Electroweak production of Z boson with two jets and determination of heavy boson densities

Dissertation
zur Erlangung des Doktorgrades an der
Fakultat für Mathematik, Informatik und
Naturwissenschaften Fachbereich Physik der
Universität Hamburg

vorgelegt von
Keila Moral Figueroa
aus Zamora, Spanien

Hamburg
2026

Eidesstattliche Versicherung / Declaration on oath

Hiermit versichere ich an Eides statt, die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt zu haben. Sofern im Zuge der Erstellung der vorliegenden Dissertationsschrift generative Künstliche Intelligenz (gKI) basierte elektronische Hilfsmittel verwendet wurden, versichere ich, dass meine eigene Leistung im Vordergrund stand und dass eine vollständige Dokumentation aller verwendeten Hilfsmittel gemäß der Guten wissenschaftlichen Praxis vorliegt. Ich trage die Verantwortung für eventuell durch die gKI generierte fehlerhafte oder verzerrte Inhalte, fehlerhafte Referenzen, Verstöße gegen das Datenschutz- und Urheberrecht oder Plagiate.

I hereby declare and affirm that this doctoral dissertation is my own work and that I have not used any aids and sources other than those indicated. If electronic resources based on generative artificial intelligence (gAI) were used in the course of writing this dissertation, I confirm that my own work was the main and value-adding contribution and that complete documentation of all resources used is available in accordance with good scientific practice. I am responsible for any erroneous or distorted content, incorrect references, violations of data protection and copyright law or plagiarism that may have been generated by the gAI.

Hamburg, February 9, 2026

Keila Moral Figueroa

Ich versichere, dass dieses gebundene Exemplar der Dissertation und das in elektronischer Form eingereichte Dissertationsexemplar (über den Docata-Upload) und das bei der Fakultät (zuständiges Studienbüro bzw. Promotionsbüro Physik) zur Archivierung eingereichte gedruckte gebundene Exemplar der Dissertationsschrift identisch sind.
Hamburg, 9. Februar, 2026

Keila Moral Figueroa

I, the undersigned, declare that this bound copy of the dissertation and the dissertation submitted in electronic form (via the Docata upload) and the printed bound copy of the dissertation submitted to the faculty (responsible Academic Office or the Doctoral Office Physics) for archiving are identical.

Hamburg, February 9, 2026

Keila Moral Figueroa

Gutachter der Dissertation:	Dr. Katarzyna Wichmann Prof. Dr. Elisabetta Gallo
Zusammensetzung der Prüfungskommission:	Prof. Dr. Sven-Olaf Moch Prof. Dr. Johannes Haller Prof. Dr. Elisabetta Gallo Dr. Katarzyna Wichmann Dr. Krisztian Peters
Vorsitzender der Prüfungskommission:	Prof. Dr. Sven-Olaf Moch
Datum der Disputation:	26.03.2026
Vorsitzender des Fach-Promotionsausschusses Physik:	Prof. Dr. Johannes Haller
Leiter des Fachbereichs Physik:	Prof. Dr. Markus Drescher
Dekan der Fakultät MIN:	Prof. Dr.-Ing. Norbert Ritter

Abstract

In this thesis the electroweak production of the Z boson in association with two jets is investigated focusing on its topology via the Vector Boson Fusion (VBF) mechanism and the determination of heavy boson parton densities.
The first part presents a theoretical study of the Drell–Yan and electroweak Zjj production mechanisms and their particle-level comparison, demonstrating the dominance of the former and motivating the definition of a VBF-enhanced phase space to isolate electroweak contributions. The impact of electroweak corrections is discussed as a motivation for introducing electroweak (heavy boson) densities. These densities, derived from extended DGLAP evolution equations, are successfully validated using HERA DIS data and applied to VBF Z production to assess their relevance for t-channel boson dynamics and future collider studies.
The second part reports fiducial and differential cross-section measurements of Z-boson production with dijets in the $Z \to e^+e^-$, $Z \to \mu^+\mu^-$ channels and the combination of both, analysed within a common phase space. Both inclusive (QCD Zjj + EW Zjj) and electroweak (EW Zjj) components are measured as functions of the dilepton transverse momentum, p_T^{ll}. The inclusive measurement uses 59.83 fb^{-1} of 13 TeV CMS data from 2018, while the main result, the electroweak Zjj, exploits the full Run 2 dataset of 138 fb^{-1} at $\sqrt{s}$ = 13 TeV. The Run 2 electroweak Zjj measurement achieves a total precision of about 6%, establishing a new benchmark for EW Zjj studies.

Zusammenfassung

In dieser Arbeit wird die elektroschwache Produktion des Z-Bosons in Verbindung mit zwei Jets untersucht mit besonderem Fokus auf deren Topologie im Rahmen des Vector-Boson-Fusion (VBF)-Mechanismus sowie auf die Bestimmung von Partondichten schwerer Bosonen.

Im ersten Teil wird eine theoretische Analyse der Drell–Yan- und elektroschwachen Zjj-Produktionsmechanismen sowie deren Vergleich auf Teilchenebene vorgestellt. Dabei wird die Dominanz des Drell–Yan-Mechanismus gezeigt, was die Definition eines VBF-optimierten Phasenraums zur Isolierung elektroschwacher Beiträge motiviert. Der Einfluss elektroschwacher Korrekturen wird erörtert und dient als Grundlage für die Einführung elektroschwacher (schwerer Bosonen-)Dichten. Diese Dichten, abgeleitet aus erweiterten DGLAP-Entwicklungsgleichungen, werden erfolgreich mit HERA-DIS-Daten validiert und anschließend auf die VBF-Z-Produktion angewendet, um ihre Relevanz für die Dynamik von t-Kanal-Bosonen sowie für zukünftige Kollider-Studien zu zeigen.

Der zweite Teil berichtet über sichtbare und differentielle Wirkungsquerschnittsmessungen von Z-Boson-Produktion mit Dijets in den Kanälen $Z \to e^+e^-$, $Z \to \mu^+\mu^-$ sowie deren Kombination, analysiert innerhalb eines gemeinsamen Phasenraums. Sowohl die inklusiven (QCD Zjj + EW Zjj) wie auch die elektroschwachen (EW Zjj) Komponenten werden als Funktionen des transversalen Dileptonimpulses p_T^{ll} gemessen. Die inklusive Messung basiert auf 59.83 fb^{-1} an 13 TeV-CMS-Daten von 2018, während das Hauptergebnis, die elektroschwache Zjj-Messung, das vollständige Run-2-Datenset von 138 fb^{-1} bei $\sqrt{s} = 13$ TeV verwendet. Die Run-2-Messung erreicht eine Gesamtexaktheit von etwa 6%, und setzt damit einen neuen Maßstab für Studien der elektroschwachen Zjj-Produktion.

Contents

Introduction

The Standard Model of particle physics stands as one of the most successful theoretical frameworks in modern science, providing a comprehensive description of fundamental particles and their interactions through quantum field theory [1]. Its formulation in the 1970s unified electromagnetism and the weak force into the electroweak theory, predicting the existence of the W and Z bosons as carriers of the weak interaction. The experimental confirmation of these particles at CERN in 1983 [2, 3] also validated the electroweak sector [4]. This progression has enabled rigorous tests of the Standard Model, including detailed investigations of electroweak processes such as vector boson fusion, which continue to play a pivotal role in refining our understanding of fundamental interactions and probing for new physics.

The study of electroweak Zjj production, especially in the vector boson fusion (VBF) topology, serves as a direct probe of the electroweak sector. Mediated by t-channel weak boson exchange, this process is sensitive to the structure of electroweak couplings. Unlike QCD-dominated Zjj production, VBF Z allows for stringent tests of anomalous triple gauge couplings [5], making it a benchmark for both theoretical calculations and experimental methodologies. The process is also highly suitable for precision phenomenology, facilitating the validation of advanced computational techniques such as the matching of fixed-order predictions with parton showers and the investigation of colour-singlet exchange, which results in reduced hadronic activity in the central region [6]. This makes VBF Z production an essential tool for studying PDF effects, initial- and final-state radiation, and resummation, all within a well-controlled environment.

From a beyond-the-Standard-Model (BSM) perspective, VBF Z production plays a key role as a calibration channel for Higgs physics and new physics searches. Sharing the same experimental topology as VBF Higgs production [7], two forward jets with a large rapidity gap and a central electroweak boson, VBF Z measurements help validate the theoretical and experimental frameworks used in searches for the Higgs boson and for BSM phenomena. The clean final state from Z-boson decays to charged leptons allows for precise differential measurements and the rigorous testing of theory calculations. Furthermore, the ability to separate VBF Z production from QCD backgrounds using characteristic kinematic cuts, such as large dijet invariant mass and rapidity separation, challenges our understanding of electroweak versus QCD processes and provides a stringent test of theoretical modelling, making it a cornerstone in the quest for new physics at high-energy colliders.

This thesis is structured as follows. Chapter 1 provides an overview of the Standard Model of particle physics and the dynamics of high-energy proton–proton collisions. Chapter 2 discusses the theoretical framework and phenomenology of Z boson production, beginning with the inclusive Z+2 jets process and subsequently focusing on the Vector Boson Fusion (VBF) topology, followed by a phenomenological comparison between both. Chapter 3 examines the electroweak contributions to these processes, with particular attention to their impact on parton distribution functions and the role of heavy boson densities in VBF Z calculations. Chapters 4 describes the Large Hadron Collider and the CMS experiment and Chapter 5 explains the reconstruction of events and physics objects within CMS. Chapter 6 presents the detailed measurement of the electroweak Z+2 jets production, while Chapter 7 reports the measured cross section for inclusive Zjj production and highlights the main result of this work, the measurement of the electroweak Zjj process. Finally, Chapter 8 summarises the conclusions.

Chapter 1

Theoretical Framework

This chapter introduces the theoretical framework relevant for describing hadronic collisions. Section 1.1 provides a brief overview of the Standard Model (SM), the theory describing the interactions of the fundamental particles. Section 1.2 explores proton collisions at high energies, this includes QCD factorization, a formal discussion on parton distributions, resummation and hadronisation.

1.1 The Standard Model of particle physics

The Standard Model (SM) is a well-established framework describing all known fundamental particles and their interactions [8]. Despite its success in reproducing experimental results, it remains incomplete: it does not incorporate gravitation and fails to explain dark matter or dark energy. These limitations indicate the existence of physics beyond the SM (BSM). To probe such phenomena and improve the precision of established measurements, high-energy particle collisions are studied at facilities such as the LHC.

From a theoretical standpoint, the SM is formulated as a Quantum Field Theory (QFT), which represents the extension of quantum mechanics to systems of dynamical fields. Within this framework, particles correspond to quantized excitations of fields defined over the entire spacetime manifold. It describes the strong, weak, and electromagnetic interactions, while the fourth fundamental force, gravitation, lies outside its scope. The SM is constructed on the principle of least action and can be expressed in terms of a Lagrangian formalism [9, 10]. The Lagrangian density, $\mathcal{L}$, encodes the fields and their interactions, from which the corresponding equations of motion are derived. As the SM is a gauge theory, its fundamental interactions arise from local internal symmetries of $\mathcal{L}$. By Noether's theorem, each such symmetry implies a conserved quantity, corresponding to a quantum number [11]. The complete SM Lagrangian density, $\mathcal{L}$ is obtained by combining the contributions of the individual interactions. Quantum Chromodynamics (QCD), based on the $SU(3)_C$ symmetry, describes the strong interaction, where the subscript C denotes the colour quantum number (see Section 1.1.3). Electromagnetism and the weak interaction are unified in the Electroweak (EW) theory (see Section 1.1.2). The EW theory is based on an $SU(2)_L \times U(1)_Y$ gauge symmetry, with associated quantum numbers given by the third component of the weak isospin T_3 and the hypercharge Y. The subscript L denotes the left-chiral nature of the weak interaction (see Section 1.1.3).

In terms of gauge symmetries, the Standard Model can thus be written as

$$SU(3)_C \times SU(2)_L \times U(1)_Y \tag{1.1}$$

where $SU(N)$ stands for special unitary algebraic groups that act on N dimensions. Each group has $N^2 - 1$ objects that generated the group, called generators. The generators can be represented through $N \otimes N$ matrices, for $SU(2)$ the three Pauli matrices and for $SU(3)$ the eight Gell-Mann matrices. These generate the corresponding internal symmetry transformations. The $SU(3)_c$ matrices act on the colour space and $SU(2)_L$ on weak isospin space.
The fundamental fields of the Standard Model (SM) (see Fig. 1.1) consist of fermion fields, with spin $\frac{1}{2}$, and vector boson fields, with spin 1. Fermions are described by spinors, which are two-component fields transforming under Lorentz symmetry. Weyl spinors are the basic building blocks: left-chiral Weyl spinors, χ_L, transform as $(\frac{1}{2}, 0)$, while right-chiral Weyl spinors, ξ_R, transform as $(0, \frac{1}{2})$. Chirality is an intrinsic property defined by the transformation under Lorentz symmetry. To describe massive fermions in the SM, these Weyl spinors are combined into four-component Dirac fields, which satisfy the Dirac equation [12]. Parity, the spacetime transformation that inverts spatial coordinates, acts on a Dirac field by exchanging its left- and right-chiral components. Accounting for chirality, the fermion fields of the SM are represented as Dirac fields:

$$\psi = \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix}, \quad \chi_L = (1/2, 0), \quad \xi_R = (0, 1/2), \tag{1.2}$$

which are four-component fields. Fermions are grouped into quark and lepton families, each organized in three generations. Quark isospin doublets contain an up-type quark with electric charge +2/3 and a down-type quark with charge −1/3. The six quark states, arising from two types across three generations, are referred to as flavours. In the lepton sector, each charged lepton (electric charge −1) has a corresponding neutrino (neutral) as its isospin partner. Finally, the EW theory requires the introduction of a scalar (spin 0) field, the Higgs field, discussed in Section 1.1.3.

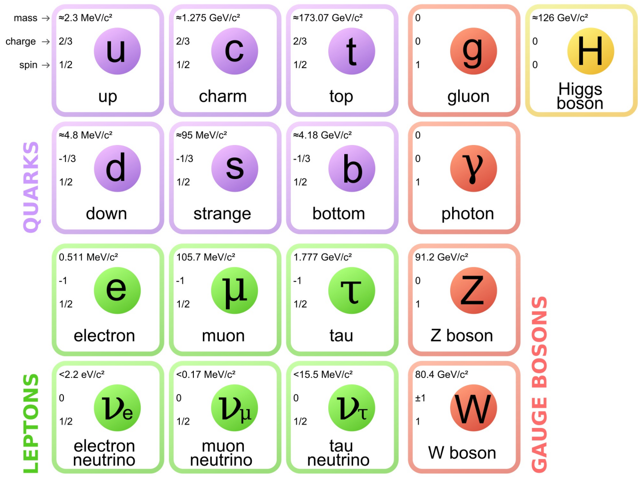


Figure 1.1: The particle content of the SM. For each particle, a measured value for its mass, its electric charge and spin quantum numbers are included [13].

1.1.1 The Feynman diagram representation

Richard Feynman introduced a graphical representation of interactions between fundamental fields, known as Feynman diagrams (see Fig. 1.2). These diagrams provide both an intuitive visualization of quantum field theory (QFT) processes and a systematic framework for computing their probability amplitudes [14].

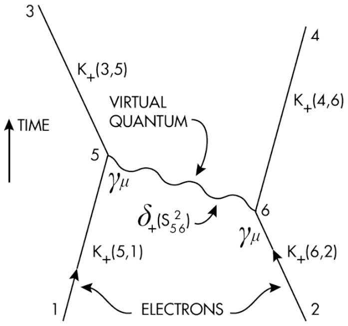


Figure 1.2: The first published Feynman diagram [14].

In this formalism, time is conventionally read from left to right. Straight lines denote fermions and antifermions, with arrows indicating fermion flow: left-to-right for fermions and right-to-left for antifermions. Curly lines represent gluons, the mediators of the strong interaction (see Section 1.2), wavy lines indicate electroweak gauge bosons, and dashed lines denote the Higgs boson (see Section 1.3). Each diagram implicitly incorporates all possible time orderings of a given process.

For a given initial and final state, an infinite number of Feynman diagrams can, in principle, be constructed, differing in the number of interaction vertices. According to the Feynman rules, each diagram is mapped onto a mathematical expression, and the total probability amplitude is obtained by summing over all possible contributions. Since this sum contains infinitely many terms, a systematic approach is required to extract finite and physically meaningful predictions.

Perturbation theory

A practical framework to address this problem is provided by perturbation theory. When the coupling of the interaction is less than unity, contributions from diagrams with additional vertices are progressively suppressed. In this regime, the amplitude can be organised as a power series in the coupling: starting with diagrams containing the minimal number of vertices (leading order, LO), and systematically improving the approximation by including higher-order corrections such as next-to-leading order (NLO) and next-to-next-to-leading order (NNLO).

Renormalization theory

Perturbation theory provides a systematically improvable framework for quantum field calculations but often yields intermediate expressions containing divergences. These arise, in particular, from loop diagrams in which virtual particles can carry unbounded momenta, leading to divergent integrals in the amplitude computation. To render the theory predictive, a regularisation procedure is introduced, such as imposing an arbitrary cutoff scale Λ, which suppresses contributions from momenta much larger than this scale. The resulting quantities must then be reformulated in terms of renormalised parameters, which are finite and correspond directly to experimentally measurable values. Renormalisation introduces a scale dependence, with physical parameters defined at a reference value known as the renormalisation scale μ_R. In practice, μ_R is chosen to be of the order of the momentum transfer characteristic of the process under consideration, thereby reducing the size of higher-order corrections in the perturbative expansion.

1.1.2 Electroweak interactions

The electroweak theory originates from the unification of quantum electrodynamics and the weak interaction. Quantum electrodynamics is a relativistic $U(1)$ gauge theory describing the interaction of charged fermions with the electromagnetic field. The weak interaction, in contrast, is a chiral non-Abelian gauge theory mediated by the charged $W^{\pm}$ and neutral Z^0 bosons, governing both flavor-changing and flavor-conserving decays of quarks and leptons [8].

Quantum electrodynamics
Quantum electrodynamics is a relativistic $U(1)$ gauge theory describing the interaction of charged fermions with the electromagnetic field, therefore it describes the interaction between matter and light.
Defining the Dirac Lagrangian as

$$\mathcal{L}_{Dirac} = \bar{\psi}\left(i\partial\!\!\!/ - m\right)\psi \quad , \tag{1.3}$$

the ψ corresponds to the doublet of spinors previously introduced in Eq. 1.2, m is the particle mass and γ are the Pauli matrices such that $\bar{\psi} = \psi^{\dagger}\gamma^0$. Due to its $U(1)$ gauge nature, the Dirac Lagrangian is invariant under $U(1)$ transformations of the spinor fields:

$$\psi \to e^{-i\alpha}\psi \quad \text{and} \quad \bar{\psi} \to \mathrm{e}^{+\mathrm{i}\alpha}\bar{\psi}, \quad \alpha \in \mathcal{R} \tag{1.4}$$

leading to a conserved current, $\gamma^{\mu}_{V} = \bar{\psi}\gamma^{\mu}\psi$, the electromagnetic charge. However, it is not invariant under a $U(1)$ local transformation, in which case the term α from Eq. 1.4 becomes a function in terms of space-time instead of a real number. Therefore, in order to restore invariance under a $U(1)$ local transformation, a gauge covariant derivative D_{μ} is introduced to describe the interaction between the fermion and the field:

$$D_{\mu} = \partial_{\mu} + ieA_{\mu}\,, \tag{1.5}$$

where e is the coupling strength understood as the electric charge of the particle ψ and A_{μ} is the electromagnetic vector potential, and under gauge transformations it acts as

$$A_{\mu} \to A_{\mu} - \frac{1}{e}\partial_{\mu}\alpha(x)\,. \tag{1.6}$$

Subsequently, by promoting the partial derivative into the gauge covariant derivative in Eq. 1.3, the theory of electromagnetism coupled to fermions is described by the following Lagrangian:

$$\mathcal{L}_{\text{QED}} = \overline{\psi}(i D\!\!\!\!/ - m)\psi - \frac{1}{4}(F_{\mu\nu})^2. \tag{1.7}$$

Here, $F_{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu}$ denotes the electromagnetic field-strength tensor. The QED Lagrangian $\mathcal{L}_{\text{QED}}$ is invariant under local $U(1)$ gauge transformations of both the fermion field ψ and the gauge field A_{μ}, with the latter identified as the photon, the mediator of electromagnetic interactions.

Weak interactions

The theory of weak interactions is a quantitative framework first proposed by E. Fermi in 1933 [15] to describe processes with a low coupling and an extremely short interaction range such as β decays [16]:

$$n \to p + e^- + \bar{\nu}^e, \quad p \to n + e^+ + \nu^e \, , \tag{1.8}$$

where n is the neutron, p is the proton, $e^-(e^+)$ is the electron (positron) and $\bar{\nu}^e(\nu^e)$ is the electron antineutrino (neutrino). The original proposal by Fermi consisted of the weak process described as a contact Lagrangian, suppressing any possible mediator:

$$\mathcal{L}_{\text{Fermi}} = -G_F \, \bar{\psi}_p \gamma^\mu \psi_n \, \bar{\psi}_e \gamma_\mu \psi_\nu \, , \tag{1.9}$$

where G_F denotes the Fermi constant, Fermi's four-fermion theory successfully described electron energy spectra, nuclear selection rules, and decay rates, while providing a consistent role for the neutrino and unifying various nuclear beta decays through a universal coupling. Nevertheless, the theory was non-renormalizable and assumed parity conservation, thus restricting its validity to low-energy processes.

Electroweak theory unification

These shortcomings were resolved with the development of the electroweak theory, formulated by S. Glashow [17], S. Weinberg [18], and A. Salam [19], which unified the weak interaction with QED. In this framework, weak processes are mediated by the exchange of massive $W^\pm$ and Z^0 bosons. The chiral V-A structure of the theory naturally incorporates the observed phenomenon of maximal parity violation.
The electroweak theory is a $SU(2)\otimes U(1)$ gauge theory described by the Lagrangian:

$$\mathcal{L}_{EW} = \mathcal{L}_{YM} + \mathcal{L}_{Higgs} \, , \tag{1.10}$$

where $\mathcal{L}_{YM}$ is the Yang-Mills Lagrangian for the $SU(2) \otimes U(1)$ gauge group with fermion matter fields, which describes only gauge bosons and fermions, and $\mathcal{L}_{Higgs}$ is the Higgs Lagrangian, which describes the Brout-Englert-Higgs mechanism, as its name suggests. The Yang-Mills Lagrangian is defined as:

$$\mathcal{L}_{\text{YM}} = -\frac{1}{4}\sum_{A=1}^{3} F^A_{\mu\nu} F^{A\,\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} + \bar{\psi}_L i\gamma^\mu D_\mu \psi_L + \bar{\psi}_R i\gamma^\mu D_\mu \psi_R \, , \tag{1.11}$$

where

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \quad \text{and} \quad F^A_{\mu\nu} = \partial_\mu W^A_\nu - \partial_\nu W^A_\mu - g\epsilon_{ABC} W^B_\mu W^C_\nu \, , \tag{1.12}$$

are the gauge antisymmetric tensors constructed out of the gauge field B_μ associated with the $U(1)$ group and $W^A_\mu, W^B_\mu, W^C_\mu$ corresponding to the three $SU(2)$ generators. Additionally, the ϵ_{ABC} is the antisymmetric Levi-Civita tensor understood as the group structure constants and the interaction of the gauge fields with the fermions is introduced by the covariant derivative:

$$D_\mu \psi_{L,R} = \left[\partial_\mu + ig \sum_{A=1}^{3} t^A_{L,R} W^A_\mu + ig' \frac{1}{2} Y_{L,R} B_\mu \right] \psi_{L,R} \, . \tag{1.13}$$

Here, g and g' are the coupling constants of the $SU(2)$ and $U(1)$ gauge fields W^a_μ and B_μ, respectively, while $t^A_{L,R}$ and $\frac{1}{2}Y_{L,R}$ are their generators in the representations

of $\psi_{L,R}$. The self-interaction of the $W_{\mu,\nu}^{A,B,C}$ fields appears from the non-Abelian structure of the $SU(2)$ group, where the generators do not commute.
The invariance of the Lagrangian under local transformations of $SU(2)_L \otimes U(1)_Y$ implies the existence of two associated conserved currents: the weak isospin current corresponding to the $SU(2)_L$ symmetry and the weak hypercharge current corresponding to the $U(1)_Y$ symmetry. The physical electric charge, Q, arises as a linear combination of these two conserved quantities, as expressed by the Gell-Mann-Nishijima relation:

$$Q = T_3 + \frac{Y}{2}, \tag{1.14}$$

where T_3 is the third component of weak isospin and Y is the weak hypercharge.
Experimentally, weak interactions couple only to left-chiral fermions, which transform as doublets under $SU(2)_L$. Right-handed fermions are singlets under $SU(2)_L$ and thus do not transform. In contrast, the $U(1)_Y$ gauge field couples to both chiralities according to their hypercharge. As a result, fermion fields are defined as:

$$\psi_{L,R} = \left[\frac{1 \mp \gamma_5}{2}\right]\psi, \qquad \overline{\psi}_{L,R} = \overline{\psi}\left[\frac{1 \pm \gamma_5}{2}\right], \tag{1.15}$$

where γ^5 is a Dirac matrix defined in [20]. These fermion fields can be re-written for leptons as:

$$L_l = \begin{pmatrix} \nu_l \\ l \end{pmatrix}_L, \tag{1.16}$$

where l represents the charged lepton and ν_l its neutrino. In the case of quarks, fermion fields are built with quark generations, for example:

$$L_q = \begin{pmatrix} u_L \\ d'_L \end{pmatrix}_L, \tag{1.17}$$

where u_L is the up quark in its mass eigenstate and d'_L is a mixed state of up quarks and down quarks respectively through the Cabibbo–Kobayashi–Maskawa matrix.
From Eq. 1.11, Eq. 1.13 it is immediate to derive all fermion couplings to the gauge bosons and their definitions. In the charged-current (CC) sector, the charged weak boson $W^{\pm}$ is defined as:

$$W^{\pm} = \frac{W^1 \mp iW^2}{\sqrt{2}}. \tag{1.18}$$

In the neutral-current (NC) sector, the photon, A_μ, and the neutral weak boson Z_μ, emerge as orthogonal and properly normalized linear combinations of the gauge fields B_μ associated with the $U(1)_Y$ symmetry and W_μ^3, the third component of the $SU(2)_L$ gauge:

$$A_\mu = \cos\theta_W B_\mu + \sin\theta_W W_\mu^3, \tag{1.19}$$

$$Z_\mu = -\sin\theta_W B_\mu + \cos\theta_W W_\mu^3, \tag{1.20}$$

where θ_W denotes the weak mixing angle. Explicit mass terms for gauge bosons are forbidden by gauge invariance, yet experimental results confirm that the $W^{\pm}$ and Z^0 bosons are massive. Hence, the $SU(2) \otimes U(1)$ symmetry is spontaneously broken. Although the Lagrangian is invariant under global and local transformations, the vacuum does not preserve the full symmetry, therefore there is a broken internal symmetry. According to the Goldstone theorem [21], each broken continuous symmetry gives rise to a massless scalar, the Goldstone boson. Englert and Brout [22] and Higgs [4] showed that in a local gauge theory these modes are absorbed, giving mass to the $W^{\pm}$ and Z^0 bosons. This process, known as the Brout-Englert-Higgs mechanism, also predicts a massive scalar particle, the Higgs boson.

The Brout-Englert-Higgs mechanism

The Brout-Englert-Higgs mechanism, originally proposed to generate gauge boson masses, was subsequently incorporated into the electroweak theory by Weinberg [18] and Salam [4], thereby providing a mechanism to explain the masses of three gauge bosons mediating the electroweak interactions.

In a first approach, two complex scalar fields are defined so that they transform as a doublet under $SU(2)$, $\Phi = (\Phi^+, \Phi^0)$. In this scenario, the Lagrangian is defined as

$$\mathcal{L} = (D_\mu \Phi)^\dagger (D^\mu \Phi) - V(\Phi^\dagger \Phi) \,, \tag{1.21}$$

where the potential $V(\Phi^\dagger \Phi)$ is the Higgs potential:

$$V(\Phi^\dagger \Phi) = -\mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2 \,, \tag{1.22}$$

where μ is the vacuum parameter and λ the quartic Higgs coupling. For $\lambda < 0$, the potential has no minimum, rendering the theory non-physical. Imposing $\lambda > 0$, there are two cases: $\mu^2 < 0$ gives a trivial minimum at $\Phi_{\text{min}} = 0$, while $\mu^2 > 0$ yields a non-trivial minimum

$$\Phi_{\text{min}} = \sqrt{\frac{\mu^2}{2\lambda}} e^{i\varphi}. \tag{1.23}$$

The potential has a continuous set of minima parametrized by φ, forming a circle of radius $\frac{\mu}{\sqrt{2\lambda}}$ in the complex Φ-plane. This produces the characteristic "sombrero" shape of $V(\Phi)$ (see Fig. 1.3), where spontaneous symmetry breaking corresponds to selecting a specific vacuum among infinitely many.

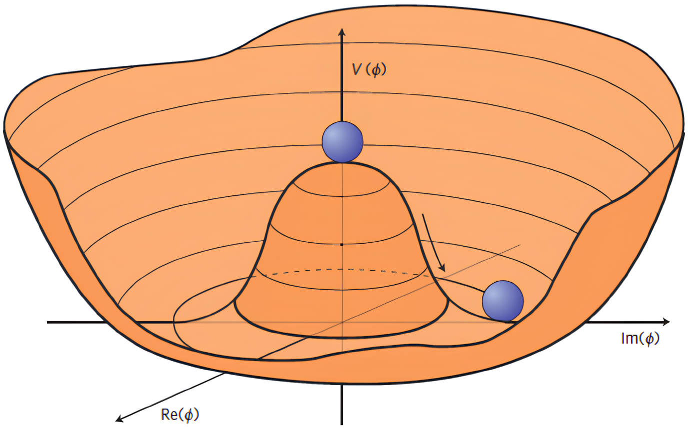


Figure 1.3: Illustration of the Higgs potential in the case of $\mu^2 < 0$. The purple marble illustrates the choice of a point at the minimum of the potential, which breaks spontaneously the symmetry [23].

Now, the Φ field can be decomposed in a radial $\sigma(x)$ and angular $\pi(x)$ fields around the non-trivial minimum, separating into components that capture amplitude variations and those corresponding to angular excitations:

$$\Phi(x) = e^{\frac{i}{v}\pi^a(x)\tau^a} \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}}\left(v + \sigma(x)\right) . \end{pmatrix} \tag{1.24}$$

In this model, $\nu = \sqrt{\mu^2/\lambda}$ is the vacuum expetation value, $\pi(x)$ corresponds to the Goldstone boson, while $\sigma(x)$ is a massive boson with mass $m_\sigma = \sqrt{2\mu^2}$. The Goldstone bosons are eliminated by fixing the exponential parametrisation to zero:

$$\Phi(x) = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}}\left(v + \sigma(x)\right) . \end{pmatrix} \tag{1.25}$$

As a result, the radial field, $\sigma(x)$ is identified as the Higgs field: $\sigma(x) = H(x)$ and the Higgs Lagrangian is:

$$\mathcal{L}_{\text{Higgs}} = (D_\mu\Phi)^\dagger(D_\mu\Phi) - \lambda\left(\Phi^\dagger\Phi - \frac{v^2}{2}\right)^2 , \tag{1.26}$$

where the covariant derivative is given by Eq. 1.13 and W_μ^A are the gauge fields corresponding to $SU(2)$ and B_μ to $U(1)$. Thus, the gauge fields of $SU(2)$ become massive following the spontaneous breaking of the $SU(2) \otimes U(1)$ symmetry. This broken symmetry induces a new vacuum with a non-zero minimum, giving rise to the massive Higgs field. The interaction between the Higgs and electroweak bosons generates their masses.

Finally the mass terms for the physical $W_\mu^\pm$ and Z_μ bosons are obtained by introducing equations 1.18 and 1.20 into the Lagrangian above. The resulting boson mass terms are:

$$m_{W^\pm} = \frac{1}{2}gv \quad \text{and} \quad m_Z = \frac{1}{2}v\sqrt{g^2 + g'^2} , \tag{1.27}$$

The Yukawa interaction

To describe the interaction between the Higgs field, Φ, and fermions, ψ, an additional contribution to the Lagrangian is introduced, known as the Yukawa interaction. This term preserves the $SU(2)$ gauge symmetry and was originally inspired by models of the nuclear force mediated by pions. The interaction is expressed as

$$\mathcal{L}_{\text{Yukawa}} = -\overline{\psi}_L\Gamma\psi_R\Phi - \overline{\psi}_R\Gamma^\dagger\psi_L\Phi^\dagger, \tag{1.28}$$

where Γ denotes the Yukawa coupling matrix, ensuring invariance under Lorentz and gauge transformations [24]. Following spontaneous symmetry breaking, this interaction provides a mechanism by which fermions acquire masses proportional to the vacuum expectation value of the Higgs field, v.

1.1.3 The strong interaction

Quantum Chromodynamics (QCD) is a non-Abelian $SU(3)$ gauge theory which describes the strong interactions, where the conserved Noether current associated with it is the colour charge.

This symmetry group describes transformations among the three colour charges carried by quarks, which transform as triplets under $SU(3)$, corresponding to three distinct colour states. The local $SU(3)$ invariance introduces eight massless gauge fields, represented by the eight Gell-Mann matrices $G_\mu^a, a \in [1, 8]$, leading to gluons being the gauge bosons of the strong interaction, each associated with a generator of the group. These gauge fields are identified as gluons, the mediators of the strong interaction [8].

The QCD Lagrangian for the strong interactions for the 4-component Dirac spinors, ψ_q^i associated with each quark field of colour i and flavour q up to gauge-fixing terms is:

$$\mathcal{L}_{\text{QCD}} = -\frac{1}{4} F_{\mu\nu}^{(a)} F^{(a)\mu\nu} + i \sum_q \bar{\psi}_q^i \gamma^\mu (D_\mu)_{ij} \psi_q^j - \sum_q m_q \bar{\psi}_q^i \psi_{qi} \tag{1.29}$$

$$F_{\mu\nu}^{(a)} = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g_s f_{abc} A_\mu^b A_\nu^c, \tag{1.30}$$

$$(D_\mu)_{ij} = \delta_{ij}\, \partial_\mu - i g_s \sum_a \frac{\lambda_{ij}^a}{2} A_\mu^a, \tag{1.31}$$

where g_s is the QCD coupling constant, A_μ^a are the Yang-Mills (gluon) fields, f_{abc} are the structure constants of the $SU(3)$ algebra and the $\lambda_{i,j}^a/2$ matrices its generators [13].

The covariant derivative is introduced to ensure local $SU(3)$ gauge invariance of the Lagrangian. Due to the non-Abelian nature of $SU(3)$, its generators do not commute, allowing gauge field self-interactions. Equation 1.29 thus represents an incomplete QCD Lagrangian, as a gauge-fixing term is needed to properly define the transformation properties of the gauge fields and fix redundant degrees of freedom.

1.2 Proton collisions at high energies

Hadronic structure is fundamental to hadronic collisions, with real electroweak corrections altering it at high energies. The discussion below focuses on the QCD aspects relevant to hadronic structure and in particular proton–proton, collisions.

Hadrons are bound states of quarks and gluons described by QCD, as outlined in Section 1.1.3. $SU(3)$ gauge invariance restricts asymptotic states to colour-neutral combinations of partons, leading to mesons, formed by a valence quark and a valence antiquark, baryons, which are three-valence-quark singlets, and more exotic configurations such as tetraquarks, pentaquarks, hybrids, and glueballs. In particular, protons are composite particles made of three valence quarks: two up and one down, giving a net charge of +1. Beyond these, they contain a dynamic sea of gluons and quark–antiquark pairs arising from quantum fluctuations. Valence quarks define the quantum numbers of the hadron, including charge and isospin. The mass of a hadron primarily arises from strong interactions, rather than from the Higgs mechanism. Notably, the observed hadron spectrum reflects confinement dynamics and approximate chiral symmetry, and is successfully reproduced by effective models and lattice QCD in the non-perturbative regime.

Confinement in QCD prevents quarks and gluons from existing as free particles, binding them into colour-singlet hadrons. This phenomenon is rooted in the non-Abelian nature of the strong interaction, where the strong coupling constant α_s

increases as the momentum scale μ decreases. At leading order, the running of α_s is given by

$$\alpha_s(\mu) = \frac{1}{\beta_0 \ln(\mu^2/\Lambda^2)} \tag{1.32}$$

As μ approaches the QCD scale Λ, α_s diverges, indicating the breakdown of perturbation theory and the onset of confinement. Conversely, at high energies, where μ is large, α_s becomes small, a property known as asymptotic freedom, which allows quarks and gluons to interact weakly and enables perturbative calculations. At hadronic scales ($\mu \sim 1\,\text{GeV}$), the coupling becomes large, rendering perturbation theory inapplicable; thus, hadron structure must be determined experimentally. In the low-energy regime, the potential between colour charges grows linearly with separation, ensuring that only colour-neutral states are physically observable.

These experimental observations motivated the development of several models, most notably the parton model, introduced by Feynman in 1969 [25], describes hadrons as composed of massless, point-like partons, later identified as quarks and gluons. The internal structure of hadrons is characterized by parton distribution functions (PDFs), which are discussed in Section 1.2.2, while Section 1.2.3 introduces transverse momentum dependent PDFs.

Proton collisions

Proton–proton colliders, such as the LHC, provide a unique environment to test Quantum Chromodynamics (QCD). In these facilities, two proton beams are accelerated in opposite directions to TeV scale energies and brought into collision at dedicated interaction points, as described in Section 4.1. At such energies, the fundamental interactions occur between the constituent partons of the colliding protons. These partons carry different momentum fractions and colour charges, making pp collisions an ideal setting to study a broad spectrum of partonic processes, and thereby a powerful tool to probe both SM and BSM physics.

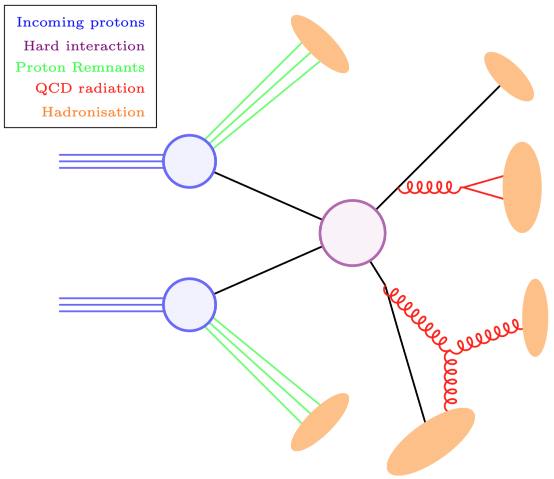


Figure 1.4: Simplified diagram of a pp collision from Ref. [26].

Figure 1.4 illustrates a simplified proton–proton collision, with colours representing distinct stages. The collision is factorized into successive phases: the initial state with two incoming protons (blue), the hard process where constituent partons interact (violet), and the subsequent QCD evolution producing hadrons (orange) and

QCD radiation (red) from initial and final states. Proton remnants (green) also hadronise. Some unstable hadrons decay further until a final stable state of leptons and hadrons is reached, which are the experimentally measurable particles. The quarks and gluons remain connected by conserved colour charge throughout.

Cross sections quantify the likelihood of a given process in these collisions. Different energy scales are involved, from the proton mass ($\sim$ 1 GeV) through the hard interaction scale (up to TeV at the LHC) down to non-perturbative hadronisation scales. Factorization theorems allow separating the cross section into components addressing these scales, which will be introduced in Section 1.2.1 along with the collinear factorization theorem.

1.2.1 QCD factorization

Factorisation theorems allow the separation of scattering processes in proton–proton collisions into distinct components associated with different energy scales. In the context of this thesis, which focuses on the electroweak production of the Z boson with 2 jets (EW Zjj), with particular emphasis on Vector Boson Fusion (VBF) processes, the factorisation approach is outlined as employed to compute this cross section.

It is important to note that the EW Zjj consists on a set of processes that produces a Z^0 boson in association with two jets solely via electroweak interactions at leading order ($\mathcal{O}(\alpha^4_{EW})$ when considering the leptonic decay). This is defined as the t-channel exchange of a weak vector boson, which includes Z-boson production via vector boson fusion (VBF), Z-boson bremsstrahlung and multi-peripheral production, forming a gauge-invariant set. In the VBF Z process (see Fig. 3.1), two quarks from the incoming protons each emit a virtual electroweak boson ($W^{\pm}$ or Z) and continue as scattered quarks, forming forward tagging jets. In a typical channel, an up-type quark emits a W^+ boson and becomes a down-type quark, while a down-type quark emits a W^- boson and becomes an up-type quark. The two vector bosons then fuse to produce an on-shell Z boson, which can decay leptonically, for instance $Z \to \mu^+\mu^-$ or $Z \to e^+e^-$. This process is purely electroweak and is characterized by a large rapidity separation between the tagging jets.

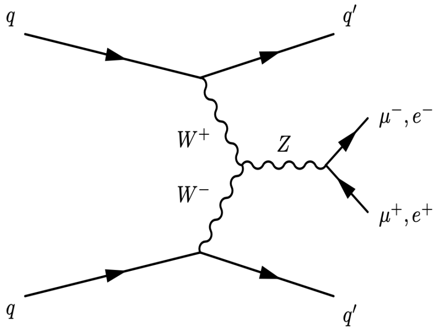


Figure 1.5: VBF Z Feynman diagram.

Collinear factorisation theorem

At QFT level, the collinear factorisation theorem for the EW Zjj process at hadron colliders like the LHC is rooted in the separation of long-distance (non-perturbative) and short-distance (perturbative) physics. This is formalised by expressing the hadronic cross section as a convolution of parton distribution functions (PDFs) and a hard-scattering matrix element that describes the electroweak interaction of the exchanging vector bosons. The PDFs describe the probability to find the incoming quark (or anti-quark) with a fraction of momentum of the proton, x, inside the proton, whereas the hard-scattering matrix element is computed in perturbation theory, therefore the EW Zjj cross section can be expressed as:

$$\sigma^{\text{EW Zjj}}_{pp\to ZJJ}(s) = \sum_{i,j}\int_0^1 dx_1 \int_0^1 dx_2\, f_i(x_1,\mu_F)\, f_j(x_2,\mu_F)\, \hat{\sigma}^{\text{EW Zjj}}_{ij\to ZJJ}(\hat{s},\mu_F,\mu_R) \tag{1.33}$$

where s is the squared center-of-mass energy of the entire proton-proton collision, $f_i(x,\mu_F)$ is the PDF for parton i with momentum fraction x at factorisation scale μ_F, $\hat{\sigma}^{\text{EW Zjj}}_{ij\to ZJJ}$ is the partonic cross section for the hard EW Zjj process whereas μ_F and μ_R are the factorisation and renormalisation scales, respectively. The partonic cross section is defined as

$$\hat{\sigma}^{\text{EW Zjj}}_{ij\to ZJJ}(\hat{s}) = \frac{1}{2\hat{s}}\int d\hat{\Phi}_{ZJJ}\left|\mathcal{M}_{ij\to ZJJ}\right|^2 \tag{1.34}$$

In this expression, $\hat{s} = x_1 x_2 s$ is the partonic centre of mass energy squared, also known as one of the Mandelstam variables, $\hat{\Phi}_{ZJJ}$ is the differential phase space for the $Z + 2$ jets final state and $\left|\mathcal{M}_{ij\to ZJJ}\right|^2$ is the squared matrix element for the process, summed/averaged over initial and final spins and colours. $\left|\mathcal{M}_{ij\to ZJJ}\right|^2$ defines the transition probability of the incoming particles into the outgoing particles, which in the EW Zjj scenario is the probability amplitude for the partonic process where two quarks (or antiquarks) exchange electroweak bosons in the t-channel and produce a Z boson plus two jets.

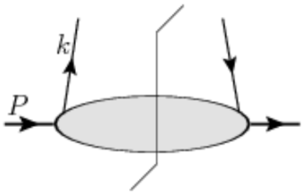
The matrix element $\mathcal{M}_{ij\to ZJJ}$ is computed perturbatively, leading to ultraviolet (UV) and infrared (IR) divergences. UV divergences from large momenta are absorbed by renormalisation, introducing the arbitrary scale μ_R, and IR divergences from small momenta are factorised into the PDFs, with μ_F as the cut-off. Both scales are unphysical, and the cross section should be independent of them, but truncating the perturbative expansion at finite order induces residual dependence, reduced at higher orders. Varying μ_F and μ_R estimates missing higher-order effects. Collinear factorisation is effective for single-scale processes but fails when disparate energy scales exist, such as in low p_T regions of the dilepton system (reconstructed Z^0 boson), where resummation becomes necessary (see Section 1.2.4).

1.2.2 Parton Distribution Functions

Parton distribution functions (PDFs) study the collinear structure of hadrons. PDFs encode the probability density for finding a parton inside a hadron carrying a fraction x of the hadron's longitudinal momentum, at a given energy scale μ.

The PDF is formally defined through deep inelastic scattering (DIS), as realized at HERA, where high-energy electrons or positrons scatter off protons [27]. In this process, the virtual photon exchanged between the lepton and the proton probes the proton's internal structure. Following the derivation by John Collins in [28], the cross section factorizes into a calculable electromagnetic current and a hadronic current, the latter encapsulating the non-perturbative structure of the proton:

$$f_{j/h}(\xi) = \mathrm{Tr}\,\frac{\gamma^+}{2}\int\frac{dk^-\,d^2\mathbf{k}_T}{(2\pi)^4} \tag{1.35}$$



This definition is exactly correct, appropriate, and valid, but only in a model QFT where the parton model is correct. In QCD it must be modified as follows: the bare (unrenormalized) parton distribution function (PDF) for a parton of flavour j in the proton H is defined as a gauge-invariant light-cone correlator of quark fields within the proton [28]:

$$f_{(0)\,j/h}(\xi) = \int \frac{dw^-}{2\pi}\, e^{-i\xi P^+ w^-}\,\langle P|\bar{\psi}_j^{(0)}(0,w^-,\mathbf{0}_T)W(w^-,0)\frac{\gamma^+}{2}\psi_j^{(0)}(0)|P\rangle_c\,, \tag{1.36}$$

where the Wilson line is

$$W(w^-,0) = \mathcal{P}\left[e^{-ig_0\int_0^{w^-}dy^-\,A^+_{(0)\alpha}(0,y^-,\mathbf{0}_T)\,t_\alpha}\right], \tag{1.37}$$

a path-ordered exponential of the gluon field along the line joining the quark and antiquark fields ensuring gauge invariance.
In Eqs. 1.36 and 1.37, $\psi^{(0)}$ denotes the bare quark field, $|P\rangle$ is an eigenstate, g_0 is the bare gauge coupling of QCD, ξ is the longitudinal momentum fraction, γ^+ is a light cone Dirac gamma matrix, P^+ is the “plus” component of the hadron’s momentum in light-cone coordinates and t_α are the generating matrices of the SU(3).
However, in QCD, ultraviolet (UV) divergences arise due to short-distance fluctuations (see Fig. 1.6). To obtain a physically meaningful, finite PDF, the bare operator must be renormalized. This is achieved by introducing a UV regulator, for instance through dimensional regularization with $n = 4 - 2\epsilon$ space-time dimensions, and by applying suitable renormalization factors to absorb the divergences. In the minimal subtraction (MS) renormalization scheme, these factors are defined by subtracting only the divergent parts of the loop corrections, expressed as poles in ϵ that arise in dimensional regularization. Specifically, order-by-order in perturbation theory, the renormalization factor $Z_{jj'}(z,g,\epsilon)$ is written as pure pole terms in ϵ at $\epsilon = 0$:

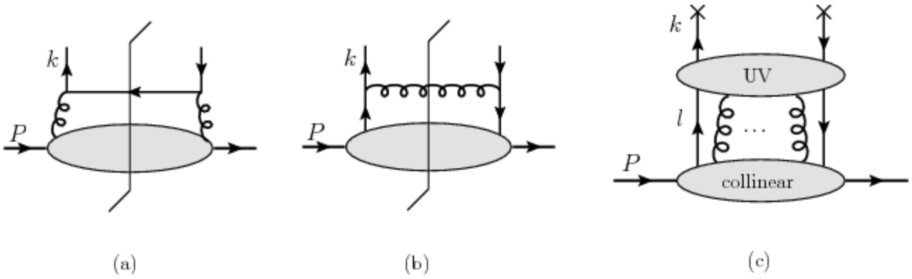


Figure 1.6: (a) and (b) Examples of graphs with UV divergences for quark density. (c) Structure of UV divergences of parton densities from Ref. [28].

$$Z_{jj'}(z,g,\epsilon) = \sum_{n=0}^{\infty}\left(\frac{g^2}{16\pi^2}\right)^n Z^{[n]}_{jj'}(z,\epsilon), \quad Z^{[n]}_{jj'}(z,\epsilon) = \sum_{l=1}^{n} Z_{n,l;jj'}(z)S_\epsilon\frac{1}{\epsilon^l}, \tag{1.38}$$

where S_ϵ is selected as:

$$S_\epsilon = \frac{(4\pi)^\epsilon}{\Gamma(1-\epsilon)} \; . \tag{1.39}$$

This renormalization procedure introduces a dependence on an arbitrary energy scale, μ, in the PDF: $f_{(0)\,j/h}(\xi) \to f_{j/h}(\xi) \equiv f_{j/h}(x,\mu)$ by performing an integration over z, the momentum fraction “transfer” between the bare and renormalized parton:

$$f_{j/h}(\xi) = \sum_{j'} \int_{\xi^-}^{1^+} \frac{dz}{z}\, Z_{jj'}(z,g,\epsilon)\, f_{(0)j/h}(\xi/z). \tag{1.40}$$

As μ is an arbitrary scale $f_{j/h}(x,\mu)$ should not depend on the choice of μ, therefore renormalisation group equations (RGEs) are derived to explain the dependence of the PDF on the energy scale:

$$\frac{\partial f_{a,h}(x,\mu)}{\partial \ln \mu} = \sum_b \int_x^1 \frac{dz}{z} P_{ab}\left(z, \alpha_s(\mu)\right) f_{b,h}\left(\frac{x}{z}, \mu\right) \; . \tag{1.41}$$

From the viewpoint of collider phenomenology, PDFs are indispensable, experimentally constrained inputs that connect the theoretical framework of QCD with measurable observables at hadron colliders such as the LHC. By encoding the probability densities of partons carrying certain momentum fractions inside the proton, PDFs enable quantitative predictions of cross sections and kinematic distributions across a wide range of processes. Measurements of differential cross sections in proton–proton collisions, performed at multiple center-of-mass energies and over broad kinematic domains, feed into global PDF analyses, thereby refining their precision and constraining the proton’s quark and gluon structure. The scale dependence governed by the DGLAP evolution equations manifests experimentally through scale- and energy-dependent modifications of these distributions, observed as scaling violations in measured cross sections. This interplay allows for increasingly precise tests of the Standard Model and enhances the sensitivity of searches for phenomena beyond it.
The RGEs for the PDFs are known as the DGLAP evolution equations and P_{ba}, the evolution kernels, are known as the splitting functions. The evolution kernels can be computed perturbatively in α_s by replacing the hadron with an on-shell parton, since they are independent of the hadron state. In this way, the kernels follow directly from

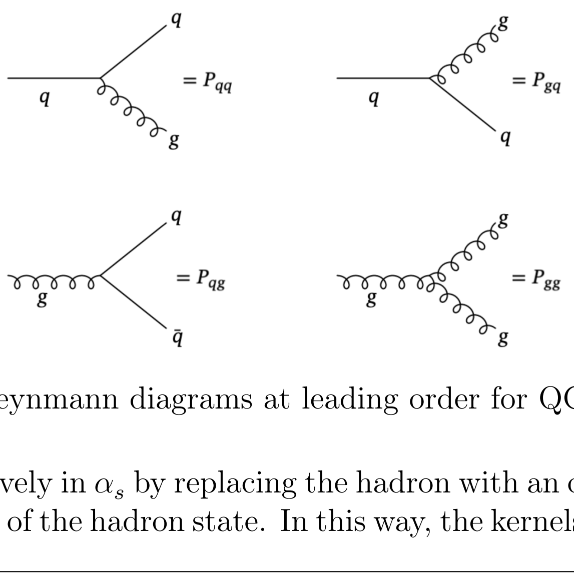


Figure 1.7: Feynmann diagrams at leading order for QCD splittings.

QCD vertices at scales μ in the perturbative regime. The leading-order diagrams are shown in Figure 1.7, and the corresponding expressions in Equation 1.42.

$$
\begin{aligned}
P_{qq}(z) &= C_F\left[\frac{1+z^2}{(1-z)_+}+\frac{3}{2}\delta(1-z)\right]\\
P_{gq}(z) &= C_F\frac{1+(1-z)^2}{z}\\
P_{qg}(z) &= T_R\left[z^2+(1-z)^2\right]\\
P_{gg}(z) &= 2C_A\left[\frac{z}{(1-z)_+}+\frac{1-z}{z}+z(1-z)\right]+\frac{11C_A-4n_fT_R}{6}\delta(1-z)
\end{aligned}
\tag{1.42}
$$

Here, C_A, C_F, and T_R denote the QCD color factors, and n_f is the number of active quark flavours.

The splitting functions exhibit divergences at specific values of z: P_{gg} and P_{qq} diverge as $z \to 1$, corresponding to gluon radiation from gluons and quarks, respectively. Near $z \sim 1$, large momentum transfer leads to emitted gluons carrying a small momentum fraction $(1-z)$, known as soft gluon radiation. P_{gg} and P_{gq} diverge as $z \to 0$. As a result, DGLAP evolution can be seen as parton splitting, where quarks lose momentum by emitting gluons via bremsstrahlung, and gluons split into $q\bar{q}$ pairs, driving PDF evolution.

While the scale dependence of PDFs can be calculated within perturbative QCD, the intrinsic structure of the hadron itself cannot be derived from first principles due to the non-perturbative nature of confinement. Instead, the partonic content of the hadron must be modelled through parametrisations. Specifically, the distribution of a given parton flavour q is represented by a general functional form of the type [29]:

$$f_q(x, Q_0^2) = Ax^{\alpha}(1-x)^{\beta}P(x), \tag{1.43}$$

where x is the fraction of momenta carried by the parton, Q_0^2 is the initial energy scale, A, α, and β are parameters fitted to experimental data, and $P(x)$ is a polynomial or other function to capture shape nuances. The starting distribution at the scale Q_0^2 are then evolved to any other scale Q^2 using the DGLAP evolution equations, which resum the leading logarithmic scale dependence governed by perturbative QCD.

Collinear PDFs characterize the hadron structure in one dimension by integrating over transverse momenta. In contrast, the three-dimensional structure of the hadron can be probed by not performing the transverse momentum integration in Eq. 1.35. These three-dimensional distributions are known as unintegrated PDFs or transverse momentum dependent PDFs (TMD PDFs).

Parton Branching-based (PB) PDFs

PDFs can also be determined by solving the DGLAP evolution equations through iterative parton branchings and renormalization group evolution in a fully exclusive way. This approach is known as the Parton Branching (PB) method. Unlike traditional inclusive PDF extraction techniques, the PB method retains the full kinematic information of each branching, making it naturally compatible with Monte

Carlo event generators (e.g. CASCADE3 [30]). This feature makes the method particularly advantageous for studying multi-scale and small-x processes.

Following the calculation from [31], the renormalization group evolution of parton distribution functions, $f_a(x,\mu^2)$ can be written in terms of DGLAP parton splitting functions, $P_{ab}(\alpha_s(\mu^2),z)$ as stated in Eq.1.41. This PDF can be expressed in its momentum-weighted form as:

$$\tilde{f}_a(x,\mu^2) \equiv x f_a(x,\mu^2) \;, \tag{1.44}$$

therefore:

$$\frac{\partial \tilde{f}_a(x,\mu^2)}{\partial \ln \mu^2} = \sum_b \int_x^1 \frac{dz}{z}\, P_{ab}(\alpha_s(\mu^2),z)\, \tilde{f}_b(x/z,\mu^2) \;. \tag{1.45}$$

Here x denotes the longitudinal momentum fraction, μ represents the evolution mass scale, α_S is the running coupling, and z stands for the splitting variable.

Because the transverse separation between emitted partons is limited by a finite resolution scale, energy–momentum conservation prevents the resolution of partons whose longitudinal momentum fractions are too close to $z = 1$. This limitation introduces a cutoff, $z > z_M$, where $1 - z_M \sim \mathcal{O}(\Lambda_{\text{QCD}}/\mu)$. Here, μ is a scale of the order of the hard-scattering energy, and $\Lambda_{\text{QCD}} \approx 1\,\text{fm}^{-1}$ represents the characteristic scale of the strong interaction. This is solved by introducing a soft-gluon resolution parameter z_M in Eq.1.45. This parameter separates the full evolution into two contributions: the resolvable range ($z < z_M$) and the non-resolvable range ($z > z_M$), where $1{=}z_M \sim O(\Lambda_{QCD}/\mu)$.

This allows the PDF to be reformulated in terms of the resolvable contributions as follows:

$$\begin{aligned}\frac{\partial \tilde{f}_a(x,\mu^2)}{\partial \ln \mu^2} = \sum_b \Bigg\{ &\int_x^{z_M} dz\, P^{(R)}_{ab}\big(\alpha_s(\mu^2),z\big) \tilde{f}_b\left(\frac{x}{z},\mu^2\right) \\ &- \int_0^{z_M} dz\, z\, P^{(R)}_{ba}\big(\alpha_s(\mu^2),z\big) \tilde{f}_a(x,\mu^2) \Bigg\} \;,\end{aligned} \tag{1.46}$$

which can be solved by an iterative Monte Carlo method if the no-branching probabilities are included. This is achieved through the implementation of the Sudakov form factor:

$$\Delta_a(z_M,\mu^2,\mu_0^2) = \exp\left(- \sum_b \int_{\mu_0^2}^{\mu^2} \frac{d\mu'^2}{\mu'^2} \int_0^{z_M} dz\, z\, P^{(R)}_{ba}\big(\alpha_s(\mu'^2),z\big) \right) \;, \tag{1.47}$$

which is defined as the probability for parton a to undergo no branching between evolution scale μ_0 and evolution scale μ, where the branchings are understood to be classified according to the given resolution z_M. As a result, including this factor leads to the PDF evolution equation rewritten as:

$$\frac{\partial \tilde{f}_a(x,\mu^2)}{\partial \ln \mu^2} = \sum_b \int_x^{z_M} dz P^{(R)}_{ab}(\alpha_s(\mu^2),z) \tilde{f}_b(x/z,\mu^2) + \frac{1}{\Delta_a(\mu^2)} \frac{\partial \Delta_a(\mu^2)}{\partial \ln \mu^2} \tilde{f}_a(x,\mu^2) \;. \tag{1.48}$$

Finally, considering $\Delta_a(\mu_0^2) = 1$, this expression can be integrated as:

$$\tilde{f}_a(x,\mu^2) = \Delta_a(\mu^2)\tilde{f}_a(x,\mu_0^2) + \sum_b \int_{\mu_0^2}^{\mu^2} \frac{d\mu'^2}{\mu'^2} \frac{\Delta_a(\mu^2)}{\Delta_a(\mu'^2)} \int_x^{z_M} dz P_{ab}^{(R)}(\alpha_s(\mu'^2), z)\tilde{f}_b(x/z, \mu'^2) \; , \tag{1.49}$$

which can be iteratively solved via a Monte Carlo method.

1.2.3 Transverse Momentum Dependent Parton Distributions

Transverse momentum dependent PDFs (TMDs) generalize collinear PDFs by encoding the full three-dimensional momentum structure of hadrons as they retain information on both the longitudinal and transverse momentum components of partons within the nucleon. This captures spin–orbit correlations and orbital angular momentum effects inaccessible in the collinear framework. Formally, TMDs arise by not integrating over the transverse momentum in Eq.1.35, requiring a joint parametrisation of longitudinal and transverse momentum dependence.

In both Parton Branching (PB) TMDs and TMDs based on the TMD factorisation framework, the hadron structure is decomposed into collinear and transverse components. While both methods are grounded in QCD factorisation, TMDs derived from TMD factorisation are defined within a formal factorisation theorem, where the transverse momentum dependence is parametrised at a low scale and evolved using dedicated evolution equations. In contrast, Parton Branching TMDs are constructed by solving QCD evolution equations with explicit generation of transverse momenta at each branching, embedding the transverse structure directly into the parton emission history.

TMD factorisation framework

The Transverse Momentum Dependent (TMD) factorization framework provides a systematic theoretical approach for describing hadronic processes in the region $q_T \ll Q$, where intrinsic transverse momentum effects become significant. This framework resumes logarithms of the form $\ln(Q^2/q_T^2)$ by introducing TMD parton distribution functions, and is essential for precision predictions of electroweak gauge boson production, such as $\mathrm{Z} + \mathrm{jets}$ at the LHC [32, 33].

TMD Factorization Formula

Within the Collins–Soper–Sterman (CSS) formalism [34], the TMD PDF in impact-parameter space b_T reads

$$\begin{aligned} f_{j/H}(x, b_T; \zeta, \mu) = \exp\Big[&- g_{j/H}(x, b_T; b_{\max}) - g_K(b_T; b_{\max}) \ln \frac{\sqrt{\zeta}}{Q_0} \Big] \\ &\times \exp\Bigg\{ K(b_*; \mu_{b_*}) \ln \frac{\sqrt{\zeta}}{\mu_{b_*}} + \int_{\mu_{b_*}}^{\mu} \frac{d\mu'}{\mu'} \Big[\gamma_j(\alpha_s(\mu')) - \tfrac{1}{2} \ln \frac{\zeta}{\mu'^2} \, \gamma_K(\alpha_s(\mu')) \Big] \Bigg\} \\ &\times \sum_{j'} \int_x^1 \frac{d\xi}{\xi} \, f_{j'/H}(\xi; \mu_{b_*}) \, C_{j/j'}^{\mathrm{PDF}}\Big(\tfrac{x}{\xi}, b_*, \mu_{b_*}, \alpha_s(\mu_{b_*}) \Big) \, . \end{aligned}$$

Here:

- x is the longitudinal momentum fraction of hadron H carried by parton j.

- b_T is the impact-parameter (Fourier-conjugate to transverse momentum k_T).
- ζ is the rapidity factorization scale, regulating rapidity divergences.
- μ is the renormalization (UV) scale.
- $K(b_T;\mu)$ is the Collins–Soper kernel, and γ_j, γ_K are the anomalous dimensions.
- g_K and $g_{j/H}$ are non-perturbative functions encoding long-distance physics.
- The b_* prescription is defined by $b_* = b_T/\sqrt{1+(b_T/b_{\max})^2}$ and $\mu_{b_*} = 2e^{-\gamma_E}/b_*$, where γ_E is the Euler–Mascheroni constant.

Evolution Equations
The rapidity evolution is given by the Collins–Soper equation [35]

$$\frac{\partial \ln f_{j/H}(x, b_T; \zeta, \mu)}{\partial \ln \sqrt{\zeta}} = K(b_T;\mu)\,, \tag{1.50}$$

with the perturbative expansion

$$K(b_T;\mu) = -8C_F\,\alpha_s(\mu)\,\ln\!\left(\frac{b_T\,\mu}{2e^{-\gamma_E}}\right) + \mathcal{O}(\alpha_s^2)\,. \tag{1.51}$$

The renormalization-group equation in μ reads

$$\frac{d\ln f_{j/H}(x, b_T; \zeta, \mu)}{d\ln \mu} = \gamma_j(\alpha_s(\mu)) - \tfrac{1}{2}\,\gamma_K(\alpha_s(\mu))\,\ln\frac{\zeta}{\mu^2}\,, \tag{1.52}$$

where γ_j and γ_K are extracted from quark electromagnetic form-factor calculations [36].

Non-perturbative Structure
For $b_T > b_*$, the Collins–Soper kernel and intrinsic-momentum behaviour become non-perturbative. The b_* prescription ensures perturbative control:

$$b_* = \frac{b_T}{\sqrt{1+(b_T/b_{\max})^2}}\,, \qquad \mu_{b_*} = \frac{2e^{-\gamma_E}}{b_*}\,. \tag{1.53}$$

The non-perturbative functions $g_K(b_T; b_{\max})$ and $g_{j/H}(x, b_T; b_{\max})$ are determined via global fits to Drell–Yan and electroweak boson data, exploiting their universality across processes [37].

This formalism underpins the interpretation of Zßboson transverse-momentum spectra, separating short-distance dynamics from universal long-distance proton structure. In brief in the TMD factorization framework there are two evolution equations: the CSS kernel, which governs the rapidity evolution, and the RG equation, which describes the evolution in μ. To determine the TMD, g_K, which is the non-perturbative function of the CSS kernel, and $g_{j/H}$, which encodes the intrinsic transverse momentum, must be fitted to experimental data.

Despite its strong theoretical basis, the TMD factorization in VBF Z production faces limitations. It is valid primarily at low $q_T \ll Q$, restricting predictive power to this region. Large b_T introduces non-perturbative functions, whose universality is uncertain, affecting precision. The framework's applicability to complex, multi-scale processes like VBF, with multiple jets and color exchanges, is less proven. Additionally, higher-order truncations, matching uncertainties, and power corrections further constrain its predictive accuracy. Consequently, careful interpretation and complementary methods are necessary for reliable phenomenology in this context [38].

Parton Branching-based (PB) TMDs
PB-based TMDs at their energy starting scale μ_0, are defined as:

$$\tilde{A}_a(x, k_{\perp,0}^2, \mu_0^2) = x f_a(x, \mu_0^2) \cdot \frac{1}{q_s^2} \exp\left(-\frac{k_{\perp,0}^2}{q_s^2}\right) , \tag{1.54}$$

where $f_a(x, \mu_0^2)$ denotes the collinear starting parton density, k_T is the transverse momentum assumed to follow a Gaussian distribution, and q_s^2 represents its width. Therefore, to determine the TMD distribution at a general energy scale μ, the parameter q_s^2 is extracted from fits to experimental Drell–Yan p_T spectra measured at various centre-of-mass energies and dilepton invariant masses, as detailed in [39]. This procedure yields a value of $q_s^2 = 1\,\text{GeV}^2$.
Then, the evolution of the PB TMDs is obtained from the extended DGLAP evolution equations [31]:

$$\begin{aligned} \tilde{A}(x, k_T, \mu^2) = \Delta_a(\mu^2)\, \tilde{A}(x, k_T, \mu_0^2) + \sum_b \int_{\mu_0^2}^{\mu^2} \frac{d^2\mu'^2}{\pi\mu'^2} \frac{\Delta_a(\mu^2)}{\Delta(\mu'^2)} \times \\ \int_x^{z_M} dz\, P_{ab}^{(R)}\big(\alpha_S(\mu'), z\big)\, \tilde{A}\Big(\frac{x}{z}, \vec{k}_T - \vec{q}_T, \mu'^2\Big) , \end{aligned} \tag{1.55}$$

which is consistent to Eq.1.49 as expected. Here Δ_a is the Sudakov form factor, as explained in Section 1.2.2; $\tilde{A}$ denotes the momentum-weighted TMD distribution; z_M represents the maximal momentum transfer from parton b to parton a; $P_{ab}^{(R)}$ is the real part of the splitting function; μ' is the scale of the radiation, and $\vec{q}_T$ corresponds to the transverse momentum of the radiated parton.

1.2.4 Resummation

The Z^0 boson is produced mainly through the Drell-Yan mechanism in proton-proton colliders: $q\bar{q} \to Z^0 \to l^+l^-$, which is a QCD-driven mechanism initiated by quark–antiquark annihilation with gluon radiation: $q\bar{q} \to Z^0 \to l^+l^-\mathbf{g}$. In this topology, the cross section with a gluon radiation according to [40] is:

$$\frac{d\sigma}{dq_T^2} \propto (d\sigma)_{Born} \times \alpha_s \frac{1}{q_T^2} \ln \frac{m_{ll}^2}{q_T^2} , \tag{1.56}$$

where $(d\sigma)_{Born}$ is the Born level cross section for the DY process, α_S is the strong coupling and m_{ll} is the mass of the dilepton system. In this scenario, the lack of higher order corrections introduces a divergent behaviour when $q_T \to 0$ and the logarithmic term and α_S increase when $q_T \ll m_{ll}^2$, introducing the need to model this region due to its non-perturbative nature.
To regularize this divergence, all order emissions are accounted for by implementing the collinear approximation together with a strong ordering condition. In the first-order approximation, the radiation angle is assumed to be small, while higher-order

terms allow the emissions to be expressed as a finite series involving n radiations. Consequently, these geometric series

$$\prod_{i=1}^{n} \int \frac{dp_{T,i}^{2}}{p_{T,i}^{2}} \ln \frac{m_{ll}^{2}}{p_{T,i}^{2}} \delta(\vec{p}_{T,n} - \vec{q}_{T}) = \exp\left(\frac{1}{2} \ln^{2} \frac{m_{ll}^{2}}{p_{t}^{2}}\right) , \tag{1.57}$$

yield a finite cross section by resumming gluon emissions to all orders in the Drell–Yan Z^0 boson production process:

$$\frac{d\sigma_{Res}}{dq_T} = \frac{d\sigma}{dq_T} \times e^{Series} , \tag{1.58}$$

where σ_{Res} is the resummed and finite cross-section. As a result, the transverse momentum spectrum of the Z^0 boson can be divided into three characteristic regimes: the non-perturbative region ($q_T \ll m_{ll}$), the resummation region ($q_T < m_{ll}$), and the hard region ($q_T > m_{ll}$). To ensure a consistent description, the resummation region must be properly matched to the hard region, thereby preventing double counting of emitted partons. The resummation of soft and collinear radiation can be performed through several approaches, traditionally using transverse momentum factorisation theorems, TMDs, or parton shower algorithms.

On the opposite, in the EW Zjj production, and particularly in the VBF Z process, does not require QCD resummation. The VBF Z process is purely electroweak at leading order and proceeds via the t-channel exchange of a weak boson between two quark lines, with no colour connection between the scattered quarks. As a consequence, the pattern of soft and collinear QCD radiation is strongly suppressed, and the large logarithms that spoil the convergence of the fixed-order expansion in QCD-induced processes, such as $\log(M_Z^2/p_T^2)$ terms at low p_T, do not arise. The fixed-order electroweak calculation therefore remains reliable across the relevant phase space.

Further advancements in the formal treatment of QCD corrections to vector boson fusion processes are discussed in detail in [41–44]. Comprehensive studies addressing the incorporation of both QCD and QED corrections in hadronic electroweak boson production can be found in [45, 46].

Parton Showers

Parton showers (PS) resum multiple quark and gluon emissions through an iterative evolution of incoming partons on an event-by-event basis. PS keep the final-state kinematics fixed to preserve momentum conservation, respect physical symmetries, and ensure compatibility with both higher-order calculations and hadronisation models. Consequently, the evolution proceeds backwards, incorporating Sudakov factors to ensure probability conservation and theoretical consistency in initial-state shower algorithms within perturbative QCD [47].

Considering the standard approach, parton showers are derived from the DGLAP evolution. In this context, the Sudakov form factor can be expressed as described in Eq. 1.47. There, a parton evolves from an initial energy scale μ_0 to a higher scale μ by generating the momentum fraction z of the emitted particle according to the splitting function $P_{ab}(z)$. In this context, if the computed μ' lies outside the interval between μ_0 and μ, the emission does not occur, and no branching takes place in the evolution, as already stated in Section 1.2.2. However, in a backward evolution, the previous procedure is inverted, meaning the evolution is performed from the final energy scale μ to the initial one μ_0, this is translated into an additional constrain in Eq. 1.47:

$$\Delta_a^{\text{Backward}} = \exp\left(-\sum_b \int_{\mu_{i-1}^2}^{\mu^2} \frac{d\mu'^2}{\mu'^2} \int_x^{z_{dyn}} dz P_{ab}^{(R)}(\alpha_s(\mu'^2), z) \frac{\tilde{f}_b(x', \mu'^2)}{\tilde{f}_a(x, \mu'^2)} \right), \qquad (1.59)$$

This procedure guarantees that the parent parton participating in the splitting process was originally part of the hadron. The total transverse momentum, obtained by combining the contributions from the emitted radiations and the intrinsic transverse momentum of the incoming partons, is subsequently transferred to the final-state particles. The non-perturbative intrinsic transverse momentum is then modelled using a Gaussian distribution.

Modern Monte Carlo event generators model QCD radiation through parton-shower algorithms that iteratively approximate the all-orders structure of soft and collinear divergences via successive 1 → 2 splittings. Variants of these algorithms differ mainly in the choice of evolution variable and the implementation of colour coherence. Virtuality-ordered showers, used in early versions of PYTHIA [48], evolve in decreasing off-shellness Q^2 and require angular vetoes to restore soft-gluon coherence. Angular-ordered showers, as in HERWIG, employ the emission angle as the evolution parameter and thus implement coherence by construction [49]. Transverse-momentum–ordered showers, adopted in PYTHIA 8 [48], evolve according to the emission hardness p_T, providing an intuitively physical ordering that facilitates consistent merging with multi-leg matrix elements. Dipole and antenna showers, implemented for instance in SHERPA [50] and VINCIA [51], treat radiation as emitted from colour-connected parton pairs, naturally capturing coherence effects and maintaining close correspondence with NLO subtraction formalisms.

For initial-state radiation (ISR), shower evolution proceeds backward from the hard scale to the hadronic scale, with the no-emission probability encoded in an ISR Sudakov factor containing ratios of parton distribution functions. This construction, formalised in the PDF2ISR framework [52], preserves the inclusive PDFs and ensures consistency between ISR evolution and DGLAP dynamics, an essential requirement for precision predictions in processes such as Drell–Yan and electroweak boson production.

1.2.5 Hadronisation

The EW Zjj process involves a dijet system as part of its final state. These jets are the result of a hadronisation process, in which the original high-energy quarks and gluons produced in the hard scattering evolve into collimated sprays of hadrons through non-perturbative QCD dynamics. This hadronisation process transforms colour-charged partons into colour-neutral hadrons, allowing them to be observed as jets in the detector [53].

Hadronisation can be studied through fragmentation functions in processes where the final states originate from QCD interactions, such as the Drell–Yan mechanism. These fragmentation functions quantify the probability density that an outgoing parton will produce a hadron h [54]. In contrast, within the EW Zjj production, particularly in the VBF Z channel, the Z^0 boson is produced via the fusion of two $W^{\pm}$ bosons exchanged between initial-state quarks. This mechanism proceeds through a colour-neutral exchange, unlike typical QCD processes involving colour-charged partons, and therefore leads to a suppressed hadronic activity in the central region. As a result, the modelling of hadronisation in this case becomes less direct, relying instead on non-perturbative techniques implemented in parton-shower Monte Carlo event generators. These generators typically employ two main approaches to simulate the hadronisation process:

- **The Lund string model:** describes the confining colour field between quarks and antiquarks separated by distances exceeding approximately one femtometer as a relativistic string characterized by a constant tension, reflecting the linearity of the confinement potential. When the energy stored in the system surpasses the energy associated with the string tension, the string breaks. This process leads to the creation of a new quark–antiquark pair from the vacuum, where the newly formed quark (or antiquark) subsequently combines with the corresponding antiquark (or quark) from the previous configuration [55].
- **The cluster model:** is founded on the principle of pre-confinement, where gluons, carrying both colour and anti-colour charges, undergo non-perturbative splitting into quark–antiquark pairs. These pairs subsequently form primordial clusters, with each quark ultimately combining with an antiquark to produce colour-neutral states [56, 57].

Additionally, in recent years, Machine Learning (ML) and data-driven approaches have begun to be explored as tools for improving hadronisation modelling. One of the first proofs of concept in this direction is provided by `MLhad` [58], an ML framework whose architecture is built upon conditional sliced-Wasserstein autoencoders (cSWAE). The cSWAE is a type of generative model designed to learn complex conditional distributions by minimizing the sliced-Wasserstein distance between the data and the model-generated samples [59]. In this context, `MLhad` is trained on simplified Pythia-generated kinematic distributions corresponding to pion emissions within the restricted Lund string model, with the training restricted to strings ending in light quark flavours.

Another recent development involves the application of deep generative models to hadronisation through a two-tier Generative Adversarial Network (GAN) framework. A GAN consists of two neural networks, a generator and a discriminator, engaged in an adversarial optimisation process in which the generator aims to reproduce the true data distribution, while the discriminator seeks to distinguish between genuine and synthetic samples [60]. In this specific configuration, the generator operates at the parton level, whereas the discriminator is defined at the hadron level. This approach has demonstrated the ability to reproduce key features of the Herwig cluster hadronisation model in a simplified configuration constrained to pion production [61]. Although ML-based and deep generative hadronidation models have demonstrated promising results in proof-of-principle studies, they have not yet been widely adopted for mainstream phenomenology applications.

Chapter 2

Theory and Phenomenology of Z Boson Production

In proton-proton colliders such as the LHC, the dominant production mode of the Z^0 boson is the Drell-Yan process, where a quark and antiquark from the colliding protons are annihilated at leading order via the s-channel, producing a Z^0 boson that subsequently decays into a lepton-antilepton pair: $q\bar{q} \rightarrow Z^0 \rightarrow l^+l^-$. This topology constitutes an excellent tool to perform PDF fits and study the PDF dependence in x, mainly for low-to-moderate x values.

In contrast, the electroweak production of the Z^0 boson with 2 jets, referred to as *EW Zjj*, involves the set of processes that produces a Z^0 boson in association with two jets solely via electroweak interactions at leading order ($\mathcal{O}(\alpha_{EW}^4)$ when considering the leptonic decay). This is defined as the t-channel exchange of a weak vector boson, which includes Z-boson production via vector boson fusion (VBF), Z-boson bremsstrahlung and multi-peripheral production, forming a gauge-invariant set. In this case, the kinematic selections typically require large dijet invariant mass and significant rapidity separation, isolating regions where initial-state quarks carry high momentum fractions, thereby providing direct sensitivity to the u and d quark PDFs at high x, which are difficult to constrain with more inclusive processes. Differential measurements enable bin-by-bin PDF validation in electroweak-enhanced regions, complementing global PDF fits especially in high-x and forward acceptance domains. Furthermore, due to its similar kinematic characteristics and sensitivity to PDFs as VBF Higgs, this approach offers a critical validation tool for PDFs in the phase space relevant to VBF Higgs and beyond the Standard Model electroweak investigations. Moreover, among the different topologies which contribute to the EW Zjj, the VBF Z constitutes an excellent tool to investigate the $W^{\pm}$ boson densities.

The differential cross section as a function of the dilepton transverse momentum, p_T^{ll} provides critical insight into the three-dimensional momentum structure of the proton through the study of Transverse Momentum Dependent (TMD) parton distribution functions.

2.1 Inclusive production of Z boson

In the current thesis, the inclusive production of the Z^0 boson refers to the inclusion of both the Drell-Yan (see Fig. 2.1) and the electroweak (EW) production of the Z^0 boson (see Fig. 2.2). Section 2.2 presents the formal calculation of the Drell-Yan

Z^0 boson inclusive cross section and differential cross section with respect to the energy scale Q^2 and the dilepton transverse momentum p_T^{ll}. Section 2.3 introduces the topologies which form the electroweak production of the Z^0 boson and displays the formal calculation of the cross sections for the VBF Z. Finally, Section 2.4 shows a comparison of Drell-Yan, EW Zjj and VBF Z at particle level.

2.1.1 Drell-Yan production of Z boson

The cross section of the Drell-Yan (see Fig. 2.1) production of the Z^0 boson is computed following the calculation presented in Refs. [1, 40]. The general partonic cross section for $q\bar{q} \to Z^0 \to l^+l^-$ is

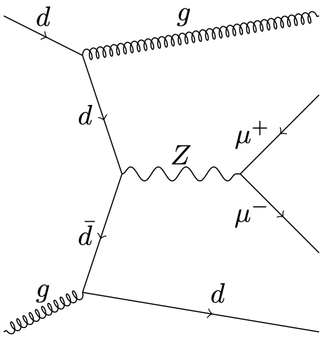


Figure 2.1: Representative Feynman diagrams for Drell-Yan production in association with dijets, from Ref. [62]

$$\hat{\sigma}(\hat{s}) = \frac{1}{2\hat{s}} \int \overline{|\mathcal{M}|^2} d\Phi_2 \ . \tag{2.1}$$

Here $\hat{s}$ is the partonic centre-of-mass energy squared, $\overline{|\mathcal{M}|^2}$ is the squared amplitude averaged over initial spins and colours and $d\Phi_2$ is the 2-body phase space element. The amplitude for $q(p_1)\bar{q}(p_2) \to Z \to \ell^+(k_1)\ell^-(k_2)$ is

$$\mathcal{M} = \bar{v}(p_2)\gamma^\mu(g_V^q - g_A^q\gamma^5)u(p_1)\frac{-ig_{\mu\nu}}{\hat{s} - M_Z^2 + iM_Z\Gamma_Z}\bar{u}(k_1)\gamma^\nu(g_V^\ell - g_A^\ell\gamma^5)v(k_2) \ , \tag{2.2}$$

where $g_V^f = \frac{g}{2\cos\theta_W}\left(T_3^f - 2Q_f \sin^2\theta_W\right)$ is the vector coupling constants for the fermion f, $g_A^f = \frac{g}{2\cos\theta_W}T_3^f$ is the axial coupling constants for the fermion f, T_3^f is weak isospin of fermion f, Q_f is fermion electric charge, M_Z and Γ_Z are Z^0 mass and width. Consequently the squared amplitude after averaging over initial spins and colours is

$$\overline{|\mathcal{M}|^2} = \frac{1}{3}\cdot\frac{1}{4}\cdot\frac{g^4}{\cos^4\theta_W}\frac{\hat{s}^2}{(\hat{s}-M_Z^2)^2 + M_Z^2\Gamma_Z^2}\left[(g_V^{q\,2} + g_A^{q\,2})(g_V^{\ell\,2} + g_A^{\ell\,2})\right] \ , \tag{2.3}$$

where the prefactor 1/3 corresponds to the colour average and 1/4 to the spin average. In the current scenario, since it involves a 2-body final state, the integration over the phase space element reads:

$$\int d\Phi_2 = \frac{1}{8\pi}\frac{|\vec{k}|}{\sqrt{\hat{s}}}d\Omega \quad \text{with } |\vec{k}| = \frac{\sqrt{\hat{s}}}{2} \ . \tag{2.4}$$

As a result, after integrating over the solid angle Ω, the partonic cross section is:

$$\hat{\sigma}(\hat{s}) = \frac{1}{3}\frac{1}{12\pi}\frac{g^4}{\cos^4\theta_W}\frac{\hat{s}}{(\hat{s}-M_Z^2)^2 + M_Z^2\Gamma_Z^2}({g_V^q}^2 + {g_A^q}^2)({g_V^\ell}^2 + {g_A^\ell}^2)\ . \tag{2.5}$$

Then, as mentioned in Chapter 1, the hadronic cross section is obtained via convolution with PDFs:

$$\sigma(s) = \sum_q \int_0^1 dx_1 \int_0^1 dx_2 f_q(x_1, \mu_F^2) f_{\bar{q}}(x_2, \mu_F^2)\hat{\sigma}(\hat{s} = x_1 x_2 s)\ . \tag{2.6}$$

Here f_q is the PDF for quark q carrying momentum fraction x, μ_F is the factorization scale, often considered $\mu_F = M_Z$ and $\hat{s} = x_1 x_2 s$ where s is hadronic centre-of-mass energy squared. Then, a change of variables is performed: $\tau = x_1 x_2 = \hat{s}/s$:

$$\sigma(s) = \sum_q \int_{\tau_{\min}}^1 d\tau \int_\tau^1 \frac{dx_1}{x_1} f_q(x_1) f_{\bar{q}}(\tau/x_1)\hat{\sigma}(\hat{s} = \tau s)\ , \tag{2.7}$$

where $\tau_{min} = 4m_l^2/s \approx 0$ for electrons or muons. Finally, the hadronic cross section for the Drell-Yan production of the Z^0 boson is

$$\sigma(s) = \sum_q \int_0^1 dx_1 \int_0^1 dx_2 f_q(x_1) f_{\bar{q}}(x_2) \frac{1}{36\pi}\frac{g^4}{\cos^4\theta_W}\frac{x_1 x_2 s({g_V^q}^2 + {g_A^q}^2)({g_V^\ell}^2 + {g_A^\ell}^2)}{(x_1 x_2 s - M_Z^2)^2 + M_Z^2\Gamma_Z^2}\ . \tag{2.8}$$

In addition, the hadronic differential cross section with respect to the invariant energy squared Q^2 can be obtained by introducing a delta function in Eq. 2.6 as:

$$\frac{d\sigma}{dQ^2} = \sum_q \int_0^1 dx_1 \int_0^1 dx_2 f_q(x_1, \mu_F^2) f_{\bar{q}}(x_2, \mu_F^2)\delta(Q^2 - x_1 x_2 s)\hat{\sigma}(\hat{s} = Q^2)\ . \tag{2.9}$$

Then, after evaluating the delta function, the differential cross section is rewritten as:

$$\begin{gathered}\delta(Q^2 - x_1 x_2 s) \implies x_2 = \frac{Q^2}{x_1 s}, \quad dx_2 = \frac{dQ^2}{x_1 s}\ , \\ \frac{d\sigma}{dQ^2} = \frac{1}{s}\sum_q \int_{Q^2/s}^1 \frac{dx_1}{x_1} f_q(x_1, \mu_F^2) f_{\bar{q}}\left(\frac{Q^2}{x_1 s}, \mu_F^2\right)\hat{\sigma}(\hat{s} = Q^2)\ .\end{gathered} \tag{2.10}$$

And by introducing the partonic cross section in it, the differential cross section finally reads as

$$\begin{gathered}\frac{d\sigma}{dQ^2} = \frac{1}{36\pi s}\frac{g^4}{\cos^4\theta_W}({g_V^q}^2 + {g_A^q}^2)({g_V^\ell}^2 + {g_A^\ell}^2) \\ \int_{Q^2/s}^1 \frac{dx_1}{x_1} f_q(x_1, \mu_F^2) f_{\bar{q}}\left(\frac{Q^2}{x_1 s}, \mu_F^2\right)\frac{Q^2}{(Q^2 - M_Z^2)^2 + M_Z^2\Gamma_Z^2}\ .\end{gathered} \tag{2.11}$$

The hadronic differential cross section with respect to the dilepton transverse momentum p_T^{ll} is obtained at NLO in QCD, since at LO the Z boson is produced with zero transverse momentum. Therefore, to get a nonzero Z^0 p_T^{ll}, at least one extra real parton is needed in the final state, therefore both channels: $q\bar{q} \to Zg$ and $qg \to Zq$ where $Z \to l^- l^+$ are needed.

First, kinematic parameters are defined: the partonic CM energy is $\hat{s} = (p_1 + p_2)^2$, the Mandelstam variables are: $\hat{t} = (p_1 - q)^2, \quad \hat{u} = (p_2 - q)^2$, the squared Z^0 boson transverse momentum: $p_T^2 = \frac{\hat{t}\hat{u} - M_Z^2\hat{s}}{\hat{s}}$. Then, solving for $\hat{t}$ in terms of p_T^2: $\hat{t} = \frac{1}{2}\left(M_Z^2 - \hat{s} \pm \sqrt{(\hat{s} - M_Z^2)^2 - 4\hat{s}p_T^2}\right)$ with integration limits: $0 \leq p_T^2 \leq (\hat{s} - M_Z^2)^2/(4\hat{s})$.

For $q\bar{q} \to Z^0 g$ the spin- and colour-averaged squared amplitude is:

$$|\overline{\mathcal{M}}|^2_{q\bar{q}\to Zg} = \frac{1}{9} g_s^2 \frac{g^2}{\cos^2\theta_W}({g_V^q}^2 + {g_A^q}^2)\frac{\hat{t}^2 + \hat{u}^2 + 2M_Z^2\hat{s}}{\hat{t}\hat{u}} \,, \tag{2.12}$$

where $g_s^2 = 4\pi\alpha_s$ and α_s is the strong coupling at scale μ_R and similarly, $g^2 = 4\pi\alpha$ and α is the weak coupling constant.

For $qg \to Z^0 q$ the spin- and colour-averaged squared amplitude is:

$$|\overline{\mathcal{M}}|^2_{gq\to Zq} = \frac{1}{32} g_s^2 \frac{g^2}{\cos^2\theta_W}({g_V^q}^2 + {g_A^q}^2)\frac{\hat{s}^2 + \hat{u}^2 + 2M_Z^2\hat{t}}{-\hat{s}\hat{u}} \,. \tag{2.13}$$

Consequently, the partonic differential cross section in p_T is

$$\frac{d\hat{\sigma}_{ij}}{dp_T^2} = \frac{1}{16\pi\hat{s}^2}\overline{|\mathcal{M}_{ij}|^2}\frac{1}{\sqrt{1 - 4p_T^2\hat{s}/(\hat{s} - M_Z^2)^2}} \,, \tag{2.14}$$

where $i, j = q\bar{q}, qg, \bar{q}g$. Then, the hadronic differential cross section reads

$$\frac{d\sigma}{dp_T^2} = \sum_{i,j}\int_0^1 dx_1 \int_0^1 dx_2 f_i(x_1, \mu_F^2) f_j(x_2, \mu_F^2)\frac{d\hat{\sigma}_{ij}}{dp_T^2}(\hat{s} = x_1 x_2 s, p_T^2, \mu_R^2) \,, \tag{2.15}$$

where μ_R is the renormalization scale and μ_F is the factorization scale, with the integration limits: $x_1 x_2 s \geq M_Z^2 + 2p_T\sqrt{M_Z^2 + p_T^2}$ which correspond to the kinematic threshold and the partonic limit: $0 < p_T^{ll} < \frac{(\hat{s} - M_Z^2)^2}{4\hat{s}}$.

2.1.2 Electroweak production of Z boson

As explained in the introduction of the current chapter, the EW Zjj process includes the contribution from the VBF Z, Z-boson bremsstrahlung and multi-peripheral production in a gauge-invariant set, as shown in Fig. 2.2. Gauge cancellations cause

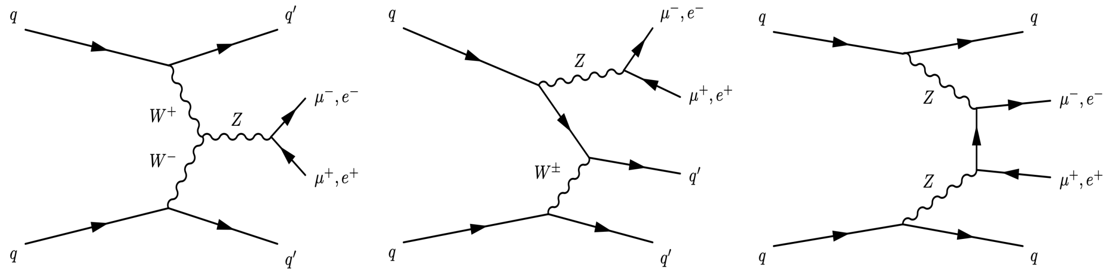


Figure 2.2: Representative leading-order Feynman diagrams for EW Zjj production at the LHC: vector boson fusion (left), Z-boson bremsstrahlung (centre) and multi-peripheral production (right), from Ref. [63].

strong negative interference between the VBF process and the other two production modes, with the bremsstrahlung-like process giving the largest contribution. Interference with multi-peripheral production only occurs when the dilepton mass is close to the Z boson mass, as stated in Ref. [64]. As a result, it is not possible to claim an absolute dominance of any of these three individual processes when considered independently due to their overlapping contributions. However, once a specific kinematic selection is applied, such as one enhancing the VBF Z production, the VBF topology becomes the dominant component in the measured electroweak Zjj yield.

2.1.2.1 Vector Boson Fusion production of Z boson

As discussed in the works of Zeppenfeld and collaborators [65, 66], the electroweak Zjj process comprises Feynman diagrams such as the ones depicted in Fig. 2.3, among which the charged-current contributions (diagrams a, b, c, d, and g) dominate due to the stronger quark coupling to the W boson compared to the photon and Z. Within the narrow-width approximation, this subset forms a gauge-invariant set, as explained in [66]. For the selection described in Section 2.2, corresponding to the experimental analysis in Chapter 6, the relative contributions at $\sqrt{s} = 13$ TeV are approximately 70% from (a)+(d), 20% from (b)+(c), and about 10% from (g), as typically computed with Monte Carlo tools [67] and illustrated in Fig. 2.9, where the η_3^* distribution shows, in green, the VBF Z contribution corresponding to 10% of the electroweak Zjj yield, shown in blue. However, under the typical vector boson fusion (VBF) phase-space selection, characterized by very forward jets, a large dijet mass threshold (~ 1 TeV), and described in detail in [65], diagram (g) provides the leading contribution, an assumption adopted in the present section.

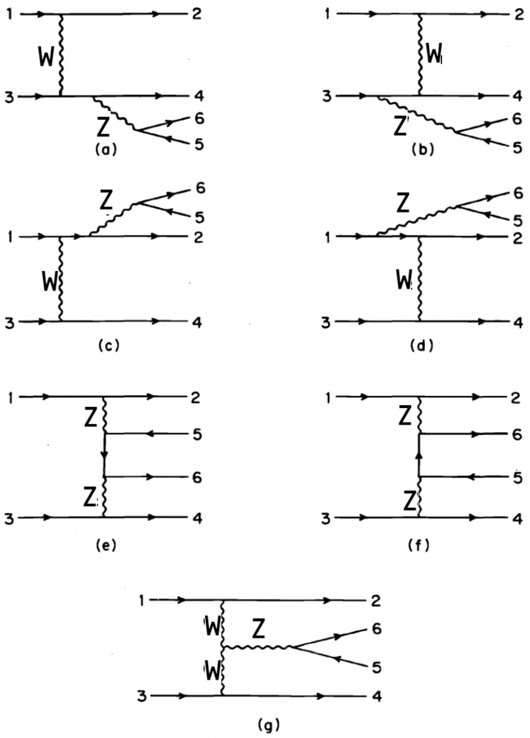


Figure 2.3: Complete set of Feynman graphs for EW Zjj production, modified from [66].

The VBF Z production at the LHC at LO is: $q(p_1) + \bar{q}'(p_2) \to q(p_3) + \bar{q}'(p_4) + Z(k)$, which from a quark perspective is: $u(p_1) + \bar{d}(p_2) \to u(p_3) + \bar{d}(p_4) + Z(k)$ with t-channel propagators $t_1 = (p_1 - p_3)^2$, $t_2 = (p_2 - p_4)^2$. This process involves t-channel $W^{\pm}$ exchange to produce the Z^0 boson, which constitutes the main charged-current contribution.
As previously explained, the full amplitude is $\mathcal{M} = \sum_{i=a}^{e} \mathcal{M}_i$. Here the gauge-invariant VBF Z is composed by two types of contributions: fermionic Z emission (diagrams (a)-(d) in Fig. 2.3) and triple gauge vertex Z emission (diagram (g)), but only the latter will be explicitly considered for the current discussion.
The matrix element for the triple gauge vertex Z emission reads:

$$\begin{aligned}\mathcal{M}^{\mu} &= \bar{u}(p_3)\,\Gamma_W^{\alpha}\,u(p_1)\,\frac{-i}{t_1 - M_W^2}\left(g_{\alpha\gamma} - \frac{q_{1\alpha}q_{1\gamma}}{M_W^2}\right)\\ &\times \bar{u}(p_4)\,\Gamma_W^{\beta}\,v(p_2)\,\frac{-i}{t_2 - M_W^2}\left(g_{\beta\delta} - \frac{q_{2\beta}q_{2\delta}}{M_W^2}\right)\\ &\times V^{\gamma\delta\mu}(q_1, q_2, k)\,\epsilon_{\mu}(k)\;.\end{aligned} \tag{2.16}$$

This matrix element factorizes into five distinct terms representing the complete t-channel W^--exchange topology with WWZ vertex emission. The term $\bar{u}(p_3)\Gamma_W^{\alpha}u(p_1)$ describes the upper quark charged-current vertex where the incoming $u(p_1)$ emits a W^- boson and scatters to the outgoing $u(p_3)$, with $\Gamma_W^{\alpha} = \frac{ig}{\sqrt{2}}\gamma^{\alpha}P_L$. The lower antiquark current $\bar{u}(p_4)\Gamma_{W\alpha}v(p_2)$ describes the charge-conjugate process $\bar{d}(p_2) \to \bar{d}(p_4) + W^+$, with index α contracted between the two currents. The t-channel $W^{\pm}$ propagators $\frac{-i(g_{\alpha\gamma} - q_{\alpha}q_{\gamma}/M_W^2)}{t_1 - M_W^2}$ carries momentum $q = p_1 - p_3$ with the standard gauge projector ensuring transversality. The WWZ triple gauge vertex $V^{\gamma\mu\rho}(q,k)$ connects the exchanged W (index γ) to the outgoing Z boson (index μ), while ϵ_{ρ} represents the Z polarization vector (summed over physical polarizations). Here the WWZ triple gauge vertex is:

$$V^{\gamma\mu\rho}(q,k) = g^{\gamma\mu}(2q-k)^{\rho} + g^{\mu\rho}(k-q)^{\gamma} + g^{\rho\gamma}(q-k)^{\mu}\;. \tag{2.17}$$

This WWZ vertex describes Z emission from a t-channel W propagator with incoming momentum q (index γ) and Z momentum k. The first term $g^{\gamma\mu}(2q-k)^{\rho}$ couples the incoming W Lorentz indices γ, μ directly to the outgoing Z index ρ through momentum conservation $2q - k = q + (q - k)$. The second term $g^{\mu\rho}(k-q)^{\gamma}$ contracts the intermediate W$^*(q-k)$ index μ with Z index ρ, carrying the momentum kick $(k-q)^{\gamma}$ back to the incoming W. The third term $g^{\rho\gamma}(q-k)^{\mu}$ completes the cyclic gauge structure by coupling incoming W index γ with outgoing Z index ρ, with $(q-k)^{\mu}$ along the continuing W* propagator, satisfying the Ward identity $q_{\gamma}V^{\gamma\mu\rho} = 0$ for the single propagator emission topology. Here the coupling is $g_{WWZ} = -e\cot\theta_W$. The spin- and colour-averaged squared matrix element is:

$$\overline{|\mathcal{M}_g|^2} = \frac{1}{2\cdot N_c^2}\sum_{\text{spins,pol}}|\mathcal{M}_g|^2 = \frac{1}{18}\sum_{\text{spins,pol}}|\mathcal{M}_g|^2\;, \tag{2.18}$$

behaving as:

$$\overline{|\mathcal{M}_g|^2} \propto \frac{1}{|t_1^2||t_2^2|}\;, \tag{2.19}$$

where for small t_i, each propagator term $(t_i - M_W^2)^{-1}$ in Eq. 2.16 effectively maps to a factor $(p_{T_i}^2 + M_W^2)^{-1}$, reflecting the characteristic behaviour of the t-channel $W^{\pm}$ fusion in the VBF Z topology, as illustrated in the distributions of Section 3.5.

2.2 Phenomenology of inclusive and electroweak production of Z boson

The production of the Z^0 boson is dominated by the Drell-Yan process, with electroweak production and especially the VBF topology being subdominant. To investigate this behaviour, both the Drell-Yan and electroweak Zjj samples were simulated using `MadGraph5_aMC@NLO` [68] with the `NNPDF31_nnlo_as_0118` PDF set, and interfaced to `Pythia` [69] for parton showering and hadronisation. The VBF Z process was simulated with `Powheg` [70] using the `PB-TMDNLO-HERAI+II-2018-set2` PDF set, and likewise interfaced to `Pythia` [69]. Then a Rivet [71] plugin was developed, implementing a series of basic kinematic cuts at particle level, and a set of additional cuts to enhance the VBF topology, denoted as **VBF Z-enhanced selection** (see Tab. 2.1).

Inclusive selection	**VBF Z-enhanced selection**
Z peak: 75 GeV < M_{ll} < 106 GeV, $N_{leptons} = 2$	
$\lvert\eta_l\rvert < 2.4$, $\lvert\eta_j\rvert < 4.7$ $p_{T_{leptons}} > 20$ GeV, $p_{T_{jets}} > 30$ GeV	
No additional cuts	$p_{Tl_{0,1}} > 40, 20$ GeV $p_{Tj_{0,1}} > 50, 50$ GeV $N_{jets} \geq 2$ $M_{jj} > 200$ GeV $\Delta\Phi_{ll} < 2.8$ $\lvert\Delta\eta_{jj}\rvert > 1.5$

Table 2.1: Fiducial selection at particle level implemented in the Rivet plugin.

The resulting particle-level distributions, shown in Figs. 2.4-2.9, illustrate the contributions of the different production topologies under these selections. In these figures, Drell-Yan Z^0 production (QCD background) is depicted in red, electroweak Z^0 production (EW Zjj) in blue, and VBF Z^0 production (VBF Z) in green, clearly highlighting the relative impact of the various cuts and topologies.

Overall, in all distributions the inclusive selection yields rates approximately one order of magnitude higher than the VBF Z-enhanced selection.

In the dilepton transverse momentum distribution, $p_{T_{\mu\mu}}$ shown in Fig. 2.4, both selections exhibit similar shapes, although the peak is shifted due to the leading lepton $p_T > 40$ GeV requirement.

For both leading (Fig. 2.5) and subleading (Fig. 2.6) jets, the absolute pseudorapidity distributions $|\eta|$ of the EW Zjj and VBF Z signals are flat in the range $0 < |\eta| < 2.5$ under the inclusive selection. In the VBF Z–enhanced selection, however, they exhibit a mild rise in this central region, followed by a steeper decrease for $2.5 < |\eta| < 4.5$. The QCD background shows the opposite trend: a slight decrease between $0 < |\eta| < 2.5$ in the inclusive selection, and an approximately flat behaviour in the same range for the VBF Z–enhanced selection, with a more pronounced fall-off beyond $|\eta| \approx 2.5$ in both cases. These features arise from the kinematic cuts, particularly the $|\Delta\eta_{jj}| > 1.5$ requirement, preferentially select events where the two jets arise from forward-backward VBF topology. In this process, each proton emits a $W^{\pm}$ boson, which then fuse to produce the Z^0 boson at central rapidities, while

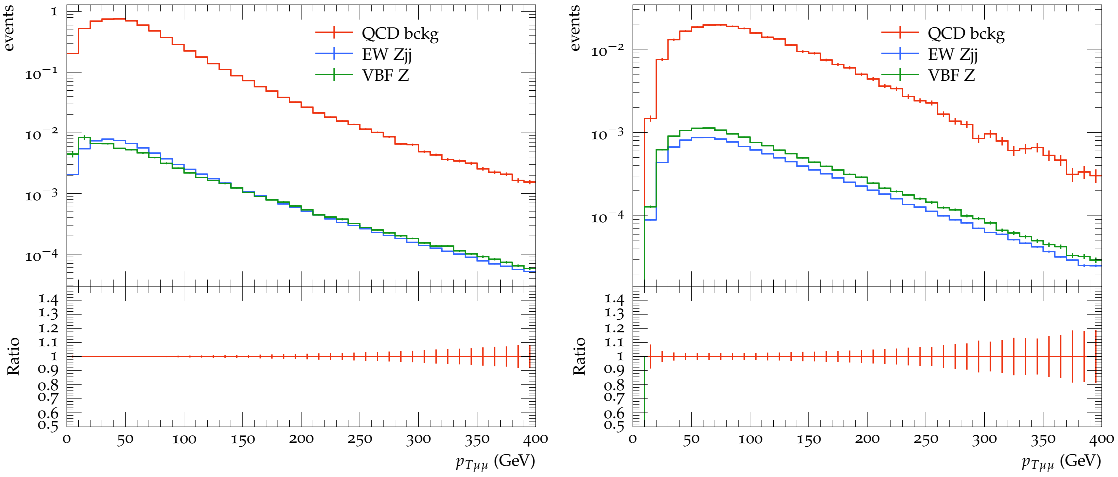


Figure 2.4: p_{Tll} distribution for the inclusive selection (left) and VBF Z-enhanced region (right) normalised to the cross section.

the initial-state jets recoil to forward and backward regions of the detector.

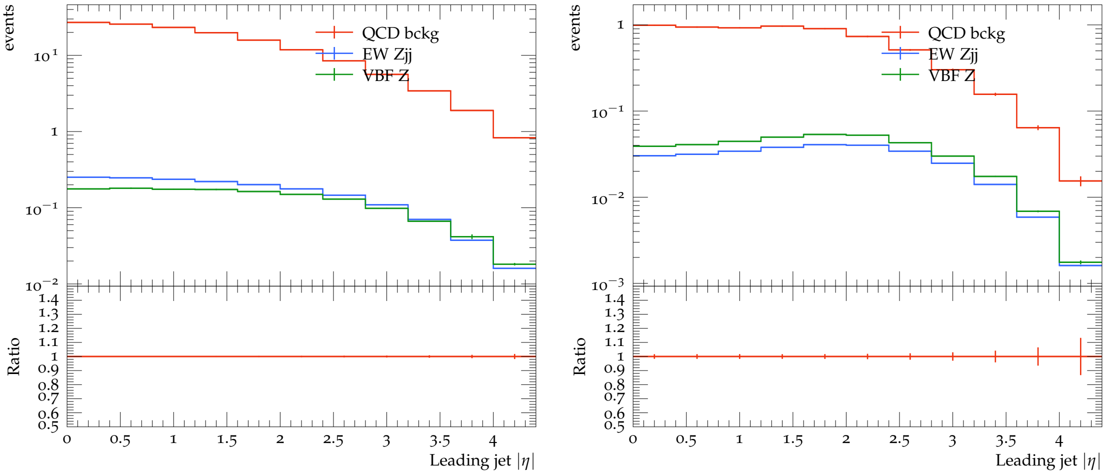


Figure 2.5: Leading jet $|\eta|$ distribution for the inclusive selection (left) and VBF Z-enhanced region (right) normalised to the cross section.

The leading (Fig. 2.7) and subleading (Fig. 2.8) jet transverse momentum distributions exhibit a similar falling behaviour for all processes and selections. The primary difference is the location of the peak: in the VBF Z-enhanced selection it occurs at $p_T \approx 80\,\text{GeV}$ for the QCD background and at $p_T \approx 130\,\text{GeV}$ for the EW Zjj and VBF Z signals, whereas in the inclusive selection all processes peak at $p_T \approx 50\,\text{GeV}$, reflecting the looser transverse momentum requirements. This shift arises because the VBF Z-enhanced region selects a phase space dominated by the VBF-like topology of the EW Zjj process, where $W\pm$ bosons radiated from the incoming protons fuse to produce the Z^0, which, through kinematic requirements favouring large dijet mass and rapidity separation, naturally selects harder jets with higher p_T.
From the distributions included in this section, one of the most relevant ones is the η_3^* defined as $\eta_3^* = y_{j2} - \frac{y_{j0}+y_{j1}}{2}$ (Fig. 2.9), which measures the rapidity position of the third jet relative to the centre of the VBF tagging-jet system. This observable clearly shows its discrimination capacity between a VBF Z enriched topology and

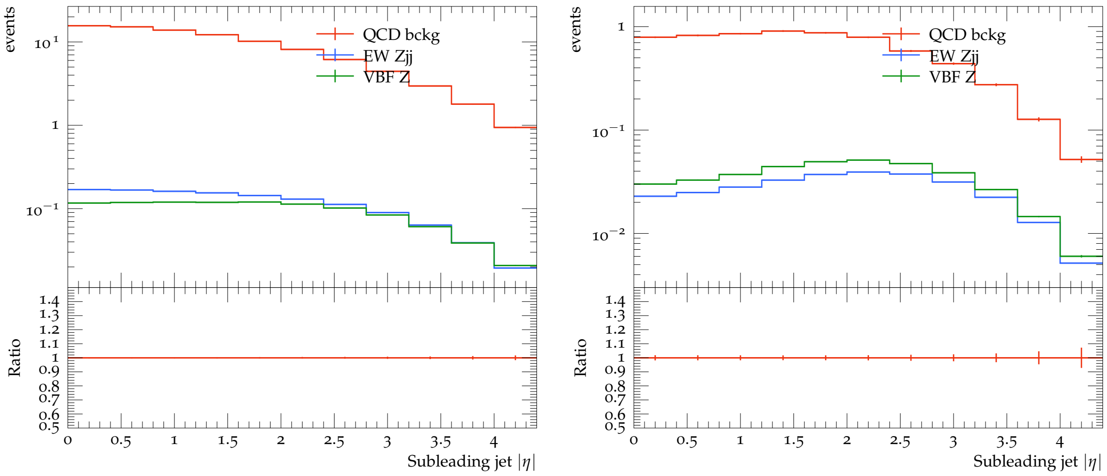


Figure 2.6: Subleading jet $|\eta|$ distribution for the inclusive selection (left) and VBF Z-enhanced region (right) normalised to the cross section.

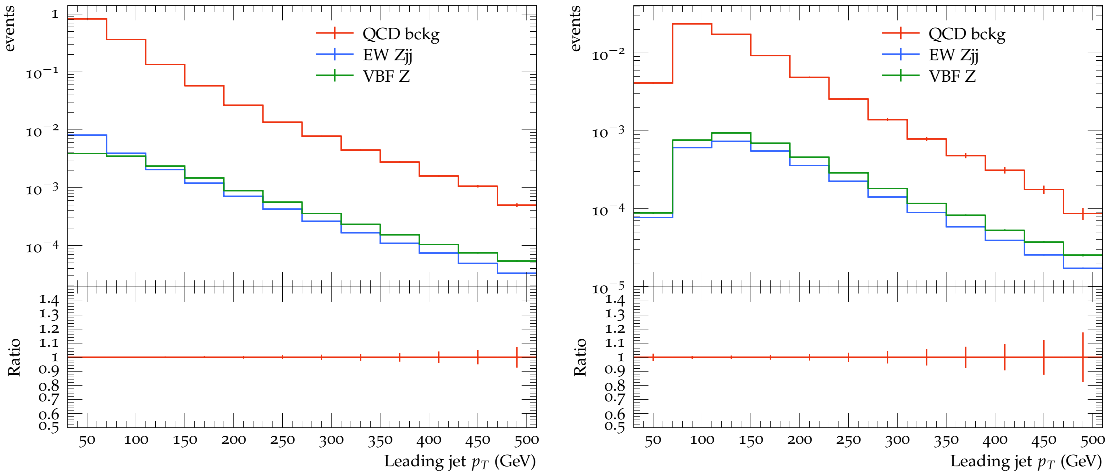


Figure 2.7: Leading jet p_T distribution for the inclusive selection (left) and VBF Z-enhanced region (right) normalised to the cross section.

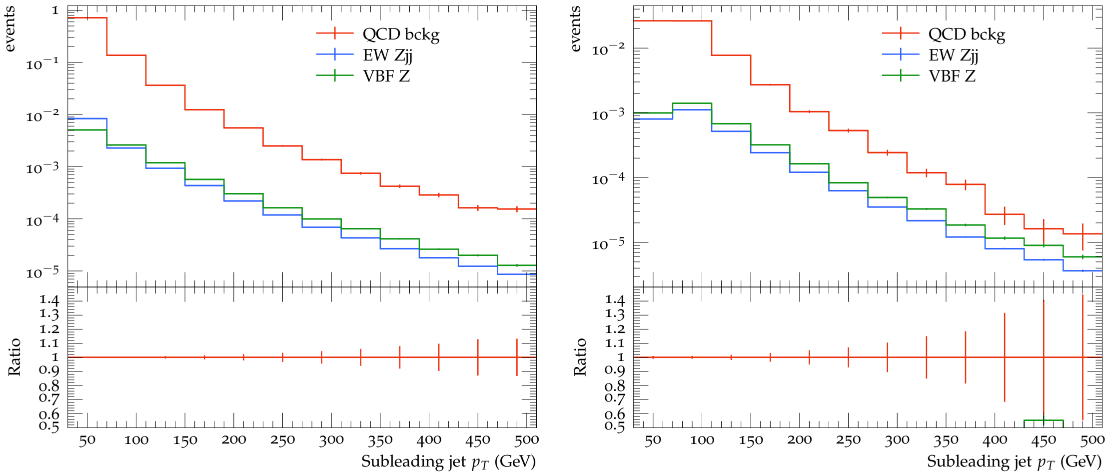


Figure 2.8: Subleading jet p_T distribution for the inclusive selection (left) and VBF Z-enhanced region (right) normalised to the cross section.

a DY dominant one when a VBF Z-enhanced kinematic selection is applied [6]. Fig. 2.9 presents how in the VBF Z process, the emission of an extra jet inside the gap is colour-suppressed leading to the characteristic gap around $\eta_3^* = 0$. In contrast with the EW Zjj curve, where a slight less pronounced suppression is observed due to the leading contribution from the VBF Z. Finally, the curve corresponding to the QCD background, shows an higher number of events around $\eta_3^* = 0$ due to the colour exchange and consequently larger central radiation.

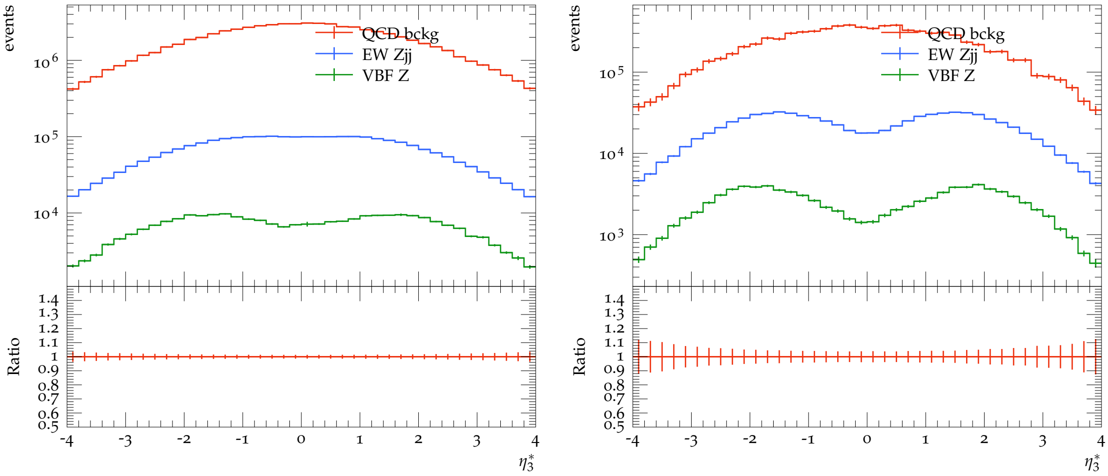


Figure 2.9: η_3^* distribution for the inclusive selection (left) and VBF Z-enhanced region (right).

Chapter 3

Electroweak corrections

Modern experimental facilities such as the LHC have reached unprecedented precision in their measurements. For several processes, the experimental uncertainties are now comparable to, or even smaller than, the theoretical ones [72]. This progress demands improved theoretical predictions to ensure a consistent interpretation of the data. Two fundamental ingredients in this context are the partonic cross section and the PDFs, introduced in Eq. 1.33 of Section 1.2.1.

As discussed in Section 1.2.2, PDFs are determined by fitting an initial distribution at a reference scale μ_0 and evolving it perturbatively to a higher scale μ, while partonic cross sections are themselves perturbative quantities. The accuracy of both components can be systematically enhanced by including higher-order perturbative corrections. Although such expansions are predominantly expressed in terms of the strong coupling α_s, electroweak effects become increasingly significant at high energies [73]. In particular, the electroweak production of the Z boson, especially via vector boson fusion (VBF), is highly sensitive to electroweak corrections, as explored throughout this work.

Corrections are classified as virtual or real. Virtual corrections arise from loop diagrams involving virtual particles, while real corrections correspond to the emission of additional real particles in the final state. These corrections apply to both massless and massive particle production processes. For massless particles, corrections introduce infrared or mass singularities, which cancel against those from real corrections; for massive particles, the mass acts as a regulator, preventing such singularities. In Abelian theories like QED, the Bloch-Nordsieck (BN) theorem [74] governs the cancellation of divergences, whereas in non-Abelian theories such as QCD, the Kinoshita-Lee-Nauenberg (KLN) theorem [75] is employed.

This chapter presents the nature of virtual and real corrections in Sections 3.1 and 3.2, then Section 3.3 explores the contribution from EW bosons in parton, photon and heavy boson densities and the appropriate treatment of heavy boson masses. Section 3.4 validates heavy boson densities through DIS HERA data. Finally Section 3.5 discusses the application of heavy boson TMDs to VBF Z production.

As presented in Section 2.1.2, the EW Zjj process, when simulated in a non-restrictive way, apart from VBF production, it includes Z^0-boson bremsstrahlung and multi-peripheral topologies. Therefore, since the VBF Z is the focus of the phenomenology studies presented in this thesis, the analytical form of the EW corrections will be explained with respect to its Feynman diagram.

3.1 Virtual corrections

Virtual electroweak (EW) corrections modify both the total VBF Z cross section and the shapes of key differential observables, such as $p_{T,j}$, $p_{T,jj}$, m_{jj}, and angular correlations between the Z boson and the tagging jets, $\Delta\Phi_{Z,jj}$ [76, 77]. In the kinematic regime relevant for VBF (large space-like momentum transfers along the t-channel lines and large invariant masses), the dominant effect is a negative correction of several percent, increasingly driven at high scales by Sudakov double logarithms of the form $\ln^2\left(\frac{Q^2}{M_V^2}\right)$, with $V = W, Z$, which grow with the typical hard scale Q in the event [78]. These logarithms originate from virtual soft-collinear W, Z exchange and are particularly relevant in the high-m_{jj} and high-p_T tails used to enhance the VBF-like topology over backgrounds.

The discussion below is structured according to the three categories of one-loop electroweak corrections contributing to the VBF Z amplitude: self-energy insertions on internal gauge-boson propagators and external fermion legs, vertex corrections to the qqW and WWZ interactions, and box diagrams linking the two quark lines, as stated in Ref. [79]. To facilitate the analysis of these contributions, the VBF Z process (see Fig. 3.1) is decomposed into its basic vertex components.

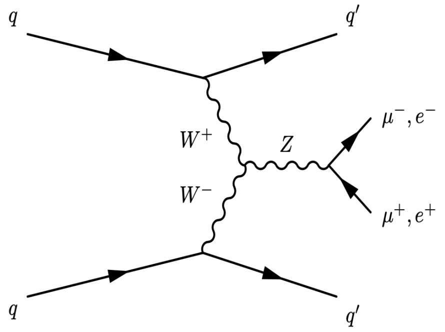


Figure 3.1: VBF Z Feynman diagram.

General expressions and renormalisation prescriptions follow the conventions of Denner and collaborators [76, 80, 81].

