

ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA
UNIVERSITÉ CLERMONT AUVERGNE
TECHNISCHE UNIVERSITÄT DORTMUND

International Master on Advanced Methods in Particle Physics

Suppressing pile-up contributions in the formation of topological clusters in ATLAS

Supervisor and first referee:
Dr. Chris Malena Delitzsch

Presented by:
Giulia Fazzino

Second referee:
Prof. Maximiliano Sioli

Academic Year 2023/2024

Abstract

The production of particles in secondary proton-proton collisions, known as pile-up, is a phenomenon which affects the measurements performed in high-luminosity particle colliders, such as the LHC.

This work explores the possibility of using deep neural networks to mitigate pile-up contributions to the measured energy of topoclusters in the ATLAS experiment, studying clusters from simulated Run 2 and Run 3 events.

The implemented networks aim to determine whether clusters contain contributions arising only from pile-up, solely from the hard scatter interaction, or from a combination of the two. After such classification, clusters identified as pile-up-only could be excluded from the jet reconstruction inputs, while pile-up contributions in the mixed clusters could be suppressed.

The results demonstrate the effectiveness of the networks in providing a classification which improves the energy response of the topoclusters. In particular, the suppression of pile-up-only clusters reduces the interquartile range of the response distribution by a factor of approximately 10^3 for clusters in Run 2 and 10^2 for those in Run 3, indicating the great improvements that pile-up mitigation can bring to cluster energy calibration.

Contents

Introduction	3
1 Theoretical Background	5
1.1 Brief Introduction to the Standard Model of Particle Physics	5
1.2 Jets	7
2 The ATLAS Experiment at the Large Hadron Collider	8
2.1 Overview of the Large Hadron Collider	8
2.1.1 Luminosity and pile-up	9
2.2 Overview of the ATLAS detector	10
3 Topoclusters in Jet Reconstruction	17
3.1 Jet Reconstruction Algorithms	17
3.2 Formation of Calorimeter Topoclusters	18
3.2.1 Cluster Energy Calibration	19
3.3 Pile-up Mitigation in Jet Reconstruction Inputs	21
4 Analysis Methods	24
4.1 Key Idea: Suppression via Topocluster Classification	24
4.1.1 Datasets	24
4.2 Deep Neural Networks	26
4.2.1 Network Architecture	27
4.2.2 Figures of Merit	28
4.2.3 Input Features Choice	30
5 Results	33
5.1 Preliminary Studies on Topocluster Response	33
5.2 Hard Scatter-Only Clusters Identification	35
5.2.1 Classification Results on Run 2	36
5.2.2 Classification Results on Run 3	40
5.3 Pile-Up-Only Clusters Identification	43

5.3.1	Classification Results on Run 2	48
5.3.2	Classification Results on Run 3	52
Conclusions		56
Bibliography		59

Introduction

Our knowledge about elementary particles and their interactions is encompassed in the Standard Model (SM) of Particle Physics. Formulated in the late 20th century, the SM is one of the most accurate theories ever produced, and experimental searches have repeatedly confirmed its predictions. Nonetheless, the SM remains the subject of tests, aimed at precisely measuring its parameters and addressing the questions it leaves unanswered.

Since very high energies are needed to probe the properties of this theory, several particle colliders have been built over the last decades. In these experimental facilities, beams of particles are brought to collision by powerful accelerators, which nowadays succeed in reaching energies in the order of the TeV. The state-of-the-art of these machines is the Large Hadron Collider (LHC), a protons and heavy ions collider located at the European Council for Nuclear Research (CERN), which can reach a collision energy of 13.6 TeV.

The very high centre of mass energy reached by the LHC allows for the search for heavy particles and rare phenomena, but at the cost of multiple interactions taking place simultaneously in each collision. This phenomenon, known as pile-up, presents a challenge to the study of any process, because it introduces uncertainty as to whether an observed signal was produced by the interaction under study or by a secondary one. While there are already several methods to address this problem, the possibility of using machine learning to minimize pile-up contributions remains to be explored.

This work focuses on the use of deep neural networks to suppress the energy contributions arising from pile-up in topoclusters, *i.e.* groups of cells in the calorimeter showing a non-negligible signal, within the ATLAS experiment. In particular, the possibility of assessing the nature of the contributions to a topocluster via a machine learning-based classification is explored, together with the possible improvements this could bring to the energy calibration. These potential improvements are particularly important in view of the crucial role played by topoclusters in the study of processes involving particles interacting via the strong force. Indeed, topoclusters are used as inputs for the reconstruction of jets, which are the experimental signatures of such particles. Therefore, improving the energy calibration of topoclusters by effectively suppressing pile-up contributions would lead to a more accurate jet reconstruction.

In the following, Chapter 1 will give an overview of the SM and jet formation, focusing

on the theoretical background necessary to understand the studies performed in this work.

Then, Chapter 2 will provide an outline of the LHC, introducing the concepts of pile-up and luminosity, and will offer a description of the ATLAS detector, including information on its main subsystems.

Topoclusters will be introduced in Chapter 3, discussing the methods for calibrating their energy and the current techniques for pile-up suppression.

Subsequently, Chapter 4 will delve into the key idea behind the cluster classification performed in this thesis. First, the data used will be presented along with its main characteristics. The chapter will then go into detail about the two networks implemented for this research.

Finally, Chapter 5 will present the results obtained from these studies. Beginning with preliminary analyses focused on exploring the motivations behind pile-up suppression, the chapter will then discuss the results obtained on the possibility of using neural networks to identify the nature of topoclusters across different data-taking periods of the LHC.

Chapter 1

Theoretical Background

1.1 Brief Introduction to the Standard Model of Particle Physics

Formulated in the 1970s, the Standard Model (SM) of Particle Physics [1] is the theoretical framework describing the interactions among elementary particles, which are interpreted as excitations of the corresponding quantum fields. Depicted in Figure 1.1, the particle content of the model is divided into two categories: fermions, which are the building blocks of matter, and bosons, which are the mediators of the three interactions described by the model. Indeed, the SM formalizes the understanding of three out of the four fundamental forces governing our Universe: the electromagnetic, weak, and strong interactions. Missing from this list is the gravitational force, for which a formulation in terms of a quantum field theory does not exist yet. However, while the understanding of this interaction remains an open question of the theory, its absence does not affect the interpretation of the physics observed in particle accelerators, as they probe energy scales where the strength of the gravitational force is negligible.

Fermions

As shown in Figure 1.1, the twelve fermions of the SM can be organized in two ways. Firstly, fermions can be either quarks or leptons, depending on whether they are strongly interacting or not, respectively. Moreover, they are divided into three generations, with particles belonging to different generations (*i.e.* fermions in the same row in the figure) sharing all their quantum numbers and having masses which increase from one generation to the next. Being the lightest, fermions of the first generation are also the most stable, since there are no particles with a smaller mass to which they could decay. Consequently, they are the particles which ordinary matter is made of, while the other fermions can only be studied in high-energy environments, such as cosmic rays [2] or particle accelerators [3].

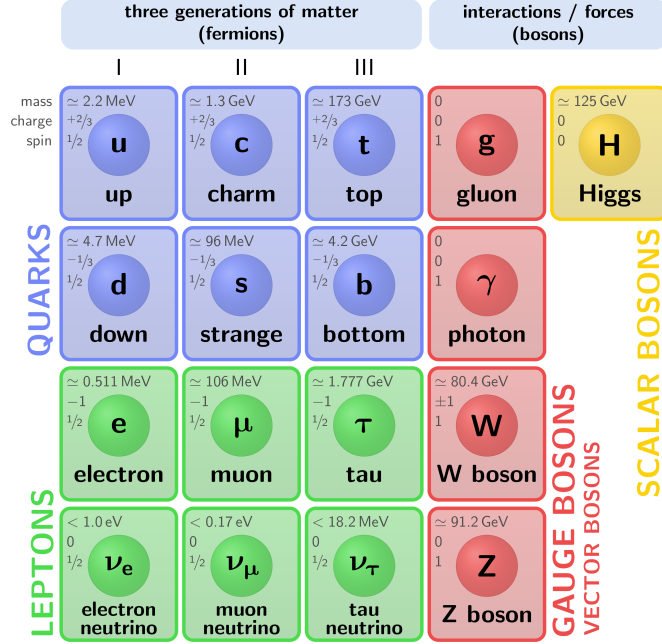


Figure 1.1: Particle content of the SM. Image generated with the code available at [4].

Bosons and Interactions

Bosons can be either vectors or scalars depending on their spin, a quantized intrinsic angular momentum which is 1 for the first category, and 0 for the second (in natural units, *i.e.* with the reduced Planck constant $\hbar = 1$). The SM features only one scalar boson: the Higgs boson, responsible for the mechanism giving mass to every other massive particle [5]. The vector bosons, on the other hand, are the carriers of the three interactions featured in the model, meaning that whenever these bosons are exchanged by two or more particles, the interaction they mediate occurs.

More specifically, the electromagnetic interaction, which happens among electrically charged particles, is carried by photons. Since its mediator is massless, electromagnetism is the only interaction, among the ones encompassed in the SM, producing long-range effects.

The strong interaction, theorized by quantum chromodynamics (QCD) [6], happens among quarks and gluons, which are the only particles carrying a colour charge (analogous to the electric charge for the electromagnetic interaction). The behaviour of this interaction at different energy scales¹ has two peculiar properties: asymptotic freedom and colour confinement. The former is an asymptotic reduction of the interaction strength as the energy scale increases, which implies that at high energies, such as those reached

¹The energy and length scales (E and L) are related by the quantum mechanical relation $L = \hbar c/E$, thus high energies correspond to small lengths and vice versa.

in colliders, quarks behave as free particles. On the other hand, because of colour confinement, coloured particles cannot exist isolated at low energies², and must therefore bind together forming colourless composite particles called hadrons. For this reason, the strong interaction does not produce long-range effects, despite having a massless force carrier.

Lastly, the weak interaction is carried by $W^\pm$ and Z bosons, the only massive gauge bosons in the theory. It is the only interaction to which all the fermions are sensitive, and it allows them to change flavour, *i.e.* it can mediate the decay of a particle of one generation into a particle of another.

1.2 Jets

Because of colour confinement, when quarks and gluons (collectively referred to as partons) are produced in high-energy processes, such as proton-proton collisions, they cannot propagate freely. Instead, they undergo a process that generates a collimated spray of colourless particles, called a jet. Information on how jets are defined in colliders can be found in Section 3.1, while here their formation process will be explained.

The formation of a jet follows two steps, shown in Figure 1.2: firstly, the initial parton emits other quarks and gluons as it propagates, in a process referred to as parton showering or fragmentation. The partons produced in this way can in turn undergo the same process, which repeats itself until the energies of the particles produced are low enough for hadronization to take place. At this stage, quarks and gluons combine into hadrons, forming a cone of colourless particles.

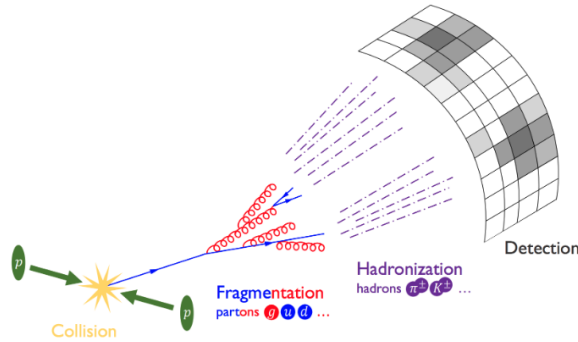


Figure 1.2: Schematic view of the formation of a jet from a parton produced in a proton-proton collision. Imagen taken from [7].

²To give a reference value, colour confinement happens below energies of the order of 1 GeV [6].

Chapter 2

The ATLAS Experiment at the Large Hadron Collider

2.1 Overview of the Large Hadron Collider

The Large Hadron Collider (LHC) [8] is the world's largest and most powerful particle accelerator, which entered operation for the first time in 2008. It is a circular hadron and ion collider, accelerator, and storage ring, located in a tunnel with a 27 km circumference at the European Council for Nuclear Research (CERN) [9] near Geneva, Switzerland. In particular, it is the last element of the CERN's accelerator complex, shown schematically in Figure 2.1.

This complex of machines can accelerate beams of protons or ions, such as lead, using each ring as an injector for the following, and currently reaches energies up to 6.8 TeV. Particle acceleration happens thanks to radio-frequency (RF) cavities [3], which are metallic chambers containing an oscillating electromagnetic field. Because of the oscillating voltage, the force exerted on each particle depends on its arrival time in the cavity. As a consequence, when the beam has reached the desired energy, particles arriving with an ideal timing aren't accelerated, while particles arriving slightly earlier (later) are decelerated (accelerated) in order to be brought to the desired energy. The result of this process is a beam in which the particles are not uniformly distributed, but rather organized in groups, called bunches.

After being accelerated by the previous machines, the particles are injected in the LHC ring in two separate beam pipes, where they circulate in opposite directions while the beam energy is increased. The two counter-rotating beams are then brought to collision in four interaction points, and, since they are both travelling with the same energy, the total centre of mass energy $\sqrt{s}$ is twice the energy of each beam. Four detectors are placed in correspondence of the interaction points: ALICE [10], dedicated to heavy ion physics, LHCb [11], focused on the study of b quarks, and two multipurpose

detectors, built in order to observe the Higgs boson and now aiming to observe any kind of Beyond the Standard Model (BSM) physics: ATLAS [12] and CMS [13].

The first collisions at the LHC took place in 2009, at the beginning of its first operation period (Run 1), which lasted until 2013. During this phase, the machine operated at centre of mass energies of $\sqrt{s} = 7$ TeV and 8 TeV. After a two-year shutdown, during which the experiments could upgrade their technologies and infrastructure, the second running period of LHC began in 2015 and lasted until 2018, colliding particles at a centre of mass energy of $\sqrt{s} = 13$ TeV. This was followed by another shutdown period and, lastly, Run 3 commenced in 2022 and is expected to continue until 2026, performing collisions at $\sqrt{s} = 13.6$ TeV.

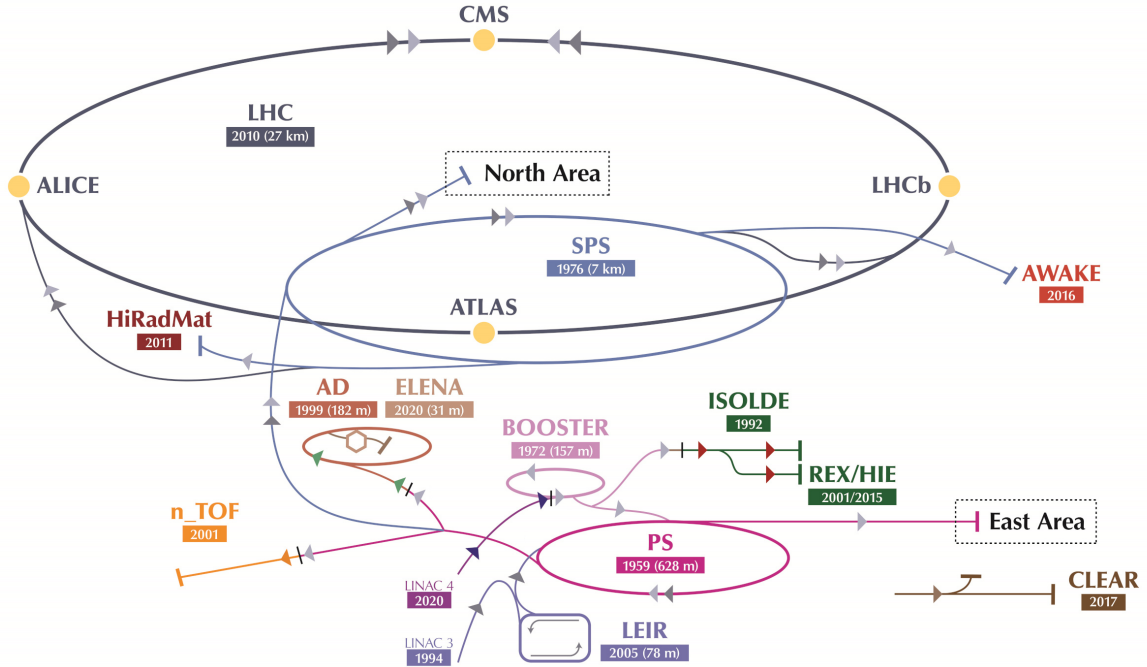


Figure 2.1: Schematic view of CERN's accelerator complex. Image taken from [9].

2.1.1 Luminosity and pile-up

When the proton bunches collide, several processes can happen, all with different probabilities. In order to verify whether the observation of a phenomenon is consistent with the expectations, it is necessary to consider how many events involving it are expected to be seen. The rate at which the process happens can be evaluated as:

$$\frac{dN}{dt} = \sigma \cdot L \cdot \epsilon \cdot A, \quad (2.1)$$

where N is the expected number of events involving the phenomenon under study, σ is its total cross-section [14], L is the instantaneous luminosity [15] of the collider, ϵ is the efficiency of the detector, *i.e.* its probability of detecting an event which has occurred, and A is the detector acceptance, which measures the range of kinematic parameters (such as η , defined below by Equation 2.3) within which particles are potentially detectable. While σ depends only on the physics entering the process of interest, the other three parameters are related to the detector and the collider configuration. In particular, the luminosity’s dependency on the beam properties is:

$$L = \frac{1}{4\pi} \cdot \frac{f_{\text{rev}} N_1 N_2 n_b}{\sigma_x \sigma_y}, \quad (2.2)$$

where f_{rev} is the beams’ revolution frequency, N_1 and N_2 the number of protons in each colliding bunch, n_b the number of bunches in each beam, and σ_x and σ_y the horizontal and vertical sizes of the beams, respectively¹.

Whenever the bunches collide, given the high amount of protons they contain², several interactions happen simultaneously. This leads to pile-up, *i.e.* the presence of unwanted extra collisions overlapping with the one of interest in the detector. As these collisions are not related to hard scatter events, they are treated as background and should therefore be mitigated. Pile-up is referred to as “in-time” if the additional interactions occur in the same bunch crossing as the collision of interest, while it is defined “out-of-time” if they happen in the bunches right before or after.

2.2 Overview of the ATLAS detector

ATLAS, dedicated to performing precision measurements of the parameters of the SM, as well as investigating BSM physics, is one of the two general-purpose detectors at the LHC.

The detector, shown in Figure 2.2, has a cylindrical shape, with a length of 44 m and a diameter of 25 m, and is centred in the interaction point. Since the particles produced in the collisions are emitted isotropically, ATLAS is a nearly 4π detector, *i.e.* it covers as much of the full solid angle as possible.

Its main subsystems, which are described more in detail below, are arranged in layers around the beam pipe. Closest to the interaction point there is the tracking system, which reconstructs the trajectory, charge and momentum of charged particles produced in the collisions. Subsequently, the energy of hadrons, electrons and photons is reconstructed in the calorimeter system, which comprises an electromagnetic calorimeter and a hadronic calorimeter. Then, the muon spectrometer reconstructs the energy and trajectory of muons. The detector also features two types of superconducting magnet systems: the

¹Equation 2.2 is derived in [15], and it is valid under the assumption of Gaussian beams.

²In both Run 2 and Run 3, the number of protons per bunch is around 10^{11} .

solenoid and the toroid magnet, which bend the trajectories of charged particles. Lastly, given the elevated number of interactions happening in the detector, a trigger system ensures that only the events of interest are stored.

The only particles able to pass through ATLAS undetected are neutrinos, which interact very weakly with matter. As a consequence, their presence can only be inferred from the transverse momentum of final state particles. Indeed, the transverse momentum when a collision happens is 0, and, because of momentum conservation, the transverse momenta of final state particles should sum up to the same value. When this doesn't happen, the additional momentum which would be needed to satisfy momentum conservation, referred to as missing transverse momentum (or missing transverse energy) E_T^{miss} [16], is an indicator of the presence of undetected final state particles.

