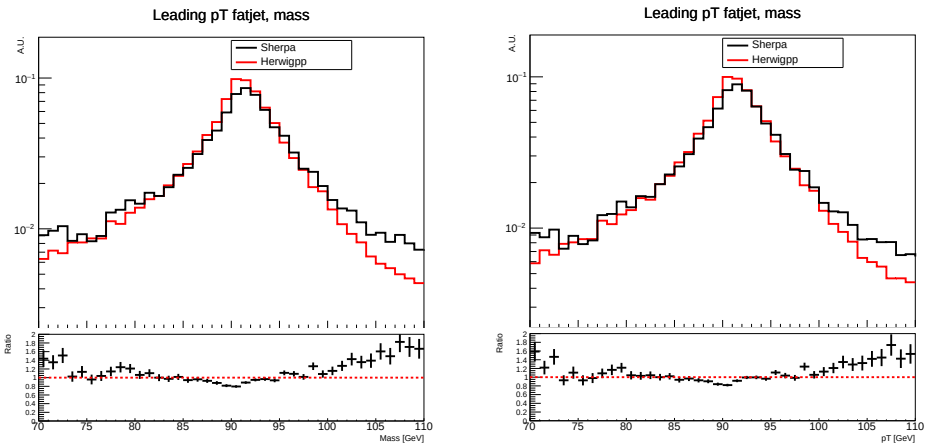


Figure A.2: Comparison of the truth leading large-R jet mass distributions for Z+jet samples generated using Sherpa (black) and Herwig (red). The $N_{large-R} = 2$ cut is used for the right plot. The distributions are normalized to the number of events between 70 and 110 GeV.

As with the truth-level studies, an extra selection with $N_{large-R} = 2$ cut is introduced to bring Sherpa and Herwig to the same order.
The comparison between Sherpa and Herwig for the signal candidate large-R jet mass for W+jets is shown in figure A.3. The agreement for the mass distributions is good around the mass peak. The same comparison for the Z boson is shown in figure A.4.

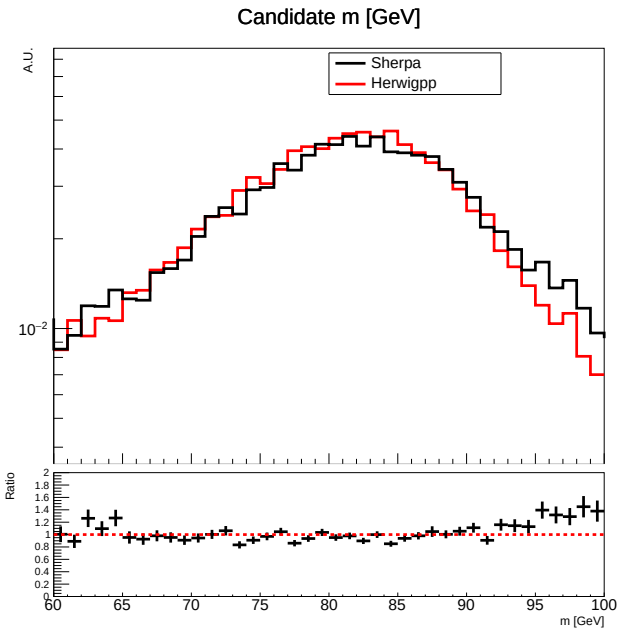


Figure A.3: Comparison of the leading large-R jet distributions for W+jet samples generated using Sherpa (black) and Herwig (red). The $N_{large-R} = 2$ cut is used for the right plots. The distributions are normalized to the number of events between 60 and 100 GeV.

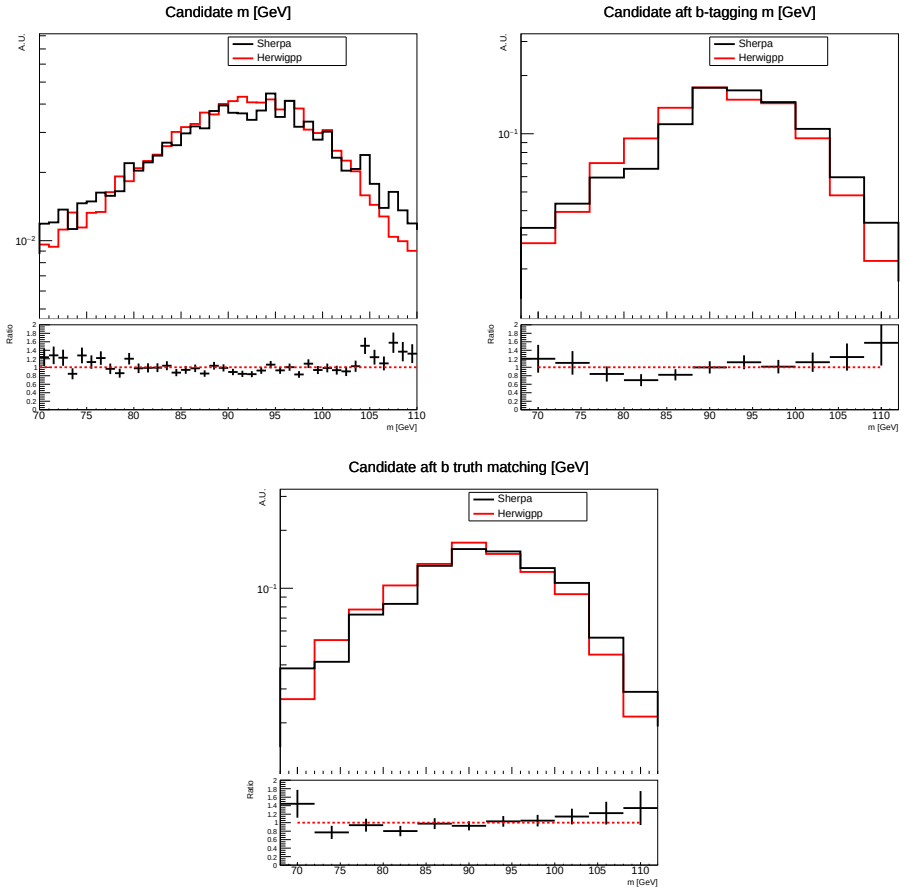


Figure A.4: Comparison of the leading large-R jet distributions for Z+jet samples generated using Sherpa (black) and Herwig (red). Selections without b-tagging (top), with regular b-tagging (middle) and truth-b-tagging (bottom) are shown. The $N_{large-R} = 2$ cut is used for the right plots. The distributions are normalized to the number of events between 70 and 110 GeV.

Appendix B

Signal Injection

All the results for the p_T binned signal regions are shown in the following section. No unexpected feature is observed. Results of the signal injection tests are shown in Figures B.1 and B.5 and in tables from B.1 to B.10.

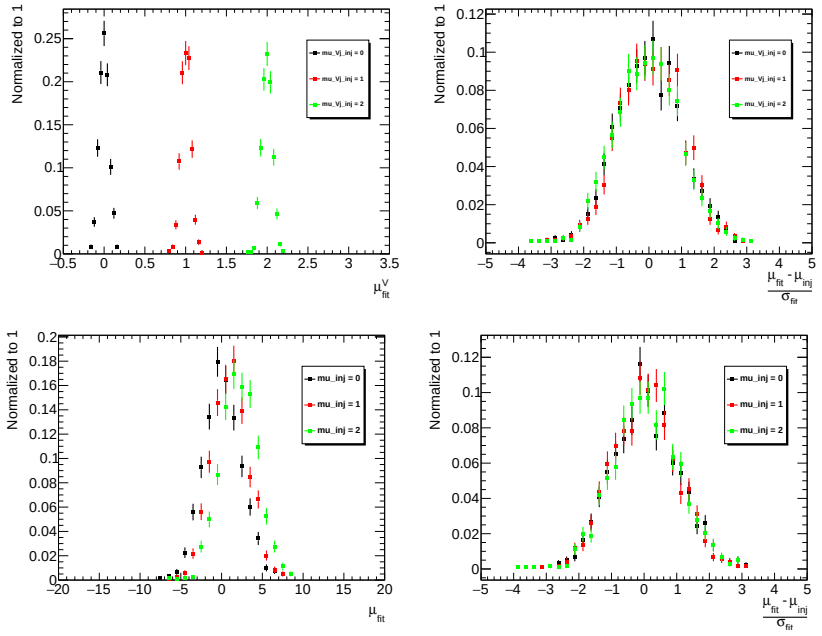


Figure B.1: The extracted signal strength distributions for a given injected signal strength in SRL1 are shown for V (top) and Higgs (bottom). μ distributions are on the left and pull distributions on the right. The pulls are centered at 0 with a width of 1.

