

Université de Montréal

**Improving the search for new physics and the identification of
electrons using machine learning at the ATLAS experiment**

par

Bruna Pascual Dias

Département de physique
Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures et post-doctorales
en vue de l'obtention du grade de
Philosophiæ Doctor (Ph.D.)
en Physique

Août, 2024

Université de Montréal

Faculté des études supérieures et postdoctorales

Cette thèse intitulée

**Improving the search for new physics and the identification of
electrons using machine learning at the ATLAS experiment**

présenté par

Bruna Pascual Dias

a été évaluée par un jury composé des personnes suivantes :

David London	Président
--------------	-----------

Alain Bellerive	Examineur externe
-----------------	-------------------

Alan Robinson	Membre du jury
---------------	----------------

Jean-François Arguin	Directeur de recherche
----------------------	------------------------

Abstract

The study of high-energy collisions by the ATLAS experiment at the Large Hadron Collider (LHC) is essential to test the validity of the Standard Model of particle physics (SM), the current theoretical framework that describes the fundamental particles and how they interact, as well as to constrain its possible extensions. In light of its third data-taking period and the next generation of accelerators, the ATLAS experiment faces challenges associated with the high-dimensionality of the recorded signals and the large amount of data still left unexplored.

In this context, the use of deep learning techniques has great potential to enhance the performance of the classification of the physics objects that originate from these signals, as well as to provide new tools to perform fast statistical inference from the data. This thesis presents applications of these deep learning techniques to improve the electron identification algorithm performance, as well as a new strategy to search for resonances in invariant mass distributions with the ATLAS experiment.

Firstly, new measurements of the efficiency of the current electron identification algorithm are presented, using data recorded at the beginning of Run 3, as well as the reprocessing of the Run 2 data with the new version of the ATLAS software. A small reduction on the discrepancies between the values obtained from MC simulated events and the data is observed, a consequence of the improvements made to the ATLAS software at the start of Run 3.

Next, the development of a new electron identification algorithm is presented, where low-level detector information is processed in the form of images via a convolutional neural network. A study of the importance of its input features shows the relevance of all the current inputs considered. Furthermore, the rejection of the larger background class is decreased when a model trained with examples from simulated events is used to reject those obtained from experimental data. We show this rejection power is recovered if these data examples are incorporated into the training.

Lastly, a novel strategy to search for resonances in invariant mass histograms is presented. It uses a neural network to predict the local statistical significance of resonances from its bin entries. The implementation of this method using realistic simulation data shows good results, with the prediction of the maximum significance within a histogram having no bias and a small variance. Work towards an implementation of this method within the ATLAS experiment is also presented, including the production of invariant mass histograms using ATLAS simulation data.

Abstract

Keywords: particle physics, LHC, ATLAS, deep learning, neural networks, electron identification, anomaly detection.

Résumé

L'étude des collisions à haute énergie par l'expérience ATLAS au Grand Collisionneur d'Hadrons (LHC, de l'anglais *Large Hadron Collider*) est essentielle pour tester la validité du modèle standard de la physique des particules (SM, de l'anglais *Standard Model of particle physics*), le cadre théorique actuel qui décrit les particules fondamentales et leurs interactions, ainsi que pour établir des limites dans ses possibles extensions. Compte tenu du début de sa troisième période de prise de données et de la prochaine génération d'accélérateurs, l'expérience ATLAS est confrontée à des défis liés à la haute dimensionnalité des signaux enregistrés et à la grande quantité de données encore inexplorées.

Dans ce contexte, l'utilisation de techniques d'apprentissage profond a un grand potentiel pour améliorer la performance de la classification des objets physiques à l'origine de ces signaux, ainsi que pour fournir de nouveaux outils pour effectuer une inférence statistique rapide à partir des données. Cette thèse présente des applications des techniques d'apprentissage profond dans l'algorithme d'identification des électrons, ainsi qu'une nouvelle stratégie de recherche de résonances dans les distributions de masse invariantes avec l'expérience ATLAS.

Tout d'abord, de nouvelles mesures de l'efficacité de l'algorithme actuel d'identification des électrons sont présentées, en utilisant les données enregistrées au début du Run 3, ainsi que le retraitement des données du Run 2 avec la nouvelle version du logiciel utilisé par ATLAS. Une légère réduction des écarts entre les valeurs obtenues à partir des événements simulés et les données expérimentales est observée, une conséquence des améliorations apportées au logiciel utilisé par ATLAS avant le début du Run 3.

Ensuite, le développement d'un nouvel algorithme d'identification des électrons est présenté, où les informations brutes du détecteur sont traitées sous forme d'images via un réseau neuronal convolutif. Une étude de l'importance de ses variables d'entrée montre la pertinence de toutes les variables considérées actuellement. En outre, le rejet de la classe de bruit de fond la plus importante diminue lorsqu'un modèle entraîné avec des exemples provenant des simulations est utilisé pour rejeter des exemples obtenus à partir de données expérimentales. Nous montrons que ce pouvoir de rejet est récupéré si ces exemples de données sont incorporés dans la formation.

Enfin, nous présentons une nouvelle stratégie de recherche de résonances dans des histogrammes de masse invariante. Elle utilise un réseau neuronal pour prédire de la signification statistique locale des résonances à partir des entrées dans l'histogramme. La mise en œuvre de cette méthode à l'aide de données de simulation réalistes donne de bons résultats, avec la

prédiction de l'importance maximale au sein d'un histogramme ne présentant aucun biais et peu de variance. Les travaux visant à mettre en œuvre cette méthode dans le cadre de l'expérience ATLAS sont également présentés, y compris la production d'histogrammes de masse invariants à l'aide de données de simulation ATLAS.

Mots-clés: physique des particules, LHC, ATLAS, apprentissage profond, réseaux neuronaux, identification des électrons, détection des anomalies.

Contents

Abstract	i
Résumé	iii
List of Tables	ix
List of Figures	xi
List of acronyms	xxiii
Acknowledgements	xxv
Statement of personal contributions	xxvii
Introduction	1
1 The Standard Model of particle physics	5
1.1 Fundamental particles	6
1.1.1 Fermions	7
1.1.2 Bosons	7
1.1.3 Hadrons	8
1.2 Quantum field theory and symmetries	9
1.3 Fermionic fields	11
1.4 Gauge fields	12
1.5 Electroweak sector and Higgs mechanism	13
1.6 Quantum chromodynamics	16
1.7 Limitations of the Standard Model and open questions	18
1.8 Beyond the Standard Model	21
2 Statistics and machine learning	25
2.1 Introduction	26
2.2 Fundamental concepts	27
2.3 Parameter estimation	29
2.3.1 Likelihood function	30
2.3.2 Maximum likelihood estimator	31

2.4	Hypothesis testing	32
2.4.1	Profile likelihood ratio test	34
2.5	Figures of merit and ROC curves	36
2.6	Machine learning	39
2.6.1	Motivation	39
2.6.2	Basic concepts	40
2.6.3	Fully-connected network (FCN)	42
2.6.4	Convolutional neural network (CNN)	45
2.7	Applications of machine learning on high-energy physics	46
2.7.1	Classification	48
2.7.2	Anomaly detection	48
3	The ATLAS experiment	51
3.1	The Large Hadron Collider (LHC)	51
3.1.1	Introduction	51
3.1.2	Creating a proton collision	53
3.2	The ATLAS experiment	54
3.2.1	Inner detector	56
3.2.2	Calorimeters	59
3.2.3	Muon spectrometer	61
3.2.4	Trigger and data acquisition	61
3.2.5	Particles and event reconstruction	62
3.2.6	Event generation and detector simulation	63
3.3	Future upgrades	64
4	Electron reconstruction and identification	67
4.1	Introduction	67
4.2	Electron reconstruction	68
4.3	Electron identification	69
4.3.1	Likelihood-based identification	72
4.4	Electron isolation	74
5	Measurement of electron identification efficiency	79
5.1	Motivation for efficiency and scale factors measurement	80
5.2	Tag-and-probe method	82
5.3	Electron identification efficiencies	83
5.4	The Z -mass method	85
5.4.1	Event selection and input preparation	86
5.4.2	Estimation of the background with a normalisation in the mass tail	90
5.4.3	Estimation of the background with a likelihood fit	93

5.4.4	Improvements with respect to the previous implementation	96
5.4.5	Efficiency and scale factors	98
5.4.6	Systematic variations	100
5.5	Results	101
5.5.1	Run 3 (2022)	102
5.5.2	Run 2 (2018)	104
5.6	Future improvements	111
6	Improving electron identification with a convolutional neural network	113
6.1	Motivation for the use of a convolutional neural network	114
6.2	Input features	115
6.2.1	High-level variables (HLV)	115
6.2.2	Track variables	116
6.2.3	Calorimeter images	116
6.3	Task and architecture	120
6.4	Datasets	123
6.4.1	Simulation data	123
6.4.2	Experimental data	125
6.5	Results	127
6.5.1	Performance over simulation data	127
6.5.2	Feature ranking	131
6.5.3	Performance in data versus simulation	136
6.6	Future improvements	141
7	Bump search with the data-directed paradigm	145
7.1	State of current searches	146
7.2	Resonance searches with anomaly detection	148
7.3	Data-directed paradigm	150
7.4	Bump search with a neural network	152
7.4.1	Task and architecture	152
7.4.2	Synthetic data generation	154
7.5	Histogram processing	157
7.5.1	Dark Machines dataset	159
7.5.2	Processing and calibration	161
7.6	Training	164
7.7	Performance over synthetic data	165
7.7.1	Impact of background shapes	173
7.7.2	Impact of systematic uncertainties	175
7.7.3	Impact of signal width	176
7.8	Performance when looking for new physics	177

Contents

7.9	Histogram production from ATLAS simulation data	183
7.10	Outlook	185
	Conclusion	187
	Bibliography	189

List of Tables

1.1	Quantum numbers of the fermion fields under the Standard Model of particle physics (SM) symmetry group. They are the colour, the third component of the weak isospin I_3 and the weak hypercharge Y , associated with the $SU(3)_C$, $SU(2)_L$ and $U(1)_Y$ symmetry groups, respectively. The electric charge $Q_{\text{em}} = I_3 + Y$, associated with the electromagnetic interaction after the electroweak symmetry breaking, is also included.	12
2.1	Examples of probability density functions.	28
4.1	High-level variables from the tracks, track-cluster matching and calorimeter that are used to discriminate electrons of interest from their background in the context of the LLH identification method ^[245]	70
4.2	Isolation working points, defined with respect to the calorimeter-based and track-based isolation variables ^[266]	76
5.1	Selections performed to construct a sample of electrons unbiased by the identification algorithm via the tag-and-probe method.	88
5.2	Selections performed to define background objects.	88
5.3	Single-electron triggers used in the ATLAS detector over the Run 2 (2015–2018) and the early Run 3 (2022) data taking periods.	89
5.4	Binning on p_T and η_{BE2}	90
6.1	Additional high-level variables included in the CNN electron identification algorithm ^[2]	116
6.2	Track variables included in the CNN electron identification algorithm. Variables for up to 5 matching tracks j are considered for each candidate and are organised in the form of a grid to form a 5×13 tracks image that is processed by a dedicated CNN layer ^[2]	117

6.3	Processes considered to obtain electron candidates from simulated MC events and the corresponding ratio of electron candidates from each class (in %), identified from its truth information. The large majority of background examples correspond to the light-flavour (LF) hadrons and their decays into e/γ , with prompt electrons also being well represented. Charge-flip, photon conversion and the heavy flavour (HF) decay classes correspond to a considerably smaller ratio of the candidates available ^[2]	124
7.1	Analytical functions used to define smoothly falling background shapes, where y corresponds to the number of entries per bin and x are the invariant mass values. The parameters a and b are defined by setting two limits $(x_{\min}, y_{\min})$ and $(x_{\max}, y_{\max})$ which completely define the shape.	154
7.2	Processes and corresponding leading-order cross-section included in the simulation datasets from the Dark Machines challenge ^[219] . These correspond to the highest cross-section SM background processes produced from pp collisions with 13 TeV centre-of-mass energy at the LHC.	160
7.3	Selections performed within the simulated Dark Machines datasets. Different channels define non-orthogonal signal regions, such that channel 1 focus on hadronic activity with a lot of missing energy, channel 2 reduces the background by requiring leptons and channel 3 is the most inclusive one. For more details on the selections, we refer to Ref. [219].	160
7.4	Kinematic cuts used in the Dark Machines histogram processing to define its sub-categories. They correspond to limits on the missing transverse energy (E_T^{miss}) or transverse momentum p_T of leading objects, namely electrons (e), muons (μ), as well as the reconstructed leptonically-decaying Z boson (Z), hadronically-decaying Z and W bosons (V_h), boosted top quark (T) and high-mass jets (HM).	161
7.5	False discovery rate in percentage of the total number of histograms and corresponding number of histograms obtained from the PLR test and the BumpNet predictions over the different background datasets. Synthetic background is generated by Poisson fluctuation of fits to Dark Machines backgrounds, while raw channel 3 and channel 2b backgrounds correspond to the original histograms.	171
7.6	Outlier rate and corresponding number of histograms obtained from BumpNet predictions over the datasets with background only (B) and signal + background (S+B), using the corresponding variance σ for each dataset. Rates are given as percentages of the total number of histograms in the testing dataset.	172

List of Figures

1.1	Fundamental particles and their interactions within the standard model of particle physics.	6
1.2	Parton distribution functions for the proton as a function of the Bjorken- x , defined as the fraction of the proton momentum carried by the parton. They are shown for energy scales of $\mu^2 = 10 \text{ GeV}^2$ (left) and $\mu^2 = 10^4 \text{ GeV}^2$ (right), and were obtained in next-to-next-to-leading order (NNLO) by the MSHT20 global analysis of available hard scattering data. Taken from Ref. [45].	9
1.3	Illustration of the Higgs potential, Eq. (1.5), when $\mu^2 < 0$. Taken from Ref. [67].	14
1.4	Illustration of the hadronisation process. When a quark and an antiquark are separated, the energy stored in the colour field is enough to produce new quark–antiquark pairs. This process continues until all quarks are bound into colourless states. Adapted from Ref. [39].	17
1.5	Event display of the production of two jets (dijet) registered by the ATLAS experiment, showing the tracks recorded at the inner detector (orange lines) and the energy deposits in the calorimeters (green and yellow blocks), which will be described in Chapter 3. Taken from Ref. [74].	18
1.6	Summary of the total production cross-section measurements performed by the ATLAS experiment, compared to the theoretical predictions from the Standard Model. Taken from Ref. [77].	19
1.7	Loop-level corrections to the Higgs boson mass from (a) fermions, (b) gauge bosons and (c) other Higgs bosons. All these loop corrections must cancel out in order to keep the Higgs mass value around the measured 125 GeV.	20
1.8	Diagrams for the (a) RPV SUSY stop pair production and (b) scalar leptoquark pair production at the LHC. Coloured lines indicate the hypothetical particles. Being both scalar massive particles, each of them decays into a charged lepton and a b -quark.	22
1.9	Leading order Feynman diagrams for leptoquark pair production at the LHC ^[134] .	23
1.10	Leading order Feynman diagrams for Z' production in the (a) 3μ and (b) 4μ final states.	24

2.1	Probability density functions of the test statistic t assuming each one of the hypothesis H_0 and H_1 . We say H_0 is rejected if t is observed to be in the critical region $t > t_{\text{cut}}$. Adapted from Ref. [151].	32
2.2	Illustration of ROC curves for different classifiers, defined as the true positive rate as a function of the false positive rate. One can alternatively write it as the background rejection $1/\epsilon_{\text{bkg}}$ as a function of the signal efficiency ϵ_{sig}	38
2.3	(a) Diagram of a perceptron where the weighted sum of the inputs $\vec{x} \in \mathbb{R}^D$ is passed through a step function to provide an output. (b) Hyperplane defined by a perceptron, here a line for a two-dimensional input, that splits the input space into two regions and thus only work with data that is linearly separable. Adapted from [170].	43
2.4	Architecture of a multi-layer perceptron (MLP), also known as a fully connected network. The input $\vec{x} \in \mathbb{R}^D$ is associated with one of C labels y_c . The network shown contains just one hidden layer with M units. Each node corresponds to a perceptron, highlighted in grey. A constant value is added to the input as well as the hidden layer to account for a bias.	43
2.5	Illustration of the convolution operation. Adapted from Ref. [43].	45
3.1	The CERN accelerator complex, highlighting the ATLAS experiment, the LHC and its injection chain. Adapted from Ref. [225].	52
3.2	Illustration of the ATLAS detector. Taken from Ref. [240].	55
3.3	Schematic illustration of the ATLAS coordinate system. The z -axis is placed along the beam line, the y -axis points to the top and the x -axis points to the centre of the collider ring. The position of the particles is determined in terms of the azimuthal angle ϕ and the pseudorapidity $\eta = -\ln[\tan(\theta/2)]$, where θ is the polar angle. Adapted from Refs. [39, 241].	55
3.4	Particle signatures in the ATLAS detector. The inner detector (green) registers the path of charged particles. When particles enter the calorimeters (orange and blue), the succession of decays that follows forms a shower of electromagnetic or hadronic nature. The energy deposits are combined to determine the energy of the particle that originated the shower. Lastly, the muon spectrometer (gray) register the path of muons, the only charged particles that can traverse the detector this far. Adapted from Ref. [244].	57
3.5	Illustration of the ATLAS inner detector. Taken from Ref. [223].	58
3.6	Illustration of the ATLAS calorimeters. Taken from Ref. [223].	59
3.7	Illustration of the ATLAS muon systems. Taken from Ref. [223].	60

4.1	Illustration of the path of a prompt electron in the ATLAS detector. A track is formed by the combination of hits in the layers of the inner detector, namely the Pixel, SCT and TRT subsystems. The electron energy is then deposited in the electromagnetic calorimeter, and no signal is expected to be deposited in the hadronic calorimeter. Taken from Ref. [245].	68
4.2	Distributions of the (a) R_ϕ and (b) f_1 observables, defined in Table 4.1, for both signal (solid black line) and background (red dashed line) electrons. Both distributions are obtained from MC simulated events, where $Z \rightarrow e^+e^-$ processes are used to obtain signal and QCD two-to-two processes are used to obtain background. The distributions have been smoothed and corrected. Taken from Ref. [245].	72
4.3	Electron identification efficiencies (top) measured as a function of (a) transverse energy E_T and (b) pseudorapidity η_{BE2} in both data (filled markers) and MC simulated events (open markers). The measurement was performed in data recorded by the ATLAS experiment in the full Run 2 data-taking period (2015–2018) from pp collisions at $\sqrt{s} = 13$ GeV, corresponding to an integrated luminosity of 139 fb^{-1} . Three working points from the likelihood identification algorithm were considered, namely, TIGHT, MEDIUM and LOOSE, with an extra B-Layer requirement on the latter. The ratio of the data efficiencies with respect to the MC ones, referred to as scale factors, is shown (middle) together with its total and statistical uncertainties (bottom). Taken from Ref. [266]. . .	73
4.4	Schema of the calorimeter isolation method, where the grid represents the second-layer calorimeter cells in the $\eta \times \phi$ space. The energy deposits (red patches) that have their barycentre within the isolation cone (purple circle) are included in the computation of the isolation variable, after removal of part of the cluster associated to the electron candidate (yellow rectangle). Taken from Ref. [245].	75
4.5	Electron isolation efficiencies (top) measured as a function of (a) transverse energy E_T and (b) pseudorapidity η . The measurement was performed in data recorded by the ATLAS experiment in the Run 2 2018 data-taking period from pp collisions at $\sqrt{s} = 13$ GeV, corresponding to an integrated luminosity of 58.5 fb^{-1} . The five isolation working points are considered, with electrons being required to pass the MEDIUM likelihood identification working point. The ratio of the data efficiencies with respect to the MC ones, referred to as scale factors, is shown (bottom). Taken from Ref. [266].	76

5.1	Examples of histograms used to estimate the background in the (a) Z -mass and (b) Z -iso methods. In the Z -mass method, the invariant-mass distributions of the electron pair m_{ee} is used. In the Z -iso method, the probe E_T^{cone30} isolation variable is used, defined as the sum of the transverse energy deposited in the calorimeters in a cone of $\Delta R \leq 0.30$. The signal regions used in the efficiency measurements, which correspond to the Z peak, are shown in gray. Taken from Ref. [266].	84
5.2	Invariant mass of the tag and probe pair m_{ee} for the denominator (left) and numerator (right) selections in selected p_T and η_{BE2} bins. The signal distributions obtained in data (black) and MC simulated events (orange) are drawn, together with the background template distribution (purple). The background estimate (blue) is obtained using the normalisation in the mass tail method, and it is combined to the MC signal shape to form the expected distribution (red).	92
5.3	Evolution of the background estimation shape obtained via the likelihood fit method as a function of the transverse momentum p_T , in particular for $p_T \in [15, 20]$ GeV (left), $p_T \in [20, 25]$ GeV (center) and $p_T \in [25, 30]$ GeV (right).	94
5.4	Invariant mass of the tag and probe pair m_{ee} for the denominator (left) and numerator (right) selections in selected p_T and η_{BE2} bins. The signal distributions obtained in data (black) and MC simulated events (orange) are drawn, together with the background template distribution (purple). The background estimate (blue) is obtained via a likelihood fit and it is combined to the MC signal shape to form the expected distribution (red).	97
5.5	Electron identification efficiencies (top) and scale factors (bottom) measured as a function of (a) transverse momentum p_T and (b) pseudorapidity η_{BE2} . The measurement was performed in data recorded by the ATLAS experiment in 2022 from pp collisions at $\sqrt{s} = 13.6$ GeV, corresponding to an integrated luminosity of 3.4 fb^{-1} . Three working points from the likelihood identification algorithm were considered, namely, TIGHT, MEDIUM and LOOSE, with an extra B-Layer requirement on the latter. The Z -mass method was performed to estimate the background, via a normalisation of the invariant mass distribution templates to the data in a mass region far from the Z peak.	102
5.6	Run 3 electron identification scale factors central values (left) and total uncertainty (right) for different values of p_T and η_{BE2} , for three likelihood working points, namely (a) TIGHT, (b) MEDIUM and (c) LOOSE, the latter with an extra B-layer requirement. The measurement was performed in data recorded by the ATLAS experiment in 2022 from pp collisions at $\sqrt{s} = 13.6$ GeV, corresponding to an integrated luminosity of 3.4 fb^{-1}	103

- 5.7 Electron identification efficiencies (top) and scale factors (bottom) measured as a function of (a) transverse momentum p_T and (b) pseudorapidity η_{BE2} . The measurement was performed in data recorded by the ATLAS experiment in 2018 from pp collisions at $\sqrt{s} = 13$ GeV, corresponding to an integrated luminosity of 58.7 fb^{-1} . Three working points from the likelihood identification algorithm were considered, namely, `TIGHT`, `MEDIUM` and `LOOSE`, with an extra B-Layer requirement on the latter. The Z -mass method was performed to estimate the background, via a likelihood fit of the invariant mass distribution templates to the data. 104
- 5.8 Electron identification scale factors central values (left) and total uncertainty (left) for different values of p_T and η_{BE2} , for three likelihood working points, namely (a) `TIGHT`, (b) `MEDIUM` and (c) `LOOSE`, the latter with an extra B-layer requirement. The measurement was performed using the software in its release 22 over data recorded by the ATLAS experiment in 2018 from pp collisions at $\sqrt{s} = 13$ GeV, corresponding to an integrated luminosity of 58.7 fb^{-1} 106
- 5.9 Relative difference on the central value of electron identification efficiencies measured with the ATLAS software release 22 with respect to their latest published results^[266] measured with its release 21, presented for different values of p_T and η_{BE2} . The measurement was performed in data recorded by the ATLAS experiment in 2018 from pp collisions at $\sqrt{s} = 13$ GeV, corresponding to an integrated luminosity of 58.7 fb^{-1} . Efficiencies were measured in both data and MC simulated events for the three likelihood working points, namely `TIGHT`, `MEDIUM` and `LOOSE`, the latter with an extra B-layer requirement. 107
- 5.10 Electron identification efficiencies measured for the `TIGHT` likelihood working point in MC simulated events with both the new ATLAS software release 22 (red) and their latest published results^[266] measured with its release 21 (black), as a function of p_T and η_{BE2} for selected bins. The ratio of the release 22 measurement with respect to the release 21 measurement is drawn, with an yellow band corresponding to its uncertainty. Both measurements were performed in data recorded by the ATLAS experiment in 2018 from pp collisions at $\sqrt{s} = 13$ GeV, corresponding to an integrated luminosity of 58.7 fb^{-1} . . . 108
- 5.11 Electron identification efficiencies measured for the `TIGHT` likelihood working point in experimental data with both the new ATLAS software release 22 (red) and their latest published results^[266] measured with its release 21 (black), as a function of p_T and η_{BE2} for selected bins. The ratio of the release 22 measurement with respect to the release 21 measurement is drawn, with an yellow band corresponding to its uncertainty. Both measurements were performed in data recorded by the ATLAS experiment in 2018 from pp collisions at $\sqrt{s} = 13$ GeV, corresponding to an integrated luminosity of 58.7 fb^{-1} . . . 109

5.12	Electron identification scale factors measured for the TIGHT likelihood working point with both the new ATLAS software release 22 (red) and their latest published results ^[266] measured with its release 21 (black), as a function of p_T and η_{BE2} for selected bins. The ratio of the release 22 measurement with respect to the release 21 measurement is drawn, with an yellow band corresponding to its uncertainty. The measurement was performed in data recorded by the ATLAS experiment in 2018 from pp collisions at $\sqrt{s} = 13$ GeV, corresponding to an integrated luminosity of 58.7 fb^{-1}	110
6.1	Average of the calorimeter images for the barrel layers for different classes. A total of 10^6 examples with $p_T \geq 4$ GeV and $ \eta \in [0.0, 2.47]$ were considered from the MC simulated events in the JF17 dataset and data recorded by the ATLAS detector in 2015. The light-flavour class corresponds to the combination of both the light-flavour hadrons and their decays into e/γ	118
6.2	Average of the calorimeter images for the end-cap layers for different classes. A total of 10^6 examples with $p_T \geq 4$ GeV and $ \eta \in [0.0, 2.47]$ were considered from the MC simulated events in the JF17 dataset and data recorded by the ATLAS detector in 2015. The light-flavour class corresponds to the combination of both the light-flavour hadrons and their decays into e/γ	119
6.3	Architecture of the CNN algorithm. The calorimeter images in both coarse and fine resolutions, as well as the track variables from up to five matched tracks are processed through separate convolution layers. The corresponding outputs are concatenated with the high-level variables and processed by a fully connected network (FCN). The output is a vector with the probability of the electron candidate to be classified as each one of the six classes.	122
6.4	Distributions of the (a) transverse energy E_T and (b) the absolute pseudorapidity $ \eta $ of the MC simulation examples for different classes. The same observables are shown in (c) and (d) for the dataset where the combined LF examples were obtained from real data.	125
6.5	Distributions of the LLH discriminant value for the LF class obtained from Data (black line) and MC (orange line), as well as the one for prompt electrons (blue line) from MC events for selected bins in E_T : (a) $10 \text{ GeV} < E_T < 15 \text{ GeV}$ and (b) $30 \text{ GeV} < E_T < 35 \text{ GeV}$	127

6.6	ROC curves from a CNN model trained and evaluated over MC simulated examples when rejecting the different background classes. The testing dataset consisted of electrons with $E_T > 15$ GeV and charge-flip electrons are considered a background class. Models were trained with different combinations of the input features, namely the HLV only (dotted red line), HLV + images (dash-dotted green line), HLV + tracks (dashed orange line) and the full CNN algorithm with HLV + tracks + images (solid blue line). The bands around the curves correspond to the statistical uncertainty associated with the size of the testing dataset. The values obtained from the LLH working points at different signal efficiencies are drawn, namely <i>Loose</i> (triangle), <i>Medium</i> (diamond) and <i>Tight</i> (circle), together with the results obtained by the LLH discriminant (solid gray line). Taken from Ref. [2].	129
6.7	Ratios of the background rejection of the (a) combined background and the (b) LF hadron class from a CNN model with respect to the LLH method and FCN network for different E_T bins. The CNN model was trained with HLV, tracks and images as input features, while the FCN values were obtained from a network trained only with HLV. Values were obtained for the signal efficiency corresponding to the <i>Loose</i> , <i>Medium</i> and <i>Tight</i> LLH working points. Taken from Ref. [2].	130
6.8	Ratios of the background rejection of the combined background from a CNN model with respect to (a) LLH method and (b) FCN network as a function of different E_T and $ \eta $ bins. The CNN model was trained with HLV, tracks and images as input features, while the FCN values were obtained from a network trained only with HLV. Values were obtained for the signal efficiency corresponding to the <i>Loose</i> LLH working point. Taken from Ref. [2].	130
6.9	Ranking of the input features according to the ratio R defined in Eq. (6.3) when discriminating the combined background and its different classes. Features include both the HLV (blue) and the images from the track-matching variables and calorimeter deposits (orange). They are presented in descending order of the ratio for the combined background, such that those listed higher have a larger impact when removed. Results are presented for electrons with $ \eta < 2.5$, corresponding to the whole coverage.	132
6.10	Ranking of the input features according to the ratio R defined in Eq. (6.3) when discriminating the combined background and its different classes. Features include both the HLV (blue) and the images from the track-matching variables and calorimeter deposits (orange). They are presented in descending order of the ratio for the combined background, such that those listed higher have a larger impact when removed. Results are presented for electrons with $ \eta < 1.3$, corresponding to the barrel region only.	133

- 6.11 Ranking of the input features according to the ratio R defined in Eq. (6.3) when discriminating the combined background and its different classes. Features include both the HLV (blue) and the images from the track-matching variables and calorimeter deposits (orange). They are presented in descending order of the ratio for the combined background, such that those listed higher have a larger impact when removed. Results are presented for electrons with $1.3 < |\eta| < 1.6$, corresponding to the gap region. 134
- 6.12 Ranking of the input features according to the ratio R defined in Eq. (6.3) when discriminating the combined background and its different classes. Features include both the HLV (blue) and the images from the track-matching variables and calorimeter deposits (orange). They are presented in descending order of the ratio for the combined background, such that those listed higher have a larger impact when removed. Results are presented for electrons with $1.6 < |\eta| < 2.5$, the end-cap region. 135
- 6.13 ROC curves from CNN models when rejecting the combined LF class. Models trained in MC are evaluated in (a) MC and in (b) Data, as well as models trained in Data are evaluated in (c) Data and in (d) MC. Different combinations of the input features are included, namely the HLV only (dash-dotted green line), HLV + tracks (dashed orange line) and the full CNN algorithm with HLV + tracks + images (solid blue line). The bands around the curves correspond to the statistical uncertainty associated with the size of the testing dataset. The values obtained from the LLH working points at different signal efficiencies, namely LOOSE (triangle), MEDIUM (diamond) and TIGHT (circle) are drawn. . . . 137
- 6.14 Discriminants of the CNN model with all input features (HLV, tracks and calorimeter images) when classifying prompt electrons from the combined light-flavour class. Results for a model trained in MC and evaluated on (a) MC and (b) Data, as well as a model trained in Data and evaluated on (c) Data and (d) MC. 138
- 6.15 Ratios of the combined LF class background rejection from a CNN model with respect to LLH methods as a function of E_T . The CNN model was trained with HLV, tracks and images as input features and results are shown for a model trained in MC and evaluated on (a) MC and (b) Data, as well as a model trained in Data and evaluated on (c) Data and (d) MC. Values were obtained for the signal efficiencies corresponding to the LLH working points, namely LOOSE, MEDIUM and TIGHT. 139

6.16	Ratios of the combined LF class background rejection from a CNN model with respect to LLH methods as a function of E_T and $ \eta $ bins. The CNN model was trained with HLV, tracks and images as input features and results are shown for a model trained in MC and evaluated on (a) MC and (b) Data, as well as a model trained in Data and evaluated on (c) Data and (d) MC. Values were obtained for the signal efficiencies corresponding to the LLH LOOSE working points.	140
6.17	Ratios of the combined LF class background rejection from a CNN model with respect to LLH methods as a function of E_T and $ \eta $ bins. The CNN model was trained with HLV, tracks and images as input features and results are shown for a model trained in MC and evaluated on (a) MC and (b) Data, as well as a model trained in Data and evaluated on (c) Data and (d) MC. Values were obtained for the signal efficiencies corresponding to the LLH TIGHT working points.	141
7.1	Distributions of the invariant mass of the diphoton candidates, presented as of the discovery of the Higgs bosons by the (a) ATLAS experiment [69] and the (b) CMS experiment [70].	146
7.2	Status of two-body resonance searches at the LHC as of 2020 as presented in Ref. [295], with each entry corresponding references to existing searches. Highlighted in orange are the final states that have not been probed yet. . . .	147
7.3	Smoothly falling shapes for the (a) inclusive jet cross-section as a function of the transverse momentum p_T and the (b) dijet cross-section as a function of m_{jj} . Taken from Ref. [75].	151
7.4	Architecture of the BumpNet neural network. A vector of the bin entries of a mass histogram is used as input and processed by 4 convolution layers in parallel. Their output is concatenated and passed through a fully connected layer that process all the representations for each bin separately. The network's output has the same size as the input and corresponds to the statistical significance value predicted for each bin.	153
7.5	Examples of (a) analytical functions and (b) fit function used as smoothly falling background shapes. The parameters of the analytical functions are obtained by fixing the limits $(x_{\min}, y_{\min})$ and $(x_{\max}, y_{\max})$. On the other hand, fits of the 4th-order log-polynomial function $\ln(y) = \sum_{n=1}^4 p_n x^n$ are obtained from the mass histograms constructed from simulation data, in this case the Dark Machines dataset.	155

7.6	Examples of invariant mass distributions generated from synthetic data. They are obtained from the combination of a smoothly falling background (orange) with a Gaussian signal (green), the latter being scaled to match a given strength. A Poisson noise is added to the combination to obtain the corresponding observed histogram (black). The profile likelihood ratio (PLR) test is used to obtain the target Z values (blue), which are the statistical significance of having a Gaussian signal centred in each bin with the background shape as the null hypothesis. The maximum Z value of the histogram, $Z_{\max}$ is quoted, matching the position and strength of the injected signal.	156
7.7	Visualisation of the significance calculation for different bins. Each significance value (blue dots) is the outcome of the PLR test, where the background (orange) corresponds to the null hypothesis. The alternative hypothesis for a given bin is the combination of this background and a test signal (blue line) corresponding to the Gaussian signal centre at the bin that maximises the likelihood (MLE). A maximum of the significance is found when the test signal overlaps exactly with the injected signal (green) that was used to generate the observed data (black).	158
7.8	Comparison of the (a) relative resolution and the (b) number of entries per bin as a function of the mass for a 3-jet and 4-jet system. The higher resolution observed at low mass is reflected in the histogram binning with smaller bin widths for this region.	163
7.9	Examples from the testing datasets with the target PLR significance (blue) and the corresponding network prediction (red). The observed histograms (black) are obtained from the combination of a smoothly falling background (orange) with a Gaussian signal (green), the latter being scaled to match a given strength. A Poisson noise is added to the combination. The maximum Z value, $Z_{\max}$, from both the target and the prediction are quoted, being expected to match the position and strength of the injected signal.	166
7.10	Distributions of (a) $Z_{\max}$ and (b) $\Delta Z_{\max}$ for BumpNet predictions compared to the PLR test results when evaluating in datasets with background-only (B) and background with signal injected (S+B).	167
7.11	Distributions of $\Delta Z_{\max}$ as a function of (a) signal strength obtained via the PLR test $Z_{\max}^{\text{PLR}}$, (b) the signal position as a ratio of the histogram binning, (c) the number of bins in the histograms and (d) the number of entries in the first bin of the histogram.	168
7.12	Distributions of $\Delta Z_{\max}$ as a function of the signal strength obtained via the PLR test $Z_{\max}^{\text{PLR}}$ when evaluating over testing datasets with backgrounds from (a) functions only and (b) fits to Dark Machines backgrounds only.	168

7.13	(a) ROC curves for both the BumpNet model and the PLR test, as well as the corresponding AUC score. (b) Visualisation of the definitions for sensitivity (TP), specificity (TN), type 1 (FP) and type 2 (FN) errors from the $Z_{\max}$ distributions.	170
7.14	Examples of false discoveries obtained by the BumpNet model when evaluating over the raw channel 3 Dark Machines backgrounds.	170
7.15	Distributions of $\Delta Z_{\max}$ over the (a) background-only (B) and (b) signal+background (S+B) testing datasets, which have mean μ and variance σ . Predictions with $ \Delta Z_{\max} > 5 \times \sigma$ (red) are considered outliers.	172
7.16	Confidence interval for the predicted $Z_{\max}^{\text{BumpNet}}$ as a function of the target $Z_{\max}^{\text{PLR}}$ from a signal+background testing dataset.	173
7.17	$\Delta Z_{\max}$ as a function of $Z_{\max}^{\text{PLR}}$ for a model trained in functions only when evaluated on (a) functions and (b) fits to channel 3 and channel 2b backgrounds.	174
7.18	Examples of predictions of model trained on backgrounds from analytical functions only and evaluated over raw backgrounds from the Dark Machines dataset.	174
7.19	Performance over testing datasets with different artificial systematic bias (top) comparing the distributions of $\Delta Z_{\max}$ as a function of the target signal strength $Z_{\max}^{\text{PLR}}$ (bottom).	176
7.20	Mean μ and variance σ of the $\Delta Z_{\max}$ distributions when evaluating over testing datasets with different signal widths.	178
7.21	Examples of predictions over histograms with the same background, either an (a) analytical function or a (b) fit to Dark Machines background, but with different signal widths, namely 1-bin width (left), 2-bin width (center) and 3-bin width (right).	178
7.22	Number of histograms as a function of the predicted $Z_{\max}$ and the predicted mass for the raw Dark Machines background-only datasets and the BSM signal datasets with an RPV stop or leptoquarks (LQ).	180
7.23	Number of histograms as a function of the predicted $Z_{\max}$ and the predicted mass for the Z' signal dataset and the raw Dark Machines channel 2b background-only dataset.	181
7.24	Examples of predictions over the Dark Machines datasets containing RPV stop particles with mass of 1 TeV from channel 3 and channel 2b.	182
7.25	Examples of predictions over the Dark Machines datasets containing leptoquarks with mass of 600 GeV decaying via $LQ \rightarrow b\ell$, $\ell = e, \mu$ from channel 2b.	182
7.26	Examples of predictions over the Dark Machines datasets containing Z' with mass of 50 GeV and decaying via $Z' \rightarrow 2\mu$ from channel 2b.	183

7.27 Preliminary invariant mass distributions obtained from the ATLAS simulation data. The different processes are scaled to match their cross-section and match the recorded luminosity	184
--	-----

List of acronyms

Acronym	Description
ADAM	adaptive moment estimation
AI	artificial intelligence
AUC	area under the curve
BSM	beyond the Standard Model
CKM	Cabibbo–Kobayashi–Maskawa
CNN	convolutional neural network
CP	charge-conjugation–parity
CSC	cathode strip chambers
DAQ	data acquisition
EM	electromagnetic
FCN	fully connected network
FPR	false positive rate
HF	heavy-flavour
HGTD	high-granularity timing detector
HL-LHC	High-Luminosity LHC
HLT	high-level trigger
HLV	high-level variables
ITk	inner tracker
KDE	kernel-density estimation
L1	Level-1 trigger
LAr	Liquid Argon
LEE	look-elsewhere effect
LEP	Large Electron-Positron
LF	light-flavour
LHC	Large Hadron Collider
Linac4	Linear accelerator 4
LLH	likelihood-based identification algorithm
LSP	lightest supersymmetric particle
MC	Monte Carlo

Acronym	Description
MDT	monitored drift tubes
Micromegas	micro-mesh gaseous structure
MLE	maximum likelihood estimator
MSE	mean squared error
MSSM	minimal supersymmetric Standard Model
NNLO	next-to-next-to-leading order
NSW	new small wheel
OS	opposite sign
p.d.f.	probability density function
PDF	parton distribution function
PLR	profile likelihood ratio
PS	Proton Synchrotron
PSB	Proton Synchrotron Booster
QCD	quantum chromodynamics
QED	quantum electrodynamics
ROC	receiver operating characteristic
RPC	resistive plate chambers
RPV	R-parity violating
SCT	semiconductor tracker
SF	scale factor
SLAC	Stanford Linear Accelerator Center
SM	Standard Model of particle physics
SPS	Super Proton Synchrotron
SR	signal region
SS	same sign
sTGC	small-strip TGC
SUSY	supersymmetry
TDAQ	trigger and data acquisition
TGC	thin gap chambers
TPR	true positive rate
TRT	transition radiation tracker
WIMP	weakly interacting massive particle

Acknowledgements

It has been a privilege to work on the projects featured in this thesis, a journey that has made me evolve in so many ways and I will always cherish it dearly. Of course, none of this would be possible without the help of many people along the way.

First and foremost, I would like to thank my supervisor, Jean-François Arguin. Your passion for this research field is contagious, and your knowledge always inspired me to push further. More importantly, your kindness and dedication reminded me that this is really a community endeavour and that everyone is welcome. You made me a better person and a better scientist, and for that I am truly grateful.

I would also like to acknowledge the financial support of the « Institut de valorisation des données (IVADO) » and the « Fonds d'excellence en recherche Apogée Canada » from the Government of Canada. Thank you for funding this research and creating a vibrant multidisciplinary community around data science and artificial intelligence.

Next, I would like to thank David London, Alan Robinson and Alain Bellerive for accepting to be part of this jury. I extend my gratitude to the whole UdeM particle physics group, a community that is passionate and kind beyond measure. Special thanks also to Sophie Tremblay, Diane Choinière, Thierry Nakache and the entire administration of the UdeM Physics department, for giving so much support to us throughout our studies.

Being a part of the ATLAS UdeM group has been an incredible journey. In particular, I would like to thank the people with which I had the pleasure to work with in the last few years: Fannie Bilodeau, Joséphine Potdevin, Émile Baril, Muhammad Usman, Hoang Dai Nghia Nguyen, Olivier Denis, Dominique Godin, Ali El Moussaouy, Ethan James Meszaros, Georges Azuelos and Claude Leroy.

I am incredibly grateful to be part of the ATLAS Collaboration, so my thanks to everyone that contributes in some way or another to make this experiment. I have learned a lot from the many people I met throughout these projects, and I am deeply grateful to them. In particular, I would like to thank the developers of the `TAGANDPROBE` software, for creating a fruitful environment of shared knowledge and growth that has helped me improve in so many ways. I extend my gratitude to the people working within the ATLAS e/γ group, which are consistently pushing to deliver their best. Special thanks to the colleagues with which I had the pleasure to share our successes and our frustrations, all of that with great kindness and good humour: Lorenzo Varriale, Joshua Puddefoot, Varsha Senthilkumar, Laura Pintucci, Léo Boudet, Susan Bataju, Veena Balakrishnan, Filip Nechanský, Stefan Richter, Ximo Poveda, Ju-

lien Maurer, Asma HadeF, Grigore Tarna, Ruggero Turra, Martina Laura Ojeda and Ludovica Aperio Bella.

I am also deeply grateful for the incredible team working on the BumpNet analysis, whose vision and dedication has inspired me greatly. In particular, special thanks to Shikma Bressler, Amit Shkuri, Etienne Dreyer, Nilotpai Kakati, Davide Melini, Michael Kwok Lam Chu, Maryna Borysova, Elad Kliger, Eva Mayer, Samuel Calvet and Julien Noce Donini. I extend my gratitude to the members of the Exotics group, for the great discussions and insights.

Moreover, I would like to take this opportunity to thank the free software community, whose contributions to the development of software suites that are accessible and respectful of our privacy are deeply appreciated. My work was greatly optimised via the use of GNU/Linux, Emacs, Org, Git, LaTeX and many other projects, all of which are the culmination of the work of many people throughout the years.

Furthermore, if I can pursue my graduate studies today and work as a scientist, it was because many women before me dared to do so in a society that had other plans for them. Thus, I thank all the women who fought against the tide and paved the way so that I could also contribute to the scientific pursuit today.

Eu não estaria aqui hoje se não fosse pelos esforços e dedicação da minha família. Obrigada por acreditarem em mim e me darem os meios de trilhar essa jornada. O seu amor e incentivo me deram a confiança necessária para seguir esse caminho e por isso serei eternamente grata. Em especial, minha profunda gratidão ao meu parceiro de vida, Igor. Obrigada por me acompanhar nessa jornada e me inspirar a ser a melhor versão de mim.

Statement of personal contributions

The ATLAS experiment is run by an international collaboration of scientists, technicians and engineers, with thousands of people contributing to different aspects of the project. For this reason, its measurements are published in the name of the whole collaboration, thus recognising the contributions of the technical and research teams located both at CERN and at the universities and laboratories worldwide. In what follows, I describe my original contribution to the projects presented in this thesis.

During my studies, I have contributed to three distinct projects: (i) the implementation and measurement of electron identification efficiencies (Chapter 5), (ii) the development of an electron identification algorithm using machine learning (Chapter 6) and (iii) the development of a neural network to identify resonances in invariant mass histograms (Chapter 7). Moreover, I also performed tracking software shifts, where I checked if there were any discrepancies on the data distributions following updates to the code. Throughout all these projects, I collaborated with different teams, always helping and being helped by my colleagues.

The first project was done in the context of my qualification as an ATLAS collaboration author. In preparation for its third data-taking period, referred to as Run 3, the ATLAS software suite had a major update when going from version 21 to version 22. As part of this update, the software framework responsible for performing the electron efficiency measurements was completely rewritten, this new implementation being called the `TAGANDPROBE` software. In this context, I was responsible for the implementation of the Z -mass method, used for the measurement of the electron efficiencies, being one of the lead developers in the initial phase of the project. This contribution also included the design and implementation of the core tag-and-probe method, the core efficiency and scale factor calculation infrastructure, as well as the user interface for producing performance plots. Given its complexity, the implementation of the Z -mass method happened in multiple stages. My contribution consisted of the implementation of the first iteration of the method into the new software, referred in this work as the “normalisation in the mass tail” method, as well as its validation against its previous implementation. Once the authorship qualification finished, I continued my involvement in this project, contributing to the software development and its documentation, as well as providing support to the other developers and users. Following these contributions, I was appointed as reconstruction and software sub-group convener for the electron-photon (e/γ) group, starting in October 2024.

Once the first implementation of the Z -mass method was finished, I performed the

first measurements of the electron identification efficiencies in Run 3. These results used the data recorded in 2022 and were featured in plots released to the public^[1]. Meanwhile, the Z -mass method was upgraded to its new version by our colleagues, using a fit instead of the normalisation in the mass tail approach. Using this new version of the method, I also performed the measurement of the electron identification efficiencies over the full Run 2 period, using the data that was reprocessed with the ATLAS software in release 22. Although these results are not final, they correspond to an important milestone in the assessment of the electron identification performance within the new ATLAS software.

My expertise in the study of electron signatures in the detector was further established via my contribution to the development of a deep learning algorithm to perform electron identification. This neural network incorporated low-level detector information using a convolutional neural network which treated the energy deposits and additional tracks information obtained from the detector as images. My first contribution was the study of the impact of the input features on the network's performance, referred to as a feature ranking study. An ATLAS public note summarising the developments of this algorithm using simulation data was published during this period^[2]. Later, I also contributed in the efforts to study the performance of the network over experimental data when compared to the use of examples from simulation, being responsible for training and evaluating the model to assess its performance.

In addition to these contributions to measure and improve the detector performance, I was also involved in the development of a novel technique to perform the analysis of the experimental data in search for new particles. In this project, we developed a neural network to identify resonances in invariant mass histograms, the ultimate goal being its use to detect such anomalies over the ATLAS experimental data. Being one of its early members, I was involved in all the different aspects of the analysis, from the software implementation to its performance studies. In particular, my main contribution involved the use of the publicly available Dark Machines simulation data to test the performance of the method over realistic simulation data. Moreover, I was responsible for implementing the software responsible for the generation of synthetic data, the training and the evaluation of the neural network, as well as assessing the performance of the network in different scenarios. Towards the end of my studies, I also contributed with the first implementation of the production of mass histograms from ATLAS simulation samples, in preparation for the use of this method in the ATLAS data. Furthermore, in this project, I served as co-analysis contact within the ATLAS collaboration, our team consisting of 15–20 members. In this role, I was responsible for the organisation of meetings, writing the documentation and reporting the status of the project to the collaboration.

*“Every judgement teeters on the brink of error”, Leto explained.
“To claim absolute knowledge is to become monstrous.
Knowledge is an unending adventure at the edge of uncertainty.”*

Frank Herbert
Children of Dune (1976)

Introduction

The description of nature at its most fundamental level is the subject of the field of particle physics, which focus on the constituents of matter and their interactions. Throughout the years, scientists have unravelled the mysteries of these phenomena by developing new experimental techniques to probe processes involving subatomic particles and formulating their description in terms of quantum fields. Its culmination is the *Standard Model of particle physics* (SM), a theoretical framework that has been shown to be very successful in describing how the fundamental particles interact via the strong, weak and electromagnetic forces.

Nevertheless, there is reason to believe the SM is incomplete. Firstly, a large number of *ad hoc* parameters are present in the model and, although some patterns are observed in the values, there are no underlying concepts to describe them. Moreover, if we assume the SM is supposed to be valid until the energy scales where gravitational effects become significant, a very precise cancellation of terms is needed to keep the Higgs boson at the low energies it has been observed. The improbability of such cancellation seems to indicate that there are phenomena beyond the SM that must be accounted for. Also, the SM does not explain a series of observations that we expect a theory of fundamental particles to describe, such as the nature of dark matter and dark energy and the violation of the charge-conjugation–parity (CP) symmetry responsible for the asymmetry between matter and antimatter in the universe. These are just a few examples of the questions that are still open in the field of particle physics and that motivate the work presented in this thesis.

That said, the study of high-energy processes has played a central role in helping describe the phenomena at the subatomic scales. To that end, particle colliders have provided valuable insights to constrain our models via the assessment of different observables from the experimental data. The largest such collider, the Large Hadron Collider (LHC), is located at the CERN⁽¹⁾ laboratory near Geneva, where multiple experiments are performed. Among them, the ATLAS experiment is one of the protagonists of this field of research, performing measurements of the products of hadronic collisions with its general-purpose detector. By registering the signatures of particles produced after the collisions, the processes that originated from them are reconstructed, providing large amounts of experimental data for inference.

In this context, the ATLAS physics programme is dedicated to precise measurements that can constrain the SM predictions, as well as the search for signs of phenomena that the current model cannot explain. Some challenges involved in these tasks include, among

(1) European Organisation for Nuclear Research, acronym from *Conseil Européen pour la Recherche Nucléaire*.

others, the proper reconstruction and identification of particles, as well as the detection of anomalies in the experimental data. The work presented in this thesis concerns both of these challenges, focusing on the detection of electrons and the search for anomalies over invariant mass distributions.

As of the writing of this thesis, ATLAS is going on its third data-taking period, referred to as Run 3, while it prepares for an upgrade in light of the next-generation accelerator, the High-Luminosity LHC (HL-LHC), which is planned to be operational in the next decade. Both these scenarios bring notable challenges, such as the increase in the number of collisions happening simultaneously at the centre of the detector, entailing a busier environment, as well as the analysis of the large amount of experimental data recorded. The benefits of having such vast experimental data at these high energies come with challenges that require the development of new techniques to fully extract physical insights from the collected data. To that end, the field of machine learning offers powerful techniques to process this data and extract relevant features. Having used many such techniques for quite some time now, the ATLAS experiment offers a great opportunity for the application and development of new machine learning methods.

In particular, some of the contemporary challenges of the ATLAS experiment require the development of algorithms capable of handling the high-dimensionality of observables measured by the detector. Combined with their non-linear correlations, the use of deep learning algorithms offers an optimal solution for tasks such as classification or anomaly detection. By exploring the large amount of experimental data already recorded, these techniques have the potential to improve the detector performance in more complex scenarios, as well as to provide novel solutions to perform statistical inference from the data. The application of deep learning methods to address challenges in both the identification of particles and the detection of anomalies in the data is a central subject of this thesis.

Before developing tools to improve the algorithms used in the detector, their performance must be properly assessed. For this reason, the first original contribution presented in this work involves the measurement of the performance of the ATLAS detector to identify electrons. These particles are involved in many of the central SM processes of interest, being produced from the decay of the Z , W and Higgs bosons as well as the top quark. As a result, the high performance of the electron identification algorithm is crucial to make precise inferences and extract valid conclusions from the data. To quantify this performance, new measurements are presented for the efficiency of the electron identification algorithm currently used in the ATLAS experiment, covering the periods of the Run 2 (2015 – 2018 and early Run 3 (2022).

Furthermore, a new algorithm is being developed to perform electron identification in ATLAS using neural networks. Given the impact of the misidentification of objects that mimic the electron signatures in the precision of ATLAS measurements involving electrons, the goal of this new algorithm is to improve the rejection of such objects for a fixed signal efficiency. To do so, this method incorporates low-level detector information into neural networks that

are capable of extracting its most relevant features. Being developed using simulation data, the evaluation of the algorithm performance over real data is an important validation of its suitability to process real experimental data. The original studies presented in this work assess the relative importance of the inputs used by this network, as well as its performance when applied to real data and its comparison to the results obtained with simulation data.

Lastly, the search for new phenomena is a central theme in the field of particle physics, and this work presents a novel approach to perform this task. One of its major limitations is the large amount of resources needed to estimate the theoretical predictions made by the SM, which considerably delay the data analysis workflow. To tackle this, a new search strategy is implemented, which uses a neural network to search for resonances in invariant mass distributions, envisioning its future application over the ATLAS experimental data. Such a network would be able to identify and guide future analyses towards the more interesting regions of the data. In this work, the method is described together with its implementation over realistic data obtained from simulation. Efforts in preparation for its application in ATLAS are also included, namely the production of the first invariant mass histograms produced from ATLAS simulation data.

In what follows, we start with a general introduction of the SM in Chapter 1, including an overview of its main concepts and the observed limitations that indicate the existence of phenomena beyond its scope. Chapter 2 is dedicated to the concepts of statistics and machine learning that are used throughout this thesis. The LHC and the ATLAS detector are described in Chapter 3, with a dedicated overview of the electron reconstruction and identification process being discussed in Chapter 4. The original contributions associated to this thesis start being presented in Chapter 5, where the methods used to measure electron identification efficiencies and the corresponding measurements performed for Run 2 and Run 3 are presented. Next, Chapter 6 is dedicated to the description of the new electron identification algorithm that incorporates low-level detector information using a convolutional neural network. Lastly, the new technique to detect resonances in invariant mass distributions using a neural network is presented in Chapter 7. The thesis concludes with a summary of the work presented and some final remarks.

Chapter 1

The Standard Model of particle physics

We do not know what the rules of the game are; all we are allowed to do is to watch the playing. Of course, if we watch long enough, we may eventually catch on to a few of the rules. The rules of the game are what we mean by fundamental physics.

Richard Feynman^[3]

1.1	Fundamental particles	6
1.1.1	Fermions	7
1.1.2	Bosons	7
1.1.3	Hadrons	8
1.2	Quantum field theory and symmetries	9
1.3	Fermionic fields	11
1.4	Gauge fields	12
1.5	Electroweak sector and Higgs mechanism	13
1.6	Quantum chromodynamics	16
1.7	Limitations of the Standard Model and open questions	18
1.8	Beyond the Standard Model	21

From all the experimental evidence gathered over the years, many theories were developed and tested to attempt a description of what happens at the microscopic level. Today, the *Standard Model of particle physics* (SM)^[4–11] is the theoretical framework that describes all the known fundamental particles and their interactions. One of its greatest achievements was the prediction of the Higgs boson, discovered at the Large Hadron Collider (LHC) in 2012^[12–15]. Since then, many more precise tests have been made of the model's predictions^[16,17]. And although no large deviations have been found, there is reason to believe the SM is not complete. In this thesis, we will present contributions to the experimental research on SM particles and the search for anomalous phenomena that might be beyond our current theoretical framework. Hence, this chapter introduces the basic elements of the SM, including a summary of its greatest challenges and future prospects.

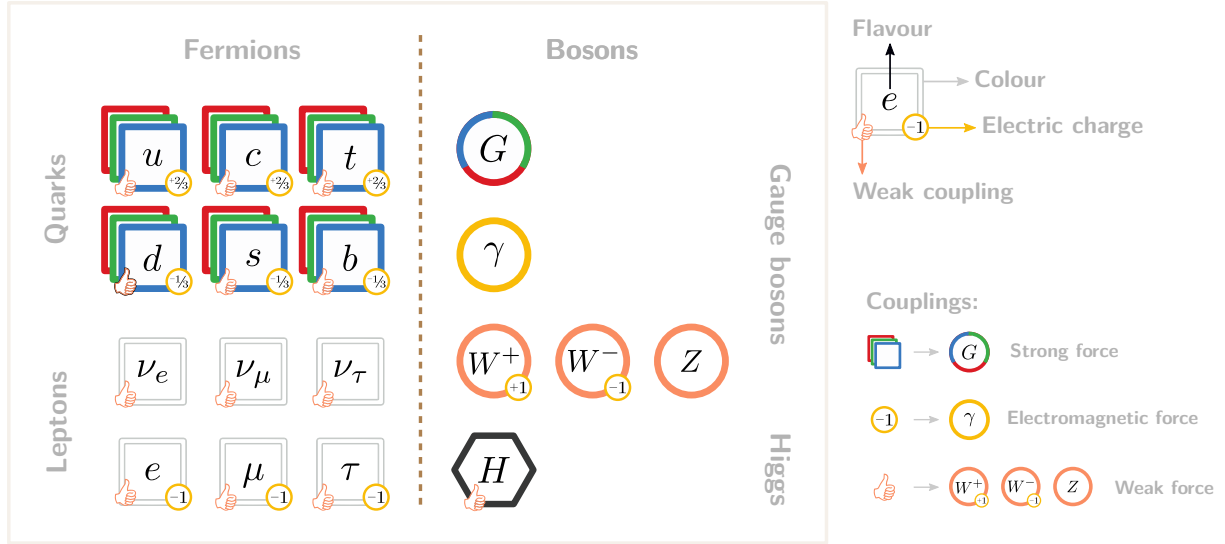


Figure 1.1: Fundamental particles and their interactions within the standard model of particle physics.

1.1 Fundamental particles

Particle physics is the scientific domain dedicated to the study of the constituents of matter and how they interact among themselves. Starting from the discovery of molecules^[18,19], the quest for more fundamental particles lead the community to investigate further within the atom. Starting with the discovery of the electron^[20], the atomic structure was then understood to be formed by a nucleus surrounded by electronic orbitals^[21–23]. In due time, our understanding of the nucleus as a composite particle lead us to also discover protons^[24] and neutrons^[25] in the early 20th century.

In the 1950s and 1960s, a multitude of new particles were discovered and catalogued, leaving physicists with many disconnected pieces and trying to make sense of it all. By the 1960s, the description of the strong force by Gell-Mann and Zweig postulated that hadrons could be in fact made of constituent particles, now referred to as quarks^[26,27]. The observation of such particles was finally confirmed in 1969 at the Stanford Linear Accelerator Center (SLAC)^[28,29]. At the same time, theoretical developments to unify the electromagnetic and weak force postulated the existence of the $W^\pm$ and Z bosons^[4–8]. After the first direct evidence of a weak neutral current was observed by the Gargamelle bubble chamber in 1973^[30,31], the Z and $W^\pm$ bosons were finally observed by the UA1 and UA2 collaborations in 1983 at a proton-antiproton collider at CERN^[32,33].

All of these discoveries were only possible due to the technological advances of the experimental systems developed to measure these particles. From taking pictures of particles in bubble chambers^[34], to particle accelerators^[35–37] and colliders^[38], today we are capable of reaching incredible energy levels and continue the pursuit of understanding matter at the fundamental level.

Our current understanding of the fundamental particles within the SM is illustrated in

Fig. 1.1. They are classified into *fermions*, spin-half particles, and *bosons*, spin-integer particles. In the first group, we find *quarks* and *leptons*, while in the second we have the *gauge bosons*, responsible for the electromagnetic, strong and weak interactions, as well as the *Higgs boson*, involved in the process that gives particles their mass.

A wide variety of *composite particles* have also been observed, corresponding to bound states of quarks. Referred to as *hadrons*, these particles are also classified as fermions or bosons depending on the number of quarks involved; In the first category we find the *mesons*, a combination of a quark and an antiquark, while in the latter we have the *baryons*, a combination of three quarks.

In what follows, we describe in more details our current understanding of these fundamental and composite particles.

1.1.1 Fermions

Quarks and leptons are the fundamental fermions of the SM, particles with spin $+1/2$ which are distinguished by their *flavour*. They form a hierarchy based on their mass in ascending order such that we identify three generations or families. In each one of these generations, we find a quark with electromagnetic charge $+2/3$, namely the up (u), charm (c) and top (t) quarks, and another quark with electromagnetic charge $-1/3$, namely the down (d), strange (s) and bottom (b) quarks. Similarly, each family has a lepton with electric charge -1 , respectively the electron (e), the muon (μ) and the tau (τ), as well as a neutrino with no electric charge, namely the electron neutrino (ν_e), the muon neutrino (ν_μ) and the tau neutrino (ν_τ).

Other than the electromagnetic charge, each fermion has an intrinsic chirality that is defined by another type of charge called its *weak isospin* (I). This charge can take either a value of $+1/2$, corresponding to a left-handed particle, or a value of 0 , such that we call it a right-handed particle. Other than that, quarks can also interact via the strong force, and thus take one of three values of *colour* charge, conventionally referred to as red, green and blue, while leptons have no colour charge. For each one of these particles, there exists another particle with the exact same mass and spin, but opposite charges. We refer to those as *antiparticles*.

Moreover, the interactions through the electromagnetic, strong and weak forces are governed according to their corresponding values of the electromagnetic, colour or weak isospin charges, respectively. This means that only particles with non-zero charge values interact through the corresponding force. These forces are mediated by other particles, namely gauge bosons, discussed in what follows.

1.1.2 Bosons

The gauge bosons are the force carriers, meaning interactions between particles happen via their exchange. Namely, the electromagnetic interaction happens through the exchange of photons γ , the strong interaction via the exchange of gluons G , and finally the weak interaction

happens through the exchange of the Z , W^+ and W^- bosons. All the gauge bosons have spin 1, and, although some of them are massless, such as the photon and the gluons, the Z , W^+ and W^- bosons have mass.

Distinct from the gauge bosons, we find the Higgs boson. This spin-0 particle is involved in the breaking of the symmetry that unifies the electromagnetic and weak forces at high energies, which will be discussed later in Section 1.5. Its field and non-zero vacuum expectation value has the important role of being responsible for giving mass to other massive particles. Among them, we find all of the ordinary matter with which we interact, in the form of composite states of quarks, which are discussed in what follows.

1.1.3 Hadrons

Due to the nature of the strong interaction, which will be discussed in Section 1.6, quarks can only be found in bound states^[39]. Hence, these states form composite particles called hadrons which have been observed to be formed by either a quark–antiquark pair, referred to as a meson, or by three quarks or three anti-quarks, referred to as baryons. More complex states of four^[40,41] and five quarks^[42] have also been observed.

A variety of mesons have been observed to date, the list going up to more than 140^[43]. The lightest ones, the π mesons, are formed by the coupling of the u and d quarks, and were also the first ones to be observed^{[44](2)}. Other notable mentions include the mesons D , K and B , that include the c , s and b quarks, respectively.

Similarly, many different baryons have been observed in the second half of the 20th century, but the great majority of them are not stable and decay rapidly. The only stable baryons are the proton and the neutron, with the exception that the latter also decays through the weak interaction in specific cases^[39].

Since the quarks within a bound state interact through the strong force by exchanging gluons, the content of the baryons can be quite dynamic. This information is described via parton distribution functions (PDFs), which describe the probability of finding a quark of a given flavour and momenta within the composite particle^[39]. As an example, the PDF for the proton are shown in Fig. 1.2, where we can see that at low energies, the valence quarks (two u quarks and one d quark) take the majority of the proton’s momentum. Other contributions come from the so-called “sea quarks” which arise from quark/anti-quark pair production from gluons within the proton.

By involving composite particles with high mass, which lose less energy to synchrotron radiation, hadron–hadron collisions can achieve high centre-of-mass energies, which allow the study of phenomena at high energy scales^[39]. These colliders play a central role in improving our understanding of the fundamental particles and their interactions. Examples of

(2) We take this opportunity to highlight the involvement of the Brazilian physicist César Lattes in this discovery. He was the main researcher and the first author of the article describing the π meson. Due to the awarding policies at the time, the 1950 Nobel Prize was solely awarded to Cecil Powell, the research head group, instead.

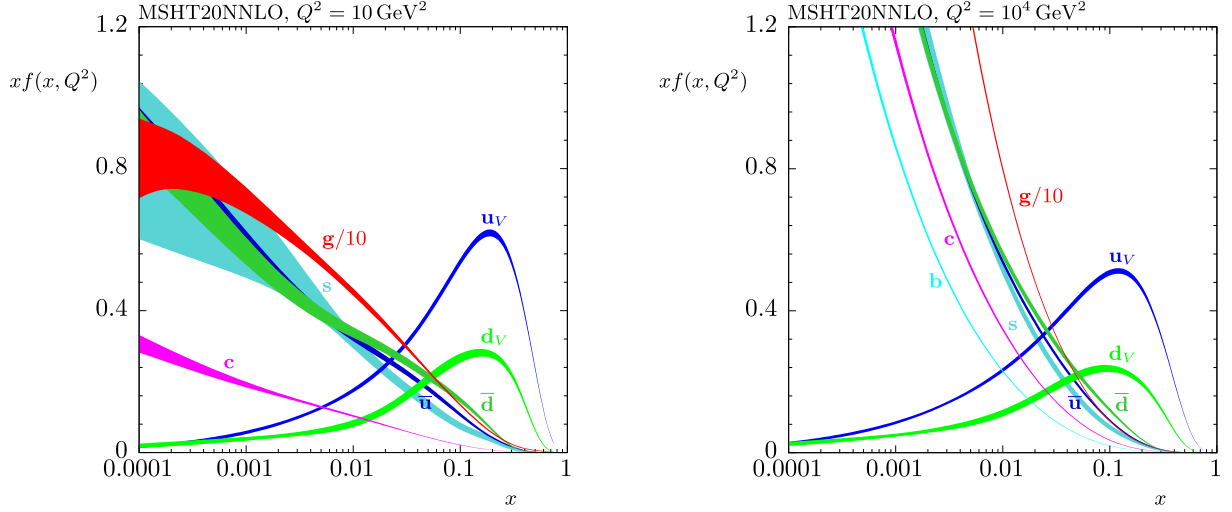


Figure 1.2: Parton distribution functions for the proton as a function of the Bjorken- x , defined as the fraction of the proton momentum carried by the parton. They are shown for energy scales of $\mu^2 = 10 \text{ GeV}^2$ (left) and $\mu^2 = 10^4 \text{ GeV}^2$ (right), and were obtained in next-to-next-to-leading order (NNLO) by the MSHT20 global analysis of available hard scattering data. Taken from Ref. [45].

their protagonism include their use in the discoveries of the W and Z bosons, as well as the discovery of the top quark and the Higgs boson. Presently, the LHC is the largest such collider, with detectors such as the ATLAS experiment, described in Chapter 3, being at the forefront of experimental research in the field. Beyond hadron colliders, many other experimental technologies have been used throughout the years and still push the scientific frontiers, such as e^+e^- colliders^[38], but also detectors dedicated to the measurement of dark matter^[46–48] and the neutrino observatories^[49–55].

Concurrently with the development of experimental techniques, new concepts and insights from the more theoretical aspects of particle physics also played a central role to the establishment of our current understanding of nature. In particular, the culmination of our knowledge in the form of the SM was greatly achieved due to the concepts of quantum fields and their symmetries. They form the basis of the theoretical framework of the SM and are briefly described in the following section.

1.2 Quantum field theory and symmetries

In the SM, particles are described as local excitations of a corresponding *quantum field* $\varphi(x)$ in space-time x . This framework allows for the description of the interactions within systems of multiple particles and the transitions between different states at the very small scale and with very large energies, while not violating neither quantum mechanics nor special relativity^[56].

These fields follow the *principle of least action*, such that their equations of motion follow

from minimising the *action* (S) of the system. This quantity is defined by

$$S = \int \mathcal{L}(\varphi, \partial_\mu \varphi) d^4x, \quad (1.1)$$

where $\mathcal{L}$ is the Lagrangian density (usually simply referred to as Lagrangian), a Lorentz invariant that is written in terms of the fields $\varphi(x)$ and its derivatives in space-time^[56].

It follows that the dynamic of the system satisfies the Euler–Lagrange equation^[56],

$$\partial_\mu \left[\frac{\partial \mathcal{L}}{\partial (\partial_\mu \varphi)} \right] - \frac{\partial \mathcal{L}}{\partial \varphi} = 0. \quad (1.2)$$

Thus, the definition of a Lagrangian for the SM encodes all the dynamics of the fundamental particles.

From Noether’s theorem^[57], we know that any transformation that leaves the Lagrangian density invariant entails a conserved quantity. Thus, it can be shown that a *global* phase transformation of a particle field, meaning its phase is changed by a constant value,

$$\varphi(x) \rightarrow \varphi'(x) = \varphi(x) \exp(iq\chi),$$

is associated with the conservation of the corresponding charge q ^[58].

Now, the existence and form of the interactions between the fundamental particles also follow from these phase transformations of the particle fields, what is referred to as the *gauge principle*^[58]. In particular, by requiring the Lagrangian density to be invariant under *local* phase transformations of the particle fields, where the phase now has a dependence on x , that is,

$$\varphi(x) \rightarrow \varphi'(x) = \varphi(x) \exp[iq\chi(x)],$$

implies the existence of a *gauge field* which couples to the associated charge q ^[39].

Thus, if a particle carries a given charge, the interaction associated with that charge follows by requiring the Lagrangian density to be invariant under a given local phase transformation. As a result, these transformations are referred to as *symmetries*. Furthermore, since the specification of a given value to this phase has no effect on the outcome of any experiment, we can specify any definite value to the gauge fields, a process called *gauging*. Since the theory originates from the existence of these degrees of freedom, we call it a *gauge theory*^[58].

In this context, the SM is a gauge theory with the symmetry group

$$\text{SU}(3)_C \times \text{SU}(2)_L \times \text{U}(1)_Y,$$

where C , L and Y correspond to charges, respectively the colour, the third component of the weak isospin I_3 (or “left-handness”) and the weak hypercharge (Y). It corresponds to the

combination of $SU(3)_C$, the symmetry group from *quantum chromodynamics* (QCD)^[9–11], and $SU(2)_L \times U(1)_Y$, the symmetry group from *electroweak theory*^[4–8]. After the breaking of the electroweak symmetry, the $U(1)_{Q_{em}}$ symmetry remains, describing *quantum electrodynamics* (QED)^[59–61].

In the following sections, we discuss in more details the characteristics of the particle fields and the theories that describe their interactions.

1.3 Fermionic fields

The quarks and leptons of the SM are described by fields called *spinors* $\psi(x)$ ^[62]. They can be decomposed into two separate fields, ψ_L and ψ_R , called *Weyl spinors*, via projection operators $P_{L,R}$, such that

$$\psi_{L,R} \equiv P_{L,R}\psi, \quad \text{where} \quad P_{L,R} \equiv \frac{(1 \pm \gamma^5)}{2}$$

and $\gamma^5 \equiv i\gamma^0\gamma^1\gamma^2\gamma^3$ is the *chirality matrix*. The latter is defined in terms of the *Dirac matrices* $\gamma^{\mu=0,1,2,3}$, which satisfy the relation $\{\gamma^{\mu\nu}\gamma^{\nu\mu}\} = 2\eta^{\mu\nu}\mathcal{I}_4$, $\eta^{\mu\nu}$ being the Minkowski metric^[63]. We refer to these Weyl spinors as the left-handed (L) and right-handed (R) components of the particle fields. Moreover, they are solutions of the Dirac equation⁽³⁾,

$$(i\hbar\gamma^\mu\partial_\mu - mc)\psi = 0,$$

where m is the particle mass, γ^μ are the Dirac matrices and $\partial_\mu \equiv \partial/\partial x^\mu$. This equation follows from the Dirac Lagrangian, given by

$$\mathcal{L}_D = \bar{\psi}(i\hbar\gamma^\mu\partial_\mu - mc)\psi, \tag{1.3}$$

where $\bar{\psi} \equiv \psi^\dagger\gamma^0$ is called the Hermitian conjugate field^[56]. This Lagrangian describes the fermion systems and, as a consequence, must be invariant under transformations of the SM symmetry group. However, we note that this is not the case for the mass term,

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

As a result, fermions are described by massless fields and gain their mass from another mechanism.

Furthermore, the fermion fields correspond to representations of the group in the form of *multiplets*. In particular, since the quarks possess colour charge, associated to the $SU(3)_C$ group, they correspond to triplets under the transformations of this group. On the other hand,

(3) We note that, in this thesis, we follow the particle physics convention of using *natural units*, where $\hbar = c = 1$ and GeV is the unit of energy, momentum and mass^[39].

	Fields	Colours	I_3	Y	Q_{em}
Quarks	u_L, c_L, t_L	r, g, b	+1/2	+1/6	+2/3
	d_L, s_L, b_L	r, g, b	-1/2	+1/6	-1/3
	u_R, c_R, t_R	r, g, b	0	+2/3	+2/3
	d_R, s_R, b_R	r, g, b	0	-1/3	-1/3
Leptons	$\nu_{eL}, \nu_{\mu L}, \nu_{\tau L}$	0	+1/2	-1	0
	e_L, μ_L, τ_L	0	-1/2	-1	-1
	e_R, μ_R, τ_R	0	0	-1	-1

Table 1.1: Quantum numbers of the fermion fields under the SM symmetry group. They are the colour, the third component of the weak isospin I_3 and the weak hypercharge Y , associated with the $SU(3)_C$, $SU(2)_L$ and $U(1)_Y$ symmetry groups, respectively. The electric charge $Q_{\text{em}} = I_3 + Y$, associated with the electromagnetic interaction after the electroweak symmetry breaking, is also included.

the left-handed fermion fields correspond to doublets under the $SU(2)_L$ symmetry, with the corresponding charge taking the values $I_3 = \pm 1/2$, while the right-handed fields are singlets with $I_3 = 0$.

In summary, we denote the *quark fields* as

$$Q_L^{i=1,2,3} \equiv \begin{pmatrix} u \\ d \end{pmatrix}_L, \begin{pmatrix} c \\ s \end{pmatrix}_L, \begin{pmatrix} t \\ b \end{pmatrix}_L; \quad u_R^{i=1,2,3} \equiv u_R, c_R, t_R; \quad d_R^{i=1,2,3} \equiv d_R, s_R, b_R;$$

and the *lepton fields* as

$$L_L^{i=1,2,3} \equiv \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L; \quad e_R^{i=1,2,3} \equiv e_R, \mu_R, \tau_R,$$

with their corresponding quantum numbers being presented in Table 1.1. Fields that have a non-zero charge interact via the corresponding force by the exchange of their mediators, the gauge bosons. These are also described with quantum fields in the SM and are presented in the following section.

1.4 Gauge fields

The interactions between fermion fields enter the SM Lagrangian via the definition of the *covariant derivative* $\mathcal{D}_\mu$ which replaces ∂_μ in Eq. (1.3) and is given by

$$\mathcal{D}_\mu \equiv \partial_\mu - ig_s G_\mu^k \frac{\lambda_k}{2} - ig W_\mu^\alpha \frac{\sigma_\alpha}{2} - ig' B_\mu \frac{Y}{2}, \quad (1.4)$$

where the Einstein summation notation is used^[58]. It defines the *gauge fields* $G_\mu^{k=1,\dots,8}$, $W_\mu^{\alpha=1,2,3}$ and B_μ and their couplings g_s , g and g' . Also, in this expression we find the generators of the $SU(3)_C$ and $SU(2)_L$ symmetry groups, $\lambda_k/2$ and $\sigma_\alpha/2$, respectively. Moreover, it is conventional to choose as representations the eight Gell-Mann matrices for the λ_k and the three Pauli matrices for the σ_α ^[58].

Now, we note that the number of gauge fields associated with each interaction depends on the number of generators of the corresponding symmetry group. For this reason, the strong interaction is mediated by eight gluon fields $G^{k=1,2,\dots,8}$ coupling to the colour charge of the group $SU(3)_C$, while the electroweak interaction has three boson fields $W^{\alpha=1,2,3}$ coupling to the third component of the weak isospin I_3 under the $SU(2)_L$ group. Lastly, we have a single boson B coupling to the weak hypercharge Y associated to $U(1)_Y$.

Similarly to the fermion fields, all these gauge fields are massless. Although one can immediately recognise G^k as the gluon fields from QCD, the bosons of the unified electroweak theory do not correspond immediately to the photon γ and the W^+ , W^- and Z bosons responsible for the electromagnetic and weak interaction, respectively. Their relations with these fields are explained by the breaking of the electroweak symmetry, the subject of the next section.

1.5 Electroweak sector and Higgs mechanism

Let us now take a closer look at the symmetries that define the SM. From one side, we have the $SU(3)_C$ symmetry group that describes the interactions via the strong force. Concurrently, we find the $SU(2)_L \times U(1)_Y$ symmetry group, that corresponds to the electroweak interactions. In the context of the gauge theory, the Lagrangian describing the physical states must be invariant under transformations of this symmetry group.

However, we note that the vacuum ground state is not symmetric under transformations of the SM symmetry group. This is a consequence of the spontaneous creation of particle–antiparticle pairs. In fact, the only observed symmetries of the vacuum are the Lorentz invariance with translational and rotational symmetries in space-time, together with neutral electromagnetic and strong interactions^[58]. We can see that in the fact that if a particle–antiparticle pair is created in the vacuum, they will have zero net electromagnetic charge and zero net colour.

As a result, the SM gauge symmetry $SU(3)_c \times SU(2)_L \times U(1)_Y$ must be reduced to a $SU(3)_c \times U(1)_{Q_{em}}$ symmetry on the physical vacuum, corresponding to the combination of the strong and electromagnetic interactions, described by QCD and QED, respectively. This process is explained by the *Brout–Englert–Higgs mechanism*^{[12–15](4)} via the introduction of a

(4) The history of this mechanism involves many antecedents, especially from condensed-matter physics in the theories of superfluidity and superconductivity^[64–66]. As J. Ellis wrote, “The wheel has turned full circle with their [superconductors] importance for the construction of the LHC.”^[67].

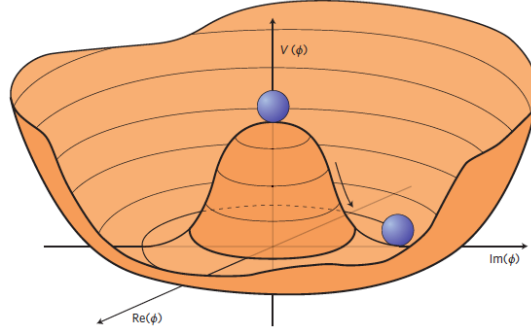


Figure 1.3: Illustration of the Higgs potential, Eq. (1.5), when $\mu^2 < 0$. Taken from Ref. [67].

complex scalar field, in particular a doublet in weak isospin with weak hypercharge $Y = 1$, given by

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix},$$

where ϕ^+ and ϕ^0 are the complex and charged neutral components, respectively^[43]. By having non-zero quantum numbers under the $SU(2)_L \times U(1)_Y$ group, this field interacts with the gauge bosons and fermions as they propagate. Moreover, the Lagrangian of this field is written

$$\mathcal{L}_H = (\mathcal{D}_\mu \phi)^\dagger (\mathcal{D}^\mu \phi) + V(\phi), \quad \text{with} \quad V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2, \quad (1.5)$$

where $\mu \in \mathbb{C}$ and $\lambda \in \mathbb{R}$ are constants and $V(\phi)$ is the most general and renormalisable Higgs potential^[43,68], which has the same form as the Ginzburg–Landau potential from the theory of superconductivity^[64–66].

If $\mu^2 < 0$, this potential takes the form illustrated in Fig. 1.3, with the minima located at^[39]

$$\phi^\dagger \phi = -\frac{1}{2} \frac{\mu^2}{\lambda} = \frac{1}{2} v^2,$$

which allows for infinite states with the minimal energy. By choosing the physical vacuum ground state to be such that

$$\langle \phi \rangle \equiv \langle 0 | \phi | 0 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix},$$

the electroweak symmetry is spontaneously broken, meaning that the interactions are still governed by the gauge principle, but this dynamic selection of the vacuum has reduced the symmetries of the vacuum^[58]. We call v the *vacuum expectation value*, a non-zero real value, that sets the energy scale in which the electroweak interaction is unified.

In this gauge, the scalar field can be expanded around the minimum by writing

$$\phi(x) = \exp[i\vec{\varepsilon}(x) \cdot \vec{\tau}] \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}.$$

Here, $\vec{\varepsilon}(x)$ is a triplet of massless scalar fields, while H is a massive scalar field called the *Higgs field*^[58].

The excitation of the Higgs field corresponds to a free-particle state, the Higgs boson, with mass given by $m_H = \sqrt{2\lambda} v$ ^[14]. Its observation was the last piece missing to validate this mechanism in the SM. The measurement performed by the ATLAS and CMS collaborations at the LHC in 2012 using the $pp \rightarrow H \rightarrow \gamma\gamma$ and $pp \rightarrow H \rightarrow ZZ^* \rightarrow 4\ell$ ($\ell = e, \mu$) channels^[69,70] consolidated the great success of this formalism. Today, the Higgs boson mass is measured to be^[43]

$$m_H = 125.25 \pm 0.17 \text{ GeV},$$

which fixes the vacuum expectation value at $v = 246.22 \text{ GeV}$ ^[43].