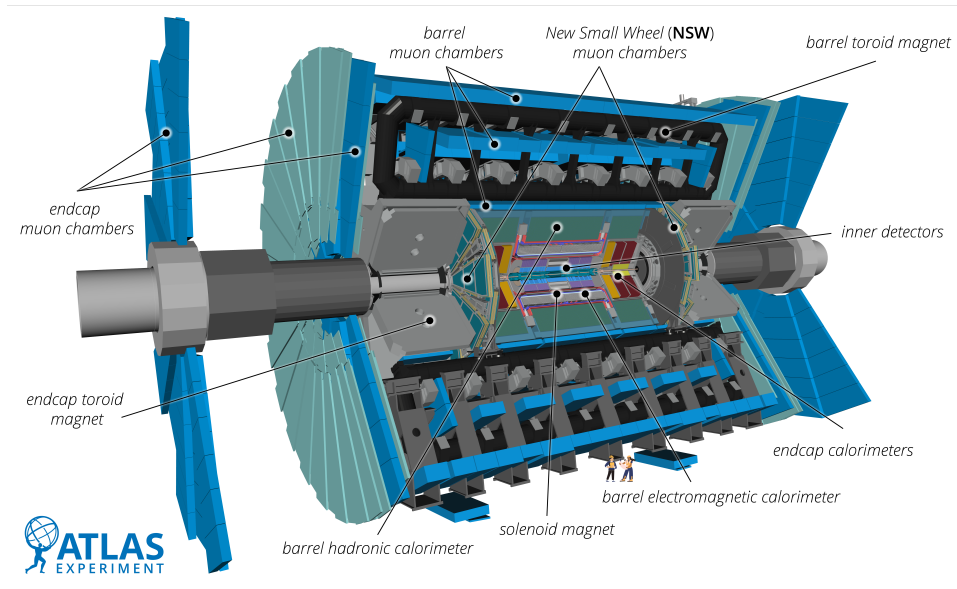


Figure 2.2: Sketch of the ATLAS detector and its main components in Run 3, taken from [9].

Coordinate System

The conventional coordinate system used in ATLAS, shown in Figure 2.3, is a right-handed frame of reference centred in the interaction point, with the z -axis pointing along the beam pipe, the y -axis pointing upwards and the x -axis pointing towards the centre of the LHC ring. The xy -plane is referred to as the transverse plane, thus the projection of momentum on it is known as transverse momentum p_T .

Positions in the xy -plane are typically described in terms of ϕ and η , where the former

is the azimuthal angle and the latter is the pseudo-rapidity, defined as:

$$\eta = -\ln \left(\tan \frac{\theta}{2} \right), \quad (2.3)$$

where θ is the polar angle between the momentum of the particle and the beam axis. The angular distance of two objects ΔR is thus defined as:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \quad (2.4)$$

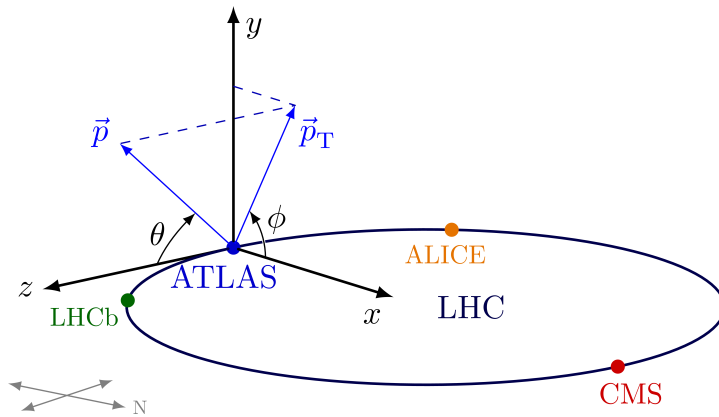


Figure 2.3: Coordinate system of the ATLAS experiment. Image generated with the code available at [17].

The Inner Detector

The Inner Detector is the innermost component of ATLAS, located just 3.3 cm around the beam pipe, and thus it is the first one to measure the collision decay products. In particular, it measures the hits of electrically charged particles, assessing their direction of motion. Moreover, it is immersed in a strong magnetic field which causes charged particles to move in a helical path with a radius r of:

$$r = \frac{mv_{\perp}}{qB}, \quad (2.5)$$

with m being the particle mass, $v_{\perp}$ its velocity in the direction perpendicular to the magnetic field B , and q its charge. Consequently, a precise measurement of the particle tracks can be used to reconstruct momentum and charge information.

The detector, depicted in Figure 2.4, comprises three subsystems: the Pixel Detector, featuring small and very precise silicon pixels, the Semiconductor Tracker, made of four layers of “micro-strips” of silicon sensors, and the Transition Radiation Tracker, composed of drift tubes which contain a gas mixture made mostly of xenon.

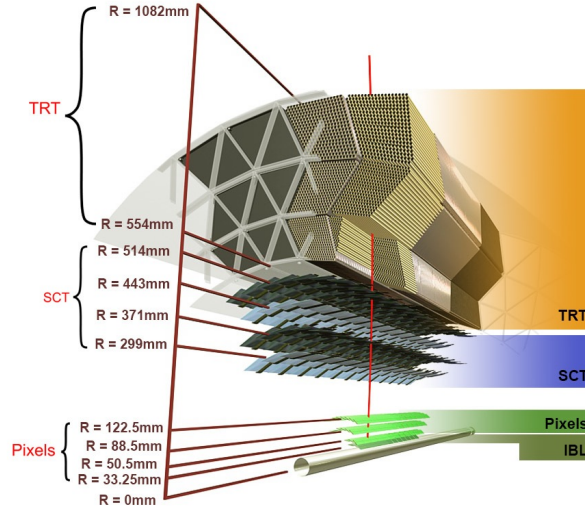


Figure 2.4: Sketch of the ATLAS inner detector, showing its components and their distances from the interaction point. Image taken from [18].

The Calorimeter

Located surrounding the tracking system is the ATLAS calorimeter. It is a sampling calorimeter, *i.e.* it consists of alternating layers of active material, dedicated to energy measurements, and layers of absorbing, or passive, material, meant to stop the incoming particles and produce showers of less energetic particles. Indeed, when a particle interacts with an atom in the calorimeter medium, it produces secondary particles having an energy lower than the primary one. These, in turn, interact with other atoms, in a process that repeats itself, generating a cascade of particles. The process stops when the particles' energy is low enough for them to be absorbed in the detector by ionization losses.

Schematically shown in Figure 2.5, the calorimeter system has three main components: a liquid-argon (LAr) electromagnetic calorimeter, designed to measure electrons, photons and hadrons, and two types of hadronic calorimeters (LAr hadronic calorimeter and Tile calorimeter), aiming to measure the particles that are not fully absorbed in the previous subsystem. As their name suggests, LAr calorimeters use liquid argon as their active material, while layers of metal are used as passive material. The Tile calorimeter, on the other hand, consists of layers of steel (used as passive material) and of plastic (used as active material) scintillating fibres.

The differences between the electromagnetic and the hadronic calorimeters arise from the different ways in which various types of particles generate showers. Indeed, when electrons (or positrons) and photons interact with the detector medium, the main processes happening are, respectively, bremsstrahlung emission, which consists of the emission of photons by a decelerating charged particle, and pair production, which occurs when a

photon decays into an electron-positron pair. As a consequence, when a shower is initiated by an electron, positron or photon, it will (mostly) contain only these three types of particles. On the other hand, hadrons also interact via the strong force, and they generate showers by undergoing inelastic interactions with the nuclei of the medium, producing mostly other hadrons. The length scale of this process is higher than that of the electromagnetic interactions described above, hence it is necessary to have a thicker calorimeter in order to absorb the shower completely.

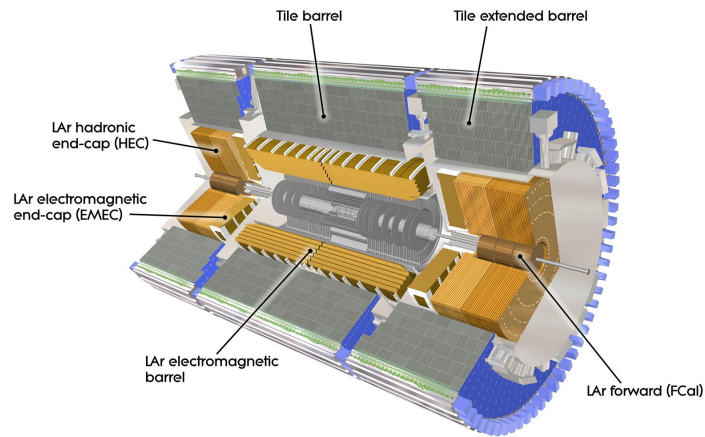


Figure 2.5: Sketch of ATLAS calorimeter system, showing its main components. Image taken from [12].

The Muon Spectrometer

Together with neutrinos, muons are the only known particles which aren't stopped by the calorimeter system. In fact, they are minimum ionizing particles, meaning that only a small fraction of their energy is deposited in the calorimeter. Hence, muons reach the outermost part of the ATLAS detector, which is dedicated to the measurement of their momenta and tracks. As shown in Figure 2.6, the Muon Spectrometer uses different detection technologies, which can be classified into two categories: the precision detectors, providing an accurate measurement of the muons tracks, and the fast-response detectors, performing a fast event selection.

Magnet System

As previously mentioned, ATLAS uses a combination of toroid and solenoid magnets, the layout of which is shown in Figure 2.7, in order to bend the trajectories of charged particles. In particular, the Inner Detector is surrounded by the Central Solenoid Magnet,

which provides a 2 T magnetic field in the core part of the experiment. Then, there are three toroids, each consisting of a series of 8 coils, able to generate a field up to 4 T, which is used to measure the momentum of muons. Two of such magnets are positioned at the ends of the experiment, while the Barrel Toroid surrounds the whole detector.

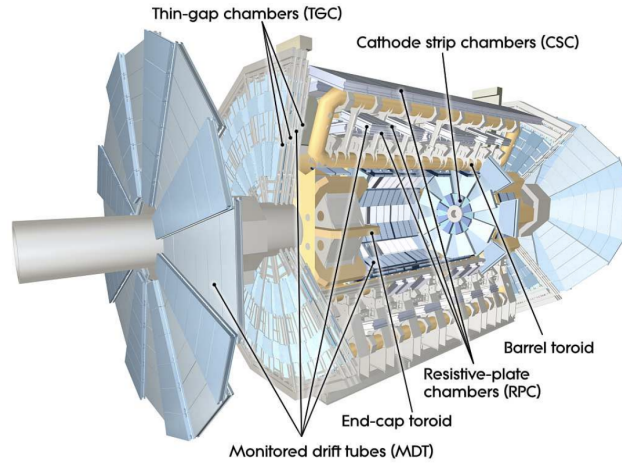


Figure 2.6: Sketch of the muon spectrometer, showing its main components. Image taken from [12].

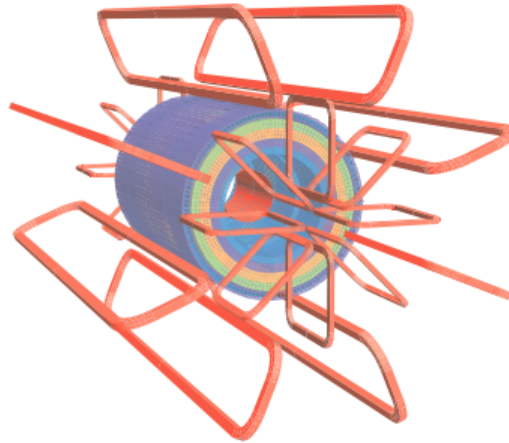


Figure 2.7: Schematic representation of the ATLAS magnet system, showing the central solenoid (purple, green and orange cylinders) and the interleaved coils of the barrel and end-cap toroids (depicted in red, the longest coils are the ones corresponding to the barrel magnet). Image taken from [12].

Trigger System

The Trigger and Data Acquisition system is responsible for ensuring that only interesting collision events are recorded for studies. Indeed, ATLAS observes nearly 2 billion proton-proton collisions every second, corresponding to a data rate of millions of megabytes per second. However, only a subset of these collisions contain processes of interest, hence the trigger applies an event selection in order to record only useful data.

The trigger system is organized in two steps. Firstly, there is a hardware-level trigger (Level-1 trigger), which uses part of the information from the calorimeter and the muon spectrometer in order to perform an initial cut-based event selection. Events passing this selection are then transferred to the software-based High-Level trigger, which can accept events at a rate of 100 kHz and analyse them by using the full detector information from specific detector regions, identified by the previous trigger subsystem. After this step, the event rate is reduced to 1 kHz and the data is transferred to a storage system, making it available for offline analysis.

Chapter 3

Topoclusters in Jet Reconstruction

3.1 Jet Reconstruction Algorithms

As discussed in Section 1.2, quarks and gluons produced in a proton-proton collision are observed as jets, *i.e.* collimated sprays of particles resulting from the fragmentation and hadronization of a high-momentum parton. Accurate jet reconstruction is thus crucial for understanding the kinematics of the originating partons.

There are two main types of objects which can be used as input for jet reconstruction [19]: topological clusters (topoclusters) and particle-flow (PFlow) objects. Topoclusters are sets of adjacent cells in the calorimeter showing a non-negligible energy signal, and thus they are constructed using only calorimeter information. Conversely, PFlow objects are a combination of calorimeter- and track-based measurements, built attempting to match each track to a topocluster.

The key idea behind a reconstruction algorithm is to group together particles that are close to each other in order to reconstruct the particle which generated the jet. When doing so, two criteria should be fulfilled: infrared and collinear safety. The former requirement ensures the stability of the set of jets if a low- p_T gluon is added to the event, while the latter secures the original set of jets against a small angle splitting. Several algorithms can be built satisfying these requirements [20], but only the one used in ATLAS will be described here.

Anti- k_T Algorithm

The anti- k_T algorithm [21] is a sequential clustering algorithm which iterates over all input particles to group them based on their mutual distance, taking into account their momentum information. In particular, two distances are computed for each pair of input objects i and j :

$$d_{ij} = \min \left(\frac{1}{p_{T,i}^2}, \frac{1}{p_{T,j}^2} \right) \frac{\Delta R_{ij}^2}{R^2}, \quad (3.1)$$

$$d_{iB} = \frac{1}{p_{T,i}^2}, \quad (3.2)$$

where $p_{T,i}$ is the transverse momentum of the i^{th} object, ΔR_{ij}^2 is the angular distance between the two objects, and R is the radius parameter of the algorithm, which determines the size of the jets it produces. Typically, $R = 0.4$ and $R = 1.0$ are used for small- and large- R jets, respectively.

The objects are then treated according to which of the two distances is smaller. If d_{ij} is smaller than d_{iB} , the two objects are combined according to a recombination scheme and the process is repeated. On the contrary, in the opposite case, the object i is called a jet and removed from the list of inputs. A typical choice for recombination is the E -scheme, in which the 4-vectors of the jet constituents are summed up.

3.2 Formation of Calorimeter Topoclusters

A possible choice for the input objects to use in jet reconstruction are clusters of topologically close signal cells in the calorimeter. The main observable controlling the formation of these clusters is the cell significance $\varsigma_{\text{cell}}^{\text{EM}}$, defined as:

$$\varsigma_{\text{cell}}^{\text{EM}} = \frac{E_{\text{cell}}^{\text{EM}}}{\sigma_{\text{noise, cell}}^{\text{EM}}}, \quad (3.3)$$

where $E_{\text{cell}}^{\text{EM}}$ is the energy measured in the cell at electromagnetic scale, *i.e.* without any calibration, and $\sigma_{\text{noise, cell}}^{\text{EM}}$ the average expected noise in the cell.

The topoclustering process [22] starts from a cell with a high significance and iteratively collects its neighbouring cells showing a signal. In order to do so, three thresholds are defined:

$$|\varsigma_{\text{cell}}^{\text{EM}}| > S \quad \text{primary seed threshold (typically } S = 4), \quad (3.4)$$

$$|\varsigma_{\text{cell}}^{\text{EM}}| > N \quad \text{threshold for growth control (typically } N = 2), \quad (3.5)$$

$$|\varsigma_{\text{cell}}^{\text{EM}}| > P \quad \text{principal cell filter (typically } P = 0). \quad (3.6)$$

Cells above the threshold in Equation 3.4 are referred to as seeds and form a proto-cluster each. Then, all the cells adjacent to a seed that satisfy Equation 3.6 are collected into the same proto-cluster. If a cell added in this way satisfies Equation 3.5 too, its neighbouring cells are also added to the proto-cluster and, if it belongs to another proto-cluster, the two are merged. This process is repeated iteratively until all the neighbouring cells satisfying Equation 3.6 are collected.

The choice to consider the absolute value of the significance for the topoclustering thresholds is made in order to improve noise reduction. In fact, out-of-time pile-up might

result in cells with $E_{\text{cell}}^{\text{EM}} < 0$ due to the negative tail of the LAr calorimeter pulse shape, shown in Figure 3.1.

If the signal in one of such cells with negative energy is high enough, it can produce a cluster with $E_{\text{cluster}}^{\text{EM}} < 0$. These clusters should not be neglected when evaluating full event observables, such as the missing transverse momentum E_T^{miss} , as they cancel out (on average) the contributions from out-of-time clusters with positive energy. Analogously, the same cancellation happens for the cells having out-of-time pile-up contributions within the same topocluster. As a consequence, only the absolute significance of the cells is used in topocluster formation, regardless of its sign.

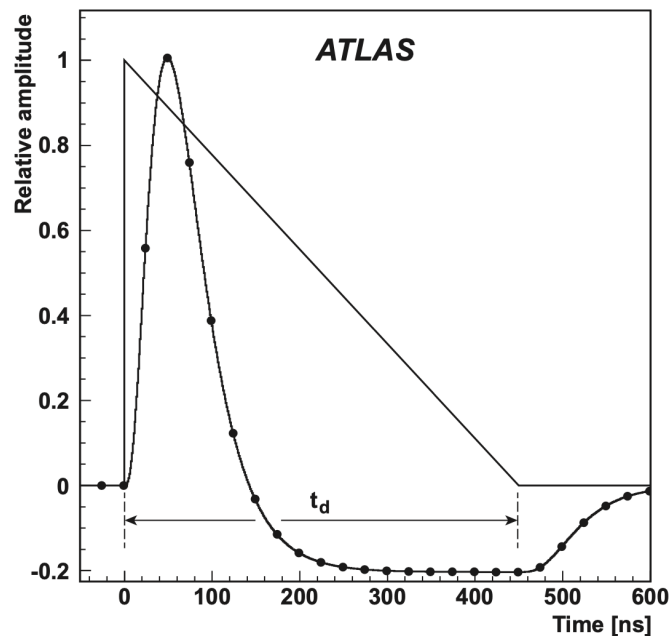


Figure 3.1: Example of the pulse shape of the LAr (central EM barrel) calorimeter. Image taken from [23].

3.2.1 Cluster Energy Calibration

When a particle deposits its energy E^{dep} in the calorimeter, several factors can degrade the detector signal, resulting in a measured energy $E_{\text{clus}}^{\text{EM}}$ different from E^{dep} . For instance, not all the calorimeter material is active, and thus some energy can be lost in inactive materials. Energy losses may also happen during the topoclustering processes, depending on how the noise thresholds are defined. In addition, ATLAS calorimeters are non-compensating, *i.e.* their yield for hadrons is lower than that for electrons and

photons of the same energy. Finally, the presence of pile-up contributions in the cluster can lead to higher measured energies.

LCW Calibration

The current way to address this issue in ATLAS is the Local Hadronic Cell Weighting (LCW) calibration [22], outlined in Figure 3.2.