Leading 2bin SR

μ_H^{inj}	μ_V^{inj}	μ_H^{fit}	$\sigma_{\mu_H}^{fit}$	μ_V^{fit}	$\sigma_{\mu_V}^{fit}$
0	0	-0.078 ± 0.117	2.225 ± 0.086	0.001 ± 0.003	0.062 ± 0.002
0	1	0.246 ± 0.126	2.346 ± 0.101	1.005 ± 0.003	0.062 ± 0.002
0	2	0.094 ± 0.118	2.268 ± 0.087	1.999 ± 0.003	0.064 ± 0.002
1	0	0.999 ± 0.107	2.020 ± 0.078	-0.003 ± 0.003	0.061 ± 0.002
1	1	1.119 ± 0.119	2.279 ± 0.094	1.003 ± 0.003	0.062 ± 0.002
1	2	0.757 ± 0.121	2.179 ± 0.094	2.001 ± 0.003	0.064 ± 0.002
2	0	2.186 ± 0.126	2.275 ± 0.099	0.003 ± 0.003	0.061 ± 0.003
2	1	2.186 ± 0.126	2.279 ± 0.104	0.999 ± 0.003	0.060 ± 0.002
2	2	1.989 ± 0.125	2.286 ± 0.102	1.993 ± 0.004	0.068 ± 0.002

Table B.1: μ and σ values are extracted with a gaussian fit to the $\mu_{H/V}$ distribution.

Leading 2bin SR

μ_H^{inj}	μ_V^{inj}	$Pull_H^{fit}$	$\sigma(pull)_{\mu_H}^{fit}$	$Pull_V^{fit}$	$\sigma(pull)_{\mu_V}^{fit}$
0	0	-0.031 ± 0.051	0.976 ± 0.035	0.056 ± 0.049	0.962 ± 0.036
0	1	0.082 ± 0.055	1.035 ± 0.045	0.094 ± 0.050	0.968 ± 0.037
0	2	-0.001 ± 0.052	1.001 ± 0.038	-0.002 ± 0.049	0.956 ± 0.036
1	0	0.009 ± 0.049	0.917 ± 0.037	-0.062 ± 0.052	1.000 ± 0.039
1	1	0.040 ± 0.053	1.014 ± 0.041	0.101 ± 0.049	0.948 ± 0.034
1	2	-0.108 ± 0.054	0.994 ± 0.044	0.014 ± 0.051	0.957 ± 0.036
2	0	0.053 ± 0.051	0.957 ± 0.035	0.015 ± 0.049	0.939 ± 0.034
2	1	0.098 ± 0.056	1.009 ± 0.046	0.009 ± 0.051	0.936 ± 0.042
2	2	0.003 ± 0.051	0.951 ± 0.037	-0.075 ± 0.054	1.008 ± 0.036

Table B.2: μ and σ values are extracted with a gaussian fit to the $pull_{H/V}$ distribution.

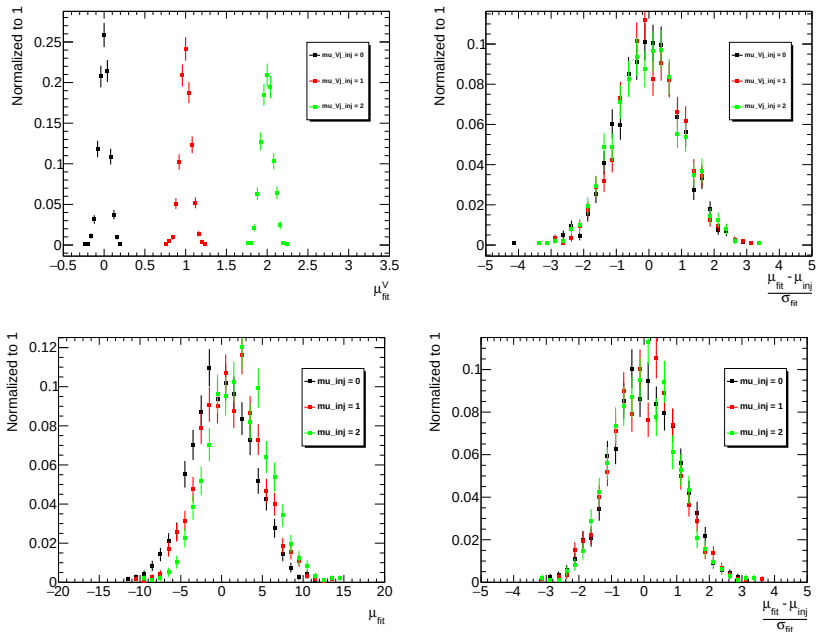


Figure B.2: The extracted signal strength distributions for a given injected signal strength in SRL2 are shown for V (top) and Higgs (bottom). μ distributions are on the left and pull distributions on the right. The pulls are centered at 0 with a width of 1.

Leading 3bin SR

μ_H^{inj}	μ_V^{inj}	μ_H^{fit}	$\sigma_{\mu_H}^{fit}$	μ_V^{fit}	$\sigma_{\mu_V}^{fit}$
0	0	0.162 ± 0.195	3.636 ± 0.163	-0.002 ± 0.003	0.063 ± 0.003
0	1	0.253 ± 0.191	3.550 ± 0.134	1.001 ± 0.003	0.068 ± 0.003
0	2	-0.161 ± 0.203	3.707 ± 0.162	2.003 ± 0.004	0.075 ± 0.003
1	0	0.937 ± 0.221	3.896 ± 0.201	-0.001 ± 0.003	0.062 ± 0.002
1	1	0.870 ± 0.178	3.252 ± 0.160	1.002 ± 0.003	0.063 ± 0.002
1	2	1.152 ± 0.210	3.752 ± 0.173	1.995 ± 0.004	0.073 ± 0.003
2	0	2.137 ± 0.201	3.580 ± 0.159	0.003 ± 0.003	0.057 ± 0.002
2	1	1.912 ± 0.228	3.773 ± 0.198	1.005 ± 0.004	0.064 ± 0.003
2	2	1.852 ± 0.271	3.662 ± 0.239	2.000 ± 0.005	0.068 ± 0.004

Table B.3: μ and σ values are extracted with a gaussian fit to the $\mu_{H/V}$ distribution.

Leading 3bin SR

μ_H^{inj}	μ_V^{inj}	$Pull_H^{fit}$	$\sigma(pull)_{\mu_H}^{fit}$	$Pull_V^{fit}$	$\sigma(pull)_{\mu_V}^{fit}$
0	0	0.025 ± 0.053	1.003 ± 0.043	-0.043 ± 0.053	1.032 ± 0.042
0	1	0.030 ± 0.051	0.966 ± 0.034	-0.021 ± 0.053	1.019 ± 0.041
0	2	-0.021 ± 0.053	0.997 ± 0.040	0.040 ± 0.056	1.038 ± 0.037
1	0	-0.042 ± 0.060	1.070 ± 0.053	-0.024 ± 0.052	1.013 ± 0.038
1	1	-0.029 ± 0.047	0.898 ± 0.036	0.038 ± 0.048	0.937 ± 0.034
1	2	-0.001 ± 0.056	1.047 ± 0.044	-0.085 ± 0.052	0.980 ± 0.036
2	0	0.002 ± 0.053	0.999 ± 0.040	0.067 ± 0.049	0.915 ± 0.037
2	1	-0.024 ± 0.057	1.004 ± 0.043	0.095 ± 0.053	0.952 ± 0.038
2	2	0.011 ± 0.065	0.940 ± 0.051	-0.033 ± 0.067	0.940 ± 0.048

Table B.4: μ and σ values are extracted with a gaussian fit to the $pull_{H/V}$ distribution.

