ETP-KA/2026-06

Karlsruhe Institute of Technology

Simultaneous measurements of $t\bar{t}$ production with additional heavy-flavor jets in the semileptonic channel with the CMS experiment

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN (Dr. rer. nat.)

von der KIT-Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT)
angenommene

DISSERTATION

von M.Sc. Rufa Kunnilan Muhammed Rafeek

aus Kerala, Indien

Tag der mündlichen Prüfung: 24. April 2026

Referent:	Prof. Dr. Ulrich Husemann	Institut für Experimentelle Teilchenphysik
Korreferent:	Prof. Dr. Jan Kieseler	Institut für Experimentelle Teilchenphysik

The work was made possible with the support of a scholarship from the German Academic Exchange Service (DAAD).

Erklärung der selbstständigen Anfertigung der Dissertationsschrift
Hiermit erkläre ich, dass ich die Dissertation mit dem Titel

Simultaneous measurements of $t\bar{t}$ production with additional heavy-flavor jets in the semileptonic channel with the CMS experiment

selbstständig angefertigt, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht habe, was aus Arbeiten anderer unverändert oder mit Abänderungen entnommen wurde.

Ich versichere außerdem, dass ich die Dissertation nur in diesem und keinem anderen Promotionsverfahren eingereicht habe und dass diesem Promotionsverfahren keine endgültig gescheiterten Promotionsverfahren vorausgegangen sind.

Karlsruhe, 18. März 2026

.....
(**Rufa K. M. Rafeek**)

Contents

1 Introduction

Understanding the fundamental constituents of matter and the forces that govern their interactions is a central goal of particle physics. Over the past century, a combination of experimental discoveries and theoretical developments has led to the formulation of the Standard Model (SM) of particle physics. It is a remarkably successful framework that describes the known elementary particle physics and three of the four fundamental forces. Despite its predictive power, the SM leaves several fundamental questions unanswered, including the nature of dark matter, the origin of the matter-antimatter asymmetry in the universe, and the hierarchy of particle masses. These open questions motivate both precision measurements of known processes and searches for deviations from SM predictions.

The Large Hadron Collider (LHC) provides an unprecedented opportunity to explore these challenges. With proton-proton collisions at center-of-mass energies reaching $\sqrt{s} = 13$ TeV during Run-II, the LHC enables both precision measurements of known processes and searches for new phenomena beyond the SM (BSM). Among the particles predicted by the SM, the top quark occupies a unique position due to its large mass and its strong coupling to the Higgs sector. Precise studies of processes involving top quarks, therefore, provide an important test of the SM and a sensitive probe for potential deviations arising from new physics.

A major milestone in the LHC program was the discovery of the Higgs boson by the CMS and ATLAS collaborations in 2012 [1, 2], completing the particle content predicted by the SM. Earlier, the top quark, the heaviest known fundamental particle, was discovered at the Tevatron collider by the CDF and DØ experiments in 1995 [3, 4].

In this context, the production of a top quark-antiquark pair ($t\bar{t}$) in association with additional heavy flavor jets or bosons, collectively referred to as $t\bar{t}$+X processes, plays a central role in understanding the dynamics of top-quark production. Processes such as $t\bar{t}$+$b\bar{b}$, $t\bar{t} + c\bar{c}$, $t\bar{t} + H(b\bar{b})$, and $t\bar{t} + Z(b\bar{b})$ contribute to complex final states characterized by high jet multi-

plicities and heavy-flavor signatures. Measurements of these processes are not only interesting in their own but are also essential for improving the modeling of backgrounds in Higgs boson measurements and searches for new physics.

The semileptonic decay channel of $t\bar{t}$ production provides a good environment for such measurements due to its relatively large branching fraction with an experimental signature consisting of one lepton, missing transverse momentum, and multiple jets. However, distinguishing the different $t\bar{t}$+X processes in this final state remains challenging due to the large combinatorial background and the similarity of kinematic signatures between jets originating from top-quark decays and additional heavy-flavor radiation.

To address these challenges, modern particle physics analyses increasingly employ machine learning (ML) techniques to exploit complex correlations among reconstructed physics objects. In this thesis, a classification strategy based on Graph Neural Networks (GNN) is developed to analyze $t\bar{t}$+X processes in the semileptonic decay channel of the $t\bar{t}$ pair using data collected by the CMS experiment during Run-II of the LHC. The analysis proceeds in two stages. First, a node-level prediction (NLP) model identifies jets that are likely to originate from additional heavy-flavor radiation. This information is then used as input to a graph-level prediction (GLP) model that performs event-level classification into different $t\bar{t}$+X categories of interest. Finally, the predicted event categories are used as inputs to a simultaneous statistical fit that extracts the signal strengths of the different processes.

The measurements presented in this thesis, as mentioned, are based on the full Run-II dataset collected by the CMS detector corresponding to p-p collisions at $\sqrt{s} = 13$ TeV. The structure of this thesis is as follows. Chapter 2 describes the contributions of the author to this analysis, which forms a part of a broader effort within our analysis group and the CMS collaboration. Chapter 3 provides an overview of the SM and introduces the physics of top quarks relevant for this study. Chapter 4 describes the experimental environment, the LHC, and the CMS detector. Chapter 5 introduces the definitions used in this analysis, including object reconstruction and event simulation.

The second part of the thesis focuses on measuring $t\bar{t}$+X processes. Chapter 6 presents the physics motivation and conceptual strategy of the analysis. Chapter 7 describes the datasets, simulated samples, reconstructed objects, and systematic uncertainties. Chapter 8 details the ML-based analysis strategy, including the GNN architecture and the classification of additional jets and the event categories. Chapter 9 presents the statistical framework used to extract the results and discusses the fit outcomes. Finally, Chapter 10 summarizes the results and outlines directions for future studies.

2 Preface

My interest in particle physics developed during my master's studies, when I was assigned a seminar on a Nobel-prize-winning discovery. I chose to present the discovery of the Higgs boson, and through this work, I became fascinated by the mechanisms underlying the electroweak symmetry breaking and the experimental efforts required to uncover such phenomena. This experience motivated me to pursue further opportunities in the field, including my studies in the "Particle physics and cosmos" program in Spain. This academic path eventually led me to pursue doctoral research in experimental particle physics.

The analysis presented in this thesis is based on data collected with the CMS detector (see Chapter 4) at the Large Hadron Collider (LHC) at CERN [5]. The CMS experiment is operated by a large international collaboration comprising several thousand scientists, engineers, and students from institutes worldwide. The design, operation, and maintenance of the CMS detector, as well as the development of reconstruction software, simulation tools, and computing infrastructure, are collective efforts of the collaboration. Modern collider analyses therefore rely on contributions of many individuals and working groups, and thus it is not the work of a single researcher or a small group.

In the early stages of my doctorate, I contributed to theoretical acceptance calculations for Drell-Yan vector-boson production using early data from the third LHC data-taking period (Run-III) at $\sqrt{s}$ =13.6 TeV, in collaboration with Dr. Xunwu Zuo. The study involved the evaluation of inclusive and fiducial cross-sections for Z and W boson production using the DYTurbo and FEWZ frameworks, and the determination of the corresponding acceptances together with theoretical uncertainties. A summary of this study is presented in Appendix A. This is part of the study contributed to the measurement done in Ref. [6].

In addition to their own research, members of the collaboration are expected to contribute to the construction, operation, maintenance, and upgrades of the detector by fulfilling Ex-

perimental Physics Responsibilities (EPR) and central data-taking shifts. This ensures the smooth and reliable operation of the detector and the long-term success of the experiment. As part of my EPR work, I contributed to the electrical characterization of silicon sensors for the Phase-2 Outer Tracker upgrade for the High-Luminosity LHC (HL-LHC). The work involved performing sensor measurements using dedicated probe stations in a temperature-controlled environment. I was trained in these measurements by Dr. Florian Wittig. The measurements included electrical tests of silicon sensors and diodes under controlled temperature and humidity conditions to verify their performance and compliance with detector specifications. In particular, the measurements focused on silicon strip sensors used in outer tracker modules, including the 2S strip sensors and the strip component of the PS modules (PS-s), which form part of the tracker system designed to operate under the high radiation and pileup conditions expected at the HL-LHC. In total, about 1250 sensors were qualified in this campaign, of which I contributed to the qualification of 142 sensors, corresponding to roughly 11% of the total.

The main research presented in this thesis concerns the analysis of $t\bar{t}$+X processes in the heavy-flavor phase space of the semileptonic decay channel. This work builds upon earlier developments within the analysis group. In particular, the work of Dr. Jan van der Linden and Dr. Emanuel Pfeffer provided the foundation for this study. Jan introduced me to the analysis framework and workflow, while the machine learning strategy and studies performed by Emanuel in the dileptonic channel served as inputs to the development of the semileptonic analysis. The dileptonic analysis was based on the 2018 LHC Run-II dataset. In this thesis, the analysis is extended to the full Run-II dataset. This required the preparation of necessary samples, including modifications to the configuration in order to incorporate all Run-II datasets, triggers, and reconstructed physics objects relevant for the semileptonic channel. The main contributions include:

- Extension of the analysis to the full Run-II dataset by preparing the necessary samples for the 2016 and 2017 data-taking periods and producing the corresponding reduced datasets (ntuples) required for the analysis.
- Implementation of the graph neural network (GNN) classification framework used in this analysis for the identification of additional jets and the event categorization.
- Preparation and validation of Run II datasets used for the training and evaluation of the machine learning models.
- Implementation of experimental and theoretical systematic uncertainties and the production of control plots to validate the modeling of kinematic observables in data and simulation.
- Statistical interpretation of the results, including the construction of the likelihood model and the extraction of signal strengths for the $t\bar{t}+B$, $t\bar{t}+C$, $t\bar{t}+H$, and $t\bar{t}+Z$ processes.

Furthermore, although a basic implementation of the statistical fit was already available, I did substantial work to refine and optimise the fitting procedure. This included improve-

ments to the event selection strategy, the addition of all relevant nuisance parameters, and detailed studies of the behavior and stability of the fit model. I have also carried out the interpretation of the results and their comparison with previous measurements. Discussions with Dr. Nicolò Trevisani helped clarify several aspects of the fit procedure and its interpretation.

During the preparation of this thesis, software tools such as ChatGPT [7], GitHub Copilot [8], and Grammarly [9] were used to assist with grammar checking, text summarization, and debugging of code. These tools were only used as supportive aids. All scientific content, analysis design, interpretation of results, and final decisions regarding the material presented in this thesis were performed by me.

3 Standard Model of particle physics

3.1 The Standard Model

The Standard Model (SM) of Particle Physics [10] describes the construction of the universe by uniting the fundamental building blocks called matter particles, which are the fermions (spin $\frac{1}{2}$) and the interactions among them through three of the four known fundamental forces mediated by carrier particles called gauge bosons (spin 1) as given in Figure 3.1.

The fermions comprise quarks (q) and leptons (l). Quarks occur in six different types based on their flavours: up (u), down (d), charm (c), strange (s), bottom (b), and top (t). The up-type quarks: up (u), charm (c) and top (t) carry an electric charge of $\frac{2}{3}e$ and the down-type quarks: down (d), strange (s) and bottom (b) carry a charge of $-\frac{1}{3}e$ where e is the electron charge. Leptons also occur in six flavours: electron (e), muon (μ), tau (τ) and their associated neutrinos: electron neutrino (ν_e), muon neutrino (ν_μ), tau neutrino (ν_τ). The charged leptons carry an electric charge of $-1e$, whereas neutrinos are chargeless.

Quarks and leptons are grouped into three generations. The second- and third-generation particles are heavier copies of the first-generation particles, with identical gauge quantum numbers but higher masses. An illustration of this characterization is given in Figure 3.1, which shows the masses, spins, and charges of the particles. All the fermions described have their own antiparticles, with the same mass but opposite charge-like quantum numbers.

The gauge bosons mediate interactions between the particles through the three fundamental forces: electromagnetic (EM), weak nuclear, and strong nuclear. The fourth fundamental force, the gravitational force, is not described in the SM as it is negligible for low mass particles as in the SM.

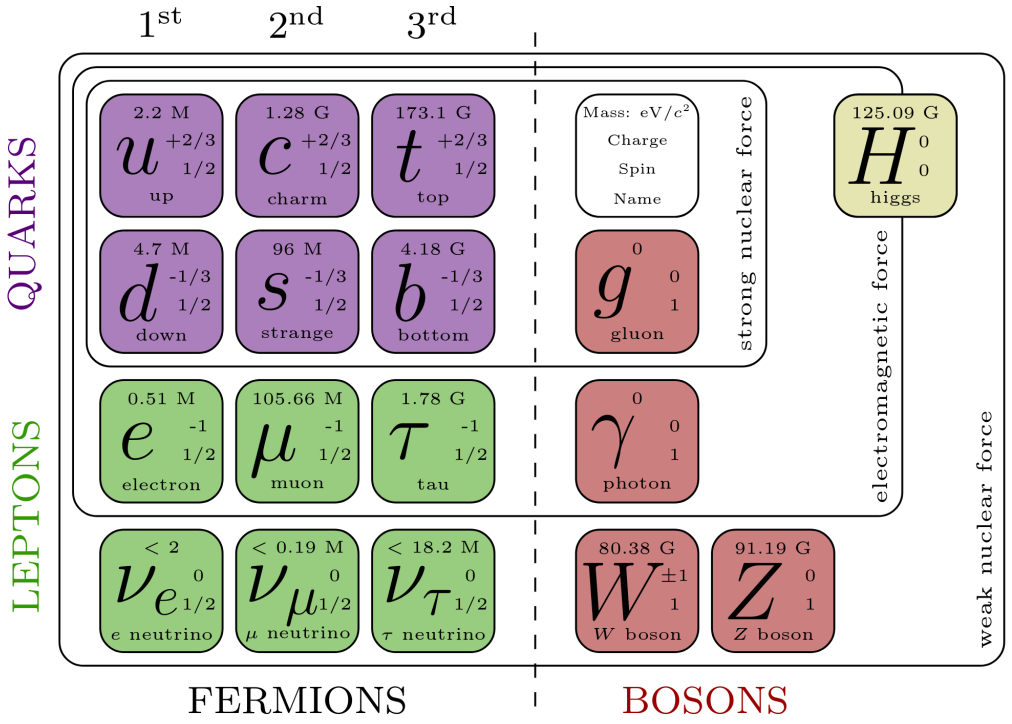


Figure 3.1: Particle content of the SM. Quarks are portrayed in purple and leptons in green. They make up the matter in the universe. The gauge bosons in red are mediators of the fundamental forces. The Higgs boson, contrary to quarks, leptons, and gauge bosons, arises from the Higgs mechanism as explained in the Sec. 3.2, which is responsible for giving mass to other particles. Figure taken from Ref. [11].

- The electromagnetic force acts between the charged particles through the absorption or emission of virtual photons, described by a quantum gauge theory, Quantum Electrodynamics (QED). The range of this force is infinite, as the photons are massless.
- The weak interaction gauge theory is mediated by the massive $W^{\pm}$ and Z^0 bosons ($M_W \approx 80$ GeV, $M_Z \approx 91$ GeV). Charged Current (CC) processes, mediated by $W^{\pm}$, change flavor quark according to the Cabibbo–Kobayashi–Maskaw (CKM) matrix (and lepton flavor in the $\ell - \nu$ sector). Neutral Current (NC) processes, mediated by Z^0, are flavor-diagonal at tree level (flavor-changing neutral currents arise only via higher order loops and are highly suppressed). Because the mediators are heavy, the weak force has a very short range, $\sim 10^{-3}$fm.
- Strong nuclear interactions act between the quarks, which are mediated by eight gluons. Gluons are massless and carry an effective color charge, as predicted by the gauge theory of Quantum Chromodynamics (QCD).

The SM explains the strong and electroweak interactions. The electroweak theory unifies the electromagnetic and weak interactions within a single framework, while QCD describes the

strong interaction of particles. So the three forces- electromagnetic, weak nuclear, and strong nuclear forces are incorporated into a single model.

The whole model is formulated by assigning a set of local gauge symmetries to the system, which governs the allowed interactions between particles. These symmetries, together with the principles of quantum mechanics and special relativity, are realized within the framework of Quantum Field Theory (QFT). In this formulation, the fundamental object is the Lagrangian density $\mathcal{L}$, which governs the theory's dynamics and kinematics. The Lagrangian is constructed to be invariant under the specified gauge symmetries, and from it the equations of motion can be derived via the Euler-Lagrangian formalism. Upon quantization, the excitations of the underlying quantum fields correspond to the particles observed in nature, with interactions governed by the terms of the Lagrangian. For a detailed introduction to the quantum field theoretic formulation, see e.g. Ref [12].

3.1.1 The Standard Model as a gauge theory

The Standard Model (SM) of particle physics is a QFT that successfully describes the electromagnetic, weak, and strong interactions among elementary particles. It is built on the principle of local gauge invariance, with the symmetry group structure

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y.$$

Each factor in this product corresponds to one of the fundamental forces: $SU(3)_C$ for the strong interaction (Quantum Chromodynamics, QCD), $SU(2)_L$ and $U(1)_Y$ for the electroweak interaction.

Gauge Symmetry and Local Invariance

At the heart of the SM lies the concept of gauge symmetry, where the fields transform under certain symmetry groups without altering physical observables. A classic example begins with the Dirac Lagrangian for a free spin-$\frac{1}{2}$ field:

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

This Lagrangian is invariant under global U(1) phase transformations:

$$\psi \to e^{i\alpha}\psi, \quad \bar{\psi} \to e^{-i\alpha}\bar{\psi}.$$

To promote this to a local gauge invariance, where the phase, $\alpha = \alpha(x)$, we introduce a gauge field A_μ and replace the derivative with a covariant derivative:

$$D_\mu = \partial_\mu - ieQA_\mu,$$

where e is the universal coupling constant of the $U(1)$ interaction, and Q is the generator of the group (the electric charge of the fermion in units of e). The gauge field transforms as

$$A_\mu \to A_\mu + \frac{1}{e}\partial_\mu \alpha(x),$$

which corresponds to the coupling of the fermion current to the gauge field. Thus, the demand for local invariance naturally introduces the photon field and leads to QED, the U(1) gauge theory of electromagnetism.

Electroweak and Strong Interactions

While the U(1) gauge describes electromagnetism, it is insufficient to explain weak and strong nuclear forces. Experimental observations indicate that weak interactions couple only to left-handed fermions, which motivates the introduction of the $\mathrm{SU}(2)_L$ gauge group. Left-handed fermions are organized into doublets, e.g.:

$$\begin{bmatrix} \nu_{eL} \\ e_L \end{bmatrix}, \quad \begin{bmatrix} u_L \\ d_L \end{bmatrix},$$

where the upper and lower components correspond to weak isospin $I = \frac{1}{2}$ states with third components $I_3 = \pm\frac{1}{2}$. Right-handed fermions do not participate in weak interactions and are singlets under $\mathrm{SU}(2)_L$ ($I = 0$).

To include electromagnetism consistently, an additional $\mathrm{U}(1)_Y$ gauge group associated with weak hypercharge, Y is introduced, resulting in the unified electroweak theory:

$$SU(2)_L \times U(1)_Y. \tag{3.1}$$

The Gell-Mann then gives the electric charge operator–Nishijima relation:

$$Q = I_3 + \frac{Y}{2}, \tag{3.2}$$

where I_3 is the third component of weak isospin, and Y is the weak hypercharge.

The gauge bosons associated with $\mathrm{SU}(2)_L$ and $\mathrm{U}(1)_Y$ mix to give the physical $W^\pm$, Z^0, and the photon (γ). Unlike the massless photon, the W and Z bosons acquire mass via the Higgs mechanism (see Section 3.2), thereby explaining the short range of the weak interaction.

Quantum Chromodynamics (QCD)

The strong interaction is described by the $\mathrm{SU}(3)_C$ gauge group, where quarks carry a charge-like quantum number called color (red, green, blue). QCD is a non-Abelian gauge theory with eight gluons as gauge bosons, corresponding to the eight generators of SU(3). The quark fields transform as triplets under SU(3), and the theory is invariant under local $\mathrm{SU}(3)_C$ transformations.

The SM is thus a renormalizable gauge theory describing all known fundamental interactions except gravity. These symmetries, together with the Higgs mechanism and spontaneous symmetry breaking, explain the mass spectrum and interaction structure of fermions and gauge bosons observed in high-energy experiments.

3.2 Electroweak symmetry breaking

At high energies (above ~ 100 GeV), the electromagnetic and weak interactions are unified under a single electroweak gauge group, given by Eq. 3.1 and 3.2. In the Standard Model, however, gauge invariance forbids the presence of explicit mass terms for fermions and gauge bosons in the Lagrangian. This presents an apparent contradiction. Experimentally, the weak gauge bosons ($W^{\pm}$ and Z^0), leptons, and quarks are massive, while the photon remains massless. The resolution is provided by spontaneous symmetry breaking (SSB) through the Higgs mechanism, which generates masses for fermions and gauge bosons while preserving the underlying gauge invariance.

The Higgs Field and Potential

The Higgs field Φ is introduced as a complex scalar $SU(2)_L$ doublet with four real degrees of freedom,

$$\Phi = \begin{bmatrix} \phi^+ \\ \phi^0 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{bmatrix}.$$

The dynamics of this field are governed by the Higgs potential,

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2. \tag{3.3}$$

where μ^2 and λ are free parameters of the theory.

For $\mu^2 < 0$ and $\lambda > 0$, the potential takes the well-known 'Mexican hat' form with a continuum of degenerate minima, as illustrated in the Fig. 3.2. The Higgs field acquires a non-zero vacuum expectation value (VEV), spontaneously breaking the electroweak symmetry,

$$\langle \Phi \rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ v \end{bmatrix}, \quad \text{where} \quad v = \sqrt{-\mu^2/\lambda}.$$

This choice of vacuum breaks the $SU(2)_L \times U(1)_Y$ symmetry down to the electromagnetic subgroup $U(1)_{EM}$.

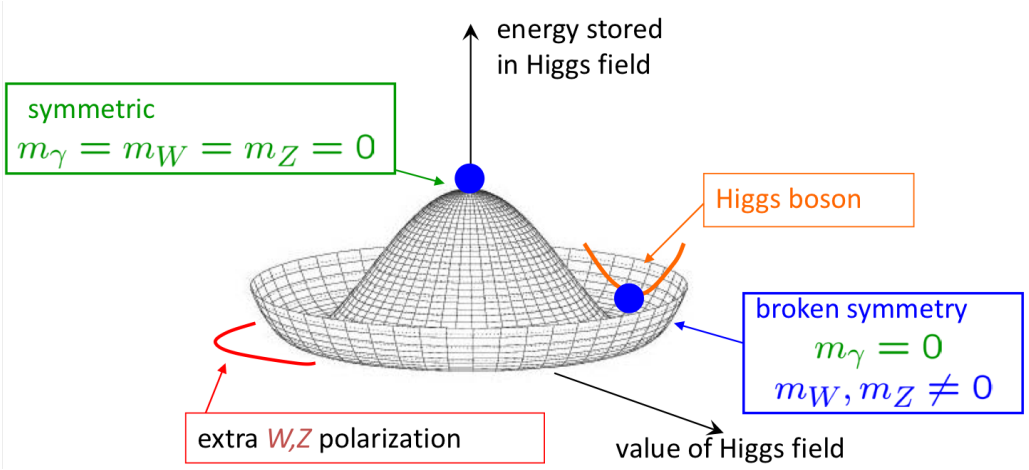


Figure 3.2: Higgs potential and spontaneous symmetry breaking in the SM, illustrating the Higgs potential, showing the energy stored in the Higgs field as a function of its value. At the symmetric point (top of the potential), the photon (M_γ), W, and Z bosons are massless ($M_\gamma = M_W = M_Z = 0$). After spontaneous symmetry breaking, the Higgs field acquires a nonzero vacuum expectation value (VEV), resulting in massive W and Z bosons ($M_W, M_Z \neq 0$) while the photon remains massless, as explained in Section 3.2. The diagram also depicts the Higgs boson and the additional W, Z polarization degrees of freedom generated during this process. Adapted from Ref. [13].

Mass Generation via the Higgs Mechanism

The breaking of the symmetry gives rise to one massive scalar particle, the Higgs boson, and 'absorbs' three would-be massless Goldstone bosons, which become the longitudinal components of the now massive vector bosons $W^\pm$ and Z^0. The photon γ remains massless, as it is associated with the unbroken $U(1)_{EM}$ symmetry.

The electroweak gauge bosons originate from the $SU(2)_L$ triplet (W^1, W^2, W^3) and the $U(1)_Y$ boson B, and after symmetry breaking, mix to form the physical fields:

$$\begin{aligned} W_\mu^\pm &= \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2), \\ Z_\mu &= \cos\theta_W W_\mu^3 - \sin\theta_W B_\mu, \\ A_\mu &= \sin\theta_W W_\mu^3 + \cos\theta_W B_\mu, \end{aligned}$$

where θ_W is the Weinberg angle.

The boson masses are then given by:

$$m_W^2 = \frac{1}{4}g^2v^2, \quad m_Z^2 = \frac{1}{4}(g^2 + g'^2)v^2, \quad m_\gamma = 0.$$

Fermion Masses and Yukawa Couplings

In the SM, fermions acquire mass through Yukawa interactions with the Higgs field. The generic term in the Lagrangian takes the form:

$$\mathcal{L}_{\text{Yukawa}} = -y_f \bar{\psi}_L \Phi \psi_R + \text{h.c.},$$

where y_f is the Yukawa coupling for fermion f, and $\psi_{L/R}$ are the left- and right-handed components. After electroweak symmetry breaking, this results in fermion masses:

$$m_f = y_f v.$$

Thus, the Higgs field is responsible for the masses of all charged fermions in the SM. Neutrino masses are not explained within this minimal framework and require extensions beyond the SM. In the general sense, the Yukawa couplings are not simply numbers but complex matrices in flavor space. Diagonalizing these matrices yields the physical fermion masses. For quarks, the misalignment between the mass eigenstates of the up and down-type quarks leads to the Cabibbo–Kobayashi–Maskaw (CKM) matrix, which encodes quark flavor mixing and CP violation.

The Higgs Boson

The Higgs boson, a neutral scalar particle predicted by this mechanism, was discovered in 2012 at the LHC with a measured mass of approximately 125 GeV [1, 2]. It is a fundamental part of the SM and a unique probe of the electroweak symmetry-breaking mechanism. Its interactions and decay channels provide critical insight into the mass generation of other particles and possible physics beyond the Standard Model.

3.3 Top quark physics

The top quark is the heaviest fundamental particle in the Standard Model (SM), with a mass of approximately 173 GeV [14]. It is the isospin partner of the bottom quark and belongs to the third generation of fermions. The existence of the top quark was predicted by Kobayashi and Maskawa in 1973 to account for CP violation in weak interactions and was confirmed experimentally in 1995 by the CDF and DØ collaborations at the Tevatron [3, 4].

The top quark carries an electric charge of $+\frac{2}{3}e$ and participates in all four fundamental interactions. However, due to its extremely short lifetime ($\sim 10^{-25}$ s), it decays before hadronization. Its dominant decay mode is via the weak interaction to a W boson and a down-type quark, with the largest branching ratio being to a bottom quark:

$$t \to W^+ + b \quad (\sim 99.8\%)$$

The relative decay probabilities to different quarks are governed by the elements of the CKM matrix, which encodes quark flavor mixing and CP violation in the SM, and is experimentally determined to be approximately [15]:

$$|V_{CKM}| = \begin{pmatrix} 0.974 & 0.225 & 0.004 \\ 0.225 & 0.974 & 0.042 \\ 0.009 & 0.041 & 0.999 \end{pmatrix}$$

Given the high top mass, the top quark couples most strongly to the Higgs field via the Yukawa interaction, with a coupling strength k_t close to unity, as shown in Fig. 3.3. It illustrates the reduced Higgs boson coupling strength modifiers for various fermions and vector bosons, including the top quark. The diagram provides further insight into how the top quark's coupling compares to other particles, and how its interaction with the Higgs field is critical for probing new physics beyond the SM [16].

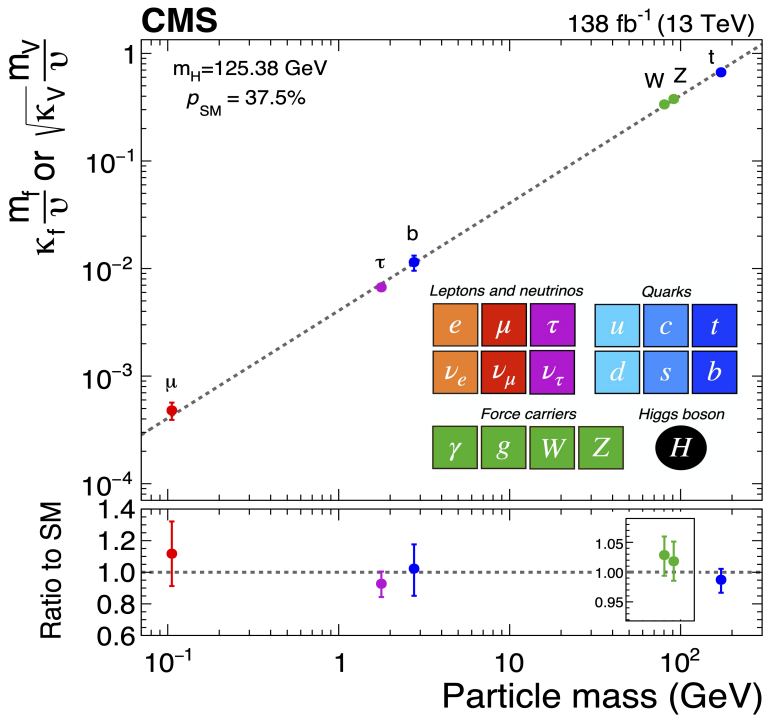


Figure 3.3: Measured Higgs boson coupling strength modifiers and their uncertainties for various fermions and vector bosons, as a function of the particle mass. For fermions, the couplings are given by $\kappa_F = \frac{m_F}{v}$, where m_F denotes the fermion mass and v is the vacuum expectation value of the Higgs field. For vector bosons, the square root of the modifier is shown, $\sqrt{\kappa_V} = \frac{m_V}{v}$, to preserve the linear proportionality with the boson mass as predicted in the SM. The plot highlights the top quark's particularly strong coupling to the Higgs boson. Taken from Ref. [16].

3.3.1 Top Quark Production at the LHC

At the LHC, top quarks are produced mainly via two SM mechanisms:

1. Top Quark–Antiquark Pair Production ($t\bar{t}$)

The dominant mode of $t\bar{t}$ production at the LHC is through the strong interaction, primarily via gluon–gluon fusion (gg), which accounts for approximately 83% of the production at the LHC at $\sqrt{13}$ TeV. The subdominant contribution originates from quark–antiquark annihilation ($q\bar{q}$), accounting for approximately 17%. The corresponding leading-order (LO) Feynman diagrams, which include s-, t-, and u-channel gluon exchange, are illustrated in Figure 3.4. In contrast, at $p\bar{p}$ colliders like the Tevatron [17], the dominant production mode was $q\bar{q}$ annihilation due to the presence of valence antiquarks in the antiproton.

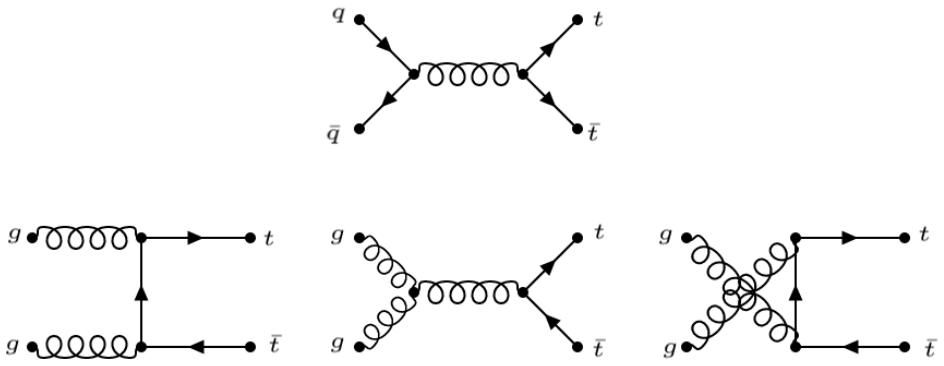


Figure 3.4: Leading-order Feynman diagrams for $t\bar{t}$ production: $q\bar{q}$ annihilation (top image), gg fusion in the t-channel (bottom first), gg fusion in the s-channel (bottom second), gg fusion in the u-channel (bottom third)

A summary of recent measurements of the inclusive $t\bar{t}$ production cross sections at various center-of-mass energies at the LHC is presented in Figure 3.5. These measurements are sensitive to fundamental parameters like the top mass, strong coupling constant α_s, and also the momentum distribution of gluons inside the proton [18]. The analysis strategies have evolved significantly over the past decade, moving from simple cut-based approaches to advanced likelihood fits that exploit b-jet multiplicity information to constrain systematics. More recently, machine learning techniques have been incorporated to enhance signal-to-background discrimination and further improve measurement precision.

2. Single Top Quark Production (t)

Single top quark production proceeds via the electroweak interaction [20]. There are three primary production mechanisms for single top quarks, mediated by the virtual exchange of a W boson. Firstly, the t-channel, which is the dominant at the LHC and accounts for approximately 70% of the cross section, involves W-gluon fusion. Secondly, the tW-channel,

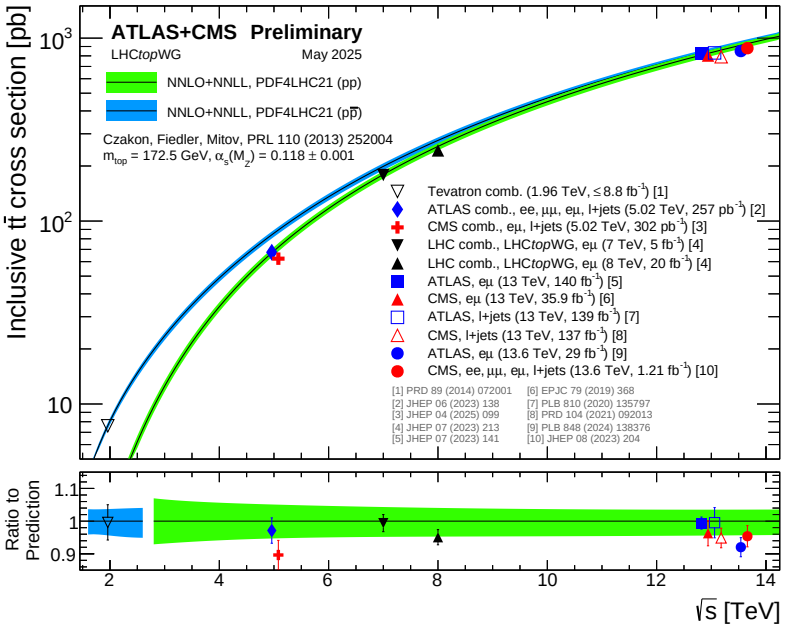


Figure 3.5: Inclusive $t\bar{t}$ production cross section measurements at different center-of-mass energies by ATLAS, CMS, and the Tevatron experiments. The results are compared to NNLO+NNLL predictions using various PDF sets. The theory band reflects uncertainties from renormalization and factorization scales, parton distribution functions, and the strong coupling constant. Measurements at identical center-of-mass energies are slightly offset for clarity. A recent measurement at 13.6 TeV is also included. The shaded bands represent theoretical uncertainties. Taken from Refs. [18, 19].

which is the associated production with a real W boson, which contributes around 25%, and the s-channel production, which makes up about 5%, is the W-mediated quark–antiquark annihilation, sensitive to new physics such as W' bosons or charged Higgs bosons. These processes are shown in Figure 3.6.

In the t-channel production mode, a virtual space-like W boson mediates the interaction. An initial-state gluon, originating from the sea quarks inside the proton, transforms into a $b\bar{b}$ pair. One of the b-quarks then interacts with the W boson to produce the top quark. This process is commonly referred to as W-gluon fusion, as illustrated in Figure 3.6 (a) and (b).

In the s-channel production mode, the interaction is mediated by a virtual time-like W boson. The W boson is produced through the fusion of two quarks ($q\bar{q}$), leading to the production of a single top quark and a b-quark. Despite its relatively low statistics, this channel is crucial for probing new physics, such as searches for charged Higgs bosons and W bosons. The s-channel process is shown in Figure 3.6 (c).

The tW-channel involves the production of a single top quark in association with a real W boson. In this channel, a b-quark from the sea quark inside the proton fuses with a gluon to produce this pair. This process is depicted in Figure 3.6 (d).

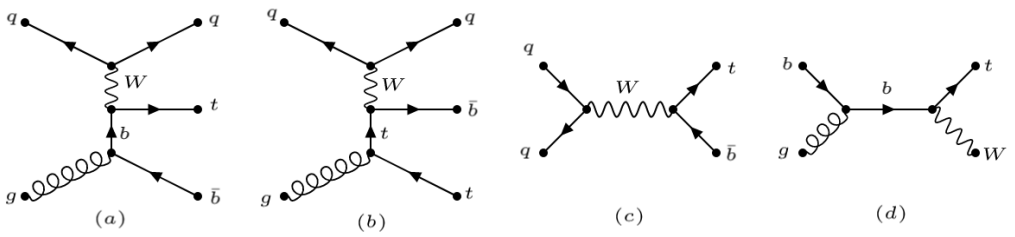


Figure 3.6: Feynman diagrams for single top quark production: (a) and (b): t-channel, (c): s-channel, and (d): tW-channel.

A comprehensive comparison of the measured inclusive cross sections for the three single top-quark production channels at $\sqrt{s} = 7$, 8, and 13 TeV is shown in Figure 3.7. These measurements, performed by both ATLAS and CMS collaborations, are compared with theoretical calculations at NNLO and NLO+NNLL precision [19]. Among the channels, t-channel production has the largest cross section, followed by tW, with the s-channel being significantly suppressed at LHC energies.

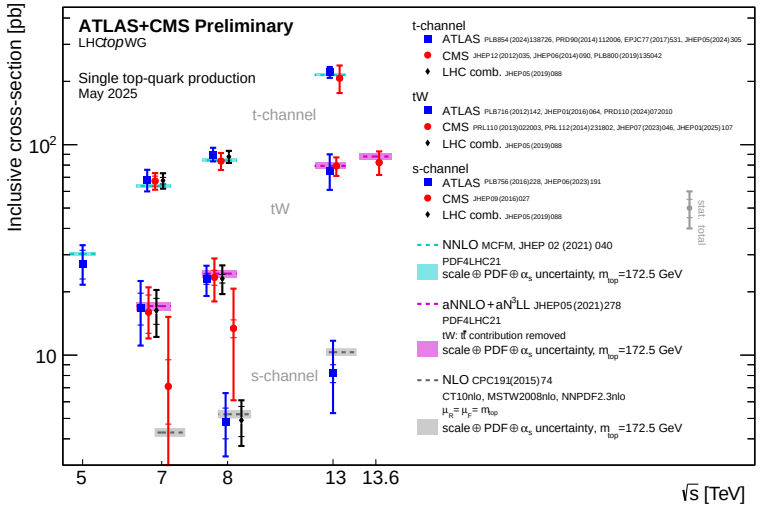


Figure 3.7: Inclusive cross-section measurements for single top-quark production in the t-channel, tW channel, and s-channel as a function of the center-of-mass energy, $\sqrt{s} = 7$, 8, 13, and 13.6 TeV by ATLAS and CMS experiments. This includes the most recent result at $\sqrt{s} = 13.6$ TeV. Theoretical predictions at NNLO and NLO+NNLL accuracies are also shown, with scale and PDF uncertainties. The measurements show good agreement with the Standard Model predictions across all channels and energies [19].

3.3.2 Top Quark Pair Decay

Each top quark decays almost exclusively into a W boson and a b-quark. Thus, a $t\bar{t}$ event yields two W bosons and two b-quarks. This decay is mediated by the charged current of the weak interaction, with the top-quark weak decays and their subsequent decay products serving as sensitive probes of the SM and beyond. The W boson itself can decay in two main ways: hadronically (into a quark-antiquark pair) or leptonically (into a charged lepton and a neutrino). The branching ratios for these modes are approximately 67% for the hadronic decay ($W \to q\bar{q}'$) and 33% for the leptonic decay ($W \to \ell\nu$), with ℓ representing an electron or muon, or tau, and ν the corresponding neutrino. The latter is shared approximately equally among the three lepton flavors (e, μ, τ) with $\approx 11\%$. Electrons and muons are directly reconstructed as isolated leptons in the detector and therefore provide clean experimental signatures. In contrast, the τ lepton has a much more complex role: it decays further, either leptonically (to e or μ and neutrinos, $\sim 35\%$ of cases) or hadronically (to hadrons and a neutrino, $\sim 65\%$). As a result, τ decays can mimic both purely leptonic and hadronic final states and must be treated carefully in analyses.

This results in three main categories of final states in $t\bar{t}$:

- Dileptonic: Both W bosons decay to leptons and neutrinos ($\ell\ell + \nu\nu + b\bar{b}$), resulting in two charged leptons and missing transverse energy. This channel, although offering the clearest signature, has a low branching fraction of about 11%, leading to significant statistical uncertainties.
- Semileptonic: One W decays leptonically, the other hadronically ($\ell\nu + q\bar{q}' + b\bar{b}$), producing one charged lepton, missing transverse energy, and four jets, two of which are b-jets from the top quarks. This channel has a branching fraction of about 44% and is characterized by a clear signature involving the lepton, which helps distinguish it from background processes.
- Hadronic: Both W bosons decay hadronically, ($q\bar{q}'q\bar{q}'' + b\bar{b}$) producing four jets. This channel has the highest branching fraction, approximately 45%, but is plagued by a significant QCD background.

In this analysis, we consider the lepton to be an isolated electron or a muon. Each decay channel has advantages and challenges, with the semileptonic final state offering a balance between a good signal-to-background ratio and a branching fraction. In this thesis, the semileptonic final state is the primary focus due to its favorable balance between signal clarity and statistical significance, as discussed in Section 3.3.3.

3.3.3 Semileptonic $t\bar{t}$ Decay

The semileptonic decay mode is characterized by one charged lepton (either an electron or a muon in this analysis), missing transverse energy from the associated neutrino, and four jets, two of which originate from b-quarks, as illustrated in the leading-order (LO) Feynman diagram in Fig. 3.8. The jets from the W boson decay are mostly light-flavored (LF), while

the jets from the top quark decay are heavy-flavored (HF) b-jets. This final state is important for its relatively high branching ratio compared to the dileptonic channel, as well as for the clean lepton signature, which is easier to identify in the detector than in hadronic decays. The presence of the charged lepton significantly enhances the ability to separate signal events from backgrounds, particularly from other processes such as QCD jets.

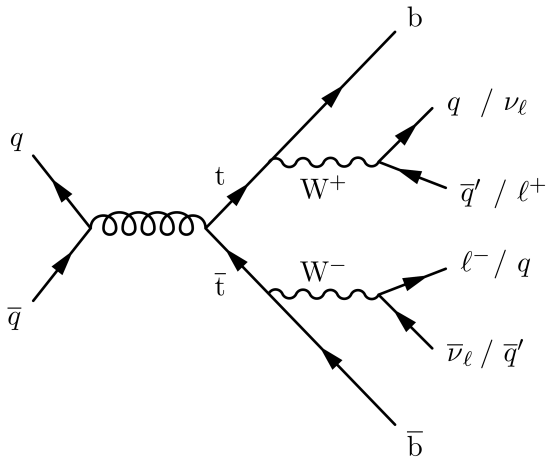


Figure 3.8: LO Feynman diagram representing the semileptonic decay channel of a $t\bar{t}$ pair produced via quark-antiquark annihilation ($q\bar{q} \rightarrow t\bar{t}$). One top quark decays hadronically ($t \rightarrow bq\bar{q}'$), while the other decays leptonically ($t \rightarrow b\ell\nu_\ell$), resulting in a final state with one charged lepton, missing transverse energy from the neutrino, and multiple jets, including two b-jets.

Several SM processes can mimic the $t\bar{t}$ semileptonic signature. These background processes often produce a similar final state, making it challenging to distinguish between signal and background. The leptons in the final state are classified as prompt leptons when they originate from the primary interaction vertex via electroweak processes, or as fake leptons, which arise from misidentifications, such as meson decays in jets, cosmic rays, unidentified leptons, or jets passing through muon chambers. One of the significant backgrounds to the channel comes from single top quark production, particularly where a single top quark is produced with a W boson. The W boson decays semileptonically, leading to a final state with a lepton, jets, and missing transverse energy, similar to the semileptonic signature. Another important background is W+jets production, where the W boson, produced in association with quarks or gluons, decays and produces a signature equivalent to the semileptonic $t\bar{t}$ channel. The gluons can produce heavy-flavor quarks, resulting in a $Wb\bar{b}$ combination that produces a lepton, two b-jets, and missing transverse energy. The Z+jets background is also a significant contributor, where the Z boson, produced in association with quarks or gluons, can decay into a $Zb\bar{b}$ combination. This process can lead to a semileptonic signature due to reconstruction errors, such as missing a lepton from the Z decay or misidentifying a jet as a lepton. Additionally, associated vector boson production ($t\bar{t}$ +V), where a W or Z boson is produced alongside a $t\bar{t}$ pair, can mimic a

semileptonic decay signature when the W or Z boson decays hadronically into light flavor jets. Diboson production (VV), including processes like WW, WZ, and ZZ, also contributes to the background. In the WW process, one W boson decays semileptonically while the other decays hadronically. In the WZ process, the W boson decays semileptonically, and the Z boson decays into heavy-flavor quark jets. Similarly, in the ZZ process, one Z boson decays leptonically, and the other decays into heavy-flavor jets, potentially resulting in a semi-leptonic final state due to a lepton being missed in the reconstruction. Finally, QCD background processes, where quarks or gluons in the detector produce a $b\bar{b}$ pair can also contribute to a semileptonic final state. These background processes need to be carefully accounted for and modeled in the analysis, as their similar signatures can lead to significant contamination of the signal region.

4 The CMS experiment at the LHC

This chapter provides the experimental context for the measurements presented in this thesis. It begins with a brief overview of the Large Hadron Collider (LHC), which delivered the $\sqrt{s} = 13$ TeV Run-II dataset (2016–2018) that underpins the analysis here, and has entered Run-III at $\sqrt{s} = 13.6$ TeV since 2022. The discussion in Section 4.1 encompasses the accelerator chain, running conditions, and the roles of luminosity and pileup in determining statistical power and reconstruction complexity, with an emphasis on how these quantities influence event yields and uncertainties. The Compact Muon Solenoid (CMS) detector is then introduced in Sec. 4.2, which is a nearly hermetic, general-purpose apparatus built around a 3.8 T superconducting solenoid, optimized for precise reconstruction of stable particles produced in collisions. It includes the coordinate system, kinematic conventions, and the detector subsystems most relevant for analyses involving multiple jets and at least one lepton. Finally, the CMS two-level trigger architecture is described in Sec. 4.3, outlining its role in selecting events of interest while reducing the high-frequency bunch-crossing rate to an analyzable data stream suitable for offline analysis. The material in this chapter establishes the detector performance assumptions, object definitions, and operational constraints that inform the selection, reconstruction, and systematic treatment used in subsequent chapters.

4.1 Large Hadron Collider (LHC)

The Large Hadron Collider (LHC) [21] at CERN is the world's largest and most powerful particle accelerator. It consists of a 27 km circumference ring of superconducting magnets installed roughly 175 m underground on the Franco–Swiss border near Geneva. It collides protons for most of the year and, during dedicated periods, also runs proton-lead and lead-lead heavy ion programs. Four interaction points host the principal detectors: ATLAS [22], CMS [23], ALICE [24], and LHCb[25], which are augmented by forward and specialized experiments like TOTEM [26], LHCf [27], MoEDAL [28], and FASER [29] that target complementary physics. A schematic top view of the LHC ring and its four main experimental points is shown in

Fig. 4.1. The LHC is designed for a proton–proton center-of-mass energy of $\sqrt{s} = 14$ TeV at an instantaneous luminosity up to $10^{34}\text{cm}^{-2}\text{s}^{-1}$. While luminosities above this design value have already been achieved in operation, collisions at the design value of $\sqrt{s}$ have not yet been reached. During Run I, the machine delivered $\sqrt{s} = 7$ TeV (2010-2011), and $\sqrt{s} = 8$ TeV (2012), enabling the Higgs boson discovery [1, 2]. Run II (2015-2018) operated at $\sqrt{s} = 13$ TeV (6.5 TeV per beam), with the 2016–2018 dataset serving as the basis for the analyses in this thesis. Since 2022, the start of Run III, the LHC has been running at $\sqrt{s} = 13.6$ TeV. Further details on the running conditions can be found in [21].

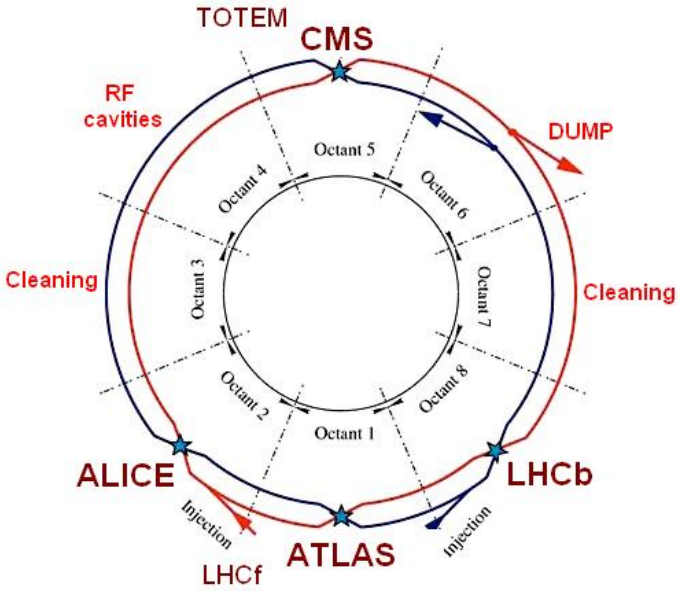


Figure 4.1: Schematic layout of the LHC, showing the 27 km ring divided into eight octants, highlighting the four main interaction points (ATLAS, CMS, ALICE, LHCb). Adapted from Ref. [21]

Beams are prepared by CERN's accelerator complex (Fig. 4.2) before being injected into the LHC. Protons, obtained by stripping electrons from hydrogen by ionization, are accelerated in LINAC2 to 0.12 GeV, boosted in the Proton Synchrotron (PS) Booster to 1.4 GeV, then in the PS to 26 GeV, and finally in the Super Proton Synchrotron (SPS) to 450 GeV, before entering the LHC where superconducting dipoles and radio-frequency cavities guide and accelerate them to multi-TeV energies. Around the ring, the general-purpose experiments: ATLAS and CMS pursue a broad program from precision Standard Model measurements to searches for new phenomena; ALICE focuses on the quark–gluon plasma in heavy-ion collisions; and LHCb specializes in heavy-flavor physics and CP-violation. While ATLAS and CMS share similar physics goals, their detector designs differ, most notably in magnet configurations, leading to complementary performance. For example, CMS emphasizes excellent momentum resolution, while ATLAS features particularly strong hadronic calorimetry. Further details on CMS are provided in Sec. 4.2.

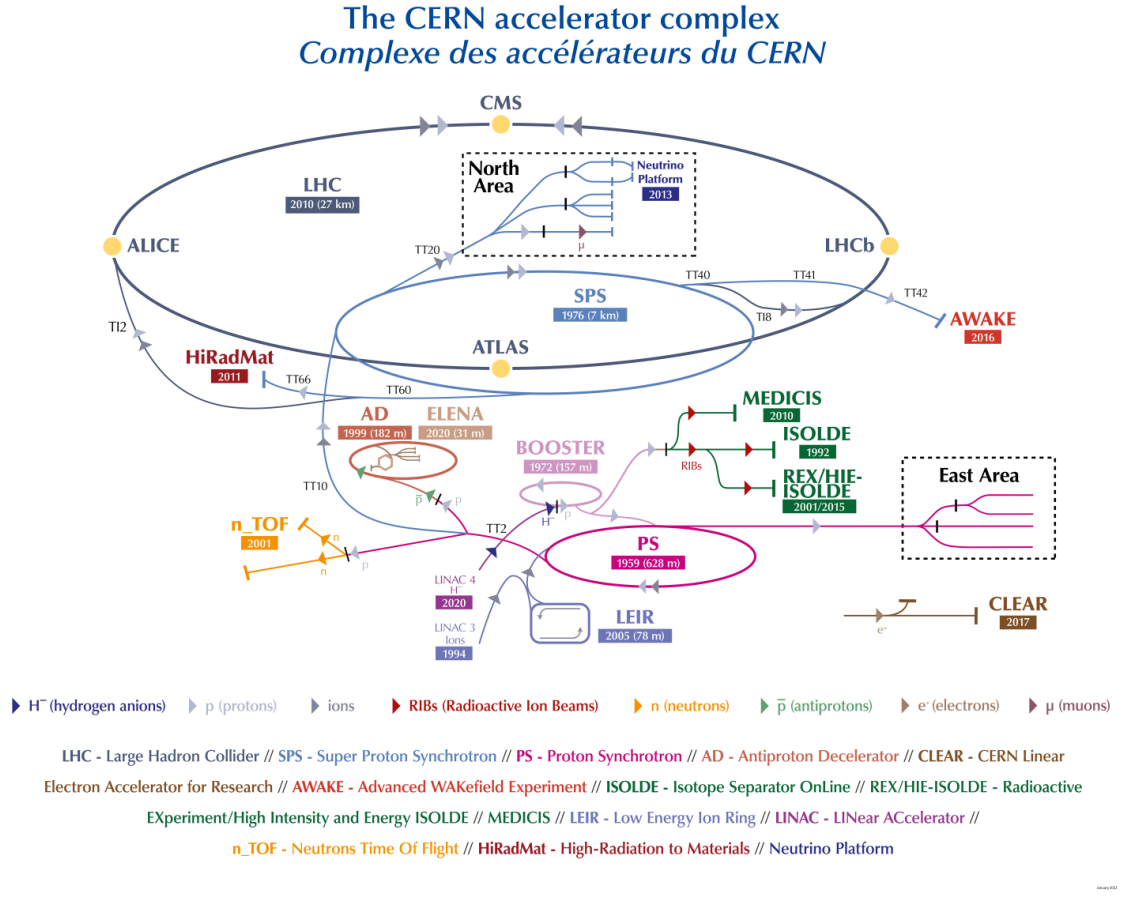


Figure 4.2: Schematic layout of the CERN accelerator complex. Adapted from Ref. [30].

4.1.1 Luminosity

While the center-of-mass energy defines the energy reach of the LHC, the amount of usable data for physics analyses is governed by its luminosity. Luminosity quantifies the collision rate and is therefore directly related to the statistical power of measurements at the experiments. The event yield for a given physics process at a collider is governed by the luminosity $\mathcal{L}$ and the process cross section σ,

$$N = \mathcal{L}_{\text{int}}\,\sigma\,\varepsilon, \tag{4.1}$$

where $\mathcal{L}_{\text{int}} \equiv \int \mathcal{L}(t)\,dt$ is the time-integrated luminosity and ε denotes the total selection efficiency and acceptance. Luminosity has units of $\text{cm}^{-2}\text{s}^{-1}$, integrated luminosity is typically quoted in inverse femtobarns (fb^{-1}), with $1\ \text{fb}^{-1} = 10^{39}\ \text{cm}^{-2}$.

For two colliding bunched beams, the (geometric) instantaneous luminosity can be expressed as

$$\mathcal{L} = \frac{f_{\text{rev}}\,n_b\,N_1 N_2}{4\pi\,\sigma_x \sigma_y} \tag{4.2}$$

where f_{rev} is the revolution frequency, n_b the number of colliding bunch pairs, $N_{1,2}$ the bunch populations, and $\sigma_{x,y}$ the transverse beam sizes at the interaction point. Operationally, higher luminosity is achieved by increasing n_b, $N_{1,2}$, and beam focusing (smaller $\sigma_{x,y}$) while controlling beam–beam effects and pileup.

The average number of inelastic interactions per bunch crossing (pileup) μ is related to the visible inelastic cross section $\sigma_{\text{inel}}^{\text{vis}}$ and the machine parameters by

$$\mu \;=\; \frac{\mathcal{L}\,\sigma_{\text{inel}}^{\text{vis}}}{n_b\,f_{\text{rev}}}, \tag{4.3}$$

illustrating that at fixed $\sigma_{\text{inel}}^{\text{vis}}$ and bunch structure, higher luminosity directly increases pileup, impacting reconstruction and analysis (see Sec. 5.2.4).

Experimentally, luminosity is measured with dedicated luminometers and online counting algorithms, normalized via dedicated beam–separation (van der Meer) scans that calibrate the visible interaction rate to an absolute scale (see Sec. 7.6.1). The calibration is repeated periodically and transferred to physics running. For LHC Runs II and III, the relative luminosity uncertainty is typically at the percent level and is propagated as a normalization uncertainty in cross-section measurements.

4.2 Compact Muon Solenoid (CMS) detector

The high-energy proton-proton (p-p) collisions delivered by the LHC can only be exploited for particle physics if the resulting particles are efficiently detected and measured. At the four main interaction points around the LHC ring, dedicated experiments have been built to serve this purpose. This thesis is based on data collected by the Compact Muon Solenoid (CMS) detector, situated at Interaction Point 5 (Fig. 4.1). CMS is one of the two general-purpose

experiments at the LHC, optimized to measure a wide range of final-state objects, including electrons, muons, photons, jets, and missing transverse momentum, across a broad kinematic range in the high-luminosity environment of the LHC. It is a large but compact apparatus (about 21.6 m in length, 15 m in diameter, and 14 kt total mass) built around a superconducting solenoid that provides a central magnetic field of approximately 3.8 T. CMS is operated by a broad international collaboration of over 5000 scientists from over 250 institutions in over 50 countries as of 2022 [31]. Its precise tracking, powerful solenoidal magnetic field, high-granularity calorimeters, and robust muon system make it particularly well suited for studies of top-quark production and associated processes such as $t\bar{t}$+X, which are the focus of this work.

CMS follows an "onion-like" layout around the interaction point: an all-silicon tracking system closest to the beam pipe, a high-granularity electromagnetic calorimeter (ECAL) made of scintillating crystals, a sampling hadronic calorimeter (HCAL), and, outside the solenoid and its steel return yoke, a redundant muon system. This nearly hermetic geometry ensures excellent momentum, energy, and angular resolutions over a wide kinematic range. A schematic cutaway drawing of the CMS detector, highlighting the main subdetectors and their geometric coverage, is presented in Fig. 4.3.

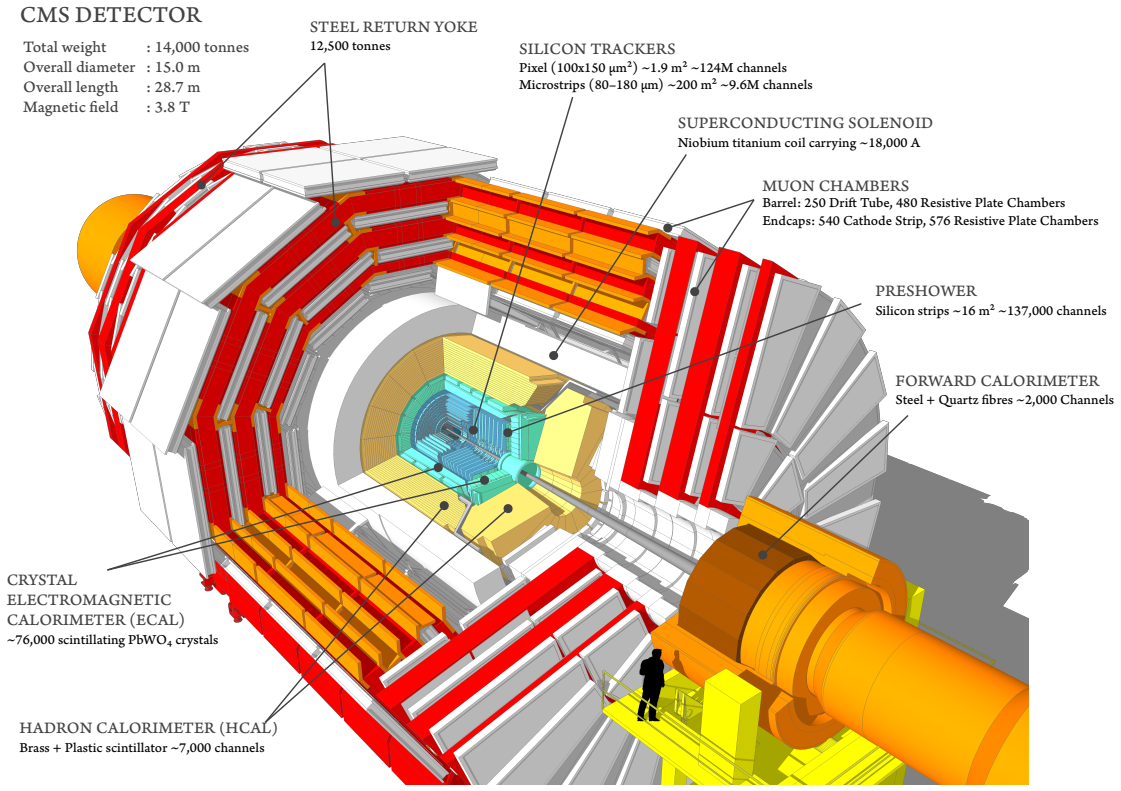


Figure 4.3: Schematic layout of the CMS detector. Taken from Ref. [32].

4.2.1 CMS coordinate system and kinematics

The CMS experiment adopts a right-handed coordinate system to describe the geometry of the detector and the kinematic properties of reconstructed particles. The origin is defined at the nominal interaction point at the center of the detector. The x-axis points radially inward toward the center of the LHC ring, the y-axis points vertically upward, and the z-axis coincides with the beam direction, pointing along the counterclockwise proton beam.

At the LHC, the colliding protons carry equal and opposite momenta along z, while the net initial transverse momentum is negligible. Because the longitudinal momentum of the partons participating in the hard scatter is a priori unknown, kinematic quantities defined in the plane transverse to the beam are preferred. The transverse momentum is

$$p_T \equiv \sqrt{p_x^2 + p_y^2}, \tag{4.4}$$

and is invariant under boosts along the beam axis. Pseudorapidity is preferred in hadron collider physics because differences in η are invariant under Lorentz boosts along the beam axis, making it a natural variable to describe the angular distribution of particles produced in high-energy collisions. Thus, for longitudinally boost-invariant angular coordinates, it is convenient to use rapidity,

$$y \equiv \frac{1}{2} \ln\left(\frac{E + p_z}{E - p_z}\right), \tag{4.5}$$

and, more commonly in reconstruction, the pseudorapidity,

$$\eta \equiv -\ln\left[\tan\left(\frac{\theta}{2}\right)\right], \tag{4.6}$$

which depends only on the polar angle θ. In the ultra-relativistic limit ($m \ll p$), one has $y \simeq \eta$, motivating the widespread use of η in detector segmentation, acceptance, and efficiency parametrizations. Distances on the detector surface are measured in the (η, ϕ) plane using

$$\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{4.7}$$

with the azimuthal difference defined modulo 2π and mapped to $(-\pi, \pi]$ as $\Delta\phi \leftarrow \text{atan2}(\sin(\phi_1 - \phi_2), \cos(\phi_1 - \phi_2))$. Moreover, particle production at the LHC tends to be approximately uniform in η, and the subdetectors of CMS are segmented in intervals of η, which simplifies acceptance and efficiency studies. The variable ΔR is rotationally invariant in ϕ and approximately boost-invariant along z, and therefore underpins jet clustering, isolation definitions, and object–object matching throughout this analysis.

Finally, the event-wide momentum imbalance in the transverse plane is quantified by the missing transverse momentum vector,

$$\vec{p}_T^{\,\text{miss}} \equiv -\sum_{\text{reco objects}} \vec{p}_{T,i} \tag{4.8}$$

whose magnitude is denoted $MET \equiv |\vec{p}_T^{\,\text{miss}}|$.

4.2.2 Beam pipe

The CMS beam pipe [33] encloses the LHC beams as they traverse the detector and intersect at the interaction point (IP), the nominal centre of the p-p collisions. It is a thin, low-mass structure designed to minimise secondary interactions before particles reach the innermost tracking layers. The LHC beams are organised in bunches containing $\mathcal{O}(10^{11})$ protons, with bunch crossings every 25 ns. At design luminosity, multiple p-p interactions can occur in the same bunch crossing, leading to an average of a few tens of collisions per crossing, as discussed in Sec. 4.1.1. This environment drives the requirements on the inner detector layout and the reconstruction algorithms used throughout this analysis.

4.2.3 Tracker system

The innermost CMS subdetectors reconstruct charged-particle trajectories and vertices in the 3.8 T solenoidal field using an all-silicon tracker [34]. Multiple concentric layers provide many space points per track, enabling robust pattern recognition and precise parameter estimation. The tracking sensors are reverse-biased silicon diodes: a charged particle traversing the depleted bulk creates electron–hole pairs whose collected charge forms a time-stamped hit. Fig. 4.4 depicts the tracker with its different subsystems.

Silicon pixel detector: Closest to the beam line sits the silicon pixel system [35], delivering three-dimensional space points with $\mathcal{O}(10\ \mu\text{m})$ intrinsic resolution. Pixels have typical cell sizes of about $100 \times 150\ \mu\text{m}^2$, enabling operation at high hit density and under significant radiation load. During Run-I and early Run-II (Phase-0), the barrel comprised three layers located roughly at radii of 44–102 mm, with coverage up to $|\eta| < 2.4$. Starting in 2017 (Phase-1), the pixel system was upgraded to four barrel layers extending the acceptance to $|\eta| < 2.5$, with the innermost layer moved inward to about 29 mm and the outer layer out to about 160 mm. The reduced first-layer radius and added granularity improve impact-parameter resolution and displaced-vertex reconstruction, which are key to heavy-flavor tagging and the reconstruction of short-lived hadrons.

Silicon strip tracker: Surrounding the pixel detector is the large-area microstrip tracker, providing a long lever arm and $\gtrsim 10$ additional space points per trajectory across the fiducial region. It is organized into barrel and endcap subsystems (Tracker Inner Barrel (TIB)/Tracker Inner Disks (TID), Tracker Outer Barrel (TOB), Tracker EndCaps (TEC$\pm$)) with ten cylindrical barrel layers and twelve endcap disks in total. Compared to pixels, the strip pitch is larger, trading per-hit granularity for extensive coverage and channel-count scalability where occupancies are lower. Together, the pixel+strip system delivers a relative transverse momentum resolution of approximately 1-2% for charged particles with $1 < p_T < 10$ GeV in the central region, and a transverse impact-parameter resolution at the level of $20 - 100\ \mu\text{m}$, depending on the data-taking era and detector configuration [35–37]. This performance enables efficient reconstruction of electrons, muons, hadrons, and displaced heavy-flavor decays throughout $|\eta| < 2.5$.

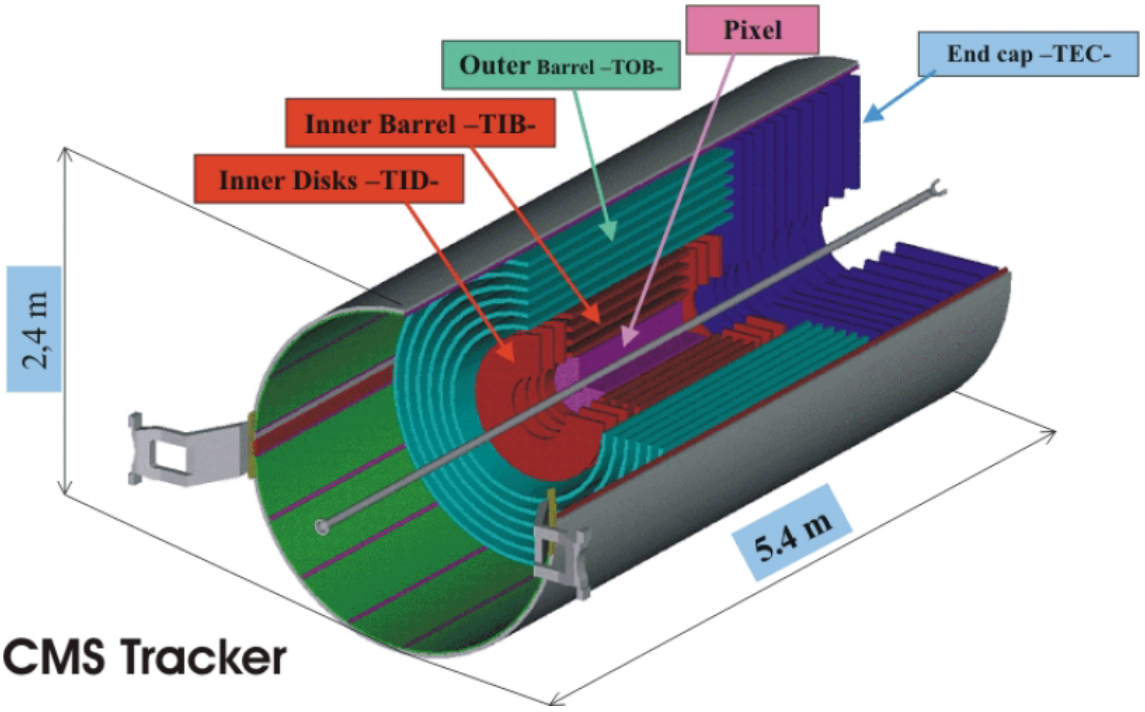


Figure 4.4: Schematic layout of the CMS tracker system. Taken from Ref. [38].

4.2.4 Calorimeters

The CMS calorimeter system measures the energies and directions of particles emerging from the pp interactions. It consists of a high-granularity homogeneous electromagnetic calorimeter (ECAL) surrounded by a sampling hadronic calorimeter (HCAL). Together with forward extensions, they provide near-hermetic coverage for precise photon/electron measurements, jet energy reconstruction, and the determination of missing transverse momentum.

Electromagnetic calorimeter (ECAL): The ECAL [39–41] surrounds the tracker and measures the energy of electrons and photons via compact electromagnetic (EM) showers initiated by bremsstrahlung and $\gamma \to e^+e^-$ pair production. It is a homogeneous crystal calorimeter built from lead–tungstate ($PbWO_4$), whose fast, radiation-hard scintillation yields a light signal approximately proportional to the incident EM energy. The system comprises a barrel (EB) and two endcaps (EE), instrumented up to $|\eta| \lesssim 3$, with a silicon-based preshower detector (ES) in front of the endcaps to initiate and sample early shower development and to help resolve close photon pairs (e.g. from $\pi^0 \to \gamma\gamma$).

The crystal lengths correspond to ∼25 radiation lengths, ensuring near-complete EM shower containment and excellent energy resolution. Typical transverse crystal sizes are a few centimeters (EB) and slightly larger in EE, matching the Molière radius to control lateral leakage. Photodiodes convert the scintillation light into electrical signals, which are then calibrated to a common energy scale. The electron momentum is reconstructed by combining the energy measured in the ECAL with the track momentum determined by the silicon tracker. For

electrons with p_T around 45 GeV from $Z \to e^+e^-$ decays, the resulting momentum resolution ranges from about 1.6% in the barrel to about 5% in the endcap regions, depending on the amount of bremsstrahlung emitted in the material upstream of the calorimeter [42, 43]. This performance is essential for precision electron measurements and for maintaining high trigger efficiency.

Hadronic calorimeter (HCAL): The HCAL [44, 45] measures the energies of hadrons and the hadronic component of jets, and together with the ECAL, provides the input for missing transverse momentum (MET). It is a sampling calorimeter composed of alternating absorber and active layers: brass (barrel/endcap) or steel (forward) absorbers interleaved with plastic scintillator tiles whose light is collected by wavelength-shifting fibers. Hadronic showers are longer and more heterogeneous than EM showers. Consequently, the HCAL is built with a large depth in units of nuclear interaction lengths, λ_I, to contain the extended and fluctuating longitudinal development of hadronic showers. In the central region ($|\eta| \lesssim 1.3$), additional sampling outside the solenoid in the Hadronic Outer (HO) increases 2 λ of material. The barrel (HB) and endcaps (HE) cover up to $|\eta| \lesssim 3$. In the forward region, the steel–quartz-fiber Hadronic Forward (HF) calorimeters extend the coverage to $|\eta| \approx 5.2$, providing a depth of approximately $10\lambda_I$, ensuring near-hermeticity for MET reconstruction.

Because of sampling fluctuations, invisible energy, and the presence of EM subshowers within hadronic cascades, the HCAL energy resolution is intrinsically worse than that of the ECAL. The combined ECAL+HCAL response is calibrated to a common particle-flow (PF) energy scale (detailed in Sec. 5.1.2) used throughout the analysis. When combining information, the energy resolution typically amounts to 15-20% at 30GeV , 10% at 100GeV , and 5% at 1TeV [46]. Fig. 4.5 depicts the layout of CMS calorimeters.

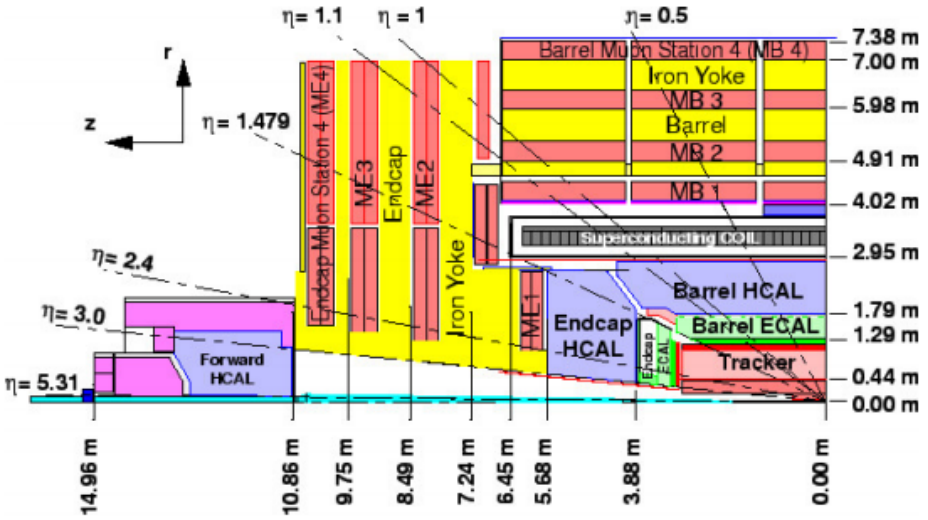


Figure 4.5: Longitudinal view of one quadrant of the CMS detector showing the ECAL barrel/endcaps, the HCAL barrel (HB), endcaps (HE), outer (HO), and the Hadronic Forward (HF) calorimeters, together with their approximate η coverage [45].

4.2.5 Superconducting solenoid magnet

A defining feature of CMS is its large superconducting solenoid that surrounds the tracking and calorimeter systems and provides a strong, nearly uniform axial magnetic field of 3.8 T [47]. The field bends the trajectories of charged particles in the transverse plane, enabling the precise determination of their momentum and charge sign through the curvature of the track,

$$r \propto \frac{p_T}{qB}, \tag{4.9}$$

where p_T is the transverse momentum, q the particle's electric charge, and B the magnetic flux density. The solenoid is wound from niobium–titanium cable and operates at cryogenic temperatures of about 4.5 K, allowing zero electrical resistance and the high current needed to sustain the strong field. To contain the magnetic flux and return the field lines, the solenoid is surrounded by a 12 kt steel yoke that also serves as absorber plates for the HCAL [48]. A map of the measured magnetic field in the r–z plane of CMS is shown in Fig. 4.6.

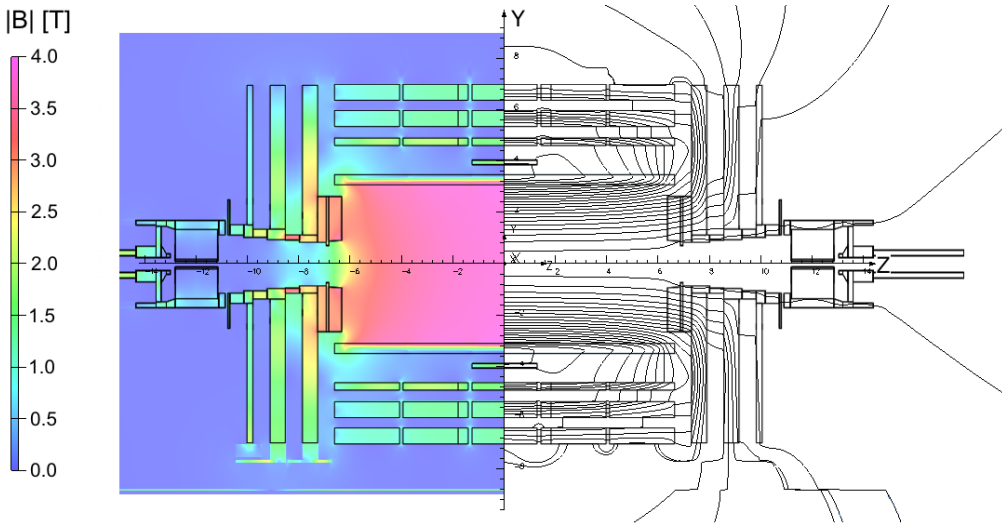


Figure 4.6: Value of the magnetic field flux density $|B|$ (left) and the corresponding field lines (right) predicted for a longitudinal r–z cross-section of the CMS detector at the nominal central field of 3.8 T. Each line denotes a magnetic flux increment of 6 Wb. Adapted from Ref. [47].

4.2.6 Muon system

Muons are key signatures in many flagship measurements at the LHC: electroweak processes, top-quark physics, Higgs boson decays, and searches for physics beyond the SM. Because muons are minimally ionizing and traverse several meters of iron, they pass through the calorimeters and are measured in a dedicated muon spectrometer placed outside the solenoid and embedded in the steel return yoke. The CMS muon system comprises four stations arranged in the barrel and endcap regions and employs three complementary technologies: drift tubes (DTs), cathode

strip chambers (CSCs), and resistive plate chambers (RPCs) [49–51]. Together, they provide precise tracking, robust pattern recognition, and fast timing for triggering at high luminosity.

Drift Tube (DT) Chambers: DTs instrument the barrel ($|\eta| \lesssim 1.2$) where the magnetic field and background rates are moderate. Chambers are arranged in four concentric stations mounted on the yoke wheels. Each chamber consists of several layers of drift cells filled with an argon–carbon–dioxide–based gas; hit positions are reconstructed from drift times, delivering a single-hit spatial resolution of about 200-250 μm in the bending plane and good z information from staggered superlayers. Combining multiple layers into track segments improves the effective position resolution to the level of $\mathcal{O}(100)\mu$m. The large lever arm from the interaction point to the outer stations provides excellent standalone momentum resolution in the barrel.

Cathode Strip Chambers (CSC): CSCs cover the endcaps ($1.0 \lesssim |\eta| \lesssim 2.4$), where the magnetic field is less uniform, and particle rates are higher. Multiwire proportional chambers, featuring finely segmented cathode strips and orthogonal anode wires, deliver precise two-dimensional space points per layer. With multiple layers per chamber, CSCs offer high-efficiency pattern recognition, fast timing, and radiation tolerance, making them suitable for both precision tracking and triggering in the forward region.

Resistive Plate Chambers (RPC): RPCs are gaseous parallel-plate detectors operated in avalanche mode, providing very fast response and $\mathcal{O}(1)$ bunch-crossing time resolution. Installed in both the barrel and endcap regions, they provide redundant timing and coarse spatial information that enhances the trigger and improves robustness to local inefficiencies in the tracking chambers.

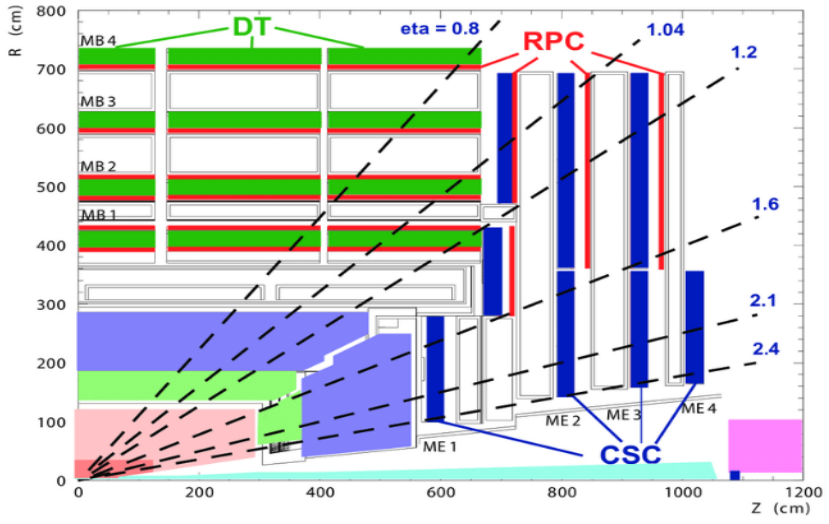


Figure 4.7: Longitudinal quarter view of the CMS muon system. The four barrel stations (MB1–MB4) are mounted on the return-yoke wheels; the endcap stations (ME1–ME4) are arranged in disks. DTs populate the barrel, CSCs the endcaps, and RPCs provide fast timing in both regions. Taken from Ref. [52].

Overall, the CMS muon system provides efficient and precise muon reconstruction over the pseudorapidity range $|\eta| < 2.4$. The single-muon trigger efficiency exceeds 90%, while the combined reconstruction and identification efficiency is above 96%. Matching muon tracks to the silicon tracker yields a relative transverse momentum resolution of about 1% in the barrel and 3% in the endcaps for p_T up to 100 GeV , with resolutions better than 7% for p_T up to 1 TeV in the central region.

In preparation for higher instantaneous luminosities in Run III and beyond, the CMS muon system has been upgraded with the installation of a Gas Electron Multiplier (GEM) detector [53] in the forward region. The GEM system enhances redundancy and improves muon triggering and reconstruction in the high-$|\eta|$ region, where background rates are highest. While the GEM detectors were not used in the reconstruction of the Run II data analyzed in this thesis, they represent an important upgrade to maintain CMS muon performance under increasingly challenging LHC conditions.

4.3 Trigger system

The LHC delivers p-p collisions at an extremely high rate, reaching approximately 40 MHz at the CMS interaction point, corresponding to one bunch crossing every 25 ns. It is neither practical nor necessary to record all of these events, as most are low-energy soft interactions with little physics interest. The CMS trigger system [54] is therefore designed to rapidly select, in real time, the small fraction of collisions that are potentially interesting for physics analyses while discarding the rest.

CMS employs a two-tier trigger architecture:

Level-1 Trigger (L1): Implemented in custom, high-speed hardware processors, the L1 trigger [55] uses coarse-granularity information from the calorimeters and the muon detectors to identify basic physics objects such as high p_T leptons, photons, jets, and missing transverse energy. It reduces the event rate from the LHC's 40 MHz down to about 100 kHz within a latency of roughly 3-4 μs, ensuring only the most promising events are passed on for further processing.

For Run II, the L1 trigger system was significantly upgraded to cope with the increased center-of-mass energy, instantaneous luminosity, and pileup. The upgraded system provides improved jet and muon momentum resolution, enhanced lepton isolation, and incorporates pileup-mitigation techniques and invariant-mass calculations at the trigger level. These improvements allow lower and more stable trigger thresholds to be maintained compared to Run I, ensuring high and uniform efficiency for processes involving top quarks and electroweak bosons. The performance of the upgraded L1 trigger remained stable throughout Run II, providing a reliable and efficient event selection for the analysis presented in this thesis.

High-Level Trigger (HLT): Events accepted by the L1 trigger are transferred to a large farm of commercial CPUs and GPUs, where the HLT performs a software-based reconstruction

using full-resolution detector data and algorithms similar to those used in offline reconstruction. The HLT further reduces the data rate to several kHz [56], selecting events based on more sophisticated criteria such as refined lepton isolation, jet substructure, or complex multi-object signatures.

This hierarchical trigger approach allows CMS to efficiently record events of interest, such as those containing high p_T leptons from electroweak bosons, top quarks, or potential new-physics processes, while operating within the limits of data acquisition and storage bandwidth. Events at this trigger level are selected by predefined HLT paths, and the HLT paths used in this analysis are explained in Sec. 7.3.1.

5 Object reconstruction and event definition

The analysis presented in the subsequent chapters relies on the accurate reconstruction of physics objects from the raw detector signals recorded in the CMS detector (4.2). The task of object reconstruction is to translate the information from the various detector components. These include tracks in the tracking system and energy deposits in the calorimeters, translated into well-defined physics objects, including leptons, jets, and missing transverse energy (MET). This step is crucial, as it forms the bridge between the low-level detector observables and the high-level quantities used in physics analyses. This enables the extraction of meaningful information about the underlying collision processes. This chapter is structured as follows. In the first part 5.1, the reconstruction algorithms and the configurations used for the relevant objects are introduced. Based on these objects, the event signatures of the signal and background processes considered in this thesis are then defined. Finally, the simulation framework for p-p collisions at the LHC is outlined, providing the basis for comparing simulated samples with recorded data, as explained in the second part 5.2.

5.1 Object reconstruction

5.1.1 Track and vertex

The charged particles traversing the CMS tracking system leave localized energy deposits, or hits, in the silicon pixel and strip detectors as they interact with the detector material. These hits, distributed along different detector layers, can be combined to reconstruct the trajectory of the particle in the presence of the solenoidal magnetic field. The trajectories follow approximately helical paths due to the magnetic field, and their reconstruction provides access to essential kinematic properties such as momentum and charge. At CMS, the track reconstruction employs a combinatorial track-finding algorithm [57] based on the Kálmán filter technique [21, 58]. The procedure is iterative:

- Seedling: initial track seeds are formed from hits in neighbouring layers of the pixel detector.

- Track finding: seeds are extrapolated across successive detector layers, and compatible hits are added to the candidate trajectory using the Kálmán filter formalism.
- Track fitting: the parameters of the reconstructed trajectory are refined to best match the observed hits, while accounting for detector resolution, multiple scattering, and energy loss in the tracker material.

The tracks that do not meet quality criteria, such as the minimum transverse momentum or a sufficient number of hits, are discarded. After each iteration, hits already associated with accepted tracks are removed, and the procedure is repeated with relaxed constraints to capture more difficult tracks. An analogous method is applied in the muon system using hits in the muon detectors. Tracks are subsequently used to reconstruct interaction vertices.

Tracks whose extrapolated z-positions cluster are assembled into candidate vertices, and their positions are refined with an adaptive vertex fitting algorithm [59]. Among the reconstructed vertices, the primary vertex is defined as the one with the largest scalar (or squared) sum of the transverse momenta of its associated tracks. This choice reflects the expectation that the hardest proton-proton interaction in the bunch crossing corresponds to the physics process of interest. The vertices must additionally satisfy quality criteria, including a location within $|z| < 24$ cm and $r < 2$ cm of the detector center, and a sufficient number of associated tracks. Events that fail to meet these requirements are discarded from the analysis. The accurate reconstruction of tracks and vertices is crucial for identifying of physics objects of interest and therefore forms the foundation of the event reconstruction strategy employed in this analysis.

5.1.2 Particle-flow (PF) algorithm

The global event reconstruction at CMS relies on the particle-flow algorithm [60]. Unlike reconstruction strategies confined to individual subdetectors, the PF approach combines information from all major detector components, the silicon tracker, electromagnetic and hadronic calorimeters, and the muon system to identify and reconstruct each particle in the event. By optimally integrating complementary measurements, the PF algorithm achieves a resolution in energy and momentum superior to that of each sub-system used in isolation. The reconstruction proceeds in an ordered sequence. Objects with the most distinct detector signatures are identified first, and the remaining information is then used for more challenging particles. In practice, the algorithm starts from the charged-particle tracks and vertices, which are linked to calorimeter clusters and muon detector hits through a dedicated linking procedure. The resulting associations, referred to as PF blocks, are subsequently interpreted by particle-specific reconstruction sequences.

- **Muons** are reconstructed by associating tracks in the tracker with segments in the muon chambers. Three categories are distinguished: firstly, standalone muons, which are built from tracks in the muon system alone, secondly, tracker muons, identified from tracks in the silicon tracker with at least one matching segment in the muon detectors, and global muons, which combine tracker and muon information for the best resolution.