After the topoclusters are formed, their probability of being generated by an electromagnetic shower $\mathcal{P}_{\text{clus}}^{\text{EM}}$ is evaluated, since the corrections applied should be higher if the energy deposit is of hadronic nature. Based on this probability, an overall weight $w_{\text{cell}}^{\text{cal}}$ is assigned to each cell:

$$w_{\text{cell}}^{\text{cal}} = \mathcal{P}_{\text{clus}}^{\text{EM}} \cdot w_{\text{cell}}^{\text{em-cal}} + (1 - \mathcal{P}_{\text{clus}}^{\text{EM}}) \cdot w_{\text{cell}}^{\text{had-cal}}, \quad (3.7)$$

where $w_{\text{cell}}^{\text{em-cal}}$ and $w_{\text{cell}}^{\text{had-cal}}$ are the weights applied to the cell signal by the electromagnetic and hadronic calibrations, respectively. Finally, the last two calibration steps correct for the energy lost in the clustering process and in inactive materials.

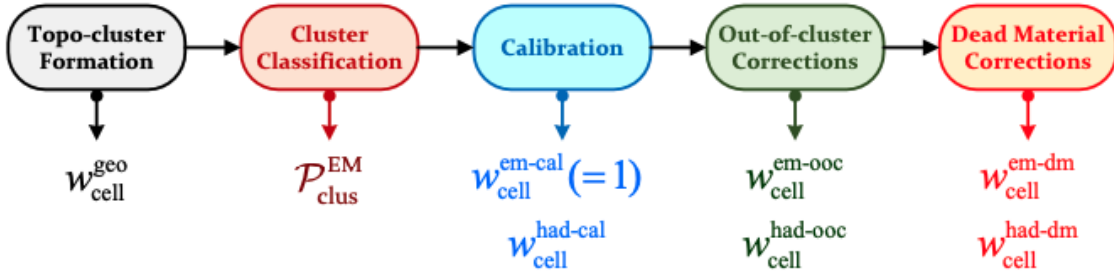


Figure 3.2: Overview of the LCW calibration scheme for topoclusters, from [22].

Machine Learning Based Calibration

An alternative way to perform energy calibration is to use machine learning techniques, which can provide insight into the correlation among cluster features and determine a smooth multidimensional calibration function. Recent studies [24] have shown the effectiveness of this method as an alternative to LCW, and in particular the improvements it provides in the calibration of low energy clusters. The calibration networks target the topocluster response, defined as:

$$R_{\text{clus}}^{\text{EM}} = \frac{E_{\text{clus}}^{\text{EM}}}{E_{\text{dep}}}, \quad (3.8)$$

which can be seen as the inverse of a weight to apply to the cluster. The input features used cover a wide set of topocluster variables, ranging from its kinematical and topological properties, to signal strength and timing, to shower characteristics and, lastly, including information on the pile-up environment.

3.3 Pile-up Mitigation in Jet Reconstruction Inputs

As anticipated in Section 2.1.1, multiple interactions happen simultaneously whenever the beams are brought to collision. Consequently, the measured energy of a topocluster $E_{\text{clus}}^{\text{EM}}$ can include contributions from pile-up, which should be mitigated in order to correctly reconstruct the energy of the hard scatter particles under study.

Firstly, it is straightforward that the use of PFlow objects already provides some degree of pile-up suppression. Indeed, as these objects are created considering tracks, clusters associated with tracks coming from secondary vertices can be removed. However, in order to have a matching track, a cluster should have originated from an electrically charged particle, which is not always the case. Thus, other methods of pile-up mitigation, which don't rely on track matching, are used too. In the following, the jet reconstruction inputs will be considered to be topoclusters. However, these methods also apply to PFlow objects.

Cut on Cell Timing in Run 3

As stated previously, pile-up can originate either from the same bunch crossing of the collision of interest (in-time pile-up) or from events outside the 25 ns bunch crossing window associated with the collision (out-of-time pile-up). In order to limit the contributions from out-of-time pile-up, a cut on cell timing has been used in ATLAS starting from Run 3 [23]. In particular, a cell satisfying the seed condition in Equation 3.4 can only be a topocluster seed, or be included in a growing cluster, if its timing t_{cell} satisfies:

$$|t_{\text{cell}}| < 12.5 \text{ ns.} \quad (3.9)$$

While this cut ensures that out-of-time pile-up is reduced, it also risks suppressing signals which could arise from BSM physics, such as the ones originating from the decays of slow-moving long-lived particles [25]. To prevent this from happening, another condition, referred to as the Upper Limit, is introduced. It ensures that the time cut is not applied if the cell signal has a high and positive significance, *i.e.* if it satisfies:

$$\varsigma_{\text{cell}}^{\text{EM}} > X_{\text{UL}}, \quad (3.10)$$

where X_{UL} is a given threshold, chosen to be $X_{\text{UL}} = 20$ in ATLAS topocluster reconstruction.

SoftKiller

The SoftKiller method [26] consists of applying an event-based p_T cut to input objects. Each event is divided into patches in the η - ϕ plane, and its pile-up density ρ is defined as the median of p_T flow density across all patches:

$$\rho = \text{median}_{i \in \text{patches}} \left\{ \frac{p_{T,i}}{A_i} \right\}, \quad (3.11)$$

where $p_{T,i}$ and A_i are the transverse momentum and area of the i^{th} patch. Then, a momentum threshold p_T^{cut} is defined by choosing the minimal cut for which $\rho = 0$. This is equivalent to requiring that half of the patches are empty after the method is applied, as shown in Figure 3.3.

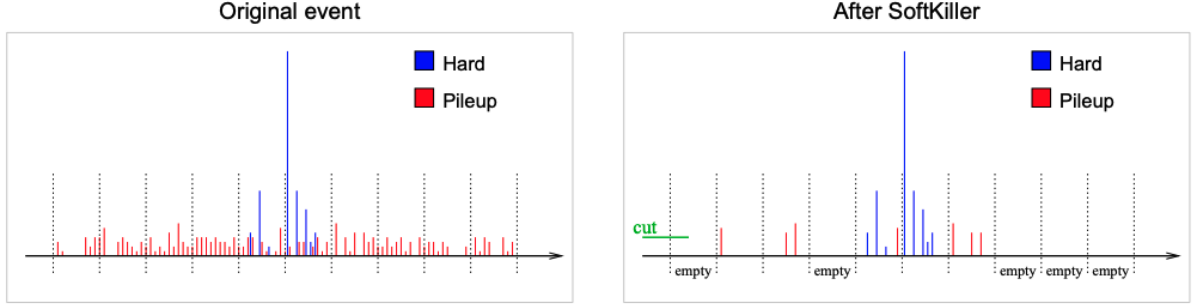


Figure 3.3: Particles in an event before (left) and after (right) SoftKiller is applied, showing the momentum threshold p_T^{cut} in green. The vertical dashed lines represent the edges of the patches in which the event was divided to evaluate ρ . Image from [26].

Constituent Subtraction

The Constituent Subtraction algorithm [27], typically applied before the SoftKiller, is an area-subtraction method which alters the inputs' momentum. After evaluating the pile-up density ρ according to Equation 3.11, massless particles called ghosts are added to the event with a uniform distribution. The ghost momentum p_T^g is determined as:

$$p_T^g = A^g \cdot \rho, \quad (3.12)$$

where A^g is the area of the ghost in the η - ϕ plane, set to 0.01 by default. Then, for each pair of cluster i and ghost k , the distance $\Delta R_{i,k}$ is computed with Equation 2.4. The pairs (i,k) are subsequently sorted in order of ascending distance and undergo an iterative process of pile-up suppression. At each step, the momenta of the cluster and

the ghost are modified as follows:

$$\begin{aligned}
\text{If } p_{T,i} \geq p_{T,k} : \quad & p_{T,i} \rightarrow p_{T,i} - p_{T,k}, \\
& p_{T,k} \rightarrow 0; \\
\text{otherwise :} \quad & p_{T,i} \rightarrow 0, \\
& p_{T,k} \rightarrow p_{T,k} - p_{T,i}.
\end{aligned} \tag{3.13}$$

This process continues until at least one $\Delta R_{i,k}$ is below a chosen value ΔR_{max} , and all the ghosts remaining after the subtraction are removed from the event.

Pile-Up Per Particle Identification

A last method that can be used for pile-up mitigation in jet inputs is Pile-Up Per Particle Identification [28], which estimates whether inputs originate from hard scatter or pile-up based on their proximity to hard scatter products. In particular, each object i is assigned a probability α_i of coming from the primary vertex, evaluated as:

$$\alpha_i = \log \left(\sum_j \frac{p_{T,j}}{\Delta R_{i,j}} \cdot \Theta(R_{\min} \leq \Delta R_{i,j} \leq R_{\max}) \right), \tag{3.14}$$

where the index j runs along the topoclusters with a matching track connecting them to the primary vertex, $R_{\min/\max}$ is the minimum/maximum distance at which two objects can be matched and Θ is the Heaviside step function. Ideally, pile-up clusters should have $\alpha \sim 0$, while hard scatter clusters $\alpha \sim 1$. Subsequently, the distribution of α is evaluated for all the objects coming from charged pile-up particles, identified thanks to their matching tracks. Denoting the mean and RMS of this distribution as $\bar{\alpha}_{PU}$ and σ_{PU} , respectively, the following quantity is evaluated for each neutral cluster:

$$\chi_i^2 = \Theta(\alpha_i - \bar{\alpha}_{PU}) \frac{\alpha_i - \bar{\alpha}_{PU}}{\sigma_{PU}^2}. \tag{3.15}$$

Finally, the four-momentum of each particle is assigned a weight given by:

$$w_i = F_{\chi^2}(\chi_i^2), \tag{3.16}$$

where F_{χ^2} denotes the cumulative distribution of the χ^2 distribution with one degree of freedom. In this way, the neutral inputs with $\alpha_i < \bar{\alpha}_{PU}$ are removed from the event. Further noise suppression can then be obtained by applying a cut on the momentum based on the number of reconstructed primary vertices N_{PV} :

$$p_T^{\text{cut}} < a + b \cdot N_{PV}, \tag{3.17}$$

where a and b are user-defined parameters.

While this method could in principle be applied to topoclusters, it is optimized for PFlow objects, as it relies on track matching for hard scatter charged particles, and is therefore not used elsewhere.

Chapter 4

Analysis Methods

4.1 Key Idea: Suppression via Topocluster Classification

The aim of this work is to suppress pile-up contributions in calorimeter topoclusters by classifying them into three categories:

- pile-up-only clusters, where all the energy comes from pile-up,
- hard scatter-only clusters, in which all the energy comes from the hard scatter,
- mixed clusters, having contributions from both hard scatter and pile-up.

Upon such classification, topoclusters that are unlikely to contain hard scatter contributions can be removed from the jet reconstruction inputs, while the mixed topoclusters can be calibrated differently from the hard scatter-only ones, to account for the effect of pile-up.

Two Deep Neural Networks (DNNs) [29] have been implemented to assess the nature of topoclusters: one which identifies hard scatter-only clusters, and one that does the same for pile-up-only clusters. The performance of the classification has been assessed both by using machine learning-related features of merit, described in Section 4.2.2, and by evaluating the improvements it provides to the cluster response (see Equation 3.8 for its definition).

4.1.1 Datasets

The studies presented here are performed separately for Run 2 and Run 3, using data from simulated di-jet events.

In particular, the simulated events are generated using PYTHIA 8 [30], with the NNPDF2.3 LO parton distribution functions [31], the Lund hadronization model [32]

and the ATLAS A14 set of tune parameters [33]. Detector effects are included by running the simulated events through the ATLAS detector simulation [34], based on the GEANT 4 toolkit [35].

The data used includes two simulations for each LHC Run: one simulating, for each event, only the hard scatter interaction and one including pile-up. The second set of simulations is obtained by overlaying the events with additional proton-proton collisions, generated with a different set of tuning parameters (A3 [36]). The number of pile-up vertices in this step is weighted to match the average number of interactions per bunch crossing $\langle\mu\rangle$ observed in data. In particular, the simulation used for Run 2 (MC20e) corresponds to 2018 data, for which $\langle\mu\rangle \sim 36$, as shown in Figure 4.1 (left), while the one for Run 3 (MC23d) corresponds to 2023 data, with $\langle\mu\rangle \sim 51$, as shown in Figure 4.1 (right).

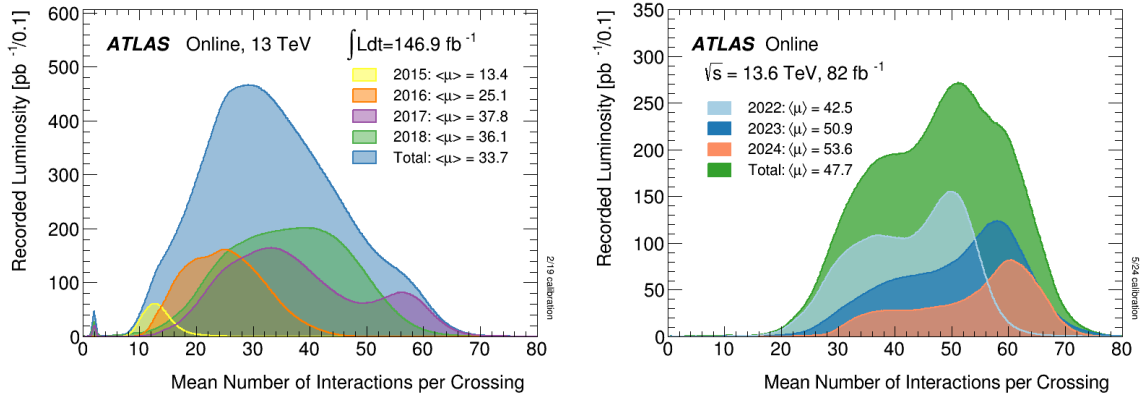


Figure 4.1: Luminosity-weighted distribution of the mean number of interactions per bunch crossing in ATLAS for Run 2 (left, from [37]) and Run 3 (right, from [38]).

Features

Since the classification implemented here is meant to be performed before jet reconstruction, only the topocluster features uncorrelated to jets can be used. In particular, the input features have been chosen, with the procedure described in Section 4.2.3, among the ones listed in Table 4.1. This set was obtained by considering all the topocluster features that are less than 75% correlated with the others, since the presence of correlated input features increases the computational resources needed without providing additional information.

Variable name	Feature Name	Description
cluster_n_cells_tot	n_{cells}	Number of cells in the cluster
clusterE	$E_{\text{clus}}^{\text{EM}}$	Cluster signal at the electromagnetic scale
cluster_time	t_{clus}	Cluster timing
cluster_SECOND_TIME	σ_t^2	Second moment of the cell timing distribution
cluster_EM_PROBABILITY	$\mathcal{P}_{\text{clus}}^{\text{EM}}$	Probability for the cluster to be generated by an electromagnetic shower
cluster_ENG_FRAC_EM	f_{emc}	Fraction of energy deposited in the electromagnetic calorimeter
cluster_FIRST_ENG_DENS	$\langle \rho_{\text{cell}} \rangle$	Energy-weighted first moment of cell signal density
cluster_CENTER_MAG	c	Distance of centre-of-gravity from nominal vertex
cluster_CENTER_LAMBDA	λ_{clus}	Distance of centre-of-gravity from calorimeter front face
cluster_LATERAL	$\langle m_{\text{lat}}^2 \rangle$	Energy dispersion perpendicular to the main cluster axis
cluster_LONGITUDINAL	$\langle m_{\text{long}}^2 \rangle$	Energy dispersion parallel to the main cluster axis
cluster_ISOLATION	f_{iso}	Measure of cluster isolation
cluster_AVG_LAR_Q	$\langle Q_{\text{LAR}} \rangle$	Average signal quality in the LAr calorimeter [39]
cluster_AVG_TILE_Q	$\langle Q_{\text{Tile}} \rangle$	Average signal quality in the Tile calorimeter [40]
cluster_CELL_SIGNIFICANCE	$\max\{s_{\text{clus}}^{\text{EM}}\}$	Highest cell signal significance in cluster
clusterEta	y_{clus}	Cluster rapidity
clusterPhi	ϕ_{clus}	Cluster azimuth
cluster_DELTA_ALPHA	$\Delta\alpha$	Angular distance between the principal shower axis and its direction from the nominal vertex
cluster_DELTA_THETA	$\Delta\theta$	Polar angle distance of the principal shower axis with respect to its direction from the nominal vertex
cluster_DELTA_PHI	$\Delta\phi$	Azimuthal distance of the principal shower axis with respect to its direction from the nominal vertex
cluster_SECOND_R	$\langle r_{\text{cell}}^2 \rangle$	Second moment of the radial distances of the cells to the principal cluster axis
cluster_SECOND_LAMBDA	$\langle \lambda_{\text{cell}}^2 \rangle$	Second moment of the distances of cells to from the cluster centre along its main axis

Table 4.1: Topocluster features considered for the pile-up suppression studies. More information on the definition of the variables can be found in [22].

4.2 Deep Neural Networks

The results presented in Chapter 5 were obtained using two fully connected DNNs implemented in the Keras software suite [41] using the Tensorflow [42] backend.

The first network, discussed in Section 5.2, distinguishes hard scatter-only clusters from mixed clusters, thus providing insight into how pile-up contributions influence the

cluster response. On the other hand, the network presented in Section 5.3 distinguishes pile-up-only clusters from clusters belonging to the other two classes, which are merged into the same label. This classification can be used to identify clusters that don't have hard scatter contributions and thus remove them from the inputs for jet reconstruction.

4.2.1 Network Architecture

For simplicity, the two networks have the same architecture, consisting of 7 hidden layers with a decreasing number of nodes ($512 \rightarrow 256 \rightarrow 128 \rightarrow 64 \rightarrow 32 \rightarrow 16 \rightarrow 8$), each followed by a batch normalization layer that rescales the data so that the input to the next layer is centred around 0 and has a standard deviation of 1. Batch normalization is employed as a way of preventing overfitting [43], a phenomenon which occurs when a network fits too closely to the training data, losing its generalization capabilities.

Another method used to avoid overfitting is the inclusion of dropout layers, here inserted after the fourth and fifth hidden layers. Their role is to randomly drop 30% of the nodes in the following layer at each training step, thus preventing the network from relying too heavily on any specific node.

All the hidden layers use a Rectified Linear Unit activation function:

$$f(x) = \max\{0, x\}, \quad (4.1)$$

where x is the input in each node and $f(x)$ its output. After them, there is an output layer consisting of only one node, as the task of each network is to perform a binary classification. This layer uses a sigmoid activation function:

$$f(x) = \frac{1}{1 + e^{-x}}, \quad (4.2)$$

and its output is the probability that a given topocluster belongs to the class labelled as "1" in the network.

Loss Function

A common choice for the loss function for binary classification tasks is binary cross-entropy:

$$\mathcal{L}(y, p) = -\frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)], \quad (4.3)$$

where i runs on the clusters in the test dataset, N is the total number of clusters, y_i is the i^{th} cluster's true label (either 0 or 1), and p_i is its probability of belonging to class 1.

Since the datasets may have unbalanced populations, meaning that more clusters belong to one class than the other, a weighted version of Equation 4.3 is used here. In

particular, each term of the sum is multiplied by a weight w_j , defined as:

$$w_j = \frac{N}{2 \cdot N_j}, \quad (4.4)$$

where $j = 0, 1$ is the true label of the cluster and N_j is the number of clusters belonging to class j . In this way, the network is prevented from being biased towards the more populated class.

Optimization

To optimize its convergence in terms of time and robustness, the network employs the Adam optimizer [44], which adjusts the learning rate using the first and second momenta of the loss function’s gradient.

In addition, the training time is optimized using the early stopping technique [45]. This method monitors the network loss on the validation data, *i.e.* a dataset that is evaluated after each training epoch to check how the model performs on new data. Training is interrupted when this loss does not decrease for a user-defined number of epochs, set to 20 for these studies. Early stopping prevents the network from overfitting, as the training is interrupted as soon as the performance on the validation set does not improve.

The remaining hyperparameters, *i.e.* the batch size and the initial learning rate, are chosen to ensure convergence and stability of the network. The selected values, which are different for the two networks and the data used (Run 2 or Run 3), are shown in Table 4.2.