The three degrees of freedom from the $\vec{\varepsilon}(x)$ field correspond to *Goldstone bosons*^[68], that are absorbed by the longitudinal polarisation of the three massive gauge bosons W^+ , W^- and Z ^[39]. Together with the photon field A , these physical fields are written in terms of the massless gauge fields $W_\mu^{1,2,3}$ and B_μ as^[58]

$$\begin{aligned} W_\mu^+ &= \frac{1}{\sqrt{2}} (W_\mu^1 - iW_\mu^2), & \text{with } m_W &= \frac{gv}{2}; \\ W_\mu^- &= \frac{1}{\sqrt{2}} (W_\mu^1 + iW_\mu^2), & \text{with } m_W &= \frac{gv}{2}; \\ Z_\mu &= W_\mu^3 \cos(\theta_W) - B_\mu \sin(\theta_W), & \text{with } m_Z &= \frac{m_W}{\cos(\theta_W)}; \\ A_\mu &= W_\mu^3 \sin(\theta_W) + B_\mu \cos(\theta_W), & \text{with } m_\gamma &= 0, \end{aligned}$$

where θ_W is the weak mixing angle defined by^[56]

$$\cos(\theta_W) = \frac{g}{\sqrt{g^2 + g'^2}} \quad \text{and} \quad \sin(\theta_W) = \frac{g'}{\sqrt{g^2 + g'^2}}.$$

Moreover, the interaction of the Higgs doublet ϕ to the fermionic fields is written in the form of the Lagrangian terms

$$\mathcal{L}_f^{\text{quarks}} = -g_f \left[\bar{Q}_L \phi d_R^i + \bar{d}_R^i \phi^\dagger Q_L \right] + g_f \left[\bar{Q}_L \phi^C u_R^i + \bar{u}_R^i \phi^{C\dagger} Q_L \right]$$

and

$$\mathcal{L}_f^{\text{leptons}} = -g_f \left[\bar{L}_L \phi e_R^i + \bar{e}_R^i \phi^\dagger L_L \right],$$

where g_f is the Yukawa coupling constant of the given fermion and ϕ^C is the Higgs doublet conjugate field^[39]. After the electroweak symmetry breaking, these Lagrangian terms become

$$\mathcal{L}_f = -m_f \bar{\psi}\psi - \frac{m_f}{v} \bar{\psi}\psi H, \quad \text{with} \quad m_f = \frac{v}{\sqrt{2}} g_f.$$

As a result, we see that the fermion masses arise from their interaction with the Higgs field. Note, however, that the coupling values are not determined by the theory, they need to be chosen to match the observed fermion masses^[39,43].

Lastly, the interaction of quarks with the $W^\pm$ bosons via charged-current weak interactions is written as

$$-\frac{g}{\sqrt{2}} (\bar{u}_L, \bar{c}_L, \bar{t}_L) \gamma^\mu W_\mu^+ V_{\text{CKM}} \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix} + \text{h.c.},$$

where V_{CKM} is the *Cabibbo–Kobayashi–Maskawa* (CKM) matrix^[71,72]. This unitary matrix can be parametrised by three mixing angles and a phase^[43], the latter being responsible for phenomena that violates the conservation of the charge-conjugation–parity (CP) symmetry in flavour changing processes^[73].

The unification of the electroweak interaction at energies above the vacuum expectation value at 246 GeV and the process with each its symmetry is broken is known as the *Glashow–Salam–Weinberg model*^[4–8]. Its combination with QCD, the subject of the following section, completes the SM.

1.6 Quantum chromodynamics

While the electroweak theory describes both the electromagnetic and weak interactions observed in nature, the remaining force described within the SM is the strong interaction. This fundamental force was first postulated in order to explain how protons and neutrons are bounded within the atomic nucleus. The idea that they were formed in fact by quarks^[26,27] led to the development of a gauge field theory describing their interaction, called quantum chromodynamics or QCD^[9–11].

In this framework, the strong interaction is described by the symmetry group $\text{SU}(3)_C$, where the colour charge takes the values conventionally referred to as red, green and blue. This choice reflects the neutrality of states with three quarks of different colours and mesons with colour–anticolour pairs^[58]. The gluon fields are the massless intermediates of the interaction and correspond to the eight generators of the group, conventionally taken to be represented by the Gell-Mann matrices, as previously mentioned. Carrying a colour and anticolour charge themselves, gluons can self-interact via three-point and four-point vertices^[58].

In particular, this self-interaction of gluons is thought to be responsible for the *colour con-*

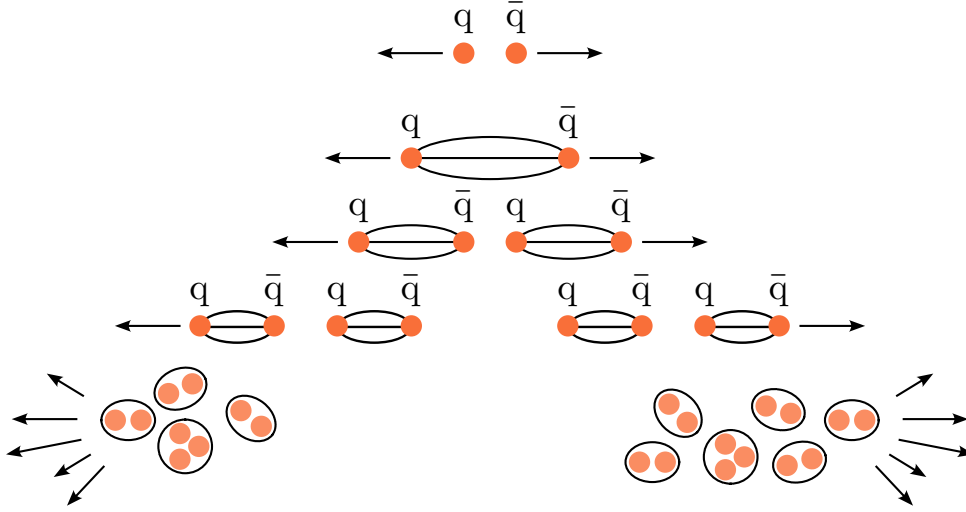


Figure 1.4: Illustration of the hadronisation process. When a quark and an antiquark are separated, the energy stored in the colour field is enough to produce new quark–antiquark pairs. This process continues until all quarks are bound into colourless states. Adapted from Ref. [39].

finement, that is, the observation that only zero-colour objects can propagate as free particles^[39]. Since the energy stored in the colour field increases linearly with the distance, particles with colour are always pulled toward each other and thus are arranged within hadrons that have zero net colour charge⁽⁵⁾. For the same reason, free quarks and gluons cannot propagate to macroscopic distances^[39], decaying into a cascade of colourless mesons and baryons, usually referred to as *jets*. The process in which a quark becomes part of a meson or baryon is called *hadronisation*^[43] and is illustrated in Fig. 1.4.

Moreover, we note that the strong coupling constant g_s , introduced in Eq. (1.4), is the only fundamental parameter of QCD^[43]. Its strength varies according to the momentum transfers, such that $\alpha_s \equiv g_s^2/4\pi \approx 0.1$ in the 0.1–1 TeV range and becomes weaker for higher energies or short enough distances^[43]. This behaviour is referred to as *asymptotic freedom*^[10,11]. Given the small values of its coupling constant at high energies, perturbation theory can be used to perform calculations and thus QCD predictions can be tested experimentally in modern colliders.

In particular, high-energy pp collisions have extensively been used in recent years to test QCD. One such example is the differential cross-sections of the production of two jets, or simply *dijet* system, the most common outcome of pp collisions. An illustration of such a final state registered by the ATLAS experiment is illustrated in Fig. 1.5. The predicted differential cross-sections for this process from QCD agrees with the measurements made at the LHC by both the ATLAS^[75] and CMS^[76] experiments. Many other tests of QCD have been performed at the LHC^[17], with an overall excellent agreement between its predictions and experimental measurements, thus consolidating its success.

(5) We note that in the same way that electromagnetic charges -1 and $+1$ add to zero, the colours (r, g, b) and the anti-colours (anti-r, anti-g and anti-b) also “add” to zero

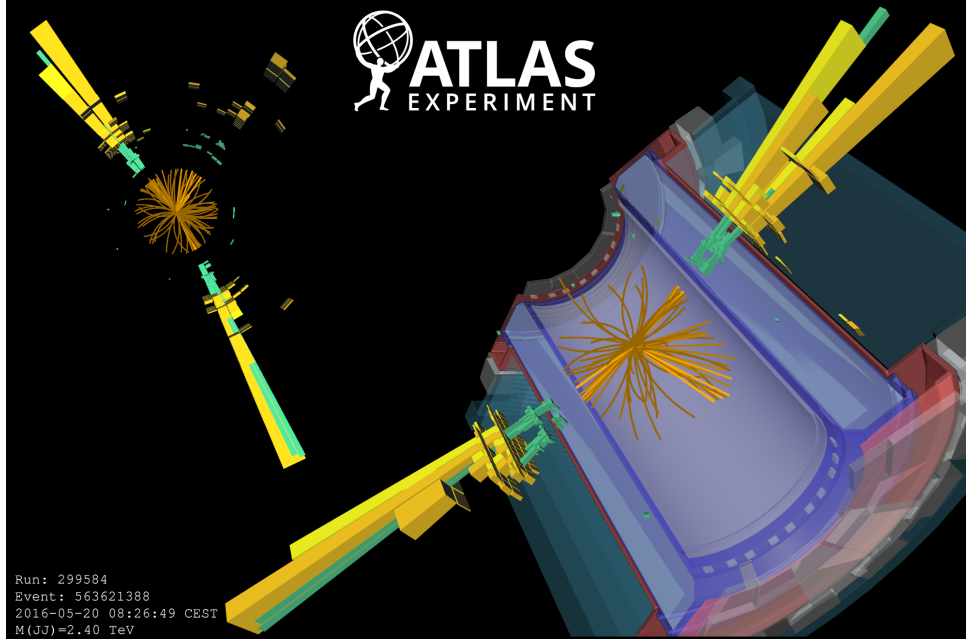


Figure 1.5: Event display of the production of two jets (dijet) registered by the ATLAS experiment, showing the tracks recorded at the inner detector (orange lines) and the energy deposits in the calorimeters (green and yellow blocks), which will be described in Chapter 3. Taken from Ref. [74].

Despite its success described in this chapter so far, the SM is far from being a complete description of the nature. In the following section, we present an overview of its limitations and why we have reasons to believe our understanding of the fundamental structure of nature is far from complete.

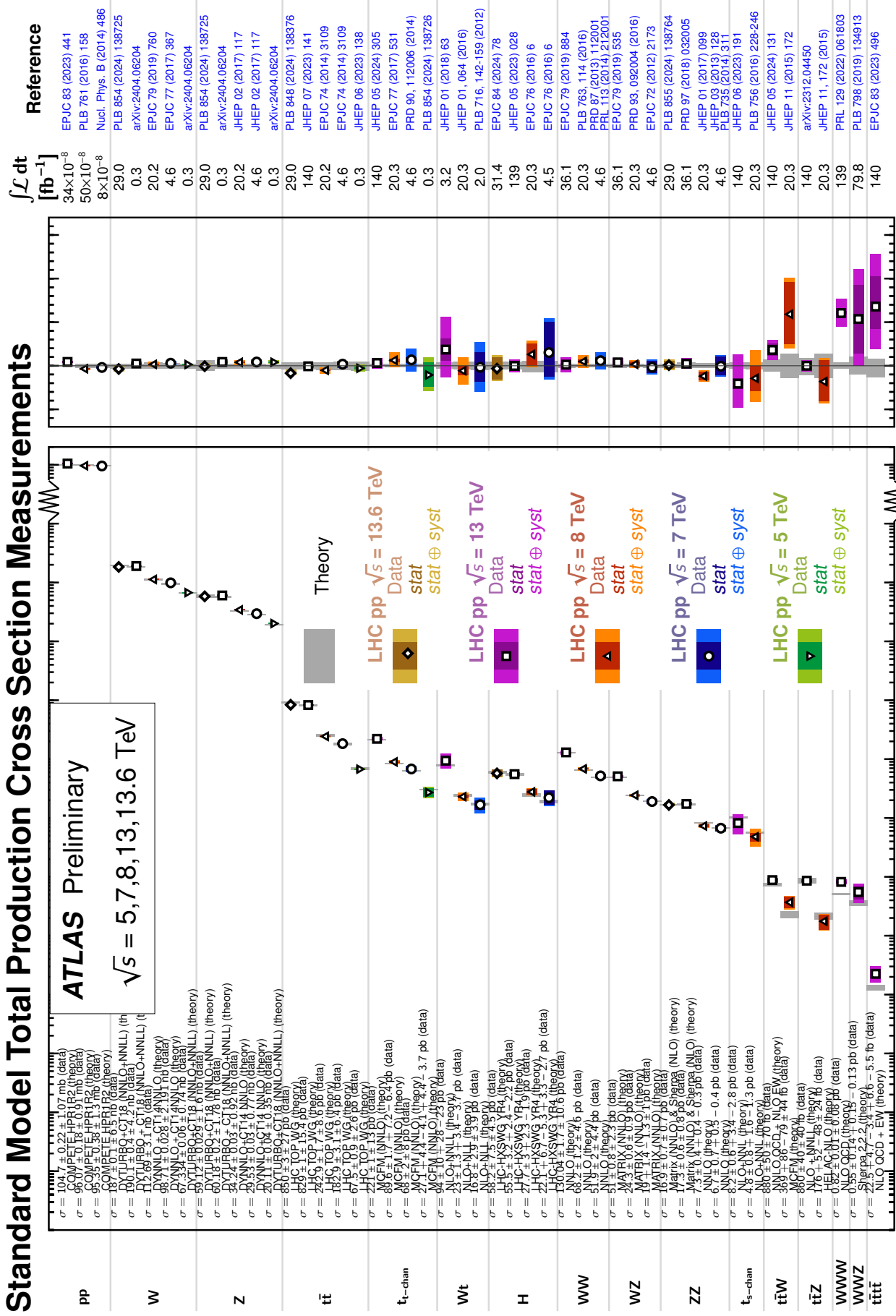
1.7 Limitations of the Standard Model and open questions

The SM has successfully withstood experimental scrutiny, its predictions agreeing with the experimental measurement in many different fronts. As shown in Fig. 1.6, for all the main processes observed so far at the LHC, the production cross-section measurements performed by the ATLAS experiment agree with the corresponding theory prediction within its uncertainty. Nevertheless, from arbitrary theoretical characteristics to unexplained experimental observations, there are many questions that the SM cannot answer.

Firstly, in its current configuration, the SM has a total of 18 free parameters that must be specified in an *ad hoc* manner. Among them, we have the 9 fermion couplings to the Higgs field that are tuned to their observed masses, the 3 coupling constants setting the strength of the gauge interactions, the 2 parameters in the Higgs potential and the 3 angles and 1 phase of the CKM matrix^[39]. There is an additional parameter, the strong CP phase $\theta_{\text{QCD}} \approx 0$, which would allow for CP violation in strong interactions. Since such violation has never been observed, θ_{QCD} is assumed to be zero^[43]. All these free parameters are measured by experiments^[78].

We note that some patterns emerge within these parameters, hinting at the existence

Standard Model Total Production Cross Section Measurements



Status: June 2024

Figure 1.6: Summary of the total production cross-section measurements performed by the ATLAS experiment, compared to the theoretical predictions from the Standard Model. Taken from Ref. [77].

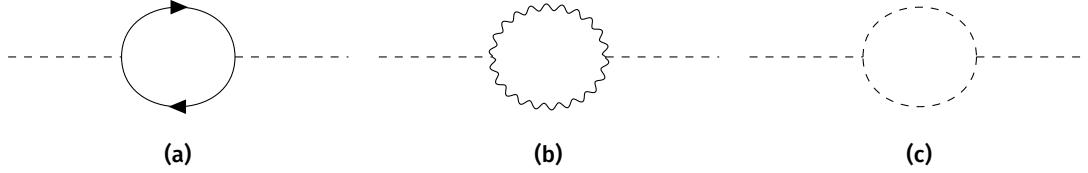


Figure 1.7: Loop-level corrections to the Higgs boson mass from (a) fermions, (b) gauge bosons and (c) other Higgs bosons. All these loop corrections must cancel out in order to keep the Higgs mass value around the measured 125 GeV.

of underlying principles behind them. For instance, the three coupling constants having the same order of magnitude could indicate the forces are unified at higher energies^[39]. One could argue that these are just numbers and the universe is as it is. But it is worth remembering that the atomic numbers also seemed arbitrary until we discovered they actually corresponded to the number of protons within the atomic nucleus. Similarly, such a large number of *ad hoc* parameters might be an indication that there is something more fundamental to be investigated.

Beyond the arbitrary nature of the SM parameters, there is also no explanation for the fine-tuning of the Higgs boson mass, known as the *hierarchy problem*^[79–81]. Its theoretical calculation has to account for loop contributions, such as the ones shown in Figure 1.7⁽⁶⁾, which depend quadratically on the energy scale up to which the SM is valid, that is, Λ_{SM} ^[83]. We could suppose this scale to be higher than the electroweak scale, at 246 GeV (i.e., the value of the vacuum expectation value) in which we have observed the SM is valid so far. For instance, it could be that the SM is in fact valid for energies up to the order of the Planck scale at $\Lambda \sim 10^{19}$ GeV, when gravitational effects become important. In this case, these loop contributions would need to cancel out in order to keep the Higgs mass around the observed value at the electroweak scale^[39]. The improbability of such a precise cancellation leads us to believe that there might be missing particles or phenomena of which we are still unaware^[84].

Furthermore, experimental observations of neutrinos oscillating between flavours indicate that neutrinos have mass^[85–87]. As previously discussed, the SM neutrinos have no mass and, as a result, the model needs to be extended to incorporate massive neutrinos. More importantly, this extension can be done in different ways, since there is no experimental evidence that neutrinos satisfy the Dirac equation, being also possible that they satisfy the Majorana equation. The determination of the neutrino nature to be either Dirac or Majorana particles is one of the open experimental questions of the field^[43,62].

Other experimental anomalies pertaining *B*-meson decays also raise suspicion that the model is not complete^[88,89]. In particular, tensions between theory and experiment have been observed in processes involving the transitions $b \rightarrow s\ell^+\ell^-$ and $b \rightarrow c\ell\nu$ ^[90]. Among these

(6) All Feynman diagrams in this thesis were made with the help of the package from Ref. [82].

disagreements, we have notably the angular observable P'_5 of the 4-body decay

$$B \rightarrow K^*(\rightarrow K\pi)\mu^+\mu^-,$$

which has small but persistent deviations^[91], together with the ratios R_D and R_{D^*} which disagree with the observed value by $\sim 3.2\sigma$ ^[92]. In some cases, the new physics scenario is preferred over the SM by 5.8σ ^[93], a discrepancy that grants further investigation. Lastly, another conflict arises in the measurement of the anomalous magnetic moment of the muon^[94–96]. Its theoretical prediction is one of the most precise ever obtained from the SM^[96]. A deviation with respect to the experimental measurement is observed only at the order of 10^{-9} , corresponding to $\sim 3.5\sigma$ ^[97].

There are, of course, all the observed phenomena that are unexplained or completely left out of the SM, but which we would expect to be included in a theory of fundamental particles. For instance, the observed CP-violation that entails an asymmetry between matter and antimatter, with a clear deficit of antimatter in the universe^[43,98], but from our current understanding should have been produced in equal quantities.

Furthermore, gravity is not accounted for in the model^[99], and although models postulating extra dimensions seem a promising solution for unifying all forces, no evidence for them has been found so far^[100–103]. Also, the SM has no description of dark matter and dark energy^[104–106]. In contemporary estimates, these account for $\approx 26.8\%$ and $\approx 68.3\%$, respectively, of the energy content of the universe^[107]. Since we know the Higgs is involved in the process that gives mass to the SM particles, it would make sense that it is also involved in the mechanism that gives mass to dark matter. Many studies are currently pursuing the search for dark matter, but nothing has been observed so far^[46–48].

With all these questions combined, one might wonder what lies beyond our current understanding of the fundamental particles and their interactions. Many interesting ideas have been proposed to address all these issues. We will briefly mention some of them in the following section.

1.8 Beyond the Standard Model

Although the SM is very successful in describing the experimental observations involving fundamental particles, many questions remain unanswered, some of which have been discussed in the previous section. To address these questions, different ideas and interesting hypothesis have been developed to modify or expand the SM framework, usually referred to as beyond the Standard Model (BSM). In this section, we present an overview of some of these theoretical models, focusing on the ones that will be revisited later on in this thesis.

The most popular extension of the SM adds the concept of supersymmetry (SUSY)^[108–113]. An idea that gained traction during the 1980s, it corresponds to a new space-time symmetry

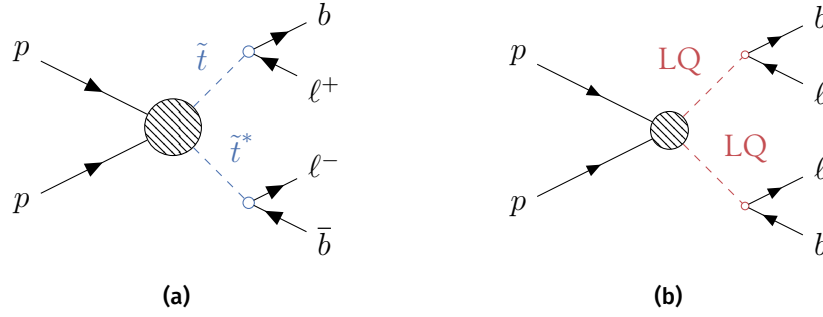


Figure 1.8: Diagrams for the (a) RPV SUSY stop pair production and (b) scalar leptoquark pair production at the LHC. Coloured lines indicate the hypothetical particles. Being both scalar massive particles, each of them decays into a charged lepton and a b -quark.

that transforms fermions into bosons, and vice-versa^[43]. Similarly to the multiplets of the electroweak theory, described in Section 1.5, there could exist *supermultiplets* formed by both fermion and boson states that have the same electric charges, weak isospin and colour degrees of freedom^[114]. Particles within the same supermultiplet are referred to as *superpartners*, and different possibilities are available on how these supermultiplets could be constructed.

In this context, extensions of the SM have been developed to incorporate this idea, one such model being the *minimal supersymmetric Standard Model* (MSSM)^[115–118]. In this model, each particle in the SM would have a superpartner with spin differing by 1/2 unit^[114]. For quarks and leptons, for instance, there would exist spin-0 particles, which are usually referred to by adding an “s”⁽⁷⁾ in front of its name and a “~” on top of its symbol (e.g. the superpartner of the top quark t would be the *stop* $\tilde{t}$). Similarly, the name of the spin-1/2 superpartner of a boson in the SM is constructed by adding the suffix “-ino” to the name of its SM counterpart, such that the superpartner of the gluon g , for instance, would be the *gluino* $\tilde{g}$ ^[114]. However, the only way to conciliate this idea with these missing particles that were never observed would be to have SUSY as a broken symmetry⁽⁸⁾.

The reason this theory is so attractive is its straightforward solution to the hierarchy problem, since the loop contributions from SM particles would be cancelled by those involving their superpartners^[119–122]. For this to work, the superpartners masses have to lie at the TeV scale^[123], which is within the reach of modern colliders. Moreover, it provides a framework for the unification of particle physics with gravity at the Planck energy scale^[114,117], with some of its variations even introducing a weakly interacting massive particle (WIMP), which is a candidate for explaining dark matter^[124–127].

As beautiful as the theory is, no experimental evidence has been found to corroborate any SUSY model as suitable to describe nature^[43]. One of its challenges relates to the concept of R-parity, which implies SM particles are R-even particles and the superpartners are R-odd^[43,128]. The conservation of this R-parity has important consequences, namely that all superpartners

(7) This “s” stands for *scalar*, since they have spin 0.

(8) How the symmetry is broken and at which energy scales give rise to a wide variety of SUSY models^[43], which are beyond the scope of this thesis.

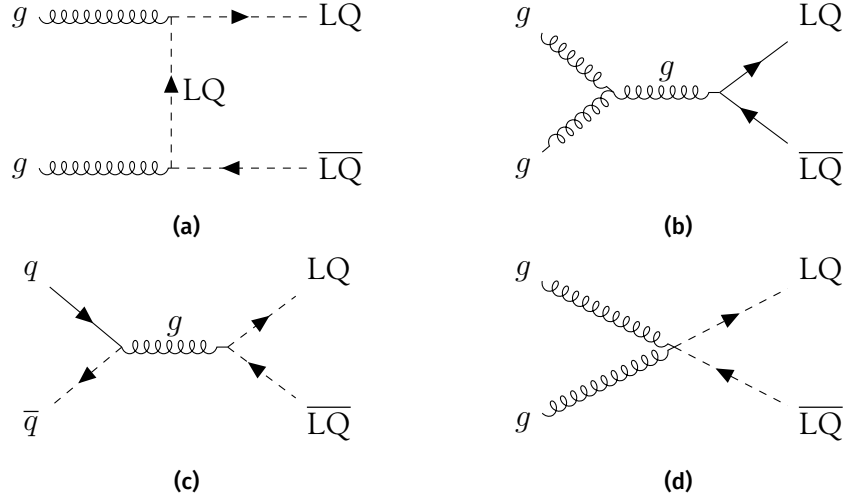


Figure 1.9: Leading order Feynman diagrams for leptoquark pair production at the LHC^[134].

eventually decay to the lightest supersymmetric particle (LSP) which is absolutely stable^[114]. To further complicate things, this LSP would have zero electric charge and no colour, and, thus, be undetectable by our current collider experiments^[127].

In this context, R-parity violating (RPV) SUSY^[123,129] models provide an alternative landscape where R-parity is not conserved, thus allowing many possibilities^[84]. Of particular interest is the stop particle, which could be the lightest squark of the model^[130–132]. If pair-produced at the LHC, it could decay into final states of b quarks and leptons^[132], as shown in Fig. 1.8a. Searches for this specific decay channel have been performed by the ATLAS experiment^[133], with mass limits being set around the 0.8–1.8 TeV range for different hypotheses. This is just one example of many searches that have been performed for particles predicted in SUSY models, which have taken a central stage within the LHC experiments in the last years. Although no experimental evidence has been found, these models will probably still be under the radar of high-energy experiments for quite some time.

But SUSY is not the only option available to expand the SM. As a matter of fact, one of the most common solutions for the semileptonic B -meson anomalies, for instance, is the introduction of *leptoquarks* (LQ). These hypothetical particles couple simultaneously to a lepton and a quark, with existing models covering the possibilities that it has either spin 0 (scalar LQ) or 1 (vector LQ). Moreover, having both colour and weak isospin charges, these particles could mediate flavour-changing neutral currents that could explain the previously discussed B -meson anomalies. Leptoquarks models also propose favourable solutions to the origin of the neutrino masses^[135–138] and the origins of CP-symmetry violation, which could explain the matter–antimatter asymmetry^[139,140].

At the LHC, the pair production of scalar LQ, for instance, is possible via gluon–gluon fusion or quark–antiquark annihilation^[134,141], as show in Fig. 1.9. Its decay into leptons or third generation quarks, shown in Fig. 1.8b, is similar to the RPV SUSY stop signature discussed above. Measurements performed by the ATLAS experiment found no significant deviations

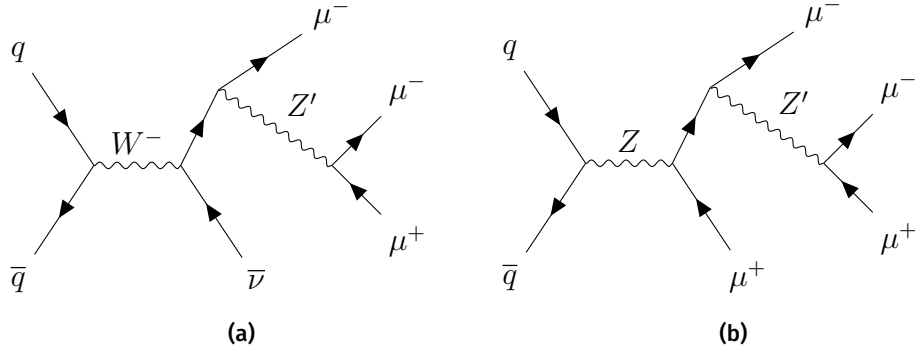


Figure 1.10: Leading order Feynman diagrams for Z' production in the (a) 3μ and (b) 4μ final states.

from the SM expectation^[142], however. These measurements exclude masses below 1.7 TeV for the scalar leptoquarks, a limit that has been corroborated by CMS as well^[143]. Nevertheless, given its versatility in addressing different problems within the SM, leptoquarks are also set to remain a focus of experimental searches at the LHC.

Lastly, another solution to the semileptonic B -meson anomalies would be a hypothetical vector boson labelled Z' . In one of its simplest variations, this particle is introduced within a lepton-flavour-violating $U(1)_{L_\mu-L_\tau}$ symmetry model, where no extra fermions are required to cancel gauge anomalies^[144,145]. Having no electric or colour charge, this spin-1 particle could have a relatively low mass. In this particular model, the Z' decays to leptons or neutrinos and, thus, could be produced and detected at the LHC. Some examples of its decay channels are shown in Fig. 1.10. In particular, the Z' could be detected from a Z boson decay via $Z \rightarrow Z' \mu \mu \rightarrow 4\mu$ (Fig. 1.10a), or from a W boson decay via the process $W \rightarrow Z' \mu \nu \rightarrow 3\mu + \nu$ (Fig. 1.10b)^[146]. Searches for such a particle with mass values below 70 GeV have been performed in the 4μ final state by both the ATLAS^[146] and CMS^[147] experiments, but no deviations were found.

All these models correspond to just a sample of all the possibilities being proposed by theoretical physicists and investigated at the LHC experiments. Among them, the ATLAS physics programme has been dedicated to the thorough search of deviations of the SM predictions, focusing, among others, in the final states of these proposed models^[148]. Having not found any deviation yet, new ways of using the detector data have emerged to identify hints of new phenomena. An example of a change in paradigm is the shift towards anomaly detection searches, where the focus is the scanning of the data and flagging of unexpected objects or statistical features, as opposed to the traditional search for a pre-established model.

Either way, these searches rely on the statistical analysis of experimental data, which is expected to increase in both quantity and complexity in the next years. To address these imminent challenges, more sophisticated tools such as machine learning can be incorporated to the task. That said, the next chapter is dedicated to an overview of the statistical methods and machine learning techniques used to interpret the experimental results and that will be employed in this thesis.

Chapter 2

Statistics and machine learning

All models are wrong, but some are useful.

George E. P. Box^[149,150]

2.1	Introduction	26
2.2	Fundamental concepts	27
2.3	Parameter estimation	29
2.3.1	Likelihood function	30
2.3.2	Maximum likelihood estimator	31
2.4	Hypothesis testing	32
2.4.1	Profile likelihood ratio test	34
2.5	Figures of merit and ROC curves	36
2.6	Machine learning	39
2.6.1	Motivation	39
2.6.2	Basic concepts	40
2.6.3	Fully-connected network (FCN)	42
2.6.4	Convolutional neural network (CNN)	45
2.7	Applications of machine learning on high-energy physics	46
2.7.1	Classification	48
2.7.2	Anomaly detection	48

Measurements performed in high-energy collider experiments, such as the LHC, rely on strong statistical inference to quantify the observables of interest. The ever-increasing amount of data requires a continuous effort to optimise its analysis and thus guarantee that its full potential is explored. In recent years, the rise of machine learning techniques have further propelled the advances in the domain by introducing novel approaches to manage larger quantities of data and tackle more complex tasks. In this chapter, we introduce many of the concepts and tools that will be used later on in this thesis. We start with basic statistical inference tools, followed by a closer look at the domain of machine learning and how these techniques are being applied to high-energy physics experiments⁽⁹⁾.

(9) Note that, unless otherwise stated, this chapter is based on Refs. [43, 151, 152].

2.1 Introduction

The principle of high-energy collider experiments, such as the LHC, is quite straightforward: we collide particles multiple times and register the products of these collisions through sophisticated detectors. Even though the initial states of the collision are always the same, the outcome will depend on the physics phenomena involved. As such, the predictions of our current model, the SM discussed in the previous chapter, can thus be tested when we compare them to the actual observations made.

In this context, it is no surprise that the domain of statistics is highly relevant to our field of study. After all, the goal of statistical methods is to make inferences about a given model from a huge amount of data, either to determine the validity of the model or estimate its parameters. As a result, the extraction of meaningful information from data is only possible via the application of these techniques for the relevant tasks. These include quantifying the accuracy of a measurement via its uncertainty, estimating a parameter of the model or testing if a hypothesis is corroborated by the observed phenomenon.

For instance, the measurement of an observable will vary both due to the resolution of the detector, but also due to a more fundamental probabilistic aspect of the system (e.g., its quantum mechanical nature). This randomness can be quantified by estimating the uncertainty of the measurement and it is a fundamental property of any statement one could make from the analysis of experimental data. Another central application of statistics in this context is the determination of our confidence when observing a phenomenon that is not predicted by the SM. More precisely, if we were to find new particles or measure a given observable to be far from its expected value, we need to guarantee that what we see is not just a consequence of the detector imperfection or due to the random nature of the observation.

Furthermore, the intersection of statistics and computer science in the field of machine learning has demonstrated great success in making inferences from large amounts of data. Its use in experimental particle physics is in rapid expansion, with applications ranging from the more technical interpretation of the detector signals to the full statistical analysis of its data^[153–156].

Being experiments of repetition by construction, high-energy colliders are the perfect target for the application of statistics methods. As a consequence, such methods will be heavily featured throughout this thesis, including techniques for hypothesis testing and the figures of merit needed to quantify the performance of a given method. In the following section, we start with a brief introduction of the fundamental concepts of statistics, upon which the methods used in this thesis will be built.

2.2 Fundamental concepts

A central concept when quantifying the degree of randomness of a process is *probability*. It is a real-valued function P that, when applied over subsets A and B of a sample space S , must follow certain axioms^[157], namely:

1. For every subset A in S , $P(A) \geq 0$;
2. If A and B are disjoint subsets ($A \cap B \neq 0$), $P(A \cup B) = P(A) + P(B)$;
3. $P(S) = 1$.

By defining the *conditional probability* of A given B as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad (2.1)$$

it follows from these axioms that

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad \text{with} \quad P(B) \neq 0, \quad (2.2)$$

known as the *Bayes' theorem*^[158], and

$$P(B) = \sum_i P(B|A_i)P(A_i), \quad (2.3)$$

such that A_i correspond to mutually exclusive events that cover all the sample space. This expression is known as the *law of total probability*.

Moreover, the assignment and interpretation of probability in data analysis follow two main approaches. In *frequentist* statistics, we interpret the probability as the limiting relative frequency of the outcome of a repeatable experiment. That is, $P(A)$ corresponds to the number of occurrences of outcome A over n measurements when $n \rightarrow \infty$. Conversely, *Bayesian* statistics has a more subjective interpretation of probability, treating it as a tool to quantify the degree of belief in a given hypothesis^[151]. Even with different interpretations of probability, both these approaches provide useful tools for experimental particle physics. It is important to highlight that neither approach is preferred over the other, with most applications using concepts from both of them.

Another important concept is that of *random variables*, which are numerical values assigned to each element of the sample space S . If we define x as the value of a single variable obtained from the outcome of an experiment⁽¹⁰⁾, the sample space S corresponds to all the values x can assume. The probability that the measurement outcome lies between x and $x + dx$ is given by $f(x; \theta) dx$, where $f(x)$ is referred to as the *probability density function* (p.d.f.), which is normalised such that the total probability is equal to one, $\int_S f(x) dx = 1$.

(10) Note that x denotes both the random variable and the values assumed by the variable.

Table 2.1: Examples of probability density functions.

Distribution	Probability density function (p.d.f.)	Mean (μ)	Variance (σ^2)
Binomial	$f(x; N, p) = \frac{N!}{x!(N-x)!} p^x (1-p)^{N-x}$ $x \in \mathbb{N}$	Np	$Np(1-p)$
Poisson	$f(x; \nu) = \frac{\nu^x}{x!} e^{-\nu}$ $x \in \mathbb{N}$ $\nu > 0$	ν	ν
Gaussian	$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2

Furthermore, we define the *expectation value* $E[x]$ of a random variable x as

$$E[x] = \int_{-\infty}^{+\infty} x f(x) dx \equiv \mu, \quad (2.4)$$

which is also referred to as the *mean* of x . We note that any function of a random variable is itself a random variable, such that $E[u(x)] = \int_{-\infty}^{\infty} u(x) f(x) dx$. Also, a particular expectation value of interest is the n th central moment of x , given by

$$E[(x - \mu)^n] = \int_{-\infty}^{\infty} (x - \mu)^n f(x) dx \equiv \mu_n. \quad (2.5)$$

The second central moment,

$$E[(x - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx \equiv V[x] \equiv \sigma^2, \quad (2.6)$$

is called the *variance* of x , and quantifies how widely x is spread around its mean value μ . We call σ its *standard deviation*.

Some notable p.d.f.s include the binomial, the Poisson and the Gaussian distributions. Their expressions, together with the corresponding mean and variances, are shown in Table 2.1. Starting with the binomial distribution, it describes the number of successes x obtained as the outcome of a binary trial performed N times, where the probability of a success is a constant value p . Likewise, the Poisson distribution corresponds to the limit where N becomes large and p becomes small. We note that x can be treated as a continuous variable if it is integrated over a range $\Delta x \gg 1$. Lastly, we have the Gaussian distribution, which describes a continuous variable in the range $-\infty < x < \infty$. Its importance follows from the *central limit theorem*, which states that the sum of n independent and continuous random variables with finite mean and variances approaches a Gaussian random variable in the limit $n \rightarrow \infty$. In other words, a large number of fluctuations will be distributed as a Gaussian even if the individual fluctuations are not.

Now that we have established some basic concepts, we can shift our focus to the uses of statistics in particle physics. One of the major tasks we want to perform is the comparison of the

SM predictions with experimental observations. In fact, when performing any measurement, we are assessing how our theory holds when confronted with the data, that is, the probability $P(\text{theory}|\text{data})$. From Bayes' theorem, we know that

$$P(\text{theory}|\text{data}) \propto P(\text{data}|\text{theory})P(\text{theory}),$$

where $P(\text{theory})$ is our prior belief on the theory and $P(\text{data}|\text{theory})$ is the probability of getting a given outcome in light of the theory, also called the *likelihood*. Thus, we say $P(\text{theory}|\text{data})$ is the *posterior probability* that our theory is correct, once it has been compared with the data.

Moreover, it is important to note that, in realistic applications, we might have high-dimensional datasets, with multiple random variables and parameters. These high-dimensional datasets often require more sophisticated tools, some of which we will discuss in Section 2.6. For simplicity, in the remainder of this chapter we will consider only the one-dimensional case where only one random variable is considered.

In the following sections, we will introduce specific uses of statistics to address the tasks explored in this thesis. These include parameter estimation and hypothesis testing.

2.3 Parameter estimation

We start considering a random variable x such that, after performing n observations using our experimental apparatus, we obtain a sample data $\vec{x} = (x_1, \dots, x_n)$. The measurements are independent and identically distributed (i.i.d.); that is, they all follow the same distribution and the probability of obtaining a value for a given measurement does not change the probability of obtaining another^[159]. This means that each component of $\vec{x}$ follows the same p.d.f. $f(x; \theta)$, where the parameter θ ⁽¹¹⁾ is a quantity we want to estimate from this data. For this purpose, we define the *estimator* $\hat{\theta}(\vec{x})$, which is a function of the data only. Since it depends on the random variable x , it is itself a random variable.

Moreover, we define the *bias* of the estimator as the difference between its expected value and its true value, $b = E[\hat{\theta}] - \theta_{\text{true}}$, while the mean squared error (MSE) is the sum of its variance with the bias squared, $\text{MSE} = E[(\hat{\theta} - \theta_{\text{true}})^2] = V[\hat{\theta}] - b^2$. The MSE is a measure of the quality of the estimator and can be interpreted as the sum of the statistical and systematic errors.

One useful concept needed to obtain the optimal estimator $\hat{\theta}$ is the likelihood function, which we introduce in what follows.

(11) For simplicity, we consider a single parameter θ , but the results are also valid for a set of parameters $\vec{\theta} = (\theta_1, \dots, \theta_m)$.

2.3.1 Likelihood function

Since the n measurements x_i are i.i.d., the probability of having $x_i \in [x_i, x_i + dx_i]$ for a given value of the parameter θ is $\prod_{i=1}^n f(x_i; \theta) dx_i$ ^[151]. In this context, since the dx_i do not depend on the parameters, we define the *likelihood function* $L(\theta)$ as

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta). \quad (2.7)$$

Now, the number of observations n is usually fixed, since the experimental measurement is over when the analysis takes place. However, if the probability of observing n events depends on the parameter θ – which is the case in most particle physics applications – the likelihood function needs to account for this dependence. If we suppose, for instance, that n follows a Poisson distribution with mean μ , meaning we observe n events when we expected μ events. In this case, we write

$$L(\theta) = \frac{\mu^n}{n!} e^{-\mu} \prod_{i=1}^n f(x_i; \theta), \quad (2.8)$$

which is referred to as the *extended likelihood function*. In plain terms, it corresponds to the likelihood function when the sample size n is also part of the result of the experiment^[43].

Additionally, we must note that, most of the time, our model is not capable of accurately describing the data even in the most optimal point of the parameters θ . In this case, the estimated parameter will have a systematic bias. To address that, it is common to consider additional parameters called *nuisance parameters*, denoted by ν , in which we are not particularly interested in analyses, but that must be included for the model to be able to accurately predict the data. In this context, one can reduce the impact of the nuisance parameters in the result, such that the parameter of interest θ can be properly estimated.

A first possibility is to have a control data $\vec{y}$ that constrain their values, with the measurement of $\vec{y}$ being statistically independent from $\vec{x}$ ^[43]. Another approach is to replace the nuisance parameters in the model with functions of the parameters of interest. Referred to as a profiling procedure, we take the nuisance parameters that maximise the likelihood function for an specific θ . The result is the *profile likelihood function*^[43,160], given by

$$L(\theta) = L(\theta, \hat{\nu}), \quad (2.9)$$

where $\hat{\nu} = \hat{\nu}(\theta)$ are the profiled values of the parameter ν ⁽¹²⁾.

Once we have constructed the likelihood function, we are ready to obtain the estimator $\hat{\theta}$ for our parameter θ , which we will describe in what follows.

(12) The choice of “ $\hat{\nu}$ ” is mainly to avoid confusion with the estimator for ν denoted $\hat{\nu}$.

2.3.2 Maximum likelihood estimator

In the method of *maximum likelihood*, the estimator $\hat{\theta}$ satisfies the equation

$$\frac{\partial L}{\partial \theta} = 0. \quad (2.10)$$

As a result, it corresponds to the value of θ for which the likelihood function L is an extreme. Since it usually corresponds to a maximum, it is referred to as the *maximum likelihood estimator* (MLE) for θ .

Now, the logarithm being monotonic, the maximum of the logarithm of a function coincides with the maximum of the function, with the advantage that the logarithm operation turns a product into a summation. For this reason, it is usually more convenient to use the logarithm of the likelihood function instead. Furthermore, the great availability of minimisation techniques in statistical packages makes it more convenient to write the estimator as

$$\hat{\theta} = \arg \min_{\theta} \{-\ln[L(\theta)]\}. \quad (2.11)$$

That is, we can estimate the parameter $\hat{\theta}$ by minimising the negative-log-likelihood function $-\ln[L(\theta)]$.

In our case, the data sample $\vec{x} = (x_i, \dots, x_n)$ is large, that is, the total number of measurements n is large. In this scenario, it is convenient to bin its values in a histogram with N bins, such that we have a vector of data $\vec{n} = (n_i, \dots, n_N)$. For each number of entries n_i in a given bin i , we have an expected value $E[n_i] \equiv \mu_i$ which is determined by our model as a function of the parameter θ . As a result, if we treat the number of entries in each bin as an independent Poisson random variable n with mean μ , the data is described as the product of Poisson probabilities

$$L(\mu) = \prod_{i=1}^N \frac{\mu_i^{n_i}}{n_i!} e^{-\mu_i}, \quad (2.12)$$

with n_i and $\mu_i \equiv E[n_i]$ corresponding to the observed and expected values for the number of entries in the bin i , respectively. It follows that, in the applications we will use, the estimator $\hat{\theta}$ is the value of the parameter that minimises the negative log-likelihood, which is given by

$$-\ln[L(\mu)] = -\sum_{i=1}^N \left[n_i \ln(\mu_i) + \ln\left(\frac{1}{n_i!}\right) - \mu_i \right]. \quad (2.13)$$

Beyond estimating the parameters of a model, an important task performed in particle physics is that of assessing the validity of a given statement regarding our understanding of the data distribution. In the following section, we discuss how this can be done via hypothesis tests.

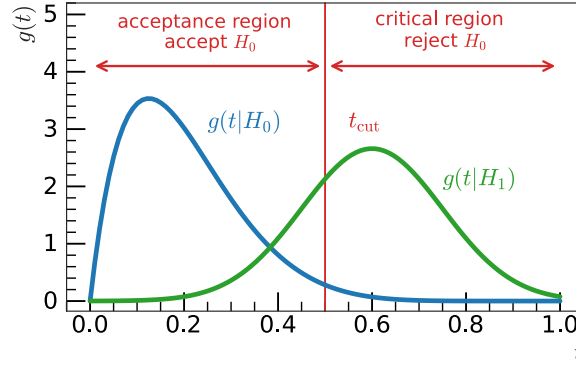


Figure 2.1: Probability density functions of the test statistic t assuming each one of the hypothesis H_0 and H_1 . We say H_0 is rejected if t is observed to be in the critical region $t > t_{\text{cut}}$. Adapted from Ref. [151].

2.4 Hypothesis testing

Our *hypothesis* is a statement about the agreement between our predicted probabilities and the observed data. In this context, *hypothesis tests* with the frequentist approach give us a framework to accept or reject a given hypothesis based on the outcome of a given measurement. In most scenarios, we want either to know if all data events correspond to a given category (e.g., only SM events) or if the data contains at least some samples of a different type (e.g., a BSM signal). For simplicity, we refer to the first case as the *background*, while the latter corresponds to the *signal*.

When claiming the discovery of novel phenomena, in particular, we need to completely falsify the hypothesis that all events observed can be explained by the SM, that means *rejecting* the background hypothesis. In statistics, the hypothesis we want to test, the background, corresponds to the *null hypothesis* H_0 , which is to be compared to the *alternative hypothesis* H_1 , where there is both background and signal. Similarly as we did before, if we denote our data corresponding to n measurements of a random variable x as $\vec{x} = (x_1, \dots, x_n)$, both hypotheses H_0 and H_1 define a p.d.f. for $\vec{x}$, respectively denoted $f(\vec{x}|H_0)$ and $f(\vec{x}|H_1)$ ⁽¹³⁾.

That said, we quantify our hypothesis test via a function of the measured variables called a *test statistic* $t(\vec{x})$, also referred to as the *discriminant*. Each hypothesis defines a p.d.f. for the statistic t , namely $g(t|H_0)$ and $g(t|H_1)$, shown in Fig. 2.1. A statement can then be made through the definition of a decision boundary t_{cut} , enclosing a *critical region*, in such a way that if the statistic t obtained from the data is found within this region, H_0 is rejected. In Chapter 4, for instance, the choice of this threshold will be referred as the definition of a *working point*.

Also, the probability for t to be observed in the critical region assuming H_0 is called the *significance level* of the test, written as

$$\alpha = \int_{t_{\text{cut}}}^{\infty} g(t|H_0) dt, \quad (2.14)$$

(13) Note that H_0 and H_1 must be simple hypothesis, that is, they do not contain any undetermined parameters.

such that we reject H_0 if $t > t_{\text{cut}}$. Likewise, we say the probability of accepting H_0 while some alternative hypothesis H_1 is true is given by

$$\beta = \int_{-\infty}^{t_{\text{cut}}} g(t|H_1) dt, \quad (2.15)$$

where $1 - \beta$ is called the *power* of the test to discriminate against hypothesis H_1 . A *type-I error* is defined as the rejection of H_0 when it is true, which might happen with a maximum probability of α (vide Eq. (2.14)). Conversely, a *type-II error* is defined when H_0 is not rejected when the alternative H_1 is true (vide Eq. (2.15)).

When performing a hypothesis test, it is also interesting to quantify the level of compatibility between our hypothesis and the data, without considering an alternative hypothesis. Referred to as *goodness-of-fit* test, its result corresponds to the *p-value*, which is the probability, under the assumption of H_0 , of obtaining a result as compatible or less with H_0 than the one actually observed. If, for instance, high values of the statistic t correspond to poor agreement with the hypothesis, the *p-value* is the probability

$$p = \int_{t_{\text{obs}}}^{\infty} f(t|H_0) dt, \quad (2.16)$$

where t_{obs} is the value of the test statistic obtained from the observed data.

An important application of such a test in particle physics is when quantifying if an observation can be used to claim a new discovery. In this context, it is useful to convert the *p-value* into an *effective significance* Z , which is defined as the equivalent number of standard deviations of a Gaussian-distributed random variable. If we are looking for a positive signal such that only positive fluctuations are accepted, it is defined as

$$Z = \Phi^{-1}(1 - p), \quad (2.17)$$

where Φ is the cumulative distribution of the standard Gaussian and Φ^{-1} is its inverse function. If both positive and negative fluctuations are accepted, we write $Z = \Phi^{-1}(1 - p/2)$ instead. In particle physics, it is convention to claim a new discovery only when the significance is at least $Z = 5$, referred to as a 5σ effect, which corresponds to a *p-value* of 2.87×10^{-7} .

So far, we have described how to test our hypothesis against an alternative one, as well as to measure the level of compatibility between our hypothesis and the data, with the help of a test statistic t . Now, the definition of this function of the data depends on the scenario and a multitude of techniques is available to address them. Beyond the statistic t itself, the choice of the critical region will also depend on the purity of events we want to accept or reject (i.e., our tolerance to misidentified examples). Fortunately, statistics provides us with one particular way to optimise the task, the profile likelihood ratio test, which will be used in Chapter 7 and is described in what follows.

2.4.1 Profile likelihood ratio test

From the *Neyman–Pearson lemma*^[161], the optimal test statistic⁽¹⁴⁾ to distinguish between hypotheses H_0 and H_1 is the *likelihood-ratio test*^[160], written as

$$\lambda(\vec{x}) = \frac{f(\vec{x}|H_1)}{f(\vec{x}|H_0)}. \quad (2.18)$$

In practice, it might be difficult to obtain λ given its dependence on the likelihood, especially in the presence of nuisance parameters. Building upon what we discussed in Section 2.3.1, one way to construct a test statistic that is approximately independent of the nuisance parameters is via a *profile likelihood ratio* (PLR). It is given by the ratio of the profile likelihood function $L(\theta, \hat{\nu})$, from Eq. (2.9), divided by the maximum likelihood,

$$\lambda(\theta) = \frac{L(\theta, \hat{\nu})}{L(\hat{\theta}, \hat{\nu})}, \quad (2.19)$$

where $\hat{\nu}$ is the value of ν that maximises L for the specific θ , with $\hat{\theta}$ and $\hat{\nu}$ being the MLE. We say $\hat{\nu}$ is the conditional MLE of ν , which is a function of θ . A convenient test statistic is then given by

$$t_\theta = -2 \ln \left[\frac{L(\theta, \hat{\nu})}{L(\hat{\theta}, \hat{\nu})} \right], \quad (2.20)$$

which we refer to as the PLR test. This method will play a central role in Chapter 7, in the context of looking for new signals in binned data. For this reason, let us derive the formula we are going to use.

Our problem statement is the search for new phenomena, such that H_0 corresponds to our prediction of the known processes (only background events), while H_1 includes both the known and new processes (background and signal events)^[152]. After the experimental observation, our data is binned in histograms with N bins, such that the bin entries correspond to $\vec{n} = (n_1, \dots, n_N)$. Thus, the expectation value of n_i can be written as

$$E[n_i] = \mu s_i + b_i, \quad (2.21)$$

where s_i and b_i are the mean number of entries on bin i corresponding to signal and background events, respectively, while μ is the strength of the signal process. In this context, the background hypothesis H_0 corresponds to $\mu = 0$, while the signal hypothesis H_1 corresponds to $\mu \neq 0$. Also note that here we suppose the background b_i is known with negligible uncertainty.

As discussed in Section 2.3.1, we treat n_i as independent random variables which are

(14) By optimal, we mean a test that maximises the power of the test $(1 - \beta)$ for a given significance level (α) ^[151].

Poisson distributed around the mean $E[n_i]$. Hence, the likelihood function is given by

$$L(E[n_i]) = \prod_{i=1}^N \frac{E[n_i]^{n_i}}{n_i!} e^{-E[n_i]}, \quad (2.22)$$

with n_i and $E[n_i]$ corresponding to the observed and expected values for the number of entries in that bin, respectively. From Eq. (2.13), we can write the negative log-likelihood function as

$$-\ln[L(E[n_i])] = -\sum_{i=1}^N \left[n_i \ln(E[n_i]) + \ln\left(\frac{1}{n_i!}\right) - E[n_i] \right]. \quad (2.23)$$

And since for both our hypotheses $E[n_i] = \mu s_i + b_i$, it follows that^[152]

$$-\ln[L(\mu)] = -\sum_{i=1}^N \left[n_i \ln(\mu s_i + b_i) + \ln\left(\frac{1}{n_i!}\right) - (\mu s_i + b_i) \right]. \quad (2.24)$$

In this context, the hypothesis test corresponds to the comparison of the background-only hypothesis H_0 , where $\mu = 0$, to an alternative hypothesis where there are signal events. Since we are interested in searching for any signal whatsoever, we have no predetermined value for the signal strength μ . As a result, we can set our signal hypothesis to correspond to the most likely scenario given the data that was collected. To do so, we apply the techniques of parameter estimation described in Section 2.3, which states that the MLE for the parameter μ is given by

$$\hat{\mu} = \arg \min_{\mu} \{-\ln[L(\mu)]\}, \quad (2.25)$$

which we define as the strength of our signal in the alternative hypothesis. In summary, the expected number of entries for each hypothesis is given by

Background-only hypothesis : $E[n_i] = b_i$;

Signal MLE hypothesis : $E[n_i] = b_i + \hat{\mu} s_i$.

In this context, using the profile likelihood ratio test from Eq. (2.20), we can test the null hypothesis $\mu = 0$ against the alternative hypothesis $\mu = \hat{\mu}$. Also, we note that, in our case, we suppose the presence of signal can only increase the event rate, such that $\mu \geq 0$. That is, if we find a data such that $\hat{\mu} < 0$, then the best level of agreement between data and any value μ occurs in our background-only hypothesis, such that we have $\ln[L(\mu = 0)/L(\hat{\mu} = 0)] = \ln(1) = 0$. As a result, our statistic is given by

$$q_0 = \begin{cases} -2 \ln \left[\frac{L(0)}{L(\hat{\mu})} \right], & \hat{\mu} \geq 0, \\ 0, & \hat{\mu} < 0. \end{cases} \quad (2.26)$$

Furthermore, from Eq. (2.16), the p -value that quantifies the compatibility of the background-only hypothesis ($\mu = 0$) with the observed data is given by

$$p_0 = \int_{q_0^{\text{obs}}}^{\infty} f(q_0|0) dq_0, \quad (2.27)$$

where $f(q_0|0)$ is the p.d.f. of the statistic q_0 under the assumption of the background-only hypothesis and q_0^{obs} is the observed value of the statistic. This distribution can be approximated in the large sample limit, i.e., when the number of entries per bin n_i is large, by the expression⁽¹⁵⁾

$$f(q_0|0) = \frac{1}{2}\delta(q_0) + \frac{1}{2} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{q_0}} e^{-q_0/2}, \quad (2.28)$$

where $\delta(q_0)$ is the delta function. As a result, by replacing Eq. (2.27) and Eq. (2.28) into Eq. (2.17), the significance Z_0 in this asymptotic limit is given by

$$Z_0 = \sqrt{q_0} = \begin{cases} \sqrt{2 \left(n_i \ln \left(\frac{n_i}{b_i} + b_i - n_i \right) \right)}, & \hat{\mu} \geq 0, \\ 0, & \hat{\mu} < 0. \end{cases} \quad (2.29)$$

So far in this chapter, we have seen statistics concepts and techniques that allow us to infer information from experimental data. Just as crucial as choosing a good test statistic for the task in question, we must also make sure that we can properly quantify the performance of the test we are using. In the following section, we investigate some figures of merit when evaluating the performance of algorithms, metrics that will be largely used in different applications throughout this thesis.

2.5 Figures of merit and ROC curves

One of the main tasks we perform in particle physics, which will be featured in different contexts in this thesis, is the idea of distinguishing signal and background events. In Chapter 4 and Chapter 6, we will discuss an algorithm that classifies a given experimental signature as either an electron or another object that shares the same signature. Moreover, in Chapter 7, we will present a technique to identify BSM signals on top of the SM background. In all these applications, we need to assess the performance of our algorithms through an appropriate figure of merit.

As discussed in Section 2.4, the task of discriminating signal and background can be performed by defining a function of the experimental data, called a test statistic t . Just as important, we need to define a critical region such that our hypothesis is rejected if the statistic

(15) For the full derivation, see Ref. [152].

value is within the region. Each hypothesis will then correspond to a distribution of the statistic. We will denote $g(t|S)$ as the distribution from the signal hypothesis and $g(t|B)$ as that from the background. Their shape, illustrated in Fig. 2.1, together with an arbitrary value of the critical region threshold t_{cut} , allow a classification of a given event based on the value of the discriminant $t(\vec{x})$ obtained from the experimental data $\vec{x}$ associated to that event.

However, from the overlap between the two distributions, it is clear that our discriminant will inevitably classify certain background events as signals, as well as some signal events as background. In the first scenario, we have a type-I error, given by the α from Eq. (2.14), which correspond to a *false positive* (FP). Likewise, the type-II error, the β in Eq. (2.15), corresponds to a *false negative* (FN). Moreover, we say we have a *true positive* (TP) and *true negative* (TN) when we correctly classify a signal and a background, respectively. From these definitions, we respectively define the *true positive rate* (TPR) and the *false positive rate* (FPR) as

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad \text{and} \quad \text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}. \quad (2.30)$$

In statistics the TPR is also referred to as the *sensitivity*, while the FPR corresponds to the *specificity* of the classifier.

From the definition of TPR, we see that it corresponds to the ratio of objects that are accepted by the test, i.e., classified as signal, from all the signal objects considered. That is, it quantifies how many signal examples the classifier got right. For this reason, it is said to correspond to the *signal efficiency* ϵ_{sig} of the classifier. Conversely, we can expect some background objects to also be accepted by the test and classified as signal. Hence, we denote the number of background events that is accepted by the classifier as the *background efficiency* ϵ_{bkg} . That is,

$$\epsilon_{\text{sig}} = \frac{\text{signal events accepted as signal}}{\text{total number of signal events}} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2.31)$$

and

$$\epsilon_{\text{bkg}} = \frac{\text{background events accepted as signal}}{\text{total number of background events}} = \frac{\text{FP}}{\text{TN} + \text{FP}}. \quad (2.32)$$

Now, in some applications, we might be interested in maximising the rejection of the background, thus guaranteeing a purer sample. In this case, it is convenient to use the concept of *background rejection*, defined as the inverse of the background efficiency,

$$\frac{1}{\epsilon_{\text{bkg}}} = 1 + \frac{\text{TN}}{\text{FP}}, \quad (2.33)$$

which has values always greater than unity. We note that in the limit $\text{TN} \gg \text{FP}$, the background rejection corresponds to the ratio TN/FP .

Of course, all these metrics will vary according to the critical region boundary t_{cut} that is

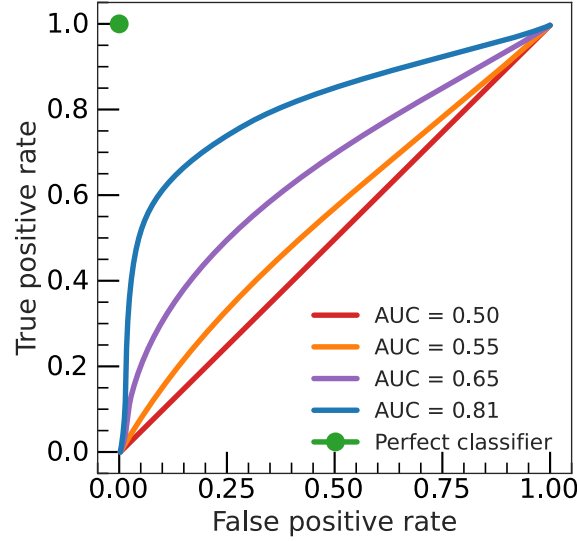


Figure 2.2: Illustration of ROC curves for different classifiers, defined as the true positive rate as a function of the false positive rate. One can alternatively write it as the background rejection $1/\epsilon_{\text{bkg}}$ as a function of the signal efficiency ϵ_{sig} .

used to define the classifier output. In this context, it is convenient to compare how these rates vary for different values of the threshold. Since both the TPR and the FPR will vary with the threshold, one convenient way to visualise this dependence is by plotting the TPR as a function of the FPR. Illustrated in Fig. 2.2, this plot is referred to as a *receiver operating characteristic* (ROC) curve⁽¹⁶⁾. For a perfect classifier, we will always have $\text{TPR} = 1$ and $\text{FPR} = 0$ for all threshold values, which means the ROC curve would be a single point in the top left of the plot. In contrast, if the classifier would just randomly choose between the signal and background classes, we would have $\text{TPR} = 0.5$ and $\text{FPR} = 0.5$ for all values of the threshold. Thus, it would present a diagonal line as its ROC curve. Hence, we expect a good classifier to have a ROC curve that approaches the top-left region of the plot.

The performance assessed via a ROC curve can be further condensed in a single number by measuring the *area under the curve* (AUC). This number is a convenient metric, since random classifiers have $\text{AUC} = 0.5$, while perfect classifiers have $\text{AUC} = 1.0$. Hence, the higher the AUC value, the better the classifier.

Lastly, in applications where we want to maximise the correct classification of background objects (as we will see in Chapter 6), it is convenient to rewrite the ROC curve as the background rejection $1/\epsilon_{\text{bkg}}$ as a function of the signal efficiency ϵ_{sig} . In this variation, we expect the best classifiers to have shapes with higher values of background rejection for any value of the signal efficiency. This is particularly convenient, since we know such models will have a good power to reduce background contamination for any value of signal efficiency that we choose.

(16) The name is a leftover from its use in the analysis of radar images, which was developed in the context of signal detection theory^[162,163].

All things considered, these metrics are essential to quantify the performance of any discriminant, independent of the choice of the critical region. They are particularly useful when implementing more sophisticated scenarios such as non-linear test statistics. The most notable example is a neural network, a powerful tool to parametrise the non-linearity within large datasets. Being a central technique used in this thesis, they are the subject of the following section, which starts with a contextualisation of the field of machine learning.

2.6 Machine learning

As we have seen so far in this chapter, statistics plays a central role in the research done in experimental particle physics. In fact, the large amount of data that needs to be processed when performing the measurements, combined with their high-dimensionality, makes the field a hub for innovation and application of state-of-the-art data mining techniques. These include, of course, the recent developments at the intersection of statistics and computer science in the domain of machine learning. In this section, we present a brief overview of this dynamic field of research and how it is being used in modern experimental particle physics, focusing on the techniques that we will revisit later in this thesis.

2.6.1 Motivation

The core principle behind machine learning techniques is the possibility that computer algorithms can *learn* to perform a specific task from examples. The term itself was popularised by Arthur Samuel in 1959^[164], but the idea was already discussed by Ada Lovelace more than 100 years before the first computer was built^[165]. This learning process would allow for the development of methods capable of detecting patterns in data and the automation of data analysis when building statistical models^[166].

Nowadays, the combination of powerful computational processing with the tools of applied statistics has opened the way for the development of artificial intelligence (AI) as we know it. However, the development of more sophisticated and complex algorithm structures was needed to deal with high-dimensional data and tackle more challenging tasks of AI, such as recognising objects or speech. These techniques concern the still-growing subfield of *deep learning* and allows computers to learn how to perform tasks that we would say are more intuitive to humans, such as identifying a face, but that cannot be passed as a set of individual mathematical instructions^[167].

The development of these machine learning methods and their success in a variety of tasks has provided a toolbox of techniques that are gradually being used more in the natural sciences, especially in physics^[156,168]. Being a tool of statistical inference by construction, particle detectors such as the ATLAS detector offer a perfect scenario for the implementation and development of machine learning techniques. Likewise, the increased amount of data

to be recorded in the next years, especially with the HL-LHC, requires an effort to find more robust ways to perform analyses of this data, with machine learning techniques having the potential to address these challenges.

In this context, this thesis presents a glimpse of the usefulness of these techniques with two applications being discussed at length in Chapters 6 and 7. That said, we start by defining some basic concepts that we hope are useful to navigate and understand the central ideas of this field of research.

2.6.2 Basic concepts

Any machine learning application starts by the definition of a task to be performed by a computer. In its simplest form, given a set of input features $\vec{x} \in \mathcal{X}$, the goal of the algorithm is to find a model $f : \mathcal{X} \rightarrow \mathcal{Y}$ that can predict the corresponding label $y \in \mathcal{Y}$ associated to a given input. For instance, for the task of identifying if a black-and-white image contains a cat or a dog, the input is the grid of pixels of the image, with values normally ranging from 0 (white) to 1 (black) and the output is one of the two labels, “cat” or “dog”. More precisely, the output is usually expressed as probabilities for the input to be associated to each one of the labels, which can then be treated to provide one single label if needed.

Now, the search for this optimal model is based on a performance metric that establishes the quality of each prediction. By having access to a huge number of examples of input features and their corresponding labels, the algorithm is able to improve its prediction capabilities by minimising the error based on this metric. This process is what we refer to as *learning*.

Moreover, the choice of the algorithm that will look for this optimal model depends, of course, on the type of data that is available. Methods that use labelled data, such as pairs of inputs and their corresponding labels, perform the so-called *supervised learning*^[167]. Tasks such as classification and regression usually require this type of data to perform well. On the other hand, some tasks do not need labels from the start, such as clustering or dimensionality reduction. In this case, the goal is usually to infer labels from the existing structures in the data^[156], performing what we call *unsupervised learning*^[166]. Similarly, techniques that require just a few entries of labelled examples in the data are often referred to as *weakly-* or *semi-supervised learning*. Tasks such as matrix completion^[166], where we try to infer values for missing entries, can benefit from techniques in this category. Lastly, approaches that exploit a system of trial and error, where the algorithm aims to maximise a given reward, are described as *reinforcement learning*^[169].

Once a given task has been established, the process of learning itself relies on quantifying how well the model performs through the definition of a *loss function*. This measure of the model success associates a value to the error depending on the prediction made and must be adapted to the task at hand. Some of the most common metrics are the misclassification rate for classification and the mean squared error for regression.

In other words, the goal of any machine learning algorithm is to find an optimal predictor that minimises this loss function^[170]. More precisely, by minimising the loss function we are also minimising the *empirical risk*, which is the average of the loss function over the data samples, that also corresponds to an estimate of the risk function in the frequentist sense^[166]. Some contexts can require a given function to be maximised instead, so this function is more generally called *objective function* or *criterion*^[167]. This process of optimisation where the algorithm goes over a set of examples and tries to minimise the loss function in the search for an optimal predictor is referred to as the *training* of the model.

During this learning process, the algorithm adjusts the *parameters* of the model, which can be seen as weights that determine how much each one of the input features influences the prediction^[167]. The optimal model is then defined as the set of parameters that minimises the loss function. Furthermore, there are a number of characteristics that one can specify in the machine-learning algorithm itself, variables that define the structure used to find the best model, called its *architecture*. Quantities such as the number of layers in a neural network or the function of choice by which the values are to be evaluated after each node are some of these variables that can be adjusted. These parameters of the network are called *hyperparameters* and must also pass through a process of optimisation. This process, however, is usually based on trial and error, and, as such, relies heavily on the experience of researchers^[156].

The main example of an optimisation technique used to minimise the loss function is called the *gradient descent*^[170]. This iterative algorithm converges to a local minimum by considering the steepest direction, that is, against the gradient of the objective function, with the size of each step being quantified by the *learning rate*. A variety of techniques exist to improve the chances that the algorithm properly finds the loss-function minimum^[166]. For instance, the method we are going to use in our applications, the *adaptive moment estimation* (ADAM) optimiser^[171], uses a variable adaptive learning rate that progressively gets smaller as the minimum of the loss function gets closer.

Once an optimal model is obtained after the training stage, its performance must be determined using examples that were not used in the training and thus never seen by the model. This *validation* over “unseen” data is used to establish the quality of a model when compared to others. This approach is particularly important in identifying if there was *overfitting*, that is, if the model performed well during training because it “memorised” the results instead of learning the important features needed to perform the task^[170] (akin to fitting a polynomial of order n over a dataset with $n + 1$ points). Some of the techniques we use to avoid this overfitting are the L_2 ^[172] and Dropout^[173] regularisation.

This validation step is crucial to determine if the optimal model will generalise well and have a good performance when deployed. For this reason, once the algorithm has made predictions for all the examples available in the training set, the metric used to measure the performance of the model is obtained from its predictions on the validation dataset instead. We refer to this step as an *epoch*, and repeat it as long as needed until we get a satisfying metric

value.

Furthermore, to ensure the generalisation of the model, it is important that the data used for training matches the distributions of the data upon which it will be used. Although having more data is helpful, it is generally not enough to guarantee a good performance if its quality is not adequate. For instance, the samples used must be independent and identically distributed (i.i.d.), as defined in Section 2.3.1, which means that all the random variables have the same probability distribution and the probability of obtaining one example does not change the probability of obtaining another^[159]. As a result, it is crucial to perform some *data engineering*, to guarantee the data we have available for training and testing the model reflects the samples on which it is expected to be used.

In addition, when applying machine learning to any real-world problem, it is important to understand the limitations of its use. For instance, due to its capabilities to extract more subtle features from data, it is natural to see the interest in using these techniques when dealing with many input features. However, difficulties arise in problems with an input space with a large number of dimensions, usually referred to as the *curse of dimensionality*^[174]. If N samples are required to describe the p.d.f. in a single dimension, scenarios working with d input variables will have an exponential need for more data, at $\mathcal{O}(N^d)$, in order to train the model^[170]. Fortunately, real-world data is usually well behaved and without discontinuities, so it can be described by a lower-dimensional space, a characteristic that most of the techniques exploit when working with high-dimensional inputs^[170].

Moreover, since there are regularities in the data that can be exploited by the algorithms when finding the optimal model, we can assume that induction is possible. This means that all machine learning algorithms make assumptions, which leads to *learning* or *inductive bias*. Finally, it is also important to consider that there is no type of algorithm that performs well for all classes of problems, stated as the *no free lunch theorem* for machine learning^[175]. A good understanding of the available techniques is the best way to have a successful application of machine learning in particle physics, and some of these tools will be described in what follows.

One of the most common tasks that are performed by machine-learning algorithms concern *classification*, which will be heavily featured in this thesis. It looks to establish a model f that best approximates the output associated with a given input when the output is part of a discrete set of categories. Another task that we will discuss in this thesis concerns anomaly detection, which tries to identify outlier events from the experimental data. Before discussing how these techniques are used in the field of particle physics, we discuss some of the architectures that define our neural networks.

2.6.3 Fully-connected network (FCN)

One of the earliest models in the history of neural networks is the *perceptron*^[176], developed in the 1960s to perform a binary classification with just one node connected to all input features,

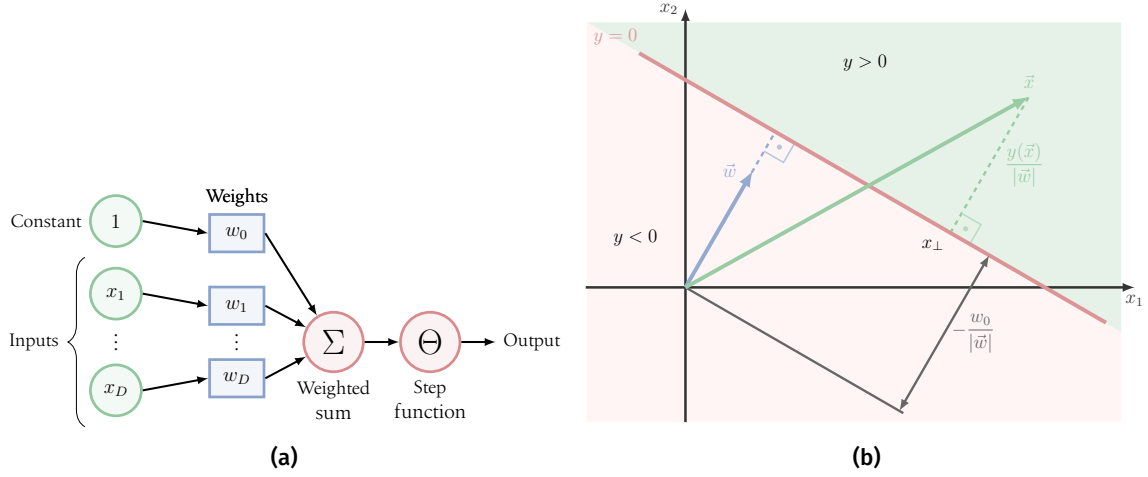


Figure 2.3: (a) Diagram of a perceptron where the weighted sum of the inputs $\vec{x} \in \mathbb{R}^D$ is passed through a step function to provide an output. (b) Hyperplane defined by a perceptron, here a line for a two-dimensional input, that splits the input space into two regions and thus only work with data that is linearly separable. Adapted from [170].

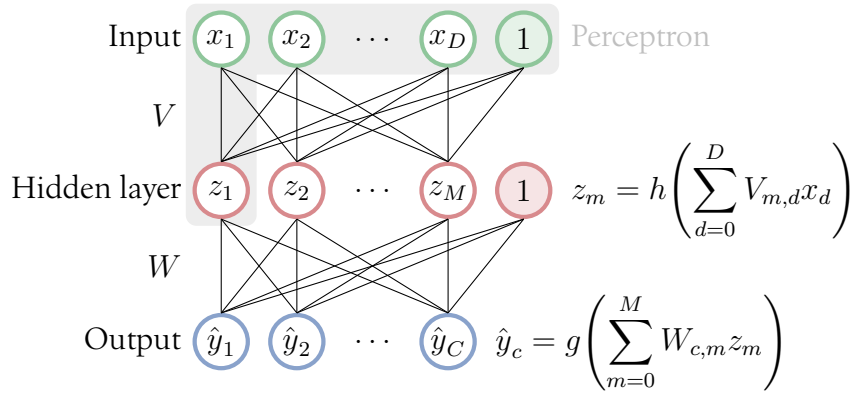


Figure 2.4: Architecture of a multi-layer perceptron (MLP), also known as a fully connected network. The input $\vec{x} \in \mathbb{R}^D$ is associated with one of C labels y_c . The network shown contains just one hidden layer with M units. Each node corresponds to a perceptron, highlighted in grey. A constant value is added to the input as well as the hidden layer to account for a bias.

thus simulating a neuron. A diagram of its structure is shown in Figure 2.3a. The learning process of a perceptron consists on determining the weights $\vec{w} \in \mathbb{R}^D$ and the bias w_0 of the model

$$y(\vec{x}) = \Theta(\vec{w} \cdot \vec{x} + w_0), \quad (2.34)$$

such that if the weighted sum of the inputs $\vec{x} \in \mathbb{R}^D$ reaches a threshold, controlled by the step function Θ , the neuron “fires”. This model defines a hyperplane with dimension $D - 1$ in the input space; a diagrammatic representation is shown in Figure 2.3b for $D = 2$. This hyperplane splits the space into two regions, with all inputs in one side corresponding to the same label. By construction, therefore, it supposes the data is linearly separable, which limits

its applications to more complex tasks⁽¹⁷⁾.

The ability of neural networks to learn non-linear functions was only achieved with the development of the *multi-layer perceptron* (MLP). As the name suggests, the idea is to combine multiple perceptrons in a *layered* structure, as shown in Figure 2.4. From a D -dimensional input $\vec{x}$, we introduce one hidden layer with M nodes z_m . These correspond to M linear combinations of the input, with weights $V_{m,d}$, each one of them transformed by a *non-linear activation function* h . Each one of the outputs of the hidden layer z_m is then linearly combined with weights $W_{c,m}$ and transformed via another activation g function to obtain the final C outputs of the network. For such a network with one hidden layer, the output label y_c that the model learns corresponds to

$$y_c = g \left[\sum_{m=0}^M W_{c,m} h \left(\sum_{d=0}^D V_{m,d} x_d \right) \right], \quad (2.35)$$

where the index 0 in the sum corresponds to the bias incorporated as a constant entry in the input vector and the hidden units. Usually, the non-linear activation functions h and g are taken to be sigmoidal functions, such as the logistic sigmoid or the hyperbolic tangent^[170], the specific choice depending on the task. In this work, for instance, we will use the ReLU and the softmax activation functions, such that we can normalise the output of the network to a probability distribution^[43].

Furthermore, this network can easily be extended to include more hidden layers, with the number of layers and units per layer corresponding to the hyperparameters of the network. Given enough hidden units, it can be shown that the MLP can approximate any suitably smooth function to any level of accuracy, stated as the *universal approximation theorem*^[178]. Moreover, it is a type of *feed-forward neural network* and it can also be referred to as a *fully connected network* (FCN).

The replacement of the step function in the perceptron to a non-linear activation function in MLPs allows for the neural network to be differentiable with respect to the model parameters w ^[170], a shift responsible for the development of the techniques of *backpropagation*^[179]. In summary, this technique performs a gradient descent, described above, that implements the chain rule from calculus at each one of the layers. This technique introduces an optimised way to determine the network's prediction errors by calculating, in a computationally efficient way, the loss in the *forward pass* and its gradient in the *backward pass*^[170]. Once this error is established, the weights are adjusted based on the gradient direction that minimises the loss function and the process is repeated iteratively until a satisfactory minimum of the loss value is reached. It is worth noting that these techniques are not limited to MLPs and are used in other types of architectures as well.

(17) A critique of this model was published in the book "Perceptrons" by Marvin Minsky and Seymour Papert^[177], which negatively impacted the research done in neural networks and it is said to be responsible for the decline in funding for neural computing research until the mid-1980s^[170].

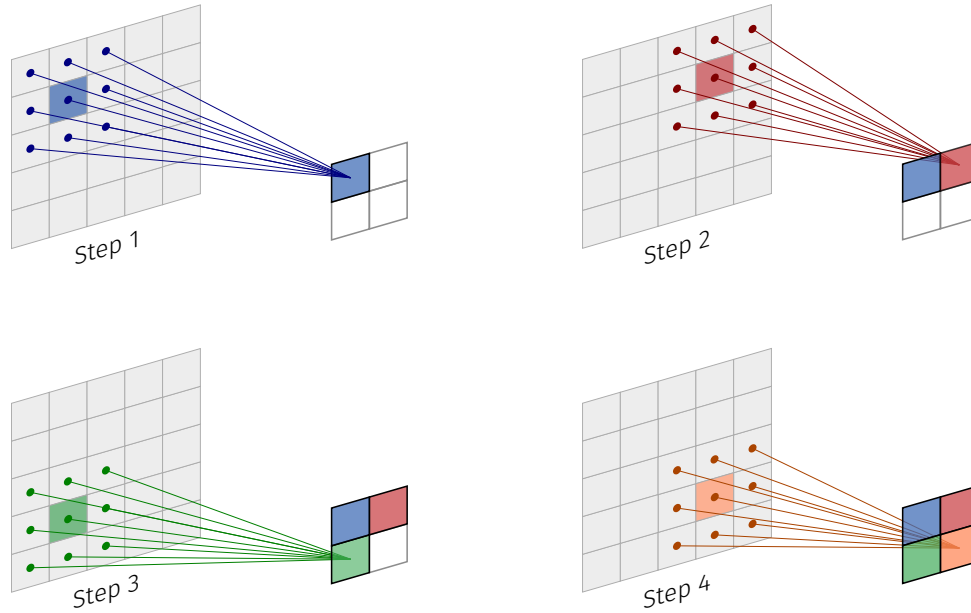


Figure 2.5: Illustration of the convolution operation. Adapted from Ref. [43].

Being one of the most straightforward architectures for neural networks, FCNs are widely used in many applications. But considering connections between all the elements of the input and its layers might not be the optimal approach in some scenarios. Especially with high-dimensional inputs, the resources needed to have so many weights in the model might be a problem. An alternative approach is to consider a different type of operation at each layer, such as the convolution, which we discuss next.

2.6.4 Convolutional neural network (CNN)

For tasks where the inputs can be structured in a grid-like form, such as an image, the surroundings of each input can be exploited to infer patterns, inspired of how the biology of vision works^[180]. Even for a single vector of numbers, the neighbouring entries can be relevant when extracting the patterns of interest. In this context, *convolutional neural network* (CNN) are specially designed for these tasks, being introduced in the 1990s to identify handwritten numbers in checks^[181,182]. They rose to prominence in the field when, in 2012, this particular architecture was used by the winning team^[183] of the ImageNet object recognition challenge^[184], a benchmark for visual classification and object detection by computers.

The principle behind the architecture of a CNN is based on the ability to identify localised elements in the input and combine them in an abstract way. This is done via the *convolution* operation, illustrated in Fig. 2.5, where the information of neighbouring cells is encoded in the form of a *kernel*. This matrix-like object stores the weights of the model and, in a way, abstract the features more useful for the task. The output of this convolution is the product of the kernel with the input entries, an operation that is repeated until all the input cells are covered. To control the size of this output, a *pooling* operating is usually performed,

such that only the maximum or average value of a region is taken instead.

The architecture of a CNN is then built as a sequence of convolution layers, each layer adding a level of abstraction to the model. In most cases, these layers are followed by a set of fully-connected layers to further incorporate the non-linearity of the features learned. Moreover, this architecture can easily be extended to work on multi-dimensional images as well. To do so, we treat each one of the dimensions as a different image, in the same way an RGB image can be decomposed in each one of the three colour channels, and then treat each channel as a separate image.

The appeal of CNNs lies on three aspects of its architecture. Firstly, different from an FCN, not all output cells are connected to all input cells, which means that the model itself will have fewer parameters. Not only that, but these parameters will be shared by all the input cells, since the kernel that runs over the image is the same across the whole input. This makes CNNs a lot more efficient on what concerns computer memory usage and processing time.

Another consequence of this parameter sharing, and an additional motivation for the use of CNNs, is that the model is equivariant to translation^[167], so the features learned in a given region of the image can be reused in another region of the image. As an example, if the model is able to identify a given feature in the middle of an image, it will be able to identify the feature if it appears in the bottom of another image. This is not true for other types of transformations, such as rotation and scaling, which would need to be handled using other techniques^[167].

Hence, it is easy to see that a natural application of CNNs is the task of image classification, where the goal is to match a given image to a label. Still, the architecture has also been shown to perform well at text detection, speech processing and caption generation^[185]. In high-energy physics experiments, one of the earliest applications of CNNs was done by the *NoVA* experiment when processing the two images produced by a scintillator^[186]. More recently, CNNs have also been used for other types of analysis in particle colliders, such as classification and anomaly detection in jets^[187].

This architecture will be used in both Chapter 6 and Chapter 7, being applied to different studies using the ATLAS detector. To further characterise the importance of machine learning techniques in particle physics, the next section presents an overview of its applications, focusing on the tasks that we will present later on.

2.7 Applications of machine learning on high-energy physics

Developments in the domain of machine learning are already being applied in particle physics in a number of different contexts^[153,155,188]. For years, physicists have been using techniques such as support vector machines, kernel density estimation, or more notably, boosted decision trees^[189]. But it was with the emergence of the deep learning techniques around the 2010s that large neural networks started being used to handle higher-dimensional and more complex

problems^[155].

