Search for BSM Higgs bosons in final states with b-quarks with the full Run 2 CMS dataset

Dissertation

zur Erlangung des Doktorgrades

an der Fakultät für Mathematik,

Informatik und Naturwissenschaften

Fachbereich Physik
der Universität Hamburg

vorgelegt von
Daina Leyva Pernía
aus
Havanna, Kuba

Hamburg
2024

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

I hereby declare, on oath, that I have written the present dissertation by my own and have not used other than the acknowledged resources and aids.

Hamburg, den 3 Juni 2024	Daina Leyva Pernía

Gutachter der Dissertation:	Dr. Roberval Walsh Prof. Dr. Elisabetta Gallo
Zusammensetzung der Prüfungskommission:	Dr. Roberval Walsh Prof. Dr. Elisabetta Gallo Prof. Dr. Sven-Olaf Moch Prof. Dr. Johannes Haller Prof. Dr. Christian Schwanenberger
Vorsitzender der Prüfungskommission:	Prof. Dr. Sven-Olaf Moch
Datum der Disputation:	2 Juli 2024
Vorsitzender des Fach-Promotionsausschusses Physik:	Prof. Dr. Markus Drescher
Leiter des Fachbereichs Physik:	Prof. Dr. Wolfgang J. Parak
Dekan der Fakultät MIN:	Prof. Dr.-Ing. Norbert Ritter

Abstract

This thesis presents the search for heavy neutral Higgs bosons, produced in association with b-quarks and decaying into a b-quark pair in proton-proton collisions at a centre-of-mass energy of 13 TeV at the LHC with the CMS detector. The process is chosen, given that a considerable enhancement of the coupling to b-quarks is possible in theories beyond the Standard Model, such as the Minimal Supersymmetric extension of the Standard Model (MSSM) and the Two Higgs Doublet Model (2HDM). A new boson would manifest itself as an excess in the invariant mass distribution of the two leading b-tagged jets. The final state is rather challenging, requiring dedicated triggers to control the large rate of multijet backgrounds. The data used in this thesis was collected during 2017 and 2018 and combined with a previous search in the same channel with 2016 data to provide results using the full Run 2 dataset, accounting for up to 127 fb^{-1} of data.

The search is performed for several mass hypotheses in two categories. In one, it is conducted in a fully-hadronic category, which is sensitive only to masses above 300 GeV and limited by trigger rates. Sensitivity to lower masses is recovered in the semi-leptonic category, in which a muon is required within a final state jet, reducing the trigger rates with low transverse-momenta jets.

Special attention is given to the background modelling, for which a data-driven approach has been used, and efforts have been made to reduce the uncertainties. In addition, a new modified event selection yields significant enhancements in the signal efficiency and overall significance compared to previous analyses.

These improvements enabled the exploration of masses up to 1.8 TeV, a significant extension compared to existing results. The sensitivity towards lower masses was extended to 125 GeV with the SL category. No significant deviation with respect to the background-only hypothesis was observed, and upper limits were set on the cross-section times branching ratio of the scrutinized process at 95% confidence level. These limits are translated into exclusion limits in the parameter space of MSSM and 2HDM benchmark scenarios. These are the most stringent limits to date on this process.

Zusammenfassung

Diese Dissertation stellt die Suche nach schweren neutralen Higgs-Bosonen vor, die in Assoziation mit b-Quarks in Proton-Proton-Kollisionen bei einer Schwerpunktsenergie von 13 TeV am LHC erzeugt werden, in ein b-Quark-Paar zerfallen und mit dem CMS-Detektor nachgewiesen werden. Dieser Prozess wurde gewählt, da in Theorien jenseits des Standardmodells, wie der minimalen supersymmetrischen Erweiterung des Standardmodells (MSSM) und dem Zwei-Higgs-Dublett-Modell (2HDM), eine erhebliche Verstärkung der Kopplung an b-Quarks möglich ist. Ein neues Boson könnte sich als Überschuss in der invarianten Massenverteilung der beiden führenden Jets manifestieren, die mit der Fragmentation von b-Quarks kompatibel sind. Der Endzustand stellt hohe Ansprüche an die Selektion bereits während der Datennahme, um die hohe Rate an Multijet-Hintergrundereignissen zu kontrollieren. Die in dieser Arbeit verwendeten Daten wurden in den Jahren 2017 und 2018 gesammelt und mit einer vorherigen Suche im selben Kanal mit Daten aus dem Jahr 2016 kombiniert, um Ergebnisse unter Verwendung des vollständigen Datensatzes von Run 2 zu liefern, der bis zu 127 fb^{-1} Daten umfasst.

Die Suche wird für mehrere Massenhypothesen in zwei Kategorien durchgeführt. In einer wird sie in einer vollhadronischen Kategorie durchgeführt, die nur für Massen über 300 GeV empfindlich ist und durch die Limitierung der Datennahmerate begrenzt ist. Die Empfindlichkeit gegenüber niedrigeren Massen wird in der semileptonischen Kategorie wiederhergestellt, in der zusätzlich ein Myon innerhalb eines Jets im Endzustand verlangt wird, wodurch die Datennahmerate bei Jets mit niedrigem Transversalimpuls erheblich reduziert wird.

Besonderes Augenmerk wird auf die Hintergrundmodellierung gelegt, für die ein datenbasierter Ansatz verwendet wurde, und es wurden Anstrengungen unternommen, um die damit verbundenen systematischen Unsicherheiten zu verringern. Darüber hinaus führt eine optimierte Ereignisauswahl zu erheblichen Verbesserungen der Signaleffizienz und der Gesamtsignifikanz im Vergleich zu früheren Analysen.

Diese Verbesserungen ermöglichten die Erforschung von Massen bis zu 1.8 TeV, eine erhebliche Erweiterung im Vergleich zu den vorhandenen Ergebnissen. Die Empfindlichkeit im Bereich niedriger Massen wurde mit der semileptonischen Kategorie bis herunter zu 125 GeV erweitert. Im Ergebnis wurde keine signifikante Abweichung von der Nur-Hintergrund-Hypothese beobachtet, und es wurden Obergrenzen für das Produkt aus Wirkungsquerschnitt und Verzweigungsverhältnis des untersuchten Prozesses mit einem Konfidenzniveau von 95% festgelegt. Diese Grenzwerte werden in Ausschlussgrenzen im Parameterraum der MSSM- und 2HDM-Benchmark-Szenarien übersetzt. Dies sind die bislang strengsten Grenzwerte für diesen Prozess.

Contents

chapter 1

Introduction

> *"In the next decades, future progress in High Energy Physics is in the hands of experimentalists whose discoveries will reveal the way Nature has solved these big questions, forcing the theorists to renounce/review/question deeply rooted bias/prejudice."*
>
> — Christophe Grojean in [1]

The quest for unveiling the intricate tapestry of reality has always been a driving force in the field of particle physics. The Standard Model (SM), an elegant and well-established theory, is a testament to this journey. The SM describes how fundamental particles and their interaction behave in the context of the strong and electroweak forces. Notable milestones in its validation include the discovery of the W [2] and Z bosons [3] and the top quark [4]. Their properties, like mass and decay width, are consistent with the Standard Model [5–11].

The discovery of a scalar particle in agreement with the Higgs boson [12, 13] in 2012 was the final accomplishment of the SM and one of the greatest triumphs of both theory and experiment. It confirmed the mechanism in which bosons gain their masses, the electroweak symmetry-breaking (EWSB), and fermions via Yukawa interaction. Its properties show remarkable consistency with the SM predictions for the observed Higgs mass. In particular, the Higgs boson signal yields as well as all the measured coupling strengths to fermions and vector bosons, are compatible with the SM [14–16], thus upholding the SM's validity up to the TeV scale with unparalleled precision.

Despite its monumental successes, the SM is not without shortcomings. It does not incorporate within its framework a description of gravity, one of the four fundamental forces, or account for the observed neutrino oscillation that directly implies the non-vanishing nature of their masses [17, 18]. The SM also cannot explain the observed imbalance between baryonic and anti-baryonic matter in the Universe, and there is no explanation for why quantum chromodynamics (QCD) seems to preserve charge parity (CP) symmetry, also known as the strong CP problem. Furthermore, the discovered Higgs boson gives rise to another issue known as the hierarchy problem [19]. The observed Higgs mass would require large radiative corrections and is only plausible by fine-tuning the SM parameters. The naturalness of these corrections is found to be rather unsatisfactory as they lack any physical motivation. Moreover, it is widely understood from cosmological observations that more than 85% of the mass content in the universe is attributed to dark matter (DM) and dark energy [20, 21]. However, the SM does not provide a suitable candidate to decipher their nature, remaining one of the most important open issues in particle physics and cosmology. These unresolved issues indicate the existence of physics Beyond the Standard Model (BSM), awaiting discovery.

Several BSM theories have been proposed to tackle these open questions, with many suggesting an enlargement of the SM Higgs sector. A relatively simple extension consists of the inclusion of a second scalar doublet, as proposed by the Two-Higgs-Doublet Models (2HDM) [22]. This addition implies a more complex phenomenology that could, for example, provide additional sources of CP violation. Furthermore, this expansion results in the Higgs sector featuring five physical Higgs bosons, three of them neutral (h, H, A) and two charged $(H^{\pm})$. The neutral states h and H are CP-even while A is CP-odd. Together they are generally referred to as ϕ.

Another well-established BSM theory is the Minimal Supersymmetric Standard Model (MSSM), which introduces a symmetry between fermions and bosons [23]. In the MSSM model, the Higgs sector also comprises two complex scalar doublets while incorporating principles from supersymmetry (SUSY). Within the MSSM framework, the particle content of the SM is doubled. Consequently, the significant radiative corrections to the Higgs mass are cancelled, thus addressing the hierarchy problem [24]. In the MSSM, it is also of particular interest that the strong and electroweak interactions unify at the scale of 10^{16} GeV [25]. Moreover, the MSSM naturally introduces possible dark matter candidates [26–29].

The existing data does not rule out the possibility that the SM is a low-energy approximation of a broader theory. The observation of the 125 GeV Higgs boson could correspond to a neutral scalar state of the 2HDM and MSSM models [30–32], with the others yet to be discovered. Given the compelling motivations discussed above, the search for signatures of an extended Higgs sector is robustly justified. High-energy colliders, such as the Large Hadron Collider (LHC) [33] at the European Organisation for Nuclear Research (CERN), offer unparalleled opportunities for such explorations, reaching energies up to the TeV scale. The Compact Muon Solenoid (CMS) [34] detector is an exemplary instrument for detecting extended Higgs sector signatures.

The present work is dedicated to the search for a neutral spin-0 resonance in a b-quark final state, which is a potential signature of an additional Higgs boson. The b-associated production process is selected over the gluon-gluon fusion to suppress the higher QCD multi-jet background involved in the latter. As in the SM, in certain MSSM and 2HDM scenarios, the Higgs bosons may feature a large coupling to bottom quarks, which can be significantly enhanced in certain regions of the parameter space of such models [35, 36].

Still, this final state is overwhelmed by a large contribution of QCD multi-jet events, and dedicated triggers designed to enhance the significance of possible signals are required. An analysis is then carried out in events featuring at least three b-quark jet candidates; the two more energetic ones are assumed to originate from the decay of the Higgs, and the third jet would correspond to the associated b-quark from the production process that would also help reduce the QCD multi-jet background. This final state will be referred hereby as a fully hadronic (FH) category. Complementarily, and benefitting from the fact that a sizable number of b-jets will contain a muon within its hadronization cone [37, 38], a muon can be selected within a jet already at trigger level. Consequently, the thresholds in transverse momentum of the jets can be lowered at trigger level while keeping the rate of recorded events under control. This allows recording events with less energetic jets and, therefore, extending the search toward lower masses. Such a final state defines a semi-leptonic (SL) category.

This investigation spans the invariant mass of the two leading jets in transverse momentum, where an excess of events in the di-jet mass distribution is anticipated, potentially indicating the presence of an additional Higgs boson. In the scenario that no signal is found, upper limits on the production cross-section times branching ratio of the process are set, as well as limits on the theoretical parameters of BSM models.

The dataset analyzed was collected by the CMS detector during the LHC Run 2 data-taking period in 2017 and 2018. This corresponds to 31.2 and 54.5 fb^{-1} of data in the 2017 FH and 2018 FH categories, respectively. Here, the mass range of ϕ is explored, starting from 300 GeV and up to 1800 GeV. Additionally, 36.7 fb^{-1} of 2017 data were analyzed in the SL category, where the mass range spans from 125 GeV up to 700 GeV. Due to issues with the muon-jet association in the triggers, the 2018 SL analysis could not be conducted. To avoid double-counting in the 2017 dataset, events selected by the SL search were excluded from the FH category. The 2017 analyses for both FH and SL categories build upon and refine previous studies [39, 40], incorporating enhanced simulations, improved calibrations, and reconstruction techniques, as well as new methodologies to enhance the sensitivity of this search. In the case of the 2018 FH category, an inaugural investigation is performed.

The results from the 2017 and 2018 analysis presented in this thesis are combined with those from a CMS 2016-only search in the same final state [35, 41] to provide the first full Run 2 results, accounting for up to 123 fb^{-1} of data, the most extensive dataset to date explored in the search for this signature. The resulting exclusion limits on the signal cross-section times branching ratio are translated into the parameter space of several 2HDM and MSSM benchmark scenarios, providing the most stringent limits set thus far in most of the parameter space studied.

This thesis is structured as follows: Chapter 2 provides an overview of the theoretical foundations of the SM and the relevant BSM theories, laying the stage for the subsequent analysis. Chapter 3 outlines the experimental apparatus, from the LHC to the CMS detector and its components. To guarantee excellent data quality for physics analysis, it is of utmost importance that proper calibrations in the relevant detector subsystems are deployed. This is the case of the alignment of the CMS tracking system [34, 42]. My contribution to the CMS tracker alignment project in preparation and during the LHC Run 3 data-taking period is briefly discussed in this Chapter and detailed in Appendix. A. The reconstruction of the physics objects that are relevant for the search discussed in this thesis is described in Chapter 4.

Chapter 5 discusses the search for additional neutral Higgs in b-quark final states, outlining the analysis strategy and statistical methods employed. My efforts were primarily concentrated on the comprehensive analysis of the 2018 data in the FH category, alongside significant contributions to the 2017 SL and FH analyses. These contributions included measuring online trigger efficiencies and deriving essential elements for background modelling. A pivotal aspect of my work was developing a new strategy to recover events triggered by the associated b-jet candidate. This approach notably amplified the sensitivity of the analysis by a factor of 2, especially in the high-masses regime. Chapter 6 presents the results from the individual channels analyzed, alongside their combination with the 2016 analysis. My contribution included preparing the combined results. I also conducted the model interpretations of these results in various benchmark scenarios, discussed in this Chapter. Furthermore, Chapter 7 provides a synthesis and outlook.

As previously noted, the analysis presented in this thesis requires specialized triggers to guarantee the recording of possible events of interest while complying with the demands of event processing time. New techniques to improve the detection and benefit the overall event triggering for this analysis were deployed during the preparatory phase and throughout the LHC Run 3 data-taking period. To benefit from them, these analysis triggers were upgraded. My contribution to their development is outlined in Appendix C.

chapter 2

Theoretical framework

This chapter provides a description of the Standard Model, its limitations, and an introduction to the phenomenological foundations of extended Higgs sectors and supersymmetry that are relevant to the main topic of this thesis.

2.1 The Standard Model of Particle Physics

The Standard Model of particle physics is a relativistic quantum field theory (QFT) based on the $SU(3) \times SU(2) \times U(1)$ symmetry group established in the 1960–1970s [43–48]. It summarizes the current understanding of the basic building blocks of matter and the fundamental sub-atomic interactions, except gravity. An extensive review of the SM can be found in Refs. [49–52]. Within the SM framework, elementary particles can be classified according to their quantum numbers, as explained in the following section.

2.1.1 Elementary particles

According to their intrinsic angular momentum, elementary particles are divided into two main groups: bosons with integer spin and fermions with half-integer spin.

Bosons

Bosons are particles that follow Bose-Einstein statistics. Particles with this property can simultaneously occupy the same quantum state.

Gauge bosons are spin-1 particles that mediate the fundamental forces in the Standard Model. The massless photon mediates the electromagnetic interaction between electrically charged particles. Eight massless gluons carry the strong interaction between particles with colour charge, i.e., quarks and gluons. The weak force is mediated by the massive W and Z bosons, which are responsible for the interaction between particles with flavour, including all quarks and leptons. There are two electrically charged W bosons, W^+ and W^-, while the Z boson is neutral. The W bosons mediate the weak charged current interactions that can alter the flavour of fermions, whereas the Z boson mediates weak neutral current interactions that do not alter particle flavours.

The Higgs boson is a massive scalar boson with spin 0 that gives mass to the other massive SM particles and itself through the interaction with the Higgs field. This fundamental process is predicted by the Higgs mechanism, a cornerstone of the Standard Model. As mentioned in the previous chapter, a particle with properties aligned with the SM Higgs boson was discovered by the ATLAS and CMS experiments at the LHC in 2012 [12,13].

Fermions

Fermions are particles that obey Fermi-Dirac statistics, meaning they adhere to the Pauli exclusion principle, which states that two fermions can not simultaneously occupy the same quantum state.

The fundamental particles classified as fermions are divided into quarks and leptons. There are six flavours of quarks: up (u), down (d), charm (c), strange (s), top (t), and bottom (b). Each quark flavour comes in three colour charges: red, green, and blue. The u, c and t quarks have a positive electric charge of $+2/3\ e$ and are called up-type quarks while the down-type quarks, d, s and b, have -1/3 e electric charges. Quarks participate in both electroweak and strong interactions, with the latter binding them into colour-neutral composite states known as hadrons (e.g., protons and neutrons in atomic nuclei).

Leptons, on the other hand, do not possess colour charge and, therefore, do not engage in strong interactions. There are six leptons: the electron (e), muon (μ), and tau (τ), which carry electric charge, each accompanied by a corresponding neutral particle, the electron neutrino (ν_e), muon neutrino (ν_μ), and tau neutrino (ν_τ). All leptons participate in weak interactions.

Fundamental fermions are divided into three generations of particles organized in isospin doublets with increasing mass. They consist of a charged lepton and its corresponding neutrino and an up-type and a down-type quark for leptons and quarks, respectively. Every fermion has a corresponding anti-particle, which possesses opposite quantum numbers but with the same mass and spin. A summary of the particle content of the SM framework is schematically represented in Fig. 2.1.

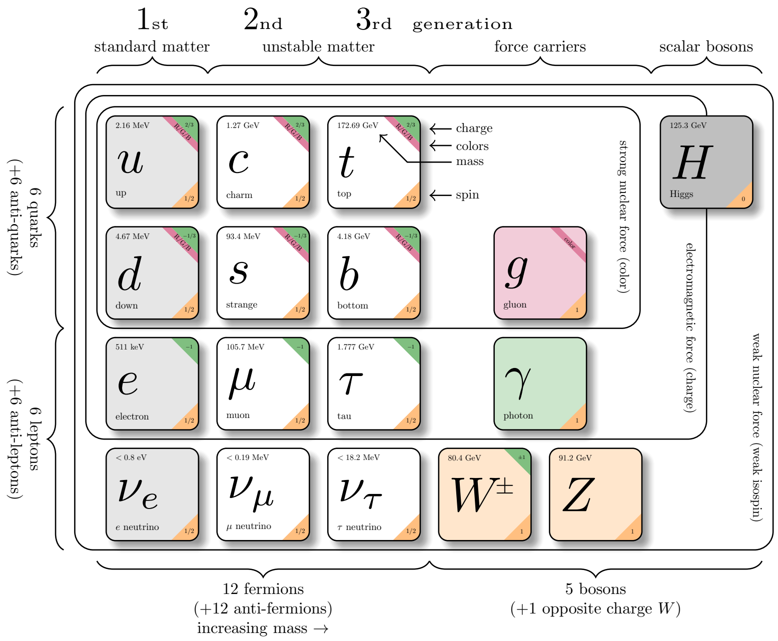


Figure 2.1: The Standard Model's elementary particle content and properties. Adapted from [53], masses taken from [54].

2.1.2 Theoretical framework

The theoretical framework of the SM is formulated on the principles of QFT and the Lagrangian formalism. In the SM, particles are represented as excitations of underlying fields denoted by ψ. The principle of least action is central in deriving the equations of motion or field configurations [55]. It consists on minimizing the action (S), which is defined by the integral of the Lagrangian density ($\mathcal{L}$) over time:

$$S = \int \mathcal{L}(\psi(\vec{x}, t), \partial_\mu \psi(\vec{x}, t))\, dt, \tag{2.1}$$

where $\mathcal{L}$ contains information about the fields and their interactions. For briefness, $\psi(\vec{x}, t)$ will be written from this point as $\psi(x)$ or simply as ψ.

The SM is also a gauge theory, meaning that the fundamental interactions are determined by imposing local symmetries on the Lagrangian density. At the same time, Noether's Theorem [56] enunciates that every symmetry of the action implies a corresponding conserved quantity. These symmetries dictate the form of $\mathcal{L}$ and the interactions among fields.

In the SM, $\mathcal{L}$ is constructed by combining the Lagrangian densities of the individual interactions, which will be detailed in the following.

2.1.3 Quantum electrodynamics

The quantum theory of electrodynamics (QED) describes the interaction of the electromagnetic field with charged matter particles in a relativistic QFT-based approach [49,51]. The QED Lagrangian density $\mathcal{L}_{\mathrm{QED}}$ can be obtained by requesting a local gauge invariance under a U(1) transformation to Dirac's Lagrangian [51]:

$$\mathcal{L}_{\mathrm{Dirac}} = i\bar{\psi}(x)\gamma^\mu \partial_\mu \psi(x) - m\psi\bar{(}x)\psi(x), \tag{2.2}$$

where the fermion field (ψ) is also referred to as the Dirac spinor, while $\bar{\psi}$ is its adjoint.

The aforementioned transformation modifies the spinor as $\psi(x) \to \psi'(x) = \psi(x)e^{iq\alpha(x)}$, and its adjoint as $\bar{\psi}(x) \to \bar{\psi}'(x) = e^{-iq\alpha(x)}\psi(x)$ with the real constant q, and the real phase $a(x)$. This results in a transformed Lagrangian in the form of:

$$\mathcal{L}' = \mathcal{L}_{\mathrm{Dirac}} - q\bar{\psi}(x)\gamma^\mu \psi(x)\partial_\mu \alpha(x). \tag{2.3}$$

To become gauge invariant, the derivative $\partial_\mu \psi(x)$ is substituted by a so-called covariant derivative $D_\mu \psi(x)$. In QED, D_μ is defined as:

$$D_\mu = \partial_\mu + iqA_\mu(x), \tag{2.4}$$

where $A_\mu(x)$ is a vector field that transforms according to $A_\mu \to A_\mu(x) + \partial_\mu \theta(x)$. The Lagrangian density now becomes:

$$\mathcal{L}_{\mathrm{QED}} = \bar{\psi}(x)i\gamma^\mu \partial_\mu \psi(x) - m\bar{\psi}(x)\psi(x) - q\bar{\psi}(x)\gamma^\mu \psi(x)A_\mu(x), \tag{2.5}$$

where the first term represents the fermionic kinetic term, the second is a mass term, and the third describes a new interaction between the vector field and the fermionic field. Assuming a massless and neutral vector boson, correspondingly with the photon's properties, is associated with the introduced vector field, a kinetic term corresponding to it is added to the Lagrangian density, similar to that of the Dirac field. This term has the form:

$$\mathcal{L}_A = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu}, \tag{2.6}$$

and represents a free vector field, where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is a so-called "field strength". which is gauge invariant for the U(1) abelian group.

The complete Lagrangian incorporating both the fermionic and gauge field parts reads:

$$\mathcal{L}_{\text{QED}} = \overline{\psi}\gamma^\mu\partial_\mu\psi - m\overline{\psi}\psi - q\overline{\psi}\gamma^\mu\psi A_\mu - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}, \tag{2.7}$$

where all terms are invariant under the selected gauge transformation. The quantity $J_\mu = q\overline{\psi}(x)\gamma^\mu\psi(x)$ is referred to as the current density, which leads to the conservation of the electric charge $q = -e$. The electric charge can be seen as a coupling strength of the electromagnetic interaction. It is related to the fine structure constant (α) as $\alpha = \frac{e^2}{4\pi}$. QED is a renormalizable theory, and α is a running coupling constant that depends on the energy scale of the process (Q^2). The fine structure constant tends to $\sim \frac{1}{137}$ [57] as the Q^2 approaches zero and increases logarithmically towards higher energies. At Q^2 in the order of the Z boson mass, $\alpha \sim \frac{1}{128}$ [58].

2.1.4 Quantum chromodynamics

Quantum chromodynamics describes the strong interaction between colour-charged particles, mediated by the gluon. In analogy to QED, quarks possess a colour charge while anti-quarks possess an anti-colour charge. This quantum number can be red, green and blue, as mentioned earlier.

The Lagrangian density of QCD can be derived in a similar approach, by enforcing gauge invariance under a symmetry of the SU(3) group. A SU(N) group can be expressed by an N×N unitary matrix with determinant 1. This results in the group being represented by $\mathrm{N}^2 - 1$ linearly independent infinitesimal transformations, also known as "generators". In the case of QCD, there are eight generators, known as the Gell-Mann matrices (t_a), which result in the introduction of eight vector fields ($G^a_{\mu\nu}$). The resulting QCD Lagrangian density can be written as:

$$\mathcal{L}_{\text{QCD}} = i\overline{\psi}(x)\gamma^\mu D_\mu\psi(x) - m\overline{\psi}(x)\psi(x) - \frac{1}{4}G^a_{\mu\nu}G_a^{\mu\nu}, \tag{2.8}$$

where vector fields represent the eight coloured, massless and electrically neutral gluons. The covariant derivative D_μ now takes the form of $D_\mu = \partial_\mu - ig_sT_aG^a_\mu$, where g_s represents a dimensionless coupling QCD constant and T_a corresponds to 3×3 matrices in the colour space related to the Gell-Mann matrices as $T_a = \frac{t_a}{2}$. It is important to note that for L_{QCD} to be gauge invariant, the mass term must be equal to zero [59].

Introducing the covariant derivative into eq. 2.8 results in:

$$\mathcal{L}_{\text{QCD}} = \overline{\psi}(x)(i\gamma^{\mu}\partial_{\mu} - m)\psi(x) - g_s\overline{\psi}(x)(\gamma^{\mu}T_a G^a_{\mu})\psi(x) - \frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a, \tag{2.9}$$

which allows to identify the first term with the kinematics and mass of the quarks, and the second term with the quark-gluon interaction. Eight conserved colour-currents $J^a_{\mu} = g_s\overline{\psi}(x)\gamma^{\mu}T_a\psi(x)$ are assiciated to the introduced gluons. The gluon field strength tensor is defined as:

$$G^a_{\mu\nu} = \partial_{\mu}G^a_{\nu} - \partial_{\nu}G^a_{\mu} + g_s f^{abc}G^b_{\mu}G^c_{\nu}, \tag{2.10}$$

which results, when expanding the last term of eq. 2.9, in a kinetic term for the introduced vector bosons, in analogy to eq. 2.6, and a component describing the gluon self-interaction coming from the last term of eq. 2.10, which is introduced to guarantee gauge invariance as the SU(3) generators do not commute. The latter implies that, contrary to QED, QCD is a non-abelian theory. The term f^{abc} is referred to as the structure constant of the SU(3) group.

An important consequence of the gluon self-interaction is that the running strong coupling constant (α_s), which is related to g_s as $\alpha_s = 4\pi g_s^2$, behaves with respect to Q^2 as follows [59]:

$$\alpha_s(Q^2) = \frac{12\pi}{(33 - 2n_f)\log\left(\frac{Q^2}{\Lambda^2_{\text{QCD}}}\right)}, \tag{2.11}$$

where n_f is the number of quark flavours with mass $m_q^2 \ll Q^2$. From eq. 2.11, it can be seen that α_s decreases towards large energy transfer. The limit where Q^2 becomes very large is referred to as "asymptotic freedom". In this limit, the strong force tends to zero, which makes it possible study QCD in a perturbative approach. On the other hand, towards low Q^2 the coupling constant becomes large, passing the perturbative regime at an energy conventionally denoted with Λ_{QCD}. This energy regime results in an effect known as "colour confinement", where quarks and gluons arrange into colour-neutral clusters (hadrons). The value of the term Λ_{QCD} depends on n_f, and ranges from about 100 to 300 MeV [54].

2.1.5 Weak interaction and electroweak unification

The weak interaction, initially postulated to explain nuclear β decays, does not conserve parity (P), as demonstrated by the Wu experiment [60]. The "vector-axial" (V-A) approach was introduced to address this issue.

The Dirac spinors are decomposed into left and right-handed components:

$$\psi(x) = \psi_L(x) + \psi_R(x), \tag{2.12}$$

where each component is obtained from the left and right-handedness projection operators, P_L and P_R, respectively, and the γ^5 Dirac matrix as follows:

$$\psi_L(x) = P_L\psi(x) = \frac{1}{2}(1-\gamma^5)\psi(x), \tag{2.13}$$

$$\psi_R(x) = P_R\psi(x) = \frac{1}{2}(1+\gamma^5)\psi(x). \tag{2.14}$$

The V-A approach exploits the distinct parity properties of vector and axial vector currents. Vector currents ($\overline{\psi}(x)\gamma^\mu\psi(x)$) are typically assigned positive parity, thus conserving parity, whereas axial vector currents ($\overline{\psi}(x)\gamma^\mu\gamma^5\psi(x)$) transform with a sign change under a parity transformation.

The only critical change to generate parity violation is to re-write the term $\overline{\psi}(x)\gamma^\mu\psi(x)$ seen in the QED Lagrangian (eq. 2.7) in the context of the weak interaction as follows [59]:

$$\overline{\psi}(x)[\gamma^\mu\frac{1}{2}(1-\gamma^5)]\psi(x) = \frac{1}{2}[\overline{\psi}(x)\gamma^\mu\psi(x) - \overline{\psi}(x)\gamma^\mu\gamma^5\psi(x)]. \tag{2.15}$$

The second part of the equality showcases the V-A structure, which explicitly violates parity conservation by the combination of both vector and axial components. From eq. 2.15, it can also be derived that the weak interaction excludes the (RH) components of fermions, involving only their left-handed (LH) counterparts. Considering the asymmetry in the parity of particles and anti-particles, weak interactions also include (exclude) RH (LH) antifermions.

The weak interaction has also been found to violate charge conjugation (C) and parity (P) conservation together [61]. This effect is beautifully incorporated in the SM through the quark-mixing Cabibbo-Kobayashi-Masakawa (CKM) matrix [62].

Electroweak unification

The electroweak (EW) theory is formulated in a SU(2)×U(1) gauge invariance that involves two conserved quantities, the weak isospin (I_3) and the hypercharge (Y) associated with the SU(2) and U(1) transformation, respectively.

LH and RH fermions are treated differently in the EW framework. LH fermions (ψ_L) are written as weak isospin doublets [59], with non-zero weak isospin and hypercharge values. On the other hand, RH fermions (ψ_R) transform as singlets with zero isospin and non-zero hypercharge. The covariant derivatives are also different for doublets and singlets. They are written, respectively, as:

$$D_{\mu,L}\psi_L = (\partial_\mu + igW^i_\mu\tfrac{\sigma^i}{2} + ig'B_\mu\tfrac{Y}{2})\psi_L \quad , \tag{2.16}$$

$$D_{\mu,R}\psi_R = (\partial_\mu + ig'B_\mu\tfrac{Y}{2})\psi_R \quad , \tag{2.17}$$

where σ^i are the three generators of the non-abelian SU(2) group, the Pauli matrices, while the abelian U(1) group has one generator. The terms g and g' are the coupling constants of the introduced fields W^i_μ and B_μ.

A fundamental aspect of the EW unification is that I and Y can be related to the electric charge q with the expression $q = I + \frac{1}{2}Y$, also known as the Gell-Mann–Nishijima formula [63, 64], integrating the electromagnetic and weak interactions under a single electroweak force.

The mixing of the vector fields W^i_μ and B_μ gives rise to the four EW gauge bosons:

$$W^\pm_\mu = \frac{1}{\sqrt{2}}(W^1_\mu \mp W^2_\mu), \tag{2.18}$$

$$Z_\mu = W^3_\mu \cos\theta_W - B_\mu \sin\theta_W, \tag{2.19}$$

$$A_\mu = W^3_\mu \sin\theta_W + B_\mu \cos\theta_W, \tag{2.20}$$

corresponding to the W$^\pm$, Z and γ, respectively. The parameter θ_W is the EW mixing angle, also known as the Weinberg angle. Depending on the exchanged electric charge, there are two types of weak interactions: the “charged current interactions” mediated by the W$^\pm$ bosons and the “neutral current interactions” mediated by the Z boson.

The EW Lagrangian density can be written as:

$$\mathcal{L}_{\mathrm{EW}} = -\frac{1}{4}W^{i\mu\nu}W^i_{\mu\nu} - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} + \overline{\psi}_L i\gamma^\mu D_{\mu,L}\psi_L + \overline{\psi}_R i\gamma^\mu D_{\mu,R}\psi_R, \tag{2.21}$$

After applying the mixing, the resulting physical states interact with vertices such as WWZ, WWγ, WWWW, WWZγ, WWZZ, and WW$\gamma\gamma$, due to the non-abelian SU(2) transformation.

The EW unification was one of the greatest achievements in particle physics, although it raised a problem. Any mass term in the $\mathcal{L}_{\mathrm{EW}}$ would break gauge invariance. The mechanism which addresses this issue in the SM is described in the following section.

2.1.6 The Brout-Englert-Higgs mechanism

The electroweak symmetry breaking, also known as the Brout-Englert-Higgs (BEH) mechanism, proposes a mechanism for providing particles with their mass [65–69]. This process involves the introduction of a scalar field to break the symmetry of the Lagrangian. A complex scalar doublet is the simplest field that can be added for this purpose, resulting in a field that is symmetric under SU(2) gauge transformations but with a non-zero vacuum expectation value. The new field is represented as:

$$\Phi = \begin{pmatrix}\phi^+ \\ \phi^0\end{pmatrix} = \begin{pmatrix}\phi_1 + i\phi_2 \\ \phi_3 + i\phi_4\end{pmatrix} \tag{2.22}$$

The spin of the doublet is assumed to be zero. Thus, Klein-Gordon's equation dictates the propagation. The Lagrangian density can be written as:

$$\mathcal{L}_\phi = (D^\mu \phi(x))^\dagger (D_\mu \phi(x)) - V(\phi(x)) = |D_\mu \phi(x)|^2 - V(\phi(x)), \tag{2.23}$$

with $|\phi|^2 = \phi^\dagger \phi = (\phi^+, \phi^0) \cdot \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$. D_μ is the SU(2)×U(1) covariant derivative from eq. 2.16. A potential, $V(\phi)$, that is gauge invariant under rotations of ϕ^+ and ϕ^0 can be written as:

$$V(\phi) = \mu^2 |\phi|^2 + \lambda |\phi|^4 \tag{2.24}$$

where μ and λ are constants. The condition that $\lambda > 0$ must be accomplished to maintain vacuum stability by guaranteeing a positive potential at large values of $|\phi|$. Analyzing the other term, μ, a positive μ^2 would result in a trivial ground state at $|\phi| = 0$. On the other hand, if $\mu^2 < 0$, the potential acquires the shape of a so-called "mexican hat", shown in Fig. 2.2. The ground state is now a circle in the complex plane located at $|\phi| = v$, where v is referred to as the vacuum expectation value, defined as:

$$v = \sqrt{-\frac{\mu^2}{\lambda}} \tag{2.25}$$

The value of $v \approx 246$ GeV [54] sets the symmetry-breaking scale.

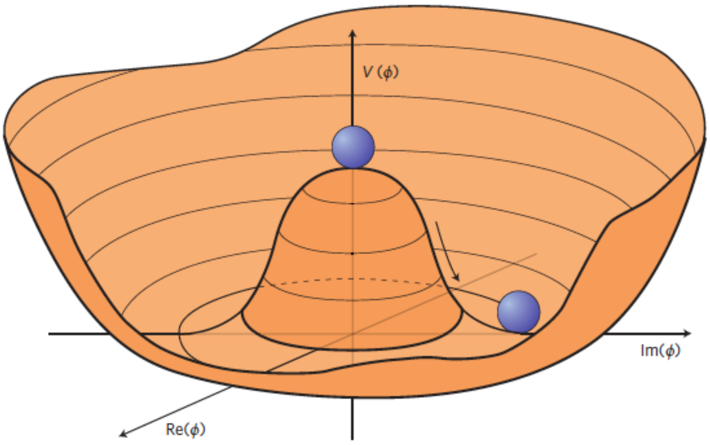


Figure 2.2: An illustration of the Higgs potential with $\mu^2 < 0$. Taken from [70]

The symmetry of this ground state is broken by fixing a minimum. For simplicity, $\phi^+_{vac} = 0$ and $\phi^0_{vac} = v$ are chosen. With this choice, the U(1) symmetry is preserved, consistent with the electric charge's conservation and the fact that the photon is massless [71].

Considering the Higgs particle to be an excitation of the Higgs field:

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \tag{2.26}$$

which, when inserted in eq. 2.24, yields:

$$V(\phi) = -\mu^2 H^2 + \lambda v H^3 + \frac{1}{4}\lambda H^4 \tag{2.27}$$

The first term in 2.27 is related to the mass of the higgs $m_H = \sqrt{-2\mu^2}$, a free parameter of the SM determined to be 125.30 GeV [54]. The second and third terms account for the self-interaction of the Higgs boson.

Returning to the Higgs field's Lagrangian density (eq. 2.23), substituting the covariant derivative yields:

$$\mathcal{L}_\phi = \left| \left(\partial_\mu + igW^i_\mu \frac{\sigma^i}{2} + ig' B_\mu \frac{Y}{2} \right) \phi(x) \right|^2 - V(\phi(x)). \tag{2.28}$$

Considering the choice of the minimum of the potential and expanding eq. 2.28, we obtain:

$$\begin{aligned} \mathcal{L}_\phi &= \frac{1}{8} v^2 g^2 \left[(W^1_\mu)^2 + (W^2_\mu)^2 \right] + \frac{1}{8} v^2 \left[g^2 (W^3_\mu)^2 - 2gg' W^3_\mu B^\mu + g'^2 B^2_\mu \right] \\ &= \frac{1}{8} v^2 g^2 \left[(W^1_\mu)^2 + (W^2_\mu)^2 \right] + \frac{1}{8} v^2 (gW^3_\mu - g' B_\mu)^2 + 0(gW^3_\mu + g' B_\mu)^2. \end{aligned} \tag{2.29}$$

Now, considering the definition of the physical states of the EW bosons (eq. 2.18, 2.19, and 2.20), and taking into account that $\tan\theta_W = g'/g$, one obtains the masses of the EW bosons:

$$\begin{aligned} m_W &= \frac{1}{2} g v, \\ m_Z &= \frac{1}{2} v \sqrt{g^2 + g'^2}, \\ m_A &= 0. \end{aligned} \tag{2.30}$$

which fulfill the relation $m_Z = m_W / \cos\theta_W$, while A_μ remains massless, as expected. The coupling of the Higgs to the gauge bosons (g_V) is proportional to their mass squared, $g_V = 2\frac{m_v^2}{v}$.

Fermion masses remain to be addressed. The Yukawa interaction with the Higgs field provides a method to do so. Additional gauge invariant interaction terms are added to the Lagrangian density in the form of:

$$\mathcal{L}_{\text{Yukawa}} = -g_Y \overline{\psi}(x) \Phi \psi(x), \tag{2.31}$$

and the fermion masses can be derived from $m_f = \frac{1}{\sqrt{2}} g_Y^f v$ after spontaneous symmetry breaking. The term g_Y^f is the Yukawa coupling constant of a Fermion to the Higgs boson, which is proportional to the fermion mass. It is relevant to note that the Yukawa interaction for quarks results in them being subject to weak flavour mixing, as per the aforementioned CKM matrix [62].

2.1.7 The Standard Model Lagrangian

Taking the above discussion into consideration, the SM Lagrangian density can be formulated from a $\mathrm{SU}(3)_c \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ gauge symmetry, combining the components of the individual interactions:

$$\begin{aligned} \mathcal{L}_{\mathrm{SM}} = & -\frac{1}{4} W^{i\mu\nu} W^i_{\mu\nu} - \frac{1}{4} B^{\mu\nu} B_{\mu\nu} + \overline{\psi}_L i\gamma^\mu D_{\mu,L} \psi_L + \overline{\psi}_R i\gamma^\mu D_{\mu,R} \psi_R \\ & + \overline{\psi}(x)(i\gamma^\mu \partial_\mu)\psi(x) - g_s \overline{\psi}(x)(\gamma^\mu T_a G^a_\mu)\psi(x) - \frac{1}{4} G^a_{\mu\nu} G^{\mu\nu}_a \\ & + \left| \left(\partial_\mu + igW^i_\mu \frac{\sigma^i}{2} + ig' B_\mu \frac{Y}{2} \right) \phi(x) \right|^2 - V(\phi(x)) \\ & - g_Y \overline{\psi}(x) \Phi \psi(x), \end{aligned} \tag{2.32}$$

resulting in a formulation that summarizes the current understanding of particle physics into one equation.

The discovery of a unique Higgs boson [12, 13], predicted by the BEH mechanism, was the last cornerstone of the SM, as mentioned in Chapter 1. A summary of the phenomenology of the Higgs boson at the LHC is given in the next section, followed by an overview of the shortcomings of the SM.

2.1.8 The Higgs Boson at the LHC

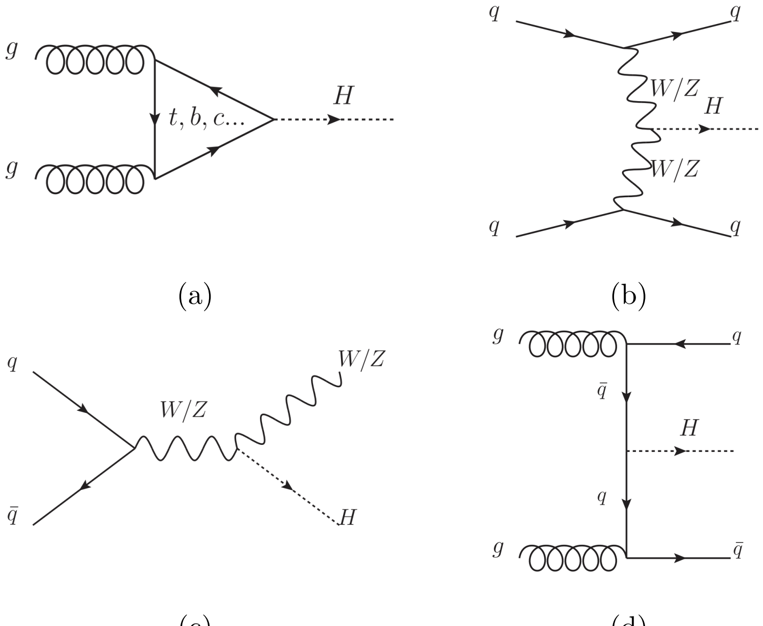
A SM Higgs boson can be produced at the LHC primarily through the following channels, listed in descending order according to their cross-section:

- **gluon-gluon fusion** (ggF) production, mediated by a virtual quark loop, as illustrated in Fig. 2.3a. The top quark is the main contributor to the loop, as the Yukawa couplings are proportional to the mass of the fermion.

- **vector boson fusion** (VBF), where two quarks each irradiate a weak boson. The two weak bosons fuse, resulting in the emission of a Higgs, as shown in Fig. 2.3b. The signature of this production process is characterized by two jets in the forward or backward direction respective to the beampipe, with large rapidity separation and high transverse momentum.

- **higgstrahlung** (VH), also known as vector boson associated production, where a quark-antiquark pair annihilates into a weak off-shell boson that emits a Higgs boson to enter an on-shell state, illustrated in Fig. 2.3c.
- **top quark associated production** ($t\bar{t}$H) and **bottom quark associated production** ($b\bar{b}$H), where a Higgs boson is produced in association with a heavy quark pair, as shown in 2.3d. Because of their higher mass, this process is predominantly driven by top [72, 73] and bottom quarks. However, the latter has not yet been measured due to the experimental challenges involved. The latest results from CMS setting limits at 3.7 times the SM predicted cross-section value [74].

The cross-sections of the aforementioned production processes in p - p collisions are shown in Fig. 2.4 (left), as a function of the centre-of-mass energy.



(a) (b)

(c) (d)

Figure 2.3: Main production Higgs boson production mechanisms at the LHC at tree level (a) ggF, (b) VBF, (c) VH and (d) $t\bar{t}$H or $b\bar{b}$H. These production processes have been observed at the LHC, except $b\bar{b}$H.

The Higgs boson is a heavy, unstable particle that may decay into a variety of channels. Their corresponding branching fraction (BR) is shown in Fig. 2.4 (right). As the Higgs boson couples only to massive particles, a massive particle loop mediates the decay into a photon pair.

The Higgs boson was discovered by the CMS and ATLAS experiments, with the most significant evidence coming from its decays into photon pairs and into two Z bosons. While other decay channels were also investigated, these two channels stood out despite their lower BR due to their lower background contribution. In contrast, the largest BR of the SM Higgs boson corresponds to the $b\bar{b}$ decay channel, which is also one of the most challenging to detect due to the large QCD multi-jet background at the LHC.

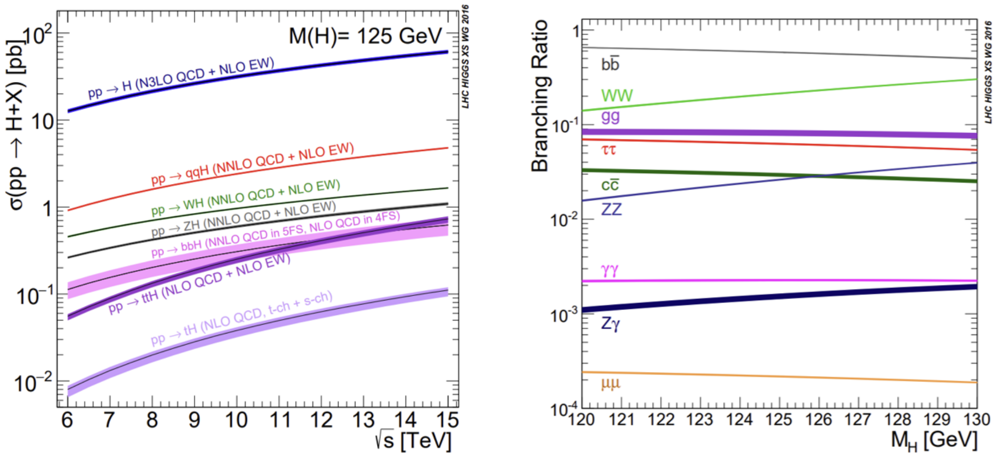


Figure 2.4: Production cross-section of the SM Higgs boson, with a mass of 125 GeV, for different production processes, depending on the centre-of-mass energy of p-p collisions (left). Branching ratios of the SM Higgs boson decay in a mass range near 125 GeV at a centre-of-mass energy of 13 TeV (right). The theoretical uncertainties are represented as bands. Taken from [75].

If kinematically accessible, the existence of additional particles may extend the number of Higgs boson decay channels. If these particles do not produce a signature in the CMS detector, they are referred to as "invisible" decays that can be associated with missing transverse energy. The latest results on direct searches for such signatures yielded a limit in the BR to invisible final states < 0.15 at 95% confidence level [76]. The SM BR to invisible, coming solely from $H \to ZZ \to 4\nu$, is expected to be about 0.1%. Given the current experimental limit, new physics contributing to invisible decays is not completely excluded.

The Higgs boson couplings to vector bosons and third and second-generation fermions that have been measured by the CMS experiment at the time of writing are shown in Fig. 2.5. These measurements highlight the striking agreement of experimental measurements with the SM expectations.

2.1.9 Shortcomings of the Standard Model

The SM has been tested experimentally to a very high precision and over a wide energy range with remarkable success. However, several issues remain unaddressed. For instance, one of the main longstanding unresolved mysteries is the nature of Dark Matter (DM) and

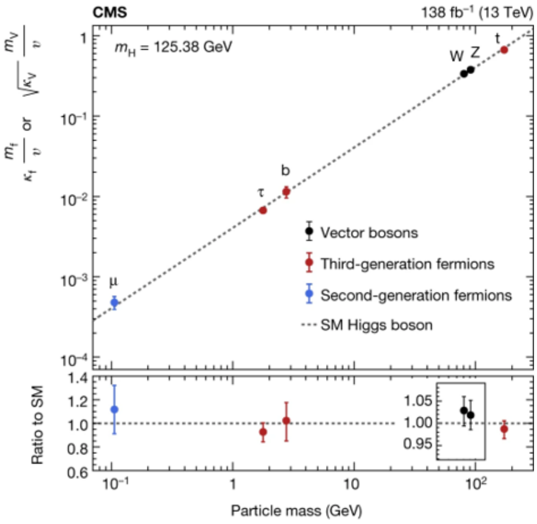


Figure 2.5: Couplings of the Higgs boson to heavy vector bosons and fermions as a function of their mass. Taken from [14].

Dark Energy, which, according to measurements from the Planck Collaboration, accounts for more than 85% of the mass content in the universe [21]. The SM also does not provide a mechanism to account for the non-zero neutrino masses (as shown in Fig. 2.1) [17, 18], and it does not include gravity within its framework. There is also no explanation for the asymmetry between matter and anti-matter observed in the universe, nor why fermion generations are organized as such, or why QCD seems to conserve the CP symmetry.

These facts indicate that the SM can be an effective field theory, applicable to a certain energy range up to a cut-off scale commonly associated with the Planck scale $\Lambda_{Planck} = \mathcal{O}(10^{19})$ GeV, where quantum gravitational effects become dominant [77, 78]. This premise, combined with the fact that the discovered Higgs has a mass of nearly 125 GeV, introduced an issue widely known as the hierarchy problem [19]. It arises from the quantum loop corrections to the Higgs boson mass from the highest scale of the theory:

$$m^2_{h_{SM}} = m^2_{h_0} + \Delta m^2_h, \tag{2.33}$$

where $\Delta m^2_h \sim \Lambda_{Planck}$, thus an unnatural tunning of the bare Higgs mass, m_{h_0}, is needed to cancel out the large contribution of the previous.

Many Beyond Standard Model (BSM) theories have been introduced to solve the open issues of the SM. In the next chapter, the BSM theories that are relevant to the study presented in this thesis are discussed.

2.2 Theories Beyond the Standard Model

In the context of BSM physics, the Higgs sector is of particular interest, as a relatively simple enlargement would lead to a much richer phenomenology that may solve some of the aforementioned problems. One example is the Two Higgs Doublet Model (2HDM), where the scalar sector is extended, as the name indicates, with one additional doublet. This enlargement results in five physical Higgs states, two of them charged, and three of them neutral. Out of the neutral particles, one is a pseudo-scalar, and the others are scalars. As will be detailed later in section 2.2.2, under certain assumptions, the parameter space of 2HDM can be reduced to three parameters. These are tanβ, the ratio of the vacuum expectation value of the two Higgs doublets; cos$(\beta - \alpha)$, where α is the mixing angle between the two scalars; and m_{A}, the mass of the pseudo-scalar particle.

Furthermore, supersymmetry is a strong candidate for new physics that provides an elegant solution to the hierarchy problem, as will be described in section 2.2.1. Moreover, the Higgs sector in the minimal supersymmetric extension of the SM (MSSM) is a particular realization of a 2HDM. At tree-level, there are only two free parameters to constrain the MSSM sector, tanβ and m_{A}.

These BSM may alter the SM couplings to up and down-type fermions, or to quarks and leptons independently. Current measurements do not completely exclude either of these proposals. For instance, constraints derived from SM Higgs measurements leave a considerable part of the parameter space of MSSM and 2HDM models still to be explored, as shown in Fig. 2.6. In this figure, the MSSM model considered is the hMSSM, which is defined later in section 2.2.1. The 2HDM models considered are Type I, II, III, and IV, which are defined later in section 2.2.2.

The next sections provide an overview of MSSM and 2HDM models.

2.2.1 The Minimal Supersymmetric Extension of the SM

A well-known BSM theory is supersymmetry (SUSY), which proposes a symmetry between fermions and bosons. As a consequence, every SM fermion has a bosonic superpartner, named as “s+fermion”, for example, the “sbottom”. In the same way, every SM boson has a fermionic superpartner named “boson+ino”, for instance, the “Higgsino”.

The richer particle landscape provides a solution to the hierarchy problem, as fermion and boson loops contribute with opposite signs to the corrections to the Higgs boson mass, thus cancelling out. In addition, SUSY provides a dark matter candidate: the lightest stable supersymmetric massive particle. Furthermore, within the SUSY framework, the strong and electroweak couplings unify at the scale of $\mathcal{O}(10^{16})$ GeV [25].

It is relevant to note that the SM particles and their supersymmetric counterparts must be mass-degenerate in an unbroken SUSY. We know this is not the case, as superparticles have not been discovered. Therefore, supersymmetry must be a broken symmetry, with superparticles heavier than their SM counterparts. This may affect the cancellation of

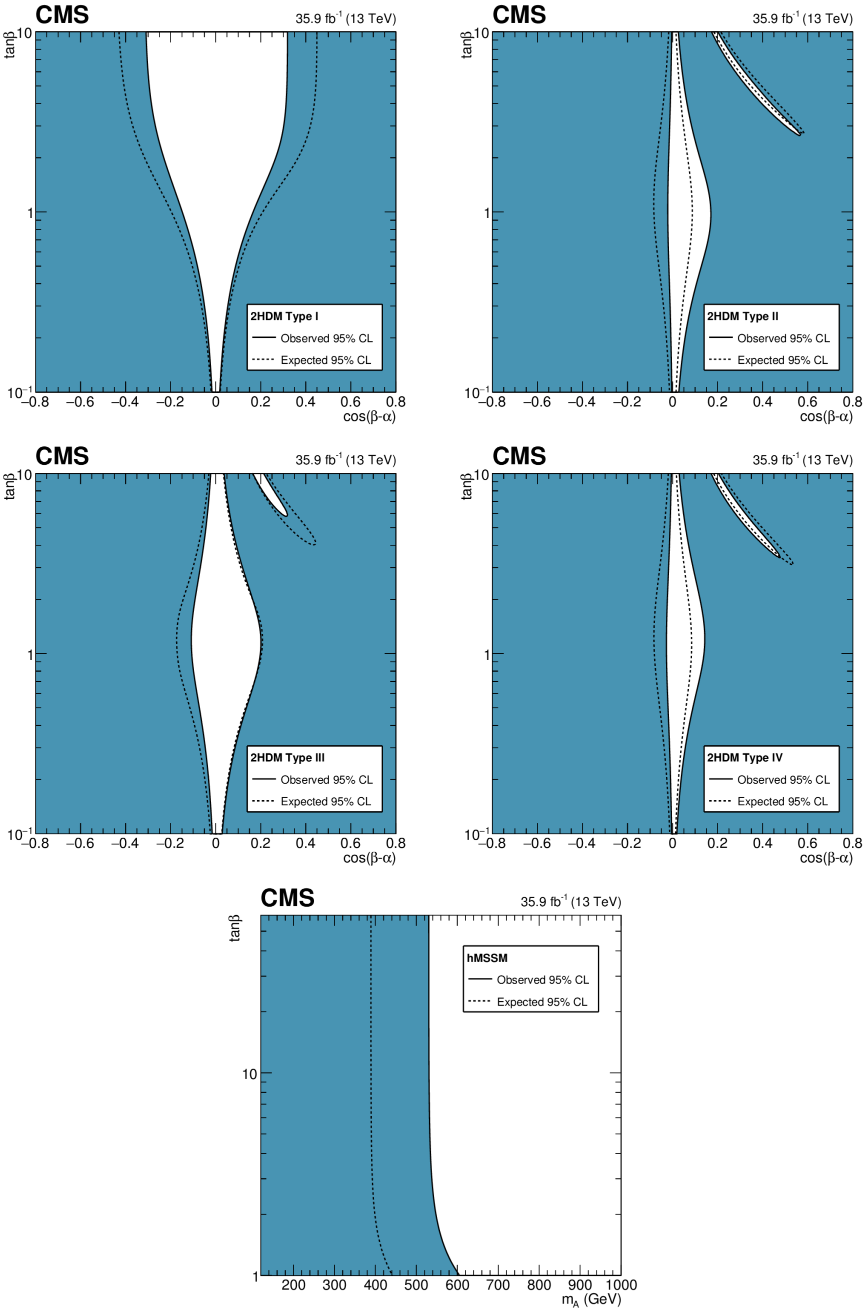


Figure 2.6: Constraints in the $\cos(\beta-\alpha)$ and tanβ parameter space of 2HDM scenarios Type I, II, III and IV, and in the m_{A} vs. tanβ parameter space of the MSSM scenario hMSSM. These scenarios are discussed in the following sections. Taken from [79].

terms that solve the hierarchy problem. However, under certain conditions [80, 81], also known as "soft SUSY breaking", the naturalness is preserved through the inclusion of additional terms to the SUSY Lagrangian density that explicitly break the symmetry but do not introduce quadratic divergences.

The simplest realization of SUSY is known as the Minimal Supersymmetric Extension of the Standard Model (MSSM) [82], which adds only the minimal number of particles to the SM needed in SUSY.

The MSSM Higgs sector

The Higgs sector in MSSM consists of two complex scalar doublets to generate the masses of up-type and down-type quarks and leptons [82]:

$$\Phi_d = \frac{1}{\sqrt{2}}\begin{pmatrix}\phi_d^0\\ \phi_d^-\end{pmatrix}, \quad \Phi_u = \frac{1}{\sqrt{2}}\begin{pmatrix}\phi_u^+\\ \phi_u^0\end{pmatrix}. \tag{2.34}$$

This configuration is similar to the Higgs sector in a type-II 2HDM, which will be discussed in section 2.2.2.

At tree-level, the Higgs potential reads:

$$\begin{aligned} V(\Phi_u, \Phi_d) = |\mu|^2\Phi_u^\dagger\Phi_u + |\mu|^2\Phi_d^\dagger\Phi_d + \frac{1}{4}(g^2 + g'^2)(\Phi_u^\dagger\Phi_u + \Phi_d^\dagger\Phi_d)^2 + \\ + \frac{1}{4}(g_w^2 - g'^2)\Phi_u^\dagger\Phi_u\Phi_d^\dagger\Phi_d - \frac{1}{2}g'^2\Phi_u^\dagger\Phi_d\Phi_d^\dagger\Phi_u, \end{aligned} \tag{2.35}$$

where μ is the higgsino mass parameter. The two complex Higgs doublets have eight degrees of freedom in total. Upon EWSB and subsequent minimization of the potential, the neutral components of the Higgs doublets acquire a vacuum expectation value:

$$\Phi_d^0 = \frac{1}{\sqrt{2}}\begin{pmatrix}v_d\\ 0\end{pmatrix}, \quad \Phi_u^0 = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v_u\end{pmatrix}, \tag{2.36}$$

such that the normalization $v^2 = v_u^2 + v_d^2 =$(246 GeV)^2. The ratio of the two vacuum expectation values defines the parameter:

$$\tan\beta \equiv \frac{v_u}{v_d}. \tag{2.37}$$

Three degrees of freedom give mass to the weak bosons, and the remaining five result in five physical Higgs states:

$$H^\pm = \Phi_d^\pm \sin\beta + \Phi_u^\pm \cos\beta, \tag{2.38}$$

$$A = \sqrt{2}(\mathrm{Im}\Phi_d^0 \sin\beta + \mathrm{Im}\Phi_u^0 \cos\beta), \tag{2.39}$$

$$h = -\left(\sqrt{2}\mathrm{Re}\Phi_d^0 - v_d\right)\sin\alpha + \left(\sqrt{2}\mathrm{Re}\Phi_u^0 - v_u\right)\cos\alpha, \tag{2.40}$$

$$H = \left(\sqrt{2}\mathrm{Re}\Phi_d^0 - v_d\right)\cos\alpha + \left(\sqrt{2}\mathrm{Re}\Phi_u^0 - v_u\right)\sin\alpha, \tag{2.41}$$

corresponding to two charged Higgs, a CP-odd pseudoscalar, and two CP-even scalars, respectively. By definition, it is assumed that $m_h \leq m_H$. The neutral states are conjointly denoted as $\phi = (h/H/A)$. The parameter α corresponds to the mixing angle of the two scalars. This is not a free parameter of the model, as it is determined by diagonalizing the squared mass matrix:

$$\mathcal{M}_0^2 = \begin{pmatrix} m_A^2 \sin^2\beta + m_Z^2 \cos^2\beta & -(m_A^2 + m_Z^2)\sin\beta\cos\beta \\ -(m_A^2 + m_Z^2)\sin\beta\cos\beta & m_A^2\cos^2\beta + m_Z^2\sin^2\beta \end{pmatrix}, \tag{2.42}$$

which yields the relation:

$$\cos^2(\beta - \alpha) = \frac{m_h^2(m_Z^2 - m_h^2)}{m_A^2(m_H^2 - m_h^2)}. \tag{2.43}$$

The squared masses of the CP-even scalars can be obtained as the eigenstates of $\mathcal{M}_0^2$ at tree level:

$$m_{H,h}^2 = \frac{1}{2}\left(m_A^2 + m_Z^2 \pm \sqrt{(m_A^2 + m_Z^2)^2 - 4m_Z^2 m_A^2 \cos^2 2\beta}\right). \tag{2.44}$$

Additionally, the following relation defines the mass of the charged Higgs as a function of m_A:

$$m_{H^\pm}^2 = m_A^2 + m_W^2. \tag{2.45}$$

Therefore, the Higgs sector parameters are defined at tree level by tanβ and one mass parameter, for which m_A is conventionally taken. After discussing the production and decay modes of neutral MSSM Higgs bosons, we will return to the relation in eq. 2.44 and its implications on the mass of the lightest scalar.

Neutral MSSM Higgs couplings and production mechanisms

The MSSM couplings of the neutral Higgs bosons to fermions and vector bosons are shown in table 2.1.

Returning to eq. 2.44, a constraint in the tree-level mass of the lightest scalar is found at:

$$m_h \leq m_Z|\cos 2\beta| \leq m_Z, \tag{2.46}$$

which is clearly incompatible with the measured mass of the discovered Higgs boson. However, the Higgs mass is subject to large radiative corrections, particularly from third-generation quarks and their superpartners, that bring this threshold towards higher values [82].

Another interesting conclusion can be derived from eq. 2.44. In the limit where $m_A \gg m_Z$, also known as the "decoupling limit", cos$(\beta - \alpha) \approx 0$, and the masses of the four

	ξ^u	ξ^d	ξ^ℓ	ξ^{V}
SM h(125)	1	1	1	1
h	$\frac{\cos\alpha}{\sin\beta}$	$-\frac{\sin\alpha}{\cos\beta}$	$-\frac{\sin\alpha}{\cos\beta}$	$\sin(\beta-\alpha)$
H	$\frac{\sin\alpha}{\sin\beta}$	$\frac{\cos\alpha}{\cos\beta}$	$\frac{\cos\alpha}{\cos\beta}$	$\cos(\beta-\alpha)$
A	$\cot\beta$	$\tan\beta$	$\tan\beta$	0

Table 2.1: Couplings of the three neutral Higgs bosons to up-type quarks (ξ^u), down-type quarks (ξ^d), leptons (ξ^l), and vector bosons (ξ^{V}), normalized to the SM values for h(125). Taken from [22].

heaviest Higgs bosons become degenerate, $m_A \approx m_H \approx m_{H^\pm}$. An important consequence of this approximation is that the couplings of h become SM-like, with exactly the same structure as the SM Higgs boson when $\cos(\beta-\alpha)=0$, also known as the alignment limit.

In the alignment limit, the coupling of H to vector bosons is suppressed. This, combined with the fact that A does not couple to vector bosons, influences the production of neutral MSSM Higgs bosons at the LHC. The production related to vector bosons (VBF and higgstrahlung) becomes less significant, and the gluon-gluon fusion and b-associated productions become the most relevant. In this thesis, a search for neutral Higgs bosons is conducted, targeting the b-associated production. This mechanism is enhanced at large tanβ, also as a consequence of radiative corrections from the SUSY sector, as will be shown in the following.

Impact of radiative corrections on Higgs production and decay modes

In the MSSM model, the tree-level bottom Yukawa coupling is subject to large one-loop radiative correction with tanβ enhanced contributions [83]. The resulting effective Lagrangian reads:

$$\begin{aligned}\mathcal{L} =& \frac{g}{2m_W}\frac{\tilde{m}_b}{1+\Delta_b}\Big[\tan\beta\, A\, i\, \bar{b}\gamma_5 b + \sqrt{2}V_{tb}\,\tan\beta\, H^+\bar{t}_L b_R \\ &+\left(\frac{\sin\alpha}{\cos\beta}-\Delta_b\frac{\cos\alpha}{\sin\beta}\right)h\bar{b}_L b_R - \left(\frac{\cos\alpha}{\cos\beta}+\Delta_b\frac{\sin\alpha}{\sin\beta}\right)H\bar{b}_L b_R\Big] + \text{h.c.},\end{aligned} \tag{2.47}$$

where Δ_b is the coupling modifier, $\tilde{m}_b$ represents the running b-quark mass, set to $\tilde{m}_b = \tilde{m}_b(m_t) \approx 2.97$ GeV. The multiplicative factor $1/(1+\Delta_b)$ originates from the resummation of the leading corrections to all orders, while the terms proportional to Δ_b stem from the mixing and coupling of the CP-even Higgs bosons to the b-quarks.

Δ_b comprises two leading contributions, corresponding to an $\mathcal{O}(\alpha_s)$ sbottom-gluino loop, and an $\mathcal{O}(\alpha_t)$ stop-higgsino loop, where α_s is the strong coupling constant and α_t is the coupling based on the top-quark Yukawa-coupling, $\alpha_t = y_t^2/4\pi$. The following definition of Δ_b applies in the limit where $\tan\beta \gg 1$:

$$\Delta_b \simeq \mu \tan\beta \left(\frac{2\alpha_s}{3\pi} \frac{m_{\tilde{g}}}{\max(m_{\tilde{b}_1}, m_{\tilde{b}_2}, m_{\tilde{g}})} + \frac{\alpha_t}{4\pi} \frac{A_t}{\max(m_{\tilde{t}_1}, m_{\tilde{t}_2}, \mu)} \right), \tag{2.48}$$

where $m_{\tilde{g},\tilde{b}_{1,2},\tilde{t}_{1,2}}$ are the gluino mass, and the bottom- and top-squarks masses derived from the coupling to each of the scalar doublets. A_t is the Higgs-stop coupling, and μ the higgsino mass parameter.