Network task	Data	Learning rate	Batch size
Hard scatter-only identification	Run 2	$1 \cdot 10^{-3}$	512
	Run 3	$1 \cdot 10^{-5}$	1024
Pile-up-only identification	Run 2	$1 \cdot 10^{-6}$	128
	Run 3	$1 \cdot 10^{-6}$	128

Table 4.2: Learning rate and batch size used for each network and each dataset.

4.2.2 Figures of Merit

Once trained, the network can be applied to new data, classifying them by imposing a threshold on the network output. In this way, all the clusters with a probability above the threshold can be assigned to class 1 (denoted “positive” in the following to match the notation used in literature), while the others will fall into class 0 (“negative”).

Information about the classifier’s performance for a given threshold can be obtained by constructing the corresponding confusion matrix. Schematically shown in Figure 4.2, it is a table illustrating four quantities: the number of true positive/negative clusters (TP and TN, respectively), *i.e* the clusters belonging to class 1/0 which have been correctly classified, and the number of false positive/negative ones (FP and FN), which have been misclassified.

		Predicted Class	
		Positive	Negative
True Class	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

Figure 4.2: Representation of the information contained in a confusion matrix for a binary classifier.

However, the network’s performance can also be assessed independently of the chosen threshold. A method to represent it is the Receiver Operating Characteristics (ROC) curve [46]. This probability curve is built by plotting, for different values of the threshold, the True Positive Rate (TPR), also called recall R or sensitivity, against the False Positive Rate (FPR), where the two variables can assume values between 0 and 1 and are defined as:

$$\text{TPR} = R = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad \text{FPR} = \frac{\text{FP}}{\text{TN} + \text{FP}}. \quad (4.5)$$

An ideal classifier is therefore expected to have a ROC curve which reaches the upper left corner of the corresponding space, whereas one that performs random guesses would result in a diagonal line, as shown in Figure 4.3. As a consequence, the closer the Area Under the Curve (AUC) is to unity, its maximum possible value, the better the classifier performs. Moreover, the ROC curve itself can be used to choose a probability threshold, since the optimal network performance is obtained in correspondence of the upper left corner of the curve.

Another possible way to visualize the network performance, useful for unbalanced datasets, is the Precision-Recall (PR) curve. Similarly to the ROC curve, it is constructed

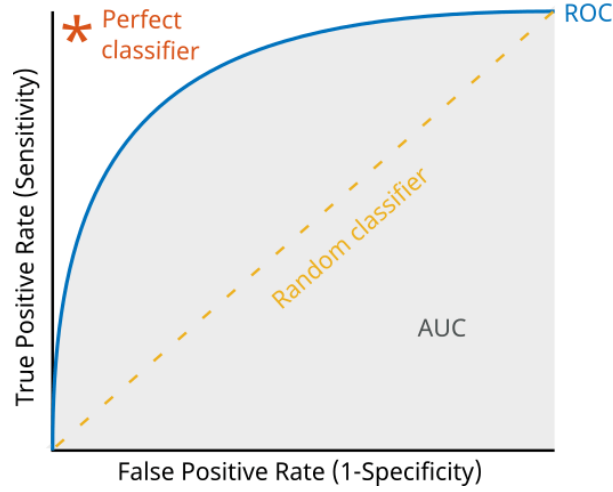


Figure 4.3: Example of a ROC curve, showing (dotted line) how it would look like for a random classifier. Image taken from [47].

by varying the classifier threshold, and plotting the precision P :

$$P = \frac{TP}{TP + FP}, \quad (4.6)$$

against the recall R , defined by Equation 4.5. Together with its AUC, a possible parameter to summarize this curve is the average precision A_P :

$$A_P = \sum_n (R_n - R_{n-1}) P_n, \quad (4.7)$$

where R_n and P_n are the recall and precision at the n^{th} threshold and n runs on all the thresholds used to build the curve.

4.2.3 Input Features Choice

As anticipated in Section 4.1.1, the input features for the networks are chosen not to be highly correlated, thereby reducing the computational resources required. Moreover, studies performed using a first version of the network for hard scatter-only recognition and a subset of the Run 2 simulation allowed a further reduction in the number of input features.

Indeed, not all the features considered are a priori useful for the network. Hence, quantifying their importance provides insight into whether they can be removed from the input features set. Such quantification was performed using the permutation feature importance method. With this algorithm, the importance i_j of each feature j is evaluated

as:

$$i_j = s - \frac{1}{K} \sum_{k=1}^K s_{k,j}, \quad (4.8)$$

where s is the score of the network (in this case the ROC-AUC) on the test data, $s_{k,j}$ is the score obtained evaluating the data after randomly shuffling the values of feature j , and K is the number of times the shuffling and evaluation should be performed, set to 10 for these studies. If a feature is highly important for the network, shuffling it would degrade the performance, resulting in a high i_j . Conversely, shuffling a feature which does not bring much information would cause minimal changes in the score, leading to a small i_j .

Eight of the features listed in Table 4.1, shown below the dashed line in Table 4.3, were found to have importances below 1% for different network architectures tried, and thus they were not used for any further studies. Consequently, the number of input features is 14 for the hard scatter-only identification network and 13 for the pile-up-only identification one. In fact, the latter excludes the cluster energy $E_{\text{clus}}^{\text{EM}}$ as an input, since it uses it to define the cluster labels, as will be discussed in Section 5.3.

Feature Normalization

To limit the numerical range of the network inputs, the selected features have been normalized with an approach analogous to the one described in [24]. In particular, each feature x , apart from t_{clus} , is transformed as:

$$v(y) = \frac{\langle y \rangle - y}{\text{std}(y)}, \quad (4.9)$$

where $y = \log_{10}(x)$ or $y = x$ depending on whether x has a smoothly falling spectrum or not. The cluster timing, on the other hand, is normalized according to:

$$y(t_{\text{clus}}) = \begin{cases} -\sqrt[3]{|t_{\text{clus}}|} & \text{if } t_{\text{clus}} < 0 \\ \sqrt[3]{t_{\text{clus}}} & \text{otherwise} \end{cases}, \quad (4.10)$$

$$v(y) = \frac{y - \langle y \rangle}{\text{std}(y)}. \quad (4.11)$$

The normalization mode used for each variable is shown in Table 4.3, where “logarithmic” refers to the case $y = \log_{10}(x)$, and “standard” to the case $y = x$.

Feature	Normalization
clusterE	Logarithmic
cluster_FIRST_ENG_DENS	Logarithmic
cluster_EM_PROBABILITY	Standard
cluster_CENTER_LAMBDA	Logarithmic
cluster_CENTER_MAG	Standard
cluster_n_cells_tot	Logarithmic
cluster_ENG_FRAC_EM	Standard
cluster_SECOND_TIME	Logarithmic
cluster_AVG_TILE_Q	Logarithmic
cluster_AVG_LAR_Q	Logarithmic
cluster_SECOND_R	Logarithmic
cluster_LATERAL	Standard
cluster_time	Equations 4.10, 4.11
cluster_ISOLATION	Standard
cluster_DELTA_ALPHA	Logarithmic
cluster_DELTA_THETA	Standard
clusterEta	Standard
cluster_DELTA_PHI	Standard
clusterPhi	Standard
cluster_SECOND_LAMBDA	Logarithmic
cluster_LONGITUDINAL	Standard
cluster_CELL_SIGNIFICANCE	Logarithmic

Table 4.3: Normalization mode used for each network input feature. Features below the dashed line have been removed from the inputs after studying their importance.

Chapter 5

Results

5.1 Preliminary Studies on Topocluster Response

The first results presented here aim to highlight the motivations for pile-up suppression by assessing which cluster variables it can improve, showing the importance of the cluster classification studies outlined in Section 4.1.

Specifically, a cluster energy calibration network, built with the strategy described in [24] and applied to MC20e data, was used to study the cluster response dependency on pile-up. As anticipated in Section 3.2.1, the network targets the cluster response $R_{\text{clus}}^{\text{EM}}$, corresponding to the ratio between the measured, uncalibrated, energy and the deposited true energy. Ideally, this variable's distribution should be narrow, as this would imply that the true energy can be obtained with small uncertainty from the measured one by multiplying it by a calibration factor (corresponding to the centre of the distribution). However, the presence of pile-up implies a measured energy higher than that deposited by hard scatter particles, resulting in a response distribution with a long tail towards high $R_{\text{clus}}^{\text{EM}}$. This behaviour can be observed in Figure 5.1, where the reconstructed response is plotted in three bins of the number of primary vertices in the event N_{PV} . As shown in the figure, the width of the distribution increases with a higher pile-up, *i.e.* for higher N_{PV} , negatively affecting the calibration of clusters from events with several pile-up interactions.

Figure 5.2 shows how the distance between the mean and the median of the response distribution increases as a function of N_{PV} , illustrating how the response worsens in high pile-up conditions. Indeed, the skewness of the distribution can be visualized by considering how much these two variables differ, as a higher mean (with respect to the median) denotes a distribution with a more pronounced right tail. Since pile-up contributions typically have low energies, this behaviour is expected to be seen mostly in clusters with small $E_{\text{clus}}^{\text{EM}}$. To verify this assumption, the mean and median response for various N_{PV} are shown separately for clusters with energy lower (left) or higher (right)

than 100 GeV (Figure 5.3). As expected, for high-energy clusters the two values remain close to each other, while for the low-energy ones the mean increases considerably with N_{PV} .

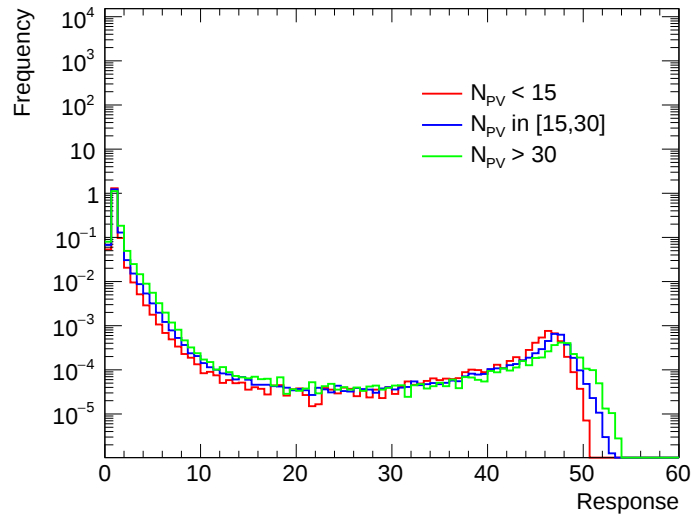


Figure 5.1: Reconstructed cluster response for three ranges of N_{PV} , corresponding to events with low (red), medium (blue) and high (green) pile-up.

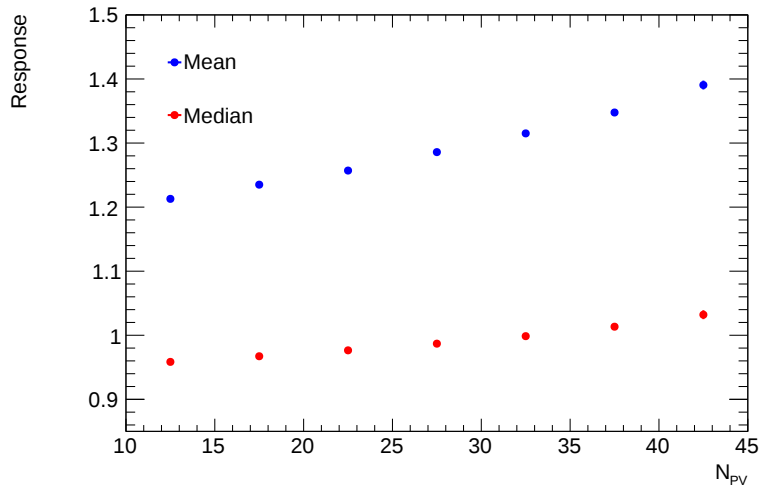


Figure 5.2: Mean and median of the predicted response distribution in bins of N_{PV} . The x coordinates of each point represent the central value of the corresponding bin.

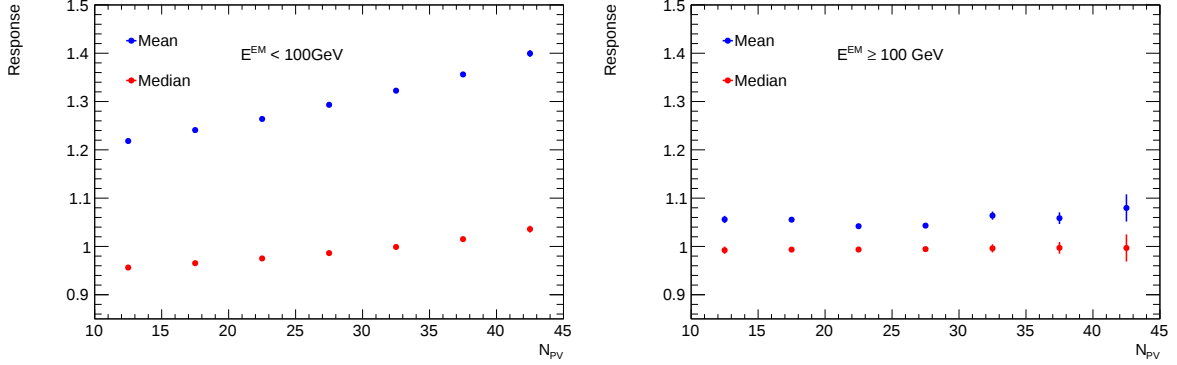


Figure 5.3: Mean and median of the predicted response distribution in bins of N_{PV} , for clusters with E_{clus}^{EM} below (left) and above (right) 100 GeV. The x coordinates of each point represent the central value of the corresponding bin.

In summary, these preliminary studies have shown how pile-up suppression can improve the topoclusters' energy calibration, as it positively affects the cluster response. In addition, they illustrated why pile-up suppression is particularly important for low energy clusters, as their response is more sensitive to the presence of contributions unrelated to the hard scatter interaction.

5.2 Hard Scatter-Only Clusters Identification

To assess the nature of the contributions to topoclusters, classifying them into the three categories described in Section 4.1, two networks were constructed: one to identify hard scatter-only clusters and one to identify pile-up-only clusters. The former, whose results are presented in this section, distinguishes hard scatter-only clusters from mixed clusters, assigning to each topocluster a probability to contain only contributions originating from the primary interaction.

Understanding to which of the two classes the clusters belong, *i.e.* defining their truth labels, is non-trivial, since the Monte Carlo simulations lack information on the origin of the energy contributions in the clusters. For this task, class labels were determined by considering clusters generated from two simulations: one simulating only hard scatter interactions, and the other including pile-up. Clusters from the former are directly classified as hard scatter-only. On the other hand, it is not possible to assess certainly whether a cluster in the full simulation contains pile-up contributions. This leads to the risk of misclassifying some clusters, including them in the mixed class while they may contain only hard scatter contributions. To try to limit this problem, the class is formed by considering only the clusters from events with an average number of interactions per bunch crossing $\langle \mu \rangle > 20$, as the presence of more pile-up increases the likelihood of it

contributing to the topoclusters. The distribution of $\langle\mu\rangle$ in the two datasets, plotted in Figure 5.4, shows that imposing the condition $\langle\mu\rangle > 20$ does not significantly reduce the data size, as most of the events have a higher number of simultaneous interactions.

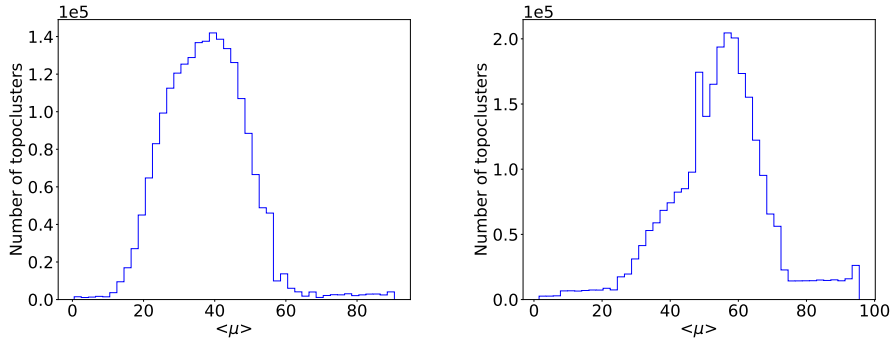


Figure 5.4: Average number of interactions per bunch crossing $\langle\mu\rangle$ for the Run 2 (left) and Run 3 (right) simulations. It can be seen that most of the clusters have $\langle\mu\rangle > 20$.

The number of clusters used for training, for each dataset, is shown in Table 5.1 and corresponds to 80% of the total number of clusters. Following the selection made in [24], only clusters with true energies $E^{\text{dep}} > 300$ MeV are considered. Moreover, only 75% of the training clusters are used for this purpose, while the remaining fraction is reserved as validation data. This allows for the evaluation of the network performance on previously unseen clusters, which is used to adjust the training duration.

Data	Hard scatter-only clusters	Mixed clusters
Run 2	900140	805429
Run 3	1017813	1104661

Table 5.1: Number of clusters, divided by class, used for the network’s training on Run 2 and Run 3 simulations. 25% of the clusters in the training set are used as validation data.

5.2.1 Classification Results on Run 2

The hard scatter identification network, whose characteristics and architecture are described in Section 4.2, was first applied to classify Run 2 data. Its losses on the training and validation sets are shown in Figure 5.5. It can be noticed how the training process is interrupted by early stopping when the validation loss stabilizes, although the loss on

the training set continues to show a decreasing trend. As described earlier, this criterion for training interruption is used as a method to avoid overfitting the training data.

The network output, depicted in Figure 5.6, shows a neat separation between the two classes for both the training and the test data, implying that hard scatter-only clusters can be identified with high precision. Indeed, the network’s ROC-AUC and average precision, shown with the corresponding curves in Figure 5.7, both assume values close to unity (respectively 0.962 and 0.966 on the test dataset).

After finding an optimal value for the classification threshold $t = 0.43$ from the ROC curve, the cluster response in the test data $R_{\text{clus}}^{\text{EM}}$ for clusters with an output score above the threshold can be compared to that of the whole dataset, as shown in Figure 5.8. The figure also shows the interquartile range (IQR) of the two distributions, *i.e.* the distance between the first and third quartiles, containing 50% of the data. It can be seen that, after selecting only clusters with a high probability of being hard scatter-only, the IQR decreases, indicating a smaller data spread. This difference points out the need to take into account the presence of pile-up contributions in topoclusters when calibrating their energies, as it affects the response.

Lastly, the feature permutation importance is depicted in Figure 5.9 and can provide information on which cluster features are more sensitive to the presence of pile-up.

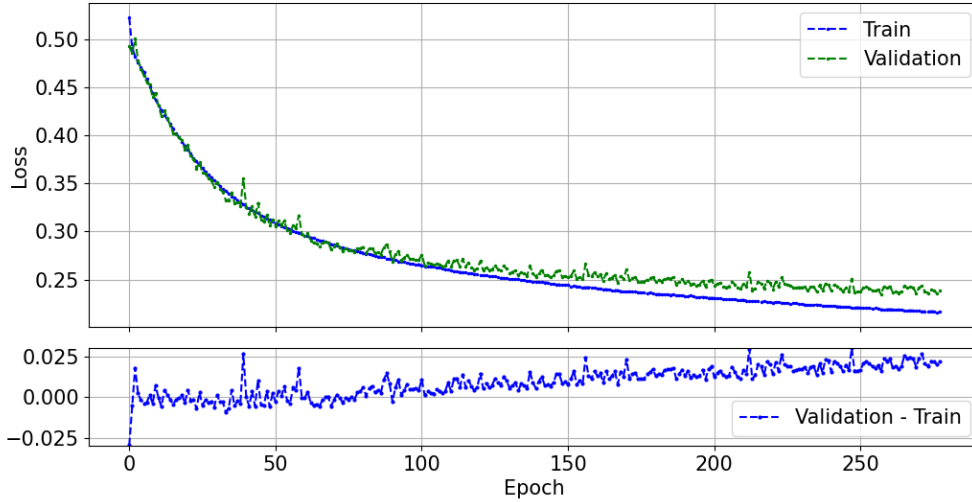


Figure 5.5: Network loss on the training (blue) and validation (green) datasets (above) for each training epoch. Difference between the two curves (below).