3.1.1 Gauge-boson and fermion self-energies

Significant electroweak corrections originate from self-energy insertions on space-like $W^\pm$ and Z^0 bosons exchanged in the t-channel. Following W. Hollik's lectures on Electroweak Theory [82], the $W^\pm$ and Z^0 boson self-energies can be expressed in terms of scalar functions $\Sigma^{W,Z}$, as given in Eq. (43) of [82]:

$$\Sigma^{W,Z}_{\mu\nu} = g_{\mu\nu}\,\Sigma^{W,Z} + \cdots \tag{3.1}$$

In the 't Hooft-Feynman gauge ($\xi = 1$), this modifies the propagator to

$$\frac{-i\,g_{\mu\nu}}{q^2 - M^2_{W,Z} + \Sigma^{W,Z}(q^2)}, \tag{3.2}$$

as shown in Eq. (45) of [82]. The one-loop self-energy corrections for gauge bosons and fermions are discussed in detail in Appendix B of Denner's review on electroweak corrections [76]. Expressed in terms of Passarino-Veltman functions [83], these self-energies exhibit logarithmic asymptotic behaviour at high energies: for $q^2 \gg M^2_{W,Z}$, the W and Z self-energies scale as $\ln(q^2/M^2_{W,Z})$, while for $p^2 \gg M^2_{W,Z}, m_f^2$, the fermionic self-energies scale as $\ln(p^2/M^2_{W,Z})$ for light fermions. Here, q^2 denotes the momentum transfer of the t-channel $W^\pm$ boson, and p^2 the momentum of the external fermion in the loop.

3.1.2 Vertex corrections: qqW and WWZ

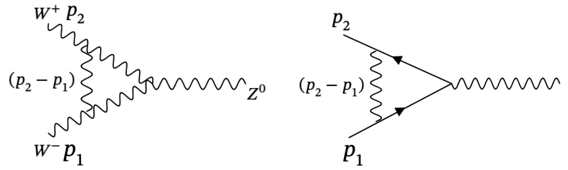


Figure 3.2: Virtual EW corrections to $W^+W^-Z^0$ vertex (left) and $qq'W^\pm$ vertex (right).

The two vertices controlling the VBF Z amplitude are the charged-current $q_iq_jW^\pm$ vertices on each quark line and the central $W^+W^-Z^0$ triple-gauge vertex (see Fig. 3.2). The $q_iq_jW^\pm$ vertex at tree level is given by

$$\Gamma^\mu_{\text{tree}} = \frac{g}{\sqrt{2}}\gamma^\mu P_L \quad \text{where} \quad \mathrm{P_L} = \frac{1-\gamma_5}{2}\ . \tag{3.3}$$

Then, a one-loop vertex correction can be introduced, mediated by an internal vector boson, while a fermion-antifermion pair may appear in loops involving Z^0 or γ. The corresponding loop amplitude is then given by the standard expression:

$$\begin{aligned}\Gamma^\nu(k) = i\frac{g^2}{16\pi^2}\int \frac{d^4k}{(2\pi)^4}\gamma^\alpha(c_V - c_A\gamma^5)\frac{\not{p}_1 - \not{k} + m_f}{(p_1-k)^2 - m_f^2}\gamma^\mu P_L \\ \frac{\not{p}_3 - \not{k} + m_f}{(p_3-k)^2 - m_f^2}\gamma^\beta(c_V - c_A\gamma^5)\left(\frac{-ig_{\alpha\beta}}{k^2 - M_V^2}\right)\ .\end{aligned} \tag{3.4}$$

Then, following Michael Melles' calculation in Sec. 2-3 of "Electroweak radiative corrections in high energy processes" [84], the leading high-energy contribution from soft/collinear regions of this integral yields the universal Sudakov double logarithm $\Delta\Gamma^\nu \propto \ln^2 \frac{q^2}{M_W^2}$.
Similarly, the renormalised WWZ vertex is parametrised as stated in Chapter 2 in Eq. 2.17 and applying similar techniques, in the high energy regime the correction is led by double logarithms as well [76, 84].

3.1.3 Box diagrams

Electroweak box diagrams connect the two quark lines through the exchange of two gauge bosons, comprising charged-current (WW), neutral-current (ZZ, γZ, $\gamma\gamma$), and mixed (WZ, $W\gamma$) topologies [80, 85] (see Fig. 3.3). As demonstrated by Darvishi and Masouminia [85], and in accordance with the explicit treatment of electroweak

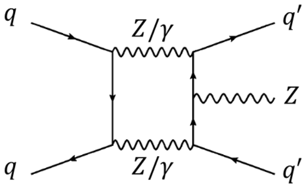


Figure 3.3: Virtual EW box correction to the VBF Z process from Ref. [85].

radiative corrections by Denner in [86] and Hollik in [87], these diagrams are ultraviolet finite and require no independent renormalisation counterterms. In the high-energy regime $s, |t| \gg M_W^2$, the weak-box contributions display leading double-logarithmic Sudakov behaviour:

$$\Sigma_{\rm EW}^{\rm box} \propto \sum_{i<j} C_{ij} \ln^2\left(\frac{|s_{ij}|}{M_W^2}\right), \tag{3.5}$$

where s_{ij} are Mandelstam invariants built from pairs of external legs and C_{ij} depend on the electroweak charges of the quarks [80, 81, 88].
Numerically, box corrections to the inclusive VBF Z cross section are typically at the few-percent level, but they can become non-negligible in specific kinematic regions, for instance, at large dijet invariant masses or in distributions sensitive to azimuthal correlations between the tagging jets and the Z boson-motivating their inclusion in precision predictions [77, 85].

3.2 Real corrections

The VBF Z process at LO is: $q_1(p_1) + q_2(p_2) \to q_3(p_3) + q_4(p_4) + Z(p_Z)$ where the real EW corrections correspond to the emission of an extra boson $q_1q_2 \to q_3q_4Z + X, \quad X = \gamma, W, Z$.

3.2.1 Real photon emission

In vector boson fusion (VBF) production of a Z boson, real photon emission from charged external quark legs can be approximated in the soft photon limit as a factorized correction to the leading-order amplitude, as derived in Ref. [89]:

$$\mathcal{M}_\gamma \approx \mathcal{M}_{\rm LO} \cdot e \sum_i Q_i \frac{p_i \cdot \epsilon^*}{p_i \cdot k}, \tag{3.6}$$

where the sum runs over charged external quark momenta p_i with electric charges Q_i, ϵ^* is the photon polarization vector, k is the photon momentum, and e is the electromagnetic coupling. The central Z boson, being neutral, does not contribute to photon emission.
The squared amplitude, including interference among photon emissions from different legs, reads

$$|\mathcal{M}_\gamma|^2 \approx |\mathcal{M}_{\rm LO}|^2 e^2 \sum_{i,j} Q_i Q_j \frac{p_i \cdot p_j}{(p_i \cdot k)(p_j \cdot k)}. \tag{3.7}$$

Integrating over the soft photon phase space below the energy cut-off ΔE yields the soft correction to the cross section:

$$d\sigma^{\text{soft}}_{\text{real},\gamma} = d\sigma_{\text{LO}} \cdot \delta_{\text{soft}}, \quad \delta_{\text{soft}} = \frac{\alpha_{\text{em}}}{\pi} \sum_{i,j} Q_i Q_j \int_{E_\gamma < \Delta E} \frac{d^3k}{2E_\gamma (2\pi)^3} \frac{p_i \cdot p_j}{(p_i \cdot k)(p_j \cdot k)}. \tag{3.8}$$

Real photon emissions become finite when combined with virtual photon corrections, as the virtual photon loops cancel the infrared divergences from real emission in agreement with the Bloch–Nordsieck and KLN theorems, as shown in Refs. [76, 90].

In VBF Z production, real photon emission constitutes a small but non-negligible QED radiative correction, typically modifying the leading-order cross section by about 1–2%, analogous to parallel studies of the VBF Higgs process with similar phase space selections [91]. This effect arises primarily from photon radiation off the charged tagging quark legs and is essential for infrared-safe, precise predictions. Accounting for these corrections significantly improves the theoretical accuracy of VBF Z processes, which serve as important benchmarks for electroweak measurements at hadron colliders.

3.2.2 Real Z^0 and $W^\pm$ boson emissions

In pure electroweak VBF Z production, additional real Z or $W^\pm$ emission can occur while preserving the characteristic t-channel $W^\pm$-boson fusion topology and tagging-jet kinematics. Such contributions arise from radiation off the external quark legs or the exchanged vector bosons in the VBF diagrams. Owing to the large weak-boson masses, these real-emission processes are suppressed by both the coupling strength and the available phase space relative to the Born-level $qq \to qqZ$ process, constituting a small higher-order electroweak correction. Although typically much smaller than real-photon emission, they remain relevant for precision studies of electroweak radiative effects [76, 78].

For real $W^\pm$ emission, the charged-current nature of the interaction allows quark-flavour transitions, leading to distinct flavour structures and kinematic distributions compared with neutral-current radiation. The overall impact on the inclusive VBF Z cross section is numerically small but becomes significant in precision electroweak analyses and in kinematic regions where charged-current effects are enhanced [76, 78].

In summary, in VBF Z production, EW corrections at next-to-leading order arise from several sources. Virtual W and Z self-energy corrections modify the t-channel propagators, reducing the cross section by a few percent, especially at high energies due to Sudakov logarithms. Vertex corrections at the $qq'Z$ and $qq'W$ interactions similarly decrease the rate, and become more relevant in events with high momentum transfer. Box diagrams introduce interference effects between quark lines, contributing negatively but at the sub-percent level, while fermion self-energies ensure proper renormalisation with negligible numerical impact. Real photon emission partially compensates the virtual photon loops in the soft limit, mostly affecting soft and collinear radiation and contributing at the percent level. Real W or Z emissions contribute positively but remain small ($< 2\%$) for inclusive cross sections [90], though they can be more visible in high-invariant-mass dijet events. Taken together, the combined effect of virtual and real EW corrections reduces the inclusive VBF Z cross section by roughly 5–10% at LHC energies, with stronger negative corrections in high-p_T or high-dijet-mass regions, demonstrating the need for an appropriate treatment of these contributions [92].

3.3 Electroweak contribution in particle densities

Complementing the broader studies presented in Chapter 2, Sections 3.1 and 3.2 further illustrate the significance of electroweak corrections in EW Zjj production processes, particularly within the VBF Z topology. This underlines the broader importance of the electroweak sector, which can also be investigated through the study of parton densities. Therefore, the current section presents a fully parametrised determination of parton densities from the extended DGLAP equation where both photon and EW boson terms are included following the work presented in Ref. [93], in which I am one of the main authors. This is implemented via the PB method due to its flexibility when implementing physical constraints and its capability to obtain both collinear and transverse momentum-dependent parton densities. The resulting densities are available in LHApdf [94] and TMDlib [95, 96] format.

As stated in Section 1.2. the DGLAP evolution equation describes the parton probability density as previously shown in Eq. 1.41. This can be re-expressed in terms of momentum-weighted PDFs, with partons carrying a momentum fraction x at an energy scale μ as

$$\mu^2 \frac{\partial f_a(x/z,\mu^2)}{\partial \ln \mu^2} = \sum_b \int_x^1 \frac{dz}{z} P_{ab}\left(z, \alpha_{eff}(\mu^2)\right) f_b\left(\frac{x}{z},\mu^2\right) . \tag{3.9}$$

Following the notation from [31], the splitting function $P_{ab}\left(z, \alpha_{eff}(\mu^2)\right)$ can be re-expressed so that non-branching probabilities and branching probabilities are disentangled, according to the PB method, as explained in Chapter 1. In addition, one of the main strengths of this method is its capability to obtain the transverse momenta of the emitted particles once the evolution scale is matched to a physical scale, which in the current derivation, is the rescaled transverse momentum from angular ordering.

Since quarks can radiate not only gluons, but photons and heavy bosons as well, the coupling α_S present in the PB PDF derivation, is now substituted by the α_{eff}, as outlined in Eq. 3.9, which will adopt different values according to the particles related to it.

In the case of QCD partons, $\alpha_{eff} = \alpha_S$, recovering the original scenario from Section 1.2.2 as a result.

In the case of the photon, studies deriving their collinear densities are vastly present in the literature [97–105]. In order to achieve a comprehensive description, both the intrinsic and dynamical components of the evolution must be taken into account. This consideration becomes particularly relevant for studies at low μ^2, where the intrinsic contribution dominates, as discussed in [99, 102, 105]. The intrinsic component is modelled via different methods: (i) through one-photon emission off valence quarks in the leading-logarithmic approximation [102], (ii) by means of a two-parameter ansatz motivated by valence quark radiation as implemented in

CT14 NLO [99], or (iii) from an input distribution derived from the master expression based on deep inelastic scattering (DIS) data and elastic proton measurements, commonly known as the "LUXqed formulation" [105]. Nevertheless, the current study focusses on the high energy regime while keeping consistency between the photon and the heavy boson treatment, and therefore the intrinsic component of the photon is neglected. The effective coupling for the photon is $\alpha_{eff} = \alpha_{em}$.
In the case of heavy bosons, effective heavy boson densities were initially explored in Refs. [106–109]. Here the Z^0 and $W^\pm$ bosons are initially considered as partons inside the proton, leading to the so-called "Effective $W^\pm/Z^0$ Approximation (EWA)", and the inclusion of a transverse momentum dependence was exposed in Ref. [110]. Additionally, treating heavy bosons similarly to gluons or massless photons in the high energy regime, has been experimentally supported by measurements performed at the LHC [26, 111].
The effective coupling for the heavy bosons can be extracted from $q_i\bar{q}_j$ cross section in terms of the Weinberg angle, θ, the vector and axial couplings of the Z^0 boson to the fermions, V_f and A_f and the CKM matrix element, V_{qq}, as:

$$Z : \alpha_{eff} = \frac{\alpha_{em}}{4\sin^2\theta cos^2\theta}(V_f^2 + A_f^2) \tag{3.10}$$

$$W^\pm : \alpha_{eff} = \frac{\alpha_{em}|V_{qq}|^2}{4\sin^2\theta} \tag{3.11}$$

In addition, in the massless limit at LO, the splitting functions for EW topologies are the same as the ones from QCD partons up to the coupling and different colour factors. For simplicity, an average over all the polarised states is considered and any longitudinal contribution is neglected in this stage. In this section, the term "particle densities" is used in place of parton densities to emphasize the inclusion of photon and heavy boson contributions.

3.3.1 QCD densities

The particle densities at the initial energy scale μ_0 are determined from a fit to deep inelastic scattering (DIS) data from HERA [112], following the methodology described in [113]. The fit is performed using UPDFEVOLV [114] within the XFITTER framework [115], where the different parton flavours are parametrised as

$$\begin{aligned}
xg(x) &= A_g x^{B_g}(1-x)^{C_g} - A_g' x^{B_g'}(1-x)^{C_g'},\\
xu_v(x) &= A_{u_v} x^{B_{u_v}}(1-x)^{C_{u_v}}(1+E_{u_v}x^2),\\
xd_v(x) &= A_{d_v} x^{B_{d_v}}(1-x)^{C_{d_v}},\\
x\bar{U}(x) &= A_{\bar{U}} x^{B_{\bar{U}}}(1-x)^{C_{\bar{U}}}(1+D_{\bar{U}}x),\\
x\bar{D}(x) &= A_{\bar{D}} x^{B_{\bar{D}}}(1-x)^{C_{\bar{D}}}.
\end{aligned} \tag{3.12}$$

To extend this approach beyond the QCD framework, a new fit is carried out to simultaneously include contributions from quarks, gluons, photons, and heavy electroweak bosons, which is further described in [93]. This generalisation enables a unified description of all particle densities within a consistent theoretical setup. The fit is performed in two regimes, as described in [113]:

- **Set 1:** employing $\alpha_S(q)$, where q denotes the propagating particle virtuality

- **Set 2:** employing $\alpha_S(q_t)$, with $q_t = q(1-z)$ corresponding to the transverse momentum of the emitted particle.

Uncertainties are evaluated following the procedure in [113]. Statistical and systematic uncertainties are derived using the Pumplin method [116], while model uncertainties are assessed by varying the initial scale, quark masses, and the minimum value of q_t in the Set 2 configuration. The resulting fit yields a slightly higher χ^2 compared to the QCD-only case, reflecting the increased complexity introduced by the additional particle degrees of freedom.

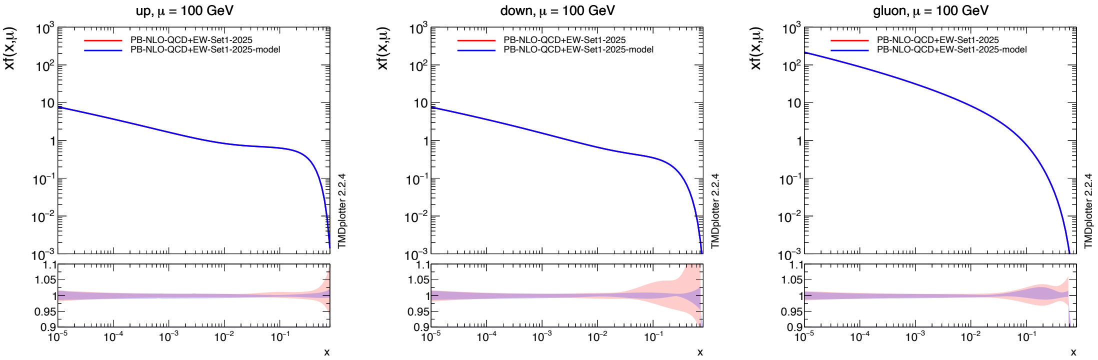


Figure 3.4: The collinear density for PB Set1 at $\mu = 100$ GeV as a function of x for the up quark (left), down quark (centre) and gluon (right).

The collinear quark and gluon densities obtained from the fit are shown in Fig. 3.4 at $\mu = 100$ GeV. Experimental and model uncertainties are displayed separately in red and blue, respectively. The uncertainty band is largest at high x, where the dynamics are dominated by quark contributions and by modelling uncertainties inherent to the QCD description. These features reflect the theoretical framework adopted for the evolution of the particle densities.

In this framework, the QCD contribution is included up to next-to-leading order (NLO), while the QED component is treated at leading order (LO), consistent with the hierarchy $\alpha_{\text{em}} \sim \alpha_S^2$. This choice ensures that the deep inelastic scattering (DIS) structure functions remain consistent with standard QCD results. Throughout the evolution, a fixed value of $\alpha_{\text{em}} = 1/137$ is used for simplicity. Furthermore, the intrinsic photon component and the interference between the photon and the Z^0 boson are neglected, as discussed previously, yielding a cleaner and theoretically consistent setup for the particle density evolution.

3.3.2 Photon densities

The collinear density for the photon is shown in Fig. 3.5 at $\mu = 100$ GeV. Here, the experimental and model uncertainties are shown separately in red and blue correspondingly, and both the total and inelastic components from the MSHT20 prediction [117] is included for comparison, illustrating that the two calculations yield very similar results, particularly in the low-x region.

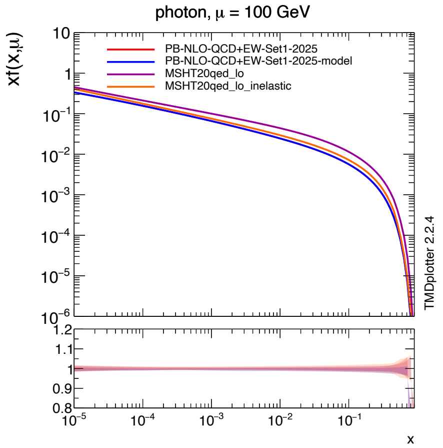


Figure 3.5: The collinear photon density for PB Set1 at $\mu = 100$ GeV as a function of x.

3.3.3 Heavy Boson densities

To obtain physically meaningful heavy boson densities, their masses must be treated appropriately. Three different approaches are discussed:

- `MASSCUTSCHEME=0`: All partons are considered massless in the Zero-Mass-Variable-Flavour-Number-Scheme (ZMVFN), used in QCD. As a result, heavy-flavours only arise once the energy scale surpasses the mass threshold. Similarly the heavy boson mass is included as a mass boundary according to the EWA [118]:

$$\alpha_{\text{eff}} \equiv \alpha_{\text{eff}} \cdot \Theta\left(q^2 - M_V^2\right) \tag{3.13}$$

- `MASSCUTSCHEME=1`: This approach is inspired by lepton-proton collisions, where the charged current cross sections lead to a non-zero contribution when $q^2 \ll M_V$. This motivates the introduction of a suppression factor related to the boson mass M_V [119]:

$$\alpha_{\text{eff}} \equiv \alpha_{\text{eff}} \cdot \left[\frac{q^2}{q^2 + M_V^2}\right]^2 \tag{3.14}$$

  This is the default approach and the one implemented in the new density fits.

- `MASSCUTSCHEME=2`: This approach is based on angular ordering [120]. Here the emitted particle mass limits the range of the z-integration by $z_M = 1 - M_V/q$, therefore the boson mass in treated via this limit:

$$\alpha_{\text{eff}} \equiv \alpha_{\text{eff}} \cdot \Theta\left(q^2(1-z)^2 - M_V^2\right) \quad \text{and} \quad z_M = 1 - \frac{M_V}{q} \tag{3.15}$$

Figure 3.6 presents the collinear density for the W^+ boson for different μ as a function of x for the previous `MASSCUTSCHEMES`. Only `MASSCUTSCHEME=1` results into a non-vanishing density across all μ scales, opposite to `MASSCUTSCHEME=0` which only arises when $q^2 \gg M_V$.
Additionally, Fig. 3.7 shows the TMD density for the W^+ boson for different μ as a function of x for the previous `MASSCUTSCHEMES`. Here the boson density obtained with `MASSCUTSCHEME=1` is physically plausible, in contrast to `MASSCUTSCHEME=2`, which displays a sharp limitation in k_T around the heavy boson mass. As a result,

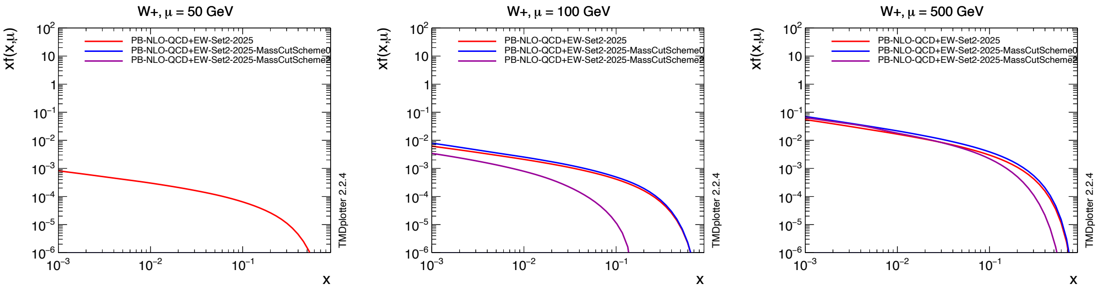


Figure 3.6: The collinear density for the W^+ boson for PB-NLO-QCD+EW-2025 Set2 at $\mu = 50, 100, 500$ GeV as a function of x for the different MASSCUTSCHEMES.

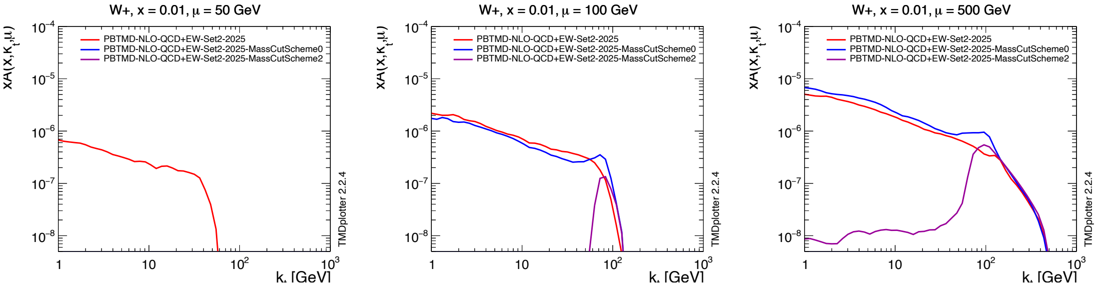


Figure 3.7: The TMD density for the W^+ boson for PB-NLO-QCD+EW-2025 Set2 at $\mu = 50, 100, 500$ GeV as a function of x for the different MASSCUTSCHEMES.

the option `MASSCUTSCHEME=1` is the only scheme capable of producing densities valid across the entire μ range and is therefore used as the default implementation.

As a result, the collinear density for the heavy bosons with `MASSCUTSCHEME=1` is shown in Fig. 3.8:

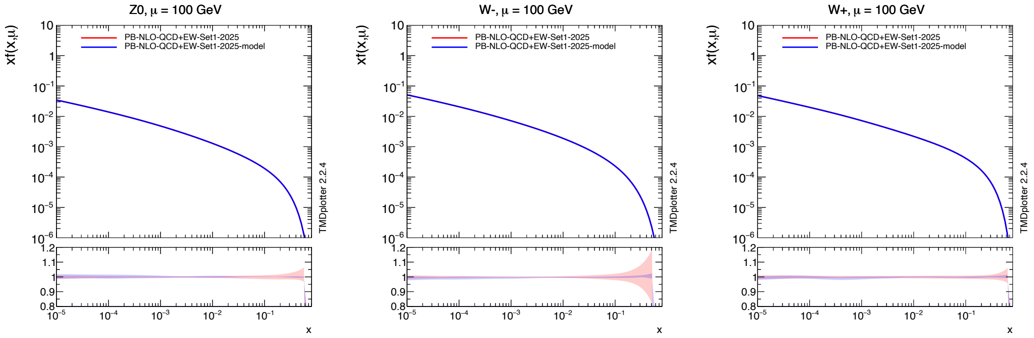


Figure 3.8: The collinear density for PB Set1 at $\mu = 100$ GeV as a function of x for the Z^0 boson (left), W^- boson (centre) and W^+ boson (right).

3.4 Validation

The collinear photon and heavy boson densities are fundamentally connected to the DIS cross section, as discussed in [98]. In that work, the photon PDF is obtained based on the proton structure functions F_L and F_2, which are extracted from measurements in electron–proton scattering experiments. In contrast, the current

study employs the DIS measurements to validate the previously obtained collinear densities.
The main objective of this section is to calculate DIS cross sections considering the HERA experimental framework, incorporating photon and heavy-boson densities. These calculations are compared with measured cross sections at HERA [112] to assess how accurately the densities describe the corresponding particle behaviour and their applicability in experimental conditions.

3.4.1 Photon and Z^0 boson

The simplest scenario is the photon case, where the differential DIS cross section can be expressed in terms of the DIS variables: Q^2, x and y for $e^+p \to e^+X$ and re-written as a function of the photon density, f_γ:

$$\frac{d\sigma_{\mathrm{NC}}}{dQ^2} = \frac{4\pi^2\,\alpha_{\mathrm{em}}}{Q^2}\,\frac{d\big(xf_\gamma(x_{\mathrm{min}},Q^2)\big)}{dQ^2}\,. \tag{3.16}$$

Here, the pure photon exchange is considered and therefore $xF_3 = 0$. Similarly, for the Z^0 boson exchange, the DIS cross section can be re-expressed in terms of the Z^0 boson density, f_Z:

$$\frac{d\sigma_{NC}}{dQ^2} = \frac{\pi^2\alpha_{\mathrm{em}}}{Q^2}\,\frac{V_e^2 + A_e^2}{\sin^2\theta_W\cos^2\theta_W}\,\frac{dx\,f_Z(x_{\mathrm{min}},Q^2)}{dQ^2} \tag{3.17}$$

Figure 3.9 displays the comparison between the measured neutral current (NC) cross section with zero polarization from [112] at HERA and the calculation using the photon and Z^0 boson densities.

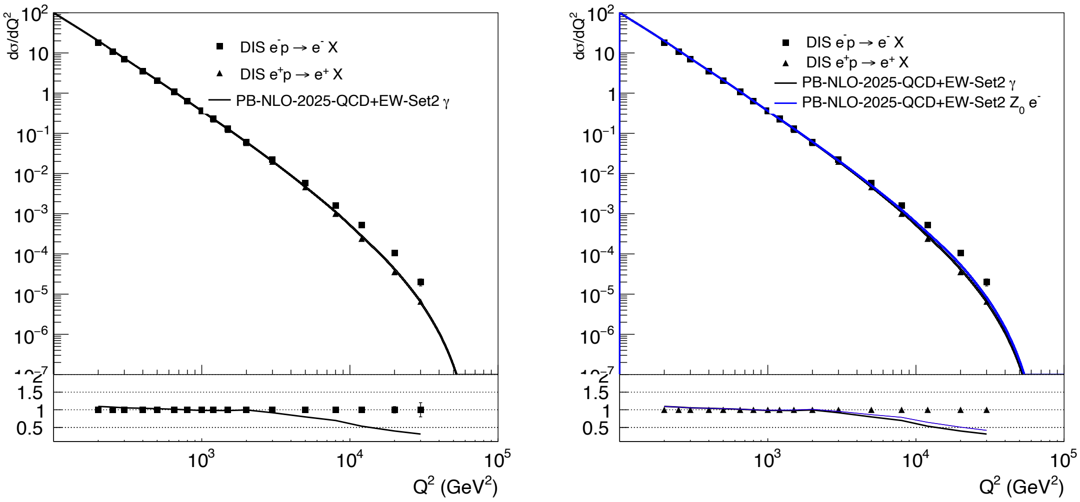


Figure 3.9: The measured DIS NC cross section [112] compared to the calculation using the collinear photon density (left) and the sum of the photon and the Z^0 boson contributions (right).

At low Q^2, where photon exchange provides the dominant contribution, the agreement between the measurements and the predicted parton densities is excellent. At higher Q^2, however, the contribution from Z^0-boson exchange becomes increasingly significant, leading to a deterioration in the agreement. This deviation arises from the omission of the interference term between photon and Z^0-boson exchange.

3.4.2 $W^\pm$ boson

In contrast to the photon and Z^0 boson cases, when comparing the charged-current (CC) deep inelastic scattering (DIS) cross section to the $W^\pm$ densities, the formulations derived in Section 3.3.3 are not applicable. This inadequacy arises from their lack of an inclusive summation over all initial-state weak isospin charges, in contrast to quark states. Consequently, all helicity configurations must be explicitly accounted for, as discussed in Ref. [118]. Under these conditions, the CC DIS cross section for $e^+p \to \nu X$ can be expressed in terms of the helicity-aware W^+ density, $f_{W^+}^{DIS}$:

$$\frac{d\sigma_{CC}}{dQ^2} = \frac{\pi^2\alpha_{em}}{Q^2}\frac{1}{\sin^2(\theta_W)}\frac{dx f_{W^+}^{\text{DIS}}(x_{\text{min}}, Q^2)}{dQ^2} \,. \tag{3.18}$$

And similarly for the helicity-aware W^- density, $f_{W^-}^{DIS}$:

$$\frac{d\sigma_{CC}}{dQ^2} = \frac{\pi^2\alpha_{em}}{Q^2}\frac{1}{\sin^2(\theta_W)}\frac{dx f_{W^-}^{\text{DIS}}(x_{\text{min}}, Q^2)}{dQ^2} \,. \tag{3.19}$$

Figure 3.10 shows the comparison between the measured charged-current (CC) cross section from Ref. [112] at HERA and the calculation using the $W^\pm$ boson densities. Overall, a good level of agreement is observed across the entire Q^2 range, with deviations below 20% for the W^- boson predictions.

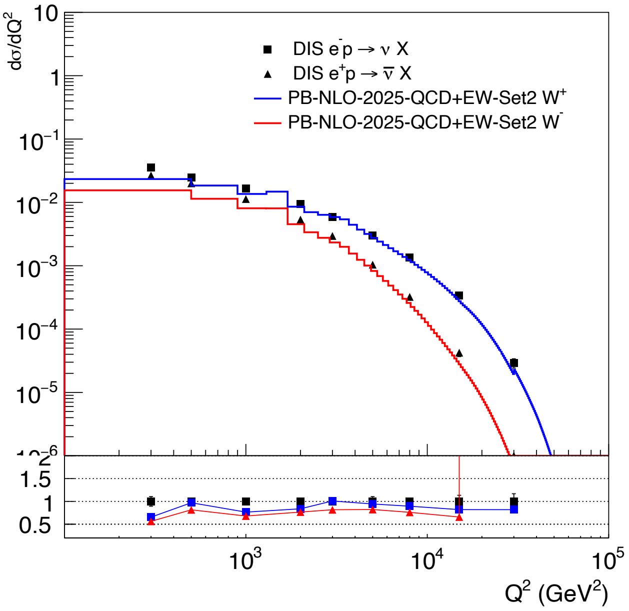


Figure 3.10: The measured DIS CC cross section [112] compared to the calculation using the collinear $W^\pm$ boson contributions.

In both the neutral and charged current regimes, the measured DIS cross sections at HERA show consistency with the predictions based on the newly derived bo-

son densities, demonstrating the robustness of the extended DGLAP framework in determining particle, and specifically heavy boson, densities.
These results pave the way for detailed investigations of heavy boson dynamics at high energies and motivate further exploration in future collider programs.

3.5 Application of heavy boson TMDs to Vector Boson Fusion production of Z boson

This section examines the Vector Boson Fusion (VBF) process through the subprocess $W^+W^- \to Z$, comparing dedicated VBF matrix-element calculations with predictions based on heavy-boson TMDs to assess their consistency and the constraints imposed by the mass schemes discussed in Section 3.3.3. Furthermore, the inclusion of heavy bosons in parton shower simulations is discussed (paper in preparation).
A wide range of future collider proposals aim to significantly increase the centre-of-mass energies achievable in the coming decades. In this regime, heavy bosons may act as initial-state particles, rendering the previously derived heavy-boson densities crucial. At such high energies, heavy bosons can interact among themselves or with QCD partons, and, following the factorization theorems introduced in Section 1.2.1, these interactions can be described in terms of a heavy-boson density and a hard scattering process.
An illustrative example to test heavy-boson densities is the production of a Z^0 boson via the triple-gauge vertex, where two $W^\pm$ bosons fuse to produce a Z^0. This process is particularly relevant as it probes the transverse momentum distribution of heavy bosons within the proton, as schematically depicted in Fig. 3.11. In the two partonic subprocesses shown in Fig. 3.11, the Z^0 boson is produced on-shell due to a hadronic collision. Considering them as $2 \to 1$ processes, in the $q\bar{q} \to Z^0$ process, the partonic cross section reads:

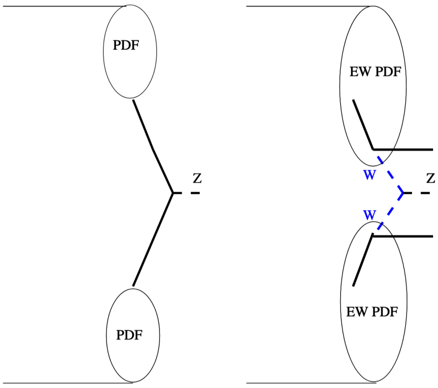


Figure 3.11: Diagrams representing Z^0 boson production induced by QCD (left) and induced by $W^\pm$ boson densities (right).

$$\hat{\sigma}_{q\bar{q}\to Z} = \frac{\pi^2 \alpha_{\rm em}}{3\, sin^2\theta_W cos^2\theta_W} \left[(T_3 - 2T_q sin^2\theta_W)^2 + T_3^2 \right] \delta(\hat{s} - m_Z^2) \;\; \text{for } q\bar{q} \to Z^0, \quad (3.20)$$

In the $W^+W^- \to Z^0$ process, applying the collinear and high-energy limit, only transverse W polarizations contribute, therefore summing over all of them, the partonic spin-summed cross section becomes:

$$\hat{\sigma}_{W^+W^-\to Z} = 8\pi^2 \alpha_{\text{em}} \cot^2\theta_W\, \delta(\hat{s} - m_Z^2) \;\text{ for } W^+W^- \to Z^0, \tag{3.21}$$

where $T_3 = \pm 1/2$ and $T_q = Q$ (the electric charge). These cross sections are implemented as hard kernels and interfaced with transverse-momentum-dependent (TMD) densities within the framework of `CASCADE` [30], encompassing both QCD partons and heavy bosons, to obtain the corresponding hadronic cross sections and, consequently, physically meaningful observables.

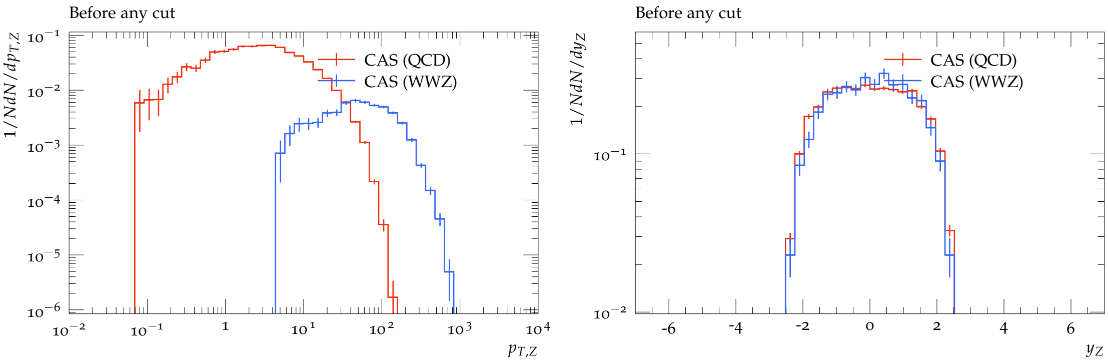


Figure 3.12: Normalized transverse momentum (left) and rapidity distribution (right) of Z bosons using the TMD densities previously obtained.

Figure 3.12 presents the normalized transverse-momentum and rapidity distributions of Z bosons obtained using the TMD densities shown in Fig. 3.7. The $q\bar{q} \to Z^0$ process is depicted in red, while the $W^+W^- \to Z^0$ production is shown in blue. These distributions are particularly informative, as the Z-boson transverse momentum directly probes the quark TMD densities in the $q\bar{q} \to Z^0$ channel and the $W^\pm$ densities in the $W^+W^- \to Z^0$ case. This is evidenced by the distinct shapes of the two production modes in the Z boson transverse momentum, reflecting the treatment of the $W^\pm$ boson mass and the evolution of the corresponding densities. In contrast, the Z boson rapidity distributions exhibit strong similarity, indicating that comparable momentum fractions are probed in both cases, primarily determined by the Z-boson mass.

Additionally, a complete calculation of the $q\bar{q} \to W^+W^- \to Zq\bar{q}$ process has been performed using `MadGraph_aMC@NLO` [121] and `Pythia` [69]. Jets are reconstructed with the anti-k_T algorithm with $R = 0.4$, excluding parton-shower, hadronisation, and primordial-k_T effects, and are analysed using a `Rivet` [71] routine. This complete calculation is compared with the approach previously described for the $W^+W^- \to Z^0$ production, employing the `MASSCUTSCHEME=1` configuration.

Figure 3.13 reveals a notable discrepancy, particularly in the jet p_T distribution, between the result obtained with `MASSCUTSCHEME=1` (blue) and the full matrix-element calculation performed with `MadGraph_aMC@NLO + Pythia` (red). This distinction is essential, as the jets originate from the TMD parton shower implemented in `CASCADE`, as described in [30]. The fluctuations at low Z-boson p_T instead reflect the limited statistical precision in that region.

This discrepancy is indeed expected, since it confirms it is not applicable introducing a suppression factor based on the evolution scale q in the $W^+W^- \to Z^0$ process,

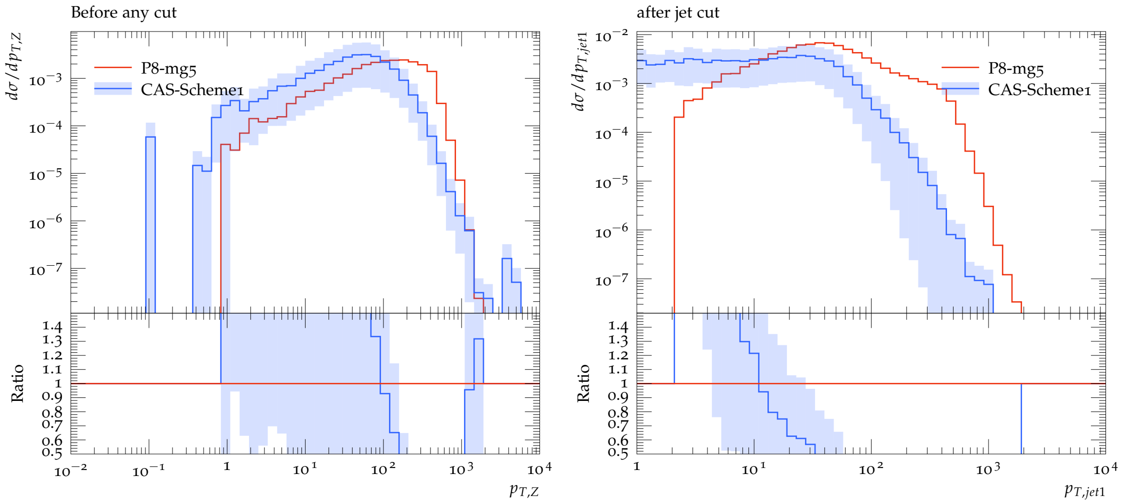


Figure 3.13: Z-boson p_T distribution (left) and jet p_T distribution (right) with `MASSCUTSCHEME=1` compared to the prediction from `MadGraph_aMC@NLO + Pythia`. The band corresponds to the variation obtained by changing the factorization scale by a factor of four.

where the physically relevant scale is the jet transverse momentum p_T, which directly enters the t-channel propagator of the VBF matrix element. As discussed in Eq. 2.16 in Chapter 2, in a VBF process, the matrix element contains propagators like $\frac{1}{t_i - M_W^2}$ for each $W^{\pm}$ boson, which is related for small t_i (Mandelstam variable) to $\frac{1}{p_T^2 + M_W^2}$ giving a factor $\frac{1}{(p_T^2+M_W^2)^2}$ at the squared matrix element level for each of the propagators.

In order to improve the description of the $W^{\pm}$-boson behaviour, a modified `MASSCUTSCHEME` where the suppression depends explicitly on the transverse momentum is implemented. This leads to a significant improvement in the description,

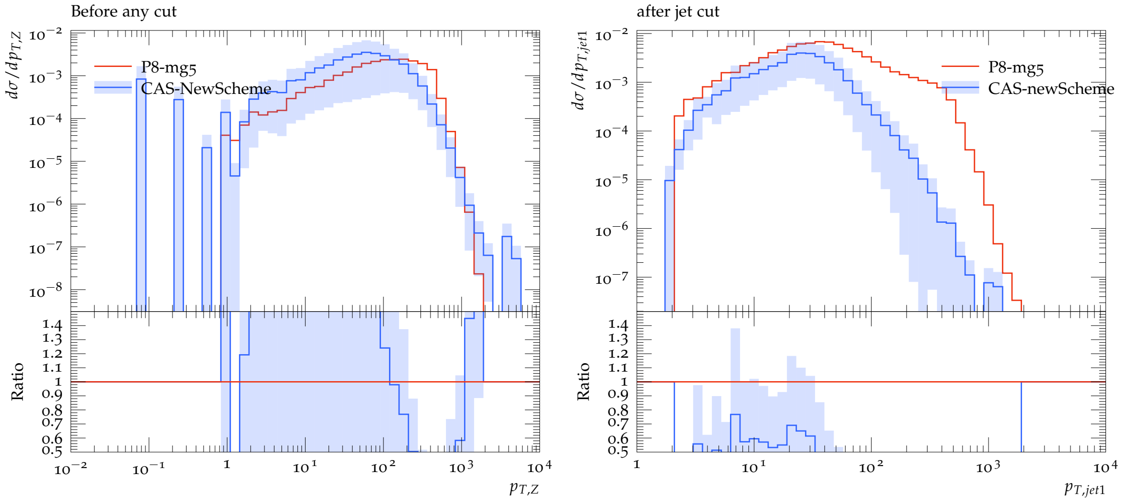


Figure 3.14: Z-boson p_T distribution (left) and jet p_T distribution (right) with a new `MASSCUTSCHEME` compared to the prediction from `MadGraph_aMC@NLO + Pythia`. The band corresponds to the variation obtained by changing the factorization scale by a factor of four.

particularly in the jet p_T distribution, where the expected decrease now appears at low p_T, and the Z-boson p_T peak exhibits a slight shift, as shown in Fig. 3.14. This new scheme introduces the t^2 propagator $\frac{p_T^4}{(p_T^2+M_W^2)^2}$ both in the TMD density and in the initial state parton shower.

Although this approach represents a substantial improvement, it remains an effective approximation. Consequently, k_T-dependent VBF matrix elements are explicitly required to accurately describe the t-channel $W^\pm$ propagator, in contrast to the DIS charged-current cross sections discussed in Section 3.4, where the $W^\pm$ boson did not play such a critical role.

These studies represent a paradigm shift in the treatment of parton densities, redefining them as particle densities through the inclusion of heavy bosons in the evolution. This framework provides the theoretical basis for exploring processes involving heavy bosons at future colliders, as exemplified by the discussion on electroweak TMDs and VBF Z production. It also lays the groundwork for developing fully k_T-dependent VBF Z matrix elements, enabling a more consistent and comprehensive description of electroweak-initiated processes in the high-energy regime.

Chapter 4

The LHC and the CMS Experiment

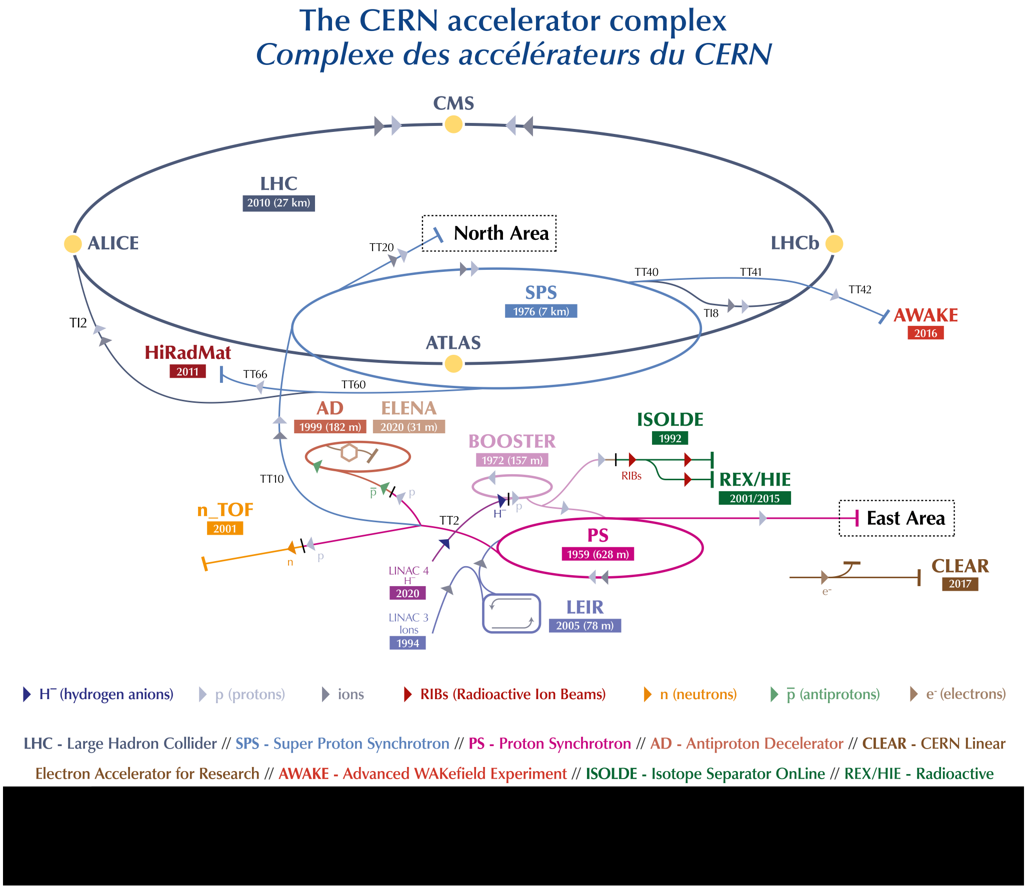
The proton–proton collision data used in this measurement were collected at the Large Hadron Collider (LHC) at the Conseil Européen pour la Recherche Nucléaire (CERN) in Switzerland. The dataset corresponds to the complete Run 2 period, recorded by the Compact Muon Solenoid (CMS) experiment between 2016 and 2018. This chapter introduces the experimental setup. Section 4.1 provides an overview of the LHC, where protons are accelerated and brought to collide at dedicated interaction points (IPs). One of these IPs houses the CMS detector, a multipurpose apparatus described in Section 4.2. The CMS coordinate system is defined in Section 4.2.1, followed by a detailed description of the detector subsystems in Sections 4.2.2 to 4.2.5. Section 4.3 presents the reconstruction of physics objects. Section 4.3.1 describes the detector simulation framework, while Section 4.3.2 address the trigger and data acquisition systems.

4.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a circular hadron collider with a circumference of 27 km, located at CERN on the border between Switzerland and France. It is designed to deliver proton–proton collisions at center-of-mass energies of up to $\sqrt{s} = 13$ TeV in Run 2, as well as heavy-ion collisions corresponding to about 5 TeV per nucleon [122]. In this thesis, the focus is on proton-proton collisions.

To achieve such high energies, protons are accelerated in stages through a sequence of accelerators, with the LHC constituting the final element of this injection chain, illustrated in Figure 4.1. The process begins at the linear accelerator Linac4, where negative hydrogen ions (H^-) are extracted from a source and accelerated to 160 MeV by means of radiofrequency (RF) cavities [123]. These ions are then stripped of their electrons, leaving protons that are subsequently injected into the Proton Synchrotron (PS), which accelerates them to 26 GeV. The protons are then transferred to the Super Proton Synchrotron (SPS), where their energy is increased to 450 GeV. At this stage, they are injected into the LHC, with beams circulating in opposite directions and further accelerated to collision energies of 13 TeV during Run 2.

The LHC proton beams consisted of up to 2556 bunches in Run 2, each containing approximately 1.15×10^{11} protons. The bunch structure is directly related to the machine's luminosity, which quantifies the collision rate of the accelerator. Luminosity is defined as the number of interactions per unit time per unit cross section. For a given physics process, the event production rate $\frac{dN_0}{dt}$ is proportional to the



[redacted]

Figure 4.1: CERN accelerator complex, from Ref [124].

process cross section σ and the instantaneous luminosity $\mathcal{L}(t)$ of the collider

$$\frac{dN_0}{dt} = \sigma \cdot \mathcal{L}(t) \ , \tag{4.1}$$

where the cross section depends on the physics process and can be calculated by applying the quantum field theory rules. Additionally, in order to measure it experimentally, it is needed to know the number of events N and the integrated luminosity $L = \int \mathcal{L} dt$ (see Fig. 4.2). The instantaneous luminosity can be known from the experimental collider parameters. As a result, for an accelerator with two beams with N_1 and N_2 number of protons in each, distributed in the transverse direction to the beam according to the 2D functions $\rho_1(x,y)$ and $\rho_2(x,y)$ and a revolution frequency ν, the instantaneous luminosity of a single interacting bunch pair can be expressed as

$$\mathcal{L}_b = \nu N_1 N_2 \iint \rho_1(x,y)\rho_2(x,y)dxdy = \nu \frac{N_1 N_2}{A_{eff}} \ , \tag{4.2}$$

where A_{eff} is the effective overlap area between the two beams. Assuming that both beams can be described by independent distributions in x and y, the previous expression can be rewritten as

$$\mathcal{L}_b \approx \nu N_1 N_2 \int f_1(x) f_2(x) dx \int g_1(y) g_2(y) dy = \nu \frac{N_1 N_2}{2\pi \sum_x \sum_y} \ , \tag{4.3}$$

where the $\sum_x$ is the effective width and $\sum_y$ is the effective height of the beam overlapping region. Finally, if the two beams are considered *ideal*, they can be considered to have equal Gaussian proton distributions, which would simplify the above expression even further as

$$\mathcal{L}_b \approx \nu \cdot \frac{N_1 N_2}{4\pi\sigma_x\sigma_y}\,, \tag{4.4}$$

where σ_x and σ_y are the root mean squared (RMS) widths of the Gaussian beams in x and y. Nevertheless, real colliders are far from perfect and as a consequence both beam distributions in the x and y directions are correlated, the beam profiles are not strictly Gaussian.

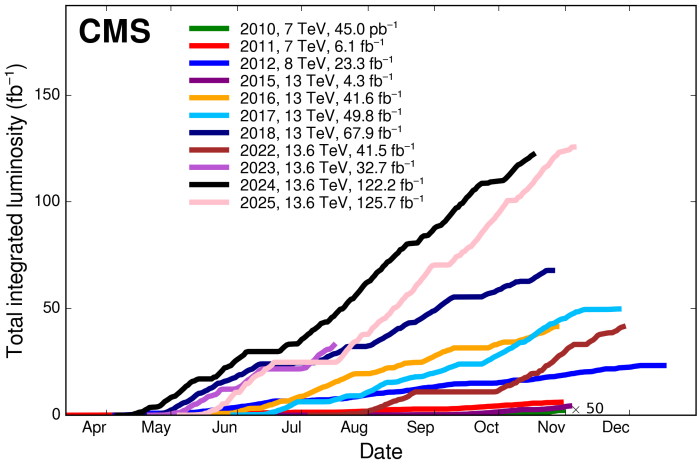


Figure 4.2: The delivered luminosity by the LHC in different years and different centre of mass energies from Ref. [125].

In addition to the previous considerations, the LHC is not a perfect circumference, since it is built by eight straight 545 m-long sections, named "insertions" connected by 2.45 km-long curved arcs called "sectors". This leads to the definition of an octant as the region which starts from the middle of an arc and ends in the middle of the following arc and thus spans a full insertion, as shown in figure 4.3. Each insertion consists of a long straight section plus two (one at each end) transition regions, called "dispersion suppressors".

The exact configuration of the straight sections of the LHC depends on their designated function, such as hosting beam collisions at interaction points (IPs), injection, beam dumping, or beam cleaning. In contrast, the arc sections follow a regular lattice structure composed of 23 arc cells per sector, each 106.9 m long and arranged in a FODO structure, consisting of main dipole magnets, quadrupole magnets, and additional higher-order multipole magnets, as illustrated in Figure 4.4.

Each sector contains 154 dipole magnets, amounting to a total of 1232 in total. These dipoles operate at the nominal magnetic field of 8.33 T, bending the proton beams along their circular trajectories. Beam focusing is achieved via quadrupole magnets, which act in pairs to maintain the tight bunch structure necessary to maximize the collision rate. The first magnet focuses the beam in the horizontal

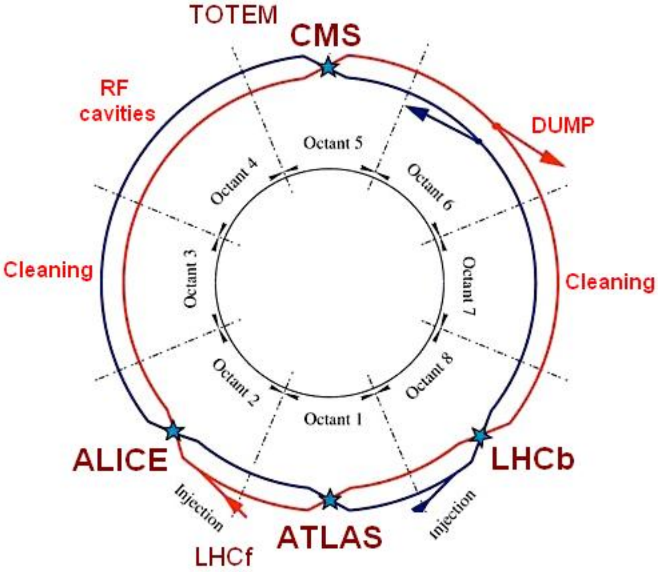


Figure 4.3: LHC layout, from Ref [122].

plane, while the second ensures focusing in the vertical plane. In total, the LHC employs 858 quadrupole magnets. Additional higher-order multipole magnets are also used to further stabilize the beams, enhance focusing, and compensate for non-linear effects such as long-range electromagnetic interactions between bunches and other beam dynamics perturbations.

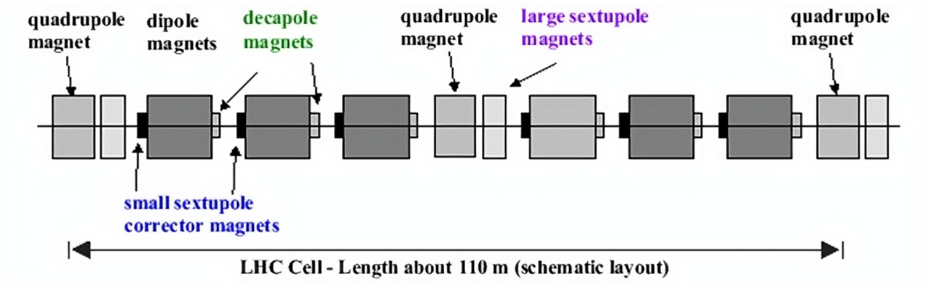


Figure 4.4: FODO structure: F focusses vertically and defocusses horizontally, D focusses horizontally and defocusses vertically and O is a space or deflection magnet, from Ref [126].

To accelerate the proton from 450 GeV up 13 TeV, 8 superconducting RF per beam are included in the insertions of the machine. These RF operate at a frequency of 400.8 MHz, which corresponds to the spacing between the bunches in the beam and their temperature is 4.5 K. There are four interactions points at the LHC where the beams collide, at a rate of every 25 ns, which is approximately 40 million collisions per second. As previously mentioned, each interaction point houses a different experiment: CMS, LHCb, ATLAS and ALICE. Both ATLAS (A Toroidal LHC ApparatuS) [127] and CMS are general purpose detectors, whereas ALICE (A Large Ion Collider Experiment) and LHCb (Large Hadron Collider beauty) have more specific goals. ALICE [128] is optimised to study the matter state in the early universe by colliding heavy ions, such as lead, which generate quark-gluon plasma. And the LHCb [129] studies the matter-antimatter asymmetry via bottom quarks.

4.2 The Compact Muon Solenoid

The CMS detector is a general-purpose experiment situated approximately 100 m underground at LHC interaction point 5 (IP5). It is designed to detect and precisely measure charged particles by employing an onion-like layered structure surrounding the interaction point [130]. From the innermost to the outermost layers, CMS comprises a silicon tracker to record charged particle trajectories, an electromagnetic calorimeter (ECAL) and a hadronic calorimeter (HCAL) to measure energy deposits from electrons or photons and hadrons respectively, and a system of muon chambers for muon tracking. A superconducting solenoid magnet generating a 3.8 T magnetic field is located between the HCAL and the muon chambers, bending charged particle paths. The outermost component is the iron return yoke, which confines the magnetic field and absorbs particles except for muons and neutrinos. The CMS detector measures 21.6 m in length, 14.6 m in diameter, and weighs approximately 14,500 tonnes.

4.2.1 The CMS coordinate system

In order to detect particles and properly measure their properties, a coordinate system of the detector is defined. As shown in Figure 4.5, the z axis corresponds to the direction along the beams. The x and y axes form the transverse plane to the beam direction. The x axis points to the centre of the LHC ring and the y axis is the vertical component. The polar coordinate system is composed by the θ and ϕ angles and the radius, r, in the xy plane. The ϕ angle measures the angle in the xy plane and θ in the zx plane.

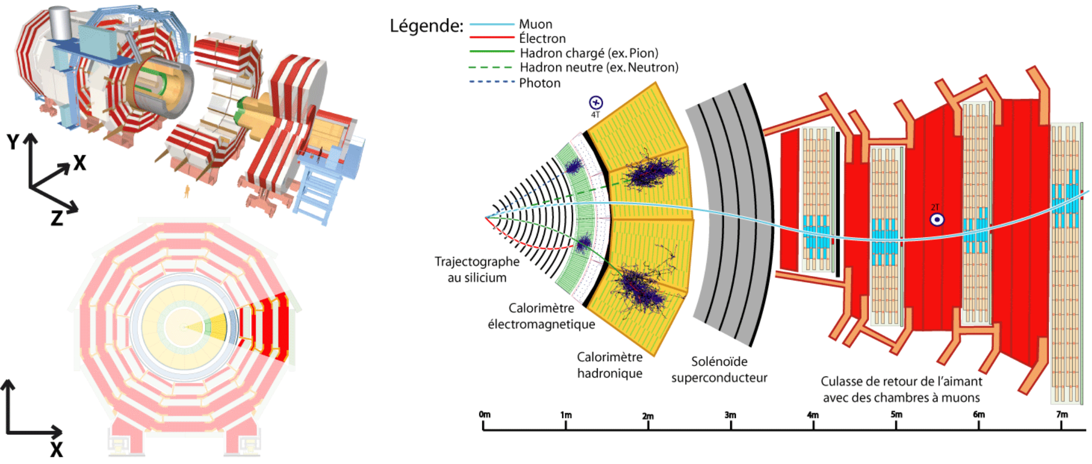


Figure 4.5: Schematic overview of the CMS detector: the left panels display the experiment's coordinate system, while the right illustrates the layered detector subsystems and typical trajectories and signatures of different particle species. from Ref. [131].

Due to the nature of proton-proton collisions, the energy of the constituents of the protons varies in each collision, therefore the measured observables need to be Lorentz invariant with respect the beam direction or z axis. For this, the transverse momentum of a particle is defined as

$$p_T = \sqrt{p_x^2 + p_y^2} \tag{4.5}$$

where p_x and p_y are the momentum in the x and y axes. The p_T sum of all the particles, both those interacting and non-interacting, should be equal to zero since the beams do not travel in the transverse plane. Additionally, another observable which is invariant under a boost in the z axis is the difference in rapidity between two particles Δy, where the rapidity is expressed as

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z} \tag{4.6}$$

This quantity can be interpreted as "how centrally a particle is produced". For massless particles or at high energy, the mass of the particle can be neglected, therefore, $E \approx |\vec{p}|$ and the rapidity converges to the pseudorapidity:

$$\eta = -\ln \tan \frac{\theta}{2} \tag{4.7}$$

4.2.2 The solenoid magnet

The solenoid magnet is one of the key elements of the CMS detector, since the magnetic field surrounding the CMS IP is used to measure both the electric charge and momenta of charged particles. These particles are bent due to a magnetic field as a consequence of the Lorentz force $\vec{F_L} = q(\vec{v} \times \vec{B})$. As a result, their curvature radius indicates the magnitude of their p_T. This solenoid magnet has a length of 12.5 m, a diameter of 6.3 m and weights 12000 tonnes, becoming the largest superconducting solenoid ever built. It is composed of four layers of niobium-titanium alloy, which are cooled by a cryostat system operating with liquid helium at 4.7 K, and provides a magnetic field of 3.8 T. It contains a steel return yoke in order to contain and return the magnetic field flux. This yoke is composed by five barrel wheels and six endcap disks, weighing in total 10000 tonnes. In addition to returning the magnetic field, the yoke behaves like a filter, allowing only weakly interacting particles to pass through. It also provides the structure for the muon detectors, opposite to the interior part of the solenoid, which confines both the tracker and calorimeter systems.

4.2.3 The tracker system

The tracker is the innermost detector of CMS. It records the charged particles path through the measurement of their curvature radius, since it is immersed in the magnetic field of the solenoid. As a result, from the curvature of the trajectories we can infer the momentum and charge of charged particles crossing the tracker system. The tracker is key for the efficient and precise primary (point of a pp interaction) and secondary (displaced vertex located outside the beam profile) vertex reconstruction in the collisions given its fine granularity. Due to its vicinity to the IP, the tracker withstands a very intense particle flux, consequently, its demands on granularity, recovery time and radiation hardness lead to the choice of silicon as its primary material. It consists on an innermost pixel detector and an outermost silicon strip detector as shown in Figure 4.6, where the pixel is the central area and the silicon strip is the rest.

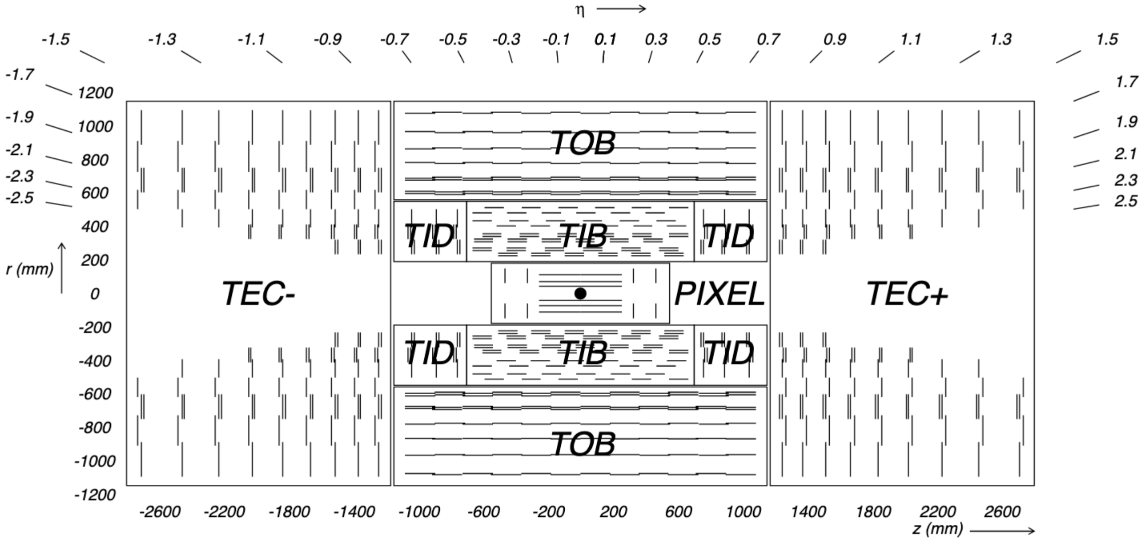


Figure 4.6: Cross section of the CMS tracking system in the zy plane from Ref. [130].

Pixel Tracker

The pixel detector is the inner part of the tracker and provides precise measurements of charged particle trajectories and collision vertices. When it was built (Phase-0), it consisted of three cylindrical barrel layers (BPIX) at radii of 4.4 cm, 7.3 cm and 10.2 cm, which cover the -270 mm to 270 mm region of the beam axis around the interaction point, and two forward/backward endcap disks (FPIX) at longitudinal positions of $z = \pm 34.5$ cm and $z = \pm 46.5$ cm and in a radius from 6 to 15 cm. Then, in 2016/2017 (Phase-1), it was upgraded to four barrel layers at radii of 2.9 cm, 6.8 cm, 10.9 cm and 16.0 cm and three endcap disks at $z = \pm 29.1$ cm, $z = \pm 39.6$ cm and $z = \pm 51.6$ cm to provide hit coverage up to an absolute value of $\eta = 2.5$, improve pattern recognition and track reconstruction, and add redundancy to cope with hit losses [132] as seen in Figure 4.7.

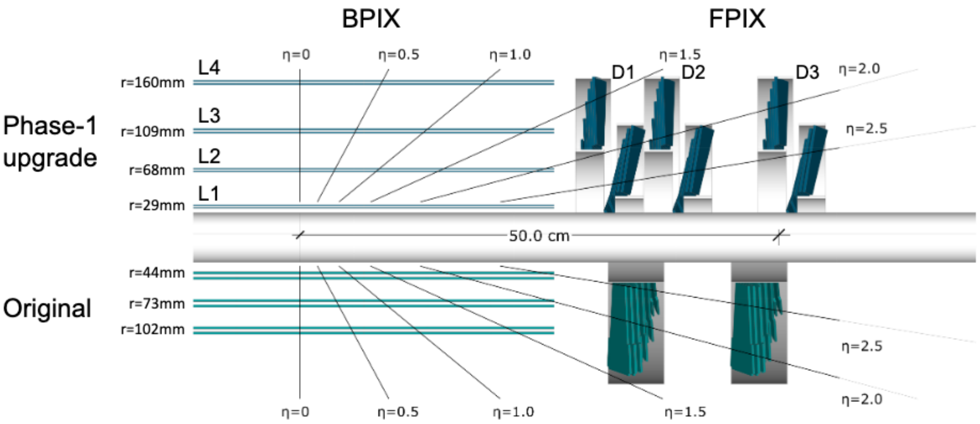


Figure 4.7: Layout of the CMS Phase-1 pixel detector compared to the original detector layout, in longitudinal view from Ref. [132].

The total silicon area of the CMS Phase-1 pixel detector is 1.9 m^2, while the total silicon area of the original pixel detector was 1.1 m^2. The current pixel detector consists on 1856 segmented silicon sensor modules, where 1184 modules are used in the barrel layers and 672 modules in the forward disks (FPIX). Each module consists of a sensor with 160 × 416 pixels with a standard size of 100 × 150 μm^2 and 285 μm thick. These pixels are connected to 16 readout chips (ROCs). In total there are 124 million readout channels which are responsible for transmitting the pulse height information produced by each hit recorded by the pixels. Finally, the resolution is measured separately for the two directions along the pixels, $r\phi$ and z in the BPIX detector and r and ϕ in the FPIX detector. The measured value depends on the track angle and on the radial position of the layer, leading to a resolution of 9.5 μm in the $r\phi$ direction and 22.2 μm in the z direction.

Strip Tracker
The Silicon Strip Tracker (SST) is the outer part of the tracker and is designed to further measure the momenta of the charged particles and to provide information for particle identification. It is divided in four parts: the inner barrel (TIB), the inner disk (TID), the outer barrel (TOB) and the outer end-caps (TEC), as seen in Figure 4.6. The TIB and TID are composed by four concentric layers and 3 disks on each side. The TOB and TEC consist of six concentric layers and nine disks on each side. It covers a total active area of about 200 m^2 and around 15 million silicon strips, each of them being 10 cm long and with a thickness of 320 μm. These strips are read out by electronics located at the end of each module, and the data are sent to the CMS data acquisition system for processing and analysis. Overall, the strip detector offers a 20-50 μm resolution in $r\phi$ and 500 μm in z, therefore their resolution is lower than the pixel sensors, but they cover a larger area, playing a complementary role.

4.2.4 The calorimeters

The calorimeters are sub-detectors whose main goal is to measure the energy of the particles involved in the collision. They are located inside the magnet structure and consist of an electromagnetic calorimeter (ECAL) and a hadronic calorimeter (HCAL).

Electromagnetic calorimeter
The aim of the electromagnetic calorimeter (ECAL) is to measure the energy of electrons and photons. It played a crucial role in the study of electroweak symmetry breaking and therefore, the discovery of the Higgs Boson through its decays to two photons or two Z bosons. This was achieved due to the ECAL excellent energy resolution. The ECAL is composed by a barrel structure (EB) and two endcaps (EC), all of them made of over eighty thousand lead tungstate ($PbWO_4$) crystals in total, weighting 93 tonnes as a whole.
The choice of ($PbWO_4$) was made based on its radiation hardness, its short radiation length ($X_0 = 0.89$ cm) and its small Molière radius of 2.19 cm. This radius reduces the effect of additional contributions from non-primary interactions to the energy measurement by reducing the area over which the energy is summed. As a result, 80% of all the light is emitted within the 25 ns between two collisions, consolidating ($PbWO_4$) as a fast scintillator.
The EB structure is divided in 36 supermodules, each with 1700 crystals. It covers the pseudorapidity range $|\eta| < 1.48$. The front face of the crystals is at a radius

of 1.29 m and each crystal has a square cross-section of $\approx 21.8 \times 21.8$ mm^2 and a length of 230 mm. The crystal cross-section reaches $\Delta\eta \times \Delta\phi = 0.0175 \times 0.0175$. The EE structure is composed by 4 "Dee" structures, each with 5382 crystals. It covers the pseudorapidity range 1.48$< |\eta| < 3$. Each crystal has a square cross-section of $\approx 24.7 \times 24.7$ mm^2 and a length of 220 mm. The crystal cross-section ranges from $\Delta\eta \times \Delta\phi = 0.0175 \times 0.0175$ to $\Delta\eta \times \Delta\phi = 0.05 \times 0.05$, as a result, the EE spatial resolution is too low. In order to fix this, in front of the EE, the preshower (ES) detector is placed, shown in dark pink in Figure 4.8. It contains lead converters followed by detector planes of silicon strips with a pitch of < 2 mm. The ES prevents low momentum photon pairs which are the decay products of pions from being detected as single photons.

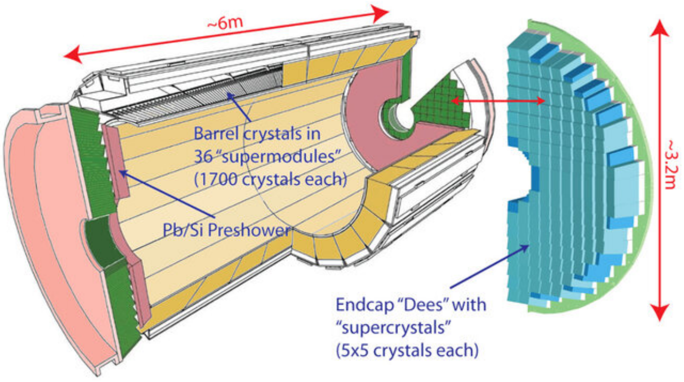


Figure 4.8: Schematic overview of the Electromagnetic Calorimeter (ECAL) of CMS from Ref. [133].

The energy resolution of the ECAL is given by:

$$\frac{\sigma}{E} = \sqrt{\left(\frac{2.8\%}{\sqrt{E/GeV}}\right)^2 + \left(\frac{12\%}{E/GeV}\right)^2 + (0.3\%)^2} \tag{4.8}$$

Where the first term models the statistical fluctuations in the shower production, the second term models the noise in electronics, and the last term accounts for the energy leakage.

Hadron calorimeter

The hadronic calorimeter (HCAL) measures the energy of hadrons using a sampling design of alternating dense absorber layers (brass or steel) and scintillators that emit light when traversed by particles. Its thick, hermetic structure captures extensive hadronic showers, which are larger and more penetrating than electromagnetic showers, ensuring accurate energy measurement crucial for jets and missing transverse energy reconstruction. The HCAL complements the tracker by measuring charged hadrons' energy and provides the only sensitivity to neutral hadrons.

The HCAL is located outside the ECAL and consists of four main parts: the hadron endcaps (HE), hadron barrel (HB), which is radially restricted due to the limited space between the ECAL and the magnet coil, an outer hadron calorimeter (HO) or "tail catcher" outside the solenoid complementing the HB, and the forward hadron calorimeters placed in the forward direction ($|\eta| < 3$) at 11.2 m from the interaction point, as sketched in Figure 4.9.

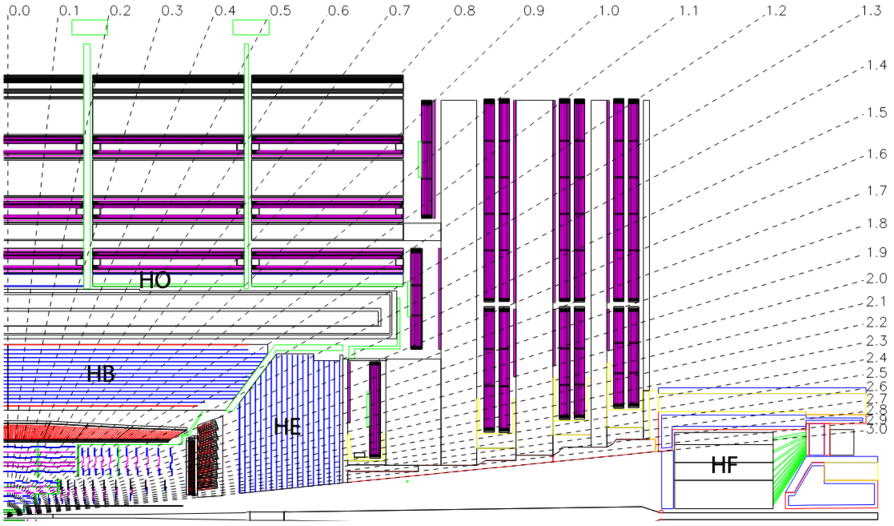


Figure 4.9: Schematic overview of the Hadronic Calorimeter (HCAL) of CMS from Ref. [130].

The HB is a sampling calorimeter covering the pseudorapidity range $|\eta| < 1.3$. It consists of 36 identical azimuthal wedges which form the two half-barrels (HB+ and HB–). It is constituted by plates of active material interlayed by an absorber. The HB and HE baseline active material is Kuraray SCSN81 plastic scintillator, due to its long-term stability and moderate radiation hardness, while the HB absorber consists of several steel and brass plates with varying thickness.

The HE consists of inner and an outer endcaps. The inner ones cover the pseudorapidity range $1.3 < |\eta| < 3$ (13.2 % of the solid angle), where about 34 % of the particles in the final state are produced. The HE absorber is C26000 cartridge brass based on its capacity to contain hadronic showers, good mechanical properties and reasonable cost.

The HO is placed outside of the solenoidal magnet to ensure an adequate sampling depth for $|\eta| < 1.3$, It employs the solenoid coil as an additional absorber, and it is used to identify late starting showers and to measure the shower energy deposited after HB. It is composed of scintillators interleaved with the steel absorber plates of the magnet return yoke seen in Figure 4.10. This configuration extends the total depth of the calorimeter system to a minimum of 11.8 λ except at the barrel-endcap boundary region.

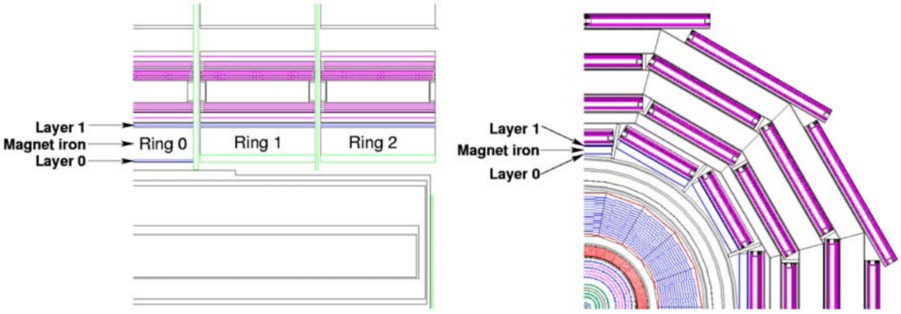


Figure 4.10: Schematic overview of the HO from Ref. [130].

Finally, the HF is located at the very forward region of the CMS detector. The outer HF covers the pseudorapidity range $2.9 < |\eta| < 4.5$, while the inner HF covers $4.5 < |\eta| < 5$. The HF measures the energy of hadrons produced at very small angles with respect the beam axis, hence, it is exposed to extremely high particle fluxes. Therefore, it requires an active material with a high radiation hardness. As a result, the HF is composed of steel absorber plates and quartz fibres (fused-silica core and polymer hard-cladding) placed behind plates of steel and borated polyethylene neutron shielding.

4.2.5 The muon chambers

The detection of muons with high precision and robustness was a core design priority for CMS, as implied by its name. Muons are present in many SM processes, but one of the main incentives was the predicted decay of the Standard Model Higgs boson into ZZ, which in turn decay into 4 muons. Additionally, the relatively long lifetime of muons (2.2 μs) allows them to traverse all layers of the CMS detector, losing only a small fraction of their energy, which motivates the placement of the muon system in the outermost region.
The muon chambers are the outermost subdetectors of CMS, located outside of the solenoidal magnet. The muon system has 3 functions: muon identification, momentum measurement, and triggering. It is composed of three types of detectors: drift tubes (DTs) in the central region $|\eta| < 1.2$, cathode strip chambers (CSCs) in the forward region $0.9 < |\eta| < 2.4$, and resistive plate chambers (RPCs) in the region $|\eta| > 1.9$ supporting the two previous detectors as seen in Figure 4.11.

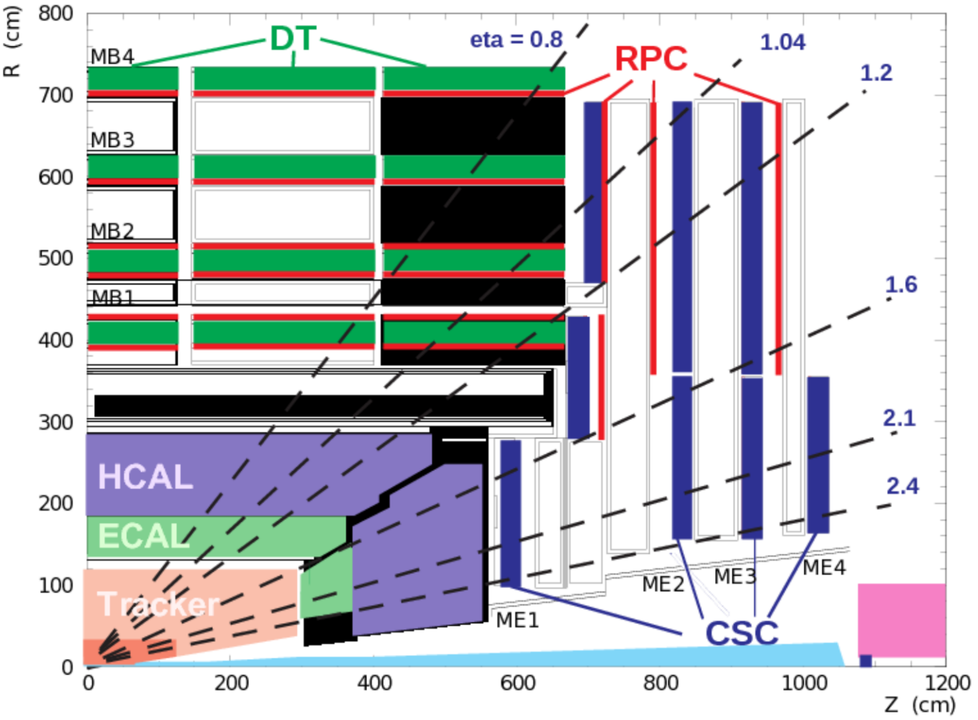


Figure 4.11: Quadrant of the Muon system, drift tubes (DT), cathode strip chambers (CSC) and resistive plate chambers (RPC). Image adapted from Ref. [134].

The DTs are located in the barrel region of CMS. These act as tracking detectors for the muon system due to the low expected rate and the relatively low strength of the local magnetic field in this area. There are 4 stations forming concentric cylinders around the beamline: the 3 inner cylinders have 60 drift chambers each and the outer cylinder has 70, leading to a total of 250 chambers in concentric cylinders around the beam line. These contain thin aluminium tubes filled with a gas mixture of argon (85%) and carbon dioxide (15%), with a central tungsten wire running through the centre of each tube.

The CSCs are placed in the endcap region, where the muon flux is higher and thus, the magnetic flux is stronger and non-uniform. As a result, a higher time and space resolution is needed. To achieve it, the muon system contains 468 CSCs, where each consists on 6 anode wire planes interleaved among 7 cathode strip panels, all contained within gas-tight chambers. Thus, the resulting signals from the anode wires and cathode strips are combined to construct 3D measurements of the muon positions.

Lastly, the RPCs play a complimentary role, which provide a large time resolution for muons passing through the detector both in the barrel and in the endcaps. They consist of two parallel plates made of a high-resistivity material, with a thin layer of gas in between. The RPCs can measure the time of an ionizing event with a resolution much shorter than the 25 ns interval between consecutive LHC bunch crossings (BX). Consequently, they can unambiguously identify the correct BX to which a muon track belongs, even in the presence of substantial background activity.

The working principle of the previous gas detectors is ionisation. Thus, when a muon passes through either a DT, CSC or RPC, it ionises the gas, creating a trail of electrons that drift towards the wire. Then, the resulting electrical signal is read out by electronics at the ends of the tube, determining the position of the muon and additionally, its momentum due the curvature of its trajectory.

4.3 Data reconstruction at CMS

In this section we discuss how data are recorded and particles are reconstructed from the information collected by the CMS detector.

4.3.1 Detector simulation

To compare the measured and simulated data, simulating the detector response is needed. At the CMS experiment, this is accomplished through the GEANT4 package [135], which simulates the geometry and the interactions of the particles with the different materials of the detector. The particles emerging from the MC simulation, called "generator-level" or "particle-level" particles, are propagated through the simulation, so GEANT4 will simulate the tracks and energy deposits that such final state will leave in the detector. This simulation is tuned to real collision data, to ensure its accuracy. Then, "detector-level" particles are obtained, which corresponds to the particle level final state that has undergone the simulation of the CMS detector. These two levels are crucial in the measurement for a first comparison between measured data and simulation and the latter correction to particle level.

4.3.2 Trigger system and data recording

The LHC produces a bunch crossing every 25 ns, leading to an amount of data too large to be stored. Thereby, a trigger system is needed to select which collisions have valuable data worth storing. This trigger system has two tiers: a Level 1 (L1) trigger and a High Level Trigger (HLT).

The L1 trigger uses custom hardware processors to select up to 100 kHz of the most interesting events with a latency of 4 μs [136]. The L1 trigger uses information of the muon system and both calorimeters to reconstruct leptons and jets. Additionally, it also computes the missing transverse energy. The selection is performed using a list of algorithms (known as “seeds”), which check events against predetermined criteria (known as “menu”). Any event that satisfies the conditions of at least one seed in the menu is accepted for further processing in the trigger chain into the HLT workflow.

The HLT hardware consists of a single processor farm composed of commodity computers, the event filter farm (EVF). These filter the events by implementing a version of the full event reconstruction software optimised for fast processing, reducing the event rate to approximately 1 kHz before data storage. This corresponds to a data stream of 90 ms per event or 1Gb/s which is permanently stored for later offline reconstruction.

Once finalised the triggering stage, accepted events are stored locally on disk and transferred to the CMS Tier-0 computing centre for offline processing, reconstruction as primary physics objects and permanent storage. Additionally, the data are classified in different streams or primary datasets according to the HLT decisions and stored in tape as well. These datasets are distributed to the Tier-1 centres, which are large centres in CMS collaborating countries, where data is further calibrated and filtered into files suitable for physics analyses, Analysis Object Data or AOD and their derivatives: miniAOD and nanoAOD. These files along with the MC samples are further distributed to Tier-2 centres, smaller institutions that enable their access through a grid system [137].

Chapter 5

Event and object reconstruction at the CMS Experiment

The current chapter explores in depth the physics object treatment and event simulation at the CMS Experiment. Section 5.1 focuses on leptonic reconstruction, Section 5.2 describes jet reconstruction, introducing the definition of jets and the algorithms employed for their clustering and identification. This section also details the jet calibration procedure, with particular emphasis on the derivation of jet energy correction factors, in which I was directly involved during Run 3, achieved through the computation of relative corrections as a function of pseudorapidity, η. Section 5.3 introduces several Monte Carlo event generators one of the main tools to simulate physical processes at the LHC, Section 5.4 explains the numerical treatment of PDF sets and finally Section 5.5 provides an overview of the main corrections applied in physics analyses.

5.1 Leptonic reconstruction

For each event, objects such as electrons, muons etc are identified and reconstructed to determine the interesting events for data analysis. This is performed based on the energy deposits in the different detector layers. In general, the Particle-Flow (PF) algorithm [138] collects the information of the different sub-detectors and combines them into PF candidates, which are the reconstructed objects: photons, muons, electrons and hadrons.

The tracks of charged particles are reconstructed based on the information from the silicon tracker. Since the tracker is immersed in the magnetic field, the trajectories of charged particles are bent, therefore both their charge and momentum can be inferred from their trajectories. This information is complemented by the observation from the muon chambers and ECAL crystal clusters for muons and electrons respectively.

In the case of muons, these are classified into “global muons” or “tracker muons”. “Global muons” arise from a combined fit based on information from both DTs and CSCs, resulting into “standalone muons”, which match tracks from the tracker. This approach infers the muon hit position and timing first, then it includes the information from the magnetic field and energy loss and finally obtains the muon p_T with a high resolution. Whereas, “tracker muons” are reconstructed inversely. First, the tracks from the inner tracker with p_T over a specific threshold are selected, and then, they are matched to hits in the muon chamber.

In the case of electrons, they produce hits in both the tracker and the ECAL, while photons deposit energy mainly in the ECAL and may produce tracks only if they convert into electron-positron pairs in the tracker. Neutral hadrons do not leave tracker signals and are measured primarily in the hadron calorimeter. Electrons emit bremsstrahlung photons in the tracker, which cause their ECAL showers to spread more compared to photons' typically more compact showers. The ECAL reconstructs energy by grouping high-energy crystals into seed clusters, then expanding these into superclusters (SC). Electron candidates are formed by matching SCs with tracks using the Gaussian Sum Filter algorithm [139, 140]. Photons are identified from SCs and further analysed to distinguish them from neutral hadrons and conversions. This combination of tracker and calorimeter information enables precise electron and photon reconstruction.

The tracks of charged particles are extrapolated to crossing points or vertices, which are defined as the exact locations where protons have collided. The sum of the transverse momentum of the tracks defines the energy of the vertex, and thus, the vertex with the largest energy deposit is considered the proton-proton collision point on the beam axis and referred to as primary vertex (PV). Other vertices in the event are then considered as coming from pileup collisions.

Finally, the hadron reconstruction is performed with the remaining tracks. A charged hadron is reconstructed when a cluster in the HCAL has a matching track in the tracker and a neutral hadron where there is none.

5.2 Jet reconstruction

When quarks and gluons are created in a partonic interaction, they emit further quarks and gluons in a narrow, concentrated stream. This stream eventually transforms into hadrons through a process called hadronisation. The original parton that initiates this sequence is known as the parent parton, and the resulting collimated stream of hadrons is referred to as a jet. When studying hadrons in general, and jets in particular, two main key points arise: multi proton-proton interaction treatment and jet reconstruction.

At the LHC, multiple proton–proton interactions typically occur within a single bunch crossing. Consequently, the detector records not only the products of the primary hard scatter but also contributions from additional interactions, collectively referred to as pileup (PU) [141]. Pileup significantly complicates event reconstruction and measurement, acting as a background that must be mitigated to ensure accurate analyses of the primary collision. During Run 2 (2016-2018), CMS operated with an average of roughly 30–35 additional interactions per bunch crossing, with even higher instantaneous values in the most challenging periods. Run 3, ongoing since 2022 and planned until July 2026, has experienced average pileup levels of roughly 50–60 interactions per crossing [125] (see Fig. 5.1).

To reduce pileup contamination in jets, the Charged Hadron Subtraction (CHS) technique is employed to remove charged particles not associated with the primary vertex [142]. In practice, tracks identified as originating from pileup vertices are excluded from jet clustering and from the computation of event observables, which already suppresses a large fraction of the pileup contribution carried by charged hadrons. The remaining neutral component, which cannot be directly associated

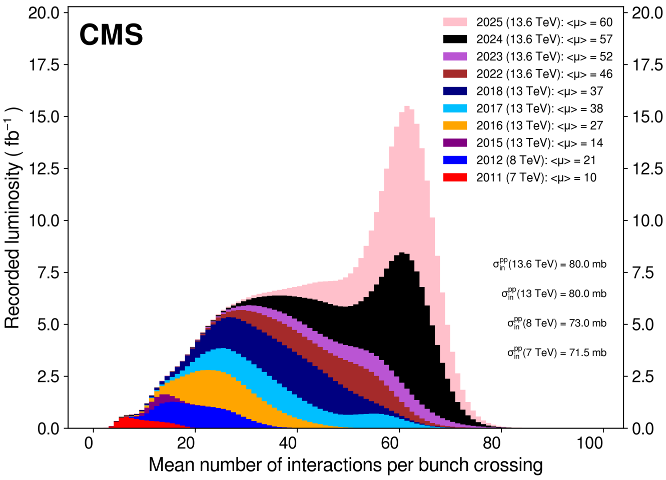


Figure 5.1: The recorded luminosity by the CMS experiment against the mean number of interactions per bunch crossing from Ref. [125].

to vertices, is then corrected statistically, for example using area-based corrections, but residual fluctuations still degrade jet energy and missing transverse momentum resolutions. This approach was the standard choice for Run 2 analyses.

For Run 3, however, the Pileup Per Particle Identification (PUPPI) algorithm [143] was recommended due to its superior performance in conditions with higher pileup densities. PUPPI assigns a weight to each reconstructed particle based on local shape and compatibility with the primary vertex, effectively down-weighting particles likely to originate from pileup, including neutrals. This per-particle weighting significantly reduces the rate of pure pileup jets and improves jet and missing transverse momentum resolutions in high-pileup environments, making it better suited for the conditions of Run 3 and the forthcoming HL-LHC.

In order to reconstruct jets, several jet algorithms are considered based on the requirement of infrared and collinear safety, which ensures their robustness. Some of them are:

- The Cambridge-Aachen (CA) algorithm: it uses angular separation and partial clustering, therefore it clearly favours the clustering of close-by particles, being advantageous for jet substructure studies [144]. Nevertheless, it makes difficult to connect the jet to the parent parton and it is prone to accept PU events. This algorithm considers the distance between the particle i and j as d_{ij}, and the distance between the particle i and the beam axis, as d_{iB} according to:

$$d_{ij} = \frac{\Delta R_{ij}}{R} \quad \text{with} \quad \Delta R_{ij} = \sqrt{\Delta\phi_{ij}^2 + \Delta\eta_{ij}^2} \quad \text{where} \quad d_{iB} = 1\ , \tag{5.1}$$

  where R is jet radius parameter.

- The k_T algorithm: it clusters particles by their relative transverse momentum, k_T, prioritizing soft radiation [145] and consequently rendering it vulnerable to accept PU events as well. In this case, d_{ij} and d_{iB} are stated as:

$$d_{ij} = \min\left(k_{T,i}^2, k_{T,j}^2\right)\frac{\Delta R_{ij}}{R} \quad \text{with} \quad \Delta R_{ij} = \sqrt{\Delta\phi_{ij}^2 + \Delta\eta_{ij}^2} \quad \text{where} \quad d_{iB} = k_{T,i}^2 \tag{5.2}$$

- The anti-k_T algorithm: it is based on the inverse relative transverse momentum of particles, $1/k_T$, which causes the algorithm to cluster hard particles first and subsequently merge softer particles around these. This approach preserves the jet boundaries and represents a significant improvement over traditional cone algorithms. In this scenario, d_{ij} and d_{iB} are stated as:

$$d_{ij} = min\left(\frac{1}{k_{T,i}^2}, \frac{1}{k_{T,j}^2}\right)\frac{\Delta R_{ij}}{R}, \quad \text{with} \quad \frac{\Delta R_{ij}}{R} = \sqrt{\Delta\phi_{ij}^2 + \Delta\eta_{ij}^2} \quad \text{where} \quad d_{iB} = \frac{1}{k_{T,i}^2}\,, \tag{5.3}$$

  the anti-k_T algorithm clusters two particles together when $d_{ij} < d_{iB}$ to form a particle k. Then, similarly, all particles with $d_{ik} < d_{iB}$ are used to reconstruct the jet till $d_{ik} > d_{iB}$, when the jet is defined and the reconstruction of additional jets starts. As a consequence, this algorithm favours the clustering of large transverse momentum particles and maintains the jet shape unaltered with respect to soft radiation. This results in the signature of the reconstructed jets being closer to the energy and transverse momentum of their corresponding parent, and thence, being suitable for theoretical comparisons.

The anti-k_T algorithm stands out in CMS due to its infrared and collinear safety and its ability to reduce pileup and underlying event contamination, limit spatial overlap in high-multiplicity events, and maintain calibration linearity over a wide p_T range. This stability simplifies jet identification, calibration, and reduces fluctuations in jet energy measurements, making anti-k_T the preferred choice for CMS and ATLAS jet reconstruction in the complex LHC environment [146–148].

Jets are reconstructed using the anti-k_T algorithm [146], where all particles reconstructed by the PF algorithm are used for clustering. As previously explained, this algorithm assumes that particles arising from the radiation of the same parent parton are close to each other. This feature is reinforced by selecting a jet radius of 0.4 in the $\eta-\phi$ plane, which is optimal for precise, stable, and efficient jet reconstruction in complex conditions such as the LHC, since it includes most of the partonic jet while properly resolving jets coming from different partons.

Although jets could in principle be clustered with a large radius (e.g. $R = 1$) followed by subclustering to recover smaller scales, this leads to significant pileup contamination, degraded energy resolution, and differs from direct small-R reconstruction. For inclusive CMS jet analyses, an empirically optimised radius $R = 0.4$ ("AK4"), balances capturing the bulk of parton shower energy and minimizing sensitivity to soft radiation. The high granularity of the CMS detector and its particle-flow reconstruction enable AK4 jets to achieve good energy resolution while mitigating soft radiation effects. Larger radii (e.g. $R = 0.8$, "AK8") are reserved for boosted object reconstruction, whereas $R = 0.4$ remains the default for inclusive jets, missing transverse momentum, and b-tagging.

5.2.1 Jet Energy Corrections (JEC)

The reconstructed jets arise from particles measured by the various sub-detectors of the CMS experiment. This makes them dependent on the technical features of each detector, leading to discrepancies between the measured jet energy and its real value. As a result, the measured jet energy is calibrated to its actual value through the jet energy corrections [149].

Jet energy corrections are derived through a factorisation scheme in the CMS experiment [150]. This streamlines the process allowing for the implementation of a correction factor for each source, as shown in Figure 5.2. These corrections, C_{JEC} are applied to the reconstructed jet momentum p_{raw} to restore its true value p_{true}

$$p_{true} = C_{JEC} \times p_{raw} \ . \tag{5.4}$$

In this section, $[X]$ denotes the bins of a given observable X, $\langle X \rangle$ the average of the observable X and (X) the dependence of a function on the observable X. The first correction to be derived is the pileup offset, also known as L1Offset. This correction is needed due to the contribution from multiple proton-proton interactions within a bunch or consecutive bunches, which affects the measured jet momentum. Even though, most of the pileup contribution has been removed by the CHS or PUPPI algorithms, it only affects the tracker covered region, and it does not remove neutral pileup contributions.

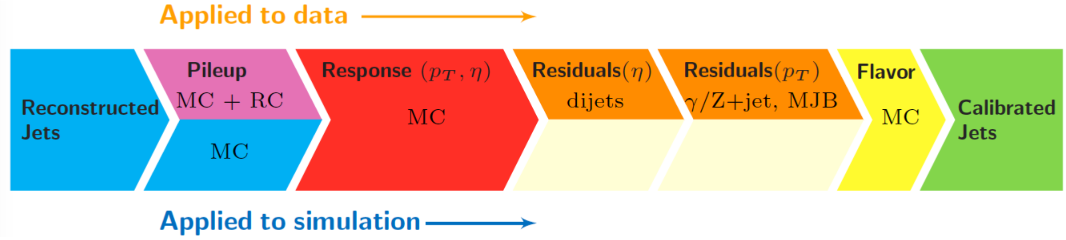


Figure 5.2: Ilustration of JEC derivation from Ref. [151]

In order to estimate the pileup offset energy, two simulation samples are needed, where the events in both arise from the same origin but only one sample includes the pileup events. Therefore, this correction is obtained in two steps. First, the offset of the transverse momentum of the jet at particle level $\langle p_{T,part,offset} \rangle$ is computed as a function of the energy density ρ in $\eta - \phi$ grids, the pseudorapidity, and the reconstructed jet momentum as

$$\langle p_{T,part,offset} \rangle \left(\langle \rho \rangle , [\eta] , p_{T,raw} \right) = \left\langle p^{NoPU}_{T,raw} - p^{WithPU}_{T,raw} \right\rangle [\mu, \eta, p_{T,part}] \ . \tag{5.5}$$

Where μ is a simulation parameter known as the number of pileup interactions in a bunch crossing. Obtaining this correction in following this approach enforces the selection of reconstructed jets with a matching jet at particle level, which are defined as those jet with $\Delta R(j_{part}, j_{reco}) < R/2$ where j_{part} and j_{reco} are the jets at particle level and detector level. Finally, the correction factor is extrapolated to ρ, η, $p_{T,Raw}$ and A, where A is the area of the jet:

$$C_{L1offset}(\rho, \eta, p_T, A) = 1 - \frac{p_{T,part,offset}}{p_{T,raw}} \ . \tag{5.6}$$

Then, the simulated response corrections are derived. The response of the detector simulation is given by

$$C_{Response}(\langle p_T \rangle , \eta) = \frac{\langle p_T \rangle}{p_{T,raw}} \left[p_{T,part}, \eta \right] \ , \tag{5.7}$$

where $p_T = C_{L1offset} \times p_{T,raw}$ and then, p_T is further rectified as $C_{Response} \times p_T$. This procedure results into fully calibrated jets in the simulation samples. In contrast

to the real detector, whose response degrades over time due to aging and radiation damage, the simulated samples do not account for such time dependence, as the detector response in GEANT4 is not time-sensitive. In order to account for this, residual corrections are derived for data using both simulated and measured data.
The residual corrections to the jet energy scale are derived employing the *tag-and-probe* technique. In this framework, the *probe jet* denotes the jet under investigation, for which the response is to be measured, while the *tag jet* (or reference object) is a well-calibrated jet used as a standard of comparison.
Jet response is assessed using two principal methods: the transverse momentum balance (p_T-balance) and the Missing transverse Momentum Projection Fraction (MPF) technique, defined respectively as

$$R_{\text{jet}}^{p_T\text{-balance}} = \frac{p_{T,\text{probe}}}{p_{T,\text{tag}}}, \quad R_{\text{jet}}^{\text{MPF}} = 1 + \frac{\vec{p}_{T,\text{miss}} \cdot \vec{p}_{T,\text{probe}}}{p_{T,\text{tag}}^2}. \tag{5.8}$$

Here $p_{T,\text{tag}}$ and $p_{T,\text{probe}}$ represent the transverse momenta of the tag and probe jets, respectively, and $\vec{p}_{T,\text{miss}}$ is the missing transverse momentum, defined as the negative vector sum of the transverse momenta of all reconstructed final-state particles in the event. The p_T-balance method exclusively utilizes the momenta of the tag and probe jets, whereas the MPF method incorporates information from the entire event.
Both approaches are susceptible to contributions from additional radiation, which can distort the measured jet response. To mitigate this effect, a topology-dependent parameter α is introduced, which quantifies the relative level of extra radiation in the event. The intrinsic jet response is subsequently extracted by extrapolating the measured response to the limit $\alpha \to 0$, corresponding to the absence of additional radiation.

- **η-dependent corrections**

  These corrections are also known as Level 2 Residuals or L2Res. In order to obtain them, dijet events are selected due to their large cross sections in hadronic collisions, which allows for the measurement of the jet response as a function of η and p_T. The two leading jets constitute the reference and probe objects, where at least one of them must have $|\eta| < 1.3$ to be considered in the event. The jet with $|\eta| < 1.3$ will be defined as the reference object and the other as the probe. When the two jets have $|\eta| < 1.3$ the reference and probe objects are randomly assigned. In this topology, α is defined as

$$\alpha = \frac{p_{T,3rdjet}}{p_{T,ave}} \ , \tag{5.9}$$

  where $p_{T,3rdjet}$ is the most energetic jet that does not originate from the event topology and $p_{T,ave} = \frac{1}{2}(p_{T,tag} + p_{T,probe})$. This average in p_T arises as a consequence of a sub-optimal resolution in the reference jet transverse momentum due to migration bias. As a result, the jet response for this topology is defined as

$$R_{jet}^{p_T-balance} = \frac{1 + \langle A \rangle}{1 - \langle A \rangle} \quad \text{where} \quad A = \frac{p_{T,probe} - p_{T,tag}}{2p_{T,ave}} \tag{5.10}$$

$$R_{jet}^{MPF} = \frac{1 + \langle B \rangle}{1 - \langle B \rangle} \quad \text{where} \quad B = \frac{\vec{p}_T^{\,miss} \cdot (\vec{p}_{T,tag}/p_{T,tag})}{2p_{T,ave}} \ , \tag{5.11}$$

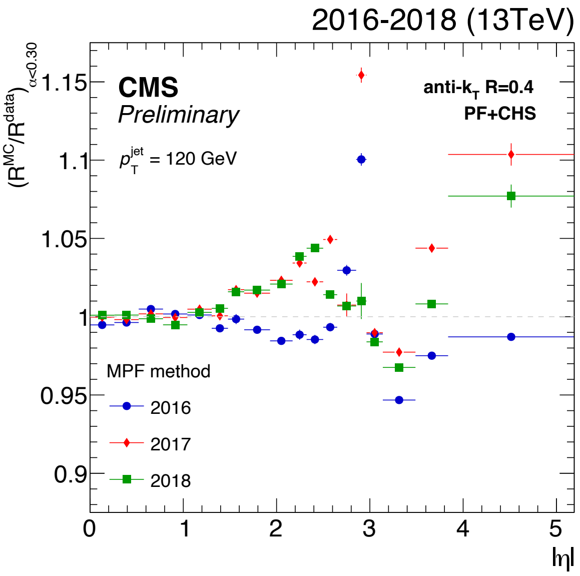


Figure 5.3: L2 Residual corrections for Full Run 2 from Ref. [152].

which converge to the expressions 5.8 when $\alpha \to 0$. As a result, the correction factor is $C_{L2Res} = R_{jet}^{Simulation}/R_{jet}^{Data}$.

Figure 5.3 shows the residual jet response corrections, normalized to the barrel response, as a function of jet $|\eta|$ for the full Run 2 dataset, taken from Ref. [152].

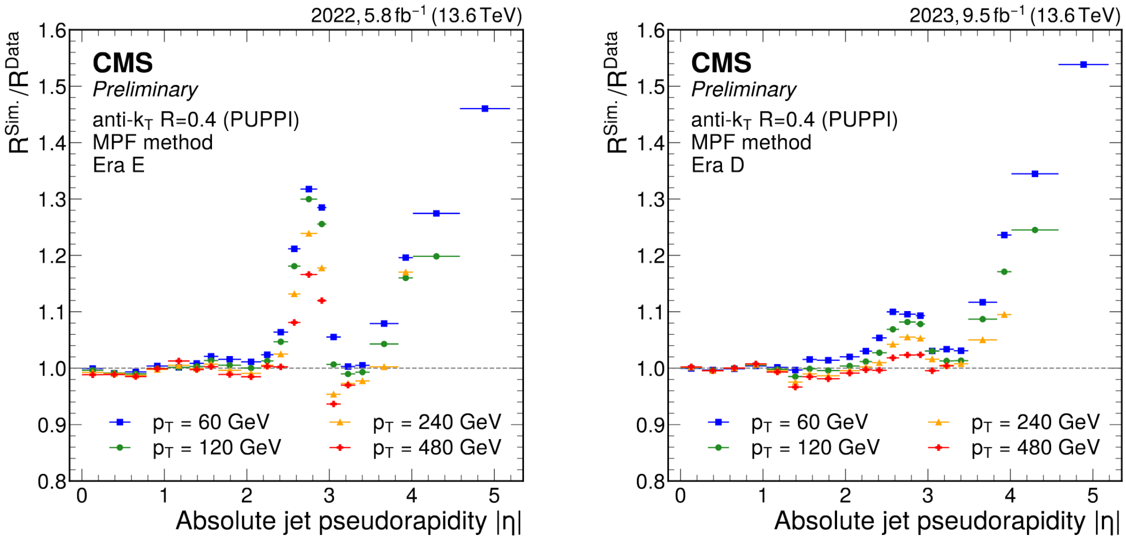


Figure 5.4: L2 Residual corrections for Run 3 (2022: Era E and 2023: Era D) from Ref. [153].

The corresponding Run 3 corrections for the 2022 and 2023 data-taking periods, in which I was directly involved, are presented in Figures 5.4 and 5.5. Figure 5.4 illustrates the corrections for Eras E (2022) and D (2023) at representative jet transverse momenta of 60, 120, 240, and 480 GeV, while Figure 5.5 summarizes the results for 60 and 480 GeV across 2022 and most of 2023. These figures illustrates the year-to-year variations in the corrections,

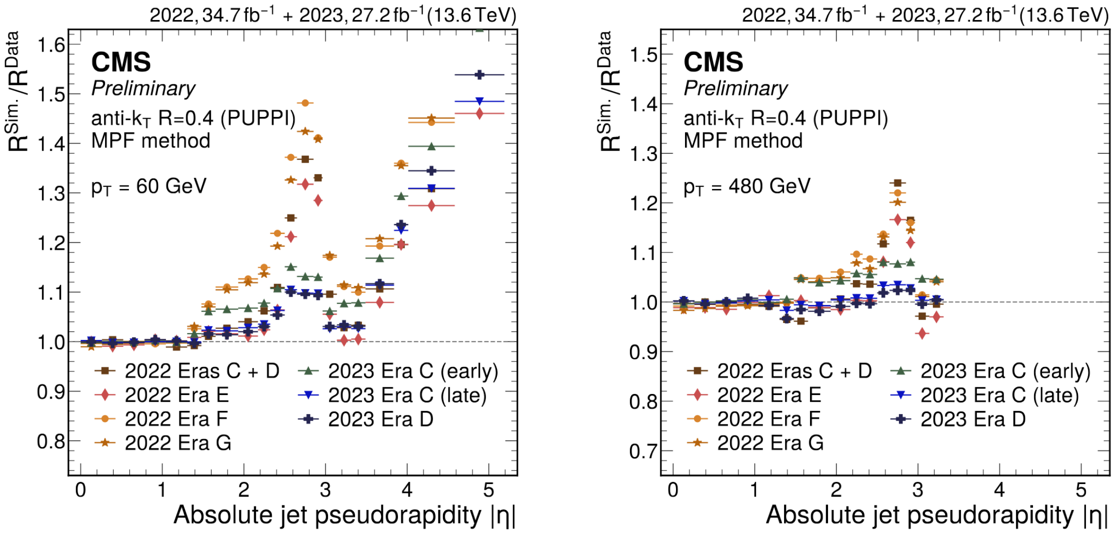


Figure 5.5: L2 Residual corrections for Run 3 (2022 and 2023) from Ref. [153].

which arise from changes in detector performance and operating conditions. Notably, one observes that the correction factor increases to values up to 1.15 in the EC-HF transition region, corresponding to $|\eta| \in [2.5, 3]$, reflecting the reduced detector response in this part of the calorimeter coverage.

- **p_T-dependent corrections**

  These corrections are also known as Level 3 Residuals or L3Res. In order to obtain them, events with different topologies covering the p_T range between 30 GeV and 1 TeV are needed. In the low-p_T regime, Z+1 jet events are utilized, as the leptonic decay channels of the Z boson provide a highly precise experimental signature, therefore the Z boson becomes the tag object. In the central-p_T region, both Z+1 jet and γ+1 jet events are used to achieve higher statistics. Finally, in the high-p_T region, these topologies do not provide enough statistics and multijet events are needed instead. For these corrections, α is defined as

  $$\alpha = \frac{p_{T,jet}}{p_{T,Z/\gamma}} . \tag{5.12}$$

  As a result, the correction factor is defined in terms of the probe jet p_T as $C^i_{L3Res} = R^{i,Simulation}_{jet} / R^{i,Data}_{jet}$, where i refers to a different event topology. Finally, all the C^i_{L3Res} are fitted, resulting in an absolute correction C_{L3Res}.

Finally, the Jet Energy Corrections can be summarised as

$$C^{Simulation}_{JEC} = C_{L1offset} \times C_{Response} \tag{5.13}$$

$$C^{Data}_{JEC} = C_{L1offset} \times C_{Response} \times C_{L2Res} \times C_{L3Res} \tag{5.14}$$

5.2.2 Jet Energy Resolution (JER)

After applying jet energy corrections, the jet energy resolution (JER) in simulation is typically better than in data due to idealized detector modelling and incomplete simulation of effects such as pileup. To account for this, the simulated jet energies are smeared post-correction by applying scale factors derived from the data, ensuring that the simulation reproduces the broader resolution observed in data according to Eq. 5.15. The jet resolution in simulation is estimated from the width, σ_{JER}, of a Gaussian fit to the ratio of reconstructed to particle-level jet response across different PU scenarios:

$$\Delta = \frac{p_{T,reco} - p_{T,part}}{p_{T,part}} \ , \tag{5.15}$$

where $p_{T,reco}$ and $p_{T,part}$ are the transverse momenta of the jet at reconstruction and particle level respectively. Hence, the $p_{T,reco}$ and $p_{T,part}$ are related as

$$p_{T,reco} = p_{T,part}(1 + \Delta) \tag{5.16}$$

Then, the smearing is applied on the simulation through the following correction

$$p_{T,reco} = p_{T,part}(1 + \Delta_{Data/Sim} \times \Delta) \ , \tag{5.17}$$

and the matching criteria between a reconstructed jet and a particle-level jet is defined as

$$\Delta R < \frac{R_{cone}}{2}, \quad |\Delta| < 3\sigma_{JER} \tag{5.18}$$

where R_{cone} is the jet radius parameter of the reconstruction algorithm, which is 0.4 for AK4PFCHS jets, which are the particle flow reconstructed jets with $R = 0.4$ and CHS substraction for PU. However, due to inefficiencies in the detector response, not all reconstructed jets can be unambiguously matched to particle-level jets. Consequently, the Jet Energy Resolution (JER) is adjusted using two complementary methods to account for these discrepancies.

If there is a matching jet at particle level, the scaling method is applied. The four-momentum of the jet is smeared with a factor

$$C_{JER} = 1 + \Delta_{Data/Sim} \times \Delta \tag{5.19}$$

On the opposite, if no matching jet at particle level is found, the stochastic method is implemented:

$$C_{JER} = 1 + \mathcal{N}(0, \sigma_{JER})\sqrt{max(\Delta_{Data/Sim} - 1, 0)} \tag{5.20}$$

where $\mathcal{N}(0, \sigma_{JER})$ is a random number drawn from a Gaussian of width σ_{JER} and negative values of C_{JER} are set to zero to prevent unphysical (negative) jet p_T [154].

5.2.3 B-tagged jets

The identification of jets originating from bottom quarks, or b-tagging, is a core technique in high-energy physics, essential for isolating rare processes and suppressing background in precision measurements [155]. At the Large Hadron Collider, where light-flavour and gluon jets from QCD processes prevail, discriminating heavy-flavour jets provides crucial access to top-quark and Higgs-boson decays as well as searches for new particles coupling preferentially to third-generation quarks. The performance of b-tagging algorithms, in terms of efficiency and purity, directly determines the sensitivity of numerous SM and BSM analyses.

The b-jet tagging relies on the distinctive properties of b-hadrons produced in b-quark fragmentation. Their relatively long lifetime ($\sim$ 1.5 ps) corresponds to a flight distance of a few millimeters for boosted hadrons, resulting in displaced secondary vertices and tracks with large impact parameters relative to the primary vertex. Additional discriminants include the large b-hadron mass ($\sim$ 5.2 GeV/c^2) leading to harder fragmentation and higher decay multiplicity, as well as the 25 % branching ratio of semi-leptonic decays producing soft leptons, particularly muons [155]. In simulation, the CMS experiment classifies jets by hadron flavour as follows: flavour 5 (b-jets containing B-hadrons), flavour 4 (c-jets with C-hadrons only), and flavour 0 (light-flavour jets from u,d,s quarks or gluons) [156].

Contemporary b-tagging algorithms, exemplified by the state-of-the-art DeepJet approach [156], employ deep neural networks, combining convolutional and recurrent layers to process low-level jet constituents (charged/neutral particle four-momenta, track parameters, secondary vertices), surpassing legacy multivariate methods by eliminating manual feature engineering and enhancing discrimination between b-, c- and light-flavour jets. These algorithms output a continuous discriminant ranking jet flavour likelihoods, from which fixed working points (WPs) are defined at standardized efficiencies (e.g., 70% on simulated $t\bar{t}$ samples) to balance signal selection against mistag rates for c- and light-flavour jets, facilitating consistent cross-analysis application and systematic uncertainty evaluation [157].

5.3 Monte Carlo event generators

The simulation of proton-proton collisions can be conceptually separated into two main stages: the generation of the hard process and the subsequent parton showering. In the study of processes such as electroweak Zjj production, the hard scattering is first generated based on the corresponding matrix elements for electroweak Z-boson production in association with two jets. This generation can be performed following two main approaches: the Vector Boson Fusion (VBF) approximation or a non-restrictive approximation.

In the VBF approximation, the generation is limited to the t-channel colour-singlet exchange in order to prevent overlap with other topologies, such as diboson production. In contrast, a non-restrictive approximation includes all Feynman diagrams consistent with the desired final state. Consequently, in addition to genuine VBF contributions, it also accounts for Z^0-boson bremsstrahlung processes and multi-peripheral l^+l^-jj production. For the present analysis, the non-restrictive approximation is adopted, since CMS does not provide a dedicated, centrally produced VBF Z-only Monte Carlo sample for any Run 2 data-taking period.

Once the hard process is generated at the parton level, the simulation of the parton shower is performed to model QCD radiation emitted from the initial- and final-state partons. Subsequently, hadronisation is simulated to convert partons into stable hadrons, together with the modelling of the underlying event.

In this context, a tune is a specific set of phenomenological parameter values, governing parton showers, multiple parton interactions, hadronisation, and colour reconnection, among others, that has been optimized to reproduce a broad set of experimental measurements based on collider data, ensuring simulated events realistically describe observed final state [158].

MC event generators employed in the current measurement

- **Madgraph** is a versatile Monte Carlo event generator at LO or "tree-level" matrix element. `MadGraph5_aMC@NLO` [68] extends it at NLO and interfaces the hard process with parton shower and hadronisation tools.

- **Pythia8** is a general-purpose Monte Carlo event generator able to model hard processes, parton showers, hadronisation and multiparton interactions. Additionally, external events at higher orders can be fed to Pythia8 as well. The parton shower is ordered in decreasing transverse momentum from the hard scattering scale down to a non-perturbative scale, the hadronisation is modelled with the Lund string model and the multiparton interactions are described as presented in [159].

- **Powheg (Positive Weight Hardest Emission Generator)** is a method and a Monte Carlo event generator implemented via the POWHEG BOX framework [70]. It generates the hard process at NLO improving its accuracy and reliability. Then, its output relies on parton shower generators such as Pythia8, to complete the full shower evolution.

5.4 Numerical Treatment of PDF Sets

Parton Distribution Functions (PDFs) are provided as numerical grids that parameterise the non-perturbative parton densities $f_a(x, Q^2)$ in a proton as functions of the momentum fraction x and energy scale Q^2. These grids are generated from global fits to experimental data using Monte Carlo replicas or Hessian eigenvector sets, enabling quantitative uncertainty propagation in high energy physics analyses. For the `NNPDF31_nnlo` sets specifically, uncertainties are quantified through 100 eigenvector pairs derived via the Hessian representation of an underlying Monte Carlo ensemble [160].

PDF sets like NNPDF31_nnlo are distributed as discrete grids in x and Q^2, typically spanning $10^{-6} \leq x \leq 1$ and Q^2 from 0.9 to 10^6 GeV^2, with interpolation (e.g., via neural networks) for intermediate points. Each set consists of a central (best-fit) PDF plus uncertainty variants: NNPDF uses 1000+ replicas for Monte Carlo sampling of the probability distribution, but the Hessian method approximates this with 100 orthogonal eigenvector sets for efficient error propagation. The Hessian eigenvectors $\Delta f_a^{(i)}(x, Q^2)$ capture deviations from the central $f_a^{(0)}(x, Q^2)$, obtained by diagonalizing the covariance matrix of replica fluctuations [161].

5.5 Corrections for particle physics analyses

Correlating the events simulated by the Monte Carlo (MC) generators to the data recorded by the CMS detector is not trivial since many factors need to be taken into consideration. Some of them are the detector limitations and malfunctions, the amount of simulated events as well as the accuracy of the simulation in reproducing the underlying physics processes among other attributes. Because of these, discrepancies between the simulation and data may persist, thereby, the current section discusses the corrections needed in order to mitigate that.