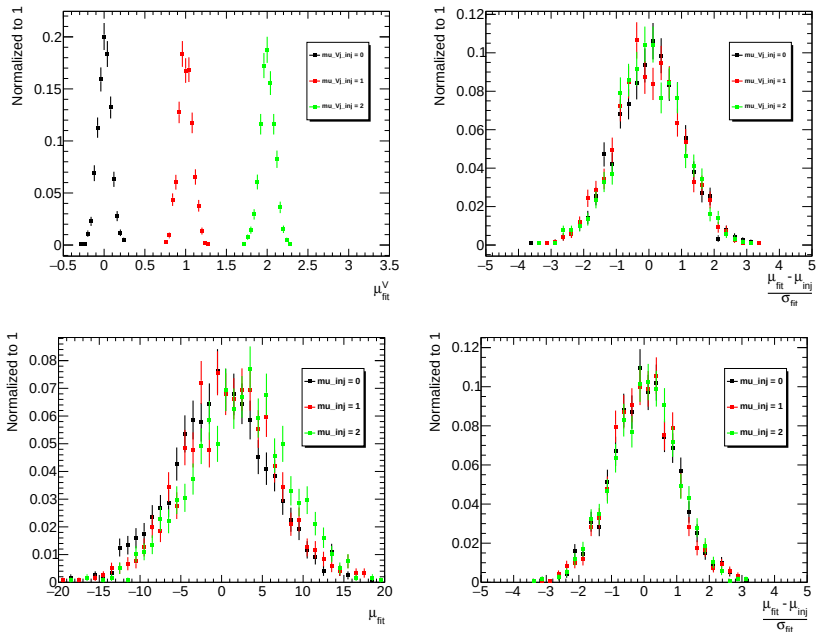


Figure B.3: The extracted signal strength distributions for a given injected signal strength in SRS0 are shown for V (top) and Higgs (bottom). μ distributions are on the left and pull distributions on the right. The pulls are centered at 0 with a width of 1.

Subleading 1bin SR

μ_H^{inj}	μ_V^{inj}	μ_H^{fit}	$\sigma_{\mu_H}^{fit}$	μ_V^{fit}	$\sigma_{\mu_V}^{fit}$
0	0	0.242 ± 0.382	5.870 ± 0.427	0.001 ± 0.004	0.076 ± 0.003
0	1	-0.368 ± 0.365	5.695 ± 0.392	0.998 ± 0.005	0.086 ± 0.003
0	2	0.455 ± 0.303	4.922 ± 0.282	1.999 ± 0.005	0.090 ± 0.004
1	0	0.638 ± 0.355	5.435 ± 0.332	0.004 ± 0.004	0.077 ± 0.003
1	1	1.353 ± 0.361	5.651 ± 0.363	0.998 ± 0.005	0.090 ± 0.004
1	2	0.547 ± 0.290	4.849 ± 0.250	2.009 ± 0.005	0.089 ± 0.004
2	0	2.533 ± 0.332	5.351 ± 0.344	0.005 ± 0.004	0.079 ± 0.003
2	1	2.790 ± 0.456	6.456 ± 0.556	0.999 ± 0.004	0.082 ± 0.003
2	2	1.647 ± 0.400	5.993 ± 0.475	2.004 ± 0.004	0.084 ± 0.003

Table B.5: μ and σ values are extracted with a gaussian fit to the $\mu_{H/V}$ distribution.

Subleading 1bin SR

μ_H^{inj}	μ_V^{inj}	$Pull_H^{fit}$	$\sigma(pull)_{\mu_H}^{fit}$	$Pull_V^{fit}$	$\sigma(pull)_{\mu_V}^{fit}$
0	0	0.030 ± 0.052	0.990 ± 0.041	0.015 ± 0.052	0.978 ± 0.038
0	1	-0.007 ± 0.051	0.969 ± 0.040	-0.020 ± 0.054	1.020 ± 0.040
0	2	-0.018 ± 0.050	0.940 ± 0.041	-0.013 ± 0.054	1.039 ± 0.043
1	0	-0.044 ± 0.049	0.950 ± 0.039	0.030 ± 0.052	1.000 ± 0.044
1	1	0.126 ± 0.052	1.002 ± 0.045	-0.028 ± 0.056	1.079 ± 0.045
1	2	-0.056 ± 0.045	0.872 ± 0.036	0.094 ± 0.054	1.024 ± 0.050
2	0	0.080 ± 0.050	0.940 ± 0.036	0.072 ± 0.051	0.977 ± 0.038
2	1	0.121 ± 0.056	1.039 ± 0.043	0.001 ± 0.052	0.987 ± 0.039
2	2	-0.032 ± 0.053	1.008 ± 0.042	0.048 ± 0.049	0.951 ± 0.036

Table B.6: μ and σ values are extracted with a gaussian fit to the $pull_{H/V}$ distribution.

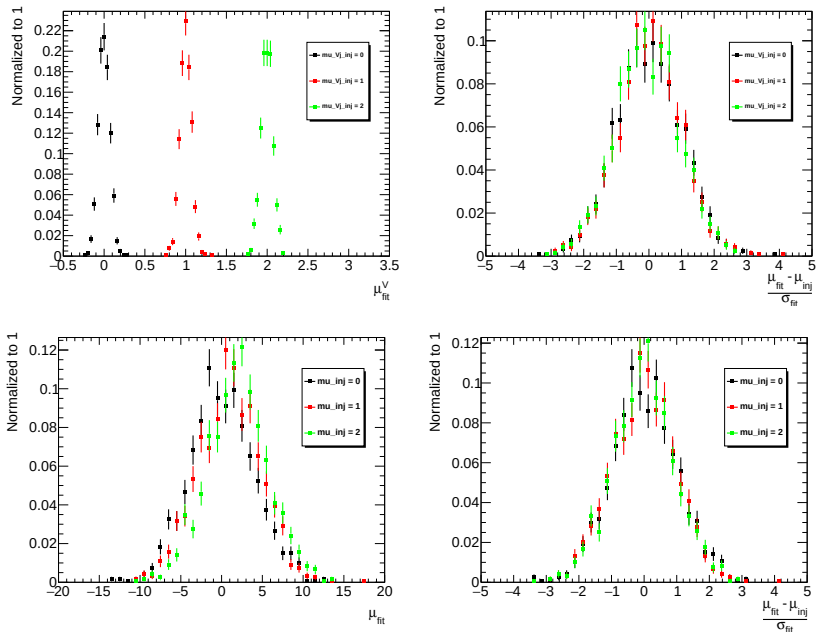


Figure B.4: The extracted signal strength distributions for a given injected signal strength in SRS1 are shown for V (top) and Higgs (bottom). μ distributions are on the left and pull distributions on the right. The pulls are centered at 0 with a width of 1.

Subleading 2bin SR

μ_H^{inj}	μ_V^{inj}	μ_H^{fit}	$\sigma_{\mu_H}^{fit}$	μ_V^{fit}	$\sigma_{\mu_V}^{fit}$
0	0	-0.100 ± 0.184	3.495 ± 0.172	0.002 ± 0.004	0.068 ± 0.002
0	1	0.441 ± 0.213	3.854 ± 0.175	1.003 ± 0.004	0.071 ± 0.003
0	2	0.155 ± 0.222	4.041 ± 0.165	2.001 ± 0.004	0.075 ± 0.003
1	0	0.938 ± 0.196	3.643 ± 0.160	-0.001 ± 0.004	0.070 ± 0.003
1	1	1.111 ± 0.211	3.921 ± 0.180	0.998 ± 0.004	0.071 ± 0.003
1	2	0.963 ± 0.204	3.776 ± 0.157	1.992 ± 0.004	0.073 ± 0.003
2	0	2.082 ± 0.210	3.702 ± 0.200	0.001 ± 0.004	0.073 ± 0.003
2	1	1.677 ± 0.178	3.363 ± 0.146	1.003 ± 0.004	0.070 ± 0.003
2	2	1.948 ± 0.206	3.830 ± 0.165	1.992 ± 0.004	0.073 ± 0.003

Table B.7: μ and σ values are extracted with a gaussian fit to the $\mu_{H/V}$ distribution.

