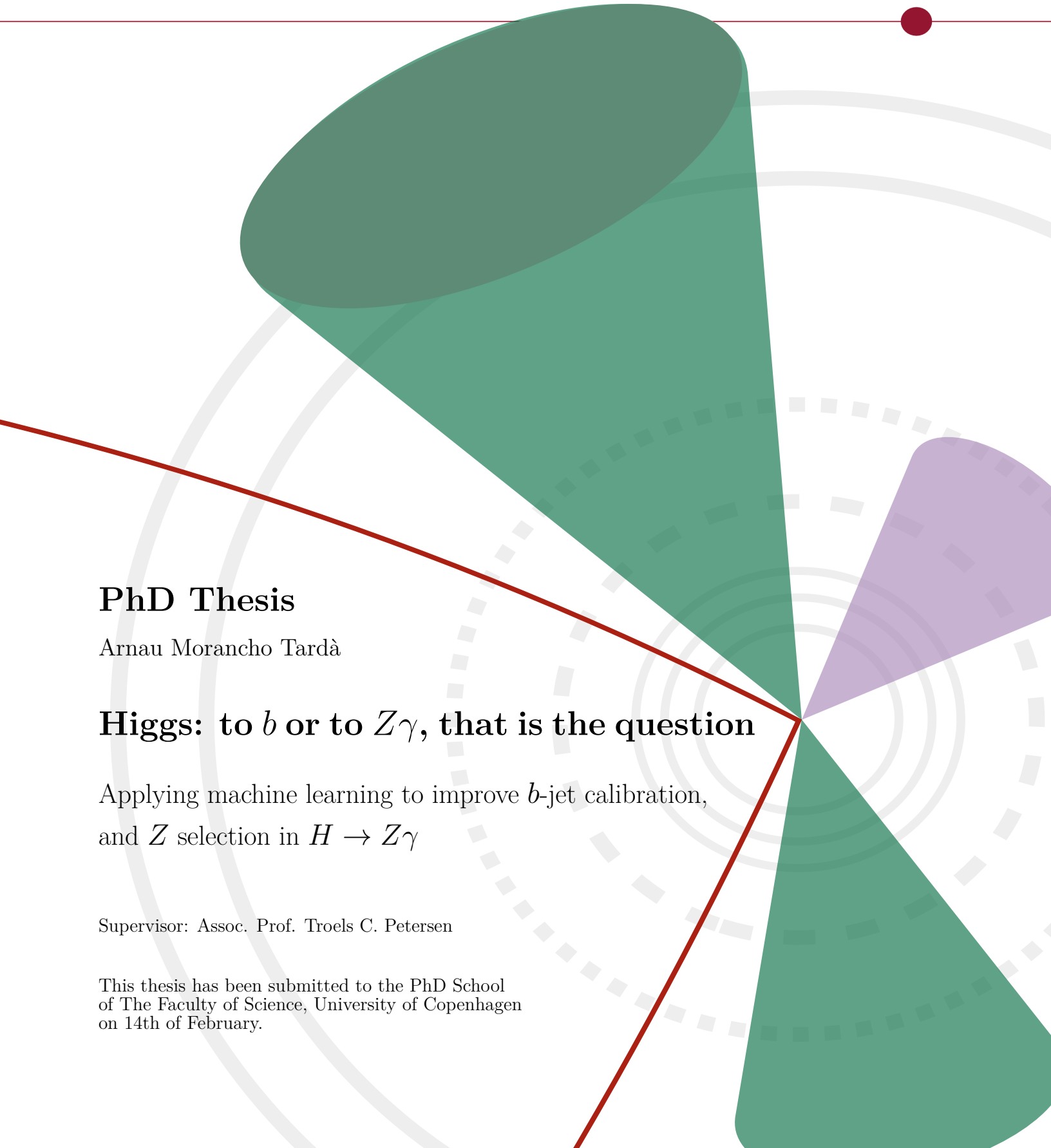
UNIVERSITY OF COPENHAGEN
FACULTY OF SCIENCE



PhD Thesis

Arnau Morancho Tardà

Higgs: to b or to $Z\gamma$, that is the question

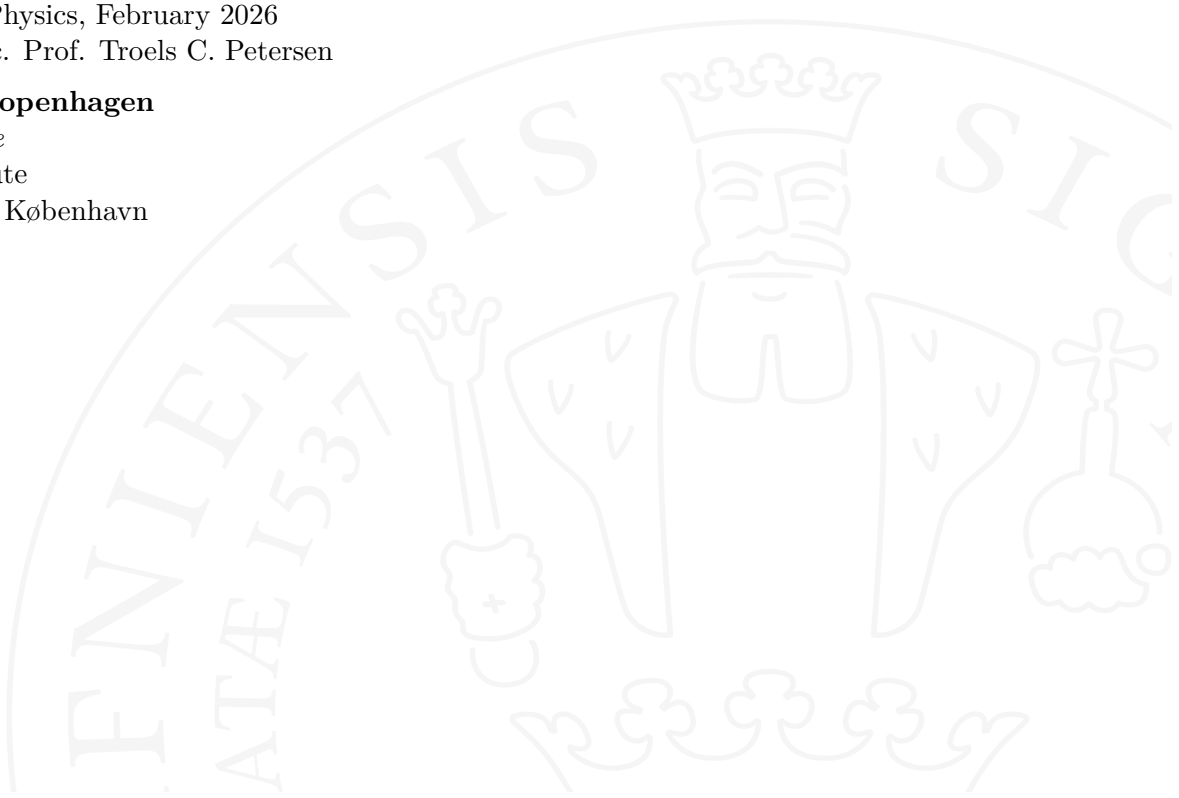
Applying machine learning to improve b-jet calibration, and Z selection in $H \to Z\gamma$

Supervisor: Assoc. Prof. Troels C. Petersen

This thesis has been submitted to the PhD School of The Faculty of Science, University of Copenhagen on 14th of February.

Arnau Morancho Tardà
Higgs: to b or to $Z\gamma$, that is the question
Applying machine learning to improve b-jet calibration, and Z selection in $H \to Z\gamma$
PhD in Particle Physics, February 2026
Supervisor: Assoc. Prof. Troels C. Petersen

University of Copenhagen
Faculty of Science
Niels Bohr Institute
Jagtvej 155, 2200 København



Abstract

The Higgs boson is one of the least experimentally characterised particles in the Standard Model. The high energy and luminosity delivered to the ATLAS detector at the LHC provide a unique opportunity to study this boson in detail. Its characterisation carries significant experimental challenges, arising from its relatively small production cross-section and consequently unfavourable signal-to-background ratio, as well as its predominantly hadronic and experimentally complex decay modes. These factors demand precise reconstruction of physics objects and efficient background discrimination, motivating the development of new analysis techniques. In recent years, machine learning (ML) techniques have emerged as powerful tools to address these complexities by exploiting non-linear correlations and high-dimensional information in collider data. This thesis investigates the application of ML techniques in two aspects of high-energy physics: b-jet calibration and the analysis of rare Higgs boson decays.

The first part of this work addresses jet calibration, a particularly challenging task due to complex showering dynamics and sensitivity to pile-up. A hybrid ML approach combining simulated and collision data is developed to mitigate the simulation–data domain shift. Boosted decision trees and transformer-based architectures are explored as alternatives to the final stages of the standard calibration procedure. While transformers show promise for domain adaptation, boosted decision trees achieve superior performance with the available training resources. The absence of truth-level information in collision data further motivates the development of methods to extract the detector-level jet energy resolution.

The second project focuses specifically on b-jet calibration. A new Monte Carlo (MC)-based calibration using a transformer architecture is introduced, resulting in an improved energy resolution of around 20% for b-jets. The detector resolution is estimated from data using a deconvolution approach combined with the Direct Balance method. While this new MC calibration has not yet been fully validated on collision data due to ongoing studies, comparisons with the nominal calibration indicate that no additional in-situ jet energy resolution correction is required for b-tagged jets.

The final part explores an ML-based lepton and Z boson selection for the rare $H \to Z\gamma$ decay, which is highly sensitive to new physics but statistically limited. A sequence of identification, isolation, and dilepton classification models increases the inclusive Z boson yield by about 20% while maintaining stable background levels. When propagated through event categorisation and signal modelling, projections using 2022 data show an improvement in the expected sensitivity from 1.81 to 2.01 for partial Run 3 luminosity, reducing the projected signal strength uncertainty from 0.70 to 0.64.

The results show that a dedicated ML-based calibration improves the energy resolution of b-jets, while in the $H \to Z\gamma$ analysis, an ML-based Z selection increases signal yield and reduces uncertainties. Overall, these findings highlight the practical impact of ML methods on boosting the signal efficiency and point toward promising directions for future developments in high-luminosity environments and broader applications within the ATLAS experiment.

Resumé

Higgs-bosonen er en af de mindst eksperimentelt karakteriserede partikler i Standardmodellen. Den høje energi og luminositet leveret til ATLAS-detektoren ved LHC giver en unik mulighed for at studere denne boson i detaljer. Dens karakterisering medfører betydelige eksperimentelle udfordringer, som skyldes den relative lille produktions-tværsnit og følgende det ikke fordelagtig signal-til-baggrund-forhold samt de overvejende hadroniske og eksperimentelt komplekse henfaldsveje. Disse faktorer kræver præcis rekonstruktion af fysiske objekter og effektiv baggrundsseparation, hvilket motiverer udviklingen af nye analysemetoder. I de senere år er machine learning (ML)-teknikker vundet frem som kraftfulde værktøjer til at håndtere disse udfordringer ved at udnytte komplekse korrelationer og højdimensionel information i collider-data. Denne afhandling undersøger anvendelsen af ML-teknikker inden for to aspekter af høj energi fysik: b-jet kalibrering og analysen af sjældne Higgs-boson henfald.

Den første del af dette arbejde omhandler jet kalibrering, en særligt udfordrende opgave på grund af komplekse showering-dynamikker og følsomhed over for pile-up. En hybrid ML-tilgang, der kombinerer simulerede- og kollisionsdata, udvikles for at reducere forkellen mellem simulations- og datadomæner. Boosted decision trees og transformer-baserede arkitekturer undersøges som alternativer til de sidste trin i den standard kalibreringsprocedure. Mens transformere viser lovende resultater for domænetilpasning, opnår boosted decision trees bedre resultater med de tilgængelige træningsressourcer. Fraværet af truth-level information i kollisionsdata motiverer yderligere udviklingen af metoder til at udtrække detector-level jet energy resolution.

Det andet projekt fokuserer specifikt på b-jet kalibrering. En ny Monte Carlo (MC)-baseret kalibrering ved brug af en transformer-arkitektur introduceres, hvilket resulterer i en forbedret energy resolution på omkring 20% for b-jets. Detektorpløsningen estimeres fra data ved hjælp af en deconvolution-tilgang kombineret med Direct Balance-metoden. Selvom denne nye MC-kalibrering endnu ikke er fuldt valideret på kollisionsdata på grund af igangværende studier, indikerer sammenligninger med den nominelle kalibrering, at der ikke kræves yderligere in-situ jet energiopløsningkorrektion for b-taggede jets.

Den sidste del undersøger en ML-baseret lepton- og Z-boson-selektion for det sjældne $H \to Z\gamma$ henfald, som er meget følsomt over for ny fysik, men statistisk begrænset. En sekvens af identification-, isolation- og to-lepton klassificationmodeller øger det inklusive Z-boson-udbytte med cirka 20%, samtidig med at baggrundsniveauerne forbliver stabile. Når dette propageres gennem event klassificering og signal modellering, viser projektioner baseret på 2022-data en forbedring i den forventede sensitivitet fra 1.81 til 2.01 for delvis Run 3 luminositet, hvilket reducerer den projicerede signalstyrekeusikkerhed fra 0.70 til 0.64.

Resultaterne viser, at en dedikeret ML-baseret kalibrering forbedrer energiopløsningen for b-jets, mens en ML-baseret Z-selektion i $H \to Z\gamma$-analysen øger signaludbyttet og reducerer usikkerheder. Samlet set fremhæver disse resultater den praktiske betydning af ML-metoder som forøger signal effektiviten og peger mod lovende retninger for fremtidig udvikling i høj-luminusitets miljøer samt bredere anvendelser inden for ATLAS-eksperimentet.

Acknowledgements

The original idea of being brief in these next lines is far gone. This thesis is dedicated to all the people who, in any way, contributed to my life over the past three years, both professional and personal. More than writing this for them, it is a letter of gratitude for this moment in time, a snapshot of both the present and the people who shaped who I am today.

Academically and personally, I have learned that nothing meaningful is accomplished alone. Progress relies on cooperation, patience, and the help you only realise you need after being stuck for a while.

First, to my supervisor, Troels. One of the most energetic and enthusiastic people I know, sometimes more motivated in my project than I am myself. Thank you for your guidance and for creating a safe environment where I could freely share my concerns and worries. I will always remember our intense discussions on the sailing boat under the perfect summer sun and the special Mont Blanc meeting.

Next, I would like to thank the brilliant minds at CERN, roughly in chronological order. First, the in-situ group and the JetEtMiss group, whom I had the privilege to meet in between karaoke at the two amazing HCWs I attended. Daniel and Magda, I have always been amazed by the hours you dedicate to every project in your subgroup and your determination to understand every detail. It has been a pleasure to learn from you, even if you were the reason I always worked until 6 pm on Fridays!

From the PhD group, I would like to thank Anubhav, Camron, Donna, and Simone. Special thanks to Laura, who took care of me when I was sick in a tent in the middle of Canada, and with whom I shared the unforgettable sight of a dead raccoon and a million stars reflected on a lake.

Also part of the in-situ group is the b-jet task force, led by Elisabeth and Brendon, who were always ready to help and share a coffee whenever I was at CERN. Thank you for your guidance over the past two years. However, practical help often comes from students, so thank you, Juan, Miguel, and Thiziri, for patiently answering my multiple questions.

Three months before my submission, I had the rush to visit part of the $H \to Z\gamma$ group in IHEP. Thank you, Yanping, for preparing everything for my stay and for providing invaluable input to my project. I am also grateful to Chenfeng, Fabio, Qiuping and especially to Shaoguang, for sharing your expertise and even showing me the best places to eat in the faculty. Your guidance was essential for my work on the analysis.

The dynamics of the Section 10 group at NBI often depend on the MSc students present each year, but over time, we have built a tight community, sharing meals and enjoying "Wednesday cake". To the professors of Section 10, who always have an open door for questions and life discussions, thank you, Alessandra, Craig, Daniel, Inar, and Jørgen. Special thanks to Stefania, who inspired my research interest while supervising my MSc. I would also like to thank the person who supports the group in countless practical ways, like cookies, and to Malene and Katrin, who ensure the bureaucracy of the university runs smoothly.

I cannot forget the core of the "pyramid": Emma, Jakob, Kat, Mark, Nina, Norman, Preet, and Tania. Thank you for making my Mondays lighter. To Ceren, Jens, and Magnus,

it has been a pleasure to be part of your MSc thesis. I hope I made your journey a little easier; you certainly made mine. Among them, I owe special thanks to three people who supported me the most. To Katinka, who would have thought that our language exchange four years ago was predicting the future? Thank you for your charm and vitality, even on days you barely slept. To Janni, who escaped from my last month of suffering. Our Pomodoro sessions, salsa dancing, and your lost keys will always hold a special place in my heart. I hope one day we can visit each other on a farm under construction. And for Kevin, Cheque, leque, panqueque! Thank you for being the brightest smile in the office, radiating energy and lifting all my concerns.

My months of academic solitude and doubt improved when I assisted the ESHEP school in Grennå, and I met incredible people. Without this experience and the friends I carried from there, I am not sure I would be writing these words today. Thank you for the piano concerts, the fires on the beach, acroyoga, karaoke, Shrek marathons, Milan vocabulary lessons, and becoming a fashion victim. Special thanks to Ana, Alberto, Chiara, Leo, Giovanni, and Graziella, and to all the other friends I may be forgetting. Many of you I had the joy to meet again in Geneva and at conferences. Thank you, your company made my days at CERN brighter. To Tiziano, for teaching me both history, politics and climbing, all while eating ice cream with Mont Blanc in the background. To Elena, my informal office mate, for being one of the warmest people I know, always prepared with a huge smile and deep hugs. Thank you for the therapy provided over the years and for sharing common worries. And for Paula, a qui, pobra, forço a parlar en català, per poder sentir-me com a casa, i amb ella s'aconsegueix fàcil. La teva manera d'estimar és a parts iguals: abraçades i ficar-te amb mi suaument. Gràcies per ensenyar-me la política de casa, sobreviure la calor anant al llac quan podíem i ensenyar-me les teves skills del hort.

During my stay at CERN, I also met my mentor Rachid, who helped me prepare for the challenges of my final year. Thanks also to Axel, Christina, and Magda for helping me navigate politics in science. Thanks to Leon and Chrysa, who brought a piece of Copenhagen to Switzerland, sharing many runs and invaluable life advice along the way.

Copenhagen has become my second home, thanks to my chosen family: Alba, Alicia, Anne, Beñat, Lasse, and Roberto. Even if we don't see each other every day, a chat with you in our improvised pub-crawls makes me laugh like I haven't in days. Also, thanks to Francesco for your invaluable advice in both climbing and life.

A *new* city comes with a new home, and I could not have been luckier to end up in such an easy-going environment that we built together, forming a huge "flamily." Thank you, Fiona, for always being ready to hear me complain about my day and for inspiring me with your personal life experiences. Thank you, Sarah, for entering our lives so quickly, filling them with energy and activities while remaining calm and rational when needed (and for blackmailing us with your bread). Thank you, Anna, for sharing both the struggles and adventures, from trips near and far, to moose hunting and unforgettable guitar concerts. Thank you for forcing me not to work on weekends in the first year, and for keeping me company during long working weekends in our last year. Many big adventures still await us. Thank you, Riv, for your warmth whenever I come to visit.

To my rope buddies, Antoine and Thibaut, whom I trust with my life countless times and

with whom I have explored spectacular places: thank you for all the escapes during CERN visits and for teaching me useful French. I look forward to many more adventures together, even if they come with a bit more calmness with age.

I now switch to Catalan for those who have been around for longer.

Per a les amistats de tota la vida, gràcies per dedicar-me temps cada cop que baixo i per posar-nos al dia de manera exprés perquè la conversa pugui acabar en banalitats, com fan els amics que es veuen cada dia. Gràcies Aida, Anna, Chantal, Núria i Sara. Especialment a la Maria, una constant a la meva vida, la meva millor amiga des que tinc memòria, amb qui els anys semblen no passar i amb qui encara ens queda molt camí fins a la jubilació.

Als amics de piragües, però també podria dir d'esport, de barri, i en general de vida. Gràcies per sempre prioritzar-me i ser tant honestos tot i que a vegades faci mal, gràcies Alan, Alba, Joan i Laura.

Pels qui sempre estan a punt per a qualsevol aventura que impliqui muntanya: gràcies Belen, Carlos, Laia, Nil i Juls. Us poso a tots al mateix sac amb l'esperança que algun dia ens puguem ajuntar tots sense haver de fer cinquanta mil canvis per veure'ns, però cadascú de vosaltres em doneu energia quan ens veiem

I als meus estimats companys de la uni i de moltes més coses, amb qui comparteixo aficions diferents (com un club de lectura, passió per la muntanya o el cel nocturn) però amb qui, junts, veiem el món de manera semblant i compartim valors similars. Infinites gràcies Hanna, Ignasi i Mariona.

Per tots els familiars que hi són, i sobretot pels que ja no. Per les noves generacions acabades de néixer i per les més grans, que sempre es preocupen que tot vagi bé. Gràcies a qui s'alegra més de la meva visita: la Pitufa. Per la família de Lleida, la de Maians i la de l'Almunia.

Especialment als pares, Carme i Josep Lluís: gràcies per no preguntar-me sempre quan torno a casa i per cuidar-me tant cada cop que hi torno i el cos diu prou. Sóc qui sóc pel que he vist en vosaltres; gràcies de tot cor. Gràcies per la vostra curiositat i per escoltar altres maneres de pensar. Espero que, si llegiu això, no tingueu por de demanar-me que us expliqui que porto fent aquests tres anys, prometo fer-ho millor que a la visita del CERN. A l'Edu, per fer-me sempre preguntes de les quals no tinc ni idea i que normalment m'he d'inventar. Gràcies per la teva curiositat pel món i per la muntanya, i per sempre trobar la diversió en el dia a dia. A la Ingrid, qui patia perquè no sabia com ajudar-me aquests últims mesos. Hi ha moltes maneres d'ajudar: algunes semblen insignificants però a vegades són les més necessàries. Gràcies simplement per ser-hi, per tenir por dels canvis però, tot i així, llançar-te a la piscina, per compartir traumes d'eficiència, d'impostor, etc., però també per compartir la manera de conviure-hi i sentir-nos-en orgullosos. Ets la meva inspiració.

La família és el més important i sé que sempre hi sereu; espero que em perdoneu que no sigueu els últims. Però hi ha dues persones que vaig conèixer alhora i que van esdevenir importantíssimes en moments diferents. Infinites gràcies als meus dos pilars, una a deu minuts de casa i l'altra a 1 000 km de distància.

Marta, crec que compartim la manera de pensar i de veure el món: sentir-nos lluny de casa però alhora amb angoixa de fer-ho, enyorar-la però sentir-nos estranys i, al mateix temps, eufòrics quan hi tornem. Admiro la teva facilitat per fer coses i aconseguir els teus

objectius quan te'ls proposes, però sobretot admiro la teva proximitat. Tenim pendents fer molts *somnis* realitats, que de moment seran en visites semestrals.

A la Neus, simplement gràcies per ser com ets, amb el teu somriure tímid i les fortíssimes abraçades que fan que tot sigui més fàcil. Has estat literalment la llum en tots els hiverns i de tu aprenc moltíssim cada dia que paso al teu costat. Gràcies per fer qualsevol conversa fàcil.

Per últim, m'agradaria dedicar-me i donar-me les gràcies a mi mateix, qui escriu aquestes paraules després de reviure els últims tres anys de la seva vida. Gràcies per no decaure quan tot es tornava fosc, per plorar quan ho necessitaves i per intentar que el teu voltant sigui un lloc més agradable amb una cosa tan aparentment insignificant com un somriure.

List of Publications

The work carried out during the three years of this PhD has led to the following publications.

I. A. Morancho Tardà, *New techniques for reconstructing and calibrating hadronic objects with ATLAS*, Proceedings of Science, EPS-HEP2025 (2026) 242, doi:10.22323/1.485.0242

II. ATLAS Collaboration, *Transformer networks for constituent-based b-jet calibration with the ATLAS detector*, CERN Technical Report ATL-PHYS-PUB-2024-015 (2024)

III. ATLAS Collaboration, *Search for the Higgs boson decay to a Z boson and a photon in pp collisions at* $\sqrt{s} = 13$ *TeV and* 13.6 *TeV with the ATLAS detector*, CERN Technical Report CERN-EP-2025-155 (2025), arXiv: 2507.12598

IV. Julia Allen *et al.* (ECFA Early-Career Researchers Panel), *The ECFA Early-Career Researchers Panel: Report for the year 2023*, arXiv: 2407.12761 (2024)

V. J.-H. Arling *et al.*, *Early Career Researcher Input to the European Strategy for Particle Physics Update: White Paper*, arXiv: 2503.19862 (2025)

The single-author article in Ref. I summarises six novel methods to improve jet reconstruction and calibration.

I also contributed directly to the ATLAS Collaboration studies reported in Refs. II and III. The research presented in this thesis builds upon these efforts and further develops the underlying methods and analyses; however, it does not constitute a direct reproduction of the corresponding collaboration publications.

During most of my PhD, I served as a Danish representative on the ECFA Early-Career Researchers Panel, which resulted in the report presented in Ref. IV. In parallel, preparations for the European Strategy for Particle Physics update were ongoing, and the European Early-Career Researchers jointly contributed to a White Paper reported in Ref. V.

Table of Contents

Chapter 1

Introduction

Particle physics aims both to deepen our understanding of the fundamental laws governing the Universe and to develop the experimental techniques required to test them. Many discoveries primarily deepen our conceptual understanding of nature. At the same time, the experimental challenges of probing the smallest scales have driven major technological advances, whose societal impact often becomes apparent only decades later.

The Standard Model (SM) [1–4] provides a remarkably successful description of the known particles and their interactions, with its predictions in excellent agreement with experimental results. Nevertheless, it is widely recognised as an incomplete description of nature, as it does not contemplate several observed phenomena. Progress in particle physics increasingly depends on precision measurements and the identification of subtle deviations from established predictions, rather than solely on the discovery of entirely new particles or interactions. In this context, improving the experimental sensitivity of analyses is crucial for guiding theoretical efforts to explain these new observations.

The Higgs boson plays a central role in this effort. First hypothesised in 1964 [5–9] and discovered in 2012 by the ATLAS and CMS collaborations [10, 11], it completed the particle content of the SM by providing a mechanism through which elementary particles acquire mass. This occurs via the Higgs field, a scalar field that permeates all of space: particles acquire mass through their interactions with this field, with the strength of the interaction determining the particle's mass. The Higgs boson itself is an excitation of the Higgs field, and its discovery provides direct evidence of the field's existence.

The Large Hadron Collider (LHC) [12] is currently the only facility capable of producing and detecting Higgs bosons significantly, given the high energies required to produce this massive particle. Although more than a decade has passed since its discovery, the experimental characterisation of the Higgs boson remains incomplete. Many of its properties are still measured with limited precision, and several rare decay modes remain statistically constrained, including those probing couplings to first-generation fermions.

Studying the Higgs boson serves two complementary purposes. First, precision measurements test the internal consistency of the SM and constrain the properties of known particles. Second, the Higgs sector is widely regarded as a potential gateway to physics beyond the SM, including possible connections to the dark sector. For these reasons, Higgs physics re-

mains an active and fruitful field, where methodological advances can directly translate into enhanced sensitivity and new insights.

Achieving these required precisions is challenging. At the LHC, roughly sixty proton–proton interactions occur per bunch crossing, making object reconstruction in this dense environment a non-trivial task. Additional interactions, collectively referred to as pile-up, can overlap with the event of interest, contaminating reconstructed objects and distorting the signal. These experimental challenges highlight the need for sophisticated reconstruction and analysis techniques capable of disentangling signal from background and correcting for detector effects.

Machine learning (ML) techniques offer a promising framework for tackling these challenges.They are well-suited for handling high-dimensional feature spaces and for learning non-linear correlations that are difficult to capture with traditional approaches. While such methods typically require large training datasets, experiments such as ATLAS provide an ideal environment, with extensive simulated samples and large volumes of collision data available for training, validating, and testing. Decision-tree-based algorithms have long been used successfully for classification tasks, such as background rejection. In recent years, neural networks have become increasingly widespread for tagging, categorisation, and identification tasks, as both computational resources and methodological understanding have improved. In this thesis, we explore ML as a methodological tool to improve both object calibration and event selection in Higgs boson analyses.

The first part of this thesis focuses on enhancing the energy resolution of b-jets, with particular emphasis on improving the reconstruction of Higgs boson decays involving jets. As the most abundantly produced objects at the LHC, their precise reconstruction is crucial for many analyses in which they can constitute either signal, background, or both simultaneously.

Jets originate from high-energy quarks or gluons produced in the collision. Due to colour confinement, these partons cannot exist freely, and their energy is radiated into a cascade of additional particles in a parton shower. The shower subsequently hadronises into unstable hadrons, which decay into lighter particles that deposit energy across an extended region of the calorimeter. Reconstructing the original parton energy is therefore highly challenging, as it requires modelling the stochastic nature of parton showering, hadronisation, hadron decays, and detector effects. The reconstructed transverse momentum of jets can deviate by 5–20% from the truth-level value in MC, depending on the jet's energy and its location in the detector. Consequently, achieving accurate jet calibration and ensuring good agreement between simulation and collision data remains a complex task, yet essential [13, 14].

Among these difficulties, b-jets play a central role, as they are present in top-quark decays and constitute the dominant decay mode of the Higgs boson. However, their relatively poor energy resolution and the presence of large backgrounds pose significant experimental challenges. Although the decay of the Higgs into two b-jets has been observed [15], further improvements in b-jet energy resolution would directly benefit both precision measurements and searches involving Higgs bosons.

This is particularly relevant for studies of the Higgs self-interaction, which provide direct access to the shape of the Higgs potential and have profound implications for the evolution

of the Universe. The most probable final state in di-Higgs production consists of four b-jets, making accurate b-jet reconstruction and calibration a key ingredient for these analyses. A distinctive feature of b-jets is the presence of B hadrons, which give rise to displaced decay vertices and, in some cases, semileptonic decays. These characteristics are already exploited through dedicated corrections in the analysis level, such as *muon-in-jet* and *PtReco* corrections [16], which partially account for the specific properties of b-jets but have some limitations.

In this thesis, the calibration of b-jets is addressed with the goal of improving their nominal energy resolution. A hybrid ML approach is investigated, trained on both simulated and collision data, targeting the jet's transverse momentum. This strategy aims to reduce the simulation-data domain shift and to provide a more accurate energy estimate by effectively replacing two standard calibration steps with a single ML-based correction.

A key challenge in data-driven calibration is the absence of truth-level information, which makes it necessary to exploit different event topologies. For this purpose, samples of Z+jets events are used, where the leading jet is approximately back-to-back with the Z boson. The Z momentum, having a narrower resolution, serves as a reference object; however, the response measured in this way does not directly correspond to the detector resolution, since certain detector-level effects are not fully captured in simulation. In this context, a calibration procedure is developed to extract the detector-level resolution from such reference measurements.

Simultaneously with the study of hybrid training, a transformer model [17] demonstrated strong performance on Monte Carlo (MC) only b-jets [18], improving their transverse momentum resolution by approximately 20%. In this work, we explore its application to collision data, using both Z+jets and γ+jets samples. Since the transformer predictions are not yet available for data, the nominal calibration procedure is used for validation and as a reference to guide the next steps in applying the model to real collision events. An improved method for extracting detector-level resolution from reference measurements is presented. By allowing the simultaneous extraction of jet energy scale and resolution from the same samples, it complements existing techniques and has the potential to reduce the associated uncertainties.

The final project addresses a different Higgs decay channel from an analysis perspective. The decay of the Higgs boson into a Z boson and a photon is a rare process that proceeds through loop diagrams, resulting in a small branching fraction of $Br(H \to Z\gamma) = (1.54^{+0.10}_{-0.11}) \times 10^{-3}$ for $m_H = 125.09$ GeV in the SM. Measuring this decay is an important step towards completing the study of Higgs boson interactions with electroweak gauge bosons and provides additional insight into the mechanism of electroweak symmetry breaking. Furthermore, the loop-induced nature of this decay makes it particularly sensitive to physics beyond the SM, as new particles entering the loop could easily modify the branching ratio.

The $H \to Z\gamma$ decay remains experimentally challenging due to its small branching ratio and its non-reducible background. For instance, the combined ATLAS and CMS Run 2 results [19] reached an observed significance of 3.4σ under the background-only hypothesis, providing evidence of the decay but still short of the 5σ threshold required for claiming a discovery. The recently published ATLAS partial Run 3 analysis [20] provides improved con-

straints on the signal strength but remains statistically limited, motivating the development of new methods. In this context, this thesis investigates an ML-based Z boson selection to increase signal yield, while explaining in detail the partial Run 3 results. The impact of this improved selection on the expected analysis sensitivity is evaluated in detail.

In both projects, ML is used as a methodological tool to achieve more precise measurements and greater statistical sensitivity. The results presented in this thesis illustrate the practical impact of advanced computational techniques on detector calibration and physics analyses, and highlight their potential to enhance discovery prospects in high-energy physics.

Thesis outline

This thesis is organised as follows.

Part I presents the theoretical and experimental framework of the studies. Chapter 2 introduces the basic concepts of particle physics, outlining how the known fundamental particles and interactions are described within the framework of the SM. The Higgs mechanism is then presented, followed by an overview of the current experimental status of Higgs boson physics, highlighting both established results and remaining limitations. The chapter concludes with a discussion of the open questions and theoretical shortcomings of the SM.

Once the theoretical framework is established, Chapter 3 focuses on the experimental context of this work. It begins with a brief overview of CERN's history and the present accelerator complex, and continues with a description of the ATLAS detector and the reconstruction of physics objects from recorded collision data.

Part II of the thesis is devoted to jet calibration studies. Chapter 4 first presents the jet reconstruction procedure and motivates the need for further calibration. The standard jet calibration methods used in ATLAS are then described, followed by a brief introduction to flavour tagging and b-jet–specific corrections.

Building on this, Chapter 5 explores a novel approach in which two calibration steps are replaced by a single ML algorithm trained using a hybrid strategy based on both simulation and collision data. The chapter presents the methodology and discusses its performance in detail.

Chapter 6 introduces a new MC–based calibration for b-jets using a transformer model. Although the application of the new calibration to collision data is still under development, its impact on the jet energy resolution is demonstrated using a novel validation method.

Part III focuses on the Higgs boson analysis. Chapter 7 reviews the recently published partial Run 3 results for the $H \to Z\gamma$ decay channel and introduces the main concepts and methodology of the analysis.

Chapter 8 then explores an ML-driven event selection aimed at improving Z boson identification and increasing its effective statistics, and studies its propagation through the full analysis chain to assess its impact on the expected sensitivity.

Finally, Chapter 9 summarises the main results of this thesis and discusses possible directions for future work.

Scope and conventions

This thesis is written for readers with a graduate-level background in particle physics. It aims to provide the level of explanation that would have been helpful at the start of this work, supporting readers new to experimental analyses within large collaborations. Several concepts and procedures that are often assumed knowledge in internal documentation or publications are therefore discussed in detail to make the text as self-contained as possible.

Clear terminology is adopted to distinguish between simulated events and data collected from real proton–proton collisions. Samples obtained from event generation and detector simulation are referred to as *simulation* or *Monte Carlo (MC)* data. Data recorded from proton–proton collisions are referred to as *real data*, *collision data*, or simply *data*, depending on the context.

A distinction is also made regarding authorship. Throughout the thesis, the pronoun *we* is used when describing work to which I have directly contributed. In contrast, results, methods, and procedures produced collectively by the groups I participated in are generally presented in the passive voice.

Part I

Theoretical and Experimental Framework

Chapter 2

Theory description

The Standard Model [1–4] (SM) of particle physics describes the fundamental constituents of matter and the interactions governing them. It represents the most successful theoretical framework currently available, with its predictions confirmed by a wide range of experimental observations.

Despite its success, the SM is not a complete description of nature. It does not incorporate gravity, nor does it explain the non-zero masses of neutrinos and their oscillations, the nature of dark matter and dark energy, or the observed baryon asymmetry of the Universe. For these reasons, it is generally believed that the SM is an effective theory [21] valid up to a specific energy scale, beyond which a more fundamental and unified theory is expected to emerge.

This chapter provides an overview of the SM and its extensions, primarily based on Refs. [22–25]. It begins with a conceptual introduction to particle physics and the gauge symmetries underlying the SM, followed by a mathematical description of its particle content. The properties of the Higgs boson are then discussed, including its production mechanisms, decay modes, and experimental measurements. The chapter concludes with a discussion of the open questions and limitations of the SM, motivating the need for physics beyond it.

2.1 The Standard Model

2.1.1 Conceptual overview

The Standard Model successfully describes three of the four fundamental interactions: the electromagnetic, weak, and strong forces. Matter is composed of fundamental particles called fermions, which have half-integer spin and obey Fermi–Dirac statistics. These particles interact via the exchange of bosons, particles with integer spin that mediate the fundamental forces.

The electromagnetic force has infinite range and is mediated by the massless photon, a neutral vector boson. The weak interaction, responsible for processes such as radioactive beta decay, is mediated by the massive charged $W^{\pm}$ bosons and the neutral Z boson. Due to

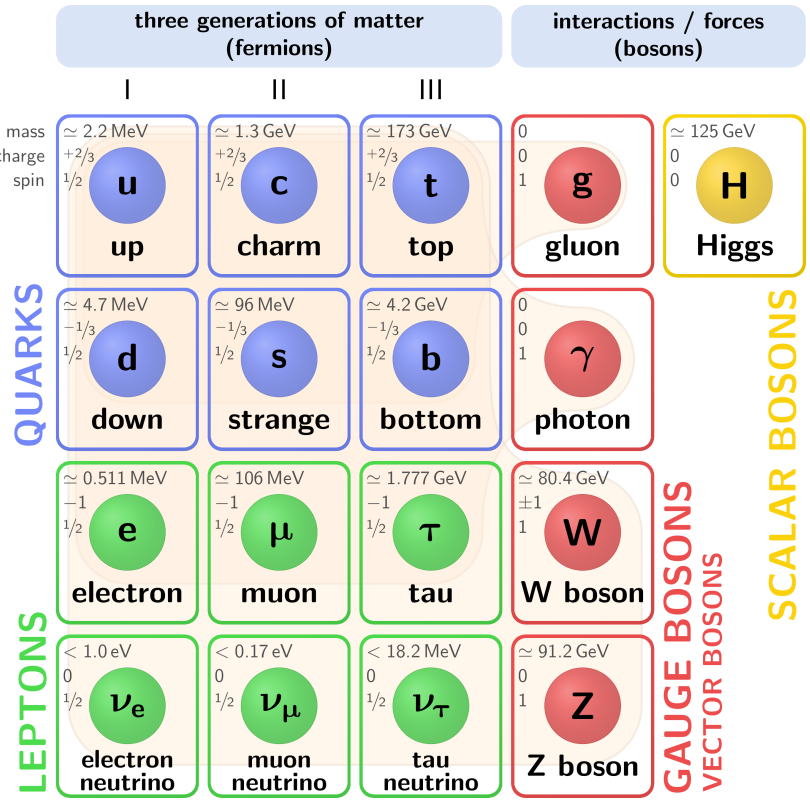


Figure 2.1: Representation of the matter particles and their interaction mediators in the Standard Model. Fermions are shown on the left, divided into quarks (purple) and leptons (green). Bosons are shown on the right, divided into vector gauge bosons (red) and the Higgs scalar boson (yellow). Figure from Ref. [26].

the large masses of these mediators, the weak force is effective only at very short distances. The strong interaction, which holds protons and neutrons together within atomic nuclei, is mediated by massless gluons. It is described by Quantum Chromodynamics (QCD), where the relevant charge is known as colour.

The gravitational interaction, while having infinite range, is many orders of magnitude weaker than the other forces at the quantum scale. It is successfully described by general relativity at macroscopic distances but is not incorporated into the SM, as a consistent quantum formulation is still lacking.

Figure 2.1 summarises the particle content of the SM and its main properties. Matter particles are fermions with spin 1/2, which are grouped into leptons and quarks, each organised into three generations. Within each generation, there are two sets of fermions with different electric charges.

In the quark sector, the up-type quarks (up, charm, and top) carry an electric charge of $Q = +2/3$, while the down-type quarks (down, strange, and bottom) carry a charge of $Q = -1/3$. Each successive generation is heavier than the previous one, although the reason for this mass hierarchy is not explained within the SM.

In the lepton sector, the charged leptons (the electron, muon, and tau) have an electric charge of $Q = -1$. Each charged lepton is accompanied by a corresponding electrically neutral and massless neutrino, named after its charged partner: the electron neutrino, muon neutrino, and tau neutrino.

Quarks and charged leptons have corresponding antiparticles that have the same mass and spin but opposite quantum numbers, most notably electric charge. For neutrinos, the situation is more subtle: in the SM, neutrinos are left-handed, while antineutrinos are right-handed, reflecting a difference in chirality rather than electric charge. However, the SM does not determine whether neutrinos are Dirac particles, distinct from their antiparticles, or Majorana particles, in which case neutrinos would be their own antiparticles. This question is closely connected to physics beyond the SM and has important implications for the origin of neutrino masses and the matter–antimatter asymmetry of the Universe.

On the right side of Fig. 2.1, the force carriers of the SM are shown; these are bosons with integer spin.

The gluon (g) mediates the strong force, which binds quarks together inside hadrons. Gluons are massless, electrically neutral spin-1 particles. Gluons themselves carry colour charge, allowing them to interact with each other, a unique property of the strong force.

The photon (γ) mediates the electromagnetic force. It is also massless, electrically neutral, and has spin 1. The photon interacts with any particle carrying electric charge, including charged leptons and quarks.

The weak force is mediated by the massive bosons $W^{\pm}$ and Z, each with spin 1. The $W^{\pm}$ bosons carry electric charges ± 1, while the Z boson is neutral. Weak interactions can change the flavour of quarks or leptons: W bosons mediate transitions between different types of fermions, whereas the Z boson mediates neutral-current interactions without changing the particle type. Additionally, the $W^{\pm}$ bosons can couple to photons, and the W and Z bosons interact among themselves.

Finally, the Higgs boson (H) is a scalar particle with spin 0, responsible for giving mass to the W and Z bosons and the fermions through the Higgs mechanism. A more detailed discussion of the Higgs mechanism and the properties of the Higgs boson is presented later in Section 2.2.

While this particle-based description provides an intuitive understanding of the SM, its predictive power arises from a precise mathematical formulation as a gauge field theory, which is introduced in the following.

2.1.2 Gauge symmetries of the Standard Model

The SM is formulated as a gauge field theory based on the symmetry group $SU(3)_C \times SU(2)_L \times U(1)_Y$. The group $U(N)$ consists of $N \times N$ complex unitary matrices. Imposing the additional condition that the determinant of these matrices is equal to one defines the special unitary group $SU(N)$. The group $U(N)$ has N^2 generators, while $SU(N)$ has $N^2 - 1$ generators. Requiring local gauge invariance under this symmetry group uniquely determines the structure of the Lagrangian, the allowed interactions, and the corresponding force carriers.

Global symmetries

Let us first focus on the $U(1)_Y$ symmetry. Within the Lagrangian formalism, the theory is invariant under a continuous global $U(1)$ transformation. As an illustrative example, we consider a Dirac fermion described by the Lagrangian

$$\mathcal{L} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi\,. \tag{2.1}$$

This Lagrangian is invariant ($\mathcal{L}' = \mathcal{L}$) under the global transformation $\psi \to \psi' = e^{i\alpha}\psi$. According to Noether's theorem [27], this global symmetry is associated with a conserved quantity, which in this case corresponds to the hypercharge Y. In addition, the transformation is also Lorentz invariant.

Local gauge invariance and gauge fields

Promoting the symmetry to a local one, $\alpha \to \alpha(x)$, breaks the invariance of the Lagrangian. Gauge invariance can be restored by introducing a vector field A_μ that transforms as

$$A_\mu \to A'_\mu = A_\mu + \frac{1}{q}\partial_\mu\alpha(x)\,. \tag{2.2}$$

The ordinary derivative can then be replaced with the covariant derivative $D_\mu = \partial_\mu + iqA_\mu$. The resulting gauge-invariant Lagrangian becomes

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}\,, \tag{2.3}$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the field-strength tensor. In four-dimensional spacetime, renormalisability[1] requires that all terms in the Lagrangian have mass dimension less than or equal to four. This condition forbids explicit mass terms for gauge fields, such as $m_A^2 A_\mu A^\mu$. As a result, gauge bosons associated with exact local symmetries must be massless. In this framework, the field A_μ can be identified with the photon.

Electroweak and strong interactions

Having discussed the Abelian $U(1)_Y$ symmetry, we now turn to the non-Abelian $SU(2)_L$ sector. The group $SU(2)$ has three generators, represented by the Pauli matrices in the fundamental representation. Gauge invariance under $SU(2)_L$ introduces three massless gauge fields, commonly denoted W^1_μ, W^2_μ, and W^3_μ, which couple exclusively to left-handed fermions. The associated conserved quantum number in the theory is the weak isospin.

Together, the $SU(2)_L \times U(1)_Y$ symmetries form the electroweak theory, which contains four gauge fields in total. Prior to symmetry breaking, gauge invariance requires all these fields to be massless. The physical electroweak bosons—the charged $W^\pm$, the neutral Z

[1] Renormalisation allows divergences to be systematically absorbed into redefinitions of physical parameters, yielding finite predictions for observables.

boson, and the photon—arise only after spontaneous symmetry breaking as linear combinations of the original gauge fields. The mechanism by which three of these bosons acquire mass while leaving the photon massless is discussed in the following subsection.

The strong interaction is described by the $SU(3)_C$ gauge symmetry. The group has eight generators, represented by the Gell-Mann matrices in the fundamental representation, corresponding to eight massless gauge bosons known as gluons. This symmetry implies the conservation of colour, a quantum number carried by quarks and gluons. Gluons transform in the adjoint representation of $SU(3)_C$ and therefore carry combinations of colour and anti-colour, allowing them to interact not only with quarks but also with each other. Observable particles must form colour-singlet states under $SU(3)_C$. As a consequence, quarks and gluons are not observed in isolation but only as colour-neutral bound states, such as mesons and baryons. This phenomenon, known as colour confinement, is a defining feature of QCD.

Spontaneous symmetry breaking

A key question is how the experimentally observed massive particles acquire their mass, given that gauge invariance forbids explicit mass terms for the gauge bosons.

The SM is a chiral theory, as established by the Wu experiment [28], meaning that left- and right-handed fermions transform differently under the electroweak symmetry. Because of this chiral structure, a direct mass term for fermions is not invariant under $SU(2)_L \times U(1)_Y$ transformations. The solution is to introduce a scalar field, the Higgs field ϕ, which transforms appropriately under the electroweak symmetry. This allows additional terms in the Lagrangian: the kinetic term for the Higgs field, its potential term $V(\phi)$, and Yukawa interactions between the Higgs and the fermion fields.

The Higgs potential is constructed with all renormalisable terms allowed by dimensional analysis, resulting in the characteristic "Mexican hat" shape:

$$V(\phi) = \mu^2\phi^*\phi + \lambda(\phi^*\phi)^2 \,. \tag{2.4}$$

The parameter λ must be positive to ensure that the potential is bounded from below. If $\mu^2 > 0$, the potential has a global minimum at $\phi = 0$ and the symmetry remains unbroken. If $\mu^2 < 0$, the minima occurs at $\phi \neq 0$.

The vacuum expectation value of the Higgs field corresponds to the minimum of this potential. Choosing a gauge such that the imaginary component vanishes ($\varphi_2 = 0$), the field acquires

$$\varphi_1 = \sqrt{\frac{2\lambda}{\mu^2}} \equiv v \,, \tag{2.5}$$

breaking the symmetry spontaneously. As a consequence, the gauge eigenstates mix, and the physical vector bosons correspond to the mass eigenstates of the theory.

The gauge bosons are given by the linear combinations

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp W_\mu^2)\,, \tag{2.6}$$

$$Z_\mu = W_\mu^3 \cos\theta_W - B_\mu \sin\theta_W\,, \tag{2.7}$$

$$A_\mu = W_\mu^3 \sin\theta_W + B_\mu \cos\theta_W\,, \tag{2.8}$$

where $\cos\theta_W = \frac{g}{\sqrt{g^2+g'^2}}$, and g and g' denote the $SU(2)_L$ and $U(1)_Y$ gauge couplings, respectively.

After spontaneous symmetry breaking, three of the four electroweak gauge bosons acquire mass, while one remains massless. The charged $W^\pm$ and the neutral Z bosons obtain masses proportional to the vacuum expectation value v of the Higgs field:

$$m_W = \frac{1}{2}gv\,, \tag{2.9}$$

$$m_Z = \frac{1}{2}\sqrt{g^2+g'^2}v\,, \tag{2.10}$$

with the two masses related by $m_Z = \frac{m_W}{\cos\theta_W}$.

The remaining combination corresponds to the massless photon A_μ, associated with the unbroken $U(1)_Q$ symmetry of electromagnetism, where $Q = T_L^3 + \frac{Y}{2}$ is the electric charge.

Simultaneously, fermions acquire mass through Yukawa couplings, which connect the fermion fields to the Higgs scalar:

$$\mathcal{L}_{\text{Yukawa}} = -y_f \bar{\psi}_L \phi \psi_R\,, \tag{2.11}$$

where y_f is the Yukawa coupling constant for each fermion. After symmetry breaking, this term generates a mass term $m_f = y_f v/\sqrt{2}$ for the corresponding fermion.

In addition, the spontaneous breaking of the electroweak symmetry predicts the existence of a physical scalar excitation of the Higgs field, known as the Higgs boson. Its properties and experimental manifestations are discussed in the following section.

It is also hypothesised that, in the early Universe, the Higgs field was initially at the top of its potential ($\varphi_1 = \varphi_2 = 0$) and subsequently rolled down to its minimum. This process may have contributed to cosmological inflation and the rapid cooling of the Universe. Understanding the dynamics of this transition, remains essential for determining the stability of the vacuum and the possibility that we reside in a metastable minimum.

2.2 Higgs boson properties

The Higgs boson is the physical excitation of the scalar Higgs field introduced in the SM to enable mass generation while preserving local gauge invariance in a chiral theory. In the SM, the Higgs boson is a spin-0 scalar particle with no electric or colour charge and positive parity.

The Higgs boson mass is not predicted a priori within the SM. However, it is related to the parameters of the Higgs potential through

$$m_H = \sqrt{2\lambda} v \,, \tag{2.12}$$

where λ is the Higgs self-coupling and v is the vacuum expectation value of the Higgs field.

Fermion masses arise from Yukawa interactions with the Higgs field. For each fermion f, the mass is given by

$$m_f = \frac{y_f v}{\sqrt{2}} \,, \tag{2.13}$$

where y_f is the corresponding Yukawa coupling. Since the Yukawa couplings are free parameters of the theory, knowledge of the fermion masses alone does not constrain the Higgs boson mass.

The SM predicts explicit interactions of the Higgs boson with fermions and massive gauge bosons. These include decays of the form $H \to f\bar{f}$ and $H \to VV^*$, where V denotes the massive weak gauge bosons $W^\pm$ or Z. The Higgs couplings are proportional to the masses of the interacting particles,

$$g_{Hff} \propto m_f, \qquad g_{HVV} \propto m_V^2 \,. \tag{2.14}$$

As a consequence, the Higgs boson couples most strongly to the heaviest fermions and to the weak gauge bosons, particularly the top and bottom quarks and the $W^\pm$ and Z bosons. In contrast, decays into light fermions are strongly suppressed and challenging to observe experimentally.

Since the Higgs boson carries no electric or colour charge, it does not couple directly to photons or gluons at tree level. Nevertheless, interactions such as $H \to \gamma\gamma$, $H \to Z\gamma$ or $H \to gg$ occur through loop diagrams involving virtual charged or coloured particles. Although loop-induced processes are suppressed relative to tree-level interactions, they are susceptible to contributions from physics beyond the Standard Model, making them powerful probes for new phenomena.

2.2.1 Production mechanisms

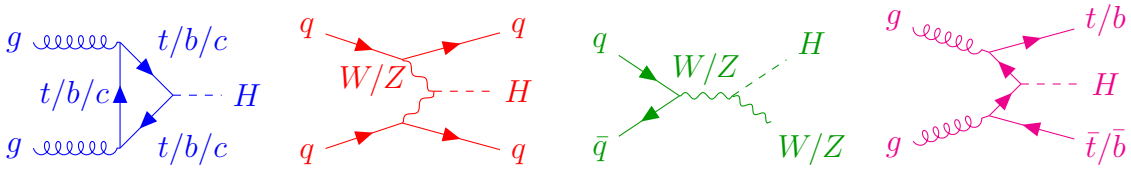
The production mechanisms of the Higgs boson depend strongly on the type of collider, the nature of the colliding particles, and the centre-of-mass energy of the interaction. At the Large Hadron Collider, described in Section 3.1.2.5, Higgs bosons are produced in proton-proton (pp) collisions at multi-TeV energies, where several production modes contribute with different cross-sections and experimental signatures.

In particle physics, the production rate of a given process is quantified by its cross-section, σ, which has units of area. Physically, the cross-section represents the effective target "size" that a particle presents to the incoming beam for a specific interaction. The integrated luminosity $\mathcal{L}$ quantifies the total particle flux delivered by the collider over a given period. It has units of inverse area (typically fb^{-1}) and represents the effective number of collisions available for producing a given process. In collider experiments, the number of events N

produced for a given process is proportional to the product of the cross-section and the integrated luminosity:

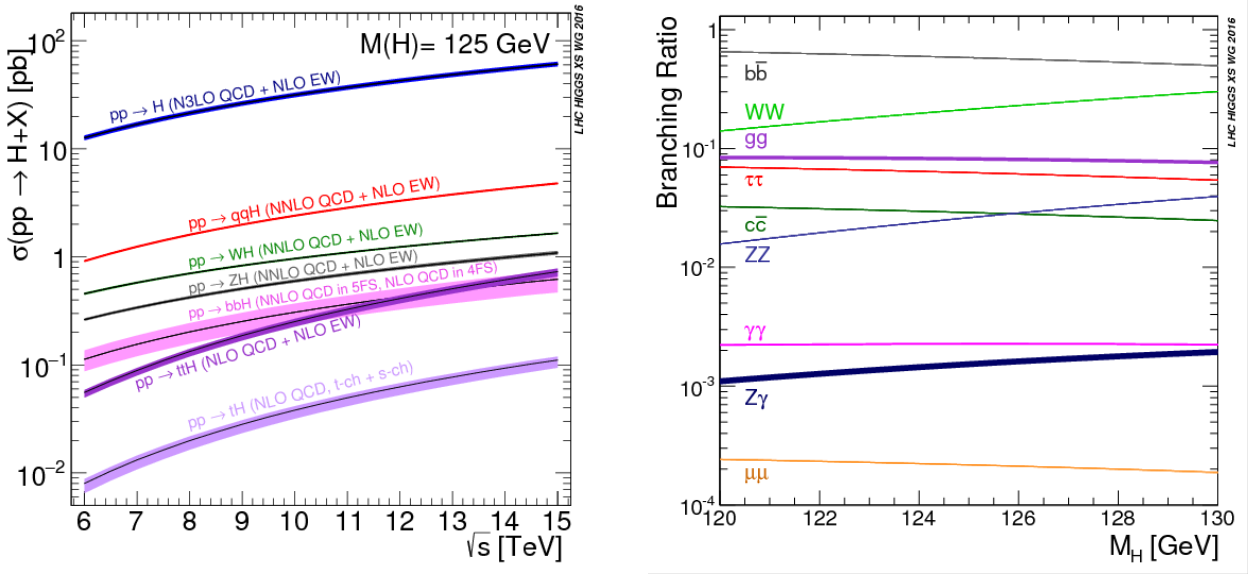
$$N = \sigma \times \mathcal{L}\,. \tag{2.15}$$

Hence, processes with larger cross-sections are more likely to occur and are generally easier to observe experimentally.



(a) ggF (b) VBF (c) VH (d) $t\bar{t}H$

Figure 2.2: Main Higgs boson production modes at the LHC: (a) gluon-gluon fusion (ggF), (b) vector boson fusion (VBF), (c) associated production with a vector boson (VH), and (d) associated production with a top-quark pair ($t\bar{t}H$).



(a) Higgs production cross-sections (b) Higgs decay modes

Figure 2.3: Summary of the Higgs boson production cross-sections at different centre-of-mass energies and of the decay mode branching ratios as a function of the Higgs boson mass. Figures from Ref. [29].

The Feynman diagrams corresponding to the four dominant Higgs production modes at a pp collider are shown in Fig. 2.2, ordered by decreasing production cross-section. The cross-section of these production modes for a pp collider to a Higgs boson for $m_H = 125$ GeV depending on the centre-of-mass energy $\sqrt{s}$ are shown in Fig. 2.3a.

The dominant mechanism is gluon-gluon fusion (ggF), which proceeds via a loop diagram with a predominantly heavy-quark circulating in the loop. Although loop-induced processes are generally suppressed relative to tree-level interactions, this suppression is compensated mainly by the large Yukawa coupling of the top quark to the Higgs boson. Furthermore, gluons carry a significant fraction of the proton's momentum at LHC energies, which enhances the ggF production rate. As a result, ggF accounts for the majority of Higgs boson production at the LHC.

The second most important production channel is vector boson fusion (VBF), whose cross-section is approximately an order of magnitude smaller than that of ggF. This process is characterised by the presence of two energetic quarks in the final state, which are observed in the detector as forward jets[2], typically separated by a large rapidity gap. The distinctive event topology of VBF provides a powerful handle for background suppression.

Another relevant production mechanism is the associated production of a Higgs boson with a vector boson ($W^{\pm}$ or Z), commonly referred to as the VH channel. Although its cross-section is smaller than those of ggF and VBF, the presence of a charged lepton or missing transverse energy from the vector boson decay offers clean experimental signatures, particularly for Higgs decays into bottom quarks.

Finally, Higgs bosons can be produced in association with heavy quarks, most notably top quark pairs ($t\bar{t}H$), and to some extent bottom quark pairs. While these processes have the smallest production rates, they provide direct sensitivity to the Yukawa couplings of the Higgs boson to the third-generation of fermions and therefore play a crucial role in testing the SM.

2.2.2 Decay modes

Unstable particles decay into lighter, more stable states. The decay of a particle is governed by the set of kinematically allowed final states and by the strength of its interactions with those states. For a given decay channel i, the branching ratio is defined as

$$\mathrm{Br}_i = \frac{\Gamma_i}{\Gamma_{\mathrm{tot}}}\,, \tag{2.16}$$

where Γ_i is the partial decay width of channel i and Γ_{tot} is the total decay width of the particle. Both quantities can be calculated theoretically using the corresponding Feynman diagrams and perturbative quantum field theory. By definition, the sum of all branching ratios satisfies $\sum_i \mathrm{Br}_i = 1$.

Figure 2.3b shows the SM predictions for the Higgs boson branching ratios as a function of its mass. For a Higgs boson mass of approximately 125 GeV, the dominant decay mode is into a bottom-antibottom quark pair ($H \to b\bar{b}$), followed by decays into pairs of weak gauge bosons, $H \to WW^*$ and $H \to ZZ^*$, where one of the bosons is off-shell. Other relevant decay channels include $H \to gg$, $H \to \tau^+\tau^-$, and $H \to c\bar{c}$.

Rare decay modes such as $H \to \gamma\gamma$, $H \to Z\gamma$, and $H \to \mu^+\mu^-$ have branching ratios below 10^{-2}. Although experimentally challenging due to their low rates, the first two channels

[2]Jets are described thoroughly in Chapter 4.

are of particular interest because they provide clean experimental signatures and sensitivity to loop-level contributions. Decays of the Higgs boson into first-generation fermions are highly suppressed, making direct measurements of their Yukawa couplings a key objective for future collider experiments.

The choice of the decay channel strongly influences the experimental sensitivity to Higgs production and properties.

2.2.3 Experimental measurements

Following the theoretical prediction of the Higgs field and the associated mechanism responsible for generating the masses of the SM particles in the 1960s by Brout, Englert, and Higgs [5–9], an experimental excess consistent with a Higgs boson was observed nearly five decades later. In July 2012, the ATLAS and CMS collaborations jointly announced the discovery of a new particle with properties compatible with those predicted by the SM with a significance of 5.9 and 5.8 standard deviations[3] [10, 11]. To date, all measurements agree with SM predictions within uncertainties. A summary of the most important experimental findings is presented below.

The SM predicts a massive scalar boson with quantum numbers $J^P = 0^+$, whose couplings to fermions and gauge bosons are proportional to their masses. Once the Higgs boson mass is fixed, its production cross-sections and decay branching ratios are fully determined within the SM.

The discovery of the Higgs boson relied primarily on the so-called golden decay channels, which provide a high signal-to-background ratio and excellent mass resolution: $H \to ZZ^* \to 4\ell$ and $H \to \gamma\gamma$. The most precise measurement of the Higgs boson mass obtained by combining Run 1 and Run 2 results from the ATLAS experiment [30] is

$$m_H = 125.11 \pm 0.09 \text{ (stat)} \pm 0.06 \text{ (syst) GeV}\,. \tag{2.17}$$

The observation of the Higgs boson in the diphoton channel constrains its spin to be either 0 or 2, since a spin-1 particle cannot decay into two photons (Landau-Yang theorem [31, 32]). Further discrimination between these hypotheses is achieved through studies of angular correlations in the $H \to ZZ^* \to 4\ell$ final state. These measurements strongly favour the spin-parity assignment $J^P = 0^+$, in agreement with the SM prediction [33].

More than ten years after its discovery, the Higgs boson has been observed to decay into massive gauge bosons, photons (through loop-induced processes), and third-generation fermions, all with high statistical significance.

Beyond the observation of individual decay modes, a central goal of the Higgs physics programme is to test the structure of the Higgs interactions and to quantify possible deviations from SM expectations. In experimental analyses, such deviations are commonly parametrised in terms of coupling strength modifiers and signal strengths.

[3] A 5σ significance corresponds to a probability of about 3×10^{-7} that the observed excess is a fluctuation of the background, which is the conventional threshold in particle physics for claiming a discovery.

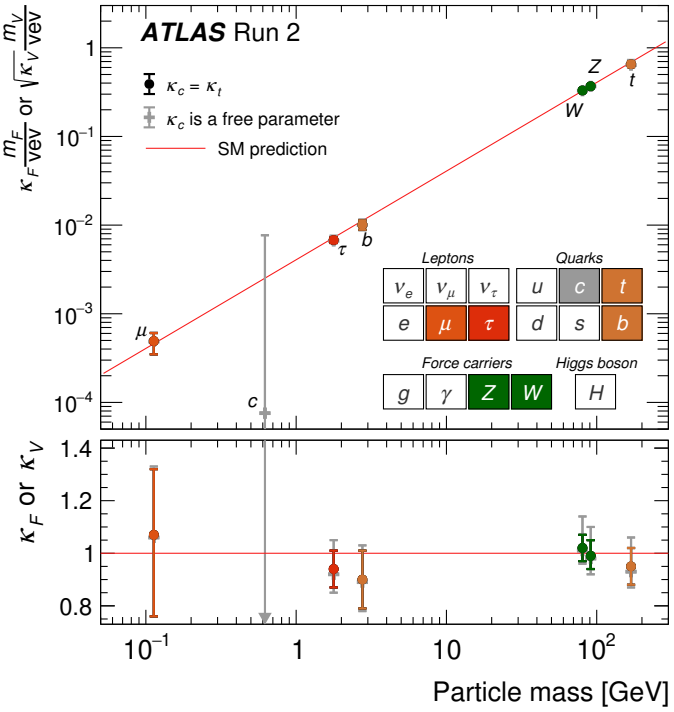


Figure 2.4: Measured Higgs coupling strength modifiers as a function of the particle mass. The couplings are scaled separately for fermions and vector bosons as $\kappa_F \frac{m_f}{v}$ and $\sqrt{\kappa_V}\frac{m_V}{v}$, respectively. Figure from Ref. [34].

The coupling strength modifier κ_i is defined as the ratio between the Higgs coupling to a given particle i and its SM expectation,

$$\kappa_i \equiv \frac{g_i}{g_i^{\text{SM}}} \,. \tag{2.18}$$

In the SM, all coupling modifiers are equal to unity, while deviations from $\kappa_i = 1$ would indicate the presence of physics beyond the Standard Model.

Figure 2.4 shows the experimentally measured Higgs coupling strengths as a function of the mass of the particle to which the Higgs couples. The couplings are rescaled such that a linear dependence on the particle mass is expected in the SM. All measured channels are consistent with this hypothesis within current experimental uncertainties. First measurements of couplings to second-generation fermions, such as the charm quark and the muon, have also been performed. While these measurements are consistent with SM predictions, they are currently limited by relatively large experimental uncertainties.

Complementary information is provided by measurements of the signal strength parameter μ, defined as the ratio of the observed production cross-section times branching ratio to

the corresponding SM prediction,

$$\mu = \frac{\sigma \times \mathrm{Br}}{(\sigma \times \mathrm{Br})_{\mathrm{SM}}} \,. \tag{2.19}$$

The signal strength can be measured for specific production and decay modes, denoted as μ_i^f, where i labels the production mechanism and f the decay channel. Figure 2.5 summarises the measured signal strengths for the main Higgs boson production mechanisms and decay channels. The results show good overall agreement with SM expectations and provide a concise overview of the current experimental precision achieved in Higgs boson measurements.