The relation in eq. 2.48 indicates that the coupling of b-quarks to the neutral Higgs bosons becomes enhanced as tanβ increases. Furthermore, negative values of μ may enhance this effect even further due to the $1/(1+\Delta_b)$ dependence. These results have a significant impact on the analysis presented in this thesis, which involves the production of neutral Higgs bosons in association with bottom quarks and decaying into a b-quark pair. The branching ratio can be approximated to [83]:

$$\mathrm{BR}(A \to b\bar{b}) \simeq \frac{9}{(1+\Delta_b)^2 + 9}, \tag{2.49}$$

yielding a total cross-section times branching ratio for $\phi = (h/H/A)$ given by:

$$\sigma(b\bar{b}\phi) \times \mathrm{BR}(\phi \to b\bar{b}) \simeq 2 \times \sigma(b\bar{b}A)_{\mathrm{SM}} \times \frac{\tan^2\beta}{(1+\Delta_b)^2} \times \frac{9}{(1+\Delta_b)^2 + 9}, \tag{2.50}$$

where $\sigma(b\bar{b}A)_{\mathrm{SM}}$ is the SM Higgs production cross-section for a mass equal to m_A.

Benchmark scenarios

In general, the MSSM sector comprises a very rich parameter space, containing several SUSY parameters that may affect the Higgs sector, such as the higgsino mass parameter μ, the SUSY breaking scale, and the masses of supersymmetric particles. However, constraining this large number of free parameters is rather impractical. Several benchmark scenarios have been developed to set the majority of these parameters while accommodating the experimental measurements [36], thereby reducing the parameter space to a more manageable and practical subset. In the case of MSSM, it is common practice to define benchmark scenarios where tanβ and m_{A} are the only free parameters. The MSSM scenarios that are employed in the model interpretation of the results presented in this thesis are discussed in the following. They have been developed by the LHC Higgs Cross-Section Working Group (LHCHXSWG) [84] and have been designed with the consideration that the mass of the lightest scalar h agrees with the SM observation of the 125 GeV Higgs boson, with a theoretical uncertainty of ±3 GeV [36, 85].

The m_{h}^{125} scenarios

The m_{h}^{125} scenarios [36, 86] have been designed considering that all superparticles are so heavy that they only have a minor impact on the production and decay modes of the predicted Higgs bosons. The loop-induced SUSY corrections to the couplings of the CP-even scalar are negligible for the lighter Higgs and for the heavier Higgs with masses up to 2 TeV. This results in a phenomenology similar to a type-II 2HDM, discussed in section 2.2.2, with MSSM-inspired couplings.

The following parameters are fixed in this scenario:

$$\begin{aligned}
&M_{Q_3} = M_{U_3} = M_{D_3} = 1.5\,\mathrm{TeV}, \quad M_{L_3} = M_{E_3} = 2\,\mathrm{TeV},\\
&M_1 = 1\,\mathrm{TeV}, \quad M_2 = 1\,\mathrm{TeV}, \quad M_3 = 2.5\,\mathrm{TeV},\\
&X_t = 2.8\,\mathrm{TeV}, \quad A_b = A_\tau = A_t,
\end{aligned}$$

where M_{Q,U,D_3} are the masses of the third generation squarks, M_{L,E_3} the third generation slepton masses, $M_{1,2,3}$ the gaugino masses, A_X is the Higgs-sfermion coupling, and the parameter $X_t = A_t - \mu \cot\beta$. In the standard m_{h}^{125} scenario, the higgsino mass parameter μ is set to 1 TeV. This value guarantees that the higgsinos are as heavy as the gauginos, resulting in the largest SUSY loop corrections arising from the coupling of the Higgs to b-quarks, which is enhanced at large tanβ as previously discussed. This enhancement is particularly stronger for the heavy Higgs bosons in the limit when $m_A \gg m_Z$. A set of alternative scenarios is proposed considering μ is set to 1, -1, -2 and -3 TeV, which enhances the bottom Yukawa couplings even further, as $\Delta_b \propto \mu \tan\beta$ [86]. The SM parameters are set following the recommendations of the LHCHXSWG [75]

The $m_{\mathrm{h}}^{\mathrm{mod+}}$ scenario

The $m_{\mathrm{h}}^{\mathrm{mod+}}$ scenario was designed to provide an update on the $m_{\mathrm{h}}^{\mathrm{max}}$ scenario [85, 87] in the light of the Higgs boson discovery at 125 GeV, as the latter had been designed to maximize the value of the mass of the light scalar boson yielding a value larger than the observed. A modification on the mixing in the stop sector is introduced, reducing $|X_t/M_{SUSY}|$, where X_t is the stop mixing parameter and M_{SUSY} the mass of the third generation sfermions. This change is motivated by the fact that $|X_t/M_{SUSY}|$ introduces a positive contribution to m_h from radiative corrections. This scenario considers the absolute value of $|X_t/M_{SUSY}|$ positive. The parameters fixed in this benchmark scenario are:

$$\begin{aligned}
&m_t = 173.2\,\mathrm{GeV}, \quad M_{SUSY} = 1000\,\mathrm{GeV}, \quad \mu = M_{gaugino} = 200\,\mathrm{GeV},\\
&m_{\tilde{g}} = m_{\tilde{q}_{1,2}} = 1500\,\mathrm{GeV}, \quad M_{\tilde{l}_3} = 1000\,\mathrm{GeV}, \quad A_b = A_t = A_\tau,\\
&X_t^{OS} = 1.5\, m_{SUSY}, \quad X_t^{\overline{MS}} = 1.6\, m_{SUSY}, \quad A_{f\neq b,t,\tau} = 0,
\end{aligned}$$

where m_t is the top-quark mass, μ is the higgsino mass parameter, $M_{gaugino}$ are the gaugino masses, $m_{\tilde{g}}$ is the gluino mass, $m_{\tilde{q}_{1,2}}$ are the masses of the first and second generation squarks, and $M_{\tilde{l}_3}$ is the mass of the third generation sleptons. The parameters X_t^{OS} and $X_t^{\overline{MS}}$ are the stop mixing parameter, calculated in an on-shell (OS) and $\overline{MS}$ schemes [85], and A_X denotes the Higgs-sfermion couplings.

It is relevant to note that this scenario considers relatively light SUSY particles, particularly charginos and neutralinos. This results in a reduced BR of the Higgs bosons in the $b\bar{b}$ and $\tau\bar{\tau}$ decay channels at tanβ values up to 20, while for large tanβ values the decay into b-quarks dominates.

The hMSSM scenario

In the hMSSM scenario [88–91] the lightest CP-even scalar also aligns with the 125 GeV observation. On the other hand, a simplified approach is employed, which results in the Higgs sector being independent from the SUSY sector. All superparticles are considered to be so heavy that they do not affect the Higgs production and decay modes, and most SUSY corrections to the Higgs couplings are negligible, except those from the top/stop loops in the neutral CP-even Higgs bosons. The mass of the heaviest CP-even scalar is defined by the relation:

$$m_H^2 = \frac{(m_A^2 + m_Z^2 - m_h^2)(m_Z^2 \cos^2\beta + m_A^2 \sin^2\beta) - m_A^2 m_Z^2 \cos^2 2\beta}{m_Z^2 \cos^2\beta + m_A^2 \sin^2\beta - m_h^2}, \qquad (2.51)$$

while the masses of the charged Higgs bosons are determined by eq. 2.45. The Higgs sector is, therefore, fully defined by tanβ and m_{A} in this scenario.

Computation of the cross-sections and Branching ratios in the MSSM benchmark scenarios

The b-associated production mechanism in a hadron collider may occur as per the so-called four-flavour scheme, where b-quarks typically originate from processes like gluon decays. The bottom quark is treated as a massive object and never appears in the initial state. This assumption provides a better description of processes where the scale, Q, is comparable to the bottom-quark mass. In contrast, in the 5FS, the b-quark is treated as a massless particle and bottom parton densities are introduced. Diagrams corresponding to these processes are shown in section 5.1, Fig. 5.1, when the signal and background processes relevant for this thesis are discussed.

One way to harmonize the predictions from 4FS and 5FS is described by the Santander Matching Scheme [92], which was employed in the inclusive cross-section computation in the $m_{\mathrm{h}}^{\mathrm{mod}+}$ and hMSSM benchmark scenarios. This method consists of a weighting procedure for calculating the total inclusive cross-section for bottom-quark associated

Higgs production combined, depending on the value of the Higgs-boson mass. The total combined 4FS and 5FS cross-section is defined as follows:

$$\sigma^{\text{Total}}_{\text{Santander}} = \frac{\sigma^{4FS} + w\sigma^{5FS}}{1+w}, \tag{2.52}$$

where σ^{4FS} (σ^{5FS}) represents the total inclusive cross-section in the 4FS (5FS), and $w = \ln\left(\frac{m_\phi}{m_b}\right) - 2$. Consequently, when the mass ratio tends to be infinite, the 5FS becomes dominant, while as the ratio becomes smaller, the 4FS dominates. Both contributions get a similar weight in the 100 GeV Higgs mass regime, consistent with the observed agreement of their contributions in this region [92].

A different approach is followed in calculating the inclusive b-associated production cross-section in the m_{h}^{125} scenarios, where the "fixed order plus next-to-leading log" or FONLL matching scheme [93, 94] was employed. In this method, the total cross-section is defined as:

$$\sigma^{\text{Total}}_{\text{FONLL}} = \sigma^{4FS} + \sigma^{5FS} - \sigma^{overlap} \tag{2.53}$$

Here, the 5FS calculations, which consist of the sum logarithms of the form $\log(\mu^2/m_b^2)$ in orders of α_s, where μ is the process scale and m_b is the bottom-quark mass, is subject to a polynomial expansion. A finite number of terms are then matched with their counterparts in the massive scheme, maintaining consistency with the 4FS results. The subtraction of the overlapping term, $\sigma^{overlap}$, ensures that the contributions unique to each flavour scheme are accurately combined. The logarithmic accuracy of the 5FS and the fixed-order precision of the 4FS are then kept.

Both the Santander and FONLL matching schemes provide robust frameworks for harmonizing the predictions from the 4FS and 5FS calculations and are, in principle, equally applicable to computing the inclusive b-associated Higgs production cross-section. In both cases, the 4FS calculations are performed up to NLO [95, 96], while the 5FS computations are done up to NNLO [97]. The Branching ratios are determined using `HDecay` [98, 99] and `FeynHiggs v.2.12.0` [100–103], which take into account one and two-loops SUSY corrections, respectively. The uncertainties concerning the value of α_s, the choice of parton distribution functions, and the uncertainty in the normalisation and factorization scales were also centrally provided.

2.2.2 The Two Higgs Doublet Model

In contrast to the MSSM model, which is attached to SUSY, the Two Higgs Doublet Model (2HDM) [22] is disconnected from a broader theory. The 2HDM, as the name indicates, is characterized by extending the SM scalar sector with another complex doublet, making

MSSM a particular realization of a 2HDM. The Higgs sector in 2HDM models will be discussed in the following.

The 2HDM Higgs sector

The Higgs potential in 2HDM, assuming CP conservation in the Higgs sector, reads:

$$\begin{aligned} V =& m_{11}^2 \Phi_1^\dagger \Phi_1 + m_{22}^2 \Phi_2^\dagger \Phi_2 - m_{12}^2 \left(\Phi_1^\dagger \Phi_2 + \Phi_2^\dagger \Phi_1 \right) + \\ &+ \frac{\lambda_1}{2} \left(\Phi_1^\dagger \Phi_1 \right)^2 + \frac{\lambda_2}{2} \left(\Phi_2^\dagger \Phi_2 \right)^2 + \\ &+ \lambda_3 \Phi_1^\dagger \Phi_1 \Phi_2^\dagger \Phi_2 + \lambda_4 \Phi_1^\dagger \Phi_2 \Phi_2^\dagger \Phi_1 + \\ &+ \frac{\lambda_5}{2} \left[\left(\Phi_1^\dagger \Phi_2 \right)^2 + \left(\Phi_2^\dagger \Phi_1 \right)^2 \right], \end{aligned} \tag{2.54}$$

where m_{11}^2, m_{12}^2 and m_{22}^2, λ_1, λ_2, λ_3, λ_4, and λ_5 are real parameters. At the same time, Φ_1 and Φ_2 are the two Higgs doublets in the 2HDM. Each Φ_i is a complex scalar field with a charged and a neutral component:

$$\Phi_i = \begin{pmatrix} \phi_i^+ \\ \phi_i^0 \end{pmatrix}, \quad i = 1, 2$$

This structure results in five physical Higgs, as in the MSSM sector. In 2HDM, flavour-changing neutral currents (FCNC) are not forbidden at tree level. However, these are strongly suppressed experimentally. This issue can be circumvented by imposing that each group of fermions can only couple to one Higgs doublet, as demonstrated by the Paschos-Glashow–Weinberg theorem [104, 105]. A Z_2 symmetry is imposed, and four types of 2HDM can be defined according to the way up-type and down-type quarks and leptons couple to the Higgs doublets. They are summarized in table 2.2.

Type	Symmetry	*up*-type quarks	*down*-type quarks	charged leptons
type-I	$\Phi_1 \to -\Phi_1$	Φ_2	Φ_2	Φ_2
lepton-specific	$\Phi_1 \to -\Phi_1$	Φ_2	Φ_2	Φ_1
type-II	$\Phi_1 \to -\Phi_1$, $d_R \to -d_R$	Φ_2	Φ_1	Φ_1
flipped	$\Phi_1 \to -\Phi_1$, $d_R \to -d_R$	Φ_2	Φ_1	Φ_2

Table 2.2: CP-conserving 2HDM models in the absence of FCNC, indicating which doublet couples to which type of fermion.

Neutral Higgs couplings

The Lagrangian density for the Yukawa coupling in 2HDM can be written as:

$$\mathcal{L}_{\text{Yukawa}}^{\text{2HDM}} = -\sum_{f=u,d,\ell} \frac{m_f}{v}\left(\xi_h^f \bar{f} f h + \xi_H^f \bar{f} f H - i\xi_A^f \bar{f}\gamma_5 f A\right) + \mathcal{L}_{\text{Yukawa},H^{\pm}}^{\text{2HDM}}, \tag{2.55}$$

where $\mathcal{L}_{\text{Yukawa},H^{\pm}}^{\text{2HDM}}$ corresponds to the Yukawa Lagrangian density of the charged Higgs bosons. The Yukawa couplings can be determined with respect to the SM values, depending on the type of 2HDM. They are summarized in table 2.3.

	Type I	Type II	Lepton-specific	Flipped
$\xi^u h$	$\cos\alpha/\sin\beta$	$\cos\alpha/\sin\beta$	$\cos\alpha/\sin\beta$	$\cos\alpha/\sin\beta$
$\xi^d h$	$\cos\alpha/\sin\beta$	$-\sin\alpha/\cos\beta$	$\cos\alpha/\sin\beta$	$-\sin\alpha/\cos\beta$
$\xi^\ell h$	$\cos\alpha/\sin\beta$	$-\sin\alpha/\cos\beta$	$-\sin\alpha/\cos\beta$	$\cos\alpha/\sin\beta$
$\xi^u H$	$\sin\alpha/\sin\beta$	$\sin\alpha/\sin\beta$	$\sin\alpha/\sin\beta$	$\sin\alpha/\sin\beta$
$\xi^d H$	$\sin\alpha/\sin\beta$	$\cos\alpha/\cos\beta$	$\sin\alpha/\sin\beta$	$\cos\alpha/\cos\beta$
$\xi^\ell H$	$\sin\alpha/\sin\beta$	$\cos\alpha/\cos\beta$	$\cos\alpha/\cos\beta$	$\sin\alpha/\sin\beta$
$\xi^u A$	$\cot\beta$	$\cot\beta$	$\cot\beta$	$\cot\beta$
$\xi^d A$	$-\cot\beta$	$\tan\beta$	$-\cot\beta$	$\tan\beta$
$\xi^\ell A$	$-\cot\beta$	$\tan\beta$	$\tan\beta$	$-\cot\beta$

Table 2.3: Yukawa couplings of *up*-type quarks, *down*-type quarks, and charged leptons to the neutral Higgs bosons h, H, A in the four 2HDM types. Taken from [22].

It can be derived from table 2.3 that the coupling of H and A to b-quarks is enhanced for large values of tanβ in the *type-II* and *flipped* models. Moreover, the analysis performed in this thesis has a unique sensitivity in the *flipped* model, where leptons are disfavoured. On the other hand, the couplings of the neutral Higgs bosons to vector bosons are similar across the four defined 2HDM types. They are defined as:

$$g_{hVV}^{\text{2HDM}}/g_{hVV}^{\text{SM}} = \sin(\beta-\alpha), \tag{2.56}$$

$$g_{HVV}^{\text{2HDM}}/g_{hVV}^{\text{SM}} = \cos(\beta-\alpha), \tag{2.57}$$

$$g_{AVV}^{\text{2HDM}}/g_{hVV}^{\text{SM}} = 0, \tag{2.58}$$

where g_{hVV}^{2HDM} is the coupling of the 2HDM h boson to vector bosons, and g_{hVV}^{SM} the SM Higgs boson coupling to vector bosons.

Benchmark scenarios

2HDM models feature a very rich phenomenology that may be simplified in certain benchmark scenarios. A well-known example is the “scenario G” [106, 107], which is inspired by the MSSM Higgs potential. The following parameters are fixed:

$$\lambda_5 = 0, \quad m_{12}^2 = \frac{1}{2} m_A^2 \sin 2\beta, \quad m_h = 125\,\mathrm{GeV}, \quad \text{and} \quad m_A = m_H = m_{H^\pm}, \tag{2.59}$$

leaving three free parameters: tanβ, m_{A}, and cos$(\beta - \alpha)$. To be able to interpret the results of the analysis performed in this thesis within 2HDM models and set constraints in the parameter space, a three-dimensional scan in the free parameters performed in the context of the studies discussed in Ref. [39,40] were employed. The b-associated production cross-sections of the neutral scalars and branching ratios of decays into a b-quark pair were derived at NNLO precision using the `SusHi v.1.6.1` [108], `2HDMC v.1.7.0` [109] and `LHAPDF v.6.1.6` [110] packages. Relevant figures showing calculated branching fractions of the A and H bosons in the *type-II* and *flipped* scenarios are summarized in Appendix G.

chapter 3

Experimental setup

The search performed in this thesis analyzes collision data produced at the largest and most powerful accelerator ever built, the Large Hadron Collider (LHC), at the European Organization for Nuclear Research (CERN). The data was collected by the Compact Muon Solenoid (CMS) detector located at the LHC ring during proton-proton (p-p) collisions at a centre-of-mass energy of 13 TeV in 2017 and 2018. This chapter introduces the LHC accelerator complex and the CMS detector apparatus.

3.1 The Large Hadron Collider

The Large Hadron Collider [33] was designed to probe the SM predictions at a higher energy scale, search for and measure the properties of the SM Higgs boson if found, and search for new physics. It is a 26.7 km radius synchrotron accelerator of protons and heavy ions, located 100 m underground in the Swiss-French border near Geneva, occupying the tunnel previously used for the Large Electron Positron collider (LEP) [111].

The LHC is the last step of CERN's current complex accelerator chain [112, 113]. Before being injected into the main LHC ring, protons are "prepared" by a pre-accelerator system that gradually increases their energy. A linear accelerator, Linac4, provides the initial acceleration to negatively charged hydrogen gas, which is stripped of its electrons, leaving only protons before entering the next stages. The protons are then injected into a series of progressively larger synchrotrons that boost them to up to 450 GeV before entering the LHC. These are, in order, the Booster, the Proton Synchrotron (PS) accelerator, and the Super Proton Synchrotron (SPS). A schematic view of the CERN's accelerator complex is shown in Fig. 3.1.

Protons enter the two LHC rings as counter-rotating beams of up to 2808 bunches, containing around 1.2×10^{11} protons each. The beams are accelerated to up to 7 TeV by superconductive radio frequency (RF) cavities that operate at 400 MHz, and are guided

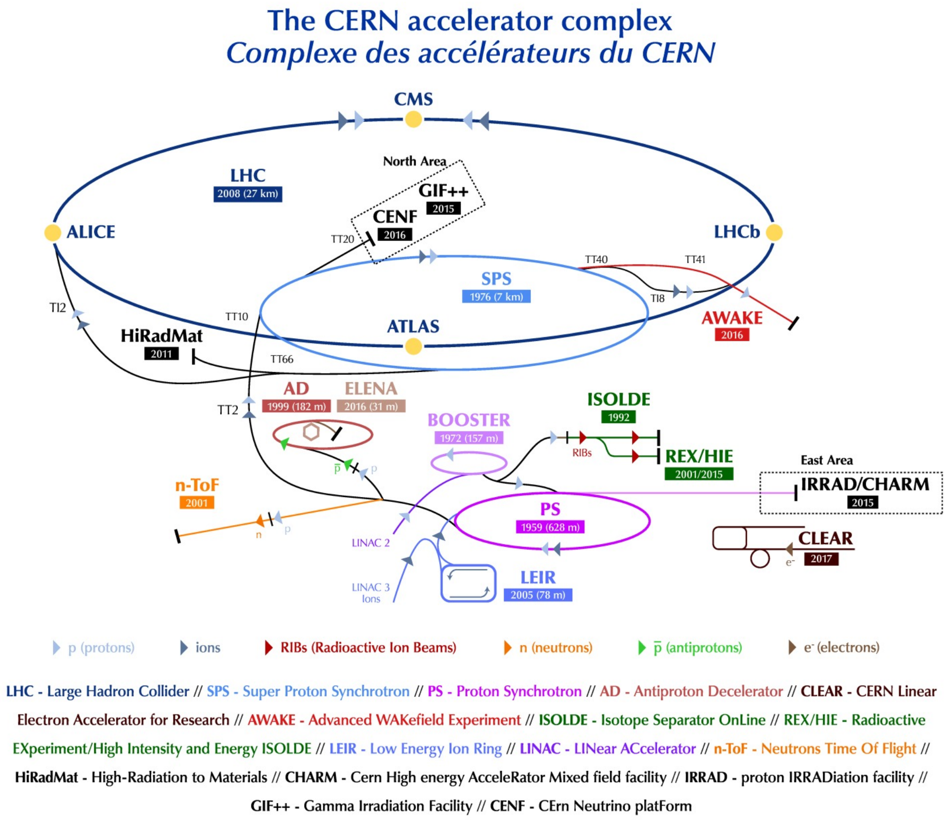


Figure 3.1: Schematic view of the LHC accelerator facility. Taken from [114].

using 8.3 T superconducting dipole magnets. A system of quadrupole magnets focuses the beam by squeezing it vertically and horizontally. To maintain their superconductive properties, the magnets are kept at a temperature of 2 K using liquid helium. Further sextupole, octupole, and decapole magnets are used to rectify imperfections in the beam.

The two beams collide at four interaction points (IPs) along the LHC tunnel every 25 ns, where the four major experiments at the LHC are located. These are: ALICE (A Large Ion Collider Experiment) [115], ATLAS (A Toroidal LHC ApparatuS) [116], CMS (Compact Muon Solenoid) [34], and LHCb (LHC beauty) [117]. The number of simultaneous collisions per particle bunch crossing is defined as pileup (PU).

In a collider of symmetric beams with energy E_{beam}, the centre-of-mass energy is defined as $\sqrt{s} = 2 \cdot E_{beam}$.

The number of events per time interval produced for a certain process with cross-section σ is related to the instantaneous luminosity $\mathcal{L}$ as follows:

$$\frac{dN}{dt} = \mathcal{L} \cdot \sigma \tag{3.1}$$

$\mathcal{L}$ is, therefore, a collider's key performance indicator. It can be determined as:

$$\mathcal{L} = \frac{N_p^2 n_b f_{rev}}{4\pi\sigma_x\sigma_y} F, \tag{3.2}$$

where N_p represents the number of protons per bunch, n_b is the number of bunches in the beam, f_{rev} denotes the revolution frequency, and σ_x and σ_y are the horizontal and vertical beam sizes at the collision point, respectively. F is a geometrical factor accounting for the crossing angle of the beams.

Special LHC runs, called Van der Meer (VdM) scans [118], are performed to estimate σ_x and σ_y in order to measure the instantaneous luminosity. These scans involve systematically varying the beam positions in both transverse directions. Dedicated detectors, known as "luminometers", record the event rate as a function of the beam separation. The values of σ_x and σ_y are then extracted from a fit to the recorded data. During 2016, the LHC reached its design peak instantaneous luminosity, namely 10^{34} cm^{-2}s^{-1}. Since then, it has regularly achieved twice this value.

The integrated luminosity $\mathcal{L}_{int}$ provides information about the amount of delivered collision data and is obtained from the instantaneous luminosity as:

$$\mathcal{L}_{int} = \int \mathcal{L} dt. \tag{3.3}$$

The LHC has completed two data-taking periods. The first period, Run 1, took place between 2010 and 2012. During this time, protons collided at a centre-of-mass energy of 7 TeV (2010-2011) and 8 TeV (2012). The second period, Run 2, took place from 2015 until 2018, during which protons collided at a centre-of-mass energy of 13 TeV. The CMS detector collected the data analyzed in this thesis during this period. The integrated luminosity delivered by the LHC and recorded by the CMS detector during Run 2 is shown in Fig. 3.2, along with the per-year average PU distributions.

At the time of writing, the third LHC data-taking period, Run 3, is ongoing, and data from p-p collisions is being taken at centre-of-mass energy of 13.6 TeV. Run 3 is planned to finalize at the end of 2025, and will be followed by the LHC's third long shutdown (LS) period that precedes the next and more powerful phase of the LHC, the High Luminosity LHC (HL-LHC) [120]. During the HL-LHC period, the instantaneous luminosity is expected to increase to up to 7.6×10^{34} cm^{-2}s^{-1}. Targeting this phase, all accelerator complex and experimental facilities, including CMS, will undergo a major upgrade [120–124]. The HL-LHC is scheduled to begin data-taking in 2029 and is expected to increase the total integrated luminosity to more than 3000 fb^{-1} by 2038. Protons will collide at 14 TeV with an anticipated average PU of 200 interactions per bunch crossing.

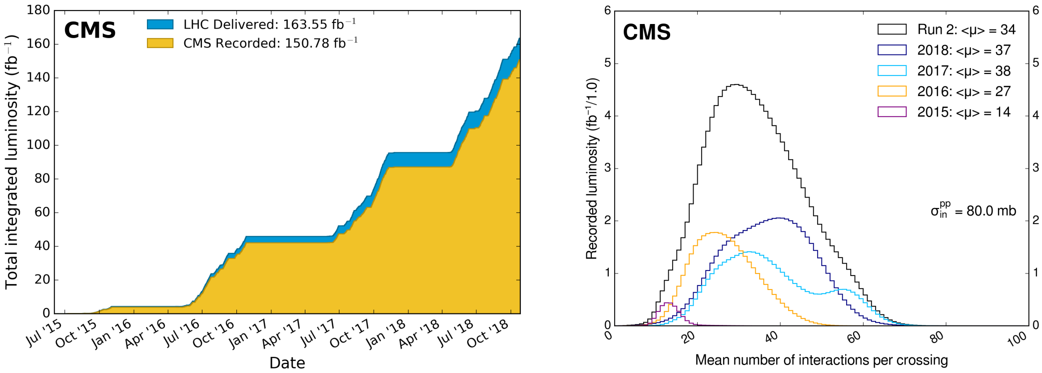


Figure 3.2: Integrated luminosity delivered by the LHC (blue) and recorded by the CMS detector (yellow) during Run 2 (left). Average number of interactions per bunch-crossing during p-p collisions in Run 2 (right). Taken from [119].

3.2 The Compact Muon Solenoid detector

The Compact Muon Solenoid is a multi-purpose detector installed 100 m underground in a cavern along the LHC ring close to the Cessy village in France. It was conceived to investigate a wide range of physics phenomena in alignment with the LHC physics program. This is reflected in its configuration, which has been foreseen to provide excellent identification of charged leptons, hadrons, and photons, as well as exceptional resolution of diphoton and two-lepton reconstructed mass, nearly 1% at 100 GeV [125], and missing transverse energy (MET). In particular, as its name indicates, the CMS detector has been optimized for the detection and measurement of muons.

The CMS detector has a cylindrical shape, measuring approximately 21 meters long and 15 meters in diameter and weighing about 14,000 tonnes. Its design facilitates large geometrical coverage to detect the products of a collision. An onion-like layering of several sub-detectors with specific functions characterizes its compact structure. Within CMS, a superconductive solenoid generates a magnetic field of 3.8 T that provides a strong bending power to charged particles that contributes to the determination of their momentum and charge. Inside it, and closest to the IP, is the silicon tracker, followed by a scintillating-crystals-based electromagnetic calorimeter (ECAL). Next is the brass/scintillator hadronic calorimeter (HCAL), whose inner configuration is directly surrounded by the solenoid, followed by an outer HCAL system. Surrounding it is an 11000-ton steel return yoke destined to confine the magnetic field and stop the remaining particles, except muons and neutrinos. The return yoke is interleaved with a multi-layered muon detection system. Figure 3.3 provides a schematic overview of the CMS detector components.

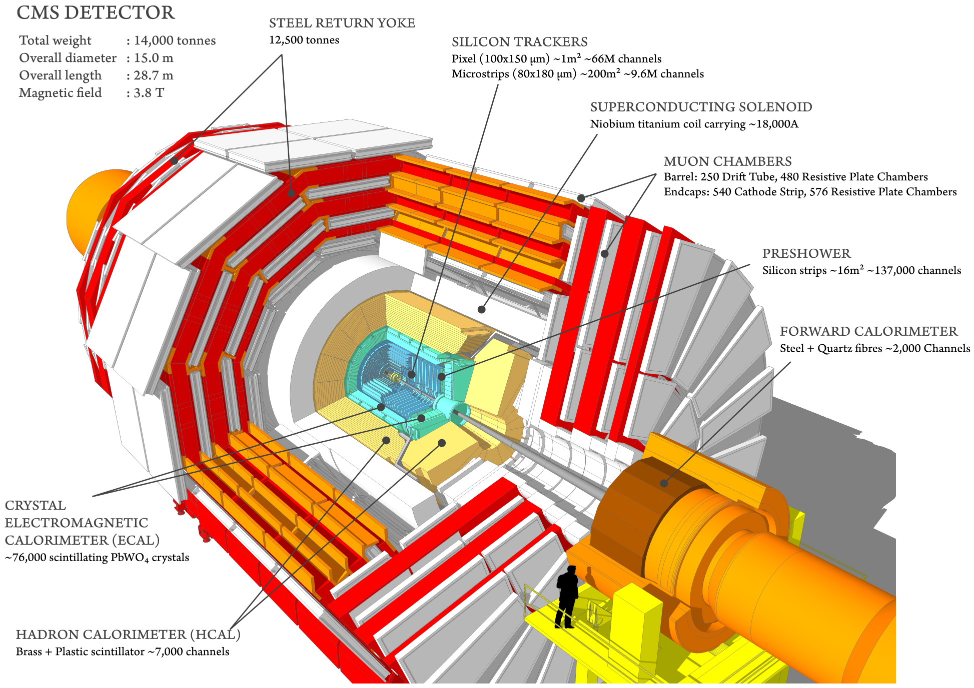


Figure 3.3: Schematic view of the CMS detector and its main features. Taken from [126].

The conventional CMS coordinate system is centred at the interaction point, with the z-axis aligned with the beam direction in a counter-clockwise orientation. The x and y-axes point vertically and horizontally towards the centre of the LHC ring, respectively. The angle ϕ is measured from the x-axis in the x-z plane, while θ is measured from the z-axis in the x-y plane, as illustrated in Fig. 3.4.

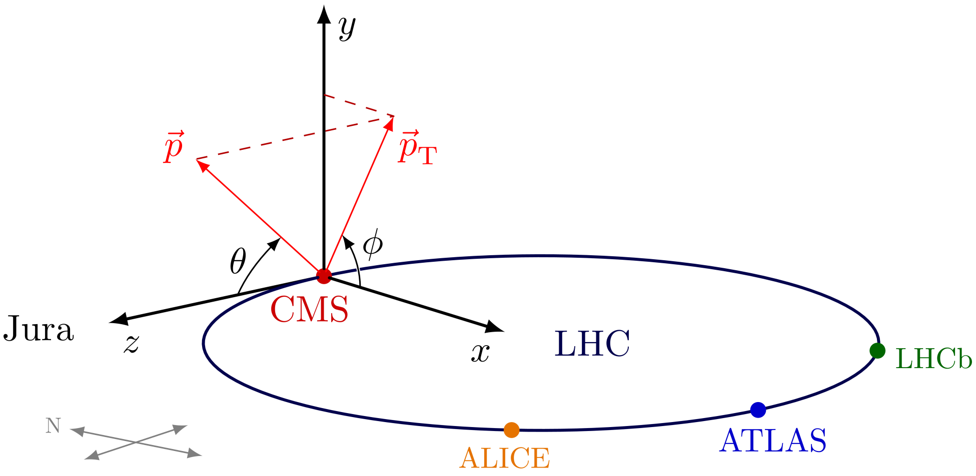


Figure 3.4: The coordinate system used by the CMS experiment. Taken from [126].

The rapidity is defined as $y = \frac{1}{2}\ln\frac{E+p_z}{E-p_z}$, where E and p_z are the energy and momentum along the beam axis, respectively. This variable is particularly useful because differences in rapidity are invariant under Lorentz boosts. The pseudorapidity, expressed as $\eta = -\ \ln(\ \tan(\theta/2))$, is generally used as an alternative to the polar angle θ. In the relativistic limit, it is equivalent to y.

The angular separation $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ is a measure combining differences in pseudorapidity $\Delta\eta$ and azimuthal angle $\Delta\phi$. This metric is invariant under Lorentz boosts along the beam axis, making it particularly useful in collider experiments for characterizing the relative distances between particles.

The momentum and energy projected in the transverse plane to the beam direction are denoted as p_{T}, and E_T, respectively. The MET in the transverse plane is denoted by E_T^{miss}.

Recording events in real-time (online) is significantly challenging due to high instantaneous luminosity and number of simultaneous interactions [127]. To address these challenges, CMS utilizes a two-step triggering system that significantly optimizes the data recording rate. The following sections cover detailed explanations of the CMS subdetectors and the triggering and data acquisition systems employed at CMS.

3.2.1 Tracking system

The CMS tracker [125, 128, 129], the innermost subdetector of the CMS apparatus, surrounds the IP and, therefore, operates under demanding conditions of high occupancy and intense radiation levels. To guarantee an efficient performance in this environment, the tracker is constructed entirely using silicon-based sensors, which also enables the construction of a highly granular detector with rapid response capabilities. The CMS tracker consists of two components covering a pseudorapidity range of $|\eta| < 2.5$: the Pixel and Silicon Strip detectors. Together, they provide an accurate reconstruction of the trajectory of charged particles (tracks) from the energy deposits in the sensitive part of the tracker modules (hits). Furthermore, the CMS tracker is instrumental in accurately reconstructing primary and secondary vertices. Figure 3.5 shows a schematic view of one-quarter of the current CMS tracker in the radial (r) and z plane.

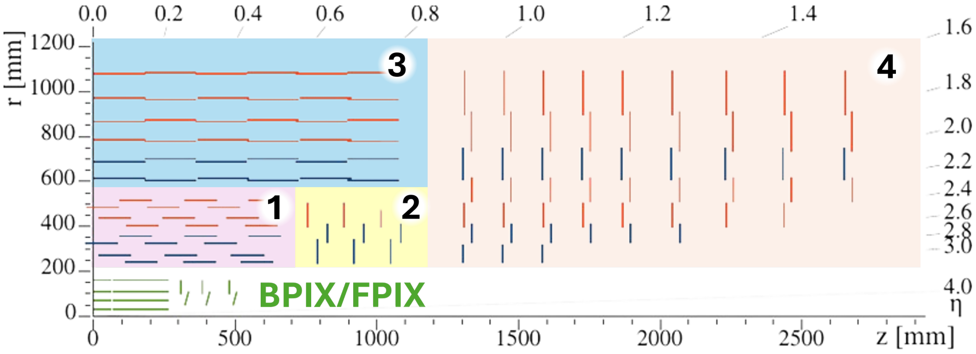


Figure 3.5: Sketch of one-quarter of the CMS tracking system in r-z view. Segments represent the detector's modules. The pixel detector, comprising the barrel pixel (BPIX) and the forward pixel (FPIX), is shown in green. The silicon strip detector is divided into sections: (1) the tracker inner barrel (TIB), (2) the tracker inner disks (TID), (3) the tracker outer barrel (TOB), and (4) the tracker endcap (TEC). Single-sided and double-sided strip modules are shown as red and blue segments, respectively. Adapted from [130].

The Pixel detector

Given its placement and highly granular construction, the Pixel detector is the main contributor to the optimal vertex reconstruction. However, this also implies it is located in a particularly harsh radiation environment.

To ensure efficient tracking, the pixel detector was upgraded in the 2016/2017 year-end technical stop. The new system is referred to as the *CMS Phase 1* pixel detector [131]. Its design includes 1856 modules arranged in four concentric barrel layers (BPIX) and three endcap disks (FPIX) on each side, comprising nearly 124 million pixel detectors in total, with a size of 100 μm × 150 μm each. A comparison with the original design, which featured three BPIX layers and two FPIX disks, is shown in 3.6. The upgraded system also includes new readout chips designed to cope with the higher occupancy foreseen in the data-taking from 2017.

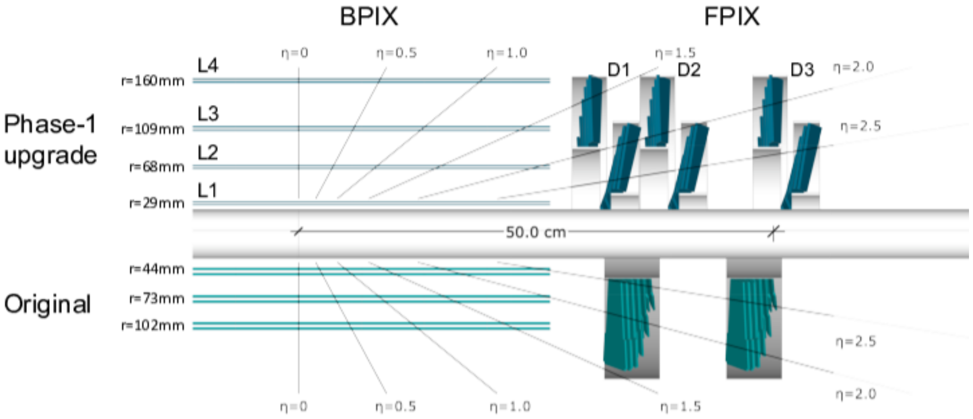


Figure 3.6: Longitudinal view of the Phase 1-upgraded pixel detector compared to the original detector layout [132].

In the pixel detector, a particle's passage induces a charge in neighbouring pixels, which is used to reconstruct hits. The shapes of these clusters are analyzed and fitted to achieve a precise hit position, typically with a resolution on the order of 10 μm [131, 133].

The Silicon Strip detector

The silicon strip detector is composed of 15,148 strip modules distributed in the four subsystems, introduced in Fig. 3.5. These are four (six) cylindrical TIB (TOB) layers and three (nine) TID (TEC) rings placed at each side of the endcap region. Strip modules can be single- or double-sided, containing about 9.3 million strip sensors in total [125]. The strip modules feature different geometries [134, 135], with an inter-strip distance (strip pitch) varying from 80 μm to 280 μm. Double-sided modules are built from two single-sided modules glued back-to-back at a stereo angle of 100 mrad between the strips [136] to enhance the spatial resolution by providing a measurement in $r-\phi$ and $r-z$ coordinates.

As in the pixel detector, the hits are reconstructed from the induced charge produced by a particle. The charge distribution across adjacent strips is analyzed and fitted to determine the hit position [136, 137]. The overall hit resolution is in the order of 10 μm in the $r-\phi$ coordinates and in the order of 100 μm in the z-direction [138].

The reconstruction of tracks and vertices with the information of both Pixel and Strip detectors will be described in section 4.2.1. The p_{T} of charged particles is measured from the curvature of the tracks due to the strong magnetic field the tracker is immersed in. The particle's charge is obtained from the sign of the track's curvature. The reconstructed track also determines the parameters ϕ and η.

The optimal performance of the CMS tracker is crucial for physics analyses, including the one presented in this thesis. The precise position and resolution of hits are essential for the accurate reconstruction of tracks and vertices. This accuracy is vital for reconstructing physics objects such as jets and muons and for various aspects of CMS data analysis, such as b-tagging algorithms. This accuracy can only be guaranteed if the position and orientation of each module are known with a precision comparable to the designed 10 μm hit resolution [139] of the CMS tracker. However, a mechanical alignment exceeds this value by a factor of 10. Furthermore, movements and distortions of the tracker modules are expected due to changes in the data-taking conditions, e.g. long-term radiation exposure, magnet cycles, or temperature changes. Corrections to the position and orientation of the sensors, also known as "tracker alignment", can only provide the desired accuracy.

Since the beginning of my PhD. project, I joined the CMS tracker alignment efforts. The strategies to derive these corrections, as well as the performance at the beginning and during the Run 3 data-taking period, are discussed in Appendix A. The work I contributed to resulted in the publications in Refs. [140–142]. A general overview of the CMS tracker performance during Run 3 is discussed in Appendix B, published in Ref. [129].

3.2.2 Calorimeters

Surrounding the tracker are the CMS calorimeter systems. They are scintillation-based detectors which provide a reliable and well-understood method of measuring the energy of a particle that interacts with the detector either through electromagnetic or strong processes [143, 144]. This interaction produces a shower of secondary particles which excite the atoms of the scintillator material, leading to the emission of photons, generally in the visible light spectra. The induced luminescence is proportional to the energy deposited by the incoming particle, allowing for precise energy measurements.

Two types of calorimeters are employed at CMS: the electromagnetic calorimeter (ECAL), to measure the energy of electrons and photons, and the hadronic calorimeter (HCAL), design to measure the energy of charged and neutral hadrons.

The Electromagnetic Calorimeter

The CMS ECAL [125, 145] consists of over 75,000 scintillating lead tungsten (PbWO4) crystals, chosen for their high density, radiation hardness, and fast response time. When interacting with the material, photons and electrons produce a particle shower that deposits their energy mainly through ionization and excitation of the material. A scintillation light is produced, and signals from individual crystals in the ECAL are identified and converted into energy measurements. Adjacent energy deposits are grouped into clusters, representing the total energy from a single particle interaction [146].

The ECAL detector is separated into a barrel section (EB) in the pseudorapidity range of $|\eta| < 1.479$, and an endcap section (EE), which extends the coverage up to $|\eta| < 3$. A preshower system (ES) featuring lead radiators and silicon strip sensors, which improve the spatial resolution, is installed in front of EE to enhance the identification of neutral pi-meson decays. Fig. 3.7 shows a sketch of one-quarter of the ECAL.

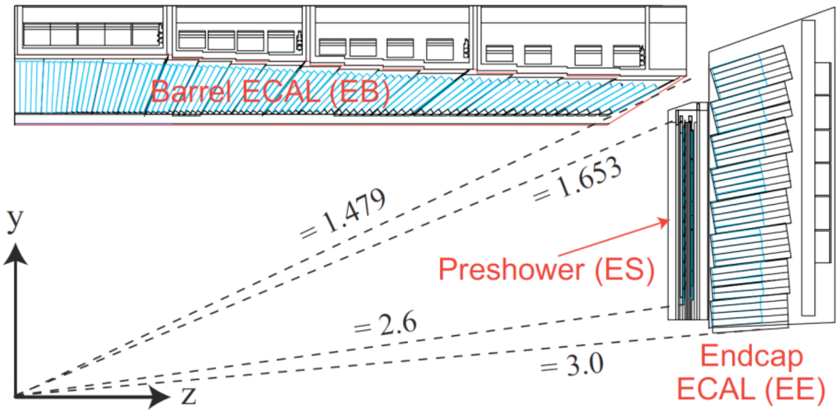


Figure 3.7: Schematic layout of one-quarter of the CMS ECAL system. Taken from [147].

The ECAL energy resolution (σ) can be parametrized as a function of the energy of the incoming electron/photon (E) as [34, 148]:

$$\frac{\sigma}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c, \tag{3.4}$$

where a accounts for stochastic contributions, b accounts for electronic noise and pile-up, and c represents the constant term that accounts for contributions independent from the energy of the particle. The symbol $\oplus$ indicates a quadratic sum. The a, b and c parameters are experimentally determined in beam-test runs and found to be 2.8%, 12% and 0.30%, respectively.

The Hadronic Calorimeter (HCAL)

Adjacent to the ECAL, the HCAL measures the energy of charged and neutral hadrons, such as protons, neutrons, and pions [125, 149, 150], by detecting the cascades of particles (hadronic showers) they create via strong interactions within the detector material [143]. Energy deposits from particles are recorded by individual scintillator tiles within the HCAL. The signals from these tiles are converted into energy measurements. As in ECAL, signals from neighbouring tiles with significant energy deposits are grouped to form clusters [151].

As the CMS solenoid coil restricts the available space immediately after the ECAL, the HCAL design features an outer detector (HO) complementing the barrel calorimeter inside the magnet (HB). HB is a sampling calorimeter, meaning that the material where the particle shower is produced differs from the one used to measure the energy deposits. It is composed of alternating brass absorbing layers and plastic scintillator tiles. HO is fully made out of plastic scintillators, as the preceding solenoid coil acts as an absorber. Both HB and HO cover the pseudorapidity range $|\eta| < 1.3$. In addition, the hadron calorimeter endcaps (HE), designed with the same materials as HB, cover the pseudorapidity range of $1.3 < |\eta| < 3$. HB, HE and HO are segmented in towers of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ $(0.09 - 0.35 \times 0.17)$ for $|\eta| < 1.7$ ($|\eta| \geq 1.7$).

The HCAL detector coverage is further extended up to $|\eta| < 5$ by incorporating a forward Cherenkov-based hadronic calorimeter system (HF). This detector is exposed to higher radiation levels due to its positioning. Therefore, the HF must employ more radiation-hard materials, namely a steel absorber with embedded quartz fibres. HF is segmented in $\Delta\eta \times \Delta\phi = 0.175 \times 0.175$ (0.175×0.35) for $|\eta| < 1.7$ ($|\eta| \geq 1.7$).

Fig. 3.8 shows a schematic layout of one quadrant of the CMS detector where HCAL subsystems are indicated.

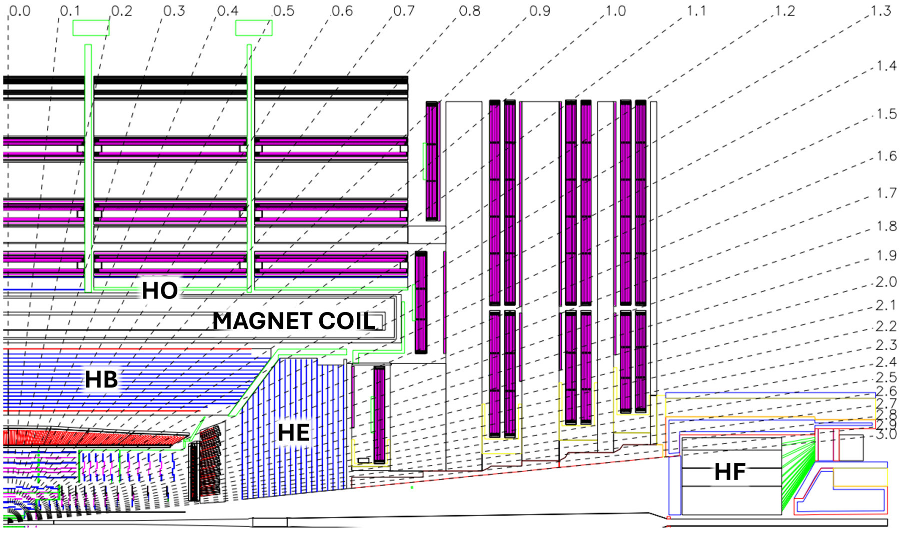


Figure 3.8: Schematic view of one-quarter of the CMS detector, where the components of the HCAL system, HO, HB, HE, and HF are indicated. Adapted from [34].

The combined ECAL and HCAL energy resolution σ is parametrized with the relation [150]:

$$\frac{\sigma}{E} = \frac{a}{\sqrt{E}} \oplus b, \tag{3.5}$$

where, in analogy to eq. 3.4, a accounts for the stochastic term, and b for effects that are non-linear with the energy. Both parameters are obtained in beam-test measurements, yielding $a = 85\%$ ($a = 198\%$) and $b = 7\%$ ($b = 9\%$) in the barrel (endcap) region. Despite the larger stochastic term in the forward region, the energy resolution remains similar, as jets tend to be much more energetic in this region [150].

3.2.3 Muon detector system

One of CMS's key design aims was a powerful muon detection system that provides robust muon identification, momentum resolution, and triggering capabilities. The CMS muon system [125, 152] comprises three gas-ionization detectors: Drift Tubes (DTs), Cathode Strip Chambers (CSCs), and Resistive Plate Chambers (RPCs). Together, they provide a muon reconstruction, identification and isolation efficiency greater than 95% [153]. The layout of these detectors in one quadrant of the CMS detector is shown in Fig. 3.9.

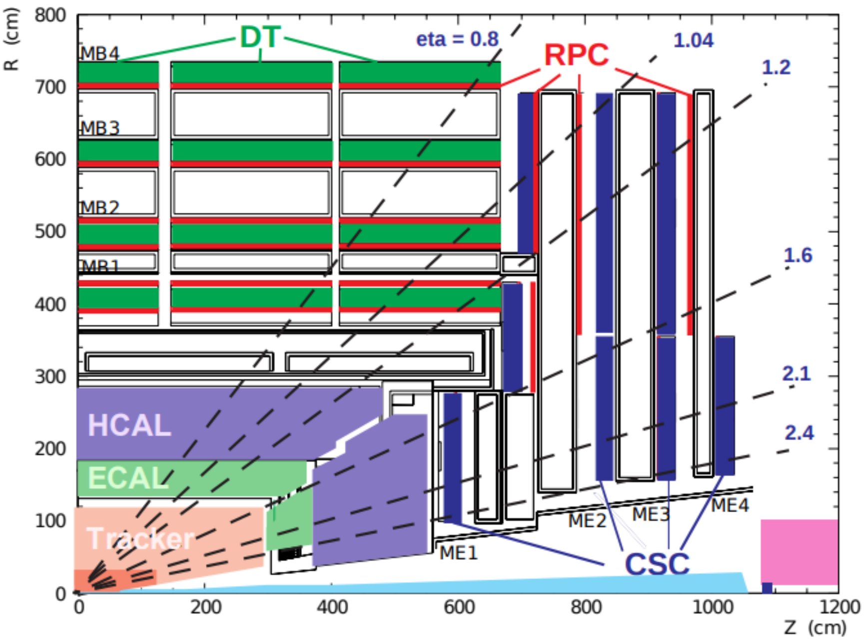


Figure 3.9: Schematic view of one-quarter of the CMS detector. The DT, CSC and RPC stations are shown in green, blue and red, respectively. Taken from [152].

The DTs are arranged in four concentric stations within the barrel region, covering the pseudorapidity range of $|\eta| < 1.2$. In this region, the signal and background rates are relatively low, so it is possible to use drift tubes to track the muons' path. Each DT station contains aluminium tubes filled with a mixture of Ar and CO_2, centred by a gold-plated stainless steel wire. The gas mixture becomes ionized by the passage of muons, and resulting electrons drift towards the wire. The drift time of these electrons is measured to determine the muon's position, enabling the calculation of its trajectory and, from the bending caused by the residual magnetic field, its momentum.

The CSCs are located in the barrel-endcap overlap region and the forward region, covering $0.9 < |\eta| < 2.4$. As this region features a larger signal and background rate, and a stronger and less uniform magnetic field, a more granular detector that provides a faster response time is employed. There are 468 CSC chambers in total. Each of them is made up of layers of positively charged anode wires that are crossed by negatively charged cathode strips within a gas mixture of Ar, CO_2 and CF_4. This configuration provides two-dimensional information of the muon's position in the r-ϕ coordinates. The fast response of the CSCs is critical for the CMS Level-1 trigger system.

The RPC detectors are distributed in six and three layers throughout the barrel and endcap regions, respectively. They offer redundancy and improved muon trigger capabilities due to a fast response time, which is in the order of 1 ns. They comprise two parallel plates with a thin gas layer in between that contains mainly $C_2H_2F_4$. This signal can be read extremely fast, in nearly 1 ns.

3.2.4 Trigger and data acquisition system

At the CMS interaction point, proton bunches cross every 25 ns, resulting in a bunch-crossing frequency of up to 40 MHz. Directly processing and storing all the data produced is not feasible due to bandwidth limitations [127]. Moreover, only a small fraction of these events are useful for physics analysis or calibration. To efficiently select events of interest, the CMS Collaboration utilises a two-level triggering system [125, 154]. This system consists of a hardware-based Level-1 Trigger (L1) and a software-based High-Level Trigger (HLT) that runs on a computing farm. Together, they reduce the initial data influx to about 1 kHz. The components of the CMS trigger system are detailed below.

The Level 1 Trigger

The firmware-based L1 trigger decides whether to keep an event based on coarsely segmented data from the calorimeters and the muon systems within roughly 4μs [125, 155]. A simplified reconstruction of physics objects, referred to as "trigger primitives", is employed for the decision-making.

The L1 system operates in two different phases. Initially, data from the muon detectors and the calorimeters are processed in iterative steps, as explained next, before being integrated and forwarded to a Global Trigger (GT).

The L1 calorimeter trigger is structured in two layers [156]. The first layer, or regional calorimeter trigger, receives inputs from ECAL and HCAL and produces an output of electron and photon candidates. The second layer, or global calorimeter trigger, utilizes this information and further sorts the input candidates, finds jets and hadronically decaying tau leptons, and calculates missing transverse energy. The outputs are the aforementioned physics object candidates.

On the other hand, the L1 Muon Trigger comprises a first step with three muon track finders that operate across three regions of the CMS detector: the barrel, the overlap, and the endcap. These track finders reconstruct muon paths which are then evaluated by a global muon trigger. The latter combines the candidates found by more than one system to eliminate duplicated objects and performs a basic quality selection before transmitting the information to the GT.

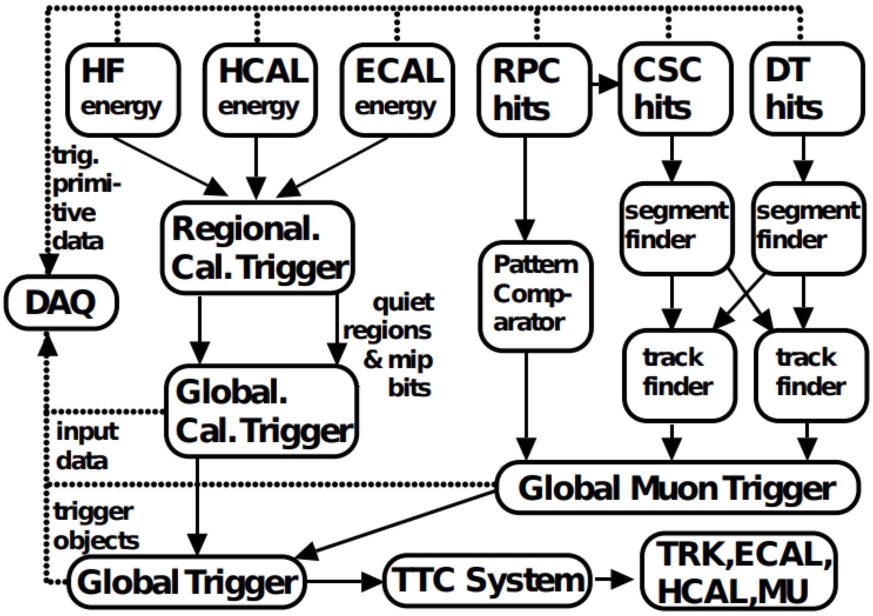


Figure 3.10: Diagram illustrating the CMS L1 trigger system. Taken from [156].

Finally, the GT combines the inputs from the calorimeter and muon triggers and runs a set of nearly 300 algorithms (or “L1 seeds”), each containing a set of selection criteria to be applied to the final set of physics objects. Figure 3.10 shows a diagram of the L1 trigger system. The output rate of the L1 trigger was limited to 100 kHz during Run 2. This value was exceeded during Run 3, reaching a maximum of 110 kHz in 2023 at the moment of writing [157].

The High Level Trigger

The HLT further processes the data that passes the L1 triggers. At this stage, the information from all CMS subsystems can be employed, and a streamlined version of the full offline event reconstruction is performed [127, 156]. The HLT algorithms run in a processor farm, executing a filtering process structured by a menu of "HLT paths". Each path is a sequence of selection steps on the reconstructed objects, which are increasingly more complex to skim the largest amount of events in the less computationally expensive steps. For instance, jets may be filtered by momentum before executing flavour tagging algorithms. The major HLT constraints are the output rate, limited the computing resources needed to fully reconstruct the offline data, and the processing time per event, also known as *timing*. The HLT output undergoes a full offline event reconstruction to be used for physics analysis.

During Run 1 and Run 2, the CMS HLT computing farms were equipped with CPU cores. A notable upgrade in preparation for Run 3 was deploying a heterogeneous scheme that incorporates GPU cores to the HLT farm. As a result, the processing time per event was reduced by 42%, from 690 ms to 397 ms at PU 55 [158].

The HLT rate was limited to 1.5 kHz during Run 2 and up to 5 kHz during Run 3. Each HLT path is assigned a certain rate bandwidth that can be achieved either by applying thresholds or pre-scaling the trigger. A pre-scaled trigger records one event after being fired a user-defined number of times. Both trigger thresholds and pre-scales must be optimized to record the most events of interest while complying with the HLT capacity.

The CMS collaboration has deployed two alternative strategies to circumvent the rate limitations. One is called "data parking", which relies on delaying the offline reconstruction until computing resources become available. The second, called "data scouting", is based on the production of a data stream for analysis in which events can be recorded at a higher rate but undergo HLT-level reconstruction only [157, 159]. Both strategies have been employed for various purposes, including Beauty physics and searches for exotic resonances in di-jet final states with loose thresholds.

The analysis triggers employed to collect the data analyzed in this thesis used the standard data collection and offline reconstruction schemes, and the rates are controlled with the online thresholds in the case of the main analysis triggers, while the control triggers used in complementary studies are also generally prescaled. These triggers are listed in section 5.5.

As part of this thesis project, I worked on the deployment and monitoring of the triggers dedicated to recording the data for the search of heavy Higgs bosons in final states with b-quarks in preparation for and during the Run 3 LHC data-taking period. This work is summarized in Appendix C.

Events that pass the trigger filters are then stored in a data storage infrastructure known as the Worldwide LHC Computing Grid (WLCG) [160], designed to handle the massive data volumes generated by LHC experiments. The “raw” data from the detectors is further processed to reconstruct physics objects that are used in physics analyses. In several instances of the data processing workflow, the data is certified as part of the CMS Data Quality Monitoring (DQM) system [161]. The reconstruction of physics objects will be detailed in the following chapter.

chapter 4

Object reconstruction and detector simulation

The passage of a detectable particle through the CMS detector results in energy deposits in its subdetectors. As shown in Fig. 4.1, each particle yields a characteristic signature. This information is employed to reconstruct the objects used in physics analysis.

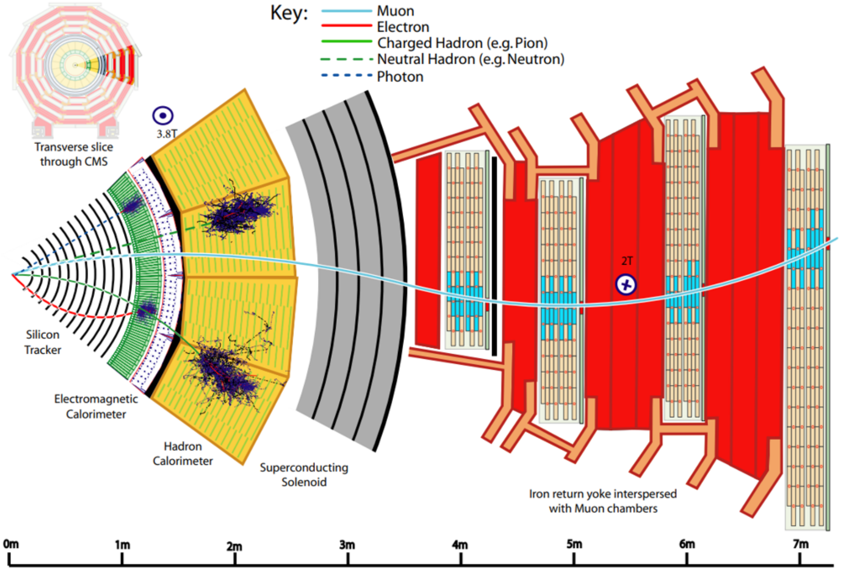


Figure 4.1: Illustration of the signature left by different particles in the CMS detector. Taken from [162].

Reconstructed vertices, jets, b-tagged jets, and muons are employed in the work presented in this thesis. This chapter summarises the detector simulation, followed by an overview of the reconstruction of the aforementioned physics objects.

4.1 Detector Simulation

Physics analyses rely on simulation-based techniques in key analysis steps, such as estimating backgrounds, trigger design, and detector calibration. Thus, an accurate simulation of events is crucial.

In the CMS experiment, detector simulations are conducted using the GEometry ANd Tracking (GEANT4) toolkit [163, 164]. This well-established tool allows a reliable model of the interactions between particles and the detector materials. This simulation creates a complete response of the detector components, mimicking the conditions of real data.

Physics processes are simulated separately, generating events based on theoretical models of particle interactions, production and decay of particles. These generated events are interfaced with the aforementioned simulated detector geometry, thus producing a response in the detector components that can be reconstructed as the real data.

In the analysis presented in this thesis, signal events are simulated with several hypothetical Higgs boson masses, and simulated QCD processes are employed to complement the data-driven background estimation, as will be described in the next chapter, in section 5.4.

4.2 Object Reconstruction

4.2.1 Tracks and primary vertex reconstruction

Tracks of charged particles and vertices are reconstructed from the hits in the pixel and silicon strip detector in an elaborate process due to the high luminosity and large PU data-taking conditions at the LHC.

First, locally reconstructed hits are translated into global coordinates, taking into account the tracker alignment corrections mentioned in Sec. 3.2.1 and detailed in Appendix A. Next, CMS employs a method called *Combinatorial Track Finder* (CTF) [165], adapted from the so-called combinatorial Kalman filter [166–169]. The track reconstruction takes place in a succession of up to six iterations of the CTF sequence, called *iterative tracking*. In the first iterations, prompt tracks are reconstructed, i.e. those produced closest to the region where the p-p collisions occur (beam spot), as well as high p_{T} tracks. The following iterations are intended to reconstruct displaced tracks and to recover those that were not reconstructed in the previous steps.

Each iteration is based on identifying hits (called "seeds") that can generate a possible track. The Kalman filter extrapolates the seed trajectory to an expected path, and additional hits corresponding to the track candidate are searched. Tracks that do not pass predefined quality criteria are discarded, and the hits of the final track candidates are removed to reduce the complexity of the following iterations. As iterations increase, the seed generation and final track selection are adjusted, and quality criteria are tightened to reduce the rate of fake or noise-related tracks and compensate for a loosening of the seed generation requirements.

During Run 2, the efficiency of reconstructing tracks ranged from 99% in the barrel region to about 95% towards higher pseudorapidity [170].

To reconstruct primary vertices, tracks are clustered with a deterministic annealing (DA) method [171], according to their point of closest approach to the beam spot. Vertex candidates with at least two associated tracks are fitted with an adaptative vertex finder [172] to determine its coordinates and covariance matrix. During this process, tracks are assigned a weight that ranges from 0 to 1, according to the probability that it is associated with the fitted vertex. The resolution of the reconstructed vertices depends on its track multiplicity and the kinematics of the associated tracks. The resolution of the measured impact parameter in the transverse plane as a function of the track transverse momentum and pseudorapidity are shown in Fig. 4.2.

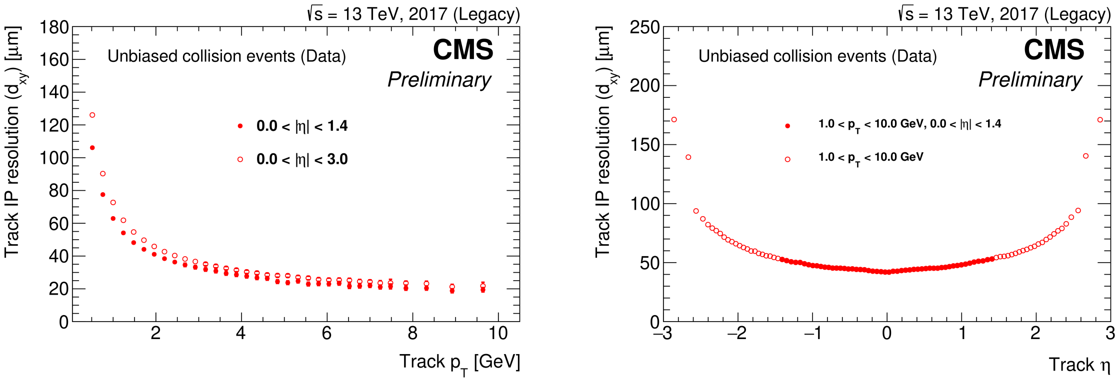


Figure 4.2: Transverse impact parameter resolution as a function of the track transverse momentum (left) in the barrel and endcap regions and as a function of the track pseudorapidity (right), obtained from collision events in 2017 data. Taken from [173].

4.2.2 The Particle Flow Algorithm

Physics objects are reconstructed at CMS with the Particle Flow (PF) algorithm [162], which aims to identify all particles utilizing the information of various subdetectors. Furthermore, it provides an efficient identification of pileup products, enabling its mitigation.

To guarantee its performance, a highly granular tracker and ECAL, as well as a hermetic HCAL, a magnetic field with strong bending power and an exceptional muon detection system, are needed. The CMS detector fulfils these conditions, and the PF algorithm has been employed for offline reconstruction since 2010, after being commissioned with data, and implemented at the HLT in 2011 [174].