Subleading 2bin SR

μ_H^{inj}	μ_V^{inj}	$Pull_H^{fit}$	$\sigma(pull)_{\mu_H}^{fit}$	$Pull_V^{fit}$	$\sigma(pull)_{\mu_V}^{fit}$
0	0	0.017 ± 0.048	0.899 ± 0.045	0.054 ± 0.054	1.009 ± 0.036
0	1	0.079 ± 0.055	1.029 ± 0.045	0.048 ± 0.051	0.981 ± 0.041
0	2	0.027 ± 0.053	1.028 ± 0.038	-0.004 ± 0.050	0.978 ± 0.034
1	0	-0.036 ± 0.049	0.953 ± 0.038	-0.015 ± 0.051	0.993 ± 0.043
1	1	0.010 ± 0.053	0.997 ± 0.043	-0.010 ± 0.049	0.964 ± 0.041
1	2	-0.024 ± 0.051	0.964 ± 0.037	-0.074 ± 0.051	0.978 ± 0.036
2	0	0.019 ± 0.049	0.931 ± 0.041	0.005 ± 0.054	1.026 ± 0.045
2	1	-0.083 ± 0.043	0.836 ± 0.034	0.007 ± 0.049	0.932 ± 0.040
2	2	0.012 ± 0.054	1.023 ± 0.044	-0.093 ± 0.049	0.961 ± 0.040

Table B.8: μ and σ values are extracted with a gaussian fit to the $pull_{H/V}$ distribution.

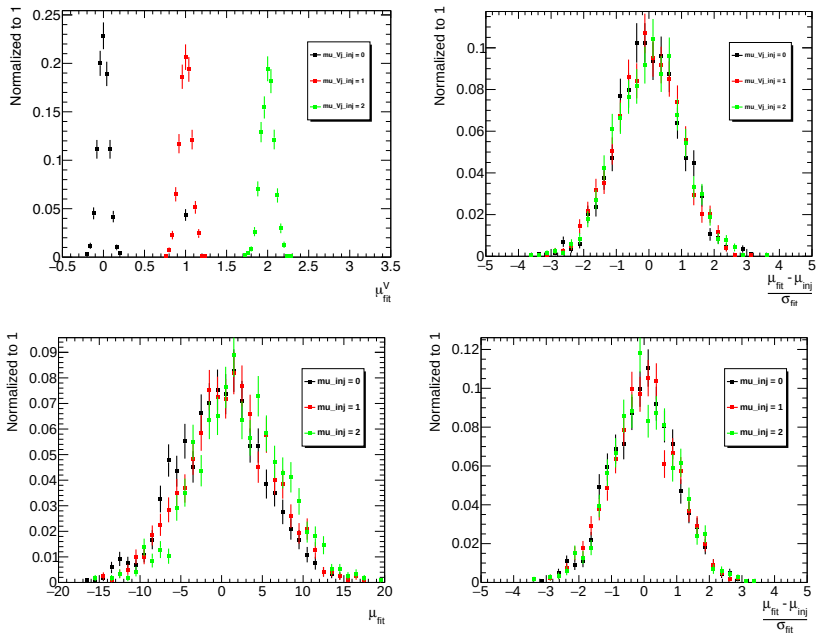


Figure B.5: The extracted signal strength distributions for a given injected signal strength in SRS2 are shown for V (top) and Higgs (bottom). μ distributions are on the left and pull distributions on the right. The pulls are centered at 0 with a width of 1.

Subleading 3bin SR

μ_H^{inj}	μ_V^{inj}	μ_H^{fit}	$\sigma_{\mu_H}^{fit}$	μ_V^{fit}	$\sigma_{\mu_V}^{fit}$
0	0	0.270 ± 0.313	5.078 ± 0.273	-0.000 ± 0.003	0.063 ± 0.003
0	1	0.388 ± 0.297	4.906 ± 0.278	1.000 ± 0.004	0.075 ± 0.003
0	2	-0.941 ± 0.318	5.100 ± 0.291	2.004 ± 0.004	0.080 ± 0.003
1	0	1.064 ± 0.322	5.148 ± 0.319	0.000 ± 0.003	0.065 ± 0.003
1	1	0.873 ± 0.355	5.543 ± 0.356	0.997 ± 0.004	0.078 ± 0.003
1	2	0.328 ± 0.294	4.831 ± 0.284	1.998 ± 0.004	0.081 ± 0.003
2	0	1.549 ± 0.302	4.992 ± 0.278	-0.004 ± 0.003	0.066 ± 0.002
2	1	2.228 ± 0.341	5.362 ± 0.315	1.000 ± 0.003	0.067 ± 0.002
2	2	2.367 ± 0.357	5.378 ± 0.348	2.000 ± 0.004	0.082 ± 0.003

Table B.9: μ and σ values are extracted with a gaussian fit to the $\mu_{H/V}$ distribution.

Subleading 3bin SR

μ_H^{inj}	μ_V^{inj}	$Pull_H^{fit}$	$\sigma(pull)_{\mu_H}^{fit}$	$Pull_V^{fit}$	$\sigma(pull)_{\mu_V}^{fit}$
0	0	0.068 ± 0.051	0.958 ± 0.038	-0.008 ± 0.048	0.928 ± 0.045
0	1	0.074 ± 0.050	0.948 ± 0.041	-0.011 ± 0.051	0.984 ± 0.039
0	2	-0.154 ± 0.054	1.021 ± 0.040	0.046 ± 0.050	0.994 ± 0.038
1	0	-0.023 ± 0.056	1.019 ± 0.041	0.011 ± 0.052	0.967 ± 0.043
1	1	0.039 ± 0.054	1.004 ± 0.039	-0.053 ± 0.054	1.035 ± 0.040
1	2	-0.056 ± 0.054	1.000 ± 0.046	0.000 ± 0.051	0.988 ± 0.035
2	0	-0.051 ± 0.053	1.009 ± 0.041	-0.032 ± 0.055	1.005 ± 0.039
2	1	0.030 ± 0.054	1.023 ± 0.041	-0.018 ± 0.050	0.949 ± 0.040
2	2	0.046 ± 0.054	1.023 ± 0.041	-0.011 ± 0.052	0.987 ± 0.037

Table B.10: μ and σ values are extracted with a gaussian fit to the $pull_{H/V}$ distribution.

Appendix C

Number of parameters choice

All the results for the p_T binned signal regions are shown in the following appendix in fig. C.1 and fig. C.2. In fig. C.3 the degradation of the modeling performance increasing statistics are shown on the leading and subleading inclusive signal regions.

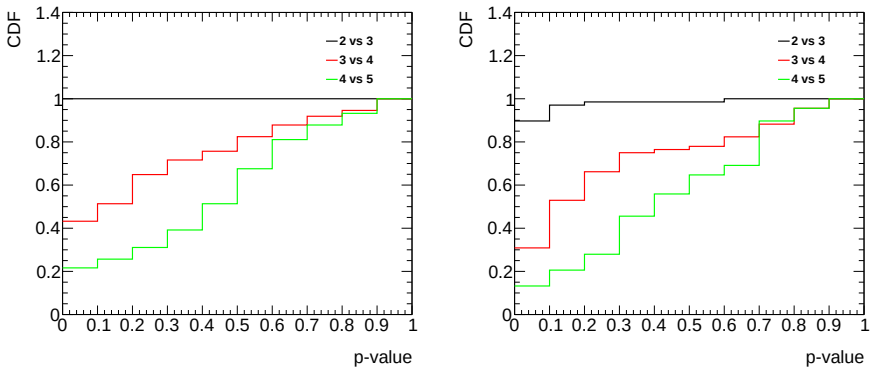


Figure C.1: Cumulative distribution of the probability for the difference of the log-likelihood obtained for exponential polynomial functions of order n and n + 1.