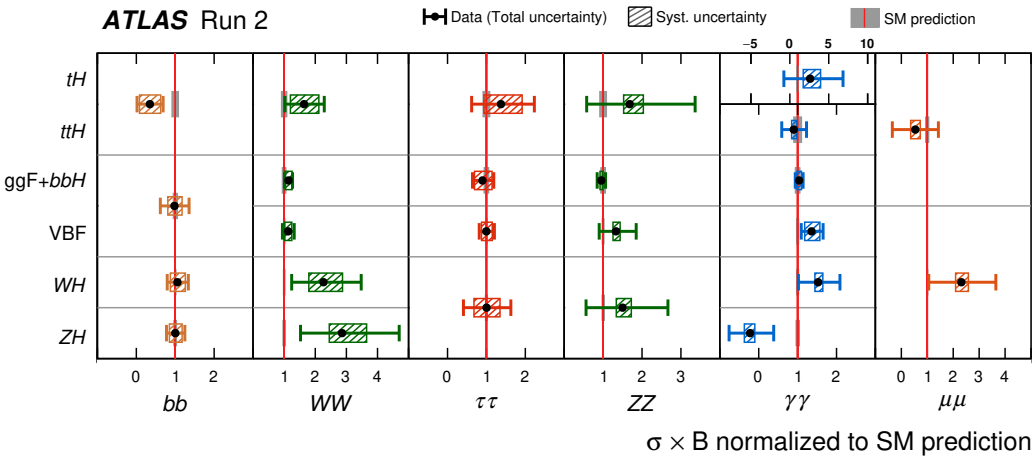


Figure 2.5: Summary of the measured signal strengths, defined as the ratio of the observed event rate to the Standard Model prediction, for different Higgs boson production and decay processes. Figure from Ref. [34].

Beyond measurements of individual decay modes and couplings, the Higgs sector also offers a unique window into the shape of the Higgs potential. The parameters λ and μ in Eq. 2.4 are closely connected to the dynamics of electroweak symmetry breaking and may have played a role in the evolution of the early Universe. While studies of Higgs self-interactions and the Higgs potential lie beyond the scope of this thesis, they constitute a major objective of future collider programmes, with important implications for vacuum stability and cosmology.

2.3 Beyond the Standard Model

Despite its remarkable success, the SM is widely regarded as an effective theory rather than a complete description of nature. Although all of its predictions have been experimentally

confirmed within their expected precision, the SM exhibits several intrinsic limitations and fails to account for several well-established physical phenomena. While the primary scope of this thesis does not focus on Beyond the Standard Model (BSM) physics itself, some of the parts discussed are sensitive to possible BSM effects. Therefore, a brief overview of the main problems of the SM is provided.

One fundamental limitation of the SM is its inability to incorporate **gravity** within the framework of quantum field theory. General relativity, which successfully describes gravity at macroscopic scales, is a non-renormalisable theory when quantised [35], preventing the calculation of meaningful observables at high energies. While general relativity provides an accurate description of gravity at macroscopic scales, its extrapolation to the quantum regime is not well-defined. Various extensions of the SM, such as supersymmetric theories, predict new particles that could mediate gravitational interactions at the quantum level. An example is the gravitino, the supersymmetric partner of the graviton. To date, however, no experimental evidence for such particles has been observed.

Another major indication of physics beyond the SM arises from **neutrino masses**. In the SM, neutrinos are assumed to be massless; however, experimental observations of neutrino oscillations demonstrate that neutrinos have non-zero masses [36, 37]. Neutrino oscillations occur because flavour eigenstates are superpositions of mass eigenstates, causing neutrinos to change flavour as they propagate. So far, the absolute mass scale of neutrinos remains unknown, but upper limits have been set by direct kinematic measurements and cosmology [38, 39].

Several cosmological and astrophysical observations provide strong evidence for the existence of **dark matter**, a form of matter that does not interact electromagnetically and is detected primarily through its gravitational effects. One of the clearest examples is the Bullet Cluster, a system of two colliding galaxy clusters [40]. Gravitational lensing measurements show that the majority of the mass is spatially separated from the hot baryonic gas observed in X-ray emission, indicating a dominant non-baryonic matter component. Further evidence comes from observations of the Cosmic Microwave Background (CMB) [41]. The angular power spectrum of its temperature anisotropies is in excellent agreement with cosmological models that include dark matter, allowing precise constraints on its density in the early Universe. At galactic scales, the rotation curves of spiral galaxies also support the presence of dark matter [42]. Stars in the outer regions rotate faster than expected from visible matter alone. Although alternative explanations such as Modified Newtonian Dynamics exist [43], dark matter remains the most widely accepted interpretation.

Little is known about the nature of **dark energy**, which is introduced to account for the observed accelerated expansion of the Universe, first inferred from measurements of Type Ia supernovae [44]. Its energy density is tightly constrained by cosmological observations, including supernova surveys, measurements of the Cosmic Microwave Background, and large-scale structure data [45, 46]. However, estimates of the vacuum energy density from quantum field theory exceed the observed value by up to 10^{50} (or more) orders of magnitude, giving rise to the so-called cosmological constant problem [47].

The **baryon asymmetry** of the Universe refers to the observed dominance of matter over antimatter. Within the SM, there is no mechanism capable of generating the observed

magnitude of this asymmetry. Any successful explanation must satisfy the three Sakharov conditions [48]: violation of baryon number, violation of charge-parity (CP), and a departure from thermal equilibrium. Baryon number violation is realised in the SM through non-perturbative electroweak processes, and CP violation is present, but its magnitude is insufficient to account for the observed asymmetry. Furthermore, for the measured Higgs boson mass, the electroweak phase transition in the SM is a smooth crossover rather than a first-order phase transition, preventing the generation of the required out-of-equilibrium conditions.

The SM includes sources of CP violation via the CKM matrix in the quark sector. However, the strong **CP problem** [49] arises from the absence of observable CP violation in QCD, despite the presence of a CP-violating term in the QCD Lagrangian. The corresponding CP-violating parameter is constrained experimentally to be extremely small, raising the question of why it is so close to zero.

Chapter 3

The experimental setup

To this point, the theoretical framework underlying this work has been presented. This chapter introduces the experimental setup, describing how data are produced, collected, and prepared for physics analyses.

The chapter begins with a brief historical overview of CERN, followed by a description of the Large Hadron Collider and the acceleration of protons up to the point of collision. The ATLAS detector is then presented in detail, with particular emphasis on its main subdetectors and trigger system. Finally, the procedures used to handle both simulated and real data are described, together with an overview of the reconstruction of the physics objects used in this thesis.

3.1 CERN

3.1.1 Early-stages: Brief history of CERN

After World War II, Europe was politically divided and economically weakened, while nuclear physics had demonstrated its decisive role in the outcome of the conflict. Within this context, several scientists began advocating for the creation of a European research laboratory that would enable scientific collaboration across borders and share the substantial costs of large-scale experimental facilities. The idea first emerged in 1949, gaining real momentum two years later. At the UNESCO intergovernmental meeting held in Paris in December 1951, a resolution was adopted to establish a **European Council for Nuclear Research**. Two months later, representatives from 11 countries signed an agreement creating a provisional council, using CERN as an acronym for its French name. Sixteen months afterwards, the CERN Convention was formally approved by twelve founding Member States.

The CERN Convention [50] explicitly stated that the organisation would not engage in research for military purposes and that its scientific results would be openly available. It also defined the financial contributions of Member States according to their national income and outlined procedures for membership. Geneva was chosen as the host city for its central location in Europe and Switzerland's political neutrality during the World Wars. While

the laboratory infrastructure was being built on the Meyrin site, the theoretical group was temporarily hosted at the Niels Bohr Institute in Copenhagen.

Over more than seven decades, CERN has grown into the world's leading centre for particle physics. Today, it includes 25 Member States and several Associate Members, maintaining its founding spirit of international cooperation in fundamental research. Its accelerator and detector programmes have shaped the field from the 1950s to the present day.

CERN's experimental program began with the Synchrocyclotron (SC) [51], which became operational in 1957. It accelerated protons up to energies of 600 MeV (about 80% of the speed of light) and served both as a training ground for early generations of scientists and as a reliable facility for obtaining initial experimental data. Among its notable achievements was the observation of the rare pion decay into an electron and a neutrino [52]. As larger and more powerful accelerators were built, the SC's focus shifted toward nuclear physics until its decommissioning in 1990.

The flagship project of CERN's early decades was the Proton Synchrotron (PS) [53], operational since 1959. The PS accelerated protons to 25 GeV in a 628 m circular ring, an impressive technical feat at the time. The accelerator supplied various particle beams to multiple experiments, playing a central role in several fundamental discoveries. Among its most significant results were the first observation of antinuclei [54] in 1965 and the findings of the Gargamelle experiment [55] in 1973. Gargamelle, a heavy-liquid bubble chamber detector, provided the first evidence of neutral currents [56], a crucial experimental confirmation of the electroweak theory. In this setup, charged particles produced visible ionisation tracks in the detector's heavy liquid ($CBrF_3$), while neutrinos interacted indirectly through their secondary products after interacting with the medium. The neutrino beam used for these studies was generated when protons from the PS collided with a beryllium target, producing pions and kaons that decayed, emitting neutrinos with energies between 1 and 10 GeV.

Parallel to these advances, detector technology was evolving rapidly. In 1968, Georges Charpak developed the multiwire proportional chamber [57], a gas-filled detector with multiple readout wires connected to amplifiers, allowing for counting rates thousands of times faster than previous techniques. This innovation revolutionised particle detection and paved the way for the modern era of high-rate experiments.

The next major milestone was the Super Proton Synchrotron (SPS) [58, 59], a 7 km circular accelerator capable of accelerating protons up to 400 GeV. Operational since 1976, the SPS has served as one of CERN's most versatile machines, accelerating not only protons but also antiprotons, electrons, positrons, and heavy ions throughout its lifetime. The SPS played a central role in several landmark discoveries in particle physics. In 1983, the UA1 and UA2 experiments observed the W and Z bosons [60], confirming the electroweak unification mechanism and leading to the 1984 Nobel Prize in Physics [61]. From 1986 onward, the SPS was also used to accelerate heavy ions to explore the formation of the quark–gluon plasma [62], a state of matter thought to have existed microseconds after the Big Bang.

Building on these successes, CERN entered the era of precision electroweak physics with the Large Electron–Positron Collider (LEP) [63] in 1989. Built in a 27 km circular tunnel, LEP collided electrons and positrons at four interaction points equipped with the ALEPH, DELPHI, L3, and OPAL detectors. Initially operating near the Z boson reso-

nance at 100 GeV, it later increased its energy to enable the production of W^+W^- pairs. The resulting measurements of the Z and W bosons' masses, widths, and couplings reached unprecedented precision and confirmed the existence of only three generations of matter particles. LEP also provided indirect hints of the Higgs boson, although its discovery would have to wait for the next generation of accelerators. After more than a decade of successful operation, LEP was decommissioned in 2000 to make room for the Large Hadron Collider (LHC), constructed in the same tunnel.

The LHC [12] began operation in 2008, initially delivering proton–proton collisions at a centre-of-mass energy of 7 TeV. Its design allows collisions at energies several times higher than those of any previous accelerator, enabling physicists to probe the most fundamental constituents of matter and the forces that govern them. Since its commissioning, the LHC has completed three major data-taking runs, each increasing both the collision energy and the luminosity. Four primary interaction points along the ring host the main detectors: ATLAS, CMS, ALICE, and LHCb [64–67]. ATLAS (A Toroidal LHC ApparatuS) and CMS (Compact Muon Solenoid) experiments are general-purpose detectors designed to explore a wide range of physics phenomena, from the search of new particles to precise measurements within the Standard Model. The ALICE (A Large Ion Collider Experiment) detector specialises in heavy-ion collisions (Pb, O, or Ne), studying the properties of the quark–gluon plasma and characterising the conditions under which this state can be created. LHCb (Large Hadron Collider bottom) focuses on flavour physics, investigating the behaviour of bottom and charm quarks to understand the matter–antimatter asymmetry in the Universe.

In 2012, the ATLAS and CMS Collaborations jointly observed a new particle consistent with the Higgs boson [10, 11], completing the final missing piece of the Standard Model and marking one of the most significant achievements in modern physics. Since then, the LHC program has continued to deliver significant results, including new exotic hadronic states such as tetraquarks and pentaquarks, studies of CP violation in baryon decays, and detailed measurements of top quark and electroweak processes.

Looking ahead, CERN is preparing the High-Luminosity LHC upgrade (HL-LHC) [68], scheduled to begin after Long Shutdown 3. The upgrade will significantly increase the collider's luminosity, expanding its reach to rare processes and enabling more precise measurements at energies up to 7 TeV per beam.

3.1.2 Following a proton from *birth* to collision

The LHC beam is not continuous but organised in discrete bunches due to the oscillating nature of the electromagnetic fields used for acceleration. Each LHC fill can contain up to 2808 bunches, separated by 25 ns, with a few empty slots reserved for calibration and machine protection. Each bunch contains roughly 1.15×10^{11} protons, meaning that a single fill (occurring approximately every eight hours) requires on the order of 3×10^{14} protons in total. To supply this enormous number of protons, CERN begins with a small container of molecular hydrogen (H_2), which contains far more protons than needed for years of LHC operation.

Protons are extracted from hydrogen gas and injected into a chain of accelerators, where

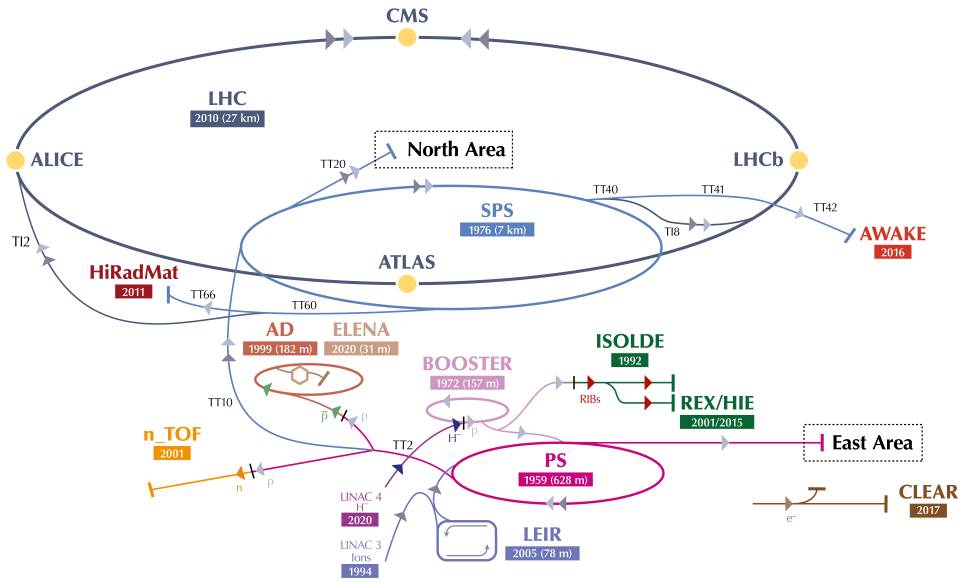


Figure 3.1: Schematic overview of the CERN accelerator complex, showing the pre-accelerators used during Run 3: LINAC4, the PS Booster, the Proton Synchrotron (PS), the Super Proton Synchrotron (SPS), and the Large Hadron Collider (LHC), along with the four main detectors. For each subsystem, the year of operation is indicated, and for the circular machines, the circumference length is also shown. Figure adapted from Ref. [69].

they are progressively brought to higher energies. As charged particles, they are accelerated using radiofrequency (RF) cavities, which generate oscillating electric fields between electrodes. When the electric field is oriented in the direction of motion as the beam passes through a cavity, the protons receive a small increment of energy. By passing through many such cavities in succession, their energy is gradually increased until they are ready for injection into the LHC ring, where they ultimately collide with a counter-rotating beam.

To help visualise the entire acceleration chain, Fig. 3.1 provides a schematic overview of the CERN accelerator complex. The figure highlights each subsystem: LINAC4, the PS Booster, the Proton Synchrotron (PS), the Super Proton Synchrotron (SPS), and the Large Hadron Collider (LHC), as well as the four main detectors. In the following subsections, we will follow a proton from its extraction from hydrogen gas, through each stage of acceleration in these machines, and finally into the high-energy collisions in the LHC.

3.1.2.1 LINAC4

Since the start of Run 3, the LHC has been fed by the LINAC4 [70] linear accelerator (purple in Fig. 3.1). The process begins by ionising molecular hydrogen through electron bombardment. Protons from the resulting plasma are directed toward a Caesium cathode surface, which enhances electron emission and facilitates the formation of H^- ions. These negatively charged ions exit the source in 400 μs pulses with an initial energy of 45 keV,

forming a low-energy beam that passes through a small magnet to remove residual free electrons. The remaining stages of LINAC4 accelerate the beam linearly up to 160 MeV.

The first element downstream of the ion source is the Low Energy Beam Transport (LEBT) line, which adjusts the beam parameters (mainly focusing and alignment) to match the acceptance of the Radio-Frequency Quadrupole (RFQ) [71,72]. The RFQ, approximately 3 m long, simultaneously focuses, accelerates, and bunches the initially continuous beam into discrete packets. It consists of a four-vane quadrupole structure with vane tips modulated in a sinusoidal pattern of gradually increasing wavelength, matching the increasing particle velocity. After the RFQ, the beam energy reaches 3 MeV and enters the chopper line, which contains a beam dump used to remove unwanted bunches and create empty time intervals, helping to stabilise the downstream beam. At these energies, the resulting radiation is minimal.

Three Drift Tube Linac (DTL) tanks then accelerate the beam to 50 MeV. The DTL uses RF fields phased to avoid decelerating the particles during the opposite half-cycle. The spacing between drift tubes increases along the accelerator to match the rising particle velocity, while permanent magnetic quadrupoles provide transverse focusing.

The next stage, a Cell-Coupled Drift Tube Linac, raises the beam energy to 102 MeV. It consists of seven modules, each containing three DTL-like cells separated by electromagnetic quadrupoles.

The final acceleration stage, the Pi-Mode Structure (PIMS), increases the H^- ion energy to 160 MeV through twelve resonant cavities. Above roughly 100 MeV, π-mode structures are more efficient than 0-mode configurations due to the relative phase advance between adjacent cells. RF power for all LINAC4 accelerating structures is supplied by klystrons [73], which convert DC power into high-frequency RF energy.

Finally, a 70 m transfer line connects the end of LINAC4 to the Proton Synchrotron Booster (PSB). Along this line, beam diagnostics continuously monitor beam parameters, and a beam dump is available for controlled beam disposal if needed.

3.1.2.2 Proton Synchrotron Booster

The Proton Synchrotron Booster (PS Booster, light pink in Fig. 3.1) [74, 75] accelerates protons to an energy of 2 GeV. Its primary purpose is to increase the proton intensity for injection into the next stage of the accelerator chain, the Proton Synchrotron (PS). At injection, a thin carbon foil strips the two electrons from the incoming H^- ions with minimal energy loss, producing a pure proton beam ready for acceleration.

The PS Booster is the first circular accelerator in the LHC injection chain, with a radius of 25 m. A circular design allows protons to pass through the same RF cavities multiple times, gaining energy on each pass. Maintaining a stable circular trajectory while increasing the particle energy is the main challenge: dipole magnets bend the beam along the curved path, and their magnetic field must increase in synchrony with the proton energy to preserve the orbit, giving the machine its name *synchrotron*. Transverse focusing is provided by quadrupole magnets to maintain a dense, stable beam and maximise luminosity.

A distinctive feature of the PS Booster is its four superimposed rings, which allow multiple

beams to be accelerated in parallel. This design increases total intensity and accommodates longer bunches. The beams are distributed among the four rings by pulsed magnets in the proton distributor system and are later recombined into a single plane for transfer.

Acceleration is achieved using a single RF cavity, while 32 dipole magnets and 48 quadrupoles bend and focus the beam, respectively. Once the target energy of 2 GeV is reached, the four beams are synchronised and recombined by a dedicated recombination system before injection into the Proton Synchrotron.

3.1.2.3 Proton Synchrotron

The Proton Synchrotron (PS, dark pink in Fig. 3.1) [76, 77] continues the acceleration of protons after the PS Booster, increasing their energy up to 26 GeV in a circular ring with a circumference of 628 m. Unlike the PS Booster, which has four stacked rings, the PS consists of a single larger ring, allowing a greater number of bunches to circulate simultaneously. Batches of bunches from the PS Booster are sequentially injected into the PS until the ring is filled, after which acceleration begins. A similar multi-step injection scheme is applied when transferring the beam to the following stages.

The PS uses an RF system to provide the accelerating electric field, while 277 conventional (room-temperature) magnets maintain the beam on its circular trajectory. This includes 100 dipoles that bend the protons along the ring and a combination of quadrupoles and higher-order multipoles that focus and shape the beam, preserving its stability and minimising emittance growth. Maintaining a high-quality, well-focused beam is crucial for subsequent acceleration stages and for achieving high luminosity at the LHC.

3.1.2.4 Super Proton Synchrotron

Protons enter the Super Proton Synchrotron (SPS, blue in Fig. 3.1) [58, 59] from the PS at an energy of 26 GeV and are further accelerated to 450 GeV in a circular ring with a circumference of 6.9 km. Each SPS fill typically requires three to four injection cycles from the PS to populate the ring fully with bunches. As in the PS, maintaining beam quality during acceleration is crucial to prevent emittance growth and to ensure a stable transfer to the LHC.

The SPS uses two RF systems: a 200 MHz system for acceleration and an 800 MHz system for longitudinal beam stabilisation. The higher-frequency system adjusts the bunch length to reduce instabilities and maintain well-defined proton packets throughout acceleration.

The accelerator contains a total of 1 317 magnets, of which 744 are dipoles that bend the beam along the circular orbit. The remaining magnets, including quadrupoles and higher-order multipoles, provide transverse focusing, beam shaping, and correction of orbit deviations.

Once protons reach 450 GeV, the beam is extracted and transferred to the LHC injection system, marking the final stage of pre-LHC acceleration.

3.1.2.5 Large Hadron Collider

The Large Hadron Collider (LHC, grey in Fig. 3.1) [12, 78] constitutes the final acceleration stage for protons in the CERN accelerator complex. It consists of a 27 km circular ring in which protons are injected in bunches from the SPS at an energy of 450 GeV over 12 successive cycles. The LHC employs a twin-aperture design that allows two counter-rotating beams to circulate within the same cryostat under ultra-high vacuum conditions, each beam accelerated synchronously.

The beams gain energy using 16 superconducting RF cavities operating at 400 MHz, reaching 6.8 TeV per proton and a total centre-of-mass energy $\sqrt{s}$ of 13.6 TeV during collisions. At this energy, protons travel at 0.999999990 times the speed of light in vacuum, completing 11 245 revolutions of the LHC ring per second. To maintain a stable circular trajectory, the beams are guided by powerful superconducting dipole magnets. Without superconducting technology, a conventional electromagnet system would require a ring many times larger to achieve the same energies; the LHC's design allows high-energy beams to be efficiently bent within the existing 27 km tunnel.

Beam steering and confinement rely on 1 232 superconducting dipole magnets, each about 15 m long and operating at 1.9 K. These niobium-titanium magnets generate magnetic fields up to 8.3 T, bending the beams with minimal energy loss due to superconductivity. Transverse focusing is provided by 392 superconducting quadrupole magnets, ensuring high beam quality and luminosity at the interaction points. The entire magnet system is cooled by a closed liquid-helium cryogenic circuit, insulated under high vacuum to minimise thermal losses.

There are four main interaction points along the ring, each equipped with a dedicated detector optimised for different physics goals. On either side of each detector, triplets of superconducting quadrupoles, known as insertion magnets, focus the beams to a transverse size of approximately 16 μm at the detector centres, maximising the probability of proton–proton collisions.

Although each bunch contains roughly 10^{11} protons, only a small fraction of these interact during each crossing. The instantaneous luminosity determines the collision rate and depends directly on beam intensity and focusing. As the beams circulate, interactions and beam degradation gradually reduce the number of collisions per crossing. After roughly eight hours, the luminosity becomes too low for efficient operation, and the remaining protons are safely extracted and dumped. This is done by first reducing the beam intensity and then directing the remaining protons into a massive graphite and concrete absorber, where their residual energy is safely dissipated.

In practice, collisions occur between the constituent partons, quarks and gluons, rather than the protons as a whole. These parton–parton interactions are responsible for the diverse range of particles observed in the detectors. The effective probability of interaction is quantified by the cross-section, which increases with energy as higher-energy partons can overcome the strong-force repulsion at small distances.

3.2 The ATLAS detector

The ATLAS detector (A Toroidal LHC ApparatuS) is one of the four large detectors located on the LHC ring. Along with CMS, ATLAS is a general-purpose detector: rather than being specialised in studying a single type of particle, it is designed to deliver high performance across a wide range of measurements, enabling it to explore many aspects of particle physics.

In its Technical report [64], the physics case was already presented clearly, highlighting its broader goals. For nucleus-nucleus collisions, the focus was on studying the hypothesised quark-gluon plasma. For proton-proton (pp) collisions, the detector requirements were driven by the need to perform precise SM measurements, while exploring new phenomena expected at the TeV scale. Accordingly, the physics program included QCD, electroweak, and flavour physics studies, as well as investigations of the top quark's properties. Regarding the Higgs boson, its mass was not yet known, so different search channels were prioritised based on its possible mass range. Other potential avenues for new physics included heavy gauge bosons (W' and Z'), supersymmetry, and gravitons.

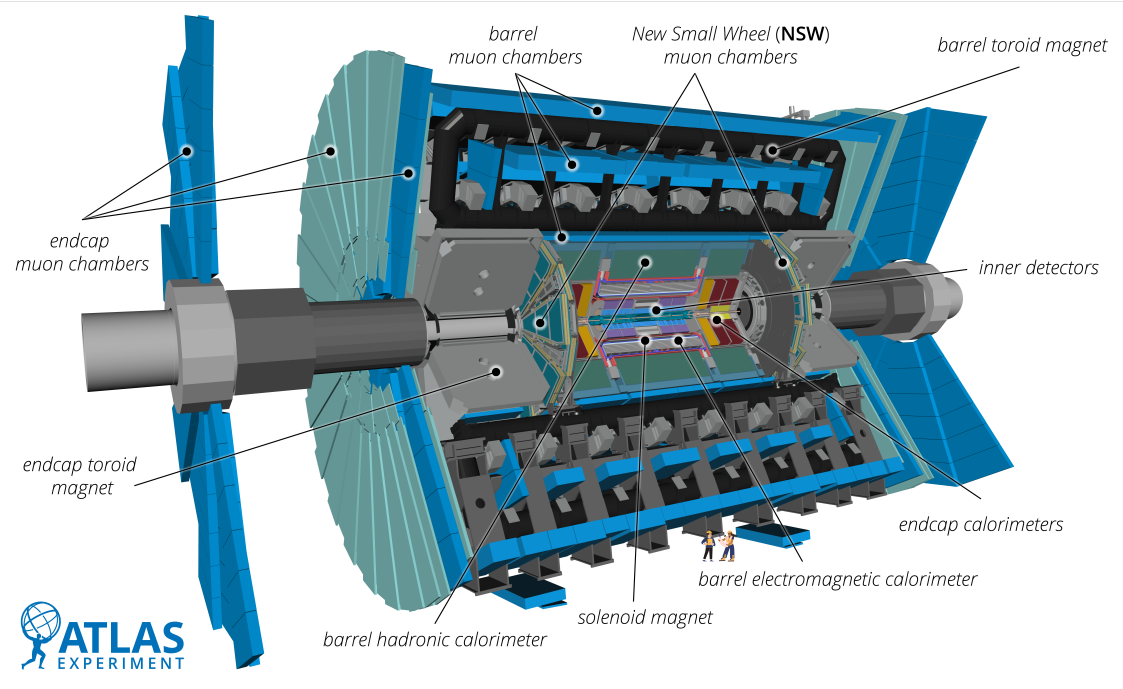


Figure 3.2: Schematic overview of the ATLAS detector and its subsystems. Figure from Ref. [79].

By merging ideas from two groups, EAGLE and ASCOT, the Technical Design reports [80] were approved, and, with that, ATLAS received formal approval in July 1997. Construction then began at individual institutions, later assembling the different parts in the ATLAS experiment pit, located at Point 1 of the LHC tunnel.

ATLAS has a cylindrical shape, with the centre coinciding with the Point of Interaction

(PoI) and surrounding the beam pipe. It is the largest detector at the LHC, with a length of 46 m, a diameter of 25 m, and a mass of 7 000 tonnes (see Fig. 3.2 for an illustration of its size and shape).

Different parts of the detector are specialised to detect various types of particles. The Inner Detector, located closest to the interaction point, measures charged-particle trajectories with very high resolution. The Calorimeter is divided into the Electromagnetic and Hadronic Calorimeters, and it is designed so that particles interact sufficiently with the medium to be stopped while recording the deposited energy. Relativistic muons have lifetimes long enough to pass through most of the detector without significant interaction. The final component of ATLAS, the Muon Spectrometer, is built specifically to detect this type of particle. The entire detector is immersed in a magnetic field, which causes charged particles to follow curved trajectories.

In this section, a summary of each detector component will be presented, along with an explanation of how the data are recorded. Many topics will not be covered, such as how the vacuum and cryogenics systems work.

3.2.1 Coordinate system

The ATLAS experiment uses a right-handed coordinate system with the origin defined as the nominal interaction point. An illustrative overview of the coordinate system is presented in Fig. 3.3. The beam direction defines the z-axis, with the positive direction pointing towards side-A of the detector (often referred to as the Airport side), and the negative direction towards side-C. The plane transverse to the beam corresponds to the x-y plane, where the positive x-axis points towards the centre of the ring, and the positive y-axis points upwards.

The azimuthal angle ϕ is measured around the beam axis, covering the range $(-\pi, \pi)$. The polar angle θ is measured from the beam axis. A commonly used variable instead of θ is the pseudorapidity, defined as $\eta = -\ln\tan(\theta/2)$. For $\theta = \pi/2$, $\eta = 0$, corresponding to a direction perpendicular to the beam pipe, while for $\theta = 0$, $\eta = \infty$. For massive particles, the rapidity is used instead, defined as $y = 1/2\ln[(E + p_z)/(E - p_z)]$. The angular distance between two particles is measured using both the azimuthal angle and the pseudorapidity, $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

The transverse momentum is often used instead of the full momentum, and is defined as the momentum component in the transverse plane $p_T = \sqrt{p_x^2 + p_y^2}$.

The Primary Vertex is chosen as the interaction vertex with the largest sum of transverse momentum, $\sum p_T$, corresponding to the hard-scatter interaction. Interactions in addition to the hard-scatter process of interest are referred to as *pile-up*.

Run 3 has so far delivered a large amount of physics data. Figure 3.4a shows the integrated luminosity delivered to and recorded by ATLAS during Run 3. The increase in luminosity is accompanied by a high level of pile-up, which poses additional challenges for event reconstruction, with an average of about 60 interactions per bunch crossing in Run 3, as illustrated in Fig. 3.4b.

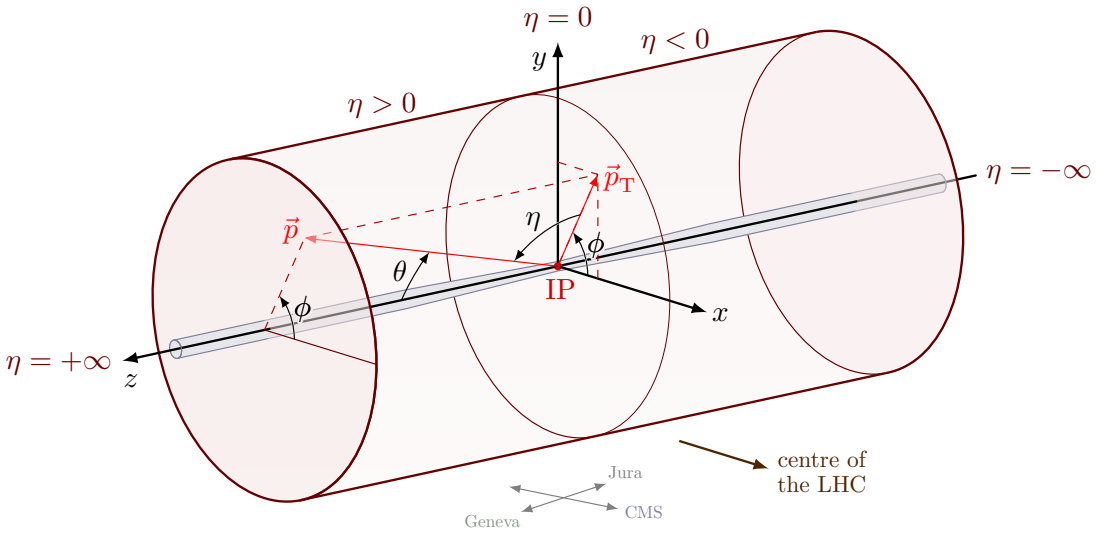
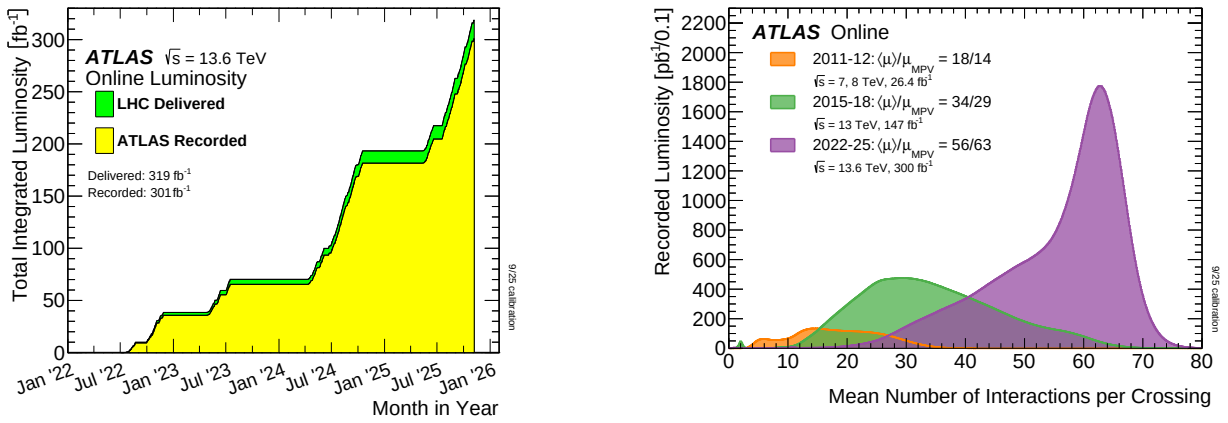


Figure 3.3: Overview of the ATLAS coordinate system, with the centre named Interaction Point (IP), the z-axis being along the beam, the positive x-axis pointing towards the centre of the LHC, and the positive y-axis towards the sky. A particle with momentum $\vec{p}$ is represented in red, its angle to the z-axis is θ, which can translate to pseudorapidity η, the projection to the xy-plane is the transverse momentum p_{T}, with an azimuthal angle of ϕ with respect to the x-axis. Figure adapted from Ref. [81].



(a) (b)

Figure 3.4: (a) Integrated luminosity over time delivered to (green) and recorded by ATLAS (yellow) during stable beams for pp collisions at 13.6 TeV centre-of-mass energy in LHC Run 3 (2022-2025). (b) Recorded luminosity of the mean number of interactions per crossing for pp collisions at 7 and 8 TeV centre-of-mass energy in Run 1, 13 TeV in Run 2, and 13.6 TeV in Run 3 (2022-2025). Figures from Ref. [82].

3.2.2 Detector characteristics

The information below is taken from Ref. [64] if not specified otherwise.

3.2.2.1 Magnet system

Magnets in the detector are used to bend the trajectories of charged particles, allowing precise determination of their momentum and charge. In ATLAS, two different types of superconducting magnet systems are employed: one solenoid and three toroids, operating at temperatures as low as 4.5 K. Both systems were optimised to minimise their interaction with the particles produced in the collisions.

The solenoid magnet is aligned along the beam axis, positioned between the Inner Detector and the Calorimeter. It provides a 2.0 T axial magnetic field to the Inner Detector. The solenoid design was optimised to maintain high calorimeter performance by minimising the material thickness between the solenoid and the calorimeter; for example, both share a common vacuum vessel. The magnet has an axial length of 5.8 m and inner and outer diameters of 2.46 m and 2.56 m, respectively. Such a strong magnetic field within a compact structure is achieved using 9 km of niobium–titanium superconductor embedded in pure aluminium stabiliser strips. The solenoid can be charged and discharged in about 30 minutes, but it requires approximately one day to cool back down to its operating temperature of 4.5 K after a warm-up cycle. The vacuum is maintained by the liquid-argon (LAr) cryogenic system.

To bend muons in the Muon Spectrometer, a barrel toroid and two end-cap toroids generate a toroidal magnetic field of up to 3.5 T.

The barrel toroid surrounds the calorimeter and consists of eight coils arranged symmetrically around the z-axis. Each coil is 25.3 m long, with inner and outer diameters of 9.4 m and 20.1 m, respectively, making it the largest toroidal magnet ever constructed. Cooling the cold mass to 4.6 K takes approximately 5 weeks. In the event of a fast dump, such as one triggered by a quench (a sudden loss of superconductivity leading to local heating), about 50 h are required to re-cool the system.

The end-cap toroids enhance the bending power in the forward regions, improving the momentum resolution for particles exiting the detector near the beam pipe. Each end-cap toroid is a rigid structure containing eight coil units, with an axial length of 5.0 m and inner and outer diameters of 1.65 m and 10.7 m, respectively.

3.2.2.2 Inner Detector

The Inner Detector (ID), shown in Fig. 3.5, is responsible for reconstructing the paths of charged particles produced in pp collisions. It is the innermost detector system in ATLAS, located closest to the beam line. As charged particles traverse the detector, they interact with its active layers, producing discrete hits in each subsystem. By combining these hits across layers, sophisticated tracking algorithms can reconstruct particle trajectories and determine their momentum and charge from the curvature of their tracks in the solenoidal magnetic field. The ID consists of three main subsystems: the Pixel detector, the Semiconductor Tracker, and the Transition Radiation Tracker. Each subsystem employs different

technologies to provide precise measurements for track reconstruction and withstand the high radiation near the interaction point.

The Pixel Detector is the innermost subsystem, positioned just a few centimetres from the beam pipe. It comprises a total of 1 744 modules arranged in three barrel layers and two end-caps, each containing three disk layers. Together, these provide continuous tracking coverage up to $|\eta| < 2.5$. Each module contains 46 080 silicon pixel sensors, each with its own readout channel. The pixels provide excellent spatial resolution, approximately 10 μm in the plane of the module and about 100 μm along the direction perpendicular to it. Before Run 2, an additional innermost layer, the Insertable B-Layer *IBL* [83], was installed at a radius of 3.3 cm to improve the vertex reconstruction and b-tagging performance.

The Semiconductor Tracker (SCT) extends the tracking volume of the Pixel detector using similar silicon sensor technology, but with reduced granularity, as it is located farther from the interaction point. It consists of four coaxial cylindrical layers in the barrel region and two end-caps, each containing nine disk layers, providing coverage up to $|\eta| < 2.5$. The SCT comprises 15 912 silicon strip sensors, each with 768 active strips. Each module is built from two pairs of single-sided sensors mounted back-to-back with a small stereo angle, allowing for precise two-dimensional space-point reconstruction. The achieved spatial resolution is approximately 17 μm in the azimuthal plane and 580 μm along the longitudinal direction.

When a charged particle crosses a silicon sensor, it ionises the material, producing electron–hole pairs. Under the influence of an applied electric field, these charges are collected by the electrodes, generating a signal proportional to the particle's energy loss as it traverses the silicon. To ensure radiation tolerance and minimise electronic noise, the Pixel and the SCT detectors operate at low temperature, typically around –7 °C to –10 °C, maintained by an evaporative CO_2 cooling system. The low temperature, together with the applied high voltage, ensures stable long-term operation of the sensors.

The Transition Radiation Tracker (TRT) is the outermost subsystem of the ID and employs a different detection technology from the Pixel and SCT. It consists of polyimide drift tubes, or *straws*, each 4 mm in diameter, filled with a xenon-based gas mixture at room temperature. When a charged particle passes through a straw, it ionises the gas, producing electron-ion pairs that drift towards a central gold-plated tungsten wire, where the resulting signal is amplified and recorded. The TRT comprises 73 straw layers parallel to the beam axis in the barrel and 160 straw planes oriented perpendicular to the beam in the end-caps, providing coverage up to $|\eta| < 2$. In addition to tracking, the TRT provides particle identification by detecting transition radiation emitted when charged particles traverse the dielectric materials surrounding the straws. This radiation depends inversely on the particle mass, allowing discrimination between light particles such as electrons and heavier hadrons.

The ID, originally designed for approximately ten years of operation under the expected radiation levels, will be entirely replaced between Run 3 and the High-Luminosity LHC (HL-LHC) era. Its successor, the Inner Tracker *ITk* [85], will be an all-silicon detector designed to withstand the substantially higher luminosity, data rates, and radiation doses anticipated at the HL-LHC, while maintaining and further improving upon the excellent

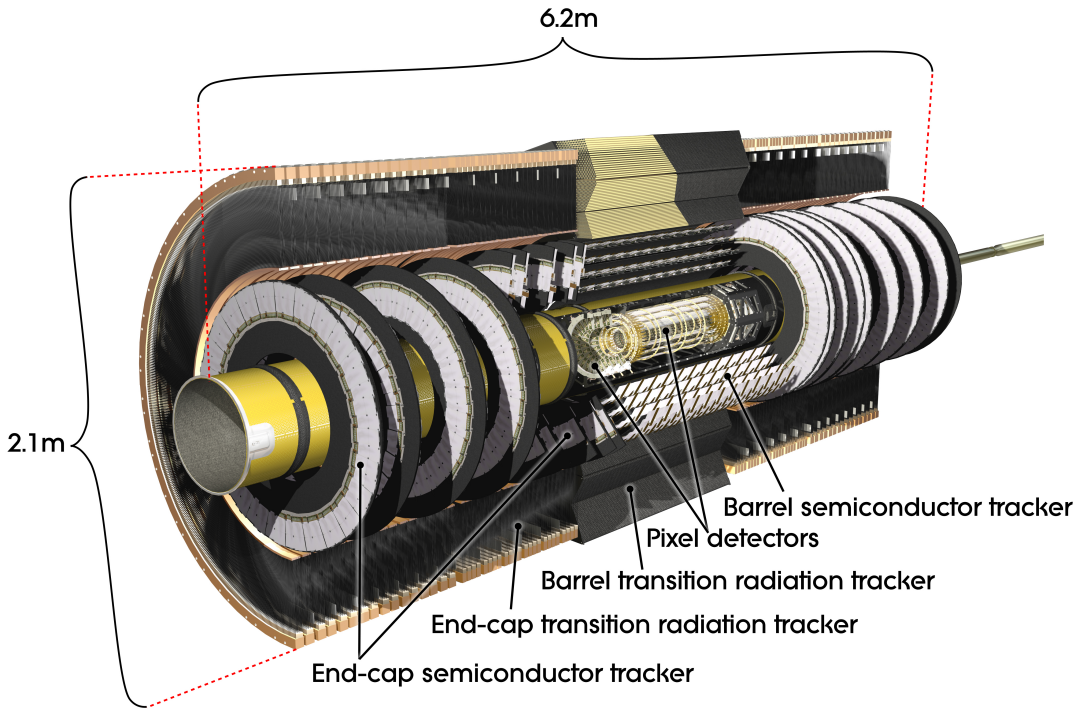


Figure 3.5: Schematic overview of the Inner Detector. Figure from Ref. [84].

tracking performance achieved by the current detector.

3.2.2.3 Calorimeter

The Calorimeter system, shown in Fig. 3.6, is located between the two magnet systems, the solenoid and the toroid. After passing through the ID, most particles deposit all their remaining energy in this subsystem, except for muons and neutrinos. Each calorimeter module consists of alternating layers of absorber and active materials. The absorber is made of a dense material, such as lead, copper, or steel, and induces interactions with the incoming particles, producing either electromagnetic or hadronic showers. The active material detects the energy deposited by shower particles and produces a measurable signal. The calorimeter measures the energy of the incoming particles and, based on the shower shape, helps identify their type. Its nearly hermetic coverage allows for a precise reconstruction of the missing transverse energy, which is essential for processes involving neutrinos or dark matter candidates. The different subsystems can be classified either by location, comprising a central barrel, two end caps, and a forward calorimeter (FCal) that covers the region closest to the beam, or by particle type, into electromagnetic (EM) and hadronic calorimeters. The latter classification will be used below.

The EM Calorimeter is divided into a barrel, two end caps, and the first layer of the FCal according to pseudorapidity coverage. All of these subsystems use liquid argon (LAr)

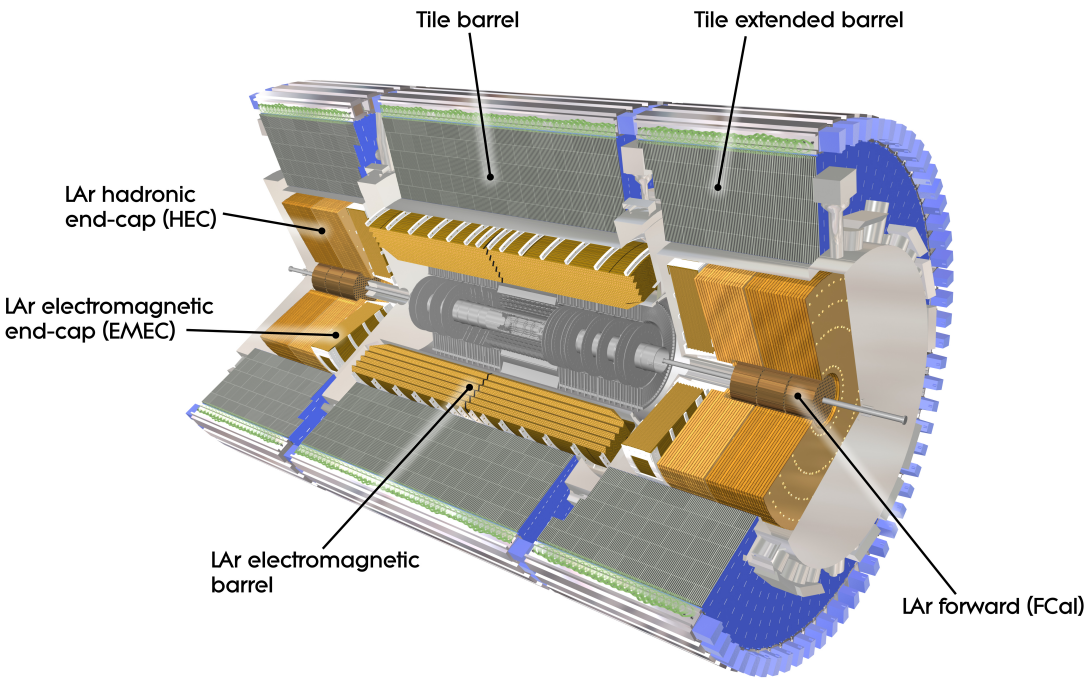


Figure 3.6: Schematic overview of the calorimeter. Figure from Ref. [86].

as the active material, in which electrons and photons produce electromagnetic showers that ionise the medium. The resulting ionisation charge is collected by electrodes arranged in an accordion geometry for both absorbers and readout plates, providing full coverage in ϕ without cracks and enabling fast signal collection. Its segmentation and granularity are optimised such that the maximum energy of the electromagnetic showers is collected in the second layer. A *presampler* is located in front of the first layer, inside the barrel cryostat, covering the same η range as the barrel. It also uses LAr to measure energy losses in upstream material, thereby improving the overall energy resolution. The barrel EM calorimeter consists of two half-barrels centred around the z-axis, covering the range $0 < |\eta| < 1.475$. The absorbers are made of lead plates of approximately 1.53 mm thickness, with two stainless-steel sheets glued to each side. The active material is LAr, kept at a temperature of about -184 °C. The two Electromagnetic End-Cap Calorimeters (EMEC) are located on either side of the barrel, covering $1.375 < |\eta| < 3.2$. Each side contains an inner and an outer wheel, with a presampler layer closer to the beam region. The first layer of the FCal, covering $3.1 < |\eta| < 4.9$, is also electromagnetic and consists of copper plates stacked along the beam axis.

The hadronic calorimeters are located in the barrel region, the end-caps (HEC), and the forward region (FCal). The tile calorimeter, located just outside the barrel EM calorimeter, covers up to $|\eta| < 1.7$. It uses steel as the absorber and plastic scintillator tiles as the active medium, from which it takes its name. The two Hadronic End-Cap Calorimeters

(HEC), located behind the EMECs, are divided into a front and a rear wheel, covering $1.5 < |\eta| < 3.2$, and use copper as an absorber. They are copper–LAr sampling calorimeters with a flat-plate geometry, in contrast to the accordion design of the EM calorimeters. In the forward region, the FCal includes two hadronic layers that use tungsten instead of copper as the absorber material, optimised for high radiation tolerance and absorption length.

3.2.2.4 Muon Spectrometer

The Muon Spectrometer (MS) is the outermost subsystem of the ATLAS detector, as shown in Fig. 3.7. Similarly to the ID, it reconstructs muon trajectories bent by the toroidal magnetic field. Muons lose, on average, about 3 GeV of energy while traversing the passive material of the detector. For this reason, a combined reconstruction using tracks from both the ID and MS is performed to improve the momentum resolution. The MS consists of several layers of detectors interleaved with and mounted on the eight coils of the superconducting barrel toroid magnet, and placed in front of and behind the end-cap toroids, providing coverage up to $|\eta| < 2.7$. The spectrometer employs two main types of detectors: precision tracking chambers that measure muon momentum and dedicated trigger chambers that provide fast information for the first-level trigger.

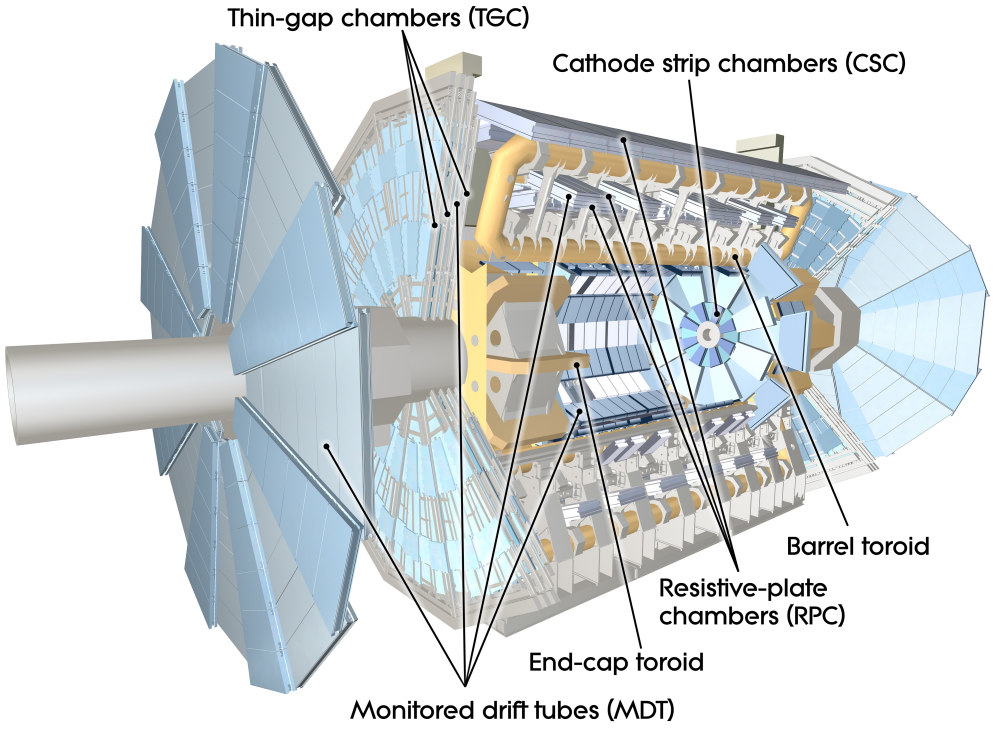


Figure 3.7: Schematic overview of the Muon Spectrometer. Figure from Ref. [87].

The Monitored Drift Tube (MDT) chambers form the main precision tracking system in the barrel and end-cap regions. Each MDT consists of drift tubes with a 29.97 mm diameter,

operated with an Ar/CO_2 gas mixture at 3 bar. This mixture is chosen for its high purity and good ageing properties. Electrons produced by ionisation of the gas drift toward a central tungsten–rhenium wire, where the signal is collected. The barrel region contains three concentric cylindrical MDT layers, and each end cap has four large wheels. Due to the large dimensions of the system, alignment and corrections for deformations caused by gravity or temperature gradients are required. In the forward region, where the counting rate exceeds the MDT limit ($|\eta| > 2.0$), Cathode Strip Chambers (CSC) are used instead, covering $2.0 < |\eta| < 2.7$. Each end-cap contains eight large and eight small CSC chambers arranged concentrically in ϕ. The CSC provides a spatial resolution of 60 μm per plane, compared to about 80 μm for MDT layers.

In addition to the precision chambers, dedicated trigger detectors provide fast information on muon trajectories and approximate transverse momentum. Two different technologies are employed. In the barrel region ($|\eta| \leq 1.05$), Resistive Plate Chambers (RPCs) are installed around the second and third MDT layers, forming a total of three RPC layers. Each RPC consists of two resistive plates separated by a 2 mm gas gap, where an electric field induces avalanches along ionising tracks towards the anode. In the end-cap region ($1.05 \leq |\eta| \leq 2.4$), Thin Gap Chambers (TGCs) are used. Each end-cap hosts one doublet layer of TGCs before the first MDT layer and three more layers surrounding the third MDT layer, for a total of seven TGC planes (two doublets and one triplet). TGCs are multi-wire proportional chambers with a wire-to-cathode spacing of 1.4 mm, enabling fast signal response.

3.2.3 Trigger system

With each bunch crossing occurring every 25 ns, about 40 million bunches of protons cross and collide every second, corresponding to a bunch rate of 40 MHz. To handle this enormous rate, a trigger system is implemented to select the most interesting events, those potentially containing signatures of new physics, and to optimise the use of available resources. At the same time, recording well-understood processes is crucial for calibrating the various reconstructed objects, assessing detector performance, and validating the modelling in simulations. The trigger system reduces the data rate from 40 MHz to about 1 kHz by analysing various signal properties, such as particle identification, charge, and transverse momentum. It consists of two parts: the Level-1 Trigger and the High-Level Trigger, which together control the flow of data that is ultimately recorded. The decision must be made extremely quickly; otherwise, the triggers could accumulate and saturate the information flow. The trigger systems are therefore located close to the ATLAS detector in the USA-15 cavern, since the speed of light limits their placement farther away, for example, on the surface.

The Level-1 (L1) trigger is hardware-based and uses signals from the calorimeter and muon detector systems. The decision is made in less than 2.5 μs, with an acceptance rate of about 100 kHz. Based on some predefined criteria, the granular information from the calorimeter and the MS is stored in buffers until the event is either accepted or rejected. If accepted, the information is passed to the High-Level Trigger (HLT). The HLT has more time (around 200 μs) to make a more informed decision about whether to accept an event. It runs on a farm of about 40 000 CPUs and processes the Regions of Interest (RoIs) identified

by the L1 trigger. The HLT analyses these areas more precisely by incorporating information from the ID. The HLT further reduces the event rate to about 1 kHz, passing the selected events to data storage for offline analysis.

The trigger menu is a list of trigger chains that define the characteristics of the events to be saved. An important distinction is made between prescaled and unprescaled triggers. Prescaled triggers correspond to collisions with a high cross-section (i.e., frequent events), where saving all of them would saturate the Data Acquisition (DAQ) system. Therefore, only one out of every N events of that type is recorded. An example is the jet trigger for low-p_{T} jets. In contrast, rarer events with low cross-sections are always saved and are thus referred to as unprescaled triggers. Additionally, a Good Run List (GRL) is produced, containing all runs in which all detector parts were operating optimally and the data met the required quality standards.

3.2.4 ATLAS Control Room

In a large collaboration such as ATLAS, all members are encouraged to contribute to the operation and maintenance of the experiment. These service tasks are essential for safe detector operation, high-quality data taking, and a shared understanding of the experiment's infrastructure. They include activities such as detector calibration, data taking, data quality monitoring, and preparations for future upgrades, such as the ITk. Contributions are recognised through a point system that also influences eligibility for responsibilities, such as conference presentations.

The ATLAS detector is located approximately 100 m underground at Point 1 of the LHC ring, while the experiment is operated from the ATLAS Control Room (ACR) on the surface. During data-taking periods, typically from April to November, the ACR is continuously staffed by shifters covering several desks. These include Shift Leader and Data Quality, Run Control and Trigger, Inner Detector, Calorimeter and Forward Detectors, and Muon Systems, all operating in 8-hour shifts. The SLIMOS desk has a distinct role focused on safety and infrastructure protection. In addition, the Run Manager and subsystem experts are often present.

The Shift Leader is responsible for coordinating detector operations, communicating with the LHC Control Room, and granting beam injection permits. They also perform first-level data quality checks using online monitoring plots. The Run Control and Trigger shifter manages run transitions, updates detector configurations when needed, and monitors the trigger system to ensure stable and efficient operation.

Subsystem shifters for the Inner Detector (ID), Calorimeter, and Muon Systems are responsible for monitoring the correct operation of their respective detectors. The Calorimeter shifter also oversees the magnet system and the LUCID luminosity detector [88]. Given my involvement as an ID shifter at various times of year, the following discussion focuses on this role.

The ID shifter monitors the correct operation of the Pixel, SCT, TRT, and BCM (Beam

Conditions Monitor) subsystems[1]. The overall detector status is monitored through the Finite State Machine (FSM) panel, which provides real-time information on key parameters such as voltage, temperature, and pressure. A central responsibility of the shifter is to ensure that all subsystems remain in a safe state and that transitions between states occur correctly in response to changes in the LHC status. The shifter also oversees data acquisition, ensuring that the ID subsystems do not block data taking, for example, due to readout or communication errors. Data quality monitoring constitutes another essential aspect of the role. The shifter inspects the first data quality monitoring (DQM) plots, both during data taking and after beam dumps, to verify correct detector behaviour before granting a new injection permit. All observations, actions, and issues are recorded in the shift logbook to ensure continuity between shifts and to allow prompt intervention by experts if needed. Overall, the ID shifter acts as an interface between the ID subsystems and the detector experts, enabling informed operational decisions to be taken under expert supervision.

The nature of the shift activity depends strongly on the data-taking period. The most intensive phase occurs before the start of the yearly physics run, when experts perform extensive tests following detector maintenance or upgrades and optimise the detector's operational parameters. During this period, warnings and errors are more frequent, although they do not prevent data taking.

Once proton collisions begin, the role becomes more operationally critical but is generally calmer under stable conditions. The most delicate procedures take place at the start of each run. The Pixel and SCT detectors must be placed in `STANDBY` mode during beam injection, during which their high voltages are switched off. As they are the subsystems closest to the beam pipe, their high voltage is ramped up only after stable beams are declared, after which data taking can begin.

During data taking, the ID shifter continuously monitors the detector to ensure that all subsystems operate within their defined thresholds. This includes verifying that triggers function correctly, that data links remain unblocked and do not become busy, and that data acquisition proceeds smoothly. In parallel, the shifter assesses data quality by inspecting the relevant Data Quality Monitoring (DQM) plots.

After a beam dump, the histograms associated with the dump must be carefully checked before issuing a new injection permit, ensuring that the dump was successful and that the detector is again in a safe operational state.

Working in the ATLAS Control Room provides valuable insight into the experiment's operation and offers a direct connection to the data-taking process that produces the datasets accessed worldwide. This experience offers essential context for offline analysis work and, with increasing familiarity, allows shifters to develop significant expertise in detector operations and data quality assessment.

[1]The BCM continuously monitors beam conditions to protect the detector from potentially harmful beam incidents and can request a beam dump if unsafe conditions are detected.

3.3 From collision to local computer

3.3.1 Real collision data

After the protons collide inside the detector, numerous particles are produced, and their trajectories are recorded by the detector subsystems. First, a rapid reconstruction of these signals is performed to allow the trigger system to decide whether an event should be kept or discarded.

If an event receives the "accepted" flag from the trigger, it is transferred to the Data Acquisition System (DAQ). The DAQ collects and transports data from the various trigger levels to a central storage. The Event Builder then gathers data from all detector subsystems to assemble a complete event record and assign a unique tag. These raw data (RAW) are subsequently sent to Tier-0, CERN's primary computing facility. Tier-0 performs the first stage of offline reconstruction, producing Event Summary Data (ESD) and monitoring plots for data quality assurance. The ESDs are then reduced to Analysis Object Data (AOD), which are smaller and optimised for analysis. The AOD files are stored on disk and distributed to Tier-1 and Tier-2 centres, forming part of the Worldwide LHC Computing Grid (WLCG). The WLCG is an international computing infrastructure that provides the resources to store, distribute, and analyse the enormous amount of data generated by the LHC experiments. As a summary, Tier-0 at CERN performs prompt reconstruction; Tier-1 centres provide large-scale storage and reprocessing capacity (when reconstruction algorithms or calibrations are updated); and Tier-2 centres are primarily used for user analyses and simulation production.

From the AODs, further data reduction steps produce Derived Analysis Object Data (DAODs). During this derivation stage, several reduction techniques are applied:

- **Skimming:** Selecting only events that satisfy certain criteria.
- **Trimming:** Removing entire collections that are not needed.
- **Thinning:** Removing some objects within collections.
- **Slimming:** Removing specific attributes of the remaining objects.

These DAODs are the format most analyses use as a starting point and are accessible worldwide to ATLAS members via the grid.

If, during a data-taking period, the detector operated under optimal conditions and the recorded data met the required quality standards, the corresponding runs or luminosity blocks are included in the Good Run List (GRL). Only data listed in the GRL is used for physics analyses.

Up until Run 2, each analysis group had its own derivation, with specific variables that were helpful to it. As statistics have increased, ATLAS has adopted a more general approach, where analyses use the `PHYS` derivation with general variables that are not necessarily specialised to each analysis. Combined Performance (CP) groups, responsible for calibrating

the different objects, need specific derivations to improve resolution, and have retained their own derivations.

Analysts typically do not work directly with DAODs but instead convert them into smaller, analysis-specific datasets called *ntuples*. The production of ntuples is computationally intensive and usually carried out on the grid. This step further reduces file size by keeping only the relevant objects and collections, applying event selection, recording cut flows and selection efficiencies, while applying the necessary calibrations to the physics objects used in the analysis.

3.3.2 Simulation data

Theoretical calculations and experimental findings are constantly evolving and being compared. It is particularly interesting when both disagree, as such discrepancies may hint at the presence of new physics. Particle physicists rely heavily on simulated data to study the expected performance of their analyses and to build hypotheses that can be compared with real measurements. Simulation data, in turn, depend directly on the underlying theory and on the computational capabilities of the available software tools.

The most widely used simulation technique is the Monte Carlo (MC) method, which produces random samples to explore the possible outcomes of complex interactions. This approach is used both for numerical integration over phase space, to evaluate cross-sections, and for event generation, where individual scattering events are simulated according to their theoretical probabilities. Typically, the initial and final particles are specified, and all possible contributions leading to this interaction are explored. This primary interaction is referred to as the *hard-scatter*. The matrix element, derived from the amplitude of the process and computed using the Feynman rules of the SM Lagrangian, is used to sample the cross-section of the process. Continuous computational advances have enabled increasingly accurate results by including higher-order loop corrections.

Several event generators are available for such simulations, including MadGraph [89], Sherpa [90], and Powheg [91–93]. Differences between these generators are typically accounted for using scale factors provided by each CP group, which are treated as additional systematic uncertainties.

Another challenging aspect of the model is understanding pp collisions themselves. Because protons are composite particles, consisting of three quarks and an indeterminate number of gluons, the energy carried by each parton inside the proton varies. Thus, when two protons collide, only a fraction of the total energy is available for the actual parton–parton interaction. To model this energy distribution, *parton distribution functions* (PDFs) [94] are computed and continuously refined.

Once the hard-scatter process is generated, the event proceeds through the decays of unstable particles and the hadronisation process. Some event generators handle both stages directly; otherwise, dedicated software such as Pythia8 [95] is used. During this stage, pile-up event simulation is also included.

Finally, Geant4 [96] is used to simulate how the resulting particles would interact with the detector material. This produces simulated hits and energy deposits similar to those

observed during real data-taking periods, which are then digitised. The standard reconstruction algorithms are applied to these simulated signals, resulting in datasets comparable in structure and format to real experimental data.

When validating MC simulations and comparing them to real data, different weights need to be applied. The per-event weights ensure the simulated process has a finite amplitude and the correct proportions. The MC weight provides the correct cross-section for the process, so the counts can be compared with data.

The advantage of simulated data lies in the availability of both truth-level information and reconstruction-level quantities, obtained after applying the detector reconstruction algorithms. This enables a direct comparison between generated and reconstructed observables. In contrast, for collision data only the reconstructed quantities are recorded. In the following section, the different reconstruction algorithms are described.

3.3.3 Reconstruction and identification of physics objects

After the detector hits are digitised, reconstruction algorithms are applied to group the detector signals and identify physics objects, providing estimates of their energy and direction.

This subsection presents a more detailed description of the reconstruction and identification of the different physics objects used in physics analyses. The reconstruction relies on combining information from the various subdetectors described in the previous sections.

Figure 3.8 illustrates the typical signatures left by different particles in a slice of the ATLAS detector. The particles are produced close to the primary interaction vertex and propagate outwards from the detector centre, interacting with the various detector components.

In the ID, only charged particles are directly reconstructed, as their trajectories are bent by the magnetic field. The direction of curvature depends on the charge of the particle. In the figure, this is illustrated for muons, electrons, and protons. Electrons and photons deposit their energy in the EM calorimeter, producing electromagnetic showers, while hadrons such as protons and neutrons interact primarily in the hadronic calorimeter, producing hadronic showers that are reconstructed as jets. Muons traverse the calorimeters with minimal energy loss and are subsequently detected in the MS, which is immersed in a magnetic field that further bends their trajectories. Neutrinos interact only very weakly with the detector material and therefore escape detection; their presence is inferred from an imbalance in the transverse momentum of the reconstructed objects in the event, referred to as missing transverse momentum.