In collider experiments such as ATLAS, examples of applications of machine learning techniques can be found across all the different areas of study. Starting at the low-level detector systems, neural networks are largely used to help the reconstruction of particle hits and their assembly into tracks^[190]. The interpretation of these signals can also be greatly improved when performing object identification^[191], an example of such application being the subject of Chapter 6.

Next, after the detector signals are interpreted, machine learning techniques are widely used to classify a given object or final state as either being expected within the SM (i.e., a background event) or corresponding to a completely new BSM phenomenon (i.e., a signal event)^[192]. Many techniques are also applying this classification in the system that decides which event signals will be recorded, called the trigger systems, such that we maximise the number of interesting events that are recorded^[193,194]. Some examples of notable analyses that rely on machine learning techniques include the searches involving t quarks^[195] and the Higgs boson^[196]. In Chapter 7, a novel machine learning approach will be presented to identify BSM resonances directly from the experimental data.

As introduced in Section 2.6.2, there are different approaches to defining which data will be used for the learning step. Given the probabilistic nature of the collisions happening inside the detectors, we cannot know which processes originated which signatures – this is what we are trying to determine in the first place. Hence, it is impossible to construct labelled examples from experimental data. Therefore, machine learning applications performing supervised learning usually relies on simulated Monte Carlo (MC) events, where the processes happening at the collision point are known and can be associated with their respective detector signatures^[154]. As we will discuss in Section 3.2.6, these simulations usually suffer from mis-modelling, and thus do not exactly correspond to a representation of the experimental data in some scenarios.

To avoid the dependence on simulation data, some alternatives have been developed. Techniques such as semi-supervised approaches incorporate the use of real data obtained with well-engineered cuts to correspond to background events^[197,198]. Weakly supervised approaches, on the other hand, allow the training of an optimal classifier from noisy experimental data instead^[199–201]. Lastly, we also have techniques performing completely unsupervised learning, where no label information is used, such as the use of autoencoders^[202–204].

In this thesis, we present applications of machine learning techniques in two particular tasks using the ATLAS detector. Firstly, Chapter 6 presents a technique to perform classification of detector signals, aiming to distinguish electrons from objects that mimic its signatures. Then, Chapter 7 presents a technique to identify resonances in the mass spectrum, thus identifying anomalies that are not explained by the SM. The remaining of this chapter is dedicated to a contextualisation of these two tasks, classification and anomaly detection, in the context of high-energy physics.

2.7.1 Classification

In many different contexts throughout experimental high-energy physics, we want to discriminate signal and background from each other. In this context, *signal* usually refers to an object for which we are looking, such as an electron we want to identify in the detector, or a particular process we are interested in studying, such as a particle decaying into a specific final state. Here, *background* corresponds to all the objects or processes that have similar signatures, but that, in truth, do not originate from the same phenomenon in which we are interested. For this reason, the use *classification* techniques are widely used in the field, where a discriminant value is obtained from a given list of observables $\vec{x}$. In this context, deep neural networks are a powerful way to approximate the optimal classifier in a high-dimensional scenario^[188].

One example of such classification task is the distinction between the signatures of objects detected by collider experiments. For instance, particles that decay hadronically are all recorded as a jet of collimated particles in the detector, making the distinction of the particle that originated the shower particularly difficult. To perform this task, deep learning networks have been used to discriminate between quark and gluon jets^[205], with other examples also considering the classification of jets originating from heavy particles such as the W , Z and Higgs bosons, as well as t quarks^[191]. In fact, many techniques are dedicated to the proper tagging of jets originated from t quarks^[206], b quarks^[207] or the distinction between the heavier b and c quarks against the lighter quarks, a task referred to as flavour tagging^[208]. Of course, the examples of applications of machine learning techniques to perform classification tasks is vast and will keep growing in the future^[153].

One thing that all these applications have in common is the need to reduce the dimensionality of the problem. To do so, physicists usually rely on the engineering of high-level features, which are functions of the low-level detector signals. However, deep learning models using low-level detector inputs directly have been shown to outperform the techniques relying on the high-level variables instead in some scenarios^[209]. In this context, the implementation of a deep learning network using low-level features on the task of electron identification will be presented in Chapter 6.

2.7.2 Anomaly detection

Another central task we want to perform is the detection of phenomena that are beyond what we expect within the SM. If they exist, these unusual processes are the key to quantifying the limitations of the SM and improving our understanding of the fundamental particles and their interactions. To address this challenge, machine learning techniques have been used to perform *anomaly detection* in particle physics^[210,211].

At the detector level, vide Chapter 3, these techniques usually involve the identification of anomalous particles, mostly jets, at the event-by-event level. Some examples include the search for anomalies within the jet substructure^[212–214], while others focus on the reconstruction of

its decay products, which were able, for instance, to rediscover the top quark^[215]. One of the main advantages of having networks capable of identifying unusual signals at the event-by-event level is their application at the trigger system used to decide if a given event will be recorded at all. Having such networks could then guarantee that interesting results are being recorded, with some of them already being developed^[193,194]. Moreover, the monitoring of detector systems can also benefit from anomaly detection techniques^[215].

Another way to incorporate anomaly detection in high-energy physics is in the search for resonances. A first approach consists on the direct comparison of simulated MC events with the recorded data^[216,217]. More recently, searches incorporating machine learning techniques have been developed to address the complexity of this task^[153,154,211,218,219]. Most of these techniques rely on simulated MC events for the training of the network, which is then used to isolate real data that is more signal-like. Other approaches incorporate experimental data to try to reduce the dependency on these simulations and have shown good performance^[210]. In Chapter 7, we will present a novel approach that tries to minimise the dependency on simulated MC events when searching for anomalies in the form of resonances in the mass distributions.

As we saw in this chapter, statistical methods are a powerful way to extract meaningful information from data and test our hypotheses. The comparison between the expectation from SM predictions and experimental measurements is the only way to identify its limitations and constrain alternative models. Moreover, machine learning techniques have shown great potential in tackling the high-dimensionality and complexities of the experimental challenges, while exploring the large amount of data already recorded by the detectors. In this thesis, we will apply many of these concepts in different aspects of the ATLAS experiment, namely the measurement of electron identification efficiencies (Chapter 5), the improvement of its classification task (Chapter 6) and searches for anomalous resonances in mass distributions (Chapter 7). Before discussing these applications, however, we present a description of the experimental apparatus in the following chapter.

Chapter 3

The ATLAS experiment

The most essential characteristic of scientific technique is that it proceeds from experiment, not from tradition.

Bertrand Russell^[220]

3.1	The Large Hadron Collider (LHC)	51
3.1.1	Introduction	51
3.1.2	Creating a proton collision	53
3.2	The ATLAS experiment	54
3.2.1	Inner detector	56
3.2.2	Calorimeters	59
3.2.3	Muon spectrometer	61
3.2.4	Trigger and data acquisition	61
3.2.5	Particles and event reconstruction	62
3.2.6	Event generation and detector simulation	63
3.3	Future upgrades	64

Modern particle colliders play a central role helping us understand the limitations of our current description of the fundamental particles and their interactions. Among them, the LHC has allowed the scientific community to venture into the highest energy scales ever achieved, with, for instance, the measurements performed by general-purpose detectors such as the ATLAS experiment. In this chapter, we introduce the ATLAS detector, the experimental apparatus in which we are interested in this thesis⁽¹⁸⁾.

3.1 The Large Hadron Collider (LHC)

3.1.1 Introduction

Being the world's largest particle accelerator, the Large Hadron Collider (LHC)^[221,226] has been a protagonist of the latest advancements in the field of particle physics since it began operating.

(18) Note that, unless otherwise stated, this chapter is based on Refs. [43, 221–224].

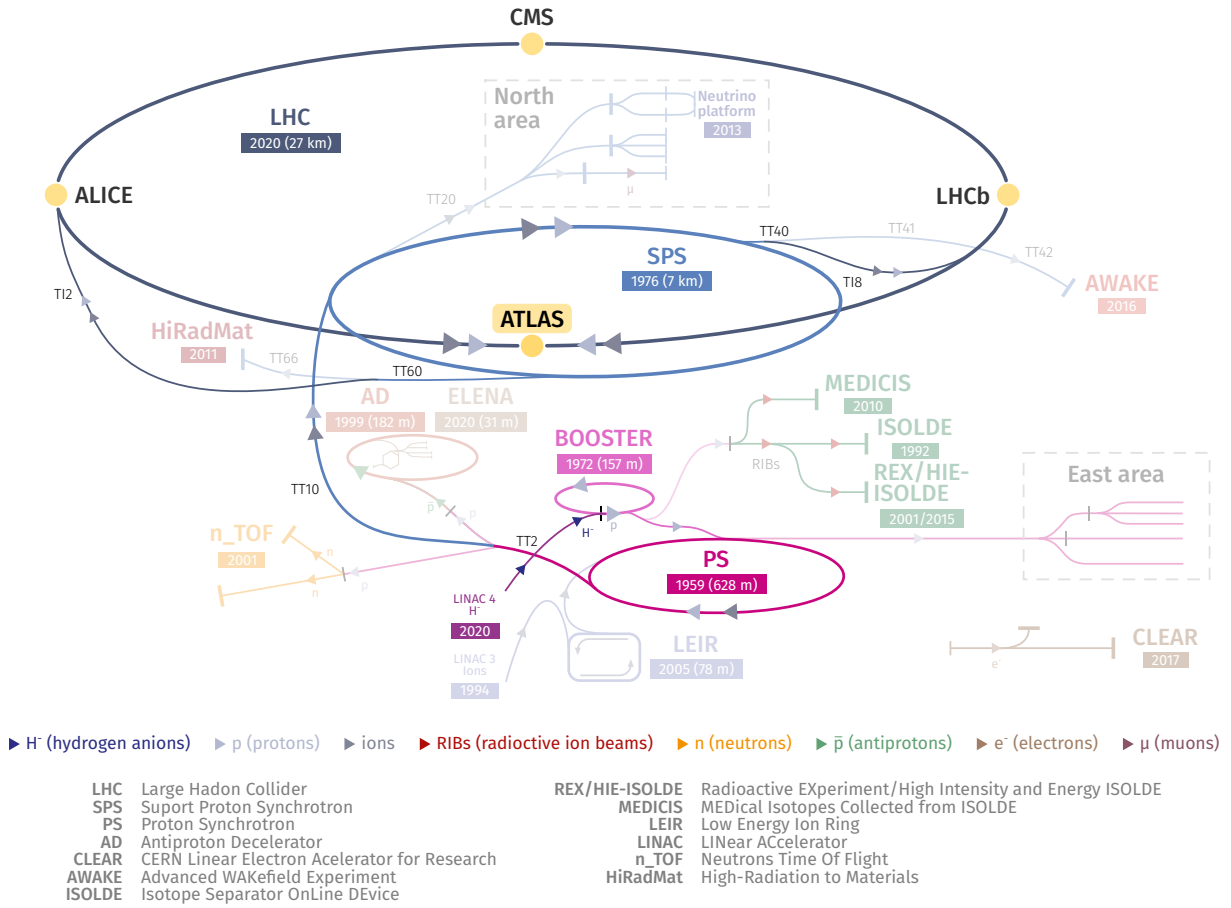


Figure 3.1: The CERN accelerator complex, highlighting the ATLAS experiment, the LHC and its injection chain. Adapted from Ref. [225].

This particle accelerator located on the border between France and Switzerland is home to detectors and experiments with different objectives, studying the collisions of high-energy particle beams. The results produced at the LHC have allowed the physics community to push the energy frontier and explore particle physics in conditions never explored before.

Constructed on the same tunnels that accelerated electrons and positrons at the Large Electron-Positron (LEP) collider (1989–2000)^[227], the LHC is part of the CERN accelerator complex, illustrated in Fig. 3.1. Its main ring has 27 km of circumference with four crossing points, where dedicated detectors are positioned. A total of nine experiments are performed at the LHC: the ATLAS^[222] and CMS^[228] experiments, large general-purpose detectors; the LHCb^[229] and ALICE^[230] experiments, detectors optimised for B -meson and heavy-ion studies, respectively; and the TOTEM^[231], MoEDAL^[232], LHCf^[233], SND^[234] and FASER^[235] experiments, which are smaller and specialised detectors located along the beam line close to the four major experiments.

Since it started operations, three data-taking periods, referred to as *runs*, were planned up to this day for the LHC, with two of them completed and the third one ongoing as of this writing. So far, many results have already been published after the successful completion of Run 1 (2009–2013) with centre-of-mass energies of 7 TeV and 8 TeV, as well as Run 2

(2015–2018) at 13 TeV. Today, the ongoing Run 3 allows the study of the collisions of protons at the record centre-of-mass energy of 13.6 TeV. Once the LHC programme is finished, a new accelerator is planned to operate in the same facilities within the next decade, the High-Luminosity LHC (HL-LHC)^[236,237]. Expected to start operations in 2029, it will allow the study of collisions with more energetic beams and increasing the amount of data recorded by factors of 5–7.

Combined with the collisions of heavy ions such as lead, the physics programme of the LHC experiments cover a wide range of high-energy phenomena, including the recreation of the hot and dense matter expected to have occurred just after the Big Bang.

In this thesis, we focus on proton–proton (pp) collisions and the particles created from these collisions that are registered by the ATLAS detector. Before colliding at the centre of the detector, the protons are accelerated throughout the accelerator chain, a process that we describe in what follows.

3.1.2 Creating a proton collision

The proton beam is accelerated in stages, starting from negative hydrogen ions H^- , which consist of hydrogen atoms with an additional electron. These are injected in the first accelerator in the chain, the Linear accelerator 4 (Linac4)^[238]. Using radiofrequency cavities^[239], the ions are pulsed through alternating positively and negatively charged conductors 400 ms at a time. At this stage, their energy ramps up from 3 MeV to 160 MeV. This principle of using electrical impulse to accelerate charged particles is used throughout the whole injection chain up to and including the LHC.

After leaving the Linac4, the electrons are removed from the ions and the remaining protons enter the Proton Synchrotron Booster (PSB). With its four superimposed synchrotron rings, it accelerates the protons up to 2 GeV before transferring them to the Proton Synchrotron (PS), where they will be further accelerated up to 26 GeV. The next stage in the accelerator chain is the Super Proton Synchrotron (SPS), where protons reach 450 GeV. At this point, the protons are ready to enter the LHC, with the beams entering two pipes, one running clockwise and the other anti-clockwise. After around 5 min, the full LHC ring is filled, with another 20 min being needed for the protons to reach their maximum energy at 6.8 TeV.

The circular trajectory of the beams is only made possible due to the strong magnetic fields that bend the path of the charged particles and are generated by superconducting technology. These dipole magnets are 15 m-long and guide the beam into tight turns along its circular path. Achieving 8.3 T, they need to be kept at a temperature of 1.9 K (-271.3°C) to function, with this cryogenics system being one of the main sources of energy consumption at the LHC. Moreover, quadrupole magnets are used to make sure the beam remains focused, especially before entering the collision points.

To maximise the chance that the protons will collide, they are grouped into *bunches*. At

the LHC, the beam is formed by 2800 bunches, each of them having approximately 10^{11} protons. They travel in a vacuum, which requires two weeks of cryogenic pumping to bring the pressure down to 10^{-13} atm. Reaching speeds of 99.9999991% of the speed of light, the beam is squeezed before colliding with the beam in the opposite direction at the collision points where the major detectors are positioned. The one in which we are interested, the ATLAS experiment, is the subject of the remainder of this chapter.

3.2 The ATLAS experiment

The ATLAS detector, acronym for *A Toroidal LHC ApparatuS*^[222], is one of the two general-purpose experiments at the LHC designed to study particles produced after pp collisions, but also heavy-ion collisions. Together with the CMS experiment, it allowed for the discovery of the Higgs boson in 2012^[69,70], a milestone that helped consolidate the SM^[4–8] as our current theoretical framework for understanding the fundamental particle interactions, vide Chapter 1.

Since then, measurements from ATLAS not only allowed for a better understanding of the properties of particles, but also served as a probe for phenomena beyond what the SM can describe. Moreover, the ATLAS experiment is the result of a global collaborative effort, a true testament to the cooperative and international nature of the scientific endeavour. It has been and it will remain a key player in the advancements of particle physics in the next decades.

As can be seen by its layout in Fig. 3.2, the ATLAS detector has a cylindrical shape around the beam line with the collision point at its centre. By registering the signatures left by particles over successive collisions, the detector allows for the measurement of the rate of events, R . This rate is directly related to the probability that an event can occur and is proportional to the cross-section σ of a process via the concept of *integrated luminosity* L via $R = L\sigma$. The value of L is usually given in units of $\text{cm}^{-2} \cdot \text{s}^{-1}$ and depends on characteristics of the collider. For two equal-size beams travelling close to the speed of light and colliding at their centre of mass, if we assume Gaussian charge distributions and densities that are uncorrelated in all planes, the integrated luminosity is given by the expression^[242]

$$L = f \frac{N_1 N_2}{4\pi\sigma_x\sigma_y}, \quad (3.1)$$

where N_1 and N_2 are the number of particles in each one of bunches travelling in opposite directions, f is the frequency with which they cross each other and σ_x and σ_y are the transverse beam sizes in the horizontal and vertical directions, respectively. It follows that rare events can be probed with higher luminosity by collecting more data. For instance, at the end of the second data-taking period, referred to as Run 2, after reaching an accumulated luminosity of 139 fb^{-1} , the ATLAS experiment was able to find strong evidence for the process with four top quarks at the final state^[243], the heaviest particle final state ever seen at the LHC.

Due to its geometry, the coordinate system used by the ATLAS detector places the z -axis

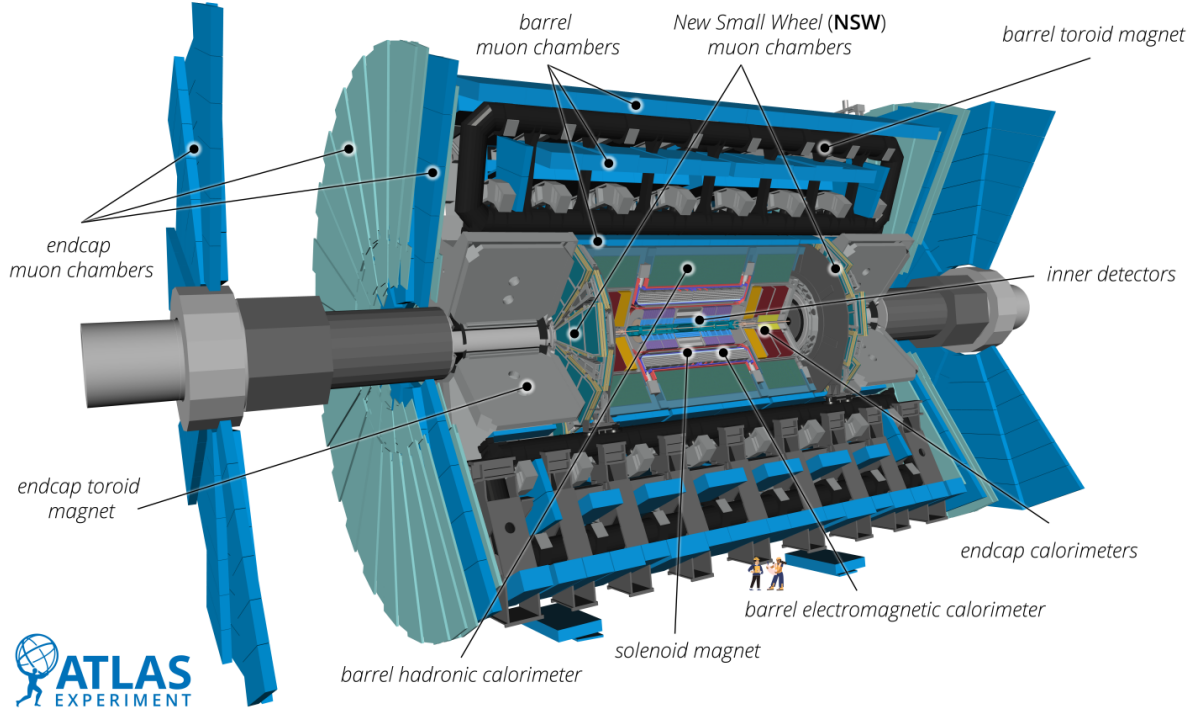


Figure 3.2: Illustration of the ATLAS detector. Taken from Ref. [240].

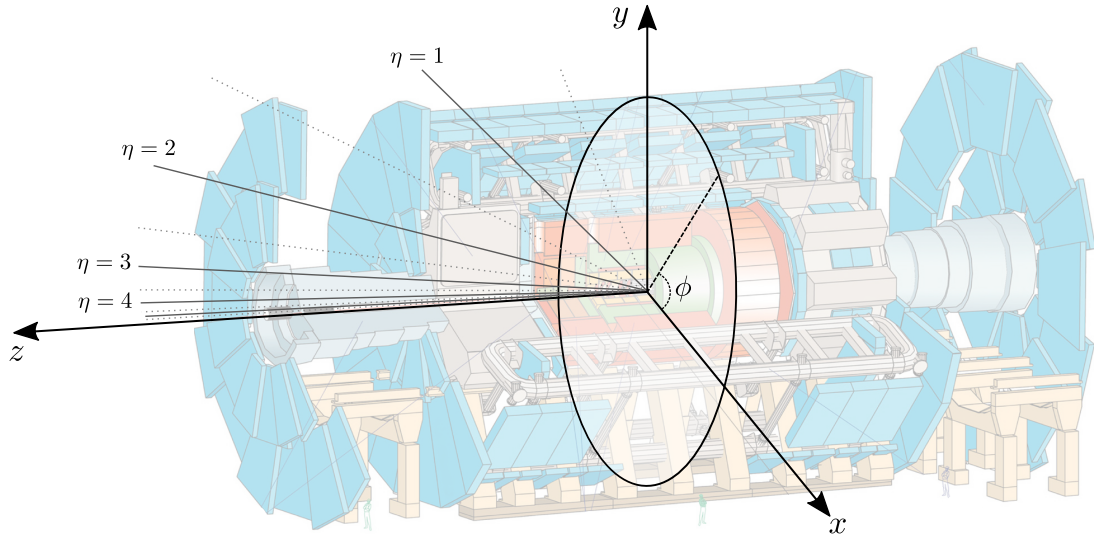


Figure 3.3: Schematic illustration of the ATLAS coordinate system. The z -axis is placed along the beam line, the y -axis points to the top and the x -axis points to the centre of the collider ring. The position of the particles is determined in terms of the azimuthal angle ϕ and the pseudorapidity $\eta = -\ln[\tan(\theta/2)]$, where θ is the polar angle. Adapted from Refs. [39, 241].

along the beam line and has its origin at the centre of the collision, called the *interaction point*. The x -axis points from the interaction point to the centre of the LHC ring, while the y -axis points upwards. The relative position of the particles is described by the azimuthal angle ϕ around the z -axis in the traverse plane, together with the *pseudorapidity* η , defined in terms of the polar angle θ as $\eta = -\ln[\tan(\theta/2)]$ ^[39]. A schematic illustration of the ATLAS coordinate system is shown in Fig. 3.3. All the angular distances are measured in units of the distance ΔR in the pseudorapidity–azimuthal angle space, defined as $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$. We also note that the part of the cylinder closer to the origin of the coordinate system is referred to as the *barrel region*, while the extremities of the cylinder are referred to as the *end-caps* or *forward region*. The region between these two is referred to as the *gap* or *transition region*.

When the two bunches travelling in opposite directions cross each other at the centre of the detector, multiple pp collisions happen. The most interesting one is referred to as the *primary interaction*, while the others correspond to the *pile-up*. The latter is quantified by the average of collisions per bunch cross, denoted $\langle\mu\rangle$. For Run 2, its value was measured at $\langle\mu\rangle \approx 30$, while for Run 3 it is expected to reach up to $\langle\mu\rangle \approx 60$ due to the increase in the centre-of-mass energy^[223].

Once a collision happens, the particles produced pass through the different layers of the detector, leaving signals in each one of these layers along its path. By studying the combination of these signatures, we can identify which particle corresponds to each path, as illustrated in Fig. 3.4. In what follows, the composition and purpose of each one of these layers are described.

3.2.1 Inner detector

The innermost part of the detector is dedicated to the detection and tracking of the path of charged particles up to 1 m in radius from the interaction point within the range of $|\eta| \leq 2.5$. It is comprised of two layers of silicon detectors, followed by an ensemble of drift tubes, all of this surrounded by a solenoid generating a strong magnetic field, as illustrated in Fig. 3.5. The whole sub-system has an extension of 5.3 m, with a diameter of 2.5 m,

The signals recorded from these sensors will then establish the points where a given particle crossed each one of the sub-layers. This system relies on the fact that a charged particle traversing the depletion zone of a semiconductor will create particle–hole pairs, as well as ionisation in the drift tubes, which can then be detected as electric current signals. When these points across sub-layers are combined, we can trace the path of the charged particle.

A high granularity of sensors is required in the first millimetres of the detector to distinguish between the tracks. For this reason, the first layers of the inner detector are formed by *silicon pixel detectors*, sensors with a nominal size of $50 \times 400 \mu\text{m}^2$ in $R - \phi \times z$. Referred to as the *Pixel* detector, it has four coaxial layers around the beam line, thus providing four hits or points of the charged particles' path up to 12 cm radially from the beam axis^[245]. Smaller pixel sizes are found in the layer closest to the beam line, the *insertable B-layer*, which has

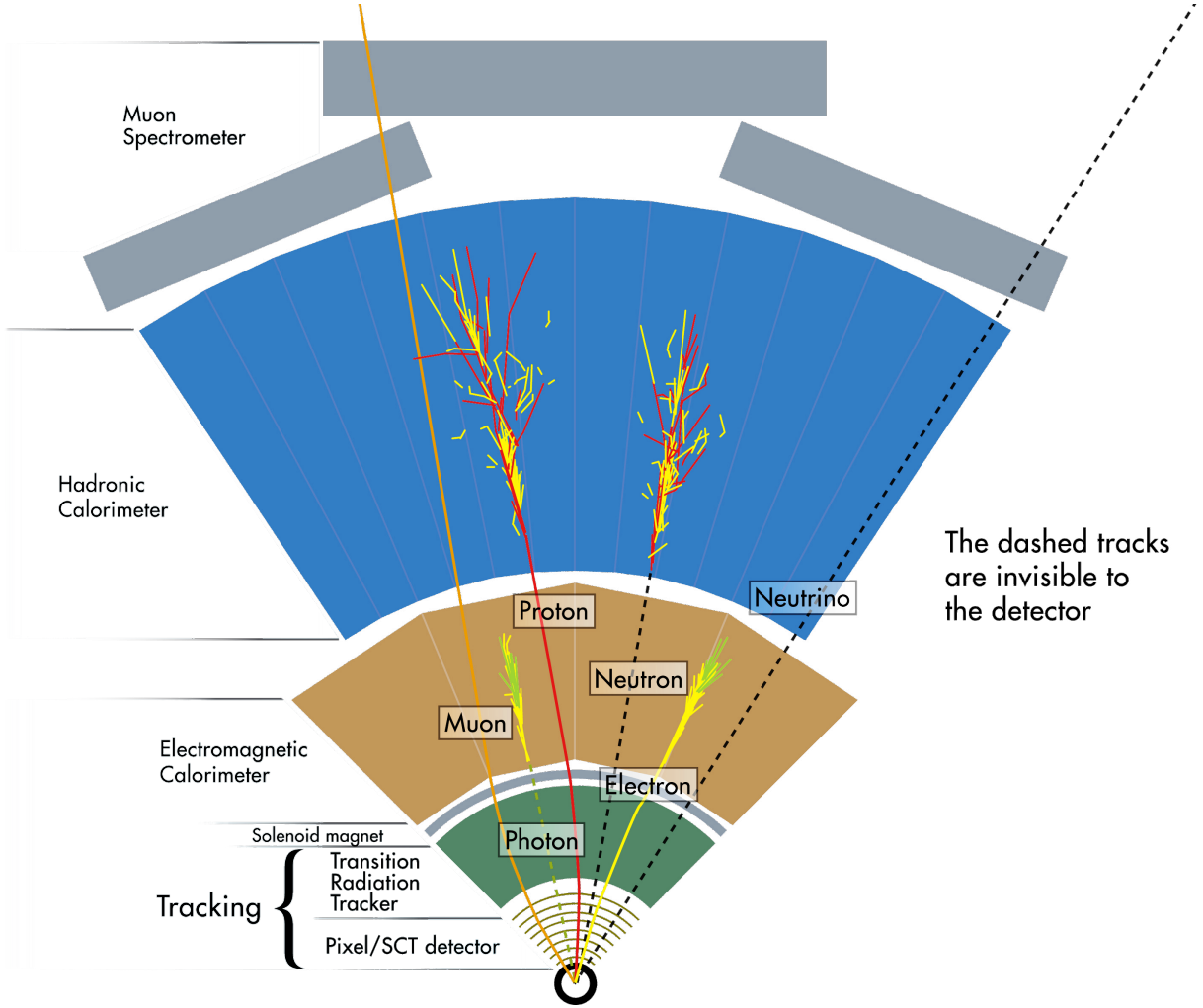


Figure 3.4: Particle signatures in the ATLAS detector. The inner detector (green) registers the path of charged particles. When particles enter the calorimeters (orange and blue), the succession of decays that follows forms a shower of electromagnetic or hadronic nature. The energy deposits are combined to determine the energy of the particle that originated the shower. Lastly, the muon spectrometer (gray) register the path of muons, the only charged particles that can traverse the detector this far. Adapted from Ref. [244].

pixels of $50 \times 250 \mu\text{m}^2$. This layer allows for a better discrimination of the tracks closer to the interaction point, thus improving the measurement of the impact parameter. This quantity is one of the important metrics to distinguish, for instance, particles coming from the collision itself or from heavy-hadron decays.

Next, the following layers correspond to the *semiconductor tracker* (SCT), which uses the same detection principle as the pixel detector, but with modules of two layers of silicon microstrip sensors. Since the strips provide information in only one direction, they are paired with an angular misalignment between strips. Thus, from its four layers in the barrel, they provide eight hits per track over the region with radius between 30 cm and 51 cm from the beam axis, while nine layers are found at the end-caps region.

The last set of tracking measurements is obtained by the *transition radiation tracker* (TRT). These are formed by straw drift tubes filled with a xenon-based gas mixture, interleaved with

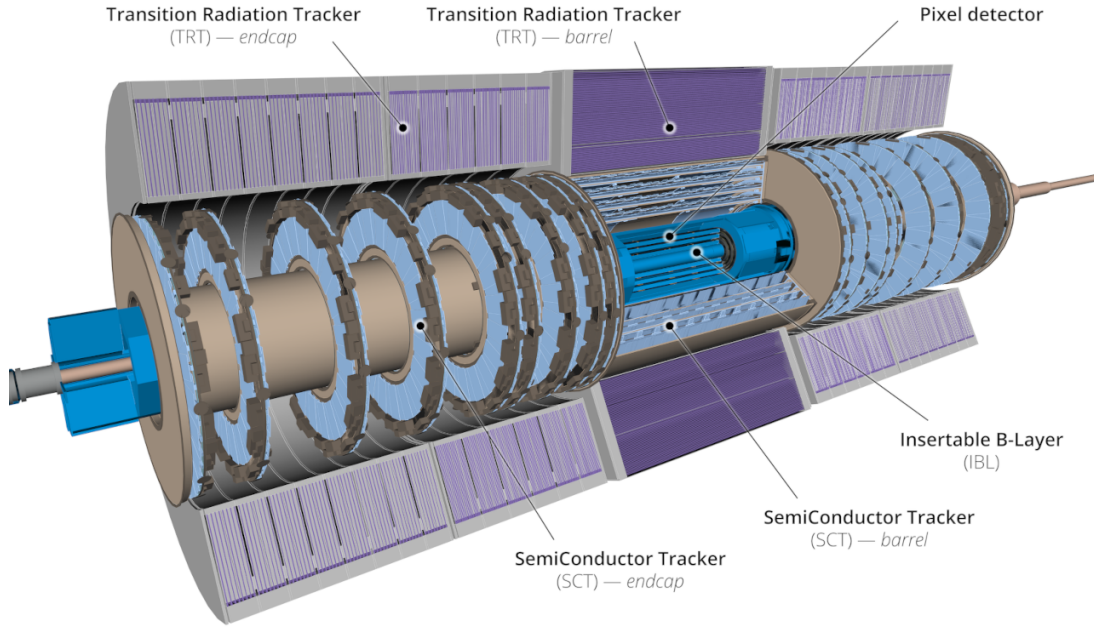


Figure 3.5: Illustration of the ATLAS inner detector. Taken from Ref. [223].

polypropylene/polyethylene radiators, and with a gold-plated tungsten wire in the middle of the tube. Each tube has a diameter of 4 mm, being 144 cm-long and parallel to the beam axis in the barrel region, while being 37 cm-long and arranged radially in the end-caps regions. Since xenon is a rare gas, some of the modules use argon-based gas mixtures instead, especially given it is difficult to fix gas leaks at such an internal part of the detector^[245]. The principle behind this detection technique is the fact that when the charged particles traverse the detector, they ionise the gas. The presence of a high electric field near the wire then generates a cascade of electrons that are registered as an electric signal in the wire. In total, the TRT provides the last 35 hits that will establish the path of charged particles, corresponding to the radial distances of 55 cm to 1 m from the beam axis.

Surrounding the inner detector, a *solenoid magnet* generates a magnet field of 2 T, responsible for bending the path of charged particles. The curvature of their path when traversing this magnetic field allows for the measurement of both their charge, depending on the direction to which their path is bent, as well as their transverse momentum p_T ⁽¹⁹⁾. Since this curvature is very subtle for the most energetic particles, which follow an almost linear path, it is crucial that the path is recorded with high precision. Thus, the use of semiconductor detectors is ideal for this task, since they are very compact and thin, providing good spatial resolution and fast signals.

(19) In this context, transverse refers to the plane perpendicular to the beam and corresponds to the xy plane in the ATLAS coordinate system, vide Fig. 3.3.

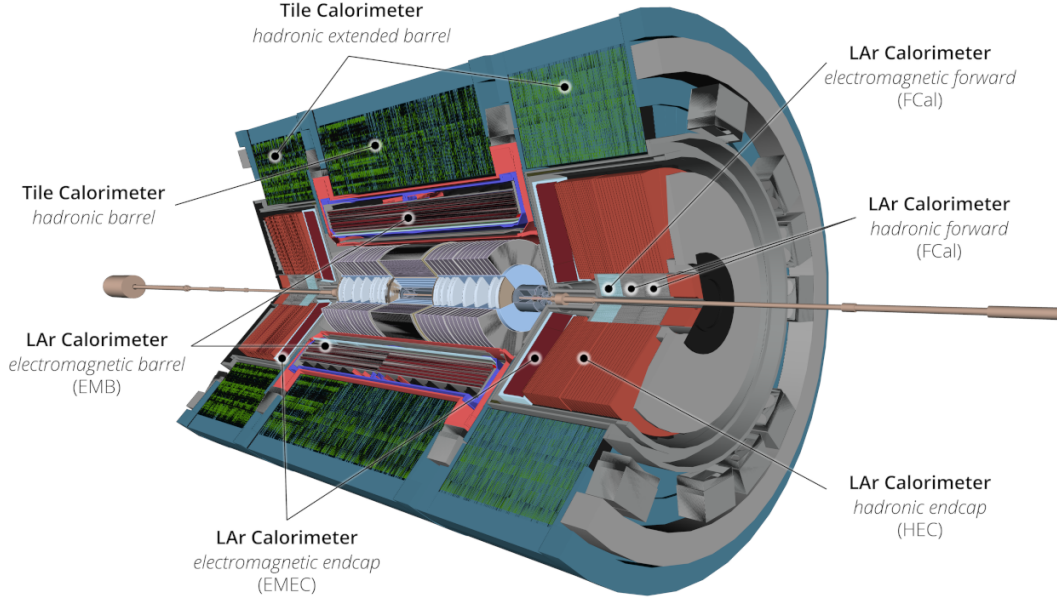


Figure 3.6: Illustration of the ATLAS calorimeters. Taken from Ref. [223].

3.2.2 Calorimeters

After the inner detector, the next layers of the detector, the *calorimeters*, are responsible for measuring the energy of both charged and neutral particles, vide Fig. 3.6. The principle of these devices is to induce a particle *shower*, as shown in Fig. 3.4. This means that when the particles interact with certain materials, referred to as the absorber material, they lose energy which is deposited in the form of ionisation or scintillation light. In the case of an electron, for instance, after interacting with the absorber material, it can emit a photon via the bremsstrahlung process, and this photon can also be converted into another electron–positron pair. The succession of these decays produces what we call an *electromagnetic shower*. A similar process also occurs with particles interacting via the strong force, with the difference that they can also produce hadrons which lose energy through nuclear interactions, forming a *hadronic shower*. By the interleaving of *absorption layers*, which induce the particles to develop these showers, and *sensitive layers*, which record the energy itself, the detector can register the energy deposited in each layer and reconstruct the total energy of the particle that initiated the shower.

In the ATLAS detector, the first calorimeter in the particle’s path is the *Liquid Argon* (LAr) electromagnetic (EM) calorimeter, responsible for registering the energy of particles decaying electromagnetically, such as electrons, positrons and photons. The absorption layers are made of tungsten, copper and lead, while the sensitive layers are chambers filled with liquid argon. These layers are interleaved in an accordion geometry, which allows for the placement of several layers in depth and offers a complete coverage over ϕ .

Moreover, this calorimeter is formed by two half-barrels covering $|\eta| < 1.475$ and two end-cap components covering the $1.375 < |\eta| < 3.2$ region, with a transition region around

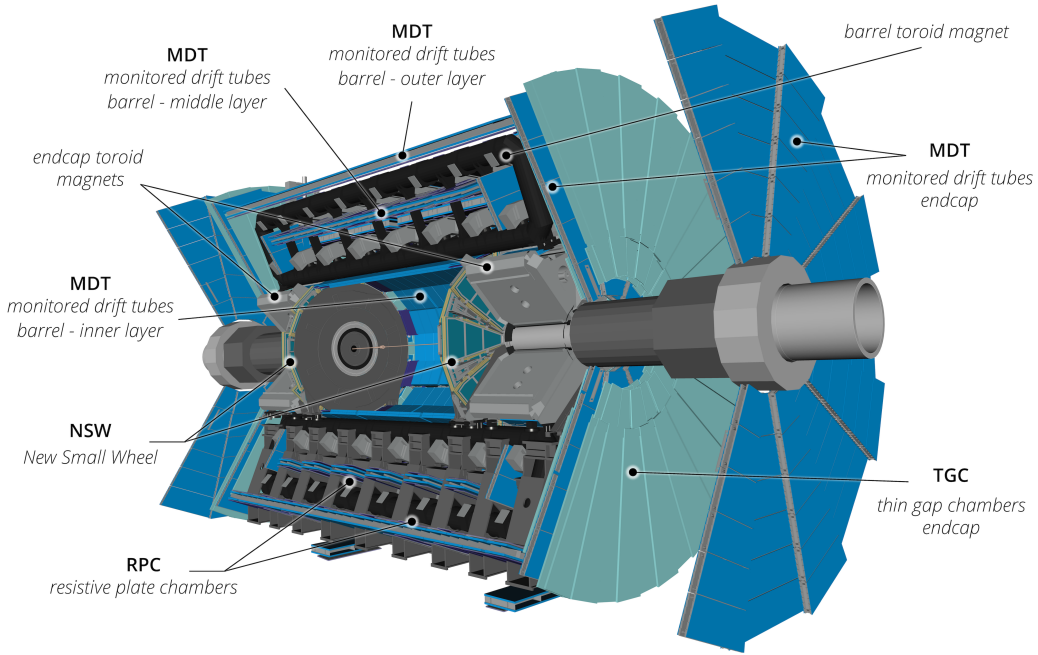


Figure 3.7: Illustration of the ATLAS muon systems. Taken from Ref. [223].

$1.37 < |\eta| < 1.52$ where the measurements are compromised. Radially, it is separated into four main regions: one pre-sampler and three sub-layers. The pre-sampler is responsible for improving the energy resolution and it measures the energy of particles that might have started the shower before the electromagnetic calorimeter. Next, the first sub-layer has a fine η segmentation and thus provides excellent information on the position, allowing for the discrimination between photons and $\pi^0 \rightarrow \gamma\gamma$ decays, for instance. The second sub-layer is the one that registers most of the shower and thus collects most of the energy. Finally, the information coming from the last sub-layer is used to establish the shower length and helps to distinguish electrons and photons from hadrons^[245]. This is possible since hadronic showers are much longer than the electromagnetic showers, reaching distances in the order of meters into the detector.

For this reason, surrounding the electromagnetic calorimeter, the *Tile hadronic calorimeter* is responsible for registering the energy of hadronic showers, although part of its energy is also deposited in the form of electromagnetic showers coming from $\pi^0 \rightarrow \gamma\gamma$ decays. The Tile uses steel in its absorption layer and a plastic scintillator for its sensitive layer, with its output light being transmitted through optical fibres to photomultiplier tubes. It covers the barrel and extended region ($|\eta| < 1.7$) and has a three-layer structure.

Lastly, in the region with $1.7 < |\eta| < 3.2$, the *Liquid Argon hadronic end-cap and forward calorimeters* are used to register the energy coming from hadronic showers in the extremities of the detector. The hadronic calorimeters also perform the energy measurement of the electromagnetic shower's tail and are important tools in discriminating events with hadronic origin^[245].

3.2.3 Muon spectrometer

The last layer of the ATLAS detector is responsible for the detection of muons, charged particles that can traverse the detector without having important interactions along the way, thus not producing any particle showers. Moreover, since they have a longer lifetime, of 2.2×10^{-6} s^[43], they typically do not decay inside the ATLAS detector. For this reason, a tracking system for muons is located beyond the calorimeters, called the *muon spectrometers*. They allow for the measurement of the muons momenta from the curving of their path by the non-homogenous magnetic field provided by *toroid magnets*. The ones located at the barrel region provide up to 2.5 T, while two of them are located at the end-caps and provide up to 3.5 T. In the barrel region, it consists of three concentric cylindrical systems, while three discs (or wheels) are found at each end-cap, as illustrated in Fig. 3.7.

In the $|\eta| < 2.7$ region, the measurement of the track coordinates of muons is performed by *monitored drift tubes* (MDT) chambers. These are formed by aluminium tubes filled with an argon-gas mixture with an axial wire serving as the anode. When charged particles traverse the gas, the electrons are knocked out of the gas molecules and drift towards the anode, thus producing an electric signal. Moreover, the azimuthal coordinate of the tracks is measured by the *resistive plate chambers* (RPC) for $|\eta| < 1.0$ and the *thin gap chambers* (TGC) for $|\eta| > 1.0$, both of which also containing dedicated trigger systems. For the larger pseudorapidities, namely $2 < |\eta| < 2.7$, *cathode strip chambers* (CSC) are used, given the demanding rate and background conditions^[222]. Referred to as the *old small wheel*, this sub-detector was removed at the start of Run 3^[223].

Additionally, in light of the increased luminosity expected for the next years, an important upgrade was introduced to the ATLAS detector before Run 3, namely the *new small wheel* (NSW)^[246]. Being the innermost wheel at the end-cap region, covering $|\eta| > 1.3$, it provides an improved p_T resolution. It uses two new chamber technologies: the small-strip TGC (sTGC) and the micro-mesh gaseous structure (Micromegas) detectors, which were designed to improve the precision in the bending direction and the azimuthal coordinate of the tracks, as well as delivering faster and more precise p_T measurements.

The assembly of all these layers forms the bulk of the detector apparatus. Throughout the whole structure, dedicated systems are designed to record the signals from all the sub-detectors and process them to decide which ones are worth saving. This is the task of the trigger and data acquisition systems, which are described in what follows.

3.2.4 Trigger and data acquisition

A total of 40×10^6 bunch crossings per second are delivered by the LHC at the centre of the ATLAS detector. Recording all these events is not only impossible, but also not really needed, since the most interesting events are usually rare (vide Fig. 1.6 in Chapter 1). For this reason, a quick and well-founded selection of the most interesting events is crucial to make

sure such rare events are not missed. This task of decision-making is assigned to the *trigger and data acquisition* (TDAQ) system, which reduces the actual data taking to a smaller rate that is recorded. This is accomplished by a combination of fast-processing hardware- and software-based algorithms that run after each event in a two-level system.

The first phase of the trigger system is called the *Level-1 trigger* (L1), which is a custom-built electronics system that processes the signals recorded by the sub-detectors with reduced granularity. The reconstruction of the event and decision-making must be completed within 2.5 μ s. At this level, the rate is reduced to around 100 kHz, which corresponds to 300 GB of data per second.

The next processing step happens at the *high-level trigger* (HLT), a software-based system that uses the full information from the detector sub-systems. This step uses a computer farm to reduce the rate to around 3 kHz rate, which corresponds to 6 GB of data per second. These final selection algorithms use multi-threaded CPU architectures since it also must operate at a very high-processing rate, spending 400 ms per collision on average.

Working in parallel to the trigger system, the *data acquisition* (DAQ) system is responsible for the transport of data from the detector to the trigger systems and finally to the offline storage. Each event that is chosen to be stored takes around 1.3 MB of data, which is made available for the whole collaboration via a dedicated cluster of servers.

Known as the LHC computing grid, simply referred to as the *Grid*^[247], this network of servers is distributed worldwide in various sites. This infrastructure provides data storage and computer processing for the whole scientific community working at the LHC. In particular, the data recorded by the experiments is distributed according to a hierarchical model, with a Tier-0 centre referring to the computer farms located at CERN, Tier-1 corresponding to 11 computer centres storing large portions of the data, followed by a larger number of Tier-2 centres scattered around the world which store a moderate fraction of the data and provide computer power for analyses. Beyond storing the data, the Grid is also used for the large-scale production of simulations, managing the distribution of jobs between the various Grid sites.

3.2.5 Particles and event reconstruction

Once the signals of the detector have been registered, the process of interpreting and making inferences from this data follows. It is by combining the different signatures measured by the detector that the particles can be distinguished, a process referred to as *reconstruction*. From its reconstruction, different algorithms are employed to further characterise and identify each object.

For electrons, for instance, this *detector signature* corresponds to the combination of a track in the inner detector with an energy deposit in the electromagnetic calorimeter. A similar energy deposit without an associated track corresponds to a photon. Similarly, charged hadrons will also leave a track in the inner detector, but their energy will be mostly deposited at

the hadronic calorimeters. Moreover, neutral hadrons have similar energy deposit signatures, but do not leave any track in the inner detector. Furthermore, tracks left in both the inner detector and the muon spectrometer allows for the reconstruction of the muons, and while neutrinos leave no signature in any part of the detector, a missing momentum in the transverse plane allows to determine the events where they might have been present.

The combination of all signatures registered after a collision defines an *event* and it is the basis of the data analysis that is done using particle collider data. Since we can suppose that all events are statistically independent, useful information can be inferred by applying the tools of probability theory to measure the rate with which events happen and compare it to the predicted values from theory, calculated in terms of cross-sections. Moreover, beyond the four-momenta obtained from the momentum-vector $\vec{p}$ and energy E measurements obtained from the track and calorimeter deposits, a number of other observables are defined. For instance, since before the collisions there is no momentum in the plane transverse to the beam line, it is convenient to work with the projections on this plane, such as the transverse momentum p_T . This allows for the determination of missing transverse energy E_T^{miss} from conservation of momentum, a metric used to indicate the presence of particles that leave no detector signature, such as neutrinos. Other variables of interest include the invariant mass m of a group of objects, an observable that will be featured in both Chapter 5 and Chapter 7. Depending on the measurement, more specific quantities can be constructed, being usually referred to as *high-level variables*.

In the ATLAS experiment, the reconstruction, simulation and analysis of the data is performed by a software framework called *Athena*^[248]. It uses Python as the configuration layer to load algorithms and objects written in C++. It is based on the Gaudi framework^[249,250], which defines the basic classes used for event processing and provides optimal algorithm execution. To take advantage of multi-thread machines, a redesign of the framework was performed in preparations for Run 3 to create a new software called *AthenaMT*. Using multi-threading, it implements parallelism at different levels, such as the processing of different events or different algorithms.

We note that the Athena software is regularly updated and new releases are made available. An important upgrade happened between major release 21 and release 22, where the use of Python 3 was adopted instead of its predecessor, Python 2, among many other upgrades. In light of this upgrade, a new software for measuring electron efficiencies was implemented, which will be discussed in Chapter 5.

3.2.6 Event generation and detector simulation

As we discussed in Chapter 1, an important part of the ATLAS physics programme is the comparison of the experimental data with the predictions made by the SM. These include, among others, the prediction on the probabilities that a given final state will be produced and

how it will interact with the detector. As described in the SM, these values can be estimated from the known interactions between the fundamental particles. In this context, the process of simulating these events and the subsequent interaction with the detector is referred to as *simulation*.

The first step when obtaining these simulations is the *event generation*, which uses the MC method to predict the probability that a given final state will be produced at the parton level^[251]. Since free quarks do not exist in nature, the process of hadronisation that follows the collisions has to be simulated after the event generation stage^[252]. The last step is then to pass these products of the collision through a simulation of the detector itself and thus obtain the predicted signatures of a given event^[224,253].

The process of obtaining data from this chain of stochastic simulations establishes the predictions for the rate at which a given event might occur. Moreover, these simulations are the only tool we have to match the true nature of particles and the quantities registered in the detector. For this reason, they help the development of algorithms aimed at the prediction of this truth information^[254], such as neural networks relying on supervised learning, as we discussed in Chapter 2. Finally, the data generated by these simulations is important to establish the efficiency of the detector, and subsequently determine the precision of the measurements performed from the experimental data. Since they are major components of the analysis chain, the quality of these simulations is of the uttermost importance and its ability to reproduce the experimental data is one of the major challenges of the physics programme of the ATLAS experiment.

However, these simulations are known to be imperfect^[255]. Already at the theoretical level, given that the calculations are based on perturbation theory, an approximation is done by keeping terms up to a given order, which introduces uncertainties to the modelling of the physics process itself. Furthermore, limitations on the precision with which we can describe the detector variables, like the energy calibration or its efficiency, adds to the mismodelling of the signatures of particles in the detector simulation. Other sources of discrepancies are also introduced by the uncertainties in our knowledge of some physical parameters, such as the mass of particles, or the limited number of events that can be generated in the first place. Mitigating these discrepancies is thus an important challenge of the physics programme of the ATLAS experiment.

3.3 Future upgrades

Since the start of data taking in 2008, the ATLAS collaboration has published over 1200 publications as of this writing, probing the physics of the SM and beyond. After Run 2 finished in 2018, the shutdown of the LHC machine allowed the installation of new infrastructures, both at the accelerator and the detectors. These included, for instance, the new linear accelerator (Linac4) and improvements to the injection beam energy of the PSB and the superconducting

radiofrequency cavities. For the ATLAS detector, this period allowed the installation of the Phase-I upgrade, which included the NSW and new front- and back-end electronics for the Liquid Argon calorimeter and the TDAQ system, among others.

In 2022, a new data-taking period started, the Run 3, where 250 fb^{-1} of data is expected to be recorded at a centre-of-mass energy of $\sqrt{s} = 13.6 \text{ TeV}$. Marking the last period of the LHC, it is expected to finish by the end of 2025 or 2026⁽²⁰⁾. In the meantime, the next generation of accelerators with more energetic beams is already being prepared. Among them, we find the HL-LHC^[236,237], which is expected to start operations in 2029. With the goal of increasing the luminosity recorded by the LHC by a factor of 5–7, some of the new technologies being developed for this upgrade include stronger superconducting magnets, precise control of the phase of the radiofrequency cavities and better beam collimation, among others^[237].

To prepare for the considerable increase in luminosity expected with the HL-LHC, the ATLAS experiment will undergo a Phase-II upgrade^[256,257]. Among the improvements, a completely new inner detector is being developed, namely the inner tracker (ITk). This all-silicon detector is designed to improve the track reconstruction and vertex identification, as well as extending the coverage to $|\eta| < 4.0$. Moreover, one of the major challenges of the HL-LHC is the much higher pile-up, which is expected to reach $\langle \mu \rangle \approx 200$. Since the calorimeter performance will be particularly degraded by this, a new high-granularity timing detector (HGTD) is being developed. With a precise timing measurement, combined with the position information of clusters, algorithms will be able to determine which cluster belongs to which pp collision within a given bunch crossing. Many other improvements are expected for the different detector sub-systems, including the muon spectrometers and the TDAQ system^[257].

In this thesis, we present measurements of the performance of the ATLAS detector when identifying electrons (Chapter 5), as well as applications of machine learning to improve the electron identification algorithm (Chapter 6) and optimise the search for new physics when analysing its data (Chapter 7). Before presenting this work, in the next chapter we take a closer look at how the reconstruction and identification of electrons is performed within the ATLAS experiment.

(20) As of this writing, running in 2026 is still uncertain, with the final decision expected to be made after the submission of this thesis.

Chapter 4

Electron reconstruction and identification

People populate the darkness; with ghosts, with gods, with electrons, with tales.

Neil Gaiman^[258]

4.1	Introduction	67
4.2	Electron reconstruction	68
4.3	Electron identification	69
4.3.1	Likelihood-based identification	72
4.4	Electron isolation	74

Being the first subatomic particle to be discovered^[20], electrons are the protagonists of the interaction between molecules and participate in a number of high-energy processes. As such, they play an important role in various measurements performed by the ATLAS experiment, especially those involving the Higgs boson, the $W^\pm$ and Z bosons, or the top quark. Part of the work in this thesis involves the study of the electron identification procedure, in particular the measurement of its efficiency (Chapter 5) and improvements on the algorithm itself (Chapter 6). Hence, in this chapter, we present the signatures of electrons in the detector, as well as the process with which they are reconstructed and identified.

4.1 Introduction

At the LHC, electrons are produced in various processes within the SM. They are the signature⁽²¹⁾ of the weak decays from Z and $W^\pm$ bosons^[245], and consequently of the Higgs boson via $H \rightarrow ZZ^* \rightarrow 4\ell$ ($\ell = e, \mu$)^[259], making them central in a number of precision measurements of the SM^[260,261]. Also, electrons are used to probe physics beyond the SM, such as searches for supersymmetry or new electroweak bosons^[262,263]. As a result, the ability to properly reconstruct the signals left by electrons and distinguish them from misidentified had-

(21) In this context, *signature* refers to the signals left on the detector, allowing the reconstruction of a given particle and, consequently, the event.

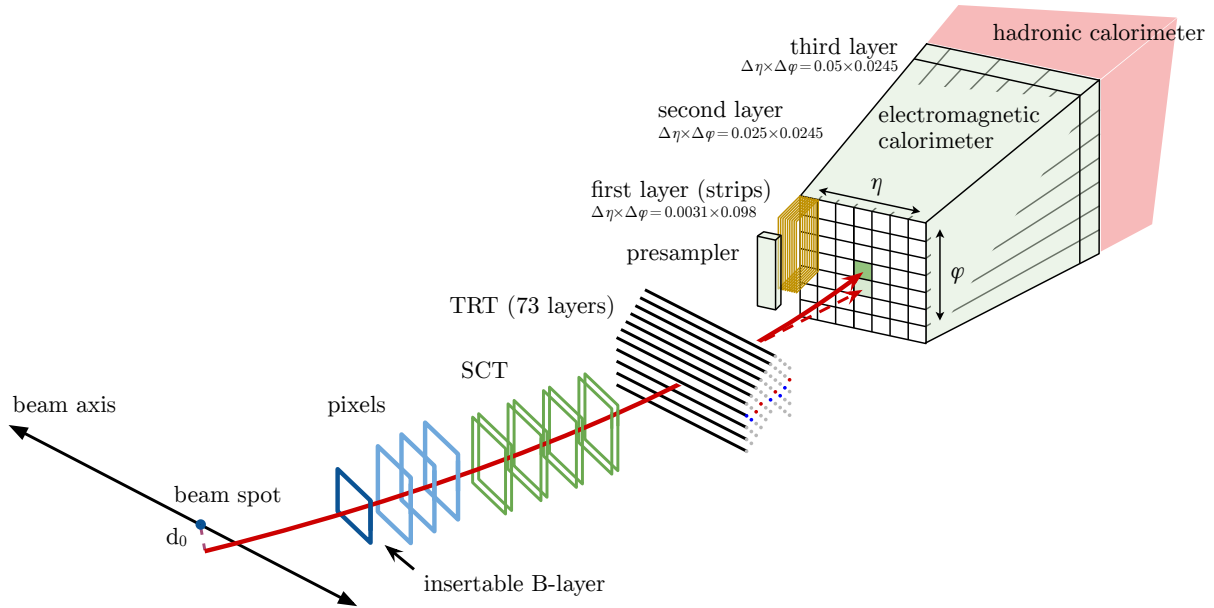


Figure 4.1: Illustration of the path of a prompt electron in the ATLAS detector. A track is formed by the combination of hits in the layers of the inner detector, namely the Pixel, SCT and TRT subsystems. The electron energy is then deposited in the electromagnetic calorimeter, and no signal is expected to be deposited in the hadronic calorimeter. Taken from Ref. [245].

rons or electrons not coming from these interesting decays are an essential part of the ATLAS physics programme.

In particular, a better understanding of the detector and the hadron collider environment has made viable a number of precision measurements⁽²²⁾ within the SM. Among them, measurements of the weak mixing angle^[264], the W boson mass^[260] and properties of the Higgs boson^[265] are all dependent on a good understanding of the electron signatures in the detector. Quantifying and reducing the systematic uncertainties associated with these objects is one of the major limiting factors in the precision of such measurements^[266].

Once an electron is produced at the collision point, it will traverse the different layers of the ATLAS experimental apparatus, described in Chapter 3. The detector registers the signatures of the products of collisions that happen at its center, and it is the task of the reconstruction procedure to characterise and interpret these signals as particles. In the following section, we discuss the specific signature left by an electron in the detector and how it is reconstructed.

4.2 Electron reconstruction

Following its production at the interaction point, an electron traverses the different layers of the ATLAS detector as depicted in Fig. 4.1. Throughout its path, signals will be recorded due to the electromagnetic interaction between the electron and the detector's subsystems. It is

(22) Precision measurements refer to those that aim the reduction on the value's uncertainty.

the goal of the reconstruction algorithm to interpret these signals and assemble the detector signatures that can be interpreted as electrons.

Starting at the inner detector, electrons will follow a curved path, given its charged nature and the strong magnetic field present in this region. This path is extrapolated from the hits recorded by the Pixel detector layers, including the insertable B-layer, together with the SCT double-sided strips and the TRT. The maximum number of hits recorded by these layers for an electron will be 4, 8 and 35 hits, respectively^[245].

From its interaction with the material of the detector, the electron will undergo bremsstrahlung processes, where the emitted photon will further decay into an electron–positron pair. This iterative conversion between electrons and photons form an *electromagnetic shower* that is recorded by the EM calorimeter. The energy deposits recorded by these calorimeter cells, referred to as *topological clusters*, are combined via an algorithm that allows the recovery of the energy of the shower^[245]. No amount of energy is expected to be deposited at the hadronic calorimeter. Using a sliding window algorithm^[267], this EM cluster is matched to tracks reconstructed from the inner detector^[266]. After undergoing position corrections and energy calibrations^[268,269], these signatures define what we call *electron candidates*.

In most analysis performed within the ATLAS experiment, we are interested in prompt electrons that originate from the interaction point. Referred to as *signal electrons*, these include electrons coming from the decay of W , Z and Higgs bosons, as well as the t quark, for instance. We must note, however, that other objects might leave a similar signature in the detector. Referred in what follows as the *electron background* or simply *background*, these other types of objects will also be reconstructed as electron candidates. These include jets, photons or electrons originating from heavy hadrons with b or c quarks, as well as charged hadrons from light quark flavours (u , d or s) such as the $\pi^\pm$ mesons. Electrons produced from the conversion of a photon, either produced at the interaction point or from a light-quark hadron decay (e.g., $\pi^0 \rightarrow \gamma\gamma$) are also considered as background. Lastly, non-prompt electrons that are not coming from the interaction point are also considered part of the background.

It is the goal of the identification algorithm to discriminate the signal electrons of interest from these background objects. In this context, the current implementation of the procedure used in the ATLAS experiment is described in the following section.

4.3 Electron identification

The proper identification of electrons relies on the understanding of the objects that mimic their signatures. By far the most common objects that populate the electron background are charged hadrons, mostly those formed by light quarks (u , d and s). Despite having a curved track and a large energy deposit in the EM calorimeter, these objects have a more diffuse shape in its energy cluster and a considerable amount of their energy is deposited in the hadronic calorimeter. Additionally, electrons coming from photon conversions into e^+e^- pairs or those

Type	Variable	Description
Track quality	$n_{\text{B-layer}}$	Number of hits in the innermost pixel layer
	n_{Pixel}	Number of hits in the pixel layer
	n_{SCT}	Number of hits in the SCT layer
	d_0	Transverse impact parameter relative to the beamline
	$d_0/\sigma(d_0)$	Significance of transverse impact parameter defined as the ratio of d_0 to its uncertainty
	$\Delta p/p$	Momentum lost by the track between the perigee and the last measurement point divided by the momentum at perigee
TRT	$e\text{ProbabilityHT}$	Likelihood probability based on transition radiation in the TRT. This variable is set to a default value for candidates with $\eta > 2.01$ due to the limited coverage of the TRT
Track-cluster matching	$\Delta\eta_1$	$\Delta\eta$ between the EM cluster position in the first layer and the extrapolated track
	$\Delta\phi_{\text{res}}$	$\Delta\phi$ between the cluster position in the second layer of the EM calorimeter and the momentum-rescaled track, extrapolated from the perigee, times the charge q
	E/p	Ratio of the EM cluster energy to the track momentum
EM calorimeter, first layer	$w_{\text{s tot}}$	Shower width, $\sqrt{(\sum E_i(i - i_{\text{max}})^2)/(\sum E_i)}$, where i runs over all strips in a window of $\Delta\eta \times \Delta\phi \approx 0.0625 \times 0.2$, corresponding typically to 20 strips in η , and i_{max} is the index of the highest-energy strip
	E_{ratio}	Ratio of the energy difference between the maximum energy deposit and the energy deposit in a secondary maximum in the EM cluster to the sum of these energies
	f_1	Ratio of the energy in the first layer to the total energy in the EM calorimeter
EM calorimeter, second layer	w_{η^2}	Lateral shower width, $\sqrt{(\sum E_i\eta_i^2)/(\sum E_i) - [(\sum E_i\eta_i)/(\sum E_i)]^2}$ where E_i is the energy and η_i is the pseudorapidity of cell i and the sum is calculated within a window of 3×5 cells
	R_ϕ	Ratio of the energy in 3×3 cells over the energy in 3×7 cells centred at the EM cluster position
	R_η	Ratio of the energy in 3×7 cells over the energy in 7×7 cells centred at the EM cluster position
EM calorimeter, third layer	f_3	Ratio of the energy in the third layer to the total energy in the EM calorimeter. Due to the known mismodelling at high $ \eta $, this variable is set to a default value for candidates with $ \eta > 2.01$
Hadronic leakage	R_{had1}	Ratio of the transverse energy E_T in the first layer of the hadronic calorimeter to E_T of the EM cluster
	R_{had}	Ratio of the transverse energy E_T in the hadronic calorimeter to E_T of the EM cluster

Table 4.1: High-level variables from the tracks, track-cluster matching and calorimeter that are used to discriminate electrons of interest from their background in the context of the LLH identification method^[245].

coming from hadronic jets from heavy-flavour decays (i.e., involving b or c quarks) will present a large impact parameter, which allows for its discrimination from the signal electrons that are coming from the interaction point.

As a result, from the recorded signature of the electron candidates, a number of discriminant variables are constructed to distinguish the signal electrons from their background. These are defined from the quantities recorded by the inner detector and calorimeter subsystems, and are listed in Table 4.1. In particular, their discriminating power relies on properties of the primary electron track and the lateral/longitudinal development of the EM shower, as well as the compatibility between these two^[268].

From the primary track associated with an electron candidate, the number of hits in each of the layers of the Pixel and SCT sub-systems is used to leverage its quality. Then, the transverse impact parameter relative to the beam line d_0 and its significance $d_0/\sigma(d_0)$ are used to quantify the displacement of the electron origin, values that will be higher for electrons coming from photon conversions and heavy hadronic jets. Moreover, the momentum lost by the electron along the track, measured as $\Delta p/p$, is used to quantify its loss of energy to bremsstrahlung, which is considerably higher for the signal electrons but not for the charged hadrons.

Likewise, the inner detector TRT subsystem, described in Chapter 3, is also optimised to distinguish electron signatures. In particular, light particles such as the electron radiate more photons when traversing the radiator foil when compared to heavier particles such as the $\pi^\pm$ ^[270]. This means more high-threshold hits are recorded in this subsystem along the reconstructed track for electrons, which, however, is not the case for its hadronic backgrounds^[271,272]. Thus, we define a variable called *eProbabilityHT* to quantify the likelihood of a given candidate to be an electron from the TRT signals^[245,273].

Additionally, since an electron produces a rather narrow cluster in the EM calorimeter, the lateral development of the shower is also an important discriminant. Quantified in variables such as the widths $w_{s\text{ tot}}$ and w_{η^2} , or the ratios R_η and R_ϕ , the lateral spread of the showers are expected to be larger for hadronic jets. Furthermore, these jets are expected to generate showers with larger depths, meaning that they can reach into the hadronic calorimeters. As such, the ratios of energy deposited in the hadronic calorimeter to the EM calorimeter, R_{had} and R_{had1} , are used to discriminate these hadronic backgrounds. Also, the depth of the showers within the EM calorimeter are quantified via ratios of the energy in its different layers, via the f_1 , f_3 and E_{ratio} variables.

Lastly, the process of matching which tracks correspond to a given EM calorimeter cluster helps identify if the pairing is appropriate. This compatibility of the calorimeter cluster with the track in the $\eta \times \phi$ space is quantified via the $\Delta\eta_1$ and $\Delta\phi_{\text{res}}$ variables. A further metric to assess if they correspond to the same object is the ratio of the EM cluster energy to the track momentum E/p .

It is via the definition of these high-level variables that the implementation of identifica-

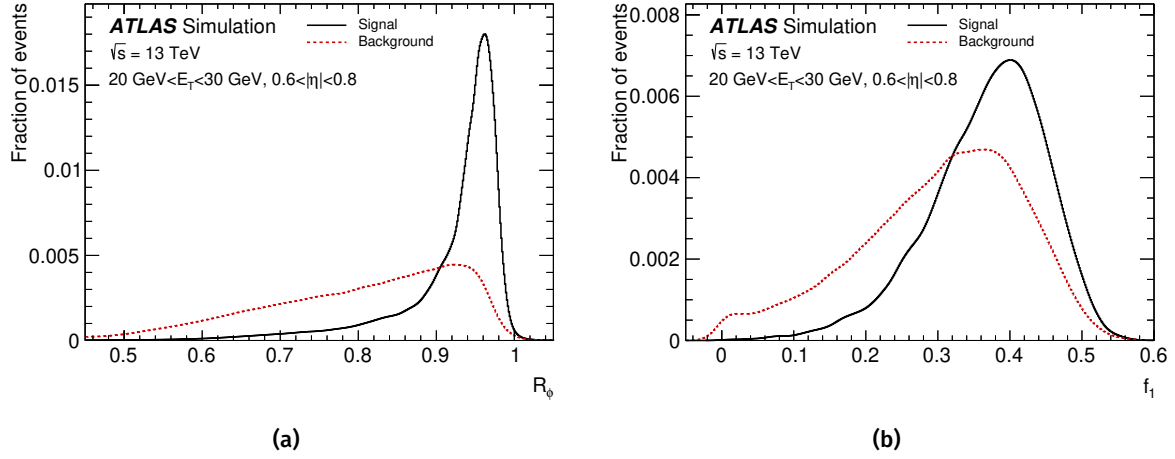


Figure 4.2: Distributions of the (a) R_ϕ and (b) f_1 observables, defined in Table 4.1, for both signal (solid black line) and background (red dashed line) electrons. Both distributions are obtained from MC simulated events, where $Z \rightarrow e^+e^-$ processes are used to obtain signal and QCD two-to-two processes are used to obtain background. The distributions have been smoothed and corrected. Taken from Ref. [245].

tion algorithms is optimised to discriminate between electrons and the objects mimicking their signatures. From 2010 to 2012, a cut-based approach^[271] was used, where simple thresholds on these variables were defined to separate signal and background. This method was implemented and optimised using the Toolkit for Multivariate Data Analysis (TMVA)^[189] with the ROOT framework. Its successor, the likelihood method^[245,274], has the advantage of combining the information of all the discriminating variables, as well as allowing the use of quantities that have similar distributions between data and MC simulation and thus could not have been used in the cut-based approach^[273]. This is the case, for instance, of the R_ϕ and f_1 variables shown in Fig. 4.2, which have significant discriminating power between signal and background. Since the likelihood method is the standard currently used for the identification of electrons in the ATLAS detector, we will describe it in what follows.

4.3.1 Likelihood-based identification

As previously discussed, the goal of the identification procedure is to decide if a given electron candidate corresponds to a signal electron, that is, one coming from the interaction point, or if it corresponds to one of its backgrounds. In the *likelihood-based identification algorithm* (LLH), the high-level variables listed in Table 4.1 are used as input to perform this task^[245]. The current ATLAS identification algorithm defines five working points, referred to as TIGHT, MEDIUM, LOOSE, VERYLOOSE and ULTRA LOOSE^[245]. These correspond to increased values of the discriminant, and, consequently, of increased background suppression at the cost of a reduced signal acceptance. The LLH threshold values depend on the E_T - η bins, and are defined to have a rather constant background efficiency with decreasing levels on the signal efficiency.

On top of the thresholds on the discriminant values, each working point has extra re-

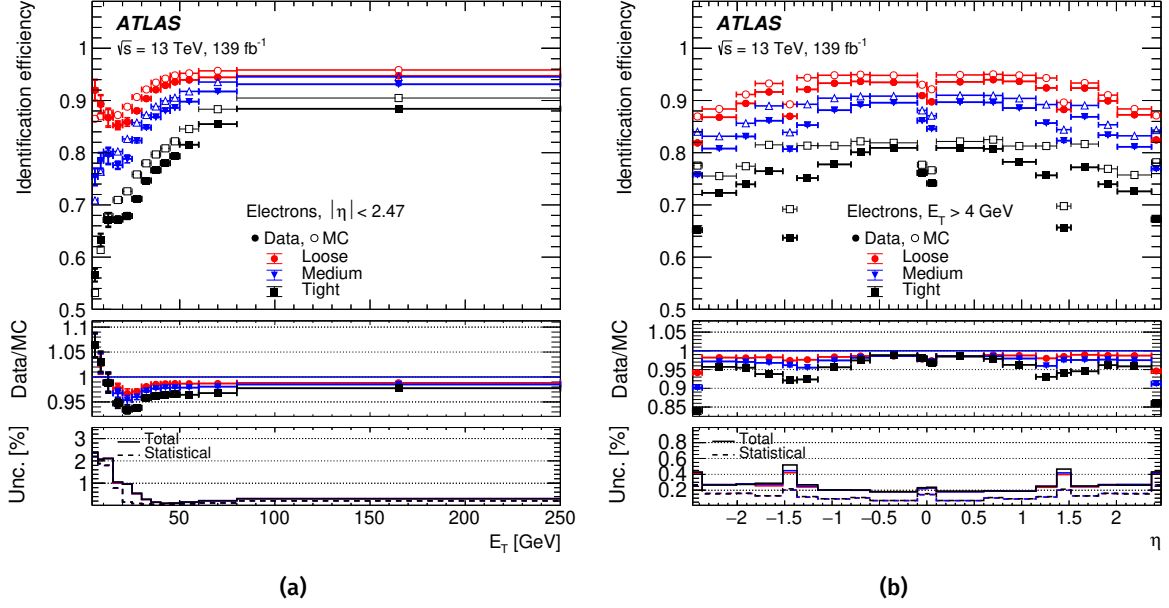


Figure 4.3: Electron identification efficiencies (top) measured as a function of (a) transverse energy E_T and (b) pseudorapidity η_{BE2} in both data (filled markers) and MC simulated events (open markers). The measurement was performed in data recorded by the ATLAS experiment in the full Run 2 data-taking period (2015–2018) from pp collisions at $\sqrt{s} = 13$ GeV, corresponding to an integrated luminosity of 139 fb $^{-1}$. Three working points from the likelihood identification algorithm were considered, namely, TIGHT, MEDIUM and LOOSE, with an extra B-Layer requirement on the latter. The ratio of the data efficiencies with respect to the MC ones, referred to as scale factors, is shown (middle) together with its total and statistical uncertainties (bottom). Taken from Ref. [266].

quirements on the number of hits in the Pixel and SCT inner detectors. For the LOOSE, MEDIUM and TIGHT working points, values of $n_{\text{Pixel}} > 2$ and $n_{\text{Pixel}} + n_{\text{SCT}} > 7$ must be satisfied, with one of the Pixel hits happening at the insertable B-layer for MEDIUM and TIGHT. Moreover, a variation of the LOOSE working point is also defined, namely the LOOSEANDBLAYER working point, which uses the same discriminant threshold for LOOSE, but adds a requirement of at least one hit in the insertable B-layer. This extra requirement is done to reduce the background coming from converted photons that occur early in the detector. Also, only one hit in the pixel detector is required on top of the VERYLOOSE working point, in particular, since it is defined to provide a more relaxed identification requirement for background studies.

The efficiency of the LLH identification working points are measured in both experimental data and MC simulated events. These quantify the rate of electrons that pass each selection criteria from a sample of signal electrons. From the data-to-simulation ratios of these efficiencies, referred to as scale factors, corrections to the MC simulation can be obtained to better reflect the detector performance in data.

The latest measurement^[266] of these efficiencies is presented in Fig. 4.3, where the full Run 2 data-taking period (2015–2018) was considered, corresponding to an integrated luminosity of 139 fb $^{-1}$. A clear decrease in the efficiencies is observed for lower values of E_T , while their

values remain symmetric across η ⁽²³⁾. Also, we note that data efficiencies are slightly lower than MC efficiencies for the TIGHT working point, with a better agreement between the two for the LOOSE working point. Overall, the data/MC ratios remain close to unity within 5%.

On top of the identification procedure, most studies performed within the ATLAS experiment also need the electron signature to not overlap with any other objects registered by the detector. This is quantified by the isolation procedure, discussed in the following section.

4.4 Electron isolation

When investigating processes involving electrons with the ATLAS detector, most of the time we are only interested in those produced at the interaction point. As previously discussed, the identification algorithm is optimised to discriminate these signal electrons from the misidentified hadrons mimicking its signature, as well as the electrons from photons and heavy-quark decays. Since these backgrounds mostly originate from hadronic processes, they will be produced within a jet and thus many energy deposits will be found around the electron candidate. As a result, the activity in a $\Delta\eta \times \Delta\phi$ region surrounding the electron candidate is a great metric to discriminate signal electrons, a concept referred to as *isolation*. Such a concept is particularly useful when investigating electrons produced in busy environments with many particles in the final state, such as the $t\bar{t}$ production, or when boosted particles decay into collimated electron–positrons pair^[245].

To that end, specific variables are defined to quantify the isolation of both the track in the inner detector and the energy deposits in the calorimeter for a given electron candidate. The vicinity of the object is quantified in the form of a cone of radius $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ around the direction of the candidate, defined from either the track or the energy cluster barycentre, and excluding the candidate itself^[245].

For the calorimeter-based isolation, the variable $E_T^{\text{cone}X}$ is defined using the method illustrated in Fig. 4.4. It is given by the sum of the transverse energy of all clusters that have their barycentre within a cone of size $\Delta R = X/100$, centred at the barycentre of the corresponding cluster associated to the electron candidate. In most applications, X is set to 20^[266]. During the estimation of this quantity, the energy of the candidate and contributions from pile-up activity⁽²⁴⁾ are subtracted^[245].

Likewise, the track-isolation variable $p_T^{\text{cone}X}$ is defined as the scalar sum of the transverse momenta of tracks within the cone of size ΔR previously defined, centred at the direction of the candidate’s track. In this case, X is usually taken to be 20 or 30^[266]. Moreover, only tracks that satisfy minimum quality requirements and originate from the primary vertex⁽²⁵⁾ of the

(23) We note that this η is obtained from the barycentre of the energy cluster in the second layer of the EM calorimeter. For this reason, it is also denoted in this work as η_{BE2} .

(24) As discussed in Section 3.2, pile-up refers to the additional pp collisions that happen together when two bunches cross the detector.

(25) Primary vertex refers to the one with the largest scalar sum of transverse momenta of the associated tracks.

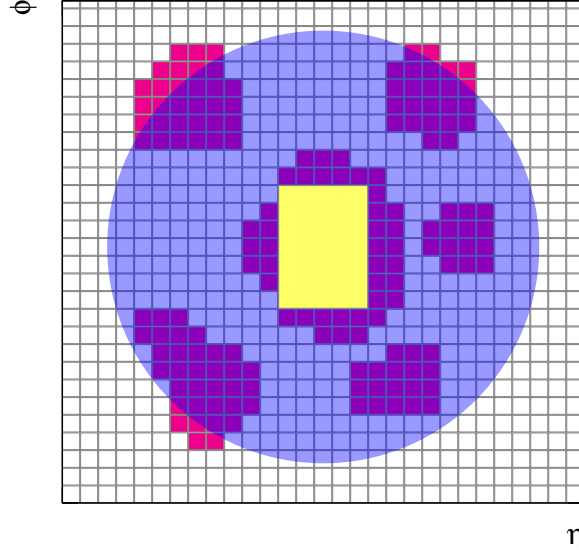


Figure 4.4: Schema of the calorimeter isolation method, where the grid represents the second-layer calorimeter cells in the $\eta \times \phi$ space. The energy deposits (red patches) that have their barycentre within the isolation cone (purple circle) are included in the computation of the isolation variable, after removal of part of the cluster associated to the electron candidate (yellow rectangle). Taken from Ref. [245].

event are considered^[268]. An alternative track-isolation variable is also defined to accommodate scenarios where the electrons are produced in cascade decays of high-momentum heavy particles^[266]. This $p_T^{\text{varconeXX}}$ quantity is obtained with a variable cone size, that gets narrower as the transverse momentum of the object increases, that is,

$$\Delta R = \min\left(\frac{10}{E_T}, \Delta R_{\text{max}}\right),$$

where E_T is given in GeV and ΔR_{max} is the maximum cone size accepted.

Similar to what is done for the identification, working points are also defined from fixed requirements on the calorimeter or the track isolation variables^[266], as listed in Table 4.2. These working points are then used on top of the identification ones to complete the identification procedure in ATLAS studies. For the TIGHTTRACKONLY_FIXEDRAD working point, in particular, a variable cone size is used at low energies ($E_T < 50$ GeV) to increase the rejection of hadronic activity.

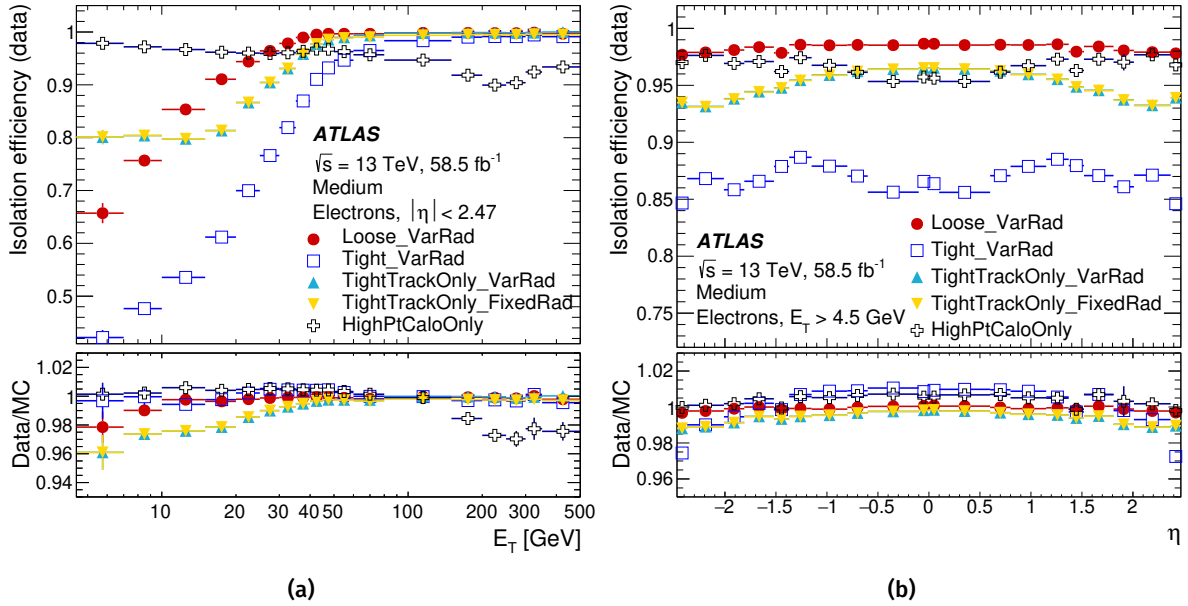
In this context, we present in Fig. 4.5 the electron isolation efficiencies latest measurements^[266], obtained in the 2018 data-taking period of Run 2 that corresponds to 58.5 fb^{-1} . Both data and MC efficiencies are presented, together with the data/MC ratios, as functions of the electron E_T and η . The five isolation working points defined above are considered, with electrons being required to pass the MEDIUM likelihood identification working point⁽²⁶⁾.

It is clear that the TIGHT_VARRAD isolation criteria presents the highest background rejection.

(26) Measurements were also performed for the LOOSE and TIGHT identification working points, but they are not presented here for simplicity.

Table 4.2: Isolation working points, defined with respect to the calorimeter-based and track-based isolation variables^[266].

Working point	Calorimeter isolation	Track isolation
HIGHPTCALOONLY	$E_T^{\text{cone20}} < \max(0.015p_T, 3.5 \text{ GeV})$	
TIGHTTRACKONLY_VARRAD		$p_T^{\text{varcone30}}/p_T < 0.06$
TIGHTTRACKONLY_FIXEDRAD		$p_T^{\text{varcone30}}/p_T < 0.06$ ($E_T < 50 \text{ GeV}$) $p_T^{\text{cone20}}/p_T < 0.06$ ($E_T > 50 \text{ GeV}$)
TIGHT_VARRAD	$E_T^{\text{cone20}}/p_T < 0.06$	$p_T^{\text{varcone30}}/p_T < 0.06$
LOOSE_VARRAD	$E_T^{\text{cone20}}/p_T < 0.20$	$p_T^{\text{varcone30}}/p_T < 0.15$

**Figure 4.5:** Electron isolation efficiencies (top) measured as a function of (a) transverse energy E_T and (b) pseudorapidity η . The measurement was performed in data recorded by the ATLAS experiment in the Run 2 2018 data-taking period from pp collisions at $\sqrt{s} = 13 \text{ GeV}$, corresponding to an integrated luminosity of 58.5 fb⁻¹. The five isolation working points are considered, with electrons being required to pass the MEDIUM likelihood identification working point. The ratio of the data efficiencies with respect to the MC ones, referred to as scale factors, is shown (bottom). Taken from Ref. [266].

tion for lower energies ($E_T < 60$ GeV), while for higher energies ($E_T > 100$ GeV) `HIGHPTCALOONLY` is preferred for this task. We note symmetric efficiencies across η for all working points, with `TIGHT_VARRAD` having significant differences along the full range. Lastly, the data/MC ratios are overall close to unity, with differences ranging below 5%–7%.