Calorimeter clusters that are linked to a track indicate charged hadrons. On the other hand, neutral hadrons and photons are identified by calorimeter clusters without an associated track. Electrons leave a signature both in the tracker and ECAL, and muons are identified with a track in the tracker linked to a track in the muon system. A detailed explanation of the reconstruction of the main physics objects used in this thesis is given in the following.

4.2.3 Muons

Muons at CMS provide a clear signature for many physics analyses. Their reconstruction relies on the information from both the CMS tracker and muon systems in two main strategies: the reconstruction of "tracker muons" and "global muons" [153, 175, 176]. The first is based on reconstructing muons in an *inside-out* approach, combining the information from high p_{T} tracks from the tracker that can be extrapolated into hits in the muon chamber. The global muon reconstruction works the other way around, reconstructing muons *outside-in*, starting from energy deposits in the muon chambers that form "standalone muons", which are matched to hits in the tracker. The combined signals in the muon and tracker systems are then fitted with a Kalman Filter to form a track. Both approaches can reconstruct muons from p-p collisions with about 99% efficiency [175].

The PF flow algorithm identifies reconstructed muons with certain identification (ID) criteria. In the study presented in this thesis, the *tight muon selection* is employed. This criterion selects muons reconstructed both as tracker and global muons, with hits in at least six tracker layers, which includes at least one pixel hit. The reconstructed muon must also satisfy the transverse (longitudinal) impact parameter selection $d_{xy} < 0.2$ cm ($d_z < 0.5$ cm). A correction to address differences in the efficiency of the tight identification of muons in data and simulations is applied to simulated samples. This correction is centrally provided by the CMS Collaboration and obtained as a function of the muon's p_{T} and η. A tag-and-probe method is employed in their determination, utilizing the $J/\psi \to \mu^+\mu^-$ and the $Z \to \mu^+\mu^-$ decay channel, corresponding to the muon's p_{T} intervals $0 < p_{\mathrm{T}} < 30$ GeV and $15 < p_{\mathrm{T}} < 120$ GeV, respectively. Fig. 4.3 shows the muon identification scale factor derived in the latter topology.

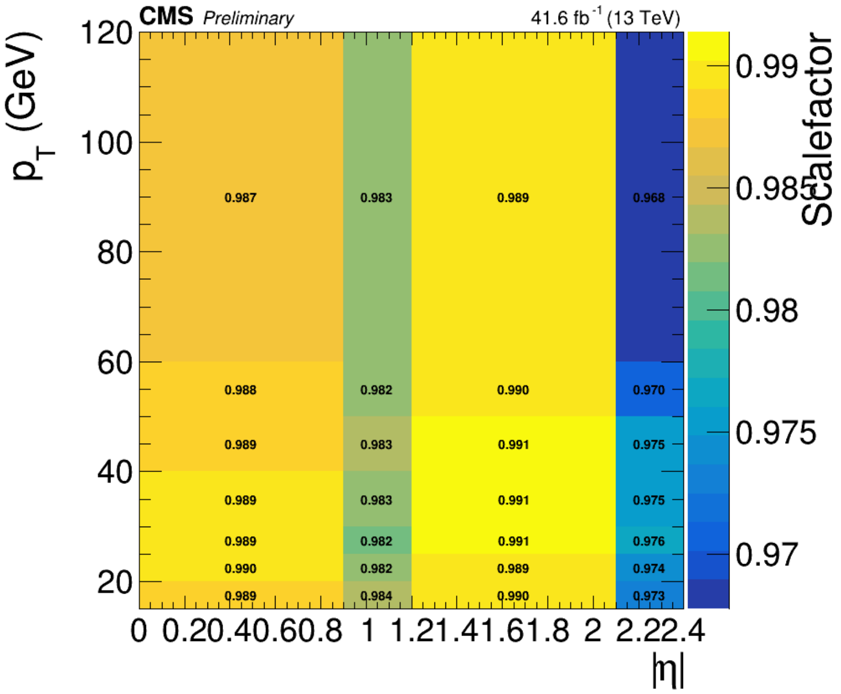


Figure 4.3: Scale factors to correct the simulations according to the muon ID efficiency, derived from the $Z \to \mu^+\mu^-$ decay channel. Taken from [177].

4.2.4 Jets

Being the result of the hadronization of quarks and gluons, jets are complex objects that may include several final state objects. The efficient reconstruction of jets is essential for many analyses, particularly in a hadron collider, as it provides information about the parton that initiated it. A well-established method, and the standard technique used at CMS to reconstruct jet candidates, is the anti-kt clustering algorithm [178]. It considers the kinematic properties of the particles that can be combined into a jet cone, as well as their angular separation. In a nutshell, this clustering algorithm computes a separation variable d_{ij} from a parton candidate or object i, which can be a calorimeter cluster, to all other objects j in an event as follows:

$$d_{ij} = \min\left(k_{ti}^{-2}, k_{tj}^{-2}\right) \cdot \frac{\Delta_{ij}^2}{R^2}, \tag{4.1}$$

where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, k_{ti} is the transverse momentum of the particle i, while y_i and ϕ_i are its rapidity and azimuthal angle, respectively. The distance to the beam B, defined as $d_{iB} = k_{ti}^{-2}$, is also computed. If $d_{ij} < d_{iB}$, the four momenta of the i and j objects are clustered into a single jet candidate. The process continues iteratively for all the objects in an event. Two values of R are generally considered, depending on the analysis performed, namely, 0.4 and 0.8. The resulting jet candidate is referred to as “AK4” or “AK8”, respectively. AK4 jets are used in the analysis presented in this thesis.

This algorithm can be executed with calorimeter information only, resulting in “calorimeter jets”, which are utilized in the triggers employed in the present work. In contrast, the Particle Flow (PF) algorithm identifies individual particles that may compose the jet by combining information from the calorimeters, the tracker, and the muon detectors [162]. These particles are then combined in the jet clustering algorithm to form “PF jets”. This approach results in improved energy resolution, an energy response closer to the true jet energy, and enhanced pile-up (PU) mitigation, as charged PF candidates not originating from the PV are removed before clustering. The offline reconstructed jets used in the present work are PF jets.

As mentioned before, the main purpose of reconstructing a jet is to access accurate information about the original parton, for instance, its energy. Several experimental effects may yield an improper energy reconstruction. Furthermore, an accurate simulation of the jets’ energy distribution is instrumental. Therefore, corrections are applied to the jets’ energy in both data and simulation by correcting their four-momentum vector. These jet energy corrections (JEC) consist of two components, namely the Jet Energy Scale (JES) and the Jet Energy Resolution corrections (JER) [179].

The JES correction is a series of multiplicative factors to the jet’s four-momentum applied sequentially to account for several effects. They include PU mitigation corrections, and corrections to account for non-linear detector response in the case of data. Disagreements between data and simulation and potential differences in the reconstructed energy of light jets (originated from u, d, s quarks and gluons) and heavy flavour jets (originated from c and b quarks) are also corrected. Figure 4.4 illustrates the JES corrections flow.

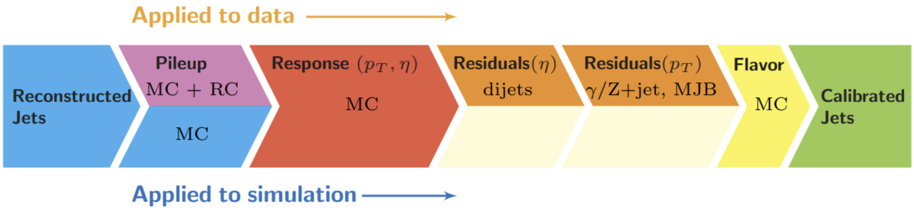


Figure 4.4: Illustration of the sequential JES corrections applied to data and simulation. Taken from [179].

The JER correction, on the other hand, aims to provide a more accurate description of the jet's p_{T} resolution in simulation by smearing it [179]. If the reconstructed jet is matched to a generator-level jet, a direct scaling to the p_{T} resolution of the jet is applied, according to the reconstructed and generated p_{T} of the jet. On the other hand, if the matching is not fulfilled, a stochastic smearing is applied according to a random distribution that follows a Gaussian pdf centred at zero and with a width derived from data.

The CMS Collaboration centrally provides both JES and JER corrections. The uncertainties related to them have a non-negligible impact on the analysis presented in this thesis, as will be discussed in section 5.7.2.

4.2.5 B-tagged jets

Jets originated from the hadronization of b-quarks, are an essential signature of many physics analyses, like the one discussed in this thesis. As a consequence of the relatively large b-hadrons lifetime ($\tau = 1.5$ ps [38]), they travel a sizable distance before decaying, producing a secondary vertex that can be reconstructed with information from the CMS tracker, as illustrated in Fig. 4.5. The secondary vertex (SV) information is critical for b-tagging techniques, therefore, they must be properly reconstructed.

At CMS, secondary vertices are reconstructed from tracks with large impact parameters with respect to a primary vertex [180] with the inclusive vertex finding (IVF) algorithm [38]. This technique also takes into account the angular distributions of the tracks, as shown in Fig. 4.6. Tracks that may introduce ambiguity between primary and secondary vertex identification are removed. The selected tracks are clustered to fit a secondary vertex.

The reconstructed secondary vertices, along with other discriminating variables like the track multiplicity, are used to identify jet flavours. Three identification algorithms have been employed in the analysis presented in this thesis, as more efficient tagging has

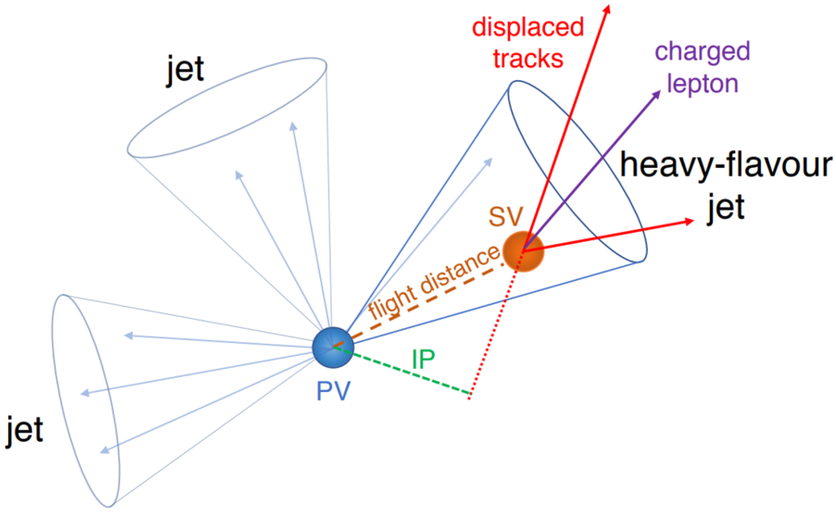


Figure 4.5: Schematic representation of a secondary vertex and heavy flavour jet initiated from the decay of a charm or beauty hadron. The secondary tracks are displaced with respect to the PV, resulting in a large impact parameter (IP). Taken from [38].

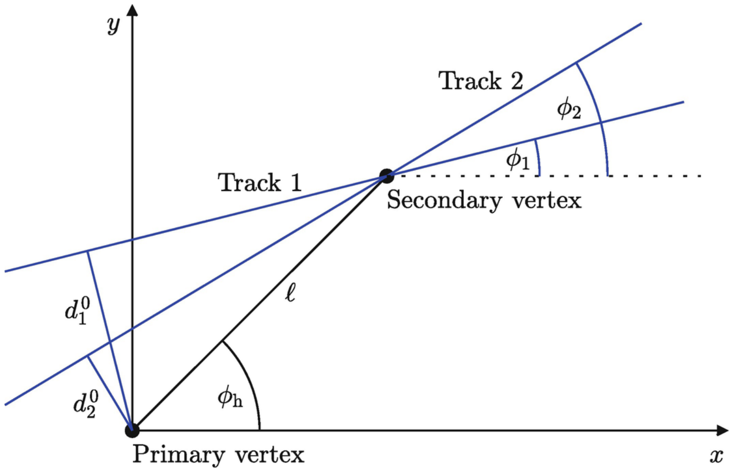


Figure 4.6: Illustration of the relation of tracks' impact parameter (d) and ϕ in the reconstruction of secondary vertices. Taken from [180].

been developed over time. In all cases, three working points are defined in CMS, according to the probability of tagging a non-b quark jet, referred to as misidentification rate. These are "tight","medium" and "loose", corresponding to 0.1%, 1%, and 10% misidentification rate, as determined in simulation.

In the present work, the Combined Secondary Vertex was employed in its second version (CSVv2) [181] at the trigger level during the 2017 data-taking. This algorithm uses a multivariate discriminator based on a neural network (NN) that employs the aforementioned information to compute binary flavour discrimination of b-jets against c-jets and light flavour. This information is combined to provide the jet-tagging information.

In the following data-taking year, the CSVv2 algorithm had been outperformed by the DeepCSV algorithm [181], and thus upgraded to the latter in the 2018 analysis triggers. The DeepCSV algorithm was based on its predecessor, although employing a deep neural network (DNN) with enhanced training techniques that allowed to obtain the probability of a jet being identified as a certain flavour.

This method was further improved by introducing the DeepJet algorithm [182], employed in this thesis at the offline b-tagging selection level in all analyses and categories. The DeepJet algorithm benefited from a much wider range of high and low and high-level input variables, nearly 650, such as the jet's and the track's kinematics, jet's track multiplicity, and primary vertex information. A comparison of the performance of the DeepJet and the DeepCSV b-tagging techniques is shown in Fig.4.7.

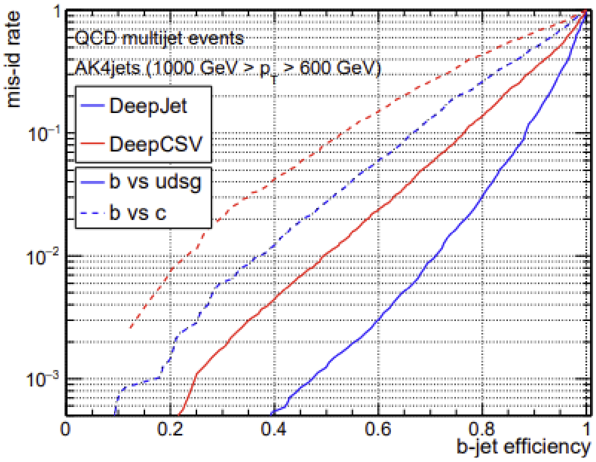


Figure 4.7: Performance of the DeepJet and DeepCSV b-tagging on high p_{T} jets from QCD multijet events. Taken from [182].

Any potential mismodelling of the b-tagging efficiency in simulated samples is corrected with scale factors. For the main analysis discussed in this thesis, the online b-tag scale factors are computed in a data-driven method (see section 5.5.2), while the offline b-tag scale factors are centrally provided by the CMS b-tag and vertex group, according to the working point employed, as a function of the jet flavour, p_{T} and η. The b-tagging scale factor is derived from combining five methods to calculate the scale factors [38] in QCD multijet events and top-quark pair events. Fig. 4.8 shows the b-tag scale factors corresponding to the medium working point of the DeepJet algorithm as a function of the jet's p_{T} in 2018.

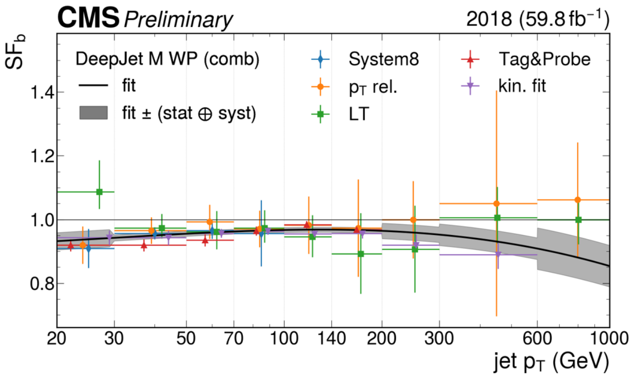


Figure 4.8: Scale factors to correct the simulations according to the b-tagging efficiency in data, corresponding to the medium working point of the DeepJet algorithm in 2018. The combination (comb) is derived from QCD and $t\bar{t}$-based scale factors. Taken from [183].

chapter 5

Search for neutral Higgs Bosons in final states with b-quarks

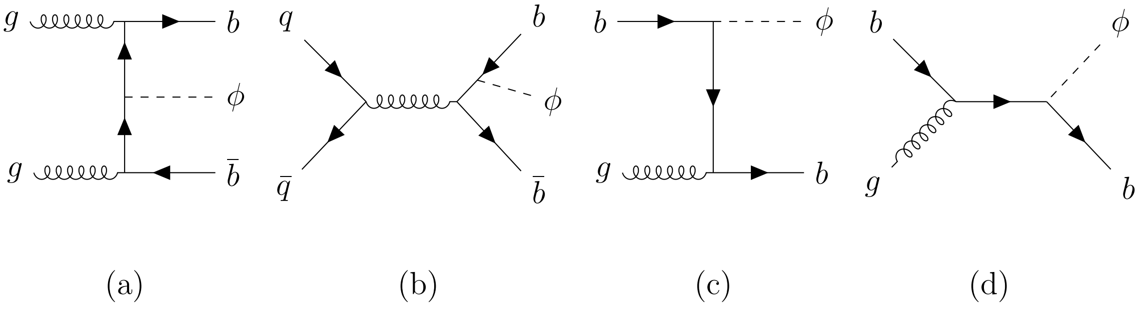
The discovery of the 125 GeV Higgs boson by the ATLAS and CMS Collaborations [12, 13] has set a new benchmark in the search for physics beyond the Standard Model. A possible interpretation is that this particle is the first observed scalar state of an extended Higgs sector. For instance, three neutral Higgs arise in 2HDM and MSSM models, conjointly denoted as $\phi = \{h, H, A\}$, as discussed in sections 2.2.2 and 2.2.1, where the 125 GeV resonance is commonly associated with h. When searching for these scalar bosons, the decay into a pair of b-quarks is rather attractive as it features a large branching ratio for a wide mass range [184]. On the other hand, the main production mechanisms of ϕ at the LHC are the b-associated production and the gluon gluon fusion. However, a search that targets either production mechanism is extremely challenging in a hadron collider, primarily due to the high multi-jet activity that complicates searches in hadronic final states. The b-associated production is less affected than its alternative. Furthermore, it is significantly enhanced across a broad region of the parameter space of 2HDM and MSSM scenarios, as stated in sections 2.2.2 and 2.2.1, resulting in an overall better sensitivity.

This chapter discusses methodologies employed in the search for an additional neutral scalar resonance in final states with at least three b-quarks, where two of them are anticipated to stem from the decay of ϕ and the third b-quark is strategically chosen to target the b-associated production. This analysis requires dedicated triggers that increase the search's significance and a thorough background study. This chapter discusses these topics in detail. Additionally, an overview of the statistical methods used for signal extraction is provided.

5.1 Signal and background processes

To understand the possible contributions to the b-associated production of signal events of a heavy neutral Higgs one has to make assumptions on the mass of the b-quark, defining flavour schemes that affect the computation of the total cross-section of the process. These are the four-flavour scheme (4FS) and five-flavour scheme (5FS), discussed in section 2.2.1, each providing distinct potential signals, as shown in Fig. 5.1.

In the process this work is focused on, the Higgs boson subsequently decays into a bottom-quark pair. A typical experimental signature consists of two very energetic b-jets that stem from the Higgs decay and at least an additional b-jet related to the production process. Although the presence of a fourth b-jet in the final state is a possible experimental outcome, the fourth jet tends to lay beyond the detector's acceptance in most signal events [35].



(a) (b) (c) (d)

Figure 5.1: Feynman diagrams illustrating the production of heavy, neutral Higgs boson (ϕ) in association with b quarks at leading order in hadron colliders. The first two diagrams (a,b) are accessible in the four-flavour scheme, while the last two (c,d) are only possible in the five-flavour scheme, where the b-quarks are considered part of the proton structure.

The background sources originate from Standard Model processes that result in a final state with a minimum of three b-jet candidates. The predominant contribution comes from QCD multi-jet production, although $t\bar{t}$ + jets production may produce a much smaller but still sizable background [35,39,41]. Nonetheless, all contributions are treated simultaneously in a data-driven approach, as will be detailed in section 5.8. Feynman diagrams of the aforementioned processes are illustrated in Fig. 5.2 in an exemplary fashion.

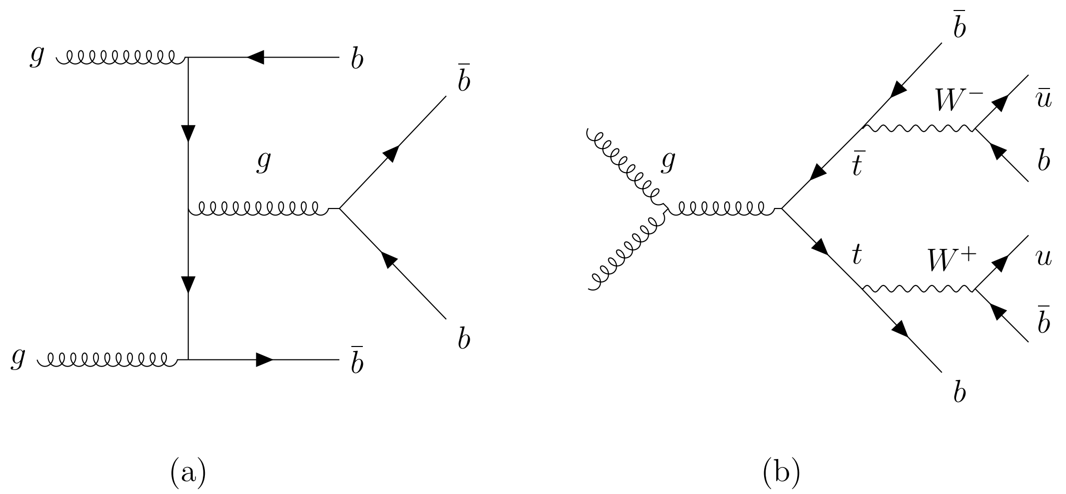


Figure 5.2: Feynman diagrams illustrating main background processes: (a) QCD multi-jet production in final states with b-quarks and (b) production of multiple b-quarks in $t\bar{t}$ events. All background sources are treated collectively in a data-driven approach.

5.2 Overview of results from previous searches

5.2.1 Searches for BSM Higgs bosons in b-quark final states at Tevatron

The search for heavy, neutral Higgs bosons in association with b quarks and decaying into a b-quark pair was initially investigated by the CDF and DØ experiments at the Tevatron collider [185]. This proton-antiproton collider was located at the Fermilab National Accelerator Laboratory and operated at a centre-of-mass energy of 1.96 TeV for nearly three decades, from 1983 to 2011. The diagrams that contribute to the b-associated production of a heavy Higgs boson in a hadron collider are shown in figure 5.1 for both 4FS and 5FS.

The combination of the searches performed by the CDF and DØ experiments with the Tevatron Run 2 data was summarized in [186], incorporating integrated luminosities of 2.6fb^{-1} and 5.2fb^{-1} from CDF and DØ, respectively. This represented approximately half of the total data collected in that period. Upper limits were set on the b-associated production cross-section multiplied by the branching ratio of the decay into a b-quark pair, varying between 44 pb and 0.7 pb in the mass range of 90 to 300 GeV. The data showed a good agreement with the background-only prediction, although a slight excess at the level of 2σ was observed at $m_\phi = 120$ GeV and 140 GeV. The results are shown in Fig. 5.3 along with their translation into exclusion limits in the tanβ and m_{A} parameter space for benchmark scenarios favouring the process studied. These model-dependent interpretations were depicted alongside those from its predecessor, the Large Electron-

Positron Collider (LEP) [87, 111], for comparison. Since the LEP was a lepton collider, the search for Higgs bosons mainly depended on their production through interactions with the Z boson, allowing to set complementary limits to those from Tevatron.

The full Run 2 CDF data was further analyzed in the search for BSM Higgs bosons and published in 2019 [187], corresponding to 5.4 fb^{-1} of integrated luminosity. This time, no evidence of a signal was found. The Tevatron ceased operations in 2011, and the search for scalar bosons continued at the LHC.

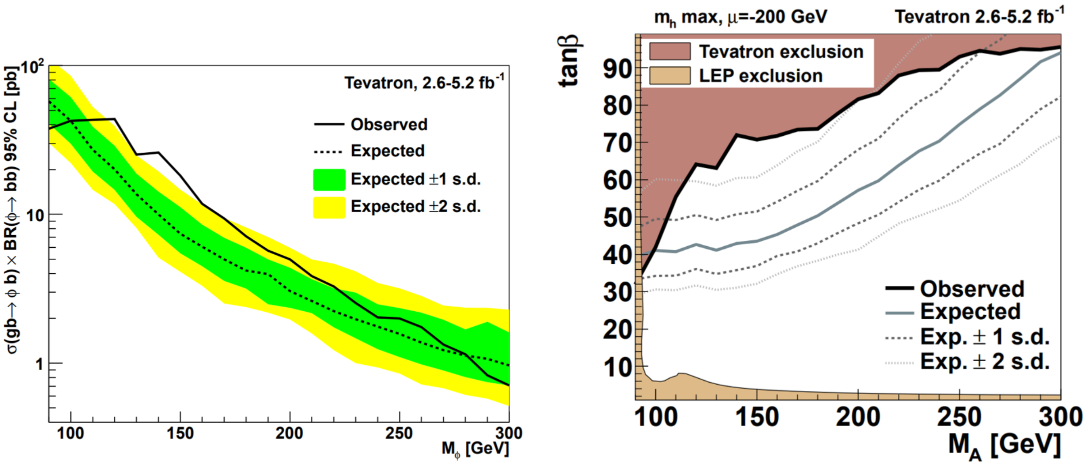


Figure 5.3: Upper limits set by TeVatron in the search for BSM Higgs bosons in the b-associated production cross-section times branching ratio of the decay into a b-quark pair (left) and in the context of the MSSM $m_{\mathrm{h}}^{\max}$ scenario, plotted alongside those from LEP (right). The two results complement each other in the explored parameter space. The observed mild excesses at 120 and 140 GeV, corresponding to deviations from the expectation of 2.5σ and 2.6σ [186], respectively, were not statistically significant and disappeared in a further study employing the full Run 2 CDF dataset [187].

5.2.2 Status of LHC searches

The Large Hadron Collider (LHC) [33] began its first operational phase, known as Run 1, in 2010, which lasted until 2013 [188], colliding protons at centre-of-mass energies of 7 and 8 TeV. A pivotal moment during this period was the discovery of a Higgs boson with a mass of about 125 GeV [12, 13]. This discovery significantly influenced the course of the extended Higgs sector predictions in MSSM and 2HDM benchmark scenarios. Although BSM theories had to account for the measured properties of the newly observed particle, many of these models were not outright discarded. Particularly, the search for supersymmetry [24] gained special interest, as it offers a theoretical mechanism to resolve the hierarchy problem that arises from the relatively low mass of the Higgs boson [24].

The CMS experiment was the first to search for heavy Higgs bosons in b-quark final states, with a dataset corresponding to an integrated luminosity of 2.7 - 4.8 fb^{-1} collected at 7 TeV centre-of-mass energy in 2011 [189]. The collected luminosity at a larger centre-of-mass energy made it possible to extend the mass range probed towards higher values. In addition to the fully-hadronic (FH), the search was conducted in a semi-leptonic (SL) final state, characterized by the presence of a non-isolated muon within the hadronization cone of one of the b-jet candidates. This approach facilitated the examination of lower mass regions by reducing the jet transverse momentum thresholds while managing trigger rates. Subsequent analyses extended to the 8 TeV dataset, corresponding to 19.7 fb^{-1} collected during 2012. This search focused on the fully-hadronic final state and the model interpretations in the context of MSSM scenarios were provided for the combination of the 7 TeV and 8 TeV results, shown in Fig. 5.4. These interpretations include a revised $m_{\mathrm{h}}^{\mathrm{max}}$ scenario and a newly formulated $m_{\mathrm{h}}^{\mathrm{mod+}}$, developed after the 125 GeV Higgs boson discovery, as discussed in section 2.2.1. Overall, the data showed a good agreement with the background-only hypothesis.

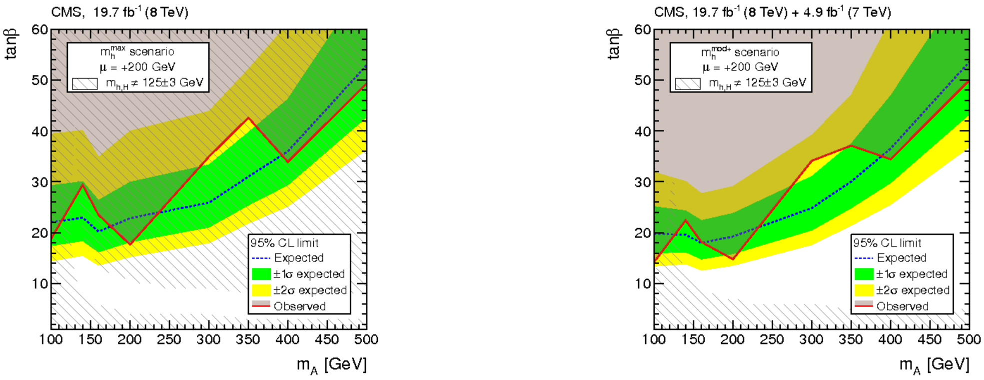


Figure 5.4: Upper limits set in the search for BSM Higgs boson in the $b\bar{b}$ decay channel in the context of the $m_{\mathrm{h}}^{\mathrm{max}}$ (left) and $m_{\mathrm{h}}^{\mathrm{mod+}}$ (right) MSSM scenarios with the combined 7 and 8 TeV CMS datasets [189, 190]. The hatched areas denote the regions where the masses of neither of the CP-even MSSM Higgs bosons, h or H, align with the 125 GeV Higgs boson discovery, allowing for a deviation of up to 3 GeV. Taken from [190].

In the following LHC data-taking period, Run 2, protons collided at centre-of-mass energy of 13 TeV from the end of 2015 until 2018. The LHC instantaneous luminosity also increased to about $2 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, exceeding about twice its design value. This allowed yet again to probe an even higher mass regime. However, the trigger rates, particularly for hadronic processes, became prohibitive, forcing the jet transverse momentum thresholds to be raised and consequently reducing the sensitivity for low-mass Higgs bosons.

Both CMS [35, 41] and ATLAS [191] experiments analyzed the 2016 dataset in an inclusive search, i.e. where no distinction was made between the fully-hadronic and semi-leptonic categories. They explored the mass range of 300-1300 GeV and 450-1400 GeV with 36.02 fb^{-1} and 27.8 fb^{-1} of data collected during 2016 by CMS and ATLAS, respectively. In addition, the 2017 dataset was studied by the CMS experiment in the semi-leptonic category [39], re-introduced to target a mass range from 125 GeV to 700 GeV, and the fully-hadronic category [40], which probed masses from 300 GeV to 1600 GeV. The CMS triggers dedicated to this search collected 36.5 fb^{-1} of data in the SL category and 31.2 fb^{-1} in the FH category. These two analyses were the results of two PhD. thesis projects and did not result in a publication.

In all Run 2 studies, the hypothetical Higgs signals were extracted from a smoothly falling background in the signal region (SR) by a data-driven approach that employs an orthogonal CR. In the case of CMS 2016 FH and 2017 SL analyses, possible functional forms of the data in SR are obtained from the CR. Moreover, in 2016 ATLAS and 2017 CMS FH analyses, a different approach is followed, relying on a simultaneous fit of SR and CR. Two components were used to describe the background in the SR: one is given by the shape of the CR, and the other is a transfer factor (TF), whose functional form is inspired by the ratio of SR and CR in simulated multijet events. In the four analyses, the upper limits on the production cross-section times branching ratio are extracted and interpreted in benchmark scenarios within MSSM, including the $m_{\mathrm{h}}^{\mathrm{mod}+}$, hMSSM and m_{h}^{125} scenario, the latter only in the 2017 FH analysis, and within 2HDM type 2 and flipped models. Overall, no significant deviation with respect to the background-only expectation was found in any of the investigations. Some of the results from the 2016 CMS and ATLAS searches are summarized in Fig. 5.5.

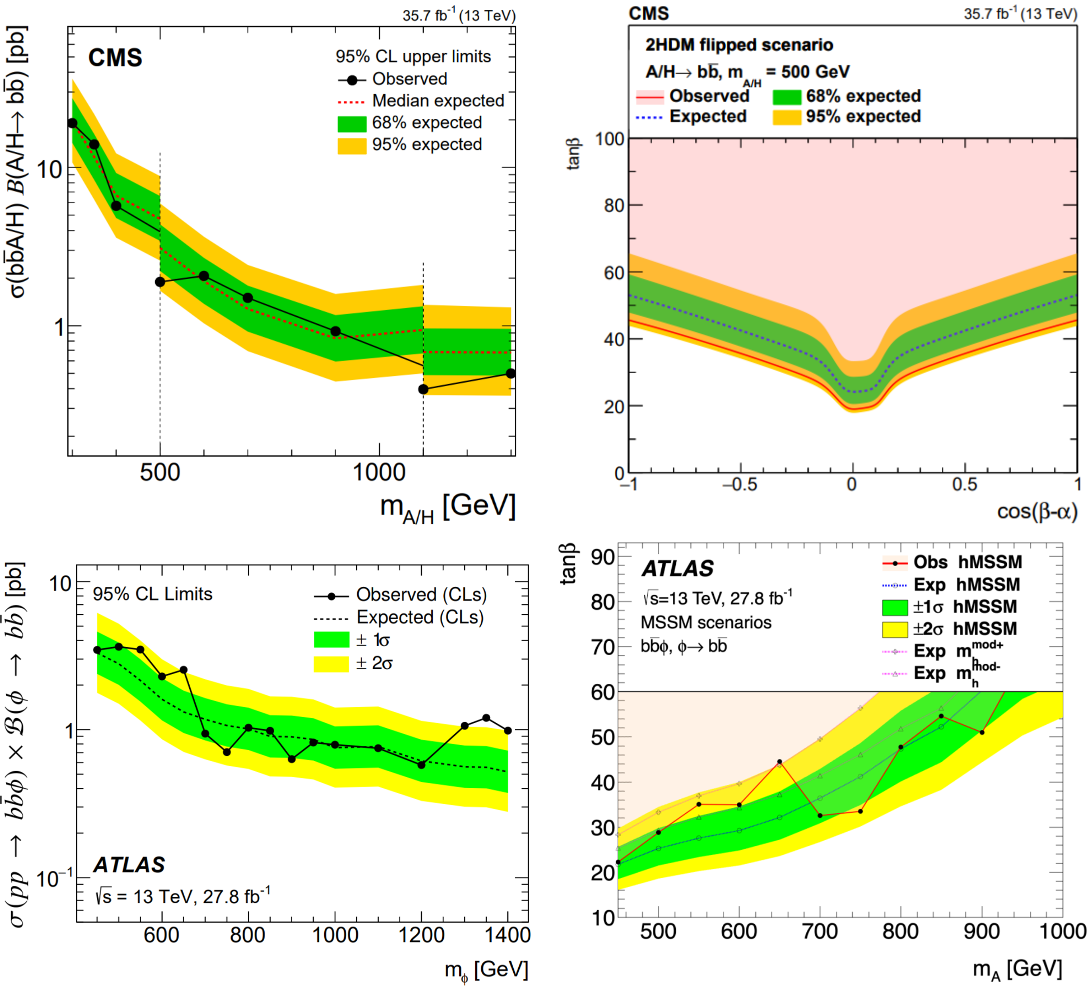


Figure 5.5: Limits on the b-associated production cross-section times branching fraction (left) set in the analysis of 2016 data by the CMS [35] (top) and ATLAS [191] experiments (bottom). Interpretations in the context of 2HDM flipped scenario in the $\cos(\beta - \alpha)$ and $\tan\beta$ plane are shown for the CMS analysis (top right) and in the context of the hMSSM benchmark scenario in the $\tan\beta$ and m_{A} parameter space for ATLAS analysis (bottom right).

5.3 Analysis strategy

As previously mentioned, the data-taking conditions at the LHC during Run 2 forced the jets p_{T} thresholds to be raised at the trigger level to cope with the high multi-jet activity. The SL category was introduced to recover the sensitivity towards lower masses, for which a trigger requesting muon within a jet was deployed specifically to target SL b-hadron decays to a muon, alongside the conventional jet triggers used in the FH category. The muon requirement reduces the trigger rates, allowing to lower the jets' p_{T} thresholds.

In the present work, the 2017 dataset is analyzed in the FH and SL categories, corresponding to 36.7 fb^{-1} and 31.2 fb^{-1} of data, respectively. Additionally, 54.5 fb^{-1} of 2018 data were analyzed for the first time, in an inclusive search due to the validity of the SL trigger this year. These analyses benefit from key enhancements with respect to their predecessors [35, 39, 40] to target substantial enhancement in the overall significance, improved systematic uncertainties evaluation, and other improvements that will be detailed in section 5.6. As a result, an unprecedented mass range from 125 GeV to 1800 GeV was explored with the SL and FH categories combined. The results obtained from these analyses were combined with those derived from the 2016-only CMS study, accounting for up to $\sim$127 fb^{-1} of data collected by the analysis triggers during the full LHC Run 2 data-taking period.

This analysis focuses on searching for an excess of events in the invariant mass distribution of the two leading b-tagged jets in p_{T}, "M_{12}", which are expected to originate from the Higgs decay. Signal events should feature at least an additional b-tagged jet related to the b-associated production process. The signal region (SR) is defined when the three leading jets in an event accomplish a triple b-tag selection criterion (*bbb*). The signal model devised in this analysis is based on simulated events in a wide range of hypothetical Higgs boson masses as will be discussed in section 5.7. The shapes of the M_{12} distributions for background and signal models are parametrized using functions chosen such that they provide a good description of the data. The arbitrariness of the choice possible leads to biases, which are addressed as systematic uncertainties.

A data-driven method comprising two components is employed to describe the complex multi-jet background shape in the SR. The first component is a signal-depleted Control Region (CR), also referred to as *bbnb*, which is defined by requesting that the third-leading jet does not fulfil the b-tag requirement. Both SR and CR feature a relatively complex shape whose parametrization covers several orders of magnitude. To assess the potential shape difference between the CR and the background in the SR, the second key component, the Transfer Factor (TF), is introduced. The transfer factor (TF) is defined as the ratio of SR and CR in simulated QCD multi-jet events. It features a relatively simple functional form that slowly varies with the dijet invariant mass, M_{12}. This method has been followed by the ATLAS and CMS collaborations [40, 191], and probed to reduce the systematic uncertainty related to the background modelling. It is relevant to note that by employing a fully data-driven background estimation, all background-contributing processes are collectively addressed.

Finally, a simultaneous signal-plus-background likelihood fit is performed in the M_{12} distributions of the SR and the CR to extract the signal production cross-section. When no signal is found, upper limits on the cross-section times branching ratio of the process are set.

A conventional blinding policy has been followed while the analysis was under development and optimization. Therefore, the invariant mass of the two leading jets is not explored until the strategies are fully deployed and validated. The validation of the signal extraction procedure is performed first with Monte Carlo (MC) datasets and later by means of a second signal-depleted region, called the Validation Region (VR), which is orthogonal to both SR and CR. This region is defined by requesting the b-tagging of the third leading jet to lie in between its definition in the SR and CR. It is thus referred to as semi-b-tagged ($bbsb$).

5.4 Data and simulated samples

The data analyzed in this work was collected by the CMS detector during 2017, over the run period, or eras, C to F, and during 2018, over eras A to D, at a centre-of-mass energy of 13 TeV. The corresponding datasets with their respective run ranges for each data collection period are listed in Table 5.1.

The Monte Carlo simulated signal samples of b-associated production of MSSM Higgs bosons and decay into a pair of b-quarks that are used in this analysis are listed in Table 5.2. In these samples, the hard-scattering events are generated at next-to-leading order with the `POWHEG 2.0` [192, 193] generator. `POWHEG 2.0` is interfaced with the `PYTHIA8` generator [194] to simulate hadronization, parton showering, and particle decay processes. The use of `POWHEG 2.0` with respect to `MADGRAPH`, which was employed in previous CMS analyses [39–41], is enforced to benefit from a lower fraction of negative weights. In comparison, studies using `MADGRAPH` to generate the $b\bar{b} \to A/H \to b\bar{b}$ signal mass points quoted a fraction of negative weight events of 36 to 40% [39, 40], while `POWHEG 2.0` yields approximately a fraction of 18%. The events with negative weight originate from interference terms in the next-to-leading order cross-section calculation and are discarded for the purpose of the analysis.

This analysis also necessitates the simulation of SM QCD multi-jet processes, which are used for the estimation of the transfer factor and for the computation of the trigger efficiency scale factors. They are produced in bins of the scalar sum of the jets transverse energies in an event, HT. Since inclusive QCD simulations possess insufficient statistics for the transfer factor and muon trigger efficiency studies, two additional groups of such simulations are employed. The first one consists of datasets enriched in b-quarks, based on NLO QCD calculations and produced with enhanced b-quark content at matrix element level or at parton-shower level. These samples are listed in Table 5.3 along with those

2017 Dataset	Run–range
Era C	299337–302029
Era D	302030–303434
Era E	303435–304826
Era F	304911–306462
Integrated luminosity (SL trigger)	36.3 fb^{-1}
Integrated luminosity (FH trigger)	31.1 fb^{-1}
2018 Dataset	Run–range
Era A	315252–316995
Era B	316998–319312
Era C	319313–320393
Era D	320394–325273
Integrated luminosity	54.5 fb^{-1}

Table 5.1: Data collected by the CMS detector in 2017 and 2018 that undergoes analysis in this study. The SL and FH triggers resulted in different total integrated luminosities during 2017 due to the occasional prescaling of the main FH L1 seed to 0 in a small number of lumi-sections.

from an inclusive QCD production. QCD samples enriched with muons were produced with PYTHIA8, and are also listed in Table 5.3.

Process	Mass point [GeV]	Events
$b\bar{b} \to A/H \to b\bar{b}$	125	1000000
	130	967000
	140	950000
	160	956000
	180	976000
	200	1000000
	250	2942000
	300	2964000
	350	2940000
	400	2892000
	450	2917000
	500	2896000
	600	3108000
	700	3099962
	800	3046973
	900	3094975
	1000	3107977
	1200	3009000
	1400	2958000
	1600	3087476
	1800	2982000

Table 5.2: Monte Carlo signal simulated samples for 2017 and 2018 configurations. The mass points from 125 to 250 GeV are exclusive of the SL analysis. Similarly, mass points from 800 GeV to 1800 GeV are only employed in the FH studies.

Sample	HT bin	Events
QCD b-Enriched	100 to 200	8894908
	200 to 300	9952918
	300 to 500	3547725
	500 to 700	2010208
	700 to 1000	891925
	1000 to 1500	267989
	1500 to 2000	61250
	2000 to Inf	302499
QCD b-GenFilter	100 to 200	9417440
	200 to 300	8611681
	300 to 500	5529691
	500 to 700	7621842
	700 to 1000	1816716
	1000 to 1500	1261997
	1500 to 2000	224332
	2000 to Inf	214088
QCD Inclusive	100 to 200	171876686
	200 to 300	58689666
	300 to 500	60316577
	500 to 700	56207744
	700 to 1000	47724800
	1000 to 1500	16595628
	1500 to 2000	11634434
	2000 to Inf	5941306
QCD μ-Enriched	50 to 80	20967952
	80 to 120	613257
	120 to 170	648944
	170 to 300	36027673
	300 to 470	494796
	470 to 600	517382
	600 to 800	17318812
	800 to 1000	17039755
	1000 to Inf	14642553

Table 5.3: List of Monte Carlo QCD multijet samples that were utilized in the present analysis.

5.5 Triggers

The intense multijet environment at the LHC makes it quite challenging to record interesting events for these analyses. Therefore, to optimally collect data for the BSM $b \to A/H/h \to b\bar{b}$ searches, in both FH and SL categories, dedicated triggers at the L1 and the HLT (see section 3.2.4) are required. They are designed to enhance the selection of events of interest and suppress background events while meeting the constraints in bandwidth and event processing time. In addition, several other general and dedicated control triggers were used to evaluate, using the data, the performance of the physics objects selected online by the analysis triggers. Eventually, any discrepancies observed in the simulation with respect to the data are corrected as discussed later in this section.

5.5.1 L1 and HLT triggers

The L1 triggering decision is based on information from the CMS calorimeter systems and muon chambers. Therefore, among the objects interesting to our analysis, only jets and muons can be employed to determine the selection of events at this stage.

For the FH category, events with at least two calorimeter jets with $p_{\mathrm{T}} \geq 100$ GeV ($p_{\mathrm{T}} \geq 112$ GeV) and $|\eta| \leq 2.3$ are selected at the L1 trigger in 2017 (2018). The reason for raising the p_{T} requirement in 2018 was due to the increase in the pileup and the resulting increase of the trigger rates. In addition, events are selected if at least one pair of jets fulfilling the requirements above has an angular separation of $\Delta\eta \leq 1.6$, which provides a strong rejection of QCD events [39, 41].

The L1 trigger requirement for the SL category comprises events with at least two calorimeter jets with $p_{\mathrm{T}} \geq 40$ GeV and $|\eta| \leq 2.3$, at least one muon with $p_{\mathrm{T}} \geq 12$ GeV, $|\eta| \leq 2.3$ and an angular separation to the jets' axis $\Delta R \leq 0.4$. Similar to the FH category, a rejection of QCD events is performed by requiring $\Delta\eta \leq 1.6$ for at least one pair of jets fulfilling the aforementioned conditions.

At the HLT, events are selected in sequence to optimize the processing time. In the FH category, a first selection of events requests at least two *anti-kt* calorimeter jets, each with a $p_{\mathrm{T}} \geq 100$ ($p_{\mathrm{T}} \geq 112$) GeV and $|\eta| \leq 2.3$ in 2017 (2018). Then, an online b-tag sequence tests at least two out of eight candidate jets, employing the CSV tagger in its second version [181] with a 0.33% misidentification rate in 2017, updated to DeepCSV with a 0.71% rate in 2018 [38]. Finally, at least two particle flow jets with $p_{\mathrm{T}} \geq 100$ ($p_{\mathrm{T}} \geq 116$) are selected in 2017 (2018). The FH HLT triggers were deployed to achieve an HLT rate of 10 Hz, and the online b-tag thresholds were adjusted accordingly. Following the CMS pixel detector upgrade in 2016/2017 [195], simulated samples were used for rate estimations in 2017. This resulted in tighter b-tag thresholds and lower rates during data collection. Conversely, in 2018, rate estimations used data from the previous year which allowed for more precise thresholds and relaxed b-tag requirements.

In the SL category, the structure of the 2017 HLT triggers mirrors that of the FH. A preselection of events with at least two $anti-kt$ calorimeter jets with $p_{\mathrm{T}} \geq 30$ GeV is first done. Additionally, a particle flow muon of $p_{\mathrm{T}} > 12$ GeV and $|\eta| < 2.3$ is required. As in its FH counterpart, the second version of the CSV tagger with a 0.33% misidentification is used next to select at least two b-tagged jets. Then, at least two particle flow jets are selected with $p_{\mathrm{T}} \geq 40$ GeV. This trigger was designed to meet an HLT rate budget of 15 Hz during data-taking. References to the full set of HLT triggers for these analyses appear in Tables 5.4, 5.5, and 5.6, complemented by L1 seed details in Table 5.7.

5.5.2 Trigger efficiencies

Understanding the efficiency of the triggers used for any physics analysis is extremely important. Trigger efficiency, together with detector acceptance, reconstruction, and overall selection efficiencies, are part of the total efficiency of recording a physics process. Therefore, trigger efficiency directly enters cross-section measurements and must be reliably measured.

In general, trigger efficiencies are affected by the detector resolution and discrepancies in the object reconstruction steps. For instance, jets are selected at L1 with only calorimeter information, while at the HLT they are reconstructed using information from all subsystems through the particle flow algorithm. Furthermore, the HLT algorithms use different inputs compared to offline reconstruction, which results in slight variations in reconstructed objects [196].

Simulating these effects is rather complex, and differences with respect to the actual performance during data-taking can occur. Therefore, simulations must be calibrated against real data. This involves computing scale factors by comparing efficiencies in data and simulations.

Different methods can be used to estimate trigger efficiencies in a data-driven way [197–199]. One example is the *orthogonal dataset* method, where a data sample triggered by one trigger is used to measure the efficiency of a different, independent trigger. For instance, a muon trigger can be used to measure a jet trigger's efficiency. Another approach is the *Boot strapping* method, where a less restrictive trigger, usually prescaled, is used to calculate the efficiency of a trigger with tighter thresholds and generally more complex.

Another well-established technique to estimate trigger efficiencies is the *tag-and-probe method*. It relies on selecting a pure sample of two objects that are related to each other by some physics process, like those stemming from a particle's decay. One of these objects, the "tag", should be well identified with stringent selection criteria to guarantee a pure sample of candidates. The other object, the "probe", is left unbiased and used to estimate the efficiency as the ratio of probes that pass the criteria under investigation divided by the total of probes. In the present work, trigger efficiencies are determined with the *tag-and-probe* method. Their determination, followed by the computations of the scale factor to correct the simulations, is discussed next.

Trigger name	N_{jets}	N_{btag}	$p_{T,jet}$ [GeV]	$\lvert\eta_{jet}\rvert$	$\lvert\Delta\eta_{jet}\rvert$	$missID_{\mathrm{btag}}$	N_{μ}	$p_{T,\mu}$ [GeV]	$\lvert\eta_{\mu}\rvert$
Main trigger									
`HLT_Mu12_DoublePFJets40MaxDeta1p6_DoubleCaloBTagCSV_p33`	2	2	40	2.3	1.6	0.33	1	12	2.3
Control triggers									
`HLT_PFJet40`	1	0	40	5	–	–	0	–	–
`HLT_PFJet60`	1	0	60	5	–	–	0	–	–
`HLT_Mu8`	0	0	–	–	–	–	1	8	2.5
`HLT_DoublePFJets40_CaloBTagCSV_p33`	2	1	40	2.3	–	0.33	0	–	–
`HLT_DoublePFJets100_CaloBTagCSV_p33`	2	1	100	2.3	–	0.33	0	–	–
`HLT_DoublePFJets200_CaloBTagCSV_p33`	2	1	200	2.3	–	0.33	0	–	–
`HLT_DoublePFJets350_CaloBTagCSV_p33`	2	1	350	2.3	–	0.33	0	–	–
`HLT_Mu12_DoublePFJets40_CaloBTagCSV_p33`	2	1	40	2.3	–	0.33	1	–	–
`HLT_Mu12_DoublePFJets100_CaloBTagCSV_p33`	2	1	100	2.3	–	0.33	1	–	–
`HLT_Mu12_DoublePFJets200_CaloBTagCSV_p33`	2	1	200	2.3	–	0.33	1	–	–
`HLT_Mu12_DoublePFJets350_CaloBTagCSV_p33`	2	1	350	2.3	–	0.33	1	–	–

Table 5.4: Summary of the HLT paths used in 2017 SL analysis. The data used in the analysis was collected with the main trigger path, whereas the control triggers listed were used in the measurement of the kinematic trigger efficiencies.

Trigger name	N_{jets}	N_{btag}	$p_{T,jet}$ [GeV]	$\lvert\eta_{jet}\rvert$	$\lvert\Delta\eta_{jet}\rvert$ [%]	$missID_{\mathrm{btag}}$
Main trigger						
`HLT_DoublePFJets100MaxDeta1p6_` `DoubleCaloBTagCSV_p33`	2	2	100	2.3	1.6	0.33
Control triggers						
`HLT_PFJet60`	1	0	60	5	–	–
`HLT_PFJet80`	1	0	80	5	–	–
`HLT_DoublePFJets100_ CaloBTagDeepCSV_p33`	2	1	100	2.3	–	0.33
`HLT_DoublePFJets200_ CaloBTagDeepCSV_p33`	2	1	200	2.3	–	0.33
`HLT_DoublePFJets350_ CaloBTagDeepCSV_p33`	2	1	350	2.3	–	0.33

Table 5.5: Summary of the HLT paths used in 2017 FH analysis. The data used in the analysis was collected with the main trigger path, whereas the control triggers listed were used in the measurement of the kinematic trigger efficiencies.

Trigger name	N_{jets}	N_{btag}	$p_{T,jet}$ [GeV]	$\|\eta_{jet}\|$	$\|\Delta\eta_{jet}\|$ [%]	$missID_{\text{btag}}$
Main trigger						
`HLT_DoublePFJets116MaxDeta1p6_ DoubleCaloBTagDeepCSV_p71`	2	2	116	2.3	1.6	0.71
Control triggers						
`HLT_PFJet60`	1	0	60	5	–	–
`HLT_PFJet140`	1	0	140	5	–	–
`HLT_DoublePFJets100_ CaloBTagDeepCSV_p71`	2	1	100	2.3	–	0.71
`HLT_DoublePFJets200_ CaloBTagDeepCSV_p71`	2	1	200	2.3	–	0.71
`HLT_DoublePFJets350_ CaloBTagDeepCSV_p71`	2	1	350	2.3	–	0.71

Table 5.6: Summary of the HLT paths used in 2018 FH analysis. The data used in the analysis was collected with the main trigger path, whereas the control triggers listed were used in the measurement of the kinematic trigger efficiencies.

Analysis	L1 seed
2017 SL	`L1_Mu12er2p3_Jet40er2p3_dR_Max0p4_DoubleJet40er2p3_dEta_Max1p6`
2017 FH	`L1_DoubleJet100er2p3_dEta_Max1p6`
2018 FH	`L1_DoubleJet112er2p3_dEta_Max1p6`

Table 5.7: Summary of the L1 seeds corresponding to 2017 SL, FH, and 2018 FH main analysis triggers.

Jet kinematic trigger efficiencies

Ideally, triggering with thresholds on the kinematics of a physical object would result in a step function efficiency at the threshold with a 100% plateau. However, due to the aforementioned effects, the trigger bias produces a turn-on curve at the threshold that is prone to mismodelling in simulations. To correct for differences in the efficiency of the jet component of the main analysis trigger, scale factors based on the corresponding jet trigger efficiency ratio in data and simulation are obtained in each year and analysis category.

The jet kinematic trigger efficiencies are determined with a tag-and-probe method as a function of the jet's kinematic variables that are biased at trigger level, p_T and η, in a sample of di-jet events in data and in QCD simulated samples. To perform this measurement, a dataset of jets selected with a more inclusive trigger is needed. Two different prescaled single-jet control triggers, listed in Tables 5.4, 5.5, 5.6, were used for this purpose in 2017 SL, 2017 FH and 2018 FH categories, respectively. The kinematic turn-on is modelled with the more inclusive control trigger, i.e. `HLT_PFJet40(60)`, while the control trigger with tighter thresholds, i.e. `HLT_PFJet60(140)` was less prescaled and was employed to extend the measurement towards high p_{T} in the SL (FH) category as detailed later.

The jet kinematic trigger efficiency can be expressed as follows:

$$\begin{aligned}\epsilon_{\mathrm{kin.}} &= f(p_T(\mathrm{jet}_1), p_T(\mathrm{jet}_2), \eta(\mathrm{jet}_1), \eta(\mathrm{jet}_2)) \\ &= f(p_T(\mathrm{jet}_1), \eta(\mathrm{jet}_1)) \cdot f(p_T(\mathrm{jet}_2), \eta(\mathrm{jet}_2)),\end{aligned} \tag{5.1}$$

where $f(p_T(jet_1), \eta(jet_1))$ and $f(p_T(jet_2), \eta(jet_2))$ are the individual kinematic trigger efficiencies of the first and the second most energetic jet, respectively. The most energetic jet is identified as the tag, while the second leading is defined as the probe. Kinematic cross-correlations between them are neglected in this equation. Previous studies have confirmed the presence of these effects. However, they are accurately accounted for in the simulations, which ensures they are cancelled in the efficiency ratio, as detailed in [35,41].

The offline selection requires two back-to-back jets with $p_T > 40, 112, 116$ GeV and $|\eta| < 2.2$ in the 2017 SL, 2017 FH and 2018 FH category, respectively. As mentioned before, the control triggers with tighter p_T threshold are used to overcome the statistical limitations at high p_T of the more inclusive ones. Studies on the single jet control trigger `HLT_PFJet60(140)` confirmed it becomes fully efficient at 85(220) GeV in the SL (FH) configuration.Therefore, the method can be utilized to extend the probe's trigger efficiency measurement from $p_T > 140(400)$ GeV up to 300(1000) GeV.

To enhance the purity of the probe sample, the tag must pass both the online trigger selection and the offline selection. The efficiency is then calculated as the ratio of events where the probe passes the online and offline requirements and the events where the probe passes the offline selection only. The matching condition between offline jets and online objects at all reconstruction levels require a phase separation $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ of no more than 0.5. Additional jet activity is suppressed by imposing an offline requirement on the p_T of the third leading jet:

$$p_T(jet_3) < 0.3 \cdot (p_T(jet_1) + p_T(jet_2))/2, \tag{5.2}$$

where the parameter values are arbitrary, but they were optimized in an analogous trigger efficiency study for PF jets performed with 2017 data [39] to remove soft jet radiation since soft QCD processes are generally more prone to mis-modelling in simulation and improve the purity of the di-jet sample.

In 2017 the kinematic efficiencies of eras C,D and E differed from F due to a failure in part of the pixel's DCDC converters modules [200]. To combine correctly in a single efficiency curve eras CDE with F for 2017 analyses, the following luminosity event-by-event weight was applied:

$$w = L^{era}_{total}/L^{era}_{trigger}, \tag{5.3}$$

where L^{era}_{total} is the total luminosity collected by the physics trigger per era, and $L^{era}_{trigger}$ corresponds to the luminosity collected by the single jet trigger in the same era. A similar procedure is followed to combine the efficiencies of the 2018 A, B, C and D eras. The resulting trigger efficiencies are shown in the top panel of Figs. 5.6, 5.7, 5.8 for 2017 SL, 2017 FH and 2018 FH respectively.

The scale factor is parametrized to smoothen out statistical fluctuations, for which a gaussian error function, defined in eq. 5.4 as "erf", is employed. The scale factors and their parametrizations are illustrated in the lower panels of Figs. 5.6, 5.7, 5.8, corresponding to 2017 SL, 2017 FH and 2018 FH respectively. This study reveals a noticeable discrepancy in the turn-on curve of data and simulation, especially in the endcap region. The kinematic scale factor's up and down 1 σ variations are obtained from the uncertainty of the scale factor, which is estimated from the covariance matrix of the fit. The effect and treatment of these uncertainties are further discussed in section 5.7.2.

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp\bigl(-t^2\bigr) dt. \tag{5.4}$$

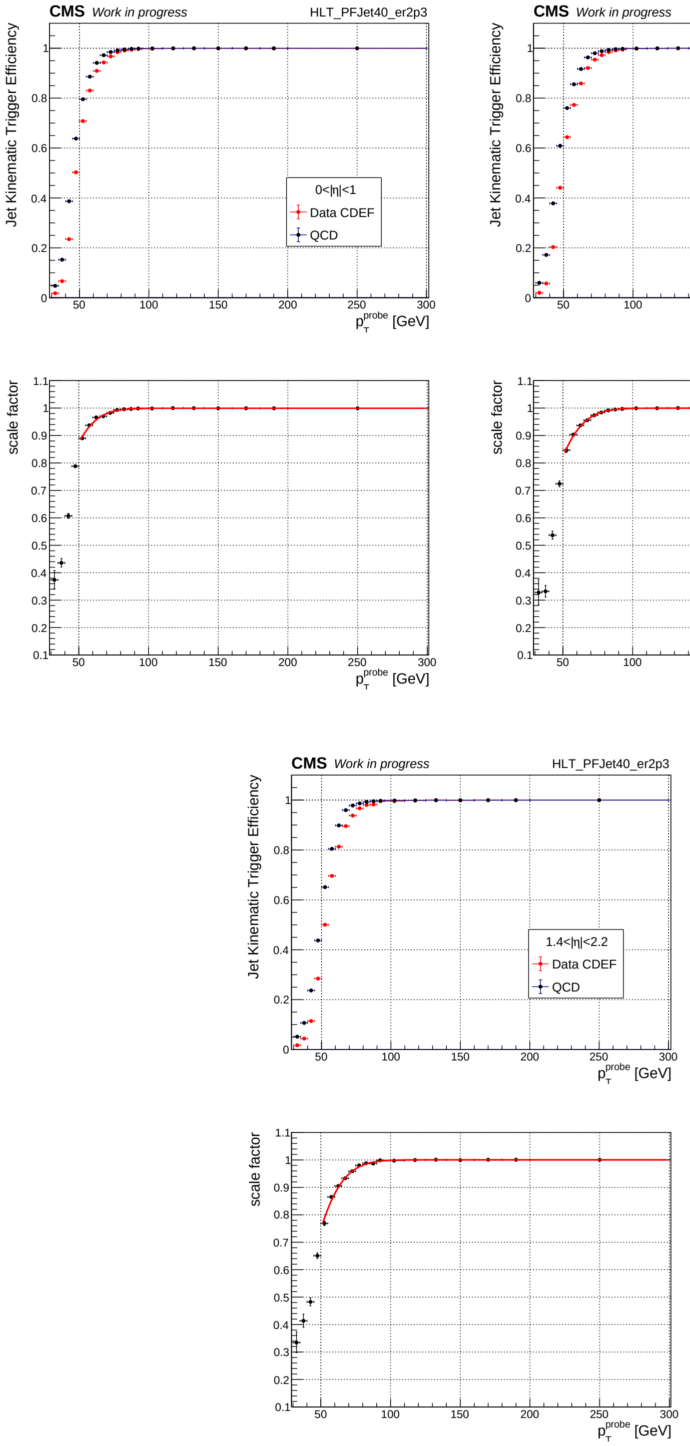


Figure 5.6: Jet kinematic trigger efficiency and scale factor as a function of p_{T} in three $|\eta|$ intervals for 2017 SL analysis, as determined for data (red) and MC (black). The scale factors obtained from the ratio are shown in the bottom panel. They are fitted with a Gaussian error function, shown as a red line.

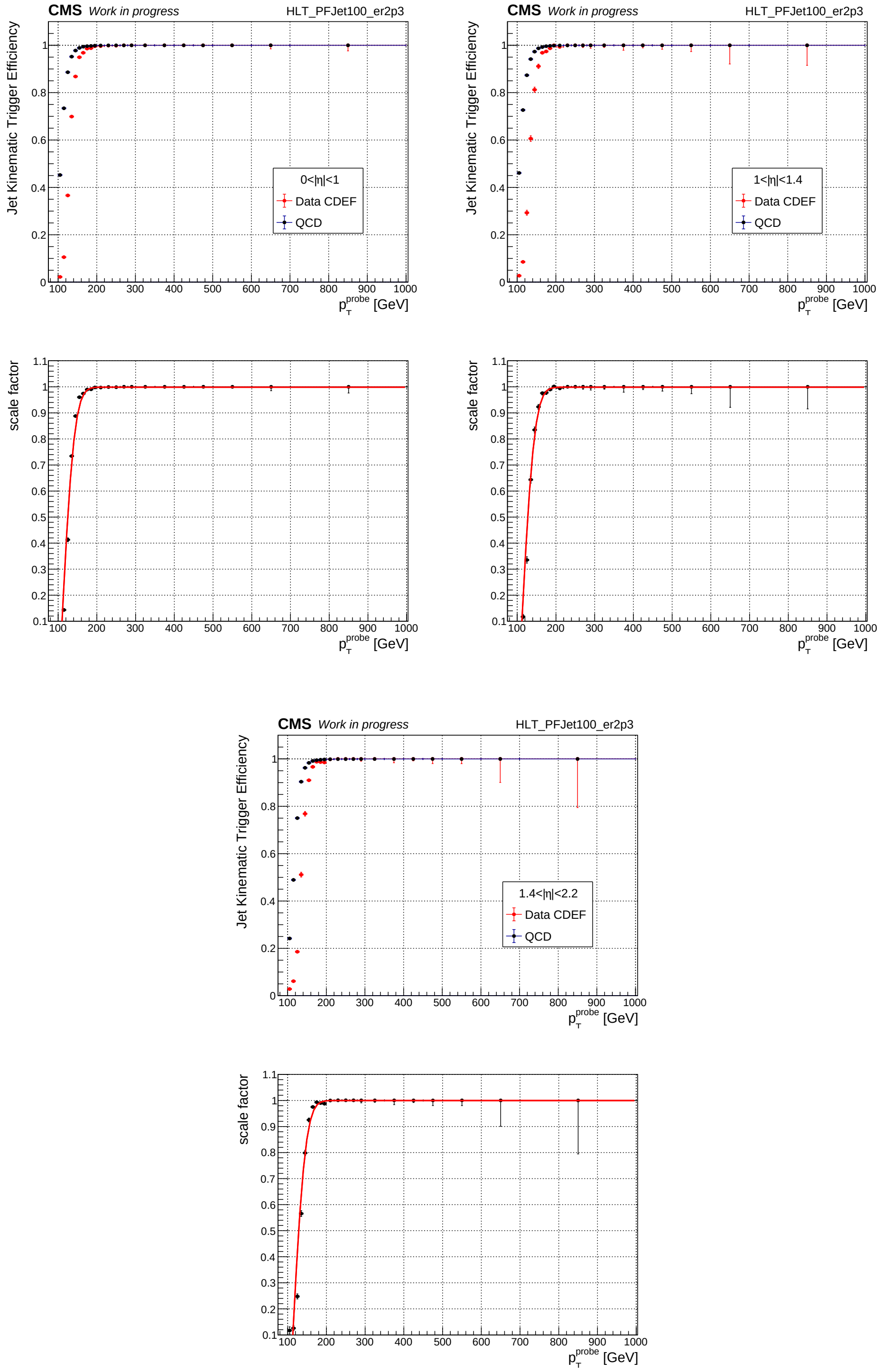


Figure 5.7: Jet kinematic trigger efficiency and scale factor as a function of p_T in three $|\eta|$ intervals for 2017 FH analysis, as determined for data (red) and MC (black). The scale factors obtained from the ratio are shown in the bottom panel. They are fitted with a Gaussian error function, shown as a red line.