Dedicated reconstruction and identification algorithms are developed by the CP groups to discriminate prompt physics objects from backgrounds and misidentified candidates. These include identification criteria and isolation requirements. Since different CP groups may reconstruct overlapping objects from the same detector signals, an overlap removal procedure is applied to resolve ambiguities and prevent double counting of energy.

This thesis focuses on the physics objects most relevant to this thesis. The reconstruction and identification of tracks, vertices, electrons, photons, and muons are described in the following sections. Jets also play a central role and are discussed separately in Chapter 4.

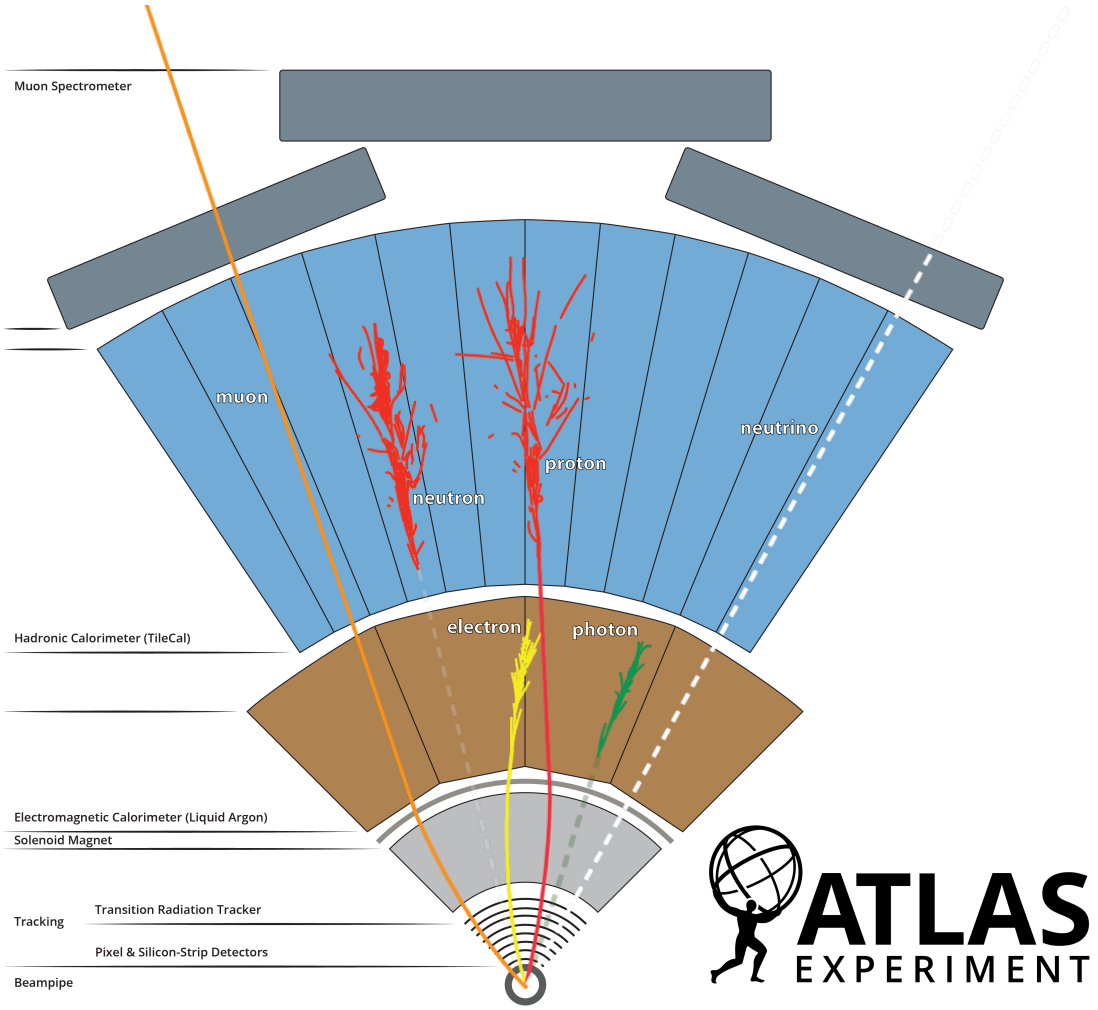


Figure 3.8: Schematic illustration of the typical signatures left by different particles in a slice of the ATLAS detector. The inner detector is shown in light grey, the electromagnetic calorimeter in brown, the hadronic calorimeter in blue, and the muon spectrometer in dark grey. Solid lines represent visible particle trajectories in a given subdetector, while dashed lines indicate the absence of a measurable signal. Muons are shown in orange, hadrons in red, electrons in yellow, protons in green, and neutrinos in white. Figure from Ref. [97].

Information on the reconstruction and calibration of taus and missing transverse momentum is available in Refs. [98–100].

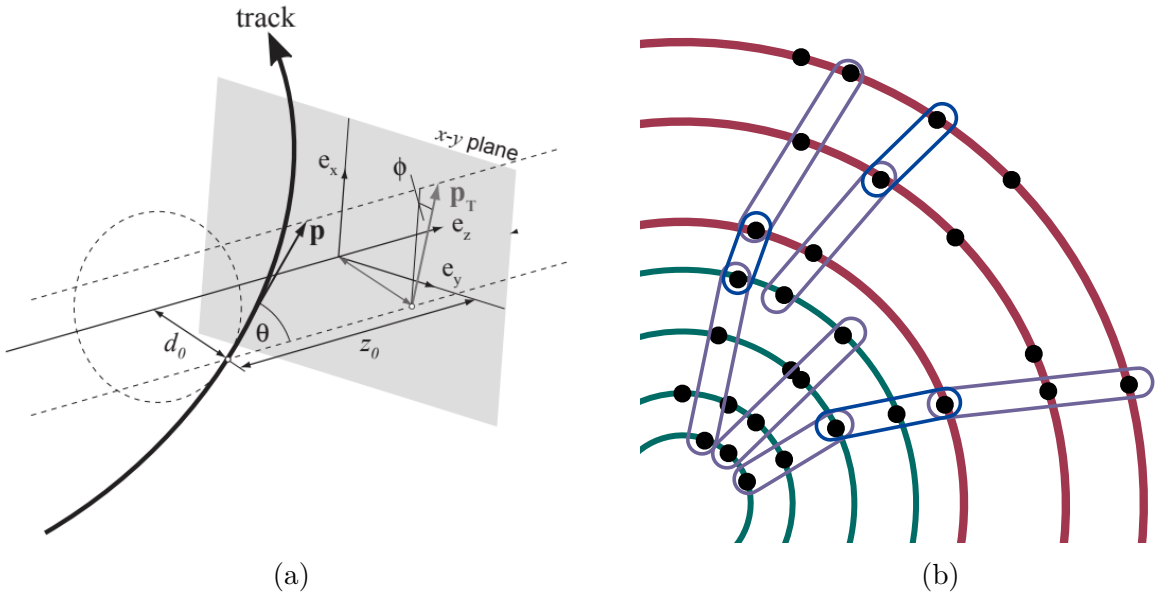
3.3.3.1 Tracks and vertices

Track reconstruction consists of determining the trajectories of charged particles as they traverse the ID. Due to the presence of the solenoidal magnetic field, charged particles follow curved paths and interact with the various detector layers, leaving discrete measurements, referred to as *hits*, in each layer. The tracking algorithms associate hits from different layers that are consistent with originating from the same particle, thereby reconstructing its trajectory. While this task is relatively straightforward in low-occupancy environments, such as events with little pile-up, it becomes significantly more challenging under the high-luminosity conditions of the LHC, where multiple interactions occur in the same bunch crossing.

The reconstructed trajectory is parametrised by five track parameters, illustrated schematically in Fig. 3.9a. The transverse impact parameter, d_0, is defined as the distance of closest approach of the particle trajectory to the interaction point in the transverse plane. The longitudinal impact parameter, z_0, describes the distance of closest approach along the beam axis. The direction of the track is characterised by the azimuthal angle ϕ and the polar angle θ. Finally, the curvature of the track is described by the ratio of the electric charge to the particle momentum, q/p. For particles with high transverse momentum, the curvature is small, making charge determination more challenging and increasing the associated uncertainty.

In ATLAS, track reconstruction [102] is performed in two stages: an inside-out and an outside-in approach. In the inside-out reconstruction, hits in the pixel and SCT detectors are first clustered. These hits are combined into three-dimensional space points, from which triplets are formed to create initial track seeds. An illustration of the initial track seeds is shown in Fig. 3.9b. The seeds are then processed by a combinatorial Kalman filter [103], which iteratively extends the track by adding compatible hit candidates from successive detector layers. A scoring algorithm is used to rank the resulting track candidates based on the number of shared hits and the χ^2 of the track fit [104]. Following this step, information from the TRT is incorporated, allowing the tracks to be further extended while continuously evaluating their quality. For particles originating from displaced vertices, such as those produced in photon conversions, an outside-in reconstruction procedure is employed. This approach uses hits that were not assigned during the inside-out reconstruction and focuses on regions of interest in the EM calorimeter. It begins by forming track segments in the TRT and subsequently extrapolates them inward, incorporating compatible hits from the SCT and pixel detectors.

Once tracks have been reconstructed, a dedicated vertex reconstruction algorithm is applied [105]. The initial vertex candidates are identified by locating regions of high track density along the beam axis. A vertex fit is then performed, and tracks incompatible with the fitted vertex are removed based on their impact parameter significance, typically requiring $|d_0|/\sigma(d_0) < 7$. This procedure is iterated until all tracks are associated with a reconstructed



(a) (b)

Figure 3.9: (a) Schematic overview of the track parameters used to describe charged-particle trajectories. (b) Illustration of hits in the pixel and SCT subdetectors and the formation of initial track seeds from hit triplets. Figures adapted from Ref. [101].

vertex. The primary vertex (PV) corresponding to the hard-scattering interaction is defined as the vertex with the highest sum of squared transverse momenta of its associated tracks, $\sum p_T^2$. The remaining reconstructed vertices are attributed to pile-up interactions occurring in the same bunch crossing. Vertices that are significantly displaced from the beamline are referred to as secondary vertices and can arise, for example, from the decays of long-lived particles such as b-hadrons.

Recent developments in track and vertex reconstruction increasingly exploit machine learning techniques [106,107]. These approaches aim to improve robustness against high pile-up conditions and enable faster, more efficient reconstruction, which is particularly important for future LHC runs with increased luminosity.

3.3.3.2 Electrons and photons

Electrons and photons are reconstructed by the same CP group, as their experimental signatures share many common features. Both objects are reconstructed within $|\eta| < 2.47$[2] using information from the EM calorimeter, complemented by tracking information from the ID [108]. Calorimeter cells are grouped into topological clusters[3], which provide a representation of the EM showers produced by electrons and photons. Tracks reconstructed in the

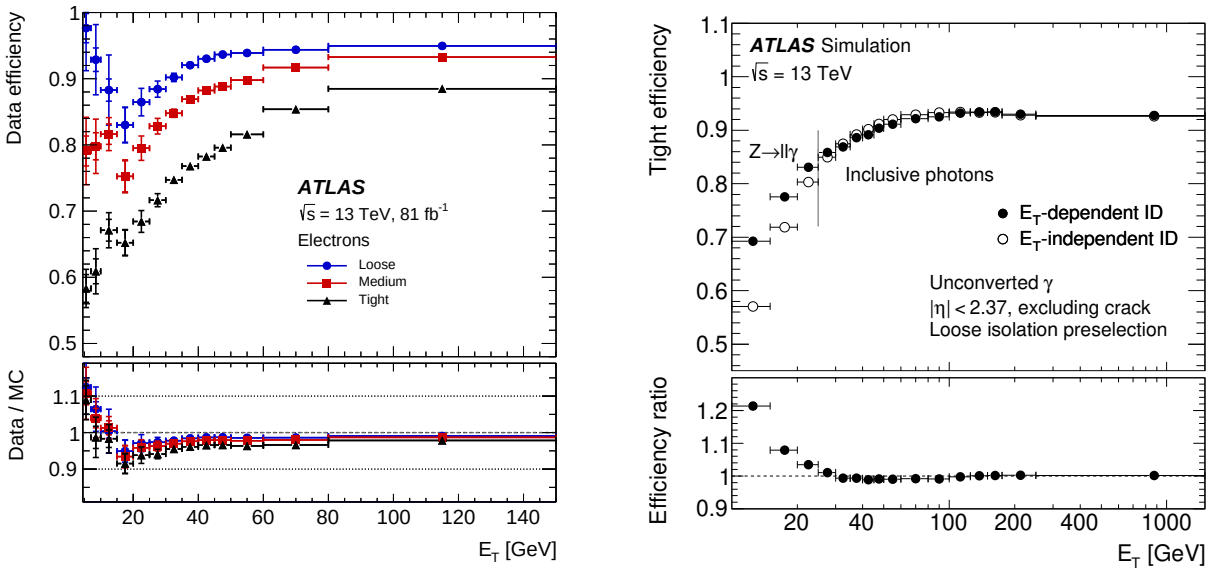
[2]Excluding the calorimeter transition (*crack*) region between $1.37 < |\eta| < 1.52$.

[3]A detailed description of topological clustering in the context of jets is provided in Section 4.2.

ID are then matched to EM clusters within a cone in $\eta \times \phi$ space to establish the particle type.

Electromagnetic clusters without an associated track are classified as unconverted photons. If the cluster is associated with one or two tracks forming a secondary vertex, it is identified as a converted photon, corresponding to a photon that undergoes pair production in the detector material upstream of the calorimeter. Electromagnetic clusters associated with a track originating from the PV are classified as electrons. For electrons, dedicated track-fitting techniques are employed to account for bremsstrahlung energy losses in the detector material.

Electrons produced directly in hard-scattering interactions are referred to as *prompt* electrons and constitute the primary signal of interest in many analyses. An electron identification algorithm is employed to suppress background contributions from jets and non-prompt leptons. It is based on a likelihood classifier using input variables such as track quality, track-cluster matching, and EM shower-shape observables. Three working points (WPs) are defined to accommodate different analysis requirements: *Loose*, *Medium*, and *Tight*, with tighter working points providing higher purity at the cost of reduced efficiency.



(a) (b)

Figure 3.10: (a) Electron identification efficiency in $Z \to ee$ events as a function of E_T, shown for the Loose, Medium, and Tight working points. The bottom panel displays the ratio of data to simulation. (b) Identification efficiency of unconverted photons for the Tight working point, as a function of photon E_T. Efficiencies for the E_T-independent and E_T-dependent selections are compared, with the bottom panel showing the ratio between them. Figures from Ref. [108].

Photon identification follows a similar strategy and is designed to suppress background from neutral hadrons and jets misidentified as photons. It relies primarily on calorimeter-

based shower-shape variables, exploiting differences between electromagnetic and hadronic energy deposits. Separate identification criteria are defined for converted and unconverted photons to account for their different detector signatures.

The identification performance for electrons and photons is shown in Fig. 3.10, where the electron efficiency for different working points and the Tight unconverted photon identification efficiency are presented as functions of transverse energy.

In addition to identification requirements, isolation criteria are applied to both electrons and photons to further suppress backgrounds from non-prompt particles and hadronic activity. Isolation variables are constructed from the energy deposited in the calorimeter and the summed transverse momentum of tracks within a cone around the object direction, after subtracting the contribution of the object itself and correcting for pile-up effects.

Energy calibration and resolution corrections are applied to both electrons and photons to account for detector effects and upstream material [109]. These corrections are derived using a combination of simulation and data-driven techniques, ensuring an accurate measurement of their kinematic properties for physics analyses.

3.3.3.3 Muons

Muons are minimum-ionising particles that typically do not decay within the detector volume. Their reconstruction [110] primarily relies on track information from both the ID and the MS. At the same time, energy deposits in the calorimeters are also used to correct for occasional large energy losses, which average around 3 GeV.

The reconstruction in the MS begins by forming segments from hits in individual muon chambers. Segments from different stations are then combined to build three-dimensional track candidates, with a global χ^2 fit applied to optimise the track parameters.

Different types of muons are defined to optimise reconstruction efficiency and kinematic coverage. The most commonly used are *combined muons*, which incorporate tracks from both the ID and MS. Other types, such as *inside-out*, *MS-extrapolated*, *segment-tagged*, and *calorimeter-tagged* muons, are designed to recover inefficiencies in regions with limited coverage or for low-p_{T} tracks.

Muon reconstruction and identification are performed within the region $|\eta| < 2.7$, where the MS provides tracking coverage, and with a typical transverse momentum threshold of $p_{\mathrm{T}} > 5$–6 GeV to ensure reliable track reconstruction and identification.

As with electrons and photons, dedicated identification and isolation algorithms are applied to muons to discriminate prompt muons from backgrounds arising from hadronic decays, secondary interactions, or misidentified tracks. The performance of the muon identification is illustrated in Figure 3.11, which shows the efficiency of the Medium working point as a function of the muon transverse momentum.

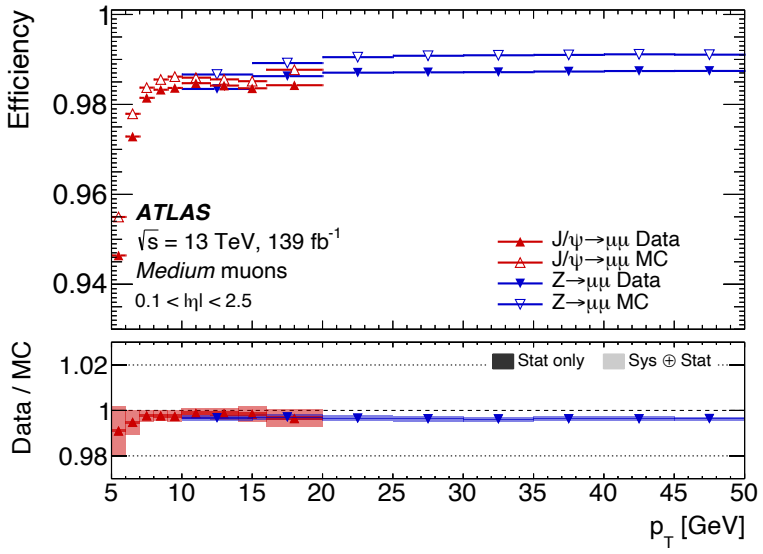


Figure 3.11: Muon identification efficiency for the Medium working point measured using $J/\psi \to \mu\mu$ and $Z \to \mu\mu$ events, shown as a function of the muon transverse momentum p_{T} for $0.1 < |\eta| < 2.5$. The use of J/ψ and Z decays provides sensitivity to low and high-p_{T} muons, respectively. The bottom panel displays the ratio of data to simulation. Figure from Ref. [110].

Part II

Calibration: b-jet energy resolution

Chapter 4

Jet reconstruction and calibration

4.1 Introduction

When protons collide in the LHC, the interactions occur between their constituent partons (quarks and gluons). Only a subset of these partons actually interacts in a hard scattering, producing new high-energy quarks or gluons. However, because of colour confinement, neither the outgoing partons from the collision nor the remnants of the protons can exist freely. As they move apart, the colour field between them stretches and eventually breaks, creating additional quark-antiquark pairs from the vacuum. This initiates a parton shower, where the high-energy quarks and gluons radiate further partons. These combine with the existing partons to form colour-neutral bound states through a non-perturbative process known as hadronisation. The resulting hadrons are typically unstable and further decay into lighter stable particles. Most of the particles originating from a single high-energy parton appear in the detector as a jet, a collimated spray of hadrons typically reconstructed within a cone of radius R in the detector's angular coordinates. A schematic illustration of the jet formation process is shown in Fig. 4.1. The reconstruction and calibration of jets is developed and maintained by the JetETMiss Combined Performance (CP) group.

The detector records the trajectories of charged particles in the inner detector (ID) and the energy deposits in the calorimeters, particularly the hadronic calorimeter, from which the constituents forming a jet are reconstructed. This information allows jets to be categorised in different ways, depending on the physics question and the available information. One common classification distinguishes quark- from gluon-initiated jets, which exhibit statistically distinct signatures and can be identified using dedicated machine learning taggers [111,112]. According to QCD, gluons radiate more partons than quarks, producing broader showers and wider jets.

Quark jets can be further categorised according to their flavour, with the majority being light-flavour jets. Heavy-flavour jets present additional subtleties. Top quarks do not produce jets directly. Due to their large mass, they are extremely short-lived and decay via the weak interaction into a b-quark and a W boson before hadronisation can occur. In contrast, bottom and charm quarks hadronise into B- and C-hadrons, which have relatively long lifetimes and

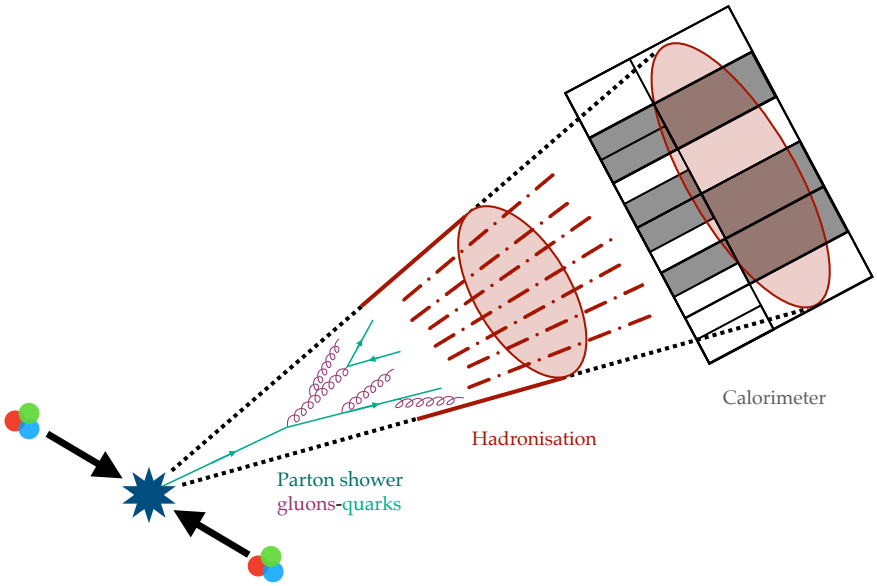


Figure 4.1: Schematic representation of a quark produced in a pp collision evolving through parton showering, hadronisation, and energy deposition in the calorimeter to form a reconstructed jet.

travel a measurable distance before decaying, leaving a secondary vertex in the detector. These hadrons can also decay leptonically, producing an electron or muon that may fall outside the reconstructed jet. Jet flavour tagging is performed using a transformer-based algorithm [113].

All these distinctions help categorise jets, which, in turn, improves the reconstruction of their energy, since the calorimeter response to different jets varies. Accurate jet measurements are crucial, as weak bosons and the Higgs frequently decay into quarks, leading to a high rate of hadronic final states. Therefore, precise measurements can aid in observing predicted processes, such as di-Higgs production, or in revealing potential connections to BSM physics.

This chapter presents the reconstruction and calibration of jets in ATLAS, with particular emphasis on in-situ calibration techniques and heavy-flavour jet identification. The limitations of current calibration strategies motivate the development of improved approaches for b-jets, which are investigated in the following chapters using machine learning algorithms.

4.2 Reconstruction

Topo-clusters [114] are constructed by combining neighbouring calorimeter cells with significant energy deposits. For each calorimeter cell, a signal significance is computed as $\zeta_{\text{cell}}^{\text{EM}} = E_{\text{cell}}^{\text{EM}}/\sigma_{\text{noise,cell}}^{\text{EM}}$, which quantifies how much the cell's energy deposit exceeds the ex-

pected noise level. Cells with $\zeta_{\text{cell}}^{\text{EM}} > 4$ are used as seeds. Their neighbouring cells with $\zeta_{\text{cell}}^{\text{EM}} > 2$ are then added, and the cluster is completed by including adjacent cells with $\zeta_{\text{cell}}^{\text{EM}} > 0$. Recently, an additional requirement has been introduced: a cell timing cut to suppress out-of-time pile-up [115], energy deposits from other bunch crossings. Cells with $|t_{\text{cell}}| > 12.5$ ns are excluded from clustering unless the signal significance exceeds a threshold of $\zeta_{\text{cell}}^{\text{EM}} > 20$, in order to preserve potential signals from long-lived particles.

Different jets can be built by using different detector inputs. *EMTopo* are built directly from topo-clusters after applying a correction to improve the reconstructed direction of the jet origin. *Particle flow objects* (PFlow) [116] combine the information from topo-clusters with tracks from the ID. To avoid double-counting the energy of the particles, a cell-based energy subtraction algorithm removes the overlap between energy measured in tracks associated with a cluster. Using track information is particularly beneficial at low-p_{T}, where the magnetic field significantly affects particle trajectories, providing improved momentum resolution. Other possible jet inputs, such as *unified flow objects* (UFO) [117], exist but are not used in this study.

Once the jet constituents have been chosen and corrected, they are passed to the anti-k_t algorithm [118], which is infrared and collinear safe[1] and produces approximately circular jets. The algorithm uses a parameter R, which effectively defines the jet radius. Currently, two main jet sizes are defined: small-R jets with $R = 0.4$ and large-R jets with $R = 1.0$. Large-R jets are typically used in analyses involving boosted particles decaying into two jets, since the decay products are closer in angular space compared to the case where the parent particle is at rest. Ongoing studies are exploring alternative R values to evaluate performance improvements [120]. Small-R jets are used in the present studies.

4.3 Calibration scheme

Once reconstructed, jets exhibit kinematic properties that differ from their true values in simulation, and there are additional differences between simulation and data. Jet calibration is a complex, sequential procedure designed to correct these effects and reduce biases in the reconstructed jet momentum and direction.

This section outlines the most recent calibration scheme for small-R jets [13], while the calibration of large-R jets is described in Ref. [14]. Small-R jets are currently calibrated through the sequential chain illustrated in Fig. 4.2, which includes:

- Jet reconstruction: PFlow jets are reconstructed using the anti-k_t algorithm with $R = 0.4$.
- Pile-up corrections: A p_{T}-density-based subtraction is applied to remove pile-up contributions, followed by an additional correction accounting for the dependency on the number of primary vertices N_{PV} and the average pile-up μ.

[1] Infrared and collinear safety refers to the property that the reconstructed jets in an event remain unchanged if the event is modified by adding a soft emission or by a collinear splitting [119].

- Absolute MC-based calibration: The reconstructed jet four-momentum is corrected to the particle-level energy scale, accounting for the non-compensating calorimeter response, energy losses in passive material, out-of-cone effects, and biases in the η reconstruction.
- Global sequential calibration (GSC): Multiplicative corrections are applied to reduce the impact of flavour and energy distribution, improving the jet energy resolution without altering the average jet energy response.
- In-situ calibration: Residual differences between data and simulation are corrected using additional techniques. The ratio of jet response in simulation to data is applied as the last correction, bringing data to the appropriate simulation scale.

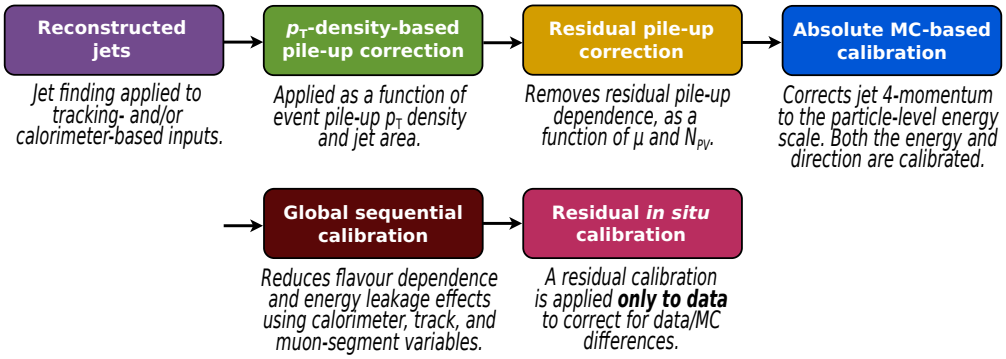


Figure 4.2: Current ATLAS calibration scheme for small-R jets. Figure from Ref. [13].

This calibration chain is derived using *inclusive* jets[2] without distinguishing between flavours. However, different quark flavours produce jets with distinct signatures, leading to flavour-dependent differences in the detector response.

In this work, we focus on improving the reconstruction of small-R jets originating from b-quarks and studying the corresponding calibration. Before presenting our studies, a more detailed explanation of the MCJES, GSC and in-situ steps is provided to clarify some of the methods employed in the following two chapters. A more detailed discussion on the previous calibration steps is not included here, as they rely on low-level detector information that is not always available at the `DAOD` level and falls beyond the scope of this thesis.

4.3.1 MCJES and Global sequential calibration

After pile-up corrections, the absolute MC-based calibration, referred to as MCJES, and the Global Sequential Calibration (GSC) are applied.

[2]Inclusive jets include all flavours, from gluon-initiated jets to jets originating from any quark type.

In simulated events, MCJES uses the truth-level jet momentum to define the jet response as

$$\mathcal{R} = \frac{p_{\mathrm{T}}^{\mathrm{reco}}}{p_{\mathrm{T}}^{\mathrm{truth}}} \,. \tag{4.1}$$

Jets are binned in transverse momentum and pseudorapidity, and the response distribution in each bin is fitted with a Gaussian. The mean of the fit represents the Jet Energy Scale (JES), indicating how close the average response is to unity, and corrections are applied to bring the average as close as possible.

The Global Sequential Calibration (GSC) applies a series of multiplicative corrections to reduce fluctuations in the calorimeter response and improve the Jet Energy Resolution (JER), without changing the overall jet energy scale. The width of the response distribution (typically the standard deviation) quantifies the JER. A response that is well-centred around unity and as narrow as possible reflects both an accurate reconstruction algorithm and a good understanding of the detector. The GSC corrections exploit several jet observables, including: the longitudinal structure of the energy depositions within the calorimeters, tracking information associated with the jet, and activity in the muon chambers behind the jet.

Figure 4.3 illustrates the effect of the GSC corrections on the jet energy resolution, where lower values indicate better resolution. The resolution naturally improves at higher p_{T} due to detector effects being more significant at low-p_{T}. Applying all corrections results in consistent improvement across the full p_{T} spectrum.

4.3.2 In-situ calibration

After the MC jet calibration steps are applied, the in-situ group derives two additional calibration factors for calibrating jets in real data. Unlike in simulation, no truth-level information is available in data. To overcome this feature of the data, in-situ calibration exploits specific event topologies in which a leading jet is back-to-back with a well-measured reference object, ensuring that their momenta are balanced. This method is known as the Direct Balance (DB) technique. The jet response in data is therefore computed as

$$\mathcal{R}_{\mathrm{DB}} = \frac{p_{\mathrm{T}}^{\mathrm{reco}}}{p_{\mathrm{T}}^{\mathrm{ref}}} \,, \tag{4.2}$$

where $p_{\mathrm{T}}^{\mathrm{reco}}$ refers to the reconstructed jet momentum and $p_{\mathrm{T}}^{\mathrm{ref}}$ to the momentum of the reference object. As in the GSC, the goal is to extract the JES and JER using this new response definition. However, discrepancies between these values in data and simulation arise and need to be corrected. This mismatch, seen in the ratio of the response in simulation to that in data, is referred to as non-closure and indicates differences in detector response or modelling. The magnitude of the non-closure quantifies how the data must be corrected to match the scale of the simulation.

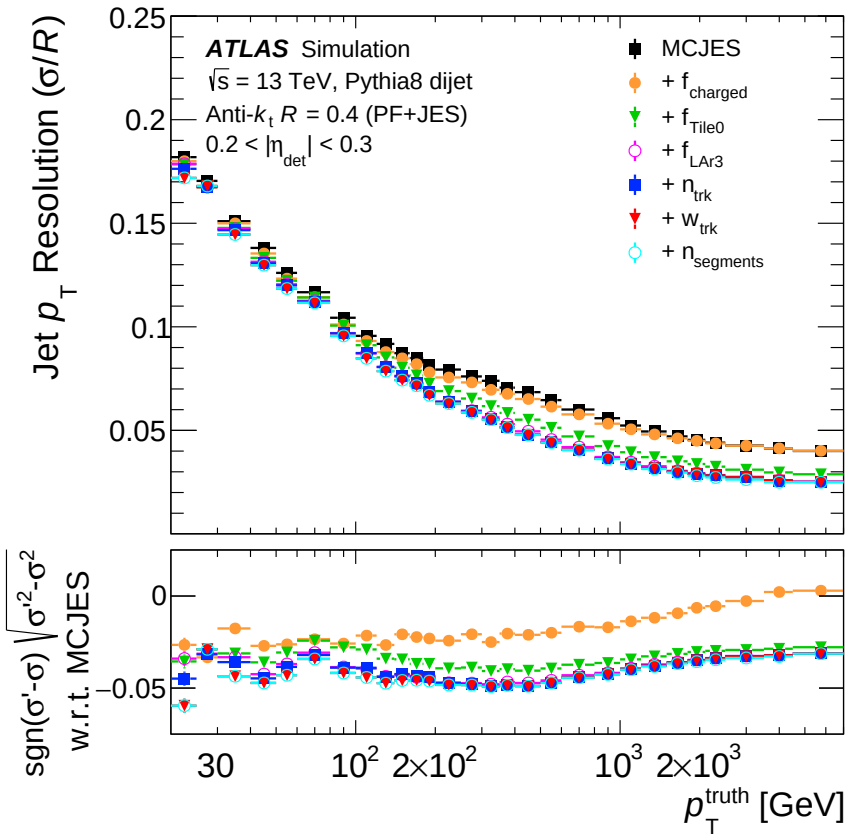


Figure 4.3: Resolution of PFlow jets with $0.2 < |\eta_{\text{det}}| < 0.3$ measured in PYTHIA8 dijet MC simulation at each stage of GSC. The lower panel shows the difference in quadrature between the resolution before GSC correction and after each step. Figure from Ref. [13].

4.3.2.1 Jet Energy Scale

To determine the JES, the in-situ methods are split into two momentum regimes. For low-p_{T} jets (from 17 GeV to 1 TeV), the calibration uses V+jets samples, where V is a well-measured object such as a Z boson (decaying leptonically to a pair of electrons or muons) or a photon, γ. A leading jet is required to be back-to-back with the recoil object, and the event must have low additional hadronic activity to ensure that R_{DB} approximates the true response. For high-p_{T} jets, multijet balance (MJB) events are used. In this topology, a single high-p_{T} jet recoils against a system of lower-p_{T} jets, which must be calibrated beforehand to calibrate the high-p_{T} regime reliably.

In the context of this project, we focus on the low-p_{T} regime, using Z+jets and γ+jets events, which together cover a broad range of transverse momenta. The direct balance (DB) technique [121] and the missing projection fraction (MPF) method [121] both exploit the momentum balance in the V+jets topology to get the JES, but differ in how the response is defined and in their sensitivity to different systematic effects.

The DB response is given by

$$\mathcal{R}_{\mathrm{DB}} = \left\langle p_{\mathrm{T}}^{\mathrm{j,reco}} / p_{\mathrm{T}}^{\mathrm{ref}} \right\rangle , \tag{4.3}$$

where $p_{\mathrm{T}}^{\mathrm{ref}} = p_{\mathrm{T}}^{Z/\gamma} |\cos \Delta\phi|$, with $\Delta\phi$ being the azimuthal angle between the reconstructed jet and the reference object. Because a reconstructed jet does not capture all particles recoiling against the Z or the photon, DB is sensitive to out-of-cone radiation.

The MPF method uses the full hadronic recoil by using other soft jets and the missing transverse momentum ($\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$), and is therefore less sensitive to out-of-cone radiation and pile-up. The MPF response is defined as

$$\mathcal{R}_{\mathrm{MPF}} = \left\langle 1 + \frac{\hat{n}_{\gamma} \cdot \vec{E}_{\mathrm{T}}^{\mathrm{miss}}}{p_{\mathrm{T}}^{\mathrm{ref}}} \right\rangle . \tag{4.4}$$

This project primarily uses the DB method, but aspects of the full hadronic activity are explored in Section 5.4.2.

4.3.2.2 Jet Energy Resolution

The jet energy resolution (JER) describes the spread of the reconstructed jet p_{T} and can be parametrised as

$$\sigma(p_{\mathrm{T}})/p_{\mathrm{T}} = N/p_{\mathrm{T}} + S/\sqrt{p_{\mathrm{T}}} + C \, , \tag{4.5}$$

where N, S, and C represent the noise, stochastic, and constant terms, respectively [13].

Dijet events are used to constrain the values of N, S, and C. However, low-p_{T} dijet events are rarely recorded due to trigger prescales (to avoid saturating the Data Acquisition system), so the noise term is further constrained using the random-cone method [13]. Recently, a study that uses pile-up events as a source of low-p_{T} events has shown improved statistics for this regime [122].

For dijet events, a back-to-back pair of jets is selected to study the momentum balance. The asymmetry distribution of the pair is used to measure the detector resolution. Calibration proceeds sequentially from the central to the more forward regions. In the central region, either jet can serve as the reference object, while in forward regions, the reference jet is taken from a previously calibrated region.

The asymmetry and its resolution are defined as,

$$\mathcal{A} \equiv \frac{p_{\mathrm{T}}^{\mathrm{probe}} - p_{\mathrm{T}}^{\mathrm{ref}}}{p_{\mathrm{T}}^{\mathrm{avg}}} \Rightarrow \sigma_{\mathcal{A}} = \left\langle \frac{\sigma_{p_{\mathrm{T}}}}{p_{\mathrm{T}}} \right\rangle^{\mathrm{probe}} \oplus \left\langle \frac{\sigma_{p_{\mathrm{T}}}}{p_{\mathrm{T}}} \right\rangle^{\mathrm{ref}} , \tag{4.6}$$

where the probe is the jet under study, and $p_{\mathrm{T}}^{\mathrm{avg}}$ is the average transverse momentum of the two jets. The JER is then defined as

$$\mathrm{JER} \equiv \left\langle \frac{\sigma_{p_{\mathrm{T}}}}{p_{\mathrm{T}}} \right\rangle^{\mathrm{probe}} = \sigma_{\mathcal{A}} \ominus \left\langle \frac{\sigma_{p_{\mathrm{T}}}}{p_{\mathrm{T}}} \right\rangle^{\mathrm{ref}} . \tag{4.7}$$

For central regions this gets reduced to JER $\equiv \frac{\sigma_{\mathcal{A}}}{\sqrt{2}}$.

The detector resolution (det) is the quantity of interest, rather than the measured (meas) width of the distribution. The detector contribution can be extracted by subtracting the true width from the measured width

$$\left(\sigma_{\mathcal{A}}^{\text{probe}}\right)_{\text{det}} = \left(\sigma_{\mathcal{A}}^{\text{probe}}\right)_{\text{meas}} \ominus \sigma_{\mathcal{A}}^{\text{true}} \,. \tag{4.8}$$

The true width $\sigma_{\mathcal{A}}^{\text{true}}$ is obtained from simulation using one of the MC generators. In general, the resolution observed in data is worse than in simulation due to detector effects. In such cases, the simulated distributions are smeared to match the resolution measured in data. Conversely, if the data response is found to be narrower in a given phase-space region, no additional smearing is applied.

In contrast to this approach, other studies have also explored JER extraction using the DB method with V+jets in Run 1 [123] and Run 2 [124].

4.4 Flavour tagging

Flavour tagging plays a crucial role in many analyses, as it enables the identification of the flavour of reconstructed jets. This information is used both to enhance the sensitivity to specific final states and to define dedicated trigger selections targeting particular physics signatures. In ATLAS, the development and calibration of flavour-tagging algorithms are performed by the Flavour Tagging CP group.

The GN2 [113] algorithm is a transformer [17] flavour-tagging algorithm, designed for small-R jets. The algorithm assigns, for each jet, probabilities corresponding to different flavour hypotheses: b-jets, c-jets, light-flavour jets (l-jets), and τ-jets. These probabilities are combined into a single discriminant optimised for a given tagging task. For instance, a jet is identified as a b-jet if the b-tagging discriminant, D_b, exceeds a predefined threshold.

Several working points (WPs) are defined to accommodate different efficiency and purity requirements and to enable calibration in the data. The standard GN2 b-tagging working points correspond to b-jet efficiencies of 65%, 70%, 77%, 85%, and 90%. Higher-efficiency working points retain a larger fraction of b-jets but provide lower rejection against non-b-jets, while tighter WPs achieve higher background rejection at the cost of reduced signal efficiency.

Figure 4.4 shows the performance of the GN2 algorithm in terms of background rejection as a function of the b-jet efficiency, and compares it to the previous DL1d tagger, demonstrating improved background rejection at the same efficiency levels.

Tables 6.3 and 6.4 in Section 6.2.1 summarise the flavour fractions for different b-tagging WPs for V+jets MC samples, and are mentioned here for reference.

4.4.1 b-jet corrections

Analyses involving b-jets in the final state often implement additional corrections to account for the specific properties of these jets. A schematic representation of a b-jet is shown in

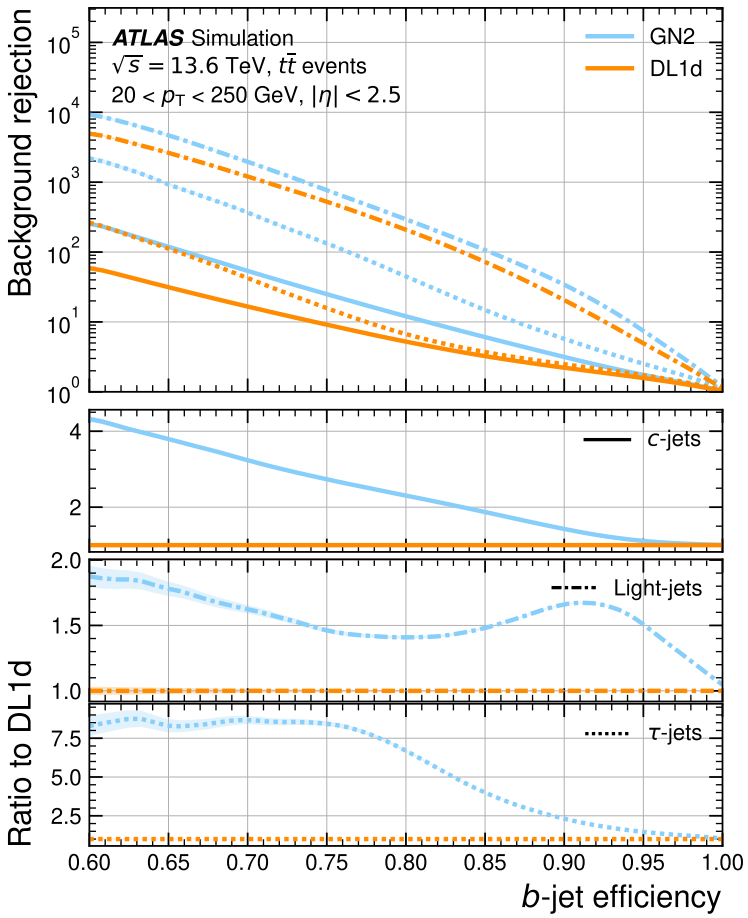


Figure 4.4: Comparison of the background rejection as a function of the b-jet efficiency for the GN2 model (blue) and the DL1d model (orange). The bottom panel shows the relative performance of GN2 with respect to DL1d for each jet flavour: c-jets (solid line), light-flavour jets (dash-dotted line), and τ-jets (dashed line). Figure from Ref. [113].

Fig. 4.5. B-hadrons frequently decay leptonically, and standard jet reconstruction algorithms typically do not include collinear leptons within the jet cone. Moreover, the calorimeter response can vary due to differences in the hadronic shower development or the displaced decay vertex associated with B-hadrons. Both methods, briefly described below, are documented in Ref. [16].

The *muon-in-jet* correction compensates for the presence of soft muons within jets. When a muon is found within $\Delta R = 0.4$ of the jet axis ($\Delta R = 1.0$ for large-R jets), its four-momentum is added to the jet. The energy deposited by the muon in the calorimeter is then subtracted to avoid double-counting. If multiple muons are present, the one closest to the jet axis is selected.

The *PtReco* correction is applied only to small-R jets. It is derived from MC as a function of p_{T} and separately for jets with or without a lepton. The MC uses a truth definition that includes the leptons and neutrinos from the decay. This correction is applied directly in both

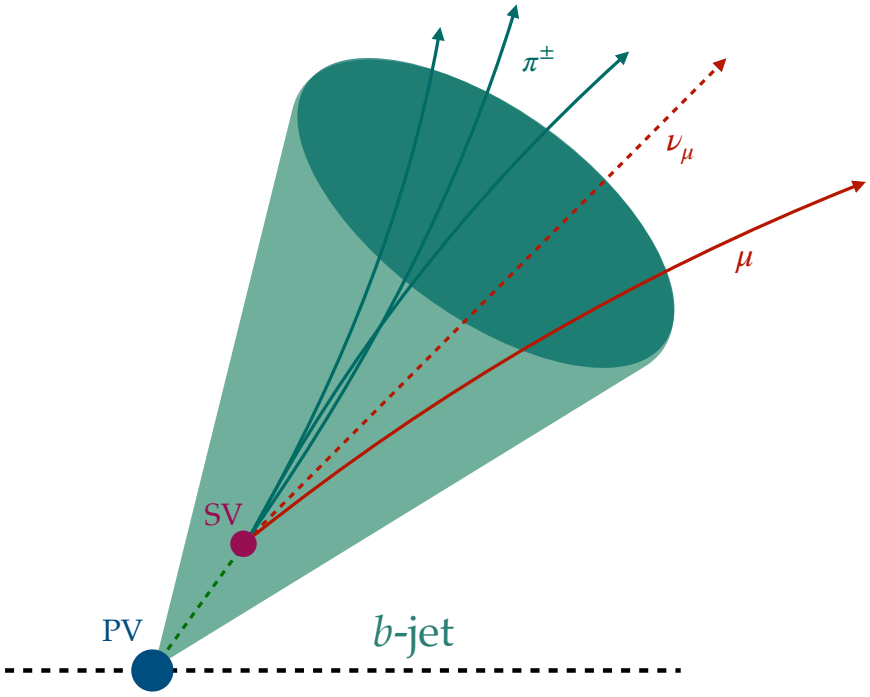


Figure 4.5: Schematic representation of a b-jet, showing a secondary vertex (SV) displaced from the primary vertex (PV) and a semileptonic decay into a muon, neutrino, and pions.

MC and real data as an extra scale factor.

In an MC sample of VH events with the Higgs boson decaying to two b-quarks, these corrections improve the resolution of the reconstructed mass peak by approximately 18% [16].

Chapter 5

Hybrid training for jet domain adaptation

5.1 Introduction

In collider physics, simulation is essential for predicting the behaviour of the experiment. In analyses, control regions are used to validate the modelling of the processes under study by demonstrating good agreement between simulation and data. However, simulation and data do not match perfectly. Simulating both the physics processes and the detector response of one of the most complex machines in the world is not a straightforward task. The reconstruction is highly sensitive to detector imperfections and operational conditions, which introduce differences between data and simulation.

Of particular relevance for Higgs and di-Higgs analyses is the precise reconstruction of jets originating from bottom quarks. The decay $H \to b\bar{b}$ was already observed in Run 2 [15], but limited jet energy resolution reduces the sensitivity by preventing a clear separation between the Higgs signal and the Z-boson background, as shown in Fig. 5.1. This challenge becomes even more pronounced in di-Higgs production due to the smaller cross-section of this process.

In the previous chapter, we described the calibration of Monte Carlo (MC) samples and the subsequent in-situ corrections that cover the remaining gap between simulation and data. In addition to the standard inclusive jet energy calibration, dedicated analysis-level corrections were often applied after identifying b-jets using flavour-tagging algorithms such as GN2. These corrections account for effects specific to heavy-flavour jets, such as semileptonic decays and the presence of undetected neutrinos, which can bias the reconstructed jet response.

The primary goal of this study is to achieve a consistent calibration of b-jets, rather than relying on analysis-level corrections. However, for this initial investigation, inclusive jets are also considered, as they can benefit from the same approach and provide substantially larger event statistics.

In this chapter, we explore a practical method to reduce the domain shift using a hybrid

machine learning (ML) approach. Hybrid training refers to training a model simultaneously on simulated and real data, allowing it to learn the underlying physics from simulation while also learning detector-specific effects directly from data.

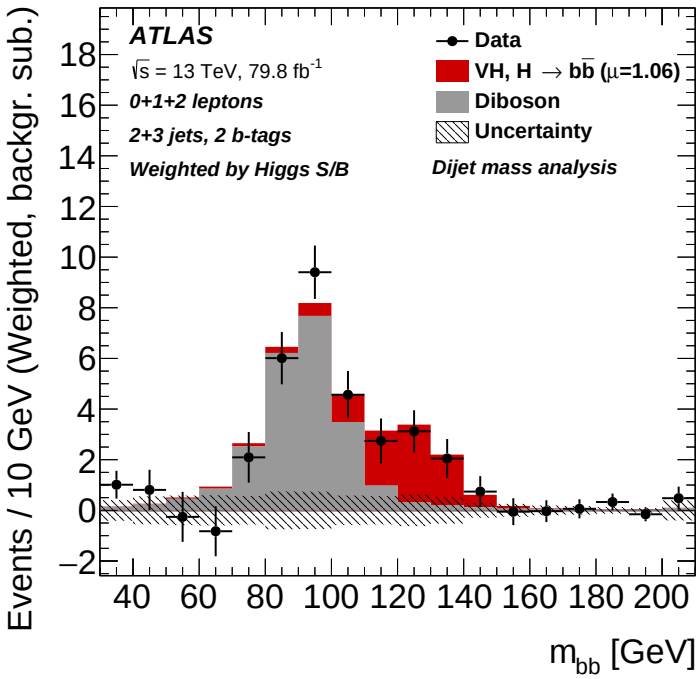


Figure 5.1: Invariant mass of a pair of b-tagged jets in the search for the VH production process, where the Higgs boson decays hadronically and the associated Z or W boson decays leptonically. Data are shown as black dots, the background (diboson sample) in grey, and the signal in red. Figure from Ref. [15].

This chapter is structured as follows. We first describe the samples and event selection, followed by the set of input variables. We then provide an overview of the machine-learning algorithm, explain how the target label is constructed in data, and finally present the results.

5.2 Data and simulation samples

The primary motivation of this project is to improve low-p_{T} b-jet energy resolution in the phase space relevant for the $H \to b\bar{b}$ analysis, as illustrated in Fig. 5.1. In particular, the jets in this channel typically have $p_{\mathrm{T}} \sim 60$ GeV[1]. We therefore focus on calibration techniques that are effective at low-p_{T}, such as the Direct Balance method using V+jets events. As discussed in Section 4.3.2, Z+jets samples populate the low-p_{T} region and are particularly well suited for this purpose. In this analysis, we focus on the $Z(\to \mu^+\mu^-)$+jets final state.

[1]The momentum of the jets is sensitive to the Higgs production mechanism.

We use Run 3 samples because their higher pile-up conditions make them well-suited for studying pile-up dependence, a key aspect of the HL-LHC upgrade. The data samples correspond to pp collisions at a centre-of-mass energy of $\sqrt{s} = 13.6$ TeV, collected by the ATLAS experiment in 2022 during LHC Run 3. The simulated samples are produced under the same detector and pile-up conditions.

The $Z(\mu\mu)$+jets MC sample is generated at next-to-leading order using POWHEG [91–93], interfaced to PYTHIA 8.307 [95] for parton showering, hadronisation, and underlying-event modelling. Decays of heavy-flavour hadrons are handled by EVTGEN (v2.1.1) [125], while final-state QED radiation is added using PHOTOS++ (v3.64). The generator configuration employs the AZNLO [126] tune together with the CTEQ6L1 [127] PDF set to ensure a consistent description of the underlying event and parton distributions across the simulation chain. The MC samples are fully simulated using the ATLAS detector simulation based on the GEANT4 framework [96].

We also considered samples from $Z \to e^+e^-$ processes to gain statistics, but ultimately did not use them due to technical issues with the ntuple production.

5.3 Event selection

The event selection proceeds in two stages, both in simulation and real collision data. First, baseline object and event requirements are applied during the ntuple production using the AnaHelpers (AH) framework [128], which calibrates and skims events[2]. These settings ensure consistent calibrations and minimal selections across all samples. A second selection is then applied during the post-processing stage, following the standard in-situ methodology used for Z+jet balance studies.

We start by applying the standard ATLAS GoodRunLists for 2022 data-taking to ensure data quality. Events are then selected using the dimuon trigger with a symmetric 14 GeV–14 GeV threshold configuration.

Jets are calibrated using the standard Run 3 particle-flow jet energy scale recommendations, including residual pile-up and global sequential corrections. This choice minimises sensitivity to framework-level changes in the other calibration steps. Since incorporating previously calibrated jets may improve performance, this decision can be revisited based on the group's needs. Operationally, the new method could ultimately be applied after in-situ calibration or in parallel with the GSC and in-situ steps, compared to Fig. 4.2.

The ntuples are produced using a configurable AnaHelpers (AH) file, which specifies both the reconstruction calibrations and the minimal kinematic requirements for each physics object. The applied kinematic selections for muons and jets are summarised in Table 5.1, while the corresponding quality requirements are listed in Table 5.2.

The ntuples are further converted to lightweight `parquet` files using the `uproot` package [130], which facilitates efficient handling of variable-length inputs for ML applications.

[2]The dedicated AH repository developed for this project is documented in Ref. [129], which uses *AnalysisBase,22.2.109*.

Object	Multiplicity	p_{T} [GeV]	$\lvert\eta\rvert$
Muons	2	> 14	< 2.4
Jets	≥ 1	> 8	< 4.5
Truth jets	≥ 1	> 8	–

Table 5.1: Baseline kinematic requirements applied at ntuple production for muons, reconstructed jets, and truth jets.

Object	Additional requirements
Jets	JVT (FixedEffPt)
Muons	Identification: Loose ID, $\lvert z_0 \sin\theta\rvert < 0.5$ mm, $\lvert d_0\rvert/\sigma_{d_0} < 3$, Isolation: PFlowLoose_VarRad

Table 5.2: Identification and quality requirements for reconstructed muons and jets.

All samples were processed locally on the computing cluster, resulting in 4.36M MC and 0.43M data events.

The physics selection closely follows the Run 2 in-situ calibration strategy explained in Section 4.3.2.1. Events are required to contain exactly two same-flavour, opposite-charge muons with a reconstructed dilepton mass in the Z window, 66 GeV$< m_{\mu\mu} <$ 116 GeV. Jets overlapping with either muon within $\Delta R < 0.35$ are removed. The remaining jets are ordered by p_{T}, and the leading jet (j_1) is required to be nearly back-to-back with the Z boson $|\Delta\phi_{Z,\mathrm{j}_1}| > 2.9$. Subleading jets ($\mathrm{j}_2$) are required to be soft in order to suppress multi-jet contamination by $p_{\mathrm{T}}^{\mathrm{j}_2} < \max(0.3 \cdot p_{\mathrm{T}}^{Z}, 12)$ GeV[3]. For MC truth-matching, the leading reconstructed jet must be matched to a truth jet within $\Delta R < 0.4$.

The cumulative efficiency of each selection step is shown in Table A.1 in Appendix A.1, where a comparison of the cutflow between data and simulation is also found.

Basic kinematic validation was performed after the selection. The $\Delta\phi$ between the leading jet and subleading jets exhibits a peak at $|\Delta\phi| \sim 0.5$, indicating that subleading jets often lie near the leading jet direction. Consequently, the Z boson typically carries larger p_{T} than the leading truth jet[4]. In addition, the Z p_{T} spectrum shows a characteristic dip around $p_{\mathrm{T}}^{Z} \sim 40$ GeV, which arises from the subleading-jet requirement: the fixed 12 GeV threshold is active until the $0.3 \cdot p_{\mathrm{T}}^{Z}$ criteria becomes dominant. Plots illustrating these characteristics in more detail are available in Appendix A.2.

5.3.1 Input variables

Since this project is exploratory, we include a broad set of variables to assess which features carry relevant information for improving the jet momentum resolution. For this reason, we

[3]The subleading jet p_{T} have to be less than 30% of the $p_{\mathrm{T}}^{\mathrm{ref}}$, or 12 GeV if the latter is smaller.

[4]It reflects on the higher right tail in Fig. 5.4.

choose to work with the DAOD_JETM derivations over the more general DAOD_PHYS, as they retain additional jet- and constituent-level detail.

The input features, summarised in Table 5.3, are organised into three main categories: global event information, variables describing the leading jet and its constituents, and objects used to reconstruct the Z boson and validate the event selection in data. The latter category is used exclusively for Z boson reconstruction and event validation and is not provided as input to the model. It includes both reconstructed and truth-level muons, as well as the kinematic properties of subleading jets. Information on the specific containers used is provided in Appendix A.3.

Jet constituents are taken from the PFlow Objects container, matched to the leading jet within $\Delta R < 0.5$. This is slightly looser than the standard small-R jet radius to account for possible misses. These variables are stored as vectors, which require specialised handling in ML algorithms.

The truth jet used for the regression includes soft leptons and neutrinos within the jet cone, providing a more accurate representation for b-jet decays.

Variables related to Jvt contain, to some extent, information about the jet momentum, which could potentially bias the models. A study of their impact is detailed in Appendix A.4, and it was decided not to include them in the final model.

5.4 Regression Strategy

The goal of this project is to develop a machine-learning–based regression that predicts the transverse momentum p_{T} of each jet. We explore a hybrid training strategy that combines simulation and data. The simulation provides access to the truth-level jet momentum, which serves as the regression target, while using data allows the algorithm to learn the residual differences between simulation and reality. The final aim is to obtain a momentum response that is centred around unity and narrower than that of the nominal ATLAS jet calibration[5]. After applying the regression, residual JES and JER corrections may still be derived in bins of p_{T} and η if needed.

There is no true information in data, so we need to construct a target variable, *label*. This label is designed to approximate the true jet momentum as closely as possible, following an approach conceptually similar to the recoil-based methods used in in-situ calibrations. In this way, the method aims to improve the existing in-situ strategy for b-jets. Since p_{T} is not included among the input features, the algorithm learns the jet response directly from low-level information, potentially replacing part of the MC-based calibration chain.

5.4.1 Machine learning

Machine learning (ML) provides a robust framework for modelling complex, multidimensional relationships. Given a set of input variables (features), an ML model learns the

[5]The nominal calibration refers to the consolidated jet energy scale and resolution calibration.

Event features	**Description**
runNumber	Run number
eventNumber	Event number
NPV	Number of primary vertices
actualInteractionsPerCrossing	Average number of interactions per crossing (LHC)
averageInteractionsPerCrossing	Average number of interactions per crossing (ATLAS)
weights	Event weights
Jet features	**Description**
p_{T}, η, ϕ	Kinematics of leading jet
EnergyPerSampling	Energy per calorimeter sampling
EMFrac	Electromagnetic energy fraction
Width	Jet cone width
NumTrkPt1000PV, TrackWidthPt1000PV	Ghost-associated track moments ($p_{\mathrm{T}} > 1$ GeV)
NumTrkPt500PV, SumPtTrkPt500PV	Ghost-associated track moments ($p_{\mathrm{T}} > 0.5$ GeV)
Jvt, JvtRpt, JvtJvfcorr *	Jet vertex tagging quantities
GN1_pb	b-tag score
Charged & Neutral PFO features	**Description**
p_{T}, η, ϕ, m, charge	Kinematics of PFlow objects
vertexType, signalType, isInDenseEnvironment	Quality and environment flags
z0, vz, d0, theta, CaloCorrectedEnergy	Track and geometry info for charged PFOs
TracksExpectedEnergyDeposit	Expected energy deposit from associated tracks (charged)
LayerEnergy	Energy in each calorimeter layer (neutral)
EM_Probability	Probability to be electromagnetic (neutral)
avg_lar_q, avg_tile_q	Average calorimeter charges (neutral)
Isolation	Isolation variable (neutral)
badCells	Flag for bad calorimeter cells (neutral)
Track features	**Description**
Truth, d0, z0, vz, qOverP	Truth-matching and basic track kinematics
Muon features	**Description**
p_{T}, η, ϕ	Kinematics of muons

Table 5.3: Summary of variables used as inputs to the ML regression and for event validation. Full details of the containers from which the variables are derived are provided in Appendix A.3.

mapping between the features and a specified target during a training phase consisting of many iterations. After each iteration, the model prediction is evaluated using a metric, and the model parameters are updated accordingly. In this work, two ML architectures are investigated: a boosted decision tree and a transformer network.

5.4.1.1 Boosted Decision Tree

A boosted decision tree (BDT) [131, 132] is a widely used ML model in particle physics due to its simplicity, interpretability, and strong performance on tabular data. In this work, BDTs serve as a baseline against which the second architecture can be compared.

The implementation used is RapidGBM [133], based on LightGBM [134] combined with Bayesian hyperparameter optimisation [135]. BDTs cannot natively process vector-valued inputs, making them less suitable for exploiting the complete particle-flow constituent information in jets.

Training is performed using 50 optimisation trials. The optimisation metric is the root-mean-squared percentage error. The optimised hyperparameters include a learning rate of 0.094, 177 leaves, a maximum depth of 61, a dropout rate of 0.161, and a minimum sum of Hessians per leaf of 2.792. The whole training procedure takes approximately 20 minutes with the available dataset.

5.4.1.2 Transformer

Transformers [17] have recently gained significant attention in the ML community. They are fully connected neural networks that rely on self-attention to capture long-range and global relationships within the input, making them particularly well suited to variable-length inputs such as particle-flow constituents.

A limitation of transformer models is the increased training time and greater difficulty in hyperparameter optimisation and feature importance estimation compared to BDTs.

The transformer architecture is implemented using `PyTorch` [136]. Figure 5.2 shows schematically the configuration used in this study. A single layer of self-attention is applied to all particle-flow constituents, and the resulting representation is combined with scalar jet-level variables in a fully connected multilayer perceptron (MLP). The MLP contains two hidden layers with 128 neurons each, using ReLU activations in the hidden layers and a hyperbolic tangent activation in the output layer. Training is performed with a learning rate of 0.001, a dropout rate of 0.2, a batch size of 500 (depending on GPU memory), and a maximum of 2 000 epochs. The total training time on local GPUs is approximately 32 hours.

5.4.2 Constructing target labels for jet p_{T} estimation

Having defined the network architectures and inputs, the next step is to construct suitable target labels for training. In Monte Carlo (MC) simulation, the target is the truth jet p_{T}, which is used both for training and for performance evaluation. In data, where true information is unavailable, approximate labels must be constructed.

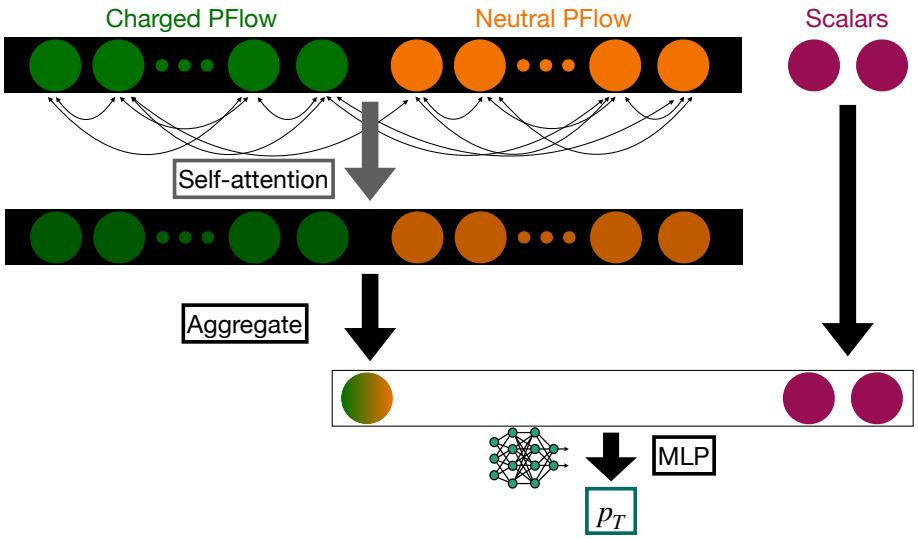


Figure 5.2: A transformer representation with one layer of self-attention applied to all PFlow features (green and orange), followed by aggregation and subsequent combination with scalar features (magenta) using an MLP.

Using the Z+jets topology (Fig. 5.3), events in which the leading jet and the Z boson are approximately back-to-back in azimuth allow an approximate label to be defined as

$$p_{\mathrm{T}}^{\mathrm{label}} = p_{\mathrm{T}}^{Z}\,, \tag{5.1}$$

where the Z boson acts as a reference object. However, most of the time, the event contains other jets, which makes this approximation inaccurate.

To better account for additional activity, we can extend this label description following an approach similar to the MPF method described in Section 4.3.2.1 of trying to use the rest of the hadronic activity of the event, by including subleading jets projected onto the leading-jet axis:

$$p_{\mathrm{T}}^{\mathrm{label}} = \sum p_{\mathrm{T}}^{\mathrm{rest}} = \cos\phi_Z \cdot p_{\mathrm{T}}^{Z} + \sum \cos\phi_{\mathrm{j}_2} \cdot p_{\mathrm{T}}^{\mathrm{j}_2}\,. \tag{5.2}$$

These labels are constructed exclusively from reconstructed objects and are validated in simulation by comparison to truth-level quantities. Figure 5.4 compares the responses obtained using Eqs. 5.1 (blue) and 5.2 (red). Including subleading jets broadens the response due to the poorer resolution of low-p_{T} jets, but yields a more symmetric distribution with a mean closer to unity. The response obtained by evaluating Eq. 5.2 using truth-level quantities is also shown (purple), illustrating the resolution that could be achieved if the input quantities were perfectly calibrated.