- **Electrons** are reconstructed by matching a track in the tracker with energy deposits in the ECAL. Bremsstrahlung photons emitted along the electron trajectory are clustered into so-called superclusters and incorporated into the electron energy measurement. To account for significant energy losses in the tracker material, a Gaussian Sum Filter (GSF) is used for track fitting.
- **Photons** are identified as ECAL energy deposits without an associated track in the silicon tracker. Additional care is taken to account for photon conversions in the tracker, where secondary tracks from e^+e^- pairs can appear displaced from the primary vertex.
- **Charged hadrons** are reconstructed from tracker tracks that are matched to calorimeter energy deposits, with the momentum primarily measured by the tracker due to its higher resolution at low energies.
- **Neutral hadrons** leave no track in the tracker and are reconstructed solely from calorimeter energy deposits, mainly in the HCAL. After the reconstruction of all other particle candidates, any remaining calorimeter energy is attributed to neutral hadrons.

To mitigate the impact of additional p-p interactions within the same bunch crossing (pileup), charged hadrons consistent with pileup vertices are removed from the PF candidate collection in a procedure referred to as charged hadron subtraction (CHS) [61]. The final set of PF candidates forms the basis for reconstructing higher-level objects, such as jets, leptons, and missing transverse energy, which are central to the analysis presented in this thesis. For the analysis, additional identification and isolation requirements are applied, as described in detail in Chapter 7.

5.1.3 Jets

In high-energy p-p collisions, quarks and gluons produced in the hard scattering process cannot exist as free particles due to color confinement. Instead, they undergo parton showering and hadronization, giving rise to collimated sprays of hadrons, which are reconstructed experimentally as jets. The properties of these jets provide access to the underlying partonic kinematics and are therefore essential in analyses involving top quarks and associated processes.

At CMS, jets are reconstructed from PF candidates using the anti-k_T sequential recombination algorithm [62]. This algorithm is widely employed because it is both infrared-safe (soft emissions do not create additional jets) and collinear-safe (collinear splittings do not alter the jet definition). The distance measure used in the clustering is defined as

$$d_{ij} = \min\left(p_{T,i}^{-2},\, p_{T,j}^{-2}\right)\frac{\Delta R_{ij}^2}{R^2}, \qquad d_i = p_{T,i}^{-2}, \tag{5.1}$$

where $p_{T,i}$ is the transverse momentum of object i, ΔR_{ij} is the distance in the (η, ϕ) plane, and R is the jet radius parameter. For this thesis, jets are clustered with $R = 0.4$, referred to as AK4 jets. The clustering proceeds iteratively until all PF candidates are assigned to jets,

resulting in cone-like jet shapes in the $\eta - \phi$ plane. Particles with higher transverse momentum dominate clustering, such that they define the jet axis, while softer constituents are added subsequently. For boosted heavy objects, larger radius parameters such as $R = 0.8$ (AK8 jets) or $R = 1.5$ (AK15 jets) are also employed in CMS analyses. In this work, however, only AK4 jets are used, as the targeted final states are characterized by moderate jet radii. Since many of the jets in the final state arise from heavy-flavor quarks, in particular bottom (b) and charm (c) quarks, dedicated tagging algorithms are employed to identify their origin. The identification of heavy-flavor jets and the associated b-tagging discriminants are discussed below.

Heavy flavor jet-tagging

The identification of the flavor of reconstructed jets is a crucial component of this analysis, as it allows the discrimination between jets originating from heavy-flavor quarks (b or c) and those initiated by light quarks or gluons. Heavy-flavor hadrons, in particular B hadrons, exhibit relatively long lifetimes of the order of $\tau \sim 10^{-12}$s. As a consequence, they travel a measurable distance of several hundred micrometers before decaying, often producing displaced secondary vertices and tracks with large impact parameters relative to the primary vertex. These distinctive signatures provide powerful handles for heavy-flavor identification. Modern jet

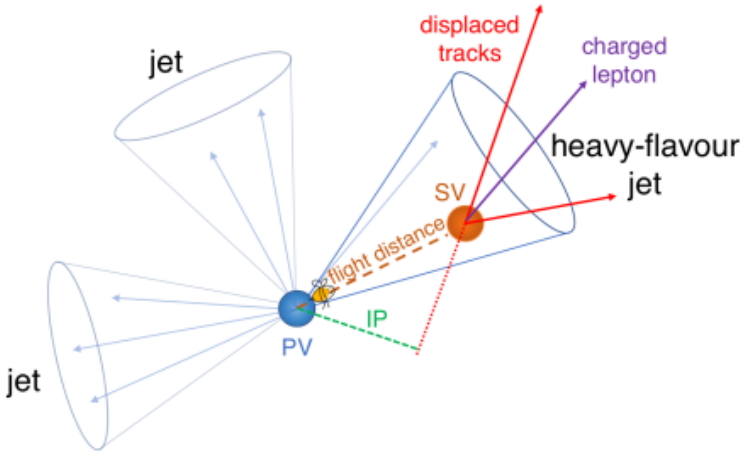


Figure 5.1: Schematic representation of a heavy-flavor jet containing a secondary vertex (SV) arising from the decay of a b or c hadron. It represents a hard-scattering interaction, in which the decay produces displaced charged particle tracks, and in some cases a soft lepton, relative to the primary vertex (PV). This displacement leads to large impact parameters (IPs). The figure is taken from the Ref. [63].

tagging techniques at CMS exploit not only the secondary-vertex information but also low-level features of the individual particle-flow (PF) constituents within the jet. The current standard algorithm used in Run II is the DeepJet tagger [64, 65], a deep neural network architecture that builds on its predecessor DeepCSV [63]. In addition to secondary vertex and track-based lifetime observables, DeepJet incorporates information from up to 25 charged and neutral PF constituents, as well as the four leading secondary vertices. Dedicated convolutional layers with

1×1 kernels are applied to these inputs, followed by recurrent layers based on Long Short-Term Memory (LSTM) cells to capture correlations among jet constituents. The outputs from these stages are concatenated with global jet observables and passed through fully connected layers to provide final probability scores for different jet flavors. The network returns six output nodes corresponding to different jet categories: three nodes for b jets (distinguishing between single-B, double-B, and leptonic-B hadron decays), one for c jets, and two for light-flavor and gluon-initiated jets. The b-tagging discriminant, commonly referred to as the b-tag in this thesis, is defined as the sum of the three b-related output probabilities. The architecture of the DeepJet tagger is illustrated in the Fig. 5.2.

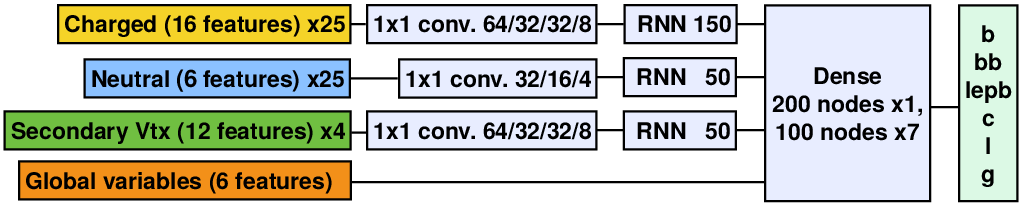


Figure 5.2: Schematic representation of the architecture of the DeepJet algorithm. Different sets of input features from charged particles, neutral particles, and secondary vertices are processed through convolutional and recurrent layers, while global event variables are included directly. The outputs are combined in dense layers to provide classification scores for different jet flavor categories. The figure is taken from the Ref. [65].

Similarly, two-dimensional discriminants can be constructed to distinguish c jets from b jets (CvB) and from light-flavor/gluon jets (CvL), defined as

$$\text{CvB} = \frac{p(c)}{p(c) + p(b) + p(bb) + p(\ell b)}, \qquad \text{CvL} = \frac{p(c)}{p(c) + p(uds) + p(g)} \tag{5.2}$$

where $p(c)$ denotes the probability assigned by the tagger that a jet originates from a c-quark, $p(b)$ corresponds to a b hadron, $p(bb)$ to two b hadrons (e.g. from gluon splitting $g \to b\bar{b}$), $p(\ell b)$ to a b hadron with a semileptonic decay, i.e. with a lepton inside the jet. Furthermore, $p(uds)$ indicates the probability for the jet to originate from a light-flavor quark (u, d, or s), and $p(g)$ represents the probability that the jet originates from a gluon.

The performance of the DeepJet algorithm is typically quantified in terms of receiver operating characteristic (ROC) curves, which evaluate the trade-off between b-jet efficiency and light-jet misidentification rate. Significant improvements in b-tagging efficiency were observed with the introduction of the upgraded Phase-1 pixel detector in 2017, which enhanced the resolution of displaced tracks and secondary vertices. The DeepJet tagger thus represents the state-of-the-art heavy-flavor tagging tool at CMS during Run II and forms a key input for the jet and event classification tasks presented in this thesis. Working points (WPs) for the DeepJet algorithm are defined at constant light-jet misidentification rates, where light jets

refer to those originating from u, d, s quarks or gluons. In CMS convention, three WPs are typically provided: tight (0.1%), medium (1%), and loose (10%) misidentification rates. The corresponding discriminator thresholds are summarized in Chapter 8. At the medium (tight) WP, b jets are correctly identified with efficiencies of about 75–80% (60%), while c jets are misidentified as b jets in approximately 15% (3%) of cases, as evaluated on simulated $t\bar{t}$ events. Furthermore, the dependence of the b-tagging efficiency on the jet transverse momentum is illustrated in Fig. 5.3. For jets with $p_T \geq 30$ GeV, the efficiency reaches 70–80% at the

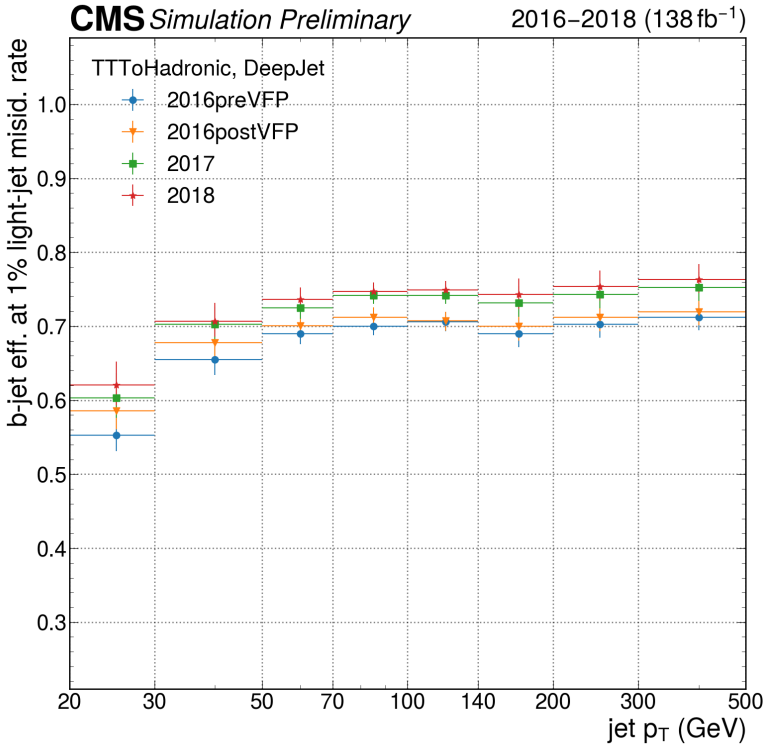


Figure 5.3: b-jet identification efficiency at a fixed light-jet misidentification rate of 1% as a function of jet transverse momentum for the DeepJet algorithm in simulated events in the fully hadronic channel. Results are shown separately for different data-taking eras (2016preVFP, 2016postVFP, 2017, and 2018). The improved efficiency in 2017 and 2018 reflects the performance gains from the CMS Phase-1 pixel detector upgrade. Taken from Ref. [66]

medium WP across all data-taking eras (2016preVFP, 2016postVFP, 2017, and 2018), while jets with $20 \leq p_T < 30$ GeV exhibit significantly reduced efficiency [66]. The 2016preVFP and 2016postVFP periods show efficiencies toward the lower end of this range, the 2017 era lies in between, while the 2018 data-taking era reaches slightly higher efficiencies within the same range. These improvements are directly correlated with the Phase-1 upgrade of the CMS pixel detector, where the enhanced tracking and secondary vertex reconstruction capabilities provided by this upgrade resulted in an increase of about 5% in b-jet identification efficiency at the same

light-jet misidentification rate, as shown in the Fig. 5.4. For this reason, and to ensure a robust tagging performance, only jets with $p_T \geq 30$ GeV are considered in the present analysis. The additional criteria on the identification and selection of jets are summarized in the Sec. 7.3.4).

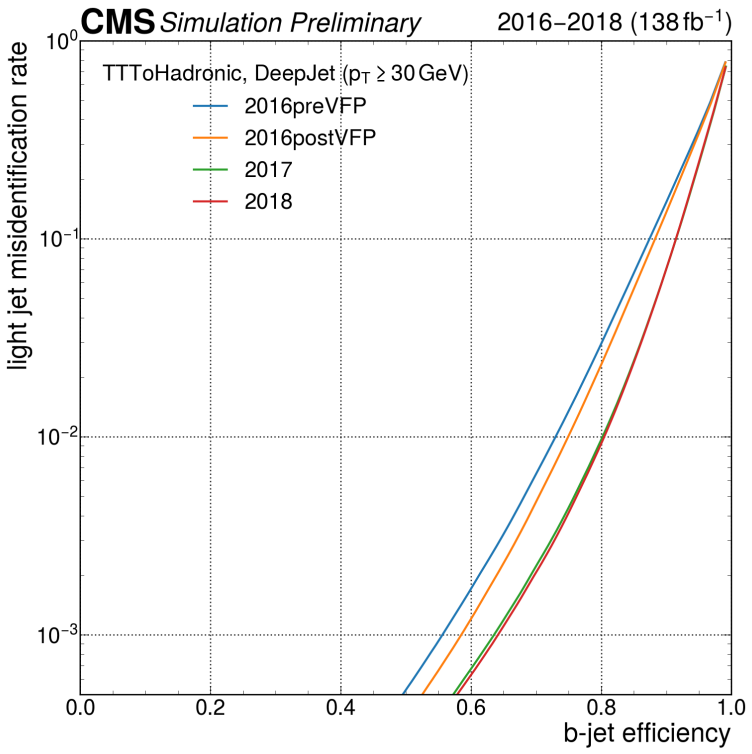


Figure 5.4: ROC curve of the DeepJet heavy-flavor jet tagging algorithm obtained from simulated $t\bar{t}$ events in the fully hadronic channel, considering jets with $p_T \geq 30$ GeV. The plot shows b-jet identification efficiency as a function of the light-jet misidentification rate, compared across different data-taking eras (2016preVFP, 2016postVFP, 2017, and 2018). Taken from Ref. [66].

Missing transverse momentum (MET)

Certain particles produced in p-p collisions cannot be directly reconstructed in the CMS detector, as they interact only weakly or not at all with the detector material. Typical examples are neutrinos, while many theories of physics beyond the SM (BSM) also predict invisible states, such as dark matter candidates. Although these particles leave no direct signal, their presence can be inferred indirectly through an imbalance of transverse momentum in the event. Since the colliding protons carry a negligible net transverse momentum, momentum conservation requires the vectorial sum of the transverse momenta of all final-state particles to vanish. A non-zero imbalance, therefore, indicates that one or more detected particles have escaped the

detector. The missing transverse momentum, $\vec{p_T}^{\text{miss}}$ is defined as,

$$\vec{p_T}^{\text{miss}} = -\sum_{i}^{\text{visible}} \vec{p_T}_{,i}, \tag{5.3}$$

and its magnitude, p_T^{miss}, is referred to as missing transverse energy (MET). The reconstruction of MET relies on summing over all PF candidates in an event. It is therefore sensitive to detector resolutions, calibration uncertainties, and detector effects such as noise or mis-measurements. While the presence of MET signals the escape of invisible particles, its precise interpretation depends on both detector performance and the underlying physics process.

5.2 Event simulation

In order to interpret the recorded data from the CMS experiment and compare it with the predictions of the Standard Model, simulated event samples are required. Since the underlying physics process of a collision cannot be uniquely identified from the detector response, particularly when different processes lead to similar final states, simulations provide the necessary reference to model both signal and background contributions. These simulated samples are processed with the same reconstruction algorithms as the data, ensuring a consistent treatment. They are indispensable not only for testing theoretical predictions and searching for possible deviations indicating new physics, but also for providing the labeled datasets required for training and validating machine learning models used in this analysis, as introduced in Chapter 8.

Event simulation plays a central role in this thesis, as it provides the reference for signal and background modeling, enables the estimation of detector effects, and supports the validation of the analysis strategy. The simulation of p-p collisions at CMS proceeds in several stages. First, the hard-scattering process is generated using Monte Carlo event generators. Next, parton showering and hadronization model the evolution of colored partons into stable hadrons. Finally, the interaction of the resulting particles with the detector is described using a detailed detector simulation. The same reconstruction algorithms as applied to real data are then run on the simulated detector signals, ensuring that simulated events can be analyzed under identical conditions to data. In the following subsections, the different steps of the simulation chain relevant to this analysis are described, including the generation of hard scattering processes, the modeling of parton showers and hadronization, and the detector simulation employed by CMS.

5.2.1 Event generation

The generation of simulated events at CMS follows the factorized description of p-p collisions in QCD. At the LHC energies, the colliding objects are not intact protons but their constituents, quarks and gluons, collectively referred to as partons. The momentum distribution of these partons inside the proton is described by parton distribution functions (PDFs), which depend on the momentum fraction x carried by the parton and the factorization scale μ_F at which the proton is probed. The PDFs are not calculable from first principles; instead, they are determined

by global fits to experimental data, and their scale dependence is predicted by perturbative QCD via the DGLAP (Dokshitzer–Gribov–Lipatov–Altarelli–Parisi) evolution equations [67–70]. In this thesis, we use the NNPDF3.1 PDF set, which is fitted with neural-network parameterizations and provides theory ingredients computed up to NNLO (Next-to-Next-to-Leading Order) in perturbative QCD [71].

Given the PDFs, the next step is to describe the hard scattering, i.e., the partonic subprocess that produces the particles of interest. The cross section for this subprocess is expressed as a perturbative expansion in the strong coupling constant α_s, with leading-order (LO) calculations including the lowest-order diagrams and higher-order corrections at next-to-leading order (NLO) and beyond, accounting for additional real emissions and virtual loops. The total proton–proton cross section for a final state X can then be written in factorized form as

$$\sigma(pp \to X) = \sum_{a,b} \int dx_1\, dx_2\, f_a(x_1, \mu_F^2)\, f_b(x_2, \mu_F^2)\, \hat{\sigma}_{ab\to X}(x_1, x_2, \mu_F^2, \mu_R^2), \tag{5.4}$$

where $f_a(x_1, \mu_F)$ and $f_b(x_2, \mu_F)$ are the PDFs of partons a and b, $\hat{\sigma}$ is the partonic cross section, and μ_F and μ_R denote the factorization and renormalization scales. These scales are not physical observables but must be chosen consistently within the simulation setup. In the POWHEG framework, an additional parameter h_{damp} is introduced to control the separation between hard emissions described by the matrix element and softer emissions described by the parton shower.

The matrix elements for the processes relevant to this thesis are generated with MadGraph5_aMC@NLO [72] and the POWHEG BOX [73, 74]. While MadGraph [75] allows the simulation of multi-leg processes at LO and NLO accuracy, POWHEG [76] provides an NLO description of inclusive processes such as $t\bar{t}$ and single-top production, generating the hardest emission with positive event weights. The decays of heavy particles, in particular top quarks and electroweak bosons, are modeled with MadSpin [77], which preserves spin correlations in the final state.

5.2.2 Parton showering and hadronization

The description of additional QCD radiation is achieved by parton showers, which resum leading soft- and collinear-logarithmic terms to all orders, modeling both initial-state radiation (ISR) and final-state radiation (FSR). The evolution of a parton shower is governed by Sudakov form factors, which describe the probability for a parton to evolve between two energy scales without emitting additional radiation. To avoid double-counting, the hardest emission is taken from the matrix-element generator, while subsequent softer emissions are produced by the parton shower. In this analysis, Pythia8 [78] is used for the parton shower simulation and for modeling the hadronization process, where colored partons form color-neutral hadrons. The Lund string fragmentation model, as implemented in Pythia [79], provides the phenomenological description of hadronization. In addition, Pythia simulates the underlying event, i.e., additional

soft parton interactions accompanying the primary scattering, using the CP5 tune [80] optimized for Run II data at $\sqrt{s} = 13$ TeV. This setup provides NLO-accurate predictions for the relevant processes while ensuring a realistic modeling of QCD radiation and hadronization.

5.2.3 Detector simulation

The response of the CMS detector to all stable particles must be simulated to enable a meaningful comparison with real collision data. This step includes interactions of particles with the detector material, energy deposits in the calorimeters, ionization in the silicon trackers, and signals in the muon chambers, as well as the effect of the 3.8 T magnetic field. The CMS experiment employs the GEANT4 framework [81] for this purpose, which provides a detailed modeling of the detector geometry, subdetectors, electronics, and material composition. GEANT4 simulates the transport and interactions of all particles through the detector, resulting in realistic detector signals that closely resemble those from actual collisions. These signals are then processed with the standard CMS reconstruction algorithms, ensuring that simulated events can be directly compared to data at the reconstructed-object level. The corrections applied to account for known differences between data and simulation are described in Chapter 7.

5.2.4 Pileup simulation

In addition to the hard scattering process, multiple p-p interactions can occur within the same or neighboring bunch crossings. These additional interactions, referred to as pileup, contribute extra energy deposits in the detector and significantly affect the reconstruction of physics objects. Pileup is therefore included in the simulation to provide a realistic description of the collision environment at the LHC. The pileup contribution is modeled by overlaying additional minimum-bias events on top of the hard-scattering event. The number of pileup interactions per bunch crossing is assumed a priori and sampled from a Poisson distribution with a mean corresponding to the expected instantaneous luminosity. Both in-time pileup (from the same bunch crossing as the hard interaction) and out-of-time pileup (from adjacent bunch crossings, due to the finite detector response time) are taken into account. The resulting simulated pileup profile is later reweighted to match the pileup distribution observed in data.

5.2.5 Jet flavor identification

For analyses involving heavy-flavor jets, it is essential to establish the flavor and origin of reconstructed jets at the generator level, i.e., before the detector simulation. This information is required to evaluate the performance of tagging algorithms and to define truth-level categories against which reconstructed observables can be compared. The procedure relies on identifying heavy hadrons produced during hadronization and associating them with particle-level jets. The identification is carried out in several steps. First, all B and C hadrons in the event are identified, and their ancestry is traced through the generator record. Each hadron is linked to its corresponding b or c quark, and the parent particle of this quark is

used to determine the hadron's origin. In this way, a hadron can be classified as originating, for example, from the decay of a top quark, a Higgs boson, or a Z boson, or from gluon splitting.

To associate the heavy hadrons with particle-level jets, the so-called *ghost clustering* technique [82, 83] is employed. In this approach, the four-momenta of the heavy hadrons are rescaled to infinitesimal values while preserving their directions, and they are clustered together with the stable final-state particles. Due to their negligible energy, these 'ghost' hadrons do not alter the kinematics of the jets, but their presence allows a direct association between hadrons and the clustered jets. A particle-level jet containing one or more ghost-clustered B hadrons is classified as a b jet, while jets without B hadrons but with at least one C hadron are classified as c jets. All remaining jets are identified as light-flavor jets, originating from u, d, or s quarks, or from gluons. In addition to the flavor classification, the generator record can also be used to determine the origin of a jet. A jet is labeled as a $t\bar{t}$ jet if any of its ghost-clustered hadrons descend from a top quark in the Monte Carlo decay chain. Conversely, jets without such ancestry are considered additional jets. This distinction is important in processes such as $t\bar{t}$ +jets, where one aims to separate the jets originating from top-quark decays from those produced in association with the $t\bar{t}$ system.

5.2.6 Event definitions in $t\bar{t}$+ jets

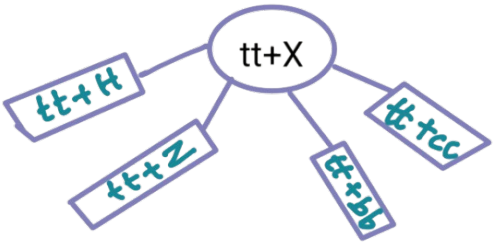
A distinction is made between jets originating from the decay of the $t\bar{t}$ system and those produced in association with it. In this thesis, a b jet that originates from the $t\bar{t}$ decay is referred to as a top b jet, and all other reconstructed jets that are not associated with such ancestry are classified as additional jets, as described in the previous section. These additional jets may arise from QCD radiation (initial (ISR) or final (FSR) state), from gluon splitting ($g \to b\bar{b}$ or $g \to c\bar{c}$), or from the decay of heavy bosons such as the Higgs or the Z boson. Based on the presence and flavor content of these additional jets, different event topologies within the $t\bar{t}$+X phase space are defined. If one or more additional b jets are present, the event is categorized as $t\bar{t}$+B. Depending on the configuration of the additional b hadrons, three sub-processes are distinguished:

- $\mathbf{t\bar{t}b\bar{b}}$: two additional B hadrons are clustered into two distinct jets
- $\mathbf{t\bar{t}2b}$: two additional B hadrons are clustered into a single jet, e.g. due to a collinear $g \to b\bar{b}$ splitting
- $\mathbf{t\bar{t}b}$: only one additional jet contains a B hadron, while the other falls outside of the detector acceptance.

These categories are collectively referred to as $\mathbf{t\bar{t}B}$. Analogous definitions apply when the additional jets contain C hadrons, leading to the category $\mathbf{t\bar{t}C}$. If neither additional b nor c hadrons are present, the event is classified as $\mathbf{t\bar{t}LF}$, where LF denotes additional light flavor or gluon jets. Beyond the QCD-induced categories, specific processes involving heavy bosons are also considered. Events with additional jets originating from a Higgs boson decay are classified

as $\mathbf{t\bar{t}H}$, while those with additional jets from a Z boson decay are categorized as $\mathbf{t\bar{t}Z}$. These are further classified by boson decay mode. Events with additional b jets from the boson decay are categorized as $\mathbf{t\bar{t}H}(\mathbf{b\bar{b}})$ or $\mathbf{t\bar{t}Z}(\mathbf{b\bar{b}})$. In contrast, $\mathbf{t\bar{t}H}(\mathbf{nonB})$ and $\mathbf{t\bar{t}Z}(\mathbf{nonB})$ refer to cases where the Higgs or Z boson decays hadronically into quarks other than b jets.

Measurement of $t\bar{t}$+X events



6 Introduction

This chapter introduces the physics motivation and conceptual framework of the $t\bar{t}$+X measurement presented in this thesis. It outlines the relevant production mechanisms and event topologies (Sec. 6.1), discusses the principal experimental and theoretical challenges associated with the separation of heavy-flavor components (Sec. 6.2), and presents the overall analysis strategy underlying the measurement (Sec. 6.3). Measurements of $t\bar{t}$ production in association with additional heavy-flavor jets or bosons have been performed by CMS and ATLAS collaborations at $\sqrt{s} = 13$ TeV. These previous measurements, including a dileptonic channel analysis within our group, provide the foundation for this work, as discussed in Sec. 6.4. The detailed experimental ingredients and implementation of the analysis are described in the subsequent chapters: the definition of reconstructed objects and event selection in Chapter 7, the machine-learning-based classification strategy in Chapter 8, and the statistical inference procedure in Chapter 9.

6.1 Physics motivation

The production of top quark-antiquark quark pairs in association with additional jets ($t\bar{t}$+X) provides a powerful laboratory for testing the SM at the highest energy scales accessible at the LHC. Owing to its large mass, the top quark couples strongly to the Higgs sector and plays a central role in electroweak symmetry breaking (Sec. 3.2). Precise measurements of associated production processes, such as $t\bar{t}$+H and $t\bar{t}$+Z, therefore, probe fundamental SM couplings and offer a better understanding of backgrounds of many SM as well as BSM processes, where these processes often constitute dominant and irreducible backgrounds. Moreover, since these interactions are not only sensitive to the properties of the Higgs or the Z bosons but also to the additional degrees of freedom entering the interaction, they offer sensitivity to possible effects of physics beyond the SM. Deviations from SM predictions in these channels could arise from anomalous couplings or higher-dimensional operators in effective field theory extensions. At the same time, these processes constitute irreducible backgrounds to one

another, making their simultaneous and consistent treatment essential for a robust measurement.

In particular, the production of top quarks with a Higgs boson ($t\bar{t}$+H) enables a direct probe of the Yukawa coupling between the Higgs boson and the top quark, the heaviest up-type fermion in the SM. This measurement is thus also a crucial test of the Higgs-fermion coupling structure. Among the Higgs decay modes, the decay $H \to b\bar{b}$ has the largest branching ratio [84] in the SM Higgs boson mass range and thus provides the highest statistical sensitivity. At leading order, $t\bar{t}$+H production occurs predominantly through gluon-gluon fusion, with the Higgs boson radiated either from the top quark or from the top quark-antiquark pair, as illustrated by the representative Feynman diagrams shown on top in Fig. 6.1. The Higgs boson can decay leptonically or hadronically, including decays to quark pairs, gluons, or electroweak bosons, with the dominant mode being to a pair of b-quarks. In contrast to $t\bar{t}$H production, the $t\bar{t}$Z and $t\bar{t}$W production processes constrain the electroweak interactions of the top quark. The $t\bar{t}$+Z can occur via the radiation of a Z boson from both initial- and final-state quarks, thereby providing direct access to the top-Z coupling through final-state radiation (FSR). This is unlike the $t\bar{t}$W process, which in the SM can only occur via the radiation of a W boson from the initial-state quarks. So far, most measurements of $t\bar{t}$+Z rely on leptonic decays of the Z boson, the hadronic decay mode $Z \to b\bar{b}$ provides complementary sensitivity and is explicitly targeted in this analysis. Consequently, this thesis focuses on the $t\bar{t}$+H and $t\bar{t}$+Z processes in final states where the associated boson decays into a pair of b-quarks.

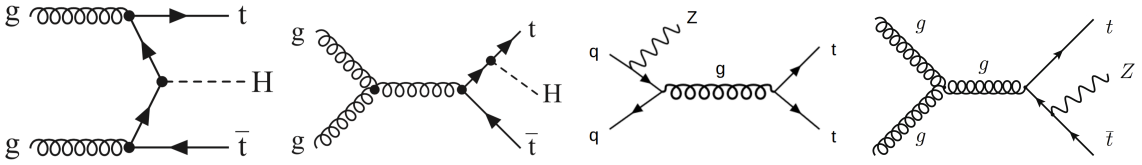


Figure 6.1: Representative leading-order Feynman diagrams for associated $t\bar{t}$+H and $t\bar{t}$+Z productions. In $t\bar{t}$+H production, the Higgs boson is radiated from either the top quark or antiquark, providing direct sensitivity to the top–Higgs Yukawa coupling. In $t\bar{t}$+Z production, the Z boson can be emitted from the initial-state quarks (ISR) or from the final-state top quarks (FSR), probing the electroweak couplings of the top quark.

The top quark pair production is frequently accompanied by additional jets arising from initial-state and final-state radiation. These jets may originate from light quarks (u, d, s), charm and bottom quarks, or gluons. Consequently, the $t\bar{t}$ system is often studied in association with jets, denoted as $t\bar{t}$ +jets. A particularly informative categorization involves defining the final state as $t\bar{t}$+$j\bar{j}$, which represents any such additional jet. Among the various jet flavors, the $t\bar{t}$+$b\bar{b}$ process is of special interest due to both theoretical implications and its role as a major background in

analyses targeting associated production of top quark-antiquark pairs with a Higgs or Z boson ($t\bar{t}$+H and $t\bar{t}$+Z, respectively), especially in final states in which the boson decays into $b\bar{b}$. The flavor identification of these additional jets at the particle level is performed using the ghost-matching technique [83], which associates bottom or charm hadrons to reconstructed jets. Jets containing bottom hadrons are defined as b-jets, those with charm hadrons as c-jets, and all others, including those from gluons and light quarks (u, d, s), as light-flavor (LF) jets.

In this analysis, the additional jets are classified into four primary categories. The $t\bar{t}$+$b\bar{b}$ category includes events with at least two additional b-jets [85], regardless of the number of bottom hadrons in each jet. The $t\bar{t}$ + bj category consists of events with one additional b-jet [86], typically arising due to detector inefficiencies or acceptance limits that prevent both b-jets from being reconstructed. This category also includes cases with a single jet containing two bottom hadrons, commonly referred to as $t\bar{t}$ +2b. The $t\bar{t} + c\bar{c}$ category encompasses events with at least two additional c-jets, regardless of the charm hadron multiplicity per jet. Finally, the $t\bar{t}$ +light-flavor (LF) category comprises events with either two additional light-flavor jets, or one light-flavor jet and one c-jet [87]. It also encompasses inclusive $t\bar{t}$ production without any extra jets. This categorization plays a vital role in separating the signal from the background in analyses of $t\bar{t} + H(b\bar{b})$, $t\bar{t} + Z(b\bar{b})$, and differential measurements of heavy-flavor jet production in association with top quark-antiquark pairs. A precise understanding of inclusive $t\bar{t}$ +jets production, including its heavy-flavor composition, is therefore a prerequisite for extracting the comparatively small signal contributions.

The events of interest in this analysis are $t\bar{t} + H(b\bar{b})$, $t\bar{t} + Z(b\bar{b})$, $t\bar{t}$+$b\bar{b}$, and $t\bar{t} + c\bar{c}$ processes, studied in the semi-leptonic decay phase space (Sec. 3.3.3) of the $t\bar{t}$ system. The $t\bar{t}$ decay modes and specific decay channel, considered in this analysis, are described in Sec. 3.3. Representative leading-order (LO) Feynman diagrams for these events are shown in Fig. 6.2.

6.2 Challenges

In previous measurements, the four processes $t\bar{t} + H(b\bar{b})$, $t\bar{t} + Z(b\bar{b})$, $t\bar{t}$+$b\bar{b}$, and $t\bar{t} + c\bar{c}$ were typically treated independently, with one process defined as signal while the others were absorbed into the background model. In contrast, this analysis aims to simultaneously measure all contributing $t\bar{t}$+X processes. Such an approach enables a consistent treatment of correlations among the processes, both at the level of event reconstruction and in the statistical interpretation. Within the machine learning strategy adopted in this thesis, no process is given special importance a priori, all signal classes are treated on equal footing. While this unified approach is a more robust interpretation of the measurement, as correlations among different $t\bar{t}$+X processes are consistently accounted for and the dependence on a specific signal-background definition is reduced, it introduces several significant challenges.

In the semileptonic channel, as discussed in Sec. 3.3.3, the final states consist of at least six reconstructed jets, four of which originate from b-quarks, sharing a very similar experimental

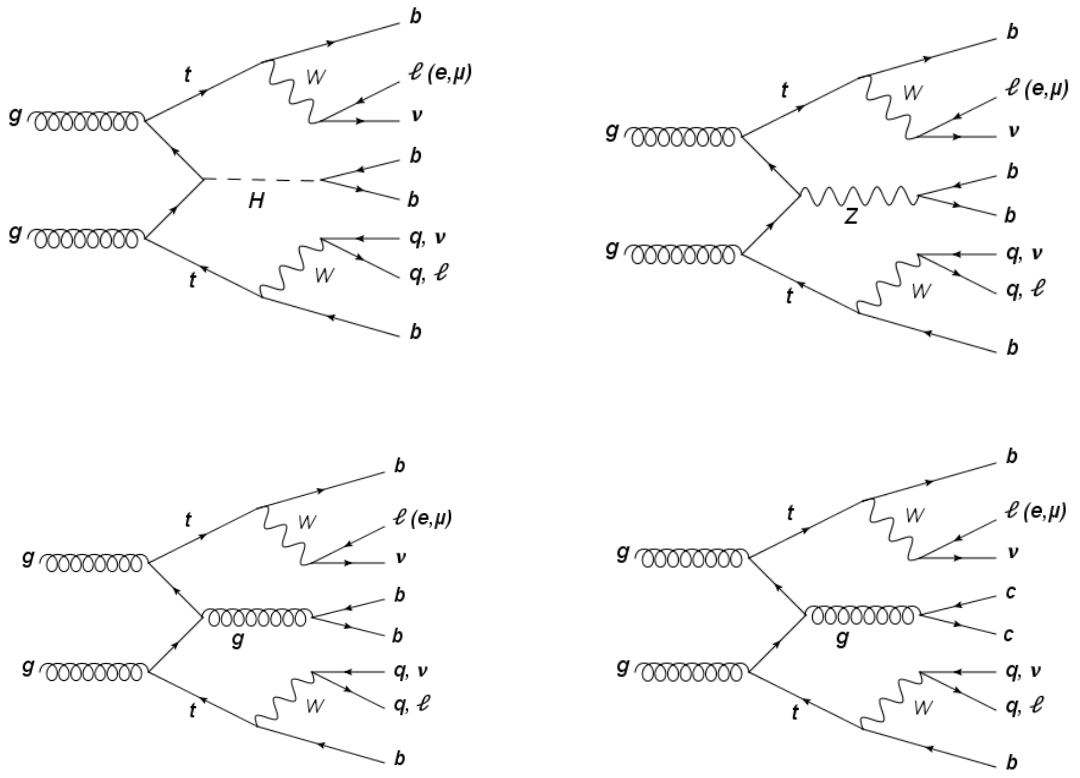


Figure 6.2: Exemplary LO Feynman diagrams for the four signal processes considered in this analysis in the semileptonic $t\bar{t}$ phase space. The signal processes include associated Higgs and Z boson production with H/Z$\rightarrow$ $b\bar{b}$ (top), as well as heavy-flavor $t\bar{t}$+$b\bar{b}$ and $t\bar{t}+c\bar{c}$ production (bottom), which constitute dominant and irreducible backgrounds.

signature, also portrayed in Fig. 6.2. These processes, characterized by high jet multiplicities and comparable kinematic properties, populate largely overlapping regions of phase space, introduce combinatorial ambiguities in the reconstruction, complicating the assignment of reconstructed jets to their underlying partons. Since they only differ in subtle correlations between reconstructed objects, it is difficult to distinguish based on simple event-level observables.

A further challenge arises from the fact that all targeted processes ultimately lead to $t\bar{t}$+$b\bar{b}$-like final-state. At leading order, final states with two top quarks and additional heavy quarks are characterized by multiple b jets. However, the reconstruction and tagging of these jets are subject to finite efficiency and non-negligible mistag probabilities. Imperfect flavor tagging results in migrations between categories: light-flavor or charm jets can be misidentified as b jets, while genuine b jets may fail the tagging requirements. The $t\bar{t} + c\bar{c}$ process is thus important in this context, as charm jets can be misidentified as bottom jets with a relatively high probability. Combined with the similarity of their kinematic properties, this results in strong confusion between these processes, making their separation experimentally demanding.

In addition, the production cross sections of events of interest differ by more than one order of magnitude, leading to highly imbalanced event yields. The broad landscape of SM cross-section measurements performed by CMS is summarized in Ref.[88], where production rates spanning more than fourteen orders of magnitude are compiled in a single overview figure (see Fig. 6.3). Processes range from the total inelastic p-p cross section at the millibarn scale down to rare EW and multi-top production modes at the femtobarn level. Within this hierarchy, inclusive $t\bar{t}$ production occurred at the level of several hundred picobarns, whereas associated processes such as $t\bar{t}$+H, and $t\bar{t}$+Z are suppressed by one to three orders of magnitude. The $t\bar{t}$+X processes investigated in this thesis, therefore, lie in a regime where comparatively small signal contributions must be extracted in the presence of substantially larger backgrounds. This imbalance leads to limited statistical precision and poses challenges for stable machine learning training and reliable statistical inference. Moreover, the theoretical description of heavy-quark and multi-boson final states requires higher-order QCD and electroweak corrections, and residual uncertainties, which affect both normalization and shape of relevant observables. Together with experimental systematic uncertainties, these effects significantly complicate precision $t\bar{t}$+X measurements.

The inclusive $t\bar{t}$+$b\bar{b}$ cross section at the LHC is significantly larger than those of $t\bar{t}$+H or $t\bar{t}$+Z, with H, Z $\rightarrow b\bar{b}$, by roughly one order of magnitude, making it the dominant irreducible background in these measurements [89, 90]. In addition, the kinematic scales relevant for the additional b-jets can differ substantially from those of the $t\bar{t}$ itself, leading to sensitivity to the modeling of hard versus soft radiation and the treatment of massive b-quarks. This scale hierarchy significantly complicates perturbative predictions and event generator descriptions [91]. Because $t\bar{t}$+$b\bar{b}$ constitutes the leading background in precision measurements of $t\bar{t} + H(b\bar{b})$ and $t\bar{t} + Z(b\bar{b})$, any mismodeling directly impacts the extraction of the corresponding signal

strengths and uncertainties in those analyses [85]. Also, theoretical predictions for heavy-flavor production in $t\bar{t}+b\bar{b}$ depend sensitively on the choice of renormalization and factorization scales and the matching of fixed-order calculations to parton showers. Variations in these scales typically dominate the perturbative uncertainties in $t\bar{t}+b\bar{b}$ simulations and are a major contributor to the systematic uncertainty in related measurements and searches [92]. Therefore, the modeling of $t\bar{t}+b\bar{b}$ represents one of the most challenging aspects in analyses targeting additional heavy-flavor jets.

Further complications arise from the sizeable theoretical and experimental uncertainties affecting jet reconstruction, flavor tagging, and background normalizations, which are explained in Sec. 7.6.

6.3 Conceptual analysis strategy

The aim of this analysis is a simultaneous measurement of the $t\bar{t}+X$ processes: $t\bar{t}+H(b\bar{b})$, $t\bar{t}+Z(b\bar{b})$, $t\bar{t}+b\bar{b}$, and $t\bar{t}+c\bar{c}$ in the semileptonic decay channel of the $t\bar{t}$ system, as discussed in previous sections. The strong overlap of the (considered) processes in both topology and kinematics, combined with high jet multiplicities and large combinatorial ambiguities, makes their separation particularly challenging. Traditional cut-based approaches or flat event representations are insufficient to fully exploit the subtle correlations between reconstructed objects that distinguish these processes. This motivates a unified analysis strategy capable of exploiting both object-level and event-level information. An advanced multivariate analysis (MVA) approach is developed for this analysis, which employs state-of-the-art machine learning techniques, based on Graph Neural Networks (GNNs), which are naturally suited to the relational structure of collider events. The experimental ingredients of the analysis, including datasets, object definitions, corrections, and systematic uncertainties, are described in Chapter 5. The strategy is organized hierarchically. First, a node-level classification identifies additional heavy-flavor jets beyond the $t\bar{t}$ decay products, providing object-level information that is robust against variations in jet multiplicity. This information is then propagated to a graph-level classifier, which performs a multiclass event classification across all $t\bar{t}+X$ processes. By combining object-level and event-level inference within a single framework, the analysis maximizes sensitivity while maintaining a consistent treatment of all signal categories. The machine-learning-based analysis strategy, in detail, is presented in Chapter 8. In the final step, the outputs of the event-level classifier are used to construct a global likelihood function, in which all $t\bar{t}+X$ processes are modeled simultaneously. A binned maximum-likelihood fit is performed to extract the signal strength parameters of the individual processes, while systematically accounting for experimental and theoretical uncertainties through nuisance parameters. This statistical inference procedure allows the relative contributions of the different $t\bar{t}+X$ components to be disentangled within the shared phase space, and forms the basis for the final measurement results. This statistical interpretation and the results are detailed in Chapter 9.

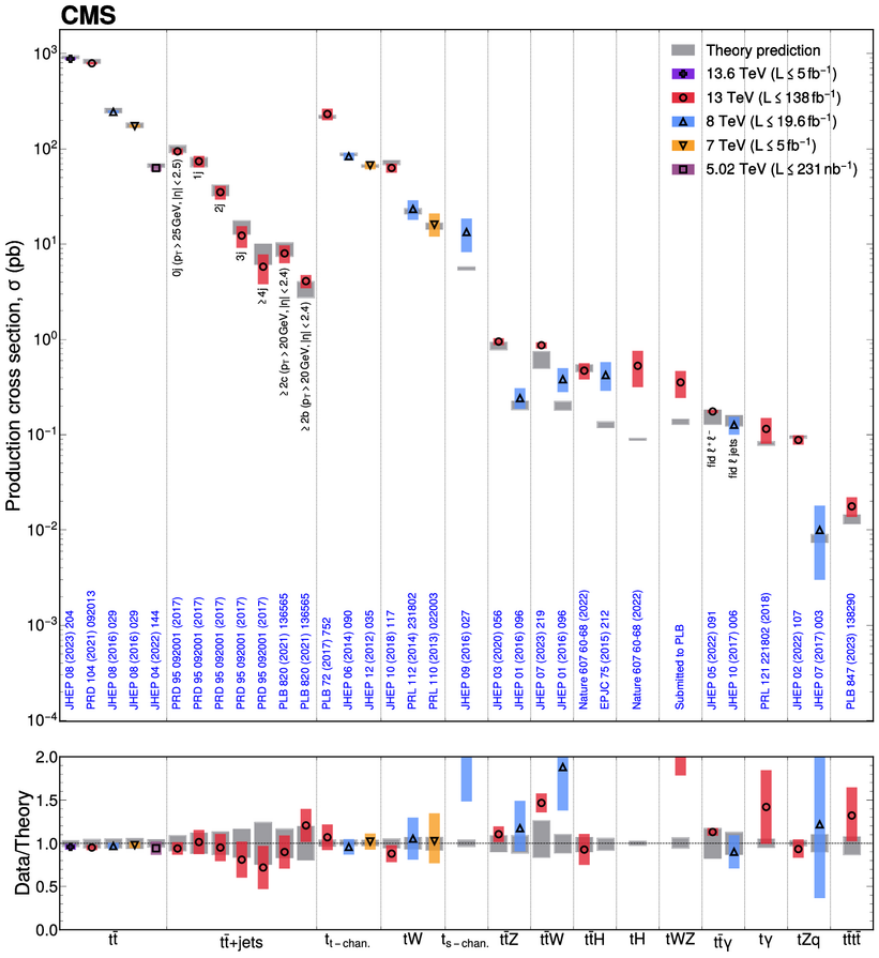


Figure 6.3: Compilation of top-quark production cross-section measurements performed by CMS at different LHC center-of-mass energies. Individual measurements are indicated by distinct markers, with colored bands representing the total experimental uncertainty (statistical and systematic combined). Grey bands correspond to the theoretical uncertainties of the SM predictions. For processes not yet observed, the shaded regions indicate the excluded cross-section range, with the solid line marking the measured 95% confidence level upper limit. Figure taken from Ref. [88].

6.4 Related analyses

Within our research group, a related $t\bar{t}$+X analysis was developed in the dileptonic channel using the 2018 CMS dataset at $\sqrt{s}$ =13 TeV, corresponding to an integrated luminosity of 59.7 fb^{-1} [93]. The analysis presented in this thesis was developed in parallel and targets the single-lepton channel, extending the strategy to the full Run-II dataset. While the dileptonic channel provides a clean experimental signature, the semileptonic channel offers larger statistical sensitivity. However, it is characterized by higher jet multiplicities and increased combinatorial ambiguity, which amplify the impact of $t\bar{t}$ +heavy-flavor backgrounds and modeling uncertainties. The various $t\bar{t}$+X processes contributing to these final states differ in heavy-flavor content and kinematic properties. In the following, the relevant processes are briefly reviewed, starting with $t\bar{t}$+$b\bar{b}$ production, which constitutes the dominant irreducible background in the signal phase space.

6.4.1 $t\bar{t}$+$b\bar{b}$ measurements

The production of $t\bar{t}$+$b\bar{b}$ has been extensively studied at the LHC due to its direct relevance in precision studies of heavy-flavor final states. In addition, this process forms an important background to measurements of associated Higgs production with the Higgs boson decaying to a pair of b-quarks. Both ATLAS and CMS have performed dedicated measurements at $\sqrt{13}$ TeV, primarily in dileptonic and semileptonic final states. The first measurement was performed by CMS using 2.3 fb^{-1} of data in the dileptonic final state [94]. It relied on NLO predictions from POWHEG interfaced with PYTHIA8, where additional b-jets originate from both matrix-element (ME) radiation and parton shower (PS) gluon splitting in the five-flavor scheme (5FS). The process involves multiple characteristic scales, including the top-quark mass and the typically softer scale associated with the additional b-jet pair, leading to sizeable theoretical uncertainties from renormalization and factorization scale choices. Variations of these scales were found to induce significant changes in the predicted cross-section. In a similar description of $t\bar{t}$+$b\bar{b}$, the ATLAS collaboration performed an analysis using 36.1 fb^{-1} of 13 TeV data collected in 2015 and 2016, conducted in both $e\mu$ dileptonic and lepton+jets channels. The measured cross sections were found to exceed the central theoretical values while remaining compatible within uncertainties. An analogous CMS measurement was performed at an integrated luminosity of 35.9 fb^{-1} [95]. The cross-section ratio between $t\bar{t}$+$b\bar{b}$ and $t\bar{t}$ + LF and inclusive $t\bar{t}$ + LF was extracted simultaneously through a binned maximum likelihood fit to the b-tagging discriminant distribution of additional jets. The results were found to be consistent with NLO QCD predictions within uncertainties, although the measured ratio was observed to be moderately larger than the nominal POWHEG prediction. Previous measurements of $t\bar{t}$+$b\bar{b}$ at $\sqrt{s}$ = 7 and 8 TeV have been extensively studied and reported by CMS [96, 97] and ATLAS [98] collaborations.

In addition to measurements in dileptonic and semileptonic final states, CMS has performed a dedicated determination of $t\bar{t}$+$b\bar{b}$ cross-section in the fully hadronic channel using 35.9 fb^{-1} of 13 TeV data [89]. Events with at least eight jets and multiple b tags were selected, and multivariate techniques were employed to suppress the overwhelming QCD multijet background

and to reconstruct the top-quark candidates. The measured total cross section was found to exceed NLO+PS predictions by a factor of approximately 1.5-2.4, corresponding to a deviation at the level of 1–2 standard deviations.

With the full Run-II dataset as well, both CMS and ATLAS have performed precision measurements of $t\bar{t}$+$b\bar{b}$ production. A measurement of inclusive and normalized differential cross sections of $t\bar{t}$+$b\bar{b}$ was performed by the CMS collaboration in the semileptonic channel at an integrated luminosity of 138 fb^{-1} [85]. The nominal signal modeling is based on NLO QCD calculations in the four-flavor scheme (4FS) with massive b quarks, using POWHEG matched to parton showers with dynamic renormalization and factorization scales. Alternative predictions in the five-flavor scheme (5FS), where additional b quarks arise predominantly from parton shower gluon splittings, are used to assess modeling uncertainties. The analysis targets multiple fiducial phase space regions, including fully resolved $t\bar{t} + b\bar{b}$, inclusive $tt + b$, and radiation sensitive $t\bar{t} + b\bar{b}j$. The results are generally consistent with NLO QCD predictions within uncertainties. In addition to the nominal simulation of $t\bar{t}$+$b\bar{b}$ production, several alternative generator configurations were tested, as described in Refs. [85, 99]. An analogous measurement was performed by the ATLAS collaboration using 140 fb^{-1} in the dileptonic $e\mu$ final state with at least 3-4 b-jets, where fiducial integrated and normalized differential cross sections were extracted at the particle level [100]. However, comparisons to state-of-the-art theoretical calculations show that no single prediction simultaneously describes all measured differential observables, although the differences among models remain smaller than the current experimental uncertainties for most distributions.

6.4.2 $t\bar{t} + c\bar{c}$ measurements

The production of top quark pairs in association with charm quark pairs ($t\bar{t} + c\bar{c}$) is a relatively unexplored area compared to $t\bar{t}$+$b\bar{b}$ and $t\bar{t}$+H processes. However, recent developments in charm jet identification and analysis techniques enabled more dedicated studies in this domain. The CMS collaboration conducted the first measurement of the inclusive cross-section for $t\bar{t} + c\bar{c}$ production using dileptonic final states in p-p collisions at $\sqrt{s} = 13$ TeV [101]. This analysis utilized data corresponding to an integrated luminosity of 41.5 fb^{-1}. A novel charm-jet identification algorithm was employed, feeding into a neural network designed to distinguish $t\bar{t}$ events with additional charm, bottom, or light-flavor or gluon jets. Through a template-fitting procedure, the inclusive $t\bar{t} + c\bar{c}$, $t\bar{t}$+$b\bar{b}$, and $t\bar{t} + LF$ cross sections were simultaneously measured, along with their ratios to the inclusive $t\bar{t}$ with two jets cross section. The measured $t\bar{t} + c\bar{c}$ cross section is found to be slightly above the central predictions of generator setups (POWHEG+PYTHIA and MadGraph5_aMC@NLO) used within one to two standard deviations. In practical terms, this corresponds to a fitted signal strength greater than unity, indicating that the simulations tend to underestimate the central value of the observed rate.

A dedicated measurement of $t\bar{t} + c\bar{c}$ was performed by ATLAS using the dataset recorded between 2015 and 2018 at 13 TeV, corresponding to an integrated luminosity of 140 fb^{-1}

[102]. The analysis separates $t\bar{t}+ \geq 2c$ and $t\bar{t}+ \geq 1c$ contributions using a custom b/c-tagging algorithm and a profile-likelihood fit across multiple signal and control regions. The analysis exploits both single-lepton and dileptonic $t\bar{t}$ decay channels, in contrast to the first CMS measurement restricted to the dileptonic final state. This measurement relies on inclusive $t\bar{t}$ samples at NLO accuracy interfaced with POWHEG+PYTHIA8 and MadGraph5_@MCNLO+Herwig7 ([103, 104]). Charm jets arise predominantly from gluon radiation and splitting as modeled by the the matrix element (ME) calculation and parton shower (PS). Comparisons with NLO+PS simulations show that all tested generator setups predict slightly lower rates than observed in data, with deviations up to approximately two standard deviations. The alternative models tested also underpredict the observed value up to 40%. The dominant limitations of the measurement arise from uncertainties in heavy-flavor modeling, b/c-tagging calibration, and the treatment of renormalization and factorization scales. These findings highlight the sensitivity of $t\bar{t}$ production with charm quarks to QCD radiation modeling and PS treatment.

CMS also performed a search for Higgs boson decays to charm quark pairs in association with $t\bar{t}$ using the full Run-II dataset corresponding to 138 fb^{-1} at 13 TeV [105]. The analysis simultaneously probes the decay of Higgs to c- and b-jets, thereby constraining the charm and bottom Yukawa couplings. $t\bar{t}$+H is simulated at NLO accuracy in QCD using POWHEG v2, interfaced with parton showering. The dominant background arises from inclusive $t\bar{t}$ +jets production, which is also generated at NLO in the five-flavor scheme (5FS) using POWHEG v2. In contrast to the inclusive treatment of charm radiation, the $t\bar{t}$+$b\bar{b}$ is modeled using a dedicated sample generated in the four-flavor scheme (4FS) with POWHEG-BOX-RES [92, 106], where additional b-quarks are explicitly included in the ME calculation. Double parton scattering contributions to $t\bar{t}$+$b\bar{b}$ are modeled separately. While $t\bar{t} + c\bar{c}$ production enters as a background component in searches for rare processes such as $t\bar{t}$+H ($H \to c\bar{c}$), its treatment in these analyses differs fundamentally from dedicated heavy-flavor measurements. In Higgs searches, the normalization of $t\bar{t} + c\bar{c}$ is typically allowed to float in the fit to absorb potential mismodeling and improve signal sensitivity. However, the extracted scale factor is not interpreted as a standalone cross-section measurement and does not provide a quantitative test of QCD predictions. In contrast, the measurements performed by CMS and ATLAS mentioned above explicitly target $t\bar{t} + c\bar{c}$ production as a parameter of interest, enabling a direct comparison with theoretical calculations. The present analysis follows the latter approach by simultaneously extracting signal strengths for the $t\bar{t}$+X processes rather than treating heavy-flavor components solely as background normalization parameters.

6.4.3 $t\bar{t} + H(b\bar{b})$ measurements

Using p-p collisions data at $\sqrt{s} = 13$ TeV, both the ATLAS and CMS collaborations have performed several measurements of $t\bar{t}$+H production with the Higgs boson decaying to $b\bar{b}$. Since these measurements target final states with high jet and *b*-tag multiplicity, where the dominant background originates from $t\bar{t}$ +heavy-flavor processes, analyses require sophisticated multivariate techniques and a precise treatment of systematic uncertainties. Previous Run-I

searches at $\sqrt{s} = 7$ and 8 TeV were statistically limited and yielded upper limits on the signal strength of approximately 3-4 times the SM prediction. [107–109].

With the start of Run-II, the increased $t\bar{t}$+H production cross-section of approximately 507 fb significantly improved the expected sensitivity. The CMS collaboration analysed the single-lepton and dileptonic $t\bar{t}$ decay channels separately from the fully hadronic final state. In the former analysis [110] performed using the 2015-2016 dataset corresponding to an integrated luminosity of 35.9 fb^{-1}, events were categorized according to jet and b-tag multiplicity, and multivariate discriminants based on boosted decision trees (BDTs) and ME techniques were employed to enhance signal separation. The $t\bar{t}$+H signal was simulated at NLO accuracy using POWHEG v2, while background processes were generated using POWHEG, MadGraph5_aMC@NLO, and PYTHIA8, depending on the process. Parton showering and hadronization were performed with PYTHIA8. For a Higgs boson mass of 125 GeV, the measured signal strength corresponds to an observed (expected) significance of 1.6 (2.2) standard deviations above the background-only hypothesis. The NNPDF3.0 PDF set was used. The fully hadronic channel [111] exploited the large BR of hadronic $t\bar{t}$ decays but faced substantial combinatorial background and $t\bar{t}$+$b\bar{b}$ contamination. The result is statistically limited and does not show a significant excess over the background-only hypothesis. Similarly, the ATLAS collaboration measured the single-lepton and dileptonic channels using 36.1 fb^{-1} of data [112]. The measured signal strength corresponds to an observed (expected) significance of 1.4 (1.6) standard deviations. In addition to inclusive single-lepton categories, the ATLAS analysis defined a dedicated boosted phase-space region targeting events with high transverse momentum top quarks and Higgs bosons. While the present analysis does not impose an explicit boosted selection, both approaches aim to enhance sensitivity in regions with improved signal-to-background separation. The dominant experimental challenge in both cases remains the modeling of $t\bar{t}$+$b\bar{b}$ production. A summary of the measured signal strengths, μ, for these analyses is summarized in the first three entries of the Table. 6.1.

Table 6.1: Summary of ATLAS and CMS measurements of $t\bar{t} + H(b\bar{b})$ production at $\sqrt{s} = 13$ TeV, including early Run-II and full Run-II results. The table lists the integrated luminosity and the measured inclusive signal strength $\mu_{t\bar{t}+H}$ and, where available, the single-lepton (SL) signal strength $\mu^{SL}_{t\bar{t}+H}$ relevant for this analysis. The abbreviations DL and FH refer to dileptonic and fully hadronic decay channels of $t\bar{t}$, respectively.

Experiment+Channel	Luminosity [fb^{-1}]	$\mu_{t\bar{t}+H}$	$\mu^{SL}_{t\bar{t}+H}$
CMS SL+DL [110]	35.9	0.72 ± 0.45	$0.84^{+0.52}_{-0.50}$
CMS FH [111]	35.9	0.9 ± 1.5	–
ATLAS SL+DL [112]	36.1	$0.84^{+0.64}_{-0.61}$	$0.67^{+0.71}_{-0.69}$
CMS SL+DL+FH [113]	138	0.33 ± 0.26	0.46 ± 0.33
ATLAS SL+DL [114]	140	$0.81^{+0.22}_{-0.19}$	$0.72^{+0.24}_{-0.21}$

Both the CMS and the ATLAS collaborations have performed significantly more precise measurements of $t\bar{t}+H(b\bar{b})$ production at $\sqrt{s} = 13$ TeV with the complete Run-II dataset available. The substantially increased luminosity of 138-139 fb^{-1}, together with refined analysis strategies, enabled a marked improvement in sensitivity compared to the individual era results. The CMS collaboration performed a combined analysis of $t\bar{t}$+H and tH production [113], corresponding to an integrated luminosity of 138 fb^{-1}. The analysis considers three mutually exclusive final states: single-lepton, dileptonic, and fully hadronic top-quark decay modes. Events are categorized according to jet and b-tag multiplicities. In the SL channel, two categories with (≥ 6 jets, ≥ 4 b) and (5 jets, ≥ 4 b) are defined, while the DL channel uses (≥ 4 jets, ≥ 3 b) and (3 jets, 3 b) categories. The FH channel is divided into three signal regions with (7 jets, ≥ 4 b), (8 jets, ≥ 4 b), and (≥ 9 jets, ≥ 4 b). Dedicated artificial neural networks (ANNs) are trained in each category to discriminate signal from background and to classify events according to the most probable production process. The measured inclusive $t\bar{t}$+H signal strength relative to the SM expectation is $\mu_{t\bar{t}+H} = 0.33 \pm 0.26$, which corresponds to an observed (expected) significance of 1.3 (4.1) standard deviations. The total uncertainty is dominated by systematic effects, in particular from the modeling of the $t\bar{t}+b\bar{b}$ background, followed by jet energy scale and b-tagging uncertainties, as also seen in early Run-II measurements. The ATLAS also measured $t\bar{t} + H(b\bar{b})$ using full Run-II data, corresponding to an integrated luminosity of 139 fb^{-1} at $\sqrt{s} = 13$ TeV [87]. The analysis targets only single-lepton and dilepton channels. In the single-lepton channel, both resolved and boosted regions are defined, the latter targeting events with Higgs bosons produced at high p_T. The measured inclusive signal strength is $\mu_{t\bar{t}+H} = 0.35^{+0.36}_{-0.34}$, corresponding to an observed (expected) significance of 1.0 (2.7) standard deviations. In the single-lepton channel, measured values are $\mu_{t\bar{t}H} = 0.30^{+0.43}_{-0.41}$ and $0.32^{+0.61}_{-0.57}$, for resolved and boosted categories respectively. The ATLAS collaboration performed a re-analysis of the same Run-II data [114], which features looser event selection requirements, improved b-jet identification, and a revised treatment of the $t\bar{t}$ +jets background, which constitutes the dominant systematic uncertainty source. The value of $\mu_{t\bar{t}+H}$ is measured to be $0.81^{+0.22}_{-0.19}$ corresponding to an observed (expected) significance of the signal over the background-only hypothesis of 4.6 (5.4) standard deviations.

The observed normalization of the dominant $t\bar{t}+b\bar{b}$ background exceeds the prediction from simulation, consistent with dedicated measurements of this process [85]. This observation highlights the need for improved modeling and continued studies of this production. A noticeable difference among the full Run-II measurements concerns the choice of renormalization and factorization scales used in the simulation of the $t\bar{t}+b\bar{b}$ background. In all analyses, $t\bar{t}+b\bar{b}$ is generated at ME using POWHEG-BOX, interfaced with PYTHIA for PS. The ATLAS analysis [87] initially adopted scale pre-factors of $\kappa_R = 1$ and $\kappa_F = 0.5$. In a subsequent re-analysis, the renormalization scale was lowered to $\kappa_R = 0.5$ while maintaining $\kappa_F = 0.5$. The CMS measurement [113] employed an even smaller factorization scale, using $\kappa_R = 0.5$ and $\kappa_F = 0.25$. Such variations alter the effective strong coupling and the parton density evaluation scale, thereby modifying both the overall normalization and the radiation pattern of the $t\bar{t}+b\bar{b}$

process. Dedicated studies of $t\bar{t}$+$b\bar{b}$ production have demonstrated that different scale choices lead to sizeable shifts in the predicted cross-section, typically at the level of 10-25%. These effects are particularly relevant in the $t\bar{t}$+H phase space, where selections on jet and b-jet multiplicities enhance sensitivity to additional heavy-flavor radiation. In the CMS analysis, the normalization of the $t\bar{t}$+B component is allowed to float in the fit and is found to exceed the nominal prediction. This behavior is consistent with the known tendency of $t\bar{t}$+$b\bar{b}$ simulations to underestimate the measured cross-section. However, since the underlying MC prediction remains tied to the ME configuration, the extracted renormalization factor does not directly translate into the signal strength parameter. The modified scale choice improves the agreement between the predicted and measured $t\bar{t}$+$b\bar{b}$ rates, and is one of the factors responsible for the increase in the measured $t\bar{t}$+H strength in Ref. [114].

A summary of the measured signal strengths, μ, for the different analyses is given in the Table. 6.1. In $t\bar{t} + H(b\bar{b})$ analyses, the overall sensitivity is generally driven by the semileptonic channel, which offers the most favorable compromise between branching fraction and background suppression, and therefore provides the largest statistical contribution to the combined result.

6.4.4 $t\bar{t} + Z(b\bar{b})$ measurements

The production of top quark pairs in association with a Z boson has been measured with high precision in multilepton final states exploiting $Z \to \ell^+\ell^-$ decays by both CMS and ATLAS [115–117]. The leptonic decay channel enables a precise reconstruction of the Z boson through its invariant mass peak and provides strong background suppression. These analyses therefore probe a phase space characterized by clean event signatures and relatively low contamination from heavy-flavor backgrounds. In contrast, the present work explicitly vetoes such leptonic Z decays, as the focus lies on a heavy-flavor–rich environment dominated by multi-jet final states. Complementary studies investigate top-associated production in kinematic regimes where the Z boson carries large transverse momentum [18]. In this boosted configuration, the quark pair from the hadronic Z decay becomes collimated and is reconstructed as a single large-radius jet, typically using jet clustering with a distance parameter $R = 0.8$. These analyses commonly impose requirements of $p_T^Z \gtrsim 200$ GeV, enhancing sensitivity to potential contributions from effective field theory operators, which preferentially modify high-momentum-transfer regions.

The measurement in the hadronic decay channel, $Z \to b\bar{b}$, remains experimentally challenging. A dedicated study of $t\bar{t}$+Z production targeting the hadron decay $b\bar{b}$ was performed in a Master's thesis using 2018 data [118]. The analysis focused on the semileptonic $t\bar{t}$ final state, which employed artificial neural networks trained separately for different jet-multiplicity categories to discriminate the signal from the dominant $t\bar{t}$ +jets background. A signal strength of $\mu = -0.77^{+1.19}_{-1.26}$ was extracted via a binned maximum likelihood fit, corresponding to a 95% CL upper limit of 1.90. In most analyses targeting this topology, the $t\bar{t}$+Z contribution is extracted as a signal-strength parameter within simultaneous fits performed primarily to

measure $t\bar{t}$+H [105], rather than as a standalone cross-section determination. In contrast, the present analysis treats $t\bar{t}$+Z along with other signals, $t\bar{t}$+H, $t\bar{t}$+$b\bar{b}$, and $t\bar{t}+c\bar{c}$ on equal footing, performing a simultaneous extraction of their signal strengths within a common heavy-flavor phase space.

7 Analysis ingredients

To establish the experimental and methodological foundation of this analysis, a well-defined set of ingredients is required. This chapter documents these inputs and analysis choices, providing a transparent definition of the data, simulation, selection strategy, and treatment of uncertainty on which the measurement is based. This chapter is organized as follows. Sec. 7.1 describes the datasets recorded by the CMS detector that is employed in the analysis and the trigger sets used to select the events in the semileptonic final state. Sec. 7.2 describes the simulated MC samples used to model the signal and background processes, including the corrections applied to ensure consistency between simulation and data. Sec. 7.3 defines the reconstructed physics objects used, including leptons (muons and electrons), jets, and missing transverse momentum, together with the corresponding identification, isolation, and calibration criteria. These object definitions ensure a consistent event reconstruction across all data-taking eras and between data and simulation. Sec. 7.4 describes the event-level and object-level corrections applied to the simulation to account for the residual differences between data and MC, ensuring an accurate modeling of detector effects, reconstruction efficiencies, and pileup conditions. Sec. 7.5 presents control distributions of key kinematic and event-level observables used to validate the modeling of data by the simulation and to assess the agreement between data and MC in background-dominated regions. Sec. 7.6 introduces the sources of systematic uncertainty, experimental and theoretical, affecting the analysis and describes their impact on event yields and distributions, as well as their treatment in the statistical interpretation.

7.1 Dataset and triggers

The analysis uses data collected by the CMS experiment, resulting from a dedicated reprocessing of p-p collision data recorded at a centre-of-mass energy of 13 TeV. It corresponds to so-called Run-II Ultra-Legacy (UL) datasets, reconstructed after the original prompt reconstruction and incorporating improved detector calibrations, alignment constants, and refined reconstruction

algorithms. These improvements lead in particular to a better performance of track reconstruction and jet energy measurements, and consequently to an enhanced heavy-flavor tagging performance, which is crucial for analysis targeting top-quark final states.

The Run-II UL data are divided into four distinct eras, between 2016 and 2018, corresponding to a total integrated luminosity of almost 138 fb^{-1}.
The dataset is divided into four eras:

- 2016preVFP (or 2016HIPM), corresponding to 19.51 fb^{-1}.
- 2016postVFP (or 2016noHIPM), corresponding to 16.81 fb^{-1}.
- 2017, corresponding to 41.48 fb^{-1}.
- 2018, corresponding to 59.83 fb^{-1}.

These eras are treated separately in the analysis to account for significant changes in detector conditions and reconstruction performance over time. In particular, the 2016 dataset is split into two sub-eras, commonly referred to as preVFP and postVFP, due to a substantial change in the operating conditions of the silicon strip tracker. During the 2016preVFP period, the strip tracker suffered from saturation effects in the readout electronics [119] under high-luminosity conditions, leading to a reduced signal-to-noise ratio and fewer hits per reconstructed track. This effect was mitigated in the latter part of 2016 by modifying the feedback preamplifier bias voltage (VFP) [120], resulting in improved tracking performance in the 2016postVFP era. For the affected dataset, a dedicated correction, known as the high-ionizing particle mitigation (HIPM), was applied during offline reconstruction. Consequently, as this correction was not available at the High-Level Trigger (HLT) level (Sec. 7.3.1), residual inefficiencies in the trigger performance for this era may be present. Owing to these differences, the two 2016 periods exhibit distinct tracking and flavor-tagging characteristics and are therefore analysed independently. In addition, a major upgrade of the CMS detector, known as the Phase-1 upgrade, was implemented between the 2016 and 2017 data-taking periods. The most significant change relevant for this analysis is the replacement of the pixel detector [35], which increases the number of pixel layers from three to four and reduces the radius of the innermost layer from 44 mm to 29 mm. This upgrade substantially improved vertex reconstruction, impact-parameter resolution, and secondary-vertex identification, further enhancing the performance of b- and c-jet tagging in the 2017 and 2018 datasets. As a result of these detector and reconstruction differences, the 2016, 2017, and 2018 data are treated as separate eras throughout the analysis, with dedicated calibrations, corrections, and systematic uncertainties applied.

Centrally-produced nanoAODv9 samples from Ultra-Legacy reprocessing are used. The nanoAOD format is the final, compact data tier of the CMS event reconstruction chain, designed for physics analyses. It contains high-level reconstructed physics objects, event-level information, and essential metadata, while significantly reducing the data size compared to higher-level formats. The nanoAODv9 version corresponds to the Ultra-Legacy Run-2 reprocessing (version 9) and ensures consistent reconstruction and calibration across all data-taking eras.

To ensure high data quality, certified luminosity sections are selected using the so-called Golden JSON files provided by the CMS Collaboration. These files list the runs and luminosity sections that satisfy all relevant detector and data-quality requirements, thereby excluding periods affected by known detector malfunctions or suboptimal data-taking conditions. The golden JSON files, listed in Table 7.1, are used in this analysis, with the corresponding integrated luminosities.

Table 7.1: Golden JSON files.

Year	Golden JSON	Int. Lumi (fb^{-1})
2016	Cert_271036-284044_13TeV_Legacy2016_Collisions16_JSON	36.31
2017	Cert_294927-306462_13TeV_UL2017_Collisions17_GoldenJSON	41.48
2018	Cert_314472-325175_13TeV_Legacy2018_Collisions18_JSON	59.83

Events with one isolated lepton (electron or muon) in the final state are focused. Data events are collected using single-lepton triggers implemented in the CMS High-Level Trigger (HLT) system (ref. Sec. 7.3.1). For each data-taking era, the lowest transverse momentum (p_T) unprescaled single-lepton trigger paths are used. An unprescaled trigger path accepts all events that satisfy the trigger requirements, without applying an additional downscaling factor, thereby enabling an unbiased selection of events above the trigger threshold. The trigger thresholds are chosen such that the offline lepton selection lies within the plateau region of the trigger efficiency. Thus, for each era, the lowest lepton p_T threshold unprescaled trigger paths are used to collect the data included in this analysis.

7.2 Simulated samples

Monte Carlo (MC) samples are used to model signal and background events. The centrally produced nanoAODv9 samples are used, reprocessed with the ultra-legacy (UL) reconstruction. Event yields are normalized to the integrated luminosity of each data-taking era, using the product of the theoretical production cross sections and the relevant branching ratios, $\sigma \times \mathcal{B}$, as given in Tab. 7.2 and 7.3.

Common simulation setup: Unless stated otherwise, hard-scattering events are generated at NLO QCD with either the matrix element (ME) generators such as POWHEG-BOX v2 [73] or MG5_aMC@NLO [75] and interfaced to PYTHIA 8 [78] for parton showering (PS), hadronization, and the underlying event using the CP5 tune [80]. The nominal PDF set is NNPDF3.1 (NNLO) [71]. ME–PS matching and merging follow the generator defaults (e.g. FxFx for multi-jet NLO samples). Detector response is modeled with GEANT4 [81]. The simulated pileup profile is reweighted to match the data.

7.2.1 Signal processes

$t\bar{t}$+jets simulation

Inclusive $t\bar{t}$ +jets production is modeled with POWHEG-BOX v2 at NLO QCD and interfaced to PYTHIA 8.240 for parton showering, hadronization, and the underlying event (CP5 tune) [73, 74]. At NLO, ME includes the $t\bar{t}$ system plus at most one additional real emission (quark or gluon) beyond the top pair. Additional radiation, including heavy flavor, is described by the parton shower (e.g. $g \to b\bar{b}$). The proton structure is modeled with the NNPDF 3.1 (NNLO) PDF set in the five-flavor scheme (5FS), where b quarks are treated as effectively massless at the generation level. Renormalization and factorization scales are chosen dynamically on an event-by-event basis,

$$\mu_R = \mu_F = \frac{1}{2}(m_{\mathrm{T},t} + m_{\mathrm{T},\bar{t}}),\ m_{\mathrm{T},i} = \sqrt{m_i^2 + p_{\mathrm{T},i}^2}, \tag{7.1}$$

which ties the perturbative scale to the hardness of the event rather than a fixed value. To regulate the hardest real emission and ensure a consistent ME–PS transition, POWHEG employs the damping parameter h_{damp}, set to $1.379\, m_t$ for Run 2 (with $m_t = 172.5$ GeV). All $t\bar{t}$ decay channels (dileptonic, semileptonic, all-hadronic) are simulated separately.

In the inclusive $t\bar{t}$ +jets samples, any extra $b\bar{b}$ pair arises from the shower (5FS $g \to b\bar{b}$). Where a dedicated $t\bar{t}$+$b\bar{b}$ ME sample is used elsewhere in the analysis, its usage and the associated systematics are documented separately. The inclusive $t\bar{t}$ +jets configuration described here serves as the common baseline for light- and charm-flavor jet activity accompanying the top pair. This is explained further in the following sections.

Table 7.2: Montecarlo signal samples used in the analysis.

Process	Sample	cross-section $\times\mathcal{B}$ [pb]
$t\bar{t}$ + jets	TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8	365.5
$t\bar{t}$ + H($b\bar{b}$)	ttHTobb_M125_TuneCP5 13TeV-powheg-pythia8	0.2953
$t\bar{t}$ + Z($b\bar{b}$)	TTZToLLNuNu_M-10_13TeV-amcatnlo-pythia8	0.2529
	TTZToQQ_13TeV-amcatnlo-pythia8	0.5297
$t\bar{t}$+$b\bar{b}$	TTbb_4f_TTToSemiLeptonic_TuneCP5-Powheg-Openloops-Pythia8	19.3
	TTbb_4f_TTTo2L2Nu_TuneCP5-Powheg-Openloops-Pythia8	4.59
	TTbb_4f_TTToHadronic_TuneCP5-Powheg-Openloops-Pythia8	19.9

$t\bar{t}b\bar{b}$ (within the $t\bar{t}$+jets) simulation

The inclusive $t\bar{t}$ +jets setup does not provide an ME-level description of the $t\bar{t}$+$b\bar{b}$ final state. To model heavy-flavor topologies accurately, we therefore supplement it with a dedicated $t\bar{t}$+$b\bar{b}$ sample generated at ME level in the four-flavor scheme (4FS) with an explicit b-quark mass ($m_b = 4.75$ GeV). The 4FS sample is produced with POWHEG-BOX RES [106] interfaced to OPENLOOPS 2 [121] and showered with PYTHIA 8 (CP5 tune), consistent with the inclusive production. The NNPDF 3.1 (NNLO) 4FS PDF set is used. The POWHEG damping parameter is kept at the Run II nominal value, $h_{\mathrm{damp}} = 1.379\, m_t$ with $m_t = 172.5$ GeV, to

ensure a stable ME–PS interface. Since the additional b quarks are generated explicitly at the matrix–element level, they are no longer treated as massless partons originating from the proton PDFs, and the calculation therefore requires a four-flavor scheme with massive b quarks.

Renormalization and factorization scales are set dynamically to reflect the kinematics of all heavy quarks in the final state (including the b quarks present at ME level), using the definitions of Eqs. 7.2, 7.3 below, with prefactors $\kappa_R = 0.5$ and $\kappa_F = 0.25$. Incorporating the b-quark mass scales at ME level typically leads to larger perturbative uncertainties than in the inclusive $t\bar{t}$ +jets setup [92].

$$\mu_R = \kappa_R \left(\prod_{i \in \{t,\bar{t},b,\bar{b}\}} m_{T,i} \right)^{1/4} = \kappa_R \Big(m_{T,t} \cdot m_{T,\bar{t}} \cdot m_{T,b} \cdot m_{T,\bar{b}} \Big)^{1/4}, \tag{7.2}$$

$$\mu_F = \kappa_F \left(\sum_{i \in \{t,\bar{t},b,\bar{b},g\}} m_{T,i} \right) = \kappa_F \Big(m_{T,t} + m_{T,\bar{t}} + m_{T,b} + m_{T,\bar{b}} + p_{T,g} \Big). \tag{7.3}$$

with $m_{T,i} = \sqrt{m_i^2 + p_{T,i}^2}$ the transverse mass and $p_{T,g}$ the transverse momentum of the radiated gluon (when present). This increase arises because the transverse mass scales of the additional b quarks enter the renormalization and factorization scales, enhancing the sensitivity to μ_R variations in the hard process.

To obtain a complete and well-modeled description of $t\bar{t}$ +jets final states, we combine the inclusive $t\bar{t}$ +jets (5FS, b massless) production with the dedicated $t\bar{t}$+$b\bar{b}$ matrix–element sample (4FS, $m_b = 4.75$ GeV). The former provides robust coverage of the bulk phase space and of light- and charm-flavor radiation, while the latter provides an explicit ME description of additional b quarks. Events are assigned exclusively to one sample using a particle-level ghost-hadron matching in $t\bar{t}$ +jets processes, as detailed in Sec. 5.2.5.

Additional jets denote particle-level jets that do not originate from top-quark decays. We match weakly decaying hadrons to these jets using the ghost technique. If an event contains at least one additional jet ghost-matched to a B hadron, it is taken from the 4FS $t\bar{t}$+$b\bar{b}$ sample. The corresponding phase-space region is vetoed from the inclusive 5FS $t\bar{t}$ +jets sample to avoid double counting. If there is at least one additional jet ghost-matched to a C hadron and no additional B-matched jet, the event is classified as $t\bar{t}$+C and taken from the inclusive 5FS $t\bar{t}$ +jets production. All remaining events (no additional B- or C-matched jet) define the $t\bar{t}$ + LF category and are taken from the 5FS $t\bar{t}$ +jets sample.

This scheme ensures that the $t\bar{t}$+C signal region benefits from an ME-level $t\bar{t}$+$b\bar{b}$ description (4FS), the $t\bar{t}$+C region is consistently modeled within the inclusive $t\bar{t}$ +jets simulation in the 5FS, and the $t\bar{t}$ + LF and $t\bar{t}$+C are smoothly modeled by the high-statistics inclusive $t\bar{t}$ +jets simulation (5FS). By construction, the three categories are mutually exclusive and cover the full $t\bar{t}$ +jets phase space relevant for this analysis.

$t\bar{t}H$ and $t\bar{t}Z$ simulation

Associated production of a top quark–antiquark pair with a Higgs or a Z boson is modeled at NLO accuracy in QCD and interfaced to the same shower/reconstruction chain used elsewhere ($\mathrm{PYTHIA}\ 8.240$ with the CP5 tune, $\mathrm{NNPDF3.1}$ NNLO PDFs). $t\bar{t}$+H events are generated with $\mathrm{POWHEG\text{-}BOX\ v2}$ at NLO. The Higgs boson mass is fixed to $m_H = 125$ GeV. Two samples are employed: one restricted to $H \to b\bar{b}$ decays and one inclusive of all non-$b\bar{b}$ Higgs decays ($H \to$ non $b\bar{b}$), enabling category-specific modeling in the analysis. $t\bar{t}$+Z events are produced with $\mathrm{MG5\_aMC@NLO}$ (v2.6.5) at NLO in QCD [72]. The decays of the Z boson are handled by MadSpin [77] to preserve spin correlations of the final-state leptons/jets. Two complementary samples are used: a leptonic sample ($Z \to \ell^+\ell^-$, $\nu\bar{\nu}$) and a hadronic sample ($Z \to q\bar{q}$). Both are showered with $\mathrm{PYTHIA}\ 8.240$ (CP5) and processed with the same detector simulation and corrections as the rest of the MC simulated samples.

7.2.2 Background processes

$t\bar{t}$LF simulation

The $t\bar{t}$ + LF category is obtained from the inclusive $t\bar{t}$ +jets simulation described in Sec. 7.2.1, after removing the phase-space regions assigned to $t\bar{t}$+C (from the dedicated $t\bar{t}$+$b\bar{b}$ sample) and $t\bar{t}$+C (events with at least one additional C-matched jet but no B). The remaining events, with no additional B- or C-matched jets, define the $t\bar{t}$ + LF category. This sample benefits from the large statistics of the inclusive production and provides a reliable modeling of light-flavor radiation accompanying the top-quark pair, serving as the main background.

Single-top quark production simulation

Single-top quark (and antiquark) contributions to the background are modeled individually for each production mode. Samples for the t-channel and the tW associated production are generated at NLO QCD with $\mathrm{POWHEG\text{-}BOX\ v2}$ and interfaced to $\mathrm{PYTHIA}\ 8.240$ (CP5 tune) for parton showering, hadronization, and the underlying event. The proton structure is described with $\mathrm{NNPDF3.1}$ (NNLO), and dynamic scales are used as in the $t\bar{t}$ setup to reflect the hardness of the event. For tW production, the interference with $t\bar{t}$ is treated with the generator default. The s-channel sample is produced at NLO QCD with $\mathrm{MG5\_aMC@NLO}$ and showered with $\mathrm{PYTHIA}\ 8.240$ (CP5).

Vector-boson production in association with jets simulation

V+jets ($V = W, Z$) backgrounds are simulated with $\mathrm{MG5\_aMC@NLO}$ and interfaced to $\mathrm{PYTHIA}\ 8.240$ (CP5). The central production uses matrix elements with up to four partons at LO, merged with the parton shower using the MLM scheme; the $\mathrm{NNPDF3.1}$ (NNLO) PDFs define the proton structure. The inclusive samples contain both light- and heavy-flavor jet components; no separate heavy-flavor enrichment is generated at the ME level for this analysis.

Diboson production simulation

Diboson backgrounds (WW, WZ, ZZ) are simulated at leading order (LO) in QCD with PYTHIA 8.240 using the CP5 tune, with separate samples for each boson combination. Only leptonic decays of the vector bosons are retained to match the analysis final state. To obtain a more realistic sample, the inclusive cross sections are rescaled to higher-order theoretical predictions, for WW to the NNLO QCD prediction, and for WZ and ZZ to NLO QCD.

$t\bar{t}W$ simulation

The $t\bar{t}W$ background is generated with MG5_aMC@NLO (v2.6.5) at NLO in QCD and interfaced to MadSpin to maintain spin correlations in the W-boson decay. Two samples are used in parallel: a leptonic sample ($W \to \ell\nu$) and a hadronic sample ($W \to q\bar{q}'$).

QCD multijet simulation

QCD multijet events can contaminate the event selection when a jet is misidentified as an isolated lepton, or when heavy-flavor decays inside jets produce non-prompt leptons that pass the selection. To model these instrumental backgrounds, PYTHIA 8 is used with the CP5 tune in exclusive p_T bins. In order to enhance the statistical coverage of regions relevant for non-prompt and misidentified leptons, dedicated generator-level filters are applied. These include samples which are enriched in muon, EM, and heavy-flavor decays ($bc \to e$), which preferentially select events containing leptons from hadron decays or jet misidentifications. The samples are stitched across p_T bins using per-event weights, $w = \frac{\sigma \mathcal{L} \epsilon_{filter}}{N_{gen}}$, with σ the nominal cross section, $\mathcal{L}$ the target integrated luminosity, ϵ_{filter} the generator filter efficiency, and N_{gen} the number of generated events. Because tight lepton identification and isolation selections suppress QCD strongly in the signal region, these samples are primarily used for shape modeling. All signal and background events are modeled using MC samples and are listed in Tabs. 7.2 and 7.3.

7.3 Objects

In this section, the object definitions used in the analysis are described.

7.3.1 Trigger Requirements

As a part of the CMS data acquisition system, High-Level Trigger (HLT) paths are used. They are responsible for reducing the data rate after initial event selection by the hardware trigger (Level-1 Trigger, L1). This analysis uses the HLT paths to select events online in real time, by requiring either a high-p_T isolated muon or a high-p_T isolated electron. The primary datasets and trigger paths used in each data-taking period are listed in Tables 7.4, 7.5, and 7.6.