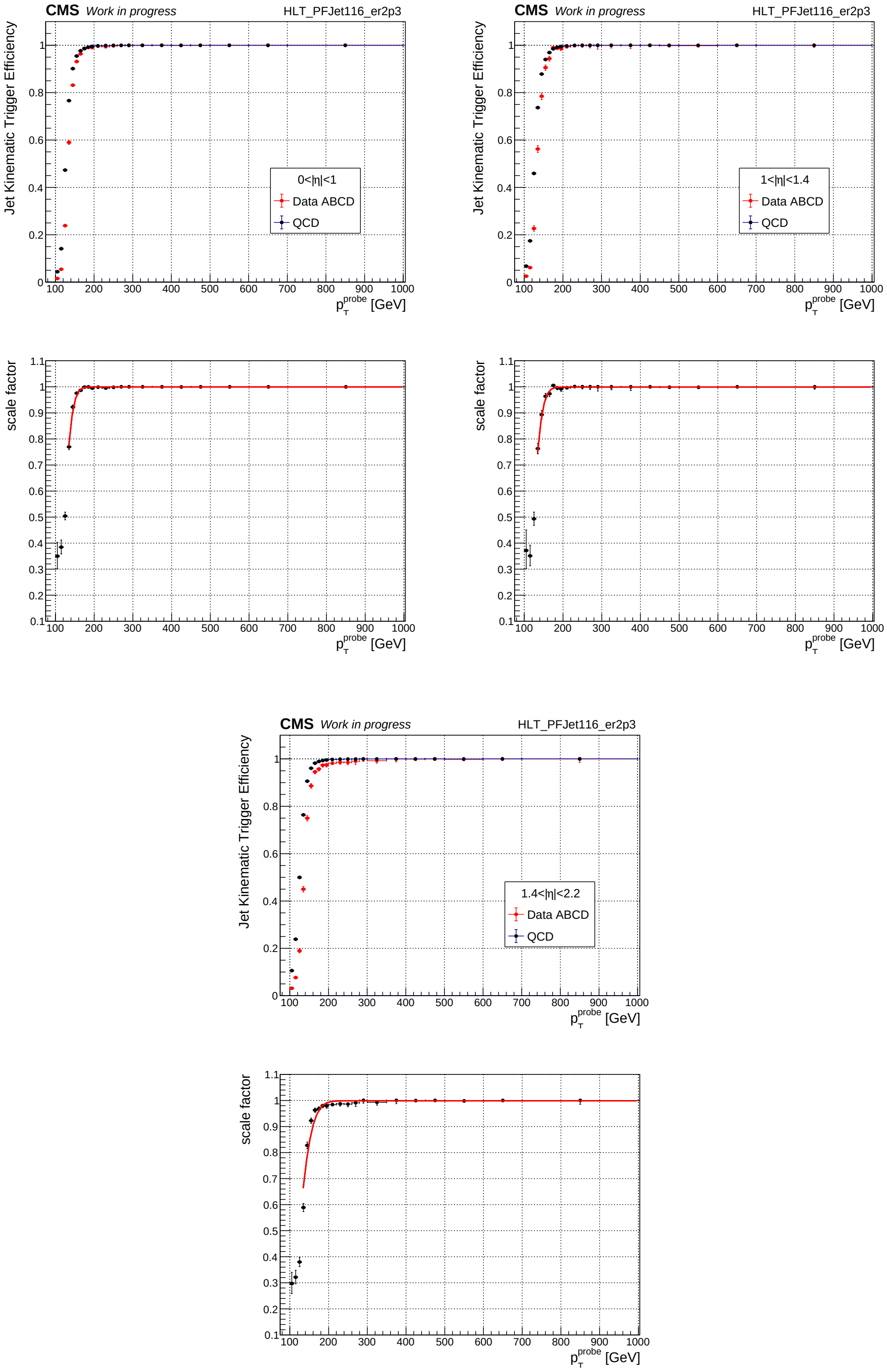


Figure 5.8: Jet kinematic trigger efficiency and scale factor as a function of p_{T} in three $|\eta|$ intervals for 2018 FH analysis, as determined for data (red) and MC (black). The scale factors obtained from the ratio are shown in the bottom panel. They are fitted with a Gaussian error function, shown as a red line.

Muon kinematic trigger efficiency

A tag and probe method was also used to evaluate the muon kinematic trigger efficiency using the J/ψ resonance peak in the $\mu^+\mu^-$ decay channel to guarantee a pure sample of muons with a soft p_T spectrum. The J/ψ resonance, peaking at 3096.900 ± 0.006 MeV [201], allows probing the kinematic turn-on region of the semi-leptonic trigger used in the 2017 analysis.

The kinematic leg in the main HLT sequence of this analysis was emulated using the more inclusive single-muon trigger `HLT_Mu8`, and the same set of kinematic cuts on the online trigger muon objects were applied to both Level-1 and Level-3 muons. The data samples utilized for this efficiency measurement in data correspond to the DoubleMuon dataset, in which single-muon triggers have the lowest prescale, and for the estimation in simulation, QCD muon-enriched samples were employed.

Further selection criteria are applied offline to the muon pair. They must be oppositely charged and pass the muon tight identification criteria. The transverse momentum of both muons must be within 5 and 30 GeV, while a slightly tighter lower threshold of 10 GeV is applied to the tag, and an angular separation ΔR between the tag and probe of $\Delta R(\mu_{probe}; \mu_{tag}) > 0.1$ required. Finally, the invariant mass of the di-muon system, $m_{\mu\mu}$, must lay within the J/ψ mass window 3.10 ± 0.15 GeV. The tag muon must have fired the emulated single-muon trigger and therefore must pass the matching condition $\Delta R(\mu_{probe}; \mu_{L1}) < 0.4$ and $\Delta R(\mu_{probe}; \mu_{L3}) < 0.005$. Applying this matching selection to the probe splits the events into two orthogonal categories, namely those where the probe passes and those where the probe fails the online matching. A simultaneous signal-plus-background fit is then performed on the reconstructed $m_{\mu\mu}$ for both event categories. The function selected for the parametrization is:

$$\mathrm{f}(x) = \{A \exp(\frac{-(x-m_0)^2}{2\sigma^2}) + k_2 x^2 + k_1 x + k_0\} \times \begin{cases} \epsilon & \text{: passing probe} \\ (1-\epsilon) & \text{: failing probe} \end{cases}, \qquad (5.5)$$

where A, m_0, and σ are the normalisation, mean, and width of the Gaussian fit used to model the J/ψ signal, respectively. The parameters k_i are introduced to account for an offset coming from background contributions, modelled by a quadratic function. The trigger efficiency ϵ is obtained from the overall normalisation factor of the simultaneous fit for each p_{T} bin, and the statistical uncertainty on the efficiency is extracted from the fit. Since all the shape parameters of the signal-plus background fit are correlated, the uncertainty related to the background model is also propagated into the efficiency measurement.

The results of the simultaneous fit are shown in Fig. 5.9 for the four p_{T} bins of the probe muon in data, and in Fig 5.10 for MC. The latter were obtained following the exact same procedure but using QCD muon-enriched samples. The resulting efficiencies are shown in Fig. 5.11 along with the scale factor obtained from their ratio.

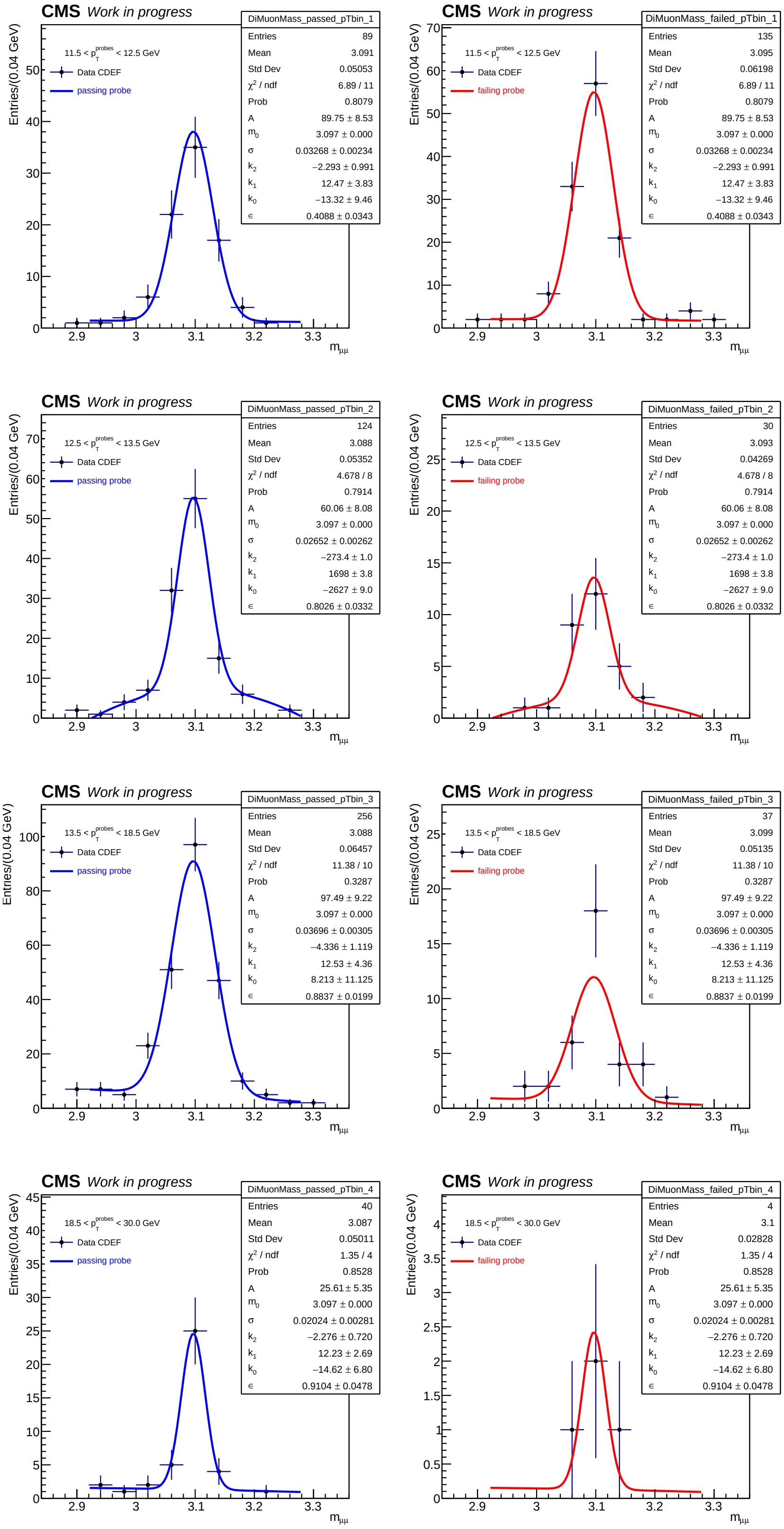


Figure 5.9: Simultaneous fit of the invariant mass di-muon spectra in the two event categories of failing and passing probes in data.

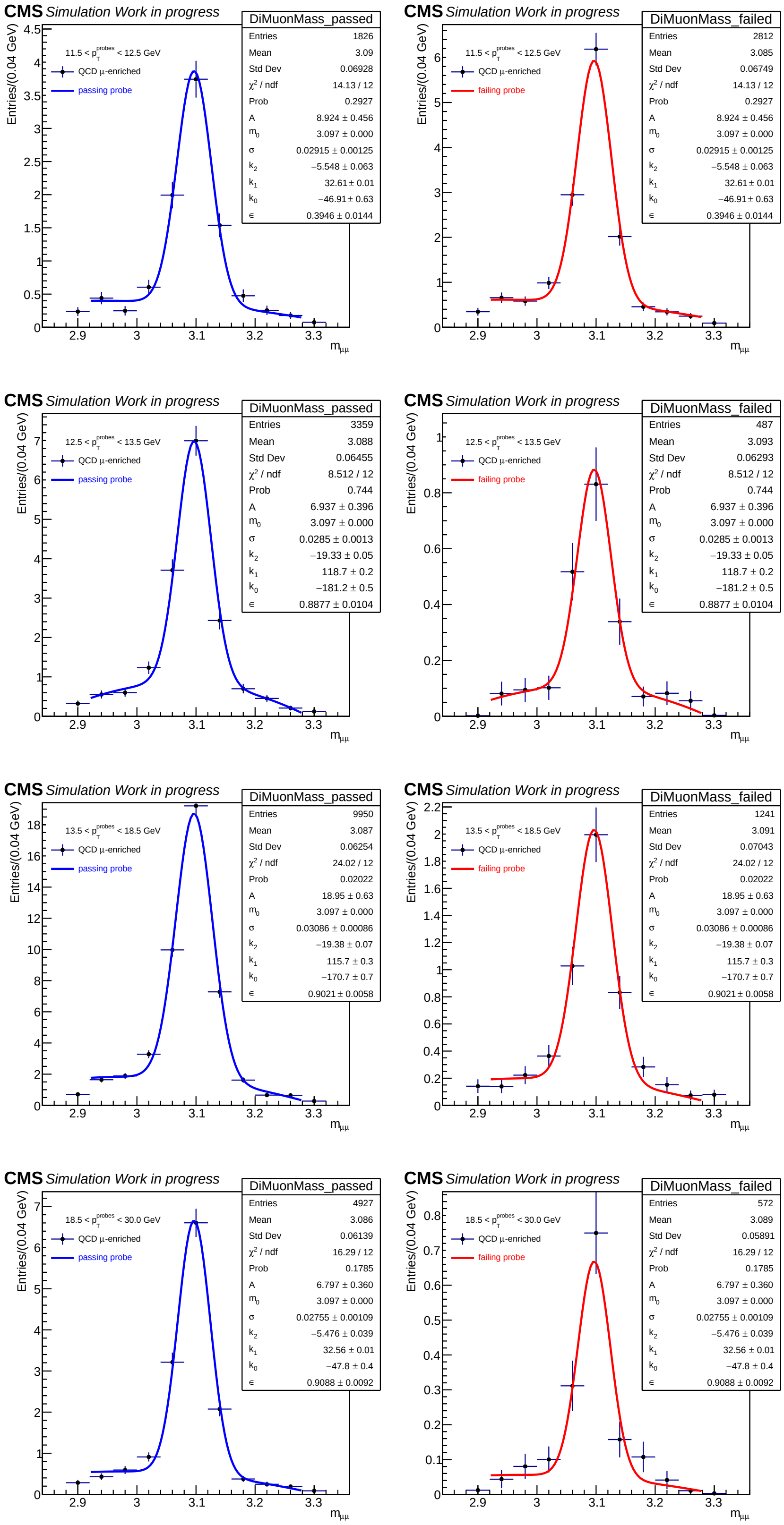


Figure 5.10: Simultaneous fit of the invariant mass di-muon spectra in the two event categories of failing and passing probes in QCD muon-enriched MC samples.

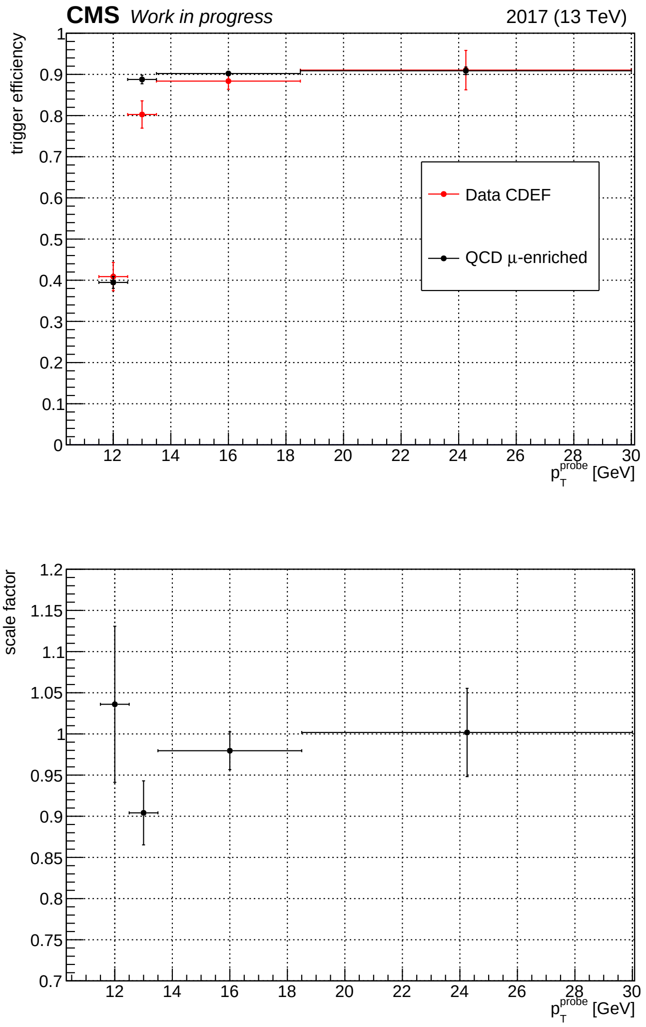


Figure 5.11: In the top panel: muon trigger efficiency measured in data (red) and simulation (black). In the bottom: the ratio plot shows the scale factor correction.

Online b-tag efficiencies

Scale factors are also needed to address inherent differences between the online b-tagging efficiencies in data and MC, which can also depend on the transverse momentum of the jets. For this study, a set of double jet control triggers was used in the FH category. These triggers feature a b-tagging component similar to the analysis trigger but with a single online b-tag requirement and different jet p_{T} thresholds. Meanwhile, in the SL category, a set of control triggers requiring a muon within the jet is also employed, applying the same thresholds to the muon as those used in the analysis triggers. The control triggers cover a wide p_{T} range and are prescaled to optimize the number of recorded events. They are listed, and the corresponding selections are detailed in Tables 5.4, 5.5 and 5.6 for 2017 SL, 2017 FH and 2018 FH categories, respectively. In addition, the corresponding luminosities and the selected p_T range of the probe-jet for each trigger are illustrated in tables in Tables 5.8, 5.9 and 5.10.

The efficiency of the online b-tag selection is estimated with respect to the offline b-tag selection using a tag-and-probe method. Two back-to-back jets passing the control triggers are selected, which are expected to come from a $gg \to b\bar{b}$ process. The b-flavour of the final jets is then verified by using an offline b-tag selection. Processes where the final state gluon(s) would be misidentified as a b quark, as in $bg \to bg$ and $gg, q\bar{q} \to gg$, contribute as possible backgrounds.

Further requirements are applied to the leading and second leading jet in p_{T}, jet_1 and jet_2, respectively:

- $|\eta(jet_i)| < 2.2$, where $i = 1, 2$
- $|\Delta\phi(jet_1, jet_2)| > 2.5$
- jet_i must pass a tight identification criteria
- jet_i must match the corresponding jet trigger objects

As in the jet trigger efficiency studies, events with a third jet are only considered if $p_T(jet_3) < 0.3[p_T(jet_1) + p_T(jet_2)]/2$ for the same purposes. Furthermore, in the SL case, a muon with $p_T(\mu) > 13$ GeV, $|\eta(\mu) < 2.2|$, matched to the muon trigger objects, $|\Delta R(\mu, jet)| < 0.4$ is also required.

In the FH case, the jet_1 and jet_2 are chosen as the probe and tag jets, whereas in the SL case, the probe-jet contains the selected muon, and the other jet is the tag. The probe jet is required to pass the offline b-tag selection medium working point of the DeepJet algorithm, while the tight working point is required for the tag jet to enhance the flavour purity of the probe. To ensure independent samples, the events are split according to the control trigger and the p_T selection of the probe-jet as displayed on the third column of Tables 5.8, 5.9 and 5.10). To suppress events that come from a final state consisting of a pair of gluons where at least one of the gluons splits into heavy flavour quarks, a

Full hadronic control triggers 2017		
HLT Path	**Luminosity** $[pb^{-1}]$	$p_T^{probe} [GeV]$
`HLT_DoublePFJets40_CaloBTagCSV_p33`	79.1	[40,160]
`HLT_DoublePFJets100_CaloBTagCSV_p33`	259.7	[160,220]
`HLT_DoublePFJets200_CaloBTagCSV_p33`	3435.9	[220,400]
`HLT_DoublePFJets350_CaloBTagCSV_p33`	19812.4	[400,1200]

Table 5.8: Control triggers for the 2017 full-hadronic analysis, their luminosity and the corresponding prob jet p_T range used for each trigger.

Semileptonic control triggers 2017		
HLT Path	**Luminosity** $[pb^{-1}]$	$p_T^{probe} [GeV]$
`HLT_Mu12_DoublePFJets40_CaloBTagCSV_p33`	90.0	[40,100]
`HLT_Mu12_DoublePFJets100_CaloBTagCSV_p33`	1538.5	[100,200]
`HLT_Mu12_DoublePFJets200_CaloBTagCSV_p33`	25483.0	[200,500]
`HLT_Mu12_DoublePFJets350_CaloBTagCSV_p33`	36674.5	[500,800]

Table 5.9: Control triggers for the 2017 semileptonic analysis, their luminosity and the corresponding prob jet p_T range used for each trigger.

Full hadronic control triggers 2018		
HLT Path	**Luminosity** $[pb^{-1}]$	$p_T^{probe} [GeV]$
`HLT_DoublePFJets40_CaloBTagCSV_p33`	6.5	[40,130]
`HLT_DoublePFJets100_CaloBTagCSV_p33`	160.4	[130,250]
`HLT_DoublePFJets200_CaloBTagCSV_p33`	779.1	[250,800]
`HLT_DoublePFJets350_CaloBTagCSV_p33`	779.1	[800,1000]

Table 5.10: Control triggers for the 2018 full-hadronic analysis, their luminosity and the corresponding prob jet p_T range used for each trigger.

cut on the tag jet probability of being a two-b-quark jet is enforced. This probability is yielded by the b-tag DeepJet algorithm, and the applied cut on the tag is optimized to maximize the probe's purity. The probe jets are then employed to derive the online b-tag efficiencies in data and simulated QCD Madgraph samples, relative to the offline b-tag selection. These efficiencies were computed as the ratio between the number of probe jets that match the online b-tag object and the number of probe jets as a function of the probe's p_{T} and η.

The efficiencies for the FH analyses of 2017 and 2018 are shown in Figure 5.12. Figure 5.13 shows the relative online b-tag efficiencies for b-jets and b-jets with a muon in the 2017 SL analysis. The scale factors are obtained from a linear fit to ratio of the efficiencies in data and in simulation. Figures 5.14 and 5.15 show the scale factors for the FH, 2017 and 2018, and semileptonic analyses, respectively. The uncertainty bands of the fits are employed to account for the relevant systematic uncertainties.

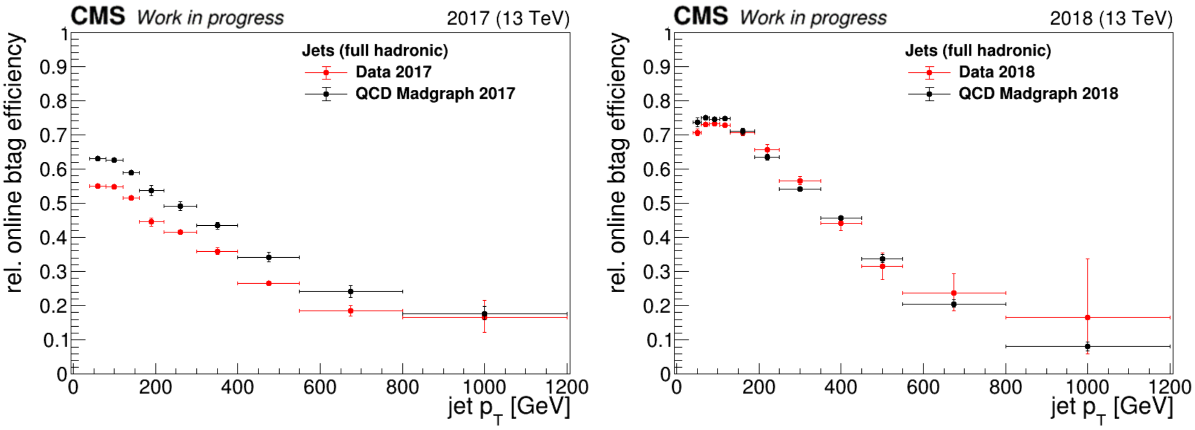


Figure 5.12: Relative online b-tag efficiencies for b jets in data and QCD simulation for the full-hadronic analyses with 2017 (left) and 2018 (right) datasets.

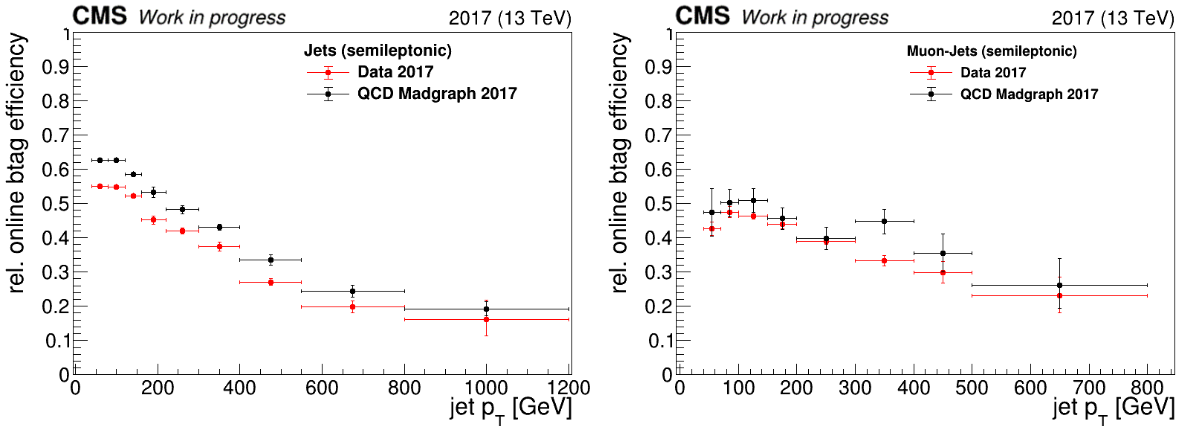


Figure 5.13: Relative online b-tag efficiencies for b jets (left) and b jets with a muon (right) in data and QCD simulation for the SL analysis with 2017 dataset.

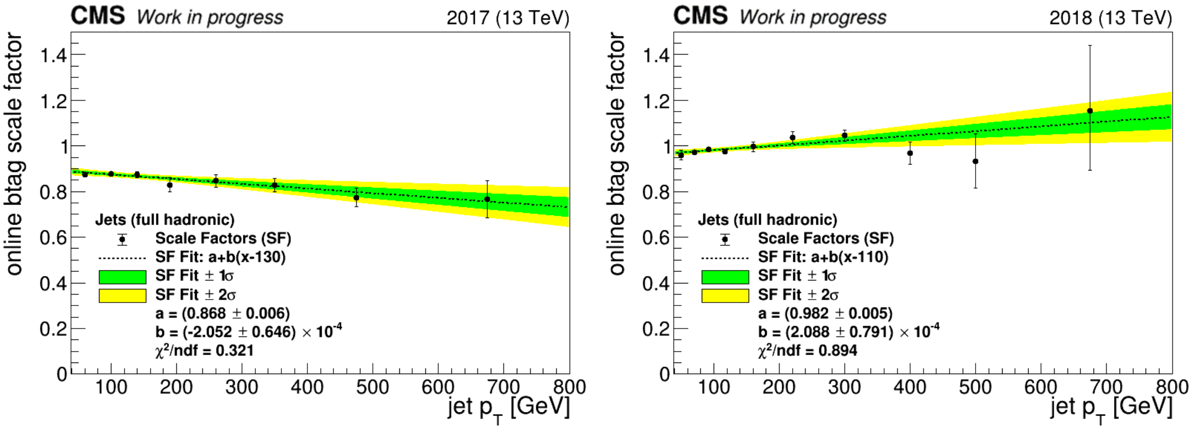


Figure 5.14: Online b-tag scale factors for b jets for the FH analyses with 2017 (left) and 2018 (right) datasets.

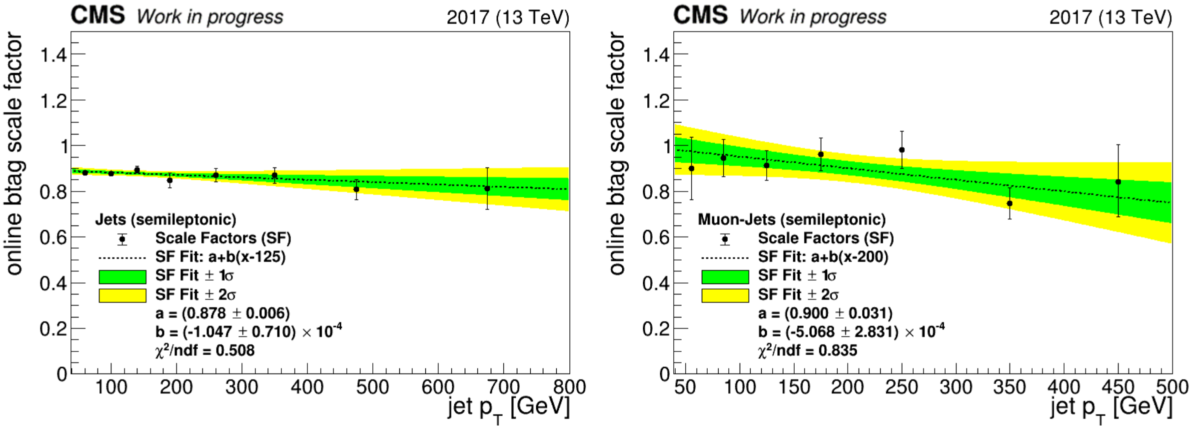


Figure 5.15: Online b-tag scale factors for b jets (left) and b jets with a muon (right) for the semileptonic analysis with 2017 dataset.

5.6 Offline event selection

Once an event meets the requirements set by the main analysis triggers described in section 5.5.1 and passes the CMS certification for physics analysis, it is considered for further analysis. This involves offline selection steps that aim to minimize background events while retaining potential signal events. This section discusses the complete selection process of the analyses presented in this thesis.

5.6.1 Primary vertex selection

All selected events must contain a well-identified and reconstructed primary vertex, corresponding to the hard-scattering process. In each event, the PV candidate with the largest p_{T}^2 sum is taken as the origin of the interaction. The PV selection is based on requiring that the PV is found in a position within 2 cm from the beampipe and within 24 cm of the interaction point. Furthermore, a minimum of four good tracks must be identified in the vertex determination fit to ensure a proper reconstruction.

5.6.2 Jet selection

Events must additionally fulfil the requirement of having at least three jets which must meet a tight identification working point criteria to guarantee the rejection of fake and noise-driven reconstructed jets. This method was deployed centrally within CMS, and requires at least two constituents within a jet, whose neutral hadron and electromagnetic components lay below 90%, muon fraction below 80% and charged electromagnetic fraction below 80% [202].

The three leading jets in p_{T} must also satisfy a loose pile-up ID requirement if they have a transverse momentum below 50 GeV, according to the internal CMS recommendations [203]. A Boosted Decision Tree is trained to discriminate prompt jets from PU jets employing several jet shape and kinematic variables. For instance, the training employs the number of primary vertices, the number of charged constituents within a jet and the jet cone characteristics.

Subsequent kinematic selections are applied, which differ for SL and FH categories and between data-taking years. In the 2017 (2018) FH category, p_{T} thresholds of 110 (130) GeV and 100 (130) GeV are applied to the leading and second leading jets, respectively. A p_{T} threshold of 40 GeV is also imposed on the third leading jet of the 2017 and 2018 FH categories. In addition, jets must lie within the pseudorapidity window of $|\eta| < 2.2$ to ensure the best b-tagging performance, and two leading should have a distance of $|\Delta\eta_{12}| < 1.5$ to improve background suppression. Furthermore, the three leading jets must always fulfil a separation of $\Delta R_{ij} > 1$ to reduce backgrounds coming from a gluon splitting into a pair of b-quarks. The well-established DeepJet b-tagger, which is described in section 4.2.5, is utilized to determine the offline b-jet identification criteria. To identify a b-jet, the medium working point is employed, which corresponds to a 1% misidentification rate, achieved at an arbitrary value of the b-tag discriminant d > 0.3040 (d > 0.2783) in 2017 (2018). In the signal region, short-handed "bbb", this b-tag selection is required for each of the three leading jets. As mentioned before on section 5.3, a reverse b-tag control region is defined to be employed in the background modelling, labelled "bbnb". This region is constructed by selecting events where the third leading jet fails the loose working point of the DeepJet algorithm, corresponding to a 10% misidentification rate, or d < 0.0532 (d < 0.0490). A validation region is defined between SR and CR, i.e. with 0.0532 > d > 0.3040 (0.0490 > d > 0.2783) in 2017 (2018). This region is utilised to test the analysis methods with real data as a final step before unblinding the data in the SR.

In the 2017 SL category, transverse momentum thresholds of 60 and 50 GeV are applied in the leading and second leading jet, respectively, while the third leading jet must pass a looser threshold of 30 GeV. Additionally, jets must also be found within $|\eta| < 2.2$, and the two leading jets should meet the same $|\Delta\eta_{12}| < 1.5$ requirement as in the FH analyses. Definitions of SR, CR and VR are equivalent to the 2017 FH case.

Further corrections to improve the mass resolution are applied to the three analyses during the selection process. One is a deep-neural-network (DNN) based technique of b-jet energy regression that allows to account for missing energy, mostly because of neutrinos emitted in semi-leptonic decays of B-hadrons [204]. A total of 43 jet kinematic, track and event-based variables are employed for the classification training. The result is a multiplicative factor to the transverse momentum of the jet which is provided centrally [205]. Furthermore, a final state radiation (FSR) recovery technique was deployed to recover possible gluon irradiations emitted by the b-jets before hadronizing. FSR candidates are considered starting from the fourth leading jet. If this candidate is in close proximity to one of the three (main) leading jets, with a $\Delta R < 0.8$, its four-momentum is added vectorially to that of the neighbouring main jet. A simple approximation of considering only one FSR candidate per main jet was considered, as it provided a better performance in our analysis. Fig. 5.16 illustrates the effects of the FSR recovery in the di-jet mass distribution for the 1 TeV signal mass point of the 2018 FH analysis.

5.6.3 Muon selection

In the 2017 SL category, one muon with tight identification in the same eta region as the jets is selected with $p_{\mathrm{T}}(\mu) > 13$ GeV. It must lie within the first or second leading b-tagged jet hadronization cone, i.e. $\Delta R(\mu, j) < 0.4$. If both jets have muons, the one with the largest muon p_{T} is considered the one that fired the muon component of the SL trigger and, therefore, chosen for the application of the muon trigger scale factors in simulation.

To avoid any overlap of events with the semi-leptonic channel, events with a muon reconstructed within one of the two leading jets ($\Delta R(\mu, j) < 0.4$), passing tight identification with $p_{\mathrm{T}}(\mu) > 13$ GeV in the pseudorapidity region $|\eta| < 2.2$, are vetoed. A previous

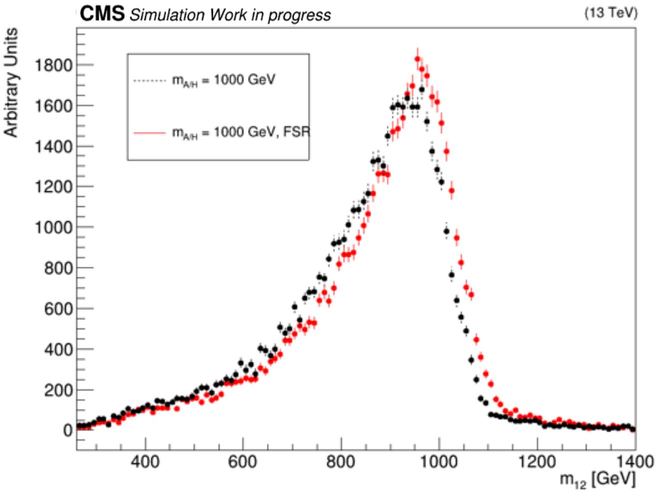


Figure 5.16: Effect of the FSR recovery technique in the 1 TeV signal mass point of the 2018 FH analysis. A clear shift toward higher masses and closer to the nominal mass is observed, as well as an improvement in the relative width of the peak.

study [40] showed that the effect of this veto on the signal shape is relatively small, with a slight upward shift and narrowing of the signal shapes being observed.

5.6.4 Matching to online objects

The matching of offline objects to the online-triggered objects, i.e., jets, muons, and b-tag objects is crucial for the proper application of the trigger efficiency scale factor corrections to simulated samples. The matching criteria are met in all cases if an angular distance of $\Delta R < 0.3$ between the offline and online object is met.

A key enhancement in the sensitivity of the present analysis resulted from revising the strategy for the online b-tagged objects matching. Previous analyses [39–41] would only test if the two leading offline b-tagged jets match the online b-tagged objects. This is rather inefficient since the online b-tagging tests for up to six jets to meet the flavour selection criteria. Furthermore, the fact that the b-tagging efficiency deteriorates at high transverse momentum, makes it more likely that a softer, and not necessarily both of the two leading jets, fired the b-tag leg of the trigger. Following this method resulted in a striking drop in the signal efficiency, particularly towards higher masses.

During the development of the analysis strategies presented in this thesis, it was studied how this undesired effect can be recovered. It was found that nearly 100% of the discarded events were recovered when, instead of considering only the two leading jets as candidates for the matching to online objects, one requires that at least two of the three leading b-tagged jets match the online b-tag objects. The effect on the signal efficiency is shown in Fig. 5.17 (left) for 2018 FH. The new strategy results in up to a factor of five improvement in signal efficiency. A similar effect is observed when running over QCD in the signal region, shown in Fig. 5.17 (right). Considering these two results, a corresponding increase in the overall significance is expected. In contrast, there is no significant effect in the other regions, i.e. CR and VR, given the third leading jet always fails the offline b-tagging criteria in these regions.

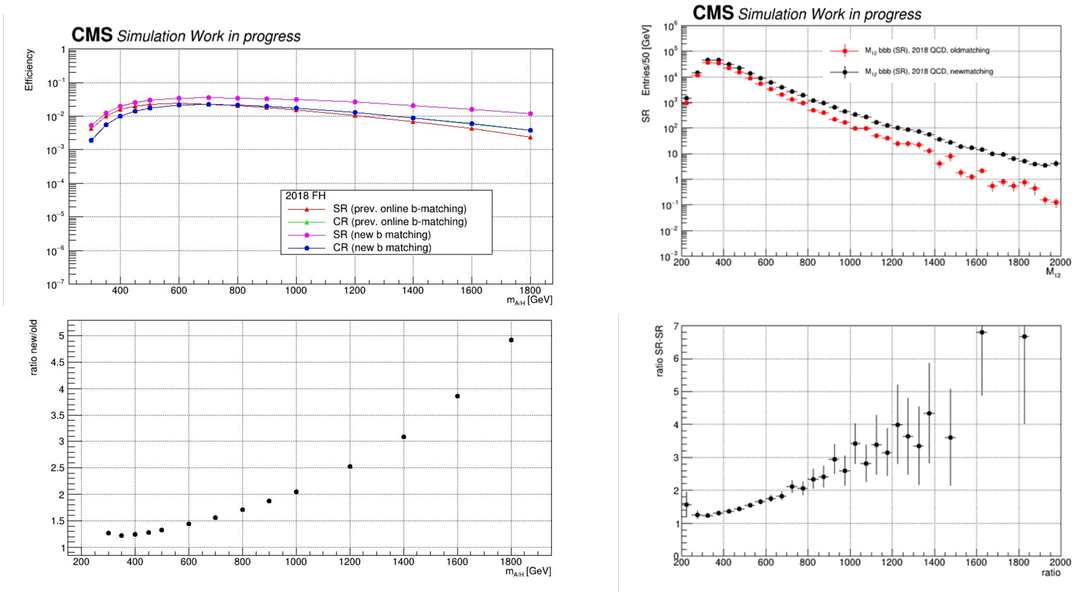


Figure 5.17: Effect of the new matching strategy of offline jets to online b-jets on the signal efficiency of 2018 FH channel (left) and on QCD in the signal region (right). In both figures, the lower plots show the ratio of events in the signal region resulting from employing the new strategy divided by the results from the previous approach. Offline b-tag scale factors (see section 5.6.7) are not considered in these plots. However, this does not have an impact on the ratio.

5.6.5 Noise filters

A further skim is performed in the SR in order to remove events with possible anomalous signals caused by instrumental and reconstruction effects. This includes, for instance, filters on events with signals induced by noise on HCAL or ECAL. The recommendations of the CMS jet and missing energy group, outlined in Ref. [206], were followed. The application of these filters in the analysis resulted in a fraction of 0.003%, 0.03% and 0.04% of events being removed from the SR in the 2017 SL, 2017 FH and 2018 FH categories, respectively.

5.6.6 Analysis cutflow

The number of events retained after the main cutflow steps for the datasets studied are summarized in Table 5.11, Table 5.12 and Table 5.13, corresponding to the three analyses: 2017 SL, 2017 FH, and 2018 FH channels, respectively. The reverse b-tag control region contains about seven times as many events as the signal region, while the population of the validation region is comparable to the signal region.

Step	CR	VR	SR
Trigger	18 620 388	18 620 388	18 620 388
Primary vertex	18 362 940	18 362 940	18 362 940
Jet ID	18 362 900	18 362 900	18 362 900
Jet kinematics	8 484 182	8 484 182	8 484 182
$\Delta\eta_{12}$	6 456 513	6 456 513	6 456 513
ΔR_{ij}	4 612 586	4 612 586	4 612 586
μ selection	3 313 220	3 313 220	3 313 220
b-tag (bbb/bbnb)	1 223 061	250 296	260 248
Online b jet match	1 195 593	244 813	254 962
Noise filters	1 195 593	244 813	254 955

Table 5.11: Cutflow table of the 2017 SL channel showing the number of events retained after each selection step in the data for the SR, CR, and VR regions separately.

Step	CR	VR	SR
Trigger	12 189 561	12 189 561	12 189 561
Primary vertex	12 055 657	12 055 657	12 055 657
Jet ID	12 055 657	12 055 657	12 055 657
Jet kinematics	5 685 338	5 685 338	5 685 338
$\Delta\eta_{12}$	4 833 397	4 833 397	4 833 397
ΔR_{ij}	3 334 646	3 334 646	3 334 646
b-tag (bbb/bbnb)	938 342	155 006	181 655
Online b jet match	931 467	153 888	180 497
Noise filters	931 467	153 888	180 446
μ veto	838 131	138 463	161 077

Table 5.12: Cutflow table of the 2017 FH channel showing the number of events retained after each selection step in the data for the SR, CR and VR regions separately.

Step	CR	VR	SR
Trigger	21 456 980	21 456 980	21 456 980
Primary Vertex	21 315 660	21 315 660	21 315 660
Jet ID	21 315 660	21 315 660	21 315 660
Jet kinematics	10 017 131	10 017 131	10 017 131
$\Delta\eta_{12}$	8 760 664	8 760 664	8 760 664
ΔR_{ij}	5 949 006	5 949 006	5 949 006
b-tag (bbb/bbnb)	1 724 211	315 199	346 019
Online b jet match	1 710 786	312 802	344 138
Noise filters	1 710 786	312 802	344 010

Table 5.13: Cutflow table of the 2018 FH channel showing the number of events retained after each selection step in the data for the SR, CR and VR regions separately.

5.6.7 Corrections to simulated samples

Monte Carlo simulations are not capable of reproducing the data perfectly for detector effects. As a result, corrections must be applied to recover the differences. Apart from the aforementioned trigger efficiency scale factors, other corrections are aimed to improve the modelling of, e.g. detector imperfections or inefficiencies that were not represented in the simulations, or data-taking conditions that varied during operation. Additionally, Monte Carlo generators have limitations in terms of theoretical accuracy and statistics. A brief description of the corrections applied to simulations is given below. These corrections introduce uncertainties that affect the signal efficiency. The treatment of such uncertainties is described in section 5.7.2.

Pile-up re-weighting

The pile-up distribution in MC simulations usually differs from the actual distribution in data. The simulations are re-weighted event-per-event to reproduce the observed pile-up conditions. As an example, figure 5.18 shows the recorded structure of the 2018 pile-up profile, along with the distribution in simulated QCD samples, and the resulting re-weighting factor in the bottom panel. A similar procedure was followed for the 2017 analyses. The uncertainties in this correction are related to the total inelastic cross-section calculations. In the latest data reprocessing, these uncertainties impacted the total signal yield by about 1% or less and were consequently neglected.

Jet energy scale and resolution

Jet energy scale and resolution are corrected to model the data in a more accurate prescription, as described in section 4.2.4. These uncertainties related to these corrections

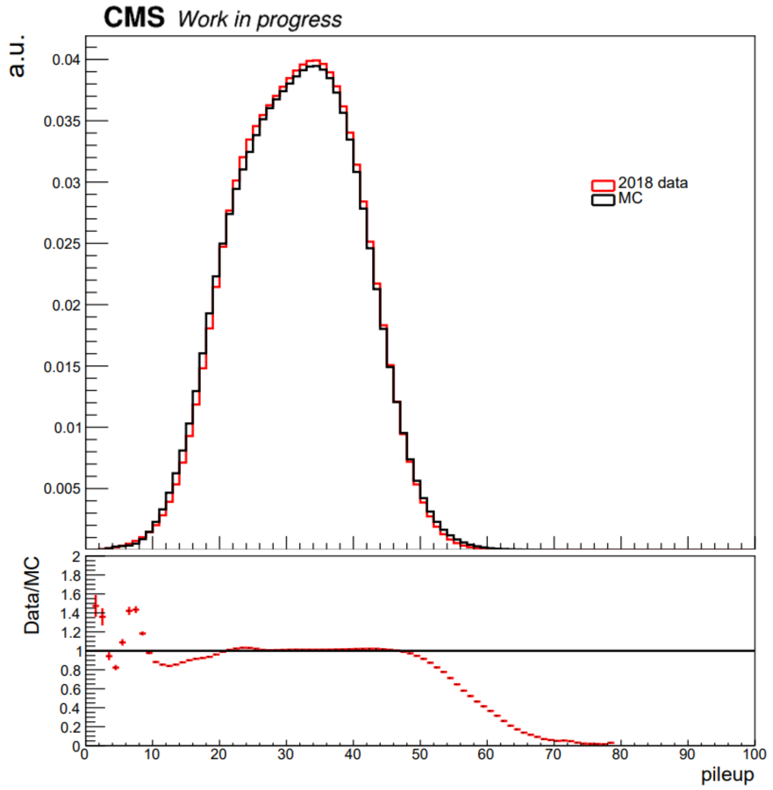


Figure 5.18: Pile-up profile in 2018 data-taking period at 13 TeV and in simulated samples. The ratio of these distributions, employed in the re-weighting procedure, is shown in the bottom panel.

have a significant impact on the main observable of the analysis which will be discussed in the treatment of the systematic uncertainties (section 5.7.2).

Trigger efficiency correction

As discussed in section 5.5, these corrections account for differences in the trigger efficiency in data and simulation concerning jet and muon kinematics and b-jet tagging at trigger level.

Offline jet-flavour identification efficiency correction

Similarly, offline jet-flavour identification performance must be corrected, as discussed in 4.2.5. These scale factors are centrally provided by the CMS Collaboration and have a generally large impact on the signal normalisation, in the order of 10%, as will be shown in section 5.7.2.

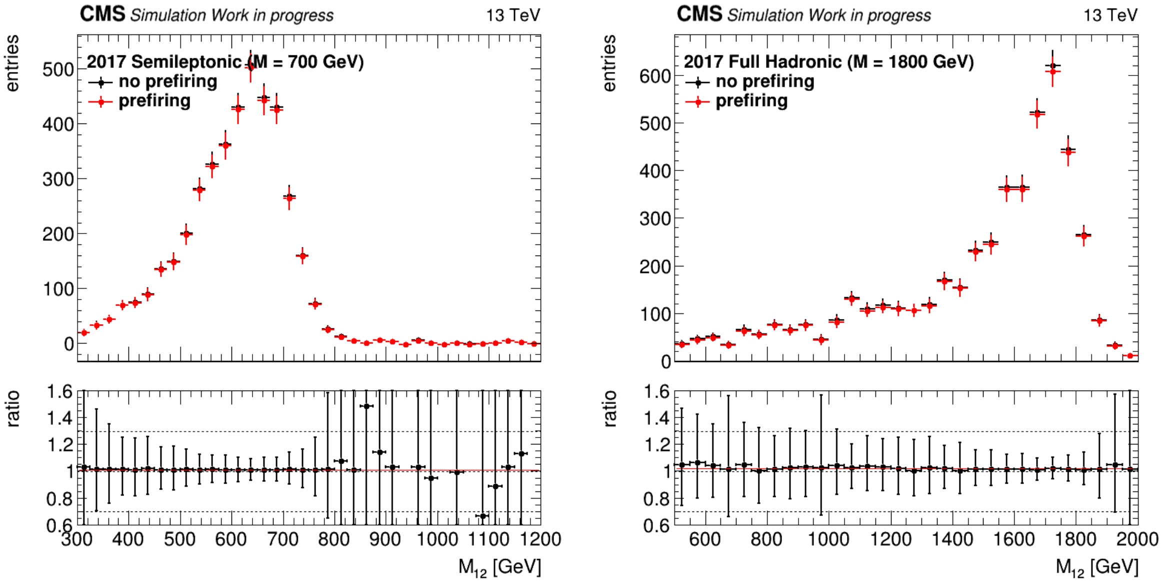


Figure 5.19: M_{12} distributions for Higgs bosons before and after applying the L1 prefiring weights [207]. On the left (right), a 700 (1800) GeV boson in the semileptonic (full hadronic) channel with 2017 data. No significant shape-altering effect due to prefiring weights is observed.

Prefiring weights

To account for the L1 ECAL and muon prefiring, an additional correction is applied. This correction takes into account the effect of a timing drift that can cause a signal for an event to be triggered in a preceding bunch crossing instead of the one containing the event. Due to Level-1 trigger rules, events can self-veto by firing two consecutive bunch crossings. The outcome is emulated in simulation by reducing the event weight according to the CMS recommendations [207]. No significant impact on the shape of the signal distributions is observed (Fig. 5.19). However, as one can see in Fig. 5.20 a 0.3% to 2% reduction in the signal efficiencies is observed in the 2017 data. Negligible effects were observed in 2018, for then no L1 ECAL prefiring occurred.

In addition, a correction is applied to 2018 MC to account for technical issues on HEM15/16, where the power supplies of two HCAL modules failed during data-taking [208]. Consequently, the energy of the jets within the region $1.57 < \phi < -0.87$ and $-2.5 < \eta < -1.3$ was scaled down by 20%. The observed effect from the HEM15/16 issue resulted to be smaller than the jet energy corrections (see section 5.7.2). As a result, the HEM15/16 issue is treated as a systematic uncertainty, following the JetMet recommendations.

The efficiency of the muon tight identification criteria imposed in the 2017 SL category

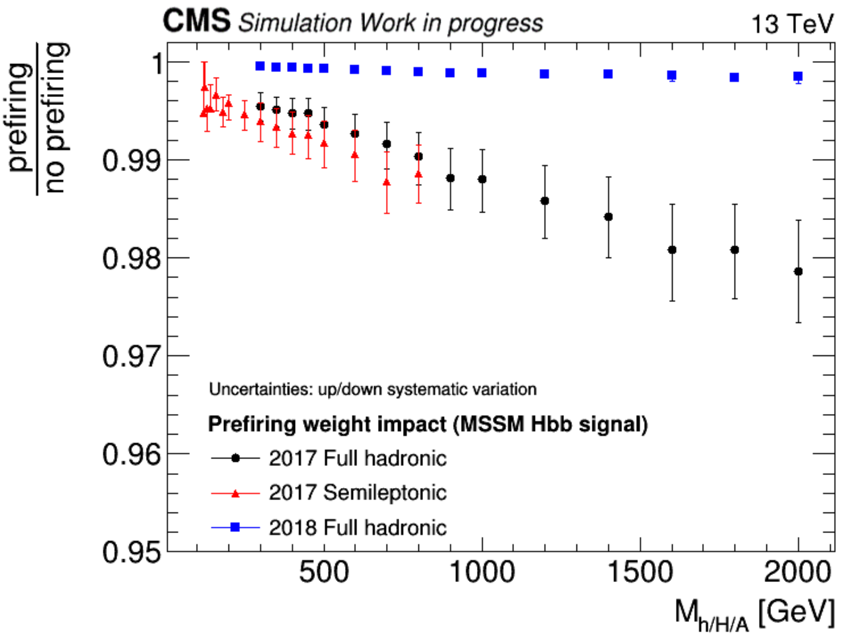


Figure 5.20: Impact of L1 ECAL and muon prefiring on the signal. The uncertainties are the estimated systematic variations as prescribed in [207].

must also be calibrated in simulations. This correction was provided centrally by the CMS collaboration and as described in section 4.2.3.

5.7 Signal modelling

Signal events of the b-associated production process of a heavy neutral Higgs and subsequent decay into a pair of b-quarks are simulated at NLO with the POWHEG generator, interfaced with PYTHIA8 to model hadronization and shower processes. In general, the natural width of the BSM Higgs boson will depend on the parameters of the underlying model. However, this width generally falls within the range of approximately 1-20% of the experimental resolution, as detailed in section 2.2.1.

Fourteen mass points are simulated per analysis category, ranging from 125 GeV to 700 GeV in the SL analysis and from 300 to 1800 GeV in the FH category. Across all samples, an arbitrary production cross-section of 1 pb is assumed to facilitate the extraction of the fitted signal strength through the signal-plus-background fit as a limit in the cross-section times branching fraction.

The signal efficiency is defined as the number of events that pass all corrections and selection steps divided by the number of generated events with positive weight. Fig. 5.21 shows the signal efficiency corresponding to the fourteen mass points in the three analysis regions (SR, CR, and VR), for 2017 SL, 2017 FH, and 2018 FH channels, respectively.

In the SL category, the maximum signal efficiency in the SR is reached for the 400 GeV mass at a value of approximately 0.2%. The signal efficiency for the lower Higgs boson masses is reduced by the imposed kinematic thresholds on the di-jet plus muon system. On the other hand, at Higgs boson masses above 500 GeV the main limiting factor to the signal yield is the b-tagging selection. Similarly, in the FH categories, the signal efficiency increases steadily in the mass range below 600 GeV, where it reaches its maximum value, reflecting the kinematic selection. The decrease beyond the mass of 600 GeV is attributed to the performance of the b-tagging, which becomes increasingly inefficient at high transverse momenta. The signal efficiency is overall slightly higher in 2018 FH compared to 2017 FH due to the loosened online b-tag thresholds and improved particle identification techniques, as detailed in section 3.2.4. Across all analyses, the VR is notably signal depleted, evidenced by an efficiency lower by approximately one order of magnitude. Similarly, the CR, comprising approximately 7 times more events in data than the signal region, further accentuates the signal depletion owing to its diminished signal efficiency.

5.7.1 Parametrization of signal templates

To execute the signal-plus-background fit, as detailed in section 5.3, a parametric description of the signal templates is employed.

The double-sided Crystal Ball (DSCB) function was found to effectively describe the signal templates for all mass hypotheses of the Higgs bosons and analysis categories. This function is characterized by a Gaussian core with two power-law tails stitched to each side. The analytical form of this function, defined by six parameters, is presented below:

$$f(x) = N \cdot \begin{cases} e^{-0.5\alpha_{\text{low}}^2} \cdot \left[\frac{\alpha_{\text{low}}}{n_{\text{low}}}\left(\frac{n_{\text{low}}}{\alpha_{\text{low}}} - \alpha_{\text{low}} - t\right)\right]^{-n_{\text{low}}} & \text{if } t < -\alpha_{\text{low}} \\ e^{-t^2/2} & \text{if } -\alpha_{\text{low}} \leq t \leq \alpha_{\text{high}} \\ e^{-0.5\alpha_{\text{high}}^2} \cdot \left[\frac{\alpha_{\text{high}}}{n_{\text{high}}}\left(\frac{n_{\text{high}}}{\alpha_{\text{high}}} - \alpha_{\text{high}} + t\right)\right]^{-n_{\text{high}}} & \text{if } \quad t > \alpha_{\text{high}} \end{cases} \tag{5.6}$$

In equation 5.6, N is a normalisation parameter, $t = (x - \mu_{\text{CB}})/\sigma_{\text{CB}}$, μ_{CB} is the peak of Gaussian distribution, σ_{CB} represents the width of the Gaussian, α_{low} (α_{high}) parametrizes the mass value where the distribution of the invariant mass becomes a power-law function on the low mass (high mass) side, together with n_{low} (n_{high}), the exponent of this function.

Figure 5.22 shows the signal parameterization of the M_{12} distribution using the double-sided Crystal Ball function for the 2017 SL category in simulation. A representative neutral Higgs boson mass point is displayed for each fit range (FR). Fig. 5.23 shows the signal parameterization for 2017 FH and Fig. 5.24 for 2018 FH channel. As can be seen in these figures, various effects impact the shape of the signal distributions. For instance,

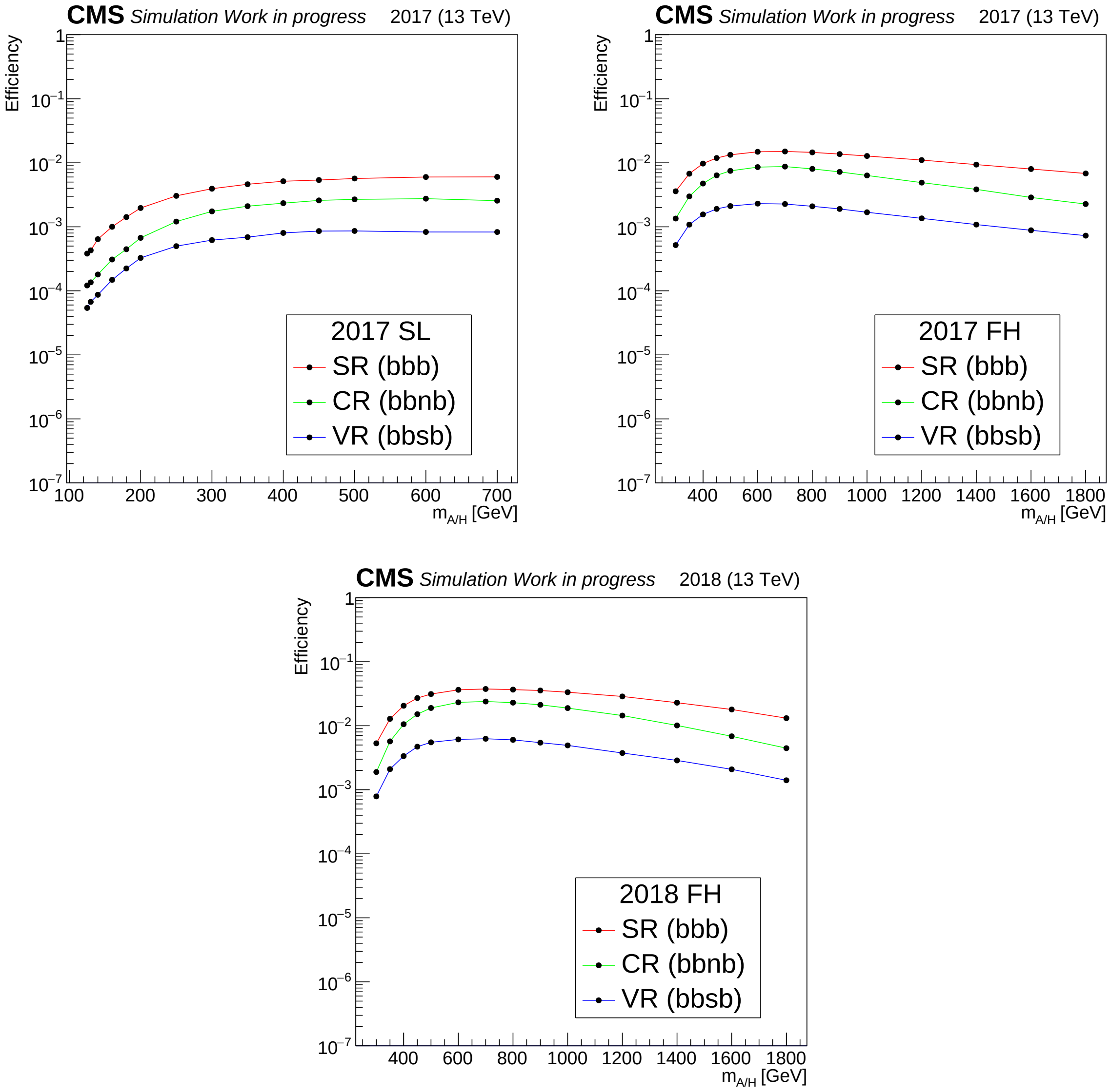


Figure 5.21: Signal efficiency per mass point after the full analysis selection and including all corrections and scale factors in Signal Region (red), Control Region (green), and Validation Region (blue) for 2017 SL (top left), 2017 FH (top right) and 2018 FH (bottom).

the impact of kinematic thresholds at the lowest mass points, indicated by 350 GeV in the 2017 and 2018 FH analyses and 160 GeV in the 2017 SL analysis, resulted in a sharp rise on the left-side tail of the distribution. This effect diminishes in significance for mid-range mass points. At higher mass points, like 1200 (600) GeV in the FH (SL) analysis, the signal distributions exhibit a pronounced left-side tail, partly attributed to missing energy stemming from neutrinos in semi-leptonic decays or unrecovered final state radiation. In addition, a background arises from the events where the pairing of b jets does not correspond to both daughters of the Higgs boson, which contributes to the tails of the distributions; in particular, the high-mass tails of the lower-mass points are attributed to this effect in all categories. The goodness of fit, described by the χ^2/ndf and the p-value, is provided on each plot, indicating that the DSCB function can perfectly model all the signal shapes in this analysis with only one analytic function.

5.7.2 Systematic effects

Systematic effects can have an impact on the total number of events, or alter the signal distribution. Depending on the effect in the signal templates, they are handled differently in the statistical inference procedure (as explained in section 5.10.1). Therefore, the sources of uncertainties should be categorized accordingly, as listed below.

Shape-altering systematic uncertainties

As the name suggests, shape-altering systematic uncertainties produce a sizable impact signal shape. Additionally, they may also affect the normalisation or total signal yield when varied by one standard deviation (upward $+1\sigma$ and downward -1σ). The relevant systematic uncertainties that produce this effect across all analysis categories are those related to the jet energy corrections, namely the jet energy scale (JES) and jet energy resolution (JER).

Not all the parameters of the signal models are necessarily affected by these systematic uncertainties. The DSCB mean (sigma) is found to be the most influenced by the JES (JER) variation. The parameterization is further done on each signal template while varying JES and JER by 1σ up and down, while fixing the DSCB parameters to those of the nominal function except the mean (sigma) in the case of JES (JER). The results are exemplified for the selected mass points in each fit range of the 2018 FH analysis in Figs. 5.25 and 5.26. As expected, one finds that an upward (downward) shift of the JES nuisance parameter shifts the signal shape towards upper (lower) masses, while the corresponding shift of the JER nuisance parameter leads to a widening (narrowing) of the shape. Similar effects are observed in the 2017 SL and FH categories, for which the relevant figures are shown in Appendix D.

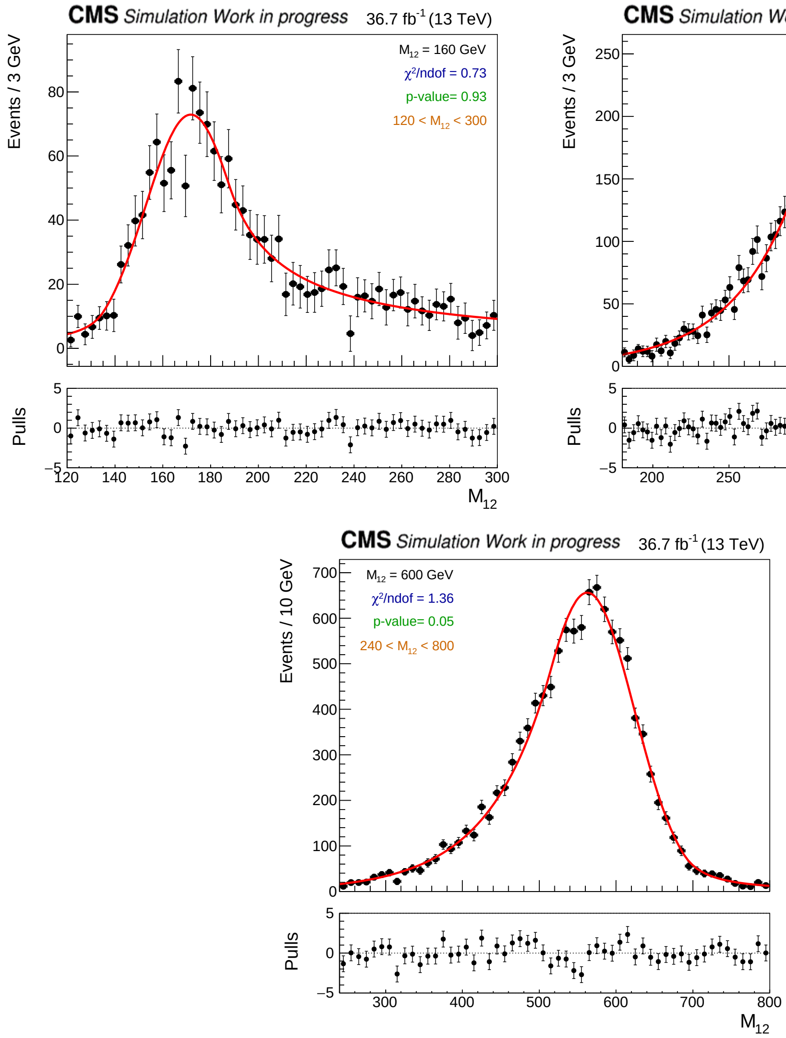


Figure 5.22: Parameterization of signal on M_{12} distribution using the DSCB function for the 2017 SL category: the top left plot represents a signal mass of 160 GeV from FR1, the top right plot illustrates the 350 GeV parametrization from FR2, and the bottom plot shows the 600 GeV signal from FR3.