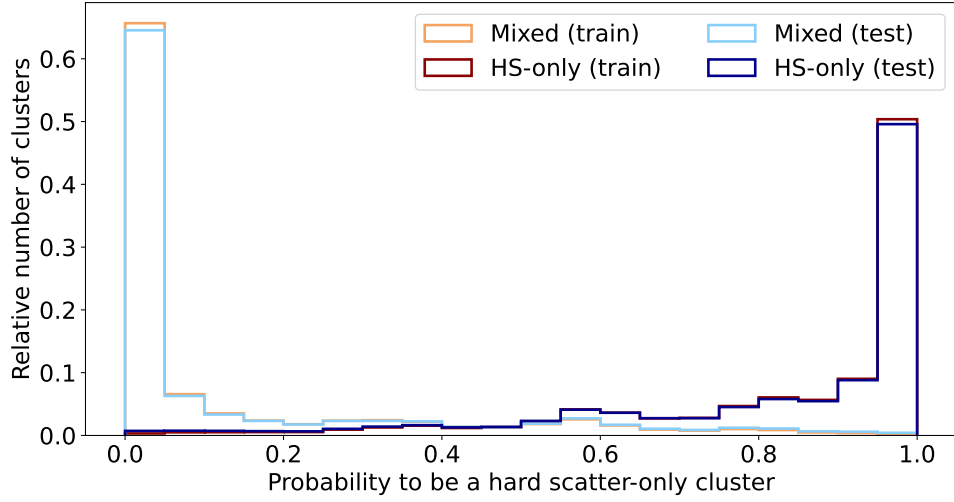


Figure 5.6: Network output on the train (light orange and red) and test (light blue and dark blue) data. The probability of being a hard scatter-only cluster is shown for mixed clusters (light orange and light blue) and hard scatter-only clusters (red and dark blue).

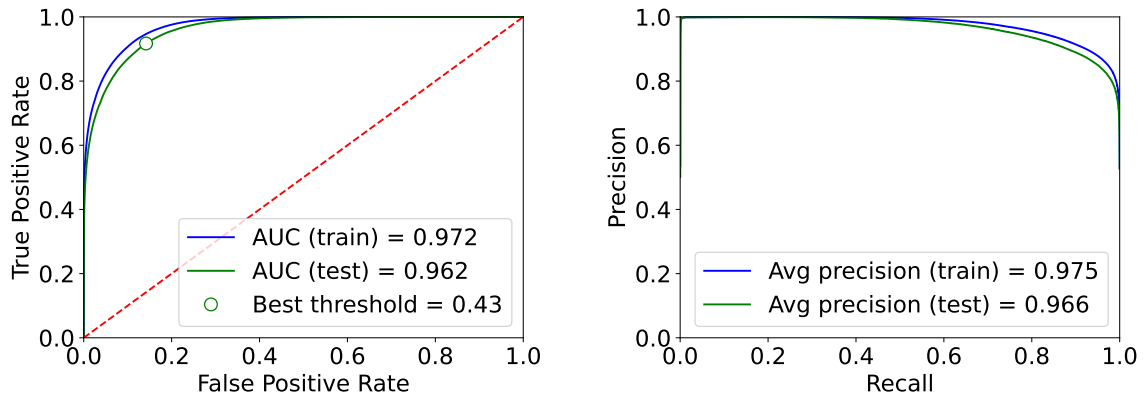


Figure 5.7: ROC (left) and PR (right) curves of the network on the train (blue) and test (green) data. The legends also show the ROC-AUC and average precision on both datasets. Moreover, the optimal threshold for classification, found as the upper corner of the train data's ROC curve, is shown in the legend on the left.

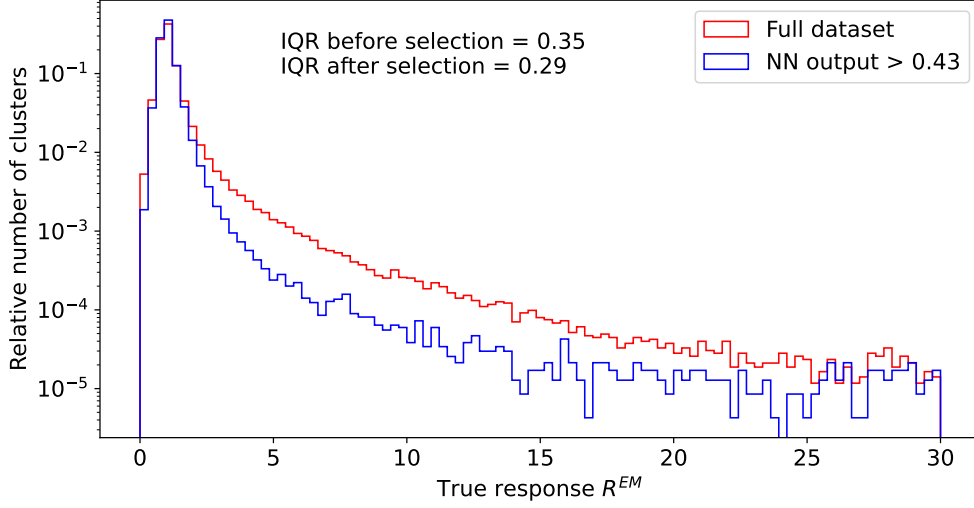


Figure 5.8: Topoclusters response R_{clus}^{EM} for the entire test dataset (red), and for only the clusters with $p > 0.43$ (blue), classified as hard scatter-only. The IQR before and after the selection (*i.e.* for the red and blue curve, respectively) is shown as a measure of the distributions' width.

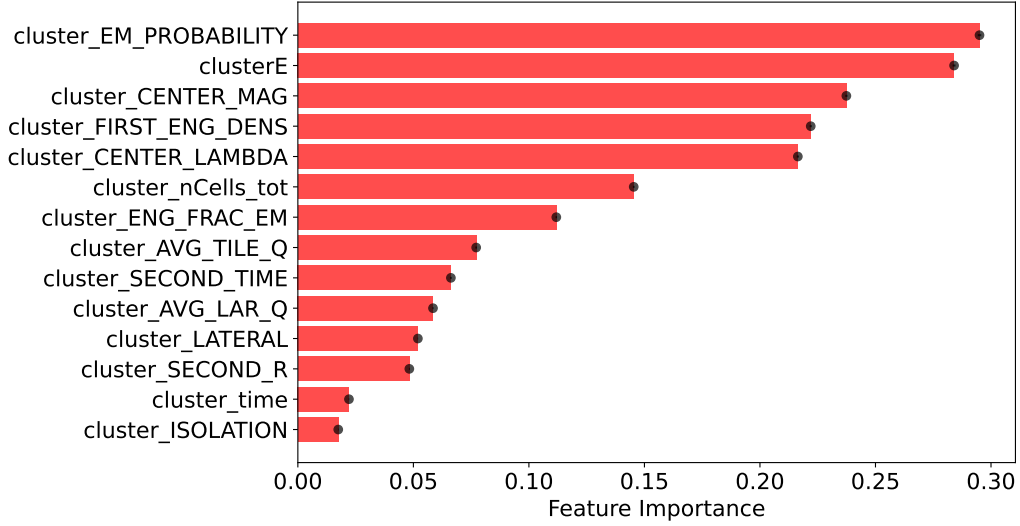


Figure 5.9: Permutation importance of each input feature (see Table 4.1 for the correspondence between variable and feature names), evaluated using Equation 4.8 with $K = 10$.

While the importance of the features for this classification is related to their pile-up sensitivity, it is not possible to draw further conclusions on the subject without having information on the amount of pile-up contributions in mixed topoclusters. Although not explored in these studies, an alternative way of defining the cluster features could be to use the DigiTruth method [48]. This would provide the cluster moments evaluated only with hard scatter contributions, allowing inference of, for example, the fraction of pile-up energy in each topocluster, and possibly improving the classification and the cluster response.

5.2.2 Classification Results on Run 3

After assessing its performance on clusters from the Run 2 Monte Carlo simulation, the network was again trained and tested on clusters from the Run 3 simulation. The corresponding training and validation losses are shown in Figure 5.10. Analogously to the previous case, when the training is interrupted the loss on the training set still shows a decreasing trend, while the one on the validation set is flat.

The output of the network is plotted, for the training and test datasets, in Figure 5.11. While the clear separation between the two classes indicates a good discriminative power, it is much lower than that observed in Figure 5.6 for Run 2 clusters. This is also reflected in the performance curves and metrics, shown in Figure 5.12. The ROC-AUC and average precision scores on the test set are 0.883 and 0.861, respectively, indicating that the network effectively discriminates between the two classes. However, they are smaller than those obtained on the Run 2 dataset. A possible reason for this discrepancy is the different way in which pile-up is treated in the two sets of simulations. Indeed, as mentioned in Section 3.3, a cut on the cell timing was introduced in Run 3 to suppress contributions from out-of-time pile-up. This additional pile-up suppression could result in more resemblances between mixed and hard scatter-only clusters in Run 3. The dissimilar behaviour of the network on the two datasets is also indicated by the different importance of the topocluster features, shown in Figure 5.13 for Run 3. Indeed, the reliance of the network on different features in this case compared to the previous one may be caused by the inherent differences in the data, particularly in clusters containing pile-up contributions.

Despite its performance being worse with respect to the one on Run 2, the network still shows a strong discriminative power on Run 3 data. Therefore, as before, its output was used to study the behaviour of the cluster response $R_{\text{clus}}^{\text{EM}}$ of the test data in the different classes. As shown in Figure 5.14, the IQR decreases if only clusters with an output score $p > 0.51$ are selected, in agreement with the expectations.

In summary, it has been shown how, despite the better pile-up suppression in Run 3, it is still possible to obtain a machine-learning-based distinction between mixed clusters and hard scatter-only clusters. Since the cluster response $R_{\text{clus}}^{\text{EM}}$ is worse for clusters belonging to the former class, this distinction could provide useful information for the

calibration of the measured topocluster energy.

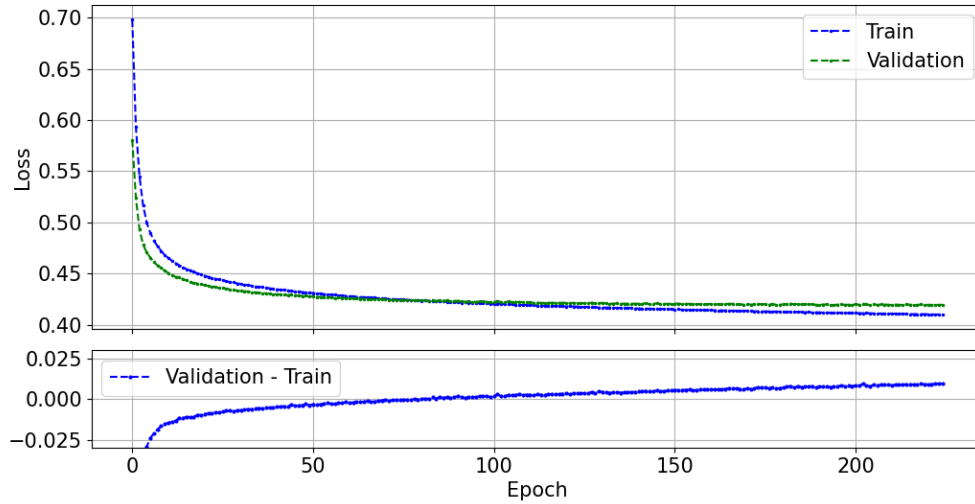


Figure 5.10: Network loss on the training (blue) and validation (green) datasets (above) for each training epoch. Difference between the two curves (below).

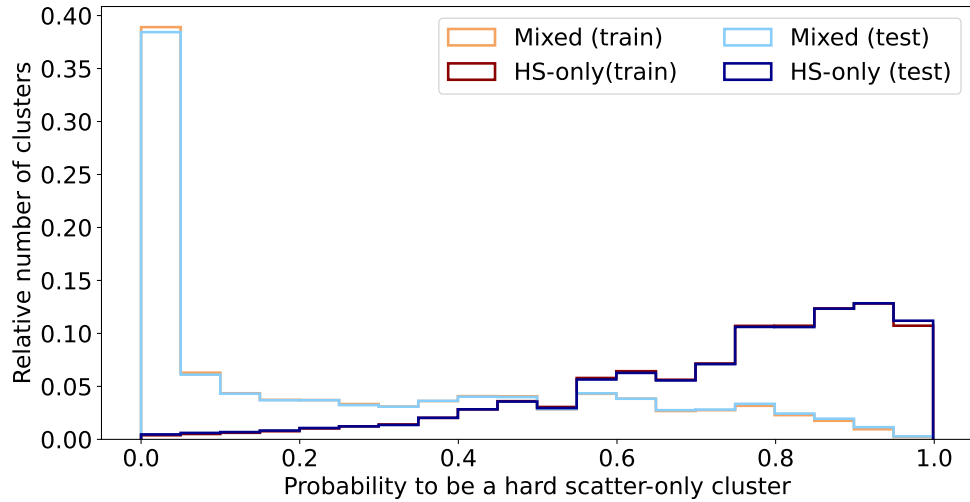


Figure 5.11: Network output on the train (light orange and red) and test (light blue and dark blue) data. The probability of being a hard scatter-only cluster is shown for mixed clusters (light orange and light blue) and hard scatter-only clusters (red and dark blue).

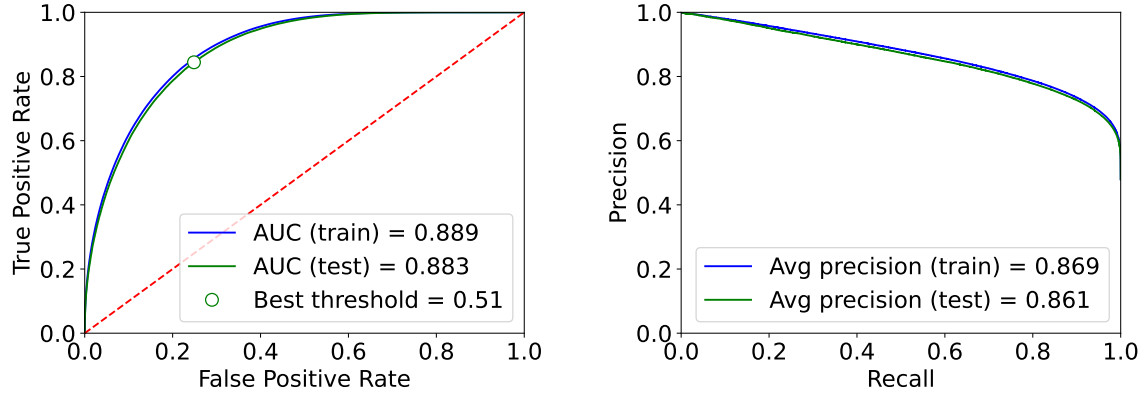


Figure 5.12: ROC (left) and PR (right) curves of the network on the train (blue) and test (green) data. The legends also show the ROC-AUC and average precision on both datasets. Moreover, the optimal threshold for classification, found as the upper corner of the train data's ROC curve, is shown in the legend on the left.

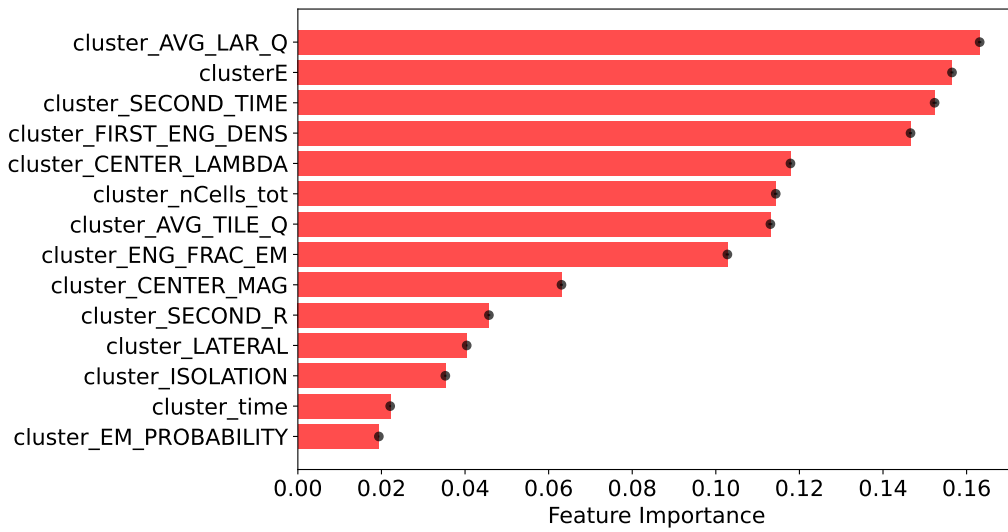


Figure 5.13: Permutation importance of each input feature (see Table 4.1 for the correspondence between variable and feature names), evaluated using Equation 4.8 with $K = 10$.

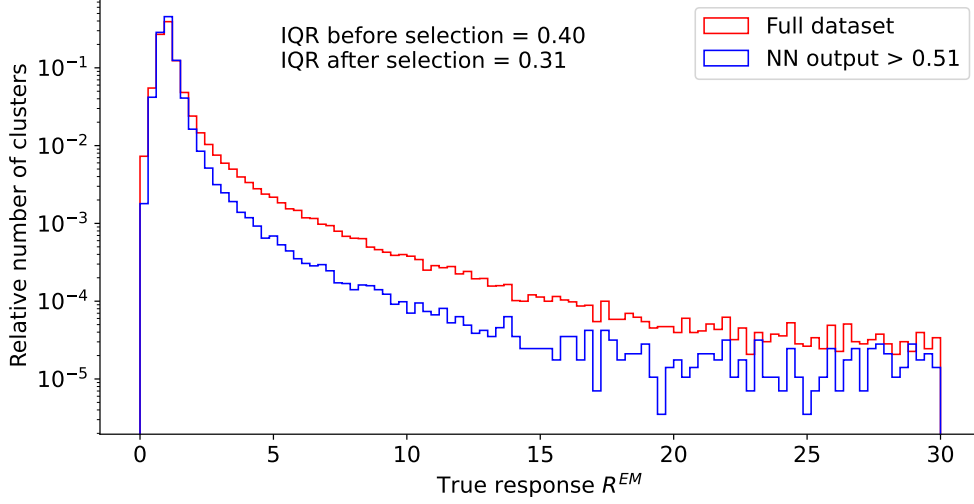


Figure 5.14: Topoclusters response $R_{\text{clus}}^{\text{EM}}$ for the entire test dataset (red), and for just the clusters with $p > 0.51$ (blue), classified as hard scatter-only. The IQR before and after the selection (*i.e.* for the red and blue curve, respectively) is shown as a measure of the distributions' width.

5.3 Pile-Up-Only Clusters Identification

As described in Section 4.1, a second network was implemented to identify pile-up-only clusters against mixed clusters. After such identification, the topoclusters with a high probability of containing only pile-up contributions could be removed from the jet inputs, improving their energy calibration.

The clusters used for these studies are generated with full simulations, that include both hard scatter and pile-up interactions. As shown in Section 5.2, the presence of pile-up leads to a high response $R_{\text{clus}}^{\text{EM}}$. Moreover, since the deposited energy E^{dep} only depends on the hard scatter products, it is expected to be very low in pile-up-only clusters. With these motivations, the pile-up-only class was defined as the set of clusters having:

$$E^{\text{dep}} < 1 \text{ MeV} \quad \text{and} \quad R_{\text{clus}}^{\text{EM}} > 4, \quad (5.1)$$

while the mixed cluster class is defined to include all the clusters which don't satisfy Equation 5.1. In addition, mixed clusters are considered here to be more hard scatter-like or pile-up-like depending on whether their response is below or above 4, respectively.

