



UNIVERSIDAD DE BUENOS AIRES
Facultad de Ciencias Exactas y Naturales
Departamento de Física

Medición de la producción de pares de jets a 13 TeV y uso de redes neuronales adversarias para identificar jets- W

Tesis presentada para optar al título de Doctor de la Universidad de Buenos Aires
en el área de ciencias físicas

Gino Marceca

Director de tesis: Ricardo Piegai

Consejero de Estudios: Gustavo Otero y Garzón

Lugar de trabajo: Departamento de Física, FCEyN, UBA

Buenos Aires, Febrero 2019



Agradecimientos

Quiero agradecer en primer lugar a mi director Ricardo Piegaiá por brindarme la oportunidad de ejercer con el este doctorado y poder participar de un experimento tan emocionante y relevante dentro de la física de partículas que es la Colaboración ATLAS. Es invaluable la cantidad de tiempo y dedicación que Ricardo invirtió en mi formación, y lo que principalmente valoro es que siempre ha estado abierto y bien predispuesto a escuchar mis dudas.

Quiero agradecer también a Gustavo Otero y Garzón por su desempeño para que los viajes a CERN resulten posibles, también por las varias reflexiones que he tenido con el cuando me preguntaba "¿cómo estás?".

Agradezco enormemente a mis amigos, Julián, Joaco, Feliciano y Federico por su contención en mis momentos de estrés y de exigencia.

A mis compañeros de trabajo, Flor, Ro, Jona y Sabri les agradezco por la buena disposición que tuvieron conmigo para trabajar en equipo, también les agradezco de corazón por la esencial compañía que fueron cuando estuvimos en el CERN.

Por último y fundamental, agradezco profundamente a mi familia por su calidez, apoyo económico y moral. Esta tesis se la dedico a ellos.

Medición de la producción de pares de jets a 13 TeV y uso de redes neuronales adversarias para identificar jets- W

Resumen

En esta tesis se presenta la medición de la sección eficaz de masa invariante entre pares de jets producidos en colisiones protón-protón a una energía de centro de masa de 13 TeV. Los datos fueron colectados por el detector ATLAS en el Gran Colisionador de Hadrones del Laboratorio CERN durante el año 2015. Las mediciones de la sección eficaz fueron comparados cuantitativamente con las predicciones teóricas del Modelo Estándar (ME) a orden siguiente al dominante (NLO) corregidas por efectos no perturbativos. Estos estudios ponen a prueba el ME por posible evidencia de nueva física. A su vez, la producción de jets contiene información acerca de la distribución partónica dentro del protón y de la constante de acoplamiento fuerte α_s , lo que permite mejorar el conocimiento de QCD a una escala de energía nunca antes alcanzada. Los resultados experimentales mostraron estar en acuerdo con las predicciones teóricas, validando por primera vez el ME a 13 TeV en lo que respecta a la producción de pares de jets.

Por otro lado, se desarrollaron técnicas de aprendizaje automático para la identificación de jets provenientes de bosones W en medio de un fondo dominante de jets de QCD, tarea fundamental para la búsqueda de partículas masivas producto de nueva física o para la mejora de precisión de las propiedades del Higgs. El modelo propuesto en esta tesis está basado en redes neuronales adversarias, el cual permite lograr un clasificador no correlacionado con ciertos observables físicos de interés, como la masa del jet. Esto es importante pues ciertas búsquedas de nueva física en ATLAS son sensibles a efectos no deseados introducidos por la correlación entre la variable discriminante y la masa del jet, lo que resulta en la reducción de la significancia estadística del análisis. Estudios realizados en simulaciones Monte Carlo muestran mejoras significativas respecto a otros métodos analíticos y de multivariable utilizados tradicionalmente en ATLAS, resultando por lo tanto prometedor para futuras búsquedas de nueva física.

Palabras clave: ATLAS, LHC, Jets, Sección eficaz, Redes neuronales

Measurement of the di-jet production at 13 TeV and use of adversarial neural-networks to identify W -jets

Abstract

The measurement of the dijet production cross-section is measured in proton-proton collisions at a center-of-mass energy of 13 TeV. The data were collected by the ATLAS detector at the Large Hadron Collider at CERN during 2015. The measurement of the cross-sections were compared quantitatively with the Standard Model (SM) theoretical predictions at next-to-leading order (NLO) corrected for non-perturbative effects. These studies test the SM for possible evidence of new physics. In addition, jets production contains information on the distributions of partons inside the proton and also on the strength of their interaction α_s , which allows to extend the knowledge of QCD to an unprecedented energy scale. The experimental results show agreement with the theoretical predictions, validating for the first time the SM dijet production at 13 TeV.

In addition, this Thesis presents the development of machine learning techniques for identification of jets arising from boosted W bosons in a dominated QCD background, an essential task for heavy particles searches beyond SM or to improve the precision of Higgs properties measurements. The technique proposed is based on adversarial neural-networks, which allows to build a non-correlated classifier with physical observables of interest, like the jet mass. Several searches in ATLAS are sensitive to undesirable effects introduced by these correlations, which results in a reduction of the statistical significance of the analysis. Studies based in Monte Carlo simulations showed significant improvements with respect to other analytical and multivariate techniques previously used in ATLAS, providing an encouraging first look for its application in future physics searches.

Keywords: ATLAS, LHC, Jets, Cross-section, Neural-Networks

Contents

1	Introduction	8
2	Theoretical Aspects	10
2.1	The Standard Model	10
2.2	Theoretical QCD Predictions	11
2.3	Monte Carlo Simulation	14
2.4	Detector simulation	15
3	The LHC and the ATLAS Detector	16
3.1	The Large Hadron Collider	16
3.2	The Long Journey of a Proton	16
3.3	Luminosity and pile-up	18
3.4	The ATLAS Detector	19
3.4.1	Calorimeter	22
3.4.2	The Muon Spectrometer	23
3.5	The Trigger System	24
3.5.1	Single Jet Triggers	26
4	Jet Reconstruction and Calibration	28
4.1	Introduction	28
4.2	From Particles to Jets	28
4.2.1	Sequential recombination algorithms	29
4.2.2	Jet Algorithm Inputs	30
4.3	Jet Substructure	31
4.4	Jet calibration	33
4.4.1	Monte Carlo based jet energy calibration	34
4.4.2	Numerical Inversion Technique	35

4.4.3	E and η calibrations	37
4.4.4	Closure Test	38
4.5	Jet systematic uncertainties	38
5	Measurement of Dijet Production at 13 TeV	40
5.1	Introduction	40
5.2	Event Selection and Trigger Analysis Strategy	41
5.3	Data vs. Monte Carlo cross-section comparisons	41
5.4	Data Unfolding	42
5.4.1	The General Unfolding Procedure	44
5.4.2	The Iterative Unsmearing Correction	45
5.4.3	Improvement of the unfolding probability matrix	47
5.4.4	Unfolding bias	47
5.5	Estimation of Statistical Uncertainties: Bootstrapping	48
5.6	Propagation of Systematics Uncertainties	50
5.6.1	Jet Energy Scale	50
5.6.2	Jet Energy Resolution	51
5.6.3	Total Experimental Uncertainty	52
5.7	Theoretical Predictions	53
5.7.1	Next-to-leading-order pQCD calculations	53
5.7.2	Non-perturbative corrections	54
5.7.3	Electroweak corrections	55
5.8	Results	56
5.8.1	Qualitative Comparisons	56
5.8.2	Quantitative Comparisons	57
6	Adversarial Neural Networks for Decorrelating Jet Observables	60
6.1	Introduction	60
6.2	Theoretical Aspects of Machine Learning	62
6.2.1	Mathematical Formulation	62
6.2.2	Why Neural Nets?	65
6.3	Sample preparation	67
6.4	Evaluation Metrics	68
6.4.1	Classification	68

6.4.2	Decorrelation	68
6.5	Adversarial Neural Networks	69
6.5.1	Implementation	71
6.5.2	Feature Engineering	73
6.5.3	Hyperparameters Tuning	74
6.6	Training Procedure	76
6.6.1	Pre-training	76
6.6.2	Joint Optimisation	77
6.7	Results	79
6.7.1	Classification	80
6.7.2	Decorrelation	82
6.7.3	Combined Performance	85
7	Conclusions	87
A	The Energy Dependence of the Jet Response	90
B	Smoothing Statistical Fluctuations in Systematic Uncertainties	93
C	Statistical Test of Agreement between Theory and Data	95
D	Decorrelated Jet Tagging Algorithms	98
D.1	Designed decorrelated taggers	98
D.2	k -NN Regression	99
D.3	Convolved Substructure	100
D.4	Adaptive boosting for uniform efficiency	101

Chapter 1

Introduction

On 20 May 2015, the ATLAS detector recorded for the first time proton-proton (pp) collisions at a centre of mass energy of 13 TeV delivered by the Large Hadron Collider (LHC). The proton collisions at 13 TeV set a new energy record, overcoming the past Run 1 at 8 TeV in 2012, and so extending the exploration of quantum chromodynamics (QCD) at scales never reached before. Precision measurements of strong interactions are crucial in understanding physics at hadron colliders, in addition, QCD provides one of the main backgrounds to many New Physics searches; furthermore, it is also through tests of QCD that New Physics may be discovered in a direct way.

The observable signatures of quarks and gluons produced at the LHC are hadronic particle showers called “jets”. By measuring jets in particle detectors, one can make predictions on the nature of the underlying parton-parton interaction and thus test the Standard Model (SM) theory. At hadron colliders, a relevant QCD jet observable is the dijet spectrum, related to the high-momentum-transfer $2 \rightarrow 2$ scattering of partons inside the colliding protons. In this kind of processes, the energy of the jet (in the partonic centre-of-mass frame) is closely related to that of the parton in the proton that underwent a hard scattering and the dijet spectrum therefore contains information on the distributions of partons inside the proton and also on the strength of their interaction α_s [1–4]. This Thesis presents, for the first time, measurements of the dijet cross-sections in proton-proton collisions at $\sqrt{s} = 13$ TeV centre of mass energy, using data collected by ATLAS in 2015 and corresponding to an integrated luminosity of 3.2 fb^{-1} . Next-to-leading-order (NLO) perturbative QCD (pQCD) predictions calculated using several parton distribution function (PDF) sets, corrected for electroweak and non-perturbative effects, are quantitatively compared to the measured results.

On the other hand, the identification of jets coming from boosted W bosons against the

large background QCD is of fundamental importance to searches for new physics. The use of multivariate analysis techniques such as boosted decision trees or neural-networks have been shown to be highly performant for the classification task. Nevertheless, these approaches are subjected to the presence of strong non-linear correlations between the discriminant variable and the jet mass observable for QCD jets, resulting in a reduction of discovery significances. A novel technique using adversarial neural-networks is presented as a way to build a mass-decorrelated classifier and applied to the identification of jets arising from W bosons.

In order to extract theoretical information from reconstructed jets it is necessary to calibrate them, that is to correct their momentum to the scale of the matching particle jets. The suite of steps necessary to achieve this calibration is discussed, focusing particularly on the jet energy scale calibration based in Monte Carlo simulation (MC-JES) which was one of the main tasks developed during this PhD.

The Thesis is organized as follows: Chapter 2 describes the SM theoretical framework, with emphasis in the theory of the strong interactions and the aspects that are important for the understanding of the hadronic final states in hadronic collisions and how they are simulated. The LHC and the ATLAS detector are described in Chapter 3, as well as the experimental conditions during the 2015 data taking. Chapter 4 concerns jet reconstruction and calibration, focusing particularly on the MC-JES calibration. On Chapter 5 and 6 the two analyses performed as a main contributor are presented. The first one is about the measurement of the dijet cross-sections at 13 TeV. The second one, relates to the performance studies using adversarial neural-networks for W -jet identification. Finally, Chapter 7 presents the conclusions.

Chapter 2

Theoretical Aspects

This chapter presents a brief overview of the Standard Model and theoretical QCD predictions, Sections 2.1 and 2.2. The Monte Carlo tools used to simulate the QCD processes and the interaction with the detector are discussed in Sections 2.3 and 2.4.

2.1 The Standard Model

The Standard Model (SM) is a quantum field theory that describes the behavior of all experimentally observed particles under the influence of the electromagnetic, weak and strong forces. Mathematically speaking, its formulation is based on group theory that obeys the global Poincarè symmetry postulated for all relativistic quantum field theories (translational, rotational symmetries and the inertial reference frame invariance), in addition to the local $SU(3) \times SU(2) \times U(1)$ gauge symmetry (internal symmetry). Roughly, the three factors of the gauge symmetry give rise to the three fundamental interactions, described as an exchange of particles (gauge bosons) between the objects affected, such as the gluon for the strong interaction, described by *Quantum Chromodynamics* (QCD), the W and Z bosons for the unified electroweak force, and the photon for the electromagnetic force, modeled by *Quantum Electrodynamics* (QED), respectively. These particles are known as the force carriers or messenger particles and all have spin 1, therefore, all known gauge bosons are vector bosons. In addition to gauge bosons, there is a scalar boson called the Higgs field. The quantum excitation of this field (the Higgs particle) was predicted to exist in the 1960s and confirmed experimentally for the first time in 2012 by the ATLAS and CMS experiments at the LHC [5]. The confirmation of the existence of the Higgs field completed the last missing piece of the SM. Unlike other known fields such as the electromagnetic field, it has a non-zero expectation value in the

vacuum and explains why some fundamental particles have mass, despite that the symmetries controlling their interactions impose that they should be massless. It also resolves several other long-standing puzzles, such as the reason for the extremely short range of the weak force [6].

The fundamental building blocks of matter described by the SM are fermions with spin 1/2:

- six leptons (and their antiparticles), organized in three families,

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix} \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix} \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}$$

- and six quarks (and their antiparticles) each in three colors, organized in three families,

$$\begin{pmatrix} u \\ d \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix}$$

The three families are ordered from lower (1st generation) to higher (3er generation) mass values. These particles are considered point-like, as there is no evidence of any internal structure of leptons or quarks (q) to date. The six types of quarks are also known as the six quark flavors. Collectively, the u (up), d (down), and s (strange) quarks are frequently referred to as light quarks. The heaviest quark of the Standard Model, the quark t (top), was the last to be found [7, 8]. The electric charge Q of quarks take fractional values, i.e. $+2/3$ for quarks u , c and t and $-1/3$ for quarks d , s and b ; yet they are only observed as the integer charge combinations of three quarks (baryons) or a quark and an antiquark (mesons)¹. Quarks and gluons are collectively called partons and are the fundamental actors of the strong interactions.

Figure 2.1 shows the elemental table which summarizes the SM.

2.2 Theoretical QCD Predictions

Renormalization of fields appearing in the SM Lagrangian is a necessary technique which allows to treat infinities arising from momentum space integral divergences in computing Feynman diagrams. Fortunately, the SM, in particular QCD, as any Gauge theory, is *renormalizable*, in the sense that all the *ultraviolet* divergences, can be absorbed into redefinition of few physical parameters. Because of renormalization and also due to the *non-Abelian* property of QCD, the coupling constant of the strong interaction, α_s , decreases as the interaction energy increases, a phenomenon called the *running* of the coupling. As a result, in high-energy interactions,

¹Although stable particles are regularly made of 3 quarks (baryons) or a quark and an antiquark (mesons), evidence for exotic baryons consisting of 5 quarks has recently been observed by the LHCb Collaboration. [9]

mass →	≈2.3 MeV/c ²	≈1.275 GeV/c ²	≈173.07 GeV/c ²	0	≈126 GeV/c ²
charge →	2/3	2/3	2/3	0	0
spin →	1/2	1/2	1/2	1	0
	u up	c charm	t top	g gluon	H Higgs boson
QUARKS	≈4.8 MeV/c ²	≈95 MeV/c ²	≈4.18 GeV/c ²	0	
	-1/3	-1/3	-1/3	0	
	1/2	1/2	1/2	1	
	d down	s strange	b bottom	γ photon	
	0.511 MeV/c ²	105.7 MeV/c ²	1.777 GeV/c ²	91.2 GeV/c ²	
	-1	-1	-1	0	
	1/2	1/2	1/2	1	
	e electron	μ muon	τ tau	Z Z boson	
LEPTONS	<2.2 eV/c ²	<0.17 MeV/c ²	<15.5 MeV/c ²	80.4 GeV/c ²	
	0	0	0	±1	
	1/2	1/2	1/2	1	
	ν_e electron neutrino	ν_μ muon neutrino	ν_τ tau neutrino	W W boson	
					GAUGE BOSONS

Figure 2.1: Elemental table of particles of the SM.

which means at short lengths ($\ll$ size of the proton), the coupling constant in QCD becomes sufficiently small to warrant a perturbative series expansion. This is known as *asymptotic freedom* and makes quantitative predictions for the strong interaction possible in particular at the energy scale probed by hadron colliders. On the contrary, partons are coupled together more strongly as the distance between them increases. This effect, known as *confinement*, “explains” why quarks and gluons are not visible in their own right, but become trapped into color-neutral hadrons (baryons or mesons). This process is known as *hadronization*, which is a non-perturbative effect and so cannot be computed from first principles.

In perturbative QCD (pQCD), each order of the expansion corresponds to an additional power in the coupling constant. By considering all possible Feynman diagrams for a given order of perturbation theory, all the terms in the calculation summed together give a definite *cross-section* at a fixed-order. In this context, leading-order diagrams (LO) are the first ones which appears in the perturbation series, without internal loops. The complexity of the process determines the precision of the calculation that has to be performed. For inclusive parton production, calculations are typically performed at next-to-leading order (NLO), although NNLO QCD predictions have recently become available for the dijet cross-sections at 7 TeV [10].

Due to the fact that the LHC does not produce simple parton-parton interactions, but instead collisions of hadrons, one may wish a tool to extend perturbative parton calculations to proton-proton collisions. Fortunately, the factorization theorem [11] allows us to account for

this. This theorem expresses the total cross section of two hadrons interaction as a weighted combination of parton cross-sections interaction. The weighting factors $f_i(x)$ are the parton distribution functions (PDFs), where $f_i(x)dx$ gives the differential number of partons of type i carrying a fraction of the total hadron momentum between x and $x + dx$. The hadronic cross-section, at some energy Q^2 that characterizes the interaction, can be written as:

$$\sigma(P_1, P_2) = \sum_{i,j} \int dx_1 dx_2 f_i(x_1, \mu_F^2) f_j(x_2, \mu_F^2) \hat{\sigma}_{ij}(p_1, p_2, \alpha_s(\mu_r^2), Q^2/\mu_r^2, Q^2/\mu_f^2). \quad (2.1)$$

Here, P_1 and P_2 are the momenta of the two incoming hadrons, x_1 and x_2 are the momentum fractions carried by the interacting partons, and $p_1 = x_1 P_1$, $p_2 = x_2 P_2$ are the interacting parton momenta. The partonic cross section $\hat{\sigma}_{ij}$, corresponding to the interaction of partons i and j , is calculated at a fixed order in α_s , which is evaluated at some renormalization scale, μ_R . The total cross section is obtained by summing over all possible parton flavors and integrating over all possible momentum fractions. The parton distribution functions, f_i and f_j , are evaluated at a factorization scale, μ_F , which can be thought of as the scale that separates short-distance, perturbative physics, from long-distance, non-perturbative physics. They are usually both taken to be equal, $\mu_F = \mu_R = \mu$, chosen at the typical scale Q^2 of the process, in order to minimize the contribution of (uncalculated) higher order terms. In the limit of a complete perturbation expansion, the hadronic cross-section would be exact and would not depend on μ_F, μ_R . Such dependence is assigned as a theoretical uncertainty at a given order expansion. Figure 2.2 shows the dijet cross-section as a function of μ_R for different μ_F values. As it can be seen, both dependences decrease as the precision of the calculation increases, which results in a significant reduction of the theoretical uncertainty of NNLO with respect to NLO [10].

The NLO predictions used in this thesis were calculated using NLOJET++ 4.1.3 [12] for various NLO PDF sets and various values of the renormalisation and factorization scales.

In order to be able to compare with the measurement, QCD perturbative calculations need to be corrected by non-perturbative (NP) and electroweak (EW) effects. NP contributions account for hadronization and underlying event (UE), additional parton interactions which are not involved in the hard scattering process.

Precise measurements of hard scattering cross-sections are used in this way not only to test predictions but also to fit (in case of agreement between data and theory) the dependence of the PDFs on x and Q^2 [13], which can be obtained only through experimental results. Cross-section measurements can also be used to determine the strength of the interaction α_s [3]. Both, the fit of the PDFs and the determination of α_s , exceed the scope of the studies developed in this measurement.

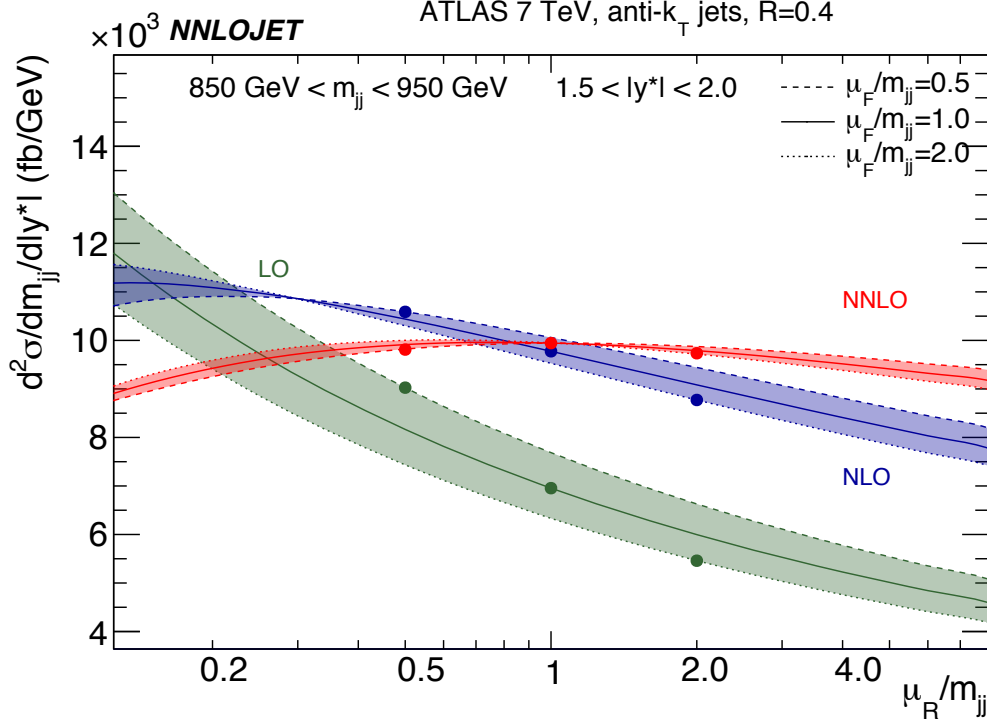


Figure 2.2: The mass dijet cross section for 7 TeV pp collisions evaluated at a specific kinematic range as a function of μ_R/m_{jj} LO (green), NLO (blue) and NNLO (red). The thickness of the bands shows the variation due to the factorization scale.

2.3 Monte Carlo Simulation

When two partons collide, they can undergo a large angle (“hard”) scattering generating two (or more) final-state partons at large momentum. This process is calculated from first principles at some order on the matrix element. The semi-empirical view of the rest of the process is as follows. Partons at these states tend to radiate more partons nearly collinear with its parents. This *parton shower* evolves, and lose momentum until the non-perturbative QCD regime is reached. At this point, due to confinement, partons hadronize, leading to a “spray” of colorless hadrons forming a cone structure that we label a *jet*.

Parton showering of jet formation exits the region of validity for pQCD, and must be therefore described by phenomenological models, to account for the length of time when showering occurs and the combination of colored partons into bound states of colorless hadrons. One example is the Lund string model [14], which is the one implemented in ATLAS *event generators*.

Event generators are based in Monte Carlos (MC) calculations and use LO (NLO) perturbative calculations of matrix elements for $2 \rightarrow 2$ ($2 \rightarrow 3$) processes, relying on the parton shower plus hadronization to produce the colorless-particle final state. The most common used parton shower generator is *Pythia*. This generator utilizes LO pQCD matrix elements, along

with a leading-logarithmic parton shower, an UE simulation with multiple parton interactions, and the Lund string model for hadronisation. The nominal Pythia samples used in this thesis were produced using a set of tuned parameters called the A14 tune [15] and the NNPDF 2.3 LO PDF [16] set. Given the good agreement between data and Pythia, and because of the vast use through the Collaboration, this generator was used as the default for both the measurement and the neural-networks performance studies. Others general purpose MC generators were used as well in the measurement such as Herwig++ [17], Powheg and Sherpa [18]. The last two are computed at NLO level in the matrix element. The specific use of each MC will be stated clearly throughout each chapter.

The MC samples are overlaid with additional Pythia simulated pp collisions (pile-up) to roughly match the data pile-up conditions with a mean number of 24 collisions per bunch crossing for the 2015 data-taking. A formal definition of pile-up is presented in Section 3.3.

2.4 Detector simulation

To have simulated events at the same level that one might observe with the detector, the passage of particles through matter and digitalisation must be simulated. The particle output from event generators is passed through a detector simulation model, resulting in the same physical signals one might observe in real data. The detailed simulation of the ATLAS detector is based on the Geant4 toolkit [19]. Geant4 is a flexible program designed to handle complex geometries and environments. It simulates the magnetic fields and all the particle interaction processes with matter such as ionization, bremsstrahlung, photon conversions, multiple scattering, scintillation, absorption and transition radiation. The software is easily adaptable to different conditions at a given time and so it contains updated information regarding dead channels and misalignments.

Finally, the simulated events are processed using the same reconstruction algorithms and analysis filters as used in real data.

Chapter 3

The LHC and the ATLAS Detector

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [20–22] is the world’s largest and most powerful particle accelerator. It is a proton-proton (pp) synchrotron located at CERN Laboratory, just outside the city of Geneva (Switzerland), approximately 100 m underground. It is designed to collide bunches of up to $\sim 10^{11}$ protons every 25 ns at a center of mass energy of $\sqrt{s} = 14$ TeV. However, in order to not delay the delivery of particle collisions for physics research for the Run 2 period, scheduled for the beginning of 2015, the decision was to operate at 13 TeV [23]. The experiments analyzing the collisions produced by the LHC are distributed around the 27 km ring mostly opposite each other. The ATLAS experiment is closest to the main CERN site and is located opposite to the other general purpose detector, CMS. ALICE and LHCb experiments are specific purpose detectors. The former is designed to investigate heavy ion collisions; the latter, to investigate rare decays of b -mesons. The layout of these four experiments along the LHC ring is shown in Figure 3.1.

3.2 The Long Journey of a Proton

Proton beams are formed, before insertion into the main LHC ring, using a succession of machines that accelerate particles to increasingly higher energies. Each machine boosts the energy of a beam of particles, before injecting the beam into the next machine in the sequence. In the LHC, the last element in this chain, particle beams are accelerated up to the record energy of 6.5 TeV per beam. Most of the other accelerators in the chain have their own experimental halls where beams are used for experiments at lower energies.

The proton source is a simple bottle of hydrogen gas. An electric field is used to strip hydrogen atoms of their electrons to yield protons. Linac 2, the first accelerator in the chain, accelerates the protons to the energy of 50 MeV. The beam is then injected into the Proton Synchrotron Booster (PSB), which accelerates the protons to 1.4 GeV, followed by the Proton Synchrotron (PS), which pushes the beam to 25 GeV. Protons are then sent to the Super Proton Synchrotron (SPS) where they are accelerated to 450 GeV.

The protons are finally transferred to the two beam pipes of the LHC. The beam in one pipe circulates clockwise while the beam in the other pipe circulates anticlockwise. It takes 4 minutes and 20 seconds to fill each LHC ring, and 20 minutes for the protons to reach their maximum energy of 6.5 TeV. Beams circulate for many hours inside the LHC beam pipes under normal operating conditions. The two beams are brought into collision inside the four detectors: ALICE, ATLAS, CMS and LHCb, where the total energy at the collision point is equal to 13 TeV.

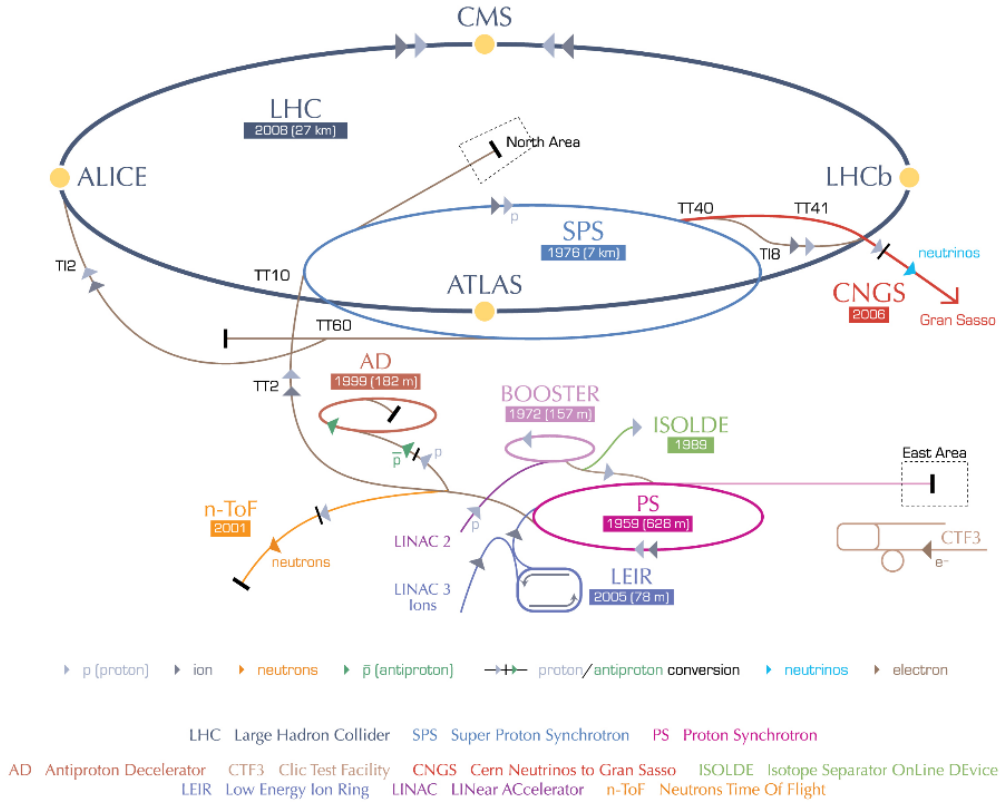


Figure 3.1: The CERN accelerator complex, showing the injection system, along with each component's date of construction, and the placement of the four main experiments.

3.3 Luminosity and pile-up

The rate of events produced by the colliding beams depends on the luminosity of the collisions, which is a measure of the number of events per second per unit cross section, typically measured in units of cm^2s^{-1} . The number of events (N) of a particular process, then, is given by the product of the integrated luminosity, $\int L dt$, and the cross-section of the process, σ_{event} :

$$N = \sigma_{\text{event}} \int L dt$$

The integrated luminosity is typically quoted in units of inverse picobarns, $\text{pb}^{-1} = 10^{-36}\text{cm}^2$. In order to measure processes with very little cross sections a very high luminosity is required.

The delivered luminosity can be written as [24]:

$$L = \frac{n_b f_r n_1 n_2}{2\pi \Sigma_x \Sigma_y}$$

where n_b is the number of colliding bunch pairs, n_1 and n_2 are the bunch populations (protons per bunch) in beam 1 and beam 2 respectively (together forming the bunch charged product), f_r is the machine revolution frequency, and Σ_x and Σ_y are the width and the height of the proton beams.

The number of protons per bunch, the number of bunches per beam, and the revolution frequency are all set by the beam operators. The widths of the proton beams are measured in a process known as a Van der Meer (*vdM*) scan [25]. In a *vdM* scan, the beams are separated by steps of a known distance. The collision rate is measured as a function of this separation, and the width of a gaussian fit to the distributions yields the width of the beams in the direction of the separation.

The total integrated luminosities provided by the LHC and recorded by ATLAS in 2015 are shown in Figure 3.2. These events form the dataset analyzed in the measurement presented in this thesis. By means of the beam-separation or *vdM* scans, as well as other techniques to measure the bunch charged product, the ATLAS Collaboration has determined that the uncertainty on the 2015 luminosity measurement was $\delta L = 2.1\%$ [26].

Due to the cross-section for interaction and the large number of protons per bunch, the possibility to observe multiple *pp* interactions per bunch crossing increases proportionally. This phenomenon, referred to as “pile-up”, can really occur in two distinct forms. The first form is the presence of multiple *pp* collisions (different from the interaction of interest) in the same bunch crossing, referred to as “in-time” pile-up. The second form of pile-up takes place due to electronic integration times within the detector. Certain detector components are actually

sensitive to multiple bunch crossings due to the long electronic signals generated in the response to energy depositions or charge collection. One or more pp collisions in neighboring bunch-crossings can then affect the measurement. This form of pile-up is referred to as “out-of-time” pile-up. The total number of reconstructed primary vertices with 2 or more tracks, N_{PV} , provides a measure of in-time pile-up activity in each event. On the other hand, the mean number of interactions per crossing $\langle\mu\rangle$ is sensitive to both in-time and out-of-time pile-up, which corresponds to the mean of the poisson distribution of the number of interactions per crossing calculated for each bunch: $\mu = L_{\text{bunch}} \times \sigma_{\text{inel}}/f_r$, where L_{bunch} is the per bunch instantaneous luminosity, σ_{inel} is the inelastic cross section which we take to be 80 mb for 13 TeV collisions, and f_r is the LHC revolution frequency. The conditions for the 2015 data were $\langle\mu\rangle \sim 13.5$ and $N_{PV} \sim 20$.

3.4 The ATLAS Detector

The ATLAS (A Toroidal LHC Apparatus) detector of 7000 tones is one of the two general purpose particle detectors built for probing pp collisions at LHC. At 13 TeV center-of-mass energy, the LHC interactions result in high particle multiplicity, requiring fine detector granularity; and, particle production at small angles with respect to the interaction line, requiring

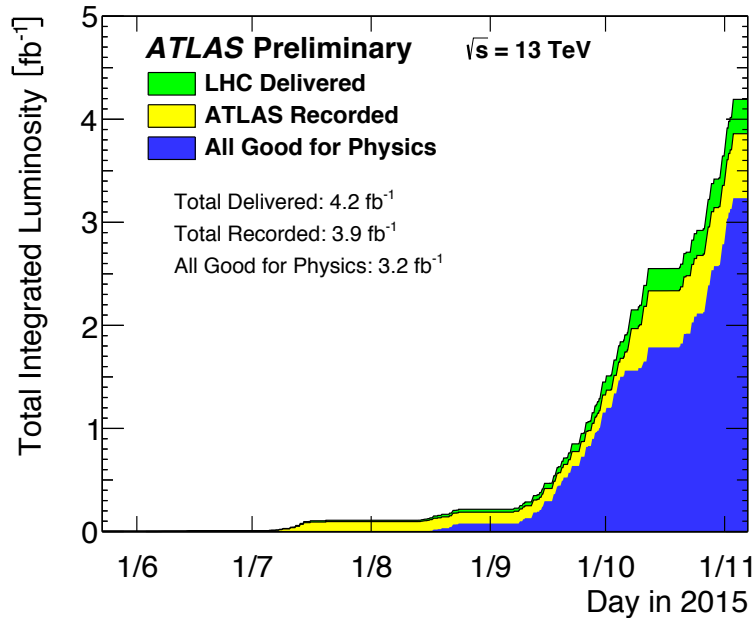


Figure 3.2: Cumulative luminosity versus time delivered to ATLAS (green), recorded by ATLAS (yellow), and certified to be good quality data (blue) during stable beams for pp collisions at 13 TeV centre-of-mass energy in 2015.

large detector angular coverage. To achieve these performance goals, a design consisting of multiple detector sub-systems with cylindrical symmetry around the incoming beams is used as shown in Figure 3.3. Closest to the interaction point the inner tracking detector is placed, providing charged particle reconstruction. The magnet configuration comprises a thin superconducting solenoid surrounding the inner detector cavity, and three large superconducting toroids (one barrel and two end-caps) arranged with an eight-fold azimuthal symmetry around the calorimeters. This fundamental choice has driven the design and size (44 m in length and 25 m in height) of the rest of the detector. Outside the solenoid, a calorimeter system performs electrons, photons, taus, and jets (defined in Section 4) energy measurements. Finally, the calorimeter is surrounded by the muon spectrometer where an array of muon drift chambers perform muon identification and momentum measurements. A detailed description of its sub-systems can be found on its technical design report [27, 28].

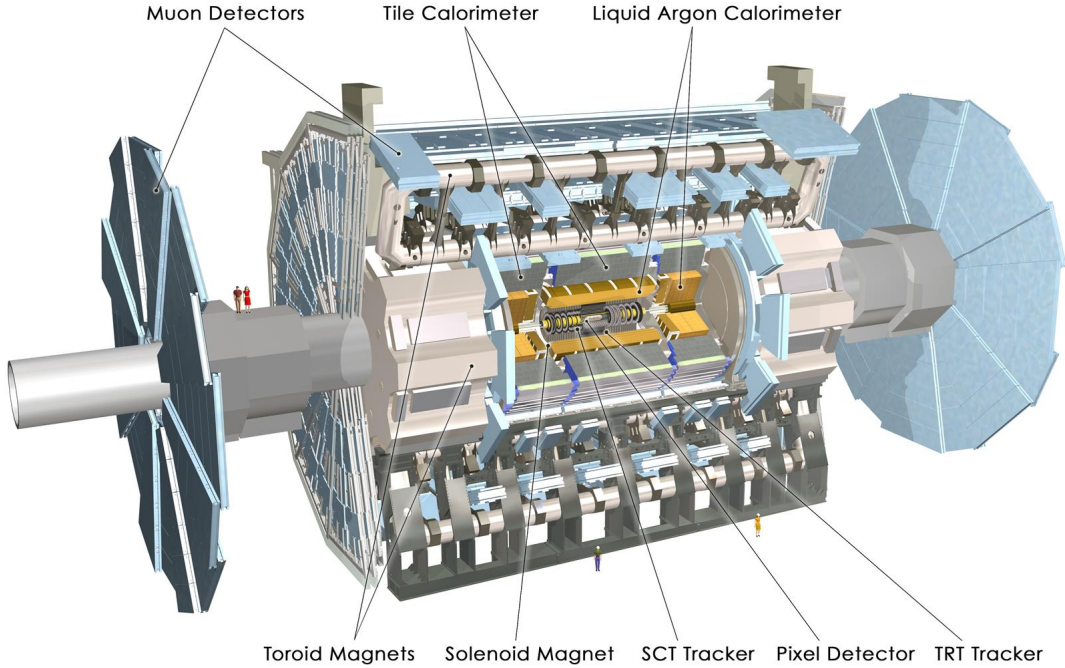


Figure 3.3: The ATLAS Detector.