For the whole electron reconstruction, identification and isolation procedures, it is crucial to properly estimate the efficiency of each one of these steps. In particular, we want to quantify how well each algorithm is capable of accepting true electrons and with what precision we know this rate. By properly quantifying the detector performance, we can properly calibrate the detector simulation to reflect this performance, as well as gauge areas of improvements for all the different methods being used. Moreover, the ratios of the efficiencies measured in data and MC simulated events allow for the correction of the latter, known to suffer from mismodelling, as discussed in Section 3.2.6. In this context, the following chapters are dedicated to the improvements introduced by this work to both the electron identification algorithm and its efficiency measurements. We start with the method used to measure the electron efficiencies and scaling factors, together with new results obtained for the identification efficiencies, in particular, which are the subject of the next chapter.

Chapter 5

Measurement of electron identification efficiency

We demand rigidly defined areas of doubt and uncertainty.

Douglas Adams^[275]

5.1	Motivation for efficiency and scale factors measurement	80
5.2	Tag-and-probe method	82
5.3	Electron identification efficiencies	83
5.4	The Z -mass method	85
5.4.1	Event selection and input preparation	86
5.4.2	Estimation of the background with a normalisation in the mass tail	90
5.4.3	Estimation of the background with a likelihood fit	93
5.4.4	Improvements with respect to the previous implementation	96
5.4.5	Efficiency and scale factors	98
5.4.6	Systematic variations	100
5.5	Results	101
5.5.1	Run 3 (2022)	102
5.5.2	Run 2 (2018)	104
5.6	Future improvements	111

As discussed in Chapter 4, the electron identification algorithm is responsible for discriminating between the electrons in which we are interested, that is, those coming from the interaction point, and those that are non-prompt or actually other objects with similar detector signatures, such as charged hadrons or electrons decaying from jets. Quantifying the efficiency of this method is an important metric for assessing the systematic uncertainty of the detector and properly correcting the MC simulated events to match the detector performance. In this chapter, we present measurements of the electron identification efficiency in the ATLAS experiment for the Run 2 data, from 2015 to 2018, and for the early Run 3 data from 2022, which both represent original contributions from the author of this thesis. Moreover, we describe the implementation and improvements made to the software used to perform these measurements.

5.1 Motivation for efficiency and scale factors measurement

The electron reconstruction and identification algorithm described in Chapter 4 is used by all ATLAS analyses involving electrons. As such, the proper estimation of the systematic uncertainties introduced by these methods impacts the precision of all the corresponding physics measurements performed. In particular, analyses that rely on the counting of the number of electrons produced are particularly sensitive to how efficient the electron identification algorithm is. This is the motivation behind the measurement of the efficiency with which electrons are identified in the detector^[245,266,268,274]. Ideally, we want all the electrons of interest that are produced at the interaction point to be properly identified as electrons. However, this is not always the case. Thus, we must make sure that we properly quantify how well these methods are performing. This is possible by measuring the electron identification efficiency and properly estimating the uncertainty we have on its value.

The precision with which these efficiencies are measured is a limiting factor for results that rely on the accurate detection of electrons. This is the case, for instance, for one of the latest ATLAS results where lepton-flavour universality was tested by comparing the decays of W -bosons into electrons or muons in top-pair events^[276]. This measurement of the ratio

$$R_{W^{e/\mu}} = B(W \rightarrow \mu\nu)/B(W \rightarrow e\nu),$$

quantifying the difference between the branching ratios $B(W \rightarrow \ell\nu)$, $\ell = e, \mu$, is the most precise to date, an achievement made possible greatly due to the improvements done to reduce the uncertainty on the electron efficiency uncertainties^[266].

Furthermore, the electron identification algorithm is known to suffer from the mismodelling in MC simulation events, as discussed in Section 3.2.6. These are known to exist from the simulation of tracking properties, calorimeter shower shapes and their isolation^[266]. In particular, the energy deposits in the calorimeter subsystems are known to be different between the simulated detector model and the real detector. This can be explained, for instance, by the presence of passive material in front of the electromagnetic calorimeter, which is imperfectly accounted for in the detector simulation^[277]. As a result, the MC simulated events must be corrected to better represent the experimental data they are intended to describe.

To address this issue, a correction is added to the MC simulated events to match what is observed in real data. This multiplicative factor, known as a *scale factor* (SF), is obtained from the ratio of the efficiency obtained in the experimental data ϵ_{data} with respect to the one obtained in the MC simulated data ϵ_{MC} , that is,

$$\text{SF} = \frac{\epsilon_{\text{data}}}{\epsilon_{\text{MC}}}.$$

Since different kinematics will impact the efficiency values, the SFs are obtained as a function of p_{T} and η . By introducing these SFs when processing MC simulated events, the final distribu-

tions better represent the behaviour that is observed in the experimental data. In particular, the definition of such ratio allows for the cancellation of residual effects coming from kinematic differences between the physics process used to obtain the efficiency measurements and the ones being studied in other data analyses. That is, even though the efficiency measurements are performed using objects coming from a particular process (namely the $Z \rightarrow e^+e^-$ decay), the corrections obtained from the SF values can be universally used by all processes studied in the ATLAS experiment⁽²⁷⁾.

As described in Chapter 4, the process of reconstructing and identifying an electron within the ATLAS detector relies on a series of sub-systems and procedures. To better quantify the impact of each one of these steps, the total electron efficiency ϵ is factorised into the measurement of reconstruction, identification, isolation and trigger efficiencies, that is,

$$\epsilon = \epsilon_{\text{EM-cluster}} \times \epsilon_{\text{reco}} \times \epsilon_{\text{ID}} \times \epsilon_{\text{iso}} \times \epsilon_{\text{trigger}}.$$

Each one of these factors is responsible for quantifying how well the given step in the reconstruction and identification chain is being performed. They are obtained individually by counting the number of electrons that satisfy the corresponding selection over the ones that passed the previous requirement,

$$\epsilon = \frac{N_{\text{EM-cluster}}}{N_{\text{all}}} \times \frac{N_{\text{reco}}}{N_{\text{EM-cluster}}} \times \frac{N_{\text{ID}}}{N_{\text{reco}}} \times \frac{N_{\text{iso}}}{N_{\text{ID}}} \times \frac{N_{\text{trigger}}}{N_{\text{iso}}}. \quad (5.1)$$

In summary, these efficiencies are defined as follows:

EM-cluster efficiency $\epsilon_{\text{EM-cluster}}$ reflects the probability to reconstruct the electromagnetic cluster of an electron. It is given by the ratio of reconstructed EM calorimeter clusters, $N_{\text{EM-cluster}}$, over all produced electrons, N_{all} ;

Reconstruction efficiency ϵ_{reco} reflects the quality of the track reconstruction and its matching to a cluster. It is calculated as the ratio of reconstructed electron candidates with a track, N_{reco} , over the number of EM-cluster candidates, N_{cluster} ;

Identification efficiency ϵ_{ID} reflects the efficiency of the menus designed to massively reject backgrounds at different signal efficiencies. It corresponds to the number of electrons that pass the identification criteria, N_{ID} , over all the reconstructed electron candidates, N_{reco} ;

Isolation efficiency ϵ_{iso} reflects the probability that an identified electron passes track and calorimeter isolation requirements. It is calculated as the ratio of electrons satisfying both identification and isolation requirements, N_{iso} , over those that pass the identification requirements, N_{ID} ;

(27) We also note that, when performing the efficiency measurements, no distinction is made between the charges of the electron (e^-) and the positron (e^+). As such, unless otherwise stated, any references to electrons in this work also includes positrons.

Trigger efficiency $\epsilon_{\text{trigger}}$ is the rate at which an electron passing all the previous requirements will also satisfy the online requirements of a specific trigger. It is given by the number of triggered (also reconstructed, identified and isolated) electron candidates N_{trigger} divided by N_{iso} ;

Except for $\epsilon_{\text{EM-cluster}}$, which is measured entirely from MC simulation^[268], all these efficiencies are estimated directly in data. To do so, a high-purity sample of electrons is needed, but it must not be biased by the algorithm of which the efficiency we are trying to quantify. In our case, we will focus on the identification efficiency, meaning we need a sample of electrons selected without using the identification algorithm. Such a selection is obtained using the tag-and-probe method, described in the following section.

5.2 Tag-and-probe method

The goal of the *tag-and-probe* method is to obtain an ensemble of objects that we are confident correspond to a given type without using any identification algorithm. To measure electron efficiencies, for instance, we need to have a sample of true electrons without using the identification methods designed to distinguish electrons from their backgrounds. Thus, we take advantage of well-known resonances in the SM, such as the $Z \rightarrow e^+e^-$ and $J/\psi \rightarrow e^+e^-$ decays, where we know the final products are electron pairs.

The procedure consists of selecting one of the electrons in the pair using strict requirements, in such a way that we are confident it is a true electron. This object is referred to as the *tag*. The second electron, referred to as the *probe*, will populate our unbiased dataset since it will have passed the EM-cluster and reconstruction procedures but not the identification requirements for which we want to measure the efficiency. By requiring the invariant mass of the tag and probe pair, m_{ee} , to be close to the Z boson mass, for instance, we will maximise the chances that the probe corresponds to an electron as well, without using any identification algorithm.

In this way, the efficiency of a given selection is obtained by counting the number of probes that pass our selection of interest (e.g., an identification criteria) among all the probes obtained in the dataset. It so happens that in some events two objects satisfy the more strict tag requirements. In such cases, both of them are considered separately as tag, allowing the other to be counted as a probe, a process referred to as *cross-probing*.

The tag-and-probe method used to measure electron efficiencies is implemented using the ATLAS software suite^[248], which is used for the reconstruction of events and processing of both simulation and experimental data. Presented in Section 3.2.5, this library is regularly updated to incorporate the latest developments to the event-processing scheme. An import-

ant release update⁽²⁸⁾ was performed before the start of Run 3, requiring a completely new implementation of the whole electron efficiency calculation software. As a result, the previous software, used to obtain the most precise measurements over the Run 2 period^[266], for instance, was replaced with a new framework, the TAGANDPROBE^[278] software suite. This new implementation was designed to perform, among other things, the efficiency measurements for the different reconstruction, identification and isolation algorithms.

Later in this chapter, we will present the measurements performed with this new software, in particular the identification efficiencies ϵ_{ID} previously defined. Before doing so, a description of this efficiency calculation is presented.

5.3 Electron identification efficiencies

As discussed in the previous chapter (Section 4.3.1), the identification of electrons in ATLAS is currently performed using a likelihood approach. This method defines different working points, namely the TIGHT, MEDIUM, LOOSE and VERYLOOSE working points, each one of them to increasing levels of signal efficiencies. This means that a sample of objects that pass the TIGHT working point will be mostly populated by true electrons, but many true electrons will be left out of the sample. Alternatively, a selection of objects that satisfy the LOOSE requirement will contain almost all electrons available, but also be contaminated by many background objects.

The efficiency of each one of these identification working point using the tag-and-probe method is measured as

$$\epsilon_{\text{ID}} = \frac{\text{Number of probes that pass the identification criteria}}{\text{Total number of probes}}, \quad (5.2)$$

such that efficiencies close to unity correspond to the ideal scenario. Since the measurement is performed with respect to reconstructed electrons, all probes are required to have a good quality track matched to an EM cluster. Moreover, these ratios are measured as a function of the transverse momentum p_{T} ⁽²⁹⁾ and the pseudorapidity η_{BE2} , the latter obtained from the barycentre in the second layer of the EM calorimeter. This layer is where the majority of the electron energy is deposited, such that it is known to give a better proxy for the location of the cluster of an electron in the EM calorimeter.

The identification efficiencies are measured in bins starting from $p_{\text{T}} \geq 4.5$ GeV. For energies of $p_{\text{T}} \geq 15$ GeV, the Z boson resonance is used to extract the probe electrons. Conversely,

(28) The ATLAS software used in the last published efficiency results^[266] was in its major version 21, and it is referred in this work as *release 21*. Now, its update to major version 22 allowed, among other things, compatibility with Python version 3, instead of the previously used version 2. In this work, other major versions of the ATLAS software were also used, namely versions 22–25. For simplicity, all these versions are taken to be equivalent and referred to as *release 22*.

(29) Since the mass of the electron is small when compared to its momentum, $E_{\text{T}} \approx p_{\text{T}}$. As such, the electron transverse momentum and transverse energy have the same value and are used interchangeably in ATLAS efficiency measurements. In this work, we have used mainly the transverse momentum p_{T} .

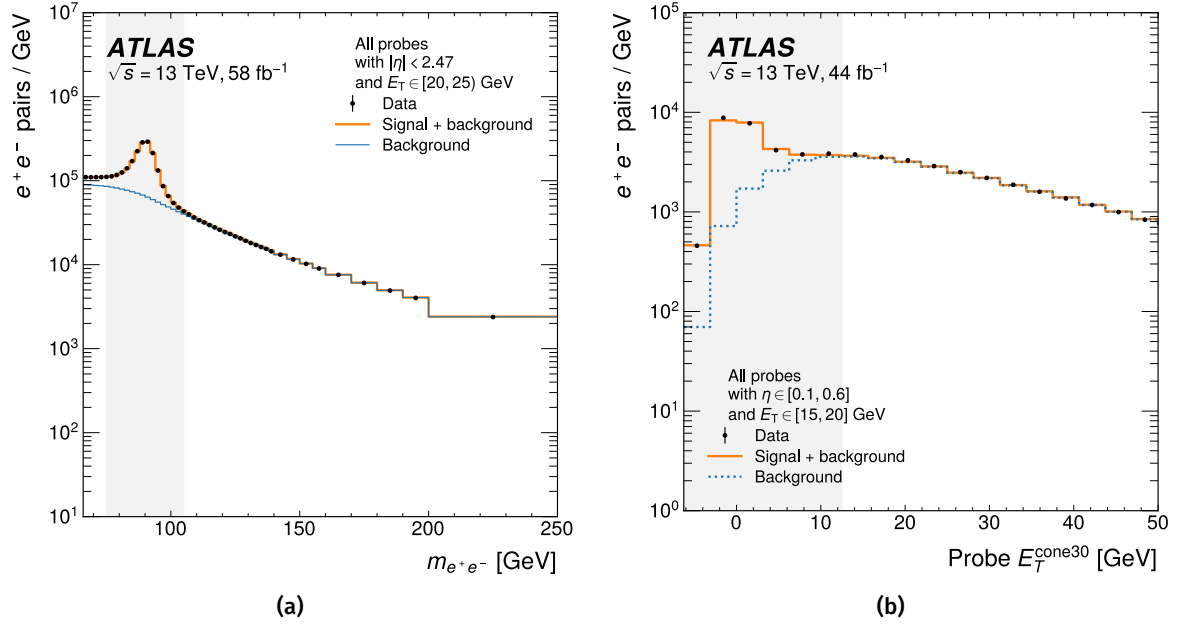


Figure 5.1: Examples of histograms used to estimate the background in the (a) Z -mass and (b) Z -iso methods. In the Z -mass method, the invariant-mass distributions of the electron pair m_{ee} is used. In the Z -iso method, the probe E_T^{cone30} isolation variable is used, defined as the sum of the transverse energy deposited in the calorimeters in a cone of $\Delta R \leq 0.30$. The signal regions used in the efficiency measurements, which correspond to the Z peak, are shown in gray. Taken from Ref. [266].

given the low statistics observed at lower energies, such as $4.5 \text{ GeV} < p_T < 20 \text{ GeV}$, the J/ψ resonance is more appropriate to the task in these ranges, with a mass of $m_{J/\psi} = 3.1 \text{ GeV}$ ^[43]. Both approaches are performed in the $15 \text{ GeV} < p_T < 20 \text{ GeV}$ window, where the results are cross-checked and combined^[245].

Furthermore, we know that among the selected probes there will be contamination from hadrons or electrons from semi-leptonic decays of heavy-flavour hadrons or photon conversions. Since these objects are not electrons produced at the interaction point, the ones we are interested, they should not be included in our efficiency calculation. Estimating the amount of these background objects among the probe electrons is one of the major challenges of the efficiency measurement.

To this end, two approaches have been developed: the Z -mass method, which uses the invariant mass of the electron pair m_{ee} to estimate the background, and the Z -iso method, which uses the energy deposits in the isolation cone of the probe electron^[245,266,268,274]. Examples of the distributions used for both these methods are shown in Fig. 5.1. For lower energies, in the range $[1.8, 4.6] \text{ GeV}$, the background estimation is done with the invariant mass distributions of the $J/\psi \rightarrow e^+e^-$ decay. Although only prompt $J/\psi \rightarrow e^+e^-$ processes produce isolated electrons, we also have non-prompt J/ψ production from the decay of b quarks, which are considered background^[266]. To suppress this non-prompt contribution, a requirement is made on the pseudo-proper time variable t_0 , defined as a function of the displacement of the J/ψ

vertex, its mass and its transverse momentum, since smaller values are associated with the prompt contribution^[245]. All these methods are combined, accounting for the statistical and systematic correlations between them⁽³⁰⁾.

For the measurements presented in this work, the Z -mass method was used. Its description and upgrades with respect to the previous iterations of these results are described in the following section.

5.4 The Z -mass method

The invariant mass distribution of the electron pair m_{ee} forms a clear peak around the Z boson mass, at $m_Z = 91 \text{ GeV}$ ^[43], as shown in Fig. 5.1a. Thus, we define the *signal region* (SR) of the measurement as a given range in the mass spectrum that includes the Z peak. It is the ratio of signal electrons in the SR that correctly pass the identification requirement, S_{ID} , among all the probe electrons that satisfy the reconstruction requirements, S_{probes} , that is used to measure the electron identification efficiency; that is

$$\epsilon_{\text{ID}} = \frac{S_{\text{ID}}}{S_{\text{probes}}} \Big|_{m_{ee} \in \text{SR}} = \frac{N_{\text{ID}} - B_{\text{ID}}}{N_{\text{probes}} - B_{\text{probes}}} \Big|_{m_{ee} \in \text{SR}}, \quad (5.3)$$

where N_{probes} and N_{ID} are the total number of electron candidates and those passing the identification requirement, respectively, while B_{probes} and B_{ID} are the respective number of objects that are not signal electrons and thus correspond to the background. These quantities are obtained as the integral of the corresponding distributions over the SR. When measuring the efficiency in MC simulated events, all events are properly selected to correspond to true electrons, so there is no background present. However, this is not the case when performing the measurement in data, where objects such as light-flavour mesons or electrons coming from a heavy flavour decay can mimic the electron signature in which we are interested.

The main challenge of the Z -mass method is to obtain this *background estimate* (solid blue line in Fig. 5.1a), from which the B_{probes} and B_{ID} values can be extracted. In this particular method, these background distributions are derived from data using m_{ee} as the discriminant variable. For this purpose, invariant mass histograms are built from objects satisfying the signal selections for both numerator and denominator. The ones obtained from data are directly used to estimate the number of events N_{ID} and N_{probes} . Histograms are also built from objects that satisfy a selection built to accept mostly background events using both data and MC, referred to in this work as *background templates*.

(30) The combination of J/ψ and Z methods is easier since the events used are different. The combination of the two Z methods, however, is a bit more complex, since they rely on the same events. For instance, since both methods use the same signal definitions, the statistical uncertainty of the signal are correlated. On the other hand, since the background estimation is based on different control regions, the background statistical uncertainties are uncorrelated.

Two approaches have been developed to obtain the background estimate. A first approach normalises the background template obtained from the data to the signal histogram in a mass region far from the Z peak, namely $m_{ee} \in [120, 250]$ GeV^[279]. A more sophisticated approach, on the other hand, fits the background templates to the signal data using a full maximum-likelihood fit, a method that was used in the latest measurements^[280]. Being easiest to implement, the first method was used to perform measurements with the early Run 3 data available in 2022. Then, once the latter was fully implemented, it was used to perform measurements in the full Run 2 data, including improvements with respect to the latest published results^[266]. We emphasise that the implementation of the first method together with these two measurements (early Run 3 data and full Run 2 data) correspond to an original contribution of the author of this thesis.

For the measurement to properly reflect the efficiency of the detector, we need to build reliable histograms. In particular, the background templates will be populated by signal events that must be removed. Furthermore, the uncertainties on the measurement need to be accounted from both the statistical size of the sample used and those introduced by detector effects. As such, *systematic variations* of the measurement are performed, where different elements of the analysis are modified. While the central value and its statistical uncertainties are obtained from the nominal variation of the measurement, its systematic uncertainty is estimated as the difference between the variation values and the nominal one, added in quadrature.

In what follows, the formalism of the Z -mass method is presented, covering the event selections used, the background estimation procedure, as well as the efficiency and SF calculation. Finally, a summary of the improvements added with respect to the previous implementation is also presented. We note that the implementation of the event selection and input preparation, which will be described in Section 5.4.1, as well as the first iteration of the background estimation, which will be described in Section 5.4.2, were performed by the author of the thesis. The implementation of its second iteration, the method described in Section 5.4.3, including its improvements with respect to the previous implementations of this method, was performed by others. Lastly, the measurements presented on this thesis, presented in Section 5.5, also correspond to an original contribution from the author of this thesis.

5.4.1 Event selection and input preparation

Datasets containing events recorded by the ATLAS detector for the full Run 2 period (2015–2018) and the early Run 3 period (2022) were used for the measurements presented in this work. Events from MC simulation were also generated and processed through the detector response using the ATLAS simulation infrastructure^[224], presented in Section 3.2.6. Both simulated and experimental data were then reconstructed using the ATLAS software^[248]. A first filter to keep only relevant events was performed, such that the input datasets were populated mainly by electrons from the $Z \rightarrow e^+e^-$ process.

The tag-and-probe method was then used to obtain samples of unbiased electron candidates for the measurement using the selections listed in Table 5.1 and Table 5.2. Only events that passed minimum reconstruction quality requirements were kept. Moreover, these events had to be originated from a primary vertex and pass one of the single-electron triggers, listed in Table 5.3.

For each event, a tag and probe pair was identified using the requirements also listed in Table 5.1⁽³¹⁾. Both objects were required to satisfy basic data-quality requirements and had its energy and resolution corrected. Moreover, the reconstruction algorithm that defines the electron candidates also defines ambiguous candidates that could be potentially either a photon or an electron^[245]. For this reason, both candidates considered as electron or ambiguously a photon or an electron are included. For MC simulated events, a selection at truth-level guaranteed that only electrons originating from a Z boson and considered as prompt were kept.

An object was accepted as a tag if it satisfied kinematic and track isolation criteria that guaranteed it was a prompt electron. In particular, this means they had $p_T \geq 27$ GeV and $|\eta_{\text{BE2}}| \leq 2.47$, with $|\eta_{\text{BE2}}| \notin [1.37, 1.52]$ such that they were not reconstructed in the transition region between the barrel and end-cap calorimeter. Also, a small value was required for the sum of the p_T from the tracks in a cone having radius $\Delta R = 0.20$ over p_T , namely $p_T^{\text{cone20}}/p_T \leq 0.1$. Also, tag objects had to satisfy requirements concerning the transverse and longitudinal impact parameter, d_0 and z_0 , respectively, recorded by the inner detector. More precisely, the maximum absolute transverse impact parameter significance of the electron primary track was required to be $|d_0/\sigma(d_0)| \leq 5.0$, while $|z_0 \sin \theta| \leq 0.5$ selected a small value for the maximum absolute longitudinal impact parameter (with respect to the hard-scattering primary vertex) times the sine of the scattering angle θ of the electron primary track, given in millimetres. Furthermore, the tag needed to match to a single electron trigger (Table 5.3) and also pass the TIGHT LLH identification working point.

On the other hand, the probe was required to have $p_T \geq 15$ GeV and $|\eta_{\text{BE2}}| \leq 2.47$, corresponding to the kinematic region of interest for the measurement. Moreover, it had to satisfy a jet veto that guaranteed an electron and a jet with $E_T \geq 20$ GeV would not overlap in a region with $\Delta R \leq 0.4$, as well as good track quality with the number of hits in the Pixel and SCT detectors satisfying $n_{\text{Pixel}} \geq 1$ and $n_{\text{Si}} \geq 8$, respectively. Furthermore, the sum of the charge of the tag and the probe were required to be zero, such that we are confident they have opposite sign (OS) and thus originate from the neutral Z boson. Events where the charge sum corresponded to either -2 or $+2$ were also used in a separate selection, since these same sign (SS) events are used when estimating the background, as will be discussed in Section 5.4.2.

From these selections, we are confident the objects satisfying the probe criteria corres-

(31) We note that for the Run 3 measurements, some of the selections were different. In particular, the tag was required to have $p_T \geq 25$ GeV and the track quality cuts had $n_{\text{Si}} \geq 7$. The values in Table 5.1 represent minor improvements made when performing the more recent Run 2 measurements, and, as such, represent the state-of-the-art values.

Table 5.1: Selections performed to construct a sample of electrons unbiased by the identification algorithm via the tag-and-probe method.

Event selection	
Event cleaning (data only)	
Included in good-run lists (GRL)	
Primary vertex selection	
Matched to a single electron trigger (see Table 5.3)	
Tag selection	Probe selection
Pass detector data-quality criteria	Pass detector data-quality criteria
Apply electron energy scale and resolution corrections	Apply electron energy scale and resolution corrections
Accepted authors: Electron, Ambiguous	Accepted authors: Electron, Ambiguous
Matched MC truth criteria (pre-2020) for electrons from the decay of a Z boson that are considered prompt (MC only)	Matched MC truth criteria (pre-2020) for electrons from the decay of a Z boson that are considered prompt (MC only)
$p_T \geq 27$ GeV	$p_T \geq 15$ GeV
$ \eta_{\text{BE2}} \leq 2.47$ and $ \eta_{\text{BE2}} \notin [1.37, 1.52]$	$ \eta_{\text{BE2}} \leq 2.47$
Matched to a single electron trigger (see Table 5.3)	Isolated from AntKt4 jet with $E_T \geq 20$ GeV within $\Delta R \leq 0.4$
$ d_0/\sigma(d_0) \leq 5.0$ and $ z_0 \sin \theta \leq 0.5$	Good track quality ($n_{\text{Pixel}} \geq 1$ and $n_{\text{Si}} \geq 8$)
Pass TIGHT LLH identification	
$p_T^{\text{cone20}}/p_T \leq 0.1$	
Multi-object requirement	
OS selection: sum of tag and probe charges equal to 0	
SS selection: sum of tag and probe charges equal to -2 or $+2$	
Numerator selection	
Pass probe selection	
Pass corresponding LLH identification (e.g. LOOSE, MEDIUM or TIGHT)	
$ d_0/\sigma(d_0) \leq 5.0$ and $ z_0 \sin \theta \leq 0.5$	

Table 5.2: Selections performed to define background objects.

Background selection in the normalisation in the mass tail method			
Variation 1		Variation 2	
$p_T < 30$ GeV	$p_T \geq 30$ GeV	$p_T < 30$ GeV	$p_T \geq 30$ GeV
Fail ULTRA LOOSE LLH $E_T^{\text{cone30}}/p_T \geq 0.02$	Fail VERY LOOSE LLH $E_T^{\text{cone40}}/p_T \geq 0.05$	Fail ULTRA LOOSE LLH $E_T^{\text{cone30}}/p_T \geq 0.02$	Fail VERY LOOSE LLH $E_T^{\text{cone40}}/p_T \geq 0.20$

Background selection in the likelihood fit method		
Nominal	Variation 1	Variation 2
Pass probe selection	Pass probe selection	Pass probe selection
Fail LOOSE LLH	Fail VERY LOOSE LLH	Fail VERY LOOSE LLH
$E_T^{\text{cone30}}/p_T \geq 0.01$	$E_T^{\text{cone30}}/p_T \geq 0.01$	$E_T^{\text{cone40}}/p_T \geq 0.07$

Table 5.3: Single-electron triggers used in the ATLAS detector over the Run 2 (2015–2018) and the early Run 3 (2022) data taking periods.

Run 2	HLT_e24_lhmedium_L1EM20VH	2015
	HLT_e60_lhmedium	2015, 2016
	HLT_e120_lhloose	2015
	HLT_e26_lhtight_nod0_ivarloose	2016, 2017, 2018
	HLT_e60_lhmedium_nod0	2016, 2017, 2018
	HLT_e140_lhloose_nod0	2016, 2017, 2018
	HLT_e300_etcut	2016, 2017, 2018
Run 3	HLT_e26_lhtight_ivarloose_L1EM22VHI	2022
	HLT_e60_lhmedium_L1EM22VHI	2022
	HLT_e140_lhloose_L1EM22VHI	2022
	HLT_e300_etcut_L1EM22VHI	2022

ponding to electron candidates without passing any identification selection. Then, further requirements were used to define the signal and background regions needed for the efficiency measurement. Objects that satisfied all the probe requirements populated the denominator of the efficiency measurement. Among them, the numerator was obtained by the number of objects that passed each LLH identification working points being assessed, namely `TIGHT`, `MEDIUM` and `LOOSEANDBLAYER`⁽³²⁾. These signal objects were also required to satisfy the impact parameter requirements, $|d_0/\sigma(d_0)| \leq 5.0$ and $|z_0 \sin \theta| \leq 0.5$, since these are usually combined to the LLH identification working points in ATLAS analyses.

Lastly, background objects that contaminated the denominator selection were identified by checking if they also failed a given LLH identification requirement, while also having poor isolation. The latter was quantified as a large value of the ratio of the sum of transverse energy deposited in the calorimeter in a cone of $\Delta R \leq 0.30(0.40)$ with respect to the particle's transverse momentum, that is, $E_T^{\text{cone30(40)}}/p_T$. Two different approaches were then used to estimate the background within the Z -mass method in the Run 2 and Run 3 measurements presented in this work. Referred to as the *likelihood fit method* and the *normalisation in the mass tail method*, they use different definitions for the background selections, which are listed in Table 5.2.

The assessment of biases introduced by the choice of these selections is performed using systematic variations. That is, the whole efficiency measurement is performed once again with a modification in one or more requirements. A summary of the systematic variations used in the measurements reported in this work will be discussed in Section 5.4.6.

Once the selections have been performed, histograms of the invariant mass of the tag and probe objects are filled with a 1 GeV-binning for $m_{ee} \in [60, 250]$ GeV. These are obtained for different p_T and η_{BE2} values, listed in Table 5.4. Objects from both MC simulation

(32) As described in the previous chapter, this working point is a combination of the `LOOSE` working point with an additional requirement of at least one hit in the insertable B-layer of the inner detector.

Table 5.4: Binning on p_T and η_{BE2} .

Observable	Binning
p_T (GeV)	[15, 20, 25, 30, 35, 40, 45, 50, 60, 80, 150, 250]
η_{BE2}	[-2.47, -2.37, -2.01, -1.81, -1.52, -1.37, -1.15, -0.80, -0.60, -0.10, 0.00, 0.10, 0.60, 0.80, 1.15, 1.37, 1.52, 1.81, 2.01, 2.37, 2.47]

and experimental data passing the denominator and numerator signal selections populate the histograms which will be used to calculate the N_{probes} and N_{ID} values, respectively. In addition, background template histograms N_{template} are drawn from the objects that satisfy these corresponding selections from experimental data.

However, these background templates are also contaminated with signal events that must be accounted for. As a result, before being used to estimate the background events B_{probes} and B_{ID} needed for the measurement, these templates must be corrected to remove the contamination from signal events. Supporting histograms are then constructed to achieve this, such as objects that satisfy the numerator selection in data, but have a tag and probe pair with the same sign, $N_{\text{ID/SS}}$, as well as objects that satisfy the background requirements in MC simulated events instead, $N_{\text{template/MC}}$.

Having discussed the selection requirements used to obtain the invariant mass distributions of both signal and background templates, the next step is to properly estimate the background contribution in the data. In the next sections, we describe the procedure with which these histograms are used to estimate the background in both the normalisation in the mass tail and the likelihood fit to data versions of the Z -mass method.

5.4.2 Estimation of the background with a normalisation in the mass tail

As previously discussed, the efficiency is calculated with values of the integral of the invariant mass distribution of the signal events in a region around the Z mass. Referred to as the *integration region*, it corresponds to $m_{ee} \in [75, 105]$ GeV. Consequently, the number of background objects in this range must be obtained from an estimate of the background shape.

One way to estimate the background shape consists in scaling the background template obtained from the experimental data, using a region of the signal mass spectrum that we know is enriched in background. This region, referred to as the *normalisation region*, must not include the Z mass resonance, and thus is chosen to be in the higher-mass region, at the *tail* of the distribution, where $m_{ee} \in [120, 250]$ GeV.

The starting point of the procedure is the template distribution of the events that satisfy the background selections in data, $N_{\text{template/raw}}$. Even though the requirements were chosen to optimise the presence of only background events, some signal events might be present in these distributions. With that in mind, a process of cleaning these templates is performed by subtracting from it the distribution of the signal events in the template, S_{template} . This

shape is estimated by performing the same background selections on the MC simulated events, $N_{\text{template/MC}}$, which should have only true electrons by construction. This distribution is then scaled by a rough estimate of an SF obtained from the objects that satisfy the TIGHT LLH requirement. That is, the SF is obtained from the ratio of events in the integration region from the signal distribution in data with respect to the same value in MC, $N_{\text{TIGHT/data}}/N_{\text{TIGHT/MC}}$. As a result, the background template after the removal of the signal contamination is given by

$$N_{\text{template}} = N_{\text{template/raw}} - S_{\text{template}},$$

where

$$S_{\text{template}} = N_{\text{template/MC}} \times \left(\frac{N_{\text{TIGHT/data}}}{N_{\text{TIGHT/MC}}} \bigg|_{m_{ee} \in [75, 105] \text{ GeV}} \right).$$

Once we are confident the background template contains only background events, it still needs to be scaled in order to match the signal distribution from data. This is done by matching the number of events in the normalisation region of the template to those of a reference distribution from data. For the numerator, the reference distribution is taken to be the histogram obtained from the objects that pass the identification requirement, but had a SS selection, $N_{\text{ID/SS}}$. This choice is based on the fact that, far away from the Z mass resonance peak, this distribution should be dominated by background, while having the same order of magnitude as the background template in the OS selection^[273]. Thus, the numerator background estimate is given by

$$B_{\text{ID}} = N_{\text{template}} \times \left(\frac{N_{\text{ID/SS}}}{N_{\text{template}}} \bigg|_{m_{ee} \in [120, 250] \text{ GeV}} \right). \quad (5.4)$$

Now, for the denominator, ideally we could use as reference the signal distribution N_{probes} , since far from the peak we expect there will not be any signal electrons. Unfortunately, this is not true and these events must be removed from the number of background events found in the normalisation region. An estimate of this value is obtained using the efficiency of the TIGHT LLH working point, ϵ_{TIGHT} , which in its first iteration is calculated without this signal contamination subtraction. Hence, to obtain the number of signal events in the normalisation region, we assume that the objects satisfying the TIGHT requirement in data, divided by this ϵ_{TIGHT} estimate, correspond to only signal electrons, with the possible background electrons that are contaminating this number being neglected. As a result, the background estimate for the denominator is obtained from the expression

$$B_{\text{probes}} = N_{\text{template}} \times \left(\frac{N_{\text{probes}} - N_{\text{TIGHT}}/\epsilon_{\text{TIGHT}}}{N_{\text{template}}} \bigg|_{m_{ee} \in [120, 250] \text{ GeV}} \right). \quad (5.5)$$

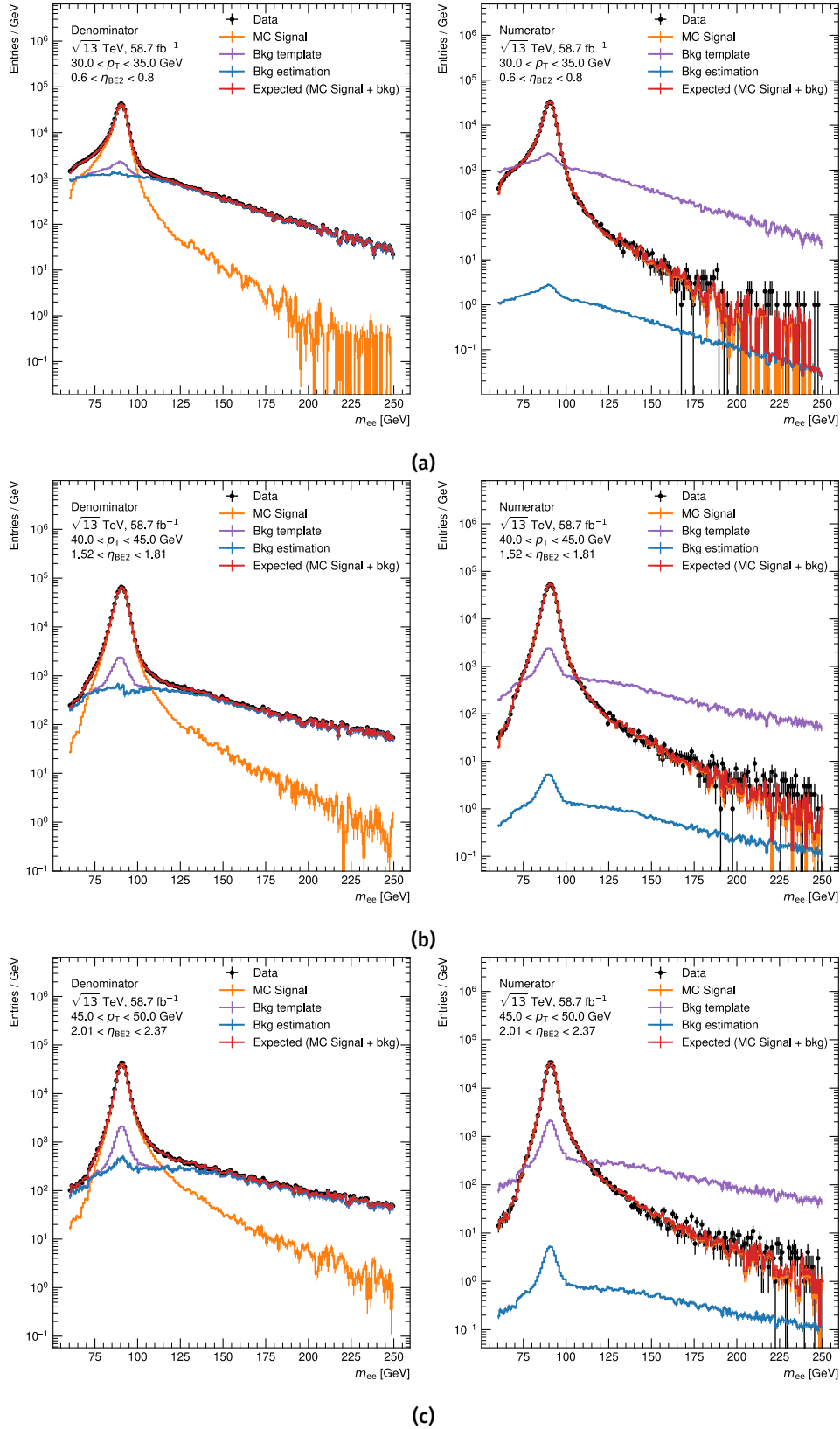


Figure 5.2: Invariant mass of the tag and probe pair m_{ee} for the denominator (left) and numerator (right) selections in selected p_T and η_{BE2} bins. The signal distributions obtained in data (black) and MC simulated events (orange) are drawn, together with the background template distribution (purple). The background estimate (blue) is obtained using the normalisation in the mass tail method, and it is combined to the MC signal shape to form the expected distribution (red).

Examples of the mass distributions obtained using this method are drawn in Fig. 5.2. We observe that, in some cases, this method allows for a good estimate of the background shape, properly accounting for the signal contamination under the peak, such as shown in Fig. 5.2a, where the background estimation shows no resonance peak for the denominator. However, we also note that sometimes the signal contamination removal in the background, obtained via the rough estimate of the SF using the TIGHT selection histograms, actually overestimates the number of signal events in the high-mass region^[281]. This generates a depletion in the background shape, which can be observed in Fig. 5.2b. Moreover, the signal contamination in the background is clearly underestimated in the numerator shapes, as well as in regions where the MC modelling is known to be problematic. This is the case of the bin before the transition region ($|\eta_{\text{BE2}}| \in [1.15, 1.37]$) and higher values with $|\eta_{\text{BE2}}| > 2.0$, as shown in Fig. 5.2c. Another limitation of the method presented in this section is that the choice of the normalisation region introduces a large systematic uncertainty, since either the low-mass or high-mass region could be considered, but their different levels of signal contamination largely affect the result^[266].

To address these issues and improve the background estimate, we can replace the simple normalisation used to obtain the signal template S_{template} and the background estimates B_{ID} and B_{probes} with a proper likelihood fit of the templates to the data. This method is described in the following section.

5.4.3 Estimation of the background with a likelihood fit

In this approach, the starting point of the background estimation remains the same: we use the mass distributions of objects satisfying the background selections obtained from data, $N_{\text{template/raw}}$. Now, as previously discussed, this distribution is expected to contain signal events that must be removed in order to not contaminate the measurement.

Similarly to what was done before, an estimate of the signal distribution S_{template} is obtained from scaling the histogram of the background selection when performed in MC events, $N_{\text{template/MC}}$. In order to reduce the dependence on the MC simulation^[266], the scaling factor is obtained using a likelihood fit, which is implemented using the RooFit package^[282]. The fit is performed for each p_{T} and η_{BE2} bin, and a model is constructed by adding this signal template obtained from MC together with a Bernstein polynomial shape to represent the background.

The choice of using this analytical function for the background allows for the flexibility to adjust the shape of the background as it varies according to the energy range in question. Since the shape of the background is highly dependent on the p_{T} of the electrons, different orders of polynomials are chosen according to the energy range. As shown in Fig. 5.3, the background changes from a concave shape at low energies with $p_{\text{T}} < 20$ GeV, assuming an S-shape at intermediate values of $p_{\text{T}} \in [20, 25]$ GeV, until finally taking a convex shape for $p_{\text{T}} > 25$ GeV. For this reason, a linear polynomial is used instead for this transition region $p_{\text{T}} \in [20, 25]$ GeV, while a second-order polynomial is used everywhere else.

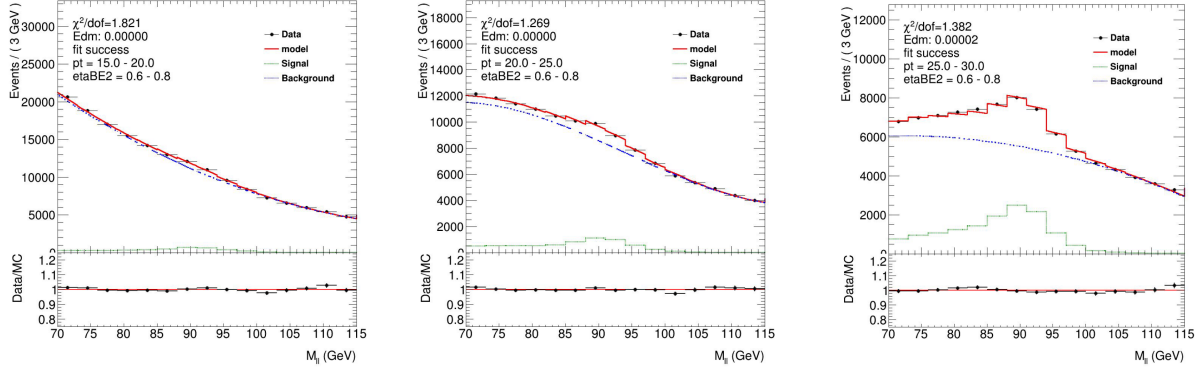


Figure 5.3: Evolution of the background estimation shape obtained via the likelihood fit method as a function of the transverse momentum p_T , in particular for $p_T \in [15, 20]$ GeV (left), $p_T \in [20, 25]$ GeV (center) and $p_T \in [25, 30]$ GeV (right).

Since this background approximation is not valid for the whole mass spectrum, the fit is restricted to the region under the Z peak, such that all histograms are rebinned according to the following mass binning:

$$\text{Signal contamination: } [70, 73, 76, 79, 82, 85, 88, 91, 94, 97, \\ 100, 103, 106, 109, 112, 115] \text{ GeV.}$$

The model combining the background polynomial shape and the signal template from MC is then fitted to match the signal denominator distribution obtained from data, $N_{\text{probes/data}}$. As a result, the background template after removal of the contamination from signal events is given by

$$N_{\text{template}} = N_{\text{template/raw}} - S_{\text{template}},$$

where

$$S_{\text{template}} = N_{\text{template/MC}} \times \left(\text{scale from fit where } \begin{cases} \text{model} = N_{\text{polynomial}} + N_{\text{template/MC}} \\ \text{data} = N_{\text{probes/data}} \end{cases} \right).$$

Once the background templates have had the signal events removed, they can be used to obtain the background estimate. To do so, they are scaled according to another fit performed once again by the `RooFit` package. In this fit, numerator and denominator are treated independently, using the corresponding signal distributions from data and MC, as well as the shared background distribution from data. These histograms are rebinned to avoid any bias that the Z mass peak could introduce in the shapes, such that the following binnings are used:

Numerator: [65, 80, 100, 250] GeV and

Denominator: [65, 70, 75, 80, 100, 120, 150, 250] GeV.

We note that the choice to have more bins below and above the peak for the denominator is a way to guarantee the influence of the shapes in the tails on the final fit^[281].

For the fit procedure, the model used is the combination of the background template with a signal model, where the latter corresponds to the distribution in question (numerator or denominator) obtained from MC simulation events. This model is then fitted to the signal distributions satisfying the numerator or denominator selections obtained from data. As a result, the final background estimate shape is obtained by scaling the templates using

$$B_{\text{probes}} = N_{\text{template}} \times \left(\text{scale from fit where } \begin{cases} \text{model} = N_{\text{template}} + N_{\text{probes/MC}} \\ \text{data} = N_{\text{probes/data}} \end{cases} \right) \quad (5.6)$$

and

$$B_{\text{ID}} = N_{\text{template}} \times \left(\text{scale from fit where } \begin{cases} \text{model} = N_{\text{template}} + N_{\text{ID/MC}} \\ \text{data} = N_{\text{ID/data}} \end{cases} \right). \quad (5.7)$$

However, since this background estimate relies on signal templates obtained from MC simulated events, residual miscalibrations and inaccuracies from the MC modelling introduce a discrepancy between the data and the simulated shape of a few percent^[266]. As discussed in Section 3.2.6, these include the material effects that lead to additional bremsstrahlung processes which the simulation cannot properly model, but also electroweak effects not accounted for in the next-to-leading order prediction of the MC generator^[281].

To address this issue, a two-step procedure is performed where, during its first iteration, weighting factors are extracted to better describe the signal shape in data. These weights are calculated using the TIGHT data distributions, being defined as the ratio of the signal distribution after removal of this first background estimate to the predicted MC signal shape, that is,

$$w = \frac{N_{\text{TIGHT/data}} - B_{\text{TIGHT}}}{N_{\text{TIGHT/MC}}} \bigg|_{m_{ee} \in [75, 105] \text{ GeV}}. \quad (5.8)$$

They are obtained for the following m_{ee} bins:

MC weight: [60, 65, 70, 75, 78, 80, 82, 84, 86, 88, 90,
92, 94, 96, 98, 100, 102, 105, 130, 250] GeV,

a choice that accounts for the low statistics in these histograms^[281].

Once this first estimate of the TIGHT efficiency is performed and the MC weights are stored, in all subsequent measurements the MC templates are scaled using these factors, while keeping the original 1 GeV-binning. These include the ones used for the signal contamination fit ($N_{\text{template/MC}}$) and the background estimate fit ($N_{\text{probes/MC}}$ and $N_{\text{ID/MC}}$). The introduction of this MC weighting before the fits makes the final invariant-mass distributions smoother and we achieve a better agreement between the combination of the background estimate and the MC signal when compared to the data.

Examples of the invariant mass distributions obtained for numerator and denominator are drawn in Fig. 5.4. The same p_T and η_{BE2} bins previously presented for the normalisation in the mass tail method (Fig. 5.2) are shown. We observe a much better signal contamination removal, especially in the problematic η_{BE2} regions mentioned before and in the numerator shapes. These results indicate that efficiencies measured using the fit method are much more accurate.

That said, we note that the complexity involved in the fit procedure was such that implementing the normalisation in the mass tail method as a first step was more appropriate, while the fit method was implemented at a second stage. In the process of re-implementing the measurement in the new TAGANDPROBE software, many improvements were made to the likelihood-fit method with respect to its previous iteration. A summary of these changes is presented in the following section.

5.4.4 Improvements with respect to the previous implementation

As previously discussed, the main challenge of the Z -mass method is to properly estimate the background distributions in the experimental data. Using an approach based on a likelihood fit instead of a normalisation in the mass tail already entails a large improvement for this task. On top of that, many improvements have been made with respect to the previous fit implementation^[281] used in the last measurements^[266] to address some remaining issues on this front. Although these improvements were not introduced by the author of the thesis, measurements using this new method will be presented in Section 5.5.2 and correspond to part of the original contribution. In what follows we will briefly discuss these changes.

Firstly, the fits performed to obtain the background estimate were not checked for convergence or fit quality. Especially in the case of failed fits, it might be useful to retry it either using a different strategy or starting with the last point of the previous fit attempt. To address this issue, automatic fit checks were implemented, allowing for retries and generating warnings and logs to make sure the fits converge.

Concerning the event selection, the anti-identification selection used to obtain the background events was changed from an anti-VERYLOOSE to an anti-LOOSE working point. The previous choice of an anti-VERYLOOSE selection is good at first glance, since it reduces the signal

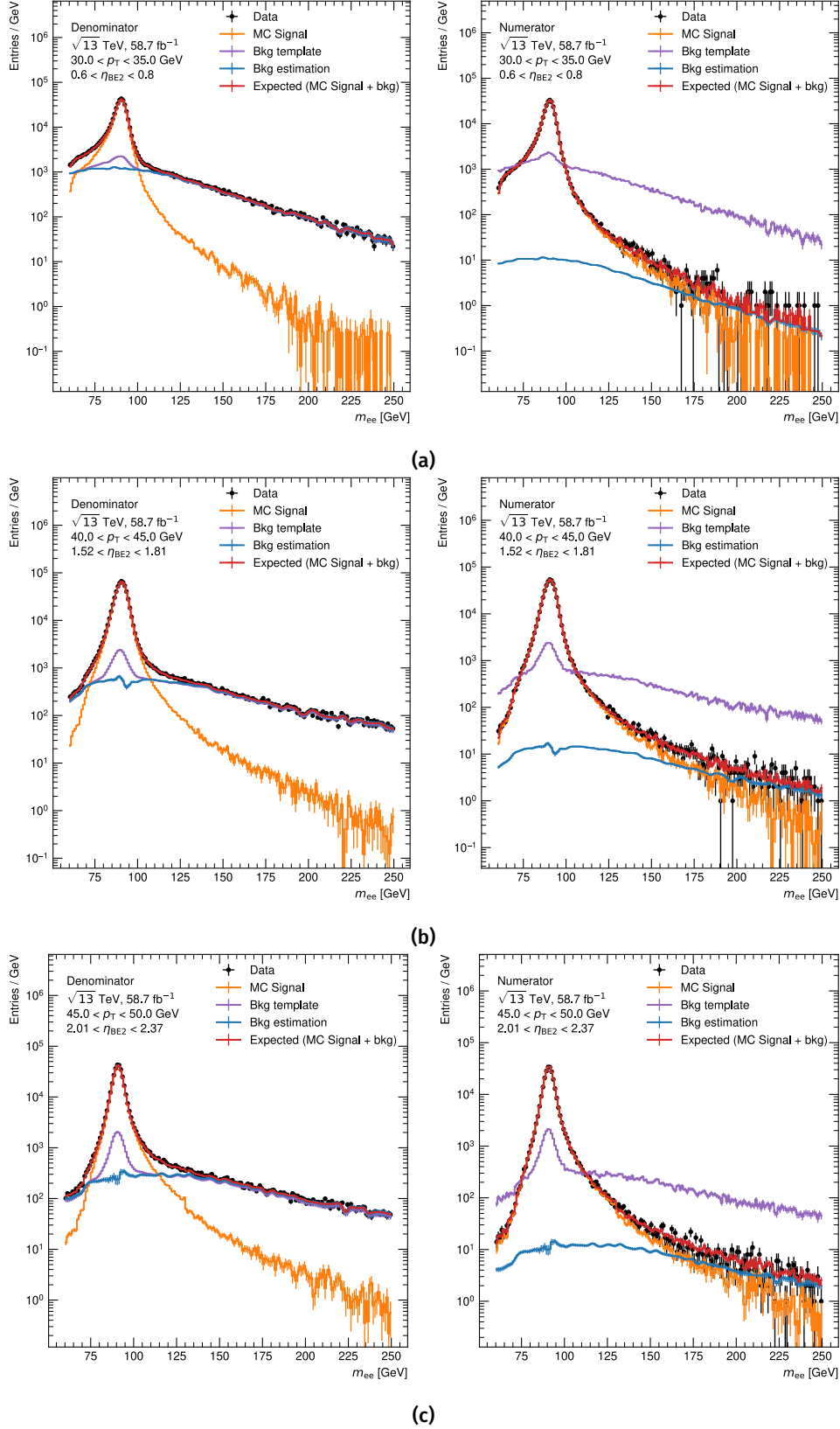


Figure 5.4: Invariant mass of the tag and probe pair m_{ee} for the denominator (left) and numerator (right) selections in selected p_T and η_{BE2} bins. The signal distributions obtained in data (black) and MC simulated events (orange) are drawn, together with the background template distribution (purple). The background estimate (blue) is obtained via a likelihood fit and it is combined to the MC signal shape to form the expected distribution (red).

contamination as much as possible. However, this impacts the estimate and subtraction of this signal contamination, given that in bins with almost no signal, the polynomial fit could completely flatten the background shape around the peak. The problem arises especially in regions where the background dominates the denominator, such as the low p_T region, where even a sub-percent amount of signal can have an impact on efficiencies of more than 1%. Moreover, in the previous implementation, to avoid fitting no signal contamination at all, a minimum requirement of 90% was made on the ratio of the MC prediction with respect to the reweighted MC shape for $p_T < 35$ GeV. This choice was based on completely arbitrary parameters. For this reason, an anti-Loose requirement is used instead, allowing for more signal contamination which can be better estimated. It must be noted that for medium p_T values and higher η_{BE2} regions, a considerable increase in signal is observed. Nevertheless, the impact on the result is not large, since the background is smaller in the measurement regions.

Furthermore, the previous iteration of the method used a Chebychev polynomial shape as the background model when estimating the signal contamination in the background template. This function can have negative values, which entailed problems in regions with low statistics. For this reason, a Bernstein polynomial is used instead, allowing for more stable fits since it only have positive values in this context. To further improve the fits in regions with very low statistics, the possibility to re-bin the histograms or force a first order polynomial was also introduced.

Lastly, as shown in Fig. 5.3, the background shape is heavily dependent on the probe transverse momentum p_T , moving from concave to convex, with the inflexion happening around the [20, 25] GeV region. It is not clear how this shape should be described, but it has an important effect on the fit of the signal electron contamination. The choice of the polynomial order, for instance, allows for changes of up to 3σ on the final scale factors. However, in the previous measurement, a third-order polynomial was used for all bins. To better address the complexity of the background shape at the inflexion-point region, now a linear function is used, while the third-order polynomial is considered as a systematic variation.

Having estimated the background shape for both numerator and denominator, the final step is to calculate the value of the efficiency and SF, discussed in the following section.

5.4.5 Efficiency and scale factors

Using the previously discussed methods to estimate the background, the values needed for the efficiency calculation can be extracted. For each (p_T, η_{BE2}) bin, the efficiency of each identification working point measured in data is given by

$$\epsilon_{\text{ID}}^{\text{data}} = \frac{S_{\text{ID}}}{S_{\text{probes}}} \bigg|_{m_{ee} \in [75, 105] \text{ GeV}} = \frac{N_{\text{ID}} - B_{\text{ID}}}{N_{\text{probes}} - B_{\text{probes}}} \bigg|_{m_{ee} \in [75, 105] \text{ GeV}}, \quad (5.9)$$

where the values are obtained from the number of objects within the integration region of each distribution. Similarly, these efficiencies are measured in MC simulated events, with the exception that no background estimation is needed and thus we have

$$\epsilon_{\text{ID}}^{\text{MC}} = \frac{S_{\text{ID}}}{S_{\text{probes}}} \Big|_{m_{ee} \in [75, 105] \text{ GeV}} = \frac{N_{\text{ID}}}{N_{\text{probes}}} \Big|_{m_{ee} \in [75, 105] \text{ GeV}}. \quad (5.10)$$

Since we use fits to determine the numerator and denominator values in these measurements and the number of objects before the ID cut (denominator) is correlated with the number of objects after the cut (numerator), the statistical uncertainties of these efficiencies cannot be calculated using simple error propagation. Instead, the statistical uncertainties are obtained using the following expression^[283]

$$\Delta \epsilon_{\text{ID}}^{\text{MC, data}} = \frac{\sqrt{(1 - 2\epsilon_{\text{ID}}^{\text{MC, data}})(\Delta S_{\text{ID}})^2 + (\epsilon_{\text{ID}}^{\text{MC, data}})^2(\Delta S_{\text{probes}})^2}}{S_{\text{probes}}}, \quad (5.11)$$

where ΔS_{ID} and ΔS_{probes} are the statistical uncertainties on the numerator and denominator values, respectively. When using the normalisation in the mass tail, these ΔS values were calculated using the expressions derived in Ref. [273], while, for the fit method, they were taken from the ROOT framework^[284] estimated error for the integrals⁽³³⁾.

Lastly, the ratio of the data and MC efficiencies is used to measure the SF, that is

$$\text{SF}_{\text{ID}} = \frac{\epsilon_{\text{ID}}^{\text{data}}}{\epsilon_{\text{ID}}^{\text{MC}}}, \quad (5.12)$$

with its statistical uncertainty being estimated via the usual error propagation of a ratio,

$$\left(\frac{\Delta \text{SF}_{\text{ID}}}{\text{SF}_{\text{ID}}} \right)^2 = \left(\frac{\Delta \epsilon_{\text{ID}}^{\text{data}}}{\epsilon_{\text{ID}}^{\text{data}}} \right)^2 + \left(\frac{\Delta \epsilon_{\text{ID}}^{\text{MC}}}{\epsilon_{\text{ID}}^{\text{MC}}} \right)^2, \quad (5.13)$$

which, combined with Eq. (5.12), leads to

$$\Delta \text{SF}_{\text{ID}} = \sqrt{\left(\frac{\Delta \epsilon_{\text{ID}}^{\text{data}}}{\epsilon_{\text{ID}}^{\text{MC}}} \right)^2 + \left(\frac{\epsilon_{\text{ID}}^{\text{data}} \Delta \epsilon_{\text{ID}}^{\text{MC}}}{(\epsilon_{\text{ID}}^{\text{MC}})^2} \right)^2}. \quad (5.14)$$

where $\Delta \epsilon_{\text{ID}}^{\text{data}}$ and $\Delta \epsilon_{\text{ID}}^{\text{MC}}$ are the statistical uncertainties of the efficiencies measured in data and MC simulated events, respectively.

We know, nevertheless, that detector effects and certain choices in the background estimation procedure may bias the final result. To account for that, systematic uncertainties are estimated and added in quadrature to these statistical uncertainties to obtain the final

(33) When the uncertainty of the numerator, ΔS_{ID} , is larger than that of the denominator, ΔS_{probes} , the correlations are ignored and we use a conservative error calculation instead to avoid raising unreasonable uncertainties.

uncertainty of the efficiency and SF measurement. That is,

$$\Delta X_{\text{total}} = \sqrt{(\Delta X_{\text{statistical}})^2 + (\Delta X_{\text{systematic}})^2}. \quad (5.15)$$

The process of estimating this systematic uncertainty consists in performing the same measurement with a given modification. A summary of how it was estimated in the context of the Z -mass method is discussed in the following section.

5.4.6 Systematic variations

Variations of the measurement were performed with modifications to different aspects of the analysis. These accounted for choices that are expected to have a large impact on the final result, such as the choices on the object selection or on the approach used to perform the background estimation. Overall, the variations covered the impact of allowing in more background, via the choice of the tag isolation or the integration region, or by modifying how the background selection and subtraction is performed^[273].

Starting with the choice of the tag selections, variations were performed with modifications on its isolation requirement. In the measurements using the normalisation in the mass tail, a calorimeter isolation requirement was added, such that the transverse energy E_T within a cone of radius $\Delta R = 0.40$ satisfied $E_T^{\text{cone40}} \leq 5.0$ GeV. This requirement reduces the backgrounds coming from multi-jet processes when compared to $W \rightarrow e\nu + \text{jet}$, since in the latter the tag would be a real electron^[273]. Conversely, in the fit method, the track isolation requirement was modified, such that one variation considered a smaller value, $p_T^{\text{cone20}}/p_T \leq 0.06$, while the other had a larger value, $p_T^{\text{cone20}}/p_T \leq 0.14$.

Now, the Z -mass method relies on the definition of a signal region around the peak formed at the Z boson mass value of 91 GeV. In both background estimation approaches, the reference range around this value was taken to be $m_{ee} \in [75, 105]$ GeV, with both wider and narrower ranges of $m_{ee} \in [70, 110]$ GeV and $m_{ee} \in [80, 100]$ GeV being added as variations.

Furthermore, the selections used to isolate background objects in both data and MC events were modified, as previously listed in Table 5.2. We note that, when normalising the background template in the mass tail, no variations are performed for the selection of the backgrounds with $p_T < 30$ GeV. Instead, the choice of normalisation region is varied from the high-mass tail ($m_{ee} \in [120, 250]$ GeV) to the low-mass region ($m_{ee} \in [60, 70]$ GeV). Alternatively, for objects with energies beyond this threshold, a variation on their isolation is performed, with variations using the $E_T^{\text{cone40}}/p_T \geq 0.05$ GeV or $E_T^{\text{cone40}}/p_T \geq 0.20$ GeV requirements. The lack of enough events at the low-mass region for these high-energy objects favours the use of only the high-mass normalisation region at $m_{ee} \in [120, 250]$ GeV.

In contrast, the use of the fit method defines three variations beyond the nominal background selections. These change the identification requirements the objects must fail (from LOOSE to VERYLOOSE), while also changing the object's isolation from $E_T^{\text{cone30}}/p_T \geq 0.01$ GeV

to $E_T^{\text{cone40}}/p_T \geq 0.07$ GeV. Beyond the choices of background objects definitions, the mass binning used to perform the background fit was also modified. In one of the variations, the high-mass bins are removed, such that,

Numerator: [65, 80, 100, 120] GeV, and

Denominator: [65, 70, 75, 80, 100] GeV,

while another variation removes the lower mass bin instead,

Numerator: [80, 100, 120, 250] GeV, and

Denominator: [80, 100, 120, 150, 250] GeV.

The uncertainties on the scaling value obtained from both the background fit and the signal contamination fit are also considered, with variations using the upper or lower limit on the scaling factors instead. Moreover, the choice of a third-order Bernstein polynomial is added as a variation when estimating the signal contamination for $p_T \in [20, 25]$ GeV, where the inflexion of the background shape happens. Lastly, a variation is performed without the reweighing of the MC distributions (Eq. (5.8)).

Once all variations have been measured, the final values of the efficiencies can be estimated. When using the normalisation in the mass tail method, the central value is given by the mean value of all the variations, while for the fit method it is obtained from the nominal variation. Moreover, the total systematic uncertainty is obtained by adding the difference of the variations to the central value in quadrature.

Using the Z -mass method described in this section, measurements of the electron identification efficiencies were performed in data and MC simulated events. The results obtained in this work are presented in the following section.

5.5 Results

Alongside its implementation in the new software suite, measurements of the electron identification efficiency were performed using the Z -mass method. Results were obtained for the efficiencies over MC simulated events and experimental data for the p_T - η_{BE2} bins previously listed in Table 5.4, as well as the corresponding SF values. All measurements quantified the performance of the LLH working points being used at the time, which were calibrated with the previous release of the ATLAS software, as described in Section 4.3.1.

Firstly, we present measurements performed with the data collected early on Run 3. This represents the first look at the collected data from this new data taking period. Next, measurements were also performed using the data recorded during Run 2 after its reprocessing with the new ATLAS software version. In what follows, we present these electron identification

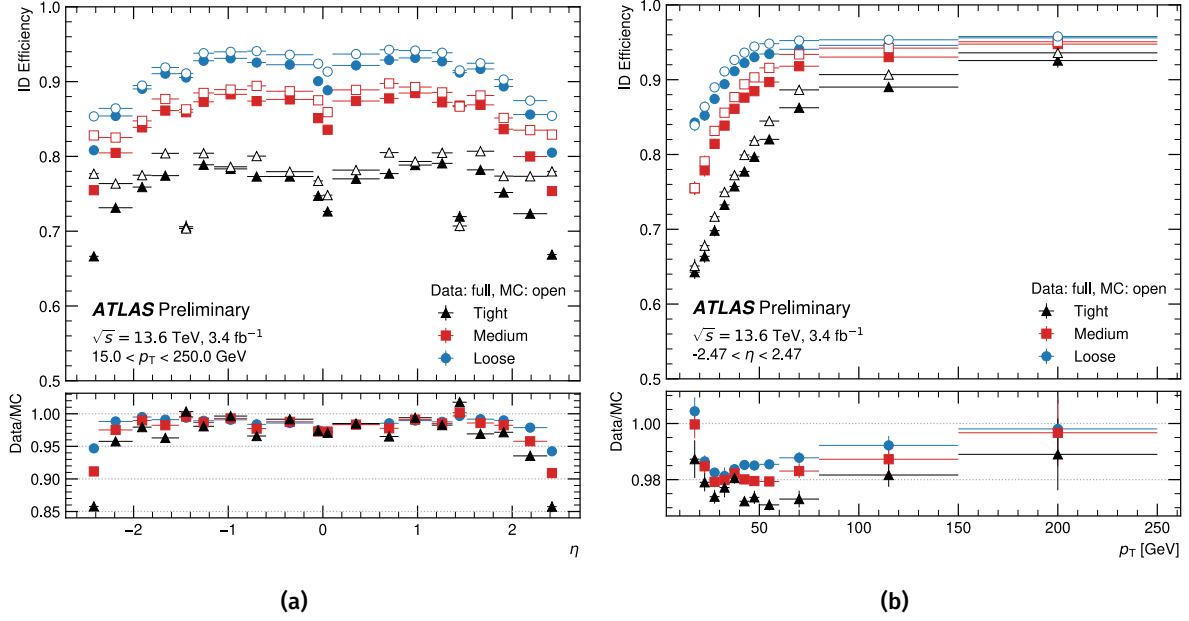


Figure 5.5: Electron identification efficiencies (top) and scale factors (bottom) measured as a function of (a) transverse momentum p_T and (b) pseudorapidity η_{BE2} . The measurement was performed in data recorded by the ATLAS experiment in 2022 from pp collisions at $\sqrt{s} = 13.6$ GeV, corresponding to an integrated luminosity of 3.4 fb^{-1} . Three working points from the likelihood identification algorithm were considered, namely, TIGHT, MEDIUM and LOOSE, with an extra B-Layer requirement on the latter. The Z -mass method was performed to estimate the background, via a normalisation of the invariant mass distribution templates to the data in a mass region far from the Z peak.

efficiencies and SF measured using the Z -mass method implemented in the new TAGANDPROBE software.

5.5.1 Run 3 (2022)

The ATLAS data taking period for Run 3 started in July 2022 and it represented the first data recorded entirely using the ATLAS software in its major version 22. As a consequence, the newly implemented software suite for efficiency measurements was needed to get a first look at how the electron identification working points were performing.

Measurements were taken over both data and MC simulated events. This data corresponded to 3.4 fb^{-1} of pp collisions with a centre of mass energy of 13.6 TeV recorded by the ATLAS detector in 2022. The background was estimated using the normalisation in the mass tail approach, described in Section 5.4.2, since the fit method was not yet implemented at the time. Three working points of the likelihood identification algorithm were considered, namely TIGHT, MEDIUM and LOOSE.

The results are presented in Fig. 5.5. Both the efficiency and SF values are symmetric and rather uniform across η_{BE2} (Fig. 5.5a), attaining their higher values in the central region and decreasing as η_{BE2} increases. Moreover, the efficiency values decrease as we decrease the

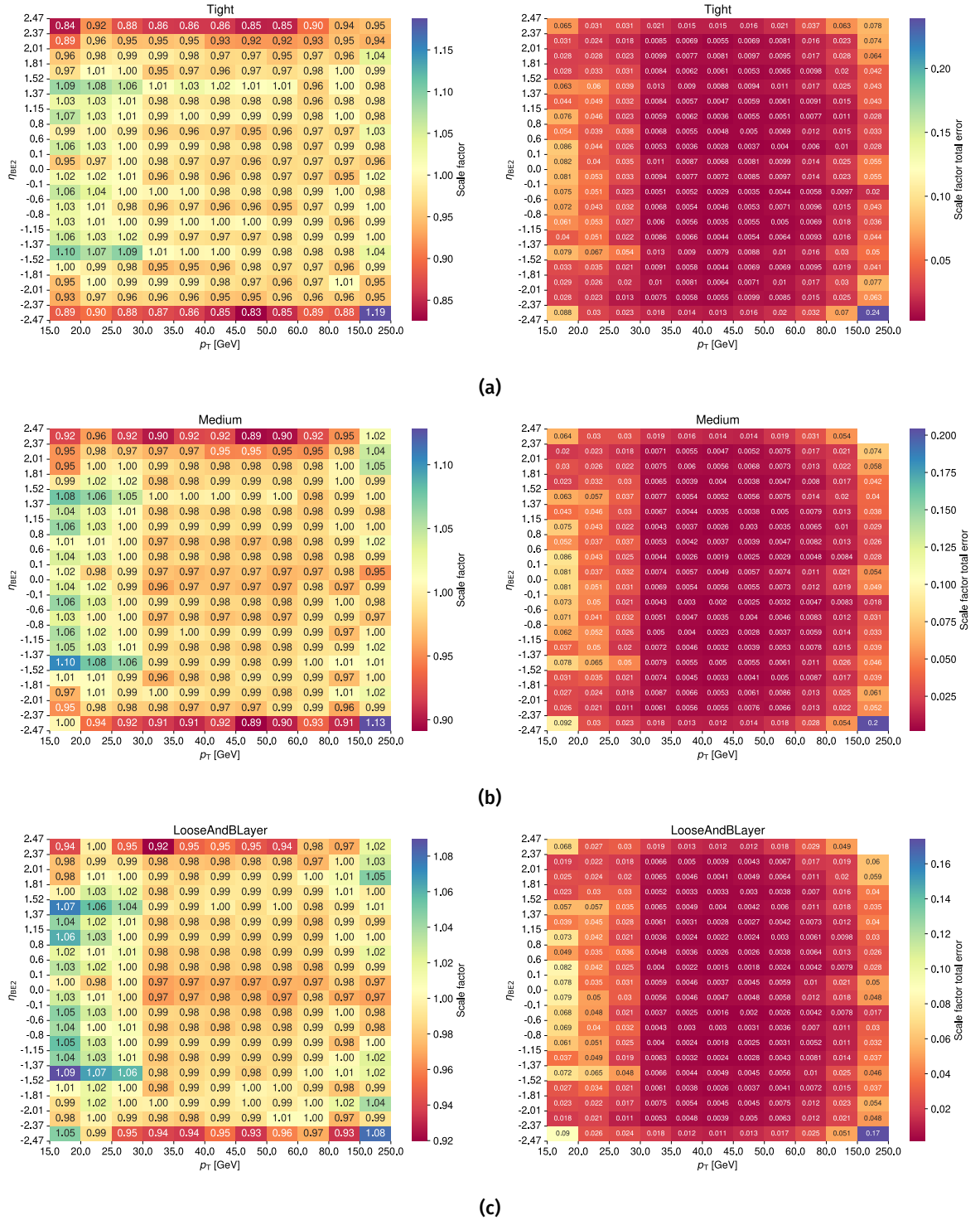


Figure 5.6: Run 3 electron identification scale factors central values (left) and total uncertainty (right) for different values of p_T and η_{BE2} , for three likelihood working points, namely (a) TIGHT, (b) MEDIUM and (c) LOOSE, the latter with an extra B-layer requirement. The measurement was performed in data recorded by the ATLAS experiment in 2022 from pp collisions at $\sqrt{s} = 13.6$ GeV, corresponding to an integrated luminosity of 3.4 fb^{-1} .

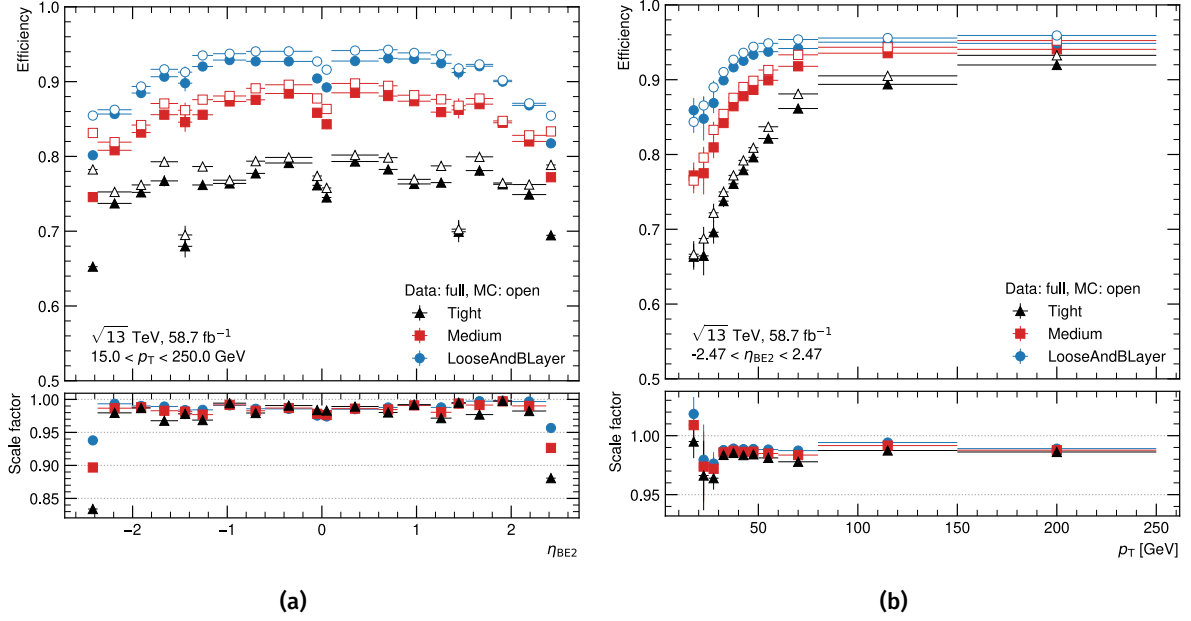


Figure 5.7: Electron identification efficiencies (top) and scale factors (bottom) measured as a function of (a) transverse momentum p_T and (b) pseudorapidity η_{BE2} . The measurement was performed in data recorded by the ATLAS experiment in 2018 from pp collisions at $\sqrt{s} = 13$ GeV, corresponding to an integrated luminosity of 58.7 fb $^{-1}$. Three working points from the likelihood identification algorithm were considered, namely, TIGHT, MEDIUM and LOOSE, with an extra B-Layer requirement on the latter. The Z -mass method was performed to estimate the background, via a likelihood fit of the invariant mass distribution templates to the data.

signal efficiency, when changing from the LOOSE to the TIGHT working points, as expected. We also note the data/MC ratios are slightly higher for the TIGHT working point, the agreement improving for the LOOSE working point. Lastly, we observe a smooth decrease in performance as the electron p_T decreases (Fig. 5.5b), while the data/MC SF values remain close to unity for the whole energy spectra.

The SF central value and total uncertainty corresponding to the TIGHT, MEDIUM and LOOSE LLH working points are presented as a function of the p_T - η_{BE2} bins in Fig. 5.6. In these intensity plots, or heatmaps, we observe that the general tendencies of the SF values are stable across the different working points. In particular, smaller values are observed in the forward region, such as $\eta_{\text{BE2}} \in [2.37, 2.47]$, while slightly higher values are observed for low- p_T ($p_T < 30$ GeV). In these low energy regions, we note the total uncertainty is in the 2%–8% range, while elsewhere the error remains at the sub-percent level.

5.5.2 Run 2 (2018)

While waiting for more Run 3 data to be recorded, measurements were also performed over the full Run 2 data that was reprocessed using the release 22 of the ATLAS software. This time, the newly implemented background estimation with the likelihood fit method was used.

As such, it allowed a comparison with the previous published results^[266], which gauged the changes between the ATLAS software versions together with the new implementation of the efficiency measurement.

During the entire Run-2 period, ATLAS recorded a total of 139.9 fb^{-1} of pp collisions at a centre of mass of 13 TeV, such that 3.2 fb^{-1} was collected in 2015, 33.4 fb^{-1} in 2016, 44.6 fb^{-1} in 2017 and 58.7 fb^{-1} in 2018. Since the measurements are performed separately for each one of these years and their results are very similar, in this work only the ones obtained for the 2018 data will be presented for simplicity.

The efficiencies measured in MC simulated events and real data are presented in Fig. 5.7, together with the corresponding SF values. They were obtained for the same LLH working points used in the previous Run 2 measurements, namely TIGHT, MEDIUM and LOOSE, the latter with an extra B-layer requirement. We observe the same symmetry and uniformity across η_{BE2} , with drops around the transition ($|\eta_{\text{BE2}}| \in [1.37, 1.52]$) and forward ($|\eta_{\text{BE2}}| \in [2.37, 2.47]$) regions. The efficiencies also decrease as the electron energy gets smaller, especially below $p_{\text{T}} < 50 \text{ GeV}$. Nevertheless, the SF values remain close to unity except for the last $|\eta_{\text{BE2}}|$, where it goes as low as 0.8 for the TIGHT LLH working point.

Heatmaps showing how the SF central values and total uncertainties vary as a function of the $p_{\text{T}}-\eta_{\text{BE2}}$ bins for the different working points are drawn in Fig. 5.8. As expected, the values are overall close to unity, with the total uncertainty at the sub-percent level. We note deviations from this, however, at the last $|\eta_{\text{BE2}}|$ bins and in the low p_{T} region, with the total error ranging within the 2%–10% range.

General tendencies on how the data and MC efficiencies changed within releases are presented in Fig. 5.9. In these heatmaps, we present the relative difference between the central values of the data (left) and MC (right) efficiencies obtained using the Z -mass method in release 21 and release 22 for the three working points. Also, we note that the comparison in the last $p_{\text{T}} \in [150, 250] \text{ GeV}$ is not reliable, since the release 21 measurement was performed over $|\eta_{\text{BE2}}|$ bins to increase statistics, while also not having the higher limit of $p_{\text{T}} < 250 \text{ GeV}$ that the new measurement has. The comparison is further presented as a function of p_{T} and η_{BE2} for the MC and data efficiencies in Figs. 5.10 and 5.11, respectively, where the TIGHT working point is shown for selected bins. The ratio with respect to the previous measurements are shown at the bottom plots, with its uncertainty drawn as a yellow band.

We observe a decrease in the efficiency in the low p_{T} region, especially in the $[20, 25] \text{ GeV}$ bin, accompanied by an increase in the corresponding uncertainties. As previously discussed, this bin contains the transition in the background shape from convex to concave, which we believe is better modelled in the new implementation. Furthermore, we note a general improvement in efficiencies for the central region ($|\eta_{\text{BE2}}| \in [0.0, 0.1]$) and the transition region ($|\eta_{\text{BE2}}| \in [1.37, 1.52]$), in particular for the MC efficiencies.

Lastly, a comparison of the SF values of the TIGHT working point is shown in Fig. 5.12 for selected bins. We observe an overall improvement across all η_{BE2} bins, with SF values being

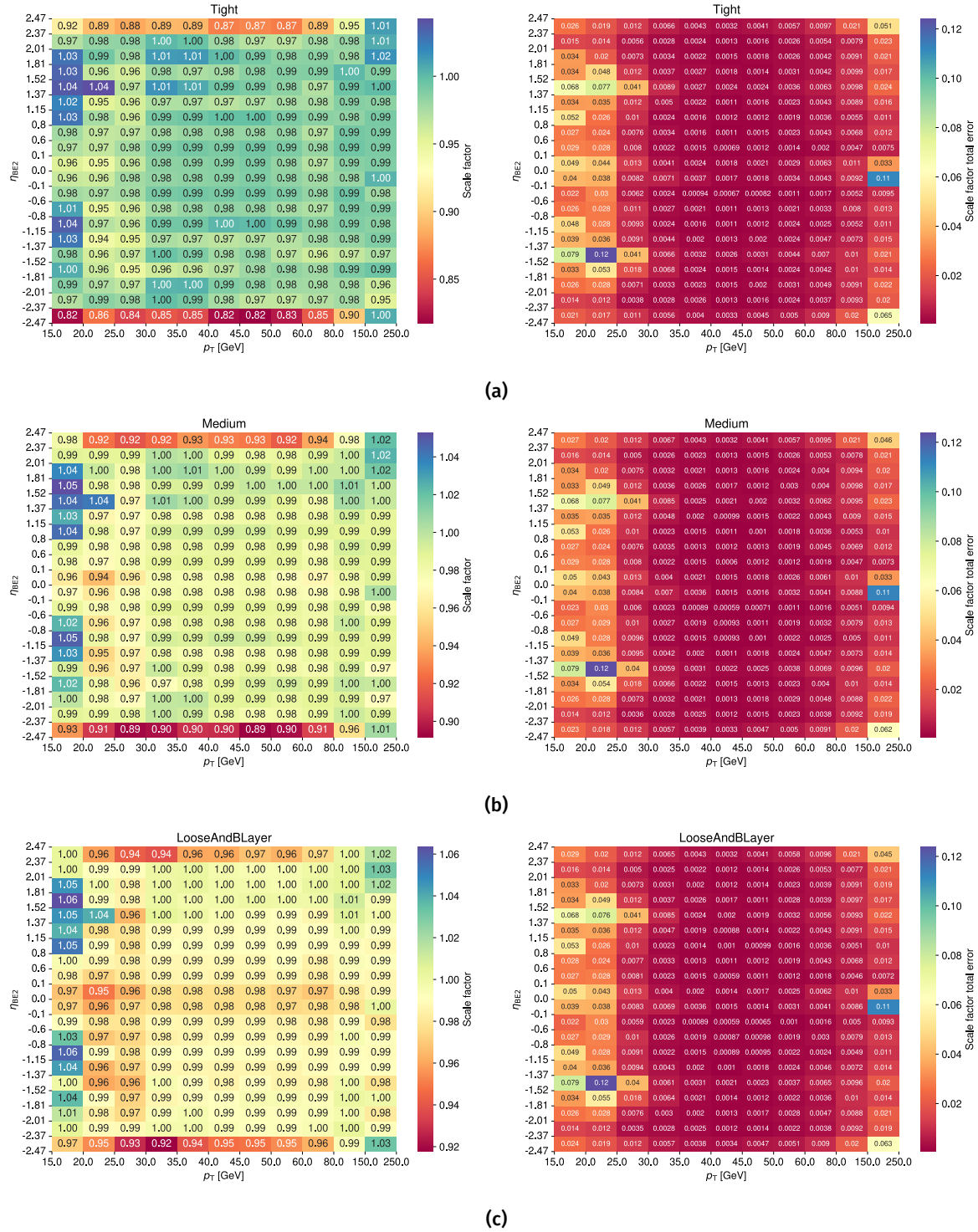


Figure 5.8: Electron identification scale factors central values (left) and total uncertainty (left) for different values of p_T and η_{BE2} , for three likelihood working points, namely (a) TIGHT, (b) MEDIUM and (c) LOOSE, the latter with an extra B-layer requirement. The measurement was performed using the software in its release 22 over data recorded by the ATLAS experiment in 2018 from pp collisions at $\sqrt{s} = 13$ GeV, corresponding to an integrated luminosity of 58.7 fb^{-1} .