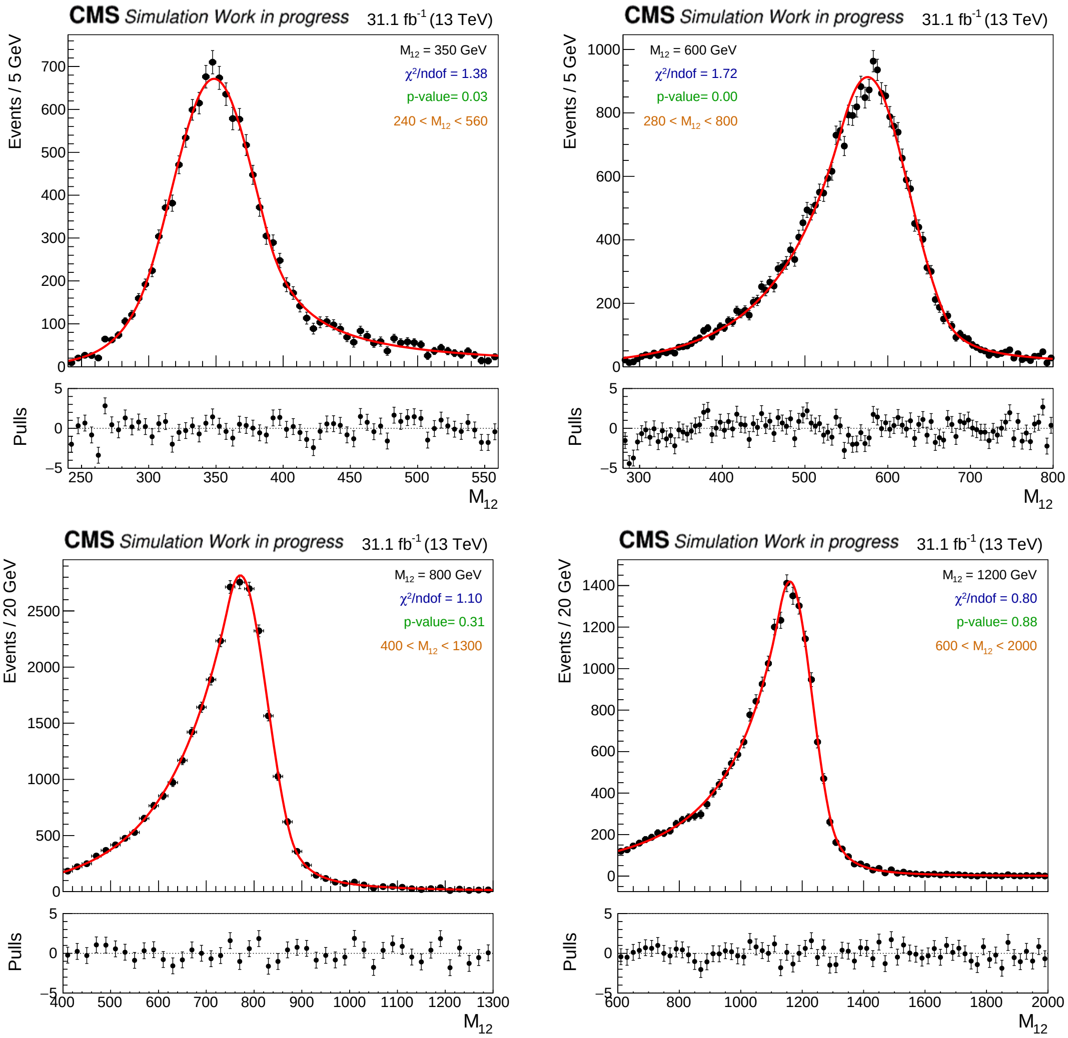


Figure 5.23: Parameterization of signal on M_{12} distribution using the DSCB function for the 2017 FH category: the top left plot represents a signal mass of 350 GeV from FR1, the top right plot illustrates the 500 GeV parametrization from FR2, the bottom left plot depicts the 600 GeV signal from FR3 while the bottom right plot shows the 1200 GeV mass-point in FR4

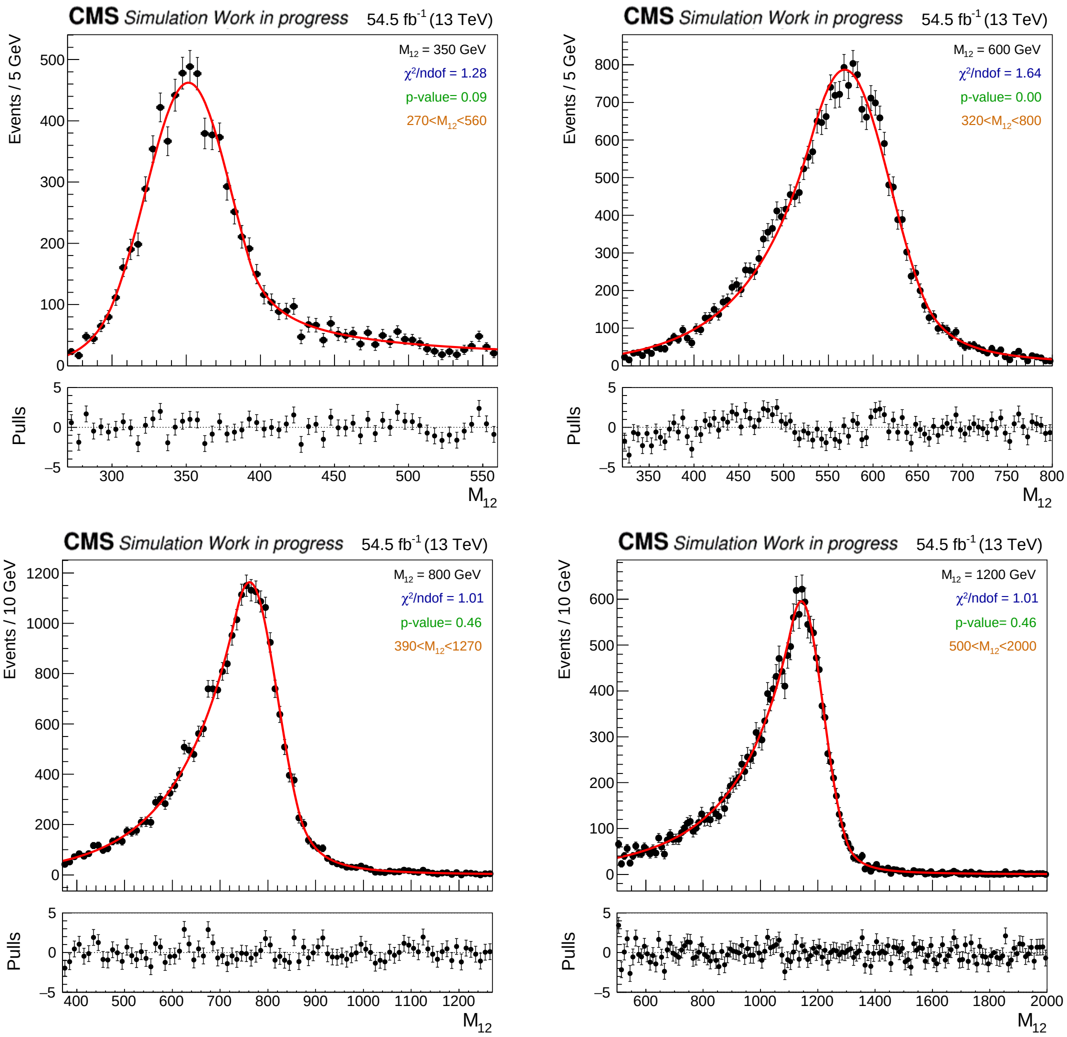


Figure 5.24: Parameterization of signal on M_{12} distribution using the DSCB function for the 2018 FH analysis: the top left plot represents a signal mass of 350 GeV from FR1, the top right plot illustrates the 500 GeV parametrization from FR2, the bottom left plot depicts the 600 GeV signal from FR3 while the bottom right plot shows the 1200 GeV mass-point in FR4

The resulting shift in the shape parameters is used for the signal extraction by introducing a corresponding nuisance parameter to account for it. This parameter is directly built into the signal model and treated as a Gaussian-constrained nuisance parameter (as detailed in section 5.10.1). The normalisation effects of the JES (JER) variations are found to be in the order of 5% (0.1%) and are also built into the signal model to be treated with the corresponding nuisance.

Normalisation systematic uncertainties

Other uncertainties only alter the signal normalisation. These are the uncertainties of the kinematic trigger scale factors, online and offline b tag scale factors, and the uncertainties related to the L1 prefiring correction, In addition, the uncertainties on the muon ID scale factor must be considered in the 2017 SL category. Similarly, the HEM15/16 [208] correction (detailed in section 5.6.7) affects the 2018 FH analysis only. As per internal CMS recommendations, this correction is treated as a 1σ systematic uncertainty, symmetrized for simplicity.

The effect of these uncertainties on the M_{12} distribution of the 2018 FH analysis is illustrated in Figs. 5.27, 5.28, 5.29, and 5.30, while the corresponding effects in the 2017 analyses can be found in Appendix D.

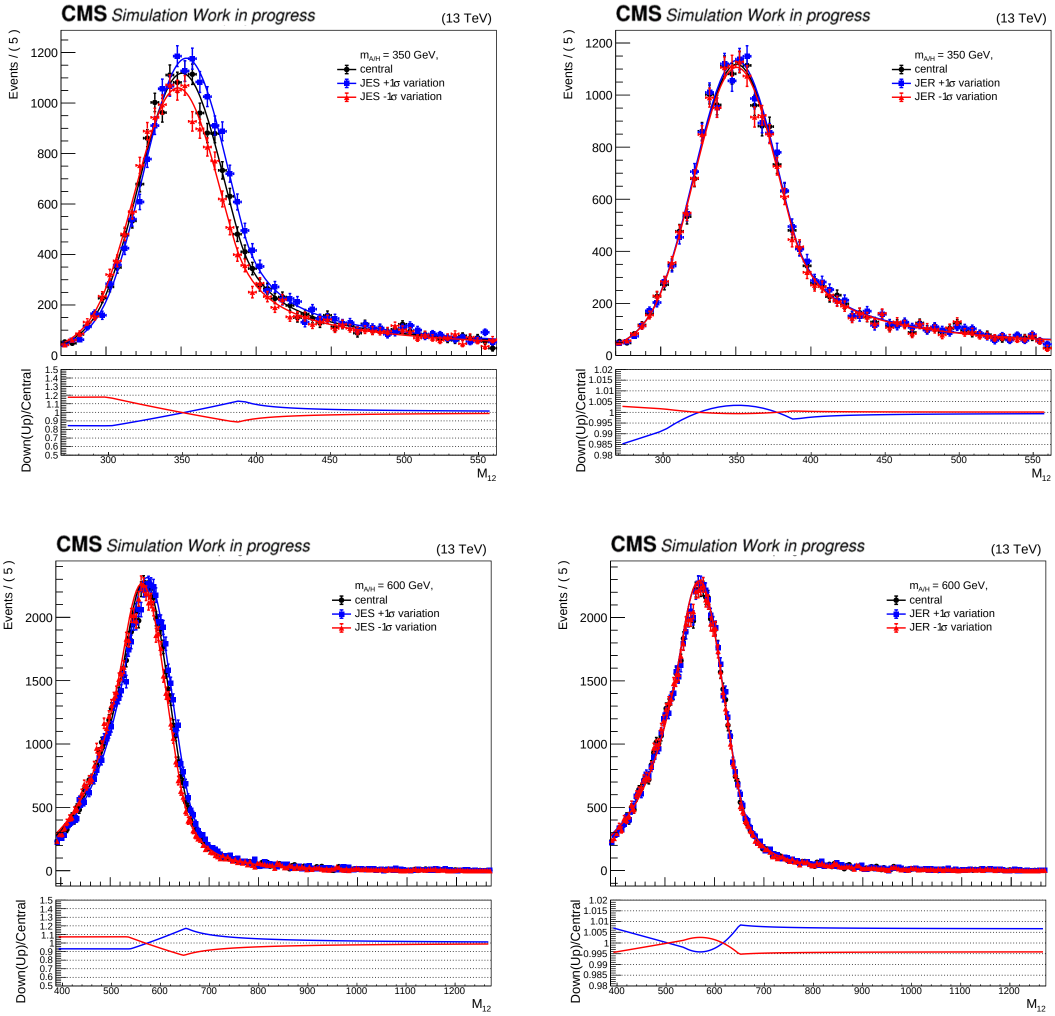


Figure 5.25: Shape-altering systematic uncertainties from JES (left) and JER (right) uncertainties in the 2018 FH channel. The plots depict representative signal masses from each fit range: 350 GeV (top) in FR1 and 600 GeV (bottom) in FR2. The parametrizations of up and down variations, as well as for the nominal, are shown. The bottom panel in each plot shows the ratio of the up/down parametrization with respect to the central distribution.

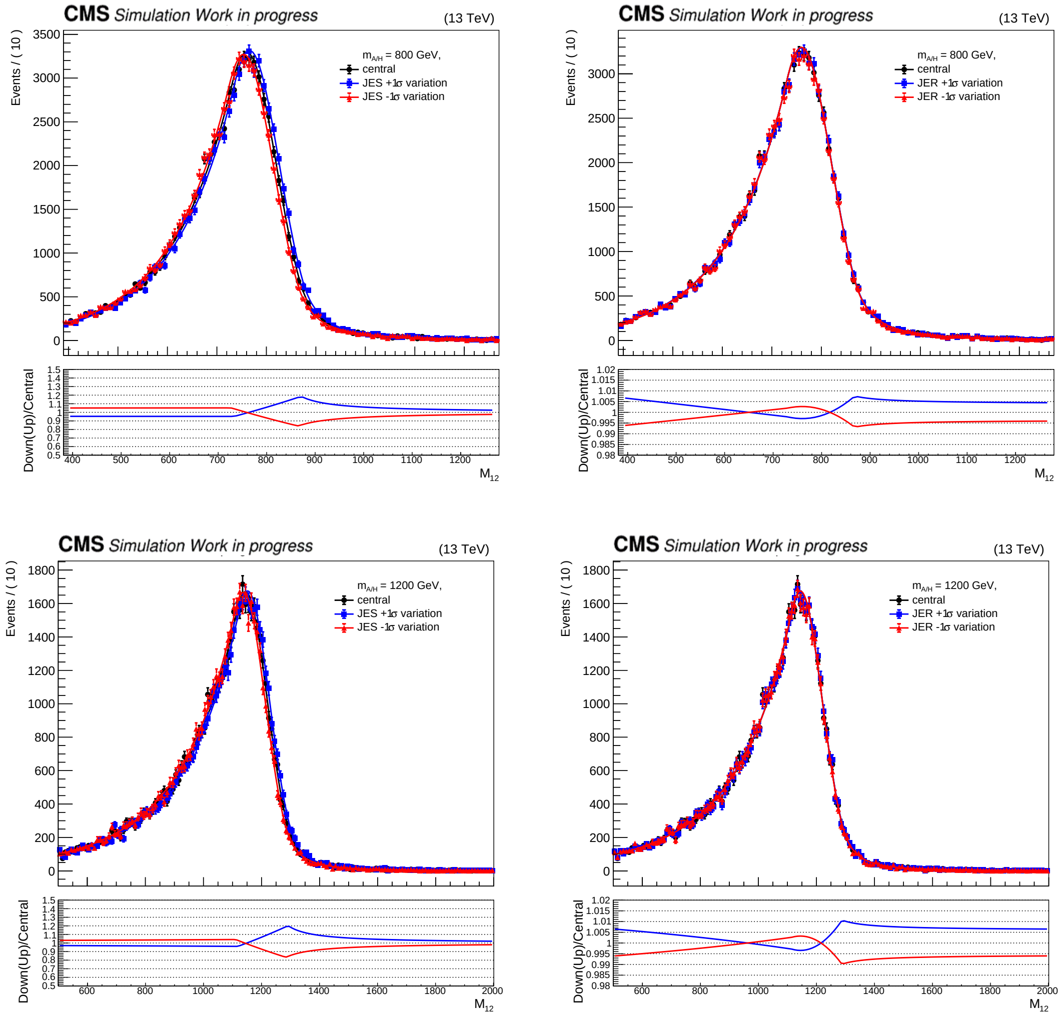


Figure 5.26: Shape-altering systematic uncertainties from JES (left) and JER (right) uncertainties in the 2018 FH channel. The plots depict representative signal masses from each fit range: 800 GeV (top) in FR3 and 1200 GeV (bottom) in FR4. The parametrizations of up and down variations, as well as for the nominal, are shown. The bottom panel in each plot shows the ratio of the up/down parametrization with respect to the central distribution.

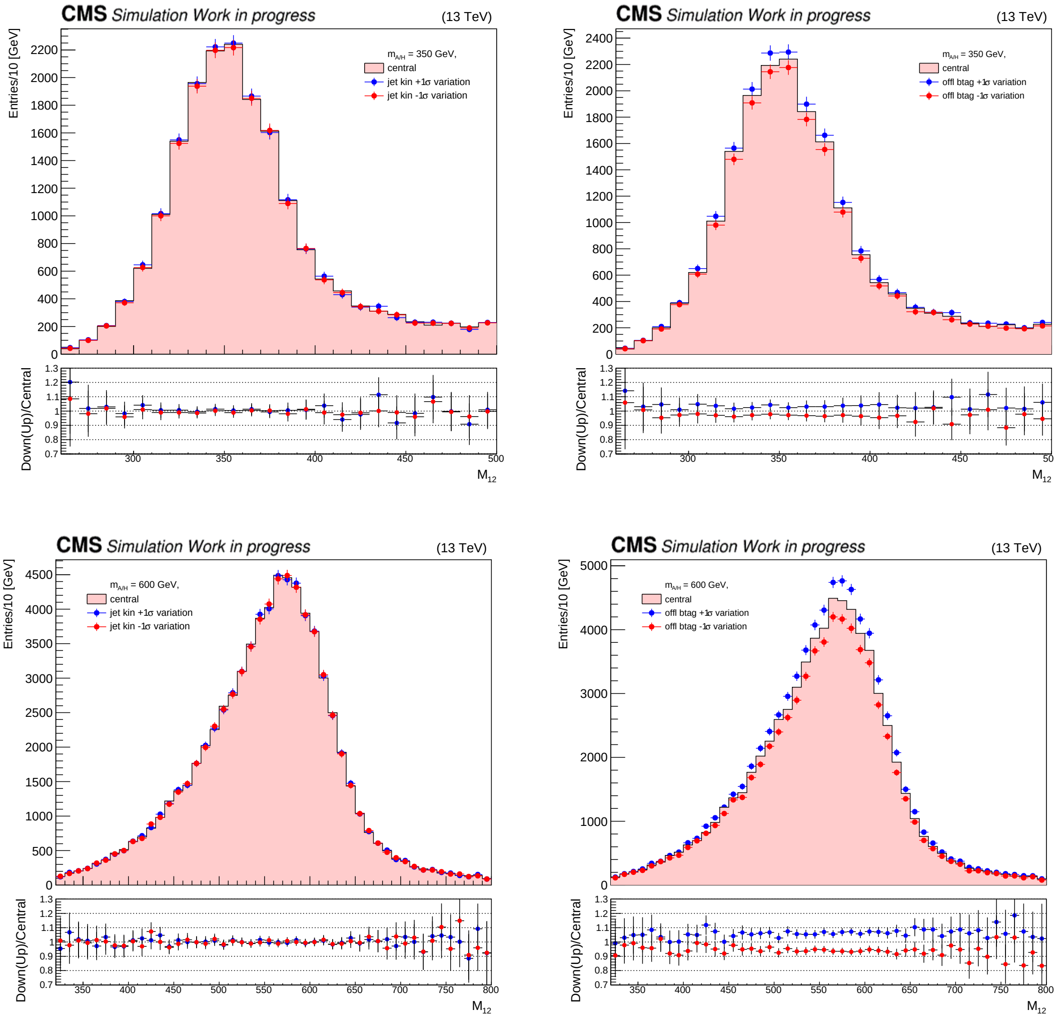


Figure 5.27: Shape-invariant systematic uncertainties from trigger kinematic scale factor (left) and offline b-tagging (right) uncertainties in 2018 FH channel, for the representative signal masses from each fit range : 350 GeV (top) in FR1, 600 GeV (bottom) in FR2, respectively.

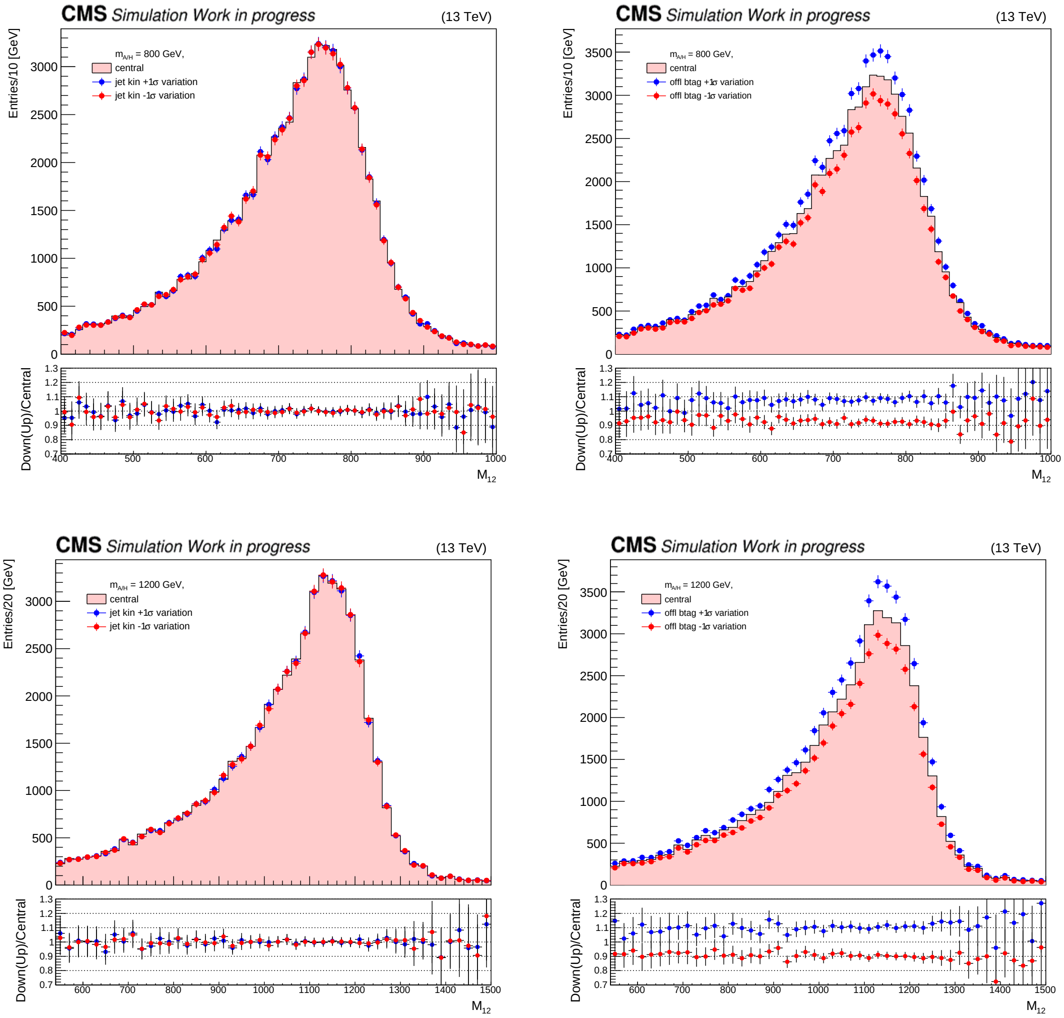


Figure 5.28: Shape-invariant systematic uncertainties from trigger kinematic scale factor (left) and offline b-tagging (right) uncertainties in 2018 FH channel, for the representative signal masses from each fit range : 800 GeV (top) in FR3, and FR3 1200 GeV (bottom) in FR4, respectively.

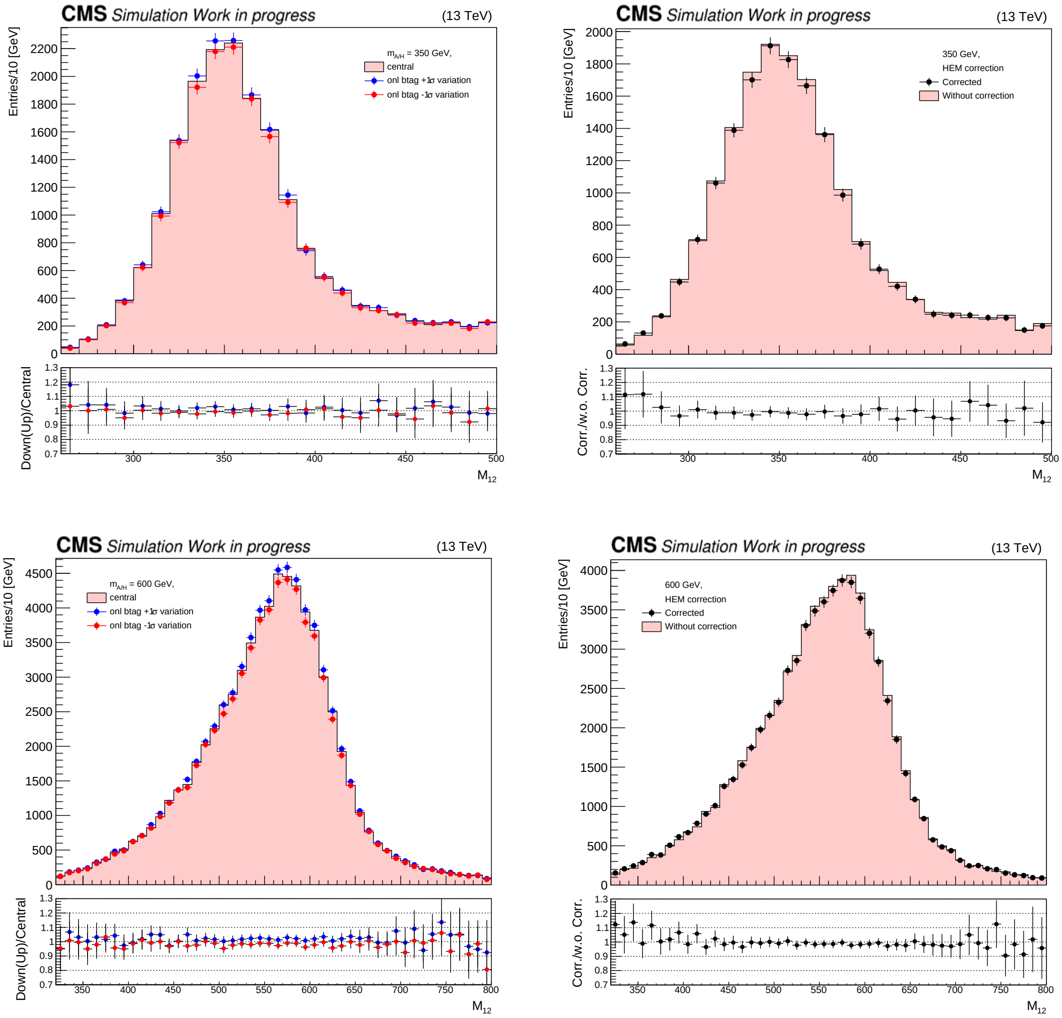


Figure 5.29: Shape-invariant systematic uncertainties from online b-tagging (left) and HEM correction (right) uncertainties in 2018 FH channel, for the representative signal masses from each fit range : 350 GeV (top) in FR1, 600 GeV (bottom) in FR2, respectively.

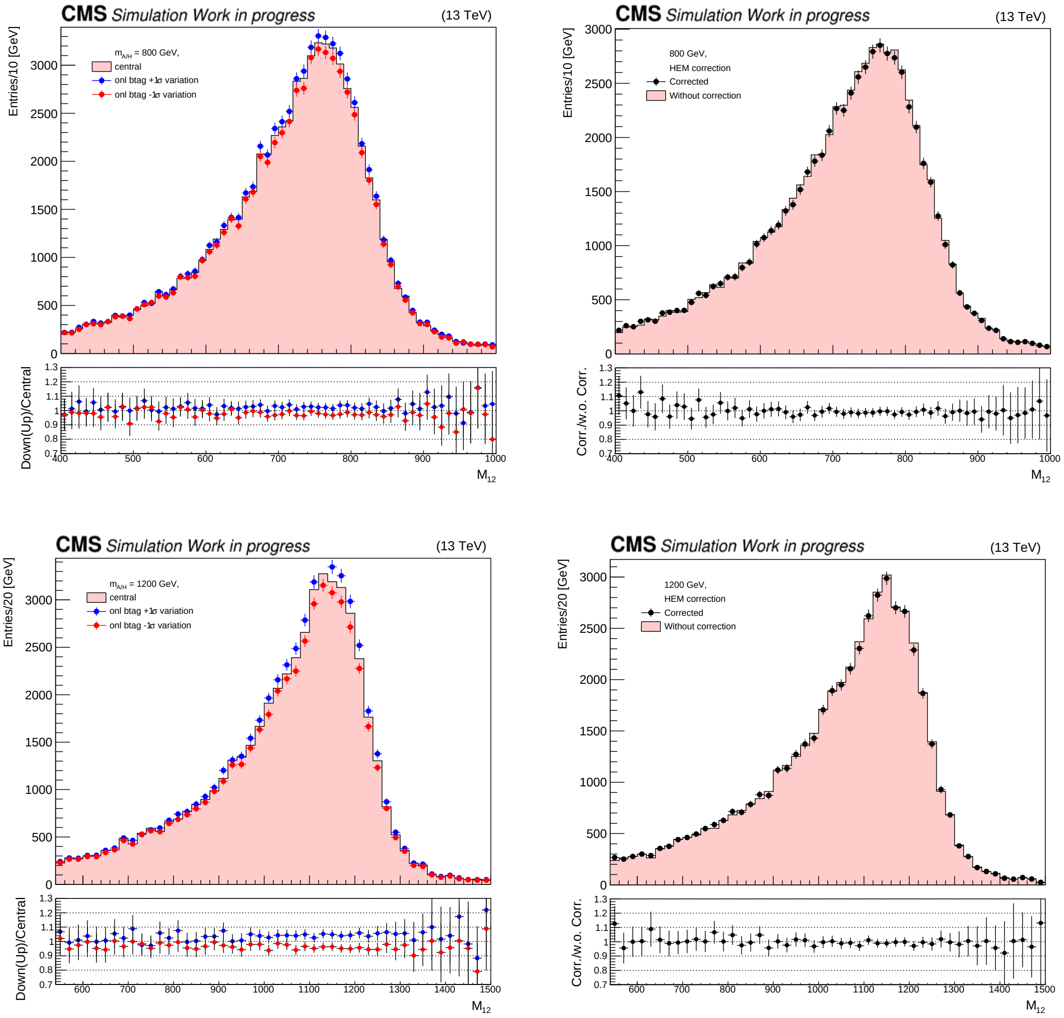


Figure 5.30: Shape-invariant systematic uncertainties from online b-tagging (left) and HEM corr uncertainties in 2018 FH channel, for the representative signal masses from each fit range : 800 GeV (top) in FR3, and FR3 1200 GeV (bottom) in FR4, respectively.

5.8 Background modelling

As mentioned in section 5.1, the dominant background affecting this analysis consists of standard model events with heavy-flavour multi-jet final states, predominantly from QCD processes resulting in three b-jets, either real or mistagged. A minor contribution coming from the production of multiple b-quarks in $t\bar{t}$ events is also expected.

An accurate description of backgrounds is crucial for distinguishing between genuine signals and background fluctuations. However, the QCD processes contributing to the background in the analysis are known to be difficult to simulate precisely, and simulations are typically statistically limited after the triple b-tag selection. Therefore, a data-driven approach is used to model the background in the SR.

The data-driven background modelling is based on two components. The first is the parametrization of a signal-depleted CR, introduced in section 5.6. The CR is then multiplied by the second component, a transfer factor (TF), which is introduced to account for differences between SR and CR and is from simulated samples. From simulation-based studies, the TF is expected to be very simple in shape, resembling roughly a linear function that increases slowly with the di-jet mass. Although simulated QCD events inspire the shape analysis of the TF, all the parameters are fitted in the data during the statistical inference procedure where a simultaneous SR and CR fit is performed. All the background sources are treated collectively, thus other contributions beyond QCD multi-jet production are not dismissed.

This section provides a description of the CR and their TF, their parametrization, and the study of the uncertainty related to the choice of the TF functional form.

5.8.1 The reverse b-tag Control Region

The Control Region, introduced in section 5.6, is defined by a selection similar to that of the Signal Region, but reverting the b-tag requirement of the third leading jet. Previous studies have found plenty of similarities between these two regions [35, 39, 41, 191]. Both are characterized by a complex shape in the M_{12} distribution, with a steep increase in the low mass region caused by the kinematic thresholds, followed by a falling trend with respect to the invariant mass.

Given the complexity of the CR, this distribution is rather difficult to parametrize in the full mass range explored. Therefore, its parameterization is performed in several ranges of the di-jet mass distribution, or fit ranges (FRs), chosen as regimes of different shape characteristics of the background mass distribution. The FRs are analysis category and year-dependent, as indicated in Table 5.14. They have been carefully chosen, optimizing the control over the background description and enhancing the sensitivity of each analysis. Moreover, the overlaps are dimensioned such that the signal shape for each mass point within a FR is well-contained in at least one of the mass regions. Table 5.14 also shows corresponding signal mass points in each fit range. The mass points at the borders are assigned to both fit ranges, which allows a comparative signal extraction in both adjacent fit ranges. As the nominal result for such overlapped mass points, the result with the better-expected limit will be chosen.

Channel	Fit Range	mass range [GeV]	signal ($m_{\mathrm{A/H}}$)
2017 SL	1	120–300	125, 130, 140, 160, 180, 200, 250
	2	180–460	250, 300, 350
	3	240–800	350, 400, 450, 500, 600, 700
2017 FH	1	240–560	300, 350, 400, 450
	2	280–800	450, 500, 600, 700
	3	400–1300	700, 800, 900, 1000
	4	600–2000	1000, 1400, 1600, 1800
2018 FH	1	270–560	300, 350, 400
	2	320–800	400, 450, 500, 600
	3	390–1270	600, 700, 800, 900, 1000
	4	500–2000	1000, 1200, 1400, 1600, 1800

Table 5.14: Definition of fit ranges for 2017 SL, 2017 FH, and 2018 FH categories and corresponding signal mass points considered in the signal extraction.

Functions used for parameterization

The functions chosen for the parametrization of the CR in fit ranges are a family of Novosibirsk functions, as introduced in [41]. In this analysis, we utilise again a so-called "extended" Novosibirsk function, which has a mathematical form as follows:

$$F^{(n)}_{\text{Extended Novosibirsk}}(x) = N \cdot \exp\left\{ -\frac{1}{2\sigma_0^2} \ln^2 \left[1 - \frac{\eta}{\sigma_E} \cdot \left(\sum_{i=1}^{n} p_i \cdot (x - x_p)^i \right) \right] - \frac{\sigma_0^2}{2} \right\}, \quad (5.7)$$

where

$$\begin{aligned} \sigma_0 &= (2/\epsilon) \sinh^{-1}(\eta\epsilon/2) \\ \epsilon &= 2\sqrt{\ln 4} = 2.36 \\ p_1 &\equiv 1 \end{aligned}$$

In the case $n = 1$ this definition simplifies to the "classic" Novosibirsk function, originally introduced to parameterise a Compton spectrum affected by detector resolution effects [209]:

$$F^{(1)}_{\text{Novosibirsk}}(x) = N \cdot \exp\left\{ -\frac{1}{2\sigma_0^2} \ln^2 \left[1 - \frac{\eta}{\sigma_E} \cdot (x - x_p) \right] - \frac{\sigma_0^2}{2} \right\}, \quad (5.8)$$

where N is a normalisation parameter, x_p is the peak position, σ_0 is the width of the peak, and the η parameter governs the tail and hence the asymmetry of the function. The sign convention of η is such that negative values produce a tail towards increasing abscissa values. In the limit where $\eta \to 0$, a symmetric Gaussian function is obtained.

The background shape in FR1 and the lower edge of FR2 is characterized by a rising edge of the background peak reflecting the effect of kinematic thresholds and momentum-dependent b-tagging efficiency, which occurs both online and offline, resulting in a marked turn-on behaviour. The Gaussian error function is further introduced to help address the turn-on behaviour by multiplying it with the extended Novosibirsk function. It is sometimes called a turn-on (T.O.) function in this analysis and defined as:

$$f(\mathrm{M}_{12}) = 0.5 \cdot (\mathrm{Erf}(p_0(\mathrm{M}_{12} - p_1)) + 1)\,, \tag{5.9}$$

where

$$\mathrm{Erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt, \tag{5.10}$$

and the parameters p_0 and p_1 describe the slope of the turn-on and the turn-on point, respectively. The turn-on function is only required for some fit ranges in which the extended Novosibirsk cannot perfectly model the background shape alone. The parameters of the turn-on function are finally fixed to the values obtained from the first fit of invariant mass (M_{12}) in the CR as shown in Table 5.15 and combined with the extended Novosibirsk function for the signal extraction. The transfer factor, described in the next subsection, will accommodate additional shape differences in the mass spectra between SR and CR.

The results from background parameterization using a combination of fixed turn-on function and classic Novosibirsk (when $n = 1$) function are shown on the top left of Fig. 5.31 and Fig. 5.32.

Fit Range	p_0	p_1
2017 SL - FR1	0.023	125.278
2017 FH - FR2	0.014	280.348

Table 5.15: The parameters of the Gaussian error function resulted from the fit of M_{12} in the bbnb reversed b tag control region.

In the 2017 SL category, the CR was parametrized with a T.O. function times an extended Novosibirsk, while FR2 and FR3 is based on the classic Novosibirsk. The resulting fits are shown in Fig. 5.31.

In the 2017 FH category, FR1 can be described by the extended Novosibirsk function with $n = 2$, while for FR2 a T.O. and an extended Novosibirsk are used. FR3 and FR4 are parametrized with the classic Novosibirsk function. Fig. 5.32 shows the CR model in 2017 FH category for all fit ranges. In the case of the 2018 FH category all fit ranges, namely FR1-4, can be modelled by the extended Novosibirsk function with $n = 2$, as shown in Fig. 5.33.

As can be seen, a very good description is obtained for each fit range using the combination of the fixed Gaussian error and extended Novosibirsk functions with at most 4 shape parameters. A summary of the function choices with the corresponding number of fit parameters, the mass range of each FR as well as the goodness of fit are shown in Table 5.16.

Channel	Fit Range	p.d.f.	parameters	fit probability	χ^2/ndf
2017 SL	1	fixed turn-on x Novosibirsk	3	41 %	$34.2/33 = 1.0$
	2	Novosibirsk	3	15 %	$63.9/53 = 1.2$
	3	Novosibirsk	3	29 %	$116.6/109 = 1.1$
2017 FH	1	Extended Novosibirsk	4	30 %	$31.4/28 = 1.1$
	2	fixed Turn-on x Novosibirsk	3	39 %	$51.3/49 = 1.0$
	3	Novosibirsk	3	22 %	$48.9/42 = 1.2$
	4	Novosibirsk	3	79 %	$57.5/67 = 0.9$
2018 FH	1	Extended Novosibirsk	4	45 %	$25.3/25 = 1.0$
	2	Extended Novosibirsk	4	15%	$53.8/44 = 1.2$
	3	Extended Novosibirsk	4	58 %	$37.6/40 = 0.9$
	4	Extended Novosibirsk	4	24 %	$63.1/56 = 1.1$

Table 5.16: Background parametrizations (probability density functions) for the three and four fit ranges in 2017 SL, 2017 FH and 2018 FH channels, respectively, together with the results from a goodness of a fit test.

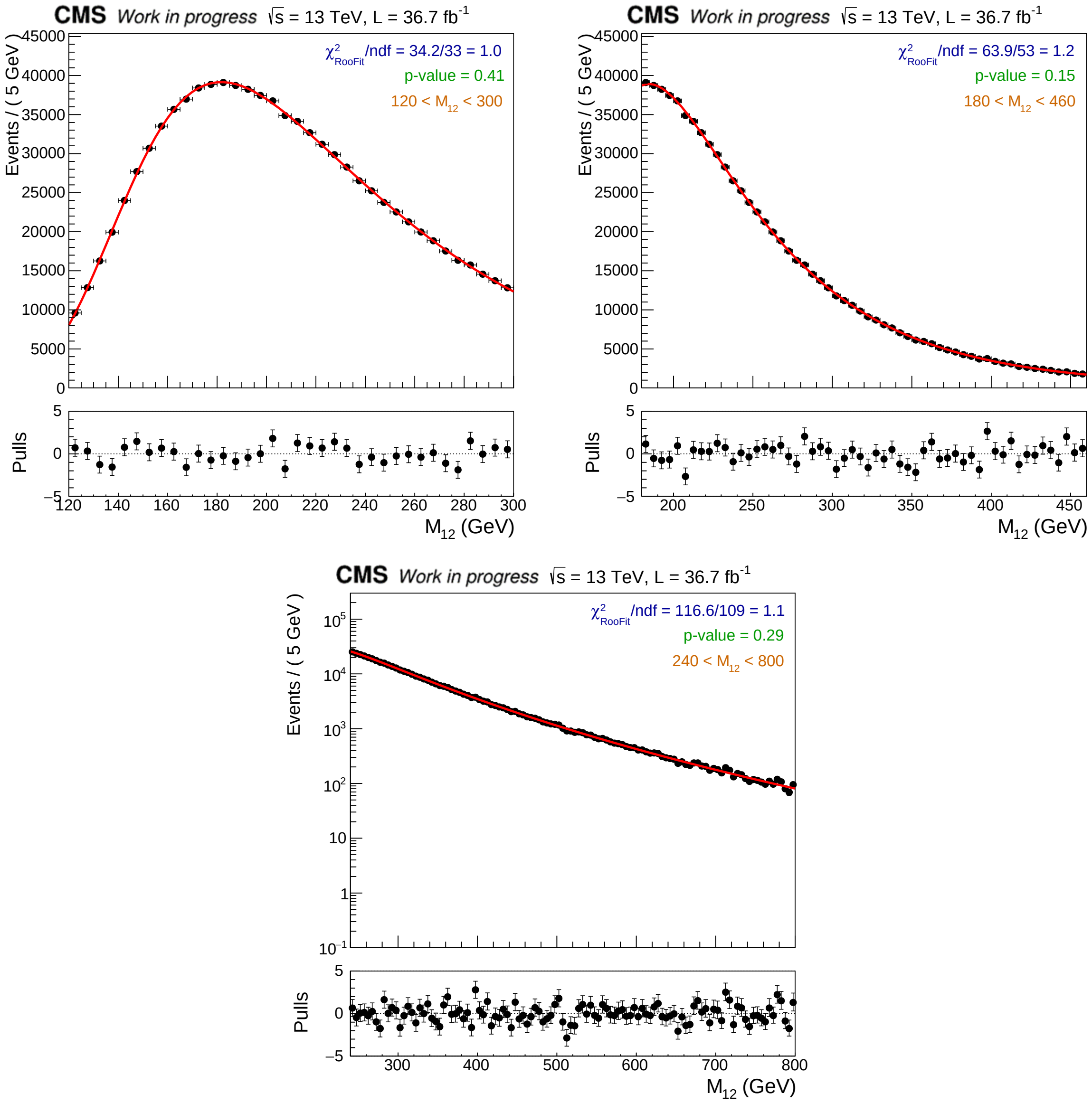


Figure 5.31: Invariant mass distributions of the three fit ranges in the reversed b tag control region (bbnb) for the 2017 SL channel, together with the fits of the functions as described in the text.

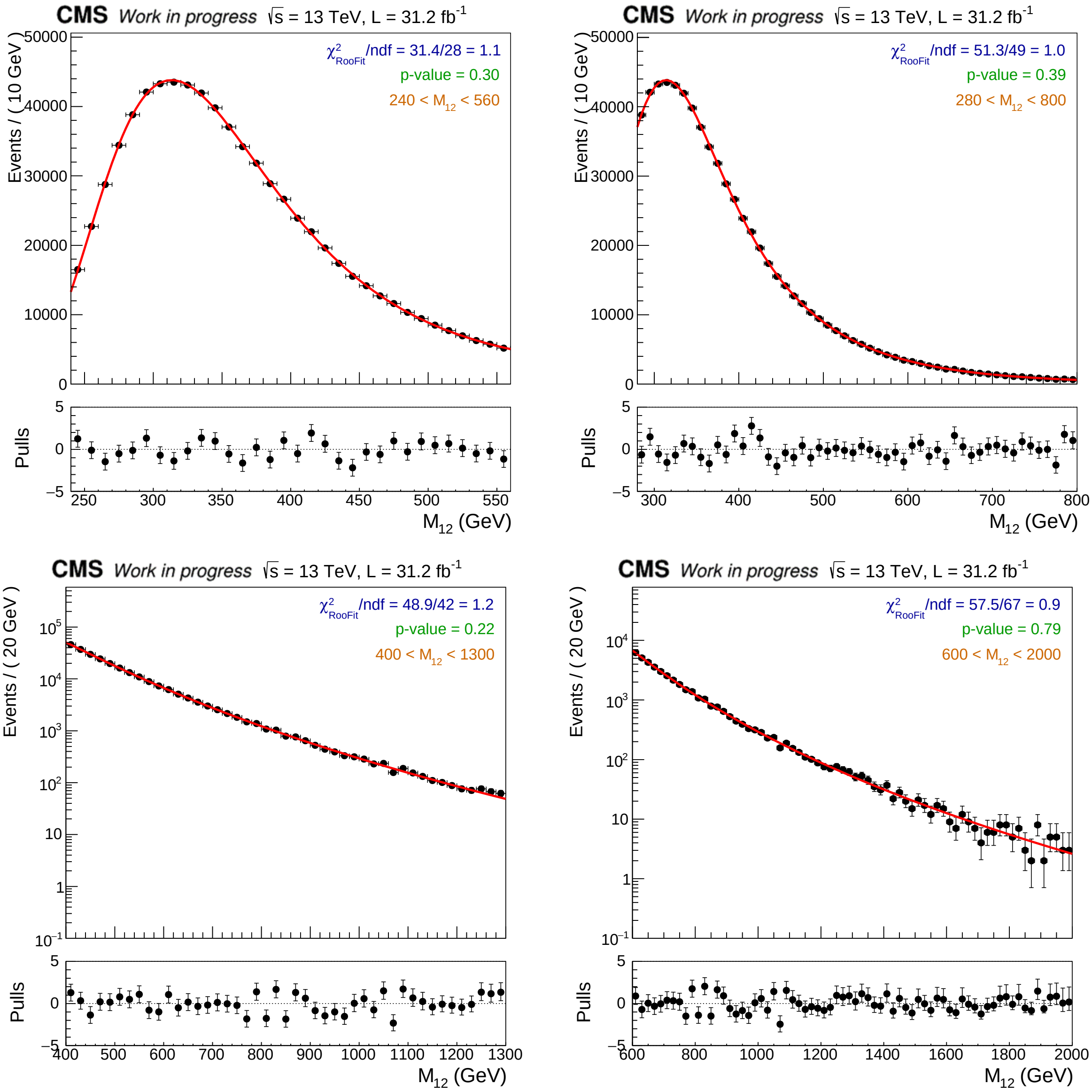


Figure 5.32: Invariant mass distributions of the three fit ranges in the reversed b tag control region (bbnb) for the 2017 FH channel, together with the fits of the functions as described in the text.

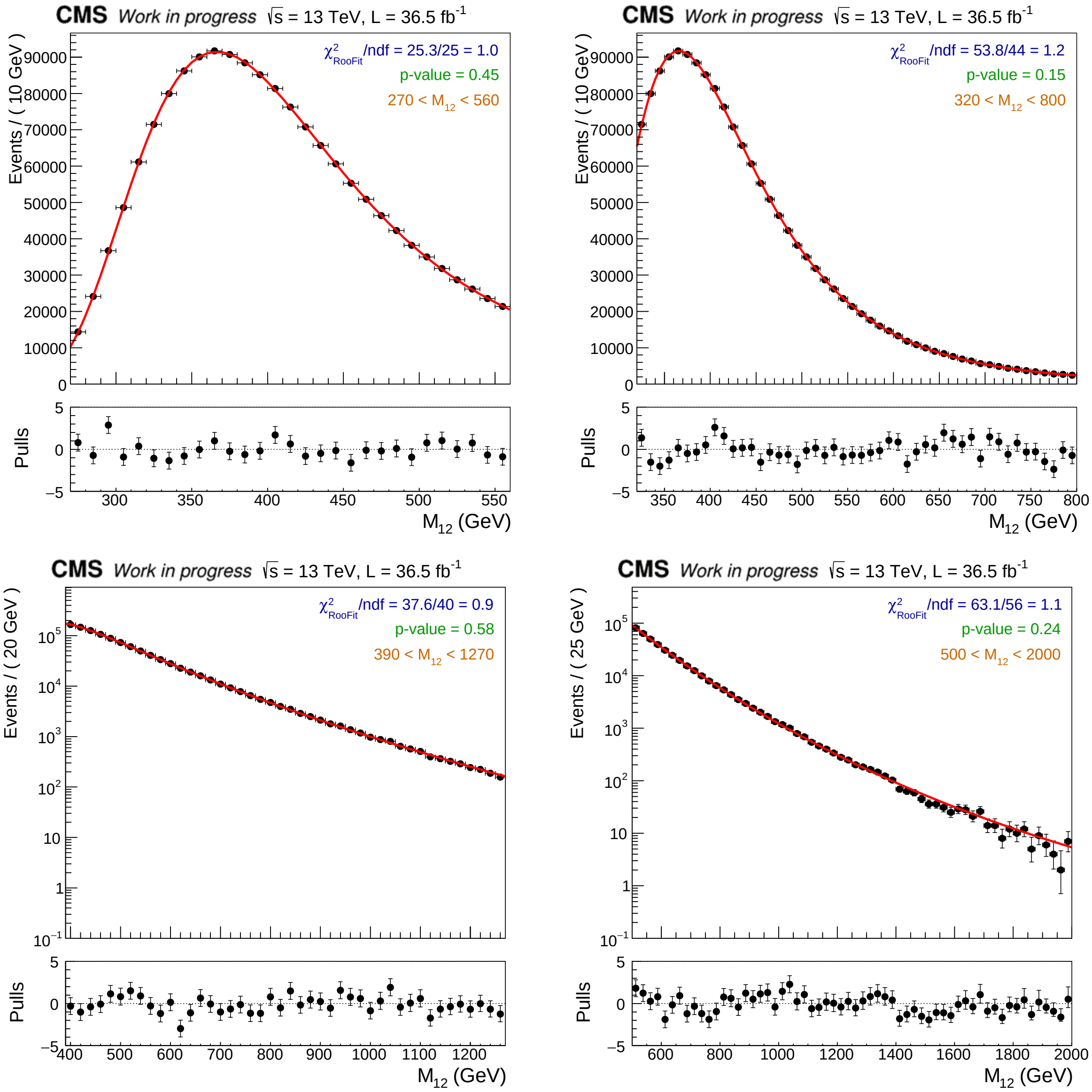


Figure 5.33: Invariant mass distributions of the four fit ranges in the reversed b tag control region (bbnb) for the 2018 FH channel, together with the fits of the functions as described in the text.

5.8.2 The Transfer Factor

The second component of the background modelling, the Transfer Factor, is incorporated to accommodate the difference in background shape that could appear between CR and SR. Its shape has been studied in b-enriched QCD multijet simulation, both at parton shower and matrix element level. These samples are notoriously limited by statistics. To overcome this, a b-tag weighting technique is employed in the estimation of the transfer factor shape. This technique substitutes hard cuts according to whether or not a jet satisfies a certain b-tag working point by assigning a weight factor to the event corresponding to the b-tag efficiency with respect to the actual jet flavour. A significant increase in the statistics is obtained, correspondingly reducing the statistical uncertainty of the TF shape. The b-tag efficiencies by flavour employed in all analyses and closure tests comparing relevant distributions with the cut-based and the b-tag weights approach can be found in Appendix E.

The resulting transfer factor in the full mass range with an arbitrary binning is shown in Fig. 5.34- 5.36 for 2017 SL, 2017 FH, and 2018 FH category, respectively. As can be deducted from the distributions, their shapes can be described with relatively simple functions.

The TF is parametrized in the fit ranges introduced in the previous subsection in each year and category. First and second-degree Chebychev polynomials are used for the parametrization. The polynomial degrees employed are motivated by an F-test, as will be described in section 5.8.3. The modelling of the TF in all fit-ranges are shown in Fig. 5.37 for 2017 SL, Fig. 5.38 for 2017 FH, and Fig. 5.39 for 2018 FH channel. The functional form is chosen during the final signal-plus-background fit by means of a discrete profiling method [210], which is also employed to address the systematic uncertainty that this decision could introduce, also described 5.8.3.

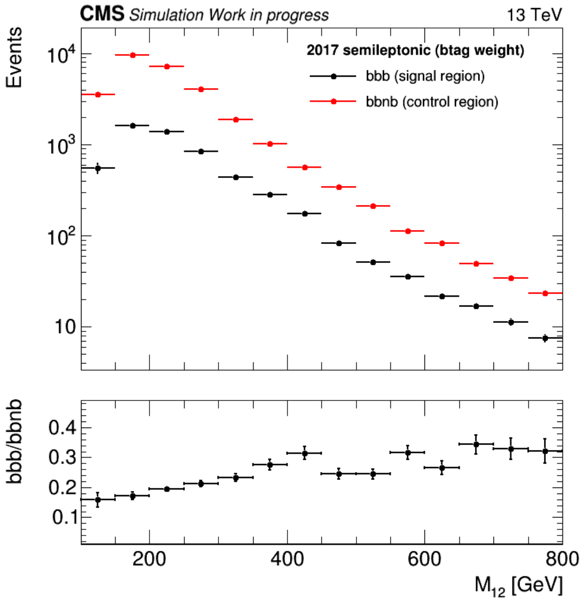


Figure 5.34: Ratios of the invariant mass distributions of the SR ("bbb") and CR("bbnb") as determined in QCD multijet simulation in the full mass-range of the 2017 SL analysis.

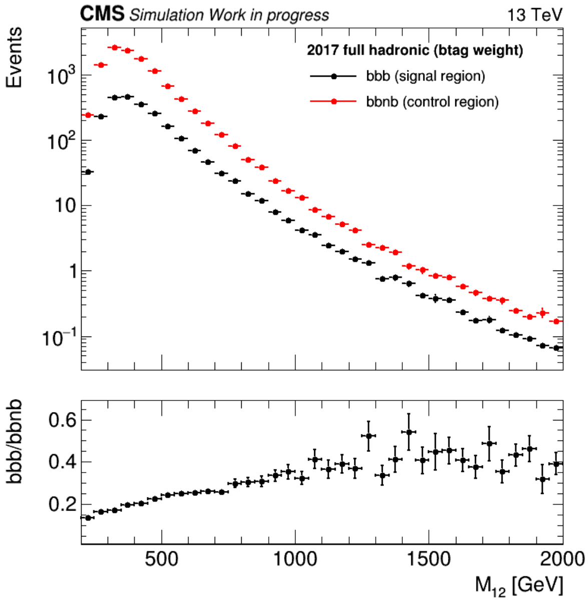


Figure 5.35: Ratios of the invariant mass distributions of the SR ("bbb") and CR("bbnb") as determined in QCD multijet simulation in the full mass-range of the 2017 FH analysis.

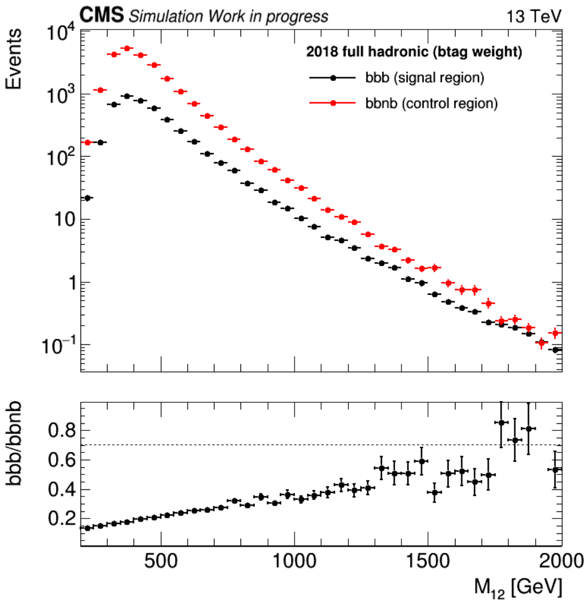


Figure 5.36: Ratios of the invariant mass distributions of the SR ("bbb") and CR("bbnb") as determined in QCD multijet simulation in the full mass-range of the 2018 FH analysis.

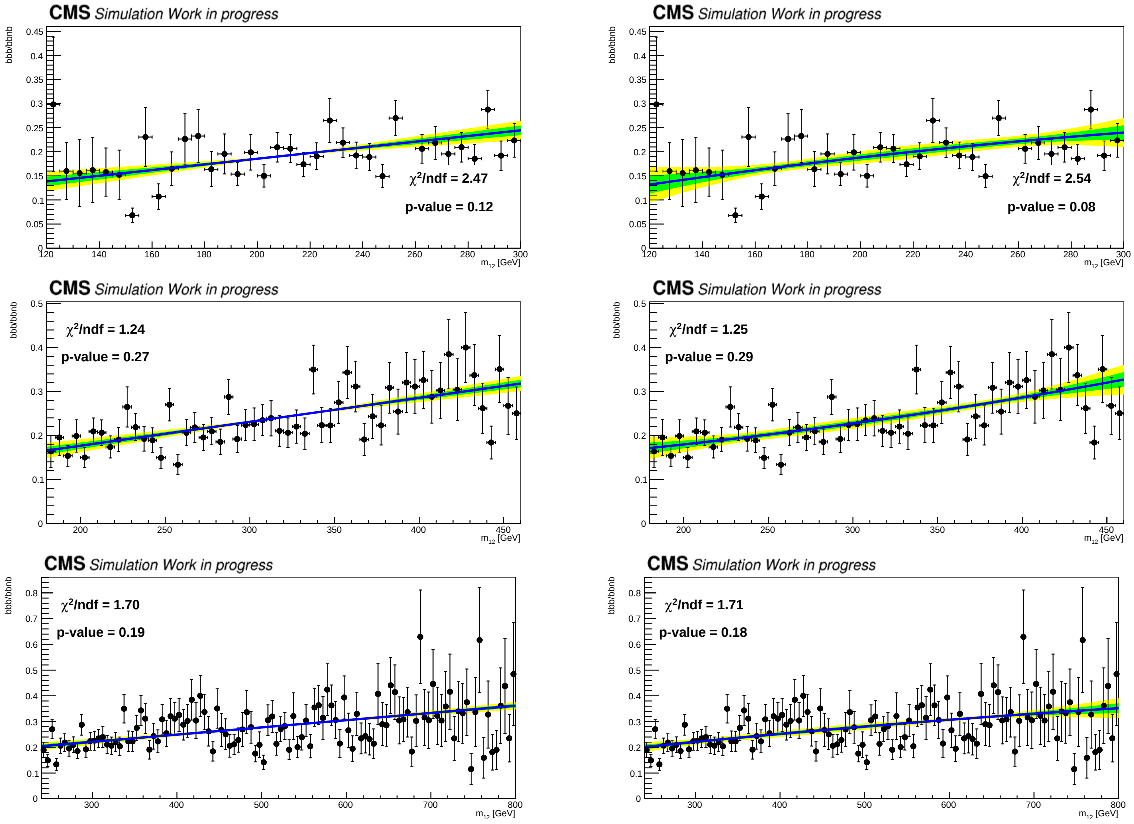


Figure 5.37: Ratios of the invariant mass distributions of the SR (“bbb”) and CR(“bbnb”) as determined in QCD multijet simulation, for the fit ranges FR1, FR2, and FR3 for 2017 SL analysis. The functions selected for the parametrization are the first (left) and second (right) degrees of Chebychev polynomials.

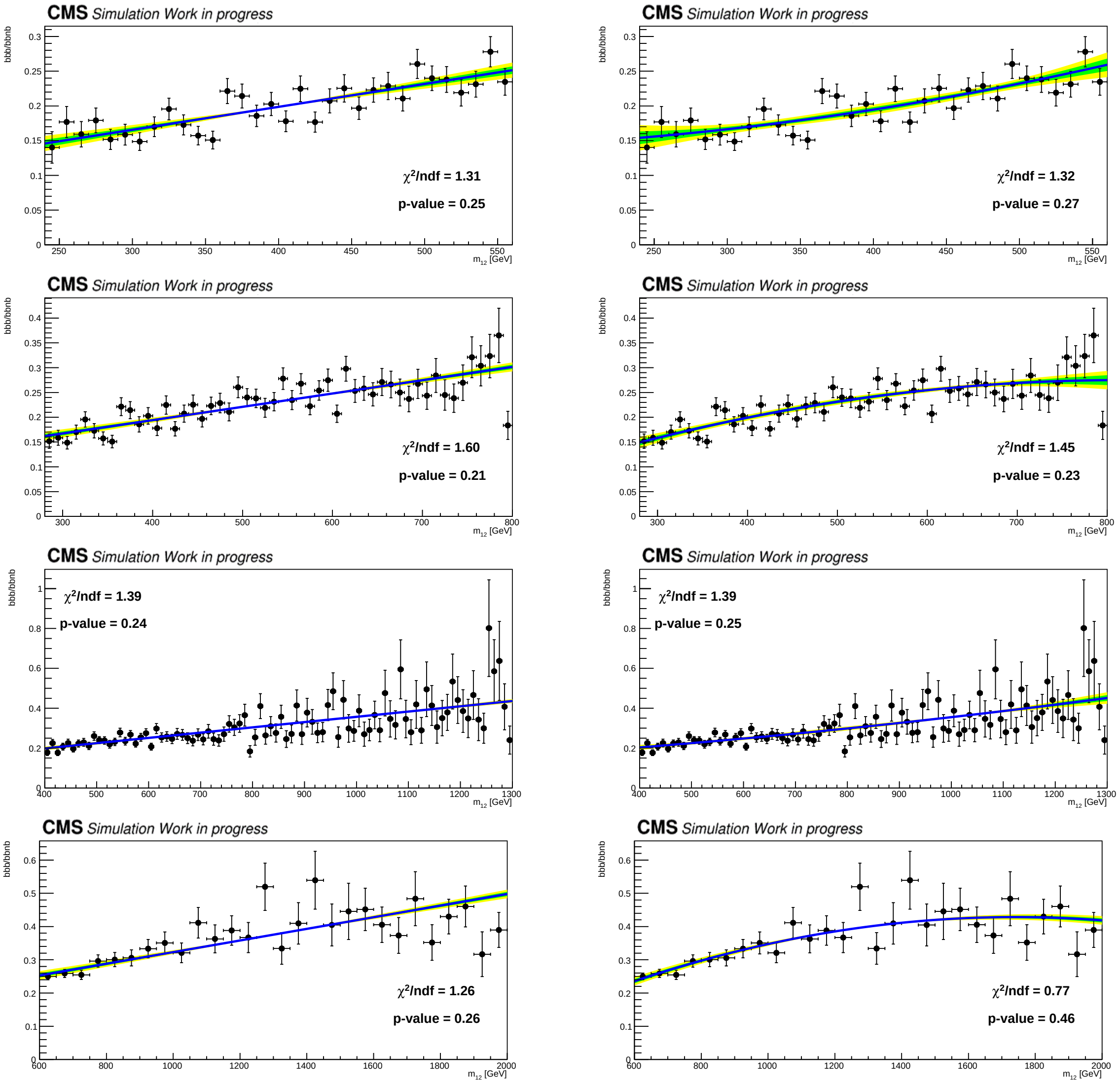


Figure 5.38: Ratios of the invariant mass distributions of the SR ("bbb") and CR("bbnb") as determined in QCD multijet simulation, for the fit ranges FR1, FR2 and FR3 for 2017 FH analysis. The functions selected for the parametrization are the first (left) and second (right) degrees of Chebychev polynomials.

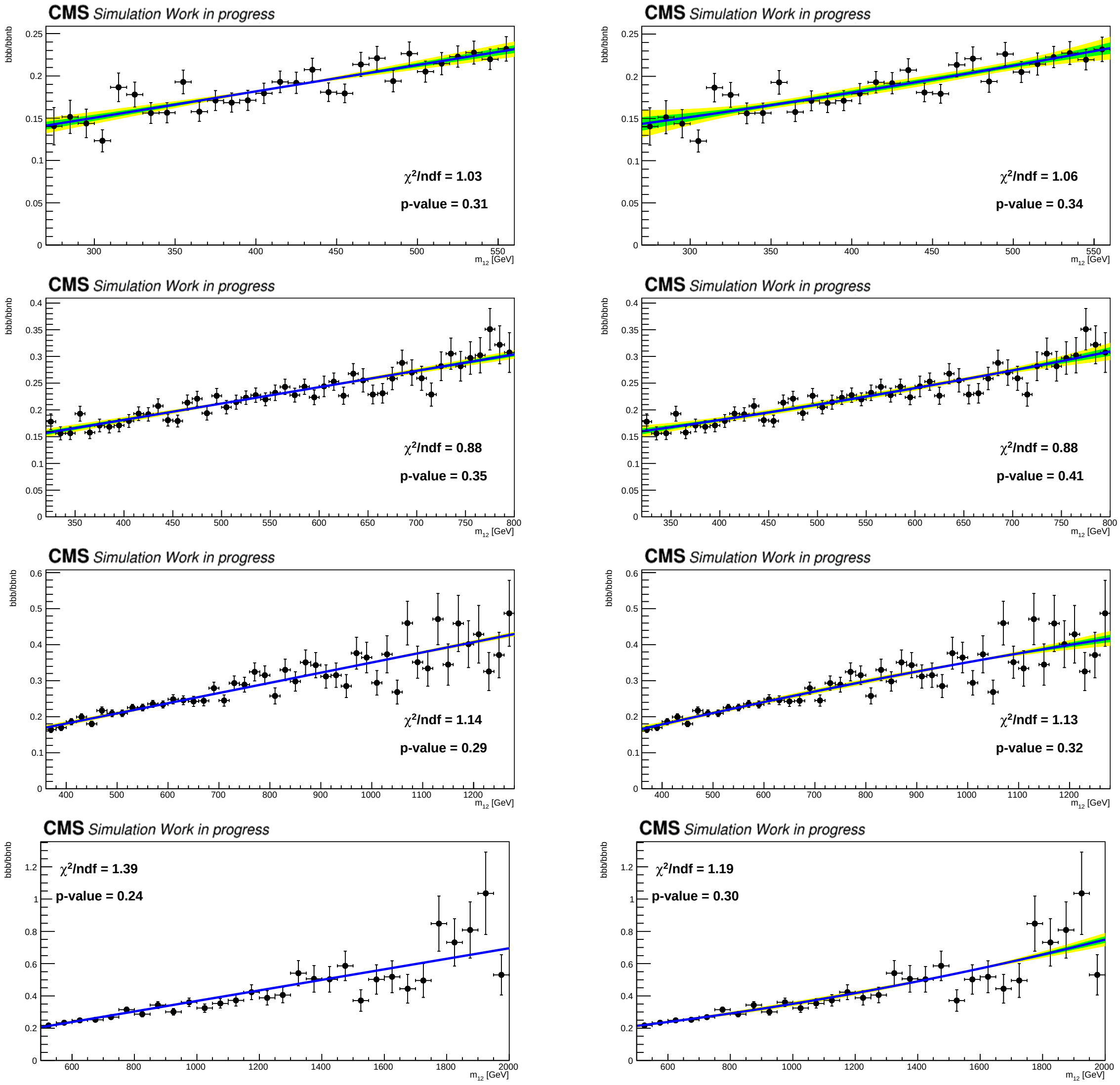


Figure 5.39: Ratios of the invariant mass distributions of the SR (“bbb”) and CR(“bbnb”) as determined in QCD multijet simulation, for the fit ranges FR1, FR2, FR3 and 4 for 2018 FH analysis. The functions selected for the parametrization are the first (left) and second (right) degrees of Chebychev polynomials.