The triggers listed select at least a single well-reconstructed lepton in each event, passing particular kinematic criteria. The triggers chosen are not pre-scaled, providing the lowest

Table 7.3: MC Background Samples

Process	Sample	cross-section $\times\mathcal{B}$ [pb]
Z+jets	DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8	18610.0
	DYJetsToLL_M-50_HT-70to100_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8	211.21
	DYJetsToLL_M-50_HT-100to200_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8	183.2
	DYJetsToLL_M-50_HT-200to400_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8	55.29
	DYJetsToLL_M-50_HT-400to600_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8	7.846
	DYJetsToLL_M-50_HT-600to800_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8	1.933
	DYJetsToLL_M-50_HT-800to1200_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8	0.831
	DYJetsToLL_M-50_HT-1200to2500_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8	0.183
	DYJetsToLL_M-50_HT-2500toInf_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8	0.004
single top	ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-amcatnlo-pythia8	3.3
	ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8	80.95
	ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8	136.02
	ST_tW_antitop_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8	19.55
	ST_tW_top_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8	19.55
$t\bar{t}$+H	ttHToNonbb_M125_TuneCP5 13TeV-powheg-pythia8	0.2118
QCD	QCD_Pt-30To50_MuEnrichedPt5_TuneCP5_13TeV-pythia8	1361000.0
	QCD_Pt-50To80_MuEnrichedPt5_TuneCP5_13TeV-pythia8	377800.0
	QCD_Pt-80To120_MuEnrichedPt5_TuneCP5_13TeV-pythia8	88620.0
	QCD_Pt-120To170_MuEnrichedPt5_TuneCP5_13TeV-pythia8	21070.0
	QCD_Pt-170To300_MuEnrichedPt5_TuneCP5_13TeV-pythia8	7019.0
	QCD_Pt-300To470_MuEnrichedPt5_TuneCP5_13TeV-pythia8	622.4
	QCD_Pt-470To600_MuEnrichedPt5_TuneCP5_13TeV-pythia8	58.86
	QCD_Pt-600To800_MuEnrichedPt5_TuneCP5_13TeV-pythia8	18.22
	QCD_Pt-800To1000_MuEnrichedPt5_TuneCP5_13TeV-pythia8	3.25
	QCD_Pt-1000_MuEnrichedPt5_TuneCP5_13TeV-pythia8	1.0
	QCD_Pt-30to50_EMEnriched_TuneCP5_13TeV-pythia8	6447000.0
	QCD_Pt-50to80_EMEnriched_TuneCP5_13TeV-pythia8	1988000.0
	QCD_Pt-80to120_EMEnriched_TuneCP5_13TeV-pythia8	367500.0
	QCD_Pt-120to170_EMEnriched_TuneCP5_13TeV-pythia8	66590.0
	QCD_Pt-170to300_EMEnriched_TuneCP5_13TeV-pythia8	16620.0
	QCD_Pt-300toInf_EMEnriched_TuneCP5_13TeV-pythia8	1104.0
	QCD_Pt_80to170_bcToE_TuneCP5_13TeV_pythia8	33700.0
	QCD_Pt_170to250_bcToE_TuneCP5_13TeV_pythia8	2125.0
	QCD_Pt_250toInf_bcToE_TuneCP5_13TeV_pythia8	562.5
$t\bar{t}$W	WJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8	0.2043
	TTWJetsToQQ_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8	0.4062
diboson	WW_TuneCP5_13TeV-pythia8	118.7
	WZ_TuneCP5_13TeV-pythia8	65.5443
	ZZ_TuneCP5_13TeV-pythia8	15.8274
W+jets	WJetsToLNu_HT-70To100_TuneCP5_13TeV-madgraphMLM-pythia8	1443.7
	WJetsToLNu_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8	1434.5
	WJetsToLNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8	383.19
	WJetsToLNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8	51.68
	WJetsToLNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8	12.53
	WJetsToLNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8	5.62
	WJetsToLNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8	1.32
	WJetsToLNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8	0.009
$t\bar{t}$ + jets	TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8	88.3
	TTToHadronic_TuneCP5_13TeV-powheg-pythia8	378.0

Table 7.4: HLT paths used to collect data in 2016.

Dataset	HLT path
SingleMuon	HLT_IsoMu24
	HLT_IsoTkMu24
SingleElectron	HLT_Ele27_WPTight_Gsf
	HLT_Ele25_eta2p1_WPTight_Gsf

Table 7.5: HLT paths used to collect data in 2017.

Dataset	HLT path
SingleMuon	HLT_IsoMu27
SingleElectron	HLT_Ele35_WPTight_Gsf

Table 7.6: HLT paths used to collect data in 2018.

Dataset	HLT path
SingleMuon	HLT_IsoMu24
SingleElectron	HLT_Ele32_WPTight_Gsf

possible lepton p_T thresholds and loose lepton requirements.

The names of the HLT paths encode the specific selection criteria applied at the trigger level. Following the Particle Flow reconstruction algorithm, these include requirements on calorimeter energy deposits, track quality, and lepton isolation (Sec. 5.1.2). The numerical value in the trigger name (`Ele32` and `Mu24`) specifies the transverse momentum threshold in GeV for the lepton. Additional tags in the name indicate further identification or isolation working points imposed at the HLT stage. Events are selected if they satisfy at least one of the single-lepton triggers applied, ensuring efficient collection of signal events. Muons are selected requiring $p_T > 24$ GeV, passing isolation cuts, and electrons are selected requiring tightly identified electrons with $p_T > 32$ GeV.

7.3.2 Muons

In this analysis, two categories of muons are defined: primary muons and veto muons. Primary muons are those used in the semileptonic event selection, while veto muons serve to reject events containing additional leptons.

Primary muons are required to have transverse momenta, $p_T > 26$ GeV, and veto muons to have $p_T > 15$ GeV. Both muon categories are constrained within a pseudorapidity range of $|\eta| < 2.4$. The selection criteria are shown in Table 7.7. Muons are reconstructed either as global muons, which combine matching tracks from both the inner tracker and the muon system, or as tracker muons, where tracks identified in the tracker are matched to short track segments (known as stubs) detected in the muon system [ref]. To ensure high-quality muon identification (ID), primary muons are selected using the tight working point (WP) criteria recommended by the CMS Collaboration [51]. These include compatibility with the primary vertex, enforced by cuts on the transverse impact parameter, $|d_{xy}| < 0.2$ cm, and the longitudinal impact parameter, $|d_z| < 0.5$ cm. These muons are required to be global muons with additional quality cuts on the track fit, matched muon chambers and stations, impact parameters, and hits in the pixel and tracker layers. In contrast, veto muons are selected with looser identification

criteria, allowing for broader acceptance of veto events with additional leptons. The criteria are summarized in Table 7.8.

Additionally, a requirement on the relative muon isolation, denoted by I_μ, is applied. These additional cuts are placed to reject muons from hadron decays and other hadrons misidentified as muons. This quantity is defined as follows:

$$I_u = \frac{1}{p_{T,\mu}} \left(\sum_{\Delta R<0.4} p_{T,ch} + max \left[0, \sum_{\Delta R<0.4} p_{T,nh} + \sum_{\Delta R<0.4} p_{T,\gamma} - \frac{1}{2} \sum_{\Delta R<0.4} p_{T,PU} \right] \right) \quad (7.4)$$

It measures the amount of additional activity surrounding the muon, helping distinguish isolated prompt muons from those embedded in jets or arising from hadron decays. This is calculated using the Particle Flow algorithm by examining all reconstructed particles within a cone of radius $\Delta R < 0.4$ around the muon's direction in the (η, ϕ) plane.
It is evaluated by considering the transverse momentum contributions of particles within a cone of radius $\Delta R < 0.4$ around the muon's direction. The calculation includes contributions from charged hadrons (ch), neutral hadrons (nh), and photons (γ). To account for additional energy from pileup (PU) interactions, the charged hadrons identified as originating from pileup are scaled by a factor of $\frac{1}{2}$, which has been validated in simulations to approximate the neutral pileup contribution, since neutral hadrons cannot be easily linked to the primary vertex. The cut on I_μ, as given in Table 7.7, is essential to reject non-isolated muons, such as those originating from semileptonic decays inside jets, and ensures the selection of clean, isolated signal muons for the analysis.

Table 7.7: Requirements used for muons in the analysis.

	primary muons	veto muons
p_T [GeV]	> 26	> 15
$\lvert\eta\rvert$	< 2.4	< 2.4
muon ID	`tight`	`loose`
I_μ	< 0.15	< 0.25

7.3.3 Electrons

Like muons, electrons are also used as primary and veto electrons in this analysis. Primary electrons are required to have transverse momenta $p_T > 34$ GeV, while veto electrons must satisfy $p_T > 15$ GeV. Both categories are limited to the pseudorapidity range $|\eta| < 2.4$. The detailed selection criteria are summarized in Table 7.9.

Electrons are reconstructed by combining calorimeter and tracking information, accounting for their tendency to emit bremsstrahlung radiation due to their light mass and magnetic field. This radiation leads to a trail of bremsstrahlung photons along the curved track in the silicon tracker, while in the electromagnetic calorimeter (ECAL), the electron produces a characteristic

Table 7.8: Identification requirements for muons, the criteria used for tight and loose identification, corresponding to the CMS recommendation [51]

Requirement	tight ID	loose ID
Global muon	True	-
PF muon	True	True
Dlobal or tracker muon	-	True
No. of segments in muon stations	> 1	-
No. of hits in muon chambers	> 0	-
No. of pixel hits	> 0	-
χ^2 /ndf of global muon track fit	< 10	-
$\lvert d_{xy}\rvert$ and $\lvert d_z\rvert$	< 0.2 cm and < 0.5 cm	-
No. of tracker layer hits	≥ 6	-

electromagnetic shower. To capture the full electron energy, the ECAL clusters photon energy deposits into so-called superclusters (SCs), which are then matched to tracks refitted using a Gaussian Sum Filter (GSF) algorithm [122, 123] that models the electron's energy loss. For quality selection, electrons are only considered in the well-calibrated detector regions, excluding SCs located in the pseudorapidity range ($1.4442 < |\eta| < 1.566$) which corresponds to the transition section between barrel and endcap regions of the ECAL. Additionally, cuts on the transverse and longitudinal impact parameters are applied, with thresholds adapted depending on whether the SC lies in the barrel or endcap, ensuring precise association of the electron with the primary vertex and reducing background from misidentified particles, given in Table 7.9.

Identification flags for electrons are used for both `tight` and `veto` working points. The CMS cut-based electron identification relies on a set of observables characterizing the electromagnetic shower shape, the matching between the reconstructed track and the associated ECAL supercluster, and the rejection of non-prompt electrons originating from photon conversions or hadrons misidentified as electrons. Different working points are defined by applying tighter or looser thresholds on these variables, trading efficiency for purity. The electron identification is based on the `cutBasedElectronID-Fall17-94X-V2` version of the electronID categories. In this analysis, the `tight` criteria are applied to primary electrons, while the `veto` criteria are used for veto electrons.

An additional identification criterion, analogous to that of muons, is applied to ensure high-quality electron identification, as recommended by the CMS Collaboration [42]. These criteria include selections on impact parameters, conversion rejection, and missing hits in the inner tracker. The exact requirements are listed separately for the ECAL barrel and endcap regions in Tables 7.10 and 7.11, reflecting the different detector responses in the two regions. A relative isolation variable, I_e, is defined like that of muons, by summing the transverse momenta of charged hadrons (CH), neutral hadrons (NH), and photons within a cone of $\Delta R < 0.3$ around

the electron's direction, expressed as,

$$I_e = \frac{1}{p_{T,e}} \left(\sum_{\Delta R<0.3} p_{T,ch} + max \left[0, \sum_{\Delta R<0.3} p_{T,nh} + \sum_{\Delta R<0.3} p_{T,\gamma} - \rho A_{\text{eff}} \right] \right) \quad (7.5)$$

To suppress pileup effects, particularly in the neutral hadron contribution, an effective area correction is applied in the form of ρA_{eff}, where ρ is the event-level energy density and A_{eff} corresponds to the electron's isolation footprint. This correction accounts for the increased spread of energy for electrons with high p_T, and the isolation thresholds are therefore parametrized as a function of the electron's transverse momentum. Unlike in the muon case, this relative isolation is integrated into the CMS-provided identification working points for electrons (Table 7.10 and 7.11).

Table 7.9: Requirements used for electrons in the analysis

	primary electrons	veto electrons
p_T [GeV]	> 34	> 15
$\lvert\eta\rvert$	< 2.4	< 2.4
$\lvert\eta_{SC}\rvert$	< 1.4442 OR > 1.556	
IP cuts [cm]	$\lvert d_{xy}\rvert < 0.05$ AND $\lvert d_z\rvert < 0.1$ $(\lvert\eta_{SC}\rvert \leq 1.479)$ $\lvert d_{xy}\rvert < 0.1$ AND $\lvert d_z\rvert < 0.2$ $(\lvert\eta_{SC}\rvert > 1.479)$	
cut-based ID	`tight`	`veto`

7.3.4 Jets

Jets used in this analysis are reconstructed from particle-flow (PF) candidates [61] using the anti=k_T clustering algorithm [62, 124] with a distance parameter of $R = 0.4$. To avoid overlaps with identified leptons, any jet candidate within a cone of $\Delta R \leq 0.4$ around a selected electron or muon is excluded from the jet selection. The selected jets are required to have a transverse momentum $p_T > 30$ GeV and lie within the pseudorapidity range $|\eta| < 2.4$. Additionally, jets must pass the `tightLepVeto` jet identification criteria. For jets with $p_T < 50$ GeV, a loose pileup jet ID is applied to suppress contributions from pileup interactions. This pileup rejection is based on a boosted decision tree discriminator trained to distinguish jets originating from the primary interaction from those due to pileup. The full set of jet selection requirements is summarized in Table 7.12.

An additional identification criterion, similar to electrons and muons, is applied to jets, as recommended by the CMS Collaboration [46, 125] to ensure well-reconstructed jets identification and veto leptons mis-reconstructed as jets. The criteria used are summarized in Table 7.13. This criterion is referred to as the `tightLepVeto` categorization, for which a flag is available in the official nanoAODs.

Table 7.10: Identification requirements for electrons in the ECAL barrel region ($|\eta_{SC}| \leq 1.479$), the criteria used for tight and veto identification, corresponding to the CMS recommendation [42]. Stricter (looser) thresholds define the tight (veto) working point. The listed variables describe electromagnetic shower shape, track-supercluster matching, conversion rejection, and isolation, with tighter (veto) thresholds defining the tight (veto) working point. The shower shape variable $\sigma_{i\eta i\eta}$ quantifies the lateral spread of the electromagnetic energy deposit in the ECAL, computed from a logarithmically weighted distribution of crystal energies, and provides discrimination between prompt electrons and hadronic backgrounds. Track-supercluster matching variables $|\Delta\eta|$ and $|\Delta\phi|$ ensure spatial consistency between the extrapolated track and the ECAL energy cluster, while the hadronic-to-electromagnetic energy ratio suppresses hadronic contamination. The quantity $|E_{\mathrm{SC}}^{-1} - p_{\mathrm{track}}^{-1}|$ enforces consistency between the track momentum and the supercluster energy. Conversion rejection is achieved by requiring a limited number of missing inner tracker hits and by vetoing candidates compatible with reconstructed photon conversions. The relative isolation variable I_e suppresses electrons embedded in jets.

Requirement	tight ID	veto ID
$\sigma_{i\eta i\eta}$	< 0.010	< 0.035
$\lvert\Delta\eta(SC, track)\rvert$	< 0.0026	< 0.0050
$\lvert\Delta\phi(SC, track)\rvert$	< 0.022	< 0.024
$E_{Hadronic}/E_{EM}$	$< 0.026 + 1.15/E_{SC} + 0.032\rho/E_{SC}$	$< 0.05 + 1.16/E_{SC} + 0.032\rho/E_{SC}$
$\lvert E_{SC}^{-1} - p_{track}^{-1}\rvert$	< 0.159	0.209
No. of missing inner hits	≤ 1	≤ 2
Pass conversion-safe veto	True	True
I_e	$< 0.0287 + 0.506/p_T$ cm	$< 0.198 + 0.506/p_T$

Table 7.11: Identification requirements for electrons in the ECAL endcap region ($|\eta_{SC}| > 1.479$), the criteria used for tight and veto identification, corresponding to the CMS recommendation [42]. The selection relies on the same classes of observables as used in the barrel region, including EM shower shape variables, track-supercluster matching, hadronic energy leakage, conversion rejection, and lepton isolation. The working points differ from those used in the barrel case due to the distinct detector geometry, reduced endcap granularity, and material budget.

Requirement	tight ID	veto ID
$\sigma_{i\eta i\eta}$	< 0.035	< 0.046
$\lvert\Delta\eta(SC, track)\rvert$	< 0.0050	< 0.0081
$\lvert\Delta\phi(SC, track)\rvert$	< 0.024	< 0.0190
$E_{Hadronic}/E_{EM}$	$< 0.019 + 2.06/E_{SC} + 0.183\rho/E_{SC}$	$< 0.05 + 2.54/E_{SC} + 0.183\rho/E_{SC}$
$\lvert E_{SC}^{-1} - p_{track}^{-1}\rvert$	< 0.019	0.132
No. of missing inner hits	≤ 1	≤ 3
Pass conversion-safe veto	True	True
I_e	$< 0.0445 + 0.963/p_T$ cm	$< 0.203 + 0.963/p_T$

Table 7.12: Requirements for jets used in the analysis

	jets
p_T [GeV]	> 30
$\lvert\eta\rvert$	< 2.4
$\Delta R_{\ell,j}$	> 0.4
jet ID	`tightLepVeto`
PileUp ID	`loose` (for $p_T < 50$ GeV)

Table 7.13: Identification requirements for jets, the criteria used for tight identification, corresponding to the CMS recommendation [46, 125]

Requirement	tight ID
Charged EM energy fraction	< 0.99
No. of neutral particles	-
Neutral hadron energy fraction	< 0.9
Neutral EM energy fraction	< 0.9
No. of constituents	≥ 2
Muon energy fraction	< 0.8
Charged hadron energy fraction	> 0
No. of charged hadrons	≥ 1

7.3.5 b-tagging

The DEEPJET heavy flavor jet tagging algorithm [66] is used for the identification of b-jets in this analysis. A jet is tagged as a b-jet if it passes the `medium` working point (WP) of the DEEPJET classifier. The medium WP corresponds to a 1% misidentification rate for light-flavor jets. The b-tagging discriminant values for the different WPs are given in the Table 7.14. There is no selection made on the b-jet multiplicity directly, however, the DEEPJET tagging scores of individual jets are utilized later in the analysis to enhance the separation between signal and background.

Table 7.14: DEEPJET b-tagging algorithm discriminant values used as working points

Year	Medium WP	Tight WP
2016preVFP	0.2598	0.6502
2016postVFP	0.2489	0.6377
2017	0.3040	0.7476
2018	0.2783	0.7100

7.3.6 Missing Transverse Momentum (MET)

To account for the neutrino produced in the e+jets and μ+jets channels, which go undetected in the detector material, a cut on missing transverse energy of MET ≥ 40 GeV is made. The MET is reconstructed using the PF algorithm, with the PUPPI (PileUp Per Particle Identification) [126] weighting scheme, and serves as a key observable in this analysis. Unless stated otherwise, the MET used throughout this analysis refers to PUPPI MET.

7.4 Event corrections

7.4.1 MET Flags

To ensure the accuracy and reliability of the MET measurement, a set of several event-level filters is applied to remove events affected by known sources of detector noise and reconstruction issues. The specific MET filters used for both simulation and data are listed in Table 7.15, which are recommended by the JET/MET POG [127]. Events that do not satisfy all required conditions are vetoed.

7.4.2 Level-1 (L1) pre-fire

L1 prefiring refers to the incorrect assignment of trigger objects to an earlier bunch crossing than the one in which they were actually produced. This misassignment leads to a veto of the true bunch crossing by the Level-1 trigger system, resulting in the loss of otherwise valid events. The effect arises due to limitations in the timing resolution of subsystems such as the ECAL and the muon detectors.

Table 7.15: MET requirements used in the event selection. Events that do not satisfy all required conditions are vetoed.

Year	Flags

2018	`Flag_goodVertices`
	`Flag_globalSuperTightHalo2016Filter`
	`Flag_HBHENoiseFilter`
	`Flag_HBHENoiseIsoFilter`
	`Flag_EcalDeadCellTriggerPrimitiveFilter`
	`Flag_BadPFMuonFilter`
	`Flag_ecalBadCalibFilter`
	`Flag_eeBadScFilter`

In the ECAL, energy deposits at high pseudorapidity can be mistakenly attributed to the previous bunch crossing, particularly when the electromagnetic energy is high. Since the L1 trigger does not allow consecutive bunch crossings to fire simultaneously, this misassignment can cause the correct one to be vetoed.

A similar issue arises in the muon system, where the limited timing resolution can lead to incorrect bunch crossing tagging of muon candidates. While the magnitude of this effect is relatively small (typically up to 3%), it can affect a broad η range and is therefore non-negligible. As these pre-firing effects are not simulated, a correction is applied to simulated events via an event weight. This weight is computed based on the pre-firing probability of each offline photon, jet, and muon in the event, as a function of their pseudorapidity and transverse momentum. The event weight is defined as the probability that none of the objects in the event causes a pre-firing:

$$\omega = 1 - P(\text{prefiring}) = \prod_{i \in \{\text{photons, jets, muons}\}} \left(1 - \varepsilon_i^{\text{pref}}(\eta, p_T)\right) \tag{7.6}$$

where ϵ_i^{pref} is the prefiring probability of object i. The tool also provides variations for systematic uncertainties by shifting these probabilities up and down. In this analysis, the prefiring weight centrally stored in nanoAODs (starting from version v9), which includes both ECAL and muon prefiring corrections, is applied. The scale of the effect is between 0% and 3% and is η dependent [128].

7.4.3 Pileup

As described in Sec. 5.2.4, pileup interactions are included in the event simulation to provide a realistic description of the LHC collision environment. However, the distribution of pileup interactions assumed in the simulation generally differs from the actual pileup profile observed in data. Variations in the instantaneous luminosity during the run, for example, can cause significant deviations. To account for this discrepancy, a correction procedure known as pileup

reweighting is applied. These are derived as scale factors between the distributions of pileup interactions in data and simulation. The weights are obtained from the ratio of the normalized pileup multiplicity distributions in data and simulation,

$$w_{\mathrm{PU}}(n) = \frac{P_{\mathrm{data}}(n)}{P_{\mathrm{MC}}(n)}, \tag{7.7}$$

where $P_{\mathrm{data}}(n)$ and $P_{\mathrm{MC}}(n)$ are the normalized distributions of the number of reconstructed vertices n in data and simulation, respectively. The weight $w_{\mathrm{PU}}(n)$ is applied to each simulated event according to its pileup multiplicity. The data distribution is obtained using measures of instantaneous luminosity, together with the assumption of a total inelastic p-p cross section at $\sqrt{s} = 13$ TeV. More information can be found in Ref. [129]. Since the pileup conditions differ across data-taking periods, independent corrections are derived for each era (2016preVFP, 2016postVFP, 2017, and 2018). Fig. 7.1 shows the distribution of the number of reconstructed primary vertices after applying the pileup reweighting. While the correction significantly improves the agreement between data and MC, residual mismodeling remains. This is particularly visible at high vertex multiplicities. A similar mismatch between data and simulation in the distribution of primary vertices has also been reported in other analyses, for example, in the semileptonic $\mathrm{t\bar{t}}+\mathrm{b\bar{b}}$ phase space [99]. Since the pileup reweighting procedure does not alter other kinematic observables relevant to this analysis, and the number of reconstructed primary vertices is not directly used in the event selection, the impact is considered negligible. Thus, no further correction is applied.

7.4.4 Trigger efficiencies

In addition to reconstruction and identification, differences in trigger performance between simulation and data are accounted for using scale factors (SFs). The event selection in this analysis relies on single-lepton triggers, as introduced in Sec. 7.3.1. To correct for residual discrepancies, trigger efficiencies are calculated separately for electrons and muons. These are provided as SFs. The efficiency is measured using a tag-and-probe method in control samples enriched in $Z \to \ell\ell$ decays, in which one well-identified lepton (the tag) ensures the event selection and the second lepton (the probe) is used to test the trigger under study. By using independent reference triggers, unbiased efficiency measurements can be obtained and translated to the analysis phase space. For electrons, dedicated studies in $\mathrm{t\bar{t}}$ enriched regions are also performed to ensure reliable performance in the semileptonic channel. In both lepton channels, the efficiencies in data and simulation are found to be very similar, resulting in SFs close to unity. These SFs are applied as a function of the lepton p_T and pseudorapidity, ensuring consistent trigger modeling in the simulated samples used for this analysis.

7.4.5 Lepton reconstruction efficiencies

Differences in the reconstruction and identification efficiencies of leptons between simulation and data are corrected using SFs provided centrally by the CMS Collaboration. These corrections are

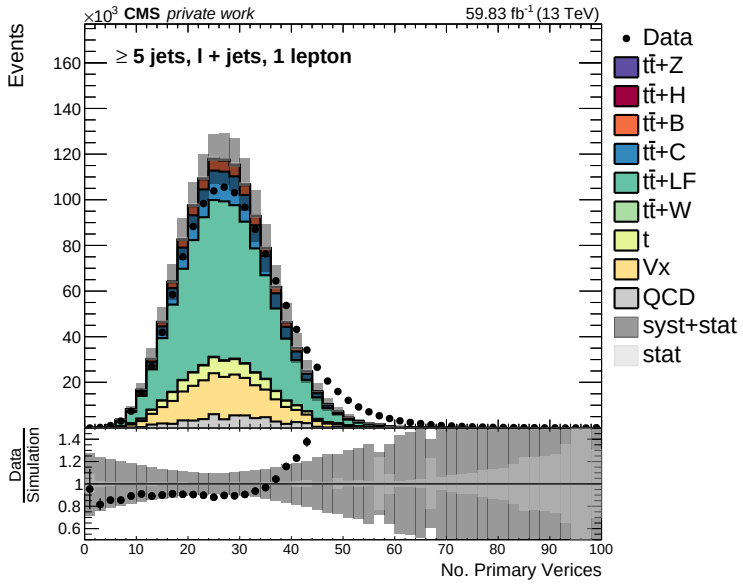


Figure 7.1: Distribution of the number of reconstructed primary vertices in data and simulation after applying pileup reweighting. The plot is shown for the 2018 data-taking era, with the simulation corrected to match the pileup profile observed in data. The shaded bands indicate the total uncertainty in the simulation, including statistical and systematic uncertainties (Sec. 7.6), evaluated for the 2018 era of Run-II. The corresponding distributions of 2017, 2016postVFP, and 2016preVFP are provided in Fig. B.

derived in control samples enriched in $Z \to \ell^+\ell^-$ decays, where here, $\ell = e, \mu$. For electrons, the total efficiency is expressed as

$$\epsilon_e = \epsilon_{\text{ID}} \cdot \epsilon_{\text{reco|ID}} \cdot \epsilon_{\text{trig|reco}}, \tag{7.8}$$

where the subscripts indicate that the efficiencies are conditional on each other, corresponding to the identification (ID), reconstruction (reco), and trigger (trig) efficiencies. The centrally provided dedicated SFs for electrons can be found in Ref. [130].
For muons, the corresponding efficiency includes an additional isolation component and is given by

$$\epsilon_\mu = \epsilon_{\text{ID}} \cdot \epsilon_{\text{iso|ID}} \cdot \epsilon_{\text{reco|iso}} \cdot \epsilon_{\text{trig|reco}}. \tag{7.9}$$

In this case, separate SFs are provided for isolation efficiencies, as the isolation criteria given in Sec. 7.3.2, are defined independently of muon identification. As with the electron efficiencies, the centrally provided SFs correct the difference between the data and simulation efficiencies and are provided by the CMS Collaboration [131]. In both cases, the SFs are applied as multiplicative correction factors to simulated events to reproduce the observed lepton efficiencies in data.

7.4.6 Top p_T spectrum

The previous CMS measurements have shown that p_T spectrum of the top quark transverse momentum in $t\bar{t}$ simulated events, when generated at next-to-leading-order (NLO) in the matrix element calculation, tends to be harder than what is observed in data [132]. This mismodeling persists across different channels, including the semileptonic final state considered in this thesis. Predictions at next-to-next-leading order (NNLO) in QCD provide a significantly improved description of the measured distributions. To mitigate this discrepancy, a reweighting procedure is applied to simulated $t\bar{t}$ and $t\bar{t}$+X events. The method maps the NLO top-quark p_T spectrum onto the corresponding NNLO prediction by assigning an event weight,

$$w_{\text{NNLO/NLO}} = \sqrt{SF(p_T^t)\, SF(p_T^{\bar{t}})}, \tag{7.10}$$

where $SF(p_T)$ denotes a scale factor derived from the ratio of NNLO to NLO predictions as a function of the top quark transverse momentum. The parameterization of $SF(p_T)$ follows the functional form

$$SF(p_T) = \alpha\, e^{-\beta\, p_T/\text{GeV}} + \gamma\, \frac{p_T}{\text{GeV}} + \delta, \tag{7.11}$$

with parameters $\alpha = 0.103$, $\beta = 0.0118$, $\gamma = -0.000134$, and $\delta = 0.973$ [133]. The correction predominantly affects the high p_T region of the top quark p_T spectrum, where the discrepancy between simulation and data is most pronounced, while leaving inclusive observables largely unaffected. This ensures a more accurate modeling of the kinematics of the $t\bar{t}$ system, which is particularly important in analyses involving additional jets from heavy-flavor or boson production, such as $t\bar{t}$+H and $t\bar{t}$+Z processes.

7.4.7 Jet energy corrections

In high-energy particle physics experiments, jets must be calibrated to accurately reflect their true energies. This calibration process, known as Jet Energy Corrections (JECs), is crucial for reducing systematic uncertainties in physics measurements and enhancing the overall quality of reconstructed events. The CMS employs a multistep JEC procedure that corrects jets from various distortions. Fig.7.2 shows the consecutive stages of the JEC applied to both simulated and real data. These include the corrections for pileup, detector response, residual data-MC discrepancies, and jet flavor effects [46]:

- The **pileup offset correction** step addresses energy contributions from pileup interactions. CMS utilizes CHS to remove charged tracks from secondary vertices and then applies a refined hybrid jet area method [124] that subtracts residual neutral energy based on jet area and event-level energy density. This technique also accounts for logarithmic dependencies on jet p_T.
- **Simulation-based response corrections** address the non-uniform response of the detector to hadrons across η and p_T. They are derived from detailed GEANT4 [81] simulations and validated through various generator tunes.

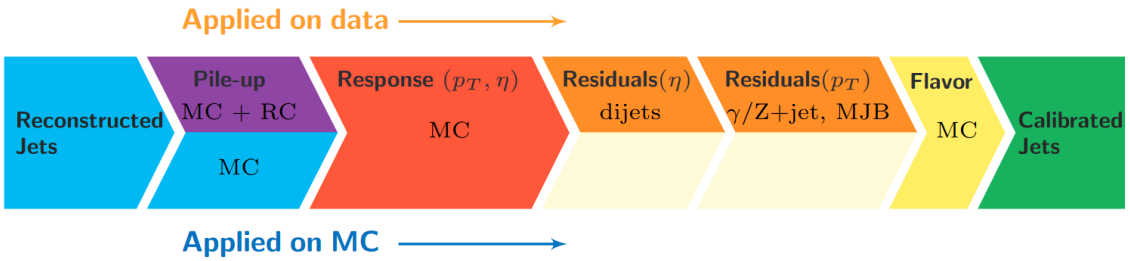


Figure 7.2: The overview of the sequential jet energy correction (JEC) steps applied in both data and simulation. Corrections labeled 'MC' originate from simulation-based studies, 'RC' refers to the random cone method, and 'MJB' indicates corrections derived from multijet balance analysis. Taken from Ref. [46].

- **Residual data corrections** are corrected using in-situ techniques, such as dijet balance for $\eta-$ dependence, and photon/$Z+$ jet events for absolute p_T scale calibration.
- **Flavor and jet size corrections** are applied, since quark and gluon jets fragment differently. This is particularly important for b-tagged jets. Adjustments also account for the distance parameter (R) used in the jet clustering algorithm.
- To manage **uncertainty propagation**, CMS provides a systematic breakdown of each source of error and its correlation across phase space. Systematic uncertainties from all correction stages are rigorously quantified. In central regions, uncertainties fall below 1% for jets with $p_T > 30$ GeV.

Together, these methods yield robust corrections with uncertainties below 3% across most of the phase space [46]. This yields a reliable jet energy calibration, which is vital for accurate cross-section measurements and particle mass measurements.

7.4.8 b-tagging efficiency

The b-tagging scale factors are provided by the BTV POG [134] to rectify the differences in the b-tagging efficiencies between data and simulation. The application of these working-point-based scale factors requires the knowledge of the b-tagging efficiency in simulation, which is corrected to match data using these scale factors [135]. In addition to correcting for the overall b-tagging efficiency, b-tagging scale factors (SFs) also account for the shape of the b-tagging discriminant. This correction is particularly important when the full discriminant distribution is used as an input feature in analyses for graph transformer networks (see Chapter 8). These corrections are typically derived using the iterative Fit (itFit) method, which compares data to simulation across different event topologies, such as $t\bar{t}$ or QCD events, that are specifically chosen for their varied jet-flavor compositions. The SFs are determined as a function of the DEEPJET discriminant, binned by p_T and η [63, 66].

It is generally expected that only the shape of distributions related to the b-tagging discriminant should change. It also affects the distributions of other observables in the analysis,

particularly when the phase space in which the SFs were derived differs from that in which they are applied. For example, observables like the number of jets and the scalar sum of transverse momenta (H_T) can be influenced by the SF application. Since this phase space can differ significantly from the one used in the scale factor measurements, the normalization may not be preserved. Therefore, the total event weights before and after applying b-tag weights, without imposing any b-tag selection, should be compared. This weight is calculated as the ratio of the event weights before and after the application of the SFs [136], expressed as:

$$r_i = \frac{\sum w_{\text{before},i}}{\sum w_{\text{after},i}}, \tag{7.12}$$

where $w_{\text{before},i}$ and $w_{\text{after},i}$ represent the event weights before and after the SFs are applied, respectively. The corrected event weight [135] is calculated as :

$$w_{b,corr} = r_i(n_{\text{jets}}, H_T). \prod_{i}^{n_{\text{jets},i}} w_b(D_i, p_{T_i}, \eta_i), \tag{7.13}$$

where D_i represents the b-tagging discriminant value for the i^{th} jet in an event. This correction compensates for phase-space differences and ensures that the application of SFs does not distort the overall event distribution. The corrected event weight is then used in the analysis, ensuring the consistency of the shape corrections applied to the b-tagging discriminant while minimizing distortions in the number of jets and H_T distributions. Fig. 7.3 shows the effects of applying these shape-correcting SFs on key observables in the 2018 era. The same procedure is applied for the other Run-II eras. These corrections are essential for achieving accurate b-tagging discriminants and for ensuring proper modeling of the data and for a simulation comparison.

7.5 Validation of simulation with control distributions

To validate the event selection and modeling of relevant physics processes, a comprehensive set of control plots is produced for all Run-II data-taking periods. These plots serve as a crucial diagnostic tool, providing visual comparisons between data and MC simulation in the semileptonic inclusive region. These plots are generated both before and after the application of final scale factors and event weights, offering insight into the performance of the selection criteria, the effectiveness of corrections, and the agreement between data and simulation. This allows to quantify the impact of corrections and to ensure that they do not distort unrelated observables, cf. the detailed discussion of shape-correcting SFs in Sec. 7.4.8 and Fig. 7.3. They also provide a sanity check on year-to-year consistency and highlight areas where mismodeling may require additional systematic treatment. The plots are broken down by year, and the relevant kinematic variables for the Jet, lepton, and MET corresponding to the era 2018 are presented in Fig. 7.4. The control distributions corresponding to the 2017, 2016postVFP and 2016preVFP eras of Run-II, are provided in Appendix C.

In the high-jet-multiplicity, heavy-flavor phase space of this analysis, $\mathrm{t\bar{t}}$ +jets in the light

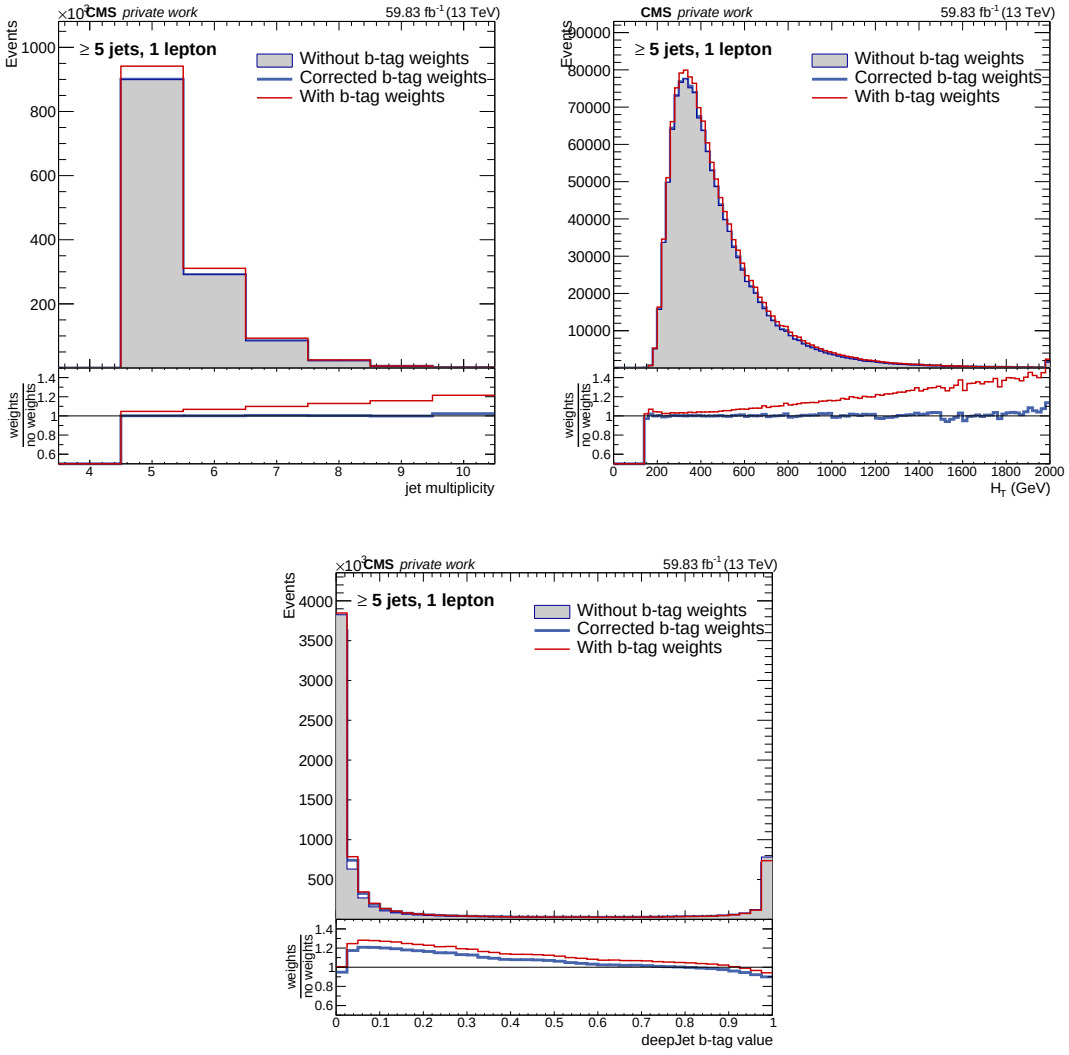


Figure 7.3: Distributions for the 2018 era of the DEEPJET b-tagging discriminant are shown for three scenarios: (i) without applying itFit weights, (ii) with nominal itFit weights, and (iii) with corrected itFit weights. The main plot illustrates the effect on the shape and normalization, while the lower ratio plot highlights the changes introduced by the weights relative to the unweighted case. The plots (top) show the jet multiplicity and the H_T distributions. The application of the itFit b-tagging weights introduces rate and shape modifications relative to the unweighted distribution, as shown in the lower ratio plot. These corrections ensure that the original structure of the observables is preserved. The plot (bottom) shows the b-tag discriminant distribution. Applying the weights modifies both the overall event yield and the distribution shape relative to the unweighted scenario, as reflected in the ratio plot.

flavor category ($t\bar{t}$ + LF) form the dominant background. Subdominant backgrounds include single top, $W/Z/\gamma^*$ + jets (Drell-Yan events), and diboson (WW/WZ/ZZ) production. Other

small 'rare' contributions include those from the associated production of $t\bar{t}$ with W bosons ($t\bar{t}W$) and from QCD multijet. All expectations are normalized to the integrated luminosity of the era using their theoretical cross sections. These represent the predictions before any fitting to the observed data.

In the semileptonic phase space, the simulation reproduces the data in the 5-jet bin (data/MC$\approx$1). At higher jet multiplicities, the prediction exceeds the observation, indicating a sensitivity to the modeling of additional radiation. Examining the yields versus the number of b-tagged jets at the medium working point shows that events with zero b-tags are dominated by background processes and are under-predicted. This is consistent with a modest underestimation of light- and charm-flavor backgrounds (and/or mistag rates) and is observed across all years. In the *b*-tagged jets (medium WP) distribution, we observe a mild over-prediction of the simulation in events with exactly one or two *b* tags for the 2018 dataset. These regions are enriched in $t\bar{t}+c\bar{c}$ and, to a lesser extent, $t\bar{t}+b\bar{b}$, $t\bar{t}+H(b\bar{b})$, and $t\bar{t}+Z(b\bar{b})$ events, including cases where one or more true *b* jets fail the medium working point or fall outside the detector acceptance. By contrast, for the 2017, 2016postVFP, and 2016preVFP datasets, the predictions in the one- and two-*b*-tag bins are broadly consistent with the data within uncertainties, showing no significant excess. At higher *b*-tag multiplicities, the statistical uncertainties increase, and the agreement remains compatible within the larger error bands.

A subset of distributions used as input features for a graph neural network (GNN), which is described in detail in Chapter 8, most notably the *b*-tag discriminants and selected edge-based observables (e.g. distance ΔR and invariant masses of reconstructed objects), are validated with dedicated control plots. The comparisons between data and simulation are quantified using goodness-of-fit (GoF) tests. In the current setup, the charm-vs-light (CvL) and charm–vs–bottom (CvB) discriminants do not satisfy the GoF criterion in the setup and are therefore excluded from the final set of inputs to the GNN.

Overall, good agreement across all eras is observed within the assigned uncertainties. The full list of uncertainties, implementation details, and the correlation model used later in the statistical inference are described in Sec. 7.6. For now, all validation is performed in the year-by-year view.

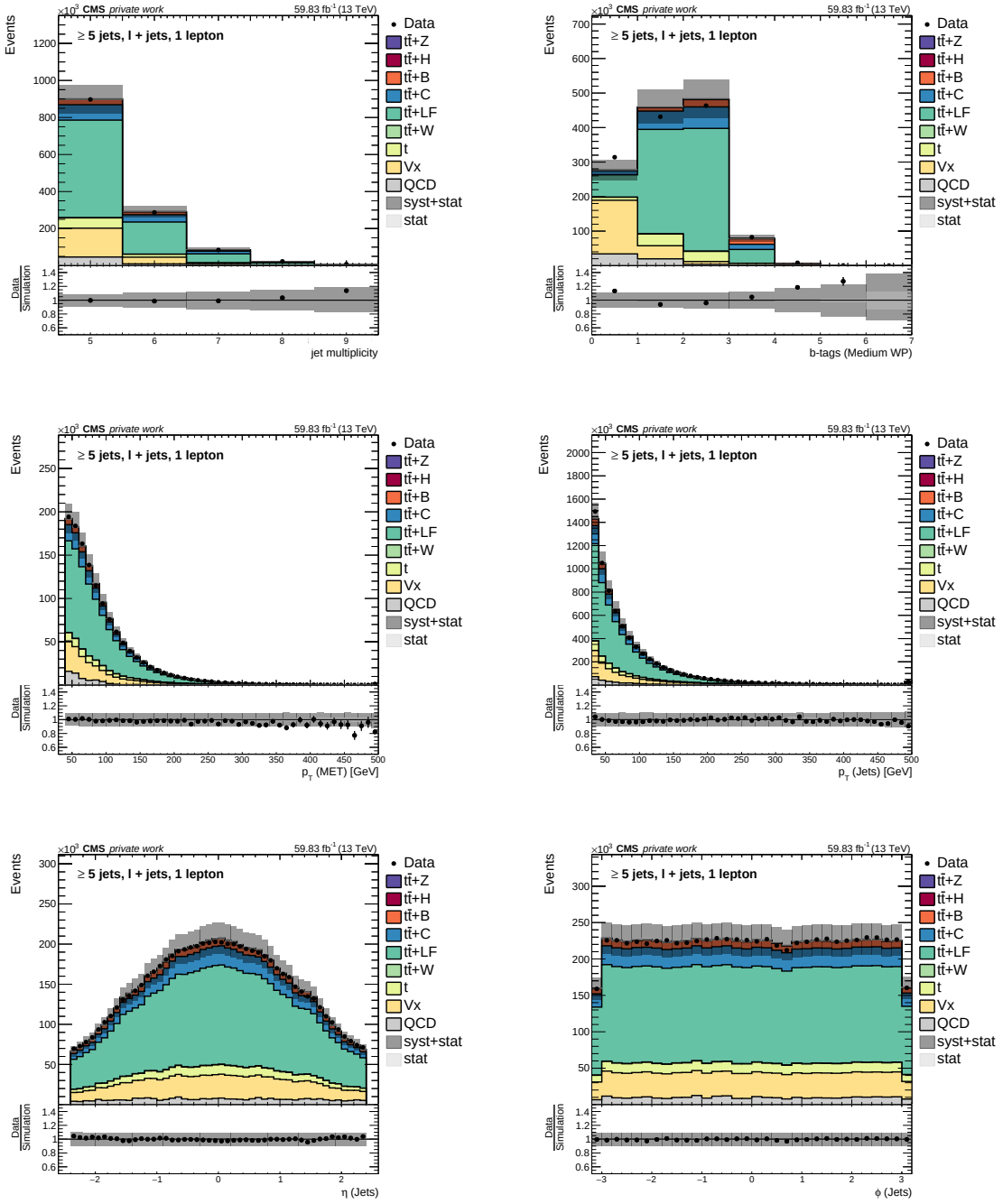


Figure 7.4: Control distributions for the 2018 era in the inclusive ≥ 5-jet phase space of the semileptonic channel. The plots show jet multiplicity, b-tag multiplicity (medium WP), missing transverse momentum, and kinematic distributions of reconstructed jets. The simulation is shown as stacked histograms of all processes, while the data are shown as black points. The gray band represents the total uncertainty on the simulation, including statistical and systematic uncertainties (Sec. 7.6). The lower panel displays the data-to-simulation ratio with the same uncertainty band. Unless stated otherwise, all expectations are pre-fit and normalized to the integrated luminosity using theoretical cross sections.

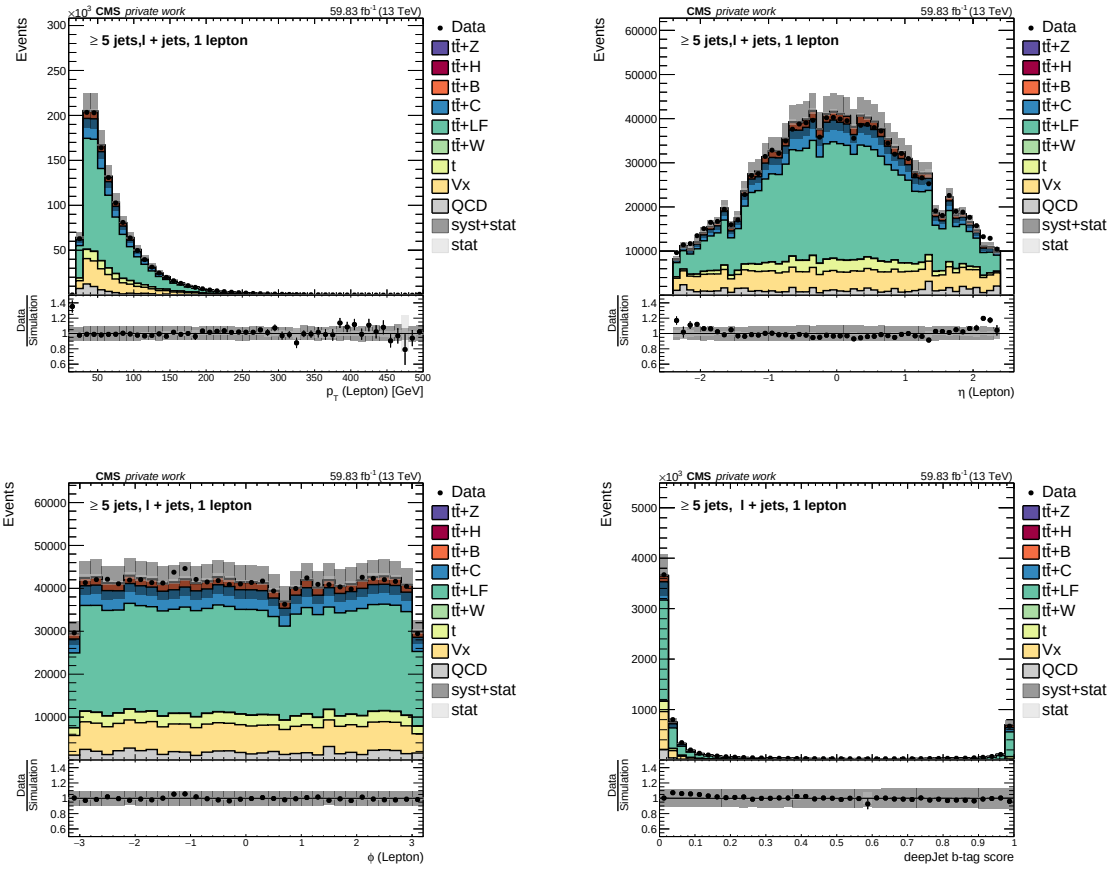


Figure 7.5: Additional control distributions for the 2018 era in the inclusive $\geq$ 5-jet phase space of the semileptonic channel. The plots show the kinematic properties of the reconstructed lepton (p_T, η, ϕ) and the DeepJet b-tagging discriminator. The simulation is shown as stacked histograms of all processes, while the data are shown as black points. The gray band represents the total uncertainty on the simulation, including statistical and systematic uncertainties.

7.6 Systematic uncertainties

Systematic uncertainties quantify imperfect knowledge of detector response, event modeling, and analysis methodology. They affect both the normalization and the shape of key observables and are incorporated as Nuisance Parameters (NPs) that vary the template histograms used in the binned maximum-likelihood fit (ref. Chapter 9). They are classified as experimental, theoretical, or modeling, and methodological sources. Each source is implemented via well-defined variations, with correlations applied across data-taking eras and processes where justified. Impacts are evaluated using pre-fit/post-fit diagnostics (nuisance pulls, constraints) and ranking plots (see Chapter 9). The uncertainty bands shown in this thesis reflect the quadratic sum of all systematic variations described in this section.

7.6.1 Experimental uncertainties

- **Integrated luminosity** measures the overall rate scale for MC-normalized processes. CMS calibrates the absolute scale using van der Meer (vdM) scans and propagates this calibration to physics running via yearly stability or linearity studies. The relative uncertainties used in this analysis are 1.2% (2016), 2.3% (2017), and 2.5% (2018). For the full Run-II dataset, a combined precision of approximately 1.6% is achieved; however, in the statistical model, we apply per-year luminosity nuisance parameters (NPs). Certain uncertainty components are correlated between years and are propagated to all events within the respective data-taking years; see Table 7.16. Each year's luminosity uncertainty is implemented as a rate-changing (log-normal) parameter applied coherently to all processes normalized to the simulation. For more details, see Refs. [137–139].

Table 7.16: Uncertainties in the measurement of the integrated luminosity per data-taking era. 'Cor. 2016/2017/2018' is a fully correlated component shared by all Run-II years, while 'Cor. 2017/2018' is correlated only between 2017 and 2018. Quadratic sums reproduce the published totals of 1.2% (2016), 2.3% (2017), and 2.5% (2018) from Refs. [137–139].

Era	Uncor. 2016	Uncor. 2017	Uncor. 2018	Cor. 2016/2017/2018	Cor. 2017/2018
2016	1.0%			0.6%	
2017		2.0%		0.9%	0.6%
2018			1.5%	2.0%	0.2%

- **Pileup reweighting**: As detailed in Sec. 7.4.3, residual differences between the observed and simulated distributions of the number of additional interactions per bunch crossing (pileup) are corrected by event reweighting. The associated systematic uncertainty is evaluated by varying the inelastic proton–proton cross section used to build the pileup profile in simulation by $\pm 4.6\%$. The resulting alternative weights are propagated to all templates, and the effect is treated as a shape variation in the fit. This variation is taken to be correlated across all simulated processes (i.e., one common pileup parameter affecting every MC process).

- **L1 pre-firing** can assign trigger objects to the wrong bunch crossing, causing events to self-veto, as discussed in Sec. 7.4.2. In simulation, we apply per-object pre-fire efficiencies (photons, muons, jets) and assess the associated uncertainty for each object by adding in quadrature the statistical uncertainty of the efficiency measurement and a flat 20% relative term. The corresponding variations are propagated through the event weights to all analysis templates.

- **Trigger efficiencies** As detailed in Sec. 7.4.4, residual differences between trigger efficiencies in data and simulation are corrected with lepton-specific SFs. The associated systematic uncertainty reflects the SF determination and is dominated by the statistical precision of the tag-and-probe measurements in the control samples. Additional subcomponents, such as possible trigger correlations, phase-space and run-era dependencies, and method variations, are included in the total SF uncertainty but are subleading. The uncertainties are implemented as shape variations (binned in lepton p_T and η) and are taken to be correlated across all simulated processes within a given lepton flavor and data-taking year, electron and muon channels, as well as different years, are treated as independent unless otherwise stated.

- **Lepton efficiencies (ID, iso, trigger)**: As outlined in Sec. 7.4.5, differences between data and simulation in lepton efficiencies are corrected with SFs. The associated systematic uncertainty comes from the SF determination (statistical precision of the control samples plus method/systematic variations). In the fit, these uncertainties are implemented as binned shape variations in lepton p_T and η, with a residual overall rate effect when applicable. NPs are correlated across all simulated processes but treated independently between lepton flavors (electron vs. muon) and between data-taking years. Although their impact is typically small compared to other sources, each component (ID, iso, reco, trig) is kept as a separate NP rather than merged into a single 'lepton' uncertainty, ensuring faithful propagation to the final discriminants.

- **Jet energy scale (JES) and resolution (JER)**: Building on the calibration strategy described in Sec. 7.4.7, uncertainties in the JES and JER are propagated by varying reconstructed jet four-momenta and re-evaluating the full event selection and observables. For each JES uncertainty source, jets and the quantities that depend on them are shifted up and down by the corresponding $\pm 1\sigma$ variation. The resulting templates enter the fit as shape NPs. JER uncertainties are treated analogously by modifying the data-MC resolution smearing within its measured uncertainty and propagating the changes to all distributions. All JEC-related NPs are correlated across simulated processes. Following CMS recommendations [140], a reduced, analysis-relevant set of JES groups rather than every fine-grained source is used. Typical groups include:
  Absolute scale, which accounts for the global jet p_T calibration and, where applicable, era-specific absolute components (denoted as `Absolute` and `AbsoluteYEAR`).
  Relative calibration terms, describe residual p_T and η-dependent effects such as dijet balance (`RelativeBal`) and sample-dependent differences (`RelativeSample`).

Flavor response (Flavor QCD) capturing residual quark/gluon/n-quark initiated jets response differences in the calorimeter.
Residuals such as `BBEC1`, which account for remaining barrel–endcap energy calibration effects not covered by other sources, and their era-specific counterparts [140].
Detector-region specific effects where relevant. For 2018, a dedicated uncertainty is applied to jets reconstructed in the Hadronic Energy Measurement (HEM) region ($-2.5 < \eta < -1.3$, $1.57 < \phi < -0.87$), which experienced an unexpected increase in hadron energy in the CMS detector's forward region during data-taking. This effect is modeled with a 20% energy shift for jets in that window and propagated to all event-level quantities.
Where the CMS central mapping specifies correlations across years, the corresponding JES sources are kept correlated across eras; sources tagged as era-specific (e.g., Absolute2018, BBEC12018) are decorrelated between years. If the analysis is restricted to a single year, all JEC nuisances are correlated within that year by construction. JER uncertainties are one (or a few) sources per era and are kept uncorrelated between years and correlated across processes.

- **b/c tagging SFs**: The response of the b-tagging discriminator in simulation is calibrated to data using the itFit method (Sec. 7.4.8), yielding per-flavor SFs for b, c, and light jets as functions of the discriminant values. Systematic uncertainties from the calibration are propagated as shape variations by reweighting jets according to their true flavor and discriminator value. The model accounts for sample-composition effects (light-flavor contamination in the heavy-flavor region and b/c contamination in the light-flavor region), statistical limitations of the itFit templates (captured by two smooth per-flavor nuisance parameters: linear and quadratic modulations across the discriminator), and charm-specific differences that cover residual c-jet modeling beyond the b-jet calibration. Variations are applied per jet and propagated through b-tag selections, affecting both shapes and effective rates where working-point thresholds enter the event selection. All b/c tagging SF uncertainties are correlated across simulated processes and, unless stated otherwise, decorrelated between data-taking years.

7.6.2 Theory and modeling uncertainties

- **Renormalization and factorization scales (ME)**: Theory uncertainties from missing higher-order QCD terms are evaluated by varying the renormalization (μ_R) and factorization (μ_F) scales used in the matrix-element (ME) calculation. Following the CMS top-physics recommendations [141], scale uncertainties from missing higher-order QCD terms are evaluated from generator matrix-element (ME) weights as two separate nuisance parameters, not via an envelope. Specifically, μ_F is varied by factors 2 and 0.5 while keeping μ_R fixed (up/down shapes), and μ_R by factors 2 and 0.5 while keeping μ_F fixed. Each variation is propagated through the full selection to produce shape and any induced rate effects in all observables. In the statistical model, these are treated as two independent NPs per process, i.e, $t\bar{t}$, $t\bar{t}$+H, $t\bar{t}$+Z, $t\bar{t}$+B, $t\bar{t}$+C, and $t\bar{t}$ + LF correlated

across categories and years for the same process and generator setup, and uncorrelated between different processes.

- **PDF and α_s**: As outlined in Sec. 5.2.1, PDFs encode the proton structure from global fits to data. In simulation, PDF uncertainties are evaluated using the variation weights shipped with each sample. The proton structure is modeled with NNPDF3.1 (NNLO), and PDF effects are propagated in two pieces:

  - **PDF shape (acceptance) uncertainty**: Using the PDF variation weights provided with each sample, per-process shape variations are constructed, retaining any induced normalization effects in accordance with the LHC working group recommendations [90, 142]. For PDF sets represented using the replica method (as in NNPDF, typical for some 4FS configurations), the up/down shapes are defined as the bin-wise RMS (root-mean-square) of the replica predictions with respect to the central value. For PDF sets represented using the Hessian (eigenvector) method (commonly used in 5FS configurations), the up/down shapes are formed by quadratic combinations of the individual eigenvector variations. This difference reflects only how variations are represented, and both probe the same underlying PDF uncertainty.
    The α_s (m_Z) variations provided with the PDF set are treated as a separate nuisance from the PDF shape. PDF-shape and α_s nuisances are correlated across categories and years for samples that share the same PDF set and perturbative order. They are kept independent across different processes only when a different set is used. This follows as per the recommendations in Ref. [143]. Among $t\bar{t}$-associated processes, this variation is taken to be correlated for $t\bar{t}$+H, $t\bar{t}$+Z,$t\bar{t}$+C, $t\bar{t}$ + LF, and $t\bar{t}$W, and uncorrelated with the $t\bar{t}$+B categories.

  - **PDF rate (cross-section) uncertainty**: Following the LHC (Run-II) recommendations for theory-cross-section uncertainties, process-level log-normal (lnN) rate uncertainties are assigned, which are common across years. These are grouped by initial state where appropriate. The values used are $\pm 3.6\%$ for $t\bar{t}$+H, $\pm 3.5\%$ for $t\bar{t}$+Z, $\pm 4.2\%$ for gg-initiated $t\bar{t}$+C, and $t\bar{t}$ + LF,$\pm 2.8\%$ for qg-initiated single-top, $\pm 3.6\%$ for $t\bar{t}$W, $\pm 5\%$ for diboson, $+0.8\%/-0.4\%$ for W+jets (asymmetric lnN), and $\pm 0.2\%$ for Z+ jets. $t\bar{t}$W, diboson, W+jets, and Z+jets are correlated as one group (qq-initiated) [90, 142]. No dedicated PDF rate is assigned to the $t\bar{t}$+B categories in our setup. The QCD multijet sample does not provide a PDF-rate variation here.
    The α_s rate for $t\bar{t}$+H and $t\bar{t}$+Z is uncorrelated between the two signals. The gg-initiated group ($t\bar{t}$ + LF, $t\bar{t}$+C) is correlated within the group, the qg single-top rate is independent, and the qq group ($t\bar{t}$W, diboson, W+jets, Z+jets) is correlated within the group. All are correlated across years for a given process.

- **Parton shower (PS)**: This analysis follows the CMS top recommendations [141] to assess uncertainties from the PS and its interface to the ME. All effects are propagated

to the final templates and treated as shape nuisances. Correlations are set per source as noted below and are correlated across Run-II years for a given process.

- **ISR/FSR scales in the shower**: The strength of initial-state and final-state QCD radiation are governed by the choice of α_s (Q^2) in the shower. We evaluate PS scale systematics using per-event weights, with default variations (factors 2 and 0.5) and NLO compensation terms, including Initial State Radiation (ISR) and Final State Radiation (FSR). In our setup, we use separate ISR NPs for the $t\bar{t}$ categories and each $t\bar{t}$+X signal, and implement them on a per-process basis. A single FSR nuisance is shared across the different $t\bar{t}$ or $t\bar{t}$+X processes and single top. ISR uncertainties are considered uncorrelated, whereas FSR is considered correlated across all processes.
- **ME–PS matching (h_{damp})**: The interface between the fixed-order ME and PS in POWHEG is controlled by the damping parameter h_{damp}, which regulates the hardest real emission and thus the ME–PS matching. The recommendation is to assess uncertainty by comparing with dedicated up/down variations of h_{damp} samples to the CP5-tuned nominal value. Rather than using these dedicated up/down samples (which suffer from limited sample size), we model the h_{damp} uncertainty via a machine–learning (ML) reweighting of the nominal simulation [144]. The nominal setting $h_{damp} = 1.379\, m_t$ is reinterpreted as two alternative configurations, $h_{damp} = 2.305\, m_t$ (up) and $h_{damp} = 0.8739\, m_t$ (down). Event-by-event ML-derived weights emulate the kinematic response to these variations, and the resulting up- and down-templates are treated as shape systematic uncertainties, which are fully propagated through the event selection and final discriminant distributions. The corresponding NP is taken as correlated across all $t\bar{t}$+X processes and across data-taking years, and is not applied to single-top production.
- **Color reconnection (CR)**: CR affects the color flow before hadronization and can modify jet kinematics and b-hadron properties [145, 146]. CR modeling is evaluated by replacing the nominal multi-parton interaction (MPI)-based model with alternative tunes: the QCD-inspired (CR1) and gluon-move (CR2) models, and the variant with early resonance decays (erdON). We take the difference from the nominal as separate shape variations (no envelope). Each CR model (CR1, CR2, and erdON) is treated as a separate, symmetric $\pm 1\sigma$ shape NP. These CR uncertainties are correlated across years. The correlations between processes follow the CMS top recommendations. In this analysis, they are applied to the $t\bar{t}$ samples using the corresponding dedicated productions.

- **Underlying event (UE) tune**: The underlying event includes multiple parton interactions, beam–beam remnants, and soft initial-state activity. It is modeled with the CP5 tune [80]. The UE systematic is assessed by comparing dedicated CP5 up/down variations to the nominal tune using $t\bar{t}$ samples, provided by CMS [80]. The differences are propagated as shape NPs that affect soft observables, such as jet multiplicity, H_T, and MET variables,

as well as any selections sensitive to them. In the statistical model, this enters as a sample-based nuisance. By convention, the UE uncertainty is correlated across data-taking years and, in general, correlated among QCD- and EWK-induced processes. In this analysis, it is applied to the $t\bar{t}$ samples following the same correlation scheme. To avoid double-counting, the UE nuisance is kept independent of color-reconnection and ISR/FSR shower-scale variations.

- **Collinear gluon-splitting**: Collinear $g \to b\bar{b}$ splittings (the $t\bar{t}$ + 2b topology) are numerically delicate in QCD for effectively massless b quarks and are imperfectly modeled in simulation. To cover this, a 50% log-normal rate uncertainty is applied to the $t\bar{t}$ + 2b component, correlated across all $t\bar{t}$+B categories and defined relative to the overall $t\bar{t}$+B normalization. Scale variations already address generic jet-multiplicity effects, but they do not specifically constrain the b-jet multiplicity. Events tagged as $t\bar{t}$ + 2b are identified with an internal classifier (ttbarID==52). Concretely, the event weights are:

$$w^{\text{up}} = \begin{cases} 1.5, & \text{if } \texttt{ttbarID} = 52, \\ 1/\kappa_y, & \text{if } \texttt{ttbarID} \neq 52, \end{cases} \qquad w^{\text{down}} = \begin{cases} 0.5, & \text{if } \texttt{ttbarID} = 52, \\ \kappa_y, & \text{if } \texttt{ttbarID} \neq 52, \end{cases}$$

with year-specific values of κ_y:

$$\kappa_{2016\text{preVFP}} = 1.121, \quad \kappa_{2016\text{postVFP}} = 1.122, \quad \kappa_{2017} = 1.123, \quad \kappa_{2018} = 1.122.$$

In this analysis, templates are not explicitly split by b-jet count. Consequently, potential heavy-flavor mismodeling effects are largely averaged in the inclusive distributions. Nevertheless, the dedicated $g \to b\bar{b}$ uncertainty is retained and propagated, while a single signal-strength parameter is used for all $t\bar{t}$+B processes to avoid over-interpreting residual differences.

- **Top p_T spectrum**: The NLO $t\bar{t}$ simulation is adjusted with an NNLO-motivated reweighting (as discussed in Sec. 7.4.6), to reproduce the measured top-quark p_T spectrum. The associated systematic is defined as a symmetric shape variation about the reweighted nominal: the up (down) template corresponds to removing (re-applying) the full correction, i.e., the size of the correction itself sets the 1σ variation. The weights are applied to all $t\bar{t}$+X samples, and the uncertainty is treated as correlated across the $t\bar{t}$ family and across analysis categories.

- **QCD rate**: Inclusive cross sections are taken from (N)NLO QCD calculations and depend on the μ_R and μ_F scales. Scale–induced shape effects are removed via ME reweighting elsewhere. The residual rate uncertainties are included as lnN terms, using the values adopted in this analysis, and are correlated across years. For $t\bar{t}$+H: $+5.8\%/-8.4\%$, $t\bar{t}$+Z: $+8.1\%/-8.6\%$, $t\bar{t}$ + LF: $+2.5\%/-3.5\%$, $t\bar{t}$W: $+25.5\%/-14.1\%$, $t\bar{t}$+B: $\pm 31\%$, $t\bar{t}$+C: $+2.5\%/-3.5\%$, single-top:$+3.1\%/-2.1\%$, diboson: $\pm 3.1\%$, W+jets: $\pm 3.8\%$, and Z+jets: $\pm 2.0\%$. These calculations are as recommended by the LHC working

groups [90, 142]. Within processes, the vector-boson group (diboson, W+jets, Z+jets) is correlated, and $t\bar{t}$ + LF with $t\bar{t}W$. All other processes are taken to be uncorrelated.

7.6.3 Statistical and methodological uncertainties

Beyond detector and theory systematics, we assign uncertainties to cover analysis-specific choices and finite sample sizes.

- **Statistical uncertainties**: Finite sample size of MC events is propagated via per-bin statistical NPs. We adopt the Barlow-Beeston-lite prescription [147, 148]. Each template bin receives an independent Poisson-constrained nuisance whose prior width is set by the effective yield in that bin. These shape uncertainties allow the fit to vary both the normalization and the distribution locally, ensuring proper coverage in bins with limited statistics. All per–bin statistical nuisances are treated as uncorrelated between bins and processes, and are correlated across categories only when they originate from the same underlying histogram bin.

- **Smoothing of shape systematics**: To mitigate artifacts from limited MC statistics in the construction of shape systematics, we apply a dedicated smoothing and pruning procedure to the variations that most often suffer from bin–to–bin noise, namely the jet–energy calibration (JES/JER), renormalization and factorization scales (μ_R, μ_F), parton–shower scales (ISR/FSR), b-tagging, color reconnection, underlying event, and the h_{damp} variation. The goal is to suppress non-physical fluctuations that can bias the fit or over-constrain NPs, while preserving genuine, slowly varying physics trends. The CMS $t\bar{t}+b\bar{b}$ analysis [85] also employed a similar procedure.
  Smoothing step [149]: For each observable and nuisance, we form an average signed ratio from the up and down templates relative to the nominal prediction N_i in bin i,

$$r_i \equiv \frac{N_i^{(+)} - N_i^{(-)}}{2\,N_i}\,, \tag{7.14}$$

  and smooth $\{r_i\}$ with a robust LOWESS regression [150] to obtain $\widehat{r}_i$. Since the smoothing acts on the average of the two variations, the resulting shapes are naturally symmetrized. The smoothed up/down templates are reconstructed as

$$\widehat{N}_i^{(\pm)} \;=\; (1 \pm s_\pm\,\widehat{r}_i)\,N_i, \tag{7.15}$$

  where the scale factors $s_\pm$ are chosen to best reproduce the original (unsmoothed) variations in an ordinary least–squares sense,

$$\chi^2_\pm \;=\; \sum_i (\widehat{N}_i^{(\pm)} - N_i^{(\pm)})^2 \quad \text{minimized w.r.t. } s_\pm. \tag{7.16}$$

  Operationally, smoothing for JES/JER is performed per era, lepton channel, process, and observable. For h_{damp}, where no strong era dependence is expected, we smooth the

distributions summed over eras and channels.

After smoothing, we test whether a nuisance exhibits a statistically significant shape effect. A χ^2 goodness–of–fit test (accounting for the template statistical uncertainties) is performed between the nominal and the unsmoothed, shape–only variation. If the resulting p–value exceeds 5%, the shape effect is deemed negligible, and the nuisance is converted to a rate–only lnN uncertainty given by the net normalization shift. If the net rate change is compatible with unity within 1%, the variation is discarded. This decision is taken independently for each observable, process, and nuisance. Consequently, the final systematic configuration can differ slightly across fits targeting different observables. This procedure preserves smooth, physics–motivated trends while removing statistical jitter from migration effects. All accepted shape variations are propagated to the final discriminants, converted rate–only terms are implemented as lnN NPs and correlated according to their sources.

7.6.4 Treatment of correlations

Correlations are set based on the underlying physics and how each calibration is derived. Global theory variations (Sec. 7.6.2) are correlated across years and processes. Era-specific calibrations (e.g., JES component splits) are uncorrelated between eras. Tagging SF shapes follow CMS correlation recommendations across flavors and p_T. A (concise) mapping of each NP to its correlation scheme and smoothing status is summarized in Table 7.17.

Table 7.17: Nuisance parameters used in the fit with their type (shape or rate), correlation schemes, and whether the shape variation undergoes the smoothing/filtering procedure of Sec. 7.6.3. "Year corr." indicates correlation across Run-II eras, "Process corr." indicates correlation across processes. Abbreviations: Y = correlated, P = partially correlated

Uncertainty source	Type	Year corr.	Process corr.	smoothing
Integrated luminosity (per year)	rate	P (Tab. 7.16)	Y	-
Pileup	shape	Y	Y	-
L1 pre-firing	shape	Y	Y	-
Trigger efficiencies (SF)	shape	-	Y (per flavor)	-
Lepton ID/iso/reco/trig (SF)	shape	Y	Y (per flavor)	-
JES groups	shape	P	Y	Y
JER	shape	N	Y	Y
b/c tagging SFs	shape	P	Y	Y
ME scales: μ_F (ME weights)	shape	Y	N (per process)	Y
ME scales: μ_R (ME weights)	shape	Y	N (per process)	Y
PDF	shape	Y	P	-
ISR (PS weights)	shape	Y	P	Y
FSR (PS weights)	shape	Y	Y	Y
ME–PS matching (h_{damp})	shape	Y	Y (within $t\bar{t}$+X)	Y
Color reconnection (CR1, CR2, ERD)	shape	Y	P	Y
Underlying event (CP5 up/down)	shape	Y	P	Y
Collinear $g \to b\bar{b}$	shape	Y	Y (within $t\bar{t}$+B)	-
Top p_T	shape	Y	Y (within $t\bar{t}$+X)	-
QCD rate (lnN)	rate	Y	P	-
Template statistical (per bin)	shape	-	-	-

8 Analysis strategy

Modern particle physics analyses increasingly rely on machine learning to compress high-dimensional detector information into compact, discriminative summaries, suitable for physics analyses. In collider experiments, events are inherently relational: physics objects carry attributes, and their pairwise geometry encodes much of the discriminating power. Representing events as graphs and learning on them with Graph Neural Networks (GNNs) provides permutation invariance to object ordering, natural support for variable multiplicities, and an explicit treatment of inter-object relations, properties that are particularly useful at high jet multiplicity in semileptonic $t\bar{t}$+X final states. Numerous studies report superior performance of GNNs over traditional architectures on similar problems [151, 152].

This chapter is organized as follows. Sec. 8.1 introduces the core ideas behind GNNs (graph representation, message passing, and readout) and motivates their use for collider events. Sec. 8.2 defines the analysis phase space and object selection(s) that determine the input graphs. Sec. 8.3 details the model design, including node and edge features, graph construction, attention-based message passing, pooling, and the training setup. Sec. 8.4 presents the node-level predictor (NLP) used to identify additional jets, providing calibrated per-jet scores. Sec. 8.5 describes the graph-level predictor (GLP) for multiclass event classification across the $t\bar{t}$+X categories, built on the same event-graph representation. Finally, Sec. 8.6 quantifies the impact of NLP pre-classification on event-level separation and on the downstream statistical fit, comparing controlled configurations with identical capacity and hyperparameters.

8.1 Introduction to Graph Neural Networks (GNNs)

In supervised learning, a model $f_\theta(x)$ is trained to predict a target y from input features x by minimizing a loss $\mathcal{L}(f_\theta(x), y)$ over labeled examples. Deep neural networks (DNNs) implement f_θ as a composition of linear maps and nonlinearities, learning expressive feature hierarchies directly from data. When inputs have strong spatial structure (e.g., images), convolutional

networks exploit translation equivariance. When inputs are orderless sets of variable size, set networks or transformers can enforce permutation invariance or learn attention over elements. Collider events, however, are neither regular grids nor arbitrary lists. They are naturally relational physics objects that carry attributes, and their geometric relations are often as informative as the objects themselves, especially in high-jet-multiplicity final states. Graph Neural Networks (GNNs) inject this relational inductive bias by operating on graphs whose nodes represent objects and whose edges capture interactions, proximity, or learned affinities. This is particularly advantageous for analyses of events like $t\bar{t}$+X, because GNNs ignore the arbitrary order of objects, handle a variable number of jets and leptons without padding heuristics, learn the important relationships between objects through their edges, and do all of this with far fewer parameters relative to fully connected networks like DNNs on flattened inputs. See [153] for a systematic comparison of GNNs with standard feed-forward DNNs.

A single perceptron, p computes a weighted sum of inputs x followed by a nonlinear activation function $\psi(z)$,

$$p \;=\; \psi\Big(w^{\top}x + b\Big)\,, \tag{8.1}$$

where w is the vector of learnable weights, and b is the bias term. This nonlinearity enables the model to learn complex decisions beyond simple linear separations. Stacking perceptrons yields a multilayer perceptron (MLP), a fully connected feed-forward network that composes affine maps with activations to learn hierarchical features. In this thesis, the hidden layers use a LeakyReLU activation function,

$$\text{LeakyReLU}(z) \;=\; \begin{cases} z, & z > 0, \\ \epsilon\, z, & z \leq 0, \end{cases} \tag{8.2}$$

where ϵ is a small constant that controls the slope for negative values. In this thesis, $\epsilon = 0.1$ is used. And, the output layer for binary classification uses a sigmoid,

$$\sigma(z) = \frac{1}{1 + e^{-z}}, \tag{8.3}$$

so that scores can be interpreted probabilistically. Models are trained by minimizing a supervised loss over labeled examples; for classification, the cross-entropy is used,

$$\mathcal{L} \;=\; -\sum_{i} y_i \,\ln \hat{y}_i, \tag{8.4}$$

and parameters are updated by (stochastic) gradient descent,

$$\theta \leftarrow \theta - \eta\, \nabla_{\theta}\, \mathcal{L}, \tag{8.5}$$

with an adaptive optimizer (Adam) and standard regularization (e.g. dropout) to reduce overfitting. Here, θ denotes the vector of all trainable parameters of the network (weights

and biases across all layers), and $\eta > 0$ is the learning rate that controls the step size of each update. The Adam optimizer maintains exponential moving averages of the first and second moments of the stochastic gradients $g_t = \nabla_{\boldsymbol{\theta}} \mathcal{L}_t$, where θ denotes the trainable parameters and $\mathcal{L}_t$ the loss evaluated at iteration t. The first moment (mean) and second moment (uncentered variance) estimates of the gradients are computed as

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1)\, g_t, \tag{8.6}$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2)\, g_t^{\odot 2}, \tag{8.7}$$

where m_t and v_t denote the exponential moving averages of the gradient and its element-wise square, respectively. Since these estimates are initialized at zero, bias-corrected versions are used,

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \qquad \hat{v}_t = \frac{v_t}{1 - \beta_2^t}, \tag{8.8}$$

The parameters are then updated according to

$$\theta_{t+1} = \theta_t \; - \; \eta \, \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}\,, \tag{8.9}$$

where η is the learning rate and ϵ is a small constant introduced for numerical stability. Typical defaults are $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$. The adaptive normalization $\sqrt{\hat{v}_t}$ scales the update of each parameter according to the observed gradient variance. This formulation follows the original description in Ref. [154].

A central risk during training is overfitting, where the model memorizes the idiosyncrasies of the training data, including statistical noise, resulting in low training loss but degraded performance on unseen data. Signals of overfitting include a widening gap between training and validation losses or unstable validation metrics. To mitigate this, standard regularization is applied. One such method is dropout [155], which randomly deactivates a fraction of hidden units during training to discourage overfitting. Another technique is early stopping based on validation loss, as well as careful data splitting and monitoring of calibration on held-out sets. These measures improve the generalization of both the node-level and graph-level predictors used later in this chapter.

In this thesis, each event is represented as a graph and uses message-passing GNNs to perform object-level classification and build event-level summaries used in downstream statistical inference. The graph formalism and the message-passing paradigm are introduced in Sec. 8.1.1 and 8.1.2, while the implementation details of the specific architectures and inputs are given later in this chapter.

8.1.1 Graph theory and properties

A finite graph is an ordered pair $G = (\mathcal{V}, \mathcal{E})$ consisting of a non-empty set of vertices $\mathcal{V} = \{1, \dots, N\}$ and a set of edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. When $(i, j) \in \mathcal{E}$, the vertices i and j are said to be adjacent. In this thesis, attributed graphs are used. An attributed graph assigns information to all parts of the graph. Each vertex i carries a feature vector $x_i \in \mathbb{R}^{d_v}$, each edge (i, j) may carry features $e_{ij} \in \mathbb{R}^{d_e}$ (pairwise measurements), and the entire graph may include a global attribute u that summarizes event-level quantities. The vertices, or interchangeably called nodes, refer to the same basic object.

With the objects defined, it is convenient to encode how vertices are linked via the adjacency matrix. For a graph with N vertices, the adjacency matrix $A \in \{0, 1\}^{N \times N}$ is

$$A_{ij} = \begin{cases} 1, & (i, j) \in \mathcal{E}, \\ 0, & \text{otherwise.} \end{cases} \tag{8.10}$$

In weighted graphs, the entry becomes $A_{ij} = w_{ij} \geq 0$, reflecting the strength of the relation. For undirected graphs, A is symmetric ($A_{ij} = A_{ji}$). Directed graphs need not be symmetric. The degree of vertex i is $d_i = \sum_j A_{ij}$ and forms the diagonal degree matrix $D = \mathrm{diag}(d_1, \dots, d_N)$. Throughout this thesis, the event graph is taken to be complete and directed without self-loops, i.e. $(i \to j) \in \mathcal{E}$ and $(j \to i) \in \mathcal{E}$ for all $i \neq j$, but $(i \to i) \notin \mathcal{E}$. For message passing on directed graphs, incoming and outgoing neighborhoods are distinguished, and normalizations are used in/out degrees as appropriate.

Each vertex i is associated with a feature vector $x_i \in R^{d_v}$, which encodes the kinematic and flavor-related properties of the corresponding reconstructed physics object. For a graph with N vertices, the node features can be arranged into a matrix $X \in R^{N \times d_v}$, where each row corresponds to the feature vector of a vertex.

Two graphs $G = (\mathcal{V}, \mathcal{E})$ and $G' = (\mathcal{V}', \mathcal{E}')$ are *isomorphic* if there exists a bijection $g : \mathcal{V} \to \mathcal{V}'$ such that

$$(i, j) \in \mathcal{E} \iff (g(i), g(j)) \in \mathcal{E}'. \tag{8.11}$$

For learning on graphs, this translates to a requirement that predictions must not depend on an arbitrary relabeling of vertices. Message-passing architectures enforce these properties by using shared learnable functions and symmetric aggregations such as sum/mean/max.

GNNs support three complementary prediction granularities: Edge-level (ELP), node-level (NLP), and graph-level (GLP) predictions. ELP infer properties of relations, for example, track-building or association tasks. NLP classifies objects based on their features and neighborhood, for example, labeling jets. GLP predicts properties of the entire graph by aggregating information from all nodes and edges into a global representation, such as determining the event class in $t\bar{t}$+X processes. In practice, these tasks differ slightly in learning setup. NLP is often collective

or semi-supervised because unlabeled vertices can benefit from their connections to labeled ones. ELP can be framed as supervised or unsupervised, depending on whether relation labels are available. GLP typically follows a standard supervised learning assumption at the graph level. This thesis employs NLP for per-jet identification and GLP for event categorization. Both tasks are trained in a supervised manner and operate on the same event graph representation, but differ in the readout, node-wise scores vs. the pooled graph embedding. The detailed explanation of the identification of additional jets using NLP is given in Sec. 8.4, and the event categorization, based on the graph as a whole, is detailed in Sec. 8.5.

8.1.2 Message passing mechanism

The learning dynamics of a GNN are driven by message passing, formalized in the Message Passing Neural Network (MPNN) framework [156]. At each iteration, vertices exchange information along edges and update their internal representations, also known as embeddings. Let $h_v^{(0)}$ be the initial embedding of vertex v, which is typically an MLP applied to features x_v. For a directed graph, incoming neighbors are $\mathcal{N}_{\text{in}}(v) = \{u \mid (u \to v) \in \mathcal{E}\}$. One iteration $\ell \mapsto \ell+1$ consists of three steps.

(1) Message construction. Each incoming edge $(u \to v)$ contributes a message computed from the sender/receiver nodes and (optionally) edge attributes:

$$m_{u\to v}^{(\ell)} \;=\; \phi^m\Big(h_u^{(\ell)},\, h_v^{(\ell)},\, e_{uv}\Big)\,, \tag{8.12}$$

where ϕ^m is a learnable function (e.g. a small MLP or a linear map with nonlinearity).

(2) Permutation-invariant aggregation (with attention). Messages arriving at v from all connected nodes are combined and aggregated, ensuring invariance to vertex relabeling. A simple choice is a sum or a mean:

$$\bar{m}_v^{(\ell)} = \text{AGG}\Big(\{m_{u\to v}^{(\ell)} \,:\, u \in \mathcal{N}_{\text{in}}(v)\}\Big)\,, \qquad \text{AGG} \in \{\textstyle\sum,\ \text{mean},\ \max\}. \tag{8.13}$$

where AGG is the aggregation function. In this analysis, a weighted aggregation is used via attention coefficients [157] $\alpha_{u\to v}^{(\ell)}$ that emphasize informative neighbors:

$$e_{u\to v}^{(\ell)} \;=\; a\Big(h_u^{(\ell)},\, h_v^{(\ell)},\, e_{uv}\Big)\,, \tag{8.14}$$

$$\alpha_{u\to v}^{(\ell)} \;=\; \frac{\exp\Big(e_{u\to v}^{(\ell)}\Big)}{\displaystyle\sum_{w\in\mathcal{N}_{\text{in}}(v)} \exp\Big(e_{w\to v}^{(\ell)}\Big)}\,, \tag{8.15}$$

$$\bar{m}_v^{(\ell)} \;=\; \sum_{u\in\mathcal{N}_{\text{in}}(v)} \alpha_{u\to v}^{(\ell)}\, m_{u\to v}^{(\ell)}. \tag{8.16}$$

Here $a(\cdot)$ is a learnable scoring function (e.g. a bilinear form or MLP). The softmax in (8.15) normalizes weights over the incoming neighborhood, which is appropriate for directed graphs.

Because Eqns. (8.13)–(8.16) are symmetric with respect to the set of neighbors, the aggregation is invariant to the ordering, and is robust to variable vertex multiplicity. For more information, see Ref. [152, 158].

(3) State update The vertex embedding is updated using the aggregated message. A generic form is

$$h_v^{(\ell+1)} \;=\; \phi^h\Big(h_v^{(\ell)},\, \bar{m}_v^{(\ell)}\Big)\,, \tag{8.17}$$

with ϕ^h realized by an MLP plus nonlinearity (e.g. LeakyReLU), optionally with normalization and residual connections. Thus, the updated vertex can be calculated using

$$h_v^{(\ell+1)} \;=\; \sigma\Big(W\,[\,h_v^{(\ell)} \,\|\, \bar{m}_v^{(\ell)}\,] + b\Big), \tag{8.18}$$

where $[\,\cdot\|\cdot\,]$ denotes concatenation, W, b are trainable parameters, and σ is a pointwise activation.

The GNN architecture performs multiple message-passing iterations inside a single layer. Thus, the index ℓ denotes the iteration count, regardless of the specific implementation. The directed, complete event graphs used here have edges $(u \to v)$ and $(v \to u)$ for $u \neq v$. Attention weights adapt the effective neighborhood at each vertex and iteration, improving discrimination in high-multiplicity events while preserving permutation invariance and supporting variable object counts.

8.2 Phase space selection

For the training, an inclusive phase space that requires exactly one lepton, at least five jets in the final state, and no additional selection on b-tagged jets is defined. This helps ensure that the focus is on the jet multiplicity and overall event topology, while leaving the b-tagging selection for the ML model to determine. This is also done to maintain a broad phase space, enabling the study of jet interactions and the overall jet behavior in the event. To refine the analysis further, preselection criteria are defined based on several kinematic variables, including lepton p_T, jet p_T, missing transverse energy (MET), jet η and ϕ, and the b-tagging discriminant. These preselections help reduce the background and ensure a signal-enriched dataset. To understand the performance of these criteria, the data-to-MC agreement of the kinematic variables in the 2018 dataset is shown in Sec. 7.5, and some of the selected edge variables are shown in Fig. 8.1.

8.3 GNN model design

In the context of Graph Neural Networks (GNNs), complex physical processes, such as those studied in high-energy physics, can be represented as graphs, in which entities are treated as nodes and interactions between them as edges. Specifically, in the case of particle physics, the particles involved in a process can be represented as nodes, and their interactions as edges connecting

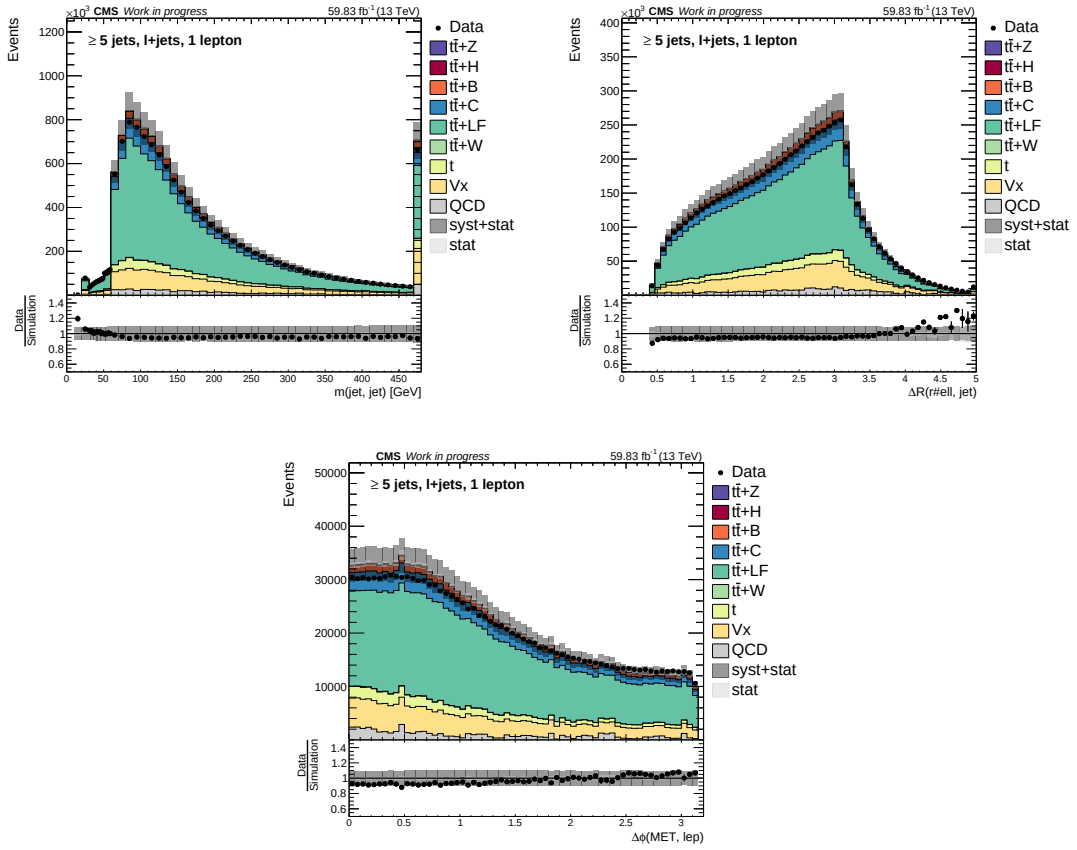


Figure 8.1: Control plots for the inclusive phase space selection, showing various kinematic distributions of the 2018 dataset. The distributions depicted are the edge variables: invariant mass of jets (top left), ΔR between lepton and jet (top right), and $\Delta\phi$ between MET and lepton (bottom).

the relevant nodes. This framework allows GNNs to learn and analyse the intricate relationships between different objects in a collision event. This representation also allows GNNs to leverage the inherent relational structure present in Feynman diagrams, making them well-suited for tasks such as particle identification, event classification, and anomaly detection in high-energy physics.