An example of this can be found in fig. C.4.

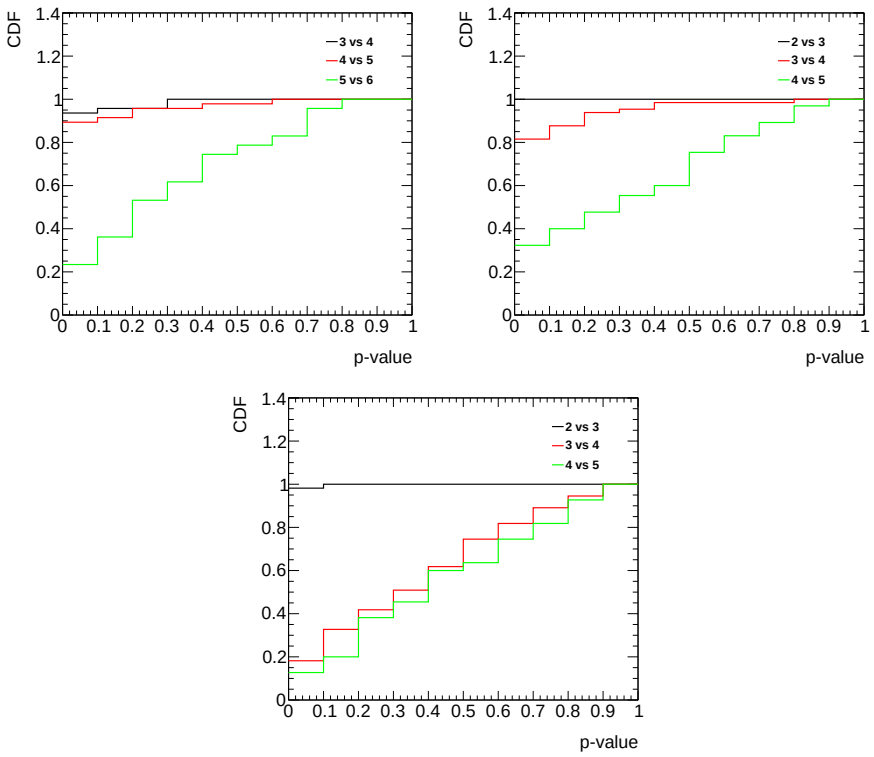


Figure C.2: Cumulative distribution of the probability for the difference of the log-likelihood obtained for exponential polynomial functions of order n and n + 1. Results for the three subleading VRs.

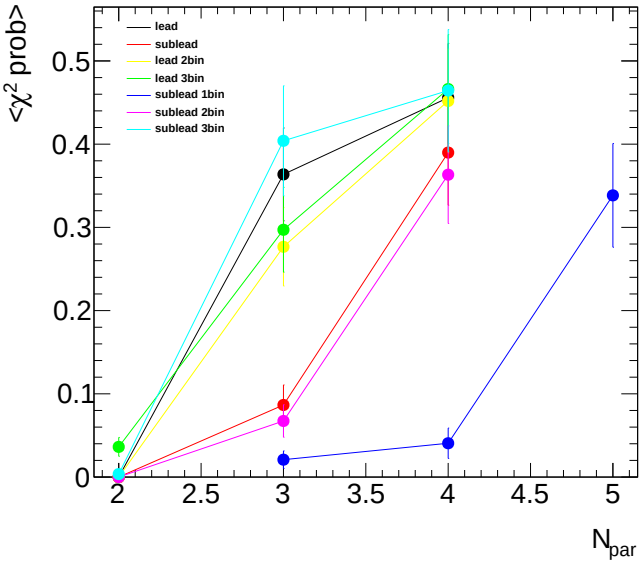


Figure C.3: The average χ^2 probability obtained by fitting all the VR data slices for different VRs and different function orders. The error bars are the RMS of the χ^2 probability for each region and each function order.

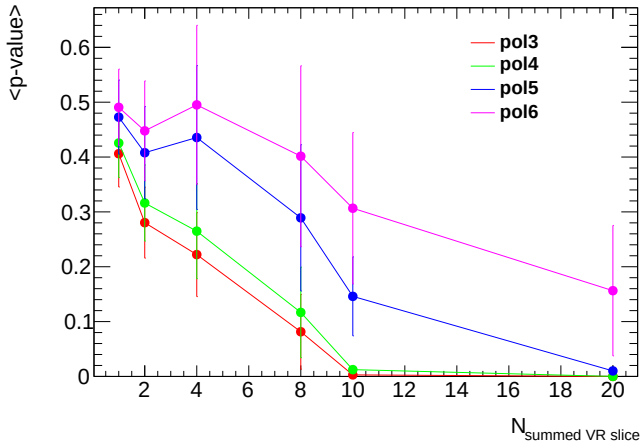


Figure C.4: The behavior of the average p-value of the fit exponential polynomial function to datasets made summing up different number of VR slices is shown. On the x-axis there is the dataset dimension in terms of summed VR slice (1 VR slice is equivalent to 140 fb^{-1} of SR). The error bars are calculated as the RMS of the p-value distribution over the square root of the number of independent in the datasets used.

Bibliography

[1] ALEPH Collaboration, DELPHI Collaboration L3, Collaboration, OPAL Collaboration, and the LEP Working Group for Higgs Boson Searches. Search for the Standard Model Higgs boson at LEP. *Physics Letters B*, 565:61–75, 2003. URL: https://doi.org/10.1016/S0370-2693(03)00614-2.

[2] Y. for the CDF Collaboration Oksuzian and for the D0 Collaboration. Searches for the Higgs boson at the Tevatron. *AIP Conference Proceedings*, 1560(43), 2013. URL: https://doi.org/10.1063/1.4826714.

[3] G. on behalf of the ATLAS Unal and CMS Collaborations. Higgs boson measurements at the LHC. 2018. URL: https://arxiv.org/abs/1811.10215.

[4] C. Degrande, B. Fuks, K. Mawatari, K. Mimasu, and V. Sanz. Electroweak Higgs boson production in the standard model effective field theory beyond leading order in QCD. *The European Physics Journal*, 77(262), 2017. URL: https://doi.org/10.1140/epjc/s10052-017-4793-x.

[5] C. Degrande, F. Maltoni, K. Mimasu, E. Vryonidou, and Zhang. C. Single-top associated production with a Z or H boson at the LHC: the SMEFT interpretation. *Journal of High Energy Physics*, (5), 2018. URL: https://doi.org/10.1007/JHEP10(2018)005.

[6] A. Azatov, C. Grojean, A. Paul, and E. Salvioni. Precise predictions for V+jets dark matter backgrounds. *Journal of High Energy Physics*, (123), 2016. URL: https://doi.org/10.1007/JHEP09(2016)123.

[7] N. Deutschmann, C. Duhr, Maltoni. F, and E. Vryonidou. Gluon-fusion Higgs production in the Standard Model Effective Field Theory. *Journal of High Energy Physics*, (63), 2017. URL: https://doi.org/10.1007/JHEP12(2017)063.

[8] M. Grazzini, A. Ilnicka, M. Spira, and M. Wiesemann. Modeling BSM effects on the Higgs transverse- momentum spectrum in an EFT approach. *Journal of High Energy Physics*, (115), 2017. URL: https://doi.org/10.1007/JHEP03(2017)115.

[9] C.N. Yang and R.L. Mills. Conservation of isotopic spin and isotopic gauge invariance. *Physical Review 96, 191*, 96(1), 1954. URL: https://doi.org/10.1103/PhysRev.96.191.

[10] M. Gell-Mann. A schematic model of baryons and mesons. *Physics Letter*, 8(3):214–215, 1964. URL: https://doi.org/10.1016/S0031-9163(64)92001-3.