To ensure that kinematic distributions, object reconstruction efficiencies, and event yields in simulation reflect the conditions present in data, targeted corrections are applied. Some of these corrections address differences in normalization (such as scaling event yields using measured cross sections), while others mitigate the impact of additional proton-proton interactions, timing-related trigger inefficiencies, detector hardware failures, or biases in momentum resolution. Still others correct for complex flavor tagging efficiency mismatches or remove problematic detector regions from analysis via veto maps.
Each of these procedures is either centrally developed and validated by CMS working groups or adapted from well established community methods. In the following subsections, these corrections are introduced in turn, beginning from global event-level adjustments and moving towards object-specific calibrations and regional exclusions.

Cross-section scale factor
In high-energy physics analyses, simulated samples must be normalized to the integrated luminosity of the corresponding data set to ensure that the expected event yields are comparable. This scaling accounts for differences between the generated Monte Carlo statistics and the amount of data collected. The cross-section weight for each simulated event i in a sample containing N events is defined as

$$w_i^{XS} = \frac{\mathcal{L}\sigma_{process} w_i}{\sum_N w_i} \,, \tag{5.21}$$

where $\mathcal{L}$ is the integrated luminosity of the data, $\sigma_{process}$ is the cross section of the simulated process and w_i is the event weight. The latter equals 1 at leading order LO but can differ from 1 at NLO due to event reweighting introduced by higher-order corrections.[162].

Multiple proton-proton interactions (pileup or PU)
As discussed previously, the presence of PU introduces additional energy deposits in the detector that can adversely affect the reconstruction of physics objects. The number of PU interactions is given by the following expression:

$$N_{PU} = \frac{\mathcal{L}_{inst}\sigma_{inelastic}}{f_{LHC}} \,, \tag{5.22}$$

where $\mathcal{L}_{inst}$ is the instantaneous luminosity of a given bunch, $\sigma_{inelastic}$ is the cross section of inelastic processes and f_{LHC} is the orbit frequency of the LHC. In simulated samples, the distribution of PU interactions may differ from that observed in data. To account for this, a pileup weight (PU weight) is applied to each simulated event. This weight adjusts the simulated PU distribution to match the one measured in data, ensuring consistent conditions between data and simulation.
Independently of PU weighting, PU mitigation techniques are applied at the object reconstruction level to reduce the impact of additional interactions. In the CMS collaboration, several such methods have been developed. In this analysis, jets are reconstructed using the charged-hadron subtraction (CHS) algorithm, which removes charged particles associated with pileup vertices from the jet clustering. Additionally, a requirement on the pileup jet identification (PU jet ID) discriminant is applied to reject jets originating primarily from pileup interactions [142].

Level 1 prefiring issue

During the 2016 and 2017 data-taking periods, the gradual timing shift of ECAL was not properly propagated to L1 trigger primitives (TP). That led to an important fraction of photons, muons, electrons and jets in the high pseudorapidity region wrongly associated to the previous bunch crossing. Since this effect is not reproduced by the simulation, it is corrected by implementing centrally-provided weights which compute the probability for an event not to prefire. As a result, the final event weight is obtained as the product of the non prefiring probability of all objects (offline photons, muons, electrons and jets) [163]:

$$w = 1 - P_{prefiring} = \prod_{i=jets,photons,muons,electrons} (1 - \epsilon_i^{prefiring}) \tag{5.23}$$

Muon momentum calibration (Rochester correction)

The reconstruction of the momentum of muons in hadron collider experiments is biased. This is due to detector misalignments, the reconstruction software, and uncertainties in the magnetic field. Initially, simulated events have no bias but this can be included later during the reconstruction of the event. This bias depends on the charge of the muon and the η and ϕ coordinates of the muon track. Thus, the muon momentum is calibrated in both data and simulation to remove the discrepancies between both. These correction factors are centrally provided by the CMS Collaboration and were initially derived at the University of Rochester, hence their name. The current measurement uses the roccor.Run2.v5 package [164].

HEM15/16 region issue

The HEM15/16 issue refers to the failure of the power supply in two sectors of the Hadron Calorimeter Endcap (HE), specifically HEM15 and HEM16, approximately halfway through the data acquisition process. These modules correspond to the region of $-3.0 < \eta < -1.3$, $-1.57 < \phi < -0.87$, which resulted in the loss of readout from the corresponding Detector Fiducial (DF) regions, covering a 40° segment of the HE^- side, which corresponds to approximately 3 % of the total HCAL coverage [165]. As a consequence, there was a loss of HCAL information from this sector related to lepton, photon, and jet reconstruction in that region; therefore, this region is excluded in the current measurement for 2018 in both data and simulated events.

B-tagging scale factor

In general, Monte Carlo simulations introduce modelling discrepancies with collision data in tracking, calorimetry, and vertex reconstruction, leading to differences in measured b-tagging efficiencies and misidentification rates for b-, c-, and light-flavour jets, which vary with jet kinematics (p_{T}, η) and working point. Dedicated calibration measurements derive scale factors, ratios of data-to-simulation tagging efficiencies (or mistag rates) parametrized by jet flavour, pseudorapidity and p_T, applied to the simulation to correct for the difference. The current measurement uses the centrally-produced b-tagging scale factors derived from the DeepJet algorithm with fixed working points (WPs). For b-tagged (hadron flavour 5) and c-tagged (flavour 4) jets, scale factors are derived exclusively from QCD-enriched samples, referred to as “mujets”, since they require a soft muon ($p_{\mathrm{T}} > 3, \mathrm{GeV}$) within the jet to ensure high heavy-flavour purity via semileptonic decays. In contrast, light-flavour jets (hadron flavour 0) employ inclusive scale factors derived from standard QCD multijet samples without a muon requirement, utilising data-driven techniques such as negative tagging, where track impact parameter significances are sign-inverted to directly measure mistag rates from the symmetric prompt-track distribution in data. This distinction accounts for the differing statistical control and background compositions in each flavour category [166].

Jet veto maps

Some regions of the calorimeters are observed to produce anomalously high or low jet rates due to radiation damage in addition to inefficiencies in some regions of the tracking system. This leads to unreliable jet reconstruction, which can be accounted for through the implementation of jet veto maps [167]. These are boolean $\eta-\phi$ grids covering the whole detector phase space derived by the Jets and Missing Transverse Energy ("JetMET") group, so that if a jet falls in a *False* evaluated cell, it is removed from the event. The current measurement only implements them for 2018 since the effect of jet veto maps in 2017 and 2016 was negligible.

Chapter 6

Measurement of Electroweak Z Boson With 2 Jets

At the Large Hadron Collider, events containing a Z boson and at least two jets (Zjj) are produced predominantly via initial-state QCD radiation from the incoming partons in the Drell-Yan process (QCD Zjj). Contrary, electroweak production of the Z boson remains a much more infrequent process, since it is suppressed by a much smaller electroweak coupling and phase-space constraints.

During Run 1, the ATLAS experiment measured the fiducial and differential cross sections of the EW Zjj process in several kinematic regions, using an integrated luminosity of 20.3 fb^{-1} of proton–proton collision data collected at $\sqrt{s} = 8$ TeV [63]. In parallel, the CMS experiment measured the fiducial cross section of the EW Zjj process and investigated the hadronic activity in the rapidity interval between the tagging jets for events with $m_{jj} > 120$ GeV, using a data sample corresponding to 5 fb^{-1} of proton–proton collisions at $\sqrt{s} = 7$ TeV [62].

In Run 2 at $\sqrt{s} = 13$ TeV, ATLAS performed measurements with dijet invariant mass requirements of $m_{jj} > 250$ GeV and $m_{jj} > 1$ TeV using 3.2 fb^{-1} of data [168], and later reported differential cross sections for $m_{jj} > 1$ TeV with the full 139 fb^{-1} dataset [169]. At the same centre-of-mass energy, CMS measured the fiducial cross section for $m_{jj} > 200$ GeV with 35.9 fb^{-1} [64]. The present analysis therefore constitutes the first fiducial measurement of the EW Zjj process at CMS using 138 fb^{-1} of 13 TeV data and the first differential measurement of this process with respect the Z-boson p_T in the CMS experiment.

This thesis uses an integrated luminosity of 59.83 fb^{-1} of proton–proton collision data at $\sqrt{s} = 13$ TeV, collected by the CMS experiment in 2018, for the inclusive Zjj measurement. The electroweak Zjj measurement, which represents the main result of this work, employs the full Run 2 dataset corresponding to an integrated luminosity of 138 fb^{-1} at $\sqrt{s} = 13$ TeV, encompassing the 2016, 2017, and 2018 data-taking periods. Both analyses combine the $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$ decay channels.

Section 6.1 define the signal and background processes and their corresponding data and simulation samples. Section 6.2 explains the selected phase space and observables, and the measurement at detector level is shown in Section 6.3. Then, the Machine Learning (ML) strategy implemented to enhance the signal process is explained in Section 6.4. The unfolding procedure is explained in Section 6.5 and finally uncertainties are discussed in Section 6.6.

6.1 Event topologies

In a high-energy physics analysis, a precise characterization of event topologies along with precise identification of signal and background processes is essential for an accurate measurement and interpretation. This section introduces a comprehensive description of the signal and background, followed by the signal topology characterisation at particle level and an overview of the data and simulated samples employed. This foundational framework establishes the basis for event selection and further interpretation efforts, explained in subsequent sections in the current chapter.

6.1.1 Signal and background definitions

The signal, EW Zjj, is defined as the t-channel exchange of a weak vector boson, which includes Z-boson production via vector boson fusion (VBF), Z-boson bremsstrahlung and multi-peripheral production, as shown in Figure 6.1 and previously explained in Chapter 2.

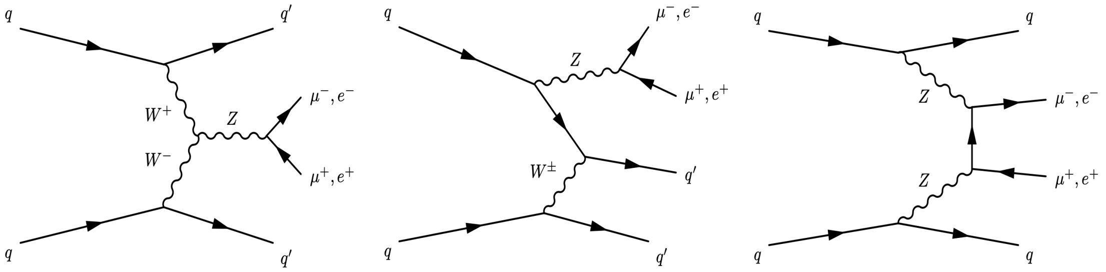


Figure 6.1: Representative leading-order Feynman diagrams for EW Zjj production at the LHC: vector boson fusion (left), Z-boson bremsstrahlung (centre) and multi-peripheral production (right), from Ref. [63] at $\mathcal{O}(\alpha_{EW}^4)$.

In particular, the VBF Z process is of great interest since it acts as a "standard candle" for other VBF processes at the LHC such as the production of Higgs bosons [170–173]. The VBF Z process is also crucial for the search for weakly interacting particles beyond the Standard Model [174–177] or for studying the sensitivity to anomalous WWZ triple gauge couplings [5, 178–181].

Background processes produce the same final state as the process under study and they are classified in three main categories:

- Drell-Yan process (DY) ($q\bar{q} \to Z^0 \to \ell\bar{\ell}$): It constitutes the largest background in the current measurement (see Fig.6.2).
- Top-antitop processes: they include top-antitop pair production where both $W^{\pm}$ bosons decay leptonically into two leptons and two neutrinos in the final state ($t\bar{t} \to 2\ell 2\nu$) and top-antitop pair production where the $W^{\pm}$ boson decay semileptonically ($t\bar{t} \to \ell\nu 4q$) into one charged lepton, one neutrino, manifested as missing transverse energy, and four quarks: two b quarks and two light quarks. These two top-antitop processes constitute the second largest background in the current measurement (see Figs. 6.3 and 6.4).
- Single top: they include single top quark production through 3 different modes: in association with a $W^{+/-}$ boson with all inclusive decay modes for the top quark

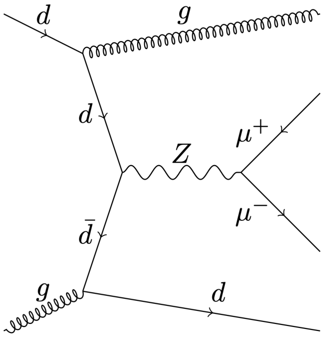


Figure 6.2: Representative leading-order Feynman diagrams for Drell-Yan production in association with dijets, from Ref. [62] at $\mathcal{O}(\alpha_s^2\alpha_{EW}^2)$.

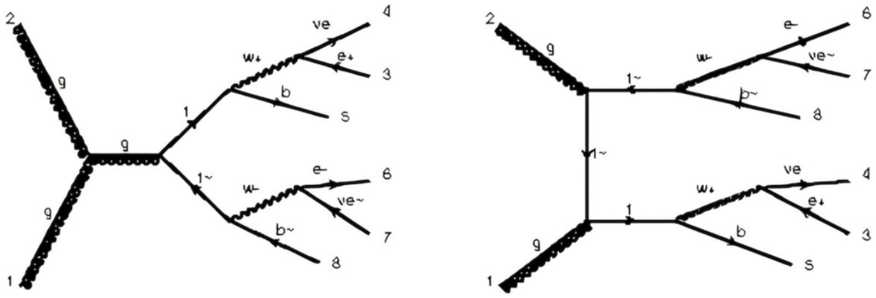


Figure 6.3: Representative leading-order Feynman diagrams for $t\bar{t} \rightarrow 2\ell 2\nu$ obtained with MadGraph5 at $\mathcal{O}(\alpha_s^2\alpha_{EW}^4)$.

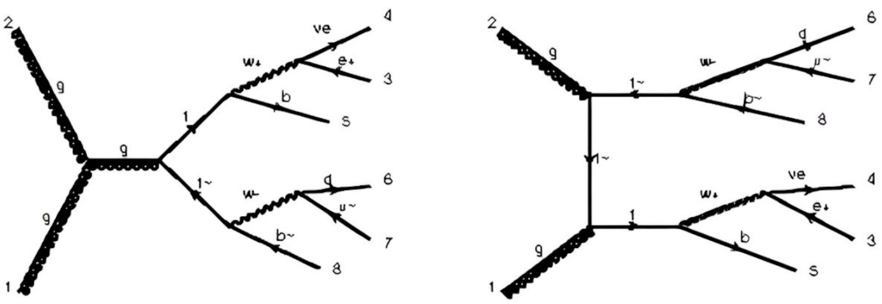


Figure 6.4: Representative leading-order Feynman diagrams for $t\bar{t} \rightarrow \ell\nu 4q$ obtained with MadGraph5 at $\mathcal{O}(\alpha_s^2\alpha_{EW}^4)$.

in the 5-flavour scheme, via t-channel in the 4-flavour scheme and via s-channel in the 4-flavour scheme as well. The difference between these flavour schemes is based on whether the b-quark is considered or not in the initial state (see Figs. 6.5 and 6.6).

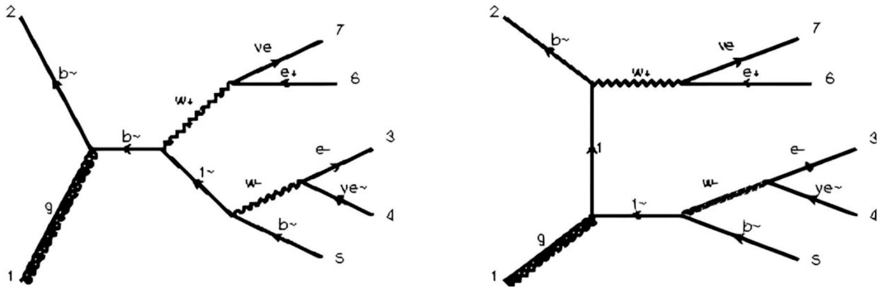


Figure 6.5: Representative leading-order diagrams for single top quark production with a $W^{+/-}$ boson with all inclusive decay modes in the 5 flavour scheme obtained with MadGraph5 at $\mathcal{O}(\alpha_s \alpha_{EW}^4)$.

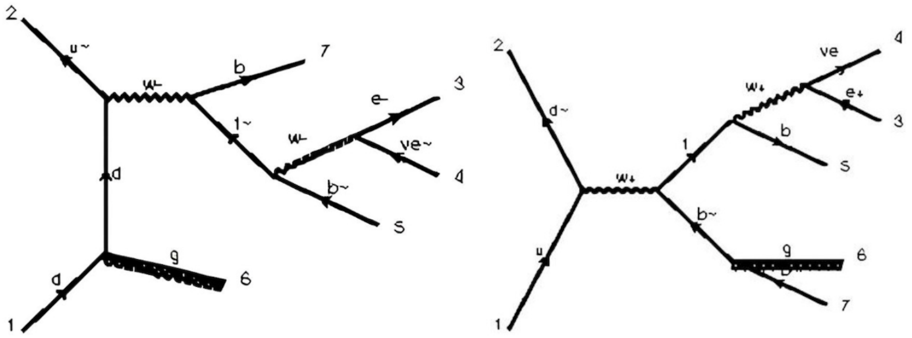


Figure 6.6: Representative leading-order diagrams for single top quark production via t-channel and s-channel in the 4-flavour scheme obtained with MadGraph5 at $\mathcal{O}(\alpha_s \alpha_{EW}^4)$.

- Vector Boson: they include W^+W^-, $W^{+/-}Z^0$ and Z^0Z^0 production. Overall, these processes constitute the smallest background contributions (see Fig. 6.7).

Background processes are modelled individually, with dedicated simulated samples used to characterise each contribution, as detailed in the following subsection. Building on the inclusive Z selection criteria described in Chapter 2, an additional set of kinematic cuts is applied to enhance the electroweak Zjj signal while suppressing the Drell-Yan background. However, as illustrated at particle level in Figure 6.8, it is challenging to isolate the signal from the background processes, and therefore a machine learning (ML) method is implemented to improve the signal vs. background ratio. Given the substantial contributions from Drell-Yan and top-antitop production, a dedicated control region for each of them is defined to both determine their normalisation and shape directly from the data and to further constrain their impact in the extraction of the electroweak measurement. This approach is also motivated by discrepancies in the CMS Collaboration regarding theoretical cross section values.

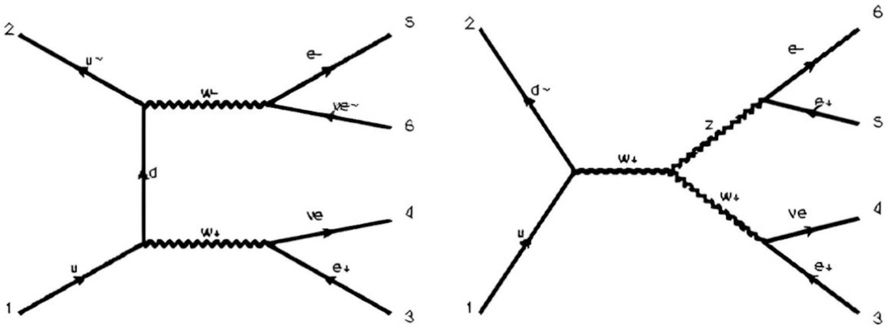


Figure 6.7: Representative leading-order diagrams for $W^{+}W^{-}$ and $W^{+/-}Z^{0}$ production obtained with MadGraph5 at $\mathcal{O}(\alpha_{EW}^{4})$.

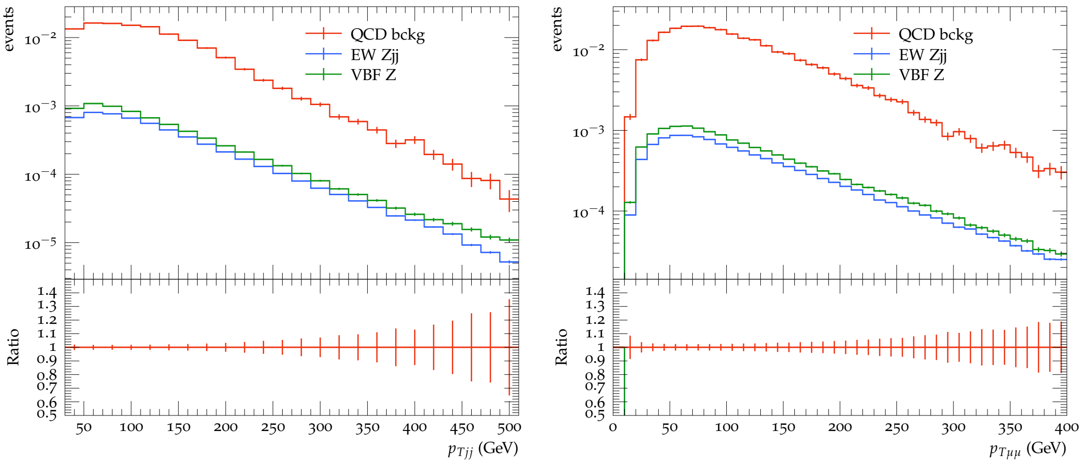


Figure 6.8: $p_{T_{ll}}$ and $p_{T_{jj}}$ at particle level, where the contribution in red "QCD bckg" corresponds to the Drell-Yan process and the EW Zjj corresponds to the signal process (normalised to the cross section).

The EW Zjj, VBF Z and DY samples used for the particle-level distributions shown in this thesis have been generated according to the footnote.[1]

6.1.2 Physical event samples

As previously mentioned, background processes are modelled individually using dedicated simulated samples for each contribution, as detailed in the following subsection. Accordingly, this section describes the data and simulated samples employed in the present measurement.

---

[1] Both the EW Zjj signal and the QCD background (DY) shown in distributions at particle level in this thesis have been simulated using the datacards available in the CMS Monte Carlo Management (McM) repository [182] corresponding to the dataset names: `EWK_LLJJ_MLL-50_MJJ-120_TuneCP5_13TeV-madgraph-pythia8_dipole` and `DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8` correspondingly, referred to as `EWK_LLJJ_MLL` and `DYJetsToLL` respectively in the current chapter, and the VBF Z process has been simulated as described in Chapter 2.

6.1.2.1 Data samples

Events in the dataset are selected by requiring them to satisfy a set of high-level trigger (HLT) (see Section 4.3.2) conditions designed to ensure the presence of charged leptons. Since the current measurement is performed by combining $Z \rightarrow e^+e^-$ and $Z \rightarrow \mu^+\mu^-$ decay channels, `SingleMuon` and `SingleElectron` or `EGamma` datasets are needed, as shown in Tables C.1 and C.2 in the Appendix.

As described in Chapter 2, the EW Zjj process is characterised by two leptons in the final state, but single-lepton triggers are particularly effective due to their relatively higher transverse momentum thresholds and their high efficiency in recording events with energetic, isolated leptons. This choice not only ensures good signal purity but also contributes to a significant reduction of background contamination. Furthermore, the performance of single-lepton triggers has proven to report a highly similar event rate across a wide range of pileup conditions and in different pseudorapidity ranges, owing to the different upgrades in the Level-1 (L1) trigger system (see Section 4.3.2) and the High-Level Trigger algorithms. Such stability is especially important when combining datasets collected in different LHC running periods. The efficiency of the CMS L1 and HLT triggers has been extensively studied using clean control samples enriched in $Z \rightarrow \mu\mu$ decays, providing robust and well-understood efficiency measurements. These properties make single-lepton triggers particularly advantageous for electroweak analyses involving final states with leptons and jets. This choice aligns with CMS Trigger system performance and selection documented in CMS trigger publications and performance notes [183].

6.1.2.2 Simulation samples

Simulation samples are employed to model the underlying physics processes, evaluate efficiencies and background contributions, estimate systematic uncertainties, and correct the data for detector effects.

In this analysis, event samples are produced using different MC generators tailored to the specific physics process and the desired level of theoretical accuracy.

Electroweak processes such as EW Zjj are not generally available in a standardised way from NLO simulation with `MadGraph5_aMC@NLO`. This is due to the computational cost and technical complexity associated with simulating electroweak Z production, particularly when compared to the inclusive (EW+QCD) case. Moreover, including NLO QCD corrections would mix the signal with the much larger QCD background, making it difficult to isolate and measure the pure EW contribution. Including NLO EW corrections would manifest as a slightly reduced theoretical uncertainty rather than a significant change in event kinematics. Consequently, in the VBF-enriched phase space, LO accuracy remains sufficient for the current measurement. Additionally, at the time of the Run 2 simulation campaign, no dedicated validation or tuning existed for NLO electroweak processes, leading to the *a priori* decision to simulate the Zjj electroweak process at leading order [68] with `MadGraph5_aMC@NLO`.

For processes with complex multi-leg final states, such as Drell-Yan production and single-top production via the s-channel, event generation is performed with `MadGraph5_aMC@NLO` [68] due to its great flexibility. The matching to the `Pythia8` parton shower is achieved through the `FXFX` matching procedure [184], a standardized framework designed to avoid double counting in events with complex final states.

For top-antitop and single-top production channels (excluding the s-channel), events are generated with `POWHEG` [70]. This choice reflects `POWHEG`'s ability to generate the hardest radiation first, yielding a more accurate description of radiation patterns and jet kinematics characteristic of these processes. Furthermore, its treatment of the interplay between initial- and final-state radiation through the Sudakov form factor provides a consistent framework for modelling interference and resonance effects. In the single-top t-channel, generated in the four-flavour scheme, `MadSpin` [185] is employed to preserve spin correlations in decays. The resulting events are then interfaced with `Pythia8` [186] for parton showering, hadronisation, and underlying-event modelling, using the CP5 tune optimized and validated by the CMS Collaboration [158].

For single-top s-channel production, the choice of `MadGraph5_aMC@NLO` over `POWHEG` is primarily motivated by the more transparent implementation of the four-flavour scheme (4f), enabling precise treatment of heavy-flavour dynamics and better control of decay channels, rather than by the matching procedure itself.

Other processes, such as diboson production, are simulated directly with `Pythia8`, which provides sufficient leading-order precision for these channels. All simulated samples employ the NNPDF3.1 parton distribution functions [187], provided at NNLO in QCD.

The event generator and PDF configurations used follow the central CMS prescriptions. Tables D.1,D.2, D.3, and D.4 summarize the simulated samples corresponding to the 2018, 2017, and 2016 datasets. The 2016 data are further divided into pre-VFP and post-VFP periods to account for a hardware modification in the CMS pixel detector, where the Vertex Finding Processor (VFP) pre-amplifier bias voltage was adjusted to enhance silicon strip tracker hit efficiency. Accordingly, four tables are presented to represent these distinct data-taking periods.

6.2 Phase space and observable selection

The electroweak production of a Z boson with two jets (EW Zjj) in a VBF-like topology displays characteristic kinematic and topological features. In this process, a t-channel exchange of electroweak bosons between incoming quarks produces two jets, referred to as "tagging jets". These jets are typically very energetic, widely separated in rapidity, and located in the forward and backward regions of the detector, with little hadronic activity in between. The Z boson decay products, forming a dilepton system, are generally found within the rapidity interval defined by the

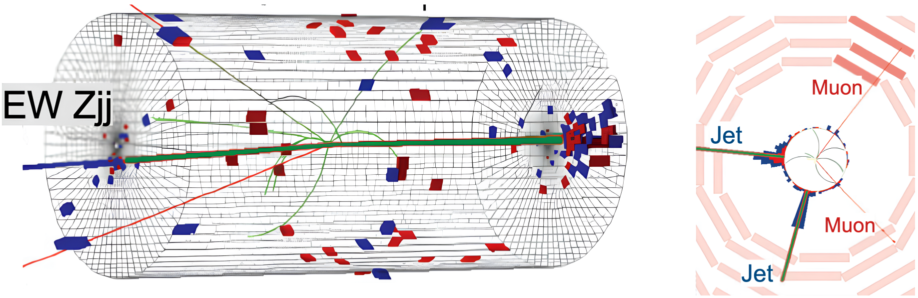


Figure 6.9: EW Zjj topology in the CMS detector: 3D view (left), transverse view (right) from [188].

tagging jets, as shown in Fig. 6.9.

In the current and following sections, observables denoted as X_{l0} and X_{j0} correspond to quantities defined with respect to the leading lepton and leading jet, respectively. Similarly, X_{l1} and X_{j1} refer to the corresponding quantities for the subleading lepton and subleading jet. Observables labelled X_{ll} and X_{jj} represent quantities defined for the dilepton and dijet systems, respectively, which will be formally introduced in the next subsection.

6.2.1 Event selection

Following the previously discussed kinematic characteristics of the electroweak Zjj process, event selection is performed by applying a set of criteria. The baseline selection comprises requirements derived from the physical features of the studied topology, denoted as *A. Physics-based criteria.* Additional conditions accounting for detector response and performance are then applied, referred to as *B. Detector-based criteria.*

A. Physics-based criteria

Events are selected after applying a set of preselection criteria summarised in Table 6.1.

Channel	$\mu\mu$	ee
Leptons	$\lvert\eta\rvert < 2.4$	$\lvert\eta\rvert < 2.1$
	$p_{Tl0,1} > 40, 20$ GeV	
Z peak	exactly 2 with opposite sign	
	$\lvert M_Z - M_{ll}\rvert < 15$ GeV	
	$min(\eta_j) < \eta_{ll} < max(\eta_j)$	
	$\Delta\Phi_{ll} < 2.8$	
Jets	at least 2	
	$p_{Tj0,1} > 50, 50$ GeV	
	$m_{jj} > 200$ GeV	
	$\lvert\eta\rvert < 4.7$	
	$\lvert\Delta\eta_{jj}\rvert > 1.5$	

Table 6.1: Summary of event preselection criteria for the muon and electron channels

In a vector boson fusion (VBF) Z-enriched topology, several categories of jets can be present:

- **Tagging jets:** the two hardest jets originating from the t-channel quark exchange. The jets with the highest and second-highest transverse momentum (p_T) are referred to as the *leading* and *subleading* jets, respectively, and together they form the *dijet system.* They are typically highly energetic, widely separated in rapidity, and located in the forward and backward regions of the detector, with limited hadronic activity between them

- **Soft gluon jets:** jets arising from Initial or Final State Radiation (ISR or FSR), usually characterized by low transverse momentum ($p_T < 30$ GeV) and predominantly central pseudorapidity

- **Extra quark or gluon jets:** jets originating from higher-order QCD radiation, which can have moderate to high p_T depending on the event topology
- **Pileup jets:** low-p_T jets uncorrelated with the hard scattering process, produced by additional proton–proton interactions within the same bunch crossing

To ensure the presence of the tagging jets while efficiently suppressing soft gluon activity, at least two jets with $p_{T,j0,1} > 50$ GeV are required. Furthermore, a minimum pseudorapidity separation between the leading and subleading jets, $|\Delta\eta_{jj}| > 1.5$, significantly reduces the QCD-induced background, which generally produces more central jets rather than the characteristic back-to-back configuration of the EW Zjj system. A conservative dijet invariant mass threshold of $m_{jj} > 200$ GeV is chosen to suppress QCD contributions while retaining sufficient event statistics.
For the reconstruction of the Z^0 boson, exactly two oppositely charged leptons are required. The leading lepton must satisfy $p_{T,0} > 40$ GeV and the subleading lepton $p_{T,1} > 20$ GeV, ensuring the suppression of non-prompt lepton contributions. The dilepton pair must be consistent with the Z^0 boson mass, fulfilling the condition $|M_Z - M_{\ell\ell}| < 15$ GeV. In addition, leptons are required to be spatially correlated, with an azimuthal separation $\Delta\Phi_{\ell\ell} < 2.8$, and located within the pseudorapidity interval defined by the tagging jets, $\min(\eta_j) < \eta_{\ell\ell} < \max(\eta_j)$, consistent with the VBF-like topology.

B. Detector-based criteria

Vertex Reconstruction and Pileup Mitigation

Reconstructed primary vertices are required to lie within 24 cm of the nominal detector center along the beam axis (z-direction) and within 2 cm of the beamspot position in the transverse plane. Vertices outside these bounds typically correspond to background events or poorly reconstructed interactions. Furthermore, the vertex fit is required to have more than four degrees of freedom to guarantee that a sufficient number of tracks contribute to the vertex reconstruction, ensuring an acceptable reconstruction quality [189].

Leptons

The lepton p_T threshold fulfils the requirements to activate the adequate triggers, which are the `HLTIsoMu24` trigger for 2018 and 2016 and `HLTIsoMu27` for 2017 in the muon channel and `HLT_Ele32_WPTight_Gsf` trigger for 2018 and 2016 and `HLT_Ele35_WPTight_Gsf` for 2017 in the electron channel.
Leptons are subject to standardised quality, positioning and isolation requirements, which ensures an optimal selection for physics analyses.

- **Quality**: muons need to pass the `mediumId` requirement, which is a set of criteria on variables such as track fitting, track hits, matches to calorimeter deposits, and isolation criteria. It is used to differentiate prompt muons from background muons originating from heavy flavour decays or misidentifications [190]. Likewise, electrons need to pass the `tightId` requirement [191]
- **Positioning**: muons need to keep a distance of $d_{xy} < 0.5$cm and $d_z < 1.0$cm relative to the primary vertex in the xy-plane and z-coordinates respectively. Electrons are required to satisfy $d_z < 0.1, \text{cm}$ and $d_{xy} < 0.05, \text{cm}$ if their $|\eta| \leq 1.479$ and $d_z < 0.2$cm and $d_{xy} < 0.1$cm otherwise [192]

- **Isolation**: In the muon channel, it is defined by requiring that the scalar sum of transverse momenta within a cone of radius $R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.4$ is below 25% of the muons' p_T, suppressing non-prompt backgrounds while preserving prompt muons [193]. Similarly, in the electron channel, it is defined by requiring that the scalar sum of transverse momenta within a cone of radius $R < 0.3$ is below 15% of the electrons' p_T [194].

Jets
Jets are subject to standardized quality requirements, isolation criteria, tagging procedures, dedicated event filtering, and data–MC efficiency mapping:

- **Quality:** Analogous to the lepton selection, jets must satisfy the `tightId` condition, ensuring genuine multi-particle contributions and suppressing noise-induced artifacts [195]
- **Isolation:** Jets within $\Delta R < 0.4$ of any preselected electron or muon are removed to avoid double counting of energy deposits
- **Tagging:** To enhance purity of the tagging jets from t-channel quark exchange in the EW Zjj topology, a b-tag veto is imposed. This analysis employs the `medium` working point (WP) of the DeepJet algorithm, corresponding to a $\sim 1\%$ misidentification rate for light-flavour jets [196]
- **Filtering:** Standard MET filters aare applied to remove events with anomalous missing transverse energy arising from detector noise, reconstruction errors, or beam-related backgrounds. These filters target events with poorly reconstructed interaction points, unusual calorimeter signals, misreconstructed muons, and known calibration issues in the electromagnetic calorimeter [197]
- **Mapping:** Jet veto efficiency maps are used to correct for residual data–MC discrepancies relevant to 2018 data, while effects in 2016 and 2017 are negligible. These corrections mitigate biases arising from jet resolution in analyses that are explicitly conditioned on the presence of reconstructed jets [167]

Furthermore, jets reconstructed in the regions $-3.2 < \eta < -1.3$ and $-1.57 < \phi < -0.87$ are excluded to address the non-functional HEM-15 and HEM-16 HCAL sectors during the 2018 data-taking period [198].

6.2.2 Observables

The present chapter details the measurement of both inclusive and differential cross sections as functions of the dilepton transverse momentum, p_T^{ll}. Alongside this primary observable, a suite of complementary variables is selected to maximize analysis sensitivity. The selection of these observables is motivated by their capacity to differentiate signal from background processes, which is reflected in distinct distribution shapes observed for signal and background contributions. Consequently, the chosen variables are those demonstrating strong discriminating power, minimal mutual correlations, and robust agreement between data and simulation at the detector level. As a result, the following observables are employed in the Boosted Decision Tree (BDT) training.

- m_{jj}: The dijet invariant mass (Fig. 6.10) is computed from the two highest-p_T jets (tagging jets) and is a key variable for identifying events with a large mass separation between the jets. In VBF Z production, a high m_{jj} is a signature of the typical topology where the two jets are produced at large rapidity intervals, reflecting the underlying electroweak process and helping to suppress the DY background
- p_{Tj_1}: The transverse momentum of the subleading jet (Fig. 6.11) provides insight into the jet kinematics characteristic of VBF Z production. In genuine VBF events, both tagging jets typically arise from hard parton scattering via colour-singlet exchange, resulting in a balanced and relatively hard jet pair even for the subleading jet. In contrast, DY Zjj events often feature softer secondary jets produced through QCD radiation. As a result, a harder p_{Tj_1} spectrum is indicative of the VBF topology and serves as an effective discriminant against the DY background
- $\Delta\Phi_{Z,jj}$: The angular opening in the azimuthal plane between the Z boson (reconstructed from the dilepton system) and the dijet system (Fig. 6.12) helps identify events with high transverse momentum balance, typical for back-to-back topologies such as the VBF Z production
- $\Delta\Phi_{Z,j_0}$: The azimuthal separation between the Z boson and the leading jet (Fig. 6.13) is sensitive to the distinct topology of VBF Z production. In genuine VBF events, the Z boson is typically produced centrally and decorrelated in azimuth from the tagging jets. DY processes, by contrast, often feature jets radiated close in azimuth to the Z boson. This variable therefore enhances the ability to isolate the VBF-like topology
- n_{l_0}: The pseudorapidity of the leading lepton (Fig. 6.14) probes the central production of the Z boson decay products in VBF events. Since the colour-singlet exchange characteristic of VBF tends to yield centrally produced leptons between the two forward-tagging jets, while DY Zjj events can produce leptons more broadly in rapidity, this observable helps capture that spatial pattern
- $\Delta\eta_{ll}$: The pseudorapidity difference between the two leading leptons (Fig. 6.15) reflects the angular spread of the dilepton system arising from the Z decay. In VBF Z events, the leptons are typically emitted closer in pseudorapidity, consistent with a central, colour-singlet–mediated production, whereas DY events show a more uniform distribution. This variable thus contributes to distinguishing the more collimated dilepton topology expected in VBF production
- $|\Delta\eta_{jj}|$: The absolute pseudorapidity separation between the two leading jets (Fig. 6.16) is one of the hallmark signatures of the VBF process. The two VBF tagging jets are usually produced widely separated in rapidity and located at large $|\eta|$, a contrast to DY Zjj events, which tend to exhibit more central and less separated jets. Large values of $|\Delta\eta_{jj}|$ therefore strongly favour the VBF topology and provide a key handle to enhance signal purity
- z^*: Zeppenfeld variable (Fig. 6.17), defined as:

$$z^* = \frac{y^*}{\Delta y_{jj}} \ ,$$

where Δy_{jj} is the rapidty separation between the leading and subleading jets and y^* is:

$$y^* = y_{ll} - \frac{y_0 - y_1}{2} \ ,$$

where y_{ll} is the Z (dilepton) rapidity, y_0 is the leading jet rapidity and y_1 is the subleading jet rapidity.

The Zeppenfeld variable is a widely used observable for distinguishing vector boson fusion (VBF) Z production from Drell-Yan (DY) Z+jets backgrounds, as it quantifies the relative centrality of the electroweak boson with respect to the dijet system. In VBF topologies, the two jets originate from a t-channel electroweak exchange and therefore tend to populate the forward regions of the detector, producing a large Δy_{jj}, while the Z boson is preferentially emitted centrally between them. This yields small values of y^* and a z^* distribution strongly peaked around zero. Conversely, in DY Z+jets events, the jets arise predominantly from QCD radiation and exhibit no particular correlation with the Z boson rapidity, leading to a much broader z^* spectrum. The variable z^* therefore provides an effective means of enhancing the purity of VBF-enriched event selections by exploiting the characteristic spatial configuration of electroweak production.

These distributions are normalized to the expected event yields, computed from theoretical cross sections and integrated luminosity per year, and corrected for all weights and detector-level scale factors (as described in *Detector-based criteria* in Section 6.2.1). The signal process is labelled as EWK $Z \to \ell^+\ell^-$, while the background processes are:

- VV: diboson processes (W^+W^-, $W^\pm Z^0$, and Z^0Z^0)
- Single top: single top production in association with a $W^\pm$ boson in the 5-flavour scheme, including all inclusive decay modes of the top quark, single top production via the t-channel and s-channel in the 4-flavour scheme
- TT: top-antitop processes involving $t\bar{t} \to 2\ell 2\nu$ and $t\bar{t} \to \ell\nu 4q$
- DY $Z \to \ell^+\ell^-$: Drell–Yan process

Figures 6.10–6.17 reveal two main differences between the $\mu\mu$ and ee channels. The data–simulation agreement is systematically slightly better in the $\mu\mu$ channel, consistent with observations across Run 2. This trend reflects the greater sensitivity of the ee channel to ECAL-related effects, as electron reconstruction and isolation depend on superclusters and bremsstrahlung photon recovery, both of which are susceptible to distortions from central hadronic activity that are difficult to reproduce in simulation. Although such activity is limited in the signal topology, it remains significant for dominant backgrounds such as Drell–Yan, which, together with $t\bar{t}$ processes, generates a higher rate of fake and non-prompt electrons than muons, amplifying residual mis-modelling effects in the ee channel. Additionally, in 2018 the Data/MC ratios are negative in the $\mu\mu$ channel but positive in the ee channel, while in 2017, 2016preVFP, and 2016postVFP both channels show ratios near unity or negative. The 2018 deviation originates from reduced ee statistics due to the HEM15/16 region, as discussed in Section 5.4.
Results from the other data-taking eras are provided in the Appendix B.

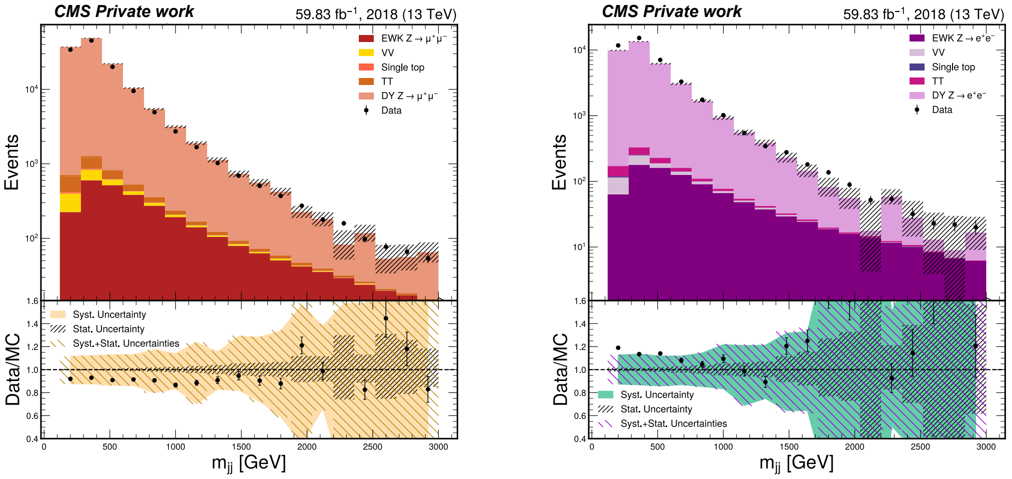


Figure 6.10: Dijet mass distribution at detector level in the $\mu\mu$ channel (left) and ee channel (right).

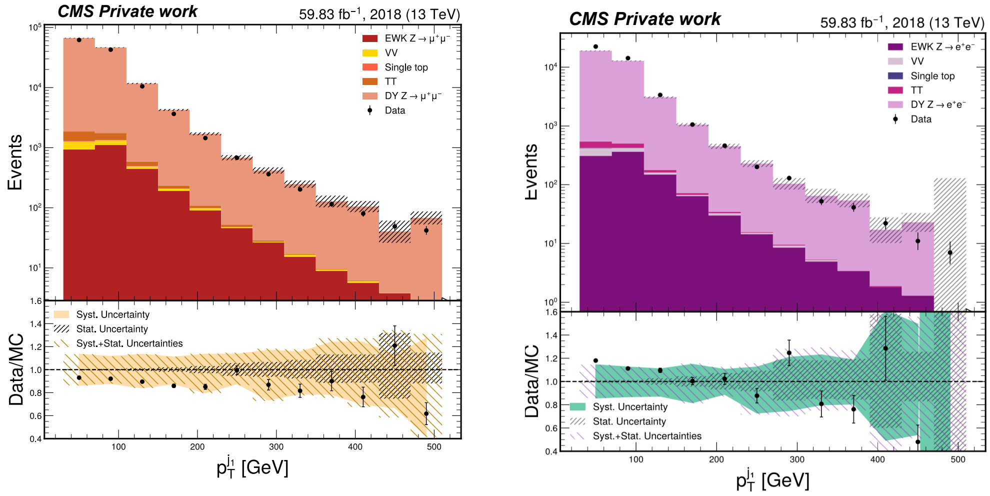


Figure 6.11: Subleading jet distribution at detector level in the $\mu\mu$ channel (left) and ee channel (right).

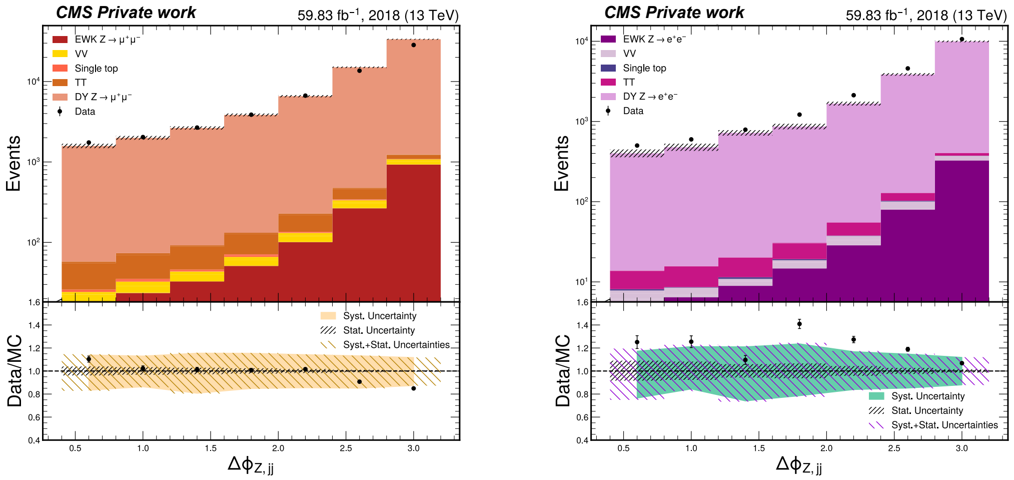


Figure 6.12: $\Delta\phi_{Z,jj}$ distribution at detector level in the $\mu\mu$ channel (left) and ee channel (right).

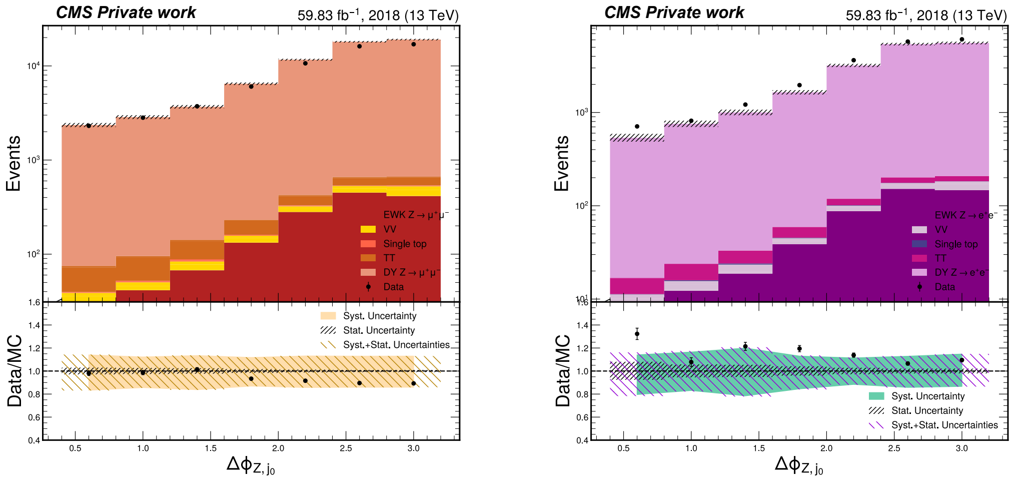


Figure 6.13: $\Delta\phi_{Z,j_0}$ distribution at detector level in the $\mu\mu$ channel (left) and ee channel (right).

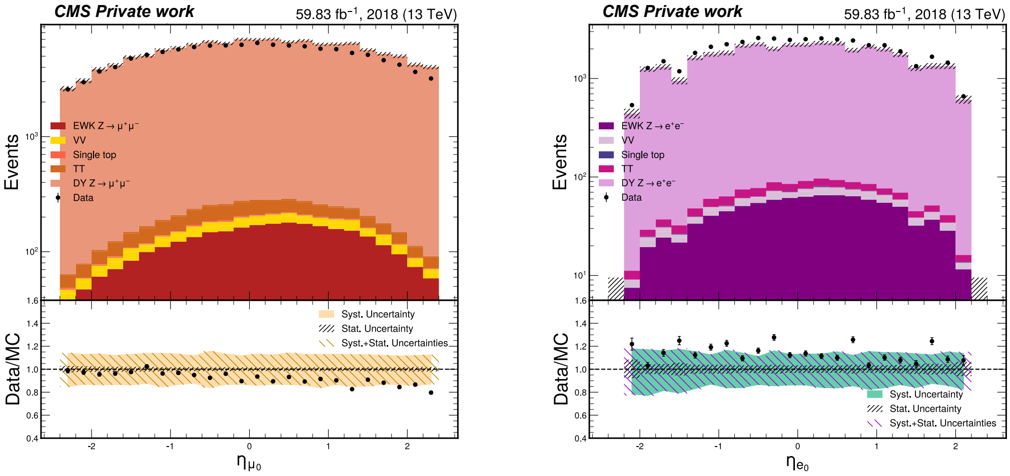


Figure 6.14: Leading lepton pseudorapidity distribution at detector level in the $\mu\mu$ channel (left) and ee channel (right).

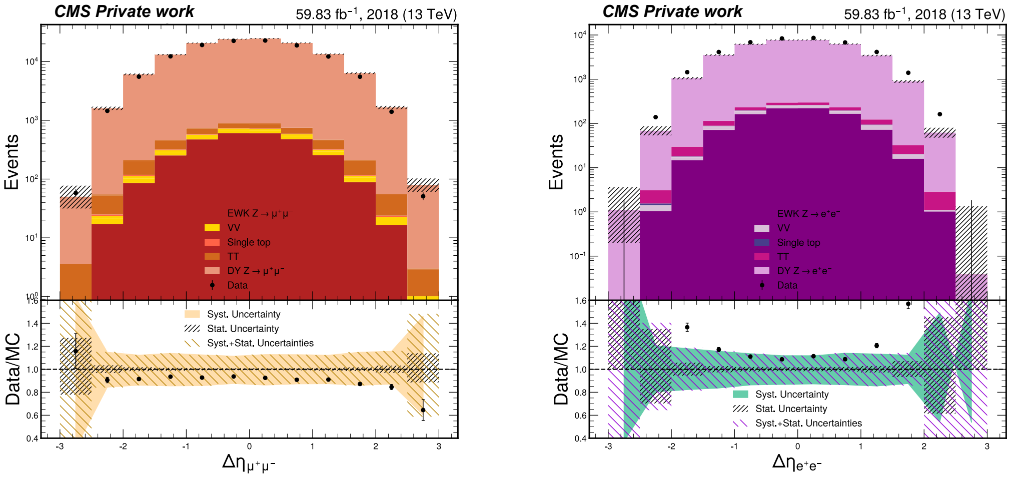


Figure 6.15: $\Delta\eta_{ll}$ distribution at detector level in the $\mu\mu$ channel (left) and ee channel (right).

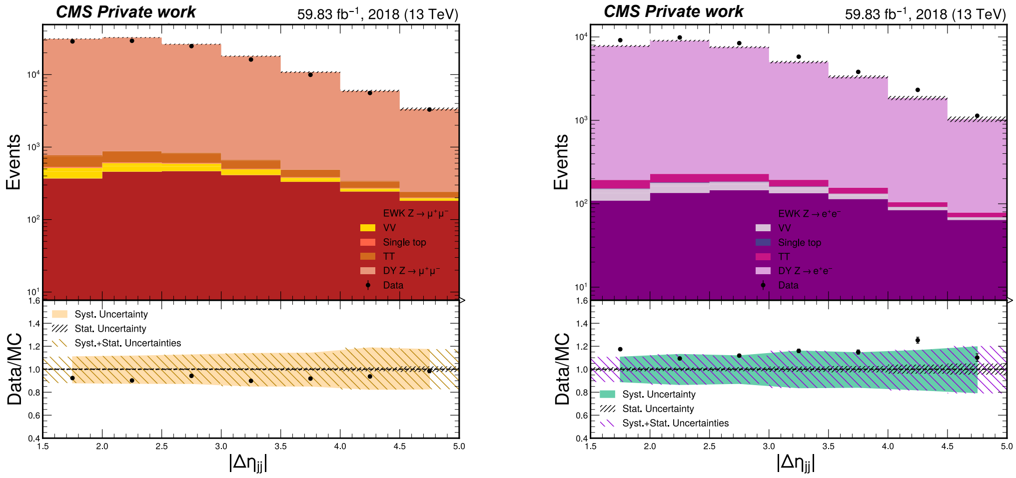


Figure 6.16: $|\Delta\eta_{jj}|$ distribution at detector level in the $\mu\mu$ channel (left) and ee channel (right).

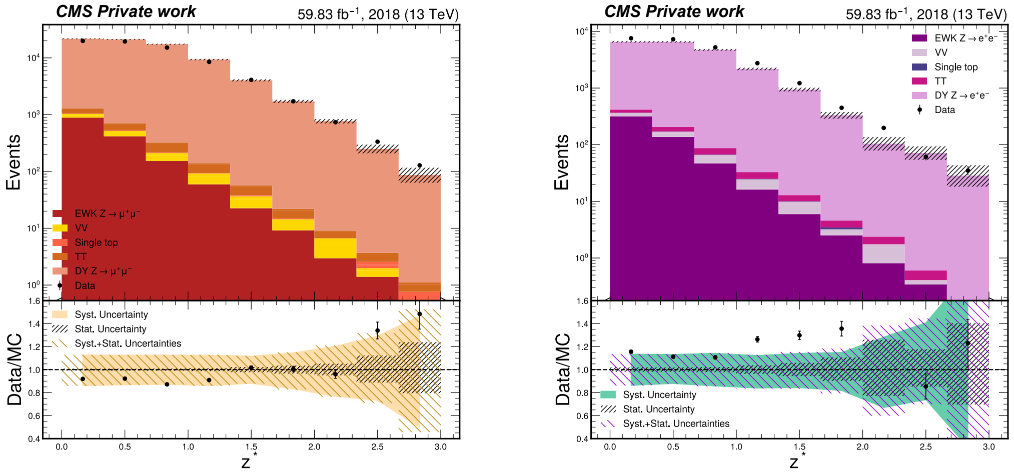


Figure 6.17: Zeppenfeld variable distribution at detector level in the $\mu\mu$ channel (left) and ee channel (right).

6.3 Measurement at detector level

This section presents the measurement of the dilepton transverse momentum, p_T^{ll}, in both channels for the 2018 data-taking period, depicted in Fig. 6.18. The distributions are shown at detector level after applying the kinematic selection criteria listed in Table 6.1 and the *detector-based criteria* presented afterwards. The comparison between data and simulation demonstrates satisfactory agreement. Results from the other data-taking eras are provided in the Appendix A.

In Fig. 6.18, the upper panels include the simulation samples stacked after applying all detector-level corrections, with the total simulation uncertainty shown as a gray dashed band.

As mentioned in Section 6.2.2, the signal process is labelled as EWK $Z \to \ell^+\ell^-$, while the background processes are:

- VV: diboson processes (W^+W^-, $W^\pm Z^0$, and Z^0Z^0)
- Single top: single top production in association with a $W^\pm$ boson in the 5-flavour scheme, including all inclusive decay modes of the top quark, single top production via the t-channel and s-channel in the 4-flavour scheme
- TT: top-antitop processes involving $t\bar{t} \to 2\ell 2\nu$ and $t\bar{t} \to \ell\nu 4q$
- DY $Z \to \ell^+\ell^-$: Drell–Yan process

The lower panel shows the data-to-simulation ratio, where the solid band corresponds to the systematic uncertainty, defined as measurement biases originating from imperfections in the detector, theoretical modelling, or analysis techniques, rather than from random statistical fluctuations. The gray dashed band to the statistical uncertainty, and the coloured dashed band to the total uncertainty, further details regarding the uncertainties can be found in Section 6.6. Additionally, Fig. 6.19 presents the same distribution illustrating the uncertainty main component. The lower panels include the contribution from Jet Energy Corrections in the uncertainty bands, which remain the dominant source of uncertainty throughout the measurement due to its final state, where the dijet system is highly sensitive to the jet resolution.

6.4 Signal purity optimization with Machine Learning

Due to the limited event statistics and the reduced cross section of the electroweak Zjj signal relative to the dominant Drell-Yan background, it is essential to employ strategies that enhance the separation between the two processes. In modern high-energy physics analyses, this task is often addressed through multivariate techniques, which provide powerful tools for exploiting correlations among kinematic variables and thereby improving measurement sensitivity [199]. In particular, Boosted Decision Trees (BDTs) have proven especially effective: they capture non-linear dependencies, remain robust against irrelevant inputs, and perform reliably even with modest Monte Carlo statistics [200]. Compared to earlier neural network approaches,

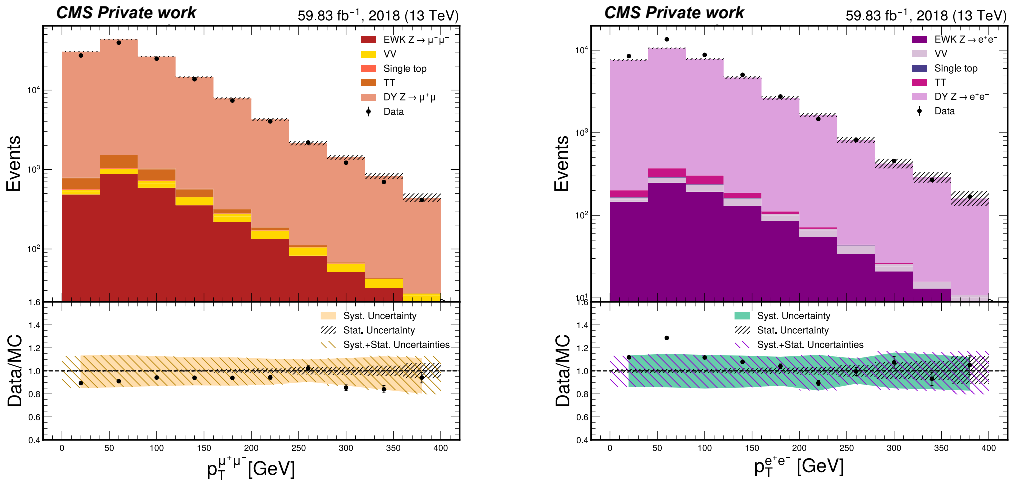


Figure 6.18: Comparison of the dilepton transverse momentum ($p_T^{\ell\ell}$) between data and simulation for the 2018 data-taking period. These distributions are normalized to the expected event yields, computed from theoretical cross sections and integrated luminosity per year, and corrected for all weights and detector-level scale factors (as described in *Detector-based criteria* in Section 6.2.1).

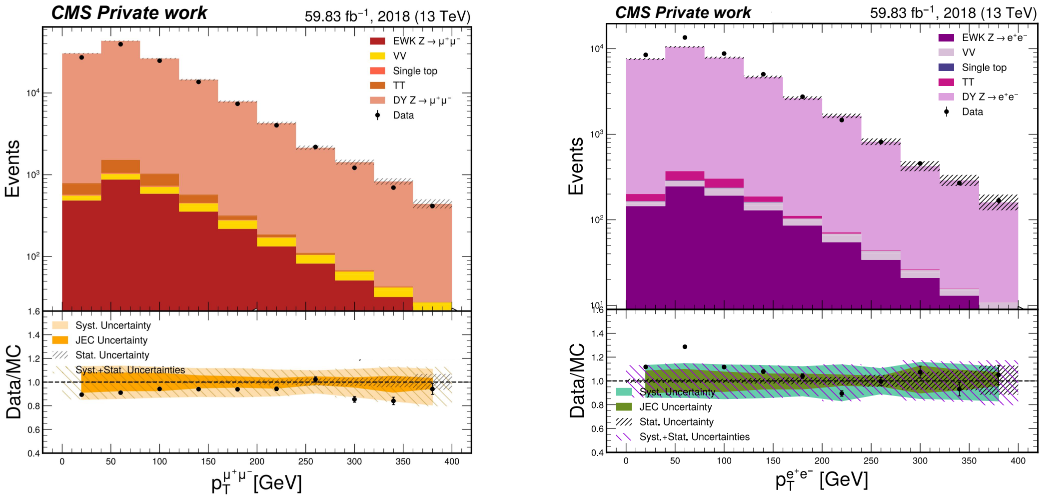


Figure 6.19: Comparison of the dilepton transverse momentum ($p_T^{\ell\ell}$) between data and simulation for the 2018 data-taking period with explicit Jet Energy Correction uncertainties in the ratio plot, for the 2018 data taking period. These distributions are normalized to the expected event yields, computed from theoretical cross sections and integrated luminosity per year, and corrected for all weights and detector-level scale factors (as described in *Detector-based criteria* in Section 6.2.1).

BDTs require less extensive tuning, exhibit more stable behaviour, and offer transparent measures of variable importance, making them well suited for collaborative analyses within CMS [201]. The following section provides a detailed description of the Boosted Decision Tree algorithm applied in this measurement.

In this analysis, the only signal of interest is electroweak Zjj production. For the unfolded measurement, this signal is later divided into bins of the dilepton transverse momentum, p_T^{ll}. These bins represent the same underlying physics process, the EW Zjj, and differ only by kinematics, meaning the p_T^{ll} bin. Training a multiclass classifier to separate individual signal bins from background would therefore not increase discrimination power. Instead, the classification task is formulated as a binary problem: signal versus background, ensuring that the model focuses solely on distinguishing the process of interest from competing processes. After classification, the signal events are assigned to p_T^{ll} bins using the relevant kinematic observable, so that the classifier output remains unbiased while preserving the differential information required for the measurement.

Since the Drell–Yan process is the dominant background, whereas top quark pair, single top, and diboson contributions are comparatively small, all backgrounds are combined into a single class for training and evaluation. To retain sensitivity to individual components, a secondary label is propagated through the machine learning workflow, allowing the backgrounds to be later separated into Drell–Yan, top quark pair, and an “others” category (single top and diboson), without influencing the BDT training.

6.4.1 Mathematical Foundations of Boosted Decision Trees

A Boosted Decision Tree (BDT) is an ensemble-based machine learning algorithm in which multiple weak decision trees are combined into a single, powerful classifier through an iterative boosting procedure. Each decision tree partitions the multidimensional feature space into regions that predominantly correspond to either signal or background events, thereby producing a piecewise-constant approximation to the underlying class probability distribution. Individually, such trees possess limited discriminative power; however, by successively training each tree to correct the misclassifications of the preceding ensemble, the boosting algorithm constructs an additive model that progressively reduces the classification error. Additionally, the adaptive nature of boosting ensures that events with ambiguous classification receive increased attention during training, leading to improved overall sensitivity even when limited by modest Monte Carlo statistics.

This subsection formalizes the mathematical framework underlying the BDT method. In order to statistically separate signal-like from background-like events in collider data, a probabilistic binary classifier based on logistic regression and gradient boosting is implemented. This classifier estimates the probability that an event with features $\mathbf{x}$ belongs to the signal or background class. The method follows the formalism introduced by J. H. Friedman in the 1999 lectures on gradient boosting machines [202].

The binary target variable is denoted as $y \in \{-1, +1\}$, corresponding to background and signal classes, respectively. The classifier output $F(\mathbf{x})$ is related to the likelihood ratio of the classes, forming the basis of the model.

$$\mathcal{L}(y, F) = \log\left(1 + e^{-2yF}\right), \tag{6.1}$$

defines the logistic loss function. Minimizing this loss corresponds to maximizing the likelihood that the predicted score $F(\mathbf{x})$ correctly classifies each event. Physically, this encourages large positive values of $F(\mathbf{x})$ for signal ($y = +1$) and large negative values for background ($y = -1$) events, enhancing signal-background discrimination. The score $F(\mathbf{x})$ can be interpreted as half the log-odds of the event belonging to the signal class:

$$F(\mathbf{x}) = \frac{1}{2}\log\left[\frac{\Pr(y = 1 \mid \mathbf{x})}{\Pr(y = -1 \mid \mathbf{x})}\right]. \tag{6.2}$$

This relates the model output directly to a likelihood ratio, which is a powerful statistical discriminator in hypothesis testing for signal presence.

At each boosting iteration m, the model is updated by fitting a base learner to the negative gradient of the loss, known as the pseudo-response:

$$\tilde{y}_i = -\left[\frac{\partial \mathcal{L}(y_i, F(\mathbf{x}_i))}{\partial F(\mathbf{x}_i)}\right]_{F=F_{m-1}(\mathbf{x}_i)} = \frac{2y_i}{1 + e^{2y_i F_{m-1}(\mathbf{x}_i)}}, \tag{6.3}$$

which indicates the direction and magnitude of corrections needed to reduce model error locally. From a physical viewpoint, $\tilde{y}_i$ down-weights well-classified events and emphasizes poorly classified ones, guiding learning towards challenging regions of phase space.

The step size ρ_m for applying the new tree model $h(\mathbf{x}; \mathbf{a}_m)$ is determined by minimizing the total loss over all training events:

$$\rho_m = \arg\min_{\rho} \sum_{i=1}^{N} \log\left(1 + e^{-2y_i(F_{m-1}(\mathbf{x}_i) + \rho h(\mathbf{x}_i; \mathbf{a}_m))}\right). \tag{6.4}$$

This ensures the update optimally improves the classifier's likelihood-based objective.

When regression trees are used as base learners, parameter updates are refined per terminal node R_{jm}. The optimal leaf value γ_{jm} is given by

$$\gamma_{jm} = \arg\min_{\gamma} \sum_{\mathbf{x}_i \in R_{jm}} \log\left(1 + e^{-2y_i(F_{m-1}(\mathbf{x}_i) + \gamma)}\right). \tag{6.5}$$

Since this problem has no closed-form solution, a single Newton-Raphson iteration provides an efficient approximation:

$$\gamma_{jm} = \frac{\sum_{\mathbf{x}_i \in R_{jm}} \tilde{y}_i}{\sum_{\mathbf{x}_i \in R_{jm}} |\tilde{y}_i|(2 - |\tilde{y}_i|)}. \tag{6.6}$$

This step can be interpreted as locally adjusting predictions by accounting for curvature of the loss, enhancing the convergence speed and stability.

Together, these steps form the gradient boosting classification framework based on maximizing the logistic likelihood. The eXtreme Gradient Boosting (XGBoost) algorithm [203, 204] builds upon this by incorporating both first- and second-order derivatives of the loss for efficient optimization and adds regularization on tree complexity to mitigate overfitting, leading to robust and powerful classification models in the context of high-energy physics analyses. Furthermore, it is the most reliable and offers a higher number of features in comparison with other major tree boosting systems, as shown in Table 1 in [203].

6.4.2 BDT training procedure

The BDT is trained using half of the simulated samples, with the observables described in Section 6.2.2 as inputs. The remaining half is used for testing. Both the training and testing are performed independently for each analysis channel and data-taking year. As previously mentioned, these observables are minimally correlated with the dilepton transverse momentum, avoiding the accidental inclusion of biases during the unfolding procedure, as shown in Fig. 6.20.

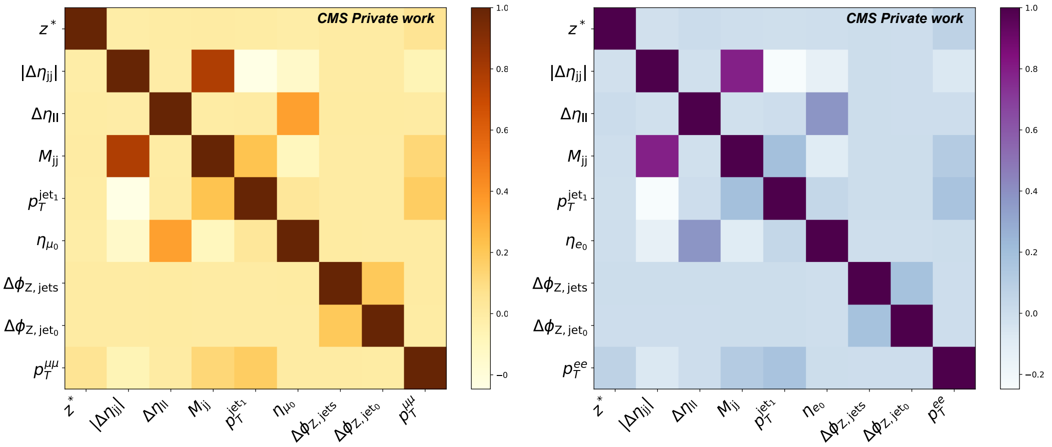


Figure 6.20: BDT inputs correlation matrices for the $\mu\mu$ channel (left) and ee channel (right), 2018.

The training process is implemented with distinct stages for nominal event values and for each up/down variation of the JEC (see Chapter 5). This separation is necessary because JEC variations modify the shape of input observables in a nontrivial way, making a BDT trained only on nominal data inaccurate when applied to these shifted samples. Treating JEC variations as generic systematic uncertainties without retraining can thus yield unphysical classifier responses. By training individual BDTs for each JEC variation per channel and year, the classifiers capture the specific kinematic changes induced by these variations, ensuring physically meaningful and statistically consistent outputs during signal extraction.

Once trained, the BDT's performance is evaluated using the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC), which from a physics perspective quantifies the classifier's ability to distinguish signal and background processes over all possible selection thresholds. The ROC curve plots the true positive rate, the signal efficiency, against the false positive rate, or background contamination, summarizing how well the BDT exploits differences in event kinematics and topology. This curve specifically measures the probability that a randomly chosen EW Zjj signal event receives a higher classifier score than a background event, directly reflecting the model's sensitivity to the underlying physical features that separate these production mechanisms. Monitoring the ROC curve during BDT training provides insight into how effectively the model captures physics-driven event features without overfitting statistical fluctuations. As shown in Figure 6.21, the BDT demonstrates robust performance in both channels, achieving an Area Under the Curve of the ROC curve over 80% and an accuracy around 75%. This indicates the classifier's strong ability to generalize the underlying physical differences between signal and background events.

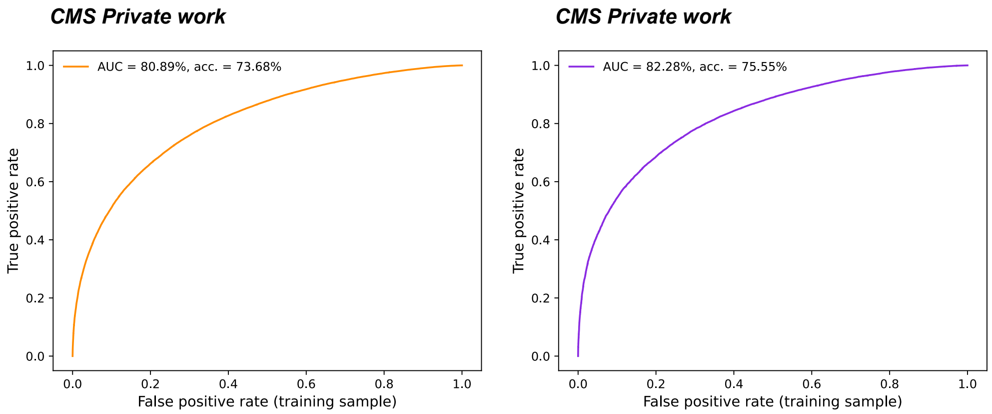


Figure 6.21: AUC values for training in the $\mu\mu$ channel (left) and ee channel (right), 2018.

6.4.3 BDT evaluation procedure

In the machine learning framework, the likelihood gradient boosting algorithm commonly uses the target variable defined as $y \in \{0, 1\}$, contrasting with the statistical convention $y \in \{-1, 1\}$. This convention facilitates the use of the logistic function, allowing the model output $f(\mathbf{x})$ to be mapped directly to a probability via $Pr(y = 1|\mathbf{x}) = \sigma(f(\mathbf{x})) = \frac{1}{1+e^{-f(\mathbf{x})}}$, which provides a straightforward probabilistic interpretation of the BDT response. Figure 6.22 shows the Boosted Decision Tree (BDT) prediction scores evaluated on simulated samples for both channels in 2018. The EW Zjj signal is represented by a line histogram overlaid on stacked background categories: Drell–Yan (DY), top-antitop (TT), and "others" (single top and diboson). The BDT effectively assigns higher scores to signal events, demonstrating good separation. However, applying a stringent threshold (e.g., $\text{BDT}_{\text{score}} > 0.9$) to extract a pure signal sample is impractical due to the resulting low statistics. Consequently, no BDT cut is applied in the analysis to maintain sufficient event counts. In the BDT output distributions, the electron channel exhibits a more peaked signal-versus-background separation compared to the muon channel, despite separate training

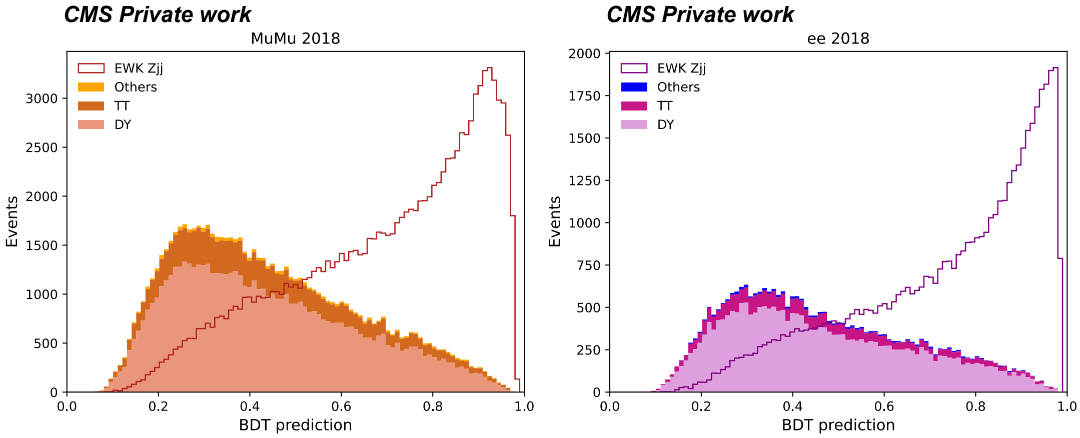


Figure 6.22: BDT prediction score distributions for the $\mu\mu$ channel (left) and ee channel (right) in 2018. The EW Zjj signal is shown as a line histogram, while background contributions are stacked by process categories.

for each. This feature arises primarily from the distinct high-level trigger selections across data-taking years: for 2018 and 2016, `HLT_Ele32_WPTight_Gsf` and `HLT_IsoMu24` are used, while for 2017, `HLT_Ele35_WPTight_Gsf` and `HLT_IsoMu27` are employed. The electron triggers impose higher p_T thresholds (32–35 GeV) and stringent identification criteria, preselecting a cleaner, harder lepton sample with reduced background contamination and enhanced kinematic separation already at the trigger level. Conversely, the muon triggers employ lower thresholds (24-27 GeV) and isolation-based selection, retaining a broader phase space with greater signal-background overlap in key input variables such as lepton p_T, η, and event shapes. Although data-MC agreement is marginally better at detector level in the muon channel, as shown in Fig. 6.18, likely reflecting superior muon reconstruction efficiencies and fewer fakes, the trigger-induced sculpting dominates the observed BDT shape difference.

In HEP analyses, the BDT output, which typically ranges between 0 and 1, is often transformed using functions based on the inverse hyperbolic tangent function [205], such as Eq. 6.7,

$$t\mathrm{BDT} = \mathrm{atanh}\left(\frac{\mathrm{BDT}+1}{2}\right) , \tag{6.7}$$

which maps the BDT score onto a wider range, in this case $(0.55, \infty)$. This transformation spreads the BDT response, making it more Gaussian-like and thus more suitable for statistical modelling such as likelihood fits. It also improves numerical stability, particularly near the boundaries, and enhances sensitivity to regions where the discrimination between the signal and the background is higher. Physically, this facilitates a more precise and continuous discriminant variable, enabling improved separation of signal and background events according to their underlying kinematic and topological properties. Figure 6.23 shows the transformed BDT prediction for both channels for 2018, where both signal and background processes are represented as stacked histograms with the same labelling as Figure 6.22.

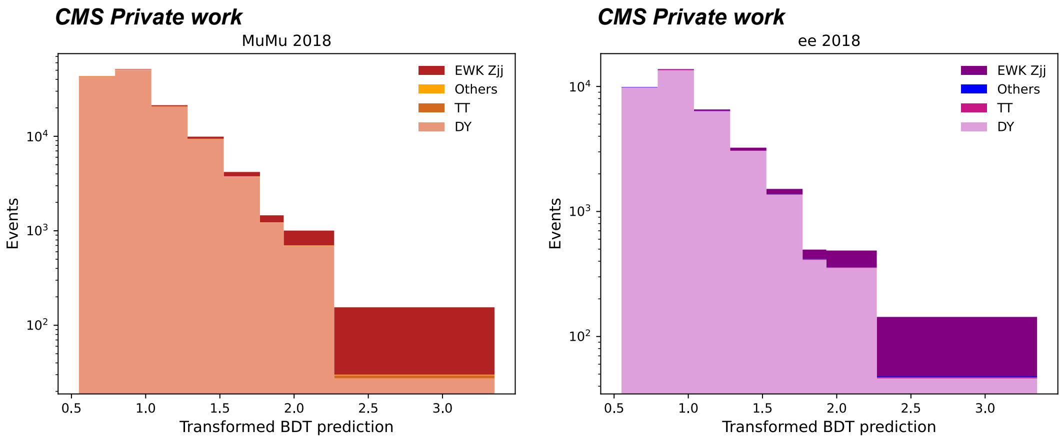


Figure 6.23: tBDT evaluation output for $\mu\mu$ channel (left) and ee channel (right), 2018. They are normalized to the expected event yields, computed from theoretical cross sections and integrated luminosity per year, and corrected for all weights and detector-level scale factors (as described in *Detector-based criteria* in Section 6.2.1).

Testing the AUC during BDT evaluation confirms that the classifier reliably generalizes the physics-driven features learned during training, ensuring robust signal-background discrimination critical for precise measurements. The similarity between

AUC values obtained in training and testing across channels and years, as illustrated in Figures 6.21 and 6.24 for 2018, demonstrates an absence of overfitting, meaning the BDT successfully avoids being biased by the training data, which would be the case if its performance captures noise or patterns only present in the training stage. This consistency validates the stability and physical interpretability of the BDT outputs throughout the analysis.

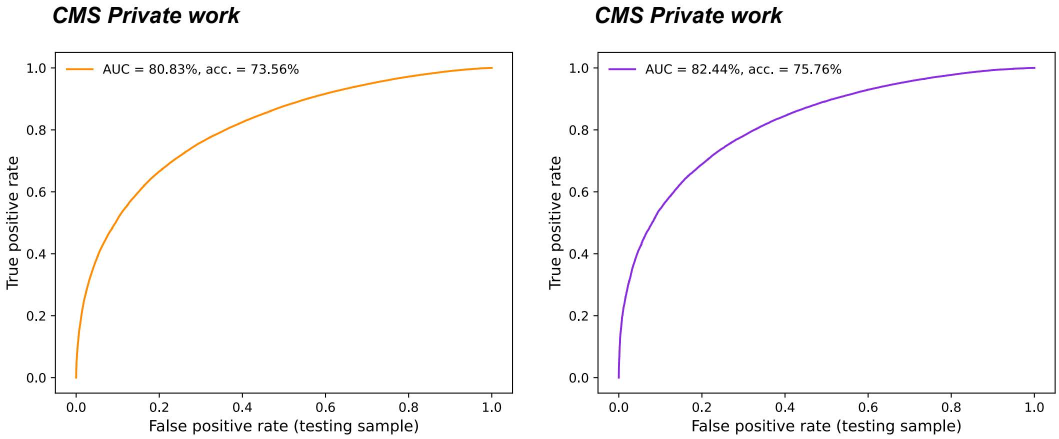


Figure 6.24: AUC values for testing in the $\mu\mu$ channnel (left) and ee channel (right), 2018.

The relevance of the observables used in the BDT training is evaluated through the frequency with which a variable is selected to split a tree node, which is referred to as F-score.

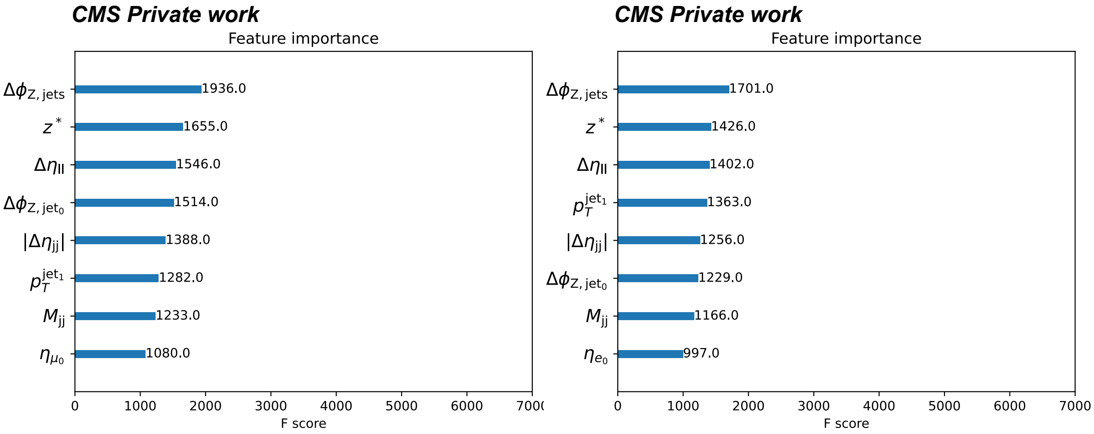


Figure 6.25: F-scores for $\mu\mu$ channel (left) and ee channel (right), 2018.

For the present measurement, the $\Delta\Phi_{Z,jj}$ observable exhibits the strongest discriminating power in both channels, as shown in Figure 6.25, a trend consistently observed throughout all Run 2 data-taking periods. This outcome further supports the choice of $\Delta\Phi_{Z,jj}$ as the key observable for defining control regions in the maximum likelihood fit unfolding procedure.

6.5 Unfolding

After applying the defined kinematic selections, a statistical analysis is performed using the RooStats framework, implemented within the Higgs Combination Tool (`Combine` [206]), to extract both inclusive and differential cross sections.

For the extraction of the EW Zjj differential cross-section, the `Combine` framework provides a distinct advantage over TUnfold [207] and RooFit-based [208] unfolding approaches. Its fully likelihood-driven statistical model allows the simultaneous treatment of signal and background processes, while consistently incorporating systematic uncertainties as nuisance parameters that can be constrained directly by the data. This capability is particularly essential for the EW Zjj process, where the signal yield is small and both backgrounds and detector effects are complex, necessitating precise handling of correlations and uncertainty propagation as will be detailed in the following subsections.

This section introduces the relevant statistical concepts, Section 6.5.1 focuses on signal extraction and Section 6.5.2 introduces the likelihood-based unfolding.

6.5.1 Signal extraction

To determine the inclusive cross section of the electroweak Zjj process, `Combine` maximizes the likelihood function by simultaneously finding the values of the model parameters, including the signal strength parameter (the parameter of interest or POI) and uncertainties (nuisance parameters) that maximize the overall likelihood. This maximization procedure yields the best-fit estimates of these parameters that make the observed data most probable under the statistical model.

Background yields are either treated as fixed contributions, as is the case for Single Top and diboson (VV) backgrounds, or they are constrained and adjusted via additional nuisance parameters derived from dedicated control regions, as implemented for the Drell-Yan and $\mathrm{t\bar{t}}$ backgrounds. This approach ensures a coherent, data-driven determination of the signal rate while consistently incorporating systematic uncertainties.

During the fit, the signal strength parameter r is treated as the POI and acts as a multiplicative scaling factor on the predicted signal yield within the statistical model, thereby quantifying deviations of the observed rate from the nominal theoretical prediction. Consequently, this fitted parameter can be identified with the conventional signal strength definition, μ, expressed as the ratio of the measured cross section to its Standard Model (SM) expectation:

$$\mu = \frac{\sigma(\text{EW Zjj})}{\sigma(EWZjj)^{\text{SM}}} \ . \tag{6.8}$$

The maximum-likelihood fit to the data is performed across N bins of the dilepton transverse momentum, p_T^{ll}. In each bin i, the probability of observing n_i events in data, given an expected yield $\lambda_i(\mu)$ (neglecting systematic uncertainties for illustration), follows a Poisson distribution:

$$P_i(n_i|\mu) = \frac{\lambda_i(\mu) e^{-\lambda_i(\mu)}}{n_i!} \ , \tag{6.9}$$

where the number of expected events $\lambda_i(\mu)$ is given by:

$$\lambda_i(\mu) = \mu \cdot s_i + \sum_j^{N_{bck}} b_{i,j} \ . \tag{6.10}$$

In this expression, s_i denotes the number of expected signal events, while $b_{i,j}$ represents the contribution from the j^{th} background source. The likelihood, which quantifies the probability of observing the data for a given value of μ, is constructed as the product of 6.10 over all N bins of the distributions.

$$\mathcal{L}(\mathbf{n}|\mu) = \prod_{i=1}^{N} \frac{\lambda_i(\mu) e^{-\lambda_i(\mu)}}{n_i!} \ . \tag{6.11}$$

Here, the vector $\mathbf{n} = (n_1, \ldots, n_N)$ represents the observed number of events per bin. When systematic uncertainties are taken into account, they are incorporated through a set of nuisance parameters, $\mathbf{\Theta}$, whose values are constrained by the data, as discussed in Ref. [209]. In this case, the expected yield λ_i becomes a function of both μ and $\mathbf{\Theta}$, i.e. $\lambda_i(\mu, \mathbf{\Theta})$, and the likelihood function is extended to include the prior knowledge of the nuisance parameters.

$$\mathcal{L}(\mathbf{n} \mid \mu, \mathbf{\Theta}) = \prod_{i=1}^{N} \frac{\lambda_i(\mu, \mathbf{\Theta}) e^{-\lambda_i(\mu, \mathbf{\Theta})}}{n_i!} \cdot \prod_{m=1}^{M} f(\Theta_m) \ . \tag{6.12}$$

The second product runs over the M nuisance parameters, with $f(\Theta_m)$ denoting the pdf of the m^{th} parameter. Most are implemented as Gaussian constraints, such that

$$f(\Theta_m) = e^{-\frac{1}{2}(\Theta_m - \Theta_{0m})^2/(\Delta\Theta_m)^2} \ , \tag{6.13}$$

where Θ_m denotes the prior uncertainty of the nuisance parameter, and Θ_{0m} its nominal value. For parameters with large uncertainties or those restricted to positive values, the Gaussian constraint is replaced by a log-normal distribution.

p-value and significance

The p-value $p(q_{\text{obs}}|\mathcal{H})$ [206] is the probability of obtaining a result at least as extreme as the observed measurement, assuming hypothesis $\mathcal{H}$. For a background-only hypothesis $\mathcal{H}_0$, this probability is

$$p_0 = \int_{q_{\text{obs}}}^{\infty} f(q_0|\mathcal{H}_0)\, dq_0 \ , \tag{6.14}$$

where $f(q_0|\mathcal{H}_0)$ is the probability density of the test statistic under $\mathcal{H}_0$. Standard rejection thresholds in HEP are 1.4×10^{-3} (3σ) and 2.9×10^{-7} (5σ), with the significance calculated as $Z = \Phi^{-1}(1 - p_0)$, where Φ^{-1} is the Gaussian inverse cumulative distribution. For signal exclusion, the criterion $p_\mu < 0.05$ corresponds to 1.64σ [206].

6.5.2 Likelihood-based unfolding

As introduced in Section 6.5.1, the unfolding procedure is based on a maximum likelihood approach. For a measurement with N generator-level bins, equation 6.8

can thus be generalized as:

$$\vec{\mu} = \begin{pmatrix} \frac{\sigma_0}{\sigma_0^{\mathrm{SM}}} \\ \frac{\sigma_1}{\sigma_1^{\mathrm{SM}}} \\ \vdots \\ \frac{\sigma_N}{\sigma_N^{\mathrm{SM}}} \end{pmatrix}, \tag{6.15}$$

where σ^{SM} denotes the expected Standard Model cross section in each generator-level bin at particle level, as defined later in this section.

The negative log-likelihood is minimized to extract the best estimates of all μ_k simultaneously, with bin-to-bin migrations caused by detector smearing or reconstruction inefficiencies leading to events generated in bin k being reconstructed in bin $j \neq k$, naturally taken into account.

This procedure is implemented in practice by defining distinct regions and data-taking periods using templates constructed separately for the 2016 pre-VFP, 2016 post-VFP, 2017, and 2018 datasets. For each period, both signal and background processes are modelled under nominal selections as well as all relevant systematic variations. The signal component includes N independent entries corresponding to the generator-level bins of the measured observable, the dilepton transverse momentum. This setup enables a simultaneous extraction of the signal yield across all years and regions without introducing additional constraints. As a result, the unfolded differential cross section of the dilepton transverse momentum is obtained in a consistent manner, with systematic uncertainties properly propagated throughout all datasets.

Signal region

The signal region is defined by selections designed to enrich events from the process of interest, in this case VBF Z production. Specifically, the signal region requires $|\Delta\eta_{jj}| > 1.5$ to select events with jets widely separated in pseudorapidity, 0 b-tagged jets to suppress top-quark backgrounds, and $|M_Z - M_{ll}| < 15$ GeV to select dilepton pairs consistent with the Z boson mass. These criteria maximize the purity of the VBF Z sample while minimizing contamination from QCD and top-quark backgrounds

Control region

Control regions are constructed to estimate and constrain the normalization of major backgrounds, specifically Drell-Yan and $t\bar{t}$, using dedicated selections summarised in Table 6.2. The Drell-Yan control region uses $|\Delta\eta_{jj}| < 1.5$ and 0 b-tagged jets, ensuring a sample enriched in Drell-Yan events while maintaining kinematic similarity to the signal region. The $t\bar{t}$ control region requires $|\Delta\eta_{jj}| > 1.5$, at least one b-tagged jet with medium working point, and $|M_Z - M_{ll}| > 15$ GeV, selecting events consistent with top-quark pair production.

A shared normalization parameter μ_{MC} (implemented as a `rateParam` in the Combine datacard) is introduced separately for each major background process (Drell-Yan and $t\bar{t}$). This parameter modifies the background component $b_{i,j}$ within $\lambda_i(\mu, \mathbf{\Theta})$ from Eq. (6.10), scaling the nominal Monte Carlo predictions N_{MC}^{SR} and N_{MC}^{CR} consistently across the signal region (SR) and its corresponding control region (CR).

Specifically, for these backgrounds the expected yield contribution becomes:

$$b^{\mathrm{SR}}_{MC,i} = \mu_{MC} \cdot N^{\mathrm{SR}}_{MC,i}, \quad b^{\mathrm{CR}}_{MC,i} = \mu_{MC} \cdot N^{\mathrm{CR}}_{MC,i} \ , \tag{6.16}$$

where $N^{\mathrm{SR/CR}}_{MC,i}$ are the *a priori* Monte Carlo event yields in bin i. Data observed in the control region constrains the best-fit value $\hat{\mu}_{MC}$ (typically $\hat{\mu}_{MC} \approx 1.0$) through the simultaneous likelihood fit of Eq. (6.10), thereby transferring the data-driven normalization to the signal region prediction.

Signal region	DY control region	$t\bar{t}$ control region
$\lvert\Delta\eta_{jj}\rvert > 1.5$ 0 b-tag jets $\lvert M_Z - M_{ll}\rvert < 15$ GeV	$\lvert\Delta\eta_{jj}\rvert < 1.5$ 0 b-tag jets $\lvert M_Z - M_{ll}\rvert < 15$ GeV	$\lvert\Delta\eta_{jj}\rvert > 1.5$ At least 1 b-tagged jet with medium WP $\lvert M_Z - M_{ll}\rvert > 15$ GeV

Table 6.2: Signal region and control region definitions at detector level

Particle level definition

Particle level corresponds to the stage in a simulation where all final-state particles are considered after the completion of parton showering, hadronisation, and the decays of short-lived particles, but prior to detector simulation. At this level, all stable or long-lived particles are included, while unstable particles are replaced by their respective decay products.

Particle-level definitions enable direct, model-independent comparisons between unfolded experimental measurements and theoretical predictions or particle-level Monte Carlo simulations.

6.6 Uncertainties

This section discusses the experimental uncertainties of the measurement and the theoretical uncertainties of the predictions.

During the various periods of Run 2 data-taking, the detector conditions evolved, requiring many corrections to be determined independently for each era. Consequently, certain systematic uncertainties must be treated as uncorrelated between different data-taking years.

6.6.1 Experimental uncertainties

The following sources of experimental uncertainty are considered:

- **Electron efficiencies:** Electron identification, reconstruction, and trigger requirements reduce the overall signal efficiency and introduce systematic uncertainties. The identification and reconstruction uncertainties are evaluated centrally [210], with the former contributing about 10% in the tBDT distribution and the latter remaining below 0.5%.

  Trigger scale factors and uncertainties are derived within each analysis and subsequently centrally validated [211]. For the present measurement, we adopt the trigger scale factors and uncertainties established in this previous analysis [212], which employ identical offline selection and ID criteria. This procedure results in a trigger uncertainty of roughly 5% in the tBDT observable, consistent with efficiencies obtained in comparable studies. Identification and isolation uncertainties are treated as correlated across data-taking years, while trigger uncertainties are considered uncorrelated between 2016, 2017, and 2018.

- **Muon efficiencies:** Muon identification, isolation, tracking and trigger implementation reduce signal efficiency and thus introduce systematic uncertainties, which are evaluated centrally [213]. Each of these sources is split into statistical and systematic components: The statistical components are uncorrelated across 2018, 2017, and 2016, while the systematic ones are correlated across years, except for the trigger, which is uncorrelated. In the tBDT observable, these uncertainties amount to approximately 0.5 % each mainly in the low p_T^{ll} region.

  Muon momenta are corrected using the Rochester calibration, with associated uncertainties below 1%. Their impact on this measurement is therefore negligible, and no systematic uncertainty from the muon momentum scale is assigned.

- **Jet Energy Scale:** As discussed in Section 5.2, jet energy scale corrections are derived in a sequential manner. This procedure naturally introduces multiple sources of systematic uncertainty, which must be retained to preserve the full statistical information of the corrections [214]. Following the official recommendations the present analysis employs the standard reduced set of 11 uncertainty sources [215]. These sources are defined to be mutually uncorrelated. These sources are:

  - *Correlated across years:* Absolute, BBEC1, EC2, FlavorQCD, HF, RelativeBal.
  - *Year-specific:* Absolute{2018,2017,2016}, BBEC1{2018,2017,2016}, EC2{2018,2017,2016}, HF{2018,2017,2016}, RelativeSample{2018,2017,2016}.

  Due to its shape-nature, source-by-source JES-aware training was required to implement in the BDT to properly account for the JES uncertainty behaviour. In some cases, both the "Up" and "Down" variations were observed to be either higher or lower than the nominal value, which is commonly referred to as "one-sided behaviour". This has previously been observed in other CMS analyses, and in the current measurement, it was observed both at detector and tBDT level, therefore, after inspecting them, JES uncertainties were symmetrised keeping the original Up variation and re-defining the "Down" variation as the "Up" variation with opposite sign. Due to the non-uniform JES behaviour observed in the tBDT observable, a smoothing technique was implemented based on 3-bin interpolation. In the current measurement this uncertainty represents the largest contribution to the total uncertainty, amounting to approximately 50% in the high tBDT region for some background processes.

- **Jet Energy Resolution:** The jet energy resolution uncertainty is evaluated by varying the JER smearing scale factors by one standard deviation and recomputing the jet four-momentum. This variation accounts for effects such as initial and final tate radiation modeling, differences between Monte Carlo (MC) tunes, and non-Gaussian resolution tails. The scale factors and uncertainties are centrally provided [216], and are treated as uncorrelated between the 2018, 2017, 2016 preVFP, and 2016 postVFP datasets. This uncertainty significantly affects the measurement, contributing about 15% across most processes in the tBDT distribution, and reaching up to 40% in certain migration effects and low-statistics backgrounds.

- **B-tagging:** The uncertainty associated with the b-tagging scale factors includes contributions from several sources: the statistical precision of the calibration, the choice of method, modelling-related effects (such as generator, parton shower, fragmentation, and heavy-flavour composition among others), detector-related effects, and the extrapolation to high jet transverse momentum. These contributions are combined into an overall uncertainty assigned to each b-tagging working point, which depends on the strictness of the identification criteria. The corresponding variations of the scale factors for bottom, charm, and light-flavour jets are provided centrally [196]. This uncertainty is divided into correlated and uncorrelated components across 2018, 2017, 2016preVFP, and 2016postVFP. The effect of this uncertainty is negligible in both the signal and DY control region since no b-tagged jets are being accepted in their kinematic regions, opposite to the $t\bar{t}$ control region, where this uncertainty amounts to 20% in the low $\Delta\phi_{Z,jj}$ range (see distribution c) in Figure 7.6.

- **Luminosity:** The total simulated yield is scaled to the integrated luminosity of the recorded data, which also enters directly in the cross-section determination. The associated luminosity uncertainties are provided centrally by the Lumi POG and are decomposed into three components [217]:
  - Year-specific: 2.5% (2018), 2.3% (2017), 1.2% (2016).
  - Correlated between 2017 and 2018: 1.6%.
  - Correlated across the full Run 2: 1.6%.

- **L1 pre-firing:** In 2017 and 2016, the gradual timing shift of the ECAL was not correctly propagated to the Level-1 trigger primitives. As a result, uncertainties due to ECAL prefiring are included and treated as uncorrelated across these years. The contribution from muon prefiring is negligible and therefore not considered [163]. The effect of this uncertainty is generally small, becoming relevant only in the low-p_T^{ll} region, where it amounts to about 3% in the electron channel.

- **Pileup (PU):** The pileup profile is primarily affected by the uncertainty in the total inelastic proton-proton cross section (Eq. 5.22). Profiles corresponding to upward and downward variations of the cross section are derived and used to obtain the respective variations in the pileup reweighting. The resulting shifts are applied to evaluate the associated systematic uncertainty. This uncertainty are treated as uncorrelated across the 2018 [218], 2017 [219], and 2016 [220] datasets. The effect of this uncertainty is generally small, becoming relevant only in the low-p_T^{ll} region, where it amounts to about 4%.

- **MC statistics:** Statistical uncertainties from the finite size of the MC samples are included for all signal and background processes and are evaluated bin by bin.

6.6.2 Theoretical uncertainties

The following sources of theoretical uncertainty are considered in the analysis:

- **Cross-section values:** EW Zjj (5%), DY (5%), $t\bar{t}$ (7%), single top (4.4%), and diboson (0.25%) [221]. Single top and diboson uncertainties are grouped in the likelihood fit and treated as correlated across years, while the DY and $t\bar{t}$ uncertainties are constrained through their respective control regions. For the EW Zjj and DY processes, a conservative 5% uncertainty is applied in place of the nominal 0.28% and 0.27% given by the `XSecAnalyzer` tool respectively, in order to prevent overconstraining the signal in the relevant phase space. These uncertainties are derived mainly from the simulated sample statistical uncertainty provided by the MC generator.

- **Factorization and renormalization scales at matrix-element level:** The factorisation scale, μ_F, separates soft from hard contributions: processes below μ_F are absorbed into the PDFs, while those above are treated perturbatively in the hard scattering. In this way, μ_F acts as an infrared cut-off, removing divergences from the hard-function calculation. This uncertainty is evaluated by fixing μ_R and varying μ_F by a factor of 2 for the "Up" variation and 0.5 for the "Down" variation.

  The renormalisation scale, μ_R, enters the hard-function calculation to absorb ultraviolet divergences from virtual corrections at large momenta. While physical cross sections should be independent of the arbitrary choices of μ_F and μ_R, the truncation of the perturbative α_S expansion leaves a residual dependence on these scales. This uncertainty is evaluated by fixing μ_F and varying μ_R by a factor of 2 for the "Up" variation and 0.5 for the "Down" variation.

  Unlike the conventional approach that combines 7- or 9-point scale variations into an envelope, the CMS analysis strategy treats the factorization and renormalization scale variations as independent nuisance parameters, meaning independent uncertainties. This methodology enables a more transparent template fit and treats the scale-related uncertainties as uncorrelated sources, thereby improving the consistency of the systematic uncertainty treatment across different data-taking periods and processes.

  These uncertainties are treated as uncorrelated across processes and correlated across years and are lower than 4%.

- **Parton showering model:** Uncertainties from parton showering are modelled with two nuisance parameters: initial-state radiation (ISR), describing radiation before the hard scattering, and final-state radiation (FSR) (see Section 1.2), describing radiation after the interaction [48]. They are derived by varying the renormalization scale for QCD emissions in ISR and FSR, effectively varying the value of α_S, in the case of ISR it is varied by a factor of 2 and in the case of the FSR it is varied by a factor of 0.5, both implemented in Pythia8. These are provided centrally in simulation samples, and are treated as uncorrelated between ISR and FSR, uncorrelated across processes and correlated across years. Similarly to the Factorization and renormalization scales uncertainties, no 7- or 9-point variation is implemented, therefore these uncertainties are treated as 2 separate nuisance parameters instead and amount to 1% approximately.

- **PDF and $\boldsymbol{\alpha_S}$:** PDF uncertainties are evaluated using the Hessian method [222] applied to the 100 eigenvectors of the `NNPDF31_nnlo` sets. These variations are treated as uncorrelated between processes but correlated across data-taking years.

  For the strong coupling, α_S, “Up” and “Down” event-weight variations are applied and assumed to be correlated both across processes and years [223].

  These uncertainties are below 8%.

In this analysis, acceptance-related uncertainties, including those from theoretical modelling of signal and background processes, are included in the systematic uncertainty. In contrast, the overall cross-section uncertainty from higher-order corrections or scale variations is excluded, as it reflects theoretical rather than experimental effects. This separation ensures a clear distinction between the measurement and its theoretical interpretation, enabling consistent comparison with different models.

In the statistical combination of the dielectron and dimuon channels, the correlation model for systematic uncertainties is implemented within the `Combine` framework as previously explained in Sections 6.5.1 and 6.5.2. The nuisance parameters associated with lepton-related uncertainties, including identification, reconstruction, and trigger efficiencies, are defined independently for each channel to account for their uncorrelated nature. Similarly, year-specific nuisance parameters, such as pileup and MC statistical uncertainties, are defined independently for each year as well. Conversely, all remaining sources of systematic uncertainty common across years and channels are represented by common nuisance parameters shared between the two channels and between all the data taking eras. This configuration ensures that `Combine` correctly propagates correlated effects in the simultaneous likelihood fit while maintaining independent treatment of channel-specific uncertainties.
A comprehensive overview of the uncertainty treatment is provided in Table 6.3. In this table, the notation is defined as follows: C:Y+P denotes uncertainties correlated across both years and processes; C:Y indicates uncertainties correlated across years but uncorrelated across processes and C:P refers to uncertainties correlated across processes but uncorrelated across years.

Uncertainty	**Type**	**Correlation**
PDF	Shape	C:Y
α_S	Shape	C:Y+P
μ_F	Shape	C:Y
μ_R	Shape	C:Y
isr	Shape	C:Y
fsr	Shape	C:Y
Cross-section normalization	Rate	C:Y
Luminosity{Year}	Rate	C:P
Luminosity{2018, 2017}	Rate	C:Y+P
Luminosity{Full Run 2}	Rate	C:Y+P
PU	Shape	C:P
L1 pre-firing	Shape	C:P
Muon ID, ISO, Reco: Stat	Shape	C:P
Muon ID, ISO, Reco: Syst	Shape	C:Y+P
Muon trigger	Shape	C:P
Electron ID, Reco	Shape	C:Y+P
Electron trigger	Shape	C:P
B-tagging{Year}	Shape	C:P
B-tagging-light{Year}	Shape	C:P
B-tagging	Shape	C:Y+P
B-tagging-light	Shape	C:Y+P
JES: Absolute, BBEC1, EC2, FlavorQCD, HF, RelativeBal	Shape	C:Y+P
JES: Absolute{Year}, BBEC1{Year}, EC2{Year}, HF{Year}, RelativeSample{Year}	Shape	C:P
JER	Shape	C:P

Table 6.3: Uncertainties sources, type and correlation across processes and years.

Chapter 7

Results

The current chapter presents both the fiducial and differential cross-section measurements of Z-boson production in association with dijets, encompassing the inclusive (QCD Zjj + EW Zjj) and electroweak (EW Zjj) components as a function of the dilepton transverse momentum, p_T^{ll}. Section 7.1 details the inclusive Zjj measurement using 59.83 fb^{-1} of 13 TeV proton-proton collision data recorded by the CMS experiment in 2018, while Section 7.2 presents the exclusive electroweak Zjj measurement, which constitutes the main result of this analysis, using the full Run 2 dataset corresponding to 138 fb^{-1} at $\sqrt{s} = 13$ TeV. Both analyses combine the $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$ channels and are performed within a common phase space.

7.1 Inclusive Zjj measurement

The measurement of the EW Zjj process is technically challenging due to its distinctive topology and the overwhelming Drell–Yan background. Therefore, first, a fiducial cross-section measurement of inclusive Z-boson + 2 jets production (EW Zjj + QCD Zjj) is first performed within the phase space defined in Section 6.2.1. For this measurement, due to the lack of statistical constraints, no machine learning techniques are needed to optimise signal purity. Therefore the cross section can extracted from any observable, in this case, for simplicity, the p_T^{ll} distribution is selected, and a maximum likelihood fit to it is performed independently for each channel and combined thereafter. This result was derived using only the 2018 dataset, corresponding to an integrated luminosity of 59.83 fb^{-1} at a centre-of-mass energy of 13 TeV. Since statistical precision was not a limiting factor, analysing the full Run 2 dataset was not required. In addition, this measurement serves mainly to illustrate the difference in magnitude between the inclusive Zjj and the electroweak Zjj measurements.

7.1.1 Choice of observable and binning

As previously discussed, the inclusive Z measurement is not performed by binning the tBDT distribution in p_T^{ll} intervals. Instead, the p_T^{ll} observable is directly binned according to the regions: $[0, 60, 110, 180, \infty)$ GeV. These p_T^{ll} regions are inherited from the EW Zjj measurement, where the tBDT distribution is divided into detector-level p_T^{ll} bins defined as $[0, 60, 110, 180, \infty)$, chosen to ensure that the significance of the electroweak Zjj signal,

$$\sigma_{EW\ Zjj} = \frac{N_{EW\ Zjj}}{\sqrt{N_{\text{bckg}}}},$$

exceeds 5σ in each bin. Correspondingly, the electroweak Zjj process is categorized into generator-level p_T^{ll} bins using the same boundaries. In the inclusive measurement, both the QCD (DY) and electroweak Zjj processes are classified into these bins, enabling the unfolding at particle level simultaneously in both. The p_T^{ll} is separated in different ranges allowing `Combine` to better characterise the background and further constrain the nuisance parameters, especially those whose behaviour modifies the overall distribution shape, deeply differ across p_T^{ll} regions or which greatly rely on the process statistics, as it is the case for jet energy corrections.
Using the p_T^{ll} observable directly instead of the tBDT distribution is feasible due to the high-statistics nature of the Drell–Yan process. The p_T^{ll} distribution is first examined within these regions prior to the maximum likelihood fit (referred to as the "pre-fit" stage). A minimal re-binning is then applied to ensure that the relative uncertainty remains below 30% approximately in the pre-fit and to reduce the number of bins for migrated processes, as shown in Appendix G. After performing the maximum likelihood fit, the post-fit distributions are examined, confirming a stable and well-behaved fit performance. This is illustrated in Figures 7.1 and 7.2, which display the p_T^{ll} distributions for the four p_T^{ll} regions in each channel for 2018. The upper panels show the stacked histograms of the signal and background contributions. Here, the label Z denotes the combined electroweak and QCD-induced Z+jj processes, TT corresponds to the top-antitop production, and *others* comprise the single top and diboson production channels. The lower panels present the post-fit ratio of data to simulation, where the grey band indicates the total uncertainty and the coloured band represents the statistical component only. These distributions are normalized to the expected event yields, computed from theoretical cross sections and integrated luminosity per year, and corrected for all weights and detector-level scale factors (as described in Detector-based criteria in Section 6.2.1).

7.1.2 Cross section for Inclusive Zjj

This section presents the inclusive cross section corrected to particle level corresponding to the fiducial phase space defined in Section 6.2.1 for the inclusive Z production.
In the dimuon channel, the signal strength (see Section 6.5) is measured to be:

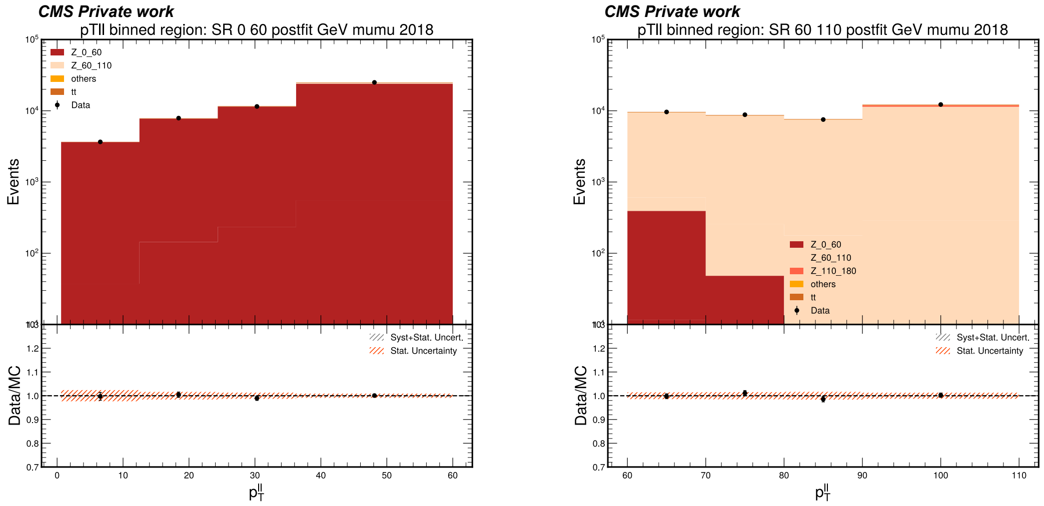
$$\mu_{Zjj}^{\mu\mu} = 1.091 \pm \binom{+0.080}{-0.074}_{\text{syst}} \pm \binom{+0.003}{-0.003}_{\text{stat}}$$

corresponding to a measured signal cross section:

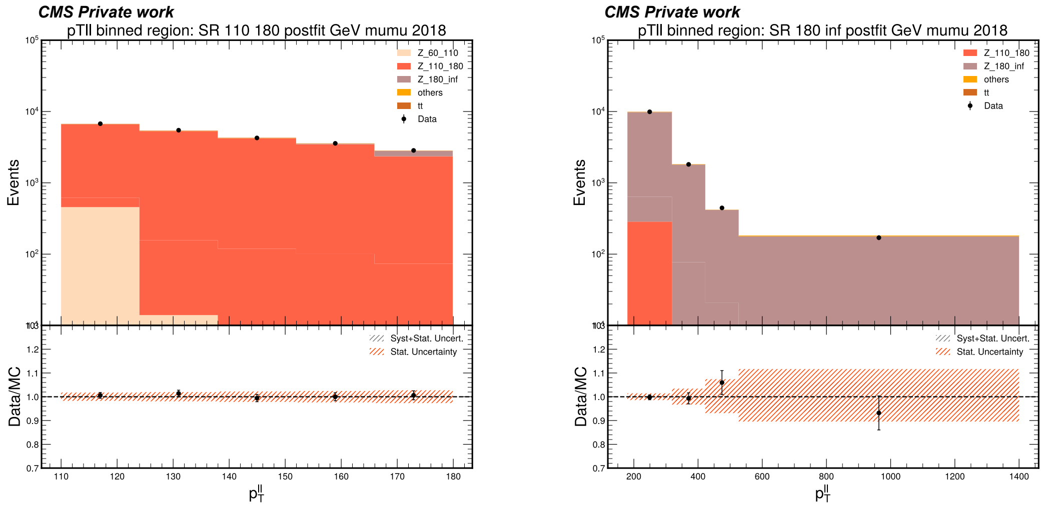
$$\sigma_{Zjj}^{\mu\mu} = 2.21 \times 10^{6} \pm \begin{pmatrix} +1.77\times10^{5} \\ -1.64\times10^{5} \end{pmatrix}_{\text{syst}} \pm \begin{pmatrix} +6.63\times10^{3} \\ -6.63\times10^{3} \end{pmatrix}_{\text{stat}} \text{ fb}$$

In the dielectron channel, the signal strength is measured to be:

$$\mu_{Zjj}^{ee} = 1.073 \pm \binom{0.094}{0.086}_{\text{syst}} \pm \binom{0.006}{0.006}_{\text{stat}}$$

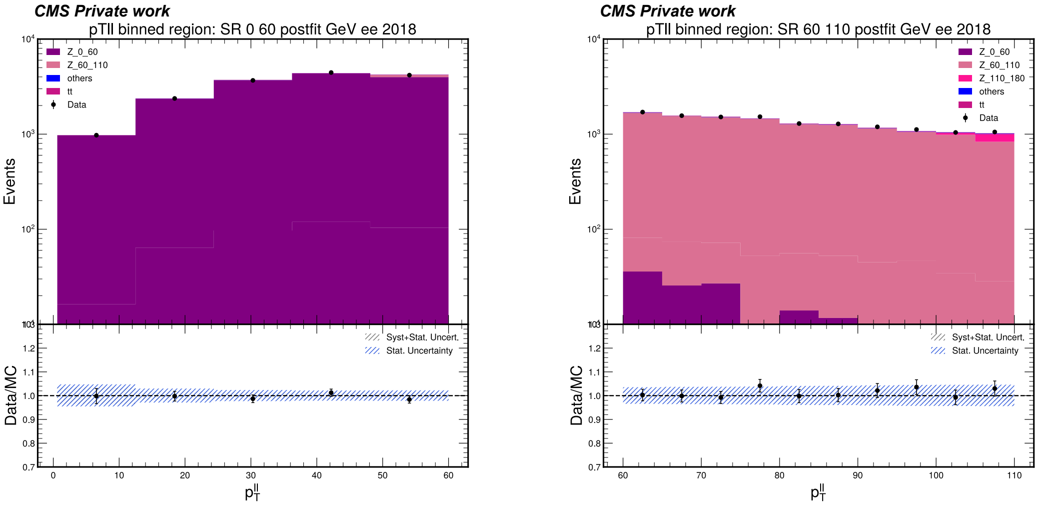


(a) $0 < p_T^{ll} < 60$ (b) $60 < p_T^{ll} < 110$

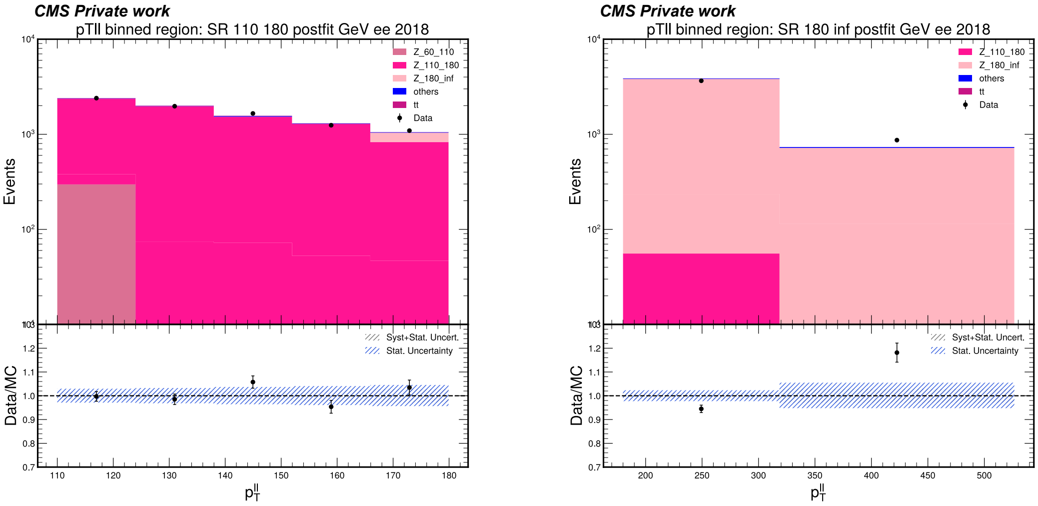


(c) $110 < p_T^{ll} < 180$ (d) $180 < p_T^{ll}$

Figure 7.1: Post-fit p_T^{ll} observable in different regions for the inclusive Zjj measurement in the $\mu\mu$ channel for 2018.



(a) $0 < p_T^{ll} < 60$ (b) $60 < p_T^{ll} < 110$



(c) $110 < p_T^{ll} < 180$ (d) $180 < p_T^{ll}$

Figure 7.2: Post-fit p_T^{ll} observable in different regions for the inclusive Zjj measurement in the *ee* channel for 2018.

corresponding to a measured signal cross section:

$$\sigma^{ee}_{Zjj} = 2.17 \times 10^6 \pm \begin{pmatrix} 2.04\times10^5 \\ 1.87\times10^5 \end{pmatrix}_{\text{syst}} \pm \begin{pmatrix} 1.31\times10^4 \\ 1.31\times10^4 \end{pmatrix}_{\text{stat}} \text{ fb}$$

From the combined fit of the two channels, the signal strength is measured to be:

$$\mu^{\mu\mu+ee}_{Zjj} = 1.013 \pm \begin{pmatrix} 0.067 \\ 0.064 \end{pmatrix}_{\text{syst}} \pm \begin{pmatrix} 0.002 \\ 0.002 \end{pmatrix}_{\text{stat}}$$

corresponding to a measured signal cross section:

$$\sigma^{\mu\mu+ee}_{Zjj} = 2.05 \times 10^6 \pm \begin{pmatrix} 1.64\times10^5 \\ 1.52\times10^5 \end{pmatrix}_{\text{syst}} \pm \begin{pmatrix} 4.11\times10^3 \\ 4.11\times10^3 \end{pmatrix}_{\text{stat}} \text{ fb}$$

The lower combined inclusive signal strength compared to the individual channels reflects the correlations and stability properties of the joint likelihood fit. In the combined fit, the higher statistical precision of the $\mu\mu$ channel constrains common nuisance parameters and stabilizes the less well-determined ee channel, thereby reducing fluctuations or upward biases that may occur when fitted independently. This interplay between channel correlations, parameter pulls, and overall fit stability leads to a statistically more robust and unbiased estimate of the signal strength, which can yield a lower combined value than those obtained from the separate channels.
Nevertheless, all the measured cross sections are in agreement with the SM prediction $\sigma_{SM}(Zjj) \approx \sigma_{SM}(EW\ Zjj) + \sigma_{SM}(QCD\ Zjj) = 2.02\times10^6 \pm \begin{pmatrix} 1.43\times10^5 \\ 1.43\times10^5 \end{pmatrix}$ fb modelled by `EWK_LLJJ`, which simulates the EW production of the Z^0 boson where it decays into a dilepton final state, with a dilepton mass threshold, $m_{ll} > 50$ GeV and a dijet mass threshold, $m_{jj} > 120$ GeV, and the `DYJetsToLL` which simulates the Drell-Yan production of the Z^0 boson decaying into a dilepton final state too with $m_{ll} > 50$ GeV. Both processes are simulated with `MadGraph5_aMC@NLO` [68], with the `EWK_LLJJ` sample generated at LO and the `DYJetsToLL` sample at NLO, both interfaced to `Pythia` [69].