To facilitate this, different objects from the detector are considered as nodes of the graph. In the semileptonic channel of $t\bar{t}$, a total of eight nodes is defined: six nodes corresponding to the jets, one node for the lepton, and another for the neutrino, accounted for by the missing transverse energy (MET). Thus, seven categories are introduced. As portrayed in Fig. 8.2, the additional jets, the jets produced from (partons) other than $t\bar{t}$ decay, are referred to as 'AddJet'. The category, 'HadTopB', the b-jet in association with the hadronically decaying top quark, is selected as the jet with the highest transverse momentum. The b-jet that is produced in association with the leptonically decaying top quark is labeled 'LepTopB', and it is the jet with the second largest p_T. The next two leading-order p_T jets, which emerge from

the hadronically decaying W boson, are denoted as 'HadTopQ'. Jets that do not fall into the leading order categories are classified as 'Unknown'. The b-jets stemming from gluon splitting, decay of the Higgs or Z boson, are equally assigned to the category of 'AddJ' for this analysis. The global attributes of the graph are not realized for graphs performing jet identification.

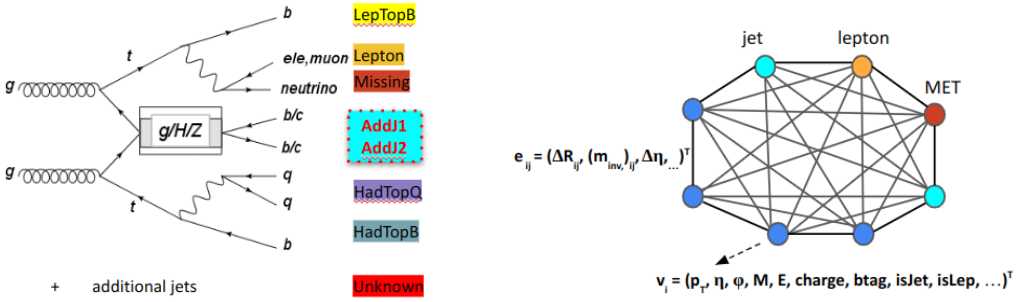


Figure 8.2: The jet categories considered for the GNN training are portrayed on the left. The graph structure illustrates the input passed to the GNNs, generated from reconstructed objects. The figure on the right illustrates the input graph representation with eight nodes, corresponding to each final state object in the semileptonic channel of $t\bar{t}$. The reconstructed objects from the detector are denoted by nodes (circles) and the relationship between the particles by edges (lines connecting circles).

8.3.1 Input features

The input features to the GNN play a crucial role in its ability to analyze and classify the physics events efficiently. These features are derived from kinematic observables of the reconstructed final-state objects in the detector, such as jets, leptons, and missing transverse energy (MET). These are accessible from the reconstructed physics objects in the nano Analysis Object Data (nanoAOD) format, except for the jet origin, which is available only in simulation. At the CMS experiment, several data formats are used to store simulated and recorded data, each with varying levels of detail and storage demands. While formats such as Analysis Object Data (AOD) and mini Analysis Object Data (miniAOD) contain more detailed event information, they can still be memory- and disk-intensive when processing billions of data points, as in this analysis. For efficient large-scale analyses and machine learning applications, the nanoAOD format is typically used, as it provides a compact representation of the most relevant reconstructed quantities. The choice of which quantities to use as input features depends on the underlying physics, but ultimately, it is the network's task to determine the relative importance of features and their combinations.

One of the key advantages of GNNs is their inductive bias, which allows them to handle variable input sizes without requiring predefined input dimensions. This means that GNNs can process any number of reconstructed objects, including all possible combinations of those

objects, without the need for padding or a fixed-size input layer. Unlike traditional Deep Neural Networks (DNNs), there is no need to set an upper limit on the number of jets or other objects in an event. The flexibility of GNNs means that the graph structure can naturally accommodate varying jet multiplicities across different events, allowing for a more dynamic and efficient representation of the data.

Each final-state object is transformed into a node of the graph. So, in a semileptonic channel, as described in Sec. 8.3, the graph contains a minimum of eight nodes. Each node is represented by a node feature vector, $\vec{v_i}$, generated from the kinematic information following a predefined order: p_T, η, ϕ, M, E, charge, and the b-tagging discriminant. The b-tagging discriminant used is DEEPJET heavy flavour tagging algorithm, as discussed in Sec. 7.3.5. In the feature vector, a value is set to zero (zero-padding) when a specific feature does not apply to a particular object, such as setting the MET value to zero for a lepton or the b-tagging value to zero for a lepton. Additionally, the vector includes flags that distinguish the type of object: the first flag represents a jet, the second identifies a lepton, and the third corresponds to the MET. The node features, which correspond to the elements in each feature vector, are listed in Table 8.1

The relationships between the physics objects are captured as edges connecting the nodes. These edges encode important relational information that allows the model to learn interactions between objects in a given event. The ability of GNNs to process and leverage such relational information is superior to that of traditional DNNs, particularly in tasks where understanding the relationships between objects is crucial. This has been demonstrated in various studies, where incorporating domain-specific features, like the geometric distance between particles or their invariant mass, significantly improves the performance of the GNN model [153]. In this analysis, multiple domain-specific features into the graph structure are embedded simultaneously, which enhances the model's ability to capture complex relationships between objects. Specifically, the edge features are constructed for every pair of nodes in the graph, incorporating kinematic variables such as the ΔR (Eq. 4.7), invariant mass (m_{ij}), transverse mass (m_T), $\Delta\phi$, and $\Delta\eta$ between the nodes. These features provide critical insights into the interactions between the particles, improving the model's discrimination power, even when individual node features (such as particle momenta or masses) may be similar.

To prevent any single feature from dominating the learning process, especially those with larger magnitudes like p_T, all continuous features are normalized to a range of [0, 1]. This normalization ensures that each feature contributes equally to the training process, which helps avoid imbalances that could affect model convergence. By embedding these edge features directly into the graph, the GNN can learn not just individual properties of particles, but also how they interact and correlate within the event. This allows for more effective classification and identification of signal versus background events. Additionally, the inclusion of relational information, such as the distance between jets or the transverse mass between particles, allows

the GNN to learn more nuanced patterns in the data that traditional methods might miss. All input features are summarized in Table 8.1.

Table 8.1: Input features and edge features for jets, lepton, and MET

Feature	**Jets**	**Lepton**	**MET**
Node features			
p_T	Yes	Yes	Yes
η	Yes	Yes	No
ϕ	Yes	Yes	No
m	Yes	Yes	No
charge	No	Yes	No
b-tagging	Yes	No	No
E	Yes	Yes	No
jet flag	Yes	No	No
lepton flag	No	Yes	No
MET flag	No	No	Yes
Edge features			
ΔR	Yes	Yes	Yes
m_{inv}	Yes	Yes	Yes
$\Delta\phi$	Yes	Yes	Yes
$\Delta\eta$	Yes	Yes	Yes
di-object p_T	Yes	Yes	Yes
m_T	Yes	Yes	Yes

The input features selected are expected to show some discriminating power between different physics objects. In the $t\bar{t}$+X case, additional jets and b-jets stemming from the top quark decays often exhibit overlapping features, making their differentiation a key challenge. Fig. 8.3 shows the node features such as p_T, η, and b-tagging value, by combining data from the relevant processes: $t\bar{t}$+H, $t\bar{t}$+Z, and $t\bar{t}$ +jets. This provides a comprehensive view of how these kinematic features vary across different jet categories. Jets originating from top quarks tend to have higher p_T compared to additional jets. As shown in Fig. 8.3, the p_T distribution of additional jets is generally shifted towards lower values, particularly when compared to jets from the $t\bar{t}$ system. For example, additional jets in $t\bar{t}$+H and $t\bar{t}$+Z events exhibit higher p_T than those in $t\bar{t}$ +jets events. This distinction is vital, as the GNN can learn to classify the higher p_T jets as being more likely associated with top quarks. The distribution of η indicates that the jets from the $t\bar{t}$ system are typically located in the more central region of the detector, compared to additional jets, which are often further away from the detector's central axis. Similarly, the b-tagging information plays a significant role in distinguishing b-jets from top quarks from additional jets. The top quarks decay most exclusively to bottom quarks, so the b-jets from the top decays tend to have larger b-tagging values due to the higher transverse momentum and more pronounced secondary vertices. Therefore, the b-tagging information is a strong discriminant feature, particularly when the additional jets are c-jets or light-flavor (LF) jets. Even a simple top-down ranking of jets by their b-tagging values can already provide useful

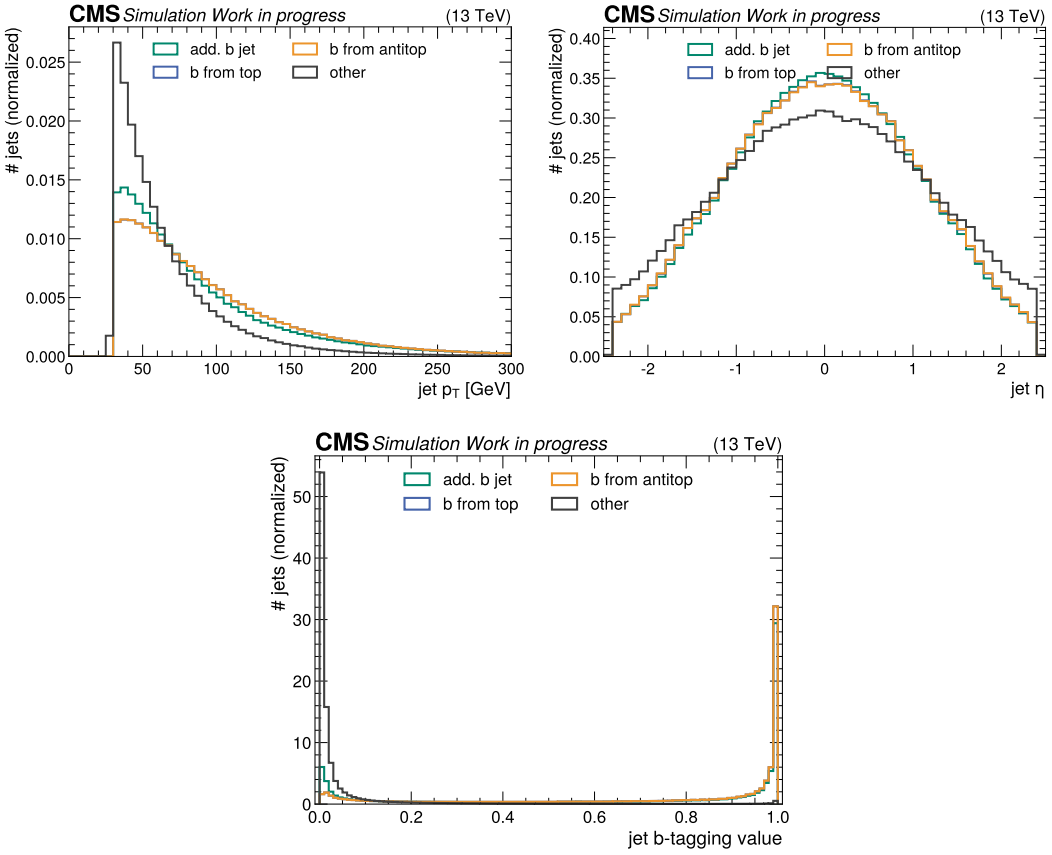


Figure 8.3: Kinematic distributions of transverse momentum (top-left), pseudorapidity (top-right), and b-tagging value (bottom) for all samples ($t\bar{t}$+H, $t\bar{t}$+Z, and $t\bar{t}$ +jets). The green line represents additional b-jets from additional heavy-flavor jets initiated by additional gluon radiation (add. b jet), the blue line represents b-jets from top quark decays (b from top), the orange line corresponds to b-jets from the top antiquark decays (b from antitop), and the dark gray line indicates 'other' jets, which include all jets from W boson decays and any jets that are not the two additional jets of interest in the analysis (other). The lines are normalized to unit area.

identification of additional jets, especially when compared with other techniques. However, this approach alone is often not enough, as additional jets may sometimes have similar b-tagging values as the top b-jets. Therefore, the inclusion of other features, such as p_T, η, and ΔR, ensures better classification. Although ϕ does not directly contribute to separation due to the detector's azimuthal symmetry, when combined with other features like ΔR, ϕ can help define spatial relationships between objects. Finally, additional node features, such as mass and energy, are included in the graph. While these quantities are related by the energy-momentum relation, including all three (mass, energy, and momentum) gives the network greater flexibility to identify which features are most relevant for classification. This approach avoids the need for the network to explicitly learn the energy-momentum relation, which could be computationally inefficient. These input features together provide the GNN with a strong foundation for identifying and differentiating between the relevant jet categories in the analysis.

The two variables, DEEPJET tagger scores CvB (c versus b) and CvL (c versus light),(Eq. 5.2) were initially considered for inclusion as part of the feature set. These two-dimensional jet flavor discriminants distinguish between b-jets and c-jets, as well as c-jets and light flavor (LF) jets, based on their CvB and CvL values, respectively. These scores provide important information for flavor identification and contribute significantly to the separation between additional jets and b-jets from the top quark decay, as demonstrated in Fig. 8.4. However, despite their

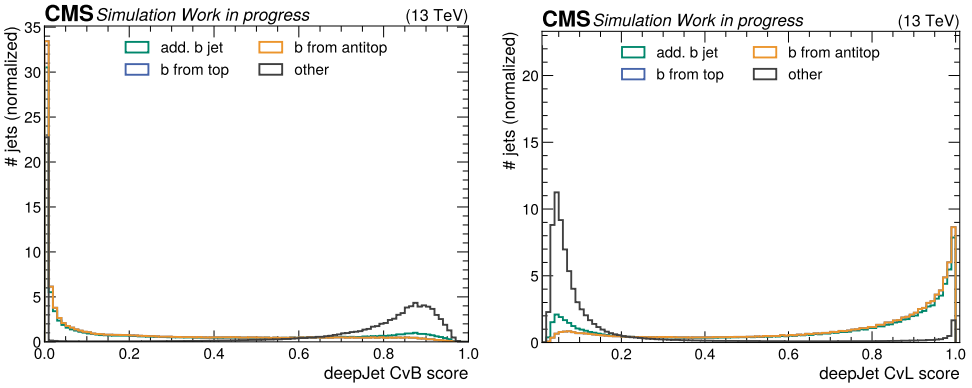


Figure 8.4: Distributions of the DEEPJET c-tagger scores for all samples ($t\bar{t}$+H, $t\bar{t}$+Z, and $t\bar{t}$ +jets): CvB distribution (left) and CvL (right) distribution. These distributions show the behavior of the two-dimensional jet flavor discriminants that distinguish between b-jets, c-jets, and light-flavor (LF) jets.

theoretical advantages, both CvB and CvL failed to pass the goodness of fit (GoF) test (Fig. 8.9), which evaluates how well the model's predictions fit known data. Fig. 8.5 demonstrates the divergence observed for CvB, and thus, the two c-jet tagger discriminants, CvB and CvL, are excluded from the training process. Specifically, these tagger scores exhibit a mismatch between simulation and data, particularly in regions with particularly high or low scores. This is shown in Fig. 8.5 for the year 2018, for example. Even attempting a calibration or applying a scale factor (SF) to address the divergence in the CvB and CvL scores did not resolve the

issue. The mismatch between simulated events and real data remains significant. In contrast, the one-dimensional b-tagging score, which is correlated with both CvB and CvL, performed better in this context. The b-tagging score not only passes the GoF test but also allows for the application of a scale factor to correct for any data-to-simulation discrepancies, as provided by the CMS Collaboration (see Sec. 7.4.8). For this reason, while incorporating CvB and CvL as node features showed an improvement in the performance of the models, they were ultimately excluded from the training process due to the persistent mismatch between simulation and data. Thus, the removal of these c-taggers (CvB and CvL) was necessary to ensure that the analysis models are based on reliable and well-calibrated features that align well with the data. The other input features are included in the GNN training.

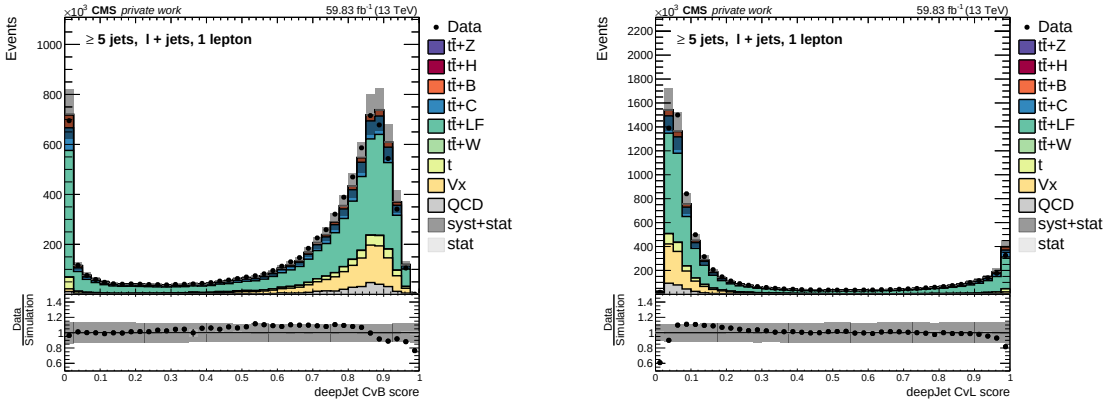


Figure 8.5: Distributions of the 2018 DEEPJET tagger scores: CvB distribution (left) and CvL distribution (right). These distributions show the behavior of the two-dimensional jet flavor discriminants, which differentiate between b-jets, c-jets, and light-flavor (LF) jets. The mismatch between these scores in simulation and data, particularly in the extreme regions of the distribution, led to their exclusion from model training because the divergence could not be resolved, despite calibration attempts.

8.3.2 Training architecture

The classification of $t\bar{t}$+X events of interest in this thesis is primarily driven by flavor composition, with the most significant differences arising from the origin of additional jets. The input to the GNN consists of a graph representation of the event, where nodes and edges are encoded with relevant input features, as explained in Sec. 8.3.1. For this analysis, the full dataset in nanoAOD format spans approximately 6.3 TB across all samples. After performing a local skimming procedure to select only the relevant events and features for our training (Sec. 8.2), the dataset is reduced to approximately 200 GB per era. Additional simulations for systematic variations further increase the storage footprint by 100 GB.

It is worth noting that information about jet flavor origin is not stored in the nanoAOD format. This label is essential for training, especially for differentiating jets from top-quark

decays from additional jets. To retrieve this, a dedicated matching step is applied using information from the miniAOD level, where generator-level truth is retained. The matching procedure is based on geometrical proximity: a reconstructed jet in nanoAOD is matched to a generator-level jet from miniAOD if the distance ΔR is below 0.4. This technique reliably transfers jet origin information, achieving a matching efficiency of approximately 99.5–99.7% across samples. The resulting dataset includes roughly 10 million unweighted events in nanoAOD format after skimming [93].

One key challenge in this analysis is distinguishing events based on their heavy-flavor content, specifically, whether the additional b-jets are part of the $t\bar{t}$ decay chain or not. So identifying the additional jets primarily could have a positive impact on the overall classification. In the context of this analysis, thus, the classification is done in two steps: Node-Level Prediction (NLP) and Graph-Level Prediction (GLP).

- **Node-Level Prediction (NLP)**: The NLP focuses on identifying individual objects and making predictions about their properties, such as whether a jet is an additional b-jet or not. This is a binary classification problem.

- **Graph-Level Prediction (GLP)**: The GLP, on the other hand, operates on the entire graph, considering the relationships between the objects and making predictions about the overall event classification, such as distinguishing between different physics processes (e.g., $t\bar{t}$+H, $t\bar{t}$+Z, or $t\bar{t}$ +jets). This is a multiclassification problem.

The summarized end-to-end representation of the two steps in the GNN architecture is portrayed in Fig. 8.6. It includes the steps from event simulation to event classification. Although the initial graph structure is identical in both training stages, the models differ in architecture and learning objective (see Sec. 8.4.1, 8.5.1). Features extracted from the GLP output are later used in the statistical fit model for signal extraction and cross-section measurements (see Chapter 9). In this work, a distinction is made between a network architecture, which refers to the overall structure and design of layers, and a model, which refers to a specific instantiation of that architecture trained with fixed hyperparameters and random initialization. To ensure robustness and mitigate potential overfitting or training instabilities, two key strategies are implemented. Each network architecture is trained eight times with different random initializations. This helps account for fluctuations in performance arising from the stochastic nature of training. Relying on a single training can lead to misleading conclusions due to accidental overfitting or underperformance. The other strategy is an even–odd splitting approach, which helps to avoid biases in the trained models when evaluating simulated samples. One model is trained on events with even event IDs and then evaluated on events with odd IDs. A second, independently trained model uses the odd IDs for training and evaluates the even ID events. This ensures that no event is used both for training and evaluation within the same model, thereby preventing statistical biases in later data–MC comparisons or statistical fits.

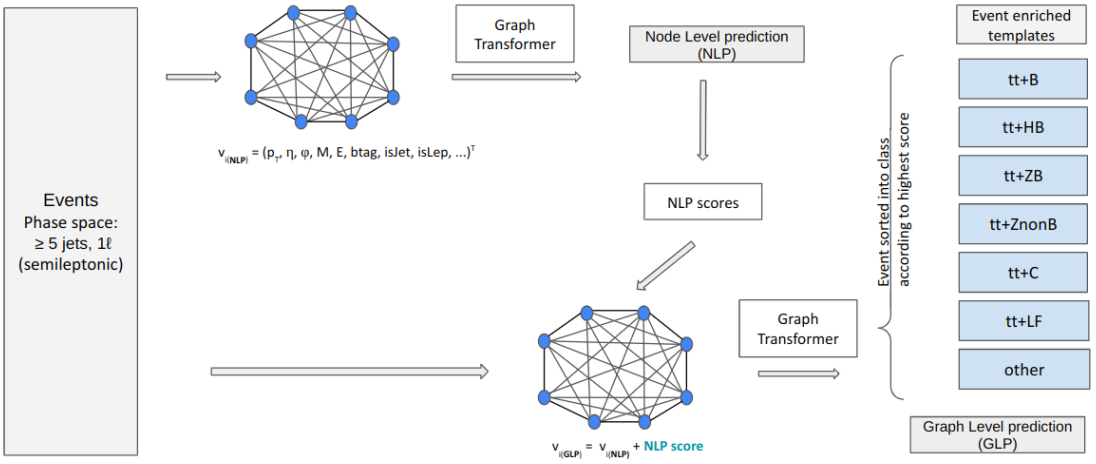


Figure 8.6: GNN strategy used for event classification. Semileptonic events in the analysis phase space (at least 5 jets, 1 lepton) are mapped to a complete graph with per-node features (Sec. 8.3.1). A graph–transformer network performs node-level prediction (NLP) to produce per-jet scores. These scores are then appended to the node features, and a second graph transformer is used to perform graph-level prediction (GLP). The GLP outputs multi-class event probabilities used to build event-enriched binned distribution (templates): $t\bar{t}$+B, $t\bar{t}+H(b\bar{b})$, $t\bar{t}+Z(b\bar{b})$, $t\bar{t}$+Z (nonB), $t\bar{t}$+C, $t\bar{t}$ + LF, and *other* events are assigned to the class with the highest GLP score (adapted from Ref. [93]).

8.4 Identification of additional jets using Node Level Prediction (NLP)

In the classification task at hand, the first step involves identifying the additional jets in $t\bar{t}$ events that do not originate from the top-quark decays in $t\bar{t}$ events. These additional jets may arise from QCD heavy-flavor radiation or from the decay of additional particles such as the Higgs or Z boson. This task is framed as a binary classification problem, which involves classifying each jet as either an additional jet or not, based on its relationship to the parton. This identification task acts as a pre-classifier for the main event classification, which categorizes the event into different $t\bar{t}$+X classes. This pre-classification step is advantageous for several reasons: it can be studied separately from the overall event classification, and its performance can be benchmarked against other jet identification methods. Additionally, alternative approaches to classify additional jets may be explored in future work. Graph Neural Networks (GNNs) are particularly well-suited to this problem, since the underlying physics processes can be naturally represented as graphs: jets are treated as nodes, and their kinematic and relational information can be encoded in the edges. For this task, the input features, both node and edge features, are explained in Sec. 8.3.1. The graph used in this task is fully connected, meaning all nodes are linked to each other in a directed manner, and no self-loops are present. Global attributes

are not utilized in this NLP task. The structure of the input graph for NLP is shown in Fig. 8.2 (right).

8.4.1 NLP architecture

NLP classification poses a combinatorial challenge, as each event contains multiple reconstructed jets whose origins must be determined. To effectively capture the complex, non-local correlations between jets, a graph transformer architecture is employed. Transformer-based models, originally developed for sequence modeling in natural language processing, have demonstrated strong performance on tasks that require modeling long-range dependencies. By incorporating a self-attention mechanism, the network can learn dynamic, context-dependent relationships between all nodes in the graph, rather than being restricted to fixed local neighborhoods. In this work, the TransformerConv [159] operator from the PyTorch Geometric library is used to implement the self-attention mechanism within the GNN framework. This operator enables the network to compute attention weights for each pair of connected nodes, thereby emphasizing the most relevant interactions for the classification objective. The resulting architecture is designed for node-level prediction, where the network outputs, for each jet in the event, a probability score indicating whether it is an additional jet or not.

The first step is to normalize the node and edge features before passing them to the GNN. The translated graph structure, as described in Sec. 8.3, is first fed into an encoding layer, which transforms the initial node vectors into an embedding space suitable for processing by the graph network/transformer (GN) blocks. This step changes the representation from raw physical quantities to learned feature vectors. The features are then normalized again. The normalization between network layers is applied to improve training stability and performance. In deep networks, it helps mitigate changes in input distributions across layers, enabling faster and more stable training. In the case of deep GNNs, normalization also counteracts the loss of distinction between node features across layers, preserving meaningful differences and preventing performance degradation. After normalization, an activation function is applied, allowing the layer to capture non-linear transformations. The output from the encoding layer is then passed through a series of GN blocks, which are the main computational units of the architecture. In the GN blocks, node features are updated using a message-passing mechanism. Each node aggregates information from its neighbors, with the relevance of each neighbor determined by learned attention weights. For example, a jet from a top quark decay will focus more on other top decay products, effectively learning which neighbors are most informative. In addition to node features, edge features are also incorporated into the attention calculation, as the relationships between objects carry important information. After several GN block iterations, each node obtains an updated representation: a weighted combination of its own features and the aggregated features of its neighbors. The final step is the classifier block, implemented as a multilayer perceptron (MLP). Here, the feature dimensions are gradually reduced, and the output is passed through a final activation to produce a probability between 0 and 1, indicating the likelihood that a given node corresponds to an additional jet. The overall structure of the

architecture is shown in Fig. 8.7, which illustrates the flow from event representation as a graph to the graph transformer layers and to the final classification output for each node. The mathematical formulation of the GN blocks is detailed in the subsequent sections.

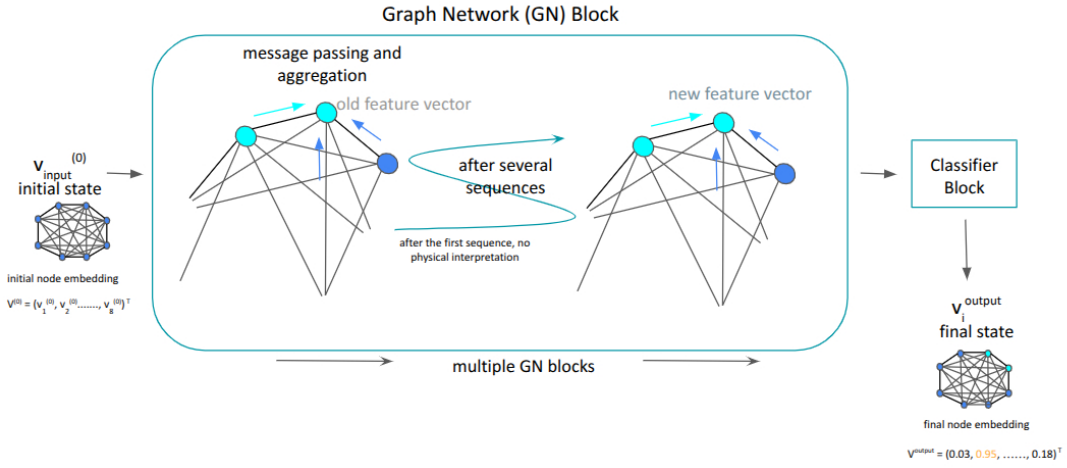


Figure 8.7: An overall representation of the schematic flow of NLP architecture. Event objects are mapped to a graph and embedded, node and edge features are normalized and passed through a stack of TransformerConv-based GN blocks that perform message passing with learned attention. The resulting node embeddings are flattened through a classifier block to yield per-node probabilities for the additional-jet label.

Graph transformer

As shown in the Fig. 8.8, a graph transformer architecture is employed to integrate both feature and relational information during training. The model uses a vanilla multi-head attention mechanism, adapted from transformer models [157] for use in GNNs. Given node features $H^{(l)} = \{h_1^l, h_2^l, ..., h_n^l\}$ for a graph with n number of nodes, the attention coefficient for each edge from node j to i is computed using learned projection matrices and bias vectors, b. For each attention head c, the query and key vectors are obtained through linear transformations of the node features, while the edge features are also projected into the same embedding space.

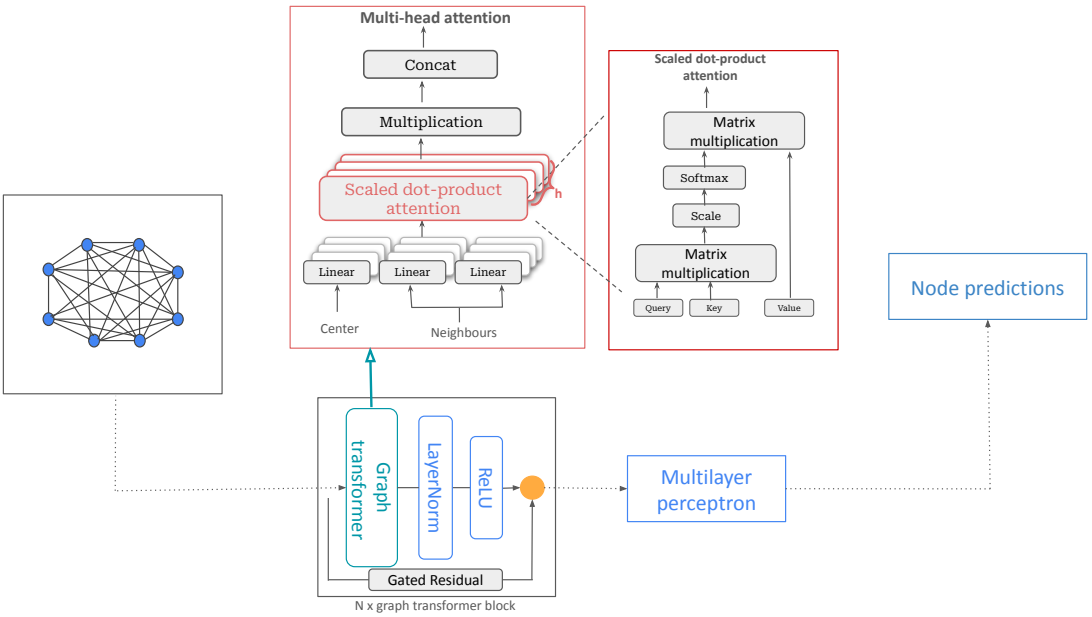


Figure 8.8: Detailed representation of the NLP architecture. Each event is processed by N stacked graph-transformer blocks. Within each block, node embeddings are updated via multi-head scaled dot-product attention: queries, keys, and values are formed from center and neighbor nodes, attention scores are computed (scale $\rightarrow$ softmax), and head outputs are concatenated and linearly projected. Blocks use LayerNorm and a gated residual connection around the attention and MLP sublayers. A final classifier head with an MLP maps the per-node embeddings to node-wise predictions. Adapted from Ref. [159].

The attention score is then computed as:

$$q_{c,i}^{(l)} = W_{c,q}^{(l)} h_i^{(l)} + b_{c,q}^{(l)}, \tag{8.19}$$

$$k_{c,j}^{(l)} = W_{c,k}^{(l)} h_j^{(l)} + b_{c,k}^{(l)}, \tag{8.20}$$

$$e_{c,ij} = W_{c,e} e_{ij} + b_{c,e}, \tag{8.21}$$

$$\alpha_{c,ij}^{(l)} = \frac{\left\langle q_{c,i}^{(l)},\, k_{c,j}^{(l)} + e_{c,ij} \right\rangle}{\sum_{u \in \mathcal{N}(i)} \left\langle q_{c,i}^{(l)},\, k_{c,u}^{(l)} + e_{c,iu} \right\rangle} \tag{8.22}$$

where $q_{c,i}^{(l)}$ is the *query* vector for node i in head c at layer l, obtained by projecting its feature vector $h_i^{(l)}$ into a head-specific subspace via the learnable weight matrix $W_{c,q}^{(l)}$ and bias $b_{c,q}^{(l)}$ [Eq. 8.19]. Similarly, $k_{c,j}^{(l)}$ is the *key* vector for node j (Eq. 8.20). Here,

$$\left\langle q_{c,i}^{(l)},\, k_{c,j}^{(l)} \right\rangle = exp \frac{(q_{c,i}^{(l)})^T (k_{c,j}^{(l)})}{\sqrt{d}} \tag{8.23}$$

is the exponential scaled dot-product of two vectors *query* and *key*, and d is the hidden size of each head. The dimension d of the query and key vectors is smaller than the embedding dimension E of the node features $h_{i,j}^{(l)}$ by a factor of C, the number of multi-heads, and is calculated as $d = E/C$. For the head attention c, the feature vector $h_i^{(l)}$ and connected feature vector $h_j^{(l)}$ are transformed into a query vector $q_{c,i}^{(l)} \in \mathbb{R}^d$ and a key vector $k_{c,j}^{(l)} \in \mathbb{R}^d$ respectively with same dimension d in all heads. The transformation is done by the multiplication of various trainable parameters, $W_{c,q}^{(l)}$, $W_{c,k}^{(l)}$, $b_{c,q}^{(l)}$, $b_{c,k}^{(l)}$ with the feature vectors. During training, all trainable parameters are optimized. Since each attention head operates independently, it can capture distinct patterns in the data. By employing multiple heads, the model learns diverse relationships in parallel, thereby increasing its expressiveness and improving its ability to generalize. This mechanism measures the relevance of the key vector concerning the query vector. When the two vectors align closely in the query-key space, their dot product is large. If the vectors are orthogonal, the dot product equals zero. Conversely, if they point in opposite directions, the dot product becomes negative, indicating low attention of the key vector for the query vector. Applying the exponential function rescales the dot product, so the more similar the query and key vectors are, the higher the dot product and, consequently, the stronger the attention node j pays to node i.
The edge feature vector e_{ij} between two nodes is calculated using Eq. 8.21 and is encoded and added into key vector as additional information for each layer, with independent trainable weight matrices, $W_{c,e}$ and biases $b_{c,e}$. The attention matrix, $\alpha_{c,ij}^{(l)}$ is thus calculated using Eq. 8.22. The summation is performed over all nodes $\mathcal{N}(i)$ connected to node i to guarantee that the attention values are properly normalized. Then, a message aggregation is made from the node j to i by defining a value vector,

$$v_{c,j}^{(l)} = W_{c,v}^{(l)} h_j^{(l)} + b_{c,v}^{(l)} \tag{8.24}$$

as,

$$\hat{h}_i^{(l+1)} = \Big\|_{c=1}^{C} \left[\sum_{j \in \mathcal{N}(i)} \alpha_{c,ij}^{(l)} \left(v_{c,j}^{(l)} + e_{c,ij}^{(l)} \right) \right] \tag{8.25}$$

where the $||$ is the concatenation function for C head attention. It replaces the original normalized vector with a new transition vector, based on the direction of the value vector $v_{c,j}^{(l)}$ and the edge vector $e_{c,ij}^{l}$. A gated residual connection is used between layers to avoid the model from over-smoothing, as given below:

$$r_i^{(l)} = W_r^{(l)} h_i^{(l)} + b_r^{(l)} \tag{8.26}$$

$$\beta_i^{(l)} = \text{sigmoid} \left(W_g^{(l)} \left[\hat{h}_i^{(l+1)}; r_i^{(l)}; \hat{h}_i^{(l+1)} - r_i^{(l)} \right] \right) \tag{8.27}$$

$$h_i^{(l+1)} = \text{ReLU} \left(\text{LayerNorm} \left((1 - \beta_i^{(l)}) \hat{h}_i^{(l+1)} + \beta_i^{(l)} r_i^{(l)} \right) \right) \tag{8.28}$$

Here, r_i^l is the residual vector defined, similar to equations 8.19, 8.20, 8.24, but with independent weight matrices $W_r^{(l)}$ and biases $b_r^{(l)}$. The gating coefficient $\beta_i^{(l)} \in [0, 1]$ is computed via a sigmoid activation, taking as input the concatenation of the intermediate updated features, $\hat{h}_i^{(l+1)}$, $r_i^{(l)}$, and their difference (Eq. 8.27). Finally, the new node representation, $h_i^{(l+1)}$ is obtained by applying a weighted combination of the intermediate features and the residual vector, controlled by the gate $\beta_i^{(l)}$, followed by layer normalization (LayerNorm) [160] and a ReLU activation (Eq. 8.28). The LayerNorm is implemented at the end of each GN block to improve training stability and performance as explained in the Sec. 8.4.1. Without the normalization, one common issue is the internal covariate shift, where the distribution of inputs to each layer changes as the parameters of preceding layers are updated. This effect accumulates in deeper architectures, making it harder for later layers to adapt and often slowing convergence by requiring smaller learning rates or careful initialization. Normalization techniques such as batch normalization (BatchNorm) [161] counteract this by recentering and rescaling activations within a batch, which helps stabilize the optimization process, accelerates training, and can improve generalization. Another challenge arises when stacking multiple layers, or referred to as oversmoothing. Here, repeated message passing can cause node representations to become increasingly similar, eventually making nodes indistinguishable and degrading performance. This phenomenon is linked to the smoothing nature of graph convolution operators. Methods like pair normalization (PairNorm) [162] address this by preserving the relative distances between node embeddings across layers, thereby maintaining representational diversity. The specific choice of normalization approach is treated as a hyperparameter in the architecture design and is optimized as part of the model selection process.

After processing through multiple graph transformer blocks, the embedded node features are passed to a prediction stage designed to determine whether each node corresponds to an additional jet. At this point, the feature dimension E of the node embeddings is progressively reduced using a multilayer perceptron (MLP). The MLP is composed of several fully connected layers, where each hidden layer decreases the feature dimension by a fixed reduction factor. This gradual compression helps the network distill relevant information for the classification task. The final layer produces a single scalar output for each node, activated by a sigmoid function, which maps the value to a probability in the range $[0, 1]$. This output can be interpreted as the model's estimated likelihood that the corresponding physics object is an additional jet.

NLP hyperparameter settings and training

The training sample for the NLP task is derived from simulated events produced for various $t\bar{t}$+X events, as detailed in Table 8.2, which covers the various scenarios that can arise at the particle level. The $t\bar{t}$+$b\bar{b}$ and $t\bar{t}$ + b events contain two and one additional b-jet, respectively. The second b-quark in $t\bar{t}$ + b is either outside the detector acceptance or fails the reconstruction requirements. In addition, a configuration referred to as $t\bar{t}$ + 2b can occur, where two additional b-quarks are produced but are reconstructed within the same jet. Similarly, for $t\bar{t}$ + $c\bar{c}$ and $t\bar{t}$ + c events, two and one additional c-jets are present, respectively, where in the latter case the

second c-quark is either outside the detector acceptance or does not satisfy the reconstruction requirements. Events from $t\bar{t} + H(b\bar{b})$ and $t\bar{t} + Z(b\bar{b})$ processes are also included in the sample, and are further subdivided for improved visualization of the network performance. The $t\bar{t} + LF$ refers to $t\bar{t}$ production without any additional heavy-flavour jets. The full composition of the training dataset is summarised in Table 8.2. A total of about 0.9 million events are used for the NLP model training.

For each network training, the available simulated events are randomly split into training,

Table 8.2: Event yields per process: NLP

Process	No. of events
$t\bar{t}$+$b\bar{b}$	95610
$t\bar{t} + b$	55421
$t\bar{t} + H(b\bar{b})$	100000
$t\bar{t}$+H (b)	107415
$t\bar{t} + Z(b\bar{b})$	178286
$t\bar{t}$+Z (b)	25147
$t\bar{t}$+Z (nonB)	75000
$t\bar{t} + c\bar{c}$	160812
$t\bar{t} + c$	85001
$t\bar{t} + LF$	100000
$\sum$(events)	982692

validation, and test datasets. The validation and test samples each correspond to 5% of the training sample size, and the three datasets are statistically independent. This ensures that the evaluation of the trained models is unbiased and representative of their performance on unseen data. The NLP training is performed using datasets from the 2017 and 2018 data-taking periods. These two years provide the largest event statistics of Run-II and thus dominate the overall sensitivity. They also correspond to the majority of the integrated luminosity and yield the most stable behavior. To ensure that a single model can be trained on the merged sample without era-dependent biases, several goodness-of-fit (GoF) tests (e.g., χ^2/ndf, Kolmogorov–Smirnov, and Anderson–Darling) are carried out on relevant kinematic distributions and on the NLP targets across eras. Fig. 8.9 summarizes the Anderson–Darling (AD) GoF comparison between 2017 and 2018 for representative inputs. Within uncertainties, the 2017 shapes are consistent with those from 2018, for example, jet p_T, N_{jets}, and deepJet scores show p-values well above typical thresholds. The p-value represents the probability of obtaining a test statistic at least as extreme as the observed one under the assumption that the two distributions are drawn from the same underlying distribution. Values above $p > 0.05$ indicate statistical compatibility. The relevant kinematic distributions are checked for consistency across these eras, allowing a single model to be trained on the merged Run-II datasets without introducing systematic biases. For completeness, the 2016 GoF test results are provided in Appendix D.

All networks are trained using the categorical cross-entropy loss function, defined for a single event i as the following:

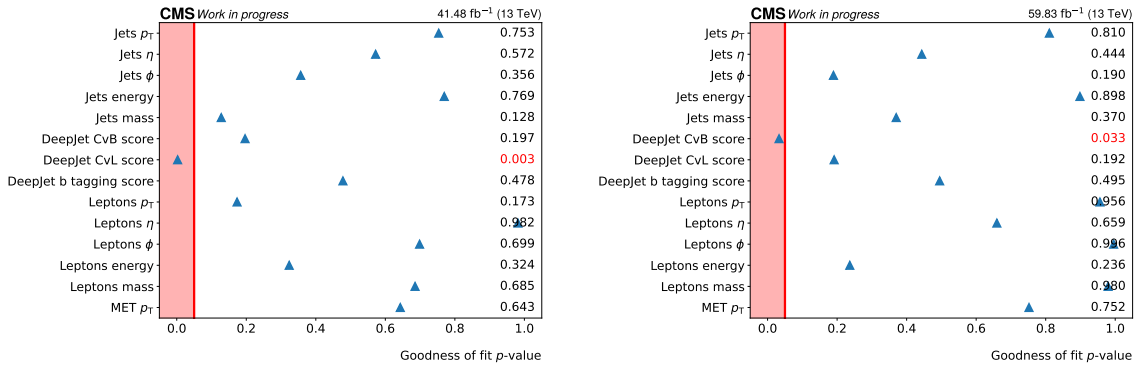


Figure 8.9: Anderson–Darling (AD) goodness-of-fit p-values for representative NLP input distributions between data and MC. On the left, 2017, and on the right, 2018 are shown. Each marker shows the AD p-value for the indicated variable, and the shaded band highlights the low-p region (e.g., $p < 0.05$). Overall, the 2017 and 2018 shapes are compatible within uncertainties, supporting a single NLP model trained on the merged 2017+2018 dataset. The CvB tagger score does not pass the GoF test in the 2018 dataset, and CvL in 2017, as observables with p-values greater than 5% are only retained as input features in the analysis.

$$l_i = -w_i \sum_c y_{i,c} \log(o_{i,c}) \tag{8.29}$$

where w_i denotes an event-specific weight, c the class index, $y_{i,c}$ the true label for class c for event i, and $o_{i,c}$ the predicted network output probability for that class. The weighing scheme compensates for variations in the number of events among the different processes and balances the relative contributions of 'true' nodes and 'false' nodes. In this analysis, true nodes refer to the additional jets and false nodes to the jets from top quark decay, lepton, and MET. Additionally, events with different jet multiplicities are weighted such that the overall loss remains independent of the number of nodes in the input graph. The Adam optimizer is employed for gradient computation and weight updates, with an initial learning rate of 10^{-2}. To mitigate overfitting, a combination of regularization strategies is employed. First, early stopping is implemented: training proceeds for up to 100 epochs but halts earlier if the AUC on the validation dataset has not improved by at least 1% within the most recent 15 epochs. The final network checkpoint corresponds to the epoch with the highest validation AUC. Second, dropout layers are introduced, randomly deactivating a fraction of the trainable parameters during training, which both improves generalization and reduces the effective network capacity. The final model performance is reported using the independent test dataset after all hyperparameters are fixed.

The choice of hyperparameters plays a critical role in the model's performance and generalization ability. The scanned parameters include the number of graph transformer layers, the

dimensionality of the node embeddings, the hidden layer sizes in the MLP, the dropout rate, the batch size, and the learning rate schedule. Each configuration is evaluated using the area under the receiver operating characteristic curve (ROC AUC) on the validation dataset, with the best-performing setting selected for the final model. The final set of hyperparameters used in training is summarized in Table 8.3.

Table 8.3: Hyperparameter settings for the NLP architecture.

Hyperparameter	Value
Batch size	2048
Normalization scheme	BatchNorm
Optimizer	Adam
Loss function	Binary cross-entropy
Activation function	leaky ReLU
Embedding dimension	32
Aggregation method	Sum
Number of multi-head attentions	5
Number of GN blocks	8
Learning rate	10^{-2}
Mutli-head consolidation	Concat
Gate residual information weighting	False
Dropout probability of normalized attention coefficients	0.2
Additive bias b	True
Root weight	True
Number of output (MLP) layers	3
Output embedding dimension	32
Output embedding dimension change factor	0.4
Output layer (MLP) dimension	1

8.4.2 NLP performance metrics and evaluation

The optimal network architecture is selected using a grid search over the hyperparameter space. Network performance, evaluated using the area under the ROC curve, illustrates the relationship between the true positive rate (signal efficiency) and the false positive rate (background misclassification) for varying classification thresholds. The AUC condenses this curve into a single value between 0 and 1, where larger values indicate better separation between classes. It provides a visual representation of the trade-off between efficiency and background rejection. The AUC quantifies this performance, where a value of 1.0 corresponds to perfect separation, and 0.5 indicates no discrimination.

Several measures, as described in Sec. 8.4.1, are applied to reduce the loss and improve generalization. The evolution of the loss and ROC AUC values during training and validation as a function of the number of training epochs is shown in Fig. 8.10. Consistent improvement on the validation dataset indicates that the network is learning meaningful, generalizable features, whereas divergence between training and validation metrics can signal overfitting. An epoch is

defined as one complete pass through the entire training dataset. In this case, training required a total of 43 epochs before early stopping was triggered. The plotted curves illustrate the reduction in loss and the corresponding increase in ROC AUC over epochs for both the training and validation datasets. The ROC AUC values reflect the network's classification accuracy, while the loss curve shows the optimization progress. If the area under the ROC curve (AUC) fails to improve by more than 5% within 10 consecutive epochs, the learning rate is reduced by a factor. Saturation in performance was observed after approximately the 18^{th} epoch, at which point the learning rate was reduced by a factor of 0.01, allowing the optimizer to perform finer adjustments to the network parameters. Training was stopped early after the second saturation point was reached. The evolution of the loss indicates that overfitting did not occur, as the training and validation curves follow each other closely without divergence. A similar pattern is observed for the ROC–AUC metric, where both training and validation scores improve steadily and saturate around a value of 0.904.

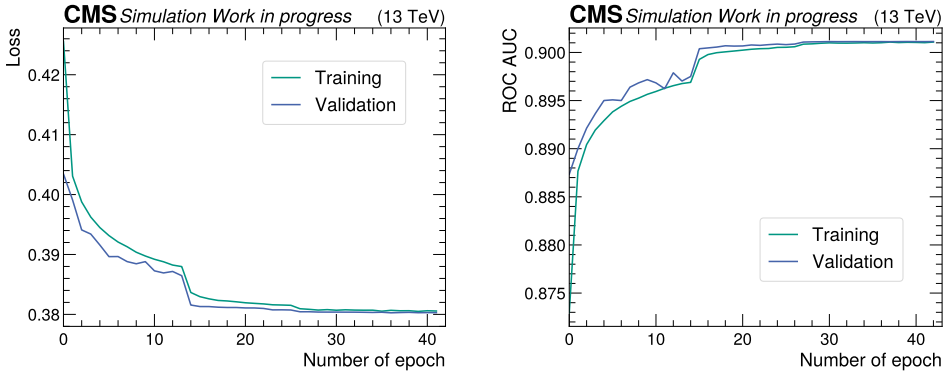


Figure 8.10: Evolution of training (green) and validation (blue) metrics over epochs. On the left, training and validation losses are shown as a function of epochs, and on the right, ROC-AUC is shown as a function of epochs. The cross-entropy loss decreases during training, and the corresponding ROC-AUC improves and stabilizes. The close agreement between training and validation curves indicates good generalization under the regularization measures of Sec. 8.4.1.

The ROC–AUC curve in Fig. 8.11 illustrates the classification performance of the trained model for each of the $t\bar{t}$ jets categories considered, as well as the combined score across all classes. The results show an overall ROC-AUC of 0.904, indicating strong discrimination power. Among individual classes, the best performance is achieved for the $t\bar{t}$ category, with an AUC of 0.96, and for $t\bar{t}$ +Z(nonB), with an AUC of 0.93, reflecting their distinct kinematic and flavor-tagging features. Good separation is also seen for $t\bar{t}+c$ (0.91) and $t\bar{t}+b$ (0.90). The more challenging categories, $t\bar{t}$+H (0.87), $t\bar{t}$+$b\bar{b}$ (0.87), and $t\bar{t}+c\bar{c}$ (0.84), show slightly lower AUC values, consistent with their higher similarity to other heavy-flavor final states and the resulting overlap in discriminating variables. Intermediate performance is obtained for $t\bar{t}$+Z (b) (0.89), where the additional b-jets provide useful discrimination but are not as unique as in the best-performing channels.

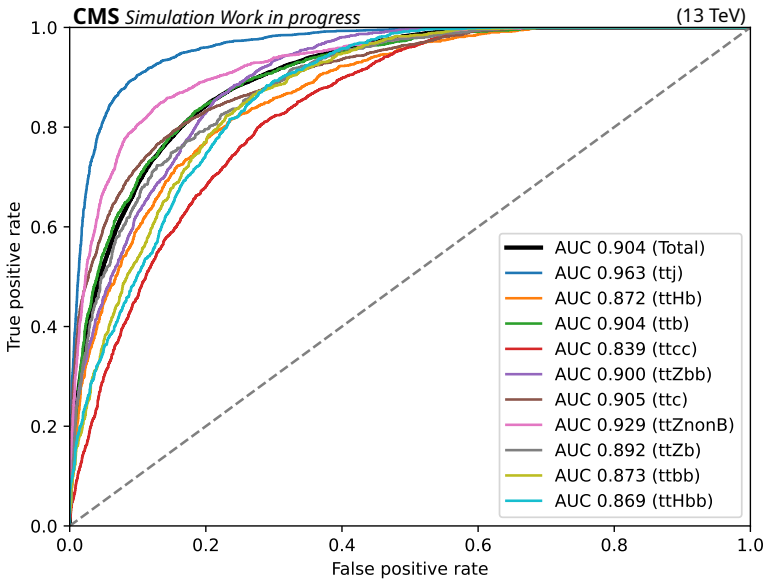


Figure 8.11: Receiver operating characteristic (ROC) curves for the NLP model, evaluated on the independent test dataset for the different $t\bar{t}$+X jet classes. The true positive rate (TPR) is plotted against the false positive rate (FPR) for each class, with the area under the curve (AUC) given in the legend. An AUC value close to 1 indicates excellent separation between the target class and the background. The total performance is indicated by a solid black line, and the lower boundary for a random classifier by a dotted gray line. Additionally, the events are further divided for different $t\bar{t}$+X classes to provide a clearer representation of classification performance for each category.

To further assess the performance and discrimination capability of the trained NLP model, the distribution of the predicted NLP scores is examined for different jet classes in the test dataset. Fig. 8.14 shows the normalized score distributions for b jets originating from the $t\bar{t}$ system, additional b jets, and other jets. An ideal classifier would yield distributions strongly peaked towards one for additional b jets (true positives) and towards 0 for the other jet categories (true negatives). The observed separation indicates that the model has learned to effectively distinguish additional b jets from jets originating from top quark decays and non-b jets. While some overlap remains between additional and top b jets, the distinct shift in score distributions reflects good discrimination power. This behavior is consistent with the classification objective and confirms that the model has successfully extracted relevant features from the input graph representation.

Another way to evaluate the performance of the model is through the assignment accuracy, defined as the fraction of events in which both additional jets are correctly identified. Fig. 8.13 shows the comparison of the fraction of events with 2/2, 1/2, and 0/2 correctly assigned jets obtained using the NLP approach compared to a more traditional method used in previous iterations of (similar) analyses, ΔR_{min} method. In the NLP approach, these fractions are obtained

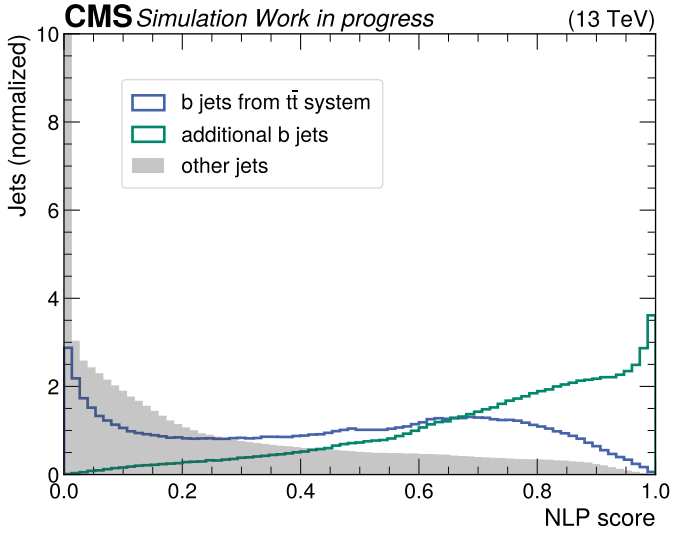


Figure 8.12: Normalized distribution of NLP scores for different jet classes in the test dataset of $t\bar{t}$+$b\bar{b}$, $t\bar{t}+H(b\bar{b})$, and $t\bar{t}+Z(b\bar{b})$. The model outputs values close to one for additional b jets and close to zero for b jets from the $t\bar{t}$ system and 'other' jets, indicating strong separation power. The 'other' jets include all jets apart from the two additional b jets of interest and the jets from W-boson decays from the $t\bar{t}$ system.

from the predictions of the trained model and their agreement with the true assignment of reconstructed jets. For the ΔR_{min} method, additional jets are selected as the two closest b or c jets in an event based on the minimal value of ΔR. This method is used in the $t\bar{t}$+$b\bar{b}$ analysis at the CMS experiment [85]. In $t\bar{t}$+$b\bar{b}$, $t\bar{t}+H(b\bar{b})$, and $t\bar{t}+Z(b\bar{b})$ events, the distances are calculated only between b-tagged jets, while in $t\bar{t}+c\bar{c}$ events, the distances are calculated between all jets in the event. The two jets with the smallest ΔR are then assigned as the additional c-jets. This procedure leads to the largest performance gap between the two methods for $t\bar{t}+c\bar{c}$ events. Across all processes, the NLP approach shows a significant improvement over the ΔR_{min} method. For example, in $t\bar{t}$+$b\bar{b}$ events, the fraction of $2/2$ assigned jets increases from 29% to 37%, and in $t\bar{t}+Z(b\bar{b})$ events from 40% to 44%. The improvement is even more pronounced for $t\bar{t}+c\bar{c}$ events, where the $2/2$ assignment accuracy nearly doubles from 17% with the ΔR_{min} approach to 33% with the NLP approach. Similar trends are observed for the $1/2$ and $0/2$ assignment categories, where the NLP model reduces the fraction of completely misassigned events $(0/2)$ compared to ΔR_{min}. These results confirm that the GNN-based NLP classifier not only has strong classification power (as shown in the score distributions) but also yields higher-quality event reconstruction. The improved identification of additional jets, particularly in challenging cases such as $t\bar{t}+c\bar{c}$ events, is expected to have a positive impact on downstream analyses and physics measurements.

Another assessment of the model performance can be made by comparing the NLP score distributions between the data and the simulation. As seen in Fig. 8.14, the simulation generally

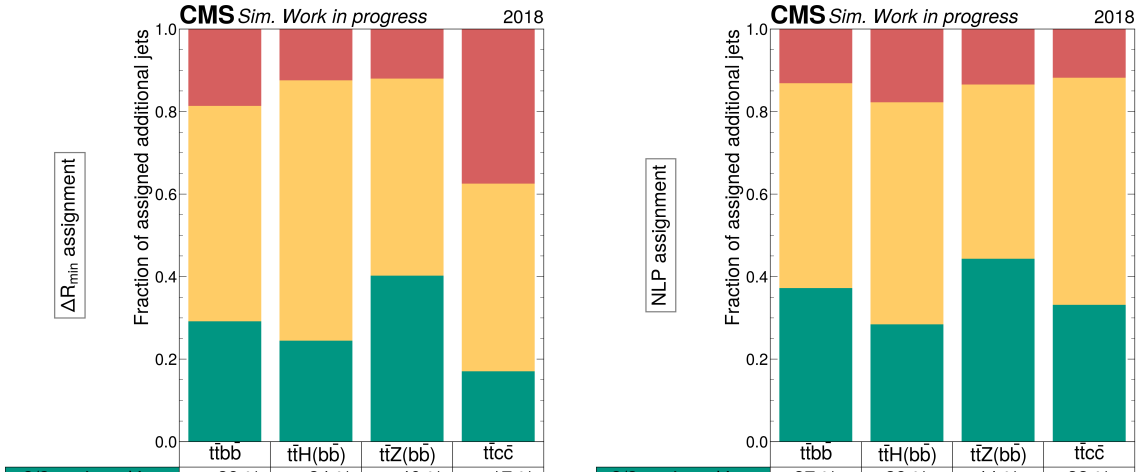


Figure 8.13: Comparison of assignment accuracy for the ΔR_{min} method (left) and the NLP-based approach (right) in the 2018 simulated data. The bars show the fraction of events in which $2/2$, $1/2$, and $0/2$ are correctly assigned additional jets for each process. The NLP approach achieves consistently higher $2/2$ assignment accuracy across all processes, with the largest improvement observed in $\mathrm{t\bar{t}} + \mathrm{c\bar{c}}$ events.

models the data well, and the overall data-MC agreement is good. A slight discrepancy is observed in the tails, however, most of this difference is accounted for by the systematic and statistical uncertainties associated with the prediction on simulated events.

8.5 Multiclass event classification using Graph Level Prediction (GLP)

The second task in the analysis is the Graph Level Prediction (GLP), where the goal is to classify entire events into different $\mathrm{t\bar{t}}$+X categories, as illustrated in the second stage of the classification strategy in the Fig. 8.6. For this task, seven different classes are defined. Unlike the NLP task, which performs a binary classification for each node to determine whether it is an additional jet or not, GLP performs a multi-class classification at the event (graph) level. While the overall framework is similar to that of the NLP, the main difference lies in the input features. In GLP, an additional feature indicating whether a node is an additional jet is included, that is, the NLP output is used as an additional input. It provides the network with higher-level information from the previous step. This also allows the model to leverage node-level information while considering the entire event phase space, including all relevant backgrounds. The accurate event-level classification is crucial for disentangling rare processes such as $\mathrm{t\bar{t}}$+H and $\mathrm{t\bar{t}}$+Z from the abundant $\mathrm{t\bar{t}}$ +jets background. Thus, it makes the GLP evaluation central to the overall analysis strategy.

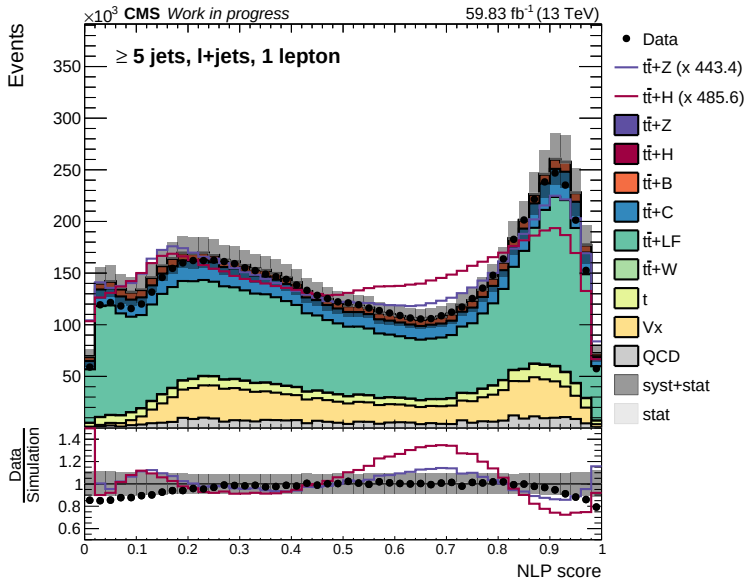


Figure 8.14: Comparison of NLP score distributions between data and simulation. Events with at least five jets in the ℓ+jets channel and one lepton are considered. The stacked histograms represent different simulated processes, and the uncertainty band includes both statistical and systematic contributions. The $t\bar{t}$+H and $t\bar{t}$+Z predictions are also overlaid as line distributions, normalized to the total yield, to enhance their visibility. Overall, a good agreement between data and simulation is observed, with small deviations in the high-score region largely accounted for by uncertainties.

8.5.1 GLP architecture

The GLP task uses a GNN framework similar to that of the NLP task, but adapted for event-level multi-class classification. As explained in Sec. 8.3.1, the input features graph consists of nodes representing reconstructed physics objects, leptons, jets, and MET, and edges encoding pairwise kinematic relationships, as in NLP. The network processes the graph through multiple message-passing layers to extract node-level features, which are then aggregated into a single graph-level representation using a pooling operation. This is achieved by applying a pooling mechanism that reduces the dimensionality of the graph while retaining the most relevant structural information. This aggregated representation captures the overall event topology and kinematics and serves as input to fully connected layers that output the classification scores for the predefined seven $t\bar{t}$ +jets categories.
The details are explained in the following paragraphs.

Graph Transformer

Similar to the NLP, the architecture is based on GN blocks, which exploit message-passing and attention mechanisms to learn correlations among reconstructed objects. However, while in

the NLP task each embedded node feature of dimension E was reduced to a scalar through a dedicated MLP to indicate the "additional-jetness" of a node, the GLP task requires a modified architecture designed for event-level classification. In particular, a downsampling procedure known as top-k pooling is integrated into the graph transformer architecture, following [163–165]. The top-k pooling procedure is conceptually related to max-pooling in convolutional neural networks (CNNs), where each feature map is reduced to size k by selecting the units with the highest activations. However, this principle cannot be directly applied to the GNNs as the nodes are not ordered and the top-k values may correspond to different nodes for each feature dimension. To address this challenge, it has been proposed to project all the node embeddings onto a single dimension by a learnable projection function. Based on the projected scores, the k nodes with the highest values are retained, while the remaining nodes are discarded [165]. With every application of the pooling layer, the number of nodes in the graph is reduced by a factor of k. The reduction is achieved by discarding $N - kN$ number of nodes from the original graph with N nodes, while retaining $\lceil kN \rceil$ nodes together with their associated edges. Fig. 8.15 portrays the operations performed by the pooling layer. The decision of which nodes to

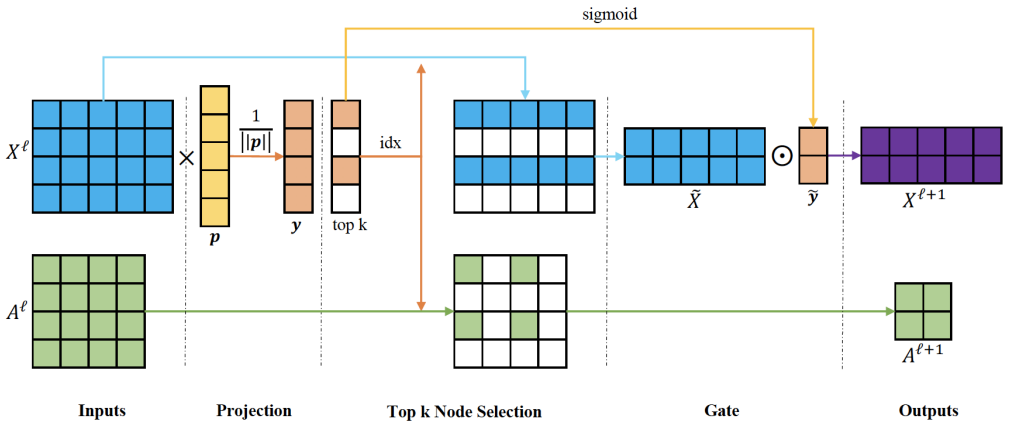


Figure 8.15: Schematic representation of the top-k pooling procedure with $k = 2$. A graph with four nodes and five features per node is projected onto a trainable vector $p \in \mathbb{R}^5$ to compute scalar scores for each node. The two nodes with the highest scores are retained to construct the pooled feature matrix $\tilde{X}^{(\ell)}$ and the corresponding adjacency matrix $A^{(\ell+1)}$. Finally, a gating operation is applied by element-wise multiplication with the selected node scores, yielding the new feature matrix $X^{(\ell+1)}$. Figure taken from Ref. [165].

retain is made using a projection of the node embeddings. At the l-th graph transformer block, the embeddings of the N nodes, each with E features, form a matrix

$$F^{(l)} \in \mathbb{R}^{N \times E}$$

This matrix is multiplied by a learnable projection vector in the l-th layer

$$p^{(l)} \in \mathbb{R}^E$$

$$y^{(l)} = \frac{F^{(l)} p^{(l)}}{||p^{(l)}||} \tag{8.30}$$

where $||p^{(l)}||$ is the L_2 norm of the vector $p^{(l)}$. The top-k selects the highest k indices from a given input vector

$$i^{(l)} = \text{top-}k(\vec{y}, k) \tag{8.31}$$

Thus, a new representation of the remaining nodes is passed to the next layer, $F^{(l+1)}$

$$F^{(l+1)} = \left(F^{(l)} \odot tanh(y^{(l)})\right)_{i^{(l)}} \tag{8.32}$$

Here, $\odot$ is elementwise multiplication, and $_{i^{(l)}}$ indicates the indexing operation that selects the rows corresponding to the retained nodes $i^{(l)}$. Finally, the feature matrix is reduced in size from $N \times E$ to $kN \times E$, keeping only the most relevant nodes for the next layer,

$$F^{(l+1)} \in \mathbb{R}^{kN \times E}$$

This ensures that the most informative subset of nodes is preserved consistently across all feature dimensions, thereby enabling an effective downsampling of the graph. At each top-k pooling step, a summary vector is constructed, and the summaries from all pooling layers are aggregated by summation before being passed to the classifier as follows:

$$s^{(l)} = \left(\frac{1}{N^{(l)}} \sum_{i=1}^{N^{(l)}} f_i^{(l)}\right) || \max_{i=1}^{N^{(l)}} f_i^{(l)}, \tag{8.33}$$

where $N^{(l)}$ denotes the number of nodes remaining after top-k pooling at layer l, $f_i^{(l)}$ feature vector of the i-th node at layer l, $\frac{1}{N^{(l)}} \sum_{i=1}^{N^{(l)}} f_i^{(l)}$ operation does the mean pooling, that is, average over all node features and the $\max_{i=1}^{N^{(l)}} f_i^{(l)}$ max pooling operation does the maximum activation across all nodes. The $||$ symbol denotes the concatenation of the two vectors (mean and max). Finally, the summary vector is calculated as the sum of all these summaries [163]

$$s = \sum_{l=1}^{L_p} s^{(l)}, \tag{8.34}$$

which aggregates information from multiple pooling stages. This allows the classifier to retain information from different levels of graph resolution, ensuring that both fine-grained node-level features and more global event-level structures contribute to the final prediction. This aggregated vector serves as an input layer for an MLP for the final event-level classification. The pooling operations and the subsequent MLP structure are illustrated in Fig. 8.16.

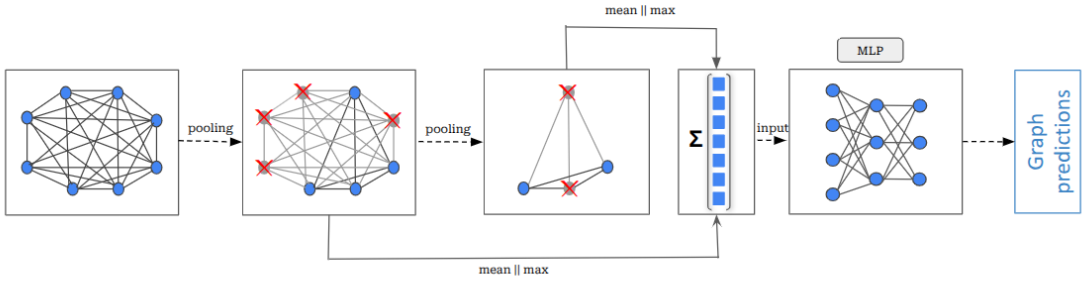


Figure 8.16: Schematic representation of additional pooling in the GLP graph transformer head. Node embeddings from the final message-passing layer are pooled (mean and/or max) after masking the selected nodes (red crosses). This is done by removing or down-sampling a subset of nodes and their incident edges, thereby producing a coarser graph representation. The pooled summaries done per-stage are concatenated with a global max over learned node embeddings for the final flattening, and the resulting summaries are aggregated (Σ) into a fixed-length vector which serves as the input to an MLP that outputs event-level predictions. Figure adapted from Ref. [163].

GLP hyperparameter settings and training

The design of the GLP architecture involves several key hyperparameters that control both the pooling operation and the subsequent classification network. Two hyperparameters govern the pooling step: the pooling ratio $k \in (0, 1]$, which specifies the fraction of nodes retained at each pooling layer, and the position of the pooling operation after a chosen number of graph transformer blocks. Although it is in principle possible to remove a fixed number of nodes, a ratio is used in this work to provide flexibility with varying jet multiplicities. In addition, the MLP applied after pooling introduces further hyperparameters, namely the embedding dimension, the number of hidden layers, and thus the overall size of the network. A complete overview of the chosen hyperparameters is provided in Table 8.4. These settings stabilize and shape learning. BatchNorm with leakyReLU improves gradient flow and training stability, while Adam with a learning rate of 10^{-2} provides fast convergence for the chosen batch size of 2048. The graph backbone uses five attention heads across eight GN blocks with sum aggregation, concatenating heads preserves complementary relation patterns, and the 'root weight' plus additive bias allows each node to retain self-information alongside messages. Regularization is applied through a 0.5 dropout on the normalized attention coefficients, the output MLP with three layers compresses a 64-dimensional embedding with a reduction factor of 0.8 to the final seven-class logits, which are explained below.

The event classes are constructed based on the simulated process type, subdivided as shown in Table 7.2, 7.3, with further distinction made by the number and flavor of additional jets. Seven

Table 8.4: Hyperparameter settings for the GLP architecture.

Hyperparameter	Value
Batch size	2048
Normalization scheme	BatchNorm
Optimizer	Adam
Loss function	Cross-entropy
Activation function	leakyReLU
Embedding dimension	32
Aggregation method	Sum
Number of multi-head attentions	5
Number of GN blocks	8
Learning rate	10^{-2}
Mutli-head consolidation	Concat
Gate residual information weighting	False
Dropout probability of normalized attention coefficients	0.5
Additive bias b	True
Root weight	True
Number of output (MLP) layers	3
Output (MLP) embedding dimension	64
Output embedding dimension reduction factor	0.6
Output layer (MLP) dimension	7

classes are defined for the GLP training in the semileptonic channel. The seven classes are $t\bar{t}$+B, $t\bar{t}$ + H(B), $t\bar{t}$ + Z(B), $t\bar{t}$ + Z(nonB), $t\bar{t}$+C, $t\bar{t}$ + LF, and *other*. The $t\bar{t}$+B class consists of simulated semileptonic $t\bar{t}$ events with one or two additional b jets. Similarly, the $t\bar{t}$H(B) ($t\bar{t}$Z(B)) class includes all simulated $t\bar{t}$+H ($t\bar{t}$+Z) events with a subsequent decay of H $\rightarrow b\bar{b}$ (Z $\rightarrow b\bar{b}$), leading to one or two additional b-jets. Events of simulated $t\bar{t}$+Z processes that do not end in Z $\rightarrow b\bar{b}$ final states are assigned to a separate class, denoted $t\bar{t}$Z(nonB). The $t\bar{t}$+C class includes events with one or two additional c-jets that are not already classified as $t\bar{t}$+B. In the absence of additional heavy-flavor jets, $t\bar{t}$+jets events are assigned to the $t\bar{t}$ + LF class. Finally, events that do not fit into any of the above categories are grouped into the *other* class. This includes, for example, miscategorized $t\bar{t}$+jets events, as well as background processes entering the semileptonic phase space. The full composition of the training dataset is summarized in Table 8.5. A total of about 0.64 million events are used for the GLP model training. Similar to the NLP training, the datasets are partitioned into 90% for training, 5% for validation, and 5% for testing. All networks are trained using the categorical cross-entropy loss function, as defined in Eq. 8.29. Gradient computation and parameter updates are performed with the Adam optimizer, as mentioned earlier. Training is performed for up to 100 epochs over 10 repetitions, with early stopping applied if the validation AUC does not improve further. The GLP model is trained on simulated samples corresponding to 2016preVFP, 2016postVFP, 2017, and 2018 data-taking periods, here, also including all relevant background processes. Since the individual classes contain different numbers of events, the loss function is weighted according to the relative class fractions to ensure that all classes contribute equally, preventing

Table 8.5: Event yields per class: GLP

Process	No. of events
$t\bar{t}B$	100000
$t\bar{t}H$	100000
$t\bar{t}Z(B)$	93764
$t\bar{t}Z(nonB)$	100000
$t\bar{t}C$	97196
$t\bar{t}LF$	100000
other	203498
$\sum$(events)	794458

the network from being biased toward the more abundant processes. In particular, the $t\bar{t}B$, $t\bar{t}C$, and $t\bar{t}LF$ categories provide a substantially larger number of simulated events compared to rarer processes like $t\bar{t}H$ or $t\bar{t}Z(B)$. While these higher statistics increase the variability of the training data for these classes, simply enlarging the imbalance further by adding more $t\bar{t}jj$ events does not lead to performance gains. On the contrary, it often degrades the classification score, even when the loss is reweighted. For this reason, the training composition summarized in Table 8.5 is chosen as a compromise. It provides sufficient statistics for the abundant classes, while preserving a balanced representation of the rarer ones. This setup yields the best overall performance in terms of discrimination power across all seven chosen GLP classes.

8.5.2 GLP performance metrics and evaluation

The evaluation of multiclass event classification with GLP is based on standard classification metrics. In particular, the model performance is assessed using multiclass ROC AUC scores and confusion matrices, which together provide a qualitative and quantitative measure of the classifier's discrimination power across all defined classes.

The loss curves are used to monitor the training stability. The evolution of the loss and ROC–AUC values during training is shown in Fig. 8.17 (left). Both the training and validation datasets are monitored as a function of the number of epochs. A smooth decrease in the loss is observed over the first 25 epochs, after which the learning rate is adjusted. Following this point, the training enters a saturation regime, with only marginal improvements in the performance. The validation curve exhibits slightly larger fluctuations compared to the training curve, but no sign of overfitting is observed. The corresponding ROC–AUC values, also displayed in Fig. 8.17 (right), confirm this behavior. The performance improves steadily during the early epochs before stabilizing after the learning rate adjustment. The final ROC–AUC on the validation dataset reaches a value of approximately 0.8, demonstrating good separation power of the GLP classifier across the event classes.

In addition to the ROC-AUC-based performance evaluation, a dedicated validation of the GLP score shapes is performed to assess the stability of the classifier and to check for possible overtraining. This is achieved by splitting events according to the parity of their event identifier,

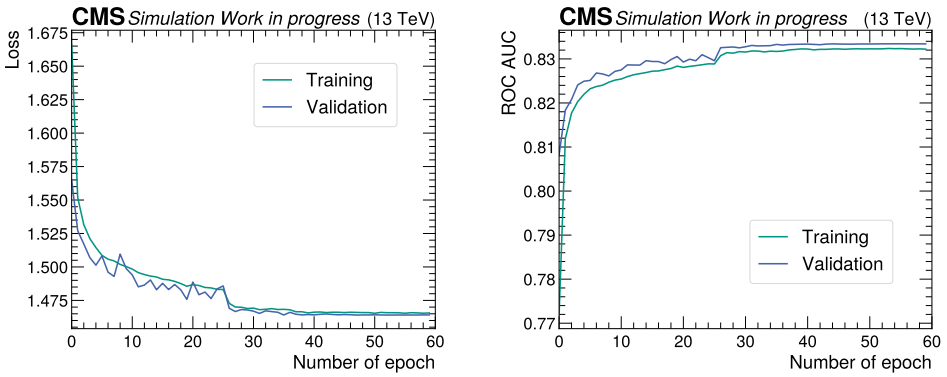


Figure 8.17: Dynamics of training of the GLP classifier. On the left, cross-entropy loss versus epochs for training (green) and validation (blue) is shown. The right plot shows ROC-AUC versus epochs for the same splits, rising steadily and saturating after the learning rate change. The final validation ROC-AUC value is 0.83, evidencing good class separation.

defining 'seen' events (even-numbered) and 'unseen' events (odd-numbered). If the classifier were overfitting, significant differences between the GLP score distributions for these subsets would be apparent. For each output node, the GLP score distributions are compared for seen and unseen events across all contributing simulated processes. The difference between the two distributions is quantified using the Kullback–Leibler (KL) divergence [166], which is widely used in probabilistic ML as a measure of discrepancy between probability distributions [167]. Across all classes, the KL values remain small, and the shapes of the two distributions are in agreement within statistical uncertainties. This demonstrates that the classifier generalizes well and does not rely on event-specific fluctuations in the training subset. An example distribution for the $t\bar{t}$+H output node is shown in Fig. 8.18, illustrating the consistency between the two event subsets. The scores of all remaining output nodes are displayed alongside the $t\bar{t}$+H score, providing a complete view of how the classifier distributes probabilities across the different classes for events classified as $t\bar{t}$+H. The corresponding distributions for the remaining output nodes are provided in Appendix E.

The model is evaluated on the test dataset. Confusion matrices (CM) complement ROC-AUC by exposing class-to-class confusions rather than a single aggregate score. Thus, CMs reveal misclassification patterns among the classes, whereas ROC AUC values provide an aggregated measure of separation power. Both the efficiency matrix (row-normalized) and the purity matrix (column-normalized) are shown, see Fig. 8.19. Efficiency entries $\epsilon_{ij} = \frac{N(true=i,pred=j)}{N(true=i)}$ quantify how events of true class i are distributed over predictions (rows sum to 1), while purity entries $\pi_{ij} = \frac{N(true=i,pred=j)}{N(pred=j)}$ quantify how clean each predicted class j is in terms of its true composition (columns sum to 1). The main misclassification arises between heavy-flavor-enriched classes, such as $t\bar{t} + H(b\bar{b})$, $t\bar{t} + Z(b\bar{b})$, and $t\bar{t}$+$b\bar{b}$, which share similar jet multiplicities and b-tag signatures, and between the charm and light-flavor categories ($t\bar{t} + c\bar{c}$ vs. $t\bar{t} + LF$), reflecting overlapping kinematics and tagger responses. Taken together with

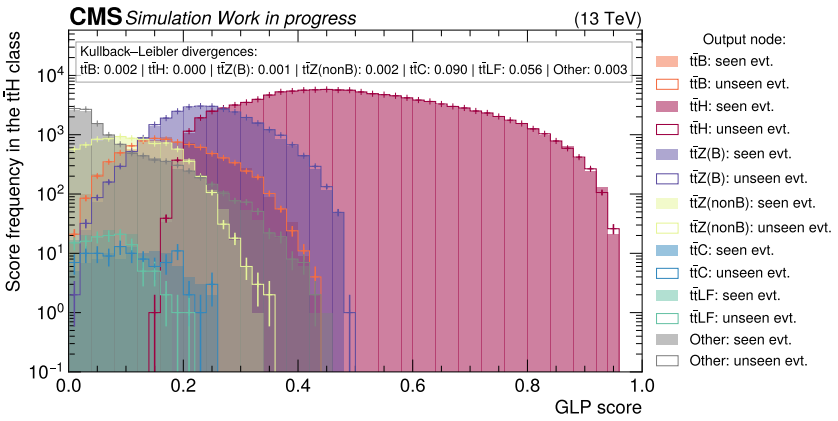


Figure 8.18: Validation of GLP score stability for the tt̄+H output node. The GLP score distributions are shown separately for even-numbered (seen) and odd-numbered (unseen) events. The agreement between the two confirms the absence of overtraining and validates the robustness of the score shapes.

the ROC-AUC, these matrices indicate overall separation while highlighting where additional features or calibration could reduce specific misclassifications.

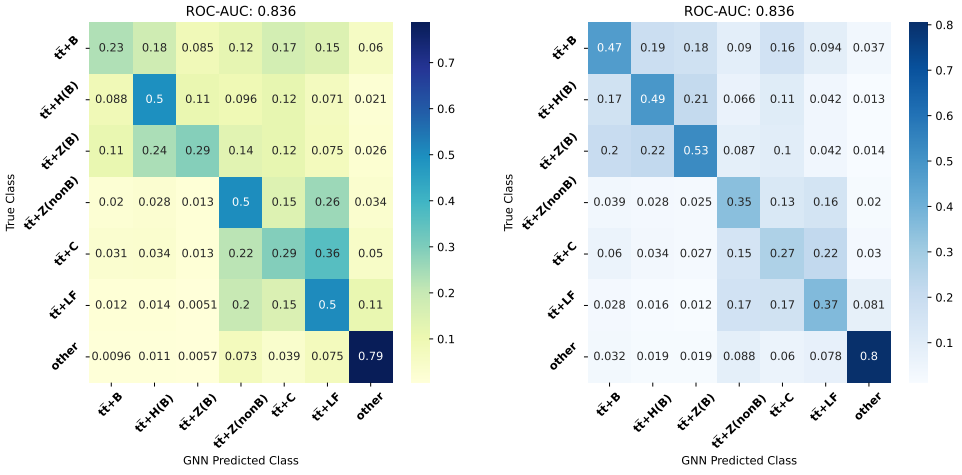


Figure 8.19: Confusion matrices for the GLP test dataset. The left matrix shows the row-normalized efficiency matrix, where for each true class, the distribution of predicted classes is displayed. The column-normalized purity matrix is shown on the right, for each predicted class, the composition of true classes. Diagonal intensity reflects correct classification, whereas off-diagonals highlight confusion between different predicted classes, notably among heavy-flavor enriched classes.

To place greater emphasis on events that are harder to classify, one strategy is to introduce a reweighting of the loss function. This can be achieved by modifying the standard cross-entropy

loss with an additional factor

$$(1 - p_t)^\gamma,$$

leading to the so-called focal loss [168], defined as

$$FL(p_t) = -(1 - p_t)^\gamma \log(p_t), \tag{8.35}$$

where p_t denotes the predicted probability of the correct class and $\gamma \geq 0$ is a tunable focusing parameter. The focal loss is designed to down-weight well-classified examples while giving more weight to misclassified ones, thereby improving the treatment of difficult cases. In the context of [168], however, applying focal loss with different values of γ did not lead to any measurable improvement in the GLP model performance.

A further reduction of confusion between classes could, in principle, be achieved by including additional discriminating features such as the two-dimensional heavy-flavor taggers, CvB and CvL. Nevertheless, as discussed in Sec. 8.3.1, these variables are not incorporated in the final model due to the failure of the goodness-of-fit test for the input feature distributions. Experimental trainings that included these observables in the node feature set indicated a reduction of confusion and a performance gain, but the inconsistency with data–MC agreement precluded their use in the final setup.

Following the assessment of the classifier performance and stability, a representative yield table for the 2018 dataset obtained in each GLP class is summarized in Table 8.6. Equivalent tables for datasets, which exhibit the same qualitative behavior, are provided in Appendix D. These yields quantify how the different processes populate the classifier-defined categories and provide a basis for evaluating signal-background ratios relevant for the statistical analysis. Each column corresponds to a GLP output node, and the process associated with that class is treated as the signal (S) and other processes as the backgrounds (B). The rows, therefore, indicate how strongly each physical process is represented in each category, illustrating the classifier's ability to separate heavy-flavor–enriched processes from the more abundant $t\bar{t}$ +jets backgrounds, particularly $t\bar{t} + \text{LF}$. The structure of the table exhibits several characteristic features. The $t\bar{t} + \text{LF}$ and 'other' columns are dominated by light-flavor and non–$t\bar{t}$ backgrounds, as expected, with the $t\bar{t} + \text{LF}$ class containing the largest event counts due to the high production rate of these processes. The $t\bar{t}$+B, $t\bar{t}$+H, and $t\bar{t}$+Z classes exhibit clear enhancements of their respective signal processes relative to the background, though with varying purity depending on the class.