While including subleading jets increases the available training statistics, it can degrade the label resolution. This trade-off can be quantified through error propagation of Eq. 5.2:

$$\sigma_{p_{\mathrm{T}}^{\mathrm{label}}} = \sqrt{\left(\cos\phi_Z \cdot \sigma_{p_{\mathrm{T}}^{Z}}\right)^2 + \sum \left(\cos\phi_{\mathrm{j}_2} \cdot \sigma_{p_{\mathrm{T}}^{\mathrm{j}_2}}\right)^2}\,. \tag{5.3}$$

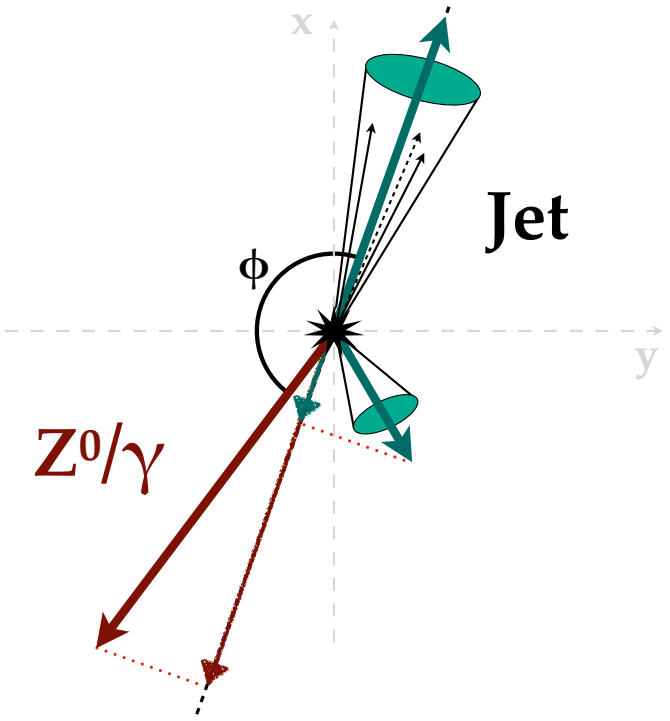


Figure 5.3: Schematic representation of a Z+jets event with a leading jet nearly back-to-back with the Z boson and one subleading jet. The Z boson and the subleading jet are projected onto the leading-jet axis to construct the label momentum, $p_{\mathrm{T}}^{\mathrm{label}}$.

The distribution of $\sigma_{p_{\mathrm{T}}^{\mathrm{label}}}$ can be used to define an optimal selection that balances label quality and statistics. In Appendix A.5, we explore some ideas of using such a label and how we could use its uncertainty. However, due to limited data statistics, no additional selections are applied in this study, and Eq. 5.2 is used directly to construct the label.

5.5 Results

This section presents the evaluation of ML-based jet p_{T} regression models. The analysis focuses on three main aspects: validating the regression in simulation, evaluating performance in real data using approximate labels, and extracting the jet energy resolution in data. Performance metrics, such as the response and resolution, are defined and applied consistently. The results are then interpreted to compare the ML models with the nominal ATLAS calibration and to assess potential improvements.

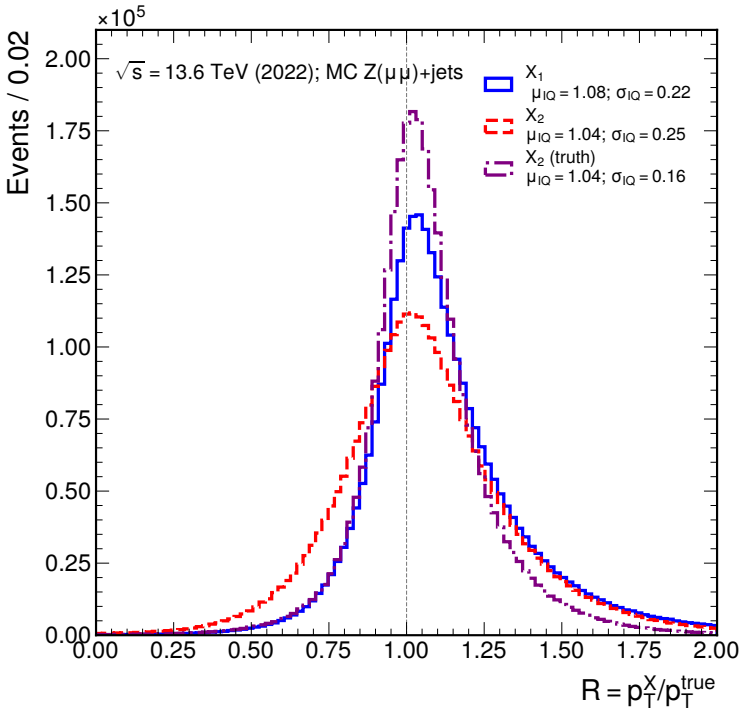


Figure 5.4: Response of different label constructions relative to the true jet p_{T}. The solid blue line uses only the Z boson ($p_{\mathrm{T}}^{\mathrm{label}} = p_{\mathrm{T}}^{Z}$), while the dashed red line includes subleading jets as defined in Eq. 5.2. The dash-dotted purple line shows the same construction evaluated using truth-level quantities instead. The mean μ_{IQ} and width σ_{IQ} are computed from the interquartile range.

5.5.1 Introduction and performance definitions

The performance of the regression models is quantified using the response:

$$\mathcal{R} = \frac{p_{\mathrm{T}}^{\mathrm{pred}}}{p_{\mathrm{T}}^{\mathrm{truth}}} \quad (\mathrm{MC})\,, \qquad \mathcal{R} = \frac{p_{\mathrm{T}}^{\mathrm{pred}}}{p_{\mathrm{T}}^{\mathrm{label}}} \quad (\mathrm{data})\,. \tag{5.4}$$

After binning in p_{T}or η, the response and resolution of a given distribution is extracted using the inter-quantile range (IQR), with the mean taken as the 50% quantile and the width as half of the difference between the 16% and 84% quantiles. This robust method is commonly used in ATLAS analyses, offering a faster alternative to Gaussian fits.

To construct resolution plots, each bin's width is normalised to its mean, $\tilde{\sigma} = \sigma/\mu$, penalising distributions with off-centred responses. Additionally, for comparison to the nominal calibration, the ratio of the predicted resolution to the nominal resolution is computed. These quantities are evaluated for both ML architectures investigated in this study: the BDT and the transformer network, after tuning their respective hyperparameters.

To avoid bias in training and evaluation, events are split by EventNumber: events ending in 0 or 1 are reserved for evaluation, while all others are used for training. This

scheme ensures reproducibility and consistency across analyses. Statistical uncertainties in the resolution are derived via error propagation from the median and IQR widths and are reported in all subsequent plots unless noted otherwise.

5.5.2 Regression performance

This subsection presents the performance of the ML regression models for jet p_{T} in simulation and real data, and in a hybrid training strategy combining both. The goal is to evaluate how well the models reproduce the truth or approximate jet momentum and to compare them with the nominal ATLAS calibration.

5.5.2.1 MC-only results

We first validate the models using simulated events, where the truth jet p_{T} is known. Jets are binned by $p_{\mathrm{T}}^{\mathrm{truth}}$, and the response and resolution are computed from $\mathcal{R} = p_{\mathrm{T}}^{\mathrm{pred}}/p_{\mathrm{T}}^{\mathrm{truth}}$.

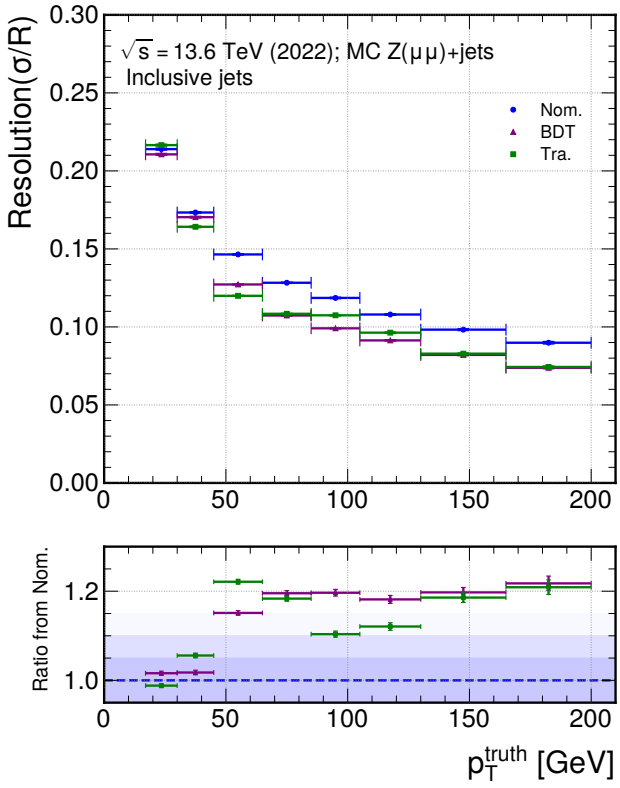


Figure 5.5: Jet p_{T} resolution for all inclusive jets in MC, comparing the nominal ATLAS calibration (blue dots), the BDT model (purple triangles), and the transformer model (green squares). Jets are binned by $p_{\mathrm{T}}^{\mathrm{truth}}$, and the bottom panel shows the ratio of the ML models to the nominal calibration.

Figure 5.5 shows the resolution for inclusive jets, comparing the BDT and transformer

models to the nominal ATLAS calibration. Both models improve the resolution by roughly 20% across the p_{T} range, with the BDT showing slightly better stability above 60 GeV.

When separating jets by flavour (light, c-, and b-jets) using HadronConeExclTruthLabelID, the most considerable improvement is observed for b-jets, with a resolution gain of ∼25% relative to the nominal calibration (Fig. 5.6). This demonstrates that b-jets benefit most from the ML regression, making them a critical target for calibration.

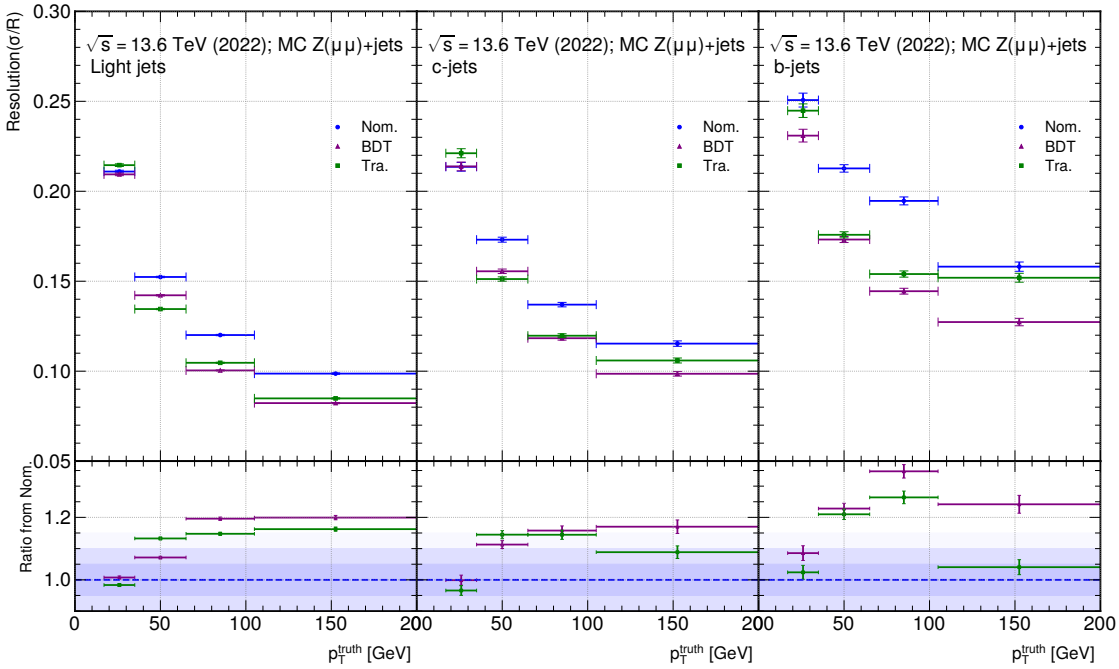
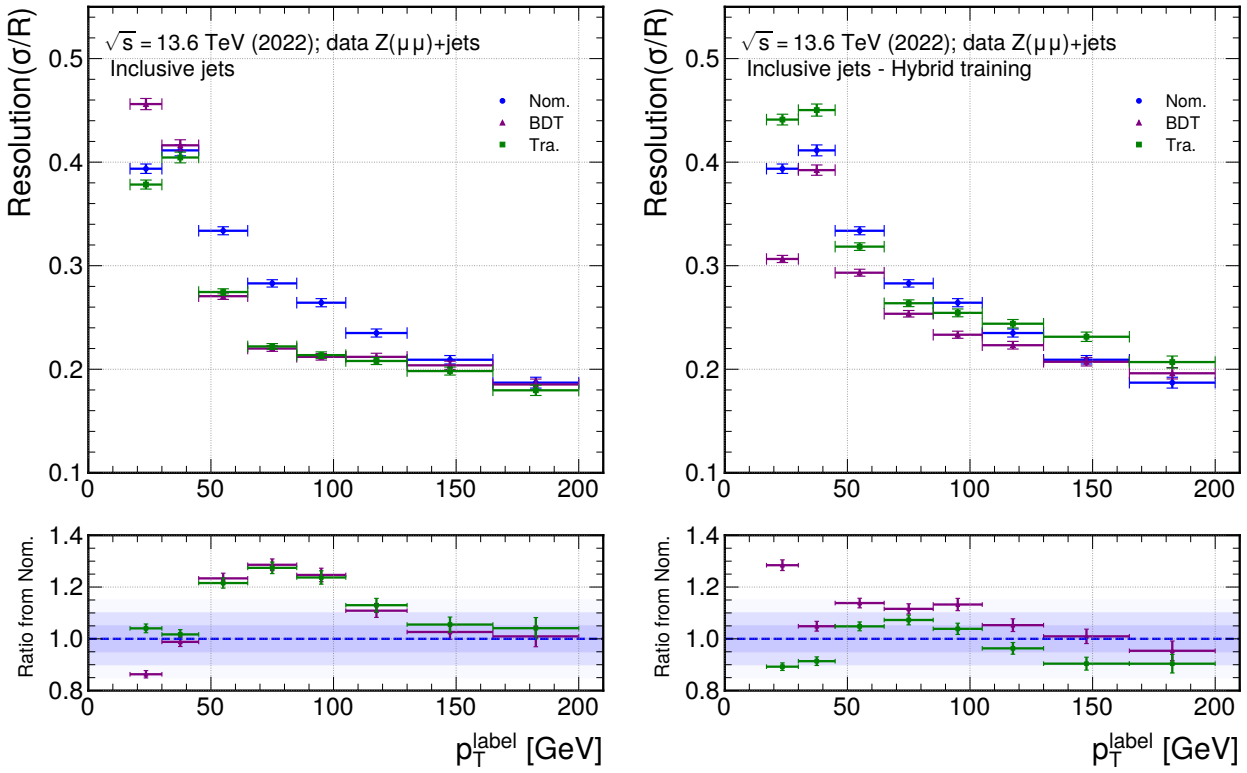


Figure 5.6: Jet p_{T} resolution in MC for light, c-, and b-jets, comparing the nominal ATLAS calibration (blue dots), the BDT model (purple triangles), and the transformer model (green squares). Jets are binned by $p_{\mathrm{T}}^{\mathrm{truth}}$, and the bottom panel shows the ratio of each ML model to the nominal calibration for each flavour.

5.5.2.2 Real data-only results

The models are next trained and evaluated on real data, where the truth jet p_{T} is not available. Approximate target labels are constructed from the Z+jets topology as defined in Eq. 5.2, and the response and resolution are computed with respect to these labels. Figure 5.7a shows the results for inclusive jets. Both the BDT and transformer perform similarly above 45 GeV, while deviations at lower p_{T} likely arise from edge effects in the training. The most significant improvement over the nominal calibration occurs in the 45–105 GeV range, with an approximate 25% reduction in resolution.

Compared with MC , the observed resolutions in data are generally larger. While some degradation is expected due to the complexity of real data, the dominant factor is the use of



(a) Data-only training

(b) Hybrid training

Figure 5.7: Jet p_{T} resolution for inclusive jets in collision data, trained using (a) data only and (b) the hybrid approach. The nominal ATLAS calibration (blue squares) is compared with the BDT model (purple triangles) and the transformer model (green squares). Jets are binned in $p_{\mathrm{T}}^{\mathrm{label}}$, and the lower panel shows the ratio of each ML model to the nominal calibration.

the approximate $p_{\mathrm{T}}^{\mathrm{label}}$ as the target, which affects both training and binning. This motivates further studies (Section 5.5.3) to refine the detector resolution in real data.

5.5.2.3 Hybrid MC + data results

Finally, we explore a hybrid training strategy combining MC and real data. The idea is for the model to learn the main physics properties from MC while simultaneously capturing differences between simulation and data.

In this study, events from MC and data are combined during training using the same event splitting scheme, but evaluation is performed on real data only. MC events provide truth-based supervision, whereas real data events contribute with approximate $p_{\mathrm{T}}^{\mathrm{label}}$ targets. No explicit weighting between MC and data is applied, although differences in dataset size may affect the training dynamics.

Figure 5.7b shows the inclusive jet resolution for the hybrid model evaluated on real

data. Compared to the data-only model, the resolution is generally higher, with some bins even performing worse than the nominal ATLAS calibration. The most stable improvement occurs in the 45–105 GeV range, which corresponds to the main target of this study. This instability likely arises from the statistical imbalance between MC and data, the combination of labels with differing precision, and the lack of weighting during training.

Results comparing the resolution with pseudorapidity or pile-up binning can be found in Appendix A.6.

Future work could explore pre-training on MC followed by fine-tuning on data, or alternative strategies to balance the influence of MC and real data to stabilise the hybrid model while preserving the benefits of both sources of information.

5.5.3 Estimating Jet Energy Resolution in Data

In real data, the truth jet transverse momentum is not accessible, and the measured jet resolution therefore contains contributions from both the regression method (ML-based or nominal calibration) and the intrinsic uncertainty of the approximate label. From this point onward in this chapter, ML will refer specifically to the BDT model. In simulation, the detector resolution can be computed directly as

$$\text{JER} \equiv \tilde{\sigma}_{\text{det}} = \tilde{\sigma}\left(p_{\text{T}}^{\text{pred}}/p_{\text{T}}^{\text{truth}}\right) . \tag{5.5}$$

To emulate the situation when truth is not present, three resolution components are defined in MC:

- $\tilde{\sigma}_{\text{obs}}$: the relative resolution observed after regression, $p_{\text{T}}^{\text{pred}}/p_{\text{T}}^{\text{label}}$,
- $\tilde{\sigma}_{\text{label}}$: the relative resolution of the label itself, $p_{\text{T}}^{\text{truth}}/p_{\text{T}}^{\text{label}}$, evaluated in MC,
- $\tilde{\sigma}_{\text{det}}$: the actual detector-level resolution, $p_{\text{T}}^{\text{pred}}/p_{\text{T}}^{\text{truth}}$.

In real data, $\tilde{\sigma}_{\text{det}}$ or $\tilde{\sigma}_{\text{label}}$ cannot be measured directly. But these resolutions are approximately related by

$$\text{JER} \equiv \tilde{\sigma}_{\text{det}} = \tilde{\sigma}_{\text{obs}} \ominus \tilde{\sigma}_{\text{label}} . \tag{5.6}$$

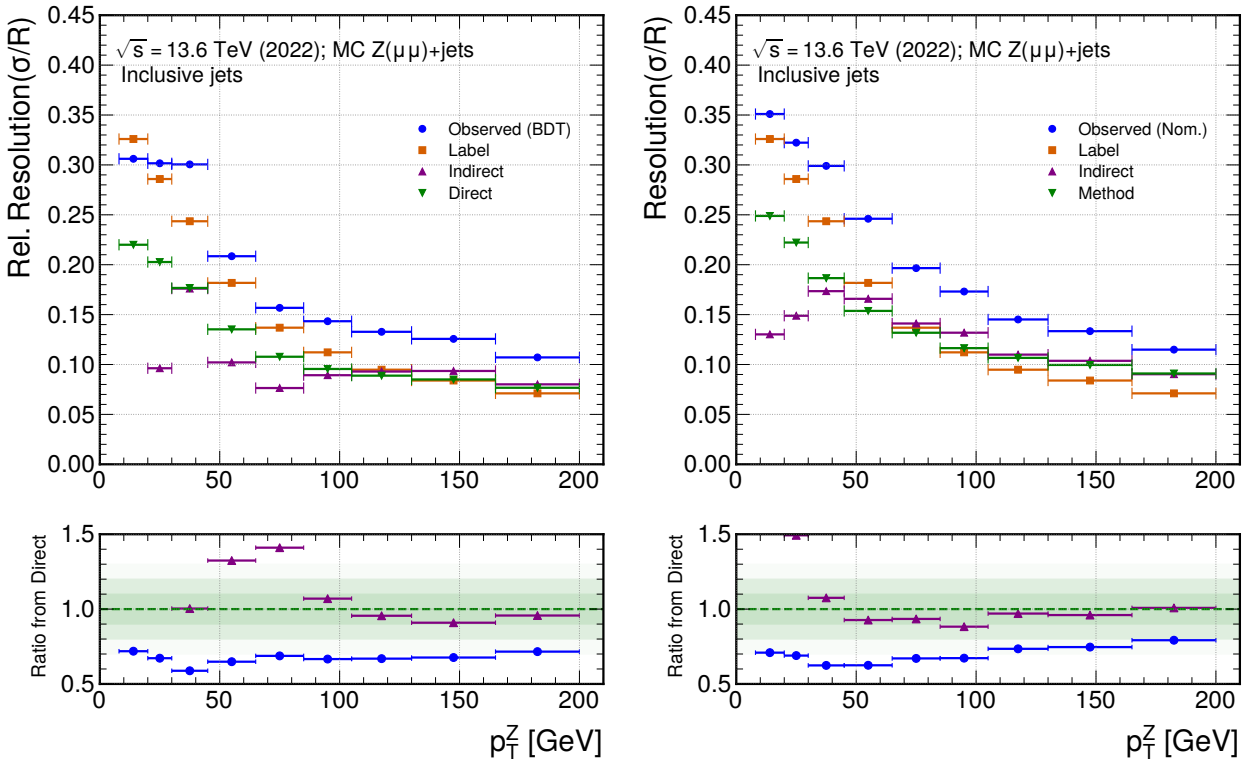
Assuming Gaussian response shapes and negligible correlation between the method and the label, an indirect detector resolution estimate can be obtained via subtraction in quadrature:

$$\text{JER} \approx \sqrt{\tilde{\sigma}_{\text{obs}}^2 - \tilde{\sigma}_{\text{label}}^2} . \tag{5.7}$$

To minimise the impact of additional radiation on the label definition, this study is restricted to events with a single back-to-back leading jet. This selection suppresses contributions from subleading jets and provides a clean topology for resolution studies.

Figure 5.8 compares the observed, label, detector indirect, and detector direct resolutions in MC. In MC, the indirect resolution is consistently closer to the truth-level resolution than the observed one. The regression (Fig. 5.8a) improves the observed resolution relative to the

nominal calibration (Fig. 5.8b). The direct resolution represents $\tilde{\sigma}_{\text{det}}$, while the indirect is the approximation using Eq. 5.7. Big deviations between the indirect and direct resolutions are observed at low p_{T}, reflecting the non-negligible contribution of the label uncertainty. At higher p_{T} (above 100 GeV), the direct and indirect resolutions converge, demonstrating good closure.



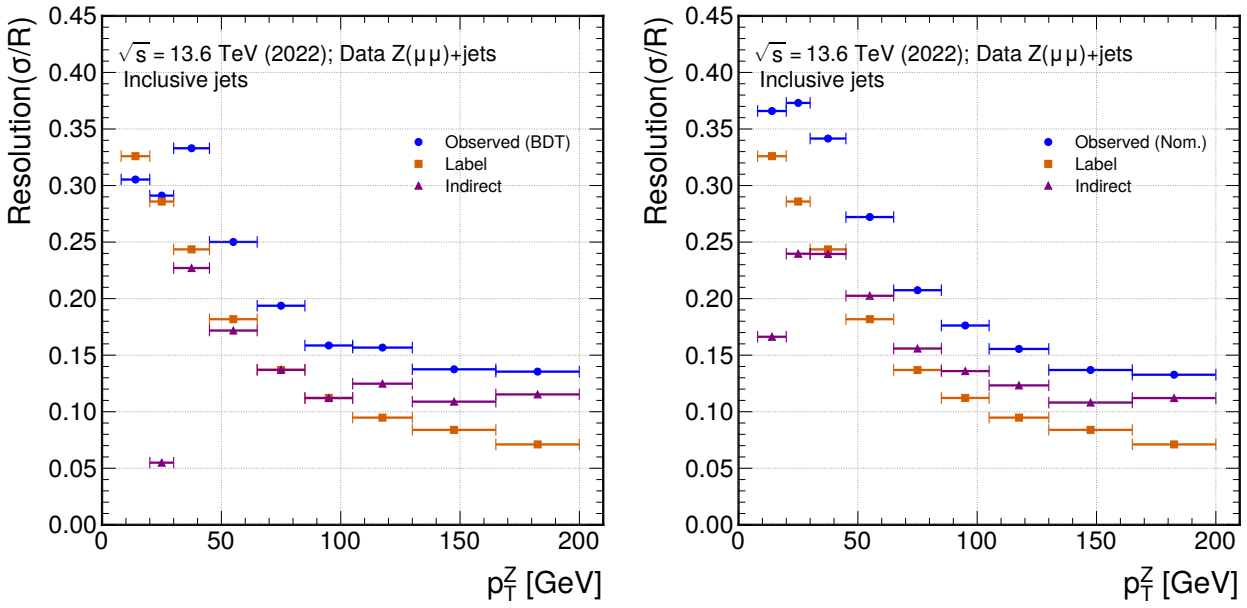
(a) BDT (b) Nominal

Figure 5.8: Jet resolution components in MC as a function of p_{T} for (a) the ML-based regression and (b) the nominal ATLAS calibration. Observed resolution ($p_{\text{T}}^{\text{pred}}/p_{\text{T}}^{\text{label}}$, blue dots); label resolution ($p_{\text{T}}^{\text{truth}}/p_{\text{T}}^{\text{label}}$, orange squares); indirect detector resolution from Eq. 5.7 (purple upward triangles); direct detector resolution ($p_{\text{T}}^{\text{pred}}/p_{\text{T}}^{\text{truth}}$,green downward triangles). The lower panel shows the ratio of each component to the direct resolution.

Figure 5.9 compares the same procedure in data. The indirect resolution provides the best available estimate of the method performance, using the label resolution derived from MC in Eq. 5.7. This approach enables a consistent determination of the JER in both MC and data, which can then be used to derive in-situ corrections.

5.5.3.1 In-Situ Corrections in Data

We extract the JER in data using the indirect method described above, which allows us to directly compare MC and real data and derive in-situ correction factors. Figure 5.10 shows



(a) BDT (b) Nominal

Figure 5.9: Jet resolution components in data as a function of p_T for (a) the ML-based regression and (b) the nominal ATLAS calibration. Observed resolution (p_T^{pred}/p_T^{label}, blue dots); label resolution from MC (p_T^{truth}/p_T^{label}, orange squares); indirect detector resolution from Eq. 5.7 (purple upward triangles, using the label resolution derived from MC).

the JER measured in MC and data using the DB method, along with their ratio, which we use to correct the MC-based calibration.

The ML-based method (left panel) generally produces a smaller JER than the nominal calibration (right panel), but the nominal method remains more stable across the full kinematic range and achieves better closure with MC. We quantify non-closure in Fig. 5.11 as

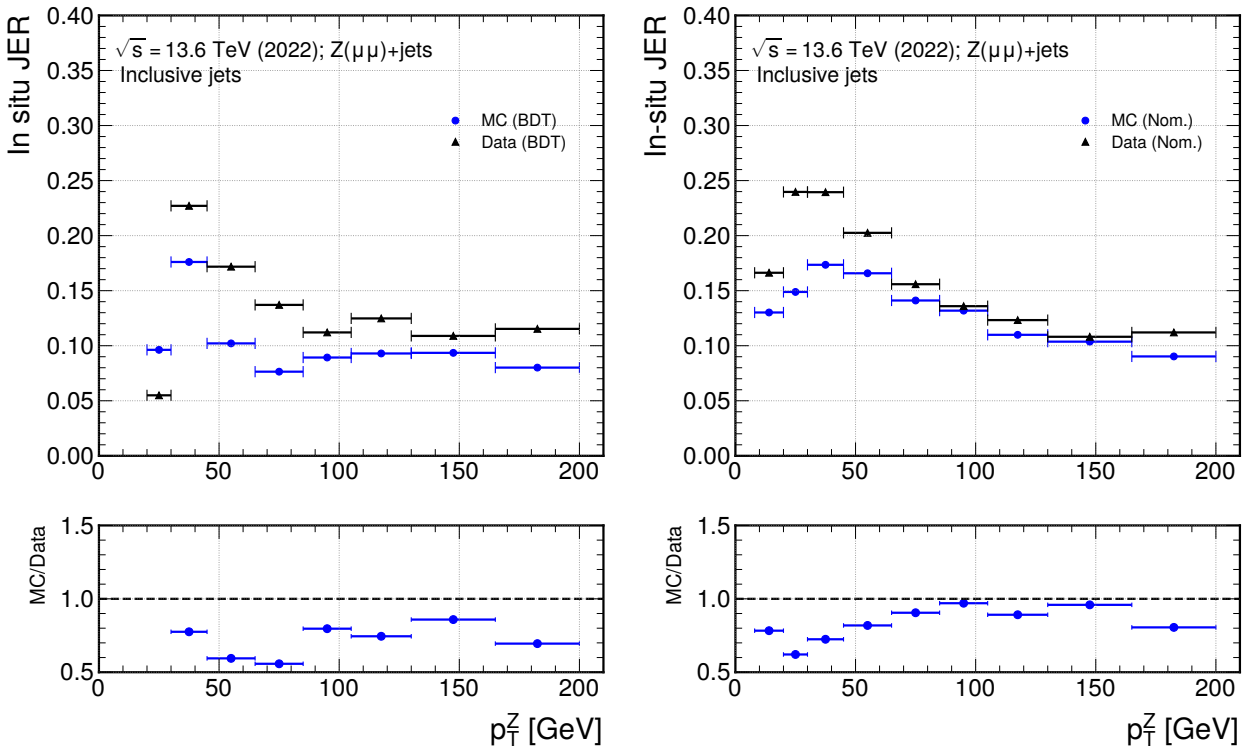
$$\text{Non-closure} = \frac{\tilde{\sigma}^{\text{MC}}_{\text{direct}} - \tilde{\sigma}^{\text{MC}}_{\text{indirect}}}{\tilde{\sigma}^{\text{Data}}_{\text{indirect}}}, \tag{5.8}$$

which confirms the improved stability of the nominal calibration.

This comparison provides a quantitative basis for correcting the resolution measured in data. Residual differences at low p_T indicate that we may need to use alternative control samples (e.g., semileptonic $t\bar{t}$ events) or improve the label definitions to apply the indirect method reliably across the full kinematic range.

5.6 Conclusions and outlook

This project has highlighted the challenges posed by the domain shift between simulation and real data. We explored the possibility of unifying the MC-based and in-situ calibrations into



(a) BDT (b) Nominal

Figure 5.10: Jet energy resolution in MC (blue dots) and data (black triangles) for the (a) ML-based regression and (b) the nominal ATLAS calibration. The lower panels show the JER ratio of MC to data.

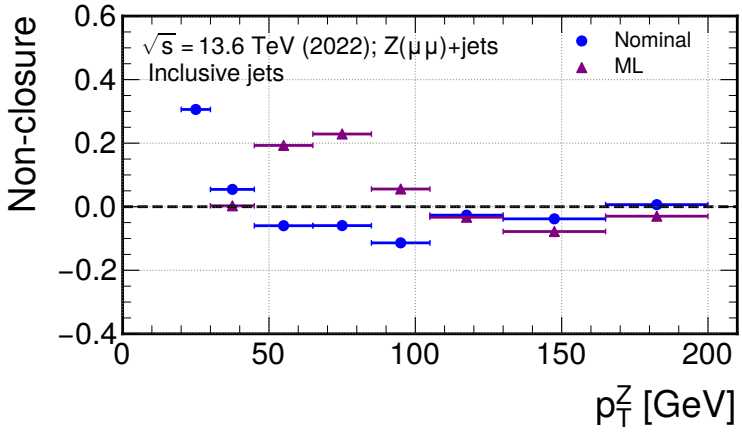


Figure 5.11: Non-closure test for the nominal calibration (blue dots) and the ML-based method (purple crosses).

a single ML procedure by using a hybrid training strategy that simultaneously incorporates

both event types.

In simulation, both ML models, the BDT and the transformer, showed substantial improvements over the nominal calibration, in line with expectations. However, when applying the models to data alone, the reconstructed scale increases, which is attributed to the label construction. In this regime, some degradation in performance is observed, although significant improvements relative to the nominal calibration remain at 45–105 GeV.

In the hybrid training configuration, the ratio to the nominal calibration fluctuates, and the transformer performs considerably worse than the BDT. Although the BDT was intended mainly as a baseline, it ultimately outperformed the transformer in almost all evaluated scenarios. Nonetheless, the transformer remains a promising architecture, particularly because it can incorporate detailed information about jet constituents. In this study, training such a model was constrained by the limited computational resources available. With current ATLAS and institute resources, large-scale transformer training would be far more feasible.

A potential improvement would be a two-step training procedure: first train on MC only, freeze the weights, and then fine-tune on real data. This would leverage the larger MC statistics without overwhelming the model during hybrid training. Exploring alternative transformer architectures is another promising direction, for example, introducing separate self-attention layers for charged and neutral jet constituents.

Although the hybrid ML approach explored in this chapter demonstrates promising performance in simulation and provides valuable insight into the interplay between MC-based and in-situ calibration strategies, it was not adopted in the final ATLAS b-jet calibration framework. The primary reasons are the observed instabilities of the hybrid approach when applied to data, as well as the availability of an alternative, well-established calibration framework (Section 6.1) that has been successfully employed in other ATLAS analyses. This baseline approach is trained exclusively on simulation, providing a more controlled environment and a more straightforward interpretation of systematic uncertainties.

In particular, the understanding gained of the domain shift between simulation and data, together with the treatment of resolution effects and label construction, directly motivates the in-situ calibration and correction procedures developed in Section 6.2.

Chapter 6

Calibration of b-jets

Alongside the developments discussed in the previous chapter, an alternative calibration framework was proposed and documented in Ref. [18], focusing exclusively on improvements at the simulation level. In this approach, ML models are trained on MC simulation to calibrate b-jets, and their performance is subsequently validated on data to assess whether a dedicated b-jet in-situ calibration is required in addition to the inclusive in-situ calibration.

This chapter presents the training strategy and performance of the MC-based ML calibration, together with the methodology used to extract the jet energy scale (JES) and jet energy resolution (JER) in data. As a reference, the JER obtained with the nominal calibration is also reported as a preliminary result[1].

6.1 Transformer MC-based b-jet calibration

In parallel with the hybrid studies described in the previous chapter, alternative machine-learning approaches were explored with a different focus. In particular, a transformer-based regression model was developed to calibrate the energy and momentum of b-jets using simulation only. This approach was ultimately adopted for the b-jet calibration. Two independent models are employed: one targeting small-R jets and another designed for large-R jets.

The following sections describe the simulated samples used for training and validation, the object definitions, the model architecture, and the resulting performance. A complete description of the project is provided in Ref. [18].

6.1.1 Samples and object definitions

The training and validation of the transformer models rely on simulated pp collision samples produced for ATLAS Run 2 at a centre-of-mass energy of $\sqrt{s} = 13$ TeV. Separate datasets are used for the small-R and large-R jet calibrations, reflecting the different physics regimes and object definitions. The training is performed on a jet-by-jet basis, allowing multiple jets from the same event to contribute independently. To enhance performance for b-jets, the samples

[1] At the time of writing, the small-R ML model was not yet available in data samples.

are enriched in this flavour while retaining jets of other flavours to ensure robustness against misidentification. A summary of all simulated samples used for training and evaluation of the small-R and large-R models, together with their generators and PDF sets, is given in Table 6.1.

Small-R jets. The small-R training dataset consists primarily of jets from $t\bar{t}$ events, with at least one top quark decaying leptonically, and from Z+jets events where the Z boson decays into a pair of muons. The combined dataset contains approximately 266 million b-jets from $t\bar{t}$ decays and 19 million jets from Z+jets production, with a flavour composition of light:charm:bottom jets in the ratio 9:1:9.

The model performance is evaluated using independent samples that are not included in the training. In particular, hadronically decaying Z and Higgs bosons produced in association with a leptonically decaying Z boson are used to study the reconstructed mass peaks, providing a robust validation of the calibration performance.

Large-R jets. The training of the large-R jet transformer requires samples containing boosted heavy objects. Signal samples consist of Higgs bosons decaying into $b\bar{b}$ or $c\bar{c}$, produced in association with a Z boson decaying into a pair of muons. To provide uniform coverage over a wide jet mass range, dedicated samples are generated with an artificially increased Higgs boson width, resulting in a flat mass distribution.

Additional training samples include QCD multijet events, which provide a large population of generic jets, as well as a dedicated multijet $b\bar{b}$ sample to enhance sensitivity to heavy-flavour jets further. The whole training dataset comprises 15 million flat-mass $H \to b\bar{b}$ jets, 15 million flat-mass $H \to c\bar{c}$ jets, 30 million QCD multijet jets, and 30 million QCD multijet $b\bar{b}$ jets.

For evaluation, approximately 800 000 jets from QCD multijet events are used to verify that the calibration introduces no artificial mass peaks. In addition, boosted resonance samples are employed, including 15 000 SM $H(b\bar{b})$ jets, 575 000 $Z \to b\bar{b}$ jets, and 725 000 top-quark jets originating from a hypothetical Z' boson with a mass of 4 TeV.

All jets considered in this study are reconstructed using the anti-k_t algorithm [118], implemented with `FastJet` [143]. For small-R jets, the clustering inputs are particle-flow objects that combine tracking and calorimeter information. Large-R jets are reconstructed using unified flow objects, which merge track-calorimeter clusters and particle-flow objects into a single collection.

Truth jets are generally constructed by clustering stable particles that deposit a significant amount of energy in the calorimeter. As a consequence, muons and neutrinos are generally excluded from the truth-jet definition. Since B-hadrons can decay semileptonically, including the charged leptons from these decays improves the jet energy resolution. Accordingly, a lepton-inclusive truth-jet definition is used for small-R jets in this study. For large-R jets, such a definition is not yet available, and the standard calorimeter-based truth-jet definition is therefore employed.

Process	**Event generator and tune**	**PDF set**
	Small-R samples	
$t\bar{t}$	Powheg [91–93] & Pythia 8.230 [95] with A14 [138]	NNPDF3.0nlo [137]
$Z(\mu\mu)$+jets	MadGraph5_aMC [89]+FxFx [139] & Pythia 8.245 with A14	NNPDF3.0nlo
$Z(\ell\ell)H(b\bar{b})$ †	Powheg Box v2 + MiNLO [140, 141] & Pythia 8.230 with AZNLO [126]	NNPDF3.0nlo
$Z(\ell\ell)Z(b\bar{b})$ †	Sherpa 2.2.11 [90]	NNPDF3.0nnlo
	Large-R samples	
$q\bar{q} \to Z(\mu\mu)H(b\bar{b})$	Pythia 8.306 with A14	NNPDF3.0nlo
$q\bar{q} \to Z(\mu\mu)H(c\bar{c})$	Pythia 8.306 with A14	NNPDF3.0nlo
QCD Multijet	Pythia 8.235 with A14	NNPDF2.3lo [142]
QCD Multijet ($b\bar{b}$)	Pythia 8.235 with A14	NNPDF2.3lo
$Z(f\bar{f}/q\bar{q})H(b\bar{b})$ †	Powheg v2 & Pythia 8.212 with AZNLO	NNPDF3.0nlo
Top $Z' \to t\bar{t}$ †	Pythia 8.235 with A14	NNPDF2.3lo
$Z \to b\bar{b}$ †	Sherpa 2.2.11	NNPDF3.0nnlo
QCD Multijet†	Pythia 8.235 with A14	NNPDF2.3lo

Table 6.1: Details of the Monte Carlo samples used for the training and evaluation (marked with a †) of the small-R and large-R jet models. In the table, $f = e, \mu, \tau, \nu$. Table adapted from Ref. [18].

The input features used for the transformer-based b-jet calibration closely follow those of the GN2 and GN2X [113, 144] flavour-tagging networks, with adaptations for regression. They include global jet kinematics, track-level information, and features derived from charged and neutral particle-flow objects for small-R jets, and from unified flow objects (UFOs) for large-R jets. For small-R jets, additional information from soft leptons associated with the jet is included.

Further details on object definitions, kinematic selections, reconstruction requirements (including transverse momentum thresholds and quality criteria), and the complete set of input features are provided in Ref. [18].

6.1.2 Model architecture

The b-jet calibration employs transformer-based architectures originally developed for flavour tagging, namely the GN2 and GN2X models for small-R and large-R jets. These architectures are adapted from classification to regression by modifying the output head, while preserving the core transformer structure.

Schematic diagrams of the small-R and large-R regression networks are shown in Figs. 6.1 and 6.2. In both cases, the models combine global jet information with constituent-level features, which are embedded and processed by a transformer encoder before being aggregated into a jet-level representation used for regression.

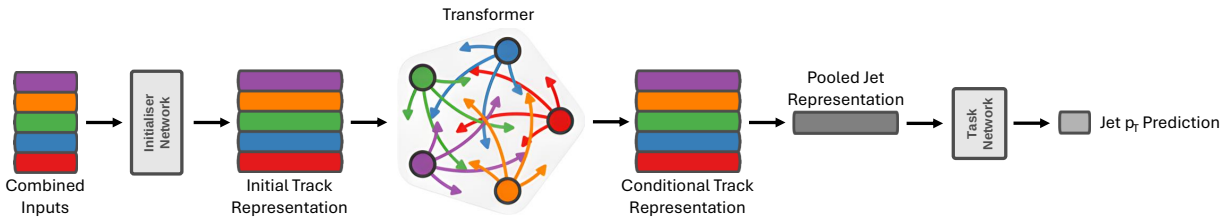


Figure 6.1: Schematic representation of the small-R jet regression network. Jet-level and constituent-level inputs are embedded by an initialiser network and processed by a transformer encoder to obtain conditional constituent representations. These are aggregated into a pooled jet-level representation, which is then used by a task network to regress the jet transverse momentum. Figure from Ref. [18].

The input features are concatenated and passed to an initialiser network. For small-R jets, these inputs include global jet kinematics and constituent-level information from tracks and charged and neutral particle-flow objects. For large-R jets, inputs from different UFO types are embedded separately using dedicated initialiser networks. Each initialiser consists

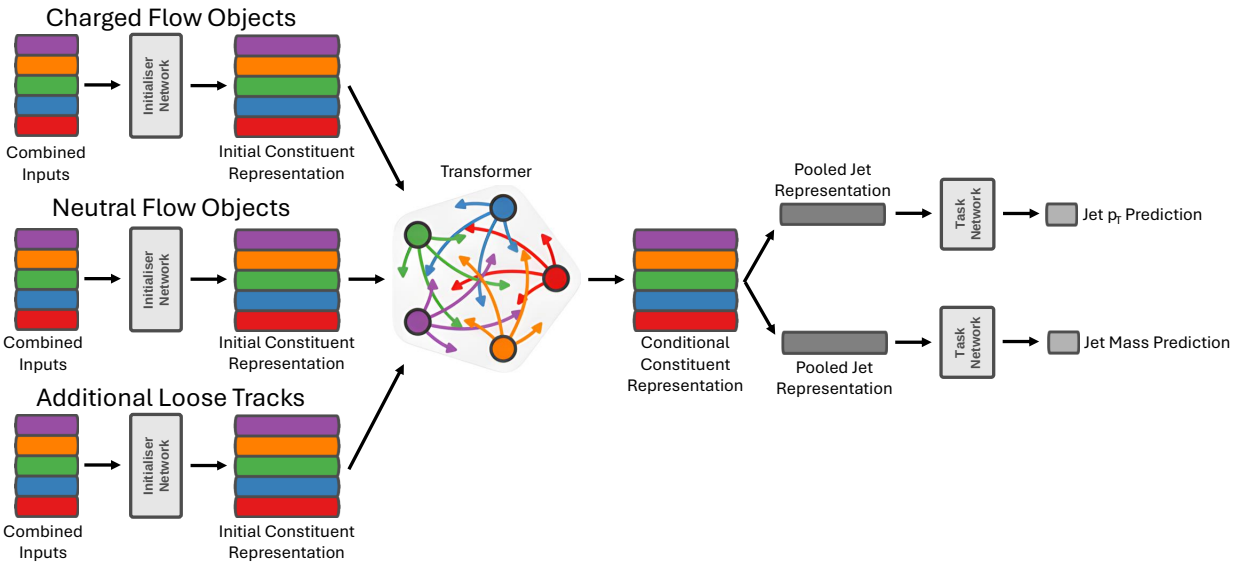


Figure 6.2: Schematic representation of the large-R jet regression network. Inputs from different object types are embedded separately and passed to a transformer encoder, producing conditional constituent representations. A pooled jet-level representation is formed and used by dedicated task networks to regress the jet transverse momentum and mass. Figure from Ref. [18].

of a fully connected layer with 256 neurons and a ReLU activation function, projecting the inputs into a 256-dimensional embedding space.

The embedded representations are processed by a transformer encoder following the architecture described in Ref. [145] and initialised with pretrained weights from the GN2 and GN2X models. To stabilise training and control the magnitude of network weights, multiple-layer normalisation [146] is applied throughout the architecture. The encoder comprises four transformer blocks with eight attention heads for small-R jets, and three blocks with two attention heads for large-R jets.

The outputs of the transformer encoder are aggregated into a single global jet representation using learned attention weights. This representation is passed to a fully connected deep neural network that performs the final regression. For small-R (large-R) jets, the regression head consists of four hidden layers with 128 (256), 128, 64, and 32 neurons, followed by a ReLU-activated output layer.

The training target is defined as the ratio of the truth to the reconstructed jet energy. For large-R jets, an additional regression task is included to predict the jet mass. The networks are trained using the mean absolute error as the loss function.

6.1.3 Regression performance in simulation

The regression performance is evaluated separately for the two models. To quantify the performance, both the jet energy response and the resolution are computed. For small-R jets, the response is defined as the ratio of the reconstructed jet p_{T} to the corresponding truth-jet p_{T}. For large-R jets, the response and resolution are evaluated for both the transverse momentum and the jet mass.

The performance metrics used in this study are the median response and the relative response resolution, defined as half the inter-quantile range between the 15.9% and 84.1% percentiles, normalised to the median $\tilde{\sigma} = \sigma/\mu$. A median response close to unity and stable across phase space, together with a smaller relative resolution, indicates improved regression performance.

To assess the regression's performance relative to the nominal calibration, the ratio of the median response from the regression model to that from the nominal calibration is computed. Improvements in the relative resolution are quantified using the root-square difference (RSD), defined as

$$\mathrm{RSD} = \mathrm{sgn}(\tilde{\sigma}' - \tilde{\sigma})\sqrt{|\tilde{\sigma}'^2 - \tilde{\sigma}^2|}\,, \tag{6.1}$$

where $\tilde{\sigma}$ denotes the relative resolution obtained with the nominal calibration and $\tilde{\sigma}'$ obtained with the regression model.

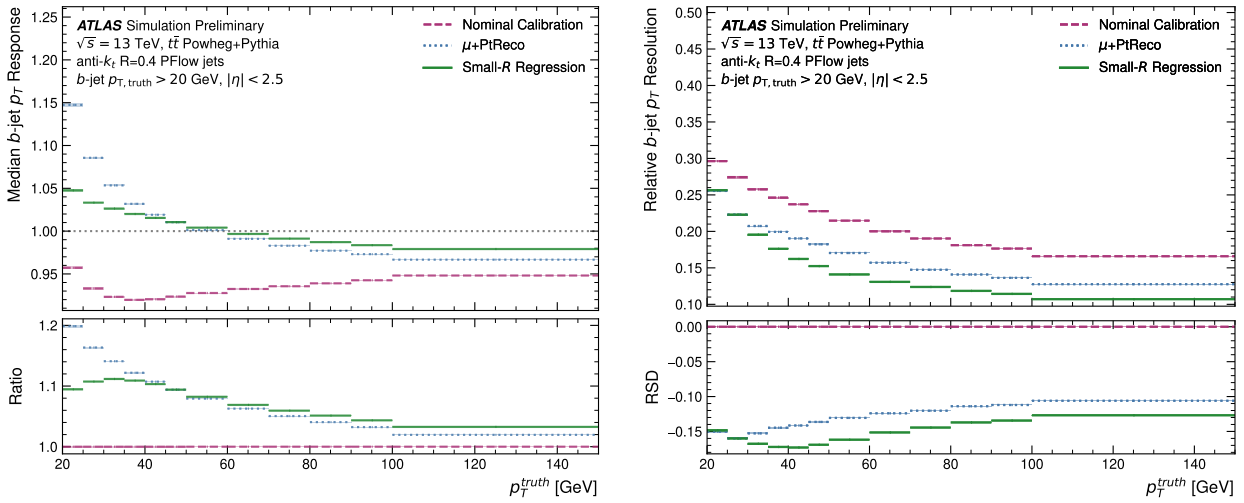
Uncertainties on the median response and the relative resolution are estimated as the standard deviation of these quantities evaluated across 100 statistically independent subsamples.

6.1.3.1 Small-R jet regression performance

For small-R jets, the regression model is compared to the nominal calibration as well as to the nominal calibration augmented with the *muon-in-jet* and *PtReco* corrections (denoted $\mu+\mathrm{PtReco}$), as described in Section 4.4.1. Jets used in the performance studies are required to be matched to truth-level b-jets with $p_{\mathrm{T}}^{\mathrm{truth}} > 20$ GeV.

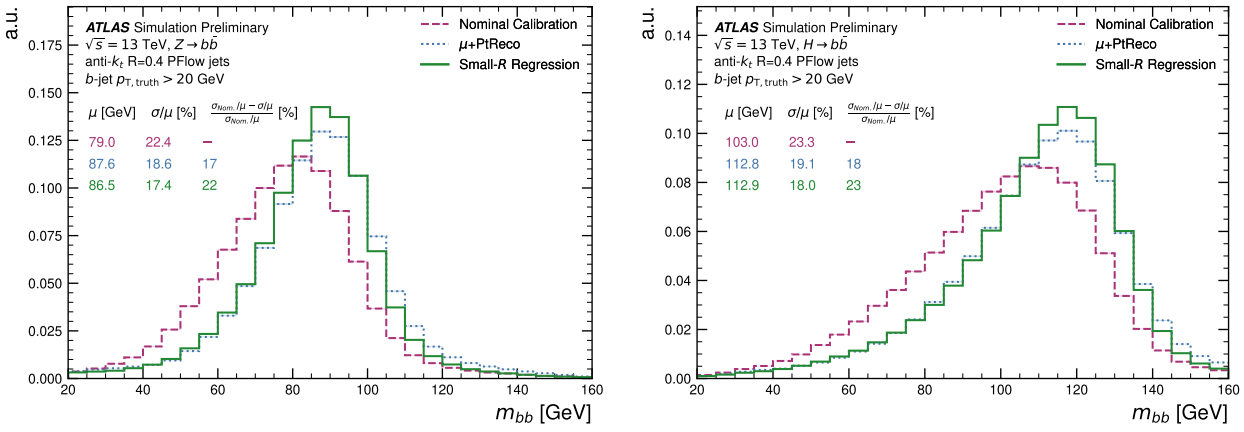
Figure 6.3 shows the median transverse momentum response of b-jets and the corresponding relative resolution for the small-R regression model. The regression yields a median response closer to unity across the full momentum spectrum, outperforming both the nominal calibration and the $\mu + \mathrm{PtReco}$ corrections (Fig. 6.3a). Similar improvements are observed in the relative resolution (Fig. 6.3b), where the regression systematically reduces the relative resolution. As expected, the resolution is poorest at low p_{T}, where detector effects dominate. In the p_{T} range of 40–100 GeV, the $\mu + \mathrm{PtReco}$ correction improves the nominal relative resolution by approximately 20%, while the regression achieves an improvement of about 30%.

To further validate the regression, the invariant masses of the Z and Higgs bosons decaying into b-quark pairs are reconstructed in Fig. 6.4. In both cases, the regression shifts the mass peak closer to the expected value while reducing mass resolution. Quantitatively, the relative mass resolution improves by 22% for the Z boson peak and 23% for the Higgs boson peak compared to the nominal calibration.



(a) Scale (b) Resolution

Figure 6.3: Performance of the small-R b-jet regression in simulation. The regression-based calibration (solid green) is compared to the nominal calibration (dashed purple) and the nominal calibration including the μ + PtReco corrections (dotted blue). The width of the coloured bands represents the statistical uncertainty. (a) Median transverse momentum response, defined as the ratio of reconstructed to truth-level jet p_{T}. The lower panel shows the ratio to the nominal calibration. (b) Relative response resolution, defined as half the inter-quantile range between the 15.9% and 84.1% percentiles, divided by the median. The lower panel shows the root-square difference (RSD) relative to the nominal calibration. Figure from Ref. [18].



(a) $Z \to b\bar{b}$ (b) $H \to b\bar{b}$

Figure 6.4: Validation of the small-R b-jet regression using reconstructed invariant mass distributions in simulation. The invariant mass of (a) the $Z \to b\bar{b}$ and (b) the $H \to b\bar{b}$ decays is shown after applying the nominal calibration (dashed purple), the nominal calibration including the μ + PtReco corrections (dotted blue), and the regression-based calibration (solid green). The width of the coloured bands represents the statistical uncertainty. The mean (μ), relative resolution (σ/μ), and relative improvement with respect to the nominal calibration are reported in the upper-left corner of each plot. Figure from Ref. [18].

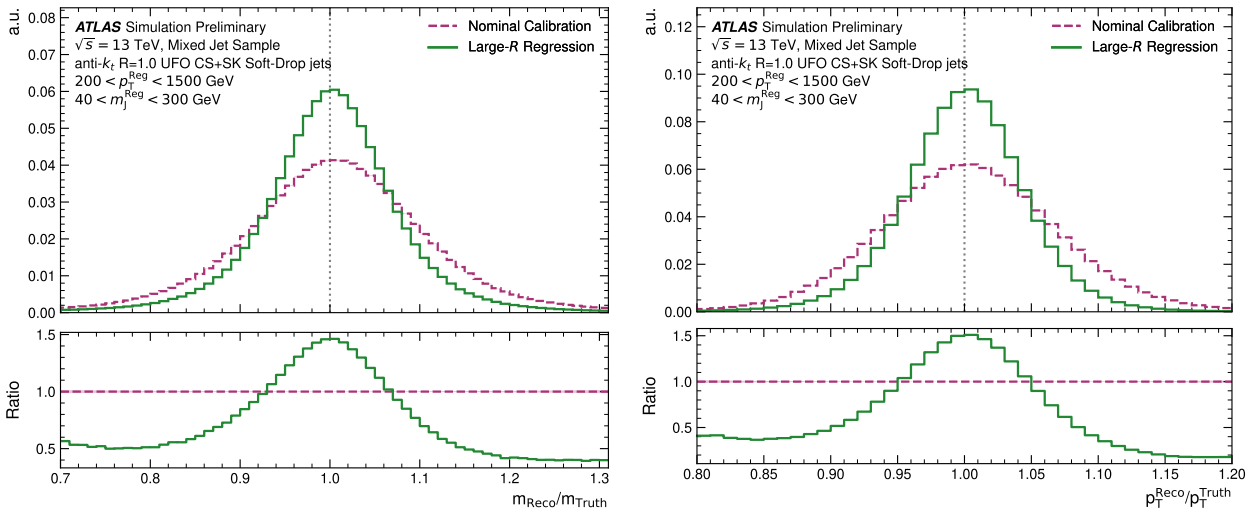
Additional studies presented in Ref. [18] show that the μ + PtReco correction relies solely on the highest-p_{T} muon. This limitation becomes significant when multiple muons originate from the same b-hadron decay chain The regression model is able to account for such scenarios.

6.1.3.2 Large-R jet regression performance

Since the primary focus of this thesis is on small-R jets, the discussion of large-R jet results is more concise. For these jets, comparisons are performed only between the regression-based calibration and the nominal calibration, as the *muon-in-jet* and *PtReco* corrections are currently applicable only to small-R jets.

Figure 6.5 shows the mass and transverse momentum response distributions integrated over the full jet p_{T} range. Differential distributions of the response and relative resolution as a function of p_{T} are provided in Ref. [18]. For both mass and p_{T}, the regression-based calibration produces response distributions that are more sharply peaked and better centred around unity. Across the p_{T} spectrum, the median response is closer to one and exhibits a flatter dependence on p_{T}. The relative resolution improves by approximately 25–35% for the transverse momentum and by 26–33% for the jet mass.

Similarly to the small-R case, the performance of the large-R jet regression is further validated using reconstructed invariant mass distributions of massive particles decaying hadronically into at least one b-quark, as shown in Fig. 6.6. For the Z and Higgs bosons, the



(a) Mass response

(b) p_{T} response

Figure 6.5: Response distributions for the large-R b-jet regression in simulation, shown for (a) the jet mass and (b) the transverse momentum p_{T}. The regression-based calibration (solid green) is compared to the nominal calibration (dashed purple). The lower panels show the ratio of the regression-based calibration to the nominal calibration. Figure from Ref. [18].

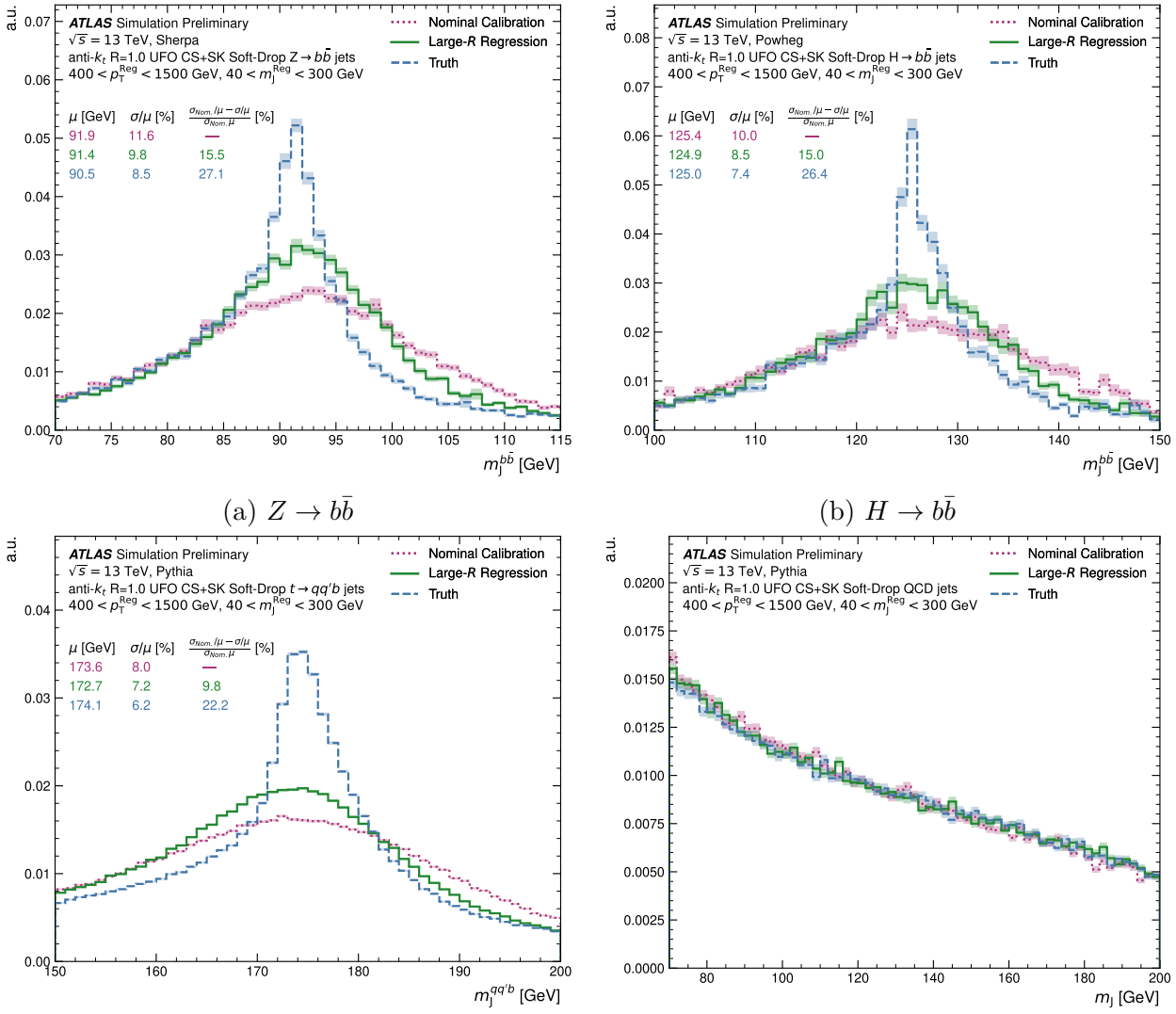
reconstructed mass peaks are closer to their true values after applying the regression-based calibration. For all three resonances, including the top quark, the relative resolution improves by approximately 10%. The QCD multijet sample confirms that the regression does not introduce artificial mass peaks.

6.2 In-situ calibration and validation in data for small-R b-jets

Building on the strong performance of the regression models presented in the previous section, this section aims to study the behaviour of regressed small-R b-jet momentum in real data and to compare it with simulations using in-situ techniques. However, the small-R model is not yet implemented in the complete ATLAS calibration framework, and therefore, its in-situ validation in data cannot be performed at this stage. As a result, the studies presented here focus on validating the methodology and the analysis workflow in the nominal calibration scenario, ensuring that the framework is ready for future application to the regression-based calibration.

For clarity, the small-R regression model is referred to as *bJR4*, while the large-R version is denoted *bJR10*.

This section includes a description of the data and simulation samples, the event and jet selection criteria, the methodology used to extract the in-situ b-jet energy scale (b-JES) and, in particular, the b-jet energy resolution (b-JER). It also includes preliminary results validat-



(a) $Z \to b\bar{b}$ (b) $H \to b\bar{b}$

(c) $t \to qq'b$ (d) QCD jets

Figure 6.6: Validation of the large-R b-jet regression using reconstructed invariant mass distributions in simulation. The invariant mass of (a) $Z \to b\bar{b}$, (b) $H \to b\bar{b}$, (c) $t \to qq'b$, and (d) QCD jets is shown for truth-level b-jets (dashed blue), after applying the nominal calibration (dotted purple), and the regression-based calibration (solid green). The width of the coloured bands represents the statistical uncertainty. The mean (μ), relative resolution (σ/μ), and relative improvement with respect to the nominal calibration are reported in the upper-left corner of the first three plots. Figure from Ref. [18].

ing the calibration in the nominal configuration for the JER. Where applicable, techniques from the hybrid ML study are used.

6.2.1 Samples and event selection

For the in-situ calibration of small-R jets, events with a V+jets topology are used, where V denotes either a Z boson or a photon, serving as the *reference* object. Different reference objects provide complementary statistical coverage across the p_{T} spectrum. Both real data and simulated pp collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV, corresponding to ATLAS Run 2 conditions, are considered.

The MC samples used for this study are summarised in Table 6.2, including the event generators, tunes, and PDF sets. For each process, multiple MC generators (Pythia 8, Sherpa, and Herwig 7 or FxFx) are used to evaluate the impact of generator modelling on the calibration. Observed differences between generators are treated as an additional source of systematic uncertainty.

For small-R b-jets, the quantities of interest are the JES and the JER. Both corrections are derived using V+jets events, rather than the inclusive dijet topologies employed in the standard JER derivation.

Process	**Event generator and tune**	**PDF set**
$Z(\ell\ell)$+jets	Powheg [91–93]+ Pythia 8.186 [95] with AZNLO [126]	CTEQ6L1 [127]
	Sherpa 2.2.11 [90]	NNPDF3.1nnlo [137]
	MadGraph5_aMC [89]+FxFx [139] & Pythia 8.245 with A14 [138]	NNPDF2.3lo [142]
γ+jets	Pythia 8.244 with A14	NNPDF2.3lo
	Sherpa 2.2.2	NNPDF3.0nnlo
	Herwig 7.2.1 [147]	MMHT2014lo [148]

Table 6.2: Details of the Monte Carlo samples used for the in-situ studies of small-R b-jets. In the table, $l = e, \mu$.

The events are selected using single- and di-lepton triggers for the Z+jets samples and single-photon triggers for the γ+jets samples. Simulated events are passed through a full Geant4-based simulation of the ATLAS detector and reconstructed with the same software as real data. The distribution of the average number of interactions per bunch crossing is reweighted in simulation to match the data.

Z+jets events are required to contain exactly two electrons or two muons with $p_{\mathrm{T}} > 20$ GeV. Electron candidates must satisfy $|\eta| < 2.37$, excluding the calorimeter transition region $1.37 < |\eta| < 1.52$, while muon candidates are required to satisfy $|\eta| < 2.5$. Both leptons must pass the *Medium* identification working point and standard isolation requirements, and

GN2 Working Point	light-quark	gluon	c-quark	b-quark
Inclusive	56.40%	27.18%	6.69%	3.97%
90%	34.43%	18.95%	37.94%	7.77%
77%+85%	19.38%	14.80%	39.69%	24.08%
65%+70%	2.00%	4.56%	4.10%	89.21%

Table 6.3: Flavour composition of jets in the $Z(\mu\mu)$ MC sample with $17 < p_{\mathrm{T}} < 1000$ GeV, generated with Pythia 8 for the full Run 2 conditions.