The ATLAS detector coordinate system is used to describe the position of particles as they traverse these subdetectors. It is a right-handed coordinate system, with z pointing along the beam direction, positive x pointing toward the center of the LHC ring, and positive y pointing up. The $x - y$ plane is referred to as the transverse plane, and the z direction as the longitudinal direction. The azimuthal angle ϕ is measured as usual around the beam axis, and the polar angle θ is the angle from the beam axis. In experimental particle physics, the

pseudorapidity η is commonly used as the polar angle instead of θ . They relate each other as following $\eta = -\ln(\tan \frac{\theta}{2})$. It can be shown that by doing a massless approximation ($m \ll |p|$), which is valid given our energy regime, η equals the rapidity $y \equiv \frac{1}{2} \ln \left(\frac{E+p_L}{E-p_L} \right)$, where p_L is the momentum across the beam line. It can be seen that the rapidity y (or pseudorapidity in the massless approximation) are Lorentz invariant under boosts along the longitudinal axis. A measurement of a rapidity difference Δy between particles (or $\Delta \eta$) is hence not dependent on the longitudinal boost of the reference frame. This is an important feature, where the colliding partons carry different longitudinal momentum fractions, which means that the rest frames of the parton-parton collisions will have different longitudinal boosts. This justifies the use of η instead of θ as the polar angle representation.

Regions of low η are referred to as “central”, and regions of high η are referred to as “forward”. The transverse momentum p_T is defined as the momentum projected in the $x - y$ plane. The distance ΔR in the pseudorapidity-azimuthal angle space is defined as $\Delta R = \sqrt{\Delta \eta^2 + \Delta \phi^2}$.

Inner Tracking System

The inner-detector system (ID), providing charged particle tracking, is immersed in a 2T axial magnetic field, generated by the central solenoid. It consists of high-granularity silicon pixel detector closest to the interaction point followed by the silicon microstrip tracker (SCT). These silicon detectors, covering the range $|\eta| < 2.5$, are complemented by the transition radiation tracker (TRT), which enables radially extended track reconstruction up to $|\eta| = 2.0$. The positions of the registered hits are combined to form tracks, with the radius of curvature of the tracks (caused by the presence of the magnetic field) providing a measurement of the particle’s transverse momentum.

The pixel detector, SCT, and TRT sensors are arranged on concentric cylinders around the beam axis, known as barrel layers, and on disks perpendicular to the beam at either end of the barrel, known as end-caps. The overall layout of the inner detector is shown in Figure 3.4

One of the upgrades from Run 1 was the insertion of a new layer between the beam pipe (where particle collisions occur) and the pixel detector. This new sub-detector, known as the Insertable B-Layer (IBL), is located at 0.2 mm from the beam tube and 1.9 mm from the pixel. Thanks to this new fourth layer and dedicated algorithms developed, the performance for reconstructing charged particles has improved with respect to Run 1, resulting in particular in a better tagging efficiency for the decay of beauty quarks (b-tagging), in which I have actively participated [29].

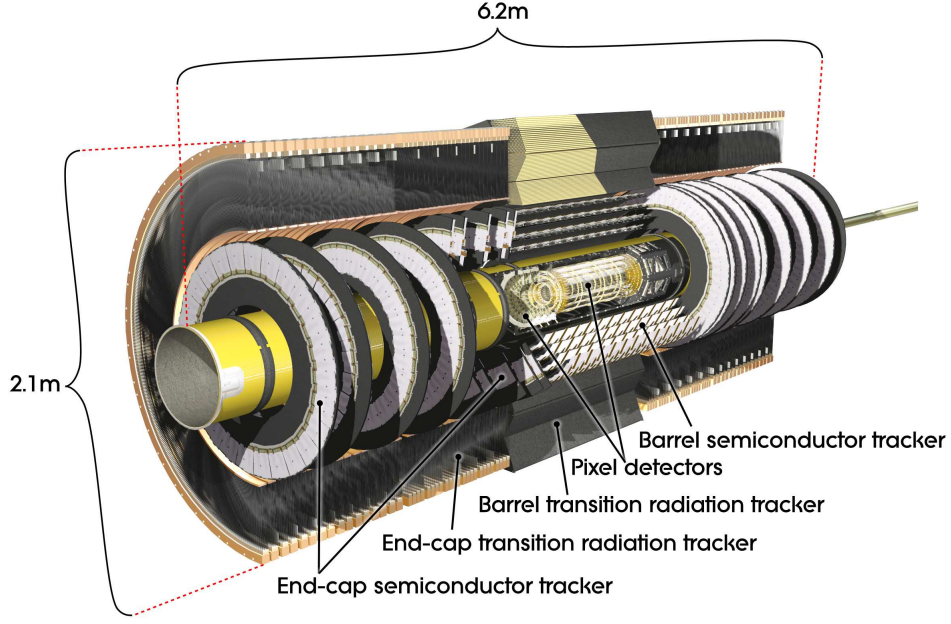


Figure 3.4: Layout of the ATLAS Inner Detector.

3.4.1 Calorimeter

For the studies presented in this thesis, the most important subsystem is the calorimetry, which covers the pseudorapidity range $|\eta| < 4.9$. It consists of an electromagnetic calorimetry (EM) which covers the region $|\eta| < 3.2$, provided by a barrel and endcap high-granularity lead/liquid-argon (LAr) detector, with an additional thin LAr presampler covering $|\eta| < 1.8$ to correct for energy loss in material upstream of the calorimeters. It only measures electrons/positrons and photons. Next to the EM calorimeter, it follows the Hadronic calorimeter, which measures the energy of hadronic particles and partially of muons. It is made of steel/scintillating-tile detectors, segmented into three barrel structures within $|\eta| < 1.7$, and two copper/LAr hadronic endcap calorimeters.

The EM and hadronic calorimeters are sampling calorimeters meaning that they utilize alternating layers of absorber material, composed of heavy atoms that interact with energetic particles and cause them to lose energy, and an active material, that produces a electrical signal in response to the deposited energy.

The calorimeters closest to the beam-line are housed in three cryostats, one barrel and two end-caps. The barrel cryostat contains the EM barrel calorimeter, and the two end-cap cryostats each contain an EM end-cap calorimeter (EMEC), a hadronic end-cap calorimeter (HEC), located behind the EMEC, and a forward calorimeter (FCal) to cover the region closest to the beam. These calorimeters use liquid argon as the active detector medium and need to be maintained at a constant temperature of $\sim 88\text{K}$. Liquid argon has been chosen for its inertness

property, which results in linear behaviour (production of ionization charge as a function of incident charge), stability of response over time and intrinsic radiation-hardness. An illustration of all these components can be found in Figures 3.5 and 3.6.

Each calorimeter is segmented both transverse to the particle direction, to give position information, and along the particle direction, to chart the development of the particle shower. This permits detailed mapping of EM and hadronic showers in the calorimeter, allowing for studies of the internal structure of hadronic jets and partially giving rise to the high resolution measurements of their energy.

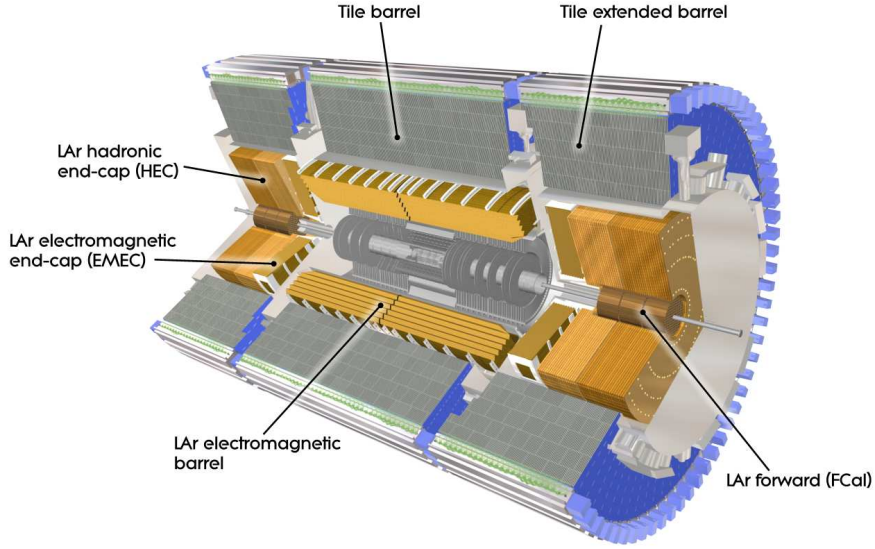


Figure 3.5: Layout of the ATLAS electromagnetic and hadronic calorimeter systems. The total length is ~ 12 m, extending to a maximum radius of 4.25 m.

3.4.2 The Muon Spectrometer

The muon system gives the ATLAS detector its overall shape and imposing nature, as depicted in Figure 3.7. Because muons have much smaller cross section to interact in material than electrons and hadrons, they do not deposit all their energy in the calorimeters. The muon spectrometer was designed to detect muons within $|\eta| < 2.7$. Given that the spectrometer is located outside the calorimeter, it is expected that only muon particles will interact with this detector.

To provide a momentum measurement, the muons trajectories are bent in a toroidal magnetic field. This field is provided by one large barrel toroid and two large end-cap toroids, each toroid consisting of eight coils arranged symmetrically around the beam axis. The toroid system produces a magnetic field that is typically oriented in the ϕ direction and that is measured

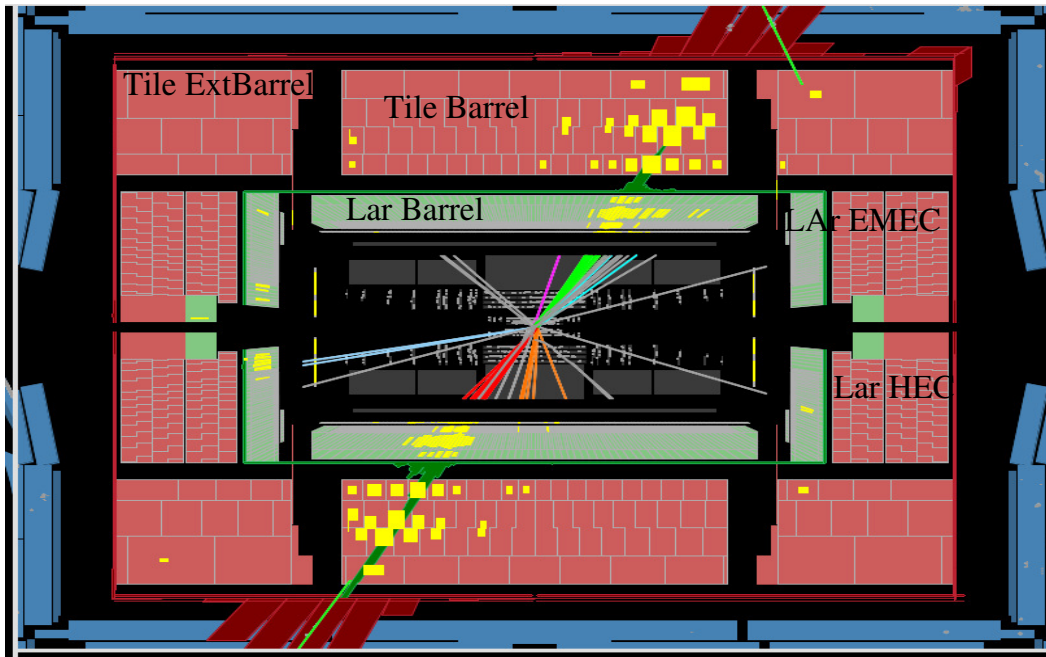


Figure 3.6: Diagram of the ATLAS detector together with a collision example in the plane $z - \eta$. Tracks are observed in different colours in the ID. The energy depositions (yellow) are observed in the em (green) and had (red) calorimeters.

with over 1800 Hall sensors placed through the magnets. Under the influence of this field, muons are deflected in the $z - \eta$ plane and the transverse momentum of the muons is given then by the radius of curvature of the tracks. The trajectory is reconstructed by measuring the time of arrival of an electron cascade produced in drift tube chambers. These consist in cylindrical tubes (cathode) filled with a reactive gas and an anode wire in the middle along its direction. The time of arrival of electrons generated in the gas to the anode allows to reconstruct a spherical surface area where the muon might have passed. A combination of these spherical surfaces from the different tubes make it possible to reconstruct the exact trajectory of the muon. An example of these chambers are the *Monitored Drift Tubes* (MDT), which provides coverage for the more central region.

3.5 The Trigger System

At design luminosity, the LHC delivers approximately 40 million collision events every second. With an average ATLAS event size of ~ 1.5 MB, this is far more information than can be saved into the finite data storage resources available. The goal of the trigger system is to move interesting physics events to permanent storage, while rejecting the vast majority of others non interesting events. Based on the technology and resources restrictions, the bandwidth kept by

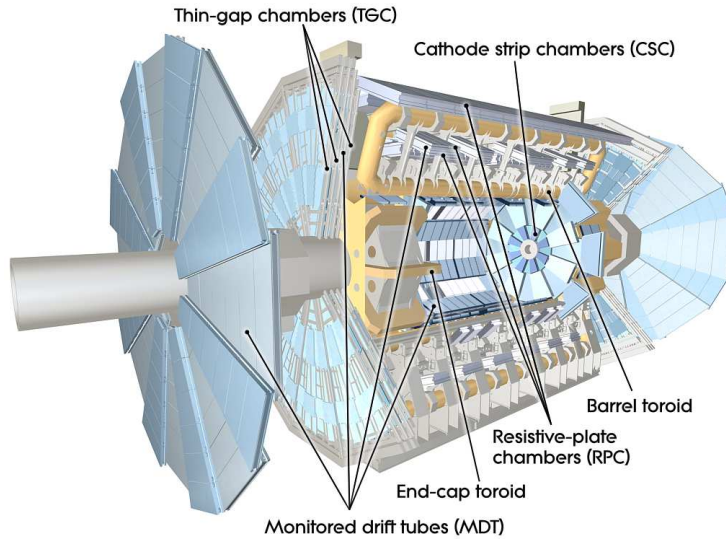


Figure 3.7: The Muon Chamber.

the trigger is approximately 1 kHz from up to 40 MHz of collisions. It is worth to mention that the first long shutdown between LHC Run 1 and Run 2 operations was used to improve the trigger system with almost no component left untouched, in order to face the increased centre of mass energy of 13 TeV and higher luminosity operations.

The current system has two sequential levels; the first level (L1) is a hardware-based system using information from the calorimeter and muon subdetectors, reducing the data rate to ~ 100 kHz. The second *High Level Trigger* (HLT) is a software-based system which operates on the events already selected from the L1 using information from all subdetectors. HLT physics output rate is 1 kHz (400 Hz in Run 1).

For each bunch crossing, the trigger system verifies if at least one of hundreds of conditions (triggers) based on different *signatures* such as electrons, photons, muons, jets, jets with b-flavour tagging (b-jets) or specific B-physics decay modes is satisfied. In this thesis we were particularly interested on jets triggers, which were built at online level using the same recombination algorithm and calibration as in offline described in Section 4.4.

One of the important upgrades for Run 2 was the inclusion of a new topological trigger module (L1Topo), consisting of two FPGA-based (Field-Programmable Gate Arrays) processors. The modules are identical hardware-wise and each is programmed to perform selections based on geometric or kinematic association between trigger objects received from the L1Calo or L1Muon systems. A complete description of the trigger system used in Run 1 and Run 2 can be found here [29, 30].

3.5.1 Single Jet Triggers

The triggers used for the measurement presented in this thesis were based on single jet transverse energy $E_T \equiv \frac{E}{\cosh(\eta)}$ thresholds. Values varying from 55 GeV to 360 GeV are used to record events with at least one jet with E_T above the threshold in the region $|\eta| < 3.2$.

Given the orders of magnitude difference on the cross-section between low and high energy jet events produced, triggers must be prescaled, meaning that only a fraction of the events fired by a trigger should be kept. This fraction depends on the prescale factor of each trigger, e.g a prescale of 1 means all events are kept, while a prescale of 10 means that just 1 over 10 events which passed the trigger is stored. The prescale factor decreases with the energy threshold to populate more the high p_T region, which are the most interesting events one wants to analyse. The highest threshold trigger (HLT_j360) was activated without prescales during the 2015 data-taking.

To ensure a fully efficient phase-space of the measurement in terms of the trigger, the efficiency for each trigger was measured as a function of jet p_T , as seen in Figure 3.8. It was computed by emulating the trigger in data using an inclusive jet kinematic selection. Each trigger corresponds in this way to a jet p_T region determined by the dash vertical line. These contiguous sets ensures an efficiency of 99.9% for each trigger.

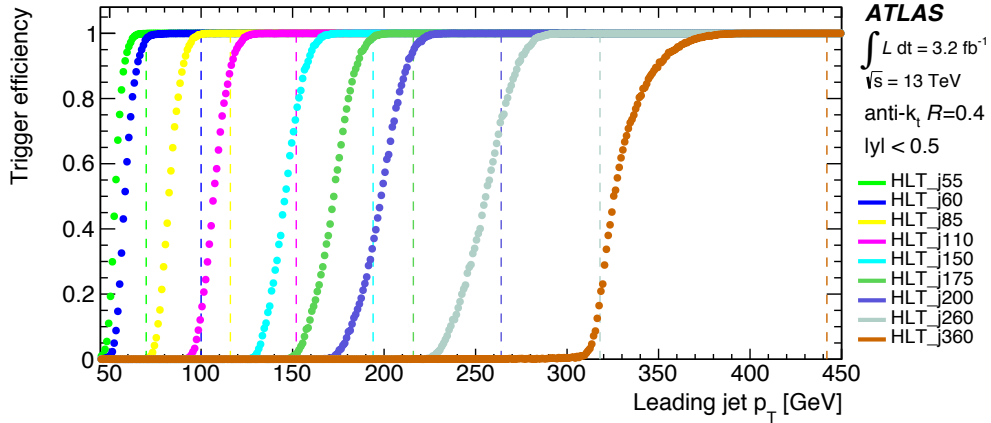


Figure 3.8: Turn-on curves of the trigger efficiency as a function of the jet p_T for the central rapidity bin, obtained by emulating the single jet triggers in data. The dashed lines represent the trigger thresholds used in measurement.

The resulted p_T ranges for each of the trigger items used in the measurement are shown in Table 3.9. These corresponds to the regions determined by the dash line thresholds in Figure 3.8. The right column represents the equivalent luminosity for each trigger defined as

p_T range, [GeV]	Trigger	Luminosity [pb^{-1}]
75–100	HLT_j55	0.0807252
100–134	HLT_j60	0.118673
134–152	HLT_j85	0.483386
152–194	HLT_j110	1.37096
194–240	HLT_j150	5.09517
240–264	HLT_j175	10.1739
264–318	HLT_j200	18.7346
318–442	HLT_j260	65.0237
442–END	HLT_j360	3158.32

Figure 3.9: p_T ranges for each trigger item determined by its region of fully efficiency. The right column shows the equivalent luminosity for each trigger item used in the measurement.

$\mathcal{L}_i = \frac{\mathcal{L}}{\pi_i}$, where π_i is the prescale factor of item i . As it can be seen, for the item HLT_j360 the prescale factor $\pi_i = 1$ and so $\mathcal{L}_i = \mathcal{L} \sim 3.2 \text{ fb}^{-1}$.

Chapter 4

Jet Reconstruction and Calibration

4.1 Introduction

Jets are a key concept in high energy physics as they represent the link between theoretical predictions, in terms of quarks and gluons, and experimental observables. In this Thesis in particular they are of utmost importance as the three major contributions were the energy calibration of jets, which is presented in the current Chapter, the measurement of the cross-section for production of di-jet events, and the identification of jets arising from the decay of boosted W bosons.

This Chapter is organized as follows. Section 4.2 discusses the different jet algorithms, used in ATLAS, and jet constituents. Section 4.3 briefly overviews the jet substructure variables used for the identification of W -jets. Finally, Sections 4.4 and 4.5 are devoted to jet energy calibration, which is one of the activities carried out during this Thesis, and to its systematic uncertainties, a key component of the analysis presented in Chapter 5, respectively.

4.2 From Particles to Jets

Because of QCD confinement, quarks and gluons cannot be directly observed. They fragment and hadronise, leading to a collimated spray of energetic hadrons, a jet. By measuring the jet energy and direction in a particle detector one can approximate to the kinematics of the originating parton. However, in order to relate the observed sprays of particles to the partons of the matrix elements in theoretical predictions, one needs a precise and robust *jet definition* algorithm both at experimental and theoretical level to perform the comparison on an equal footing.

A jet algorithm is made of a set of parameters and rules on how to combine four-momentum objects. It takes a list of partons, hadrons, tracks or calorimeter depositions, and returns parton, particle, track or calorimeter jets, respectively. To avoid theoretical issues and allow a proper comparison of theory and experiment, a jet definition should be *infrared* and *collinear* safe (IRC) [31]. This property means that if one modifies an event by a collinear splitting or the addition of a soft emission, the set of hard jets that are found in the event with the algorithm should remain unchanged, as shown in Figure 4.1.

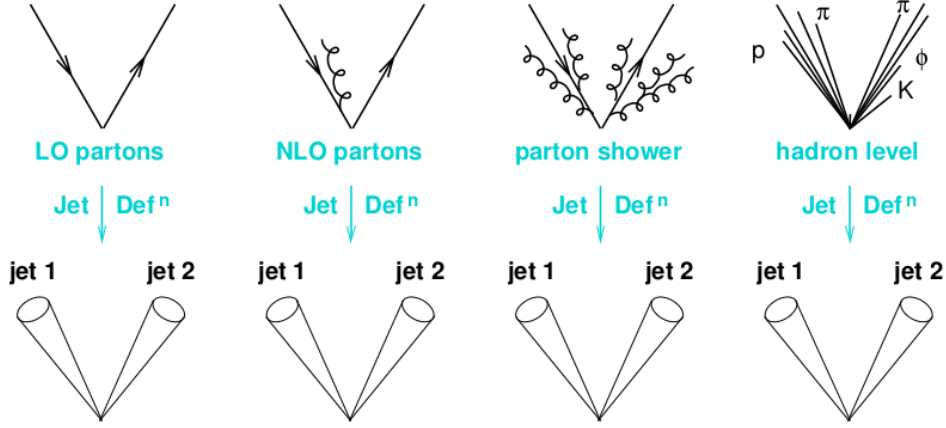


Figure 4.1: The application of a jet definition at different theoretical levels. Events that differ through soft/collinear branching and hadronization should give identical jets in all cases [32]

4.2.1 Sequential recombination algorithms

Recombination algorithms are both collinear and infrared safe. For this reason, they can be used in calculations to any order in perturbation theory. They operate by successively combining four-momentum objects using a distance metric d_{ij} between pairs of constituents:

$$d_{ij} = \min(p_{T,i}^{2p}, p_{T,j}^{2p}) \frac{\Delta R_{ij}^2}{R^2} \quad (4.1)$$

where ΔR is the angular distance $\Delta R^2 = \Delta y^2 + \Delta \phi^2$, and R is a parameter that represents the radius of the jet. The choice of the parameter p results in three different algorithms. The k_t algorithm with $p = 1$ [33, 34], the Cambridge-Aachen (C/A) algorithm [35], with $p = 0$, and the anti- k_t algorithm, with $p = -1$ [36].

The general algorithm procedure consists in:

1. Compute the minimum d_{ij} and d_{ii} among all particles, where $d_{ii} := p_{T,i}^{2p}$.

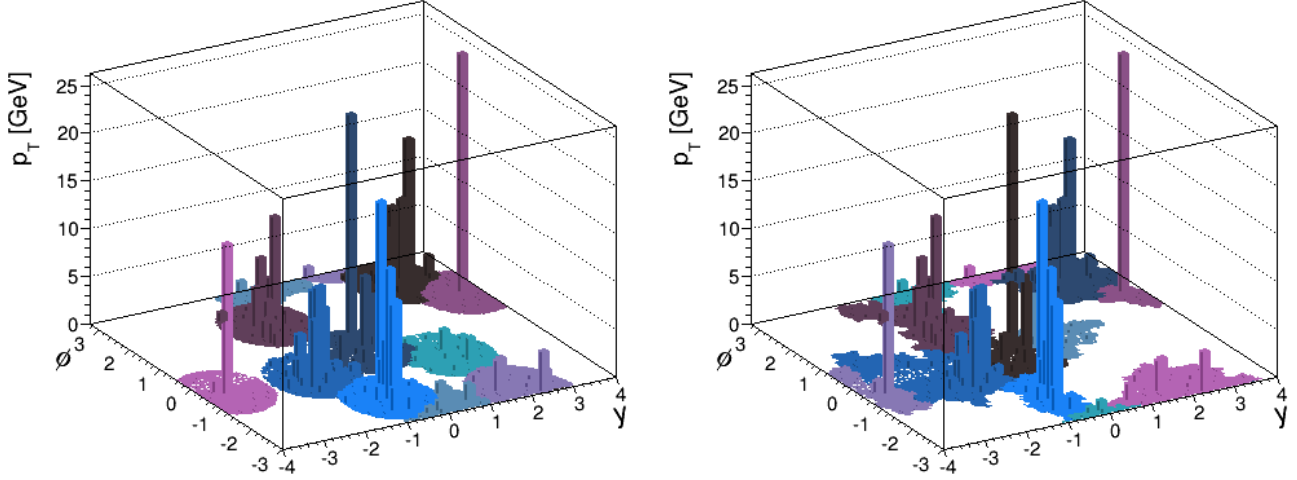


Figure 4.2: Active area of jets reconstructed with the anti- k_t (left) and k_t algorithms (right) with $p_T^{\text{jet}} > 10$ GeV.

2. If the minimum is a d_{ij} , then recombine i and j into a single particle by summing their four-momenta and repeat from step 1.
3. Otherwise, if it is d_{ii} , then identify i as a jet and remove it from the list of particles. Return to step 1.
4. Continue until no particles remain.

The jet algorithm used in this thesis was the anti- k_t algorithm, which is generally the one adopted for the reconstruction of physics objects in ATLAS. The anti- k_t algorithm first clusters hard objects together which results in more regular jets, in contrast with the k_t algorithm, which starts with soft particles, and so it results in irregular shapes (Figure 4.2). This quality makes the anti- k_t algorithm less sensitive to pile-up effects [37]. The choice of the radius parameter was limited by those used to derive the jet calibrations, as it takes long time and lot of effort to calibrate jets for a given radius. The values used were $R = 0.4$ and $R = 1.0$.

These algorithms were implemented using the *FastJet 2.4.3* [38, 39] software package.

4.2.2 Jet Algorithm Inputs

Hadronic jets used for ATLAS analyses are reconstructed starting from the energy depositions of electromagnetic and hadronic showers in the calorimeters. The input to jet reconstruction can be topological cell clusters, charged particle tracks or any object which can be described by a four-momentum. A collection of three-dimensional, massless, positive-energy topological

clusters (topo-clusters) made of calorimeter cell energies are used as input to the anti- k_t algorithm to form the calorimeter jets. Topo-clusters are built from neighboring calorimeter cells containing a significant energy above a noise threshold that is estimated from measurements of calorimeter electronic noise and simulated pile-up noise [40]. The calorimeter cell energies are measured at the EM scale, which assumes the energy is deposited by electromagnetically interacting particles. The resulting jets formed by these clusters are named EM-scale calorimeter jets (EM jets).

A second topo-cluster collection is built by taking the EM scale cluster collection, classifying them as electromagnetic or hadronic, and calibrating the cells in these clusters accordingly such that the response of the calorimeter to hadrons is correctly taken into account. This calibration uses the so called local cell signal weighting (LCW) method that aims at an improved resolution compared to the EM scale by further correcting for a variety of effects in the calorimeter [41]. LCW corrections are determined from Monte Carlo simulations of charged and neutral pions. The jets built by these clusters are known as LCW-scale calorimeter jets (LCW jets).

Another topo-cluster collection built from EM scale cells, i.e without the LCW correction, are the particle flow clusters (p-flow). This is a novel technique implemented for the first time in Run 2 which is still evolving. In this case, tracking and calorimetric information are combined in order to remove overlaps between the momentum and energy measurements made in the inner detector and calorimeters, respectively. As a result, this improves the accuracy of the charged-hadron measurement, while retaining the calorimeter measurements of neutral-particle energies, and so improving the resolution of p-flow jets at low momenta [42].

Jets referred to as truth jets are reconstructed using stable, final-state particles from MC generators as input. Candidate particles are required to have a lifetime of $c\tau > 10$ mm. Muons, neutrinos, and particles from pile-up activity are excluded to derive the calibration but used in the di-jet analysis. Truth jets are therefore measured at the particle-level energy scale.

4.3 Jet Substructure

A major field of jet physics is the identification of hadronically decaying boosted top quarks or W bosons, reconstructed as a single jet in the calorimeter. These boosted particles are the product of many beyond SM particle decays and so they are crucial to searches for new physics and measurements at the energy frontier. Such high-mass resonances result in high p_T particles which decay in collimated products. The angular separation of the decay products is $R \sim 2M/p_T$, where M and p_T are the mass and transverse momentum of the boosted, top or

W particle. This boost makes it possible to reconstruct the entire hadronic decay in a single, large-radius jet ($R = 1.0$).

The identification of resonances decaying to jets relies on jet substructure, which is generally derived from the correlations between jet constituents. These correlations reflect differences in the radiation patterns for jets produced through resonant (signal) and non-resonant (background) events, and can therefore be used to distinguish between the two. A substantial number of jet substructure variables have been proposed based on theoretical considerations, many of which have been used for jet classification in ATLAS [43–45]. Furthermore, it has been shown that improvements in the identification of hadronically decaying W bosons versus light quark- and gluon-initiated multijets can be achieved through a combination of several jet substructure variables using multivariate analysis (MVA) approaches, in particular using boosted decision trees (BDT) and neural networks (NN) [46, 47].

Table 4.1 shows the substructure variables, together with its references, used in the MVA studies described in Chapter 6.

Variable	Type	Reference
C_2, D_2	Energy correlation ratios	[48]
τ_{21}	N -subjettiness	[49]
R_2^{FW}	Fox–Wolfram moment	[50]
$\mathcal{P}$	Planar flow	[51]
a_3	Angularity	[52]
A	Aplanarity	[53]
$Z_{\text{cut}}, \sqrt{d_{12}}$	Splitting scales	[54, 55]
$KtDR$	k_t -subjettiness ΔR	[34]

Table 4.1: Substructure variables used for the neural network- and BDT-based MVA jet classifiers. The choice of these variables are motivated by [47].

4.4 Jet calibration

Calorimeter jets need to be calibrated in order to account for various known limitations of the detector as well as the inherent experimental setup, such as:

- **Calorimeter non-compensation:** correction for the different scales of the energy measured from hadronic and electromagnetic showers. Hadronic interactions deposit less ionization energy in the detector than EM ones due to the energy used to break nuclei.
- **Dead material:** energy deposited in non-instrumented regions, such as wires or services (support structures and cryostats).
- **Leakage:** showers reaching and escaping the outer edge of the calorimeters.
- **Energy deposits below noise thresholds:** clusters are only formed by energy deposits which are well above the background noise. Therefore it is required to correct for particles that do not form clusters.
- **Out of cone radiation:** particles which are included in the truth jet but not in the reconstructed jet. Part of the shower in the calorimeter develop far from the jet cone yielding topo-clusters that fail to be added to the jet in the clustering process.
- **Pile-up:** energy deposition in jets is affected by the presence of multiple proton-proton collisions in the same bunch crossing (in-time pile-up) and in previous bunch crossings (out-of-time pile-up).

In order to account for these effects, a combination of methods based on Monte-Carlo (MC) and data-driven techniques are employed to calibrate detector signals (reco jets) to physics-level objects (truth jets). Figure 4.3 presents an overview of the 2015 ATLAS calibration scheme for EM-scale calorimeter jets. Each stage of the calibration corrects the full four-momentum unless otherwise stated, scaling the jet p_T , energy, and mass.

The jet calibration techniques are discussed in detail in [56]. The following presents a summary about the jet calibration steps.

First, the origin correction recalculates the four-momentum of jets to point to the hard-scatter primary vertex¹ rather than the center of the detector, while keeping the jet energy constant. This correction improves the η resolution of jets, as measured from the difference

¹The “primary vertex” is defined by that which has the highest $\sum p_T^2$ of tracks associated with it.

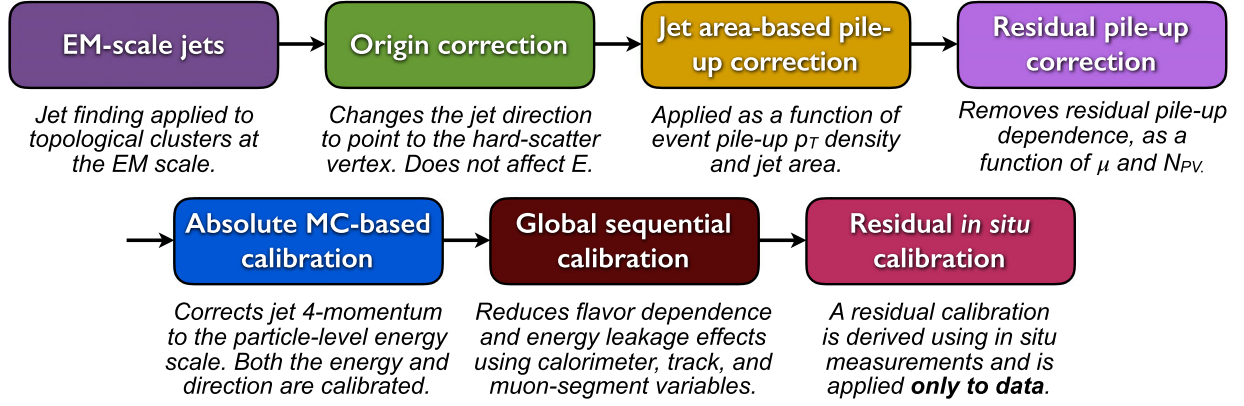


Figure 4.3: Calibration stages for EM-scale jets. Other than the origin correction, each stage of the calibration is applied to the four-momentum of the jet.

between reco jets and truth jets in MC simulation. The η resolution improves from roughly 0.06 to 0.045 at a jet p_T of 20 GeV and from 0.03 to below 0.006 above 200 GeV.

Next, the pile-up correction removes the excess energy due to in-time and out-of-time pile-up. It consists of two components: an area-based method subtraction based on the median p_T density (ρ) [37] and a residual correction as a function of μ and N_{PV} of the event derived from MC simulation.

The MC-JES calibration corrects, on average, the jet four-momentum to the particle-level energy scale, as derived using truth jets in dijet MC events, and is discussed in detail in the next section. Further improvements to the reconstructed energy and related uncertainties are achieved through the use of calorimeter information of longitudinal development of the jet, muon segments, and track-based variables in the global sequential calibration (GSC).

Finally, a residual in situ (data-driven) calibration is applied to correct jets in data using well-measured reference objects, including photons, Z bosons, and calibrated jets.

4.4.1 Monte Carlo based jet energy calibration

This section describes the technique used to implement the jet energy scale (JES) based on MC in order to recover the jet particle-level (true) scale from the reconstructed (reco) one. The calibration is reduced to energy- (E) and η -dependent corrections. It consists in determining the jet response as a function of the jet reco energy, and applying it as $E^{reco}/R(E^{reco})$ on a jet-by-jet basis, where the response function is defined as:

$$R(E^{reco}) = \left\langle \frac{E^{reco}}{E^{true}} \right\rangle$$

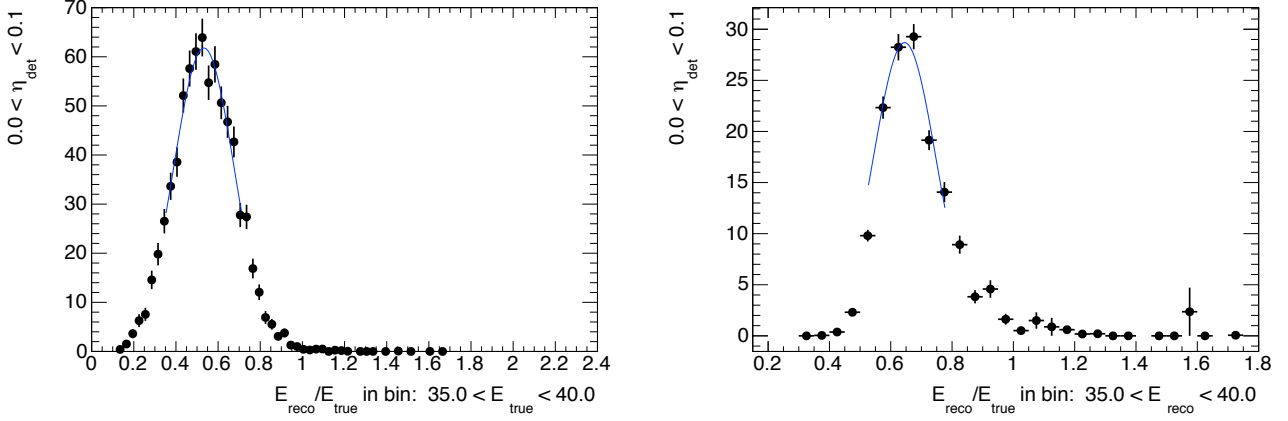


Figure 4.4: Response distribution for a given E^{true} (left) and E^{reco} (right) bin. The response in the left is mostly gaussian, while the one in the right present a bias towards higher values.

The E correction is derived in η bins of size 0.1 from -4.5 to 4.5 . Calibrations using different reconstruction jet scale choices such as EM (nominal), LCW and pflow and various jet radii $R = (0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 1.0)$ were derived and released for the use of the whole collaboration. We will focus on the results obtained for jets reconstructed at EM scale with $R = 0.4$, which were the default recommendation proposed by the calibration team to use in jet analyses. The studies were performed using a high statistics di-jet MC sample simulated with *Pythia*.

4.4.2 Numerical Inversion Technique

To estimate the true energy of a jet based on reco information only, we need to estimate the response as a function of E^{reco} , $R(E^{reco})$. With it, the true energy of a jet can be estimated as $\frac{E^{reco}}{R(E^{reco})}$. However, it is not straightforward to evaluate the response as a function of E^{reco} , because for fixed E^{reco} the response distribution is not Gaussian. On the other hand, the calorimeter response for jets of fixed E^{true} is Gaussian, as illustrated in Figure 4.4. This is related to the fact that the cross-section is a steeply falling function with E^{true} , then for a fixed E^{reco} bin there will be more jets from low E^{true} (i.e: higher response), causing a bias towards higher values of the response distribution.

The *numerical inversion* technique attempts to solve this issue by determining $R(E^{true})$ and then $R(E^{reco})$ from it. It can be summarized as follows:

1. Calculate $R(E^{true})$: First, the distributions of $\frac{E^{reco}}{E^{true}}$ are evaluated in E^{true} bins. Next, each distribution is fit by a Gaussian and the mean value is computed as a function of

the E^{true} bin center. Finally, in order to get a continuous response function to be applied on a jet-by-jet basis, a polynomial of the form of Equation 4.2 is fitted.²

2. Estimate the reco energy of a jet from its true energy as: $E_{est}^{reco} = R(E^{true}) \cdot E^{true}$, which is an unbiased estimator for E^{reco} assuming $R(E^{true})$ to be Gaussian.
3. Plot the response points obtained in step 1, but now in E_{est}^{reco} bins and fit it with Equation 4.2.