The signal-to-background ratios (S/B) provide a quantitative measure of this separation. While S/B measures the purity of a category, $S/\sqrt{B}$ quantifies its statistical power. The $t\bar{t}$+B class achieves a moderate purity of about 0.77, indicating a useful enrichment of $t\bar{t}$+B events. The $t\bar{t}$+H and $t\bar{t}$+Z classes have lower purities ($S/B \approx 0.035$ and $S/B \approx 0.019$, respectively), consistent with the smaller production cross sections and similar kinematic features shared with $t\bar{t}$+jets backgrounds. Despite the lower purity, these categories still contribute significantly to

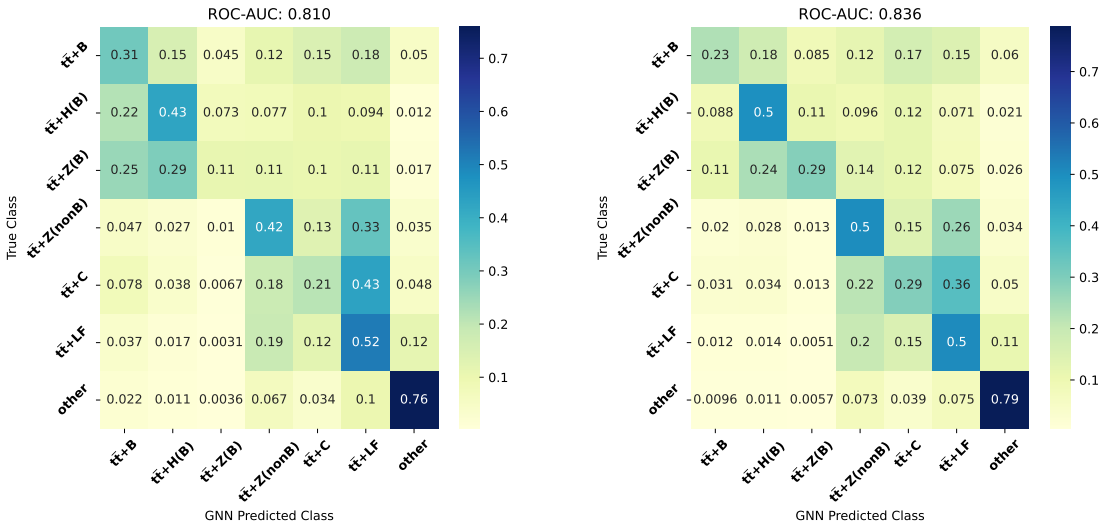
Table 8.6: Event yields per GLP class after the full selection for the 2018 dataset. For each column, the process matching the GLP class label is treated as a signal, while all remaining processes contribute to the background in that column. The event yields for the years 2017, 2016postVFP, and 2016preVFP are provided in Appendix F.

	GLP class							
Process	$t\bar{t}$+B	$t\bar{t}$+H	$t\bar{t}$+Z (B)	$t\bar{t}$+Z (nonB)	$t\bar{t}$+C	$t\bar{t}$ + LF	other	Total
Data	14641	12245	5204	75520	82030	189927	920364	1299931
$t\bar{t}B$	5649	3861	1722	3578	5376	5288	27631	53105
$t\bar{t}H$	97	376	85	335	221	273	1027	2414
$t\bar{t}Z$	64	101	86	501	202	351	1437	2742
$t\bar{t}C$	1853	2179	696	13092	17673	25738	82308	143539
$t\bar{t}$ + LF	2820	3610	1182	50312	45785	148607	511230	763546
$t\bar{t}W$	36	51	24	407	228	348	1426	2520
single top	1793	775	460	4920	4095	8490	56484	77017
VV	8	5	5	152	64	211	3302	3747
W+jets	283	103	145	5403	3052	7510	156527	173023
Z+jets	85	25	27	715	590	1941	22190	25573
QCD	320	25	77	705	1243	8260	48076	58706
Data/MC	1.126	1.102	1.157	0.943	1.045	0.917	1.010	0.995
S/B	0.768	0.035	0.019	–	0.290	–	–	–
$S/\sqrt{B}$	65.8	3.63	1.29	–	71.6	–	–	–

the overall sensitivity because they isolate regions of phase space where the signal is present and where systematic uncertainties differ from those in background-dominated regions. The Data/MC ratios shown in Table 8.6 provide a cross-check of the overall modeling accuracy in each GLP class. A value close to 1 indicates good agreement between the observed yield and the simulated expectation after the full selection. In the heavy-flavor enriched signal classes, the Data/MC ratios lie close to unity, within approximately 5-15%, suggesting that the simulation describes the overall normalization of the top-quark plus heavy-flavor processes at the level expected before the statistical fit.

8.6 Impact of NLP pre-classification

This section quantifies the performance of the NLP pre-classifier and how it affects the GLP predictions and the physics-target discrimination at the event level. The effect of the NLP classifier on the statistical inference is also studied and explained in Sec. 9.2.1. To understand the contribution of NLP classification, a controlled study is performed with two GLP models trained under identical conditions. The first one is trained with NLP-derived guidance, injecting the additional-jet prediction score itself as a GLP input, whereas the second model is trained without any information derived from the NLP stage. Both models use the same backbone architecture, the same number of trainable parameters, identical hyperparameters (see Table 8.4), the same training and validation splits, and the same non-NLP inputs. This design ensures that any observed performance difference is attributable to the structural information



(a) GLP without NLP information

(b) GLP with NLP guidance

Figure 8.20: Confusion matrices on the test set for the GLP trained (left) without and (right) with NLP-derived information. Matrices are row-normalized so each row sums to 1 (darker diagonal indicates better per-class accuracy).

sharing introduced by NLP, rather than to capacity or optimization artifacts. Because the additional-jet score is not provided as a GLP feature in this model (without NLP) and all other inputs remain unchanged, any gains reflect improved use of relational structure induced by the NLP stage, rather than a direct propagation of high-level tagging information into the GLP.

The GLP model that does not use the NLP information achieves an ROC-AUC score of 0.810, while the model with NLP score reaches 0.836. The corresponding confusion matrices are shown in Fig. 8.20, summarizing per-class performance for both configurations. There is a clear improvement in GLP's aggregate separation. The gains concentrate in heavy-flavor enriched classes, where the diagonal increases markedly. $t\bar{t}$+H (B) improves from ~ 0.43 to ~ 0.50, roughly 16% improvement in identifying those events correctly. $t\bar{t}$+C shows a 38% increase, and $t\bar{t}$+Z (B) shows a pronounced jump in correctly identifying events, and the 'other' class becomes cleaner, indicating fewer non-other events are misclassified. Though a modest decrease is observed for $t\bar{t}$+B, NLP guidance reduces the symmetric confusions while leaving the global architecture and training setup unchanged overall. The net effect is higher diagonal mass and a higher ROC-AUC. The impact of NLP pre-classifier on the fit model is explained in the Sec. 9.2.1.

This chapter introduced a graph-based learning strategy tailored to the inclusive 5-jet semileptonic $t\bar{t}$ phase-space. Starting from a principled graph formalism and message-passing updates, an NLP and a GLP were designed and trained with stable optimization and regularization settings. The NLP stage identifies additional jets at the object level and, by providing calibrated relational cues, improves the accuracy of the subsequent GLP event classification. Controlled

tests, keeping the model size and hyperparameters fixed, showed that sharing NLP information with the GLP improves overall event classification. The GLP model is then applied across the full analysis phase space to produce class probabilities for all events. These classified events form the basis of the statistical measurement; their templated distributions feed the likelihood fit from which the signal-strength parameters and corresponding cross sections are extracted (see Chapter 9).

9 Statistical data analysis

In high-energy physics, statistical data analysis provides the bridge between reconstructed event counts and quantitative statements about the underlying physics, such as the production cross sections and the presence or absence of a signal process. This chapter presents the statistical framework used to extract the physical quantities of interest from the data. The objective of the statistical analysis is to quantify the compatibility between the observed data and the theoretical model under study, estimate the parameters of interest (POIs), and determine their associated uncertainties with rigorous treatment of systematic effects. In measurements of rare or subdominant processes, the signal contribution is often small compared to backgrounds and affected by detector effects and systematic uncertainties. A quantitative interpretation of the observed event yields relies on well-defined statistical inference methods.

The structure of the chapter is as follows. Sec. 9.1 summarizes the statistical methodology underlying the measurement in this analysis, covering the formulation of cross-section measurements, the maximum likelihood method, and the application of hypothesis tests for the extraction of the POIs. Sec. 9.2 introduces the construction of the fit model, including the likelihood definition, parameterization of signal and background contributions, and implementation details. Sec. 9.3 presents validation studies based on pseudodata, evaluating the performance of the statistical model. Sec. 9.4 describes the application of the fit model to data. The extracted signal strengths, correlation matrices, and impact of systematic uncertainties are presented and discussed. Sec. 9.5 compares the obtained results with other related measurements and theoretical predictions to assess consistency and relative precision.

9.1 Introduction

The central object of inference is a parametric likelihood that encodes the probability of the observed data given a physics model. POIs are accompanied by nuisance parameters (NPs) that represent systematic effects. Systematic uncertainties entering the statistical inference are

discussed in detail in Sec. 7.6. In the present chapter, these uncertainties are incorporated into the statistical model through constrained NPs and propagated to the final measurement. In binned template analysis, which is standard for multijet, semileptonic $t\bar{t}$+X final states, the likelihood factorizes over categories and bins. It is typically augmented with constraint terms that encode prior knowledge about systematic uncertainties.

9.1.1 Cross-section measurements

Cross-section measurements aim to convert event-level yields into a physically meaningful rate within a well-defined phase space. For a process with cross-section σ, the expected number of selected signal events is governed by the collected data volume, integrated luminosity $\mathcal{L}$, and the probability for events to fall into, and be retained by, the analysis phase space. The latter is summarized by an overall acceptance-efficiency factor $A\varepsilon$. In a single counting region, a convenient first expression is

$$\sigma \;=\; \frac{N_{\text{sig}}}{\mathcal{L}_{int}\, A\,\varepsilon}, \qquad N_{\text{sig}} \;=\; N_{\text{obs}} - N_{\text{bkg}}\,, \tag{9.1}$$

where N_{obs} is the observed yield and N_{bkg} the expected number of background events. In realistic measurements, the application of the above equation is complex. The measurements seek small effects and must incorporate random fluctuations as well as main sources of systematic uncertainty (Sec. 7.6). For more information on the methodologies used and the statistical foundations discussed in subsequent sections, the reader is referred to [169, 170]. Hence, it is often practical to reparameterize the rate in terms of a signal strength

$$\mu = \frac{\sigma}{\sigma_{pred}}, \tag{9.2}$$

defined relative to a chosen prediction σ_{pred}, the SM. Then $\mu = 1$ corresponds to agreement with the prediction, while $\mu \neq 1$ indicates an excess or deficit.

Maximum likelihood fits

To exploit more information than a single count, the analysis typically uses binned distributions (templates) of discriminating observables, such as multivariate outputs or kinematic variables, which enhance separation between signal and backgrounds. The statistical model is constructed as a likelihood $\mathcal{L}(\text{data}|\mu,\theta)$ that factorizes over bins and categories and includes NPs θ that describe systematic effects (see Sec. 7.6). The POIs (μ) encode the physics rates as given in Eq. 9.2. Prior knowledge about individual NPs θ enters via constraint terms $\pi_k(\theta_k)$, enabling coherent uncertainty propagation. This knowledge is also specified in that section. Parameter estimation proceeds through maximum-likelihood fits, typically using the parametric likelihood to handle nuisances,

$$\mathcal{L}(\text{data} \mid \mu,\theta) = \prod_{i=1}^{N_{\text{bins}}} \mathcal{P}(n_i \mid \lambda_i(\mu,\theta)) \prod_{k=1}^{N_{\text{NP}}} \pi_k(\theta_k)\,, \tag{9.3}$$

where $\mathcal{P}$ is the Poisson probability for observing n_i events when λ_i are expected. It is described by

$$\mathcal{P}(n_i \,|\, \lambda_i) = \frac{\lambda_i^{n_i}\, e^{-\lambda_i}}{n_i!}\,. \tag{9.4}$$

The expected bin yield is modeled as

$$\lambda_i(\mu, \theta) = \sum_{m=1}^{N_{\text{sig}}} \mu_m\, s_{mi}(\theta) \;+\; b_i(\theta)\,, \tag{9.5}$$

with s_{mi} and b_i the signal and background templates after all selections. When quoting results relative to a reference SM prediction, the parametrization chosen is Eq. 9.2. The dependence of s_{mi} and b_i on θ covers both normalization and shape variations, following the implementation choices documented in Sec. 7.6.

Point estimates $(\hat{\mu}, \hat{\theta})$ are obtained by maximizing $\mathcal{L}(\text{data}|\mu, \theta)$, equivalently, minimizing the negative log-likelihood (NLL).

$$\text{NLL}(\mu, \theta) \;=\; -\ln \mathcal{L}(\text{data} \mid \mu, \theta) \tag{9.6}$$

Uncertainties on μ that account for systematic uncertainties are obtained with the profile likelihood. For fixed μ, the conditional maximizers $\hat{\theta}(\mu)$ define the likelihood ratio

$$\lambda(\mu) \equiv \frac{L\Big(\mu,\, \hat{\theta}(\mu)\Big)}{L\Big(\hat{\mu},\, \hat{\theta}\Big)}\,. \tag{9.7}$$

This profiling propagates correlations and prior information contained in π_k to the POIs consistently.

Near the optimum, the NLL is well approximated by a quadratic form. The covariance of the fitted parameters $\alpha \in \{\mu, \theta\}$ is estimated from the inverse observed information [171] in the large sample limit,

$$\text{Cov}(\hat{\alpha}) = \left[-\frac{\partial^2 \ln \mathcal{L}}{\partial \alpha\, \partial \alpha^\top} \right]^{-1}_{\alpha=\hat{\alpha}} \tag{9.8}$$

Diagonal elements provide one-standard-deviation uncertainties; off-diagonal elements give linear correlations. One–standard–deviation uncertainties, σ and linear correlation coefficients ρ_{ij} follow from the covariance:

$$\sigma(\alpha_i) = \sqrt{\text{Cov}_{ii}}, \qquad \rho_{ij} = \frac{\text{Cov}_{ij}}{\sqrt{\text{Cov}_{ii}\,\text{Cov}_{jj}}}\,. \tag{9.9}$$

Confidence intervals can also be obtained by inverting the profile-likelihood ratio [172],

$$q_\mu \equiv -2 \ln \lambda(\mu) = -2 \left[\ln \mathcal{L}\Big(\mu,\, \hat{\theta}(\mu)\Big) - \ln \mathcal{L}\Big(\hat{\mu},\, \hat{\theta}\Big) \right], \tag{9.10}$$

and using Wilks' theorem [173] to set thresholds. In one dimension, $q_\mu = 1$ and $q_\mu = 3.84$ correspond approximately to 68% and 95% confidence intervals, respectively. For two parameters, the corresponding thresholds are $q = 2.30$ (68%) and $q = 5.99$ (95%).

These asymptotic results hold when regularity conditions are met. Near physical boundaries (e.g. $\mu \geq 0$), with small samples, or with weakly constrained nuisances, intervals from full profile scans (and, if needed, toy studies) are preferred. Asymptotic results applied to the profile likelihood yield fast, accurate uncertainty estimates and expected sensitivities. This completes the likelihood construction and estimation procedure. Tests of hypotheses are then formulated from profile-likelihood-based test statistics built from $\lambda(\mu)$.

9.1.2 Hypothesis testing and significance

Hypothesis testing and significance use the sampling distribution of test statistics derived from the likelihood ratios. It is derived from the profile likelihood $\lambda(\mu)$ of Eq. (9.7). The background-only case ($\mu = 0$) serves as the null hypothesis (H_0), while 'signal+background' ($\mu > 0$) is treated as the alternative (H_1). The goal is to quantify whether the data are compatible with the null or favor the presence of a signal, and to assign a statistical significance to any observed excess.

To test for the presence of a signal, a one-sided test statistic is used. It is given as,

$$q_0 = \begin{cases} -2 \ln \lambda(0), & \hat{\mu} \geq 0, \\ 0, & \hat{\mu} < 0, \end{cases} \tag{9.11}$$

which is large for data incompatible with $\mu = 0$. Under regularity conditions, q_0 follows the usual asymptotic mixture distribution, yielding the background p-value and Gaussian equivalent significance

$$p_0 = 1 - \Phi(\sqrt{q_0}), \qquad Z = \Phi^{-1}(1 - p_0) = \sqrt{q_0}, \tag{9.12}$$

where Φ denotes the cumulative distribution function (CDF) of the standard normal distribution and Z is the Gaussian equivalent significance. By convention, $Z = 5$ corresponds to "discovery" (corresponding to p-value of 3×10^{-7}) and $Z = 3$ to "evidence".

When setting an upper limit on a specified signal strength $\mu > 0$, the one-sided statistic

$$q_\mu = \begin{cases} -2 \ln \lambda(\mu), & \hat{\mu} \leq \mu, \\ 0, & \hat{\mu} > \mu, \end{cases} \tag{9.13}$$

which yields a p-value for the signal+background hypothesis at that μ,

$$p_\mu = 1 - \Phi(\sqrt{q_\mu}). \tag{9.14}$$

A $(1-\alpha)$ confidence level (CL) upper limit μ_{up} solves the equation $p_{\mu_{\text{up}}} = \alpha$ (e.g. $\alpha = 0.05$ for a 95% CL limit). If the modified frequentist prescription is adopted, the criterion

$$\text{CL}_s = \frac{p_{s+b}}{1-p_b} < \alpha \tag{9.15}$$

is applied, where p_{s+b} is computed from q_μ under the signal+background assumption and p_b from the distribution of the same statistic under the background hypothesis. This protects against excluding models in the presence of downward background fluctuations.

For expected significances and limits, the Asimov dataset provides accurate and fast approximations [169]. It is defined as a representative dataset in which the observed counts in each bin are set equal to their expectation values under a given hypothesis, eliminating statistical fluctuations. By construction, fits to this dataset reproduce the true parameter values, enabling the direct computation of median expected significances or limits without generating ensembles of pseudo-experiments. In the Gaussian regime, one may write

$$q_{\mu,\text{A}} = \frac{(\mu-\mu')^2}{\sigma^2}, \tag{9.16}$$

where μ' is the data-generating value (often 0 for background-only expectations) and σ is the conditional standard deviation of $\hat{\mu}$. The asymptotic formulae of Wilks and Wald underpin Eqs. (9.12)–(9.16) and provide fast approximations for significances and limits. Their validity can be verified using profile-likelihood scans or toy Monte Carlo studies when necessary. Throughout this thesis, results obtained from the collision data are labeled 'observed'. Results derived from pseudo-experiments or from the Asimov construction, which represent the median expectation under a specified hypothesis, are labeled 'expected'.

9.2 Fit model

The statistical interpretation of the analysis is performed using a binned maximum-likelihood fit constructed from the output of GLP scores (see Sec. 8.5). The GLP assigns each event a set of prediction scores corresponding to the seven targeted final states, namely $t\bar{t}$+B, $t\bar{t}$+H (B), $t\bar{t}$+Z (B), $t\bar{t}$+Z (nonB), $t\bar{t}$+C, $t\bar{t}$ + LF, and 'other' background events. For each of these analysis categories, the GLP output scores are histogrammed to produce binned templates before the fit, referred to as pre-fit distributions. These distributions define the expected shapes of signal and background components and form the basis of the likelihood model. A schematic illustration of the pre-fit GLP score distributions for representative categories of the era 2018 is shown in Figs. 9.1, 9.2. These plots highlight the expected separation between signal-like and background-like regions before the fit. These GLP prediction scores serve as the primary discriminants between the different $t\bar{t}$+X processes and constitute the basis of the statistical model. The distributions corresponding to eras 2017, 2016preVFP, and 2016postVFP are portrayed in appendix G.

The event categorization strategy (Sec. 8.5) exploits differences in the multijet structure,

flavor composition, and kinematic features of the final states, allowing the fit to separately constrain the dominant backgrounds and extract the signal strengths across all regions simultaneously. The signal extraction is performed through a simultaneous template fit across all categories. The parameters of interest (POIs) are defined as the signal strengths $\mu_{t\bar{t}H}$, $\mu_{t\bar{t}Z}$, $\mu_{t\bar{t}B}$, and $\mu_{t\bar{t}C}$, which scale the expected yields of the corresponding $t\bar{t}$+X processes relative to their SM predictions as defined by the Eq. 9.2,

$$\mu_i = \frac{\sigma_i}{\sigma_i^{SM}}, \qquad i \in \{t\bar{t}H, t\bar{t}Z, t\bar{t}B, t\bar{t}C\} \tag{9.17}$$

These parameters represent the ratio between the measured cross-section and the SM expectation, and are simultaneously determined in the fit. The likelihood is constructed as the product of Poisson probabilities over all bins and categories, with nuisance parameters (NPs) introduced to account for systematic uncertainties affecting both rate and shape (see Sec. 7.6).

The binning of the GLP output score distributions is chosen to maximize the sensitivity of the fit to pseudodata while avoiding overly fine binning that could introduce artificial constraints and inflate the expected sensitivity. For the $t\bar{t}$+B, $t\bar{t}$+H, and $t\bar{t}$+Z classes, the observed data are consistent with the predictions within the combined statistical and systematic uncertainties. Variations in the predicted distributions among different classes within the same class indicate that this differential information contributes to the discrimination power of the maximum likelihood fit. In contrast, the $t\bar{t}$+C discriminant demonstrates the increased difficulty of identifying $t\bar{t}$+C events, as only a small fraction of events achieve high classifier scores. This behavior is primarily attributed to the absence of dedicated c-tagging information in the classifier, limiting its separation capability. A slight deficit of data relative to the prediction is observed in the high classifier score region. However, the data remain compatible within the assigned uncertainties. In the background class discriminants, the $t\bar{t}$ + Z(nonB) and $t\bar{t}$ + LF classes show small normalization offsets of approximately 6% and 9% respectively, as seen in the ratio of data to simulated events (Table 8.6), while the 'other' class is well described within the uncertainty bands.

Several alternative fit configurations were investigated to ensure a robust and well-behaved likelihood model. These studies included masking individual categories, merging bins with limited statistical precision, introducing normalization (rate) parameters for dominant backgrounds, and freezing selected NPs exhibiting strong correlations or anti-correlations. The configurations were evaluated using goodness-of-fit tests, parameter impact studies, and correlation matrices. The $t\bar{t}$ + LF category was found to be insufficiently modeled and provided no significant improvement in sensitivity in pseudo-data studies. Its inclusion also led to degraded goodness-of-fit (GoF) performance and increased fit instability. It was therefore excluded from the final fit to avoid potential biases arising from mismodeled shapes. However, removing this category alone did not fully resolve the observed tensions, as the $t\bar{t}$ + Z(nonB) category also exhibited residual mismodeling effects. Masking both categories improved the fit behavior but resulted in a loss

of important normalization constraints on $t\bar{t}$+Z (and its separation from $t\bar{t}$+C and partly $t\bar{t}$+B).

To address this while preserving sensitivity, normalization (rate) parameters were introduced, in addition to the uncertainties in Sec. 7.6, for dominant background processes exhibiting mild tension between prediction and data. Since $t\bar{t}$ + LF constitutes a major background in multiple regions, even small normalization mismodeling can significantly influence the global likelihood. Allowing its normalization, together with that of the V+jets background, provides controlled flexibility for the fit to accommodate these differences without biasing other parameters. With this additional flexibility, the $t\bar{t}$ + Z(nonB) category could be reintroduced as a single-bin normalization constraint, retaining its valuable information while mitigating shape-driven mismodeling effects. The resulting configuration yields a stable fit with well-behaved NPs and satisfactory GoF performance. Therefore, this configuration represents the best compromise between statistical sensitivity and modeling reliability and is selected as the baseline model. Additional validation through correlation and impact studies, presented in the following sections, confirms the robustness of the adopted statistical model.

9.2.1 Impact of NLP pre-classifier on fit model

The NLP pre-classifier described in Chapter 9 is not only beneficial in terms of ML performance metrics, but also has a direct and measurable impact on the statistical inference of the signal processes. By providing the object-level information on the likelihood of jets originating from heavy-flavor radiation, the NLP enhances the separation power of the GLP classifier, which serves as the primary discriminant in the fit. To assess the impact of the NLP pre-classifier on the fit model, two otherwise identical configurations are compared, models trained with and without access to NLP-derived information as explained in Sec. 8.6.

Table 9.1 summarizes the fitted signal strengths for the four processes of interest, $t\bar{t}$+H, $t\bar{t}$+Z, $t\bar{t}$+B, and $t\bar{t}$+C, before and after the inclusion of NLP information. It should be noted that the results correspond to statistical-only fits, where all systematic uncertainties are disabled. For this comparison, pseudodata based on the 2018 simulation is used to isolate and illustrate the impact of the NLP pre-classifier on the fit model. The improved event-level separation achieved with NLP guidance translates into a clearer differentiation of the $t\bar{t}$ flavor categories in the final discriminant distributions. As a consequence, the likelihood fit exhibits increased sensitivity to the individual signal-strength parameters. This is reflected in a systematic reduction of the post-fit uncertainties on the signal strengths when NLP information is included. In particular, substantial improvements are observed for processes with smaller cross sections and stronger overlap with background contributions such as $t\bar{t}$+H and $t\bar{t}$+Z, where the uncertainty reduction reaches the level of several tens of percent. Even more for abundant processes such as $t\bar{t}$+$b\bar{b}$ and $t\bar{t} + c\bar{c}$, the NLP-guided model yields a noticeable tightening of the constraints.

Importantly, including NLP information does not introduce additional fit instabilities or biases. The fit converges reliably in both configurations, and the post-fit NP pulls and constraints remain well-behaved. Since the NLP-derived scores are treated as standard input features

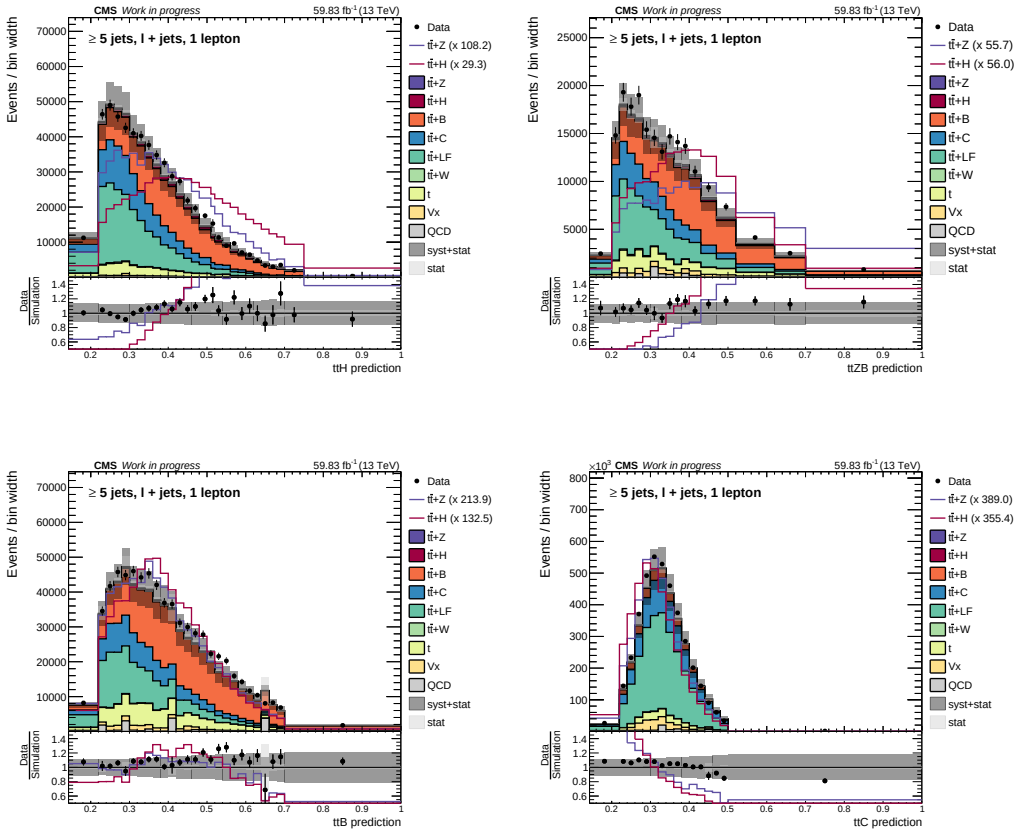


Figure 9.1: Pre-fit GLP score distributions for the signal classes $t\bar{t}$+H, $t\bar{t}$+Z, $t\bar{t}$+B, and $t\bar{t}$+C in the 2018 data-taking era. The top row portrays $t\bar{t}$+H, $t\bar{t}$+Z, and $t\bar{t}$+B scores, while the bottom row portrays the $t\bar{t}$+C score. The simulation is shown as stacked histograms of all processes, and the data is represented by black markers. The $t\bar{t}$+H and $t\bar{t}$+Z predictions are additionally overlaid as line distributions normalized to the total yield to enhance their visibility. The gray band indicates the total uncertainty on simulation, including both statistical and systematic components described in Sec. 7.6. The lower panel displays the data-to-simulation ratio.

Table 9.1: Comparison of fitted signal-strength parameters on pseudodata (2018) obtained using GLP templates trained without and with NLP-derived information. The results also correspond to fits with only statistical uncertainties, with all systematic uncertainties disabled.

Process	Without NLP	With NLP
$\mu^{\text{pseudo}}_{t\bar{t}+B}$	$1.00^{+0.05}_{-0.05}$	$1.00^{+0.02}_{-0.02}$
$\mu^{\text{pseudo}}_{t\bar{t}+C}$	$1.00^{+0.32}_{-0.26}$	$1.00^{+0.03}_{-0.03}$
$\mu^{\text{pseudo}}_{t\bar{t}+H}$	$1.00^{+0.78}_{-0.78}$	$1.00^{+0.28}_{-0.27}$
$\mu^{\text{pseudo}}_{t\bar{t}+Z}$	$1.00^{+2.89}_{-2.89}$	$1.00^{+0.63}_{-0.62}$

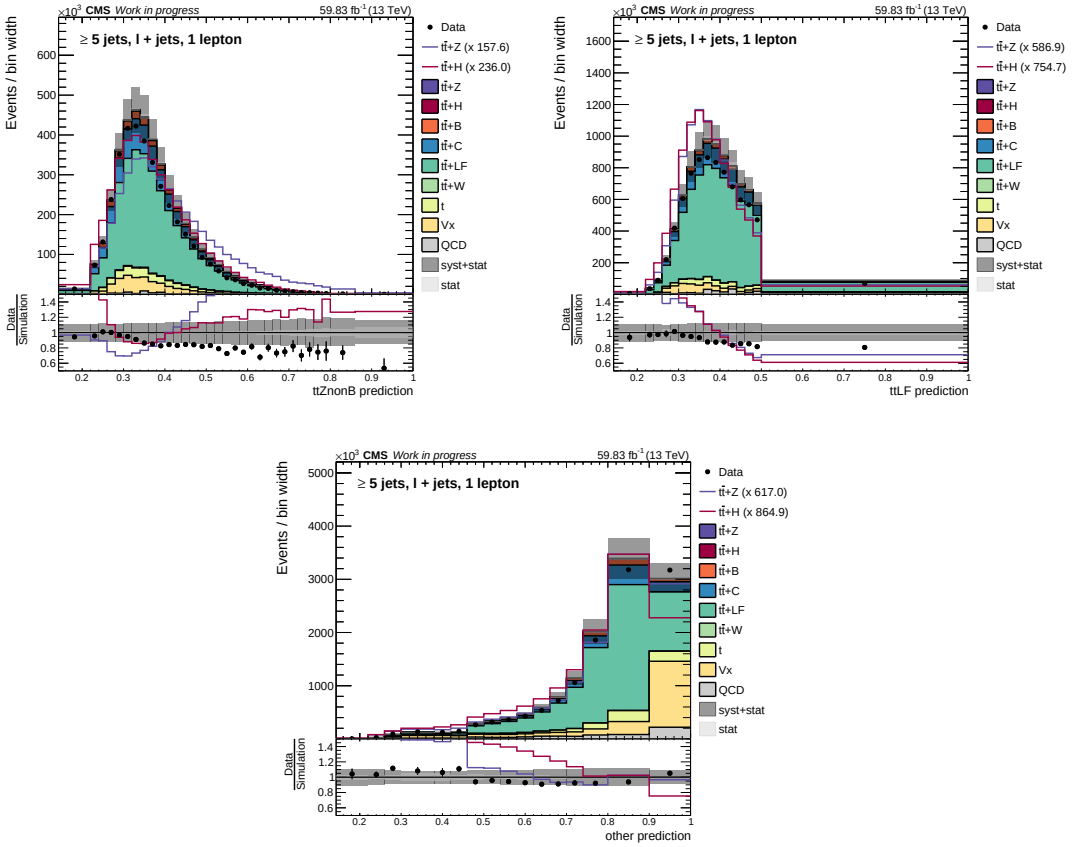


Figure 9.2: Pre-fit GLP score distributions for the background classes $t\bar{t}+Z(nonB)$, $t\bar{t}+LF$, and 'other' in the 2018 data-taking era. The simulation is shown as stacked histograms of all processes, and the data is represented by black markers. The $t\bar{t}+H$ and $t\bar{t}+Z$ predictions are additionally overlaid as line distributions normalized to the total yield to enhance their visibility. The gray band indicates the total uncertainty on simulation, including both statistical and systematic components described in Sec. 7.6. The lower panel displays the data-to-simulation ratio.

to the GLP, all associated experimental and theoretical uncertainties are consistently propagated through the event classification and encoded in the shape variations used in the likelihood.

Overall, the NLP classifier enhances the measurement's statistical power by sharpening event-level discrimination, resulting in tighter constraints on the signal-strength parameters and, consequently, more precise cross-section measurements. Since the uncertainty reduction shown here reflects the intrinsic improvement in the statistical sensitivity of the fit model, the overall impact in the full analysis, including systematic uncertainties, is expected to be more moderate. Nevertheless, the enhanced event separation from the NLP pre-classifier still contributes to improved measurement precision.

9.3 Fit results: Pseudodata

To assess the expected performance of the analysis in a realistic configuration, the full statistical model is evaluated using eras of Run-II pseudodata: the 2018, 2017, 2016postVFP, and 2016preVFP. In contrast to studies in Sec. 9.2.1, where statistical-only fits were used to isolate methodological effects, the results presented here include all experimental and theoretical systematic uncertainties described in Sec. 7.6. The pseudodata correspond to an Asimov dataset constructed from the nominal simulation, with event yields matching the integrated luminosity for each year, and the fits are performed simultaneously across all categories. In practice, the content of each bin is obtained by summing the simulated event weights, so that the dataset reflects the full SM prediction. It therefore realizes the alternative hypothesis in which the targeted signal processes are present at their expected cross-sections, corresponding to $\mu = 1$ for all measured components [169]. The fits to pseudodata are performed using the profile likelihood based on the test statistic $-2\Delta\ln\mathcal{L}$. The model contains four parameters of interest $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, $\mu_{t\bar{t}+Z}$, while all remaining parameters enter as nuisance parameters describing all the systematic and statistical uncertainties. The total number of NPs depends on the data-taking era due to era-specific calibrations and uncertainties: 231 (2018), 216 (2017), 243 (2016postVFP), and 215 (2016preVFP). All NPs are profiled simultaneously with the POIs in the likelihood maximization, ensuring a consistent propagation of correlations and constraints. The resulting best-fit values of the four POIs for each Run-II era are summarized in Table 9.2, while the expected significances derived from the same fits are summarized in Table 9.3.

The expected significances obtained from pseudodata quantify the intrinsic sensitivity of the simultaneous fit to each process. As anticipated from the relative production rates and the chosen heavy-flavor enriched phase space, the $t\bar{t}$+B contribution exhibits a very large sensitivity, reaching approximately 35σ in the combined Run-II dataset. Such a large significance is produced by a steep increase in the NLL as the signal strength approaches zero, indicating that the hypothesis of vanishing $t\bar{t}$+B contribution is strongly disfavored by the model and the dataset. A similar behavior was also observed in the dileptonic channel [93]. The per-era significances combine approximately in quadrature, indicating a statistically consistent scaling with luminosity. The sensitivity to $t\bar{t}$+C is moderate, at about 5σ in Run-II, reflecting its smaller cross-section and the more limited separation power between additional c-jets and other heavy-flavor components. For $t\bar{t}$+H, the expected sensitivity is around 3σ, consistent with its comparatively small production rate and the partial correlation with other heavy-flavor processes in this high jet multiplicity phase space. The $t\bar{t}$+Z contribution shows the weakest sensitivity in this topology, with an expected significance slightly above 1σ, as the selected phase space is not specifically optimized for this process. Overall, the hierarchy of expected sensitivities follows the relative cross sections and the discriminating power of the classifier, and provides a consistent picture of the analysis reach before examining the observed data.

The fitted signal rates are all compatible with the injected SM values of unity, demonstrating the closure of the statistical model and the absence of biases in the extraction procedure. As

Table 9.2: Best-fit signal strengths obtained from pseudodata fits including all systematic uncertainties. Results are shown separately for each Run-II data-taking era in the inclusive (≥ 5 jets) category. Uncertainties correspond to the profiled 68% confidence intervals.

	2018	2017	2016postVFP	2016preVFP	Run-II
$\mu^{\text{pseudo}}_{\text{t}\bar{\text{t}}+\text{B}}$	$1.00^{+0.45}_{-0.30}$	$1.00^{+0.38}_{-0.27}$	$1.00^{+0.47}_{-0.31}$	$1.00^{+0.47}_{-0.31}$	$1.00^{+0.35}_{-0.26}$
$\mu^{\text{pseudo}}_{\text{t}\bar{\text{t}}+\text{C}}$	$1.00^{+0.43}_{-0.33}$	$1.00^{+0.50}_{-0.40}$	$1.00^{+0.68}_{-0.51}$	$1.00^{+0.66}_{-0.50}$	$1.00^{+0.29}_{-0.24}$
$\mu^{\text{pseudo}}_{\text{t}\bar{\text{t}}+\text{H}}$	$1.00^{+0.50}_{-0.48}$	$1.00^{+0.56}_{-0.52}$	$1.00^{+0.78}_{-0.79}$	$1.00^{+0.73}_{-0.68}$	$1.00^{+0.39}_{-0.36}$
$\mu^{\text{pseudo}}_{\text{t}\bar{\text{t}}+\text{Z}}$	$1.00^{+1.21}_{-1.20}$	$1.00^{+1.26}_{-1.21}$	$1.00^{+1.85}_{-1.71}$	$1.00^{+1.85}_{-1.80}$	$1.00^{+0.84}_{-0.82}$

Table 9.3: Expected significances for each signal process obtained from pseudodata, including all systematic uncertainties.

	2018	2017	2016postVFP	2016preVFP	Run-II
$Z^{\text{pseudo}}_{\text{t}\bar{\text{t}}+\text{B}}$	21.94	21.14	13.71	13.78	35.42
$Z^{\text{pseudo}}_{\text{t}\bar{\text{t}}+\text{C}}$	3.50	2.68	2.07	2.05	5.02
$Z^{\text{pseudo}}_{\text{t}\bar{\text{t}}+\text{H}}$	2.19	2.02	1.27	1.48	3.08
$Z^{\text{pseudo}}_{\text{t}\bar{\text{t}}+\text{Z}}$	0.83	0.82	0.58	0.55	1.22

expected, the precision follows the hierarchy of production rates. Although the $\text{t}\bar{\text{t}}$+B component has the largest effective yield (for example, Table 8.6) in the selected phase-space, the $\text{t}\bar{\text{t}}$+C signal strength is measured with slightly better precision. This indicates that the precision is influenced not only by signal yields but also by correlations with backgrounds and NPs. The rarer EW processes $\text{t}\bar{\text{t}}$+H and $\text{t}\bar{\text{t}}$+Z show substantially larger intervals, reaching about $\sim$50% for $\text{t}\bar{\text{t}}$+H per era, improving to about 38% in Run-II, while $\text{t}\bar{\text{t}}$+Z remains close to unity, reflecting their smaller cross-sections and stronger correlations with background contributions. The comparison of the statistical only (Table 9.4) and full systematic fits (Table 9.2) indicates that the measurement of $\text{t}\bar{\text{t}}$+B and $\text{t}\bar{\text{t}}$+C remains strongly systematically dominated, while the statistical uncertainties are at the percent level and therefore contribute only marginally to the total uncertainty. The impact of systematic uncertainties is pronounced for $\text{t}\bar{\text{t}}$+H and $\text{t}\bar{\text{t}}$+Z, whose total uncertainties are further amplified by limited yields. This behavior is consistent across all eras and reflects the stronger dependence of the rare processes on modeling, background composition, and NP correlations. The statistical advantage of the semileptonic channel is evident in comparison with the dileptonic analysis of Ref. [93]. For 2018, the statistical uncertainty on the $\text{t}\bar{\text{t}}$+H component is more than a factor of two smaller in this analysis than in the dileptonic case, confirming that the present measurement is significantly less statistics-limited.

The combined Run-II results exhibit a significant improvement in precision with respect to

the individual data-taking eras, as expected from the simultaneous fit of all years. The Run-II signal strengths reach relative precisions of about 30% for $t\bar{t}$+B and 27% for $t\bar{t}$+C, computed from the average of the asymmetric 68% confidence intervals. The uncertainties are reduced to

Table 9.4: Best-fit signal strengths obtained from pseudodata fits including statistical uncertainties only. Results are shown separately for each Run-II data-taking era in the inclusive (≥ 5 jets) category.

	2018	2017	2016postVFP	2016preVFP	Run-II
$\mu^{\text{stat}}_{t\bar{t}+B}$	$1.00^{+0.02}_{-0.02}$	$1.00^{+0.03}_{-0.03}$	$1.00^{+0.04}_{-0.04}$	$1.00^{+0.04}_{-0.04}$	$1.00^{+0.01}_{-0.01}$
$\mu^{\text{stat}}_{t\bar{t}+C}$	$1.00^{+0.03}_{-0.03}$	$1.00^{+0.03}_{-0.03}$	$1.00^{+0.06}_{-0.06}$	$1.00^{+0.06}_{-0.06}$	$1.00^{+0.02}_{-0.02}$
$\mu^{\text{stat}}_{t\bar{t}+H}$	$1.00^{+0.28}_{-0.27}$	$1.00^{+0.33}_{-0.33}$	$1.00^{+0.56}_{-0.54}$	$1.00^{+0.52}_{-0.50}$	$1.00^{+0.18}_{-0.18}$
$\mu^{\text{stat}}_{t\bar{t}+Z}$	$1.00^{+0.63}_{-0.62}$	$1.00^{+0.75}_{-0.74}$	$1.00^{+1.23}_{-1.20}$	$1.00^{+1.21}_{-1.18}$	$1.00^{+0.42}_{-0.41}$

about 38% for $t\bar{t}$+H in the Run-II combination compared to typical single-era uncertainties of 50-80% for $t\bar{t}$+H. For $t\bar{t}$+Z, the Run-II uncertainty reduced to roughly 80% compared to values exceeding unity for $t\bar{t}$+Z. This indicates that these measurements remain constrained primarily by limited signal yields and strong correlations with background normalizations and modeling uncertainties. In particular, the $t\bar{t}$+Z component remains close to the sensitivity threshold, with an expected Run-II significance below 2σ, despite the gain from the full dataset. Overall, the Run-II fit demonstrates the statistical consistency of the model across all eras and confirms that the dominant limitations of the measurement differ by process. While $t\bar{t}$+B and $t\bar{t}$+C benefit substantially from the increased dataset, the precision of $t\bar{t}$+H and $t\bar{t}$+Z remains driven by systematic effects and a small cross-section. This behavior is consistent with expectations based on the relative production cross-sections and validates the robustness of the combined Run-II statistical framework. This behavior is further quantified in the Run-II pseudodata summary presented in Fig. 9.3, which displays the best-fit signal strength parameters and their decomposed statistical and systematic uncertainties, thereby highlighting the process-dependent interplay between statistical precision and systematic limitations in the combined Run-II fit.

In the current implementation, the fit to pseudodata employs the pure GLP output score as the sole discriminating observable, as explained in previous sections. While this choice already provides substantial separation power between the $t\bar{t}$+X processes and the dominated backgrounds, the GLP model offers additional possibilities for constructing more dedicated discriminants. Following strategies successfully applied in the CMS $t\bar{t} + H(b\bar{b})$ analysis (Ref. [113]), a ratio-based discriminant can be derived from the individual class probabilities predicted by the GLP. Such ratios enhance sensitivity by exploiting the relative likelihood of competing hypotheses, for example, between $t\bar{t}$+H, $t\bar{t}$+Z, and $t\bar{t}$ + jets categories, rather than relying on the absolute score of a single node. These alternative observables can be incorporated either as one–dimensional templates replacing the current GLP score or in a multi–dimensional fit configuration. Studies with pseudodata indicate that ratio discriminants can improve the

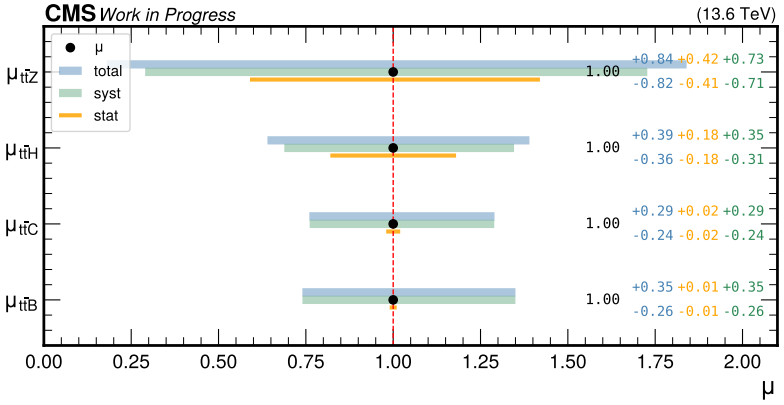


Figure 9.3: Summary of the best-fit signal strength parameters obtained from the combined Run-II fit using asimov pseudo-data. The markers indicate the best-fit values, while the horizontal bands represent the total, statistical, and systematic uncertainties, respectively. The figure highlights the process-dependent interplay between statistical and systematic effects in the measurement.

stability of the signal strength extraction in specific scenarios, as demonstrated in Ref. [113], which improves the sensitivity on $t\bar{t}$+H by 18%.

Three representative configurations are discussed in detail in the following. Table 9.5 summarizes the expected performance of these configurations, where the corresponding signal strength uncertainties obtained from pseudodata fits are compared to the nominal approach based solely on the pure GLP score. The HvB model introduces a likelihood ratio discriminant, R_{HvB}, for events in $t\bar{t}$+H and $t\bar{t}$+B categories, and is defined as,

$$R_{HvB} = \frac{GLP(t\bar{t}+H)}{GLP(t\bar{t}+H) + GLP(t\bar{t}+B)} \tag{9.18}$$

It is used together with the pure GLP scores of the remaining classes ($t\bar{t}$+C, $t\bar{t}Z$(B), $t\bar{t}Z$(nonB)) and designed to address the challenging discrimination between $t\bar{t}$+H and $t\bar{t}$+B processes.

Table 9.5: Expected signal strength modifiers of the era 2018 under different statistical model configurations.

	Pure GLP	HvB	HvZB	2D-unrolled
μ_{ttB}^{pseudo}	$1.00^{+0.45}_{-0.30}$	$1.00^{+0.46}_{-0.30}$	$1.00^{+0.47}_{-0.31}$	$1.00^{+0.43}_{-0.29}$
μ_{ttC}^{pseudo}	$1.00^{+0.43}_{-0.33}$	$1.00^{+0.67}_{-0.52}$	$1.00^{+0.53}_{-0.43}$	$1.00^{+0.51}_{-0.39}$
μ_{ttH}^{pseudo}	$1.00^{+0.50}_{-0.48}$	$1.00^{+0.57}_{-0.51}$	$1.00^{+0.88}_{-0.94}$	$1.00^{+0.56}_{-0.50}$
μ_{ttZ}^{pseudo}	$1.00^{+1.21}_{-1.20}$	$1.00^{+1.46}_{-1.42}$	$1.00^{+1.67}_{-1.67}$	$1.00^{+1.05}_{-0.99}$

The HvZB model is built analogously, targeting the separation between $t\bar{t}$+H and $t\bar{t}Z$(B) processes. The corresponding ratio is,

$$R_{\mathrm{HvZB}} = \frac{\mathrm{GLP}(t\bar{t}+\mathrm{H})}{\mathrm{GLP}(t\bar{t}+\mathrm{H}) + \mathrm{GLP}(t\bar{t}\mathrm{Z}(\mathrm{B}))} \tag{9.19}$$

where the ratio is used instead of the GLP score of $t\bar{t}$Z(B) and $t\bar{t}$+H single nodes. Finally, the 2D-unrolled model represents a combination of the two ratio-based approaches. Both HvB and HvZB discriminants are exploited simultaneously by constructing a two-dimensional template, complemented by the pure GLP scores of the remaining classes. Within each interval of the ratio R_{HvB}, the distribution of the complementary observable R_{HvZB} is fitted. In this configuration, the HvB ratio acts as a categorization variable, and for every HvB interval, a dedicated shape of the HvZB discriminant is built. The collection of these HvZB spectra is subsequently unrolled to form the final template used in the likelihood fit. The unrolled bins, therefore, correspond to regions of similar HvB response, each containing a dedicated HvZB spectrum.

The comparison in Table 9.5 shows that no single ratio-based configuration outperforms the nominal pure GLP model for all signal strengths. For $\mu_{t\bar{t}+\mathrm{B}}$, the 2D-unrolled model achieves a slightly improved precision relative to the pure GLP score, while the HvB and HvZB configurations lead to marginally larger uncertainties. A clear benefit of the 2D-unrolled approach is observed for $\mu_{t\bar{t}+\mathrm{Z}}$, where it provides the strongest constraint among all tested models, indicating that the simultaneous exploitation of both ratio observables enhances the separation power for this particularly challenging process. For $\mu_{t\bar{t}+\mathrm{C}}$, the pure GLP model provides the best sensitivity, whereas both ratio-based configurations show a noticeable degradation. This is expected since neither of the models is designed to separate $t\bar{t}$+C from other event classes. A similar trend is observed for $\mu_{t\bar{t}+\mathrm{H}}$, where the targeted ratios do not improve the precision with respect to the nominal GLP output. In particular, the HvZB model exhibits a significant loss, probably due to the strong correlation between $t\bar{t}$+H and $t\bar{t}$Z(B) in this observable.

Overall, the results indicate that the pure GLP score remains the most balanced single discriminant, delivering either the best or near-optimal precision for the majority of signal strengths while avoiding the additional complexity introduced by ratio observables. Although the 2D-unrolled model demonstrates competitive performance and even surpasses the baseline for $t\bar{t}$+Z, its advantages are not uniform across processes and come at the cost of a more involved statistical model with stronger correlations between templates. The study therefore demonstrates that ratio observables are valuable as complementary inputs, but do not fully replace the information encoded in the original GLP classifier (Chapter 8). Although the 2018 dataset provides the largest statistics of Run-II, the present study only includes this single era. The performance of the 2D-unrolled model is expected to benefit from the full Run-II combination, where the increased event yield may better constrain the correlations introduced by the ratio observables. However, within the scope of this analysis and considering time constraints, the Pure GLP

model is adopted as the nominal choice, as it delivers the best or comparable precision for all signal strengths in 2018 while offering a simpler and more robust statistical interpretation.

9.3.1 Bias (and closure) test

A bias test is performed to validate the statistical model and fitting procedure used to extract the signal strength. The goal is to verify that the fit does not systematically overestimate or underestimate the signal in the presence of statistical fluctuations or modeling assumptions. This is achieved by generating pseudodata samples from a known model and refitting them with potentially different background parameterizations. Pseudodata samples ("toy data") are generated from a predefined background model with injected signal strength r_{true}. In this analysis, the nominal value $r_{true} = 1$ is used as a reference scenario, corresponding to the SM expectation. The generated pseudodata are subsequently fitted with an alternative background model to obtain the fitted signal strength r_{fit} and its associated uncertainty σ_{fit}. For each toy dataset, a pull value is defined as

$$\mathrm{P} = \frac{r_{true} - r_{fit}}{\sigma_{fit}} \tag{9.20}$$

By repeating this procedure for a large ensemble of toys, a pull distribution is constructed. In the absence of bias and with a well-estimated uncertainty, the pull distribution is expected to follow a Gaussian distribution centered at zero with unit width. In addition to the nominal scenario with the injected signal, bias studies are repeated for the case $r_{true} = 0$, corresponding to a background-only hypothesis. This test evaluates the potential tendency of the fit model to artificially create a signal in the absence of a true signal. Such studies are essential to ensure that the analysis does not produce spurious excesses due to background mismodeling. The absence of significant deviations from zero indicates that the fit procedure is unbiased and statistically well-behaved.

A total of $\mathrm{N} = 5000$ toy datasets are generated for each choice of background model. The bias is quantified as the mean of the pull distribution. Since the pull is defined relative to the fitted uncertainty, the bias is expressed in units of the total uncertainty on the signal strength. To assess whether the bias is acceptable, a predefined threshold is required. Typical values used in high-energy physics analyses range between 0.1 and 0.2, corresponding to a relative impact of approximately 1-2% on the total uncertainty when added in quadrature. A maximum value of 0.14 is set for the mean bias before performing the test, since a bias below this threshold would change the total uncertainty by less than 1%. When added in quadrature, this corresponds to $\sqrt{1^2 + 0.14^2} = 1.0098$, i.e a relative change of about 0.98%. The stability of the bias estimate is verified by increasing the number of toys, ensuring that statistical fluctuation in the pull mean is negligible.

Here, the bias studies are performed using the nominal statistical model based on the GNN discriminant templates. The pull distribution as portrayed in Fig. 9.4 exhibits a mean of 0.14, which lies within the predefined tolerance range of 0.1 and 0.2. The width of 0.97 suggests

slightly conservative uncertainty estimates. Overall, the distribution is consistent with the expectation of an unbiased fit with well-calibrated uncertainties.

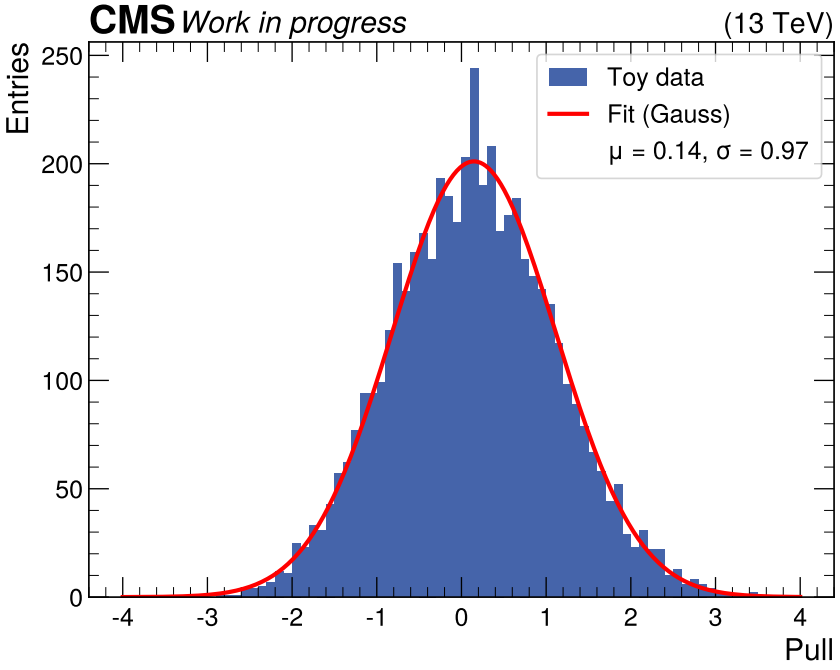


Figure 9.4: Distribution of the pull obtained from 5000 toy datasets. The blue histogram shows the pull distribution, while the red curve represents the fitted Gaussian function. The fit parameters, the mean μ and standard deviation σ, are indicated in the legend.

9.3.2 Goodness-of-Fit (GoF) test

To assess the compatibility between the data and the fitted statistical model, a goodness-of-fit (GoF) test is performed. Unlike closure and bias tests, which validate the behavior of the fit procedure, the GoF test explicitly evaluates the compatibility of the observed data with the fitted model prediction. It provides a quantitative measure of the agreement between the data and the post-fit model prediction. It serves as an important validation of the likelihood model used in the signal extraction. The GoF test is performed using the saturated model approach implemented in the *Combine* framework [174], where the test statistic is defined as the negative log-likelihood ratio between the nominal model and a saturated model in which the expected yields in each bin are allowed to independently match the observed data [175]. This test statistic can be interpreted as a generalization of the χ^2 test for likelihood-based fits to binned distributions.

To obtain the expected distribution of the test statistic, an ensemble of pseudo-experiments is generated. For each pseudo-dataset, the test statistic is computed after profiling the signal

strength parameters and NPs. The resulting distribution is compared with the observed value obtained from the data. The GoF p-value is defined as the probability of obtaining a test statistic equal to or larger than the observed value under the model hypothesis, assuming the fitted model is correct. Fig. 9.5 shows the distribution of the test statistic derived from 10000 pseudo-experiments, together with the observed value indicated by the vertical line. The resulting p-value is found to be $p = 0.109$, which indicates that the observed data are well described by the fitted model, with no evidence of significant mismodeling. This value is well above the commonly used reference level of 0.05. The observed test statistic is indicated by the vertical line in the figure. To ensure the robustness of the result, the test is repeated with an independent ensemble of pseudo-experiments, yielding consistent p-values within statistical fluctuations. The corresponding GoF tests performed separately for each Run-II data-taking period are provided in Appendix X, where consistent p-values are observed across all years.

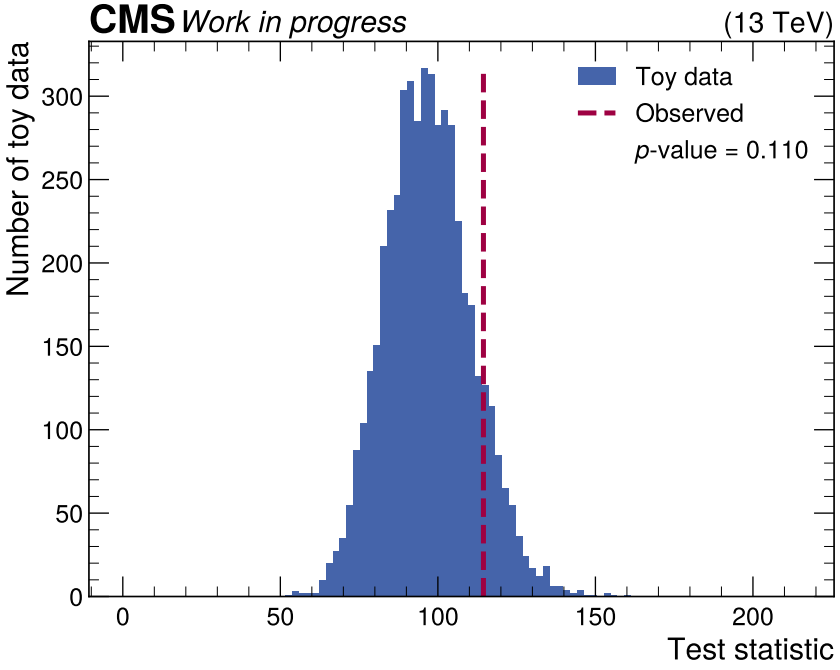


Figure 9.5: Distribution of the goodness-of-fit test statistic obtained from pseudo-experiments in the inclusive category of at least 5 jets. The histogram corresponds to 10000 toy datasets, while the vertical dashed red line indicates the observed test statistic obtained from data. While a p-value below 0.05 is often considered indicative of potential mismodeling, the observed value of $p = 0.109$ shows no such evidence.

9.3.3 Correlations

In addition to validating the fitted signal strengths, the NPs' correlation structure is examined using pseudodata. Fig. 9.6 shows the post-fit correlation matrix of the NPs and all POIs obtained from the pseudodata fit for the 2018 data-taking period. The corresponding correlation matrices for the 2017, 2016postVFP, and 2016preVFP eras are provided in Appendix I. The

correlations shown are derived from the post-fit covariance matrix of the likelihood. NPs associated with the limited size of simulated event samples are omitted from the visualization to improve readability. The matrix provides insight into the interplay between different sources of systematic uncertainty and their impact on the likelihood model. Overall, the correlation pattern is dominated by small off-diagonal elements, indicating that most NPs remain only weakly correlated after the fit. This behavior reflects the design of the statistical model, in which many experimental and theoretical uncertainties affect different processes or kinematic regions.

Moderate correlations are observed among groups of related uncertainties. In particular, the components of the jet energy scale (JES) uncertainties exhibit visible correlations, reflecting their common impact on jet kinematics and event categorization. Similarly, correlations are visible among flavor-tagging uncertainties, such as those associated with heavy-flavor, charm, and light-flavor tagging efficiencies, due to their shared dependence on jet flavor composition and tagging calibrations. Theory-driven uncertainties, such as QCD scale variations and parton distribution function (PDF) uncertainties, also show moderate correlations due to their simultaneous impact on multiple signals and background processes.

Furthermore, a noticeable correlation is present between the $t\bar{t}$+H signal strength parameter, $\mu_{t\bar{t}+H}$, and the uncertainty associated with gluon splitting to heavy flavor, with a weaker effect observed for $t\bar{t}$+Z. This behavior likely reflects the sensitivity of the analysis to additional heavy flavor production, which can alter jet flavor composition and partially resemble signal-like topologies. Importantly, no large or pathological correlations or pronounced degeneracies are observed among the four signal strength parameters and dominant NPs. This indicates that the simultaneous extraction of the POIs is not driven by large systematic degeneracies and that the likelihood model is well constrained.

Since the correlations are derived from a fit to pseudodata generated under the signal-plus-background hypothesis, the observed structure primarily reflects the intrinsic behavior of the statistical model rather than statistical fluctuations in data. The absence of pathological correlations therefore supports the robustness of the fit configuration and provides confidence in the stability of the signal extraction.

9.4 Fit results

In this section, the results of the simultaneous profile likelihood fit to the Run-II data-taking eras in the semileptonic channel are presented. The fit is performed using the full statistical model described in Sec. 9.2, including all experimental and theoretical systematic uncertainties outlined in Sec. 7.6. The extracted signal strengths of the four POIs and their corresponding significances are reported and interpreted. The best-fit signal strengths obtained from data are summarized in Table 9.6, with the corresponding significances reported in Table 9.7. The results are further illustrated in the Run-II summary plot presented in Fig. 9.7, which displays the

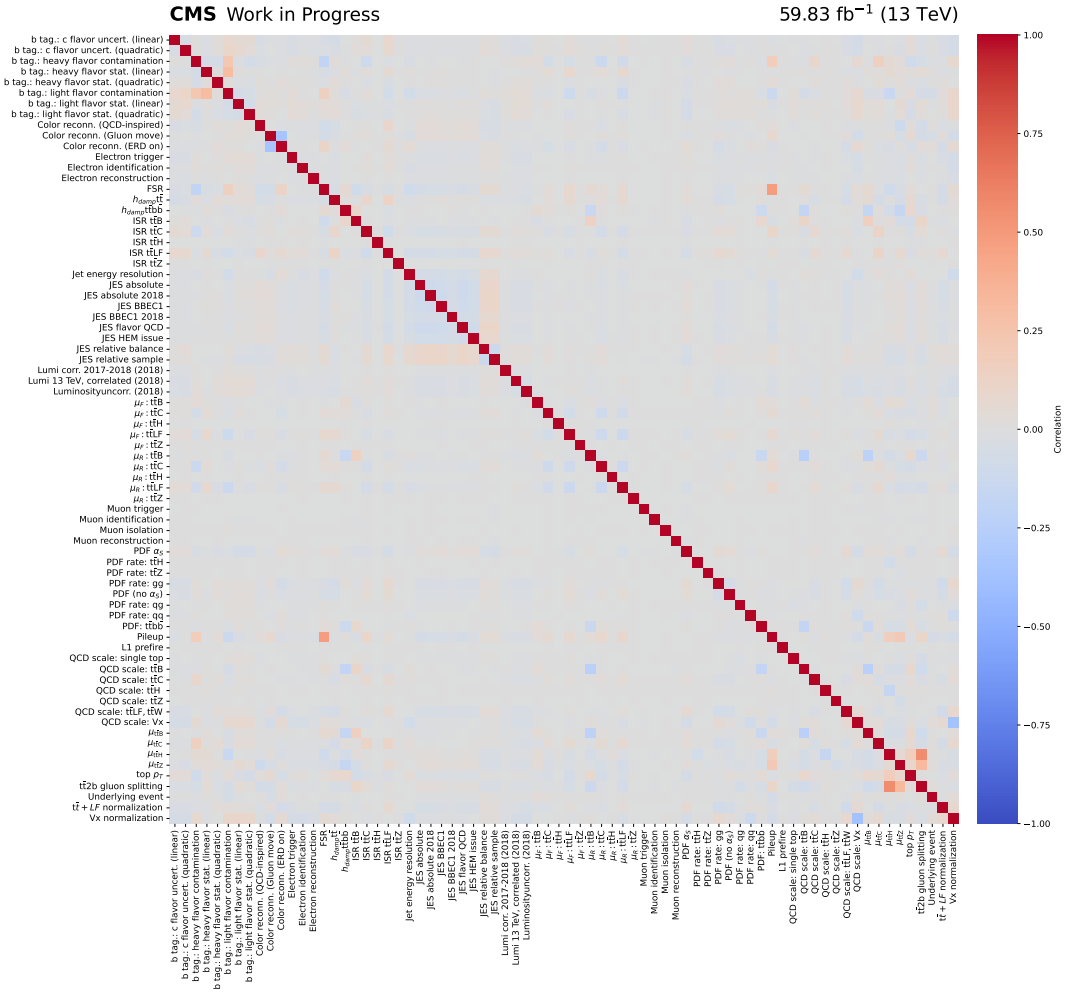


Figure 9.6: Correlation matrix of the profiled NPs and POIs obtained from the pseudodata fit of era 2018. The correlations are derived from the post-fit covariance matrix of the likelihood model.

best-fit signal strength parameters and their decomposed statistical and systematic uncertainties.

Among the four POIs, the $t\bar{t}$+B signal strength is measured with the highest precision and is found to be compatible with unity across the individual eras as well as in the combined Run-II result, indicating good agreement between data and the predicted heavy-flavor production rate. The $t\bar{t}$+C signal strength shows notable variations between eras, with the 2017 result closer to unity, while the remaining eras prefer smaller central values accompanied by larger uncertainties. The combined Run-II constraint shows a lower measured value with respect to the SM prediction. For the rarer EW processes $t\bar{t}$+H and $t\bar{t}$+Z, the fitted signal strengths are observed to exceed unity. While the combined Run-II results correspond to upward fluctuations relative to the SM prediction, they remain statistically compatible once the full uncertainty model and parameter correlations are taken into account.

The observed significances quantify the level of compatibility of the data with the background-only hypothesis for each process. The $t\bar{t}$+B process is firmly established with an observed (expected) significance exceeding 38σ (35σ), reflecting its comparatively large production rate. As explained in Sec. 9.3, such a large unusual significance is produced by a steep increase in the NLL as the signal strength approaches zero. A similar behavior was also observed in the dileptonic channel [93]. Evidence is obtained for $t\bar{t}$+H production with an observed (expected) significance of 4.8σ (3.1σ), while the $t\bar{t}$+Z process is observed with a significance of approximately 3σ (1.2σ). In contrast, the observed (expected) significance to $t\bar{t}$+C remains limited, with a significance of 1.7σ (5σ). This hierarchy reflects the differing production rates and background compositions of the processes, and is consistent with the uncertainty structure identified in the impact studies.

The fitted signal strengths are generally consistent with the expectations derived from the Asimov studies, presented in Sec. 9.3. Although the central values for $t\bar{t}$+H and $t\bar{t}$+Z exceed unity, the relatively large confidence intervals indicate that these deviations are statistically consistent with fluctuations rather than evidence of mismodeling. The $t\bar{t}$+B measurement closely reproduces the projected sensitivity, indicating that the statistical model performs as anticipated in a high-yield regime. The $t\bar{t}$+H and $t\bar{t}$+Z signal strengths are observed above their expected values, while $t\bar{t}$+C exhibits a downward fluctuation relative to the SM prediction. However, given the relatively large uncertainties, particularly for the rarer processes, these deviations remain statistically compatible with the expectations. Uncertainties in the signal strength parameters are dominated by different sources depending on the process. For $t\bar{t}$+B, the dominant contribution arises from the theoretical uncertainties, in particular the QCD scale variation, while parton shower and h_{damp} uncertainties provide subleading contributions. For $t\bar{t}$+C, experimental effects play a larger role, with the b-tagging and jet energy calibration uncertainties representing the largest contributions. The uncertainty on $t\bar{t}$+H is mainly driven by theoretical modeling uncertainties, particularly those associated with gluon splitting, the top-quark p_T reweighting, and QCD scale variations. In contrast, the measurement of $t\bar{t}$+Z

signal strength parameter is dominated by statistical uncertainty, with additional sizeable contributions from theoretical modeling uncertainties as PDF, QCD scale, h_{damp}, and correlations with other signal strength parameters. Furthermore, post-fit distributions, correlation matrices, and impact plots are presented to assess the quality and robustness of the fit.

Table 9.6: Best-fit signal strengths including all systematic uncertainties. Results are shown separately for each Run-II data-taking era in the inclusive ($\geq$ 5 jets) category. Uncertainties correspond to the profiled 68% confidence intervals.

	2018	2017	2016postVFP	2016preVFP	Run-II
$\mu^{obs}_{t\bar{t}+B}$	$1.12^{+0.48}_{-0.34}$	$1.13^{+0.43}_{-0.31}$	$1.46^{+0.71}_{-0.45}$	$1.25^{+0.61}_{-0.40}$	$1.09^{+0.39}_{-0.28}$
$\mu^{obs}_{t\bar{t}+C}$	$0.26^{+0.29}_{-0.32}$	$1.00^{+0.41}_{-0.39}$	$0.60^{+0.65}_{-0.83}$	$0.42^{+0.97}_{-1.28}$	$0.31^{+0.19}_{-0.18}$
$\mu^{obs}_{t\bar{t}+H}$	$1.56^{+0.81}_{-0.69}$	$2.85^{+1.01}_{-0.84}$	$0.82^{+1.11}_{-1.02}$	$2.00^{+1.11}_{-1.05}$	$1.99^{+0.65}_{-0.55}$
$\mu^{obs}_{t\bar{t}+Z}$	$3.16^{+1.73}_{-1.57}$	$3.49^{+1.84}_{-1.67}$	$0.89^{+2.47}_{-2.26}$	$4.82^{+2.69}_{-2.50}$	$2.90^{+1.18}_{-1.04}$

Table 9.7: Observed significances for each signal process obtained from data, including all systematic uncertainties.

	2018	2017	2016postVFP	2016preVFP	Run-II
$Z^{obs}_{t\bar{t}+B}$	25.57	23.07	15.98	14.73	38.55
$Z^{obs}_{t\bar{t}+C}$	0.82	2.85	0.71	0.44	1.68
$Z^{obs}_{t\bar{t}+H}$	2.24	4.04	0.80	1.99	4.76
$Z^{obs}_{t\bar{t}+Z}$	1.97	2.15	0.38	1.96	2.97

9.4.1 Post-fit distributions

To further validate the performance of the statistical model, post-fit distributions of the GLP output scores are examined for the primary discriminating observables in the signal-enriched categories. Fig. 9.8 shows representative post-fit distributions for the four signal classes $t\bar{t}$+H, $t\bar{t}$+Z, $t\bar{t}$+B, and $t\bar{t}$+C after the maximum likelihood fit, shown for the 2018 data-taking era. Corresponding distributions for the remaining Run-II eras are provided in Appendix J. Overall, good agreement between data and the post-fit model prediction is observed across all categories, indicating that the statistical model provides an adequate description of the data. Compared to the pre-fit distributions shown in Fig. 9.1, the post-fit model exhibits improved consistency with the data, reflecting the constraining power of the likelihood fit on both normalization and shape-related NPs. The total uncertainty is visibly reduced after the fit, reflecting the constraining power of the data on the NPs. The remaining bin-by-bin fluctuations are consistent with statistical variations and do not indicate any significant mismodeling.

9.4.2 Correlations among fit parameters

The correlations between the POIs and the NPs are examined to assess the interplay between the signal processes and the systematic uncertainties using the post-fit covariance matrix

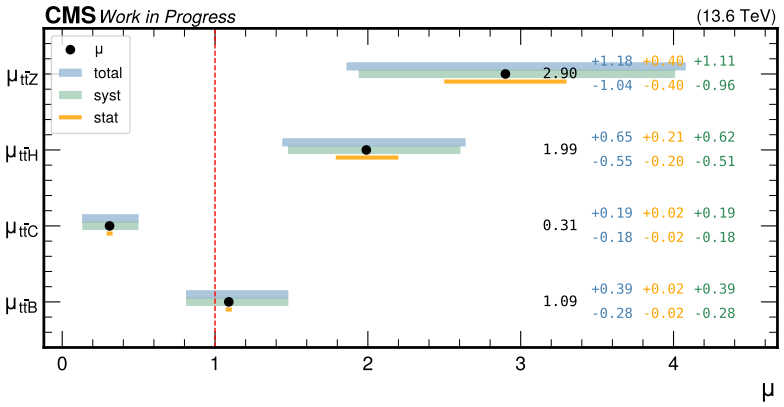


Figure 9.7: Summary of the best-fit signal strength parameters obtained from the combined Run-II fit using data. The markers indicate the best-fit values, while the horizontal bands represent the total, statistical, and systematic uncertainties, respectively. The figure highlights the process-dependent interplay between statistical and systematic effects in the measurement.

obtained from the fit to the 2018 data. Fig. 9.9 shows the corresponding correlation matrix for the 2018 data-taking period. For clarity, NPs associated with the finite size of simulated samples are not displayed. Additional correlation matrices for other eras are provided in Appendix K. The observed correlations reflect the expected degeneracies between signal processes and background normalizations, as well as the impact of flavor-tagging and modeling uncertainties. The matrix is, in general, characterized by predominantly small off-diagonal elements, indicating that most NPs remain only weakly correlated after the fit. This behavior suggests that the likelihood model is well constrained by the data and does not exhibit strong parameter degeneracies. The global correlation pattern closely resembles that observed in the pseudodata fit. Compared to the pseudodata fit, the correlations in data appear slightly more pronounced in localized regions of the matrix. This behavior is expected, as the fit to real data incorporates statistical fluctuations that can enhance the effective constraints on certain NPs and thereby modify their mutual correlations. In contrast, pseudodata are generated exactly according to the model hypothesis, such that their correlation structure primarily reflects the intrinsic shape of the likelihood function rather than statistical fluctuations in the data.

Moderate correlations are observed within groups of related systematic uncertainties. A noticeable correlation between the $\mu_{t\bar{t}+H}$ and the gluon-splitting uncertainty ($t\bar{t}2b$) remains visible in the data fit, with a weaker effect observed for $\mu_{t\bar{t}+Z}$. This pattern is consistent with the sensitivity of the analysis to additional heavy-flavor production and confirms that modeling uncertainties affecting heavy-flavor composition are properly accounted for in the simultaneous fit. A moderate correlation is also present between the ISR variation in $t\bar{t}$+B production and $\mu_{t\bar{t}+B}$. This reflects the sensitivity of the analysis to additional heavy-flavor radiation in $t\bar{t}$

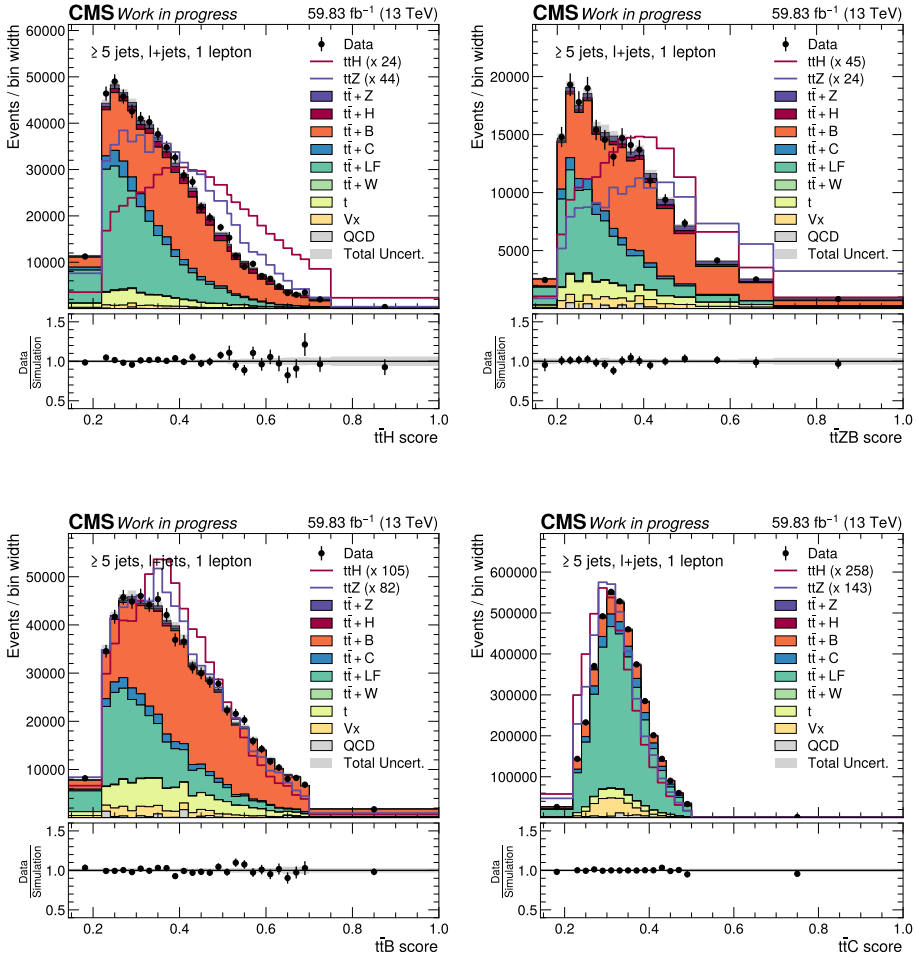


Figure 9.8: Post-fit GLP score distributions for the signal classes $t\bar{t}$+H, $t\bar{t}$+Z, $t\bar{t}$+B, and $t\bar{t}$+C in the 2018 data-taking era. The top row portrays $t\bar{t}$+H, $t\bar{t}$+Z, and $t\bar{t}$+B scores, while the bottom row portrays the $t\bar{t}$+C score. The post-fit prediction is represented by stacked histograms of all processes, and the observed data is represented by black markers. The $t\bar{t}$+H and $t\bar{t}$+Z predictions are additionally overlaid as line distributions normalized to the total post-fit yield to enhance their visibility. The gray band indicates the total post-fit uncertainty, including both statistical and systematic components described in Sec. 7.6. The lower panels show the ratio of data to the post-fit prediction. Corresponding distributions for the remaining Run-II eras are provided in Appendix J.

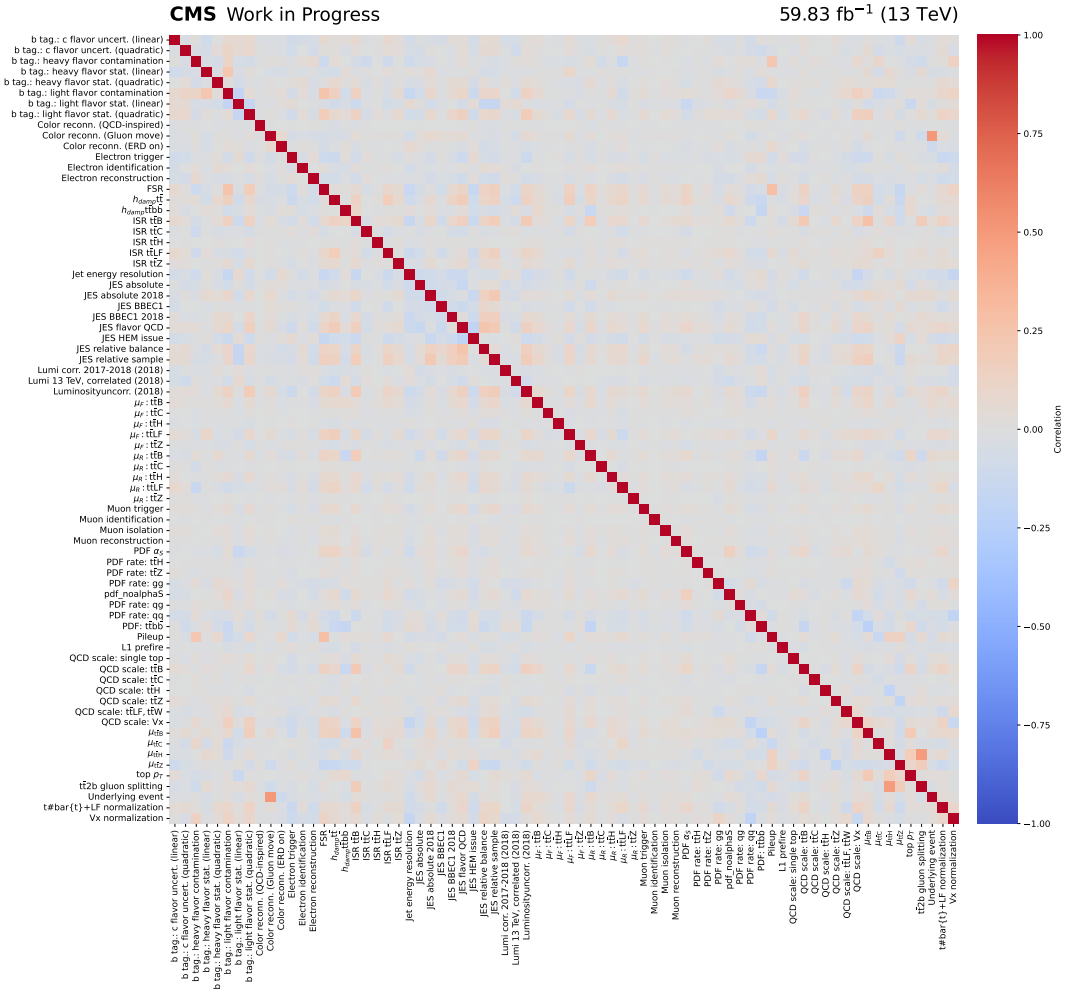


Figure 9.9: Correlation matrix of the profiled NPs and POIs obtained from the fit of era 2018. The correlations are derived from the post-fit covariance matrix of the likelihood model.

events. Variations in initial-state radiation can modify the jet multiplicity and flavor composition of the final state, partially affecting the same phase-space regions that constrain the $t\bar{t}$+B signal strength. A visible correlation is observed between the underlying event uncertainty and the color-reconnection variation corresponding to the gluon-move model. This is expected, as both uncertainties affect the modeling of soft QCD activity and hadronization effects, influencing the overall jet structure and event topology in a correlated manner. Additionally, small positive correlations of the top p_T modeling with $\mu_{t\bar{t}+B}$ and $\mu_{t\bar{t}+H}$ are visible. These effects are consistent with the sensitivity of the analysis categories to jet multiplicity and flavor composition, which are influenced both by top-quark kinematics and flavor-tagging performance.

Small anti-correlations are observed between the QCD scale variations of $t\bar{t}$+H and $t\bar{t}$+Z and their corresponding signal strength parameters. This behavior is expected from the structure of the likelihood. An upward variation in the theoretical cross-section can be partially compensated by a downward adjustment of the fitted signal strength, leading to a negative correlation between the scale uncertainty and the associated POI. Smaller anti-correlations are also visible among certain JES and JER components. These arise from the fact that different jet energy scale variations can have partially compensating effects on reconstructed jet kinematics and event categorization. A weak anti-correlation is observed between the top-quark p_T reweighting uncertainty and the statistical uncertainty associated with light-flavor b-tagging. Finally, mild anti-correlations involving the h_{damp} variation in $t\bar{t}$+$b\bar{b}$ production and the PDF uncertainty associated with $t\bar{t}$+$b\bar{b}$ are observed. Such patterns are typical for modeling uncertainties affecting similar kinematic regions, where different theoretical variations can partially compensate each other within the fit.

Importantly, no strong correlations or pronounced degeneracies are observed among the signal strength parameters or between the parameters of interest and dominant NPs. This indicates that the simultaneous extraction of the $t\bar{t}$+B, $t\bar{t}$+C, $t\bar{t}$+H, and $t\bar{t}$+Z signal strengths is not driven by large systematic degeneracies and that the fit remains statistically well conditioned. The similarity between the correlation patterns observed in data and those obtained from pseudodata further supports the robustness of the statistical model and indicates that the fit behaves as expected when confronted with data.

9.4.3 Impact of nuisance parameters (NPs)

A detailed understanding of the sources of uncertainty affecting the measurement is essential for assessing the robustness of the statistical model and identifying the dominant limitations of the analysis. In the profile likelihood framework, systematic uncertainties are incorporated through nuisance parameters (NPs), which encode imperfect knowledge of detector effects, theoretical modeling, and background normalizations (see Sec. 7.6).

The impact of an NP, θ, on a POI μ is defined as the variation $\Delta\mu$ induced when θ is fixed to its post-fit value shifted by $\pm\sigma$, while all remaining parameters are profiled according

to the likelihood maximization procedure. This definition follows the approach described in Ref. [176]. The resulting shift provides a direct measure of the correlation between the NP and the POI, thereby identifying which uncertainties contribute most strongly to the total uncertainty of the measurement.

Impact evaluation is performed by first performing an initial global fit to determine the best-fit values of the POIs and NPs, followed by a series of dedicated fits in which the NP is varied individually. Since multiple POIs are extracted simultaneously in this analysis, the remaining POIs are treated as floating parameters during the procedure to preserve the correlations encoded in the likelihood. The outputs of these fits are collected and summarized in a structured format that allows the ranking of NPs according to their effect (impact) on the measurement. The graphical representation of the impacts provides complementary information on both the behavior of the NPs and their influence on the POIs. The parameters are ordered according to the magnitude of their effect on the μ, ensuring that the dominant uncertainty sources appear at the top of the ranking.

The figure is divided into two panels. The left panel characterizes the response of each NP through the normalized shift between its pre-fit expectation and post-fit estimate. For an NP, θ, this quantity is defined as $(\hat{\theta}-\theta_1)/\sigma_1$, where $\hat{\theta}$ denotes the post-fit value, θ_1 the nominal (prefit) value, and σ_1 the associated pre-fit uncertainty. Expressing the displacement in units of the prior uncertainty provides a direct measure of the compatibility between the observed data and the initial model assumptions. Values close to zero indicate that the fit preserves the nominal parameter value, whereas larger deviations imply that the likelihood favors a shifted parameter to better describe the data. The markers correspond to the results obtained from the fit to data, whereas shaded bands represent the expectation derived from the Asimov dataset. The horizontal error bars reflect the post-fit uncertainties, allowing the constraining power of the data to be evaluated. NPs whose post-fit uncertainties are noticeably smaller than their pre-fit counterparts are effectively constrained by the fit, demonstrating that the measurement provides information beyond the prior assumptions. The right panel presents the impact of each NP on the signal strengths of each POI, $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$,. The impact is defined as the change in the fitted POI when a given NP is fixed to its post-fit value shifted by $\pm\sigma$, while all other parameters remain profiled. Solid lines indicate impacts obtained from data, and the shaded bands denote the expected variations. The separate visualization of upward and downward shifts enables asymmetric responses of the likelihood to be identified. The sign of the variation reveals whether the NP is correlated or anti-correlated with the signal strength, while the magnitude quantifies its contribution to overall uncertainty. Parameters producing the largest shifts, therefore, represent the principal systematic limitations of the measurement.

Taken together, the normalized parameter shifts and the corresponding impacts provide a comprehensive diagnostic of the statistical model. Large impacts accompanied by small shifts typically point to intrinsic sensitivity to a given uncertainty, whereas substantial displacements

may signal regions where the data prefer adjustments relative to the nominal prediction. This combined view is essential for validating the stability of the likelihood fit and for identifying the dominant effects governing the precision of the measurement. Because the NPs are determined from a single global likelihood fit, their post-fit values remain identical regardless of which POI is monitored. The impacts, however, differ for each signal strength parameter, reflecting the process-dependent sensitivity to individual systematic uncertainties. Figs. 9.10 and 9.11 show the impact of the dominant systematic uncertainties on the four POIs for the combined Run-II data. Each figure portrays the 15 most important parameters for expected and observed results. The full set of impact plots is presented in Appendix L.

The dominant uncertainties affecting the extraction of $\mu_{t\bar{t}+B}$ are driven by theoretical modeling of the $t\bar{t}$+B process. The highest impact arises from the QCD scale, indicating a strong dependence of the measurement on the perturbative QCD description of additional heavy-flavor jets. Subleading effects originate from the modeling of h_{damp} and the choice of parton distribution functions (PDFs) for $t\bar{t}$+B events. Normalization uncertainties of background processes, particularly vector boson production and $t\bar{t}$ + LF events, also contribute noticeably, reflecting residual degeneracies between signal and background yields that are only partially resolved by the likelihood fit. Additional modeling uncertainties, including initial-state radiation (ISR) in $t\bar{t}$+B events, the top-quark p_T spectrum, and final-state radiation (FSR), further affect the measurement, emphasizing the sensitivity to the description of extra jet activity. Experimental uncertainties related to flavor tagging also play a relevant role. In particular, the treatment of light flavor jets contamination and the calibration of charm-jet identification contribute to the total uncertainty, underscoring the importance of accurate jet-flavor discrimination in analyses targeting additional b-quark production. Pileup modeling contributes subdominantly, exhibiting post-fit shifts slightly larger than expected. However, their impacts on the signal strength remain moderate, and no coherent pattern indicative of a modeling deficiency is observed.