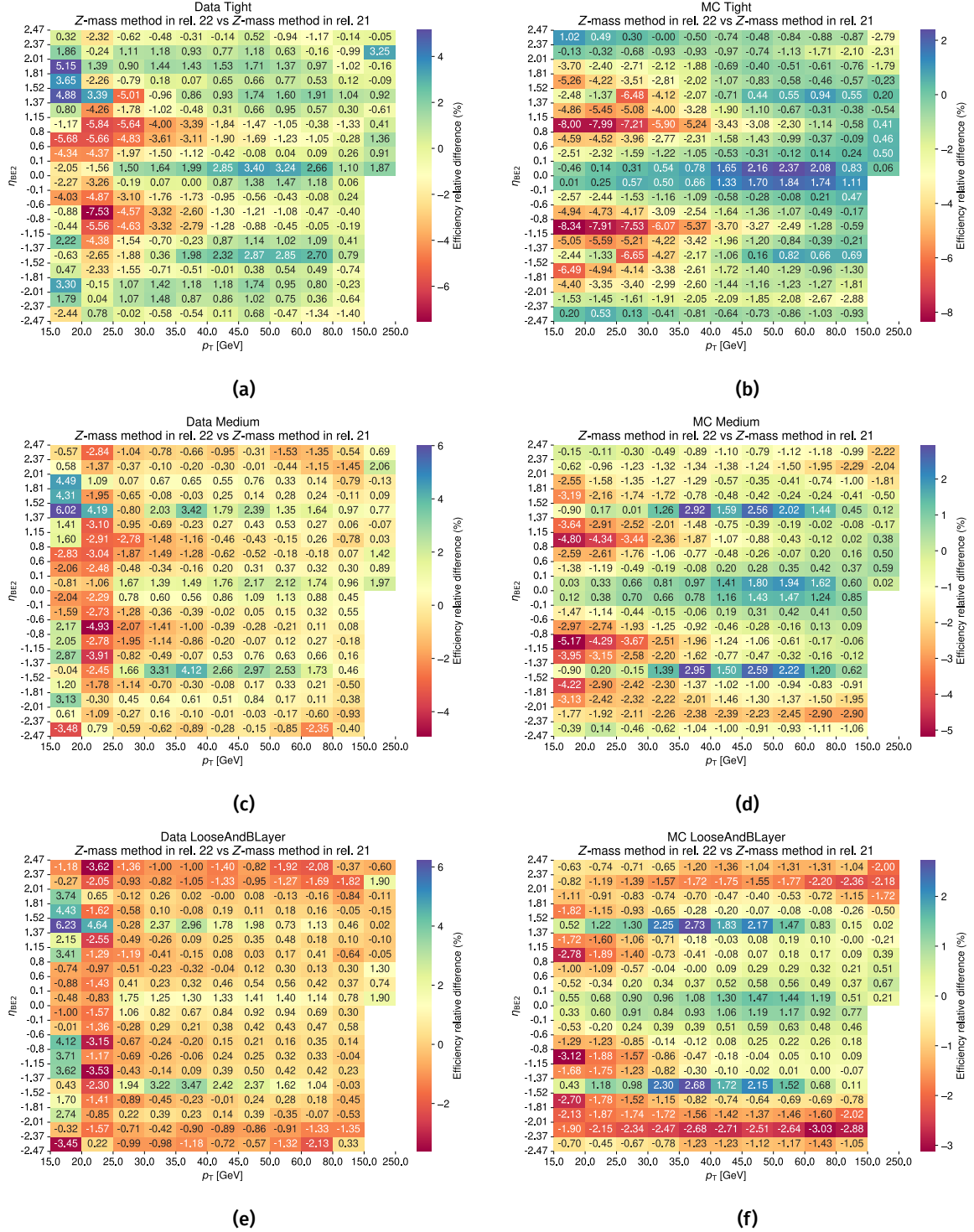


Figure 5.9: Relative difference on the central value of electron identification efficiencies measured with the ATLAS software release 22 with respect to their latest published results^[266] measured with its release 21, presented for different values of p_T and η_{BE2} . The measurement was performed in data recorded by the ATLAS experiment in 2018 from pp collisions at $\sqrt{s} = 13$ GeV, corresponding to an integrated luminosity of 58.7 fb^{-1} . Efficiencies were measured in both data and MC simulated events for the three likelihood working points, namely TIGHT, MEDIUM and LOOSE, the latter with an extra B-layer requirement.

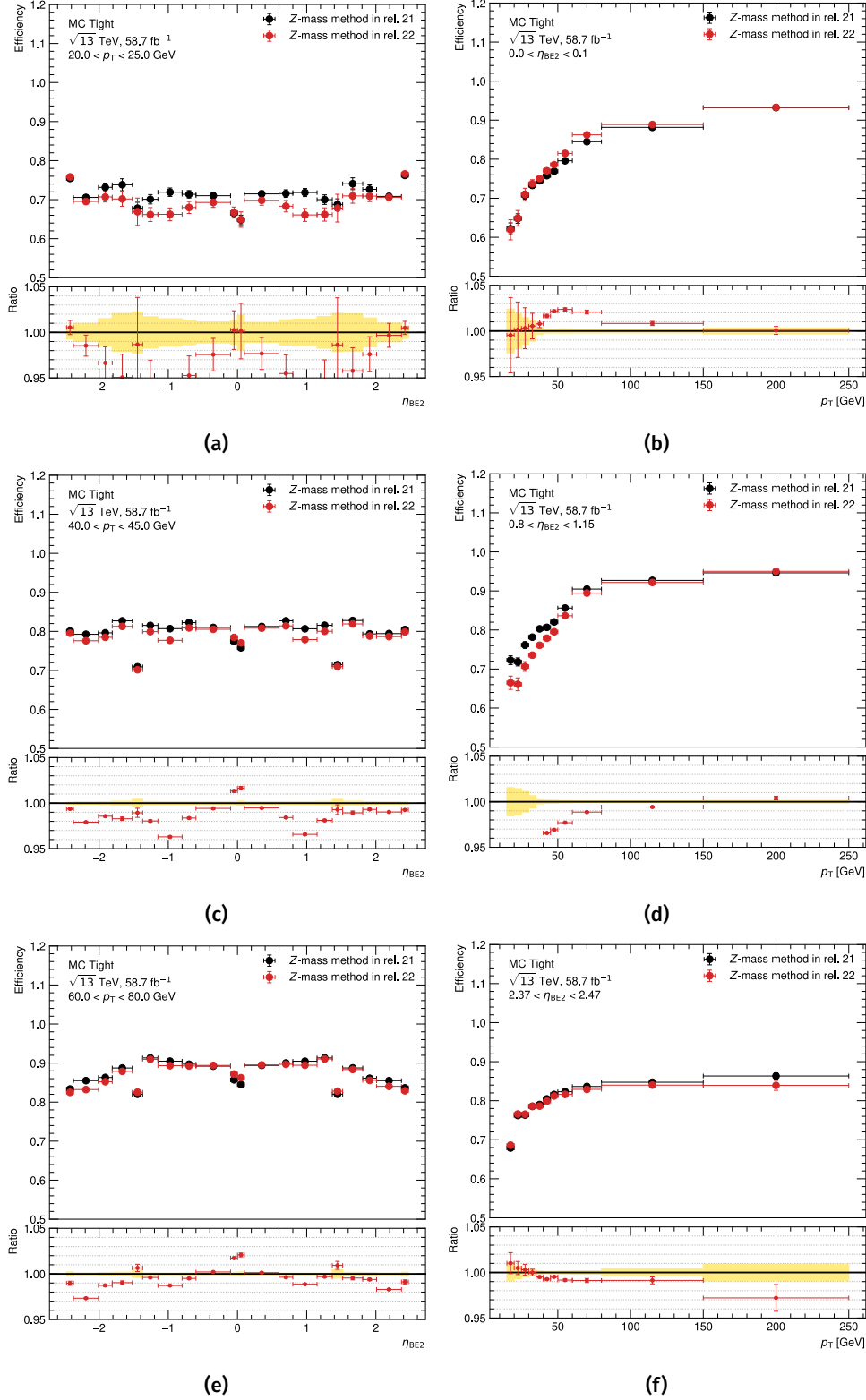


Figure 5.10: Electron identification efficiencies measured for the TIGHT likelihood working point in MC simulated events with both the new ATLAS software release 22 (red) and their latest published results^[266] measured with its release 21 (black), as a function of p_T and η_{BE2} for selected bins. The ratio of the release 22 measurement with respect to the release 21 measurement is drawn, with a yellow band corresponding to its uncertainty. Both measurements were performed in data recorded by the ATLAS experiment in 2018 from pp collisions at $\sqrt{s} = 13$ GeV, corresponding to an integrated luminosity of 58.7 fb^{-1} .

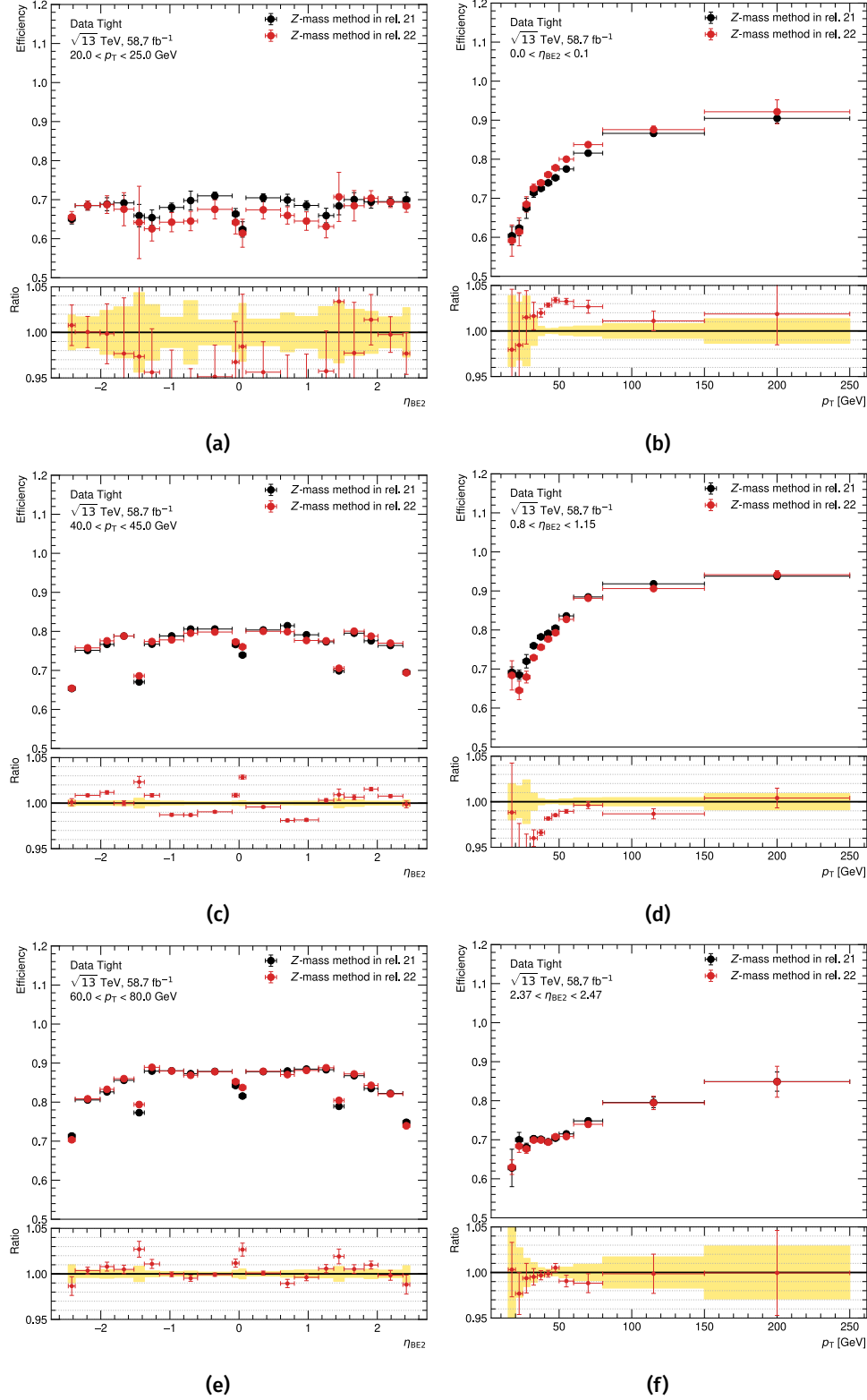


Figure 5.11: Electron identification efficiencies measured for the TIGHT likelihood working point in experimental data with both the new ATLAS software release 22 (red) and their latest published results^[266] measured with its release 21 (black), as a function of p_T and η_{BE2} for selected bins. The ratio of the release 22 measurement with respect to the release 21 measurement is drawn, with a yellow band corresponding to its uncertainty. Both measurements were performed in data recorded by the ATLAS experiment in 2018 from pp collisions at $\sqrt{s} = 13$ GeV, corresponding to an integrated luminosity of 58.7 fb $^{-1}$.

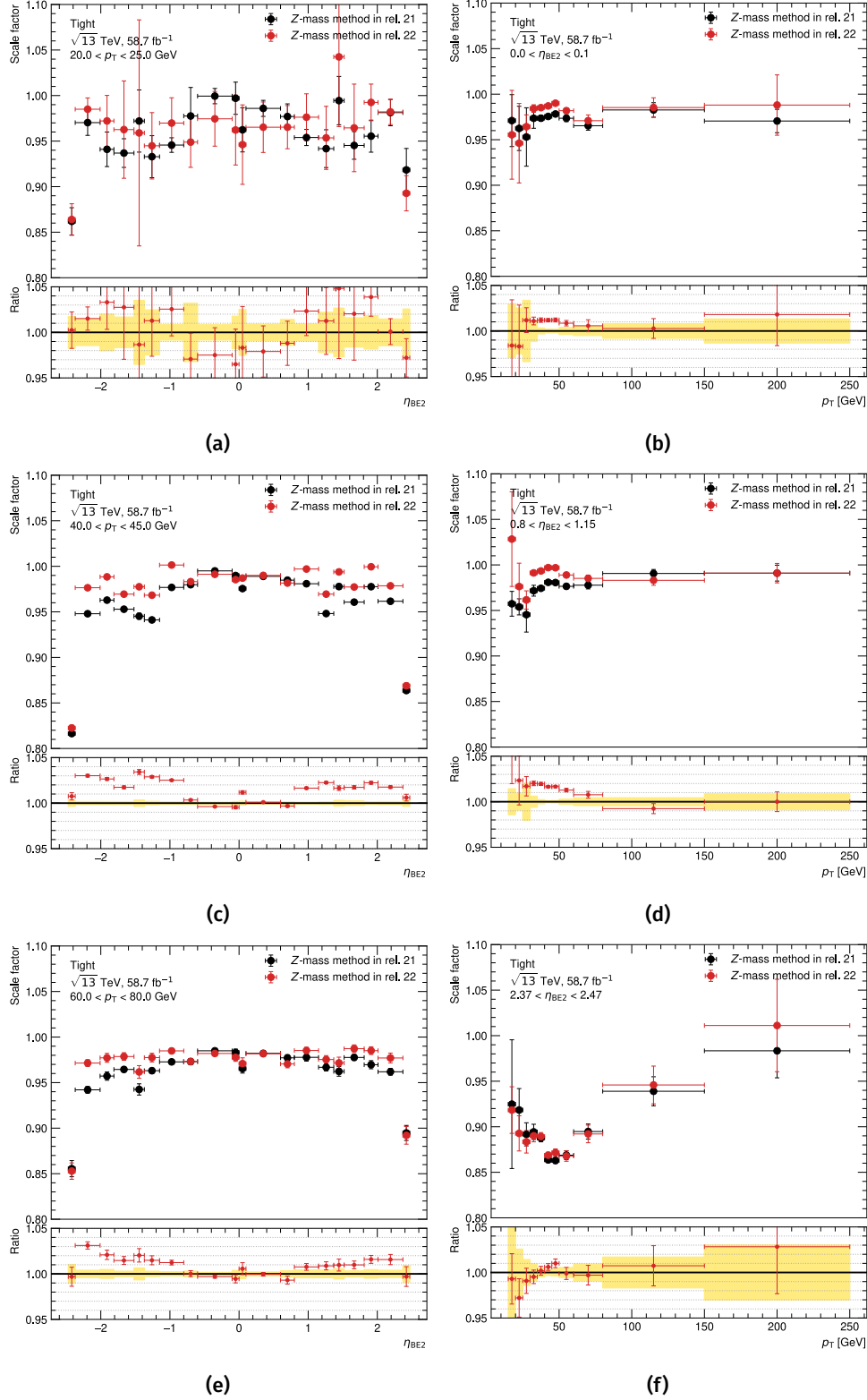


Figure 5.12: Electron identification scale factors measured for the TIGHT likelihood working point with both the new ATLAS software release 22 (red) and their latest published results^[266] measured with its release 21 (black), as a function of p_T and η_{BE2} for selected bins. The ratio of the release 22 measurement with respect to the release 21 measurement is drawn, with an yellow band corresponding to its uncertainty. The measurement was performed in data recorded by the ATLAS experiment in 2018 from pp collisions at $\sqrt{s} = 13$ GeV, corresponding to an integrated luminosity of 58.7 fb^{-1} .

closer to unity. This aligns with the improvements made in the ATLAS software release 22 version, and show that the discrepancies between MC simulated events and the experimental data have decreased in this context. The improvement is not so clear for lower energies, since results overlap within the uncertainty of the measurements. No major decrease in the SF is observed outside the quoted error, further establishing the general tendencies of improvement in the ATLAS performance.

5.6 Future improvements

In this chapter, we presented measurements of the electron identification efficiency and SF using the Z -mass method. These were obtained using its new implementation in the `TAGANDPROBE` software, which was completely rewritten in light of the update of the ATLAS software to its major version 22. Alongside with the software development to implement the two background estimation approaches of the method, measurements were performed using data taken in early Run 3, as well as the full Run 2 data reprocessed using the new software version.

The general behaviour of the identification efficiencies in both Run 3 and Run 2 follow the previously observed patterns^[266], namely symmetrical and rather uniform values across η_{BE2} and an increased performance as p_{T} increases. Both data and MC efficiencies are smaller for the `TIGHT` working point and have a smaller data/MC ratio, while we observe increasing efficiency values and a better agreement for the `LOOSE` working point. Additionally, we observe scale factor values are within the 0.95–1.0 range, except for higher values of $|\eta_{\text{BE2}}|$, where the MC simulation mismodelling is known to be large.

Many challenges were encountered in the new implementation, both in the technical and physics side. As previously discussed, the lack of enough statistics in certain $p_{\text{T}} - \eta_{\text{BE2}}$ intervals made it hard to properly perform a fit, a challenge that was addressed by allowing more signal contamination and a creating a better method to estimate it. Despite not being discussed in this work, similar improvements were developed in the other complimentary methods used to measure the identification efficiencies, namely the Z -iso method and the J/ψ method previously mentioned. Moreover, the new implementation presented in this work was also used to measure isolation and trigger efficiencies, covering all the different elements of the full electron efficiency measurement.

The measurements presented here allowed the assessment of the performance of the electron identification algorithm in light of the new data-taking period of Run 3. We note that it is hard to make a fair comparison between the Run 2 and Run 3 results, given the changes in the detector configuration and running conditions between the two data-taking periods. For instance, during Run 3, the TRT gas was modified and a new pile-up modelling was introduced^[223]. Nevertheless, we observe an overall stability between the identification efficiencies and SF values (Fig. 5.5 and Fig. 5.7). With no major surprises observed in the Run 3 results, the challenge now is to further decrease the uncertainty on these values in order to

continue improving the precision of the affected ATLAS measurements.

As presented in this work, some changes were introduced to the Z -mass method during its re-implementation. Among the improvements, we highlight the use of a more accepting anti-identification working point that allows a better signal contamination fit, as well as the use of a linear Bernstein polynomial in the non-trivial region where the background changes its curvature. Nevertheless, some further improvements could be addressed in the method, especially concerning how the fits are performed.

Firstly, the main challenge is the proper understanding of what populates the background. While at low energies we are confident they correspond to objects mimicking the electron signature, it is not as clear for higher energies. Some leading backgrounds that are not currently accounted for are objects from $t\bar{t}$ and multi-boson productions. Even though the contribution is small, they should be properly understood in order to achieve sub-percent precision over all efficiencies.

The great dependence of the background shape on p_T could also be further investigated in order to improve the method. Especially at low energies where, the change in curvature happens, we could attempt a finer binning to avoid the inflexion point in the $p_T \in [20, 25]$ GeV region. Also, different background functions could be considered, allowing the extension of the signal fit to bins with $m_{ee} < 65$ GeV, which represents the current lower limit of the range where the signal fit is performed. Concerning the background estimation itself, we could consider a fit in both the background and signal regions. This could further improve the signal estimation in the background shape.

Other challenges are expected when we combine these Z -mass values with those obtained from methods covering other kinematic regions. In particular, the correlations between the Z -mass and Z -iso methods must be properly studied, since they both rely on the same signal objects from the same Z boson decays. And while the background estimations use different control regions, and thus are uncorrelated, there is an overlap on the systematics introduced when performing the tag and probe selections (for instance, variations on the mass windows and the isolation requirement of the tag). Properly identifying these correlations between the systematic variations is thus needed to improve the precision of future measurements.

Beyond improvements in the measurement of the electron efficiencies, we could also consider improvements in the identification algorithm itself. One possibility to do so involves the use of machine learning techniques to develop an electron classifier. An example of such algorithm is the subject of the following chapter.

Chapter 6

Improving electron identification with a convolutional neural network

We are stuck with technology when what we really want is just stuff that works.

Douglas Adams^[285]

6.1	Motivation for the use of a convolutional neural network	114
6.2	Input features	115
6.2.1	High-level variables (HLV)	115
6.2.2	Track variables	116
6.2.3	Calorimeter images	116
6.3	Task and architecture	120
6.4	Datasets	123
6.4.1	Simulation data	123
6.4.2	Experimental data	125
6.5	Results	127
6.5.1	Performance over simulation data	127
6.5.2	Feature ranking	131
6.5.3	Performance in data versus simulation	136
6.6	Future improvements	141

As discussed in Chapter 4, the current electron identification method used in the ATLAS experiment relies on a likelihood approach. But machine learning techniques, as discussed in Chapter 2, have the potential to further improve its performance, in particular by exploiting low-level inputs from the detector and the correlation between high-level variables. In this chapter, we present the implementation of a convolutional neural network for electron identification, together with studies of its input features and the performance when comparing MC simulation and experimental data.

6.1 Motivation for the use of a convolutional neural network

The electron identification algorithm used in the ATLAS detector, described in Section 4.3, is optimised to discriminate electrons produced at the interaction point, referred to as *prompt electrons*, from its different types of backgrounds. These backgrounds correspond to objects that mimic the prompt electron signature, but are in fact jets, photons or electrons originating from decays of b - or c -hadrons. This identification procedure relies on observables defined from the shower shape, track and track-cluster variables registered by the detector. Using the LLH approach, these variables are assumed to be uncorrelated, which simplifies the process of retuning the identification algorithm in case a given variable must be added or removed.

However, this assumption does not fully hold because many of these variables are correlated. For instance, the ratio of transverse energy in the hadronic calorimeter to the EM cluster energy, quantified via the variable R_{had} , will increase if the ratio of the energy in the third layer to the total energy in the EM calorimeter, f_3 , increases, since large energy deposits in the hadronic calorimeter follow large energy deposits in the third EM layer. Instead of ignoring these correlations, machine learning techniques can be used to take advantage of them, as seen in Chapter 2.

Since the electron identification algorithm is used in both the trigger and during data analysis, the chosen method needs to be highly performant. One of the metrics used to gauge this performance is the background rejection, introduced in Section 2.5. It quantifies how well the method can correctly identify the background as a function of its ability to correctly identify the signal. As such, improving the background rejection for a fixed value of the signal efficiency is particularly useful in analyses that suffer from specific types of background contamination. Thus, to increase the signal/background sensitivity of certain analyses, an identification that maximises background rejection is preferred.

Moreover, since the trigger system requires a fast response and must contemplate as many scenarios as possible, a binary classification between signal and background is implemented at this level. Now, for the physics analyses studies, it might be useful to calibrate the performance to better reject a particular type of background. In this scenario, these studies would benefit from a multinomial classification where different types of background targets are identified.

For instance, in analyses relying on single-lepton final states^[262,286], fake electrons are coming from hadronic backgrounds, while there are no other leptons that could be used to reject multi-jet backgrounds (such as Z and W boson production or $t\bar{t}$ production). In such cases, an identification method that could be tuned to mostly reject this particular type of background could greatly increase the sensitivity of the measurement. On the other hand, some measurements suffer from the similarities of the photon and electron signatures, such as in $t\bar{t}H$ production^[259], where the main Higgs signatures are two photons ($H \rightarrow \gamma\gamma$) or four leptons ($H \rightarrow ZZ^* \rightarrow 4\ell, \ell = e, \mu$). The precision of this measurement could be improved with a better discrimination of prompt electrons coming from the collision vertex and those

coming from photons converted into electromagnetic showers in the calorimeter.

In this context, neural networks are good candidates to improve electron identification, as they address these limitations. Thriving over input correlations, they can exploit the non-linear relationships between the detector's inputs and generate a multinomial output with enhanced performance. Not only that, but they present the possibility to directly use low-level detector signatures, incorporating all the information the detector has available to the task. As a result, a lot of research has been performed, in recent years, to develop machine learning techniques for the electron identification task.

A first approach was developed using a deep neural network (DNN) with input variables similar to the likelihood method, while also introducing a multinomial classification into a signal electron and different types of backgrounds^[287]. This method already offers a boost in performance and outperforms the current likelihood approach. However, this approach is still limited to the use of scalar variables, such as the ones already used by the likelihood method.

Since neural networks can manage different types of inputs, such as images, they allow for the use of additional information recorded by the detector. In particular, both the calorimeter energy deposits and information from tracks in the inner detectors can be processed as images using CNNs^[288], a technique described in Section 2.6.4. In what follows, the development of a CNN algorithm for electron identification is described, including optimisation and performance studies performed incorporating real experimental data.

6.2 Input features

6.2.1 High-level variables (HLV)

Once an electron candidate is reconstructed via the procedure described in Section 4.2, a number of high-level variables (HLV) are defined. The ones used by the LLH identification method were listed in Table 4.1. They were engineered to enhance the background discrimination, for instance, of hadronic showers and converted photons. Hence, they are important features to be included in any electron identification algorithm.

Now, some new HLV are selected^[2], as listed in Table 6.1, incorporating more track information as well as kinematic variables. Among them, the charge of the best matched track is considered in order to improve identification of electrons that have the reconstructed charge opposite to the true charge^[245], as well as the number of degrees of freedom of the track fit and the number of tracks matched to the candidate. Moreover, the transverse energy E_T , its ratio to the transverse momentum E_T/p_T and the pseudorapidity η of the electron candidates are also included as input features.

Before being processed by the CNN algorithm, a non-linear transformation is applied to all the HLVs to allow each one of them to follow a normal distribution. This process spreads out the most frequent values of the variables and reduces the impact of outliers. It is implemented

Table 6.1: Additional high-level variables included in the CNN electron identification algorithm^[2].

Type	Variable	Description
Track quality	$q \times d_0 / \sigma(d_0)$	Charge of the best matched track times the significance of the transverse impact parameter
	$\langle q \rangle_{\text{SCT}w}$	Average charge of all tracks matched to the electron weighted by their number of hits in the SCT detector
	n_{dof}	Number of degrees of freedom in the track fit
Track-cluster matching	E_T / p_T	Ratio of the cluster transverse energy to the track transverse momentum p_T
	n_{tracks}	Number of tracks matching the electron candidate
Kinematics	E_T	Transverse energy of the electron measured by the calorimeter system. This variable is not used for discrimination purposes, but for the CNN to infer the E_T dependence of other variables.
	η	Pseudorapidity of the electron as measured by the calorimeter system. This variable is not used for discrimination purposes, but for the CNN to infer the η dependence of other variables.

by the `QuantileTransformer` class from the scikit-learn Python library^[289] and the network has been observed to perform better when this pre-processing of variables is performed^[2].

6.2.2 Track variables

The track information incorporated into these high-level variables considers only the track that best matches the electron candidate. However, during the reconstruction of electron candidates, multiple tracks can be associated to its main energy cluster. We note, in particular, that prompt electrons have on average 1.3 matched tracks, while background classes have 2.8 tracks^[2]. Thus, information on these extra tracks could be useful when discriminating prompt electrons from their backgrounds.

To do so, different features are considered for up to 5 tracks associated with each electron candidate. These variables, listed in Table 6.2, are ordered according to the distance of the track to the candidate position in $\Delta\eta \times \Delta\phi$. A grid of 5×13 features for the corresponding tracks is then used as input to the algorithm, and is referred to as the tracks image in what follows.

6.2.3 Calorimeter images

Beyond the extra tracks information, other low-level features registered by the detector that can be exploited by the electron identification algorithm are the energy deposits themselves. As described in Section 4.2, the electron candidate crosses different layers of both the EM and the hadronic calorimeters. In each one of these layers, energy deposits are recorded in a grid-like structure over η and ϕ , which can, in turn, be encoded as a two-dimensional image.

Table 6.2: Track variables included in the CNN electron identification algorithm. Variables for up to 5 matching tracks j are considered for each candidate and are organised in the form of a grid to form a 5×13 tracks image that is processed by a dedicated CNN layer^[2].

Variable	Description
p_j/E	Ratio of the momentum of track j to the energy of the electron candidate
$\Delta\eta_i$	$\Delta\eta$ between the track j and the electron candidate position
$\Delta\phi_i$	$\Delta\phi$ between the track j and the electron candidate position
n_{Pixel}^j	Number of hits in the Pixel detector
n_{SCT}^j	Number of hits in the SCT detector
n_{TRT}^j	Number of hits in the TRT detector
d_0^j	Transverse impact parameter relative to the beamline
$\sigma(d_0)_j$	Uncertainty on d_0
z_0^j	Longitudinal impact parameter relative to the beamline
q_j	Charge of the track
χ_j^2	χ^2 of the track fit
ndof_j	Number of degrees of freedom of the track fit
vtx_j	Matched vertex index number

The main challenge in this task is the fact that the number of layers traversed by the candidates varies as a function of $|\eta|$. This is observed, in particular, in the transition region with $1.37 < |\eta| < 1.52$, also referred to as the gap region, where there is an overlap of the barrel and the end-cap calorimeters. Moreover, the granularity of the energy cells is not the same for every layer and η region, being less finely arranged as we move away from the collision point.

Thus, calorimeter images are constructed by creating 2D arrays of finite values that correspond to the energy deposits in a given layer. If there is no deposit on a given region, the corresponding image is filled with zeros instead. In this way, all layers are included and can be used as input for the neural network. As a result, if the electron candidate is in the barrel region ($|\eta| < 1.3$), images associated with the end-cap region ($1.6 < |\eta| < 2.5$) are also generated, being completely filled with zeros.

Furthermore, all images are centred at the electron candidate cluster position and produced with a fixed granularity where each pixel corresponds to $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$, such that the final images cover $\eta \times \phi = 7 \times 11$. This value corresponds to the granularity of the second layer of the EM calorimeter, where most of the electron energy is deposited. In

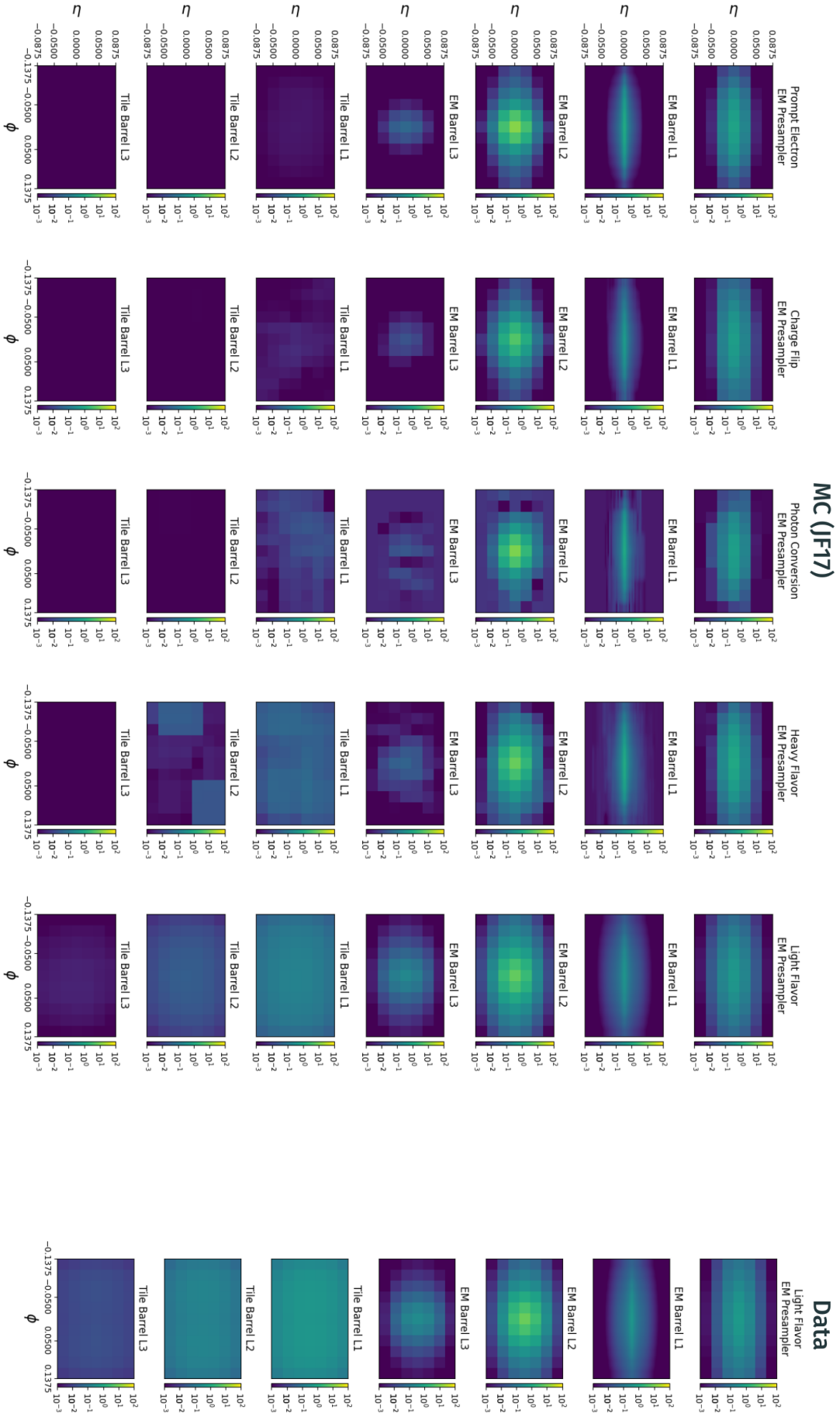


Figure 6.1: Average of the calorimeter images for the barrel layers for different classes. A total of 10^6 examples with $p_T \geq 4$ GeV and $|\eta| \in [0.0, 2.47]$ were considered from the MC simulated events in the JF17 dataset and data recorded by the ATLAS detector in 2015. The light-flavour class corresponds to the combination of both the light-flavour hadrons and their decays into e/γ .

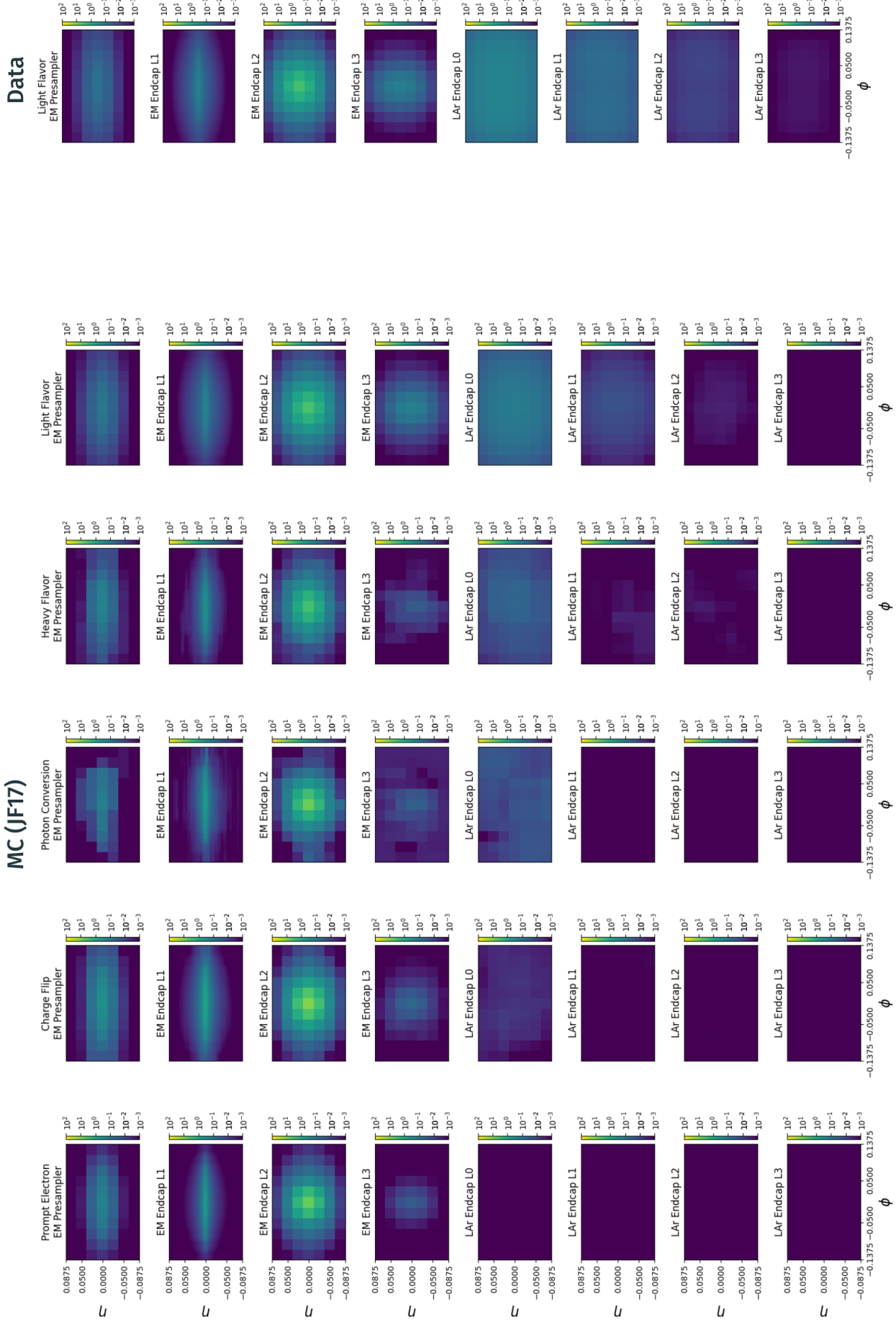


Figure 6.2: Average of the calorimeter images for the end-cap layers for different classes. A total of 10^6 examples with $p_T \geq 4$ GeV and $|\eta| \in [0.0, 2.47]$ were considered from the MC simulated events in the JF17 dataset and data recorded by the ATLAS detector in 2015. The light-flavour class corresponds to the combination of both the light-flavour hadrons and their decays into e/γ .

layers where the granularity is coarser, such as the case for the Tile calorimeter, the energy is equally distributed in the smaller pixels that would correspond to the same area. By including all layers with the same granularity, the network can capture the longitudinal development of the shower through the calorimeters. Additionally, the first layer of the EM calorimeter is available in a finer granularity, namely pixels of $0.025/8$ size, such that the total image corresponds to $\eta \times \phi = 56 \times 11$. This particular layer is also included in its full granularity to make sure no information is lost. As such, two fine-granularity images are included in the inputs, one for the barrel region and another for the end-cap region.

An average of the calorimeter images obtained for 10^6 electron candidates is shown in Fig. 6.1 and Fig. 6.2. The narrower shape of the prompt electrons in the second layer of the EM calorimeter (third row in both Fig. 6.1 and Fig. 6.2) is clear when compared to the different types of backgrounds. Moreover, the deposits in the Tile and LAr hadronic calorimeters (rows 5–7 in Fig. 6.1 and rows 5–8 in Fig. 6.2) that can be observed for all backgrounds are completely absent in the prompt electrons candidates. Meanwhile, the background classes present different levels of broader deposits that extend into the calorimeter layers. These observations further motivate the use of these images to improve the electron identification algorithm, which is described in more details in the following section.

6.3 Task and architecture

The inputs for the CNN algorithm are the high-level variables, the tracks image and the calorimeter images described above. By undergoing supervised training, the network performs a multinomial classification of the electron candidates, each class being defined using the truth information available in simulation data. The same classes used by the DNN implementation in ATLAS^[287] are used. They are defined as follows:

Prompt electrons: electrons produced at the interaction point, that are isolated and have the reconstructed charge matching the true charge. These include electrons from short-lived particles decays, processes such as $Z \rightarrow e^+e^-$, $W^\pm \rightarrow e^\pm\nu$, $J/\psi \rightarrow e^+e^-$ and Higgs boson decays, for instance, as well as electrons from a final state radiation or bremsstrahlung photons if they originate from a prompt electron;

Charge-flip electrons: prompt isolated electrons with the reconstructed charge opposite to the true charge. These include electrons whose charge is mis-measured due to material effects or from an error in the electron curvature measurement;

Prompt photon conversion: electrons from prompt photons which convert into a e^+e^- pair, as well as prompt photons which are reconstructed as an electron;

Electrons from heavy-flavour (HF) hadron decays: electrons produced in the decay of a b - or c -hadron (except the J/ψ), where these hadrons originate from the parton showering and hadronisation processes of jets;

Light-flavour (LF) hadrons: charged light-quark hadrons (e.g., $\pi^\pm$) that fake the electron signature. These correspond to the dominant background after the electron reconstruction is performed^[266].

Light-flavour (LF) e/γ : electrons produced from the decay of a light-quark hadron. This include, for instance, electrons from converted photons in decays such as $\pi^0 \rightarrow \gamma\gamma$;

These classes were engineered to address the background objects that most impact the measurements from the ATLAS physics programme. In the majority of cases, both the prompt electron class and the charge-flip electron class are considered signal electrons. Nevertheless, analyses featuring pairs of same-sign leptons^[290,291] would benefit from rejecting the contribution of Drell–Yan processes that leads to two opposite-sign leptons^[287]. As a result, considering the charge-flip electrons as a background class instead can reduce the contamination from these background events.

The architecture of the CNN algorithm is presented in Fig. 6.3. The matched track variables, arranged in a 5×13 grid, and the calorimeter images are processed in 3 independent convolutional sequences with 2 layers, such that the coarse resolution images ($\eta \times \phi = 7 \times 11$) from all layers and the fine resolution images ($\eta \times \phi = 56 \times 11$) from the first EM layer are processed separately. When processing the calorimeter images, the different layers are treated as different channels, similar to how coloured images are split into RGB channels. For both the coarse and the fine resolution calorimeter images, a kernel size of (3, 5) is used, while a (1, 1) kernel is used over the tracks image. A Max pooling layer reduces the dimensionality of the CNN layers output, an operation that keeps only the maximum value of the feature map in the region covered by the filter. The downscaling of the (x, y) dimensions is done by factors (4, 1) and (2, 1) for the coarse images and (1, 1) for all the layers of the other networks.

The outputs of the three convolutional layers are then flattened and concatenated with the HLV and processed through an FCN. This network is constructed with two hidden layers of 100 neurons and has as its output a vector of the probabilities of the electron candidate to be classified as each one of the six classes.

To implement the framework, Tensorflow version 2.1.0^[292] was used. After each hidden layer, a ReLU activation function is used, except for the last layer of the FCN which uses a Softmax activation function to normalise the output to a probability distribution over the output classes. The ADAM optimiser^[171] is used to perform the gradient descent, with the L_2 ^[172] and Dropout^[173] regularisation being used to reduce the risk of overfitting. A categorical cross-entropy is used as the loss function and the training convergence is assessed from the accuracy. The final model is obtained from the weights of the best epoch once the accuracy has not improved for five consecutive epochs.

The final discriminant $\mathcal{D}$ is obtained from the weighted probabilities for an electron

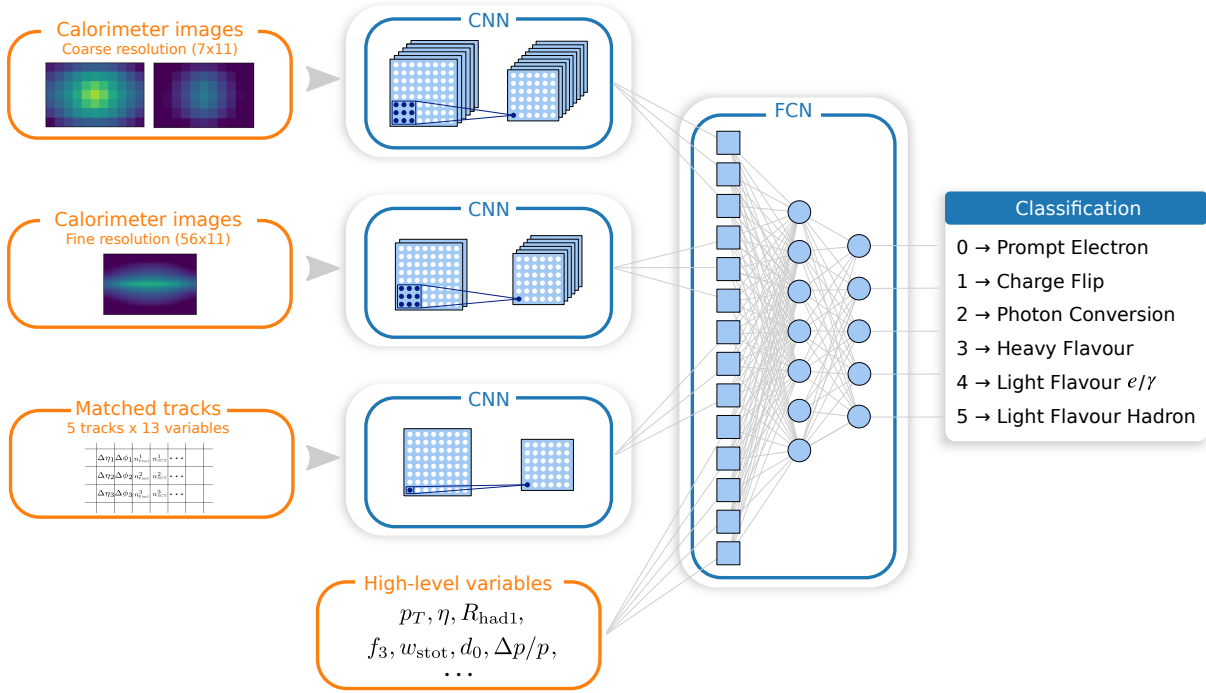


Figure 6.3: Architecture of the CNN algorithm. The calorimeter images in both coarse and fine resolutions, as well as the track variables from up to five matched tracks are processed through separate convolution layers. The corresponding outputs are concatenated with the high-level variables and processed by a FCN. The output is a vector with the probability of the electron candidate to be classified as each one of the six classes.

candidate to be classified into a signal p_{sig} or background p_{bkg} class, that is,

$$\mathcal{D} = \frac{\sum_c w_{sig}^c p_{sig}^c}{\sum_c w_{bkg}^c p_{bkg}^c}, \quad (6.1)$$

where the sums are performed over the corresponding classes c considered as either signal (numerator) or background (denominator), and w_{sig}, w_{bkg} are adjustable weights which add to unity. This discriminant is transformed via

$$\mathcal{D} \rightarrow \frac{\mathcal{D}}{1 + \mathcal{D}}, \quad (6.2)$$

such that the final discriminant is in the $[0, 1]$ interval.

This discriminant definition introduces a flexibility that allows it to be tuned according to each analysis' needs. By increasing the value of the weight associated with a given background, the corresponding background rejection for the class is increased. The optimal performance against the combined background classes has been found when the weights are set to the fractions that each class is found in the sample^[2].

Once the classification task is defined and the neural network is implemented, it is time to apply them to electron candidate examples. These are obtained from both the MC simulated

events and real experimental data, as described in the following section.

6.4 Datasets

As discussed in Chapter 2, the choice of examples of electron candidates for the training of the CNN algorithm is crucial to achieve a good performance. Not only is the quantity of examples provided important, but also its quality and resemblance to the final application examples are all determinant factors to be assessed when preparing the datasets.

Since a supervised learning is performed, the use of MC simulation events is convenient due to the availability of the truth information that allows the assignment of a true label to the electron candidates. Additionally, the large number of events generated by the ATLAS simulation pipeline guarantees large statistics for training. This is especially important given the large impact that the number of examples in the training sample has in determining a good performance of the algorithm^[2].

However, the known mismodelling of the MC simulation, presented in Section 3.2.6, introduces discrepancies between the examples seen during the training and the experimental data on which the network will be used. In particular, the simulation of the fake electron objects is known to be imperfect, especially those involving hadronic showers, since the calorimeter shower shapes obtained in the detector simulation are largely affected. Since the CNN inputs incorporate these low-level features of the detector, the trained model might be more sensitive to these discrepancies than those observed at the high-level features.

Some of the proposed solutions available to address this issue attempt to correct this effect on the high-level variables, similarly to what is done for the LLH approach described in Section 4.3.1. However, doing this at the shower-shape level adds considerable difficulty to the task, given the different cell configurations across the detector layers. An alternative approach would be to incorporate examples coming directly from experimental data. Since there is no truth information available from the experimental data, obtaining perfect labelling is impossible. However, a reliable labelling can be obtained by performing selection cuts that enhance the acceptance of examples from a particular class while rejecting the other classes.

Both these approaches, using MC simulated data and real experimental data, are used in this work. The labelling of the former is obtained from the truth information available in the MC simulation, while the latter was performed with requirements to select only the light-flavour hadrons and light-flavour e/γ background classes. The construction of these datasets is described in what follows.

6.4.1 Simulation data

Simulation of events that are known to produce electron candidates were used to obtain examples from all the pertinent classes. These events were processed using the ATLAS simulation

Table 6.3: Processes considered to obtain electron candidates from simulated MC events and the corresponding ratio of electron candidates from each class (in %), identified from its truth information. The large majority of background examples correspond to the light-flavour (LF) hadrons and their decays into e/γ , with prompt electrons also being well represented. Charge-flip, photon conversion and the heavy flavour (HF) decay classes correspond to a considerably smaller ratio of the candidates available^[2].

	Prompt $e^\pm$	Charge-flip	Photon conversion	HF decay	LF e/γ	LF hadron
JF17, JF35, JF50	2.5×10^{-3}	9.5×10^{-5}	0.05	0.20	7.67	31.28
$t\bar{t}$ ($W \rightarrow \ell\nu$)	2.19	0.07	0.05	1.22	5.44	25.89
W, Z production	10.45	0.36	0.04	0.03	1.48	3.08
$Z^*/\gamma^* \rightarrow e^+e^-$	3.93	0.21	8.6×10^{-3}	0.01	0.54	1.76
γ + jets	3.9×10^{-5}	3.8×10^{-6}	0.73	0.02	0.69	2.59
Total	16.57	0.64	0.88	1.49	15.82	64.60

infrastructure^[2,224], the software suite responsible for generating events and simulating the detector response, presented in Section 3.2.6. The processes considered include W and Z production decaying to leptons, the Drell–Yan $Z^*/\gamma^* \rightarrow e^+e^-$ process as well as $t\bar{t}$ production with at least one $W \rightarrow \ell\nu$ decay and the photon + jets production.

To further increase the number of background examples available, the dedicated JF17, JF35 and JF50 datasets were also included. They simulate $2 \rightarrow 2$ QCD processes, including multi-jet, $qg \rightarrow q\gamma$, $q\bar{q} \rightarrow g\gamma$, electroweak and top production processes. To guarantee the sample is enriched in electron candidates backgrounds, a requirement is made that the sum of the transverse energy of the particles (with the exception of muons and neutrinos) in a window of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ exceeds a lower energy threshold of 17, 35 and 50 GeV, for JF17, JF35 and JF50, respectively. These selections increase the acceptance of high- p_T background examples that mimic the highly localised energy deposits of electrons^[273]. Prompt electrons are not explicitly removed when constructing this sample, but after the cuts only very few remain.

In total, approximately 180×10^6 electron candidates are available from MC simulated events. The class composition of each process included is shown in Table 6.3. Note that the majority of the examples belong to the light-flavour hadron and light-flavour e/γ classes, with prompt electrons being also well represented. Each class has at least 10^6 electrons, guaranteeing that all classes are well represented in the training data. The distributions of E_T and $|\eta|$ for the different classes in the MC sample are shown in Figures 6.4a and 6.4b. We note that the LF examples are dominant at low energies, with prompt electrons and charge flips being dominant at high energy. Further, $|\eta|$ has an uniform distributions for all classes except charge-flip, with a prominent drop in the gap region $1.3 < |\eta| < 1.6$ and the last bin.

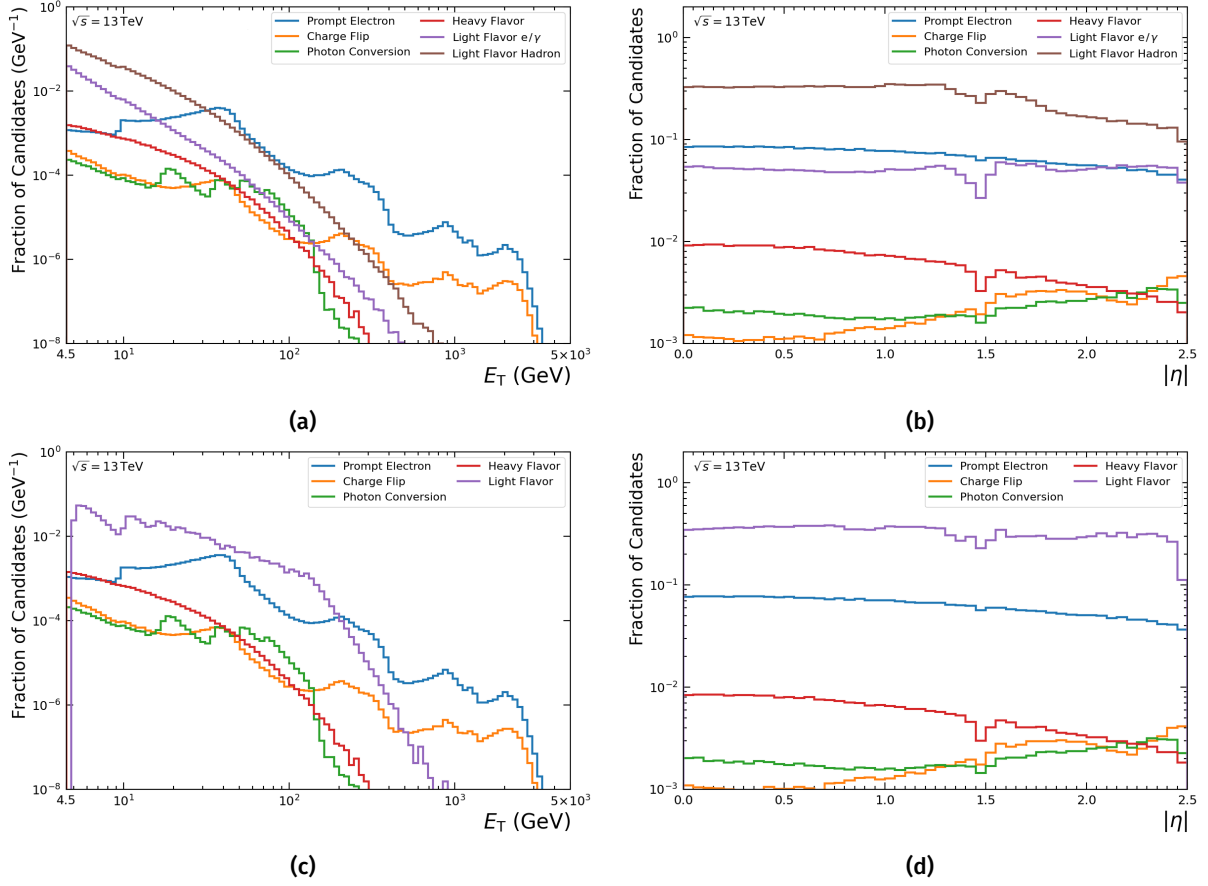


Figure 6.4: Distributions of the (a) transverse energy E_T and (b) the absolute pseudorapidity $|\eta|$ of the MC simulation examples for different classes. The same observables are shown in (c) and (d) for the dataset where the combined LF examples were obtained from real data.

6.4.2 Experimental data

By construction, it is impossible to obtain labelled examples from data since there is no information on the exact process that originated an electron candidate at the collision point. However, as mentioned earlier, we can obtain a sample enriched in a given type of electron candidate by performing well-engineered selections that maximise the acceptance of electron candidates from a specific class.

Such a sample was obtained for the most common electron backgrounds: the light-flavour hadron and e/γ (LF) classes. These classes are particularly difficult to simulate, especially hadrons faking the electron signature, due to the badly modelled calorimeter shower shapes^[266,277], as previously mentioned in Section 5.1. Nevertheless, they can be isolated using selections that maximise the presence of electron candidates while also reducing the number of prompt electrons.

This LF-enriched data sample was obtained from experimental data registered by the ATLAS detector in 2015. In particular, the dataset used contains a subset of events that are enriched in backgrounds that match the electron signature. These are selected by requiring that the event passes at least one of the HLT triggers designed to have a loose selection and,

as such, be completely dominated by hadronic backgrounds. Such triggers require a lower energy threshold on either the transverse energy E_T or passing the `Loose` LLH identification working point. Furthermore, the events are also required to have at least one reconstructed electron with $p_T > 4.5$ GeV.

When constructing this data sample, a pre-selection was made requiring good reconstruction quality, as well as a minimum number of hits in the inner detector, namely $n_{\text{Pixel}} \geq 1$ and $n_{\text{Pixel}} + n_{\text{SCT}} \geq 7$. Furthermore, only the candidates with $|\eta_{\text{BE2}}| < 2.47$ are kept⁽³⁴⁾. Next, to guarantee that the majority of examples included correspond to LF backgrounds, further selections were incorporated^[293]. To remove electrons coming from the W boson decay, events are required to have small missing transverse energy, $E_T^{\text{miss}} < 25$ GeV, and transverse mass of the electron candidate with the missing energy of $m_T^{e\nu} < 40$ GeV. Similarly, prompt electrons coming from the Z boson decay are excluded from the sample by requiring that no more than one electron passes the medium LLH working point and, in the case one passes the medium LLH working point, it is required that the invariant mass of its combination with another electron candidate in the event, the latter passing a `Tight` LLH working point, to be $m^{ee} < 70$ GeV or $m^{ee} > 110$ GeV.

Events that pass all the imposed requirements correspond to LF candidates, with a purity estimated at 99%^[293]. Thus, these candidates are labelled as LF class and considered as examples for the CNN network. A dataset is then constructed by replacing both the LF classes obtained from MC simulation with these examples retrieved directly from experimental data. The E_T and $|\eta|$ distributions obtained in this dataset are shown in Figures 6.4c and 6.4d. This dataset containing LF examples from data and the remaining examples from MC simulation are referred in the following sections as the “Data” dataset, for simplicity, while the one containing only examples from MC simulation is referred to as the “MC” dataset.

To assess the agreement between the LF samples obtained from Data and MC, distributions of the LLH discriminant value (described in Section 4.3.1) are shown in Fig. 6.5 for selected bins in E_T . We observe an overlap between the distributions of the LF class from MC and Data, with both of them having a clear distinction from the prompt electrons distribution, as expected. However, for lower energies (Fig. 6.5a), the LF class distribution from Data is slightly shifted towards the signal region, an effect that is not observed as we increase the energy (Fig. 6.5b). In summary, these LLH discriminant distributions further corroborate that the examples obtained from experimental data are indeed mainly populated by the LF class and thus a comparison between MC and Data examples of this class can be made.

There are two main advantages in producing a sample containing examples from experimental data: benchmarking the model trained in MC simulation in a realistic scenario and incorporating more realistic examples in the training. Both these studies will be discussed in

(34) The η_{BE2} variable is the pseudorapidity of the barycentre in the second layer of the EM calorimeter, where the majority of the electron energy is deposited. This variable is known to give a better proxy for the location of the cluster of an electron in the EM calorimeter. It is the same one used in the electron efficiency measurements discussed in Chapter 5.

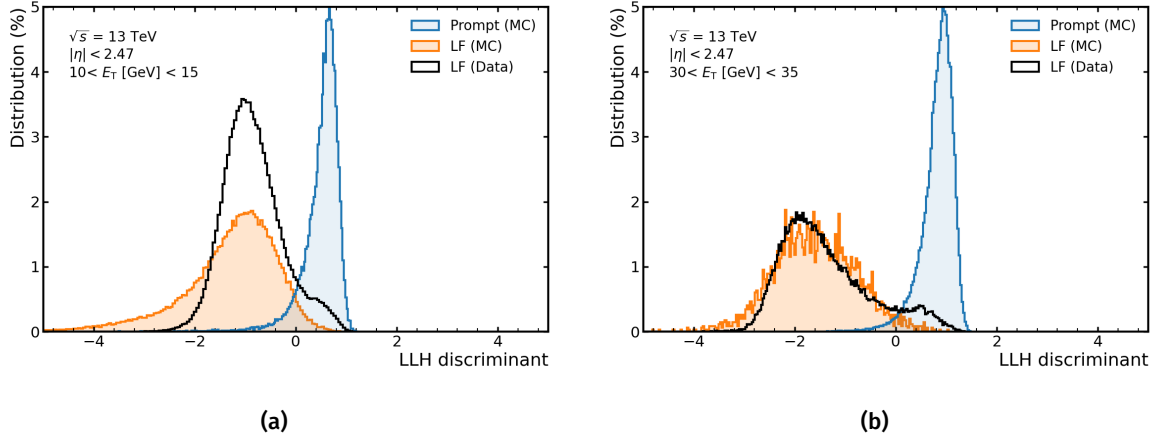


Figure 6.5: Distributions of the LLH discriminant value for the LF class obtained from Data (black line) and MC (orange line), as well as the one for prompt electrons (blue line) from MC events for selected bins in E_T : (a) $10 \text{ GeV} < E_T < 15 \text{ GeV}$ and (b) $30 \text{ GeV} < E_T < 35 \text{ GeV}$.

more details in the following section.

In particular, we highlight that the contribution from the author of this thesis to this project consisted on the study and optimisation of the input features used by the algorithm, a work that is described in Section 6.5.2. Moreover, the author also contributed to the comparison of the network’s performance when evaluating over examples from MC and Data, as described in Section 6.5.3.

6.5 Results

6.5.1 Performance over simulation data

The performance of the CNN algorithm is obtained by evaluating examples on a testing dataset that is statistically independent to the one used for the training. The available electrons are randomly split with approximately 90% being used for the training and 10% being used for the testing dataset. Track quality is required for the testing dataset, with the number of hits in the pixel and SCT detector satisfying $n_{\text{Pixel}} \geq 2$ and $n_{\text{Pixel}} + n_{\text{SCT}} \geq 7$.

When comparing the performance of the CNN algorithm in different studies, the main figure of merit considered is the *background rejection*. As discussed in Section 2.5, this quantity is defined as the inverse of the efficiency of the background, that is,

$$\frac{1}{\epsilon_{\text{bkg}}} = \frac{\text{TN} + \text{FP}}{\text{FP}} = 1 + \frac{\text{TN}}{\text{FP}}$$

where TN and FP correspond to the number of true negatives (specificity) and false positives (type-1 error), respectively. Since the background rejection increases monotonically with the number of candidates correctly classified as background, when compared at a fixed signal

efficiency, a greater value of the background rejection corresponds to a better performance of the network. As a result, ROC curves are drawn showing how the background rejection varies as a function of the signal efficiency for each one of the background classes.

In Fig. 6.6, the performance of the CNN algorithm is presented when evaluated over a testing dataset with MC simulated electrons with $E_T > 15$ GeV and considering only prompt electrons as signal. Models were trained with different combinations of the input features (HLV, tracks image and/or calorimeter images), allowing for the quantification of their individual impact. The background p_T distributions were reweighed to match a 4:1 ratio with respect to the prompt electron distributions. This was done to guarantee that classes that have fewer examples (charge flip, photon conversion and heavy flavours) will be balanced and achieve a comparable performance. The performance was then obtained over a sample statistically independent from the training sample, having the same class proportions but with no kinematic reweighing applied.

As shown in the ROC curves, the model with all input features (solid blue line) outperforms all the others when rejecting both the combined background and the individual backgrounds separately. Moreover, all CNN models considerably improve the background rejection when compared to the LLH algorithm (solid gray line) and the corresponding working points (black markers).

This improvement is further quantified in Figs. 6.7 and 6.8 via the ratios of the background rejection of the combined background for the full CNN model, trained with HLV, tracks image and calorimeter images, with respect to the one obtained from the LLH method and an FCN with the HLV as input features. In Fig. 6.7, ratios are shown when rejecting the combined background (Fig. 6.7a) and the LF hadron class (Fig. 6.7b), considering different E_T bins and the LOOSE, MEDIUM and TIGHT working points. Furthermore, ratios with respect to the LLH method (Fig. 6.8a) and the FCN (Fig. 6.8b) are also shown for different E_T and $|\eta|$ bins for a signal efficiency corresponding to the LOOSE LLH working point, of around 90%.

We observe that the CNN outperforms both the LLH and FCN methods in all bins by factors of at least 1–2. When rejecting the combined background, an improvement of order 3–7 is obtained when comparing to the LLH, while the introduction of the images is responsible for factors 1–2 when compared to the FCN results. The ratios are still higher when rejecting the LF hadron class, with factors around 10 for LOOSE and factors around 5 for TIGHT. Lastly, we highlight the stability of the improvement across E_T and $|\eta|$ for the LOOSE working point. At this value of the signal efficiency, most prompt electrons are accepted by the classifier, so the improvement in background rejection shows that the contamination from the LF class is considerably reduced by when using the CNN model.

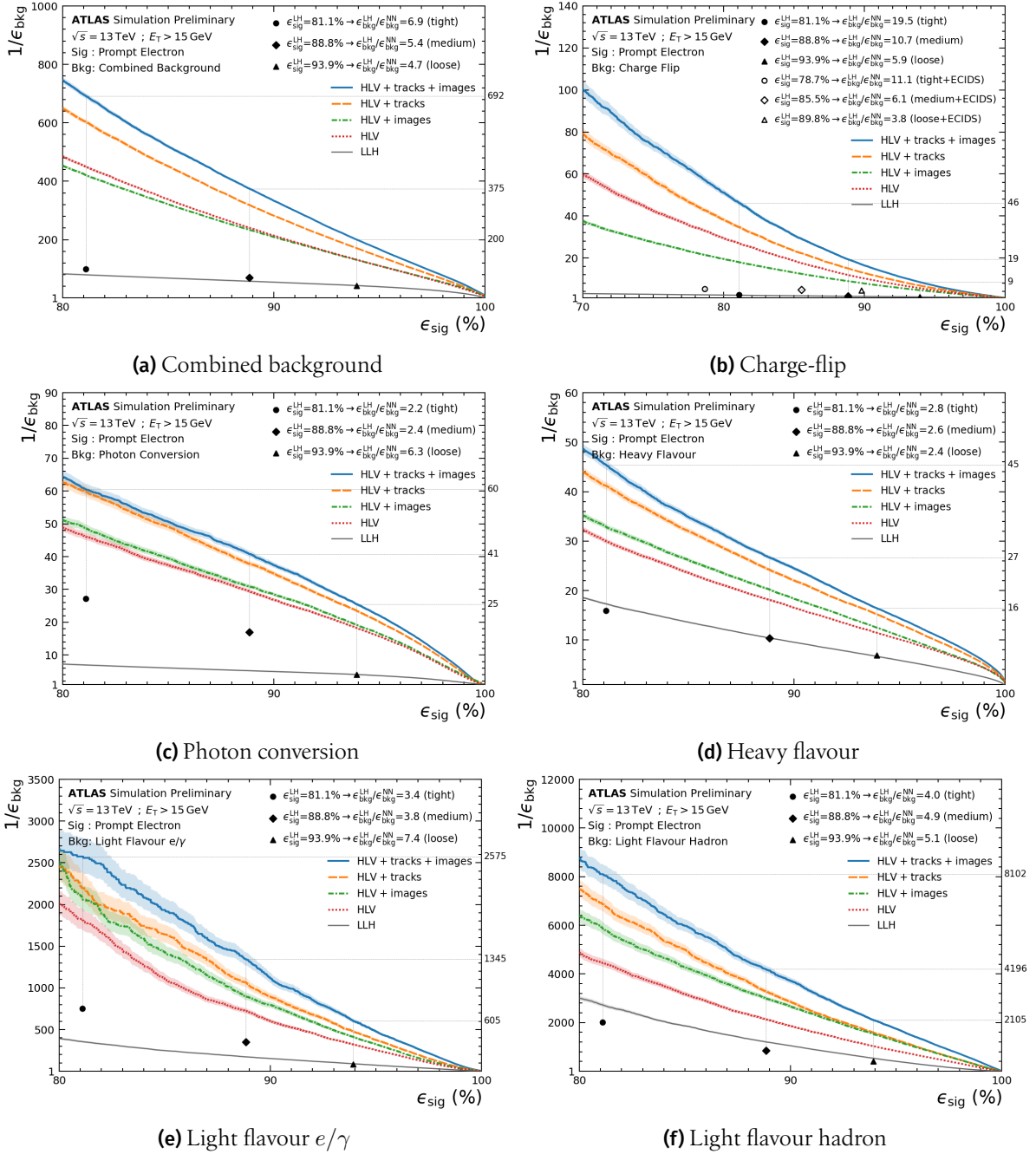


Figure 6.6: ROC curves from a CNN model trained and evaluated over MC simulated examples when rejecting the different background classes. The testing dataset consisted of electrons with $E_T > 15$ GeV and charge-flip electrons are considered a background class. Models were trained with different combinations of the input features, namely the HLV only (dotted red line), HLV + images (dash-dotted green line), HLV + tracks (dashed orange line) and the full CNN algorithm with HLV + tracks + images (solid blue line). The bands around the curves correspond to the statistical uncertainty associated with the size of the testing dataset. The values obtained from the LLH working points at different signal efficiencies are drawn, namely LOOSE (triangle), MEDIUM (diamond) and TIGHT (circle), together with the results obtained by the LLH discriminant (solid grey line). Taken from Ref. [2].

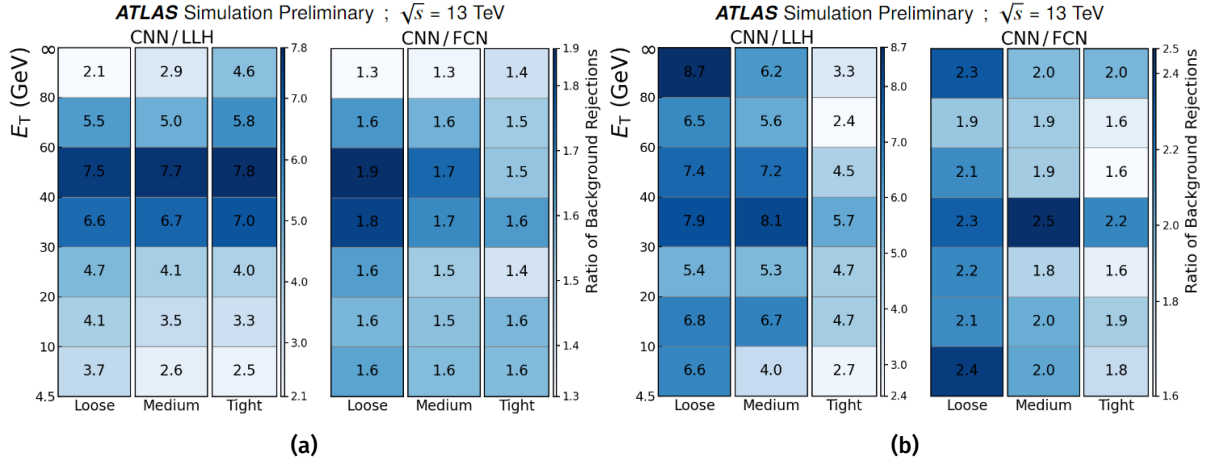


Figure 6.7: Ratios of the background rejection of the (a) combined background and the (b) LF hadron class from a CNN model with respect to the LLH method and FCN network for different E_T bins. The CNN model was trained with HL ν , tracks and images as input features, while the FCN values were obtained from a network trained only with HL ν . Values were obtained for the signal efficiency corresponding to the LOOSE, MEDIUM and TIGHT LLH working points. Taken from Ref. [2].

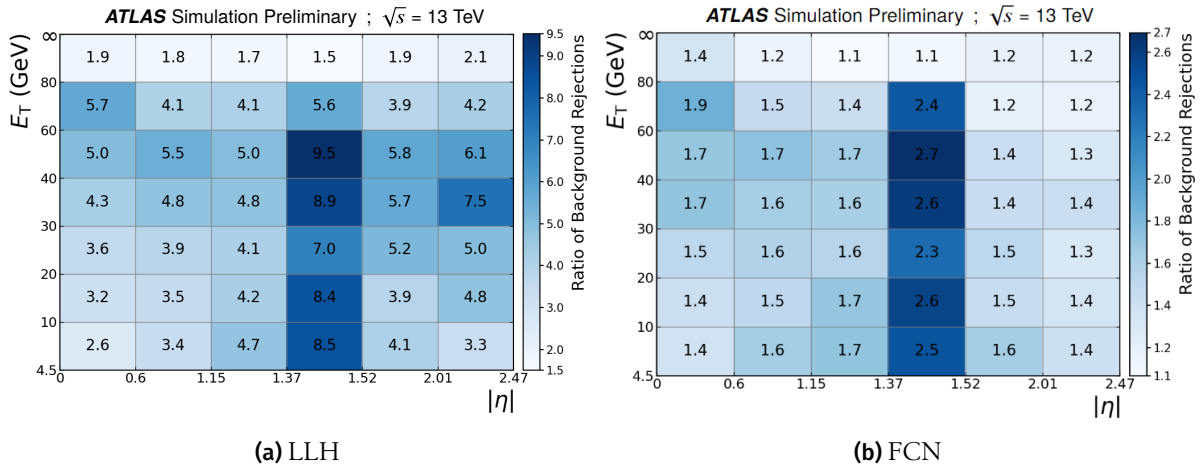


Figure 6.8: Ratios of the background rejection of the combined background from a CNN model with respect to (a) LLH method and (b) FCN network as a function of different E_T and $|\eta|$ bins. The CNN model was trained with HL ν , tracks and images as input features, while the FCN values were obtained from a network trained only with HL ν . Values were obtained for the signal efficiency corresponding to the LOOSE LLH working point. Taken from Ref. [2].

6.5.2 Feature ranking

When developing a neural network, the choice of input features is central in determining how well the task in question is performed, as discussed in Chapter 2. In particular, the process of feature engineering can be optimised by studying the impact of each individual feature on the performance figure of merit. Such a study allows not only for a better understanding of what drives the performance of the network, but also a potential removal of superfluous inputs, entailing a reduction of computational resources needed by the algorithm.

In the CNN algorithm, we have a total of 45 input features, namely 26 scalars from the HLV, 16 coarse images, 2 fine images and 1 tracks image. To assess the influence of each one of these features in the network's background rejection, a feature-ranking study was performed. In this study, we measured the ratio of the background rejection at a fixed signal efficiency obtained from a model trained with all the features with respect to the one obtained from a model trained without the respective feature, that is,

$$R = \frac{\text{background rejection with all the features}}{\text{background rejection without the feature}}, \quad (6.3)$$

such that

- if $R > 1$, the network performs better with the feature;
- if $R = 1$, the removal of the feature does not affect the performance;
- if $R < 1$, the network performs better without the feature.

Models were trained to discriminate the six classes with 20×10^6 electrons, a smaller training dataset size that allowed for faster training and less computing resources without compromising the baseline performance. Moreover, the background p_T distributions were also reweighed to match a 4:1 ratio with respect to the prompt electron distributions. The performance was evaluated over a testing dataset with 100×10^6 examples, where the background rejection was obtained at 90% signal efficiency over examples within the whole η region ($|\eta| < 2.5$), but also separately for the barrel ($|\eta| < 1.3$), gap ($1.3 < |\eta| < 1.6$) and end-cap ($1.6 < |\eta| < 2.5$) regions.

The results are shown in Figures 6.9, 6.10, 6.11 and 6.12. The variables are sorted in descending order of the ratio R when discriminating the combined background, while also showing their corresponding values for the individual background classes. Variables for which $R > 1$, drawn with a bar to the right of the reference value ($R = 1$, dotted red line), are important for the performance of the network, since their removal decreases the background rejection. In contrast, variables for which $R < 1$, with bars drawn to the left of the reference, actually improve the performance when removed, since the background rejection is higher without them.

We note that the pattern of influential features is kept across the different classes and

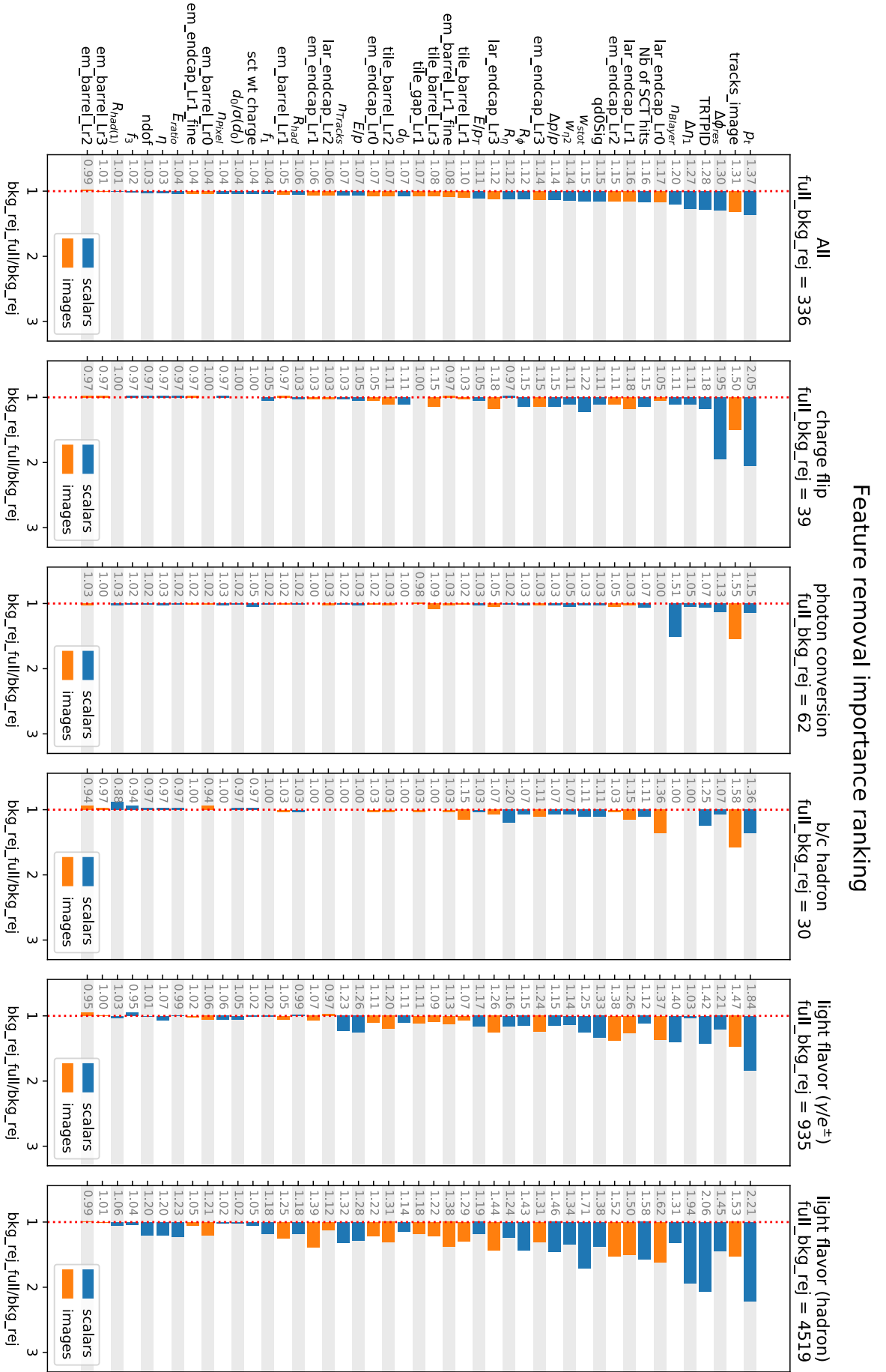


Figure 6.9: Ranking of the input features according to the ratio R defined in Eq. (6.3) when discriminating the combined background and its different classes. Features include both the HLIV (blue) and the images from the track-matching variables and calorimeter deposits (orange). They are presented in descending order of the ratio for the combined background, such that those listed higher have a larger impact when removed. Results are presented for electrons with $|\eta| < 2.5$, corresponding to the whole coverage.