GN2 Working Point	light-quark	gluon	c-quark	b-quark
Inclusive	68.53%	17.83%	11.76%	1.86%
90%	39.86%	12.71%	44.64%	2.79%
77%+85%	32.43%	13.08%	44.35%	10.13%
65%+70%	10.26%	17.42%	8.55%	63.77%

Table 6.4: Flavour composition of jets in the γ+jets MC sample with $150 < p_{\mathrm{T}} < 1000$ GeV, generated with Pythia 8 for the full Run 2 conditions.

must have opposite electric charge. The invariant mass of the dilepton system is required to be consistent with the Z boson mass, $66 < m_{\ell\ell} < 116$ GeV.

γ+jets events are required to contain a photon with $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.37$, excluding the transition region. Photon candidates must satisfy the *Tight* identification and standard isolation requirements.

Small-R jets are reconstructed using the anti-k_t algorithm with $R = 0.4$. Selected events are required to contain a leading jet (j_1) with $p_{\mathrm{T}} > 10$ GeV and $|\eta_{\mathrm{det}}| < 0.8$, which is required to be back-to-back with the reference object: $\Delta\phi(Z, j_1) > 2.6$ or $\Delta\phi(\gamma, j_1) > 2.8$. To suppress additional hard radiation, subleading jets (j_2) are required to satisfy $p_{\mathrm{T}}^{\mathrm{j}_2} < \max(12\ \mathrm{GeV}, 0.1 \cdot p_{\mathrm{T}}^{\mathrm{ref}})$ for Z+jets events, and $p_{\mathrm{T}}^{\mathrm{j}_2} < \max(15\ \mathrm{GeV}, 0.1 \cdot p_{\mathrm{T}}^{\mathrm{ref}})$ for γ+jets events. Jets with $20 < p_{\mathrm{T}} < 60$ GeV are required to pass the Jet Vertex Tagger (JVT) selection to suppress contamination from pile-up interactions.

Jets are b-tagged using the GN2 algorithm [113], applied to jets with $p_{\mathrm{T}} > 17$ GeV. To avoid biases when using b-tagging, jets are additionally required to have truth-level $p_{\mathrm{T}}^{\mathrm{truth}} > 17$ GeV. Different b-tagging working points (WPs) are considered to study the calibration performance based on flavour. Some WPs are combined to improve the statistical precision. The resulting flavour compositions for Z+jets and γ+jets events are shown in Tables 6.3 and 6.4. The inclusive WP is dominated by light-flavour jets, the intermediate WPs contain roughly 40% c-jets, and the 65%+70% WP achieves the highest b-jet purity, reaching approximately 90% in the Z+jets sample.

6.2.2 In-situ calibration methodology

After the jet momentum has been calibrated in bins of p_{T} and η using MC samples, residual differences between simulation and real data may remain. To account for these effects, an additional in-situ calibration is applied to align the jet momentum scale in simulation with that observed in data.

The methodology used to derive the in-situ JES follows closely the approach described in Section 4.3.2.1, and is therefore only briefly summarised here for completeness.

For the JER, this analysis adopts a hybrid strategy that combines techniques of the standard in-situ method described in Section 4.3.2.2 with the samples and elements introduced in Section 5.5.3. Since the JER determination constitutes the novel contribution of this work, the methodology is presented in more detail in the following sections.

Jet Energy Scale

This analysis determines the in-situ JES using events in which a well-measured reference object recoils against a leading jet [121]. The procedure evaluates the jet response in both simulation and data using the same observable and derives the in-situ JES correction from the ratio of the two responses.

The jet momentum response is computed in bins of jet transverse momentum and pseudorapidity as

$$\mathcal{R}_{\mathrm{DB}} = p_{\mathrm{T}}^{\mathrm{j,reco}} / p_{\mathrm{T}}^{\mathrm{ref}} \,, \tag{6.2}$$

where $p_{\mathrm{T}}^{\mathrm{ref}} = p_{\mathrm{T}}^{Z/\gamma} |\cos\Delta\phi|$, and $\Delta\phi$ is the azimuthal separation between the reconstructed jet and the reference object.

The mean response is extracted by fitting the response distribution with a Gaussian function. To reduce the impact of non-Gaussian tails, the fit is restricted to the central region of the peak. In the low-p_{T} regime ($p_{\mathrm{T}}^{\mathrm{ref}} < 30$ GeV), which is dominated by Z+jets events, a modified Poisson function is used instead to better describe the asymmetric shape of the response distribution.

In this low-p_{T} region, both Pythia 8 simulation and data exhibit a double-peak structure in the response distribution. The additional peak is attributed to fake jets, which predominantly originate from pile-up interactions. To mitigate this effect, the fake-jet contribution is modelled using simulation and subsequently subtracted from the data.

The in-situ JES correction factor is then computed as the ratio of the jet response measured in data to that obtained from simulation,

$$c_{\mathrm{JES}} = \frac{\mathcal{R}_{\mathrm{DB}}^{\mathrm{data}}}{\mathcal{R}_{\mathrm{DB}}^{\mathrm{MC}}} \,. \tag{6.3}$$

This correction is derived for the inclusive jet sample as well as for jets tagged using the 90% WP, the combined 85%+77% WPs, and the combined 70%+65% WPs.

To isolate potential flavour-dependent effects, a double ratio is constructed between the b-tagged and inclusive calibrations, giving the following calibration factor,

$$\tilde{\mathcal{R}}_{b\text{JER}} = \frac{\mathcal{R}^{\text{data}}_{b\text{-tagged}}/\mathcal{R}^{\text{MC}}_{b\text{-tagged}}}{\mathcal{R}^{\text{data}}_{\text{inclusive}}/\mathcal{R}^{\text{MC}}_{\text{inclusive}}}\,. \tag{6.4}$$

Compatibility of this double ratio with unity indicates that the inclusive in-situ calibration already accounts for response differences between inclusive and b-jets, and that no additional flavour-specific calibration is required.

Both Z+jets and γ+jets samples are used to determine the JES over a wide range of jet transverse momentum, exploiting their complementary statistical reach. The individual in-situ measurements are subsequently combined into a single JES calibration using a χ^2-based procedure.

In the standard in-situ JES calibration, the $Z(\to ee)$+jets, $Z(\to \mu\mu)$+jets, γ+jets, and multijet balance methods are combined to provide coverage over a wide jet p_{T} range. In the context of the small-R b-jet calibration, the $Z(\to ee)$+jets and $Z(\to \mu\mu)$+jets samples are instead combined into a single Z+jets measurement to improve statistical precision. The multijet balance method is not included, as the calibration targets the jet momentum range $17 < p_{\text{T}} < 1000$ GeV. Furthermore, the reduced event yields resulting from the application of b-tagging requirements motivate the use of larger p_{T} bins than in the standard configuration.

The final JES calibration is obtained by combining the individual in-situ measurements from the different reference channels as a function of jet transverse momentum [149]. The combination employs a χ^2 minimisation that accounts for both statistical and systematic uncertainties of each input measurement. The combination effectively performs a weighted average in each jet p_{T} bin, with the relative contribution based on the precision of each input.

Several MC generators are used to assess the calibration's sensitivity to the modelling of the hard process and the parton shower. Observed differences between generators are treated as an additional source of systematic uncertainty. Unless stated otherwise, the Pythia 8 sample is used as the nominal reference.

Jet Energy Resolution

The JER is determined using the same V+jets samples employed for the JES extraction. Unlike the JES, however, the detector resolution cannot be inferred directly from the width of the reconstructed response distribution, and a more elaborate procedure is required.

In Section 5.5.3, we obtained the JER by subtracting the label resolution from the observed resolution in quadrature. This approach assumes that both response distributions are well described by Gaussian distributions. In the present case, the label response exhibits significant non-Gaussian features, making the quadratic subtraction an invalid approximation.

We therefore determine the JER using the convolution-based methodology described in Section 4.3.2.2 for dijet events. In this framework, the detector response is defined as

$$\mathcal{R}_{\text{det}} \equiv \mathcal{R}_{\text{obs}} \ominus \mathcal{R}_{\text{label}}\,, \tag{6.5}$$

where the symbol $\ominus$ denotes a deconvolution operation. The response quantities are defined as $\mathcal{R}_{\text{label}} = p_{\text{T}}^{\text{j,truth}}/p_{\text{T}}^{\text{ref}}$, $\mathcal{R}_{\text{reco}} = p_{\text{T}}^{\text{j,reco}}/p_{\text{T}}^{\text{ref}}$, and $\mathcal{R}_{\text{det}} = p_{\text{T}}^{\text{j,truth}}/p_{\text{T}}^{\text{j,reco}}$. The JER corresponds to the width of the detector response that, when convolved with the label response, reproduces the observed response distribution.

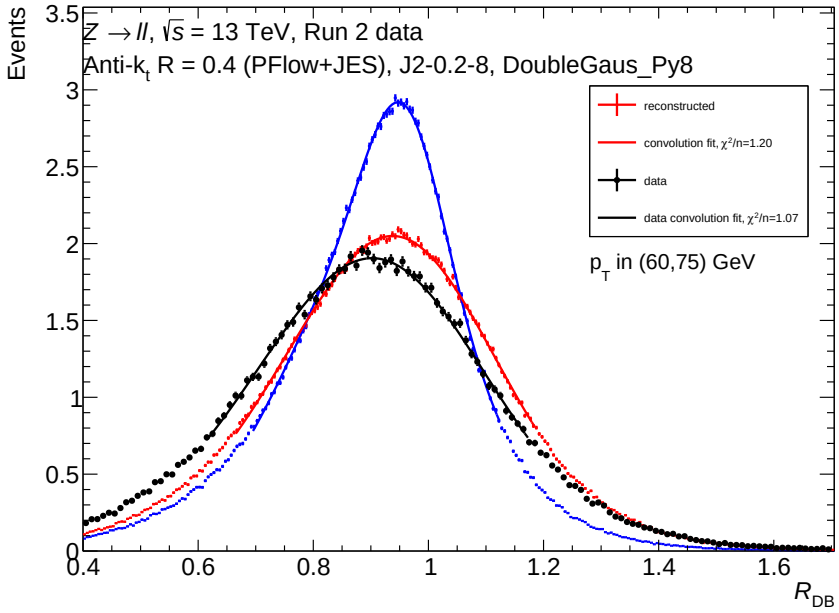


Figure 6.7: Example of the convolution-based JER extraction for the p_{T} bin $60 < p_{\text{T}} < 75$ GeV in the Z+jets samples. The label response in simulation (blue) is fitted in the central region with a double-Gaussian function. This fitted response is then convolved with a Gaussian function and fitted to the reconstructed jet response in simulation (red) and in data (black). The legend reports the corresponding reduced χ^2 values of the fits.

The procedure is illustrated in Fig. 6.7. First, we model the label response (in blue) in simulation by fitting the central region of the $\mathcal{R}_{\text{label}}$ distribution with a double-Gaussian function[2]. We then convolve this fitted label response with a Gaussian resolution function and fit the result to the reconstructed jet response, $\mathcal{R}_{\text{obs}}$ (in red for MC and in black in data). The width of the Gaussian resolution function extracted from the fit represents the detector contribution, the jet energy resolution.

To quantify the difference between data and simulation, we compute the in-situ JER correction using the root-squared difference (RSD). Throughout this section, the symbol $\tilde{\sigma}$ denotes the relative width $\sigma/\langle\mathcal{R}\rangle$, which factorises calibration effects on the response mean from those on the resolution. The RSD is defined as

$$\text{RSD} = \sqrt{|\sigma^2_{\text{data}} - \sigma^2_{\text{MC}}|}\,. \tag{6.6}$$

[2]Alternative functions, including Double-Sided Crystal Ball, Bukin, and exponential-plus-Gaussian functions, were tested and found to provide inferior agreement.

A reference MC generator is also chosen both to define the nominal label response and to serve as the baseline for the data–MC comparison, while alternative generators are used to estimate modelling uncertainties. For consistency with the JES determination, we use the Pythia 8 sample as the nominal reference. In the low-p_{T} region, we model and subtract the fake-jet contribution identified in this sample before extracting the resolution.

The JER determination is repeated for b-tagged jets using the GN2 working points introduced earlier. To probe potential flavour-dependent effects, we construct a double ratio between the RSD obtained for b-tagged jets and that for the inclusive jet sample,

$$\tilde{\mathcal{R}}_{b\mathrm{JER}} = \frac{\sqrt{|\sigma^2_{\mathrm{data}} - \sigma^2_{\mathrm{MC}}|_{\mathrm{bTag}}}}{\sqrt{|\sigma^2_{\mathrm{data}} - \sigma^2_{\mathrm{MC}}|_{\mathrm{Inclusive}}}} \,. \tag{6.7}$$

Compatibility of this ratio with unity indicates that the inclusive in-situ calibration adequately describes the resolution of b-jets, while deviations would point to the need for an additional flavour-dependent correction.

As for the JES, the individual JER measurements obtained from the different reference channels can be combined using a χ^2-based procedure to yield a single in-situ resolution calibration, while fitting Eq. 4.5. If compatibility with the inclusive calibration is achieved, the nominal JER retrieved using the dijet asymmetry method will be directly used.

6.2.3 Results

This section presents the results of the in-situ JER analysis developed as described in the previous sections. Although the same framework can be used to derive the corresponding JES corrections, only the JER results are discussed here, as they constitute the primary original contribution of this work.

The results are presented separately for Z+jets and γ+jets events, which probe complementary jet transverse momentum (p_{T}) regimes and partially overlapping kinematic ranges.

Figure 6.8 shows the in-situ JER measured in Run 2 Z+jets events, expressed as the relative resolution defined in Eq. 6.5, as a function of jet transverse momentum for different b-tagging WPs. The results obtained with the nominal Pythia 8 simulation are shown as red dots and compared with data as black dots[3]. Predictions from two additional generators, Sherpa (green squares) and FxFx (blue triangles), are also included to illustrate the measurement's sensitivity to the modelling of the hard process and the parton shower. The lower panels display the RSD (Eq. 6.6) of each generator with respect to data.

The inclusive results, shown in Fig. 6.8a, exhibit good stability and relatively small statistical uncertainties across the full probed p_{T} range. As expected, the JER decreases gradually with increasing jet transverse momentum. The largest differences between data and simulation are observed in the lowest p_{T} bins, where detector effects and modelling uncertainties are more pronounced.

The corresponding results for b-tagged jets are shown in Figs. 6.8b, 6.8c, and 6.8d, for the 90%, 85%+77%, and 70%+65% WPs, respectively. Applying b-tagging requirements

[3] As a reminder, data JER is computed by convolving the $\mathcal{R}_{\mathrm{obs}}$ from Pythia 8 simulation.

significantly reduces event yields, leading to larger statistical uncertainties across all p_{T} bins. A local maximum in the JER is observed in the $p_{\mathrm{T}} = (35, 45)$ GeV bin, which is already less stable in the inclusive measurement. At higher p_{T}, the JER decreases similarly to the inclusive case. At low p_{T}, however, the extracted JER values are significantly reduced, an effect that is likely driven by edge effects associated with the minimum jet transverse momentum requirement.

Figure 6.9 presents the double ratio of the JER measured for b-tagged jets to that obtained for inclusive jets in the Pythia 8 sample where the first bin has been removed. This observable provides a direct test of potential flavour-dependent differences in the jet energy resolution.

A linear fit to the distribution is found to be consistent with unity. The reduced χ^2 indicates good agreement with the chosen fit function for all three working point configurations. Within the current statistical uncertainties, the fitted values deviate by less than two standard deviations from unity, indicating that the inclusive in-situ JER calibration provides an adequate description of the resolution of b-tagged jets in the phase space probed by this analysis.

Similar results are observed in the γ+jets analysis. Figure 6.10 shows the JER and the relative standard deviation (RSD) measured in data and compared to three MC generators: Pythia 8, Sherpa, and Herwig 7. The JER decreases with increasing jet transverse momentum, stabilising at about 0.7, even when a b-tagging requirement is applied. At low p_{T}, although the resolution is larger, the overall trend is smoother and exhibits a more regular behaviour than in the Z+jets case.

Figure 6.11 displays the corresponding double ratio for γ+jets events in the Pythia 8 sample. The distribution remains compatible with a constant behaviour across the probed kinematic range. A linear fit yields values consistent with unity, and the reduced χ^2 confirms a satisfactory description of the data for all three working point configurations.

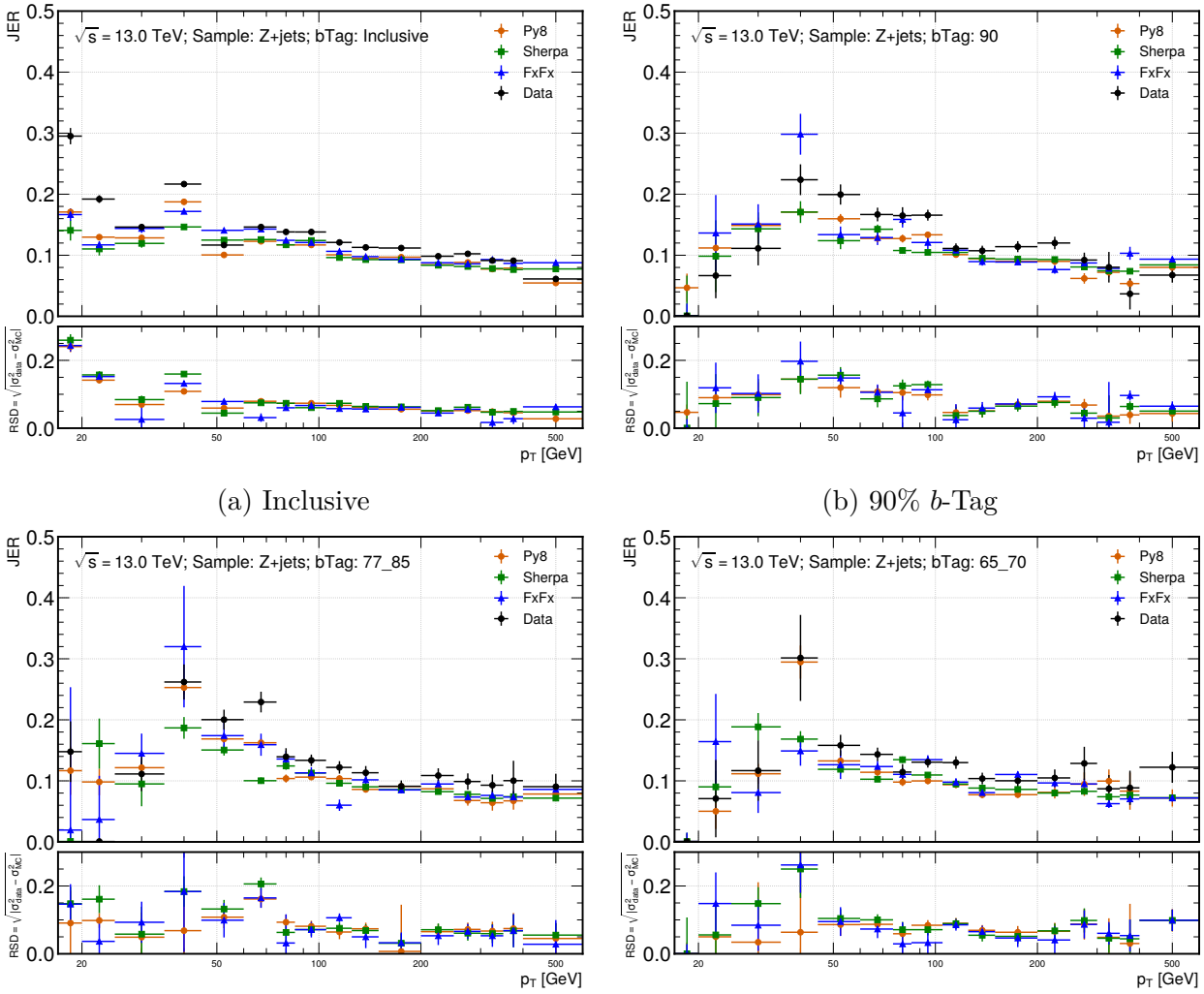
The strongest agreement is observed for the 65% and 70% working points, where the fitted values differ from unity by about one standard deviation. In the remaining configurations, the deviations do not exceed two standard deviations.

For completeness, figures showing $\tilde{\mathcal{R}}_{b\mathrm{JER}}$ for the other generators are provided in Appendix B.

6.3 Conclusions and outlook

These preliminary results demonstrate that the nominal in-situ calibration can be consistently applied to b-tagged jets. The observed fluctuations of the double ratio around unity indicate no significant deviation from the inclusive calibration. However, the achieved statistical precision does not allow a determination of the jet energy resolution at the per-mille level, with the uncertainty being dominated by the available data statistics.

Further improvements are therefore limited by statistical power. Ongoing studies focus on optimising the fitting strategy, although this is intrinsically challenging, as it requires a careful balance between sufficiently large p_{T} bins to ensure adequate statistical precision and



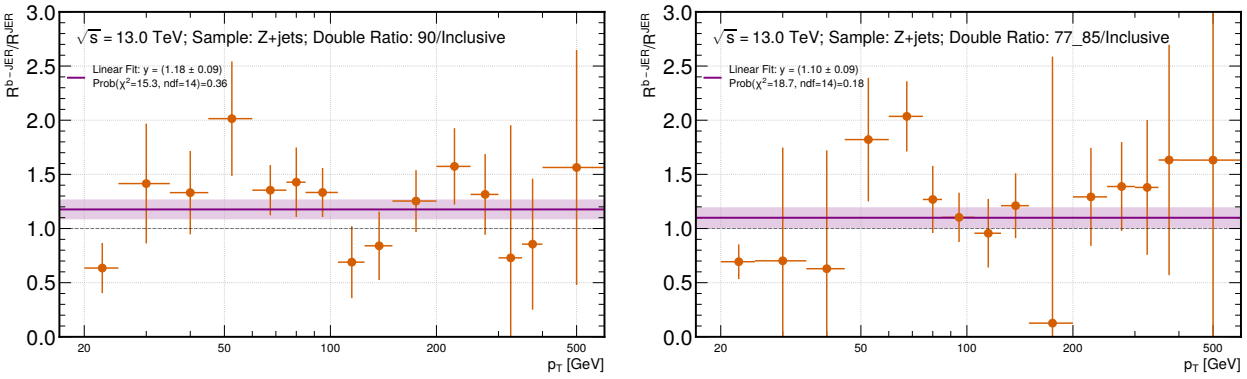
(a) Inclusive

(b) 90% b-Tag

(c) 85%+77% b-Tag

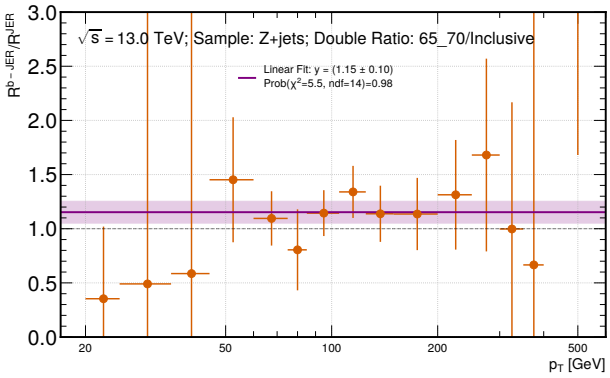
(d) 70%+65% b-Tag

Figure 6.8: In-situ jet energy resolution (JER) as a function of jet transverse momentum in Run 2 Z+jets events for (a) the inclusive jet sample and for the b-tagging working points (b) 90%, (c) 85%+77%, and (d) 70%+65%. The results obtained from data are shown in black and are compared to predictions from simulation using Pythia 8 (orange dots), Sherpa (green squares), and FxFx (blue triangles). The lower panels show the relative standard deviation (RSD) of each simulation with respect to the data. Only statistical uncertainties are shown.



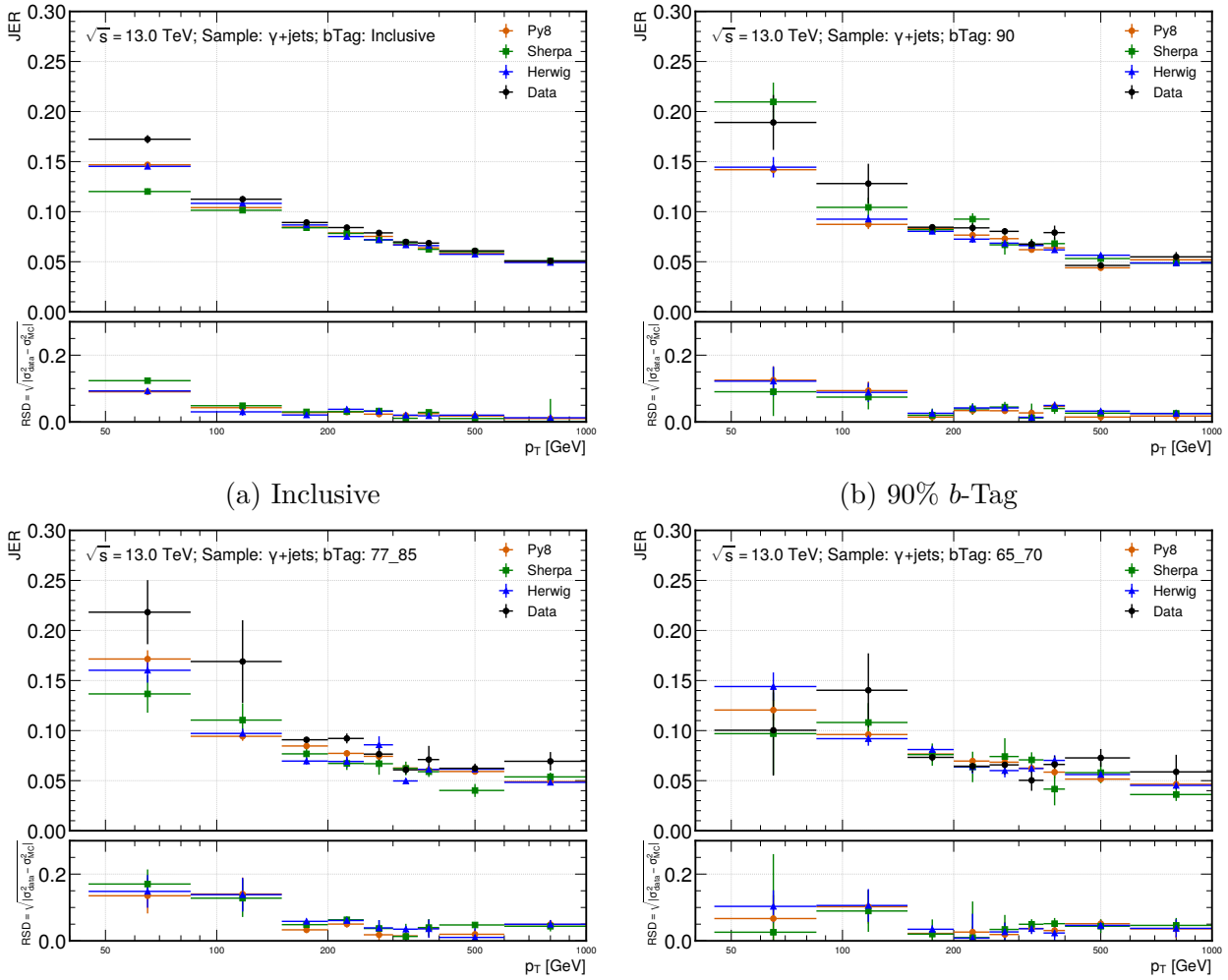
(a) 90% b-Tag

(b) 85%+77% b-Tag



(c) 70%+65% b-Tag

Figure 6.9: Ratio of the in-situ jet energy resolution (JER) relative standard deviation between b-tagged and inclusive jets as a function of jet transverse momentum in Run 2 Z+jets events, shown for the b-tagging working points (a) 90%, (b) 85%+77%, and (c) 70%+65%. The ratio is constructed from measurements in data and the corresponding predictions from simulation using Pythia 8; statistical uncertainties are shown. A linear fit is overlaid in purple, with the corresponding 1σ uncertainty band displayed as a shaded region.



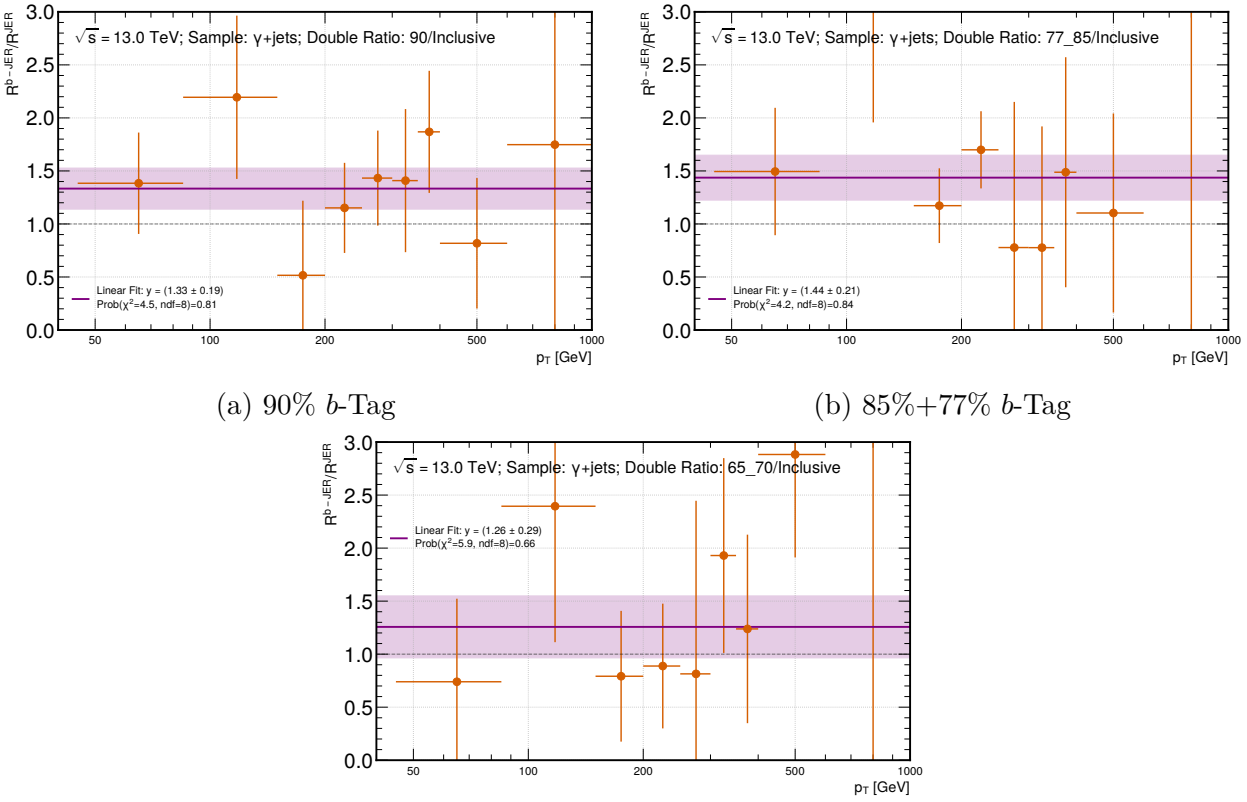
(a) Inclusive

(b) 90% b-Tag

(c) 85%+77% b-Tag

(d) 70%+65% b-Tag

Figure 6.10: In-situ jet energy resolution (JER) as a function of jet transverse momentum in Run 2 γ+jets events for (a) the inclusive jet sample and for the b-tagging working points (b) 90%, (c) 85%+77%, and (d) 70%+65%. The results obtained from data are shown in black and are compared to predictions from simulation using Pythia 8 (orange dots), Sherpa (green squares), and Herwig 7 (blue triangles). The lower panels show the relative standard deviation (RSD) of each simulation with respect to the data. Only statistical uncertainties are shown.



(a) 90% b-Tag

(b) 85%+77% b-Tag

(c) 70%+65% b-Tag

Figure 6.11: Ratio of the in-situ jet energy resolution (JER) relative standard deviation between b-tagged and inclusive jets as a function of jet transverse momentum in Run 2 γ+jets events, shown for the b-tagging working points (a) 90%, (b) 85%+77%, and (c) 70%+65%. The ratio is constructed from measurements in data and the corresponding predictions from simulation using Pythia 8; statistical uncertainties are shown. A linear fit is overlaid in purple, with the corresponding 1σ uncertainty band displayed as a shaded region.

sufficiently fine binning to capture detector-resolution effects.

Future work will reanalyse these samples using jets calibrated with the bJR4 results in Run 2. In parallel, these studies are being extended to Run 3 data to provide updated calibrations for use in upcoming analyses. If a similar level of agreement is observed, a full b-jet calibration strategy will be enabled by applying the transformer-based corrections in simulation and subsequently correcting the data using the nominal in-situ calibration. Nevertheless, even in the case where the double ratio remains consistent with unity, the limited precision motivates the assignment of an additional systematic uncertainty. A conservative uncertainty of 5% is therefore foreseen to account for the observed fluctuations with respect to the nominal calibration.

Part III

Analysis: Improving Z selection for $H \to Z\gamma$

Chapter 7

Nominal $H \to Z\gamma$ analysis of partial Run 3

7.1 Introduction

Since the discovery of the Higgs boson in 2012 [10, 11], a major effort has been devoted to precisely measuring its properties, including its mass, width, spin and couplings. Because the Standard Model is known to be incomplete, any deviation between theoretical predictions and experimental measurements could hint at new physics beyond the Standard Model (BSM).

The decay of the Higgs boson into a Z boson and a photon ($H \to Z\gamma$) has a very small branching ratio, as it occurs only via loop-induced diagrams rather than at tree level. Measuring this decay is an important step in completing the study of Higgs boson interactions with electroweak gauge bosons, providing further insight into the mechanism of electroweak symmetry breaking. Figure 7.1 shows the SM Feynman diagrams for the $H \to Z\gamma$ decay. The predicted branching ratio of this decay in the SM is $Br(H \to Z\gamma) = (1.54^{+0.10}_{-0.11}) \times 10^{-3}$ for $m_H = 125.09$ GeV.

This decay is particularly sensitive to contributions from BSM physics since new particles running in the loop could alter the branching ratio relative to the SM prediction. These new particles could be colourless charged scalars, fermions, or vector bosons [150–156], or modifications arising from composite Higgs scenarios [157–159]. Therefore, any significant deviation from the SM expectation could provide sensitivity to BSM physics.

The $H \to Z\gamma$ decay was studied by the ATLAS and CMS collaborations using the full Run 2 pp dataset, corresponding to an integrated luminosity of 140 fb^{-1} at $\sqrt{s} = 13$ TeV [160, 161]. In this analysis, the Z boson is reconstructed in its decays to charged lepton pairs, $Z \to \ell^+\ell^-$, where $\ell = e, \mu$. ATLAS reported an observed significance of 2.2σ with a signal strength $\mu = 2.0^{+1.0}_{-0.9}$, while CMS observed a significance of 2.7σ with $\mu = 2.4 \pm 0.9$. The combination of both searches [19] gave a significance of 3.4σ, representing the first evidence for the $H \to Z\gamma$ decay, though not sufficient to claim a discovery.

More recently, ATLAS has reported new results using early Run 3 pp collision data corresponding to an integrated luminosity of 165 fb^{-1} at $\sqrt{s} = 13.6$ TeV [20]. This chapter

presents the methodology and results of this analysis, providing additional insight into procedures that are often taken for granted. It serves as a foundation for understanding and motivating the novel studies introduced in the following chapter.

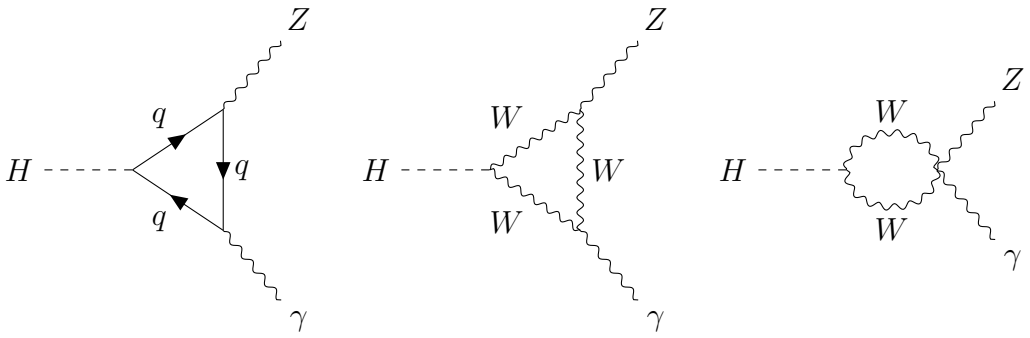


Figure 7.1: Feynman diagram contributions to the Standard Model $H \to Z\gamma$.

7.2 Data and simulation samples

This analysis uses pp collision data collected by the ATLAS detector during Run 3 of the LHC, between 2022 and 2024, with a centre-of-mass energy of $\sqrt{s} = 13.6$ TeV.

Events are selected using single-lepton and dilepton triggers. Single-electron (single-muon) triggers require a transverse momentum $p_{\mathrm{T}} > 26$ GeV ($p_{\mathrm{T}} > 24$ GeV). Dilepton triggers include symmetric thresholds of 17-17 GeV for electrons and 14-14 GeV for muons, as well as an asymmetric dimuon trigger with thresholds of 22 and 8 GeV. The average number of interactions per bunch crossing rises from approximately 42 in 2022 to about 58 in 2024.

A key aspect of this analysis is the optimisation of the signal selection and the accurate modelling of the background processes. Monte Carlo (MC) simulated samples are processed through the full detector simulation based on the GEANT4 framework [96], and event reconstruction chain, ensuring a consistent treatment with collision data.

Signal MC samples include the main Higgs boson production modes with the decay $H \to Z\gamma$, where the Z boson decays leptonically. In simulation, the Higgs boson is generated with a mass of $m_H = 125$ GeV and a decay width of $\Gamma_H = 4.1$ MeV. The samples are subsequently normalised to the corresponding production cross-sections and to the more precise mass value of $m_H = 125.09$ GeV. Table 7.1 summarises the signal MC samples for each Higgs boson production mode, including the event generator and the PDF set used. All signal events are generated using POWHEG [91–93], while the Higgs decay, parton showering, hadronisation, and pile-up are simulated with PYTHIA 8.310 [95] (A14 tune [138], PDF4LHC21 [162]).

Backgrounds are divided into several categories and are listed in Table 7.2. They include non-resonant $Z\gamma$ production, Z+jets, electroweak $Z\gamma jj$, diboson processes, and $t\bar{t}$ events.

The dominant background contribution arises from non-resonant $Z\gamma$ events, where the photon originates from initial- or final-state radiation (ISR/FSR). A significant contribution also arises from Z+jets events, in which a jet is misidentified as a photon. The EW $Z\gamma jj$ background becomes relevant in the *VBF* production mode. Diboson backgrounds primarily affect the associated VH production channel, whereas $t\bar{t}$ events contribute only marginally.

Corrections provided by the ATLAS Combined Performance groups are applied to account for differences between data and simulation in trigger, reconstruction, identification, and isolation efficiencies of reconstructed objects.

Process	**Event generator and tune**	**PDF set**
ggF	Powheg [91–93] & Pythia 8 [95] with A14 [138]	NNPDF2.3lo [142]
VBF	Powheg & Pythia 8 with A14	NNPDF2.3lo
$q\bar{q} \to ZH$	Powheg & Pythia 8 with A14	NNPDF2.3lo
$gg \to ZH$	Powheg & Pythia 8 with A14	NNPDF2.3lo
WH	Powheg & Pythia 8 with A14	NNPDF2.3lo
$t\bar{t}H$	Powheg & Pythia 8 with A14	NNPDF2.3lo
$b\bar{b}H$	Powheg & Pythia 8 with A14	NNPDF2.3lo

Table 7.1: Signal MC samples used in this analysis for Higgs boson decays into a Z boson and a photon ($H \to Z\gamma$), including the production mode, event generator, tune, and PDF set.

Process	**Event generator and tune**	**PDF set**
$Z \to ee\gamma$	Sherpa 2.2.14 [90]	NNPDF3.0nnlo [137]
$Z \to \mu\mu\gamma$	Sherpa 2.2.14	NNPDF3.0nnlo
$Z(\to ee)$ + jets	Powheg [91–93] & Pythia 8 [95] with AZNLO [126]	CTEQ6L1 [127]
$Z(\to \mu\mu)$ + jets	Powheg & Pythia 8 with AZNLO	CTEQ6L1
$Z(\to ee\gamma)jj$	MadGraph [89] & Pythia 8 with A14 [138]	NNPDF2.3lo [142]
$Z(\to \mu\mu\gamma)jj$	MadGraph & Pythia 8 with A14	NNPDF2.3lo
$llll$	Sherpa 2.2.16	NNPDF3.0nnlo
$lll\nu$	Sherpa 2.2.16	NNPDF3.0nnlo
$lljj$	Sherpa 2.2.16	NNPDF3.0nnlo
$lll\nu jj$	Sherpa 2.2.16	NNPDF3.0nnlo
$Z(\to qq)Z(\to ll)$	Sherpa 2.2.14	NNPDF3.0nnlo
$W(\to qq)Z(\to ll)$	Sherpa 2.2.14	NNPDF3.0nnlo
$t\bar{t}$	Powheg & Pythia 8 with A14	NNPDF2.3lo
$t\bar{t}\gamma$	MadGraph & Pythia 8 with A14	NNPDF2.3lo

Table 7.2: Background MC samples used in this analysis, including non-resonant $Z\gamma$, Z+jets, EW $Z\gamma jj$, diboson, and $t\bar{t}$ processes, together with the event generator, tune, and PDF set used for each sample.

7.3 Event selection and reconstruction

Events are required to pass the trigger requirements described in Section 7.2. Selected events must contain at least one photon and two oppositely charged, same-flavour leptons (electrons or muons). They are required to originate from the primary vertex of the event, defined as the vertex with the highest sum of squared transverse momenta, $\sum p_{\mathrm{T}}^2$, of the associated tracks.

Reconstructed electrons are required to satisfy the *Loose* likelihood identification criteria, have $p_{\mathrm{T}} > 10$ GeV, and lie within $|\eta| < 2.47$, excluding the calorimeter transition region $1.37 < |\eta| < 1.52$. Muons are required to satisfy the *Medium* identification criteria, have $p_{\mathrm{T}} > 5$ GeV, and satisfy $|\eta| < 2.5$. Both electrons and muons must fulfil track-based impact parameter requirements, with $|\Delta z_0 \sin\theta| < 0.5$ mm, ensuring compatibility with the primary vertex. In addition, the transverse impact parameter significance is required to satisfy $|d_0|/\sigma_{d_0} < 3$ for electrons and $|d_0|/\sigma_{d_0} < 5$ for muons, which suppresses backgrounds from non-prompt leptons.

Photon candidates are required to satisfy the *Tight* identification criteria, have $p_{\mathrm{T}} > 10$ GeV, and lie within $|\eta| < 2.37$, excluding the calorimeter transition region. This stringent selection is motivated by the fact that misidentified photons constitute one of the dominant background sources. All selected leptons and photons are required to satisfy the standard ATLAS isolation criteria.

Jets are used to categorise events targeting the VBF production mode. They are reconstructed from particle-flow objects using the anti-k_t algorithm [118] with a radius parameter $R = 0.4$. Jets are required to have $p_{\mathrm{T}} > 25$ GeV and rapidity $|y| < 4.4$. To suppress pile-up contributions, jets with $p_{\mathrm{T}} < 60$ GeV are required to pass the jet-vertex tagger criteria [163, 164], applied to both central and forward jets.

To avoid double-counting of reconstructed objects, an overlap removal procedure is applied. Electrons within $\Delta R < 0.02$ of a muon are removed. Photon candidates within $\Delta R < 0.3$ of a selected lepton are discarded to suppress photons originating from bremsstrahlung[1]. Finally, jets within $\Delta R < 0.2$ of selected leptons or photons are removed.

The Z boson is reconstructed from a pair of oppositely charged, same-flavour leptons compatible with the trigger. Offline p_{T} thresholds exceeding the trigger requirements by 1-2 GeV are applied to ensure full trigger efficiency. In the muon channel, collinear photons found within $\Delta R < 0.15$ of a muon are added to the muon four-momentum to account for final-state radiation, improving the dimuon invariant mass resolution. The dilepton invariant mass must satisfy $|m_{\ell\ell} - m_Z| < 10$ GeV, where $m_Z = 91.2$ GeV. If multiple dilepton candidates are present, the pair with invariant mass closest to m_Z is selected. A kinematic fit is then applied to constrain the dilepton system to the known Z boson mass.

The reconstructed Z boson candidate is combined with the highest-p_{T} photon in the event to form the Higgs boson candidate. The invariant mass of the $\ell\ell\gamma$ system is required

[1] Bremsstrahlung radiation is emitted when a charged particle is accelerated or deflected by the electromagnetic field of another charged particle, leading to the emission of photons collinear with the parent lepton.

to lie in the range $115 < m_{\ell\ell\gamma} < 160$ GeV, suppressing backgrounds from on-shell Z boson production with final-state radiation photons. The photon transverse momentum is required to satisfy $p_{\mathrm{T}}^{\gamma}/m_{\ell\ell\gamma} > 0.09$ to further suppress non-resonant $Z\gamma$ background.

The overall reconstruction and selection efficiency for the $H \to Z(\to \ell\ell)\gamma$ signal ranges between 20% and 26%, depending on the Higgs boson production mode.

7.4 Event categorisation

The categorisation of events into distinct regions of phase space is used to maximise sensitivity to the Higgs boson signal relative to the various backgrounds. A total of 13 mutually exclusive categories are defined, based on lepton and jet multiplicities and on kinematic properties of the reconstructed objects. Several of these categories are constructed using XGBoost [165] BDTs trained to enhance the separation between signal and background while minimising correlations with the $m_{\ell\ell\gamma}$ invariant mass to avoid mass-sculpting effects. The input variables used in the BDTs are described in Ref. [20].

Each category is designed to target one or more specific Higgs boson production modes, although some degree of overlap between production modes is unavoidable. A schematic overview of the event categorisation is shown in Fig. 7.2.

The first category, referred to as the *Lepton* category and shown in yellow, requires the presence of more than three identified and isolated leptons in the event. This category is primarily enriched in VH and $t\bar{t}H$ production processes (see Fig. 2.2c and Fig. 2.2d for the corresponding Feynman diagrams). The dominant background contributions arise from Z+jets, diboson processes, and non-resonant $Z\gamma$ production.

Events with fewer than three leptons but containing at least two jets and a minimum BDT score are assigned to the *VBF* category, shown in green. This category targets Higgs boson production via vector-boson fusion (Fig. 2.2b). A dedicated BDT is used to separate the VBF signal from non-resonant $Z\gamma$ and electroweak $Z\gamma jj$ backgrounds, as well as to suppress contamination from ggF production. Based on the BDT output score, events are further classified into a Tight (*VBFT*) and a Loose (*VBFL*) subcategory. Due to its stringent selection criteria, this category provides the highest signal purity and the strongest discrimination against background.

Events that do not satisfy the *VBF* category requirements are categorised according to the relative transverse momentum of the photon. Events with $p_{\mathrm{T}}^{\gamma}/m_{\ell\ell\gamma} \geq 0.4$ are assigned to the high relative-p_{T} category (*HRelpT*), shown in pink, while the remaining events populate the low relative-p_{T} category (*LRelpT*), shown in orange. A high relative transverse momentum of photons is indicative of a boosted Higgs boson topology, which is less common in background processes.

The *HRelpT* category is divided into electron and muon channels. Within each channel, a dedicated BDT is used to further separate events into Tight and Loose subcategories. These categories are primarily sensitive to gluon–gluon fusion (ggF) production (Fig. 2.2a).

The remaining events, characterised by a low relative photon transverse momentum, define the *LRelpT* category, which is also dominated by ggF production. As in the HRelpT

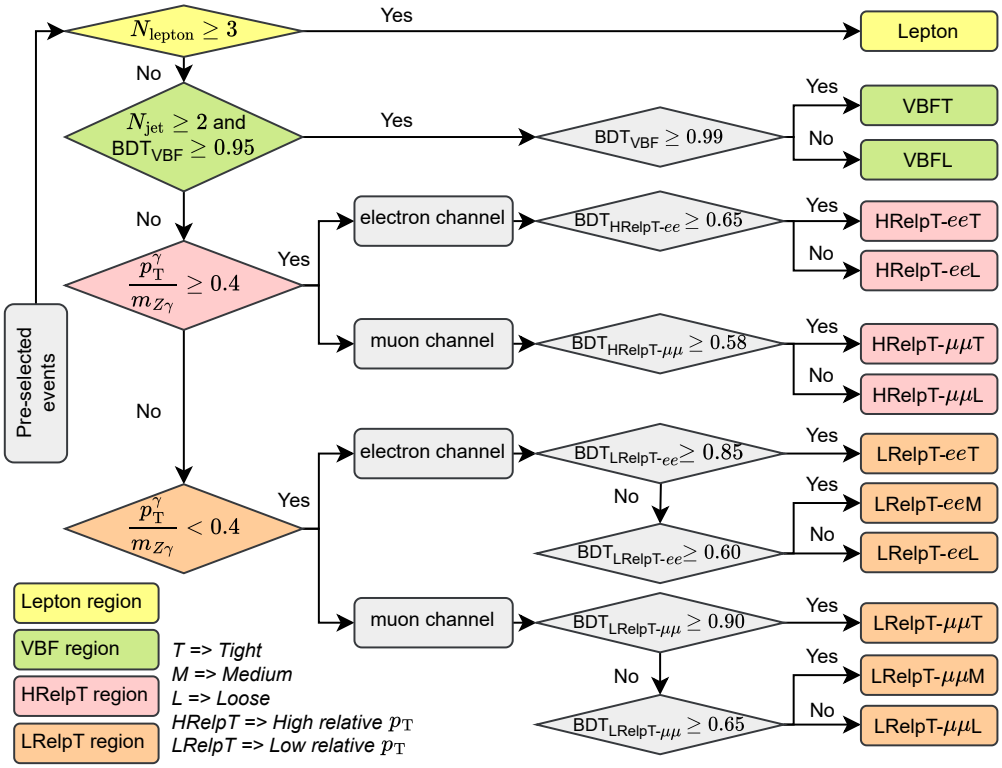


Figure 7.2: Schematic overview of the event categorisation based on lepton and jet multiplicities, kinematic selections, and multivariate algorithms. The *Lepton* category is shown in yellow. The *VBF* category, divided into Tight (VBFT) and Loose (VBFL) subcategories, is shown in green. The high relative-p_{T} region is shown in pink and is subdivided into four categories according to lepton flavour and BDT score (Tight and Loose). The low relative-p_{T} region is shown in orange and is further subdivided into six categories by lepton flavour and BDT score (Tight, Medium and Loose). Figure from Ref. [20].

case, events are split by lepton flavour, and a BDT is used to define Tight, Medium, and Loose subcategories based on the BDT score. This category, particularly the loose subcategory, contains the largest background contribution, mainly from Z+jets and non-resonant $Z\gamma$ processes.

Although each BDT is trained to enhance sensitivity to a specific production mode, some signal mixing between categories persists. Figure 7.3 shows the signal MC population across the different categories.

The BDT score thresholds are determined by scanning the expected signal significance as a function of the score, where the significance quantifies the expected sensitivity to the signal relative to the background. The threshold corresponding to the maximum expected significance is therefore selected, providing an optimal balance between signal efficiency and background rejection.

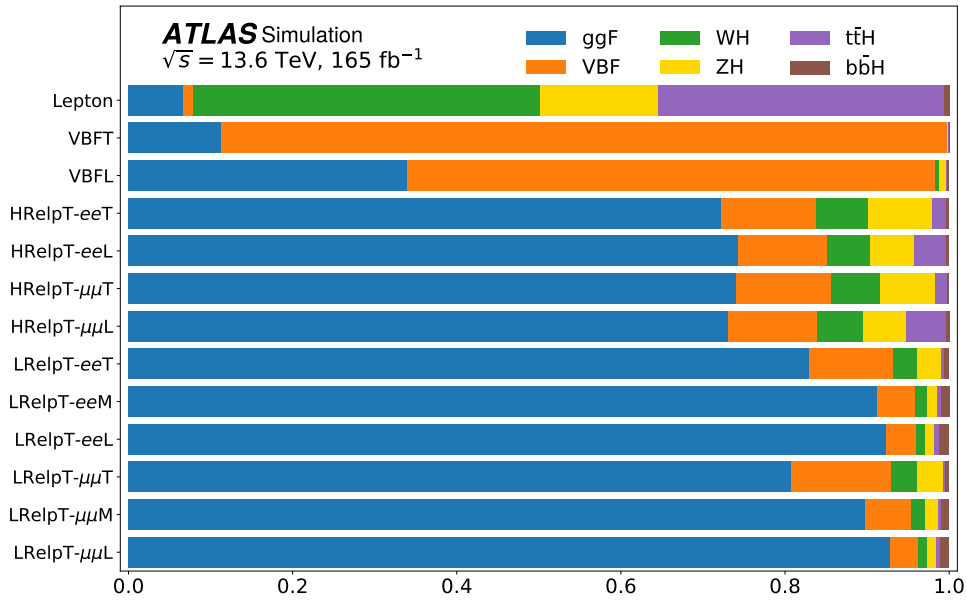


Figure 7.3: Signal MC population in the different event categories. The panels correspond to the categories shown in Fig. 7.2, while the horizontal axis shows the percentage contribution of each Higgs boson production mode in a given category. The *ggF* process is shown in blue, *VBF* in orange, *VH* in green and yellow for W and Z bosons respectively, $t\bar{t}H$ in purple, and $b\bar{b}H$ in brown. Figure from the auxiliary material of Ref. [20].

Motivation for separate category treatment The benefit of treating categories independently, rather than merging them into a single inclusive selection, can be understood analytically.

Considering two statistically independent samples, each containing signal and background events. Sample 1 is characterised by a signal yield S_1, a background yield B_1, and a total event count $N_1 = S_1 + B_1$. Similarly, sample 2 is characterised by S_2, B_2, and $N_2 = S_2 + B_2$.

If the two samples are merged and analysed inclusively, the resulting sensitivity is given by

$$Z_{\text{pooled}} = \frac{S_1 + S_2}{\sqrt{N_1 + N_2}} \,. \tag{7.1}$$

In contrast, when the samples are analysed separately and the sensitivities are combined statistically, the total sensitivity becomes

$$Z_{\text{sep}} = \sqrt{Z_1^2 + Z_2^2} = \sqrt{\left(\frac{S_1}{\sqrt{N_1}}\right)^2 + \left(\frac{S_2}{\sqrt{N_2}}\right)^2} = \sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}} \,. \tag{7.2}$$

The difference between the two approaches is defined as

$$\Delta Z = Z_{\text{sep}} - Z_{\text{pooled}} \,. \tag{7.3}$$

Since both Z_{sep} and Z_{pooled} are positive, the inequality $\Delta Z \geq 0$ can be studied by squaring both sides. After algebraic simplification, the difference can be written as

$$\Delta Z = \frac{(S_1 N_2 - S_2 N_1)^2}{N_1 N_2 (N_1 + N_2)} \,. \tag{7.4}$$

The denominator consists exclusively of event counts and is therefore strictly positive. The numerator is a perfect square and is thus non-negative. It follows immediately that $\Delta Z \geq 0$.

This result shows that a separate analysis yields a higher sensitivity than a pooled analysis whenever the signal-to-total ratios S_1/N_1 and S_2/N_2 differ between the samples. By treating categories independently, samples with higher signal purity receive a larger statistical weight, whereas merging categories dilutes the contribution of cleaner samples relative to those with a larger background fraction.

7.5 Signal and background modelling

Modelling the signal and background is a key step in setting expectations for the observed data and in testing the signal hypothesis without introducing bias. In this analysis, the signal and background components are modelled separately, with a relatively simple, well-constrained description of the signal and a more flexible approach for the background to account for its complex shape. The signal strength is extracted using an unbinned extended maximum-likelihood fit to the $m_{Z\gamma}$ distribution, which allows the full event-by-event information to be used and reduces sensitivity to binning choices and statistical fluctuations.

For the signal modelling, the double-sided Crystal Ball (DSCB) function is used. It consists of a Gaussian core with power-law tails on both sides, accounting for the detector resolution and non-Gaussian effects. The shape of the DSCB is fixed from simulation in each category after combining all signal samples. The mean is shifted by 90 MeV to account for the measured Higgs boson mass of $m_H = 125.09$ GeV, while the simulation was generated at 125 GeV. In simulation, signal normalisation determines the expected signal yield, whereas in data, it is the only free parameter that determines the observed signal yield. A contribution up to 3.8% in some categories arises from $H \to \mu\mu$ decays. This contribution is modelled using a separate DSCB template, with its normalisation fixed to the SM prediction.

The background modelling starts with a dedicated inclusive background template that accounts for the main backgrounds, non-resonant $Z\gamma$ and Z+jets. Its relative fraction is determined using a two-dimensional sideband method, commonly referred to as the ABCD method [166]. This approach exploits regions defined by loosening the nominal photon identification and isolation requirements, comparing event yields in the different sidebands to extract the $Z\gamma$ purity in data under the nominal photon selection. The underlying assumption is that in the sidebands, data is enriched in Z+jets events, and that the photon identification and isolation variables are approximately uncorrelated. The number of $Z\gamma$

events in the signal region (region A) is

$$N_A^{Z\gamma} = N_A^{\text{data}} - N_A^{Z+\text{jets}} = N_A^{\text{data}} - \frac{N_B^{\text{data}} \cdot N_C^{\text{data}}}{N_D^{\text{data}}}, \tag{7.5}$$

where B, C, and D are the sideband regions. The fraction of $Z\gamma$ is found to be $0.49^{+0.05}_{-0.10}$ in Run 3, compared to $0.78^{+0.04}_{-0.09}$ in Run 2. This reduction in purity is due to several factors, including increased pile-up, lowering the photon momentum in the event selection (from 15 GeV to 10 GeV), which allows more jets to fake photons, and the smaller SM ratio of $Z\gamma$ to Z+jets at the Run 3 centre-of-mass energy.

The fraction is then extrapolated to the different categories and smoothed to reduce statistical fluctuations by fitting it with a low-order polynomial or exponential-polynomial function. The extrapolation relies on $Z\gamma$ fast-simulation for the signal component and data-driven Z+jets control samples for the background component. For the *VBF* categories, a similar smoothing procedure is applied, after adding the EW $Z\gamma jj$ background. The diboson background contribution is included in the template of the *Lepton* category. Any residual differences between the simulation template and the data outside the signal region are accounted for by scaling the simulation accordingly.

The spurious signal (SS) study quantifies the extent to which a chosen background model can induce an artificial signal peak in the signal region when fitted to a background-only sample. Multiple analytic functions are tested to model the background. For each category, the template is selected by identifying the functions that introduce the least bias. A function is considered acceptable if the |SS| is sufficiently small compared to the expected signal uncertainty and if the probability of the background-only fit exceeds 1%. If multiple functions satisfy these criteria, the one with the fewest free parameters is chosen. If none pass, the fit range is narrowed down until this criterion is fulfilled, with a minimum window of 35 GeV. To account for uncertainties in the relative fractions of the two background components ($Z\gamma$ and Z + jets), these fractions are varied up and down within their uncertainties to build alternative templates. The spurious signal study is repeated for each variation, and the maximum |SS| observed across all templates is assigned as the systematic uncertainty of the background model. Finally, the background parameters are left free and are extracted directly from a fit to the invariant mass distribution in data.

Once the fits are defined, the sensitivity of the different categories is estimated. Table 7.3 summarises the resulting sensitivities together with the intermediate quantities entering their calculation. For each category, the mass window containing 68% of the expected signal, w_{68}, is first determined. Within this window, the expected signal (S_{68}^{exp}) and background (B_{68}^{exp}) yields are extracted from a fit to the Asimov dataset [167]. The sensitivity is then computed using these quantities.

As expected, categories with a high signal-to-background ratio yield the largest sensitivities. In particular, the most sensitive category is *VBFT*, followed by *LRelpT-μμM*. The inclusive sensitivity is obtained by combining the individual categories statistically. This combined result shows a clear improvement compared to a fully inclusive approach without categorisation, for which the expected sensitivity would be around 0.78.

Category	w_{68} [GeV]	S_{68}^{exp}	B_{68}^{exp}	N_{68}	$S_{68}^{\text{exp}}/\sqrt{S_{68}^{\text{exp}}+B_{68}^{\text{exp}}}$
Lepton	4.4	1.5 ± 1.1	75.4 ± 2.8	70	0.17
VBFT	3.8	1.5 ± 1.1	1.3 ± 0.4	3	0.92
VBFL	4.0	2.8 ± 2.0	27.2 ± 1.8	18	0.52
HRelpT-eeT	3.2	1.2 ± 0.8	6.6 ± 0.9	9	0.43
HRelpT-eeL	4.0	2.9 ± 2.1	53.9 ± 1.8	61	0.39
HRelpT-$\mu\mu$T	3.9	2.4 ± 1.7	20.3 ± 1.7	25	0.51
HRelpT-$\mu\mu$L	4.1	2.4 ± 1.7	56.5 ± 1.7	61	0.32
LRelpT-eeT	3.8	8.9 ± 6.2	231 ± 6	240	0.57
LRelpT-eeM	4.1	29 ± 20	$2\,562 \pm 19$	2 587	0.56
LRelpT-eeL	4.5	24 ± 17	$13\,122 \pm 50$	13 074	0.21
LRelpT-$\mu\mu$T	3.9	4.9 ± 3.4	95 ± 4	107	0.49
LRelpT-$\mu\mu$M	4.1	34 ± 24	$2\,527 \pm 19$	2 583	0.67
LRelpT-$\mu\mu$L	4.4	36 ± 25	$16\,844 \pm 40$	16 642	0.28
Inclusive	4.0	150 ± 110	$35\,625 \pm 70$	35 480	1.81

Table 7.3: Expected sensitivity of the individual event categories and of the inclusive case, computed as $S_{68}^{\text{exp}}/\sqrt{S_{68}^{\text{exp}}+B_{68}^{\text{exp}}}$. The expected signal events (S_{68}^{exp}) and background events (B_{68}^{exp}) obtained from an MC Asimov dataset within a mass window of width w_{68}, defined to contain 68% of the signal. The number of observed events in the data for the same window, N_{68}, is also reported. The uncertainty on the signal yield reflects the impact of the signal strength modifier μ, while the background uncertainty accounts for the statistical uncertainty from the fit. For the inclusive case, the sensitivity is obtained by combining the individual category sensitivities in quadrature. Table adapted from Ref. [20].

7.6 Statistical and systematic uncertainties

The measurement is affected by both statistical and systematic uncertainties, which affect signal extraction and ultimately determine the analysis's sensitivity.

Table 7.4 summarises the different sources of uncertainty affecting the signal strength for both the expected and observed results, ordered from the largest to the smallest contribution in the expected case. The expected uncertainties reported in this section are evaluated using an Asimov dataset constructed under the SM signal and background assumptions. The Asimov dataset corresponds to the median experimental outcome in the absence of statistical fluctuations.

The dominant source of uncertainty arises from statistical effects, with an expected (observed) uncertainty of 0.68 (0.65). Systematic uncertainties are evaluated by varying each nuisance parameter by $\pm 1\sigma$ from its nominal value and propagating the effect to the final measurement. The total systematic uncertainty is found to be less than one-third of the statistical uncertainty, totalling 0.16 and 0.17 for the expected and observed cases, respectively. The largest systematic contribution originates from the spurious signal associated with the background modelling being 0.11, followed by theoretical modelling uncertainties. Uncertainties in the electron, muon, and photon energy scale and resolution are found to be below 0.2.

Uncertainty source	$\Delta\mu$	
	Expected	Observed
Statistical uncertainty	0.68	0.65
Systematic uncertainty	0.16	0.17
Spurious signal (background modelling)	0.11	0.11
QCD scale, PDF+α_S, parton shower	0.09	0.06
Branching ratio ($H \to Z\gamma$)	0.05	0.05
Luminosity	0.03	0.03
Photon efficiency	0.02	0.03
Jet	0.02	0.07
Electron and photon energy scale and resolution	0.02	0.06
Electron efficiency	0.02	0.02
Muon	< 0.01	0.02
Trigger	< 0.01	0.02
Total	0.70	0.67

Table 7.4: Symmetrised impact of the individual sources of statistical and systematic uncertainty on the expected and observed signal strengths, ordered by decreasing impact in the expected case. Table from Ref. [20].

7.7 Results

The results are obtained from a simultaneous unbinned maximum-likelihood fit performed across all analysis categories. The likelihood function includes the signal strength as the parameter of interest, along with nuisance parameters constrained by auxiliary measurements from control regions. These nuisance parameters model the systematic uncertainties affecting both the normalisation and the shape of the signal and background distributions.

The partial Run 3 analysis is first performed independently and then combined with the previous ATLAS search for the $H \to Z\gamma$ decay conducted using Run 2 data [160]. The combination accounts for which sources of systematic uncertainty are correlated or uncorrelated between the two datasets. However, the impact of these correlations on the final results is negligible, as the total uncertainty is dominated by statistical contributions.

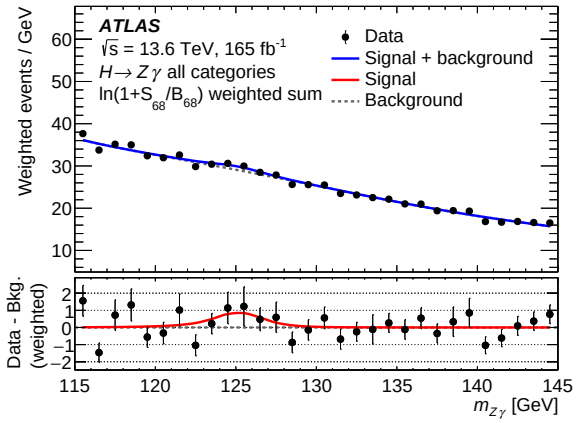
Figure 7.4 shows the distribution of the three-body invariant mass $m_{\ell\ell\gamma}$, where each event is weighted by $\ln(1+S_{68}/B_{68})$ according to its category. The weighted distributions are then summed to form an inclusive spectrum. The partial Run 3 results (left) are compared to the combined Run 2 and Run 3 results (right), where an increased deviation of the data from the background-only fit is observed in the signal region. The complete analysis procedure is applied as well to an Asimov dataset to derive the expected signal strength, uncertainties, and significance, which are then compared with the observed results from data.