5.8.3 Bias study

Searches for a signal on top of a large background are sensitive to variations in the background shape which may significantly bias the fitted signal. For this reason, the choice of the functional form of the background must be carefully considered, and any possible bias must be accounted for.

As explained at the beginning of section 5.8, in the analyses presented in this thesis the background in the signal region is constructed by combining the shape of the CR, which holds the majority of complexities, with a simple TF. The CR shape parameters are largely constrained by the simultaneous fit in CR and SR. On the other hand, as the TF only changes slowly with M_{12}, the freedom in altering its shape is rather small, resulting in a much-reduced bias effect [40].

To evaluate the systematic uncertainty related to the TF choice, alternative functions are explored using a discrete profiling method [210]. This involves treating the TF selection as a discrete parameter that is profiled to get the best-fit value as part of the minimization of the negative logarithm of the likelihood ratio (NLL) function. The specifics of the likelihood function formulation will be presented in section 5.10. Different TFs lead to varying NLL distributions, and an "envelope" is constructed around them, thus widening the uncertainties in the fit.

The functions considered for the bias study of the TF were first and second-degree Chebychev polynomials. The decision to halt at the second degree was motivated by a Fisher test (F-test) [211], given that for all years and fit ranges, increasing in complexity does not provide a significantly better fit to the TF distributions. Further details on this methodology and the outcomes of determining the optimal polynomial degree are available in Appendix F. It is important to note that these results are based on simulated samples and may not fully capture the complexity of the actual data.

Profiled NLL scans for an exemplary fit in a hypothetical Higgs mass point of the 2017 SL, 2017 FH, and 2018 FH analyses can be found in Fig. 5.40. The envelope and the individual scans corresponding to each function are shown. In these plots, Asimov datasets are generated with the function preferred by the aforementioned F-test and fitted with all alternatives to determine the envelope. An Asimov dataset is a representative dataset that is generated according to the expected values of certain distribution, in this case, the background *pdf*.

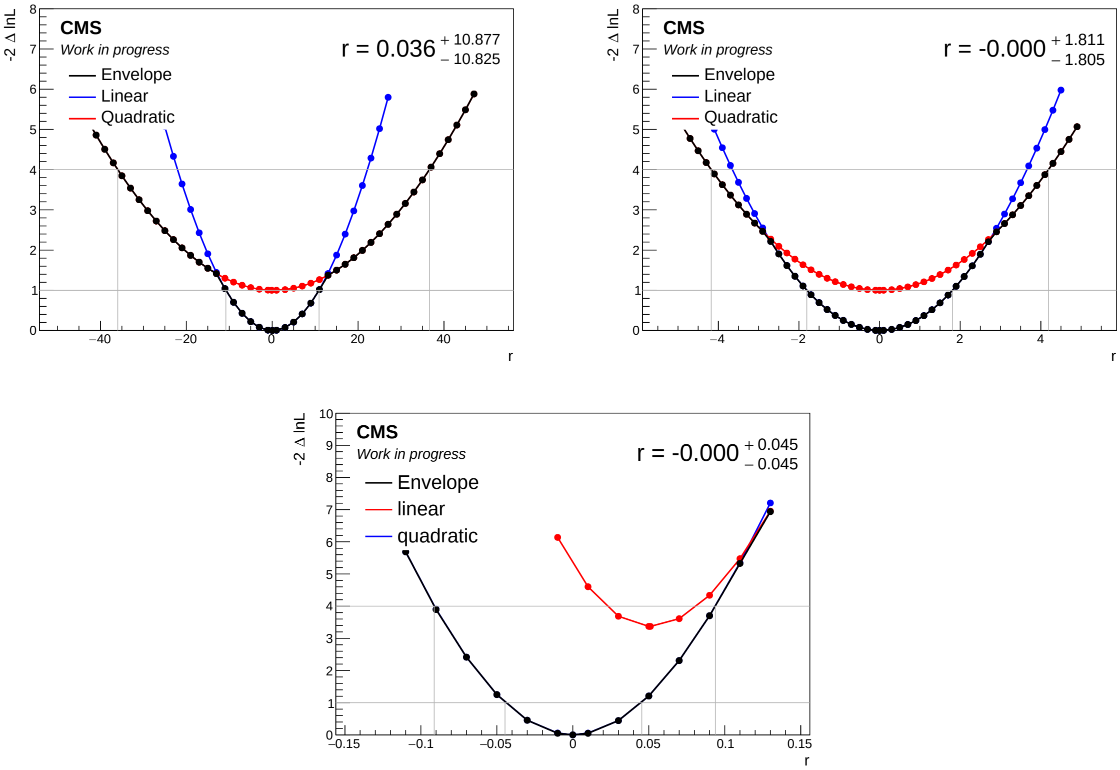


Figure 5.40: Profiled scans of twice the negative of the logarithm of the likelihood ratio function at 160 GeV in 2017 SL analysis (top-left), 400 GeV in 2017 FH analysis (top-right), and 1400 GeV in 2018 FH analysis (bottom).

5.9 Signal extraction

As mentioned earlier, the signal extraction is performed based on a simultaneous maximum likelihood signal-plus-background fit of the SR and CR of each fit range and analysis category. The functions fitted to the CR (f_{CR}) and SR (f_{SR}) are defined as:

$$\begin{aligned} F_{\mathrm{CR}}(M_{12}) &= N_{\mathrm{cr}} \cdot f_{\mathrm{cr}}(M_{12};\theta_{\mathrm{cr}}), && (5.11) \\ F_{\mathrm{SR}}(M_{12}) &= \mu \cdot S\ f_{\mathrm{s}}(M_{12};\theta_{\mathrm{s}}) + B \cdot f_{\mathrm{b}}(M_{12};\theta_{\mathrm{b}}), && (5.12) \\ f_{\mathrm{b}}(M_{12};\theta_{\mathrm{b}}) &= f_{\mathrm{cr}}(M_{12};\theta_{\mathrm{cr}}) \cdot f_{\mathrm{tf}}(M_{12};\theta_{\mathrm{tf}}). && (5.13) \end{aligned}$$

In equation 5.12 N_{cr} is the CR normalisation, f_{cr} its parametric description, and θ_{cr} the set of systematic uncertainties on the description of the CR. In equation 5.13 S and B are the expected signal and background yields in the signal region, respectively. The parameter μ is referred to as the signal strength modifier. In the present study, the signal distributions are normalised to 1 pb, so μ accounts for the cross-section of possible b-associated produced Higgs times branching fraction of decay into a pair of b-quarks. The parametric form describing the signal distribution is represented as f_s, while the background in the SR is represented as f_b, which can be rewritten as the multiplication of the functions describing the CR (f_{cr}) and the TF (f_{tf}), as expressed in equation 5.13. The set of systematic uncertainties on the description of the signal and the transfer factor shapes are represented as θ_{s} and θ_{tf}, respectively.

The TF parameters are allowed to freely float in the fit, and the CR parameters are similarly floating unless stated otherwise. It is relevant to note that the CR is about seven times more statistically populated than the SR and is thus heavily constrained in the simultaneous fit. The simultaneous fit allows the uncertainties of the modelling of the CR to be propagated to the signal strength.

5.10 Statistical methods

The nature of collision events and the detector response is intrinsically random. Therefore, statistical methods are employed to infer meaningful results from the data. In the case of a BSM search, such as the one presented in this analysis, we want to probe the presence of new signal events and, if no signal is present, set upper limits (UL) on the cross-section times branching ratio of the process.

The "Asymptotic CLs method" [212, 213], based on a modified frequentist approach, provides the means of setting these limits. The corresponding statistical calculations are performed with the COMBINE TOOL package [214], originally developed for the combination of the results on the searches for the Higgs boson by the CMS experiment. This versatile and powerful tool is used to perform the fits to extract the signal strength, which is then converted into cross-sections multiplied by the branching fractions or its 95% confidence level upper limits. It is also used to perform goodness-of-fit tests, fit diagnostics, as well as treatment of systematic uncertainties as nuisance parameters and determine their pulls and impacts on the parameters of interest.

5.10.1 The likelihood function

The likelihood function is defined as the probability density function (*pdf*) of the observed data as a function of specific model parameters, quantifying the agreement between the model expectation and the measurement.

The likelihood function, $\mathscr{L}(data|\mu,\theta)$ can be constructed as follows [214] for the observed events (*data*), composed of the BSM Higgs boson events s and the background yields b:

$$\mathscr{L}(\mathrm{data}|\mu,\theta) = \mathrm{Poisson}(\mathrm{data}|\mu\cdot s(\theta)+b(\theta))\cdot\prod p_j(\tilde{\theta_j}|\theta_j), \tag{5.14}$$

where *data* can either be the observed experimental data or the pseudo-data used to construct sampling distributions.

The first component in equation 5.14 stands for a Poisson distribution, in the case of unbinned shape analysis, with a total number of events $N = S + B$, it can be written as:

$$\mathrm{Poisson}(\mathrm{data}|\mu\cdot s(\theta)+b(\theta)) = \frac{e^{-(\mu S+B)}}{N!}\cdot\prod_{i=1}^{N}(\mu\cdot Sf_{\mathrm{s}}(\bar{x_i},\bar{\theta})+Bf_{\mathrm{b}}(\bar{x_i},\bar{\theta})), \tag{5.15}$$

where $f_s(\bar{x_i},\bar{\theta})$ and $f_b(\bar{x_i},\bar{\theta})$ are the *pdfs* of signal and background of the set of observables $\bar{x}$ in the event i, while S and B are total event rates expected for signal and backgrounds. These *pdfs* are dependent on the systematic uncertainties, represented as $\bar{\theta}$, also called nuisance parameters. In statistical models, systematic uncertainties are dealt with by introducing nuisance parameters. The $p_j(\tilde{\theta_j}|\theta_j)$ term in equation 5.14 represents the *pdfs* of observing the first estimate of the nuisance parameters, $\tilde{\theta}_j$, for a true value of θ_j, corresponding to a systematic uncertainty j.

The form of $p(\tilde{\theta}|\theta)$ will depend on the type of systematic uncertainty. As outlined in section 5.7.2, two types of systematic uncertainties affecting the signal can be identified: normalisation-uncertainties and shape-uncertainties. Those that are classified within the first category are parametrized with a log-normal distribution:

$$p(\tilde{\theta}|\theta) = \frac{1}{\sqrt{2\pi}\ln(\kappa)} \exp\left(-\frac{(\ln \tilde{\theta}/\theta)^2}{2(\ln \kappa)^2} \right) \frac{1}{\tilde{\theta}}, \tag{5.16}$$

where κ characterizes the width of the distribution and can be approximated to $\kappa = e^{\tilde{\sigma}}$, with $\tilde{\sigma}$ connected to the width of the Gaussian distribution [214].

The second group of systematic uncertainties, i.e. those influencing the signal shape distributions, is treated by determining the impact on the shape parameters of the signal samples parametrization and incorporating this effect into the signal model. The corresponding θ is treated as a Gaussian-constrained nuisance parameter, whose *pdf* is defined as follows:

$$p(\tilde{\theta}|\theta) = \frac{1}{\sqrt{2\pi}\tilde{\sigma}} \exp\left(-\frac{(\tilde{\theta}-\theta)^2}{2\tilde{\sigma}^2} \right), \tag{5.17}$$

where $\tilde{\theta}$ corresponds to the initial estimate of the nuisance parameter, and $\tilde{\sigma}$ is the estimate of its standard deviation.

In addition, unconstrained nuisance parameters are assigned flat priors within a range much larger than their validity scope. An example of those in the present analysis is the CR fit parameters. Furthermore, the uncertainty related to the choice of the functional form of the transfer factor is addressed with the discrete profiling method [210], as detailed in section 5.8.3.

5.10.2 The maximum likelihood method

In this study, an unbinned analysis of the shape of the invariant mass distribution of the two leading b-tag jet candidates is performed. The signal strength (μ) of the searched BSM process is evaluated by maximizing the likelihood function for the observed events in the data, which substituting all terms in equation 5.14 is written as:

$$\mathscr{L}(\text{data}|\mu,\theta) = \frac{e^{-(\mu S+B)}}{N!} \cdot \prod_i (\mu S f_s(\bar{x_i},\bar{\theta}) + B f_b(\bar{x_i},\bar{\theta})) \cdot \prod_j p_j(\tilde{\theta}_j|\theta_j). \tag{5.18}$$

For simplicity, the negative logarithmic likelihood (NLL) is minimized instead. This process is equivalent to maximizing the likelihood function, but more computationally efficient given the product terms in equation 5.18 are treated as a sum. The minimization is performed numerically by MINUIT [215], which, in addition to finding the minimum, analyzes the NLL shape in its vecinity to find the best-fit values. Furthermore, nuisance parameter impacts are determined by profiling the NLL function around its minimum value. In the case of discrete nuisances, a likelihood scan is also performed as a function of the Higgs signal strength, by leaving the discrete index that identifies the selected functional form floating. Each point in the likelihood will give the lowest $-2 * \ln\mathrm{N}$ plus a correction factor that accounts for the different number of parameters in the proposed functions [210]. Then, a full scan is performed, and the "envelope" around the minimum of all the curves is utilised to address the uncertainty on the functional form of the distribution under scrutiny, as explained in section 5.8.3.

5.10.3 Profile likelihood ratio as test statistics

The goal of a statistical test is to make a statement of how well the observed data is in agreement with a certain null hypothesis [216]. In the case of an analysis focused on the search for a certain signal, it is relevant to examine the compatibility to a background-only and a signal-plus-background hypothesis. The profile likelihood ratio is a well-established test statistic for this kind of hypothesis testing. Its yield $\tilde{q}_\mu$ is defined as:

$$\tilde{q}_\mu = -2 \ln \frac{L(\text{data}|\mu, \hat{\theta}_\mu)}{L(\text{data}|\hat{\mu}, \hat{\theta})}, \tag{5.19}$$

for a certain signal strength μ under test, which is constraint by $0 \leq \hat{\mu} \leq \mu$. In this equation, $\hat{\mu}$ and $\hat{\theta}$ are the unconstrained estimators of the signal strength and the nuisance parameters from the fit to the data, respectively. On the other hand, $\hat{\theta}_\mu$ is the estimator of the nuisance parameter conditioned by μ.

According to Wilk's theorem the distribution of the log-likelihood ratio adheres to a chi-squared distribution with n degrees of freedom, where n represents the difference in the number of parameters between the numerator and denominator of the ratio [217]. This premise is valid in the assumption of an infinitely large sample size, also referred to as the asymptotic approximation. For equation 5.19, n is exactly 1.

The test statistics can be calculated for the two hypotheses under consideration as:

$$p_\mu = p_{s+b} = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{obs}|\hat{\mu} = \mu) = \int_{\tilde{q}_\mu^{obs}}^{\infty} f(\tilde{q}_\mu|\mu, \tilde{q}_\mu^{obs}) d\tilde{q}_\mu, \tag{5.20}$$

in the case of the signal-plus-background hypothesis, where $\tilde{q}_\mu^{obs}$ is the measured value of the test statistics for the signal strength μ, p_μ represents the probability of obtaining a test statistic value in a confidence interval $[\tilde{q}_\mu^{obs}; \infty]$, and $f(\tilde{q}_\mu|\mu, \tilde{q}_\mu^{obs})$ is the *pdf* that follows the one degree of freedom χ^2 distribution. Similarly, for the background-only conjecture:

$$p_0 = 1 - p_b = P(\tilde{q}_\mu \geq \tilde{q}_0^{obs} | \hat{\mu} = 0) = \int_{\tilde{q}_0^{obs}}^{\infty} f(\tilde{q}_\mu | 0, \tilde{q}_0^{obs}) d\tilde{q}_\mu, \tag{5.21}$$

where p_0 is the probability of rejecting this hypothesis, and the remaining components have an equivalent meaning to those from equation 5.20.

Overall, the χ^2 distributions of $f(\tilde{q}_\mu|\mu, \tilde{q}_\mu^{obs})$ and $f(\tilde{q}_\mu|0, \tilde{q}_0^{obs})$ can be generated with toy pseudo-data, which will be used for the next steps that include setting upper limits in case no significant signal is found.

5.10.4 Upper limits setting and confidence intervals

In the search for new phenomena, one can calculate the observed value $\tilde{q}\mu^{obs}$ to exclude a signal-plus-background hypothesis at a predetermined confidence level, typically 95%. For this purpose, a modified frequentist approach known as the CLs method is employed. This method tests both signal-plus-background and background-only hypotheses by testing the ratio of these two probabilities:

$$CLs(\mu) = \frac{p_\mu}{p_0} = \frac{p_{s+b}}{1 - p_b} \tag{5.22}$$

Upper limits are set by finding the value of μ at which CLs is equal to a certain value. By convention, CLs is required to be below or equal to 0.05 so that the signal strength value identified in this manner is excluded with a 95% confidence level.

Exemplary distributions of the *pdfs* $f(\tilde{q}_\mu|0, \tilde{q}_0^{obs})$ generated with pseudo-data are shown in figure 5.41, as well as an arbitrary value of $CLs(\mu)$ for illustrational purposes.

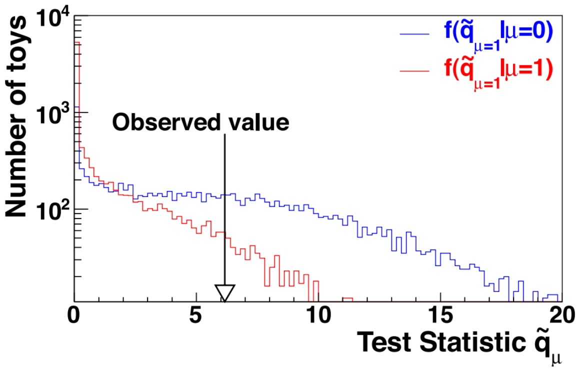


Figure 5.41: Pseudo-data-based distributions of test statistics corresponding to signal-plus-background and background-only hypotheses. The integral of these distributions in the interval from ∞ to the observed value marked with an arrow is employed to obtain the value of $CLs(\mu)$. Taken from [214]

On the other hand, to determine expected upper limits and $\pm 1\sigma$ and $\pm 2\sigma$ uncertainty bands, the procedure is slightly different. First, a large set of background-only hypothesis pseudo-data is generated. Then, the CLs and μ at 95%CL are calculated for each of them, as if they were real data. A cumulative probability distribution of results can be built by integrating from ∞. The point at which this distribution crosses the quantile of 50% is extracted as the expected value, while the crossings of the 16% (84%) define the $-(+)1\sigma$ band lower (upper) limit. Similarly, the crossings at 2.5% and 97.5% define the $\pm 2\sigma$ uncertainty bands lower and upper limit, respectively.

5.10.5 Combination across years and channels

The results from the analysis of 2017 data in the SL and FH categories and 2018 data in the FH category presented in this thesis will be combined with those from the published CMS analysis of the 2016 data in the FH category, targetting to provide results accounting for the full Run 2 CMS dataset, or up to nearly 127 fb^{-1} of data.

When more than one analysis is combined, the likelihoods of the individual components are also combined. Therefore, it is relevant to address whether there is a correlation between the systematic uncertainties across years. By convention, the uncertainties are treated as 100% (anti-)correlated or fully de-correlated. In this analysis, the uncertainties related to cross-year correlations in the luminosity measurement and cross-year correlations in the offline b-tagging uncertainty are treated as fully correlated across all analyses. When setting exclusion limits in the parameter space of benchmark scenarios, the uncertainties concerning theoretical calculations, i.e. the QCD renormalisation and factorization scale, the uncertainty on the definition of the proton's parton distribution function and on the definition of the strong coupling constant (see section 2.2.1), are also treated as fully correlated across years and categories.

The next chapter will present the results of analyzing the individual years and categories examined in this work, before and after unblinding, and the results of the full Run 2 combination.

chapter 6

Results

This chapter presents the results of the search for BSM Higgs bosons in final states with b-quarks, following the methodologies and strategies outlined in the previous chapter. The CMS 2017 dataset is analyzed in the semi-leptonic (SL) and fully-hadronic (FH) categories, as well as the 2018 dataset in the FH category. The combination of these analyses with a previous 2016-only search is also presented. A blinding policy is followed, meaning that the M_{12} distribution in the signal region (SR) is not inspected until the analysis methods are fully deployed and validated. Thus, the extraction of limits in the cross-section times branching ratio is first exercised using Asimov datasets and data in a signal-depleted validation region orthogonal to the signal region. Once the statistical methods are validated and the impacts of the main nuisance parameters are assessed, the signal region is unblinded, and the observed limits are obtained. These results are further interpreted as exclusion limits in the parameter space of several scenarios within MSSM and 2HDM models, discussed in sections 2.2.1 and 2.2.2.

6.1 Prior to unblinding

6.1.1 Test with Asimov data

Expected upper limits are initially calculated at 95% confidence level (C.L.) employing background-only Asimov datasets. These datasets are generated based on the Control Region (CR) shape multiplied by the transfer factor (TF) preferred by an F-test in each fit range (FR), defined for background modelling (see section 5.8). The Asimov dataset normalization is matched to the SR normalization in each range.

In the 2018 FH category, a few highly correlated parameters of the CR parametrization have been frozen to ensure a stable fit when extracting the limits. These are the parameters accounting for the extension of the Novosibirsk function with respect to the classic Novosibirsk (see Section 5.8.1) in all fit ranges, and the "peak" parameter of the extended Novosibirsk function, the latter only in FR3 and FR4. All other parameters are freely floating in the fits.

The expected upper limits for the cross-section times branching fraction for the 2017 SL, 2017 FH and 2018 FH channels are displayed individually in Figs. 6.1, 6.2, and 6.3, respectively. The resulting expected limits reflect the signal efficiency trend and the background shape. As discussed in chapter 5, an optimization of the M_{12} subrange division is performed to maximize the sensitivity. The signal hypotheses that lie in the overlap region between adjacent sub-ranges are tested in both regions, and the final decision on the sub-range assignment is made based on the range that yields the best expected limit. The aforementioned figures also show the limits of evaluating these signal hypotheses in both ranges. An improved performance is seen in the 2018 FH category compared with the 2017 counterpart. This is mainly driven by higher luminosity collected that year, the improved b-tagging algorithm at the trigger level, and the slightly loosened online b-tag thresholds described in section 3.2.4.

In Fig. 6.4 (left), the results employing the full Run 2 dataset are shown, combining the results of the 2017 SL, 2017 FH, and 2018 FH analyses with those from 2016-only. The latter is directly obtained from the results in Ref. [35]. The combination of both categories extends the limits to masses down to 125 GeV with the 2017 SL analysis and up to 1.8 TeV with the FH category, which starts contributing at 300 GeV. The Run 2 combination is compared to the individual contributions and the 2017 SL + FH combination in Fig. 6.4 (right). As expected, the results of the Run 2 combination are driven by the 2017 SL analysis in the region where it is unique, and mostly driven by the 2018 FH towards medium and higher masses.

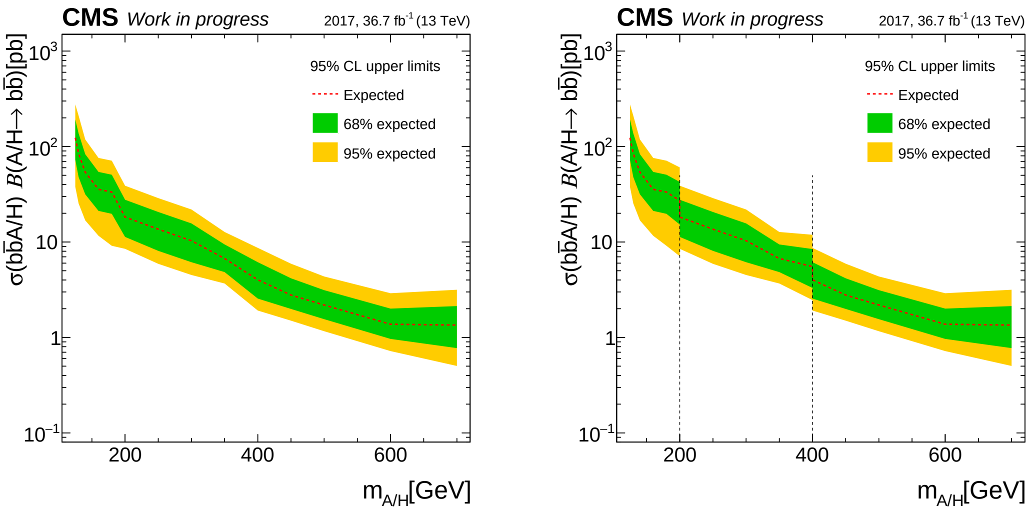


Figure 6.1: Expected limits for cross-section times branching fraction in the mass range 125–700 GeV, at 95% C.L., as determined before unblinding in the 2017 SL category. The distribution of signal hypothesis into fit ranges was optimized (left), after evaluating those peaking in the overlap of two ranges in both (right)

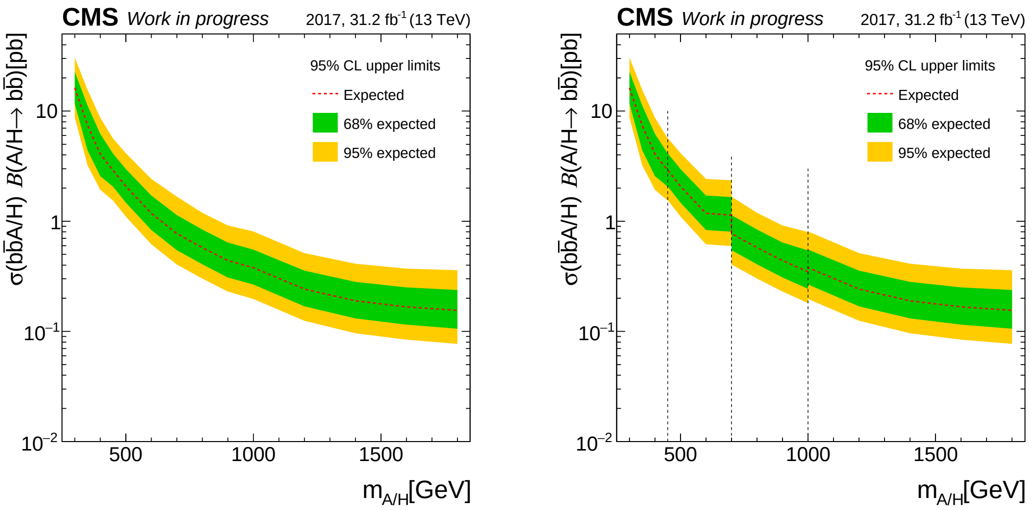


Figure 6.2: Expected limits for cross-section times branching fraction in the mass range 300–1800 GeV, at 95% C.L., as determined before unblinding in the 2017 FH category. The distribution of signal hypothesis into fit ranges was optimized (left), after evaluating those peaking in the overlap of two ranges in both (right)

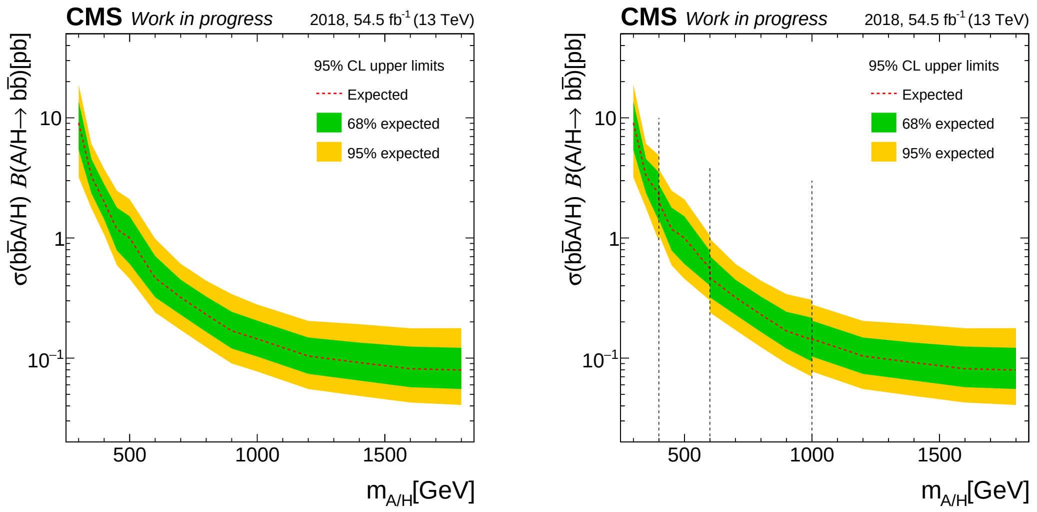


Figure 6.3: Expected limits for cross-section times branching fraction in the mass range 300–1800 GeV, at 95% C.L., as determined before unblinding in the 2018 FH category. The distribution of signal hypothesis into fit ranges was optimized (left), after evaluating those peaking in the overlap of two ranges in both (right)

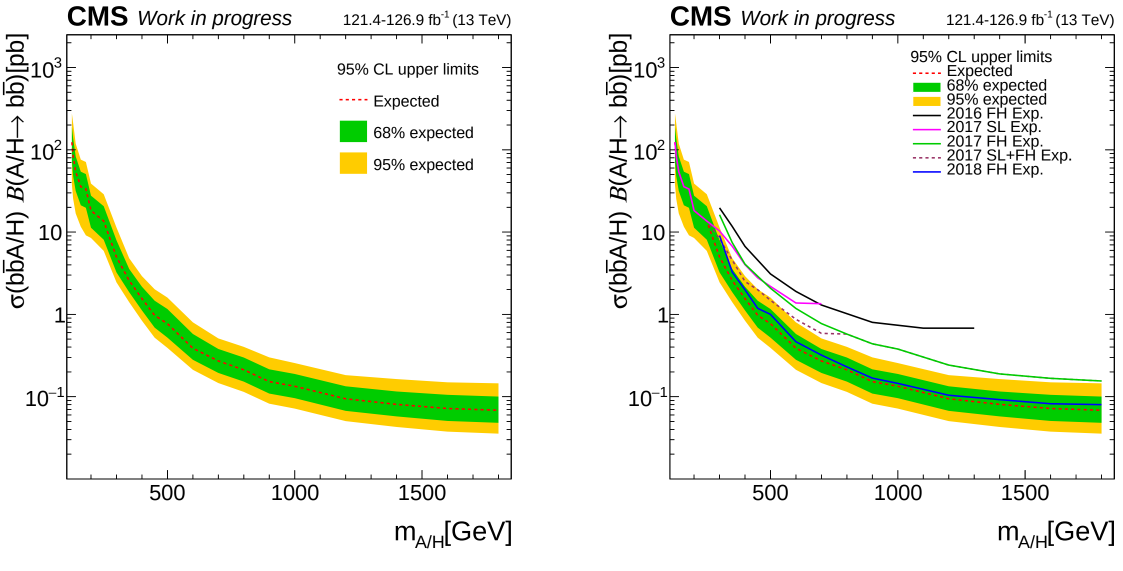


Figure 6.4: Expected limits for cross-section times branching fraction in the mass range 125–1800 GeV (left), at 95% C.L., as determined prior to unblinding in the 2017SL, 2017 FH and 2018 FH channel, combined with the 2016 limits from Ref. [35]. Asimov dataset are also used for the latter to be in agreement with the 2017 and 2018 results. The full Run 2 combination is plotted alongside the individual contributions, as well as the 2017 SL+FH combined (right).

The impacts and pulls of the nuisance parameters are obtained based on signal-plus-background Asimov data with an injected signal strength $r = 1$. The generated dataset is fitted, letting one nuisance parameter float at the time around their estimated value from a global fit and the influence on the fitted signal strength and potential constraints are assessed.

Figures 6.5, 6.6, and 6.7 show the pulls and impacts of the most relevant nuisance parameters for a representative mass point per fit range: 180, 400, and 600 GeV in the 2017 SL analysis; 350, 500, 900, and 1600 GeV in the 2017 FH analysis; and 400, 600, 800, and 1800 GeV in the 2018 FH analysis, respectively. The columns correspond, from left to right, to the name of the nuisance parameter, the pulls of the nuisance parameter or the fitted value with its uncertainty, and the impact on the signal strength. The pulls and impacts of the main nuisance parameters in four representative mass points of the Run 2 combination can be found in Fig. 6.8.

Overall, no parameters are found to be significantly constrained, apart from those freely floating in the fit. This is expected for Asimov datasets, as the log-normal nuisance parameters are centred at zero by construction.

One of the largest sources of uncertainty across all years and channels arises from the statistics of the data in the SR. Additionally, uncertainties related to background modelling are particularly significant. This includes uncertainties on the fitted parameters of the CR and the TF. The uncertainties associated with the online and offline b-tagging also have a substantial impact, especially at intermediate and higher mass points.

As mentioned in section 5.10.5, in the Run 2 combination, all uncertainties are treated as fully de-correlated, except cross-year luminosity and offline b-tagging uncertainties, which are treated as fully correlated.

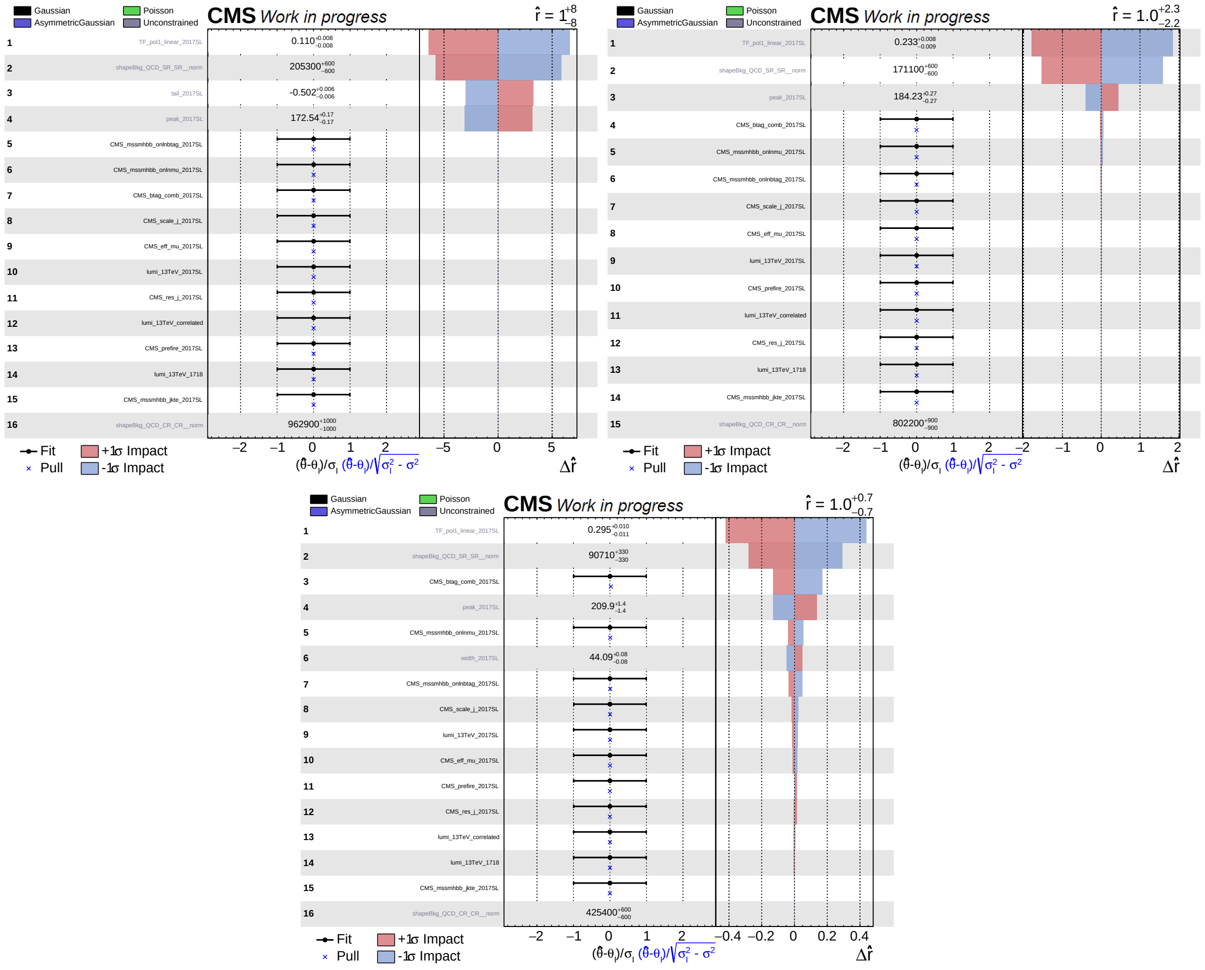


Figure 6.5: Pulls and impacts for the three representative mass points of the 2017 SL analysis: 180, 400, and 600 (top left to bottom) in their respective fit ranges FR1, FR2 and FR3. Obtained with Asimov data with an injected signal strength $r = 1$.

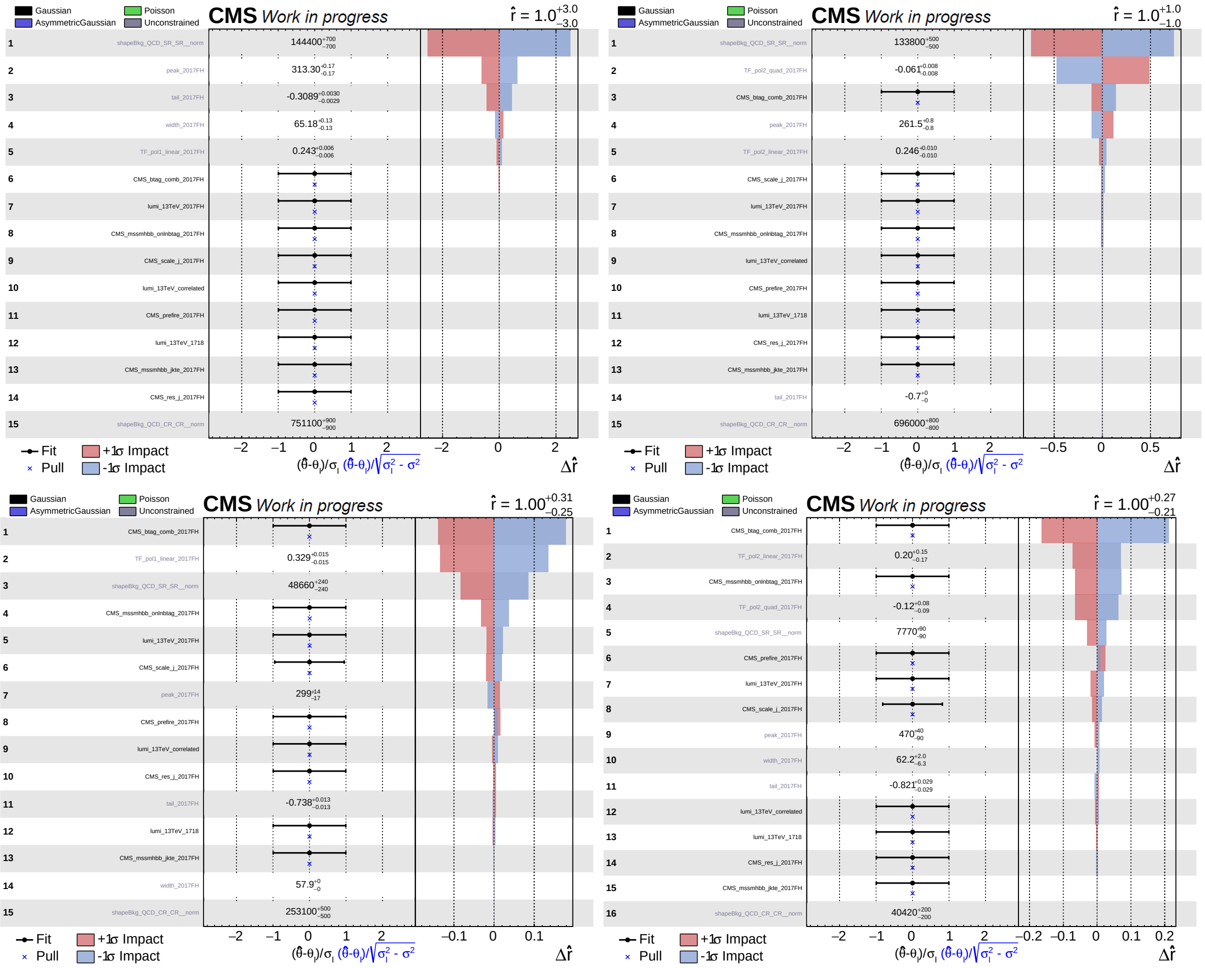


Figure 6.6: Pulls and impacts for the four representative mass points of the 2017 FH analysis: 350, 500, 900 and 1600 (top left to bottom) in their respective fit ranges FR1, FR2, FR3 and FR4. Obtained with Asimov data and an injected signal strength $r = 1$.

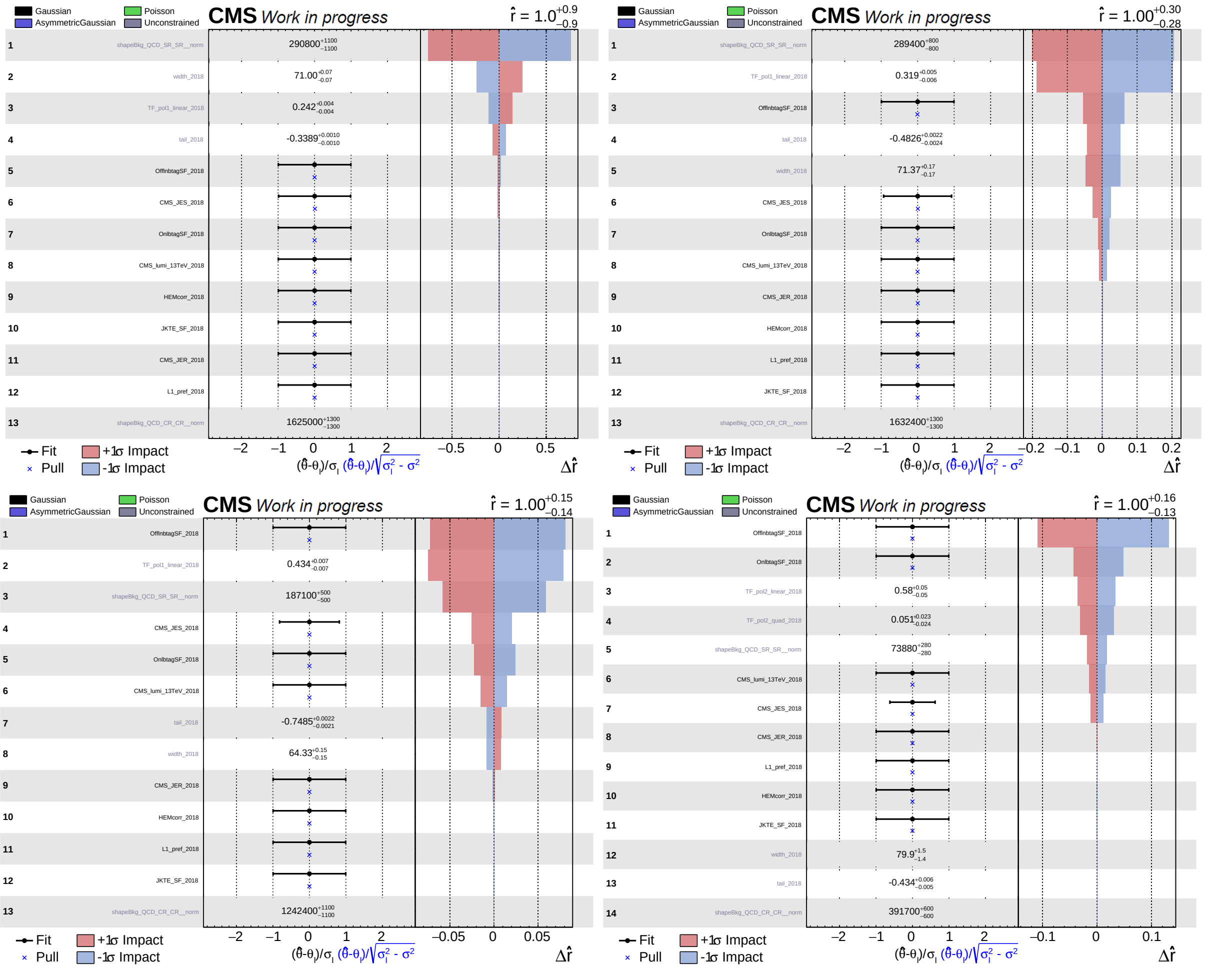


Figure 6.7: Pulls and impacts for the four representative mass points of the 2018 FH analysis: 400, 600, 800, and 1400 GeV (top left to bottom right) in their respective fit ranges FR1, FR2, FR3, and FR4. Obtained with Asimov data and an injected signal strength $r = 1$.

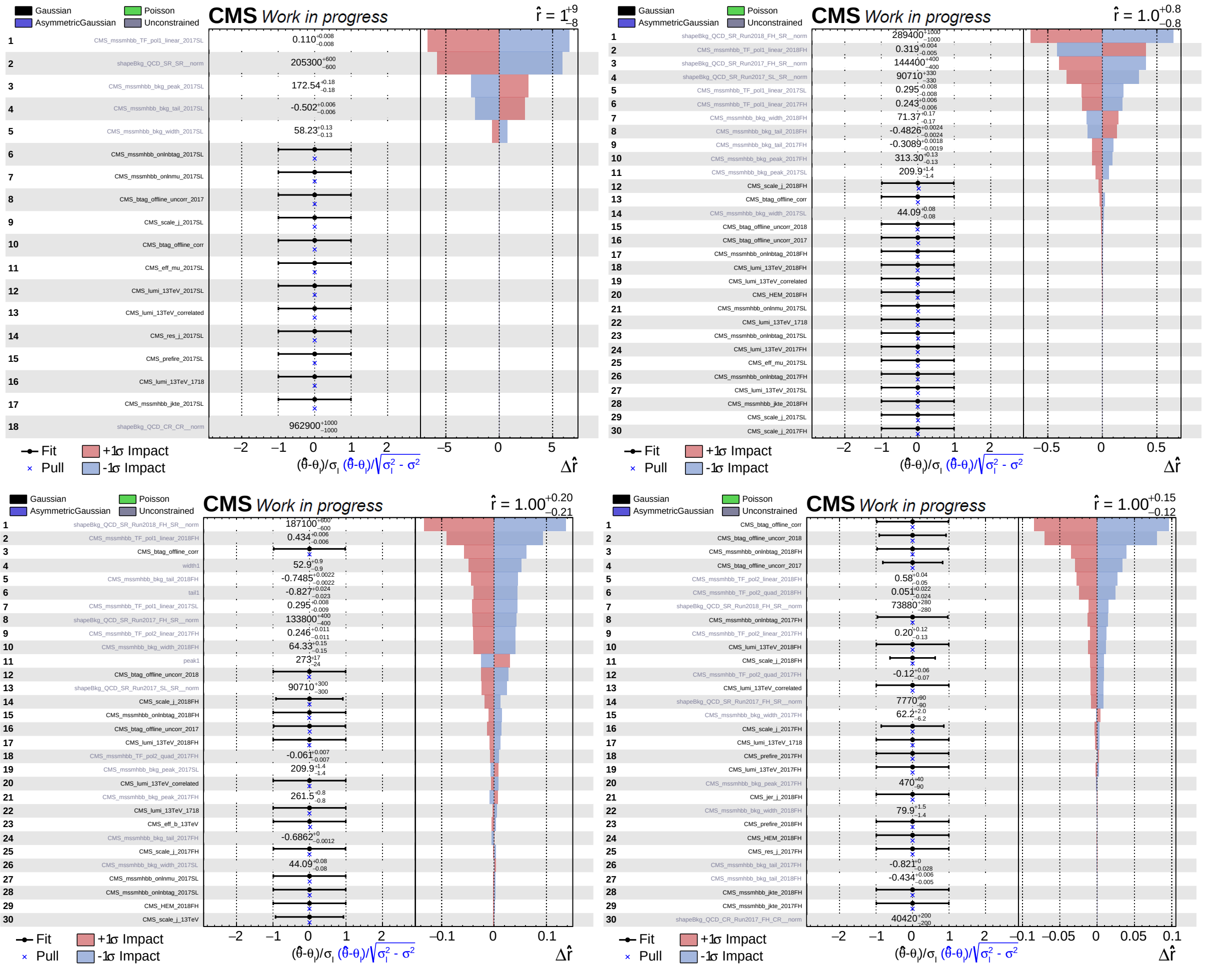


Figure 6.8: Main pulls and impacts for the combination in four representative mass points: 180 GeV, where the 2017 SL channel is the only contribution; 400 GeV and 600 GeV, which includes the contributions from 2016-only, 2017 SL, 2017 FH and 2018 FH; and 1400 GeV which includes the contributions from the 2017 FH and 2018 FH analyses. Obtained with Asimov data and an injected signal strength $r = 1$.

6.1.2 Test with data in the Validation Region

The signal extraction procedure is further scrutinized with real data before unblinding the signal region, exploiting the validation region. This region features a similar number of events as the SR, a similar kinematic and b-tagging selection of the two leading jets, and only differs on the offline b-tag selection of the third leading jet as discussed in section 5.6.

The methodologies foreseen for the signal extraction in the SR are applied to the validation region (VR), namely, the background modelling and fitting strategy. However, the signal templates are taken from the SR. Because of this, while the statistical inference procedure and background modelling can be validated, upper limits do not correspond to a cross-section measurement in the VR. Under these grounds, expected and observed upper limits in the cross-section are obtained. The results can be found in Fig. 6.9, 6.10, and 6.11 corresponding to 2017 SL, 2017 FH and 2018 FH categories, respectively. All signal mass points are found to be within or near the $\pm 2\sigma$ margin of the expected limit.

A Goodness-of-Fit (GoF) test is performed to probe the compatibility between the statistical model and the observed data [218]. The statistical model is the null hypothesis, according to which pseudo-data is generated. A saturated test statistic is evaluated for the observed data and pseudo-data, determining a quantity comparable to a χ^2 distribution [219]. The results of the test statistics are used to obtain a p-value under the null hypothesis. Figs. 6.12, 6.13, and 6.14 for 2017 SL, 2017 FH and 2018 FH, respectively. In all cases, good compatibility of the model with the data is observed.

In addition, Figures 6.15, 6.15, and 6.17 show the main pulls and impacts for a representative mass point per fit range in the 2017 SL, 2017 FH and 2018 FH analysis respectively. As for the Asimov dataset, no significant constraint was observed for any nuisance parameter apart from those freely floating.

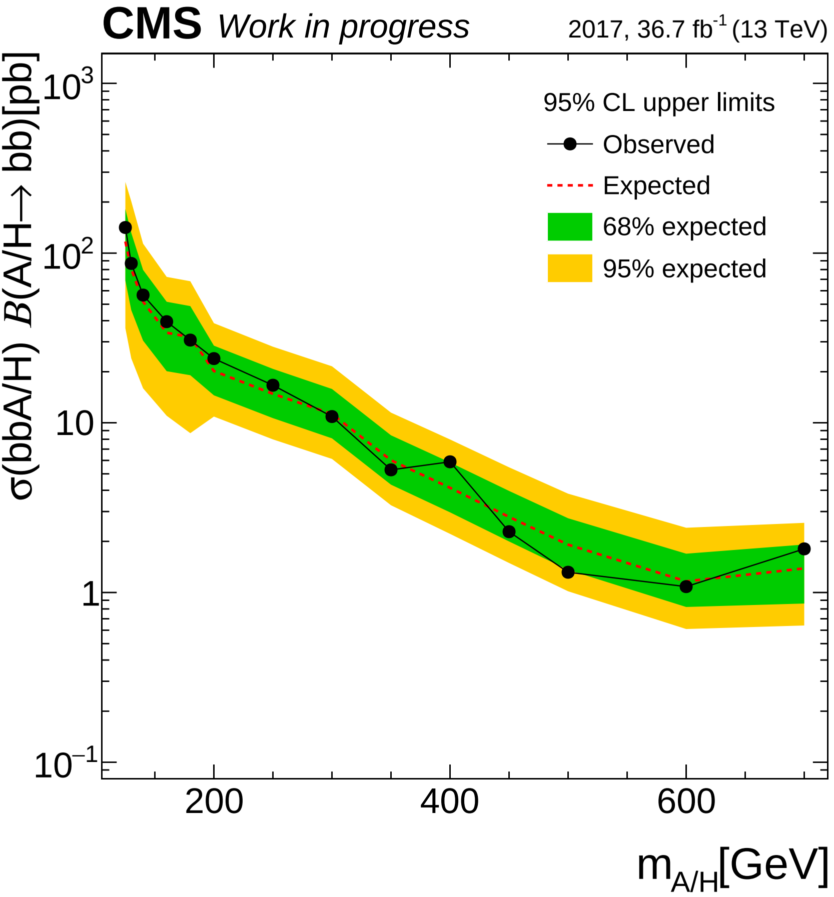


Figure 6.9: Expected and observed limits for cross-section times branching fraction at 95% CL, as determined in the validation region (VR) of the 2017 SL analysis. It is relevant to note that the VR is only a tool to test our statistical methods.

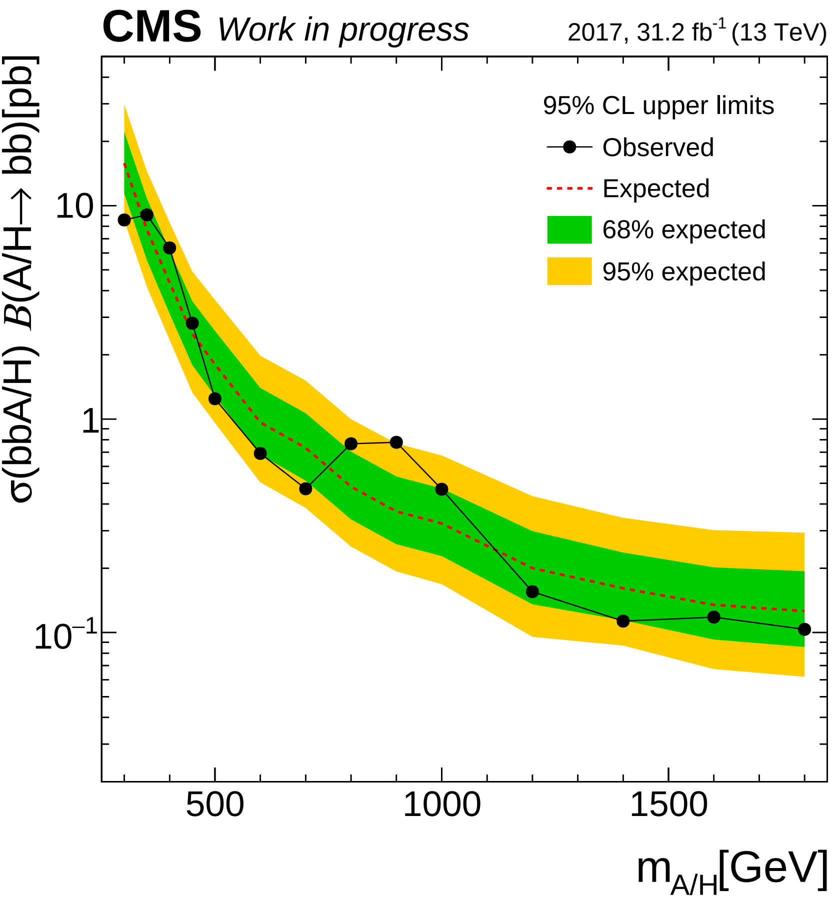


Figure 6.10: Expected and observed limits for cross-section times branching fraction at 95% CL, as determined in the validation region (VR) of the 2017 FH analysis. It is relevant to note that the VR is only a tool to test our statistical methods.

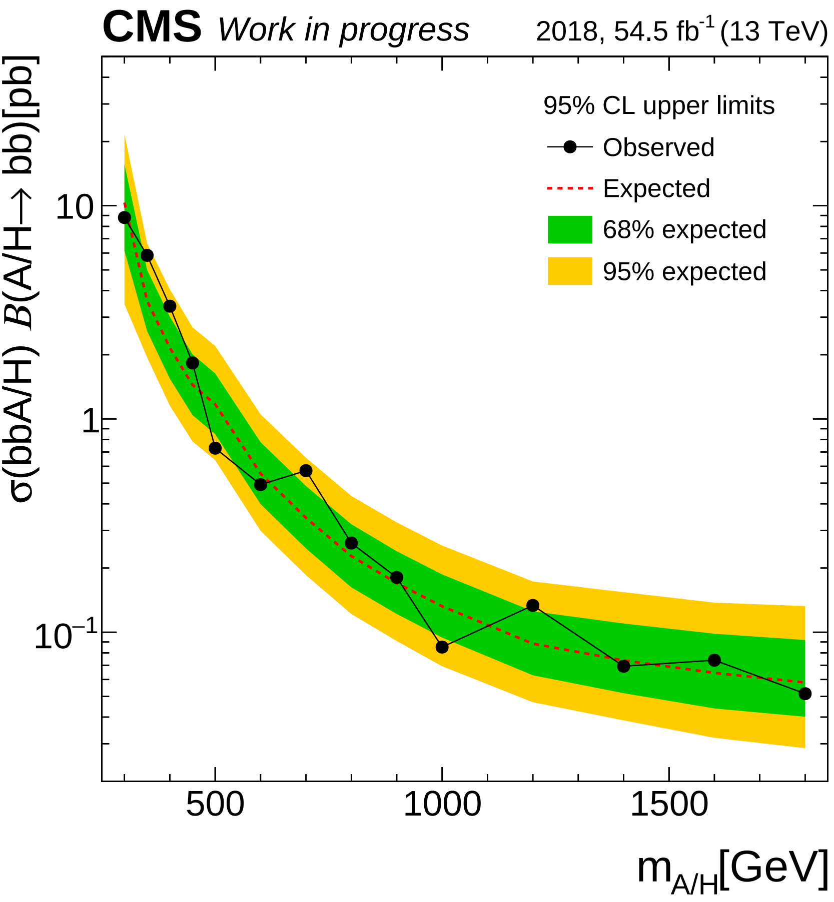


Figure 6.11: Expected and observed limits for cross-section times branching fraction at 95% CL, as determined in the validation region (VR) of the 2018 FH analysis. It is relevant to note that the VR is only a tool to test our statistical methods.

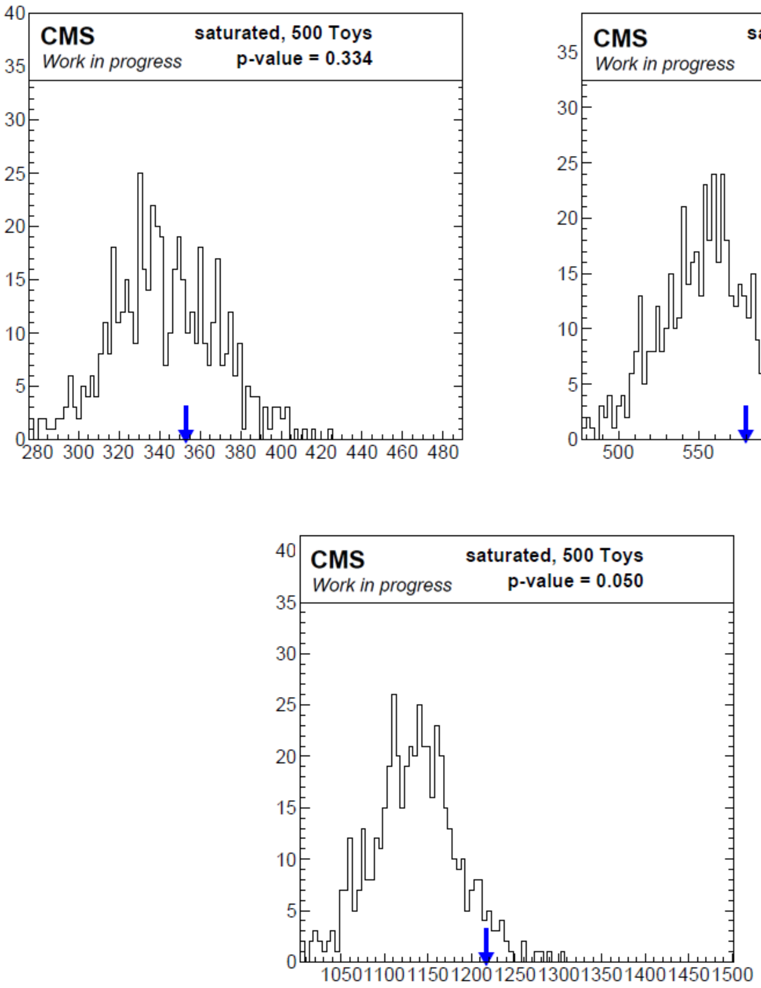


Figure 6.12: Goodness of fit tests corresponding to four representative mass-points in 2017 SL analysis: 180 GeV in FR1 (top left), 300 GeV FR2 (top right), and 500 GeV FR3 (bottom). The number of toy data sets that are generated and fitted for all mass points, and the respective GoF test results are shown in the histograms.

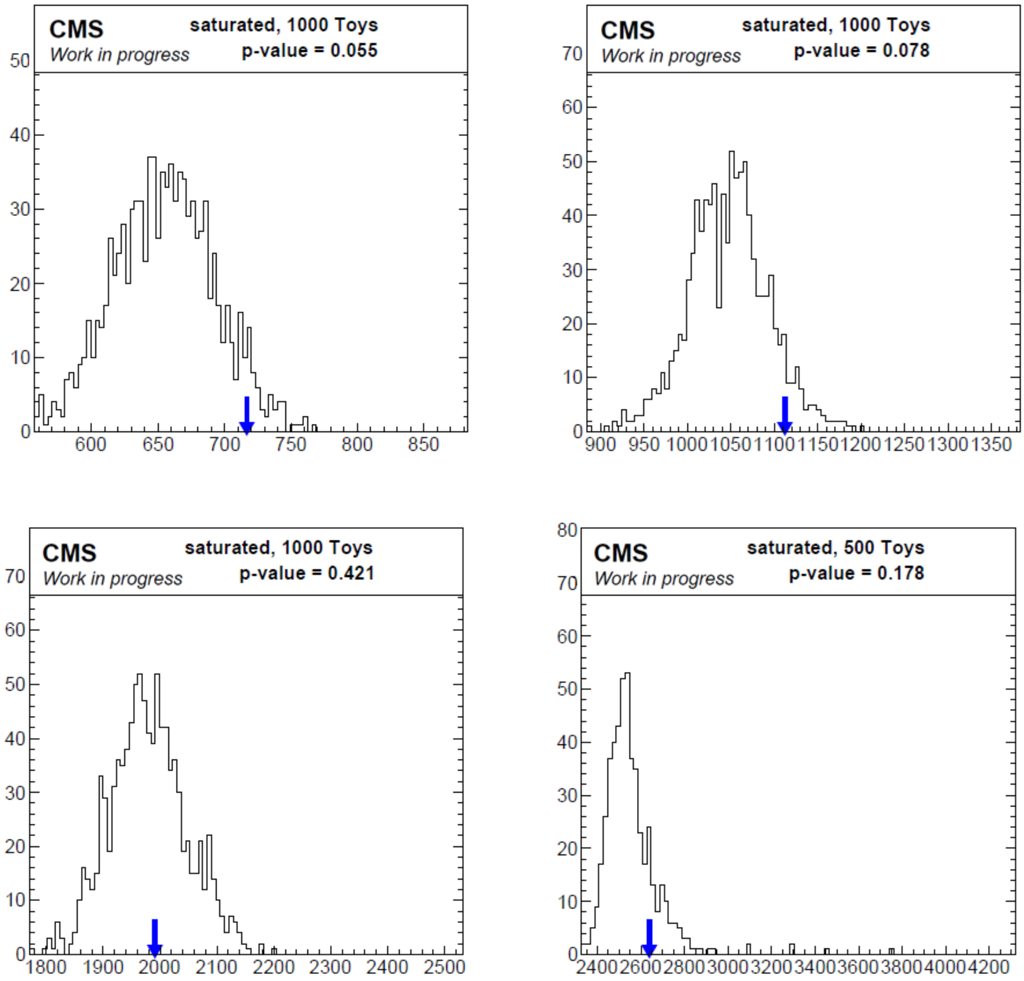


Figure 6.13: Goodness of fit tests corresponding to four representative mass-points in 2017 FH analysis: 400 GeV in FR1 (top left), 600 GeV FR2 (top right), 800 GeV FR3 (bottom left) and 1000 GeV FR4 (bottom right). The number of toy data sets that are generated and fitted for all mass points, and the respective GoF test results are shown in the histograms.