Notice that the estimated $R(E^{reco})$ was in fact obtained from E_{est}^{reco} . However, when used as part of the calibration it is evaluated on a jet at E^{reco} .

$$R(E) = a_0 + \sum_{i=1}^n a_i (\log E)^i \quad (4.2)$$

The effect of the numerical inversion technique is illustrated in Figure 4.5. The response function (left) is used to get the final calibration curve (right). The transformation corresponds to displacing the points parallel to the x -axis to the left ($E^{reco} < E^{true}$), leaving the y -axis unchanged.

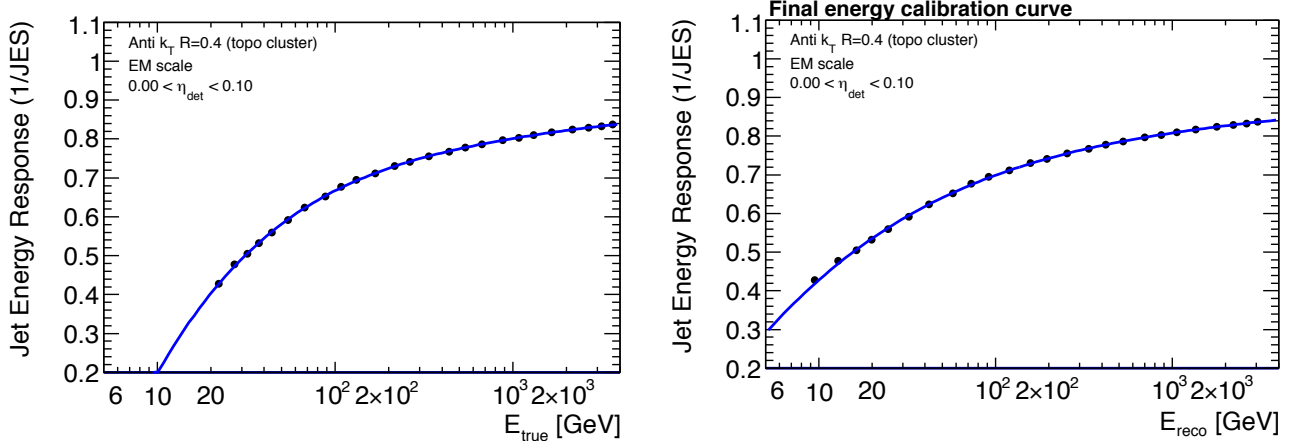


Figure 4.5: Average response as a function of E^{true} (left) and E^{reco} (right). The right plot is the final calibration curve. Statistical uncertainties are negligible since this was obtained from MC simulation.

The shape of the response function depends on the clusters used as inputs to the jet algorithm and on the jet radius parameter R . Responses for different jet collections can be appreciated in Figure 4.6. As it can be seen in the figure on the left, LCW jets have a higher response (closer to the particle-level) than EM jets. This is expected as LCW jets are built from calorimeter clusters partially calibrated to the hadronic scale [41]. Both curves increase

²The physics motivation behind this functional dependence can be found in Appendix A.

with energy, as expected from the rationale described in Appendix A. On the other hand, p-flow jets present a better response than EM and LCW jets for $E^{true} < 100$ GeV and then decreases smoothly up to the EM response. This is expected given that in p-flow jets, the energy measured in the calorimeter of EM topo clusters that have associated tracks is replaced using the momentum of tracks measured in the inner detector, yielding an improvement of the jet resolution and energy response particularly for jets with $p_T < 100$ GeV. The figure in the right shows the jet response for different jet radii. It is clearly seen how the response decreases as the jet radius increases. This could be related to the lower energy density in the outer part of the for larger radii jets, which brings about two effects, namely a lower probability of the energy to be clustered, and a lower calorimeter response.

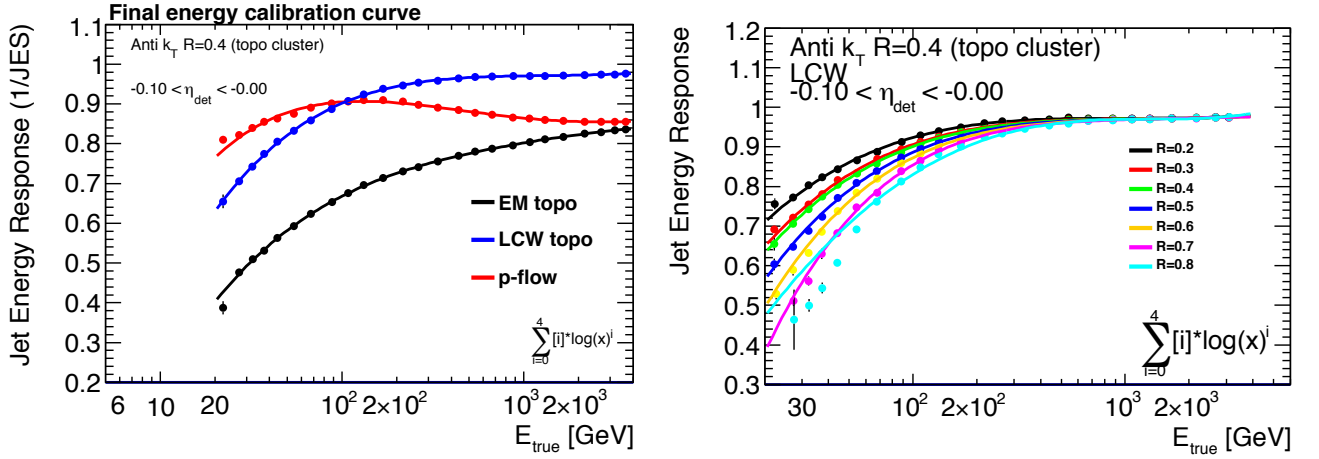


Figure 4.6: Response function for EM, LCW and p-flow jets (left), and different jet radius parameters (right).

4.4.3 E and η calibrations

With an expression for $R(E^{reco})$, the jet can be calibrated by applying the correction factor $R(E^{reco})^{-1}$ to E^{reco} :

$$E^{calib} = E^{reco} / R(E^{reco})$$

The jet four-momentum is calibrated in an equivalent way by using $R(E^{reco})^{-1}$ as a scale factor for each of its components.

Once the energy calibration derived, the following step is to correct for the jet η bias caused by the jet energy response and the calorimeter granularity dependences on η [56]. The η response is defined as $\langle \eta^{reco} - \eta^{true} \rangle$ and is parametrized also using Equation 4.2 in terms of E^{true} . There is no need of numerical inversion in this case since the jets have already been

calibrated on average to the true energy scale. The correction is applied on a jet-by-jet basis:

$$\eta^{calib} = \eta^{reco} - \langle \eta^{reco} - \eta^{true} \rangle$$

4.4.4 Closure Test

The derived E and η calibrations were validated by comparing E^{calib} and η^{calib} with their respective true values. In concrete, the distribution of E^{calib}/E^{true} was computed in different E^{true} and η bins. The averages of these distributions are then plotted for each E^{true} bin as a function of η , as shown in Figure 4.7. The average response agrees with unity within 1% for jets with $p_T > 30$ GeV, presenting a non-closure below that range, similar to what obtained in the Run 1 calibration [41]. This is traced back to the non-gaussianity of the response distribution at low p_T , arising from a $p_T < 5$ GeV cut imposed at the jet constituent level. This was done in order to reduce the size of the dataset, although nowadays lower thresholds are being used precisely to achieve better calibrations at low p_T .

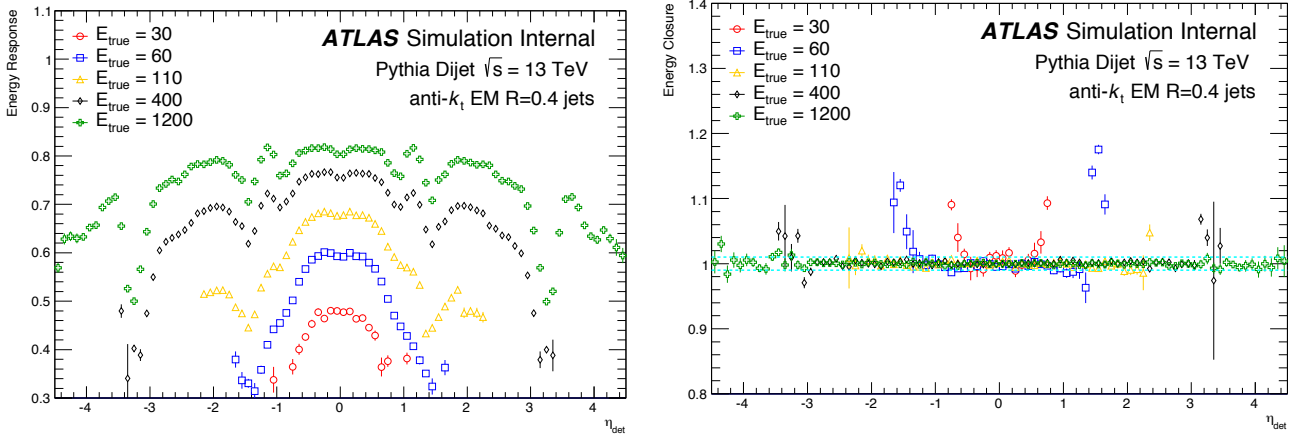


Figure 4.7: Average response in different E^{true} bins as a function of η before (left) and after (right) the calibration.

4.5 Jet systematic uncertainties

The most relevant jet uncertainty comes from the Jet Energy Scale (JES) determination based on the calibration chain described in Fig. 4.3. The final calibration ends up with a set of 80 JES uncertainty terms propagated from each calibration stage. Most of them (70) arise from the *in situ* corrections and account for assumptions made in the event topology, MC simulation, sample statistics, and electron, muon, and photon energy scale propagation uncertainties.

The remaining ones correspond to pile-up corrections, potential MC mismodelling of μ , N_{pv} and ρ , *punch-through* GSC correction, differences in the flavour response between gluon and light-quarks initiated jets and high p_{T} single-particle response studies. A detailed list of each uncertainty source can be found in [56].

The combination (sum in quadrature) of all relative jet p_{T} uncertainty terms as a function of jet p_{T} and η is shown in Figure 4.8.

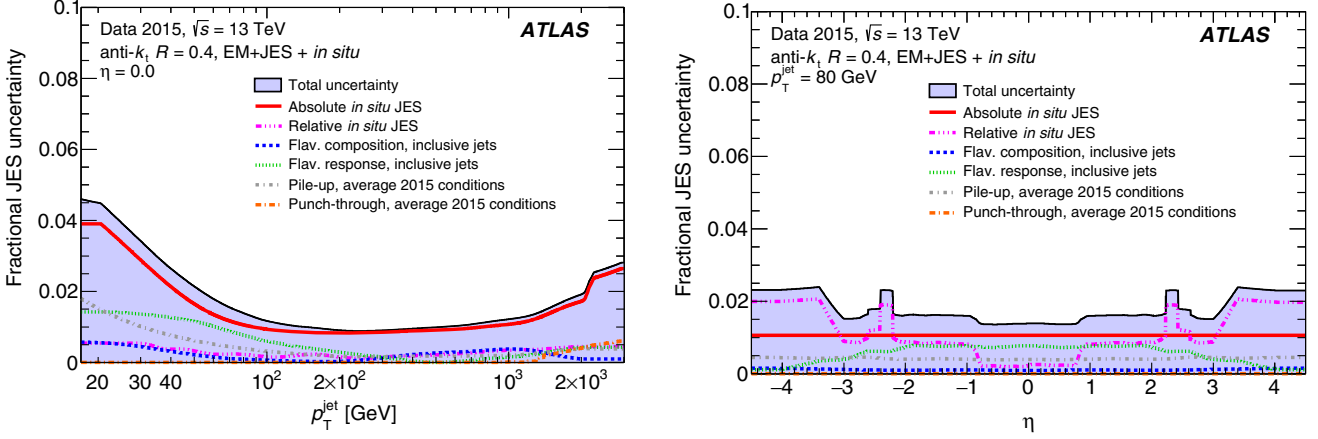


Figure 4.8: JES p_{T} uncertainty of fully calibrated jets as a function of jet p_{T} at $\eta = 0$ (left) and as a function of η at $p_{\text{T}} = 80$ GeV (right).

The uncertainty from the statistical variations of the relative *in situ* calibration (η -Intercalibration) was split in 247 single components corresponding to the statistical uncertainty of each $p_{\text{T}}-\eta$ bins. The use of this decomposition is very important particularly for precise measurements, because a better modelling of the uncertainties results in a more accurate estimation of the degree of agreement between data and theory.

Another lesser but still significant contribution arises from the uncertainty on the precision of the measurement of the energy of a jet. In other words, the Jet Energy Resolution (JER) uncertainty. This will be covered in detail in Section 5.6.2.

Chapter 5

Measurement of Dijet Production at 13 TeV

5.1 Introduction

Precise measurements of jet cross-sections are crucial in understanding QCD physics at hadron colliders. In particular they provide valuable information about the strong coupling constant, α_s , and the structure of the proton (PDFs). What is more, dijet events represent a background to many other processes at hadron colliders. The predictive power of fixed-order QCD calculations is therefore also relevant in many searches for new physics and so it must be tested with experimental data.

In this chapter the measurement of double-differential dijet production cross-sections is presented as a function of the invariant mass of the dijet system, m_{jj} , and as a function of half the absolute rapidity separation, denoted y^* , defined as $|y_1 - y_2|/2$, where the subscripts 1,2 label the highest and second highest- p_T jets in the event satisfying $|y| < 3.0$, respectively. This quantity is invariant under a Lorentz boost along the z -direction and is equal to the absolute rapidity of each jet in the dijet rest frame.

The double-differential dijet cross-section is calculated as the number of dijet events in a given m_{jj} and y^* bin after correcting for detector effects (Section 5.4), N_{dijet} , divided by the integrated luminosity of the data, and the size of the intervals in invariant mass and y^* , Δm_{jj} and Δy^* respectively:

$$\frac{d^2\sigma}{dm_{jj}dy^*} = \frac{N_{\text{dijet}}}{L\Delta m_{jj}\Delta y^*}$$

This Chapter is structured as follows. Section 5.2 presents the event selection and trigger strategy. Cross-section comparisons between data and MC (Pythia) are featured in Section 5.3. Section 5.4 describes the experimental unfolding technique used to deconvolute the data spectrum. Section 5.5 and Section 5.6 present the statistical and the systematics uncertainties

respectively. The theoretical predictions and non-perturbative corrections are shown in Section 5.7. Finally, the results on the comparison with NLO perturbative QCD predictions, corrected for electroweak and non-perturbative effects, are detailed in Section 5.8.

5.2 Event Selection and Trigger Analysis Strategy

Events selected for analysis require at least two jets with $p_T > 75$ GeV and $|y| < 3$. In addition, the scalar sum of the p_T of the first and second leading jets, $H_{T,2} = p_{T1} + p_{T2}$, must be above 200 GeV. This was required to avoid known instabilities at NLO jet production cross-sections in the presence of symmetric cuts, i.e $p_{T1} > p_{T\text{ cut}}$ and $p_{T2} > p_{T\text{ cut}}$ [58].

As mentioned in Section 3.5.1, single-jet triggers with E_T thresholds varying from 55 GeV to 360 GeV were used to record events. The trigger strategy for the dijet production accounts for different prescale combinations for dijet events in a given (m_{jj}, y^*) bin, which can be accepted by up to two jet triggers depending on the transverse momenta and pseudorapidities of the leading and sub-leading jets. Each pairing of triggers has a unique corresponding luminosity, which is used to calculate the differential cross section for that pairing. The separate cross sections from all pairings are then summed to obtain the physical distribution. This strategy is described as the *Inclusive method for fully efficient combinations* in [59]. As an example, consider an event which has a leading and a sub-leading jet with $318 \text{ GeV} < p_T^{(1)} < 442 \text{ GeV}$ and $264 \text{ GeV} < p_T^{(2)} < 318 \text{ GeV}$. Each jet belongs to a different fully efficient trigger region described in Table 3.9. In this case the corresponding triggers are: t1=HLT_j260 and t2=HLT_j200 respectively. When the two jets fall in different trigger regions, an equivalent luminosity of the “OR” from the two triggers is calculated by weighting the event by $w = 1/P$, where P is the probability of the event to pass either t1 or t2 prescales, determined by their prescale factors (π_{t1}, π_{t2}) respectively:

$$P(t1 \cup t2) = P(t1) + P(t2) - P(t1 \cap t2) = \frac{1}{\pi_{t1}} + \frac{1}{\pi_{t2}} - \frac{1}{\pi_{t1}} \times \frac{1}{\pi_{t2}} \quad (5.1)$$

When the leading and sub-leading jets belong to the same trigger region, i.e: $318 \text{ GeV} < p_T^{(1)}, p_T^{(2)} < 442$, the equation 5.1 is reduced to $P = \frac{1}{\pi_{t1}}$.

5.3 Data vs. Monte Carlo cross-section comparisons

Before entering in the description of the unfolding procedure, let’s show some results for the m_{jj} cross-sections in data compared to those obtained in Pythia MC simulation at detector

level. The distributions for the six y^* bins are depicted in Figure 5.1. Although the agreement looks reasonable over the six y^* bins, there are some shape differences particularly at high p_T . Those differences will enter as a systematic uncertainty from the unfolding procedure, described later in Section 5.4.4. Notice how the first m_{jj} bin moves to higher values as y^* increases. This arises from the fact that at high y^* the jets are more separated in rapidity and for a fixed energy (imposed by the $p_T > 75$ GeV cut from the analysis selection) the m_{jj} becomes higher. Mathematically it can be easily derived from the following fundamental kinematic relation:

$$m_{jj}^2 = (P_1 + P_2)_\mu (P_1 + P_2)^\mu = m_1^2 + m_2^2 + 2(E_1 E_2 - |\vec{P}_1| |\vec{P}_2| \cos \theta) \quad (5.2)$$

where the indexes (1) and (2) label the leading and sub-leading jets respectively. The angle θ is the separation between the two jet momenta, which is proportional to y^* . Here it can be seen that when the separation increases the third term of the equation becomes higher.

The comparison with Powheg+Pythia 8 Monte Carlo simulation has also been made, resulting in a worse agreement than Pythia. This was one of the main reasons of the choice of Pythia as the nominal MC used in the unfolding.

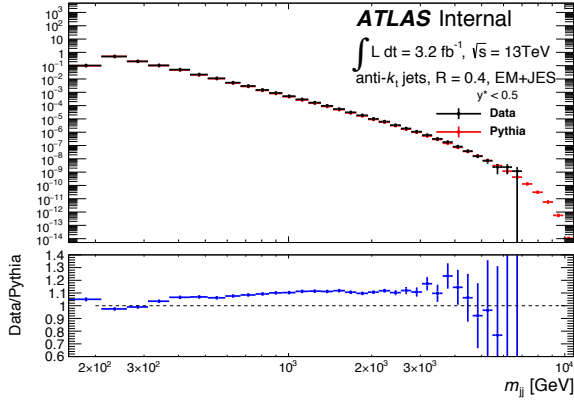
5.4 Data Unfolding

The comparison of data and theoretical predictions at particle level rather than at reconstruction level has the advantage that data can be used to test any theoretical model without the need for further detector simulation. In other words, unfolded distributions provide a result which can directly be compared to other experiments as well as with theoretical predictions. Another possibility, clearly much less preferred, would be to avoid the unfolding and publish instead the reconstructed cross-sections and the detector resolutions so that theory groups can smear their predictions before comparing to our data.

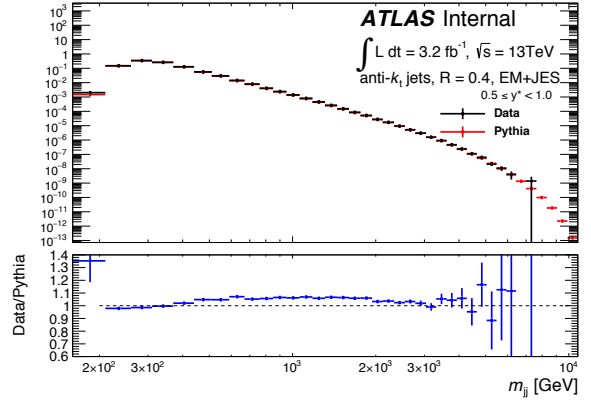
It is worth to mention that the *unfolding*, also called *de-convolution* or *unsmearing*, can be found in mathematics under the general heading of *inverse problems*, which is the process of calculating from a set of observations the causal factors that produced them. As one can imagine, this has many applications not only in particle physics, but in fields such as optical image reconstruction, radio astronomy, crystallography and medical imaging.

In this section, the description of the unfolding technique used to de-convolute the data by detector effects is presented.

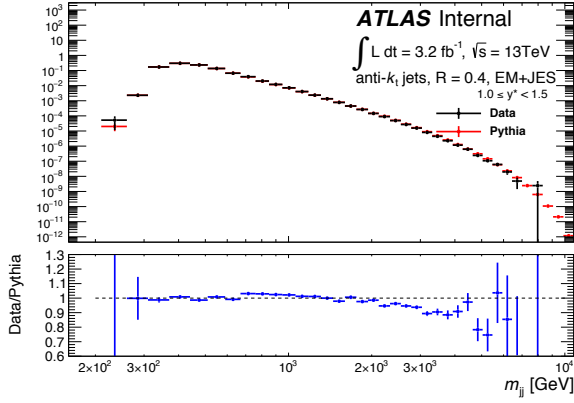
It is based on an iterative method which is detailed in [60, 61]. This technique is very general and is capable to be applied in analyses when the data contain structures which are not present



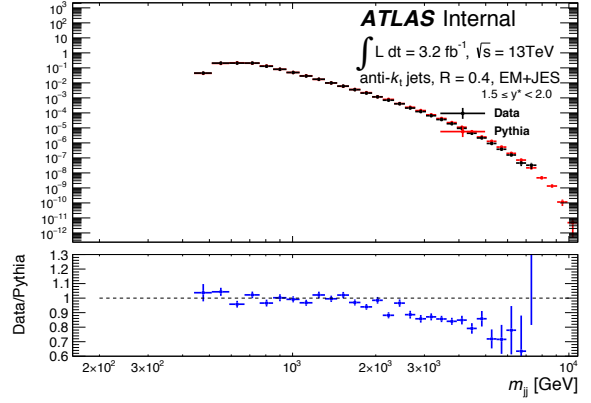
(a) $y^* < 0.5$



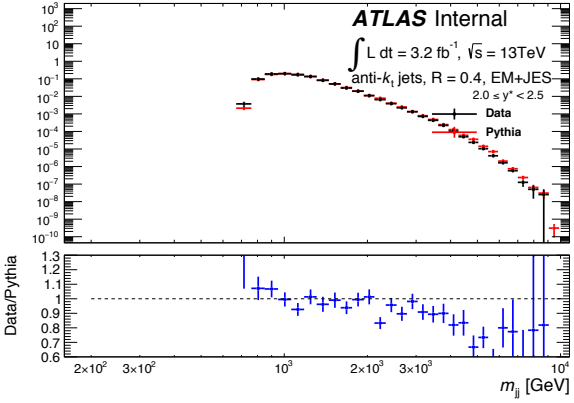
(b) $0.5 < y^* < 1.0$



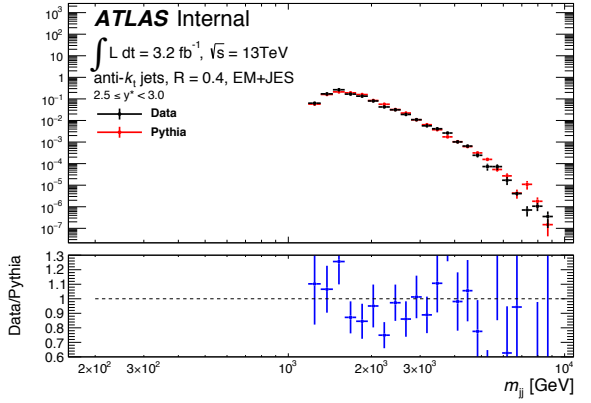
(c) $1.0 < y^* < 1.5$



(d) $1.5 < y^* < 2.0$



(e) $2.0 < y^* < 2.5$



(f) $2.5 < y^* < 3.0$

Figure 5.1: The m_{jj} cross-section at detector level for anti- k_t $R = 0.4$ jets in different bins of y^* in data (black points) and Pythia (red points) normalised to the total area. The ratio of the two (blue points) is shown in the bottom panel.

in the MC simulation. In order to facilitate the explanation, the description of the technique is limited to the use for this particular measurement, and so issues such as unexpected structures

in the data, background subtraction effects and different initial/final binnings will be omitted.

The unfolding attempts to correct physics observables by detector effects due to limited acceptance, finite resolution, or other systematic effects producing a transfer of events between different regions of the spectra. This method is based on the idea that the MC simulation provides a relatively good description of the data and of the detector response, which is justified by the calibrations previously applied, covered in Section 4.4, such as the MC-JES and *in-situ* calibrations. Nevertheless, as it will be shown, the technique can be iterated until the MC shape gets closer to the data. In addition, a systematic uncertainty due to possible residual MC/data mismodelling is derived which will be described in detail on Sec. 5.4.4.

In order to describe the migrations of events across m_{jj} bins between the particle- (part) and reconstructed- (reco) level, a transfer matrix was built from MC. It is determined by the number of events generated in bin i and reconstructed in bin j (A_{ij}). It was filled with events that pass the selection cuts both at reco and part level and belong to the same y^* bin, fulfilling in this way the *matching condition*. Figure 5.2 shows A_{ij} for the central y^* bin.

5.4.1 The General Unfolding Procedure

The whole deconvolution technique can be summarized in three steps.

First, the data d_i , where i labels a m_{jj} bin index, is corrected by the matching inefficiency at reco level, also known as purity, which is the conditional probability that a given reco event matches a true event $P(\text{true} \mid \text{reco})_i$. The conditional probability was obtained by using the definition $P(A|B) = \frac{P(A \cap B)}{P(B)}$. In this way, $P(\text{true} \mid \text{reco})_j = \frac{\sum_{k=1}^{n_d} A_{kj}}{R_j}$, where R_j are all reco events which pass the selection but are not necessarily matched by a true event, and n_d is the number of bins in data. The corrected data $\tilde{d}_i$ spectra results in:

$$\tilde{d}_i = d_i \times P(\text{true} \mid \text{reco})_i$$

Next, the data is corrected by migrations of events between different reco m_{jj} bins due to resolution effects. For this, folding and unfolding matrices are computed: $P_{ij} = \frac{A_{ij}}{\sum_{k=1}^{n_d} A_{ik}}$, $\tilde{P}_{ij} = \frac{A_{ij}}{\sum_{k=1}^{n_d} A_{kj}}$. The folding (unfolding) probability matrix P_{ij} ($\tilde{P}_{ij}$) corresponds to the probability of an event generated (reconstructed) in bin i (j) to be reconstructed (generated) in bin j (i). The μ_i spectra corrected by matching inefficiency at reco level and resolution effects (unsmearing) is then obtained as described in Sec 5.4.2.

Finally, a correction due to matching inefficiency at part level (ϵ_i) is applied:

$$\tilde{\mu}_i = \mu_i / \epsilon_i = \mu_i / P(\text{reco} \mid \text{true})_i$$

where $\tilde{\mu}_i$ is in this way the unfolded result ready to be compared to the theoretical predictions.

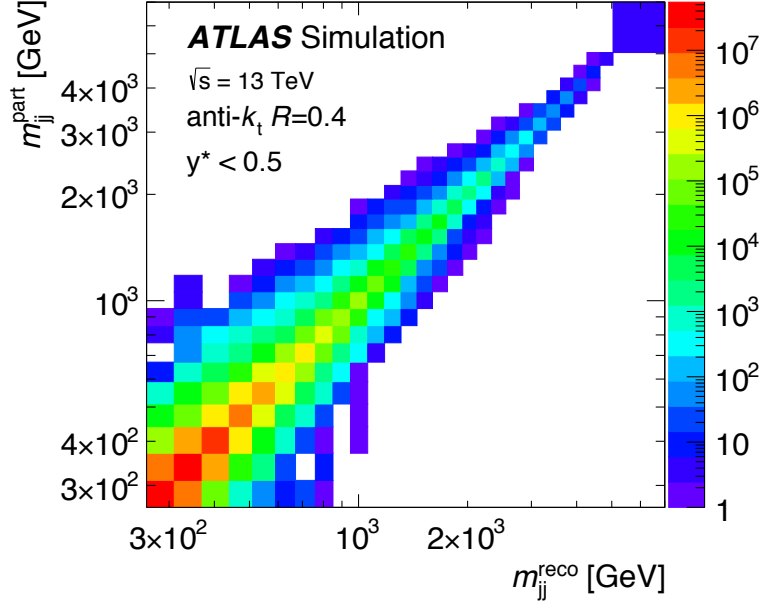


Figure 5.2: Transfer matrix used for unfolding for the central y^* bin.

5.4.2 The Iterative Unsmearing Correction

From the definitions of P_{ij} and $\tilde{P}_{ij}$, it is easy to see that the reconstructed and true MC distributions are related through:

$$r_j = \sum_{k=1}^{n_d} P_{kj} t_k$$

$$t_i = \sum_{k=1}^{n_d} \tilde{P}_{ik} r_k \quad (5.3)$$

where, for a given bin i , t_i and r_i are the number of true and reco MC events respectively.

Equation 5.3 can be seen as a “basis for the unfolding”, in the sense that it describes this operation performed on MC. If the differences between data and reconstructed MC are relatively small, then those between the unfolding result and true MC will be small too. A zero order result would then be given by the following relation:

$$\mu_i^{(0)} = \sum_{k=1}^{n_d} \tilde{P}_{ik} d_k \quad (5.4)$$

where the superscript labels the iteration step number.

From this first result, the transfer matrix is iteratively improved as described in Sec 5.4.3, which yields to a better MC true estimate and so a smaller unfolding uncertainty (Sec. 5.4.4).

The result after n iterations is then given by:

$$\mu_i^{(n)} = t'_i N + \sum_{k=1}^{n_d} f(\Delta d_k, \sigma, \lambda) \tilde{P}'_{ik} \Delta d_k \quad (5.5)$$

where $N = \frac{N_d}{N_{MC}}$ is a normalization coefficient defined as the ratio between the total events in data and MC respectively after purity correction. t'_i is the modified MC true spectra resulting from a projection of the improved transfer matrix. $\tilde{P}'_{ik}$ is the improved unfolding matrix and $\Delta d_k = d_k - N r_k$ is the difference between data and reco MC spectra normalized to the data. $f(\Delta d_k, \sigma, \lambda)$ is a regularization function defined as:

$$f(\Delta d_k, \sigma, \lambda) = 1 - e^{-\left(\frac{\Delta d_k}{\lambda \sigma}\right)^2}$$

Notice that by considering $f(\Delta d_k, \sigma, \lambda) = 1$, Equation 5.5 can be expressed as

$$\mu_i^{(n)} = t'_i N + \sum_{k=1}^{n_d} \tilde{P}'_{ik} \Delta d_k = \sum_{k=1}^{n_d} \tilde{P}'_{ik} r_k N + \sum_{k=1}^{n_d} \tilde{P}'_{ik} (d_k - r_k N) = \sum_{k=1}^{n_d} \tilde{P}'_{ik} d_k$$

where we have used that $t'_i = \sum_{k=1}^{n_d} \tilde{P}'_{ik} r_k$. By comparing with Equation 5.4, this is what one would expect to get after performing n iterations (i.e n improvements of the transfer matrix).

The regularization function is used in order to control dynamically the smoothness of the solution and avoid unfolding events which could be due to statistical fluctuations, which are naturally amplified through each iteration. σ is the statistical error of data and MC (the latter being negligible). $\frac{\Delta d_k}{\sigma}$ represents in this way the significance of the absolute deviation between data and simulation in a given bin. The parameter λ is a scaling factor which is chosen to be $\lambda = 0$ for the zero iteration, i.e $f(\Delta d_k, \sigma, \lambda) = 1$, and $\lambda = 0.5$ for the following iterations. The number of iterations was chosen as the minimum number that gives an unfolding uncertainty at the percent level. This was achieved using only one iteration, as shown in Sec. 5.4.4.

Figure 5.3 shows the ratio of number of events between unfolded and reco data as a function of m_{jj} for the central y^* bin. As it can be seen, the unfolding correction is relatively small (less than 4%). Also, the fluctuations with respect to one iteration are higher for both zero and ten iterations. This is because for zero iterations no regularization was used. On the other hand, for the ten iterations scenario, the scaling parameter λ is not big enough to attenuate the statistical amplification. The green dashed line shows the first reliable bin in which the final measurement is presented. This was chosen as the first bin not affected by the selection cut $H_{T,2} > 200$ GeV, defined in Sec. 5.1. However, the unfolding was performed starting from previous bins in order to consider both positive and negative migrations for the first measured bin, i.e getting an unbiased correction.

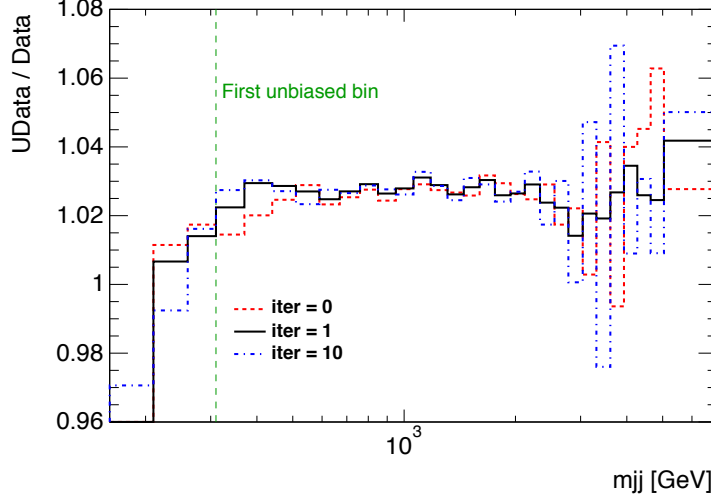


Figure 5.3: Ratio of number of events between unfolded data and data as a function of m_{jj} for a central y^* bin.

5.4.3 Improvement of the unfolding probability matrix

The initial true MC distribution might not contain or badly describe some structures which are present in the data. One can iteratively improve it, and hence the transfer matrix, by using a better (weighted) true MC distribution with the same folding matrix describing the physics of the detector, which yields an improved unfolding matrix.

The improvement is performed for one bin i at the time, exploiting the difference between an intermediate unfolding result and the true MC

$$A'_{ij} = A_{ij} + \Delta\mu_i P_{ij} \frac{N_{MC}}{N_{\mu d}}, \quad \text{for } j \in \{1; n_d\}$$

with $\Delta\mu_i = \mu_i - t_i \frac{N_{\mu d}}{N_{MC}}$ and $N_{\mu d}$ the total number of unfolded events, which in our case is equal to the total number of data events N_d after purity correction.

This method allows an efficient improvement of the unfolding matrix, without introducing spurious fluctuations. The amplification of small fluctuations can be prevented at this step by using the regularization function discussed in the previous section.

5.4.4 Unfolding bias

The transfer matrix was improved by one iteration to account for shape differences between data and MC on the m_{jj} spectra. In addition, a data-driven closure test was performed to account for any residual MC mismodelling.

In this study, the part level m_{jj} spectrum in the MC simulation is reweighted and multiplied by the folding matrix in order to get a modified reconstructed spectrum. The weights are chosen

ad hoc such that a significantly improved agreement between the resulting reconstructed spectra and data is attained. The reweighted reco MC is then unfolded with the nominal transfer matrix as used for the data. Comparison of this result with the reweighted part level spectra provides an estimate of the bias, which is interpreted as the unfolding systematic uncertainty. Figure 5.4 shows the uncertainty as a function of m_{jj} for different iteration steps for the central y^* bin. As it can be seen, an uncertainty at the percent level was achieved with just one iteration.

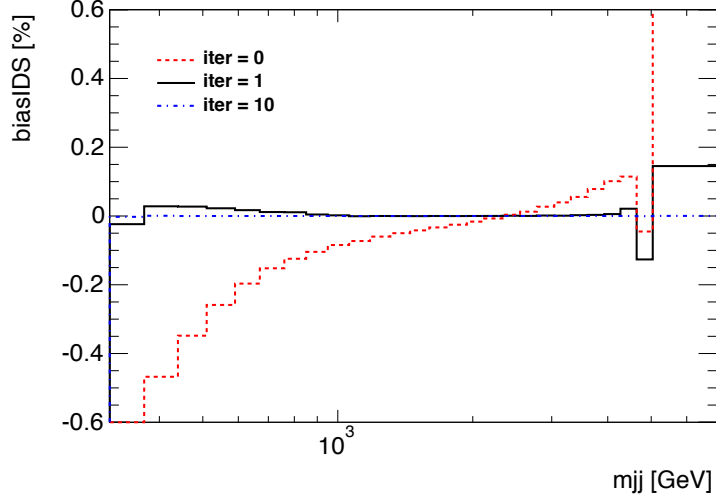


Figure 5.4: Unfolding systematic uncertainty as a function of m_{jj} for different iteration steps obtained from the data-driven closure test for the central y^* bin.

Due to lack of data and MC statistics, particularly for the two last y^* bins, the uncertainty was smoothed by using a Gaussian kernel technique, which is detailed in Appendix B.

5.5 Estimation of Statistical Uncertainties: Bootstrapping

Because of the unfolding procedure, correlated statistical uncertainties are introduced in the measurement. As a result, a covariance matrix has to be determined. Due to the complexity of the unfolding method, it may result impractical, if not impossible, to compute the covariance matrix analytically through the law of error propagation. The only option is to estimate it empirically. A naive solution would be to collect more data and repeat the experiment multiple times, or divide the sample space in N subsamples and compute the variability σ in order to get the error as $\sigma/\sqrt{N}$. The first approach is not possible since we have limited data. The second one will be a poor estimate of the variance, in the sense that it will present high statistical fluctuations. A more sensible solution, which does not need any subsampling division (or collecting more data), is the *bootstrapping* technique, a computational statistical method that

simulates new samples to help determine how estimates from replicate experiments might be distributed, and so it allows to answer questions about precision and bias. What is more, the bootstrap technique used here does not make any assumption on the underlying distributions of the measurement, and so it can be used generally for any kind of observables. The method consists basically in generating N replicas from the original sample by weighting each event by a random variable k extracted from a *Poisson* distribution with $\mu = 1$.

$$\text{event}_j^{\text{replica}} = \text{event}_j \times k$$

We obtain in this way, in addition to the nominal distribution ($k = 1, \forall j$), N replicas of the same observable. The unfolding is performed for each replica and a covariance matrix of the cross-section is constructed in each y^* bin. The total statistical uncertainty for each m_{jj} bin is obtained from the diagonal elements of the covariance matrix (i.e the variance). The separate contributions from the data and from the MC statistics are obtained from the same procedure by varying either the data or the simulated events. The numbers of replicas used for this analysis was $N = 1000$.