In contrast to $t\bar{t}$+B measurement, the uncertainty on $\mu_{t\bar{t}+C}$ is not dominated by a single theoretical source but instead arises from a broader interplay of experimental and modeling effects. The most influential impact originates from b-tagging uncertainties related to heavy-flavor contamination, highlighting the challenge of distinguishing c-jets from b-jets. Given the similarity of their tagging signatures, even small miscalibrations can propagate directly into the signal extraction. The modeling of h_{damp} in $t\bar{t}$ events, color reconnection (QCD-inspired) modeling, and the treatment of gluon splitting also rank among the leading systematic sources, emphasizing the dependence of the measurement on the accurate simulation of heavy-flavor production mechanisms. These effects modify the jet flavor composition and event topology, thereby affecting the separation between $t\bar{t}$+C signal events and backgrounds. Normalization uncertainties, particularly those associated with vector-boson production, introduce additional variations of order 0.1 in the signal strength. Several contributions from limited simulated event samples also appear among the dominant parameters, indicating that the measurement precision is partially constrained by the available MC statistics in specific regions of phase space.

The observed hierarchy of uncertainties reflects the intrinsic difficulty of isolating the $t\bar{t}$+C contribution, where the separation relies heavily on flavor-tagging performance and detailed modeling of heavy-flavor production. Consequently, further improvements in the measurement are expected to depend primarily on advances in c-tagging calibration and increased simulated event samples rather than solely on larger data statistics.

The extraction of the $\mu_{t\bar{t}+H}$ is primarily driven by theoretical modeling of heavy-flavor production. In particular, the treatment of gluon splitting in the $t\bar{t}2b$ process emerges as the leading systematic effect, indicating a strong dependence of the analysis on the accurate simulation of additional heavy-flavor jets. This behavior is expected, given the similarity between the $t\bar{t}$+H final state with $H \to b\bar{b}$ and background processes containing extra b-quarks. $\mu_{t\bar{t}+Z}$ parameter also contributes significantly, reflecting the similarity of these final states to $t\bar{t}$+H events. The top-quark transverse momentum modeling also introduces sizeable variations in the fitted signal strength, contributing shifts of the order of a few tenths. Further contributions arise from QCD scale variations in the $t\bar{t}$+H simulation and from flavor-tagging uncertainties. Experimental uncertainties related to flavor tagging also produce noticeable impacts. Light flavor tagging, including statistical precision and heavy-flavor contamination, are among those uncertainties, producing noticeable impacts, while c-tagging uncertainties contribute moderately. Additional theoretical uncertainties, such as h_{damp} in $t\bar{t}$+B events, renormalization-scale (μ_R) and factorization-scale (μ_F) variations in $t\bar{t}$ + LF, FSR, and PDF uncertainties in $t\bar{t}$+H, further affect the measurement. Pileup and (uncorrelated) 2017 luminosity uncertainty have smaller, yet non-negligible, effects. Overall, the impacts are distributed across multiple sources rather than being dominated by a single parameter. This pattern suggests that the measurement is constrained by both systematic effects and the intrinsically small production rate of the $t\bar{t}$+H process.

The uncertainty primarily affecting $\mu_{t\bar{t}+Z}$ is the modeling of heavy-flavor production, with gluon splitting to 2 b-quarks. This dependence is expected, as the presence of additional bottom quarks plays a crucial role in distinguishing the $t\bar{t}$+Z signal from background processes. A sizeable impact is observed from POI representing the $t\bar{t}$+H signal strength, indicating a non-negligible correlation between the $t\bar{t}$+H and $t\bar{t}$+Z processes. Contributions from limited simulated event samples and global event modeling further impact the measurement. Significant impacts also arise from the modeling of the top-quark transverse momentum, pileup, and from flavor-tagging uncertainties, including light-flavor misidentification and heavy-flavor contamination. Additional theoretical uncertainties originate from renormalization and factorization scale variations in $t\bar{t}$ + LF production and QCD scale variations in the $t\bar{t}$+Z simulation. Other subleading effects arise from the h_{damp} parameter in $t\bar{t}+b\bar{b}$ production, ISR modeling in $t\bar{t}$+B, and luminosity uncertainties for the 2017 dataset. The presence of several uncertainties with comparable impact suggests that the measurement is not limited by a single dominant effect but rather by the combined influence of multiple systematic sources together with the intrinsically small production rate. Overall, the observed hierarchy of uncertainties indicates that the precision

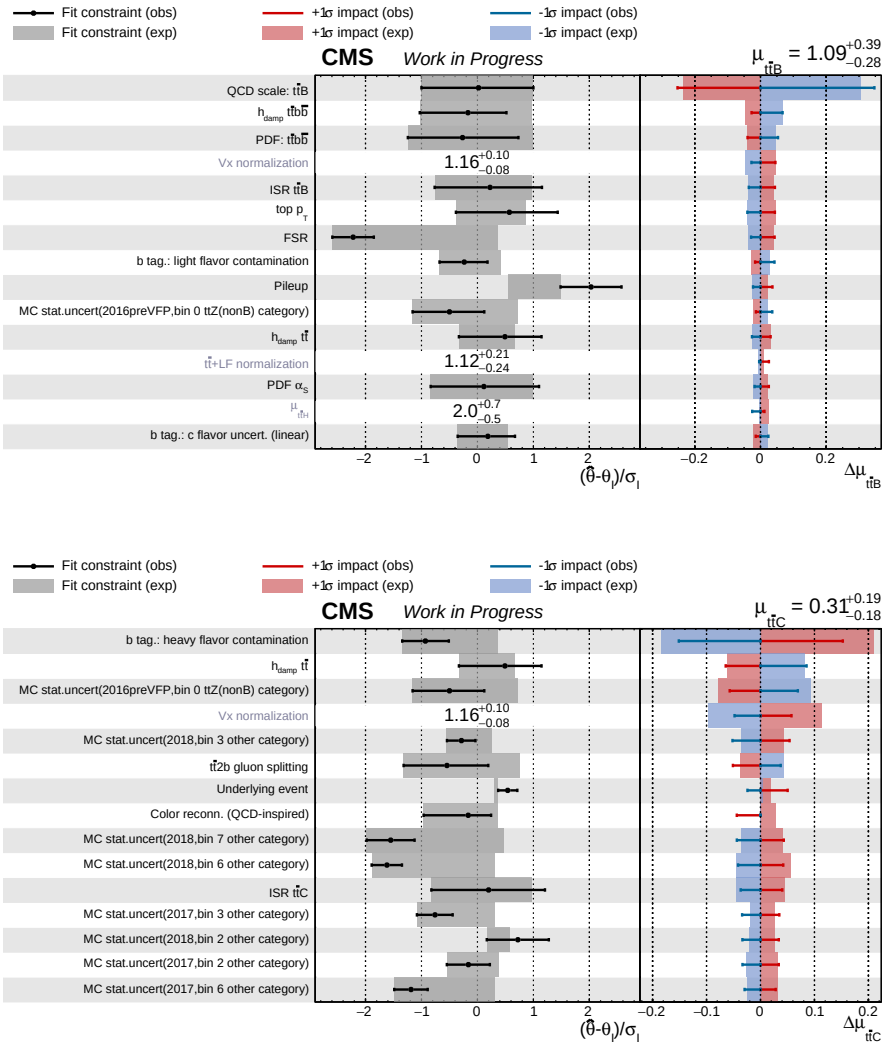


Figure 9.10: Impact plots of the dominant nuisance parameters (NPs) on the signal strength parameters, $\mu_{t\bar{t}+B}$ (top) and $\mu_{t\bar{t}+C}$ (bottom), obtained from the profile likelihood fit. The left-most panel lists the NPs. The middle panel shows the normalized shift of each NP relative to its pre-fit value, defined as $(\hat{\theta} - \theta_1)/\sigma_1$, where $\hat{\theta}$ is the post-fit value, θ_1 the pre-fit value, and σ_1 the corresponding pre-fit uncertainty. The right panel displays the corresponding variation induced in the POI when the NP is shifted by $\pm 1\sigma$. The markers and solid lines represent the observed results, whereas the shaded bands indicate the expectation from the Asimov dataset. The parameters (1-15) are ordered according to the magnitude of their impact.

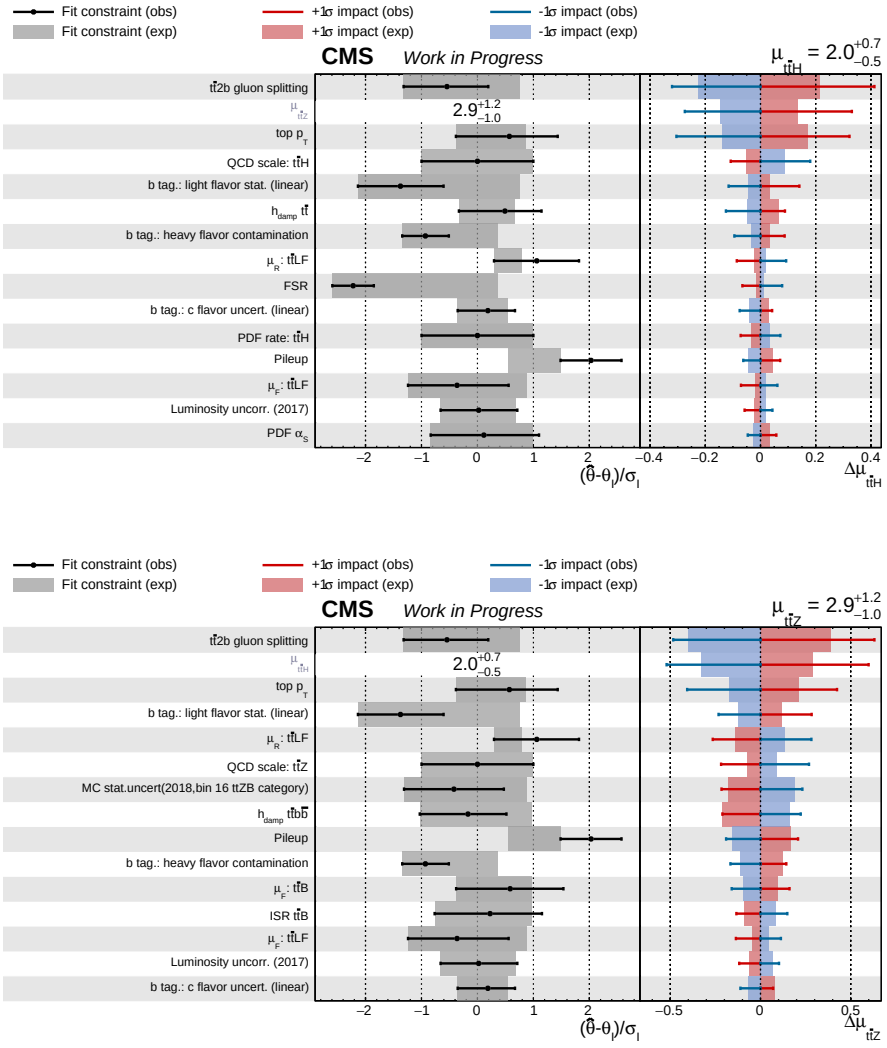


Figure 9.11: Impact plots of the dominant nuisance parameters (NPs) on the signal strength parameters, $\mu_{t\bar{t}+H}$ (top) and $\mu_{t\bar{t}+Z}$ (bottom), obtained from the profile likelihood fit. The left-most panel shows the names of the NPs. The middle panel shows the normalized shift of each NP relative to its pre-fit value, while the right panel displays the corresponding variation induced in the POI when the NP is shifted by $\pm 1\sigma$. The markers and solid lines represent the observed results, whereas the shaded bands indicate the expectation from the Asimov dataset. The parameters (1-15) are ordered according to the magnitude of their impact.

of the $t\bar{t}$+Z measurement is primarily constrained by the available statistical power, with systematic effects playing a secondary but non-negligible role.

A similar behavior for pileup modeling is observed across multiple signal strength parameters, which exhibit moderate post-fit displacements while maintaining a limited impact on the extracted signal strengths. This pattern further supports the stability of the global likelihood fit. The correction applied to simulated events to mitigate the known discrepancy between data and simulation in the pileup distribution is described in Sec. 7.4.3. Overall, the observed parameter displacements support the stability of the fit and provide no evidence of significant tension between simulation and data.

9.5 Comparison with related analyses

9.5.1 $t\bar{t}$+B signal strength

The measured Run-II signal strength for $t\bar{t}$+B production is,

$$\mu_{t\bar{t}+B}^{\text{Run-II}} = 1.09^{+0.39}_{-0.28}$$

which is statistically consistent with the SM expectation. The 68% confidence interval fully contains unity, and no significant deviation is observed. The individual years show central values around unity (see Table 9.6), with larger fluctuations in the lower-luminosity 2016 eras. The combined Run-II result shows a smaller uncertainty than any individual year due to the larger effective dataset and the simultaneous treatment of NPs across all eras. This result is compatible with the measurements of $t\bar{t}$+$b\bar{b}$ production in the semileptonic channel by the CMS collaboration [85, 99]. In the $\geq$ 5j, $\geq$ 4b phase space, the ratio of measured to predicted cross-section was found to be close to unity [85]. The present inclusive normalization agrees with this observation. In contrast to that dedicated fiducial measurement, the current analysis extracts $t\bar{t}$+B simultaneously with $t\bar{t}$+C, $t\bar{t}$+H, and $t\bar{t}$+Z contributions in a common $\geq$ 5j phase space. Furthermore, the result is compatible with the post-fit $t\bar{t}$+B normalization reported in the CMS $t\bar{t} + H(b\bar{b})$ Run-II analysis, where a value of $\mu_{t\bar{t}+B} = 1.19^{+0.13}_{-0.12}$ was obtained. The difference between central values corresponds to less than 0.3σ of the uncertainty of the present measurement, indicating no tension between the analyses.
For the 2018 dataset, a simultaneous $t\bar{t}$+X analysis performed in the dileptonic channel [93] obtained $\mu_{t\bar{t}+B}^{2018} = 0.98^{+0.34}_{-0.25}$. The corresponding result of the present semileptonic analysis, $\mu_{t\bar{t}+B}^{2018} = 1.12^{+0.48}_{-0.34}$ is compatible within uncertainties. The difference between the central values corresponds to less than 0.6σ, indicating no statistically significant discrepancy.

9.5.2 $t\bar{t}$+C signal strength

The combined Run-II result for $t\bar{t}$+C production is

$$\mu_{t\bar{t}+C}^{\text{Run-II}} = 0.31^{+0.19}_{-0.18}$$

which lies approximately 3-4 standard deviations below unity. The year-by-year results exhibit sizeable fluctuations, particularly in the 2016 eras, reflecting the limited discriminating power between charm and light-flavor jets and the strong correlation with $t\bar{t}$+B. The Run-II combination significantly reduces the uncertainty due to the larger integrated luminosity and the simultaneous profiling of NPs. The individual 2017 and 2016 era results are compatible with the SM expectation within uncertainties, whereas the 2018 dataset drives the lower central value in the Run-II combination (see Table 9.6). The combined Run-II value obtained in this analysis shows a lower measured cross-section compared to the prediction. The observed significance of 1.68σ, as shown in the Table 9.7, indicates that the data does not provide strong evidence for $t\bar{t}$+C production in this phase space. Although the expected sensitivity is approximately 5σ under the SM assumption, the reduced observed significance reflects a downward statistical fluctuation.

Measurements of $t\bar{t} + c\bar{c}$ are experimentally challenging, and dedicated measurements as explained in Sec. 6.4.2 typically report measured cross sections exceeding the nominal simulation prediction in specific fiducial phase spaces. A direct comparison with the present result is not straightforward, as those analyses rely on different definitions and modeling schemes. However, for the 2018 dataset, the simultaneous $t\bar{t}$+X fit in the dileptonic final state reports $\mu^{2018}_{t\bar{t}+C} = 0.74^{+0.41}_{-0.41}$. The corresponding result obtained in this analysis is $\mu^{2018}_{t\bar{t}+C} = 0.26^{+0.29}_{-0.32}$. Although both analyses yield central values below unity, the interpretation differs. In the present semileptonic analysis, the best-fit value is significantly below one, indicating that the data prefer a reduced $t\bar{t}$+C contribution relative to the nominal simulation. Since both measurements favor a reduced $t\bar{t}$+C normalization in 2018, it suggests that the effect is not specific to a particular final-state topology.

A signal strength parameter below unity for the $t\bar{t}$+C process can arise from the limited discriminating power between c- and LF-jets in the analysis phase space. There is confusion in discriminating between b- and c-jets as well. In the GLP classifier (Sec. 8.5), challenges in the $t\bar{t}$+C event classification were observed, which may have been caused by the lack of c-jet tagging input information. This can lead to a non-negligible overlap between $t\bar{t}$+C and $t\bar{t} + LF$ event topologies. As a consequence, misclassification of $t\bar{t}$+C events as $t\bar{t} + LF$ may occur at an enhanced rate relative to processes containing additional b-jets. In the analysis of at least 5 jet phase space, the $t\bar{t} + LF$ contribution constitutes the dominant background component. A mild overestimation of events with additional light-flavor jet radiation is observed, consistent with trends reported in previous heavy-flavor measurements. In a simultaneous multi-parameter fit, such effects can induce anti-correlated shifts between $t\bar{t}$+C and $t\bar{t} + LF$ contributions, resulting in a reduced fitted normalization for $t\bar{t}$+C.

9.5.3 $t\bar{t}$+H signal strength

The combined Run-II result for $t\bar{t}$+H production obtained in this analysis is

$$\mu^{\text{Run-II}}_{t\bar{t}+H} = 1.99^{+0.65}_{-0.55}$$

The central value lies above unity, corresponding to an enhancement of approximately 1.6σ relative to the SM prediction. For the 2018 dataset, the semileptonic result $\mu^{2018}_{t\bar{t}+H} = 1.56^{+0.81}_{-0.69}$ is statistically compatible within uncertainties. The corresponding dileptonic measurement, which reported $\mu^{2018}_{t\bar{t}+H} = 0.89^{+0.95}_{-0.93}$ also shows compatible results with this analysis. The difference between the two central values corresponds to less than one standard deviation. Inclusive Run-II measurements of $t\bar{t} + H(b\bar{b})$ production by the CMS and ATLAS Collaborations report signal strengths closer to unity, as summarized in Table 6.1, except for the CMS measurement including all decay channels. In particular, the most precise ATLAS measurement yields $\mu^{\text{Run-II}}_{t\bar{t}+H} = 0.81^{+0.19}_{-0.22}$, while the CMS combination reports $\mu_{t\bar{t}+H} = 0.33 \pm 0.26$. The present result is higher in central value and is statistically compatible within approximately 2σ when uncertainties are taken into account. The differences can be attributed to the distinct analysis strategies and phase-space definitions. These earlier measurements combine multiple final states and employ dedicated $t\bar{t} + H(b\bar{b})$ event categorization, whereas the present analysis is restricted to the semileptonic channel and performs a simultaneous extraction. The stronger correlations between $t\bar{t}$+H and $t\bar{t}$+Z in the high jet-multiplicity semileptonic topology can amplify statistical fluctuations in the fitted normalization.

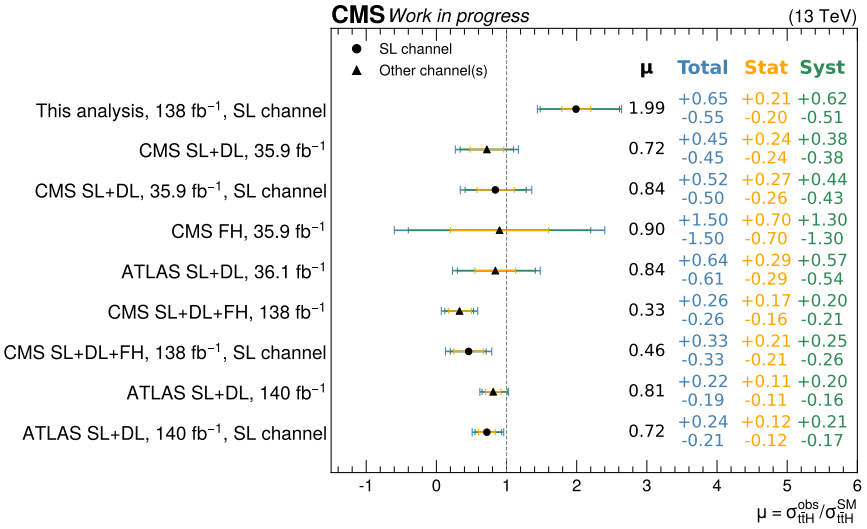


Figure 9.12: Summary of (selected) ATLAS and CMS measurements of $t\bar{t} + H(b\bar{b})$ production at $\sqrt{s} = 13$ TeV, including early Run-II and full Run-II results. The measurements include CMS analyses: SL+DL 35.9fb^{-1} [110], FH 35.9fb^{-1} [111], SL+DL+FH 138 fb^{-1} [113] and ATLAS analyses: SL+DL 36.1 fb^{-1} [112], SL+DL 140 fb^{-1} (re-analysis) [114]. The abbreviations DL, SL, and FH refer to dileptonic, semileptonic, and fully hadronic decay channels of $t\bar{t}$, respectively.

9.5.4 $t\bar{t}$+Z signal strength

The combined Run-II result for $t\bar{t}$+Z production is

$$\mu^{\text{Run-II}}_{t\bar{t}+Z} = 2.90^{+1.18}_{-1.04}$$

The central value lies above unity, corresponding to an enhancement of approximately 1.8σ relative to the SM prediction. Inclusive $t\bar{t}$+Z cross-section measurements performed by the CMS collaboration in the leptonic decay channel of the Z boson report a measured cross-section of approximately 13% above the predicted value, corresponding to $\mu \sim 1.1$. While the present analysis focuses on $Z \to b\bar{b}$ decay mode in a high multiplicity phase space and therefore is not directly comparable to inclusive measurements, both results exhibit central values above the SM expectation. A study of $t\bar{t}$+Z production performed in a thesis (Ref. [118]) using 2018 data in the semileptonic channel reported $\mu^{2018}_{t\bar{t}+Z} = -0.77^{+1.19}_{-1.26}$, with large statistical uncertainty. This result relies only on 2018 data and also does not include all the systematic uncertainties. The present 2018 result, $\mu^{2018}_{t\bar{t}+Z} = 3.16^{+1.73}_{-1.57}$ differs in central value. The corresponding result obtained in the dileptonic channel in the $t\bar{t}$+X analysis is $1.28^{+1.06}_{-1.06}$. While the semileptonic channel yields a larger central value, both measurements are characterized by sizeable uncertainties. In the recent CMS analysis in Ref. [105], the $t\bar{t} + Z(b\bar{b})$ signal strength was measured within measurements performed primarily to constrain $t\bar{t}$+H, and the value is $\mu^{\text{Run-II}}_{t\bar{t}+Z} = 1.47^{+0.45}_{-0.41}$. This value corresponds to a combined fit of fully hadronic, semileptonic, and dileptonic channels, and is compatible within uncertainties. The difference in central values can be attributed to statistical fluctuations and to the different phase-space sensitivities and fit correlations of these different final states, in particular the stronger overlap between $t\bar{t} + Z(b\bar{b})$ and $t\bar{t} + H(b\bar{b})$ topologies in this high jet-multiplicity semileptonic channel.

10 Summary and outlook

Understanding the fundamental constituents of matter and their interactions is a central goal of particle physics. At the energy frontier, this is pursued through high-energy proton-proton collisions provided by the Large Hadron Collider (LHC) at the Organisation européenne pour la recherche nucléaire (CERN). Among the experiments operating at the LHC, the Compact Muon Solenoid (CMS) experiment is designed to study a broad range of phenomena predicted by the Standard Model (SM) and to search for possible signs of new physics.

In this context, processes involving the production of top-quark pairs ($t\bar{t}$) in association with additional heavy-flavor jets provide an important testing ground for the SM. In particular, final states containing heavy-flavor jets offer a sensitive probe of quantum chromodynamics (QCD) in complex multi-jet environments and play a crucial role in improving the modeling of backgrounds in Higgs boson measurements and searches for physics beyond the SM. This thesis presented a simultaneous measurement of top quark-antiquark production in association with heavy flavor jets in the semileptonic decay channel using the full LHC Run-II dataset recorded by the CMS experiment at a center-of-mass energy of 13 TeV, corresponding to an integrated luminosity of 138 fb^{-1}.

In this analysis, several processes contributing to final states with multiple heavy-flavor jets, collectively referred to as $t\bar{t}$+X processes, are considered simultaneously. These include $t\bar{t}$ production with additional bottom quarks ($t\bar{t}$+$b\bar{b}$) and charm quarks ($t\bar{t} + c\bar{c}$), as well as associated production with a Higgs boson with decay $H \to b\bar{b}$ ($t\bar{t} + H(b\bar{b})$) or a Z boson with $Z \to b\bar{b}$ ($t\bar{t} + Z(b\bar{b})$). The $t\bar{t}$+H production provides direct access to the Yukawa coupling between the top quark and the Higgs boson. Measurements of this process, therefore, play a key role in testing the mechanism of electroweak symmetry breaking. In addition, previous analyses using Run-II data reported a signal strength for $t\bar{t}$+H production that was below the SM expectation, motivating this study with improved analysis strategies and a more precise treatment of heavy-flavor backgrounds. The $t\bar{t}$+Z process further probes the electroweak interaction of

the top quark and provides sensitivity to the top-Z coupling. While most measurements of this process rely on leptonic decays of the Z boson, the hadronic decay channel $Z \to b\bar{b}$ offers complementary sensitivity in a challenging multi-jet final state. These processes lead to final states containing multiple b-jets that closely resemble those arising from $t\bar{t}$+$b\bar{b}$ production, which constitutes one of the dominant and most challenging backgrounds. Furthermore, the $t\bar{t} + c\bar{c}$ process exhibits similar kinematic properties, and due to the limited efficiency and non-negligible mistag probability of heavy-flavor tagging, migration between these processes can occur, making their separation experimentally demanding.

To address these challenges posed by the large combinatorial background and the similarity of kinematic properties between different jet sources, modern analyses increasingly employ machine learning techniques to exploit correlations among reconstructed objects in collider events. In this analysis, a strategy based on Graph Neural Networks (GNNs) is developed. GNNs provide a natural framework for representing collider events as graphs, where physics objects such as jets, leptons, and missing transverse momentum are treated as nodes and their relationships are encoded through edges. This approach allows the model to exploit individual properties of objects and their relational information. The graph representation is processed through a series of graph neural network blocks that perform message passing between connected objects. These blocks incorporate multi-head attention mechanisms, allowing the network to learn complex correlations between objects and identify the most relevant interactions within the event topology. The architecture is designed to perform a simultaneous measurement of multiple $t\bar{t}$+X processes using a two-stage classification strategy. In the first stage, a node-level prediction (NLP) model performs a binary node-classification task, assigning each jet a probability of belonging to the additional heavy-flavor category rather than from top-quark decays. The resulting node-level information is then propagated to a graph-level prediction (GLP) model, which performs event-level classification of the full event into the different $t\bar{t}$+X categories. This hierarchical design allows the model to combine local object-level information with global event topology, improving the separation of signal and background processes. The impact of the NLP stage on the overall classification performance is explicitly evaluated in this analysis. Including the additional jet identification as an intermediate step improves the event-level discrimination compared to a model without NLP information, with gains of about 16% for $t\bar{t}$+H, about 40% for $t\bar{t}$+C, as well as substantial improvement for $t\bar{t}$+Z events. This demonstrates the importance of incorporating physically motivated object-level information into the event-level classifier.

Using the outputs of the GLP classifier, a simultaneous binned maximum likelihood fit is performed to extract the signal strengths of the relevant processes. In this framework, the production cross sections of the processes of interest are parametrized in terms of signal-strength parameters defined as $\mu = \sigma_{\text{obs}}/\sigma_{\text{SM}}$, where σ_{SM} denotes the SM prediction. Four signal strength parameters, $\mu^{obs}_{t\bar{t}+B}$, $\mu^{obs}_{t\bar{t}+C}$, $\mu^{obs}_{t\bar{t}+H}$, and $\mu^{obs}_{t\bar{t}+Z}$, are simultaneously extracted in a combined fit. This approach ensures a consistent treatment of all processes and their associated uncertainties. Systematic effects affecting both the normalization and shape of the distributions

are incorporated in the fit through nuisance parameters (NPs). Prior to applying the fit to collision data, the statistical model is extensively validated using pseudodata. The combined LHC Run-II results obtained in this analysis for the signal strengths in heavy-flavor phase space are,

$$\mu^{obs}_{t\bar{t}+B} = 1.09^{+0.39}_{-0.28}\,(\text{tot}) = 1.09^{+0.02}_{-0.02}\,(\text{stat})^{+0.39}_{-0.28}\,(\text{syst}), \tag{10.1}$$

$$\mu^{obs}_{t\bar{t}+C} = 0.31^{+0.19}_{-0.18}\,(\text{tot}) = 0.31^{+0.02}_{-0.02}\,(\text{stat})^{+0.19}_{-0.18}\,(\text{syst}), \tag{10.2}$$

$$\mu^{obs}_{t\bar{t}+H} = 1.99^{+0.65}_{-0.55}\,(\text{tot}) = 1.99^{+0.21}_{-0.20}\,(\text{stat})^{+0.62}_{-0.51}\,(\text{syst}), \tag{10.3}$$

$$\mu^{obs}_{t\bar{t}+Z} = 2.90^{+1.18}_{-1.04}\,(\text{tot}) = 2.90^{+0.40}_{-0.40}\,(\text{stat})^{+1.11}_{-0.96}\,(\text{syst}). \tag{10.4}$$

The results show that the measured value of $\mu_{t\bar{t}+B}$ is consistent with the SM prediction within the experimental uncertainties. The extracted signal strength for $t\bar{t}$+C is found to be significantly below unity, indicating that the observed rate of events attributed to its production is lower than predicted by the current simulation within the analyzed phase space. In contrast, the measurements of $t\bar{t}$+H and $t\bar{t}$+Z production are observed to be above the SM expectation. In particular, the measured value of $\mu_{t\bar{t}+H}$ lies above the central values of most previous measurements. Nevertheless, the uncertainties remain sufficiently large such that the result is compatible with the SM prediction and earlier measurements within approximately two standard deviations. The dominant sources of uncertainty differ across the measured processes. The $t\bar{t}$+B and $t\bar{t}$+H measurements are mainly limited by theoretical modeling uncertainties, the $t\bar{t}$+C result is more sensitive to experimental effects such as heavy-flavor tagging, and the $t\bar{t}$+Z is primarily limited by statistical uncertainty. The analysis also highlights the importance of accurate modeling of heavy-flavor production in $t\bar{t}$ events, as theoretical modeling and heavy-flavor tagging uncertainties constitute important limitations in precise measurements of $t\bar{t}$ production with heavy-flavor jets.

Looking ahead, several avenues exist for improving the precision of such measurements. Continued developments in machine learning techniques may further enhance the ability to exploit correlations between reconstructed objects and improve the separation between processes with similar topologies. Improvements in heavy-flavor tagging algorithms, especially for charm-jet identification, will play a crucial role in reducing systematic uncertainties and enhancing the sensitivity to $t\bar{t}$+C production, which remains one of the least precisely measured components of $t\bar{t}$+X processes. Improved modeling of heavy-flavor production in $t\bar{t}$ events will also be essential for reducing theoretical uncertainties and achieving more precise measurements. Further progress can be achieved by refining the categorization of heavy-flavor processes. For example, separating the $t\bar{t}$+B contribution into its underlying components such as $t\bar{t}+b\bar{b}$, $t\bar{t}$ +2b, and $t\bar{t}+b$ would allow a more detailed understanding of the different underlying production mechanisms. A similar approach could be applied to $t\bar{t}+c\bar{c}$ production, providing improved sensitivity to the modeling of charm-quark radiation and gluon splitting processes in perturbative QCD. In addition, including processes such as $t\bar{t}$+H with the decay H $\rightarrow$cc may benefit from the improved tagging algorithms and may provide further constraints on the charm Yukawa coupling.

These larger datasets will also provide stringent tests of perturbative QCD calculations and MC modeling. For the $t\bar{t}$+Z measurement, the sensitivity is partly limited by the size of the available simulated event samples. Increasing the size of these samples, particularly for QCD background processes, and developing data-driven estimates of the QCD background could further improve the precision of the measurement. Finally, combining measurements from different $t\bar{t}$ decay channels, such as the dileptonic and semileptonic final states, may further improve the overall precision and provide complementary sensitivity to the different $t\bar{t}$+X processes. These could enable more precise measurements, leading to a deeper understanding of SM and potentially providing hints of physics beyond the SM.

Bibliography

[1] The ATLAS collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020.

[2] The CMS collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 30–61. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2012.08.021. URL: https://www.sciencedirect.com/science/article/pii/S0370269312008581.

[3] CDF collaboration. "Observation of Top Quark Production in $\overline{p}p$ Collisions with the Collider Detector at Fermilab". In: *Phys. Rev. Lett.* 74 (14 Apr. 1995), pp. 2626–2631. DOI: 10.1103/PhysRevLett.74.2626. URL: https://link.aps.org/doi/10.1103/PhysRevLett.74.2626.

[4] DØ Collaboration. "Observation of the Top Quark". In: *Phys. Rev. Lett.* 74 (14 Apr. 1995), pp. 2632–2637. DOI: 10.1103/PhysRevLett.74.2632. URL: https://link.aps.org/doi/10.1103/PhysRevLett.74.2632.

[5] CERN. *The European Organization for Nuclear Research*. Accessed: 2026-03-10. 2024. URL: https://home.cern.

[6] CMS. "Measurements of the inclusive W and Z boson production cross sections and their ratios in proton-proton collisions at $\sqrt{s} = 13.6$ TeV". In: *Journal of High Energy Physics* 2026.1 (Jan. 2026). ISSN: 1029-8479. DOI: 10.1007/jhep01(2026)047. URL: http://dx.doi.org/10.1007/JHEP01(2026)047.

[7] OpenAI. *ChatGPT*. https://chat.openai.com. Large language model used for grammar checking, text summarisation, and coding assistance. Accessed: 2026-03-10. 2026.

[8] GitHub, Inc. *GitHub Copilot*. https://github.com/features/copilot. AI-based code completion tool. Accessed: 2026-03-10. 2026.

[9] Grammarly Inc. *Grammarly*. https://www.grammarly.com. AI-based grammar and writing assistance tool. Accessed: 2026-03-10. 2026.

[10] S. F. Novaes. "Standard Model: An Introduction". In: (2000). arXiv: hep-ph/0001283 [hep-ph]. URL: https://arxiv.org/abs/hep-ph/0001283.

[11] M. Lubej. "Study of $B \to KK\ell\nu_\ell$ decays at Belle". PhD thesis. 2018. URL: https://repozitorij.uni-lj.si/IzpisGradiva.php?lang=slv&id=105432.

[12] F. Halzen and A. D. Martin. *Quarks And Leptons: An Introdutory Course in Modern Particle Physics*. 1984. ISBN: 978-0-471-88741-6.

[13] M. Neubauer. *Higgs Boson Physics*. Accessed: 05 January 2026. 2025. URL: https://msneubauer.github.io/projects/4_project/.

[14] ATLAS, CDF, CMS and D0 collaborations. *First combination of Tevatron and LHC measurements of the top-quark mass*. 2014. arXiv: 1403.4427 [hep-ex].

[15] R. W. et al. (Particle Data Group). "Review of Particle Physics". In: *Prog. Theor. Exp. Phys.* 2024 (2024), p. 083C01. DOI: 10.1093/ptep/ptae053.

[16] CMS Collaboration. "A portrait of the Higgs boson by the CMS experiment ten years after the discovery". In: *Nature* 607 (July 2022), pp. 1–9. DOI: 10.1038/s41586-022-04892-x.

[17] S. Holmes, R. S. Moore, and V. Shiltsev. "Overview of the Tevatron collider complex: goals, operations and performance". In: *Journal of Instrumentation* 6.08 (Aug. 2011), T08001–T08001. ISSN: 1748-0221. DOI: 10.1088/1748-0221/6/08/t08001. URL: http://dx.doi.org/10.1088/1748-0221/6/08/T08001.

[18] CMS. "Search for new physics using effective field theory in 13 TeV collision events that contain a top quark pair and a boosted Z or Higgs boson". In: *Physical Review D* 108.3 (Aug. 2023). ISSN: 2470-0029. DOI: 10.1103/physrevd.108.032008. URL: http://dx.doi.org/10.1103/PhysRevD.108.032008.

[19] LHCTopWG. *LHCtopWG Joined Experimental and Theory LHC Working Group*. https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCTopWG, Accessed on 2025-06-23.

[20] V. M. Abazov, B. Abbott, and et. al. "Evidence for Production of Single Top Quarks and First Direct Measurement of $|V_{tb}|$". In: *Phys. Rev. Lett.* 98 (18 May 2007), p. 181802. DOI: 10.1103/PhysRevLett.98.181802. URL: https://link.aps.org/doi/10.1103/PhysRevLett.98.181802.

[21] O. S. Brüning et al. *LHC Design Report*. CERN Yellow Reports: Monographs. Geneva: CERN, 2004. DOI: 10.5170/CERN-2004-003-V-1. URL: https://cds.cern.ch/record/782076.

[22] The ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08003–S08003. DOI: 10.1088/1748-0221/3/08/s08003. URL: https://doi.org/10.1088/1748-0221/3/08/s08003.

[23] G. Bayatian. "CMS Physics Technical Design Report, Volume II: Physics Performance". In: *J.Phys.* G34 (June 2007), pp. 995–1579. DOI: 10.1088/0954-3899/34/6/S01.

[24] K. Aamodt et al. "The ALICE experiment at the CERN LHC". In: *JINST* 3 (2008), S08002. DOI: 10.1088/1748-0221/3/08/S08002.

[25] A. A. Alves Jr. et al. "The LHCb Detector at the LHC". In: *JINST* 3 (2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005.

[26] G. Anelli et al. "The TOTEM experiment at the CERN large hadron collider". In: *Journal of Instrumentation* 3 (Aug. 2008). DOI: 10.1088/1748-0221/3/08/S08007.

[27] LHCf Collaboration. "The LHCf detector at the CERN Large Hadron Collider". In: *Journal of Instrumentation* 3 (Aug. 2008), S08006. DOI: 10.1088/1748-0221/3/08/S08006.

[28] M. Staelens. *Recent Results and Future Plans of the MoEDAL Experiment*. 2019. arXiv: 1910.05772 [hep-ex].

[29] FASER Collaboration. *FASER: ForwArd Search ExpeRiment at the LHC*. 2019. arXiv: 1901.04468 [hep-ex].

[30] E. Lopienska. "The CERN accelerator complex, layout in 2022". In: (2022). General Photo. URL: https://cds.cern.ch/record/2800984.

[31] *CMS experiment. CERN Grey Book*. CERN. URL: https://greybook.cern.ch/experiment/detail?id=CMS (visited on 09/28/2025).

[32] T. Sakuma. "Cutaway diagrams of CMS detector". In: (2019). URL: https://cds.cern.ch/record/2665537.

[33] J. H. Fichet. "CMS beam pipe installation". In: (Apr. 2021). General Photo. URL: https://cds.cern.ch/record/2764445.

[34] V. Karimäki. "The CMS tracker system project: Technical Design Report". In: (1997). Ed. by M. Mannelli et al.

[35] W. Adam et al. "The CMS Phase-1 Pixel Detector Upgrade". In: *JINST* 16 (2021), P02027. DOI: 10.1088/1748-0221/16/02/P02027. arXiv: 2012.14304 [physics.ins-det].

[36] S. Chatrchyan et al. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". In: *JINST* 9 (2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009. arXiv: 1405.6569 [physics.ins-det].

[37] *Track impact parameter resolution for the full pseudo rapidity coverage in the 2017 dataset with the CMS Phase-1 Pixel detector*. CMS Detector Performance Summary CMS-DP-2020-049. 2020. URL: https://cds.cern.ch/record/2743740.

[38] K. Z. Gumus. "Search for new physics in the Compact Muon Solenoid (CMS) experiment and the response of the CMS calorimeters to particles and jets". PhD thesis. Lubbock, Texas Tech U., 2008. DOI: 10.2172/936638.

[39] *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical design report. CMS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/349375.

[40] P. Bloch et al. *Changes to CMS ECAL electronics*. Technical design report. CMS. Geneva: CERN, 2002. URL: https://cds.cern.ch/record/581342.

[41] CMS Collaboration. *Technical proposal for the upgrade of the CMS detector through 2020*. Tech. rep. 2011. URL: https://cds.cern.ch/record/1355706.

[42] The CMS collaboration. "Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC". In: *Journal of Instrumentation* 16.05 (May 2021), P05014. DOI: 10.1088/1748-0221/16/05/P05014. URL: https://dx.doi.org/10.1088/1748-0221/16/05/P05014.

[43] CMS Collaboration. *ECAL 2016 refined calibration and Run2 summary plots*. CMS Detector Performance Summary CMS-DP-2020-021. 2020. URL: https://cds.cern.ch/record/2717925.

[44] *The CMS hadron calorimeter project: Technical Design Report*. Technical design report. CMS. The following files are from http://uscms.fnal.gov/pub/hcal-tdr and may not be the version as printed, please check the printed version to be sure. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/357153.

[45] J. Mans et al. *CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter*. Tech. rep. Additional contact persons: Jeffrey Spalding, Fermilab, spalding@cern.ch, Didier Contardo, Universite Claude Bernard-Lyon I, contardo@cern.ch. 2012. URL: https://cds.cern.ch/record/1481837.

[46] V. e. Khachatryan. "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV". In: *Journal of Instrumentation* 12.02 (Feb. 2017), P02014. DOI: 10.1088/1748-0221/12/02/P02014. URL: https://dx.doi.org/10.1088/1748-0221/12/02/P02014.

[47] The CMS Collaboration. "Precise mapping of the magnetic field in the CMS barrel yoke using cosmic rays". In: *Journal of Instrumentation* 5.03 (Mar. 2010), T03021. DOI: 10.1088/1748-0221/5/03/T03021.

[48] I. Seth. "Phase I Upgrade of the CMS Hadron Calorimeter". In: *Nuclear and Particle Physics Proceedings* 273-275 (Apr. 2016), pp. 1002–1007. DOI: 10.1016/j.nuclphysbps.2015.09.157.

[49] J. G. Layter. *The CMS muon project: Technical Design Report*. Technical design report. CMS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/343814.

[50] The CMS Collaboration. *The CMS muon project: Technical Design Report*. Technical design report. CMS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/343814.

[51] The CMS Collaboration. "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s}$=13 TeV". In: *Journal of Instrumentation* 13.06 (June 2018), P06015. DOI: 10.1088/1748-0221/13/06/P06015. URL: https://dx.doi.org/10.1088/1748-0221/13/06/P06015.

[52] F. Bechtel. "The underlying event in proton-proton collisions". PhD thesis. Hamburg U., 2009. DOI: 10.3204/DESY-THESIS-2009-015.

[53] F. Sauli. "The gas electron multiplier (GEM): Operating principles and applications". In: *Nucl. Instrum. Meth. A* 805 (2016), pp. 2–24. DOI: 10.1016/j.nima.2015.07.060.

[54] The CMS Collaboration. "The CMS trigger system". In: *Journal of Instrumentation* 12.01 (Jan. 2017), P01020–P01020. ISSN: 1748-0221. DOI: 10.1088/1748-0221/12/01/p01020. URL: http://dx.doi.org/10.1088/1748-0221/12/01/P01020.

[55] A. M. Sirunyan et al. "Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$". In: *JINST* 15 (2020), P10017. DOI: 10.1088/1748-0221/15/10/P10017. arXiv: 2006.10165 [hep-ex].

[56] A. Hayrapetyan et al. "Performance of the CMS high-level trigger during LHC Run 2". In: *JINST* 19 (2024), P11021. DOI: 10.1088/1748-0221/19/11/P11021. arXiv: 2410.17038 [physics.ins-det].

[57] The CMS Collaboration. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". In: *Journal of Instrumentation* 9.10 (Oct. 2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009. URL: https://dx.doi.org/10.1088/1748-0221/9/10/P10009.

[58] P. Billoir. "Progressive track recognition with a Kalman-like fitting procedure". In: *Computer Physics Communications* 57.1 (1989), pp. 390–394. ISSN: 0010-4655. DOI: https://doi.org/10.1016/0010-4655(89)90249-X. URL: https://www.sciencedirect.com/science/article/pii/001046558990249X.

[59] W. Waltenberger, R. Frühwirth, and P. Vanlaer. "Adaptive vertex fitting". In: 34.12 (Nov. 2007), N343. DOI: 10.1088/0954-3899/34/12/N01. URL: https://dx.doi.org/10.1088/0954-3899/34/12/N01.

[60] CMS. "Particle-flow reconstruction and global event description with the CMS detector". In: *Journal of Instrumentation* 12.10 (Oct. 2017), P10003. DOI: 10.1088/1748-0221/12/10/P10003. URL: https://dx.doi.org/10.1088/1748-0221/12/10/P10003.

[61] CMS. "Particle-flow reconstruction and global event description with the CMS detector". In: *Journal of Instrumentation* 12.10 (Oct. 2017), P10003. DOI: 10.1088/1748-0221/12/10/P10003. URL: https://dx.doi.org/10.1088/1748-0221/12/10/P10003.

[62] M. Cacciari, G. P. Salam, and G. Soyez. "The anti-kt jet clustering algorithm". In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. URL: https://dx.doi.org/10.1088/1126-6708/2008/04/063.

[63] CMS. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". In: *Journal of Instrumentation* 13.05 (May 2018), P05011. DOI: 10.1088/1748-0221/13/05/P05011. URL: https://dx.doi.org/10.1088/1748-0221/13/05/P05011.

[64] CMS. "Inclusive and differential cross section measurements of $\mathrm{t\bar{t}b\bar{b}}$ production in the lepton+jets channel at $\sqrt{s} = 13$ TeV". In: *Journal of High Energy Physics* 2024.5 (May 2024). ISSN: 1029-8479. DOI: 10.1007/jhep05(2024)042. URL: http://dx.doi.org/10.1007/JHEP05(2024)042.

[65] E. Bols et al. "Jet flavour classification using DeepJet". In: *Journal of Instrumentation* 15.12 (Dec. 2020), P12012. DOI: 10.1088/1748-0221/15/12/P12012. URL: https://dx.doi.org/10.1088/1748-0221/15/12/P12012.

[66] "Performance summary of AK4 jet b tagging with data from proton-proton collisions at 13 TeV with the CMS detector". In: (2023). URL: https://cds.cern.ch/record/2854609.

[67] V. N. Gribov and L. N. Lipatov. "Deep inelastic e p scattering in perturbation theory". In: *Sov. J. Nucl. Phys.* 15 (1972), pp. 438–450.

[68] G. Altarelli and G. Parisi. "Asymptotic Freedom in Parton Language". In: *Nucl. Phys. B* 126 (1977), pp. 298–318. DOI: 10.1016/0550-3213(77)90384-4.

[69] H. D. Politzer. "Asymptotic Freedom: An Approach to Strong Interactions". In: *Phys. Rept.* 14 (1974), pp. 129–180. DOI: 10.1016/0370-1573(74)90014-3.

[70] Y. L. Dokshitzer. "Calculation of the Structure Functions for Deep Inelastic Scattering and e^+e^- Annihilation by Perturbation Theory in Quantum Chromodynamics". In: *Sov. Phys. JETP* 46 (1977). Translated from Zh. Eksp. Teor. Fiz. 73 (1977) 1216–1240, pp. 641–653.

[71] R. D. e. Ball. "Parton distributions from high-precision collider data". In: *The European Physical Journal C* 77.10 (Oct. 2017), p. 663. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-017-5199-5. URL: https://doi.org/10.1140/epjc/s10052-017-5199-5.

[72] J. e. Alwall. "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 (July 2014). ISSN: 1029-8479. DOI: 10.1007/jhep07(2014)079. URL: http://dx.doi.org/10.1007/JHEP07(2014)079.

[73] S. 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 (June 2010). ISSN: 1029-8479. DOI: 10.1007/jhep06(2010)043. URL: http://dx.doi.org/10.1007/JHEP06(2010)043.

[74] S. Frixione, P. Nason, and C. Oleari. "Matching NLO QCD computations with parton shower simulations: the POWHEG method". In: *Journal of High Energy Physics* 2007.11 (Nov. 2007), pp. 070–070. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2007/11/070. URL: http://dx.doi.org/10.1088/1126-6708/2007/11/070.

[75] J. Alwall et al. "MadGraph 5: going beyond". In: *Journal of High Energy Physics* 2011.6 (June 2011). ISSN: 1029-8479. DOI: 10.1007/jhep06(2011)128. URL: http://dx.doi.org/10.1007/JHEP06(2011)128.

[76] S. Frixione and B. R. Webber. "Matching NLO QCD computations and parton shower simulations". In: *Journal of High Energy Physics* 2002.06 (June 2002), pp. 029–029. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2002/06/029. URL: http://dx.doi.org/10.1088/1126-6708/2002/06/029.

[77] P. 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: 1029-8479. DOI: 10.1007/jhep03(2013)015. URL: http://dx.doi.org/10.1007/JHEP03(2013)015.

[78] T. Sjöstrand, S. Ask, and et. al. "An introduction to PYTHIA 8.2". In: *Computer Physics Communications* 191 (2015), pp. 159–177. ISSN: 0010-4655. DOI: https://doi.org/10.1016/j.cpc.2015.01.024. URL: https://www.sciencedirect.com/science/article/pii/S0010465515000442.

[79] B. Andersson et al. "Parton fragmentation and string dynamics". In: *Physics Reports* 97.2 (1983), pp. 31–145. ISSN: 0370-1573. DOI: https://doi.org/10.1016/0370-1573(83)90080-7. URL: https://www.sciencedirect.com/science/article/pii/0370157383900807.

[80] CMS. "Extraction and validation of a new set of CMS pythia8 tunes from underlying-event measurements". In: *The European Physical Journal C* 80.1 (Jan. 2020). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-019-7499-4. URL: http://dx.doi.org/10.1140/epjc/s10052-019-7499-4.

[81] S. Agostinelli et al. "GEANT4 - A Simulation Toolkit". In: *Nucl. Instrum. Meth. A* 506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8.

[82] CMS Collaboration. *Object definitions for top quark analyses at the particle level*. Tech. rep. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2267573.

[83] M. Cacciari, G. P. Salam, and G. Soyez. "The catchment area of jets". In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), p. 005. DOI: 10.1088/1126-6708/2008/04/005. URL: https://dx.doi.org/10.1088/1126-6708/2008/04/005.

[84] D. de Florian et al. "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector". In: *CERN Yellow Rep. Monogr.* 2 (2017), pp. 1–869. DOI: 10.23731/CYRM-2017-002. arXiv: 1610.07922 [hep-ph].

[85] CMS. "Inclusive and differential cross section measurements of $\mathrm{t\bar{t}b\bar{b}}$ production in the lepton+jets channel at $\sqrt{s} = 13$ TeV". In: *Journal of High Energy Physics* 2024.5 (May 2024). ISSN: 1029-8479. DOI: 10.1007/jhep05(2024)042. URL: http://dx.doi.org/10.1007/JHEP05(2024)042.

[86] ATLAS. "Measurements of fiducial cross-sections for $t\bar{t}$ production with one or two additional b-jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector". In: *The European Physical Journal C* 76.1 (Jan. 2016). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-015-3852-4. URL: http://dx.doi.org/10.1140/epjc/s10052-015-3852-4.

[87] ATLAS. "Measurement of Higgs boson decay into b-quarks in associated production with a top-quark pair in pp collisions at $\sqrt{s}$= 13 TeV with the ATLAS detector". In: *Journal of High Energy Physics* (2022). ISSN: 1029-8479. DOI: 10.1007/JHEP06(2022)097. URL: https://doi.org/10.1007/JHEP06(2022)097.

[88] In: *Physics Reports* 1115 (Apr. 2025), pp. 3–115. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2024.11.005. URL: http://dx.doi.org/10.1016/j.physrep.2024.11.005.

[89] A. Sirunyan et al. "Measurement of the $t\bar{t}b\bar{b}$ production cross section in the all-jet final state in pp collisions at $\sqrt{s} = 13$ TeV". In: *Physics Letters B* 803 (Apr. 2020), p. 135285. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2020.135285. URL: http://dx.doi.org/10.1016/j.physletb.2020.135285.

[90] A. Kulesza et al. "Associated production of a top quark pair with a heavy electroweak gauge boson at NLO+NNLL accuracy". In: *The European Physical Journal C* 79.3 (Mar. 2019). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-019-6746-z. URL: http://dx.doi.org/10.1140/epjc/s10052-019-6746-z.

[91] L. Ferencz et al. *Study of ttbb and ttW background modelling for ttH analyses*. 2023. arXiv: 2301.11670 [hep-ex]. URL: https://arxiv.org/abs/2301.11670.

[92] T. Ježo et al. "New NLOPS predictions for $t\bar{t} + b$-jet production at the LHC". In: *The European Physical Journal C* 78.6 (June 2018). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-018-5956-0. URL: http://dx.doi.org/10.1140/epjc/s10052-018-5956-0.

[93] E. L. Pfeffer. *Simultaneous cross section measurements of top quark-antiquark pair production with additional heavy flavor jets at the CMS experiment* /. Karlsruhe, [2024]. URL: https://doi.org/10.5445/IR/1000178528.

[94] CMS. "Measurements of $t\bar{t}$ cross sections in association with b jets and inclusive jets and their ratio using dilepton final states in pp collisions at $\sqrt{s} = 13$ TeV". In: *Physics Letters B* 776 (Jan. 2018), pp. 355–378. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2017.11.043. URL: http://dx.doi.org/10.1016/j.physletb.2017.11.043.

[95] CMS. "Measurement of the cross section for $t\bar{t}$ production with additional jets and b jets in pp collisions at $\sqrt{s} = 13$ TeV". In: *Journal of High Energy Physics* 2020.7 (July 2020). ISSN: 1029-8479. DOI: 10.1007/jhep07(2020)125. URL: http://dx.doi.org/10.1007/JHEP07(2020)125.

[96] "Measurement of the cross section ratio ttbb/ttjj in pp Collisions at 8 TeV". In: *Physics Letters B* 746 (2015), pp. 132–153. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2015.04.060. URL: https://www.sciencedirect.com/science/article/pii/S0370269315003263.

[97] CMS. "Measurement of $t\bar{t}$ production with additional jet activity, including $b\bar{b}$ quark jets, in the dilepton decay channel using pp collisions at $\sqrt{s} = 8$ TeV". In: *The European Physical Journal C* 76.7 (July 2016). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-016-4105-x. URL: http://dx.doi.org/10.1140/epjc/s10052-016-4105-x.

[98] ATLAS. "Measurements of fiducial cross-sections for $t\bar{t}$ production with one or two additional b-jets in pp collisions at $\sqrt{s}$=8 TeV using the ATLAS detector". In: *The European Physical Journal C* 76.1 (Jan. 2016). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-015-3852-4. URL: http://dx.doi.org/10.1140/epjc/s10052-015-3852-4.

[99] J. Van Der Linden. "Inclusive and differential cross section measurement of $t\bar{t}b\bar{b}$ production and studies of $t\bar{t}$ production with additional jet radiation". PhD thesis. Karlsruher Institut für Technologie (KIT), 2023. 221 pp. DOI: 10.5445/IR/1000160616.

[100] ATLAS. "Measurement of $t\bar{t}$ production in association with additional b-jets in the e final state in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Journal of High Energy Physics* 2025.1 (Jan. 2025). ISSN: 1029-8479. DOI: 10.1007/jhep01(2025)068. URL: http://dx.doi.org/10.1007/JHEP01(2025)068.

[101] A. e. Sirunyan. "First measurement of the cross section for top quark pair production with additional charm jets using dileptonic final states in pp collisions at $\sqrt{s} = 13\,TeV$". In: *Physics Letters B* 820 (Sept. 2021), p. 136565. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2021.136565. URL: http://dx.doi.org/10.1016/j.physletb.2021.136565.

[102] ATLAS Collaboration. "Measurement of top-quark pair production in association with charm quarks in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Physics Letters B* 860 (Jan. 2025), p. 139177. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2024.139177. URL: http://dx.doi.org/10.1016/j.physletb.2024.139177.

[103] M. Bähr et al. "Herwig++ physics and manual". In: *The European Physical Journal C* 58.4 (Nov. 2008), pp. 639–707. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-008-0798-9. URL: http://dx.doi.org/10.1140/epjc/s10052-008-0798-9.

[104] J. Bellm et al. *Herwig 7.1 Release Note*. 2017. arXiv: 1705.06919 [hep-ph]. URL: https://arxiv.org/abs/1705.06919.

[105] CMS. "Simultaneous Probe of the Charm and Bottom Quark Yukawa Couplings Using $t\bar{t}$+H Events". In: *Physical Review Letters* 136.1 (Jan. 2026). ISSN: 1079-7114. DOI: 10.1103/9nwb-splk. URL: http://dx.doi.org/10.1103/9nwb-splk.

[106] T. Ježo and P. Nason. "On the treatment of resonances in next-to-leading order calculations matched to a parton shower". In: *Journal of High Energy Physics* 2015.12 (Dec. 2015), pp. 1–47. ISSN: 1029-8479. DOI: 10.1007/jhep12(2015)065. URL: http://dx.doi.org/10.1007/JHEP12(2015)065.

[107] ATLAS. "Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in *pp* p-p collisions at $\sqrt{\mathbf{s}} = 8\ TeV$ with the ATLAS detector". In: *The European Physical Journal C* 75.7 (July 2015). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-015-3543-1. URL: http://dx.doi.org/10.1140/epjc/s10052-015-3543-1.

[108] CMS. "Search for the associated production of the Higgs boson with a top-quark pair". In: *Journal of High Energy Physics* 2014.9 (Sept. 2014). ISSN: 1029-8479. DOI: 10.1007/jhep09(2014)087. URL: http://dx.doi.org/10.1007/JHEP09(2014)087.

[109] CMS. "Search for a standard model Higgs boson produced in association with a top-quark pair and decaying to bottom quarks using a matrix element method". In: *The European Physical Journal C* 75.6 (June 2015). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-015-3454-1. URL: http://dx.doi.org/10.1140/epjc/s10052-015-3454-1.

[110] CMS. "Search for $t\bar{t}H$ production in the $H \to b\bar{b}$ decay channel with leptonic $t\bar{t}$ decays in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Journal of High Energy Physics* 2019.3 (Mar. 2019). ISSN: 1029-8479. DOI: 10.1007/jhep03(2019)026. URL: http://dx.doi.org/10.1007/JHEP03(2019)026.

[111] CMS. "Search for $t\bar{t}H$ production in the all-jet final state in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Journal of High Energy Physics* 2018.6 (June 2018). ISSN: 1029-8479. DOI: 10.1007/jhep06(2018)101. URL: http://dx.doi.org/10.1007/JHEP06(2018)101.

[112] ATLAS. "Search for the standard model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Physical Review D* 97.7 (Apr. 2018). ISSN: 2470-0029. DOI: 10.1103/physrevd.97.072016. URL: http://dx.doi.org/10.1103/PhysRevD.97.072016.

[113] CMS Collaboration. "Measurement of the $t\bar{t}H$ and tH production rates in the $H \to b\bar{b}$ decay channel using proton-proton collision data at $\sqrt{s} = 13$ TeV". In: *Journal of High Energy Physics* 2025.2 (Feb. 2025). ISSN: 1029-8479. DOI: 10.1007/jhep02(2025)097. URL: http://dx.doi.org/10.1007/JHEP02(2025)097.

[114] ATLAS. "Measurement of the associated production of a top-antitop-quark pair and a Higgs boson decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC". In: *The European Physical Journal C* 85.2 (Feb. 2025). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-025-13740-x. URL: http://dx.doi.org/10.1140/epjc/s10052-025-13740-x.

[115] CMS. "Measurement of top quark pair production in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Journal of High Energy Physics* 2020.3 (Mar. 2020). ISSN: 1029-8479. DOI: 10.1007/jhep03(2020)056. URL: http://dx.doi.org/10.1007/JHEP03(2020)056.

[116] CMS. "Measurements of inclusive and differential cross sections for top quark production in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Journal of High Energy Physics* 2025.2 (Feb. 2025). ISSN: 1029-8479. DOI: 10.1007/jhep02(2025)177. URL: http://dx.doi.org/10.1007/JHEP02(2025)177.

[117] ATLAS Collaboration. "Inclusive and differential cross-section measurements of $t\bar{t}Z$ production in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, including EFT and spin-correlation interpretations". In: *Journal of High Energy Physics* 2024.7 (July 2024). ISSN: 1029-8479. DOI: 10.1007/jhep07(2024)163. URL: http://dx.doi.org/10.1007/JHEP07(2024)163.

[118] J. Van Der Linden. *Limit on $t\bar{t} + Z$ production in the $Z \to b\bar{b}$ channel at the CMS experiment, Master's thesis*. May 2019. URL: https://cds.cern.ch/record/2703587.

[119] M. French et al. "Design and results from the APV25, a deep sub-micron CMOS front-end chip for the CMS tracker". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 466.2 (2001). 4th Int. Symp. on Development and Application of Semiconductor Tracking Detectors, pp. 359–365. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(01)00589-7. URL: https://www.sciencedirect.com/science/article/pii/S0168900201005897.

[120] "Simulation of the Silicon Strip Tracker pre-amplifier in early 2016 data". In: (2020). URL: https://cds.cern.ch/record/2740688.

[121] F. Buccioni et al. "OpenLoops 2". In: *The European Physical Journal C* 79.10 (Oct. 2019). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-019-7306-2. URL: http://dx.doi.org/10.1140/epjc/s10052-019-7306-2.

[122] R. Frühwirth. "Track fitting with non-Gaussian noise". In: *Computer Physics Communications* 100.1 (1997), pp. 1–16. ISSN: 0010-4655. DOI: https://doi.org/10.1016/S0010-4655(96)00155-5. URL: https://www.sciencedirect.com/science/article/pii/S0010465596001555.

[123] W. Adam et al. *Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC*. Tech. rep. Geneva: CERN, 2005. URL: https://cds.cern.ch/record/815410.

[124] M. Cacciari, G. P. Salam, and G. Soyez. "FastJet user manual: (for version 3.0.2)". In: *The European Physical Journal C* 72.3 (Mar. 2012). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph]. URL: http://dx.doi.org/10.1140/epjc/s10052-012-1896-2.

[125] CMS. *Jet algorithms performance in 13 TeV data*. Tech. rep. https://cds.cern.ch/record/2256875. Geneva: CERN, 2017.

[126] D. Bertolini et al. "Pileup per particle identification". In: *Journal of High Energy Physics* 2014.10 (Oct. 2014). ISSN: 1029-8479. DOI: 10.1007/jhep10(2014)059. URL: http://dx.doi.org/10.1007/JHEP10(2014)059.

[127] CMS. *CMS Internal TWiki: MET Filter Recommendations for Run II*. https://twiki.cern.ch/twiki/bin/viewauth/CMS/MissingETOptionalFiltersRun2, Accessed on 2025-05-14.

[128] CMS twiki. *Reweighting recipe to emulate Level 1 ECAL and Muon prefiring*. https://twiki.cern.ch/twiki/bin/viewauth/CMS/L1PrefiringWeightRecipe, Accessed on 2025-05-14.

[129] The CMS Collaboration. "Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV". In: *Journal of High Energy Physics* 2018.7 (July 2018). ISSN: 1029-8479. DOI: 10.1007/jhep07(2018)161. URL: http://dx.doi.org/10.1007/JHEP07(2018)161.

[130] The CMS collaboration. "Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC". In: *Journal of Instrumentation* 16.05 (May 2021), P05014. DOI: 10.1088/1748-0221/16/05/P05014. URL: https://dx.doi.org/10.1088/1748-0221/16/05/P05014.

[131] The CMS Collaboration. "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at s=13 TeV". In: *Journal of Instrumentation* 13.06 (June 2018), P06015–P06015. ISSN: 1748-0221. DOI: 10.1088/1748-0221/13/06/p06015. URL: http://dx.doi.org/10.1088/1748-0221/13/06/P06015.

[132] The CMS Collaboration. "Differential cross section measurements for the production of top quark pairs and of additional jets using dilepton events from pp collisions at $\sqrt{s}=$ 13 TeV". In: *Journal of High Energy Physics* 2025.2 (Feb. 2025). ISSN: 1029-8479. DOI: 10.1007/jhep02(2025)064. URL: http://dx.doi.org/10.1007/JHEP02(2025)064.

[133] M. Czakon et al. "Top-pair production at the LHC through NNLO QCD and NLO EW". In: *Journal of High Energy Physics* 2017.10 (Oct. 2017). ISSN: 1029-8479. DOI: 10.1007/jhep10(2017)186. URL: http://dx.doi.org/10.1007/JHEP10(2017)186.

[134] CMS. *Internal TWiki: Recommendation for Using b-tag Objects in Physics Analyses*. https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation, Accessed on 2025-05-13.

[135] CMS. *Internal TWiki: Methods to apply b-tagging efficiency scale factors*. https://twiki.cern.ch/twiki/bin/view/CMS/BTagSFMethods, Accessed on 2025-05-14.

[136] The CMS Collaboration. *CMS Internal TWiki: Top BTV Information for Analyses (RunII)*. https://twiki.cern.ch/twiki/bin/view/CMS/TopBTV, Accessed on 2025-05-14.

[137] CMS. "Precision luminosity measurement in proton–proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS". In: *The European Physical Journal C* 81.9 (Sept. 2021). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-021-09538-2. URL: http://dx.doi.org/10.1140/epjc/s10052-021-09538-2.

[138] *CMS luminosity measurement for the 2017 data-taking period at* $\sqrt{s} = 13$ TeV. Tech. rep. Geneva: CERN, 2018. URL: https://cds.cern.ch/record/2621960.

[139] *CMS luminosity measurement for the 2018 data-taking period at* $\sqrt{s} = 13$ TeV. Tech. rep. Geneva: CERN, 2019. URL: https://cds.cern.ch/record/2676164.

[140] The CMS Collaboration. "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV". In: *Journal of Instrumentation* 12.02 (Feb. 2017), P02014–P02014. ISSN: 1748-0221. DOI: 10.1088/1748-0221/12/02/p02014. URL: http://dx.doi.org/10.1088/1748-0221/12/02/P02014.

[141] CMS Top Quark Physics Group. *TOP Systematic Uncertainties (Run2)*. https://twiki.cern.ch/twiki/bin/viewauth/CMS/TopSystematics, Accessed on 2025-09-20.

[142] F. Demartin et al. *Higgs production in association with a single top quark at the LHC*. 2015. arXiv: 1504.00611 [hep-ph]. URL: https://arxiv.org/abs/1504.00611.

[143] J. Butterworth et al. "PDF4LHC recommendations for LHC Run II". In: *Journal of Physics G: Nuclear and Particle Physics* 43.2 (Jan. 2016), p. 023001. DOI: 10.1088/0954-3899/43/2/023001. URL: https://dx.doi.org/10.1088/0954-3899/43/2/023001.

[144] CMS. "Reweighting simulated events using machine-learning techniques in the CMS experiment". In: *The European Physical Journal C* 85.5 (May 2025). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-025-14097-x. URL: http://dx.doi.org/10.1140/epjc/s10052-025-14097-x.

[145] S. Argyropoulos and T. Sjöstrand. "Effects of color reconnection on $t\bar{t}$ final states at the LHC". In: *Journal of High Energy Physics* 2014.11 (Nov. 2014). ISSN: 1029-8479. DOI: 10.1007/jhep11(2014)043. URL: http://dx.doi.org/10.1007/JHEP11(2014)043.

[146] J. R. Christiansen and P. Z. Skands. "String formation beyond leading colour". In: *Journal of High Energy Physics* 2015.8 (Aug. 2015). ISSN: 1029-8479. DOI: 10.1007/jhep08(2015)003. URL: http://dx.doi.org/10.1007/JHEP08(2015)003.

[147] R. J. Barlow and C. Beeston. "Fitting using finite Monte Carlo samples". In: *Comput. Phys. Commun.* 77 (1993), pp. 219–228. DOI: 10.1016/0010-4655(93)90005-W.

[148] J. Conway. "Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra". In: (2011). Comments: Presented at PHYSTAT 2011, CERN, Geneva, Switzerland, January 2011, to be published in a CERN Yellow Report, pp. 115–120. DOI: 10.5170/CERN-2011-006.115. arXiv: 1103.0354. URL: https://cds.cern.ch/record/1333496.

[149] W. S. Cleveland. "Robust Locally Weighted Regression and Smoothing Scatterplots". In: *Journal of the American Statistical Association* 74.368 (1979), pp. 829–836. DOI: 10.1080/01621459.1979.10481038. eprint: https://www.tandfonline.com/doi/pdf/10.1080/01621459.1979.10481038. URL: https://www.tandfonline.com/doi/abs/10.1080/01621459.1979.10481038.

[150] W. S. Cleveland and S. J. Devlin. "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting". In: *Journal of the American Statistical Association* 83.403 (1988), pp. 596–610. DOI: 10.1080/01621459.1988.10478639. eprint: https://www.tandfonline.com/doi/pdf/10.1080/01621459.1988.10478639. URL: https://www.tandfonline.com/doi/abs/10.1080/01621459.1988.10478639.

[151] J. Shlomi, P. Battaglia, and J.-R. Vlimant. "Graph neural networks in particle physics". In: *Machine Learning: Science and Technology* 2.2 (Jan. 2021), p. 021001. ISSN: 2632-2153. DOI: 10.1088/2632-2153/abbf9a. URL: http://dx.doi.org/10.1088/2632-2153/abbf9a.

[152] W. Hamilton. "Graph Representation Learning". In: *Synthesis Lectures on Artificial Intelligence and Machine Learning* 14 (Sept. 2020), pp. 1–159. DOI: 10.2200/S01045ED1V01Y202009AIM046.

[153] E. Pfeffer et al. "A Case Study of Sending Graph Neural Networks Back to the Test Bench for Applications in High-Energy Particle Physics". In: *Computing and Software for Big Science* 8.1 (July 2024). ISSN: 2510-2044. DOI: 10.1007/s41781-024-00122-3. URL: http://dx.doi.org/10.1007/s41781-024-00122-3.

[154] D. P. Kingma and J. Ba. *Adam: A Method for Stochastic Optimization*. 2017. arXiv: 1412.6980 [cs.LG]. URL: https://arxiv.org/abs/1412.6980.

[155] N. Srivastava et al. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: *Journal of Machine Learning Research* 15.56 (2014), pp. 1929–1958. URL: http://jmlr.org/papers/v15/srivastava14a.html.

[156] J. Gilmer et al. *Neural Message Passing for Quantum Chemistry*. 2017. arXiv: 1704.01212 [cs.LG]. URL: https://arxiv.org/abs/1704.01212.

[157] A. Vaswani et al. *Attention Is All You Need*. 2023. arXiv: 1706.03762 [cs.CL]. URL: https://arxiv.org/abs/1706.03762.

[158] W. L. Hamilton, R. Ying, and J. Leskovec. *Representation Learning on Graphs: Methods and Applications*. 2018. arXiv: 1709.05584 [cs.SI]. URL: https://arxiv.org/abs/1709.05584.

[159] Y. Shi et al. *Masked Label Prediction: Unified Message Passing Model for Semi-Supervised Classification*. 2021. arXiv: 2009.03509 [cs.LG]. URL: https://arxiv.org/abs/2009.03509.

[160] J. L. Ba, J. R. Kiros, and G. E. Hinton. *Layer Normalization*. 2016. arXiv: 1607.06450 [stat.ML]. URL: https://arxiv.org/abs/1607.06450.

[161] S. Ioffe and C. Szegedy. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. 2015. arXiv: 1502.03167 [cs.LG]. URL: https://arxiv.org/abs/1502.03167.

[162] L. Zhao and L. Akoglu. *PairNorm: Tackling Oversmoothing in GNNs*. 2020. arXiv: 1909.12223 [cs.LG]. URL: https://arxiv.org/abs/1909.12223.

[163] C. Cangea et al. *Towards Sparse Hierarchical Graph Classifiers*. 2018. arXiv: 1811.01287 [stat.ML]. URL: https://arxiv.org/abs/1811.01287.

[164] B. Knyazev, G. W. Taylor, and M. R. Amer. *Understanding Attention and Generalization in Graph Neural Networks*. 2019. arXiv: 1905.02850 [cs.LG]. URL: https://arxiv.org/abs/1905.02850.

[165] H. Gao and S. Ji. *Graph U-Nets*. 2019. arXiv: 1905.05178 [cs.LG]. URL: https://arxiv.org/abs/1905.05178.

[166] S. Kullback and R. A. Leibler. "On information and sufficiency". In: *Annals of Mathematical Statistics* 22.1 (1951), pp. 79–86.

[167] K. P. Murphy. *Probabilistic Machine Learning: Advanced Topics*. MIT Press, 2023. URL: https://probml.github.io/pml-book.

[168] T.-Y. Lin et al. *Focal Loss for Dense Object Detection*. 2018. arXiv: 1708.02002 [cs.CV]. URL: https://arxiv.org/abs/1708.02002.

[169] G. Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *The European Physical Journal C* 71.2 (Feb. 2011). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-011-1554-0. URL: http://dx.doi.org/10.1140/epjc/s10052-011-1554-0.

[170] G. Cowan. *Topics in statistical data analysis for high-energy physics*. 2010. arXiv: 1012.3589 [physics.data-an]. URL: https://arxiv.org/abs/1012.3589.

[171] M. Jansen and G. Claeskens. "Cramér–Rao Inequality". In: *International Encyclopedia of Statistical Science*. Ed. by M. Lovric. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 322–323. ISBN: 978-3-642-04898-2. DOI: 10.1007/978-3-642-04898-2_197. URL: https://doi.org/10.1007/978-3-642-04898-2_197.

[172] J. Neyman and E. S. Pearson. "On the Problem of the Most Efficient Tests of Statistical Hypotheses". In: *Phil. Trans. Roy. Soc. Lond. A* 231.694-706 (1933), pp. 289–337. DOI: 10.1098/rsta.1933.0009.

[173] S. S. Wilks. "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses". In: *Annals Math. Statist.* 9.1 (1938), pp. 60–62. DOI: 10.1214/aoms/1177732360.

[174] CMS Collaboration. "The CMS Statistical Analysis and Combination Tool: Combine". In: *Computing and Software for Big Science* 8.1 (2024). DOI: 10.1007/s41781-024-00121-4. URL: https://doi.org/10.1007/s41781-024-00121-4.

[175] R. D. Cousins. *Generalization of χ^2 Goodness-of-Fit Test for Binned Data Using Saturated Models*. Tech. rep. Technical report. University of California, Los Angeles, Mar. 2013. URL: https://www.physics.ucla.edu/~cousins/stats/cousins_saturated.pdf.

[176] The ATLAS Collaboration. "Search for the $b\bar{b}$ decay of the Standard Model Higgs boson in associated (W/Z)H production with the ATLAS detector". In: *Journal of High Energy Physics* 2015.1 (Jan. 2015). ISSN: 1029-8479. DOI: 10.1007/jhep01(2015)069. URL: http://dx.doi.org/10.1007/JHEP01(2015)069.

[177] S. Camarda et al. "DYTurbo: fast predictions for Drell–Yan processes". In: *The European Physical Journal C* 80.3 (Mar. 2020). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-020-7757-5. URL: http://dx.doi.org/10.1140/epjc/s10052-020-7757-5.

[178] S. Camarda, L. Cieri, and G. Ferrera. "Fiducial perturbative power corrections within the q_T subtraction formalism". In: *The European Physical Journal C* 82.6 (June 2022). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-022-10510-x. URL: http://dx.doi.org/10.1140/epjc/s10052-022-10510-x.

[179] R. Gavin et al. "FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order". In: *Computer Physics Communications* 182.11 (Nov. 2011), pp. 2388–2403. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2011.06.008. URL: http://dx.doi.org/10.1016/j.cpc.2011.06.008.

[180] R. Gavin et al. *W physics at the LHC with FEWZ 2.1*. 2012. arXiv: 1201.5896 [hep-ph]. URL: https://arxiv.org/abs/1201.5896.

[181] Y. Li and F. Petriello. "Combining QCD and electroweak corrections to dilepton production in the framework of the FEWZ simulation code". In: *Physical Review D* 86.9 (Nov. 2012). ISSN: 1550-2368. DOI: 10.1103/physrevd.86.094034. URL: http://dx.doi.org/10.1103/PhysRevD.86.094034.

Appendix

A Acceptance study for early Run-III analysis

The LHC has started its third data-taking period (Run-III) in July 2022. The measurement of Drell-Yan (DY) production, $q\bar{q} \to Z(W) \to \ell\ell(\ell\nu)$ to a pair of leptons, with the Run-III commissioning dataset, represents one of the first physics analyses performed with the CMS detector after Long Shutdown2 (LS2). The theoretical prediction of cross sections and their uncertainties is essential for the modeling and interpretation of this dataset. The goal of this study is to measure the inclusive and fiducial cross sections of the DY process, and to evaluate the corresponding acceptance for Z and W boson production at a centre of mass energy of 13.6 TeV.

The Drell-Yan Turbo (DYTurbo) tool [177, 178] is used to calculate and compare the Z and W boson production cross sections at different orders of QCD perturbation theory by transverse momentum (q_T) resummation. DYTurbo calculates the cross section at the NNLO+NNLL (next-to-next-leading-order and next-to-next-leading logarithmic) accuracy with leading-order (LO) QED corrections. The cross section is calculated as a function of q_T, rapidity, y, and the invariant mass, m, of the lepton pair. For the inclusive cross-section calculations, a Fortran-based tool called Fully Exclusive W and Z Production (FEWZ) [179–181] is used for the theory prediction (with scale and PDF uncertainties). It calculates the DY process cross section at next-to-next-leading-order (NNLO) in QCD perturbation theory, no resummation, but with NLO (next-to-leading-order) QED corrections. The theoretical uncertainties associated with muon and electron channel acceptance are also studied. The comparison between these tools allows the evaluation of modeling differences and their impact on the predicted SM cross-section and acceptances. Tables A.1, A.2, A.3, A.4 provide the values of the predicted cross sections obtained for center-of-mass energies of 13 TeV and 13.6 TeV. The total phase space is defined as $50 < m_Z < 13000$ GeV, the inclusive phase space $60 < m_Z < 120$ GeV, the electron fiducial region $60 < m_Z < 120$ GeV with $p_T^e > 25$ GeV and $|\eta^e| < 2.5$, and the muon fiducial region $60 < m_Z < 120$ GeV with $p_T^\mu > 25$ GeV and $|\eta^\mu| < 2.4$. For the W boson, the total phase

space is defined without any kinematic constraint, the mW cut refers to the requirement of $m_W > 20$ GeV, the electron fiducial region is defined as $p_T^e > 29$ GeV and $|\eta^e| < 2.5$, and the muon fiducial region is defined as $p_T^\mu > 25$ GeV and $|\eta^\mu| < 2.4$. The total cross section is not reported for DYTurbo for the W boson, as the calculation is performed only for the phase space defined by $m_W > 20$. This inclusive phase space is used as the reference for the acceptance calculation.

Table A.1: Production cross sections of the $pp \to Z \to \ell\ell$ (each flavor) process at 13TeV calculated with DYTurbo and FEWZ. Uncertainties listed in the table are only integration uncertainties.

σ (pb)	DYTurbo	FEWZ
Total	1995.2 ± 1.2	2017.62 ± 0.17
Inclusive	1925.00 ± 0.75	1926.34 ± 0.16
Electron fid.	768.8 ± 1.2	758.96 ± 0.60
Muon fid.	735.68 ± 0.92	723.48 ± 0.63

Table A.2: Production cross sections of the $pp \to Z \to \ell\ell$ (each flavor) process at 13.6TeV calculated with DYTurbo and FEWZ. Uncertainties listed in the table are only integration uncertainties.

σ (pb)	DYTurbo	FEWZ
Total	2081.6 ± 1.2	2115.33 ± 0.18
Inclusive	2018.09 ± 0.80	2019.81 ± 0.17
Electron fid.	700.8 ± 1.0	682.18 ± 0.66
Muon fid.	766.6 ± 1.1	753.05 ± 0.62

Table A.3: Production cross sections of the $pp \to W \to \ell\nu$ (each flavor) process at 13TeV calculated with DYTurbo and FEWZ.

σ (pb)	W^+		W^-	
	DYTurbo	FEWZ	DYTurbo	FEWZ
Total	-	11531.3 ± 310.0	-	8281.3 ± 249.9
mW cut	11447.6 ± 7.7	11427.60 ± 0.91	8457.8 ± 6.2	8439.78 ± 0.66
Mu fid. & mW	5270.2 ± 5.1	5245.2 ± 3.6	4016.9 ± 3.6	3981.0 ± 2.4

From the dedicated inclusive and fiducial cross sections, the acceptance for vector boson production can be determined. The acceptance is defined as the ratio between the fiducial cross-section and the corresponding inclusive cross-section. In addition, different MC generators are compared in order to assess possible modeling effects on the acceptance calculation. Table A.5 summarizes the fiducial acceptance for Z and W boson production obtained with DYTurbo

Table A.4: Production cross sections of the $W \to \ell\nu$ (each flavor) process at 13.6TeV calculated with DYTurbo and FEWZ.

σ (pb)	W^+		W^-	
	DYTurbo	FEWZ	DYTurbo	FEWZ
Total	-	12783.7 ± 920.1	-	8710.16 ± 246.19
mW cut	11961.9 ± 9.3	11939.2 ± 0.98	8889.0 ± 6.1	8857.42 ± 0.70
Ele fid. & mW	4733.5 ± 5.7	4713.91 ± 3.64	3724.6 ± 6.3	3689.79 ± 2.62
Mu fid. & mW	5454.1 ± 4.5	5427.8 ± 3.9	4189.0 ± 3.6	4155.3 ± 2.6

and FEWZ at $\sqrt{s} = 13.6$ TeV. For the inclusive cross-section calculation, the lepton pairs are

Table A.5: Fiducial acceptance for vector boson production at 13.6 TeV collisions, based on the calculations listed in Tables 12 and 14. The acceptance for the Z boson is calculated as the ratio between the lepton fiducial cross section and the inclusive cross section. The acceptance for the W boson is calculated as the ratio between the lepton fiducial cross section with m_W cut to the cross section with only the m_W cut.

Acceptance	DYTurbo ($\epsilon_{\text{integration}}$)	FEWZ ($\epsilon_{\text{integration}} + \epsilon_{\text{PDF}} + \epsilon_{\text{scale}}$)
Z	$37.99\% \pm 0.06\%$	$37.28\% \pm 0.03\% \pm 0.21\%\ ^{+0.23\%}_{-0.18\%}$
W^+	$45.60\% \pm 0.05\%$	$45.46\% \pm 0.03\% \pm 0.21\%\ ^{+0.28\%}_{-0.21\%}$
W^-	$47.13\% \pm 0.05\%$	$46.91\% \pm 0.03\% \pm 0.20\%\ ^{+0.12\%}_{-0.14\%}$

selected with invariant mass (m_{ll}) between 60 GeV and 120 GeV. For acceptance calculations, the scale and PDF uncertainties are calculated using MadGraph (MG) by counting the number of generated events; the resummation uncertainty is taken from the kinematic acceptance with MADGRAPH5_aMC@NLO and is compared to the DYTurbo acceptance predictions with NLL (next-to-leading-logarithmic) accuracy matched with NLO accuracy. The fiducial region for Z boson acceptance is defined with leptons with $p_T > 25$ GeV and a pseudorapidity of $|\eta| < 2.4$ for the muon channel and $|\eta| < 2.5$ for the electron channel and dilepton invariant mass, $60\text{GeV} < m_{ll} < 120$ GeV. The acceptance comparison within DYTurbo between NLO+NLL and NNLO+NNLL tells us whether we should apply a correction for the higher-order corrections to the acceptance. Since samples from other generators are not available for 13.6 TeV, samples of 13 TeV are used to estimate some uncertainties.

The MG sample is generated at NLO+LL accuracy. The comparison between MG and DYTurbo tells us whether the resummation makes a significant difference and whether there should be a correction. The acceptance obtained with MadGraph is consistent with the DYTurbo prediction within the estimated theoretical and statistical uncertainties, indicating that resummation

effects do not introduce a significant change in the acceptance. The acceptance obtained with different generator configurations is shown in Table A.6.

Table A.6: Acceptance of $Z \to \ell\ell$ production calculated with different event generators.

Generator configuration	Acceptance (Muon fiducial)
MadGraph + Pythia	37.22%
Powheg + Pythia (dressed)	37.12%
Powheg + Photos (dressed)	37.17%
Powheg + Pythia (undressed)	36.75%
Powheg + Photos (undressed)	36.81%

The corresponding acceptance values for $W \to \ell\nu$ production are summarized in Table A.7. Since the differences between the generators are found to be negligible, no additional systematic

Table A.7: Fiducial acceptance of $W \to \ell\nu$ events calculated with different simulation samples. To evaluate the effect of the m_W cut on the fiducial acceptance, the acceptances listed in this table are calculated with respect to the total cross section (without any cut on m_W). Calculations are not split by charge, as the FSR uncertainty is expected to be the same for leptons of both charges.

Generator configuration	Muon fid.	Muon fid. & m_W
Powheg + Pythia (dressed)	44.8703%	44.8697%
Powheg + Photos (dressed)	44.8875%	44.8870%
Powheg + Pythia (undressed)	44.2984%	44.2977%
Powheg + Photos (undressed)	44.4040%	44.4036%

uncertainty from generator modeling is assigned. For more information and results, see Ref. [6].

B Pileup

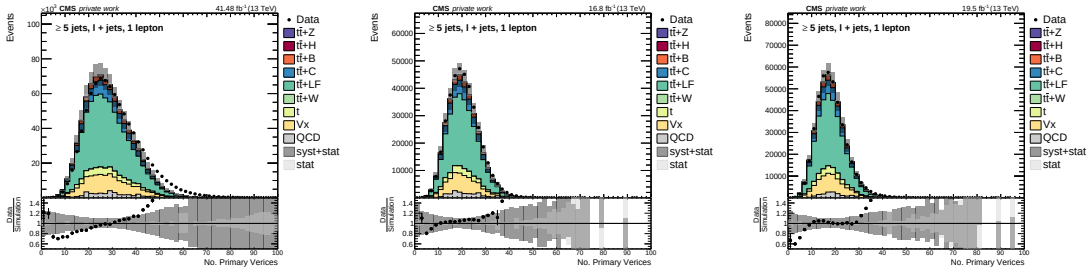


Figure B.1: Distribution of the number of reconstructed primary vertices in data and simulation after applying pileup reweighting. The comparison is shown for different data-taking eras, with the simulation corrected to match the pileup profile observed in data. The shaded bands indicate the total uncertainty on simulation, including the statistical and systematic uncertainties (Sec. 7.6) evaluated for the 2017, 2016postVFP, and 2016preVFP data-taking eras of Run II, respectively.