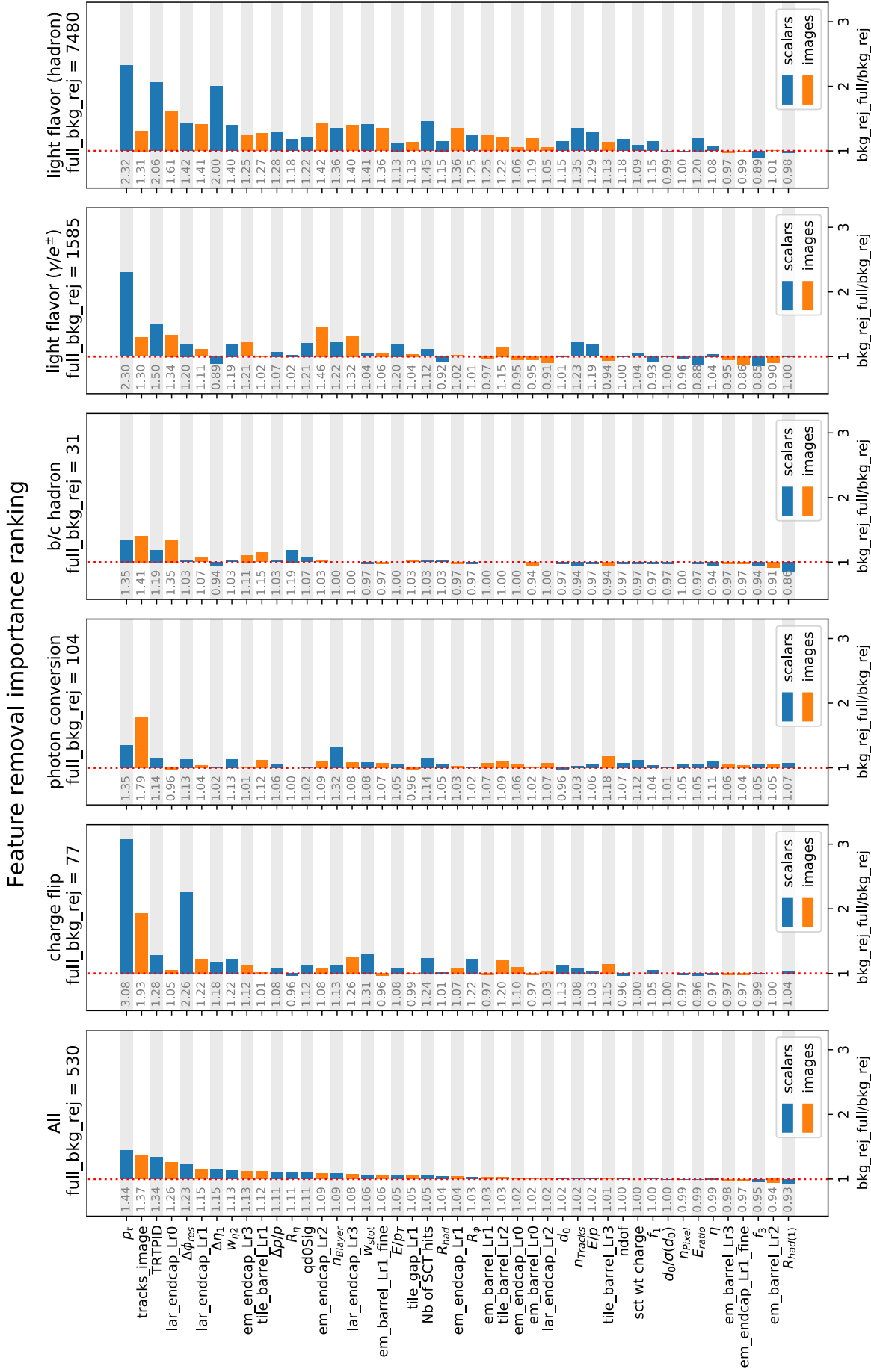


Figure 6.10: Ranking of the input features according to the ratio R defined in Eq. (6.3) when discriminating the combined background and its different classes. Features include both the HLV (blue) and the images from the track-matching variables and calorimeter deposits (orange). They are presented in descending order of the ratio for the combined background, such that those listed higher have a larger impact when removed. Results are presented for electrons with $|\eta| < 1.3$, corresponding to the barrel region only.

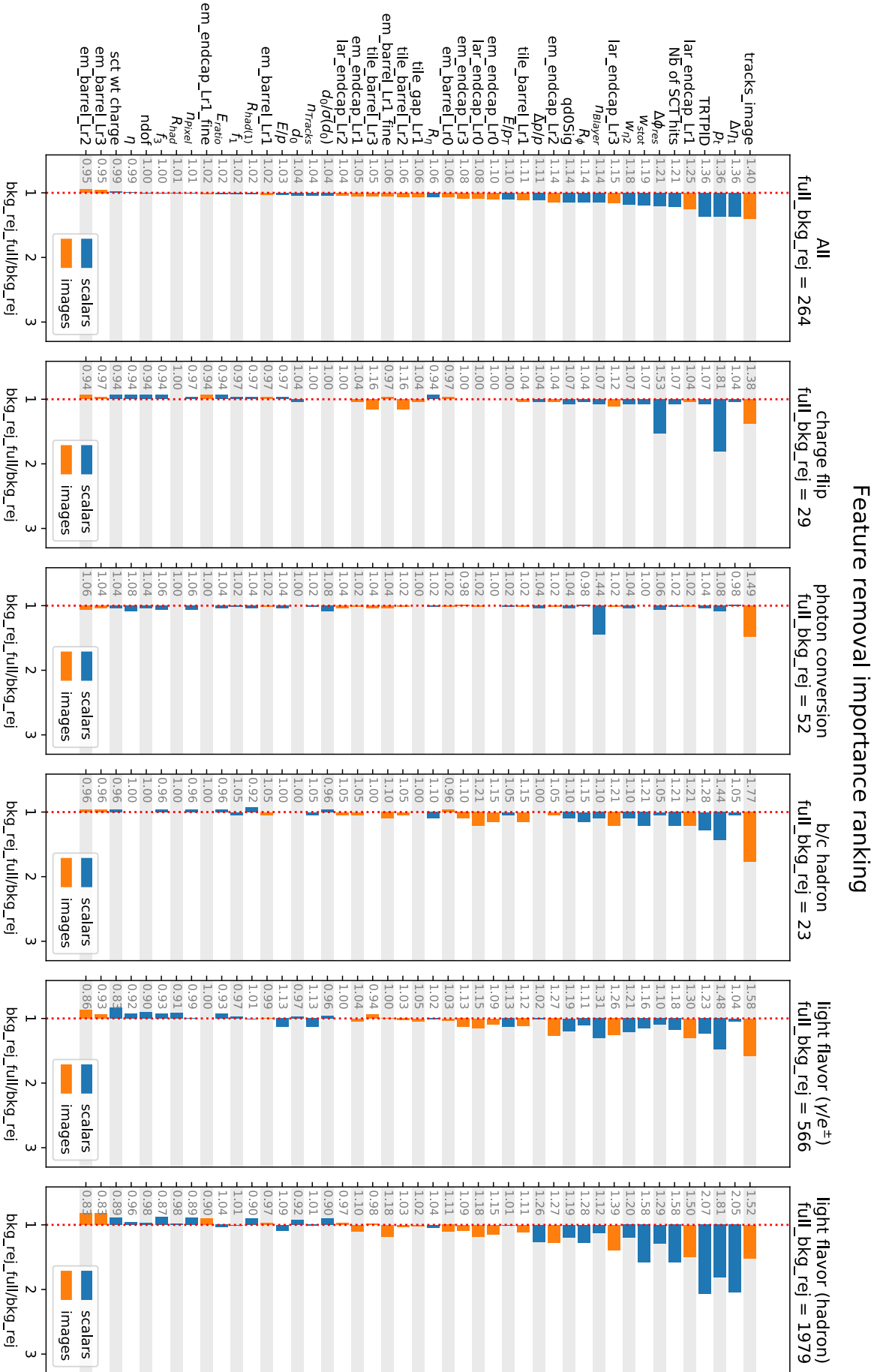


Figure 6.11: Ranking of the input features according to the ratio R defined in Eq. (6.3) when discriminating the combined background and its different classes. Features include both the HLIV (blue) and the images from the track-matching variables and calorimeter deposits (orange). They are presented in descending order of the ratio for the combined background, such that those listed higher have a larger impact when removed. Results are presented for electrons with $1.3 < |\eta| < 1.6$, corresponding to the gap region.

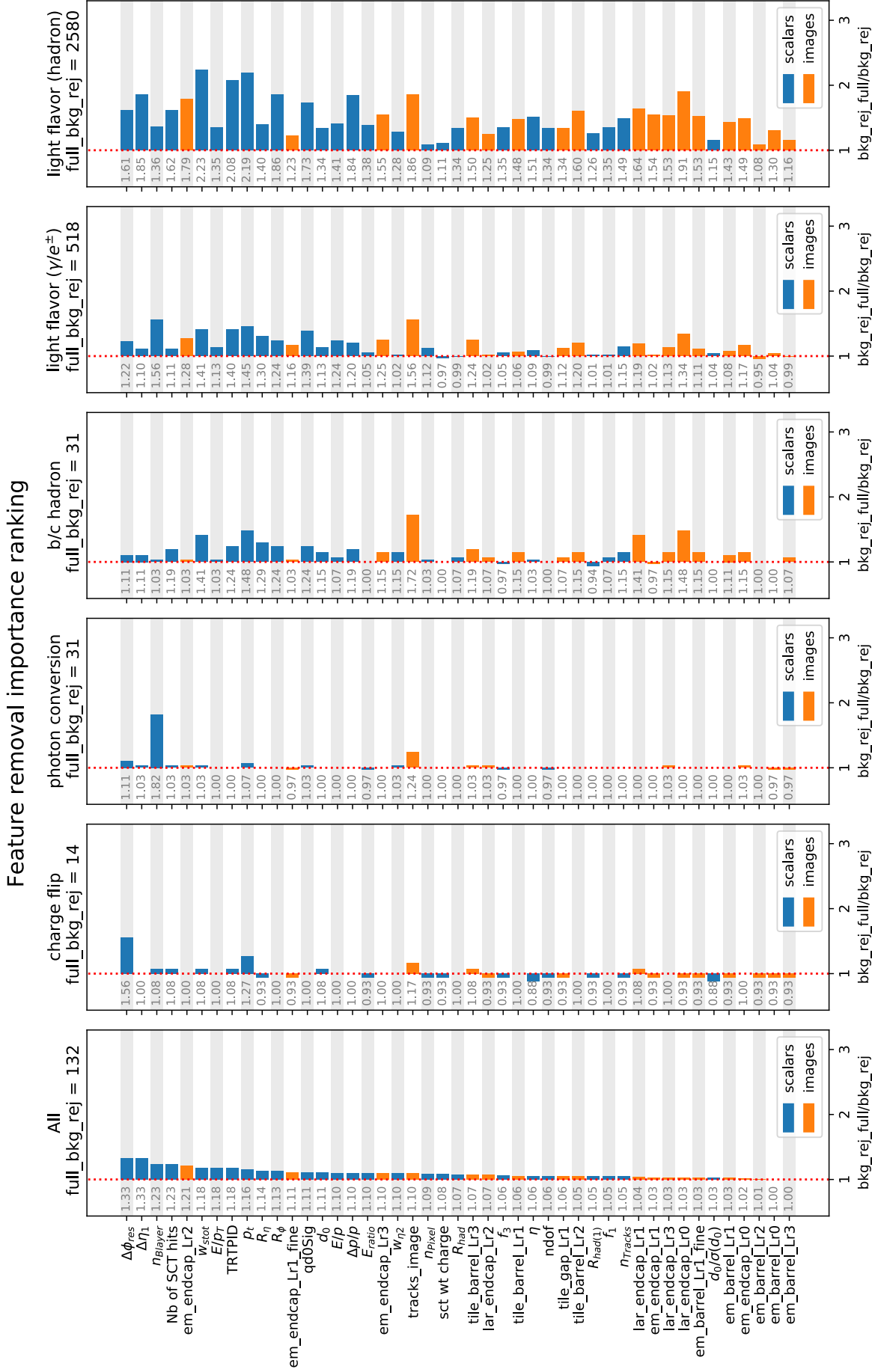


Figure 6.12: Ranking of the input features according to the ratio R defined in Eq. (6.3) when discriminating the combined background and its different classes.

Features include both the HLV (blue) and the images from the track-matching variables and calorimeter deposits (orange). They are presented in descending order of the ratio for the combined background, such that those listed higher have a larger impact when removed. Results are presented for electrons with $1.6 < |\eta| < 2.5$, the end-cap region.

η regions. Also, no feature has both minimum and maximum ratios across all classes and all regions around $R = 1$, meaning that all features are important in at least one scenario. Moreover, there is no feature that has a non-negligible detrimental contribution ($R < 0.9$) that is not important in another scenario ($R > 1.0$).

We observe that the tracks image is particularly important, having a large impact ($R > 1.3$) in all regions considered and for all backgrounds. Since they have a dedicated CNN to process them, their removal is associated with less weights for the model to learn. In fact, the total number of parameters goes from 3.5×10^6 to 0.9×10^6 with their removal, a considerable reduction of the model size that matches the observation of a worse performance. Also, the calorimeter images are shown to be important features especially for both the LF classes.

In light of these results, all input features were kept in the CNN algorithm. This study also further established that the addition of the tracks and calorimeter images are important elements in improving the performance of the algorithm when rejecting backgrounds from the different classes.

6.5.3 Performance in data versus simulation

The results presented so far show that the CNN algorithm considerably improves the discrimination of electrons from its different types of background with respect to the LLH method. However, the ultimate goal of the algorithm is to be applied to real experimental data collected by the ATLAS detector and, as such, it is important to test the CNN algorithm in a realistic scenario. This was achieved by comparing the performance over the dataset containing only MC simulated examples and the one where the LF class was replaced with examples obtained from experimental data.

Since the background examples we have from the experimental data are coming from the LF class, the figure of merit in which we are mostly interested is the background rejection with respect to this class. To guarantee a fair comparison, the classifications of the two light-flavour classes were combined for the model trained in simulated MC examples. Moreover, only examples coming from the JF17 simulated dataset were included, a choice that was made since their distributions have been observed to be closer to those coming from real data^[293]. Different models were trained, either over the MC simulated dataset or with the dataset that had LF examples from data, both including approximately 90% of the available examples. The remaining 10% examples were considered as testing datasets, with both models being evaluated to reject either LF examples from the MC or Data samples, described in Section 6.4.

The ROC curves corresponding to the background rejection of the LF class in the different scenarios are shown in Fig. 6.13. Results are presented for models trained with HLV only, a combination of HLV and the tracks image, as well as the full CNN algorithm with the HLV, the tracks image and the calorimeter images. The corresponding values obtained from the LLH working points are also included for comparison. Also, the discriminant distributions

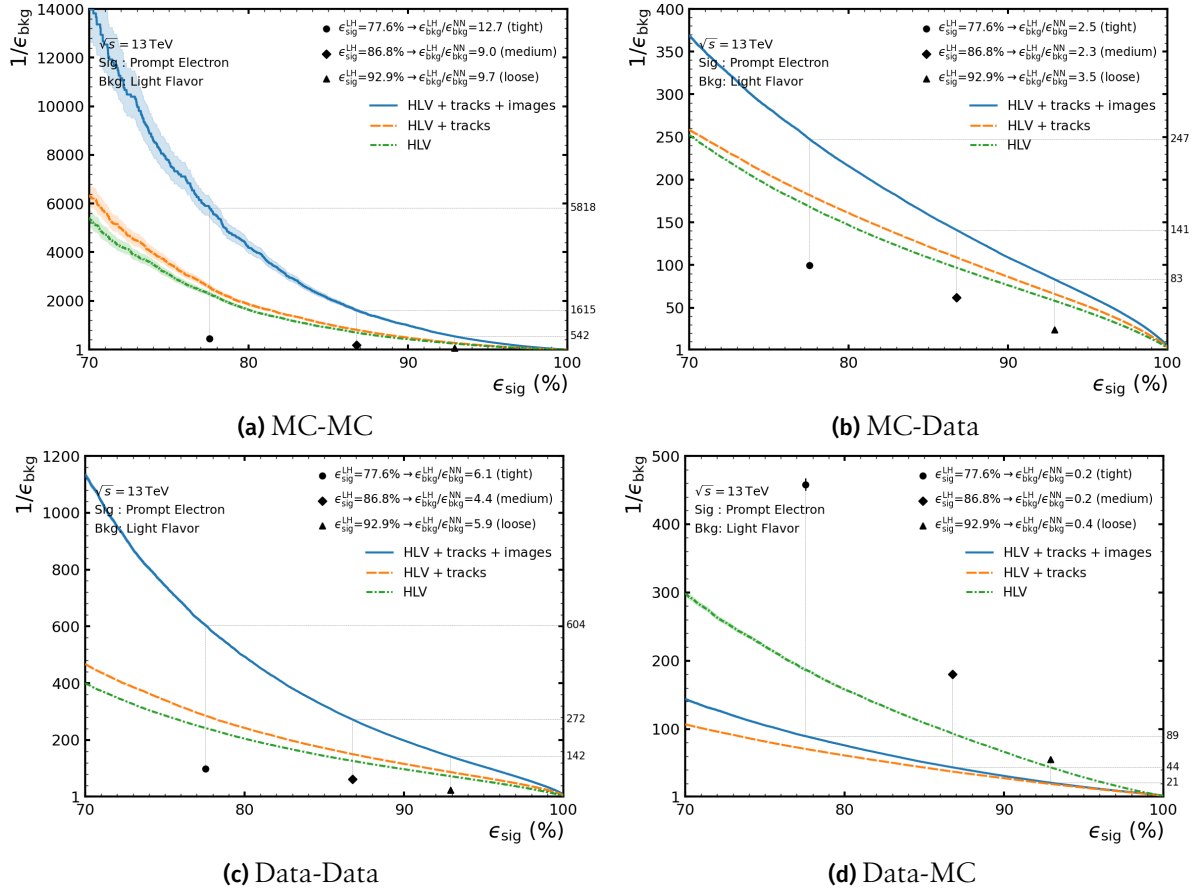


Figure 6.13: ROC curves from CNN models when rejecting the combined LF class. Models trained in MC are evaluated in (a) MC and in (b) Data, as well as models trained in Data are evaluated in (c) Data and in (d) MC. Different combinations of the input features are included, namely the HLV only (dash-dotted green line), HLV + tracks (dashed orange line) and the full CNN algorithm with HLV + tracks + images (solid blue line). The bands around the curves correspond to the statistical uncertainty associated with the size of the testing dataset. The values obtained from the LLH working points at different signal efficiencies, namely LOOSE (triangle), MEDIUM (diamond) and TIGHT (circle) are drawn.

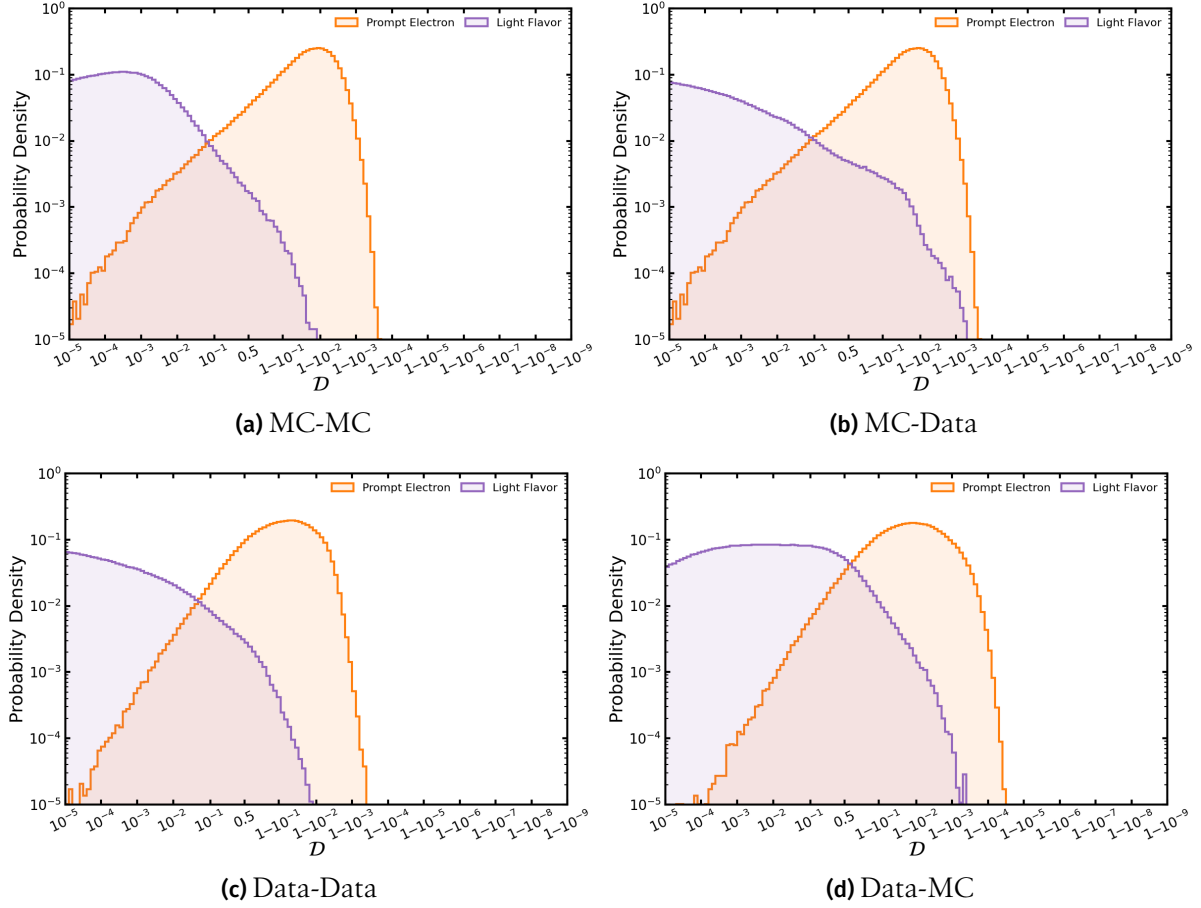


Figure 6.14: Discriminants of the CNN model with all input features (HLV, tracks and calorimeter images) when classifying prompt electrons from the combined light-flavour class. Results for a model trained in MC and evaluated on (a) MC and (b) Data, as well as a model trained in Data and evaluated on (c) Data and (d) MC.

obtained by the full CNN model for the prompt electrons and the LF class are presented in Fig. 6.14.

The first test in which we are interested is how the performance of the model trained in MC examples is impacted when it is evaluated over examples coming from Data. We observe that the great performance of this model over the MC simulated examples (Fig. 6.13a) considerably drops when discriminating the examples coming from Data (Fig. 6.13b). Nevertheless, the values obtained for all these models are still considerably higher than the ones obtained from the LLH working points.

If instead we incorporate the examples from data into the training, the rejection of the LF class in data (Fig. 6.13c) is recovered for all models. In particular, at the signal efficiency associated with the Loose LLH working point (at 92%), the CNN performance with all the input features is 6 times the value obtained by the LLH method, compared to a ratio of 3.5 obtained for the model trained with simulated examples.

However, when the model trained with LF examples from data is tested to discriminate the MC simulated examples (Fig. 6.13d), a worse background rejection is observed. This

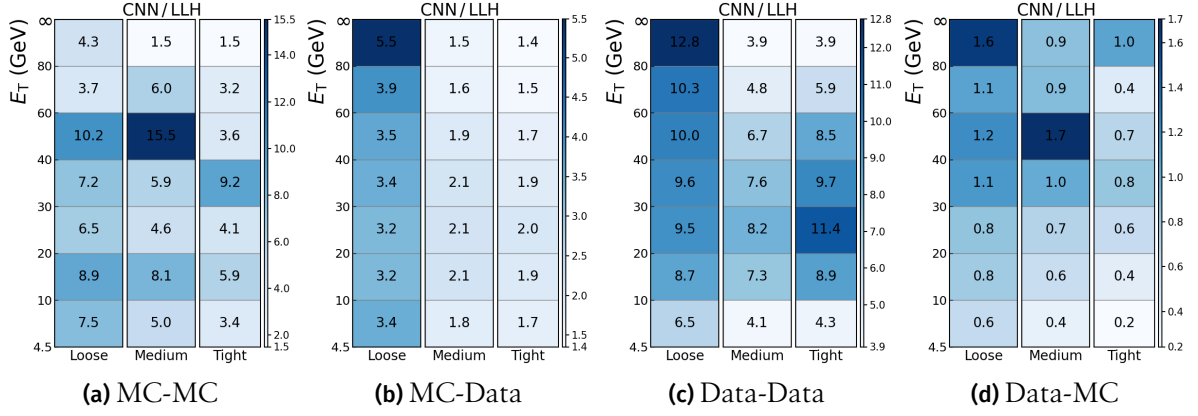


Figure 6.15: Ratios of the combined LF class background rejection from a CNN model with respect to LLH methods as a function of E_T . The CNN model was trained with HLX, tracks and images as input features and results are shown for a model trained in MC and evaluated on (a) MC and (b) Data, as well as a model trained in Data and evaluated on (c) Data and (d) MC. Values were obtained for the signal efficiencies corresponding to the LLH working points, namely LOOSE, MEDIUM and TIGHT.

happens for all the combinations of input features, with the values being below the LLH ones. Although not ideal, this result shows that the trained models are not as general as they could be and the model does not seem to be completely robust against any example, independently if they come from MC simulation or experimental data. This lack of generalisation, however, does not demerit the good performance achieved for the model when evaluated over data, since this is the final application dataset for which we expect the algorithm to be deployed.

Additionally, the ratios of the LF-class background rejection of the full CNN algorithm with respect to the LLH values as a function of E_T are shown in Fig. 6.15. An improvement when using the CNN algorithm is observed throughout the energy spectrum and for all the working points. For the LOOSE working point, in particular, the observed improvement when applying on data goes from a factor of 3 to a factor of 10 when comparing the model trained in simulation MC (Fig. 6.15b) to the one trained with Data examples (Fig. 6.15c). Moreover, the ratios obtained for the model trained and evaluated in data surpass the ones obtained when training and evaluating with simulated examples. These results demonstrate that incorporating real data examples in the training can considerably improve the rejection of the LF class in data applications.

The improvement with respect to the LLH method is further quantified in terms of both E_T and $|\eta|$ bins, as shown in Figs. 6.16 and 6.17 for the LOOSE and TIGHT working points, respectively. The LF class background rejection obtained at the signal efficiency corresponding to both the LOOSE and TIGHT working points is consistently greater than the LLH ones across $|\eta|$. In other words, this shows that in applications that need more electrons to be accepted (i.e., higher signal efficiency), the contamination of LF class is considerably reduced. Improvements with factors 4–8 for LOOSE (Fig. 6.16c) and 3–6 for TIGHT (Fig. 6.17c) are observed in the barrel region ($|\eta| < 1.3$) when training and evaluating in Data, while the training in MC presents

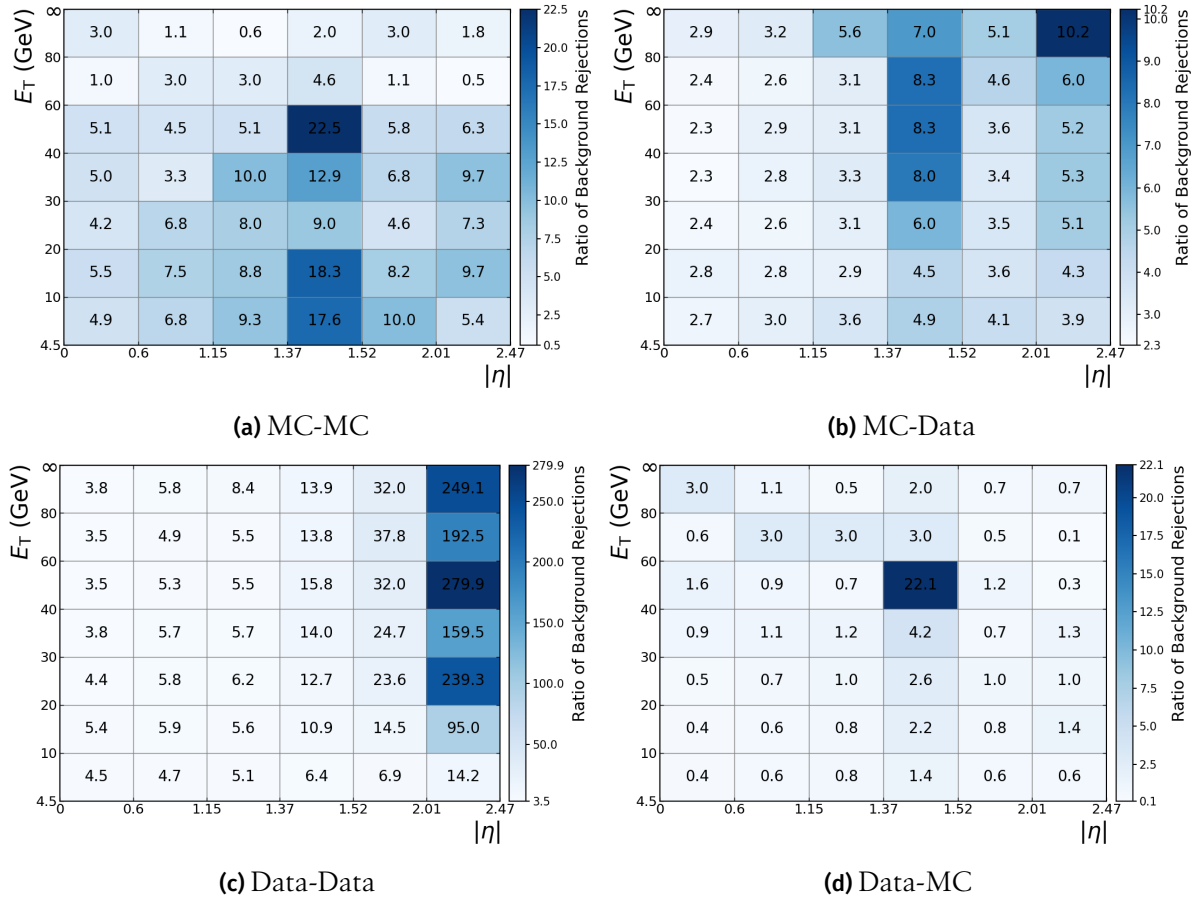


Figure 6.16: Ratios of the combined LF class background rejection from a CNN model with respect to LLH methods as a function of E_T and $|\eta|$ bins. The CNN model was trained with HL ν tracks and images as input features and results are shown for a model trained in MC and evaluated on (a) MC and (b) Data, as well as a model trained in Data and evaluated on (c) Data and (d) MC. Values were obtained for the signal efficiencies corresponding to the LLH Loose working points.

improvements of factors 2–6 for Loose (Fig. 6.16b) and 1–2 for Tight (Fig. 6.17b) when rejecting Data examples.

We note, in particular, that in the forward region, where $|\eta| > 2.0$, we observe improvements in the order of factors 200 when training and evaluating over Data, a result that grants further investigation. While no differences were observed in the calorimeter images for these η values, when checking the η distributions of the LF class, we observe fewer examples in this last bin. This under-representation of the LF class in both the training and testing datasets makes this last bin’s results less reliable than the others.

In summary, the CNN algorithm presents a great improvement with respect to the LLH method when rejecting backgrounds from the LF class. Even though the performance observed with the model trained with simulated examples is reduced when rejecting examples of the background from real data, the values still remain above the LLH reference by factors of 1.5–2 for the Tight working point and factors of 3–5 for the Loose working point. When incorporating LF class examples from experimental data in the training, the improvement

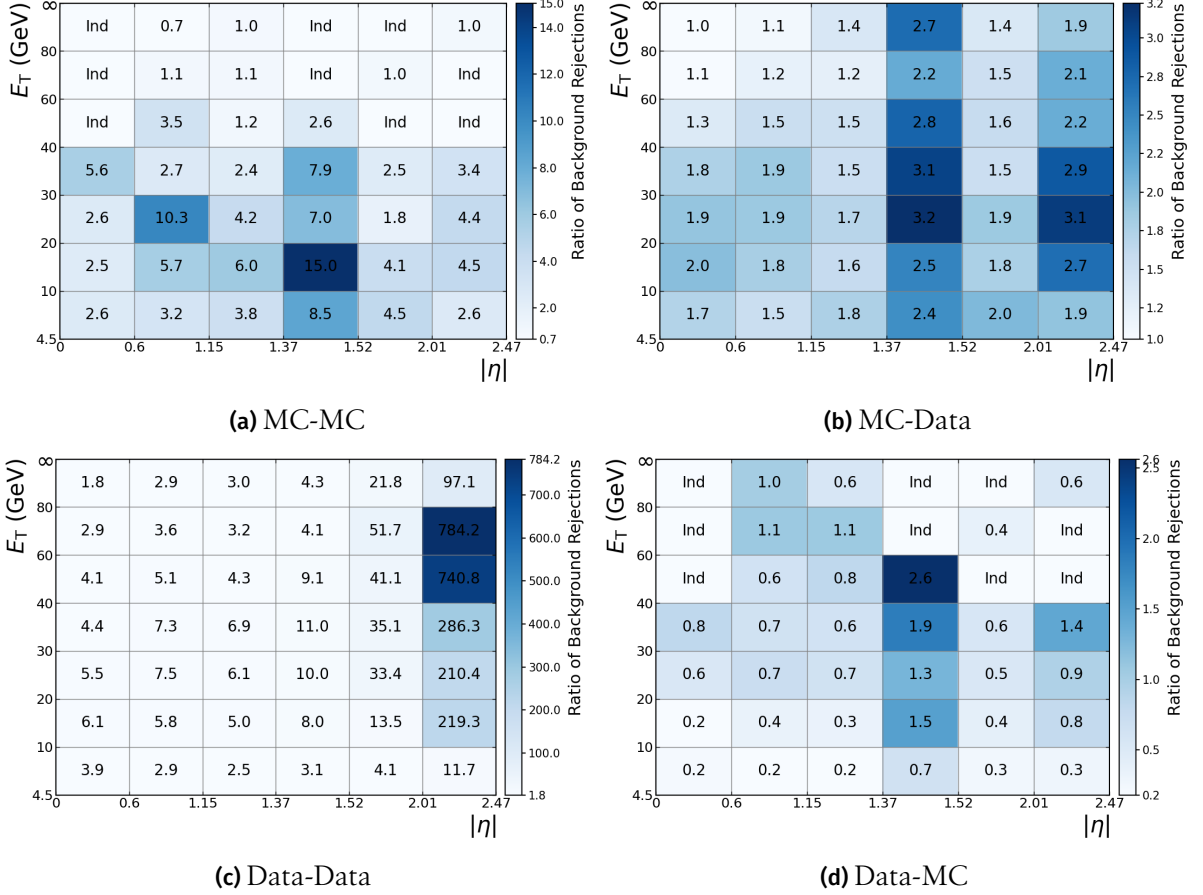


Figure 6.17: Ratios of the combined LF class background rejection from a CNN model with respect to LLH methods as a function of E_T and $|\eta|$ bins. The CNN model was trained with HL ν tracks and images as input features and results are shown for a model trained in MC and evaluated on (a) MC and (b) Data, as well as a model trained in Data and evaluated on (c) Data and (d) MC. Values were obtained for the signal efficiencies corresponding to the LLH TIGHT working points.

ranges from 3–10 across the different working points and E_T bins, values that surpass those obtained when both training and evaluating with simulated examples. As a result, the CNN algorithm is a potential alternative to the LLH method for the identification of electrons in the ATLAS detector, providing an improved rejection of its most prominent background, the LF class.

6.6 Future improvements

The results presented of a CNN algorithm for electron identification are encouraging. By introducing low-level features such as the calorimeter energy deposits and track-matching variables to the task, it improves the background rejection with respect to the current LLH method across different backgrounds. In particular, it presents a consistent improvement when evaluated in real experimental data, more so when incorporating such data in the training

process.

One must note, however, that the performance of neural networks is highly dependent on the availability of a large amount of examples with great quality. The use of MC simulation events guarantees large statistics to do so, but its known mismodelling remains one of the major problems of such applications in the ATLAS detector. As a result, the mitigation of these discrepancies between MC and data is a core element in future developments of any electron identification algorithm.

One way to address this was discussed in this work: the incorporation of real experimental data examples in the training dataset. This approach has shown great results when considering the largest electron background, the LF class, but could just as well be expanded to incorporate signal electrons from data. As discussed in Chapter 5, a sample enriched in prompt electrons can be obtained from the $Z \rightarrow e^+e^-$ decays to measure the electron efficiency in data. Training a neural network with both signal and background coming from real data could guarantee the robustness needed when discriminating electrons in data, especially for trigger applications.

Another approach would be to correct the input features of the CNN in MC simulated examples to compensate the mismodelling. Inspired by the one currently used in the LLH method, vide Section 4.3.1, it would consist of reweighing the energy deposit cells according to the distributions observed in data. This method, however, would involve a more complex and technically challenging procedure, in part due to the different cell sizes in the layers and transition regions, as well as the presence of missing cells.

A more sophisticated architecture could also be incorporated to the algorithm using an adversarial network^[294]. This way, a supporting network could be trained to distinguish if examples were generated from MC or real data, with this discrimination power incorporated into the main network responsible for discriminating prompt electrons from backgrounds. The trade-off between these classification tasks could create a network that is robust against the discrepancies observed between the MC simulated and real data examples.

Furthermore, the architecture of the neural networks used to perform the electron identification could be revisited. Many advancements in the field of machine learning could be exploited, such as the use of graph neural networks (GNN), which could also take advantage of the energy deposit cells^[190].

Nevertheless, it is important to mention that any neural network approach is less adaptable regarding changes in the input features. While the assumption of uncorrelated variables in the LLH method allows for the easy removal of a given variable, this is not true when dealing with neural networks. That is, if any change needs to be made to the input features, a new model needs to be trained. On the other hand, this disadvantage is counterbalanced by the flexibility introduced by the multinomial aspect of the discriminant, which allows for a tuning that improves the rejection of the background associated with a given class.

Moving forward, the results presented in this work indicate that the use of low-level

features such as calorimeter images and track-matching information are potentially beneficial to the electron identification task. In particular, the ability of the network to incorporate the correlations between the input features into the classification task shows great potential in improving its capabilities to discriminate the different types of electron background. As such, the development of any electron identification algorithm would benefit from the use of low-level features and the incorporation of examples from experimental data, as shown in this work.

Chapter 7

Bump search with the data-directed paradigm

We take for granted so much of what surrounds us that we must take a step back to recognize what it is about the physical world that requires explanation.

Robert N. Cahn^[78]

7.1	State of current searches	146
7.2	Resonance searches with anomaly detection	148
7.3	Data-directed paradigm	150
7.4	Bump search with a neural network	152
7.4.1	Task and architecture	152
7.4.2	Synthetic data generation	154
7.5	Histogram processing	157
7.5.1	Dark Machines dataset	159
7.5.2	Processing and calibration	161
7.6	Training	164
7.7	Performance over synthetic data	165
7.7.1	Impact of background shapes	173
7.7.2	Impact of systematic uncertainties	175
7.7.3	Impact of signal width	176
7.8	Performance when looking for new physics	177
7.9	Histogram production from ATLAS simulation data	183
7.10	Outlook	185

In this chapter, a novel approach to search for new physics in the ATLAS data is introduced. The technique uses neural networks to search for resonances in invariant mass distributions obtained from experimental data. A proof of concept is developed and tested using MC simulation data, achieving great performance in identifying both synthetic Gaussian signals and simulated BSM models. Moreover, a preliminary implementation to process invariant mass histograms is discussed in preparation for its application to ATLAS data.

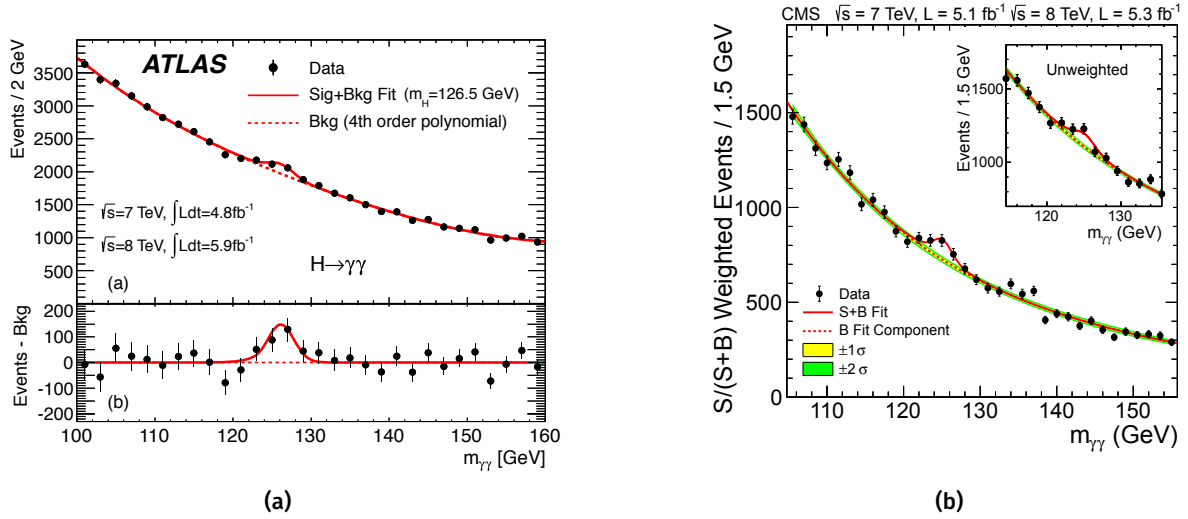


Figure 7.1: Distributions of the invariant mass of the diphoton candidates, presented as of the discovery of the Higgs bosons by the (a) ATLAS experiment [69] and the (b) CMS experiment [70].

7.1 State of current searches

One of the main goals of the ATLAS experiment physics programme is to search for signs of new physics. These can be observed as either quantitative deviations from theoretical predictions or qualitatively different phenomena that are not yet understood. In the first scenario, a deviation in a given observable could be coming from higher-order contributions involving unknown particles or interactions. Thus, measuring a major deviation from the SM predictions would indicate the existence of physics beyond the SM. Another more straightforward way to detect a new particle is by directly reconstructing its decay products, in what is called a *direct search*. If a particle is produced *on shell*⁽³⁵⁾, a resonance is produced in the invariant mass distribution of its decay products. This is how the Higgs boson was discovered, for instance, via the reconstruction of the $H \rightarrow ZZ^* \rightarrow 4\ell$ and $H \rightarrow \gamma\gamma$ final states^[69,70].

Since the resonance appears in the form of a bump in the invariant mass distributions, the search for resonances is sometimes characterised as a *bump search*. This name refers to the shape that is observed around the particle's mass in the invariant mass distribution of its decay products, as we can see in Fig. 7.1 for the case of the $H \rightarrow \gamma\gamma$ decay. In the context of these two-body resonances, where the new particle decays into two other particles, this bump will depend only on the mass, width and production cross-section of the new particle^[295]. If the decay products happen to be SM particles, they should be detectable by the ATLAS detector. However, many of these searches have not yet been performed, as can be seen in Fig. 7.2, which shows the coverage of existing searches of two-body resonances at the LHC as of 2020^[295]. Many final states are still uncovered, such as those involving a lepton and a

(35) We say a particle is produced *on (mass) shell* when it satisfies its equation of motion, in this case $E^2 - p^2c^2 = m_0^2c^4$, meaning a real particle was produced. This is not the case for a *virtual particle*, which can contribute to a process through diagrams, but cannot exist beyond a short time or space interval.

	e	μ	τ	q/g	b	t	γ	Z/W	H	BSM $\rightarrow$ SM ₁ $\times$ SM ₁			BSM $\rightarrow$ SM ₁ $\times$ SM ₂			BSM $\rightarrow$ complex		
										q/g	γ/π^0 s	b	tZ/H	bH	$\dots$	$\tau qq'$	eqq'	$\mu qq'$
e	[37, 38]	[39, 40]	[39]	∅	∅	∅	[41]	[42]	∅	∅	∅	∅	∅	∅	∅	∅	[43, 44]	∅
μ		[37, 38]	[39]	∅	∅	∅	[41]	[42]	∅	∅	∅	∅	∅	∅	∅	∅	∅	[43, 44]
τ			[45, 46]	∅	[47]	∅	∅	∅	∅	∅	∅	∅	∅	∅	∅	[48, 49]	∅	∅
q/g				[29, 30, 50, 51]	[52]	∅	[53, 54]	[55]	∅	∅	∅	∅	∅	∅	∅	∅	∅	∅
b					[29, 52, 56]	[57]	[54]	[58]	[59]	∅	∅	∅	[60]	∅	∅	∅	∅	∅
t						[61]	∅	[62]	[63]	∅	∅	∅	[64]	[60]	∅	∅	∅	∅
γ							[65, 66]	[67–69]	[68, 70]	∅	∅	∅	∅	∅	∅	∅	∅	∅
Z/W								[71]	[71]	∅	∅	∅	∅	∅	∅	∅	∅	∅
H								[72, 73]		[74]	∅	∅	∅	∅	∅	∅	∅	∅
BSM $\rightarrow$ SM ₁ $\times$ SM ₁										∅	∅	∅	∅	∅	∅	∅	∅	∅
q/g										∅	[75]	∅	∅	∅	∅	∅	∅	∅
γ/π^0 s												[76, 77]	∅	∅	∅	∅	∅	∅
b													∅	∅	∅	∅	∅	∅
$\dots$																		
$\dots$																		

Figure 7.2: Status of two-body resonance searches at the LHC as of 2020 as presented in Ref. [295], with each entry corresponding references to existing searches. Highlighted in orange are the final states that have not been probed yet.

quark/gluon or Higgs boson. If we go one step further and consider all the possible topologies and particle combinations, it becomes clear that there is still a lot of potential to find new particles via resonances from the LHC data^[296].

The traditional way of performing a data analysis in particle physics is based on the *blind-analysis paradigm*^[297]. The principle is to estimate the predicted observable from the SM without looking at the data to avoid any bias. In a bump hunt, for instance, this would correspond to building the invariant mass distribution in question from SM simulated events. Once this background shape is properly estimated, thus defining the null hypothesis, the data is *unblinded*, meaning the corresponding invariant mass distribution is obtained from the experimental data. If the background hypothesis is rejected when compared to the observed data, the discovery of a new physics phenomenon can be claimed.

When searching for specific BSM models, one can engineer certain selections on the data to reduce the SM background and thus enhance the sensitivity to the given model. These cuts are usually referred to as the *signal region* and the corresponding experimental data remains unseen until the whole background estimation is properly conducted. By defining *control regions* which are orthogonal to this signal region, many checks can be developed to make sure the background estimate is properly calibrated and the uncertainties accounted for. Only once all checks are performed and the researchers are confident that the background is correctly estimated, the experimental data can be processed and compared to the theory prediction.

Although this method is great to enhance sensitivity to BSM models without biasing the results, it requires an elaborate workflow. When building the background estimate, all the different processes that contribute to the observable must be considered. Not only that, but we must make sure that the simulation correctly accounts for detector effects, which means correctly calibrating the outputs and correcting for mismodelling. Finally, all the different systematic uncertainties introduced in the theory prediction of the observable must be properly estimated, meaning the whole process must be repeated with variations of the different

parameters. This effort requires many resources, from the computational side to generate all the simulated events, but especially from the human effort side, since the time needed from the start of an analysis until its results publication usually spans from 3 to 4 years.

Most ATLAS resonance searches are performed in this blind-analysis paradigm, the wide majority of them being model-dependent and focusing on a specific final state. Since the final state is chosen to optimise the search for a particular BSM model, the analysis is largely dependent on our imagination to think of these new models. More importantly, the effort needed to perform these analyses leaves a large portion of the data unexplored. In this context, model-agnostic approaches offer an alternative to explore the data by prioritising the search for deviations from the SM without any assumption on the signal. Many such searches have been performed^[153,210,211], taking advantage, in particular, of anomaly detection techniques.

7.2 Resonance searches with anomaly detection

A plethora of machine learning techniques has been developed to address the anomaly detection task in high-energy physics, as discussed in Section 2.7.2. These techniques frame the task as the search for something that is not expected from the SM, either as an outlier event or object, or an unexpected shape of a differential cross-section. For resonance searches, in particular, we expect that the new physics will cluster in the form of a bump and many anomaly detection techniques can take advantage of this localised signal.

For instance, one way to perform such an anomaly detection search is to directly compare the data to the background estimate, which can be a fit over the distributions from simulated MC events. This technique was performed by ATLAS when searching for new resonances over the dijet invariant mass spectrum^[102] or the mass of objects produced together with a high-energy leptonically-decaying Z boson^[298]. Another approach is to train an autoencoder to distinguish signal from simulated background events, and perform the usual bump hunt over the data that gets classified as more signal-like, a technique that has been used by ATLAS in two-body invariant mass distributions^[299]. Furthermore, one can incorporate a data-driven approach by exploiting the localised nature of the anomaly. If we probe a given region of the mass spectrum at a time, we can suppose the surrounding areas, referred to as the side bands, are enriched in background. Thus, we can train a weakly supervised classifier, such as done in the CWoLA method^[200,300], a technique that was successfully applied in an ATLAS di-jet resonance search^[301], although no deviations were observed.

One of the main applications of anomaly detection is in the context of general searches, where many final states are checked for deviations. Such a scan of the phase space can identify regions that show potential signs of new physics that could be further investigated with more precise methods. One such analysis was performed by ATLAS^[302] and looked for statistical deviations from the SM expectation. This strategy used pseudo-experiments to estimate the statistical significance of excesses in data when compared to the MC simulation, which means

it also relied heavily on the MC description of the processes and the simulations of the detector response. But the fact that it allows for a global interpretation of finding deviations allows for the detection of small independent excesses that combined could be worth investigating.

However, one of the main limitations of general searches is their reduced sensitivity. When looking for a particular new physics model, it is common to define selection cuts that reduce the expected SM background in the signal region. For instance, in the ATLAS search for leptoquarks in the $b\bar{b}l\bar{l}$ channel^[141], a subset of the cuts required $m_{\ell\ell} > 130$ GeV and $p_T^{\ell\ell} > 75$ GeV to reduce the background from Drell–Yan and on-shell Z boson productions, while requiring $E_T^{\text{miss}}/\sqrt{H_T} < 3.5$ GeV^{1/2} suppressed the $t\bar{t}$ background. In a general search, however, we want to avoid performing such cuts since no assumption is made in the signal model and these cuts might remove the signal events altogether. But without cuts, the dataset can be overwhelmed by background events and the signal might not be detectable at all. That said, this should not discourage general searches, but it should be kept in mind when developing and interpreting their results. After all, we do not have the luxury of dismissing general searches, since they can play an important role in identifying deviations between data and SM, especially in light of the many unexplored regions of the phase space.

Another important thing to keep in mind when performing model-agnostic searches is the problem of multiple comparisons, or *look-elsewhere effect* (LEE)^[303]. When calculating the statistical significance associated with a given excess, one must consider the significance of finding an excess anywhere. That is, the excess might as well be a fluctuation of the SM background that could happen in any given range of the observable values. This consideration naturally reduces the significance of any local deviation by a factor of roughly the order of independent signal regions that are considered^[219]. To properly estimate this effect, one would need to repeat many simulations of background-only events and quantify how many times the given excess appears. This, however, is incredibly time consuming, since to achieve a 5σ significance, for instance, one would need to perform $\mathcal{O}(10^7)$ tests^[303]. As a result, any attempt to cover many different signatures will lose sensitivity to individual deviations. Consequently, any excess found in such a way needs a dedicated follow-up analysis, performed on a statistically uncorrelated dataset, to confirm the excess over background.

At present, most anomaly detection analyses suffer from the main limitation as traditional analyses: they rely on MC simulations of background SM events. Thus, the known existing mismodelling of these simulations, discussed in Section 3.2.6, entails considerable MC/data discrepancies and has a direct impact on the reliability of the background estimation. In particular, for machine learning models that rely on the discrimination of signal from background events, if the training data comes entirely from simulation, the output of the network will be embedded with this mismodelling. Furthermore, the known MC/data discrepancies affect the estimation of uncertainties by introducing nuisance parameters that are hard to quantify. This is especially problematic for machine learning applications, since the distributions from which these models learn would not match the true distribution it is

expected to describe^[304]. Therefore, a heavy dependence on MC simulations is a limiting factor in many analyses which attempt to perform model-agnostic anomaly detection.

One possibility to mitigate the limitations of MC mismodelling is to incorporate data into the learning of a classification task. This can be done either by training directly in a background-enriched region of the data, such as the case of the CWoLa^[200,300] and CURTAINS method^[305], or using autoencoders trained to reconstruct data and learning to ignore noise^[299]. Going one step further, would be to reduce as much as possible the dependence on MC simulation, with a shift in paradigm that focus on the experimental data instead. An example of such strategy is described in the following section.

7.3 Data-directed paradigm

When defining search strategies, one can move away from the blind-analysis paradigm and consider looking for anomalies directly in the data. To do that, we need a reliable method to efficiently search the vast parameter space of mass values and decay channels of new particles, such that we can quantify how anomalous the data is with respect to a given reference. This is what the *data-directed paradigm*^[306] proposes: to search the experimental data for deviations of a well-established property of the SM. This can be achieved by constructing a reliable tool that allows for a fast scan of the available phase space and identifies regions of interests to be explored with more sensitive methods further on.

The first ingredient in this approach is to identify a well-known property of the SM such that any deviations from this property would correspond to new physics. Once the target deviation metric is established, we need a tool that can efficiently search for these deviations. Neural networks are great candidates for this task since they can map the observable in question to a deviation score. Also, their ability to encode nuanced features and correlations when mapping input and output has great potential when dealing with more intricate scenarios.

Once such a tool is constructed, it can be directly used in the experimental data to flag deviations. By not depending on MC simulations to estimate SM backgrounds, this method is unaffected by the MC/data discrepancies that greatly limit recent searches. Furthermore, the time and computing resources are reduced, since there is no need for large statistics datasets or systematic uncertainties computations. With fewer resources needed, we obtain not only faster outputs but a survey of a larger number of final states and object combinations than traditional analyses.

It must be noted, of course, that the modelling of the background hypothesis will not be as precise without a full background estimate. Furthermore, if different final states probed are correlated, the calculation of a global significance that accounts for the LEE will be a challenge. That said, this approach still offers a great alternative to tap into the discovery potential of an experiment. It will not replace traditional analyses, but it can serve as a supporting tool when exploring the data.

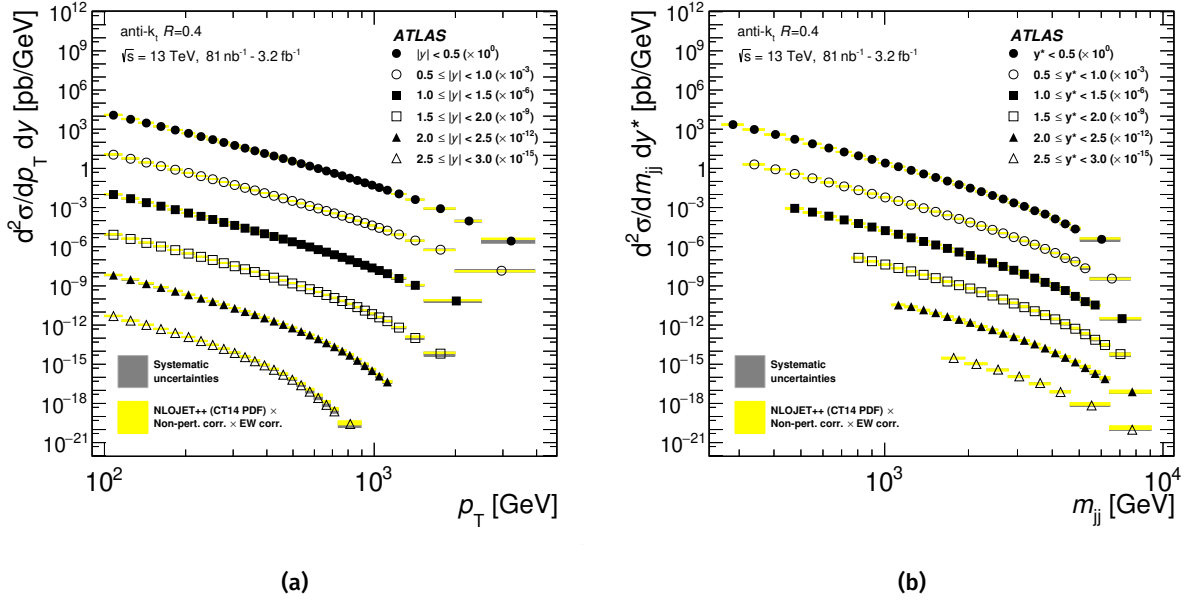


Figure 7.3: Smoothly falling shapes for the (a) inclusive jet cross-section as a function of the transverse momentum p_T and the (b) dijet cross-section as a function of m_{jj} . Taken from Ref. [75].

In particular, the goal is to construct a reliable tool that can give us enough confidence that an anomaly is present. By identifying which final states and selections are most promising in the search for new physics based only on what the data shows us, it serves as a guide to where a traditional analysis should be made. This way, we accelerate and optimise the search for new physics and narrow down the existing efforts to searches that have more chances of actually discovering anomalous phenomena.

Proof of concepts using the data-directed paradigm already show promising results. One approach in particular considered the symmetries of the SM as the property in question^[307]. The idea was to compare two samples that are mutually exclusive in the data. Since a SM symmetry would assume both datasets originate from the same distribution, any deviation from this behaviour would indicate the presence of new physics. This technique can be useful when investigating the apparent lepton-flavour symmetry in the SM, where there is no preference in processes involving electrons and muons, for instance, but some observables seem to indicate slight experimental deviations, as described in Section 1.7.

Another approach which has been shown to be very promising is to perform a bump hunt. In this case, we take advantage of the fact that in a region of the mass spectra where no SM particle can be found, we expect a smoothly falling shape for the invariant mass distribution. This behaviour is directly related to the same smoothly falling behaviour observed in cross-section shapes^[75], as can be seen in Fig. 7.3. For all processes happening at the collision point, the cross-sections get smaller as the energy increases, a property that is linked to different phenomena that happen at high-energy collisions. Firstly, as the centre-of-mass energy gets higher, many different kinematic regions become accessible. Thus, processes that require less energy will contribute less as the energies increase, resulting in smaller cross-sections.

Furthermore, as shown in Fig. 1.2, the proton PDFs are dominated by gluons at higher energies. This means that processes relying on quarks with high fractions of the proton momentum are suppressed and will also present a smaller cross-section.

In this context, a new search strategy was presented in Ref. [306], where a neural network is trained to estimate the statistical significance of resonances over invariant mass distributions, supposing smoothly falling background shapes. In what follows, this novel approach is described and expanded to incorporate more realistic applications, together with preparations for its application in ATLAS experimental data.

7.4 Bump search with a neural network

7.4.1 Task and architecture

As previously discussed, if a particle is created on-shell during a collision, a resonance will appear around its mass in the invariant mass distribution of its decay products. If a bump that appears in the histogram cannot be explained by the SM, it may constitute a new particle. Thus, the focus of a bump search is to identify statistically significant bumps in the invariant mass distribution of a given object combination. In particular, our goal is to explore many final states and assess the mass of different object combinations within these final states. By considering many unexplored physics object combinations, we can make sure to cover as much of the possibilities available in the experimental data.

Hence, we expect a BSM particle to create a bump in a mass region where the SM events would produce a smoothly falling shape. This can be used to probe the existence of the new particle if the statistical significance of this bump can reject our null hypothesis, i.e., the background. In particular, the statistical significance of having a bump centred at a given bin can be used as a metric for how much deviation we have from the smoothly falling shape.

So, if we consider an idealised analysis, where we suppose no uncertainty in either the background or the signal invariant mass shapes, the background can be described by a smoothly falling shape, for instance coming from an analytical function, while the signal can be described by a narrow Gaussian centred at the mass value of the particle. When compared to real data, a hypothesis test can be performed comparing the background-only hypothesis to a signal hypothesis where the signal is centred at each bin. If we map each bin to the statistical significance that a signal bump is centred at that bin, high values of this statistical significance indicate considerable deviations from the smoothly falling shape.

Neural networks are the perfect candidates for performing this task. By construction, they can learn to map an input to a given output through exposure to many different examples. The fact that they can abstract away correlations between the features involved make them especially appealing, as the concept of “smoothly falling” can be encoded by the network from a well-constructed training dataset. In this dataset, examples have as input a vector of the

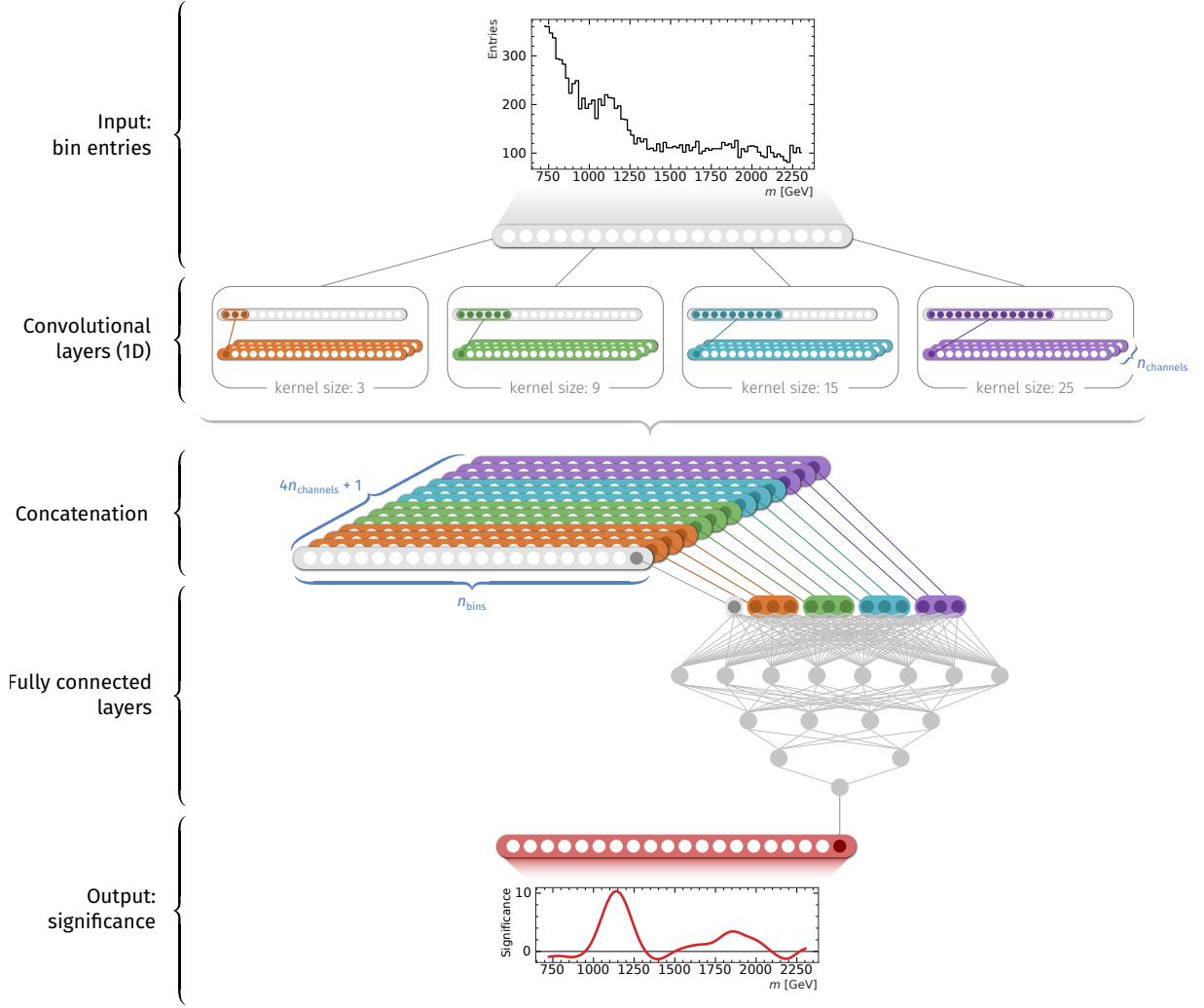


Figure 7.4: Architecture of the BumpNet neural network. A vector of the bin entries of a mass histogram is used as input and processed by 4 convolution layers in parallel. Their output is concatenated and passed through a fully connected layer that process all the representations for each bin separately. The network's output has the same size as the input and corresponds to the statistical significance value predicted for each bin.

bin entries of an invariant-mass histogram, while the output is a vector of the corresponding statistical significance of having a Gaussian signal in each bin, calculated using the hypothesis test described in Section 2.4.

To perform this task, a neural network called *BumpNet* is constructed using the architecture shown in Fig. 7.4. Since the input is a vector of size $(n_{\text{bins}}, 1)$, it is processed by one-dimensional CNNs that are run in parallel, where each convolution has a kernel with different sizes, namely 3, 9, 15 and 25. Each convolution sequence has five combinations of a convolution layer followed by a ReLU layer. A padding of zeros is performed to guarantee that the output of each convolution layer has the same size as the input. The output of each one of these convolutions is a tensor of size $(n_{\text{bins}}, n_{\text{channels}})$, with $n_{\text{channels}} = 64$ ⁽³⁶⁾. These are

(36) There is no particular reason for this number other than it is commonly used in machine learning applications. The higher the number, the more complex the network becomes.

Table 7.1: Analytical functions used to define smoothly falling background shapes, where y corresponds to the number of entries per bin and x are the invariant mass values. The parameters a and b are defined by setting two limits $(x_{\min}, y_{\min})$ and $(x_{\max}, y_{\max})$ which completely define the shape.

Linear	$y = ax + b$	Exponential	$y = ae^{-bx}$
One over x	$y = \frac{1}{ax} + b$	Half parabola	$y = a(x - x_{\max})^2 + y_{\min}$
One over x squared	$y = \frac{1}{ax^2} + b$	Negative logarithm	$y = -a \ln(x) + b$
One over x cubed	$y = \frac{1}{ax^3} + b$	Cosine quarter	$y = \Delta y \cos[a(x - b)] + y_{\max}$
One over x to the 4th	$y = \frac{1}{ax^4} + b$	Half-hyperbolic cosine	$y = \cosh[a(x - x_{\max})] + b$
One over x to the n th	$y = \frac{1}{ax^n} + b, n \in [0.01, 10]$		

concatenated together with the input vector into a tensor of size $(n_{\text{bins}}, 4n_{\text{channels}} + 1)$. It is, in turn, passed to an FCN, which treats each one of the bins independently, but mixes all their representations. It is formed by four internal layers of linear transformations followed by ReLU transformations, having 128, 64, 32 and 1 nodes respectively. The output of this FCN will be a vector of size $(n_{\text{bins}}, 1)$, corresponding to the statistical significance of each bin.

The main advantage of this particular architecture is that the network is agnostic to the number of bins in the histogram. This puts no constraints on the type of histogram that can be checked or if the binning must be changed to accommodate the network. Furthermore, the use of convolution layers performing scans of neighbouring bins is intuitive since we know the bump will appear as an excess in successive bins of the histogram. Of course, the network's performance depends heavily on having great quality examples provided during training, the creation of which is discussed in the next section.

7.4.2 Synthetic data generation

Synthetic datasets are created to teach the neural network how to properly estimate the statistical significance of a bump in the invariant-mass distribution. An example histogram is constructed via the combination of a background and a signal shape with an added noise to form the equivalent of the observed data. Since both the background and signal shapes are known, the statistical significance in each observed bin can be calculated using the hypothesis test discussed in Section 2.4.

Starting with the background shapes, these must be smoothly falling to reflect the property of the SM we want the network to learn. The first type of background considered are analytical functions, shown in Table 7.1. These 11 functions have 2 parameters (a, b) that are defined by the limits $(x_{\min}, x_{\max})$ and $(y_{\min}, y_{\max})$, taken randomly from the intervals $x \in [10, 5 \cdot 10^3]$ and $y \in [1, 10^5]$. These values represent the ranges we expect to see in real data for the mass

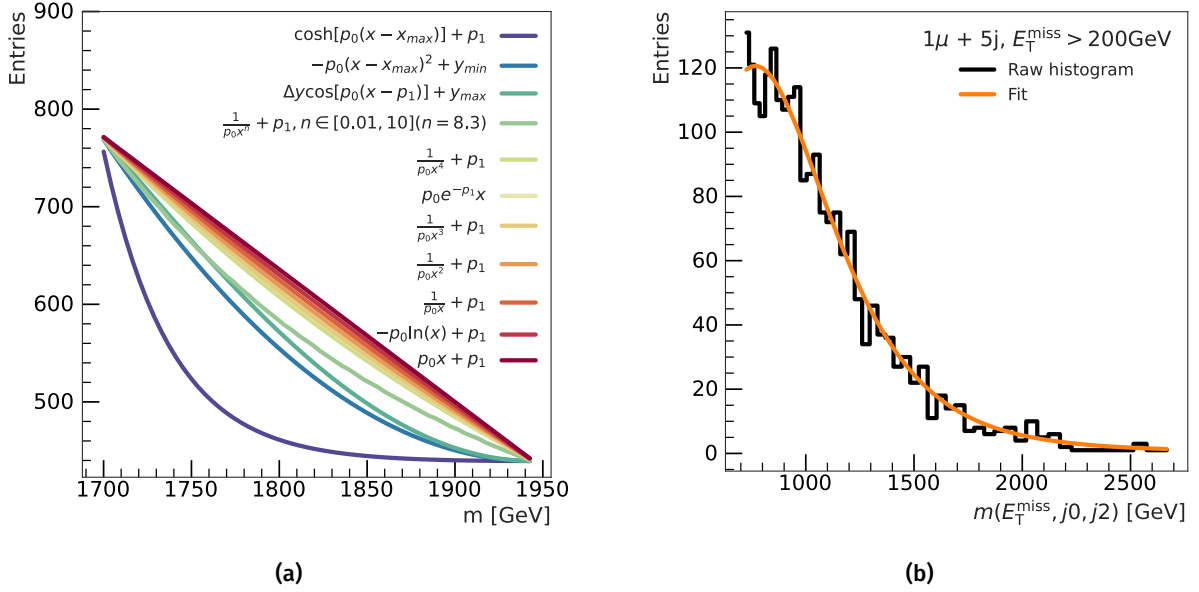


Figure 7.5: Examples of (a) analytical functions and (b) fit function used as smoothly falling background shapes. The parameters of the analytical functions are obtained by fixing the limits $(x_{\min}, y_{\min})$ and $(x_{\max}, y_{\max})$. On the other hand, fits of the 4th-order log-polynomial function $\ln(y) = \sum_{n=1}^4 p_n x^n$ are obtained from the mass histograms constructed from simulation data, in this case the Dark Machines dataset.

values and number of entries per bin, respectively. A minimum separation of 100 is required between the $x_{\min}$ and $x_{\max}$ to avoid mass ranges that are too small. A visualisation of how the different functions change for a fixed limit is shown in Fig. 7.5a.

As it will be shown in Section 7.7.1, these shapes cannot fully capture the histograms we encounter in real data. In particular, kinematic cuts can sculpt the invariant mass shapes in such a way that flattens the lower-mass range of the histograms. This behaviour is observed in MC simulation histograms, which can in turn be incorporated to enhance the shapes we consider for background. That is, by fitting a function to these simulated histograms, more unconventional background shapes are obtained, as shown in Fig. 7.5b, where we have an example including both the raw simulated SM background histogram and the corresponding fit. After some optimisation, a 4th-order log-polynomial function, defined as $\ln(y) = \sum_{n=1}^4 p_n x^n$, was chosen to perform the fit and thus obtain a smooth background.

Next, a signal is assumed to be narrow with respect to the detector resolution, such that the excess happens over a few bins. We assume a Gaussian shape with variance corresponding to approximately one bin. After scaling this signal to match a given strength, both background and signal histograms are added to create the equivalent of the observed histogram. Statistical noise from a Poisson distribution is added to this histogram to make it more realistic. Some synthetic data examples are shown in Fig. 7.6.

Since in these synthetic data examples we know the true background, the statistical significance of having a Gaussian signal in a given bin can be calculated using the PLR test^[152], as discussed in Section 2.4. Therefore, for each bin, we perform the hypothesis test to obtain

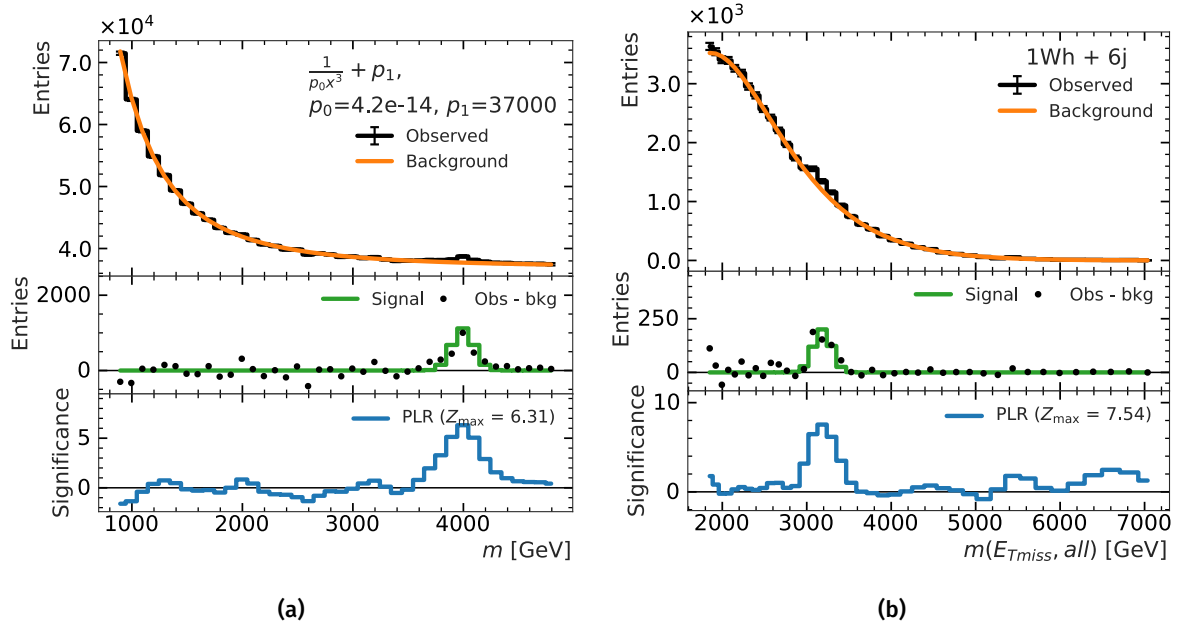


Figure 7.6: Examples of invariant mass distributions generated from synthetic data. They are obtained from the combination of a smoothly falling background (orange) with a Gaussian signal (green), the latter being scaled to match a given strength. A Poisson noise is added to the combination to obtain the corresponding observed histogram (black). The PLR test is used to obtain the target Z values (blue), which are the statistical significance of having a Gaussian signal centred in each bin with the background shape as the null hypothesis. The maximum Z value of the histogram, $Z_{\max}$ is quoted, matching the position and strength of the injected signal.

the statistical significance to reject the null hypothesis (background-only) when we measure the observed histogram. For each bin i , the null hypothesis predicts $N_i^{\text{exp}} = b_i$, while the alternative hypothesis predicts $N_i^{\text{exp}} = b_i + \hat{\mu}s_i$, where s_i is a Gaussian signal centred at bin i , that is,

$$s_i = \int_{\text{bin } i} f(x; x_o, \sigma) dx \approx f(x; x_o, \sigma_s) \Delta x,$$

where x_o is the centre of bin i and σ_s is the width of the signal, both in units of GeV, the latter given by the product of the width of the bin i and how many bins the signal spans. We choose the signal width to correspond to only one bin. The signal is scaled by the MLE $\hat{\mu}$, as previously discussed in Section 2.4, and given by

$$\hat{\mu} = \arg \min \{ -\ln[L(\mu)] \}, \quad \text{where} \quad -\ln[L(\mu)] = -\sum_{i=1}^{n_{\text{bins}}} N_i^{\text{obs}} \ln(N_i^{\text{exp}}) - N_i^{\text{exp}} + \ln\left(\frac{1}{N_i^{\text{obs}}!}\right),$$

with N_i^{obs} and $N_i^{\text{exp}} = b_i + \mu s_i$ corresponding to the observed and expected values for the number of entries in that bin, respectively.

Using the likelihood for each one of the hypotheses, namely,

$$\begin{aligned} -\ln[L(\mu = 0)] &\rightarrow \text{Background-only hypothesis } (N_i^{\text{exp}} = b_i), \\ -\ln[L(\mu = \hat{\mu})] &\rightarrow \text{Signal MLE hypothesis } (N_i^{\text{exp}} = b_i + \hat{\mu}s_i), \end{aligned}$$

the significance Z_i for the corresponding bin is given by

$$\begin{aligned} \text{If } \hat{\mu} \geq 0 : \quad Z_i &= \sqrt{q_0} \quad \text{with } q_0 = -2 \ln \left[\frac{L(0)}{L(\hat{\mu})} \right], \\ \text{If } \hat{\mu} < 0 : \quad Z_i &= -\sqrt{-q_0} \quad \text{with } q_0 = \begin{cases} -2 \ln \left[\frac{L(\hat{\mu})}{L(0)} \right], & \text{if two-sided,} \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

In the case of a two-sided significance, a Gaussian-shaped depletion is also considered a signal which corresponds to a negative significance. Alternatively, one could consider only positive significances by setting any depletion to zero significance instead.

A visualisation of this statistical significance calculation is shown in Fig. 7.7. As can be seen, the PLR target Z is constructed by sliding the Gaussian signal shape that maximises the likelihood at each bin and obtaining the significance of rejecting the background shape. The maximum value of this significance shape occurs when the signal in the hypothesis test overlaps exactly with the signal shape that was injected in the synthetic histogram.

7.5 Histogram processing

As previously discussed, the ultimate goal of this analysis is to assess the statistical significance in invariant mass histograms for many final states and for each physics object combination within these final states. An essential part of that is to obtain these histograms by processing the events, assessing the final states of interest, calculating the mass of different object combinations and filling the histograms.

When looking for narrow resonances, we assume the signal spreads over just a few bins. However, for that to happen, the binning of the histogram must be properly calibrated such that the width of the histogram's bin correspond to the experimental resolution of the ATLAS detector. This will be mainly limited by the resolution of the detector, which is a measure of the variance on the value obtained from the detector's inputs. Thus, if the detector precision is able to capture such a signal, the histogram binning must be calibrated to better reflect this precision, namely matching one bin to approximately 1σ of the experimental resolution on the observable. This is particularly important since the network must be trained with data that reflects its application use-cases. Also, if the detector is precise enough to discriminate the signal, but we calibrate the histograms wrong, we are not using it to its full capabilities.

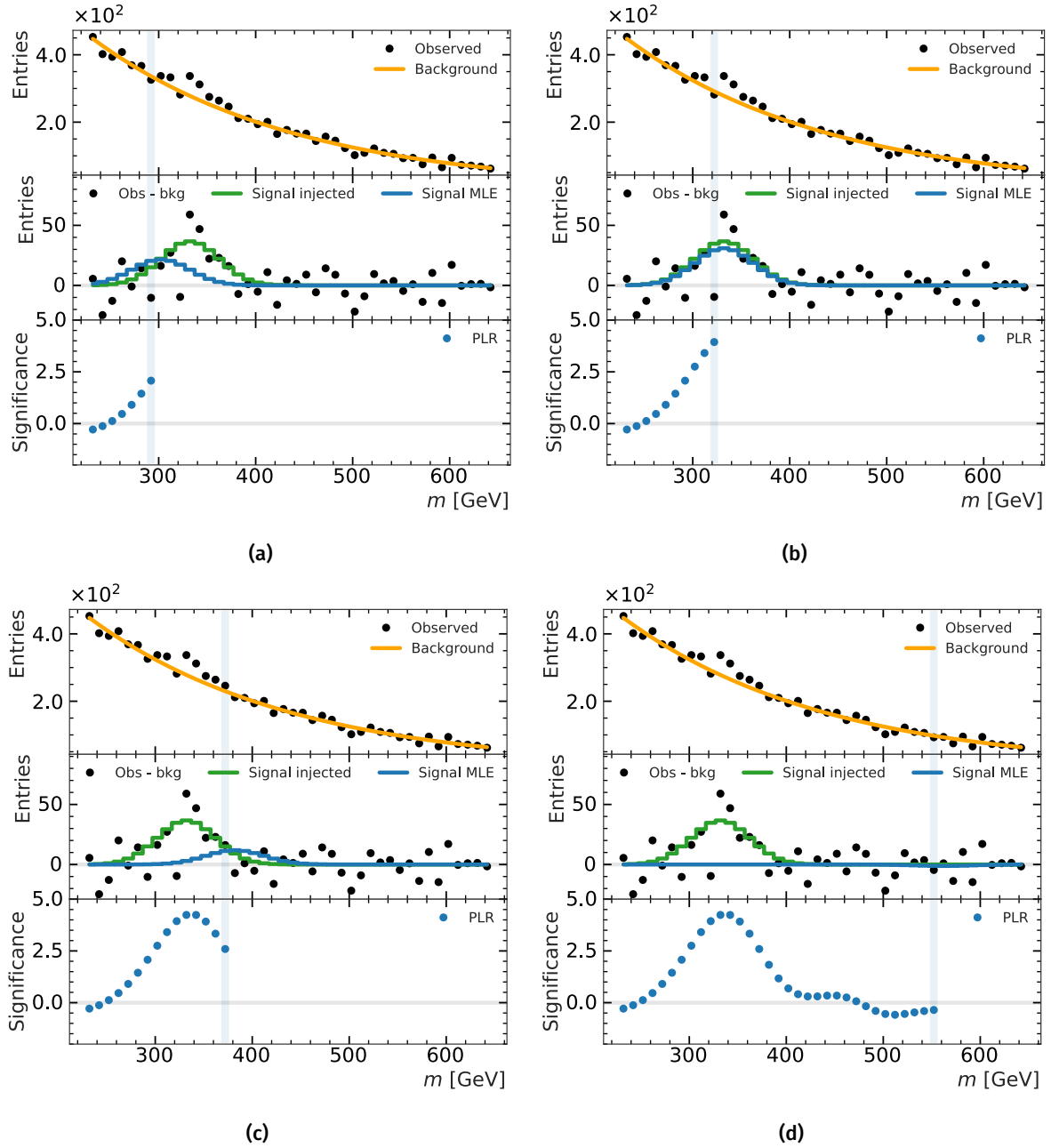


Figure 7.7: Visualisation of the significance calculation for different bins. Each significance value (blue dots) is the outcome of the PLR test, where the background (orange) corresponds to the null hypothesis. The alternative hypothesis for a given bin is the combination of this background and a test signal (blue line) corresponding to the Gaussian signal centre at the bin that maximises the likelihood (MLE). A maximum of the significance is found when the test signal overlaps exactly with the injected signal (green) that was used to generate the observed data (black).