Experiment	Year	$\mathcal{L}\,[\text{pb}^{-1}]$	Channel	$\sigma(Zjj)\,[10^6\text{ fb}]$
ATLAS no m_{jj} constraint	2016	81	$\ell^+\ell^-$	$1.981 \pm 0.007_{\text{stat}} \pm 0.038_{\text{syst}} \pm 0.042_{\text{lumi}}$
CMS no m_{jj} constraint	2016	$2.3 \cdot 10^{-3}$	$\mu^+\mu^-$	$1.870 \pm 0.002_{\text{stat}} \pm 0.035_{\text{syst}} \pm 0.051_{\text{lumi}}$
CMS no m_{jj} constraint	2017	206	$\ell^+\ell^-$	$1.952 \pm 0.004_{\text{stat}} \pm 0.018_{\text{syst}} \pm 0.045_{\text{lumi}}$
CMS $m_{jj} > 200$ GeV	**2018**	$\mathbf{59.83 \cdot 10^{-3}}$	$\ell^+\ell^-$	$\mathbf{2.05 \pm^{0.0041}_{0.0041\,\text{stat}} \pm^{0.164}_{0.152\,\text{syst}}}$

Table 7.1: Summary of selected measurements of the inclusive $pp \to Z + X$ production cross section with $Z \to l^+l^-$ at $\sqrt{s} = 13$ TeV [224–226] with the current measurement shown in the last row.

Several measurements of the process $pp \to Z + X$ with $Z \to l^+l^-$ have been performed at $\sqrt{s} = 13$ TeV, as summarised in Table 7.1. The CMS and ATLAS inclusive measurements, based on integrated luminosities of 206 and 81 pb^{-1}, respectively, serve as illustrative examples of results obtained in similar analyses under different phase-space definitions. Although these analyses employ different selections and methodologies, their results together provide valuable context for the present study and illustrate the diversity of experimental approaches to Zjj measurements at 13 TeV.

7.1.3 Differential cross section for Inclusive Zjj

This section presents the differential cross sections corrected to the particle level, corresponding to the fiducial phase space defined in Section 6.2.1 for the inclusive Z boson production. Fig. 7.3 shows the differential cross section for each channel separately and Fig. 7.4 for the combination of both. The upper panel illustrates the differential cross section as a function of four distinct p_T^{ll} intervals. The black markers represent the measured values, where the vertical error bars indicate the systematic uncertainties, which also account for the theoretical modelling of the signal acceptance. The coloured shaded band denotes the total uncertainty, combining both statistical and systematic components. For comparison, the Standard Model prediction is shown by the gray markers, obtained from the `EWK_LLJJ` and `DYJetsToLL` simulated samples as described in the previous chapter. The corresponding theoretical cross-section uncertainty is represented by the gray dashed band. The lower panel displays the ratio between the measured cross section and the Standard Model prediction, where the coloured band includes both statistical and systematic uncertainties, and the gray band represents the statistical component only. Observing the highly suppressed statistical uncertainty highlights the significant impact of systematic uncertainties, as well as the dominant contribution from the QCD process.

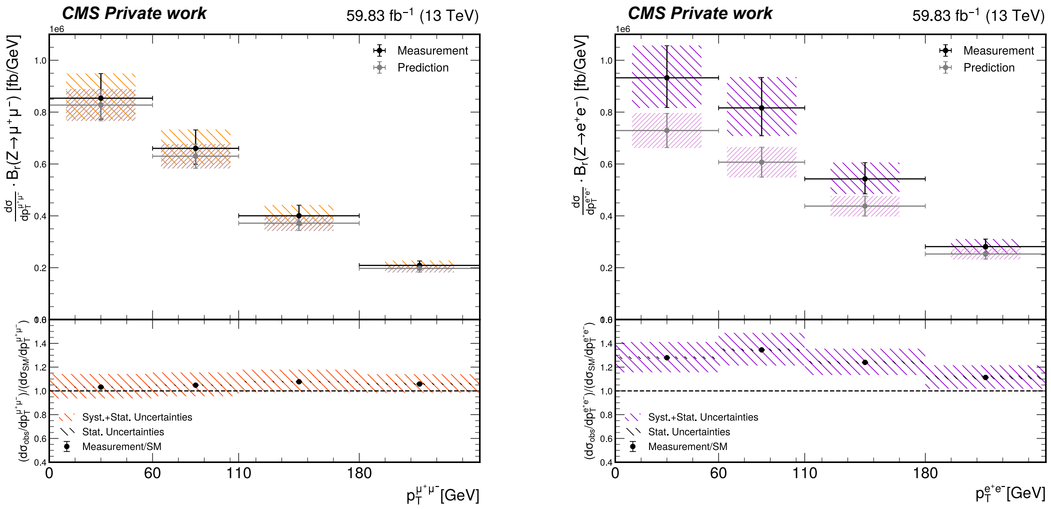


Figure 7.3: Differential cross section for inclusive Zjj with respect the Z-boson p_T: $\mu\mu$ channel (left), ee channel (right), 2018.

Figs. 7.3 and 7.4 demonstrate good agreement between data and simulation for both the individual channels and their combination. The agreement is notably better in the $\mu\mu$ channel than in the ee channel due to the higher statistics in the former, which stabilises the maximum likelihood fit, an important factor given the relatively small contribution from the EW Zjj component. This behaviour is particularly evident in the low-$p_{\mathrm{T},\ell\ell}$ region due to how `Combine` performs multi-signal fits and the physics nature of the processes involved.

In this analysis, the inclusive measurement is defined by the combined contribution of the Drell-Yan (DY) and electroweak (EW) processes, which are treated jointly as signal despite their distinct physical behaviours. This can result in unphysical

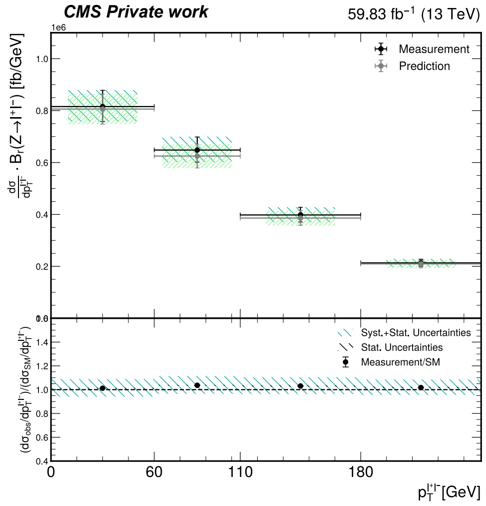


Figure 7.4: Differential cross section for inclusive Zjj with respect the Z-boson p_T in the combined $\mu\mu$ and ee channels, 2018.

event migrations between p_T bins and between the EW and DY components of the signal, manifesting as a higher differential cross section in the ee channel relative to the $\mu\mu$ channel, even when the inclusive signal strength is smaller in ee. The fit stability can be mitigated when both channels and signal components have adequate statistics and share a correlated, constrained set of nuisance parameters. However, in regions with limited statistics, small data fluctuations can lead to large deviations in the smaller signal, as observed in the ee channel.
This interplay is also visible in the discrepancy between data and simulation in Figure 6.18 for the ee channel in 2018, as well as in the limited EW Zjj statistics within the first two p_T^{ll} bins (0–60 GeV and 60–110 GeV), shown in Appendix F through the $p_{\mathrm{T},\ell\ell}$-binned tBDT distributions. These deviations are therefore expected. When both channels are combined, the maximum likelihood fit constrains the less stable ee channel using information from the more precise $\mu\mu$ channel, mitigating overestimation in the ee component and yielding a robust and well-balanced agreement between data and theoretical expectation.

7.2 Electroweak Zjj measurement

In this section the inclusive and differential cross sections corrected to particle level corresponding to the fiducial phase space defined in Section 7.2.1 are presented for the EW Zjj process, defined as the t-channel exchange of a weak vector boson, including Z-boson production via vector boson fusion (VBF), Z-boson bremsstrahlung and multi-peripheral production.

7.2.1 Choice of observable and binning

The differential cross-section of the EW Zjj process is measured as a function of the dilepton transverse momentum, p_T^{ll}. As outlined in Chapter 2, this observable is sensitive to QCD dynamics, and experimentally, the dilepton channel provides a clean and well-defined signature, making it a robust and precise observable for phenomenological studies in high-energy physics.

Signal region

As introduced in the previous section, the fitting variable, the tBDT distribution, is divided into detector-level p_T^{ll} bins defined as $[0, 60, 110, 180, \infty)$, chosen to ensure that the significance of the electroweak Zjj signal,

$$\sigma_{EW\ Zjj} = \frac{N_{EW\ Zjj}}{\sqrt{N_{\mathrm{bckg}}}},$$

where N refers to the number of events, exceeds 5σ in each bin. Correspondingly, the electroweak Zjj process is also categorized into generator-level p_T^{ll} bins with the same boundaries.

The binning of the tBDT distribution is determined so as to optimally retain the statistical information required for the maximum likelihood fit. This choice serves to minimize statistical uncertainties while preventing an artificial enhancement of the JEC uncertainties. The sensitivity of the tBDT observable to JEC variations is particularly pronounced due to two factors. Firstly, JEC uncertainties are of shape type, modifying the global form of jet–kinematic distributions rather than entering as overall weight variations. Secondly, as the majority of the BDT input variables are functions of jet kinematics, JEC variations propagate non-linearly into the BDT response. This induces distortions in the BDT output that are more substantial than those affecting individual detector-level jet observables.

As a result, the following criterion is applied to determine the optimal tBDT binning, inspired by the concept of the Area Under the Curve (AUC), which is widely used in machine learning to evaluate model performance and was previously introduced in Section 6.4.2:

1. The tBDT distributions are examined prior to the application of the maximum-likelihood fit (referred to as “pre-fit”).

2. To evaluate the efficiency of the tBDT with respect to the EW Zjj process, the AUC is calculated. From this, the fraction of the AUC where $\sigma_{\mathrm{EW\,Zjj}}$ lies below 4σ is extracted:

   $$\mathcal{F}_{\mathrm{EW\,Zjj}}\,[\%] \;=\; \frac{AUC\big(\sigma_{\mathrm{EW\,Zjj}} < 4\sigma\big)}{AUC\big(\sigma_{\mathrm{EW\,Zjj}}\big)} \times 100\,. \tag{7.1}$$

   An extreme value, i.e. $\mathcal{F}_{\mathrm{EW\,Zjj}} \to 100\%$, indicates that the current tBDT binning is not optimal. Consequently, in practice, whenever $\mathcal{F}_{\mathrm{EW\,Zjj}} > 90\%$ and the number of bins exceeds three, the tBDT distribution is reduced to a single bin. This situation occurs only in very specific and rare cases, which in this measurement only arises in the ee channel for 2018 in the $p_T^{\ell\ell}$-binned tBDT distribution within 0 GeV $< tBDT <$ 60 GeV.

3. If $\mathcal{F}_{\mathrm{EW\,Zjj}} < 90\%$, a rebinning with minimal modifications is performed to ensure that the total uncertainty remains below 60% in the least populated bin of the given tBDT region (pre-fit), while also achieving a smooth behaviour of the total uncertainty band. Minimal modifications are defined as:

  - the final number of bins must not be reduced below half the initial number of bins if the initial count was even, or below $(N_{\text{init}} - 1)/2$ if the initial count was odd, for cases with $\mathcal{F}_{\text{EW Zjj}} < 70\%$;
  - the final number of bins must be at least two when the initial configuration corresponds to $\mathcal{F}_{\text{EW Zjj}} > 70\%$.

4. Finally, $\mathcal{F}_{\text{EW Zjj}}$ is recalculated on the re-binned tBDT distributions to verify that it remains below 90% in all cases where the number of bins exceeds three.

The complete set of tBDT distributions for both channels, covering all eras together with their respective binning selection studies, is provided in the Appendix F.

Control regions
As described in Section 6.5.2, control regions are employed to constrain the normalization of the Drell-Yan and $t\bar{t}$ backgrounds during the unfolding procedure. Their impact is evaluated using the $\Delta\Phi_{Z,jj}$ observable in each control region, chosen for its high separation capability due to the different shape behaviour between the EW Zjj signal and the background processes, as quantified by the F-score in Section 6.4.3. This observable is not divided in detector-level p_T^{ll} bins and it maintains the same binning in $\Delta\Phi_{Z,jj}$ in both channels through all eras:
$[0.50, 0.84, 1.18, 1.51, 1.85, 2.19, 2.53, 2.86, 3.20]$ (see Figs. 7.5 and 7.6).
These distributions are normalized to the expected event yields, computed from theoretical cross sections and integrated luminosity per year, and corrected for all weights and detector-level scale factors (as described in Detector-based criteria in Section 6.2.1).

7.2.2 Cross section for Electroweak Zjj

This section presents the inclusive cross section corrected to particle level corresponding to the fiducial phase space defined in Section 6.2.1 for the EW Zjj.
In the dimuon channel, the signal strength is measured to be:

$$\mu^{\mu\mu}_{\text{EW Zjj}} = 1.06 \pm \binom{0.09}{0.09}_{\text{syst}} \pm \binom{0.04}{0.04}_{\text{stat}}$$

corresponding to a measured signal cross section:

$$\sigma^{\mu\mu}_{\text{EW Zjj}} = 609.1 \pm \binom{56.6}{54.8}_{\text{syst}} \pm \binom{21.3}{21.3}_{\text{stat}} \text{ fb}$$

In the dielectron channel, the signal strength is measured to be:

$$\mu^{ee}_{\text{EW Zjj}} = 1.01 \pm \binom{0.12}{0.11}_{\text{syst}} \pm \binom{0.05}{0.05}_{\text{stat}}$$

corresponding to a measured signal cross section:

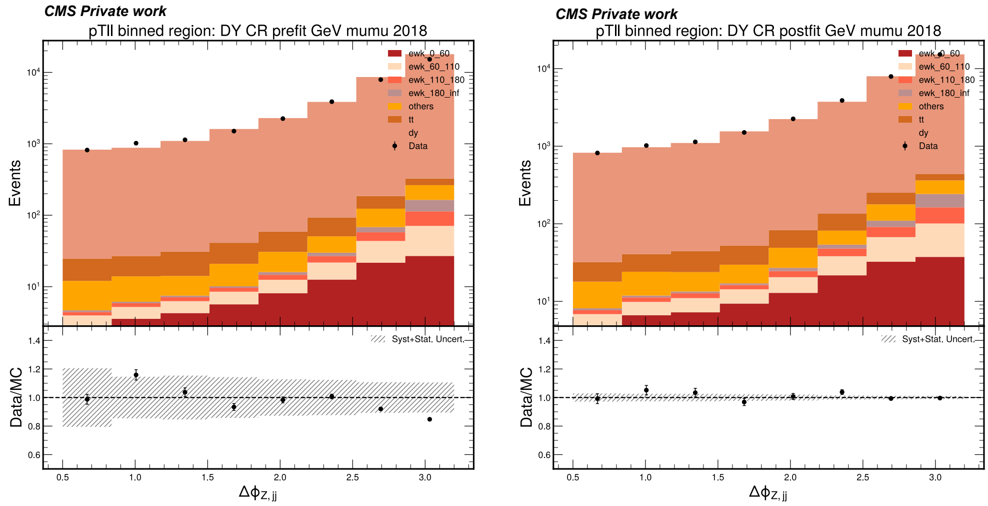
$$\sigma^{ee}_{\text{EW Zjj}} = 579.3 \pm \binom{67.2}{63.1}_{\text{syst}} \pm \binom{29.0}{29.0}_{\text{stat}} \text{ fb}$$

From the combined fit of the two channels, the signal strength is measured to be:

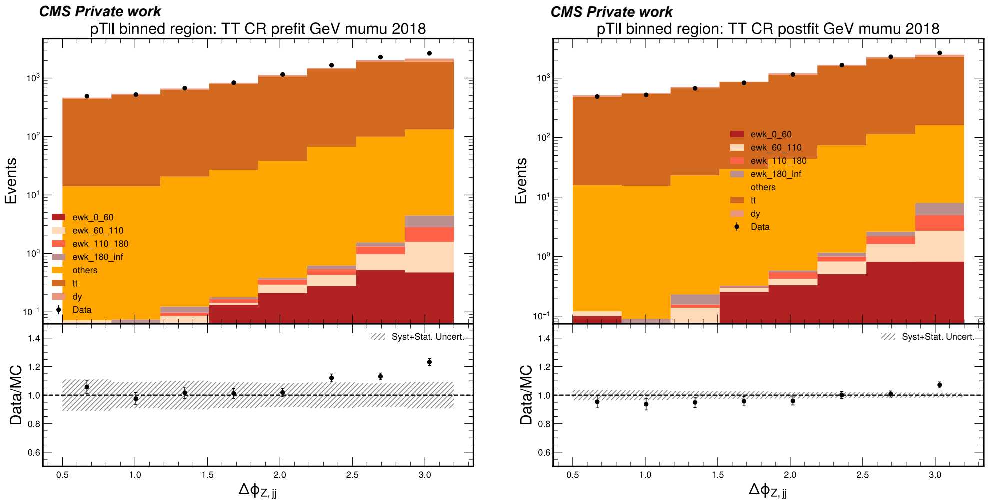
$$\mu^{\mu\mu+ee}_{\text{EW Zjj}} = 1.04 \pm \binom{0.06}{0.06}_{\text{syst}} \pm \binom{0.03}{0.03}_{\text{stat}}$$

corresponding to a measured signal cross section:

$$\sigma^{\mu\mu+ee}_{\text{EW Zjj}} = 596.5 \pm \binom{35.8}{34.6}_{\text{syst}} \pm \binom{17.3}{17.3}_{\text{stat}} \text{ fb}$$



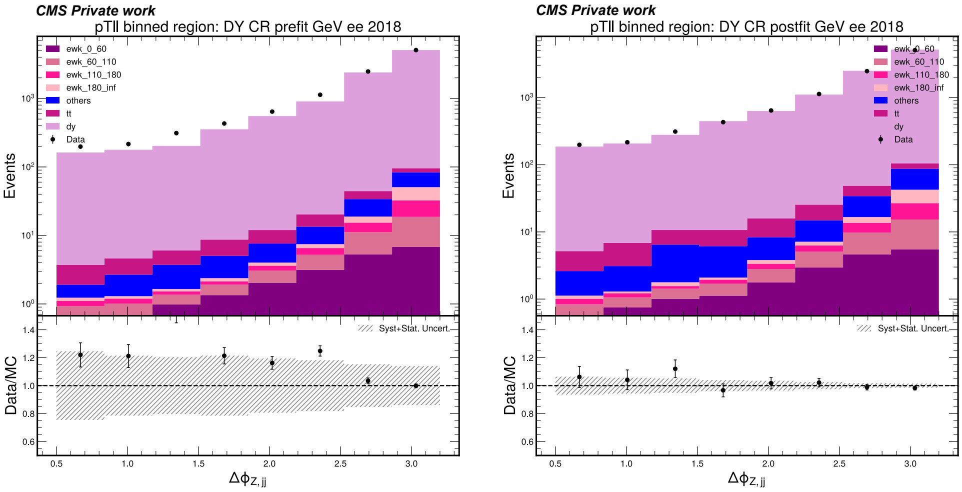
(a) DY CR pre-fit (b) DY CR post-fit



(c) TT CR pre-fit (d) TT CR post-fit

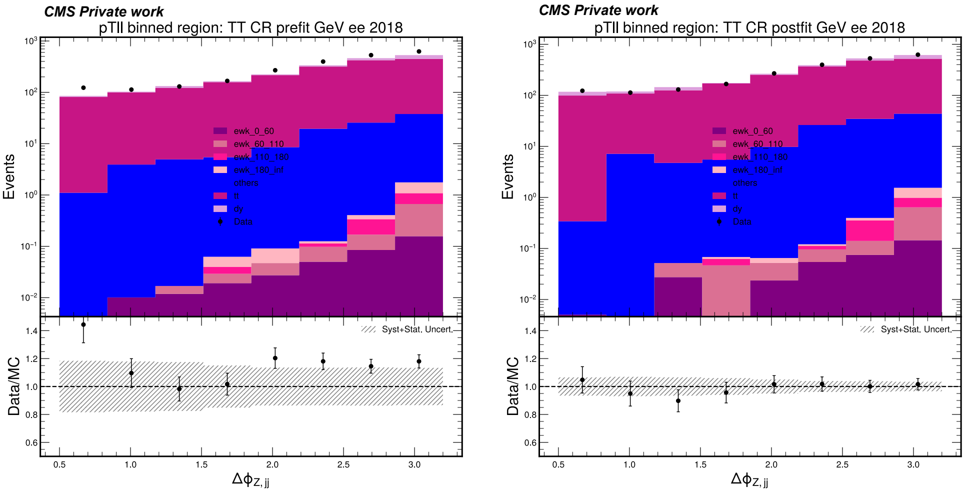
(e) MC labels

Figure 7.5: $\Delta\Phi_{Z,jj}$ observable for $\mu\mu$ channel, 2018. Top row: Drell-Yan CR pre- and post-fit. Bottom row: $t\bar{t}$ CR pre- and post-fit. Labels are defined as: `ewk_0_60`: EW Zjj signal process for particle level $0 < p_{Tll} < 60$ GeV bin, `ewk_60_110`: EW Zjj signal process for particle level $60 < p_{Tll} < 110$ GeV bin, `ewk_110_180`: EW Zjj signal process for particle level $110 < p_{Tll} < 180$ GeV bin, `ewk_180_inf`: EW Zjj signal process for particle level $180 < p_{Tll}$ GeV bin, `others`: single top and diboson processes $(W^+W^-, W^\pm Z^0, Z^0Z^0)$, `tt`: top-antitop processes and `dy`: Drell-Yan process.



(a) DY CR pre-fit

(b) DY CR post-fit



(c) TT CR pre-fit

(d) TT CR post-fit

(e) MC labels

Figure 7.6: $\Delta\Phi_{Z,jj}$ observable for ee channel, 2018. Top row: Drell-Yan CR pre- and post-fit. Bottom row: $t\bar{t}$ CR pre- and post-fit. Labels are defined as: `ewk_0_60`: EW Zjj signal process for particle level $0 < p_{Tll} < 60$ GeV bin, `ewk_60_110`: EW Zjj signal process for particle level $60 < p_{Tll} < 110$ GeV bin, `ewk_110_180`: EW Zjj signal process for particle level $110 < p_{Tll} < 180$ GeV bin, `ewk_180_inf`: EW Zjj signal process for particle level $180 < p_{Tll}$ GeV bin, `others`: single top and diboson processes $(W^+W^-, W^\pm Z^0, Z^0Z^0)$, `tt`: top-antitop processes and `dy`: Drell-Yan process.

As a result, all the measured cross sections are in agreement with the SM prediction $\sigma_{SM}(EW\ Zjj) = 573.0 \pm 28.7$ fb modelled by `EWK_LLJJ`. This simulates the EW production of the Z^0 boson decaying into a dilepton final state with $m_{ll} > 50$ GeV and a dijet mass threshold $m_{jj} > 120$ GeV. The process is generated with `MadGraph5_aMC@NLO` [68] at LO and interfaced to `Pythia` [69].

Furthermore, the current signal strengths are compatible with the ones in the measurement reported by the publication using the 36.3 fb^{-1} of data collected in 2016 by the CMS Collaboration with the same m_{jj} threshold requirement [205]:

$$\mu_{\text{EW Zjj}}^{\mu\mu,\text{CMS}-2016} = 0.97 \pm 0.11_{\text{syst}}\ \pm\ 0.04_{\text{stat}}$$

$$\mu_{\text{EW Zjj}}^{ee,\text{CMS}-2016} = 0.96 \pm 0.13_{\text{syst}}\ \pm\ 0.06_{\text{stat}}$$

$$\mu_{\text{EW Zjj}}^{\mu\mu+ee,\text{CMS}-2016} = 0.98 \pm 0.10_{\text{syst}}\ \pm\ 0.04_{\text{stat}}$$

A direct extrapolation of the 2016 cross-section uncertainties is not possible due to differing fiducial phase spaces: the 2016 analysis employed looser jet and lepton selections, lower-threshold leptonic triggers, and less restrictive jet and pileup identification criteria. While these differences have limited impact on the systematic component, they affect the statistical precision. Despite these stricter criteria, the full Run 2 measurement of electroweak Zjj production achieves an improved precision where the relative uncertainties amount to $^{6.66\%}_{6.48\%}$ in the combined $\mu\mu + ee$ measurement.

Table 7.2 summarizes the measured and predicted fiducial cross sections for electroweak Zjj production across different centre-of-mass energies and phase-space definitions.

Experiment	**Year**	**Measured Cross Section [fb]**
CMS 7 TeV, $m_{jj} > 120$ GeV	2011	$154 \pm 46_{\text{exp}} \pm 27_{\text{th}} \pm 24_{\text{stat}}$
ATLAS 8 TeV, $m_{jj} > 250$ GeV	2012	$54.7 \pm ^{9.8}_{10.4\text{syst}} \pm 4.6_{\text{stat}} \pm 1.5_{\text{lumi}}$
ATLAS 8 TeV, $m_{jj} > 1$ TeV	2012	$10.7 \pm 1.9_{\text{syst}} \pm 0.9_{\text{stat}} \pm 0.3_{\text{lumi}}$
ATLAS 13 TeV, $m_{jj} > 250$ GeV	2015	$119 \pm 20_{\text{syst}} \pm 16_{\text{stat}} \pm 2_{\text{lumi}}$
ATLAS 13 TeV, $m_{jj} > 1$ TeV	2015	$34.2 \pm 5.5_{\text{syst}} \pm 5.8_{\text{stat}} \pm 0.7_{\text{lumi}}$
CMS 13 TeV, $m_{jj} > 200$ GeV	2016	$534 \pm 57_{\text{syst}} \pm 20_{\text{stat}}$
ATLAS 13 TeV, $m_{jj} > 1$ TeV (139 fb^{-1})	2015-2018	$37.4 \pm 5.5_{\text{syst}} \pm 3.5_{\text{stat}}$
CMS 13 TeV, $m_{jj} > 200$ GeV	**2016-2018**	$\mathbf{596.5} \pm \binom{\mathbf{35.8}}{\mathbf{34.6}}_{\textbf{syst}} \pm \binom{\mathbf{17.3}}{\mathbf{17.3}}_{\textbf{stat}}$

Table 7.2: Measured and theoretical fiducial cross sections for EW Zjj production at different energies and phase space regions [62, 168, 169, 205] with the current measurement shown in the last row

A compilation of electroweak Zjj cross-section measurements from CMS and ATLAS, summarized in Table 7.2, contextualizes the present analysis. Despite variations in fiducial selections, dijet mass thresholds, and centre-of-mass energies, these results collectively outline the complementary approaches adopted by both collaborations. The table is not intended for numerical comparison but to illustrate the expanding experimental coverage and the methodological refinements achieved throughout successive LHC runs, demonstrating the increasing precision in studying electroweak Zjj production across diverse phase-space definitions.

7.2.3 Differential cross section for Electroweak Zjj

This section presents the signal strength and differential cross section corrected to the particle level, corresponding to the fiducial phase space defined in Section 6.2.1 for the EW Zjj process. Tables 7.3 and 7.4 summarizes the measured EW Zjj signal strengths for each individual channel as well as for their combined fit, reported for both the differential and inclusive measurements.

	$0 < p_{Tll} < 60$ GeV	**$60 < p_{Tll} < 110$ GeV**
$\mu\mu$ channel	1.26 (+0.23/ − 0.20) syst (+0.07/ − 0.07) stat	1.02 (+0.13/ − 0.13) syst (+0.06/ − 0.06) stat
ee channel	0.99 (+0.24/ − 0.22) syst (+0.10/ − 0.10) stat	1.00 (+0.20/ − 0.18) syst (+0.09/ − 0.09) stat
both channels	0.93 (+0.13/ − 0.11) syst (+0.05/ − 0.05) stat	1.12 (+0.09/ − 0.08) syst (+0.05/ − 0.05) stat

Table 7.3: Observed signal strengths in the low p_T^{ll} region (Full Run 2).

	$110 < p_{Tll} < 180$ GeV	**$180 < p_{Tll}$ GeV**	**Inclusive**
$\mu\mu$ channel	1.04 (+0.18/ − 0.17) syst (+0.08/ − 0.08) stat	0.99 (+0.12/ − 0.12) syst (+0.07/ − 0.07) stat	1.06 (+0.09/ − 0.09) syst (+0.04/ − 0.04) stat
ee channel	1.07 (+0.18/ − 0.17) syst (+0.09/ − 0.09) stat	0.99 (+0.17/ − 0.18) syst (+0.10/ − 0.09) stat	1.01 (+0.12/ − 0.11) syst (+0.05/ − 0.05) stat
both channels	1.18 (+0.10/ − 0.10) syst (+0.06/ − 0.06) stat	0.93 (+0.08/ − 0.08) syst (+0.06/ − 0.06) stat	1.04 (+0.06/ − 0.06) syst (+0.03/ − 0.03) stat

Table 7.4: Observed signal strengths in the high p_T^{ll} region and inclusive (Full Run 2).

Figure 7.7 shows the differential cross section for each channel separately and Figure 7.8 for the combination of both. The upper panel illustrates the differential cross section as a function of four distinct p_T^{ll} intervals, the black markers represent the measured values, where the vertical error bars indicate the systematic uncertainties, which also account for the theoretical modelling of the signal acceptance. The coloured shaded band denotes the total uncertainty, combining both statistical and systematic components. For comparison, the Standard Model prediction is shown by the gray markers, obtained from the `EWK_LLJJ` simulated sample, previously described in Section 7.2.2. The corresponding theoretical cross-section uncertainty is represented by the gray dashed band. The lower panel displays the ratio between the measured cross section and the Standard Model prediction, where the coloured band includes both statistical and systematic uncertainties, and the gray band represents the statistical component only.

Tables 7.5 and 7.6 summarizes the measured EW Zjj cross sections for each individual channel as well as for their combined fit, reported for both the differential and inclusive analyses.

Overall, the differential cross sections presented in Figures 7.7 and 7.8 show good agreement with the theoretical prediction. However, it is worth noting a deviation between the observed measurement and the theoretical expectation in the first p_T^{ll} bin ($0 < p_{Tll} < 60$ GeV) in the $\mu\mu$ channel. In this region, the signal strength μ exceeds unity in the $\mu\mu$ channel ($\mu = 1.26$), while it falls below unity in the ee

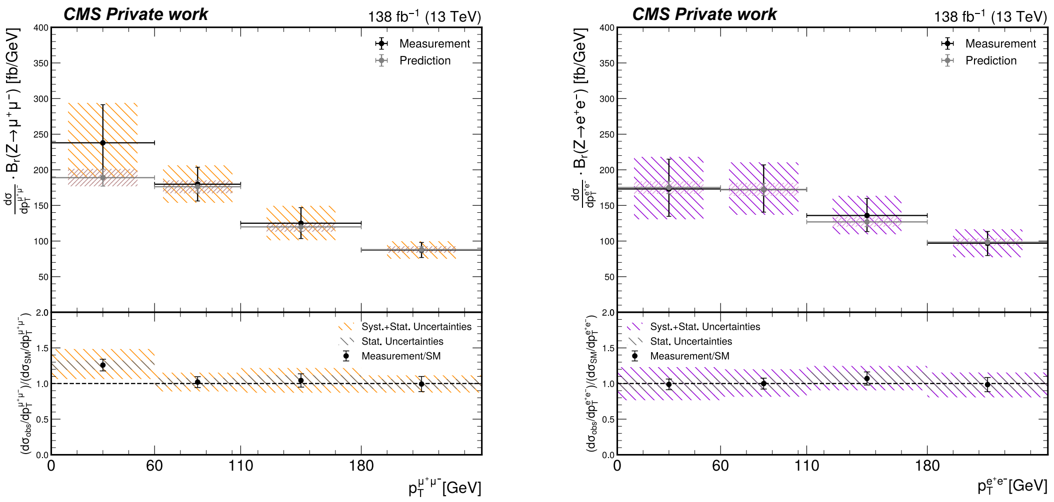


Figure 7.7: Differential cross section for EW Zjj with respect the Z-boson p_T: $\mu\mu$ channel (left) and ee channel (right) for the Full Run 2.

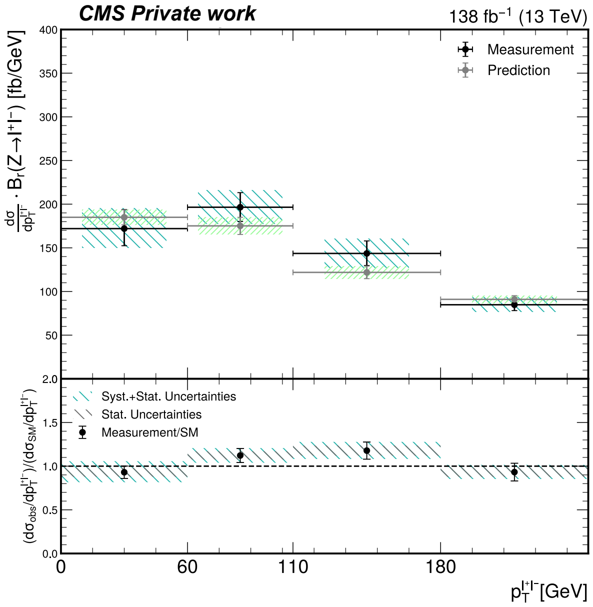


Figure 7.8: Differential cross section of EW Zjj with respect the Z-boson p_T: in the combined $\mu\mu$ and ee channels for the Full Run 2.

	$0 < p_{Tll} < 60$ GeV	**$60 < p_{Tll} < 110$ GeV**
$\mu\mu$ channel	237.9 (+53.5/ − 46.6) syst (+15.7/ − 15.7) stat fb	179.8 (+23.9/ − 23.5) syst (+10.8/ − 10.6) stat fb
ee channel	173.1 (+41.9/ − 38.4) syst (+16.8/ − 16.6) stat fb	172.3 (+34.6/ − 31.7) syst (+15.2/ − 14.8) stat fb
both channels	172.0 (+21.8/ − 19.6) syst (+8.9/ − 8.9) stat fb	196.5 (+16.7/ − 16.3) syst (+9.6/ − 9.4) stat fb

Table 7.5: Observed differential cross sections (fb) for low p_T^{ll} bins for the Full Run 2.

	110 < p_{Tll} < 180 GeV	**180 < p_{Tll} GeV**	**Inclusive**
$\mu\mu$ channel	125.0 (+21.9/ − 21.5) syst (+9.4/ − 9.4) stat fb	87.3 (+10.7/ − 10.5) syst (+6.0/ − 5.9) stat fb	609.1 (+56.6/ − 54.8) syst (+21.3/ − 21.3) stat fb
ee channel	135.9 (+23.9/ − 22.6) syst (+12.5/ − 12.4) stat fb	97.0 (+16.5/ − 17.4) syst (+9.9/ − 8.2) stat fb	579.3 (+67.2/ − 63.1) syst (+29.0/ − 29.0) stat fb
both channels	143.6 (+14.4/ − 13.8) syst (+8.5/ − 8.5) stat fb	84.8 (+7.0/ − 6.7) syst (+4.7/ − 4.7) stat fb	596.5 (+35.8/ − 34.6) syst (+17.3/ − 17.3) stat fb

Table 7.6: Observed differential and inclusive cross sections (fb) for high p_T^{ll} for the Full Run 2.

channel ($\mu = 0.99$). Consequently, the maximum likelihood fit adjusts the statistical model in opposite directions for the two channels, leading to a slight degree of tension when both are combined. Such behaviour is common in multi-channel analyses, where nuisance parameters and statistical fluctuations can affect channels differently, and variations in detector performance across data-taking periods can further contribute. In the combined fit, the resulting signal strength is $\mu = 0.930$, which is lower than unity. Although this result might appear counter-intuitive, it reflects the fact that the maximum likelihood fit operates beyond a simple linear combination, effectively compensating for potential overestimations of the signal strength present in individual channel fits, in addition this difference remains fully covered by the total uncertainty band.

The present analysis provides the first fiducial determination of the electroweak Zjj production at CMS using the full Run 2 dataset of 138 fb^{-1} collected at 13 TeV. It extends beyond the previous result based on 35.9 fb^{-1} and represents the first differential study of this process with respect the Z-boson p_T in the CMS experiment. A direct extrapolation of the 2016 uncertainties is not possible due to differing fiducial phase spaces although it is remarkable the higher precision achieved by the current measurement. These advancements enable the investigation of more selective kinematic regions, where higher dijet-mass thresholds and forthcoming luminosity increases will allow for more precise experimental and theoretical exploration of the VBF Z topology. In the present measurement, this topology, associated with the triple-gauge vertex diagram shown in Figure 6.1, contributes about 10% of the total electroweak Zjj yield, as illustrated in Figure 2.9. A dedicated experimental extraction of the VBF Z component would constitute a key step toward validating the heavy-boson TMDs and further developing theoretical descriptions of heavy-boson dynamics at future high-energy colliders.

Chapter 8

Conclusion

This thesis investigates the role of the electroweak sector of the Standard Model from both theoretical and experimental perspectives. The Drell-Yan process and electroweak Z^0jj production processes are formally studied and compared phenomenologically at particle level. This comparison highlights the several-orders-of-magnitude dominance of the Drell-Yan topology and underscores the necessity of defining a VBF Z-enhanced phase space to effectively distinguish between the two configurations. The impact of electroweak corrections is assessed and serves to motivate a deeper study of the determination and characteristics of parton distribution functions. In this context, heavy boson densities are derived from the extended DGLAP evolution equations and validated using DIS HERA data. Overall, the HERA measurements in both NC and CC regimes are reliably reproduced by the newly derived boson densities, demonstrating the robustness and predictive strength of the extended DGLAP framework in describing particle, and especially, heavy-boson interactions.

A particularly suitable process to probe heavy-boson densities is vector boson fusion (VBF) Z production via the triple-gauge vertex, as it directly tests the transverse momentum distribution of heavy bosons within the proton. Therefore, a VBF Z calculation based on heavy-boson densities is compared to a full matrix-element computation, establishing a solid foundation for future developments in the field.

In heavy-boson densities determination, further refinements are achievable such as more precise treatment of the photon-Z^0 interference. Moreover, a consistent inclusion of t-channel $W^{\pm}$ bosons in VBF processes, supported by the derivation of fully k_T-dependent VBF matrix elements, would be essential for an accurate description of the VBF Z process. Such improvements would enable a robust comparison with current and forthcoming experimental VBF Z measurements, particularly under the higher luminosity regimes.

The experimental analysis strategy for the electroweak Z^0jj process is then presented. A phase-space selection is first implemented to suppress background contributions, particularly from Drell–Yan Z^0 production. A Boosted Decision Tree is subsequently trained and optimized to enhance signal extraction. The measurement is determined through a maximum-likelihood fit, beginning with the inclusive Z^0jj cross section using 59.83fb $^{-1}$ of proton–proton collision data at $\sqrt{s} = 13$ TeV collected by the CMS experiment in 2018, and culminating in the study of the electroweak Z^0jj topology performed with the full Run 2 dataset corresponding to 138 fb^{-1} recorded between 2016 and 2018.

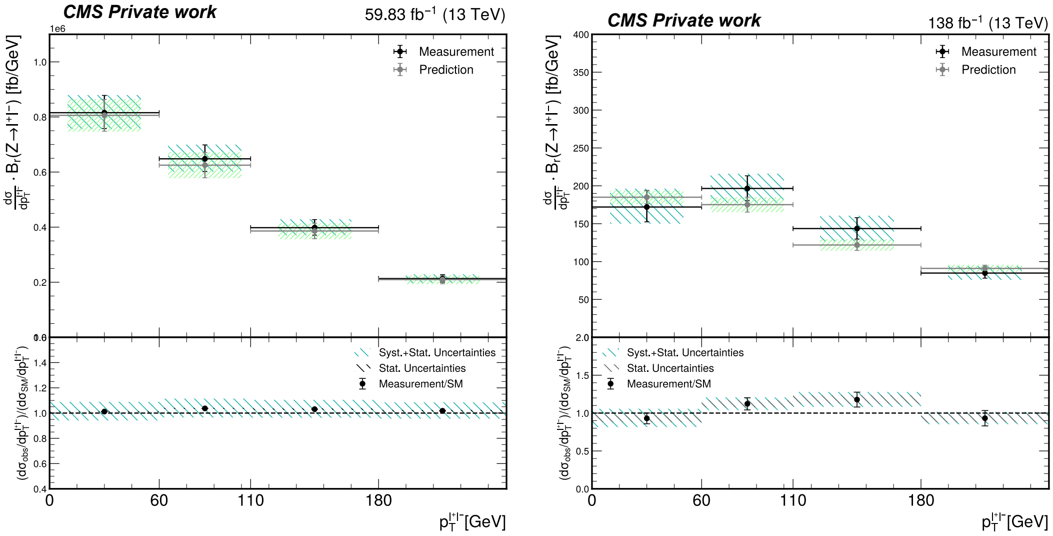


Figure 8.1: Differential cross sections with respect the Z-boson p_T: inclusive Z^0jj for $\mu\mu$ and ee channels combined for 2018 (left) and EW Z^0jj for $\mu\mu$ and ee channels combined for the Full Run 2 (right).

The inclusive Z^0jj cross section combining both channels is

$$\sigma^{\mu\mu+ee}_{Zjj} = 2.1 \times 10^6 \pm \binom{1.6{\times}10^5}{1.5{\times}10^5}_{\text{syst}} \pm \binom{4.1{\times}10^3}{4.1{\times}10^3}_{\text{stat}} \text{ fb}$$

and the electroweak Z^0jj cross section combining both channels is

$$\sigma^{\mu\mu+ee}_{\text{EW Zjj}} = 596.5 \pm \binom{35.8}{34.6}_{\text{syst}} \pm \binom{17.3}{17.3}_{\text{stat}} \text{ fb}$$

Both results show excellent agreement with the Standard Model predictions and with previous measurements, as detailed in Chapter 7.

In particular, the full Run 2 measurement of the electroweak Zjj process presents the first fiducial measurement of electroweak Zjj production in the CMS experiment utilizing the complete Run 2 proton-proton collision dataset corresponding to an integrated luminosity of 138 fb^{-1} at a center-of-mass energy of 13 TeV. It extends the previous CMS study based on 35.9 fb^{-1} and constitutes the first differential measurement of this process with respect to the Z-boson p_T performed within the CMS experiment. Due to differences in fiducial definitions, a direct comparison with the 2016 uncertainties is not feasible. Nevertheless, the total precision in the current analysis improved to a relative uncertainty of approximately 6%, establishing Run 2 as a new benchmark for EW Zjj analyses. This paves the way for exploring more stringent kinematic regions, where a higher dijet-mass threshold combined with increased luminosity at the High-Luminosity LHC (HL-LHC) will enable both experimental and theoretical studies of the VBF Z topology with greater accuracy. Both Run 3 [227] and the forthcoming HL-LHC phase [228] will deliver significantly larger datasets, enabling more precise measurements and improved stability in the simultaneous unfolding of the leptonic channels. The increased statistics will particularly mitigate the impact of shape-driven uncertainties such as those from the Jet Energy Scale (JES), to which low-statistics samples are highly sensitive.

The HL-LHC is scheduled to commence after Long Shutdown 3 (2026-2029), during which the CMS detector will undergo major Phase-2 upgrades [229]. From 2030 onward [230], it will operate at an instantaneous luminosity of up to $\mathcal{L} = 7.5 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ in pp collisions at $\sqrt{s} = 14$ TeV [229]. The Phase-2 upgrade includes the MIP Timing Detector (MTD), providing $\sim$ 30–35 ps timing resolution for charged particles, thus associating tracks with the correct vertex within a $\sim$ 190 ps pileup window [231]. This capability will reduce pileup jet rates, especially in the forward endcap regions crucial for VBF Z tagging, and will enhance signal extraction in high-pileup conditions.

Complementary timing information from upgraded calorimeters will further improve object reconstruction in both the $\mu\mu$ and ee channels [232, 233], with the former benefiting additionally from the upgraded muon spectrometers [234]. The new silicon tracker will provide higher granularity, extended coverage up to $|\eta| = 4$, and improved resolution [235], substantially advancing forward jet reconstruction and tracking efficiency in VBF processes.

Furthermore, the High-Granularity Calorimeter (HGCAL) [236] will deliver superior jet resolution and identification. Its fine segmentation enables enhanced discrimination between electromagnetic and hadronic shower shapes, improving separation of quark-initiated jets typical of EW Zjj production from gluon-initiated jets dominating the DY Zjj background.

Collectively, the theoretical developments and the phenomenological and experimental results presented in this thesis establish a foundation for more precise predictions and innovative measurement strategies at the HL-LHC opening new frontiers in the exploration of previously uncharted sectors of particle physics.

Bibliography

[1] R. K. Ellis, W. J. Stirling, and B. R. Webber. *QCD and Collider Physics.* 1st. Cambridge, UK: Cambridge University Press, 1996. ISBN: 978-0521563733.

[2] G. Arnison et al. "Experimental observation of isolated large transverse energy electrons with associated missing energy at s=540 GeV". In: *Physics Letters B* 122.1 (1983), pp. 103–116. ISSN: 0370-2693. DOI: `https://doi.org/10.1016/0370-2693(83)91177-2`. URL: `https://www.sciencedirect.com/science/article/pii/0370269383911772`.

[3] UA1 Collaboration et al. "Experimental observation of lepton pairs of invariant mass around 95 GeV/c^2 at the CERN SPS collider". In: *Phys. Lett. B* 126 (1983). First direct observation of the Z^0 boson, pp. 398–410. DOI: `10.1016/0370-2693(83)90188-0`.

[4] P. W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (1964), pp. 508–509. DOI: `10.1103/PhysRevLett.13.508`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.13.508`.

[5] CMS Collaboration. "Search for anomalous triple gauge couplings in WW and WZ production in lepton + jet events in proton-proton collisions at $\sqrt{s}$ = 13 TeV". In: *J. High Energy Phys.* 2019.12 (2019), p. 062. DOI: `10.1007/JHEP12(2019)062`. arXiv: `1907.08354 [hep-ex]`.

[6] Pietro Govoni and Chiara Mariotti. "Z production via Vector Boson Fusion at LHC". In: *arXiv preprint* (2010). eprint: `arXiv:1001.4357`. URL: `https://arxiv.org/abs/1001.4357`.

[7] ATLAS Collaboration. "Observation of vector-boson-fusion production of Higgs bosons in proton-proton collisions at $\sqrt{s}$ = 13 TeV". In: *CERN CDS Record* 2728055 (2020). Observation of Higgs boson production via vector boson fusion with the ATLAS detector at the LHC. URL: `https://cds.cern.ch/record/2728055`.

[8] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to Quantum Field Theory.* Westview Press, 1995. ISBN: 978-0-201-50397-5.

[9] Francis Halzen and Alan D. Martin. *Quarks and Leptons: An Introductory Course in Modern Particle Physics.* John Wiley & Sons, 1984. ISBN: 978-0-471-88741-6.

[10] Steven Weinberg. *The Quantum Theory of Fields, Volume 1: Foundations.* Cambridge University Press, 1995. ISBN: 978-0-521-55001-7. DOI: `10.1017/CBO9781139644167`.

[11] Emmy Noether. “Invariante Variationsprobleme”. In: *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* (1918), pp. 235–257. URL: `https://eudml.org/doc/59024`.

[12] P. A. M. Dirac. “The quantum theory of the electron”. In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 117.778 (1928), pp. 610–624. DOI: `10.1098/rspa.1928.0023`.

[13] S. Navas et al. (Particle Data Group). “Review of Particle Physics”. In: *Phys. Rev. D* 110 (2024). 2025 Update, p. 030001. URL: `https://pdg.lbl.gov/`.

[14] R P Feynman. *Space-Time Appxoach to Quantum Electxotiynamics*. Tech. rep. 1940.

[15] G. Rajasekaran. “Fermi and the theory of weak interactions”. In: *Resonance* 19.1 (Jan. 2014), pp. 18–44. ISSN: 0973-712X. DOI: `10.1007/s12045-014-0005-2`. URL: `http://dx.doi.org/10.1007/s12045-014-0005-2`.

[16] Stephen P. Martin and James D. Wells. *Elementary Particles and Their Interactions*. Graduate Texts in Physics. Cham: Springer, 2022. ISBN: 9783031143670. DOI: `10.1007/978-3-031-14368-7`.

[17] Sheldon L. Glashow. “Partial-symmetries of weak interactions”. In: *Nuclear Physics* 22 (1961), pp. 579–588. DOI: `10.1016/0029-5582(61)90469-2`.

[18] Steven Weinberg. “A Model of Leptons”. In: *Physical Review Letters* 19 (1967), pp. 1264–1266. DOI: `10.1103/PhysRevLett.19.1264`.

[19] Abdus Salam. “Weak and Electromagnetic Interactions”. In: *Elementary Particle Theory: Relativistic Groups and Analyticity (Nobel Symposium No. 8)*. Ed. by N. Svartholm. Stockholm: Almqvist & Wiksell, 1968, pp. 367–377.

[20] Palash B. Pal. “Representation-independent manipulations with Dirac matrices and spinors”. In: *arXiv:physics/0703214* (2007). URL: `https://arxiv.org/abs/physics/0703214`.

[21] J. Goldstone. “Field Theories with Superconductor Solutions”. In: *Nuovo Cim.* 19 (1961). Original Goldstone theorem paper, pp. 154–164.

[22] F. Englert and R. Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons”. In: *Phys. Rev. Lett.* 13 (1964), pp. 321–323. URL: `https://www.baumanrarebooks.com/rare-books/f-englert-peter-higgs/broken-symmetry-and-the-mass-of-gauge-vector-mesons-three-papers-/89273.aspx`.

[23] John Ellis. “Higgs Physics”. In: *arXiv preprint arXiv:1312.5672* (2013). Lectures presented at the ESHEP 2013 School of High-Energy Physics, to appear as part of the proceedings in a CERN Yellow Report. URL: `https://arxiv.org/pdf/1312.5672`.

[24] Arcadi Santamaria. *Masses, Mixings, Yukawa Couplings and their Symmetries*. Tech. rep. 1993.

[25] Richard P. Feynman. “Very High-Energy Collisions of Hadrons”. In: *Phys. Rev. Lett.* 23 (1969), pp. 1415–1417. DOI: `10.1103/PhysRevLett.23.1415`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.23.1415`.

[26] Mikel Mendizabal Morentin. “Z boson production in association with jets: measurement and phenomenology”. PhD thesis. Universität Hamburg, 2023.

[27] Cédric Lorcé et al. “Parton Distribution Functions and their Generalizations”. In: (July 2025). URL: `http://arxiv.org/abs/2507.12664`.

[28] John Collins. *Parton Distribution Functions (definition).* Accessed: 2025-09-01. 2012. URL: `https://www.scholarpedia.org/article/Parton_distribution_functions_(definition)`.

[29] F. Cornet et al. *A new 5 Flavour LO Analysis and Parametrization of Parton Distributions in the Real Photon.* Report CERN-TH-2002-362, DESY 02-118, IFT-22/2002, UG-FT-138/02, CAFPE-8/02. hep-ph/0212160. CERN, DESY, IFT, UG / CAFPE, 2002. URL: `https://cds.cern.ch/record/596343/files/0212160.pdf`.

[30] S. Baranov et al. “CASCADE3 A Monte Carlo event generator based on TMDs”. In: *European Physical Journal C* 81.5 (May 2021). ISSN: 14346052. DOI: `10.1140/epjc/s10052-021-09203-8`.

[31] F. Hautmann et al. “Collinear and TMD Quark and Gluon Densities from Parton Branching Solution of QCD Evolution Equations”. In: (Aug. 2017). DOI: `10.1007/JHEP01(2018)070`. URL: `http://arxiv.org/abs/1708.03279%20http://dx.doi.org/10.1007/JHEP01(2018)070`.

[32] John Collins and Ted C. Rogers. “Connecting Different TMD Factorization Formalisms in QCD”. In: (June 2017). DOI: `10.1103/PhysRevD.96.054011`. URL: `http://arxiv.org/abs/1705.07167%20http://dx.doi.org/10.1103/PhysRevD.96.054011`.

[33] R. Angeles-Martinez et al. “Transverse Momentum Dependent (TMD) parton distribution functions: status and prospects”. In: *Acta Phys. Pol. B* 46 (2015), pp. 2501–2534. eprint: `arXiv:1507.05267`.

[34] John C. Collins, Davison E. Soper, and George Sterman. “Transverse momentum distribution in Drell-Yan pair and W and Z boson production”. In: *Nucl. Phys. B* 250 (1985), pp. 199–224. DOI: `10.1016/0550-3213(85)90479-1`.

[35] John C. Collins. “CSS Equation, etc, Follow from Structure of TMD Factorization”. In: *Int. J. Mod. Phys. Conf. Ser.* 25 (2014), pp. 146–156. arXiv: `arXiv:1212.5974 [hep-ph]`.

[36] S Moch, J A M Vermaseren, and A Vogt. *The Quark Form Factor at Higher Orders.* Tech. rep. 2005.

[37] M. Boglione M. Anselmino and S. Melis. “A Phenomenological analysis of transverse momentum dependent distributions in proton-proton Drell-Yan processes”. In: *Phys. Rev. D* 89 (2014), p. 114026. eprint: `arXiv:1404.6465`.

[38] A. Vladimirov D. Scimemi. “Analysis of vector boson production within TMD factorization”. In: *Eur. Phys. J. C* 78 (2018), p. 89. eprint: `arXiv:1706.01473`.

[39] I. Bubanja et al. “The small kTregion in Drell–Yan production at next-to-leading order with the parton branching method”. In: *European Physical Journal C* 84.2 (Feb. 2024). ISSN: 14346052. DOI: `10.1140/epjc/s10052-024-12507-0`.

[40] H Jung. *QCD at Hadron Colliders Introduction and Perspectives*. Tech. rep. 2024.

[41] Barbara Jäger, Carlo Oleari, and Dieter Zeppenfeld. "Next-to-leading order QCD corrections to Z boson pair production via vector-boson fusion". In: *arXiv preprint* (Apr. 2006). arXiv: `hep-ph/0604200 [hep-ph]`. URL: `https://arxiv.org/pdf/hep-ph/0604200.pdf`.

[42] Frank Verstraete et al. "Renormalization algorithms for Quantum-Many Body Systems in two and higher dimensions". In: *arXiv preprint* (July 2004). arXiv: `cond-mat/0407066 [cond-mat]`. URL: `https://arxiv.org/pdf/cond-mat/0407066.pdf`.

[43] Dieter Zeppenfeld. *NLO QCD Corrections to VBF and VVV Production*. CTEQ Meeting, Argonne, Dec. 5-6, 2008. Universität Karlsruhe, Dec. 2008. URL: `https://web.pa.msu.edu/people/huston/cteq_anl/cteq_zeppenfeld.pdf`.

[44] T. Figy, C. Oleari, and D. Zeppenfeld. "Next-to-leading order jet distributions for Higgs boson production via weak-boson fusion". In: *arXiv preprint* (June 2003). arXiv: `hep-ph/0306109 [hep-ph]`. URL: `https://arxiv.org/pdf/hep-ph/0306109.pdf`.

[45] Leandro Cieri, Giancarlo Ferrera, and Germán F. R. Sborlini. "Combining QED and QCD transverse-momentum resummation for Z boson production at hadron colliders". In: *arXiv preprint* (May 2018). arXiv: `1805.11948 [hep-ph]`. URL: `https://arxiv.org/pdf/1805.11948.pdf`.

[46] Andrea Autieri et al. "Combining QED and QCD transverse-momentum resummation for W and Z boson production at hadron colliders". In: *arXiv preprint* (Feb. 2023). arXiv: `2302.05403 [hep-ph]`. URL: `https://arxiv.org/pdf/2302.05403.pdf`.

[47] Paolo Nason. "Shower Monte Carlo programs". In: *Proceedings of the MCWS Workshop / Lecture Notes*. Ed. by F. Frixione and R. Tenchini. Revisors: F. Frixione and R. Tenchini. 2006. URL: `https://virgilio.mib.infn.it/~nason/mcws/Proceedings/MCWSShower.pdf`.

[48] S. Mrenna and P. Skands. "Automated parton-shower variations in pythia 8". In: *Physical Review D* 94.7 (Oct. 2016). ISSN: 24700029. DOI: `10.1103/PhysRevD.94.074005`.

[49] Jonathan Bellm et al. "Herwig 7.0 / Herwig++ 3.0 release note". In: *Eur. Phys. J. C* 76 (2016), p. 196. DOI: `10.1140/epjc/s10052-016-4018-8`. arXiv: `1512.01178 [hep-ph]`.

[50] Enrico Bothmann et al. "Event generation with Sherpa 2.2". In: *SciPost Phys.* 7.3 (2019), p. 034. DOI: `10.21468/SciPostPhys.7.3.034`. arXiv: `1905.09127 [hep-ph]`.

[51] Andrew J. Larkoski, Juan J. Lopez-Villarejo, and Peter Skands. "Helicity-Dependent Showers and Matching with VINCIA". In: *Phys. Rev. D* 87.054033 (2013), p. 054033. DOI: `10.1103/PhysRevD.87.054033`. arXiv: `1301.0933 [hep-ph]`.

[52] H. Jung et al. "A parton shower consistent with parton densities at LO and NLO: Pdf2Isr". In: *European Physical Journal C* 85.8 (Aug. 2025). ISSN: 14346052. DOI: `10.1140/epjc/s10052-025-14595-y`.

[53] Bryan Webber. "Hadronization". In: *arXiv preprint* (1994). Lectures at Summer School on Hadronic Aspects of Collider Physics, Zuoz, Switzerland, August 1994. arXiv: `hep-ph/9411384`. URL: `https://arxiv.org/abs/hep-ph/9411384`.

[54] O. Biebel et al. *Fragmentation Functions in e^+e^-, ep, and pp Collisions.* Review 19.1. Revised August 2019. Particle Data Group (PDG), Lawrence Berkeley National Laboratory, Aug. 2019. URL: `https://pdg.lbl.gov/2019/reviews/rpp2019-rev-frag-functions.pdf`.

[55] Division of History of Physics Lund University. *The Lund model for high energy collisions. Theoretical ideas meet experimental reality.* Accessed: 27 Oct 2025. Lund, Sweden: Lund University, 2025. URL: `http://history.fysik.lu.se/EN/EN_Bok_17_LB_web.pdf`.

[56] Gurpreet Singh Chahal and Frank Krauss. "Cluster Hadronisation in SHERPA". In: *arXiv preprint* arXiv:2203.11385 (2022). v2, `https://arxiv.org/abs/2203.11385`.

[57] Jan-Christopher Winter, Frank Krauss, and Gerhard Soff. "A modified cluster-hadronization model". In: *arXiv e-prints* (Nov. 6, 2003). CERN-TH/2003-272; v1 at arXiv. eprint: `hep-ph/0311085`. URL: `https://arxiv.org/abs/hep-ph/0311085`.

[58] Philip Ilten et al. "Modeling Hadronization Using Machine Learning". In: *arXiv preprint* (2022). v2, `https://arxiv.org/abs/2203.04983`. arXiv: `2203.04983 [hep-ph]`.

[59] Soheil Kolouri et al. "Sliced-Wasserstein Autoencoder: An Embarrassingly Simple Generative Model". In: *arXiv preprint* (2018). v3, `https://arxiv.org/abs/1804.01947`. arXiv: `1804.01947 [cs.LG]`.

[60] Ian J. Goodfellow et al. "Generative Adversarial Nets". In: *Advances in Neural Information Processing Systems 27*. Ed. by Z. Ghahramani et al. Curran Associates, Inc., 2014, pp. 2672–2680.

[61] Jay Chan et al. "Fitting a Deep Generative Hadronization Model". In: *Journal of High Energy Physics* 2023.9 (2023), p. 084. DOI: `10.1007/JHEP09(2023)084`. arXiv: `2305.17169 [hep-ph]`. URL: `https://arxiv.org/abs/2305.17169`.

[62] S. Chatrchyan et al. "Measurement of the hadronic activity in events with a Z and two jets and extraction of the cross section for the electroweak production of a Z with two jets in pp collisions at sqrt(s)= 7 TeV". In: *Journal of High Energy Physics* 2013.10 (Oct. 2013). ISSN: 10298479. DOI: `10.1007/JHEP10(2013)062`.

[63] ATLAS Collaboration. "Measurement of the electroweak production of dijets in association with a Z-boson and distributions sensitive to vector boson fusion in proton-proton collisions at sqrt(s) = 8 TeV using the ATLAS detector". In: (Apr. 2014). DOI: `10.1007/JHEP04{\%}282014{\%}29031`. URL: `http://arxiv.org/abs/1401.7610%20http://dx.doi.org/10.1007/JHEP04%282014%29031`.

[64] Albert M Sirunyan et al. "Electroweak production of two jets in association with a Z boson in proton–proton collisions at $\sqrt{s} = 13$ TeV". In: *Eur. Phys. J. C* 78.7 (2018), p. 589. DOI: `10.1140/epjc/s10052-018-6049-9`. arXiv: `1712.09814 [hep-ex]`.

[65] David L. Rainwater, R. Szalapski, and Dieter Zeppenfeld. "Probing color-singlet exchange in +2-jet events at the CERN LHC". In: *Phys. Rev. D* 54.11 (1996), pp. 6680–6689. DOI: `10.1103/PhysRevD.54.6680`. arXiv: `hep-ph/9605444`.

[66] H. Chehime and D. Zeppenfeld. "Single W- and Z-boson production as a probe for rapidity gaps at the SSC". In: *Phys. Rev. D* 47 (1993), pp. 3898–3908. DOI: `10.1103/PhysRevD.47.3898`.

[67] POWHEG BOX collaboration. *Manual for Zjj Production via VBF in the POWHEG BOX*. Accessed: 2026-01-22. POWHEG BOX. unknown. URL: `https://usermanual.wiki/Pdf/manual.2069603384.pdf`.

[68] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *Journal of High Energy Physics* 2014.7 (2014). ISSN: 10298479. DOI: `10.1007/JHEP07(2014)079`.

[69] Torbjörn Sjöstrand et al. "An Introduction to PYTHIA 8.2". In: *arXiv preprint* (Oct. 2014). DOI: `10.1016/j.cpc.2015.01.024`. URL: `http://arxiv.org/abs/1410.3012%20http://dx.doi.org/10.1016/j.cpc.2015.01.024`.

[70] Simone Alioli et al. "A general framework for implementing NLO calculations in shower Monte Carlo programs: The POWHEG BOX". In: *Journal of High Energy Physics* 2010.6 (2010). ISSN: 10298479. DOI: `10.1007/JHEP06(2010)043`.

[71] Christian Bierlich et al. "Robust Independent Validation of Experiment and Theory: Rivet version 3". In: *SciPost Physics* 8.2 (2020), p. 026. DOI: `10.21468/SciPostPhys.8.2.026`. arXiv: `1912.05451 [hep-ph]`. URL: `https://scipost.org/SciPostPhys.8.2.026/pdf`.

[72] CMS Collaboration. *Precision measurements using top quarks at CMS.* Accessed: 2025-10-28. CERN. 2022. URL: `https://cms.cern/news/precision-measurements-using-top-quarks-cms`.

[73] P. Ciafaloni and D. Comelli. "Sudakov enhancement of electroweak corrections". In: *Phys. Lett. B* 446 (1999), pp. 278–284. arXiv: `hep-ph/9809321 [hep-ph]`.

[74] F. Bloch and A. Nordsieck. "Note on the radiation field of the electron". In: *Phys. Rev.* 52 (1937), pp. 54–59. DOI: `10.1103/PhysRev.52.54`.

[75] T. Kinoshita. “Mass singularities of Feynman amplitudes”. In: *J. Math. Phys.* 3 (1962), pp. 650–677. DOI: `10.1063/1.1724268`.

[76] Ansgar Denner. “Techniques for the calculation of electroweak radiative corrections at the one-loop level and results for W-physics at LEP200”. In: *Fortschritte der Physik* 41 (1993). Review paper re-uploaded to arXiv in 2007 (original 1993 publication), pp. 307–420. DOI: `10.1002/prop.2190410402`. arXiv: `0709.1075 [hep-ph]`.

[77] Roberto Frederix et al. “The automation of next-to-leading order electroweak calculations”. In: *JHEP* 2018.4 (2018), p. 103. DOI: `10.1007/JHEP04(2018)103`.

[78] E. Accomando, A. Denner, and C. Meier. “Electroweak corrections to $W\gamma$ and $Z\gamma$ production at the LHC”. In: *Eur. Phys. J. C* 47 (2006), pp. 125–146. DOI: `10.1140/epjc/s2006-02521-y`. arXiv: `hep-ph/0509234 [hep-ph]`.

[79] Terrance Figy, Sophy Palmer, and Georg Weiglein. “Higgs Production via Weak Boson Fusion in the Standard Model and the MSSM”. In: *arXiv preprint* (2010). eprint: `1012.4789`. URL: `https://arxiv.org/abs/1012.4789`.

[80] Ansgar Denner and Stefan Dittmaier. “Electroweak Radiative Corrections for Collider Physics”. In: *Phys. Rept.* 864 (2020), pp. 1–163. DOI: `10.1016/j.physrep.2020.01.004`. arXiv: `1912.06823 [hep-ph]`.

[81] Ansgar Denner and Stefano Pozzorini. “One-loop leading logarithms in electroweak radiative corrections, I. Results”. In: *Eur. Phys. J.* C18 (2001), pp. 461–480. DOI: `10.1007/s100520100551`. arXiv: `hep-ph/0010201 [hep-ph]`.

[82] W. Hollik. *Electroweak Theory.* arXiv:hep-ph/9602380. Lectures given at the 5th Hellenic School and Workshops on Elementary Particle Physics, Corfu, Greece, 3–24 Sept 1995. 1996. URL: `https://arxiv.org/abs/hep-ph/9602380`.

[83] M. Nowakowski and A. Pilaftsis. “On gauge invariance of Breit-Wigner propagators”. In: *Zeitschrift fur Physik C - Particles and Fields* 60.1 (1993), pp. 121–125. DOI: `10.1007/BF01650437`. arXiv: `hep-ph/9305321 [hep-ph]`.

[84] Michael Melles. “Electroweak radiative corrections in high energy processes”. In: *Physics Reports* 375.4 (2003), pp. 219–326. DOI: `10.1016/S0370-1573(02)00550-1`. URL: `https://www.sciencedirect.com/science/article/abs/pii/S0370157302005501`.

[85] Neda Darvishi and M. R. Masouminia. “Electroweak Radiative Corrections in Precision LHC Measurements of $W/Z+$ Jets”. In: *Nuclear Physics B* 985 (2022). Preprint arXiv:2112.15487 (2021), p. 116025. DOI: `10.1016/j.nuclphysb.2022.116025`. arXiv: `2112.15487 [hep-ph]`.

[86] M. Böhm et al. “Electroweak radiative corrections to $e^+e^- \to W^+W^-$”. In: *Nuclear Physics B* 304 (1988), pp. 463–499. DOI: `10.1016/0550-3213(88)90638-4`. URL: `https://doi.org/10.1016/0550-3213(88)90638-4`.

[87] Wolfgang Hollik. “Radiative Corrections in the Standard Model and Their Role for Precision Tests of the Electroweak Theory”. In: *Fortschritte der Physik* 38 (1990), pp. 165–260. DOI: `10.1002/prop.2190380302`.

[88] P. Ciafaloni and D. Comelli. “Sudakov Effects in Electroweak Corrections”. In: *arXiv preprint* (1998). arXiv:hep-ph/9809321. arXiv: `hep-ph/9809321 [hep-ph]`.

[89] Stefan Dittmaier. “A General Approach To Photon Radiation Off Fermions”. In: *Nucl. Phys. B* 565 (2000). submitted 23 April 1999, revised 16 July 1999, pp. 69–122. DOI: `10.1016/S0550-3213(99)00563-5`. arXiv: `hep-ph/9904440 [hep-ph]`.

[90] M. Ciccolini, S. Dittmaier, and M. Krämer. “Electroweak and QCD corrections to Higgs production via vector-boson fusion at the LHC”. In: *Phys. Rev. D* 68 (2003). Published version available at `https://link.aps.org/doi/10.1103/PhysRevD.68.073003`, p. 073003. DOI: `10.1103/PhysRevD.68.073003`. arXiv: `hep-ph/0306234 [hep-ph]`.

[91] Ansgar Denner. *Electroweak and QCD corrections to Higgs production in vector-boson fusion at the LHC*. Talk at *Radcor 2007*, Florence, Italy, October 1–5 2007. Slides (PDF) available at `https://www.ggi.infn.it/talkfiles/slides/talk421.pdf`. 2007.

[92] Stefan Dittmaier and Michael Krämer. “Electroweak radiative corrections to W-boson production at hadron colliders”. In: *Phys. Rev. D* 65.7 (2002), p. 073007. DOI: `10.1103/PhysRevD.65.073007`. arXiv: `hep-ph/0109062 [hep-ph]`.

[93] K. Moral Figueroa et al. “Determination of Parton Densities for QCD partons and Electroweak Bosons”. In: *arXiv preprint* (2025). arXiv: `arXiv:2511.08321 [hep-ph]`.

[94] Andy Buckley et al. “LHAPDF6: parton density access in the LHC precision era”. In: *Eur. Phys. J. C* 75.3 (2015), p. 132. DOI: `10.1140/epjc/s10052-014-3309-9`. arXiv: `1412.7420 [hep-ph]`.

[95] F. Hautmann et al. “TMDlib and TMDplotter: library and plotting tools for transverse-momentum-dependent parton distributions”. In: *Eur. Phys. J. C* 74 (2014), p. 3220. DOI: `10.1140/epjc/s10052-014-3220-9`. arXiv: `1408.3015 [hep-ph]`.

[96] S. et al. Abdulov. “TMDlib2 and TMDplotter: a platform for 3D hadron structure studies”. In: *Eur. Phys. J. C* 81 (2021), p. 763. DOI: `10.1140/epjc/s10052-021-09508-8`. arXiv: `2103.09741 [hep-ph]`.

[97] H. Jung, S. T. Monfared, and T. Wening. “Determination of collinear and TMD photon densities using the Parton Branching method”. In: *Phys. Lett. B* 817 (2021), p. 136299. DOI: `10.1016/j.physletb.2021.136299`. arXiv: `2102.01494 [hep-ph]`.

[98] A. V. Manohar et al. “The Photon Content of the Proton”. In: *JHEP* 12 (2017), p. 046. DOI: `10.1007/JHEP12(2017)046`. arXiv: `1708.01256 [hep-ph]`.

[99] C. Schmidt et al. “CT14QED parton distribution functions from isolated photon production in deep inelastic scattering”. In: *Phys. Rev. D* 93 (2016), p. 114015. DOI: `10.1103/PhysRevD.93.114015`. arXiv: `1509.02905 [hep-ph]`.

[100] A. Manohar et al. “How bright is the proton? A precise determination of the photon parton distribution function”. In: *Phys. Rev. Lett.* 117 (2016), p. 242002. DOI: `10.1103/PhysRevLett.117.242002`. arXiv: `1607.04266 [hep-ph]`.

[101] NNPDF Collaboration. “Parton distributions with QED corrections”. In: *Nucl. Phys. B* 877 (2013), pp. 290–320. DOI: `10.1016/j.nuclphysb.2013.10.010`. arXiv: `1308.0598 [hep-ph]`.

[102] A. D. Martin et al. “Parton distributions incorporating QED contributions”. In: *Eur. Phys. J. C* 39 (2005), pp. 155–161. DOI: `10.1140/epjc/s2004-02070-5`. arXiv: `hep-ph/0411040`.

[103] M. Roth and S. Weinzierl. “QED corrections to the evolution of parton distributions”. In: *Phys. Lett. B* 590 (2004), pp. 190–198. DOI: `10.1016/j.physletb.2004.03.042`. arXiv: `hep-ph/0403200`.

[104] M. Gluck, C. Pisano, and E. Reya. “The Polarized and unpolarized photon content of the nucleon”. In: *Phys. Lett. B* 540 (2002), pp. 75–82. DOI: `10.1016/S0370-2693(02)02048-7`. arXiv: `hep-ph/0206126`.

[105] L. A. Harland-Lang et al. “Ad Lucem: QED Parton Distribution Functions in the MMHT Framework”. In: *Eur. Phys. J. C* 79.10 (2019), p. 811. DOI: `10.1140/epjc/s10052-019-7331-0`. arXiv: `1907.02750 [hep-ph]`.

[106] Sally Dawson. “The Effective W Approximation”. In: *Nuclear Physics B* 249 (1985), pp. 42–60. DOI: `10.1016/0550-3213(85)90038-0`.

[107] G. L. Kane, Wayne Repko, and W. Rolnick. “The Effective $W^{\pm}$, Z^{0} Approximation for High-Energy Collisions”. In: *Physics Letters B* 148 (1984), pp. 367–372. DOI: `10.1016/0370-2693(84)90007-9`.

[108] Z. Kunszt and D. E. Soper. “On the Validity of the Effective W Approximation”. In: *Nuclear Physics B* 296 (1988), pp. 253–286. DOI: `10.1016/0550-3213(88)90047-4`.

[109] M. S. Chanowitz and M. K. Gaillard. “The TeV Physics of Strongly Interacting W’s and Z’s”. In: *Nuclear Physics B* 261 (1985), pp. 379–431. DOI: `10.1016/0550-3213(85)90580-2`.

[110] Alsu G. Bagdatova and Sergey P. Baranov. “Polarization and kinematic properties of the splitting functions $q \to W^{\pm} + q'$ and $q \to Z^{0} + q$”. In: *arXiv preprint* (2024). Version v2, revised 17 Mar 2025. DOI: `10.48550/arXiv.2404.10832`. arXiv: `2404.10832 [hep-ph]`.

[111] ATLAS Collaboration. *Measurements of the production cross section of a Z boson in association with high transverse momentum jets in pp collisions at* $\sqrt{s} = 13\,TeV$ *with the ATLAS detector*. Technical Report. CDS record: `https://cds.cern.ch/record/2777239`. Geneva, Switzerland: CERN, July 2021.

[112] H1 and ZEUS Collaborations. “Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data”. In: *Eur. Phys. J. C* 75.12 (2015), p. 580. DOI: `10.1140/epjc/s10052-015-3710-4`. arXiv: `1506.06042 [hep-ex]`.

[113] A. Bermudez Martinez et al. “Collinear and TMD parton densities from fits to precision DIS measurements in the parton branching method”. In: (Apr. 2018). DOI: `10.1103/PhysRevD.99.074008`. URL: `http://arxiv.org/abs/1804.11152%20http://dx.doi.org/10.1103/PhysRevD.99.074008`.

[114] F. Hautmann, H. Jung, and S. Taheri Monfared. “The CCFM uPDF evolution: uPDFevolv Version 1.0.00”. In: *Eur. Phys. J. C* 74.3082 (2014). DOI: `10.1140/epjc/s10052-014-3082-1`.

[115] H. Abdolmaleki et al. “xFitter: An Open Source QCD Analysis Framework”. In: *arXiv preprint* (2022). arXiv: `2206.12465 [hep-ph]`.

[116] J. Pumplin et al. “Uncertainties of predictions from parton distribution functions. 2. The Hessian method”. In: *Phys. Rev. D* 65 (2001), p. 014013. DOI: `10.1103/PhysRevD.65.014013`. arXiv: `hep-ph/0101032`.

[117] L. Cridge et al. “QED parton distribution functions in the MSHT20 fit”. In: *Eur. Phys. J. C* 82 (2022), p. 1. DOI: `10.1140/epjc/s10052-022-10137-0`. arXiv: `2111.05357 [hep-ph]`.

[118] Richard Ruiz et al. “The Effective Vector Boson Approximation in high-energy muon collisions”. In: *JHEP* 06 (2022), p. 114. DOI: `10.1007/JHEP06(2022)114`. arXiv: `2111.02442 [hep-ph]`.

[119] Bartosz Fornal, Aneesh V. Manohar, and Wouter J. Waalewijn. “Electroweak Gauge Boson Parton Distribution Functions”. In: *Journal of High Energy Physics* 05 (2018), p. 106. DOI: `10.1007/JHEP05(2018)106`. arXiv: `1803.06347 [hep-ph]`.

[120] Christian W. Bauer, Nicolas Ferland, and Bryan R. Webber. “Standard Model Parton Distributions at Very High Energies”. In: *Journal of High Energy Physics* 08 (2017), p. 036. DOI: `10.1007/JHEP08(2017)036`. arXiv: `1703.08562 [hep-ph]`.

[121] Michelangelo L Mangano, Mauro Moretti, and Michele Treccani. *Matching matrix elements and shower evolution for top-quark production in hadronic collisions.* Tech. rep. 2006.

[122] Xabier Cid Vidal and Ramon Cid Manzano. *Taking a closer look at LHC: LHC layout.* Accessed: 2025-09-16. 2023. URL: `https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.lhc_layout`.

[123] “Linac4 design report”. In: (). DOI: `10.23731/CYRM-2020-006`. URL: `https://doi.org/10.23731/CYRM-2020-006`.

[124] CERN. *Facts and figures about the LHC.* 2025. URL: `https://home.cern/resources/faqs/facts-and-figures-about-lhc`. Accessed: 2025-09-16.

[125] CMS Collaboration. *CMS Luminosity Public Results.* Accessed: 2025-09-16. 2022. URL: `https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults`.

[126] Francisco R. Villatoro. *Dibujo20170828 LHC FODO structure cell.* Accessed: 2025-09-16. 2017. URL: `https://francis.naukas.com/files/2017/08/Dibujo20170828-LHC-FODO-structure-cell.png`.

[127] G Aad et al. *The ATLAS Collaboration.* Tech. rep. 2008, p. 8003.

[128] *The ALICE experiment at the CERN LHC*. Tech. rep. URL: `http://www.iop.org/EJ/jinst/`.

[129] *LHCb A Large Hadron Collider Beauty Experiment for Precision Measurements of CP Violation and Rare Decays*. Tech. rep. 1998.

[130] *The CMS experiment at the CERN LHC*. Tech. rep. 2008. URL: `http://www.iop.org/EJ/jinst/`.

[131] Siona Ruth Davis and CMS Collaboration. *CMS slice image view (transverse/longitudinal/3-D)*. English. Accessed: 2025-10-05. 2016. URL: `https://cds.cern.ch/record/2204863/` (visited on 10/05/2025).

[132] The Tracker Group of the CMS Collaboration. "The CMS Phase-1 Pixel Detector Upgrade". In: (Dec. 2020). URL: `http://arxiv.org/abs/2012.14304`.

[133] Paul Begovatz. "Investigation of New Technologies to Improve Light Collection from Scintillating Crystals for Fast Timing". Accessed: 2025-10-05. PhD thesis. Aachen, Germany: RWTH Aachen University, 2017. URL: `https://www.researchgate.net/publication/364997605_Investigation_of_new_technologies_to_improve_light_collection_from_scintillating_crystals_for_fast_timing`.

[134] G. L. Bayatian et al. CMS Collaboration. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. English. CERN-LHCC-2006-001, CMS-TDR-8.1. 2006. URL: `https://cds.cern.ch/record/922757` (visited on 10/05/2025).

[135] Makoto Asai and Richard Mount. "Geant4 — Current and Future: A Snowmass 2013 White Paper". In: *arXiv preprint* (2013). White Paper, Snowmass 2013. eprint: `1308.1994`. URL: `https://arxiv.org/abs/1308.1994` (visited on 10/06/2025).

[136] The CMS Collaboration. "Performance of the CMS Level-1 Trigger in Proton-Proton Collisions at $\sqrt{s} = 13\,\mathrm{TeV}$". In: *arXiv preprint* (2020). Published in JINST as 10.1088/1748-0221/15/10/P10017. arXiv: `2006.10165 [hep-ex]`.

[137] The CMS Collaboration. "The CMS trigger system". In: *JINST* 12.01 (2017), P01020. DOI: `10.1088/1748-0221/12/01/P01020`. arXiv: `1609.02366 [physics.ins-det]`.

[138] Milos Dordevic. "The CMS Particle Flow Algorithm". In: *EPJ Web of Conferences*. Vol. 191. EDP Sciences, Oct. 2018. DOI: `10.1051/epjconf/201819102016`.

[139] W Adam et al. *Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at LHC*. Tech. rep. 2003.

[140] Wolfgang Adam. *Gaussian Sum Fitter*. URL: `https://twikiai.cern.ch/twiki/bin/view/CMSPublic/SWGuideGsfFitter`. (accessed: 05.07.2024).

[141] Xabier Cid Vidal and Ramón Cid Manzano. *Pile-up*. Accessed: 2025-10-06. Taking a Closer Look at LHC. 2025. URL: `https://lhc-closer.es/taking_a_closer_look_at_lhc/0.pile_up`.

[142] CMS Collaboration. *Study of Pileup Removal Algorithms for Jets*. `https://cds.cern.ch/record/1751454/files/JME-14-001-pas.pdf`. CERN-PAS-JME-14-001. 2014. URL: `https://cds.cern.ch/record/1751454/files/JME-14-001-pas.pdf`.

[143] *Pileup mitigation at CMS in 13 TeV data (PAS)*. Tech. rep.

[144] CMS Collaboration. *A Cambridge-Aachen (C-A) based Jet Algorithm for boosted top-jet tagging*. `https://cds.cern.ch/record/1194489/files/JME-09-001-pas.pdf`. CERN-PAS-JME-09-001. 2009. URL: `https://cds.cern.ch/record/1194489/files/JME-09-001-pas.pdf`.

[145] S. Catani et al. "Longitudinally invariant k_T clustering algorithms for hadron hadron collisions". In: *Nuclear Physics B* 406 (1993), pp. 187–224. DOI: `10.1016/0550-3213(93)90166-M`.

[146] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. *The anti-k t jet clustering algorithm*. Tech. rep. 2008, p. 63.

[147] Gavin P. Salam. "Towards Jetography". In: *The European Physical Journal C* 67 (2009), pp. 637–686. DOI: `10.1140/epjc/s10052-009-1199-4`. arXiv: `0906.1833 [hep-ph]`.

[148] Vardan Khachatryan et al. (CMS Collaboration). "Measurement of inclusive jet cross sections in pp collisions at $\sqrt{s} = 7$TeV". In: *The European Physical Journal C* 74 (2014), p. 2913. DOI: `10.1140/epjc/s10052-014-2913-1`. arXiv: `1406.0324 [hep-ex]`.

[149] Garvita Agarwal. "Jet Energy Scale and Resolution Measurements in CMS". In: (Jan. 2023). DOI: `10.22323/1.414.0652`. URL: `http://arxiv.org/abs/2301.02175%20http://dx.doi.org/10.22323/1.414.0652`.

[150] V. Khachatryan et al. "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV". In: *Journal of Instrumentation* 12.2 (Feb. 2017). ISSN: 17480221. DOI: `10.1088/1748-0221/12/02/P02014`.

[151] CMS Collaboration. *Jet energy scale and resolution measurement with Run 2 Legacy Data Collected by CMS at 13 TeV J*. Tech. rep. 2021.

[152] CMS Collaboration. *Jet Energy Scale and Resolution Performance with 13 TeV Data Collected by CMS in 2016–2018*. CMS Performance Note DP-2020/019. Accessed: 2026-01-16. CERN, Apr. 9, 2020. URL: `https://cds.cern.ch/record/2715872/files/DP2020_019.pdf`.

[153] CMS Collaboration. *Jet Energy Scale and Resolution Measurements Using Run3 Data Collected by CMS in 2022 and 2023 at 13.6 TeV*. CMS Performance Note DP-2024/039. Accessed: 2026-01-16. CERN, June 24, 2024. URL: `https://cds.cern.ch/record/2902862/files/DP2024_039.pdf`.

[154] CMS Collaboration. *Jet Energy Resolution and Smearing Procedures*. 2024. URL: `%5Curl%7Bhttps://twiki.cern.ch/twiki/bin/viewauth/CMS/JetResolution#Smearing_procedures%7D`. Accessed: 2025-09-04.

[155] CMS Collaboration. "Identification of b-quark jets with the CMS experiment". In: *arXiv preprint* (2013). Submitted to *Journal of Instrumentation* JINST **8** (2013) P04013. DOI: `10.1088/1748-0221/8/04/P04013`. arXiv: `1211.4462 [hep-ex]`. URL: `https://arxiv.org/pdf/1211.4462.pdf`.

[156] Swapneel Mehta et al. *DeepJet: A Machine Learning Environment for High-energy Physics*. Tech. rep. URL: `http://cds.cern.ch/record/2263802`.

[157] Leonardo Giannini. "Machine Learning and b-tagging in CMS". In: *Scientific Data Analysis School 2019*. Presented at Scientific Data Analysis School, SNS Pisa, Italy. Nov. 28, 2019. URL: `https://indico.sns.it/event/8/contributions/20/attachments/71/160/Machine_Learning_and_b-tagging_in_CMS.pdf`.

[158] A. M. Sirunyan et al. "Extraction and validation of a new set of CMS pythia8 tunes from underlying-event measurements". In: *European Physical Journal C* 80.1 (Jan. 2020). ISSN: 14346052. DOI: `10.1140/epjc/s10052-019-7499-4`.

[159] T. Sjostrand and M. van Zijl. "A Multiple Interaction Model for the Event Structure in Hadron Collisions". In: *Phys. Rev. D* 36 (1987), p. 2019.

[160] Stefano Carrazza et al. "An unbiased Hessian representation for Monte Carlo PDFs". In: *arXiv preprint* (2015). arXiv: `1504.03842 [hep-ph]`. URL: `https://arxiv.org/abs/1504.03842`.

[161] NNPDF Collaboration. "Parton distributions from high-precision collider data". In: *Eur. Phys. J. C* 77.5 (2017). NNPDF3.1 NNLO sets with 100 Hessian eigenvectors, p. 663. DOI: `10.1140/epjc/s10052-017-5191-3`. arXiv: `1706.00428 [hep-ph]`.

[162] CMS Open Data Group. *Background Modelling*. Accessed: 2025-10-06. CMS Open Data Group. 2025. URL: `https://cms-opendata-guide.web.cern.ch/analysis/backgrounds/`.

[163] EGamma POG. *L1PreFiring recipe*. URL: `https://twiki.cern.ch/twiki/bin/viewauth/CMS/L1PrefiringWeightRecipe`. (accessed: 05.07.2024).

[164] Muon POG. *Rochester corrections*. URL: `https://twiki.cern.ch/twiki/bin/view/CMS/RochcorMuon`. (accessed: 05.07.2024).

[165] K. Borras. *Status of the CMS Experiment*. Presentation at the RDMS Annual Workshop. Accessed: 2025-10-06. 2018. URL: `https://indico.cern.ch/event/754760/contributions/3127685/attachments/1722906/2782166/1_Borras.pdf`.

[166] CMS Collaboration. *Performance summary of AK4 jet b tagging with data from proton-proton collisions at 13 TeV with the CMS detector*. Accessed: 2025-10-06. 2023. URL: `https://btv-public.docs.cern.ch/DPNotesRun2/DP-2023-005/DP-2023-005/`.

[167] CMS Jet Energy Reconstruction and Calibration (JERC) Group. *CMS JERC Recommendations for Run 2 Analyses*. Accessed: 2025-10-06. 2023. URL: `https://cms-jerc.web.cern.ch/Recommendations/#run-2_2`.

[168] M. Aaboud et al. "Measurement of the cross-section for electroweak production of dijets in association with a Z boson in pp collisions at s=13 TeV with the ATLAS detector". In: *Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics* 775 (Dec. 2017), pp. 206–228. ISSN: 03702693. DOI: `10.1016/j.physletb.2017.10.040`.

[169] G. Aad et al. "Differential cross-section measurements for the electroweak production of dijets in association with a Z boson in proton–proton collisions at ATLAS". In: *European Physical Journal C* 81.2 (Feb. 2021). ISSN: 14346052. DOI: `10.1140/epjc/s10052-020-08734-w`.

[170] ATLAS Collaboration. "Search for invisible decays of the Higgs boson produced via vector boson fusion at the LHC with the ATLAS Run-2 data". In: *J. High Energy Phys.* 2021.7 (2021), p. 088. DOI: `10.1007/JHEP07(2021)088`. arXiv: `2103.01802 [hep-ex]`.

[171] ATLAS Collaboration. "Search for $H \to c\,\bar{c}$ in vector boson fusion production with Run-3 data at the LHC". In: *J. High Energy Phys.* 2025.8 (2025), p. 055. DOI: `10.1007/JHEP08(2025)055`. arXiv: `2504.12345 [hep-ex]`.

[172] CMS Collaboration. "Search for invisible decays of the Higgs boson produced via vector boson fusion at the LHC with Run-2 data". In: *Phys. Lett. B* 819 (2021), p. 136429. DOI: `10.1016/j.physletb.2021.136429`. arXiv: `2103.01802 [hep-ex]`.

[173] CMS Collaboration. "Search for $H \to c\,\bar{c}$ in vector boson fusion production with Run-3 data at the LHC". In: *J. High Energy Phys.* 2025.8 (2025), p. 055. DOI: `10.1007/JHEP08(2025)055`. arXiv: `2504.12345 [hep-ex]`.

[174] ATLAS Collaboration. "Search for new physics in final states with large missing transverse momentum and jets using 139 fb^{-1} of pp collision data at $\sqrt{s}$ = 13 TeV with the ATLAS detector". In: *J. High Energy Phys.* 2020.8 (2020), p. 202. DOI: `10.1007/JHEP08(2020)202`. arXiv: `2003.08089 [hep-ex]`.

[175] ATLAS Collaboration. "Search for anomalous triple gauge couplings in WW and WZ production in lepton + jet events in proton-proton collisions at $\sqrt{s} =$ 13.6 TeV with the ATLAS detector". In: *J. High Energy Phys.* 2024.8 (2024), p. 123. DOI: `10.1007/JHEP08(2024)123`. arXiv: `2406.12345 [hep-ex]`.

[176] CMS Collaboration. "Search for new physics in final states with large missing transverse momentum and jets using 137 fb^{-1} of pp collision data at $\sqrt{s} = 13$ TeV with the CMS detector". In: *J. High Energy Phys.* 2020.8 (2020), p. 202. DOI: `10.1007/JHEP08(2020)202`. arXiv: `2003.08089 [hep-ex]`.

[177] CMS Collaboration. "Search for anomalous triple gauge couplings in WW and WZ production in lepton + jet events in proton-proton collisions at $\sqrt{s} =$ 13.6 TeV with the CMS detector". In: *J. High Energy Phys.* 2024.8 (2024), p. 123. DOI: `10.1007/JHEP08(2024)123`. arXiv: `2406.12345 [hep-ex]`.

[178] U. Baur and D. Zeppenfeld. *Measuring three vector boson couplings in $qq \to qqW$ at the SSC.* 1993. arXiv: `hep-ph/9309227 [hep-ph]`.

[179] ATLAS Collaboration. "Measurement of the WW production cross section and limits on anomalous triple gauge couplings in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Phys. Lett. B* 772 (2017), pp. 232–253. DOI: `10.1016/j.physletb.2017.06.037`. arXiv: `1702.04519 [hep-ex]`.

[180] ATLAS Collaboration. "Search for long-lived particles decaying into displaced leptons with the ATLAS detector at $\sqrt{s} = 13.6$ TeV". In: *J. High Energy Phys.* 2024.7 (2024), p. 123. DOI: `10.1007/JHEP07(2024)123`. arXiv: `2406.12345 [hep-ex]`.

[181] CMS Collaboration. "Search for long-lived particles decaying into displaced leptons with the CMS detector at $\sqrt{s} = 13.6$ TeV". In: *J. High Energy Phys.* 2024.7 (2024), p. 123. DOI: `10.1007/JHEP07(2024)123`. arXiv: `2406.12345 [hep-ex]`.

[182] CMS Collaboration. *McM (Monte Carlo Management)*. Accessed: 2 Nov 2025. 2025. URL: `https://cms-pdmv-prod.web.cern.ch/mcm/`.

[183] Darin Acosta. *The CMS Trigger System*. Tech. rep.

[184] Rikkert Frederix and Stefano Frixione. "Merging meets matching in MC@NLO". In: *Journal of High Energy Physics* 2012.12 (Dec. 2012). ISSN: 10298479. DOI: `10.1007/JHEP12(2012)061`.

[185] Pierre Artoisenet et al. "Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations". In: *Journal of High Energy Physics* 2013.3 (Mar. 2013). ISSN: 10298479. DOI: `10.1007/JHEP03(2013)015`.

[186] Christian Bierlich et al. "A comprehensive guide to the physics and usage of PYTHIA 8.3". In: *SciPost Physics Codebases* (Nov. 2022). DOI: `10.21468/scipostphyscodeb.8`.

[187] Richard D. Ball et al. "Parton distributions from high-precision collider data: NNPDF Collaboration". In: *European Physical Journal C* 77.10 (Oct. 2017). ISSN: 14346052. DOI: `10.1140/epjc/s10052-017-5199-5`.

[188] CERN Courier. *First measurements of electroweak boson fusion*. Accessed: 3 Nov 2025. Feb. 2013. URL: `https://cerncourier.com/a/first-measurements-of-electroweak-boson-fusion/`.

[189] CMS Collaboration. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". In: *JINST* 9.10 (2014), P10009. DOI: `10.1088/1748-0221/9/10/P10009`.

[190] Muon POG. *Muon ID UL18*. URL: `https://twiki.cern.ch/twiki/bin/view/CMS/MuonUL2018#ID_efficiencies_AN1`. (accessed: 05.07.2024).

[191] EGamma POG. *Electron MVA ID Run 2*. URL: `https://twiki.cern.ch/twiki/bin/view/CMS/MultivariateElectronIdentificationRun2#MVA_Recipe_V1_for_regular_users`. (accessed: 05.07.2024).

[192] *Cut-Based Electron Identification Run 2*. CMS Collaboration. 2023. URL: `https://twiki.cern.ch/twiki/bin/viewauth/CMS/CutBasedElectronIdentificationRun2` (visited on 09/03/2025).

[193] Muon POG. *Muon ISO UL18*. URL: `https://twiki.cern.ch/twiki/bin/view/CMS/MuonUL2018#ISO_efficiencies`. (accessed: 05.07.2024).

[194] CMS Collaboration. "Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC". In: *JINST* 15.12 (2020), P12005. DOI: `10.1088/1748-0221/15/12/P12005`.

[195] CMS Collaboration. *JetID: Recommendations for 13 TeV 2018, 2017 and 2016*. Accessed: 2025-10-13. 2018. URL: `https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetID#Recommendations_for_13_TeV_2018`.

[196] BTV POG. *B-tagging uncertainty guideline for the Full Run 2.* URL: `https://btv-wiki.docs.cern.ch/PerformanceCalibration/SFUncertaintiesAndCorrelations/#ak4-working-point-based-sfs-fixedwp-sfs`. (accessed: 23.05.2025).

[197] JetMET POG. *MET filters.* URL: `https://twiki.cern.ch/twiki/bin/viewauth/CMS/MissingETOptionalFiltersRun2#Run_2_recommendations`. (accessed: 05.07.2024).

[198] JERC Collaboration. *Jet Energy Resolution and Calibration (JERC) Meeting, 9 September 2019, Version 2.* Accessed: 2025-10-13. CERN. 2019. URL: `https://indico.cern.ch/event/846608/contributions/3556096/attachments/1904250/3144533/JERCMeeting_9_9_2019_V2.pdf`.

[199] Helge Voss, Andreas Höcker, and Jörg Stelzer. *TMVA, the Toolkit for Multivariate Data Analysis with ROOT PoS(ACAT)040 TMVA, the Toolkit for Multivariate Data Analysis.* Tech. rep. URL: `http://pos.sissa.it/`.

[200] CMS Collaboration. *Performance of Boosted Decision Trees in the Separation of Electroweak Zjj Signal from Drell-Yan Background.* `https://cds.cern.ch/record/2300206/files/DP2017_060.pdf`. Accessed: 2025-10-13. 2017.

[201] Yann Coadou. “Boosted Decision Trees and Applications in High-Energy Physics”. In: *arXiv preprint arXiv:2206.09645* (2022). Accessed: 2025-10-13. URL: `https://arxiv.org/pdf/2206.09645.pdf`.

[202] Jerome H Friedman. *1999 REITZ LECTURE GREEDY FUNCTION APPROXIMATION: A GRADIENT BOOSTING MACHINE.* Tech. rep. 5. 2001, pp. 1189–1232.

[203] Carlos Guestrin Tianqi Chen. “XGBoost: A Scalable Tree Boosting System”. In: (2018). DOI: `10.1145/2939672.2939785`. arXiv: `1603.02754 [cs.LG]`.

[204] XGBoost. *XGBoost documentation.* URL: `https://xgboost.readthedocs.io/en/stable/`. (accessed: 05.07.2024).

[205] CMS Collaboration. *Electroweak production of two jets in association with a Z boson in proton-proton collisions at sqrt(s) = 13 TeV.* Tech. rep.

[206] Glen Cowan et al. “Asymptotic formulae for likelihood based tests of new physics”. In: *European Physical Journal C* 71.2 (Feb. 2011). ISSN: 14346052. DOI: `10.1140/epjc/s10052-011-1554-0`.

[207] Stefan Schmitt. “TUnfold: an algorithm for correcting migration effects in high energy physics”. In: *JINST* 7 (2012), T10003. DOI: `10.1088/1748-0221/7/10/T10003`. arXiv: `1205.6201 [physics.data-an]`.

[208] Wouter Verkerke and David Kirkby. “The RooFit toolkit for data modeling”. In: *arXiv preprint* (2003). arXiv: `physics/0306116 [physics.data-an]`.

[209] J. S. Conway. “Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra”. In: (Mar. 2011). URL: `http://arxiv.org/abs/1103.0354`.

[210] EGamma POG. *Electron SF uncertainties guidelines for the Full Run 2.* URL: `https://twiki.cern.ch/twiki/bin/view/CMS/EgammaRunIIRecommendations`. (accessed: 23.05.2025).

[211] CMS Collaboration. *EgHLT Scale Factor Measurements.* 2024. URL: `https://twiki.cern.ch/twiki/bin/viewauth/CMS/EgHLTScaleFactorMeasurements`. Accessed: 2025-09-04.

[212] CMS Collaboration. "Search for the rare decay of the W boson into a pion and a photon in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Phys. Lett. B* 843 (2022), p. 137979. DOI: `10.1016/j.physletb.2023.137979`. URL: `https://cms-results.web.cern.ch/cms-results/public-results/publications/SMP-20-008/index.html`.

[213] Muon POG. *Muon SF uncertainties guidelines for the Full Run 2.* URL: `https://twiki.cern.ch/twiki/bin/view/CMS/MuonUL2018#Note_on_correlations`. (accessed: 23.05.2025).

[214] JetMET Group. *General JEC uncertainty recommendations for Full Run 2.* URL: `https://twiki.cern.ch/twiki/bin/view/CMS/JECAnalysesRecommendations`. (accessed: 23.05.2025).

[215] JetMET Group. *Run 2 reduced set of JEC uncertainty sources.* URL: `https://twiki.cern.ch/twiki/bin/view/CMS/JECUncertaintySources#Run_2_reduced_set_of_uncertainty`. (accessed: 23.05.2025).

[216] JetMET Group. *JER uncertainty guideline for the Full Run 2.* URL: `https://cms-jerc.web.cern.ch/Recommendations/#jet-energy-resolution_1`. (accessed: 23.05.2025).

[217] LUMI POG. *Luminosity uncertainties for the Full Run 2.* URL: `https://twiki.cern.ch/twiki/bin/viewauth/CMS/LumiRecommendationsRun2#Luminosity_for_pp_13_TeV_data`. (accessed: 23.05.2025).

[218] LUMI POG. *2018 Pileup Scenarios for Ultra-Legacy.* URL: `https://indico.cern.ch/event/826107/contributions/3465964/attachments/1861686/3059846/PileupPPD20190613.pdf`. (accessed: 23.05.2025).

[219] LUMI POG. *2017 Pileup Scenarios for Ultra-Legacy.* URL: `https://indico.cern.ch/event/815042/contributions/3406023/attachments/1834447/3004998/PileupPPD20190425.pdf`. (accessed: 23.05.2025).

[220] LUMI POG. *2016 Pileup Scenarios for Ultra-Legacy.* URL: `https://indico.cern.ch/event/870037/contributions/3670166/attachments/1960095/3257291/Pileup20191210.pdf`. (accessed: 23.05.2025).

[221] Tau POG. *Cross-section uncertainties.* URL: `https://twiki.cern.ch/twiki/bin/view/CMS/TauIDMeasurementsHelp#2017_measurements_MC2017_v2_samp`. (accessed: 23.05.2025).

[222] Jon Butterworth et al. "PDF4LHC recommendations for LHC Run II". In: *arXiv preprint* (2015). see section 6 for usage guidelines and uncertainties. arXiv: `1510.03865 [hep-ph]`.

[223] Top Group. *PDF uncertainties guidelines.* URL: `https://twiki.cern.ch/twiki/bin/viewauth/CMS/TopSystematics#PDF`. (accessed: 23.05.2025).

[224] ATLAS Collaboration. "Measurement of $W^{\pm}$ and Z-boson production cross sections in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Phys. Lett. B* 759 (2016), pp. 601–621. DOI: `10.1016/j.physletb.2016.06.023`. arXiv: `1603.09222 [hep-ex]`. URL: `https://cds.cern.ch/record/2142492/files/arXiv%3A1603.09222.pdf`.

[225] CMS Collaboration. *Measurements of inclusive and differential Z boson production cross sections in pp collisions at $\sqrt{s} = 13$ TeV*. CMS Physics Analysis Summary CMS-PAS-SMP-15-011. CERN, Mar. 2016. URL: `https://cds.cern.ch/record/2140105/files/SMP-15-011-pas.pdf`.

[226] CMS Collaboration. "Measurement of the inclusive cross sections for W and Z boson production in proton-proton collisions at $\sqrt{s} = 5.02$ and 13 TeV". In: *JHEP* 04 (2025), p. 162. DOI: `10.1007/JHEP04(2025)162`. arXiv: `2408.03744 [hep-ex]`. URL: `https://cds.cern.ch/record/2906542/files/CMS-SMP-20-004-arXiv.pdf`.

[227] Srećko Morović and CMS Collaboration. "CMS detector: Run 3 status and plans for Phase-2". In: *arXiv* (2023). Preprint; Presented at DIS2023, XXX International Workshop on Deep-Inelastic Scattering and Related Subjects, 27–31 March 2023. eprint: `2309.02256`. URL: `https://arxiv.org/abs/2309.02256`.

[228] CERN. *High-Luminosity LHC*. Accessed: 2026-01-24; Overview of the High-Luminosity Large Hadron Collider upgrade project at CERN. 2025. URL: `https://home.cern/science/accelerators/high-luminosity-lhc`.

[229] Thiago R. F. P. Tomei. "CMS Upgrades for the High-Luminosity LHC Era". In: *arXiv preprint* (2025). arXiv:2501.03412v1 [hep-ex], presented at the 12th Large Hadron Collider Physics Conference (LHCP2024). DOI: `10.48550/arXiv.2501.03412`. arXiv: `2501.03412 [hep-ex]`.

[230] CERN Courier. *Revised schedule for the High-Luminosity LHC*. Accessed: 2026-01-24. IOP Publishing / CERN Courier. Nov. 13, 2024. URL: `https://cerncourier.com/a/revised-schedule-for-the-high-luminosity-lhc/`.

[231] CMS Collaboration. *A MIP Timing Detector for the CMS Phase-2 Upgrade*. Tech. rep. Geneva, Switzerland: CERN, 2019. URL: `https://cds.cern.ch/record/2667167`.

[232] CMS Collaboration. *The Phase-2 Upgrade of the CMS Endcap Calorimeter*. Tech. rep. Geneva, Switzerland: CERN, 2017. URL: `http://cds.cern.ch/record/2293646`.

[233] CMS Collaboration. *The Phase-2 Upgrade of the CMS Barrel Calorimeters*. Tech. rep. Geneva, Switzerland: CERN, 2017. URL: `https://cds.cern.ch/record/2283187`.

[234] CMS Collaboration. *The Phase-2 Upgrade of the CMS Muon Detectors*. Tech. rep. This is the final version, approved by the LHCC. Geneva, Switzerland: CERN, 2017. DOI: `10.17181/CERN.5T9S.VPMI`. URL: `https://cds.cern.ch/record/2283189`.

[235] CMS Collaboration. *The Phase-2 Upgrade of the CMS Tracker*. Tech. rep. Geneva, Switzerland: CERN, 2017. DOI: `10.17181/CERN.QZ28.FLHW`. URL: `https://cds.cern.ch/record/2272264`.

[236] Chiara Amendola and CMS Collaboration. “An overview of the CMS High-Granularity Calorimeter”. In: *JINST* 20.01 (2025). Presented at the 25th International Workshop on Radiation Imaging Detectors (iWoRiD 2024), Lisbon, Portugal, 30 June–4 July 2024, p. C01031. DOI: `10.1088/1748-0221/20/01/C01031`. URL: `https://cds.cern.ch/record/2918593`.

Acknowledgments

I would like to acknowledge Prof. Elisabetta Gallo and Dr. Katarzyna Wichmann for their supervision during my doctoral research. I appreciate the institutional support and resources provided during my time at DESY and the University of Hamburg, which were instrumental in bringing my work to fruition.

I extend my special gratitude to Hannes Jung, who supported me since the very beginning till the end and gave me the strength and courage to face all the difficulties I encountered during all these years. Thanks for your knowledge, advice and all the projects and lunches together. Thanks for helping make science a more equal and safer space where collaboration is encouraged, diversity is cherished, effort is valued and all voices are listened to, thanks for making me feel like I belonged at the table. I see in you not only a colleague, but a friend too.

I am grateful for numerous idea exchanges with A. Bagdatova, S. Baranov, A. Kotikov, A. Lipatov, M. Malyshev and the other participants of the WeeklyOfflineMeeting during the last years.

I express my appreciation to the QCD group: Mikel, Qun, Sara and Luis Ignacio, thanks for welcoming me from the very start and helping me during countless hours. Mikel, thanks for all your invaluable discussions, your support and for opening up your community to me, I have found a friend in you with no doubt at all. I am deeply thankful to all the current and former CMS PhD students and postdocs, especially Aliya, Dominic, Hugo, Laurids, Moritz, Gabriele, Ana, Lovisa, Beatriz and Federica, who always answered all my questions, and pushed me through these years.

Jacobo, Stepan and Mathilde, thanks for your friendship, warmth and offering a coffee (or even a whole jar!) anytime I needed it, which was honestly very often. My sincere thanks to Jörn, Daina and Saverio, who cheered me up through my competitive sports journey. Overall, I am deeply appreciative of the whole CMS group, for all their support through my PhD.

Finally, Raimundo, thanks for being by my side since 2022, I have grown as a person by your side and I truly see a brother in you.

Appendix A

Detector-Level Distributions of Dilepton Transverse Momentum for Additional Eras

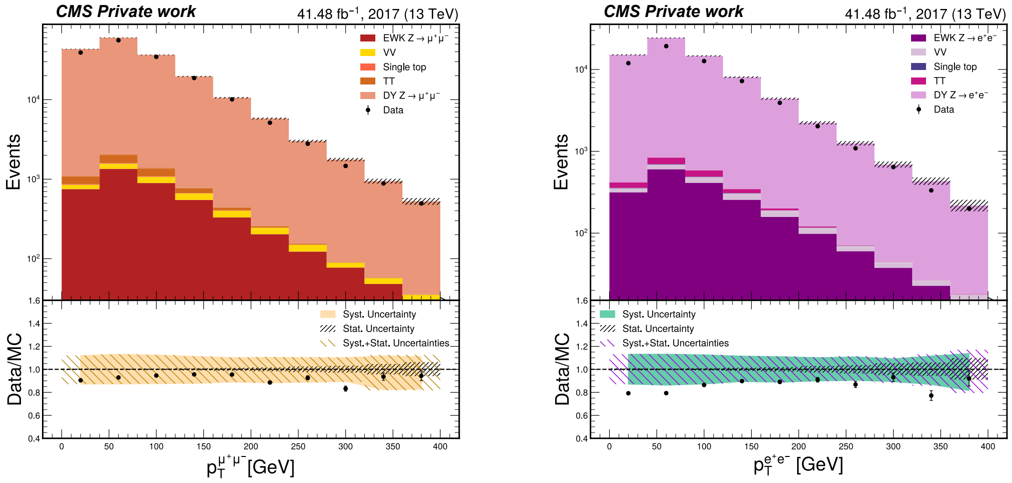


Figure A.1: $p_T^{\ell\ell}$ comparison between data and simulation for the 2017 data taking period. These distributions are normalized to the expected event yields, computed from theoretical cross sections and integrated luminosity per year, and corrected for all weights and detector-level scale factors.

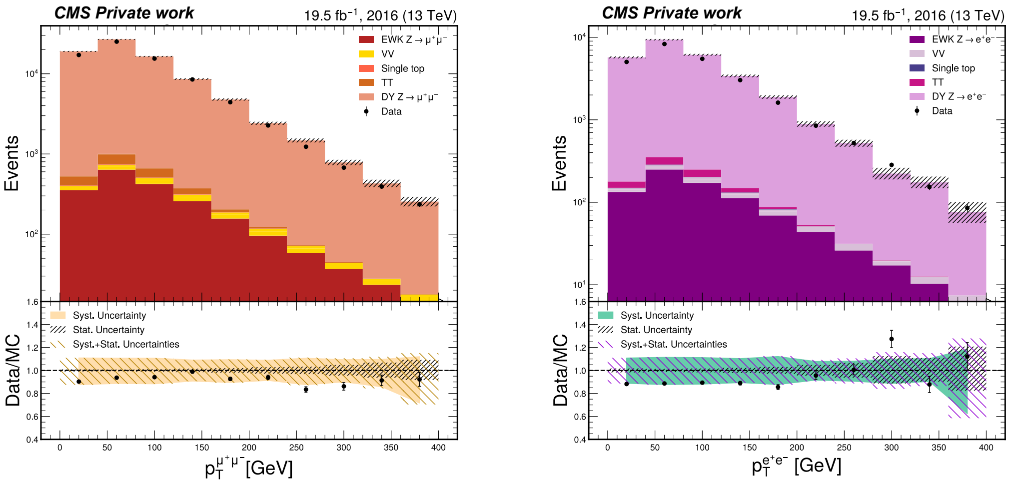


Figure A.2: $p_T^{\ell\ell}$ comparison between data and simulation for the 2016preVFP data taking period. These distributions are normalized to the expected event yields, computed from theoretical cross sections and integrated luminosity per year, and corrected for all weights and detector-level scale factors.

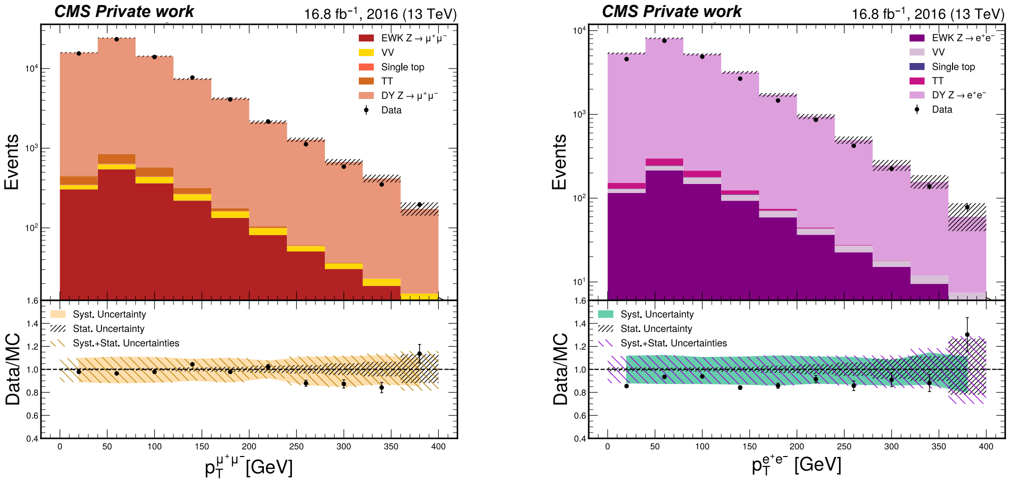


Figure A.3: $p_T^{\ell\ell}$ comparison between data and simulation for the 2016postVFP data taking period. These distributions are normalized to the expected event yields, computed from theoretical cross sections and integrated luminosity per year, and corrected for all weights and detector-level scale factors.

Appendix B

Extended Set of Detector-Level Observables Across Eras

All these distributions are normalized to the expected event yields, computed from theoretical cross sections and integrated luminosity per year, and corrected for all weights and detector-level scale factors.

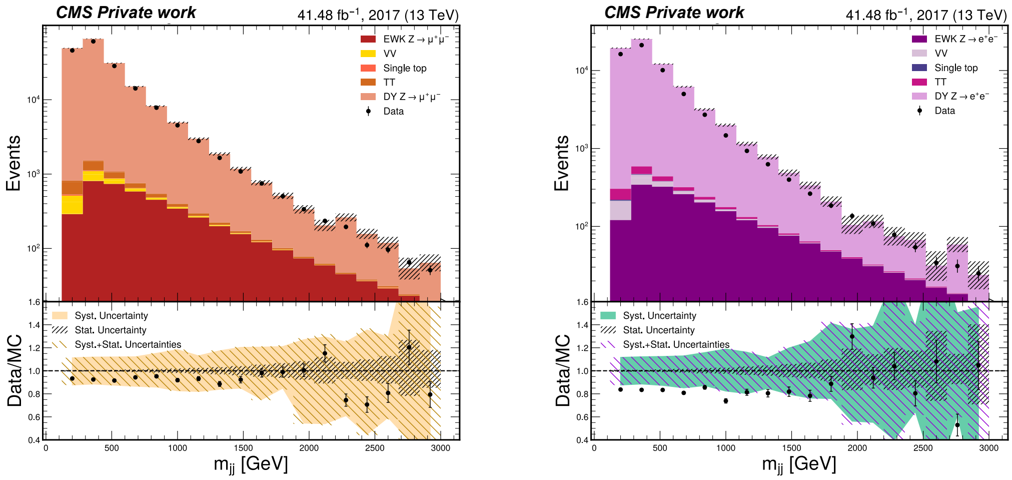


Figure B.1: Dijet mass distribution at detector level in the $\mu\mu$ channel (left) and ee channel (right), 2017.

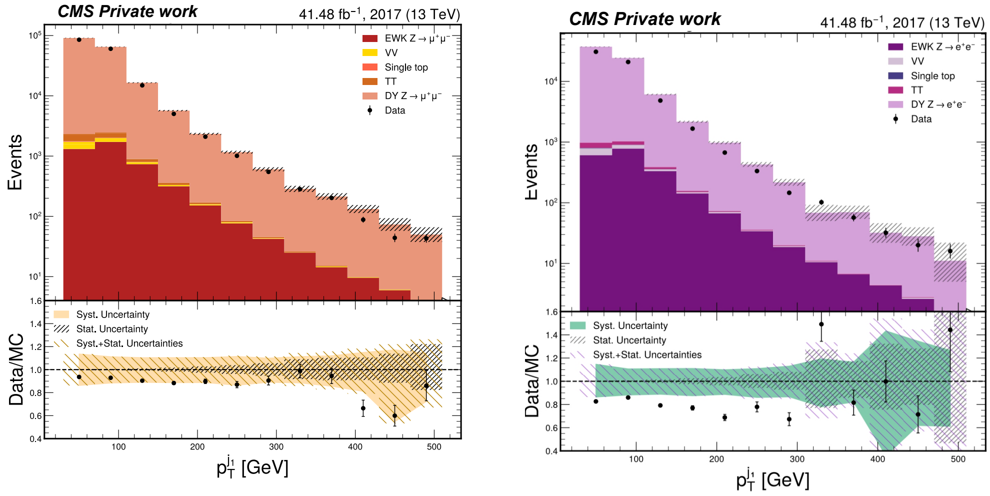


Figure B.2: Subleading jet distribution at detector level in the $\mu\mu$ channel (left) and ee channel (right), 2017.

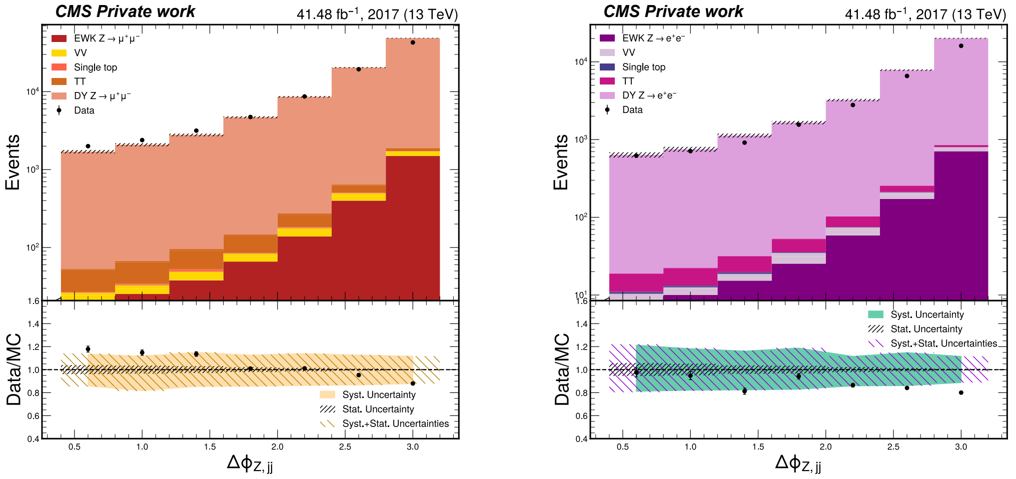


Figure B.3: $\Delta\phi_{Z,jj}$ distribution at detector level in the $\mu\mu$ channel (left) and ee channel (right), 2017.