To verify the correctness of using Equation 5.1 to define pile-up-only clusters, it was checked how many clusters satisfied it in a simulation including only hard scatter interactions. Indeed, a priori there could also be hard scatter-only clusters with small deposited energies or high responses, and they should not be labelled as pile-up-only.

As expected, only a small fraction (in the order of 10^{-5}) of hard scatter-only clusters met the condition, indicating that the class definition prevents it from including hard scatter-only clusters.

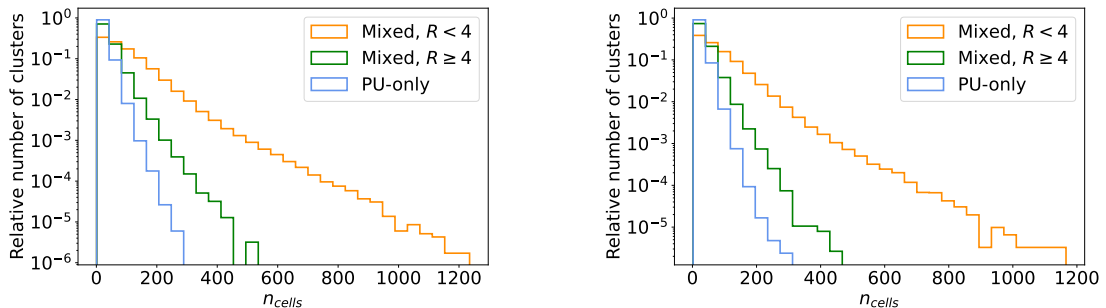
Table 5.2 shows the number of clusters used to train the network on Run 2 and Run 3 data, which corresponds to 80% of the total number of clusters. As with the hard scatter-only identification network, only 75% of the training clusters are used for this purpose, with the remainder reserved for validation. Lastly, the remaining clusters were used for testing. Since most clusters in each simulation contained both pile-up and hard scatter, the mixed class was downsampled to 3% of its original size. This resulted in the two classes having closer sizes, as shown in the table.

	Pile-up-only clusters	Mixed clusters
Run 2	668467	1184781
Run 3	1009156	1036820

Table 5.2: Number of clusters, divided by class, used for the network’s training on Run 2 and Run 3 simulations. 25% of the clusters in the training set are used as validation data.

The input features used by the network are shown, for clusters in Run 2 (left) and Run 3 (right) simulations, in Figure 5.15. In the figure, the feature distribution for the mixed class is shown separately for the two types of clusters it contains: the hard scatter-like ones, with $R_{\text{clus}}^{\text{EM}} < 4$, and the pile-up-like ones, with $R_{\text{clus}}^{\text{EM}} \geq 4$.

It can be noticed that there are some clear differences between the features of the hard scatter-like mixed clusters and the pile-up-only ones. On the other hand, this is not always the case when comparing pile-up-like mixed clusters to pile-up-only clusters. As a result, mixed clusters with pile-up-like characteristics are more difficult to classify, as will be seen in the discussion of the classification results below.



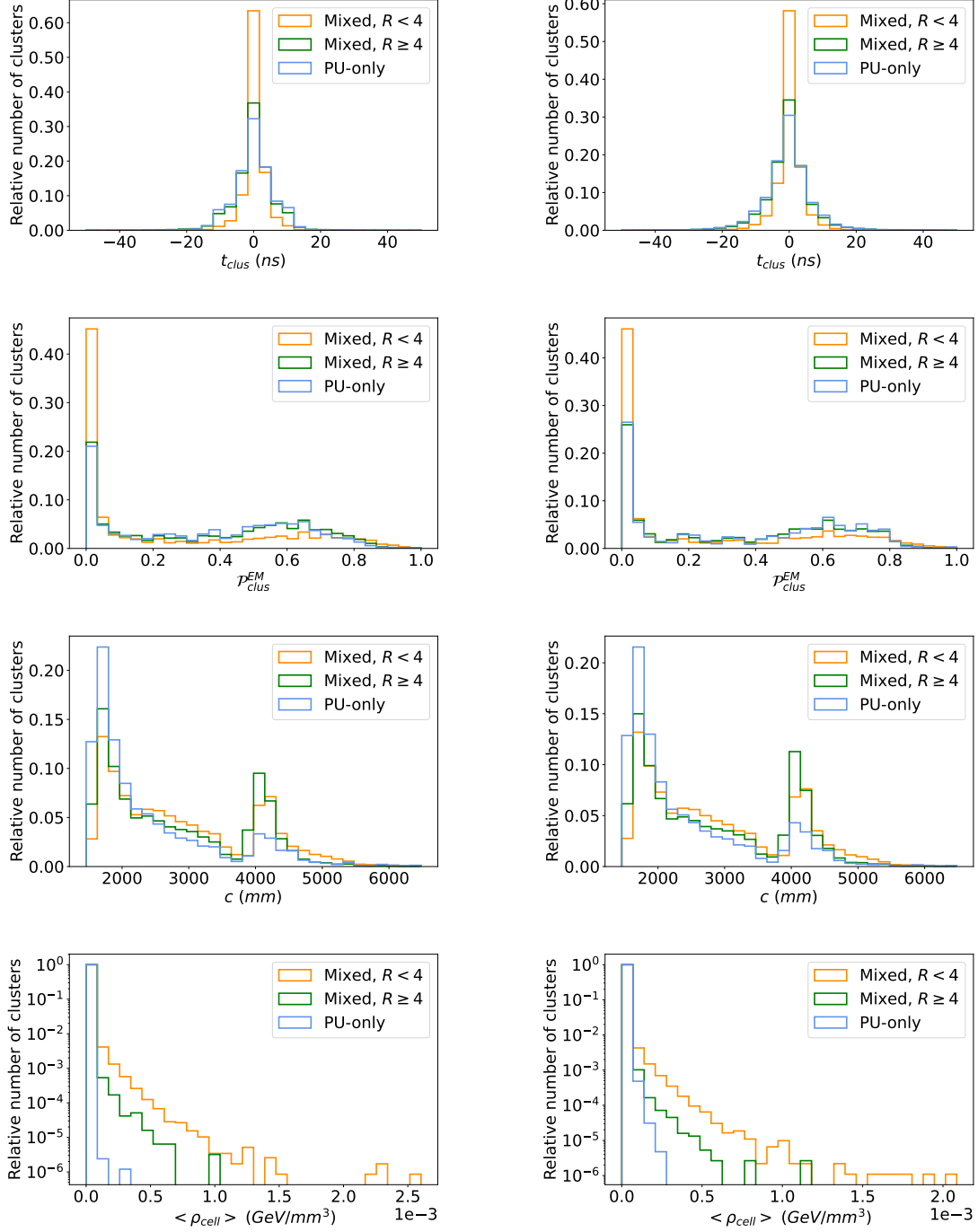


Figure 5.15: Distribution of the topocluster features used as network input, shown for pile-up-only clusters (light blue) and mixed clusters. The latter class is divided, based on R_{clus}^{EM} , into hard scatter-like clusters (orange) and pile-up-like clusters (green). The figures on the left represent the Run 2 Monte Carlo simulation, and the ones on the right the Run 3 one.

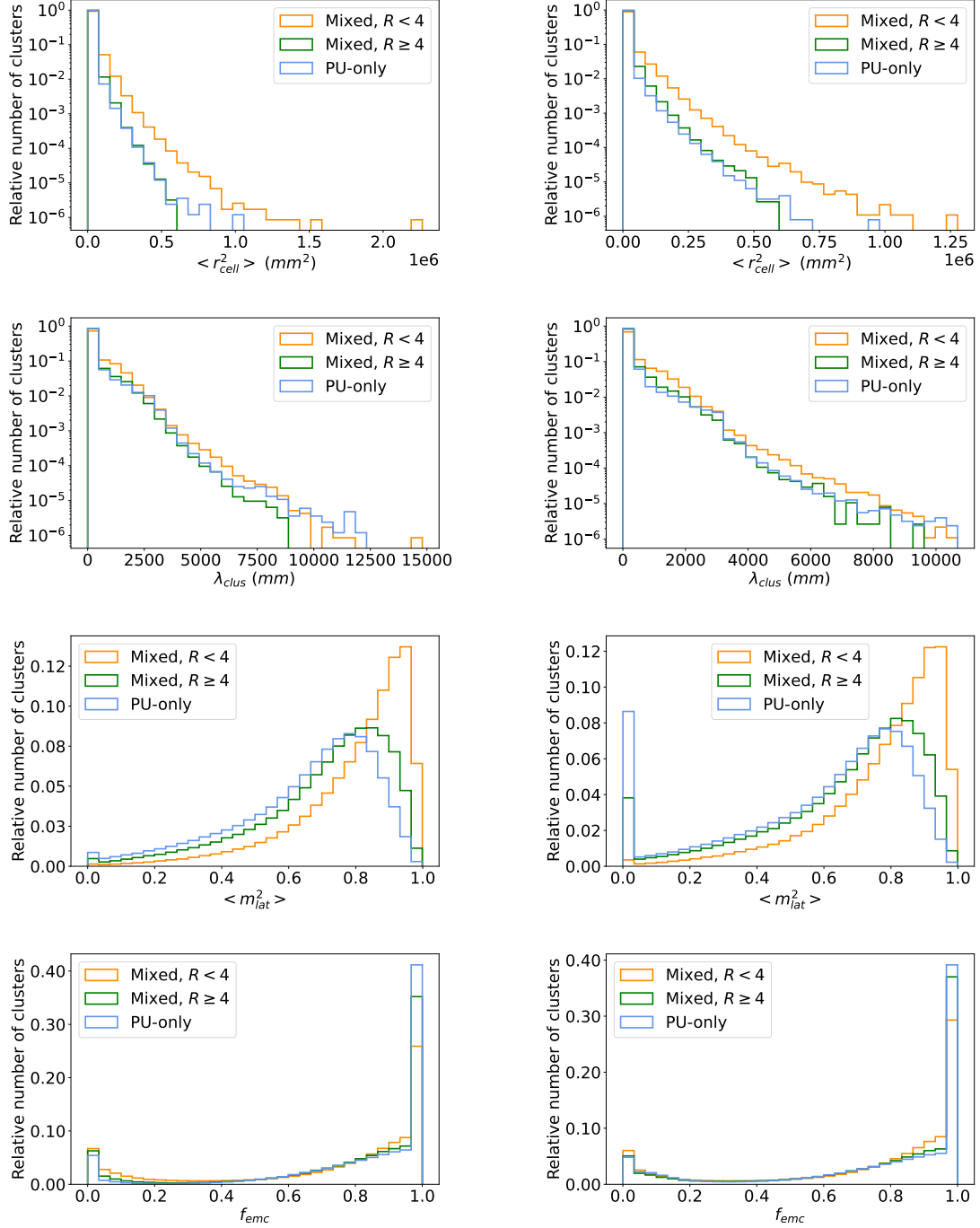


Figure 5.15: Distribution of the topocluster features used as network input, shown for pile-up-only clusters (light blue) and mixed clusters. The latter class is divided, based on R_{clus}^{EM} , into hard scatter-like clusters (orange) and pile-up-like clusters (green). The figures on the left represent the Run 2 Monte Carlo simulation, and the ones on the right the Run 3 one – continued.

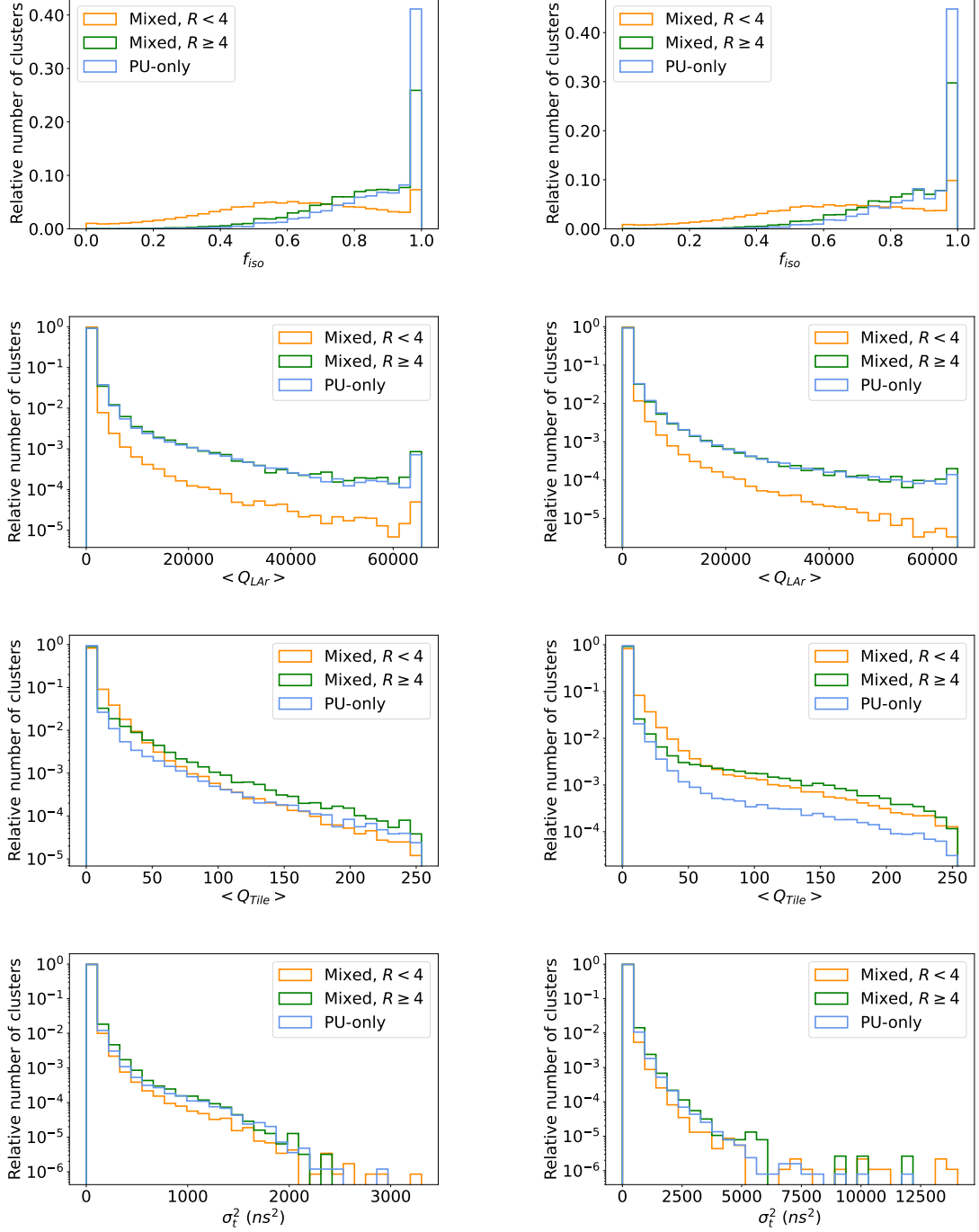


Figure 5.15: Distribution of the topocluster features used as network input, shown for pile-up-only clusters (light blue) and mixed clusters. The latter class is divided, based on R_{clus}^{EM} , into hard scatter-like clusters (orange) and pile-up-like clusters (green). The figures on the left represent the Run 2 Monte Carlo simulation, and the ones on the left the Run 3 one – continued.

5.3.1 Classification Results on Run 2

The pile-up-only identification network, built as described in Section 4.2, was used separately on Run 2 and Run 3 simulations to assess the possibility of identifying clusters with no hard scatter contributions.

The network losses on the training and validation sets, when applied to Run 2 data, are shown in Figure 5.16. In this case, both curves have reached a plateau when the network is fully trained, indicating a stable performance.

The probability of being a pile-up-only cluster, *i.e.* the output of the network, is shown for the test and training clusters in Figure 5.17. Although the output never reaches its maximum possible value of 1, there is a clear separation between the two classes, indicating a good discriminative power. Moreover, the separation improves if only mixed clusters with small responses are considered, as shown in Figure 5.18, where the output distribution is shown for the two types of mixed clusters in the test set. Indeed, pile-up-like mixed clusters tend to have features similar to the pile-up-only ones and are therefore likely to be associated with higher probabilities. Subsequently, further studies on this classification could be conducted by changing the definition of the mixed class to include only hard scatter-like clusters. Nonetheless, the figure also shows that the vast majority of mixed clusters have $R_{\text{clus}}^{\text{EM}} < 4$, thus the presence of a few mixed clusters that behave very similarly to the pile-up-only ones should not significantly affect the performance.

Indeed, the ROC and PR curves, plotted in Figure 5.19, show that the network yields good classification results, with a ROC-AUC of 0.876 and an average precision of 0.760 on the test data. Pile-up suppression can thus be achieved by selecting only clusters associated with an output $p < t$, where the threshold used as a cut for the classification $t = 0.49$ is found from the ROC curve.

The vast improvement in the cluster response distribution brought by this selection is shown in Figure 5.20, where the response is plotted for the entire test dataset and for only the clusters with an output below the threshold. As the dataset includes clusters with very low energies and high responses, the initial distribution has a very high IQR when compared to the ones shown in Section 5.2, where a lower bound was imposed on the cluster energy. However, the spread of the distribution decreases drastically when the selection is applied, with the IQR decreasing by three orders of magnitude.

Finally, the permutation importance of the input features is shown in Figure 5.21. A comparison with Figure 5.15 shows that the features that mostly influence the network also have visibly different distributions for the two classes. For example, the number of cells in the cluster n_{cells} tends to be far smaller for pile-up-only clusters than for the mixed ones. Also, clusters with no hard scatter contributions tend to have higher isolation f_{iso} .

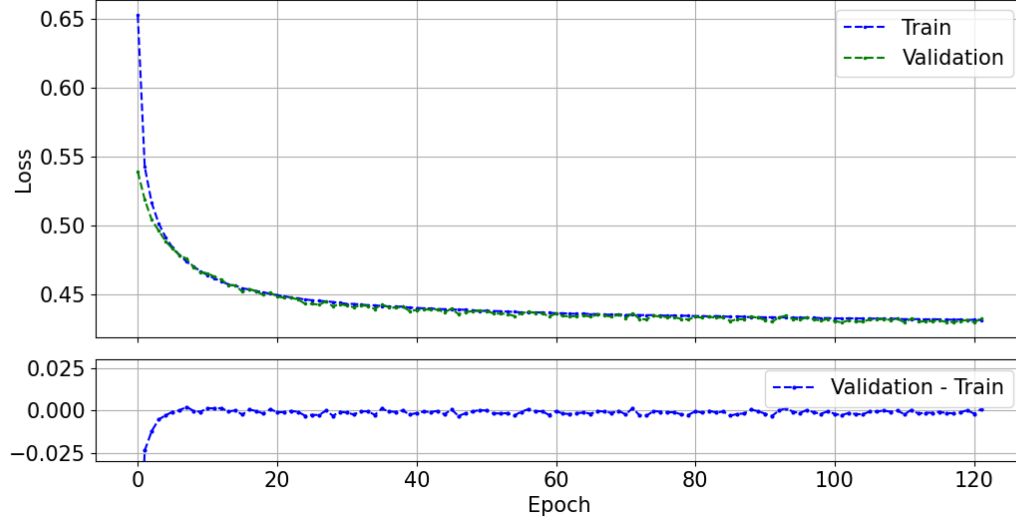


Figure 5.16: Network loss on the training (blue) and validation (green) datasets (above) for each training epoch. Difference between the two curves (below).