[11] P. W. Higgs. Broken symmetries, massless particles and gauge fields. *Physics Letter*, 12(2):132–133, 1964. URL: https://doi.org/10.1016/0031-9163(64)91136-9.

[12] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Physical Review Letters*, 13(321), 1964. URL: https://doi.org/10.1103/PhysRevLett.13.321.

[13] J. Goldstone. Field theories with " Superconductor " solutions. *Il Nuovo Cimento (1955-1965)*, 19:154–164, 1961. URL: https://doi.org/10.1007/BF02812722.

[14] S. Windberg. A model of leptons. *Physical Review Letters*, 19(1264), 1967. URL: https://doi.org/10.1103/PhysRevLett.19.1264.

[15] A. Salam. Elementary particle physics: Relativistic groups and analyticity. *Eighth Nobel Symposium*, 1968.

[16] C.S. Wu, E. Ambler, R.W. Hayward, D.D. Hoppes, and R.P. Hudson. Experimental test of parity conservation in beta decay. *Physical Review Journals Archive*, 105(1413), 1957. URL: https://doi.org/10.1103/PhysRev.105.1413.

[17] J.H. Christenson, J.W. Cronin, V.L. Fitch, and R. Turlay. Evidence for the 2π decay of the k_0^2 meson system. *Physical Review Letters*, 13:138–140, 1964. URL: https://doi.org/10.1103/PhysRevLett.13.138.

[18] UA2 Collaboration. Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN anti-p p collider. *Physics Letter B*, 122(5-6):476–485, 1983. URL: https://doi.org/10.1016/0370-2693(83)91605-2.

[19] UA2 Collaboration. Evidence for Z0 $\rightarrow e^+e^-$ at the CERN anti-p p collider. *Physics Letter B129*, 129(1-2):130–140, 1983. URL: https://doi.org/10.1016/0370-2693(83)90744-X.

[20] CDF Collaboration. Observation of top quark production in $p\bar{p}$ collisions with the collider detector at Fermilab. *Physical Review Letters*, 74:2626–2631, 1995. URL: https://doi.org/10.1103/PhysRevLett.74.2626.

[21] S.W. et al. Herb. Observation of a dimuon resonance at 9.5 gev in 400 GeV proton-nucleus collisions. *Physical Review Letters*, 39:252–255, 1977. URL: https://doi.org/10.1103/PhysRevLett.39.252.

[22] M. et al. Breidenbach. Observed behavior of highly inelastic electron-proton scattering. *Physical Review Letters*, 23:935–939, 1969. URL: https://doi.org/10.1103/PhysRevLett.23.935.

[23] SLAC-SP-017 Collaboration. Discovery of a narrow resonance in e+ e- annihilation. *Physical Review Letters*, 33:1453–1455, 1974. URL: https://doi.org/10.1103/PhysRevLett.33.1453.

[24] F. Reines and C.L. Cowan. Detection of the free neutrino. *Physical Review*, 92(3):830, 1953. URL: https://doi.org/10.1103/PhysRev.92.830.

[25] G. et al. Danby. Observation of high-energy neutrino reactions and the existence of two kinds of neutrinos. *Physical Review Letters*, 9(1):36–44, 1962. URL: https://doi.org/10.1103/PhysRevLett.9.36.

[26] M.L. et al. Perl. Evidence for anomalous lepton production in e^+e^- annihilation. *Physical Review Letters*, 35(22):1489–1492, 1975. URL: https://doi.org/10.1103/PhysRevLett.35.1489.

[27] N. Cabibbo. Unitary symmetry and leptonic decays. *Physical Review Letters*, 10(531), 1963. URL: https://doi.org/10.1103/PhysRevLett.10.531.

[28] K.G. Wilson. Confinement of quarks. *Physical Review D*, 10(8):2445–2459, 1974. URL: https://doi.org/10.1103/PhysRevD.10.2445.

[29] Particle Data Group. Review of particle physics. *Chinese Physics C*, 38(9), 2014-2015. URL: http://pdg.lbl.gov/2014/download/rpp2014-Chin.Phys.C.38.090001.pdf.

[30] E. Neother. Invariant variations problem. *Transport Theory and Statistical Physics*, 1(3):183–207, 1971. URL: https://doi.org/10.1080/00411457108231446.

[31] M. Cepeda, S. Gori, M. Kado, and F. et al. Riva. Higgs physics at the hl-lhc and he-lhc. *CERN Yellow Reports: Monographs*, 7(CERN-LPCC-2018-04):221–584, 2019. URL: http://dx.doi.org/10.23731/CYRM-2019-007.221.

[32] ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs Boson with the atlas detector at the LHC. *Physics Letters B*, 716(1):1–29, 2012. URL: https://doi.org/10.1016/j.physletb.2012.08.020.

[33] CMS collaboration. Observation of a new boson at a mass of 125 gev with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30–61, 2012. URL: https://doi.org/10.1016/j.physletb.2012.08.021.

[34] ATLAS Collaboration. Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector. *Europena Physical Journal C*, 75(10):476, 2015. URL: https://doi.org/10.1140/epjc/s10052-015-3685-1.

[35] S. Di Vita, C. Grojean, G. Panico, M. Riembau, and T. Vantalon. A global view on the Higgs self-coupling. *Journal of High Energy Physics*, (69), 2017. URL: https://doi.org/10.1007/JHEP09(2017)069.

[36] Y. Hamada, H. Kawai, and K. Oda. Bare higgs mass at planck scale. *Physical Review D*, 87(5), 2015. URL: https://doi.org/10.1103/PhysRevD.87.053009.

[37] Daya Bay collaboration. Observation of electron-antineutrino disappearance at Daya Bay. *Physical Review Letters*, 108(17), 2012. URL: https://doi.org/10.1103/PhysRevLett.108.171803.

[38] S. on behalf of the LHCb collaboration Klaver. Lepton flavour universality in charged-current B decays. *17th Conference on Flavor Physics and CP Violation*, 2019. URL: https://arxiv.org/abs/1907.01500.

[39] Particle Data Group. Muon anomalous magnetic moment. *Physical Review D*, 98, 2019. URL: http://pdg.lbl.gov/2019.

[40] Planck Collaboration. Planck 2013 results. xv. CMB power spectra and likelihood. *Astronomy and Astrophysics*, 571(A16), 2014. URL: https://doi.org/10.1051/0004-6361/201321573.

[41] ATLAS Collaboration. Constraints on new phenomena via Higgs boson couplings and invisible decays with the ATLAS detector. *Journal of High Energy Physics*, 1511(206), 2016. URL: https://doi.org/10.1007/JHEP11(2015)206.

[42] ATLAS Collaboration. Constraints on mediator-based dark matter and scalar dark energy models using $\sqrt{s}$= 13 TeV pp collision data collected by the atlas detector. *Journal of High Energy Physics*, 1905(142), 2019. URL: https://doi.org/10.1007/JHEP05(2019)142.

[43] J. Wess and B. Zumino. Supergauge transformations in four-dimensions. *Nuclear Physics B*, 70(1):39–50, 1974. URL: https://doi.org/10.1016/0550-3213(74)90355-1.

[44] G. R. Farrar and S. Weinberg. Supersymmetry at ordinary energies. 2. r invariance, goldstone bosons, and gauge fermion masses. *Physics Review D*, 27(2732), 1983. URL: https://doi.org/10.1103/PhysRevD.27.2732.

[45] H. Baer and X. Tata. The minimal supersymmetric standard model. *Weak Scale Supersymmetry: From Superfields to Scattering Events, Cambridge university press*, pages 127–189, 2006. URL: https://doi.org/10.1017/CBO9780511617270.009.

[46] J. Wess. Supersimmetry and supergravity. *22nd International Conference on High Energy Physics. Vol. 1*, 1984. URL: https://doi.org/10.1007/978-3-642-71760-4_2.

[47] C.F. Berger, J.S. Gainer, Hewett J.L., and T. G. Rizzo. Supersymmetry without prejudice. *Journal of High Energy Physics*, 2(23), 2009. URL: https://doi.org/10.1088/1126-6708/2009/02/023.