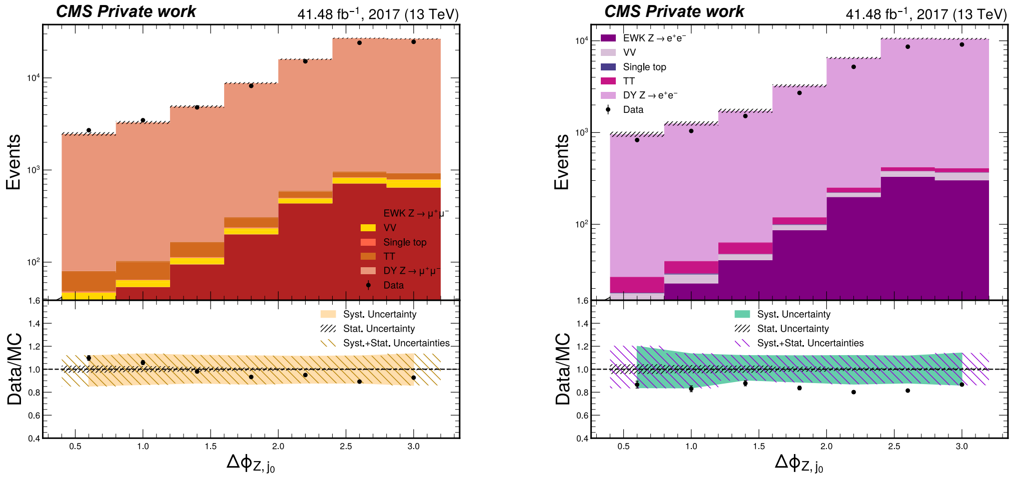


Figure B.4: $\Delta\phi_{Z,j_0}$ distribution at detector level in the $\mu\mu$ channel (left) and ee channel (right), 2017.

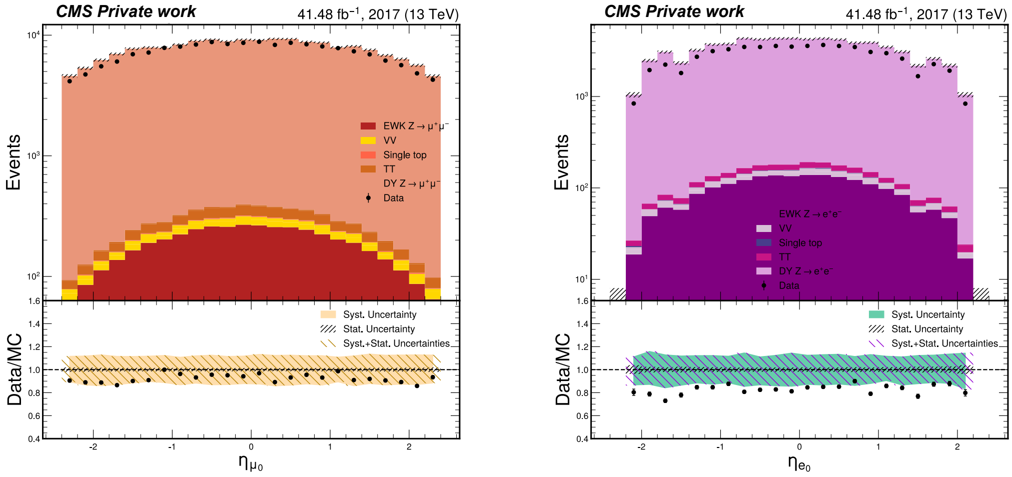


Figure B.5: Leading lepton pseudorapidity distribution at detector level in the $\mu\mu$ channel (left) and ee channel (right), 2017.

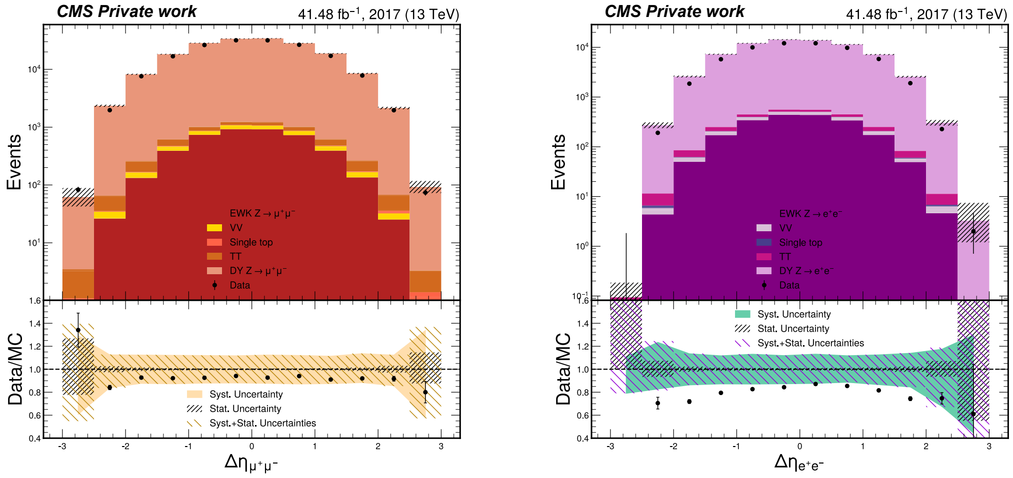


Figure B.6: $\Delta\eta_{ll}$ distribution at detector level in the $\mu\mu$ channel (left) and ee channel (right), 2017.

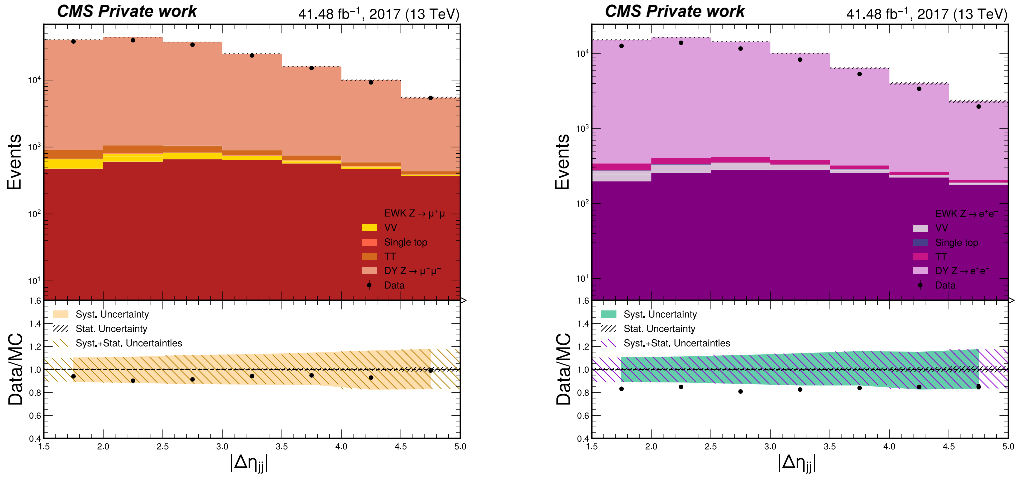


Figure B.7: $|\Delta\eta_{jj}|$ distribution at detector level in the $\mu\mu$ channel (left) and ee channel (right), 2017.

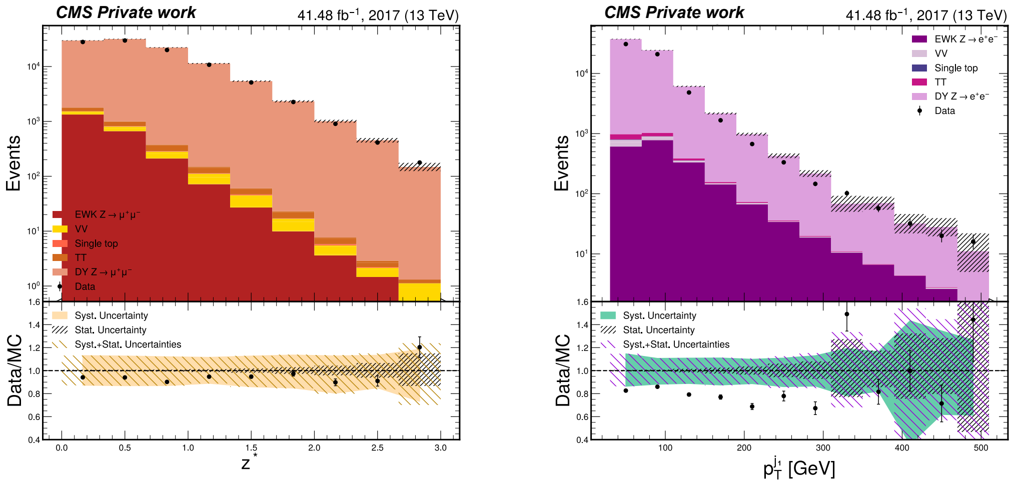


Figure B.8: Zeppenfeld variable distribution at detector level in the $\mu\mu$ channel (left) and ee channel (right), 2017.

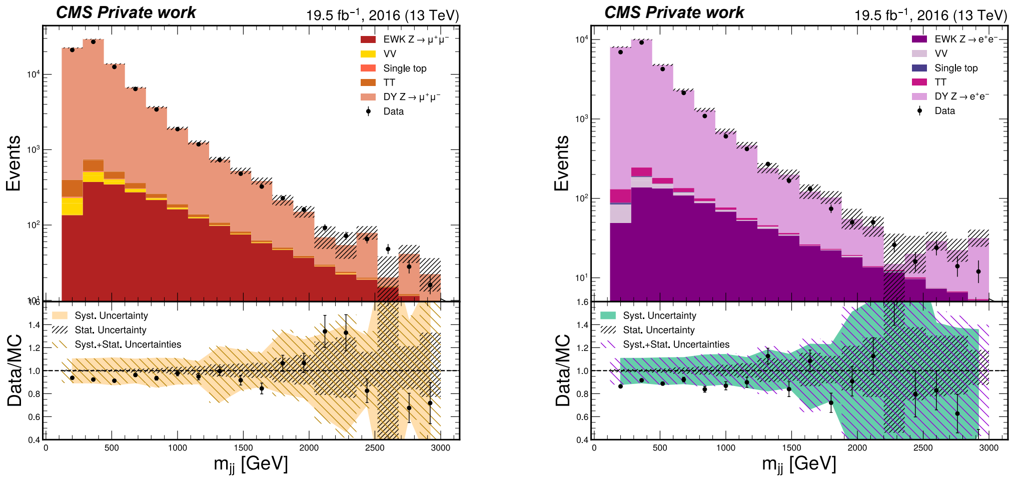


Figure B.9: Dijet mass distribution at detector level in the $\mu\mu$ channel (left) and ee channel (right), 2016preVFP.

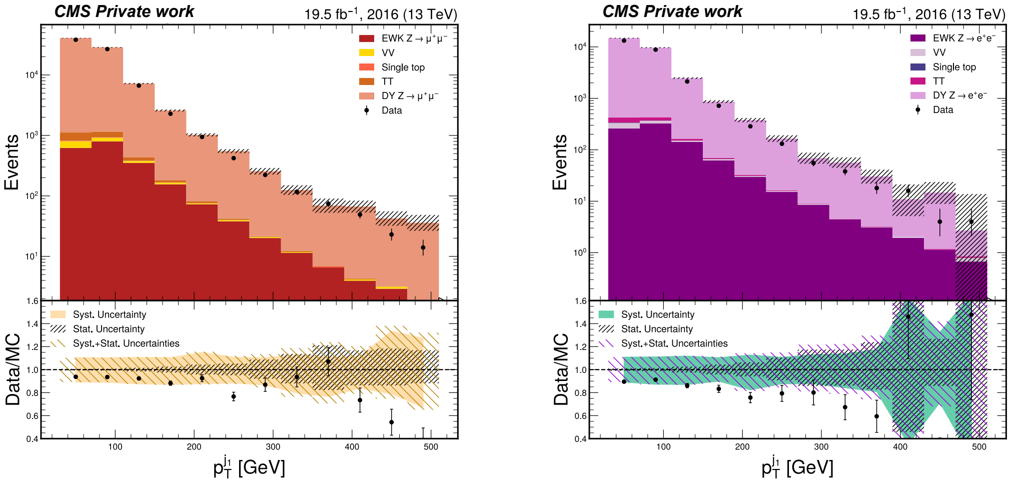


Figure B.10: Subleading jet distribution at detector level in the $\mu\mu$ channel (left) and ee channel (right), 2016preVFP.

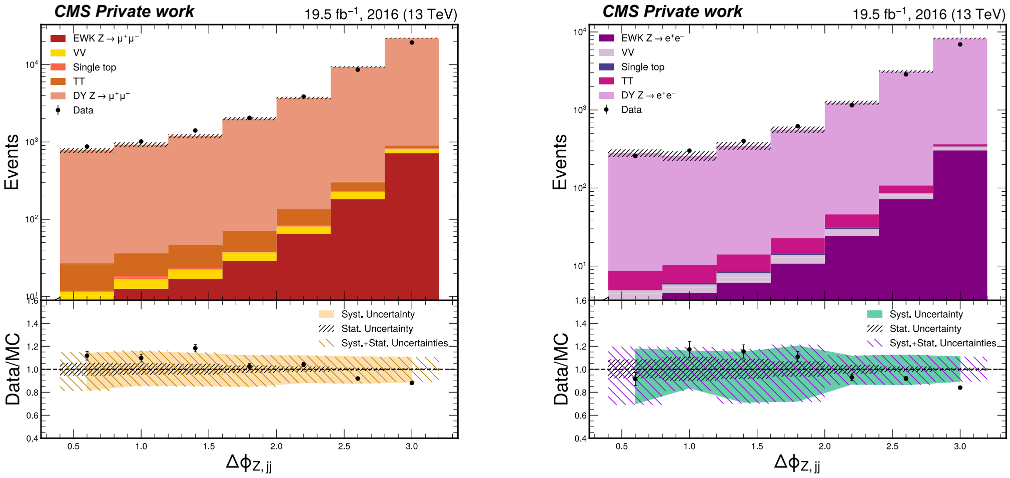


Figure B.11: $\Delta\phi_{Z,jj}$ distribution at detector level in the $\mu\mu$ channel (left) and ee channel (right), 2016preVFP.

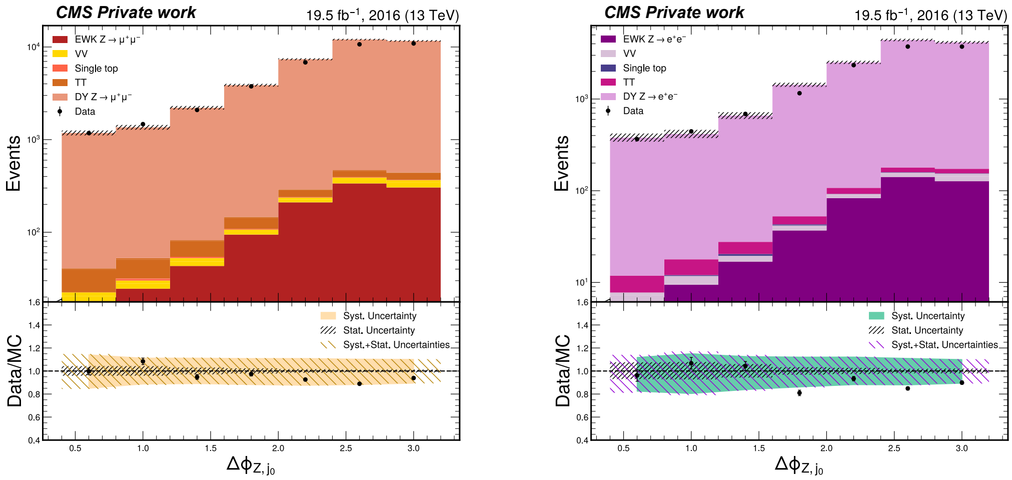


Figure B.12: $\Delta\phi_{Z,j_0}$ distribution at detector level in the $\mu\mu$ channel (left) and ee channel (right), 2016preVFP.

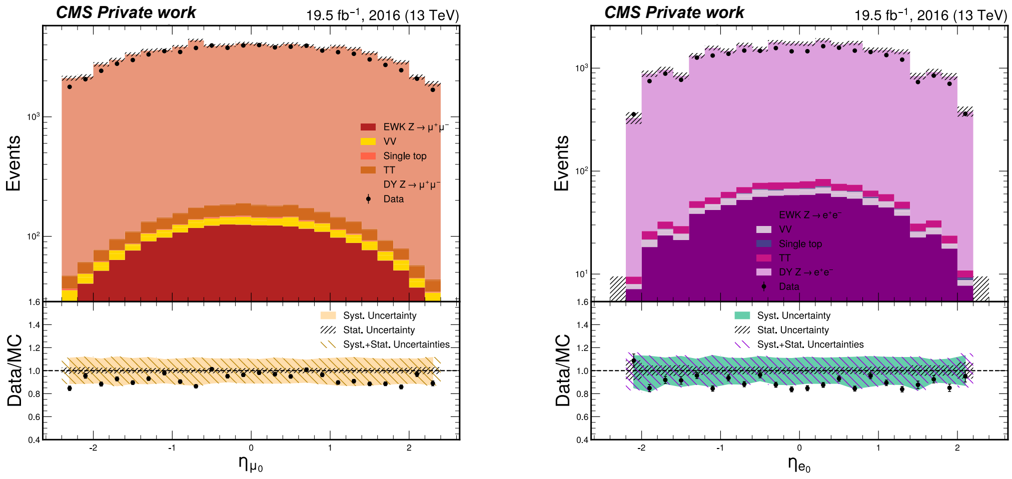


Figure B.13: Leading lepton pseudorapidity distribution at detector level in the $\mu\mu$ channel (left) and ee channel (right), 2016preVFP.

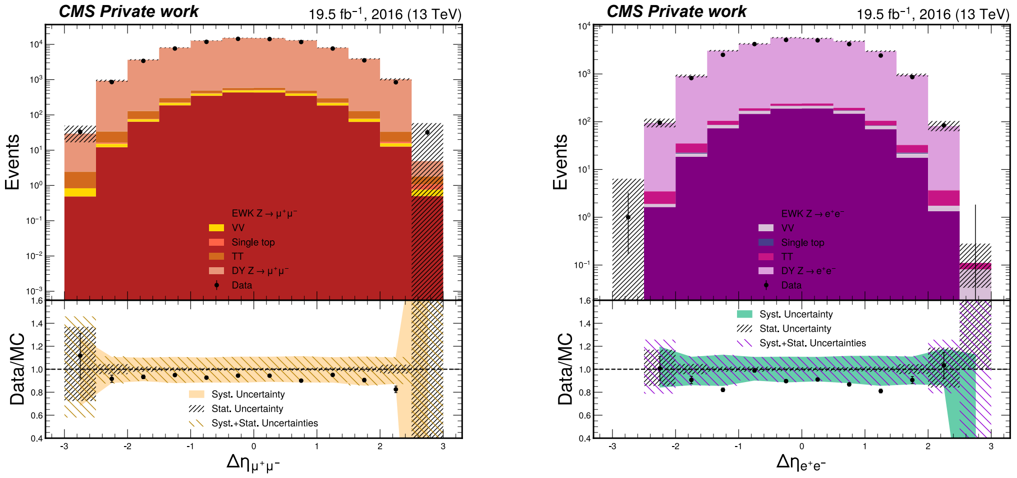


Figure B.14: $\Delta\eta_{ll}$ distribution at detector level in the $\mu\mu$ channel (left) and ee channel (right), 2016preVFP.

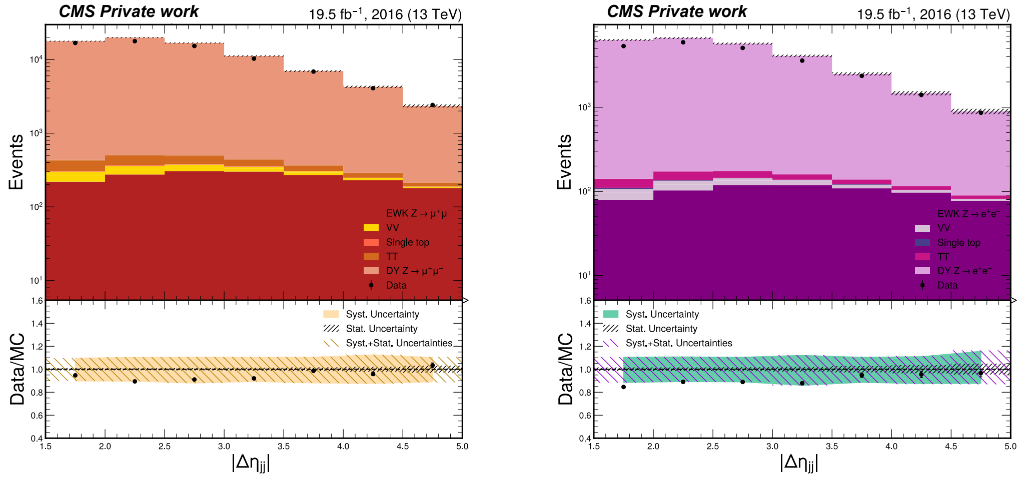


Figure B.15: $|\Delta\eta_{jj}|$ distribution at detector level in the $\mu\mu$ channel (left) and ee channel (right), 2016preVFP.

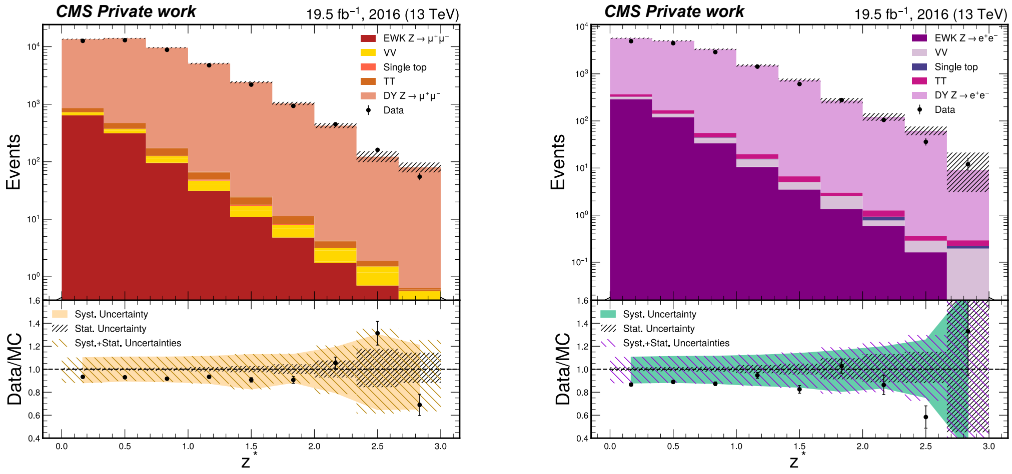


Figure B.16: Zeppenfeld variable distribution at detector level in the $\mu\mu$ channel (left) and ee channel (right), 2016preVFP.

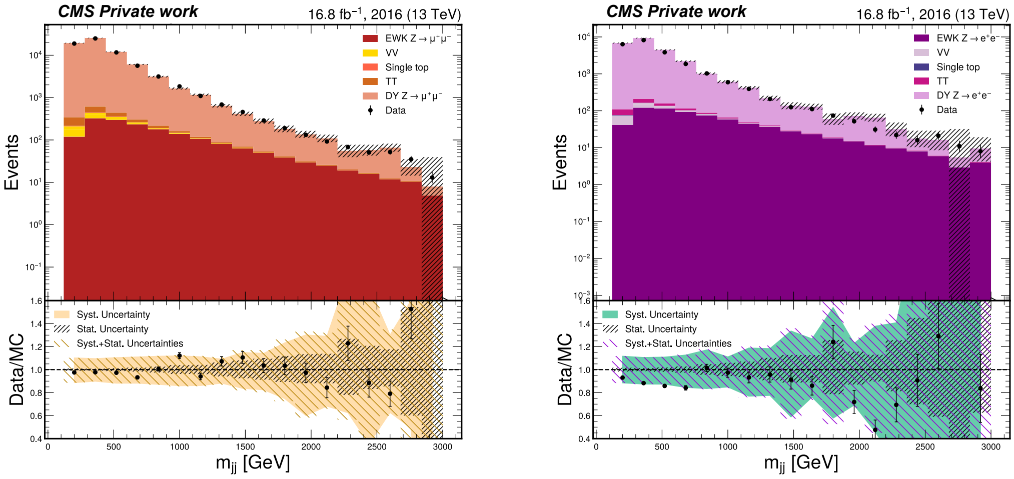


Figure B.17: Dijet mass distribution at detector level in the $\mu\mu$ channel (left) and ee channel (right), 2016postVFP.

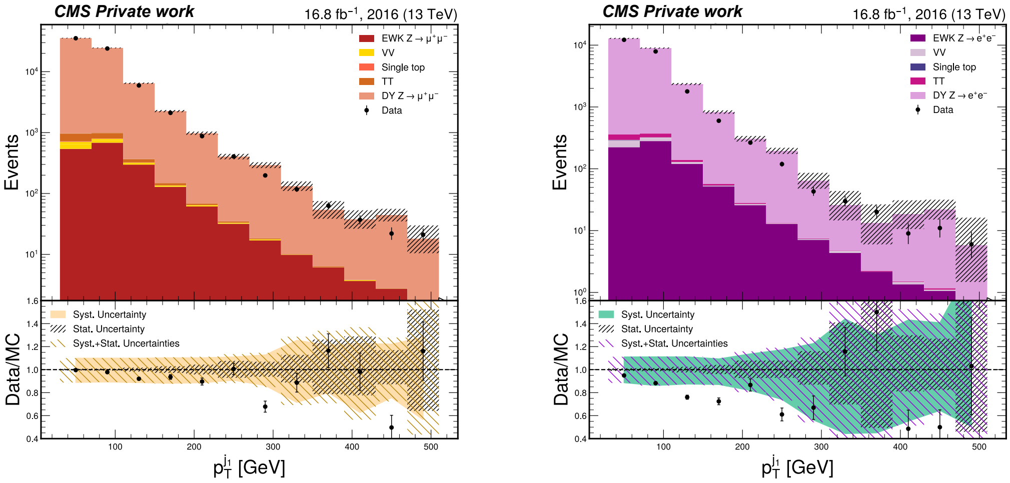


Figure B.18: Subleading jet distribution at detector level in the $\mu\mu$ channel (left) and ee channel (right), 2016postVFP.

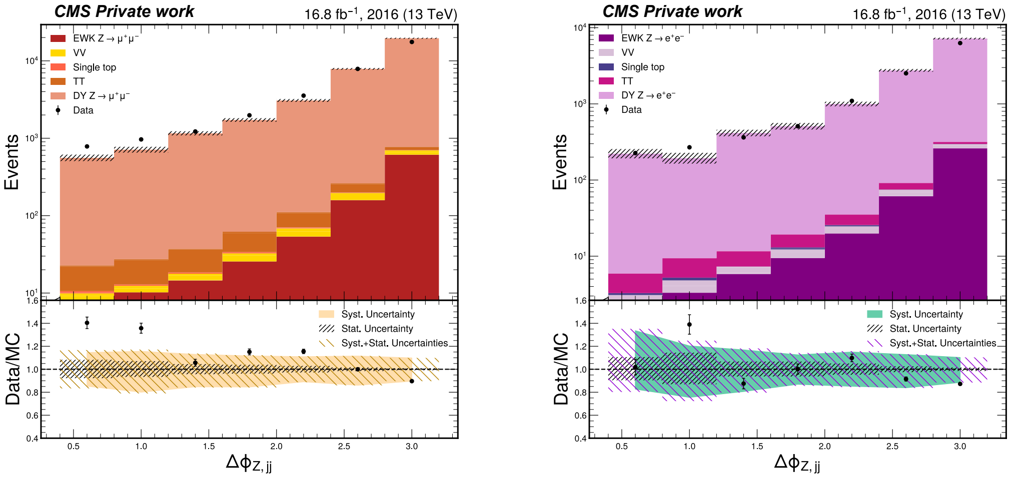


Figure B.19: $\Delta\phi_{Z,jj}$ distribution at detector level in the $\mu\mu$ channel (left) and ee channel (right), 2016postVFP.

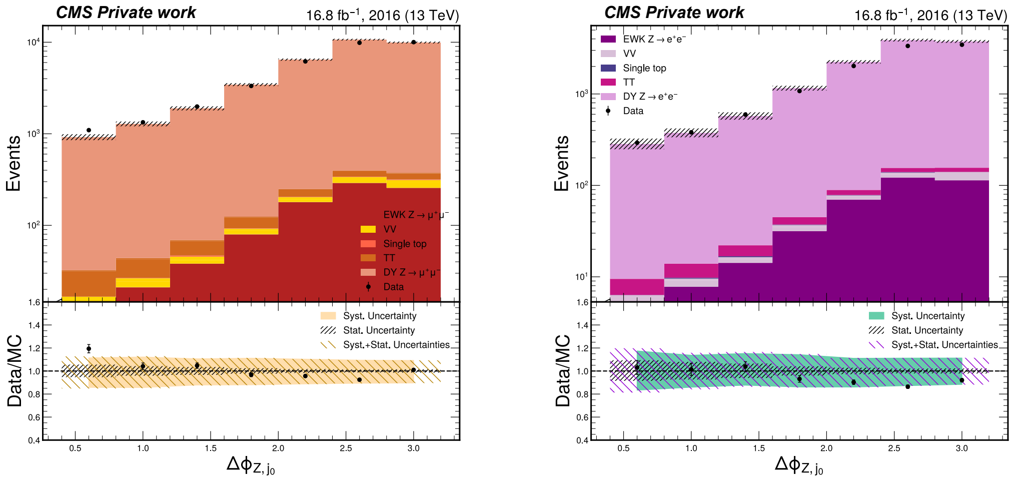


Figure B.20: $\Delta\phi_{Z,j_0}$ distribution at detector level in the $\mu\mu$ channel (left) and ee channel (right), 2016postVFP.

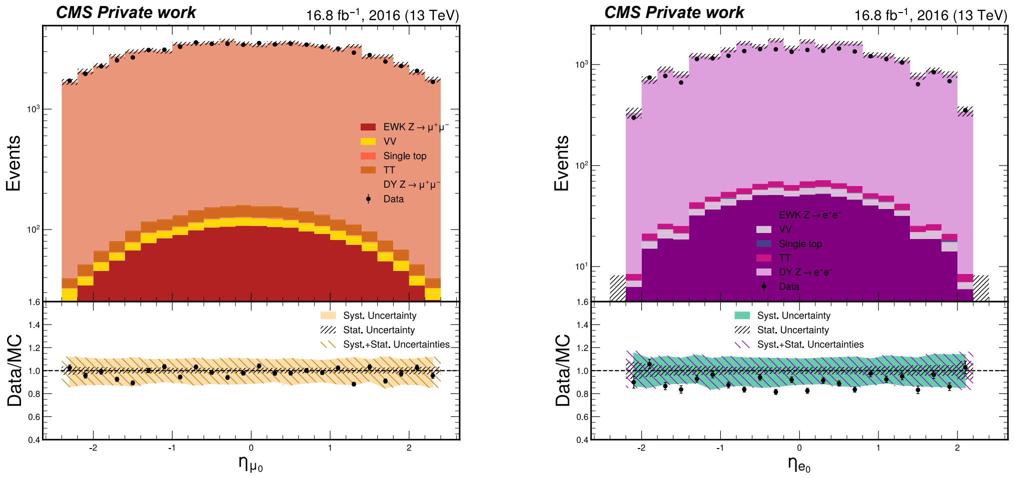


Figure B.21: Leading lepton pseudorapidity distribution at detector level in the $\mu\mu$ channel (left) and ee channel (right), 2016postVFP.

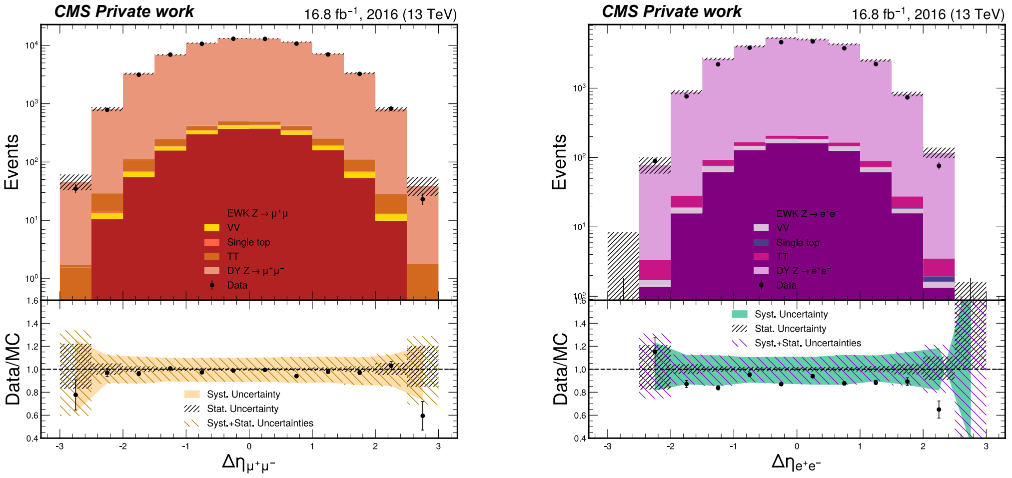


Figure B.22: $\Delta\eta_{ll}$ distribution at detector level in the $\mu\mu$ channel (left) and ee channel (right), 2016postVFP.

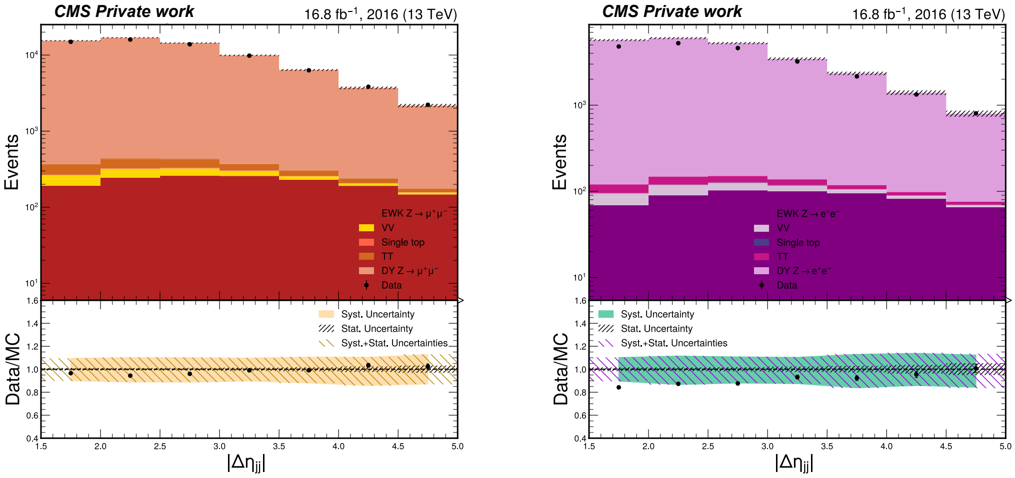


Figure B.23: $|\Delta\eta_{jj}|$ distribution at detector level in the $\mu\mu$ channel (left) and ee channel (right), 2016postVFP.

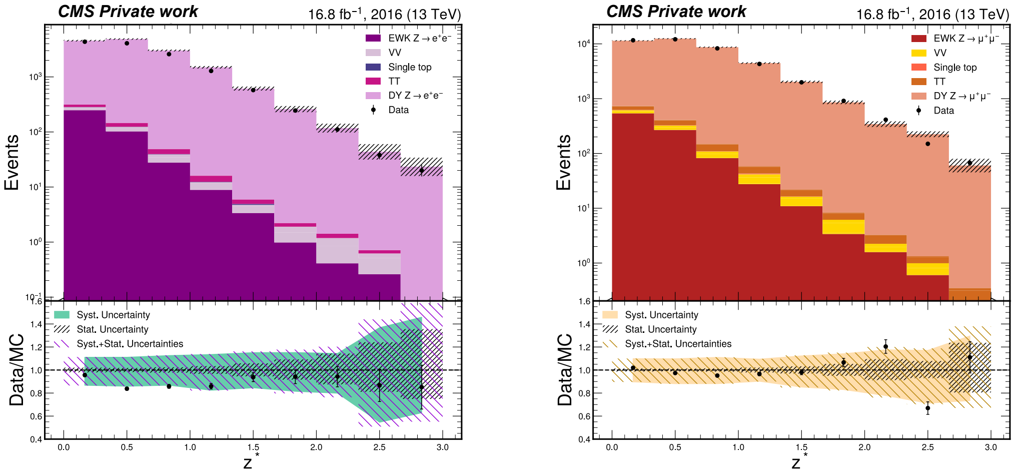


Figure B.24: Zeppenfeld variable distribution at detector level in the $\mu\mu$ channel (left) and ee channel (right), 2016postVFP.

Appendix C

Data samples

Ultra-legacy (UL) datasets and simulations in the NanoAODv9 tier have been used for the analysis. These datasets are listed in Tables C.1 and C.2, which were filtered with the Golden JSONs listed in Table C.3.

In the muon chanel, for both the 2018 and the 2017 data-taking periods, the `GT36-v1` version has been used. This is due to a significant data-to-MC disagreement observed in the UL2018 `NanoAODv9` version, this resolution loss has a geometrical origin caused by some misalignment of the muon system.

Era	Muon channel
2018	/SingleMuon/Run2018A-UL2018_MiniAODv2_NanoAODv9_GT36-v1/NANOAOD /SingleMuon/Run2018B-UL2018_MiniAODv2_NanoAODv9_GT36-v1/NANOAOD /SingleMuon/Run2018C-UL2018_MiniAODv2_NanoAODv9_GT36-v1/NANOAOD /SingleMuon/Run2018D-UL2018_MiniAODv2_NanoAODv9_GT36-v1/NANOAOD
2017	/SingleMuon/Run2017B-UL2017_MiniAODv2_NanoAODv9_GT36-v1/NANOAOD /SingleMuon/Run2017C-UL2017_MiniAODv2_NanoAODv9_GT36-v1/NANOAOD /SingleMuon/Run2017D-UL2017_MiniAODv2_NanoAODv9_GT36-v1/NANOAOD /SingleMuon/Run2017F-UL2017_MiniAODv2_NanoAODv9_GT36-v1/NANOAOD /SingleMuon/Run2017G-UL2017_MiniAODv2_NanoAODv9_GT36-v1/NANOAOD /SingleMuon/Run2017H-UL2017_MiniAODv2_NanoAODv9_GT36-v1/NANOAOD
2016preVFP	/SingleMuon/Run2016B-ver2_HIPM_UL2016_MiniAODv2_NanoAODv9-v2/NANOAOD /SingleMuon/Run2016C-HIPM_UL2016_MiniAODv2_NanoAODv9-v2/NANOAOD /SingleMuon/Run2016D-HIPM_UL2016_MiniAODv2_NanoAODv9-v2/NANOAOD /SingleMuon/Run2016E-HIPM_UL2016_MiniAODv2_NanoAODv9-v2/NANOAOD /SingleMuon/Run2016F-HIPM_UL2016_MiniAODv2_NanoAODv9-v2/NANOAOD
2016postVFP	/SingleMuon/Run2016F-UL2016_MiniAODv2_NanoAODv9-v1/NANOAOD /SingleMuon/Run2016G-UL2016_MiniAODv2_NanoAODv9-v1/NANOAOD /SingleMuon/Run2016H-UL2016_MiniAODv2_NanoAODv9-v1/NANOAOD

Table C.1: Ultralegacy data samples used in the muon channel for the full Run 2.

Era	Electron channel

2018	`/EGamma/Run2018B-UL2018_MiniAODv2_NanoAODv9-v1/NANOAOD`
	`/EGamma/Run2018C-UL2018_MiniAODv2_NanoAODv9-v1/NANOAOD`
	`/EGamma/Run2018D-UL2018_MiniAODv2_NanoAODv9-v1/NANOAOD`
	`/EGamma/Run2018E-UL2018_MiniAODv2_NanoAODv9-v1/NANOAOD`

2017	`/SingleElectron/Run2017B-UL2017_MiniAODv2_NanoAODv9-v1/NANOAOD`
	`/SingleElectron/Run2017C-UL2017_MiniAODv2_NanoAODv9-v1/NANOAOD`
	`/SingleElectron/Run2017D-UL2017_MiniAODv2_NanoAODv9-v1/NANOAOD`
	`/SingleElectron/Run2017E-UL2017_MiniAODv2_NanoAODv9-v1/NANOAOD`

2016preVFP	`/SingleElectron/Run2016B-ver2_HIPM_UL2016_MiniAODv2_NanoAODv9-v2/NANOAOD`
	`/SingleElectron/Run2016C-HIPM_UL2016_MiniAODv2_NanoAODv9-v2/NANOAOD`
	`/SingleElectron/Run2016D-HIPM_UL2016_MiniAODv2_NanoAODv9-v2/NANOAOD`
	`/SingleElectron/Run2016E-HIPM_UL2016_MiniAODv2_NanoAODv9-v2/NANOAOD`
	`/SingleElectron/Run2016F-HIPM_UL2016_MiniAODv2_NanoAODv9-v2/NANOAOD`

2016postVFP	`/SingleElectron/Run2016F-UL2016_MiniAODv2_NanoAODv9-v1/NANOAOD`
	`/SingleElectron/Run2016G-UL2016_MiniAODv2_NanoAODv9-v1/NANOAOD`
	`/SingleElectron/Run2016H-UL2016_MiniAODv2_NanoAODv9-v1/NANOAOD`

Table C.2: Ultralegacy data samples used in the electron channel for the full Run 2.

Era	Golden Lumisections JSON	Integrated luminosity (fb^{-1})
2018	`Cert_314472-325175_13TeV_Legacy2018_Collisions18_JSON.txt`	59.83
2017	`Cert_294927-306462_13TeV_UL2017_Collisions17_GoldenJSON.txt`	41.48
2016preVFP	`Cert_271036-284044_13TeV_Legacy2016_Collisions16_JSON.txt`	19.5
2016postVFP	`Cert_271036-284044_13TeV_Legacy2016_Collisions16_JSON.txt`	16.8

Table C.3: Certified Golden JSON files used to filter datasets

Appendix D

Simulation samples

Simulation sample	σ (fb)
`/EWK_LLJJ_MLL-50_MJJ-120_TuneCP5_13TeV-madgraph-pythia8_dipole/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOAODSIM`	1719
`/DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v2/NANOAODSIM`	6077220
`/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOAODSIM`	88290
`/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOAODSIM`	366290
`/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v2/NANOAODSIM`	34910
`/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v2/NANOAODSIM`	34970
`/ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOAODSIM`	10320
`/ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOAODSIM`	136020
`/ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOAODSIM`	80950
`/WW_TuneCP5_13TeV-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOAODSIM`	75800
`/WZ_TuneCP5_13TeV-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOAODSIM`	27600
`/ZZ_TuneCP5_13TeV-pythia8/RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOAODSIM`	12140

Table D.1: Monte Carlo samples and cross sections used for the 2018 data-taking period.

Simulation sample	σ (fb)
/EWK_LLJJ_MLL-50_MJJ-120_TuneCP5_13TeV-madgraph-pythia8_dipole /RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1 /NANOAODSIM	1719
/DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8 /RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v2 /NANOAODSIM	6077220
/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8 /RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1 /NANOAODSIM	88290
/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8 /RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1 /NANOAODSIM	365350
/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8 /RunIISummer20UL17NanoAODv9-106X_upgrade2017_realistic _v16_L1v1-v2/NANOAODSIM	34910
/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8 /RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v2 /NANOAODSIM	34910
/ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-amcatnlo-pythia8 /RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1 /NANOAODSIM	3549
/ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg -madspin-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017 _realistic_v9-v1/NANOAODSIM	113400
/ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg -madspin-pythia8/RunIISummer20UL17NanoAODv9-106X_mc2017 _realistic_v9-v1/NANOAODSIM	67930
/WW_TuneCP5_13TeV-pythia8 /RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1 /NANOAODSIM	76250
/WZ_TuneCP5_13TeV-pythia8 /RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1 /NANOAODSIM	27550
/ZZ_TuneCP5_13TeV-pythia8 /RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1 /NANOAODSIM	12230

Table D.2: Monte Carlo samples and cross sections used for the 2017 data-taking period.

Simulation sample	σ (fb)
`/EWK_LLJJ_MLL-50_MJJ-120_TuneCP5_13TeV-madgraph-pythia8_dipole/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOAODSIM`	1719
`/DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOAODSIM`	6077220
`/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOAODSIM`	88290
`/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOAODSIM`	365350
`/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOAODSIM`	34910
`/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOAODSIM`	34910
`/ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOAODSIM`	3549
`/ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOAODSIM`	113400
`/ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOAODSIM`	67930
`/WW_TuneCP5_13TeV-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOAODSIM`	75950
`/WZ_TuneCP5_13TeV-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_V11-v1/NANOAODSIM`	27590
`/ZZ_TuneCP5_13TeV-pythia8/RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOAODSIM`	12170

Table D.3: Monte Carlo samples and cross sections used for the 2016preVFP data-taking period.

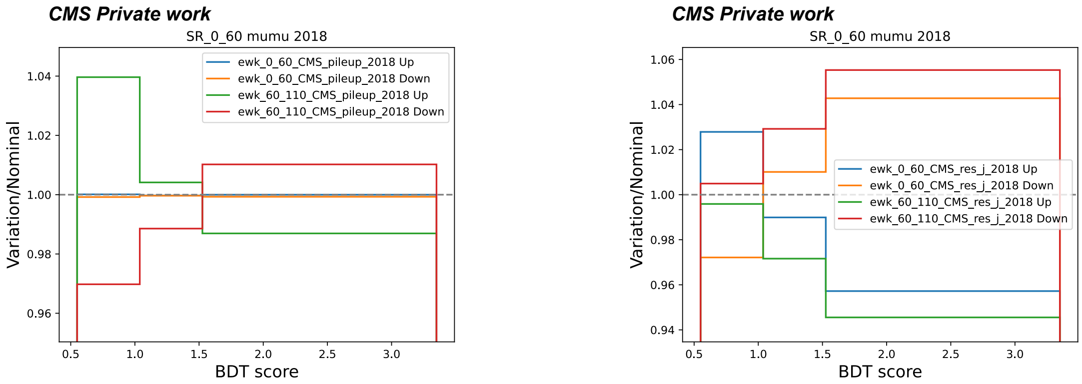
Simulation sample	σ (fb)
/EWK_LLJJ_MLL-50_MJJ-120_TuneCP5_13TeV-madgraph-pythia8_dipole/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM	1719
/DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM	6077220
/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM	88290
/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM	365350
/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v2/NANOAODSIM	34910
/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v2/NANOAODSIM	34910
/ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-amcatnlo-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM	3549
/ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM	113400
/ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM	67930
/WW_TuneCP5_13TeV-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM	75950
/WZ_TuneCP5_13TeV-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM	27590
/ZZ_TuneCP5_13TeV-pythia8/RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v/NANOAODSIM	12170

Table D.4: Monte Carlo samples and cross sections used for the 2016postVFP data-taking period.

Appendix E

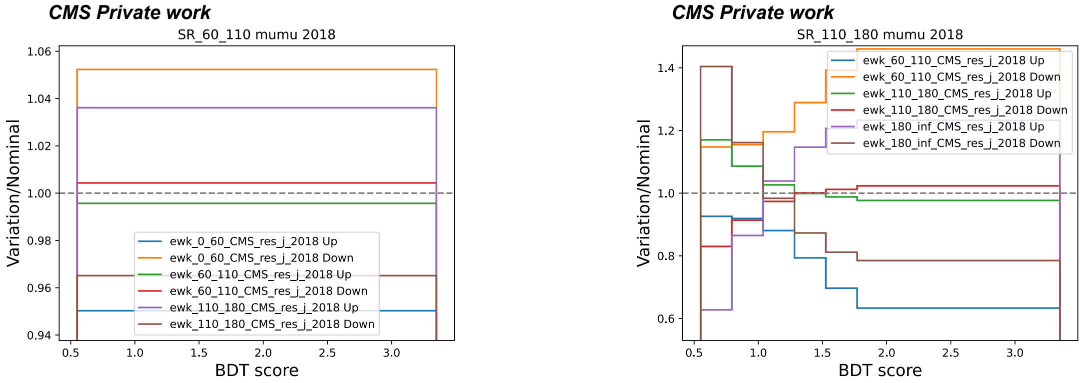
Uncertainties breakdown

This appendix provides a detailed breakdown of the uncertainties contributing to the tBDT distribution. For brevity, only the 2018 data corresponding to the $\mu\mu$ channel are presented, and only uncertainties exceeding 1% are included. The distributions in this appendix refer to the tBDT score as “BDT score” for simplicity.



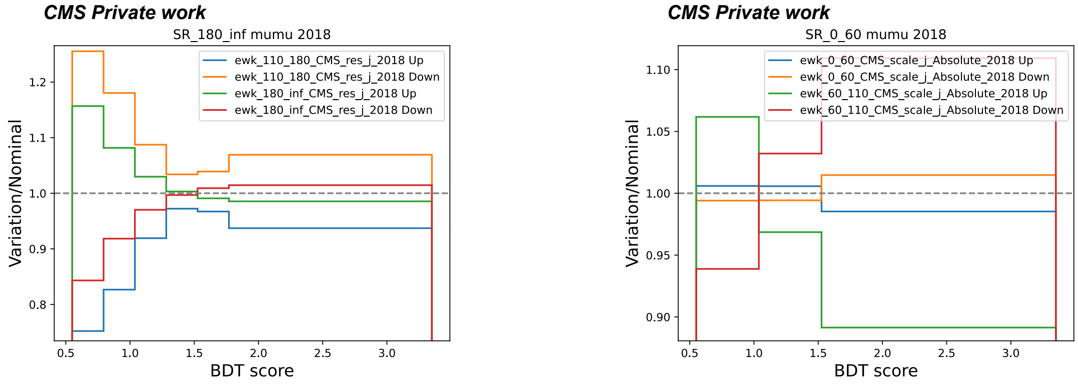
(a) Pileup 2018, $0 < p_T^{ll} < 60$ GeV

(b) Jet resolution 2018, $0 < p_T^{ll} < 60$ GeV



(c) Jet resolution 2018, $60 < p_T^{ll} < 110$ GeV

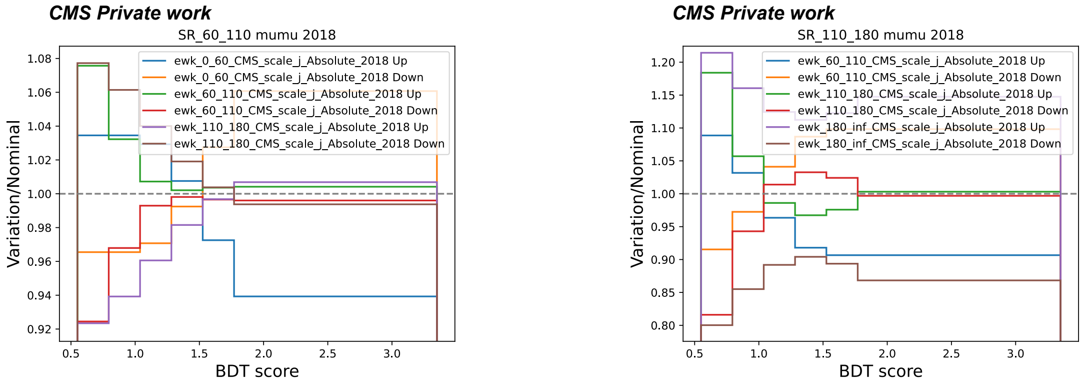
(d) Jet resolution 2018, $110 < p_T^{ll} < 180$ GeV



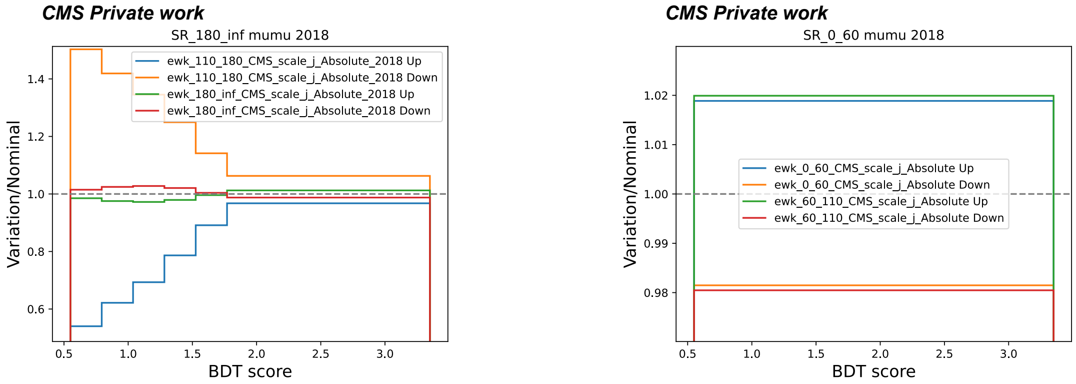
(e) Jet resolution 2018, $p_T^{ll} > 180$ GeV

(f) Absolute 2018, $0 < p_T^{ll} < 60$ GeV

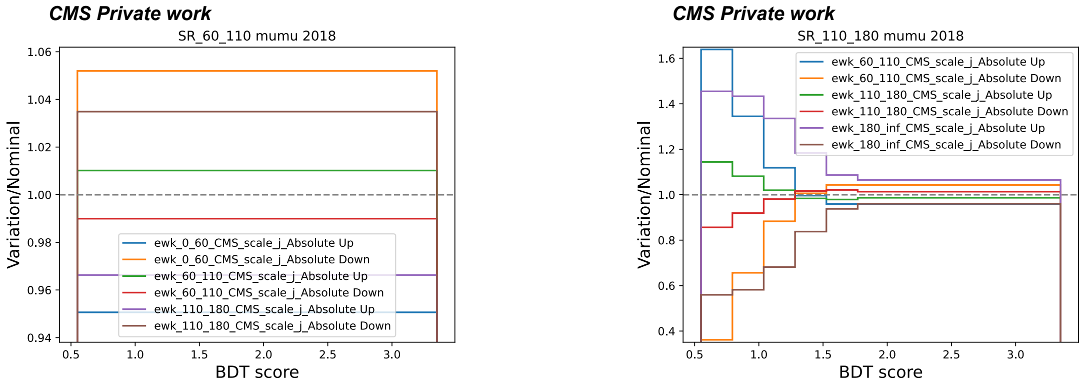
Figure E.1: Uncertainties breakdown for $\mu\mu$ channel, 2018



(a) Absolute 2018, $60 < p_T^{ll} < 110$ GeV (b) Absolute 2018, $110 < p_T^{ll} < 180$ GeV

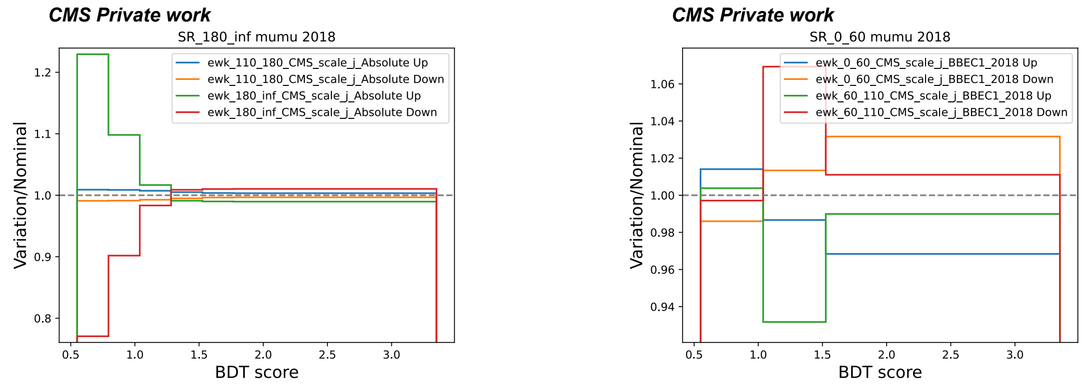


(c) Absolute 2018, $p_T^{ll} > 180$ GeV (d) Absolute, $0 < p_T^{ll} < 60$ GeV

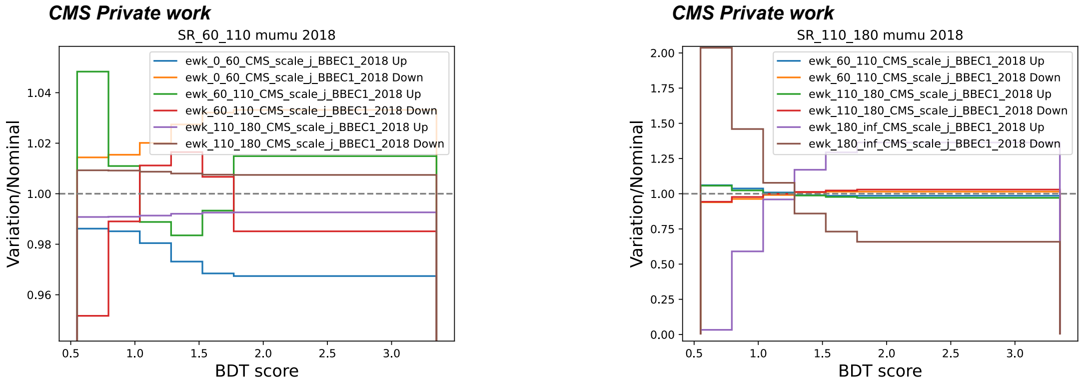


(e) Absolute, $60 < p_T^{ll} < 110$ GeV (f) Absolute, $110 < p_T^{ll} < 180$ GeV

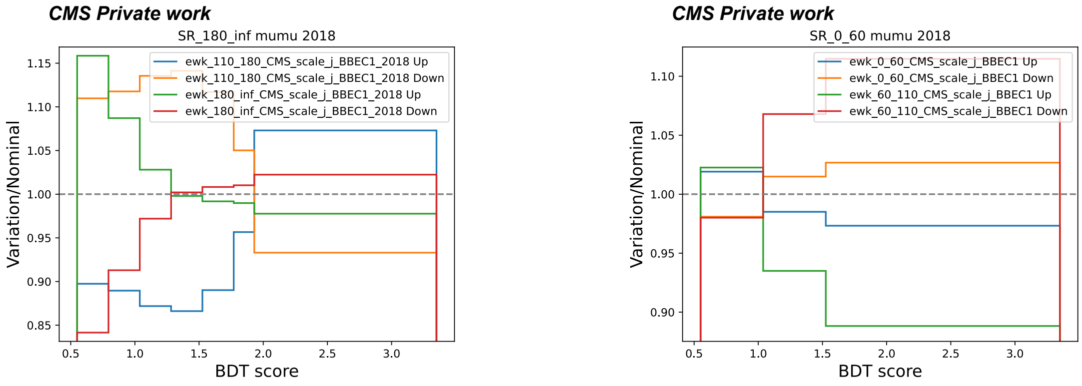
Figure E.2: Uncertainties breakdown for $\mu\mu$ channel, 2018



(a) Absolute, $p_T^{ll} > 180$ GeV (b) BBEC1 2018, $0 < p_T^{ll} < 60$ GeV

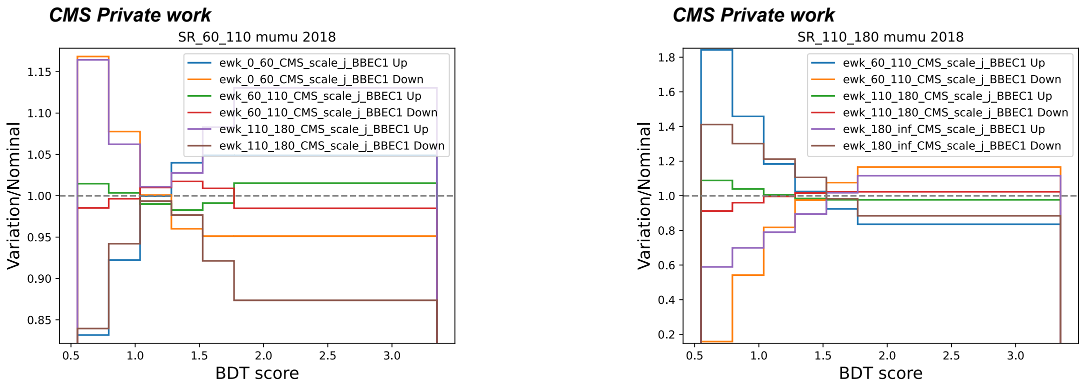


(c) BBEC1 2018, $60 < p_T^{ll} < 110$ GeV (d) BBEC1 2018, $110 < p_T^{ll} < 180$ GeV



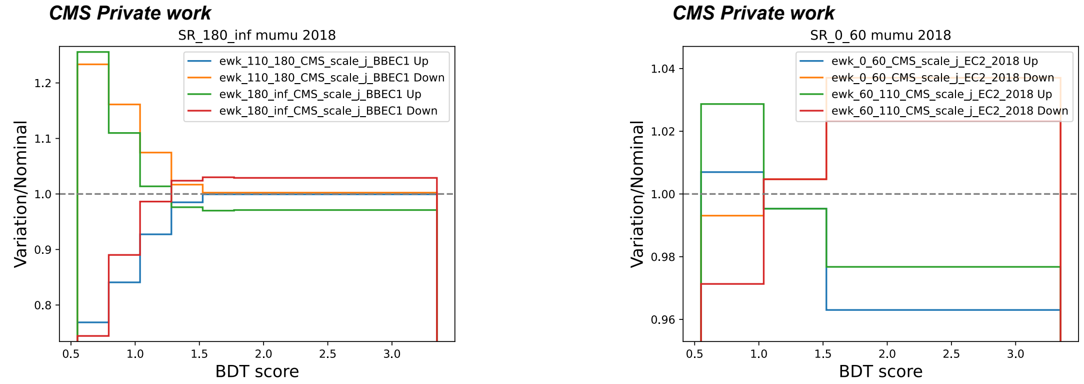
(e) BBEC1 2018, $p_T^{ll} > 180$ GeV (f) BBEC1, $0 < p_T^{ll} < 60$ GeV

Figure E.3: Uncertainties breakdown for $\mu\mu$ channel, 2018



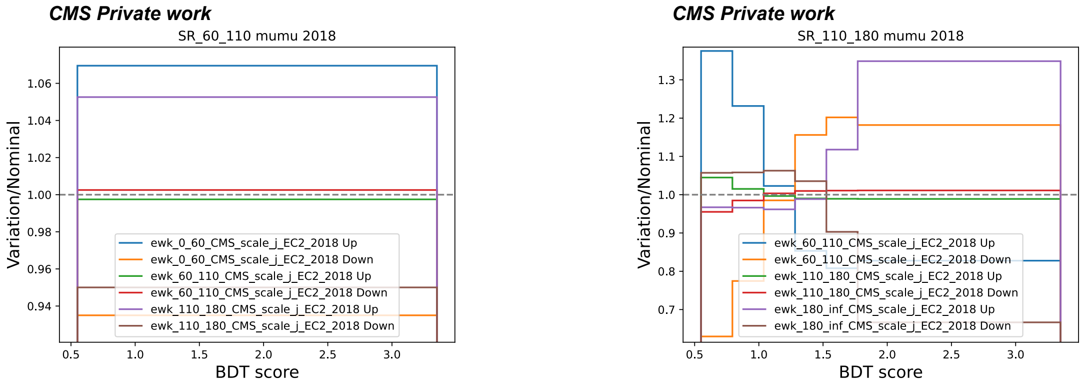
(a) BBEC1, $60 < p_T^{ll} < 110$ GeV

(b) BBEC1, $110 < p_T^{ll} < 180$ GeV



(c) BBEC1, $p_T^{ll} > 180$ GeV

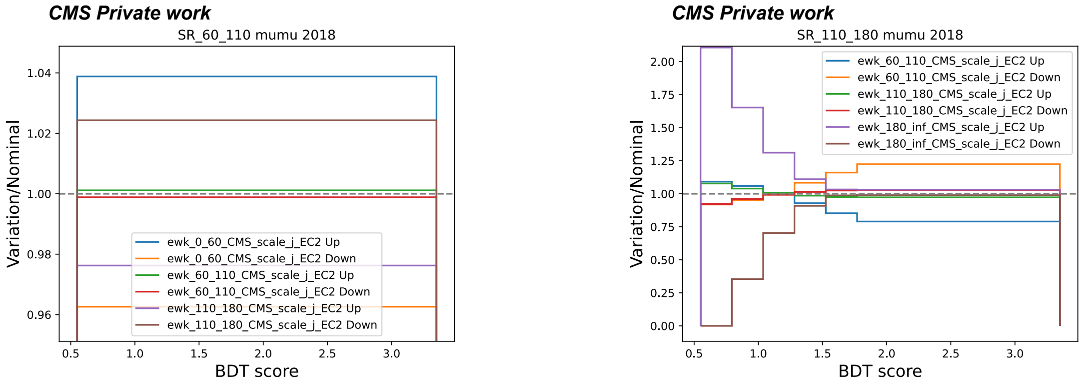
(d) EC2 2018, $0 < p_T^{ll} < 60$ GeV



(e) EC2 2018, $60 < p_T^{ll} < 110$ GeV

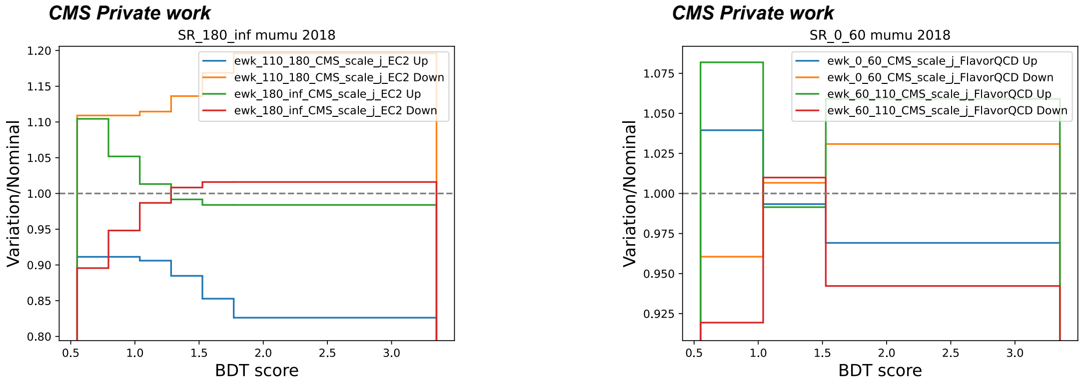
(f) EC2 2018, $110 < p_T^{ll} < 180$ GeV

Figure E.4: Uncertainties breakdown for $\mu\mu$ channel, 2018



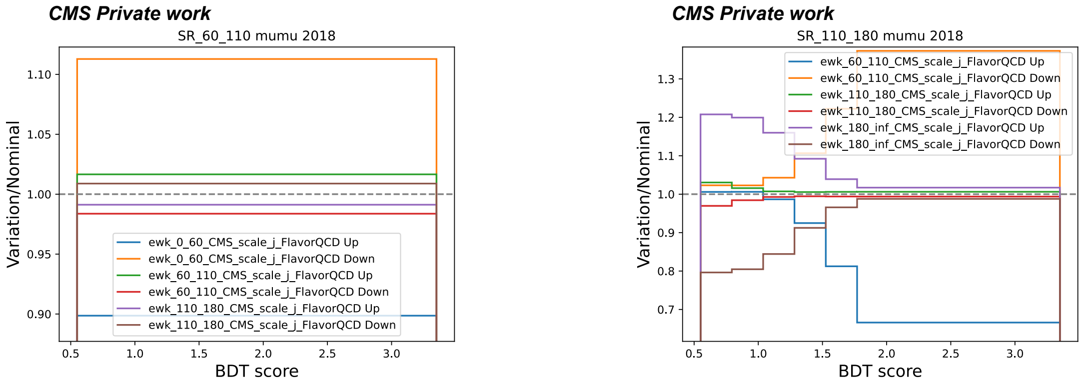
(a) EC2, $60 < p_T^{ll} < 110$ GeV

(b) EC2, $110 < p_T^{ll} < 180$ GeV



(c) EC2, $p_T^{ll} > 180$ GeV

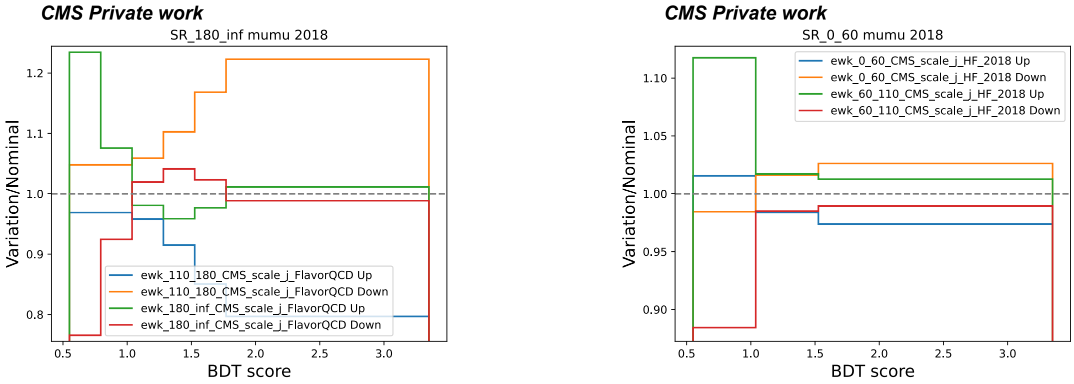
(d) FlavorQCD, $0 < p_T^{ll} < 60$ GeV



(e) FlavorQCD, $60 < p_T^{ll} < 110$ GeV

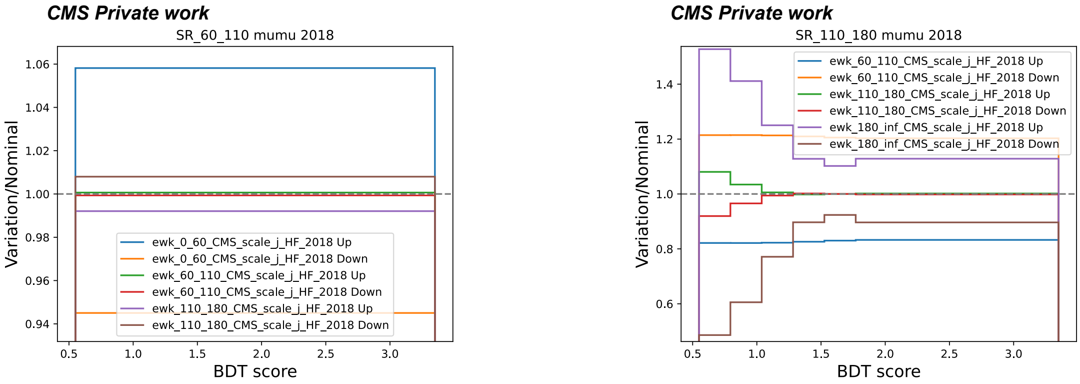
(f) FlavorQCD, $110 < p_T^{ll} < 180$ GeV

Figure E.5: Uncertainties breakdown for $\mu\mu$ channel, 2018



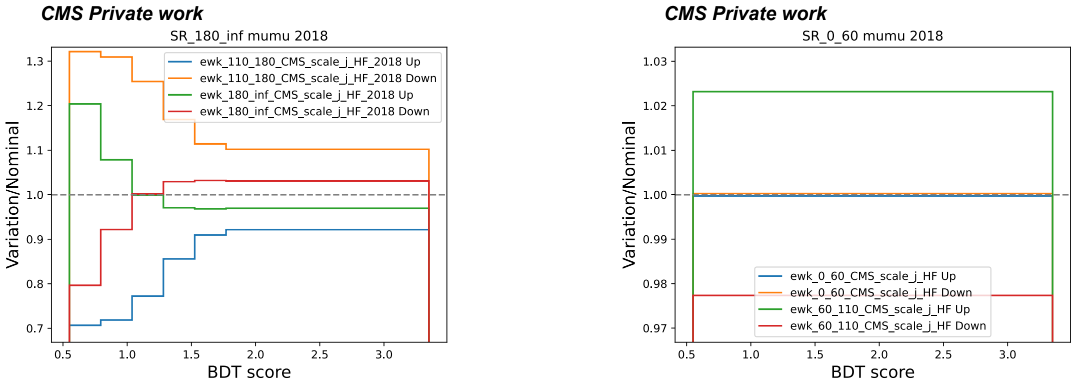
(a) FlavorQCD, $p_T^{ll} > 180$ GeV

(b) HF 2018, $0 < p_T^{ll} < 60$ GeV



(c) HF 2018, $60 < p_T^{ll} < 110$ GeV

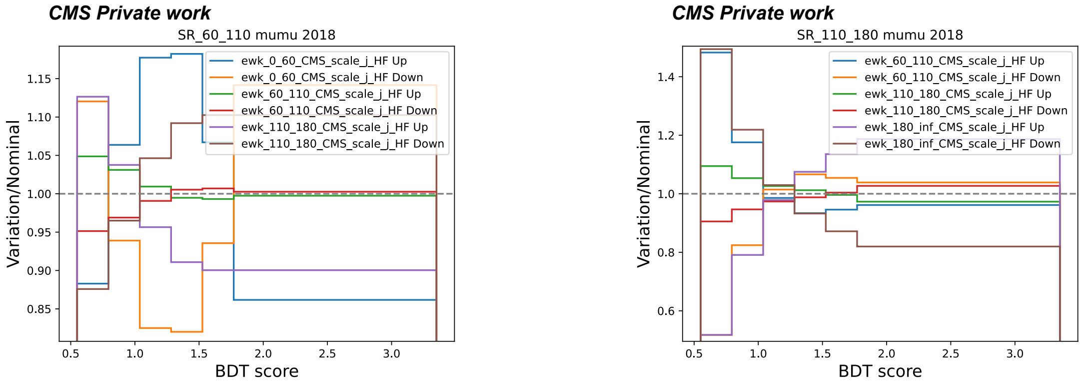
(d) HF 2018, $110 < p_T^{ll} < 180$ GeV



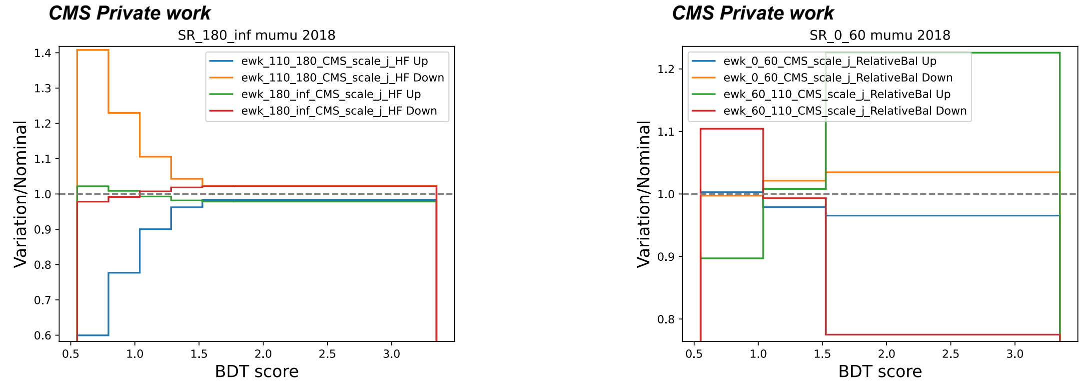
(e) HF 2018, $p_T^{ll} > 180$ GeV

(f) HF, $0 < p_T^{ll} < 60$ GeV

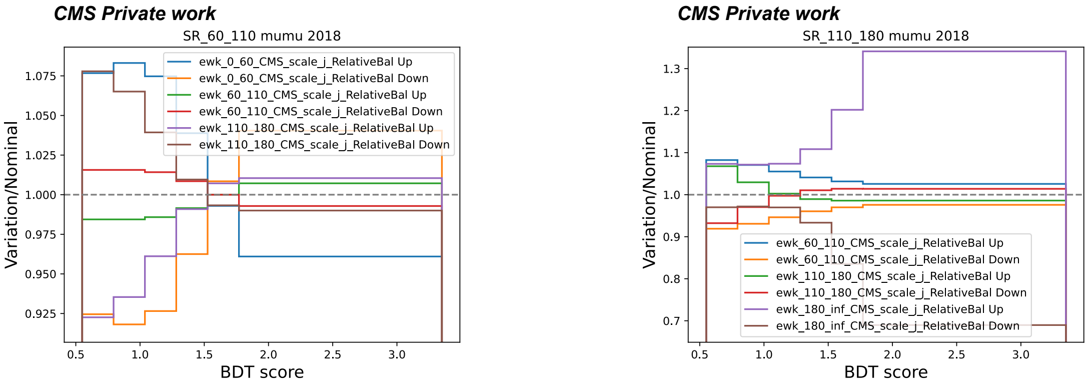
Figure E.6: Uncertainties breakdown for $\mu\mu$ channel, 2018



(a) HF, $60 < p_T^{ll} < 110$ GeV (b) HF, $110 < p_T^{ll} < 180$ GeV

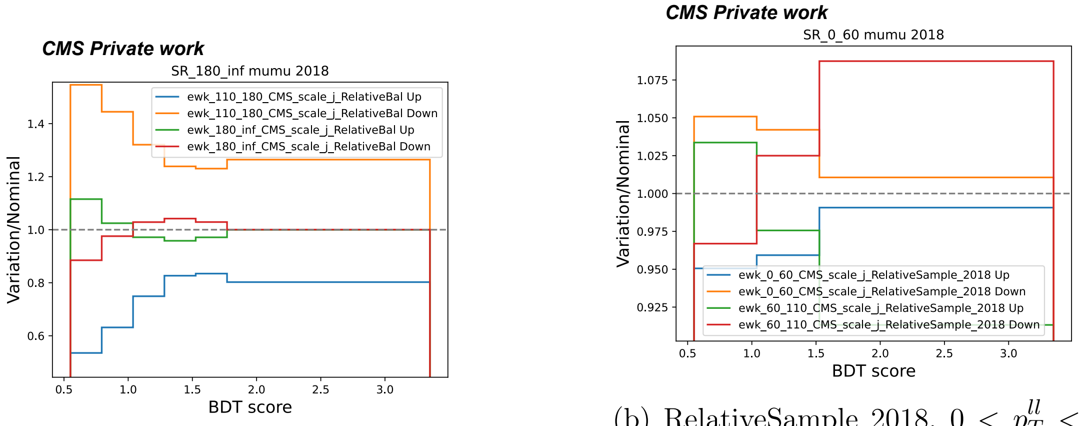


(c) HF, $p_T^{ll} > 180$ GeV (d) RelativeBal, $0 < p_T^{ll} < 60$ GeV



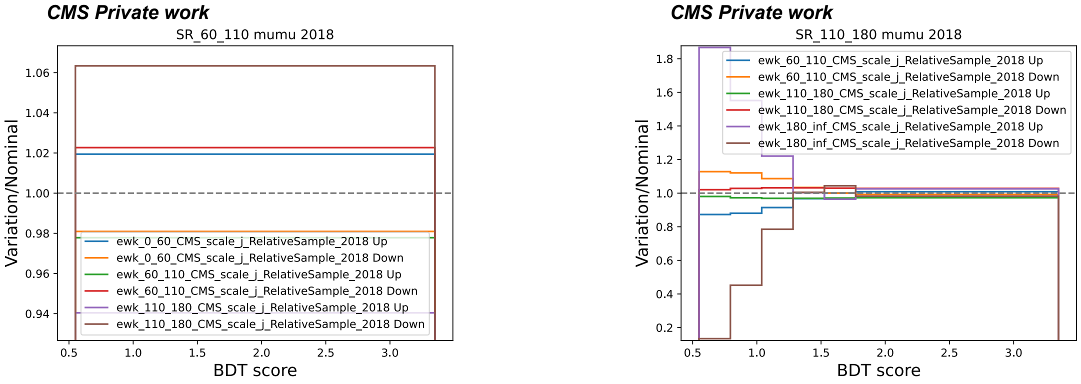
(e) RelativeBal, $60 < p_T^{ll} < 110$ GeV (f) RelativeBal, $110 < p_T^{ll} < 180$ GeV

Figure E.7: Uncertainties breakdown for $\mu\mu$ channel, 2018



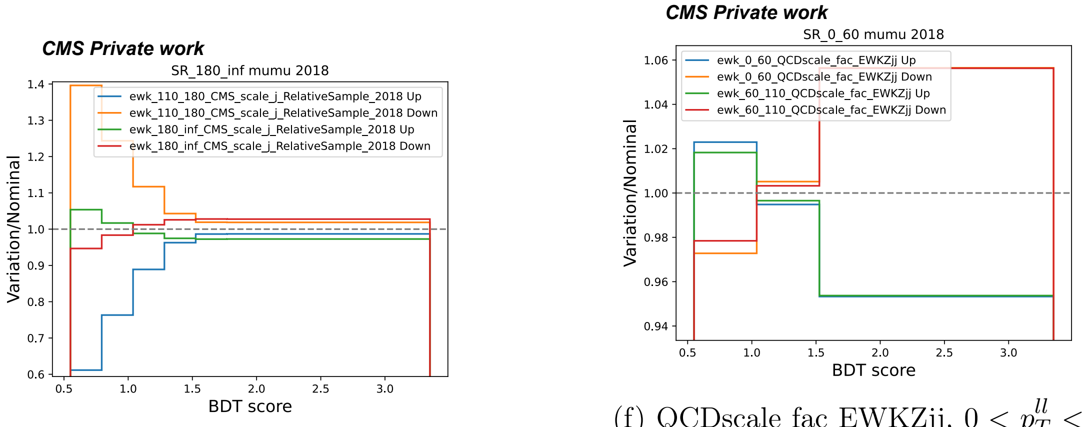
(a) RelativeBal, $p_T^{ll} > 180$ GeV

(b) RelativeSample 2018, $0 < p_T^{ll} < 60$ GeV



(c) RelativeSample 2018, $60 < p_T^{ll} < 110$ GeV

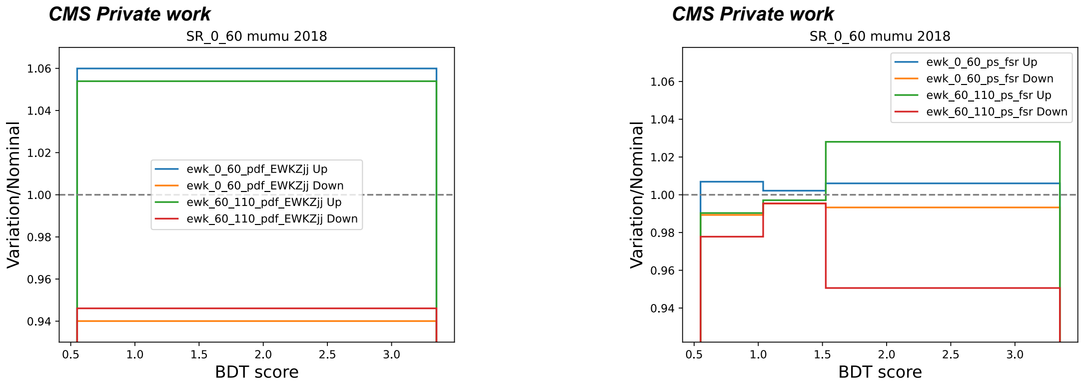
(d) RelativeSample 2018, $110 < p_T^{ll} <$ 180 GeV



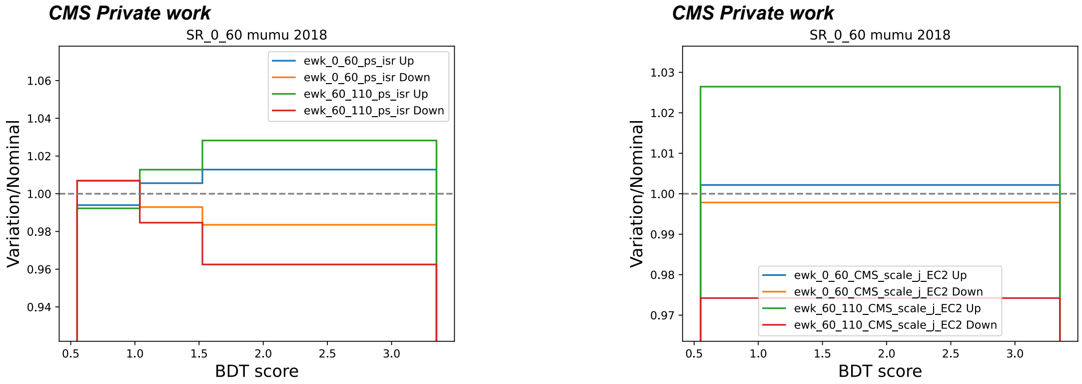
(e) RelativeSample 2018, $p_T^{ll} > 180$ GeV

(f) QCDscale fac EWKZjj, $0 < p_T^{ll} < 60$ GeV

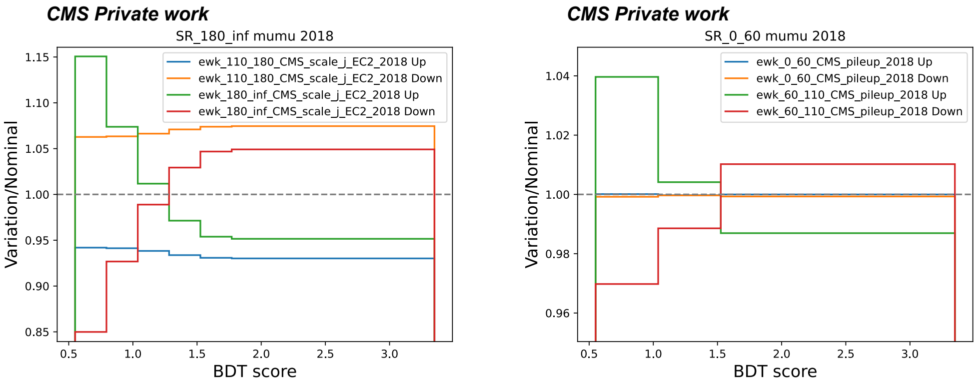
Figure E.8: Uncertainties breakdown for $\mu\mu$ channel, 2018



(a) PDF EWKZjj, $0 < p_T^{ll} < 60$ GeV (b) PS FSR, $0 < p_T^{ll} < 60$ GeV

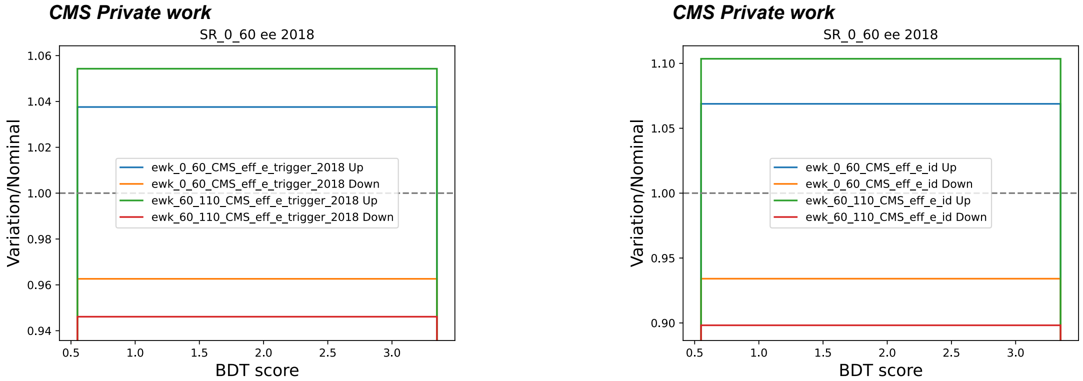


(c) PS ISR, $0 < p_T^{ll} < 60$ GeV (d) EC2, $0 < p_T^{ll} < 60$ GeV



(e) EC2 2018, $p_T^{ll} > 180$ GeV (f) Pileup, $0 < p_T^{ll} < 60$ GeV

Figure E.9: Uncertainties breakdown for $\mu\mu$ channel, 2018



(a) Electron trigger efficiency, $0 < p_T^{ll} < 60$ GeV

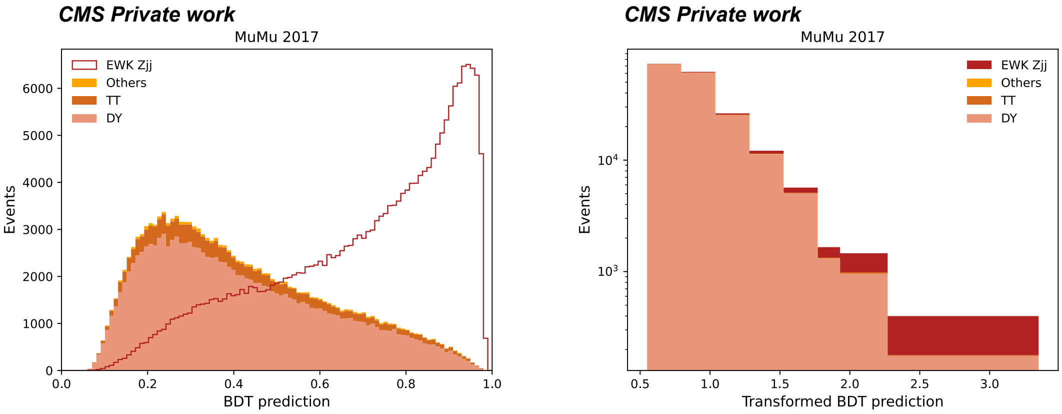
(b) Electron ID efficiency, $0 < p_T^{ll} < 60$ GeV

Figure E.10: Uncertainties breakdown for $\mu\mu$ channel, 2018

Appendix F

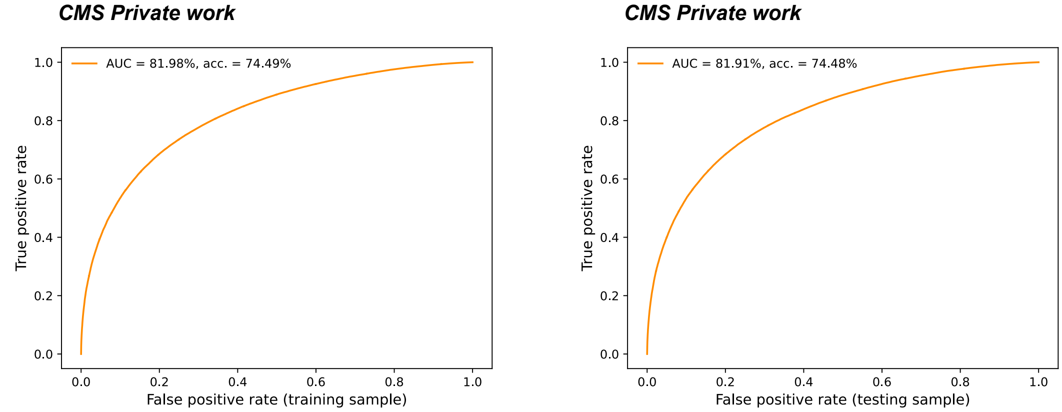
BDT Supplementary Material

F.1 BDT and tBDT distributions and AUC



(a) BDT evaluation output.

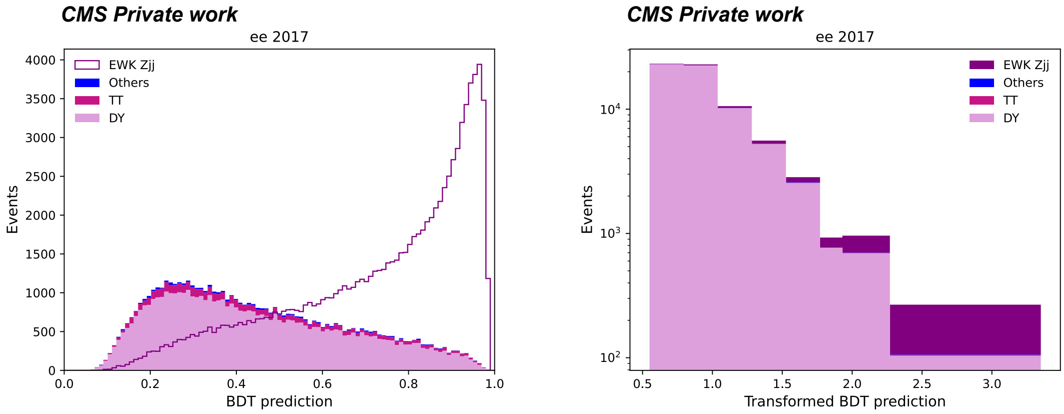
(b) tBDT evaluation output.



(c) AUC values for training.

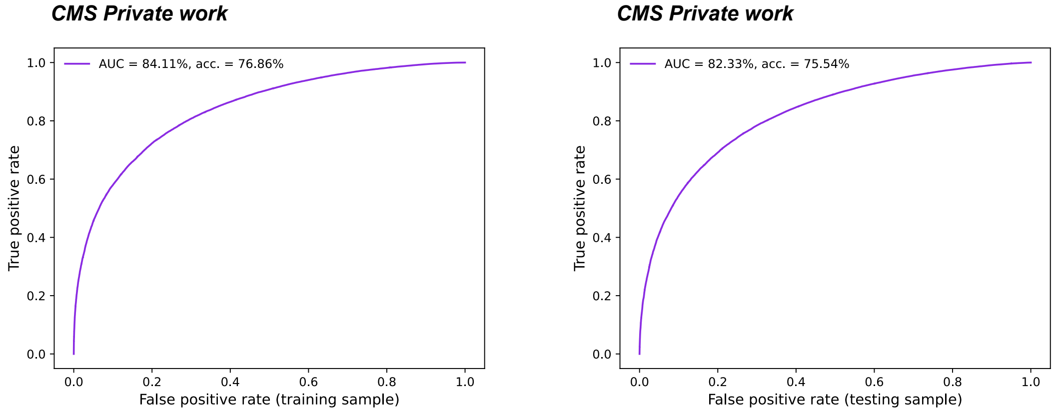
(d) AUC values for testing

Figure F.1: Complementary BDT distributions for $\mu\mu$ channel, 2017.



(a) BDT evaluation output.

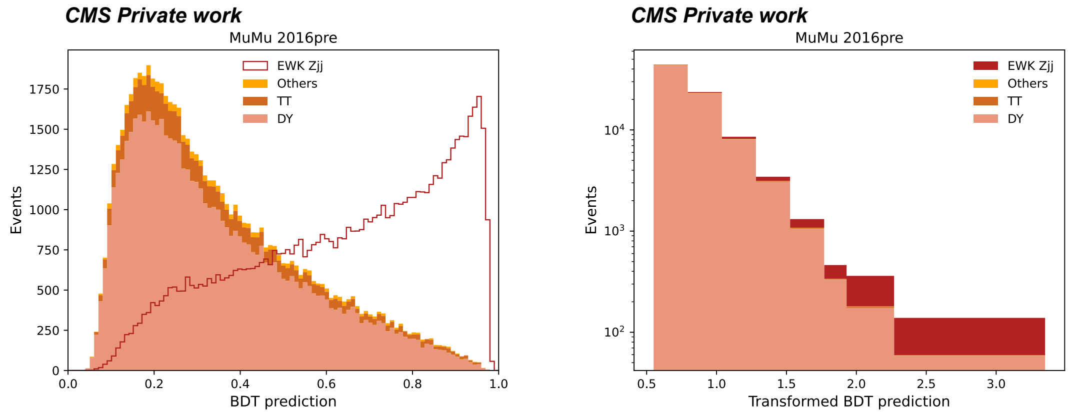
(b) tBDT evaluation output.



(c) AUC values for training.

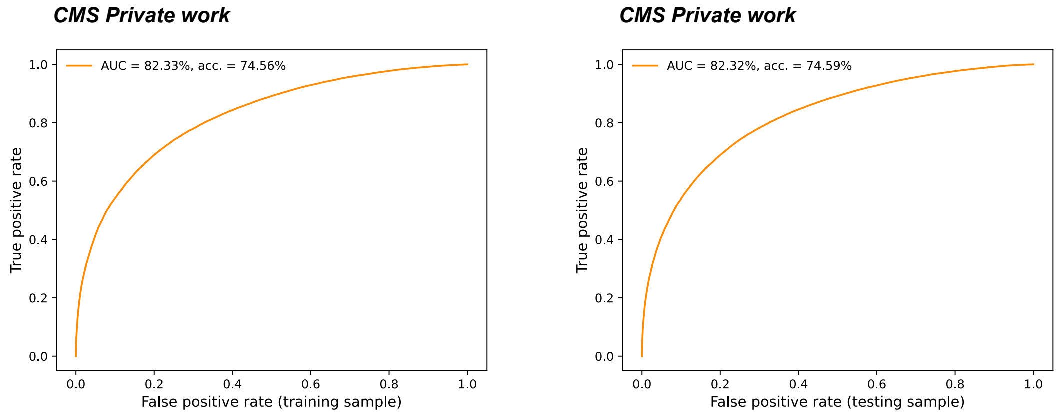
(d) AUC values for testing

Figure F.2: Complementary BDT distributions for *ee* channel, 2017.



(a) BDT evaluation output.

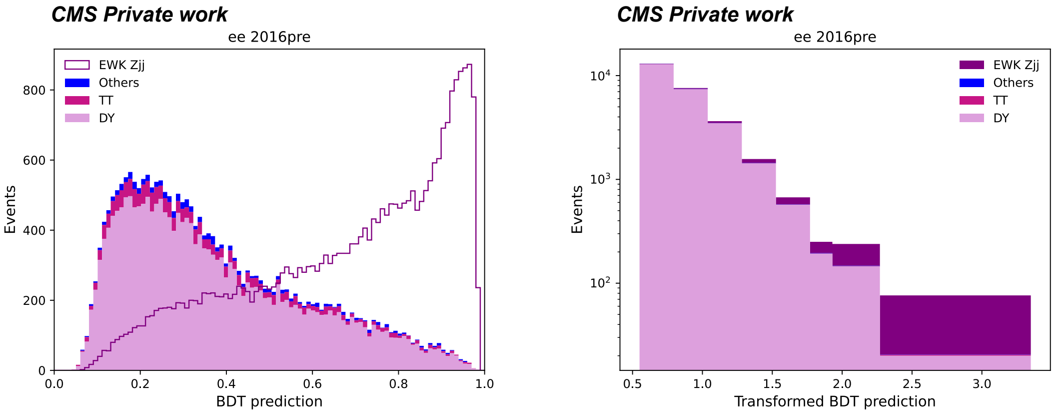
(b) tBDT evaluation output.



(c) AUC values for training.

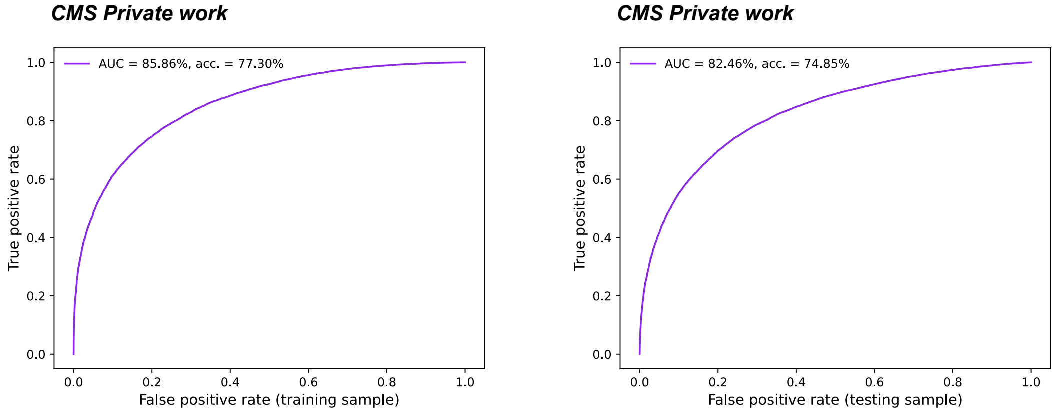
(d) AUC values for testing

Figure F.3: Complementary BDT distributions for $\mu\mu$ channel, 2016preVFP.



(a) BDT evaluation output.

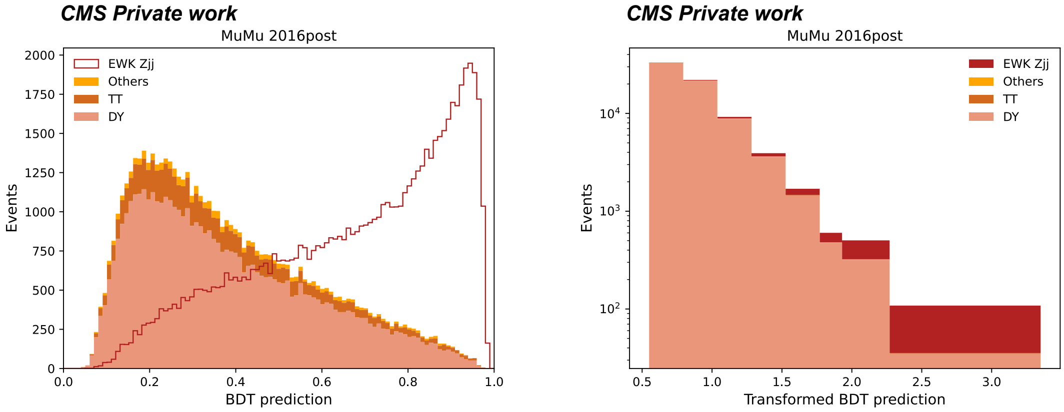
(b) tBDT evaluation output.



(c) AUC values for training.

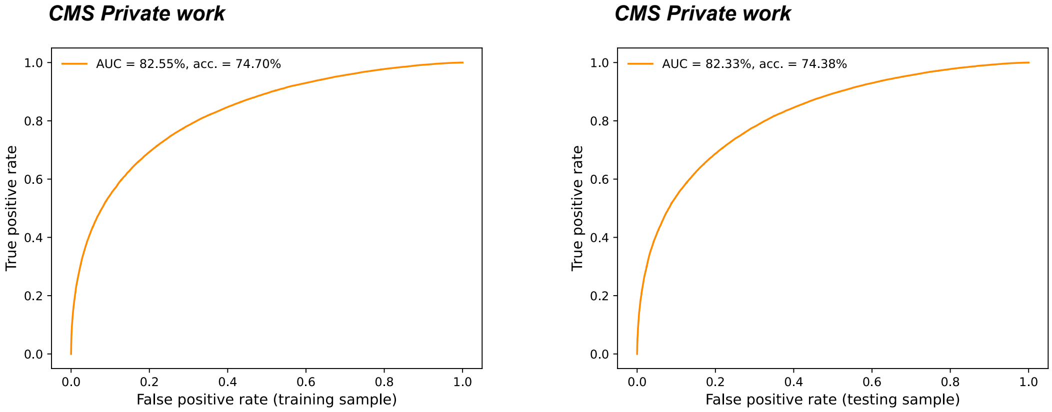
(d) AUC values for testing

Figure F.4: Complementary BDT distributions for *ee* channel, 2016preVFP.



(a) BDT evaluation output.

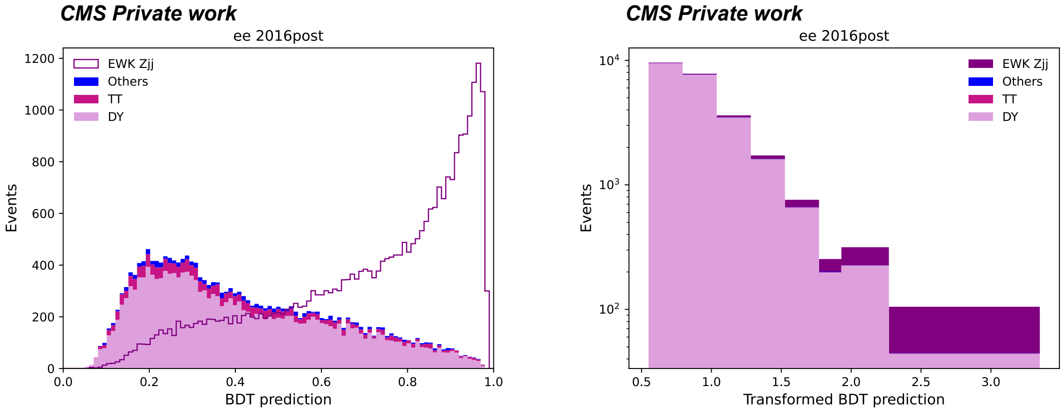
(b) tBDT evaluation output.



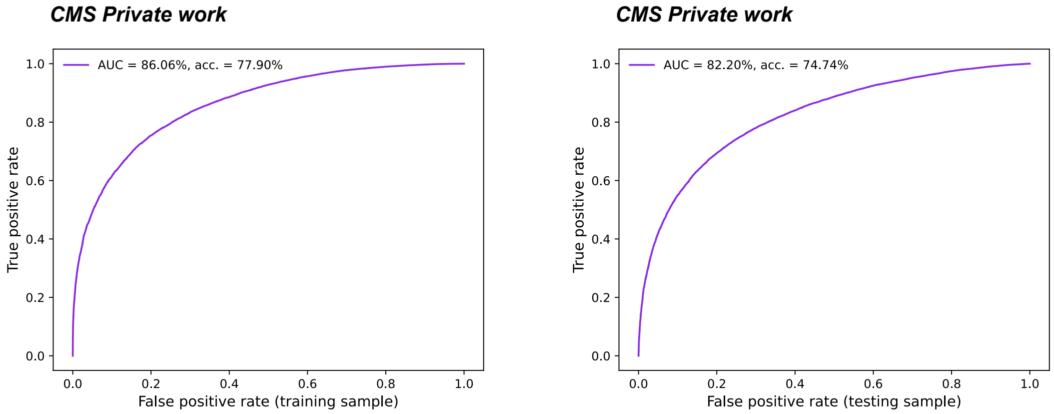
(c) AUC values for training.

(d) AUC values for testing

Figure F.5: Complementary BDT distributions for $\mu\mu$ channel, 2016postVFP.



(a) BDT evaluation output. (b) tBDT evaluation output.



(c) AUC values for training. (d) AUC values for testing

Figure F.6: Complementary BDT distributions for *ee* channel, 2016postVFP.

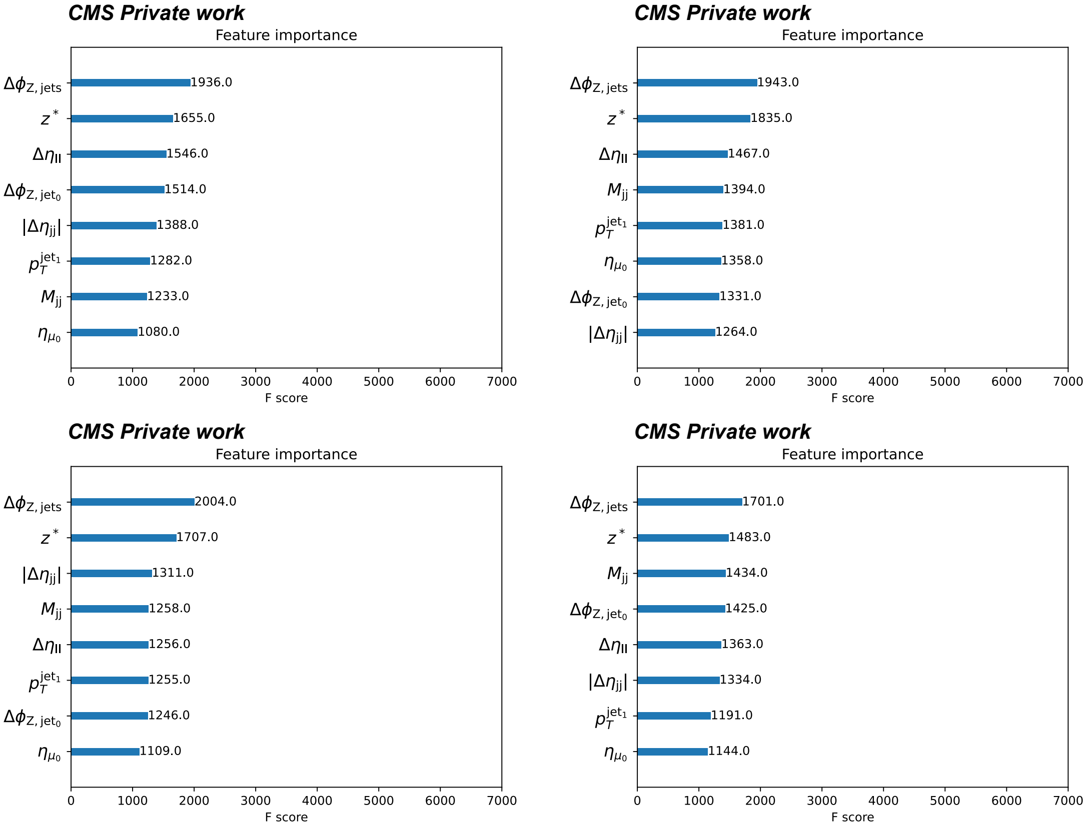


Figure F.7: F score for $\mu\mu$ channel for all eras in Run 2. Top row: 2018 (right), 2017 (left), bottom row: 2016preVFP (right), 2016postVFP (left).

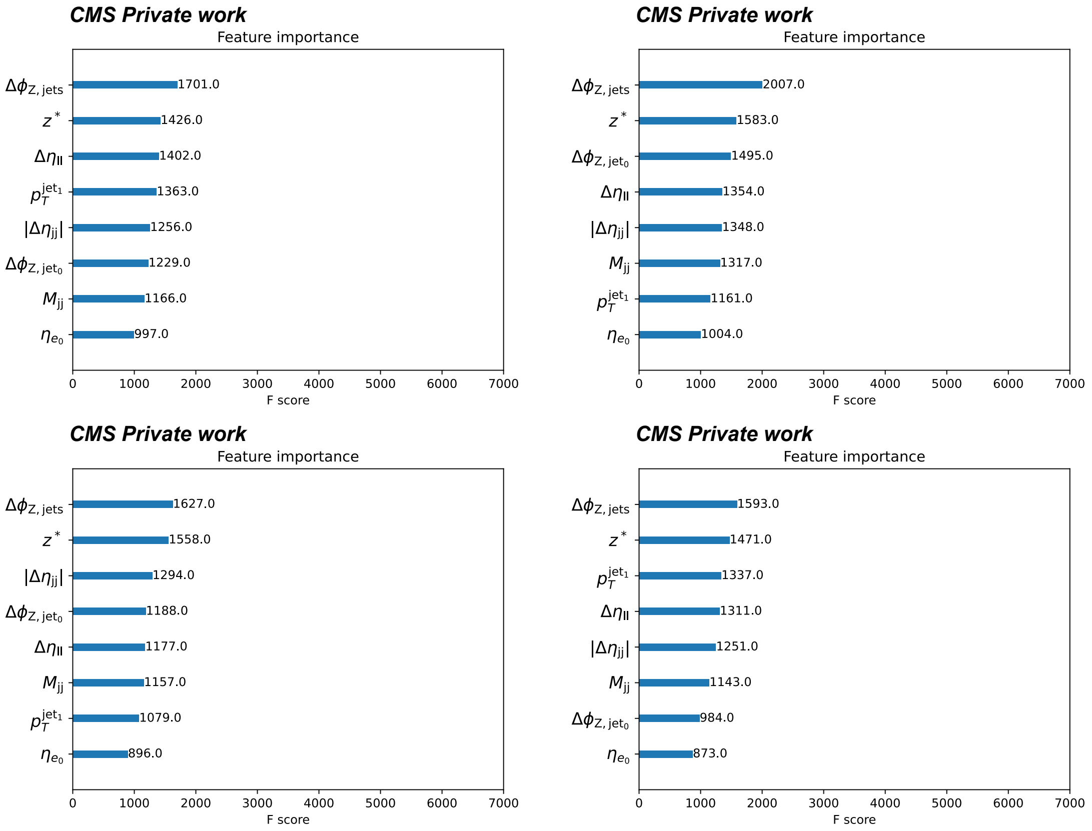


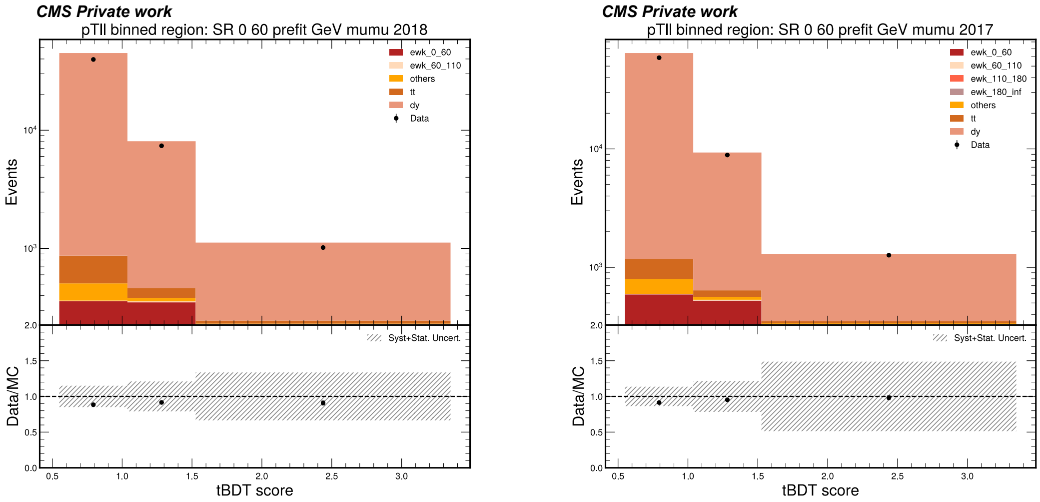
Figure F.8: F score for *ee* channel for all eras in Run 2. Top row: 2018 (right), 2017 (left), bottom row: 2016preVFP (right), 2016postVFP (left).

The BDT training is performed in the $\mu\mu$ and ee channel separately, where in each training sample, around 54% are signal events. The hyperparameters which lead to the best performance are:

- Learning rate: 0.05
- Maximum depth of a tree: 7
- Fraction of events to train tree on: 0.75
- Fraction of features to be randomly sampled for each tree: 0.65
- Minimum loss reduction required to make a further partition on a leaf node of the tree: 0.5
- L2 regularization: 1.5
- Minimum child weight: 2
- Objective function: binary-logistic
- Number of trees: 100

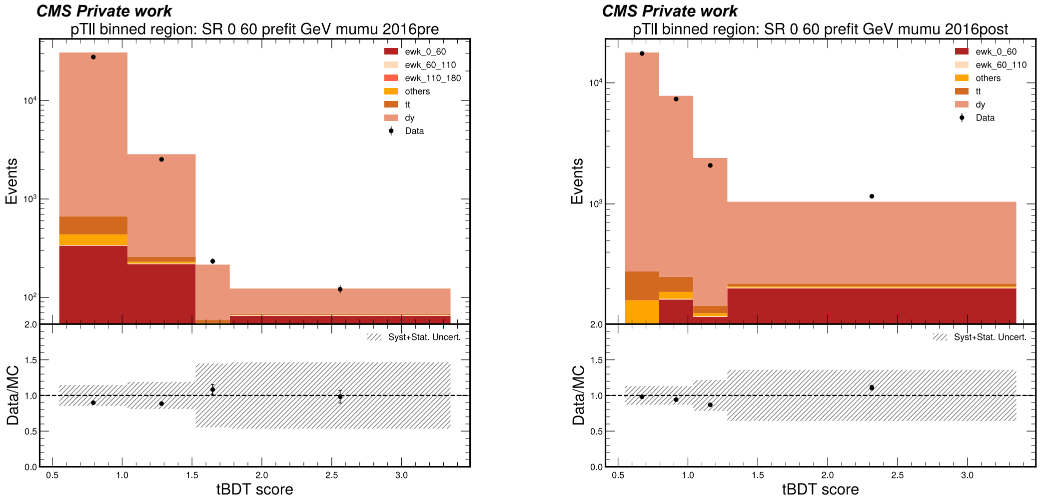
In the BDT prediction score distributions, the EW Zjj signal is shown as a line histogram, while background contributions are stacked by process categories.
The tBDT evaluation outputs are normalized to the expected event yields, computed from theoretical cross sections and integrated luminosity per year, and corrected for all weights and detector-level scale factors (as described in *Detector-based criteria* in Section 6.2.1).

F.2 Pre-fit p_T^{ll} binned tBDT distributions



(a) 2018

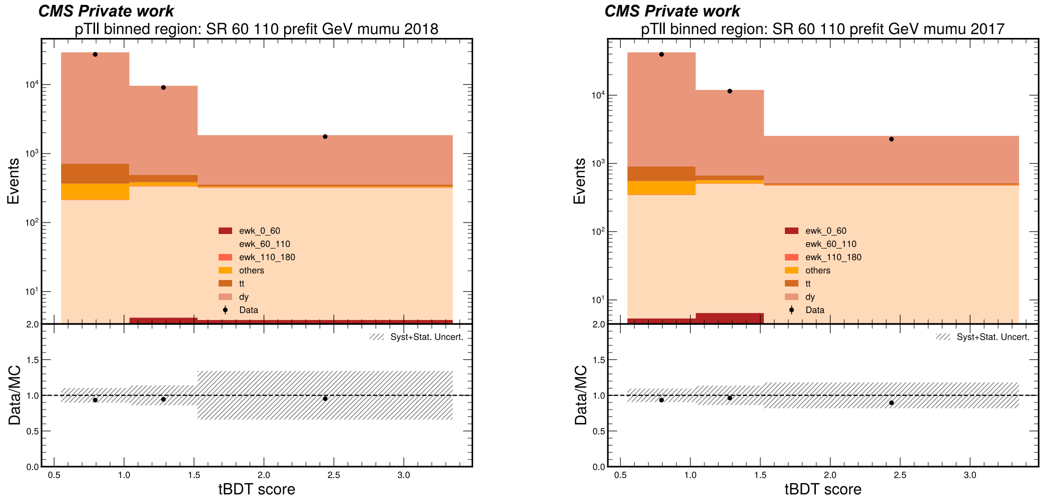
(b) 2017



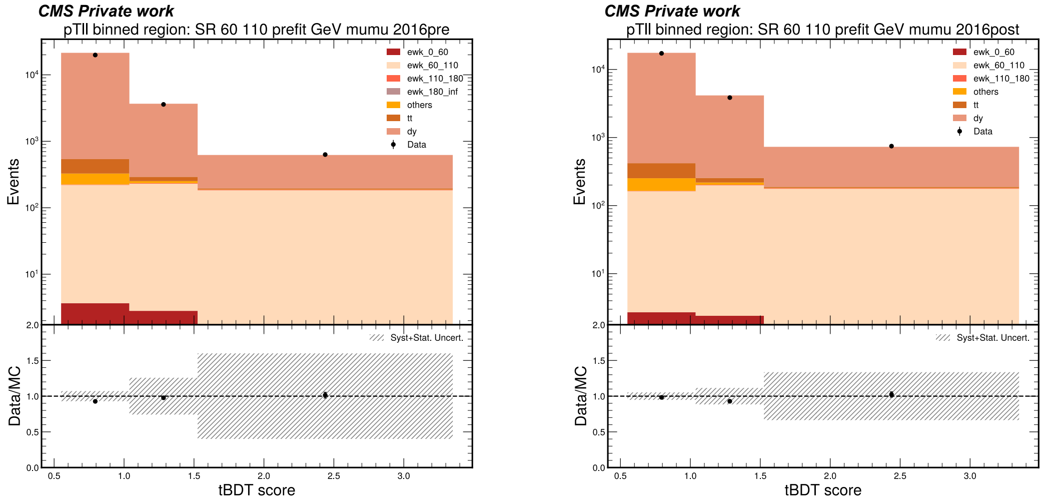
(c) 2016preVFP

(d) 2016postVFP

Figure F.9: Pre-fit p_T^{ll} binned tBDT for $0 < p_T^{ll} < 60$ GeV in $\mu\mu$ channel for all eras in Run 2.