In order to validate this approach, a MC sample was divided in 32 subsamples and the variance, for one of the systematic errors, was obtained via the usual equation $\sum_{i=1}^{32} (x_i - \bar{x})^2 / 31$. The uncertainty was then scaled by $1/\sqrt{32}$ to represent an estimate of the full sample statistics. This is the naive approach one would take to compute the statistical uncertainty. The second one, and much better possibility, is to bootstrapping the full sample. The result in this way will be smoother because the bootstrap algorithm is statistical superior. Figure 5.5 shows the comparison between the first approach (red curve) and the second one (black). The blue curve corresponds to bootstrapping one subsample, multiplied by the factor $1/\sqrt{32}$. As expected, the three curves overlap each other, but the fluctuations of the red and blue curves are larger than the black one.

The bootstrapping method was used to estimate the statistical uncertainties of data and MC spectra, as well as for computing the variance of the systematic uncertainties needed by the smoothing procedure (Appendix B). The covariance matrix was used to test the level of agreement between data and theory (Appendix C).

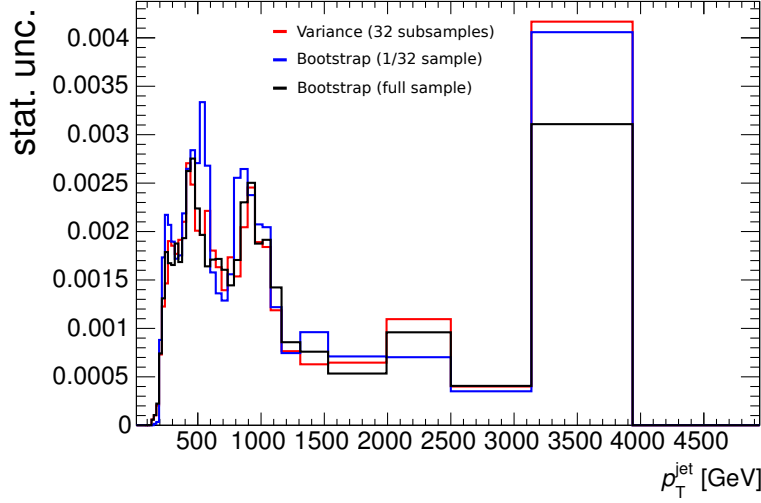


Figure 5.5: Statistical uncertainty of one of the systematic errors as a function of jet p_T computed with three different approaches. Arithmetic calculation from the 32 MC subsamples (red curve), bootstrap in 1 subsample (blue curve), and bootstrap in the full sample (black curve). As expected, fluctuations of the red and blue curves are larger than the black one.

5.6 Propagation of Systematics Uncertainties

5.6.1 Jet Energy Scale

All the components of the JES uncertainty (Section 4.5) are propagated throughout the unfolding process. Each jet p_T is scaled up and down by one standard deviation for each component.

$$p_T'^{\text{jet}} = p_T^{\text{jet}} \pm \Delta$$

where Δ is the jet uncertainty from a given source of systematic.

The p_T shifted jets are used to extract a shifted reco cross-section and the resulting spectra are unfolded using the nominal transfer matrix. The difference between the nominal unfolded cross-section and the systematically shifted unfolded cross-section is taken as the systematic uncertainty.

A breakdown of the propagated JES uncertainties to the dijet cross-section can be seen in Figure 5.6. The dominant contributions come from the LAr, jet flavour, E/P (single-particle response) at high mass, η -intercalibration and the absolute *in situ* (γ/Z +jets) uncertainties. One point to remark is that the total cross-section relative uncertainty in a given m_{jj} bin is much higher than the total m_{jj} uncertainty in that bin. This arises because of the steeply falling dijet cross-section as a function of m_{jj} : $\sigma(m_{jj}) \propto m_{jj}^{-a}$, which results in an amplification

of the relative uncertainty:

$$\frac{\Delta\sigma}{\sigma} = a \frac{\Delta m_{jj}}{m_{jj}} \sim a \frac{\Delta p_T}{p_T}$$

where a is the slope in log scale of the falling spectra and $\Delta\sigma$, Δm_{jj} and Δp_T are the absolute uncertainties of the cross-section, m_{jj} and p_T respectively. For the last two terms of the equation we have used the approximation $m_{jj} \sim 2p_T$ which is only valid for the central region, i.e $y^* \sim 0$ or equivalently $y \sim 0$ in the center of mass frame. This can be easily obtained by setting $\theta = 0$ in Equation 5.2. By looking at the left plot in Figure 4.8, it can be seen that the relative p_T uncertainty due to the JES in the central region doubles between $p_T = 300$ GeV and $p_T = 2000$ GeV. On the other hand, the total JES uncertainty in the dijet cross-section for the central y^* bin, black curve from Figure 5.6, triples between the corresponding m_{jj} bins (i.e between $m_{jj} = 2 \times 300$ GeV and $m_{jj} = 2 \times 2000$ GeV). This is consistent with an increase in the exponent of the dijet cross-section spectra by a factor of 1.5, as shown in Figure 5.11.

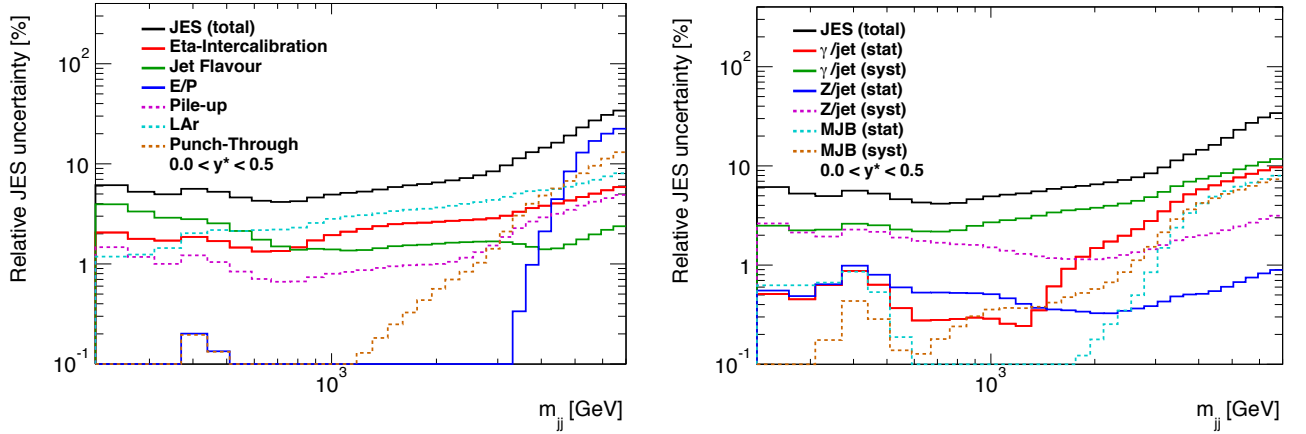


Figure 5.6: Breakdown of the relative JES systematic uncertainties in the dijet cross-section for the first y^* bin. As a reference, the black line shows the total uncertainty (sum in quadrature of each contribution).

5.6.2 Jet Energy Resolution

Because the final result relies on the resolution dependent unfolding procedure, an uncertainty on the jet energy resolution (JER) must be taken into account. It was computed based on the JER derived in 2012 and so an additional nuisance parameter (NP) was considered, which captures the changes in detector conditions from 2012 to 2015 and differences in the underlying event tune due to the increase from 8 TeV to 13 TeV in center-of-mass energy.

There are 11 JER systematic uncertainty components, which is the result of 9 NPs from Run 1, one reduced NP from the Run1-Run2 cross calibrations and one NP for the forward

region derived also in Run 1. Some of the components are positive in part of the phase-space and negative in the others, which allows to account for (anti-)correlations. This requires an enhancement of the JER in some phase-space regions and a reduction of the JER in the complementary parts. The effect of each positive NP contribution is evaluated by smearing the p_T of the reconstructed jets in the MC simulation such that the resolution is enhanced (where relevant) by the size of each uncertainty component. To do so, the jet p_T is smeared by the factor σ_{smeared} , calculated as:

$$\sigma_{\text{smeared}}^2 + \sigma_{\text{nominal}}^2 = (\sigma_{\text{nominal}} + |\Delta\sigma|)^2$$

where σ_{nominal} is the 2012 fractional resolution and $\Delta\sigma$ the resolution uncertainty for a given NP component.

Since a consistent reduction (on average) of the resolution is currently not possible to implement, an enhancement of the resolution is performed but applied to the jets used in building the transfer matrix. The enhancement of the MC JER is accounted by constructing a new transfer matrix using smeared jets. The ratio between the MC spectrum with enhanced JER unfolded with the varied transfer matrix and the nominal unfolded MC is taken as the systematic uncertainty.

5.6.3 Total Experimental Uncertainty

In order to assess the statistical precision of the systematic uncertainties, each component is evaluated using a set of pseudo-experiments from bootstrapping. The bin-to-bin fluctuations of the systematic uncertainties are minimised using a smoothing procedure (Appendix B).

The individual components of the systematic uncertainties are taken as uncorrelated and added in quadrature. The systematic uncertainties for the first and second last y^* bins are shown in Figure 5.7. The jet energy scale is the dominant uncertainty for $m_{jj} < 4000$ ($m_{jj} < 6000$) GeV in the first(second last) y^* bins. In the complementary regions, the dominant source of uncertainty is statistics. The total uncertainty on the dijet measurement is about 5% (7%) in the medium m_{jj} range of 500-1000 GeV (2000-3000 GeV) in the first (second last) y^* bin. The uncertainty increases both towards lower and higher m_{jj} values reaching 8% at low- m_{jj} (second last y^* bin) and 30% at high m_{jj} (first y^* bin).

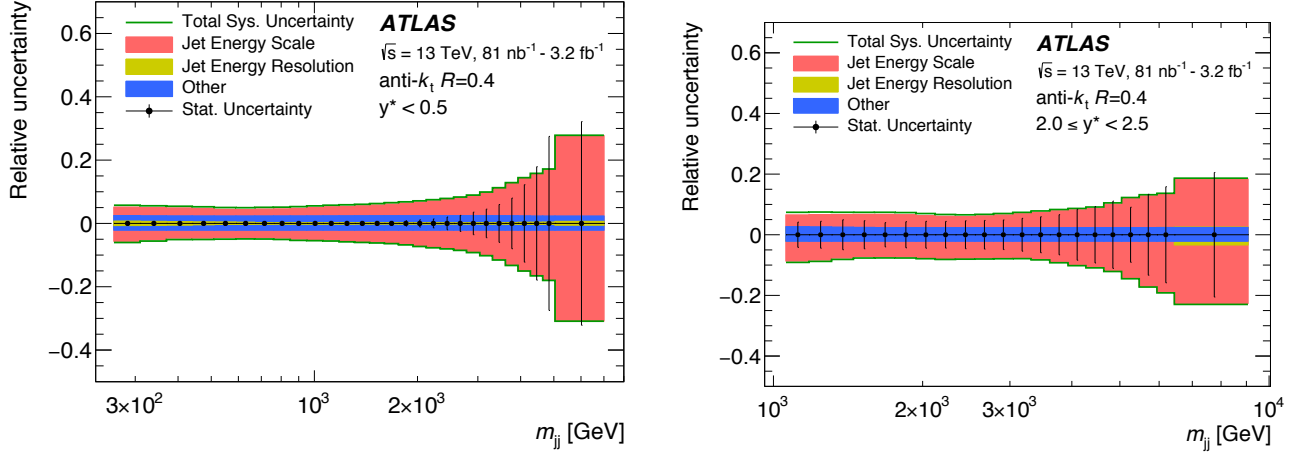


Figure 5.7: Relative systematic uncertainty for the first and second last y^* bins (left and right respectively). The total systematic uncertainty is shown as a black line. The individual uncertainties are shown in colors: the jet energy scale (red), jet energy resolution (green) and the other uncertainties (jet cleaning, luminosity and unfolding bias) are added up in quadrature (blue). The statistical uncertainty is shown as a vertical black line.

5.7 Theoretical Predictions

Theoretical predictions of the cross-sections are obtained using NLO pQCD calculations with corrections for non-perturbative (NP) and electroweak (EW) effects.

5.7.1 Next-to-leading-order pQCD calculations

The NLO pQCD predictions are calculated using NLOJET++ 4.1.3 [12] interfaced to APPLGRID [62] for fast and flexible calculations with various PDF sets and various values of the renormalisation and factorisation scales. The dijet cross-sections are calculated using the renormalisation and factorisation scales $\mu_R = \mu_F = p_T^{\max} e^{0.3y^*}$, where $p_T^{\max}$ is the leading jet p_T , as suggested in [63] and previously used in the ATLAS dijet analysis at 7 TeV [2]. The predictions are computed using several PDFs provided by the LHAPDF6 [64] library: the NLO CT14 [65], MMHT 2014 [66], NNPDF 3.0 [67], HERAPDF 2.0 [13] and the NNLO ABMP16 [68] set.

The main uncertainties in the NLO predictions come from uncertainties associated with the PDFs, the choice of renormalisation and factorisation scales, and the uncertainty in the value of α_s . The PDF uncertainties are defined at the 68% CL and propagated through the calculations following the prescription given for each PDF set, as recommended by the PDF4LHC group for PDF-sensitive analyses [69]. Calculations are redone with varied μ_R and μ_F to estimate the uncertainty due to missing higher-order terms in the pQCD expansion, as shown in Figure 2.2.

The scales are independently varied up or down by a factor of two in both directions excluding opposite variations. The envelope of resulting variations of the prediction is taken as the scale uncertainty. The uncertainty from α_s is evaluated by calculating the cross-sections using two PDF sets that differ only in the value of α_s used and then scaling the cross-section difference corresponding to an α_s uncertainty $\Delta\alpha_s = 0.0015$ as recommended in [69].

The uncertainties in the NLO QCD cross-section predictions obtained with the CT14 PDF set are shown in Figure 5.8 for first and last y^* bins. The uncertainty due to the choice of μ_R and μ_F scales is dominant in most phase-space regions, rising from 20% at about $m_{jj} = 300$ GeV in the central y^* bin to about 60% in the highest m_{jj} bin in the most forward rapidity y^* region. The PDF uncertainties vary from 2% to 20% depending on m_{jj} and y^* . The contribution from the α_s uncertainty is about 2% at low m_{jj} and negligible for the highest m_{jj} bin in each y^* .

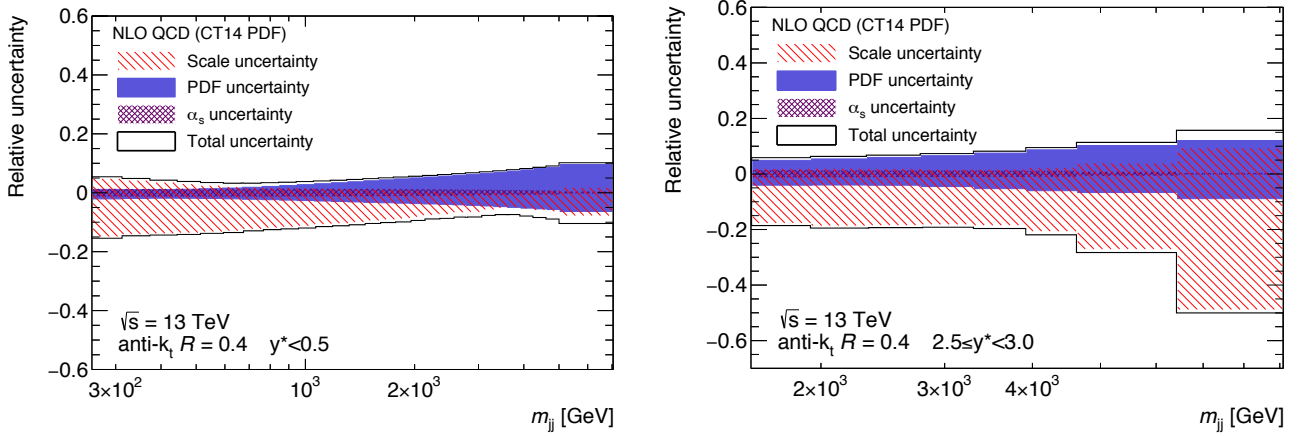


Figure 5.8: Relative NLO QCD uncertainties in the dijet cross-sections calculated using the CT14 PDF set for the first (left) and last (right) y^* bins. The uncertainties due to μ_R and μ_F scales, the α_s , the PDF and the total uncertainty are shown. The total uncertainty, calculated by adding the individual uncertainties in quadrature, is shown as a black line.

5.7.2 Non-perturbative corrections

Non-perturbative (NP) corrections are applied to the parton-level cross-sections from the NLO pQCD calculations to account for the differences between matrix-element level and particle level cross-sections. These arise mainly from hadronization and underlying event. The correction factors are calculated using LO MC event generators, as the bin-by-bin ratio of the nominal particle-level MC cross-sections to the MC cross-section derived from the partons remaining after showering, when the modelling of hadronisation and the underlying event are switched off. The correction factors are evaluated using several event generators and tunes, which are

listed in Table 5.1. The baseline correction is taken from `Pythia 8` using the A14 tune with the NNPDF2.3 LO PDF set. The envelope of all corrections is considered as a systematic uncertainty.

	CTEQ6L1 [70]	CTEQ6L1 [70]	MSTW2008LO [71]	CT10	NNPDF2.3LO	NNPDF2.3LO	CTEQ6L1 [70]
<code>Pythia 8</code>	4C [72]	AU2 [73]	A14 [15]	AU2 [73]	MONASH [74]	A14 [15]	A14 [15]
<code>Herwig++</code>	UE-EE-5 [75, 76]	UE-EE-4 [75, 76]	UE-EE-5 [75, 76]				

Table 5.1: Summary of the soft-physics model tunes used for the evaluation of the non-perturbative corrections for each event generator and PDF set.

The correction factors for a set of representative event generators and tunes are shown in Figure 5.9 in illustrative y^* bins as a function of m_{jj} . The values of the correction are in the range $0.94 - 1.01$ ($0.98 - 0.99$) at low (high) m_{jj} for the first y^* bin. For the last y^* bin, a fixed range $0.92 - 1.07$ is conservatively taken for all m_{jj} bins due to lack of statistical precision at large m_{jj} .

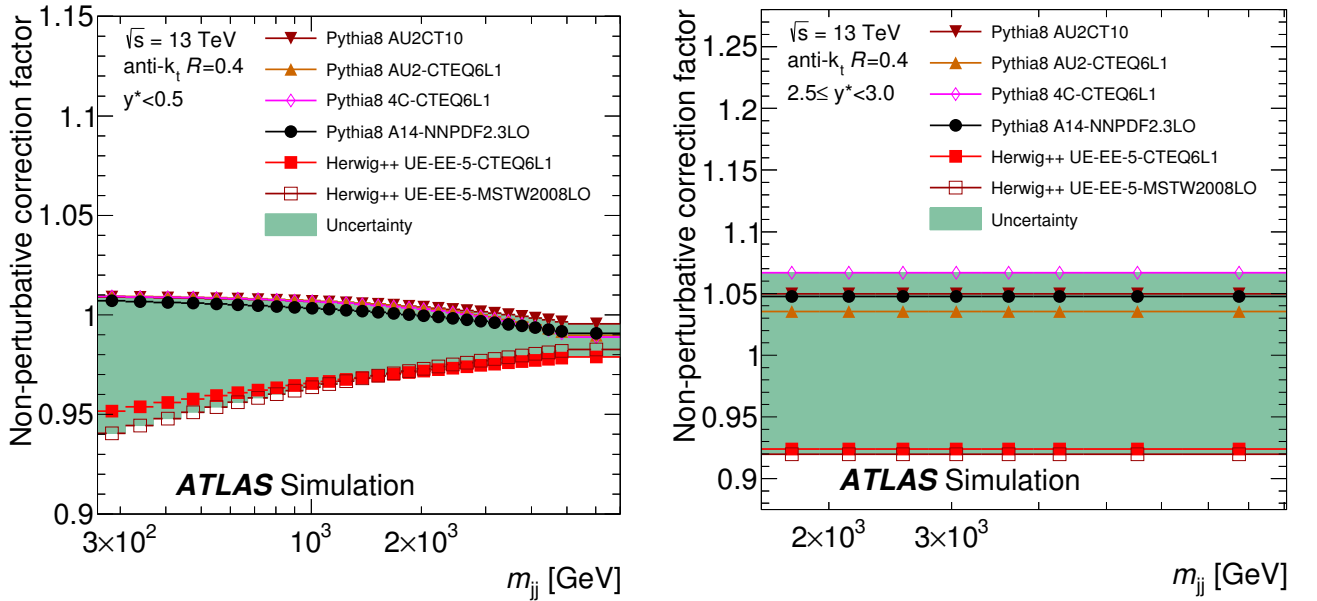


Figure 5.9: NP correction factors for the dijet NLO pQCD prediction as a function of m_{jj} for the first (left) and last (right) y^* bins. The corrections are derived using `Pythia 8` with the A14 tune with the NNPDF2.3 LO PDF set. The envelope of all MC configuration variations is shown as a band.

5.7.3 Electroweak corrections

The NLO pQCD predictions do not include the contributions of diagramans containing γ or $W^\pm/Z$ propagators. Electroweak (EW) corrections account for these effects. They are derived

using an NLO calculation of EW contributions to the LO pQCD process. The correction is defined as the ratio of a $2 \rightarrow 2$ calculation including tree-level effects of order α_s^2, α^2 , and $\alpha_s \alpha$ (from interference of QCD and EW diagrams), plus weak loop corrections of order $\alpha_s^2 \alpha$ to the LO QCD $2 \rightarrow 2$ calculation.

The correction factors are derived in the phase space considered for the measurements presented here and were specifically calculated by the authors in [77] for our measurement. No uncertainty associated with these corrections is presently estimated.

The electroweak correction factors for the dijet cross-section as a function of the m_{jj} in bins of y^* are shown in Figure 5.10. They are small for low m_{jj} , but reach 11% at $m_{jj} = 7$ TeV for the central y^* bin. For the rest of the y^* bins the correction is less than 3%.

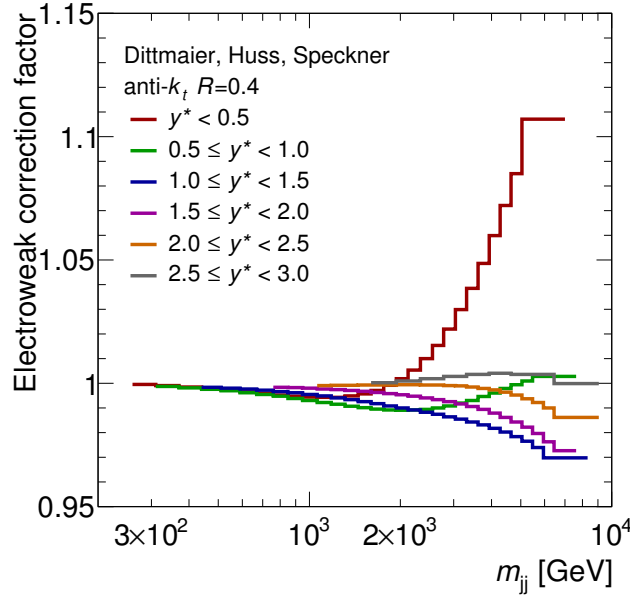


Figure 5.10: EW correction factors for the dijet cross-section as a function of m_{jj} for all y^* bins.

5.8 Results

This section presents the comparisons of the measured dijet cross-sections and the NLO pQCD predictions as a function of m_{jj} for the six y^* bins.

5.8.1 Qualitative Comparisons

The measured double-differential dijet cross-sections are shown in Figure 5.11. The measurement covers the m_{jj} range from 300 GeV to 9 TeV for $y^* < 3.0$, thus attaining a significantly

higher reach than the previous ATLAS measurement [78]. Superimposed are the NLO predictions computed with the CT14 PDF set.

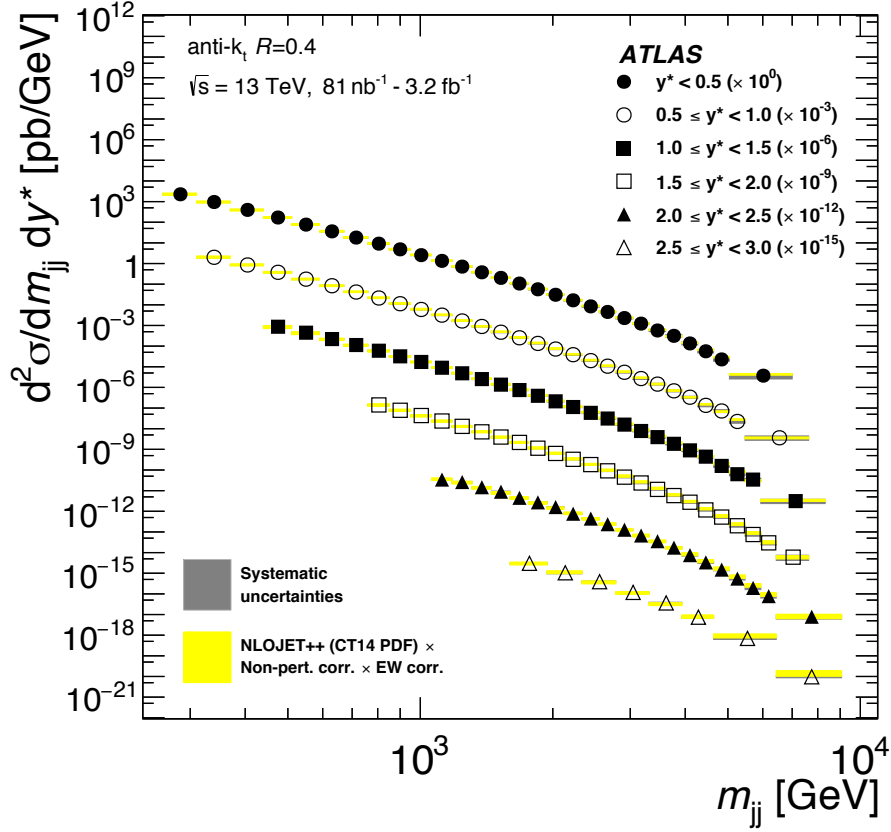


Figure 5.11: Dijet cross-sections as a function of m_{jj} and y^* , for anti- k_t jets with $R = 0.4$. The statistical uncertainties are smaller than the size of the symbols used to plot the cross-section values. The dark gray shaded areas indicate the experimental systematic uncertainties. The data are compared to NLO pQCD predictions calculated using NLOJET++ with $\mu_R = \mu_F = p_T^{\max} e^{0.3y^*}$ and the CT14 NLO PDF set, to which non-perturbative and electroweak corrections are applied. The yellow shaded areas indicate the predictions with their uncertainties. In most m_{jj} bins the experimental systematic uncertainty is smaller than the theory uncertainties and is therefore not visible.

The ratios of the NLO pQCD predictions to the measured dijet cross-sections are shown in Figure 5.12. No significant deviation of the data points from the predictions is seen, the NLO pQCD predictions and data agree within uncertainties.

5.8.2 Quantitative Comparisons

The NLO pQCD predictions, corrected for NP and EW effects, are quantitatively compared to the measurement using the method described in Appendix C. The χ^2 statistic and the corresponding observed p -value, P_{obs} , are computed taking into account the asymmetric errors and the correlations of the experimental and theoretical uncertainties. The individual components

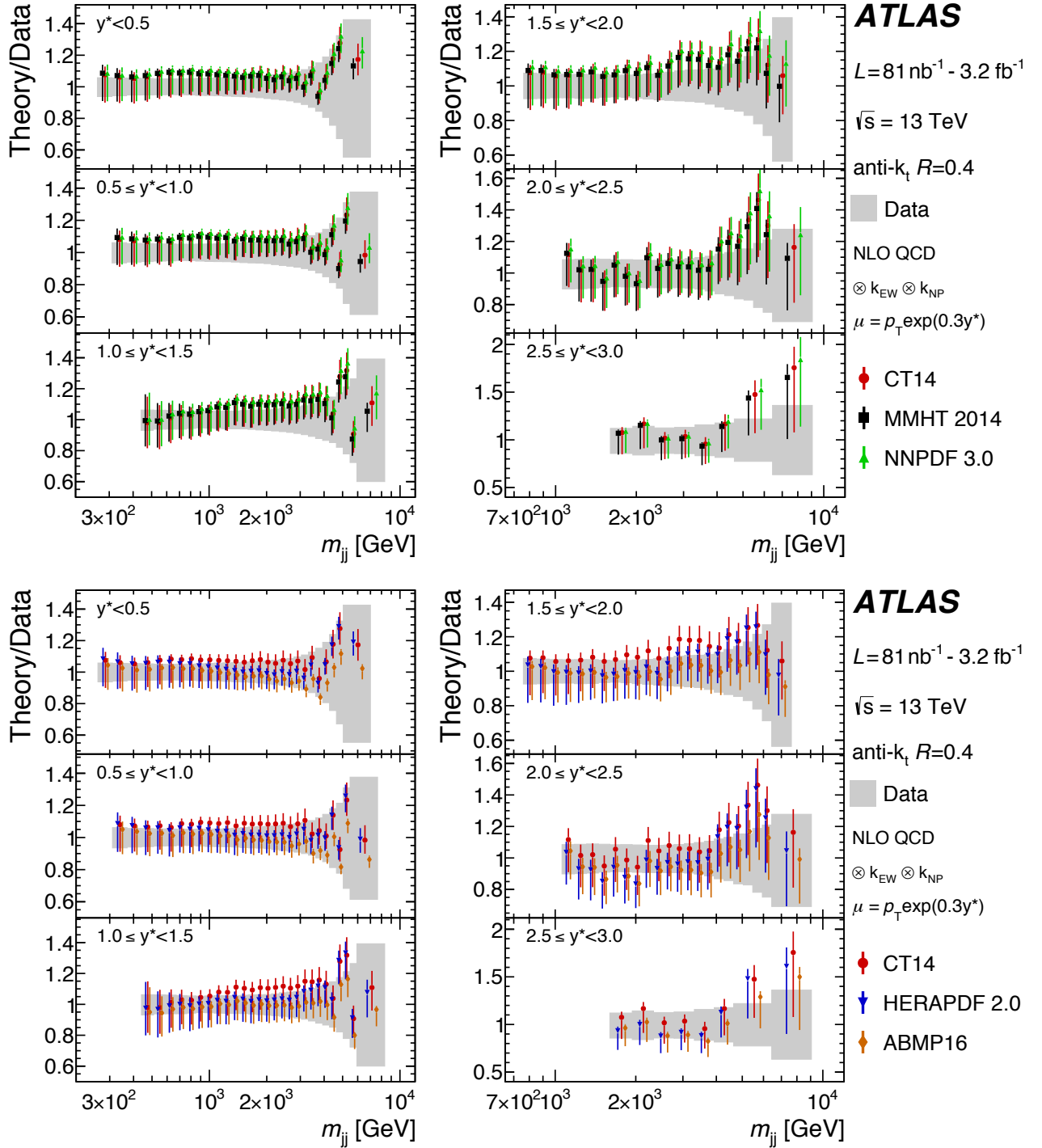


Figure 5.12: Ratios of predictions to the measured cross-sections shown as a function of the jet m_{jj} in six y^* bins. The predictions are calculated using NLOJET++ with 5 different PDF sets (CT14, MMHT 2014, NNPDF 3.0) in upper plot, and (CT14, HERAPDF 2.0, ABMP16) in lower plot. NP and EW corrections are applied. The uncertainties of the predictions are shown in coloured lines. The gray bands show the total data uncertainty including both the systematic and statistical ones.

are assumed to be uncorrelated among each other and fully correlated across the m_{jj} and y^* bins. The correlations of the statistical uncertainties across different phase-space regions are

taken into account using covariance matrices derived from 1000 pseudo-experiments obtained by fluctuating the data and the MC simulation, as described in Section 5.5.

Table 5.13 shows the summary of the observed P_{obs} values for each y^* bin of the dijet measurement, as well as those from a global χ^2 -test using all the m_{jj} and y^* bins. Fair agreement is seen (with p -values in the percent range) when considering events from individual, as well as in aggregate, y^* regions. The null hypothesis that the SM theory describes well the data is confirmed, without indication of new physics. The result is consistent with previous ATLAS measurement at $\sqrt{s} = 7$ TeV [78].

y^* ranges	P_{obs}				
	CT14	MMHT 2014	NNPDF 3.0	HERAPDF 2.0	ABMP16
$y^* < 0.5$	79%	59%	50%	71%	71%
$0.5 \leq y^* < 1.0$	27%	23%	19%	32%	31%
$1.0 \leq y^* < 1.5$	66%	55%	48%	66%	69%
$1.5 \leq y^* < 2.0$	26%	26%	28%	9.9%	25%
$2.0 \leq y^* < 2.5$	41%	34%	29%	3.6%	20%
$2.5 \leq y^* < 3.0$	45%	46%	40%	25%	38%
all y^* bins	9.4%	6.5%	11%	0.1%	5.1%

Figure 5.13: Summary of the observed P_{obs} values obtained from the comparison of the dijet cross-section and the NLO pQCD predictions for various PDF sets and for each individual y^* range. The last row of the table corresponds to a global χ^2 -test using all m_{jj} and y^* bins of the measurement.

Chapter 6

Adversarial Neural Networks for Decorrelating Jet Observables

6.1 Introduction

Hadronic decay of heavy particles, such as top quarks or W bosons, are very important processes for searches of new physics. As pointed in Section 4.3, such resonances result in collimated decay products reconstructed as a large-radius jet in the calorimeter. However, multijet production (QCD) represent a vast background which makes the identification of jets originating from heavy particle decay a crucial and non-trivial task to do.

Multivariate analysis (MVA) approaches using jet substructure variables, in particular boosted decision trees (BDT) and neural networks (NN) have been shown to be highly performant for the identification of W bosons compared to the use of single jet substructure variable discriminants [46, 47]. Nevertheless, jet substructure variables exhibit non-linear correlations with the reconstructed large- R jet mass, effectively making MVA classifiers to extract the mass as a powerful feature for discriminating against the non-resonant background, even though mass is explicitly not used as an input. As a result, fixed-value cuts on such MVA classifier observables tend to distort the shape of the large- R jet mass distribution for non-resonant jet production, making it resemble the resonance jet mass distribution [46]. Such effect is illustrated in Figure 6.1. The background distribution before the classifier cut (green) approximates the resonance mass peak (yellow) after selection is applied. This effect complicates the background estimation which results in a reduction of discovery significances. An optimal discriminant at the same time non-correlated with the jet mass is therefore desired.

In this Thesis, a novel technique based in adversarial neural networks (ANN), a network

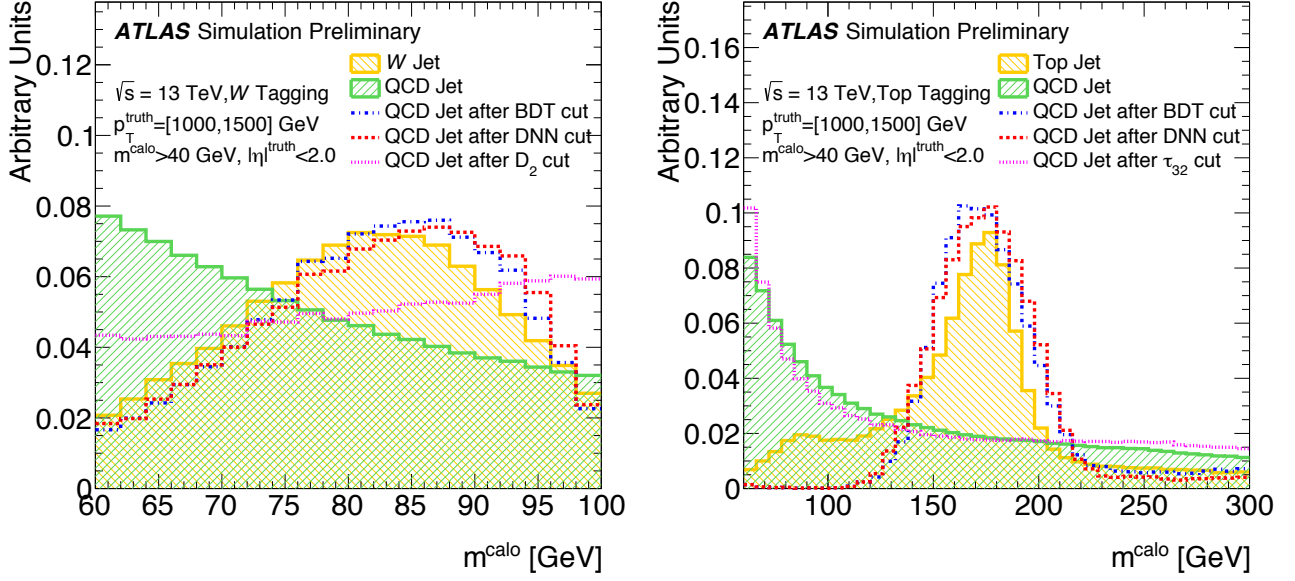


Figure 6.1: Jet mass distributions showing the signal (yellow) and the background before and after a discriminant cut on BDT, DNN and D2 (W -boson tagging) or τ_{32} (top-quark tagging) for jets with $1000 < p_T^{\text{true}} < 1500$ GeV. It can be clearly seen how the background jet mass distribution before the selection (green) resembles the signal one after classification is performed (pink, red and blue) for both W -boson (left) and top-quark tagging (right).

architecture designed by Google in 2014, is developed as a way to identify boosted W bosons and leverage classification power while reducing their correlation with the jet mass for QCD jets. As a benchmark, this method is compared against analytical single-variable decorrelated taggers, as well as MVA-based ones. In particular, designed decorrelated taggers (DDT) [79], fixed-efficiency k -nearest neighbours (k -NN) regression [80], and convolved substructure (CSS) [81] provide analytical methods for decorrelating a single jet substructure observable from the jet mass. Similarly, adaptive boosting for uniform efficiency (uBoost) [82] is used as an extension of standard boosted decision trees (BDT), as a MVA technique to manage decorrelation power as well.

This Chapter is organized as follows. Section 6.2 offers a theoretical introduction about Machine Learning and Neural Networks. The dataset used is described briefly in Section 6.3. Section 6.4 presents the metrics chosen to quantify classification and mass-decorrelation power. Section 6.5 focuses in describing the adversarial NN technique. The other decorrelated methods attempted are briefly detailed in Appendix D. Finally, results for the various taggers are presented in Section 6.7, evaluated according to the metrics selected.

6.2 Theoretical Aspects of Machine Learning

This section, presents a brief introduction to Machine Learning, with focus on neural nets, which is the main subject treated in this Chapter. This will serve in particular as the basis to the main topics developed in Section 6.5.

Machine Learning (ML) has become a very popular paradigm in computer science during the last decades. The reason of its success is mainly because, in contrast with traditional computer programs, ML does not need explicit rules for a given specific task, but it has the flexibility to *learn* its own set of rules based on experience, which might be too complex to program in a computer. What is more, ML has the properties of adaptation and generalization. The first one means that the rules it learns can adapt easily for a given change in the inputs, by just retraining on them, and the second one that it can learn, in the best case, not by memorization, and so behaving performant also for unseen data of the same nature. ML is closely related to computational statistics and mathematical optimization, which focus in making data-driven predictions, decisions, statistical and causal inferences.

ML has a broad range of applications and there exist many paradigms for learning. In this thesis we focused in *supervised learning* paradigm, which aims to predict an output when given an input vector based on example input-output pairs. In other words, it learns a function that maps an input to an output based on labeled training data consisting of a set of training examples.