[48] ATLAS Collaboration. Summary of the ATLAS experiment's sensitivity to supersymmetry after LHC Run 1 — interpreted in the phenomenological mssm. *Journal of High Energy Physics*, 10(134), 2015. URL: https://doi.org/10.1007/JHEP10(2015)134.

[49] L. Evans and P. Bryant. LHC machine. *Journal of Instrumentation*, 3(08), 2008. URL: https://doi.org/10.1088%2F1748-0221%2F3%2F08%2Fs08001.

[50] ATLAS Collaboration. The ATLAS experiment at the CERN Large Hadron Collider. *Journal of Instrumentation*, 3, 2008. URL: https://doi.org/10.1088/1748-0221/3/08/S08003.

[51] ATLAS Collaboration. Insertable B-Layer Technical Design Report. *ATLAS TDR 19, CERN/LHCC 2010-013*, 2010. URL: http://inspirehep.net/record/1609508/files/ATLAS-TDR-019.pdf.

[52] ATLAS collaboration. Alignment of the ATLAS Inner Detector with the initial LHC data at $\sqrt{s} = 13$ TeV. *ATL-PHYS-PUB-2015-031*, 2015. URL: https://cds.cern.ch/record/2038139.

[53] ATLAS collaboration. Study of alignment-related systematic effects on the ATLAS Inner Detector tracking. *ATLAS-CONF-2012-141*, 2012. URL: http://cdsweb.cern.ch/record/1483518.

[54] F. M. Follega, E. Ricci, and R. Iuppa. Updates on long-term alignment monitoring and diagnostics for atlas id misalignments. *ATLAS-CONF-2012-141*, 2019. URL: https://www.sif.it/riviste/sif/ncc/econtents/2019/042/04/article/57.

[55] ATLAS Collaboration. Calorimeter Performance Technical Design Report. *CERN-LHCC-96-40*, 1996.

[56] ATLAS Collaboration. The ATLAS Data Acquisition and High Level Trigger system. *Journal of Instrumentation*, 11, 2016. URL: https://doi.org/10.1088/1748-0221/11/06/P06008.

[57] G. Apollinari, O. Brüning, T. Nakamoto, and L. Rossi. High-luminosity Large Hadron Collider (HL-LHC). *CERN Yellow Reports: CERN-2015-005*, pages 1–19, 2017. URL: https://cds.cern.ch/record/2120673.

[58] L. on behalf of the ATLAS Collaboration Peilian. Expected performance of the upgrade ATLAS experiment for HL-LHC. *13th Conference on the Intersections of Particle and Nuclear Physics*, 2018. URL: https://cds.cern.ch/record/2637860/files/1809.02181.pdf.

[59] R. Pöschl. Future colliders - Linear and circular. *Proceedings of science - ALPS 2018*, 330, 2018. URL: https://doi.org/10.22323/1.330.0042.

[60] J. Shelton. TASI lectures on jet substructure. 2013. URL: https://arxiv.org/pdf/1302.0260.pdf.

[61] ATLAS Collaboration. Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1. *The European Physical Journal C*, 77(490), 2017. URL: https://doi.org/10.1140/epjc/s10052-017-5004-5.

[62] ATLAS Collaboration. Variable radius, exclusive-k_t, and center-of-mass subjet reconstruction for Higgs($\to b\bar{b}$) tagging in ATLAS. *ATL-PHYS-PUB-2017-010*, 2017. URL: http://cds.cern.ch/record/2268678/files/ATL-PHYS-PUB-2017-010.pdf?version=1.

[63] R. Atkin. Review of jet reconstruction algorithms. *Journal of Physics: Conference Series 645*, 2015. URL: https://doi.org/10.1088/1742-6596/645/1/012008.

[64] M. Cacciari, G.P. Salam, and G. Soyez. The anti-k_t jet clustering algorithm. *Journal of High Energy Physics*, 0804(063), 2008. URL: https://doi.org/10.1088/1126-6708/2008/04/063.

[65] A.J. Larkoski, I. Moult, and D. Neill. Power counting to better jet observables. *Journal of High Energy Physics*, 1412(009), 2014. URL: https://doi.org/10.1007/JHEP12(2014)009.

[66] ATLAS Collaboration. Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$=13 TeV. *The European Physics Journal C*, 76(292), 2016. URL: https://doi.org/10.1140/epjc/s10052-016-4120-y.

[67] ATLAS collaboration. Optimisation of the ATLAS b-tagging performance for the 2016 LHC run. *ATLAS Public Note*, 2016. URL: http://cds.cern.ch/record/2160731/?ln=it.

[68] ATLAS Collaboration. Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV. *Journal of High Energy Physics*, (89), 2018. URL: https://doi.org/10.1007/JHEP08(2018)089.

[69] ATLAS collaboration. ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ tev. *The European Physical Journal C*, 79(970), 2020. URL: https://arxiv.org/pdf/1907.05120.pdf.

[70] B.P. et al. Roe. Boosted decision trees as an alternative to artificial neural networks for particle identification. *Nuclear Instruments and Methods in Physics Research Section A*, 543(1-2):577–584, 2005. URL: https://doi.org/10.1016/j.nima.2004.12.018.

[71] ATLAS Collaboration. Search for boosted resonances decaying to two b-quarks and produced in association with a jet at $\sqrt{s} = 13$ TeV with the ATLAS detector. *ATLAS-CONF-2018-052*, 2018. URL: http://inspirehep.net/record/1705606/files/ATLAS-CONF-2018-052.pdf?version=1.

[72] ATLAS Collaboration. Observation of $h \to b\bar{b}$ decays and VH production with the ATLAS detector. *Physics Letters B*, 786:59–86, 2018. URL: https://doi.org/10.1016/j.physletb.2018.09.013.

[73] M.R. Buckley, D. Hooper, and J.L. Rosner. A leptophobic z' and dark matter from grand unification. *Physics Letters B7*, 703(3):343–347, 2011. URL: https://doi.org/10.1016/j.physletb.2011.08.014.

[74] ATLAS Collaboration. Search for light resonances decaying to boosted quark pairs and produced in association with a photon or a jet in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Lett. B*, 788:316–335, 2019. URL: https://doi.org/10.1016/j.physletb.2018.09.062.

[75] ATLAS Collaboration. Search for new phenomena in dijet events using 37 fb^{-1} of pp collision data collected at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Physics Review D*, 96(052004), 2017. URL: https://doi.org/10.1103/PhysRevD.96.052004.

[76] ATLAS Collaboration. Constraints on mediator-based dark matter and scalar dark energy models using $\sqrt{s} = 13$ TeV pp collision data collected by the ATLAS detector. *Journal of High Energy Physics*, 05(142), 2019. URL: https://doi.org/10.1007/JHEP05(2019)142.

[77] K. Hamilton, P. Nason, and G. Zanderighi. Finite quark-mass effects in the NNLOPS POWHEG+MiNLO Higgs generator. *Journal of High Energy Physics*, 05(140), 2015. URL: https://doi.org/10.1007/JHEP05(2015)140.

[78] E. Re et all. Single-top Wt-channel production matched with parton showers using the powheg method. *The European Physical Journal C*, 71(1547), 2011. URL: https://doi.org/10.1088/1126-6708/2009/02/007.

[79] R.D. Ball, V. Bertone, and S. et al. Carrazza. Parton distributions for the LHC Run II. *Journal of High Energy Physics*, 1504(40), 2015. URL: https://doi.org/10.1007/JHEP04(2015)040.

[80] T. et al Sjöstrand. An introduction to PYTHIA 8.2. *Computer Physics Communication*, 191:159–177, 2015. URL: https://doi.org/10.1016/j.cpc.2015.01.024.

[81] ATLAS Collaboration. Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Journal of High Energy Physics*, 1409(145), 2014. URL: https://doi.org/10.1007/JHEP09(2014)145.