C Control distributions

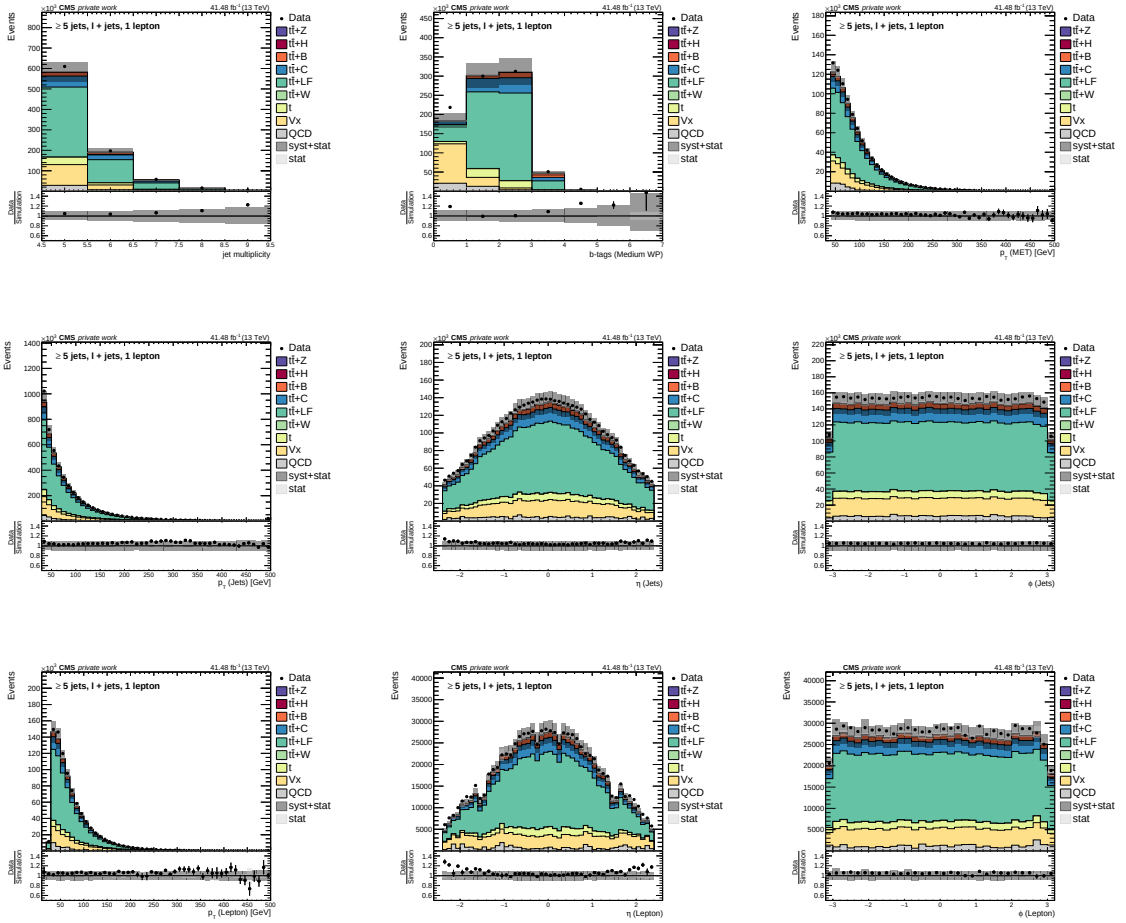


Figure C.2: Control distributions for the 2017 era in the inclusive $\geq$ 5-jet phase space of the semileptonic channel. The plots are displayed for jet multiplicity, b-tag multiplicity (medium WP), leading-object kinematics: p_T of jet and lepton, angular variables η and ϕ respectively. The simulation is shown as stacked histograms of all processes, and the data are shown as black points. The legend lists all simulated processes in the same order as the stack, from top to bottom. All relevant background processes are included. The gray band shows the total uncertainty on simulation, including the statistical and systematic uncertainties (Sec. 7.6) evaluated for the era 2017. The lower panel displays the data-to-simulation ratio with the same uncertainty band. Unless stated otherwise, all expectations are pre-fit and normalized to the era's integrated luminosity using theoretical predictions of cross sections of the processes.

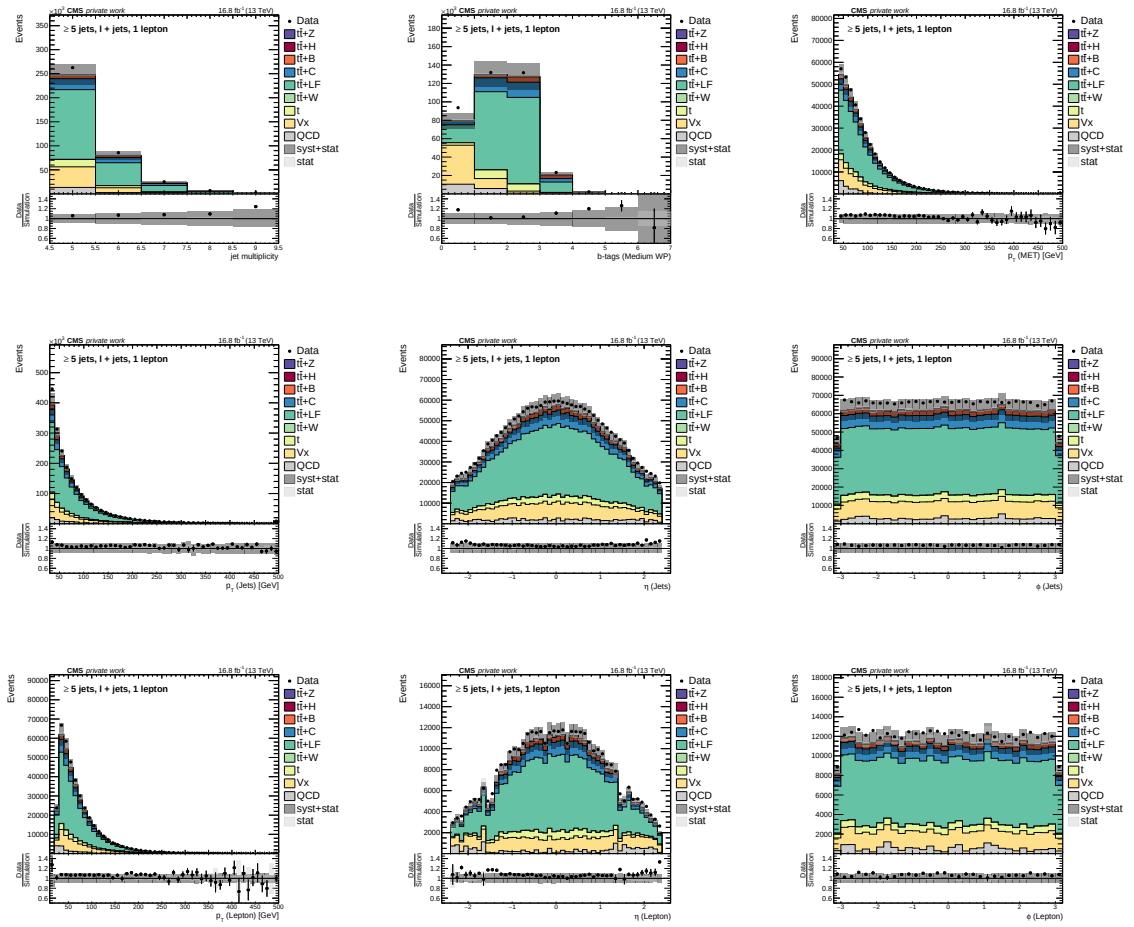


Figure C.3: Control distributions for the 2016postVFP era in the inclusive $\geq$ 5-jet phase space of the semileptonic channel. The plots are displayed for jet multiplicity, b-tag multiplicity (medium WP), leading-object kinematics: p_T of jet and lepton, angular variables η and ϕ respectively. The simulation is shown as stacked histograms of all processes, and the data are shown as black points. The legend lists all simulated processes in the same order as the stack, from top to bottom. All relevant background processes are included. The gray band shows the total uncertainty on simulation, including the statistical and systematic uncertainties (Sec. 7.6) evaluated for the era 2016postVFP. The lower panel displays the data-to-simulation ratio with the same uncertainty band. Unless stated otherwise, all expectations are pre-fit and normalized to the era's integrated luminosity using theoretical predictions of cross sections of the processes.

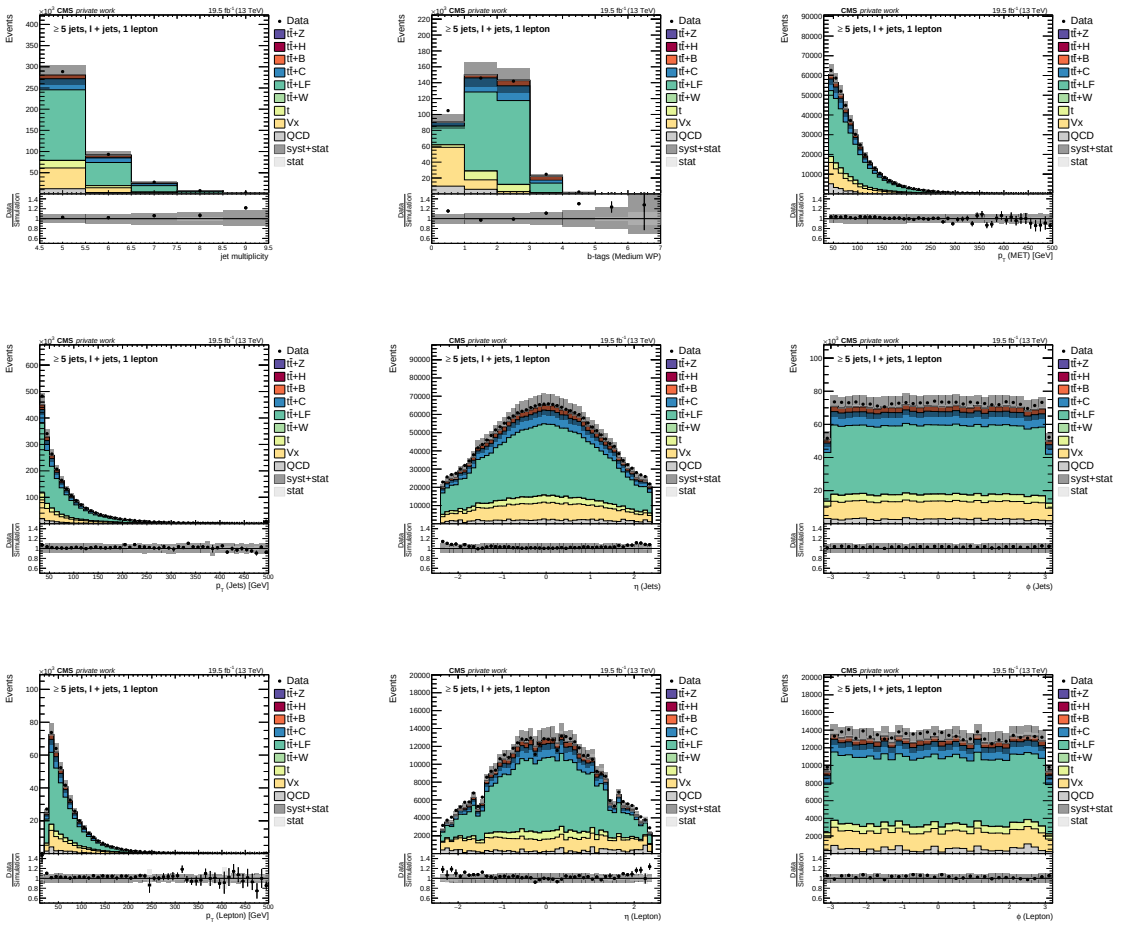


Figure C.4: Control distributions for the 2016preVFP era in the inclusive ≥ 5-jet phase space of the semileptonic channel. The plots are displayed for jet multiplicity, b-tag multiplicity (medium WP), leading-object kinematics: p_T of jet and lepton, angular variables η and ϕ respectively. The simulation is shown as stacked histograms of all processes, and the data are shown as black points. The legend lists all simulated processes in the same order as the stack, from top to bottom. All relevant background processes are included. The gray band shows the total uncertainty on simulation, including the statistical and systematic uncertainties (Sec. 7.6) evaluated for the era 2016preVFP. The lower panel displays the data-to-simulation ratio with the same uncertainty band. Unless stated otherwise, all expectations are pre-fit and normalized to the era's integrated luminosity using theoretical predictions of cross sections of the processes.

D GoF tests - 2016

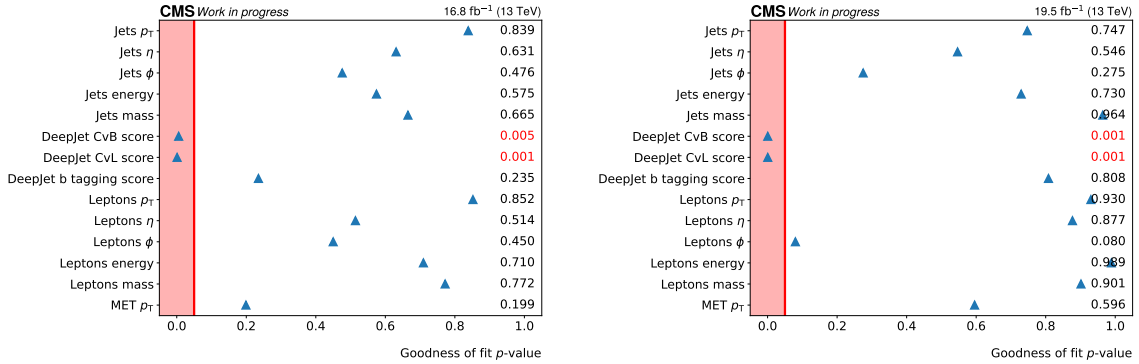
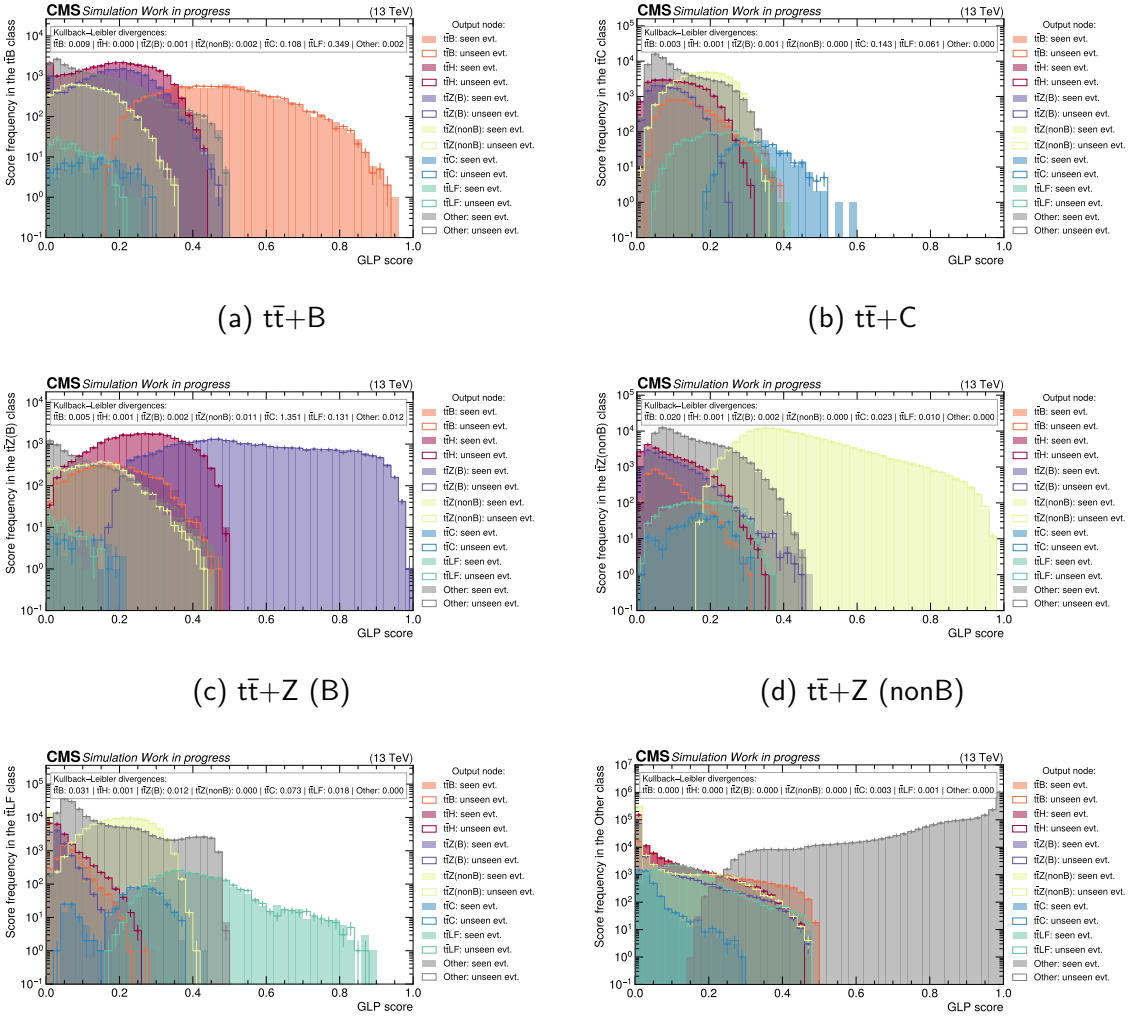


Figure D.5: Anderson–Darling (AD) goodness-of-fit p-values for representative NLP input distributions between data and MC. On the left, 2016postVFP, and on the right, 2016preVFP are shown. Each marker shows the AD p-value for the indicated variable, and the shaded band highlights the low-p region (e.g., $p < 0.05$). The CvB and CvL tagger scores do not pass the GoF test, as observables with p-values greater than 5% are only retained as input features in the analysis. The plot demonstrates the divergence observed for CvB, and thus, the two c-jet tagger discriminants, CvB and CvL, are excluded from the training process. The other input features are included in the GNN training.

E GLP validation



(a) $t\bar{t}$+B

(b) $t\bar{t}$+C

(c) $t\bar{t}$+Z (B)

(d) $t\bar{t}$+Z (nonB)

(e) $t\bar{t}$ + LF

(f) other

Figure E.6: Validation of GLP score stability for the six output nodes. The score distributions are shown separately for even-numbered (seen) and odd-numbered (unseen) events. The agreement between the two subsets demonstrates the absence of significant overtraining and the robustness of the classifier response.

F GLP yields

Table F.8: Event yields per GLP class after the full selection for the 2017 dataset. For each column, the process matching the GLP class label is treated as a signal, while all remaining processes contribute to the background in that column.

	GLP class							
Process	$t\bar{t}B$	$t\bar{t}H$	$t\bar{t}Z$(B)	$t\bar{t}Z$(nonB)	$t\bar{t}C$	$t\bar{t}+$LF	Other	Total
Data	10039	8511	3744	50118	57159	129588	627568	886727
$t\bar{t}B$	3730	2544	1113	2245	3528	3265	17832	34257
$t\bar{t}H$	64	248	52	211	146	173	664	1558
$t\bar{t}Z$	41	68	52	319	138	226	936	1780
$t\bar{t}C$	1283	1490	479	8254	11747	16315	53240	92808
$t\bar{t}+$LF	1895	2468	768	31725	30313	95652	333047	495868
$t\bar{t}W$	25	37	16	259	143	221	909	1610
single top	1206	530	306	3122	2722	5468	37011	50365
VV	7	3	4	109	50	126	2222	2521
W+jets	165	71	87	3458	2064	4741	101421	112007
Z+jets	59	17	21	478	417	1188	14711	16891
QCD	94	56	44	401	1470	4368	31004	37437
Data/MC	1.172	1.130	1.273	0.991	1.084	0.984	1.058	1.047
S/B	0.771	0.034	0.018	–	0.287	–	–	–
$S/\sqrt{B}$	53.6	2.9	1.0	–	58.0	–	–	–

Table F.9: Event yields per GLP class after the full selection for the 2016postVFP dataset. For each column, the process matching the GLP class label is treated as a signal, while all remaining processes contribute to the background in that column.

	GLP class							
Process	$t\bar{t}B$	$t\bar{t}H$	$t\bar{t}Z$(B)	$t\bar{t}Z$(nonB)	$t\bar{t}C$	$t\bar{t}+$LF	Other	Total
Data	3515	3174	1318	24329	24618	57134	268619	382707
$t\bar{t}B$	1357	1002	439	1189	1726	1503	7982	15198
$t\bar{t}H$	24	101	22	109	67	78	304	705
$t\bar{t}Z$	14	26	21	156	63	98	422	800
$t\bar{t}C$	388	512	167	4171	5019	7563	23750	37570
$t\bar{t}+$LF	645	948	306	15966	13878	41853	147701	221297
$t\bar{t}W$	8	13	6	134	62	99	419	741
single top	400	188	103	1416	1189	2299	15510	21105
VV	1	1	1	52	18	55	912	1040
W+jets	54	24	35	1589	914	2036	43275	47927
Z+jets	21	15	6	187	158	477	5817	6681
QCD	73	17	12	320	472	3191	13223	17308
Data/MC	1.177	1.115	1.179	1.001	1.091	0.964	1.049	1.033
S/B	0.833	0.037	0.019	–	0.286	–	–	–
$S/\sqrt{B}$	33.6	1.93	0.63	–	37.9	–	–	–

Table F.10: Event yields per GLP class after the full selection for the 2016preVFP dataset. For each column, the process matching the GLP class label is treated as a signal, while all remaining processes contribute to the background in that column.

	GLP class							
Process	$t\bar{t}B$	$t\bar{t}H$	$t\bar{t}Z$(B)	$t\bar{t}Z$(nonB)	$t\bar{t}C$	$t\bar{t}+$LF	Other	Total
Data	3559	3629	1536	28941	25985	61880	294375	421905
$t\bar{t}B$	1485	1133	490	1454	1878	1705	9083	17228
$t\bar{t}H$	25	113	23	131	73	90	350	802
$t\bar{t}Z$	15	29	23	184	66	112	485	914
$t\bar{t}C$	422	566	181	5072	5427	8698	27255	47421
$t\bar{t}+$LF	712	1079	333	19398	15015	48148	169187	253872
$t\bar{t}W$	9	15	7	156	73	108	480	848
single top	448	208	126	1711	1279	2662	17898	24332
VV	2	1	3	55	19	69	1042	1191
W+jets	69	29	31	1935	945	2344	49328	54681
Z+jets	18	6	7	231	169	608	6505	7544
QCD	36	19	11	526	644	2691	12462	164389
Data/MC	1.098	0.860	1.241	0.969	1.016	0.949	0.998	0.973
S/B	0.845	0.028	0.019	–	0.269	–	–	–
$S/\sqrt{B}$	35.5	1.76	0.67	–	38.2	–	–	–

G Pre-fit distributions

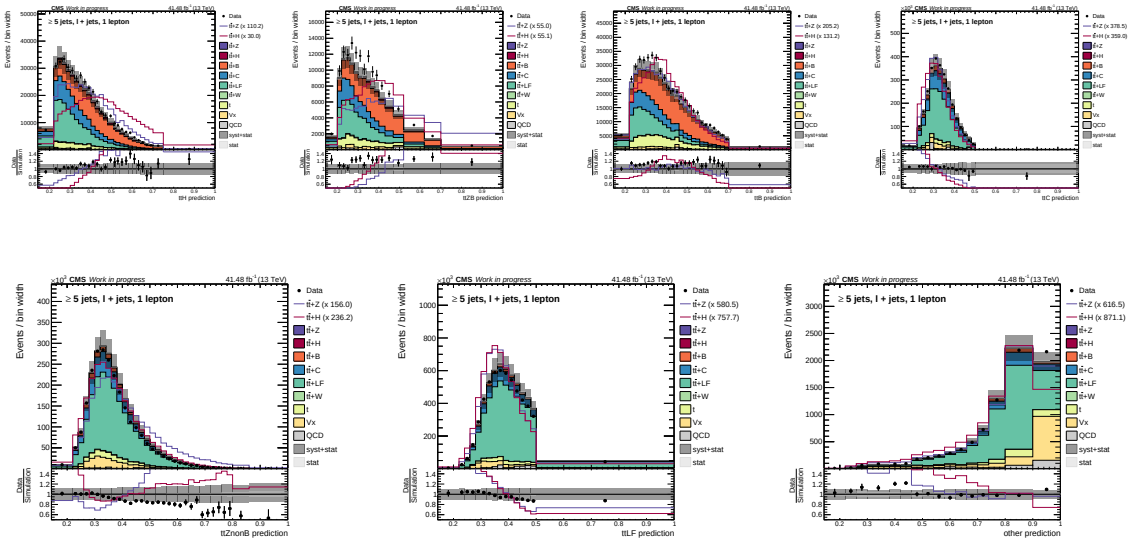


Figure G.7: 2017 pre-fit GLP score distributions per event node. Top row shows signal processes ($t\bar{t}$+H, $t\bar{t}$+Z, $t\bar{t}$+B, and $t\bar{t}$+C); bottom row shows the pre-fit GLP score distributions for the background classes $t\bar{t}$ + Z(nonB), $t\bar{t}$ + LF, and 'other'. The simulation is shown as stacked histograms of all processes, and the data is represented by black markers. The $t\bar{t}$+H and $t\bar{t}$+Z predictions are additionally overlaid as line distributions normalized to the total yield to enhance their visibility. The gray band indicates the total uncertainty on simulation, including both statistical and systematic components described in Sec. 7.6. The lower panel displays the data-to-simulation ratio.

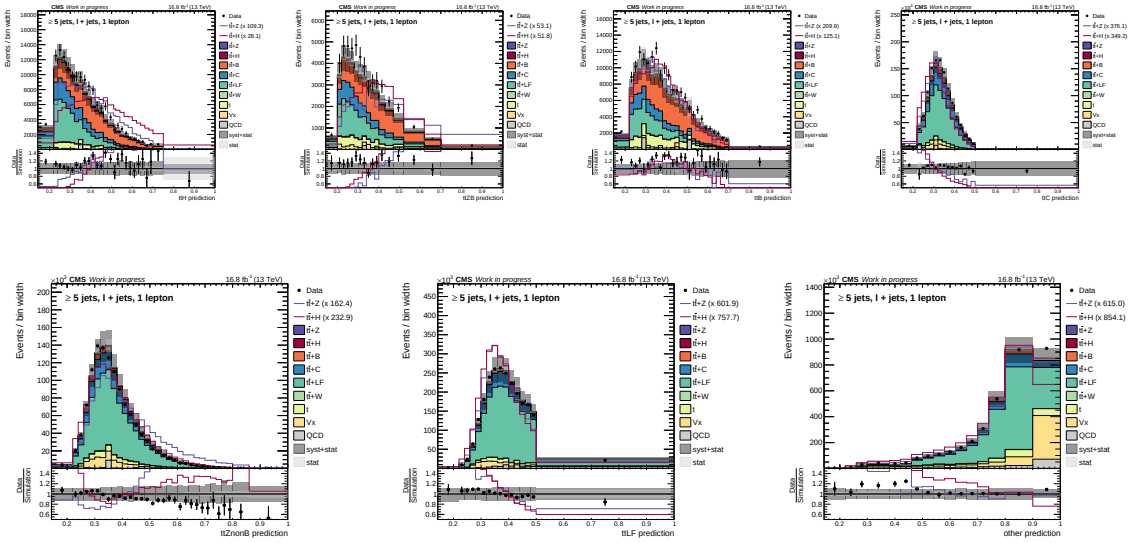


Figure G.8: 2016postVFP pre-fit GLP score distributions per event node. The top row shows signal processes ($t\bar{t}$+H, $t\bar{t}$+Z, $t\bar{t}$+B, and $t\bar{t}$+C); bottom row shows the pre-fit GLP score distributions for the background classes $t\bar{t}$ + Z(nonB), $t\bar{t}$ + LF, and 'other'. The simulation is shown as stacked histograms of all processes, and the data is represented by black markers. The $t\bar{t}$+H and $t\bar{t}$+Z predictions are additionally overlaid as line distributions normalized to the total yield to enhance their visibility. The gray band indicates the total uncertainty on simulation, including both statistical and systematic components described in Sec. 7.6. The lower panel displays the data-to-simulation ratio.

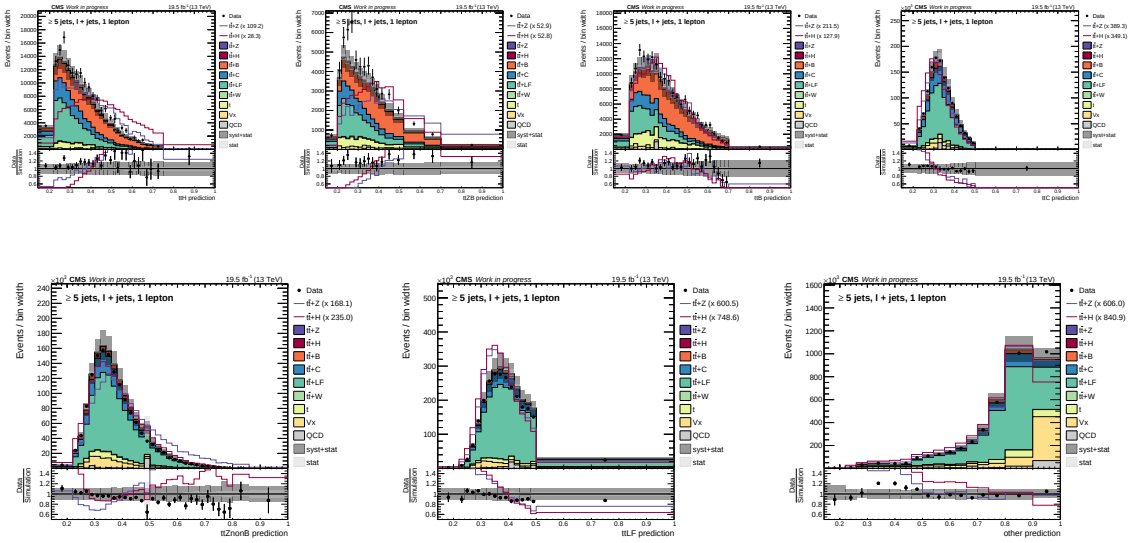


Figure G.9: 2016preVFP pre-fit GLP score distributions per event node. Top row shows signal processes ($t\bar{t}$+H, $t\bar{t}$+Z, $t\bar{t}$+B, and $t\bar{t}$+C); bottom row shows the pre-fit GLP score distributions for the background classes $t\bar{t}$ + Z(nonB), $t\bar{t}$ + LF, and 'other'. The simulation is shown as stacked histograms of all processes, and the data is represented by black markers. The $t\bar{t}$+H and $t\bar{t}$+Z predictions are additionally overlaid as line distributions normalized to the total yield to enhance their visibility. The gray band indicates the total uncertainty on simulation, including both statistical and systematic components described in Sec. 7.6. The lower panel displays the data-to-simulation ratio.

H GoF test results: year-wise comparison

The goodness-of-fit test evaluated independently for each Run-II data-taking era is shown in Fig. H.10. It assesses the level of agreement between the observed data and the postfit model. The obtained p-values are 0.109, 0.075, 0.256, and 0.036 for years 2018, 2017, 2016postVFP, and 2016preVFP, respectively. All years exhibit values greater than the threshold of 0.5, except for the 2016preVFP era, and thus lie within the range expected from statistical fluctuations, indicating no significant tension between the data and the model prediction. While the 2016preVFP dataset shows the lowest p-value below the commonly used threshold of 0.05, it is not alarming since this era typically has fewer events and thus larger statistical variations. Overall, the GoF results demonstrate a satisfactory description of the data across the full Run-II period.

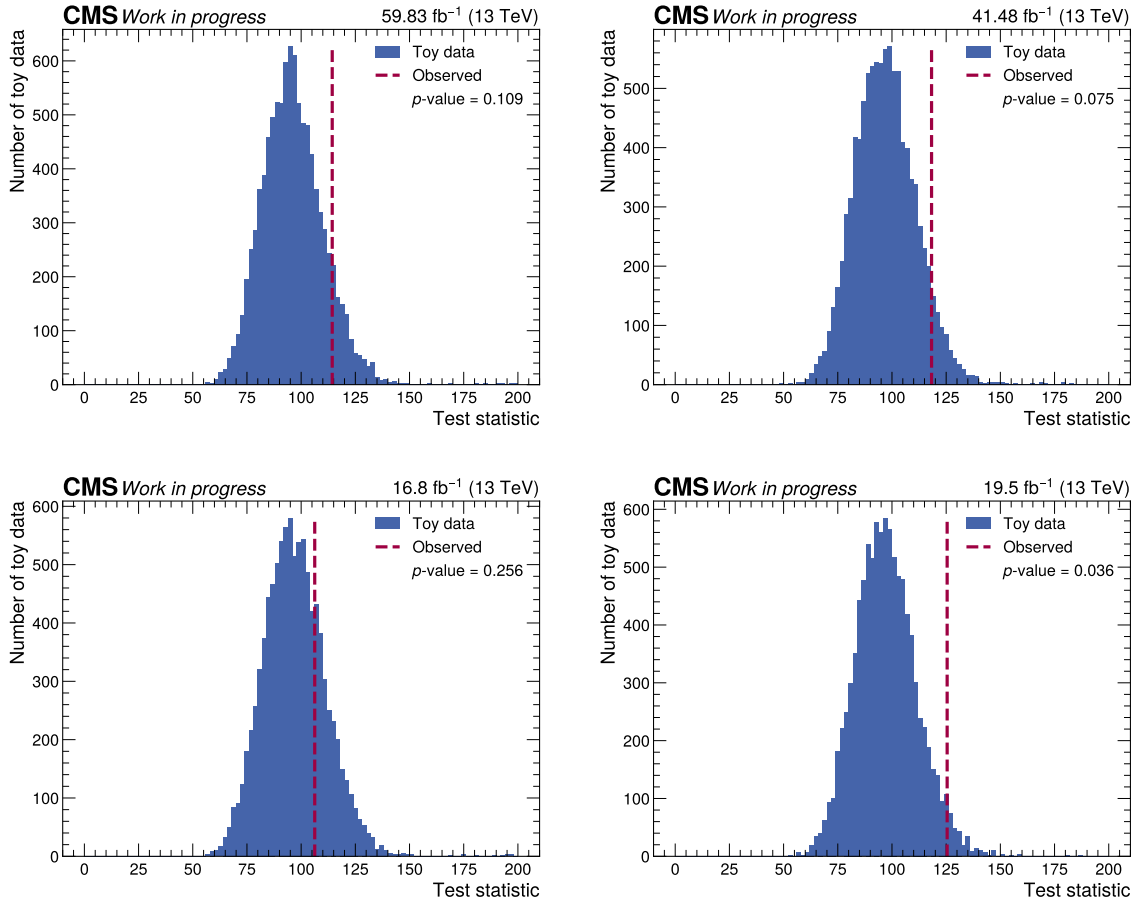


Figure H.10: GoF test for the different data-taking periods.

I (Asimov) Fit results: Correlation matrices

The post-fit correlation matrix obtained from the pseudodata fit for the 2017 data-taking period is shown in Fig. I.11, and for the 2016postVFP and 2016preVFP eras are shown in Fig. I.12. The correlations are derived from the covariance matrix of the profiled likelihood, with NPs associated with limited simulation statistics omitted for clarity. The overall structure is characterized by predominantly small off-diagonal elements, consistent with that observed for 2018, indicating that most NPs remain only weakly correlated after the fit. Moderate (anti-)correlations are visible among groups of related uncertainties, such as JES components, flavour-tagging uncertainties, and theory-driven modeling uncertainties. As in the 2018 case, a correlation between the FSR and pileup uncertainties can be observed. Additional localized correlations include the expected interplay between the QCD scale uncertainty of $t\bar{t}$+B production and the corresponding $\mu_{t\bar{t}+B}$, as well as a visible correlation between the $t\bar{t}$+H signal strength and the gluon-splitting $t\bar{t}$2b uncertainty, with a smaller effect present for $t\bar{t}$+Z. This is comparatively more pronounced in 2017 than in 2016. Anti-correlations are also observed between the vector-boson normalization parameter and its associated QCD scale uncertainty.

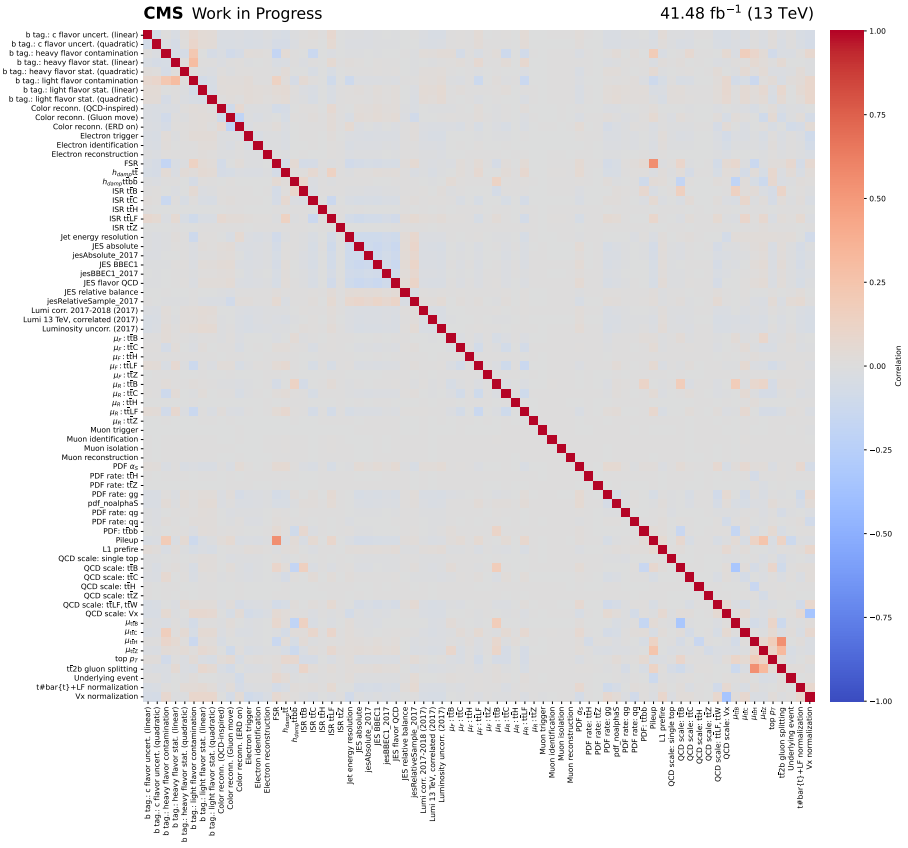


Figure I.11: Correlation matrices of the nuisance parameters (NPs) and parameters of interest (POIs) obtained from the pseudodata fits for the 2017 data-taking era.

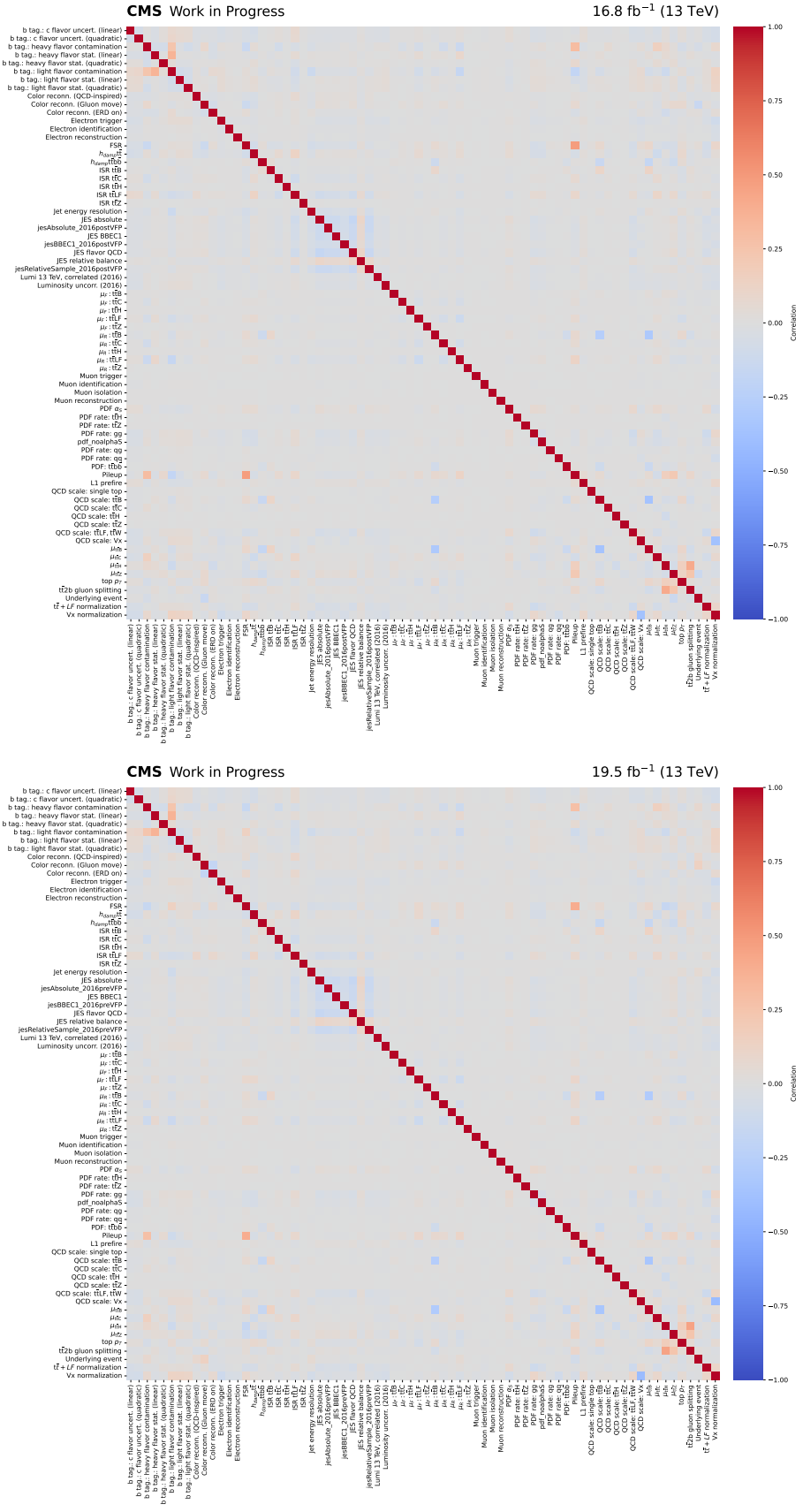


Figure I.12: Correlation matrices of the nuisance parameters (NPs) and parameters of interest (POIs) obtained from the pseudodata fits for the 2016postVFP and 2016preVFP data-taking eras.

No pronounced correlation bands or near-degenerate parameter pairs are observed, confirming the stability of the statistical model across all Run-II eras.

J Post-fit distributions

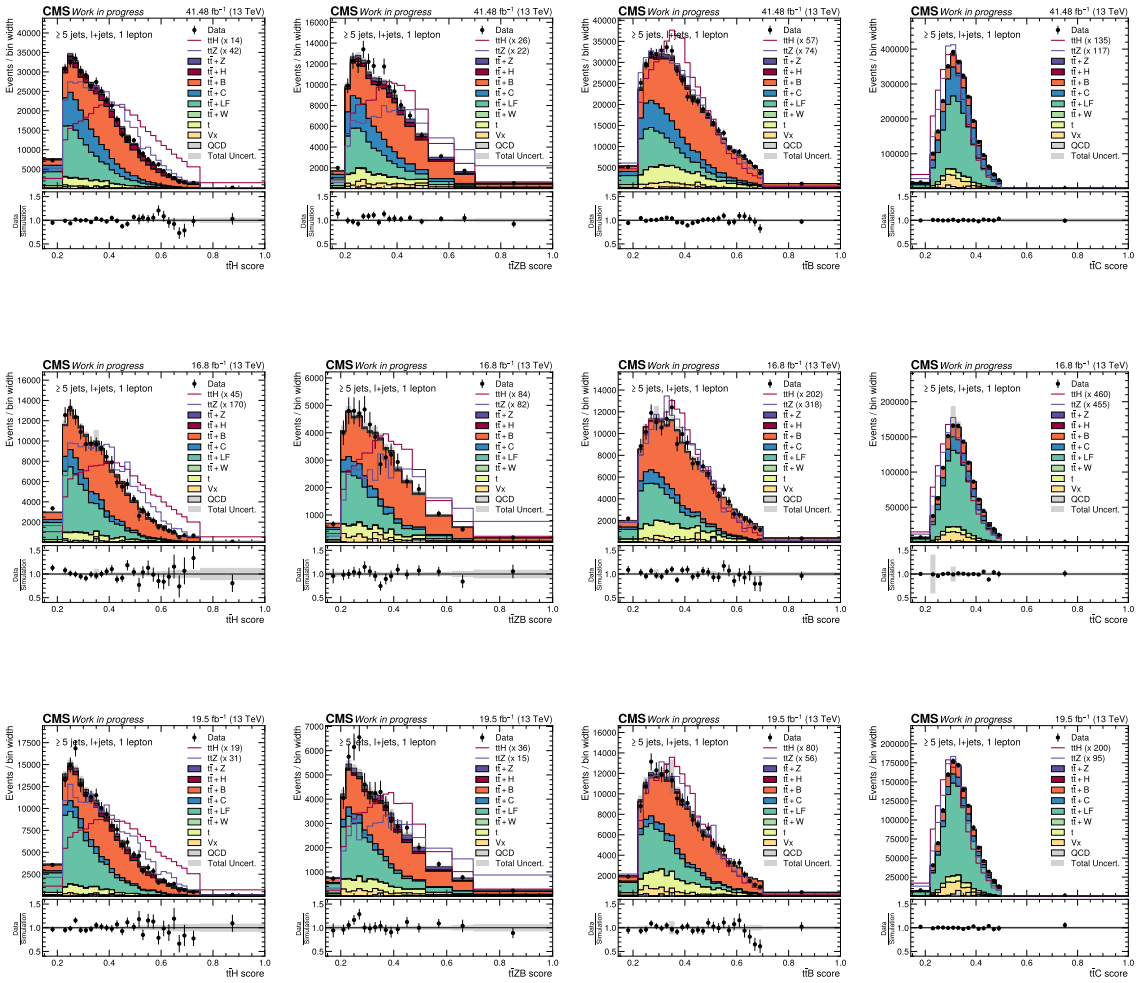


Figure J.15: Post-fit GLP score distributions for the signal classes $t\bar{t}$+H, $t\bar{t}$+Z, $t\bar{t}$+B, and $t\bar{t}$+C for 2017, 2016postVFP, and 2016preVFP eras respectively. The post-fit prediction is represented by stacked histograms of all processes, and the observed data is represented by black markers. The $t\bar{t}$+H and $t\bar{t}$+Z predictions are additionally overlaid as line distributions normalized to the total post-fit yield to enhance their visibility. The gray band indicates the total post-fit uncertainty, including both statistical and systematic components described in Sec. 7.6.

K Fit results: Correlation matrices

The post-fit correlation matrix obtained from the fit to the 2017 data is shown in Fig. K.16, and for the 2016postVFP and 2016preVFP eras are shown in Fig. K.17. The correlations are derived from the covariance matrix of the profiled likelihood, with NPs associated with limited simulation statistics omitted for clarity. The overall structure is characterized by predominantly small off-diagonal elements, consistent with that observed for 2018, indicating that most NPs remain only weakly correlated after the fit. Moderate correlations are observed among systematic uncertainties, including JES components, flavor-tagging uncertainties, and modeling variations. Correlations between pileup and FSR are visible, as well as localized correlations involving QCD scale parameters and signal strength parameters, such as the interplay between QCD scale variations and the corresponding $t\bar{t}$+B and $t\bar{t}$+H signal strengths. Anti-correlations between normalization parameters and their associated QCD scale uncertainties are also observed. The 2016postVFP and 2016preVFP data correlation matrices also show patterns similar to what was observed for 2017 and 2018.

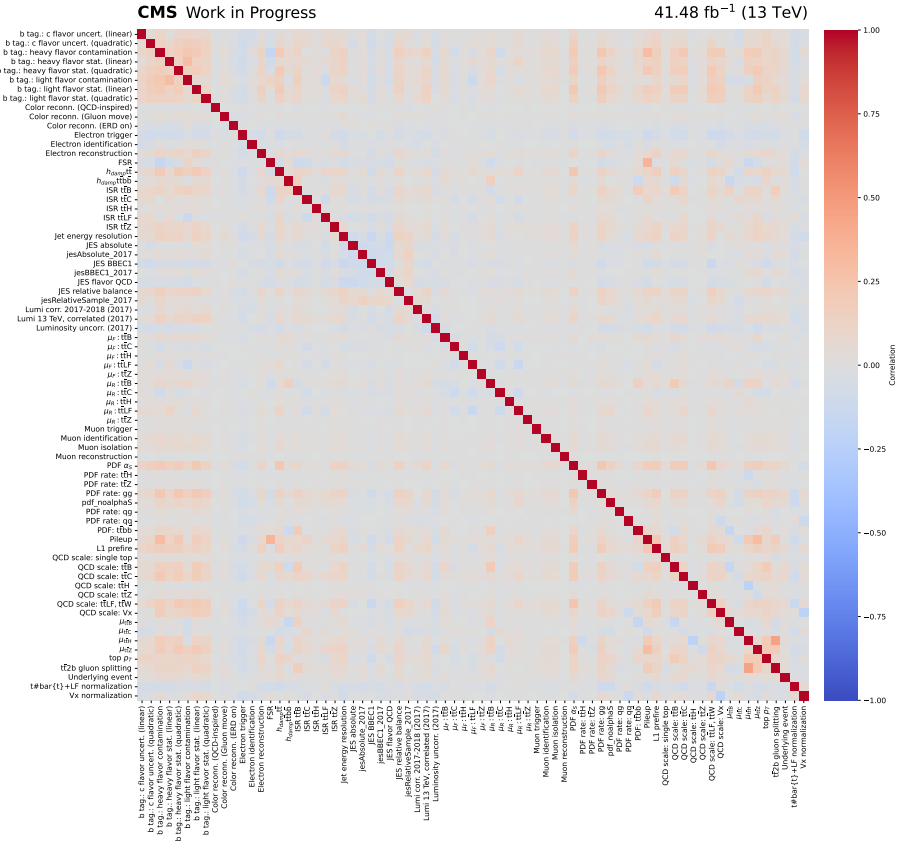


Figure K.16: Correlation matrices of the nuisance parameters (NPs) and parameters of interest (POIs) obtained from the data fits for the 2017 data-taking era.

No pronounced correlation bands or near-degenerate parameter pairs are observed, confirming the stability of the statistical model across all Run-II eras.

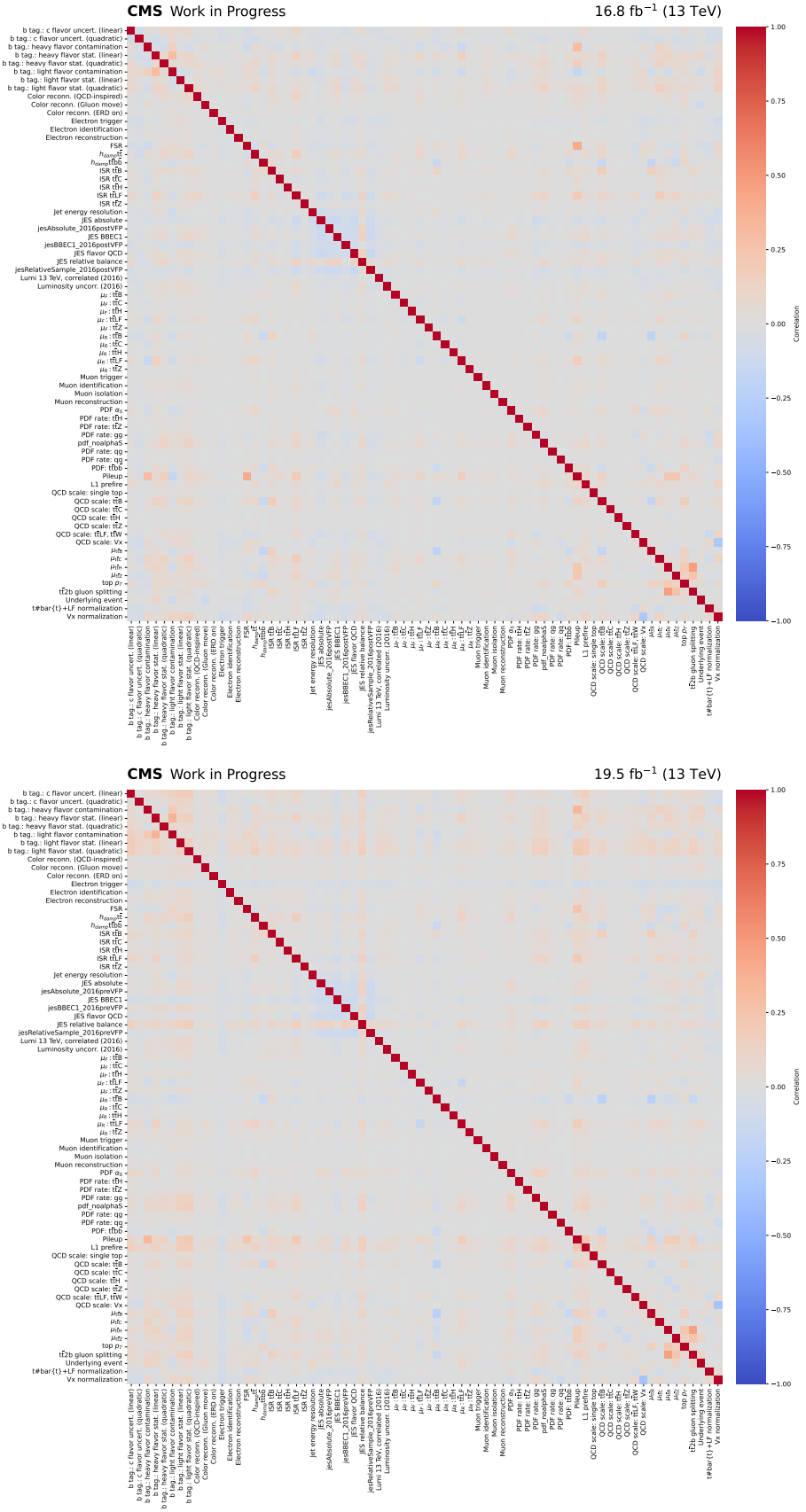


Figure K.17: Correlation matrices of the nuisance parameters (NPs) and parameters of interest (POIs) obtained from the data fits for the 2016postVFP and 2016preVFP data-taking eras.

L Impact plots

In the following pages, impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the individual Run-II datasets are given. For the combined Run-II dataset, all NPs except the first 15 are included, while the remaining NPs are presented across the subsequent pages up to page 10. For all eras, 2018, 2017, 2016postVFP, and 2016preVFP, the full set of NPs is shown.

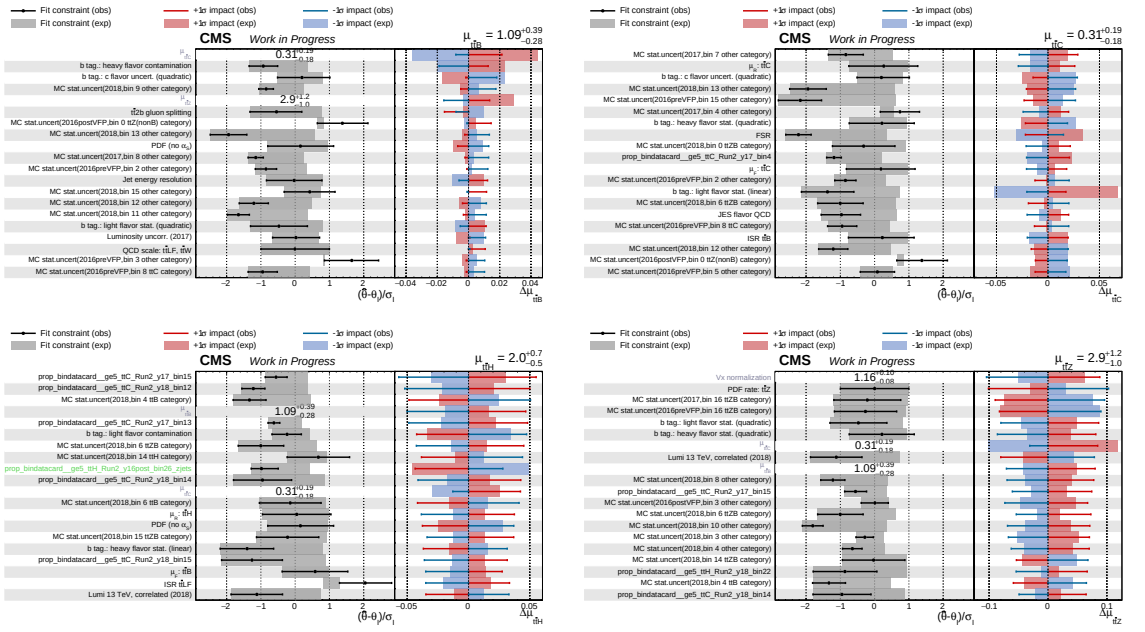


Figure L.18: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the combined Run-II dataset. The middle panels show the normalized displacement $(\hat{\theta} - \theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm 1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (16-35) are ordered according to the magnitude of their impact.

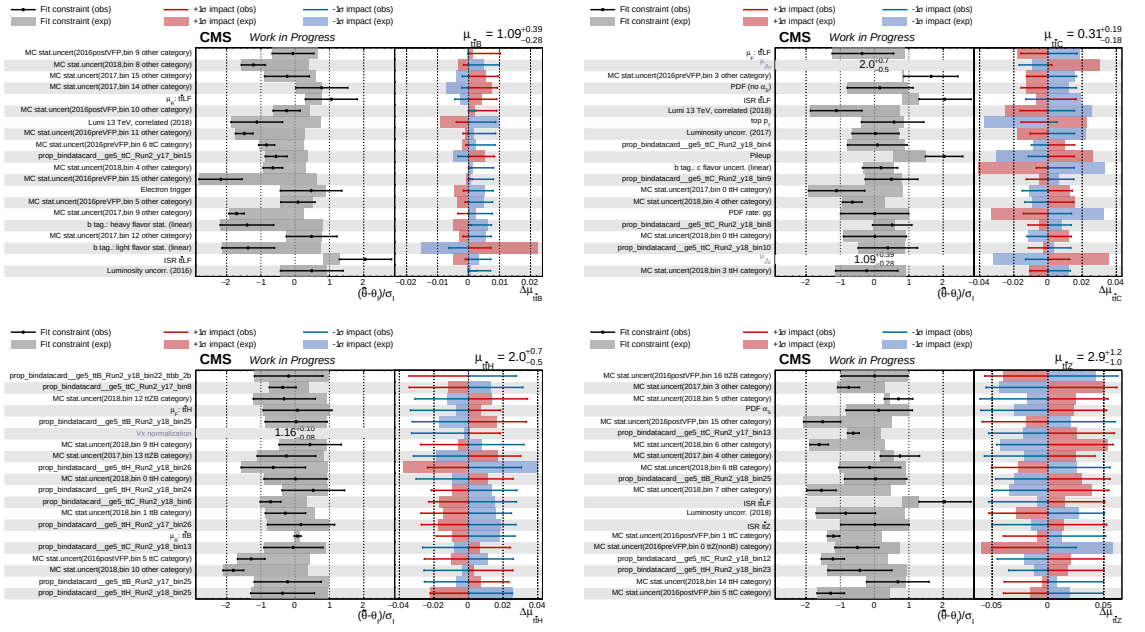


Figure L.19: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the combined Run-II dataset. The middle panels show the normalized displacement $(\hat{\theta}-\theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (36-55) are ordered according to the magnitude of their impact.

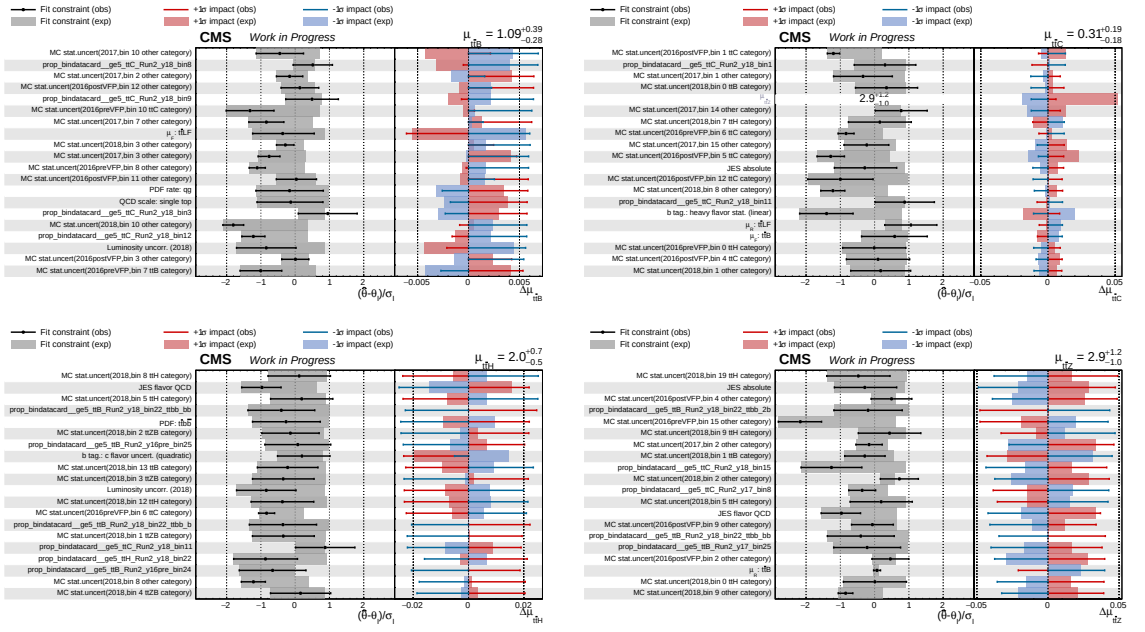


Figure L.20: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the combined Run-II dataset. The middle panels show the normalized displacement $(\hat{\theta} - \theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm 1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (56-75) are ordered according to the magnitude of their impact.

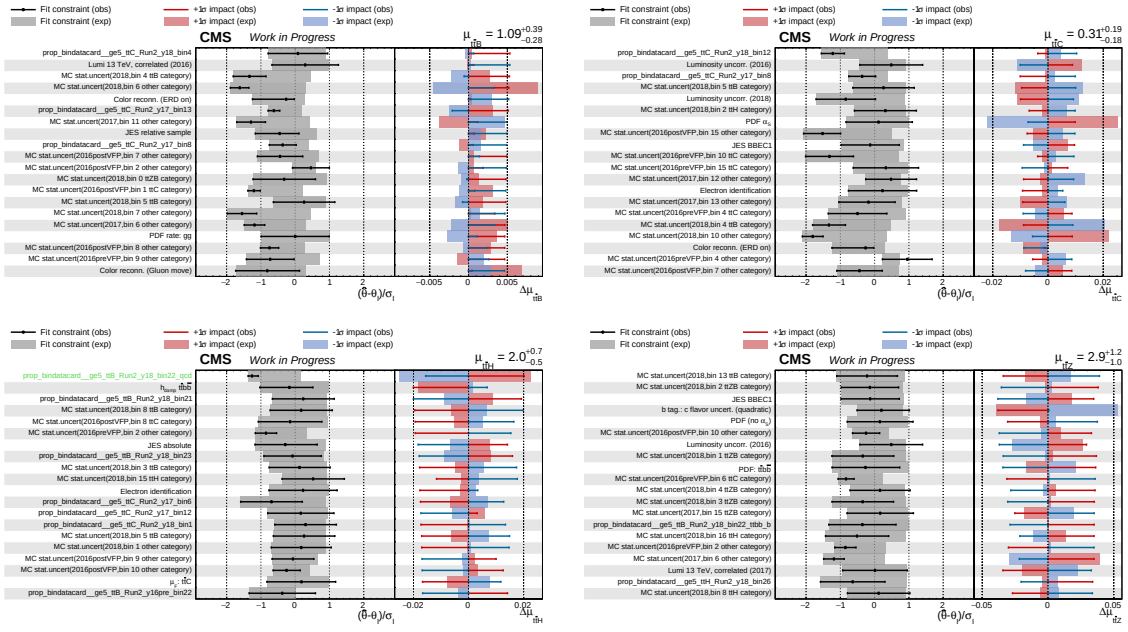


Figure L.21: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the combined Run-II dataset. The middle panels show the normalized displacement $(\hat{\theta}-\theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (76-95) are ordered according to the magnitude of their impact.

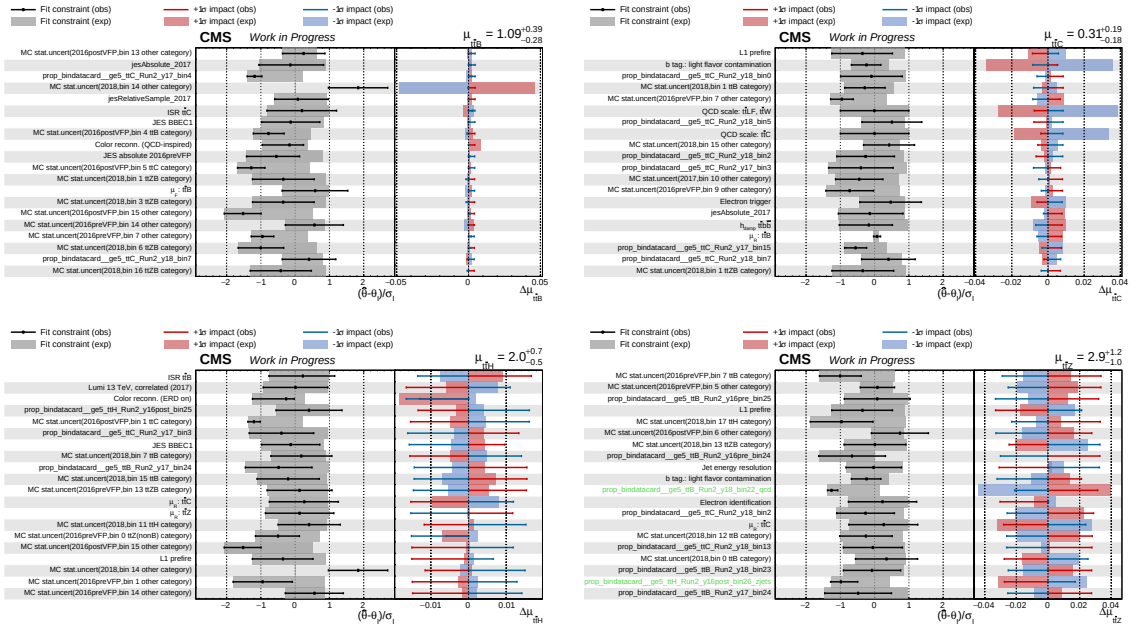


Figure L.22: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the combined Run-II dataset. The middle panels show the normalized displacement $(\hat{\theta} - \theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm 1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (96-115) are ordered according to the magnitude of their impact.

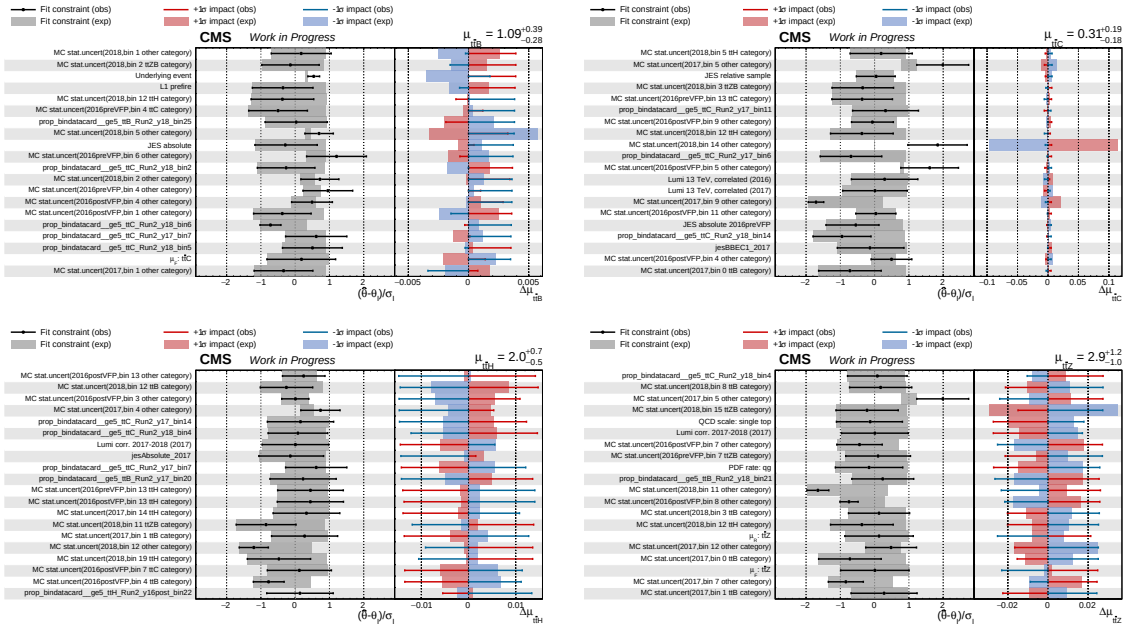


Figure L.23: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the combined Run-II dataset. The middle panels show the normalized displacement $(\hat{\theta} - \theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm 1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (116-135) are ordered according to the magnitude of their impact.

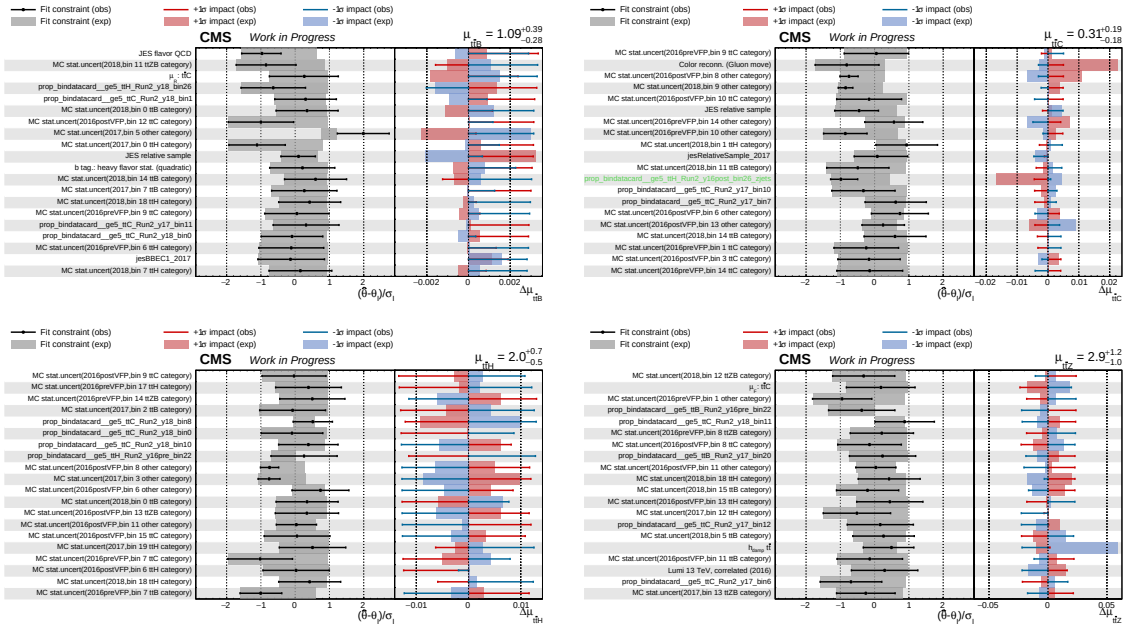


Figure L.24: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the combined Run-II dataset. The middle panels show the normalized displacement $(\hat{\theta} - \theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm 1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (136-155) are ordered according to the magnitude of their impact.

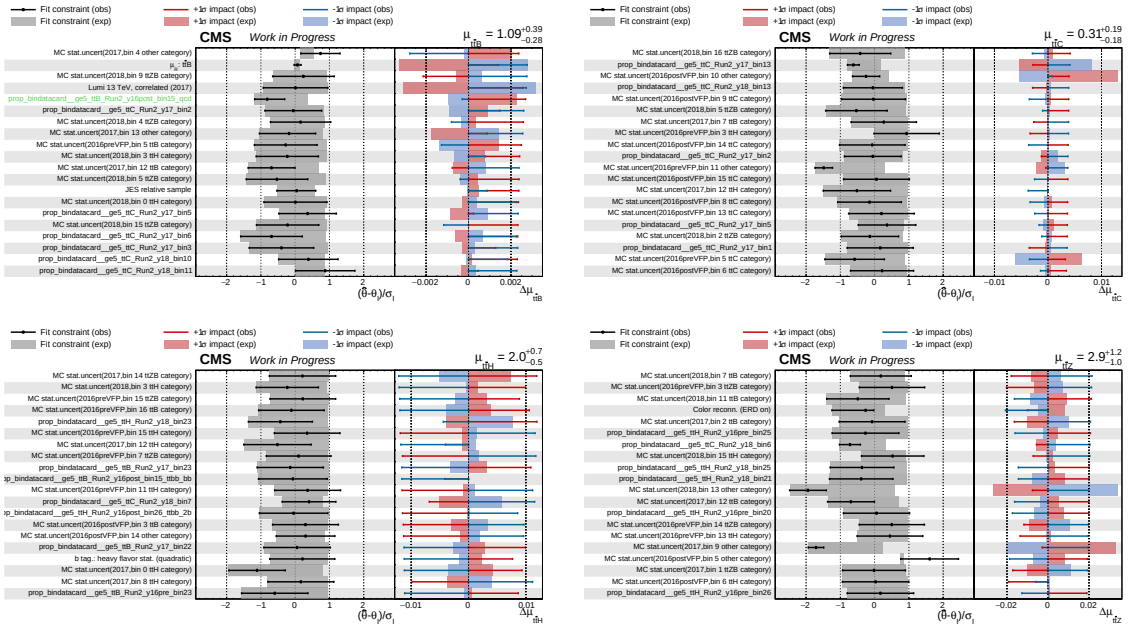


Figure L.25: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the combined Run-II dataset. The middle panels show the normalized displacement $(\hat{\theta}-\theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (156-175) are ordered according to the magnitude of their impact.

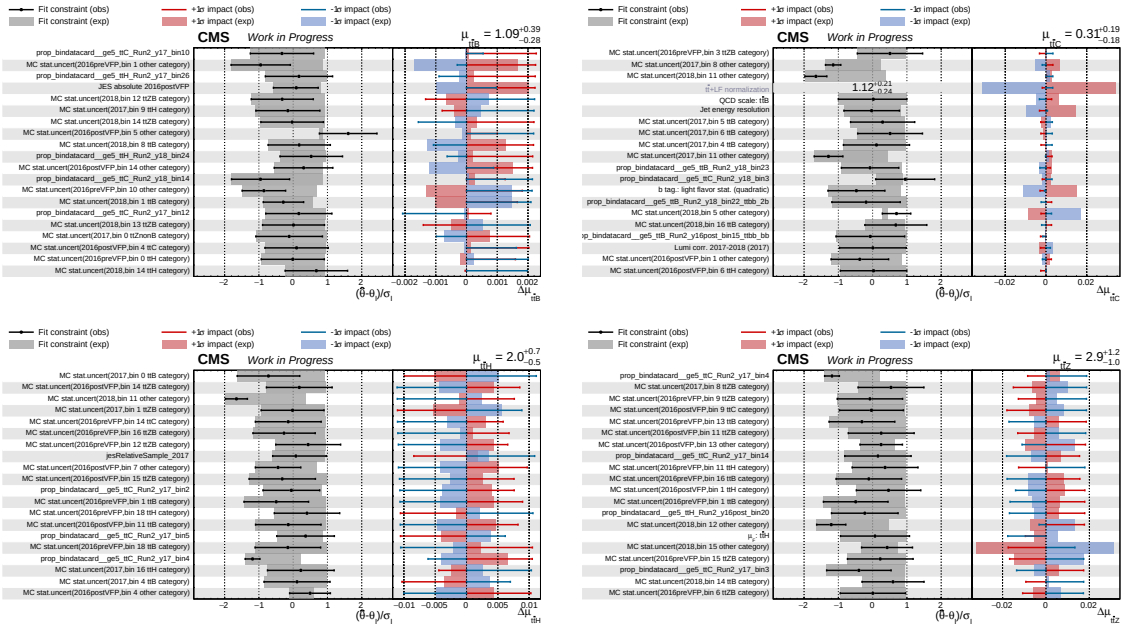


Figure L.26: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the combined Run-II dataset. The middle panels show the normalized displacement $(\hat{\theta}-\theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (176-195) are ordered according to the magnitude of their impact.

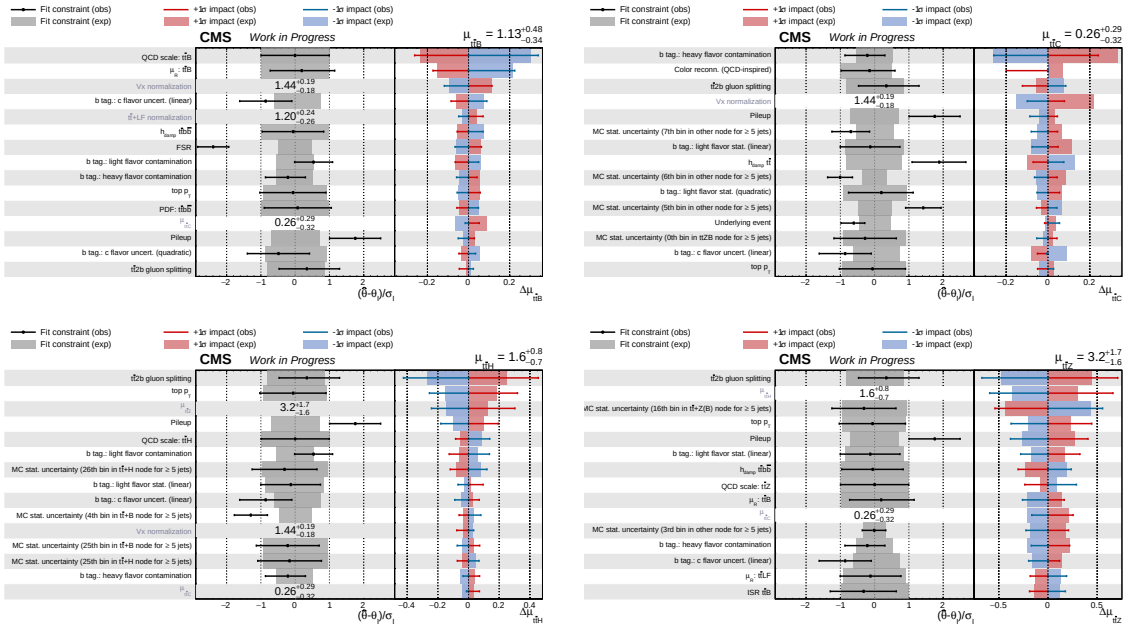


Figure L.27: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2018 dataset. The middle panels show the normalized displacement $(\hat{\theta}-\theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (1-20) are ordered according to the magnitude of their impact.

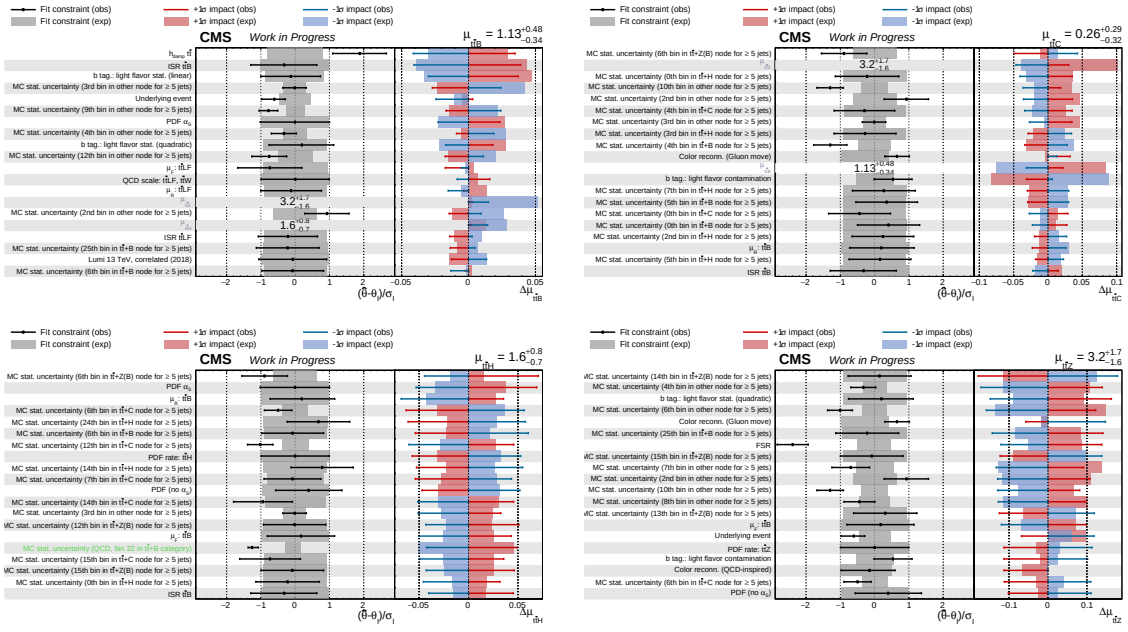


Figure L.28: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2018 dataset. The middle panels show the normalized displacement $(\hat{\theta}-\theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (21-40) are ordered according to the magnitude of their impact.

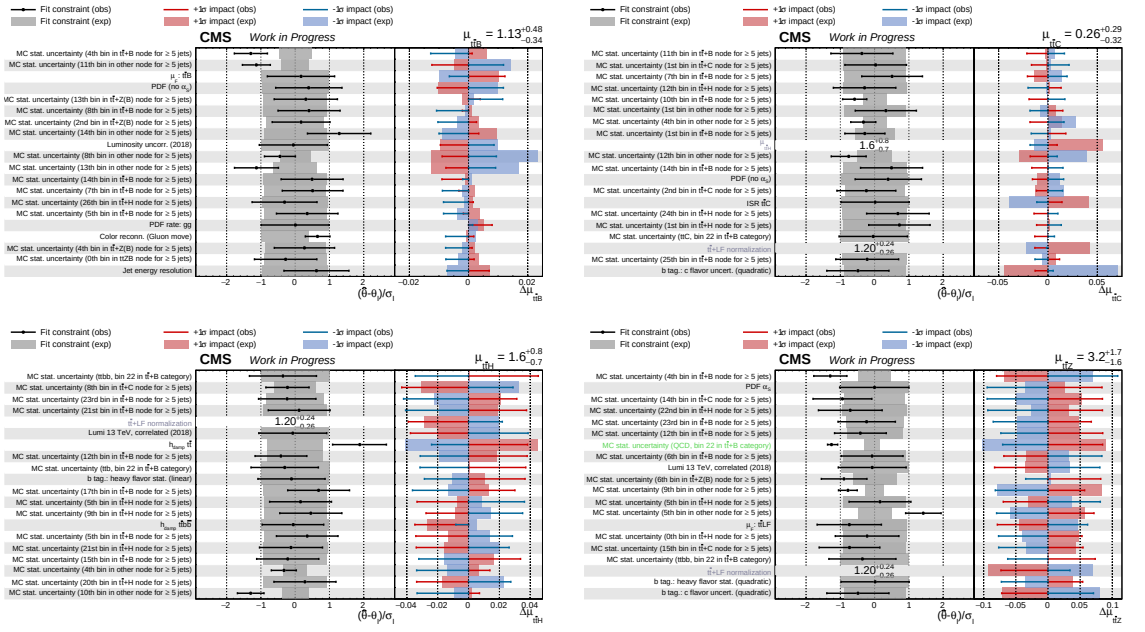


Figure L.29: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2018 dataset. The middle panels show the normalized displacement $(\hat{\theta}-\theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (41-60) are ordered according to the magnitude of their impact.

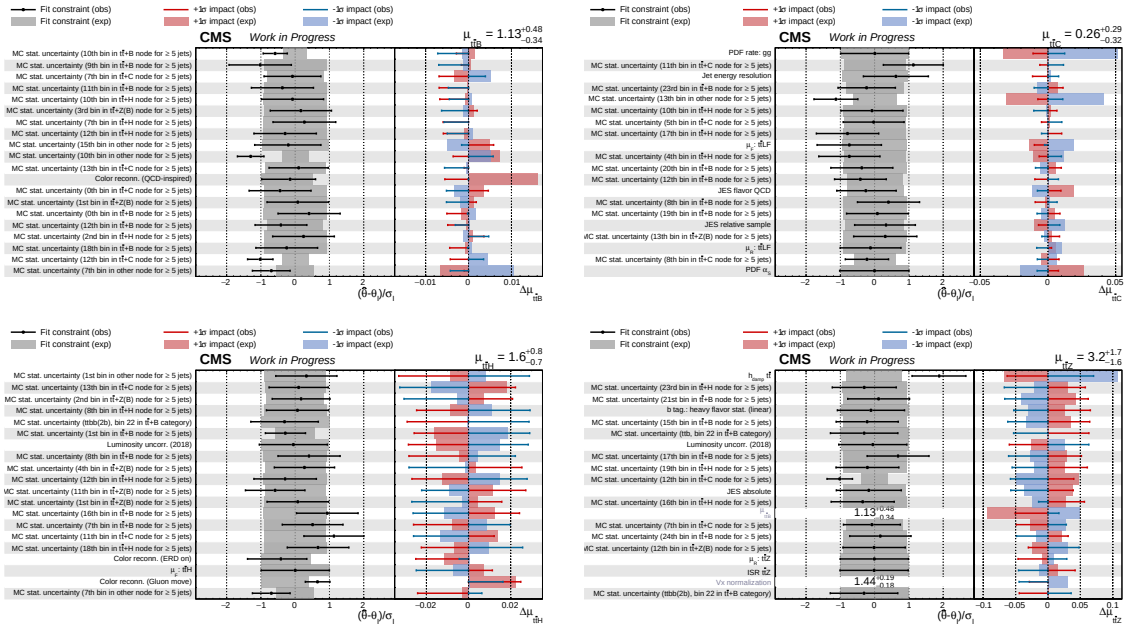


Figure L.30: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2018 dataset. The middle panels show the normalized displacement $(\hat{\theta}-\theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (61-80) are ordered according to the magnitude of their impact.

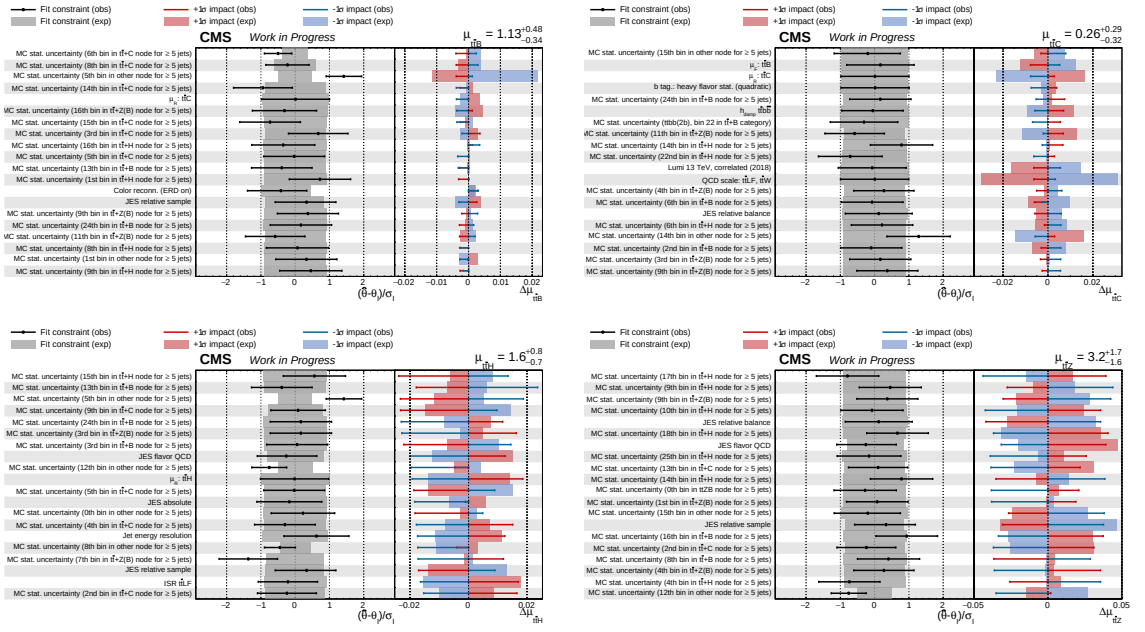


Figure L.31: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2018 dataset. The middle panels show the normalized displacement $(\hat{\theta}-\theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (81-100) are ordered according to the magnitude of their impact.