Supervised learning deals with two kind of problems. In *regression* the target output is a real number. Given a training set of points x with its labels y , $(x_i, y_i)_{i=1}^n \in \mathbb{R}^M \times \mathbb{R}$, the task of the learning algorithm is to determine a function $f : \mathbb{R}^M \rightarrow \mathbb{R}$ which would be as close as possible to the corresponding y value for an x example not present in the training set. In *classification* the target output is a class label, and so $f(x)$ will predict an integer number that represents some category, $f : \mathbb{R}^M \rightarrow \mathbb{Z}$.

6.2.1 Mathematical Formulation

For simplicity, let's restrict the ML formulation to the classification problem, the rationale for regression is similar. Considering pairs of examples-categories modeled as random variables:

$$(X, Y) \sim p_{XY}(x, y), (x, y) \in \mathcal{X}, \mathcal{Y}$$

with $\mathcal{Y} = 1, 2, \dots, K$ a finite set of labels and $p_{X,Y}(x, y)$ a probability model, which in general is unknown. The goal is to find the best estimator $\hat{\xi} : \mathcal{X} \rightarrow \mathcal{Y}$, known as the *decision rule*, such

that

$$P\{\hat{\xi}(X) \neq Y\} := \int_{\mathcal{X}} p_X(x) p_{Y|X}(Y \neq \hat{\xi}(X) | X = x) dx$$

is as small as possible, where $p_{Y|X}(Y = y | X = x) = \frac{p_{XY}(x, y)}{\sum_{y \in \mathcal{Y}} p_{XY}(x, y)}$. By abuse of notation, the symbol p will denote either a probability density or mass function. Any ambiguity should be easy to resolve from the context. The *Bayes classifier theorem* guarantees the existence of the best solution for the stated problem. The optimal decision rule $\hat{\xi}^*(X)$ is to pick the hypothesis that is most probable, given by:

$$\hat{\xi}^*(X) = \arg \max_{y \in \mathcal{Y}} p_{Y|X}(Y = y | X = x)$$

The optimal decision function $\hat{\xi}^*(X)$ is also known as the *maximum a posteriori* (MAP) rule. From Bayes theorem, we have that:

$$p_{Y|X}(y|x) = \frac{p_Y(y)}{p_X(x)} p_{X|Y}(x|y)$$

It follows that if all classes $y \in \mathcal{Y}$ have uniform prior, i.e. $p_Y(y) = \frac{1}{K} \forall y \in \mathcal{Y}$, the MAP rule becomes the *maximum likelihood* decision rule. In the simplest case of binary classification, i.e. $K = 2$, the maximum likelihood solution is related with a likelihood ratio test statistic for a given significance level [83].

Since the true $p_{XY}(x, y)$ is not known in practice, otherwise no machine learning algorithm would be needed, the solution to the problem corresponds to the determination of $K - 1$ functions $\eta_i^* : \mathcal{X} \rightarrow [0, 1]$ defined as $\eta_i^*(x) := p_{Y|X}(y = i | x)$ with $i = 1, \dots, K$. For the $K = 2$ case, the optimal rule is determined by $\eta^*(x) = p_{Y|X}(y = 1 | x)$. The decision problem can be cast as:

$$\text{If } \eta^*(x) > c \text{ decide for } y = 1. \text{ Otherwise decide for } y = 0. \quad (6.1)$$

where c is a free parameter threshold which determines a given working point of the classifier. The curve defined by $\eta^*(x) = c$ is known as the *decision boundary* in $\mathcal{X}$.

In order to find these functions, a parametric model is used:

$$\alpha(x; \theta) = \sum_{l=1}^L \theta_l f_l(x) + \theta_{L+1} = \boldsymbol{\theta}^T \mathbf{f}(x), \boldsymbol{\theta} \in \mathbb{R}^{L+1}$$

We can express $\eta^*(x)$ in terms of the optimal solution $\alpha^*(x)$, as $\eta^*(x) := \sigma(\alpha^*(x))$ where $\sigma(x)$ is the *logistic sigmoid* function defined as:

$$\sigma(x) := \frac{1}{1 + \exp(-x)} \quad (6.2)$$

which ensures a result between 0 and 1. The functions $\{f_l(x)\}_{l=1}^L$ are some fixed basis (chosen in advance) where in general $L \geq M$, being M the input feature dimension. With this parametric model, known as *logistic regression*, one looks for the best function $\alpha(x, \hat{\theta})$ that explains the training data. In this sense $\hat{\theta}$ is a function of the training data $(x_i, y_i)_{i=1}^n$. By setting $L = M$ and $f_l(x) = x_l$ for $l = 1, \dots, M$, with some proper rule for selecting the coefficients $\{\theta_i\}_{i=1}^{M+1}$ in terms of the training set, we obtain the different *linear discriminant models*, e.g., Fisher discriminant rule.

Given a training set $(x_i, y_i)_{i=1}^n$ and functions $\{f_l(x)\}_{l=1}^L$, a natural approach to find the coefficients $\{\theta_i\}_{i=1}^{L+1}$ would be to minimize the *empirical risk* defined as:

$$\hat{\theta} = \arg \min_{\theta \in \mathbb{R}^{L+1}} \sum_{i=1}^n \mathbb{1}(\hat{\xi}(x_i; \theta) \neq y_i)$$

where for binary classification $\hat{\xi}$ is defined, in accordance to the decision boundary 6.1, as:

$$\hat{\xi}(x_i; \theta) = \begin{cases} 1 & \text{if } \eta(x; \theta) > c \\ 0 & \text{otherwise} \end{cases}$$

where $\eta(x; \theta) := \sigma(\alpha(x))$ is a parametric solution which is expected to be as close as possible to the optimal one $\eta^*(x)$. This criterion looks for the parameters that minimize the missclassification error in the training set. Theoretical analysis of this criterion is very important. However, from the practical point of view it is seldom used because of its computational complexity specially when L and n are large. For this reason, in practice a *surrogate loss* function is used. Assuming the training set to be independent and identically distributed random variables according to some $p_{X,Y}(x, y)$, the true conditional probability of $\{y_i\}_{i=1}^n$ given $\{x_i\}_{i=1}^n$, where y_i can take only two possible outcomes 1 or 0, can be written as:

$$p_{Y_1, \dots, Y_n | X_1, \dots, X_n}(\{y_i\}_{i=1}^n | \{x_i\}_{i=1}^n) = \prod_{i=1}^n [p_{Y|X}(1|x_i)]^{y_i} [1 - p_{Y|X}(1|x_i)]^{1-y_i}$$

As we look for $\eta(x; \theta)$ to be a good approximation to the true $p_{Y|X}(1|x)$ we can set the following as the surrogate criterion:

$$\hat{\theta} = \arg \max_{\theta \in \mathbb{R}^{L+1}} \prod_{i=1}^n \left[\sigma(\theta^T \mathbf{f}(x_i)) \right]^{y_i} \left[1 - \sigma(\theta^T \mathbf{f}(x_i)) \right]^{1-y_i}$$

By taking the *log* and multiplying by -1 , it is obtained the *empirical risk function*, also known as the *binary cross-entropy loss*:

$$\mathcal{L}(\{(x_i, y_i)\}_{i=1}^n; \theta) = - \sum_{i=1}^n y_i \log \left(\sigma(\theta^T \mathbf{f}(x_i)) \right) + (1 - y_i) \log \left(1 - \sigma(\theta^T \mathbf{f}(x_i)) \right) \quad (6.3)$$

The surrogate loss function, contrary to the $\mathbb{1}(y \neq \hat{\xi})$ function, is differentiable with respect to θ (in fact infinitely differentiable) and can be optimized efficiently using iterative methods (as *gradient descent* or the *Newton-Raphson* methods). Notice that the functions $\{f_l(x)\}_{l=1}^L$ are a fixed basis predefined beforehand. In the next section we will see how this basis can also be adapted with training data using neural nets.

6.2.2 Why Neural Nets?

From the mathematical point of view, neural nets (NNs) became extremely popular because of their capacity to estimate or learn very complex functions, as it usually happens in real optimization problems. In the previous section we considered parametric families of the form:

$$\mathcal{F} = \left\{ \sum_{l=1}^L \theta_l f_l(x) + \theta_{L+1} : (\theta_1, \dots, \theta_{L+1}) \in \mathbb{R}^{L+1} \right\} \quad (6.4)$$

where $\{f_l(x)\}_{l=1}^L$ are some fixed functions (chosen in advance). As we consider only linear combinations of $\{f_l(x)\}_{l=1}^L$, $\mathcal{F}$ is the subspace generated by those functions. If the true objective function is not in this subspace then it will not be reconstructed exactly. A possible solution would be to also let the algorithm learn the basis functions $\{f_l(x)\}_{l=1}^L$ from the data. As it will be seen shortly, this is accomplished in NNs. Let's show this in a *multilayer perceptron*, also known as *feed-forward* neural-network. We will make the functions $\{f_l(x)\}_{l=1}^L$ depend on parameters and allow to adjust them along with $\{\theta_i\}_{i=1}^{L+1}$ during training. This can be described by a series of functional operations. We start from the input $\mathbf{x} \in \mathbb{R}^M$ and construct L_1 linear combinations from them:

$$u_{1i} = \mathbf{W}_{1i}^T \mathbf{x} + b_{1i}, \quad i = 1, \dots, L_1$$

where $\{\mathbf{W}_{1i}\}_{i=1}^{L_1} \in \mathbb{R}^M$ and $\{b_{1i}\}_{i=1}^{L_1} \in \mathbb{R}$ are respectively called *weights* and *biases* of the first layer of the network to be optimized. Notice that we made the following change of notation for the parameters:

$$\boldsymbol{\theta} \rightarrow (\mathbf{W}, \mathbf{b})$$

The quantities u_{1i} are known as the *activations*. Each of the activations are transformed using a differentiable, **nonlinear** *activation function* $g : \mathbb{R} \rightarrow \mathbb{R}$, e.g. rectified linear unit (ReLU) [84], to give:

$$x_{2i} = g(u_{1i}), \quad i = 1, \dots, L_1$$

which are known as the *hidden* units, and their values are again linearly combined to obtain the activations of the second layer:

$$u_{2i} = \mathbf{W}_{2i}^T \mathbf{x}_2 + b_{2i}, \quad i = 1, \dots, L_2, \quad \{\mathbf{W}_{2i}\}_{i=1}^{L_2} \in \mathbb{R}^{L_1}, \quad \{b_{2i}\}_{i=1}^{L_2} \in \mathbb{R}$$

In a NN with just one hidden layer, L_2 corresponds to K and the values $\{u_{2i}\}_{i=1}^{L_2}$ are called the *output units activations* which conform the last layer of the network. Finally, these units are transformed by another activation function $h : \mathbb{R}^{\mathbb{K}} \rightarrow \mathbb{R}^{\mathbb{K}}$ to give the set of network predictions $\hat{y}_k$ with $k = 1, \dots, K$. For the binary classification problem the function h is a sigmoid. The output of the network $\hat{y}$ as a function of the inputs $\mathbf{x}$, for the simple case where $K = L_2 = 1$, is then:

$$\hat{\mathbf{y}} = \sigma \left(\sum_{j=1}^{L_1} W_{2j} g \left(\sum_{i=1}^M W_{1ji} x_i + b_{1j} \right) + b_2 \right)$$

The following change in notation has been made with respect Section 6.2.1: $\eta(x; \theta) \rightarrow \hat{y}$. Comparing this expression with the previous one in Eq. 6.4 it can be realized that NNs induce also a parametrization on the basis functions.

The case described above is known as the *single hidden layer network*. An illustrative workflow diagram is shown in Figure 6.2. It can be seen how the information is propagated *forward* from the input layer vector $\mathbf{x}$ to the linear combination resulting in the vector $\{u_{1i}\}_{i=1}^{L_1}$ passing through the activation function to get the hidden units vector $\mathbf{x}_2$ and so on to get finally the outputs $\{y_k\}_{k=1}^K$. When $K = 1$ we are in the binary-classification case, otherwise this belongs to the multi-classification problem.

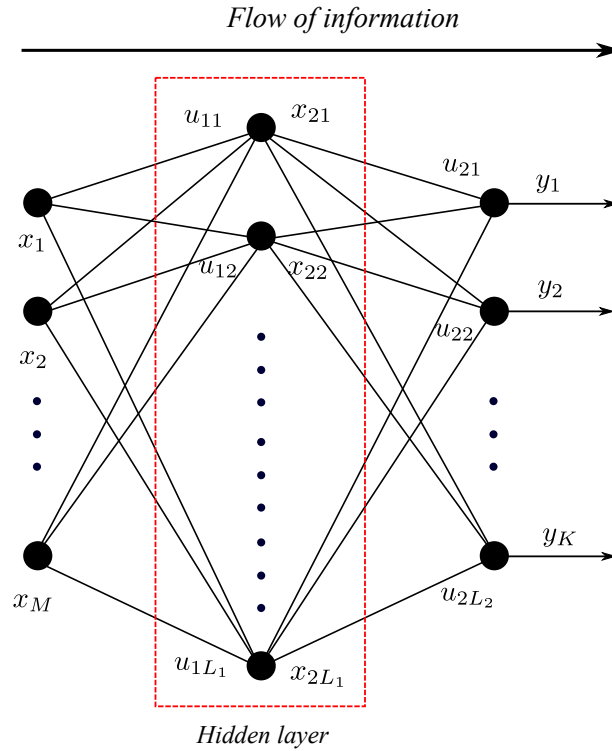


Figure 6.2: Workflow of a single hidden layer feed-forward network for a multi-classification problem.

The same procedure can be repeated through many hidden layers which results in a *deep*

neural network. A workflow representation is shown in Figure 6.3. As one can expect, the output of a deep net will be a more complex function of the inputs, given by many nested nonlinear functions:

$$\hat{\mathbf{y}} = h(g(g(\dots g(\mathbf{x}; \mathbf{W}_1) \dots; \mathbf{W}_{P-1}); \mathbf{W}_P); \mathbf{W}_{out})$$

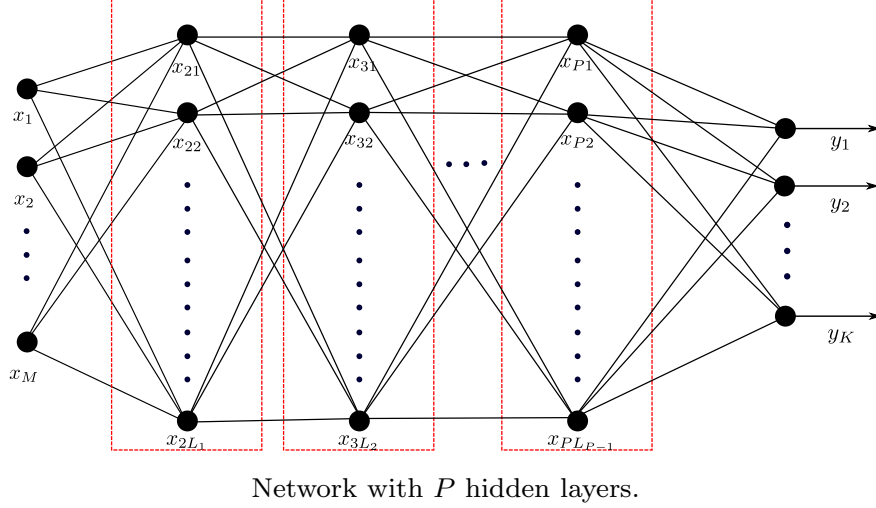


Figure 6.3: Workflow of a deep (P hidden layers) feed-forward network for a multi-classification problem.

One of the main successes of deep representations is their flexibility to learn very complex functions. Let's take as an example the problem of image recognition. Intuitively, one can think of earlier layers of the NN as detecting simple features in small portions of the image, like edges or orientation, and later layers composing them together so that it can learn more and more complex features to build up a fully reconstructed image. This visualization is indeed what happens in a neural nets used for image recognition. This type of simple to complex hierarchical or composition representation applies in other types of data than images and it was the main quality which made deep nets succeed. It is worth to add that in order to achieve this, the activation functions must be nonlinear.

6.3 Sample preparation

The study was done with MC simulation using hadronic decay of W bosons originating from exotic Z' particles as signal and multijets QCD as background simulated with `Pythia 8` [14] generator using the A14 tune [15] with the NNPDF2.3 LO PDF set [16]. The analysis was restricted to the jet mass range $50 \text{ GeV} < m < 300 \text{ GeV}$ and reconstructed jet transverse

momentum in the range $p_T \in [200, 2000]$ GeV, where boosted final states are guaranteed and where resonances searches with masses below 500 GeV are currently being performed [85]. In addition, all jets are required to have more than two constituent clusters in order to ensure that all substructure moments are well-defined. After this baseline selection, 10^6 signal and 10^6 background jets each are retained for training, ensuring balance between the classes. In addition, a separate set of 10^6 signal and 10^7 background jets are used for final performance evaluation, in order to ensure sufficient statistics for differential studies.

6.4 Evaluation Metrics

6.4.1 Classification

Classification performance is measured by the relative selection efficiency of signal jets $\varepsilon_{\text{sig}}^{\text{rel}}$ and the associated relative background rejection factor $1/\varepsilon_{\text{bkg}}^{\text{rel}}$ for different values of the cut on the classification variables. These measures are defined as:

$$\varepsilon_{\text{sig}}^{\text{rel}} = \frac{N_{\text{sig}}^{\text{tagged}}}{N_{\text{sig}}^{\text{total}}}, \quad \frac{1}{\varepsilon_{\text{bkg}}^{\text{rel}}} = \frac{N_{\text{bkg}}^{\text{total}}}{N_{\text{bkg}}^{\text{tagged}}}$$

where N^{total} refers to the total number of events (in this case, jets) of a given class passing the baseline selection and N^{tagged} refers to the number of events passing both the baseline selection and the tagging cut(s). The relative background rejection at $\varepsilon_{\text{sig}}^{\text{rel}} = 50\%$ is used as a summary metric to evaluate the classification performance, where 50% is chosen as a typical working point for jet tagging studies [46, 47], indicative of the level of selection efficiency used by searches in this kinematic regime [86, 87].

6.4.2 Decorrelation

The linear correlations between the jet mass, substructure variables, and MVA jet classifier outputs were studied in [46]. However, a linear correlation coefficient has insufficient capacity to express the highly non-linear correlations observed between the jet mass and existing MVA jet taggers [46, 47]. Instead, a figure of merit which directly quantifies the sculpting of the background jet mass distribution, resulting from a cut on a jet tagger variable, is proposed.

The Kullback-Leibler divergence [88] for discrete probability distributions P, Q is defined as:

$$\text{KL}(P \parallel Q) = - \sum_i P_i \log_n \left(\frac{Q_i}{P_i} \right) \quad (6.5)$$

where i enumerates the discrete bins of the distributions. The Kullback-Leibler divergence measures the relative entropy of P with respect to Q , and can therefore be used to quantify the similarity between discrete distributions.

To avoid divergences and to treat P and Q on equal footing, Jensen-Shannon proposed a symmetrized metric [89]:

$$\text{JSD}(P \parallel Q) = \frac{1}{2}(\text{KL}(P \parallel M) + \text{KL}(Q \parallel M)), \quad \text{with } M = \frac{P + Q}{2}$$

In this study, the Jensen-Shannon divergence is used to measure the difference between the normalised mass distributions of the background jets passing and failing, respectively, a given jet tagger cut:

$$\text{JSD} \equiv \text{JSD} \left(N_{\text{bkg}}^{\text{pass}}(m) / \sum_i N_{\text{bkg},i}^{\text{pass}}(m) \parallel N_{\text{bkg}}^{\text{fail}}(m) / \sum_i N_{\text{bkg},i}^{\text{fail}}(m) \right) \quad (6.6)$$

For brevity, this metric will be simply referred as JSD for any given signal or background selection efficiency. Using a logarithm base $n = 2$ in Equation 6.5, JSD will be in the range $[0, 1]$ with smaller values indicating a smaller relative entropy between the jet mass distributions and therefore a greater degree of mass-decorrelation; and *vice versa*. For summary performance evaluation, JSD or $1/\text{JSD}$ at $\varepsilon_{\text{sig}}^{\text{rel}} = 50\%$ will be used to quantify the mass-decorrelation of a given jet classifier.

6.5 Adversarial Neural Networks

It was proposed in [90] the use of adversarial neural networks to build a decorrelated discriminant with the jet mass. This technique is based on the *Generative Adversarial Networks* (GANs) model, a supervised learning algorithm created by Google in 2014 [91]. It was already proposed in ATLAS as a way to reduce the sensitivity with systematic uncertainties of a given predictive model [92].

GANs have become a very powerful technique within artificial intelligence and its application is growing very fast within high energy physics (HEP) and beyond. A particular example in HEP is the speed up of MC detector simulations, in order to meet the growing analysis demands, without major losses in precision [93]. It was found recently to have promising results also in signal processing, such as deconvolution of astrophysical images, demonstrating the possibility of going beyond the deconvolution limit imposed by the Shannon–Nyquist sampling theorem [94]. GANs have also been shown to successfully deal with inverse problems as the unfolding approach described in Section 5.4 [95].

This analysis presents the first use of GANs as a decorrelation method with the jet mass in ATLAS, and is a generalization of a standard classification one might do with a neural network. The problem can be stated formally as follows.

Given the data $x \in \mathcal{X}$, the target $y \in \mathcal{Y}$, and $d \in \mathcal{D}$ an observable in which the *mutual information* shared with x is different from zero, i.e $p(x, d) \neq p(x)p(d)$ where $p(x, d)$ is the joint distribution of x and d , the goal is to find a predictive function $f : \mathcal{X} \rightarrow \mathcal{Z}$ with parameters θ_f , where $z \in \mathcal{Z} = \mathbb{R}^{\dim(\mathcal{Y})}$ corresponds to the categorical scores used to make predictions in $y \in \mathcal{Y}$, such that the distribution of $f(x; \theta_f)$ is invariant with the observable d :

$$p(f(x; \theta_f) = z | d) = p(f(x; \theta_f) = z | d')$$

for all $d, d' \in \mathcal{D}$ and all values $z \in \mathcal{Z}$ of $f(x; \theta_f)$. In the particular case of our problem, x corresponds to the jet substructure observables, y the label of a jet ($y = 1$ for signal and $y = 0$ for background), z the binary classifier predictions and d the jet mass observable m . The technique discussed here is limited to continuous random variables, x and d , but it holds for discrete observables as well.

To find such an invariant in terms of a configuration in θ_f parameters, two neural networks, the second conditioned to the first one, are used. The first net is a standard classifier parametrised in θ_f . It receives N substructure variables as input and outputs a jet tagging variable z in the range $[0, 1]$. Its purpose is to minimise the classification loss L_{clf} with respect to θ_f , which introduces the correlation with the jet mass. The second net, called the adversary (adv) takes as input realizations z of $f(x, \theta_f)$ and produces as output a function modeling a posterior probability density $p(d | f(x, \theta_f) = z)$. In this sense, the adv may take any form, i.e. it does not need to be a neural network, as long as it exposes a function $p_{\text{adv}}(d | f(x, \theta_f) = z)$ of sufficient capacity to represent the true distribution $p(d | f(x, \theta_f) = z)$. A *Gaussian Mixture Model* (GMM) is proposed to model the adv posterior using a *mixture density network* [96]. The GMM parameters are found by minimising an adversary loss L_{adv} with respect to a set of weights θ_{adv} . Intuitively, if $p(f(x; \theta_f) = z | d)$ varies with d , then the corresponding correlation can be captured by the adv. In the limiting case where the classifier becomes invariant with d , the adv posterior will converge to the jet mass distribution $p(d)$, having captured the entire correlation between x and d . The adv training therefore acts as a way to constrain the parameters θ_{clf} in order to achieve decorrelation. Since we are interested in a mass-decorrelated tagger for background jets, the adversary loss L_{adv} is evaluated only for multijets, as it is the case for the mass-decorrelation metric in Equation 6.6.

The combined neural network is trained by min-max optimisation of the joint objective:

$$\min_{\theta_{\text{clf}}} \max_{\theta_{\text{adv}}} L_{\text{clf}}(\theta_{\text{clf}}) - \lambda L_{\text{adv}}(\theta_{\text{clf}}, \theta_{\text{adv}})$$

balancing the classification task (first term) with the decorrelation task (second term). The adversary loss L_{adv} depends on the weights of both networks, since the adversary’s task is conditional on the classifier configuration. The inner optimisation ($\max_{\theta_{\text{adv}}}$) leads the adversary to minimise $L_{\text{adv}}(\theta_{\text{clf}}, \theta_{\text{adv}})$ with respect to the adversary network weights θ_{adv} . The outer optimisation ($\min_{\theta_{\text{clf}}}$) leads the classifier to minimise the effective loss $L_{\text{eff}}(\theta_{\text{clf}}) = L_{\text{clf}} - \lambda L_{\text{adv}}$ with respect to the classifier network weights θ_{clf} .

The difference in sign between the two terms means that the classifier is trained to maximise L_{adv} , making it harder for the adversary to infer the jet mass distribution. The trade-off between the two competing objectives is controlled by the regularization parameter λ : For $\lambda \rightarrow \infty$, the classifier is allowed only to retain information “orthogonal” to the jet mass; for $\lambda \rightarrow 0$ the standard neural network classifier is recovered. The adversarial neural net architecture is in this way a generalization of a traditional classifier. The parameter λ can be considered as an additional hyperparameter of the adversarial neural network, to be optimised according to some use case-specific figure of merit, such as a search sensitivity metric $S/\sqrt{B}$ which would depend on a each specific analysis.

6.5.1 Implementation

The classifier objective, parametrised by weights θ_{clf} , is defined as a binary cross-entropy loss (introduced in Equation 6.3) optimised to perform binary classification of the jet labels Y based on input features X :

$$L_{\text{clf}}(\theta_{\text{clf}}) = \mathbb{E}_{x \sim X} \mathbb{E}_{y \sim Y} \left[-y \log p_{\text{clf}}(y = 1 | x, \theta_{\text{clf}}) - (1 - y) \log (1 - p_{\text{clf}}(y = 1 | x, \theta_{\text{clf}})) \right]$$

where $\mathbb{E}_{x \sim X}$ means the expected value with respect to variable x , which according to *the law of large numbers* approximates the arithmetic mean. To simplify the notation, we will denote $p_{\text{clf}}(y = 1 | x, \theta_{\text{clf}})$ as $p_{\text{clf}}(y | x, \theta_{\text{clf}})$. On the other hand, the task of the adversarial is to infer the jet mass distribution, implemented by parametrising a probability density function (p.d.f.) in m conditioned to the classifier output $z = p_{\text{clf}}(y | x, \theta_{\text{clf}})$ and an auxiliary variable $a = \log p_{\text{T}}$ ¹.

The output of the adversary is therefore the conditional probability $p_{\text{adv}}(m | z, a, \theta_{\text{adv}})$, evaluated at the mass value m for each non-resonant jet in the training sample. The adversary,

¹The use of $\log p_{\text{T}}$ as an auxiliary variable to the adv input yields robust results as a function of jet p_{T} as well as ensuring a better mass-decorrelation.

tasked with decorrelating the classifier from the general set of features D , is trained with the heuristic loss [97]:

$$L_{\text{adv}}(\theta_{\text{clf}}, \theta_{\text{adv}}) = \mathbb{E}_{z \sim p_{\text{clf}}(X, \theta_{\text{clf}})} \mathbb{E}_{d \sim D} \mathbb{E}_{a \sim A} [-\log p_{\text{adv}}(d | z, a, \theta_{\text{adv}})] \quad (6.7)$$

In practice, the posterior $p_{\text{adv}}(m | z, a, \theta_{\text{adv}})$ is constructed by specifying the coefficients $\pi_k(l)$, means $\mu_k(l)$ and widths $\sigma_k(l)$ components of a GMM, such that:

$$p_{\text{adv}}(t = m | l) := \sum_{k=1}^{N_{\text{GMM}}} \pi_k(l) \mathcal{N}(t = m | \mu_k(l), \sigma_k(l))$$

where $l = (z, a, \theta_{\text{adv}})$ represents the adversarial inputs and parameters.

The objective of the adversary is thus to find the best $\pi_k(l)$, $\mu_k(l)$, $\sigma_k(l)$ functions (from $k = 1$ to N_{GMM}), by minimising L_{adv} with respect to θ_{adv} , such that the GMM prediction of the adversary converges to the true distribution of the mass conditioned to z and a :

$$p_{\text{adv}}(d | z, a, \theta_{\text{adv}}) \rightarrow p(d | z, a) \quad (6.8)$$

From the definition of the differential entropy H of a continuous random variable x with a p.d.f. f , i.e $H(x) := \mathbb{E} [-\log f(x)]$, Equation 6.8 can be stated as:

$$L_{\text{adv}}(\theta_{\text{clf}}, \theta_{\text{adv}}) \rightarrow H(d | z, a)$$

where θ_{clf} is kept fixed.

The classifier and adversary neural networks are connected in a single, feed-forward architecture as shown in Figure 6.4. Both models are constructed as densely connected networks in `keras` 2.1.5 [98] using the `tensorflow` 1.4.1 backend [99].

The classifier network consists in three hidden layers, each with 64 nodes with rectified linear unit (ReLU) activation [84], and a single output node with sigmoid activation (Equation 6.2) which represents the classification probability.

The adversary consists in a single hidden layer comprising 64 nodes with ReLU activation, parametrising the GMM posterior p.d.f. $N_{\text{GMM}} = 20$ components was found to have sufficient capacity to perform the decorrelation. To model the mixing coefficients $\pi_k(l)$ a *softmax* activation function [101] was used, which ensures that $\sum_k \pi_k(l) = 1$. For the means and widths, the *tanh* and *softplus* activation functions were used, both making sure that $-1 < \mu_k(l) < 1$ and $\sigma_k(l) > 0$, respectively.

For the implementation of the joint optimisation, a gradient reversal layer [100] is used. It consists in applying a gradient scaling operation to the connection between the classifier and the

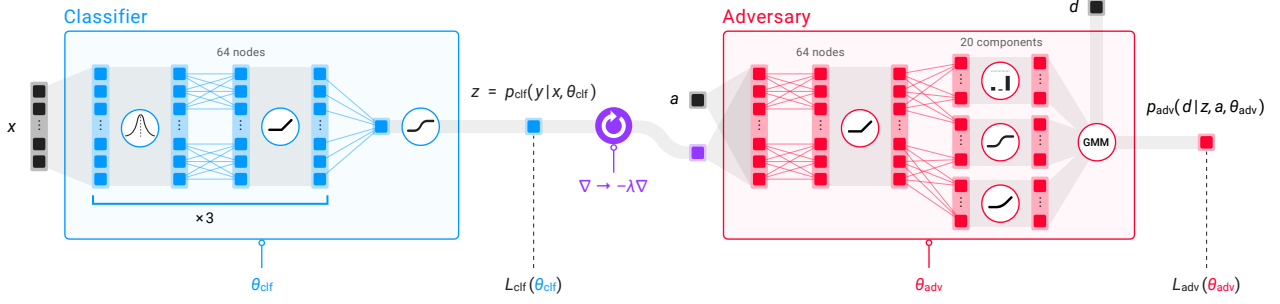


Figure 6.4: Adversarial neural network architecture. The classifier network is tasked with predicting jet labels (y) based on jet substructure variable inputs (x), outputting a tagger variable (z). The adversary network is tasked with inferring a probability density function of the variables from which the classifier is to be decorrelated (d ; here the jet mass m), optionally aided by auxiliary features (a ; here $\log p_T$), by parametrising a posterior p.d.f. as a Gaussian mixture model (GMM). The adversarial training was technically implemented in keras using a gradient reversal layer [100] shown in purple. The trade-off between L_{clf} and L_{adv} is controlled by the parameter λ .

adversary. In the feed-forward mode (computation of the loss function) it acts as the identify operations and during back-propagation (computation of the derivatives of the loss with respect to θ parameters) the gradient from the adversary is scaled by $-\lambda$. The minus sign means that the gradient flowing backwards from adversary has the inverse effect for the classifier weights as for the adversary. That is, whereas the gradient acts on θ_{adv} to minimise L_{adv} , it will have the effect of maximising L_{adv} for θ_{clf} . The pseudo-code of the joint optimisation algorithm is described in Section 6.6.2.

6.5.2 Feature Engineering

The ten jet substructure variables used in the training of the neural network-based jet classifier are listed in Table 4.1. This choice was motivated from optimisation studies on feature selection performed in [47] due to the similarity of the problem and the datasets used. In contrast to [47], the jet mass and p_T are not used as input features for the classifier. The reason for not using the jet mass is obvious, on the contrary, jet p_T could have been used since the training was performed on a flat jet p_T spectrum.² Using only substructure moments as input features was intended to provide a more direct comparison with the analytical taggers described in Appendix D, all of which are based on single substructure moment as their base observable.

To validate this selection of inputs features, a comparison in terms of classification power

²Classifiers trained on the physical jet p_T distribution, would depend on a particular physics process, resulting unusable in a different physics analysis.

was performed between a neural network trained on the inputs from Table 4.1 and one trained on an expanded set of 20 substructure variables, including the nominal ten. No significant improvement in performance was found. Similarly, no improvements were found by performing a principal component analysis (PCA) [102] on the expanded set and retaining 13 or 10 features.

6.5.3 Hyperparameters Tuning

The optimal architecture and configuration of the neural network, such as learning rate, hidden layers, nodes per hidden layer, etc. for both the classifier and adversary, collectively referred to as Hyperparameters (HP), are chosen using Bayesian Optimisation (BO) implemented with **Spearmin** [103]. In contrast with standard HP search techniques, such as grid search or random search, BO keeps track of past evaluation results to build a probabilistic model in order to propose a better HP candidate for the next evaluation, based on an expected improvement criterion quantified by the probabilistic model. The idea behind is that even though the algorithm spends more time selecting the next HP by maximizing the expected improvement, this is much cheaper in terms of computational cost than evaluating the model randomly many times. In other words, BO reduces the chances of spending time evaluating poor hyperparameter choices, being able to find better HP than random search in the same number of trials. However, BO has no convergence criteria and is not guaranteed to yield the most optimal choice, but it is capable to find efficiently an optimal configuration with few evaluations. The probability model selected was based on *Gaussian Processes* [104]. A more detailed explanation on BO can be found in [105].

In order to evaluate the goodness for each HP configuration, an objective function is needed.

The classifier is optimised according to the classification loss L_{clf} . In order to ensure stability of the results, a *3-fold cross-validation* technique was used both for the classifier and adversary (hyper)parameters optimisation, as well as for the joint optimisation (Sec. 6.6.2). This is used to prevent *overfitting*, i.e behaving performant in the training set but poorly in the test set. N -fold cross-validation is widely used to avoid this issue and consists in splitting the training set in N contiguous subsamples (folds). The training and evaluation is done from scratch N times for each partition, consisting in $N - 1$ subsets for training and the remaining one for testing. This results in N predictions which are then averaged and their standard deviation computed in order to get the final result. This makes the choice of HP more robust to the test set evaluation. The training was done for 50 *epochs*, one epoch defined as one pass-through of the entire training dataset. In addition, between each epoch the order of the training samples

were shuffled. The mean and standard deviation of the 3 validation splits ($L_{\text{clf}}^{\text{val}}$) are evaluated, and the optimisation metric chosen as the mean plus one standard deviation.

The HP of the adversary network, in contrast with the stand-alone classifier optimisation, cannot be meaningfully optimised according to the loss in Equation 6.7 alone. This loss measures only the capacity of the adversary to construct the posterior p.d.f. for the jet mass, not the quality of the resulting decorrelated tagger. Therefore, to tune the adversarial neural network, the optimisation is performed by maximising the metric $1/\varepsilon_{\text{bkg}}^{\text{rel}} + \lambda/\text{JSD}$ computed at $\varepsilon_{\text{sig}}^{\text{rel}} = 50\%$ for a fixed λ , chosen to be $\lambda = 10$ since this value is found to yield robust results for mass-decorrelation, as it will be shown in Section 6.7.3. The choice of this metric allows to capture the joint optimisation nature of the problem. Throughout the optimisation of the adversary, a classifier pre-trained with optimal HP is used.

The optimisation process, run for 100 bayesian HP configuration evaluations in order to check stability for extended training duration, is shown in Figure 6.5.

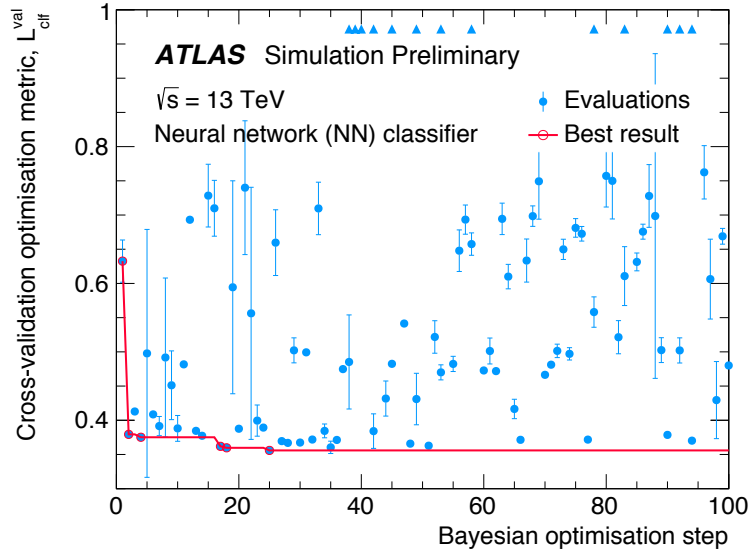


Figure 6.5: Neural network classifier Bayesian hyperparameter optimisation. Blue markers indicate the mean and standard deviation for each evaluation. The red line indicates the running optimisation metric minimum. Open red markers indicate improvements. Triangular markers indicate evaluation metrics outside of the axis range.

6.6 Training Procedure

6.6.1 Pre-training

The training of both classifier and adversary nets was performed on a cluster of NVIDIA Tesla K80 GPUs. To give adequate initial conditions to the adversarial joint optimisation, the classifier and adversary are first pre-trained stand-alone with the chosen hyperparameter configuration using 3-fold cross-validation.

First, the classifier is pre-trained for 200 epochs. The evolution of the classifier loss L_{clf} as a function of the number of training epochs is shown in Figure 6.6.

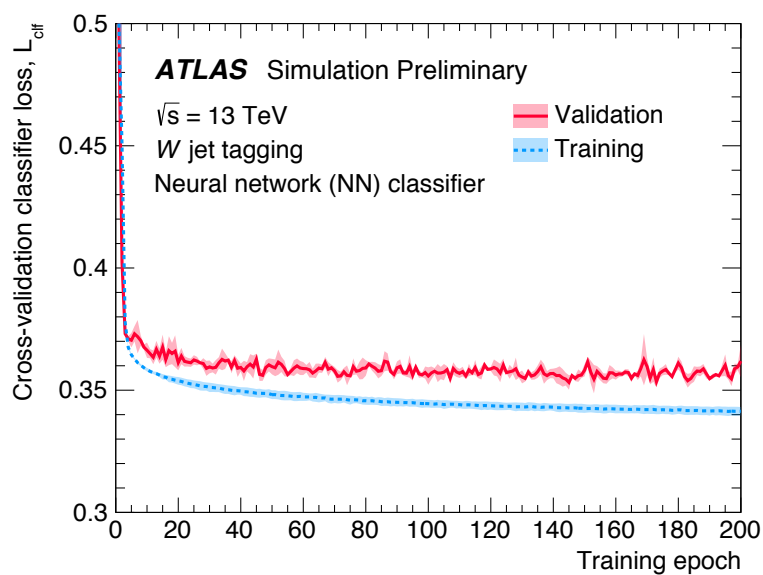


Figure 6.6: Stand-alone neural network classifier loss during 3-fold cross validation training as a function of the number of training epochs, for training and validation splits. Lines indicate mean loss across folds. Shaded bands indicate standard deviation of losses across folds.