In a way, adjusting the binning of the histogram is equivalent to adjusting the focus on a microscope: the signal might as well be there, but we will not be able to discriminate it from the background as everything will be blurred.

When constructing the mass histograms, MC simulation is a useful proxy for what realistic data looks like. Hence, it is used to implement the workflow and tune the binning calibration according to detector resolution. Moreover, it allows us to get an idea of how many final states and object combinations we expect to probe. Finally, simulation data is used to test the network in more realistic scenarios and to enrich its training dataset as to match the distributions it will encounter when applied to real data. One source of MC simulation data that was used in this project, the Dark Machines dataset, is described in what follows.

7.5.1 Dark Machines dataset

The Dark Machines^[219] dataset is a publicly available MC simulation data created in the context of developing and benchmarking anomaly detection techniques applicable to LHC data. In the context of the project, a billion simulated events were generated, coming from the highest cross-section SM processes at the LHC, listed in Table 7.2, as well as signal events corresponding to different BSM models. The goal of this challenge was to test if model-agnostic techniques could find new physics signals hidden within these events.

The simulated events in this dataset were generated using MadGraph MG5_AMC@NLO^[308] and Pythia version 8.239^[309], passed through a fast detector simulation using Delphes version 3.4.2^[310] with a version of the ATLAS detector card⁽³⁷⁾. The total dataset is equivalent to approximately 10 fb^{-1} of LHC data. It corresponds to pp collisions at 13 TeV whose final state objects include jets, b -jets, electrons, muons and photons. Table 7.3 summarises the cuts performed in the sample. For processes with higher cross-sections, such as $\gamma/W^\pm/Z$ + jets production and QCD jet production, an additional event-level selection is performed on the scalar sum of the transverse momenta of all jets H_T , that is

$$H_T = \sum_i |p_{T,j_i}|, \quad \text{where } p_{T,j_i} > 20 \text{ GeV} \quad \text{and} \quad |\eta| < 2.8,$$

with j_i corresponding to each jet in the event.

Signal events were also generated, considering different BSM scenarios. New physics particles include a 50 GeV Z' in lepton-violating $U(1)_{L_\mu-L_\tau}$ decaying to leptons and neutrinos, as well as RPV SUSY particles, such as a 1 TeV stop pair production, decaying to leptons and b -quarks, both these models being previously discussed in Section 1.8.

Different channels were defined, corresponding to non-exclusive signal regions, as shown in Table 7.3. Channel 1 was defined to have hadronic activity with a lot of missing energy, while channel 2a and 2b require leptons and, as a result, considerably reduce the background. Finally,

(37) A link to the Delphes card used can be found in Ref. [219].

Table 7.2: Processes and corresponding leading-order cross-section included in the simulation datasets from the Dark Machines challenge^[219]. These correspond to the highest cross-section SM background processes produced from pp collisions with 13 TeV centre-of-mass energy at the LHC.

Process	σ (pb)		Process	σ (pb)
$pp \rightarrow jj(+2j)$	19718	$(H_T > 600 \text{ GeV})$	$pp \rightarrow Z\gamma(+2j)$	29.9
$pp \rightarrow \ell^\pm \nu_\ell(+2j)$	10537	$(H_T > 100 \text{ GeV})$	$pp \rightarrow ZZ(+2j)$	9.91
$pp \rightarrow \gamma j(+2j)$	7927	$(H_T > 100 \text{ GeV})$	$pp \rightarrow h(+2j)$	1.94
$pp \rightarrow \ell^+ \ell^- (+2j)$	3753	$(H_T > 100 \text{ GeV})$	$pp \rightarrow t\bar{t}\gamma(+2j)$	1.55
$pp \rightarrow t\bar{t}(+2j)$	541		$pp \rightarrow t\bar{t}Z$	0.59
$pp \rightarrow t + \text{jets}(+2j)$	130		$pp \rightarrow t\bar{t}h(+1j)$	0.46
$pp \rightarrow \bar{t} + \text{jets}(+2j)$	112		$pp \rightarrow \gamma t(+2j)$	0.39
$pp \rightarrow W^+W^- (+2j)$	82.1		$pp \rightarrow t\bar{t}W^\pm$	0.35
$pp \rightarrow W^\pm t(+2j)$	57.8		$pp \rightarrow \gamma \bar{t}(+2j)$	0.27
$pp \rightarrow W^\pm \bar{t}(+2j)$	57.8		$pp \rightarrow Zt(+2j)$	0.26
$pp \rightarrow \gamma\gamma(+2j)$	47.1		$pp \rightarrow Z\bar{t}(+2j)$	0.15
$pp \rightarrow W^\pm \gamma(+2j)$	45.1		$pp \rightarrow t\bar{t}t\bar{t}$	0.0097
$pp \rightarrow ZW^\pm(+2j)$	31.6		$pp \rightarrow t\bar{t}W^+W^-$	0.0085

Table 7.3: Selections performed within the simulated Dark Machines datasets. Different channels define non-orthogonal signal regions, such that channel 1 focus on hadronic activity with a lot of missing energy, channel 2 reduces the background by requiring leptons and channel 3 is the most inclusive one. For more details on the selections, we refer to Ref. [219].

Object level selections			
Jet/b-jet	Electron	Muon	Photon
$p_T > 60 \text{ GeV}$ $ \eta < 2.8$	$p_T > 25 \text{ GeV}$ $ \eta < 2.7, \notin [1.37, 1.52]$	$p_T > 25 \text{ GeV}$ $ \eta < 2.7$	$p_T > 25 \text{ GeV}$ $ \eta < 2.37$
Event level selections			
Channel 1	Channel 2a	Channel 2b	Channel 3
214×10^3 events	20×10^3 events	340×10^3 events	8.5×10^6 events
$H_T \geq 600 \text{ GeV}$	-	$H_T \geq 50 \text{ GeV}$	$H_T \geq 600 \text{ GeV}$
$E_T^{\text{miss}} \geq 200 \text{ GeV}$	$E_T^{\text{miss}} \geq 50 \text{ GeV}$	$E_T^{\text{miss}} \geq 50 \text{ GeV}$	$E_T^{\text{miss}} \geq 100 \text{ GeV}$
$E_T^{\text{miss}}/H_T \geq 0.2$	-	-	-
Minimum of 4 (b-)jets with $p_T > 50 \text{ GeV}$	Minimum of 3 leptons with $p_T > 15 \text{ GeV}$	Minimum of 2 leptons with $p_T > 15 \text{ GeV}$	-
Minimum of 1 (b-)jet with $p_T > 200 \text{ GeV}$			

Table 7.4: Kinematic cuts used in the Dark Machines histogram processing to define its sub-categories. They correspond to limits on the missing transverse energy (E_T^{miss}) or transverse momentum p_T of leading objects, namely electrons (e), muons (μ), as well as the reconstructed leptonically-decaying Z boson (Z), hadronically-decaying Z and W bosons (V_h), boosted top quark (T) and high-mass jets (HM).

Leading object type	Selections (GeV)
E_T^{miss}	$E_T^{\text{miss}} \geq 200, 500$
e	$p_T \geq 100, 200$
μ	$p_T \geq 100, 200, 400$
Z	$p_T \geq 100, 200$
V_h	$p_T \geq 200, 500, 1000$
T	$p_T \geq 500, 1000$
HM	$p_T \geq 500, 1000$

channel 3 was defined to be more inclusive, catching most of the signals except soft electroweak ones.

7.5.2 Processing and calibration

To construct the mass histograms, the processing starts by looping over the events and storing information on the objects and their kinematics. The objects considered were electrons, muons, photons, jets and b -tagged jets. Moreover, some particles were reconstructed from their decay products, namely leptonically-decaying Z boson, boosted hadronically-decaying W and Z bosons (V_h), boosted top quarks (T), as well as high-mass jets (HM). The interest in doing so is to reduce the background in processes that might involve these particles. For instance, if a new physics particle decays to a Z boson, the presence of an identified Z boson will reduce the $t\bar{t}$ SM background events in a final state such as $qql\nu\ell\nu$ where it is expected to be dominant. In particular, if two leptons with the same flavour and opposite sign have a mass within $m_Z = 91.18 \pm 15$ GeV, they are reconstructed as a Z boson. Also, jets with mass in the $[60, 110]$ GeV range are considered as a boosted hadronic vectorial boson, either a W (≈ 80 GeV) or a Z (≈ 90 GeV). Similarly, jets with mass in the $[110, 200]$ GeV range are tagged as boosted top quarks. Furthermore, jets with $m \geq 200$ GeV are considered as independent objects, as they could originate from boosted high-mass new particles. Finally, the value of the missing transverse energy (E_T^{miss}) is also considered.

Once the kinematic information from the objects is stored, the events are grouped in categories according to the number of objects in the final state, except for light jets ($m < 60$ GeV). Final states with two leptons were considered separately if they had same-sign or opposite-sign electric charge. Sub-categories were then further defined from cuts in the leading

objects, as summarised in Table 7.4. Finally, events are further split according to their jet multiplicity, meaning the number of exclusive light jets in the event. Thus, jet bins of exactly $1, 2, \dots, (n - 1)$ jets, or at least n jets were constructed, where n was defined such that the inclusive bin had at least 1500 events.

All the events sharing the same signature (category, sub-category and jet multiplicity) are grouped together in a $\tau\tau$ tree using the ROOT framework^[284]. For each one of those, all possible combinations of the objects are considered and, for each combination, histograms are produced for two observables: (i) the invariant mass, defined as the norm of the sum of the four-momenta of the objects, and (ii) the invariant mass with E_T^{miss} , defined as the norm of the sum of the four-momenta of the objects and the four-momentum of the missing transverse energy, the latter having p_z and m set to 0. Additionally, the missing energy of the event is added to the mass and considered separately.

Then, histograms are filled for the mass observables using a 1 GeV binning. To match the detector resolution, these histograms are rebinned such that one bin corresponds to approximately 0.5σ of the detector resolution. This choice was made to increase the number of bins in the histogram, since the network presented difficulties when dealing with histograms that were too small.

To do so, firstly the maximum of the histogram is identified with the corresponding mass value defining the lower edge of the first bin. The following bin edges are defined from $m_i = m_{i-1}(1 + \sigma_{\text{res}})/2$ where m_{i-1} is the preceding bin edge and σ_{res} is the estimated detector resolution. This resolution depends on the objects considered and is defined as

$$\sigma_{\text{res}} = \sqrt{\sum_n N_n \sigma_{\text{res}}(n)^2},$$

where n corresponds to the object types in the combination, with their corresponding resolution $\sigma_{\text{res}}(n)$ and number of objects involved, N_n . The reconstructed objects T , V_h and HM are considered as 3, 2 and 2 jets respectively, while the Z is considered as 2 electrons.

We note that the experimental resolution varies with the p_T and η of the particles. The jet energy resolution, for instance, can be parameterised as^[311]

$$\frac{\sigma_{p_T}}{p_T} = \sqrt{\left(\frac{N}{p_T}\right)^2 + \left(\frac{S}{\sqrt{p_T}}\right)^2 + C^2} \quad \text{with} \quad \begin{cases} N = 3.33, \\ S = 0.71, \\ C = 0.03, \end{cases} \quad (7.1)$$

where N parametrises the effect of noise coming from the electronics and the pile-up, S parameterises the stochastic effect arising from the sampling nature of the calorimeters and C is a constant term. These terms are obtained experimentally and define the calibration scheme to be used when dealing with jets.

Moreover, in the Delphes card used in the Dark Machines dataset^[219], the electron, muon

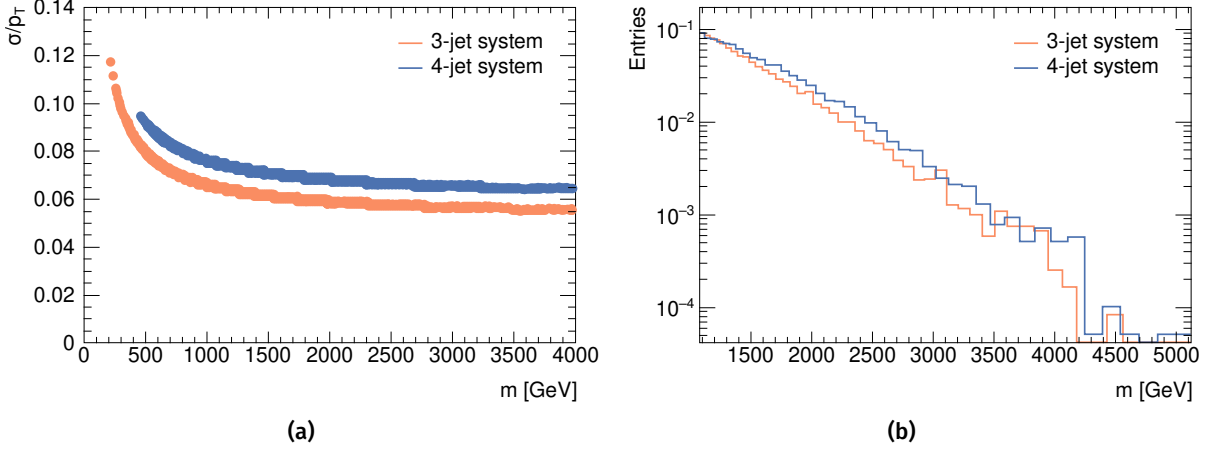


Figure 7.8: Comparison of the (a) relative resolution and the (b) number of entries per bin as a function of the mass for a 3-jet and 4-jet system. The higher resolution observed at low mass is reflected in the histogram binning with smaller bin widths for this region.

and photon resolutions are described as functions of p_T , supposing $|\eta| \leq 0.5$, by the formulas

$$\sigma_{\text{res}}(e) = \sigma_{\text{res}}(\gamma) = \sqrt{\left(\frac{0.03}{p_T}\right)^2 + (1.3 \times 10^{-3})^2} \quad \text{and} \quad \sigma_{\text{res}}(\mu) = \sqrt{\left(\frac{0.01}{p_T}\right)^2 + (1.0 \times 10^{-4})^2}. \quad (7.2)$$

Hence, the definition of the bin edges uses these expressions to estimate the resolution, using the approximation $p_T \approx m/2$.

Also, mass histograms with no jet in the combination had their bin width doubled to avoid spikes due to very small resolutions, while histograms including E_T^{miss} had their bin widths halved to increase the number of bins in these histograms. We can see the impact of the calibration with respect to the experimental resolution for histograms from the Dark Machines dataset in Fig. 7.8. We observe that the detector has a better resolution for the lower-mass region and for combinations with more jets, which corresponds to smaller bin widths in the low-mass region.

Furthermore, when constructing the histograms, it is expected that some peaks will naturally appear due to known SM resonances. For instance, the Z boson is expected to create a peak at around 90 GeV in the reconstruction of two leptons or two jets. To address this issue and guarantee that the histograms will have no resonance, histograms that are expected to have the Z peak are divided into two separate histograms: one *before* Z containing the lower-mass bins up to 90 GeV, and another *after* Z , containing the higher-mass bins starting at 90 GeV.

7.6 Training

A neural network model is as good as its training dataset. Therefore, a training dataset that contains a large variety of background types, histogram sizes and dynamic ranges is constructed, to make sure it is representative of the real data on which the network will be applied. Background shapes are obtained from the 11 analytical functions, shown in Table 7.1, and from fits to the Dark Machines background histograms from channel 3 and channel 2b, since the former channel is more inclusive and the latter improves the sensitivity of certain BSM signals by requiring leptons. A total of approximately 1.5×10^5 different background shapes is obtained, being 1.1×10^5 shapes coming from analytical functions, 31662 from fits to channel 3 and 8104 from fits to channel 2b.

Synthetic data is generated for the training dataset, following the process described in Section 7.4.2. Namely, each example is constructed from the combination of one of the background shapes with a Gaussian signal centred at a random bin and scaled to correspond to significances randomly chosen from the interval $[1, 10]$. The combined background and signal is then Poisson-fluctuated, meaning a noise taken from a Poisson distribution is added for each bin. Finally, the significance is calculated with respect to the original background shape.

A total of approximately 2.3×10^6 examples are included in the training dataset, with backgrounds corresponding to a ratio of 1:0.5:0.5 of analytical functions, channel 3 and channel 2b fits, respectively. To match these ratios, each background shape was included multiple times, but with different signal and Poisson fluctuations, guaranteeing an independent histogram. Moreover, a validation dataset was generated with approximately 2×10^5 examples each⁽³⁸⁾. These examples were generated with the same procedure, but with different random seeds, such that the position and strength of the signal, as well as the noise, would be different.

The model was trained for 300 epochs with a learning rate of 0.001 and a batch size of 5000. The loss function used is the MSE, that is, a squared L_2 norm given by

$$L(x^{\text{pred}}, x^{\text{target}}) = \{l_1, \dots, l_N\}^T, \quad l_n = (x_n^{\text{pred}} - x_n^{\text{target}})^2,$$

where x^{pred} is the network's prediction and x^{target} is the target value, in units of significance. Moreover, a linear transformation is used to scale the values range to the $[0, 1]$ interval, a choice done to optimise calculations in computers. The total loss is then defined as the sum of the mean of two losses, one obtained from the prediction of the target significance Z and another from the prediction of the background shape B , that is, $L = \overline{L_Z} + \overline{L_B}$. During training, this loss is minimised using the ADAM optimiser^[171]. The inclusion of the term L_B in the loss corresponds to the network's prediction for the background shape. This attempts to further quantify how the network is estimating the significance of each histogram and it has been

(38) A validation dataset is used at the end of each epoch to estimate the loss, while a testing dataset is used to obtain the network's performance after the training is finished.

shown that its inclusion does not impact the performance of the significance prediction. That being said, we will focus on the performance of the significance prediction from the network.

7.7 Performance over synthetic data

To assess the performance of the network, predictions are obtained from synthetic datasets and compared to their corresponding PLR target values. These are generated using the same procedure and parameters as the training and validation datasets, but with a different seed to guarantee that the histograms are different. All testing datasets were generated with approximately 2×10^5 examples, all having different compositions.

In the testing datasets, three different types of background shapes were considered: analytical functions, fits to channel 3 and fits to channel 2b. A dataset with all backgrounds was generated, together with two separate datasets with analytical functions only and the Dark Machines fits only. For each one of these background compositions, a sample with signal in all samples was generated (signal sample), while another one had no signal injected (background-only sample). Note that even if no signal was injected in the background-only sample, a Poisson fluctuation was performed and the corresponding PLR significance was calculated. Finally, a testing sample with the raw channel 3 Dark Machines background histograms was considered.

Some examples of histograms in the testing dataset are shown in Fig. 7.9. The observed histogram (black dots) is the input given to the network, being obtained by the combination of the background (orange line) and the signal (green line) with a Poisson noise. The network's prediction (red line) is compared to the PLR test result (blue line), corresponding to the network's target.

It is clear from these samples that the network's predicted significance has a very good agreement with the one obtained via the PLR test, which corresponds to the true significance as obtained with the optimised test statistic (vide Section 2.4.1). Not only it can correctly identify the position and strength of injected signals, but it also captures the oscillations of the noise, even if the strength of these oscillations is not always perfectly matched.

In particular, we encounter scenarios where the signal is so faint that it gets lost within the noise, such as the case found in Fig. 7.9a, while in other cases the noise creates fluctuations above 5σ even when there is no signal injected, as is the case in Fig. 7.9b. Overall, BumpNet is shown to properly match the expected significance shape, as in Fig. 7.9c, even if in some cases the predicted values are found within a certain range of the target ones, as we see in Fig. 7.9d.

Now, the main output of the network we are interested in is the maximum value of the significance $Z_{\max}$, since it would indicate the position and strength of the signal peak. As a result, the performance of the network is quantified in terms of how close the maximum predicted significance $Z_{\max}^{\text{BumpNet}}$ approaches the maximum target PLR significance $Z_{\max}^{\text{PLR}}$. Furthermore, the difference between these values $\Delta Z_{\max} = Z_{\max}^{\text{BumpNet}} - Z_{\max}^{\text{PLR}}$ is the main figure of

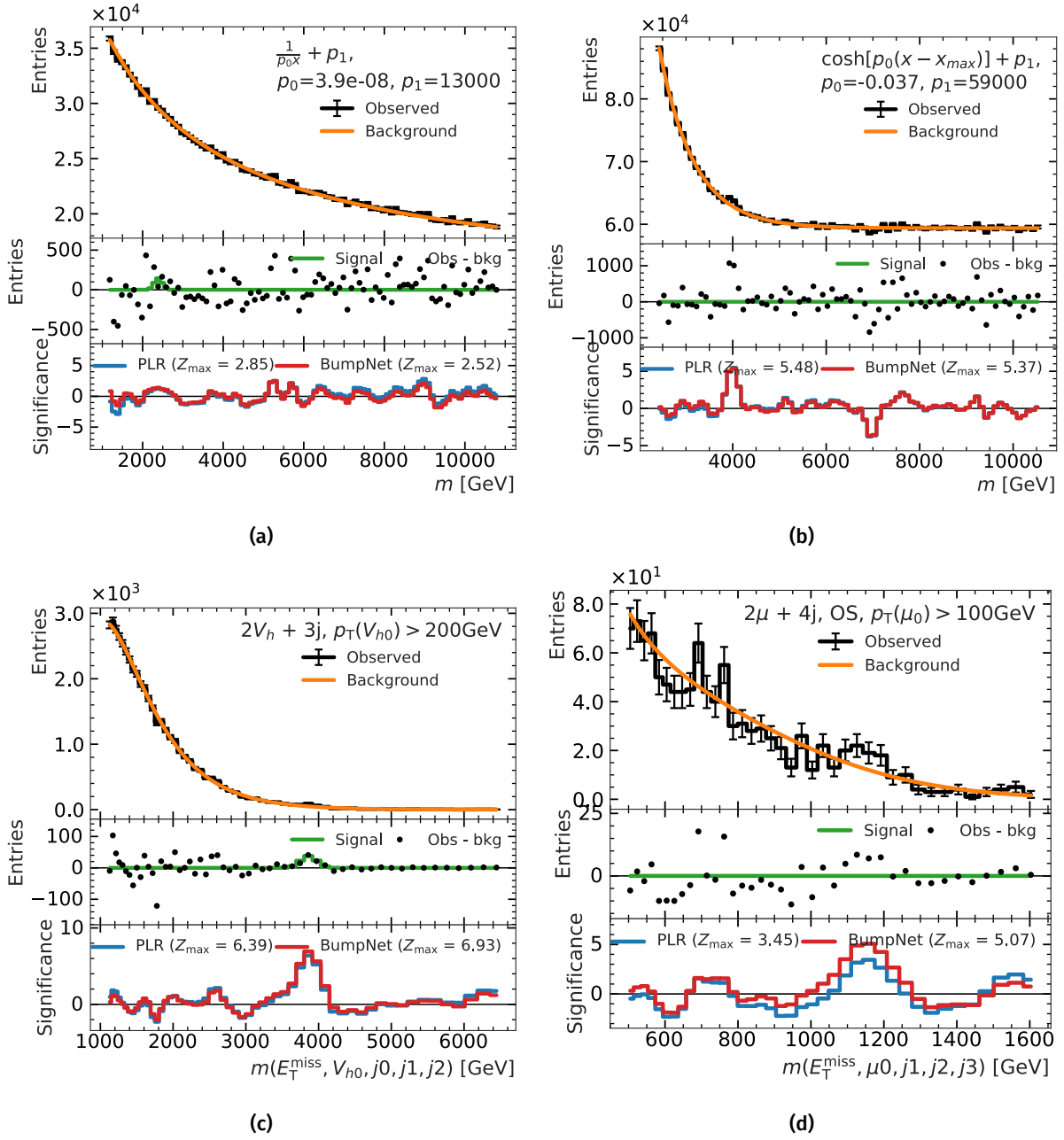


Figure 7.9: Examples from the testing datasets with the target PLR significance (blue) and the corresponding network prediction (red). The observed histograms (black) are obtained from the combination of a smoothly falling background (orange) with a Gaussian signal (green), the latter being scaled to match a given strength. A Poisson noise is added to the combination. The maximum Z value, $Z_{\max}$, from both the target and the prediction are quoted, being expected to match the position and strength of the injected signal.

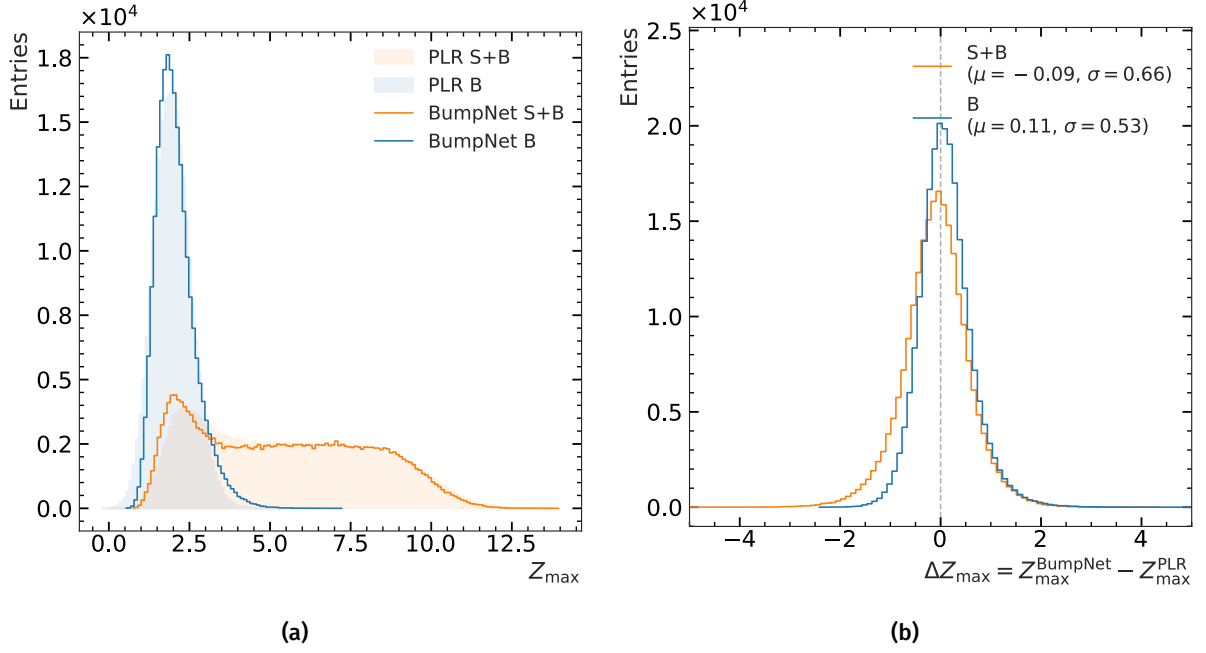


Figure 7.10: Distributions of (a) $Z_{\max}$ and (b) $\Delta Z_{\max}$ for BumpNet predictions compared to the PLR test results when evaluating in datasets with background-only (B) and background with signal injected (S+B).

merit considered, as it should be close to zero with a small variance. It must be noted that, after some investigation, we observed that edge effects make predictions in the first 10% of the bins not reliable. To account for this limitation, $Z_{\max}$ is taken as the maximum value within bins after this threshold.

A comparison between the distributions of $Z_{\max}$ from the network's prediction and the PLR test is shown in Fig. 7.10a. There is a good agreement between them when testing over datasets with either background-only and background with signal injected. Similarly, the distributions of $\Delta Z_{\max}$ are shown in Fig. 7.10b. We note that the network presents no bias, having a mean of $\mu \approx 0.1$ and a variance of $\sigma \approx 0.5\text{--}0.6$.

This performance is stable across different variables, as shown in Fig. 7.11. For different values of the injected signal strength and signal position, Figs. 7.11a and 7.11b, the network's prediction presents no bias. However, there is a greater variance when the signal is injected closer to the edges, with $\Delta Z_{\max}$ ranging between -4 and $+4$. No bias is observed for different histogram characteristics such as the number of bins and entries per bin in the histograms, as shown in Figs. 7.11c and 7.11d. A slight increase in variance, nevertheless, is observed for histograms with fewer than 50 bins, with $|\Delta Z_{\max}|$ getting up to 4. When tested separately over the different types of backgrounds, the performance is also stable, as shown in Fig. 7.12.

Moreover, the discriminating power of the network can be assessed by constructing ROC curves, as discussed in Section 2.5. For a given value of significance Z , a prediction is considered a true positive if there was a signal in the histogram and the predicted maximum is above the considered threshold, $Z_{\max} \geq Z_{\text{thr}}$, while we have a false positive if the same happens

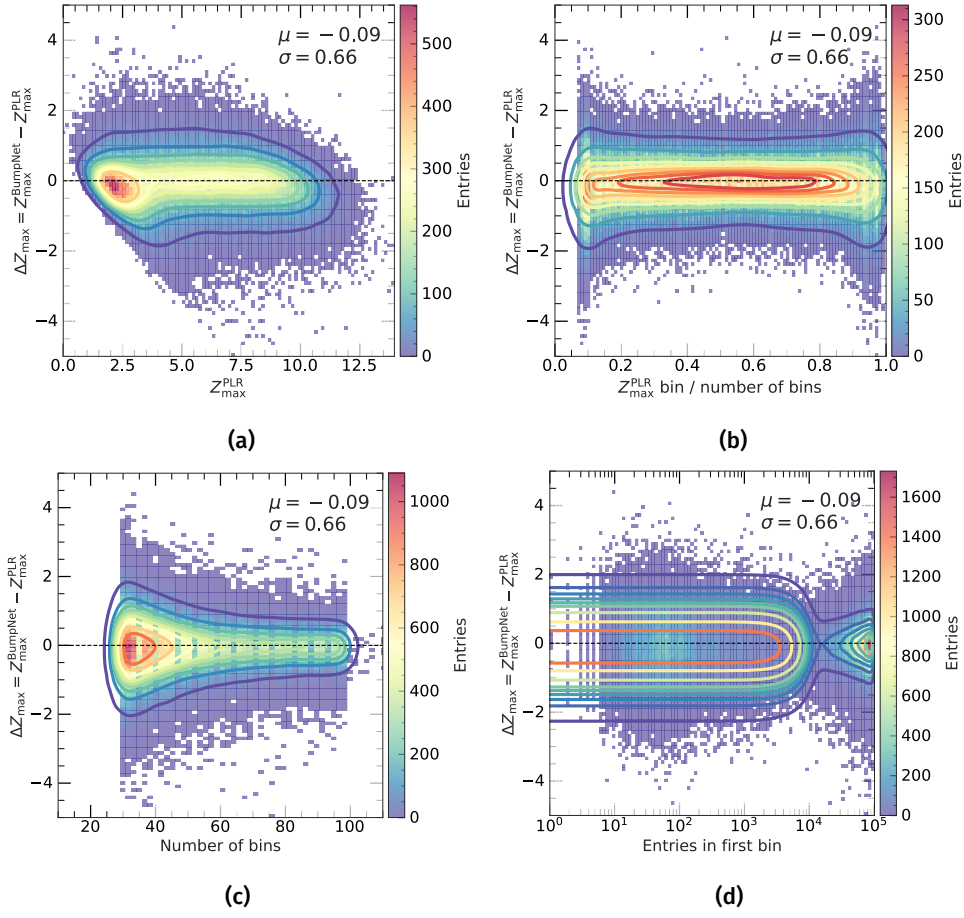


Figure 7.11: Distributions of $\Delta Z_{\max}$ as a function of (a) signal strength obtained via the PLR test $Z_{\max}^{\text{PLR}}$, (b) the signal position as a ratio of the histogram binning, (c) the number of bins in the histograms and (d) the number of entries in the first bin of the histogram.

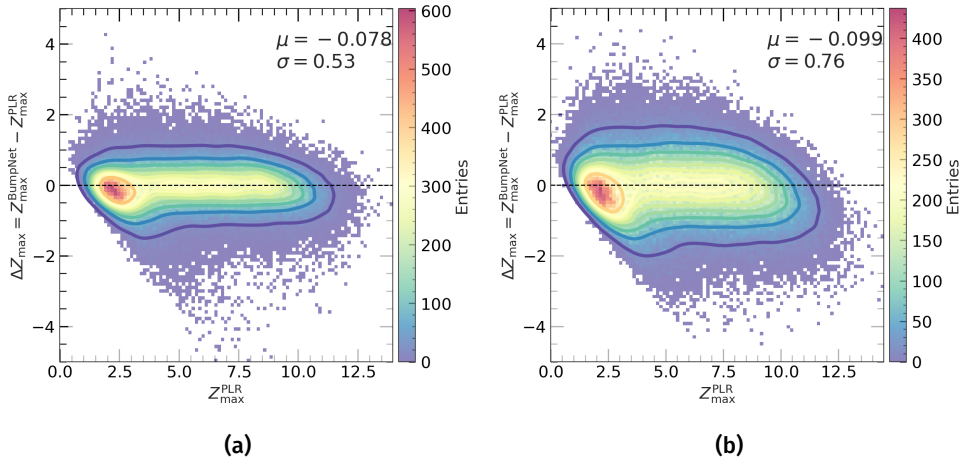


Figure 7.12: Distributions of $\Delta Z_{\max}$ as a function of the signal strength obtained via the PLR test $Z_{\max}^{\text{PLR}}$ when evaluating over testing datasets with backgrounds from (a) functions only and (b) fits to Dark Machines backgrounds only.

in a background-only histogram. In this way, the TPR and the FPR, defined from Eq. (2.30) are given by

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad \text{and} \quad \text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}},$$

where

TP = number of S+B histograms with $Z_{\max} \geq Z_{\text{thr}}$ (sensitivity),

TN = number of B histograms with $Z_{\max} < Z_{\text{thr}}$ (specificity),

FP = number of B histograms with $Z_{\max} \geq Z_{\text{thr}}$ (type 1 error),

FN = number of S+B histograms with $Z_{\max} < Z_{\text{thr}}$ (type 2 error).

Using these definitions, ROC curves are constructed for both the BumpNet prediction and the target PLR significance values, as shown in Fig. 7.13a. Each point in the curve corresponds to a bin from the $Z_{\max}$ distributions, shown in Fig. 7.13b. Since there are very few examples with $Z_{\max} \geq 5$ in the background-only sample, the rates for bins beyond this threshold correspond to the high density of points with low false-positive rate. As we go lower, with $Z_{\max} \leq 3$, the overlap of the signal + background and background-only distributions entails the ambiguity that increases the false- and true-positive rates.

An important metric to quantify the discrimination performance of the network is the value of the AUC, which is also shown in Fig. 7.13b. The PLR performance value, given by 0.9, represents the upper limit the network can achieve, due to the LEE effect. By achieving a score of 0.88, the network has an excellent discrimination power, already very close to the ideal scenario of the PLR test.

When performing a search with BumpNet, it is crucial to define a threshold above which we will define a histogram as *interesting*. Therefore, it is important to assess the rate of *false discoveries* the network would claim in a background-only dataset for different threshold values. These values were obtained for the background-only dataset with Poisson-fluctuated functions and Dark Machines fits, but also for the raw backgrounds from Dark Machines channels 3 and 2b, as shown in Table 7.5. It is a convention in particle physics to claim discovery only when the threshold of 5σ significance is obtained, while results with 3σ are referred to as evidence of phenomena. Over the fluctuated backgrounds, BumpNet predicts 0.1% of histograms with $Z_{\max} \geq 5$, to be compared to an expected 0.003% rate coming from the PLR values. When predicting over the more realistic raw channel 3 and channel 2b backgrounds, we observe 0.2% and 0.5%, respectively. Some examples of predictions with $Z_{\max} \geq 5$ obtained in the Dark Marchines' channel 3 raw backgrounds are shown in Fig. 7.14. Many of these histograms actually present a bump-like structure in the background, as is the case in Figs. 7.14a and 7.14b. In some other cases, such as Fig. 7.14c, it is not as clear if there is a physical explanation for the shape. These results demonstrate that the network is powerful enough to catch features of the

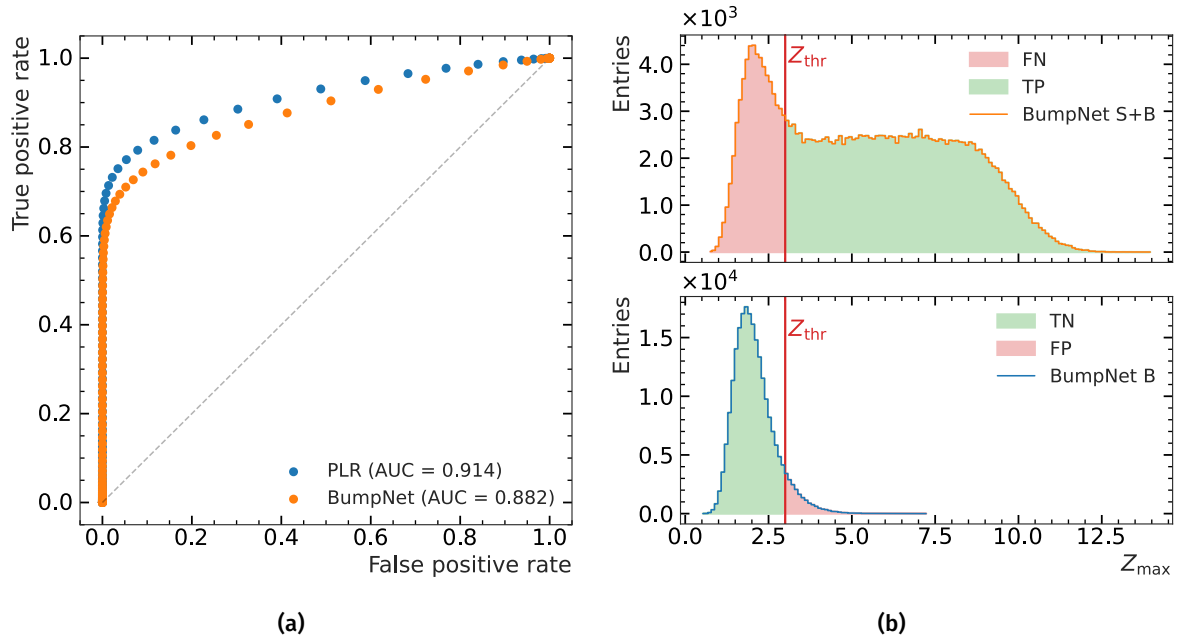


Figure 7.13: (a) ROC curves for both the BumpNet model and the PLR test, as well as the corresponding AUC score. (b) Visualisation of the definitions for sensitivity (TP), specificity (TN), type 1 (FP) and type 2 (FN) errors from the $Z_{\max}$ distributions.

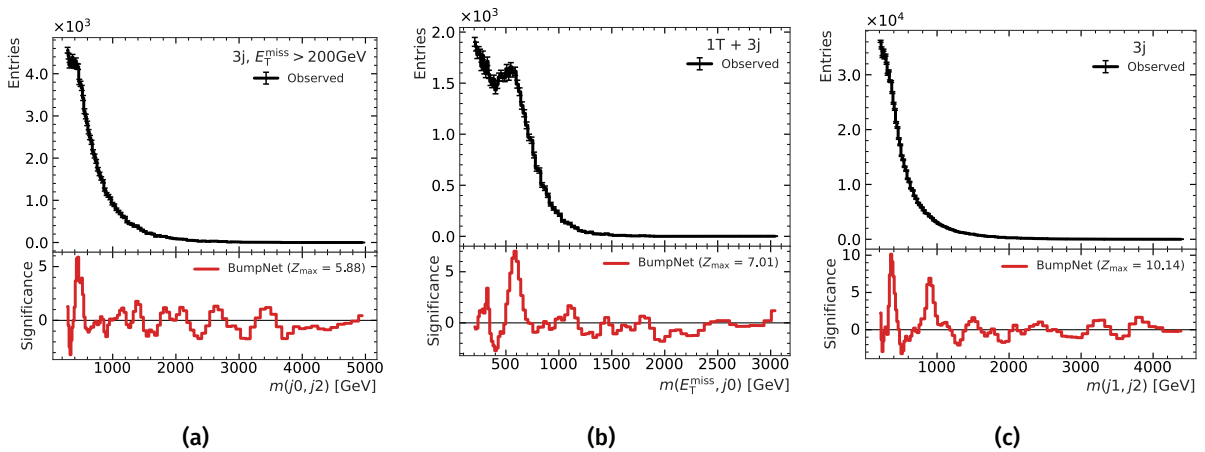


Figure 7.14: Examples of false discoveries obtained by the BumpNet model when evaluating over the raw channel 3 Dark Machines backgrounds.

Table 7.5: False discovery rate in percentage of the total number of histograms and corresponding number of histograms obtained from the PLR test and the BumpNet predictions over the different background datasets. Synthetic background is generated by Poisson fluctuation of fits to Dark Machines backgrounds, while raw channel 3 and channel 2b backgrounds correspond to the original histograms.

	PLR synthetic	BumpNet synthetic	BumpNet raw channel 3	BumpNet raw channel 2b
$Z_{\max} \geq 3$	4.6% (10539)	8.22% (18849)	7.06% (2387)	8.62% (728)
$Z_{\max} \geq 4$	0.123% (281)	0.997% (2287)	1.1% (371)	1.92% (162)
$Z_{\max} \geq 5$	0.00349% (8)	0.102% (234)	0.228% (77)	0.509% (43)
$Z_{\max} \geq 6$	0% (0)	0.00829% (19)	0.0799% (27)	0.272% (23)
$Z_{\max} \geq 7$	0% (0)	0.000436% (1)	0.0473% (16)	0.166% (14)

background, which were sculpted by the cuts performed when selecting the events. However, the false-discovery rate cannot be used to fully assess how many bad predictions the network makes, thus another supporting metric needs to be constructed.

One way to address this problem is to incorporate the intrinsic uncertainty on the predicted $Z_{\max}$. In the absence of any bias, that is, since $\mu \approx 0$, we quantify this uncertainty as the variance on the $\Delta Z_{\max}$ distribution, given by $\sigma \approx 0.5\text{--}0.6$ (vide Fig. 7.10b). If we suppose this distribution to be Gaussian, predictions with $\Delta Z_{\max}$ beyond a given threshold in units of this variance can be considered *outliers*, as they go beyond the expected error of the network, as shown in Fig. 7.15. Hence, we calculate an *outlier rate* as the number of histograms with $\Delta Z_{\max}$ above a given threshold divided by the total number of histograms in the sample, as shown in Table 7.6. We observe that, when evaluating over samples with signal+backgrounds and background only, the $\Delta Z_{\max}$ distribution is approximately Gaussian for up to $\Delta Z_{\max} \leq 3 \times \sigma$. The rate of histograms beyond this threshold is of 1.02% when there is signal and 1.21% for background only. If we consider only histograms that have a $Z_{\max}^{\text{BumpNet}} \geq 5$ discovery claim, we observe that only 0.08% of histograms with background-only are outliers, while we have 0.46% for examples with signal injected. This means, in summary, that when evaluating over a sample with only fluctuated backgrounds, BumpNet’s model predicts 0.1% of histograms with $Z_{\max}^{\text{BumpNet}} \geq 5$, but only 0.08% have $|\Delta Z_{\max}| \geq 3 \times \sigma$.

Lastly, another way to quantify the reliability of the network’s prediction is the confidence interval on the predicted $Z_{\max}$ value. For this purpose, the mean and variance of the predicted $Z_{\max}^{\text{BumpNet}}$ for different bins of the target significance $Z_{\max}^{\text{PLR}}$ are shown in Fig. 7.16. On the other hand, it is clear that the narrow confidence interval in the network predictions in the signal + background dataset remains stable as the strength of the signal increases.

In summary, the BumpNet network is robust enough to properly predict statistical significance for bumps in invariant mass distributions, with a quantifiable variance on how far

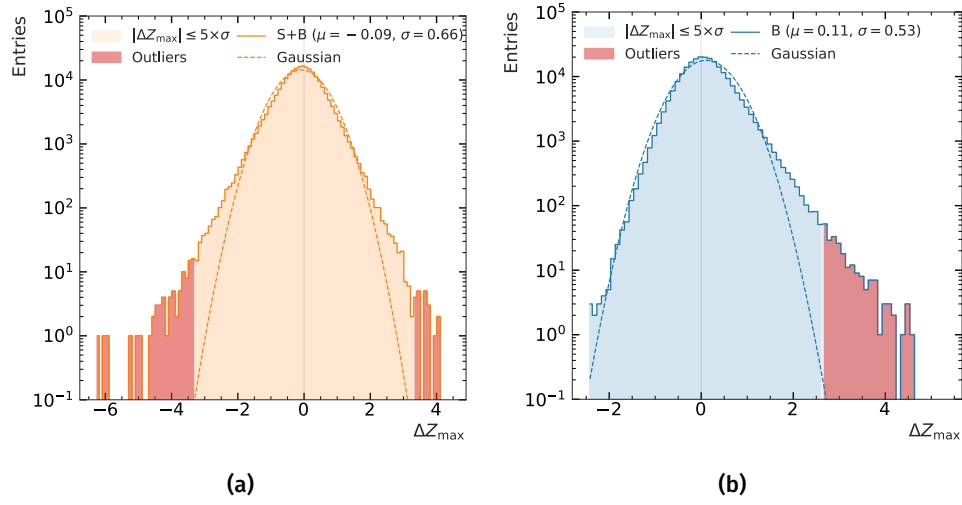


Figure 7.15: Distributions of $\Delta Z_{\max}$ over the (a) background-only (B) and (b) signal+background (S+B) testing datasets, which have mean μ and variance σ . Predictions with $|\Delta Z_{\max}| > 5 \times \sigma$ (red) are considered outliers.

Table 7.6: Outlier rate and corresponding number of histograms obtained from BumpNet predictions over the datasets with background only (B) and signal + background (S+B), using the corresponding variance σ for each dataset. Rates are given as percentages of the total number of histograms in the testing dataset.

	All histograms		$Z_{\max}^{\text{BumpNet}} \geq 5$	
	B ($\sigma = 0.53$)	S+B ($\sigma = 0.66$)	B ($\sigma = 0.53$)	S+B ($\sigma = 0.66$)
$ \Delta Z_{\max} \geq 1 \times \sigma$	27.2% (62252)	27.4% (62719)	0.0977% (2244)	14.5% (33178)
$ \Delta Z_{\max} \geq 2 \times \sigma$	5.44% (12465)	5.52% (12668)	0.0907% (208)	2.84% (6515)
$ \Delta Z_{\max} \geq 3 \times \sigma$	1.21% (2778)	1.02% (2329)	0.0798% (183)	0.465% (1067)
$ \Delta Z_{\max} \geq 4 \times \sigma$	0.313% (7180)	0.213% (489)	0.0637% (146)	0.0842% (193)
$ \Delta Z_{\max} \geq 5 \times \sigma$	0.0986% (226)	0.0458% (105)	0.0427% (98)	0.014% (32)

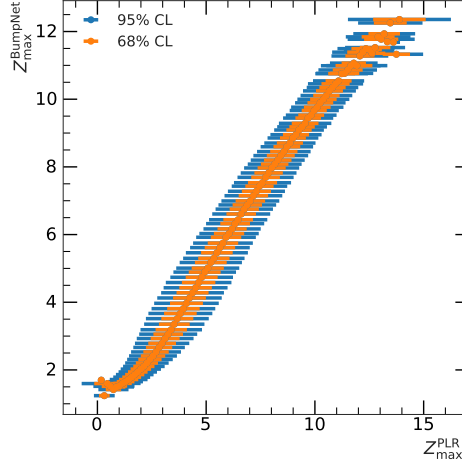


Figure 7.16: Confidence interval for the predicted $Z_{\max}^{\text{BumpNet}}$ as a function of the target $Z_{\max}^{\text{PLR}}$ from a signal+background testing dataset.

the predicted values are from the target PLR test. As a result, a threshold on $Z_{\max}$ values can be defined, based on the number of fake predictions we are ready to accommodate. Once this threshold is agreed upon, the BumpNet network can reliably select mass distributions that present potential for excess in the form of a Gaussian bump. In the following sections, we discuss studies performed to test the robustness of the network when varying certain aspects of its development, such as the background composition of the training dataset, the signal width and the introduction of systematic uncertainties between the training and application datasets.

7.7.1 Impact of background shapes

If we assume the background to be smoothly falling, one could expect that using only analytical functions in the training dataset would be enough to obtain a performing network. However, a model trained only in analytical functions does not perform well when evaluated on Dark Machines backgrounds, as shown in Fig. 7.17.

To guarantee a fair comparison, the model was obtained from training datasets with the same number of training examples, but using analytical functions only. When evaluated on a functions-only testing dataset, the model present no bias and a variance of 0.5, as shown in Fig. 7.17a. But when evaluating over a testing dataset including Dark Machines fits only, shown in Fig. 7.17b, the model trained over functions only has a poor performance, especially with its many predictions with $\Delta Z_{\max}$ up to 30σ . This is not observed in the model that had Dark Machines fits included in the training, as previously shown.

After close inspection, we see that the network identifies the change in concavity present especially in the beginning of the histograms as a part of a large bump, as shown in Fig. 7.18. This feature of the Dark Machines histograms can be explained by the cut on H_T present in both the channel 3 and channel 2b histograms. Removing this cut would reduce this sculpting,

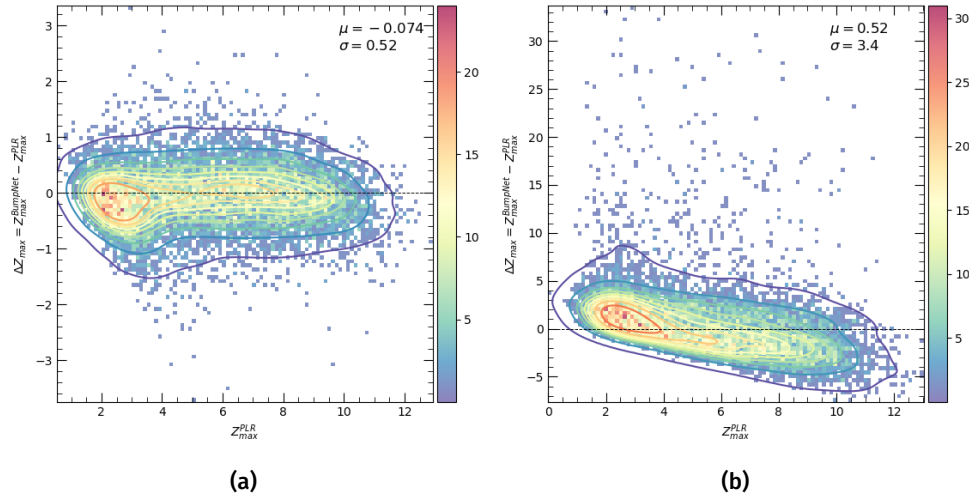


Figure 7.17: $\Delta Z_{\max}$ as a function of $Z_{\max}^{\text{PLR}}$ for a model trained in functions only when evaluated on (a) functions and (b) fits to channel 3 and channel 2b backgrounds.

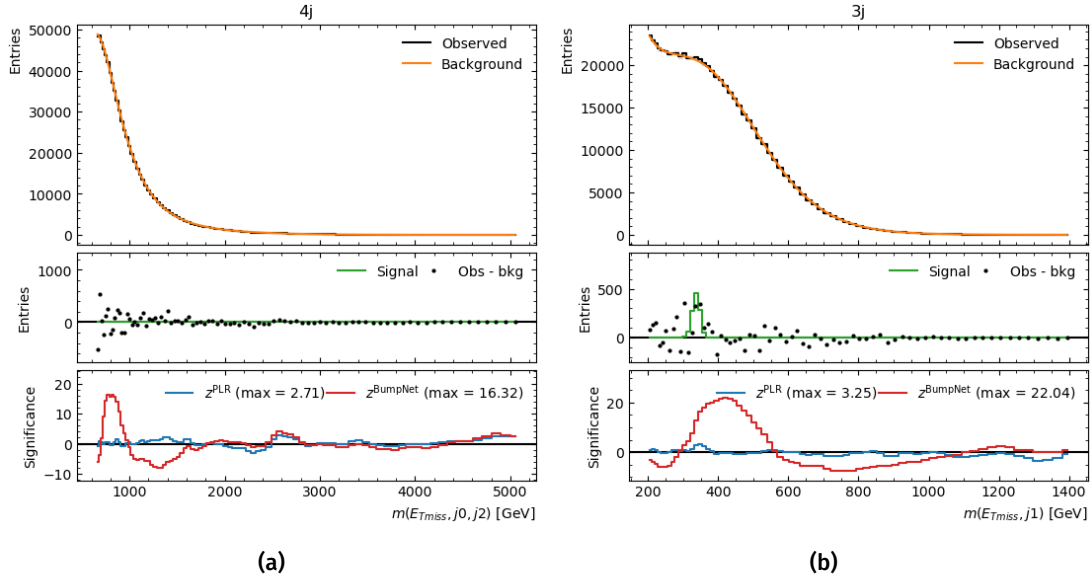


Figure 7.18: Examples of predictions of model trained on backgrounds from analytical functions only and evaluated over raw backgrounds from the Dark Machines dataset.

but we do expect that some data histograms will have some form of delayed smoothly falling shape, due to the fact that the histogram starts at the bin with the maximum number of entries.

Hence, the inclusion of fits to simulation data into the training dataset makes it more representative of the real data on which the network will be applied, helping to make it robust against this kind of behaviour. It is important to emphasise that since the shapes that are used in the training are fitted functions to the MC simulation, they will introduce some of the problematic mismodelling between MC simulation and data into the analysis. Nevertheless, the network will be exposed to many different background shapes during the training, shapes which could eventually even account for different systematics. This means that, after the training, the network should be robust against the systematic uncertainties of individual shapes, as will be shown in the following section. Thus, by compromising with this dependence on MC, the network can achieve good performance in more realistic applications.

7.7.2 Impact of systematic uncertainties

As we have seen, incorporating fits to simulation data into the training is important to guarantee a good performance. These histograms are obtained with the exact same selections as the real data, but since they rely on MC simulation, they reflect our understanding of the ATLAS data. As a result, the systematic uncertainties that parameterise our knowledge of detector effects will bias our training data with respect to the real data. Thus, to guarantee that the network's prediction is reliable, we must also make sure it is robust against these systematic uncertainties.

To test this, different testing datasets were created where a series of distortions were introduced into the background shapes obtained from the fits to the Dark Machines backgrounds. These distortions attempted to mimic and exaggerate effects such as the jet energy scale, where we would see an increase on the number of events at lower mass together with a deficit at the higher-mass region. In this study, the following transformations were applied to the background functions:

$$\begin{aligned} \text{Standard bias: } y'_i &= 1.5 \frac{N_{\text{bins}} - x_i}{N_{\text{bins}}} y_i, \\ \text{Reverse bias: } y'_i &= 0.5 \frac{N_{\text{bins}} - x_i}{N_{\text{bins}}} y_i + y_{20}, \\ \text{Inferior bias: } y'_i &= 0.5 y_i, \\ \text{Superior bias: } y'_i &= 1.5 y_i, \end{aligned}$$

where x_i and y_i respectively are the bin number and the number of entries in bin i , with N_{bins} being the total number of bins in the histogram. In the reverse bias, the number of entries in the 20th bins y_{20} was added to ensure there would be no zero entries, since all histograms had

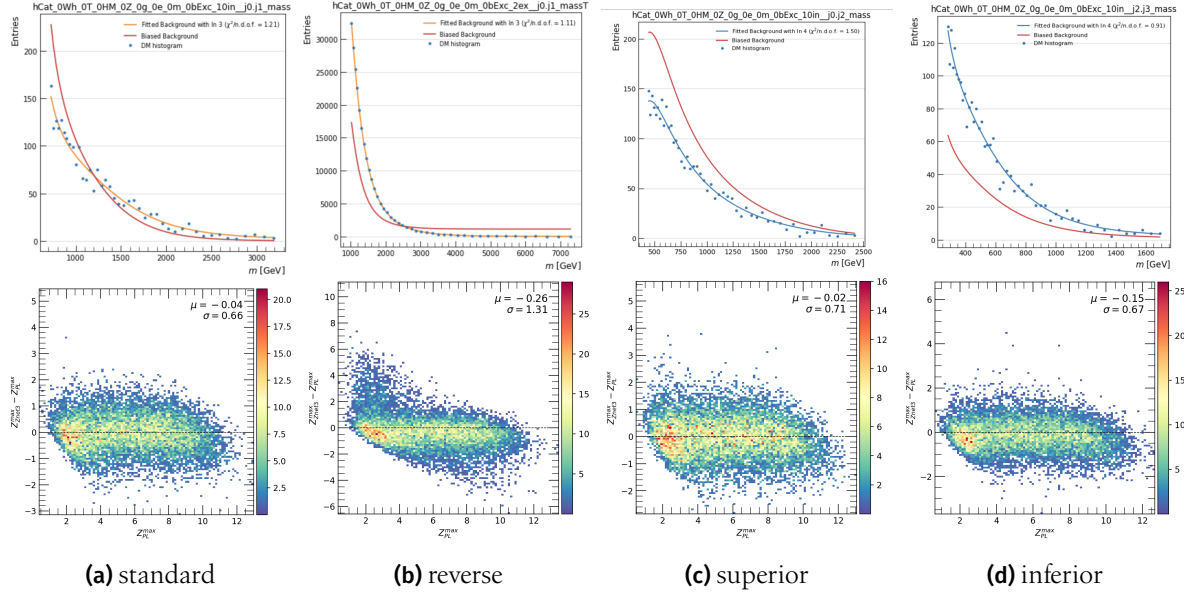


Figure 7.19: Performance over testing datasets with different artificial systematic bias (top) comparing the distributions of $\Delta Z_{\max}$ as a function of the target signal strength $Z_{\max}^{\text{PLR}}$ (bottom).

at least 20 bins.

Thus, biased testing datasets were obtained using the synthetic data generation procedure, as described in Section 7.4.2, and contained background shapes coming only from the Dark Machines fits. A model was then trained using examples generated from backgrounds coming from both analytical functions and Dark Machines fits, with injected signals of 1-bin width. We note that the training examples were nominal histograms, where no bias was injected, such that the systematic bias was introduced between the training and the testing examples.

When tested against these biases, the model presented the same good performance, as shown in Fig. 7.19. Thus, these studies show that the network is robust against these linear distortions applied over the testing datasets. It must be noted that all histograms with the distortions could still be fitted by the same 4th-order log-polynomial that is familiar to the network and this might not be the case for a more complex systematic bias. But the systematic distortions studied here represent a very large extrapolation of the realistic systematic uncertainties within ATLAS data. Hence, these results are encouraging as they show the stability of the network against the systematic uncertainties that might exist between the training sample and the application data.

7.7.3 Impact of signal width

As discussed in Section 7.5.2, the binning of invariant-mass histograms plays an important role in defining if the signal will appear as a bump. By calibrating it to match the detector resolution, we expect narrow signals to span over just a few bins. For this reason, the network is trained with signals of a fixed width of one bin.

In order to test the robustness of this choice, the BumpNet model was tested over datasets

following the same composition of backgrounds as the training sample, but where the injected signal had different widths, ranging from 1–5 bins. The values of the mean μ and variance σ of $\Delta Z_{\max}$ on these different datasets can be found in Fig. 7.20.

As the signal width increases, there is an increase in the bias from the network. A comparison of the predicted significance for histograms with the same background shape, but with different signal widths injected is shown in Fig. 7.21. As the signal gets wider, the network is no longer capable of distinguishing it from the background noise. Thus, by choosing to use the model trained with 1-bin width signals, the network might be able to catch some wider signals, but its $Z_{\max}$ prediction will considerably underestimate its strength. Also, this observation reiterates the importance of properly calibrating the histograms and adjusting their bin widths to represent the realistic resolution observed in the data.

7.8 Performance when looking for new physics

So far, we have shown that a trained BumpNet model has a great performance when predicting the significance of bumps in synthetic data generated from both functions and fits to simulation data. Now, it is also important to test how the network would perform in realistic signals. To do so, datasets were prepared by adding MC-simulated signal events to the backgrounds available from the Dark Machines dataset.

Firstly, an RPVSUSY^[123,129] stop pair production is studied. Events which were generated within the Dark Machines project were added to both channel 3 and channel 2b backgrounds. As described in Section 1.8, the stop squark is the top quark superpartner in a version of the SUSY theory where the R-parity symmetry is violated, meaning that processes do not need to conserve the number of sparticles. It was generated with 1 TeV mass and is produced in pairs, each of them decay to a lepton (e or μ) and a b -jet.

Another BSM signal available from the Dark Machines datasets is a Z' in a lepton-violating $U(1)_{L_\mu-L_\tau}$ model^[144,145]. Generated with a mass of 50 GeV, this particle decays to two opposite-sign muons and is produced in association to either one or two additional muons, vide Fig. 1.10 in Chapter 1. In this dataset, only the $3\mu + \nu$ final state was included and, since it is expected to be reconstructed in histograms with the mass of 2 muons, it was combined only with channel 2b backgrounds.

Finally, two datasets including leptoquarks decaying to $b\ell b\ell$ ($\ell = e, \mu$) final states were also created specifically to benchmark the BumpNet performance. Signal events were simulated using the same configurations that were used to produce the Dark Machines datasets, with the new particle having a mass of 600 GeV. These events were then added to channel 2b backgrounds, since the particle is produced in pairs, each one of them decaying to a lepton and a b -jet, vide Fig. 1.9 in Chapter 1. Since the particle is expected to be reconstructed in histograms of the mass of a lepton and a jet, this scenario is very similar to the RPV stop, but testing another mass value.

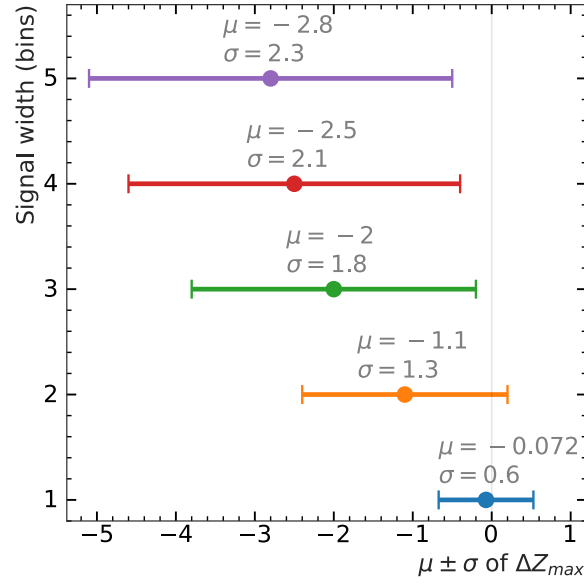


Figure 7.20: Mean μ and variance σ of the $\Delta Z_{\max}$ distributions when evaluating over testing datasets with different signal widths.

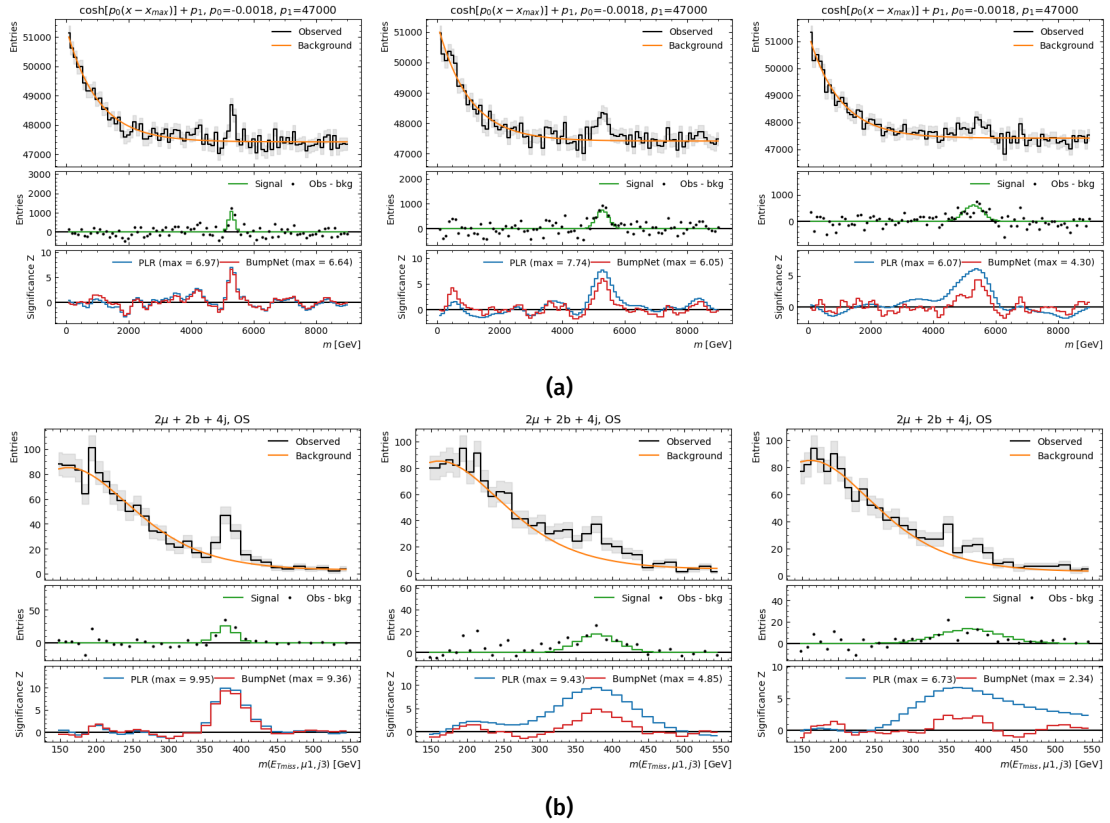


Figure 7.21: Examples of predictions over histograms with the same background, either an (a) analytical function or a (b) fit to Dark Machines background, but with different signal widths, namely 1-bin width (left), 2-bin width (center) and 3-bin width (right).

By testing the network in these simulated signals, we can prepare for what a real analysis on data would look like. For this reason, testing datasets were obtained by processing background and signal events together, following the processing and calibration described in Section 7.5.2, as it would be done to real experimental data. Once the histograms were obtained, they were given as input to the network and the predictions obtained were analysed.

To measure the performance of the network, heatmaps of the number of histograms as a function of the predicted $Z_{\max}$ and the predicted mass were drawn and are shown in Figs. 7.22 and 7.23. As a reference, predictions were also obtained from the channel 3 and channel 2b raw background-only datasets. As can be seen respectively in Figures 7.22a and 7.22b, the network predicts some excesses in both background-only datasets, mainly in the low-mass region. For the channel 3 background-only histograms, we observe a feature at around 500 GeV, while for the channel 2b histograms, a similar hint of excess is observed at masses smaller than 100 GeV. These excesses observed by the network correspond to real features in the background, but not to new physics.

When evaluating over the BSM signal datasets, an excess of predictions with $Z_{\max} \geq 5$ is observed around the corresponding injected signal mass of all particles considered. For the 1 TeV RPV stop, the excess is visible in both channel 3 (Fig. 7.22c) and channel 2b (Fig. 7.22d). Same is true for the 600 GeV leptoquarks decaying to both the *bebe* (Fig. 7.22e) and *b μ b μ* (Fig. 7.22f) final states in channel 2b. Finally, an excess is also observed for the Z' signal at 50 GeV, as shown in Fig. 7.23, although this is much less clear to see as many predictions with high significance are also claimed in the background-only datasets.

The great performance of the network is further established when we look at examples of histograms from these different datasets. In examples from the RPV stop datasets in channel 3 and channel 2b, shown in Fig. 7.24, the clear peak around the injected mass of 1 TeV was identified by the network in the mass of both 1 electron and 1 muon together with *b*-jet in both channels (Figs. 7.24a, 7.24b, 7.24d and 7.24e). Furthermore, an excess was also observed in histograms of the mass of 1 electron, 1 muon and 2 *b*-jets on channel 3 (Fig. 7.24c), as well as for 2 muons and 2 *b*-jets on channel 2b (Fig. 7.24f). The excess is found around 2 TeV, corresponding to the reconstruction of the entire stop pair. These examples reiterate the ability of the network to identify the resonances associated to the decay products of the new particles, being an important milestone for the establishment of this method.

Comparable performance is also found for the leptoquarks datasets. Resonances are found in both the *b ℓ b ℓ* ($\ell = e, \mu$) decay channels in the channel 2b, as shown in Fig. 7.25. A peak around 600 GeV is observed in histograms of the mass of an electron with a *b*-jet (Fig. 7.25a and Fig. 7.25b), as well as the mass of a muon and a *b*-jet (Fig. 7.25d and Fig. 7.25e). Once again, the pair is also reconstructed in the complete final state *bebe* (Fig. 7.25c) and *b μ b μ* (Fig. 7.25d), with a peak around 1.2 TeV.

A more challenging signal to find, the 50 GeV Z' , is hidden under a lot of background at low mass. Nevertheless, the network clearly identifies the peak in the mass of two muons,

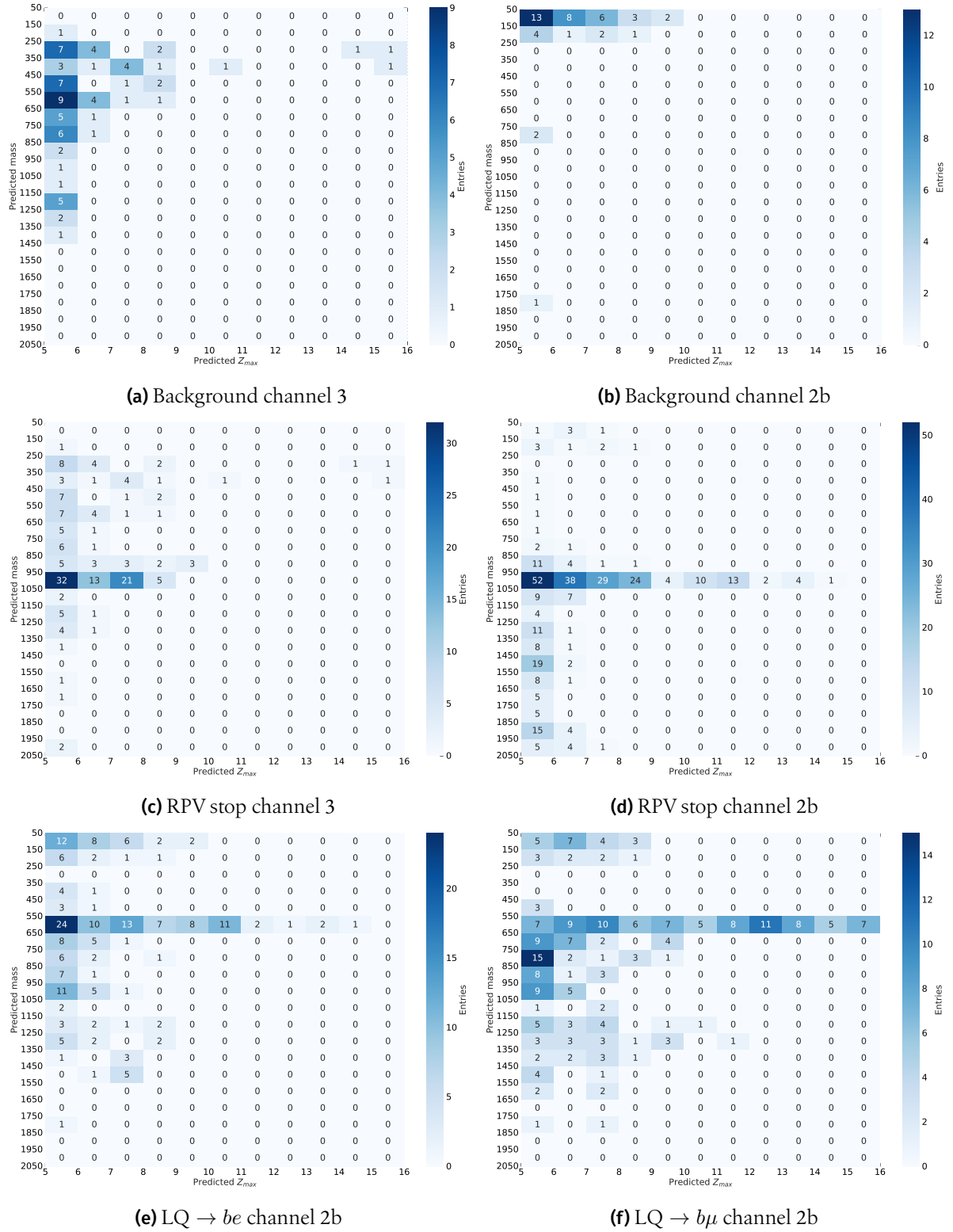


Figure 7.22: Number of histograms as a function of the predicted $Z_{\max}$ and the predicted mass for the raw Dark Machines background-only datasets and the BSM signal datasets with an RPV stop or leptoquarks (LQ).

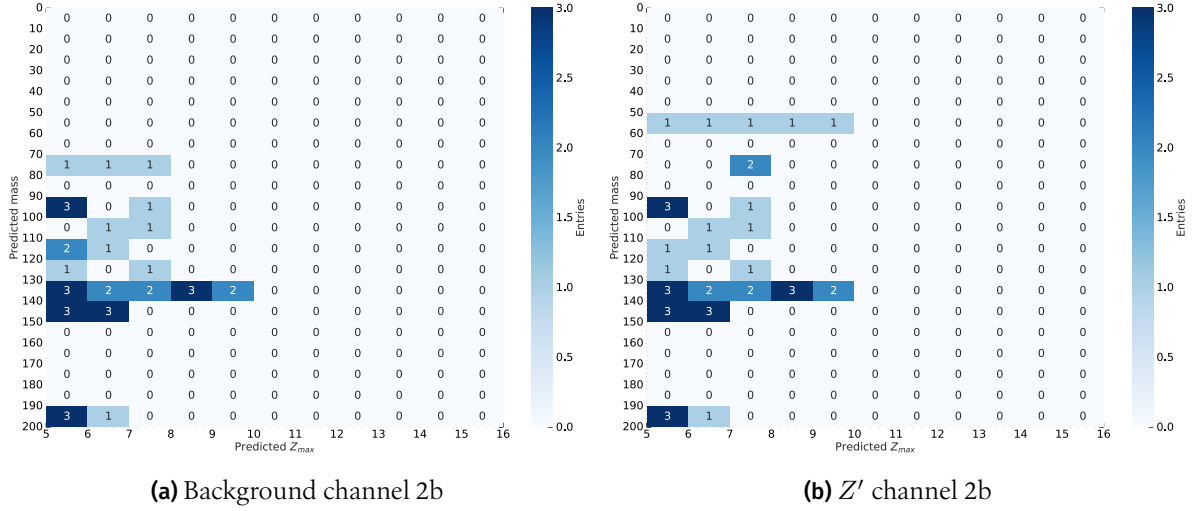


Figure 7.23: Number of histograms as a function of the predicted $Z_{\max}$ and the predicted mass for the Z' signal dataset and the raw Dark Machines channel 2b background-only dataset.

as shown in Fig. 7.26. Not only is the bump identified, but this happens in histograms that are slightly smoothly ascending, since this mass region is close to the beginning of the Z boson peak at 100 GeV.

An important observation throughout the different BSM signal datasets is that the resonant peaks are all found in histograms of the mass of the decay products with final states corresponding to multiple exclusive jet selections. In the case of the Z' , for instance, the peak is found in selections with exactly 1 jet (Fig. 7.26a), exactly 2 jets (Fig. 7.26b) and exactly 3 jets (Fig. 7.26c). The same observation can be made for the RPV stop examples (Fig. 7.24) and leptoquarks (Fig. 7.25). The chances that such an excess is observed in exclusive jet bins at the exact same place simply due to statistical fluctuations of the SM background are extremely unlikely. That said, if statistically significant resonances were to be observed in the experimental data occurring in multiple exclusive jet bins, the probability that the signal found would indeed be coming from a new physics resonance are considerably enhanced. For this reason, the splitting of the data into exclusive jet bins is an important strategy to reduce the LEE and help navigate the predictions performed by BumpNet.

The author of this thesis was involved in many different aspects of the work presented so far in this Chapter, with the main contributions including the implementation of the software for synthetic data generation, the training and evaluation of BumpNet models, performance studies on the false discoveries and outliers, as well as the analysis of the predictions over the BSM samples. Once established the potential of this method, the author also contributed to its implementation over ATLAS data, which is the subject of the following section.

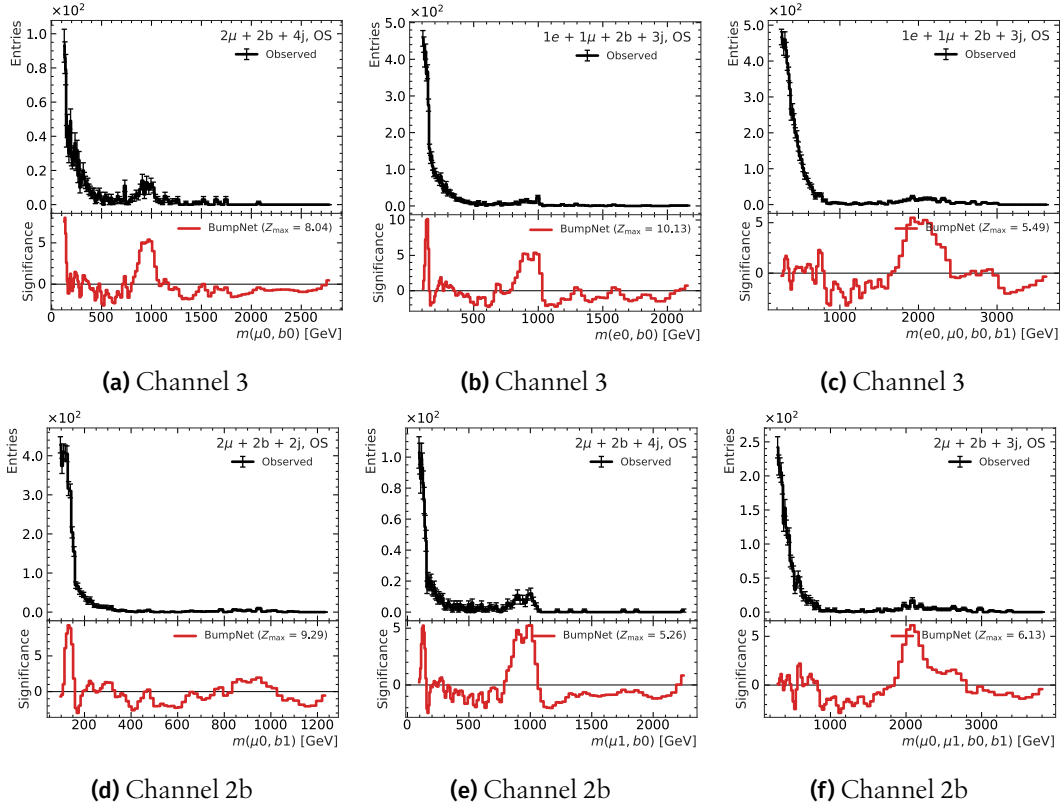


Figure 7.24: Examples of predictions over the Dark Machines datasets containing RPV stop particles with mass of 1 TeV from channel 3 and channel 2b.

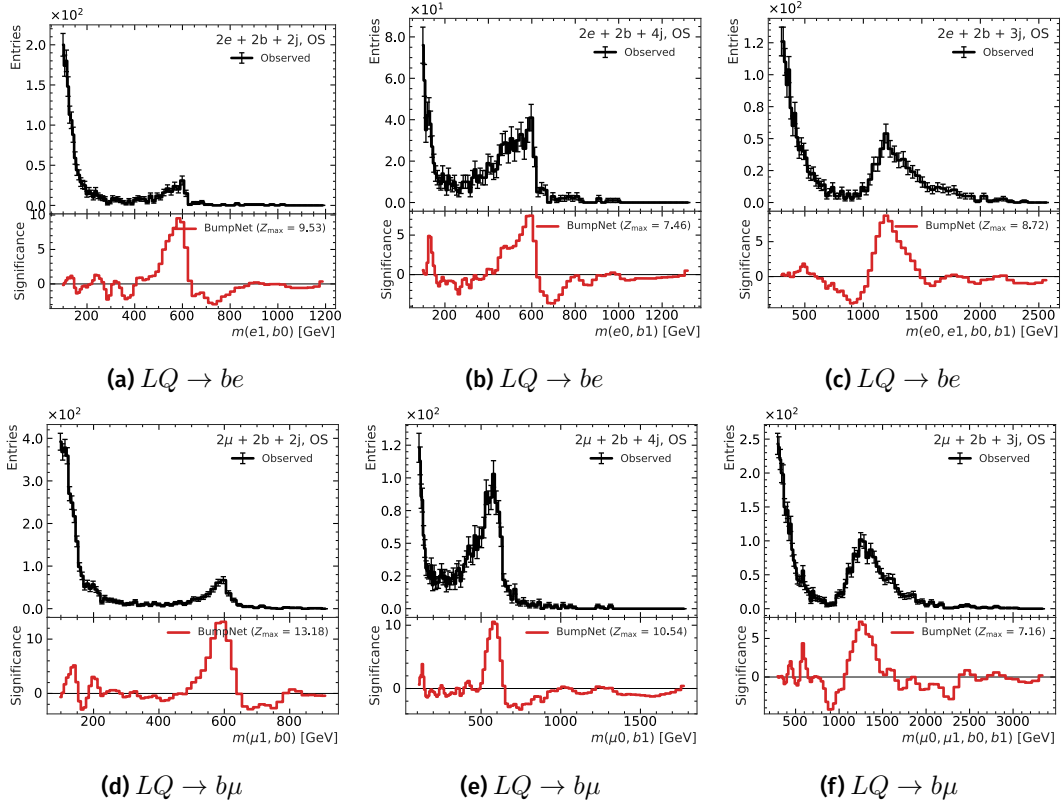


Figure 7.25: Examples of predictions over the Dark Machines datasets containing leptoquarks with mass of 600 GeV decaying via $LQ \rightarrow b\ell$, $\ell = e, \mu$ from channel 2b.