Figure 7.5 presents the expected and observed profile likelihood scans of the signal strength for Run 2, partial Run 3, and their combination. The combined likelihood is obtained as the product of the individual likelihood functions. The minimum of the profile likelihood corresponds to the best-fit value of the signal strength, while the total uncertainty is determined by the points where $-2\ln\Lambda = 1$, allowing for asymmetric uncertainties by construction. The statistical significance of the observed excess with respect to the background-only hypothesis ($\mu = 0$) is evaluated as $Z = \sqrt{-2\ln\Lambda_0}$.

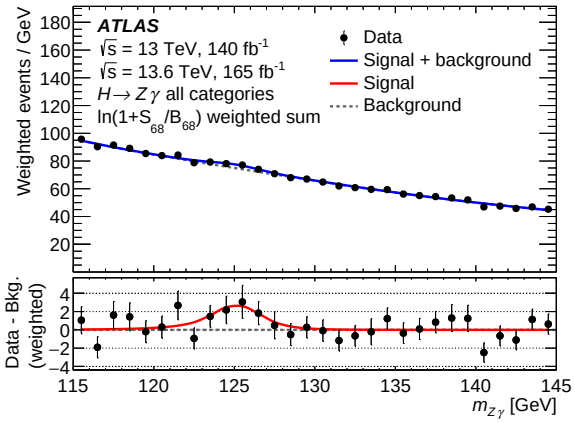
For the partial Run 3 dataset, the observed signal strength is $\mu_{\text{obs}} = 0.9^{+0.7}_{-0.6}$, , while the expected signal strength is $\mu_{\text{exp}} = 1.0 \pm 0.7$. These results are consistent with the SM expectation. The individual signal strength measurements are also consistent with the global signal strength, yielding a p-value of 0.37. The significance of the excess above the background-only hypothesis is 1.4σ for the observed result and 1.5σ for the expected result[2]. Compared to Run 2, the expected significance improves by 28%, driven by improved event selection and categorisation, a higher signal-to-background ratio at the Run 3 centre-of-mass energy, and an increased dataset size.

When combining the Run 2 and Run 3 results, the expected combined signal strength is $\mu_{\text{exp}} = 1.3^{+0.6}_{-0.5}$, while the observed combined signal strength is $\mu_{\text{obs}} = 1.0^{+0.6}_{-0.5}$. The observed significance of the excess above the background-only hypothesis is 2.5σ, compared to an expected significance of 1.9σ. The Run 2 and Run 3 measurements are mutually compatible, with a corresponding p-value of 0.33.

[2] A discovery claim requires a significance of at least 5σ.

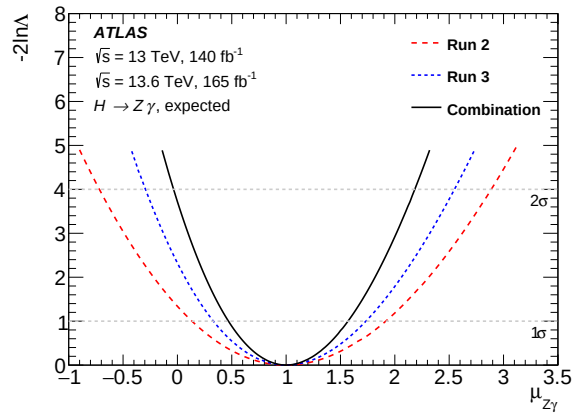


(a) Partial Run 3

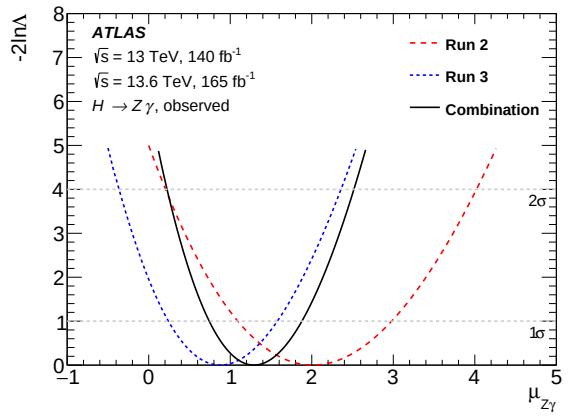


(b) Combined Run 2 + partial Run 3

Figure 7.4: Invariant mass of the three-body system $ll\gamma$ in data, weighted by $\ln(1+S_{68}/B_{68})$ based on the signal and background yield in each category, for (a) partial Run 3 and (b) the combination of Run 2 and partial Run 3. The black points represent the data with statistical uncertainty. The solid blue curve shows the signal-plus-background fit, while the dashed grey curve represents the background-only model. The bottom panel displays the residual between the data and the background model, with the signal fit indicated by a solid red line. Figure from Ref. [20].



(a) Expected



(b) Observed

Figure 7.5: Profiled likelihood scan of the Higgs signal strength for (a) expected and (b) observed results. The dashed blue line corresponds to the Run 3 dataset, the dashed red line shows the previous Run 2 analysis, and the solid black line represents the combined dataset. Dashed grey lines indicate the 1σ and 2σ confidence thresholds. Figure from Ref. [20].

7.8 Conclusions

In this analysis, a search for the SM Higgs boson decaying to the $ll\gamma$ final state is performed using partial Run 3 data from pp collisions at $\sqrt{s} = 13.6$ TeV, collected during the 2022–2024

data-taking period and corresponding to an integrated luminosity of 165 fb^{-1}. The results are obtained using an unbinned maximum-likelihood fit and are subsequently combined with the Run 2 ATLAS results, yielding the most stringent sensitivity to date, with an observed (expected) signal strength of $\mu_{\text{exp}} = 1.3^{+0.6}_{-0.5}$ ($\mu_{\text{obs}} = 1.0^{+0.6}_{-0.5}$). The individual and combined measurements are mutually consistent and in agreement with the Standard Model expectation within uncertainties.

Compared to the Run 2 analysis, the expected significance improves by 61% in the combined result. This gain is driven by the enhanced event selection and categorisation, the increased centre-of-mass energy of Run 3, and the larger integrated luminosity. Nevertheless, the measurement remains statistically limited.

Several strategies to further improve the sensitivity are currently under investigation. In particular, optimised event selection and background discrimination techniques are expected to provide the largest gains. Approaches that exploit correlations among multidimensional observables offer significant potential to enhance the signal-to-background ratio. This motivates the exploration of ML-based event selection strategies, which are investigated in the following chapter.

Chapter 8

Enhancing $H \to Z\gamma$ sensitivity via machine learning Z selection

8.1 Introduction

The results presented in the previous chapter, hereafter referred to as the *nominal* analysis, show a significant improvement in analysis sensitivity compared to earlier studies. Nevertheless, the observed signal significance of 2.5σ remains far from the discovery threshold. This limitation is driven primarily by statistical uncertainties. While increased integrated luminosity will naturally improve sensitivity, further gains can be achieved by identifying and addressing weaknesses in the current event selection to enhance signal acceptance. Additional improvements could also arise from the inclusion of alternative final states, such as hadronic decays of the Z boson or the invisible decay into neutrinos.

An inspection of the event-selection cutflow summarised in Table 8.1 reveals that the trigger, lepton selection, and photon selection steps exhibit the lowest relative efficiencies, all below 90%. Trigger optimisation studies are beyond the scope of this thesis, although improvements in this area would be highly relevant. The lepton and photon selections, on the other hand, could in principle be optimised using similar strategies to improve their efficiencies.

Any modification to the object selection beyond the recommended working points requires deriving new efficiency scale factors to correct for differences between simulation and data. For leptons, this procedure is commonly and robustly performed using the reconstructed Z boson mass peak in data, making such an approach experimentally well controlled. In contrast, determining photon efficiency scale factors is more challenging, as photons do not originate from the decay of a narrow on-shell resonance, making data-driven efficiency measurements more subtle. Improvements to the photon selection are therefore left for future work.

The focus of the novel work presented in this chapter is the improvement of the lepton selection using machine-learning (ML) algorithms to increase the acceptance of Z boson signal events. The training procedure, model implementation, and the resulting ML-based event

Selection cut	Rel. Efficiency	Total Efficiency
Event Quality	100	100
Triggers	60.3	60.3
Two leptons	83.7	50.5
Lepton PID	79.4	40.1
Lepton ISO	94.1	37.7
Overlap Removal	99.7	37.6
llg_size>0	99.0	37.2
2l+γ	76.3	28.4
Trigger match	99.7	28.3
Photon Overlap Removal	95.2	27.0
m_{ll}	91.3	24.6
p_{T}^{γ}	98.5	24.2
Photon PID	83.1	20.1
Photon ISO	88.6	17.8
$m_{ll\gamma}$	99.1	17.7

Table 8.1: Cutflow of the nominal $H \to Z\gamma$ selection for a simulated *VBF* signal sample, showing the relative efficiency with respect to the previous cut and the total efficiency.

selection are described, together with preliminary results of the improved Z selection and its impact on subsequent analysis steps, including event categorisation, modelling, systematic uncertainties, and expected significance.

However, a complete reproduction of the analysis using the selection presented below was not possible due to time constraints. Although the methodology described here was not directly implemented in the nominal analysis, it informed and helped motivate some analysis choices. This study is ongoing and is intended to be incorporated into future iterations of the analysis using the full Run 3 dataset.

8.2 Training and framework

To improve the selection steps in the cutflow, the hard-cut-based approach of the nominal analysis is replaced with ML models, which can be optimised to maximise the expected Z signal significance.

The samples we use for the Z boson models are the same Z+jets and $t\bar{t}$ background samples listed in Table 7.2. These samples are exploited to construct training targets corresponding to identified and non-identified leptons, isolated and non-isolated leptons, as well as genuine Z boson decays and combinatorial or fake Z candidates.

The training strategy is divided into two stages. In the first stage, we train single lepton models to assess lepton identification and isolation performance independently. In the second stage, we train a dedicated Z boson model using the outputs of the single lepton models as inputs, together with additional kinematic information, to evaluate the likelihood that a

given lepton pair originates from a Z boson decay.

8.2.1 Single lepton models

Identification (PID) and isolation (ISO) models are constructed separately for electrons and muons. For electrons, we use the output score from an already-trained deep neural network available in the DAOD instead of training a new PID model.

The input variables correspond to standard PID- or ISO-related observables, and the target is defined using the lepton `TruthType`. The features used in the `mu_PID` model include variables such as `pseudorapidity`, `energyLoss`, `CaloMuonScore`, `CaloMuonIDTag`, as well as quantities related to the number of hits and holes in the associated tracks. The ISO models use variables such as `topoEtCone`, `ptvarcone`, `ptcone`, `neflowisol`, and `pile-up` observables. The latter is included to provide the model with information about the event environment, allowing it to account for the increased activity expected in bunch crossings with a high number of simultaneous pp interactions. Good agreement between data and simulation is observed for all training variables, indicating that the MC samples provide an adequate description of the data.

We train the `lep_PID` and `lep_ISO` models using the XGBoost framework [165], with 1 000 boosting trees, a binary cross-entropy loss, and accuracy for evaluation. Four-fold cross-validation is employed to ensure sufficient statistical power and robustness for the subsequent meta-models.

Model performance is assessed using receiver operating characteristic (ROC) curves and accuracy metrics. Feature importance scores are used to identify the most relevant inputs and to assess whether any variables can be removed in order to reduce the model complexity. All models demonstrate strong performance in their respective tasks, achieving accuracies consistently above 0.95.

A detailed definition of `TruthType`, along with the distributions of the input variables, the feature importance rankings, and the ROC curves, is provided in Appendix C.1.1.

8.2.2 Z boson models

Once individual leptons are assigned `lep_PID` and `lep_ISO` scores, we pass these quantities to the `Z_ID` model together with the relative longitudinal impact parameter significance of the two leptons, $\Delta z_0/\Delta\sigma_{z_0}$. We train separate `Z_ID` models for electrons and muons to account for the different background contributions affecting each lepton type.

The Z label is constructed from same-flavour lepton pairs originating from a Z boson, as defined using the `TruthOrigin` information. Most background events used for training originate from $t\bar{t}$ samples, while the signal is dominated by well-reconstructed Z bosons from Z+jets samples.

We intentionally restrict the set of input variables to prevent the explicit reconstruction of the Z boson invariant mass. This choice ensures that selections based on the `Z_ID` score do not introduce sculpting effects in the dilepton invariant mass distribution.

The `Z_ID` models use the same framework, architecture, and initial loss function as the single lepton models. Corresponding validation plots are provided in Appendix C.1.2. The model outputs a probability that a dilepton pair originates from a Z boson, which we transform into a logit score to enhance discrimination in the tails of the distribution. This logit score is used throughout the analysis for selection and optimisation studies and is defined as

$$logit(x) = ln\left(\frac{x}{1-x}\right) . \tag{8.1}$$

To determine the optimal working point of the `Z_ID` score, the expected statistical significance is evaluated as a function of the score threshold. Since our goal is to recover signal events that fail the nominal lepton selection, we apply the ML-based requirement only to events that do not pass the nominal selection. In the limiting case of a maximally restrictive ML requirement, no additional events are recovered beyond the nominal selection. As the ML score threshold is progressively relaxed, the ML-based approach accepts more events, enhancing signal acceptance while also introducing additional background.

Because the MC samples do not perfectly model the `Z_ID` score distribution in data (mainly due to missing specific backgrounds), we adopt a data-driven procedure to optimise the score threshold. We select dilepton pairs in data and study the invariant mass distribution for different values of the `Z_ID` logit score. Figure 8.1a shows the resulting logit distribution.

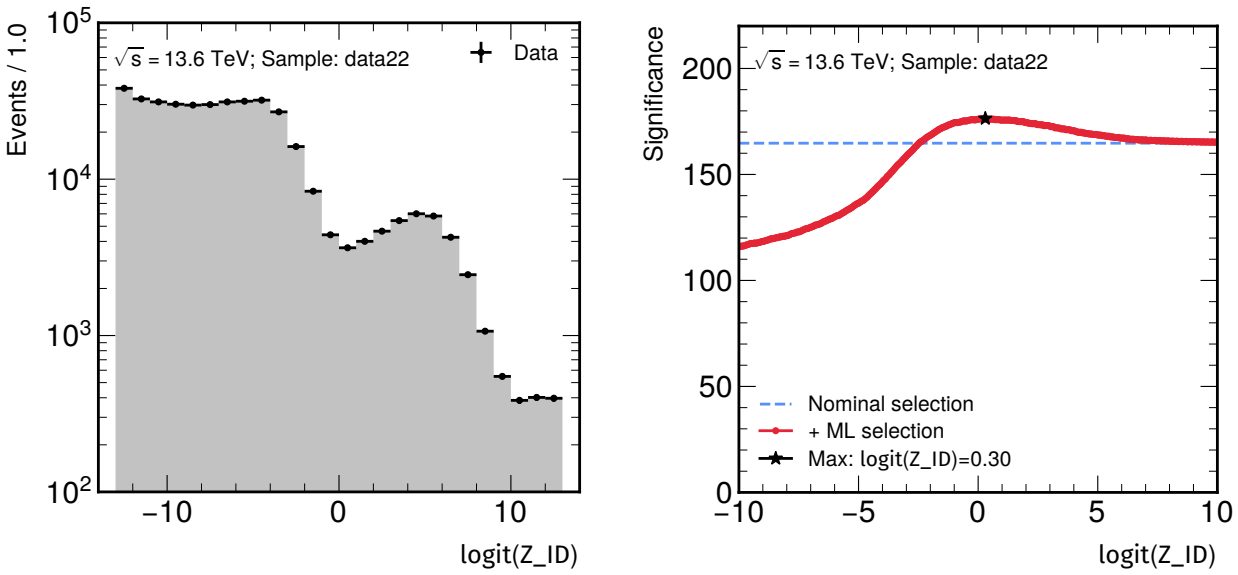
For each score value, we perform a two-step fitting procedure. First, we fit the background component using an exponential of a quadratic , with the Z boson mass window blinded. Subsequently, the Z boson peak is modelled using a double-sided Crystal Ball function, while a fraction of the background contribution is assumed to survive the background subtraction.

From this procedure, we extract the number of Z signal (N_{sig}) and background (N_{bkg}) for each threshold and compute the corresponding significance as

$$Z = \frac{N_{\text{sig}}}{\sqrt{N_{\text{sig}} + N_{\text{bkg}}}} . \tag{8.2}$$

Figure 8.1b shows the resulting Z boson significance as a function of the `Z_ID` logit score. At high score values, the significance approaches that of the nominal selection, while the maximum significance is observed at logit(`Z_ID`)$= 0.3$.

To visualise the impact of the ML-based selection, we also present a ROC-like curve in Fig. 8.2, where we normalise the signal and background yields to the nominal selection. The horizontal axis shows the relative background contribution, while the vertical axis shows the relative signal yield above the nominal. From this figure, we see that the signal yield saturates at roughly 20% improvement, with the maximal significance corresponding to an approximately 50% increase in the Z background.



(a) (b)

Figure 8.1: Illustration of the Z boson significance procedure using a subset of 2022 data. (a) Distribution of the `Z_ID` logit score for dilepton pairs, scanned to determine the optimal ML selection. (b) Resulting significance of the Z signal as a function of the score threshold, obtained by fitting the background and signal components. The nominal selection is indicated with a dashed blue line, the ML-augmented selection with a solid red line, and the score corresponding to maximum significance is marked with a black star.

8.3 Additional machine learning event selection

8.3.1 Implementation

We extend the `H2Zy` analysis framework [168] for this study in a dedicated development branch, `dev2_nbi` [169]. Two main modifications are required. First, additional information is retrieved from the `DAOD` events and propagated to the analysis ntuples to provide the necessary input features for the ML algorithms. Second, the ML inference itself is implemented using the ONNX framework [170]. As a consequence, the event selection cutflow is slightly modified to incorporate decisions based on the ML output.

The updated ntuple production follows the standard cutflow up to the lepton selection stage. For events failing the nominal lepton identification or isolation requirements, dedicated ML algorithms are used to compute lepton identification (`lep_PID`) and isolation (`lep_ISO`) scores. These scores are then combined to form dilepton candidates, which are evaluated by an additional ML classifier trained to identify Z boson decays (`Z_ID`).

Figure 8.3 shows the full cutflow for a representative subset of a *VBF* signal sample. The definition of each cutflow step is given in Appendix C.2. The red contribution illustrates the additional selection step, in which the standard lepton identification and isolation re-

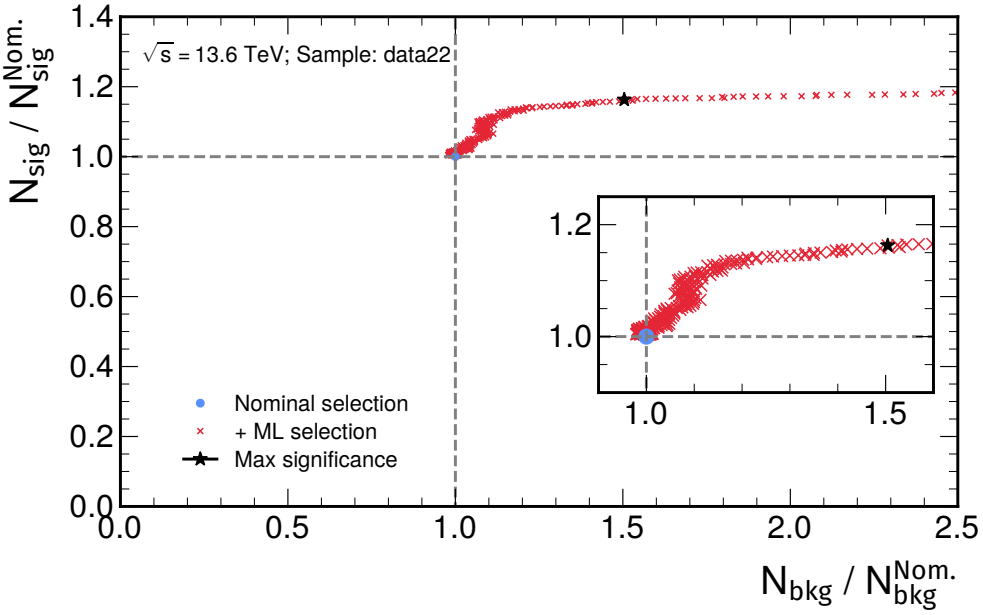


Figure 8.2: Data ROC curve for the ML-augmented Z selection. The number of signal and background events is normalised to the nominal selection, corresponding to the point $(1, 1)$ shown in blue, while adding the ML selection for different score thresholds is shown with red crosses. The point of maximum significance is highlighted with a black star. A zoom-in is provided in the lower-right panel.

quirements are effectively replaced by ML-based selection for events that fail the nominal criteria. After the nominal lepton or the ML-based selection, the event proceeds through the remainder of the nominal analysis selection; as a result, a larger number of events survive the full cutflow, but not all that fulfil the ML cut. Events are stored and labelled according to whether they pass the nominal selection or the ML-based selection, enabling a direct comparison between the two approaches. Extra leptons stored in the event must fulfil the nominal identification and isolation criteria to be considered as possibly originating from VH production.

In summary, the ML-based lepton and Z boson selection provides a controlled way to recover signal events failing the nominal selection, while preserving the dilepton mass shape and enabling the data-driven scale factor procedure described in Section 8.7.

8.3.2 Z selection results

To assess the impact of the ML-based lepton selection, we examine its effect on the reconstruction of the Z boson. The Z boson exhibits a distinctive and narrow resonance at a mass of 91.2 GeV, making it an ideal reference object for this comparison. Figure 8.4a shows the dilepton invariant mass distribution in a combined MC sample, comparing the true Z signal, the nominal selection, and the selection obtained after applying the ML-based approach.

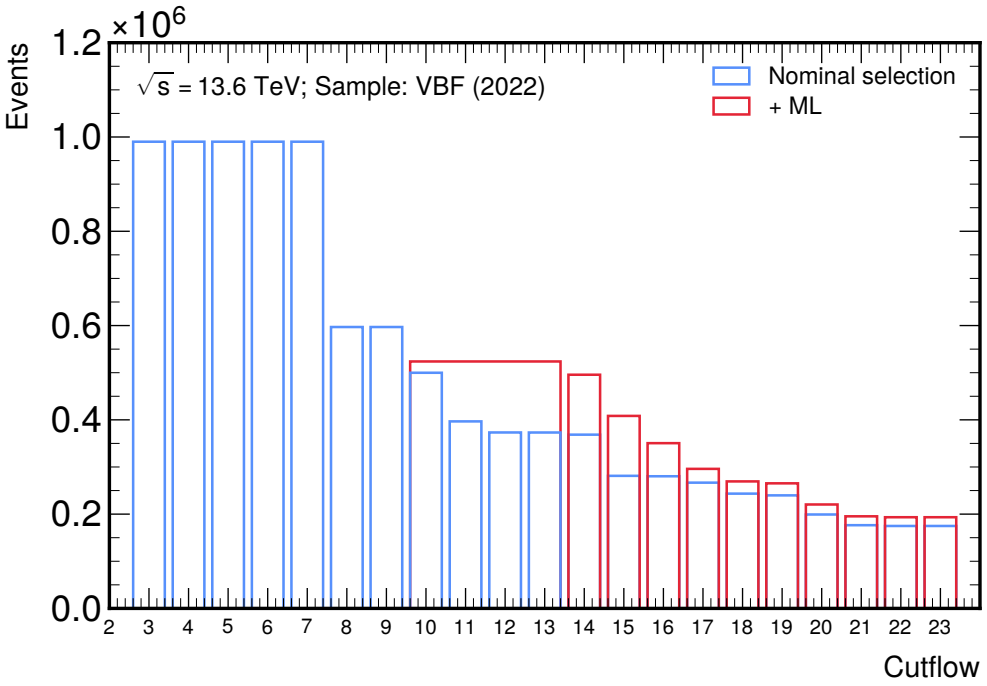


Figure 8.3: Cutflow of the different selection steps for a *VBF* signal sample. The nominal selection is shown in blue, indicating the number of events surviving each sequential step. The additional events selected by the `Z_ID` model are shown in red; this contribution corresponds to the step represented by the wide bin encompassing cuts 10 to 13.

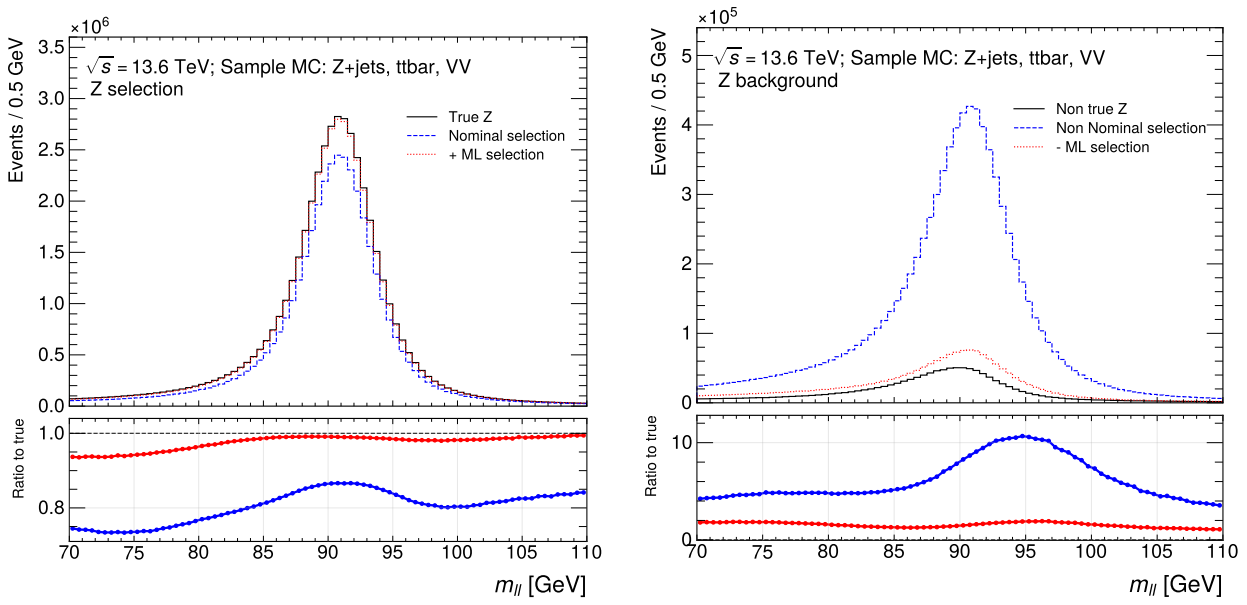
We observe that the `Z_ID` model increases the signal selection efficiency from approximately 80% in the nominal selection to about 97%. The ratio panel shows a uniform improvement across the mass spectrum, demonstrating that our ML-based selection does not introduce any sculpting effects around the Z boson mass peak.

This improvement is even more evident when considering the events failing the selection shown in Fig. 8.4b, which should represent the background (in this case mainly $t\bar{t}$ events). While the nominal selection leaves a pronounced peak around the Z boson mass in the background distribution, the `Z_ID` model correctly recovers this contribution as signal, effectively reducing the residual Z-like structure in the background.

By looking at 2022 data in Fig. 8.5, we see that the ML-based selection improves the signal yield by approximately 20% compared to the nominal selection, while preserving the MC shape of the dilepton mass distribution and maintaining background rejection.

8.4 Impact on event categorisation

The event categorisation follows the same strategy described in Section 7.4 and is left unchanged in this study. However, since the ML-based lepton selection alters the composition of the selected event sample and may populate regions of phase space previously suppressed,



(a) Signal selection

(b) Background selection

Figure 8.4: Dilepton invariant mass distribution for (a) signal selection and (b) background selection in a combined MC dataset of Z+jets, $t\bar{t}$, and dilepton samples. The solid black histogram shows the true Z signal and background, the dashed blue histogram corresponds to the nominal selection, and the dotted red histogram corresponds to the ML-based selection. The lower panel shows the ratio with respect to the truth.

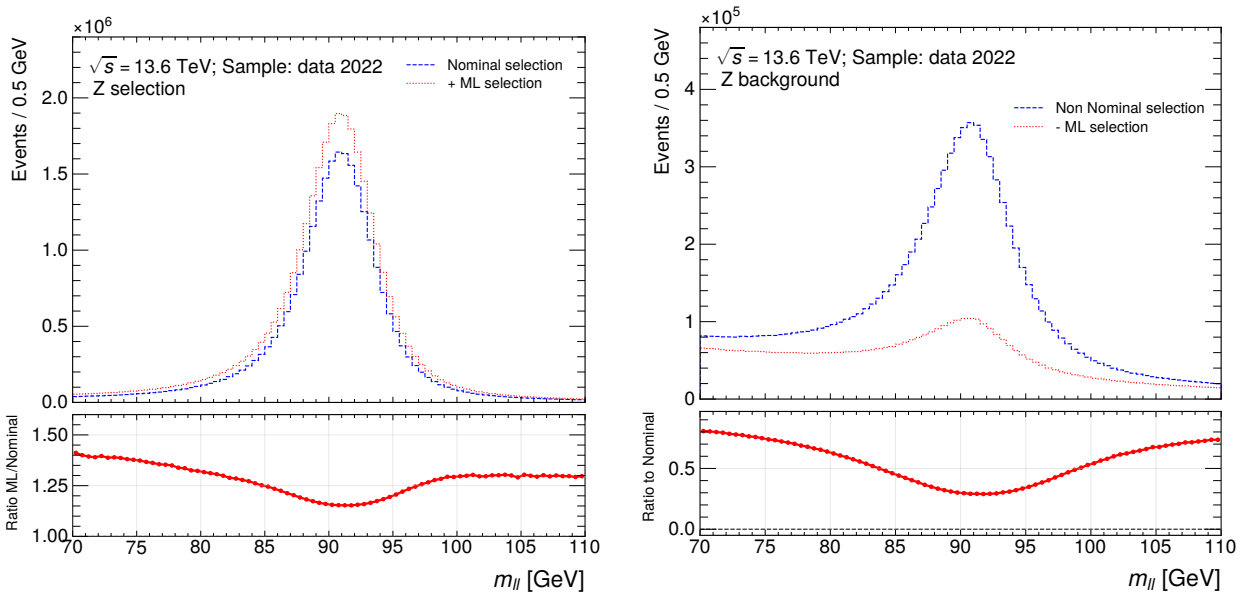
the efficiency of the existing categorisation BDTs is not guaranteed to remain optimal.

We apply the nominal categorisation to the events selected by the ML-based approach to evaluate the impact of the new selection on the category populations. The ML selection targets exclusively the reconstruction of the Z boson, while all subsequent steps in the cutflow remain unchanged. As a consequence, we do not expect an increase in the number of backgrounds associated with photon reconstruction.

For processes with a genuine Z boson in the final state, including the Higgs signal and irreducible Z containing backgrounds, the ML-based selection is expected to increase the yield with comparable efficiency. Figure 8.6 shows the fraction yield of different signal samples per category. We observed proportions comparable to the nominal selection from Fig 7.3, showcasing a general and stable improvement in Z selection for the Higgs signal.

However, because the nominal selection already achieves a very high Z purity, any additional contribution from non-Z processes, even if small in absolute terms, represents a significant relative increase and must be carefully evaluated.

From the nominal analysis, diboson processes were found to contribute predominantly to the *Lepton* category, electroweak $Z\gamma jj$ events mainly populated the *VBF* categories, and the contribution from $t\bar{t}$ processes was negligible. In this study, particular attention is therefore given to assessing whether the ML-based selection leads to a non-negligible increase in non-Z



(a) Signal selection

(b) Background selection

Figure 8.5: Dilepton invariant mass distribution for (a) signal selection and (b) background selection in a data 2022 sample. The dashed blue histogram corresponds to the nominal selection, and the dotted red histogram corresponds to the ML-based selection. The lower panel shows the ratio of the ML selection over the nominal.

backgrounds, specifically from diboson and $t\bar{t}$ processes, and how this propagates through the existing categorisation.

Figures 8.7 and 8.8 show the invariant mass distribution of the $ll\gamma$ system in MC for both signal (in purple) and background. The ratio panel highlights the improvement over the nominal Z selection. The *LRelpT-eeM* and *LRelpT-eeL* categories exhibit the largest statistical gains, at around 20%. The remaining categories show improvements of 10%-20%, with the *Lepton* category showing the smallest increase.

8.5 Signal and background modelling

As for the event categorisation, the signal and background modelling strategy follows that of the nominal analysis in terms of functional forms and parametrisation. In the nominal analysis, an unbinned maximum-likelihood fit is employed, which is optimal given the limited statistics in some categories. However, due to the limited scope of this study, the fits presented here are instead performed using a binned χ^2 minimisation. No changes to the underlying signal or background models are introduced, and no data-driven corrections or scaling to match data are applied. This approach allows us to isolate the effect of the ML-based selection on the MC distributions.

The signal is modelled independently in each category using a double-sided Crystal Ball

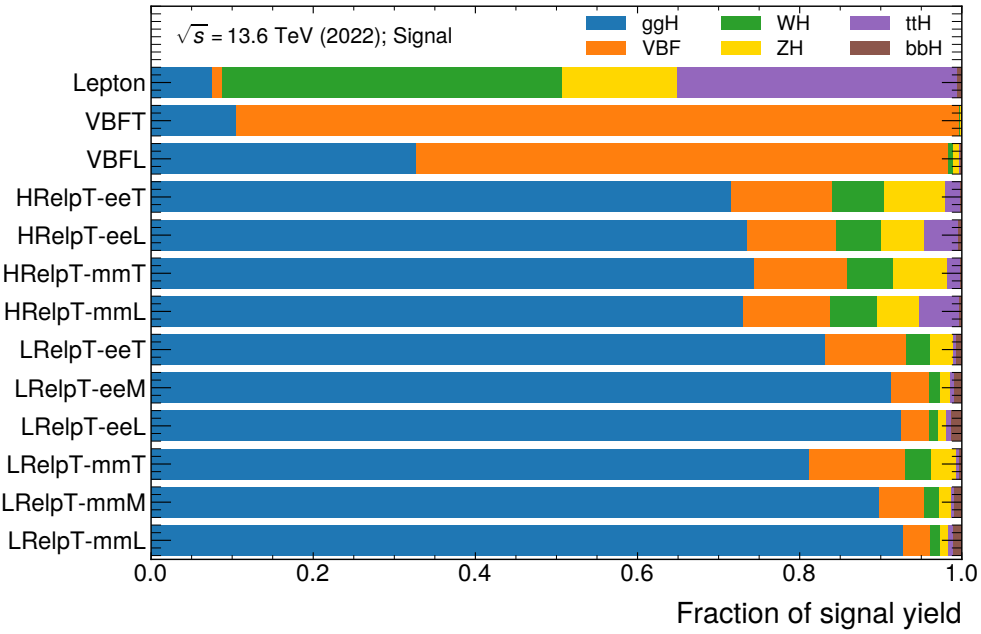


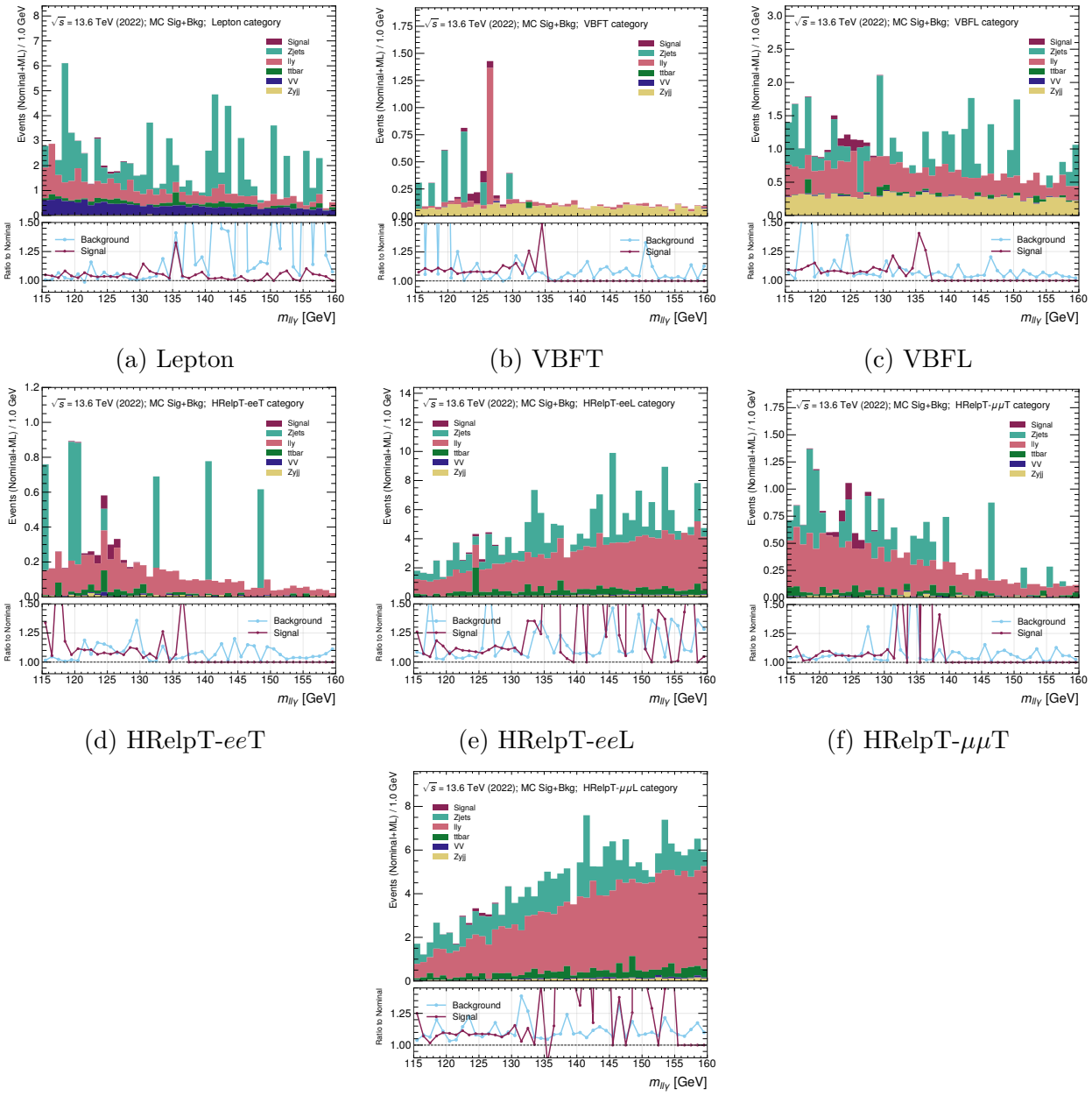
Figure 8.6: Signal population in the different event categories for the ML selection for the 2022 MC sample. The panels correspond to the different categories, while the horizontal axis shows the fraction of signal yield contribution of each Higgs boson production mode in a given category. The *ggF* process is shown in blue, *VBF* in orange, *VH* in green and yellow for W and Z bosons respectively, $t\bar{t}H$ in purple, and $b\bar{b}H$ in brown.

(DSCB) function. Figure 8.9 shows the invariant mass distribution and corresponding fit for the inclusive signal samples. The signal shapes observed after the ML-based selection are very similar to those obtained with the nominal selection, with the primary effect being a change in normalisation rather than in shape. The fit demonstrates proper agreement with the simulated distribution for this study, and the ratio between the distribution and the fitted model remains close to unity across the full mass range. This indicates that the ML-based lepton selection does not distort the signal mass shape and that the nominal signal modelling remains valid.

The signal fits for each category are shown in Appendix C.3. In all categories, the fitted models provide a good description of the corresponding mass distributions. Categories characterised by a boosted Higgs boson topology, such as *VBFT* and *HRelpT*, exhibit a narrower mass resolution, quantified by a smaller Crystal Ball width σ_{CB}, compared to the other categories. This behaviour is expected, as higher-momentum particles are generally reconstructed and calibrated with better precision[1].

For the background, we do not perform a category-by-category optimisation of the functional forms. Instead, we adopt the functions determined in the nominal analysis, as reported

[1]At very high muon momenta, the resolution can degrade, since the momentum measurement relies predominantly on track curvature, which becomes increasingly small for high-energy muons.



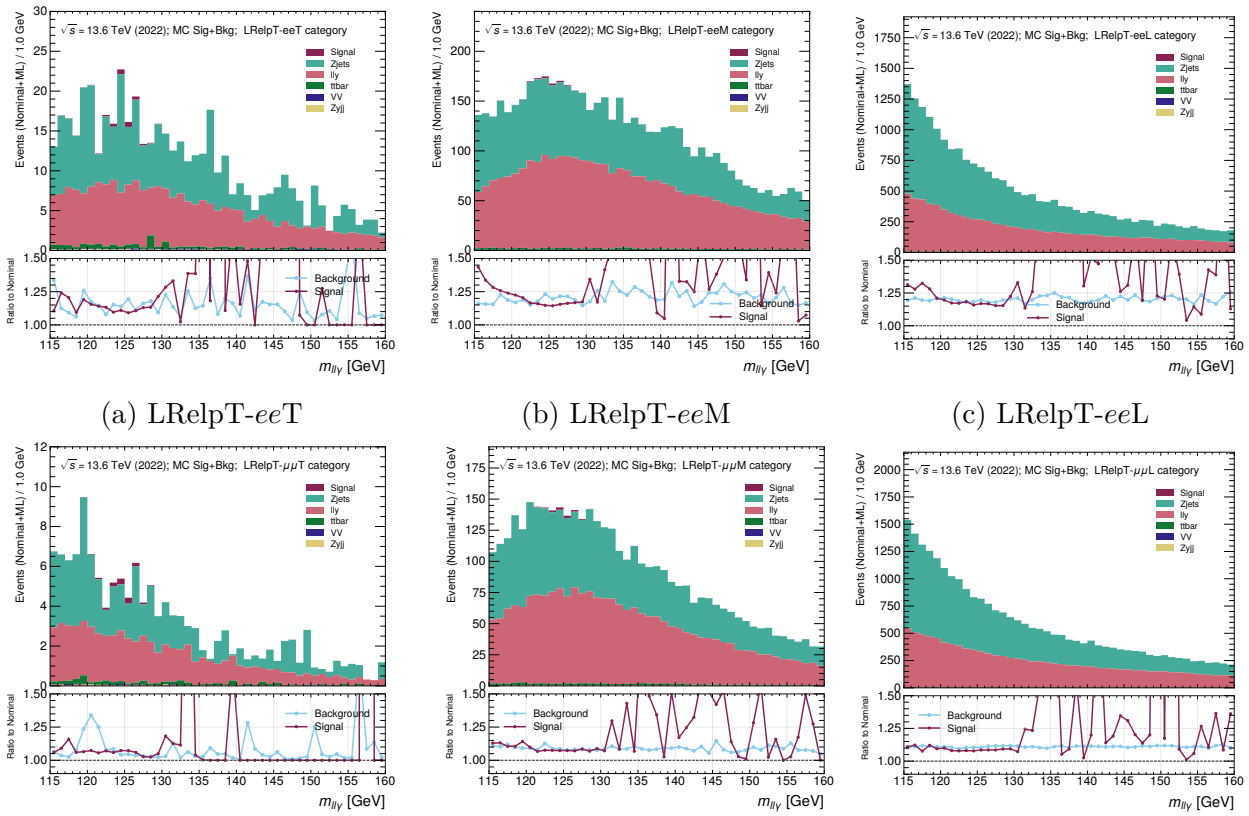
(a) Lepton (b) VBFT (c) VBFL

(d) HRelpT-eeT (e) HRelpT-eeL (f) HRelpT-$\mu\mu$T

(g) HRelpT-$\mu\mu$L

Figure 8.7: Invariant mass distributions of nominal plus ML selected MC signal and background for the *Lepton*, *VBF* and *HRelpT* categories. The lower panel shows the bin-by-bin ratio to the only nominal selection, combining the individual backgrounds into a single background.

in the auxiliary material of Ref. [20], but using a uniform fit window of [115,160] GeV for all categories. The functions considered include: exponential, FK0, second-, third-, and



(a) LRelpT-eeT (b) LRelpT-eeM (c) LRelpT-eeL

(d) LRelpT-$\mu\mu$T (e) LRelpT-$\mu\mu$M (f) LRelpT-$\mu\mu$L

Figure 8.8: Invariant mass distributions of nominal plus ML selected MC signal and background for the *LRelpT* categories. The lower panel shows the bin-by-bin ratio to the only nominal selection, combining the individual backgrounds into a single background.

fourth-order exponential polynomials, and fourth- and fifth-order Bernstein polynomials.

Figure 8.10 shows the weighted event distributions for the *Lepton*, *VBFT*, and *VBFL* categories after applying the nominal plus ML selection, together with the fit probability. The Z+jets sample exhibits sizeable statistical fluctuations across all three categories. In the *Lepton* category, the diboson background is significant, but Z+jets and non-resonant $Z\gamma$ remain the dominant contributions. The $t\bar{t}$ background is small but present throughout the spectrum. In the *VBFT* category, the background is minimal and dominated by EW $Z\gamma jj$, although statistical fluctuations in the Z+jets sample produce visible contributions in some bins. The non-resonant $Z\gamma$ background contributes significantly near the Higgs mass. For the *VBFL* category, the EW $Z\gamma jj$, non-resonant $Z\gamma$, and Z+jets backgrounds provide comparable contributions, with small $t\bar{t}$ contributions in a few bins.

Figure 8.11 shows the invariant mass distributions and fits for the *HRelpT* categories, with the electron and muon channels separated, and each further divided into Tight and Loose subcategories. Comparing the Tight selections among them, as well as the Loose selections, we observe that they follow a similar overall trend, with the *HRelpT-*$\mu\mu$*T* category exhibiting

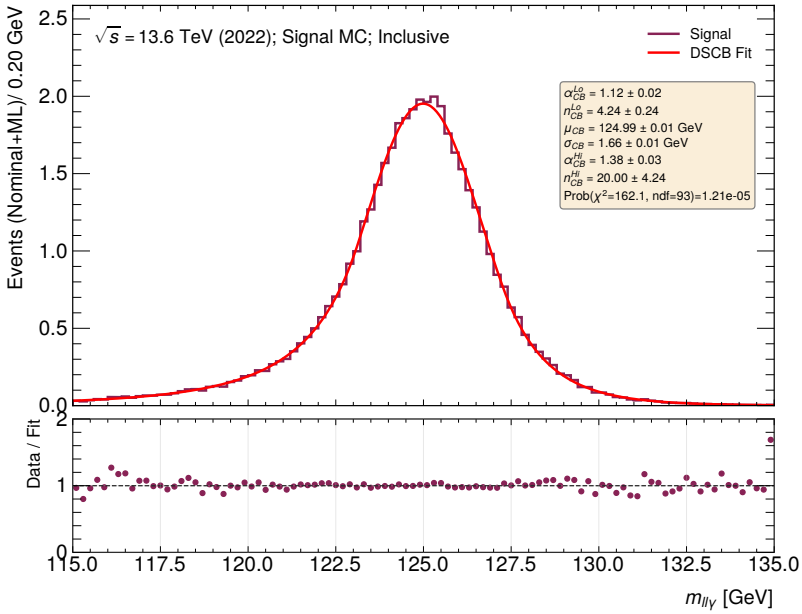


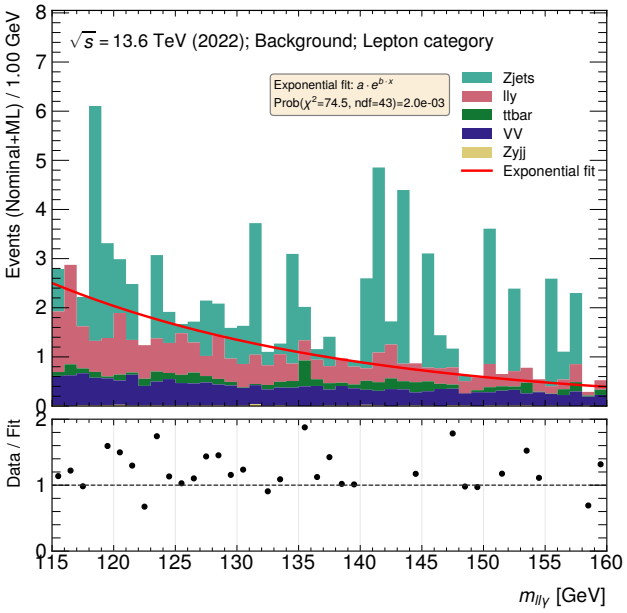
Figure 8.9: Invariant mass distribution of the $\ell\ell\gamma$ system for the inclusive 2022 signal MC sample, shown in purple. The distribution is fitted with a double-sided Crystal Ball function, shown in red, with the best-fit parameters, χ^2 value, and corresponding fit probability reported in the inset. The lower panel shows the bin-by-bin ratio of the distribution to the fit.

a higher yield than its electron counterpart. The dominant background contribution is from non-resonant $Z\gamma$, while Z+jets are more prominent in the Loose categories and appear reduced but sharply peaked in the Tight categories. The $t\bar{t}$ background remains small but is present throughout the spectrum.

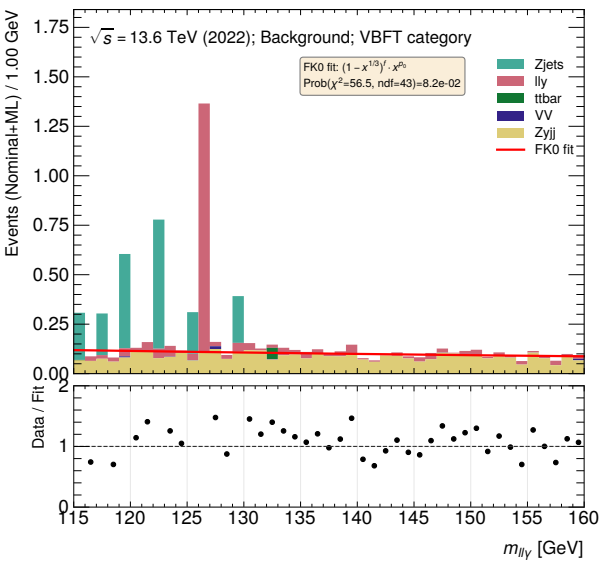
Figure 8.12 shows the invariant mass distributions and fits for the $LRelpT$ categories, with the electron and muon channels separated and each further divided into Tight, Medium, and Loose subcategories. These categories contain the largest number of background events passing our selection. We observe that the two Tight have quite different shapes, while Medium and Loose selections resemble each other. The dominant backgrounds are non-resonant $Z\gamma$ and Z+jets, which contribute at roughly the same magnitude, consistent with the results obtained using the nominal data-driven method to estimate the purity and relative fractions of these two backgrounds.

8.6 Projections on sensitivity

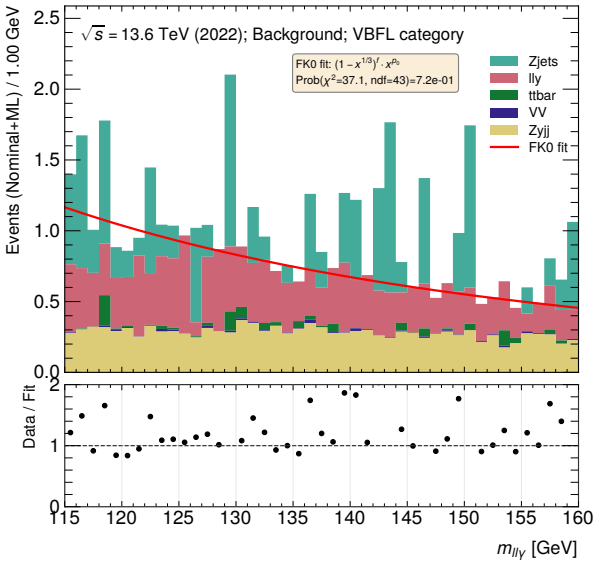
From these fits, we determine the mass window containing 68% of the expected signal in each category, w_{68}, together with the corresponding expected signal yield S_{68} within this window. Using the same window, we evaluate the expected background yield B_{68} for each category. We obtain the signal uncertainties by propagating the fitted model parameter



(a) Lepton

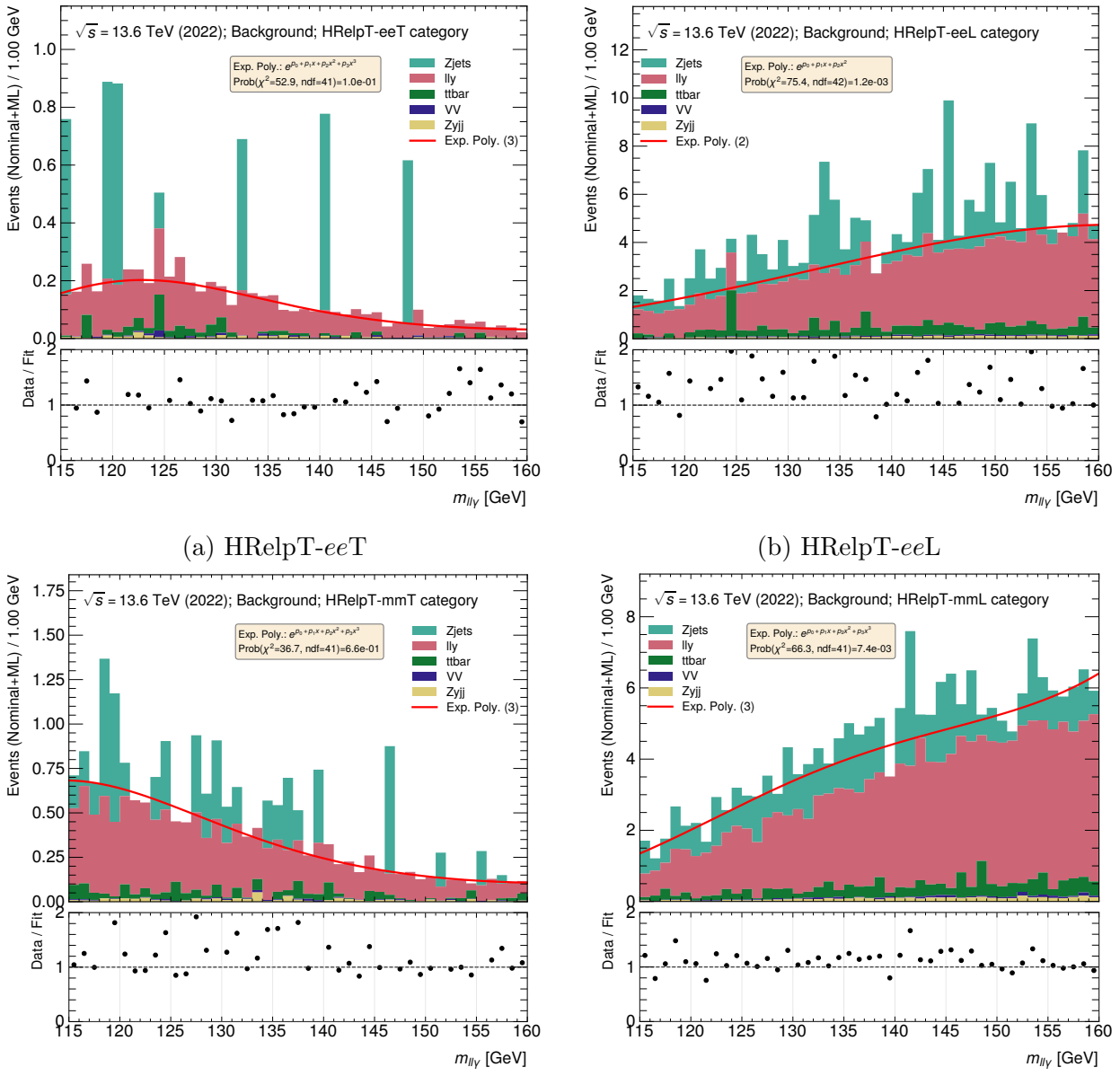


(b) VBFT



(c) VBFL

Figure 8.10: Invariant mass distribution of nominal plus ML selected MC background for the (a) *Lepton*, (b) *VBFT*, and (c) *VBFL* categories. The distributions are fitted with the best-fit function in red: an exponential for the *Lepton* category and an FK0 for the *VBF* categories. The fit function and the fit probability are reported in the inset. The lower panel shows the bin-by-bin ratio of the distribution to the fit.



(a) HRelpT-eeT

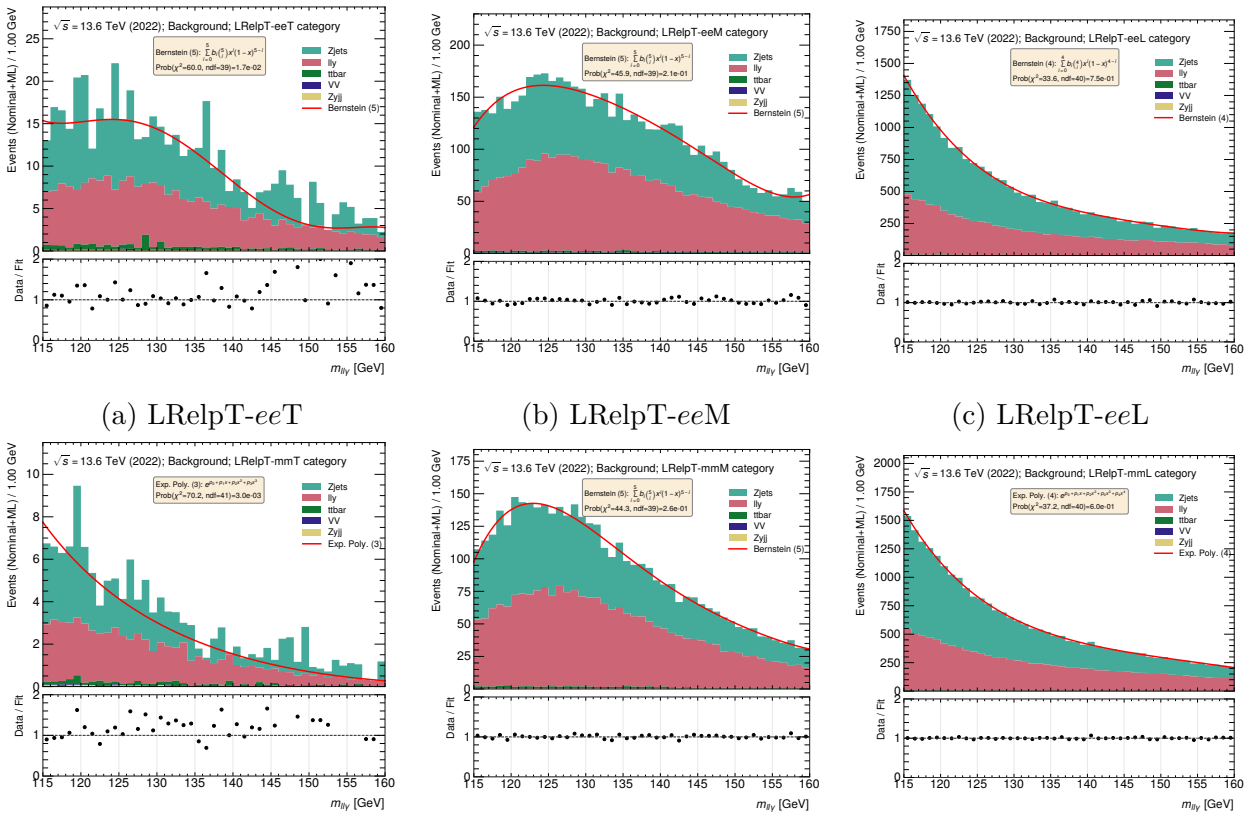
(b) HRelpT-eeL

(c) HRelpT-$\mu\mu$T

(d) HRelpT-$\mu\mu$L

Figure 8.11: Invariant mass distributions of nominal plus ML selected MC background for the (a) *HRelpT-eeT*, (b) *HRelpT-eeL*, (c) *HRelpT-μμT*, and (d) *HRelpT-μμL* categories. The distributions are fitted with the best-fit second- and third-order exponential polynomial functions shown in red. The fit function and the fit probability are reported in the inset. The lower panel displays the bin-by-bin ratio of the distribution to the fit.

uncertainties through pseudo-experiments, ensuring they reflect the statistical component of the fit. Background uncertainties are determined either from the propagation of the fitted background model parameter uncertainties through the integral over the signal region, or



(a) LRelpT-eeT (b) LRelpT-eeM (c) LRelpT-eeL

(d) LRelpT-$\mu\mu$T (e) LRelpT-$\mu\mu$M (f) LRelpT-$\mu\mu$L

Figure 8.12: Invariant mass distributions of nominal plus ML selected MC background for the *LRelpT* categories, with electron and muon channels each divided into Tight (T), Medium (M), and Loose (L) subcategories. The distributions are fitted with the best-fit third- and fourth-order exponential polynomials and fourth- and fifth-order Bernstein polynomials, shown in red. The fit function and the fit probability are reported in the inset. The lower panel shows the bin-by-bin ratio of the distribution to the fit.

via bootstrap resampling when the fit covariance matrix is ill-conditioned.

We then compute the sensitivity from these quantities according to Eq. 8.2. Since the categories are mutually exclusive, we obtain the inclusive sensitivity by combining the individual category sensitivities in quadrature.

Table 8.2 shows the sensitivities obtained by applying the ML-based selection on top of the nominal selection, after projecting the 2022 MC distributions to an integrated luminosity of 165 fb^{-1}. This projection enables a direct comparison with the nominal sensitivity reported in Table 7.3 for the partial Run 3 dataset. Appendix C.4 provides the sensitivities derived directly from the 2022 MC samples only.

Although the *VBFT* category loses some of its previous significance, it remains the most sensitive category under the new selection. The small improvements observed across most other categories combine to yield an inclusive expected sensitivity of 2.01. These results,

while promising, are based on 2022 MC data only and should be interpreted with caution, as statistical fluctuations remain significant due to the limited sample size.

Category	w_{68} [GeV]	S_{68}^{exp}	B_{68}^{exp}	$S_{68}^{\text{exp}}/\sqrt{S_{68}^{\text{exp}}+B_{68}^{\text{exp}}}$
Lepton	3.93	1.57 ± 0.03	37 ± 2	0.25
VBFT	3.64	1.80 ± 0.06	2.26 ± 0.11	0.89
VBFL	3.68	3.14 ± 0.10	19.2 ± 0.9	0.66
HRelpT-eeT	3.02	1.26 ± 0.07	3.4 ± 2.7	0.58
HRelpT-eeL	3.51	3.15 ± 0.12	43 ± 2	0.46
HRelpT-$\mu\mu$T	3.55	2.51 ± 0.11	11 ± 3	0.69
HRelpT-$\mu\mu$L	3.68	2.55 ± 0.08	56 ± 8	0.33
LRelpT-eeT	3.67	10.7 ± 0.3	4319 ± 16	0.59
LRelpT-eeM	3.88	35.5 ± 0.7	3 517 ± 56	0.59
LRelpT-eeL	4.37	30.5 ± 1.1	17 390 ± 154	0.23
LRelpT-$\mu\mu$T	3.70	5.23 ± 0.16	87 ± 16	0.55
LRelpT-$\mu\mu$M	3.87	37.1 ± 0.5	3 059 ± 34	0.67
LRelpT-$\mu\mu$L	4.10	40.6 ± 0.6	19 426 ± 215	0.29
Inclusive	4.01	176.5 ± 1.2	43 971 ± 509	2.01

Table 8.2: Expected sensitivity projection from 2022 MC to an integrated luminosity of 165 fb^{-1} after ML Z selection. The expected signal events (S_{68}^{exp}) and background events (B_{68}^{exp}) are obtained by integrating with the fit in a mass window of width w_{68}, defined to contain 68% of the signal. The signal uncertainties represent the statistical uncertainty on the yield in the signal region, obtained by propagating the fitted model parameter uncertainties via pseudo-experiments. The background uncertainties represent the uncertainty on the estimated background yield in the signal region from the fitted background model. For the inclusive case, the sensitivity is obtained by combining the individual category sensitivities in quadrature.

8.7 Projections on signal strength uncertainties

The uncertainty in the signal strength is derived from a profile likelihood scan of the Higgs signal strength parameter. Systematic uncertainties are evaluated by varying each nuisance parameter by $\pm 1\sigma$ with respect to its nominal value and propagating the resulting variations to the fitted signal strength. The statistical uncertainty is then obtained from the total uncertainty, defined by the points where the profile likelihood crosses the 1σ threshold, after subtracting the total systematic uncertainty in quadrature.

Due to the limited scope of this study, we do not perform a full re-evaluation of all systematic uncertainties or their propagation through the analysis. For the ML-based lepton selection, we expect changes in the statistical uncertainty, while modifications to the electron- and muon-related systematic uncertainties may also arise.

This section outlines the proposed methodology for deriving scale factors and assessing the energy scale, resolution, and efficiency associated with ML-based lepton selection. Based on this procedure, we present qualitative projections to illustrate the potential impact of the new selection on the overall uncertainty, with particular emphasis on the expected improvement in the statistical component.

In general, electron and muon calibration and efficiency corrections are obtained by reconstructing leptonic Z boson decays using a tag-and-probe method, which allows the lepton reconstruction, identification, and calibration efficiencies to be measured independently for each lepton [108–110].

We could use a similar strategy for the ML-based selection by comparing the reconstructed Z boson mass peak in simulation and data for both the nominal working point and the ML approach. We first measure the efficiency of the nominal selection in simulation using truth-matched Z boson decays. We then obtain the corresponding efficiencies in data by applying the nominal scale factors provided by the CP groups.

For the ML-based selection, efficiencies can be evaluated analogously in simulation and data. By comparing yields obtained with the nominal and ML selections, we can derive a global scale factor for the ML-based lepton selection. This approach relies on nominal calibration and scale factors as inputs and would likely require assigning additional systematic uncertainties to account for the modified selection.