L_{clf} decreases monotonically during training, for both training and cross-validation splits, indicating no overfitting given the chosen hyperparameter configuration. In addition, the final classifier loss is comparable to [46], suggesting that the classifier network architecture is performant.

The adversary is then pre-trained for 10 epochs at the fixed θ_{clf} configuration found previously, resulting in a minimal perturbation around the optimal neural network classifier. The adversary is allowed in this way to condition its posterior on the pre-trained classifier.

6.6.2 Joint Optimisation

The pseudo-code of the joint optimisation procedure is described in Algorithm 1. The network operates on a jet-by-jet basis. It takes as inputs the jet substructure features x , the jet class labels y , the jet mass d and the auxiliary variable $a = \log p_T$. The outputs are the classifier probability of a jet belonging to signal, $z = f(x; \theta_{\text{clf}})$, and the adversary posterior from the GMM, $p_{\text{adv}}(d|z, a)$. In order to achieve a faster convergence, the parameters θ_{clf} and θ_{adv} are updated per batches consisting in $M = 8192$ jets for training (called a minibatch). First, the gradient of the adversary loss with respect to θ_{adv} is computed (θ_{clf} are fixed). Next, the adversary parameters are updated K times until convergence is achieved (here we assume that K minibatches is enough to make the adversary converge). Finally, for a given minibatch (t), θ_{clf} are updated after computing the gradient of the effective loss at fixed θ_{adv} . The procedure is repeated T times until achieving a classifier invariant with the jet mass.

Algorithm 1 Adversarial training of a classifier clf against an adversary adv

```

▷ %Classifier inputs:  $\{x_i, y_i\}_{i=1}^N$ , Classifier outputs:  $z = f(x; \theta_{\text{clf}})$ %
▷ %Adversary inputs:  $\{z_i, a_i, d_i\}_{i=1}^N$ , Adversary outputs:  $p_{\theta_{\text{adv}}}(d|z, a)$ %

▷ %Loop over  $T$  minibatches%
for  $t = 0$  to  $T$  do
  ▷ %Loop over  $K$  minibatches%
    for  $k = 0$  to  $K$  do
      ▷ %Sample minibatch  $\{x_m, d_m, z_m\}_{m=1}^M$  of size  $M$ %
      ▷ %With fixed  $\theta_{\text{clf}}$ , compute adversary loss gradient with respect to  $\theta_{\text{adv}}$  and update  $\text{adv}$  parameters:%
        
$$\nabla_{\theta_{\text{adv}}} L_{\text{adv}}(\theta_{\text{clf}}, \theta_{\text{adv}}) := -\nabla_{\theta_{\text{adv}}} \sum_{m=1}^M \log p_{\theta_{\text{adv}}}(d_m|z_m, a_m)$$

        
$$\theta_{\text{adv}} = \theta_{\text{adv}} - \alpha_{\text{adv}} \nabla_{\theta_{\text{adv}}} L_{\text{adv}}$$

      end for
      ▷ %Sample minibatch  $\{x_m, y_m, d_m, z_m\}_{m=1}^M$  of size  $M$ %
      ▷ %With fixed  $\theta_{\text{adv}}$ , compute effective loss gradient with respect to  $\theta_{\text{clf}}$  and update  $\text{clf}$  parameters:%
        
$$\nabla_{\theta_{\text{clf}}} L_{\text{eff}} := \nabla_{\theta_{\text{clf}}} \sum_{m=1}^M [L_{\text{clf}}(\theta_{\text{clf}}) + \lambda \log p_{\theta_{\text{adv}}}(d_m|z_m, a_m)]$$

        
$$\theta_{\text{clf}} = \theta_{\text{clf}} - \alpha_{\text{clf}} \nabla_{\theta_{\text{clf}}} L_{\text{eff}}$$

    end for
  end for

```

Notice that in general T is not equal to K , since the number of updates (at fixed θ_{adv}) the

clf needs to achieve invariance with the mass is not necessarily the same number of updates (at fixed θ_{clf}) the adv needs to converge. Another point to remark is that different step sizes of the gradient (learning rates), α_{clf} and α_{adv} , were used to update the clf and adv parameters. Since the adversary is tasked with estimating the parameters of a large number of GMM components, convergence will necessarily be slower than for the classifier, a difference in learning rate that must be account for. As a result, we used the following relation $\frac{\alpha_{clf}}{\alpha_{adv}} = 2 \times 10^{-7}$. This ratio was treated as an hyperparameter of the network.

The training dynamics of the classifier, adversary and effective losses are shown in Figure 6.7. As it can be seen, during the adversary pre-training the classifier is kept fixed and thus the classifier loss L_{clf} remains constant. At the same time L_{adv} decreases monotonically from 0, corresponding to the uniform distribution between 0 and 1 which represents random guessing of the jet mass³, up to a minimum. After the adversary pre-training, the two networks are trained simultaneously.

L_{clf} gets worse to allow for the adversary being unable to infer the jet mass, thus rising at the same time L_{adv} . In other words, the classifier makes the variable z to be more independent with m at the cost of loosing classification power. The net effect results in a minimisation of the effective loss $L_{clf} - \lambda L_{adv}$ as it can be seen in the last panel of the figure.

Ideally, for every parameter update of the classifier, the adversary should be allowed to fully converge (as presented in Algorithm 1). In practice, the optimisation could typically be done using alternating [91] or simultaneous [106] gradient updates of the classifier and adversary weights. Since the latter was found to lead to better convergence, it was chosen as the method for the joint neural network training.

In the limit of full mass-decorrelation, the variable z will be independent with m and so the adversary posterior will be equal to the prior, i.e the mass conditioned to the auxiliary variable a p.d.f. $p(m|a)$. In other words, updates in θ_{clf} will produce the conditional entropy $H(m|z, a)$ to reach its maximum at $H(m|a)$.

The optimal classifier $\hat{f}(x, \theta_{clf})$, if it exists, corresponds to fulfill both classifier and adversary ideal situations, where $\hat{f}$ is either an optimal and invariant classifier, i.e $L_{clf} = H(y|x)$ and $L_{adv} = H(m|a)$. In this case the following relation holds:

$$H(y|x) - H(m|a) \leq L_{clf} - H(m|z, a)$$

where the equality is attained for the optimal $\hat{\theta}_{clf}$. In practice, the assumption of existence of an optimal and invariant classifier may not be valid, resulting in a strict lower bound. In

³The mass was scaled to be between 0 and 1.

particular, the deviation between the final L_{adv} and $H(\text{prior})$, as seen in Figure 6.7, could be due to either a trade-off between L_{adv} and L_{clf} that is beneficial to the overall loss and/or a lack of capacity in the parametrisation of the adversary posterior p.d.f. as a Gaussian Mixture Model. Nevertheless, a situation very close to the ideal case was attained.

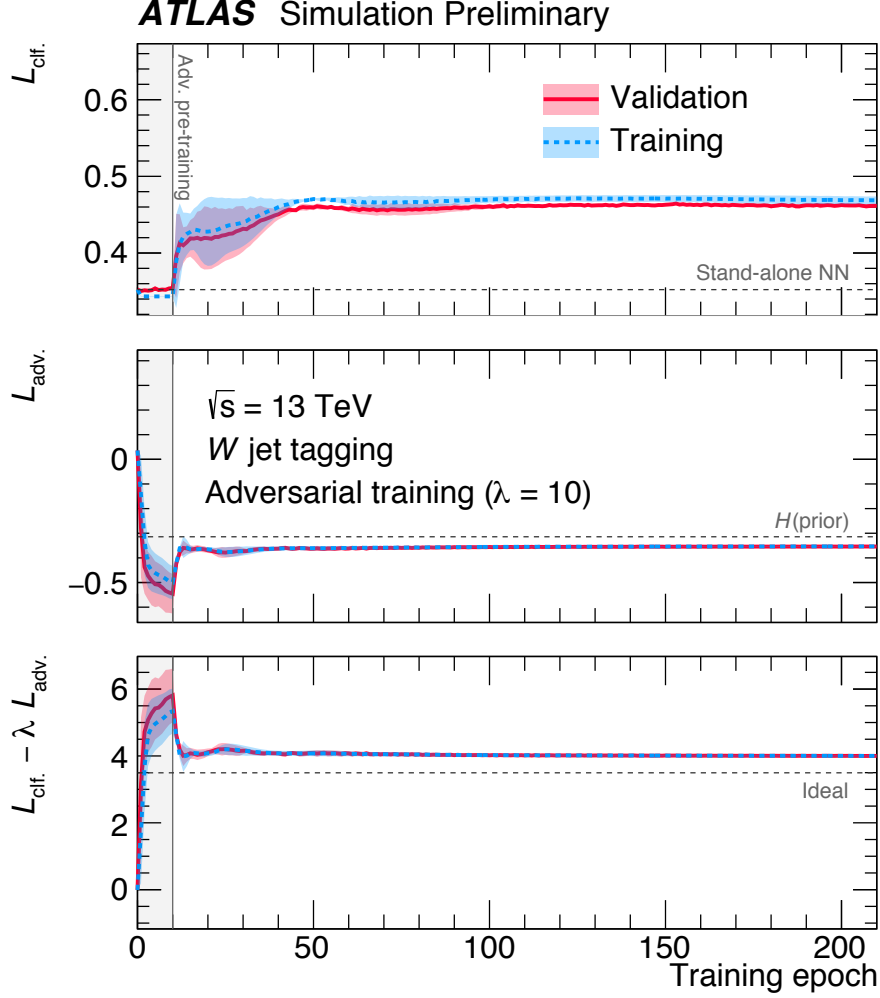


Figure 6.7: Classification (top), adversary (middle), and effective classifier losses (bottom) associated with the adversarial neural network tagger during 3-fold cross validation training as a function of the number of epochs, for training and validation splits. The first 10 epochs are spent on pre-training the adversary network (grey band). Lines indicate mean loss across folds, shaded bands indicate standard deviation of losses across folds. References for the optimal (stand-alone) NN classifier loss, the entropy H of the adversary prior, and the ideal, effective loss are shown by the grey dashed line.

6.7 Results

The various mass-decorrelated jet taggers (Appendix D), together with the adversarial network treated in the previous section, result in single discriminant variables which classify jets as

either W jets or non-resonant multijets while keeping the shape of the background jet mass distribution unaltered to a certain degree.

As a representative example, the distributions of the NN and ANN tagger discriminants are shown in Figure 6.8 for signal and background jets. It can be seen that the NN tagger powerfully separates the two classes of jets, while for the ANN case this happens to a lesser level as a result of the mass-decorrelation compromise.

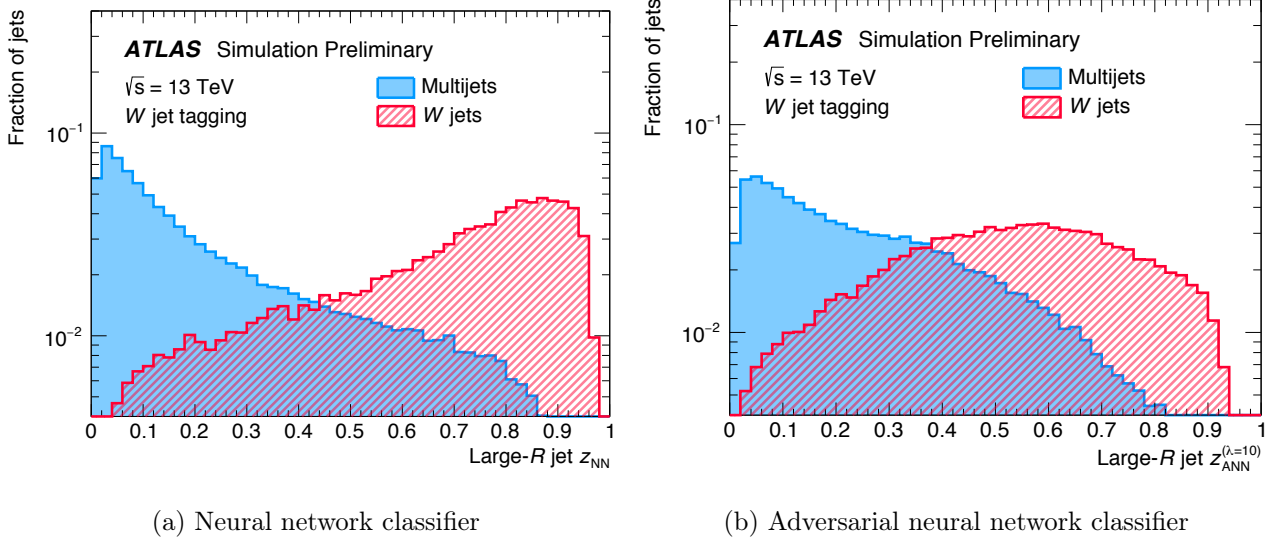


Figure 6.8: Distributions of the jet tagger variables for the standard neural network classifier and the adversarially trained neural network classifier, for multijets and W jets.

The performance of each tagger is evaluated for both classification and mass-decorrelation metrics. The applications for which mass-decorrelated taggers are immediately useful are typically characterised by a jet p_T determined by the photon and jet trigger thresholds, which are around 200 and 500 GeV, respectively. Therefore, results are studied in two different p_T bins: $[200, 500]$ GeV and $[500, 1000]$ GeV.

6.7.1 Classification

The classification performance, displayed in terms of receiver operating characteristic (ROC) curves, shows for each tagger in Figure 6.9, clearly a deterioration in the discrimination power of the ANN with respect to the NN tagger.

For the $200 \text{ GeV} < p_T < 500 \text{ GeV}$ case, in Figure 6.9a, the best discrimination power is achieved by the standard MVA taggers, NN and AdaBoost, characterized by the background rejection at a fixed signal efficiency of $\epsilon_{\text{sig}}^{\text{rel}} = 50\%$. The two standard MVA taggers are seen to

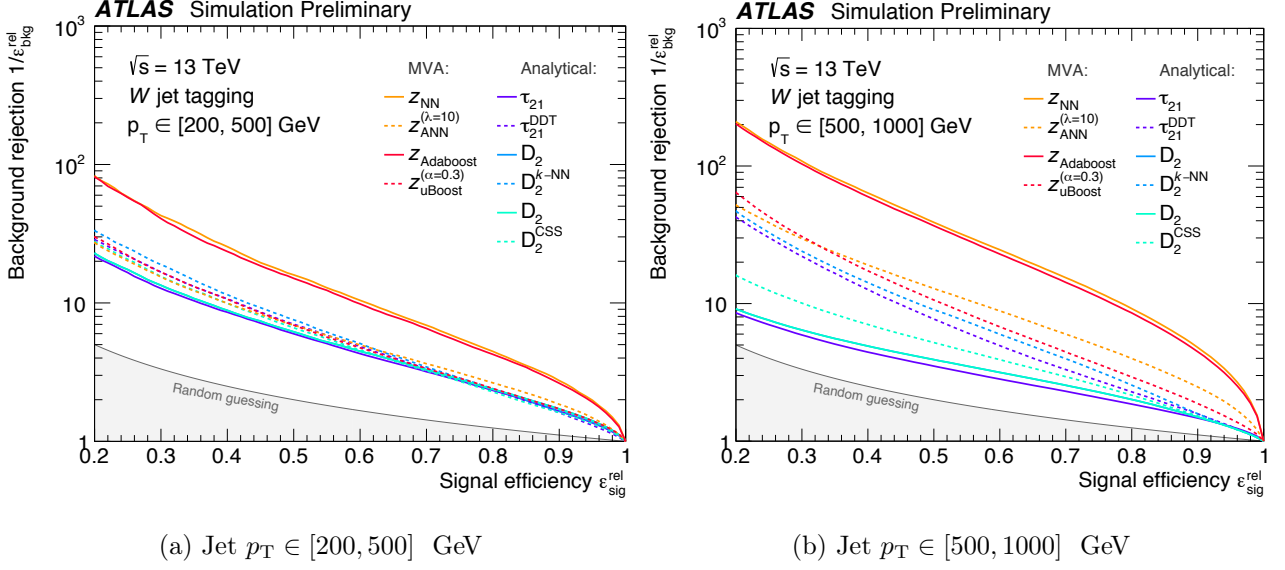


Figure 6.9: Rejection of multijets (background) as a function of W jet (signal) selection efficiency, for standard and mass-decorrelated versions of analytical and multivariate (MVA) jet taggers in two regions of jet p_T .

yield background rejections which are up to four times larger than those of the standard, single-variable taggers. In this kinematic region, for a particular choice of decorrelation parameters ($\lambda = 10, \alpha = 0.3$), all mass-decorrelated taggers are seen to perform almost identically in terms of classification.

Among the analytical mass-decorrelation methods, k -NN is seen to provide the best classification; better than CSS, even though both based on the D_2 substructure variable [48].

For all single-variable taggers, the observed background rejection after mass-decorrelation is greater than for the original taggers, an effect which is particularly appreciated in the high- p_T bin. The same effect is also seen in the low- p_T bin, although to a much smaller extent. In the case of DDT, an instructive parallel is provided by the observation that the linear transformation acts similarly to a Fisher discriminant, extracting additional information from ρ^{DDT} which slightly improves classification.

The standard single-variable taggers are seen to lose classification power with p_T , which is an expression of relative shifts of the substructure variable distributions for multijets and W jets as a function of p_T . On the other hand, since the DDT and k -NN methods both decorrelate the substructure variable not only from mass, but also from p_T through ρ and ρ^{DDT} , the fact that these methods recover almost exactly the lost performance by removing the p_T -dependence is not surprising. Indeed, the k -NN method removes the mass and p_T -dependence by construction at $\epsilon_{\text{sig}}^{\text{rel}} = 50\%$.

In contrast to the mass-decorrelated single-variable taggers, the mass-decorrelated MVA taggers exhibit a considerable decrease in classification performance relative to their standard variants. This effect is expected and acceptable as part of a trade-off between the two cases. In addition, the greater initial degree of correlation (see Figure 6.10 below) with the jet mass observed for standard MVA taggers compared to the standard single-variable taggers means that there is more potential for degradation of classification power.

6.7.2 Decorrelation

The most direct measure of the mass-decorrelation effect produced for each tagger, is the inspection of the normalised multijet mass distribution before and after the cut on the discriminant variable. Such comparisons, together with the signal mass distribution before tagging, are shown in Figure 6.10.

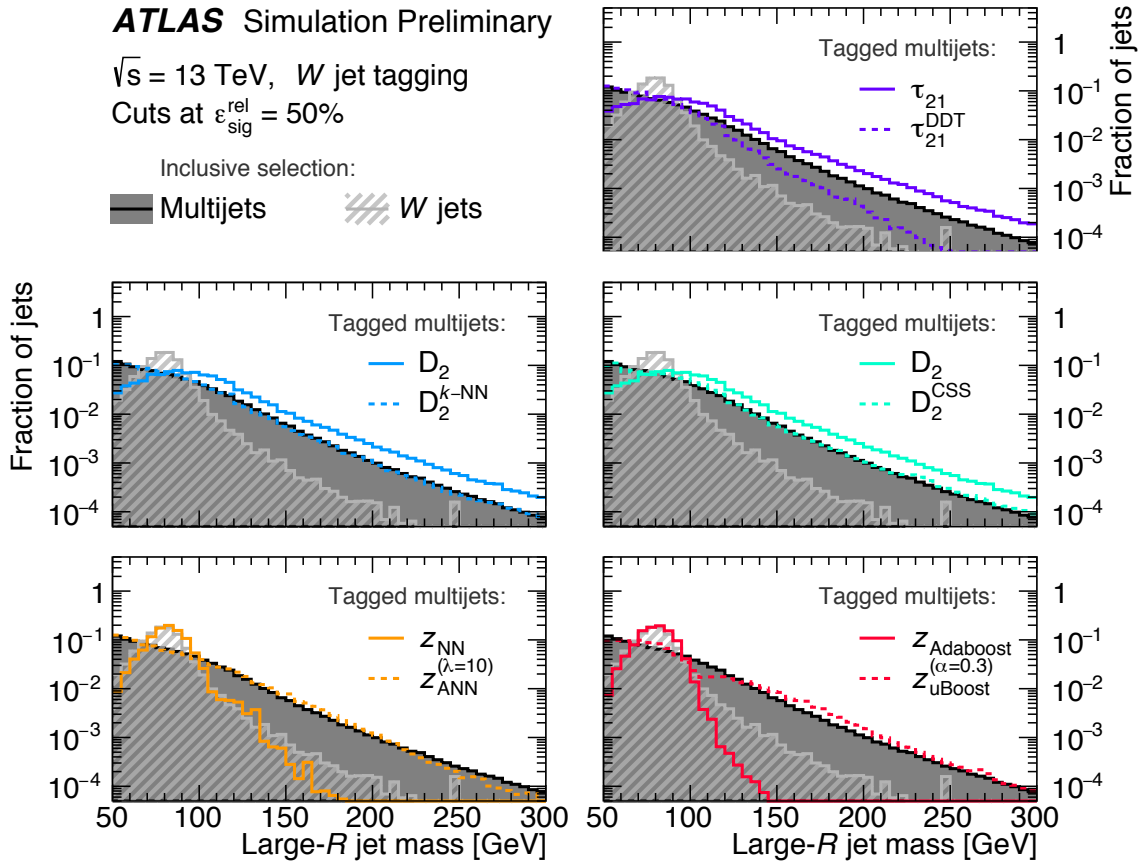


Figure 6.10: Normalised jet mass distribution for inclusive multijets before cuts, compared to the same distributions after cuts on the studied jet discriminants. Also shown for reference is the jet mass distribution for W jets before tagging. Cuts are chosen to correspond to a W jet (signal) selection efficiency of $\varepsilon_{\text{sig}}^{\text{rel}} = 50\%$.

Cutting on each of the standard taggers sculpts the background jet mass distribution, thereby introducing artificial structures not present in the original spectrum. Such sculpting may directly impact the sensitivity of physics searches by reducing the ability to constrain systematic uncertainties on e.g. multijet backgrounds. In particular, the standard MVA taggers sculpt the background jet mass distribution to resemble the W jet mass peak (solid red and orange lines).

Each of the mass-decorrelation procedures mitigates such sculpting. For the mass-decorrelated single-variable taggers, CSS and in particular k -NN regression are both seen to remove most sculpting effects. The DDT transform removes the sculpting in the lower jet mass region, while some disagreement persists at higher masses as a result of the limitations of the assumption of linearity underlying the method.

The mass-decorrelation of both MVA taggers is considerably improved following the application of their respective training methods for decorrelation. In particular, the ANN tagger yields a largely smooth, un-sculpted distribution of multijet events passing the cut. On the other hand, the mass distribution after the cut on $z_{\text{uBoost}}^{(\alpha=0.3)}$ retains some residual sculpting, particularly around $m \approx 100$ GeV.

Similar residual sculpting is not observed for the ANN tagger. This is evident from the local background efficiency as a function of the jet mass, for a range of inclusive background efficiencies shown in Figure 6.11 for the neural network taggers. Figure 6.11a shows the sculpting of the multijet background around the W jet mass peak, and the mass-decorrelation effect is evident in Figure 6.11b, where the background efficiency profiles are roughly uniform as a function of the jet mass. The deviation from uniformity in Figure 6.11b, most evident between jet masses of 150 and 200 GeV, reflects the residual correlation with the jet mass, which may be reduced by more fine-tuned optimisation and longer adversarial training.

Another advantage of ANN is that, in contrast with uBoost, it provides uniform selection efficiencies for all inclusive selections, i.e not only at $\varepsilon_{\text{sig}}^{\text{rel}} = 50\%$.

The quantitative summary metric for the mass-decorrelation, JSD, is shown in Figure 6.12 as a function of the background efficiency at which the cut is performed. As expected, the JSD values for the mass-decorrelated taggers are consistently and considerably lower — i.e. have lower degrees of correlation with the jet mass — than the standard ones.

k -NN regression is seen to lead to the greatest degree of mass-decorrelation, especially in the vicinity of the background efficiency percentile at which the regression is performed (16%). The other methods exhibit more uniform mass-decorrelation across $\varepsilon_{\text{bkg}}^{\text{rel}}$. Among the MVA

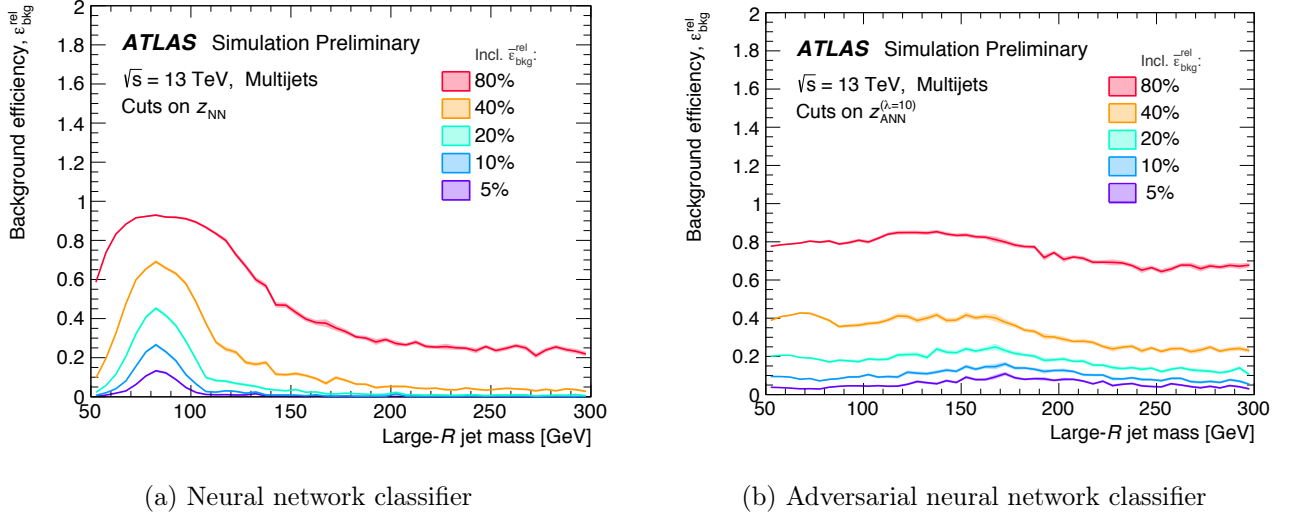


Figure 6.11: Jet mass-dependent multijet selection efficiencies for various inclusive efficiencies for the standard neural network tagger and the adversarially trained neural network tagger.

taggers, the ANN performs considerably better than *uBoost* for the chosen values of λ and α .

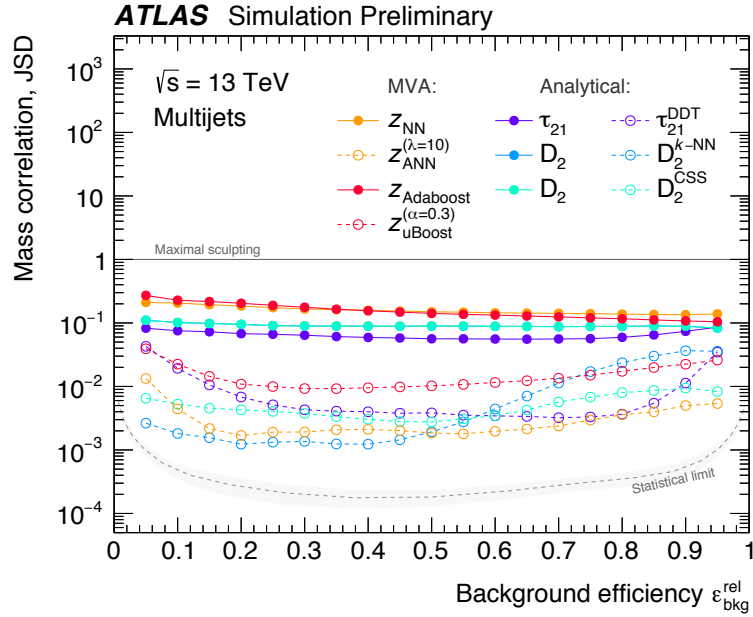


Figure 6.12: Profiles of the Jensen-Shannon divergence (JSD) for cuts corresponding to various multijet (background) selection efficiencies. Standard classifiers are indicated with filled markers, and mass-decorrelated classifiers indicated with open markers. The shaded grey band indicates the statistical limit on JSD due to the finite number of simulated jets.

The fact that the JSD is computed from histograms with finite statistics imposes a lower bound for the mass-decorrelation given the chosen testing dataset. Drawing two finite samples from the same underlying distribution will still result in non-zero JSD when evaluated on the

two sampled distribution. This statistical limit on JSD can be estimated using bootstrap sampling [107], which is shown as a function of the background selection efficiency in Figure 6.12 as a smoothed, dashed line and shaded band, indicating the mean and standard deviation, respectively, of the JSD limit. The procedure can be summarized as follows: at each fixed background efficiency ε , two bootstrapped samples $S_{1,2}$ are built by drawing samples (randomly) of $\varepsilon \times N$ and $(1 - \varepsilon) \times N$ jets, respectively, from the testing dataset D comprising N jets. The value $\text{JSD}(S_1 \parallel S_2)$ is computed and a distribution of it is obtained by repeating the procedure many times. The statistical limit is then obtained by taking the mean and variance of the bootstrap replicas.

It is worth to note that for a given physics task, full mass-decorrelation might not be necessary or optimal. Depending e.g. on the level of systematic uncertainty, some intermediate level of correlation with the jet mass might provide a balance with classification power which is better suited for the physics task at hand than full mass-decorrelation.

6.7.3 Combined Performance

A combined metric, reflecting both classification performance and mass-decorrelation, is necessary to assess the trade-off balance in each of the mass-decorrelation procedures. A more complete picture of the two metrics performance is found by plotting the two metrics together. Figure 6.13 shows the mass-decorrelation ($1/\text{JSD}$) versus the background rejection ($1/\varepsilon_{\text{bkg}}^{\text{rel}}$) for tagger cuts at $\varepsilon_{\text{sig}}^{\text{rel}} = 50\%$, in two p_{T} bins. The x -axis measures classification power and the y -axis measures mass-decorrelation, with larger values along each one indicating better performance. Depending on the given task, a specific direction in the plane of Figure 6.13 will correspond to the best trade-off.

For each of the mass-decorrelated MVA taggers, several working points are evaluated by scanning λ for the ANN tagger and α for uBoost. For high values of λ ($\gtrsim 10$), the ANN method starts to saturate given the chosen network configurations, training procedures, and datasets.

The dashed line and shaded band at high $1/\text{JSD}$ indicate the statistical limit of the mass-decorrelation, estimated using the bootstrap sampling described previously.

Figure 6.13 shows that for equal levels of mass-decorrelation, the ANN tagger generally provides the greatest background rejection. The BDT-based MVA taggers have performance comparable to the NN-based taggers for the standard variants, but the adversarial training mass-decorrelation method is seen to perform better than the *uBoost* method for the chosen configurations. From Figure 6.13b, the effect for the analytical, single-variable mass-decorrelation

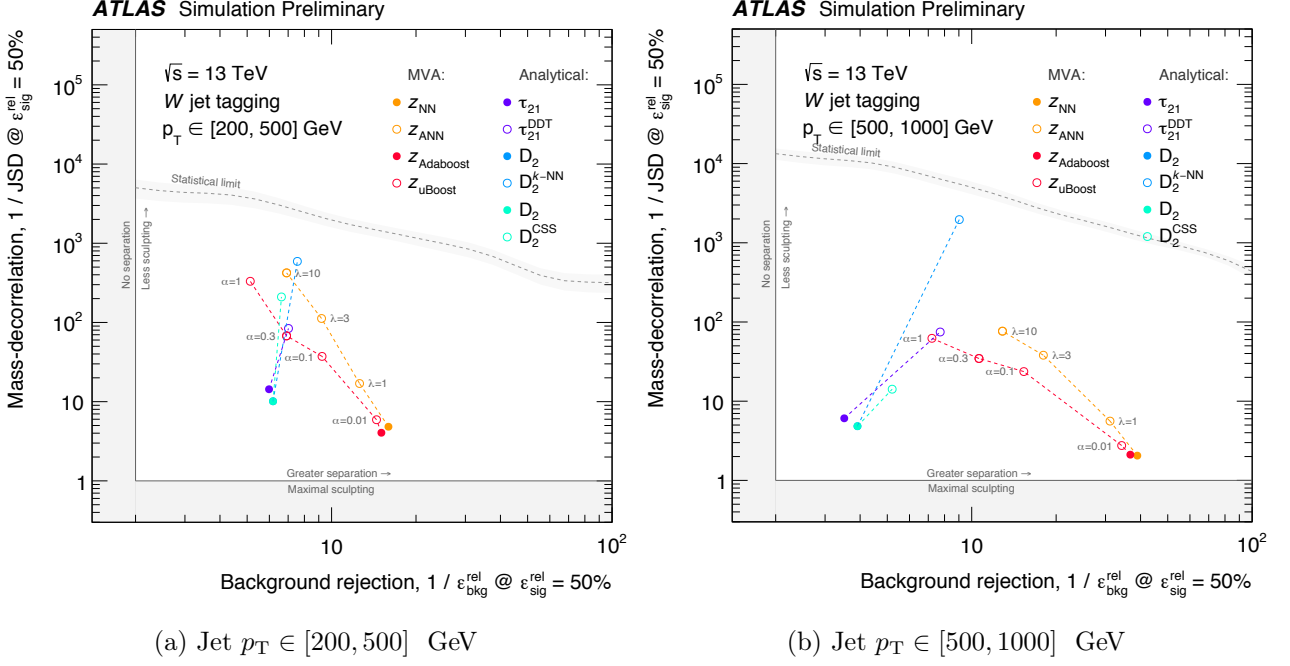


Figure 6.13: Unified plot of the metrics for classification (background rejection, $1/\epsilon_{\text{bkg}}^{\text{rel}}$) and mass-decorrelation (inverse Jensen-Shannon divergence, $1/\text{JSD}$), for cuts corresponding to $\epsilon_{\text{sig}}^{\text{rel}} = 50\%$, in two p_T bins. Greater values along each axis indicate better performance. Standard classifiers are indicated with filled markers. Mass-decorrelated classifiers are indicated with open markers, with parameter scans traced out by dashed lines. The shaded grey band indicates the statistical limit on $1/\text{JSD}$ from the finite number of simulated jets.

methods of simultaneously improving the classification power and the decorrelation from the jet mass, as discussed above, is particularly evident.

The k -NN regression method is seen to be the most effective of the analytical decorrelation methods, leading to a tagger variable which is close to fully decorrelated from the jet mass within statistical uncertainties. The saturation of the ANN tagger at high λ means that an upper limit on $1/\text{JSD}$ exists for $\lambda \gtrsim 10$ with the chosen configuration. This is a reasonable observation, considering the complexity of the ANN training procedure and the fact that the tagger is evaluated on a dataset 10 times larger than the training dataset. This discrepancy in data sample sizes means that complex decorrelation methods like the ANN are likely decorrelated at, or close to, the statistical limit of the training dataset, but not necessarily at the statistical limit for the much larger evaluation dataset. The simplicity of e.g. the k -NN method means that it will be more robust to such differences in statistics. For these reasons, raising the upper limit on $1/\text{JSD}$ for ANN will likely be possible by increasing training statistics or performing a more fine-tuned model architecture optimisation.

Chapter 7

Conclusions

The presented thesis was done within the Ph.D program in Physics during the period March 2014 – December 2018. The work centered on jets, particularly in jet calibration, jet cross-sections measurement and performance development to improve new physics searches based on jets. A wide range of software-based topics within the ATLAS Collaboration was covered. The main studies performed were the Monte Carlo based jet energy scale calibration, the measurement of the dijet cross-section at 13 TeV and the development of adversarial neural-networks to identify jets from the hadronic decay of W -bosons without altering the jet mass spectrum of the background.

The dijet measurement used data collected at the LHC with the ATLAS detector during 2015. It was the first time such a measurement was attempted at the energy scale of 13 TeV, testing the SM at an unprecedented energy regime. Quantitative comparisons to fixed-order NLO QCD calculations, corrected for non-perturbative and electroweak effects was done, yielding good agreement with pQCD, and so showing no hints of new physics, in all the kinematic regions explored. Several techniques used through the measurement were described in detail, such as data unfolding to get rid of detector effects, bootstrapping to estimate statistical uncertainties and χ^2 calculation to test the agreement with SM.

The use of adversarial neural-networks for jet identification coming from boosted W bosons against the large QCD background, is a novel machine learning technique attempted for the first time in ATLAS. It was proposed as a way to build a classifier independent with the jet mass, motivated by the fact that standard classifiers based on jet substructure variables exhibit strong non-linear correlations with the mass. Therefore, fixed-value cuts on such classifiers tend to distort the shape of the jet mass distribution for non-resonant jet production, resulting in a reduction of the statistical significance for searches of new physics. The studies were based on

Monte Carlo simulation. Results using adversarial networks show a considerable improvement with respect to others techniques such as adaptive boosted decision trees.

The studies performed in this thesis, not only helped the ATLAS Collaboration to fulfill its physics program, but also provided me with invaluable knowledge in software computing, statistical data analysis and machine learning. The results achieved hopefully will be used in the future for the scientific high energy physics community. More precisely, the dijet cross-sections measurement might be used to further constrain the PDFs of the proton with **HERAFitter** package [108]. On the other hand, the machine learning studies for jet identification provide an encouraging first look at the prospects of using adversarial neural-networks for mass-decorrelated substructure taggers in physics searches.

Appendix A

The Energy Dependence of the Jet Response

The calorimeter jet response is defined as the fraction of the true jet energy that is measured by the calorimeter. This value is smaller than one due to the presence of dead material in the supporting structures of the inner detector, the magnet and the calorimeters, the uninstrumented regions housing the criogenic and electronics conexions, and the different behaviour of hadronic (had) and electromagnetic (em) particles which comprise the jet. Electrons and photons produce electromagnetic showers to which the calorimeter is calibrated in test beams to have an electromagnetic particle response $r_e = 1$. Hadronic particles (charged pions, kaons, protons, neutrons) interact with nuclei, producing less ionization than em particles and thus having a lower response $r_h < 1$. Neutral pions (π^0) decay in two photons with 99.9 branching ratio and thus behave like em particles. The intrinsic response of the calorimeter depends thus on f_e , the fraction of the shower that is electromagnetic:

$$R = f_e + f_h r_h$$

where f_h is the hadronic fraction of the jet and thus $f_h + f_e = 1$.