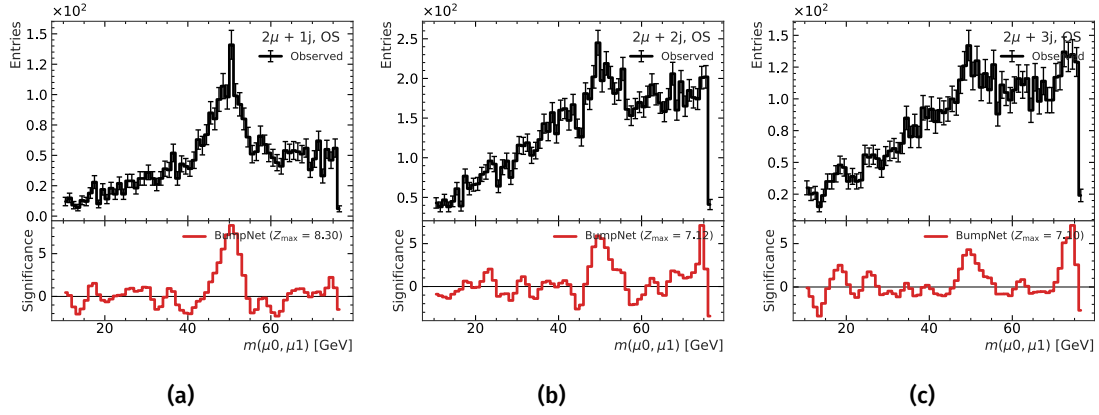


Figure 7.26: Examples of predictions over the Dark Machines datasets containing Z' with mass of 50 GeV and decaying via $Z' \rightarrow 2\mu$ from channel 2b.

7.9 Histogram production from ATLAS simulation data

Given the promising results of this technique when applied to the Dark Machines simulated data, the development of a BumpNet model to be applied over ATLAS experimental data started being implemented. To better reflect the real data on which the network will be applied, MC simulated events that account for the interaction with the ATLAS detector are used. As presented in Section 3.2.6, these datasets are produced using the ATLAS simulation infrastructure^[224] and, in most cases, Geant4^[253], covering the event-level generation up to the simulation of detector effects. These simulated events are reconstructed events using the same algorithms used for data, thus being available in the same format as the real data.

Hence, before applying the BumpNet network to real experimental data, simulation data can be used to implement and validate the histogram production chain. More importantly, it can be used to benchmark its performance and produce background fits to be included in the training data, which were shown to increase the chances that the network will perform well when evaluating over real data.

In this first implementation, we considered only events that pass the single-lepton triggers, that is, that pass a trigger which requires the presence of at least one lepton. Thus, simulated events of the processes with the highest cross-sections that produce at least one lepton are included, namely the productions of the W and Z bosons, $t\bar{t}$, single-top and diboson final states. Datasets for the whole Run-2 period, from 2015 to 2018, are considered, having been generated according to the respective data-taking conditions of the detector at the time. These include pp collisions at 13 TeV and are scaled to match the total amount of 139 fb^{-1} of data recorded for the period. Basic data-quality selections are performed over the events to guarantee good reconstruction of objects, together with the requirement of at least one electron or muon with $p_T \geq 25 \text{ GeV}$.

Similar to the processing and calibration procedure described in Section 7.5.2, an initial implementation of the construction of invariant mass histograms from ATLAS simulation was

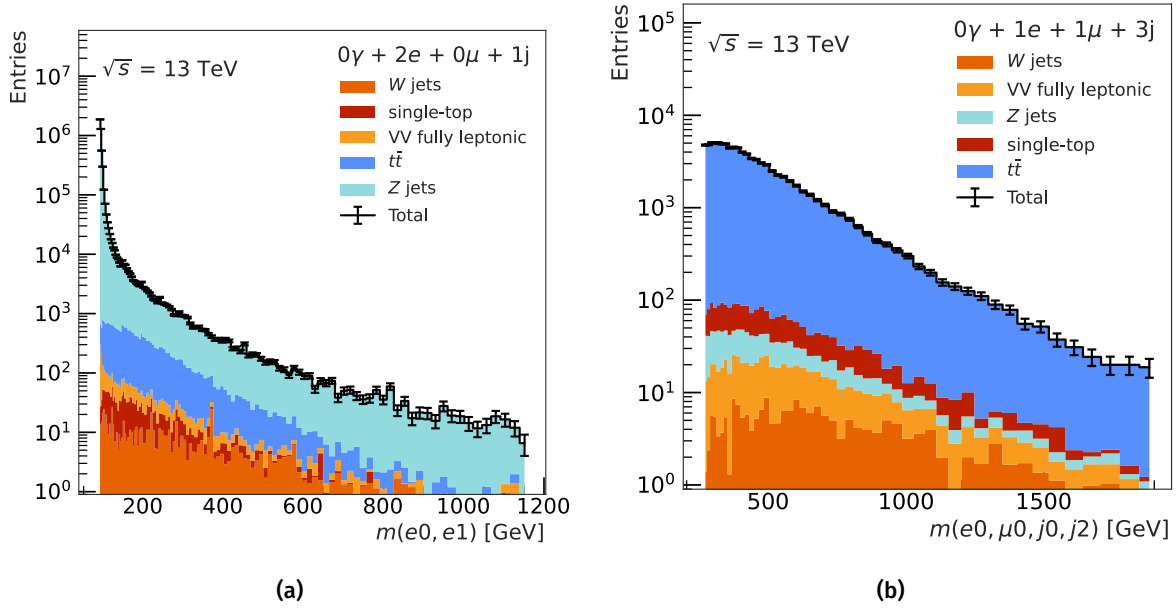


Figure 7.27: Preliminary invariant mass distributions obtained from the ATLAS simulation data. The different processes are scaled to match their cross-section and match the recorded luminosity.

developed. In this initial iteration, only electrons, muons, photons and jets were considered, such that no reconstructed objects or b -tagged jets were implemented. Histogram categories were constructed from all possible combinations of up to four of each object type, with exclusive jet bins being constructed from 1 to 8 jets and an inclusive bin with 9 or more jets. No cuts on kinematic variables were included. For the events corresponding to each one of these signatures, histograms were filled for the invariant mass of the different combinations of objects, as well as the invariant mass with the missing transverse energy E_T^{miss} .

All events corresponding to the different SM processes are scaled according to their corresponding cross-section and luminosity, and stacked together. The calibration of the binning of the histogram is performed using constant values for the objects' resolutions, namely 3% for electrons and photons, 2% for muons and 5% for jets. The bin widths values were determined from the total histogram combining all events, but the processes were kept separate to allow the visualisation of their relative contribution to the background in each scenario. Some of the produced histograms are shown in Fig. 7.27. In particular, Fig. 7.27a shows the invariant mass of the two most-energetic electrons in events containing exactly two electrons and one jet. Similarly, Fig. 7.27b shows the invariant mass distribution of the combination of one electron, one muon and two jets in events with exactly one electron, one muon and three jets. The different processes used to generate the events are indicated and combined to form the total histograms.

The work performed by the author of the thesis included the first implementation of this histogram production workflow over ATLAS simulation data. From this first iteration of the code, many developments are still needed to construct a reliable BumpNet model. Firstly,

the implementation of the reconstructed objects, such as boosted V_h and T , is still missing, as well as the tagging of b -jets. Moreover, the proper calibration of the histograms must be made using the correct ATLAS resolution of the objects. Lastly, given the large amount of data to be processed and the many combinations of objects and final states, optimisations are needed to make sure production is doable using the time and computing resources available. Once this histogram production is implemented, the next step is to train a BumpNet and quantify its performance before applying it to the experimental data.

7.10 Outlook

In this chapter, we presented the development of a neural network that predicts the position and statistical significance of bumps in invariant-mass distributions. Referred to as BumpNet, it is trained over synthetic data generated by injecting Gaussian signals of different strengths and positions into smoothly falling backgrounds obtained from analytical functions and fits to simulation data. We showed it has a very good performance over synthetic data, also being robust against many histogram features. Moreover, BumpNet was able to find fully simulated BSM particles injected over SM backgrounds, such as RPV stop, leptoquark and a Z' particles.

From these results, it is clear that this method has the potential to help identify regions of the phase space that have not yet been explored in the ATLAS experimental data. The predictions obtained by BumpNet when evaluating over histograms constructed from the real data can guide traditional analysis efforts, such that histograms that achieve a given threshold of significance predicted by the network could be flagged as interesting for further investigation. To that end, the implementation of the processing of ATLAS datasets into histograms has been started via the use of MC simulation datasets.

Moving forward, it is important to invest in the improvement of the binning calibration in terms of the objects resolution. The approximation of $p_T \approx m/2$ does not hold in all scenarios, hence enforcing the need to obtain resolution values that better reproduce what is observed in the real data. That said, the use of ATLAS simulation data can help assess how to improve this calibration.

In Section 7.7.1, we have shown how important the inclusion of the fits to simulation data is in making sure the network performs well in these shapes. The 4th-order log-polynomial function was good enough when dealing with the Dark Machines histograms, but this procedure could be further optimised for the ATLAS data. For instance, it seems clear that different statistics behave differently and thus different fit functions could be considered for different dynamic ranges. Moreover, methods such as kernel-density estimation (KDE) seems to be a good candidate to fit the wide range of shapes that is likely to result from the analysis workflow, since this method estimates the probability density function in a non-parametric way.

Furthermore, the BumpNet analysis could take advantage of the already existing simulation data from ATLAS. In particular, the many BSM signal samples available can be used to

test the performance of the network in realistic scenarios. Also, the sensitivity of the method could be gauged by testing it in already unblinded results and comparing them with the corresponding ATLAS analyses.

A crucial element of any application of this method would be the establishment of a general analysis procedure to interpret the network's prediction. As we have discussed, if a bump is found at the same mass values of the same object combinations in several exclusive final states, such as the different jet multiplicity histograms, we would have greater confidence that it is coming from a new physics signal. This procedure considerably reduces the LEE, since it is extremely unlikely that this occurs due to statistical fluctuations. Another approach to reduce the LEE would be to perform the analysis over just 50% of the data, such that any excesses that might be found on this first half would need to also appear after the processing of the whole dataset to represent convincing signs of new physics. Other physics-motivated considerations can further be explored, such as comparisons of histograms with leptons of different flavours, where we replace an electron with a muon, and vice-versa, in the final state.

Of course, before applying a BumpNet model over experimental data, it is important to make sure its predictions are reliable. After all, we need to establish that we are confident enough in the network such that, if a bump is found, it is worth committing a full analysis to investigate it. This can be achieved by increasing the threshold of $Z_{\max}$ values one wants to investigate or by finding ways to reduce the outliers rate. Either way, the final network's limitations must be properly quantified so that a positive outcome can be trusted. In this way, a full background estimation can be performed to obtain a precise estimation of the statistical significance of an excess. In contrast, if no bump is found, it would be interesting to try and set limits based on the analysis outcome. Either way, sharing the invariant mass distributions from the real data with the scientific community would already be a valuable outcome of such an analysis.

In summary, a BumpNet neural network has been developed to predict the statistical significance of bumps in invariant mass histograms. By achieving a very good performance over both synthetic data and realistic BSM scenarios, it has the potential of being an important ally in the expansion of the phase space explored by the ATLAS experiment. If applied to data, such a network could guide us through unknown realms into the future of particle physics.

Conclusion

This thesis presented contributions to the assessment and improvement of the performance of the electron identification algorithm, as well as the search for new particles, in the ATLAS experiment. On the detector front, measurements of the electron identification efficiencies were presented in light of the new data-taking period of Run 3. Furthermore, a new identification algorithm was presented, which incorporated low-level detector information in the form of images that are processed by a convolutional neural network. The development of techniques to search for new phenomena was also featured via the development of a neural network to detect anomalies in the invariant mass distributions obtained from the experimental data.

With the beginning of the data-taking period of Run 3, the performance of the current electron identification algorithm used in ATLAS was assessed via new measurements, as presented in Chapter 5. Results were obtained for the data recorded from 2022 in the beginning of Run 3, where collisions happened at a centre-of-mass energy of $\sqrt{s} = 13.6$ TeV. Measurements were also performed using the Run 2 data, where collisions happened at $\sqrt{s} = 13$ TeV, after the reprocessing of this data using the latest ATLAS software version. Efficiencies were obtained in both experimental data and MC simulated events, allowing the measurement of the Data/MC ratios. Referred to as scale factors, these ratios quantify the discrepancies between the simulation and the recorded data, such that values closer to unity represent a good agreement between the two.

In the measurement performed over Run 3 data, efficiencies were observed to be uniform in η , while they decrease with smaller values of p_T , features that are expected from previous measurements. The scale factors were found to be mostly close to unity, except in the very forward region, where $|\eta| > 2.37$, and for low energies, with $p_T < 20$ GeV. Given these results are within the expected values observed in previous years, they indicate a great performance of the detector at the start of the Run 3 data-taking period.

Moreover, the Run 2 results presented focused on the 2018 period, with similar patterns being found for the years 2015–2017. A comparison of the values obtained with the new ATLAS software when compared to the previous values of these measurements showed a general improvement in the Data/MC ratios across η , with values closer to unity overall. This can be understood from the improvements performed in the software and indicates a better agreement between the MC simulation and the data in this context. Although these measurements are not final, they correspond to an important milestone in the assessment of the electron identification performance in the Run 2 using the new software.

Next, a new algorithm for electron identification was presented in Chapter 6, where we described a neural network used to incorporate low-level features from the detector. The rejection of the electron background is improved when the energy deposits and additional tracks information are processed as images via CNNs. Moreover, studies of the importance of the different input variables indicated all the currently used features are relevant to discriminate the different backgrounds. In particular, the calorimeter images were shown to play an important role in discriminating the most common background, the light-flavour classes. Furthermore, a comparison of the performance of the network was done when rejecting these light-flavour backgrounds obtained from real data and those obtained from MC simulated events. A significant decline in the performance was observed for a model trained over MC simulation examples when tested over examples gathered from data. Even so, the presented algorithm still outperforms the current likelihood approach by factors of 2–5. This performance was shown to be recovered if the examples obtained from data are incorporated in the training, with the new algorithm improving the background rejection by factors of 3–10 with respect to the likelihood values.

Lastly, Chapter 7 was dedicated to a novel technique that can be used to search for resonances in the invariant mass distributions obtained from experimental data. The use of a neural network trained to predict the local significance of having a Gaussian resonance from the histogram bin entries was shown to effectively predict the presence of such resonances in invariant mass distributions. The generation of synthetic data using Gaussian signals over functional forms obtained from fits to simulation data allowed for the training of a model that was capable of predicting the maximum value of the significance in a histogram with no bias and a variance of 0.5–0.6. Moreover, the trained model was able to detect some selected BSM signals in simulation data, namely an RPV stop of 1 TeV, a leptoquark of 600 GeV and a Z' of 50 GeV. In light of these promising results, the application of this technique within the ATLAS experiment was started via the implementation of the production of invariant mass histograms using ATLAS simulation data.

As Run 3 reaches its end and the preparations for the the HL-LHC are in development, the ATLAS experiment has still plenty of unexplored data to be analysed. Machine learning techniques have shown a great potential to help address many of the current challenges of the field, namely the high-dimensionality of the data and the complexity of busy detector environments with many simultaneous collisions. To that end, the work presented in this thesis included contributions to the assessment of the electron identification algorithm performance, as well as its improvement via the application of deep learning techniques. It also included contributions to a new strategy in the search for resonances in invariant mass distributions, using neural networks to identify the regions of the data that have a larger potential to have signs of new particles. These contributions further advance the research performed by the ATLAS experiment, in the hope that its exploration of high-energy processes keeps enlightening our understanding of the more fundamental particles and interactions in nature.

Bibliography

- [1] ATLAS Collaboration, *Electron identification efficiency with Run 3 early data*, 2022, <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/EGAM-2022-04/>.
- [2] ATLAS Collaboration, *Electron identification with a convolutional neural network in the ATLAS experiment*, 2023, <https://cds.cern.ch/record/2850666>.
- [3] R. P. Feynman, R. B. Leighton and M. L. Sands, *The Feynman Lectures on Physics* (Basic Books, 2011).
- [4] S. L. Glashow, “Partial-symmetries of weak interactions”, *Nuclear Physics* **22**, 579 (1961) [10.1016/0029-5582\(61\)90469-2](#).
- [5] S. Weinberg, “A Model of Leptons”, *Physical Review Letters* **19**, 1264 (1967) [10.1103/PhysRevLett.19.1264](#).
- [6] A. Salam, “Weak and electromagnetic interactions”, *Conf. Proc. C* **680519**, 367 (1968) [10.1142/9789812795915_0034](#).
- [7] S. L. Glashow, J. Iliopoulos and L. Maiani, “Weak Interactions with Lepton-Hadron Symmetry”, *Physical Review D* **2**, 1285 (1970) [10.1103/PhysRevD.2.1285](#).
- [8] C. Bouchiat, J. Iliopoulos and Ph. Meyer, “An anomaly-free version of Weinberg’s model”, *Physics Letters B* **38**, 519 (1972) [10.1016/0370-2693\(72\)90532-1](#).
- [9] D. J. Gross, “Asymptotic Freedom”, *Physics Today* **40**, 39 (1987) [10.1063/1.881111](#).
- [10] D. J. Gross and F. Wilczek, “Ultraviolet Behavior of Non-Abelian Gauge Theories”, *Physical Review Letters* **30**, 1343 (1973) [10.1103/PhysRevLett.30.1343](#).
- [11] H. D. Politzer, “Reliable Perturbative Results for Strong Interactions?”, *Physical Review Letters* **30**, 1346 (1973) [10.1103/PhysRevLett.30.1346](#).
- [12] P. W. Anderson, “Plasmons, Gauge Invariance, and Mass”, *Physical Review* **130**, 439 (1963) [10.1103/PhysRev.130.439](#).
- [13] F. Englert and R. Brout, “Broken Symmetry and the Mass of Gauge Vector Mesons”, *Physical Review Letters* **13**, 321 (1964) [10.1103/PhysRevLett.13.321](#).
- [14] P. W. Higgs, “Broken Symmetries and the Masses of Gauge Bosons”, *Physical Review Letters* **13**, 508 (1964) [10.1103/PhysRevLett.13.508](#).

- [15] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble, “Global Conservation Laws and Massless Particles”, *Physical Review Letters* **13**, 585 (1964) [10.1103/PhysRevLett.13.585](#).
- [16] J. Erler and M. Schott, “Electroweak precision tests of the Standard Model after the discovery of the Higgs boson”, *Progress in Particle and Nuclear Physics* **106**, 68 (2019) [10.1016/j.pnpnp.2019.02.007](#), [arXiv:1902.05142 \[hep-ph\]](#).
- [17] T. Gehrmann and B. Malaescu, “Precision QCD Physics at the LHC”, *Annual Review of Nuclear and Particle Science* **72**, 233 (2022) [10.1146/annurev-nucl-101920-014923](#).
- [18] R. Brown, “XXVII. A brief account of microscopical observations made in the months of June, July and August 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies”, *The Philosophical Magazine* **4**, 161 (1828) [10.1080/14786442808674769](#).
- [19] A. Einstein, “Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen”, *Annalen der Physik* **322**, 549 (1905) [10.1002/andp.19053220806](#).
- [20] J. J. Thomson, “XL. Cathode Rays”, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **44**, 293 (1897) [10.1080/14786449708621070](#).
- [21] H. Geiger and E. Marsden, “On a diffuse reflection of the α -particles”, *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **82**, 495 (1909) [10.1098/rspa.1909.0054](#).
- [22] H. Geiger, “The scattering of α -particles by matter”, *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **83**, 492 (1910) [10.1098/rspa.1910.0038](#).
- [23] E. Rutherford, “LXXIX. The scattering of α and β particles by matter and the structure of the atom”, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **21**, 669 (1911) [10.1080/14786440508637080](#).
- [24] E. Rutherford, “Collision of α particles with light atoms. IV. An anomalous effect in nitrogen”, *Philosophical Magazine* **90**, 31 (1919) [10.1080/14786431003659230](#).
- [25] J. Chadwick, “The existence of a neutron”, *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **136**, 692 (1932) [10.1098/rspa.1932.0112](#).
- [26] M. Gell-Mann, “A schematic model of baryons and mesons”, *Physics Letters* **8**, 214 (1964) [10.1016/S0031-9163\(64\)92001-3](#).
- [27] G. Zweig, “An SU_3 model for strong interaction symmetry and its breaking”, [10.17181/CERN-TH-412](#) (2002) [10.17181/CERN-TH-412](#).

- [28] E. D. Bloom et al., “High-Energy Inelastic e–p Scattering at 6° and 10°”, *Physical Review Letters* **23**, 930 (1969) [10.1103/PhysRevLett.23.930](#).
- [29] M. Breidenbach et al., “Observed Behavior of Highly Inelastic Electron-Proton Scattering”, *Physical Review Letters* **23**, 935 (1969) [10.1103/PhysRevLett.23.935](#).
- [30] F. Hasert et al., “Search for elastic muon-neutrino electron scattering”, *Physics Letters B* **46**, 121 (1973) [10.1016/0370-2693\(73\)90494-2](#).
- [31] F. Hasert et al., “Observation of neutrino-like interactions without muon or electron in the gargamelle neutrino experiment”, *Physics Letters B* **46**, 138 (1973) [10.1016/0370-2693\(73\)90499-1](#).
- [32] G. Arnison et al., “Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s} = 540$ GeV”, *Physics Letters B* **122**, 103 (1983) [10.1016/0370-2693\(83\)91177-2](#).
- [33] UA2 Collaboration, “Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider”, *Physics Letters B* **122**, 476 (1983) [10.1016/0370-2693\(83\)91605-2](#).
- [34] D. A. Glaser, “Some Effects of Ionizing Radiation on the Formation of Bubbles in Liquids”, *Physical Review* **87**, 665 (1952) [10.1103/PhysRev.87.665](#).
- [35] J. Cockcroft and E. Walton, “Experiments with high velocity positive ions. — (I) Further developments in the method of obtaining high velocity positive ions”, *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **136**, 619 (1932) [10.1098/rspa.1932.0107](#).
- [36] J. Cockcroft and E. Walton, “Experiments with high velocity positive ions. II. — The disintegration of elements by high velocity protons”, *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **137**, 229 (1932) [10.1098/rspa.1932.0133](#).
- [37] E. O. Lawrence and M. S. Livingston, “The Production of High Speed Light Ions Without the Use of High Voltages”, *Physical Review* **40**, 19 (1932) [10.1103/PhysRev.40.19](#).
- [38] V. Shiltsev and F. Zimmermann, “Modern and future colliders”, *Reviews of Modern Physics* **93**, 015006 (2021) [10.1103/RevModPhys.93.015006](#).
- [39] M. Thomson, *Modern Particle Physics* (Cambridge University Press, 2013).
- [40] LHCb Collaboration, “Observation of structure in the J/ψ -pair mass spectrum”, *Science Bulletin* **65**, 1983 (2020) [10.1016/j.scib.2020.08.032](#), [arXiv:2006.16957 \[hep-ex\]](#).

- [41] LHCb Collaboration, “Observation of New Resonances Decaying to $J/\psi K^+$ and $J/\psi \phi$ ”, *Physical Review Letters* **127**, 082001 (2021) [10.1103/PhysRevLett.127.082001](#), [arXiv:2103.01803 \[hep-ex\]](#).
- [42] LHCb Collaboration, “Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \rightarrow J/\psi k^- p$ Decays”, *Physical Review Letters* **115**, 072001 (2015) [10.1103/PhysRevLett.115.072001](#), [arXiv:1507.03414 \[hep-ex\]](#).
- [43] Particle Data Group, “Review of Particle Physics”, *Progress of Theoretical and Experimental Physics* **2022**, 083C01 (2022) [10.1093/ptep/ptac097](#).
- [44] C. M. G. Lattes et al., “Processes involving charged mesons”, *Nature* **159**, 694 (1947) [10.1038/159694a0](#).
- [45] S. Bailey et al., “Parton distributions from LHC, HERA, Tevatron and fixed target data: MSHT20 PDFs”, *The European Physical Journal C* **81**, 341 (2021) [10.1140/epjc/s10052-021-09057-0](#), [arXiv:2012.04684 \[hep-ph\]](#).
- [46] R. J. Gaitskell, “Direct detection of dark matter”, *Annual Review of Nuclear and Particle Science* **54**, 315 (2004) [10.1146/annurev.nucl.54.070103.181244](#).
- [47] A. Boveia and C. Doglioni, “Dark Matter Searches at Colliders”, *Annual Review of Nuclear and Particle Science* **68**, 429 (2018) [10.1146/annurev-nucl-101917-021008](#).
- [48] G. Bertone and T. M. P. Tait, “A new era in the search for dark matter”, *Nature* **562**, 51 (2018) [10.1038/s41586-018-0542-z](#).
- [49] K. Abe et al., “The T2K experiment”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **659**, 106 (2011) [10.1016/j.nima.2011.06.067](#).
- [50] S. Fukuda et al., “The Super-Kamiokande detector”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **501**, 418 (2003) [10.1016/S0168-9002\(03\)00425-X](#).
- [51] N. Jelley, A. B. McDonald and R. H. Robertson, “The Sudbury Neutrino Observatory”, *Annual Review of Nuclear and Particle Science* **59**, 431 (2009) [10.1146/annurev.nucl.55.090704.151550](#).
- [52] R. Acciarri et al., “Long-Baseline Neutrino Facility (LBNF) and Deep Underground Neutrino Experiment (DUNE) Conceptual Design Report Volume 1: The LBNF and DUNE Projects”, 20th Jan. 2016, [arXiv:1601.05471 \[hep-ex, physics:physics\]](#).
- [53] S. Andringa et al., “Current Status and Future Prospects of the SNO+ Experiment”, *Advances in High Energy Physics* **2016**, 1 (2016) [10.1155/2016/6194250](#).
- [54] A. Bellerive, “Review of solar neutrino experiments”, *International Journal of Modern Physics A* **19**, 1167 (2004) [10.1142/S0217751X04019093](#).

- [55] C. Bemporad, G. Gratta and P. Vogel, “Reactor-based neutrino oscillation experiments”, *Reviews of Modern Physics* **74**, 297 (2002) [10.1103/RevModPhys.74.297](#).
- [56] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory* (Addison-Wesley Pub. Co, 1995).
- [57] E. Noether, “Invariante Variationsprobleme”, *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* **1918**, 235 (1918), [arXiv:physics/0503066](#).
- [58] C. G. Tully, *Elementary Particle Physics in a Nutshell* (Princeton University Press, 2011).
- [59] R. P. Feynman, “The Theory of Positrons”, *Physical Review* **76**, 749 (1949) [10.1103/PhysRev.76.749](#).
- [60] R. P. Feynman, “Space-Time Approach to Quantum Electrodynamics”, *Physical Review* **76**, 769 (1949) [10.1103/PhysRev.76.769](#).
- [61] R. P. Feynman, “Mathematical Formulation of the Quantum Theory of Electromagnetic Interaction”, *Physical Review* **80**, 440 (1950) [10.1103/PhysRev.80.440](#).
- [62] S. Braibant, G. Giacomelli and M. Spurio, *Particles and Fundamental Interactions: an Introduction to Particle Physics*, Undergraduate lecture notes in physics (Springer, 2012).
- [63] A. Zee, *Quantum Field Theory in a Nutshell* (Princeton University Press, 2003).
- [64] V. L. Ginzburg and L. D. Landau, “On the Theory of Superconductivity”, in *On Superconductivity and Superfluidity* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2009), pp. 113–137, [10.1007/978-3-540-68008-6_4](#).
- [65] L. N. Cooper, “Bound Electron Pairs in a Degenerate Fermi Gas”, *Physical Review* **104**, 1189 (1956) [10.1103/PhysRev.104.1189](#).
- [66] J. Bardeen, L. N. Cooper and J. R. Schrieffer, “Theory of Superconductivity”, *Physical Review* **108**, 1175 (1957) [10.1103/PhysRev.108.1175](#).
- [67] J. Ellis, “Higgs physics”, 2015, [10.5170/CERN-2015-004.117](#), [arXiv:1312.5672 \[hep-ph\]](#).
- [68] J. Goldstone, “Field theories with « Superconductor » solutions”, *Il Nuovo Cimento* **19**, 154 (1961) [10.1007/BF02812722](#).
- [69] ATLAS Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”, *Physics Letters B* **716**, 1 (2012) [10.1016/j.physletb.2012.08.020](#), [arXiv:1207.7214 \[hep-ex\]](#).
- [70] CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”, *Physics Letters B* **716**, 30 (2012) [10.1016/j.physletb.2012.08.021](#), [arXiv:1207.7235 \[hep-ex\]](#).

- [71] N. Cabibbo, “Unitary Symmetry and Leptonic Decays”, *Physical Review Letters* **10**, 531 (1963) 10.1103/PhysRevLett.10.531.
- [72] M. Kobayashi and T. Maskawa, “CP Violation in the Renormalizable Theory of Weak Interaction”, *Progress of Theoretical Physics* **49**, 652 (1973) 10.1143/PTP.49.652.
- [73] A. Höcker and Z. Ligeti, “CP Violation and the CKM Matrix”, *Annual Review of Nuclear and Particle Science* **56**, 501 (2006) 10.1146/annurev.nucl.56.080805.140456.
- [74] ATLAS Collaboration, *Search for resonances with boson-tagged jets in 15.5 fb⁻¹ of pp collisions at $\sqrt{s} = 13$ TeV collected with the ATLAS detector*, ATLAS-CONF-2016-055 (CERN, 2016), <https://cds.cern.ch/record/2206137>.
- [75] ATLAS Collaboration, “Measurement of inclusive jet and dijet cross-sections in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”, *Journal of High Energy Physics* **2018**, 195 (2018) 10.1007/JHEP05(2018)195, arXiv:1711.02692 [hep-ex].
- [76] CMS Collaboration, “Measurement of the triple-differential dijet cross section in proton-proton collisions at $\sqrt{s} = 8$ TeV and constraints on parton distribution functions”, *The European Physical Journal C* **77**, 746 (2017) 10.1140/epjc/s10052-017-5286-7.
- [77] ATLAS Collaboration (ATLAS), *Standard model summary plots june 2024*, ATL-PHYS-PUB-2024-011 (CERN, Geneva, 2024), <https://cds.cern.ch/record/2903866>.
- [78] R. N. Cahn, “The eighteen arbitrary parameters of the standard model in your everyday life”, *Reviews of Modern Physics* **68**, 951 (1996) 10.1103/RevModPhys.68.951.
- [79] S. Weinberg, “Implications of dynamical symmetry breaking: An addendum”, *Physical Review D* **19**, 1277 (1979) 10.1103/PhysRevD.19.1277.
- [80] E. Gildener, “Gauge-symmetry hierarchies”, *Physical Review D* **14**, 1667 (1976) 10.1103/PhysRevD.14.1667.
- [81] L. Susskind, “Dynamics of spontaneous symmetry breaking in the Weinberg-Salam theory”, *Physical Review D* **20**, 2619 (1979) 10.1103/PhysRevD.20.2619.
- [82] J. P. Ellis, “TikZ-Feynman: Feynman diagrams with TikZ”, *Computer Physics Communications* **210**, 103 (2017) 10.1016/j.cpc.2016.08.019, arXiv:1601.05437 [hep-ph].
- [83] P. Langacker, *The Standard Model and Beyond*, 2nd ed., Series in high energy physics, cosmology and gravitation (CRC Press, Taylor & Francis Group, 2017).
- [84] J. L. Feng, “Naturalness and the Status of Supersymmetry”, *Annual Review of Nuclear and Particle Science* **63**, 351 (2013) 10.1146/annurev-nucl-102010-130447.
- [85] Y. Fukuda et al., “Evidence for Oscillation of Atmospheric Neutrinos”, *Physical Review Letters* **81**, 1562 (1998) 10.1103/PhysRevLett.81.1562.

- [86] M. C. Gonzalez-Garcia and Y. Nir, “Neutrino masses and mixing: evidence and implications”, *Reviews of Modern Physics* **75**, 345 (2003) [10.1103/RevModPhys.75.345](#), [arXiv:hep-ph/0202058](#).
- [87] M. Gonzalez-Garcia and M. Maltoni, “Phenomenology with massive neutrinos”, *Physics Reports* **460**, 1 (2008) [10.1016/j.physrep.2007.12.004](#).
- [88] V. G. Lüth, “Semileptonic B Meson Decays”, *Annual Review of Nuclear and Particle Science* **61**, 119 (2011) [10.1146/annurev.nucl.012809.104421](#).
- [89] J. Dingfelder and T. Mannel, “Leptonic and semileptonic decays of B mesons”, *Reviews of Modern Physics* **88**, 035008 (2016) [10.1103/RevModPhys.88.035008](#).
- [90] D. London, “Anomalies in B Decays: A Sign of New Physics?”, 2019, [arXiv:1911.06238 \[hep-ph\]](#).
- [91] LHCb Collaboration, “Angular analysis of the $B^+ \rightarrow K^* \mu^+ \mu^-$ decay”, *Physical Review Letters* **126**, 161802 (2021) [10.1103/PhysRevLett.126.161802](#).
- [92] M. Algueró et al., “To (b)e or not to (b)e: no electrons at LHCb”, *The European Physical Journal C* **83**, 648 (2023) [10.1140/epjc/s10052-023-11824-0](#), [arXiv:2304.07330 \[hep-ph\]](#).
- [93] B. Capdevila, A. Crivellin and J. Matias, “Review of semileptonic B anomalies”, *The European Physical Journal Special Topics*, [10.1140/epjs/s11734-023-01012-2](#) (2023) [10.1140/epjs/s11734-023-01012-2](#), [arXiv:2309.01311 \[hep-ph\]](#).
- [94] J. P. Miller, E. de Rafael and B. Lee Roberts, “Muon ($g - 2$): experiment and theory”, *Reports on Progress in Physics* **70**, 795 (2007) [10.1088/0034-4885/70/5/R03](#).
- [95] F. Jegerlehner and A. Nyffeler, “The Muon $g - 2$ ”, *Physics Reports* **477**, 1 (2009) [10.1016/j.physrep.2009.04.003](#), [arXiv:0902.3360 \[hep-ph\]](#).
- [96] J. P. Miller et al., “Muon ($g - 2$): Experiment and Theory”, *Annual Review of Nuclear and Particle Science* **62**, 237 (2012) [10.1146/annurev-nucl-031312-120340](#).
- [97] ETM Collaboration, “Strange and charm HVP contributions to the muon ($g - 2$) including QED corrections with twisted-mass fermions”, *Journal of High Energy Physics* **2017**, 157 (2017) [10.1007/JHEP10\(2017\)157](#).
- [98] Y. Amhis et al., “Averages of b -hadron, c -hadron, and τ -lepton properties as of 2018: Heavy Flavor Averaging Group (HFLAV)”, *The European Physical Journal C* **81**, 226 (2021) [10.1140/epjc/s10052-020-8156-7](#), [arXiv:1909.12524 \[hep-ex\]](#).
- [99] M. Ishak, “Testing general relativity in cosmology”, *Living Reviews in Relativity* **22**, 1 (2019) [10.1007/s41114-018-0017-4](#).

- [100] ATLAS Collaboration, “Search for new phenomena in events with an energetic jet and missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”, *Physical Review D* **103**, 112006 (2021) [10.1103/PhysRevD.103.112006](#), [arXiv:2102.10874 \[hep-ex\]](#).
- [101] CMS Collaboration, “Search for new particles in events with energetic jets and large missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *Journal of High Energy Physics* **2021**, 153 (2021) [10.1007/JHEP11\(2021\)153](#), [arXiv:2107.13021 \[hep-ex\]](#).
- [102] ATLAS Collaboration, “Search for new resonances in mass distributions of jet pairs using 139 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”, *Journal of High Energy Physics* **2020**, 145 (2020) [10.1007/JHEP03\(2020\)145](#), [arXiv:1910.08447 \[hep-ex\]](#).
- [103] CMS Collaboration, “Search for high mass dijet resonances with a new background prediction method in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *Journal of High Energy Physics* **2020**, 33 (2020) [10.1007/JHEP05\(2020\)033](#), [arXiv:1911.03947 \[hep-ex\]](#).
- [104] G. Bertone, D. Hooper and J. Silk, “Particle dark matter: evidence, candidates and constraints”, *Physics Reports* **405**, 279 (2005) [10.1016/j.physrep.2004.08.031](#), [arXiv:hep-ph/0404175](#).
- [105] J. G. de Swart, G. Bertone and J. van Dongen, “How dark matter came to matter”, *Nature Astronomy* **1**, 0059 (2017) [10.1038/s41550-017-0059](#), [arXiv:1703.00013 \[astro-ph.CO\]](#).
- [106] G. Bertone and D. Hooper, “History of dark matter”, *Reviews of Modern Physics* **90**, 045002 (2018) [10.1103/RevModPhys.90.045002](#), [arXiv:1605.04909 \[astro-ph.CO\]](#).
- [107] Planck Collaboration, “Planck 2013 results. I. Overview of products and scientific results”, *Astronomy & Astrophysics* **571**, A1 (2014) [10.1051/0004-6361/201321529](#), [arXiv:1303.5062 \[astro-ph.CO\]](#).
- [108] Y. Golfand and E. Likhtman, “Extension of the algebra of poincare group generators and violation of P invariance”, *JETP Letters* [translation of Pisma v Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki] **13**, 323 (1971).
- [109] D. Volkov and V. Akulov, “Is the neutrino a goldstone particle?”, *Physics Letters B* **46**, 109 (1973) [10.1016/0370-2693\(73\)90490-5](#).
- [110] J. Wess and B. Zumino, “Supergauge transformations in four dimensions”, *Nuclear Physics B* **70**, 39 (1974) [10.1016/0550-3213\(74\)90355-1](#).
- [111] J. Wess and B. Zumino, “Supergauge invariant extension of quantum electrodynamics”, *Nuclear Physics B* **78**, 1 (1974) [10.1016/0550-3213\(74\)90112-6](#).

- [112] S. Ferrara and B. Zumino, “Supergauge invariant Yang-Mills theories”, *Nuclear Physics B* **79**, 413 (1974) [10.1016/0550-3213\(74\)90559-8](#).
- [113] A. Salam and J. Strathdee, “Super-symmetry and non-Abelian gauges”, *Physics Letters B* **51**, 353 (1974) [10.1016/0370-2693\(74\)90226-3](#).
- [114] S. P. Martin, “A supersymmetry primer”, in G. L. Kane, *Advanced Series on Directions in High Energy Physics*, Vol. 18 (World Scientific, July 1998), pp. 1–98, [10.1142/9789812839657_0001](#).
- [115] P. Fayet, “Supersymmetry and weak, electromagnetic and strong interactions”, *Physics Letters B* **64**, 159 (1976) [10.1016/0370-2693\(76\)90319-1](#).
- [116] P. Fayet, “Spontaneously broken supersymmetric theories of weak, electromagnetic and strong interactions”, *Physics Letters B* **69**, 489 (1977) [10.1016/0370-2693\(77\)90852-8](#).
- [117] H. Nilles, “Supersymmetry, supergravity and particle physics”, *Physics Reports* **110**, 1 (1984) [10.1016/0370-1573\(84\)90008-5](#).
- [118] H. Haber, “The search for supersymmetry: Probing physics beyond the standard model”, *Physics Reports* **117**, 75 (1985) [10.1016/0370-1573\(85\)90051-1](#).
- [119] E. Witten, “Dynamical breaking of supersymmetry”, *Nuclear Physics B* **188**, 513 (1981) [10.1016/0550-3213\(81\)90006-7](#).
- [120] S. Dimopoulos and H. Georgi, “Softly broken supersymmetry and SU(5)”, *Nuclear Physics B* **193**, 150 (1981) [10.1016/0550-3213\(81\)90522-8](#).
- [121] N. Sakai, “Naturalnes in supersymmetric GUTS”, *Zeitschrift für Physik C Particles and Fields* **11**, 153 (1981) [10.1007/BF01573998](#).
- [122] L. Susskind, “The gauge hierarchy problem, technicolor, supersymmetry, and all that”, *Physics Reports* **104**, 181 (1984) [10.1016/0370-1573\(84\)90208-4](#).
- [123] B. Fuks, “Beyond the minimal supersymmetric standard model: from theory to phenomenology”, *International Journal of Modern Physics A* **27**, 1230007 (2012) [10.1142/S0217751X12300074](#).
- [124] F. D. Steffen, “Dark-matter candidates: Axions, neutralinos, gravitinos, and axinos”, *The European Physical Journal C* **59**, 557 (2009) [10.1140/epjc/s10052-008-0830-0](#), [arXiv:0811.3347 \[hep-ph\]](#).
- [125] D. Hooper, *TASI 2008 Lectures on Dark Matter*, (2009) pre-published.
- [126] H. Goldberg, “Constraint on the photino mass from cosmology”, *Physical Review Letters* **50**, 1419 (1983) [10.1103/PhysRevLett.50.1419](#); Erratum: “Erratum: Constraint on the Photino Mass from Cosmology [Phys. Rev. Lett. 50, 1419 (1983)]”, *Physical Review Letters* **103**, 099905 (2009) [10.1103/PhysRevLett.103.099905](#).

- [127] J. Ellis et al., “Supersymmetric relics from the big bang”, in *IN *BATAVIA 1984, PROCEEDINGS, INNER SPACE/OUTER SPACE**, 458-459, in *srednicki, M.A. (ed.): Particle physics and cosmology* 223-246, Vol. 238, Nuclear Physics B SLAC-PUB-3171 (1984), p. 453, [10.1016/0550-3213\(84\)90461-9](#).
- [128] G. R. Farrar and P. Fayet, “Phenomenology of the production, decay, and detection of new hadronic states associated with supersymmetry”, *Physics Letters B* **76**, 575 (1978) [10.1016/0370-2693\(78\)90858-4](#).
- [129] R. Barbier et al., “R-Parity-violating supersymmetry”, *Physics Reports* **420**, 1 (2005) [10.1016/j.physrep.2005.08.006](#), [arXiv:hep-ph/0406039](#).
- [130] K. Inoue et al., “Aspects of grand unified models with softly broken supersymmetry”, *Progress of Theoretical Physics* **68**, 927 (1982) [10.1143/PTP.68.927](#).
- [131] J. R. Ellis and S. Rudaz, “Search for supersymmetry in toponium decays”, *Physics Letters B* **128**, 248 (1983) [10.1016/0370-2693\(83\)90402-1](#).
- [132] C. Boehm, A. Djouadi and Y. Mambrini, “Decays of the lightest top squark”, *Physical Review D* **61**, 095006 (2000) [10.1103/PhysRevD.61.095006](#), [arXiv:hep-ph/9907428](#).
- [133] ATLAS Collaboration, “A search for R -parity violating supersymmetric decays of the top squark to a b -jet and a lepton in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector”, *Phys. Rev. D* **110**, 092004 (2024) [10.1103/PhysRevD.110.092004](#), [arXiv:2406.18367 \[hep-ex\]](#).
- [134] I. Doršner and A. Greljo, “Leptoquark toolbox for precision collider studies”, *Journal of High Energy Physics* **2018**, 126 (2018) [10.1007/JHEP05\(2018\)126](#).
- [135] H. Päs and E. Schumacher, “Common origin of R_K and neutrino masses”, *Physical Review D* **92**, 114025 (2015) [10.1103/PhysRevD.92.114025](#), [arXiv:1510.08757 \[hep-ph\]](#).
- [136] I. Doršner, S. Fajfer and N. Košnik, “Leptoquark mechanism of neutrino masses within the grand unification framework”, *The European Physical Journal C* **77**, 417 (2017) [10.1140/epjc/s10052-017-4987-2](#), [arXiv:1701.08322 \[hep-ph\]](#).
- [137] O. Popov and G. White, “One leptoquark to unify them? Neutrino masses and unification in the light of $(g - 2)$, μ , $R_D^{(*)}$ and R_K anomalies”, *Nuclear Physics B* **923**, 324 (2017) [10.1016/j.nuclphysb.2017.08.007](#), [arXiv:1611.04566 \[hep-ph\]](#).
- [138] K. Babu and J. Julio, “Two-loop neutrino mass generation through leptoquarks”, *Nuclear Physics B* **841**, 130 (2010) [10.1016/j.nuclphysb.2010.07.022](#), [arXiv:1006.1092 \[hep-ph\]](#).
- [139] V. Barger and K. Cheung, “Atomic parity violation, leptoquarks, and contact interactions”, *Physics Letters B* **480**, 149 (2000) [10.1016/S0370-2693\(00\)00401-9](#), [arXiv:hep-ph/0002259](#).

- [140] P. Herczeg, “CP-violating electron-nucleon interactions from leptoquark exchange”, *Physical Review D* **68**, 116004 (2003) [10.1103/PhysRevD.68.116004](#).
- [141] ATLAS Collaboration, “Search for pairs of scalar leptoquarks decaying into quarks and electrons or muons in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector”, *Journal of High Energy Physics* **2020**, 112 (2020) [10.1007/JHEP10\(2020\)112](#), [arXiv:2006.05872 \[hep-ex\]](#).
- [142] ATLAS Collaboration, “Combination of searches for pair-produced leptoquarks at $\sqrt{s} = 13$ TeV with the ATLAS detector”, *Physics Letters B* **854**, 138736 (2024) [10.1016/j.physletb.2024.138736](#), [arXiv:2401.11928v2 \[hep-ex\]](#).
- [143] CMS Collaboration, “Search for pair production of scalar and vector leptoquarks decaying to muons and bottom quarks in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *Physical Review D* **109**, 112003 (2024) [10.1103/PhysRevD.109.112003](#), [arXiv:2402.08668 \[hep-ex\]](#).
- [144] X.-G. He et al., “New- Z' phenomenology”, *Physical Review D* **43**, R22 (1991) [10.1103/PhysRevD.43.R22](#).
- [145] X.-G. He et al., “Simplest Z' model”, *Physical Review D* **44**, 2118 (1991) [10.1103/PhysRevD.44.2118](#).
- [146] ATLAS Collaboration, “Search for a new Z' gauge boson in 4μ events with the ATLAS experiment”, *Journal of High Energy Physics* **2023**, 90 (2023) [10.1007/JHEP07\(2023\)090](#), [arXiv:2301.09342 \[hep-ex\]](#).
- [147] CMS Collaboration, “Search for an $L_\mu - L_\tau$ gauge boson using $Z \rightarrow 4\mu$ events in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *Physics Letters B* **792**, 345 (2019) [10.1016/j.physletb.2019.01.072](#), [arXiv:1808.03684 \[hep-ex\]](#).
- [148] ATLAS Collaboration, *Exploration at the high-energy frontier: ATLAS Run 2 searches investigating the exotic jungle beyond the Standard Model*, (Mar. 2024) [arXiv:2403.09292 \[hep-ex\]](#), pre-published.
- [149] G. Box, “Robustness in the Strategy of Scientific Model Building”, in *Robustness in Statistics* (Elsevier, 1979), pp. 201–236, [10.1016/B978-0-12-438150-6.50018-2](#).
- [150] G. E. P. Box, “Science and Statistics”, *Journal of the American Statistical Association* **71**, 791 (1976) [10.1080/01621459.1976.10480949](#).
- [151] G. Cowan, *Statistical data analysis*, Oxford science publications (Clarendon Press, 1998).
- [152] G. Cowan et al., “Asymptotic formulae for likelihood-based tests of new physics”, *The European Physical Journal C* **71**, 1554 (2011) [10.1140/epjc/s10052-011-1554-0](#), [arXiv:1007.1727 \[hep-ex, physics:physics\]](#), Erratum: [10.1140/epjc/s10052-013-2501-z](#), *Eur. Phys. J. C*, **73**, (2013) 2501.

- [153] M. Feickert and B. Nachman, “A Living Review of Machine Learning for Particle Physics”, Feb. 2021, [arXiv:2102.02770 \[hep-ex, physics:hep-ph, physics:physics, stat\]](#).
- [154] G. Karagiorgi et al., “Machine learning in the search for new fundamental physics”, *Nature Reviews Physics* **4**, 399 (2022) [10.1038/s42254-022-00455-1](#), [arXiv:2112.03769 \[hep-ph\]](#).
- [155] D. Guest, K. Cranmer and D. Whiteson, “Deep Learning and Its Application to LHC Physics”, *Annual Review of Nuclear and Particle Science* **68**, 161 (2018) [10.1146/annurev-nucl-101917-021019](#), [arXiv:1806.11484 \[hep-ex\]](#).
- [156] G. Carleo et al., “Machine learning and the physical sciences”, *Reviews of Modern Physics* **91**, 045002 (2019) [10.1103/RevModPhys.91.045002](#), [arXiv:1903.10563 \[physics.comp-ph\]](#).
- [157] Kolmogorov, *Foundations of the theory of probability* (Chelsea Publishing Company, 1950).
- [158] Bayes, “LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S”, *Philosophical Transactions of the Royal Society of London* **53**, 370 (1763) [10.1098/rstl.1763.0053](#).
- [159] M. H. DeGroot and M. J. Schervish, *Probability and statistics*, 4th ed (Addison-Wesley, 2012).
- [160] W. A. Rolke, A. M. Lopez and J. Conrad, “Limits and confidence intervals in the presence of nuisance parameters”, *Nucl. Instrum. Meth. A* **551**, edited by L. Lyons and M. Karagoz, 493 (2005) [10.1016/j.nima.2005.05.068](#), [arXiv:physics/0403059](#).
- [161] J. Neyman and E. S. Pearson, “IX. On the problem of the most efficient tests of statistical hypotheses”, *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* **231**, 289 (1933) [10.1098/rsta.1933.0009](#).
- [162] W. W. T. G. Peterson, T. Birdsall and W. Fox, “The theory of signal detectability”, *Transactions of the IRE professional group on information theory* **4**, 171 (1954).
- [163] J. Marcum, “A statistical theory of target detection by pulsed radar”, *IRE Transactions on Information Theory* **6**, 59 (1960).
- [164] A. L. Samuel, “Some studies in machine learning using the game of Checkers”, *IBM Journal of Research and Development*, 71 (1959).
- [165] A. Lovelace, “Notes upon L. F. Menabrea’s “Sketch of the Analytical Engine invented by Charles Babbage””, (1842).

- [166] K. P. Murphy, *Machine learning: a probabilistic perspective*, Adaptive computation and machine learning series (MIT Press, 2012).
- [167] Ian Goodfellow, Yoshua Bengio and Aaron Courville, *Deep Learning* (MIT Press, 2016).
- [168] P. Mehta et al., “A high-bias, low-variance introduction to Machine Learning for physicists”, *Physics Reports* **810**, 1 (2019) [10.1016/j.physrep.2019.03.001](#), [arXiv:1803.08823](#).
- [169] R. S. Sutton and A. G. Barto, *Reinforcement learning: an introduction* (MIT Press, 1998).
- [170] C. M. Bishop, *Pattern recognition and machine learning*, Information science and statistics (Springer, 2006).
- [171] D. P. Kingma and J. Ba, “Adam: a method for stochastic optimization”, in (Dec. 2014), [arXiv:1412.6980 \[cs.LG\]](#).
- [172] C. Cortes, M. Mohri and A. Rostamizadeh, *L2 Regularization for Learning Kernels*, (May 2012) [arXiv:1205.2653 \[cs.LG\]](#), pre-published.
- [173] N. Srivastava et al., “Dropout: a simple way to prevent neural networks from overfitting”, *Journal of Machine Learning Research* **15**, 1929 (2014).
- [174] R. E. Bellman, *Adaptive control processes: A Guided Tour* (Princeton University Press, 1961).
- [175] D. H. Wolpert, “The Lack of A Priori Distinctions Between Learning Algorithms”, *Neural Computation* **8**, 1341 (1996) [10.1162/neco.1996.8.7.1341](#).
- [176] F. Rosenblatt, *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms* (Spartan, 1962).
- [177] M. Minsky and S. A. Papert, *Perceptrons: an introduction to computational geometry* (MIT Press, 1972).
- [178] K. Hornik, “Approximation capabilities of multilayer feedforward networks”, *Neural Networks* **4**, 251 (1991) [10.1016/0893-6080\(91\)90009-T](#).
- [179] D. E. Rumelhart, G. E. Hinton and R. J. Williams, “Learning representations by back-propagating errors”, *Nature* **323**, 533 (1986) [10.1038/323533a0](#).
- [180] D. H. Hubel and T. N. Wiesel, “Receptive fields and functional architecture of monkey striate cortex”, *The Journal of Physiology* **195**, 215 (1968) [10.1113/jphysiol.1968.sp008455](#).
- [181] Y. LeCun et al., “Backpropagation applied to handwritten zip code recognition”, *Neural Computation* **1**, 541 (1989).
- [182] Y. LeCun et al., “Handwritten digit recognition with a back-propagation network”, in *Advances in neural information processing systems* (NIPS 1989), Vol. 2, edited by D. Touretzky (1990).

- [183] A. Krizhevsky, I. Sutskever and G. E. Hinton, “ImageNet classification with deep convolutional neural networks”, *Communications of the ACM* **60**, 84 (2017) [10.1145/3065386](#).
- [184] O. Russakovsky et al., “ImageNet Large Scale Visual Recognition Challenge”, *International Journal of Computer Vision* **115**, 211 (2015) [10.1007/s11263-015-0816-y](#).
- [185] J. Gu et al., “Recent Advances in Convolutional Neural Networks”, Oct. 2017, [arXiv:1512.07108 \[cs\]](#).
- [186] A. Aurisano et al., “A convolutional neural network neutrino event classifier”, *Journal of Instrumentation* **11**, P09001 (2016) [10.1088/1748-0221/11/09/P09001](#).
- [187] M. Kagan, “Image-Based Jet Analysis”, Dec. 2020, [arXiv:2012.09719 \[hep-ex, physics:hep-ph, physics:physics\]](#).
- [188] B. Nachman, “A guide for deploying Deep Learning in LHC searches: How to achieve optimality and account for uncertainty”, *SciPost Physics* **8**, 090 (2020) [10.21468/SciPostPhys.8.6.090](#), [arXiv:1909.03081 \[hep-ph\]](#).
- [189] A. Hoecker et al., “TMVA - Toolkit for Multivariate Data Analysis”, July 2009, [arXiv:physics/0703039](#).
- [190] J. Duarte and J.-R. Vlimant, “Graph Neural Networks for Particle Tracking and Reconstruction”, Dec. 2020, [arXiv:2012.01249 \[hep-ex, physics:hep-ph, physics:physics\]](#).
- [191] CMS Collaboration, “Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques”, *Journal of Instrumentation* **15**, P06005 (2020) [10.1088/1748-0221/15/06/P06005](#), [arXiv:2004.08262 \[hep-ex\]](#).
- [192] A. Butter et al., “Machine learning and LHC event generation”, *SciPost Physics* **14**, 079 (2023) [10.21468/SciPostPhys.14.4.079](#), [arXiv:2203.07460 \[hep-ph\]](#).
- [193] T. Q. Nguyen et al., “Topology Classification with Deep Learning to Improve Real-Time Event Selection at the LHC”, *Computing and Software for Big Science* **3**, 12 (2019) [10.1007/s41781-019-0028-1](#).
- [194] J. Duarte et al., “Fast inference of deep neural networks in FPGAs for particle physics”, *Journal of Instrumentation* **13**, P07027 (2018) [10.1088/1748-0221/13/07/P07027](#), [arXiv:1804.06913 \[physics.ins-det\]](#).
- [195] G. Kasieczka et al., “The Machine Learning landscape of top taggers”, *SciPost Physics* **7**, 014 (2019) [10.21468/SciPostPhys.7.1.014](#), [arXiv:1902.09914 \[hep-ph\]](#).
- [196] J. Lin et al., “Boosting $H \rightarrow b\bar{b}$ with machine learning”, *Journal of High Energy Physics* **2018**, 101 (2018) [10.1007/JHEP10\(2018\)101](#), [arXiv:1807.10768 \[hep-ph\]](#).
- [197] T. Li et al., “Semi-supervised graph neural networks for pileup noise removal”, *The European Physical Journal C* **83**, 99 (2023) [10.1140/epjc/s10052-022-11083-5](#), [arXiv:2203.15823 \[hep-ex\]](#).

- [198] B. Lieberman et al., *Trials Factor for Semi-Supervised NN Classifiers in Searches for Narrow Resonances at the LHC*, (2024) [arXiv:2404.07822 \[hep-ph\]](#), pre-published.
- [199] L. M. Dery et al., “Weakly supervised classification in high energy physics”, *Journal of High Energy Physics* **2017**, 145 (2017) [10.1007/JHEP05\(2017\)145](#), [arXiv:1702.00414 \[hep-ph\]](#).
- [200] E. M. Metodiev, B. Nachman and J. Thaler, “Classification without labels: learning from mixed samples in high energy physics”, *Journal of High Energy Physics* **2017**, 174 (2017) [10.1007/JHEP10\(2017\)174](#), [arXiv:1708.02949 \[hep-ph\]](#).
- [201] J. Hernández-González, I. Inza and J. A. Lozano, “Weak supervision and other non-standard classification problems: A taxonomy”, *Pattern Recognition Letters* **69**, 49 (2016) [10.1016/j.patrec.2015.10.008](#).
- [202] J. Bardhan et al., *Unsupervised and lightly supervised learning in particle physics*, (2024) [arXiv:2403.13676 \[hep-ph\]](#), pre-published.
- [203] A. Blance, M. Spannowsky and P. Waite, “Adversarially-trained autoencoders for robust unsupervised new physics searches”, *Journal of High Energy Physics* **2019**, 47 (2019) [10.1007/JHEP10\(2019\)047](#), [arXiv:1905.10384 \[hep-ph\]](#).
- [204] E. Govorkova et al., “Autoencoders on field-programmable gate arrays for real-time, unsupervised new physics detection at 40 MHz at the Large Hadron Collider”, *Nature Machine Intelligence* **4**, 154 (2022) [10.1038/s42256-022-00441-3](#), [arXiv:2108.03986 \[physics.ins-det\]](#).
- [205] *Quark versus gluon jet tagging using jet images with the ATLAS detector*, ATL-PHYS-PUB-2017-017 (CERN, Geneva, 2017), <https://cds.cern.ch/record/2275641>.
- [206] ATLAS Collaboration, *Accuracy versus precision in boosted top tagging with the ATLAS detector*, (July 2024) [arXiv:2407.20127 \[hep-ex\]](#), pre-published.
- [207] ATLAS Collaboration, “Fast b-tagging at the high-level trigger of the ATLAS experiment in LHC Run 3”, *Journal of Instrumentation* **18**, P11006 (2023) [10.1088/1748-0221/18/11/P11006](#), [arXiv:2306.09738 \[hep-ex\]](#).
- [208] ATLAS Collaboration, “ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset”, *The European Physical Journal C* **83**, 681 (2023) [10.1140/epjc/s10052-023-11699-1](#), [arXiv:2211.16345 \[physics.data-an\]](#).
- [209] P. Baldi, P. Sadowski and D. Whiteson, “Searching for exotic particles in high-energy physics with deep learning”, *Nature Communications* **5**, 4308 (2014) [10.1038/ncomms5308](#).
- [210] B. Nachman, “Anomaly Detection for Physics Analysis and Less than Supervised Learning”, Oct. 2020, [arXiv:2010.14554 \[hep-ex, physics:hep-ph, physics:physics\]](#).

- [211] V. Belis, P. Odagiu and T. K. Årrestad, “Machine Learning for Anomaly Detection in Particle Physics”, *Reviews in Physics* **12**, 100091 (2024) [10.1016/j.revip.2024.100091](#), [arXiv:2312.14190 \[hep-ex, physics:physics, physics:quant-ph\]](#).
- [212] B. M. Dillon, D. A. Faroughy and J. F. Kamenik, “Uncovering latent jet substructure”, *Physical Review D* **100**, 056002 (2019) [10.1103/PhysRevD.100.056002](#), [arXiv:1904.04200 \[hep-ph\]](#).
- [213] E. A. Moreno et al., “JEDI-net: a jet identification algorithm based on interaction networks”, 10th Sept. 2019, [arXiv:1908.05318 \[hep-ex, physics:hep-ph\]](#).
- [214] T. Cheng et al., “Variational Autoencoders for Anomalous Jet Tagging”, 29th Sept. 2020, [arXiv:2007.01850](#).
- [215] O. Knapp et al., “Adversarially Learned Anomaly Detection on CMS Open Data: re-discovering the top quark”, 4th May 2020, [arXiv:2005.01598 \[hep-ex, physics:hep-ph\]](#).
- [216] D0 Collaboration, “Search for New Physics in e mu X Data at D0 Using Sleuth: A Quasi-Model-Independent Search Strategy for New Physics”, 2000, [10.48550/arXiv.hep-ex/0006011](#), [arXiv:hep-ex/0006011](#).
- [217] CMS Collaboration, “MUSiC: a model-unspecific search for new physics in proton–proton collisions at $\sqrt{s} = 13$ TeV”, *Eur. Phys. J. C* **81**, 629 (2021) [10.1140/epjc/s10052-021-09236-z](#), [arXiv:2010.02984 \[hep-ex\]](#).
- [218] G. Kasieczka et al., “The LHC Olympics 2020: a community challenge for anomaly detection in high energy physics”, *Reports on Progress in Physics* **84**, 124201 (2021) [10.1088/1361-6633/ac36b9](#), [arXiv:2101.08320 \[hep-ph\]](#).
- [219] T. Aarrestad et al., “The Dark Machines Anomaly Score Challenge: Benchmark Data and Model Independent Event Classification for the Large Hadron Collider”, *SciPost Physics* **12**, 043 (2022) [10.21468/SciPostPhys.12.1.043](#), [arXiv:2105.14027 \[hep-ph\]](#).
- [220] B. Russell, *The scientific outlook* (Routledge, 1931).
- [221] L. Evans and P. Bryant, “LHC machine”, *Journal of Instrumentation* **3**, S08001 (2008) [10.1088/1748-0221/3/08/S08001](#).
- [222] ATLAS Collaboration, “The ATLAS Experiment at the CERN Large Hadron Collider”, *Journal of Instrumentation* **3**, S08003 (2008) [10.1088/1748-0221/3/08/S08003](#).
- [223] ATLAS Collaboration, “The ATLAS experiment at the CERN Large Hadron Collider: a description of the detector configuration for Run 3”, *Journal of Instrumentation* **19**, P05063 (2024) [10.1088/1748-0221/19/05/P05063](#).

- [224] ATLAS Collaboration, “The ATLAS Simulation Infrastructure”, *The European Physical Journal C* **70**, 823 (2010) [10.1140/epjc/s10052-010-1429-9](#), [arXiv:1005.4568 \[physics.ins-det\]](#).
- [225] E. Lopienska, *The CERN accelerator complex, layout in 2022*, 2022, <https://cds.cern.ch/record/2800984>.
- [226] L. Evans, “The Large Hadron Collider”, *Annual Review of Nuclear and Particle Science* **61**, 435 (2011) [10.1146/annurev-nucl-102010-130438](#).
- [227] R. Aßmann, M. Lamont and S. Myers, “A brief history of the LEP collider”, *Nuclear Physics B - Proceedings Supplements* **109**, 17 (2002) [10.1016/S0920-5632\(02\)90005-8](#).
- [228] CMS Collaboration, “The CMS experiment at the CERN LHC”, *Journal of Instrumentation* **3**, S08004 (2008) [10.1088/1748-0221/3/08/S08004](#).
- [229] LHCb Collaboration, “The LHCb Detector at the LHC”, *Journal of Instrumentation* **3**, S08005 (2008) [10.1088/1748-0221/3/08/S08005](#).
- [230] ALICE Collaboration, “The ALICE experiment at the CERN LHC”, *Journal of Instrumentation* **3**, S08002 (2008) [10.1088/1748-0221/3/08/S08002](#).
- [231] TOTEM Collaboration, “The TOTEM Experiment at the CERN Large Hadron Collider”, *Journal of Instrumentation* **3**, S08007 (2008) [10.1088/1748-0221/3/08/S08007](#).
- [232] MoEDAL collaboration, *Technical design report of the MoEDAL experiment*, CERN-LHCC-2009-006, MoEDAL-TDR-001 (2009), <https://cds.cern.ch/record/1181486>.
- [233] LHCf Collaboration, “The LHCf detector at the CERN Large Hadron Collider”, *Journal of Instrumentation* **3**, S08006 (2008) [10.1088/1748-0221/3/08/S08006](#).
- [234] SND Collaboration, “SND@LHC: the scattering and neutrino detector at the LHC”, *Journal of Instrumentation* **19**, P05067 (2024) [10.1088/1748-0221/19/05/P05067](#), [arXiv:2210.02784 \[hep-ex\]](#).
- [235] J. L. Feng et al., “ForwArd Search ExpeRiment at the LHC”, *Physical Review D* **97**, 035001 (2018) [10.1103/PhysRevD.97.035001](#), [arXiv:1708.09389 \[hep-ph\]](#).
- [236] “CERN Yellow Reports: Monographs, Vol. 10 (2020): High-Luminosity Large Hadron Collider (HL-LHC): Technical design report”, [10.23731/CYRM-2020-0010](#) (2020) [10.23731/CYRM-2020-0010](#).
- [237] G. Apollinari et al., “High Luminosity Large Hadron Collider HL-LHC”, [10.5170/CERN-2015-005.1](#) (2015) [10.5170/CERN-2015-005.1](#).
- [238] G. Arduini et al., “LHC Upgrades in preparation of Run 3”, *Journal of Instrumentation* **19**, P05061 (2024) [10.1088/1748-0221/19/05/P05061](#).

- [239] H. S. Padamsee, “Superconducting Radio-Frequency Cavities”, *Annual Review of Nuclear and Particle Science* **64**, 175 (2014) [10.1146/annurev-nucl-102313-025612](#).
- [240] ATLAS Collaboration, *ATLAS experiment schematic or layout illustration*, 2022, <https://cds.cern.ch/record/2837191>.
- [241] CERN AC, *Layout of ATLAS*, Mar. 1998, <https://cds.cern.ch/record/39038>.
- [242] W. Herr and B. Muratori, “Concept of luminosity”, [10.5170/CERN-2006-002.361](#) (2006) [10.5170/CERN-2006-002.361](#).
- [243] ATLAS Collaboration, “Measurement of the $t\bar{t}t\bar{t}$ production cross section in pp collisions at $\sqrt{s}=13$ TeV with the ATLAS detector”, 22nd June 2021, [arXiv:2106.11683 \[hep-ex\]](#).
- [244] J. Pequeno and P. Schaffner, *How ATLAS detects particles: diagram of particle paths in the detector*, Jan. 2013, <https://cds.cern.ch/record/1505342>.
- [245] ATLAS Collaboration, “Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at $\sqrt{s} = 13$ TeV”, *The European Physical Journal C* **79**, 639 (2019) [10.1140/epjc/s10052-019-7140-6](#), [arXiv:1902.04655 \[physics.ins-det\]](#).
- [246] T. Kawamoto et al., *New small wheel technical design report*, CERN-LHCC-2013-006, ATLAS-TDR-020 (2013), <https://cds.cern.ch/record/1552862>.
- [247] K. Bos et al., *LHC computing grid: Technical design report*, Technical design report. LCG (CERN, 2005).
- [248] *The ATLAS collaboration software and firmware*, ATL-SOFT-PUB-2021-001 (CERN, Geneva, 2021), <https://cds.cern.ch/record/2767187>.
- [249] G. Barrand et al., “GAUDI — A software architecture and framework for building HEP data processing applications”, *Computer Physics Communications* **140**, 45 (2001) [10.1016/S0010-4655\(01\)00254-5](#).
- [250] M. Cattaneo et al., *Status of the GAUDI event-processing framework* (), <https://cds.cern.ch/record/2810842>.
- [251] S. Alioli et al., “NLO vector-boson production matched with shower in POWHEG”, *Journal of High Energy Physics* **2008**, 060 (2008) [10.1088/1126-6708/2008/07/060](#), [arXiv:0805.4802](#).
- [252] T. Sjöstrand, S. Mrenna and P. Skands, “A brief introduction to PYTHIA 8.1”, *Computer Physics Communications* **178**, 852 (2008) [10.1016/j.cpc.2008.01.036](#), [arXiv:0710.3820](#).
- [253] S. Agostinelli et al., “Geant4—a simulation toolkit”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506**, 250 (2003) [10.1016/S0168-9002\(03\)01368-8](#).

- [254] P. T. Komiske et al., “Learning to classify from impure samples with high-dimensional data”, *Physical Review D* **98**, 011502 (2018) [10.1103/PhysRevD.98.011502](https://arxiv.org/abs/10.1103/PhysRevD.98.011502).
- [255] T. Dorigo and P. de Castro, “Dealing with Nuisance Parameters using Machine Learning in High Energy Physics: a Review”, 17th Jan. 2021, [arXiv:2007.09121](https://arxiv.org/abs/2007.09121) [[hep-ex](#), [physics:hep-ph](#), [physics:physics](#), [stat](#)].
- [256] ATLAS Collaboration, *Letter of intent for the phase-II upgrade of the ATLAS experiment*, CERN-LHCC-2012-022, LHCC-I-023 (CERN, Geneva, 2012), <https://cds.cern.ch/record/1502664>.
- [257] ATLAS Collaboration, *ATLAS phase-II upgrade scoping document*, CERN-LHCC-2015-020, LHCC-G-166 (CERN, Geneva, 2015), <https://cds.cern.ch/record/2055248>.
- [258] N. Gaiman, *American gods*, 1st ed (W. Morrow, 2001).
- [259] ATLAS Collaboration, “Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector”, *Physics Letters B* **784**, 173 (2018) [10.1016/j.physletb.2018.07.035](https://arxiv.org/abs/10.1016/j.physletb.2018.07.035), [arXiv:1806.00425](https://arxiv.org/abs/1806.00425) [[hep-ex](#)].
- [260] ATLAS Collaboration, “Measurement of the W-boson mass in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector”, *The European Physical Journal C* **78**, 110 (2018) [10.1140/epjc/s10052-017-5475-4](https://arxiv.org/abs/10.1140/epjc/s10052-017-5475-4), [arXiv:1701.07240](https://arxiv.org/abs/1701.07240) [[hep-ex](#)].
- [261] ATLAS Collaboration, “Measurement of the total and differential Higgs boson production cross-sections at $\sqrt{s} = 13$ TeV with the ATLAS detector by combining the $H \rightarrow ZZ^* \rightarrow 4\ell$ and $H \rightarrow \gamma\gamma$ decay channels”, *Journal of High Energy Physics* **2023**, 28 (2023) [10.1007/JHEP05\(2023\)028](https://arxiv.org/abs/10.1007/JHEP05(2023)028), [arXiv:2207.08615](https://arxiv.org/abs/2207.08615) [[hep-ex](#)].
- [262] ATLAS Collaboration, “Search for new phenomena with top-quark pairs and large missing transverse momentum using 140 fb^{-1} of pp collision data at $\sqrt{s} = 13$ TeV with the ATLAS detector”, *Journal of High Energy Physics* **2024**, 139 (2024) [10.1007/JHEP03\(2024\)139](https://arxiv.org/abs/10.1007/JHEP03(2024)139), [arXiv:2401.13430](https://arxiv.org/abs/2401.13430) [[hep-ex](#)].
- [263] ATLAS Collaboration, “Search for lepton-flavour violation in high-mass dilepton final states using 139 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”, *Journal of High Energy Physics* **2023**, 82 (2023) [10.1007/JHEP10\(2023\)082](https://arxiv.org/abs/10.1007/JHEP10(2023)082), [arXiv:2307.08567](https://arxiv.org/abs/2307.08567) [[hep-ex](#)].
- [264] ATLAS Collaboration, “Measurement of the forward-backward asymmetry of electron and muon pair-production in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector”, *Journal of High Energy Physics* **2015**, 49 (2015) [10.1007/JHEP09\(2015\)049](https://arxiv.org/abs/10.1007/JHEP09(2015)049), [arXiv:1503.03709](https://arxiv.org/abs/1503.03709) [[hep-ex](#)].

- [265] ATLAS Collaboration, “Measurement of the Higgs boson mass in the $H \rightarrow ZZ^* \rightarrow 4\ell$ decay channel using 139 fb^{-1} of $\sqrt{s} = 13 \text{ TeV}$ pp collisions recorded by the ATLAS detector at the LHC”, *Physics Letters B* **843**, 137880 (2023) [10.1016/j.physletb.2023.137880](#), [arXiv:2207.00320 \[hep-ex\]](#).
- [266] ATLAS Collaboration, “Electron and photon efficiencies in LHC Run 2 with the ATLAS experiment”, *Journal of High Energy Physics* **2024**, 162 (2024) [10.1007/JHEP05\(2024\)162](#), [arXiv:2308.13362](#).
- [267] W. Lampl et al., *Calorimeter clustering algorithms: Description and performance*, ATL-LARG-PUB-2008-002, ATL-COM-LARG-2008-003 (CERN, Geneva, 2008), <https://cds.cern.ch/record/1099735>.
- [268] ATLAS Collaboration, “Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data”, *Journal of Instrumentation* **14**, P12006 (2019) [10.1088/1748-0221/14/12/P12006](#), [arXiv:1908.00005 \[hep-ex\]](#).
- [269] ATLAS Collaboration, “Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton-proton collision data”, *Journal of Instrumentation* **14**, P03017 (2019) [10.1088/1748-0221/14/03/P03017](#), [arXiv:1812.03848 \[hep-ex\]](#).
- [270] ATLAS Collaboration, “Performance of the ATLAS Transition Radiation Tracker in Run 1 of the LHC: tracker properties”, *Journal of Instrumentation* **12**, P05002 (2017) [10.1088/1748-0221/12/05/P05002](#), [arXiv:1702.06473 \[hep-ex\]](#).
- [271] ATLAS Collaboration, “Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data”, *The European Physical Journal C* **72**, 1909 (2012) [10.1140/epjc/s10052-012-1909-1](#), [arXiv:1110.3174 \[hep-ex\]](#).
- [272] ATLAS Collaboration, “Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton–proton collision data”, *The European Physical Journal C* **74**, 2941 (2014) [10.1140/epjc/s10052-014-2941-0](#), [arXiv:1404.2240 \[hep-ex\]](#).
- [273] Kurt Brendlinger, “Physics with Electrons in the ATLAS Detector”, PhD thesis (University of Pennsylvania, 2016), <https://cds.cern.ch/record/2228644?ln=en>.
- [274] ATLAS Collaboration, “Electron efficiency measurements with the ATLAS detector using 2012 LHC proton–proton collision data”, *The European Physical Journal C* **77**, 195 (2017) [10.1140/epjc/s10052-017-4756-2](#), [arXiv:1612.01456 \[hep-ex\]](#).
- [275] D. Adams, *The Hitchhiker’s guide to the Galaxy*, Pan original (Pan Books, 1979).

- [276] ATLAS Collaboration, “Precise test of lepton flavour universality in W-boson decays into muons and electrons in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”, *The European Physical Journal C* **84**, 993 (2024) [10.1140/epjc/s10052-024-13070-4](#), [arXiv:2403.02133 \[hep-ex\]](#).
- [277] ATLAS Collaboration, “Electron and photon energy calibration with the ATLAS detector using LHC Run 2 data”, *Journal of Instrumentation* **19**, P02009 (2024) [10.1088/1748-0221/19/02/P02009](#), [arXiv:2309.05471 \[hep-ex\]](#).
- [278] ATLAS Collaboration, *TagAndProbe*, ATLAS collaboration, 2019, <https://tagandprobe.readthedocs.io>.
- [279] ATLAS Collaboration, *Support Note for Electron ID: efficiency measurements*, 5th Sept. 2017, <https://cds.cern.ch/record/2282430>.
- [280] ATLAS Collaboration, *Electron identification efficiencies using the Zmass method*, 4th Mar. 2021, <https://cds.cern.ch/record/2798044>.
- [281] E. M. Rüttinger, “Electron identification efficiency and W^+W^- cross section measurements with the ATLAS detector”, PhD thesis (University of Sheffield, 2020), <http://etheses.whiterose.ac.uk/28606/>.
- [282] W. Verkerke and D. Kirkby, *The RooFit toolkit for data modeling*, (2003) [arXiv:physics/0306116](#), pre-published.
- [283] C. Blocker, “Uncertainties on Efficiencies”, *CDF/MEMO/STATISTICS/PUBLIC/7168* (2004).
- [284] R. Brun et al., *root-project/root: v6.18/02*, version v6-18-02, Zenodo, 23rd Aug. 2019, [10.5281/ZENODO.3895860](#), <https://zenodo.org/record/3895860>.
- [285] D. Adams, *The Salmon of Doubt: Hitchhiking the Galaxy One Last Time* (William Heinemann Ltd., 2002).
- [286] ATLAS Collaboration, “Search for new phenomena in a lepton plus high jet multiplicity final state with the ATLAS experiment using $\sqrt{s} = 13$ TeV proton-proton collision data”, *Journal of High Energy Physics* **2017**, 88 (2017) [10.1007/JHEP09\(2017\)088](#), [arXiv:1704.08493 \[hep-ex\]](#).
- [287] ATLAS Collaboration, “Identification of electrons using a deep neural network in the ATLAS experiment”, *ATL-PHYS-PUB-2022-022* (2022).
- [288] S. Datta, “A Review on Convolutional Neural Networks”, in *Advances in Communication, Devices and Networking*, Vol. 662, edited by R. Bera et al. (Springer Singapore, Singapore, 2020), pp. 445–452, [10.1007/978-981-15-4932-8_50](#).
- [289] F. Pedregosa et al., “Scikit-learn: Machine learning in Python”, *Journal of Machine Learning Research* **12**, 2825 (2011).

- [290] CMS Collaboration, “Observation of Electroweak Production of Same-Sign W Boson Pairs in the Two Jet and Two Same-Sign Lepton Final State in Proton-Proton Collisions at $s = 13$ TeV”, *Physical Review Letters* **120**, 081801 (2018) [10.1103/PhysRevLett.120.081801](#), [arXiv:1709.05822 \[hep-ex\]](#).
- [291] ATLAS Collaboration, “Search for new phenomena in events with same-charge leptons and b-jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”, *Journal of High Energy Physics* **2018**, 39 (2018) [10.1007/JHEP12\(2018\)039](#), [arXiv:1807.11883 \[hep-ex\]](#).
- [292] Martín Abadi et al., *TensorFlow: Large-scale machine learning on heterogeneous systems*, 2015, <https://www.tensorflow.org/>.
- [293] Olivier Denis, “Identification des électrons dans l’expérience ATLAS à l’aide de réseaux de neurones convolutifs entraînés dans les données expérimentales”, MSc dissertation (Université de Montréal, Nov. 2023), <https://hdl.handle.net/1866/33044>.
- [294] G. Louppe, M. Kagan and K. Cranmer, “Learning to Pivot with Adversarial Networks”, 1st June 2017, [arXiv:1611.01046 \[physics, stat\]](#).
- [295] J. H. Kim et al., “The motivation and status of two-body resonance decays after the LHC Run 2 and beyond”, *Journal of High Energy Physics* **2020**, 30 (2020) [10.1007/JHEP04\(2020\)030](#), [arXiv:1907.06659 \[hep-ph\]](#).
- [296] S. V. Chekanov, *Estimation of the chances to find new phenomena at the LHC in a model-agnostic combinatorial analysis*, (Nov. 2023) [arXiv:2311.09012 \[hep-ph\]](#), <http://arxiv.org/abs/2311.09012>, pre-published.
- [297] J. R. Klein and A. Roodman, “Blind analysis in nuclear and particle physics”, *Annual Review of Nuclear and Particle Science* **55**, 141 (2005) [10.1146/annurev.nucl.55.090704.151521](#).
- [298] ATLAS Collaboration, “A search for new resonances in multiple final states with a high transverse momentum Z boson in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector”, *Journal of High Energy Physics* **2023**, 36 (2023) [10.1007/JHEP06\(2023\)036](#), [arXiv:2209.15345 \[hep-ex\]](#).
- [299] ATLAS Collaboration, “Search for New Phenomena in Two-Body Invariant Mass Distributions Using Unsupervised Machine Learning for Anomaly Detection at $s = 13$ TeV with the ATLAS Detector”, *Physical Review Letters* **132**, 081801 (2024) [10.1103/PhysRevLett.132.081801](#), [arXiv:2307.01612 \[hep-ex\]](#).
- [300] J. H. Collins, K. Howe and B. Nachman, “Extending the search for new resonances with machine learning”, *Physical Review D* **99**, 014038 (2019) [10.1103/PhysRevD.99.014038](#).

- [301] ATLAS Collaboration, “Dijet resonance search with weak supervision using $\sqrt{s} = 13$ TeV pp collisions in the ATLAS detector”, *Physical Review Letters* **125**, 131801 (2020) [10.1103/PhysRevLett.125.131801](#), [arXiv:2005.02983 \[hep-ex\]](#).
- [302] ATLAS Collaboration, “A strategy for a general search for new phenomena using data-derived signal regions and its application within the ATLAS experiment”, *Eur. Phys. J. C* **79**, 120 (2019) [10.1140/epjc/s10052-019-6540-y](#), [arXiv:1807.07447 \[hep-ex\]](#).
- [303] E. Gross and O. Vitells, “Trial factors for the look elsewhere effect in high energy physics”, *The European Physical Journal C* **70**, 525 (2010) [10.1140/epjc/s10052-010-1470-8](#).
- [304] B. Nachman, “A guide for deploying Deep Learning in LHC searches: How to achieve optimality and account for uncertainty”, *SciPost Physics* **8**, 090 (2020) [10.21468/SciPostPhys.8.6.090](#).
- [305] J. A. Raine et al., “CURTAINS for your sliding window: Constructing unobserved regions by transforming adjacent intervals”, *Frontiers in Big Data* **6**, 899345 (2023) [10.3389/fdata.2023.899345](#), [arXiv:2203.09470 \[hep-ph\]](#).
- [306] S. Volkovich, F. De Vito Halevy and S. Bressler, “A data-directed paradigm for BSM searches: the bump-hunting example”, *The European Physical Journal C* **82**, 265 (2022) [10.1140/epjc/s10052-022-10215-1](#), [arXiv:2107.11573 \[hep-ex\]](#).
- [307] M. Birman et al., “Data-Directed Search for New Physics based on Symmetries of the SM”, *The European Physical Journal C* **82**, 508 (2022) [10.1140/epjc/s10052-022-10454-2](#), [arXiv:2203.07529 \[hep-ph\]](#).
- [308] J. Alwall et al., “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”, *Journal of High Energy Physics* **2014**, 79 (2014) [10.1007/JHEP07\(2014\)079](#), [arXiv:1405.0301 \[hep-ph\]](#).
- [309] T. Sjöstrand et al., “An introduction to PYTHIA 8.2”, *Computer Physics Communications* **191**, 159 (2015) [10.1016/j.cpc.2015.01.024](#).
- [310] DELPHES Collaboration, “DELPHES 3: a modular framework for fast simulation of a generic collider experiment”, *Journal of High Energy Physics* **2014**, 57 (2014) [10.1007/JHEP02\(2014\)057](#).
- [311] ATLAS Collaboration, “Determination of jet calibration and energy resolution in proton–proton collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector”, *The European Physical Journal C* **80**, 1104 (2020) [10.1140/epjc/s10052-020-08477-8](#), [arXiv:1910.04482 \[hep-ex\]](#).