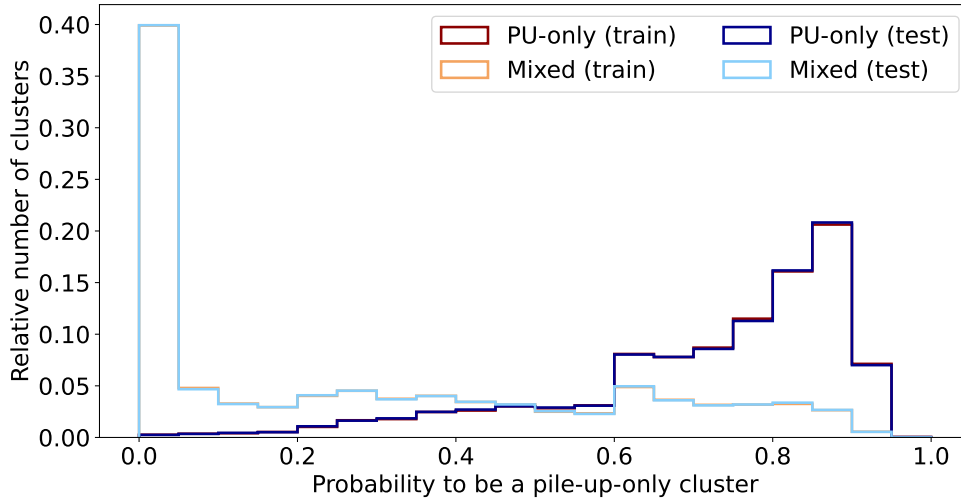


Figure 5.17: Network output on the train (light orange and red) and test (light blue and dark blue) data. The probability of containing only pile-up contributions is shown for mixed clusters (light orange and light blue) and pile-up-only clusters (red and dark blue).

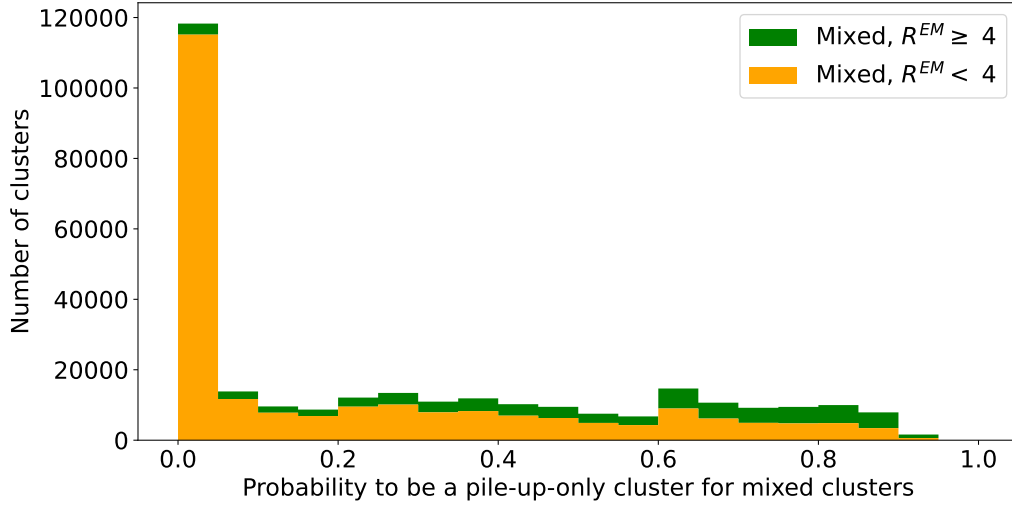


Figure 5.18: Stacked distributions of the probability to be a pile-up-only cluster for mixed clusters with $R_{\text{clus}}^{\text{EM}} < 4$ (orange) and $R_{\text{clus}}^{\text{EM}} \geq 4$ (green) in the test dataset.

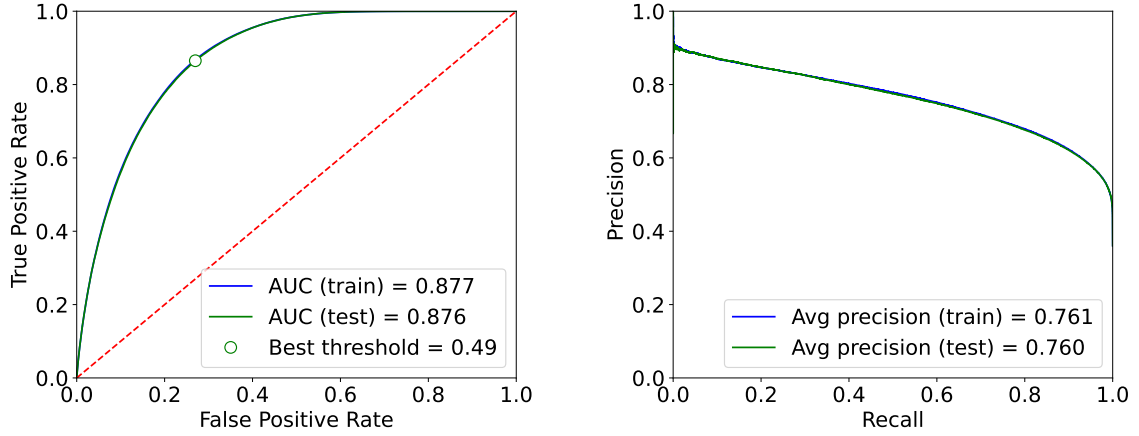


Figure 5.19: ROC (left) and PR (right) curves of the network on the train (blue) and test (green) data. The legends also show the ROC-AUC and average precision on both datasets. Moreover, the optimal threshold for classification, found as the upper corner of the train data's ROC curve, is shown in the legend on the left.

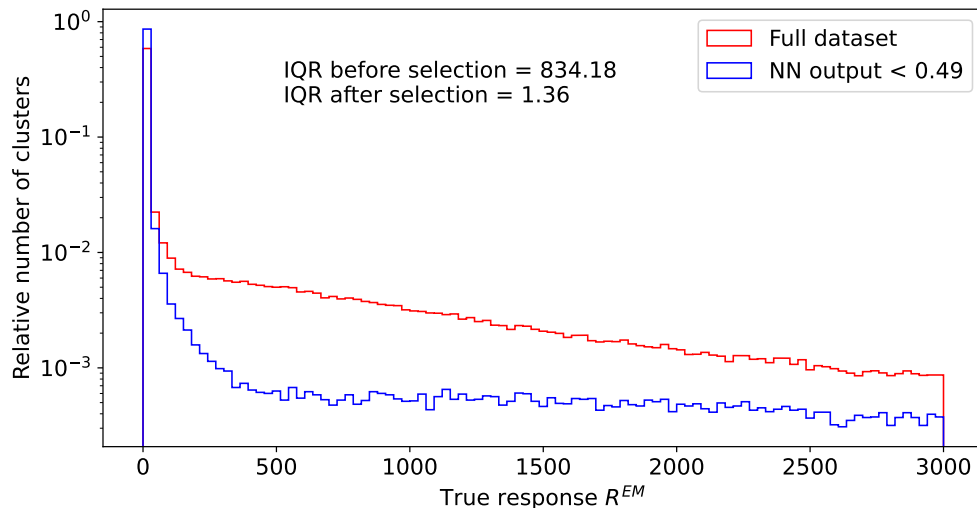


Figure 5.20: Topoclusters response R_{clus}^{EM} for the entire test dataset (red), and for only the clusters with $p < 0.49$ (blue), classified as mixed. The IQR before and after the selection (*i.e.* for the red and blue curve, respectively) is shown as a measure of the distributions' width.

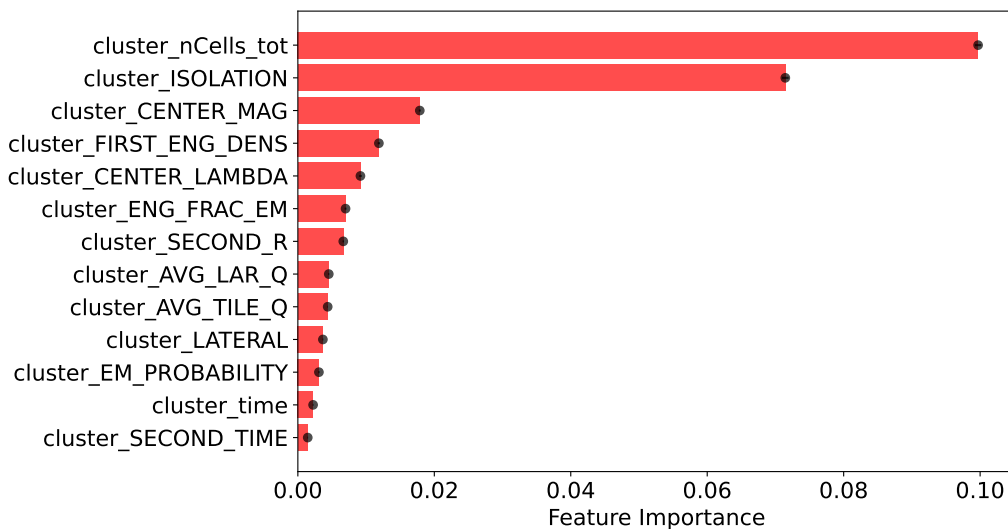


Figure 5.21: Permutation importance of each input feature (see Table 4.1 for the correspondence between variable and feature names), evaluated using Equation 4.8 with $K = 10$.

5.3.2 Classification Results on Run 3

The pile-up-only identification network was then studied with Run 3 clusters. Its losses on the training and validation data throughout the training are shown in Figure 5.22. As it happened with Run 2 data, both variables are stable at the end of the training process.

The network output, shown in Figure 5.23 for the two classes and in Figure 5.24 for the two types of mixed clusters, has a less pronounced probability distribution for pile-up-only clusters compared to that observed with clusters in the simulation of Run 2. Nevertheless, the overall behaviour of the network is very similar for the two simulations, thus the considerations on its output made in Section 5.3.1 continue to hold.

Figure 5.25 shows that the performance scores are comparable to those obtained on Run 2, with the network having a ROC-AUC of 0.853 and an average precision of 0.818 on the test dataset. The optimal classification threshold, found as the upper corner of the ROC curve, is $t = 0.50$. This is used as an upper bound to select mixed clusters, allowing the construction of the cluster response distribution without pile-up-only clusters, shown in Figure 5.26. The initial value of the distribution's IQR is very high due to the significant presence of pile-up in Run 3. However, it decreases by two orders of magnitude when only mixed clusters are selected, showing the importance of suppressing contributions from pile-up-only clusters.

In conclusion, the feature permutation importance, shown in Figure 5.27, has a lot of similarities with that shown in Figure 5.21 for Run 2.

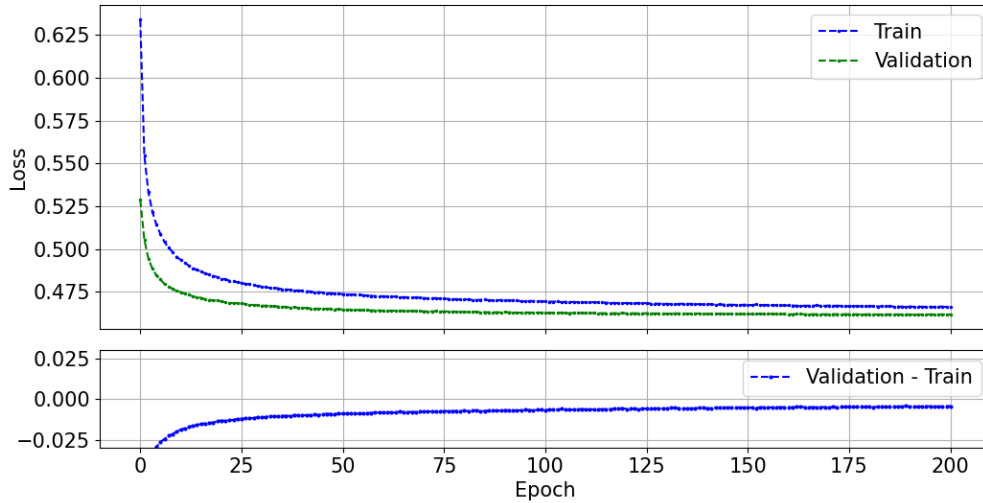


Figure 5.22: Network loss on the training (blue) and validation (green) datasets (above) for each training epoch. Difference between the two curves (below).

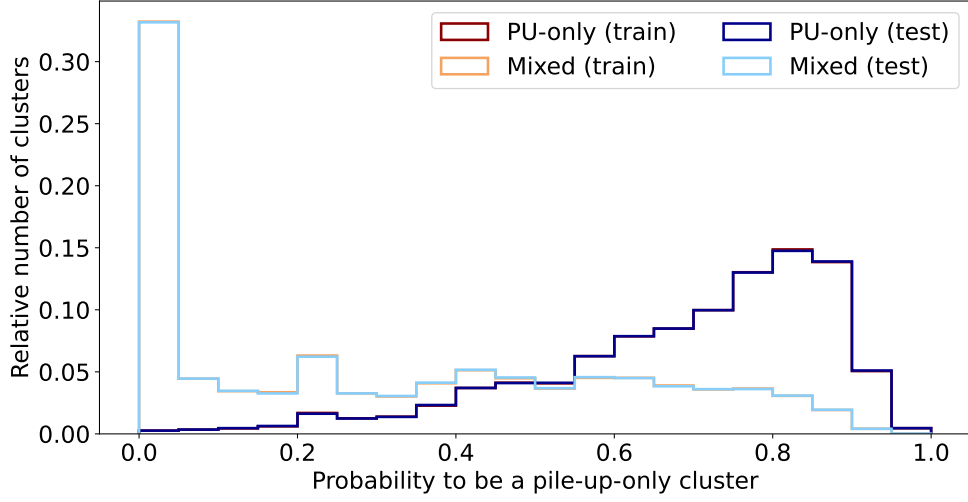


Figure 5.23: Network output on the train (light orange and red) and test (light blue and dark blue) data. The probability of containing only pile-up contributions is shown for mixed clusters (light orange and light blue) and pile-up-only clusters (red and dark blue).

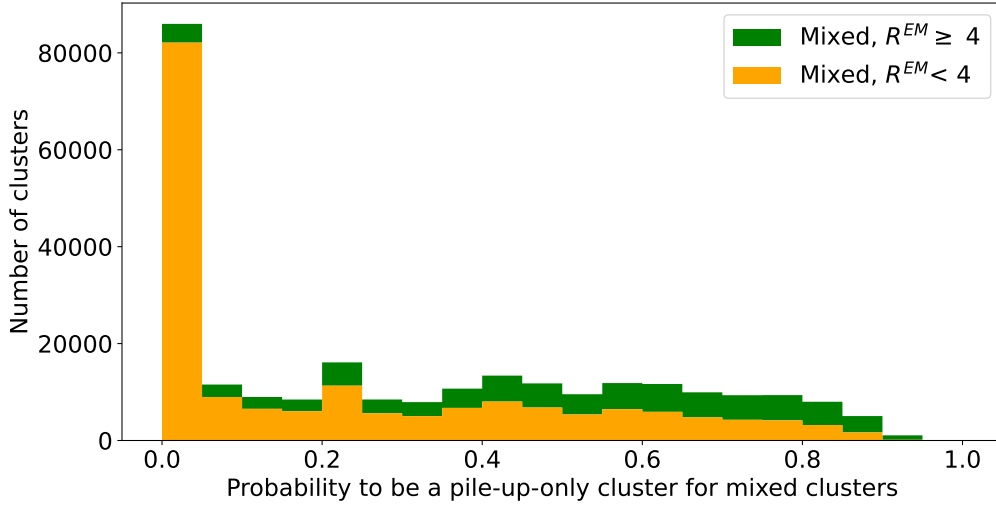


Figure 5.24: Stacked distributions of the probability to be a pile-up-only cluster for mixed clusters with $R_{\text{clus}}^{\text{EM}} < 4$ (orange) and $R_{\text{clus}}^{\text{EM}} \geq 4$ (green) in the test dataset.

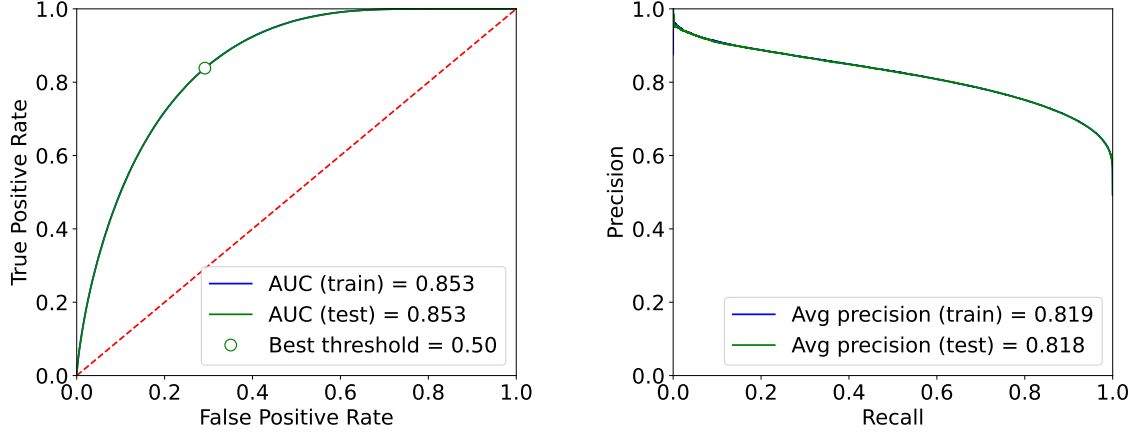


Figure 5.25: ROC (left) and PR (right) curves of the network on the train (blue) and test (green) data. The legends also show the ROC-AUC and average precision on both datasets. Moreover, the optimal threshold for classification, found as the upper corner of the train data's ROC curve, is shown in the legend on the left.

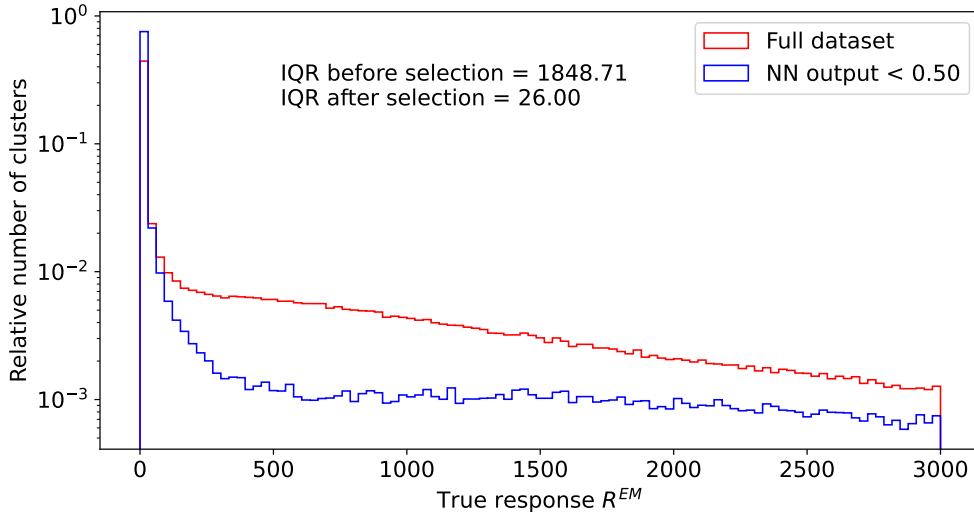


Figure 5.26: Topoclusters response R^{EM}_{clus} for the entire test dataset (red), and for only the clusters with $p < 0.50$ (blue), classified as mixed. The IQR before and after the selection (*i.e.* for the red and blue curve, respectively) is shown as a measure of the distributions' width.