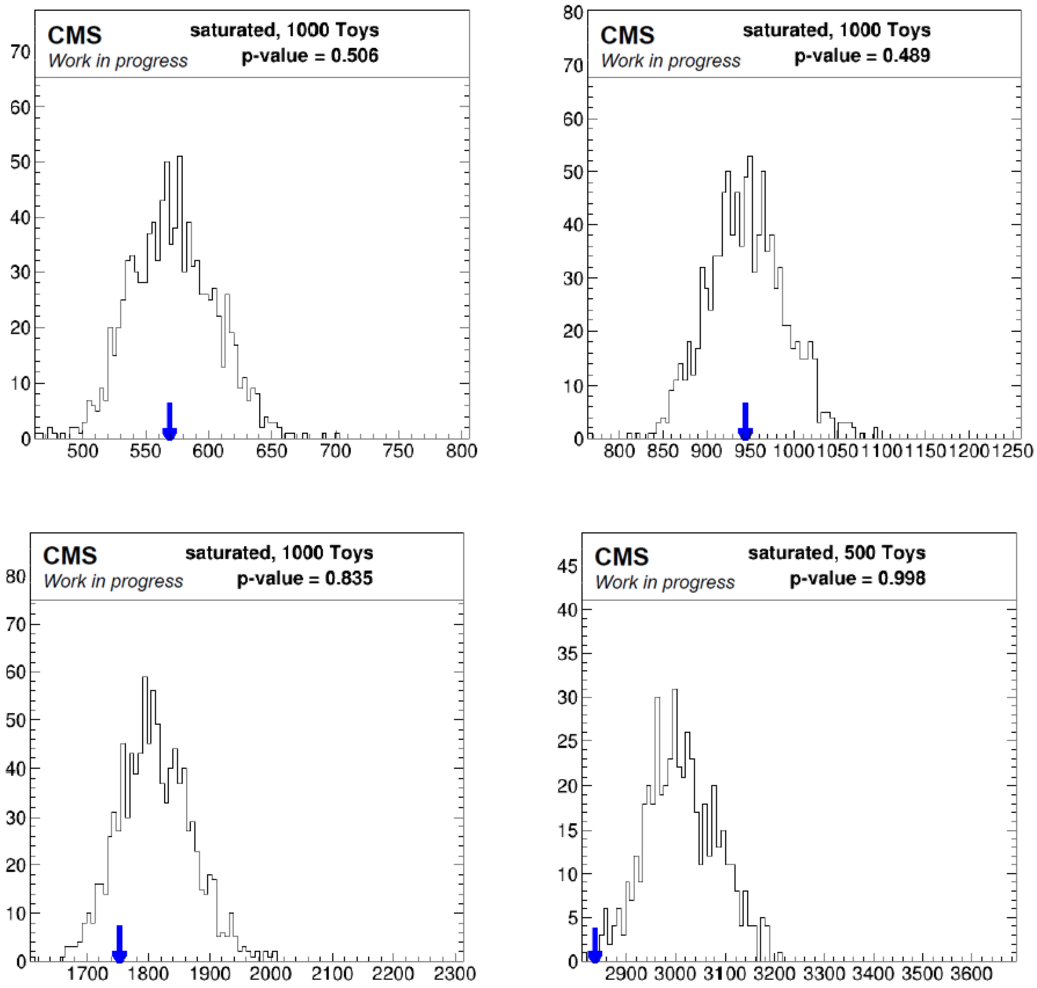


Figure 6.14: Goodness of fit tests corresponding to four representative mass-points in 2018 FH analysis: 400 GeV in FR1 (top left), 600 GeV FR2 (top right), 800 GeV FR3 (bottom left) and 1400 GeV FR4 (bottom right). The number of toy data sets that are generated and fitted for all mass points, and the respective GoF test results are shown in the histograms.

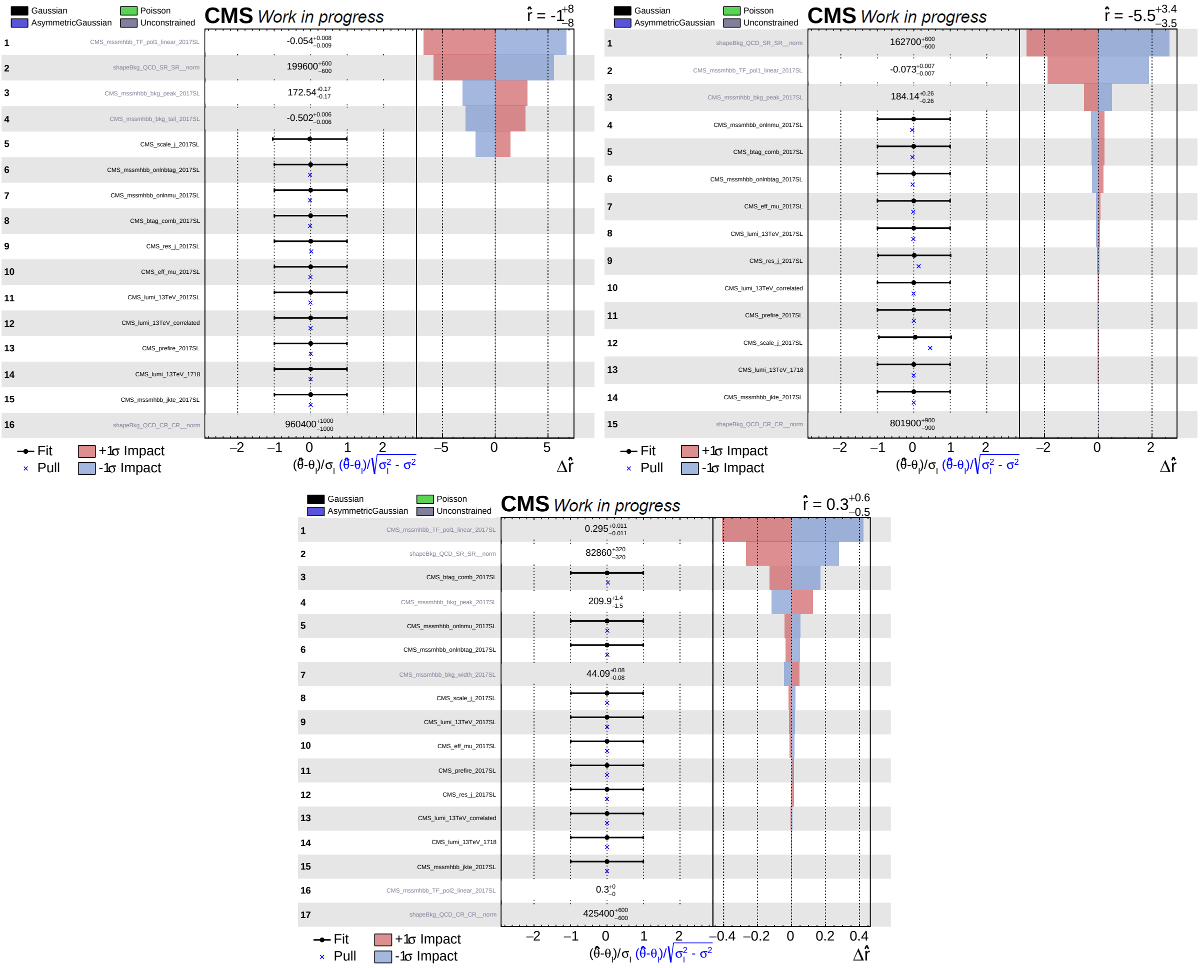


Figure 6.15: Pulls and impacts determined in the VR, corresponding to three representative mass points of the 2017 SL analysis: 180, 300, and 600 (top left to bottom) in their respective fit ranges FR1, FR2 and FR3.

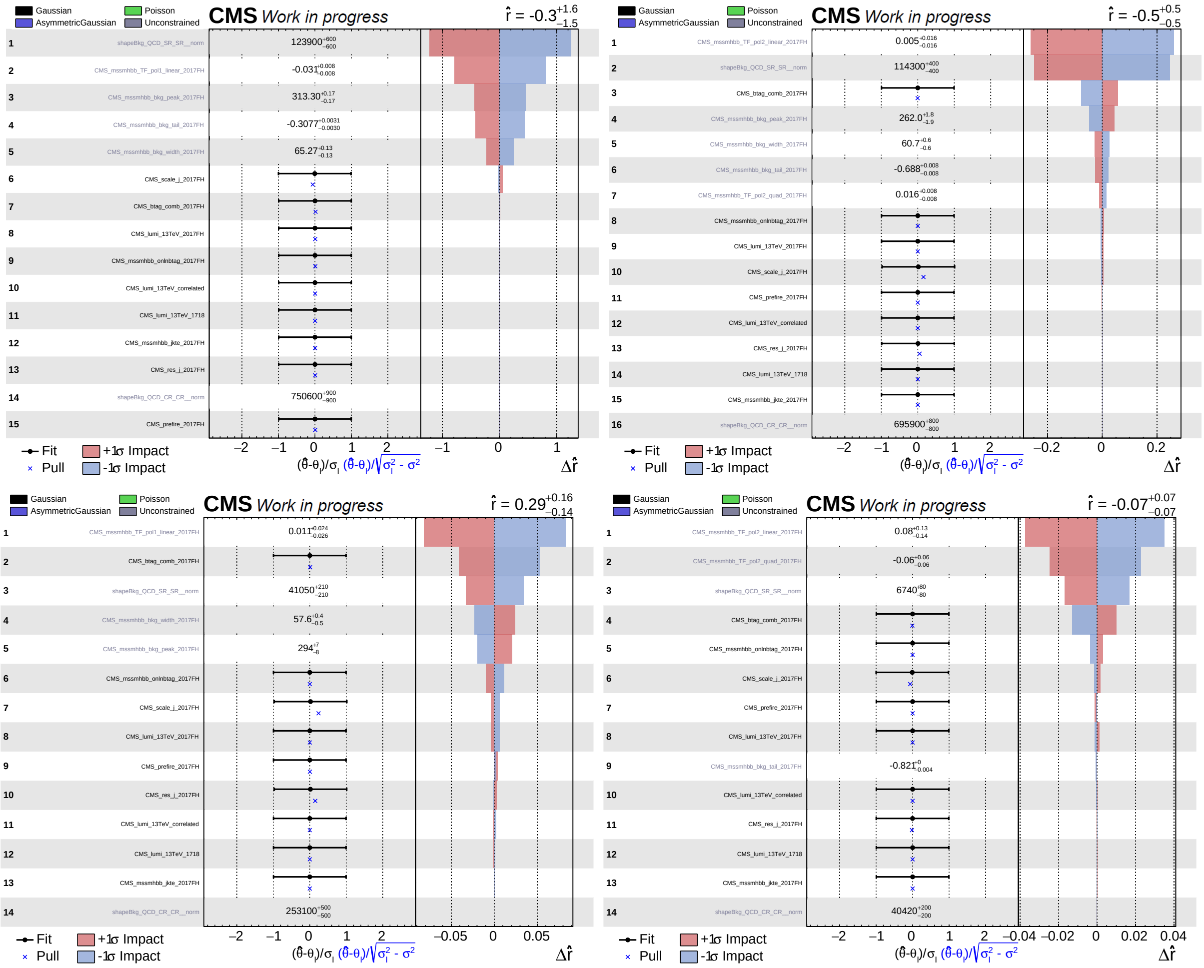


Figure 6.16: Pulls and impacts determined in the VR, corresponding to four representative mass points of the 2017 FH analysis: 400, 600, 1000 and 1400 (top left to bottom) in their respective fit ranges FR1, FR2, FR3 and FR4.

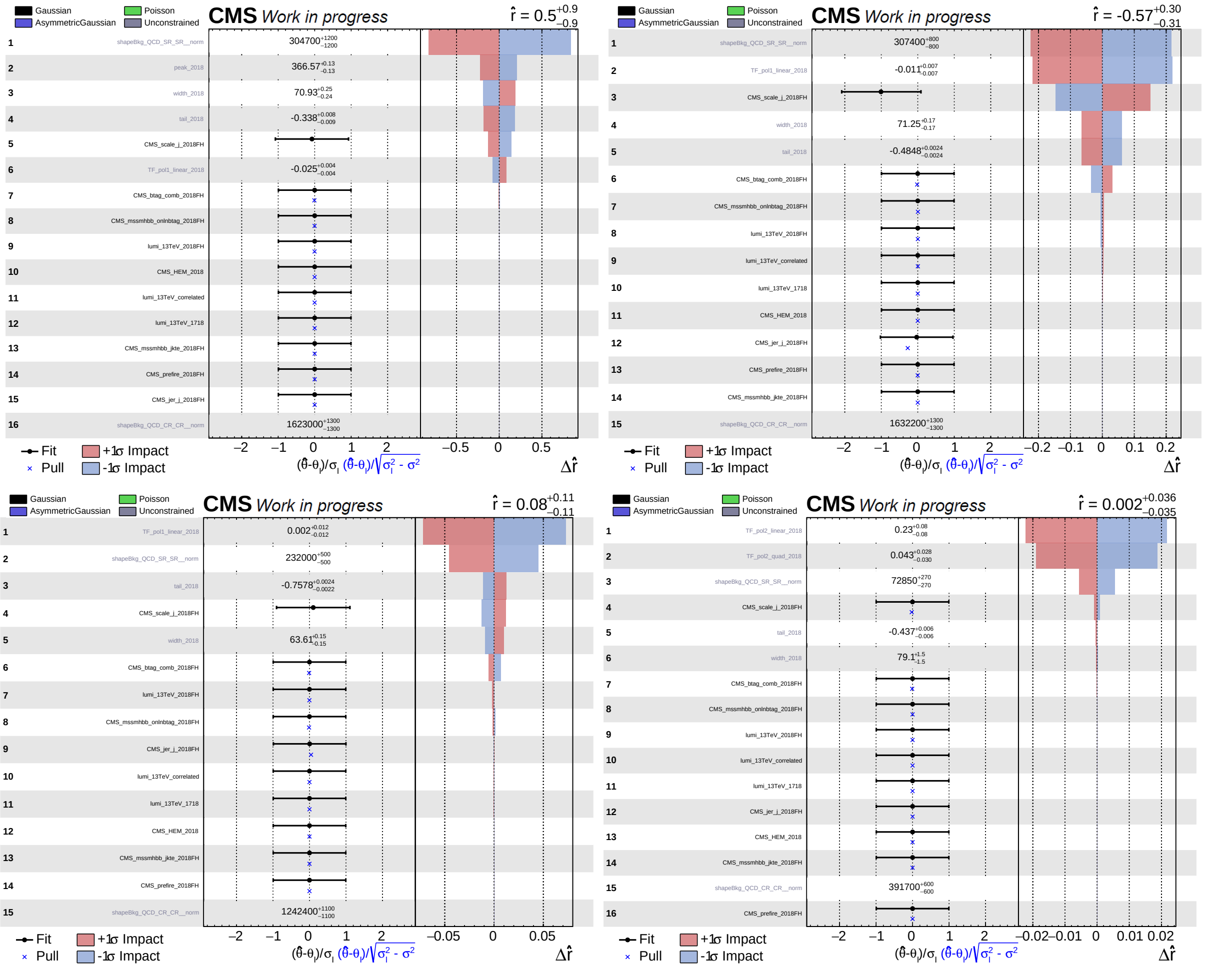


Figure 6.17: Pulls and impacts determined in the VR, corresponding to four representative mass points 400, 600, 800, and 1400 GeV (top left to bottom right) in their respective fit ranges FR1, FR2, FR3, and FR4.

6.2 Results from unblinded data

Once the analysis strategy has been validated, as discussed in the previous sections, the M_{12} distribution is unblinded in the SR. The procedure is conventionally done in two steps. First, the GoF tests, based on the saturated model, are obtained without examining the fitted signal strengths or the upper limits. Tables 6.1, 6.2, and 6.3 show the GoF test results for the 2017 SL, 2017 FH, and 2018 FH analyses, respectively.

An unexpectedly low p-value was found in the fully hadronic channels, particularly in the first fit range of the 2018 FH analysis and in the first and last fit range of the 2017 FH analysis. As mentioned in section 5.8.3, the discrete profiling method included a first and second-degree Chebychev polynomial. However, this expectation was estimated with an F-test performed for the parametrization of the TF based on simulated QCD samples, which are statistically limited and might not reflect the full complexity of the data. Aiming to improve the GoF test, a third-degree Chebychev polynomial was included for the discrete profiling in the FH category to provide more flexibility to the background shape. The observed GoF test significantly recovered afterwards. This is also shown in Table 6.2 and 6.3 for all mass-points of the 2017 FH and the 2018 FH analyses, respectively.

The inclusion of additional nuisance parameters may deteriorate the upper limits. However, in the 2018 FH (2017 FH) analysis, it was found that the expected limits only increased by up to a factor of 15% (30%) in two mass-points that were not affected by the small GoF test. The inclusion of a third-degree Chebyshev polynomial for the parametrization of the TF was done consistently for all fit-ranges of the FH category in the following stages of the unblinding. Similar to the parametrizations with polynomials of the first and second-degree described in section 5.8.2, the initial parameters were derived from simulated QCD samples. The corresponding fits and new scans of the negative logarithm of the likelihood ratio (NLL) can be found in Appendix I.

The unblinding procedure continued with the examination of the background-only and signal-plus-background fits of the M_{12} distribution, as will be discussed in the following.

FR	**Mass (GeV)**	**TF1+TF2**
FR1	125	0.404
FR1	130	0.404
FR1	140	0.400
FR1	160	0.400
FR1	180	0.402
FR1	200	0.402
FR2	250	0.412
FR2	300	0.390
FR2	350	0.328
FR2	400	0.328
FR3	450	0.104
FR3	500	0.106
FR3	600	0.124
FR3	700	0.124

Table 6.1: Summary of the p-values resulting from the GoF tests for mass hypotheses in the three fit-ranges of the 2017 SL analysis, based on the saturated model. The results correspond to using a TF with first and second-degree (TF1+TF2 column) in the discrete profiling method. A good p-value was found across all mass-points.

FR	**Mass (GeV)**	**TF1+TF2**	**TF1+TF2+TF3**
FR1	300	0.052	0.076
FR1	350	0.052	0.078
FR1	400	0.056	0.078
FR2	450	0.214	0.212
FR2	500	0.204	0.200
FR2	600	0.190	0.192
FR3	700	0.202	0.196
FR3	800	0.208	0.196
FR3	900	0.208	0.194
FR4	1000	0.020	0.070
FR4	1200	0.020	0.072
FR4	1400	0.016	0.074
FR4	1600	0.016	0.072
FR4	1800	0.018	0.070

Table 6.2: Summary of the p-values resulting from the GoF tests for mass hypotheses in the four fit-ranges of the 2017 FH analysis, based on the saturated model. The results correspond to using a TF with first and second-degree (TF1+TF2 column) in the discrete profiling method and also including the third-degree polynomial (TF1+TF2+TF3 column). Using the latter setting, the fit quality in fit-range 1 and 4 recovered.

FR	**Mass (GeV)**	**TF1+TF2**	**TF1+TF2+TF3**
FR1	300	0.12	0.30
FR1	350	0.07	0.30
FR1	400	0.07	0.36
FR2	450	0.69	0.70
FR2	500	0.69	0.70
FR2	600	0.69	0.68
FR3	700	0.88	0.88
FR3	800	0.88	0.88
FR3	900	0.89	0.88
FR4	1000	0.95	0.92
FR4	1200	0.94	0.92
FR4	1400	0.94	0.92
FR4	1600	0.94	0.93
FR4	1800	0.94	0.92

Table 6.3: Summary of the p-values resulting from the GoF tests for mass hypotheses in the four fit-ranges of the 2018 FH analysis, based on the saturated model. The results correspond to using a TF with first and second-degree (TF1+TF2 column) in the discrete profiling method and also including the third-degree polynomial (TF1+TF2+TF3 column). Using the latter setting, the fit quality in fit-range 1 recovered.

6.2.1 Background-only fits of M_{12} distributions in the SR

The results of the simultaneous CR and SR fit under the background-only hypothesis are shown in Figs. 6.18, 6.19, and 6.20, corresponding to the 2017 analyses in the SL and FH categories, and the 2018 analysis in the FH category, respectively. The normalised difference between the data and the estimated background shown in the bottom panel shows a good agreement with the background-only hypothesis across all analysis categories and fit-ranges.

6.2.2 Limits on the production cross-section times branching fraction

The signal extraction is performed in a signal-plus-background fit, and the expected and observed upper limits in the cross-section times branching fraction are obtained at 95% confidence level. The results are shown in Figs. 6.21, 6.22, and 6.23 corresponding to 2017 SL, 2017 FH and 2018 FH categories, respectively. All signal mass points are found to be within or near the $\pm 2\sigma$ margin of the expected limit. Out of the 42 mass-points probed, only three lie outside $\pm 2\sigma$. These include an excess of events in the 250 GV and 300 GeV mass-points of the 2017 SL analysis, with a local significance of 3.2 σ and 2.7 σ, respectively, and an excess of events of the order of 2.0 σ in the 450 GeV mass-point of the 2017 FH analysis.

The results of the Run 2 combination are shown in Fig 6.24. The limits range from 123 *pb* at 125 GeV down to 0.73 *pb* at 1800 GeV. As the excess at 300 in the 2017 SL analysis and at 450 GeV in the 2017 FH category are not seen in the other channels, they are not visible in the combination. However, the excess at 250 GeV prevailed since this mass-point is exclusively probed in the 2017 SL analysis. The results obtained considerably extend the exclusion limits and the scrutinized mass range in comparison to the 2016-only results [35], also shown in Fig. 6.24. The observed and expected upper limits set by this analysis range from 123 pb at 125 GeV towards 0.077 pb at 1.8 TeV, and the results for all mass-points are summarized in Appendix H.

The pulls and impacts of the nuisance parameters are shown Figs. 6.25, 6.26, 6.27, and 6.28 for a representative mass point per fit range in the 2017 SL, 2017 FH and 2018 FH analysis, and in the Run 2 combination, respectively. None of the nuisance parameters was found to be significantly constrained over all mass-range, apart from those allowed to float freely in the fit. In addition, there is no substantial pull in any of the nuisance parameters. As observed in the studies with Asimov data, the largest impacts arise from the statistics of the signal region, the uncertainty on the background modelling, and the uncertainties related to the online and offline b-tagging.

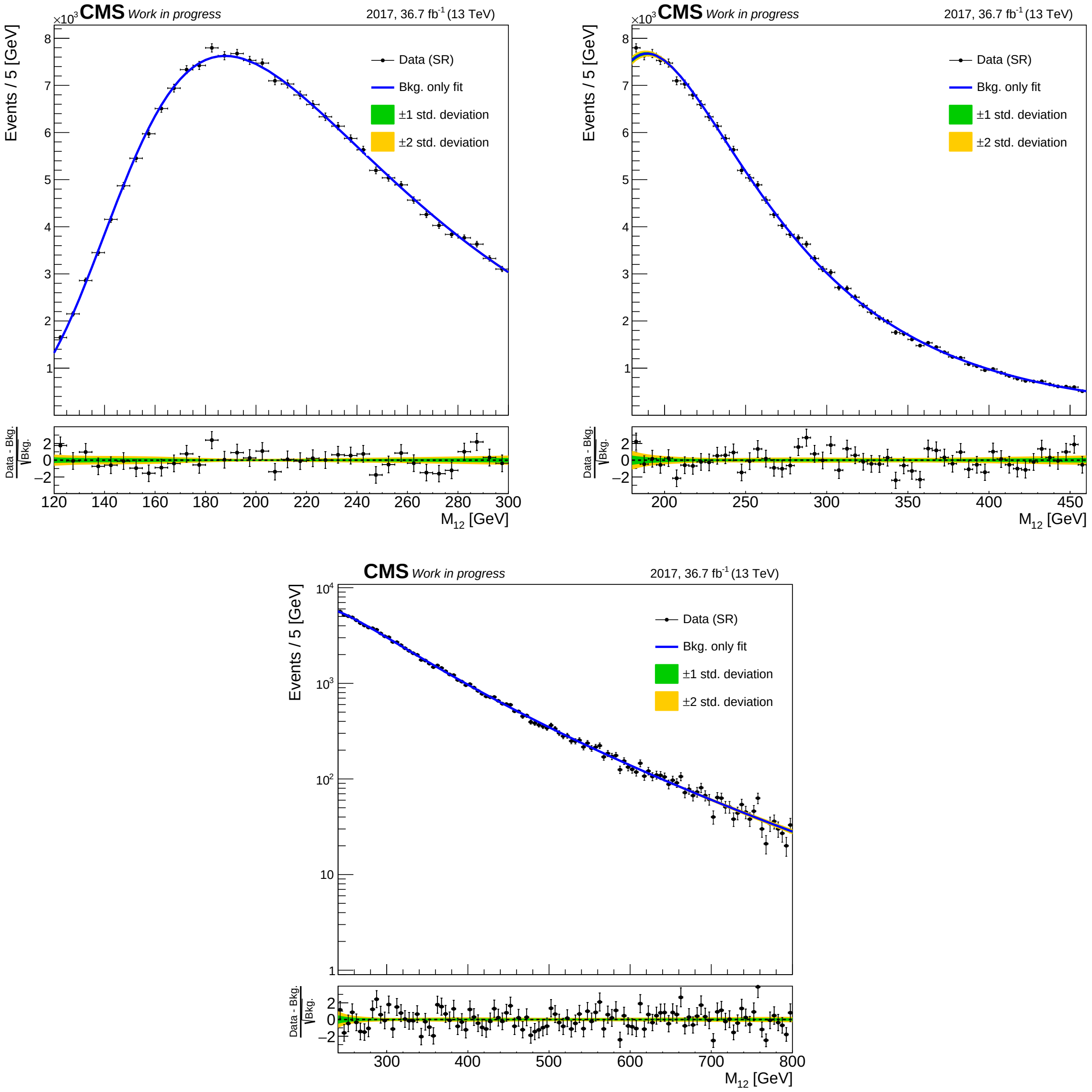


Figure 6.18: Background-only fits of the M_{12} distribution in each FR of the 2017 analysis in the SL category, shown together with $\pm 1{,}2\sigma$ uncertainty bands extracted from the fit. The pulls with respect to the estimated background in each FR are shown in the bottom panel.

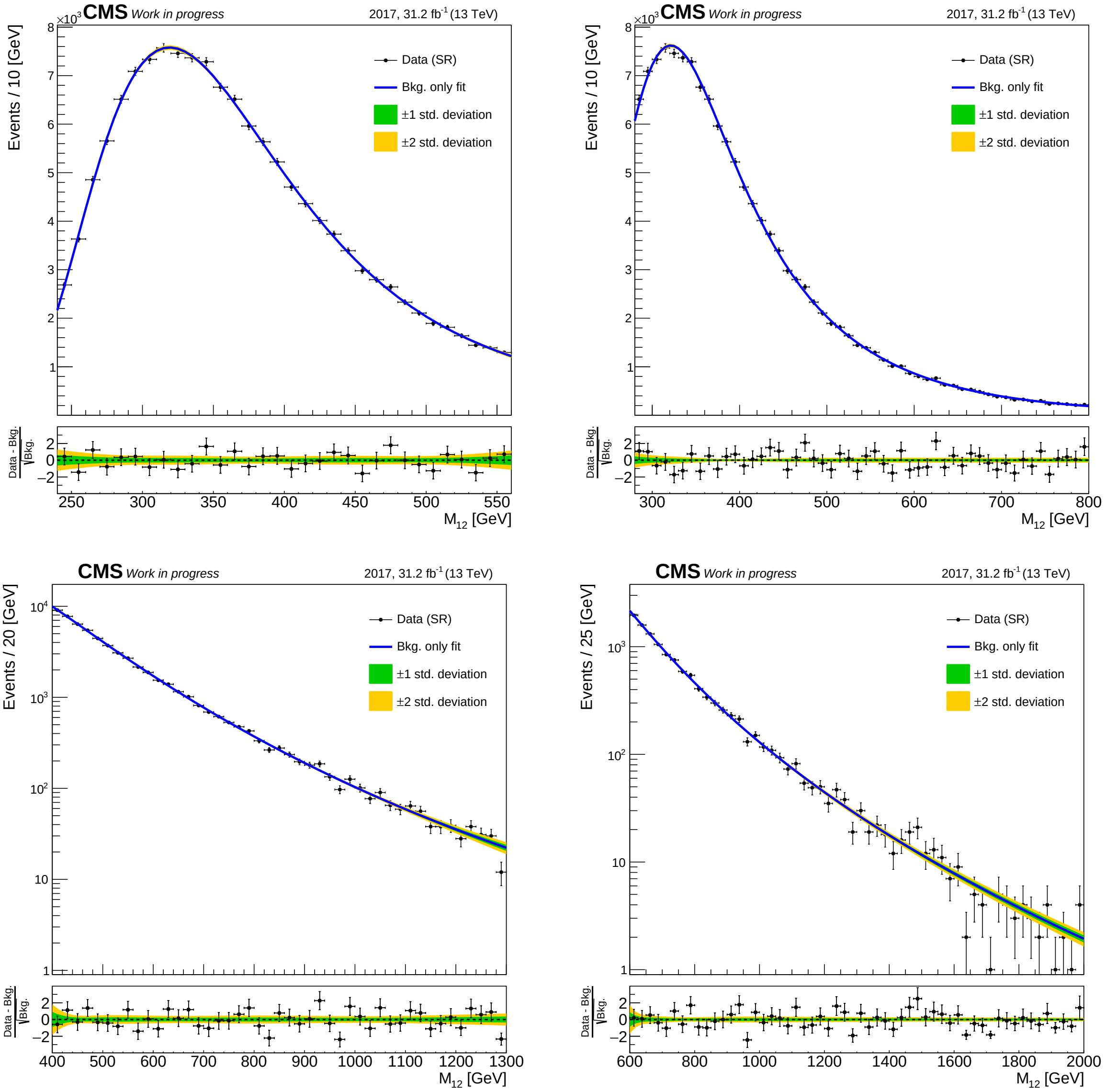


Figure 6.19: Background-only fits of the M_{12} distribution in each FR of the 2017 analysis in the FH category, shown together with $\pm 1{,}2\sigma$ uncertainty bands extracted from the fit. The pulls with respect to the estimated background in each FR are shown in the bottom panel.

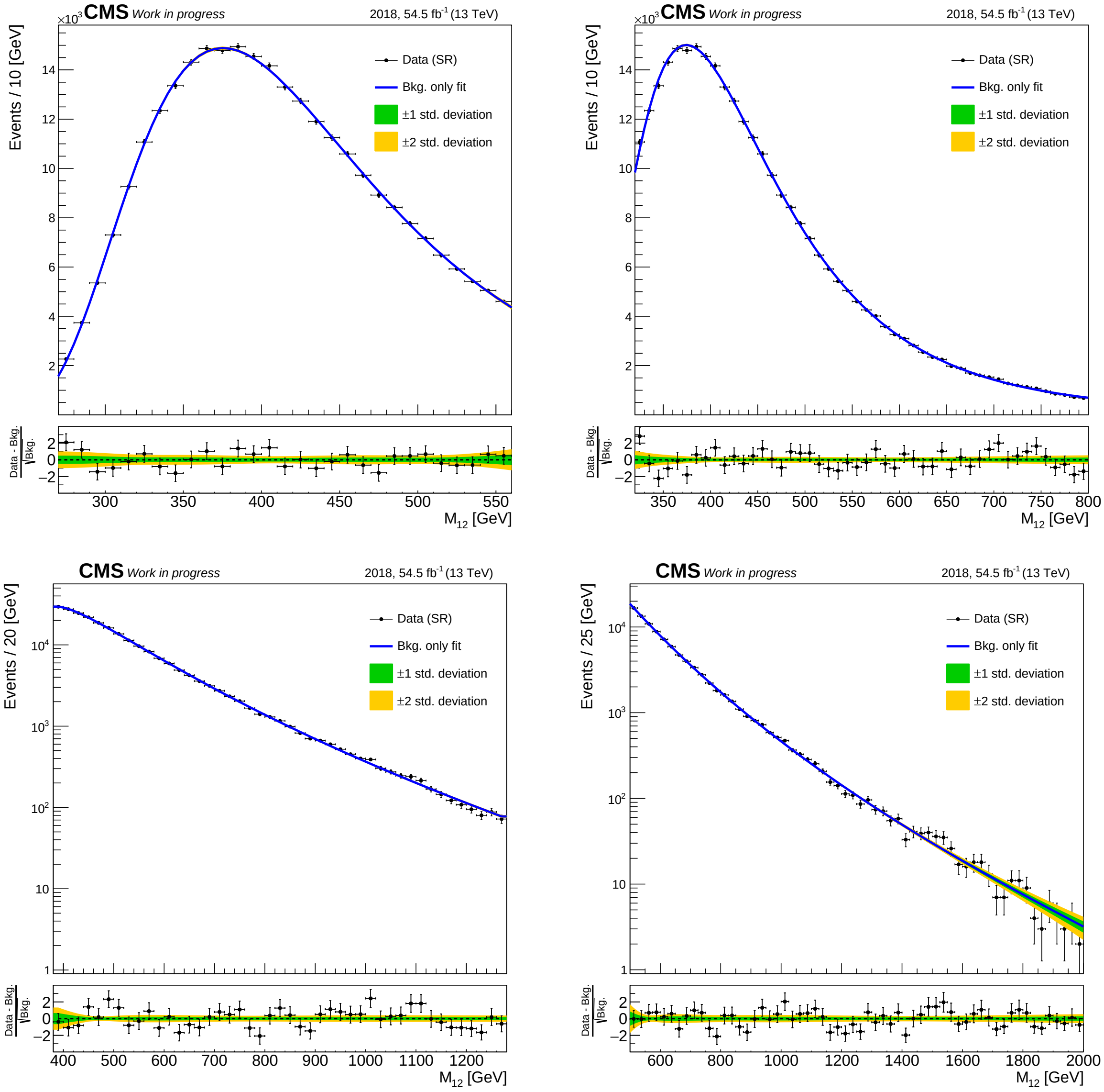


Figure 6.20: Background-only fits of the M_{12} distribution in each FR of the 2018 analysis in the FH category, shown together with $\pm 1,2\sigma$ uncertainty bands extracted from the fit. The pulls with respect to the estimated background in each FR are shown in the bottom panel.

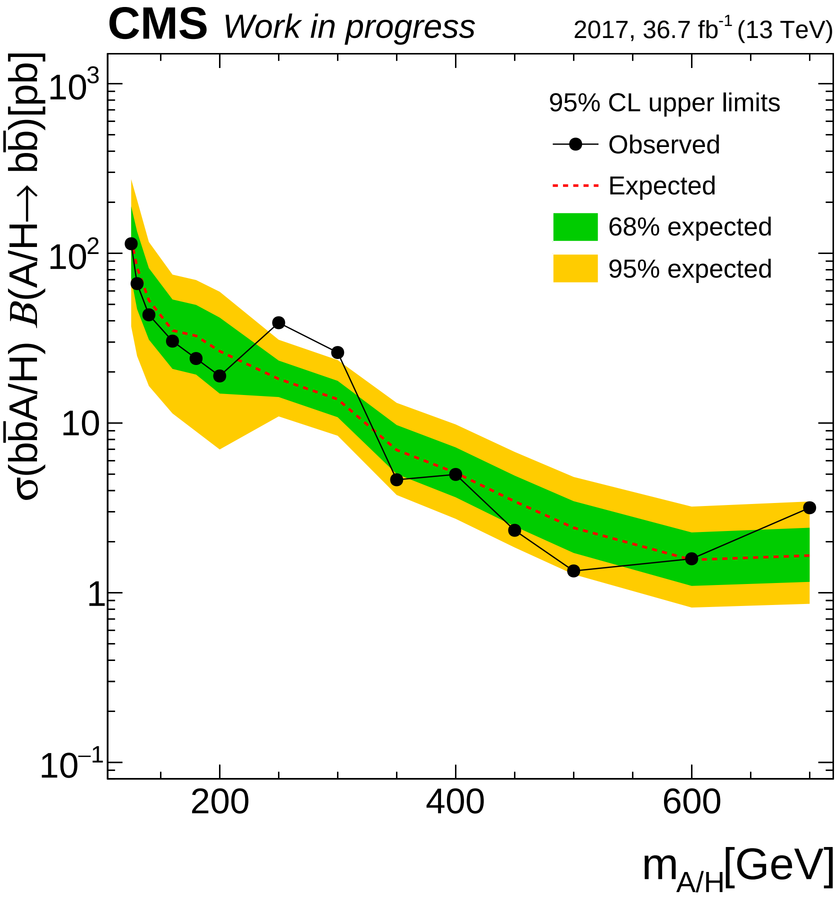


Figure 6.21: Expected and observed upper limits for the Higgs b-associated production cross-section times branching fraction of the decay into a b-quark pair at 95% CL as a function of M_{12} for the 2017 SL category. The green (yellow) bands correspond to $\pm 1\sigma$ ($\pm 2\sigma$) bands.

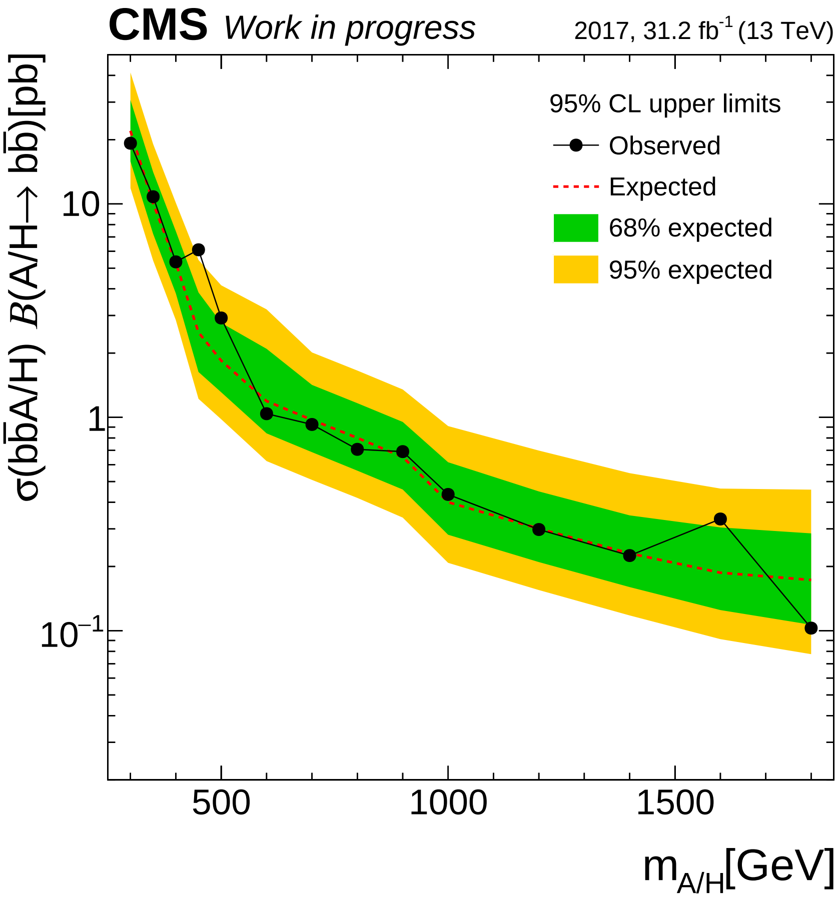


Figure 6.22: Expected and observed upper limits for the Higgs b-associated production cross-section times branching fraction of the decay into a b-quark pair at 95% CL as a function of M_{12} for the 2017 FH category. The green (yellow) bands correspond to $\pm 1\sigma$ ($\pm 2\sigma$) bands.

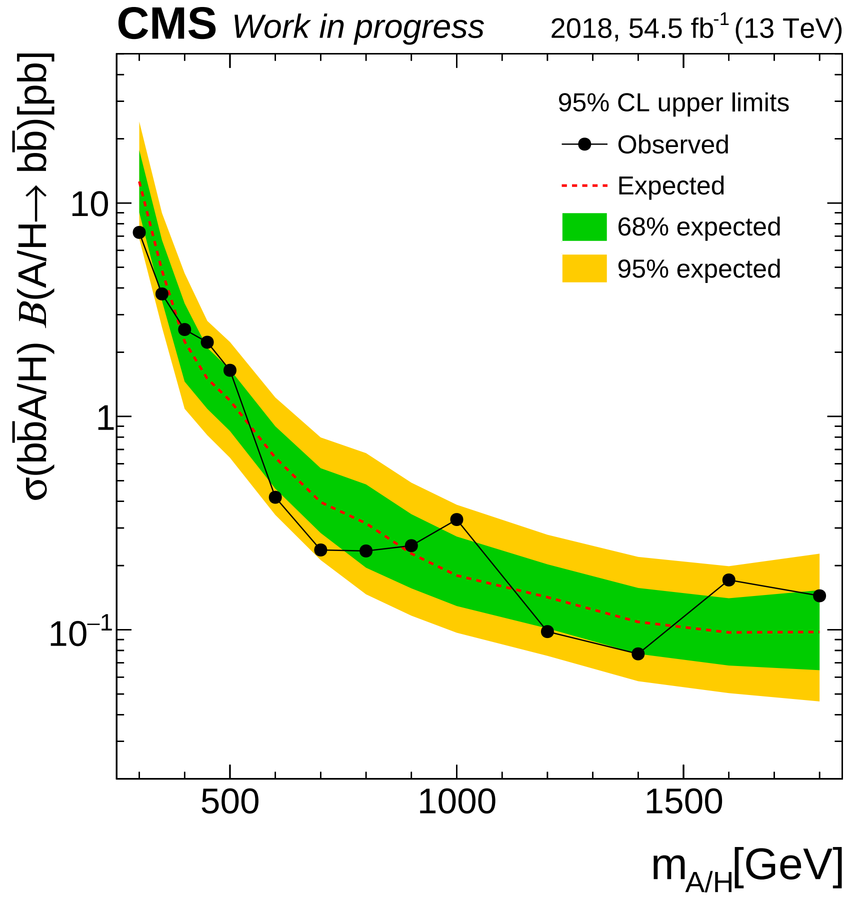


Figure 6.23: Expected and observed upper limits for the Higgs b-associated production cross-section times branching fraction of the decay into a b-quark pair at 95% CL as a function of M_{12} for the 2018 FH category. The green (yellow) bands correspond to $\pm1\sigma$ ($\pm2\sigma$) bands.

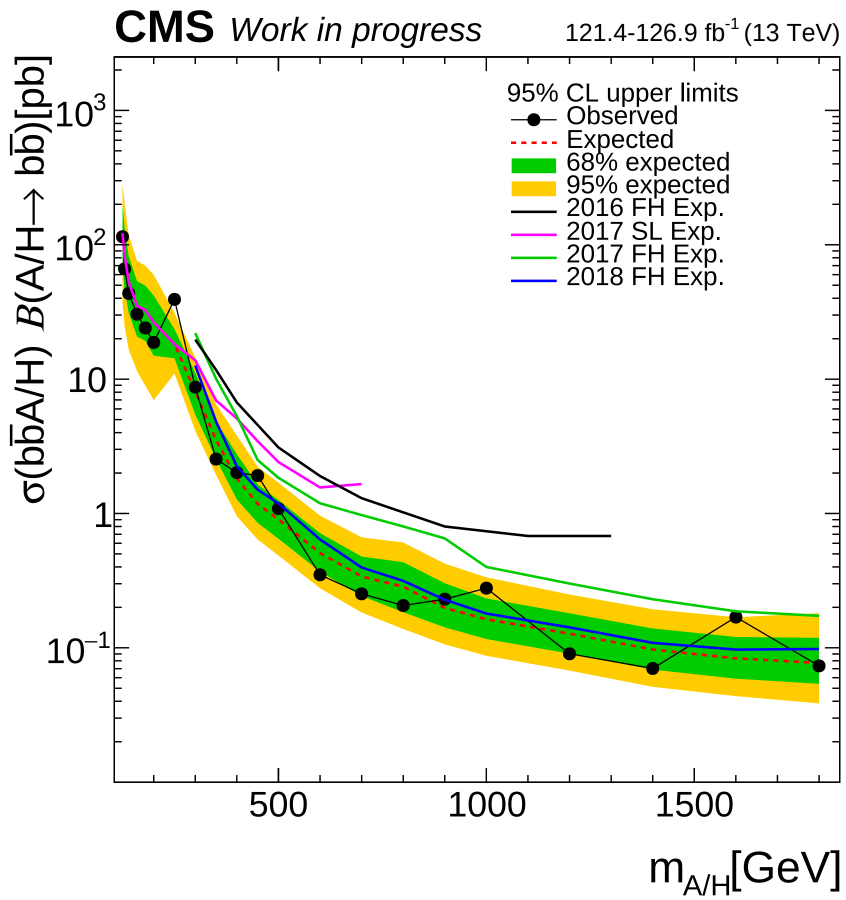


Figure 6.24: Expected and observed upper limits for the Higgs b-associated production cross-section times branching fraction of the decay into a b-quark pair at 95% CL as a function of M_{12}, corresponding to the Run 2 combination. The green (yellow) bands correspond to $\pm1\sigma$ ($\pm2\sigma$) bands. The contributions of the individual analyses are also shown for comparison.

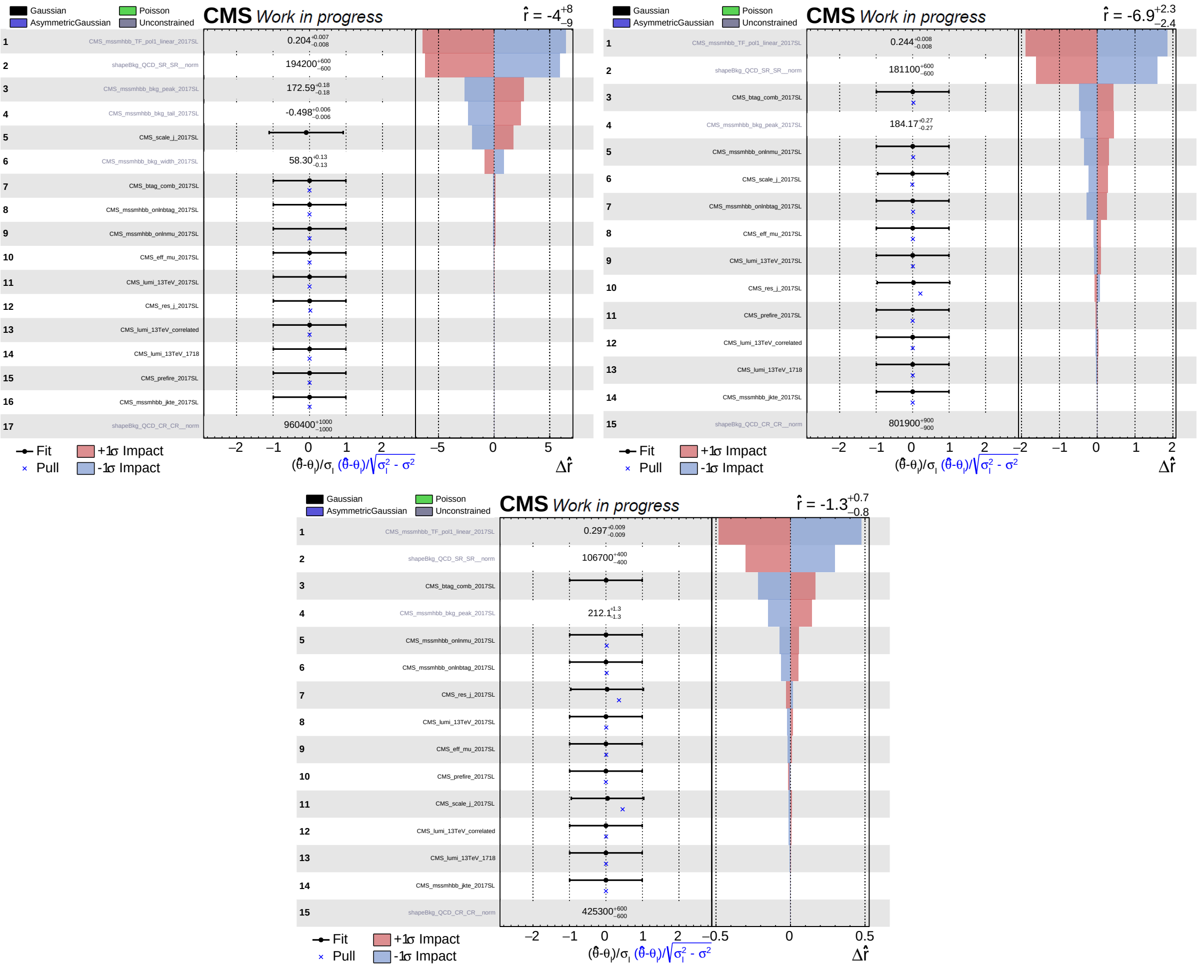


Figure 6.25: Pulls and impacts (SR data), corresponding to the three representative mass points of the 2017 SL analysis: 180, 400, and 600 GeV (top left to bottom) in their respective fit ranges FR1, FR2 and FR3.

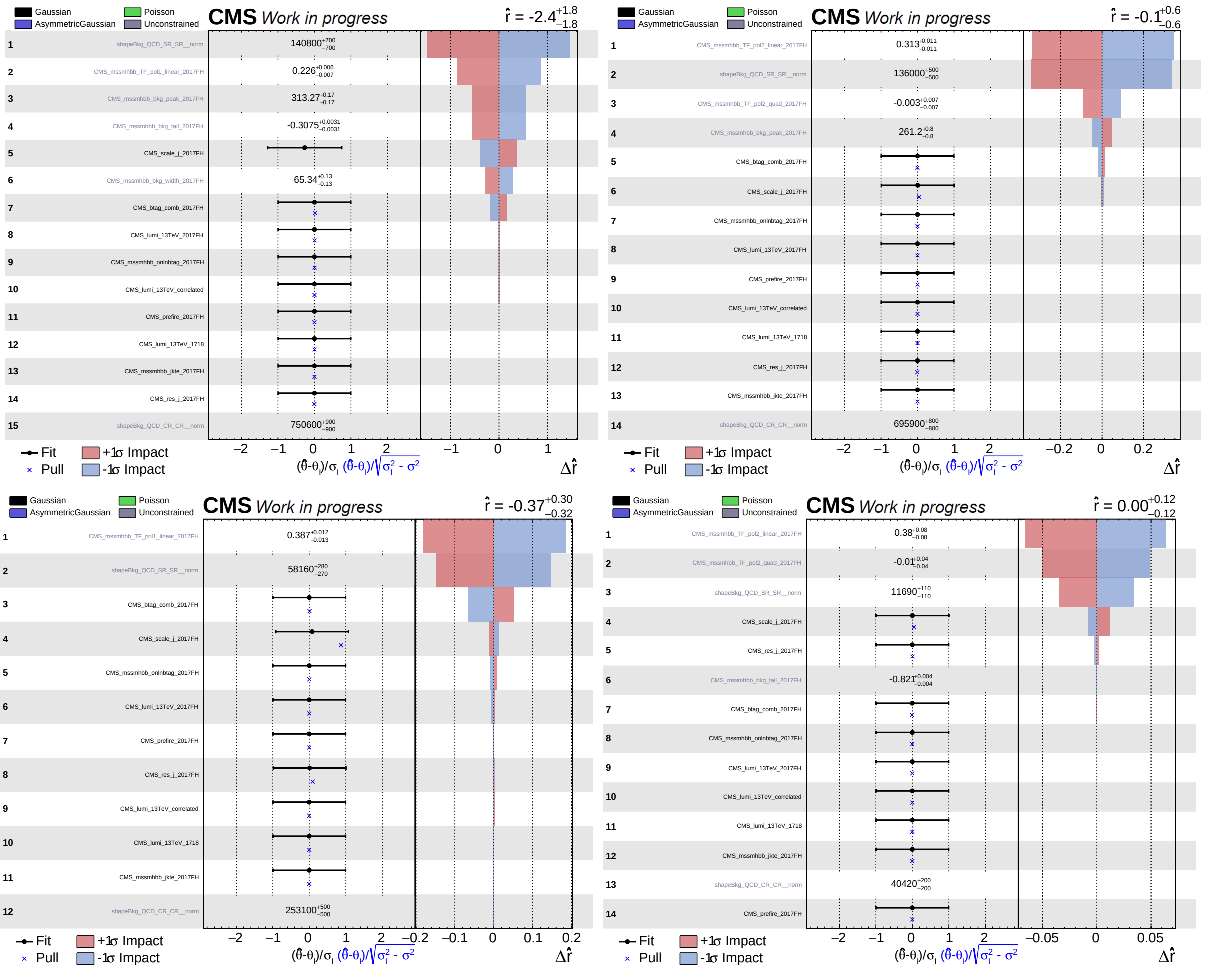


Figure 6.26: Pulls and impacts (SR data), corresponding to the four representative mass points of the 2017 FH analysis: 400, 600, 800 and 1400 GeV (top left to bottom) in their respective fit ranges FR1, FR2, FR3 and FR4.

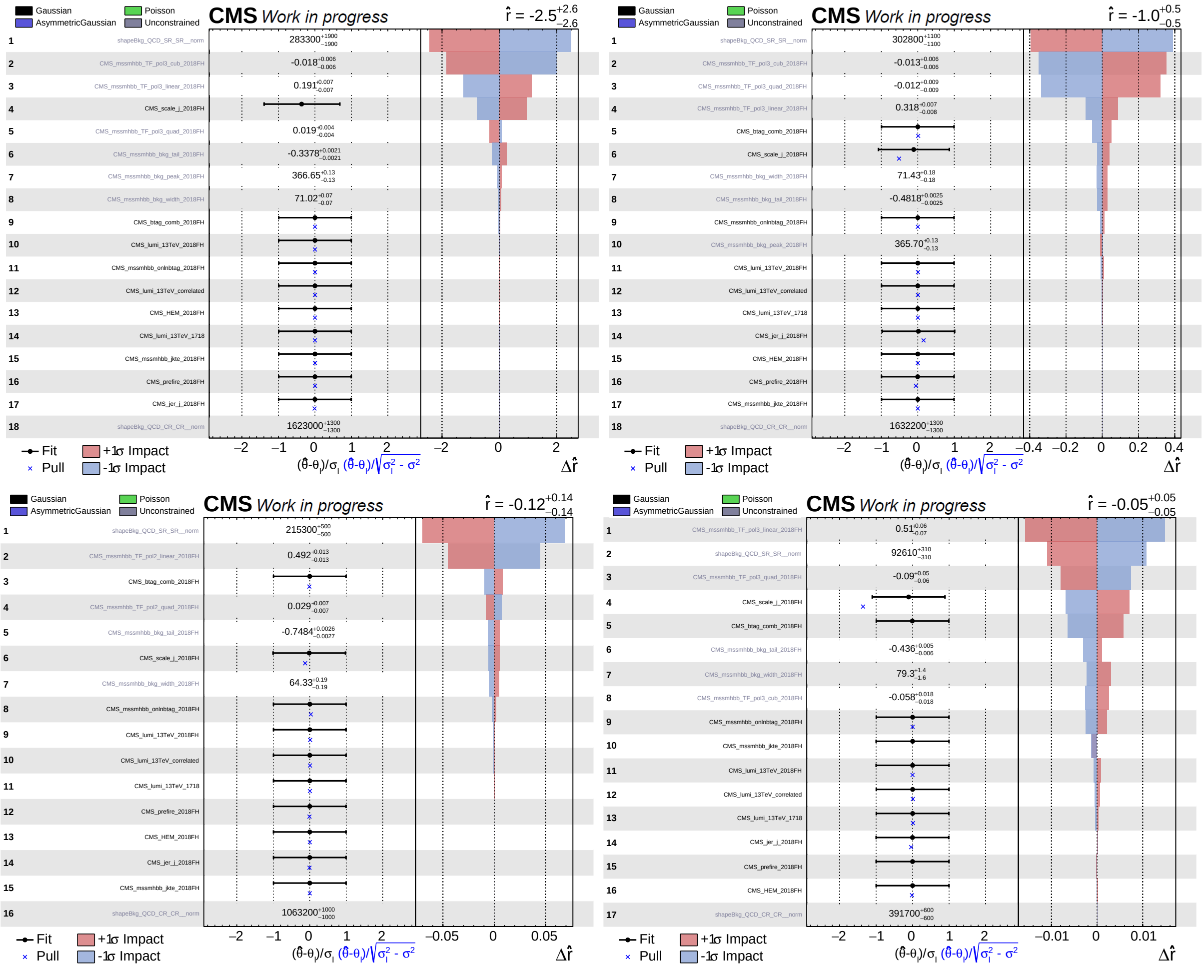


Figure 6.27: Pulls and impacts (SR data), corresponding to the four representative mass points of the 2018 FH analysis: 400, 600, 800, and 1400 GeV (top left to bottom) in their respective fit ranges FR1, FR2, FR3 and FR4.

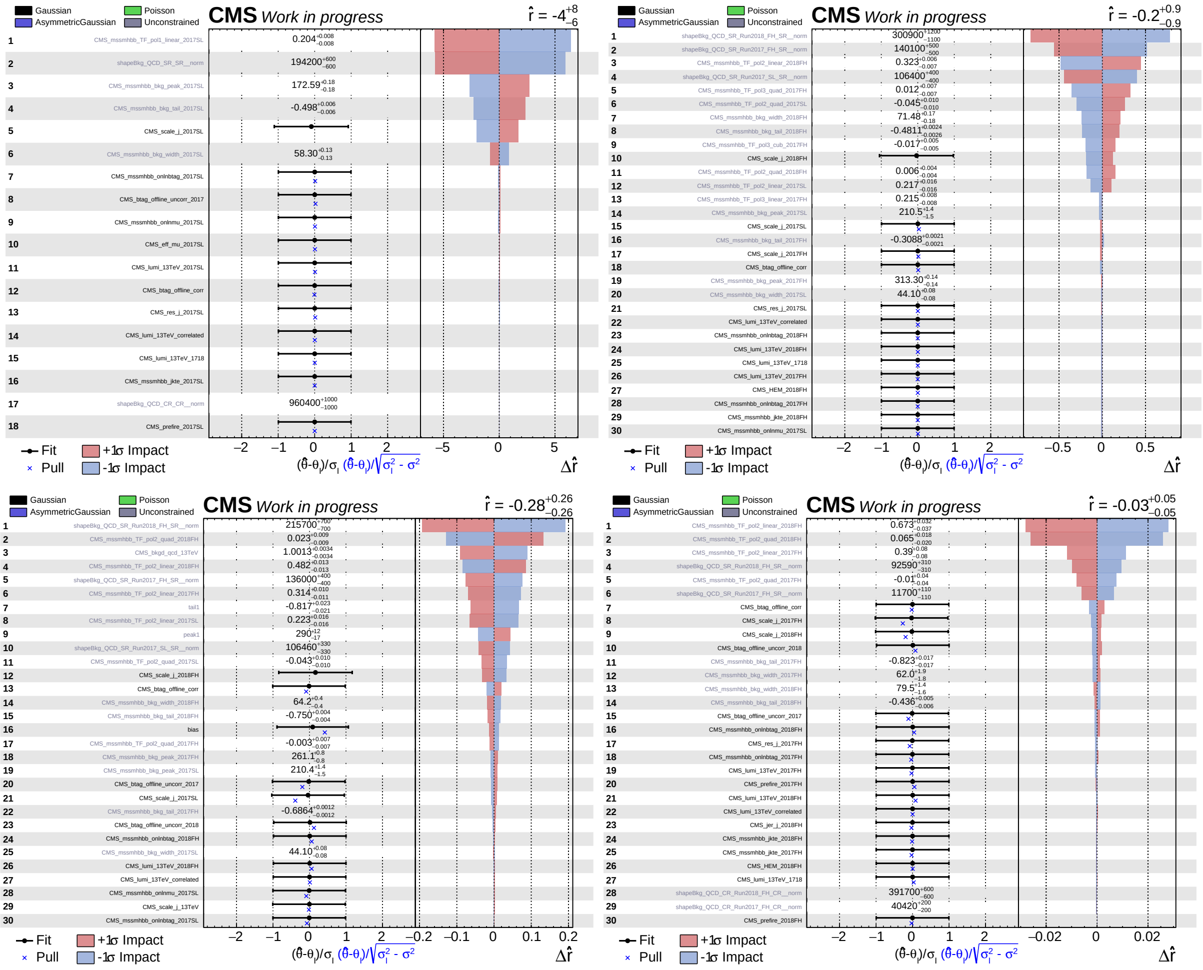


Figure 6.28: Pulls and impacts (SR data), corresponding to the four representative mass points of the Run 2 Combination: 180, 400, 600, and 1400 GeV (top left to bottom).

6.3 Interpretation of the results in the context of the MSSM scenarios

The UL in the production cross section times branching ratio are interpreted in the MSSM benchmark scenarios discussed in 2.2.1, namely m_{h}^{125} scenarios considering both positive and negative Higgsino mass parameter μ, and the $m_{\mathrm{h}}^{\mathrm{mod+}}$ and hMSSM scenarios where a default $\mu = 200$ GeV is selected. For each benchmark scenario, the model-dependent b-associated production cross-sections were calculated centrally by the LHC Higgs Cross Section Working Group (LHCHXSWG) [84] for each of the three neutral Higgs bosons in the tanβ and m_{A} parameter space, according to the 4FS (5FS) prescriptions at NLO [96, 220] (NNLO [97]). As discussed in section 2.2.1, the predictions of both schemes are harmonized in the total cross-section according to the Santander matching scheme [92] in the hMSSM and the $m_{\mathrm{h}}^{\mathrm{mod+}}$ scenarios, while for the m_{h}^{125} scenarios the FONLL [93, 94] was employed. Branching ratios to the $b\bar{b}$ decay channel were determined centrally using the *FeynHiggs* framework for all scenarios except the hMSSM. For the hMSSM scenario, the *HDecay* framework was employed.

The UL from Fig. 6.24 are mapped to those of the model-dependent calculations to translate them into limits on the tanβ vs. m_{A} plane. As mentioned in section 2.2.1, the uncertainties related to the value of α_s, the choice of parton distribution functions, and the uncertainty in the normalisation and factorization scales were centrally provided by the LHC Higgs Cross-Section Working Group (LHCHXSWG) [84]. These uncertainties affect only the normalization of the signal mass and are treated as fully correlated across years and categories.

Figure 6.29 shows the interpretation results in the m_{h}^{125} scenario, where a $\mu = 1$ TeV is considered. These limits display a structure at 180 GeV and 300 GeV, reflecting the regions where the background peaks in the SL and FH categories. The exclusion ranges from tanβ values of 10.6 at m_A= 160 GeV to 50.4 at about 1 TeV. The $b\bar{b}$ decay channel is particularly sensitive to negative values of the Higgsino mass parameter, since it directly impacts the bottom Yukawa-coupling correction Δ_b, as discussed in section 2.2.1. Interpretations in m_{h}^{125} scenarios considering a negative Higgsino mass parameter $\mu = -1, -2$ and -3 TeV are shown in figure 6.30. The limits extend down to 8.5 at 160 GeV and to 51.5 at 1800 GeV in the $\mu = -1$TeV case. The hashed region representing the parameter space incompatible with the observed Higgs boson mass of 125 ± 3 GeV due to theoretical constraints increases with the absolute value of μ in these scenarios. In the $\mu = -2(-3)$ TeV scenario, tanβ values of 7.9 (7.3) are excluded at 160 GeV, ranging up to nearly 30 (20) in the allowed parameter space, at about 1400 (1200) GeV. These are the most stringent limits to date in tanβ set from a search of heavy Higgs bosons in final states with bottom quarks.

Figure 6.31 displays the results in the hMSSM and $m_{\mathrm{h}}^{\mathrm{mod+}}$ scenarios, compared to the latest CMS publication on the search for BSM Higgs bosons in the same final state

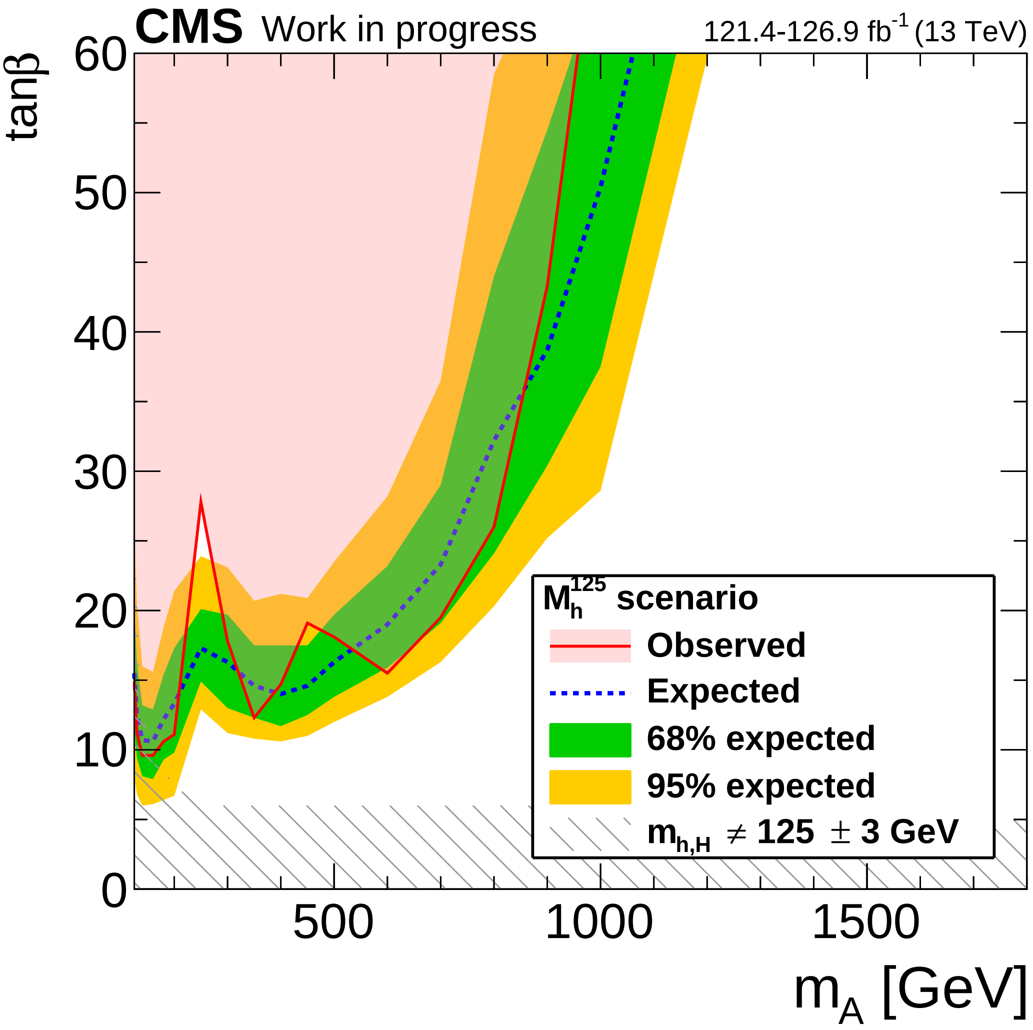


Figure 6.29: Observed and expected upper limits on tanβ as a function of the pseudoscalar mass m_A in the m_{h}^{125} benchmark scenario, considering $\mu = 1\mathrm{TeV}$. The hashed region indicates the parameter space excluded due to incompatibility with the observed 125 GeV Higgs boson within a range of 3 GeV.