A full calibration of the ML-based lepton selection could be performed in a dedicated future study. The precision of such a calibration could improve the standard one by developing a dedicated Z boson tagger, enabling calibration at the dilepton level rather than through the deconvolution of individual leptons.

Table 8.3 shows the expected uncertainty on the signal strength reported for the nominal Run 3. We have added a column which shows the projection and the strength of improving statistics by the ML Z selection. The electron and muon efficiencies have been updated to reflect the new selection. For this study, a conservative inflation factor of three is applied to these uncertainties. All other systematic uncertainties remain unchanged.

To estimate the projected improvement in the statistical uncertainty on the signal strength, we consider an approximate relation between the signal strength uncertainty and the expected statistical significance: $\Delta\mu \propto 1/Z$, where Z is the significance of the process.

The signal strength is defined as

$$\mu = \frac{N_{\text{sig}}}{N_{\text{sig}}^{\text{SM}}}\,, \tag{8.3}$$

where N_{sig} denotes the signal yield inferred from a signal-plus-background model.

We assume the relation

$$\Delta\mu \propto \frac{1}{Z} = \frac{\sqrt{S+B}}{S}\,, \tag{8.4}$$

and use it solely as a relative estimator of the statistical sensitivity; it does not replace a full profile-likelihood treatment.

We treat the ratio of uncertainties between the two methods as a consistent and reliable quantity for comparison. Therefore,

$$\frac{\Delta\mu^{\text{Nom.}}_{\text{stat.}}}{\Delta\mu^{\text{ML}}_{\text{stat.}}} \simeq \frac{Z_{\text{ML}}}{Z_{\text{Nom.}}} = \frac{2.01}{1.81} \simeq 1.11\,, \tag{8.5}$$

where the expected sensitivity Z_{ML} and $Z_{\text{Nom.}}$ are retrieved from Tab. 8.2 and 7.3 respectively.

Given that $\Delta\mu^{\text{Nom.}}_{\text{stat.}} = 0.68$, the projected uncertainty for the ML selection is

$$\Delta\mu^{\text{ML}}_{\text{stat.}} \simeq \frac{\Delta\mu^{\text{Nom.}}_{\text{stat.}}}{1.11} \simeq 0.61\,. \tag{8.6}$$

Propagating these numbers in Tab. 8.3, yields a total uncertainty of 0.64, corresponding to a 9% reduction compared to the nominal selection, even with the conservative overestimation of the electron and muon efficiencies.

Uncertainty source	$\Delta\mu$	
	Expected	**ML projection**
Statistical uncertainty	0.68	**0.61**
Systematic uncertainty	0.16	**0.18**
Spurious signal (background modelling)	0.11	0.11
QCD scale, PDF+α_S, parton shower	0.09	0.09
Branching ratio ($H \to Z\gamma$)	0.05	0.05
Luminosity	0.03	0.03
Photon efficiency	0.02	0.02
Jet	0.02	0.02
Electron and photon energy scale and resolution	0.02	**0.06**
Electron efficiency	0.02	**0.06**
Muon	< 0.01	**0.03**
Trigger	< 0.01	< 0.01
Total	0.70	**0.64**

Table 8.3: Symmetrised impact of the individual sources of statistical and systematic uncertainty on the expected strengths using the nominal and a projection adding the ML selection. The numbers being projected in the ML case are in bold. Table adapted from Ref. [20].

8.8 Conclusions and outlook

In this study, we demonstrate through simulation the potential to replace the nominal cut-based lepton selection with a machine learning approach that identifies Z bosons from lepton pairs. We select the optimal working point of the ML-based Z score by maximising the expected Z boson significance, and propagate the resulting selection through the remainder of the nominal analysis chain. We then study the impact of the updated selection on the

event yields across all categories, verifying that no uncontrolled increase in the Z boson background is introduced. We extract signal and background yields in each category using binned χ^2 fits, performed independently for each category and using the same functional forms as in the nominal analysis, which was validated through spurious signal tests.

We evaluate the expected sensitivity to the $H \to Z\gamma$ signal by defining, for each category, a mass window that contains 68% of the signal and combining the corresponding sensitivities in quadrature. Because the available MC samples are limited to the 2022 dataset, we project the results to the integrated luminosity of the nominal analysis, 165 fb^{-1}, to enable a fair comparison. Under this projection, we observe an improvement in the total expected sensitivity to 2.01, compared with the 1.81 reported in the published analysis.

We explore the relationship between the signal strength uncertainty and the statistical sensitivity to quantify the impact of the ML-based selection. Using this approach, we expect the statistical uncertainty on the signal strength to improve from $\Delta\mu^{\text{Nom.}}_{\text{stats.}} = 0.68$ to $\Delta\mu^{\text{ML}}_{\text{stats.}} = 0.61$. Even under a conservative scenario in which the electron and muon scale factor uncertainties are increased by a factor of three, we still observe a reduction in the total uncertainty on the signal strength, from $\Delta\mu^{\text{Nom.}} = 0.70$ to $\Delta\mu^{\text{ML}} = 0.64$. We emphasise that these results are based on projections and do not replace a full evaluation using a profile likelihood fit.

Currently, several strategies are being explored to address weaknesses in the previous analysis, to improve sensitivity, and reduce uncertainties for the full Run 3 results. As part of this project, we are developing a data-driven procedure to derive precise scale factors for the ML-based lepton selection, enabling the assignment of concrete systematic uncertainties rather than relying on projections.

Simultaneously, efforts are underway to prepare a large public dataset containing Z+jets in both MC and data. Encouraging the community to address the domain shift between MC and data could significantly improve object reconstruction and calibration. A well-calibrated, fully reconstructed Z boson would benefit all analyses involving Z bosons and could enhance the resolution of Higgs signals, analogous to how improvements in $H \to b\bar{b}$ resolution motivated our work on b-jet calibration.

Future work could extend the approach explored in this chapter to photons. Unlike leptons, photons do not originate from a narrow resonance, and photons from final-state radiation (e.g., in $Z\gamma$ events) often have low transverse momentum, making a calibration across a broad p_{T} range more challenging. One potential strategy is to exploit the similarities between electrons and photons in the electromagnetic calorimeter, as both produce similar shower shapes, although photons typically penetrate slightly deeper, on average 9/7, into the calorimeter. By shaping electron showers to mimic photon signatures, it might be possible to reconstruct a Z boson peak and find its scale factors.

Chapter 9

Summary and Discussion

During this thesis, we have developed and applied ML techniques in two central areas of experimental particle physics: object calibration and physics analysis. The primary objective has been to enhance Higgs boson searches, with a particular focus on improving b-jet calibration and increasing the sensitivity of the rare $H \to Z\gamma$ decay channel.

The first part of this work focused on jet calibration, aiming to improve the energy resolution of b-jets using ML techniques. A hybrid training strategy combining simulated and collision data was investigated to reduce the domain shift and improve the jet energy resolution. As an initial step, a suitable data-driven training target was defined using event topologies in which a well-calibrated reference object recoils against the leading jet. Z+jets samples were first employed, as they provide a kinematic regime comparable to that of b-jets originating from Higgs boson decays. Two approaches were then explored: a BDT, used as a proof of concept, and a transformer architecture capable of exploiting detailed detector-level information, such as jet constituents.

Training solely on MC simulation already led to a significant improvement in jet energy resolution compared to the nominal calibration, with slightly better performance from the BDT. The transformer, while promising, proved computationally intensive and difficult to train. When moving to hybrid training with both simulation and real data, the algorithm did not converge sufficiently toward the target derived from data. This was primarily due to the imbalance between the large number of simulation events and the limited data, leading to underrepresentation of data labels in the loss function.

We observed that the resolution scale in data was, in general, larger than that observed in simulation, motivating an initial attempt to deconvolve the measured resolution into its detector component. A first approach based on subtraction in quadrature was tested in simulation, providing a better estimate of the detector resolution than the directly observed resolution. This method was then applied to collision data, yielding results consistent with simulation, although the outcome proved noisy depending on the regression method used. These initial in-situ studies served as the foundation for subsequent investigations.

Following this, a transformer trained exclusively on simulation demonstrated substantial improvements in both the b-jets energy scale and resolution. Building on the data-driven resolution studies expertise developed in the previous project, we investigated whether an

additional in-situ calibration would be required in addition to the nominal in-situ calibration for this new MC-based approach. For the first time, Z+jets and γ+jets samples were employed in combination with a deconvolution method, assuming a Gaussian detector resolution, to extract the JER in collision data. At the time of this work, the transformer-based calibration `bJR4` was not yet available for application to real collision data. Preliminary results obtained using the full nominal calibration indicate compatibility between the inclusive in-situ calibration and the b-jet-specific component, suggesting a generally robust description of the in-situ calibration and, at the same time, the presence of still large uncertainties.

Moving to the explicit analysis work, the final project applied ML techniques to improve the selection of Z bosons, thereby enhancing the statistical sensitivity of the $H \to Z\gamma$ analysis. Separate BDT models were trained for electron and muon channels, combining particle identification and isolation information to estimate the probability that a lepton pair originates from a Z boson. A significance-based optimisation was then performed to determine the optimal selection threshold for Z candidates.

This improved selection was applied to the 2022 simulation, with subsequent evaluation using the nominal event categorisation and fitting functions from the partial Run 3 analysis. Projections to the integrated luminosity of the same period were performed to compare the expected sensitivity of the $H \to Z\gamma$ analysis with and without the ML-based selection. The projected sensitivity increased from 1.81 for the nominal selection to a projected 2.01 using the ML-enhanced Z selection.

In terms of signal strength uncertainty, the statistical component improved from $\Delta\mu^{\text{Nom.}}_{\text{stats.}} = 0.68$ to $\Delta\mu^{\text{ML}}_{\text{stats.}} = 0.61$. Systematic uncertainties related to leptons were conservatively inflated by a factor of three to account for uncomputed scale factors. Combining these contributions, the total uncertainty decreased from $\Delta\mu^{\text{Nom.}} = 0.70$ to $\Delta\mu^{\text{ML}} = 0.64$, corresponding to a 9% reduction.

Overall, this thesis demonstrates the potential of ML techniques to enhance both jet calibration and event selection in Higgs boson analyses. The b-jet studies explored the feasibility of hybrid training to reduce the domain shift between simulation and data. The results of an MC-only ML calibration were presented, along with studies assessing the need for an additional in-situ correction. At the same time, the $H \to Z\gamma$ analysis shows that a targeted ML-based Z selection can increase signal yield and reduce uncertainties. Together, these results highlight the practical impact of advanced computational methods on precision measurements and provide a foundation for future improvements and broader applications within the ATLAS experiment.

The results of this thesis point towards several concrete next steps, both for consolidating the methods developed here and for extending their impact within ATLAS. In the near term, the highest priority is completing the in-situ calibration studies for the `bJR4` model. Finalising this calibration would make the model available for physics analyses, improving b-jet energy resolution and enhancing the sensitivity of a wide range of measurements, including $H \to b\bar{b}$ decays and di-Higgs production where one of the Higgs bosons decays to $b\bar{b}$. In parallel, a complete propagation of the ML-based Z selection through the $H \to Z\gamma$ analysis chain using a profile likelihood approach is foreseen. This would enable a more complete and robust evaluation of the expected signal strength sensitivity. If the projected improvements

are confirmed, the ML-based Z selection is expected to become the primary choice in the next full Run 3 $H \to Z\gamma$ analysis.

Beyond these immediate goals, revisiting the hybrid training strategy remains a promising direction. If successful, such an approach could be generalised to reconstructed objects beyond jets. One of the main limitations encountered in this thesis was the simultaneous use of simulation and collision data during training without an appropriate weighting scheme, leading to the dominance of simulation statistics. A potential improvement would involve a two-step training procedure: first, training the model to regress towards truth-level quantities in simulation, allowing it to learn the correlations among input variables; followed by a fine-tuning stage using collision data to adapt the learned representation and reduce residual discrepancies.

More broadly, addressing the simulation-data domain shift remains an open and active area of research. In this context, a combined simulation and collision dataset from Run 2 has recently been released to the general public [171]. Work is ongoing to prepare a derived dataset containing low-level particle information suitable for ML applications from the same data. Such a dataset would enable the wider community to perform studies of domain adaptation techniques and facilitate the development of object-level taggers, such as a dedicated leptonic Z boson tagger. This would allow analyses involving Z bosons to operate directly at the dilepton level, rather than relying on independent single-lepton selections, potentially improving performance across analyses.

With the large amount of data already delivered to ATLAS during Run 3, exceeding a total integrated luminosity of 0.5 ab^{-1}, when combined with previous runs, and with the High-Luminosity LHC scheduled to begin operation in the early 2030s, the field is entering an era of unprecedented statistical power. This transition opens the door to increasingly precise measurements and the study of rare processes previously inaccessible. Building on these opportunities, continued advances in reconstruction, calibration, and analysis techniques, such as those explored in this thesis, will be essential for fully exploiting the available data and, ultimately, deepening our understanding of the fundamental physics underlying the Universe.

Appendices

Appendix A

Hybrid training

This Appendix provides additional material supporting Chapter 5.

A.1 Postprocess cutflow

After the ntuples are produced with AnaHelpers, a series of postprocessing selections is applied to refine the event sample. In Table A.1, a list of each of these postprocessing cuts is shown together with the efficiency for a $Z(\mu\mu)$+jets MC sample. Similar selections are

Selection step	Efficiency [%]
Initial sample	100
Two truth leptons	99.98
Truth jet ≥ 1	76.45
Opposite-charge muons	75.66
Jet–muon overlap removal	75.66
$66 < m_{\mu\mu} < 116$ GeV	73.34
$\lvert\Delta\phi(Z, \mathrm{j}_1)\rvert > 2.9$	19.72
Subleading jet veto	6.22
Jet truth matching	5.55

Table A.1: Cutflow for the $Z(\mu\mu)$+jets Monte Carlo sample after applying the physics selection with the total efficiency. The quoted efficiencies are cumulative.

applied in Fig. A.1, which shows the cutflow for a subset of the MC and real data samples.

A.2 Sample validation

This subsection presents a set of distributions used to validate the MC and data samples employed in the analysis.

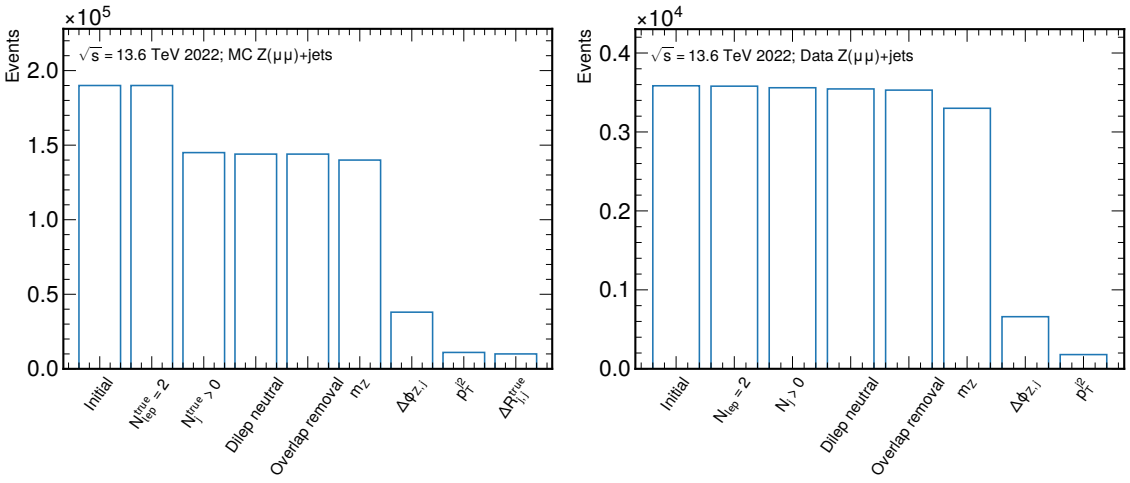


Figure A.1: Cutflow of the selection cuts applied to the ntuple with a subset of the dataset for MC (left) and real data (right).

Figure A.2a shows the azimuthal separation $\Delta\phi$ between the leading jet and subleading jets in a $Z(\mu\mu)$+jets MC sample. A higher density of entries is observed around $\Delta\phi \simeq 0.5$, indicating that subleading jets tend to be more aligned with the leading jet than with the reference object. Such configurations can lead to an overestimation of the constructed reference momentum, $p_{\mathrm{T}}^{\mathrm{ref}}$.

The leading jet p_{T} distribution in Fig. A.2b exhibits a depletion around 40 GeV, which is even more pronounced in the p_{T}^{Z} distribution. This behaviour originates from the subleading jet veto, which requires the subleading jet to have $p_{\mathrm{T}} < 12$ GeV until the leading jet momentum exceeds 40 GeV.

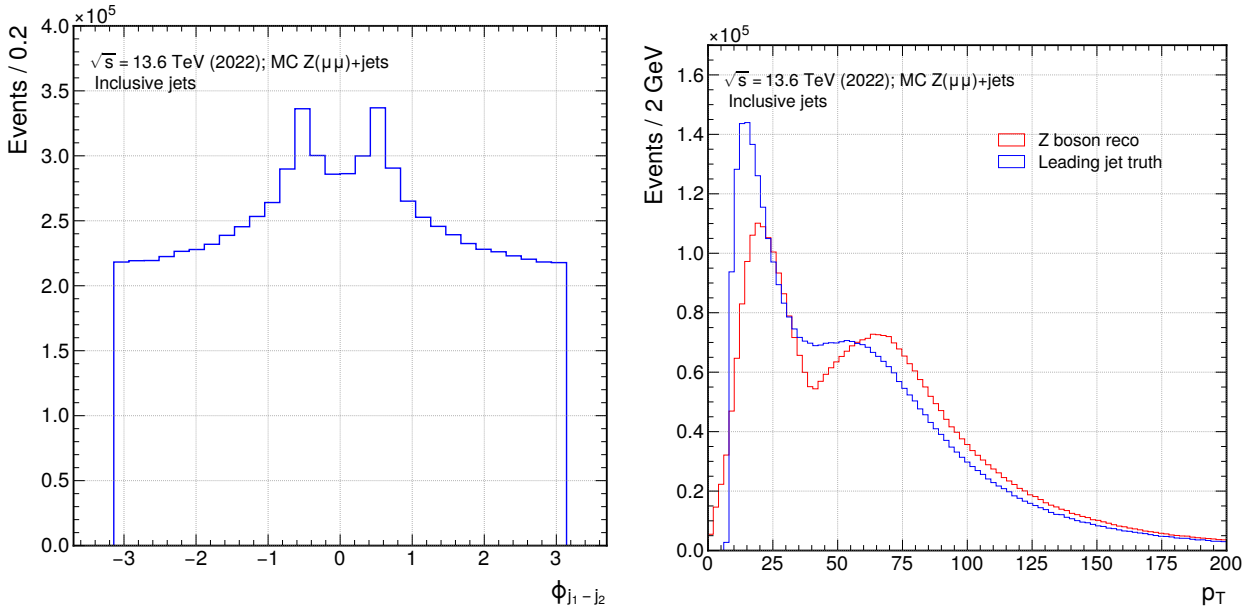
Figure A.3 compares the distribution of the number of reconstructed jets per event between simulation and real data after the full event selection. The overall agreement indicates a consistent modelling of jet multiplicities between data and simulation.

A.3 DAOD Containers

Table 5.3 summarises the variables used as inputs to the ML regression and for event validation. In this subsection, we specify the data containers from which these variables are extracted and motivate their inclusion.

With Athena Release 22, calorimeter cell-level information is no longer stored, which limits access to fine-grained energy deposition patterns. While `JETM3` does not provide explicit jet constituents, it does include the GlobalNeutralParticleFlowObjects and GlobalChargedParticleFlowObjects containers. These particle-flow (PFlow) objects are matched to the leading jet within $\Delta R < 0.5$ to obtain an approximate set of jet constituents suitable for use as ML inputs.

Future dataset versions (newer p-tags) may reduce the availability of these containers, in



(a) (b)

Figure A.2: (a) Azimuthal separation between the leading jet and subleading jets in a Monte Carlo sample of $Z(\mu\mu)$+jets events. (b) Momentum distributions of the reference object Z boson (green) and the truth leading jet (blue).

which case a transition to track-based quantities would be required. For the present work, however, the PFlow information provides the only available constituent-level information and is therefore retained.

The variables are sourced from the following containers:

- Track-level quantities from InDetTrackParticles,
- Reconstructed jets from AntiKt4EMPFlowJets,
- Jet constituents from GlobalNeutralParticleFlowObjects and GlobalChargedParticleFlowObjects,
- Reconstructed muons from Muons, and
- The truth jet p_{T} target from AntiKt4TruthDressedWZJets, which provides a more accurate representation of the b-jet momentum.

Truth jet flavour labelling follows the FTag group recommendation using HadronConeExclTruthLabelID.

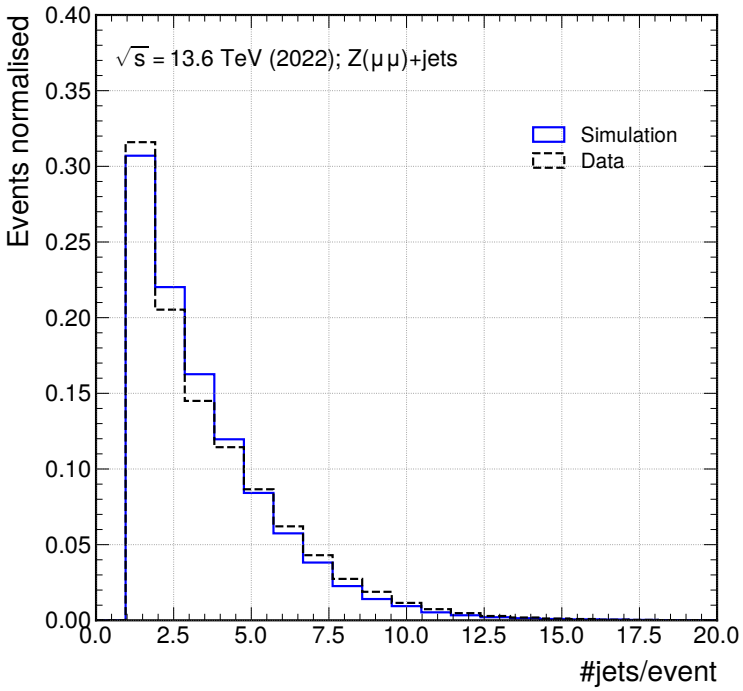


Figure A.3: Normalised distribution of the number of jets per event in MC (solid blue) and real data (dashed black).

A.4 Removing JVT variables

Since the JVT variables are computed from the final stages of the jet calibration, their inclusion can produce a biased regression. Its impact on the regression performance is studied by removing it from the input feature set.

This cross-check is important to ensure the model remains independent of other calibration steps. The three JVT-related variables removed, as defined in Table 5.3, are: Jvt, JvtRpt, and JvtJvfcorr.

Figure A.4 compares the resolution obtained with and without the JVT variables in the BDT model. As expected, the resolution degrades slightly when these variables are removed, but the effect is less than 3%. This highlights the trade-off between achieving marginally better resolution and maintaining independence from other calibration procedures.

A.5 Resolution studies: analytical vs numerical

Some additional considerations arose when determining the optimal label to use, which are described below. The uncertainty associated with the label can be computed from

$$p_{\mathrm{T}}^{\mathrm{label}} = \sum p_{\mathrm{T}}^{\mathrm{rest}} = \cos\phi_{Z/\gamma} \cdot p_{\mathrm{T}}^{Z/\gamma} + \sum \cos\phi_{j_2} \cdot p_{\mathrm{T}}^{j_2} \,, \tag{A.1}$$

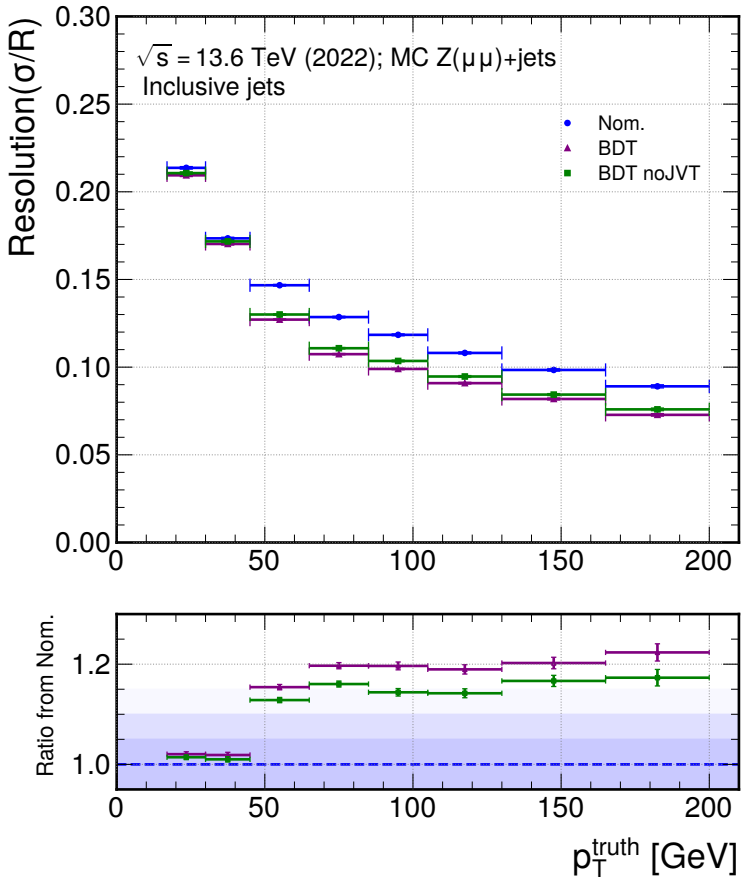


Figure A.4: Resolution of inclusive jets binned by p_{T} in MC samples, comparing the nominal ATLAS calibration (blue dots), a BDT regression with all variables (purple triangles), and a BDT regression without JVT variables (green squares). The bottom panel shows the corresponding ratios from the nominal calibration.

and, assuming the uncertainties are uncorrelated, standard error propagation can be applied

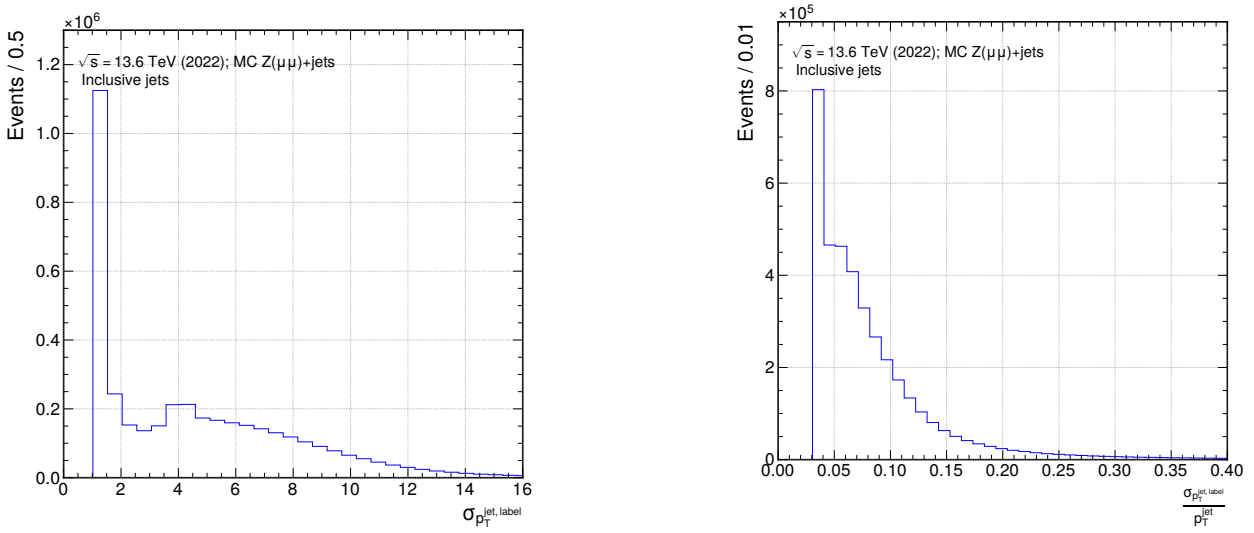
$$\sigma_{p_{\mathrm{T}}^{label}} = \sqrt{\left(\cos\phi_Z \cdot \sigma_{p_{\mathrm{T}}^Z}\right)^2 + \sum \left(\cos\phi_{j_2} \cdot \sigma_{p_{\mathrm{T}}^{j_2}}\right)^2}. \tag{A.2}$$

This expression depends on the uncertainty of the subleading jets momentum $\sigma_{p_{\mathrm{T}}^{j_2}}$, and the Z boson momentum $\sigma_{p_{\mathrm{T}}^Z}$. We can try to parametrise these uncertainties. These uncertainties can be parametrised as functions of p_{T}. The observed uncertainty for the Z boson is relatively constant across the p_{T} range, and fitting it with a power-law function yields $\sigma_{p_{\mathrm{T}}^Z} = \frac{3.32}{p_{\mathrm{T}}} - \frac{0.90}{\sqrt{p_{\mathrm{T}}}} + 0.09$. For the subleading jets, a similar power-law fit is performed in bins of truth p_{T}, resulting in $\sigma_{p_{\mathrm{T}}^{j_2}} = \frac{0.01}{p_{\mathrm{T}}} + \frac{0.98}{\sqrt{p_{\mathrm{T}}}} + 0.02$, which is larger than $\sigma_{p_{\mathrm{T}}^Z}$.

Two strategies naturally follow from this study. The first is to impose a cut on the label uncertainty, selecting only events with low uncertainty. For instance, this would preferentially select events where the subleading jets are in the central region of the detector and have

relatively high energy, compared to the broader cuts typically applied. The second approach leverages Monte Carlo (MC) simulation, where the actual uncertainty can be computed directly. In this case, the value obtained from Eq. A.2 can be compared with the label resolution defined as $R_{\text{label}} = p_{\text{T}}^{\text{truth}}/p_{\text{T}}^{\text{label}}$.

After applying the parametrisation, the per-p_{T} uncertainties can be propagated through Eq. A.2. The resulting uncertainty on the label is shown in Fig. A.5a. high-p_{T} jets may exhibit a good relative resolution while still having a large absolute $\sigma_{p_{\text{T}}^{label}}$. A more appropriate comparison can therefore be obtained by normalising the uncertainty to the reconstructed jet momentum, shown in Fig. A.5b. This distribution yields a more continuous distribution.



(a) (b)

Figure A.5: Distribution of (a) the uncertainty of the label, and (b) the uncertainty of the label divided by the $p_{\text{T}}^{\text{label}}$.

Following the first approach, using this relative $\sigma_{p_{\text{T}}^{label}}$, we can remove events where we expect the error to be big, based on the uncertainties of the other objects of the event, so we only use for training those with a low uncertainty. The optimal threshold can be determined by scanning its impact on the $p_{\text{T}}^{\text{label}}$ response; an example with a cut at 0.08 is shown in Fig. A.6. With this requirement, the distribution becomes more centred around zero and narrower compared to both the $X1$ and $X2$ methods; however, the improvement remains smaller than anticipated. However, the event yield is significantly reduced, implying a trade-off between improved systematic behaviour and statistical precision.

Building on the first approach, the second strategy evaluates how well the numerically computed uncertainties reflect the actual label resolution. Figure A.7 compares the uncertainties computed numerically using Eq. A.2 with those obtained from the resolution of the label R_{label}. The purple line indicates equality between the two methods. As observed, the numerical estimate is significantly smaller in the simulation. This difference could be quan-

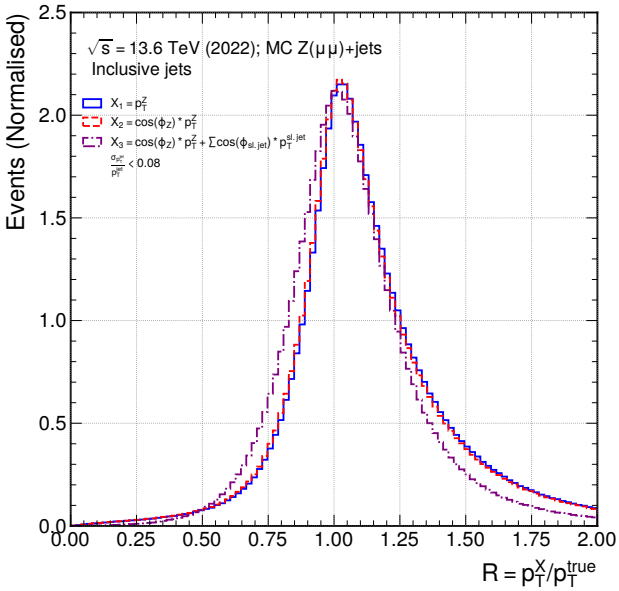


Figure A.6: Normalised response distributions for the different label constructions. The dash–dotted purple curve shows the result after applying the uncertainty cut.

tified in MC, validated using independent samples, and subsequently propagated to collision data to obtain a more realistic estimate of the detector resolution.

These studies were instrumental in understanding how the label is constructed and how its uncertainty is estimated. However, the focus of the project gradually shifted toward clarifying what this resolution represents physically and how it propagates to the jet energy resolution.

A.6 Results for pseudorapidity

In Section 5.5, the performance of two ML algorithms was compared as a function of jet transverse momentum. Here, the differences between the BDT and transformer results and the nominal calibration are examined as a function of pseudorapidity, η, and are shown in Fig. A.8. In all cases, a degradation of performance is observed when moving from MC to data, with a further reduction seen for the hybrid training.

The resolution is best in the central region of the calorimeter as expected, and gradually worsens for forward jets, maintaining approximate symmetry between the two sides.

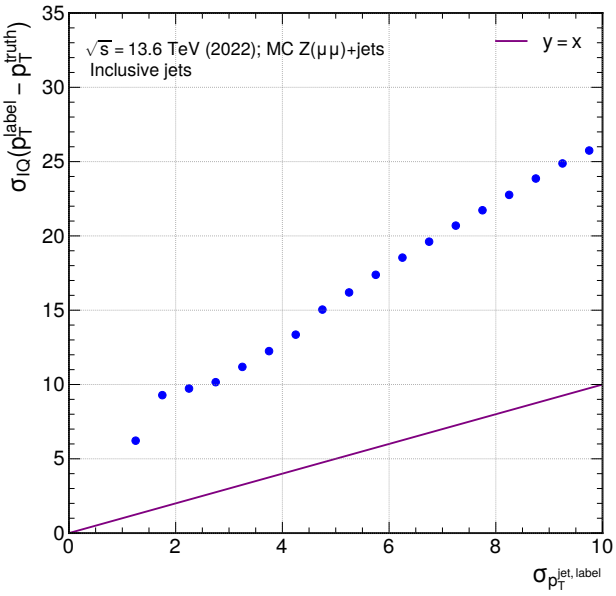
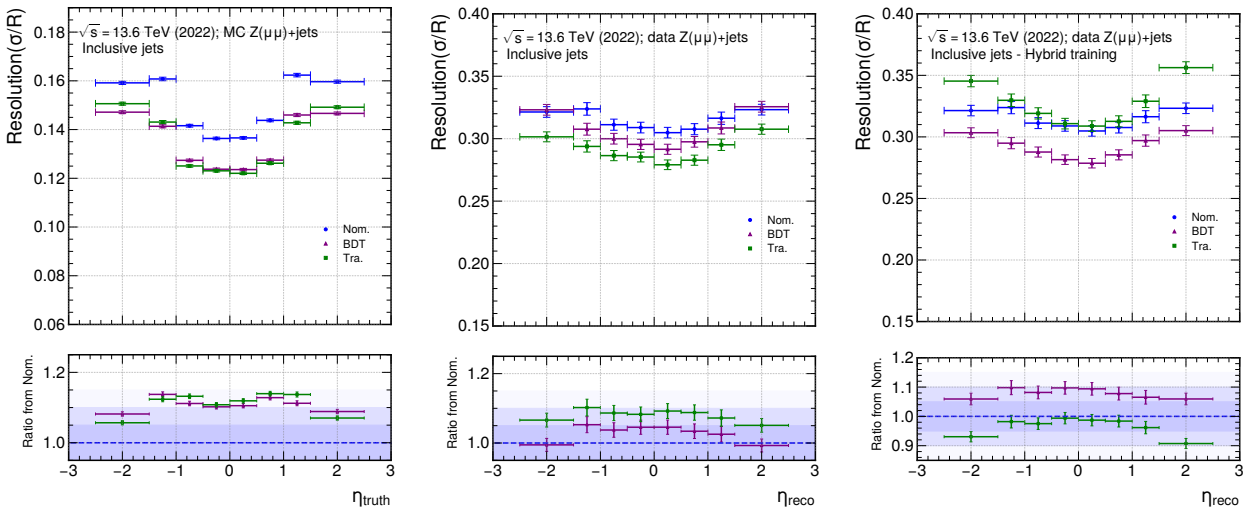


Figure A.7: Comparison of the analytically computed resolution with the truth resolution, shown for each bin of $\sigma_{p_{\mathrm{T}}^{\mathrm{label}}}$. Perfect agreement would correspond to the purple line.



(a) MC-only (b) Data-only (c) Hybrid training

Figure A.8: Observed resolution as a function of pseudorapidity for the nominal calibration (blue dots), the BDT model (purple triangles), and the transformer model (green squares), shown for different training schemes: (a) MC-only, (b) data-only, and (c) hybrid training. The lower panel displays the ratio of each ML algorithm from the nominal calibration.

Appendix B

In-situ b-jet results with different generators

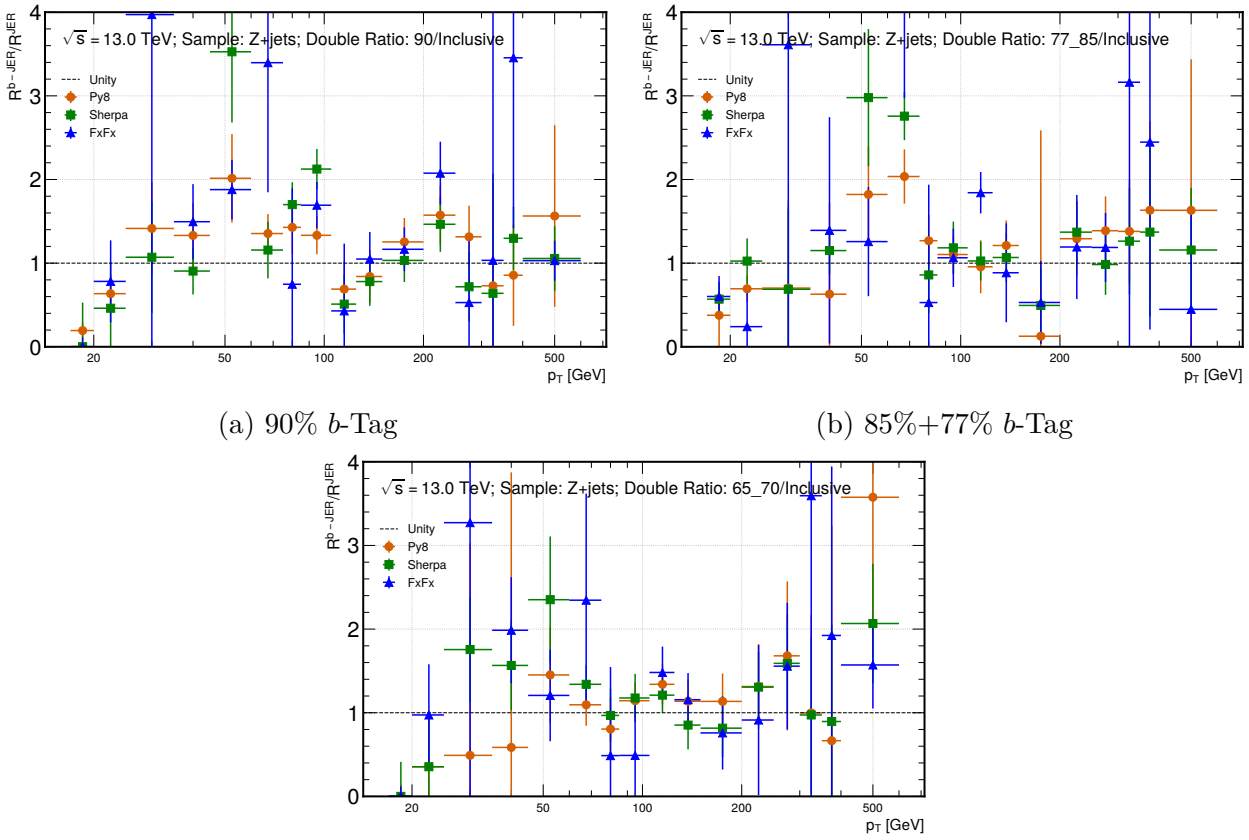
This Appendix provides additional material supporting Chapter 6.

After studying the JER for different generators, the quantity $\tilde{\mathcal{R}}_{b\text{JER}}$ was evaluated for Pythia 8, and the best fit was found to be consistent with a constant line.

Figures B.1 and B.2 show $\tilde{\mathcal{R}}_{b\text{JER}}$ for Z+jets and γ+jets events, respectively, for the three generators studied in each case.

Large fluctuations are observed across all working points and samples; however, the results are generally consistent within their respective uncertainties. The differences between generators are commonly treated as an additional systematic uncertainty and are studied here as a validation.

Since the ratio is consistent with unity, no additional calibration appears to be required. Nevertheless, these differences can serve to better quantify the additional uncertainty that should be assigned to b-jets.

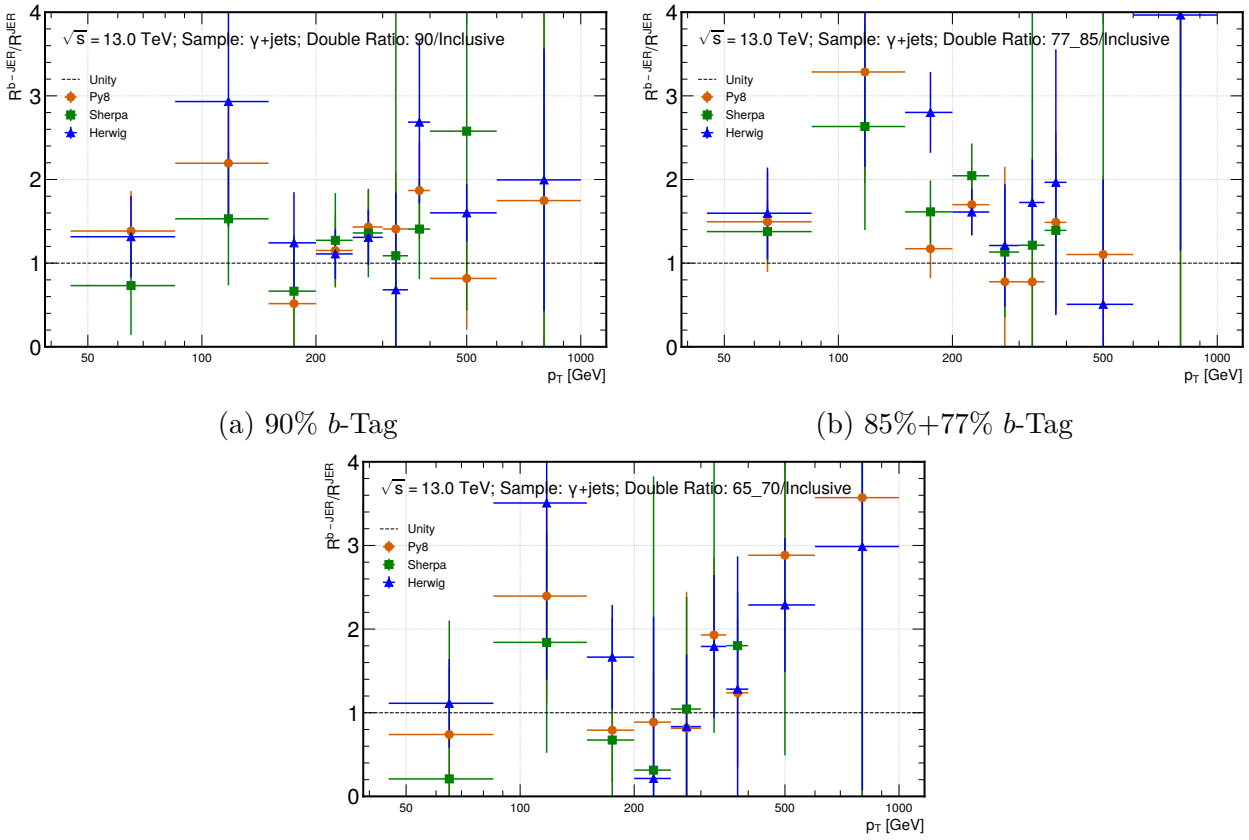


(a) 90% b-Tag

(b) 85%+77% b-Tag

(c) 70%+65% b-Tag

Figure B.1: Ratio of the in-situ jet energy resolution (JER) relative standard deviation between b-tagged and inclusive jets as a function of jet transverse momentum in Run 2 Z+jets events, shown for the b-tagging working points (a) 90%, (b) 85%+77%, and (c) 70%+65%. The ratio is constructed from measurements in data and the corresponding predictions from simulation using Pythia 8 (orange dots), Sherpa (green squares), and FxFx (blue triangles); statistical uncertainties are shown.



(a) 90% b-Tag

(b) 85%+77% b-Tag

(c) 70%+65% b-Tag

Figure B.2: Ratio of the in-situ jet energy resolution (JER) relative standard deviation between b-tagged and inclusive jets as a function of jet transverse momentum in Run 2 γ+jets events, shown for the b-tagging working points (a) 90%, (b) 85%+77%, and (c) 70%+65%. The ratio is constructed from measurements in data and the corresponding predictions from simulation using Pythia 8 (orange dots), Sherpa (green squares), and Herwig 7 (blue triangles); statistical uncertainties are shown.

Appendix C

$H \to Z\gamma$ with ML Z selection

This appendix provides additional material supporting Chapter 8.

C.1 ML Z selection validation plots

C.1.1 Single lepton models

In order to classify lepton candidates as identified or isolated, the `TruthType` variable is used as detailed in Tab. C.1. For electrons, a value of 2 corresponds to an isolated electron. For muons, identification is indicated by `TruthType` values 5, 6, 7, or 8, while isolation is specifically indicated by a value of 6.

truthType value	Definition
1	Unknown Electron
2	Isolated Electron
3	Non isolated Electron
4	Background Electron
5	Unknown Muon
6	Isolated Muon
7	Non isolated Muon
8	Background Muon

Table C.1: Details of the `truthType` meaning for electrons and muons.

Figure C.1 shows the distributions of MC signal and background for the input features of the `e_ISO` model. Most signal distributions are clustered closer to zero compared to the background, while the pile-up distribution behaves similarly for both, as expected. Figure C.2 presents the feature importance ranking, with `topoetcone20` providing the most discriminating power. The ROC curve on the right demonstrates strong model performance, with an area under the curve of AUC = 0.981.

For muons, two models are used: one for identification and another for isolation.

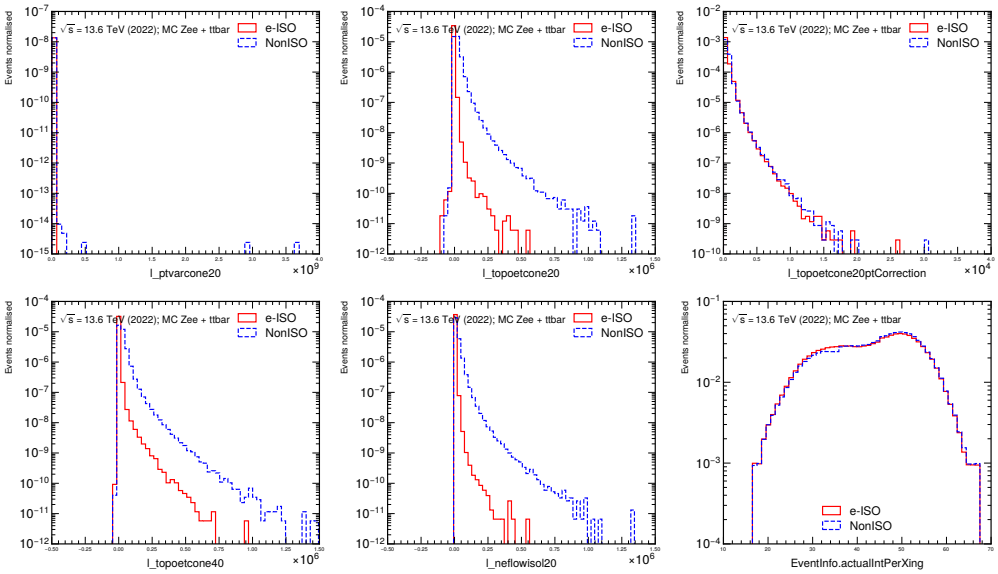


Figure C.1: Normalised distributions of the `e_ISO` input parameters in 2022 MC samples ($Z \to ee$ and $t\bar{t}$). The distributions for eISO are shown in solid red, while non-eISO are shown in dashed blue.

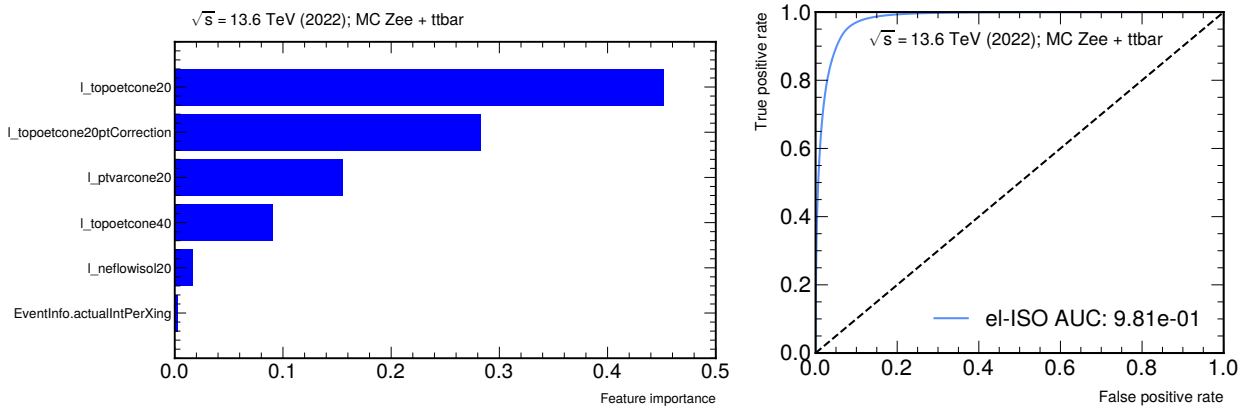


Figure C.2: Left: Feature importance for the `e_ISO` model. Right: ROC curve for the `e_ISO` model using 2022 MC samples ($Z \to ee$ and $t\bar{t}$), with the corresponding accuracy score.

Figure C.3 shows the distributions of MC signal and background for the identification-related variables used in the `mu_PID` model. The variable `segmentDeltaEta` provides the largest 1D visual separation between signal and background. Following the example of electrons, Figure C.4 presents the feature importance ranking and ROC curve. The most important variable is η, followed by `energyLoss`, with the model achieving an area under the curve of AUC = 0.997.

For muon isolation, the distributions of the variables used for training are shown in Fig. C.5. Similar to the electron isolation variables, signal events generally have lower values.

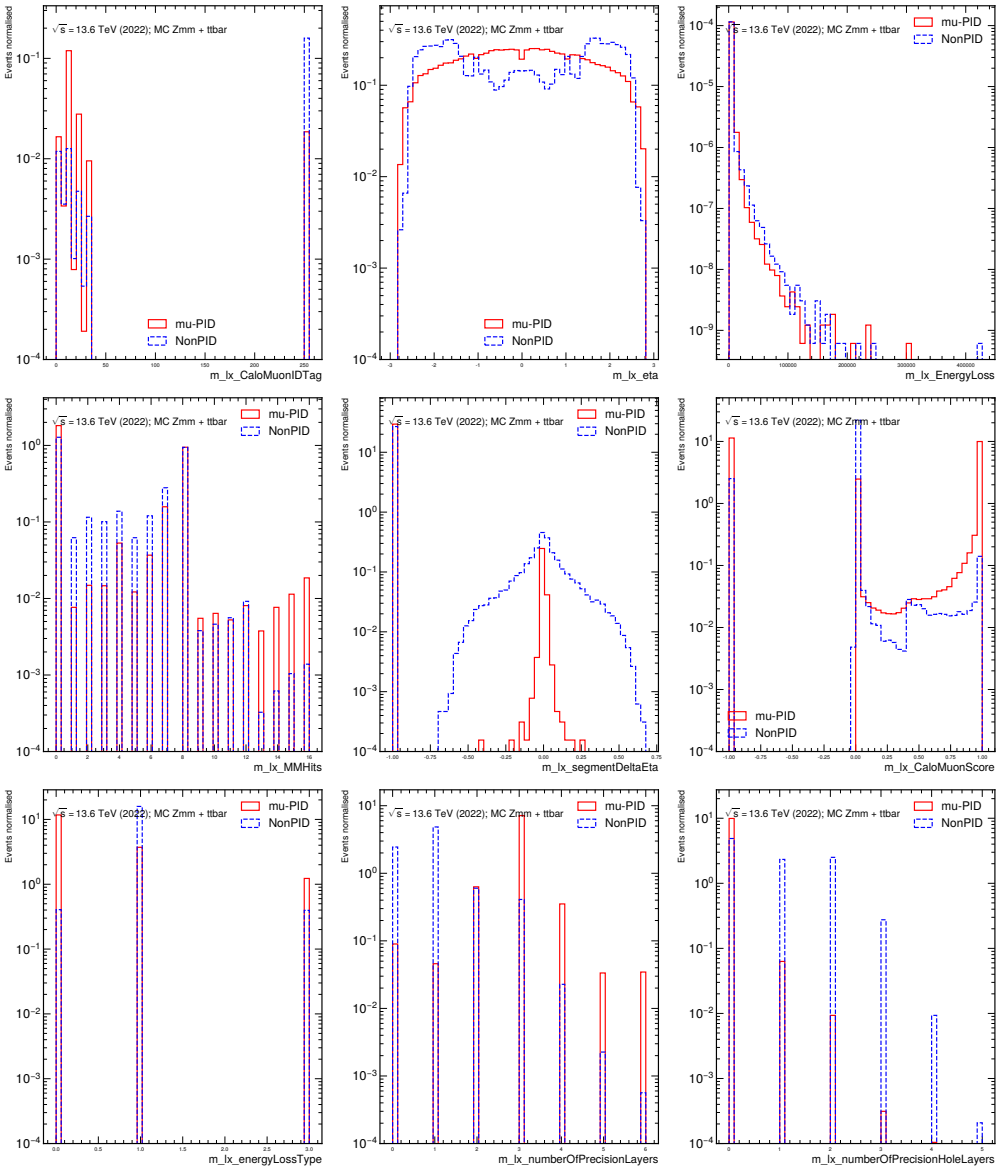


Figure C.3: Normalised distributions of the `mu_PID` input parameters in 2022 MC samples ($Z \to \mu\mu$ and $t\bar{t}$). The distributions for μPID are shown in solid red, while non-μPID are shown in dashed blue.

Figure C.6 shows that the most important features in this model are `DFCommonJetDr` and `newflowisol20`, although all variables contribute with useful information. The ROC curve, shown on the right, has the lowest area among the models presented so far, with an AUC = 0.964.

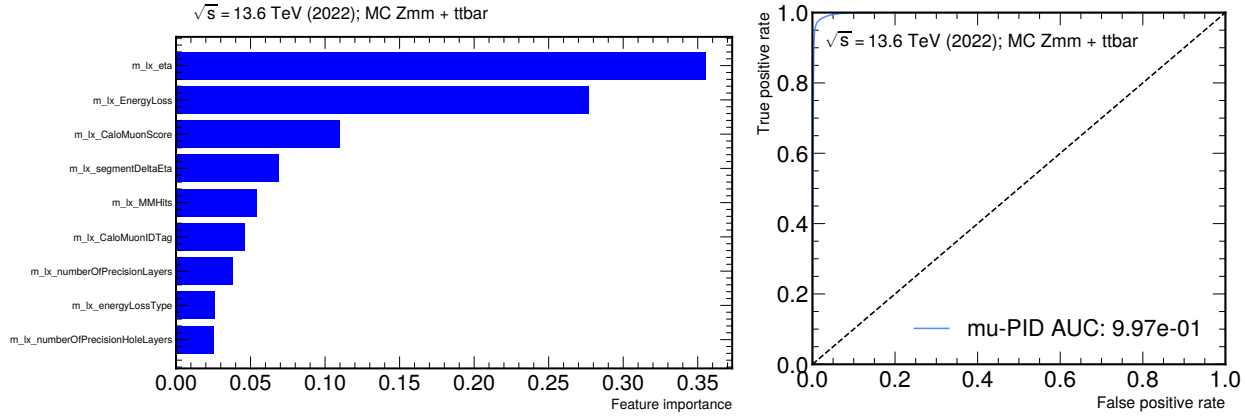


Figure C.4: Left: Feature importance of the `mu_PID` model. Right: ROC curve for the `mu_PID` model using 2022 MC samples ($Z \to \mu\mu$ and $t\bar{t}$), with the corresponding accuracy score.

C.1.2 Z boson models

Once the individual leptons have been assigned a `lep_PID` and `lep_ISO` score, this information is combined for lepton pairs along with the relative impact parameter significance to train two Z-boson models, one for each lepton type.

Figures C.7 and C.8 show the distributions of the input variables for the `Zee_ID` model, along with the feature importance and ROC curve. As noted earlier, for electron identification, the variables produced by the DNN (`DNN_pel`) are used. The most important feature appears to be the isolation score of the first lepton, followed by the identification score, although all variables contribute. The model achieves strong performance, with an area under the ROC curve of AUC = 0.987.

The corresponding plots for the `Zmm_ID` model are shown in Figs. C.9 and C.10. In this case, the feature importance is more evenly distributed across all input variables, with the relative impact parameter significance having a slightly higher weight. The model also achieves strong performance, with an area under the ROC curve of AUC = 0.993.

Overall, both models show significant improvement, which is reflected in the Z mass peak for signal and background, as shown in MC and data in Section 8.3.2 after their implementation in the analysis framework.

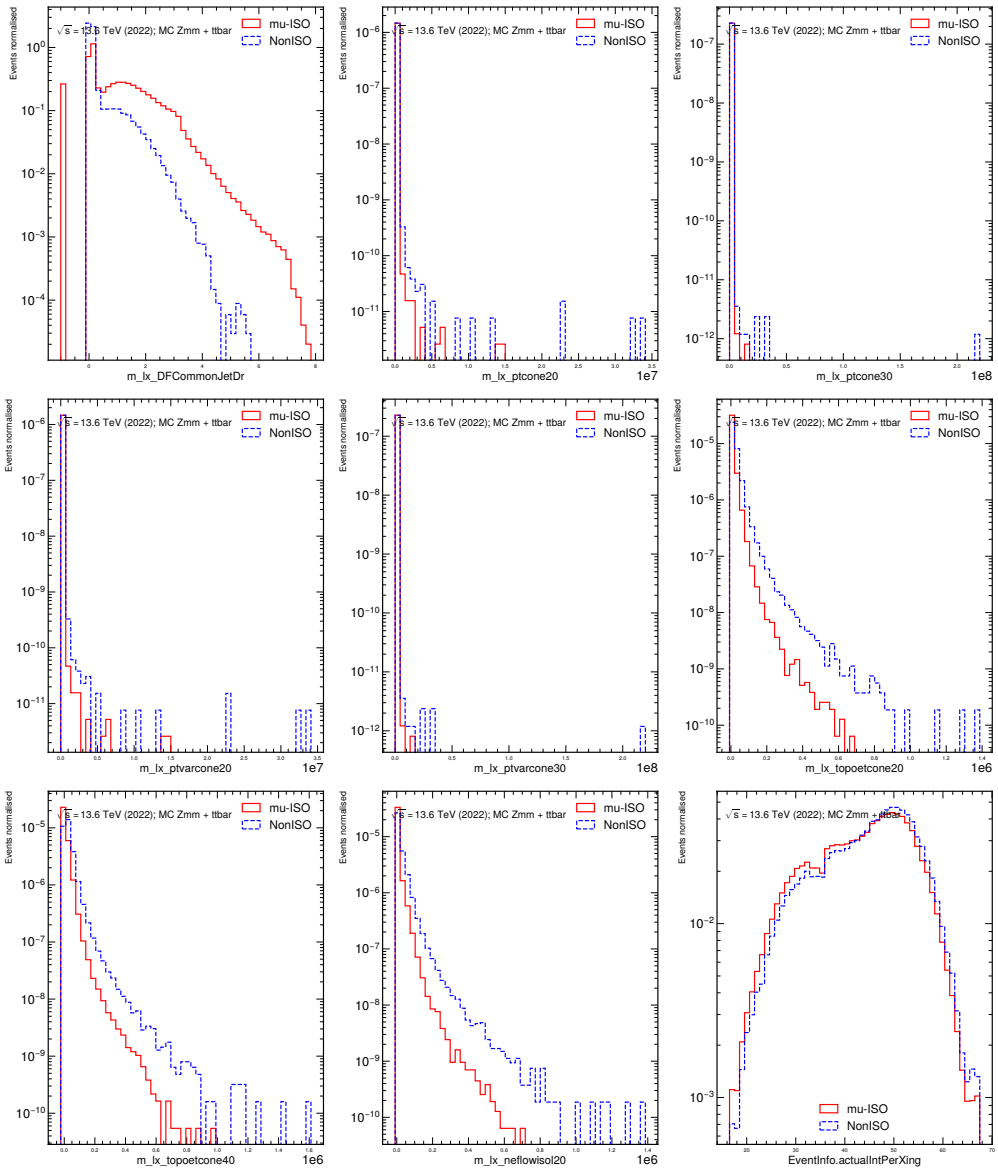


Figure C.5: Normalised distributions of the `mu_ISO` input parameters in 2022 MC samples ($Z \to \mu\mu$ and $t\bar{t}$). The distributions for μISO are shown in solid red, while non-μISO are shown in dashed blue.

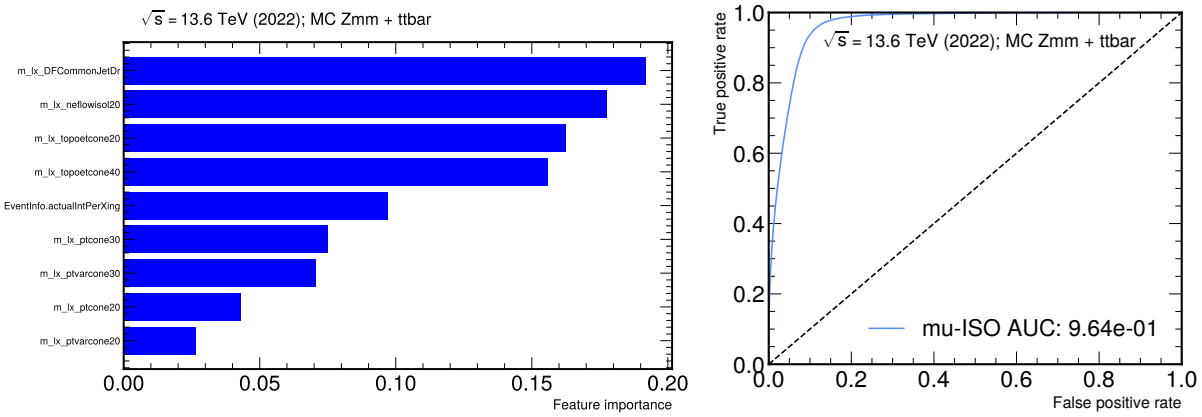


Figure C.6: Left: Feature importance of the `mu_ISO` model. Right: ROC curve for the `mu_ISO` model using 2022 MC samples ($Z \to \mu\mu$ and $t\bar{t}$), with the corresponding accuracy score.

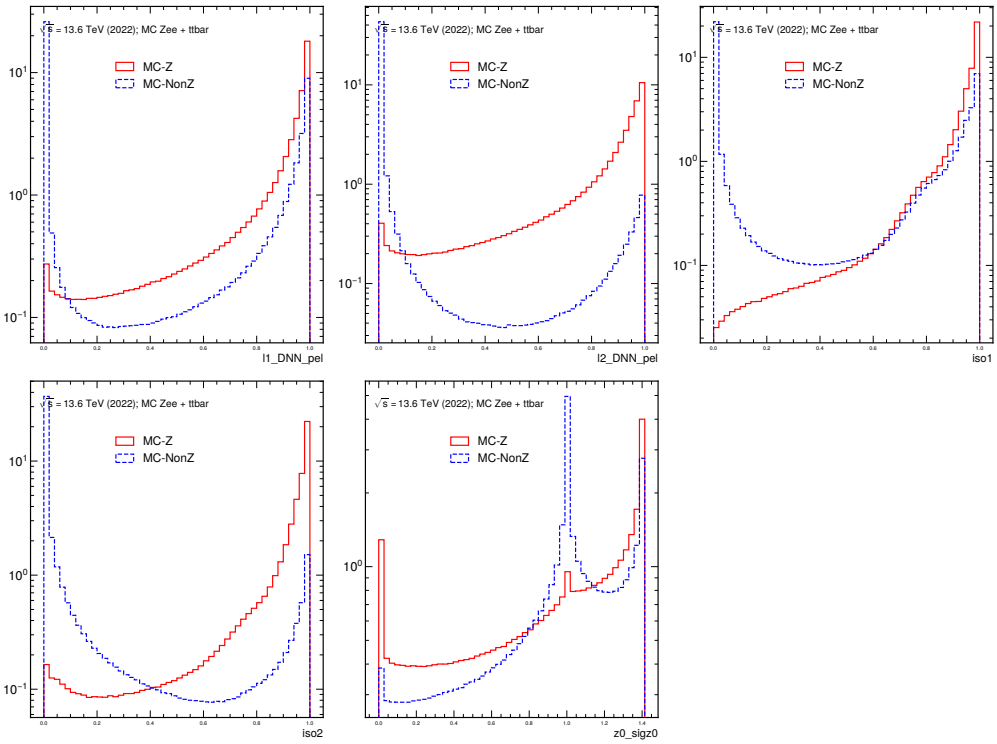


Figure C.7: Normalised distributions of the `Zee_ID` input parameters in 2022 MC samples ($Z \to ee$ and $t\bar{t}$). The distributions for Z_ID are shown in solid red, while non-Z_ID are shown in dashed blue.