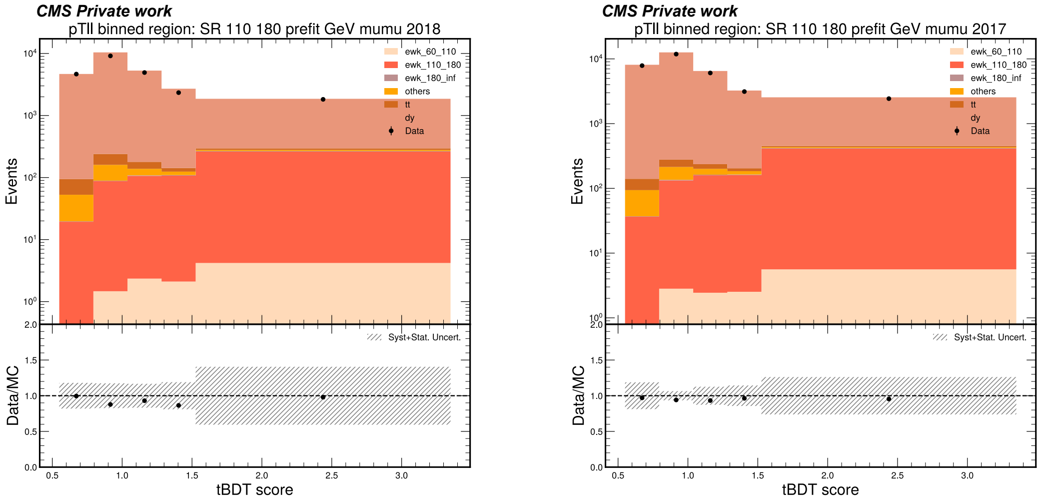


(a) 2018 (b) 2017



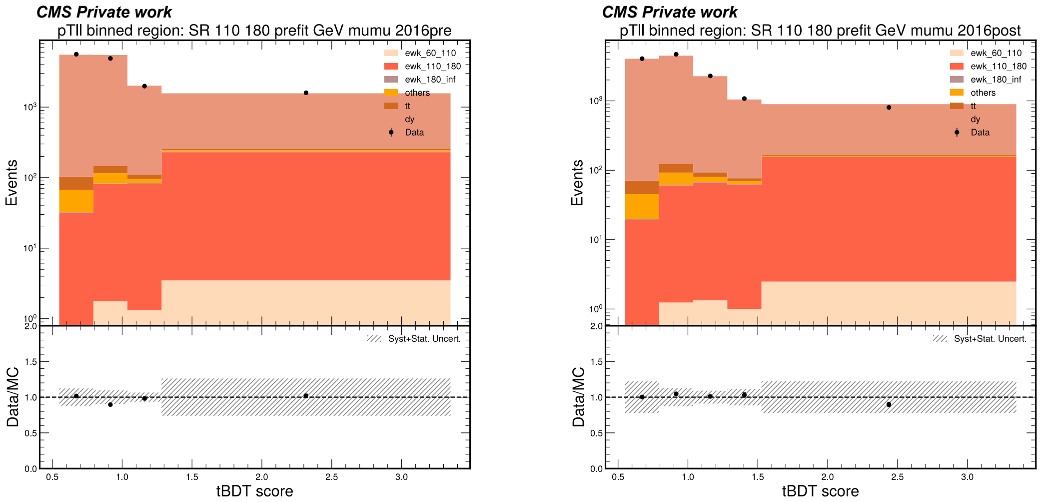
(c) 2016preVFP (d) 2016postVFP

Figure F.10: Pre-fit p_T^{ll} binned tBDT for $60 < p_T^{ll} < 110$ GeV in $\mu\mu$ channel for all eras in Run 2.



(a) 2018

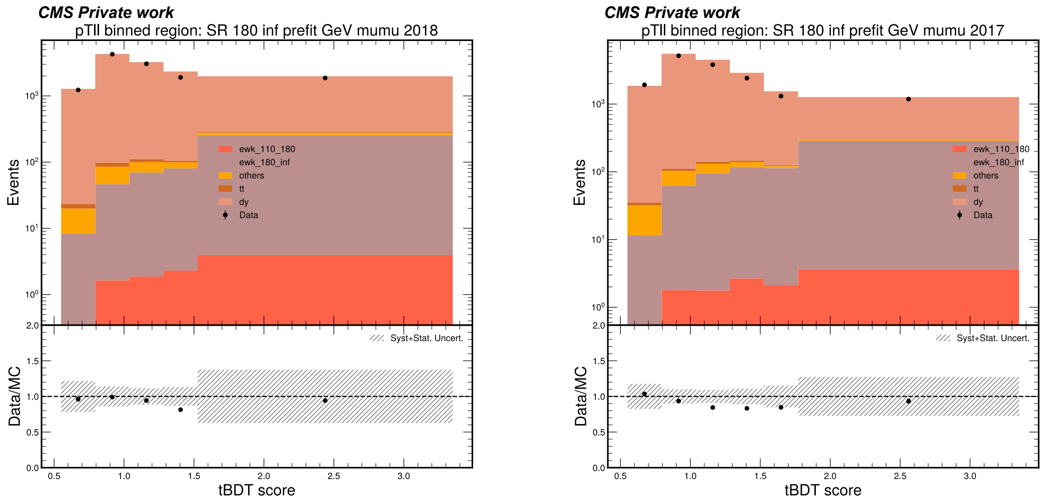
(b) 2017



(c) 2016preVFP

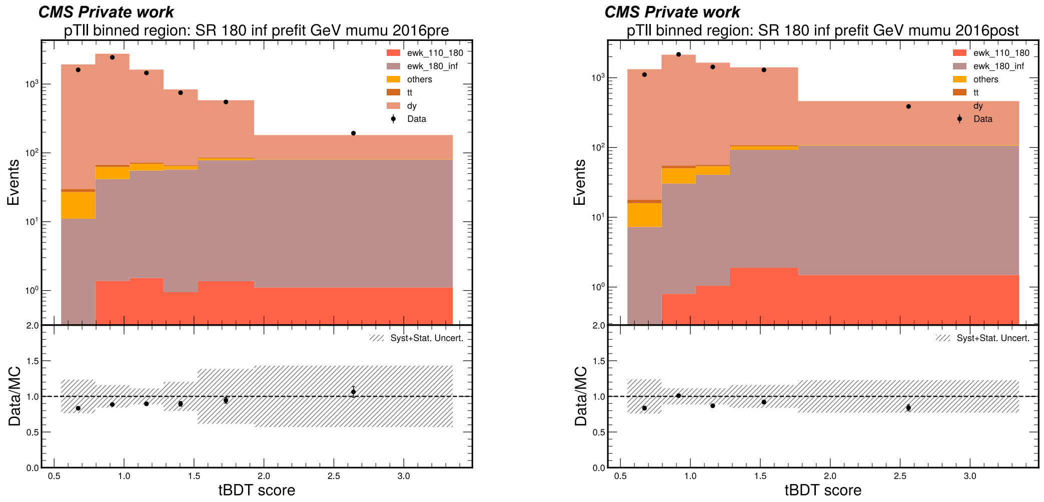
(d) 2016postVFP

Figure F.11: Pre-fit p_T^{ll} binned tBDT for $110 < p_T^{ll} < 180$ GeV in $\mu\mu$ channel for all eras in Run 2.



(a) 2018

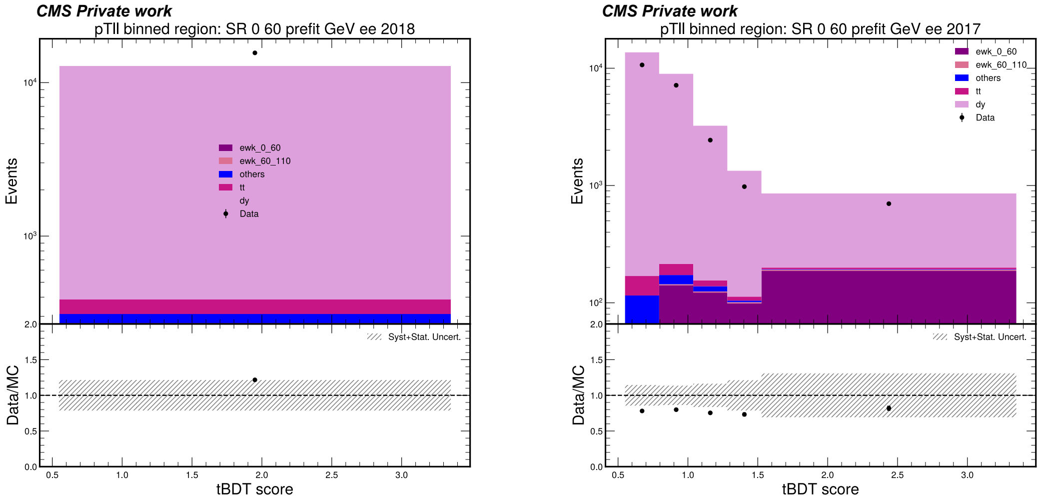
(b) 2017



(c) 2016preVFP

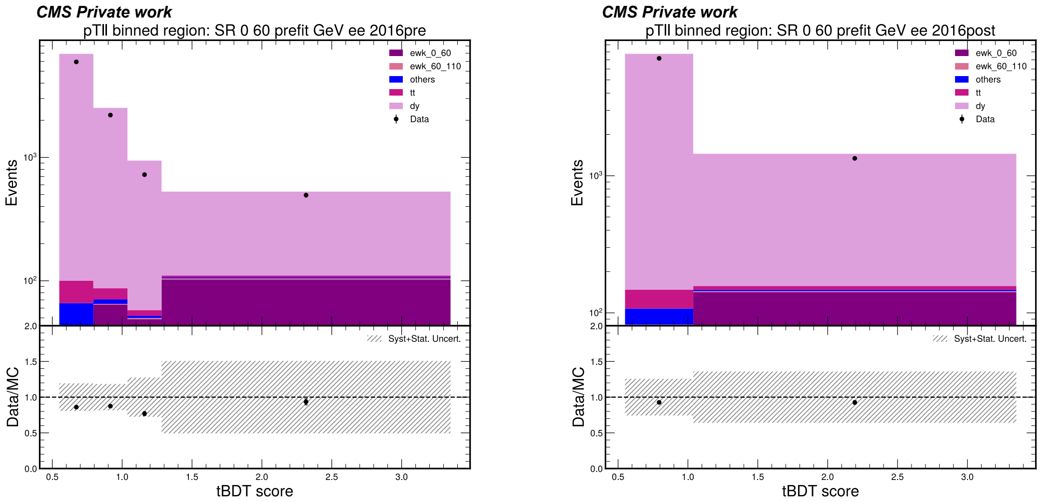
(d) 2016postVFP

Figure F.12: Pre-fit p_T^{ll} binned tBDT for $180 < p_T^{ll}$ GeV in $\mu\mu$ channel for all eras in Run 2.



(a) 2018

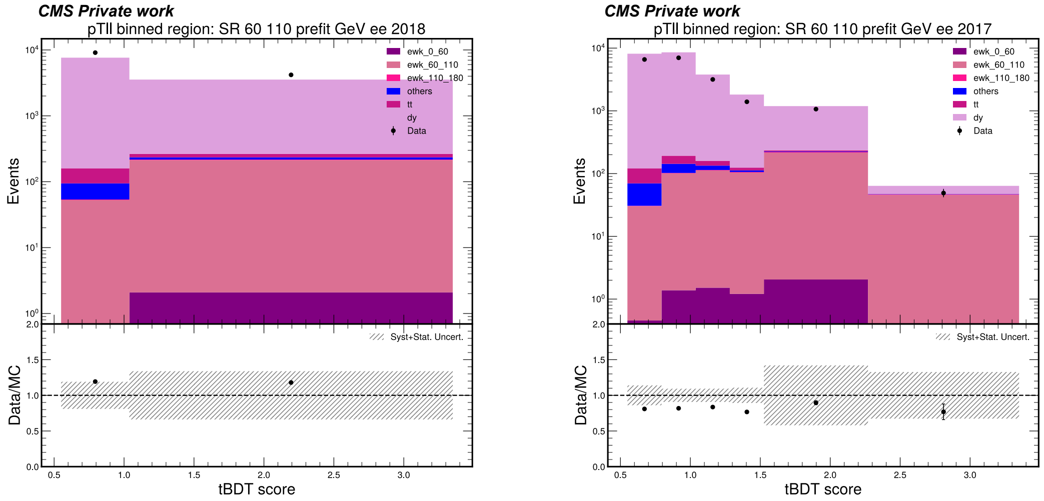
(b) 2017



(c) 2016preVFP

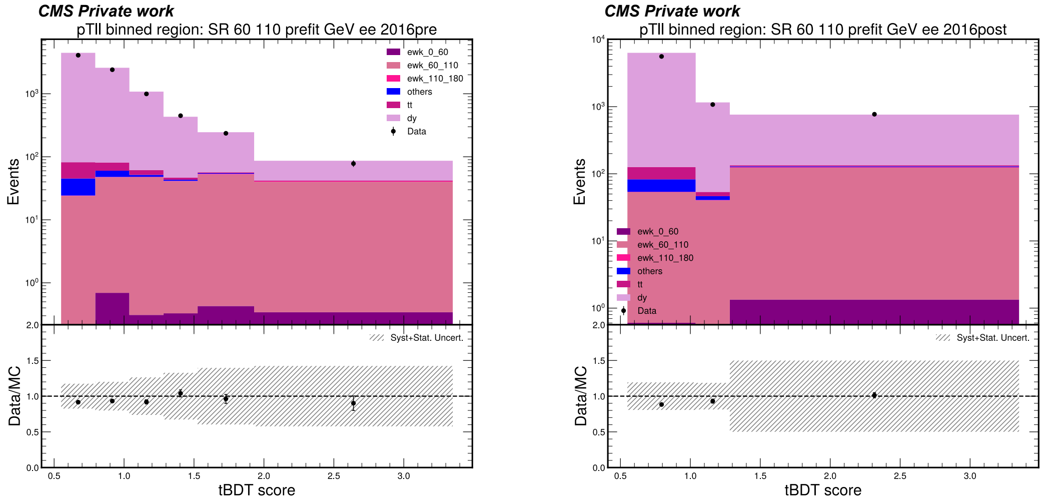
(d) 2016postVFP

Figure F.13: Pre-fit p_T^{ll} binned tBDT for $0 < p_T^{ll} < 60$ GeV in $\mu\mu$ channel for all eras in Run 2.



(a) 2018

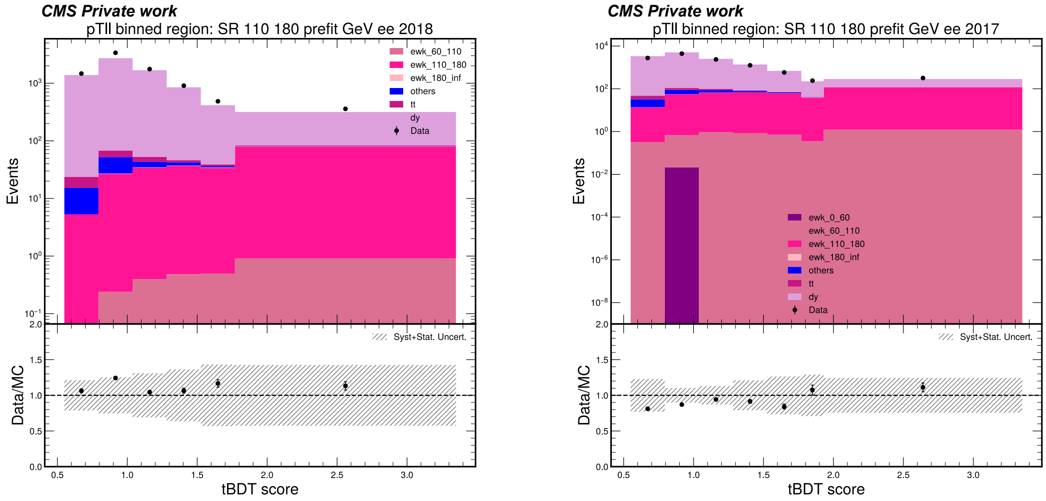
(b) 2017



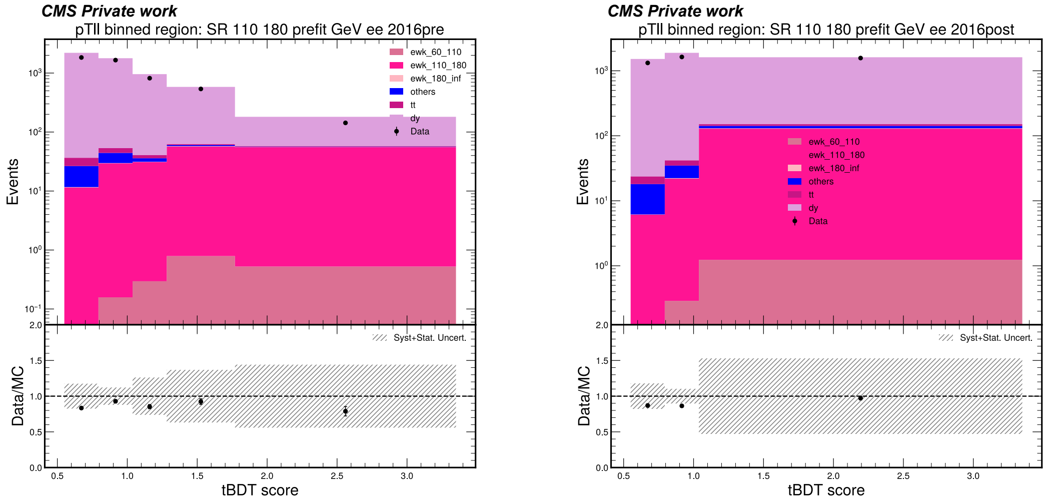
(c) 2016preVFP

(d) 2016postVFP

Figure F.14: Pre-fit p_T^{ll} binned tBDT for $60 < p_T^{ll} < 110$ GeV in $\mu\mu$ channel for all eras in Run 2.

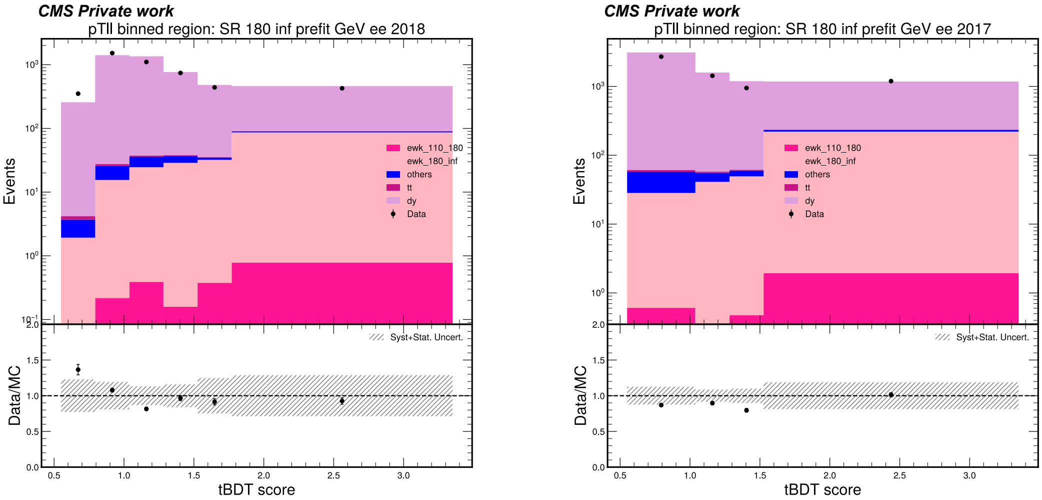


(a) 2018 (b) 2017



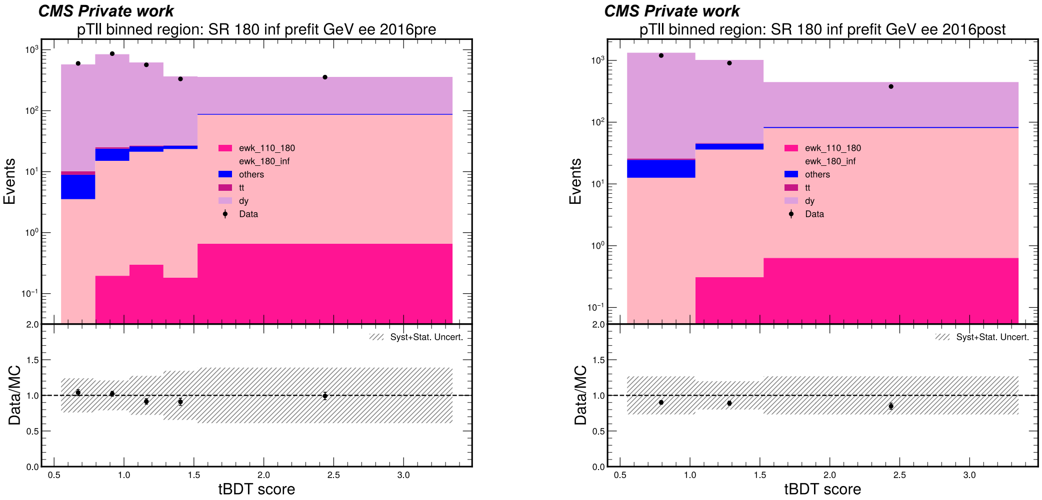
(c) 2016preVFP (d) 2016postVFP

Figure F.15: Pre-fit p_T^{ll} binned tBDT for $110 < p_T^{ll} < 180$ GeV in $\mu\mu$ channel for all eras in Run 2.



(a) 2018

(b) 2017



(c) 2016preVFP

(d) 2016postVFP

Figure F.16: Pre-fit p_T^{ll} binned tBDT for $180 < p_T^{ll}$ GeV in $\mu\mu$ channel for all eras in Run 2.

F.3 Post-fit p_T^{ll} binned tBDT distributions

Post-fit tBDT distributions for both $\mu\mu$ and ee channel for the Full Run 2 and σ_{EWZjj} AUC for the measurement of the EW Zjj process.

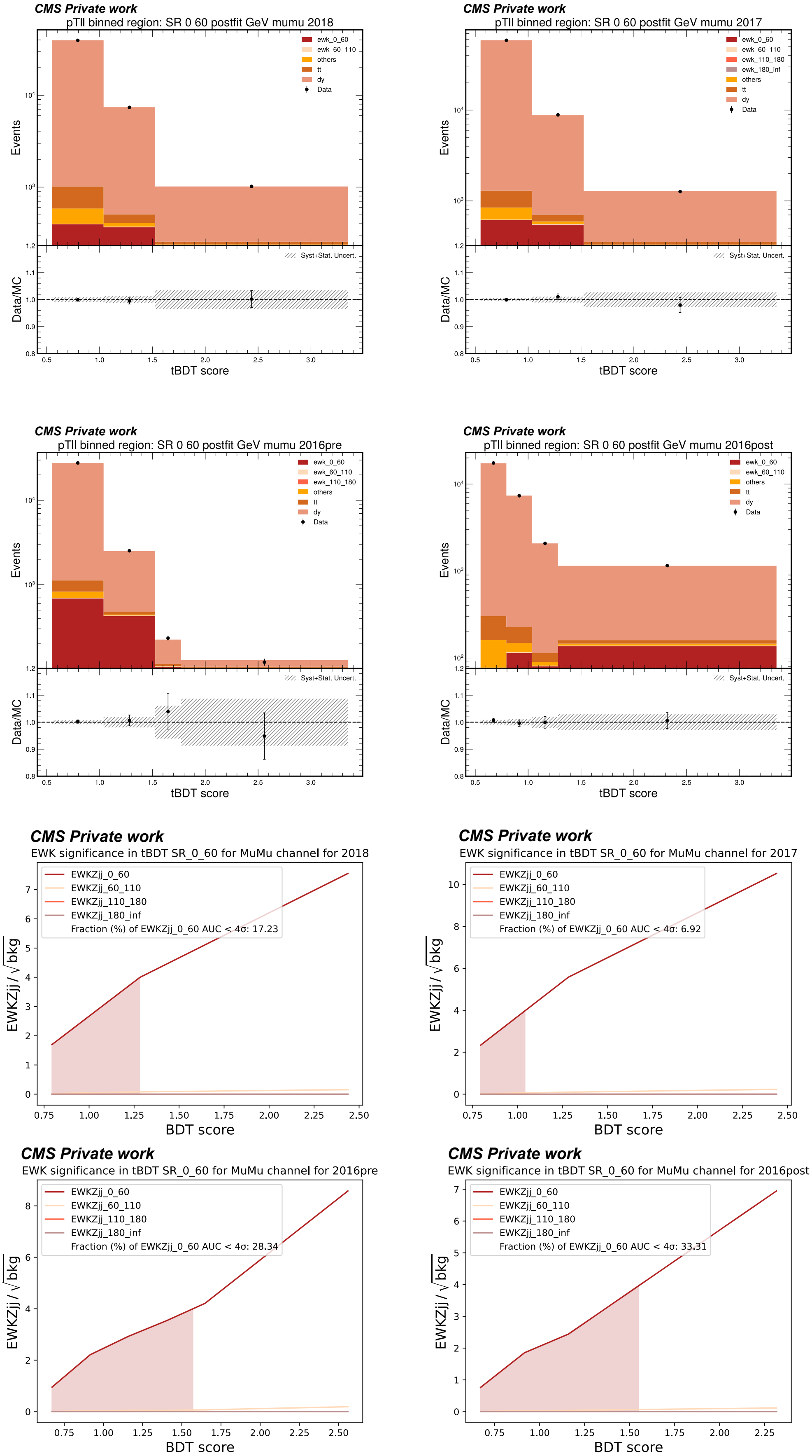


Figure F.17: Post-fit p_T^{ll} binned tBDT and σ_{EWZjj} AUC for $0 < p_T^{ll} < 60$ GeV in $\mu\mu$ channel for all eras in Run 2.

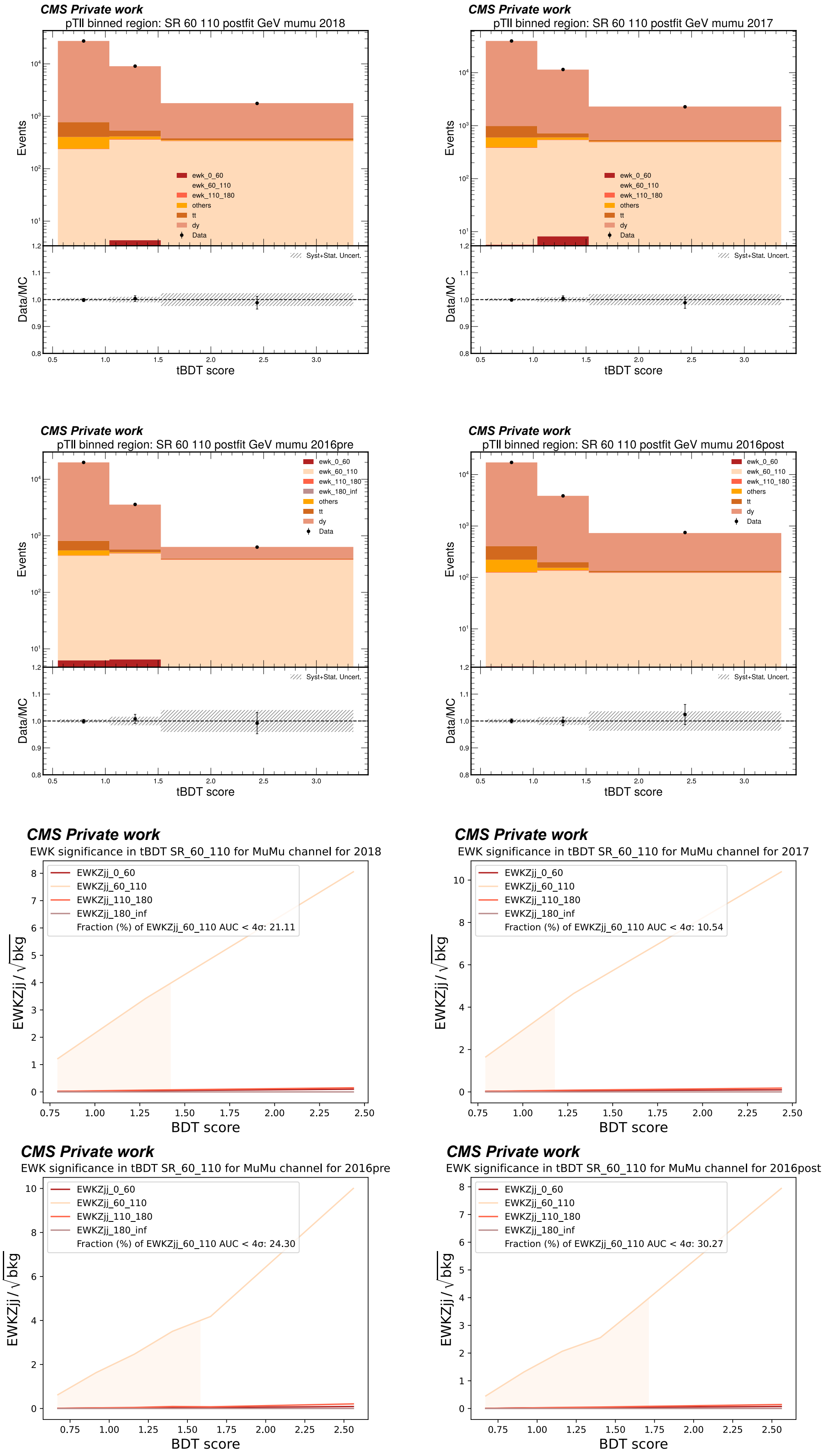


Figure F.18: Post-fit p_T^{ll} binned tBDT and σ_{EWZjj} AUC for $60 < p_T^{ll} < 110$ GeV in $\mu\mu$ channel for all eras in Run 2.

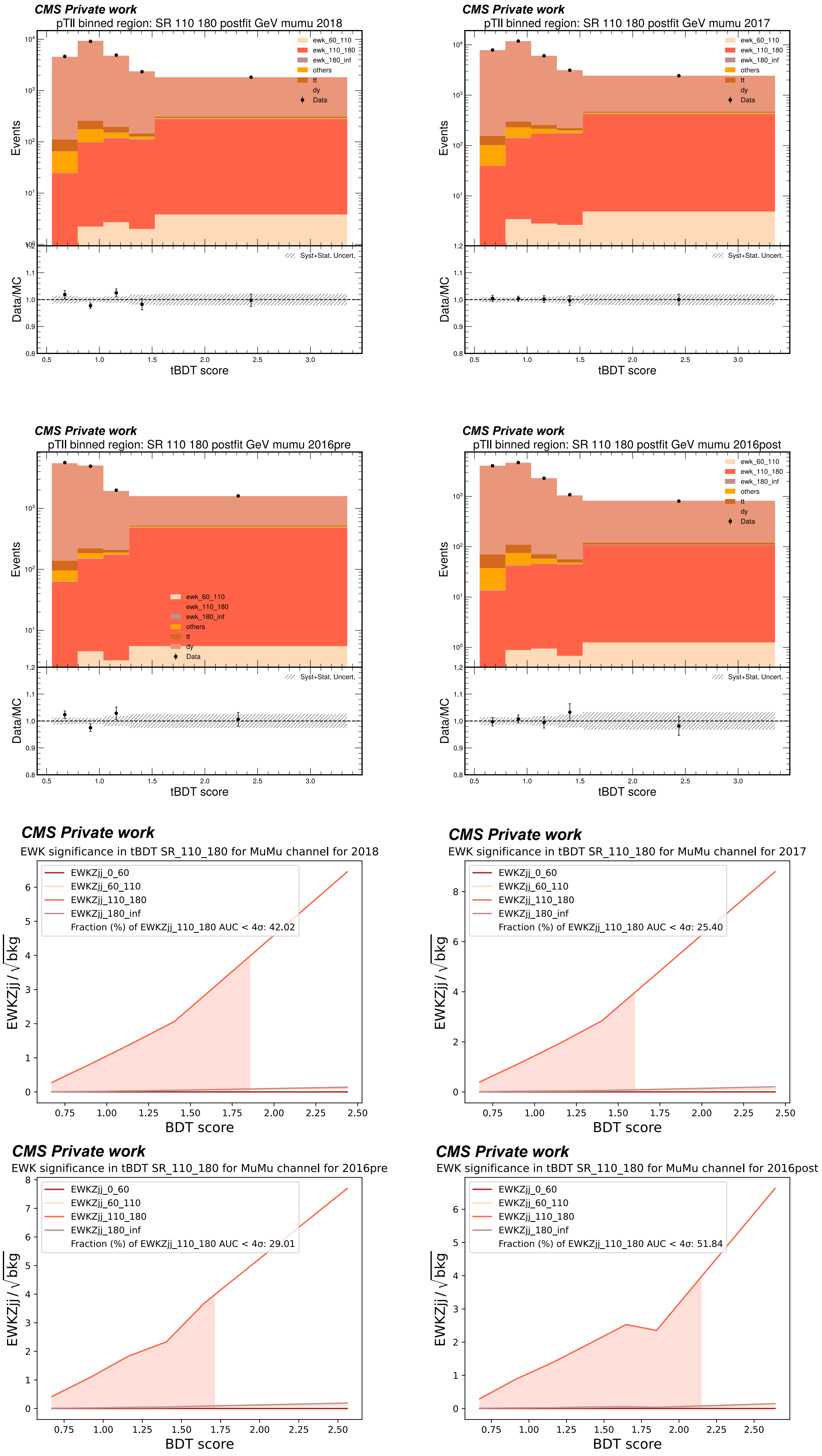


Figure F.19: Post-fit p_T^{ll} binned tBDT and σ_{EWZjj} AUC for $110 < p_T^{ll} < 180$ GeV in $\mu\mu$ channel for all eras in Run 2.

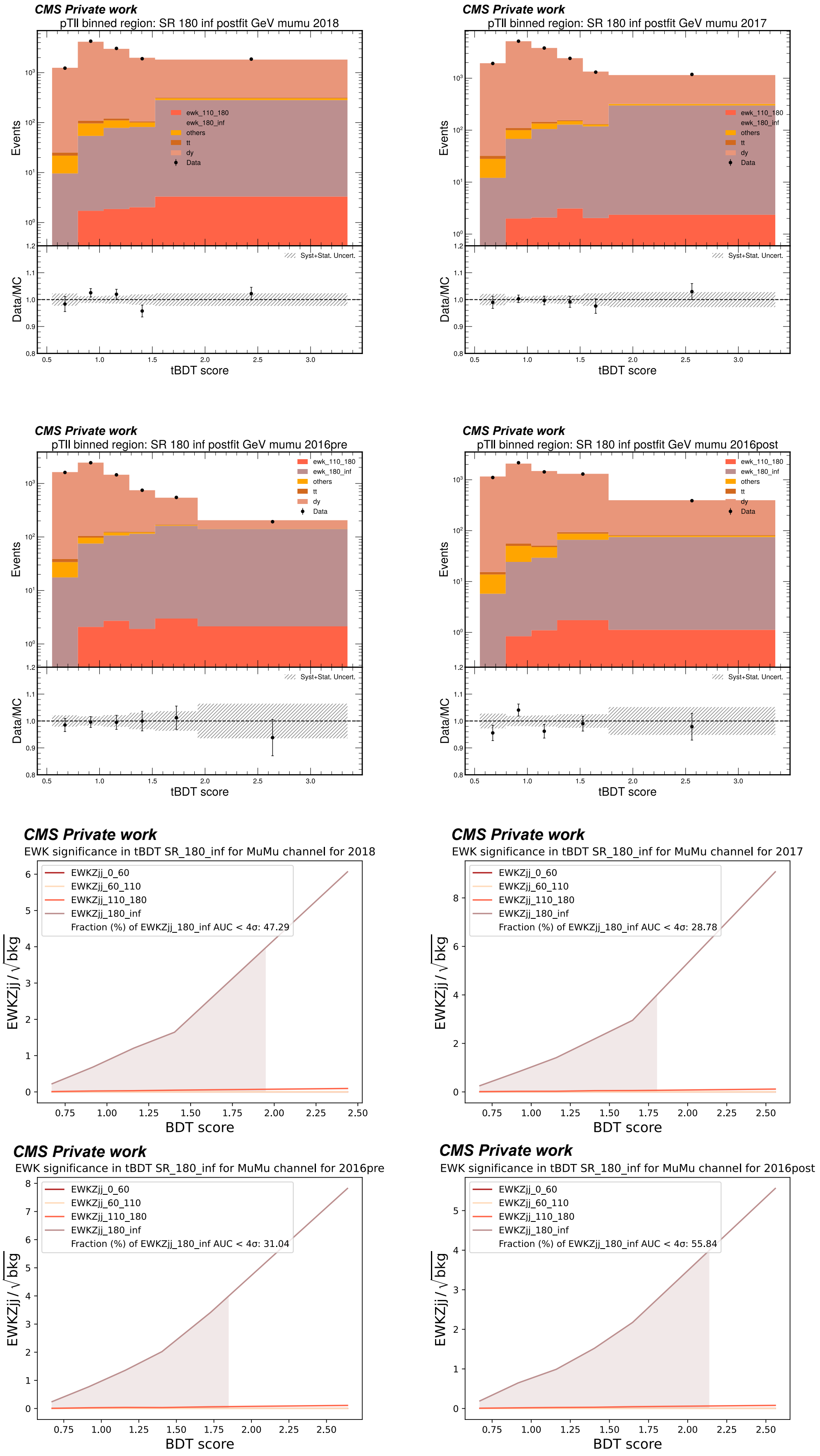


Figure F.20: Post-fit p_T^{ll} binned tBDT and σ_{EWZjj} AUC for $180 < p_T^{ll}$ GeV in $\mu\mu$ channel for all eras in Run 2.

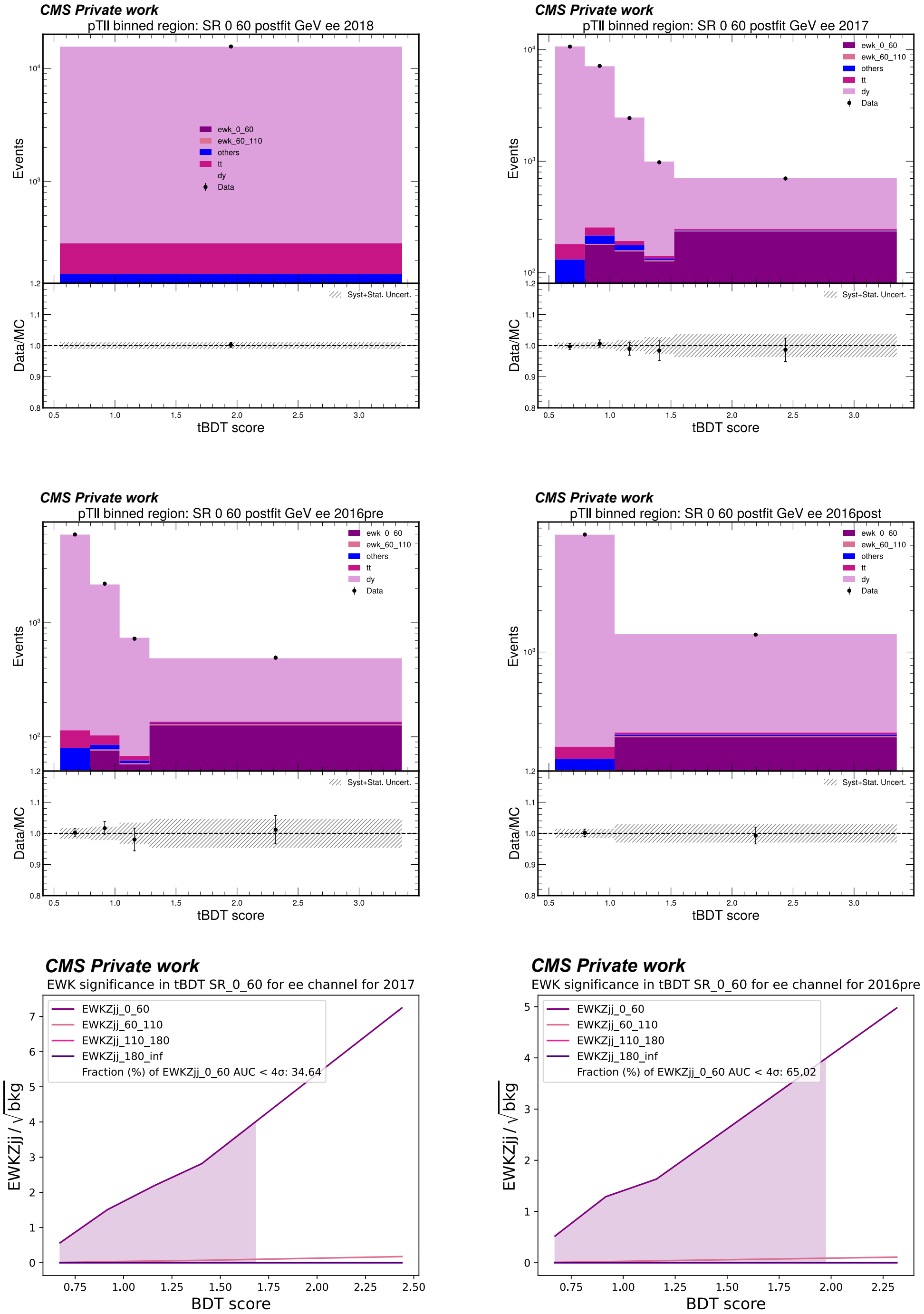


Figure F.21: Post-fit p_T^{ll} binned tBDT for all eras in Run 2 and σ_{EWZjj} AUC for $0 < p_T^{ll} < 60$ GeV in ee channel for 2017 and 2016preVFP.

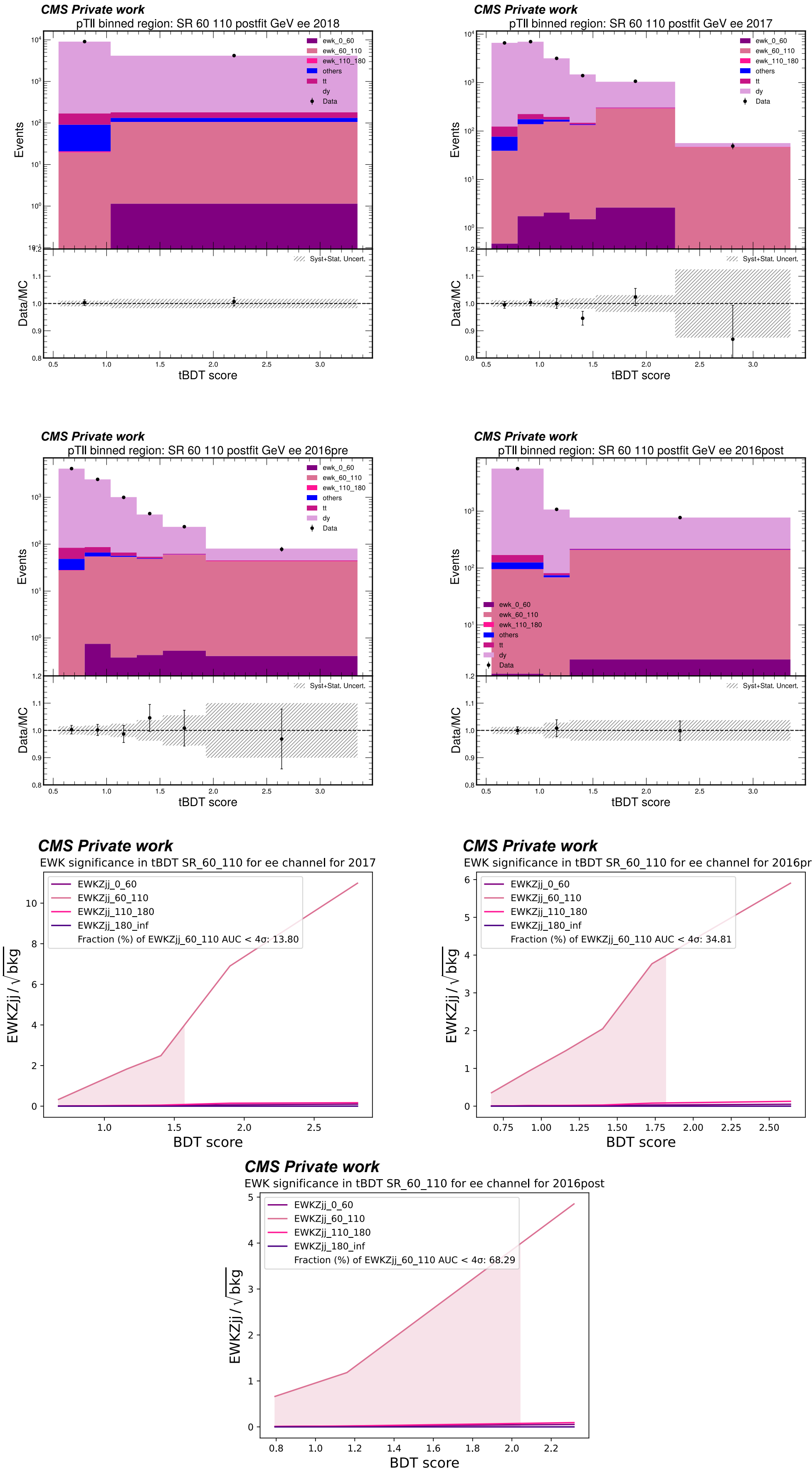
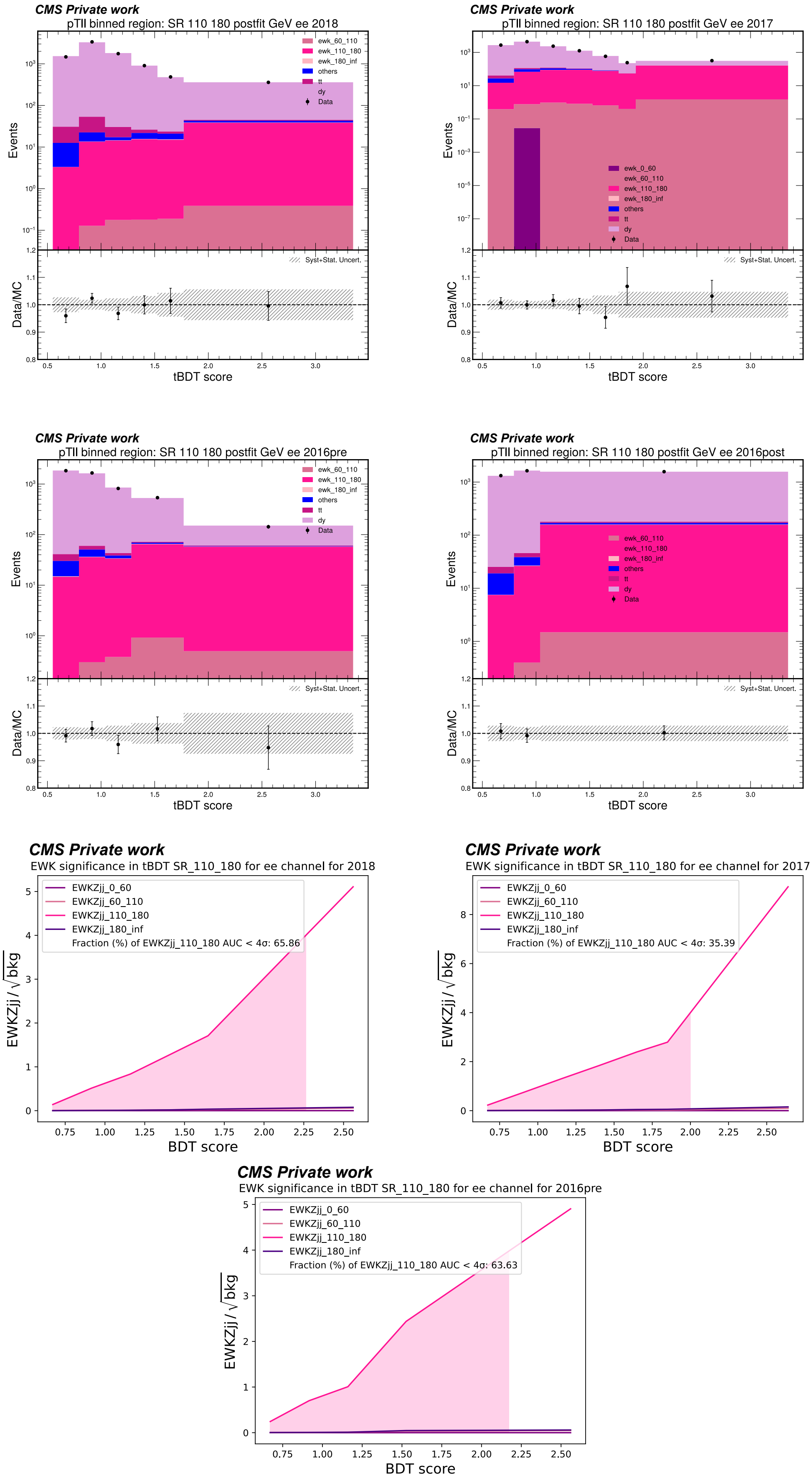


Figure F.22: Post-fit p_T^{ll} binned tBDT for all eras in Run 2 and σ_{EWZjj} AUC for $60 < p_T^{ll} < 110$ GeV in ee channel for 2017, 2016preVFP and 2016postVFP.


CMS Private work
pTll binned region: SR 110 180 postfit GeV ee 2018

CMS Private work
pTll binned region: SR 110 180 postfit GeV ee 2017

CMS Private work
pTll binned region: SR 110 180 postfit GeV ee 2016pre

CMS Private work
pTll binned region: SR 110 180 postfit GeV ee 2016post

CMS Private work
EWK significance in tBDT SR_110_180 for ee channel for 2018

CMS Private work
EWK significance in tBDT SR_110_180 for ee channel for 2017

CMS Private work
EWK significance in tBDT SR_110_180 for ee channel for 2016pre

Figure F.23: Post-fit p_T^{ll} binned tBDT for all eras in Run 2 and σ_{EWZjj} AUC for $110 < p_T^{ll} < 180$ GeV in ee channel for 2018, 2017 and 2016preVFP.

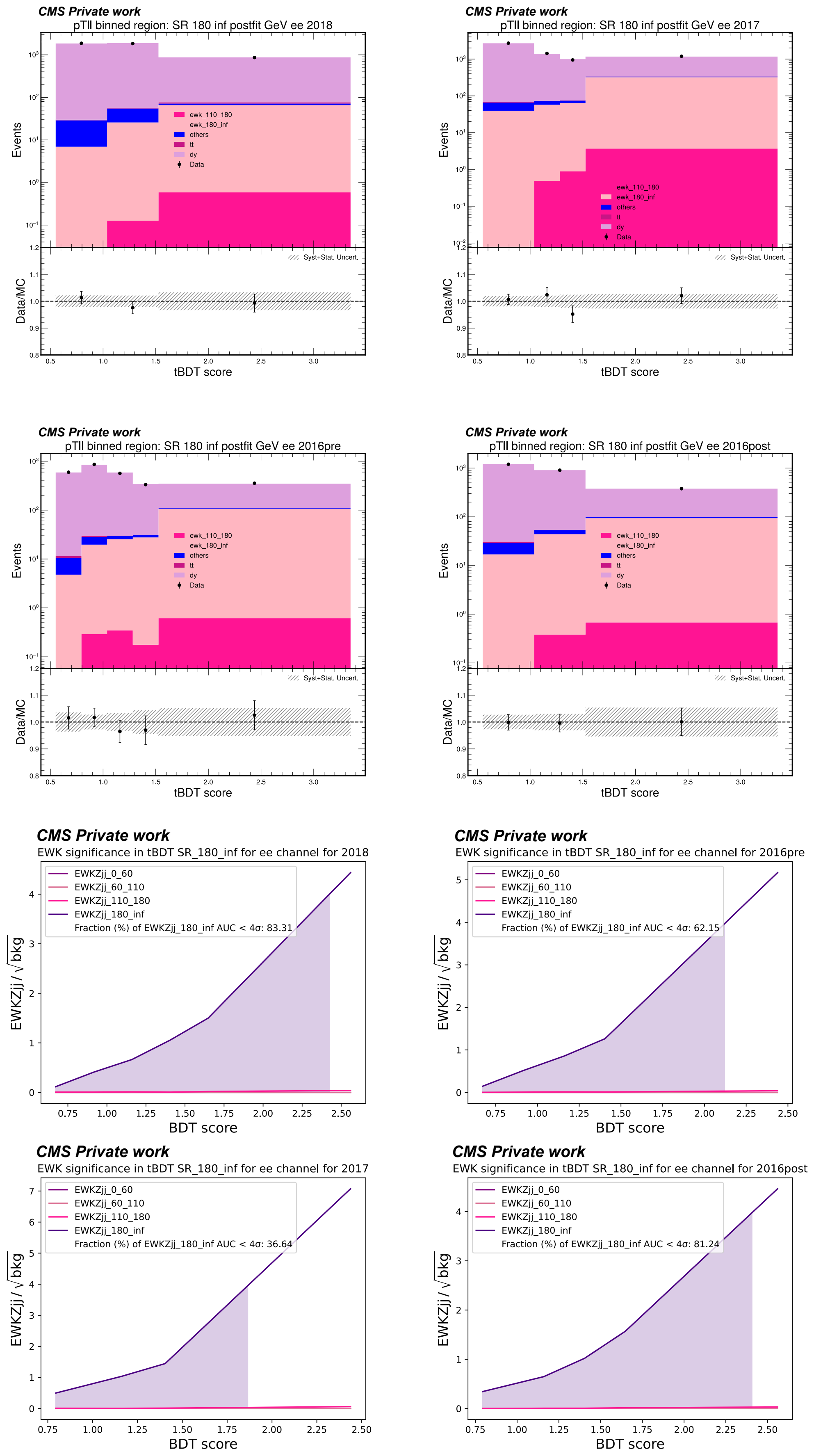
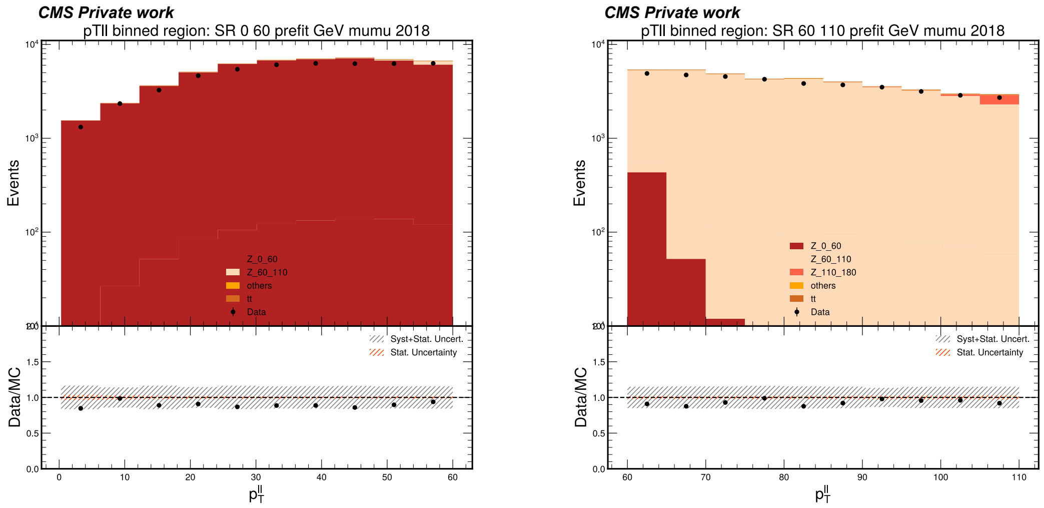


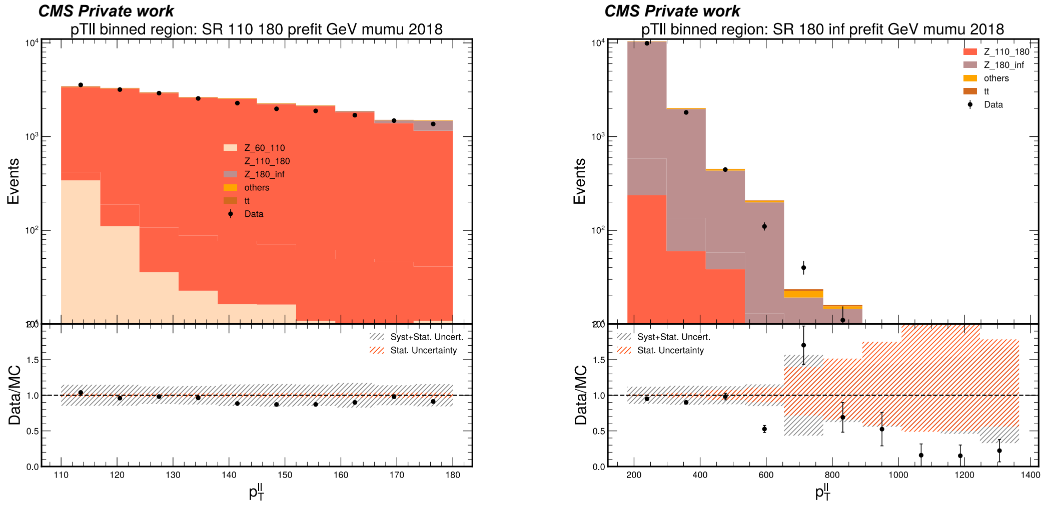
Figure F.24: Post-fit p_T^{ll} binned tBDT for all eras in Run 2 and σ_{EWZjj} AUC for $180 < p_T^{ll}$ GeV in ee channel for for all eras in Run 2.

Appendix G

Inclusive Z p_T^{ll} distributions

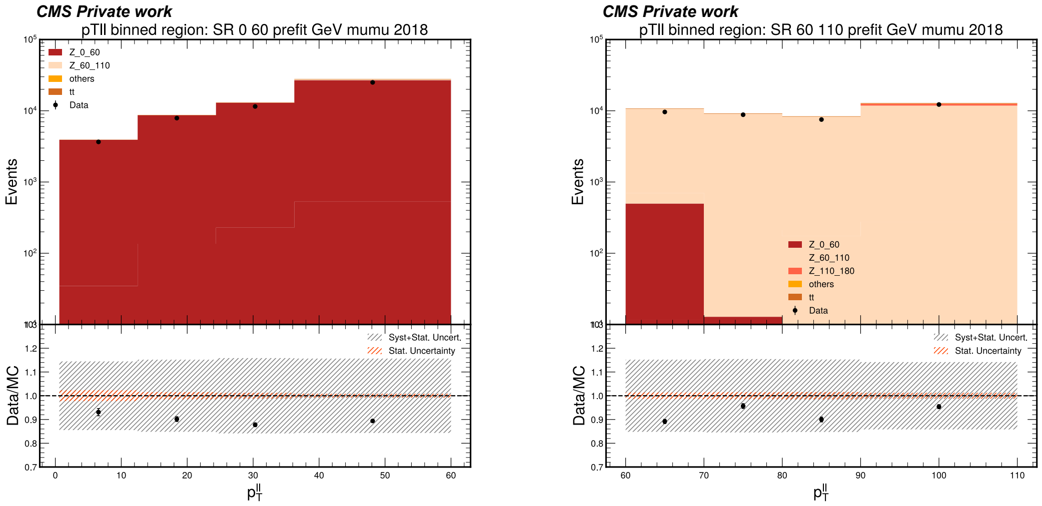


(a) $0 < p_T^{ll} < 60$ (b) $60 < p_T^{ll} < 110$

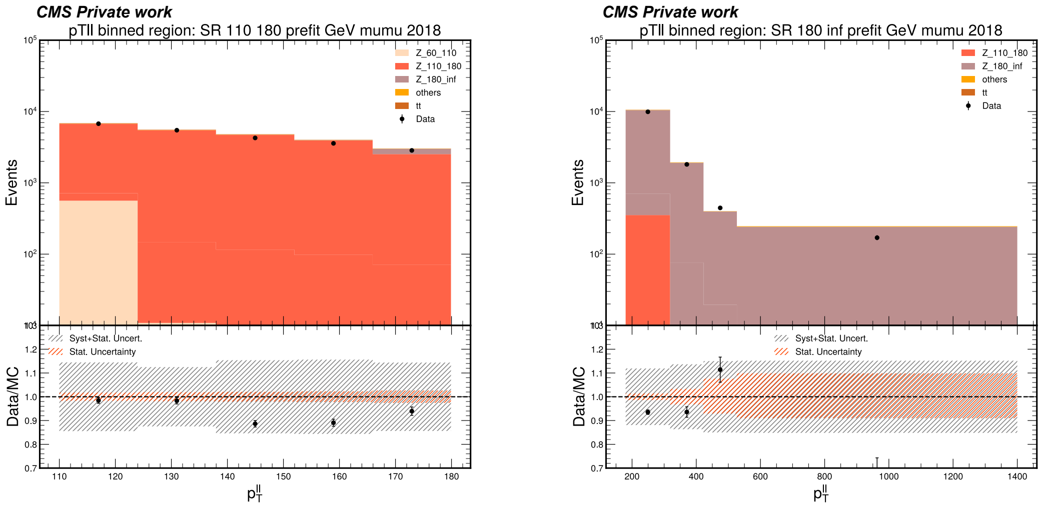


(c) $110 < p_T^{ll} < 180$ (d) $180 < p_T^{ll}$

Figure G.1: Pre-fit and pre-binning p_T^{ll} observable in different regions for the inclusive Z measurement in the $\mu\mu$ channel for 2018.

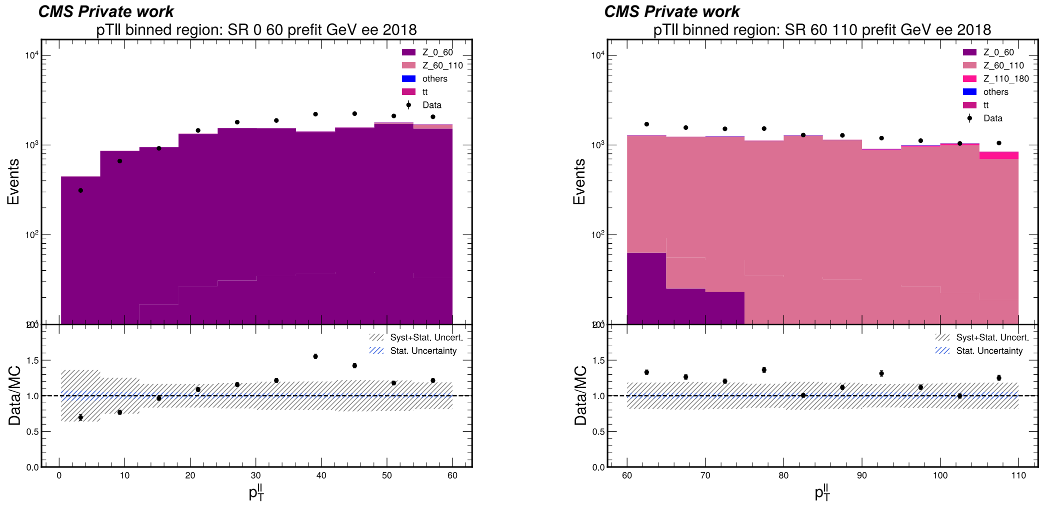


(a) $0 < p_T^{ll} < 60$ (b) $60 < p_T^{ll} < 110$

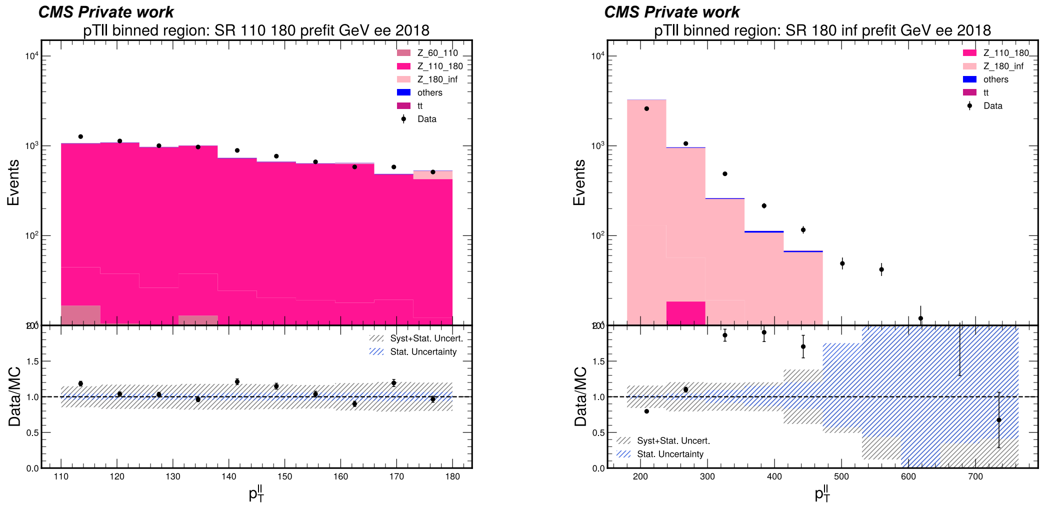


(c) $110 < p_T^{ll} < 180$ (d) $180 < p_T^{ll}$

Figure G.2: Pre-fit and post-binning p_T^{ll} observable in different regions for the inclusive Z measurement in the $\mu\mu$ channel for 2018.

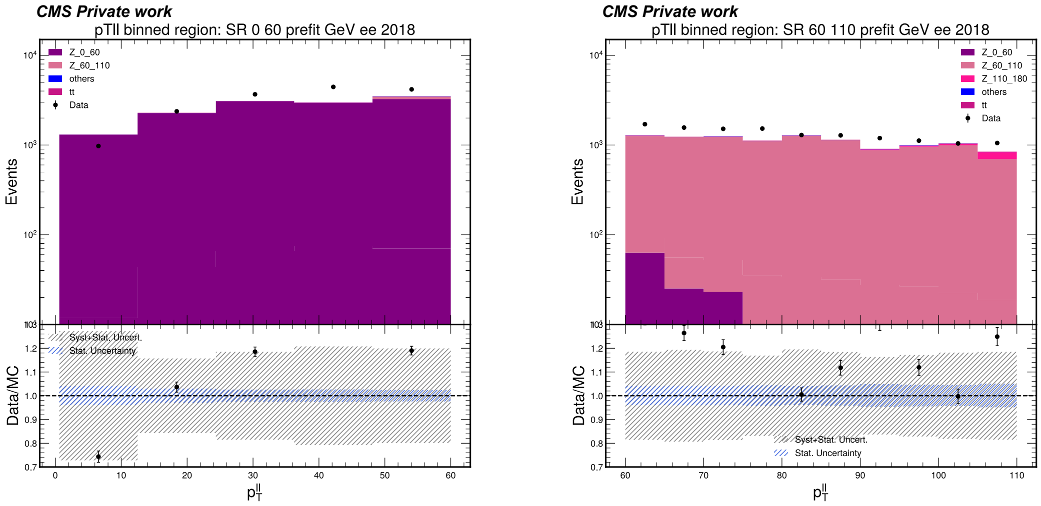


(a) $0 < p_T^{ll} < 60$ (b) $60 < p_T^{ll} < 110$

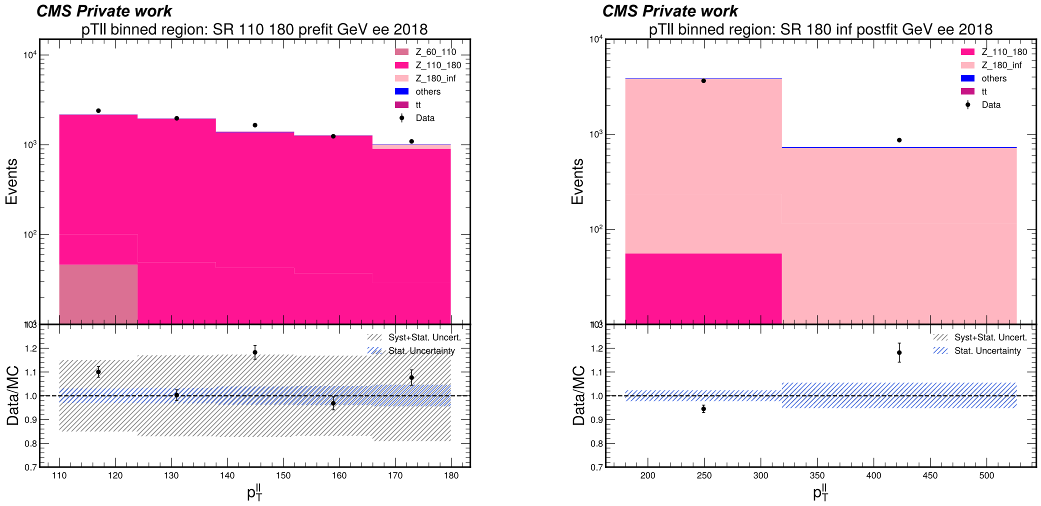


(c) $110 < p_T^{ll} < 180$ (d) $180 < p_T^{ll}$

Figure G.3: Pre-fit and pre-binning p_T^{ll} observable in different regions for the inclusive Z measurement in the ee channel for 2018.



(a) $0 < p_T^{ll} < 60$ (b) $60 < p_T^{ll} < 110$



(c) $110 < p_T^{ll} < 180$ (d) $180 < p_T^{ll}$

Figure G.4: Pre-fit and post-binning p_T^{ll} observable in different regions for the inclusive Z measurement in the ee channel for 2018.

Appendix H

Production card for a centrally produced VBF Z sample

```
! Production parameters for VBF Z production with Powheg
 idvecbos 23     ! PDG code for vector boson to be produced ( Z:23 )
vdecaymode -13  ! PDG code for charged decay product of vector boson
zmass 91.188d0  ! mass of the Z boson [GeV]
wmass 80.419d0  ! mass of the W boson [GeV]
zerowidth 1       ! 0 (offshell Zs); if 1 generate Zs on shell
numevts NEVENTS
ih1   1     ! hadron 1 (1 for protons, -1 for antiprotons)
ih2   1     ! hadron 2 (1 for protons, -1 for antiprotons)
ebeam1 6500d0     ! energy of beam 1
ebeam2 6500d0     ! energy of beam 2
! To be set only if using LHA pdfs
lhans1 325300
lhans2 325300
! To be set only if using different pdf sets for the two incoming hadrons
! Parameters to allow or not the use of stored data
use-old-grid    1 ! if 1 use old grid if file pwggrids.dat is present
use-old-ubound  1 ! if 1 use norm of upper bounding function in pwgubound.dat
ncall1  50000 ! number of calls for initializing the integration grid
itmx1    1     ! number of iterations for initializing the integration grid
ncall2  50000  ! number of calls for the integral and finding upper bound
itmx2    3      ! number of iterations for the integral and finding upper bound
foldcsi   5     ! number of folds on csi integration
foldy     5     ! number of folds on  y  integration
foldphi   10    ! number of folds on phi integration
nubound 100000       ! number of bbarra calls to setup norm of upper bounding func
icsimax  1      ! <= 100, number of csi subdivision when computing the upper bound
iymax    1      ! <= 100, number of y subdivision when computing the upper bounds
xupbound 2d0    ! increase upper bound for radiation generation
! OPTIONAL PARAMETERS
bornonly   0      ! (default 0) if 1 do Born only
smartsig   1      ! (default 1) remember equal amplitudes (0 do not remember)
withdamp    1      ! (default 0, do not use) use Born-zero damping factor
```

```
bornsuppfact 1     ! (default 0, do not use) use Born suppression factor
mll_gencut  30d0   ! mll generation cut
testplots  1      ! (default 0, do not) do NLO and PWHG distributions
iseed SEED
rand1     -1      ! initialize random number sequence
rand2     -1      ! initialize random number sequence
fakevirt 1    ! uses Born for determining virt. grid
withnegweights 1   ! default 1 allow negative weights
runningscales 0 ! default 0 (no running scales); 1 or 2 use running scales
btlscalereal  1 ! flag for running scales
btlscalect 1      ! flag for running scales
pdfreweight 0
storeinfo_rwgt 0
rwl_group_events 2000
 lhapdf6maxsets 50
rwl_file 'pwg-rwl.dat'
rwl_format_rwgt 1
```

Appendix I

Rivet plugin

```
// -*- C++ -*-
#include "Rivet/Analysis.hh"
#include "Rivet/Projections/FinalState.hh"
#include "Rivet/Projections/FastJets.hh"
#include "Rivet/Projections/VetoedFinalState.hh"
#include "Rivet/Projections/ZFinder.hh"
#include <cmath>

namespace Rivet {

  /// @brief Electroweak Z+jj analysis
  class Rivet_ewzjj_endorsed : public Analysis {
  public:

    /// Constructor
    RIVET_DEFAULT_ANALYSIS_CTOR(Rivet_ewzjj_endorsed);

    /// @name Analysis methods
    /// @{

    /// Book histograms and initialise projections before the run
    void init() {

      // Basic final state projection
      FinalState fs(Cuts::pT > 20*GeV && Cuts::abseta < 4.7);
      declare(fs, "FS");

      // Z boson finder for muons
      Cut muon_cuts = Cuts::abseta < 2.4 && Cuts::pT > 20*GeV;
      const ZFinder zfinder_mu(fs, muon_cuts, PID::MUON, 75*GeV, 106*GeV,
                               0.1, ZFinder::ClusterPhotons::ALL);
      declare(zfinder_mu, "zfinder_mu");

      // Define vetoed final state to prevent Z-decay products entering
```

```
// jet algorithm
VetoedFinalState had_fs;
had_fs.addVetoOnThisFinalState(getProjection<ZFinder>("zfinder_mu"));

// Jet reconstruction
FastJets jetfs(had_fs, FastJets::ANTIKT, 0.4, JetAlg::Muons::NONE,

JetAlg::Invisibles::NONE);
declare(jetfs, "jets");

// Book histograms - Jet observables
book(_h["histPtJet1"], "histPtJet1", 12, 30, 510);
book(_h["histPtJet2"], "histPtJet2", 12, 30, 510);
book(_h["histAbsetaJet1"], "histAbsetaJet1", 11, 0, 4.4);
book(_h["histAbsetaJet2"], "histAbsetaJet2", 11, 0, 4.4);
book(_h["histEtaJet1"], "histEtaJet1", 22, -4.4, 4.4);
book(_h["histEtaJet2"], "histEtaJet2", 22, -4.4, 4.4);
book(_h["histAbsPhiJet1"], "histAbsPhiJet1", 15, 0, 3);
book(_h["histAbsPhiJet2"], "histAbsPhiJet2", 15, 0, 3);
book(_h["histAbsRapJet1"], "histAbsRapJet1", 23, 0, 4.6);
book(_h["histAbsRapJet2"], "histAbsRapJet2", 23, 0, 4.6);
book(_h["histRapJet1"], "histRapJet1", 46, -4.6, 4.6);
book(_h["histRapJet2"], "histRapJet2", 46, -4.6, 4.6);

// Dijet observables
book(_h["histPtDijet"], "histPtDijet", 24, 30, 510);
book(_h["histAbsetaDijet"], "histAbsetaDijet", 11, 0, 4.4);
book(_h["histAbsRapDijet"], "histAbsRapDijet", 11, 0, 4.4);
book(_h["histAbsDRapidityDijet"], "histAbsDRapidityDijet", 25, 0, 0.5);
book(_h["histAbsDetajj"], "histAbsDetajj", 7, 1.5, 5);
book(_h["histMassDijet"], "histMassDijet", 18, 120, 3000);
book(_h["histAbsPhiDijet"], "histAbsPhiDijet", 31, 0, 6.2);
book(_h["histDphiDijet"], "histDphiDijet", 14, 0, 3.2);

// Muon observables
book(_h["histPtMuon1"], "histPtMuon1", 9, 30, 390);
book(_h["histPtMuon2"], "histPtMuon2", 7, 20, 300);
book(_h["histAbsetaMuon1"], "histAbsetaMuon1", 12, 0, 2.4);
book(_h["histAbsetaMuon2"], "histAbsetaMuon2", 12, 0, 2.4);
book(_h["histEtaMuon1"], "histEtaMuon1", 24, -2.4, 2.4);
book(_h["histEtaMuon2"], "histEtaMuon2", 24, -2.4, 2.4);
book(_h["histAbsRapMuon1"], "histAbsRapMuon1", 12, 0, 2.4);
book(_h["histAbsRapMuon2"], "histAbsRapMuon2", 12, 0, 2.4);
book(_h["histRapMuon1"], "histRapMuon1", 24, -2.4, 2.4);
book(_h["histRapMuon2"], "histRapMuon2", 24, -2.4, 2.4);
book(_h["histAbsPhiMuon1"], "histAbsPhiMuon1", 15, 0, 3);
book(_h["histAbsPhiMuon2"], "histAbsPhiMuon2", 15, 0, 3);
```

```
// Dimuon/Z observables
book(_h["histPtDimuon"], "histPtDimuon", 40, 0, 400);
book(_h["histAbsetaDimuon"], "histAbsetaDimuon", 12, 0, 2.4);
book(_h["histAbsRapDimuon"], "histAbsRapDimuon", 12, 0, 2.4);
book(_h["histAbsPhiDimuon"], "histAbsPhiDimuon", 15, 0, 3);
book(_h["histMassDimuon"], "histMassDimuon", 15, 76, 106);
book(_h["histAbsDetaDimuon"], "histAbsDetaDimuon", 6, 0, 3);
book(_h["histAbsDRapidityDimuon"], "histAbsDRapidityDimuon", 20, 0, 0.04);
book(_h["histDphiDimuon"], "histDphiDimuon", 14, 0, 3.2);

// Angular correlations
book(_h["histDPhiMuJet1"], "histDPhiMuJet1", 14, 0.4, 3.2);
book(_h["histDPhiMuJet2"], "histDPhiMuJet2", 14, 0.4, 3.2);
book(_h["histDPhiMuJets"], "histDPhiMuJets", 14, 0.4, 3.2);
book(_h["histDPhiJetsMu"], "histDPhiJetsMu", 14, 0.4, 3.2);
book(_h["histDRJetsMuon1"], "histDRJetsMuon1", 60, 0, 6.0);
book(_h["histDRJetsMuon2"], "histDRJetsMuon2", 60, 0, 6.0);
book(_h["histDRMuJet3"], "histDRMuJet3", 60, 0, 6.0);

// VBF-specific observables
book(_h["histRpThard"], "histRpThard", 25, 0, 1);
book(_h["histZstar"], "histZstar", 15, 0, 3);
book(_h["histEtaThreeStar"], "histEtaThreeStar", 40, -4, 4);
book(_h["histAlphajj"], "histAlphajj", 20, 0, 1);
book(_h["histEtaThreeStarAlternative"], "histEtaThreeStarAlternative",

40, -4, 4);
book(_h["histSumpzjets"], "histSumpzjets", 25, 35, 550);
book(_h["histSumjets_pT_pTjets"], "histSumjets_pT_pTjets", 50, 0, 1);
book(_h["histAlphajj_Z"], "histAlphajj_Z", 20, 0, 1);
book(_h["histAlphajj_Z_1jet"], "histAlphajj_Z_1jet", 20, 0, 1);
book(_h["histNjets"], "histNjets", 6, 1.5, 7.5);

// Energy fraction observables
book(_h["histZ_Et_JJ_Et"], "histZ_Et_JJ_Et", 20, 0, 1);
book(_h["histZ_Et_ZJJ_Et"], "histZ_Et_ZJJ_Et", 20, 0, 1);
book(_h["histJ1_Et_ZJJ_Et"], "histJ1_Et_ZJJ_Et", 20, 0, 1);
book(_h["histZ_Et2_JJ_Et2"], "histZ_Et2_JJ_Et2", 20, 0, 1);
book(_h["histZ_Et2_ZJJ_Et2"], "histZ_Et2_ZJJ_Et2", 20, 0, 1);
book(_h["histJ1_Et2_ZJJ_Et2"], "histJ1_Et2_ZJJ_Et2", 20, 0, 1);
book(_h["histEt_total"], "histEt_total", 24, 30, 510);
book(_h["histpT_total"], "histpT_total", 24, 30, 510);

// 2D histograms
book(_t["histjetpT_total_Jet0_pT"],
"histjetpT_total_Jet0_pT", 24, 30, 510, 25, 35, 550);
book(_t["histjetpT_total_Jet0_pT_reduced"],
"histjetpT_total_Jet0_pT_reduced", 17, 30, 200, 37, 30, 400);
```

```
    book(_t["histJet0_pT_jetpT_total"],
    "histJet0_pT_jetpT_total", 37, 35, 405, 24, 30, 510);
    book(_t["histJet0_pT_reduced_jetpT_total"],
    "histJet0_pT_reduced_jetpT_total", 34, 30, 200, 25, 50, 200);
    book(_t["histjetpT_total_Jet1_pT"],
    "histjetpT_total_Jet1_pT", 24, 30, 510, 25, 35, 550);
    book(_t["histjetpT_total_Jet1_pT_reduced"],
    "histjetpT_total_Jet1_pT_reduced", 17, 30, 200, 37, 30, 400);
    book(_t["histJet1_pT_jetpT_total"],
    "histJet1_pT_jetpT_total", 37, 35, 405, 24, 30, 510);
    book(_t["histJet1_pT_reduced_jetpT_total"],
    "histJet1_pT_reduced_jetpT_total", 34, 30, 200, 25, 50, 200);
  }

  /// Perform the per-event analysis
  void analyze(const Event& event) {

    // Basic event selection
    const FinalState& fs = applyProjection<FinalState>(event, "FS");
    if (fs.size() <= 2) vetoEvent;

    // Z boson reconstruction
    const ZFinder& zfinder_mu = apply<ZFinder>(event, "zfinder_mu");

    // Z boson selection
    if (zfinder_mu.empty()) vetoEvent;
    if (zfinder_mu.constituents().size() != 2) vetoEvent;
    if (zfinder_mu.bosons().size() != 1) vetoEvent;

    const Particles& zfmuons = sortBy(zfinder_mu.constituents(), cmpMomByPt);
    FourMomentum z_boson = zfinder_mu.bosons()[0].momentum();
    FourMomentum leading_muon = zfmuons[0].momentum();
    FourMomentum subleading_muon = zfmuons[1].momentum();

    // Muon acceptance cuts
    if (leading_muon.absrapidity() > 2.4 ||

    subleading_muon.absrapidity() > 2.4) vetoEvent;
    if (!inRange(z_boson.mass(), 75.0*GeV, 106.0*GeV)) vetoEvent;
    if (leading_muon.pT() < 40.0*GeV) vetoEvent;
    if (subleading_muon.pT() < 20.0*GeV) vetoEvent;

    // Dimuon kinematics
    const double DphiDimuon = deltaPhi(leading_muon, subleading_muon);
    if (DphiDimuon > 2.8) vetoEvent;

    // Jet reconstruction and selection
    const FastJets& jetfs = apply<FastJets>(event, "jets");
```

```
const Jets& jetss = jetfs.jetsByPt(Cuts::pT > 30*GeV

&& Cuts::abseta < 4.7);
const Jets jets = discardIfAnyDeltaRLess(jetss, zfmuons, 0.4);

if (jets.size() < 2) vetoEvent;

// Define jet system
FourMomentum leading_jet = jets[0].momentum();
FourMomentum subleading_jet = jets[1].momentum();
FourMomentum dijet = leading_jet + subleading_jet;

// Jet acceptance cuts
if (leading_jet.absrapidity() > 4.7) vetoEvent;
if (subleading_jet.absrapidity() > 4.7) vetoEvent;
if (leading_jet.pT() < 50.0*GeV) vetoEvent;
if (subleading_jet.pT() < 50.0*GeV) vetoEvent;

// Isolation requirements
if (deltaR(dijet, leading_muon) < 0.4) vetoEvent;
if (deltaR(dijet, subleading_muon) < 0.4) vetoEvent;

// VBF selection cuts
const double dijet_mass = dijet.mass()/GeV;
const double abs_deta_jj = abs(leading_jet.pseudorapidity() -
subleading_jet.pseudorapidity());

if (dijet_mass < 200*GeV) vetoEvent;
if (abs_deta_jj < 1.5) vetoEvent;

// Z boson rapidity gap requirement
if (z_boson.rapidity() < min(subleading_jet.rapidity(),

leading_jet.rapidity())) vetoEvent;
if (z_boson.rapidity() > max(subleading_jet.rapidity(),

leading_jet.rapidity())) vetoEvent;

// Calculate observables
const unsigned int Njets = jets.size();

// Jet kinematics
const double jet_pt_0 = leading_jet.pT()/GeV;
const double jet_pt_1 = subleading_jet.pT()/GeV;
const double jet_eta_0 = leading_jet.pseudorapidity();
const double jet_eta_1 = subleading_jet.pseudorapidity();
const double jet_rap_0 = leading_jet.rapidity();
const double jet_rap_1 = subleading_jet.rapidity();
```

```
const double jet_phi_0 = leading_jet.phi();
const double jet_phi_1 = subleading_jet.phi();

// Muon kinematics
const double muon_pt_0 = leading_muon.pT()/GeV;
const double muon_pt_1 = subleading_muon.pT()/GeV;
const double muon_eta_0 = leading_muon.pseudorapidity();
const double muon_eta_1 = subleading_muon.pseudorapidity();
const double muon_rap_0 = leading_muon.rapidity();
const double muon_rap_1 = subleading_muon.rapidity();
const double muon_phi_0 = leading_muon.phi();
const double muon_phi_1 = subleading_muon.phi();

// Z boson kinematics
const double z_boson_pt = z_boson.pT()/GeV;
const double z_boson_eta = z_boson.pseudorapidity();
const double z_boson_rap = z_boson.rapidity();
const double z_boson_phi = z_boson.phi();
const double z_boson_mass = z_boson.mass()/GeV;

// Calculate global jet properties
double pT_total = 0.0;
double Et_total = 0.0;
FourMomentum jets_total;

for (size_t n = 0; n < Njets; ++n) {
  jets_total += jets[n].momentum();
  Et_total += jets[n].momentum().Et();
  pT_total += jets[n].momentum().pT()/GeV;
}

// VBF-specific observables
const double RpThard = (leading_jet + subleading_jet +
leading_muon + subleading_muon).pT() /
                      (leading_jet.pT() + subleading_jet.pT() +
                      leading_muon.pT() + subleading_muon.pT());

const double Zstar = (z_boson_rap - 0.5*(jet_rap_0 + jet_rap_1)) /
(jet_rap_0 - jet_rap_1);

const double EtaThreeStarAlternative = jets_total.rapidity() -
0.5*(jet_rap_0 + jet_rap_1);

// Angular separations
const double DPhiMuJet1 = deltaPhi(z_boson, leading_jet);
const double DPhiMuJet2 = deltaPhi(z_boson, subleading_jet);
const double DPhiMuJets = deltaPhi(z_boson, dijet);
const double DRJetsMuon1 = deltaR(dijet, leading_muon);
```

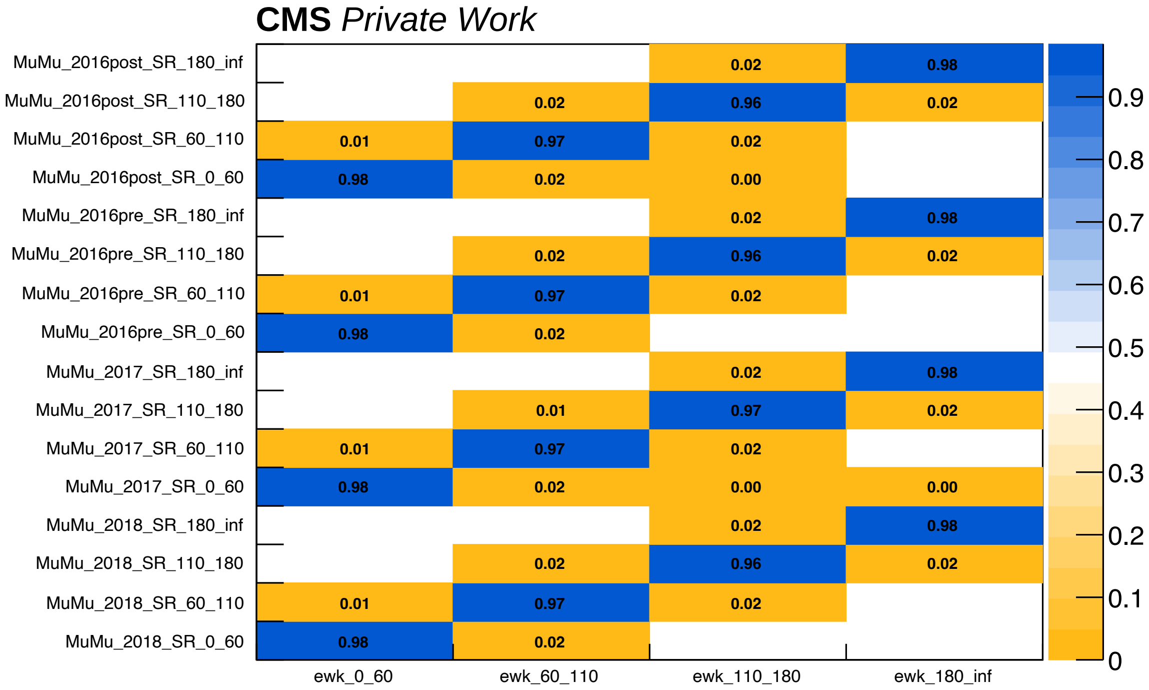
```
const double DRJetsMuon2 = deltaR(dijet, subleading_muon);

// Energy fractions
const double z_Et = z_boson.Et();
const double jet_Et_0 = leading_jet.Et();
const double jet_Et_1 = subleading_jet.Et();

const double Z_Et_JJ_Et = z_Et / (jet_Et_0 + jet_Et_1);
const double Z_Et_ZJJ_Et = z_Et / (z_Et + jet_Et_0 + jet_Et_1);
const double J1_Et_ZJJ_Et = jet_Et_0 / (z_Et + jet_Et_0 + jet_Et_1);

// Fill histograms - Muon observables
_h["histPtMuon1"]->fill(muon_pt_0);
_h["histPtMuon2"]->fill(muon_pt_1);
_h["histAbsetaMuon1"]->fill(abs(muon_eta_0));
_h["histAbsetaMuon2"]->fill(abs(muon_eta_1));
_h["histEtaMuon1"]->fill(muon_eta_0);
_h["histEtaMuon2"]->fill(muon_eta_1);
_h["histAbsRapMuon1"]->fill(abs(muon_rap_0));
_h["histAbsRapMuon2"]->fill(abs(muon_rap_1));
_h["histRapMuon1"]->fill(muon_rap_0);
_h["histRapMuon2"]->fill(muon_rap_1);
_h["histAbsPhiMuon1"]->fill(abs(muon_phi_0));
_h["histAbsPhiMuon2"]->fill(abs(muon_phi_1));

// Z/dimuon observables
_h["histPtDimuon"]->fill(z_boson_pt);
_h["histAbsetaDimuon"]->fill(abs(z_boson_eta));
_h["histAbsPhiDimuon"]->fill(abs(z_boson_phi));
_h["histMassDimuon"]->fill(z_boson_mass);
_h["histAbsRapDimuon"]->fill(abs(z_boson_rap));
_h["histAbsDetaDimuon"]->fill(abs(muon_eta_0 - muon_eta_1));
_h["histAbsDRapidityDimuon"]->fill(abs(muon_rap_0 - muon_rap_1));
_h["histDphiDimuon"]->fill(DphiDimuon);

// Jet observables
_h["histPtJet1"]->fill(jet_pt_0);
_h["histPtJet2"]->fill(jet_pt_1);
_h["histAbsetaJet1"]->fill(abs(jet_eta_0));
_h["histAbsetaJet2"]->fill(abs(jet_eta_1));
_h["histEtaJet1"]->fill(jet_eta_0);
_h["histEtaJet2"]->fill(jet_eta_1);
_h["histAbsRapJet1"]->fill(abs(jet_rap_0));
_h["histAbsRapJet2"]->fill(abs(jet_rap_1));
_h["histRapJet1"]->fill(jet_rap_0);
_h["histRapJet2"]->fill(jet_rap_1);
_h["histAbsPhiJet1"]->fill(abs(jet_phi_0));
_h["histAbsPhiJet2"]->fill(abs(jet_phi_1));
```

```
// Dijet observables
_h["histPtDijet"]->fill(dijet.pT()/GeV);
_h["histAbsetaDijet"]->fill(abs(dijet.pseudorapidity()));
_h["histMassDijet"]->fill(dijet_mass);
_h["histAbsPhiDijet"]->fill(abs(dijet.phi()));
_h["histAbsRapDijet"]->fill(abs(dijet.rapidity()));
_h["histAbsDetajj"]->fill(abs_deta_jj);
_h["histAbsDRapidityDijet"]->fill(abs(jet_rap_0 - jet_rap_1));
_h["histDphiDijet"]->fill(deltaPhi(leading_jet, subleading_jet));

// Angular correlations
_h["histDPhiMuJet1"]->fill(DPhiMuJet1);
_h["histDPhiMuJet2"]->fill(DPhiMuJet2);
_h["histDPhiMuJets"]->fill(DPhiMuJets);
_h["histDPhiJetsMu"]->fill(DPhiMuJets);
_h["histDRJetsMuon1"]->fill(DRJetsMuon1);
_h["histDRJetsMuon2"]->fill(DRJetsMuon2);

// VBF observables
_h["histRpThard"]->fill(RpThard);
_h["histZstar"]->fill(Zstar);
_h["histEtaThreeStarAlternative"]->fill(EtaThreeStarAlternative);
_h["histNjets"]->fill(Njets);
_h["histEt_total"]->fill(Et_total);
_h["histpT_total"]->fill(pT_total);

// Energy fractions
_h["histZ_Et_JJ_Et"]->fill(Z_Et_JJ_Et);
_h["histZ_Et_ZJJ_Et"]->fill(Z_Et_ZJJ_Et);
_h["histJ1_Et_ZJJ_Et"]->fill(J1_Et_ZJJ_Et);

// Additional observables for 3+ jet events
if (Njets >= 3) {
  const double EtaThreeStar = jets[2].momentum().rapidity() -
  0.5*(jet_rap_0 + jet_rap_1);
  const double Alphajj = jets[2].momentum().pT() /
  (0.5*(jet_pt_0 + jet_pt_1));
  const double Alphajj_Z = jets[2].momentum().pT() /
  (0.5*(jet_pt_0 + jet_pt_1) + z_boson_pt);
  const double DRMuJet3 = deltaR(z_boson, jets[2].momentum());

  _h["histEtaThreeStar"]->fill(EtaThreeStar);
  _h["histAlphajj"]->fill(Alphajj);
  _h["histAlphajj_Z"]->fill(Alphajj_Z);
  _h["histDRMuJet3"]->fill(DRMuJet3);
}
```

```
      // Fill 2D histograms
      _t["histjetpT_total_Jet0_pT"]->fill(pT_total, jet_pt_0);
      _t["histJet0_pT_jetpT_total"]->fill(jet_pt_0, pT_total);
      _t["histjetpT_total_Jet1_pT"]->fill(pT_total, jet_pt_1);
      _t["histJet1_pT_jetpT_total"]->fill(jet_pt_1, pT_total);
    }

    /// Normalize histograms after the run
    void finalize() {
      const double scale_factor = crossSection() / picobarn / sumW();

      // Scale all 1D histograms
      for (auto& hist : _h) {
        scale(hist.second, scale_factor);
      }

      // Note: 2D histograms typically don’t need scaling in the same way
    }

    /// @}

    /// @name Histograms
    /// @{
    map<string, Histo1DPtr> _h;
    map<string, Histo2DPtr> _t;
    /// @}
  };

  RIVET_DECLARE_PLUGIN(Rivet_ewzjj_endorsed);
}
```

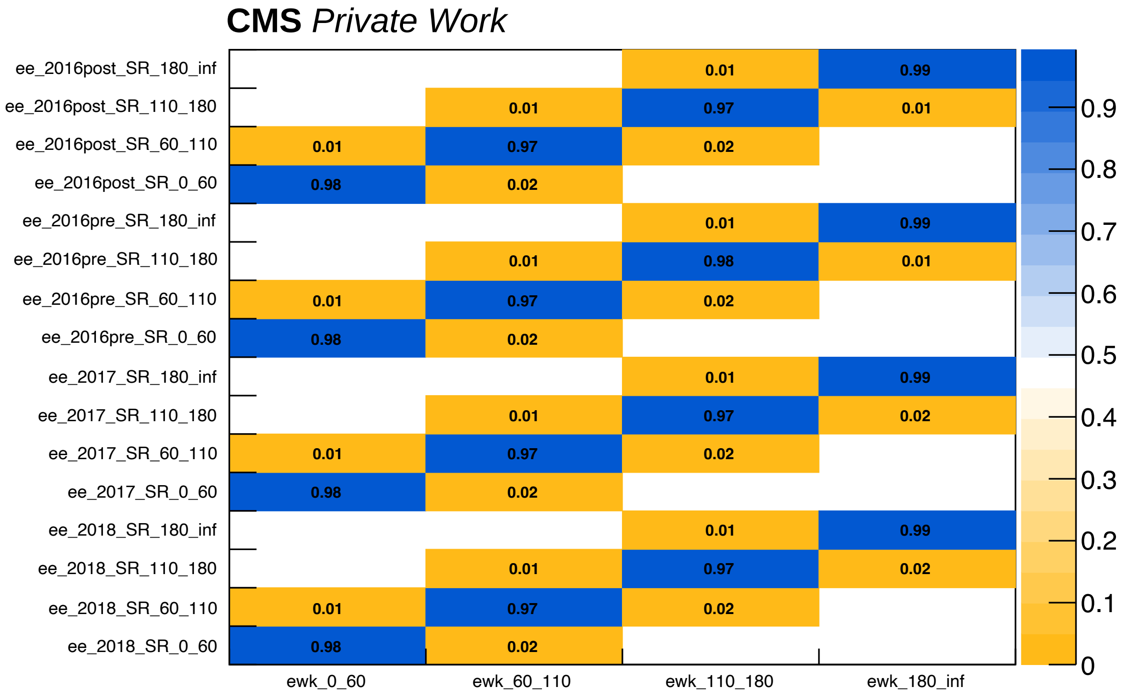
Appendix J

Signal migration matrices for EW Zjj measurement



	ewk_0_60	ewk_60_110	ewk_110_180	ewk_180_inf
MuMu_2016post_SR_180_inf			0.02	0.98
MuMu_2016post_SR_110_180		0.02	0.96	0.02
MuMu_2016post_SR_60_110	0.01	0.97	0.02	
MuMu_2016post_SR_0_60	0.98	0.02	0.00	
MuMu_2016pre_SR_180_inf			0.02	0.98
MuMu_2016pre_SR_110_180		0.02	0.96	0.02
MuMu_2016pre_SR_60_110	0.01	0.97	0.02	
MuMu_2016pre_SR_0_60	0.98	0.02		
MuMu_2017_SR_180_inf			0.02	0.98
MuMu_2017_SR_110_180		0.01	0.97	0.02
MuMu_2017_SR_60_110	0.01	0.97	0.02	
MuMu_2017_SR_0_60	0.98	0.02	0.00	0.00
MuMu_2018_SR_180_inf			0.02	0.98
MuMu_2018_SR_110_180		0.02	0.96	0.02
MuMu_2018_SR_60_110	0.01	0.97	0.02	
MuMu_2018_SR_0_60	0.98	0.02		

Figure J.1: Signal migration matrix in the $\mu\mu$ channel for the Full Run 2 generated via `Combine`.



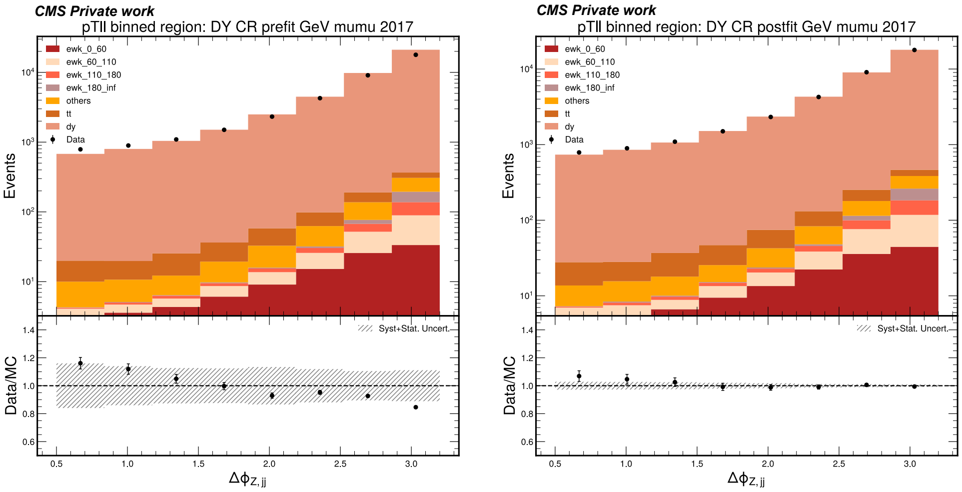
	ewk_0_60	ewk_60_110	ewk_110_180	ewk_180_inf
ee_2016post_SR_180_inf			0.01	0.99
ee_2016post_SR_110_180		0.01	0.97	0.01
ee_2016post_SR_60_110	0.01	0.97	0.02	
ee_2016post_SR_0_60	0.98	0.02		
ee_2016pre_SR_180_inf			0.01	0.99
ee_2016pre_SR_110_180		0.01	0.98	0.01
ee_2016pre_SR_60_110	0.01	0.97	0.02	
ee_2016pre_SR_0_60	0.98	0.02		
ee_2017_SR_180_inf			0.01	0.99
ee_2017_SR_110_180		0.01	0.97	0.02
ee_2017_SR_60_110	0.01	0.97	0.02	
ee_2017_SR_0_60	0.98	0.02		
ee_2018_SR_180_inf			0.01	0.99
ee_2018_SR_110_180		0.01	0.97	0.02
ee_2018_SR_60_110	0.01	0.97	0.02	
ee_2018_SR_0_60	0.98	0.02		

Figure J.2: Signal migration matrix in the *ee* channel for the Full Run 2 generated via `Combine`.

Appendix K

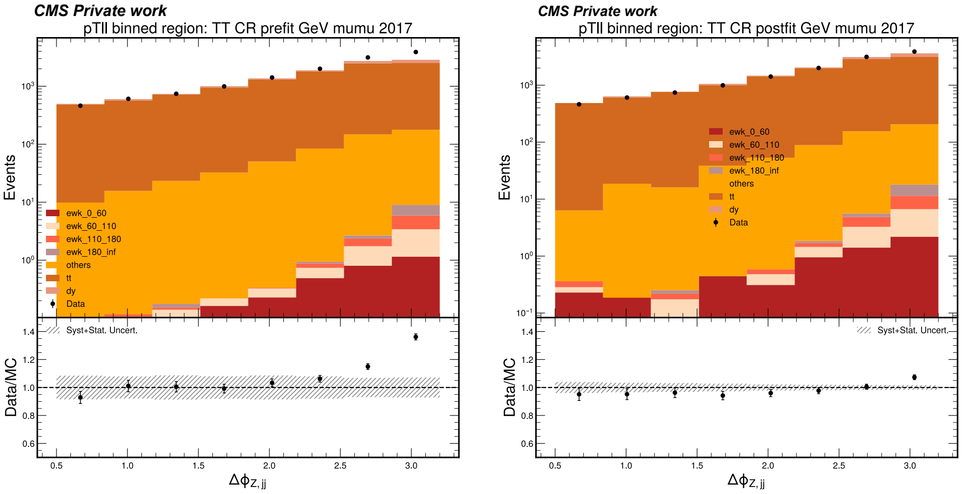
Control Region Distributions of $\Delta\Phi_{Z,jj}$ for EW Zjj Measurement

This appendix presents the $\Delta\phi_{Z,jj}$ detector-level distributions in the phase spaces defined by the DY and TT control regions for the EW Zjj cross-section measurement. Labels are defined as: `ewk_0_60`: EW Zjj signal process for particle level $0 < p_{Tll} <$ 60 GeV bin, `ewk_60_110`: EW Zjj signal process for particle level $60 < p_{Tll} <$ 110 GeV bin, `ewk_110_180`: EW Zjj signal process for particle level $110 < p_{Tll} <$ 180 GeV bin, `ewk_180_inf`: EW Zjj signal process for particle level $180 < p_{Tll}$ GeV bin, `others`: single top and diboson processes $(W^+W^-, W^\pm Z^0, Z^0Z^0)$, `tt`: top-antitop processes and `dy`: Drell-Yan process.



(a) DY CR pre-fit

(b) DY CR post-fit

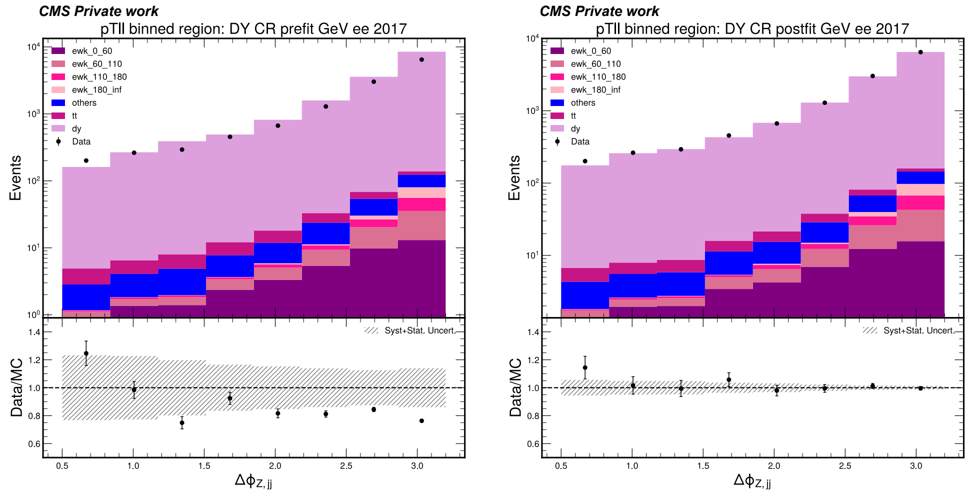


(c) TT CR pre-fit

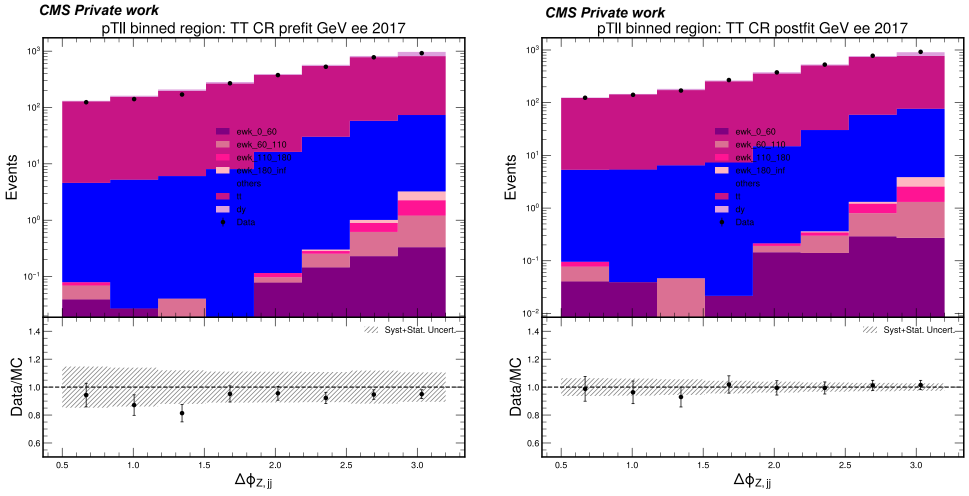
(d) TT CR post-fit

(e) MC labels

Figure K.1: $\Delta\Phi_{Z,jj}$ observable for $\mu\mu$ channel, 2017. Top row: Drell-Yan CR pre- and post-fit. Bottom row: $t\bar{t}$ CR pre- and post-fit.



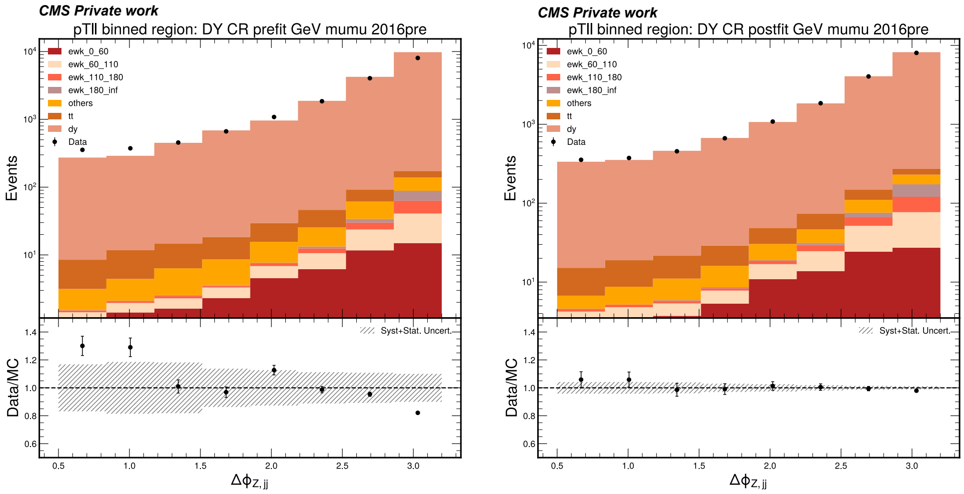
(a) DY CR pre-fit (b) DY CR post-fit



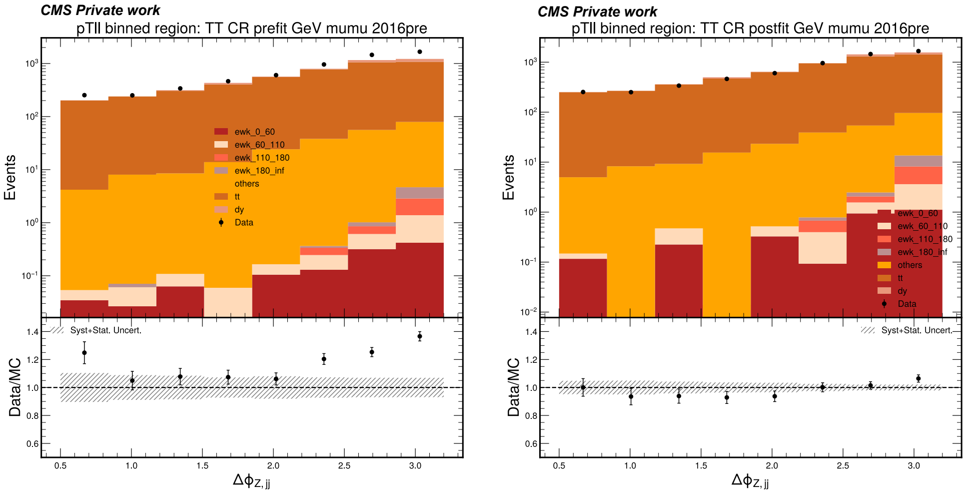
(c) TT CR pre-fit (d) TT CR post-fit

(e) MC labels

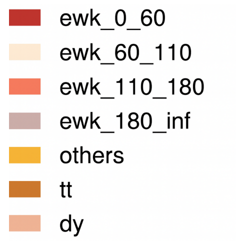
Figure K.2: $\Delta\Phi_{Z,jj}$ observable for ee channel, 2017. Top row: Drell-Yan CR pre- and post-fit. Bottom row: $t\bar{t}$ CR pre- and post-fit.



(a) DY CR pre-fit (b) DY CR post-fit

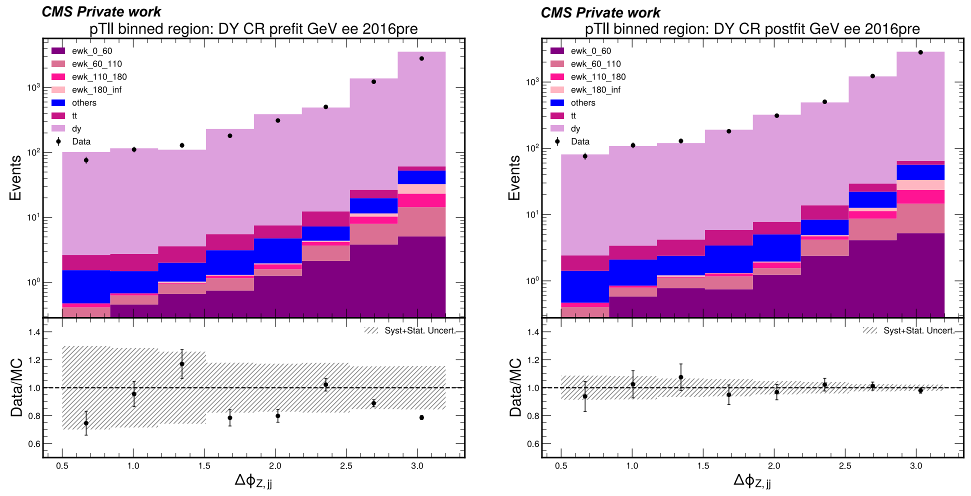


(c) TT CR pre-fit (d) TT CR post-fit

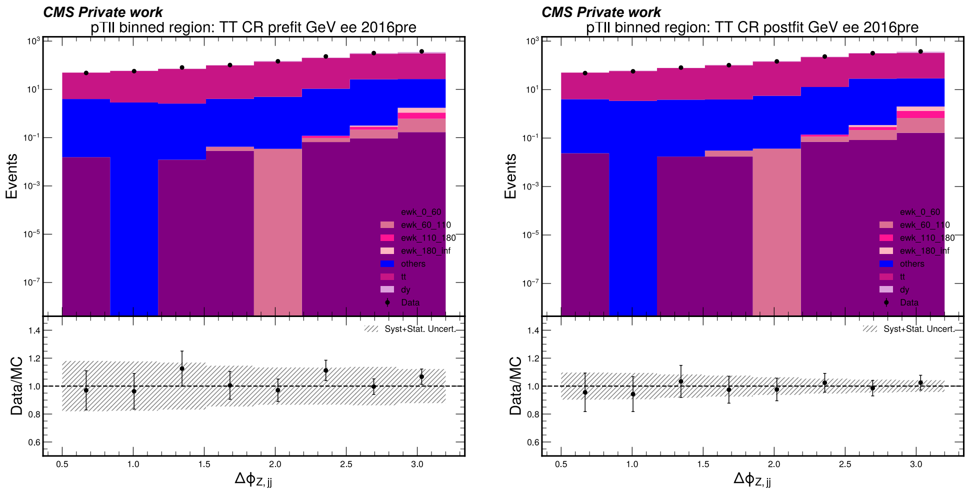


(e) MC labels

Figure K.3: $\Delta\Phi_{Z,jj}$ observable for $\mu\mu$ channel, 2016preVBF. Top row: Drell-Yan CR pre- and post-fit. Bottom row: $t\bar{t}$ CR pre- and post-fit.



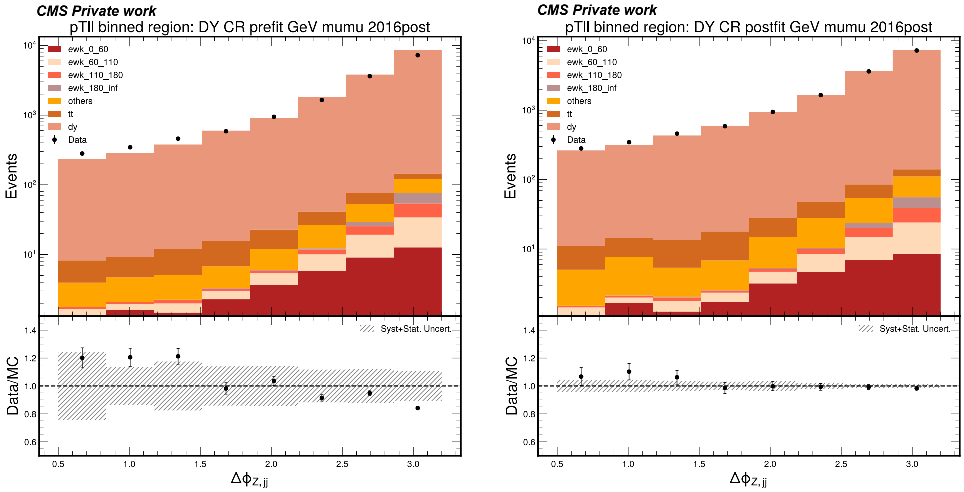
(a) DY CR pre-fit (b) DY CR post-fit



(c) TT CR pre-fit (d) TT CR post-fit

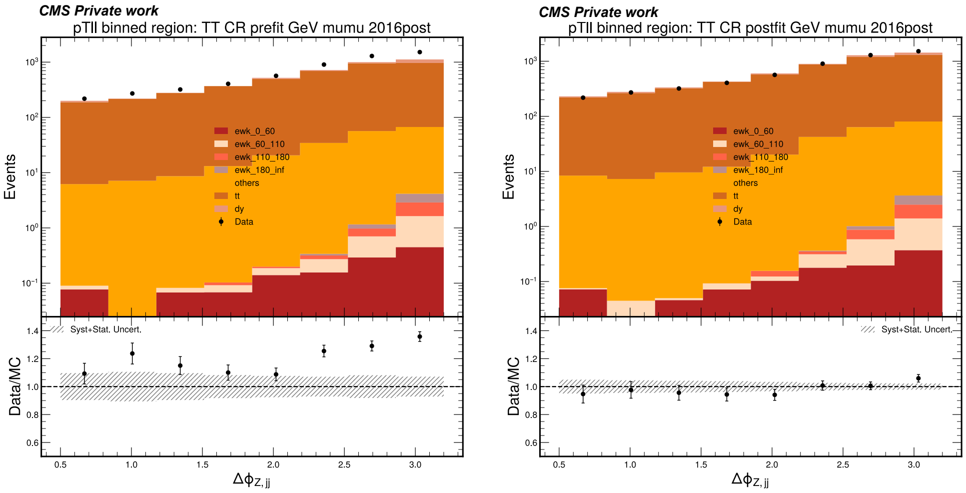
(e) MC labels

Figure K.4: $\Delta\Phi_{Z,jj}$ observable for ee channel, 2016preVBF. Top row: Drell-Yan CR pre- and post-fit. Bottom row: $t\bar{t}$ CR pre- and post-fit.