The jet energy response grows with the energy mainly due to the increase of the number of nuclear interactions as the energy increases, which results in a higher f_e . In order to visualize this, let us suppose a jet composed by three pions (π^-, π^0, π^+). Before interacting with any nuclei (0th generation) the electromagnetic (em) and hadronic (had) components are 1/3 and 2/3 respectively, the em fraction will be then $f_e = 1/3$. As far as π^- and π^+ particles find nuclei, each of them will perform a nuclear interaction and generate new pions (π^-, π^0, π^+)

with equal probability ($1/3$ each)¹. As the number of generations grows, the production of π^0 increases, resulting in a higher em fraction (f_{em}) of the jet as it can be seen in Figure A.1. It can be clearly realized from the diagram that f_{em} follows a potential law:

$$f_{\text{em}} = \frac{1}{3} \sum_{k=0}^n \left(1 - \frac{1}{3}\right)^k = 1 - \left(1 - \frac{1}{3}\right)^{n+1}$$

Here we are making the assumptions that every nuclear interaction produces just pions, while in reality other particles (like kaons) are also produced, and that every pion type carries $1/3$ of the energy of the parent pion, which is equivalent to assume that there is no energy loss on each nuclear interaction. In reality, f_{em} will grow actually slower than a power law with the number of generations (n).

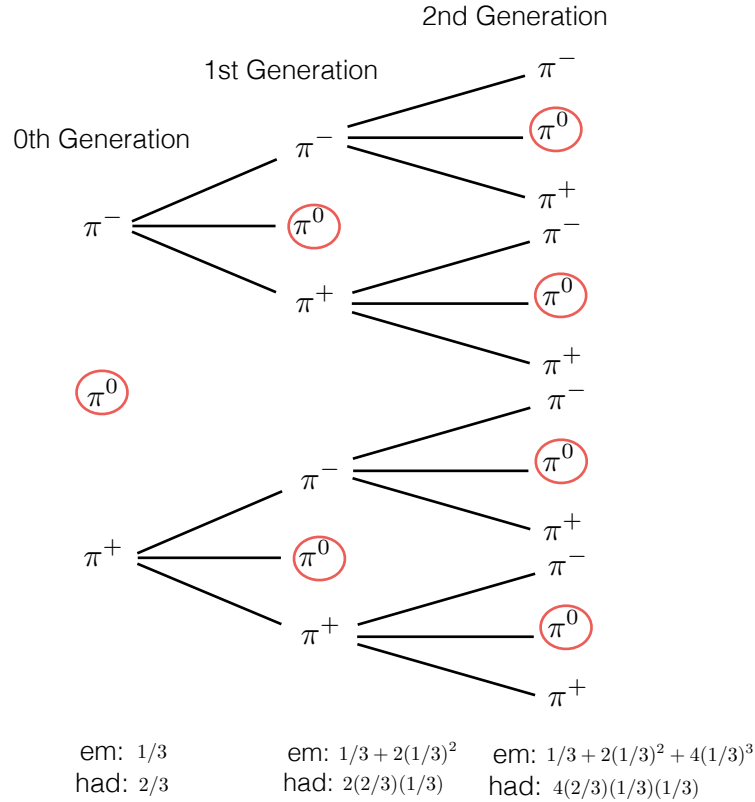


Figure A.1: Diagram of pions generation. The produced neutral pions do not contribute to the further development of the hadronic shower.

We can go deeper with this toy model and relate f_{em} with the initial energy of the jet (E). For this, we assume that the number of particles produced per interaction m (multiplicity) does not depend on the energy of the parent particle, and that the energy is proportional to

¹The three pions are the components of an isospin $I = 1$ triplet and so strong interactions do not distinguish between them.

the number of particles produced: $\frac{E}{E_0} = m^n$, where E_0 is the energy necessary to produce a π^0 , which is around 1.3 GeV [109]. As a result, we get that

$$n \sim \log\left(\frac{E}{E_0}\right)$$

Finally, we can conclude (at zero order approximation) the following relation:

$$f_{\text{em}} \sim 1 - \left(1 - \frac{1}{3}\right)^{\log\left(\frac{E}{E_0}\right)} \quad (\text{A.0.1})$$

This relation follows a similar pattern as the jet energy response from Figure 4.5. Although it is worth to note some differences with the expression we derived. It can be seen that when $E \rightarrow \infty$, $f_{\text{em}} \rightarrow 1$, which is not consistent with Figure 4.5 and clearly does not happen in reality. There are a lot of effects we are not taking into account which will result in a decrease of the em response. As we mentioned, we are neglecting the energy lost in breaking the nucleus, which is the contribution of r_h . A more detailed description of hadronic showers can be found in [109].

Appendix B

Smoothing Statistical Fluctuations in Systematic Uncertainties

To avoid including statistical fluctuations in the reported systematics, a smoothing technique was applied. The method takes the relative uncertainty for each systematic component as a function of m_{jj} whose statistical fluctuations were calculated using the bootstrap method introduced in Section 5.5. The steps of the procedure can be summarized as follows.

First, bins are combined iteratively until their weighted significance is larger than 2σ (in some cases significances larger than 1σ), starting from both the right and left sides of the mass range. The side with most remaining bins is chosen.

Second, a rebinned histogram is built from the original one where each bin is now statistically significant and the content y_j of each bin j corresponds to the weighted average:

$$y_j = \frac{\sum_{i=j}^{j+\Delta j} \frac{s_i}{\sigma_i^2}}{\sum_{i=j}^{j+\Delta j} \frac{1}{\sigma_i^2}}$$

where i labels the original binning and it moves from j until $j + \Delta j$ whose width is determined until the bin becomes significant. Here s_i is the relative systematic uncertainty and σ_i its statistical uncertainty.

Finally, a smoothed histogram μ_i with the original binning is built by applying a Gaussian kernel sliding window. Particularly, for each bin i , the following weighted average is computed:

$$\mu_i = \frac{\sum_l t_l K_{il}}{\sum_l K_{il}}$$

where i and l label the original binning, $t_{l=j\dots j+\Delta j} = y_j$, and

$$K_{il} = \exp\left(-\frac{(x_i - x_l)^2}{\Delta_i^2/4}\right)$$

where x_i is the central mass value of each bin and $\Delta_i = a + bx_i$, being $a = 10$ and $b = 0.3/\text{GeV}$.

The net effect is a reduction of statistical fluctuations in the systematic uncertainty components, with minimal added bias. This also has the effect of reducing artificial asymmetries, which are shown to cause problems when performing statistical tests for agreement between data and theory (Appendix C).

Results from this technique can be seen in Figure B.1. The plot on the left shows a relative uncertainty that has high statistical fluctuations. As a consequence, the reweighted distribution is a smoothed version from the original uncertainties. On the other hand, on the right plot all bins are already significant, and so the smoothing does not make any substantial improvement.

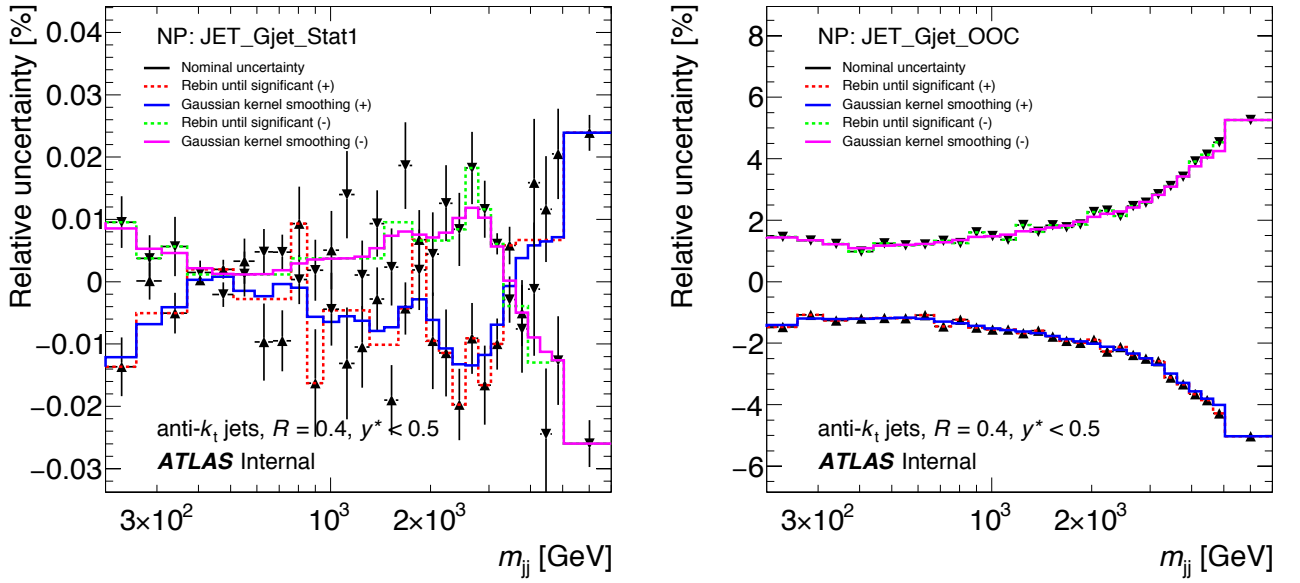


Figure B.1: Relative systematic uncertainty for up (blue) and down (pink) variations. On the left (right) uncertainty components with high (low) statistical fluctuations are shown. Markers corresponds to the nominal uncertainty, the dashed line to the rebinned histogram for the intermediate smoothing step and the solid line to the resulting smoothed uncertainty.

Appendix C

Statistical Test of Agreement between Theory and Data

The statistical method used to test the agreement of SM predictions to the measured cross-sections is based on a modification of the standard χ^2 test hypothesis.

The simplest version of the χ^2 test statistic is constructed from a sum of squares representing the degree of deviation of one spectrum from another, and is defined as:

$$\chi^2(d; t) = \sum_i \left(\frac{d_i - t_i}{\sigma_i} \right)^2$$

where d_i and t_i label the measured and theoretical cross-sections on bin i respectively, and σ_i^2 is the total variance (both data and theory uncertainties) in that bin. Under the assumption that the SM describes well the data (null hypothesis is true), that the bin contents are independent, and that the data are normally distributed, which is valid in many cases due to the central limit theorem, this test statistic follows a χ^2 distribution.

This expression accounts only for the theoretical and experimental uncertainties (including both statistical and systematics) on individual bins (σ_i), which ignores the statistical (due to the unfolding of resolution effects) and even stronger systematic correlations between bins. As a result, the standard definition results in a less powerful test. To account for this, the covariance matrix (C) is introduced in the χ^2 definition as:

$$\chi^2(d; t) = \sum_{i,j} (d_i - t_i) \cdot [C^{-1}]_{i,j} \cdot (d_j - t_j)$$

which also follows a χ^2 distribution. However, the covariance matrix is built from symmetrized uncertainties and thus cannot account for the asymmetries between positive and negative uncertainty components.

A modified version, which we keep denoting χ^2 , was proposed motivated from [110] and [111]. It accounts for asymmetric uncertainties by separating them from the symmetric ones in the test statistic, which is now:

$$\chi^2(d; t) = \min_{\beta_a} \left\{ \sum_{i,j} \left[d_i - \left(1 + \sum_a \beta_a \cdot (\epsilon_a^\pm(\beta_a))_i \right) t_i \right] \cdot [C_{\text{su}}^{-1}(t)]_{i,j} \right. \\ \left. \cdot \left[d_j - \left(1 + \sum_a \beta_a \cdot (\epsilon_a^\pm(\beta_a))_j \right) t_j \right] + \sum_a \beta_a^2 \right\} \quad (\text{C.0.1})$$

where C_{su} is the covariance matrix built using only the *symmetric* uncertainties, and β_a are the profiled coefficients resulting from a fit on the *asymmetric* uncertainties that minimize the χ^2 . Here $\epsilon_a^\pm$ is the positive component of the a^{th} asymmetric relative uncertainty if the fitted value of β_a is positive, or the negative component otherwise. For this analysis, an uncertainty component was considered asymmetric if the absolute difference between the magnitudes of the positive and negative portions is larger than 1% of the cross-section in at least one bin. Due to the large asymmetric uncertainties, coming mainly from the theoretical predictions, this χ^2 definition is a better motivated choice. Although it is not proved to be the optimal one, it provides stronger statistical power than the two approaches mentioned above.

As explained in Section 5.5, 1000 replicas of unfolded data were obtained for the estimation of the statistical uncertainty. The statistical covariance matrix is then built from these replicas following the standard definition. To build the symmetric systematic covariance, the uncertainties are assumed to be non correlated between each other and fully correlated across m_{jj} bins. As a result, the covariance for a given symmetric component k is then defined as $\text{cov}(x_i^k, x_j^k) = \sigma_i^k \sigma_j^k$, which comes from the fact that they are fully correlated across bins. The assumption that the systematics are independent between each other implies that one can sum linearly their covariance matrices. The result is added to the statistical covariance to get the total uncertainty matrix: $C_{\text{su}} = \sum_k \sigma_i^k \sigma_j^k + \text{stat. cov.}$

The χ^2 distribution for hypothetical data originating from a given theoretical prediction is required in order to calculate the probability of measuring a specific χ^2 value under that theory hypothesis. Since Eq. C.0.1 does not follow a standard χ^2 distribution, an empirical approach based on pseudo-experiments (Section 5.5) was performed. They represent fluctuations of data around a given theory hypothesis due to the full set of experimental and theoretical uncertainties. To do this, the statistical and symmetric systematic uncertainties were taken as Gaussian-distributed components, while a two-sided Gaussian distribution is used for the asymmetric ones. For each pseudo-experiment, the χ^2 value is computed between the pseudo-data (theory fluctuated spectra) and the theory hypothesis. In this way, the χ^2 distribution

that would be expected for experiments drawn from the theory hypothesis is obtained without making assumptions about its shape.

The observed χ^2 (χ_{obs}^2) value is computed using the actual data and the theory hypothesis. To quantify the compatibility of the data with the theory, the ratio of the area of the χ^2 distribution with $\chi^2 > \chi_{\text{obs}}^2$ to the total area is used. This fractional area, called the p -value, is the probability, under the assumption that the data follows the theory hypothesis, to find a value of χ^2 with equal or less compatibility than the observed one. If the observed p -value is smaller than 5%, the theoretical prediction is considered to poorly describe the data at the 95% CL.

Appendix D

Decorrelated Jet Tagging Algorithms

A variety of approaches to constructing mass-decorrelated jet taggers, from simple linear transforms to complex MVA architectures, are detailed below.

D.1 Designed decorrelated taggers

The original method relies on the jet scaling variable $\rho = \log(m^2/p_T^2)$ [79]. It is observed that profiling the substructure variable τ_{21} as a function of ρ exposes a linear relationship, which can be exploited to perform a linear transform, decorrelating τ_{21} with respect to ρ .

In practice [87, 112] it is found that the jet scaling variable

$$\rho^{\text{DDT}} = \log\left(\frac{m^2}{p_T \times \mu}\right)$$

more robustly removes the residual dependence on the jet p_T , and therefore leads to better decorrelation across the jet kinematic phase space. Here, the parameter μ balances the dimensions and is usually taken to be $\mu = 1 \text{ GeV}$.

In Figure D.1, a linear relationship between $\langle\tau_{21}\rangle$ and ρ^{DDT} is observed, roughly in the range $\rho^{\text{DDT}} \in [1.5, 4.0]$.

Towards higher values, the linearity breaks down due to effects arising from the fixed-radius jet clustering algorithm used, while it breaks down towards low values of ρ^{DDT} where soft QCD effects become important.

A linear fit is performed to the τ_{21} profile in the range $\rho^{\text{DDT}} \in [1.5, 4.0]$. From this fit, the transform $\tau_{21} \mapsto \tau_{21}^{\text{DDT}}$ is defined as

$$\tau_{21}^{\text{DDT}} = \tau_{21} - a \times (\rho^{\text{DDT}} - 1.5)$$

where a is the measured slope of the fit. Figure D.1 shows how the DDT transformation removes the linear correlation of τ_{21} with ρ^{DDT} (blue dots). Consequently, since ρ^{DDT} effectively encodes

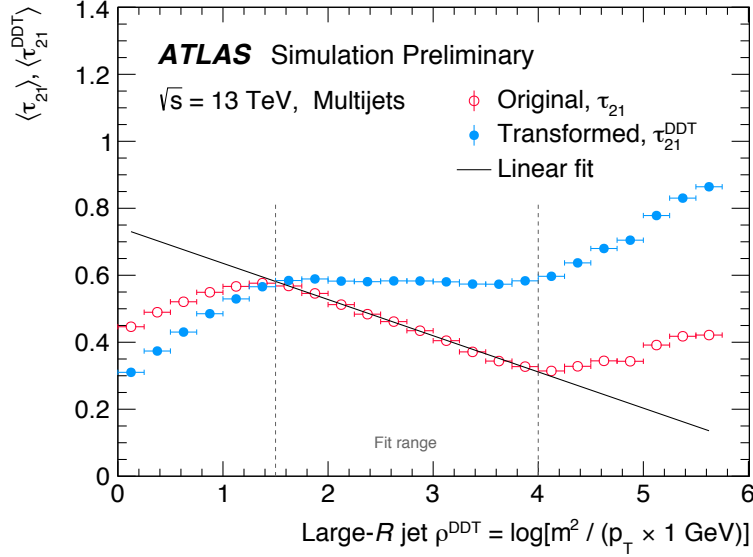


Figure D.1: Mean values of τ_{21} and τ_{21}^{DDT} as functions of ρ^{DDT} for the multijet background. A linear fit to the τ_{21} -profile, used in the definition of τ_{21}^{DDT} , is performed on the indicated range.

information about the kinematics of the jet (through m and p_T), the DDT transform yields a jet substructure discriminant which is decorrelated from the jet mass.

D.2 k -NN Regression

Whereas the DDT transform removes linear correlations between a substructure variable and kinematic variable(s), more general strategies able to face non-linear relationships exist. One of these methods relies in a k -NN regression and can be summarized as follows:

First, a fixed percentile of inclusive background efficiency $\varepsilon_{\text{bkg}}^{\text{rel}}$ for the D_2 [48] distribution is computed for multijets in bins of $\rho = \log(m^2/p_T^2)$ and p_T , resulting in a two-dimensional profile. The efficiency selected in this case is $\varepsilon_{\text{bkg}}^{\text{rel}} = 16\%$, corresponding roughly to a signal efficiency of $\varepsilon_{\text{sig}}^{\text{rel}} = 50\%$. Second, in order to determine its functional dependence, a two-dimensional regression fit using the k -nearest neighbours (k -NN) algorithm [80] is performed, yielding the continuous distribution $D_2^{(16\%)}(\rho, p_T)$. Finally, for each jet a new observable $D_2 \mapsto D_2^{k\text{-NN}}$ is constructed by subtracting the predicted values from D_2 :

$$D_2^{k\text{-NN}} = D_2 - D_2^{(16\%)}$$

The profiles of the fixed-efficiency D_2 cut as a function of ρ and p_T are shown in Figure D.2. The k -NN technique generalises the central concept behind DDT, making the method admissible to a more general class of substructure variables.

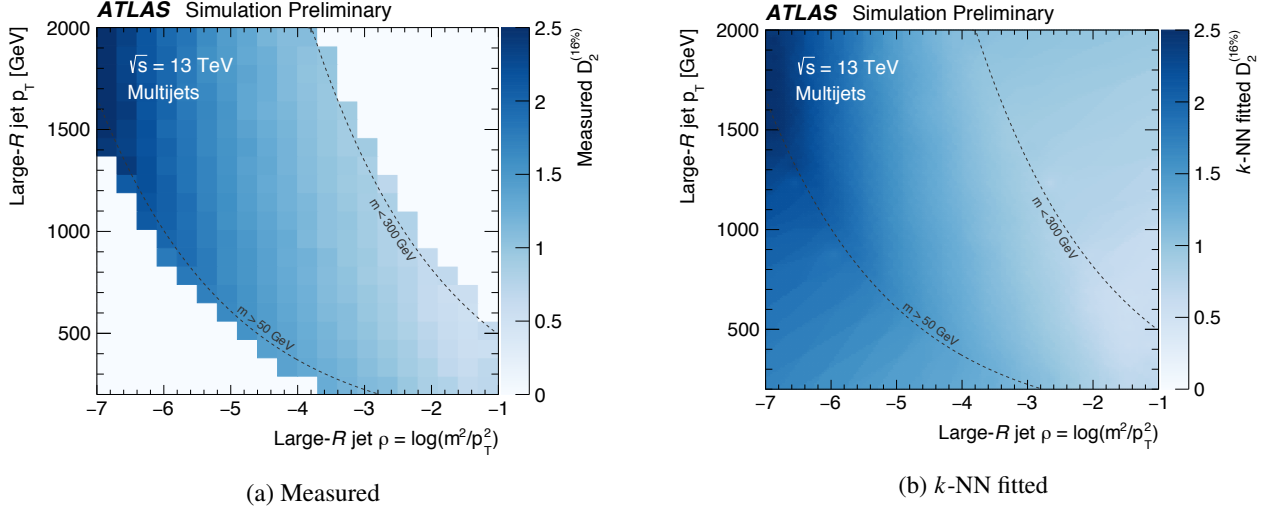


Figure D.2: Profiles of D_2 at $\varepsilon_{\text{bkg}}^{\text{rel}} = 16\%$ for multijets as measured in the training dataset (a) and fitted using k -NN regression (b). Dashed lines indicate the phase space limits arising from the jet mass selection.

D.3 Convolved Substructure

Another analytical, more general, technique than the DDT described above is the convolved substructure (CSS) method [81]. It is capable to remove the dependence with the mass on the entire shape by convolving the distribution of the substructure variable of interest, $c(x) = \frac{1}{\sigma} \frac{d\sigma}{dx}$, with a shape function mapping it to the distribution at a reference mass:

$$\frac{1}{\sigma} \frac{d\sigma}{dx} \mapsto \frac{1}{\sigma} \frac{d\sigma}{dx_{\text{CSS}}} = \frac{1}{\sigma} \frac{d\sigma}{dx} \otimes F_{\text{CSS}}(x|\alpha, \beta), \quad (\text{D.3.1})$$

where $F_{\text{CSS}}(x|\alpha, \beta)$ is a Gamma distribution. The parameter α specifies the functional form and β depends on the actual mass and the reference mass to which shape is desired to evolve.

$$F_{\text{CSS}}(x|\alpha, \beta) = \left(\frac{\alpha}{\beta}\right)^\alpha \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{\alpha x}{\beta}}. \quad (\text{D.3.2})$$

Equation (D.3.1) is defined only at the level of distributions; in practice, the convolution, indicated by $\otimes$, is estimated on a jet basis by transforming $x \mapsto x_{\text{CSS}} = G^{-1}(C(x)|\alpha, \beta)$, where $C(x)$ is the cumulative distribution function (CDF) of the distribution $c(x)$ and G is the CDF corresponding to the distribution of x_{CSS} (i.e. $\frac{1}{\sigma} \frac{d\sigma}{dx_{\text{CSS}}}$).

The parameters α and β are optimised inclusively and per bin of jet mass, respectively, using a two-step optimisation procedure as described in [81].

In order to facilitate the comparison with the k -NN technique, $x = D_2$ was selected as the substructure variable. The transformed D_2^{CSS} distribution in the highest p_T bin is shown in Figure D.3. The convolution procedure from Equation (D.3.1) evolves the distribution of D_2

at one mass to the distribution at a reference mass. Therefore, using the lowest mass bin as the reference, the D_2^{CSS} in this bin is approximately unchanged. In contrast, the distribution of D_2^{CSS} in the high mass bin visibly differs from D_2 in the same mass bin. Not only has the mean value of D_2 shifted towards the reference mass, but the entire distribution is now coherently closer. Using the CSS technique, the D_2 distribution is transformed to be similar for all jet masses, thereby directly reducing correlation with the jet mass.

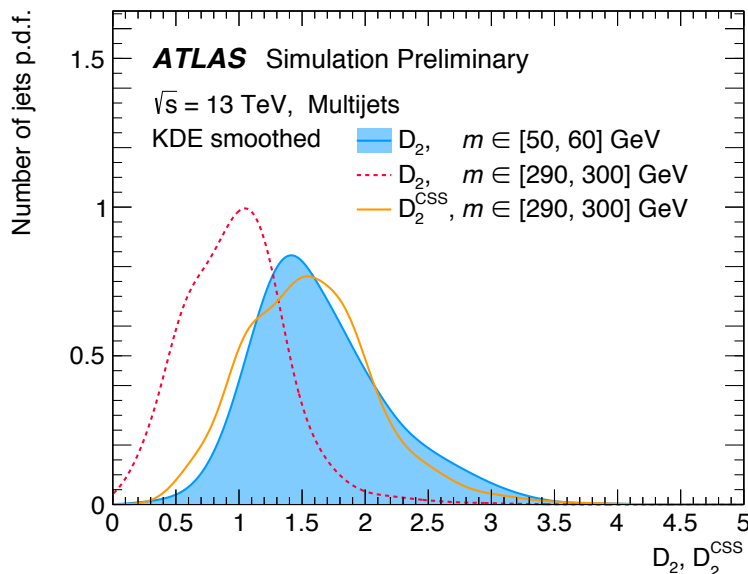


Figure D.3: Distributions of D_2 and D_2^{CSS} for multijets in the highest mass bin, as well as the reference distribution in the lowest mass bin. A kernel-density estimation (KDE) based smoothing is applied to all training distributions.

D.4 Adaptive boosting for uniform efficiency

Boosted decision trees (BDTs) algorithms like AdaBoost (from “adaptive boosting”) [113] are playing an increasingly prominent role in HEP physics applications. As an example, the observation of $H \rightarrow b\bar{b}$ has recently been achieved thanks to the use of BDTs [114]. They are also being used in object calibration [115], MC reweighting [116], b -tagging [117], and software triggers. In contrast with the analytical taggers described above, based on a single feature, BDTs are multivariate (MVA) taggers that combine information of multiple features in order to achieve a single predictor. Other examples of MVA based techniques are Neural-Networks. BDTs possess an additional property called *boosting*, where an ensemble of *weak* classifiers are combined in order to obtain a single *strong* learner.

Despite its success, BDTs are also found to yield selection efficiencies which are non-uniform

with respect to the jet mass, sculpting the mass background spectra after jet classification. The *adaptive boosting for uniform efficiency* (uBoost) method [82] is proposed to mitigate this non-uniformity behaviour.

Boosting works by assigning training events weights based on classification errors made by previous members of the series; events that are misclassified are given more weight. The uBoost technique augments this procedure by not just considering classification error, but also the uniformity of the selection. Events in regions of m where the selection efficiency is lower (higher) than the mean are given larger (smaller) weights. In this way, uBoost is able to drive the BDT towards a fixed uniform selection efficiency in the m observable at the cost of loosing discrimination power.

The BDT classifiers use the substructure variables listed in Table 4.1, the same used for the neural network classifiers, as input features. The hyperparameter configuration adopted for AdaBoost is the same as the one used for the BDT classifier in [46]. For the remaining uBoost hyperparameters, the default values in [118] are used. The degree of mass-decorrelation for uBoost is controlled by a hyperparameter α , called the uniforming rate. For $\alpha \rightarrow 0$, the adaptive boosting only takes classification loss into account, and the standard AdaBoost classifier is retrieved. Conversely, for larger α , the uniform efficiency boosting becomes gradually more important.

The AdaBoost and uBoost classification losses as a function of the number of boostings (epochs) are shown in Figure D.4 for the training and testing set.

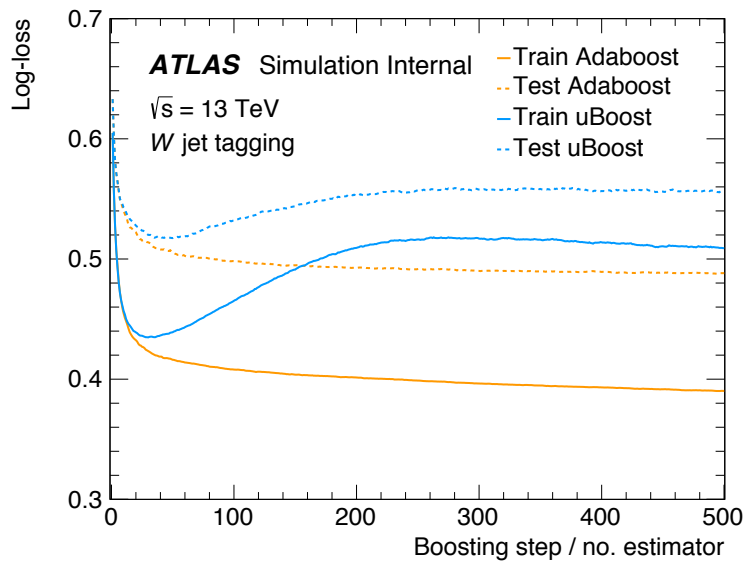


Figure D.4: Classification losses of AdaBoost and uBoost classifiers for training and testing datasets at $\alpha = 0.1$.

The AdaBoost classification loss is seen to decrease monotonically and reach a plateau for the testing dataset at ~ 100 epochs of training. In contrast, the classification loss for uBoost initially decreases due to improved discriminating power, and then rebounds as the adaptive boosting for uniform efficiency takes effect. The plateau is achieved at ~ 250 epochs.

Resúmenes de Capítulos en Castellano

Capítulo 1: Introducción

El 20 de Mayo del 2015, el detector ATLAS recolectó por primera vez datos de colisiones protón-protón generadas por el Gran Colisionador de Hadrones (LHC) a una energía de centro de masa de 13 TeV, la mas alta alcanzada por un acelerador de partículas al dia de la fecha. Mediciones de la sección eficaz de producción de jets resulta fundamental para extender el conocimiento de QCD a escalas de energía nunca antes exploradas. En particular, la producción de pares de jets brinda información acerca de la estructura interna partónica del protón y de la constante de acoplamiento fuerte α_s .

Esta tesis presenta, por primera vez, mediciones de la sección eficaz de pares de jets a 13 TeV. Dicha medición es comparada cuantitativamente con las predicciones teóricas del Modelo Estándar (ME) a orden siguiente dominante (NLO) corregidas por efectos no perturbativos.

Los jets son también importantes para la búsqueda de resonancias de partículas masivas, las cuales podrian dar indicios de nueva física. En particular, los jets que provienen de una partícula pesada como el quark top o el bosón W . Resulta fundamental por ende la identificación de este tipo de jets en fondos de QCD dominantes. En esta tesis se presenta una técnica novedosa basada en redes neuronales adversarias para lograr un clasificador de bosones W no correlacionado con la masa del jet. Esto muestra una mejora en el rendimiento con respecto a otras técnicas utilizadas en ATLAS hasta la fecha, resultando por lo tanto prometedor para el uso en futuros analisis de física.

Capítulo 2: Aspectos Teóricos

En este capítulo se presenta un breve resumen del Modelo Estándar (ME) y de las predicciones de QCD. A su vez, se describen las técnicas de Monte Carlo utilizadas para las simulaciones de las colisiones protón-protón y la interacción con el detector ATLAS.

El ME es una teoría cuántica de campos que describe el comportamiento de las partículas experimentalmente observadas bajo la interacción de la fuerza electromagnética (EM), débil y fuerte. El modelo está descrito por simetrías fundamentales de la naturaleza como el grupo de Poincaré (homogeneidad, isotropía y covarianza) del espacio-tiempo y simetrías de gauge local (resultante de las interacciones fundamentales) descritas por los grupos $SU(3) \times SU(2) \times U(1)$. Todo el modelo de simetrías e interacciones está representado por una densidad Lagrangiana, la cuál describe 3 familias de quarks, 3 familias de leptones, 4 intermediarios de las 3 fuerzas fundamentales (EM, débil y fuerte) y el bosón de Higgs, descubierto en 2012 por ATLAS y CMS.

Debido a que las colisiones en el LHC son hadrónicas (formadas por muchos partones), se necesita de un método para extender las predicciones de nivel partónico a nivel hadrónico. El teorema de factorización descrito por la ecuación 2.1 permite la realización de tal propósito, indispensable para comparar las predicciones teóricas con las mediciones experimentales. Esta relación se utiliza para poder ajustar con un fit la dependencia de las distribuciones partónicas (PDFs) en función de la fracción de momento de los partones dentro del protón x y la escala de energía Q^2 . De esta manera, las mediciones permiten mejorar el conocimiento de las funciones de estructura del protón, las cuales no pueden ser predichas de primeros principios.

Simulaciones Monte Carlo basadas en lluvia partónicas son fundamentales para comparar con los observables físicos medidos y reconstruidos en el detector. Dichos generadores parten de la simulación a nivel de elemento de matriz (LO o NLO) y radican en una lluvia partónica para facilitar la simulación del proceso de hadronización permitiendo obtener las partículas estables medidas por un detector. El generador más utilizado por ATLAS es *Pythia*, y dado el buen acuerdo que presenta con los datos reales, fue el elegido por defecto en esta tesis.

Una vez generados las partículas estables del evento, se simula la interacción con la materia resultando en el mismo tipo de señales que uno mediría con el detector ATLAS en un experimento real. Para tal propósito, se utiliza el software **Geant4**, el cual es un programa flexible diseñado para simular geometrías complejas en entornos complejos como es el caso del detector ATLAS.

Capítulo 3: El LHC y el Detector ATLAS

El Gran Colisionador de Hadrones (LHC) es el acelerador de partículas mas grande del mundo. Es un sincrotrón de protones ubicado en el Laboratorio CERN en la frontera Franco-Suiza (Ginebra). Esta ubicado 100 m bajo tierra y tiene un perímetro de 27 km. El mismo alberga varios detectores, entre ellos ATLAS, CMS, Alice y LHCb. Los dos primeros son detectores multipropósito, cuyo objetivo es buscar cualquier tipo de anomalía que haya en los datos. Por otro lado, Alice y LHCb son mas específicos y buscan comprender mejor el quark-gluon plasma y la naturaleza de los quarks b , respectivamente.

Los protones, previos a ser inyectados en el LHC, son generados por una botella de hidrógeno y separados mediante un campo eléctrico. Estos son acelerados por diversos aceleradores de menor energía hasta finalmente ser inyectado al anillo central del LHC.

El LHC colisiona protones a una frecuencia de 40 MHz. El numero de eventos total generados es proporcional a la sección eficaz del proceso (dependiente de la naturaleza) y a la luminosidad integrada en el tiempo (dependiente de las condiciones experimentales). Un efecto no deseado inherente del proceso de generación de datos es el denominado *pile-up*. Este consiste en múltiples colisiones protón-protón dentro de un mismo cruce (*in-time* pile-up) y efectos introducidos por cruces vecinos (*out-of-time* pile-up). Estos efectos ocasionan pérdida de precisión en las mediciones y resulta un desafío experimental desarrollar técnicas que atenuen dicho fenómeno.

El detector ATLAS esta formado por varios sub-detectores. En orden de cercanía a la zona de interacción se encuentran: el detector de trazas, el cual esta formado por un detector de pixeles de alta granularidad, seguido por discos de microfibra de silicio (SCT) y complementado por un detector de radiación (TRT). Estos se encargan de medir la curvatura de las trazas producidas por partículas cargadas mediante la cual se obtiene el momento de las mismas.

Luego del detector de trazas se encuentran los calorímetros electromagnético (em) y hadrónico (had). El calorímetro em se encarga de medir solo la energía depositada por fotones o electrones. Ambos pueden ser distinguidos al saber si dejaron o no una traza en el detector de

trazas. Por otro lado, el calorímetro had mide la energía depositada por los hadrones, los cuales interactúan con la materia luego de pasar el calorímetro em.

Por último se encuentra el detector de muones, los cuales miden exclusivamente el momento de los muones.

Dada la alta frecuencia que opera el LHC (40 MHz) y el tamaño de cada evento (~ 1.5 MB), tecnológicamente no es posible almacenar todos los datos. Por tal motivo ATLAS cuenta con un sistema de disparo (trigger), el cual está optimizado para de forma rápida decidir si un evento es o no interesante para ser almacenado. Este sistema funciona como un filtro de eventos, y reduce la frecuencia de almacenamiento de datos de 40 MHz a 1 KHz.

Capítulo 4: Reconstrucción y Calibración de Jets

Debido a la propiedad de confinamiento de QCD, los quarks y gluones no pueden ser observados de forma directa, sino a través de lluvia de partículas incoloras, que reciben el nombre de *jets*. Para relacionar las señales observadas con el proceso partónico que tuvo lugar a nivel de elemento de matriz, es necesario un algoritmo de recombinación de partículas (algoritmo de jet). Este toma como input una lista de cuadrivectores (partones, partículas, trazas, clusters calorimétricos) y los recombina para obtener un jet el cual representa a la partícula originaria de la lluvia.

El algoritmo de jet mas utilizado en ATLAS, y en particular el utilizado en esta tesis, es el denominado *anti- k_t* , dado a su no sensibilidad con el pile-up. Este algoritmo se basa en una definición de distancia entre elementos representados por cuadrivectores dado por la ecuación 4.1.

Los inputs del algoritmo de jets pueden ser cualquier lista de objetos representados por un cuadrivector. En particular clusters calorimétricos. La recombinación de estos clusters da lugar a jets calorimétricos. Aquellos clusters en donde se asume que la energía es depositada es mediante partículas electromagnéticas se los denomina EM clusters, dando lugar a EM jets. Otro tipo de jets son los llamados LCW, los cuales fueron formados por clusters corregidos por la respuesta del calorímetro a hadrones. Por último, se encuentran los denominados p-flow jets, formados por EM clusters utilizando además la información de las trazas asociadas a los mismos para determinar su momento. Esta técnica, novedosa en ATLAS, permite mejorar la resolución a bajo momento (≤ 100 GeV) de los jets. Por último se encuentran los jets de partículas, los cuales fueron reconstruidos por las partículas estables simuladas por los Monte Carlos.

La calibración de los jets resulta fundamental para su posterior comparación con los modelos teóricos. La corrección consiste en llevar la energía de los jets reconstruidos con el detector

a su correspondiente jet de partícula. Unas de las etapas de la calibración es la basada en simulaciones Monte Carlo denominada *MC Jet Energy Scale (MC-JES)*. Dicha calibración fue una de las tareas principales ejercidas durante el doctorado y se describe en detalle en este capítulo.

La calibración de los jets introduce incertezas sistemáticas en su cuadrimomento que son luego propagadas a la medición final de la sección eficaz. La más relevante en esta medición es la que proviene en la determinación de la escala de energía de los jets, la denominada *Jet Energy Scale* que proviene en mayor medida de correcciones en los datos por diferencias con las simulaciones Monte Carlo.

Capítulo 5: Medición de la Producción de Pares de Jets a 13 TeV

En este capítulo se describe en detalle la medición de la sección eficaz de producción de pares de jets a 13 TeV. Esta fue medida para distintas regiones de separación angular entre jets, y^* , en función de la masa invariante del sistema, m_{jj} .

Poseer una medición de la sección eficaz al nivel partícula en vez de al nivel del detector tiene la ventaja de poder utilizar las mediciones para la comparación con futuros experimentos sin necesidad de requerir de los detalles del detector utilizado. A su vez, las predicciones teóricas no necesitan de la posterior simulación del detector. Para lograr una medición al nivel partícula, se uso una técnica de de-convolución conocida también como *unfolding*. La misma consiste en un método iterativo basado en una matriz de transferencia (*transfer matrix*) que contiene simultaneamente información de la resolución del detector y del espectro a nivel partícula (*truth*) previo a ser pasado por la simulación del detector. La transfer matrix cuantifica la probabilidad de un evento reconstruido en el detector provenga de uno a nivel partícula y vice versa, la cual permite llevar los datos reconstruidos a los generados a nivel partícula.