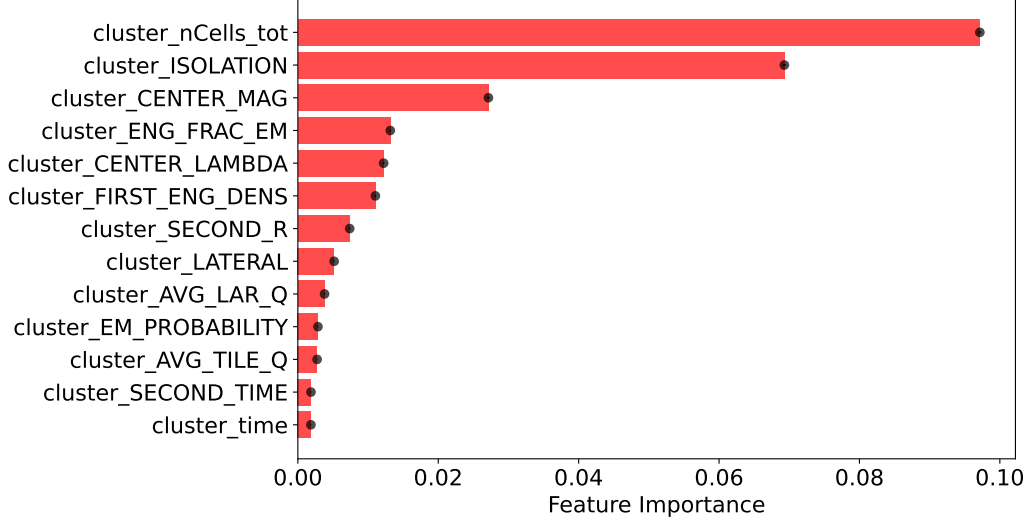


Figure 5.27: Permutation importance of each input feature (see Table 4.1 for the correspondence between variable and feature names), evaluated using Equation 4.8 with $K = 10$.

The several similarities in the behaviour of the network on Run 2 and Run 3 data, which were not observed in the network identifying hard scatter-only clusters, should not be seen as a cause for concern. In fact, the differences observed in Section 5.2 may be attributed to the presence of the timing cut as an additional source of pile-up suppression in Run 3. Since this cut only affects clusters in the full simulation, it could be the reason for the differences between Run 2 and Run 3 when such clusters are classified against those generated in hard scatter-only simulations. In contrast, all clusters used for the pile-up-only identification task are generated in full simulations, meaning that the timing cut affects both classes the network aims to recognize. As a consequence, the ability to discriminate between these two classes may not depend on the presence of this additional suppression, thus explaining the similar behaviour observed for the two LHC runs only for the network discussed in this section.

Conclusions

This work focused on how DNNs can be used to mitigate pile-up contributions in calorimeter topoclusters, in order to improve the energy response of the clusters and, consequently, their energy calibration. In particular, the possibility of determining the origin of the energy contributions to each cluster has been explored as a way of providing the information necessary to perform pile-up suppression at topocluster level.

The aim of these studies is to classify clusters according to whether they contain contributions arising solely from pile-up interactions, solely from the hard scatter one, or from a combination of the two. After such identification, the pile-up-only clusters could be removed from the inputs for jet reconstruction, while the pile-up contributions in mixed clusters could be mitigated.

Two DNNs were implemented to assess the origin of the energy contributions to calorimeter topoclusters in ATLAS, using simulated events for the LHC Run 2 and Run 3. The importance of this suppression was investigated with the preliminary studies on the topocluster energy response $R_{\text{clus}}^{\text{EM}}$. It was shown how the distribution of this variable widens under high pile-up conditions, implying a worse cluster energy calibration in these scenarios.

A DNN was then built to distinguish hard scatter-only from mixed clusters. A clear separation in the network output for the two classes was observed, showing the possibility of a good classification. This was also verified by evaluating the ROC-AUC and average precision scores of the network, which reached 0.972 and 0.966 for Run 2, and 0.883 and 0.861 for Run 3. The performance differences between the two datasets could be due to the different pile-up mitigation methods used in the data-taking periods considered. Indeed, a cut on the cluster cells' timing was introduced in Run 3 to limit the influence of out-of-time pile-up. Hence, this additional suppression could be the reason for the increased similarity between hard scatter-only and mixed clusters in Run 3 simulations. However, this motivation needs to be investigated further. Nevertheless, the network showed a very high discriminative power on both datasets, and its output was used to observe the difference in the width of the response distribution after the suppression of mixed clusters. As expected, for both simulations the IQR of the distribution decreases significantly after selecting solely the clusters classified as hard scatter-only, showing how information about the presence of pile-up in topoclusters can be useful to improve their

energy calibration.

Successively, after choosing a definition of pile-up-only clusters as those with $R_{\text{clus}}^{\text{EM}} > 4$ and $E^{\text{dep}} < 1 \text{ MeV}$, a second network was built to recognize them. For both datasets, the network performed well in distinguishing pile-up-only and mixed clusters when the latter had $R_{\text{clus}}^{\text{EM}} < 4$, but failed to identify mixed clusters with higher responses. Since these clusters behave very similarly to the pile-up-only ones, as shown by their features, future studies could be conducted by varying the thresholds used for the class definition. However, as only a few mixed clusters had responses above 4, their presence did not prevent the network from achieving high performance scores, with a ROC-AUC and average precision of 0.876 and 0.761 for Run 2, and 0.853 and 0.818 for Run 3. In this case, the performance scores and the importances of the network features were very similar for the two datasets. This can be explained by considering that, differently from the previous classification, here all clusters come from full simulations, so the cell timing cut in Run 3 is applied to both classes considered. As a result, the timing constraints may be irrelevant when distinguishing between the two classes.

With the same approach used before, the output of the network was used to build the cluster response distribution without pile-up-only clusters. This resulted in a significant reduction in the IQR range of the distribution, which decreased by a factor of approximately 10^3 for clusters in Run 2 and 10^2 for those in Run 3, consistently with the observation that the presence of pile-up results in higher values of $R_{\text{clus}}^{\text{EM}}$.

Finally, the feature importance studies provided insight into which topocluster features are more sensitive to the presence of pile-up. For example, for the identification of pile-up-only clusters, the number of cells and the cluster isolation have a very high separation power. This can be explained by considering that pile-up-only clusters have very low energies and don't contain contributions from multiple vertices, unlike mixed clusters. As a consequence, they tend not to include too many cells (since the presence of several cells with a high signal would result in a high cluster energy) and to be more isolated.

In summary, this thesis has shown how DNNs can be effectively used to assess whether the energy contributions in calorimeter topocluster arise from pile-up, hard scatter or from a mixture of the two. Accessing this information can be used as a new method for pile-up suppression, as clusters identified as pile-up-only could be removed from the set of jet inputs and mixed clusters could be calibrated differently from the hard scatter-only ones.

Several future studies could be carried out on this topic. For example, this suppression method could be compared to the current pile-up mitigation algorithms used in ATLAS, to see if it could be an alternative to them. Furthermore, the way in which a truth label is assigned to pile-up-only clusters could be modified by studying the optimal energy and response thresholds. In addition, the topocluster features could be defined using the DigiTruth method [48], a procedure which builds the topocluster moments using only hard scatter contributions. This would provide a way to quantify the pile-up

contributions in each cluster, potentially improving the classification performed here. Finally, a single network to discriminate among the three classes of clusters could be implemented and combined with the topocluster energy calibration network, thus building a DNN that both provides information on the cluster origin and calibrates its energy.

Bibliography

- [1] D. Perkins, *Introduction to High Energy Physics*. Cambridge University Press, 2000.
- [2] M. Kachelrieß and D. Semikoz, “Cosmic ray models,” *Progress in Particle and Nuclear Physics*, vol. 109, p. 103710, 2019. <https://www.sciencedirect.com/science/article/pii/S0146641019300456>.
- [3] K. Hübner, “Accelerators for high-energy physics,” 2006. <https://cds.cern.ch/record/823752>.
- [4] I. Neutelings, “Standard model,” accessed in 2024. https://tikz.net/sm_particles/.
- [5] F. Englert and R. Brout, “Broken Symmetry and the Mass of Gauge Vector Mesons,” *Phys. Rev. Lett.*, vol. 13, pp. 321–323, 1964. <https://link.aps.org/doi/10.1103/PhysRevLett.13.321>.
- [6] P. Skands, “Introduction to QCD,” in *Searching for New Physics at Small and Large Scales*, WORLD SCIENTIFIC, 2013. http://dx.doi.org/10.1142/9789814525220_0008.
- [7] E. M. Metodiev, “The Fractal Lives of Jets,” 2020. <https://www.ericmetodiev.com/post/jetformation/>.
- [8] L. Evans and P. Bryant, “LHC Machine,” *Journal of Instrumentation*, vol. 3, no. 08, p. S08001, 2008. <https://dx.doi.org/10.1088/1748-0221/3/08/S08001>.
- [9] “CERN: Accelerating Science,” accessed in 2024. <https://home.cern>.
- [10] E. Vercellin, “The ALICE experiment at the LHC,” *Nuclear Physics A*, vol. 805, no. 1, pp. 511c–518c, 2008. <https://doi.org/10.1016/j.nuclphysa.2008.02.289>.
- [11] The LHCb Collaboration, “The LHCb Detector at the LHC,” *Journal of Instrumentation*, vol. 3, no. 08, p. S08005, 2008. <https://dx.doi.org/10.1088/1748-0221/3/08/S08005>.

- [12] The ATLAS Collaboration, “The ATLAS Experiment at the CERN Large Hadron Collider,” *Journal of Instrumentation*, vol. 3, no. 08, p. S08003, 2008. <https://dx.doi.org/10.1088/1748-0221/3/08/S08003>.
- [13] The CMS Collaboration, “The CMS experiment at the CERN LHC,” *Journal of Instrumentation*, vol. 3, no. 08, p. S08004, 2008. <https://dx.doi.org/10.1088/1748-0221/3/08/S08004>.
- [14] P. Grafström, “Lifetime, cross-sections and activation,” 2007. <https://cds.cern.ch/record/1047067>.
- [15] W. Herr and B. Muratori, “Concept of luminosity,” 2006. <https://cds.cern.ch/record/941318>.
- [16] “Object-based missing transverse momentum significance in the ATLAS detector,” tech. rep., CERN, Geneva, 2018. <https://cds.cern.ch/record/2630948>.
- [17] I. Neutelings, “CMS coordinate system,” accessed in 2024. https://tikz.net/axis3d_cms/.
- [18] The ATLAS Collaboration, “Track Reconstruction Performance of the ATLAS Inner Detector at $\sqrt{s} = 13$ TeV,” tech. rep., CERN, Geneva, 2015. <https://cds.cern.ch/record/2037683>.
- [19] G. Aad *et al.*, “Optimisation of large-radius jet reconstruction for the ATLAS detector in 13 TeV proton–proton collisions,” *The European Physical Journal C*, vol. 81, no. 4, 2021. <http://dx.doi.org/10.1140/epjc/s10052-021-09054-3>.
- [20] R. Atkin, “Review of jet reconstruction algorithms,” *Journal of Physics: Conference Series*, vol. 645, no. 1, p. 012008, 2015. <https://dx.doi.org/10.1088/1742-6596/645/1/012008>.
- [21] M. Cacciari, G. P. Salam, and G. Soyez, “The anti-kt jet clustering algorithm,” *Journal of High Energy Physics*, vol. 2008, no. 04, p. 063, 2008. <https://dx.doi.org/10.1088/1126-6708/2008/04/063>.
- [22] G. Aad *et al.*, “Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1,” *The European Physical Journal C*, vol. 77, no. 7, 2017. <http://dx.doi.org/10.1140/epjc/s10052-017-5004-5>.
- [23] G. Aad *et al.*, “Improving topological cluster reconstruction using calorimeter cell timing in ATLAS,” *The European Physical Journal C*, vol. 84, no. 5, 2024. <http://dx.doi.org/10.1140/epjc/s10052-024-12657-1>.

- [24] “The application of neural networks for the calibration of topological cell clusters in the ATLAS calorimeters,” tech. rep., CERN, Geneva, 2023. <https://cds.cern.ch/record/2866591>.
- [25] J. Alimena *et al.*, “Searching for long-lived particles beyond the Standard Model at the Large Hadron Collider,” *Journal of Physics G: Nuclear and Particle Physics*, vol. 47, no. 9, p. 090501, 2020. <https://dx.doi.org/10.1088/1361-6471/ab4574>.
- [26] M. Cacciari, G. P. Salam, and G. Soyez, “SoftKiller, a particle-level pileup removal method,” *The European Physical Journal C*, vol. 75, no. 2, 2015. <http://dx.doi.org/10.1140/epjc/s10052-015-3267-2>.
- [27] P. Berta, M. Spousta, D. W. Miller, and R. Leitner, “Particle-level pileup subtraction for jets and jet shapes,” *Journal of High Energy Physics*, vol. 2014, no. 6, 2014. [http://dx.doi.org/10.1007/JHEP06\(2014\)092](http://dx.doi.org/10.1007/JHEP06(2014)092).
- [28] D. Bertolini, P. Harris, M. Low, and N. Tran, “Pileup per particle identification,” *Journal of High Energy Physics*, vol. 2014, no. 10, 2014. [http://dx.doi.org/10.1007/JHEP10\(2014\)059](http://dx.doi.org/10.1007/JHEP10(2014)059).
- [29] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [30] T. Sjöstrand, S. Mrenna, and P. Skands, “A brief introduction to PYTHIA 8.1,” *Computer Physics Communications*, vol. 178, no. 11, pp. 852–867, 2008. <https://www.sciencedirect.com/science/article/pii/S0010465508000441>.
- [31] R. D. Ball *et al.*, “Parton distributions with LHC data,” *Nuclear Physics B*, vol. 867, no. 2, pp. 244–289, 2013. <https://doi.org/10.1016/j.nuclphysb.2012.10.003>.
- [32] T. Sjöstrand, “Jet fragmentation of multiparton configurations in a string framework,” *Nuclear Physics B*, vol. 248, no. 2, pp. 469–502, 1984. [https://doi.org/10.1016/0550-3213\(84\)90607-2](https://doi.org/10.1016/0550-3213(84)90607-2).
- [33] “ATLAS Pythia 8 tunes to 7 TeV data,” tech. rep., CERN, Geneva, 2014. <https://cds.cern.ch/record/1966419>.
- [34] G. Aad *et al.*, “The ATLAS Simulation Infrastructure,” *The European Physical Journal C*, vol. 70, no. 3, p. 823–874, 2010. <http://dx.doi.org/10.1140/epjc/s10052-010-1429-9>.
- [35] S. Agostinelli *et al.*, “GEANT4—a simulation toolkit,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, no. 3, pp. 250–303, 2003. [https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8).

- [36] “The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie-Landshoff diffractive model,” tech. rep., CERN, Geneva, 2016. <https://cds.cern.ch/record/2206965>.
- [37] “Public ATLAS Luminosity Results for Run-2 of the LHC,” accessed in 2024. <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>.
- [38] “Public ATLAS Luminosity Results for Run-3 of the LHC,” accessed in 2024. <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun3>.
- [39] The ATLAS collaboration, “Monitoring and data quality assessment of the atlas liquid argon calorimeter,” *Journal of Instrumentation*, vol. 9, p. P07024, jul 2014. <https://dx.doi.org/10.1088/1748-0221/9/07/P07024>.
- [40] J. M. Seixas *et al.*, “Quality Factor for the Hadronic Calorimeter in High Luminosity Conditions,” *Journal of Physics: Conference Series*, vol. 608, no. 1, p. 012044, 2015. <https://dx.doi.org/10.1088/1742-6596/608/1/012044>.
- [41] F. Chollet *et al.*, “Keras,” 2015. <https://keras.io>.
- [42] M. Abadi *et al.*, “TensorFlow: Large-scale machine learning on heterogeneous systems,” 2015. <https://www.tensorflow.org/>.
- [43] X. Ying, “An Overview of Overfitting and its Solutions,” *Journal of Physics: Conference Series*, vol. 1168, no. 2, p. 022022, 2019. <https://dx.doi.org/10.1088/1742-6596/1168/2/022022>.
- [44] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” 2017. <https://arxiv.org/abs/1412.6980>.
- [45] L. Prechelt, *Early Stopping — But When?*, pp. 53–67. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. https://doi.org/10.1007/978-3-642-35289-8_5.
- [46] A. P. Bradley, “The use of the area under the ROC curve in the evaluation of machine learning algorithms,” *Pattern Recognition*, vol. 30, no. 7, pp. 1145–1159, 1997. [https://doi.org/10.1016/S0031-3203\(96\)00142-2](https://doi.org/10.1016/S0031-3203(96)00142-2).
- [47] “Compare Deep Learning Models Using ROC Curves – MathWorks Documentation,” accessed in 2024. <https://it.mathworks.com/help/deeplearning/ug/compare-deep-learning-models-using-ROC-curves.html>.
- [48] J. K. Roloff, *Exploring the Standard Model and beyond with jets from proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS Experiment*. PhD thesis, Harvard U., 2019. Appendix C. <https://cds.cern.ch/record/2677419>.

Eidesstattliche Versicherung

(Affidavit)

FAZZINO, GIULIA

Name, Vorname
(surname, first name)

243968

Matrikelnummer
(student ID number)

☐ Bachelorarbeit
(Bachelor's thesis)

☒ Masterarbeit
(Master's thesis)

Titel
(Title)

SUPPRESSING PILE-UP CONTRIBUTIONS IN THE
FORMATION OF TOPOLOGICAL CLUSTERS IN ATLAS

Ich versichere hiermit an Eides statt, dass ich die vorliegende Abschlussarbeit mit dem oben genannten Titel selbstständig und ohne unzulässige fremde Hilfe erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate kenntlich gemacht. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

I declare in lieu of oath that I have completed the present thesis with the above-mentioned title independently and without any unauthorized assistance. I have not used any other sources or aids than the ones listed and have documented quotations and paraphrases as such. The thesis in its current or similar version has not been submitted to an auditing institution before.

DORTMUND, 04/09/26

Ort, Datum
(place, date)

Giulia Fazzino

Unterschrift
(signature)

Belehrung:

Wer vorsätzlich gegen eine die Täuschung über Prüfungsleistungen betreffende Regelung einer Hochschulprüfungsordnung verstößt, handelt ordnungswidrig. Die Ordnungswidrigkeit kann mit einer Geldbuße von bis zu 50.000,00 € geahndet werden. Zuständige Verwaltungsbehörde für die Verfolgung und Ahndung von Ordnungswidrigkeiten ist der Kanzler/die Kanzlerin der Technischen Universität Dortmund. Im Falle eines mehrfachen oder sonstigen schwerwiegenden Täuschungsversuches kann der Prüfling zudem exmatrikuliert werden. (§ 63 Abs. 5 Hochschulgesetz - HG -).

Die Abgabe einer falschen Versicherung an Eides statt wird mit Freiheitsstrafe bis zu 3 Jahren oder mit Geldstrafe bestraft.

Die Technische Universität Dortmund wird ggf. elektronische Vergleichswerkzeuge (wie z.B. die Software „turnitin“) zur Überprüfung von Ordnungswidrigkeiten in Prüfungsverfahren nutzen.

Die oben stehende Belehrung habe ich zur Kenntnis genommen:

Official notification:

Any person who intentionally breaches any regulation of university examination regulations relating to deception in examination performance is acting improperly. This offense can be punished with a fine of up to EUR 50,000.00. The competent administrative authority for the pursuit and prosecution of offenses of this type is the Chancellor of TU Dortmund University. In the case of multiple or other serious attempts at deception, the examinee can also be unenrolled, Section 63 (5) North Rhine-Westphalia Higher Education Act (*Hochschulgesetz, HG*).

The submission of a false affidavit will be punished with a prison sentence of up to three years or a fine.

As may be necessary, TU Dortmund University will make use of electronic plagiarism-prevention tools (e.g. the "turnitin" service) in order to monitor violations during the examination procedures.

I have taken note of the above official notification:*

DORTMUND, 04/09/24

Ort, Datum
(place, date)

Giulia Fazzino

Unterschrift
(signature)

*Please be aware that solely the German version of the affidavit ("Eidesstattliche Versicherung") for the Bachelor's/ Master's thesis is the official and legally binding version.