(a) DY CR pre-fit

(b) DY CR post-fit

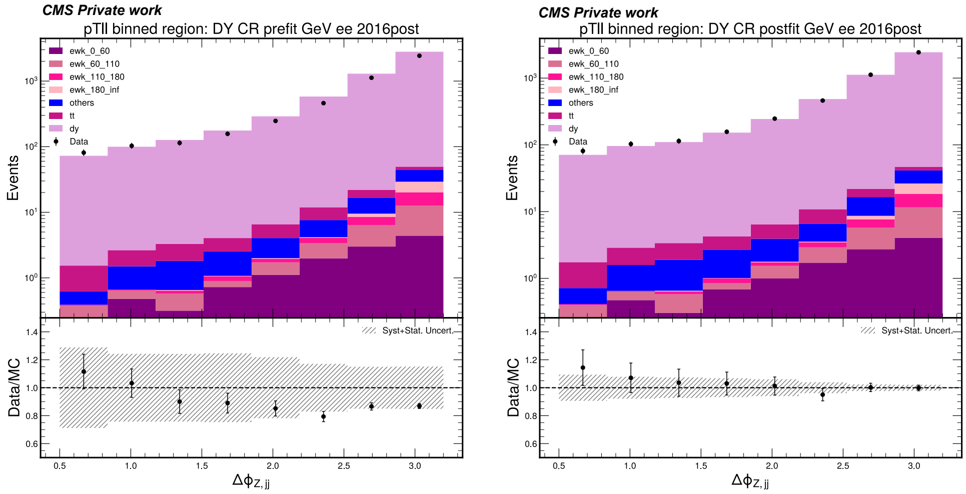


(c) TT CR pre-fit

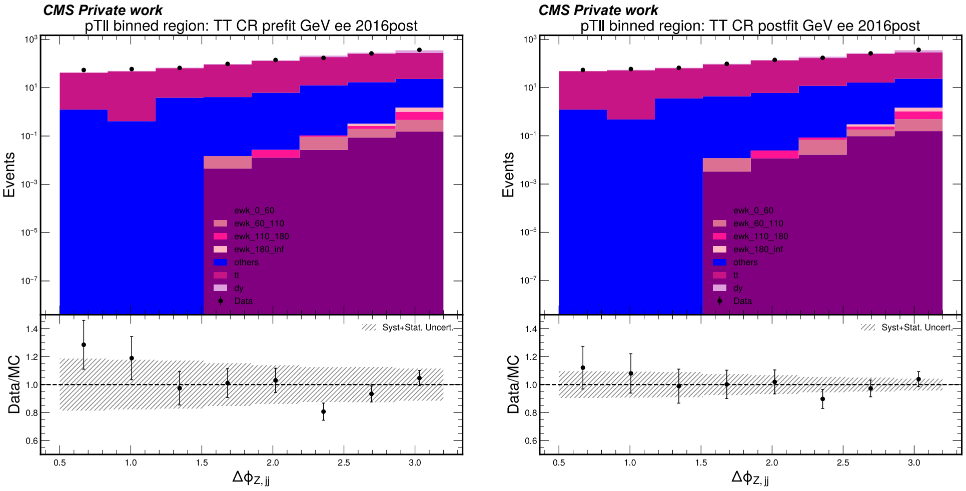
(d) TT CR post-fit

(e) MC labels

Figure K.5: $\Delta\Phi_{Z,jj}$ observable for $\mu\mu$ channel, 2016postVBF. Top row: Drell-Yan CR pre- and post-fit. Bottom row: $t\bar{t}$ CR pre- and post-fit.



(a) DY CR pre-fit (b) DY CR post-fit



(c) TT CR pre-fit (d) TT CR post-fit

(e) MC labels

Figure K.6: $\Delta\Phi_{Z,jj}$ observable for ee channel, 2016postVBF. Top row: Drell-Yan CR pre- and post-fit. Bottom row: $t\bar{t}$ CR pre- and post-fit.

Appendix L

Publications

L.1 Back-to-back azimuthal correlations in Z+jet events at high transverse momentum in the TMD parton branching method at next-to-leading order

Eur. Phys. J. C (2022) 82:755
https://doi.org/10.1140/epjc/s10052-022-10715-0

THE EUROPEAN
PHYSICAL JOURNAL C

Letter

Back-to-back azimuthal correlations in Z+jet events at high transverse momentum in the TMD parton branching method at next-to-leading order

H. Yang[1,2], **A. Bermudez Martinez**[2], **L. I. Estevez Banos**[2], **F. Hautmann**[3,4,5], **H. Jung**[2,a], **M. Mendizabal**[2], **K. Moral Figueroa**[6], **S. Prestel**[7], **S. Taheri Monfared**[2], **A. M. van Kampen**[4], **Q. Wang**[1,2], **K. Wichmann**[2]

[1] School of Physics, Peking University, Beijing, China
[2] Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany
[3] Theoretical Physics Department, CERN, Geneva, Switzerland
[4] Elementary Particle Physics, University of Antwerp, Antwerp, Belgium
[5] University of Oxford, Oxford, UK
[6] University of Edinburgh, Edinburgh, UK
[7] Department of Astronomy and Theoretical Physics, Lund University, Lund, Sweden

Received: 7 April 2022 / Accepted: 16 August 2022 / Published online: 29 August 2022

Abstract Azimuthal correlations in Z+jet production at large transverse momenta are computed by matching Parton-Branching (PB) TMD parton distributions and showers with NLO calculations via MCatNLO. The predictions are compared with those for dijet production in the same kinematic range. The azimuthal correlations $\Delta\phi$ between the Z boson and the leading jet are steeper compared to those in dijet production at transverse momenta $\mathcal{O}(100)$ GeV, while they become similar for very high transverse momenta $\mathcal{O}(1000)$ GeV. The different patterns of Z+jet and dijet azimuthal correlations can be used to search for potential *factorization-breaking* effects in the back-to-back region, which depend on the different color and spin structure of the final states and their interferences with the initial states. In order to investigate these effects experimentally, we propose to measure the ratio of the distributions in $\Delta\phi$ for Z+jet- and multijet production at low and at high transverse momenta, and compare the results to predictions obtained assuming factorization. We examine the role of theoretical uncertainties by performing variations of the factorization scale, renormalization scale and matching scale. In particular, we present a comparative study of matching scale uncertainties in the cases of PB-TMD and collinear parton showers.

1 Introduction

The description of jet production in association with a Z boson in hadron–hadron collisions is an important test of predictions obtained in Quantum Chromodynamics (QCD), and provides a relevant background to Higgs boson studies and to new physics searches. The associated Z boson plus jet production has been measured by CDF and D0 in proton-antiproton collisions at a center-of-mass energy $\sqrt{s} = 1.96$ TeV [1,2]. At the LHC, the ATLAS and CMS collaborations have published measurements in proton-proton (pp) collisions at a center-of-mass energy $\sqrt{s} = 7$ TeV [3–5], 8 TeV [6] and 13 TeV [7,8]. Azimuthal correlations between Z bosons and jets have been measured at 8 TeV [6] and 13 TeV [8].

The distribution in the azimuthal angle $\Delta\phi$ between the Z boson and the jet is an especially sensitive observable, probing several aspects of QCD physics. At leading order in the strong coupling α_s, one has $\Delta\phi = \pi$. The smearing of this delta-like distribution is a measure of higher order QCD radiation. In the region near $\Delta\phi = \pi$, this is primarily soft gluon radiation, while in the region of small $\Delta\phi$ it is primarily hard QCD radiation. The large-$\Delta\phi$ region of nearly back-to-back Z boson and jet is influenced by both perturbative and non-perturbative QCD contributions. The relative significance of these contributions depends on the scale of the transverse momentum imbalance between the boson and the jet. Importantly, the resummation of soft multi-gluon emissions in the nearly back-to-back region probes the transverse momenta

---

[a] e-mail: hannes.jung@desy.de (corresponding author)

of the initial state partons, which can be described by transverse momentum dependent (TMD) [9] parton distribution functions (PDFs). Theoretical predictions for Z boson + jet production including soft gluon resummation have recently been given in Refs. [10–16].

All the experimental measurements of boson–jet azimuthal correlations that have been performed so far are in the kinematical range of transverse momenta of the Z boson and the jets of the order $p_{\mathrm{T}} \approx \mathcal{O}(100)$ GeV. In this kinematical range, fixed-order perturbative corrections beyond next-to-leading order (NLO) are sizeable, and at small $\Delta\phi$ NLO calculations are usually not sufficient for reliable predictions. For the large-$\Delta\phi$ region of nearly back-to-back Z boson and jet, the boson–jet p_{T} imbalance scale is of order a few GeV, which is significantly influenced by both perturbative resummation and non-perturbative effects. It is worth noting that all the experimental measurements performed up to now do not cover the large $\Delta\phi$, nearly back-to-back, region with sufficiently fine binning to investigate detailed features of QCD.

With the increase in luminosity at the LHC, it becomes possible to measure Z+jet production in the high p_{T} range, with $p_{\mathrm{T}} \approx \mathcal{O}(1000)$ GeV. In this work, we observe that in this kinematical range the resummation of soft gluons and TMD dynamics in the nearly back-to-back region can be explored in a new regime, characterized by boson–jet p_{T} imbalance scales on the order of a few ten GeV. The large-$\Delta\phi$ region, involving deviations of the order of the experimental angular resolution of about 1 degree from $\Delta\phi = \pi$, can be investigated by analyzing jets with measurable transverse momenta.

Based on the above observation, in this paper we propose experimental investigations of back-to-back azimuthal correlations in the $p_{\mathrm{T}} \approx \mathcal{O}(1000)$ GeV region, with a systematic scan of the large-$\Delta\phi$ regime from this high p_{T} region down to $p_{\mathrm{T}} \approx \mathcal{O}(100)$ GeV – a regime which is completely unexplored experimentally up to now. We present dedicated phenomenological studies of this $\Delta\phi$ region as a function of p_{T}, enabling one to explore boson–jet transverse momentum imbalances from a jet scale of several ten GeV down to the few GeV scale. To perform these studies, we use the Parton Branching (PB) approach [17,18] to TMD evolution, matched to NLO calculations of Z+jet production with MADGRAPH5_AMC@NLO [19]. This approach has already been successfully applied, across a wide energy and mass range, to the Z boson p_{T} spectrum at the LHC [20] and the Drell–Yan (DY) p_{T} spectrum at lower fixed-target energies [21], so that the investigation of the same method in the Z+jet case is compelling. The $\Delta\phi$ correlation in the kinematical range proposed in this paper allows one to study the interplay of perturbative and non-perturbative contributions to TMD dynamics (see e.g. [22] for the DY case) as a function of both the boson–jet p_{T} imbalance and the evolution scale of the TMD distribution itself, of the order of the hard scale of the process, given by the transverse momenta of the Z boson or the jet.

In a previous publication [23] we have investigated the $\Delta\phi_{12}$ correlation in high-p_{T} dijet events by applying TMD PDFs and parton shower together with NLO calculations of the hard scattering process. In multijet events the azimuthal correlation between two jets has been measured at the LHC by ATLAS and CMS [24–28]. The region of $\Delta\phi_{12} \to \pi$ is of special interest, since so-called *factorization-breaking* [29–31] effects could become important in the case of colored final states. Multijet production is believed to be sensitive to such effects, as well as vector boson + jet production [32]. In order to investigate factorization-breaking effects, we propose to compare the theoretical description of the azimuthal correlation $\Delta\phi_{12}$ in multijet production with the one in Z+jet production. A thorough investigation of azimuthal correlations in the back-to-back region in Z+jet events has been also performed in Ref. [11], addressing the issue of factorization-breaking.

In this report we compare in detail high-p_{T} dijet and Z+jet production by applying the PB TMD method [17,18] matched with NLO. In Ref. [23] the NLO PB TMD predictions have been found to describe well the measurements of dijet azimuthal correlations [27,28]. In the present paper we apply the same method to the calculation of Z+jet production, and present the corresponding predictions. We propose to use the same kinematic region for the high-p_{T} dijet and Z+jet production to allow a direct comparison of the angular observables in the two cases.

We will see that, in the region of leading transverse momenta of the order $p_{\mathrm{T}} \approx \mathcal{O}(100\ \mathrm{GeV})$, the boson–jet final state is more strongly correlated azimuthally than the jet–jet final state. As the transverse momenta increase above the electroweak symmetry breaking scale, $p_{\mathrm{T}} \approx \mathcal{O}(1000\ \mathrm{GeV})$, this difference is reduced, and the boson–jet and jet–jet become more similarly correlated. We connect this behavior to features of the partonic initial state and final state radiation in the boson–jet and jet–jet cases. Since potential factorization-breaking effects arise from color interferences of initial-state and final-state radiation, different breaking patterns can be expected for strong and weak azimuthal correlations, influencing differently the boson–jet and jet–jet cases. We therefore propose to systematically compare measurements of dijet and Z+jet distributions, scanning the phase space from low transverse momenta $p_{\mathrm{T}} \approx \mathcal{O}(100\ \mathrm{GeV})$ to high transverse momenta $p_{\mathrm{T}} \approx \mathcal{O}(1000\ \mathrm{GeV})$.

In the following, we start by describing the basic elements of the PB TMD method and the Z+jet calculation in Sect. 2. In Sect. 3 we present results for the Z+jet azimuthal correlations and compare them with the multijet case. We summarize in Sect. 4. In an Appendix we discuss technical details on the use of MCatNLO+CASCADE3.

2 Basic elements of the calculation

In this section we first recall the salient features of the PB TMD approach, summarizing the main concepts of the approach and its applications; then we describe the calculation of Z+jet production by the PB TMD method matched with NLO matrix elements in MADGRAPH5_AMC@NLO.

2.1 PB-TMD method

The PB approach [18] provides a formulation for the evolution of TMD parton distributions in terms of perturbatively calculable Sudakov form factors and real-emission splitting kernels, with angular ordering phase space constraints and with non-perturbative distributions at the initial scale of the evolution to be determined from fits to experiment. This formulation uses a soft-gluon resolution scale z_M [17] to separate resolvable and non-resolvable branchings. An important feature of the PB TMD evolution equation [18] concerns its collinear limits: upon integration over all transverse momenta, the PB TMD evolution equation returns the DGLAP [33–36] equation for resolution scale $z_M \rightarrow 1$, while it coincides with the CMW [37,38] coherent branching equation for angular-ordered z_M [39]. The PB TMD method is based on the "unitarity" picture [40] of parton evolution usually employed in parton showering Monte Carlo (MC) algorithms [41,42]. The PB evolution equation for the TMD distributions is matched by a corresponding TMD parton shower for the spacelike parton cascade, generated by "backward evolution" [43]. A significant difference with respect to ordinary parton showers is that in the PB TMD method TMD distributions are defined and determined from fits to experimental data, which places constraints on fixed-scale inputs to evolution, while in ordinary parton showers instead non-perturbative physics parameters and showering parameters are tuned. No MC tuning is performed in the PB TMD case.

The NLO PB collinear and TMD parton distributions were obtained in Ref. [44] from QCD fits to precision DIS data from HERA [45] using the xFitter analysis framework [46,47]. Two different sets, PB-NLO-2018-Set 1 and PB-NLO-2018-Set 2, were obtained, with PB-NLO-2018-Set 1 corresponding at collinear level to HERAPDF 2.0 NLO [45]. In PB-NLO-2018-Set 2 the transverse momentum (instead of the evolution scale in Set 1) is used as the scale in the running coupling α_s which corresponds to the angular ordering of soft gluon emissions in the initial-state parton evolution [38,39,48,49]. It has been shown in [20,21] that Set 2 provides a better description of experimental measurements for the Z-boson spectrum at low-p_T. Also, it has been shown in [23] that the transverse momentum scale in the running coupling α_s is important for a good description of data on di-jet angular correlations. In this paper we will concentrate on Set 2 only.

In Fig. 1 we show the TMD PDF distributions for up quarks and gluons at $x = 0.01$ and $\mu = 100$ and 1000 GeV for PB-NLO-2018-Set 2. The transverse momentum distribution of gluons is broader than that of quarks, due to gluon self-coupling and the different color factors. In Fig. 1 also the uncertainties of the distributions, as obtained from the fit [44], are shown. The differences in the transverse momentum spectra of quarks and gluons will show up in differences in azimuthal correlation distributions.

The PB TMD evolution equation resums Sudakov logarithms. Current calculations in the PB TMD approach are performed with leading-logarithm (LL) and next-to-leading-logarithm (NLL) accuracy. The accuracy can be systematically improved, and the extension to next-to-next-to-leading logarithmic (NNLL) accuracy is being studied. In this respect, the approach can be compared [50] with analytic resummation methods [51,52]. The extraction of TMD distributions from the PB TMD fits described above could be compared with extractions, such as [53,54], based on [51,52]. The TMDlib tool [55,56] is designed as an aid for such studies. On the other hand, while analytic resummation approaches apply to the inclusive transverse momentum spectrum, the PB TMD approach works at exclusive level and can be applied to make predictions not only for the inclusive spectrum but also for the structure of the final states.

A framework to compute theoretical predictions combining the PB TMD resummation with fixed-order NLO matrix elements in MADGRAPH5_AMC@NLO has been developed in [20,21]. The predictions [20] have been successfully compared with LHC measurements of Z boson p_T and ϕ^* distributions [57–59]. Predictions by this method have also been successful in describing [21] DY p_T spectra at lower masses and energies [60–63]. The significance of this result is enhanced by the recent observation [64] that fixed-order NNLO corrections are not extremely large in the kinematic region of the data. This framework has also been applied to di-jet production [23], and predictions for di-jet correlations have been found in good agreement with LHC measurements [27,28]. We will employ this framework for Z+jet production in the next subsection.

As a method which is applicable at the level of exclusive final states, the PB TMD approach can be used in the context of multi-jet merging algorithms. A TMD multi-jet merging method has been developed in [65]. Its application to Z boson + multi-jets production [65–67] illustrates that trans-

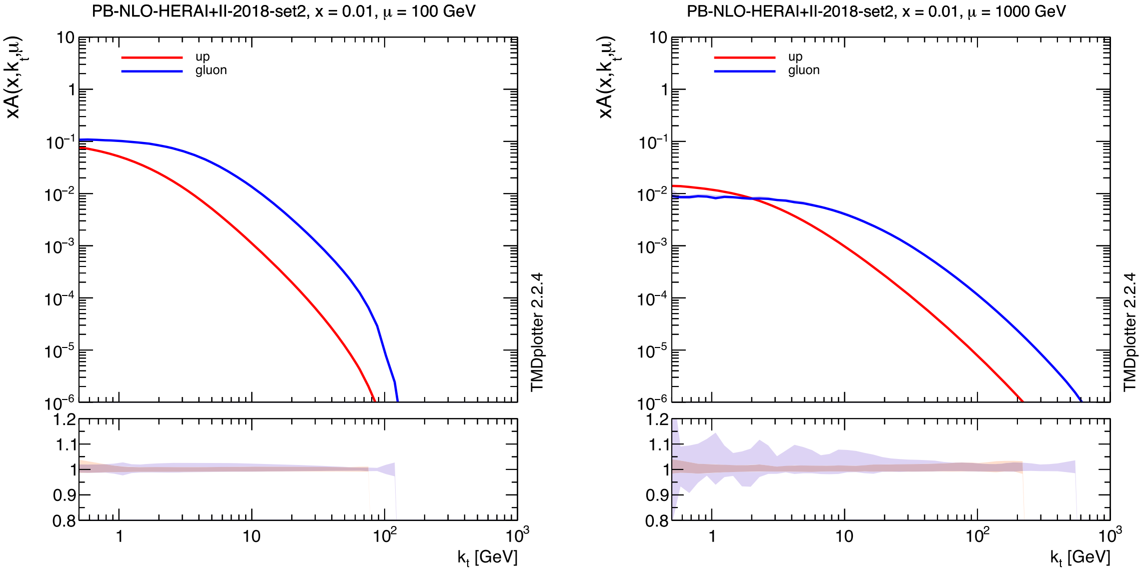


Fig. 1 TMD parton density distributions for up quarks and gluons of PB-NLO-2018-Set 2 as a function of k_T at $\mu = 100$ and 1000 GeV and $x = 0.01$. In the lower panels show the full uncertainty of the TMD PDFs, as obtained from the fits [44]

verse momentum recoils in the initial-state showers [68–70] influence significantly the theoretical systematics associated with the merging parameters. In the present paper, we will concentrate on the Z+jet back-to-back region, rather than the multi-jet production region, and we will therefore not use the TMD merging procedure.

Recently, the PB TMD evolution equation has been generalized to include TMD splitting functions [71,72], defined through high-energy factorization [73]. This generalization is important particularly for processes sensitive to TMD distributions at small values of longitudinal momentum fractions x. In this paper we focus on processes at mid to large x, and thus we do not consider this in the following.

2.2 Calculation of Z+jet distributions

The process Z+jet at NLO is calculated with MADGRAPH5_AMC@NLO using the collinear PB-NLO-2018-Set 2, as obtained in Ref. [44] applying $\alpha_s(M_Z) = 0.118$. The matching of NLO matrix elements with PB TMD parton distributions is described in Refs. [20,21,43]. The extension to multijet production is illustrated in Ref. [23]. Predictions are obtained by processing the MADGRAPH5_AMC@NLO event files in LHE format [74] through CASCADE3 [43] for an inclusion of TMD effects in the initial state and for simulation of the corresponding parton shower (labeled MCatNLO+CAS3 in the following).

Fixed order NLO Z+jet production is calculated with MADGRAPH5_AMC@NLO in a procedure similar to the one applied for dijet production described in [23] (labeled MCatNLO(fNLO)). For the MCatNLO mode, the HERWIG6 [75,76] subtraction terms are calculated, as they are best suited for the use with PB-parton densities, because both apply the same angular ordering condition. The use of HERWIG6 subtraction terms together with CASCADE3 is justified in the Appendix for final state parton shower as well as initial and final state showers by a comparison of the predictions obtained with CASCADE3 and with HERWIG6. The matching scale μ_m = SCALUP limits the contribution from PB-TMDs and TMD showers.

In the calculations, the factorization and renormalization scales are set to $\mu = \frac{1}{2}\sum_i p_{T,i}$, where the index i runs over all particles in the matrix element final state. This scale is also used in the PB-TMD parton distribution $\mathcal{A}(x, k_T, \mu)$. The scale uncertainties of the predictions are obtained from variations of the scales around the central value in the 7-point scheme avoiding extreme cases of variation.

In Fig. 2 we show the distributions of the transverse momentum of the Z+jet system, $p_{T,Zj}$, and the azimuthal correlation in the Z+jet system, $\Delta\phi_{Zj}$, for a fixed NLO calculation, for the full simulation including PB-TMD PDFs and parton showers as well as for the MCatNLO calculation at the level where subtraction terms are included without addition from parton shower (LHE-level). We require a transverse momentum $p_T > 200$ GeV for the Z boson and define jets with the anti-k_T jet-algorithm [77], as implemented in the FASTJET package [78], with a distance parameter of R = 0.4. The effect of including PB-TMD PDFs and parton showers can be clearly seen from the difference to the fixed NLO and LHE-level calculations.

In the low $p_{T,Zj}$ region one can clearly see the expected steeply rising behavior of the fixed NLO prediction. In the $\Delta\phi_{Zj}$ distribution one can observe the limited region for fixed NLO at $\Delta\phi_{Zj} < 2/3\pi$, since at most two jets in

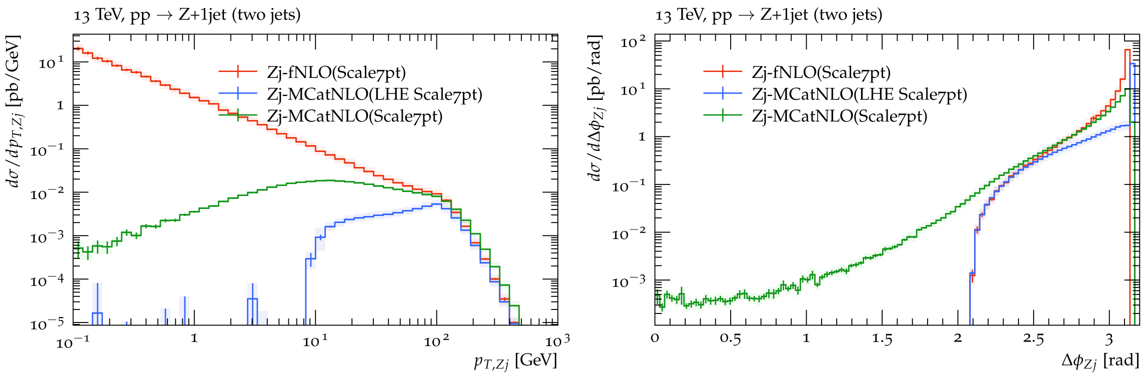


Fig. 2 Transverse momentum spectrum of the Z+jet-system $p_{T,Zj}$ (left) and $\Delta\phi_{\mathrm{Zj}}$ distribution (right). Shown are predictions from fixed NLO (fNLO), the (unphysical) distribution at LHE-level and the full simulation (after inclusion of PB-TMDs and TMD showers, MCatNLO+CAS3)

addition to the Z boson appear in the calculation. At large $\Delta\phi_{\mathrm{Zj}}$, the fixed NLO prediction rises faster than the full calculation including resummation via PB-TMDs and parton showers. In the following we concentrate on the large $\Delta\phi_{\mathrm{Zj}}$ region.

3 Back-to-back azimuthal correlations in Z+jet and multijet production

We now present predictions, obtained in the framework described above, for Z+jet and multijet production.[1] The selection of events follows the one of azimuthal correlations $\Delta\phi_{12}$ in the back-to-back region ($\Delta\phi_{12} \to \pi$) in multijet production at $\sqrt{s} = 13$ TeV as obtained by CMS [28]: jets are reconstructed with the anti-k_{T} algorithm [77] with a distance parameter of 0.4 in the rapidity range of $|y| < 2.4$. We require either two jets with $p_{\mathrm{T}}^{\mathrm{leading}} > 200$ GeV or a Z boson and a jet as leading or subleading objects with a transverse momentum $p_{\mathrm{T}}^{\mathrm{leading}} > 200$ GeV.

We consider distributions of the azimuthal correlation between the Z boson and the leading jet, $\Delta\phi_{\mathrm{Zj}}$, for $p_{\mathrm{T}}^{\mathrm{leading}} > 200$ GeV as well as for the very high p_{T} region of $p_{\mathrm{T}}^{\mathrm{leading}} > 1000$ GeV.

The calculations are performed with MCatNLO+CAS3 using PB-NLO-2018-Set 2 as the collinear and TMD parton densities with running coupling satisfying $\alpha_s(m_Z) = 0.118$ and PB-TMD parton shower.

In Fig. 3, the prediction for the azimuthal correlations $\Delta\phi_{\mathrm{Zj}}$ for Z+jet production in the back-to-back region is shown.[2] We also show, for comparison, the prediction of azimuthal correlations $\Delta\phi_{12}$ for multijet production in the same kinematic region, compared to the measurement of dijet production obtained by CMS [28]. We observe that the distribution of azimuthal angle $\Delta\phi_{\mathrm{Zj}}$ in Z+jet-production for $p_{\mathrm{T}}^{\mathrm{leading}} > 200$ GeV is more strongly correlated towards π than the distribution of angle $\Delta\phi_{12}$ in multijet production. This difference is reduced for $p_{\mathrm{T}}^{\mathrm{leading}} > 1000$ GeV.

Differences in $\Delta\phi$ between Z+jet and multijet production can result from the different flavor composition of the initial state and therefore different initial state transverse momenta and initial state parton shower, as well as from differences in final state showering since both processes have a different number of colored final state partons. Effects coming from factorization-breaking, interference between initial and final state partons, will depend on the final state structure and the number of colored final state partons.

We first investigate the role of initial state radiation and the dependence on the transverse momentum distributions coming from the TMD PDFs, which gives a large contribution to the decorrelation in $\Delta\phi$. The k_{T}-distribution obtained from a gluon TMD PDF is different from the one of a quark TMD PDF as shown in Fig. 1 for $x = 0.01$ and scales of $\mu = 200(1000)$ GeV. In Fig. 4 we show the probability of gg, qg and qq initial states (q stands for quark and antiquark) as a function of $p_{\mathrm{T}}^{\mathrm{leading}}$ for Z+jet and multijet production obtained with MCatNLO+CAS3. At high $p_{\mathrm{T}}^{\mathrm{leading}} > 1000$ GeV the qq channel becomes important for both Z+jet and multijet final states, while at lower $p_{\mathrm{T}}^{\mathrm{leading}} > 200$ GeV the gg channel is dominant in multijet production, leading to larger decorrelation effects, since gluons radiate more compared to quarks.

The role of final state radiation in the correlation in $\Delta\phi_{12}$ distributions is more difficult to estimate, since the subtraction terms for the NLO matrix element calculation also depend on the structure of the final state parton shower. In order to estimate the effect of final state shower we compare a calculation of the azimuthal correlations in the back-to-back region obtained with MCatNLO+CAS3 with the one obtained with MCatNLO+Pythia8 (Fig. 5). For the calcula-

---

[1] A framework based on CCFM evolution [79] was described in [80,81] for multi-jet and vector boson + jet correlations.

[2] Predictions for the region of small $\Delta\phi$ require including the contribution of higher parton multiplicities, e.g. via multi-jet merging [65].

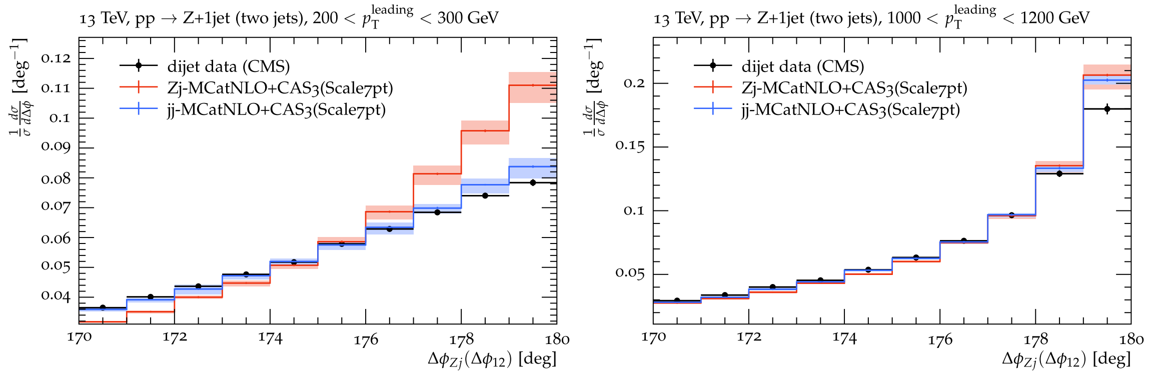


Fig. 3 Predictions of the azimuthal correlation $\Delta\phi_{\mathrm{Zj}}(\Delta\phi_{12})$ for Z+jet and multijet processes in the back-to-back region for $p_{\mathrm{T}}^{\mathrm{leading}} > 200$ GeV (left) and $p_{\mathrm{T}}^{\mathrm{leading}} > 1000$ GeV (right) obtained from MCatNLO+CAS3. Shown are the uncertainties obtained from scale variation (as described in the text). The measurements of dijet correlations as obtained by CMS [28] are shown as data points, for comparison

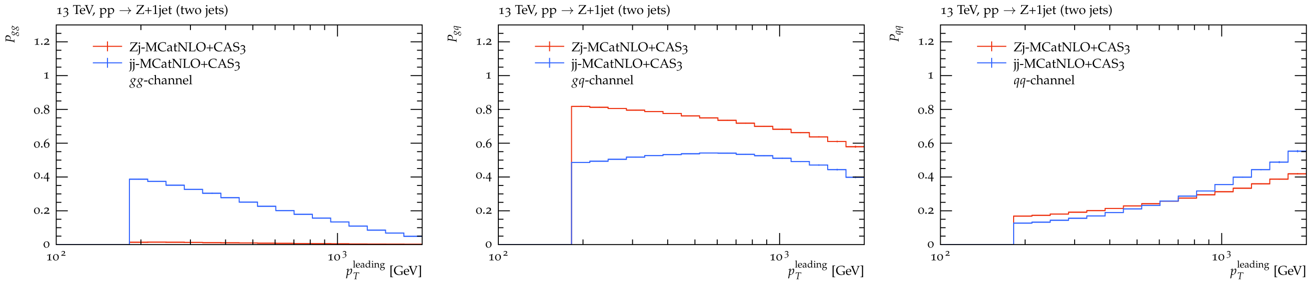


Fig. 4 The probability of gg, qg and qq initial states in Z+jet and multijet production (q stands for quark and antiquark) as a function of $p_{\mathrm{T}}^{\mathrm{leading}}$. The predictions are calculated with MCatNLO+CAS3

tion MCatNLO+PYTHIA8 we apply the PYTHIA8 subtraction terms in the MADGRAPH5_AMC@NLO calculation, use the NNPDF3.0 [82] parton density and tune CUETP8M1 [83].

As shown in Fig. 5, the distributions are different because of the different parton shower in CASCADE3 and PYTHIA8, but the ratio of the distributions for Z+jet and multijet production are similar: Z+jet-production gives a steeper (more strongly correlated) distribution at low $p_{\mathrm{T}}^{\mathrm{leading}}$, while at high $p_{\mathrm{T}}^{\mathrm{leading}}$ the distributions become similar in shape. We conclude, that the main effect of the $\Delta\phi$ decorrelation comes from initial state radiation, and the shape of the $\Delta\phi$ decorrelation in the back-to-back region becomes similar between Z+jet and dijet processes at high $p_{\mathrm{T}}^{\mathrm{leading}}$ where similar initial partonic states are important.

The matching scale μ_m limits the hardness of parton-shower emissions, and is thus typically a non-negligible source of variation in matched calculations (see e.g. [84] for a detailed discussion). It is thus interesting to assess the robustness of the previous findings under variations of the matching scale. Assessing matching scale variations in both an angular-ordered shower – such as CASCADE3 – and a transverse-momentum-ordered shower – such as PYTHIA8 – additionally tests the *interpretation (role)* of the matching scale. In transverse-momentum ordered showers, the matching scale sets the maximal transverse momentum of the first shower branchings, while branchings beyond the first emission are not explicitly affected by the matching scale. In an angular-ordered shower, however, the matching scale is applied as “veto scale” to avoid larger transverse momenta for any branching, i.e. the matching scale directly affects all branchings. The result of changing the matching scale to half or twice the central value is shown in Fig. 6. As expected, the value of the matching scale has an impact on the prediction (∼ 5%). This is particularly apparent when μ_m is used to set the maximal transverse momentum of the first emission in PYTHIA8. Overall, we find that interpreting the matching scale as veto scale in CASCADE3 leads to apparently more robust predictions. Interestingly, the matching scale uncertainty becomes smaller for higher-$p_{\mathrm{T}}^{\mathrm{leading}}$ jet configurations in CASCADE3. The size of the matching scale variation is comparable to scale variations, and should thus be carefully studied when designing uncertainty estimates.

In dijet production the measurements are rather well described with predictions obtained with MCatNLO+CAS3, as shown in Fig. 3 and discussed in detail in Ref. [23]. Only in the very high $p_{\mathrm{T}}^{\mathrm{leading}}$ region, a deviation from the mea-

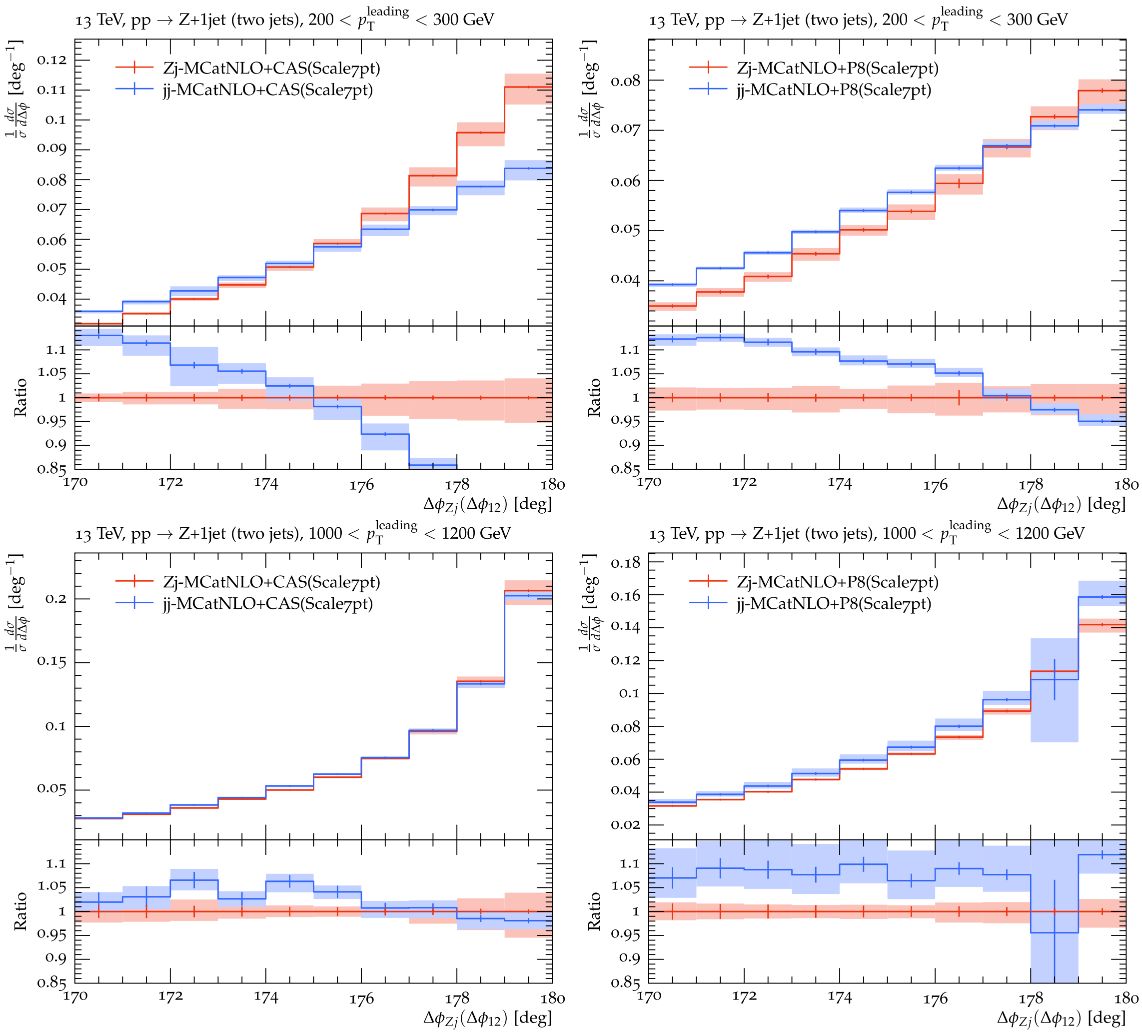


Fig. 5 Predictions for the azimuthal correlation $\Delta\phi_{\mathrm{Zj}}(\Delta\phi_{12})$ in the back-to-back region for Z+jet and multijet production obtained with MCatNLO+CAS3 (left column) and MCatNLO+PYTHIA8 (right column). Shown are different regions in $p_{\mathrm{T}}^{\text{leading}} > 200$ GeV (upper row) and $p_{\mathrm{T}}^{\text{leading}} > 1000$ GeV (lower row). The bands show the uncertainties obtained from scale variation (as described in the text)

surement is observed, which could be perhaps interpreted as coming from a violation of factorization. It is therefore very important to measure $\Delta\phi$ distributions in other processes, where factorization is expected to hold.

In order to experimentally probe effects which could originate from factorization-breaking in the back-to-back region we propose to measure the ratio of distributions in $\Delta\phi_{\mathrm{Zj}}$ for Z+jet and $\Delta\phi_{12}$ for multijet production at low and very high $p_{\mathrm{T}}^{\text{leading}}$, and compare the measurement with predictions assuming that factorization holds. The number of colored partons involved in Z+jet and multijet events is different, and deviations from factorization will depend on the structure of the colored initial and final state. In order to minimize the effect of different initial state configurations, a measurement at high $p_{\mathrm{T}}^{\text{leading}}$, hint more clearly at possible factorization-breaking effects.

In Ref. [11] a detailed study on Z+jet azimuthal correlations is reported, applying TMD-factorization and the "winner-takes-all" jet recombination scheme, with the aim to reduce potential factorization breaking contributions. We have checked that our main results remain largely unchanged when the "winner-takes-all" jet recombination scheme [85, 86] is applied and only in the last bin of the $\Delta\phi_{\mathrm{Zj}}$ distributions the cross section is reduced. We find that multijet events are more affected by the "winner-takes-all" jet recombination scheme in the back-to-back region at high p_{T} than Z+jet events.

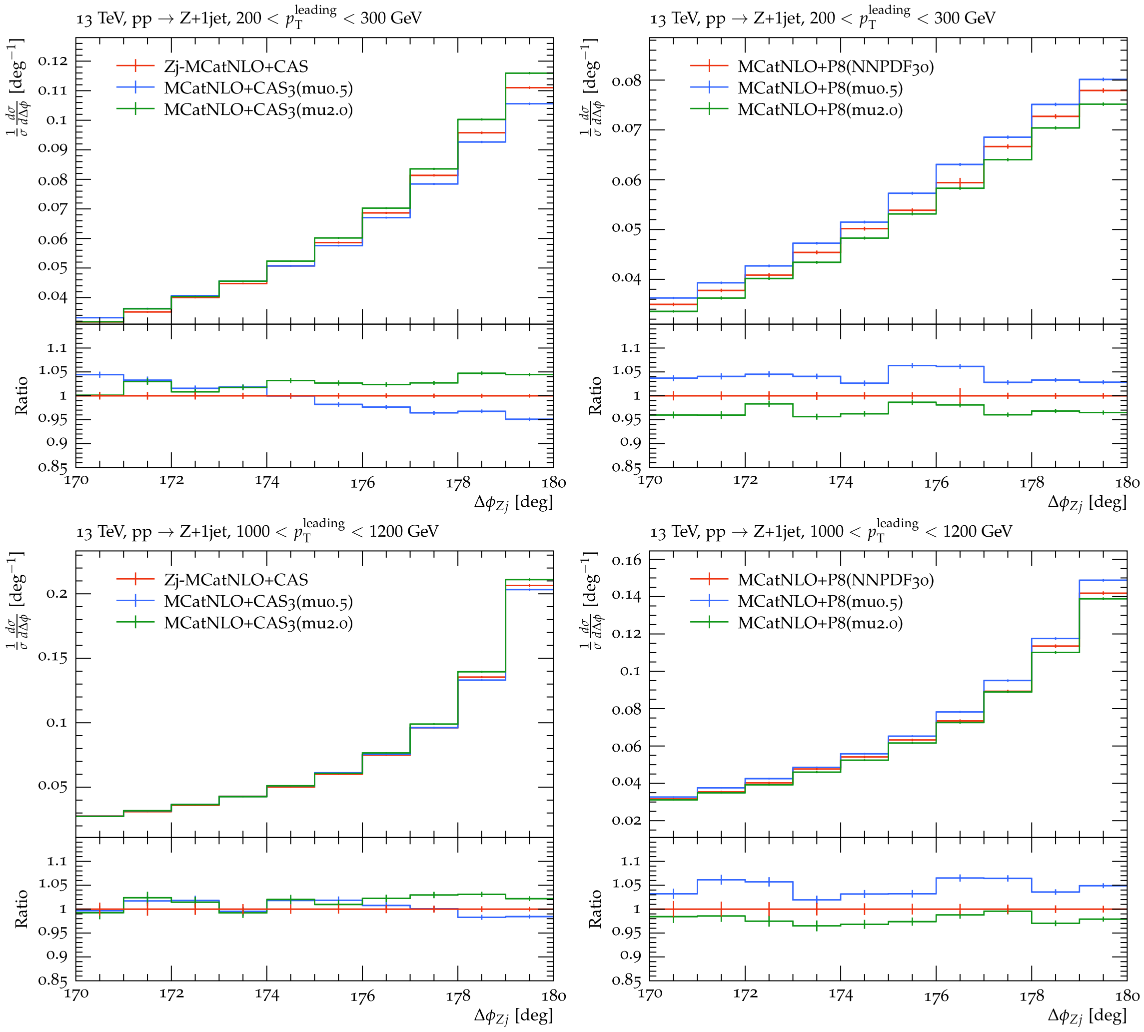


Fig. 6 The dependence on the variation of the matching scale μ_m in predictions for the azimuthal correlation $\Delta\phi_{\mathrm{Zj}}(\Delta\phi_{12})$ in the back-to-back region. Shown are predictions obtained with MCatNLO+CAS3 (left column) and MCatNLO+PYTHIA8 (right column) for $p_T^{\text{leading}} >$ 200 GeV (upper row) and $p_T^{\text{leading}} >$ 1000 GeV (lower row). The predictions with different matching scales μ_m varied by a factor of two up and down are shown

4 Summary and conclusions

We have investigated azimuthal correlations in Z+jet production and compared predictions with those for multijet production in the same kinematic range. The predictions are based on PB-TMD distributions with NLO calculations via MCatNLO supplemented by PB-TMD parton showers via CASCADE3. The azimuthal correlations $\Delta\phi_{\mathrm{Zj}}$, obtained in Z+jet production are steeper compared to those in multijet production ($\Delta\phi_{12}$) at transverse momenta $\mathcal{O}(100)$ GeV, while they become similar for very high transverse momenta, $\mathcal{O}(1000)$ GeV, which is a result of similar initial parton configuration of both processes.

In Z+jet production the color and spin structure of the partonic final state is different compared to the one in multijet production, and differences in the azimuthal correlation patterns can be used to search for potential factorization - breaking effects, involving initial and final state interferences. In order to experimentally investigate those effects, we propose to measure the ratio of the distributions in $\Delta\phi_{\mathrm{Zj}}$ for Z+jet- and $\Delta\phi_{12}$ for multijet production at low and at very high p_T^{leading}, and compare the measurements to predictions obtained assuming that factorization holds.

We have studied the matching scale dependence in the PB-TMD predictions and compared it with the case of NLO-matched calculations based on the PYTHIA8 collinear shower. We find that variations of the matching scale lead to more stable predictions in the PB-TMD case, with the relative reduction of the matching scale theoretical uncertainty becoming more pronounced for increasing p_T^{leading} transverse momenta.

Acknowledgements We are grateful to Olivier Mattelaer from the MadGraph5_aMC@NLO team for discussions, help and support with the LHE option for fixed NLO calculations in MCatNLO.

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors' comment: This is a phenomenological study without producing experimental data. The results of the Monte Carlo simulations are presented in the figures.]

Funded by SCOAP^3. SCOAP^3 supports the goals of the International Year of Basic Sciences for Sustainable Development.

Appendix: Comparison of CASCADE3 and HERWIG6

The calculations presented here apply the MCatNLO method using HERWIG6 (H6) subtraction terms, as implemented in MADGRAPH5_AMC@NLO. The NLO accuracy of the calculations is preserved by construction, since the use of PB-TMD distributions and TMD shower, as well as the ordinary parton shower, does not change the inclusive cross section.

Since HERWIG6 (H6) subtraction terms are used in the MCatNLO+CAS3 calculations, we investigate here in detail the contribution of the parton shower used in CASCADE3. We compare predictions obtained with MCatNLO+CAS3 with the corresponding ones obtained with MCatNLO+H6, using LHE files produced with MADGRAPH5_AMC@NLO for Z production. The Z boson is reconstructed from two oppositely charged leptons with $p_T > 20$ GeV in $|\eta| < 2.4$. We also study jet distributions obtained with the anti-k_T algorithm with distance parameter 0.4 with $p_T > 30$ GeV and $|\eta| < 5$.

In H6 the allowed region of z for a branching $q \to qg$ in the final state shower is $Q_q/Q < z < 1 - Q_g/Q$ (e.g. A.2.2 in Ref. [87]), with $Q_q = m_q + \texttt{VQCUT}$ and $Q_g = m_g + \texttt{VGCUT}$, and m_q, m_g being the quark and gluon effective masses, and $\texttt{VQCUT}, \texttt{VGCUT}$ the minimum virtuality parameters. Similar cuts are applied for initial state shower.

First we investigate final state parton showers. We compare distributions of the first and second jet in Z+jet events: the first (highest p_T) jet is part of the lowest order process, while the second (highest p_T) jet is the real correction and therefore subject to subtraction terms (keeping in mind that the highest p_T jet in the NLO calculation can also come from the α_s^2 real emission diagram). In CASCADE3, the PYTHIA6 final state shower is used (since the PB-method has not yet been applied for final state radiation), with the angular ordering veto condition. Since final state radiation is independent of parton densities, a direct comparison of MCatNLO+CAS3 and MCatNLO+H6, using the same LHE files, while only simulating final state radiation, is possible. In Fig. 7 we show a comparison of predictions for the transverse momentum of the first two highest p_T jets in Z+jet events (using identical LHE files).

The uncertainty coming from different parameter settings in the H6 final state parton shower is estimated by changing the light quark masses from the default to 0.32 GeV (`Rmas =`

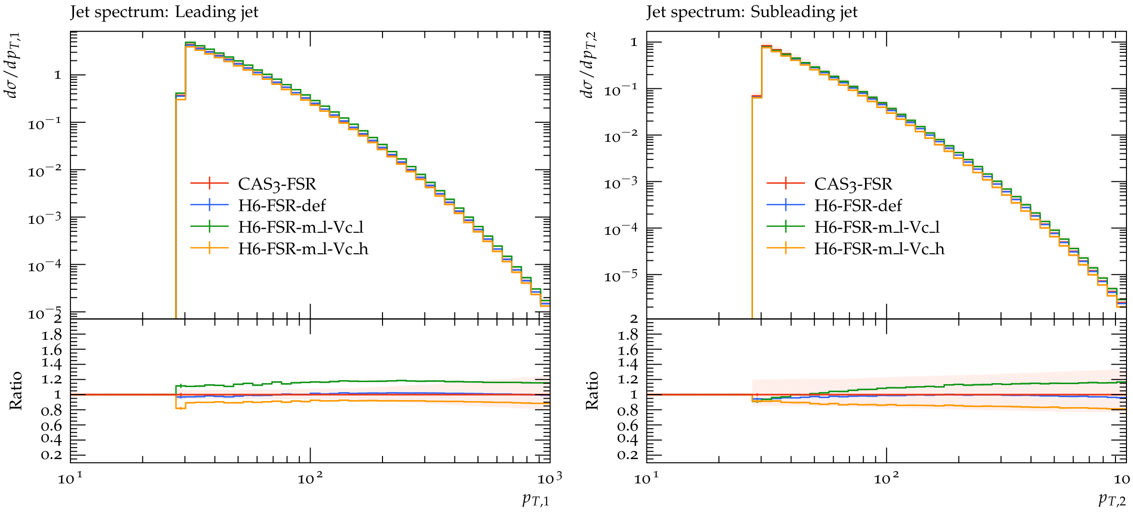


Fig. 7 Comparison of predictions obtained with MCatNLO+CAS3 and MCatNLO+H6 for Z+jet obtained with MCatNLO. Shown are predictions using only final state parton shower. The band of MCatNLO+CAS3 shows the uncertainties obtained from scale variation (as described in the text)

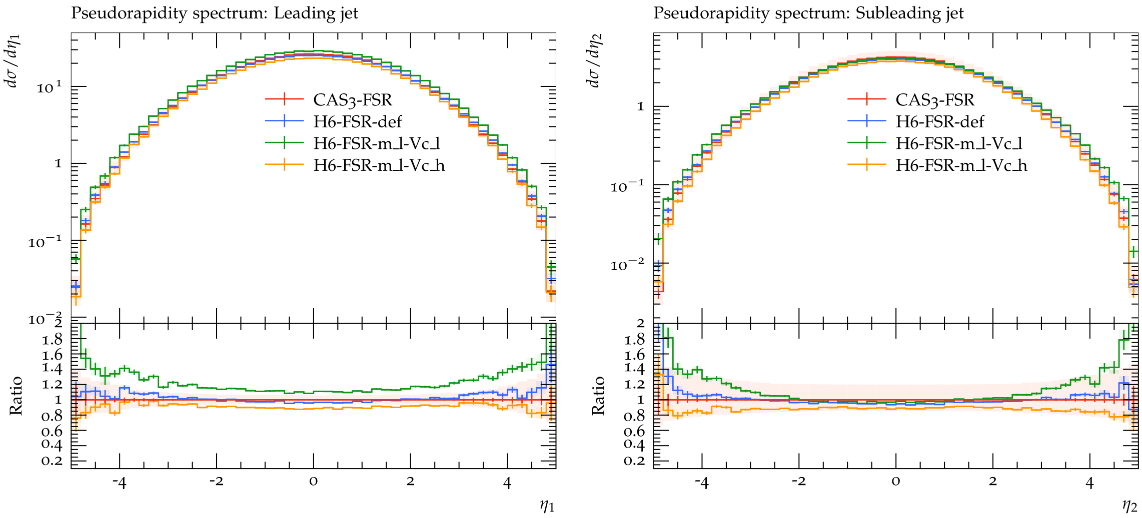


Fig. 8 Comparison of predictions obtained with MCatNLO+CAS3 and MCatNLO+H6 for Z+jet obtained with MCatNLO. Shown are predictions using only final state parton shower. The band of MCatNLO+CAS3 shows the uncertainties obtained from scale variation (as described in the text)

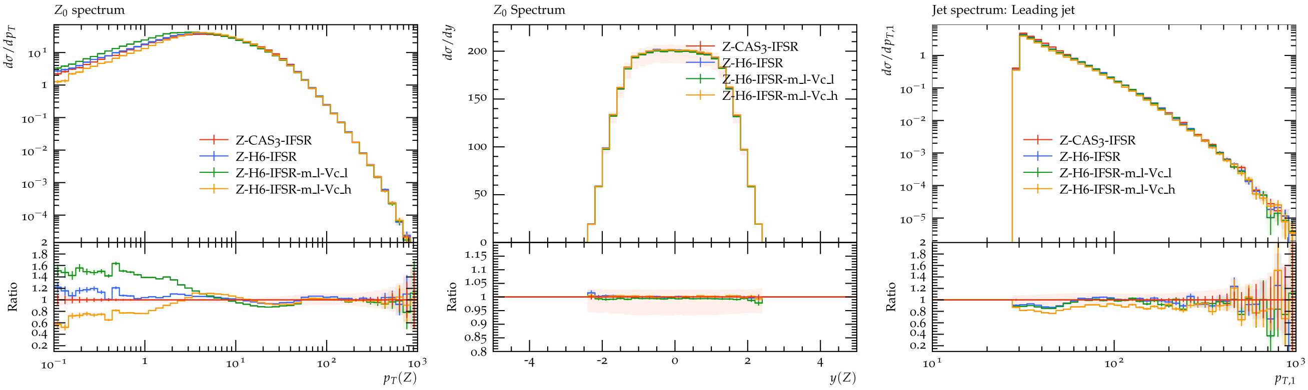


Fig. 9 Comparison of predictions obtained with MCatNLO+CAS3 and MCatNLO+H6 for Z production obtained with MCatNLO. Shown are predictions using initial state parton shower. The band of MCatNLO+CAS3 shows the uncertainties obtained from scale variation (as described in the text)

`0.32`, labelled as m_l) and `VQCUT`, `VGCUT` from the default to `0.1(1.5)`, labelled as $Vc_l(Vc_h)$, respectively (the lowest values chosen are those for which H6 is still working).

In Fig. 8 a comparison is shown for the pseudorapidity η of the first two highest p_{T} jets. Within the variation of the parameters, the prediction of MCatNLO+CAS3 agrees well with the one of MCatNLO+H6, justifying the application of the PYTHIA6 final state parton shower algorithm.

Next we investigate the contribution of PB-TMD PDFs and the PB-TMD parton shower in the initial state and compare the predictions with the ones from H6. We study Z production generated by MADGRAPH5_AMC@NLO which is essentially driven by initial state radiation. In Fig. 9 we show the transverse momentum of the Z boson, its rapidity distribution and the transverse momentum of the first reconstructed jet with $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 5$. Here the rapidity y of the Z boson is used, since it is related to the momentum fractions of the initial partons (instead of the pseudorapidity η which is used for jets as it is related to the scattering angle θ). We show a comparison of MCatNLO+CAS3 and MCatNLO+H6 predictions (including the same parameter variations for H6 as for the final state shower). In the region of low transverse momentum of the Z boson one can clearly see the sensitivity to the parameter choice in H6.

While at low p_{T} the parton shower matters, and the CAS3 prediction lies in between the ones from H6 with parameter variation, we observe good matching of the parton shower to the real emission at higher transverse momentum. The rapidity distribution obtained from CAS3 also lies within the one predicted by H6 with parameter variation. The p_{T} distribution of the first jet also agrees well within the band given by the uncertainties.

Finally we investigate Z+jet events, when both initial and final state radiation is important. In Fig. 10 we show a comparison of MCatNLO+CAS3 and MCatNLO+H6 predictions (including the same parameter variations for H6 as for the

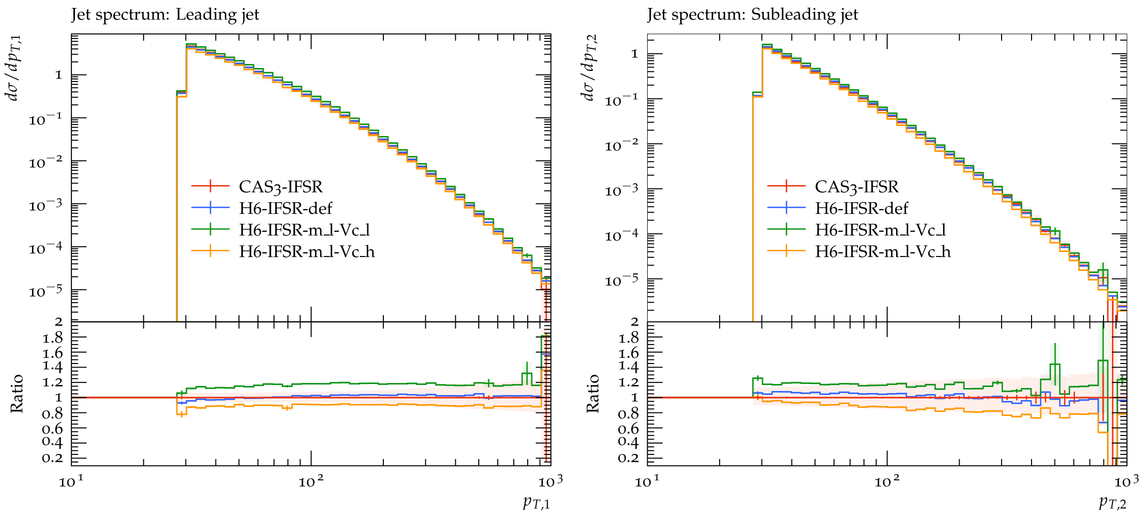


Fig. 10 Comparison of predictions obtained with MCatNLO+CAS3 and MCatNLO+H6 for Z+jet obtained with MCatNLO. Shown are predictions using initial and final state parton shower. The band of MCatNLO+CAS3 shows the uncertainties obtained from scale variation (as described in the text)

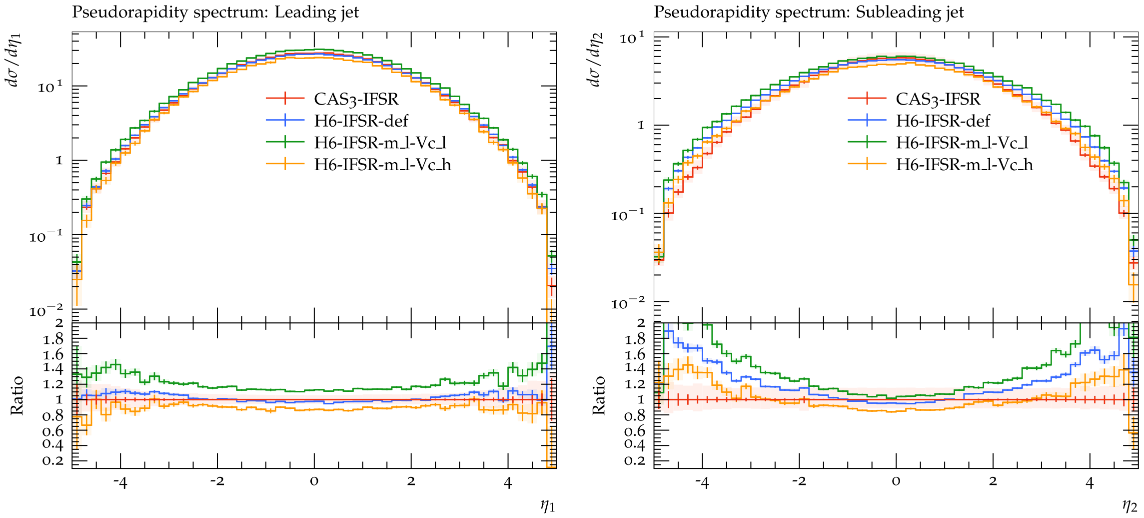


Fig. 11 Comparison of predictions obtained with MCatNLO+CAS3 and MCatNLO+H6 for Z+jet obtained with MCatNLO. Shown are predictions using initial and final state parton shower. The band of MCatNLO+CAS3 shows the uncertainties obtained from scale variation (as described in the text)

final state shower) for the transverse momentum of the first two highest p_T jets. In Fig. 11 the corresponding comparison is shown for the pseudorapidity distributions. The transverse momentum distributions agree well within the uncertainties coming from parameter variations, while for the η-distributions some differences in the very forward/backward regions are seen. However, one can see, that a variation of `VQCUT`, `VGCUT` has a significant effect especially in the forward/backward region.

In conclusion, we observe agreement between predictions obtained by MCatNLO+CAS3 and MCatNLO+H6 within the band of parton shower parameter variation in H6, confirming the use of H6 subtraction terms in MCatNLO together with PB - TMD PDFs, and PB - TMD initial state parton shower, as applied in MCatNLO+CAS3.

References

1. CDF Collaboration, Measurement of inclusive jet cross-sections in $Z/\gamma^* \to e^+e^-$ + jets production in $p\bar{p}$ collisions at $\sqrt{s}$ = 1.96-TeV. Phys. Rev. Lett. **100**, 102001 (2008). arXiv:0711.3717
2. D0 Collaboration, Measurement of differential Z/γ^* + jet + X cross sections in $p\bar{p}$ collisions at $\sqrt{s}$ = 1.96-TeV. Phys. Lett. B **669**, 278–286 (2008). arXiv:0808.1296
3. ATLAS Collaboration, Measurement of the production cross section of jets in association with a Z boson in pp collisions at

$\sqrt{s}$ = 7 TeV with the ATLAS detector. JHEP **07**, 032 (2013). arXiv:1304.7098

4. ATLAS Collaboration, Measurement of the production cross section for Z/γ^* in association with jets in pp collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector. Phys. Rev. D **85**, 032009 (2012). arXiv:1111.2690
5. CMS Collaboration, Measurements of jet multiplicity and differential production cross sections of $Z+$ jets events in proton–proton collisions at $\sqrt{s} = 7$ TeV. Phys. Rev. D **91**, 052008 (2015). arXiv:1408.3104
6. CMS Collaboration, Measurements of differential production cross sections for a Z boson in association with jets in pp collisions at $\sqrt{s} = 8$ TeV. JHEP **04**, 022 (2017). arXiv:1611.03844
7. ATLAS Collaboration, Measurements of the production cross section of a Z boson in association with jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. Eur. Phys. J. C **77**, 361 (2017). arXiv:1702.05725
8. CMS Collaboration, Measurement of differential cross sections for Z boson production in association with jets in proton–proton collisions at $\sqrt{s} = 13$ TeV. Eur. Phys. J. C **78**, 965 (2018). arXiv:1804.05252
9. R. Angeles-Martinez et al., Transverse Momentum Dependent (TMD) parton distribution functions: status and prospects. Acta Phys. Pol. B **46**(12), 2501 (2015). arXiv:1507.05267
10. H. Bouaziz, Y. Delenda, K. Khelifa-Kerfa, Azimuthal decorrelation between a jet and a Z boson at hadron colliders. arXiv:2207.10147
11. Y.-T. Chien et al., Precision boson-jet azimuthal decorrelation at hadron colliders. arXiv:2205.05104
12. L. Buonocore, M. Grazzini, J. Haag, L. Rottoli, Transverse-momentum resummation for boson plus jet production at hadron colliders. Eur. Phys. J. C **82**, 27 (2022). arXiv:2110.06913
13. Y.-T. Chien et al., Recoil-free azimuthal angle for precision boson-jet correlation. Phys. Lett. B **815**, 136124 (2021). arXiv:2005.12279
14. Y.-T. Chien, D.Y. Shao, B. Wu, Resummation of boson-jet correlation at hadron colliders. JHEP **11**, 025 (2019). arXiv:1905.01335
15. M.G.A. Buffing, Z.-B. Kang, K. Lee, X. Liu, A transverse momentum dependent framework for back-to-back photon+jet production. arXiv:1812.07549
16. P. Sun, B. Yan, C.P. Yuan, F. Yuan, Resummation of high order corrections in Z boson plus jet production at the LHC. Phys. Rev. D **100**, 054032 (2019). arXiv:1810.03804
17. F. Hautmann et al., Soft-gluon resolution scale in QCD evolution equations. Phys. Lett. B **772**, 446 (2017). arXiv:1704.01757
18. F. Hautmann et al., Collinear and TMD quark and gluon densities from Parton Branching solution of QCD evolution equations. JHEP **01**, 070 (2018). arXiv:1708.03279
19. J. Alwall et al., The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. JHEP **1407**, 079 (2014). arXiv:1405.0301
20. A. Bermudez Martinez et al., Production of Z-bosons in the parton branching method. Phys. Rev. D **100**, 074027 (2019). arXiv:1906.00919
21. A. Bermudez Martinez et al., The transverse momentum spectrum of low mass Drell–Yan production at next-to-leading order in the parton branching method. Eur. Phys. J. C **80**, 598 (2020). arXiv:2001.06488
22. F. Hautmann, I. Scimemi, A. Vladimirov, Non-perturbative contributions to vector-boson transverse momentum spectra in hadronic collisions. Phys. Lett. B **806** 135478 (2020)
23. M.I. Abdulhamid et al., Azimuthal correlations of high transverse momentum jets at next-to-leading order in the parton branching method. Eur. Phys. J. C **82**, 36 (2022). arXiv:2112.10465
24. ATLAS Collaboration, Measurement of dijet azimuthal decorrelations in pp collisions at sqrt(s) = 7 TeV. Phys. Rev. Lett. **106**, 172002 (2011). arXiv:1102.2696
25. CMS Collaboration, Dijet azimuthal decorrelations in pp collisions at sqrt(s) = 7 TeV. Phys. Rev. Lett. **106**, 122003 (2011). arXiv:1101.5029
26. CMS Collaboration, Measurement of dijet azimuthal decorrelation in pp collisions at $\sqrt{s} = 8$ TeV. Eur. Phys. J. C **76**, 536 (2016). arXiv:1602.04384
27. CMS Collaboration, Azimuthal correlations for inclusive 2-jet, 3-jet, and 4-jet events in pp collisions at $\sqrt{s} = 13$ TeV. Eur. Phys. J. C **78**, 566 (2018). arXiv:1712.05471
28. CMS Collaboration, Azimuthal separation in nearly back-to-back jet topologies in inclusive 2- and 3-jet events in pp collisions at $\sqrt{s} = 13$ TeV. Eur. Phys. J. C **79**, 773 (2019). arXiv:1902.04374
29. J. Collins, J.-W. Qiu, k_T factorization is violated in production of high-transverse-momentum particles in hadron–hadron collisions. Phys. Rev. D **75**, 114014 (2007). arXiv:0705.2141
30. W. Vogelsang, F. Yuan, Hadronic dijet imbalance and transverse-momentum dependent parton distributions. Phys. Rev. D **76**, 094013 (2007). arXiv:0708.4398
31. T.C. Rogers, P.J. Mulders, No generalized TMD-factorization in hadro-production of high transverse momentum hadrons. Phys. Rev. D **81**, 094006 (2010). arXiv:1001.2977
32. T.C. Rogers, Extra spin asymmetries from the breakdown of transverse-momentum-dependent factorization in hadron–hadron collisions. Phys. Rev. D **88**, 014002 (2013). arXiv:1304.4251
33. V.N. Gribov, L.N. Lipatov, Deep inelastic *ep* scattering in perturbation theory. Sov. J. Nucl. Phys. **15**, 438 (1972) [Yad. Fiz. **15**, 781 (1972)]
34. L.N. Lipatov, The parton model and perturbation theory. Sov. J. Nucl. Phys. **20**, 94 (1975) [Yad. Fiz. **20**, 181 (1974)]
35. G. Altarelli, G. Parisi, Asymptotic freedom in parton language. Nucl. Phys. B **126**, 298 (1977)
36. Y.L. Dokshitzer, Calculation of the structure functions for Deep Inelastic Scattering and e^+e^- annihilation by perturbation theory in Quantum Chromodynamics. Sov. Phys. JETP **46**, 641 (1977) [Zh. Eksp. Teor. Fiz. **73**, 1216 (1977)]
37. G. Marchesini, B.R. Webber, Monte Carlo simulation of general hard processes with coherent QCD radiation. Nucl. Phys. B **310**, 461 (1988)
38. S. Catani, B.R. Webber, G. Marchesini, QCD coherent branching and semiinclusive processes at large x. Nucl. Phys. B **349**, 635 (1991)
39. F. Hautmann, L. Keersmaekers, A. Lelek, A.M. Van Kampen, Dynamical resolution scale in transverse momentum distributions at the LHC. Nucl. Phys. B **949**, 114795 (2019). arXiv:1908.08524
40. B.R. Webber, Monte Carlo simulation of hard hadronic processes. Ann. Rev. Nucl. Part. Sci. **36**, 253–286 (1986)
41. J. Bellm et al., Herwig 7.0/Herwig++ 3.0 release note. Eur. Phys. J. C **76**, 196 (2016). arXiv:1512.01178
42. T. Sjöstrand et al., An introduction to PYTHIA 8.2. Comput. Phys. Commun. **191**, 159 (2015). arXiv:1410.3012
43. S. Baranov et al., CASCADE3 a Monte Carlo event generator based on TMDs. Eur. Phys. J. C **81**, 425 (2021). arXiv:2101.10221
44. A. Bermudez Martinez et al., Collinear and TMD parton densities from fits to precision DIS measurements in the parton branching method. Phys. Rev. D **99**, 074008 (2019). arXiv:1804.11152
45. ZEUS, H1 Collaboration, Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data. Eur. Phys. J. C **75**, 580 (2015). arXiv:1506.06042
46. xFitter Developers' Team Collaboration, H. Abdolmaleki et al., xFitter: an open source QCD analysis framework. A resource and reference document for the Snowmass study (2022). arXiv:2206.12465

47. S. Alekhin et al., HERAFitter. Eur. Phys. J. C **75**, 304 (2015). arXiv:1410.4412
48. A. Bassetto, M. Ciafaloni, G. Marchesini, Jet structure and infrared sensitive quantities in perturbative QCD. Phys. Rep. **100**, 201–272 (1983)
49. Y.L. Dokshitzer, V.A. Khoze, S.I. Troian, A.H. Mueller, QCD coherence in high-energy reactions. Rev. Mod. Phys. **60**, 373 (1988)
50. A.M. van Kampen, Drell-Yan transverse spectra at the LHC: a comparison of parton branching and analytical resummation approaches SciPost Phys. Proc. **8**, 151 (2022)
51. J.C. Collins, D.E. Soper, G.F. Sterman, Transverse momentum distribution in Drell–Yan pair and W and Z boson production. Nucl. Phys. B **250**, 199 (1985)
52. J. Collins, *Foundations of Perturbative QCD. Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology*, vol. 32 (2011)
53. A. Bacchetta et al., Unpolarized transverse momentum distributions from a global fit of Drell–Yan and semi-inclusive deep-inelastic scattering data. arXiv:2206.07598
54. M. Bury et al., PDF bias and flavor dependence in TMD distributions. arXiv:2201.07114
55. N.A. Abdulov et al., TMDlib2 and TMDplotter: a platform for 3D hadron structure studies. Eur. Phys. J. C **81**, 752 (2021). arXiv:2103.09741
56. F. Hautmann et al., TMDlib and TMDplotter: library and plotting tools for transverse-momentum-dependent parton distributions. Eur. Phys. J. C **74**(12), 3220 (2014). arXiv:1408.3015
57. ATLAS Collaboration, Measurement of the transverse momentum and ϕ^*_η distributions of Drell–Yan lepton pairs in proton–proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. Eur. Phys. J. C **76**, 291 (2016). arXiv:1512.02192
58. CMS Collaboration, Measurements of differential Z boson production cross sections in proton–proton collisions at $\sqrt{s} = 13$ TeV. JHEP **12**, 061 (2019). arXiv:1909.04133
59. CMS Collaboration, Measurement of the mass dependence of the transverse momentum of lepton pairs in Drell–Yan production in proton–proton collisions at $\sqrt{s} = 13$ TeV. arXiv:2205.04897
60. PHENIX Collaboration, Measurements of $\mu\mu$ pairs from open heavy flavor and Drell–Yan in $p + p$ collisions at $\sqrt{s} = 200$ GeV. Phys. Rev. D **99**, 072003 (2019). arXiv:1805.02448
61. D. Antreasyan et al., Dimuon scaling comparison at 44-GeV and 62-GeV. Phys. Rev. Lett. **48**, 302 (1982)
62. NuSea Collaboration, Absolute Drell–Yan dimuon cross sections in 800-GeV/c pp and pd collisions. arXiv:hep-ex/0302019
63. J.C. Webb, Measurement of continuum dimuon production in 800-GeV/c proton nucleon collisions. arXiv:hep-ex/0301031
64. R. Gauld et al., Transverse momentum distributions in low-mass Drell–Yan lepton pair production at NNLO QCD. Phys. Lett. B **829**, 137111 (2022). arXiv:2110.15839
65. A. Bermudez Martinez, F. Hautmann, M.L. Mangano, TMD evolution and multi-jet merging. Phys. Lett. B **822**, 136700 (2021). arXiv:2107.01224
66. A. Bermudez Martinez, F. Hautmann, M.L. Mangano, Multi-jet merging with TMD parton branching. arXiv:2208.02276
67. A. Bermudez Martinez, F. Hautmann, M.L. Mangano, Multi-jet physics at high-energy colliders and TMD parton evolution (2021). arXiv:2109.08173
68. F. Hautmann, QCD and jets. Acta Phys. Pol. B **44**, 761–778 (2013)
69. S. Dooling, P. Gunnellini, F. Hautmann, H. Jung, Longitudinal momentum shifts, showering and nonperturbative corrections in matched NLO-shower event generators. Phys. Rev. D **87**, 094009 (2013). arXiv:1212.6164
70. F. Hautmann, H. Jung, Collinearity approximations and kinematic shifts in partonic shower algorithms. Eur. Phys. J. C **72**, 2254 (2012). arXiv:1209.6549
71. F. Hautmann et al., A parton branching with transverse momentum dependent splitting functions. arXiv:2205.15873
72. L. Keersmaekers, Implementing transverse momentum dependent splitting functions in parton branching evolution equations. SciPost Phys. Proc. **8**, 127 (2022). arXiv:2109.07326
73. S. Catani, F. Hautmann, High-energy factorization and small x deep inelastic scattering beyond leading order. Nucl. Phys. B **427**, 475 (1994). arXiv:hep-ph/9405388
74. J. Alwall et al., A standard format for Les Houches event files. Comput. Phys. Commun. **176**, 300 (2007). arXiv:hep-ph/0609017
75. G. Corcella et al., HERWIG 6.5 release note. arXiv:hep-ph/0210213
76. G. Marchesini et al., HERWIG: a Monte Carlo event generator for simulating hadron emission reactions with interfering gluons. Version 5.1-April 1991. Comput. Phys. Commun. **67**, 465 (1992)
77. M. Cacciari, G.P. Salam, G. Soyez, The anti-k_t jet clustering algorithm. JHEP **04**, 063 (2008). arXiv:0802.1189
78. M. Cacciari, G.P. Salam, G. Soyez, FastJet user manual. Eur. Phys. J. C **72**, 1896 (2012). arXiv:1111.6097
79. H. Jung et al., The CCFM Monte Carlo generator CASCADE version 2.2.03. Eur. Phys. J. C **70**, 1237 (2010). arXiv:1008.0152
80. S. Dooling, F. Hautmann, H. Jung, Hadroproduction of electroweak gauge boson plus jets and TMD parton density functions. Phys. Lett. B **736**, 293 (2014). arXiv:1406.2994
81. F. Hautmann, H. Jung, Angular correlations in multi-jet final states from kt-dependent parton showers. JHEP **10**, 113 (2008). arXiv:0805.1049
82. NNPDF Collaboration, Parton distributions for the LHC Run II. JHEP **04**, 040 (2015). arXiv:1410.8849
83. CMS Collaboration, Event generator tunes obtained from underlying event and multiparton scattering measurements. Eur. Phys. J. C **76**, 155 (2016). arXiv:1512.00815
84. J. Bellm et al., Parton shower uncertainties with Herwig 7: benchmarks at leading order. Eur. Phys. J. C **76**, 665 (2016). arXiv:1605.01338
85. G. Salam, E_t^∞ scheme. Unpublished
86. D. Bertolini, T. Chan, J. Thaler, Jet observables without jet algorithms. JHEP **04**, 013 (2014). arXiv:1310.7584
87. S. Frixione, P. Nason, B.R. Webber, Matching NLO QCD and parton showers in heavy flavour production. JHEP **08**, 007 (2003). arXiv:hep-ph/0305252

L.2 Boson-jet and jet-jet azimuthal correlations at high transverse momenta

PROCEEDINGS OF SCIENCE

Boson-jet and jet-jet azimuthal correlations at high transverse momenta

A.M. van Kampen,[a,*] **A. Bermudez Martinez,**[b] **L.I. Estevez Banos,**[b] **F. Hautmann,**[a,c,d] **H. Jung,**[b] **M. Mendizabal,**[b] **K. Moral Figueroa,**[e] **S. Prestel,**[f] **S. Taheri Monfared,**[b] **Q. Wang,**[b,g] **K. Wichmann**[b] **and H. Yang**[b,g]

[a] *Elementary Particle Physics, University of Antwerp, Antwerp, Belgium*
[b] *Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany*
[c] *CERN, Geneva, Switzerland*
[d] *University of Oxford, Oxford, UK*
[e] *University of Edinburgh, Edinburgh, UK*
[f] *Department of Astronomy and Theoretical Physics, Lund University, Lund, Sweden*
[g] *School of Physics, Peking University, Beijing, China*

E-mail: `AronMees.vanKampen@uantwerpen.be`

We discuss our recent results on azimuthal distributions in vector boson + jets and multi-jet production at the LHC, obtained from the matching of next-to-leading order (NLO) perturbative matrix elements with transverse momentum dependent (TMD) parton branching. We present a comparative analysis of boson-jet and jet-jet correlations in the back-to-back region, and a study of the theoretical systematic uncertainties associated with the matching scale in the cases of TMD and collinear parton showers.

41st International Conference on High Energy physics - ICHEP2022
6-13 July, 2022
Bologna, Italy

---

[*] Speaker

https://pos.sissa.it/

Experiments at the Large Hadron Collider (LHC) carry out accurate measurements of azimuthal correlations in vector-boson plus jets [1, 2] and multi-jet [3–7] final states. When a boson and a jet, or two jets, recoil nearly back-to-back, reliable QCD predictions call for soft-gluon resummation of Sudakov processes — see recent studies of Refs. [8–14] in the boson-jet case and Refs. [15–18] in the di-jet case. This region probes the transverse momentum dependent (TMD) [19–21] parton distribution of the initial state.

With the increase in luminosity at the LHC, it becomes possible to explore this region experimentally over a wide kinematic range in the hard scale of the process, set by the highest transverse momentum p_T^{leading} produced into the final state, from $p_T^{\text{leading}} \approx O(100 \text{ GeV})$ to $p_T^{\text{leading}} \approx O(1000$ GeV). In particular, at the highest scales the nearly back-to-back region accessible with the experimental angular resolution of about 1 degree is characterized by transverse momentum imbalances of a few ten GeV, which can be investigated by analyzing jets with measurable transverse momenta.

The combined study of the p_T^{leading} and p_T-imbalance dependence of TMD dynamics is especially important, because the production of colored states near the back-to-back region may be influenced by factorization-breaking effects [22–25], due to interferences of gluon exchange in the soft region [26] with collinear radiation.

The present article is based on our work [27], in which we perform studies of azimuthal correlations as a function of p_T^{leading}, enabling one to explore p_T imbalances from a jet scale of several ten GeV down to the few GeV scale, and propose systematic measurements of ratios of boson-jet to jet-jet distributions, for varying transverse momenta, to investigate potential effects of soft-gluon interferences.

The studies [27] employ the parton branching (PB) approach [28, 29] to TMD evolution, and its matching [30–32] to next-to-leading-order (NLO) matrix-element calculations. The PB TMD evolution and corresponding parton shower are implemented in the Monte Carlo event generator CASCADE3 [33], and the NLO matching is performed using MADGRAPH5_AMC@NLO [34] (we label this calculational framework as MCatNLO+CAS3 in the following). TMD parton distributions at the starting scale of evolution are obtained from fits [35] to precision deep-inelastic scattering data [36] using the xFitter analysis framework [37, 38].

The comparison of NLO PB-TMD predictions and collinear-shower predictions with the measurements of di-jet azimuthal correlations [6, 7] is discussed in Refs. [32, 39]. The description of jet correlation data by NLO PB-TMD is good, and the comparison underlines the importance of transverse momentum recoils in the parton showers [40–43] and of angular ordering [44–46] in achieving this. For the boson-jet case, the measurements performed so far do not yet have TeV-scale transverse momenta and sufficiently fine binning to investigate detailed QCD features.

Fig. 1 [27] shows NLO-matched predictions from TMD shower (left) and collinear shower [47] (right), illustrating spectra in the $\Delta\phi$ azimuthal separation in Z + jet and di-jet systems. We concentrate on the large-$\Delta\phi$ region, as the low-$\Delta\phi$ decorrelation region requires going beyond the framework of the present calculation to include the contributions of higher jet multiplicities, e.g. via multi-jet merging techniques [48, 49]. For low p_T^{leading} the boson-jet final state is more strongly correlated azimuthally than the jet-jet final state (top panels). When the transverse momenta increase, the boson-jet and jet-jet states become more similarly correlated (bottom panels). One can connect this behavior to features of the partonic initial state and final state radiation in the two cases [27]. Since potential factorization-breaking effects arise from color interferences of initial-

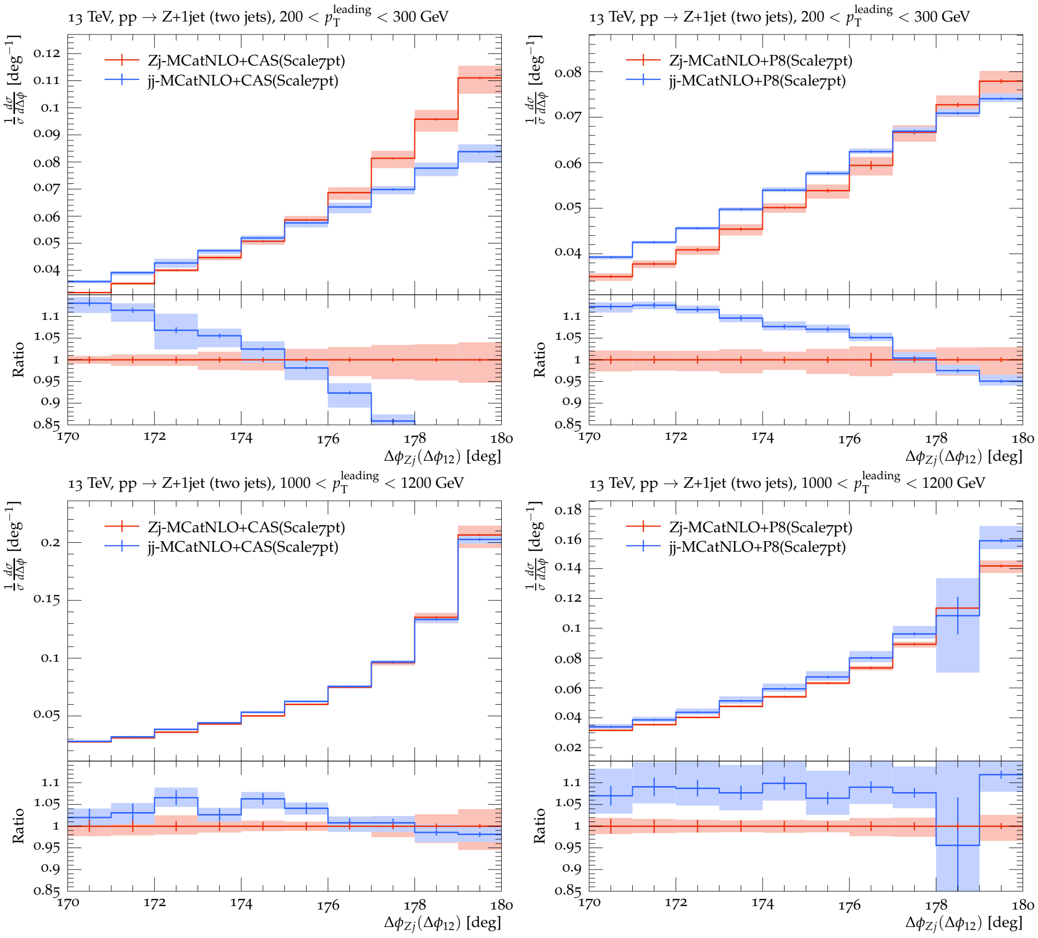


Figure 1: Predictions for the azimuthal correlation in the back-to-back region for Z+jets and multijet production obtained with MCatNLO+CAS3 (left column) and MCatNLO+Pythia8 (right column) [27]. Shown are different regions in $p_T > 200$ GeV (upper row) and $p_T > 1000$ GeV (lower row). The bands give an estimate of theoretical uncertainties obtained from scale variations as described in [27].

state and final-state radiation, different breaking patterns can be expected for strong and weak azimuthal correlations, influencing differently the boson-jet and jet-jet cases. Thus we propose to compare measurements of di-jet and Z + jet distributions systematically, scanning the phase space from low to high p_T^{leading}.

An important source of theoretical systematic uncertainties is given by the matching scale μ_m, limiting the hardness of parton shower radiation. In Fig. 2 we study this theoretical systematics for TMD shower (left) and collinear shower (right) calculations. Variations of the matching scale lead to more stable predictions in the TMD case, with the relative reduction of the matching scale theoretical uncertainty becoming more pronounced for increasing transverse momenta [27].

In conclusion, azimuthal distributions provide useful observables to gain insight into TMD dynamics and factorization — see e.g. [9, 32]. By varying the leading transverse momentum, the relative influence on the p_T imbalance can be studied from perturbative components and non-perturbative effects, e.g. [50, 51]. Sensitivity to TMD splitting probabilities [52] could also be

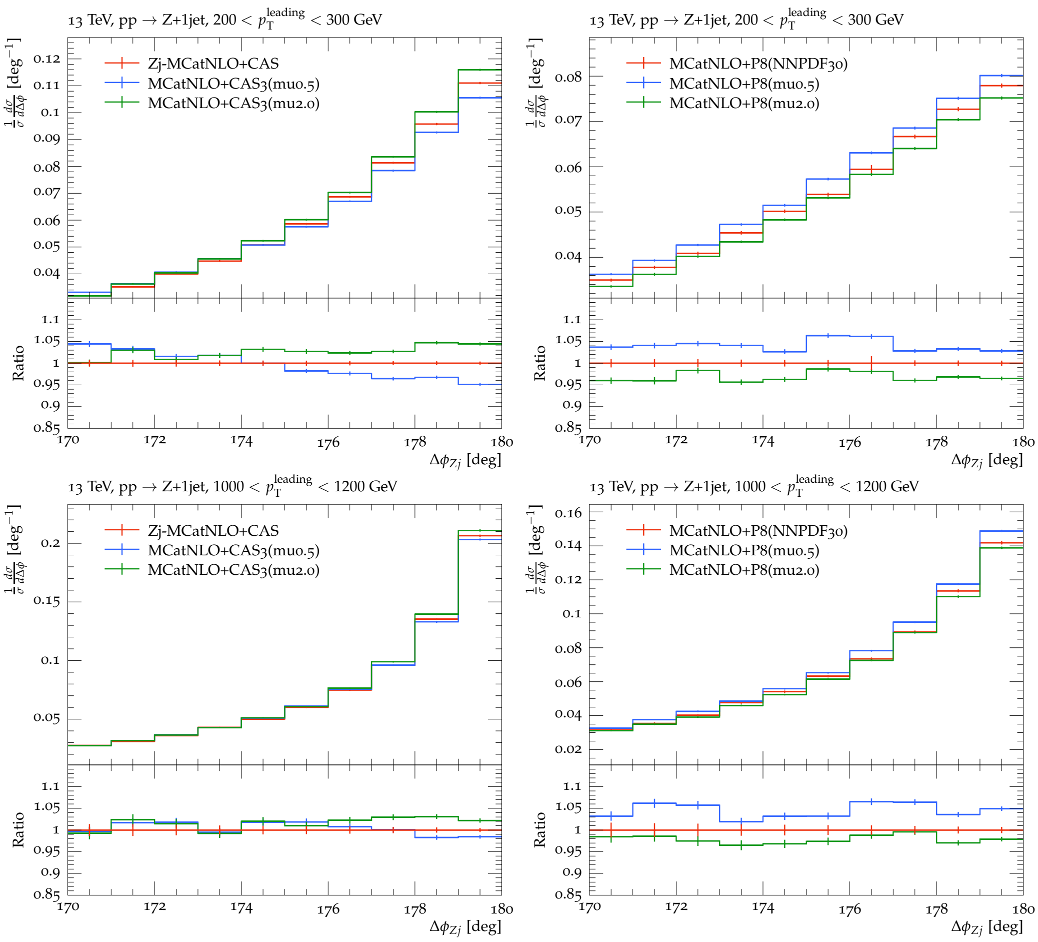


Figure 2: The dependence on the variation of the matching scale μ_m in predictions for the azimuthal correlation in the back-to-back region [27]. Shown are predictions obtained with MCatNLO+CAS3 (left column) and MCatNLO+Pythia8 (right column) for $p_T > 200$ GeV (upper row) and $p_T > 1000$ GeV (lower row). The predictions with different matching scales μ_m varied by a factor of two up and down are shown.

observed through azimuthal correlations.

Acknowledgments. We thank the conference organizers and convenors for putting up a great conference and for the invitation.

References

[1] CMS Coll., JHEP **04** (2017) 022 [arXiv:1611.03844 [hep-ex]].

[2] CMS Coll., Eur. Phys. J. C **78** (2018) 965 [arXiv:1804.05252 [hep-ex]].

[3] ATLAS Coll., Phys.Rev.Lett. **106** (2011) 172002 [arXiv:1102.2696 [hep-ex]].

[4] CMS Coll., Phys. Rev. Lett. **106** (2011) 122003 [arXiv:1101.5029 [hep-ex]].

[5] CMS Coll., Eur. Phys. J. C **76** (2016) 536 [arXiv:1602.04384 [hep-ex]].

[6] CMS Coll., Eur. Phys. J. C **78** (2018) 566 [arXiv:1712.05471 [hep-ex]].

[7] CMS Coll., Eur. Phys. J. C **79** (2019) 773 [arXiv:1902.04374 [hep-ex]].

[8] H. Bouaziz, Y. Delenda and K. Khelifa-Kerfa, arXiv:2207.10147 [hep-ph].

[9] Y.-T. Chien *et al.*, arXiv:2205.05104 [hep-ph].

[10] L. Buonocore, M. Grazzini, J. Haag and L. Rottoli, Eur. Phys. J. C **82** (2022) 27 [arXiv:2110.06913].

[11] Y.-T. Chien et al., Phys. Lett. B **815** (2021) 136124 [arXiv:2005.12279].

[12] Y.-T. Chien, D. Y. Shao and B. Wu, JHEP **11** (2019) 025 [arXiv:1905.01335].

[13] M. G. A. Buffing, Z.-B. Kang, K. Lee and X. Liu, arXiv:1812.07549.

[14] P. Sun, B. Yan, C. P. Yuan and F. Yuan, Phys. Rev. D **100** (2019) 054032 [arXiv:1810.03804].

[15] P. Sun, C. P. Yuan and F. Yuan, Phys. Rev. Lett. **113** (2014) 232001 [arXiv:1405.1105].

[16] P. Sun, C. P. Yuan and F. Yuan, Phys. Rev. **D92** (2015) 094007 [arXiv:1506.06170].

[17] Y. Hatta, B.-W. Xiao, F. Yuan, and J. Zhou, Phys. Rev. D **104** (2021) 054037 [arXiv:2106.05307].

[18] Y. Hatta, B.-W. Xiao, F. Yuan and J. Zhou, Phys. Rev. Lett. **126** (2021) 142001 [arXiv:2010.10774].

[19] N. A. Abdulov *et al.*, Eur. Phys. J. C **81** (2021) 752 [arXiv:2103.09741 [hep-ph]].

[20] F. Hautmann *et al.*, Eur. Phys. J. C **74** (2014) 3220 [arXiv:1408.3015 [hep-ph]].

[21] R. Angeles-Martinez *et al.*, Acta Phys. Pol. B 46 (2015) 2501 [arXiv:1507.05267].

[22] T. C. Rogers and P. J. Mulders, Phys. Rev. **D81** (2010) 094006 [arXiv:1001.2977].

[23] T. C. Rogers, Phys. Rev. D **88** (2013) 014002 [arXiv:1304.4251 [hep-ph]].

[24] J. Collins and J.-W. Qiu, Phys. Rev. **D75** (2007) 114014 [arXiv:0705.2141 [hep-ph]].

[25] W. Vogelsang and F. Yuan, Phys. Rev. D **76** (2007) 094013 [arXiv:0708.4398 [hep-ph]].

[26] J.C. Collins and F. Hautmann, Phys. Lett. B 472 (2000) 129 [hep-ph/9908467].

[27] H. Yang *et al.*, Eur. Phys. J. C **82** (2022) 755 [arXiv:2204.01528 [hep-ph]].

[28] F. Hautmann *et al.*, Phys. Lett. B 772 (2017) 446 [arXiv:1704.01757 [hep-ph]].

[29] F. Hautmann *et al.*, J. High Energy Phys. **01** (2018) 070 [arXiv:1708.03279 [hep-ph]].

[30] A. Bermudez Martinez *et al.*, Phys. Rev. D **100** (2019) 074027 [arXiv:1906.00919].

[31] A. Bermudez Martinez *et al.*, Eur. Phys. J. C **80** (2020) 598 [arXiv:2001.06488 [hep-ph]].

[32] M. I. Abdulhamid *et al.*, Eur. Phys. J. C **82** (2022) 36 [arXiv:2112.10465 [hep-ph]].

[33] S. Baranov *et al.*, Eur. Phys. J. C **81** (2021) 425 [arXiv:2101.10221 [hep-ph]].

[34] J. Alwall *et al.*, JHEP **1407** (2014) 079 [arXiv:1405.0301 [hep-ph]].

[35] A. Bermudez Martinez *et al.*, Phys. Rev. D99 (2019) 074008 [arXiv:1804.11152].

[36] ZEUS, H1 Coll., H. Abramowicz et al., Eur. Phys. J. C75 (2015) 580 [arXiv:1506.06042].

[37] H. Abdolmaleki *et al.* [xFitter Developers' Team], arXiv:2206.12465 [hep-ph].

[38] S. Alekhin *et al.*, Eur. Phys. J. C75 (2015) 304 [arXiv:1410.4412 [hep-ph]].

[39] A. Bermudez Martinez and F. Hautmann, arXiv:2208.08446 [hep-ph].

[40] S. Dooling *et al.*, Phys. Rev. D **87** (2013) 094009 [arXiv:1212.6164 [hep-ph]].

[41] F. Hautmann, Acta Phys. Polon. B **44** (2013) 761; arXiv:1304.8133 [hep-ph].

[42] H. Jung *et al.*, Eur. Phys. J. C **70** (2010) 1237 [arXiv:1008.0152 [hep-ph]].

[43] F. Hautmann and H. Jung, JHEP **10** (2008) 113; arXiv:0804.1746 [hep-ph].

[44] G. Marchesini and B. R. Webber, Nucl. Phys. B **310** (1988)461.

[45] S. Catani, B. R. Webber, and G. Marchesini, Nucl. Phys. B **349** (1991) 635.

[46] F. Hautmann *et al.*, Nucl. Phys. B949 (2019) 114795 [arXiv:1908.08524 [hep-ph]].

[47] T. Sjöstrand *et al.*, Comput. Phys. Commun. **191** (2015) 159 [arXiv:1410.3012 [hep-ph]].

[48] A. Bermudez Martinez, F. Hautmann and M. L. Mangano, Phys. Lett. B **822** (2021) 136700 [arXiv:2107.01224 [hep-ph]]; arXiv:2109.08173 [hep-ph].

[49] A. Bermudez Martinez, F. Hautmann and M. L. Mangano, JHEP **09** (2022) 060 [arXiv:2208.02276 [hep-ph]].

[50] M. Bury *et al.*, arXiv:2201.07114 [hep-ph].

[51] F. Hautmann, I. Scimemi and A. Vladimirov, Phys. Lett. B **806** (2020) 135478 [arXiv:2002.12810 [hep-ph]]; arXiv:2109.12051 [hep-ph].

[52] F. Hautmann *et al.*, Phys. Lett. B **833** (2022) 137276 [arXiv:2205.15873 [hep-ph]].

L.3 The small kT region in Drell-Yan production at next-to-leading order with the parton branching method

Eur. Phys. J. C (2024) 84:154
https://doi.org/10.1140/epjc/s10052-024-12507-0

THE EUROPEAN
PHYSICAL JOURNAL C

Regular Article - Theoretical Physics

The small k_{T} region in Drell–Yan production at next-to-leading order with the parton branching method

I. Bubanja[1,2,a], **A. Bermudez Martinez**[3], **L. Favart**[2,b], **F. Guzman**[4], **F. Hautmann**[3,5,6], **H. Jung**[7,8,c], **A. Lelek**[5], **M. Mendizabal**[7], **K. Moral Figueroa**[7], **L. Moureaux**[9,d], **N. Raicevic**[1,e], **M. Seidel**[10], **S. Taheri Monfared**[7,f]

[1] Faculty of Science and Mathematics, University of Montenegro, Podgorica, Montenegro
[2] Interuniversity Institute for High Energies (IIHE), Université libre de Bruxelles, Brussels, Belgium
[3] CERN, Geneva, Switzerland
[4] InSTEC, Universidad de La Habana, Havanna, Cuba
[5] Elementary Particle Physics, University of Antwerp, Antwerp, Belgium
[6] University of Oxford, Oxford, UK
[7] Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany
[8] II. Institut für Theoretische Physik, Universität Hamburg, Hamburg, Germany
[9] Institut für Experimentalphysik, Universität Hamburg, Hamburg, Germany
[10] Riga Technical University, Riga, Latvia

Received: 15 December 2023 / Accepted: 29 January 2024 / Published online: 15 February 2024

Abstract The Parton Branching (PB) method describes the evolution of transverse momentum dependent (TMD) parton distributions, covering all kinematic regions from small to large transverse momenta k_{T}. The small k_{T}-region is very sensitive both to the contribution of the intrinsic motion of partons (intrinsic k_{T}) and to the resummation of soft gluons taken into account by the PB TMD evolution equations. We study the role of soft-gluon emissions in TMD as well as integrated parton distributions. We perform a detailed investigation of the PB TMD methodology at next-to-leading order (NLO) in Drell–Yan (DY) production for low transverse momenta. We present the extraction of the nonperturbative "intrinsic-k_{T} " distribution from recent measurements of DY transverse momentum distributions at the LHC across a wide range in DY masses, including a detailed treatment of statistical, correlated and uncorrelated uncertainties. We comment on the (in)dependence of intrinsic transverse momentum on DY mass and center-of-mass energy, and on the comparison with other approaches.

1 Introduction

The measurement of the vector boson transverse momentum, p_{T}, in Drell–Yan (DY) production [1] allows one to investigate in detail many different aspects of the strong interaction sector of the Standard Model, and their impact on precision electroweak measurements. The very low p_T region of the DY cross section is sensitive to the contribution from the non-perturbative transverse motion of partons inside the hadrons; additionally at low transverse momentum multiple soft gluon emissions have to be resummed; at larger transverse momenta perturbative higher-order contributions become dominant. The precise description of the Z/γ boson transverse momentum distribution has been investigated since the 1980's, and approaches like CSS [2] analytic resummation and parton-shower [3–6] numerical algorithms have been applied with different success.

In this work we explore the approach [7,8] to DY p_{T} spectra based on the parton branching (PB) TMD methodology in momentum space proposed in [9,10], and we perform a detailed analysis of the small-p_{T} region for wide ranges in center of mass energies and in DY masses. Though fitted only on deep-inelastic scattering (DIS) data from HERA experiments, the PB-TMD methodology has been shown to be capable of describing DY p_{T} spectra at LHC energies [7] and at low energies [8] without any need for adjustment of parameters. This approach takes into account simultaneously

---

[a] e-mail: itana.bubanja@cern.ch
[b] e-mail: Laurent.Favart@ulb.be
[c] e-mail: hannes.jung@desy.de (corresponding author)
[d] e-mail: louis.moureaux@cern.ch
[e] e-mail: natasar@ucg.ac.me
[f] e-mail: sara.taheri.monfared@desy.de

soft gluon radiations and the transverse momentum recoils in the parton branchings along the QCD cascade. It provides a successful natural treatment of the multiple-scale problem of the DY transverse momentum for transverse momenta much smaller than DY masses but also of the DY with hard jet production [11]. It also confirms the universality of the TMDs being able to describe both DIS and DY cross sections at all available center of mass energies [12]. Alternative approaches based on parton showers in standard Monte Carlo event generators like PYTHIA8 [3] can also describe multi-differential DY cross section but it has been observed that they require intrinsic transverse momentum distributions strongly dependent on $\sqrt{s}$ [13,14]. In order to describe the measurements at LHC energies, a Gaussian width exceeding the Fermi motion kinematics is needed. Approaches based on CSS [15] provide very precise analytic predictions for inclusive enough observables like the Drell–Yan cross section transverse momentum. In this paper we study in detail the low k_T behavior of the PB-TMD parton distributions where both very soft gluon emission and intrinsic-k_T contribute significantly and interplay. The results presented here provide a multi-scale economical and coherent approach demonstrating the sensitivity to nonperturbative TMD contributions and first steps in disentangling the intrinsic-k_T contribution from the nonperturbative Sudakov one [16]. We compare DY theoretical predictions with experimental measurements in wide ranges in center-of-mass energies, $\sqrt{s}$ and in DY masses, m_{DY} , to extract the intrinsic-k_T parameter from the transverse momentum distributions. We are carefully taking into account systematic and statistical uncertainties using the breakdown of experimental uncertainties provided by the full set of covariance matrices available in the recent Drell–Yan differential cross section measurement at 13 TeV [17] and we treat for the first time the scale uncertainties in the theoretical predictions as correlated uncertainties within a given mass bin.

The results for TMD parameters such as intrinsic-k_T obtained from the DY analysis in this paper can be compared with analogous results obtained from TMD fits in the CSS coordinate-space framework, see e.g. the recent studies [18,19]. A significant difference between these approaches and the approach of this paper concerns the treatment of collinear parton distribution functions (PDFs). As shown in Refs. [7,8], in the approach of this paper the inclusive DGLAP limit is recovered and fits of collinear distributions are made, e.g. from inclusive DIS structure functions, along with TMD distributions [20–22]. In contrast, CSS approaches do not recover inclusive DGLAP and rather use an ansatz based on the operator product expansion of TMD distributions in terms of collinear PDFs, assuming collinear PDFs to be given by standard PDF sets. The PDF bias effect [19] which results from this has been shown to influence significantly the central values of the extracted distributions and dominate the systematic uncertainties in all the existing TMD determinations based on CSS approaches. The possibility to treat collinear and TMD distributions on the same footing and determine them without having to rely on existing PDF fits is a distinctive feature of the PB TMD approach. We believe that in the long run this could bring significant advantages in pursuing TMD phenomenology.

On the other hand, the results of this paper for intrinsic-k_T can also be compared with the case of parton shower Monte Carlo event generators, such as PYTHIA [3] and HERWIG [4]. Monte Carlo tuning to experimental data shows that parton shower approaches require intrinsic-k_T distributions dependent on the center-of-mass energy $\sqrt{s}$ [13,14], and a Gaussian width exceeding the Fermi motion kinematics. In contrast, in the approach of this paper we find that the width of the intrinsic-k_T distribution has a much milder center-of-mass energy $\sqrt{s}$ dependence. We obtain more natural Gaussian width σ, $\sigma = q_s/\sqrt{2}$, with q_s close to 1 GeV resulting from fits to DY measurements from fixed-target to LHC energies. We propose in this paper that the different behavior, concerning intrinsic-k_T distributions, between PB TMD and parton-shower approaches can be ascribed to the different treatment of the contributions to parton evolution from the nonperturbative Sudakov region, near the soft-gluon resolution boundary. See also [23] for a discussion of this and comparison of PB TMD and parton-shower results.

The paper is organized as follows. In Sect. 2 we briefly recall the basic elements of the calculational framework [7,9,10,21–23]: we start with the PB TMD approach; next we give a few comments on the treatment of the small transverse momentum region in this approach; then we discuss the Monte Carlo computation of DY differential distributions. Section 3 is the central section of the paper, in which we perform fits to DY data and present results for the intrinsic-k_T TMD parameter. We give conclusions in Sect. 4.

2 PB TMDs and DY production

To study the different contributions to the low-p_T spectrum, at different m_{DY} and different $\sqrt{s}$, we calculate DY production cross section in the PB TMD method, which proceeds as described in Refs. [7,23]. NLO hard-scattering matrix elements are obtained from the MADGRAPH5_AMC@NLO [24] next-to-leading (NLO) event generator and matched with TMD parton distributions and showers obtained from PB evolution [9,10,20], using the subtractive matching procedure proposed in [7] and further analyzed in [25].

We will show that the application of PB TMD distributions leads to a non negligible contribution of pure intrinsic-k_T, even if most of the small-k_T contribution comes from the PB-evolution. We also show that the proper treatment of photon radiation from the DY decay leptons is rather important, especially in the DY mass region below the Z boson peak. The contribution of intrinsic-$k_{\rm T}$ of heavy flavor partons is found to be negligible over the whole range since heavy quarks are not present in the initial configuration of the proton.

2.1 TMD distributions from the PB method

The PB evolution equations for TMD parton distributions $\mathcal{A}_a(x, \mathbf{k}, \mu^2)$ of flavor a are given by [9]

$$\begin{aligned}\mathcal{A}_a(x, \mathbf{k}, \mu^2) &= \Delta_a(\mu^2)\, \mathcal{A}_a(x, \mathbf{k}, \mu_0^2)\\ &+ \sum_b \int \frac{d^2\mathbf{q}'}{\pi \mathbf{q}'^2} \frac{\Delta_a(\mu^2)}{\Delta_a(\mathbf{q}'^2)}\, \Theta(\mu^2 - \mathbf{q}'^2)\, \Theta(\mathbf{q}'^2 - \mu_0^2)\\ &\times \int_x^{z_M} \frac{dz}{z}\, P_{ab}^{(R)}(\alpha_s, z)\, \mathcal{A}_b\left(\frac{x}{z}, \mathbf{k} + (1-z)\mathbf{q}', \mathbf{q}'^2\right),\end{aligned} \tag{1}$$

where $\mathbf{k}$ and $\mathbf{q}$ are 2-dimensional momentum vectors, z_M is the soft resolution scale [10], z is the longitudinal momentum transferred at the branching, $P_{ab}^{(R)}(\alpha_s, z)$ are the resolvable splitting functions[1] (whose explicit expressions for all flavor channels are given in [9]), and Δ_a are the Sudakov form factors

$$\begin{aligned}&\Delta_a(z_M, \mu^2, \mu_0^2)\\ &= \exp\left(-\sum_b \int_{\mu_0^2}^{\mu^2} \frac{d\mathbf{q}'^2}{\mathbf{q}'^2} \int_0^{z_M} dz\, z\, P_{ab}^{(R)}(\alpha_s, z)\right).\end{aligned} \tag{2}$$

The branching evolution (1) fulfills soft-gluon angular ordering [27–29], with the branching variable $\mathbf{q}'^2$ being related to the transverse momentum $q_{\rm T}$ of the parton emitted at the branching by

$$q_{\rm T} = (1 - z)\, |\mathbf{q}'|. \tag{3}$$

It is shown in [10] that angular ordering is essential for the TMD distribution arising from the solution of Eq. (1) to be well-defined and independent of the choice of the soft-gluon resolution scale $z_M = 1 - \varepsilon$ for $\varepsilon \to 0$. In contrast, p_T ordering leads, for instance, to ambiguities in the definition of the TMD from the $z \to 1$ region.

Analogously to the case of ordinary (collinear) parton distribution functions, the distribution $\mathcal{A}_a(x, \mathbf{k}, \mu_0^2)$ at the starting scale μ_0 of the evolution, in the first term on the right hand side of Eq. (1), is a nonperturbative boundary condition to the evolution equation, and is to be determined from experimental data. For simplicity we parameterize $\mathcal{A}_a(x, \mathbf{k}, \mu_0^2)$ in the form

$$\mathcal{A}_{0,a}(x, \mathbf{k}, \mu_0^2) = f_{0,a}(x, \mu_0^2) \cdot \exp\left(-|\mathbf{k}|^2/2\sigma^2\right)/(2\pi\sigma^2), \tag{4}$$

with the width of the Gaussian distribution given by $\sigma = q_s/\sqrt{2}$, independent of parton flavor and x, where q_s is the intrinsic-$k_{\rm T}$ parameter.

The scale at which the strong coupling α_s is to be evaluated in Eqs. (1) and (2) is a function of the branching variables. Two scenarios are studied in Refs. [9,20]:

$$\begin{aligned}&\text{i) :} \quad \alpha_s = \alpha_s(\mathbf{q}'^2)\\ &\text{ii) :} \quad \alpha_s = \alpha_s(\mathbf{q}'^2(1-z)^2) = \alpha_s(q_{\rm T}^2)\end{aligned} \tag{5}$$

In scenario i), it is shown in [9] that Eq. (1), in the collinear case, i.e. once it is integrated over all transverse momenta, reproduces exactly the DGLAP evolution [30–33] of parton densities. In scenario ii), it is discussed in [34] how, upon integration over transverse momenta and suitable treatment of the resolution scale, Eq. (1) returns the CMW coherent branching evolution [29].

In Ref. [20], fits to precision DIS HERA measurements [35] based on Eqs. (1) and (4), combined with NLO DIS matrix elements, are performed for both scenarios i) and ii), using the fitting platform `xFitter` [36,37]. It is found that fits to DIS measurements with good χ^2 values can be achieved in either case. Correspondingly, PB-NLO-HERAI+II-2018 set 1 (abbreviated as PB-NLO-2018 Set1) (with the DGLAP-type $\alpha_s(\mathbf{q}'^2)$) and PB-NLO-HERAI+II-2018 set2 (abbreviated as PB-NLO-2018 Set2) (with the angular-ordered CMW-type $\alpha_s(q_{\rm T}^2)$) are obtained, both having intrinsic-$k_{\rm T}$ parameter in Eq. (4) set to $q_s = 0.5$ GeV [20]. All PB TMD parton distributions (and many others) are accessible in TMDlib and via the graphical interface TMDplotter [38,39].

On the other hand, it is found that PB-NLO-2018 Set2 provides a much better description, compared to PB-NLO-2018 Set1, of measured Z/γ transverse momentum spectra at the LHC [7], in low-energy experiments [23], and of dijet azimuthal correlations near the back-to-back region at the LHC [40]. This underlines the relevance of the angular-ordered coupling $\alpha_s(q_{\rm T}^2)$ in regions dominated by soft-gluon emissions.

Based on this observation, in the following we will focus on the PB-NLO-2018 Set2 approach and perform fits to DY transverse momentum measurements to investigate the sensitivity of these measurements to the nonperturbative TMD intrinsic-$k_{\rm T}$ parameter q_s, and perform determinations of its value.

As discussed in [7,20], in order to complete the definition of the PB-NLO-2018 Set2 scenario the treatment of the coupling α_s needs to be specified in the region of small transverse momenta $q_{\rm T} \lesssim q_0$, where q_0 is a semi-hard scale on the order of a GeV. As in [7,20], we take

$$\alpha_s = \alpha_s(\max(q_0^2, q_{\rm T}^2)), \tag{6}$$

---

[1] Using transverse momentum dependent splitting functions as described in Ref. [26] would require using off-shell matrix elements and a completely new fit to inclusive structure functions.

setting $q_0 = 1$ GeV, which may be regarded as similar in spirit to the "pre-confinement" proposal in the context of infrared-sensitive QCD processes [41,42].[2] In the present study, we will perform a determination of the nonperturbative TMD parameter q_s from DY transverse spectra by assuming the above behavior for α_s.

To better illustrate the underlying physical picture, we give next a few further comments on nonperturbative contributions and the treatment of the small transverse momentum region in the PB TMD approach.

As implied by Eqs. (1) and (2), the PB TMD method incorporates Sudakov evolution via phase space integrations of appropriate kernels over the resolvable region, i.e. over momentum transfers z up to the soft-gluon resolution scale z_M. For each branching evolution scale $\mathbf{q}'^2$, it is instructive to examine separately parton emissions with transverse momenta above the semi-hard scale q_0, $q_T > q_0$, and below q_0, $\Lambda_{QCD} < q_T \lesssim q_0$. Using the angular ordering relation (3), these emissions are mapped respectively on the regions

$$\begin{aligned} (a):&\quad z < z_{dyn} = 1 - q_0/|\mathbf{q}'|, \\ (b):&\quad z_{dyn} \lesssim z < z_M, \end{aligned} \tag{7}$$

where $z_{dyn} = 1 - q_0/|\mathbf{q}'|$ is the dynamical resolution scale associated with the angular ordering [27–29,34]. In region (a), the strong coupling (6) is evaluated at the scale of the emitted transverse momentum, $\alpha_s(q_T^2)$; the contribution from region (a) to the evolution in Eqs. (1), (2) corresponds to the perturbative Sudakov resummation (see e.g. [43,44]). In region (b), the strong coupling (6) freezes around the semi-hard scale q_0; the contribution from region (b) to the evolution is the nonperturbative Sudakov form factor in the PB TMD approach.

It is worth noting that the PB-NLO-2018 Set 2 framework provides a very natural and economical description of nonperturbative Sudakov effects, based on perturbative modeling of the Sudakov form factor (2) combined with the infrared α_s behavior (6): it does not contain any additional nonperturbative functions and parameters, besides the scale q_0.

In Fig. 1 we show parton distributions obtained with the PB approach using the starting distributions from PB-NLO-2018 Set2. We show distributions for the gluon and down quark parton densities for different values of z_M: $z_M \rightarrow 1$ (default - regions $(a + b)$ - red curve [20]) and $z_M = z_{dyn} = 1-q_0/\mathbf{q}'$ (region (a) only - blue curve obtained with the same parameters as PB-NLO-2018 Set2 except z_M using UPDFEVOLV [45]). The distributions obtained from PB-NLO-2018 set2 with $z_M \rightarrow 1$ are significantly different from those applying $z_M = z_{dyn}$, illustrating the importance of soft contributions even for collinear distributions. In Ref. [46] it was found that limiting the z-integration leads to inconsistencies. In Ref. [47] a procedure to correct the z limitation is discussed. A detailed discussion on the role of soft gluons and the nonperturbative Sudakov form factor is given in Ref. [48]. Please note that the intrinsic-k_T distribution, since not part of the collinear calculation, does not affect the collinear parton densities.

In the transverse momentum distributions obtained with the PB-approach, the effect of the z_M cut-off is even more visible. In Fig. 2 the transverse momentum distributions obtained for down and charm quarks are shown for PB-NLO-2018 Set2, with $z_M \rightarrow 1$, i.e. regions $(a + b)$, with (red curve) and without intrinsic-k_T distribution applied (blue curve – a Gauss distribution with $q_s = 0.00001$ GeV). We also show the transverse momentum distribution contribution from region (a) alone, i.e. for $z_M = z_{dyn} = 1 - q_0/\mu'$, without intrinsic-$k_T$ (corresponding to the magenta curve of Fig. 1). The importance of the large z-region on the transverse momentum distributions is seen in the comparison with the predictions without intrinsic-k_T distribution (blue and magenta curves).

The transverse momentum distributions show very clearly the large effect of the choice of z_M for the soft region, while in the perturbative region $k_T > q_0$ the effect becomes smaller with increasing k_T. Applying such a scale, $z_M = z_{dyn} = 1 - q_0/\mu'$, removes emissions with $q_T < q_0$ (there are still low-k_T contributions, which come from adding vectorially all intermediate emissions). However, very soft emissions are automatically included with $z_M \rightarrow 1$.

As shown in Fig. 2, the effect of the intrinsic-k_T distribution is much reduced at large scales, but the contribution of the region $z_{dyn} < z < 1$ stays important for small k_T.

It is interesting to observe that the charm density shows essentially no effect of an intrinsic-k_T distribution: this is because charm is generated dynamically from gluons only, and there is no intrinsic charm density.

2.2 Transverse momentum distributions of PB-NLO-2018

After having discussed the importance of the soft nonperturbative region to the transverse momentum distribution, we turn now to a discussion of the transverse component of the PB parton distributions of Ref. [20], which are used for comparison with measurements.

In previous investigations on Z-boson production at the LHC [7], as well as for low DY mass, m_{DY}, and at low $\sqrt{s}$ [8], it was found that PB-NLO-2018 Set2 describes the measurements much better, while PB-NLO-2018 Set 1 gives too large a cross section at small DY lepton pair transverse momenta, $p_T(\ell\ell)$.

The difference between PB-NLO-2018 Set1 and Set2, which comes from the choice of renormalization scale (argument in α_s), is seen essentially in the low-k_T region, where the

---

[2] Different forms of the extension to small q_T could be considered. However, this will entail new fits both to precision DIS data and to DY data.