La estimación de las incertezas estadísticas se hizo mediante la técnica de *bootstrapping*, la cuál consiste en pesar los eventos por una variable aleatoria proveniente de una distribución de *Poisson* con $\mu = 1$ para generar *toys* y propagar los mismos al observable final. Se construye con los toys una matriz de covariancia, obteniendo de esta forma una estimación de la incerteza estadística empírica sin necesidad de realizar mas experimentos (lo cual resultaria imposible).

Las incertezas sistemáticas mas significativas son debidas a la *Jet Energy Scale*, y se obtienen variando el cuadri-momento del jet por su respectiva incerteza y propagando el jet con el momento modificado a la sección eficaz. Las incertezas del momento del jet son naturalmente amplificadas al ser propagadas, debido a la dependencia exponencial decreciente de la sección eficaz con la masa invariante.

Habiendo cuantificado todas las incertezas, se procede a la comparación con los cálculos

teóricos a orden siguiente al dominante (NLO) corregidos por efectos no perturbativos. El nivel de acuerdo entre datos y teoría se cuantifica mediante un test de χ^2 , el cual tiene en cuenta las varianzas (estadísticas y sistemáticas) y sus correlaciones (covarianzas) experimentales y teóricas. El estadístico utilizado es sensible también a la asimetría presentes en algunas incertezas (principalmente debida a los cálculos teóricos).

Las mediciones dieron consistentes con el Modelo Estándar (p-values del orden porcentual), validando el ME de producción de pares de jets a 13 TeV.

Capítulo 6: Redes Neuronales Adversarias para Identificar Jets- W

Este capítulo comienza con una descripción matemática general del aprendizaje automático, fundamentando el uso de las redes neuronales como modelo predictivo capaz de parametrizar distribuciones de probabilidad complejas.

En este análisis se propone el uso de redes neuronales adversarias (ANN) para construir un clasificador de jets masivos entre señal (bosones W) y fondo (QCD) a su vez no correlacionado con la masa del jet. Los clasificadores tradicionales utilizados en ATLAS, como redes neuronales y arboles de decisión, presentan correlaciones no lineales con la masa del jet, lo que provoca problemas en la determinación del fondo produciendo una reducción de la significancia estadística de varios analisis de búsqueda de nueva física. El método propuesto basado en ANN, presenta una solución a este tipo de problemas. El mismo es comparado con otras técnicas de descorrelación, como *uBoost*, regresión k -NN, DDT y CSS. Todos estos métodos se basan en variables de sub-estructura de los jets, los cuales son características poderosas que permite distinguir entre jets de señal y de fondo.

El análisis se restringe a la región de masa $50 \text{ GeV} < m < 300 \text{ GeV}$. El momento transverso del jet se limitó al rango $200 < p_T < 2000 \text{ GeV}$. Con estos criterios cinemáticos se garantiza que el jet sea colimado (de forma de contener los decaimientos de un boson W dentro del mismo jet). Para el entrenamiento de la red se utilizaron 10^6 eventos de señal y de fondo. Por otro lado, la evaluación del rendimiento se hizo con un set de testeo consistente de 10^6 y 10^7 eventos de señal y de fondo respectivamente.

Para medir el poder de clasificación se utilizaron curvas ROC, en donde se grafica el rechazo del fondo a una dada eficiencia de señal. El punto de trabajo elegido para la optimización fue de 50%, típicamente el utilizado en varios analisis. Por otro lado, se midió el poder de descorrelación mediante la divergencia de *Kullback-Leibler*. La misma mide la entropía relativa entre las distribuciones de fondo antes y luego del corte aplicado por el clasificador.

La optimización conjunta de los parámetros consiste en dos redes neuronales que compiten entre si. Por un lado una red clasificadora se encarga de optimizar el poder de clasificación usando una función de costo de entropía cruzada. Por el otro, una red adversarial condicionada al output de la clasificadora, se encarga de aprender la distribución de masa del jet mediante la parametrización por un modelo de mixtura Gaussiano. Haber aprendido la distribución de masa del jet, se traduce a introducir una regularización en los parámetros del clasificador y restringirlos a un sub-espacio tal que la predicción del clasificador, para los jets de fondo, resulta independiente con la masa del jet. En otras palabras, esto permite lograr una eficiencia de selección de jets de fondo uniforme con la masa del jet, como se muestra en la Figura 6.11.

El rendimiento final se presenta en un gráfico bidimensional, donde en el eje vertical se indica el poder de descorrelación y en el eje horizontal el poder de clasificación alcanzado, evaluado en el set de testeo. Se puede observar que ANN logra el mayor poder de clasificación a un poder de descorrelación dado en comparación con uBoost. Por otro lado, el máximo poder de descorrelación se logra con la técnica analítica de regresión k -NN.

Los resultados obtenidos utilizando redes neuronales adversarias para la identificación de bosones W resultan alentadores para aplicar dicha técnica en futuros analisis de búsqueda de nueva física.

Bibliography

- [1] F. Abe et al. “Measurement of the dijet mass distribution in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV”. In: *Phys. Rev. D* 48 (1993), pp. 998–1008. DOI: [10.1103/PhysRevD.48.998](#).
- [2] ATLAS Collaboration. “Measurement of inclusive jet and dijet cross sections in proton–proton collisions at 7 TeV centre-of-mass energy with the ATLAS detector”. In: *Eur. Phys. J. C* 71 (2011), p. 1512. DOI: [10.1140/epjc/s10052-010-1512-2](#). arXiv: [1009.5908 \[hep-ex\]](#).
- [3] Morad Aaboud et al. “Determination of the strong coupling constant α_s from transverse energy–energy correlations in multijet events at $\sqrt{s} = 8$ TeV using the ATLAS detector”. In: *Eur. Phys. J. C* 77.12 (2017), p. 872. DOI: [10.1140/epjc/s10052-017-5442-0](#). arXiv: [1707.02562 \[hep-ex\]](#).
- [4] C. Adloff et al. “Measurement of dijet cross-sections in photoproduction at HERA”. In: *Eur. Phys. J. C* 25 (2002), pp. 13–23. DOI: [10.1007/s10052-002-1004-0](#). arXiv: [hep-ex/0201006 \[hep-ex\]](#).
- [5] Georges Aad et al. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Phys. Lett. B* 716 (2012), pp. 1–29. DOI: [10.1016/j.physletb.2012.08.020](#). arXiv: [1207.7214 \[hep-ex\]](#).
- [6] P.W. Higgs. “Broken symmetries, massless particles and gauge fields.” In: *Physics Letters* 12(2) (1964), pp. 132–133.
- [7] F. Abe et al. “Observation of top quark production in pp collisions with the collider detector at fermilab.” In: *Phys. Rev. Lett.* 74 (Apr 1995.), 2626–2631.
- [8] S. Abachi et al. “Search for high mass top quark production in pp collisions at $\sqrt{s} = 1.8\text{TeV}$.” In: *Phys. Rev. Lett.* 74 (Mar 1995.), 2422–2426.
- [9] Roel Aaij et al. “Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \rightarrow J/\psi K^- p$ Decays”. In: *Phys. Rev. Lett.* 115 (2015), p. 072001. DOI: [10.1103/PhysRevLett.115.072001](#). arXiv: [1507.03414 \[hep-ex\]](#).

- [10] James Currie et al. “Precise predictions for dijet production at the LHC”. In: *Phys. Rev. Lett.* 119.15 (2017), p. 152001. DOI: [10.1103/PhysRevLett.119.152001](#). arXiv: [1705.10271 \[hep-ph\]](#).
- [11] John C. Collins Davison E. Soper and George F. Sterman. “Factorization of Hard Processes in QCD.” In: *Adv.Ser.Direct.High Energy Phys.* 5 (1988), pp. 1–91.
- [12] Zoltan Nagy. “Next-to-leading order calculation of three jet observables in hadron hadron collision”. In: *Phys. Rev.* D68 (2003), p. 094002. DOI: [10.1103/PhysRevD.68.094002](#). arXiv: [hep-ph/0307268 \[hep-ph\]](#).
- [13] H. Abramowicz et al. “Combination of measurements of inclusive deep inelastic $e^\pm p$ scattering cross sections and QCD analysis of HERA data”. In: *Eur. Phys. J.* C75.12 (2015), p. 580. DOI: [10.1140/epjc/s10052-015-3710-4](#). arXiv: [1506.06042 \[hep-ex\]](#).
- [14] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. “A Brief Introduction to PYTHIA 8.1”. In: *Comput. Phys. Commun.* 178 (2008), pp. 852–867. DOI: [10.1016/j.cpc.2008.01.036](#). arXiv: [0710.3820 \[hep-ph\]](#).
- [15] Andrew Buckley. “ATLAS Pythia 8 tunes to 7 TeV data”. In: *Proceedings of the Sixth International Workshop on Multiple Partonic Interactions at the Large Hadron Collider*. CERN. Geneva: CERN, 2014, p. 29.
- [16] Richard D. Ball et al. “Parton distributions with QED corrections”. In: *Nucl. Phys.* B877 (2013), pp. 290–320. DOI: [10.1016/j.nuclphysb.2013.10.010](#). arXiv: [1308.0598 \[hep-ph\]](#).
- [17] M. Bahr et al. “Herwig++ Physics and Manual”. In: *Eur. Phys. J.* C58 (2008), pp. 639–707. DOI: [10.1140/epjc/s10052-008-0798-9](#). arXiv: [0803.0883 \[hep-ph\]](#).
- [18] Steffen Schumann and Frank Krauss. “A Parton shower algorithm based on Catani-Seymour dipole factorisation”. In: *JHEP* 03 (2008), p. 038. DOI: [10.1088/1126-6708/2008/03/038](#). arXiv: [0709.1027 \[hep-ph\]](#).
- [19] S. Agostinelli et al. “GEANT4: A simulation toolkit”. In: *Nucl. Instrum. Meth.* A 506 (2003), pp. 250–303.
- [20] Oliver S. Bruning et al. “LHC Design Report Vol.1: The LHC Main Ring”. In: (2004).
- [21] O. Buning et al. “LHC Design Report. 2. The LHC infrastructure and general services”. In: (2004).
- [22] M. Benedikt et al. “LHC Design Report. 3. The LHC injector chain”. In: (2004).

- [23] URL: <https://home.cern/about/engineering/restarting-lhc-why-13-tev>.
- [24] Georges Aad et al. “Luminosity Determination in pp Collisions at $\sqrt{s} = 7$ TeV Using the ATLAS Detector at the LHC”. In: *Eur. Phys. J. C* 71 (2011), p. 1630. DOI: [10.1140/epjc/s10052-011-1630-5](https://doi.org/10.1140/epjc/s10052-011-1630-5). arXiv: [1101.2185](https://arxiv.org/abs/1101.2185) [hep-ex].
- [25] S. van der Meer. “Calibration of the Effective Beam Height in the ISR”. In: (1968).
- [26] Morad Aaboud et al. “Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC”. In: *Eur. Phys. J. C* 76.12 (2016), p. 653. DOI: [10.1140/epjc/s10052-016-4466-1](https://doi.org/10.1140/epjc/s10052-016-4466-1). arXiv: [1608.03953](https://arxiv.org/abs/1608.03953) [hep-ex].
- [27] A. Airapetian et al. “ATLAS: Detector and physics performance technical design report. Volume 1”. In: (1999).
- [28] G. Aad et al. “The ATLAS Experiment at the CERN Large Hadron Collider”. In: *JINST* 3 (2008), S08003. DOI: [10.1088/1748-0221/3/08/S08003](https://doi.org/10.1088/1748-0221/3/08/S08003).
- [29] Morad Aaboud et al. “Performance of the ATLAS Trigger System in 2015”. In: *Eur. Phys. J. C* 77.5 (2017), p. 317. DOI: [10.1140/epjc/s10052-017-4852-3](https://doi.org/10.1140/epjc/s10052-017-4852-3). arXiv: [1611.09661](https://arxiv.org/abs/1611.09661) [hep-ex].
- [30] Georges Aad et al. “Performance of the ATLAS Trigger System in 2010”. In: *Eur. Phys. J. C* 72 (2012), p. 1849. DOI: [10.1140/epjc/s10052-011-1849-1](https://doi.org/10.1140/epjc/s10052-011-1849-1). arXiv: [1110.1530](https://arxiv.org/abs/1110.1530) [hep-ex].
- [31] Gavin P. Salam. “Towards Jetography”. In: *Eur. Phys. J. C* 67 (2010), pp. 637–686. DOI: [10.1140/epjc/s10052-010-1314-6](https://doi.org/10.1140/epjc/s10052-010-1314-6). arXiv: [0906.1833](https://arxiv.org/abs/0906.1833) [hep-ph].
- [32] Gavin P. Salam. “Elements of QCD for hadron colliders”. In: *High-energy physics. Proceedings, 17th European School, ESHEP 2009, Bautzen, Germany, June 14-27, 2009*. 2010. arXiv: [1011.5131](https://arxiv.org/abs/1011.5131) [hep-ph]. URL: <https://inspirehep.net/record/880643/files/arXiv:1011.5131.pdf>.
- [33] Stephen D. Ellis and Davison E. Soper. “Successive combination jet algorithm for hadron collisions”. In: *Phys. Rev D* 48 (1993), 3160–3166.
- [34] S. Catani et al. “Longitudinally-invariant k_t -clustering algorithms for hadron-hadron collisions”. In: *Nuclear Physics B* 406.1 (1993), pp. 187–224. ISSN: 0550-3213. DOI: [https://doi.org/10.1016/0550-3213\(93\)90166-M](https://doi.org/10.1016/0550-3213(93)90166-M). URL: <http://www.sciencedirect.com/science/article/pii/055032139390166M>.

- [35] Yuri L. Dokshitzer et al. “Better jet clustering algorithms”. In: *JHEP* 08 (1997), p. 001. DOI: [10.1088/1126-6708/1997/08/001](https://doi.org/10.1088/1126-6708/1997/08/001). arXiv: [hep-ph/9707323](https://arxiv.org/abs/hep-ph/9707323) [hep-ph].
- [36] M. Cacciari, G.P. Salam, and G. Soyez. “The anti- k_t jet clustering algorithm”. In: *JHEP* 04 (2008), p. 063. DOI: [10.1088/1126-6708/2008/04/063](https://doi.org/10.1088/1126-6708/2008/04/063). arXiv: [0802.1189](https://arxiv.org/abs/0802.1189) [hep-ex].
- [37] Matteo Cacciari and Gavin P. Salam. “Pileup subtraction using jet areas”. In: *Phys. Lett. B* 659 (2008), pp. 119–126. DOI: [10.1016/j.physletb.2007.09.077](https://doi.org/10.1016/j.physletb.2007.09.077). arXiv: [0707.1378](https://arxiv.org/abs/0707.1378) [hep-ph].
- [38] Matteo Cacciari and Gavin P. Salam. “Dispelling the N^3 myth for the k_t jet-finder”. In: *Phys. Lett. B* 641.1 (2006), pp. 57–61. ISSN: 0370-2693. DOI: [10.1016/j.physletb.2006.08.037](https://doi.org/10.1016/j.physletb.2006.08.037).
- [39] M. Cacciari, G. P. Salam and G. Soyez. *The FastJet Project*. <http://fastjet.fr/>.
- [40] “Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1”. In: *Eur. Phys. J. C* 77 (2017), p. 490. URL: <https://doi.org/10.1140/epjc/s10052-017-5004-5>.
- [41] Georges Aad et al. “Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 7$ TeV”. In: *Eur. Phys. J. C* 73.3 (2013), p. 2304. DOI: [10.1140/epjc/s10052-013-2304-2](https://doi.org/10.1140/epjc/s10052-013-2304-2). arXiv: [1112.6426](https://arxiv.org/abs/1112.6426) [hep-ex].
- [42] Morad Aaboud et al. “Jet reconstruction and performance using particle flow with the ATLAS Detector”. In: *Eur. Phys. J. C* 77.7 (2017), p. 466. DOI: [10.1140/epjc/s10052-017-5031-2](https://doi.org/10.1140/epjc/s10052-017-5031-2). arXiv: [1703.10485](https://arxiv.org/abs/1703.10485) [hep-ex].
- [43] ATLAS Collaboration. “A new method to distinguish hadronically decaying boosted Z bosons from W bosons using the ATLAS detector”. In: *Eur. Phys. J. C* 76 (2016), p. 238. DOI: [10.1140/epjc/s10052-016-4065-1](https://doi.org/10.1140/epjc/s10052-016-4065-1). arXiv: [1509.04939](https://arxiv.org/abs/1509.04939) [hep-ex].
- [44] ATLAS Collaboration. “Identification of high transverse momentum top quarks in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector”. In: *JHEP* 06 (2016), p. 093. DOI: [10.1007/JHEP06\(2016\)093](https://doi.org/10.1007/JHEP06(2016)093). arXiv: [1603.03127](https://arxiv.org/abs/1603.03127) [hep-ex].
- [45] ATLAS Collaboration. “Identification of Boosted, Hadronically Decaying W Bosons and Comparisons with ATLAS Data Taken at $\sqrt{s} = 8$ TeV”. In: *Eur. Phys. J. C* 76 (2016), p. 154. DOI: [10.1140/epjc/s10052-016-3978-z](https://doi.org/10.1140/epjc/s10052-016-3978-z). arXiv: [1510.05821](https://arxiv.org/abs/1510.05821) [hep-ex].

- [46] Tatsumi Nitta and. “Identification of hadronically decaying W bosons and top quarks using multivariate techniques at ATLAS”. In: *Journal of Physics: Conference Series* 1085 (Sept. 2018), p. 042012. DOI: [10.1088/1742-6596/1085/4/042012](https://doi.org/10.1088/1742-6596/1085/4/042012). URL: <https://doi.org/10.1088/1742-6596/1085/4/042012>.
- [47] Morad Aaboud et al. “Performance of top-quark and W-boson tagging with ATLAS in Run 2 of the LHC”. In: (2018). arXiv: [1808.07858](https://arxiv.org/abs/1808.07858) [hep-ex].
- [48] Andrew J. Larkoski, Ian Moulton, and Duff Neill. “Power Counting to Better Jet Observables”. In: *JHEP* 12 (2014), p. 009. DOI: [10.1007/JHEP12\(2014\)009](https://doi.org/10.1007/JHEP12(2014)009). arXiv: [1409.6298](https://arxiv.org/abs/1409.6298) [hep-ph].
- [49] Jesse Thaler and Ken Van Tilburg. “Identifying Boosted Objects with N-subjettiness”. In: *JHEP* 03 (2011), p. 015. DOI: [10.1007/JHEP03\(2011\)015](https://doi.org/10.1007/JHEP03(2011)015). arXiv: [1011.2268](https://arxiv.org/abs/1011.2268) [hep-ph].
- [50] Geoffrey C. Fox and Stephen Wolfram. “Observables for the Analysis of Event Shapes in e^+e^- Annihilation and Other Processes”. In: *Phys. Rev. Lett.* 41 (23 Dec. 1978), pp. 1581–1585. DOI: [10.1103/PhysRevLett.41.1581](https://doi.org/10.1103/PhysRevLett.41.1581). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.41.1581>.
- [51] Leandro G. Almeida et al. “Top Jets at the LHC”. In: *Phys. Rev. D* 79 (2009), p. 074012. DOI: [10.1103/PhysRevD.79.074012](https://doi.org/10.1103/PhysRevD.79.074012). arXiv: [0810.0934](https://arxiv.org/abs/0810.0934) [hep-ph].
- [52] ATLAS Collaboration. “ATLAS measurements of the properties of jets for boosted particle searches”. In: *Phys. Rev. D* 86 (2012), p. 072006. DOI: [10.1103/PhysRevD.86.072006](https://doi.org/10.1103/PhysRevD.86.072006). arXiv: [1206.5369](https://arxiv.org/abs/1206.5369) [hep-ex].
- [53] Chunhui Chen. “New approach to identifying boosted hadronically-decaying particle using jet substructure in its center-of-mass frame”. In: *Phys. Rev. D* 85 (2012), p. 034007. DOI: [10.1103/PhysRevD.85.034007](https://doi.org/10.1103/PhysRevD.85.034007). arXiv: [1112.2567](https://arxiv.org/abs/1112.2567) [hep-ph].
- [54] Jesse Thaler and Lian-Tao Wang. “Strategies to Identify Boosted Tops”. In: *JHEP* 07 (2008), p. 092. DOI: [10.1088/1126-6708/2008/07/092](https://doi.org/10.1088/1126-6708/2008/07/092). arXiv: [0806.0023](https://arxiv.org/abs/0806.0023) [hep-ph].
- [55] ATLAS Collaboration. “Measurement of k_T splitting scales in $W \rightarrow \ell\nu$ events at $\sqrt{s} = 7$ TeV with the ATLAS detector”. In: *Eur. Phys. J. C* 73 (2013), p. 2432. DOI: [10.1140/epjc/s10052-013-2432-8](https://doi.org/10.1140/epjc/s10052-013-2432-8). arXiv: [1302.1415](https://arxiv.org/abs/1302.1415) [hep-ex].

- [56] M. Aaboud et al. “Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *Phys. Rev. D* 96.7 (2017), p. 072002. DOI: [10.1103/PhysRevD.96.072002](https://doi.org/10.1103/PhysRevD.96.072002). arXiv: [1703.09665](https://arxiv.org/abs/1703.09665) [hep-ex].
- [57] Stefano Frixione and Giovanni Ridolfi. “Jet photoproduction at HERA”. In: *Nucl. Phys. B* 507 (1997), pp. 315–333. DOI: [10.1016/S0550-3213\(97\)00575-0](https://doi.org/10.1016/S0550-3213(97)00575-0). arXiv: [hep-ph/9707345](https://arxiv.org/abs/hep-ph/9707345) [hep-ph].
- [58] Simone Alioli et al. “Jet pair production in POWHEG”. In: *JHEP* 04 (2011), p. 081. DOI: [10.1007/JHEP04\(2011\)081](https://doi.org/10.1007/JHEP04(2011)081). arXiv: [1012.3380](https://arxiv.org/abs/1012.3380) [hep-ph].
- [59] Victor Lendermann et al. “Combining Triggers in HEP Data Analysis”. In: *Nucl. Instrum. Meth. A* 604 (2009), pp. 707–718. DOI: [10.1016/j.nima.2009.03.173](https://doi.org/10.1016/j.nima.2009.03.173). arXiv: [0901.4118](https://arxiv.org/abs/0901.4118) [hep-ex].
- [60] Bogdan Malaescu. “An Iterative, Dynamically Stabilized(IDS) Method of Data Unfolding”. In: *Proceedings, PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, CERN, Geneva, Switzerland 17-20 January 2011*. 2011, pp. 271–275. DOI: [10.5170/CERN-2011-006.271](https://doi.org/10.5170/CERN-2011-006.271). arXiv: [1106.3107](https://arxiv.org/abs/1106.3107) [physics.data-an]. URL: <https://inspirehep.net/record/913906/files/arXiv:1106.3107.pdf>.
- [61] Bogdan Malaescu. “An Iterative, dynamically stabilized method of data unfolding”. In: (2009). arXiv: [0907.3791](https://arxiv.org/abs/0907.3791) [physics.data-an].
- [62] Tancredi Carli et al. “A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: The APPLGRID Project”. In: *Eur. Phys. J. C* 66 (2010), pp. 503–524. DOI: [10.1140/epjc/s10052-010-1255-0](https://doi.org/10.1140/epjc/s10052-010-1255-0). arXiv: [0911.2985](https://arxiv.org/abs/0911.2985) [hep-ph].
- [63] Stephen D. Ellis, Zoltan Kunszt, and Davison E. Soper. “Two-jet production in hadron collisions at order α_s^3 in QCD”. In: *Phys. Rev. Lett.* 69 (10 Sept. 1992), pp. 1496–1499. DOI: [10.1103/PhysRevLett.69.1496](https://doi.org/10.1103/PhysRevLett.69.1496). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.69.1496>.
- [64] Andy Buckley et al. “LHAPDF6: parton density access in the LHC precision era”. In: *Eur. Phys. J. C* 75 (2015), p. 132. DOI: [10.1140/epjc/s10052-015-3318-8](https://doi.org/10.1140/epjc/s10052-015-3318-8). arXiv: [1412.7420](https://arxiv.org/abs/1412.7420) [hep-ph].

- [65] Sayipjamal Dulat et al. “New parton distribution functions from a global analysis of quantum chromodynamics”. In: *Phys. Rev. D* 93.3 (2016), p. 033006. DOI: [10.1103/PhysRevD.93.033006](#). arXiv: [1506.07443 \[hep-ph\]](#).
- [66] L. A. Harland-Lang et al. “Parton distributions in the LHC era: MMHT 2014 PDFs”. In: *Eur. Phys. J. C* 75.5 (2015), p. 204. DOI: [10.1140/epjc/s10052-015-3397-6](#). arXiv: [1412.3989 \[hep-ph\]](#).
- [67] Richard D. Ball et al. “Parton distributions for the LHC Run II”. In: *JHEP* 04 (2015), p. 040. DOI: [10.1007/JHEP04\(2015\)040](#). arXiv: [1410.8849 \[hep-ph\]](#).
- [68] S. Alekhin et al. “Parton distribution functions, α_s , and heavy-quark masses for LHC Run II”. In: *Phys. Rev. D* 96.1 (2017), p. 014011. DOI: [10.1103/PhysRevD.96.014011](#). arXiv: [1701.05838 \[hep-ph\]](#).
- [69] Jon Butterworth et al. “PDF4LHC recommendations for LHC Run II”. In: *J. Phys. G* 43 (2016), p. 023001. DOI: [10.1088/0954-3899/43/2/023001](#). arXiv: [1510.03865 \[hep-ph\]](#).
- [70] J. Pumplin et al. “New generation of parton distributions with uncertainties from global QCD analysis”. In: *JHEP* 07 (2002), p. 012. DOI: [10.1088/1126-6708/2002/07/012](#). arXiv: [hep-ph/0201195](#).
- [71] A. Martin et al. “Parton distributions for the LHC”. In: *Eur.Phys.J. C* 63 (2009) 189–285 (). eprint: [arXiv:0901.0002\[hep-ph\]](#).
- [72] Richard Corke and Torbjorn Sjostrand. “Interleaved Parton Showers and Tuning Prospects”. In: *JHEP* 03 (2011), p. 032. DOI: [10.1007/JHEP03\(2011\)032](#). arXiv: [1011.1759 \[hep-ph\]](#).
- [73] *Summary of ATLAS Pythia 8 tunes*. Tech. rep. ATL-PHYS-PUB-2012-003. Geneva: CERN, Aug. 2012. URL: <https://cds.cern.ch/record/1474107>.
- [74] Peter Skands, Stefano Carrazza, and Juan Rojo. “Tuning PYTHIA 8.1: the Monash 2013 Tune”. In: *Eur. Phys. J. C* 74.8 (2014), p. 3024. DOI: [10.1140/epjc/s10052-014-3024-y](#). arXiv: [1404.5630 \[hep-ph\]](#).
- [75] Stefan Gieseke, Christian Rohr, and Andrzej Siodmok. “Colour reconnections in Herwig++”. In: *Eur. Phys. J. C* 72 (2012), p. 2225. DOI: [10.1140/epjc/s10052-012-2225-5](#). arXiv: [1206.0041 \[hep-ph\]](#).

- [76] Michael H. Seymour and Andrzej Siodmok. “Constraining MPI models using σ_{eff} and recent Tevatron and LHC Underlying Event data”. In: *JHEP* 10 (2013), p. 113. DOI: [10.1007/JHEP10\(2013\)113](https://doi.org/10.1007/JHEP10(2013)113). arXiv: [1307.5015](https://arxiv.org/abs/1307.5015) [hep-ph].
- [77] Stefan Dittmaier, Alexander Huss, and Christian Speckner. “Weak radiative corrections to dijet production at hadron colliders”. In: *JHEP* 11 (2012), p. 095. DOI: [10.1007/JHEP11\(2012\)095](https://doi.org/10.1007/JHEP11(2012)095). arXiv: [1210.0438](https://arxiv.org/abs/1210.0438) [hep-ph].
- [78] Georges Aad et al. “Measurement of dijet cross sections in pp collisions at 7 TeV centre-of-mass energy using the ATLAS detector”. In: *JHEP* 05 (2014), p. 059. DOI: [10.1007/JHEP05\(2014\)059](https://doi.org/10.1007/JHEP05(2014)059). arXiv: [1312.3524](https://arxiv.org/abs/1312.3524) [hep-ex].
- [79] James Dolen et al. “Thinking outside the ROCs: Designing Decorrelated Taggers (DDT) for jet substructure”. In: *JHEP* 05 (2016), p. 156. DOI: [10.1007/JHEP05\(2016\)156](https://doi.org/10.1007/JHEP05(2016)156). arXiv: [1603.00027](https://arxiv.org/abs/1603.00027) [hep-ph].
- [80] S. A. Dudani. “The Distance-Weighted k-Nearest-Neighbor Rule”. In: *IEEE Transactions on Systems, Man, and Cybernetics* SMC-6.4 (Apr. 1976), pp. 325–327. ISSN: 0018-9472. DOI: [10.1109/TSMC.1976.5408784](https://doi.org/10.1109/TSMC.1976.5408784).
- [81] Ian Moulton, Benjamin Nachman, and Duff Neill. “Convolved Substructure: Analytically Decorrelating Jet Substructure Observables”. In: *JHEP* 05 (2018), p. 002. DOI: [10.1007/JHEP05\(2018\)002](https://doi.org/10.1007/JHEP05(2018)002). arXiv: [1710.06859](https://arxiv.org/abs/1710.06859) [hep-ph].
- [82] Justin Stevens and Mike Williams. “uBoost: A boosting method for producing uniform selection efficiencies from multivariate classifiers”. In: *JINST* 8 (2013), P12013. DOI: [10.1088/1748-0221/8/12/P12013](https://doi.org/10.1088/1748-0221/8/12/P12013). arXiv: [1305.7248](https://arxiv.org/abs/1305.7248) [nucl-ex].
- [83] <https://stats.stackexchange.com/bayes-optimal-classifier-vs-likelihood-ratio>.
- [84] Abien Fred Agarap. “Deep Learning using Rectified Linear Units (ReLU)”. In: *CoRR* abs/1803.08375 (2018). arXiv: [1803.08375](https://arxiv.org/abs/1803.08375). URL: <http://arxiv.org/abs/1803.08375>.
- [85] Morad Aaboud et al. “Search for low-mass dijet resonances using trigger-level jets with the ATLAS detector in pp collisions at $\sqrt{s} = 13$ TeV”. In: *Phys. Rev. Lett.* 121.8 (2018), p. 081801. DOI: [10.1103/PhysRevLett.121.081801](https://doi.org/10.1103/PhysRevLett.121.081801). arXiv: [1804.03496](https://arxiv.org/abs/1804.03496) [hep-ex].
- [86] CMS Collaboration. “Inclusive search for a highly boosted Higgs boson decaying to a bottom quark-antiquark pair”. In: *Phys. Rev. Lett.* 120.7 (2018), p. 071802. DOI: [10.1103/PhysRevLett.120.071802](https://doi.org/10.1103/PhysRevLett.120.071802). arXiv: [1709.05543](https://arxiv.org/abs/1709.05543) [hep-ex].

- [87] 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”. In: (2018). arXiv: [1801.08769 \[hep-ex\]](#).
- [88] S. Kullback and R. A. Leibler. “On Information and Sufficiency”. In: *Ann. Math. Statist.* 22.1 (Mar. 1951), pp. 79–86. DOI: [10.1214/aoms/1177729694](#). URL: <https://doi.org/10.1214/aoms/1177729694>.
- [89] J. Lin. “Divergence measures based on the Shannon entropy”. In: *IEEE Transactions on Information Theory* 37.1 (Jan. 1991), pp. 145–151. ISSN: 0018-9448. DOI: [10.1109/18.61115](#).
- [90] Chase Shimmin et al. “Decorrelated Jet Substructure Tagging using Adversarial Neural Networks”. In: *Phys. Rev. D* 96.7 (2017), p. 074034. DOI: [10.1103/PhysRevD.96.074034](#). arXiv: [1703.03507 \[hep-ex\]](#).
- [91] Ian J. Goodfellow et al. “Generative Adversarial Networks”. In: (2014). arXiv: [1406.2661](#).
- [92] Gilles Louppe, Michael Kagan, and Kyle Cranmer. “Learning to Pivot with Adversarial Networks”. In: *Advances in Neural Information Processing Systems 30*. Ed. by I. Guyon et al. Curran Associates, Inc., 2017, pp. 981–990. URL: <http://papers.nips.cc/paper/6699-learning-to-pivot-with-adversarial-networks.pdf>.
- [93] Michela Paganini, Luke de Oliveira, and Benjamin Nachman. “CaloGAN : Simulating 3D high energy particle showers in multilayer electromagnetic calorimeters with generative adversarial networks”. In: *Phys. Rev. D* 97.1 (2018), p. 014021. DOI: [10.1103/PhysRevD.97.014021](#). arXiv: [1712.10321 \[hep-ex\]](#).
- [94] K. Schawinski et al. “Generative Adversarial Networks recover features in astrophysical images of galaxies beyond the deconvolution limit”. In: (2017). arXiv: [1706.04156](#).
- [95] Kaustuv Datta, Deepak Kar, and Debarati Roy. “Unfolding with Generative Adversarial Networks”. In: (2018). arXiv: [1806.00433 \[physics.data-an\]](#).
- [96] Christopher M. Bishop and Nasser M. Nasrabadi. “Mixture Density Networks”. In: *Pattern Recognition and Machine Learning*. Vol. 16. 4. 2007. Chap. 5, p. 272. DOI: [10.1117/1.2819119](#). URL: <https://doi.org/10.1117/1.2819119>.
- [97] Ian J. Goodfellow. “NIPS 2016 Tutorial: Generative Adversarial Networks”. In: *CoRR* (2017). arXiv: [1701.00160](#).

- [98] François Chollet et al. *Keras*. <https://github.com/fchollet/keras>. Version 2.1.5. 2018.
- [99] Martín Abadi et al. “TensorFlow: A System for Large-scale Machine Learning”. In: *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation*. OSDI’16. Savannah, GA, USA: USENIX Association, 2016, pp. 265–283. ISBN: 978-1-931971-33-1. URL: <http://dl.acm.org/citation.cfm?id=3026877.3026899>.
- [100] Yaroslav Ganin et al. “Domain-Adversarial Training of Neural Networks”. In: *Journal of Machine Learning Research* 17.59 (2016), pp. 1–35. URL: <http://jmlr.org/papers/v17/15-239.html>.
- [101] Softmax function. *Softmax function — Wikipedia, The Free Encyclopedia*. URL: https://en.wikipedia.org/wiki/Softmax_function.
- [102] K. Pearson. “LIII. On lines and planes of closest fit to systems of points in space”. In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2.11 (1901), pp. 559–572. DOI: [10.1080/14786440109462720](https://doi.org/10.1080/14786440109462720). URL: <https://doi.org/10.1080/14786440109462720>.
- [103] Jasper Snoek, Hugo Larochelle, and Ryan P Adams. “Practical Bayesian Optimization of Machine Learning Algorithms”. In: *Advances in Neural Information Processing Systems* 25. Ed. by F. Pereira et al. Curran Associates, Inc., 2012, pp. 2951–2959. URL: <http://papers.nips.cc/paper/4522-practical-bayesian-optimization-of-machine-learning-algorithms.pdf>.
- [104] Carl Edward Rasmussen. “Gaussian processes for machine learning”. In: MIT Press, 2006.
- [105] URL: <https://towardsdatascience.com/a-conceptual-explanation-of-bayesian-model-based-hyperparameter-optimization-for-machine-learning-b8172278050f>.
- [106] V. Nagarajan and J. Z. Kolter. “Gradient descent GAN optimization is locally stable”. In: (). arXiv: [1706.04156](https://arxiv.org/abs/1706.04156).
- [107] B. Efron. “Bootstrap Methods: Another Look at the Jackknife”. In: *Ann. Statist.* 7.1 (Jan. 1979), pp. 1–26. DOI: [10.1214/aos/1176344552](https://doi.org/10.1214/aos/1176344552). URL: <https://doi.org/10.1214/aos/1176344552>.
- [108] S. Alekhin et al. “HERAFitter”. In: *Eur. Phys. J. C* 75.7 (2015), p. 304. DOI: [10.1140/epjc/s10052-015-3480-z](https://doi.org/10.1140/epjc/s10052-015-3480-z). arXiv: [1410.4412](https://arxiv.org/abs/1410.4412) [hep-ph].

- [109] R. Wigmans. “Calorimetry: Energy measurement in particle physics”. In: *Int. Ser. Monogr. Phys.* 107 (2000), p. 57.
- [110] G. D’Agostini. “On the use of the covariance matrix to fit correlated data”. In: *Nuclear Instruments and Methods in Physics Research A* 346 (1994), pp. 306–311.
- [111] V. Blobel. *Some Comments on χ^2 Minimization Applications*. <http://www.slac.stanford.edu/econf/C030908/proceedings.html>. 2003.
- [112] CMS Collaboration. “Search for Low Mass Vector Resonances Decaying to Quark-Antiquark Pairs in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV”. In: *Phys. Rev. Lett.* 119.11 (2017), p. 111802. DOI: [10.1103/PhysRevLett.119.111802](https://doi.org/10.1103/PhysRevLett.119.111802). arXiv: [1705.10532](https://arxiv.org/abs/1705.10532) [hep-ex].
- [113] Yoav Freund and Robert E Schapire. “A Short Introduction to Boosting”. In: 14 (Oct. 1999), pp. 771–780.
- [114] Morad Aaboud et al. “Observation of $H \rightarrow b\bar{b}$ decays and VH production with the ATLAS detector”. In: (2018). arXiv: [1808.08238](https://arxiv.org/abs/1808.08238) [hep-ex].
- [115] Martin Flechl. “Identification and energy calibration of hadronic tau lepton decays at the LHC”. In: *5th Large Hadron Collider Physics Conference (LHCP 2017) Shanghai, China, May 15-20, 2017*. 2017. arXiv: [1709.01351](https://arxiv.org/abs/1709.01351) [hep-ex].
- [116] Roel Aaij et al. “Measurement of angular and CP asymmetries in $D^0 \rightarrow \pi^+\pi^-\mu^+\mu^-$ and $D^0 \rightarrow K^+K^-\mu^+\mu^-$ decays”. In: *Phys. Rev. Lett.* 121.9 (2018), p. 091801. DOI: [10.1103/PhysRevLett.121.091801](https://doi.org/10.1103/PhysRevLett.121.091801). arXiv: [1806.10793](https://arxiv.org/abs/1806.10793) [hep-ex].
- [117] Francesco Armando Di Bello. “Optimisation of the ATLAS b -tagging algorithms for the 2017-2018 LHC data-taking”. In: *PoS EPS-HEP2017* (2017), p. 733. DOI: [10.22323/1.314.0733](https://doi.org/10.22323/1.314.0733).
- [118] Alex Rogozhnikov et al. *hep_ml: Machine Learning for High Energy Physics*. https://github.com/arogozhnikov/hep_ml. Version 0.5.0. 2017.