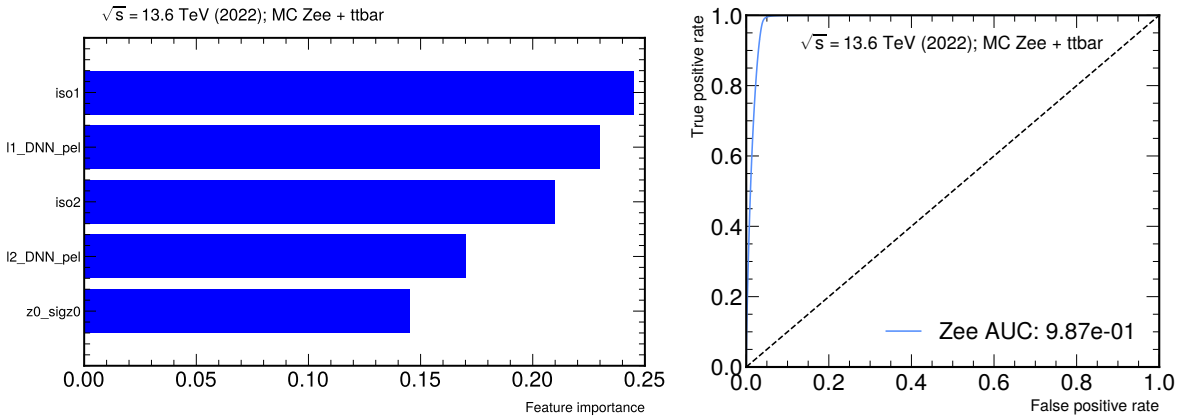


Figure C.8: Left: Feature importance of the `Zee_ID` model. Right: ROC curve for the `Zee_ID` model using 2022 MC samples ($Z \to ee$ and $t\bar{t}$), with the corresponding accuracy score.

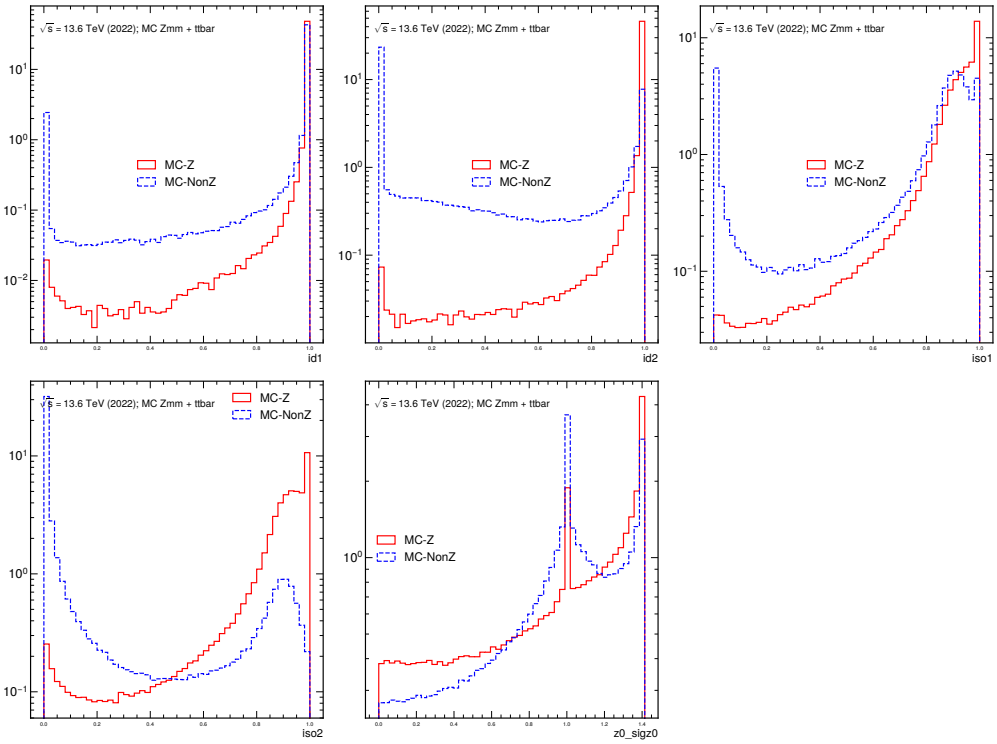


Figure C.9: Normalised distributions of the `Zmm_ID` input parameters in 2022 MC samples ($Z \to \mu\mu$ and $t\bar{t}$). The distributions for Z_ID are shown in solid red, while non-Z_ID are shown in dashed blue.

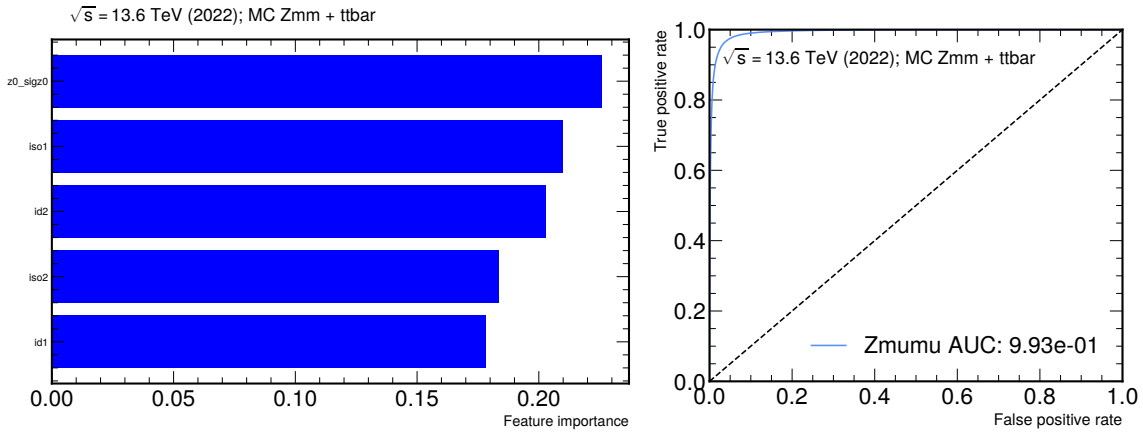


Figure C.10: Left: Feature importance of the `Zmm_ID` model. Right: ROC curve for the `Zmm_ID` model using 2022 MC samples ($Z \to \mu\mu$ and $t\bar{t}$), with the corresponding accuracy score.

C.2 Cutflow detailed

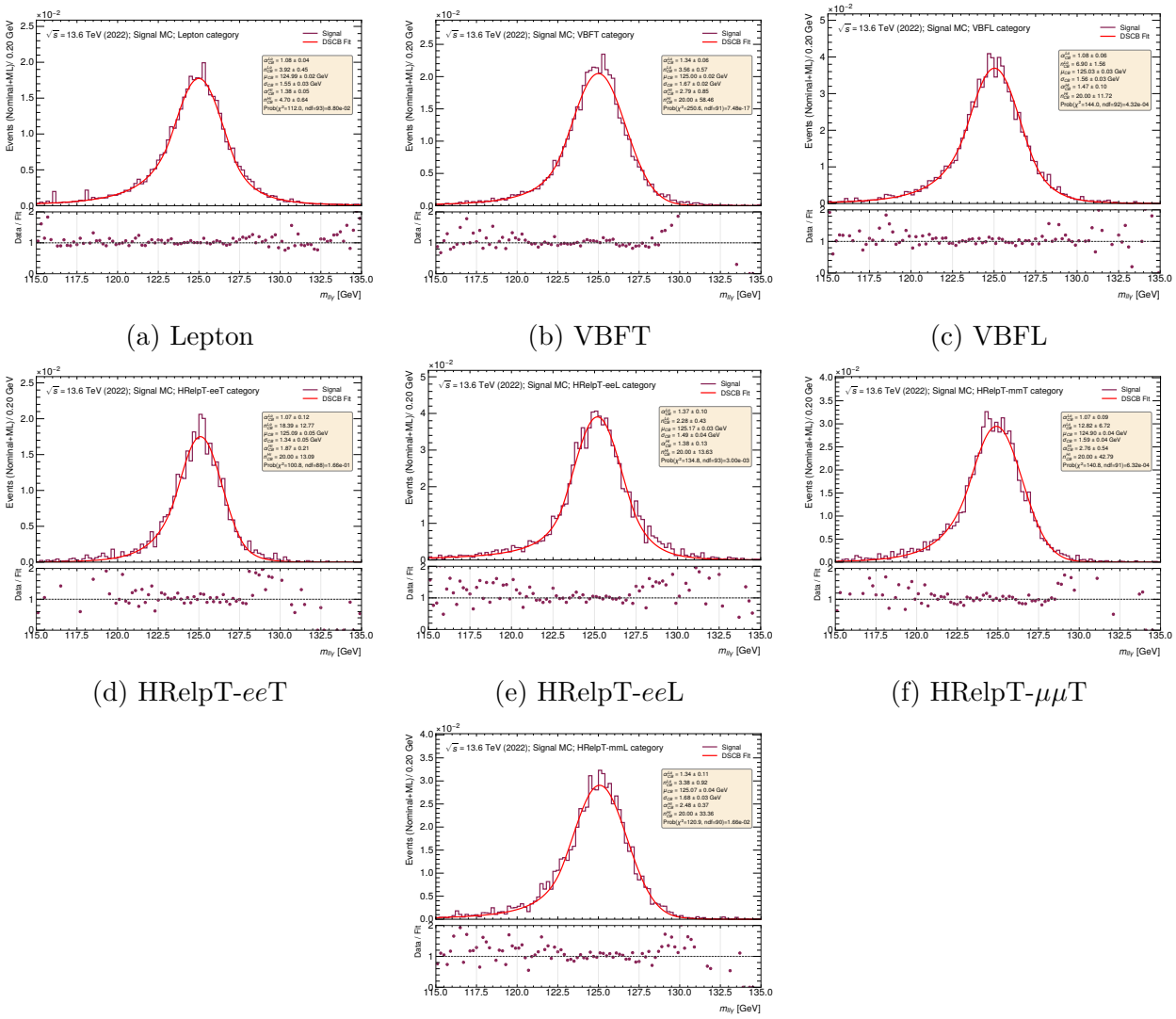
Table C.2 explicitly defines the sequence of selection steps applied in the nominal $H \to Z\gamma$ analysis; its numbers are equivalent to those shown in Fig. 8.3.

Cutflow	**Selection requirement**
3	Events/sumweight in DAOD
4	Events/sumweight in DAOD
5	GRL
6	Primary-Vertex
7	Event-Quality
8	Triggers
9	(No selection applied)
10	Minimum two lepton requirement
11	Lepton ID
12	Lepton ISO
13	Object level overlap removal
14	Two lepton opposite sign
15	Loose photon selection
16	Trigger match
17	Overlap removal on photon and jets
18	Di-lepton mass final cut
19	Photon p_{T} low-mass cut
20	Photon ID final cut
21	Photon ISO final cut
22	Three body mass cut
23	Photon p_{T} high-mass cut

Table C.2: Definition of the nominal cutflow steps used in the baseline $H \to Z\gamma$ event selection.

C.3 Signal and background modelling

The fits described in Section 8.5 are shown for both signal and background per category, together with the corresponding best-fit parameters. The signal fits are presented in Figs. C.11 and C.12, while the background fits are presented in Figs. C.13, C.14, and C.15. Although the background fits were already discussed in the main text, the corresponding fit parameters are included here for completeness and future reference.

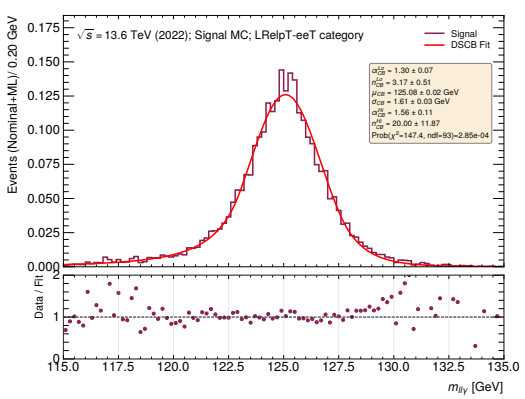


(a) Lepton (b) VBFT (c) VBFL

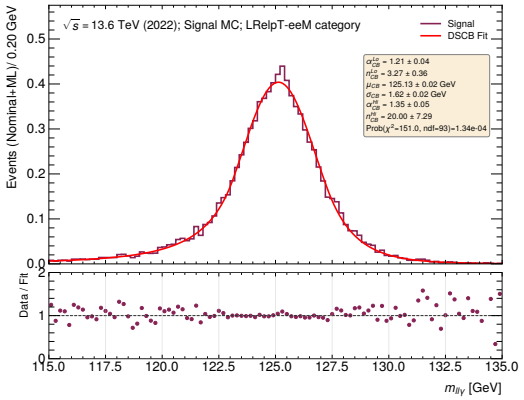
(d) HRelpT-eeT (e) HRelpT-eeL (f) HRelpT-$\mu\mu$T

(g) HRelpT-$\mu\mu$L

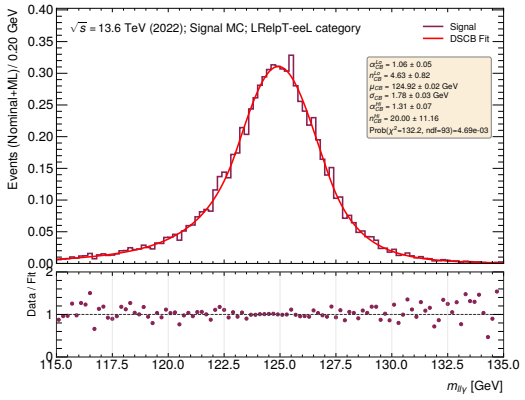
Figure C.11: Invariant mass distribution of nominal plus ML selected MC signal for the (a) *Lepton*, (b) *VBFT*, (c) *VBFL*, (d) *HRelpT-eeT*, (e) *HRelpT-eeL*, (f) *HRelpT-$\mu\mu$T*, and (g) *HRelpT-$\mu\mu$L* categories, shown in purple. The distributions are fitted with a double-sided Crystal Ball function, shown in red, with the best-fit parameters, the χ^2 and the fit probability are reported in the inset. The lower panels show the bin-by-bin ratio of the distribution to the fit.



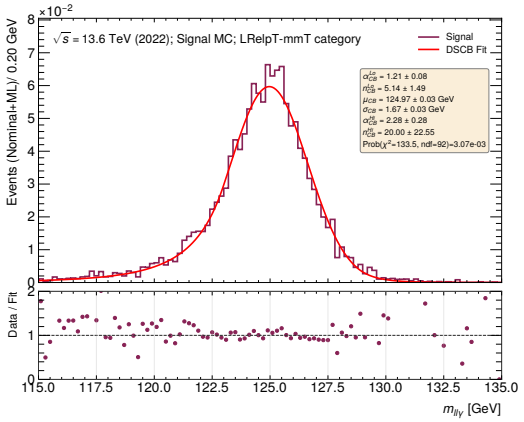
(a) LRelpT-eeT



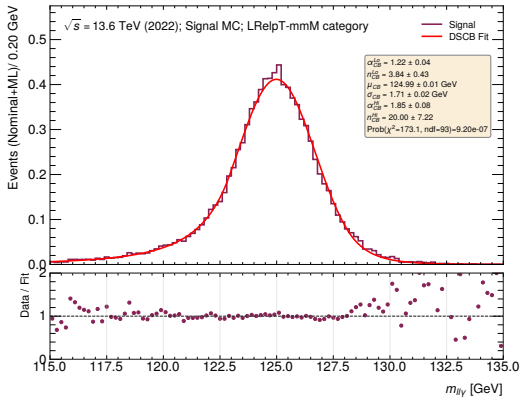
(b) LRelpT-eeM



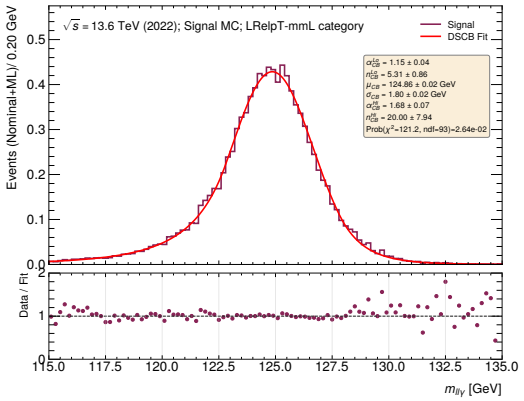
(c) LRelpT-eeL



(d) LRelpT-$\mu\mu$T

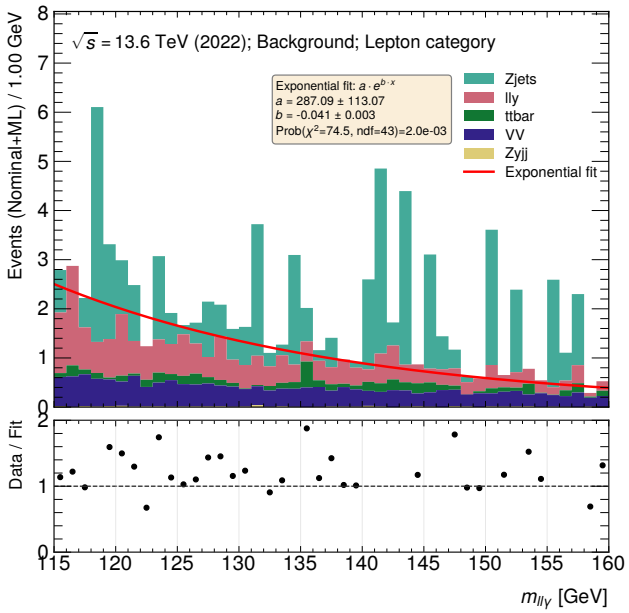


(e) LRelpT-$\mu\mu$M

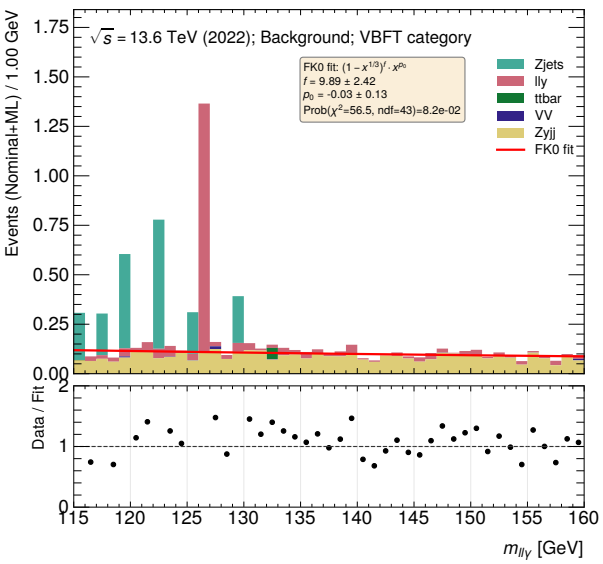


(f) LRelpT-$\mu\mu$T

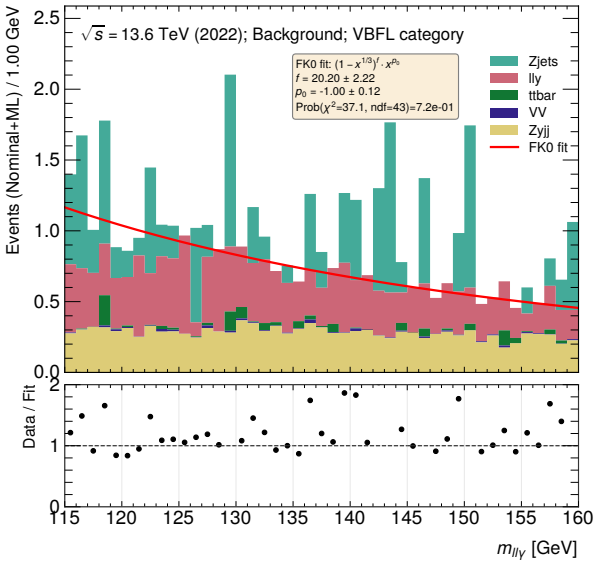
Figure C.12: Invariant mass distribution of nominal plus ML selected MC signal for the (a) *LRelpT-eeT*, (b) *LRelpT-eeM*, (c) *LRelpT-eeL*, (d) *LRelpT-μμT*, (e) *LRelpT-μμM*, and (f) *LRelpT-μμL* categories, shown in purple. The distributions are fitted with a double-sided Crystal Ball function, shown in red, with the best-fit parameters, the χ^2 and the fit probability are reported in the inset. The lower panels show the bin-by-bin ratio of the distribution to the fit.



(a) Lepton

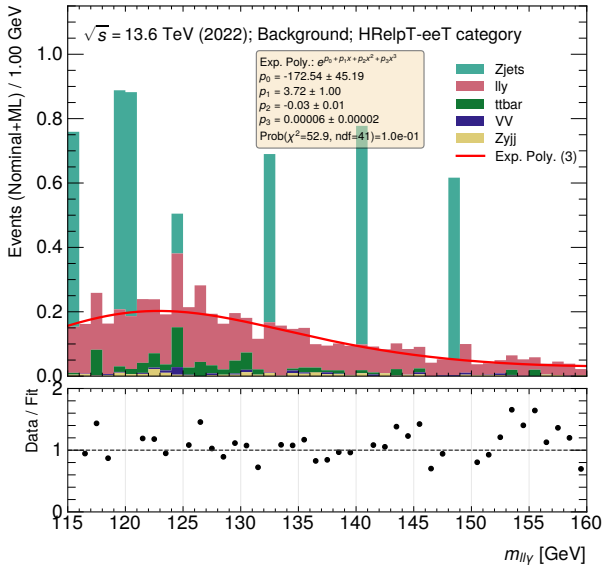


(b) VBFT

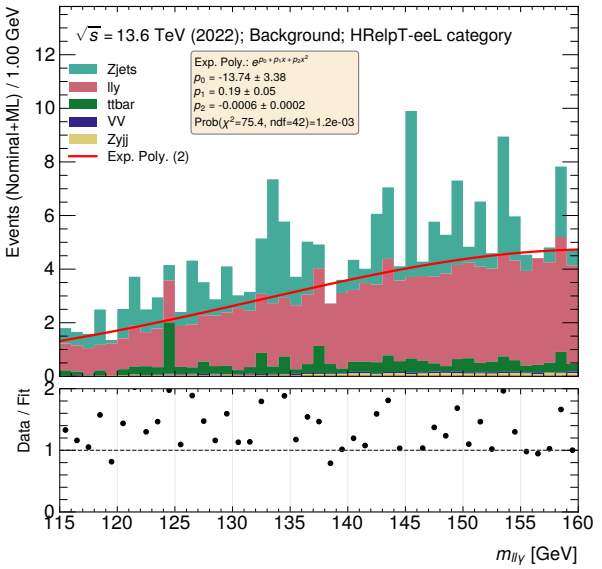


(c) VBFL

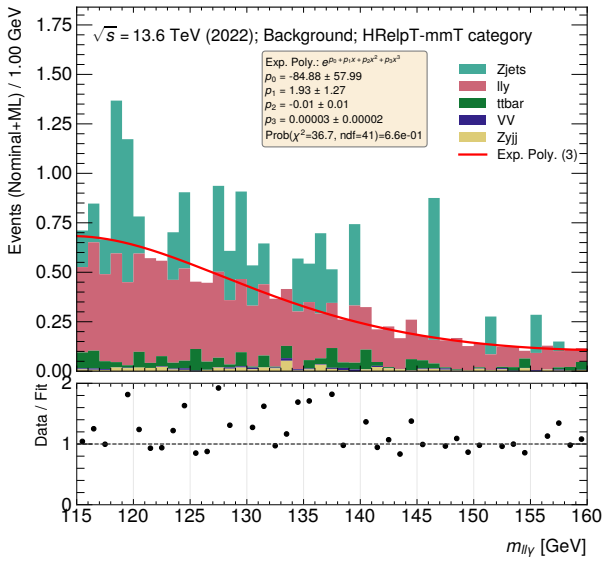
Figure C.13: Invariant mass distribution of nominal plus ML selected MC background for the (a) *Lepton*, (b) *VBFT*, and (c) *VBFL* categories. The distributions are fitted with the best-fit function in red: an exponential for the *Lepton* category and an FK0 for the *VBF* categories. The best-fit parameters, the χ^2 and the fit probability are reported in the inset. The lower panel shows the bin-by-bin ratio of the distribution to the fit.



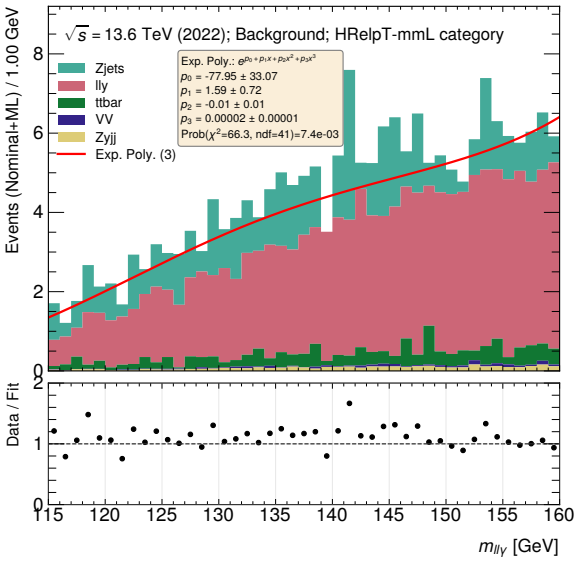
(a) HRelpT-eeT



(b) HRelpT-eeL

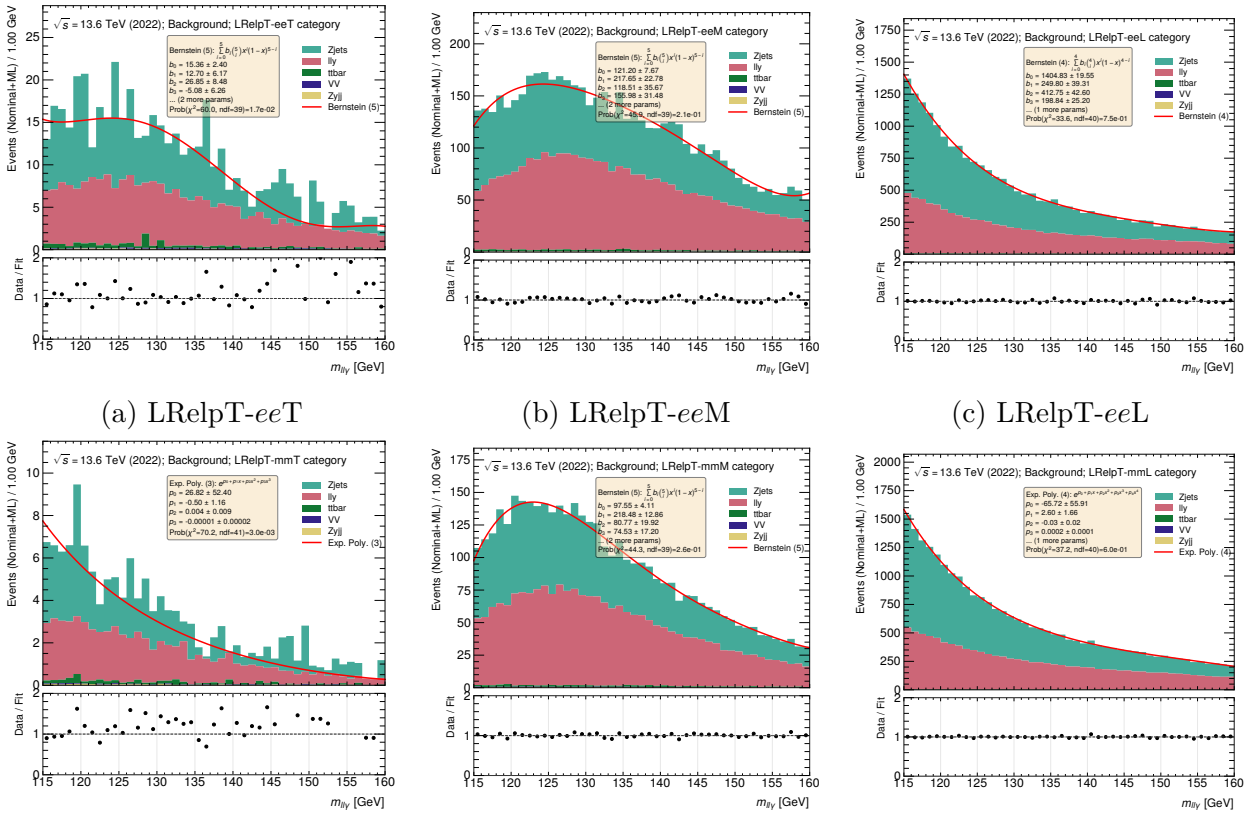


(c) HRelpT-$\mu\mu$T



(d) HRelpT-$\mu\mu$L

Figure C.14: Invariant mass distributions of nominal plus ML selected MC background for the (a) *HRelpT-eeT*, (b) *HRelpT-eeL*, (c) *HRelpT-μμT*, and (d) *HRelpT-μμL* categories. The distributions are fitted with the best-fit second- and third-order exponential polynomial functions shown in red. The best-fit parameters, the χ^2 and the fit probability are reported in the inset. The lower panel displays the bin-by-bin ratio of the distribution to the fit.



(a) LRelpT-eeT (b) LRelpT-eeM (c) LRelpT-eeL

(d) LRelpT-$\mu\mu$T (e) LRelpT-$\mu\mu$M (f) LRelpT-$\mu\mu$L

Figure C.15: Invariant mass distributions of nominal plus ML selected MC background for the *LRelpT* categories, with electron and muon channels each divided into Tight (T), Medium (M), and Loose (L) subcategories. The distributions are fitted with the best-fit third- and fourth-order exponential polynomials and fourth- and fifth-order Bernstein polynomials, shown in red. The best-fit parameters, the χ^2 and the fit probability are reported in the inset. The lower panel shows the bin-by-bin ratio of the distribution to the fit.

C.4 Sensitivity 2022

After fitting the signal and background in each category using the available 2022 MC samples, the signal window containing 68% of the signal is determined, and the expected number of signal events within this range, S_{68}^{exp}, is extracted from the fit. The same window is then applied to the background to obtain the corresponding number of background events, B_{68}^{exp}. Signal uncertainties are estimated by propagating the uncertainties of the fitted model parameters through pseudo-experiments to capture the statistical component of the fit. Background uncertainties are evaluated either by propagating the fitted parameter uncertainties through the integral over the signal region or, when the fit covariance matrix is ill-conditioned, via bootstrap resampling. The sensitivity is then computed using both S_{68}^{exp} and B_{68}^{exp}.

To allow a fair comparison with the published partial Run 3 analysis, the distributions were scaled using the corresponding luminosity weight, as described in Section 8.6. Table C.3 shows the (*raw*) results using only the 2022 MC samples. The values are consistent with the scaled results, demonstrating overall agreement and validating the procedure.

Category	w_{68} [GeV]	S_{68}^{exp}	B_{68}^{exp}	$S_{68}^{\text{exp}}/\sqrt{S_{68}^{\text{exp}}+B_{68}^{\text{exp}}}$
Lepton	3.93	0.28 ± 0.01	6.5 ± 0.4	0.11
VBFT	3.64	0.32 ± 0.01	0.40 ± 0.02	0.38
VBFL	3.68	0.56 ± 0.02	3.41 ± 0.16	0.28
HRelpT-eeT	3.02	0.22 ± 0.01	0.60 ± 0.5	0.24
HRelpT-eeL	3.51	0.56 ± 0.02	7.6 ± 0.4	0.20
HRelpT-$\mu\mu$T	3.55	0.45 ± 0.02	1.9 ± 0.6	0.29
HRelpT-$\mu\mu$L	3.68	0.45 ± 0.02	10.0 ± 1.4	0.14
LRelpT-eeT	3.67	1.89 ± 0.05	57 ± 3	0.25
LRelpT-eeM	3.88	6.30 ± 0.10	624 ± 10	0.25
LRelpT-eeL	4.37	5.42 ± 0.16	3 088 ± 27	0.10
LRelpT-$\mu\mu$T	3.70	0.93 ± 0.03	15.4 ± 3	0.23
LRelpT-$\mu\mu$M	3.87	6.58 ± 0.08	543 ± 6	0.28
LRelpT-$\mu\mu$L	4.10	7.21 ± 0.13	3 449 ± 38	0.12
Inclusive	4.01	31.3 ± 0.2	7 515 ± 90	0.85

Table C.3: Sensitivity from ML Z selected events from 2022 MC. The expected signal events (S_{68}^{exp}) and background events (B_{68}^{exp}) are obtained by integrating with the fit in a mass window of width w_{68}, defined to contain 68% of the signal. The signal uncertainties represent the statistical uncertainty on the yield in the signal region, obtained by propagating the fitted model parameter uncertainties via pseudo-experiments. The background uncertainties represent the uncertainty on the estimated background yield in the signal region from the fitted background model. For the inclusive case, the sensitivity is obtained by combining the individual category sensitivities in quadrature.

Bibliography

[1] S.L. Glashow, *Partial Symmetries of Weak Interactions*, *Nucl. Phys.* **22** (1961) 579. (Cited on pages 1 and 7)

[2] S. Weinberg, *A Model of Leptons*, *Phys. Rev. Lett.* **19** (1967) 1264. (Cited on pages 1 and 7)

[3] A. Salam and J.C. Ward, *Weak and electromagnetic interactions*, *Nuovo Cim.* **11** (1959) 568. (Cited on pages 1 and 7)

[4] G. 't Hooft and M.J.G. Veltman, *Regularization and Renormalization of Gauge Fields*, *Nucl. Phys. B* **44** (1972) 189. (Cited on pages 1 and 7)

[5] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, *Phys. Rev. Lett.* **13** (1964) 321. (Cited on pages 1 and 16)

[6] P.W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, *Phys. Rev. Lett.* **13** (1964) 508. (Cited on pages 1 and 16)

[7] P.W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, *Phys. Rev.* **145** (1966) 1156. (Cited on pages 1 and 16)

[8] G.S. Guralnik, C.R. Hagen and T.W.B. Kibble, *Global Conservation Laws and Massless Particles*, *Phys. Rev. Lett.* **13** (1964) 585. (Cited on pages 1 and 16)

[9] T.W.B. Kibble, *Symmetry breaking in nonAbelian gauge theories*, *Phys. Rev.* **155** (1967) 1554. (Cited on pages 1 and 16)

[10] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012) 1 [1207.7214]. (Cited on pages 1, 16, 23, and 100)

[11] CMS Collaboration, *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, *Phys. Lett. B* **716** (2012) 30 [1207.7235]. (Cited on pages 1, 16, 23, and 100)

[12] L. Evans and P. Bryant, eds., *LHC Machine*, *JINST* **3** (2008) S08001. (Cited on pages 1, 23, and 27)

[13] ATLAS Collaboration, *Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Eur. Phys. J. C* **81** (2021) 689 [2007.02645]. (Cited on pages 2, 51, 52, 54, and 55)

[14] ATLAS Collaboration, *In situ calibration of large-radius jet energy and mass in 13 TeV proton–proton collisions with the ATLAS detector*, *Eur. Phys. J. C* **79** (2019) 135 [1807.09477]. (Cited on pages 2 and 51)

[15] ATLAS Collaboration, *Measurements of WH and ZH production in the $H \to b\bar{b}$ decay channel in pp collisions at 13 TeV with the ATLAS detector*, *Eur. Phys. J. C* **81** (2021) 178 [2007.02873]. (Cited on pages 2, 59, and 60)

[16] ATLAS Collaboration, *Evidence for the $H \to b\bar{b}$ decay with the ATLAS detector*, *JHEP* **12** (2017) 024 [1708.03299]. (Cited on pages 3, 57, and 58)

[17] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez et al., *Attention Is All You Need*, in *31st International Conference on Neural Information Processing Systems*, 6, 2017 [1706.03762]. (Cited on pages 3, 56, and 65)

[18] ATLAS Collaboration, *Transformer networks for constituent-based b-jet calibration with the ATLAS detector*, Tech. Rep. ATL-PHYS-PUB-2024-015, CERN, Geneva (2024). (Cited on pages 3, 77, 79, 80, 81, 83, 84, 85, and 86)

[19] ATLAS and CMS Collaborations, *Evidence for the Higgs Boson Decay to a Z Boson and a Photon at the LHC*, *Phys. Rev. Lett.* **132** (2024) 021803 [2309.03501]. (Cited on pages 3 and 100)

[20] ATLAS Collaboration, *Search for the Higgs boson decay to a Z boson and a photon in pp collisions at $\sqrt{s} = 13$ TeV and 13.6 TeV with the ATLAS detector*, 2507.12598. (Cited on pages 3, 100, 104, 105, 106, 109, 110, 112, 124, and 132)

[21] I. Brivio and M. Trott, *The Standard Model as an Effective Field Theory*, *Phys. Rept.* **793** (2019) 1 [1706.08945]. (Cited on page 7)

[22] M.D. Schwartz, *Quantum Field Theory and the Standard Model*, Cambridge University Press (3, 2014). (Cited on page 7)

[23] M. Thomson, *Modern particle physics*, Cambridge University Press, New York (10, 2013), 10.1017/CBO9781139525367. (Cited on page 7)

[24] M.E. Peskin and D.V. Schroeder, *An Introduction to quantum field theory*, Addison-Wesley, Reading, USA (1995), 10.1201/9780429503559. (Cited on page 7)

[25] S. Weinberg, *The Quantum theory of fields. Vol. 1: Foundations*, Cambridge University Press (6, 2005), 10.1017/CBO9781139644167. (Cited on page 7)

[26] I. Neutelings, "Figure Standard Model." https://tikz.net/sm_particles/. (Cited on page 8)

[27] E. Noether, *Invariant Variation Problems*, *Gott. Nachr.* **1918** (1918) 235 [physics/0503066]. (Cited on page 10)

[28] C.S. Wu, E. Ambler, R.W. Hayward, D.D. Hoppes and R.P. Hudson, *Experimental Test of Parity Conservation in β Decay*, *Phys. Rev.* **105** (1957) 1413. (Cited on page 11)

[29] LHCHiggsCrossSectionWorkingGroup collaboration, *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, Tech. Rep. CERN-2017-002-M, CERN-2017-002, Geneva (2017), DOI. (Cited on page 14)

[30] ATLAS Collaboration, *Combined Measurement of the Higgs Boson Mass from the H→γγ and H→ZZ*→4ℓ Decay Channels with the ATLAS Detector Using s=7, 8, and 13 TeV pp Collision Data*, *Phys. Rev. Lett.* **131** (2023) 251802 [2308.04775]. (Cited on page 16)

[31] L.D. Landau, *On the angular momentum of a system of two photons*, *Dokl. Akad. Nauk SSSR* **60** (1948) 207. (Cited on page 16)

[32] C.-N. Yang, *Selection Rules for the Dematerialization of a Particle Into Two Photons*, *Phys. Rev.* **77** (1950) 242. (Cited on page 16)

[33] ATLAS Collaboration, *Evidence for the spin-0 nature of the Higgs boson using ATLAS data*, *Phys. Lett. B* **726** (2013) 120 [1307.1432]. (Cited on page 16)

[34] ATLAS Collaboration, *A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery*, *Nature* **607** (2022) 52 [2207.00092]. (Cited on pages 17 and 18)

[35] M.H. Goroff and A. Sagnotti, *QUANTUM GRAVITY AT TWO LOOPS*, *Phys. Lett. B* **160** (1985) 81. (Cited on page 19)

[36] Super-Kamiokande collaboration, *Evidence for oscillation of atmospheric neutrinos*, *Phys. Rev. Lett.* **81** (1998) 1562 [hep-ex/9807003]. (Cited on page 19)

[37] SNO collaboration, *Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory*, *Phys. Rev. Lett.* **89** (2002) 011301 [nucl-ex/0204008]. (Cited on page 19)

[38] KATRIN collaboration, *Direct neutrino-mass measurement based on 259 days of KATRIN data*, *Science* **388** (2025) adq9592 [2406.13516]. (Cited on page 19)

[39] Atacama Cosmology Telescope collaboration, *The Atacama Cosmology Telescope: DR6 constraints on extended cosmological models*, *JCAP* **11** (2025) 063 [2503.14454]. (Cited on page 19)

[40] D. Paraficz, J.P. Kneib, J. Richard, A. Morandi, M. Limousin, E. Jullo et al., *The Bullet cluster at its best: weighing stars, gas, and dark matter*, *Astron. Astrophys.* **594** (2016) A121 [1209.0384]. (Cited on page 19)

[41] Planck collaboration, *Planck 2018 results. I. Overview and the cosmological legacy of Planck*, *Astron. Astrophys.* **641** (2020) A1 [1807.06205]. (Cited on page 19)

[42] A.G. Doroshkevich, V.N. Lukash and E.V. Mikheeva, *A solution to the problems of cusps and rotation curves in dark matter halos in the cosmological standard model*, *Phys. Usp.* **55** (2012) 3 [1209.0388]. (Cited on page 19)

[43] M. Milgrom, *A Modification of the Newtonian dynamics: Implications for galaxies*, *Astrophys. J.* **270** (1983) 371. (Cited on page 19)

[44] Supernova Search Team collaboration, *Observational evidence from supernovae for an accelerating universe and a cosmological constant*, *Astron. J.* **116** (1998) 1009 [astro-ph/9805201]. (Cited on page 19)

[45] Planck collaboration, *Planck 2018 results. VI. Cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6 [1807.06209]. (Cited on page 19)

[46] Pan-STARRS1 collaboration, *The Complete Light-curve Sample of Spectroscopically Confirmed SNe Ia from Pan-STARRS1 and Cosmological Constraints from the Combined Pantheon Sample*, *Astrophys. J.* **859** (2018) 101 [1710.00845]. (Cited on page 19)

[47] S. Weinberg, *The Cosmological Constant Problem*, *Rev. Mod. Phys.* **61** (1989) 1. (Cited on page 19)

[48] A.D. Sakharov, *Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe*, *Pisma Zh. Eksp. Teor. Fiz.* **5** (1967) 32. (Cited on page 20)

[49] R.D. Peccei, *The Strong CP problem and axions*, *Lect. Notes Phys.* **741** (2008) 3 [hep-ph/0607268]. (Cited on page 20)

[50] CERN, "Convention for the establishment of a European organization for nuclear research." https://cds.cern.ch/record/330625, 1971. (Cited on page 21)

[51] U. Mersits, *Construction of the CERN SynchroCyclotron (1952-1957)*, Tech. Rep. CERN-CHS-13, CERN, Geneva (1984). (Cited on page 22)

[52] G. Fidecaro, *The Discoveries of Rare Pion Decays at the CERN Synchrocyclotron*, *Adv. Ser. Direct. High Energy Phys.* **23** (2015) 397. (Cited on page 22)

[53] CERN, "Proton Synchrotron." https://cds.cern.ch/record/1479602, 1956. (Cited on page 22)

[54] T. Massam, T. Müller, B. Righini, M. Schneegans and A. Zichichi, *Experimental observation of antideuteron production*, *Nuovo Cimento* **39** (1965) 10. (Cited on page 22)

[55] WA14, BARI-CERN-ECOLE POLY-MILAN-ORSAY collaboration, *Results from gargamelle neutrino experiment at CERN SPS*, in *1979 EPS High-Energy Physics Conference*, vol. 1, (Geneva, Switzerland), pp. 167–176, CERN, 1979. (Cited on page 22)

[56] D. Haidt, *The Discovery of Weak Neutral Currents*, *Adv. Ser. Dir. High Energy Phys.* **23** (2015) 165. (Cited on page 22)

[57] G.Charpak and et al., *The use of multiwire proportional counters to select and localize charged particles*, Tech. Rep. CERN-2018-06 (1968). (Cited on page 22)

[58] T. Mastoridis and P. Baudrenghien, *CERN's Super Proton Synchrotron 200 MHz cavity regulation upgrade: Modeling, design optimization, and performance estimation*, *Phys. Rev. Accel. Beams* **25** (2022) 021002. (Cited on pages 22 and 26)

[59] N. Doble, L. Gatignon, K. Hübner and E. Wilson, *The Super Proton Synchrotron (SPS): A Tale of Two Lives*, *Adv. Ser. Direct. High Energy Phys.* **27** (2017) 135. (Cited on pages 22 and 26)

[60] L. Di Lella and C. Rubbia, *The Discovery of the W and Z Particles*, *Adv. Ser. Dir. High Energy Phys.* **23** (2015) 137. (Cited on page 22)

[61] Carlos Rubbia and Simon van der Meer, "Nobel prize in physics 1984." https://www.nobelprize.org/prizes/physics/1984/summary/. (Cited on page 22)

[62] E.V. Shuryak, *Quantum Chromodynamics and the Theory of Superdense Matter*, *Phys. Rept.* **61** (1980) 71. (Cited on page 22)

[63] S. Myers, *The LEP Collider, from design to approval and commissioning*, John Adams' memorial lecture, CERN, Geneva (1991), 10.5170/CERN-1991-008. (Cited on page 22)

[64] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08003. (Cited on pages 23, 28, and 31)

[65] CMS Collaboration, *The CMS experiment at the CERN LHC*, *JINST* **3** (2008) S08004. (Cited on page 23)

[66] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, *JINST* **3** (2008) S08002. (Cited on page 23)

[67] LHCb Collaboration, *The LHCb Detector at the LHC*, *JINST* **3** (2008) S08005. (Cited on page 23)

[68] O. Aberle, I. Béjar Alonso, O. Brüning, P. Fessia, L. Rossi, L. Tavian et al., *High-Luminosity Large Hadron Collider (HL-LHC): Technical design report*, CERN Yellow Reports: Monographs, CERN, Geneva (2020), 10.23731/CYRM-2020-0010. (Cited on page 23)

[69] E. Mobs, "The CERN accelerator complex in 2019." https://cds.cern.ch/record/2684277, 2019. (Cited on page 24)

[70] M. Vretenar, J. Vollaire, R. Scrivens, C. Rossi, F. Roncarolo, S. Ramberger et al., *Linac4 design report*, vol. 6 of *CERN Yellow Reports: Monographs*, CERN, Geneva (2020), 10.23731/CYRM-2020-006. (Cited on page 24)

[71] A.M. Lombardi, "The radio frequency quadrupole (RFQ)." https://doi.org/10.5170/CERN-2006-012.201, 2006. (Cited on page 25)

[72] A. Schempp, "Radio-frequency quadrupole linacs." https://doi.org/10.5170/CERN-2005-003.305, 2005. (Cited on page 25)

[73] G. Caryotakis, *The Klystron: A Microwave source of surprising range and endurance*, Tech. Rep. SLAC-PUB-7731, SLAC, Stanford, CA (1998). (Cited on page 25)

[74] K. Hanke, *Past and present operation of the CERN PS Booster*, *Int. J. Mod. Phys. A* **28** (2013) 1330019. (Cited on page 25)

[75] U. Bigliani, G. Nassibian, K.H. Reich and D. Zanaschi, "The RF accelerating system for the CERN PS booster." https://cds.cern.ch/record/351687, 1971. (Cited on page 25)

[76] J.-P. Burnet, C. Carli, M. Chanel, R. Garoby, S. Gilardoni, M. Giovannozzi et al., *Fifty years of the CERN Proton Synchrotron*, CERN Yellow Reports: Monographs, CERN, Geneva (2011), 10.5170/CERN-2011-004. (Cited on page 26)

[77] D. Cundy and S. Gilardoni, *The Proton Synchrotron (PS): At the Core of the CERN Accelerators*, *Adv. Ser. Direct. High Energy Phys.* **27** (2017) 39. (Cited on page 26)

[78] L. Rossi, *The LHC Main Dipoles and Quadrupoles toward Series Production*, Tech. Rep. LHC-Project-Report-623, CERN-LHC-Project-Report-623 (2003). (Cited on page 27)

[79] R.M. Bianchi and A. Collaboration, "ATLAS experiment schematic or layout illustration." https://cds.cern.ch/record/2837191, 2022. (Cited on page 28)

[80] ATLAS Collaboration, *ATLAS*, LHC technical proposal, CERN, Geneva (1994), 10.17181/CERN.NR4P.BG9K. (Cited on page 28)

[81] I. Neutelings, "Figure CMS coordinate system." https://tikz.net/axis3d_cms/. (Cited on page 30)

[82] ATLAS Collaboration, "Public ATLAS Online Luminosity Plots for Run-3 of the LHC." https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun3. (Cited on page 30)

[83] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*, Tech. Rep. CERN-LHCC-2010-013, ATLAS-TDR-19 (2010). (Cited on page 32)

[84] J. Pequenao, "Computer generated image of the ATLAS inner detector." https://cds.cern.ch/record/1095926, 2008. (Cited on page 33)

[85] ATLAS Collaboration, *Technical Design Report for the ATLAS Inner Tracker Pixel Detector*, Tech. Rep. CERN-LHCC-2017-021, ATLAS-TDR-030, CERN, Geneva (2017), DOI. (Cited on page 32)

[86] J. Pequenao, "Computer Generated image of the ATLAS calorimeter." https://cds.cern.ch/record/1095927, 2008. (Cited on page 34)

[87] J. Pequenao, "Computer generated image of the ATLAS Muons subsystem." https://cds.cern.ch/record/1095929, 2008. (Cited on page 35)

[88] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, *JINST* **13** (2018) P07017. (Cited on page 37)

[89] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079 [1405.0301]. (Cited on pages 40, 79, 87, and 102)

[90] Sherpa collaboration, *Event Generation with Sherpa 2.2*, *SciPost Phys.* **7** (2019) 034 [1905.09127]. (Cited on pages 40, 79, 87, and 102)

[91] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, *JHEP* **11** (2004) 040 [hep-ph/0409146]. (Cited on pages 40, 61, 79, 87, 101, and 102)

[92] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, *JHEP* **11** (2007) 070 [0709.2092]. (Cited on pages 40, 61, 79, 87, 101, and 102)

[93] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **06** (2010) 043 [1002.2581]. (Cited on pages 40, 61, 79, 87, 101, and 102)

[94] D.E. Soper, *Parton distribution functions*, *Nucl. Phys. B Proc. Suppl.* **53** (1997) 69 [hep-lat/9609018]. (Cited on page 40)

[95] C. Bierlich et al., *A comprehensive guide to the physics and usage of PYTHIA 8.3*, *SciPost Phys. Codeb.* **2022** (2022) 8 [2203.11601]. (Cited on pages 40, 61, 79, 87, 101, and 102)

[96] GEANT4 Collaboration, *GEANT4 - A Simulation Toolkit*, *Nucl. Instrum. Meth. A* **506** (2003) 250. (Cited on pages 40, 61, and 101)

[97] S. Mehlhase, "ATLAS detector slice (and particle visualisations)." https://cds.cern.ch/record/2770815, 2021. (Cited on page 42)

[98] ATLAS Collaboration, *Reconstruction, Identification, and Calibration of hadronically decaying tau leptons with the ATLAS detector for the LHC Run 3 and reprocessed Run 2 data*, Tech. Rep. ATL-PHYS-PUB-2022-044, CERN, Geneva (2022). (Cited on page 43)

[99] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at* $\sqrt{s}$ *= 13 TeV*, *Eur. Phys. J. C* **78** (2018) 903 [1802.08168]. (Cited on page 43)

[100] ATLAS Collaboration, *The performance of missing transverse momentum reconstruction and its significance with the ATLAS detector using 140* fb^{-1} *of* $\sqrt{s} = 13$ *TeV pp collisions*, *Eur. Phys. J. C* **85** (2025) 606 [2402.05858]. (Cited on page 43)

[101] ATLAS tracking CP group, "Atlas tracking software tutorial." https://atlassoftwaredocs.web.cern.ch/internal-links/tracking-tutorial/index.html. (Cited on page 44)

[102] ATLAS Collaboration, *Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2*, *Eur. Phys. J. C* **77** (2017) 673 [1704.07983]. (Cited on page 43)

[103] R. Fruhwirth, *Application of Kalman filtering to track and vertex fitting*, *Nucl. Instrum. Meth. A* **262** (1987) 444. (Cited on page 43)

[104] T.G. Cornelissen, M. Elsing, I. Gavrilenko, J.F. Laporte, W. Liebig, M. Limper et al., *The global* χ^2 *track fitter in ATLAS*, *J. Phys. Conf. Ser.* **119** (2008) 032013. (Cited on page 43)

[105] ATLAS Collaboration, *Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC*, *Eur. Phys. J. C* **77** (2017) 332 [1611.10235]. (Cited on page 43)

[106] ATLAS Collaboration, *Tracking in Dense Environments with Transformers*, Tech. Rep. ATL-PHYS-PUB-2025-045, CERN, Geneva (2025). (Cited on page 44)

[107] ATLAS Collaboration, *Primary Vertex identification using deep learning in ATLAS*, Tech. Rep. ATL-PHYS-PUB-2023-011, CERN, Geneva (2023). (Cited on page 44)

[108] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data*, *JINST* **14** (2019) P12006 [1908.00005]. (Cited on pages 44, 45, and 131)

[109] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using LHC Run 2 data*, *JINST* **19** (2024) P02009 [2309.05471]. (Cited on pages 46 and 131)

[110] ATLAS Collaboration, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at* $\sqrt{s} = 13$ *TeV*, *Eur. Phys. J. C* **81** (2021) 578 [2012.00578]. (Cited on pages 46, 47, and 131)

[111] ATLAS Collaboration, *Performance and calibration of quark/gluon-jet taggers using 140* fb^1 *of pp collisions at TeV with the ATLAS detector*, *Chin. Phys. C* **48** (2024) 023001 [2308.00716]. (Cited on page 49)

[112] ATLAS Collaboration, *Constituent-Based Quark Gluon Tagging using Transformers with the ATLAS detector*, Tech. Rep. ATL-PHYS-PUB-2023-032, CERN, Geneva (2023). (Cited on page 49)

[113] ATLAS Collaboration, *Transforming jet flavour tagging at ATLAS*, 2505.19689. (Cited on pages 50, 56, 57, 80, and 88)

[114] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, *Eur. Phys. J. C* **77** (2017) 490 [1603.02934]. (Cited on page 50)

[115] ATLAS Collaboration, *Improving topological cluster reconstruction using calorimeter cell timing in ATLAS*, *Eur. Phys. J. C* **84** (2024) 455 [2310.16497]. (Cited on page 51)

[116] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, *Eur. Phys. J. C* **77** (2017) 466 [1703.10485]. (Cited on page 51)

[117] ATLAS Collaboration, *Optimisation of large-radius jet reconstruction for the ATLAS detector in 13 TeV proton–proton collisions*, *Eur. Phys. J. C* **81** (2021) 334 [2009.04986]. (Cited on page 51)

[118] M. Cacciari, G.P. Salam and G. Soyez, *The anti-*k_t *jet clustering algorithm*, *JHEP* **04** (2008) 063 [0802.1189]. (Cited on pages 51, 78, and 103)

[119] G.P. Salam, *Towards Jetography*, *Eur. Phys. J. C* **67** (2010) 637 [0906.1833]. (Cited on page 51)

[120] B. Malaescu, F. Napolitano, M.F. Daneri, G. Otero y Garzon, J.D. Bossio Sola, L. Henkelmann et al., *In situ calibrations for R-scan jets*, Tech. Rep. ATL-COM-PHYS-2018-407, CERN, Geneva (2018). (Cited on page 51)

[121] ATLAS Collaboration, *Determination of jet calibration and energy resolution in proton-proton collisions at $\sqrt{s}$ = 8 TeV using the ATLAS detector*, *Eur. Phys. J. C* **80** (2020) 1104 [1910.04482]. (Cited on pages 54 and 89)

[122] ATLAS Collaboration, *Using pile-up collisions as an abundant source of low-energy hadronic physics processes in ATLAS and an extraction of the jet energy resolution*, *JHEP* **12** (2024) 032 [2407.10819]. (Cited on page 55)

[123] ATLAS Collaboration, *Determination of jet calibration and energy resolution in proton-proton collisions at $\sqrt{s}$ = 8 TeV using the ATLAS detector*, *Eur. Phys. J. C* **80** (2020) 1104 [1910.04482]. (Cited on page 56)

[124] D. Camarero Muñoz, *Measurements of the inclusive isolated-photon and photon-plus-jet production in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, Tech. Rep. CERN-THESIS-2021-240 (2021). (Cited on page 56)

[125] D.J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth. A* **462** (2001) 152. (Cited on page 61)

[126] ATLAS Collaboration, *Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector*, *JHEP* **09** (2014) 145 [1406.3660]. (Cited on pages 61, 79, 87, and 102)

[127] W.-K. Tung, *New generation of parton distributions with uncertainties from global QCD analysis*, *Acta Phys. Polon. B* **33** (2002) 2933 [hep-ph/0206114]. (Cited on pages 61, 87, and 102)

[128] G. Stark, M. Milesi, J. Alison, G. Facini, K. Krizka, J. Dandoy et al., "xAODAnaHelpers, v1.0.0." https://doi.org/10.5281/zenodo.7335128, 2022. (Cited on page 61)

[129] A. Morancho Tardà, "Local xAODAnaHelpers repository under JetMET area." https://gitlab.cern.ch/atlas-jetetmiss/insitu/bJEScalibrations/AnaHelpers_Arnau/-/tree/master. (Cited on page 61)

[130] J. Pivarski, H. Schreiner, A. Hollands, P. Das, K. Kothari, A. Roy et al., "Uproot." https://doi.org/10.5281/zenodo.18435288, Jan., 2026. (Cited on page 61)

[131] T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning*, Springer (2009), 10.1007/978-0-387-84858-7. (Cited on page 65)

[132] J.H. Friedman, *Greedy function approximation: a gradient boosting machine*, *Annals of statistics* (2001) . (Cited on page 65)

[133] D. Hans Munk, "RapidGBM documentation." https://dhmunk.github.io/rapidgbm/. (Cited on page 65)

[134] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma et al., *Lightgbm: A highly efficient gradient boosting decision tree*, . (Cited on page 65)

[135] J. Mockus, V. Tiesis and A. Zilinskas, *The application of bayesian methods for seeking the extremum*, *Towards global optimization* (1978) . (Cited on page 65)

[136] J. Ansel, E. Yang, H. He, N. Gimelshein, A. Jain, M. Voznesensky et al., *PyTorch 2: Faster Machine Learning Through Dynamic Python Bytecode Transformation and Graph Compilation*, Tech. Rep. 10.1145/3620665.3640366 (2024). (Cited on page 65)

[137] NNPDF collaboration, *Parton distributions for the LHC Run II*, *JHEP* **04** (2015) 040 [1410.8849]. (Cited on pages 79, 87, and 102)

[138] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, Tech. Rep. ATL-PHYS-PUB-2014-021, Geneva (2014). (Cited on pages 79, 87, 101, and 102)

[139] R. Frederix and S. Frixione, *Merging meets matching in MC@NLO*, *JHEP* **12** (2012) 061 [1209.6215]. (Cited on pages 79 and 87)

[140] K. Hamilton, P. Nason and G. Zanderighi, *MINLO: Multi-Scale Improved NLO*, *JHEP* **10** (2012) 155 [1206.3572]. (Cited on page 79)

[141] K. Hamilton, P. Nason, C. Oleari and G. Zanderighi, *Merging H/W/Z + 0 and 1 jet at NLO with no merging scale: a path to parton shower + NNLO matching*, *JHEP* **05** (2013) 082 [1212.4504]. (Cited on page 79)

[142] R.D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys. B* **867** (2013) 244 [1207.1303]. (Cited on pages 79, 87, and 102)

[143] M. Cacciari, G.P. Salam and G. Soyez, *FastJet User Manual*, *Eur. Phys. J. C* **72** (2012) 1896 [1111.6097]. (Cited on page 78)

[144] ATLAS Collaboration, *Transformer Neural Networks for Identifying Boosted Higgs Bosons decaying into $b\bar{b}$ and $c\bar{c}$ in ATLAS*, Tech. Rep. ATL-PHYS-PUB-2023-021, CERN, Geneva (2023). (Cited on page 80)

[145] S. Shleifer, J. Weston and M. Ott, *NormFormer: Improved Transformer Pretraining with Extra Normalization*, 2110.09456. (Cited on page 81)

[146] J.L. Ba, J.R. Kiros and G.E. Hinton, *Layer Normalization*, 1607.06450. (Cited on page 81)

[147] J. Bellm et al., *Herwig 7.2 release note*, *Eur. Phys. J. C* **80** (2020) 452 [1912.06509]. (Cited on page 87)

[148] L.A. Harland-Lang, A.D. Martin, P. Motylinski and R.S. Thorne, *Parton distributions in the LHC era: MMHT 2014 PDFs*, *Eur. Phys. J. C* **75** (2015) 204 [1412.3989]. (Cited on page 87)

[149] ATLAS Collaboration, *Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Eur. Phys. J. C* **81** (2021) 689 [2007.02645]. (Cited on page 90)

[150] M. Carena, I. Low and C.E.M. Wagner, *Implications of a Modified Higgs to Diphoton Decay Width*, *JHEP* **08** (2012) 060 [1206.1082]. (Cited on page 100)

[151] C.-W. Chiang and K. Yagyu, *Higgs boson decays to $\gamma\gamma$ and $Z\gamma$ in models with Higgs extensions*, *Phys. Rev. D* **87** (2013) 033003 [1207.1065]. (Cited on page 100)

[152] C.-S. Chen, C.-Q. Geng, D. Huang and L.-H. Tsai, *New Scalar Contributions to $h \to Z\gamma$*, *Phys. Rev. D* **87** (2013) 075019 [1301.4694]. (Cited on page 100)

[153] A. Djouadi, V. Driesen, W. Hollik and A. Kraft, *The Higgs photon - Z boson coupling revisited*, *Eur. Phys. J. C* **1** (1998) 163 [hep-ph/9701342]. (Cited on page 100)

[154] H.T. Hung, T.T. Hong, H.H. Phuong, H.L.T. Mai and L.T. Hue, *Neutral Higgs decays $H \to Z\gamma, \gamma\gamma$ in 3-3-1 models*, *Phys. Rev. D* **100** (2019) 075014 [1907.06735]. (Cited on page 100)

[155] P. Archer-Smith, D. Stolarski and R. Vega-Morales, *On new physics contributions to the Higgs decay to $Z\gamma$*, *JHEP* **10** (2021) 247 [2012.01440]. (Cited on page 100)

[156] X.-G. He, Z.-L. Huang, M.-W. Li and C.-W. Liu, *The SM expected branching ratio for $h \to \gamma\gamma$ and an excess for $h \to Z\gamma$*, *JHEP* **10** (2024) 135 [2402.08190]. (Cited on page 100)

[157] A. Azatov, R. Contino, A. Di Iura and J. Galloway, *New Prospects for Higgs Compositeness in $h \to Z\gamma$*, *Phys. Rev. D* **88** (2013) 075019 [1308.2676]. (Cited on page 100)

[158] I. Low, J. Lykken and G. Shaughnessy, *Singlet scalars as Higgs imposters at the Large Hadron Collider*, *Phys. Rev. D* **84** (2011) 035027 [1105.4587]. (Cited on page 100)

[159] I. Low, J. Lykken and G. Shaughnessy, *Have We Observed the Higgs (Imposter)?*, *Phys. Rev. D* **86** (2012) 093012 [1207.1093]. (Cited on page 100)

[160] ATLAS Collaboration, *A search for the $Z\gamma$ decay mode of the Higgs boson in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Lett. B* **809** (2020) 135754 [2005.05382]. (Cited on pages 100 and 111)

[161] CMS Collaboration, *Search for Higgs boson decays to a Z boson and a photon in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **05** (2023) 233 [2204.12945]. (Cited on page 100)

[162] PDF4LHC Working Group collaboration, *The PDF4LHC21 combination of global PDF fits for the LHC Run III*, *J. Phys. G* **49** (2022) 080501 [2203.05506]. (Cited on page 101)

[163] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, *Eur. Phys. J. C* **76** (2016) 581 [1510.03823]. (Cited on page 103)

[164] ATLAS Collaboration, *Forward jet vertex tagging using the particle flow algorithm*, Tech. Rep. ATL-PHYS-PUB-2019-026, CERN, Geneva (2019). (Cited on page 103)

[165] T. Chen and C. Guestrin, *XGBoost: A Scalable Tree Boosting System*, 1603.02754. (Cited on pages 104 and 116)

[166] ATLAS Collaboration, *Measurement of the inclusive isolated prompt photon cross section in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *Phys. Rev. D* **83** (2011) 052005 [1012.4389]. (Cited on page 107)

[167] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J. C* **71** (2011) 1554 [1007.1727]. (Cited on page 108)

[168] ATLAS Collaboration, "H2Zy Framework." https://gitlab.cern.ch/atlas-hgam-sw/h-Zy/H2Zy. (Cited on page 118)

[169] A. Morancho Tardà, "Development branch from H2Zy Framework." https://gitlab.cern.ch/atlas-hgam-sw/h-Zy/H2Zy/-/tree/dev2_nbi?ref_type=heads. (Cited on page 118)

[170] J. Bai, F. Lu, K. Zhang et al., "ONNX: Open Neural Network Exchange." https://github.com/onnx/onnx, 2021. (Cited on page 118)

[171] ATLAS Collaboration, "ATLAS simulated samples collection for jet reconstruction training, as part of the 2020 open data release." https://doi.org/10.7483/OPENDATA.ATLAS.L806.5CKU, 2020. (Cited on page 136)