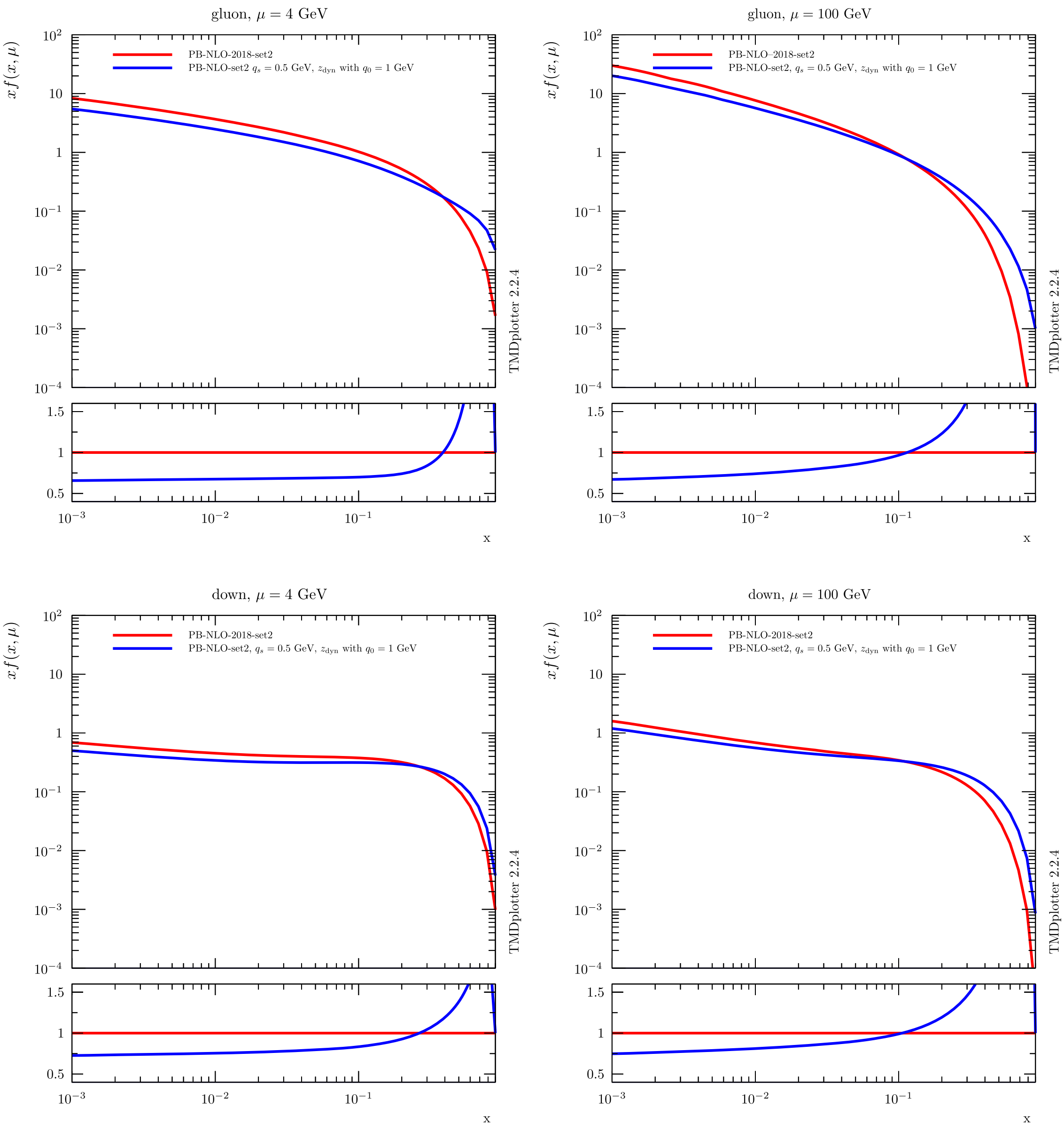


Fig. 1 Integrated gluon and down-quark distributions at $\mu = 4$ GeV (left column) and $\mu = 100$ GeV (right column) obtained from the PB approach based on PB-NLO-2018 Set2. The red curve is the published PB-NLO-2018 Set2 [20] and corresponds to $z_M \to 1$ (regions $(a+b)$ in text). The blue curve corresponds to $z_M = z_{\rm dyn}$ with $q_0 = 1$ GeV (region (a) only). The ratio plots show the ratios to the one for $z_M \to 1$

nonperturbative Sudakov form factor (region (b)), with the integral $z_M \to 1$, plays an important role. In Fig. 3 (upper row) the distributions for up and charm quarks are shown when no intrinsic-$k_{\rm T}$ distribution is included, the lower row shows distributions including the default intrinsic Gauss $k_{\rm T}$ distributions of widths $q_s = 0.5$ GeV. It is very interesting to observe that the differences between the sets setting $q_s = 0$ or not are very much reduced for heavy flavors since they are only generated dynamically (since heavy flavors are not present at the starting scale in the VFNS which is applied here). In principle an intrinsic charm contribution can be included in PB densities, however, this is not required from inclusive DIS data [35] used in the fit of PB-NLO-2018 Set2.

In Fig. 4 the transverse momentum distribution for down quarks, with and without an intrinsic-$k_{\rm T}$ distribution, is shown at different scales μ. While at low scales $\mu \sim 50$ GeV a significant effect of the intrinsic-$k_{\rm T}$ distribution is observed for very small $k_{\rm T}$, at large scales $\mu \sim 350$ GeV this effect is much reduced. This scale dependence will result in a much smaller sensitivity to the intrinsic-$k_{\rm T}$ distribution at high $m_{\rm DY}$.

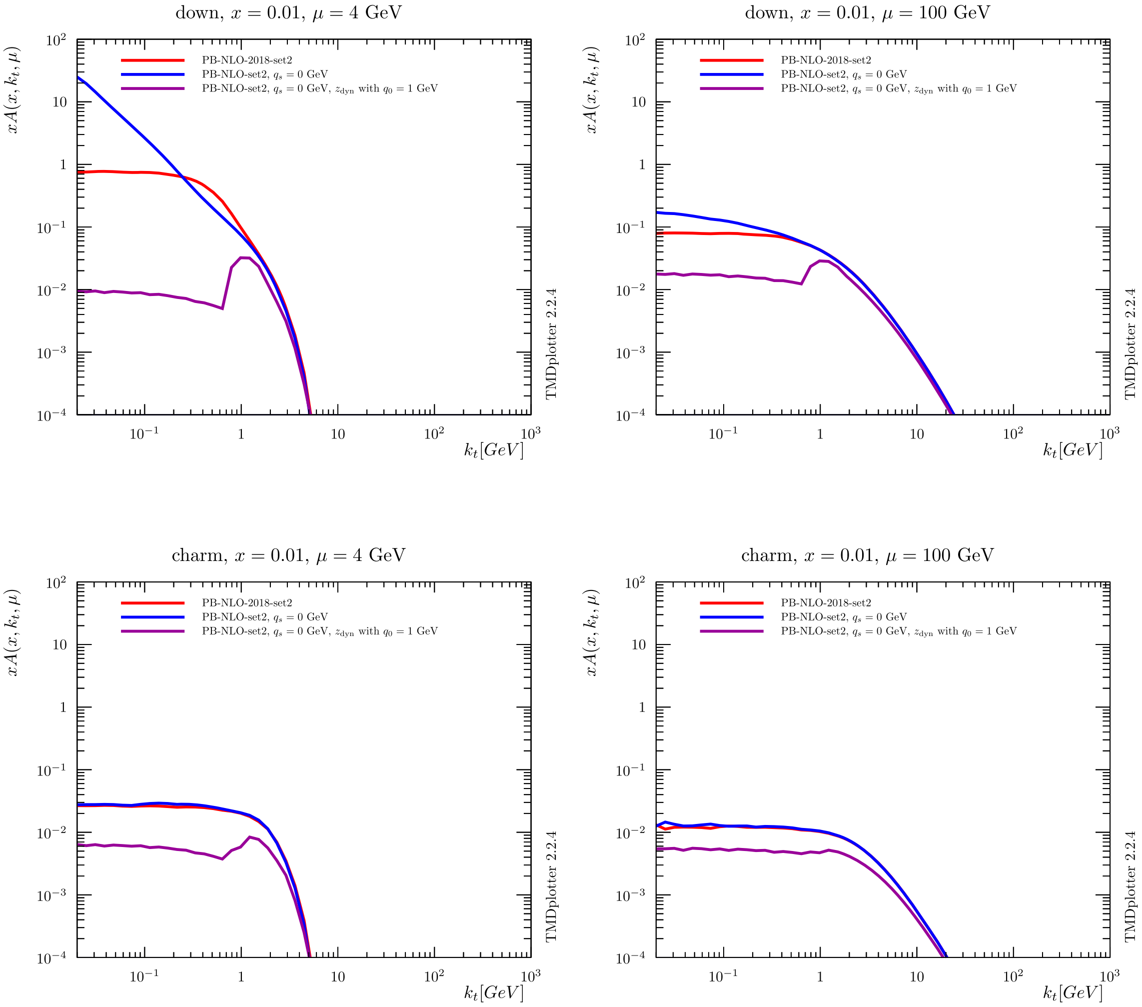


Fig. 2 Transverse momentum distributions of down and charm quarks at $\mu = 4$ GeV (left column) and $\mu = 100$ GeV (right column) obtained from the PB approach based on PB-NLO-2018 Set2. Two distributions do not include intrinsic-k_T: the blue curve corresponds to $z_M \to 1$ (regions a+b in text) and the magenta curve to $z_M = z_{\rm dyn} = 1 - q_0/q$ (region a only). The red curve is the published one PB-NLO-2018 Set2 [20] and including intrinsic-k_T and $z_M \to 1$

2.3 Calculation of the DY cross section

The cross section of DY production is calculated at NLO with MadGraph5_aMC@NLO[24]. In the MCatNLO method, the collinear and soft contributions of the NLO cross section are subtracted, as they will be later included when parton shower, or as in our case, TMD parton densities are applied. As in earlier studies, we use Cascade3 [49] to include TMD parton distributions and parton shower to the MCatNLO calculation (a detailed investigation of the effect of TMD parton distributions and parton showers applied in the Cascade3 Monte Carlo generator is given in Ref. [25]). We use the Herwig6 subtraction terms in MCatNLO, since they are based on the same angular ordering conditions as the PBTMD parton distribution sets, PB-NLO-2018 Set1 and PB-NLO-2018 Set2, described in the previous section. The validity and consistency using Herwig6 subtraction terms in MCatNLO together with PB TMD distributions has been studied in detail in the appendix of Ref. [25]. The predicted cross sections (labeled as MCatNLO+CAS3 in the following) are calculated using the integrated versions of the NLO parton densities PB-NLO-2018 Set1 and PB-NLO-2018 Set2 together with $\alpha_s(m_Z) = 0.118$ at NLO.

The factorization scale μ, used in the calculation of the hard process is set to $\mu = \frac{1}{2}\sum_i \sqrt{m_i^2 + p_{t,i}^2}$, with the sum running over all final state particles, in case of DY production over all decay leptons and the final jet. For the generation of transverse momentum according to the PB-TMD distributions, the factorisation scale μ in the hard process is set to $\mu = m_{\rm DY}$, in the case of a real emission it is set to $\mu = \frac{1}{2}\sum_i \sqrt{m_i^2 + p_{t,i}^2}$. The generated transverse momentum is limited by the matching scale μ_m = `SCALUP` [49]. Since there are no PB-fragmentation functions available yet,

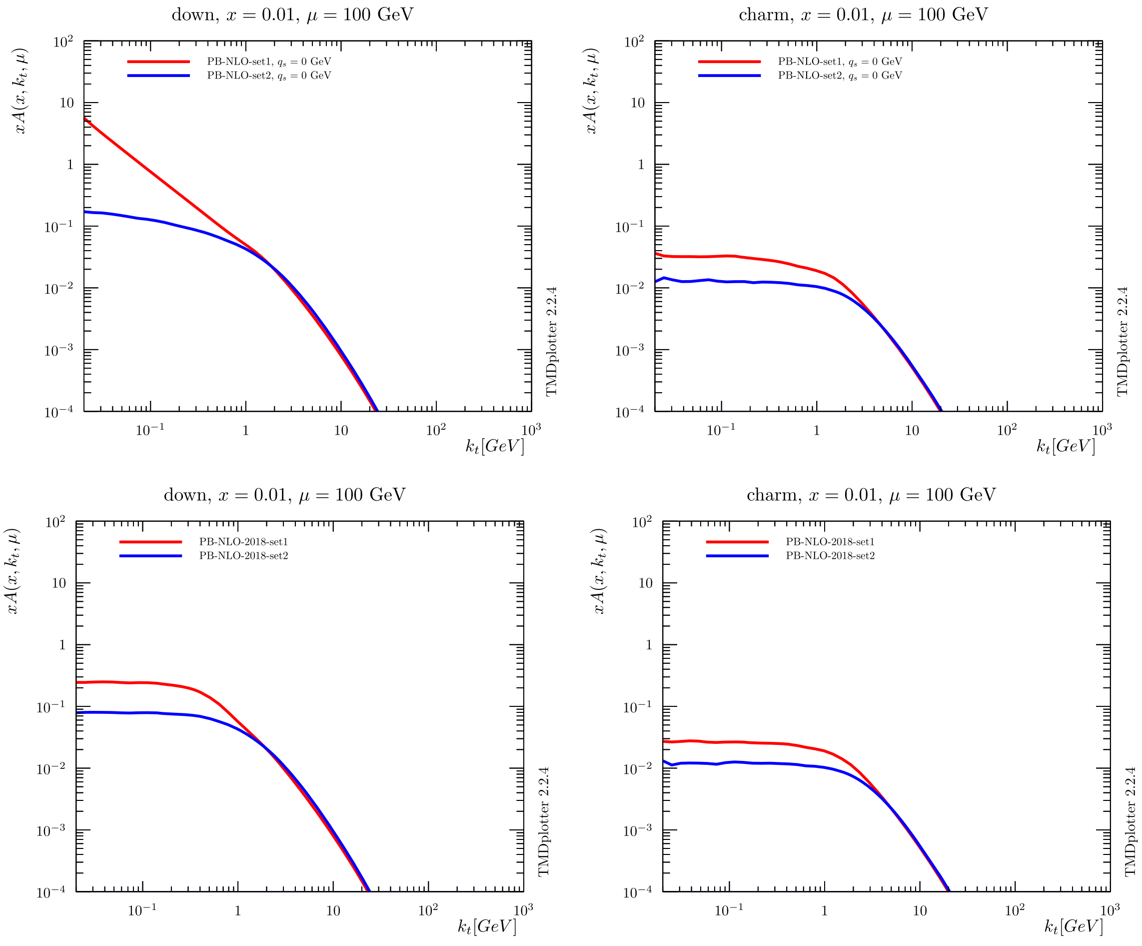


Fig. 3 TMD parton density distributions for down and charm quarks of the published PB-NLO-2018 Set1 (red curve) and PB-NLO-2018 Set2 (blue curve) [20] as a function of k_T at $\mu = 100$ GeV and $x = 0.01$. In the upper row are shown distributions when no intrinsic-k_T distribution is included ($q_s = 0.00001$ GeV), and the lower row shows the default distributions with $q_s = 0.5$ GeV

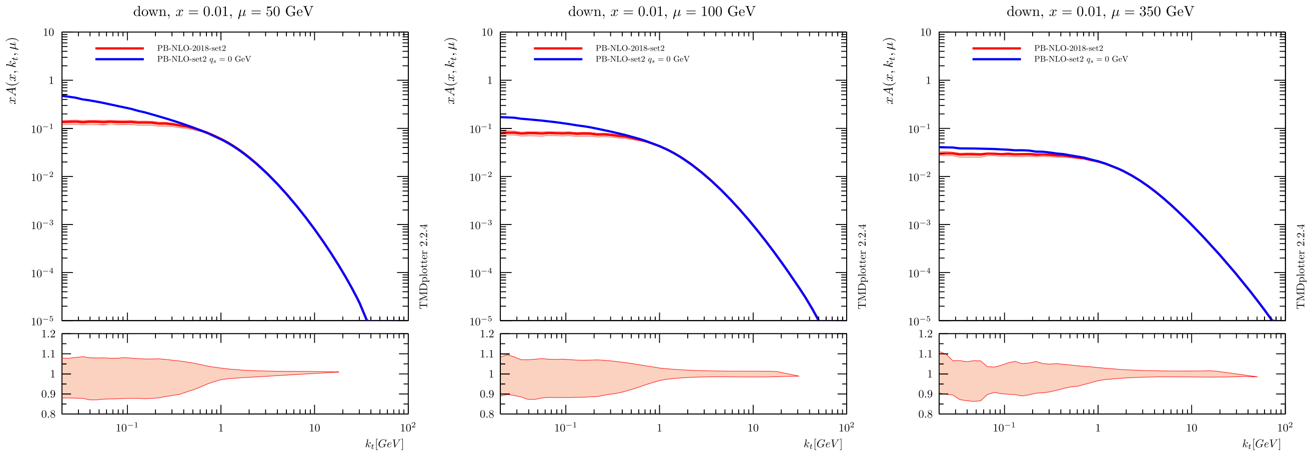


Fig. 4 TMD parton density distributions for down quarks of PB-NLO-2018 Set2 with (red curve) and without (blue curve) intrinsic-k_T distribution as a function of k_T at different scales μ and $x = 0.01$. The lower panels show the full uncertainty of the TMD PDFs, as obtained from the fits [20]. Shown is the ratio to each central value. The red band shows the uncertainty of PB-NLO-2018 Set2, the blue line has no uncertainty band

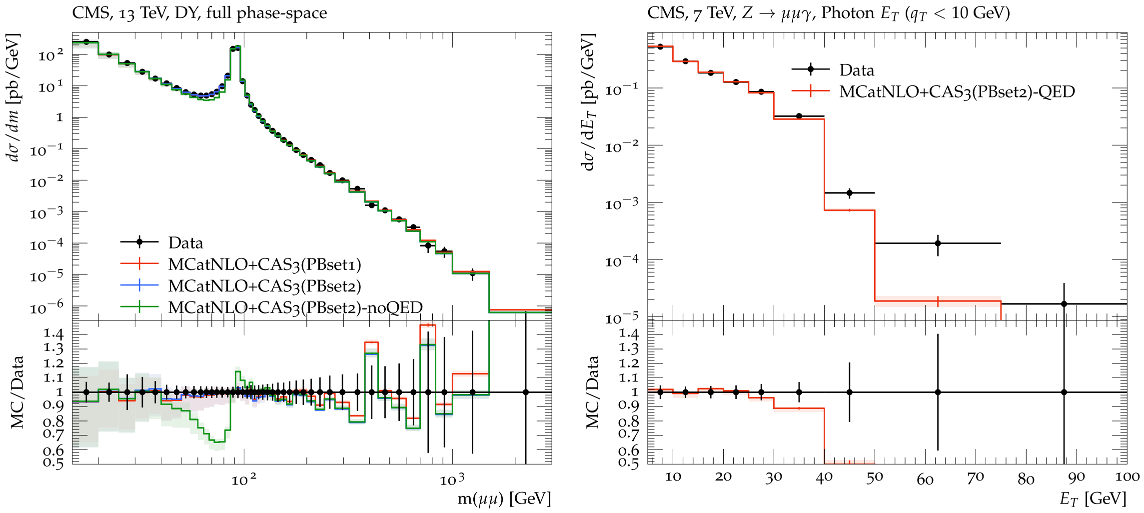


Fig. 5 Left: The mass distribution of DY lepton pairs at 13 TeV [51] compared to predictions of MCatNLO+CAS3 with PB-NLO-2018 Set1 (red curve), PB-NLO-2018 Set2 (blue curve) and without QED corrections (green curve). Right: The spectrum of photons transverse momentum in $Z \to \mu^+\mu^-\gamma$ at 7 TeV [55] compared to MCatNLO+CAS3 PB-NLO-2018 Set2 including QED radiation for a transverse momentum of the DY pair $p_T(\ell\ell) < 10$ GeV. The bands show the scale uncertainty

the final state parton shower in CASCADE3 is generated from PYTHIA[50], including photon radiation of the lepton pair.

A good description of the final state QED corrections, and in particular the kinematic effect of the real photon radiations, is essential in order to achieve a precise description of the DY transverse momentum. Figure 5 (left) shows the DY mass distribution as measured by CMS [51] at 13 TeV together with predictions of MCatNLO+CAS3.[3] The bands show the scale uncertainty coming from a variation of the renormalization and factorization scale by a factor of two up and down, avoiding the extreme values (7-point variation). The DY mass is calculated from the so-called *dressed-leptons* (see for example [53,54]), where photons radiated within a cone of radius of $R < 0.1$ are merged to the lepton before the momenta are calculated. We show predictions based on PB-NLO-2018 Set1 and Set2, and also, for illustration, when photon radiation is turned off in the final-state shower (labeled as "noQED"). A rather good description of the DY mass spectrum over a large range on m_{DY} is obtained both with PB-NLO-2018 Set1 and Set2. Only at m_{DY} greater than a few hundred GeV the predictions tend to become smaller than the measurement (while still within the uncertainties). However, this is the region where the partonic x becomes large and not well constrained by the fit to HERA data [35] used for the PB-NLO-2018 TMD extraction [20]. In the region of m_{DY} below the Z-pole, one can observe the importance of QED corrections. In Fig. 5 (right) we show the photon transverse momentum spectrum in Z-production as measured by CMS [55] at 7 TeV in comparison with MCatNLO+CAS3 including QED radiation. The photon spectrum is well described at low $E_T < 40$ GeV, while the high E_T spectrum predicted by the parton shower falls below the measurement, since the precision of parton showers are limited for the high p_T region.

3 The transverse momentum spectrum of DY lepton pairs

The transverse momentum spectrum of DY pairs at $\sqrt{s} =$ 13 TeV has been measured for a wide m_{DY} range by CMS [17]. We use this measurement for comparison with predictions of MCatNLO+CAS3 based on PB-NLO-2018 Set1 and PB-NLO-2018 Set2, as shown in Fig. 6. As already observed in previous investigations [7,23,25,40], the PB-NLO-2018 Set1 gives too high a contribution at small transverse momenta $p_T(\ell\ell)$, while PB-NLO-2018 Set2 describes the measurements rather well, without any further adjustment of parameters,[4] underlining the role of evaluating the strong coupling at the transverse momentum scale. In order to illustrate the importance of QED corrections, we show in addition a prediction based on PB-NLO-2018 Set2 without including QED final state radiation (labeled noQED). Especially in the low m_{DY} region, the inclusion of QED radiation is essential, not only changing the total cross section but rather strongly modifying the shape of the transverse momentum distribu-

---

[3] We use the Rivet package [52] for the calculation of the final distributions.

[4] The predictions shown here are slightly different compared to the predictions in [17] because we use here a lower minimum k_T cut and because of a bug in the treatment of QED radiation in Rivet, corrected in version 3.1.8.

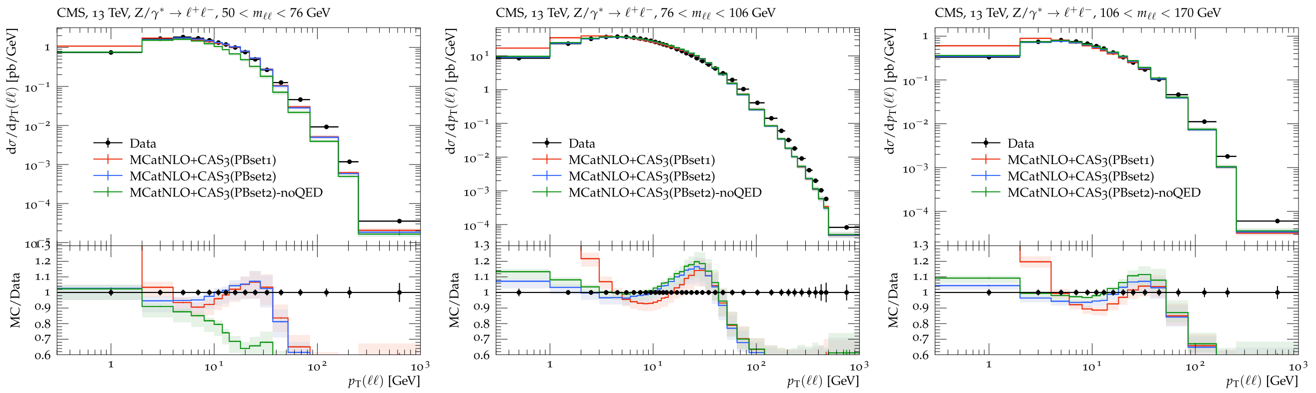


Fig. 6 The $p_{\mathrm{T}}(\ell\ell)$ dependent DY cross section for different m_{DY} regions as measured by CMS [17] compared to MCatNLO+CAS3 predictions based on PB-NLO-2018 Set 1 (red curve) and Set 2 (blue curve). Also shown are predictions without the inclusion of final state QED radiation from the leptons (green curve). The band shows the 7-point variation of the renormalization and factorization scales

tion $p_{\mathrm{T}}(\ell\ell)$. All calculations predict too low a cross section at large transverse momentum due to missing higher-order contributions in the matrix element. In Refs. [11,56,57] it is shown explicitly that including higher orders in the matrix element through the TMD multi-jet merging technique gives an excellent description even for largest $p_{\mathrm{T}}(\ell\ell)$. For all further distributions, we restrict the investigations to $p_{\mathrm{T}}(\ell\ell)$ below the peak region (i.e. $p_{\mathrm{T}}(\ell\ell) \lesssim 8$ GeV).

3.1 Influence of the intrinsic-k_{T} distribution on DY transverse momentum distributions

Given the rather successful description of the DY $p_{\mathrm{T}}(\ell\ell)$-spectrum with MCatNLO+CAS3 using PB-NLO-2018 Set 2 in the low $p_{\mathrm{T}}(\ell\ell)$-region, we investigate below the importance of the intrinsic-k_{T} distribution. In PB-NLO-2018 the intrinsic-k_{T} distribution is parameterized as a Gauss distribution with zero mean and a width $\sigma^2 = q_s^2/2$ [20] (see Eq. (4)), where q_s was fixed by default at $q_s = 0.5$ GeV.

In order to illustrate the sensitivity range of the intrinsic-k_{T} distribution, we show in Fig. 7 the MCatNLO+CAS3 predictions for the low $p_{\mathrm{T}}(\ell\ell)$-spectrum of DY production at different DY masses m_{DY} for different intrinsic-k_{T} distribution (with different q_s parameter values) compared to the CMS measurement [17]. We observe that sensitivity to intrinsic-k_{T} is more pronounced at small $p_{\mathrm{T}}(\ell\ell)$ values. This sensitivity decreases with increasing mass, as expected from Fig. 4

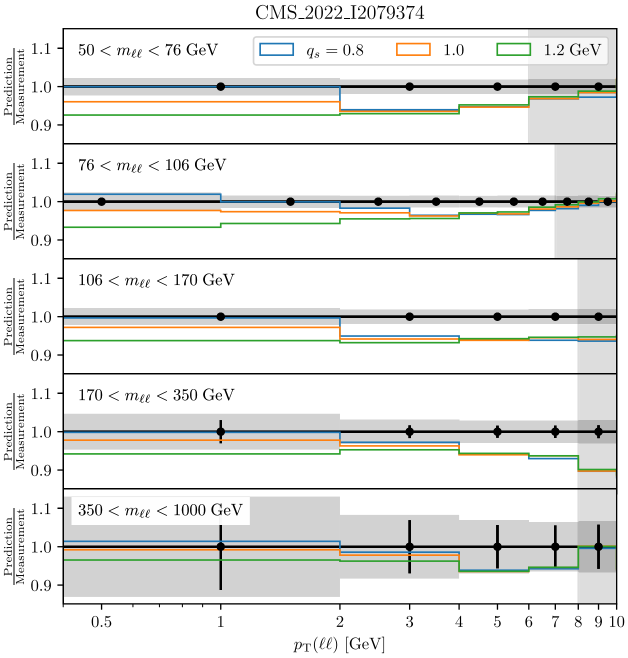


Fig. 7 Drell–Yan cross section ratios of MCatNLO+CAS3 predictions for different q_s values over CMS measurement [17] as a function of $p_{\mathrm{T}}(\ell\ell)$ for different m_{DY} regions. Only the lowest $p_{\mathrm{T}}(\ell\ell)$ values are shown. The points error bar show the statistical uncertainties and the gray bands the total experimental uncertainties. The gray area at the highest $p_{\mathrm{T}}(\ell\ell)$ values show the maximal values included in the fit described in Sect. 3.2

In the following we describe a determination of the Gaussian width q_s for different DY masses, m_{DY}, at different $\sqrt{s}$. The prediction is obtained from a calculation of MCatNLO+CAS3 using TMD distributions obtained with the PB-NLO-2018 Set2 parameters for the collinear distribution, but with different q_s values. We scan for each m_{DY}-bin q_s in steps of 0.1 to 0.3 GeV in the range $q_s = 0.1, \ldots, 2.0$ GeV. At higher DY transverse momenta, higher order contributions have to be taken into account (a study using multijet merging is given in Refs. [11,56,57]).

3.2 Fit of the Gauss width q_s in pp at $\sqrt{s} = 13$ TeV

The transverse momentum distribution of DY leptons has been measured by the CMS collaboration [17]. This is the basic measurement for the determination of the intrinsic-k_{T}

parameter q_s, since it covers a wide m_{DY}-range with high precision and that a detailed uncertainty breakdown, discussed in Sect. 3.2.1, is provided. The measurements of Z-production obtained from LHCb [58] are discussed in Sect. 3.2.2, while measurements at lower center-of-mass energies are shown in Sect. 3.3.

3.2.1 DY production over a wide DY mass range

The CMS collaboration has measured Drell–Yan production at 13 TeV [17] covering a range of DY mass $m_{\mathrm{DY}} = [50, 76, 106, 170, 350, 1000]$ GeV. The measurement is provided with a detailed uncertainty breakdown, corresponding to a complete treatment of experimental uncertainties including correlations between bins of the measurement for each uncertainty source separately. Note that we use the fully detailed breakdown of the experimental uncertainties provided on the CMS public website.[5]

In order to determine the intrinsic-k_{T} we vary the q_s parameter and calculate a χ^2 to quantify the model agreement with the measurement. We evaluate the following expression,[6]

$$\chi^2 = \sum_{i,k} (m_i - \mu_i) C_{ik}^{-1} (m_k - \mu_k), \tag{8}$$

with m_i being the measurement and μ_i being the prediction for data point i. The covariance matrix C_{ik} is decomposed into a component describing the uncertainty in the measurement, $C_{ik}^{\text{meas.}}$, and the statistical and scale uncertainties in the prediction,

$$C_{ik} = C_{ik}^{\text{meas.}} + C_{ik}^{\text{model-stat.}} + C_{ik}^{\text{scale}}. \tag{9}$$

The covariance matrix of the measurement is taken directly from the supplementary material provided by CMS. The statistical uncertainty in the prediction, arising from the use of a Monte Carlo simulation, is accounted for as a small diagonal contribution without correlations between bins,

$$C_{ik}^{\text{model-stat.}} = \sigma_{i,\text{stat.}}^2 \, \delta_{ik}, \tag{10}$$

where $\sigma_{i,\text{stat.}}$ is the bin-by-bin statistical uncertainty. We also treat, for the first time, the scale uncertainties of the theoretical prediction as a correlated uncertainty, for a given m_{DY} range, allowing for a global shift of all bins together within the band defined by the symmetrized envelope of the scale uncertainties. This contribution to the covariance matrix is constructed as follows,

$$C_{ik}^{\text{scale}} = \sigma_{i,\text{scale}} \, \sigma_{k,\text{scale}}, \tag{11}$$

where $\sigma_{i,\text{scale}}$ encodes the scale variation for each bin.

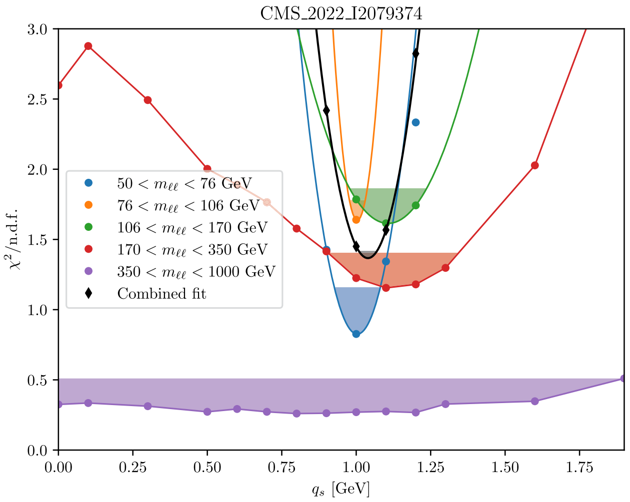


Fig. 8 The reduced χ^2/n.d.f distribution as a function of q_s for different m_{DY} regions obtained from a comparison of the MCatNLO+CAS3 prediction with the measurement by CMS [17]. The points represent the obtained χ^2 values. The lines represent the curves used for the uncertainty estimate (see text), which is materialized by the shaded areas

We first extract independent values of q_s for each invariant mass region considered in the measurement, considering only the region most sensitive to q_s, $p_{\mathrm{T}}(\ell\ell) < 8$ GeV. We reduce this range further in the first two regions to $p_{\mathrm{T}}(\ell\ell) < 6$ GeV for $50 < m_{\mathrm{DY}} < 76$ GeV and $p_{\mathrm{T}}(\ell\ell) < 7$ GeV for $76 < m_{\mathrm{DY}} < 106$ GeV to stay in the region of sensitivity and not be biased by missing higher orders in the predictions affecting high $p_{\mathrm{T}}(\ell\ell)$ shape. The obtained χ^2/n.d.f (reduced χ^2)) values are shown in Fig. 8 as a function of q_s.

Within each region, we consider the value of q_s for which the smallest χ^2 is obtained as our "best fit" value. We construct a one-sigma confidence region as the set of all q_s values for which $\chi^2(q_s) < \min(\chi^2) + 1$. When possible, this region is determined graphically using a linear interpolation between scan points. When the minimum is too narrow for a reliable determination of the uncertainty using this method, we use instead a quadratic interpolation between the lowest three points and add an uncertainty equal to one half-bin-width (0.05 GeV) in quadrature. In addition, we include an uncertainty derived by repeating the procedure with modified fit boundaries. The values obtained using this method are listed in Table 1 and a comparison is shown in Fig. 9.

The values derived from each m_{DY} interval are compatible with each other. The most precise determination is obtained from the Z peak region, $76 < m_{\mathrm{DY}} < 106$ GeV, followed by the regions around it. The sensitivity at high mass suffers from larger statistical uncertainties in the measurement. This independence of the intrinsic-k_{T} with the DY mass contrasts with the need to tune the Parton Shower parameters for different masses in standard Monte Carlo events

---

[5] The corresponding HEPdata records only contain summarised information.

[6] The code used with the full covariance matrix is available in Ref. [59], an earlier version to be used directly with Rivet is in Ref. [60].

generators (see [17] - Fig. 6 for a data comparison with MADGRAPH5_AMC@NLO interfaced with PYTHIA Parton Shower).

Having obtained compatible results, we proceed to deriving a combined fit by calculating a joint χ^2 including the considered bins in all mass ranges. For this, we construct a new covariance matrix $C_{ik}^{\text{comb.}}$ as a sum over the 650 uncertainty sources included in the detailed breakdown. We consider that each systematic uncertainty is fully correlated between m_{DY} regions and construct their covariance matrices in the same way as in Eq. (11). The statistical uncertainties (data and Monte Carlo) in the measurement feature nontrivial correlations due to the use of unfolding but are independent in each m_{DY} region, and therefore we construct a block-diagonal matrix from the covariance matrices in each m_{DY} region. The statistical uncertainty in the prediction is diagonal. We consider that the uncertainties in the QCD scales are not correlated between m_{DY} regions and use a block-diagonal matrix.

The χ^2 values obtained using the combined covariance matrix are shown in Fig. 8. The best fit value, extracted in the same way as for separate regions, is,

$$q_s = 1.04 \pm 0.03(\text{data}) \pm 0.05(\text{scan}) \pm 0.05(\text{binning})\ \text{GeV}\,.$$

This value and its uncertainty are shown as a black line and shaded area on Fig. 9 for comparison with the individual m_{DY} bins. A cross-check has been performed by interpolating the prediction for each bin between q_s values and searching for the minimum of the χ^2 distribution using a finer q_s grid. It returned values within the uncertainties quoted above. The TMD distributions including the new q_s value are available in TMDlib and TMDplotter [38,39].

To make consistency checks of the obtained value of q_s and to examine possible trends of its dependence on DY mass and centre-of mass energy, the DY measurements at high rapidity and lower collision energies have been analysed. Since for these measurements no full error breakdown are available, we treat all uncertainties as being uncorrelated and do not include systematic uncertainty coming from the scale variation in the theoretical calculation.

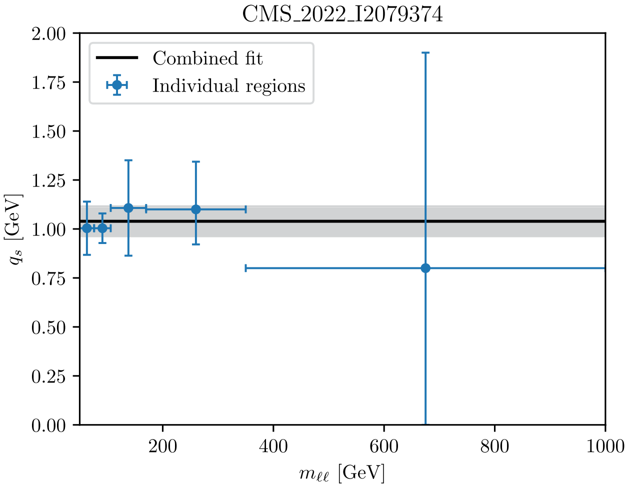


Fig. 9 The values of q_s in each m_{DY} -bin as obtained from Ref. [17]. Indicated is also the combined fit value of q_s

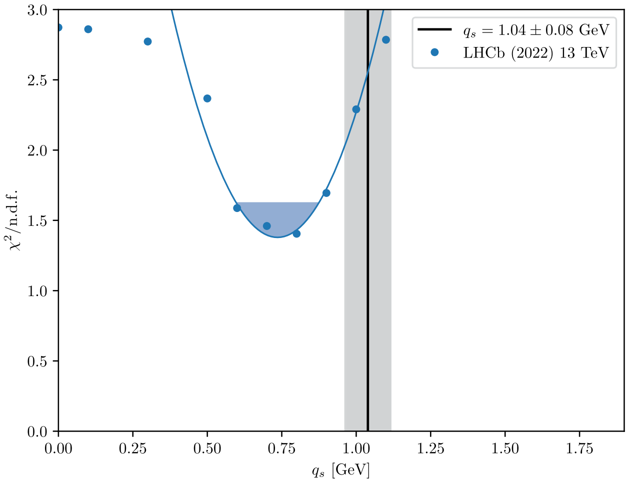


Fig. 10 The reduced χ^2/n.d.f distribution as a function of q_s summed over all rapidity regions obtained from a comparison of the MCatNLO+CAS3 prediction with the measurement by LHCb [58]. The shaded area corresponds to $\chi^2 + 1$. The best fit value is $q_s = 0.74 \pm 0.15$ GeV. The value of $q_s = 1.04 \pm 0.08$ GeV as obtained from the measurements in Ref. [17] is indicated by a black vertical line

Table 1 Results of the fit on individual m_{DY} intervals for the CMS measurement [17]. The "data" uncertainty is the one estimated using $\min(\chi^2) + 1$, the "scan" uncertainty accounts for the step size of the q_s scan, and the "bins" uncertainty is derived by varying the number of bins included in the fit. The number of bins used in the fit gives n.d.f

m_{DY} region	Best χ^2	n.d.f	Best fit q_s [GeV]
50–76 GeV	2.45	3	$1.00 \pm 0.08(\text{data}) \pm 0.05(\text{scan}) \pm 0.1(\text{bins})$
76–106 GeV	11.4	7	$1.03 \pm 0.03(\text{data}) \pm 0.05(\text{scan}) \pm 0.05(\text{bins})$
106–170 GeV	6.46	4	$1.11 \pm 0.13(\text{data}) \pm 0.05(\text{scan}) \pm 0.2(\text{bins})$
170–350 GeV	4.62	4	$1.1^{+0.24}_{-0.18}(\text{data})$
350–1000 GeV	1.04	4	< 1.9

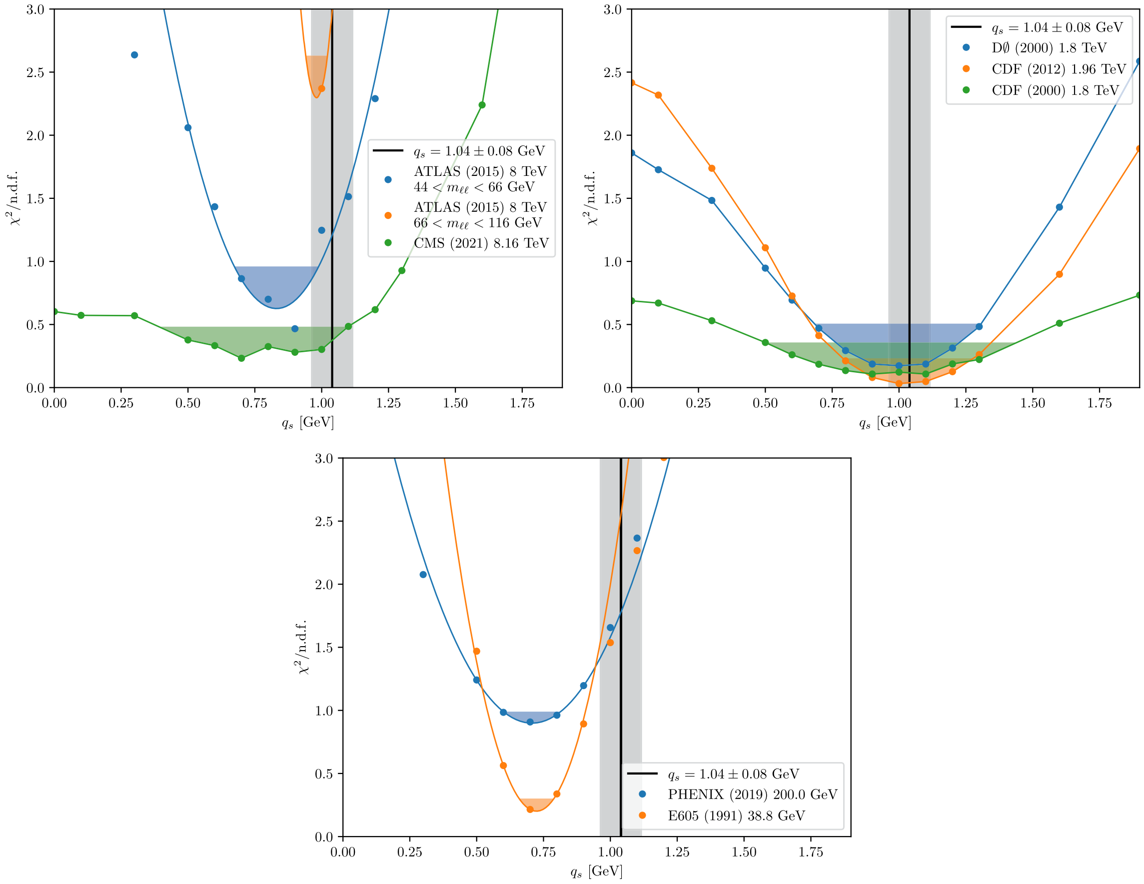


Fig. 11 The reduced χ^2/n.d.f. distribution as a function of q_s obtained from a comparison of the MCatNLO+CAS3 PB-NLO-2018 Set2 prediction with the measurements at lower center-of-mass energies. The colored shaded band shows the χ^2 variation of one unit for each data set. The value of $q_s = 1.04 \pm 0.08$ GeV is shown as the grey band. **Left-top**: ATLAS measurement in 2 mass bins that we analysed at $\sqrt{s} = 8$ TeV (n.d.f. = 4 for each mass bin) [61] and CMS in pPb at $\sqrt{s} = 8.1$ TeV [67]. **Right-top**: Tevatron measurements - D0 at $\sqrt{s} = 1.8$ TeV (n.d.f. = 4) [62], CDF at $\sqrt{s} = 1.8$ TeV (n.d.f. = 5) [63] and $\sqrt{s} = 1.96$ TeV (n.d.f. = 6) [64] **Bottom**: Measurements at lower energies - PHENIX at $\sqrt{s} = 200$ GeV (n.d.f. = 12) [65] and E605 at $\sqrt{s} = 38.8$ GeV (n.d.f. = 11) [66]

3.2.2 *Z production at high rapidities at 13 TeV*

The LHCb collaboration [58] has measured Z-production at $\sqrt{s} = 13$ TeV in the forward region, covering a rapdity range of $2 < |y| < 4.5$.

The χ^2 distribution is shown in Fig. 10 summed over the rapidity range of the DY lepton pair as a function of q_s. A minimum is obtained for $q_s = 0.74 \pm 0.15$ GeV, where the uncertainty comes from a variation of χ^2 by one unit and from the step size of the q_s scan.

3.3 The Gauss widths q_s from lower center of mass energies

The ATLAS collaboration has measured the production of DY from collisions at $\sqrt{s} = 8$ TeV in several DY mass bins, out of which only the two with $44 < m_{\rm DY} < 66$ GeV and $66 < m_{\rm DY} < 116$ GeV are relevant for $p_{\rm T}(\ell\ell) < 10$ GeV [61]. In Fig. 11 we show the χ^2/n.d.f as a function of q_s obtained from these two mass bins (n.d.f = 8).

The Tevatron experiments D0 [62] and CDF have measured transverse momenta of DY lepton pairs created in $p\bar{p}$ collisions at lower center-of-mass energies (1.8 TeV [63] and 1.96 TeV [64]). The PHENIX collaboration measured DY production at $\sqrt{s} = 200$ GeV [65], and E605 [66] at $\sqrt{s} = 38.8$ GeV. The Drell–Yan differential cross section in $p_{\rm T}(\ell\ell)$ has also been measured in pPb data at $\sqrt{s} = 8.1$ TeV by CMS [67]. Figure 11 shows the impact that the q_s choice has on χ^2/n.d.f for these different measurements.

3.4 Consistency between determinations of intrinsic $k_{\rm T}$ width

A global fit of q_s is obtained by calculating χ^2 for different measurements, as shown in Table 2, including the corresponding center-of-mass energies, collision types and the

Table 2 All data sets with the corresponding center-of-mass energies, collision types and the number of degrees of freedom used for the global fit of q_s

Analysis	$\sqrt{s}$	Collision types	n.d.f
CMS_2022_I2079374 [17]	13 TeV	pp	25
LHCb_2022_I1990313 [58]	13 TeV	pp	5
CMS_2021_I1849180 [67]	8.1 TeV	pPb	5
ATLAS_2015_I1408516 [61]	8 TeV	pp	8
CDF_2012_I1124333 [64]	1.96 TeV	p$\bar{\text{p}}$	6
CDF_2000_S4155203 [63]	1.8 TeV	p$\bar{\text{p}}$	5
D0_2000_I503361 [62]	1.8 TeV	p$\bar{\text{p}}$	4
PHENIX_2019_I1672015 [65]	200 GeV	p$\bar{\text{p}}$	12
E605_1991_I302822 [66]	38.8 GeV	pp	11
Total			81

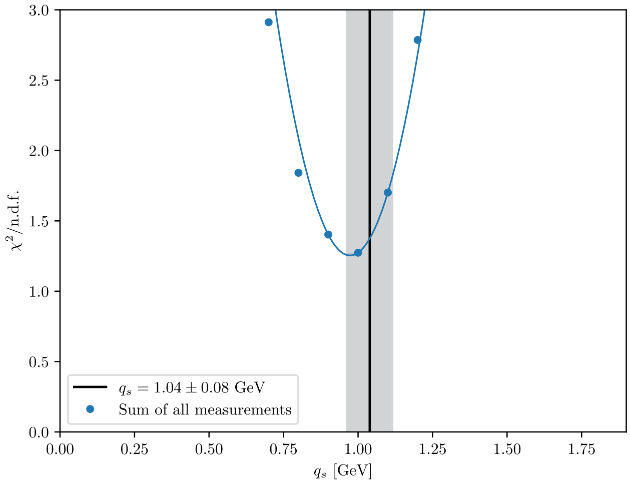


Fig. 12 The reduced χ^2/n.d.f. distribution (n.d.f. = 81) as a function of q_s obtained from a comparison of the MCatNLO+CAS3 PB-NLO-2018 Set2 prediction with the measurement of Refs. [17,58,61–67]. The minimum of global DY data fit is close to $q_s = 1$ GeV and consistent with the CMS measurement [17] shown separately by a black line

number of fitted data points, resulting in a total of 81 data points.

The impact of intrinsic-k_T distribution at lower collision energies has been analyzed using the entire range of $p_\text{T}(\ell\ell)$, while at higher center-of-mass energies we investigate up to the peak region in the transverse momentum distribution.

The χ^2/n.d.f distribution as a function of q_s, for all the data together, is shown in Fig. 12. The χ^2 distribution exhibits a minimum at around $q_s = 1.0$ GeV, which is consistent with the value obtained as described above.

Figure 13 displays the value of q_s as a function of m_DY and $\sqrt{s}$ obtained from the different measurements in Refs. [17,58,61–67]. For the data which do not provide detailed uncertainty breakdown and are mainly used for the cross checks and comparison purpose, the uncertainty bars of q_s shown in the figures are obtained from the χ^2 variation of one unit and step size of the q_s scan only. The value of $q_s = 1.04 \pm 0.08$ GeV as derived from the measurements in Ref. [17], is compatible for all ranges of m_DY, and also holds true for various values of $\sqrt{s}$. The obtained value is also found to be compatible for pPb data.

To summarize, we have obtained a value for the width of the Gauss distribution for modeling the intrinsic-k_T distribution inside protons of $q_s = 1.04 \pm 0.08$ GeV. This value, in contrast to standard Monte Carlo event generators, has no strong dependence on the center-of-mass energy as well as on the mass of the produced Drell–Yan lepton pair m_DY. The results of this section indicate that the treatment of soft emissions in the region $z_\text{dyn} \lesssim z < z_M$ with the strong coupling of Eq. (6) applied in PB-NLO-2018 Set2 leads to intrinsic-k_T distributions with width parameter q_s consistent with Fermi motion kinematics, and mildly varying with energy.

4 Conclusion

In this paper we have carried out a detailed application of the PB-TMD methodology, which is reviewed in the first part of the paper, and used it to describe the DY low transverse momentum distributions across a wide range of DY masses. Within this methodology, we have presented the extraction of the intrinsic-k_T nonperturbative TMD parameter from fits to the measurements of DY p_T differential cross sections performed recently at the LHC at $\sqrt{s} = 13$ TeV, for DY masses between 50 GeV and 1 TeV. We have compared this with extractions from other DY measurements at different center-of-mass energies and masses.

As shown previously, the measured DY cross section at low p_T favours a choice of the strong coupling α_s scale to be taken as the transverse momentum of each parton emission, as in angular-ordered CMW parton cascades. This corresponds to the TMD parton distribution set PB-NLO-2018 Set2. In this paper we use PB-NLO-2018 Set2 with "pre-confinement" scale q_0 of 1 GeV. The strong coupling is evaluated at the emitted transverse momentum q_T for emissions with $q_T > q_0$, populating the phase space region $z < z_\text{dyn}$ (where $z_\text{dyn} = 1 - q_0/|\mathbf{q}'|$, with $|\mathbf{q}'|$ being the scale of the branching), while it is evaluated at the semi-hard scale q_0 for emissions with $q_T \lesssim q_0$. The contribution to the Sudakov evolution from the parton branching in the phase space region $z < z_\text{dyn}$ gives the perturbative resummation of Sudakov logarithms, while the contribution from the parton branching in the phase space region $z_\text{dyn} \lesssim z < z_M$ gives the nonperturbative Sudakov form factor. Therefore the PB-NLO-2018 Set2 contains two sources of nonperturbative effects: i) the nonperturbative TMD distribution at low evolution scale μ_0, and ii) the nonperturbative Sudakov form factor, specified by the "pre-confinement" scale prescription to continue the branching evolution to the infrared region

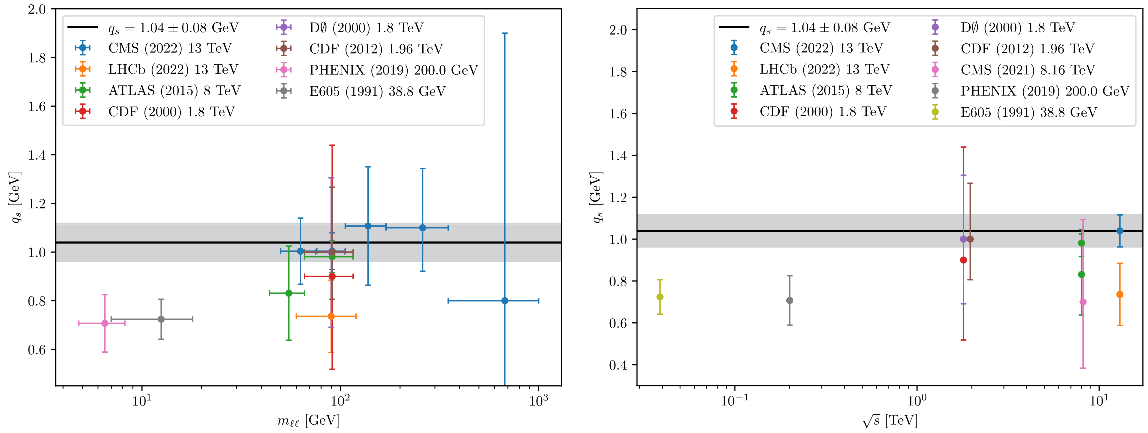


Fig. 13 **Left**: the value of q_s as a function of the DY-mass as obtained from the measurements in Refs. [17,58,61–67]. **Right**: same as a function of $\sqrt{s}$. The value of $q_s = 1.04 \pm 0.08$ GeV as obtained from the measurements in Ref. [17] is indicated

$z_{\rm dyn} \lesssim z < z_M$. The former includes the intrinsic-k_T width parameter q_s, corresponding to Fermi motion in the hadron beam, while the latter is characterized by the semi-hard scale parameter q_0. At low $k_{\rm T}$, the contribution of nonperturbative Sudakov form factor interplays with the contribution of the intrinsic transverse momentum.

The main result of the present work is the extraction, within the PB-NLO-2018 Set2 framework, of the intrinsic-k_T Gauss distribution with zero mean and width parameter $q_s = \sqrt{2}\sigma$ from the measured $p_{\rm T}$ dependence of the DY cross sections obtained recently at the LHC at $\sqrt{s} = 13$ TeV [17], for different DY masses $m_{\rm DY}$, between 50 GeV and 1 TeV. These measurements provide a complete decomposition of the different systematic uncertainties and their covariance matrices. To compare to the data, we have used DY production at NLO obtained with the MadGraph5_aMC@NLO event generator matched with the PB TMD distributions PB-NLO-2018 Set2, with a given parameter q_s value. We performed a scan over a large range of values q_s on the transverse momentum spectrum below the peak, i.e. the sensitive part to intrinsic-$k_{\rm T}$, and considering separately each experimental source of uncertainty and their correlations. The theory scale uncertainties have been considered to be fully correlated inside each $m_{\rm DY}$ bin and uncorrelated between $m_{\rm DY}$ bins. We found the value $q_s = 1.04 \pm 0.08$ GeV, consistent for the different $m_{\rm DY}$. The obtained value is in agreement to the expected value from Fermi-motion in protons. It has been cross checked that this value is compatible with q_s values obtained from other DY measurements at different center-of-mass energies $\sqrt{s}$ and for a variety of DY masses. The global picture shows no strong dependence of the intrinsic-$k_{\rm T}$ on the center-of-mass energy or on the DY mass, which contrasts with tuned standard Monte Carlo event generators that need a strongly increasing intrinsic Gauss width with $\sqrt{s}$ and with $m_{\rm DY}$.

We suggest that the remarkably stable value of q_s that we obtain in our study can be attributed to the contribution of the nonperturbative Sudakov form factor and the treatment of the $z_{\rm dyn} \lesssim z < z_M$ region near the soft-gluon resolution boundary.

Acknowledgements We are grateful to M. Abdullah Al-Mashad, L.I. Estevez Banos, L. Keersmaekers, A.M. van Kampen, K. Wichmann, Q. Wang, H. Yang from the CASCADE developer group for interesting and productive discussions. We are grateful to Louie Corpe for developing a code to calculate χ^2 including correlations within the Rivet frame. We also thank Marius Ambrozas for many discussion on the treatment of systematic uncertainties. This article is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement STRONG 2020 - No 824093. LF is supported by the F.R.S.-FNRS of Belgium. A. L. acknowledges funding by Research Foundation-Flanders (FWO) (application number: 1272421N). LM acknowledges the support of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2121 "Quantum Universe" - 390833306.

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors' comment: The PB TMD grid files are stored for TMDlib (the source is available from https://tmdlib.hepforge.org/). The data grid files are available from the DESY cloud: https://syncandshare.desy.de/index.php/s/ncL62ttzreJ83ZT.]

Funded by SCOAP^3.

References

1. S. Drell, T.-M. Yan, Massive lepton pair production in hadron-hadron collisions at high-energies. Phys. Rev. Lett. **25**, 316 (1970)
2. J.C. Collins, D.E. Soper, G.F. Sterman, Transverse Momentum Distribution in Drell-Yan Pair and W and Z Boson production. Nucl. Phys. B **250**, 199 (1985)
3. T. Sjöstrand et al., An introduction to PYTHIA 8.2. Comput. Phys. Commun. **191**, 159 (2015). arXiv:1410.3012
4. J. Bellm et al., Herwig 7.0/Herwig++ 3.0 release note. Eur. Phys. J. C **76**, 196 (2016). arXiv:1512.01178
5. M. Bahr et al., Herwig++: physics and manual. Eur. Phys. J. C **58**, 639–707 (2008). arXiv:0803.0883
6. T. Gleisberg et al., Event generation with SHERPA 1.1. JHEP **0902**, 007 (2009). arXiv:0811.4622
7. A. Bermudez Martinez et al., Production of Z-bosons in the parton branching method. Phys. Rev. D **100**, 074027 (2019). arXiv:1906.00919
8. A. Bermudez Martinez et al., The transverse momentum spectrum of low mass Drell–Yan production at next-to-leading order in the parton branching method. Eur. Phys. J. C **80**, 598 (2020). arXiv:2001.06488
9. F. Hautmann et al., Collinear and TMD quark and gluon densities from Parton Branching solution of QCD evolution equations. JHEP **01**, 070 (2018). arXiv:1708.03279
10. F. Hautmann et al., Soft-gluon resolution scale in QCD evolution equations. Phys. Lett. B **772**, 446 (2017). arXiv:1704.01757
11. A. Bermudez Martinez, F. Hautmann, M. L. Mangano, TMD evolution and multi-jet merging. Phys. Lett. B **822**, 136700 (2021). arXiv:2107.01224
12. R. Angeles-Martinez et al., Transverse Momentum Dependent (TMD) parton distribution functions: status and prospects. Acta Phys. Polon. B **46**(12), 2501 (2015). arXiv:1507.05267
13. S. Gieseke, M. H. Seymour, A. Siodmok, A Model of non-perturbative gluon emission in an initial state parton shower. JHEP **06**, 001 (2008). arXiv:0712.1199
14. T. Sjöstrand, P. Skands, Multiple interactions and the structure of beam remnants. *JHEP* **03**, 053 (2004). arXiv:hep-ph/0402078
15. J. Isaacson, Y. Fu, C. P. Yuan, Improving ResBos for the precision needs of the LHC. arXiv:2311.09916
16. F. Hautmann, I. Scimemi, A. Vladimirov, Non-perturbative contributions to vector-boson transverse momentum spectra in hadronic collisions. Phys. Lett. B **806**, 135478 (2020). arXiv:2002.12810
17. CMS Collaboration, Measurement of the mass dependence of the transverse momentum of lepton pairs in Drell-Yan production in proton-proton collisions at $\sqrt{s} = 13$ TeV. Eur. Phys. J. C **83**(7), 628 (2023). arXiv:2205.04897
18. A. Bacchetta et al., Unpolarized Transverse Momentum Distributions from a global fit of Drell-Yan and Semi-Inclusive Deep-Inelastic Scattering data. arXiv:2206.07598
19. M. Bury et al., PDF bias and flavor dependence in TMD distributions. JHEP **10**, 118 (2022). arXiv:2201.07114
20. A. Bermudez Martinez et al., Collinear and TMD parton densities from fits to precision DIS measurements in the parton branching method. Phys. Rev. D **99**, 074008 (2019). arXiv:1804.11152
21. H. Jung, S. T. Monfared, T. Wening, Determination of collinear and TMD photon densities using the Parton Branching method. Phys. Lett. B **817**, 136299 (2021). arXiv:2102.01494
22. H. Jung, S. T. Monfared, TMD parton densities and corresponding parton showers: the advantage of four- and five-flavour schemes. arXiv:2106.09791
23. A. Bermudez Martinez et al., The transverse momentum spectrum of low mass Drell–Yan production at next-to-leading order in the parton branching method. Eur. Phys. J. C **80**, 598 (2020). arXiv:2001.06488
24. J. Alwall et al., The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. JHEP **1407**, 079 (2014). arXiv:1405.0301
25. H. Yang et al., Back-to-back azimuthal correlations in Z+jet events at high transverse momentum in the TMD parton branching method at next-to-leading order. Eur. Phys. J. C **82**, 755 (2022). arXiv:2204.01528
26. F. Hautmann et al., A parton branching with transverse momentum dependent splitting functions. Phys. Lett. B **833**, 137276 (2022). arXiv:2205.15873
27. B.R. Webber, Monte Carlo Simulation of Hard Hadronic Processes. Ann. Rev. Nucl. Part. Sci. **36**, 253 (1986)
28. G. Marchesini, B.R. Webber, Monte Carlo Simulation of General Hard Processes with Coherent QCD Radiation. Nucl. Phys. B **310**, 461 (1988)
29. S. Catani, B.R. Webber, G. Marchesini, QCD coherent branching and semiinclusive processes at large x. Nucl. Phys. B **349**, 635 (1991)
30. V. N. Gribov, L. N. Lipatov, Deep inelastic *ep* scattering in perturbation theory. Sov. J. Nucl. Phys. **15**, 438 (1972). [Yad. Fiz.15,781(1972)]
31. L. N. Lipatov, The parton model and perturbation theory. Sov. J. Nucl. Phys. **20**, 94 (1975). [Yad. Fiz.20,181(1974)]
32. G. Altarelli, G. Parisi, Asymptotic freedom in parton language. Nucl. Phys. B **126**, 298 (1977)
33. Y. L. Dokshitzer, Calculation of the structure functions for Deep Inelastic Scattering and e^+e^- annihilation by perturbation theory in Quantum Chromodynamics. Sov. Phys. JETP **46**, 641 (1977). [Zh. Eksp. Teor. Fiz.73,1216(1977)]
34. F. Hautmann, L. Keersmaekers, A. Lelek, A. M. Van Kampen, Dynamical resolution scale in transverse momentum distributions at the LHC. Nucl. Phys. B **949**, 114795 (2019). arXiv:1908.08524
35. ZEUS, H1 Collaboration, Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data. Eur. Phys. J. C **75**, 580 (2015). arXiv:1506.06042
36. xFitter Developers' Team Collaboration, H. Abdolmaleki et al., xFitter: An Open Source QCD Analysis Framework. A resource and reference document for the Snowmass study. **6** (2022). arXiv:2206.12465
37. S. Alekhin et al., HERAFitter, Open Source QCD Fit Project. Eur. Phys. J. C **75**, 304 (2015). arXiv:1410.4412
38. F. Hautmann et al., TMDlib and TMDplotter: library and plotting tools for transverse-momentum-dependent parton distributions. Eur. Phys. J. C **74**(12), 3220 (2014). arXiv:1408.3015
39. N. A. Abdulov et al., TMDlib2 and TMDplotter: a platform for 3D hadron structure studies. Eur. Phys. J. C **81**, 752 (2021). arXiv:2103.09741
40. M. I. Abdulhamid et al., Azimuthal correlations of high transverse momentum jets at next-to-leading order in the parton branching method. Eur. Phys. J. C **82**, 36 (2022). arXiv:2112.10465
41. D. Amati et al., A treatment of hard processes sensitive to the infrared structure of QCD. Nucl. Phys. B **173**, 429 (1980)
42. A. Bassetto, M. Ciafaloni, G. Marchesini, Jet structure and infrared sensitive quantities in perturbative QCD. Phys. Rept. **100**, 201–272 (1983)
43. A. M. van Kampen, Drell-Yan transverse spectra at the LHC: A comparison of parton branching and analytical resummation approaches. SciPost Phys. Proc. **8**, 151 (2022). arXiv:2108.04099
44. A. Bermudez Martinez et al. to be published
45. H. Jung et al., The Parton Branching evolution for collinear and TMD parton densities - uPDFevolv2. to be published, 2023
46. Z. Nagy, D. E. Soper, Evolution of parton showers and parton distribution functions. Phys. Rev. D **102**(1), 014025 (2020). arXiv:2002.04125

47. S. Frixione, B. R. Webber, Correcting for cutoff dependence in backward evolution of QCD parton showers. arXiv:2309.15587
48. M. Mendizabal, F. Guzman, H. Jung, S. Taheri Monfared, On the role of soft gluons in collinear parton densities. arXiv:2309.11802
49. S. Baranov et al., CASCADE3 A Monte Carlo event generator based on TMDs. Eur. Phys. J. C **81**, 425 (2021). arXiv:2101.10221
50. T. Sjöstrand, S. Mrenna, P. Skands, PYTHIA 6.4 physics and manual. JHEP **05**, 026 (2006). arXiv:hep-ph/0603175
51. CMS Collaboration, Measurement of the differential Drell-Yan cross section in proton-proton collisions at $\sqrt{s}$ = 13 TeV. JHEP **12**, 059 (2019). arXiv:1812.10529
52. A. Buckley et al., Rivet user manual. Comput. Phys. Commun. **184**, 2803 (2013). arXiv:1003.0694
53. CMS Collaboration, Measurements of differential Z boson production cross sections in proton-proton collisions at $\sqrt{s}$ = 13 TeV. JHEP **12**, 061 (2019). arXiv:1909.04133
54. ATLAS Collaboration, Measurement of the low-mass Drell-Yan differential cross section at $\sqrt{s}$ = 7 TeV using the ATLAS detector. JHEP **06**, 112 (2014). arXiv:1404.1212
55. CMS Collaboration, Study of Final-State Radiation in Decays of Z Bosons Produced in *pp* Collisions at 7 TeV. Phys. Rev. D **91**, 092012 (2015). arXiv:1502.07940
56. A. Bermudez Martinez, F. Hautmann, M. L. Mangano, Multi-jet merging with TMD parton branching. JHEP **09**, 060 (2022). arXiv:2208.02276
57. A. Bermudez Martinez, F. Hautmann, M. L. Mangano, Multi-jet physics at high-energy colliders and TMD parton evolution (2021). arXiv:2109.08173
58. LHCb Collaboration, Precision measurement of forward *Z* boson production in proton-proton collisions at $\sqrt{s} = 13$ TeV. JHEP **07**, 026 (2022). arXiv:2112.07458
59. L. Moureaux, I. Bubanja, Fits with covariance matrices. https://github.com/lmoureaux/CovarianceFits
60. L. Corpe, *Correlations Library* (Jet and electroweak bosons, Contribution to yellow report of LHCEW working group, 2019)
61. ATLAS Collaboration, Measurement of the transverse momentum and ϕ^*_η distributions of Drell–Yan lepton pairs in proton–proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. Eur. Phys. J. C **76**, 291 (2016). arXiv:1512.02192
62. D0 Collaboration, Measurement of the inclusive differential cross section for *Z* bosons as a function of transverse momentum in $\bar{p}p$ collisions at $\sqrt{s} = 1.8$ TeV. Phys. Rev. D **61**, 032004 (2000). arXiv:hep-ex/9907009
63. CDF Collaboration, The transverse momentum and total cross section of e^+e^- pairs in the *Z* boson region from $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV. Phys. Rev. Lett. **84**, 845 (2000). arXiv:hep-ex/0001021
64. CDF Collaboration, Transverse momentum cross section of e^+e^- pairs in the *Z*-boson region from $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. Phys. Rev. D **86**, 052010 (2012). arXiv:1207.7138
65. PHENIX Collaboration, Measurements of $\mu\mu$ pairs from open heavy flavor and Drell-Yan in $p + p$ collisions at $\sqrt{s} = 200$ GeV. Phys. Rev. D **99**, 072003 (2019). arXiv:1805.02448
66. G. Moreno et al., Dimuon production in proton - copper collisions at $\sqrt{s}$ = 38.8 GeV. Phys. Rev. D **43**, 2815 (1991)
67. CMS Collaboration, Study of Drell-Yan dimuon production in proton-lead collisions at $\sqrt{s_{\mathrm{NN}}}$ = 8.16 TeV. JHEP **05**, 182 (2021). arXiv:2102.13648

L.4 The Parton Branching evolution package UPDFEVOLV2

DESY-24-083
July 23, 2024

The Parton Branching evolution package uPDFevolv2*

H. Jung †[1,2,3], A. Lelek‡[3], K. Moral Figueroa§[1], and S. Taheri Monfared¶[1]

[1]Deutsches Elektronen-Synchrotron DESY, Germany
[2]II. Institut für Theoretische Physik, Universität Hamburg, Germany
[3]Elementary Particle Physics, University of Antwerp, Belgium

Abstract

uPDFevolv2 is a software package designed for evolving collinear and Transverse Momentum Dependent (TMD) parton densities using the DGLAP evolution equation. A comprehensive description of both the theoretical framework and technical implementation is given, accompanied by a detailed guide on program usage, focusing on customizable parameters.

This report is a technical release note for uPDFevolv2 version 2.5.03.

---

[*] This article is dedicated the memory to M. Botje, author of the QCDnum package.

[†] hannes.jung@desy.de
[‡] aleksandra.lelek@uantwerpen.be
[§] keila.moral.figueroa@desy.de
[¶] sara.taheri.monfared@desy.de

PROGRAM SUMMARY

Title of Program: uPDFevolv2 2.5.03
Computer for which the program is designed and others on which it is operable: Any with standard Fortran 77 (gfortran) and C++, tested on Linux, MAC.
Programming Language used: FORTRAN 77, C++
High-speed storage required: No
Separate documentation available: No
Keywords: QCD, DGLAP evolution equation, NLO and NNLO splitting functions, transverse momentum dependent pdf (TMD)
Nature of physical problem: The evolution equations for parton densities cannot be solved analytically and numerical methods need to be applied. Transverse Momentum Dependent parton (TMD) densities can be obtained from the inclusive DGLAP evolution equations once the evolution scale q' is associated with the transverse momentum $q_{\rm t}$ of the emitted parton.
Method of solution: The evolution equations for parton densities are solved numerically using a formulation involving Sudakov form factors. The integral equations are of Fredholm type and can be solved iteratively using a Monte Carlo technique. The iterative solution allows for a treatment of the kinematic relations at each individual branching process, and thus allows directly to calculate the transverse momentum of the emitted partons, leading to a direct calculation of the TMD parton densities.
Restrictions on the complexity of the problem: None
Other Program used: QCDnum for splitting functions and α_s. Root for plotting the result.
Download of the program: `https://updfevolv.hepforge.org`
Unusual features of the program: None

1 Introduction

uPDFevolv2 is a versatile software tool rooted in the Smallx [1,2] program[*] and built upon the Parton Branching (PB) method [4, 5]. It serves as a robust platform for evolving parton densities, both collinear and Transverse Momentum Dependent (TMD), utilizing the widely-used DGLAP evolution equation [6–9]. With its foundation in Smallx and the innovative PB method, uPDFevolv2 offers researchers a powerful framework for studying the intricate dynamics of parton evolution in high-energy physics.

The uPDFevolv2 package described here is a significant extension of uPDFevolv1 [10], which is based on the evolution of gluon densities using the CCFM evolution equation [11–14].

2 Theoretical Input

2.1 The Parton Branching solution of the DGLAP equation

The DGLAP evolution equation can be solved with the PB method, as detailed in Refs. [4,5].

We begin by expressing the DGLAP evolution equation for the momentum-weighted parton density $xf_a(x, \mu^2)$ of parton a with momentum fraction x at scale μ:

$$\mu^2 \frac{\partial (x f_a(x,\mu^2))}{\partial \mu^2} = \sum_b \int_x^1 dz \; P_{ab}\left(z, \alpha_s\right) \; \frac{x}{z} f_b\left(\frac{x}{z}, \mu^2\right), \tag{1}$$

where P_{ab} represents the regularized DGLAP splitting functions governing the transition of parton b into parton a. The function P_{ab} can be decomposed as follows (following Ref. [4]):

$$P_{ab}(z, \alpha_s) = D_{ab}(\alpha_s)\delta(1-z) + K_{ab}(\alpha_s)\frac{1}{(1-z)_+} + R_{ab}(z, \alpha_s). \tag{2}$$

Here, D_{ab} and K_{ab} are coefficients expressed as $D_{ab}(\alpha_s) = \delta_{ab} d_a(\alpha_s)$ and $K_{ab}(\alpha_s) = \delta_{ab} k_a(\alpha_s)$, respectively, while R_{ab} encompasses terms non-singular as $z \to 1$. Each coefficient can be expanded in powers of α_s:

$$d_a(\alpha_s) = \sum_{n=1}^{\infty} \left(\frac{\alpha_s}{2\pi}\right)^n d_a^{(n-1)}, \quad k_a(\alpha_s) = \sum_{n=1}^{\infty} \left(\frac{\alpha_s}{2\pi}\right)^n k_a^{(n-1)}, \quad R_{ab}(z, \alpha_s) = \sum_{n=1}^{\infty} \left(\frac{\alpha_s}{2\pi}\right)^n R_{ab}^{(n-1)}(z). \tag{3}$$

Introducing the real-emission branching probabilities $P_{ab}^{(R)}(z, \alpha_s)$:

$$P_{ab}^{(R)}(z, \alpha_s) = K_{ab}(\alpha_s)\frac{1}{1-z} + R_{ab}(z, \alpha_s) \; , \tag{4}$$

---

[*] The Smallx source code is available from Ref. [3]

the solution to the evolution equation for the momentum-weighted parton density $xf_a(x,\mu^2)$ at scale μ is given by:

$$xf_a(x,\mu^2) = \Delta_a(\mu^2)xf_a(x,\mu_0^2) + \sum_b \int_{\mu_0^2}^{\mu^2} \frac{dq'^2}{q'^2} \frac{\Delta_a(\mu^2)}{\Delta_a(q'^2)} \int_x^{z_M} dz\, P_{ab}^{(R)}(z,\alpha_s) \frac{x}{z} f_b\left(\frac{x}{z}, q'^2\right), \quad (5)$$

where μ_0 is the starting scale, $\Delta_a(\mu^2) := \Delta_a(\mu^2,\mu_0^2)$ is the Sudakov form factor and $\mathbf{q}'$ is a 2-dimensional vector with $\mathbf{q}'^2 = q'^2$.

2.1.1 Determination of inclusive Parton Densities using the Parton-Branching Method

The approach outlined above has been employed to determine collinear (k_T-integrated) parton densities using Monte Carlo techniques with details provided in [4] and briefly summarised in Section 2.2.2.

Figure 1 shows predictions of parton densities evolved to a large scale using the PB method and compares them with calculations obtained using the conventional tool, QCDNUM [15]. The PB predictions are presented for various values of z_M. Notably, when z_M is sufficiently large, the predictions align precisely with semi-analytical calculations, as seen in prior studies [16–20].

These findings hold significance in two respects:

- the DGLAP evolution equation, solved with the concept of resolvable branchings, faithfully reproduces other DGLAP solutions when the "soft resolution" scale z_M is sufficiently large.
- An iterative solution of the DGLAP equation using Monte Carlo techniques based on resolvable branchings, such as the PB method, is equivalent to other DGLAP solutions (e.g. QCDNUM).

The PB method has been applied to determine collinear parton densities by fitting them to inclusive DIS measurements [21]. This resulted in two sets of densities, PB-NLO-HERAI+II-set1 and PB-NLO-HERAI+II-Set2, depending on which scale in α_s is used. In Set1, the evolution scale q' is utilized, while in Set2, the transverse momentum $q'(1-z)$ is employed.

It is crucial to note the significance of z_M: it must be close to one to avoid omitting a significant part of the parton evolution. Here, we illustrate the effect of $z_M = z_{\text{dyn}} = 1 - q_0/q'$, with q_0 being a free resolution parameter of the order of a GeV, on inclusive collinear distributions. The results, obtained using the same parameters and starting distributions as in [21] but with a different cut on z_M, are shown in Figure 2 (for details, refer to [22]).

Figure 2 depicts distributions for down-quarks at different scales, demonstrating the impact when z_M is dynamical: soft gluons, with transverse momenta even below a resolution scale of $q_0 = 1$ GeV, play a significant role in collinear distributions and cannot be neglected.

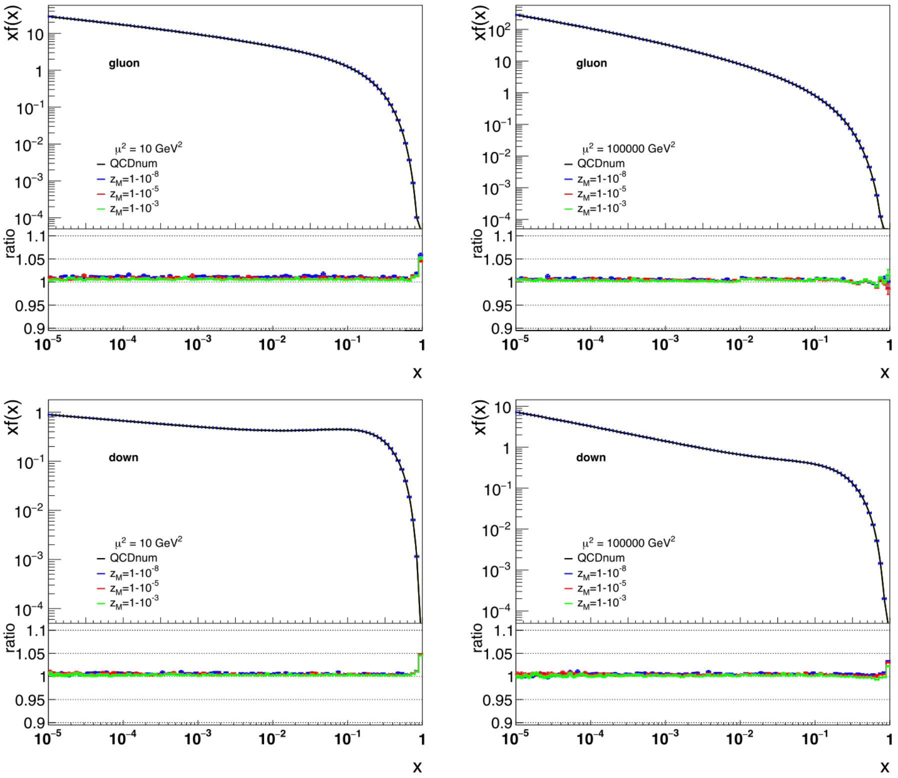


Figure 1: Integrated gluon and down-quark distributions at $\mu^2 = 10$ GeV^2 (left column) and $\mu^2 = 10^5$ GeV^2 (right column) obtained from the PB solution for different values of z_M, compared with the result from QCDNUM. The ratio plots show the ratio of the results obtained with the PB method to the result from QCDNUM. Figure taken from [4].

2.2 Transverse Momentum Dependent parton densities: Parton Branching TMD

Below, we present an extension of the DGLAP evolution equation incorporating the dependence on transverse momenta, as outlined by the PB method [4]. Solving the evolution equation iteratively offers the advantage of treating each splitting explicitly, allowing the application of kinematic relations in every branching, akin to a parton shower process. Consequently, parton distributions can be obtained not only depending on x and μ (as in $f(x, \mu^2)$), but also on the transverse momentum $\mathbf{k}$ of the propagating parton (as in Transverse Momentum Dependent (TMD) parton distributions $\mathcal{A}(x, \mathbf{k}, \mu)$).

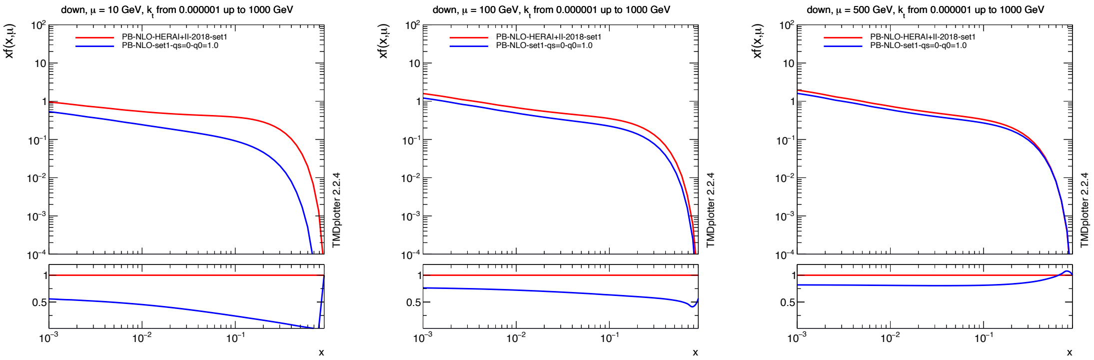


Figure 2: Integrated down-quark distributions at $\mu = 10$, 100, and 500 GeV obtained from the PB approach for different values of z_M: PB-NLO-HERAI+II-set1 applies $z_M \to 1$, and PB-NLO-set1-qs=0-q0=1.0 applies $z_M = z_{\rm dyn}$ with $q_0 = 1$ GeV. The ratio plots show the ratios to the one for $z_M \to 1$ (figures from [22]).

2.2.1 The PB-TMD evolution equation

We can now extend the DGLAP evolution equation to incorporate the dependence on transverse momenta as described by the PB method [4]. The extended evolution equation for the transverse momentum dependent parton density $\mathcal{A}(x, \mathbf{k}, \mu^2)$ is given by:

$$\begin{aligned} x\mathcal{A}_a(x,\mathbf{k},\mu^2) &= \Delta_a(\mu^2)\, x\mathcal{A}_a(x,\mathbf{k},\mu_0^2) + \sum_b \int \frac{d^2\mathbf{q}'}{\pi\mathbf{q}'^2}\, \frac{\Delta_a(\mu^2)}{\Delta_a(\mathbf{q}'^2)}\, \Theta(\mu^2 - \mathbf{q}'^2)\, \Theta(\mathbf{q}'^2 - \mu_0^2) \\ &\times \int_x^{z_M} dz\, P_{ab}^{(R)}(z,\alpha_s)\, \frac{x}{z}\mathcal{A}_b\left(\frac{x}{z}, \mathbf{k} + (1-z)\mathbf{q}', \mathbf{q}'^2\right) \quad , \end{aligned} \tag{6}$$

where the transverse momentum vectors (2-dimensional) $\mathbf{k}$ and $\mathbf{q}$ are used to fully account for the transverse momentum dependence. Here we implicitly assume angular ordering (the default option in uPDFevolv2) which relates the transverse momentum of the emitted parton $q_{\rm t}$ to the evolution scale q' via $q_{\rm t}^2 = (1-z)^2 q'^2$. The final transverse momentum of the propagating parton is calculated as the vectorial sum over intrinsic transverse momentum of the initial parton and all transverse momenta of the emitted partons i:

$$\mathbf{k} = \mathbf{k}_0 - \sum_i \mathbf{q}_{t,i}\ . \tag{7}$$

This enables the determination of the corresponding Transverse Momentum Dependent (TMD) parton distribution $\mathcal{A}(x, \mathbf{k}, \mu^2)$, in addition to the inclusive distribution $f(x, \mu^2)$, integrated over $\mathbf{k}$:

$$\int \mathcal{A}(x, \mathbf{k}, \mu^2) d^2\mathbf{k} = f(x, \mu^2)\ . \tag{8}$$

Similarly to equation (5), also equation (6) can be solved iteratively using a Monte Carlo method [4].

In the literature (e.g. [23]p.69), also virtuality ordering is being discussed, which gives a slightly different relation between the evolution scale and the transverse momentum $q_{\rm t}^2 = (1-z)q'^2$ (available in uPDFevolv2 as option `Iqord=1`). In Figure 3 the transverse momentum distribution for d-quarks is shown for angular ordering (default) as well as for virtuality ordering for different values of the evolution scale μ. It is interesting to observe differences, while the integrated distributions are identical.

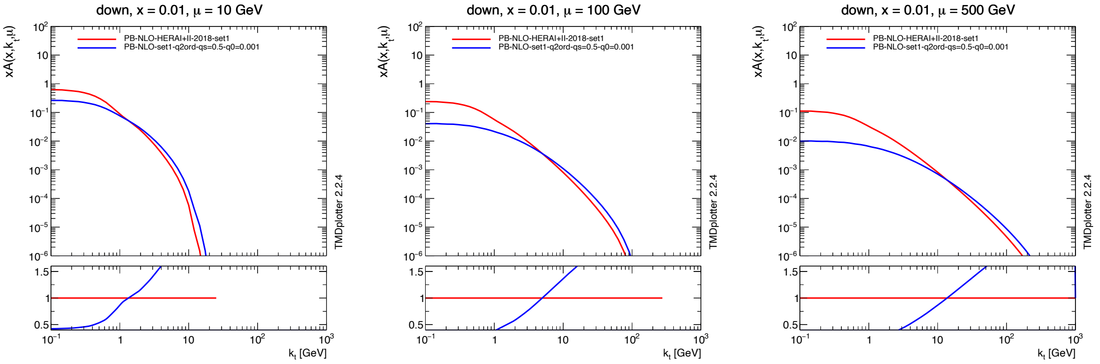


Figure 3: Transverse momentum distributions of down quarks at $\mu = 10, 100$ GeV (left, middle column) and $\mu = 500$ GeV (right column). The red curve shows PB-NLO-HERAI+II-set1, with angular ordering condition applied to relate branching scale q'^2 and the emitted transverse momentum q_t^2, the blue curve shows a prediction obtained with virtuality ordering.

2.2.2 Monte Carlo solution of the evolution equations

As described above, the evolution equations (5, 6) are integral equations of the Fredholm type

$$f(x) = f_0(x) + \lambda \int_a^b K(x,y) f(y) dy$$

and can be solved by iteration as a Neumann series

$$f(t) = \lim_{n\to\infty} \sum_{i=0}^{n} \lambda^i u_i(t) \quad , \tag{9}$$

where

$$
\begin{aligned}
u_0(t) &= f_0(t) \quad , \\
u_1(t) &= \int_a^b K(t,y) f_0(y) dy \quad , \\
u_2(t) &= \int_a^b \int_a^b K(t,y_1) K(y_1,y_2) f_0(y_2) dy_2 dy_1 \quad , \\
\cdots & \\
\vdots & \\
u_n(t) &= \int_a^b \int_a^b \int_a^b K(t,y_1) \cdots K(y_{n-1},y_n) f_0(y_n) dy_n \cdots dy_2 dy_1 \quad . \qquad (10)
\end{aligned}
$$

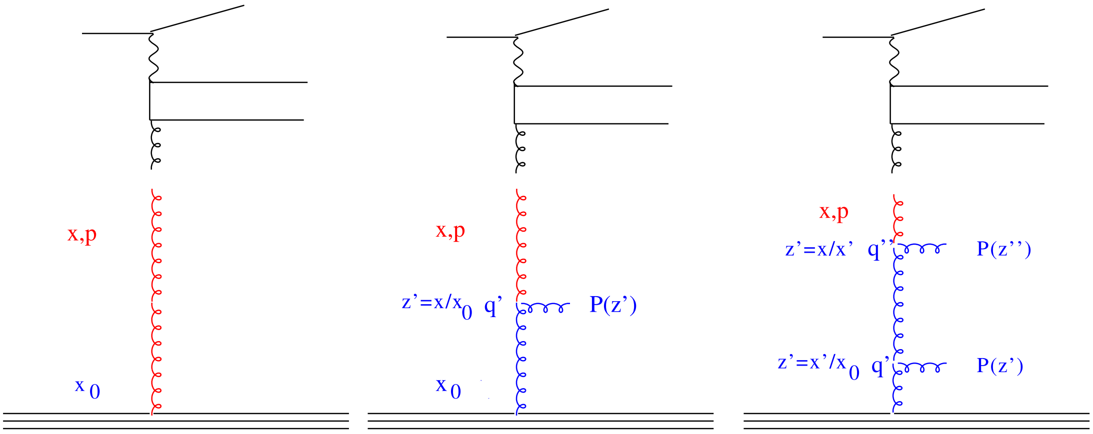


Figure 4: Evolution by iteration

In a procedure termed forward evolution, a Monte Carlo technique [1, 2, 4, 5, 10, 24] is employed to evolve from μ_0 to a value q' determined by the Sudakov factor $\Delta(q')$. This factor represents the probability of evolving from μ_0 to q' without encountering any resolvable branching.

Let us consider a simplified one-flavor scenario for integrated distribution. Utilizing the Sudakov factor Δ and the relation

$$
\frac{\partial}{\partial q'} \Delta(q') \quad = \quad -\Delta(q') \left[\frac{1}{q'} \right] \int^{z_M} dz P^R(z, \alpha_s),
$$

the iterative solution of the evolution equation is expressed as

$$
\begin{aligned}
& xf(x,\mu^2) = xf(x,\mu_0^2)\Delta(\mu^2) \\
+ & \int_{\mu_0^2}^{\mu^2} \int_x^1 dz' (-d\Delta(q'^2)) \frac{\Delta(\mu^2)}{\Delta(q')} P^R(z',\alpha_s) \frac{x}{z'} f(x/z', \mu_0'^2)) \left[\int^{z_M} dz P^R(z,\alpha_s) \right]^{-1} \\
& + \ldots \, . \qquad (11)
\end{aligned}
$$

The value q' is obtained by solving the equation:

$$R_1 = \int^{q'^2} \frac{\partial \Delta(q''^2)}{\partial q''^2} dq''^2 = \Delta(q'^2) \, , \tag{12}$$

where R_1 is a random number sampled from the interval $[0, 1]$.

If $q' > \mu$, the scale μ is reached, and the evolution halts, leaving only the initial term without any resolvable branching. Conversely, if $q' < \mu$, a branching is generated at q' according to the splitting function $P^R(z', \alpha_s)$:

$$\int_{z_{min}}^{z} dz' P^R(z', \alpha_s) \quad = \quad R_1 \int_{z_{min}}^{z_M} dz' P^R(z', \alpha_s), \tag{13}$$

where R_2 is a random number in $[0, 1]$.

The evolution then continues using the Sudakov factor $\Delta(q'', q')$. If $q'' > \mu$, the evolution stops, leaving a single resolvable branching at q'. Otherwise, the evolution continues as described above. This iterative process repeats until q's larger than μ^2 are generated. Through this method, all kinematically allowed contributions in the series are summed, yielding an Monte Carlo estimate of the parton distribution function.

2.3 Splitting functions, α_s, and starting distributions

The evolution in uPDFevolv2 can be carried out at leading order (LO), next-to-leading order (NLO), and next-to-next-to-leading order (NNLO). Splitting functions and α_s are sourced from QCDnum [15], while the initial parton density distributions are obtained in the same format. The NLO and NNLO splitting functions are calculated in Refs [25–35].

Alternatively, LHApdf [36] offers an alternative source for the starting distributions, as well as for the parametrization of α_s.

2.4 Computational Techniques: uPDFevolv2 Grid

In fitting programs where the DGLAP evolution is employed to determine the starting distribution $\mathcal{A}_0(x)$, a full Monte Carlo solution [10] becomes impractical due to its time-consuming nature and susceptibility to numerical fluctuations. Instead, a convolution method introduced in [37,38] is utilized.

The kernel $\mathcal{K}\left(x'', k_{\mathrm{T}}, \mu^2\right)$, where $|\mathbf{k}| = k_{\mathrm{T}}$, is determined from the Monte Carlo solution of the PB evolution equation and then convoluted with the non-perturbative starting distribution $\mathcal{A}_0(x)$:

$$\begin{aligned} x\mathcal{A}(x, k_{\mathrm{T}}, \mu^2) \quad &= \quad x \int dx' \int dx'' \mathcal{A}_0(x') \mathcal{K}\left(x'', k_{\mathrm{T}}, \mu^2\right) \delta(x'x'' - x) \\ &= \quad \int dx' \mathcal{A}_0(x') \cdot \frac{x}{x'} \, \mathcal{K}\left(\frac{x}{x'}, k_{\mathrm{T}}, \mu^2\right). \end{aligned} \tag{14}$$

The kernel $\mathcal{K}$ encapsulates all dynamics of the evolution, including Sudakov form factors and splitting functions. It is determined on a grid of $50 \otimes 50 \otimes 50$ bins in x, k_{T}, μ^2. The grid's binning is logarithmic, with 40 bins in logarithmic spacing below 0.1 for the longitudinal variable x, and 10 bins in linear spacing above 0.1.

The starting distributions can be obtained from fits to measurements via the xFitter platform [39, 40]. The user has to provide the grid files for gluons, as well as for u- and d-type quarks obtained from uPDFEVOLV2, both collinear and TMD grids are needed. These grid files are being used inside xFitter convoluted with the starting distributions to provide evolved parton distribution functions (pdfs).

After a fit is performed, the resulting collinear pdfs are written in LHApdf format, the TMD parton densities are written in TMDlib format [41]. Both collinear and TMD parton distributions can be plotted using the graphical web interface TMDPLOTTER [41].

2.5 Determination of electroweak particle densities

The DGLAP evolution equation can be extended to include also photons and other electroweak particles [42–46]. Since the quarks carry different electric and weak charges, it is necessary to split the evolution into u-type and d-type quarks.

The evolution of the photon density inside a hadron has been described in Ref. [47, 48]. In Figure 5 the collinear photon density as a function of x is shown for different values of the evolution scales μ (from Ref. [47]). For comparison also the prediction from CT14qed-proton [49] is shown.

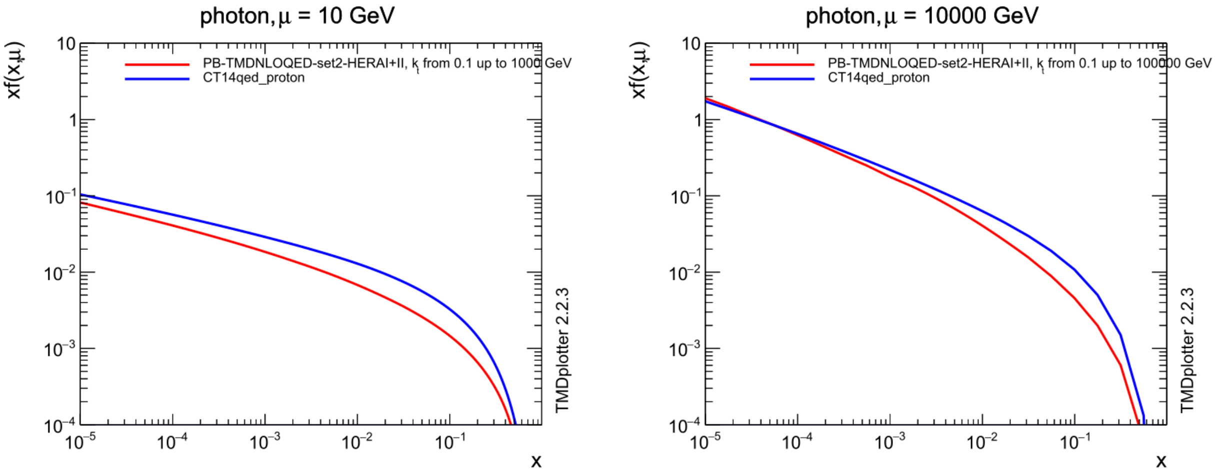


Figure 5: The collinear photon density at $\mu = 10$ GeV and $\mu = 10000$ GeV as a function of x. For comparison the prediction from CT14qed-proton [49] is also shown.

In Figure 6 the TMD density of photons is shown (from Ref. [47]).

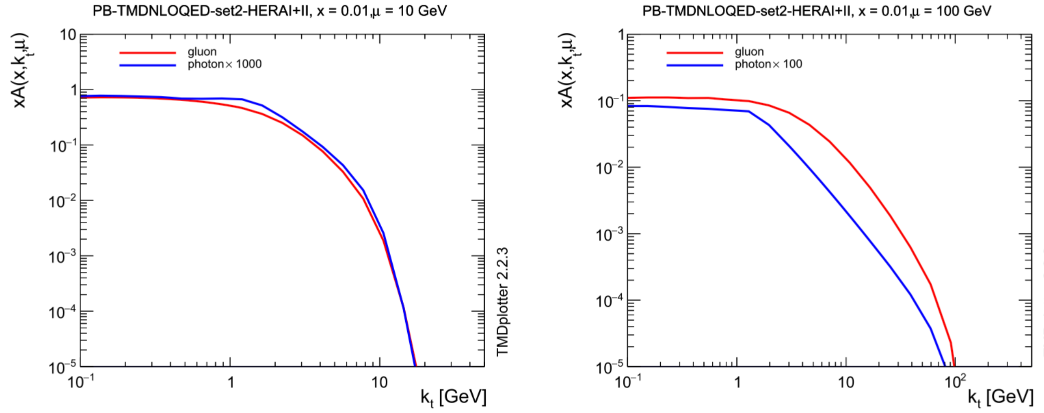


Figure 6: The TMD photon density at $\mu = 10$ GeV and $\mu = 100$ GeV as a function of k_{T}. For comparison the gluon density is also shown. Plot from [47].

The determination of effective W densities has been discussed already in Refs. [50–58]. In recent years, this ideas has been picked up again in Refs. [42, 43, 45, 59, 60].

The approach to determine the photon densities within the PB-method can be easily extended to calculate the collinear and TMD densities of Z and W. The straightforward application of the method gives the collinear and TMD densities as shown in Figures 7,8, compared also to the photon density.

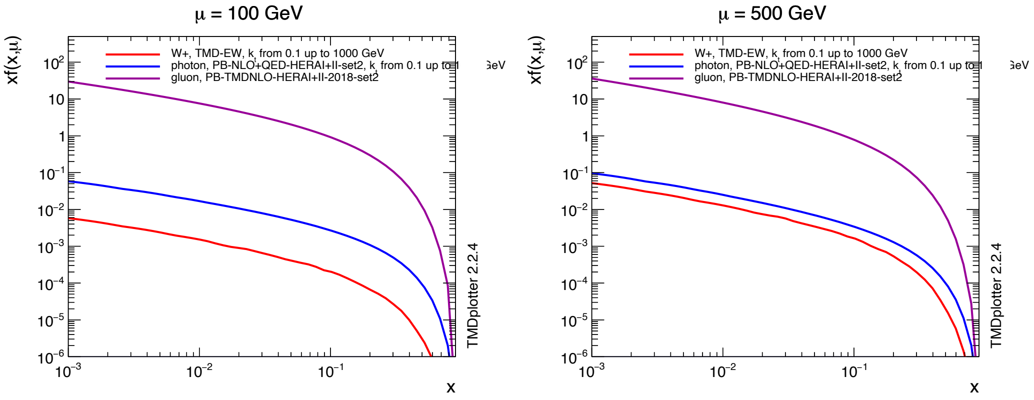


Figure 7: The collinear vector boson densities at $\mu = 100$ GeV and $\mu = 500$ GeV as a function of x.

The heavy vector-boson density vanishes (in the approach applied here) for scales $\mu < m_{\mathrm{W}}$ therefore the densities are only shown for larger scales. For higher scales, the photon and W densities approach each other, as they should.

In the transverse momentum distribution (Figure 8), one can observe the similarity of

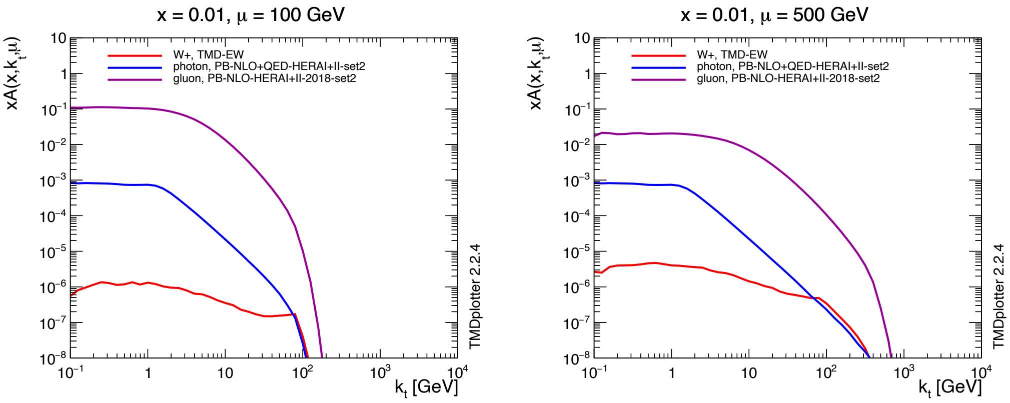


Figure 8: The TMD vector boson densities at $\mu = 100$ GeV and $\mu = 500$ GeV as a function of k_{T}.

photon and W densities at large k_{T}, while significant differences are visible at smaller k_{T}.

2.6 Application: Predictions using PB collinear and TMD distributions

The PB TMDs were fitted to the DIS precision data [61], as described in detail in Ref. [21].

The CASCADE3 Monte Carlo generator [62], engineered specifically to conform to the PB TMD approach, is the only generator with parton shower based on TMDs. Special strength of this approach is that CASCADE3 allows to simulate parton showers fully consistent with TMD parton distribution functions. The PB method and CASCADE3 provide successful predictions for Deep-Inelastic Scattering (DIS) [21, 63], inclusive Drell-Yan (DY) at different $\sqrt{s}$ and mass ranges [64–67], Z+jets [68, 69] and dijets [70, 71]. Furthermore, the PB method has also been included in various publications of CMS [70, 72, 73] and H1 [63] showing a great potential for becoming commonly used in the experimental community.

3 Description of the program components

3.1 Subroutines and functions

The source code of uPDFevolv2 and this manual can be found under:
`https://updfevolv.hepforge.org/`

`sminit`	to initialise
`sminfn`	to generate starting distributions in x and k_T
`smbran`	to simulate perturbative branchings
`szval`	to calculate z values for the splitting
`smqtem`	to generate q' from the corresponding Sudakov factor
`updfgrid`	to build, fill and normalise the updf grid.

`asbmy(q)`	to calculate $\frac{\alpha_s(q)}{2\pi}$ at LO, NLO or NNLO using the utility from QCDNUM or to take it from LHApdf.

Utility routines:

`evolve tmd`	Main routine to perform parton evolution
`updfread`	example program to read and plot the results
`gadap`	1-dimensional Gauss integration routine
`gadap2`	2-dimensional Gauss integration routine
`divdif`	linear interpolation routine (CERNLIB)
`ranlux`	Random number generator `RANLUX` (CERNLIB)

3.2 Parameter in steering files

`Ipdf = 60500`	set name for collinear valence quark starting distribution IPDF=-1: QCDNUM parametrization used as the starting distribution IPDF>0: LHApdf set name for the starting distribution
`Qg = 1.3`	starting value μ_0 for perturbative evolution
`Qs = 1.0`	Gaussian width for the intrinsic k_T distribution
`Iordas = 2`	order in perturbation theory for splitting functions (iordas = 1 is LO, iordas=2 is NLO, iordas=3 is NNLO). If QCDNUM is used as a starting distribution, Iordas defines also the order of α_s (for IPDF>0, α_s is accessed from LHApdf)
`Iqord = 2`	ordering definition to calculate scale in α_s, the transverse momentum q_{t} and z_M. Iqord=0 : Angular ordering with the scale of $(1-z)^2 {q'}^2$ in α_s and $z_M = z_{\mathrm{dyn}} = 1 - q_0/q'$ Iqord=1 : Virtuality ordering with the scale of $(1-z){q'}^2$ in α_s and $z_M = z_{\mathrm{dyn}} = 1 - (q_0/q')^2$ Iqord=2: Angular ordering with the scale of ${q'}^2$ in α_s and fixed z_M
`zmaxfixed = 0.999999`	active only for Iqord=2
`Nev = 1000000`	Number of generated events
`Q0ord=0.01`	active only for Iqord=0 and 1 q_0 value for dynamical $z_M = z_{\mathrm{dyn}}$
`Ikern=0`	
	Ikern=0 : full evolution with starting distribution (run ends up with two grid files for TMD and collinear distribution) Ikern=1 : only kernel to be used in xFitter (run ends up with four grid files, two for gluon and quark TMD and two collinear kernels)
`mc=1.47`	charm mass
`mb=4.5`	bottom mass
`mt=173.`	top mass

`asZ=0.118` $\alpha_s(m_Z)$

3.3 Storing of the outputs

`updf-grid.dat`	name of the grid file for TMD outcome.
`updf-grid_int.dat`	name of the grid file for collinear outcome.
`test.root`	name of the root file containing histograms with collinear pdfs.

In order to have enough statistics, usually 900 jobs with 10^6 events are needed. The results of each job are then added to produce the final grid files. The code for doing this can be obtained from the authors.

4 Program Installation

UPDFEVOLV2 follows the standard AUTOMAKE convention. To install the program, do the following

```
1) Get the source

tar xvfz uPDFevolv2-XXXX.tar.gz
cd uPDFevolv2-XXXX

2) Generate the Makefiles (do not use shared libraries)
./configure

3) Compile the binary
make

4) Install the executable
make install

4) The executable is in bin

run it with:
bin/updf_evolve < steering.txt

plot the result with:

bin/updfread
```

5 Acknowledgments

We are grateful to R. Zlebcik and L. Keersmaekers for many discussions during the evolution of the UPDFEVOLV2 code. A. Lelek acknowledges funding by Research Foundation-Flanders (FWO) (application number: 1272421N).

References

[1] G. Marchesini and B. R. Webber, “Final states in heavy quark leptoproduction at small x”, *Nucl. Phys. B* **386** (1992) 215.

[2] G. Marchesini and B. R. Webber, “Simulation of QCD initial state radiation at small x”, *Nucl. Phys.* **B349** (1991) 617.

[3] B. Webber, “The SMALLX program package”. Available from updfevolv download page.

[4] F. Hautmann et al., “Collinear and TMD quark and gluon densities from Parton Branching solution of QCD evolution equations”, *JHEP* **01** (2018) 070, `arXiv:1708.03279`.

[5] F. Hautmann et al., “Soft-gluon resolution scale in QCD evolution equations”, *Phys. Lett. B* **772** (2017) 446, `arXiv:1704.01757`.

[6] V. N. Gribov and L. N. Lipatov, “Deep inelastic ep scattering in perturbation theory”, *Sov. J. Nucl. Phys.* **15** (1972) 438. [Yad. Fiz.15,781(1972)].

[7] L. N. Lipatov, “The parton model and perturbation theory”, *Sov. J. Nucl. Phys.* **20** (1975) 94. [Yad. Fiz.20,181(1974)].

[8] G. Altarelli and G. Parisi, “Asymptotic freedom in parton language”, *Nucl. Phys. B* **126** (1977) 298.

[9] Y. L. Dokshitzer, “Calculation of the structure functions for Deep Inelastic Scattering and e^+e^- annihilation by perturbation theory in Quantum Chromodynamics.”, *Sov. Phys. JETP* **46** (1977) 641. [Zh. Eksp. Teor. Fiz.73,1216(1977)].

[10] F. Hautmann, H. Jung, and S. T. Monfared, “The CCFM uPDF evolution uPDFevolv”, *Eur. Phys. J. C* **74** (2014) 3082, `arXiv:1407.5935`.

[11] M. Ciafaloni, “Coherence effects in initial jets at small Q^2/s.”, *Nucl. Phys. B* **296** (1988) 49.

[12] S. Catani, F. Fiorani, and G. Marchesini, “QCD coherence in initial state radiation”, *Phys. Lett. B* **234** (1990) 339.

[13] S. Catani, F. Fiorani, and G. Marchesini, “Small x behavior of initial state radiation in perturbative QCD”, *Nucl. Phys. B* **336** (1990) 18.

[14] G. Marchesini, “QCD coherence in the structure function and associated distributions at small x”, *Nucl. Phys. B* **445** (1995) 49, `arXiv:hep-ph/9412327`.

[15] M. Botje, “QCDNUM: fast QCD evolution and convolution”, *Comput.Phys.Commun.* **182** (2011) 490, `arXiv:1005.1481`.

[16] K. J. Golec-Biernat et al., “Markovian Monte Carlo solutions of the one-loop CCFM equations”, *Acta Phys. Polon.* **B38** (2007) 3149, `arXiv:hep-ph/0703317`.

[17] K. J. Golec-Biernat, S. Jadach, W. Placzek, and M. Skrzypek, “Markovian Monte Carlo solutions of the NLO QCD evolution equations”, *Acta Phys. Polon.* **B37** (2006) 1785, `arXiv:hep-ph/0603031`.

[18] S. Jadach et al., “Constrained MC for QCD evolution with rapidity ordering and minimum kT*”, *Comput. Phys. Commun.* **180** (2009) 675, `arXiv:hep-ph/0703281`.

[19] S. Jadach and M. Skrzypek, “QCD evolution in the fully unintegrated form”, *Acta Phys. Polon. B* **40** (2009) 2071, `arXiv:0905.1399`.

[20] S. Jadach and M. Skrzypek, “Exact solutions of the QCD evolution equations using Monte Carlo method”, *Acta Phys. Polon.* **B35** (2004) 745, `arXiv:hep-ph/0312355`.

[21] A. Bermudez Martinez et al., “Collinear and TMD parton densities from fits to precision DIS measurements in the parton branching method”, *Phys. Rev. D* **99** (2019) 074008, `arXiv:1804.11152`.

[22] M. Mendizabal, F. Guzman, H. Jung, and S. Taheri Monfared, “On the role of soft gluons in collinear parton densities”, `arXiv:2309.11802`.

[23] C. Bierlich et al., “A comprehensive guide to the physics and usage of PYTHIA 8.3”, `arXiv:2203.11601`.

[24] F. James, “Monte Carlo Theory and Practice”, *Rept.Prog.Phys.* **43** (1980) 1145.

[25] W. L. van Neerven and A. Vogt, “Improved approximations for the three loop splitting functions in QCD”, *Phys. Lett. B* **490** (2000) 111–118, `arXiv:hep-ph/0007362`.

[26] S. Moch, J. A. M. Vermaseren, and A. Vogt, “The Three loop splitting functions in QCD: The Nonsinglet case”, *Nucl. Phys.* **B688** (2004) 101, `arXiv:hep-ph/0403192`.

[27] A. Vogt, S. Moch, and J. A. M. Vermaseren, “The Three-loop splitting functions in QCD: The Singlet case”, *Nucl. Phys.* **B691** (2004) 129, `arXiv:hep-ph/0404111`.

[28] J. Vermaseren, A. Vogt, and S. Moch, “The Third-order QCD corrections to deep-inelastic scattering by photon exchange”, *Nucl.Phys.* **B724** (2005) 3, `arXiv:hep-ph/0504242`.

[29] J. Blümlein, P. Marquard, C. Schneider, and K. Schönwald, “The three-loop unpolarized and polarized non-singlet anomalous dimensions from off shell operator matrix elements”, *Nucl. Phys. B* **971** (2021) 115542, `arXiv:2107.06267`.

[30] J. Blümlein, P. Marquard, C. Schneider, and K. Schönwald, “The massless three-loop Wilson coefficients for the deep-inelastic structure functions F_2, F_L, xF_3 and g_1”, *JHEP* **11** (2022) 156, `arXiv:2208.14325`.

[31] J. Ablinger et al., “The transition matrix element Agq(N) of the variable flavor number scheme at $\mathcal{O}(\alpha_s^3)$”, *Nucl. Phys. B* **882** (2014) 263, `arXiv:1402.0359`.

[32] J. Ablinger et al., “The three-loop splitting functions $P_{qg}^{(2)}$ and $P_{gg}^{(2,N_F)}$”, *Nucl. Phys. B* **922** (2017) 1, `arXiv:1705.01508`.

[33] S. Moch, J. A. M. Vermaseren, and A. Vogt, "The Three-Loop Splitting Functions in QCD: The Helicity-Dependent Case", *Nucl. Phys. B* **889** (2014) 351, `arXiv:1409.5131`.

[34] A. Behring et al., "The Polarized Three-Loop Anomalous Dimensions from On-Shell Massive Operator Matrix Elements", *Nucl. Phys. B* **948** (2019) 114753, `arXiv:1908.03779`.

[35] J. Blümlein, P. Marquard, C. Schneider, and K. Schönwald, "The three-loop polarized singlet anomalous dimensions from off-shell operator matrix elements", *JHEP* **01** (2022) 193, `arXiv:2111.12401`.

[36] A. Buckley et al., "LHAPDF6: parton density access in the LHC precision era", *Eur. Phys. J. C* **75** (2015) 132, `arXiv:1412.7420`.

[37] H. Jung and F. Hautmann, "Determination of transverse momentum dependent gluon density from HERA structure function measurements", in *Proceedings, 20th International Workshop on Deep-Inelastic Scattering and Related Subjects (DIS 2012): Bonn, Germany, March 26-30, 2012*, pp. 433–436. Verlag Deutsches Elektronen Sychrotron, Hamburg, Germany, 2012. `arXiv:1206.1796`.

[38] F. Hautmann and H. Jung, "Transverse momentum dependent gluon density from DIS precision data", *Nuclear Physics B* **883** (2014) 1, `arXiv:1312.7875`.

[39] xFitter Developers' Team Collaboration, H. Abdolmaleki et al., "xFitter: An Open Source QCD Analysis Framework. A resource and reference document for the Snowmass study", 6, 2022. `arXiv:2206.12465`.

[40] S. Alekhin et al., "HERAFitter, Open Source QCD Fit Project", *Eur. Phys. J. C* **75** (2015) 304, `arXiv:1410.4412`.

[41] N. A. Abdulov et al., "TMDlib2 and TMDplotter: a platform for 3D hadron structure studies", *Eur. Phys. J. C* **81** (2021) 752, `arXiv:2103.09741`.

[42] B. Fornal, A. V. Manohar, and W. J. Waalewijn, "Electroweak Gauge Boson Parton Distribution Functions", *JHEP* **05** (2018) 106, `arXiv:1803.06347`.

[43] C. W. Bauer, N. Ferland, and B. R. Webber, "Standard Model Parton Distributions at Very High Energies", *JHEP* **08** (2017) 036, `arXiv:1703.08562`.

[44] T. Veness, "The effective W approximation". DESY summerstudent program, 2012.

[45] P. Ciafaloni and D. Comelli, "Electroweak evolution equations", *JHEP* **11** (2005) 022, `arXiv:hep-ph/0505047`.

[46] M. Ciafaloni, P. Ciafaloni, and D. Comelli, "Towards collinear evolution equations in electroweak theory", *Phys. Rev. Lett.* **88** (2002) 102001, `arXiv:hep-ph/0111109`.

[47] H. Jung, S. T. Monfared, and T. Wening, "Determination of collinear and TMD photon densities using the Parton Branching method", *Physics Letters B* **817** (2021) 136299, `arXiv:2102.01494`.

[48] T. Wening, "Transverse Momentum Dependent Parton Density Function for the Photon", masterthesis, University of Hamburg, 2020. Masterarbeit, University of Hamburg, 2020.

[49] C. Schmidt, J. Pumplin, D. Stump, and C. P. Yuan, "CT14QED parton distribution functions from isolated photon production in deep inelastic scattering", *Phys. Rev.* **D93** (2016), no. 11, 114015, `arXiv:1509.02905`.

[50] G. L. Kane, W. Repko, and W. Rolnick, "The Effective W+-, Z0 Approximation for High-Energy Collisions", *Phys. Lett. B* **148** (1984) 367.

[51] J. Lindfors, "Distribution Functions for Heavy Vector Bosons Inside Colliding Particle Beams", *Z. Phys. C* **28** (1985) 427.

[52] R. N. Cahn, "Production of Heavy Higgs Bosons: Comparisons of Exact and Approximate Results", *Nucl. Phys. B* **255** (1985) 341. [Erratum: Nucl.Phys.B 262, 744 (1985)].

[53] S. Dawson, "The Effective W Approximation", *Nucl. Phys. B* **249** (1985) 42.

[54] M. S. Chanowitz and M. K. Gaillard, "The TeV Physics of Strongly Interacting W's and Z's", *Nucl. Phys. B* **261** (1985) 379–431.

[55] R. Kleiss and W. J. Stirling, "Anomalous High-energy Behavior in Boson Fusion", *Phys. Lett. B* **182** (1986) 75.

[56] S. Dawson and S. S. D. Willenbrock, "Heavy Fermion Production in the Effective W Approximation", *Nucl. Phys. B* **284** (1987) 449.

[57] G. Altarelli, B. Mele, and F. Pitolli, "Heavy Higgs Production at Future Colliders", *Nucl. Phys. B* **287** (1987) 205–224.

[58] Z. Kunszt and D. E. Soper, "On the Validity of the Effective W Approximation", *Nucl. Phys. B* **296** (1988) 253.

[59] C. W. Bauer, N. Ferland, and B. R. Webber, "Combining initial-state resummation with fixed-order calculations of electroweak corrections", *JHEP* **04** (2018) 125, `arXiv:1712.07147`.

[60] P. Ciafaloni, G. Co', D. Colferai, and D. Comelli, "Electroweak Evolution Equations and Isospin Conservation", `arXiv:2403.08583`.

[61] H1 and ZEUS Collaboration, "Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data", *Eur. Phys. J. C* **75** (2015) 580, `arXiv:1506.06042`.

[62] S. Baranov et al., "CASCADE3 A Monte Carlo event generator based on TMDs", *Eur. Phys. J. C* **81** (2021) 425, `arXiv:2101.10221`.

[63] H1 Collaboration, "Measurement of Lepton-Jet Correlation in Deep-Inelastic Scattering with the H1 Detector Using Machine Learning for Unfolding", *Phys. Rev. Lett.* **128** (2022), no. 13, 132002, `arXiv:2108.12376`.

[64] A. Bermudez Martinez et al., "The transverse momentum spectrum of low mass Drell–Yan production at next-to-leading order in the parton branching method", *Eur. Phys. J. C* **80** (2020) 598, `arXiv:2001.06488`.

[65] CMS Collaboration, "Measurement of the mass dependence of the transverse momentum of lepton pairs in Drell-Yan production in proton-proton collisions at $\sqrt{s}$ = 13 TeV", *Eur. Phys. J. C* **83** (2023) 628, `arXiv:2205.04897`.

[66] A. Bermudez Martinez et al., "Production of Z-bosons in the parton branching method", *Phys. Rev. D* **100** (2019) 074027, `arXiv:1906.00919`.

[67] I. Bubanja et al., "The small k_{T}region in Drell–Yan production at next-to-leading order with the parton branching method", *Eur. Phys. J. C* **84** (2024) 154, `arXiv:2312.08655`.

[68] H. Yang et al., "Back-to-back azimuthal correlations in Z+jet events at high transverse momentum in the TMD parton branching method at next-to-leading order", *Eur. Phys. J. C* **82** (2022) 755, `arXiv:2204.01528`.

[69] CMS Collaboration, "Measurement of differential cross sections for the production of a Z boson in association with jets in proton-proton collisions at $\sqrt{s}$ = 13 TeV", *Phys. Rev. D* **108** (2023) 052004, `arXiv:2205.02872`.

[70] CMS Collaboration, "Measurements of jet multiplicity and jet transverse momentum in multijet events in proton–proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$", *Eur. Phys. J. C* **83** (2023), no. 8, 742, `arXiv:2210.13557`.

[71] M. I. Abdulhamid et al., "Azimuthal correlations of high transverse momentum jets at next-to-leading order in the parton branching method", *Eur. Phys. J. C* **82** (2022) 36, `arXiv:2112.10465`.

[72] CMS Collaboration, "Azimuthal correlations in Z +jets events in proton–proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$", *Eur. Phys. J. C* **83** (2023), no. 8, 722, `arXiv:2210.16139`.

[73] CMS Collaboration, “Measurement of double-parton scattering in inclusive production of four jets with low transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV”, `arXiv:2109.13822`.