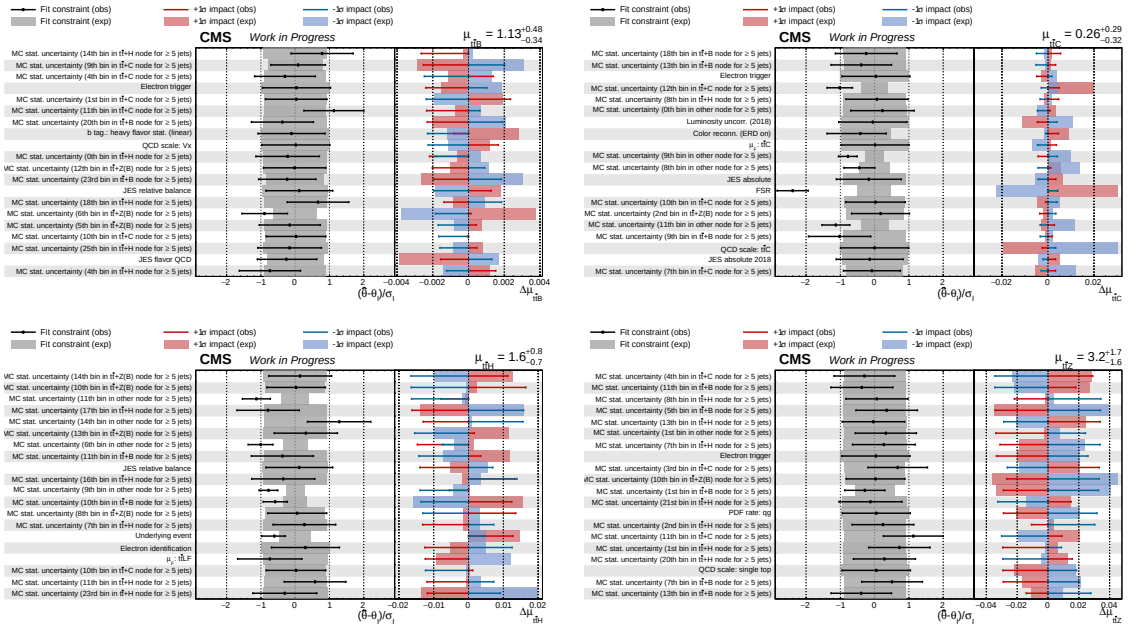


Figure L.32: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2018 dataset. The middle panels show the normalized displacement $(\hat{\theta}-\theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (101-120) are ordered according to the magnitude of their impact.

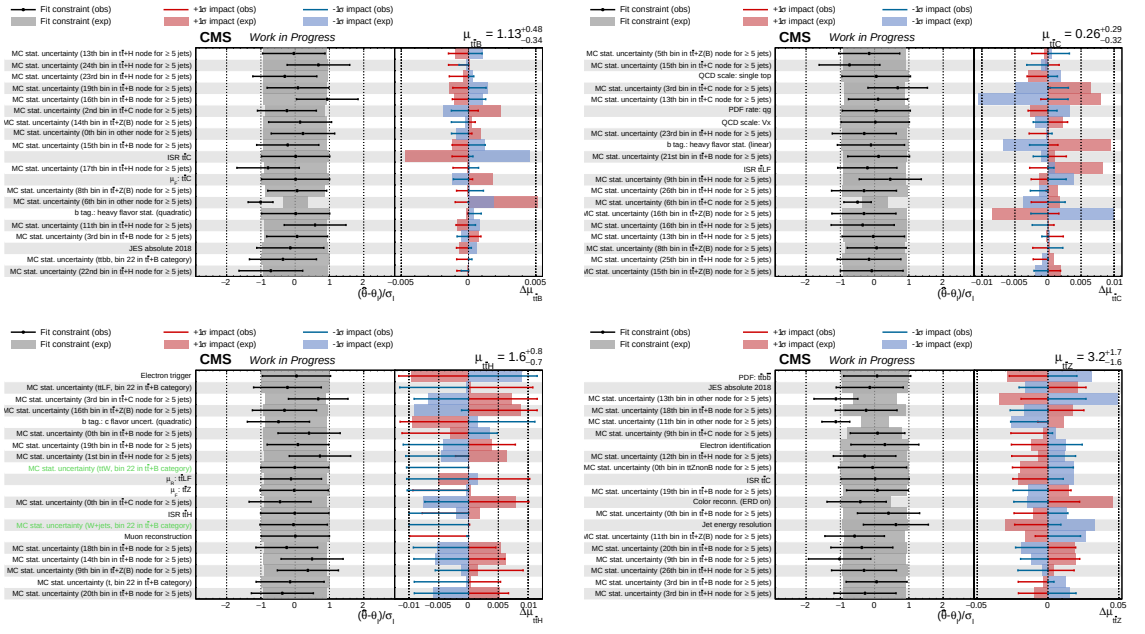


Figure L.33: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2018 dataset. The middle panels show the normalized displacement $(\hat{\theta}-\theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (121-140) are ordered according to the magnitude of their impact.

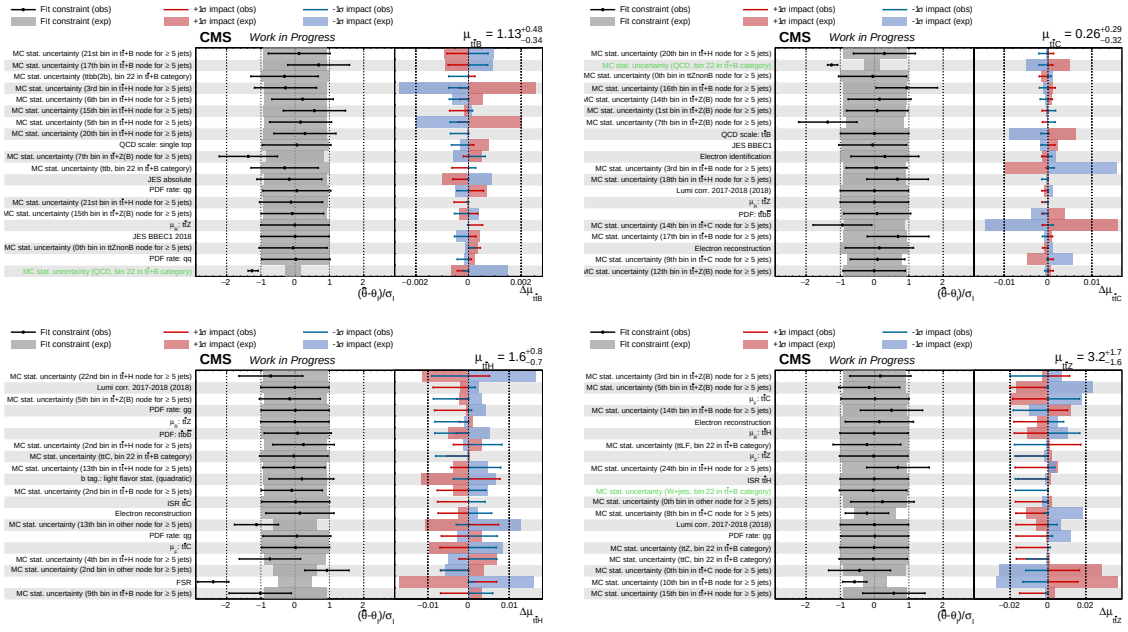


Figure L.34: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2018 dataset. The middle panels show the normalized displacement $(\hat{\theta}-\theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (141-160) are ordered according to the magnitude of their impact.

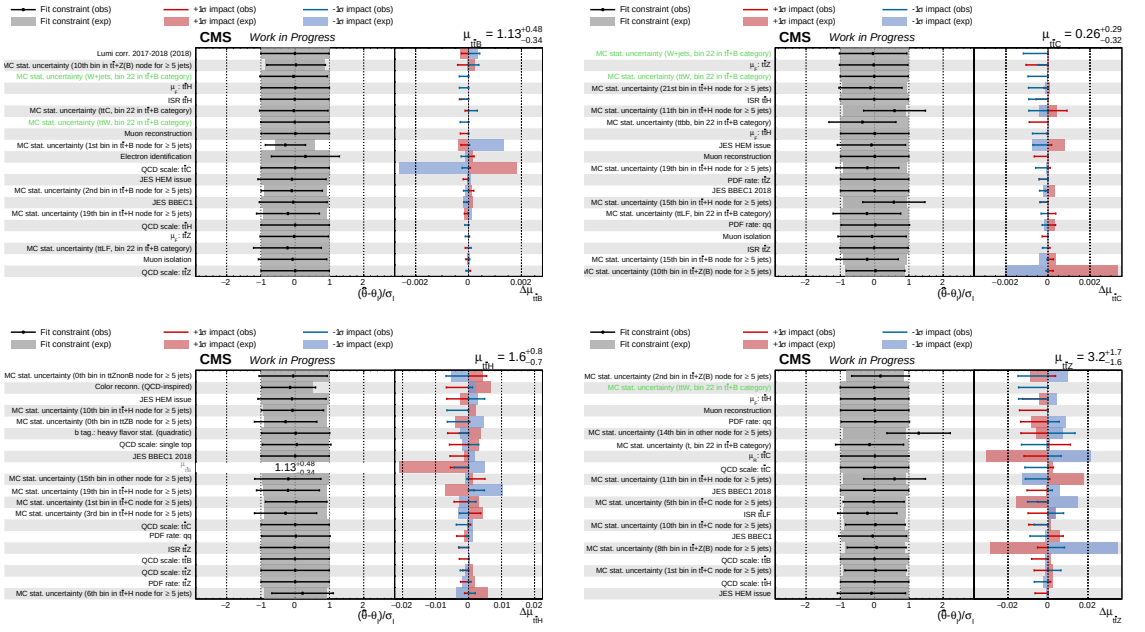


Figure L.35: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2018 dataset. The middle panels show the normalized displacement $(\hat{\theta}-\theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (161-180) are ordered according to the magnitude of their impact.

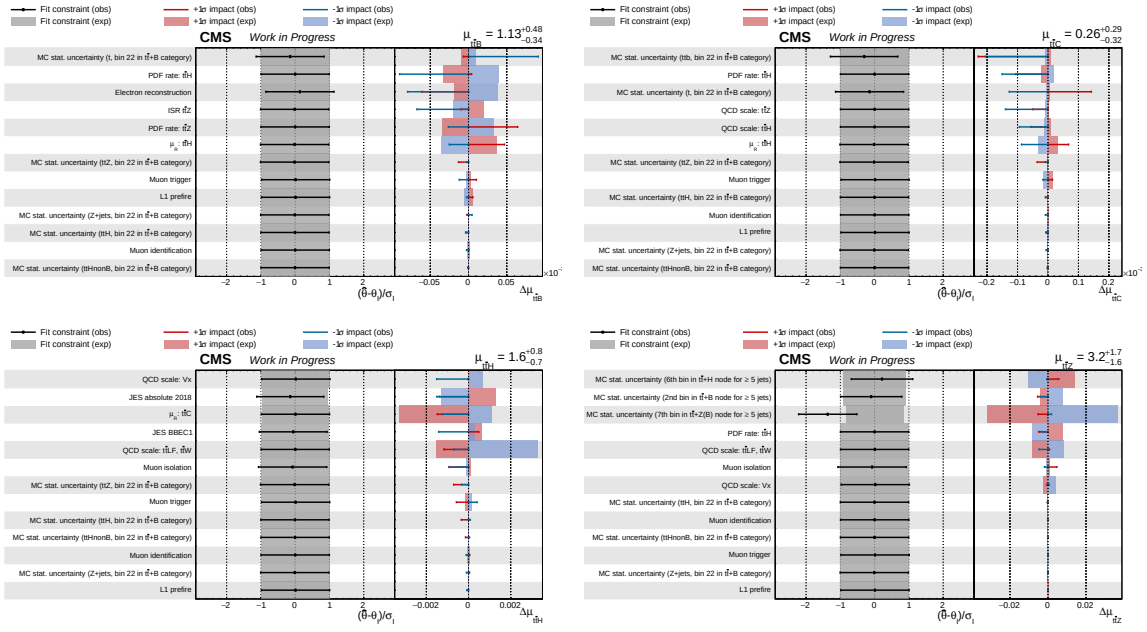


Figure L.36: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2018 dataset. The middle panels show the normalized displacement $(\hat{\theta}-\theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (181-200) are ordered according to the magnitude of their impact.

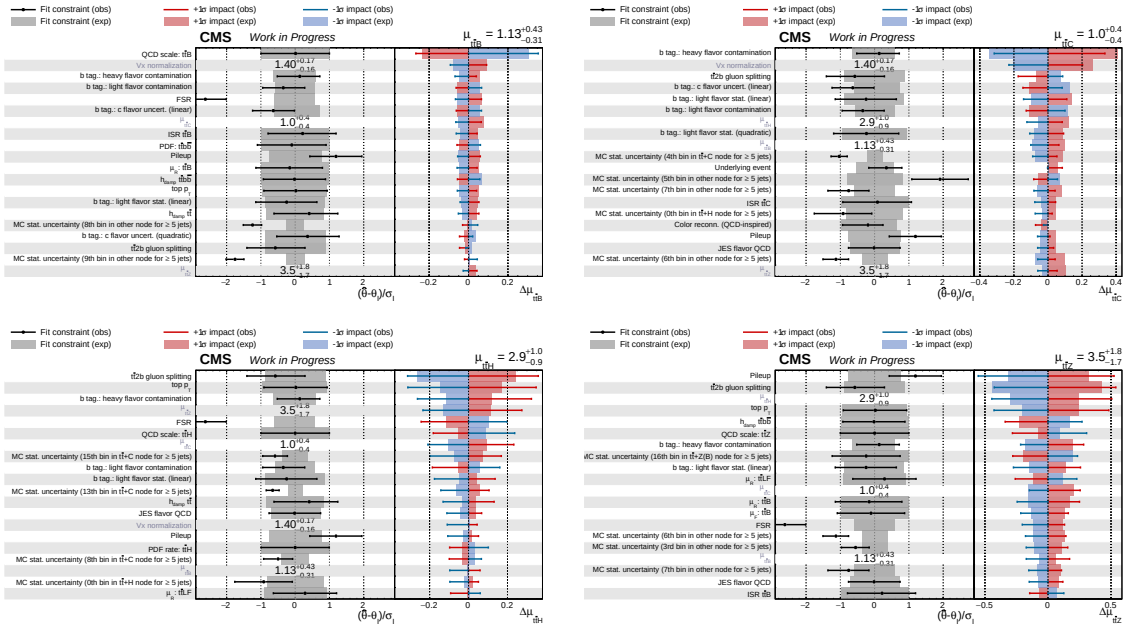


Figure L.37: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2017 dataset. The middle panels show the normalized displacement $(\hat{\theta}-\theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (1-20) are ordered according to the magnitude of their impact.

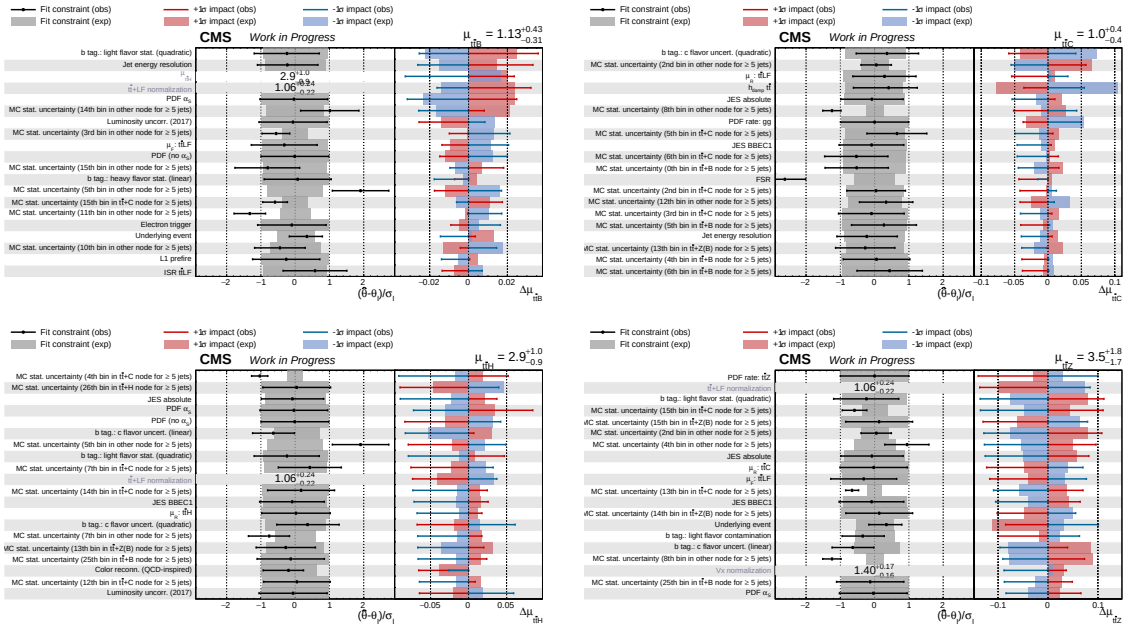


Figure L.38: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2017 dataset. The middle panels show the normalized displacement $(\hat{\theta}-\theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (21-40) are ordered according to the magnitude of their impact.

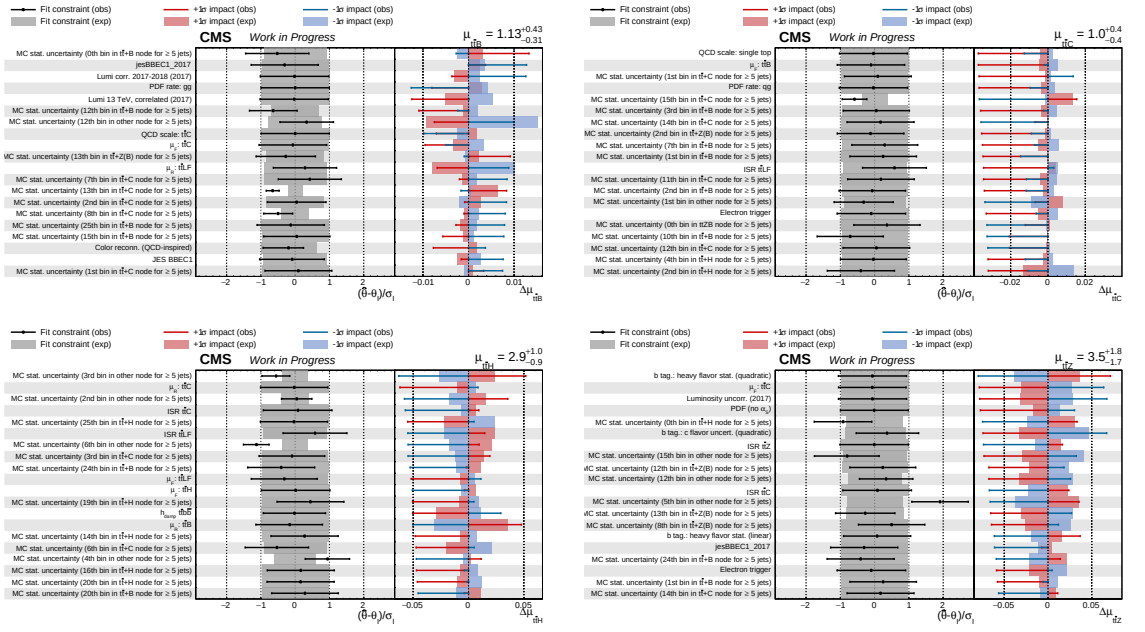


Figure L.39: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2017 dataset. The middle panels show the normalized displacement $(\hat{\theta}-\theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm 1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (41-60) are ordered according to the magnitude of their impact.

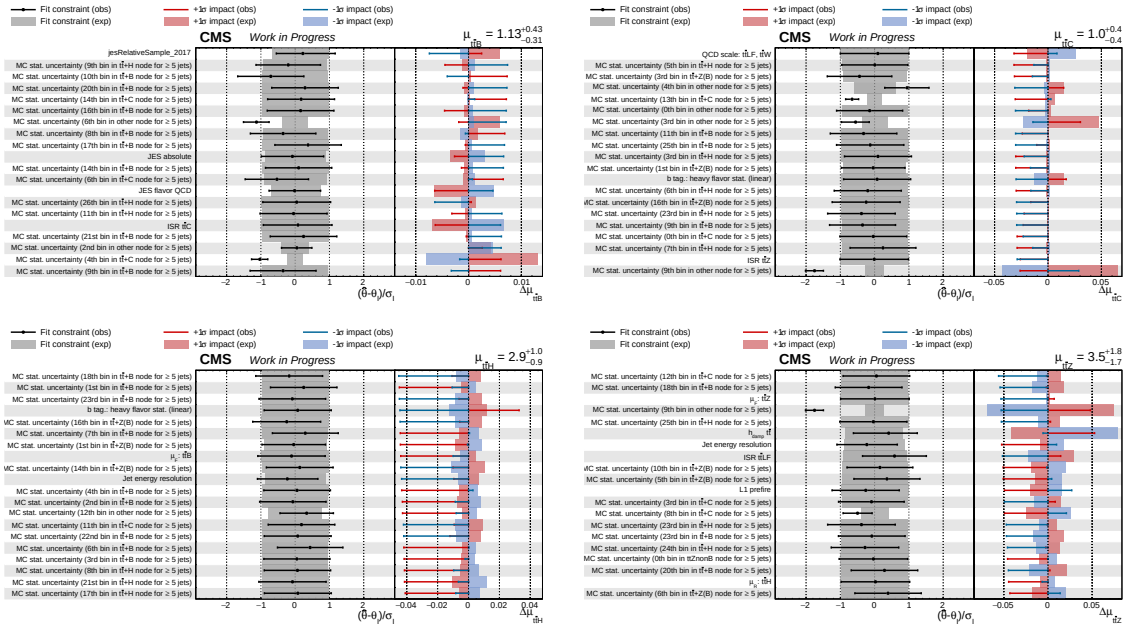


Figure L.40: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2017 dataset. The middle panels show the normalized displacement $(\hat{\theta}-\theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (61-80) are ordered according to the magnitude of their impact.

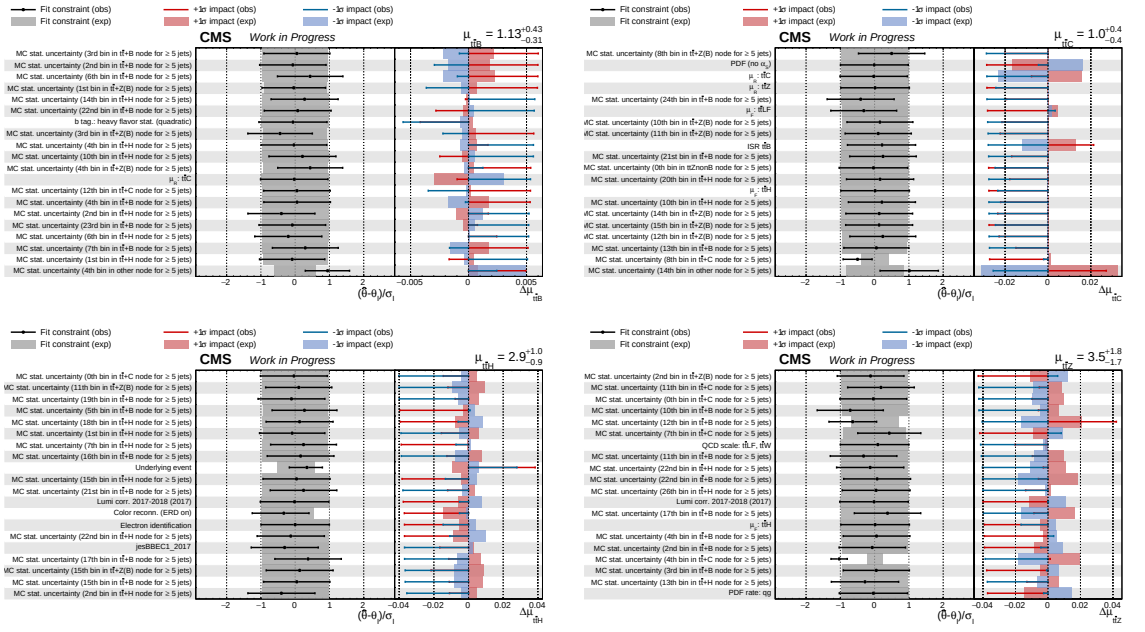


Figure L.41: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2017 dataset. The middle panels show the normalized displacement $(\hat{\theta}-\theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm 1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (81-100) are ordered according to the magnitude of their impact.

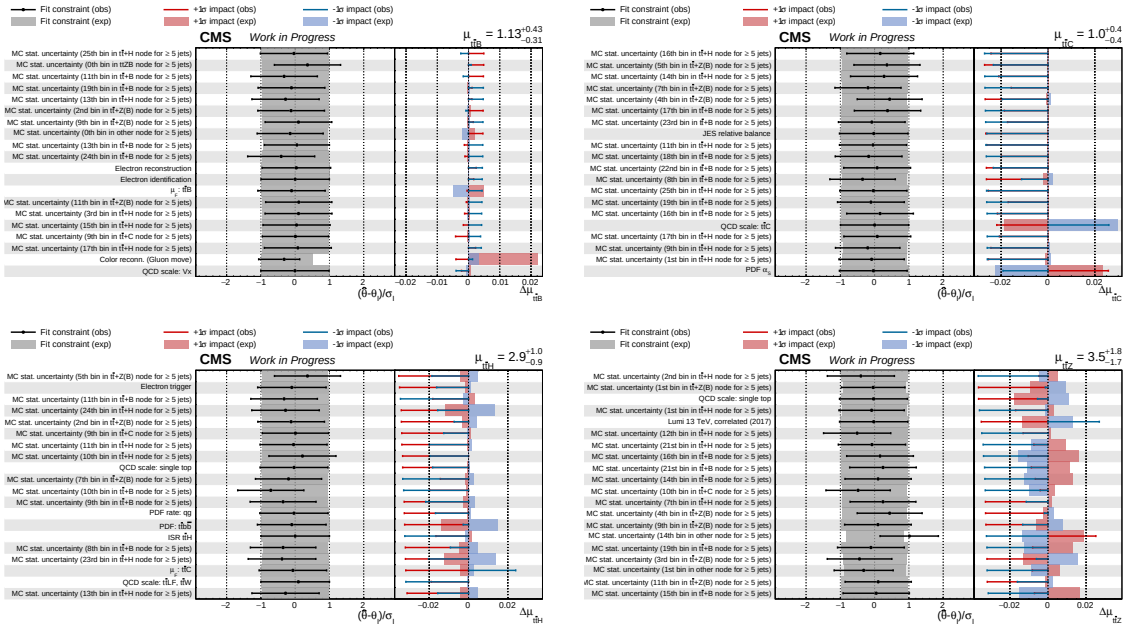


Figure L.42: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2017 dataset. The middle panels show the normalized displacement $(\hat{\theta}-\theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (101-120) are ordered according to the magnitude of their impact.

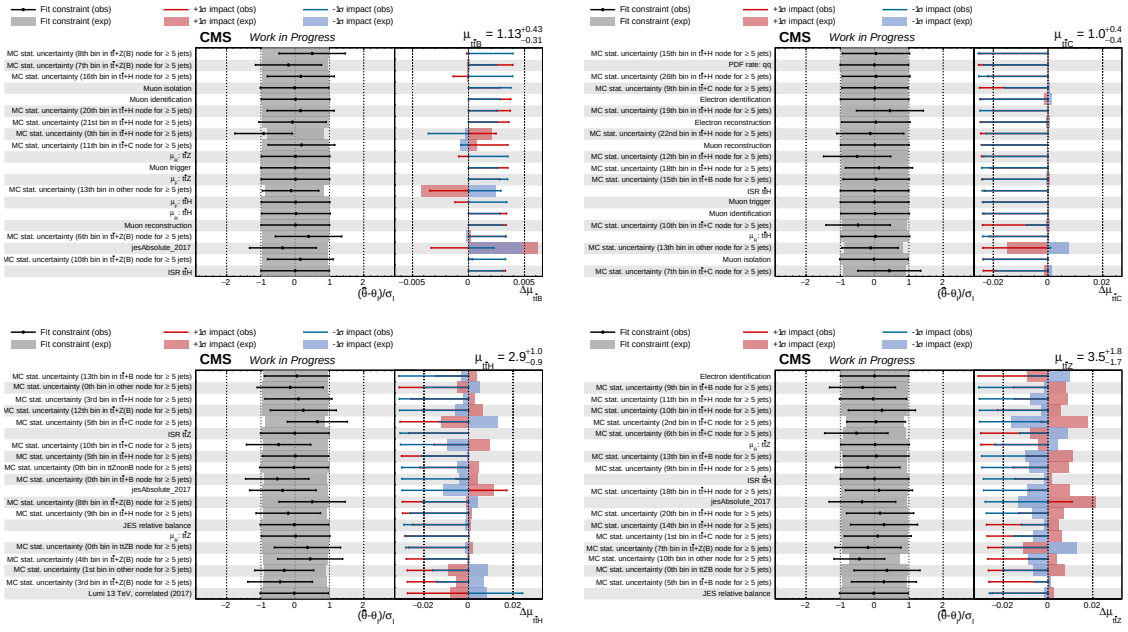


Figure L.43: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2017 dataset. The middle panels show the normalized displacement $(\hat{\theta}-\theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (121-140) are ordered according to the magnitude of their impact.

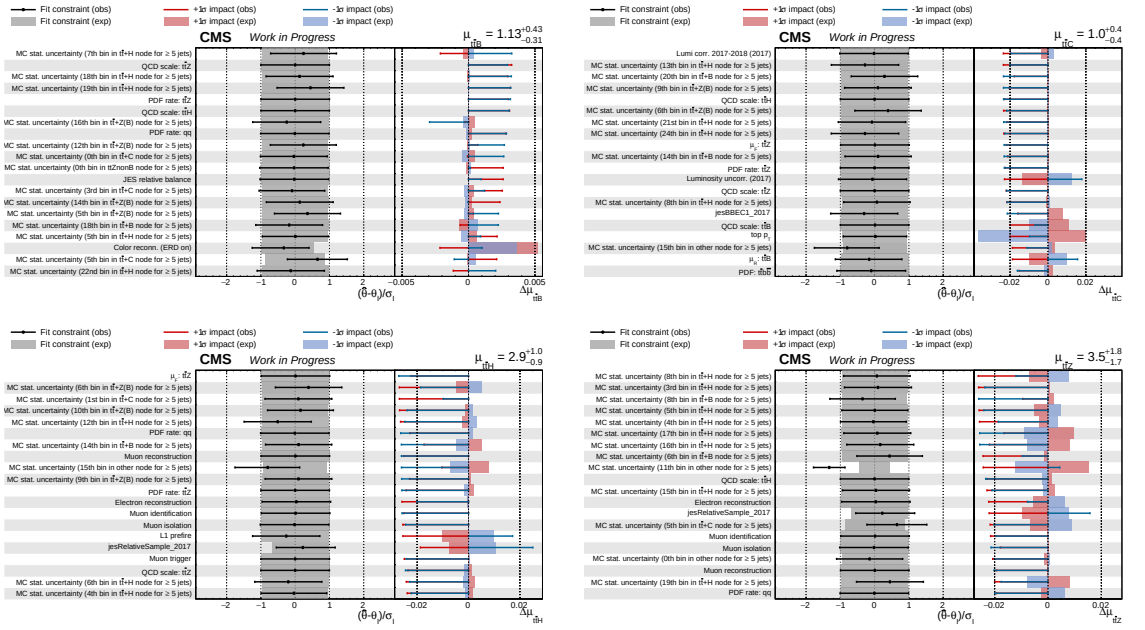


Figure L.44: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2017 dataset. The middle panels show the normalized displacement $(\hat{\theta}-\theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm 1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (141-160) are ordered according to the magnitude of their impact.

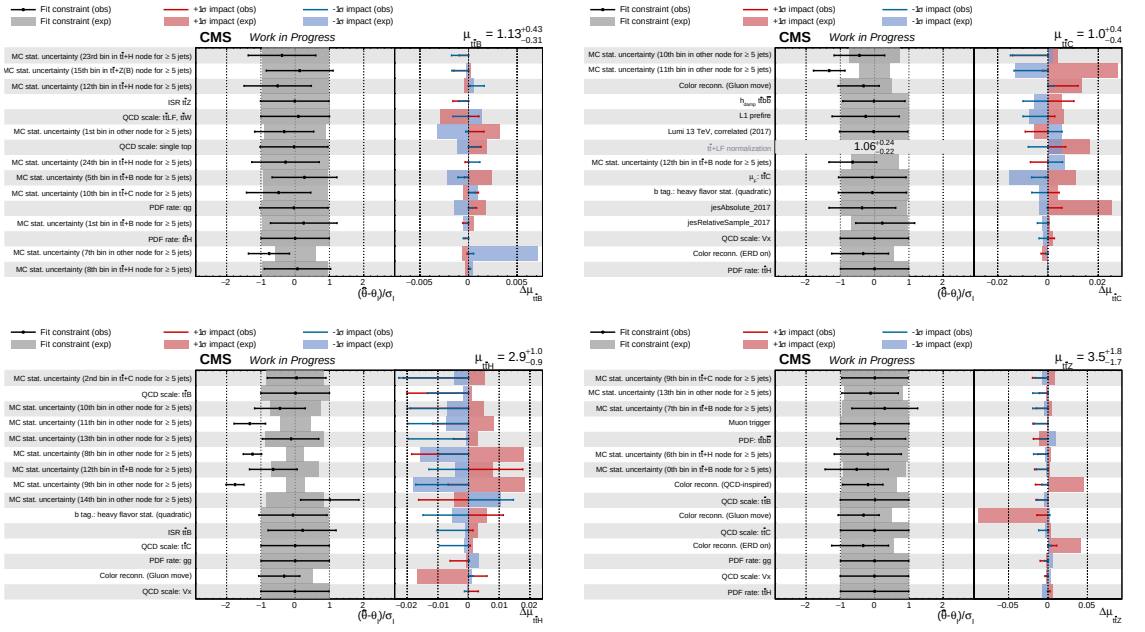


Figure L.45: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2017 dataset. The middle panels show the normalized displacement $(\hat{\theta}-\theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (161-180) are ordered according to the magnitude of their impact.

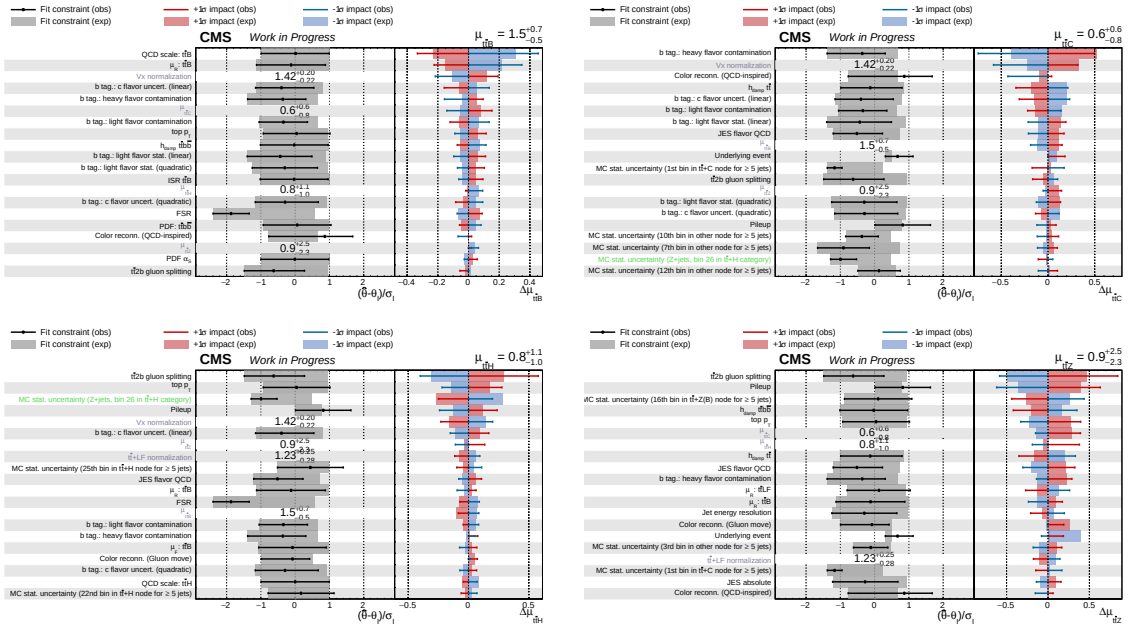


Figure L.46: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2016postVFP dataset. The middle panels show the normalized displacement $(\hat{\theta} - \theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm 1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (1-20) are ordered according to the magnitude of their impact.

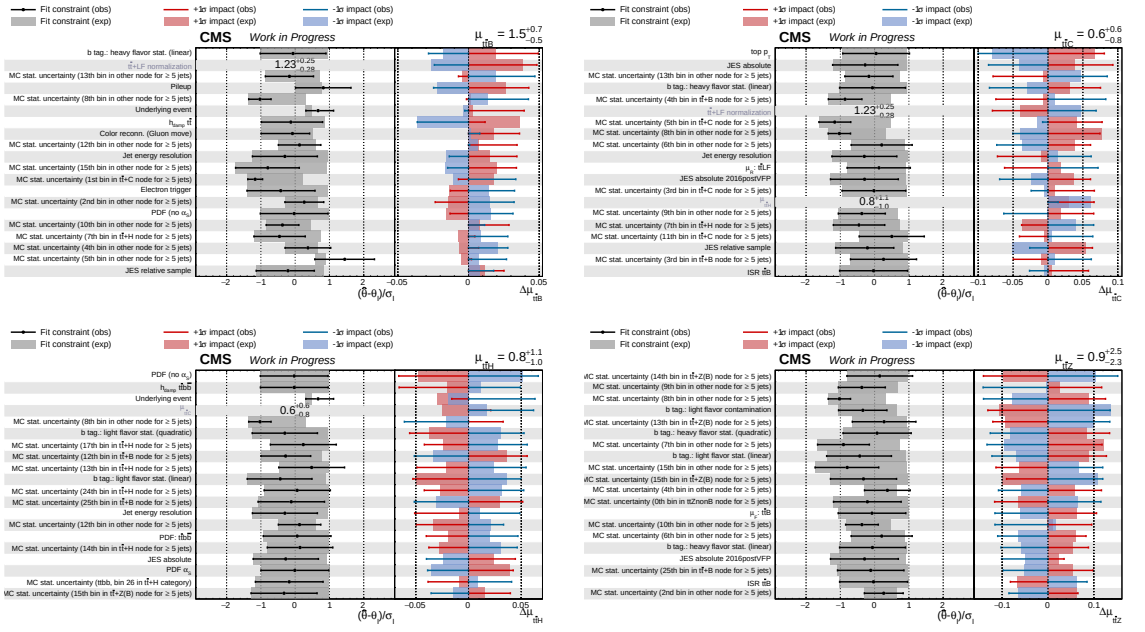


Figure L.47: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2016postVFP dataset. The middle panels show the normalized displacement $(\hat{\theta}-\theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (21-40) are ordered according to the magnitude of their impact.

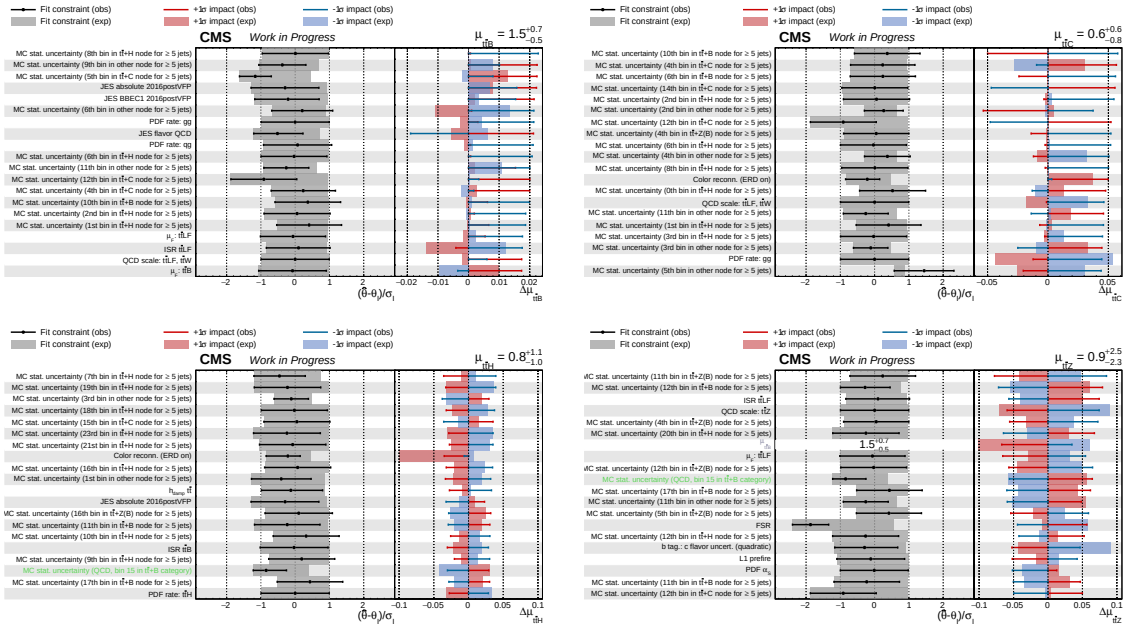


Figure L.48: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2016postVFP dataset. The middle panels show the normalized displacement $(\hat{\theta}-\theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (41-60) are ordered according to the magnitude of their impact.

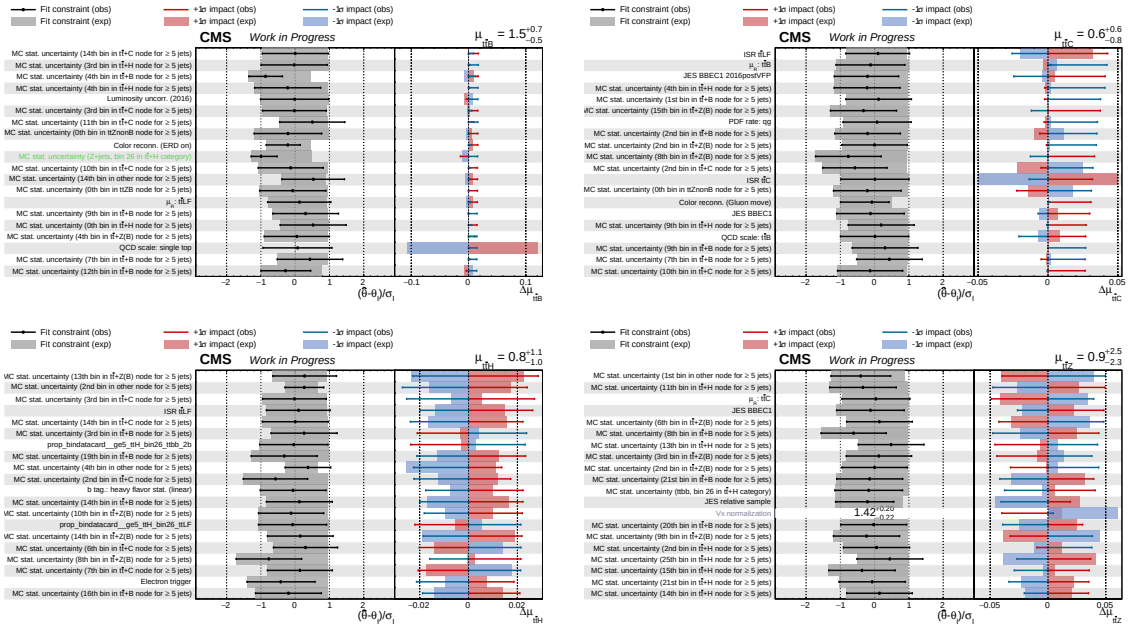


Figure L.49: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2016postVFP dataset. The middle panels show the normalized displacement $(\hat{\theta}-\theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (61-80) are ordered according to the magnitude of their impact.

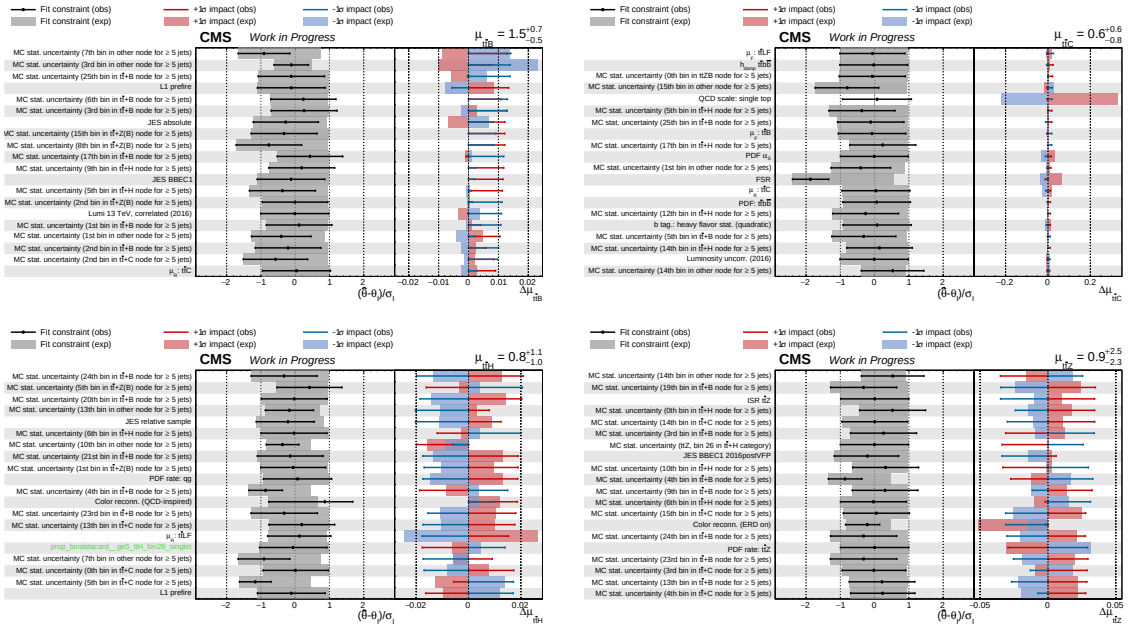


Figure L.50: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2016postVFP dataset. The middle panels show the normalized displacement $(\hat{\theta} - \theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm 1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (81-100) are ordered according to the magnitude of their impact.

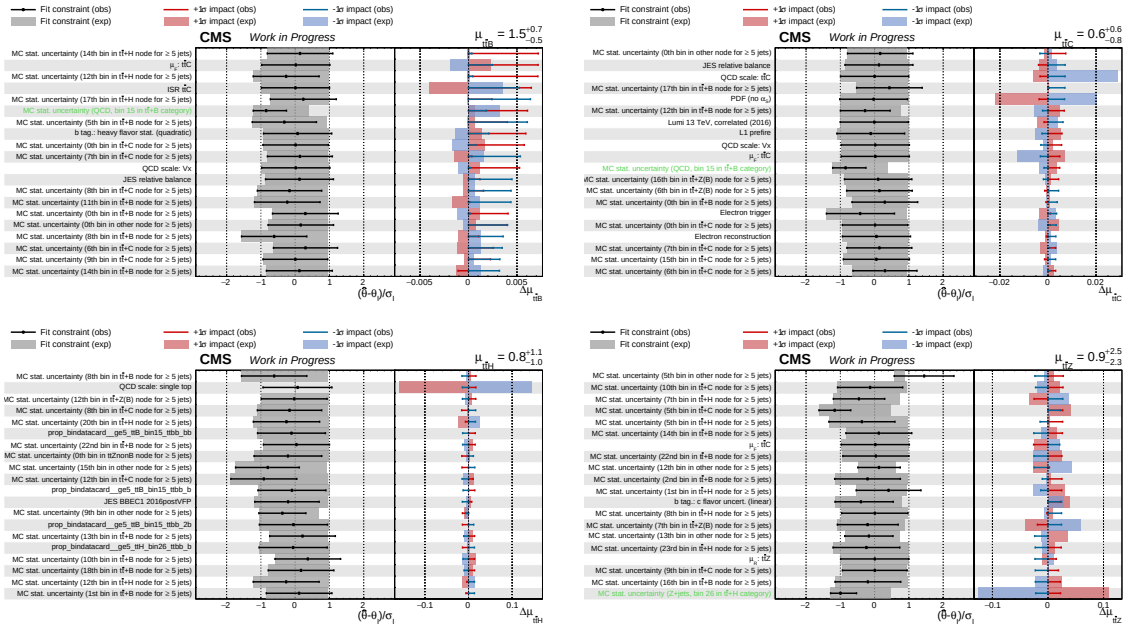


Figure L.51: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2016postVFP dataset. The middle panels show the normalized displacement $(\hat{\theta}-\theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (101-120) are ordered according to the magnitude of their impact.

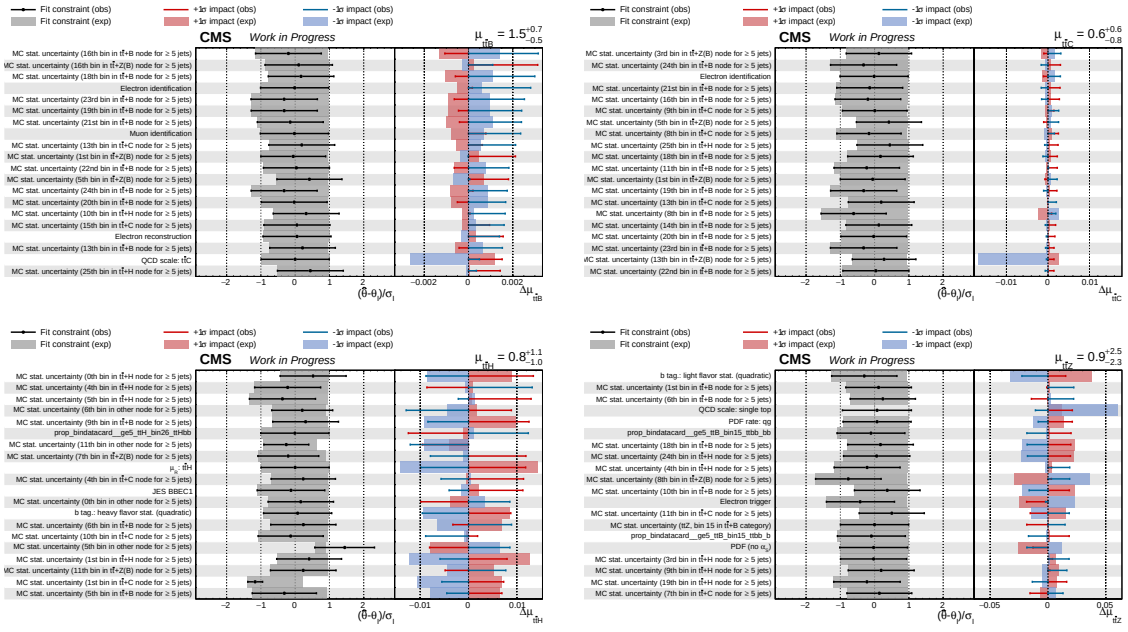


Figure L.52: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2016postVFP dataset. The middle panels show the normalized displacement $(\hat{\theta} - \theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm 1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (121-140) are ordered according to the magnitude of their impact.

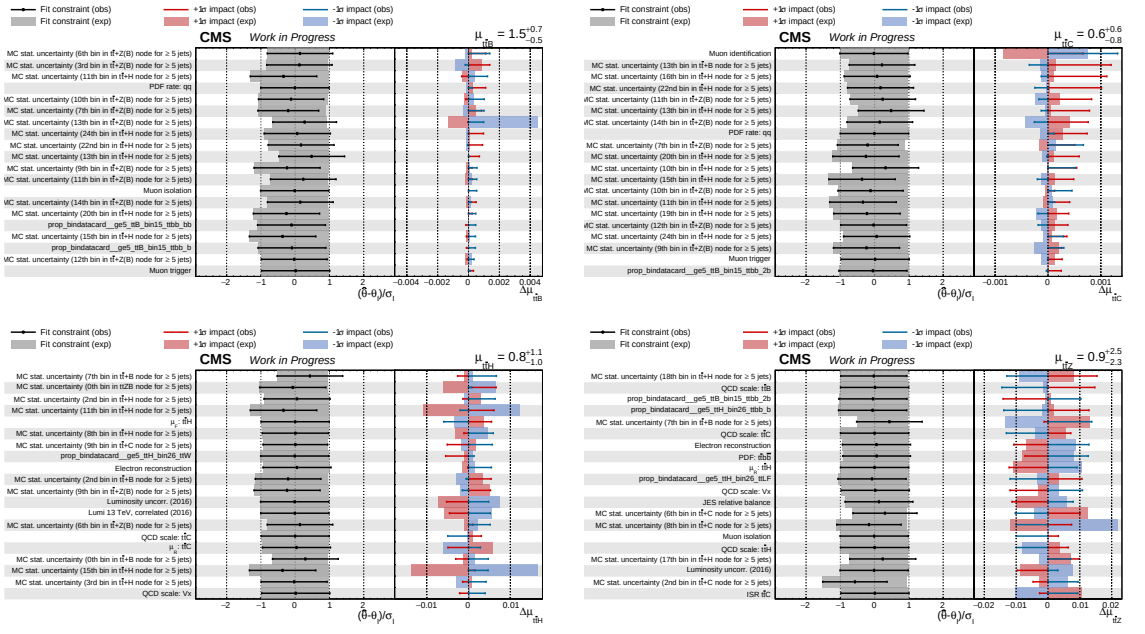


Figure L.53: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2016postVFP dataset. The middle panels show the normalized displacement $(\hat{\theta}-\theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (141-160) are ordered according to the magnitude of their impact.

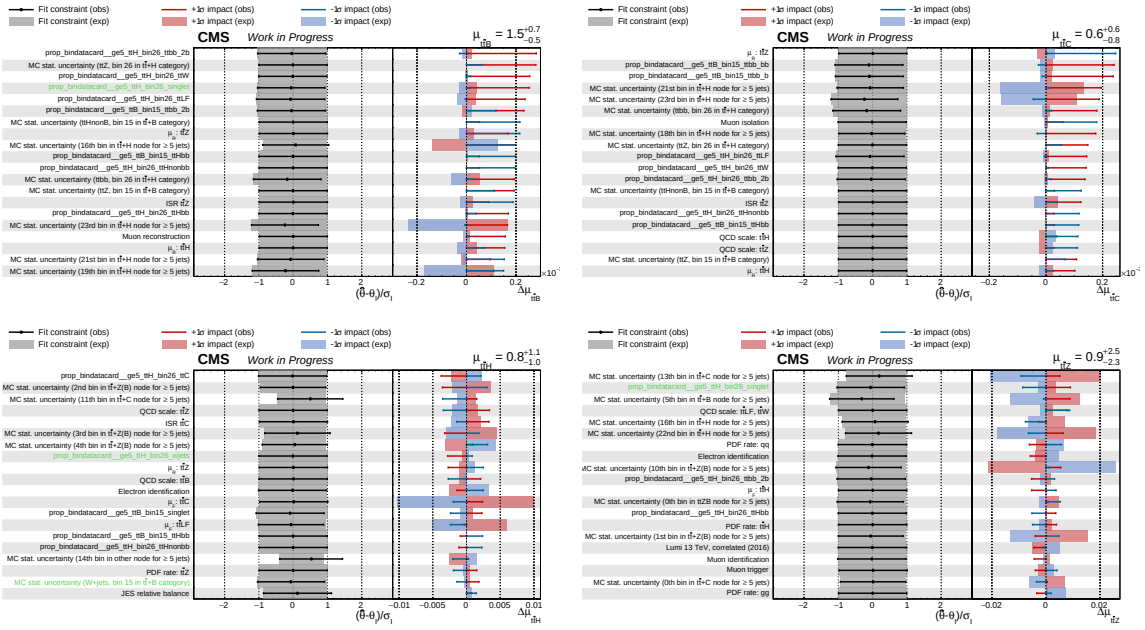


Figure L.54: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2016postVFP dataset. The middle panels show the normalized displacement $(\hat{\theta}-\theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (161-180) are ordered according to the magnitude of their impact.

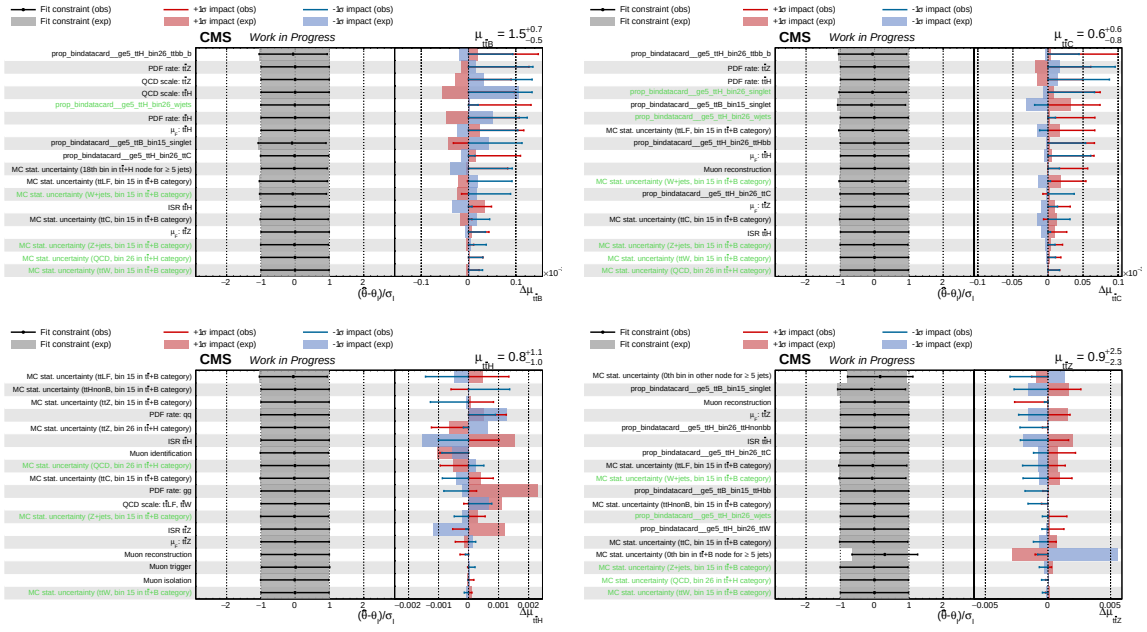


Figure L.55: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2016postVFP dataset. The middle panels show the normalized displacement $(\hat{\theta} - \theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm 1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (181-200) are ordered according to the magnitude of their impact.

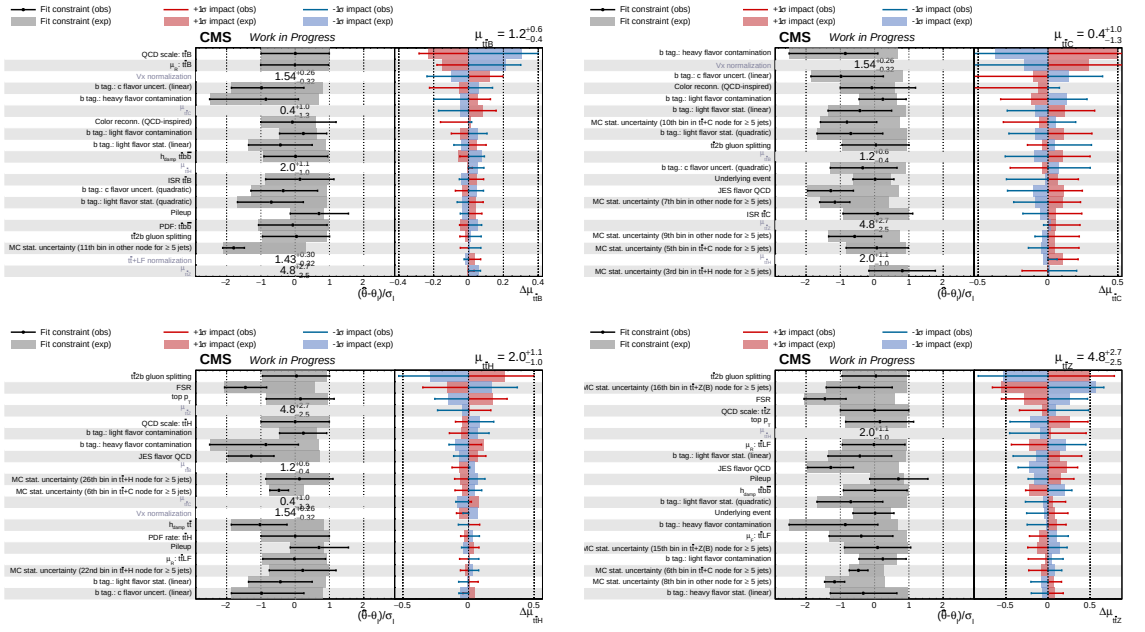


Figure L.56: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2016preVFP dataset. The middle panels show the normalized displacement $(\hat{\theta} - \theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm 1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (1-20) are ordered according to the magnitude of their impact.

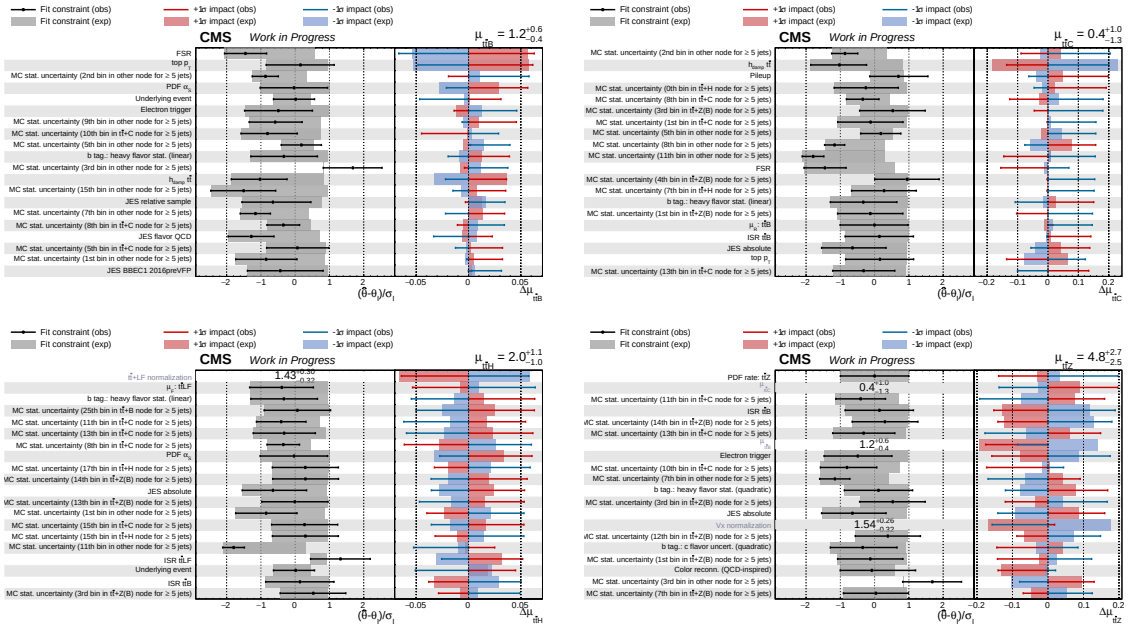


Figure L.57: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2016preVFP dataset. The middle panels show the normalized displacement $(\hat{\theta} - \theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm 1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (21-40) are ordered according to the magnitude of their impact.

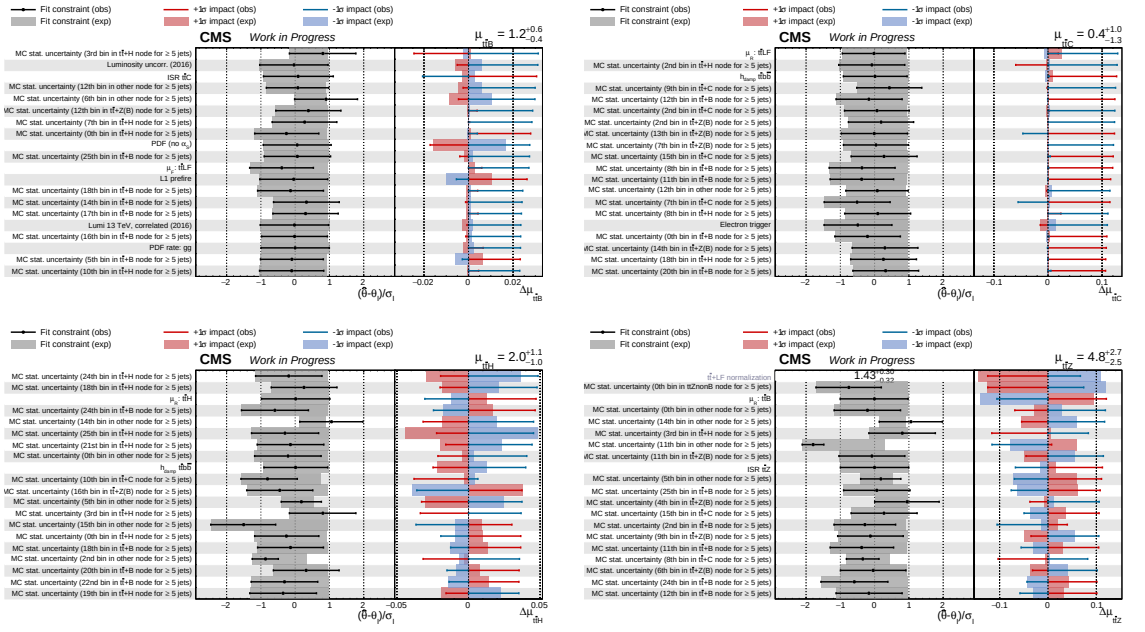


Figure L.58: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2016preVFP dataset. The middle panels show the normalized displacement $(\hat{\theta} - \theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm 1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (41-60) are ordered according to the magnitude of their impact.

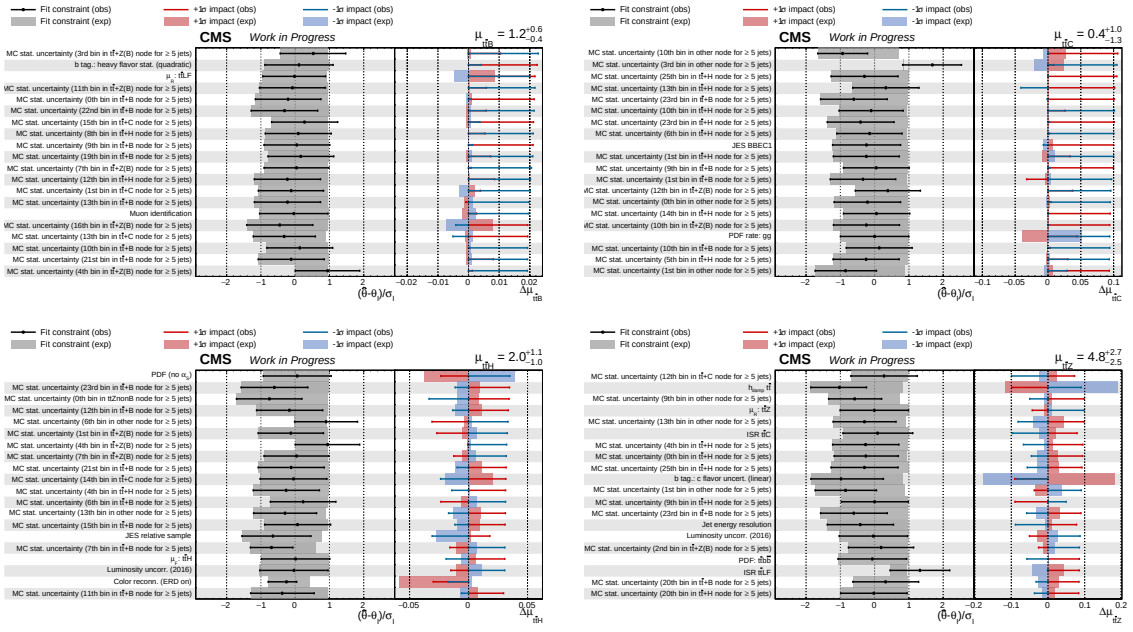


Figure L.59: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2016preVFP dataset. The middle panels show the normalized displacement $(\hat{\theta} - \theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm 1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (61-80) are ordered according to the magnitude of their impact.

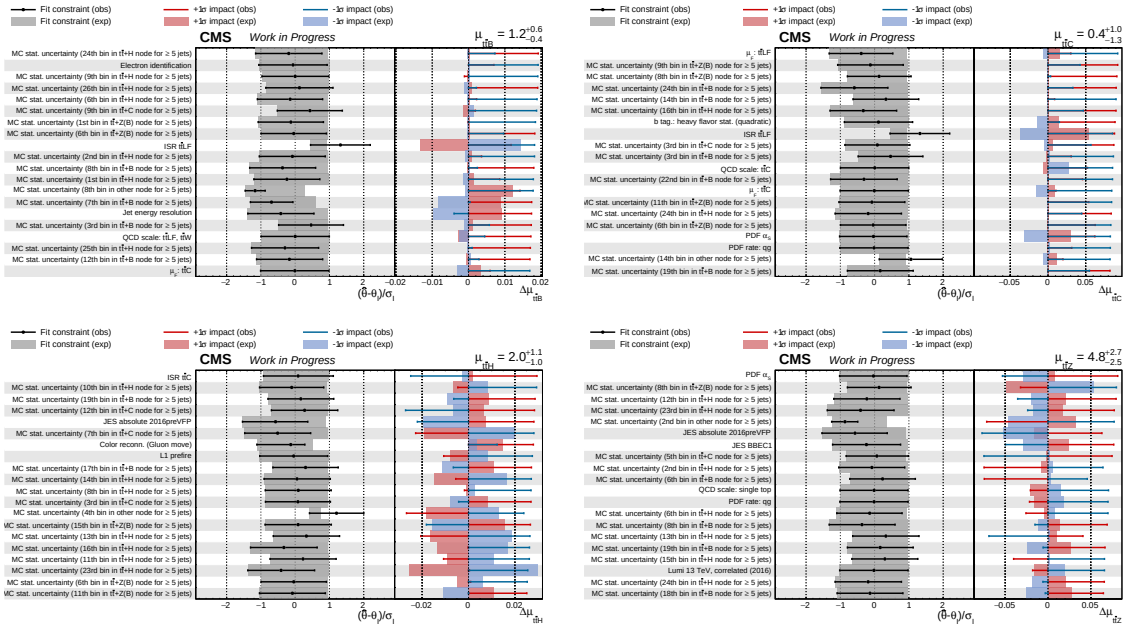


Figure L.60: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2016preVFP dataset. The middle panels show the normalized displacement $(\hat{\theta}-\theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (81-100) are ordered according to the magnitude of their impact.

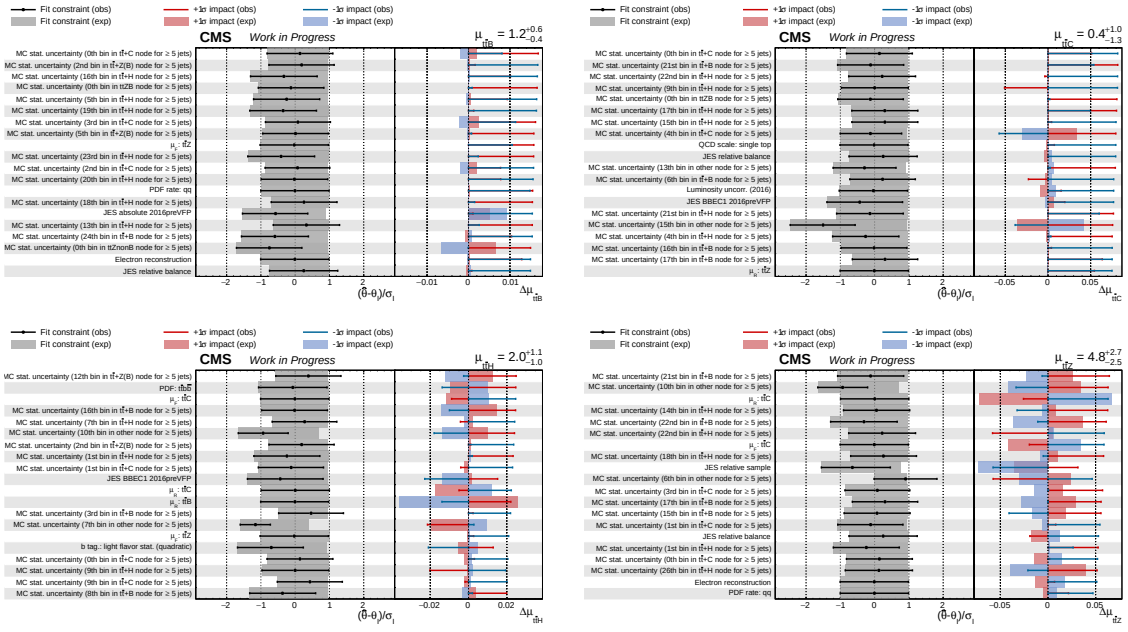


Figure L.61: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2016preVFP dataset. The middle panels show the normalized displacement $(\hat{\theta}-\theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (101-120) are ordered according to the magnitude of their impact.

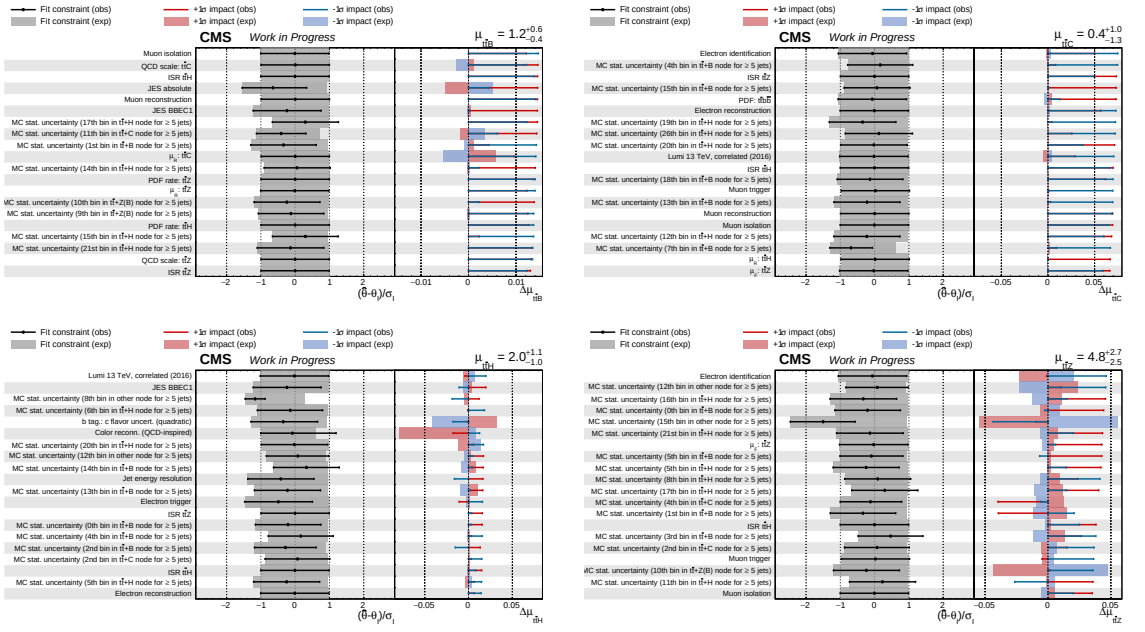


Figure L.62: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2016preVFP dataset. The middle panels show the normalized displacement $(\hat{\theta} - \theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm 1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (121-140) are ordered according to the magnitude of their impact.

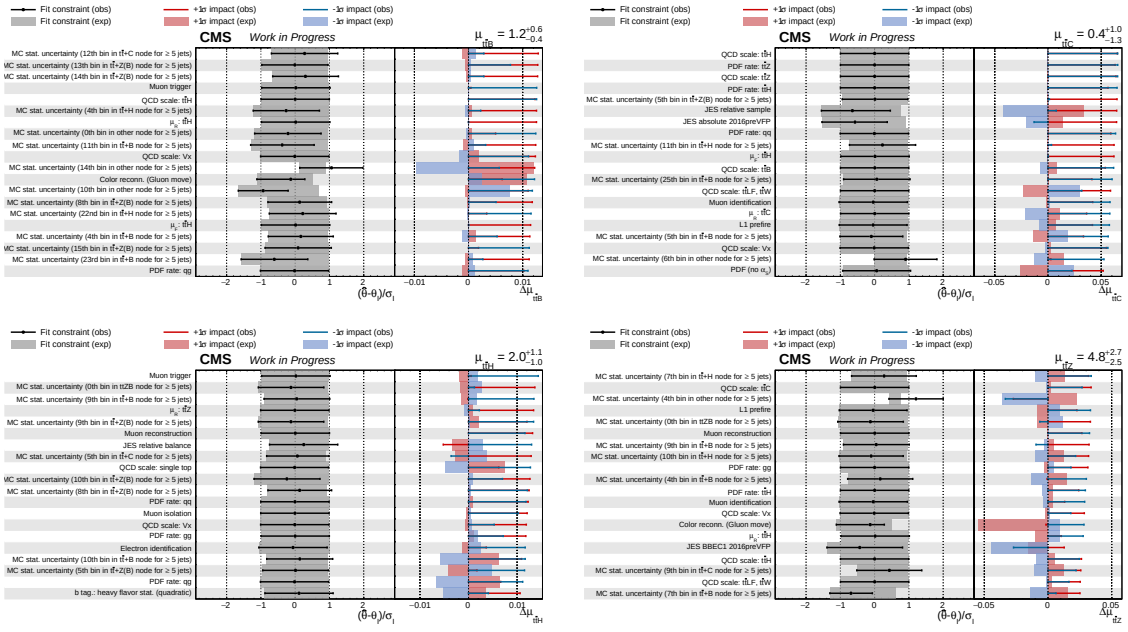


Figure L.63: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2016preVFP dataset. The middle panels show the normalized displacement $(\hat{\theta} - \theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm 1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (141-160) are ordered according to the magnitude of their impact.

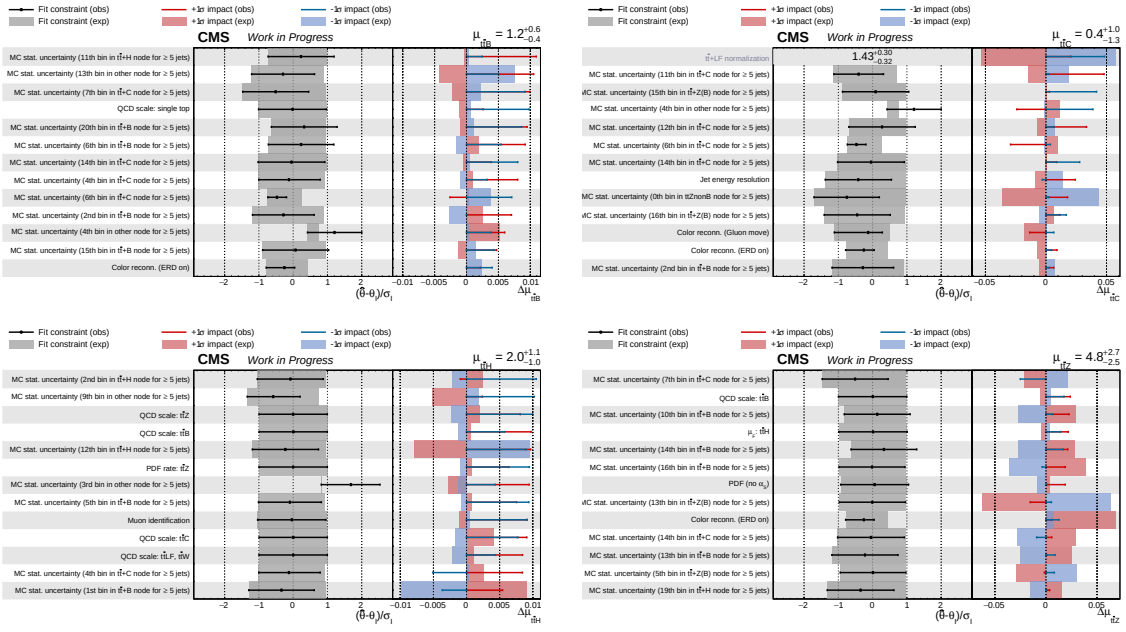


Figure L.64: Impact plots of the NPs on the signal strength modifiers $\mu_{t\bar{t}+B}$, $\mu_{t\bar{t}+C}$, $\mu_{t\bar{t}+H}$, and $\mu_{t\bar{t}+Z}$ obtained from the profile likelihood fit to the 2016preVFP dataset. The middle panels show the normalized displacement $(\hat{\theta} - \theta_1)/\sigma_1$, while the right panels display the corresponding variation induced in the POI for $\pm 1\sigma$ shifts of each nuisance parameter. Markers and solid lines represent observed results, and shaded bands indicate the expectation from the Asimov dataset. The parameters (161-180) are ordered according to the magnitude of their impact.

M Uncertainty breakdown

Table M.11: Breakdown of the uncertainties on the signal strength parameters of Run-II data. Uncertainties are obtained using the sequential nuisance group freezing procedure. Entries are shown as + up / − down. The statistical uncertainty is obtained by fixing all NPs to their post-fit values and performing the fit again. Small differences between the total uncertainty and the quadratic sum of the components arise from correlations and rounding effects. The miscellaneous category includes NPs not assigned to any other group, such as the luminosity uncertainty.

Source of uncertainty	μ_{ttB}	μ_{ttC}	μ_{ttH}	μ_{ttZ}
Experimental				
b tagging	+0.05 / -0.03	+0.10 / -0.09	+0.14 / -0.12	+0.23 / -0.23
JEC	+0.01 / -0.02	+0.08 / -0.08	+0.07 / -0.07	+0.25 / -0.19
Lepton	±0.01	+0.01 / -0.01	+0.04 / -0.04	+0.12 / -0.12
Theory				
QCD scale	+0.34 / -0.26	+0.04 / -0.04	+0.21 / -0.12	+0.38 / -0.26
tt2b gluon splitting	+0.04 / -0.02	+0.07 / -0.06	+0.30 / -0.25	+0.29 / -0.32
PS	+0.09 / -0.05	±0.05	+0.06 / -0.05	+0.21 / -0.22
h_{damp}	+0.06 / -0.05	+0.05 / -0.04	±0.01	+0.33 / -0.33
top p_T correction	+0.04 / -0.02	+0.05 / -0.05	+0.27 / -0.24	+0.33 / -0.23
PDF	±0.01	±0.01	+0.18 / -0.17	+0.32 / -0.30
Color reconnection	±0.00	±0.01	±0.00	+0.04 / -0.05
Miscellaneous				
Miscellaneous	±0.02	+0.02 / -0.02	+0.06 / -0.06	+0.24 / -0.23
Other POIs	±0.01	±0.02	±0.10	+0.35 / -0.34
Statistical	±0.02	±0.02	+0.21 / -0.28	±0.48
Total	+0.39 / -0.28	+0.19 / -0.18	+0.65 / -0.55	+1.18 / -1.04

Table M.12: Breakdown of the uncertainties on the signal strength parameters of the 2018 era. Uncertainties are obtained using the sequential nuisance group freezing procedure.

Source of uncertainty	μ_{ttB}	μ_{ttC}	μ_{ttH}	μ_{ttZ}
Experimental				
b tagging	+0.07 / 0.04	+0.18 / 0.20	+0.25 / 0.23	+0.33 / 0.33
JEC	±0.03	+0.07 / 0.06	+0.06 / 0.04	+0.34 / 0.28
Lepton	±0.01	+0.02 / 0.02	+0.12 / 0.11	+0.52 / 0.50
Theory				
QCD scale	+0.41 / 0.30	+0.07 / 0.06	+0.20 / 0.11	+0.52 / 0.37
tt2b gluon splitting	±0.00	+0.13 / 0.16	+0.41 / 0.35	+0.52 / 0.50
PS	+0.08 / 0.06	+0.04 / 0.05	+0.03 / 0.03	+0.28 / 0.28
h_{damp}	+0.11 / 0.06	+0.06 / 0.06	+0.05 / 0.05	+0.48 / 0.48
top p_T correction	+0.10 / 0.06	+0.07 / 0.11	+0.32 / 0.23	+0.48 / 0.30
PDF	+0.07 / 0.06	+0.04 / 0.04	+0.27 / 0.27	+0.71 / 0.68
Color reconnection	±0.00	±0.00	+0.04 / 0.05	±0.05
Miscellaneous				
Miscellaneous	±0.00	+0.02 / 0.02	+0.07 / 0.09	+0.38 / 0.39
Other POIs	±0.01	±0.00	+0.09 / 0.08	±0.00
Statistical	±0.01	±0.02	+0.25 / 0.25	+0.56 / 0.57
Total	+0.48 / 0.34	+0.29 / 0.31	+0.74 / 0.64	+1.58 / 1.46

Acknowledgments

First and foremost, I would like to express my gratitude to my supervisor, Prof. Ulrich Husemann, for his guidance, encouragement, and continuous support throughout my PhD. His insight and constructive feedback have greatly contributed to the development of this work and to my growth as a researcher. I am particularly grateful for his prompt and thoughtful feedback on draft chapters of this thesis. Despite his demanding responsibilities as Director of particle physics at DESY, he consistently made time for discussions and valuable feedback, which I greatly appreciate. I would also like to thank Prof. Jan Kieseler for kindly agreeing to serve as my second supervisor and for his support. My thanks also go to Prof. Bharat Kumar, Prof. Rajeev Joshi, Prof. Deepak Samuel, and Dr. Javier Brochero for their inspiration and guidance throughout my Master's studies, without their encouragement, this journey would not have been possible.

I gratefully acknowledge the financial support provided by the German Academic Exchange Service (DAAD), which made this research possible. I would also like to acknowledge the Karlsruhe School of Elementary Particle and Astroparticle Physics (KSETA) and the Institute of Experimental Particle Physics (ETP) at KIT for their support and for providing a stimulating research environment to pursue my PhD. My thanks also go to the secretariat, Diana Fellner, Bärbel Bräunling, and Chen Peichen, for their reliable support in organizational matters. The work presented in this thesis was carried out within the CMS collaboration at CERN, whose members, together with the engineers, technicians, and support staff, make analyses such as this possible through their efforts in designing, operating, and maintaining the detector and computing infrastructure.

The journey towards completing this thesis was not always straightforward, it came with its share of challenges, moments of uncertainty, and periods of adjustment. Moving to a new country and starting a new life in a different academic environment was both exciting and challenging, and it took time to adapt and find my footing. I am grateful to Ulrich for fostering

an open and supportive atmosphere within the group, which made it possible to pursue research in a collaborative environment. I would like to thank my colleagues for their support and for the many discussions that helped me progress throughout my PhD. I am particularly grateful to Emanuel, whose earlier work formed an important foundation for my analysis and helped me continue on this project. I also acknowledge the work, particularly frameworks developed for the smooth flow of analysis, by Jan, Michael, and Sebastian, which were used in this work. I am thankful to my colleagues Nikita, Moritz, and Kylian for the pleasant working atmosphere and for our discussions along the way. A special mention goes to Nicolò, who joined the group towards the later stage of my PhD but was always willing to help whenever I had questions, including many technical ones. His readiness to discuss problems and offer support during difficult moments was greatly appreciated. I also enjoyed many group activities outside of work, such as group dinners, bowling, mini-golf, and archery, which added many memorable moments to my PhD years.

The years of my doctorate were not only a period of scientific development but also a time of significant personal change. Like many long journeys, it included moments that were challenging and required resilience and reflection. During those times, the support and presence of friends became especially meaningful. The many trips, conversations, sleepovers, and moments spent together reminded me that I was never alone and helped me maintain perspective beyond the challenges of research and everyday life. I am particularly thankful to my close friends Amrutha, Manju, and Shiva, who have been a constant presence in my life since my Master's studies in India. The many trips we shared together, the lively discussions, and countless small moments along the way, including recurring debates about who should cook during our trips, created memories that I always cherish. I am also thankful for the friends I made after moving to Karlsruhe, whose companionship made my time here far more enjoyable. In particular, I would like to thank Pavi, Vani, Milin, Niyathi, and Rose for their friendship, the many conversations we shared, and the moments, including the many fun reels we recorded, that added joy and balance to these years. I would also like to thank Alex for his companionship during the final months of writing this thesis. Sharing many late evenings spent working together, along with the frequent takeouts and dinners outside, made the long writing process a little easier. I also thank everyone who supported me during this journey, but whom I may have unintentionally not mentioned by name.

Finally, I would like to thank my parents, Rafeek and Shameema, for their love and support, and my siblings, Raees, Shifa, and Aamir, whose occasional calls, often sharing their doubts and adventures, always reminded me of home and made this achievement possible.