[82] J. Pumplin et al. New generation of parton distributions with uncertainties from global QCD analysis. *Journal of High Energy Physics*, 0207(12), 2002. URL: https://doi.org/10.1088/1126-6708/2002/07/012.

[83] D.J. Lange. The evtgen particle decay simulation package. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 462(1-2):152–155, 2001. URL: https://doi.org/10.1016/S0168-9002(01)00089-4.

[84] B. Mellado Garcia, P. Musella, M. Grazzini, and R. Harlander. Cern report 4: Part I Standard Model Predictions. *LHCHXSWG-DRAFT-INT-2016-008*, 2016. URL: http://cds.cern.ch/record/2150771/?ln=it.

[85] D. et al Abercrombie. Dark matter benchmark models for early LHC Run-2 searches: report of the ATLAS/CMS dark matter forum. *Physics of the Dark Universe*, 27, 2020. URL: https://doi.org/10.1016/j.dark.2019.100371.

[86] J. et al Alwall. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics*, 1407(79), 2014. URL: https://doi.org/10.1007/JHEP07(2014)079.

[87] Andrew Nelson. Dark matter width calculator. URL: https://github.com/LHC-DMWG/DarkMatterWidthCalculation.

[88] ATLAS and CMS collaborations. Dark Matter Benchmark Models for Early LHC Run-2 Searches: Report of the ATLAS/CMS Dark Matter Forum. 2020. URL: https://doi.org/10.1016/j.dark.2019.100371.

[89] T. et al Gleisberg. Event generation with SHERPA 1.1. *Journal of High Energy Physics*, 02, 2009. URL: https://doi.org/10.1088/1126-6708/2009/02/007.

[90] M. et all Bahr. Herwig++ physics and manual. *The European Physical Journal C*, 58, 2008. URL: https://doi.org/10.1140/epjc/s10052-008-0798-9.

[91] ATLAS Collaboration. Selection of jets produced in 13 TeV proton-proton collisions with the ATLAS detector. *ATLAS-CONF-2015-029*, 2015. URL: http://inspirehep.net/record/1384980/files/ATLAS-CONF-2015-029.pdf?version=1.

[92] ATLAS Collaboration. Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 7$ TeV. *The European Physical Journal*, 73(2304), 2013. URL: https://doi.org/10.1140/epjc/s10052-013-2304-2.

[93] ATLAS Collaboration. Jet mass reconstruction with the ATLAS detector in early run 2 data. *ATLAS-CONF-2016-035*, 2016. URL: http://inspirehep.net/record/1477266/files/ATLAS-CONF-2016-035.pdf.

[94] G. Snecdecor and W. Cochran. Statistical methods. 1991. URL: https://doi.org/10.1177/001316446902900247.

[95] S. Wilk, S. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *The Annals of Mathematical Statistics*, 9(1):60–62, 2019. URL: http://cdsweb.cern.ch/record/2677054.

[96] ATLAS collaboration. Luminosity determination in pp collisions at $\sqrt{s}$=13 tev using the ATLAS detector at the LHC. *ATLAS-CONF-2019-021*, 2019. URL: http://cdsweb.cern.ch/record/2677054.

[97] A. Caldwell, D. Kollar, , and K. Kroninger. BAT - The Bayesian Analysis Toolkit. *Computer Physics Communications*, 180, 2009. URL: https://doi.org/10.1016/j.cpc.2009.06.026.

[98] G. Choudalakis. On hypothesis testing, trials factor, hypertests and the BumpHunter. *PHYSTAT2011 Proceedings*, 2011. URL: https://arxiv.org/abs/1101.0390.

[99] E. Gross and O. Vitells. Trial factors for the look elsewhere effect in high energy physics. *European Physics Journal C*, 70, 2010. URL: https://doi.org/10.1140/epjc/s10052-010-1470-8.

[100] E. Gross and O. Vitells. Estimating the significance of a signal in a multi-dimensional search. *Astroparticle Physics*, 35, 2011. URL: https://doi.org/10.1016/j.astropartphys.2011.08.005.

[101] ATLAS Collaboration. Data quality information for data. 2018. URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/RunStatsPublicResults2010.

[102] ATLAS Collaboration. Total integrated luminosity and data quality in 2018. 2018. URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/.

[103] ATLAS Collaboration. Atlas b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s}$=13 TeV. *The European Physical Journal C*, 79(970), 2019. URL: https://doi.org/10.1140/epjc/s10052-019-7450-8.

[104] K. et al. Becker. Recommended predictions for the boosted-Higgs cross section. *tech. rep. LHCHXSWG-2019-002*, 2019.

[105] ATLAS Collaboration. Search for new phenomena in dijet mass and angular distributions from pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector. *Physics Letter B*, 754:302–322, 2016. URL: https://doi.org/10.1016/j.physletb.2016.01.032.

[106] ATLAS Collaboration. Measurements of $t\bar{t}$ differential cross-sections of highly boosted top quarks decaying to all-hadronic final states in pp collisions at $\sqrt{s}$ = 13 TeV using the ATLAS detector. *Physical Review*, 98(012003), 2018. URL: https://doi.org/10.1103/PhysRevD.98.012003.

[107] ATLAS Collaboration. Measurements of top-quark pair differential and double-differential cross-sections in the l+jets channel with pp collisions at $\sqrt{s}$ = 13 TeV using the ATLAS detector. *The European Physical Journal C*, 79(1028), 2019. URL: https://doi.org/10.1140/epjc/s10052-019-7525-6.

[108] H. Yang. Xml analytic workspace builder twiki page. 2019. URL: https://twiki.cern.ch/twiki/bin/view/Sandbox/HongtaoYangSandbox.

[109] W. et al. Verkerke. The roofit toolkit for data modeling. *Talk from Computing in High Energy and Nuclear Physics (CHEP03)*, 2003. URL: https://arxiv.org/abs/physics/0306116v1.

[110] K. et al. Cranmer. HistFactory: A tool for creating statistical models for use with RooFit and RooStats. *CERN-OPEN-2012-016*, 2012. URL: https://cds.cern.ch/record/1456844/files/CERN-OPEN-2012-016.pdf.

[111] R. Brun and F. Rademakers. ROOT - an object oriented data analysis framework. *Nuclear Instruments and Methods in Physics Research Section A*, 389(1-2):81–86, 1997. URL: https://doi.org/10.1016/S0168-9002(97)00048-X.

[112] ATLAS Collaboration. Quickfit. URL: https://gitlab.cern.ch/atlas_higgs_combination/software/quickFit.

[113] G. Cowan, K. Cranmer, E. Gross, and O. Vitells. Asymptotic formulae for likelihood-based tests of new physics. *The European Physical Journal C*, 71(1554), 2011. URL: https://doi.org/10.1140/epjc/s10052-011-1554-0.

[114] ATLAS Jet/Etmiss Group. Large-R (R=1.0) jet energy scale uncertainties for full run 2 analyses (release 21) which points to the preliminary uncertainties for JMR given on this twiki. 2018. URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/JetUncertaintiesRel21Moriond2018LargeR.

[115] ATLAS Jet/Etmiss Group. Large-R (R=1.0) jet energy scale uncertainties for full run 2 analyses (release 21). 2019. URL: https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/JetUncertaintiesRel21Summer2019LargeR.

[116] ATLAS Jet/Etmiss Group. Talk pointed to on JMR recommendation TWiki to source of 20 percent relative JMR error. 2008. URL: https://indico.cern.ch/event/441619/contributions/1099656/attachments/1148550/1647846/Jet_Substructure_Status_09_02_2015-Updated.pdf.

[117] G. Cowan, K. Cranmer, E. Gross, and O. Vitells. Asymptotic formulae for likelihood-based tests of new physics. *The European Physical Journal C*, 71(1554), 2010. URL: https://doi.org/10.1140/epjc/s10052-011-1554-0.

[118] J. M. et al. Lindert. Precise predictions for V+jets dark matter backgrounds. *CERN-LPCC-2017-02*, 2017. URL: http://inspirehep.net/record/1609518/files/ATLAS-CONF-2017-041.pdf?version=1.