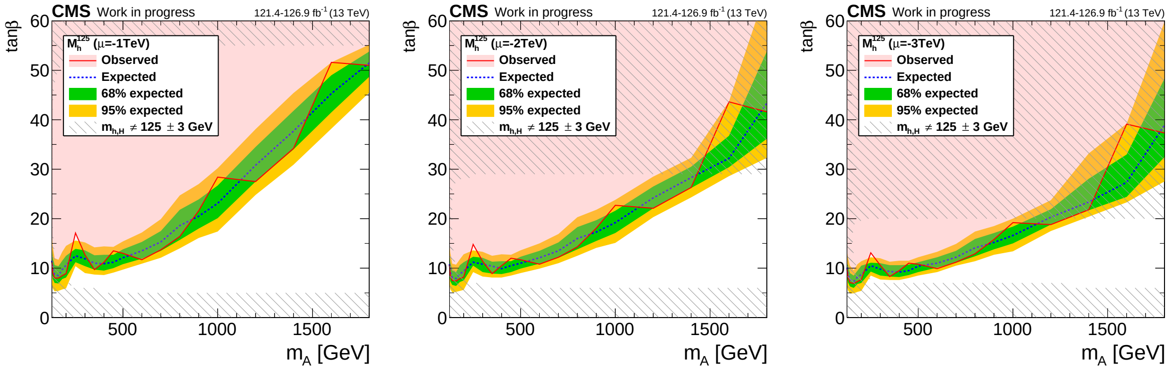


Figure 6.30: Observed and expected upper limits on tanβ as a function of the pseudoscalar mass m_A in the m_{h}^{125} benchmark scenario, considering $\mu = -1$ TeV (left) $\mu = -2$ TeV (center) and $\mu = -3$ TeV (right).

2016-only data. The improvement with respect to previous results is significant. Values of tanβ can be excluded down to 9.6 (10.10) at $m_A = 160$ GeV, and up to tan$\beta = 60$ for masses nearly up to 1300 GeV (1100 GeV) in the hMSSM ($m_{\mathrm{h}}^{\mathrm{mod}+}$) scenario. The limits in the hMSSM scenario are more stringent because the masses of SUSY particles in this model are large enough that the Higgs decay rate is almost unaffected by them, even at high Higgs masses.

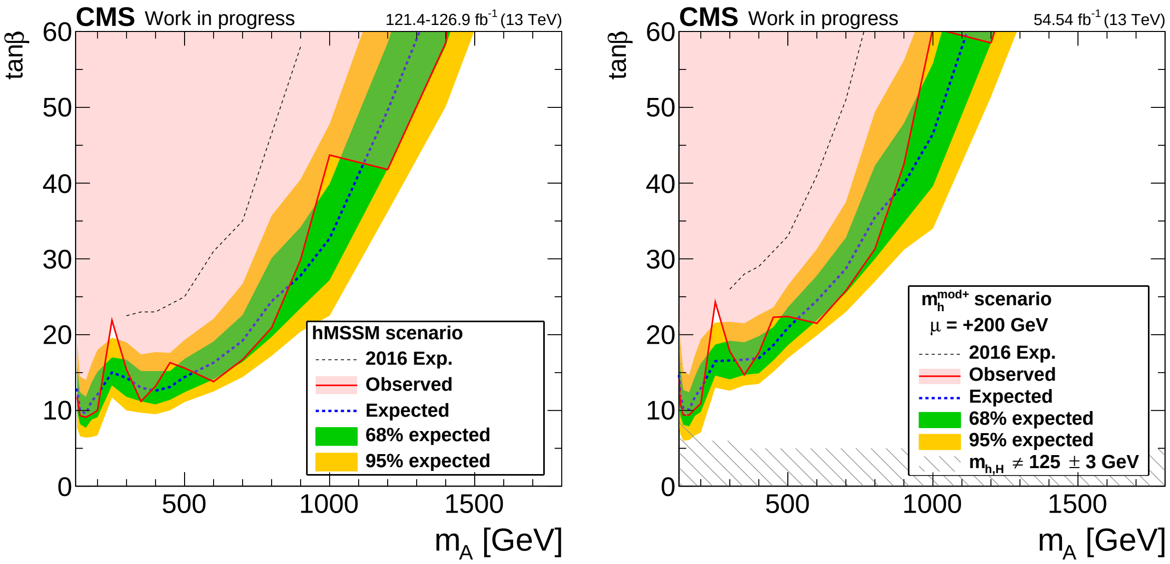


Figure 6.31: Observed and expected upper limits on tanβ as a function of the pseudoscalar mass m_A in the hMSSM (left) and $m_{\mathrm{h}}^{\mathrm{mod}+}$ benchmark scenarios.

6.4 Interpretation within 2HDM scenarios

Exclusion limits in 2HDM scenarios are also obtained by mapping the model-independent cross-sections to that of the relevant benchmark scenarios, computed as a function of their free parameters. As discussed in section 2.2.2, 2HDM models feature a large number of free parameters, which can be reduced down to three parameters according to the so-called scenario-G [106], namely m_A, tanβ and cos$(\beta-\alpha)$. The interpretations are performed for the benchmark scenarios where the coupling of the Higgs to b-quarks is enhanced, i.e. the *type-II* and *flipped* benchmark scenarios. One parameter is fixed, and limits are provided in the other two. The methodologies to compute the cross-sections and branching rations employed were discussed in section 2.2.2.

Figures 6.32 and 6.33 show the limits in the tanβ and cos$(\beta-\alpha)$ parameter space in the *type-II* and *flipped*, respectively, for a fixed value of m_A = 140, 300, 600 and 1200 GeV. Values of cos$(\beta-\alpha)$ further from the alignment limit display a visible deterioration on the limits that increases with mass. This is due to the increase in the decay width of bosonic decay channels such as $A \to Zh$ and $H \to hh$. Near the alignment limit, the exclusion limits are slightly asymmetrical. An observed asymmetry in the branching ratio of the CP-even H boson, shown in Appendix G, causes this. The limits in both benchmark scenarios are rather similar, and tanβ can be excluded down to slightly less than 15 at 140 GeV, nearly 17 at 300 GeV, 18 at 600 GeV, 30 at 900 GeV and 50 at 1200 GeV.

By fixing cos$(\beta-\alpha)$ to 0.1, limits can be obtained for tanβ as a function of m_{A}. This value is selected considering that it is not excluded by SM Higgs measurements, as shown in Fig. 2.6. The results are shown in Fig. 6.34. This analysis can exclude the mass range from 125 GeV up to nearly 1500 GeV, where tanβ = 100 is excluded. This represents a remarkable improvement with respect to the 2016-only results, also shown for comparison in both plots. The lowest tanβ excluded is nearly 9 at 160 GeV in both scenarios considered. As in the MSSM interpretations, these are the most stringent limits to date derived from a search of additional Higgs bosons in final states with b-quarks.

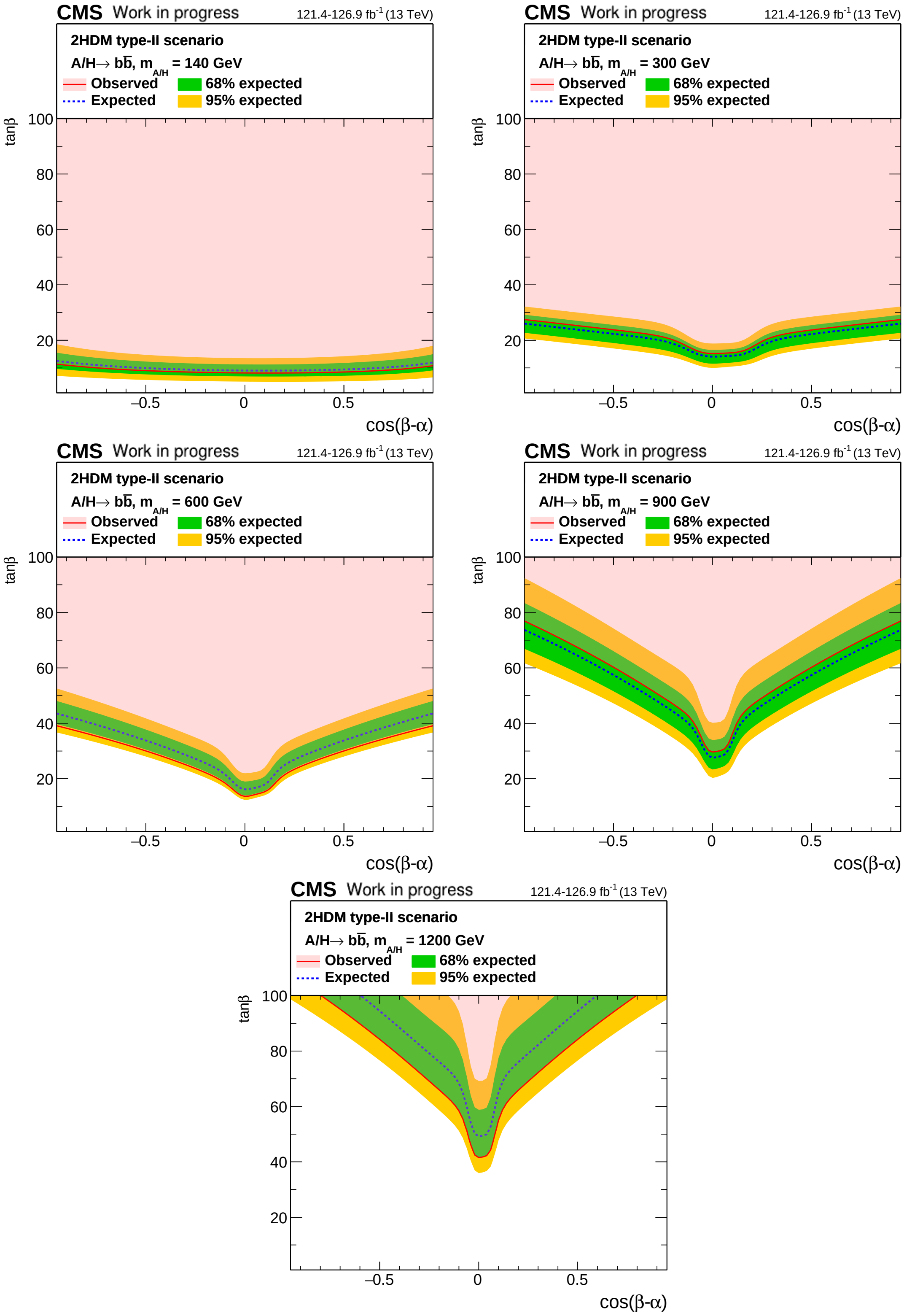


Figure 6.32: Observed and expected upper limits on tanβ as a function of $\cos(\beta-\alpha)$ in the 2HDM *type-II* benchmark scenario, corresponding to a pseudoscalar mass $m_A = 140 GeV$ (top left), 300 GeV (top right), 600 GeV (centre left), 900 GeV (centre right), and 1200 GeV (bottom).

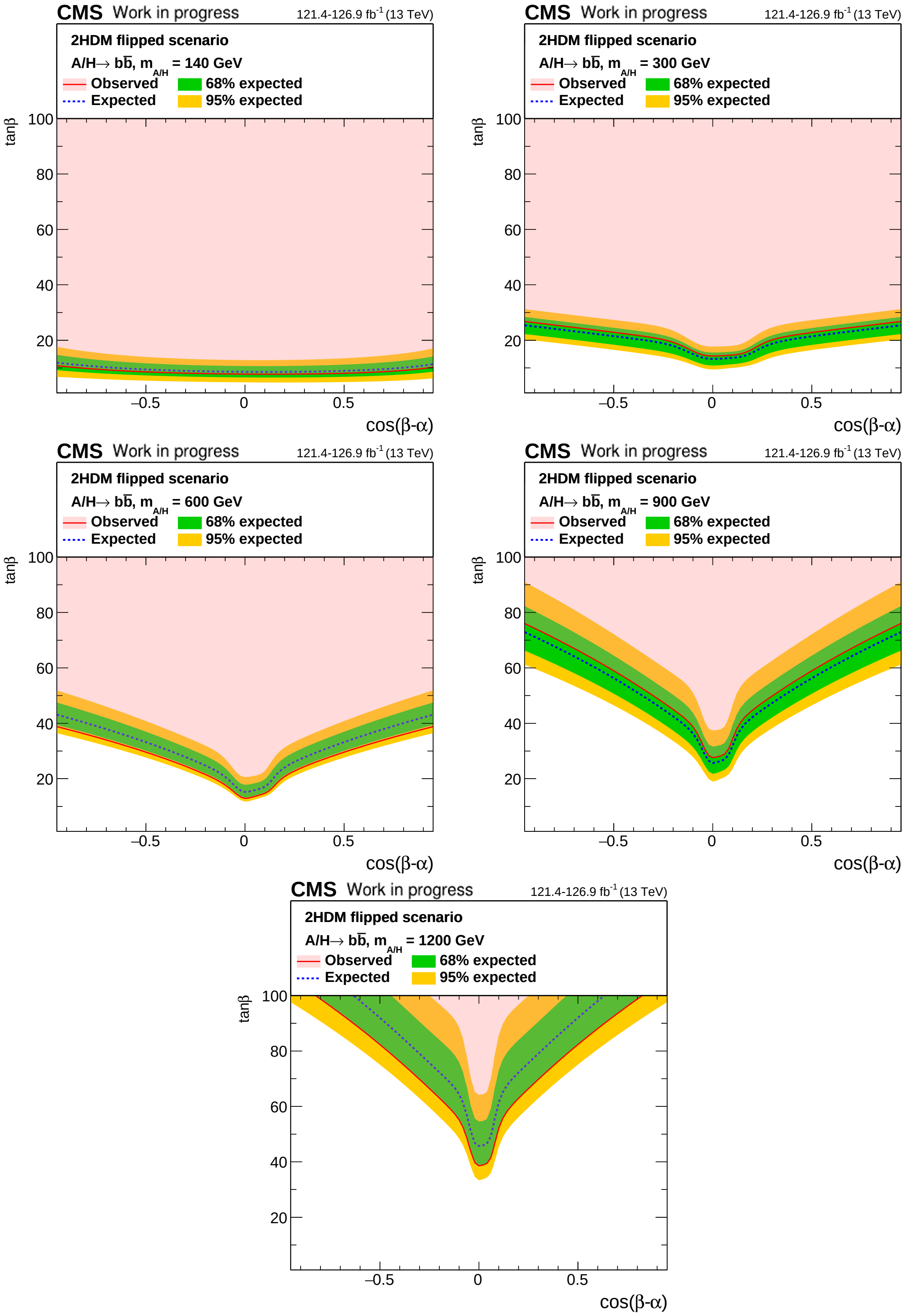


Figure 6.33: Observed and expected upper limits on tanβ as a function of cos$(\beta-\alpha)$ in the 2HDM *flipped* benchmark scenario, corresponding to a pseudoscalar mass $m_A = 140 GeV$ (top left), 300 GeV (top right), 600 GeV (centre left), 900 GeV (centre right), and 1200 GeV (bottom).

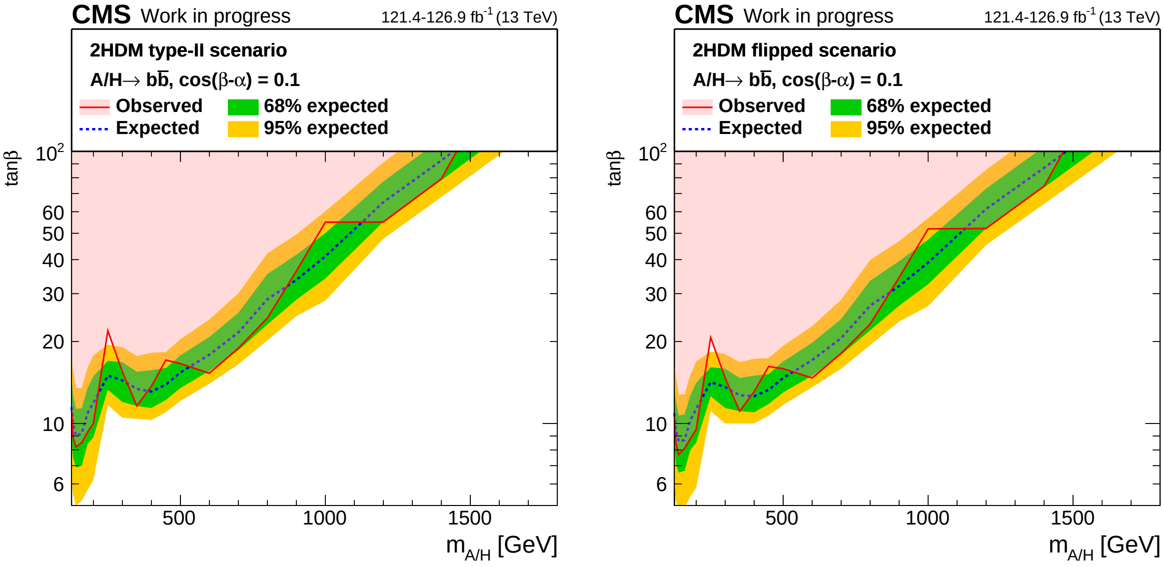


Figure 6.34: Observed and expected upper limits on tanβ as a function of m_{A} in the 2HDM *type-II* (left) and *flipped* (right), corresponding to $\cos(\beta-\alpha) = 0.1$.

chapter 7

Summary and Outlook

A search for BSM Higgs bosons in the b-quark decay channel and produced in association with b-quarks was conducted. Run 2 CMS data collected during p-p collisions was scrutinized. The 2017 and 2018 datasets were analyzed in the fully-hadronic (FH) category, corresponding to 31.2 and 54.5 fb^{-1} of data, respectively. This category targeted hypothetical masses from 300 GeV to 1800 GeV. The 2017 data was also analyzed in the semi-leptonic (SL) category, corresponding to 36.7 fb^{-1} of data, where masses from 125 GeV to 700 GeV were probed. Due to trigger validity, there was no 2018 SL analysis. My efforts were mainly focused on the comprehensive analysis of the 2018 data in the FH category and the derivation of key elements for the 2017 SL and FH analyses, as outlined below.

The Higgs bosons searched in this work are predicted in theories with an extended Higgs sector, such as the 2HDM, or the MSSM, which is a minimal realization of SUSY. An enhancement of the coupling to b-quarks is foreseen in certain scenarios within these models, particularly at large tanβ values, thus enhancing the sensitivity of a search in final states with b-quarks. A signal from a BSM Higgs is expected to rise in the invariant mass distribution of the two most energetic b-jets, which is ultimately scrutinized.

The search is affected by a large QCD multi-jet activity at the LHC, resulting in a considerable background in the signal region. Targeting the b-associated production by selecting final states with at least three b-quarks contributes to a partial background reduction. Still, this analysis requires dedicated triggers that are designed to enhance the analysis sensitivity while meeting the trigger bandwidth constraints. These triggers select events with at least two b-tagged high p_{T} jets in the FH category. The SL trigger has a similar structure but requests a muon within a jet, allowing to lower the p_{T} thresholds.

After the trigger selection, an offline selection is performed, designed for signal enhancement and background reduction. Object reconstructed offline must be matched to online objects to guarantee a proper application of online trigger scale factors that correct the simulated samples. My contribution included deriving these scale factors for all the data-taking years and categories. These were obtained using a data-driven approach with the tag-and-probe method.

The present work revised the strategy of matching offline b-jet candidates to online b-tagged objects since a substantial inefficiency was present in events where one of the two most energetic jets was not tagged by the trigger, but rather, the quite softer third jet passed the selection. This occurs more prominently at large p_{T} due to the expected deterioration of the b-tagging performance. I proposed a new strategy to recover those events, resulting in a substantial improvement in the signal efficiency, which determined a considerable gain in the significance of the search.

An offline b-tag selection determines the regions of the analysis. A signal region (SR) is defined when the three leading jets in an event pass the b-tag requirement. The main observable of the search is not viewed in this region as per a standard blinding policy until all analysis methods are fully developed and validated. A control region (CR), which is employed in the background modelling, is defined with the condition that the third most energetic jet fails the b-tagging requirement. Between the two regions, a validation region (VR) is defined. The VR is utilized for the final commissioning of the methodology before unblinding the SR.

Signal events are modelled in several hypothetical Higgs boson masses, ranging from 125 to 1800 GeV. They are simulated at NLO with the `POWHEG 2.0` generator, interfaced with `PYTHIA8`. The signal distributions are parametrized with the double-sided Crystal Ball.

Background events originate primarily from QCD processes, which are challenging to simulate accurately. Therefore, special attention is given to the background modelling. The background in the SR is modelled with the aforementioned CR multiplied by a transfer factor (TF), which maps the CR distribution into the SR. The functional form of the TF is obtained from the ratio of SR and CR in simulated samples enriched with b-jets, although the parameters are allowed to float freely for the final fit. My efforts included deriving the TF for all the years and categories of the presented work. The background is then modelled in ranges of the invariant mass distribution, optimized for the best and simplest description of the data in the CR.

A simultaneous fit of the signal plus background models is performed to the CR and the SR to extract a potential signal in the latter. The COMBINE TOOL package [214,218] was employed for this purpose. As no considerable excess was observed with respect to the background-only hypothesis, upper limits were set in the production cross-section times branching ratio of the process under study at 95% confidence level for the individual years and categories. I performed the statistical inference for the 2018 FH analysis and, with the acquired expertise in the COMBINE TOOL package, assisted in the 2017 SL and 2017 FH analyses. Additionally, I performed the statistical combination of the aforementioned analyses with those from a previous partial Run 2 CMS search, where only 2016 data was analysed [35]. As a result, 123 fb^{-1} of data were analysed in the full Run 2 combination. The final limits extend from 123 *pb* at 125 GeV down to 0.77 *pb* 1800 GeV, representing a considerable improvement beyond the increase in luminosity with respect to 2016-only results.

I performed the interpretation of these results into constraints in the parameter space of MSSM scenarios, i.e. tanβ and m_A, and in the parameter space of 2HDM *type-II* and *flipped* scenarios, i.e. tanβ, cos$(\beta - \alpha)$, and m_A.

The MSSM scenarios considered were m_h^{125} with higgsino mass parameter hypothesis, μ, set at -3, -2, -1 and 1 TeV, and hMSSM and $m_\mathrm{h}^{\mathrm{mod}+}$, which set this parameter at 200 GeV. A wide range of the parameter space is excluded in all cases, going down to about tanβ ~10 towards lower masses. The tightest constraints are set in scenarios with μ = -3, excluding the tanβ region from 7.3 to 20 in the explored mass range that is consistent with the 125 GeV Higgs boson observation. The constraints in hMSSM and $m_\mathrm{h}^{\mathrm{mod}+}$ scenarios are compared with the 2016-only. The Run 2 combination is able to exclude a considerably larger parameter space, ranging from tan$\beta \approx 10$ at 160 GeV and up to tan$\beta = 60$ at nearly 1300 GeV.

In 2HDM models, a large improvement in the excluded region is also observed with respect to the 2016-only results. The tanβ vs. cos$(\beta - \alpha)$ plane is explored for masses up to 1200 GeV. In the region close to the alignment limit, tanβ values are excluded down to approximately 15 at 140 GeV, and down to 50 at 1200 GeV. The tanβ vs. m_A plane is also excluded from nearly tan$\beta = 9$ at 160 GeV up to tan$\beta = 100$ at 1500 GeV.

The successful LHC data-taking during the Run 2 period at the large centre-of-mass energy of 13 TeV allowed to considerably extend the limits in searches for BSM physics. Particularly, the search presented in this thesis has set the most stringent limits thus far in the search for additional Higgs bosons in final states with b-quarks. These limits are unique for 2HDM models in the *flipped* scenario and particularly sensitive for MSSM scenarios with negative μ. These results have been pre-approved by the CMS collaboration and are close to publication.

At the time of writing, the Run 3 data-taking is ongoing. As a result, the current dataset is expected to be doubled in size. The triggers dedicated to extending the analysis presented in this work with Run 3 data were developed within the scope of this thesis and have been successfully taking data since the beginning of Run 3. Furthermore, after Run 3, the LHC will enter the high-luminosity period and is expected to increase the total integrated luminosity to more than 3000 fb^{-1} of p-p collision data at 14 TeV by 2038, which is expected to significantly enhance the sensitivity of this search in the future.

appendix A

Alignment of the CMS Tracker

An incorrect tracker geometry biases the track reconstruction. Therefore, to guarantee excellent performance of the CMS tracker, a precise knowledge of the position and orientation of its sensors is critical, as mentioned in section 3.2.1. The desired precision must be comparable to the design hit resolution of the silicon sensors, of the order of 10 μm. However, a mechanical alignment can only yield a precision that is 10 times larger [139]. Moreover, systematic movements are anticipated due to changes in the operational conditions, such as a ramping down of the CMS magnet, followed by a ramp-up (magnet cycles), temperature changes, and radiational damage. Therefore, corrections to the position and orientation of the sensors, also known as “tracker alignment” must be derived. This chapter describes the methodologies and strategies followed in the tracker alignment process in preparation for and during Run 3.

A.1 Local and global coordinate system

Complementary to the CMS global coordinate system defined in section 3.2, denoted as (x, y, z), a local reference a local reference system is defined for each module. It is denoted as (x', y', z'), and defined with the origin centred in the active region of the module, the y-axis perpendicular to the LHC-plane, and the z-axis is set anti-clockwise beam direction. The angles α, β, and γ indicate rotations over the x', y', and z' axes, respectively [221].

A.2 Track-based alignment

To derive the alignment corrections, the CMS collaboration follows a “track-based alignment” approach that employs a large set of reconstructed charged particles’ trajectories [221]. This method relies on the fact that misaligned modules would pull away the measured hit positions from the fitted trajectory. The residuals between the two are thus a measurement of the misalignment, as illustrated in Fig. A.1. The true module's position and orientation, also known as alignment parameters ($\mathbf{p}$), are derived by minimizing the sum of squares of the normalized track-hit residuals:

$$\chi^2(\mathbf{p},\mathbf{q}) = \sum_j^{\text{tracks}} \sum_i^{\text{hits}} \left(\frac{m_{ij} - f_{ij}(\mathbf{p},\mathbf{q}_i)}{\sigma_{ij}} \right)^2 , \tag{A.1}$$

where $\mathbf{q}$ accounts for track-related parameters, for instance, its curvature or deflection. The parameters m and f account for the measurements and predictions of a hit, while σ_{ij} represents the uncertainty on the measured hit.

Figure A.1: Schematic representation of the fitted track prediction according to the alignment of the tracker modules. The local hit positions are constant, and only the global position will change according to the alignment of the modules. The increase in the track-hit residuals due to misaligned modules is also illustrated [222].

Assuming a set of initial track and alignment parameters, $\mathbf{q_0}$ and $\mathbf{p_0}$, that can be estimated for instance, from prior alignment studies, eq. A.1 can be adjusted, and a linear system must be solved to obtain the alignment parameters [139]. Nearly $\mathcal{O}(10^5)$ parameters must be determined, a number that might be even higher, depending on the number of tracks used for the alignment study and the complexity of the detector geometry to be aligned. This computationally difficult task is performed within CMS with the `MillePede-II` [223, 224] software, which offers either an exact or a numerical solution to the aforementioned problem.

A.3 Weak modes

An additional major challenge in the track-based alignment arises due to the fact that the χ^2 may remain unchanged by a set of linear combinations of the alignment parameters, known as "weak modes". The weak modes occur when coherent changes in the alignment parameters are compensated by changes in the track parameters due to the inherent symmetry of the tracker system, particularly in collision events with tracks passing through the beam line [225]. The presence of weak modes introduces systematic distortions that may affect physics measurements. To control them, it is crucial to provide additional information to the fit, which is accomplished by employing tracks with different topologies, such as cosmic ray tracks, which traverse the detector, and tracks produced by a resonant decay, like the Z boson decaying into a muon pair.

A.4 Alignment strategies

Hierarchical and differential alignment

The CMS tracker detector has a hierarchical structure that goes from a most general level, comprising large mechanical structures, down to a single module, as shown in Fig. A.2. If the number of available tracks is insufficient to align at a module level, the hierarchical arrangement is exploited, and the tracker's less granular substructures are aligned. This method is referred to as "hierarchical alignment". In the alignment algorithm, the mechanical structures are aligned by introducing six alignment parameters, three translational and three rotational.

The outcome of the alignment procedure can be time-dependent or time-independent. Both conditions are determined in an approach called "differential alignment" [225]. The time-dependent parameters are valid exclusively in a time period called "interval of validity" (IOV) and must be derived with tracks recorded during that IOV.

The time-dependent determination of the position and orientation of large mechanical structures is called "High-level structures" (HLS) alignment and can be performed in any of the sub-structures shown in Fig. A.2 above the level of single modules, like the pixel half-barrels, or the strips endcap disks. On the other hand, the module-level alignment is performed relative to the position of the HLS.

The time-dependent conditions are interfaced with the outcome from a hierarchical alignment so that the alignment of larger structures varies with time while the sensor's position remains constant relative to the corresponding larger structure.

Regular movements

The definition of an IOV is critical to account for expected movements during data collection. Magnet cycles, for instance, may introduce HLS movements of the order of 1 mm, while a temperature change after ceasing the tracker cooling may result in movements of the order of 10 μm [139].

Movements are also originated by the effect of the highly radiational environment that the CMS tracker is exposed to. This effect can be estimated by determining the drift of

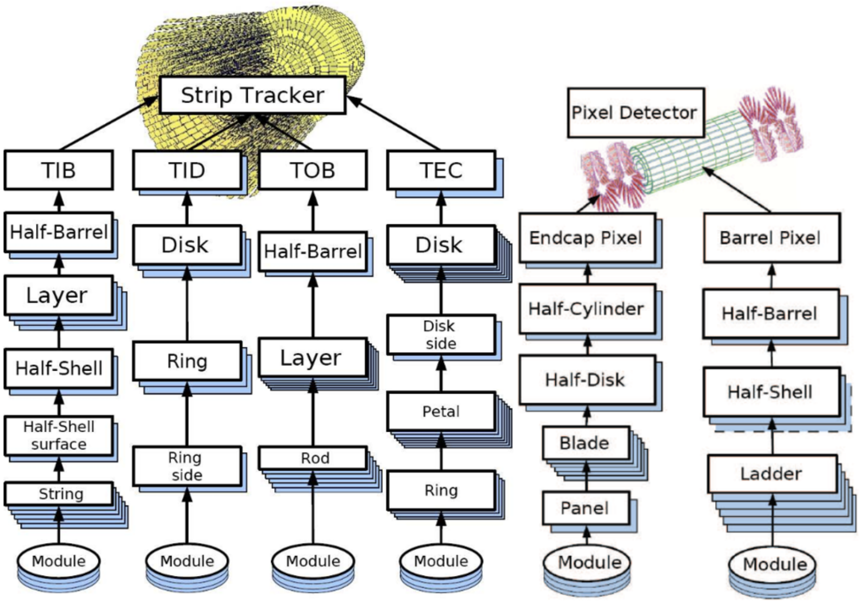


Figure A.2: Schematic representation of the hierarchical structure of the CMS tracker. Taken from [226].

the charge carriers produced by the ionization in the active region of the sensors when moving towards the electrodes, which is induced by the magnetic field. This parameter, also known as the Lorentz drift, changes with the mobility of the charge carriers, which depends on the irradiation dose. Corrections based on the Lorentz drift are determined in dedicated calibrations, which are combined with further alignment corrections.

The alignment parameters need to be updated regularly to account for these effects and possible weak modes. Small movements are corrected during data-taking itself (online) with an automated, streamlined version of the full alignment that runs offline. This automated procedure is complementary to regular offline alignment iterations and can align the pixel HLS at the level of ladders and panels [227].

Datasets used for tracker alignment

Tracks from different datasets are used for tracker alignment. Depending on the availability, several datasets can be simultaneously employed. They include:

- Tracks from events selected with loose trigger conditions aiming to record inelastic collision events with the least possible trigger bias [228], also known as minimum bias events.

- Tracks from isolated muons produced in W bosons' decays.

- Tracks from di-muon resonances such as the Z boson or the Υ meson.
- Cosmic ray tracks, collected with either the CMS magnet on or off. These datasets are known as to as Cosmic RUn at ZEro Tesla (CRUZET) and Cosmic Run At Four Tesla (CRAFT).

The event selection criteria of these datasets are discussed in Refs [139] and [225].

A.5 Validation techniques

The derived alignment corrections result in an updated geometry of the tracker's components, whose performance must be validated. Different methods have been implemented for this purpose. They rely on comparing the measurement of certain sensible variables under the alignment condition(s) under scrutiny with the performance in an ideal scenario with perfect alignment. This section outlines a subset of these validations.

It is relevant to note that to avoid possible bias in the results of the validation, it is preferred to use a different set of tracks than the one used for alignment.

Distribution of median track-hit residuals validation

The distribution of median track-hit residuals (DMR) is obtained by refitting each track with the alignment parameters under study. Then, the hit residuals are obtained relative to the track prediction, for which the hit under study is removed, and a histogram is filled. This histogram is ideally centred at zero, and deviations indicate possible bias. Moreover, the width of the distribution measures the precision of the alignment. Given the large number of events available, minimum bias events are generally employed to derive this validation.

Track-vertex residuals validation

This method, also known as primary vertex (PV) validation, scrutinizes the vertex reconstruction performance based on minimum bias data.

All tracks are iterated, and the PV is reconstructed without considering the track under scrutiny. The distance between the PV and the track at the point of closest approach is obtained. Ideally, this distribution is equal to zero across the track's coordinates. Deviations from the ideal distribution indicate movements in the pixel detector, which is mostly responsible for the PV reconstruction.

Primary vertex resolution validation

Another way to validate the alignment geometry related to the vertexing performance is by investigating the PV resolution. For this, the set of tracks that are assigned to a vertex is separated into two. The tracks are sorted based on their p_{T}, and paired up starting from the highest p_{T}. Each pair of tracks is then split, with one track randomly assigned to one group and the other track to the other group, so that both groups have similar characteristics. Then, the vertex is reconstructed with both new groups of tracks. The difference between the vertex positions obtained from these two groups is calculated, which gives a measurement of the residual of the PV resolution. The process is repeated for different values of $\sum p_{\mathrm{T}}$ for all tracks in the dataset. The residuals are then analyzed and fitted with a Gaussian function, whose mean value defines the PV resolution. Vertices are categorized based on the total p_{T} assigned to them. This is because the accuracy of the vertex position improves with a higher total p_{T} in the typical range analysed since very low p_{T} tracks are more difficult to reconstruct accurately, as they are more prone to multiple scattering.

Muon track-split validation

The muon track-split (MTS) validation exploits the topology of tracks from cosmic ray muons, which cross the tracker detector. These tracks are split into two at the point of the closest approach to the origin. The two new tracks are refitted with the alignment parameters under scrutiny, and the difference between them is analyzed. The width of the distribution is a measurement of the alignment precision, while the mean deviating from zero suggests a potential bias, as both of the new tracks should be identical. Although relatively uncomplicated, this method is extremely powerful. It is sensitive to systematic misalignments and the presence of weak modes. Since collision tracks are not needed for this validation, this validation method can be exploited at the beginning of data-taking before the startup of the LHC operation.

A.6 Alignment of the CMS tracker during Run 3

The beginning of Run 3 data-taking posed a challenging scenario for the CMS tracker alignment. During the Long Shutdown period that preceded it, the Pixel detector was extracted for a series of refurbishments and upgrades [229]. After re-installation, large movements were expected, and several iterations were performed to mitigate it, as more data became available.

The last tracker geometry derived during Run 2 was taken as a starting point. The first two iterations made use of CRUZET data during 2021. This comprises an early alignment at the level of half-barrels and half-cylinders of the pixel and strip subdetectors, derived from about 120,000 cosmic ray tracks. Shortly after, employing 1,5 million tracks from CRUZET data, a refined alignment (red) was obtained, allowing an increase in the granularity in the barrel pixel detector to the level of single ladders. The refined alignment was critical to recovering the pixel reinstallation misalignments. Fig. A.3 and A.4 compare the performance of the startup geometry to the early alignment and the refined alignment with the DMR and MTS validation. These results were published in Ref. [140].

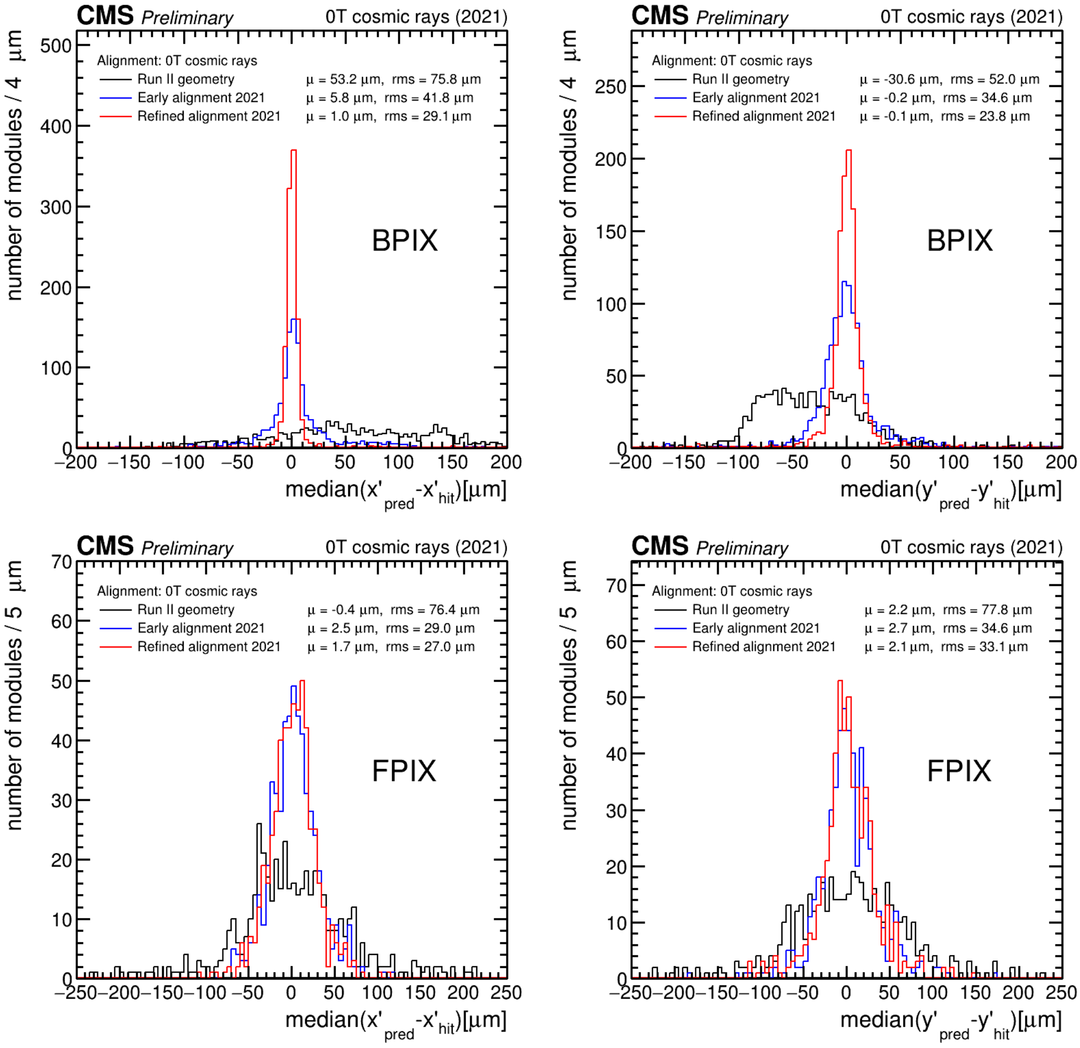


Figure A.3: DMR distributions in the barrel pixel (BPIX) (top), and in the forward pixel (FPIX) (bottom). The startup geometry (black) is compared to the first CRUZET alignment iteration (blue) and the refined alignment with CRUZET data (red). The results along the local x (left) and local y (right) coordinates are shown. A significant improvement was achieved with the alignment campaign with 2021 data in both BPIX and FPIX. The enhanced performance of the refined alignment in BPIX is achieved with the increased granularity in this sub-detector.

Further cosmic ray data was collected after the magnet ramp-up, which was followed by the first collisions of Run 3 at 900 GeV centre-of-mass energy. Both datasets were em-

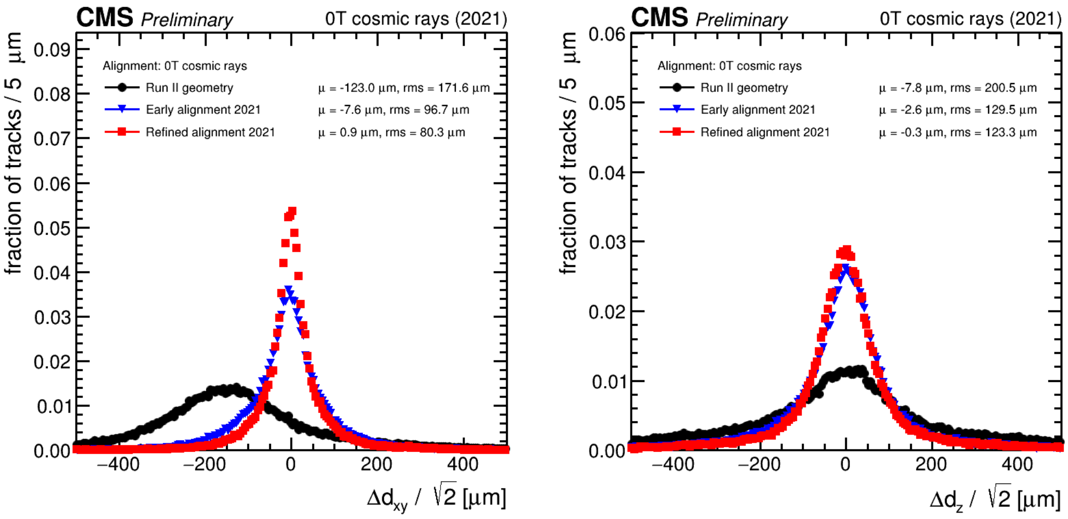


Figure A.4: MTS validation results comparing the startup geometry (black) to the first CRUZET alignment iteration (blue) and the refined alignment with CRUZET data (red). The difference in the track impact parameter in the transverse plane (left) and in the longitudinal direction (right) with respect to the beam pipe is shown. A noticeable improvement was achieved iteratively.

ployed to derive alignment corrections. A subset of validations are shown in Figs. A.5, A.6, and A.7. The latest of the aforementioned geometries derived with CRUZET data was taken as a starting point to derive the following. An alignment was derived using cosmic rays collected at 3.8 T, comprising 765,000 tracks. The corrections were obtained at the level of single modules for BPIX, and at the level of half-barrels and half-cylinders for FPIX and the strips sub-detectors. On the other hand, an alignment with cosmic ray and collision events was derived with 3.6 million cosmic ray tracks and with 255.2 million collision tracks. With this dataset, it was possible to obtain the alignment corrections for the pixel and the strip sub-detectors at the level of single modules, providing a considerable refinement of the tracker alignment corrections for the beginning of data-taking in 2022. These results were published in Ref. [141].

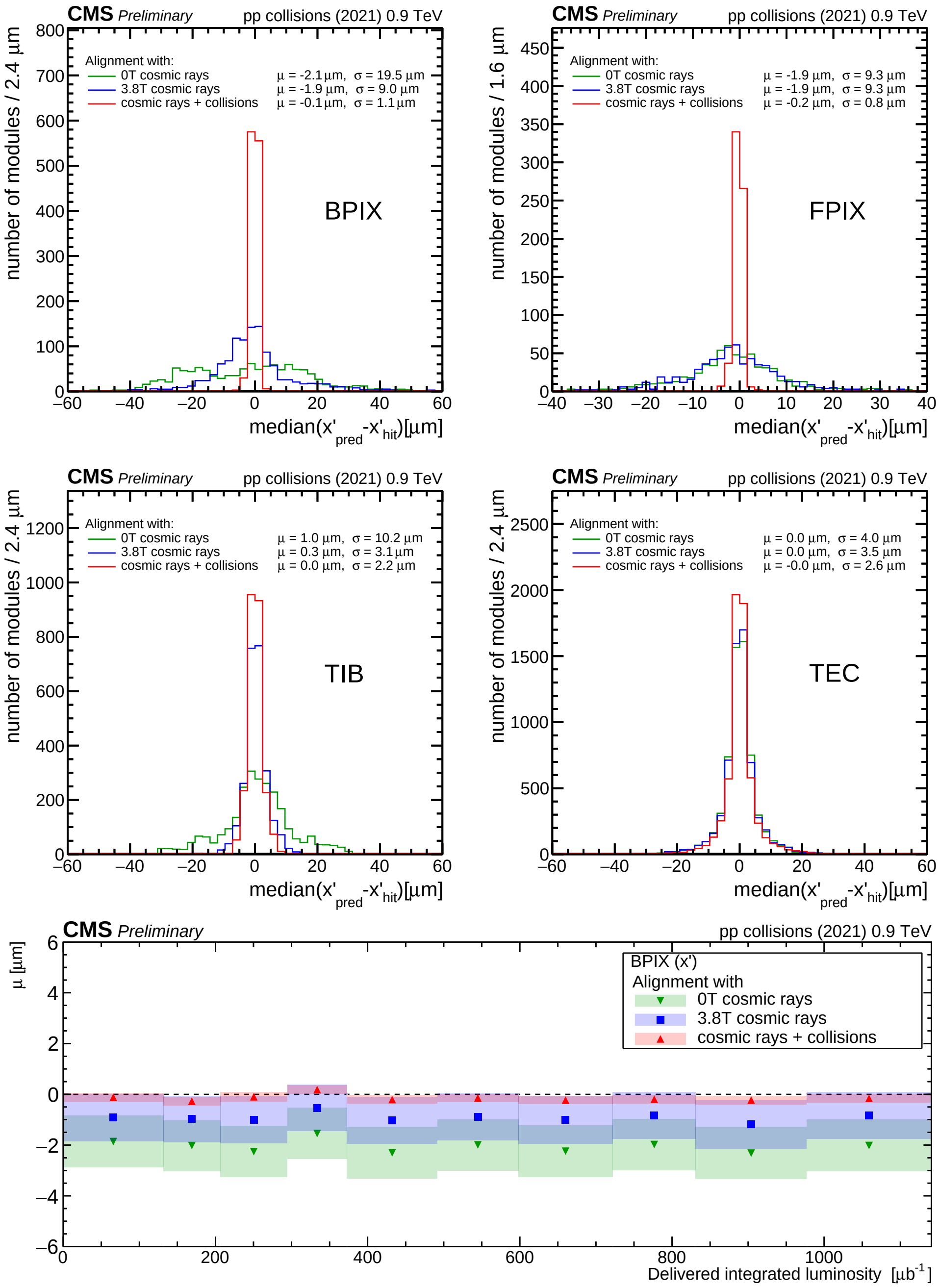


Figure A.5: The DMR distributions for the local x coordinate in BPIX (top left), FPIX (top right), and the strips tracker inner barrel (TIB) (centre left) and tracker endcap (TEC) (centre right) detectors are shown. The performance of the alignment geometries derived with CRUZET (green), CRAFT (blue) and cosmic rays + collisions at 900 GeV (red) is compared. The mean value of the DMR distributions in BPIX as a function of the delivered integrated luminosity during 2021 is also shown (bottom). The uncertainty bands correspond to the standard mean error of the depicted quantity. After the latest alignment iteration the DMR distributions are consistently closer to zero.

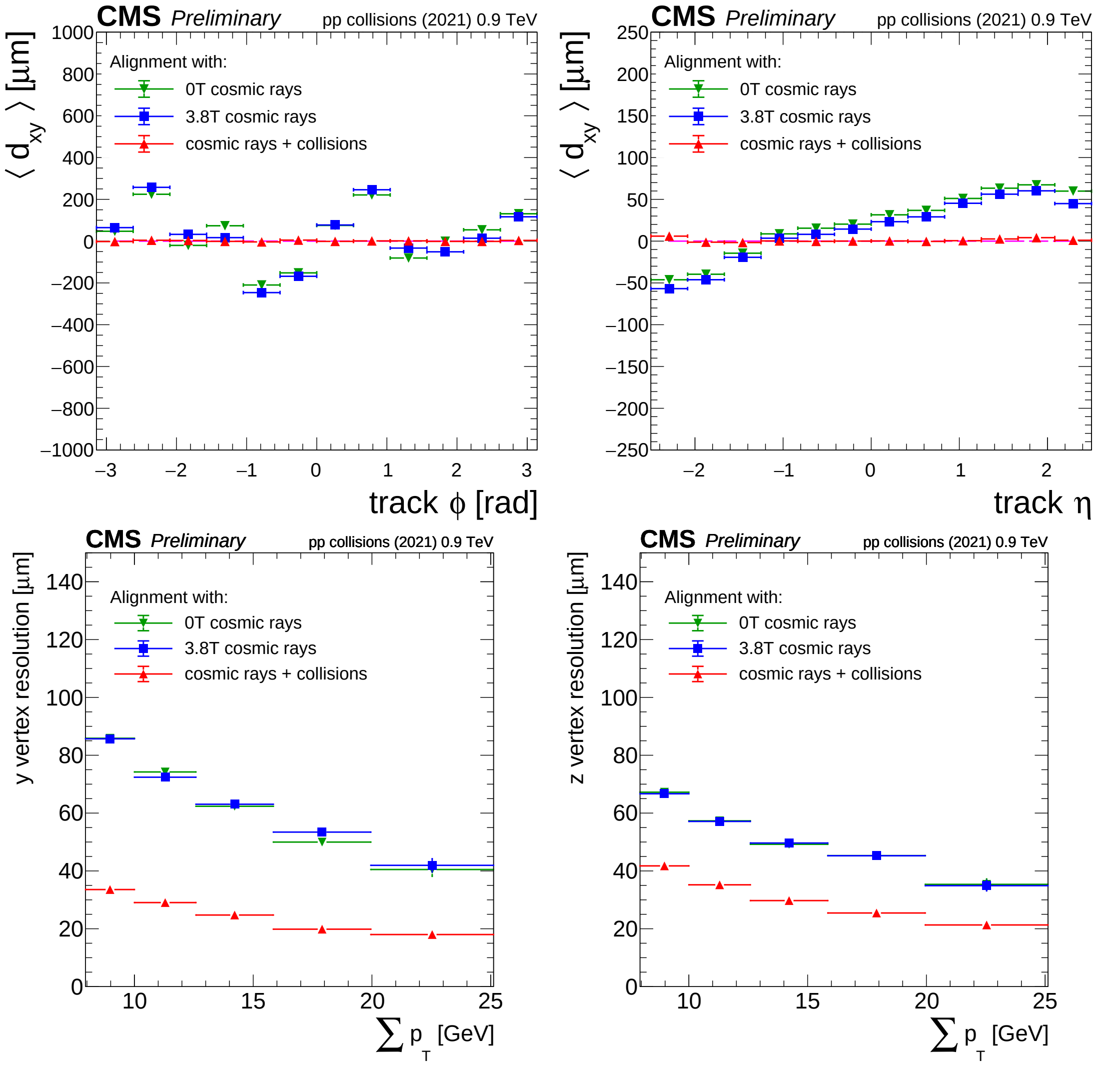


Figure A.6: The PV validation (top) of the final three 2021 tracker alignment campaigns with respect to the track ϕ (left) and η (right) coordinates in the transverse (d_{xy}) plane is shown. The PV resolution validation (bottom) in the y (left) and z (right) coordinates are also shown. The performance of the alignment geometries derived with CRUZET (green), CRAFT (blue), and cosmic rays + collisions at 900 GeV (red) is compared. These distributions showcase the improvement in the performance of the 2021 alignment campaigns, achieved iteratively.

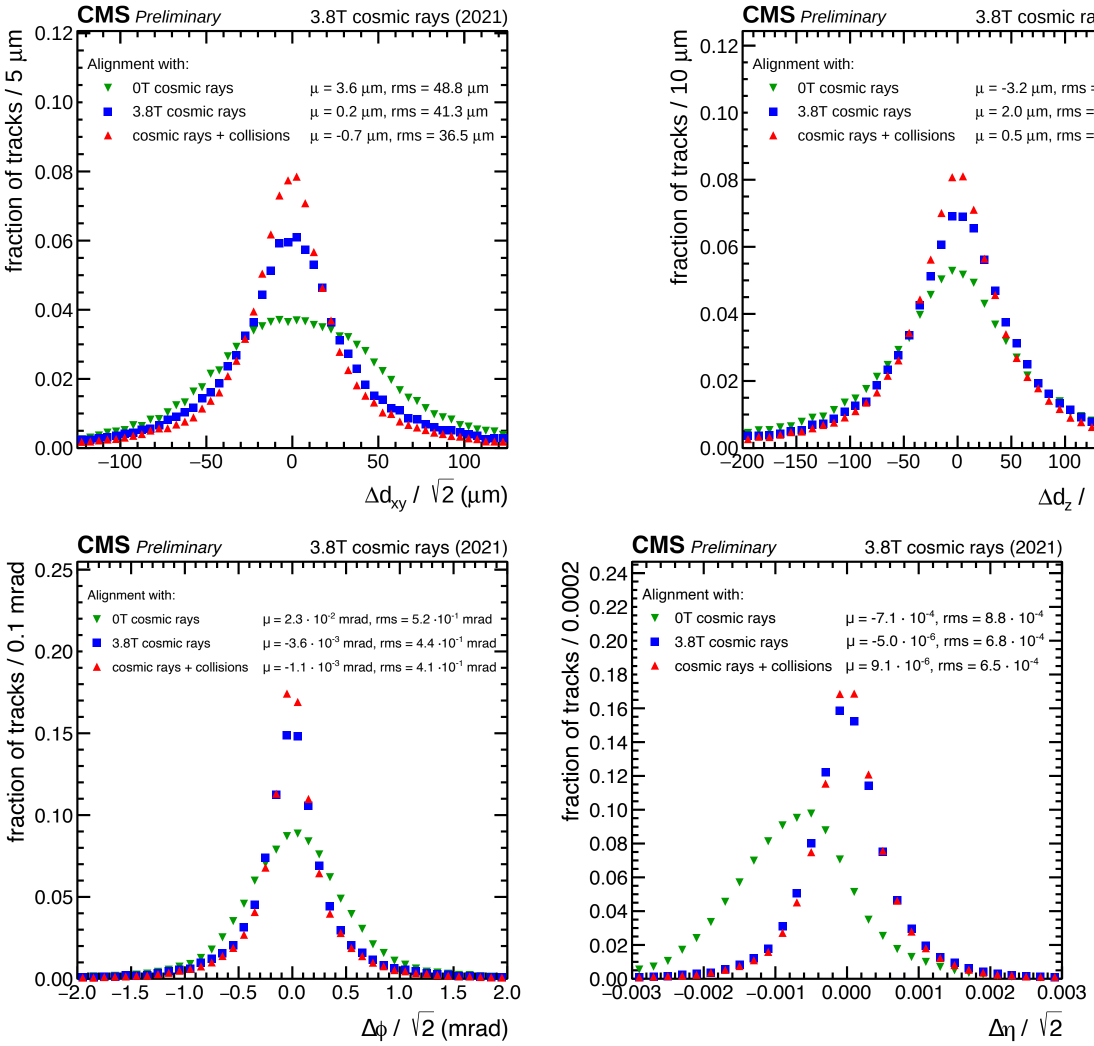


Figure A.7: MTS validation results comparing the performance of the alignment geometries derived with CRUZET (green), CRAFT (blue), and cosmic rays combined with collisions tracks recorded at $\sqrt{s}$ = 900 GeV (red). The differences in the track impact parameter in the transverse plane (top left), in the longitudinal direction (top right), in the ϕ coordinate (bottom left)l, and in the η (bottom right) coordinate are shown. A noticeable improvement was achieved iteratively.

As data-taking continued, so did the tracker alignment efforts. A summary of the performance of the CMS tracker during 2023, including the alignment undertaking, is provided in Appendix B. Fig. A.8 shows complementary tracker alignment results. The first 2023 alignment geometry, derived with 186,900 tracks from CRAFT data, is compared to two further iterations. The first was performed with 4,53 million collision tracks recorded at a centre-of-mass energy of 900 GeV and 71,200 cosmic ray tracks recorded during collisions. The next iteration was derived with 16.1 million tracks collected at a centre-of-mass energy of 13 TeV and 161,700 cosmic ray tracks recorded during collisions. These results were published in [142].

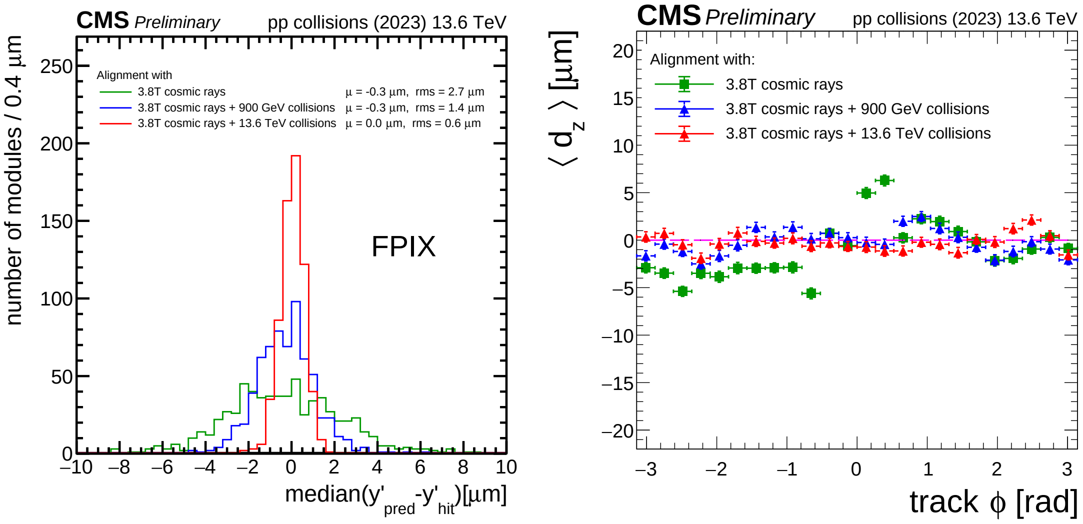


Figure A.8: The DMR distribution in the local y coordinate of the FPIX detector (left) and the PV validation for the longitudinal impact parameter (d_z) relative to the track ϕ coordinate (right) are shown. The first 2023 alignment, derived with cosmic rays (green), is compared to the next iteration, employing cosmic rays and collision tracks recorded at $\sqrt{s} = 900$ GeV (blue), followed by an alignment with cosmic rays and collision tracks recorded at $\sqrt{s} = 13.6$ TeV (red). The latest campaign delivers a considerable improvement.

A.7 Summary

The alignment of the CMS tracker is crucial for ensuring high-quality data for physics analysis. In particular, analyses involving heavy flavour identification techniques, such as the one discussed in this thesis, rely on an accurate reconstruction of tracks and vertices directly affected by the alignment of the tracker components.

Deriving the tracker alignment geometry is a challenging process due to the extensive number of calculations required. The `MillePede-II` algorithm provides an optimal way to solve the large linear system involved. Several validations are in place to ensure the quality of the alignment geometries derived. An excellent performance during Run 3 has been achieved.

appendix B

CMS Tracker performance in Run 3

In the present section, an overview of the CMS tracker performance during Run 3 is outlined, focusing on the pixel and silicon strip detectors as well as the alignment of the CMS tracker.

PROCEEDINGS OF SCIENCE

Performance of the CMS Tracker during Run 3

Daina Leyva Pernia[a] **and Maryam Bayat Makou**[a,*] **for the CMS collaboration**

[a] *Deutsches Elektronen Synchrotron (DESY),*
Notkestraße 85, Hamburg, Germany

E-mail: daina.leyva.pernia@desy.de, maryam.bayat.makou@desy.de

The current CMS silicon tracker consists of two tracking devices: the inner pixel and the outer strip detectors. The tracker occupies the region around the center of CMS, where the LHC beams collide, and therefore, operates in a high-occupancy and high-radiation environment produced by the particle collisions within the LHC tunnel.

This article provides an overview of the excellent performance of the CMS silicon tracker during the ongoing Run 3 data-taking period. It discusses the behavior of local observables, such as hit reconstruction efficiency, their response to the accumulated integrated luminosity, and the precision achieved in aligning the detector components.

European Physical Society Conference on High Energy Physics (EPS-HEP)
20-25 August 2023
Hamburg , Germany

---

[*] Speaker

1. The CMS tracker detector

Comprising 1856 silicon pixel detector modules and 15 148 silicon strip detector modules, the CMS tracker [1, 2] plays a crucial role in physics research. The Pixel detector, located closest to the interaction point, is particularly susceptible to radiation damage. Its modules are arranged in four cylindrical layers around the beampipe and three endcap disks on each side of the detector. It is surrounded by the Silicon Strip detector, which features ten cylindrical layers and twelve endcap disks. Together, they deliver robust tracking and contribute with a pivotal role in CMS vertex reconstruction. This article highlights the remarkable performance attained in the face of challenging conditions during the Run 3 LHC data-taking period, including managing up to 62 interactions per beam crossing.

2. Tracker detector performance during Run 3

2.1 Pixel detector performance

Being the closest component to the interaction point, the Pixel detector is much more likely to suffer from radiation damage effects. These can lead to inefficiencies or instabilities, impacting the data quality. Consequently, during the second LHC Long Shutdown (LS2) period, from 2018 to 2022, the Pixel detector was extracted for a series of improvements and refurbishments [3]. This includes the installation of a whole new pixel barrel layer to replace the one nearest to the interaction point and the repair of modules and electronics in the other layers and disks. A measure of its performance during the present data-taking period is shown in Fig. 1 (left), showing the hit efficiency with instantaneous luminosity during Run 3 [4]. The distribution exhibits rather stable performance, which slightly deteriorates towards larger instantaneous luminosity for all layers, with the layer one efficiency being the most affected. This is mostly caused by the saturation of the readout buffer in the chips [5]. The improvements in preparation for Run 3 allowed for a hit efficiency higher than 96% at 22×10^{33} cm^{-2}s^{-1}. A summary of the hit efficiency in the barrel layers for the delivered integrated luminosity in Run 3 is shown in Fig. 1 (right). Here again, Layer 1 efficiency decreases rather rapidly with accumulated radiation. The effect can be partially recovered by increasing the application voltage for the sensors and through continuous calibrations [6], which can be seen as the discontinuities where the efficiency increases rapidly in the figure.

2.2 Silicon Strip detector performance

The integrity of the strip detector is essential for data-taking. The stability during Run 3 can be seen in the fraction of bad module components trend with the integrated luminosity, reflecting the integrity of its components to maintain an excellent tracking performance. This trend is shown in Fig. 2 for 2022 and 2023 proton-proton collisions [7], showing a rather stable trend, with a fraction of active channels of about 96%. The jumps at 205 fb^{-1} are caused by the recovery of a cooling loop on the endcap region. Furthermore, some of the module power supplies in the Tracker Inner Disks were turned off during 2023 because of technical issues with the Front-End Drivers after 245 fb^{-1}[7]. As can be seen, the trend returned to usual values after the power to the modules was restored. Overall, no major issues have affected data quality.

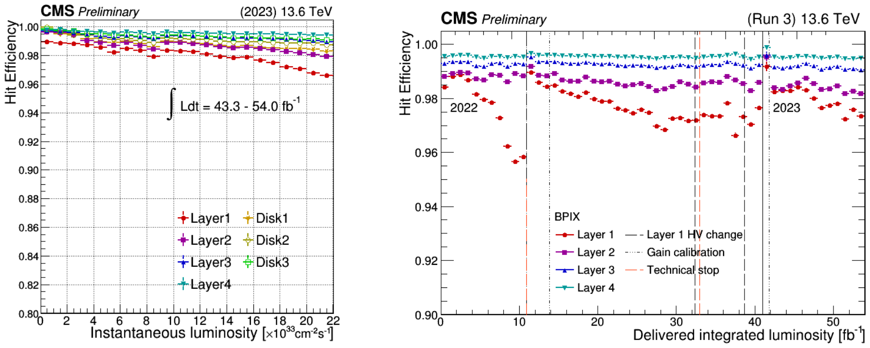


Figure 1: Pixel detector Hit Efficiency vs Instantaneous Luminosity during data-taking runs in May and June 2023 (left) and vs delivered integrated luminosity during Run 3 (right).

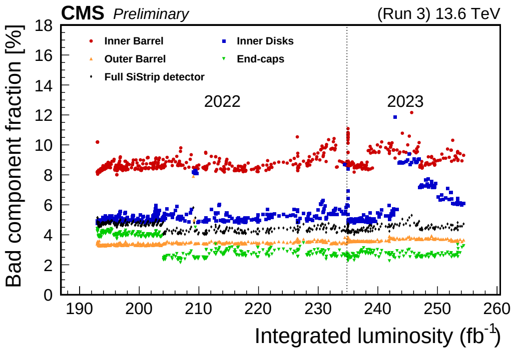


Figure 2: Evolution of fraction of modules flagged as bad vs delivered integrated luminosity during Run 3.

2.3 Tracker alignment performance

A feature of the CMS tracker detector is its outstanding hit resolution, of about 10 μm. However, after installation, a mechanical alignment can only yield a precision on the position and orientation of the modules of about 0.1 mm [8]. Furthermore, it has been observed that changes in the conditions, like magnet cycles and temperature changes, as well as the long-term exposure to a high-radiation environment, can cause real or apparent movements of the detectors [8, 9]. To improve the precision of the knowledge of the component's geometry, a track-based alignment approach, relying on the minimization of the sum of squares of normalized track-hit residuals, is performed. This process allows us to obtain changes to alignment parameters, which describe the geometrical location of the components.

After the technical stop at the end of 2022 and the beginning of 2023 (Year End Technical Stop or YETS), significant movements were expected as explained before. To overcome this, alignment geometries were iteratively derived using cosmic rays and proton-proton collision data at 900 GeV and 13.6 TeV, as data became available [10]. The performance achieved, continuously improving the mean and reducing the width of the track hit residuals to guarantee the accuracy for data-taking, is shown in Fig. 3.

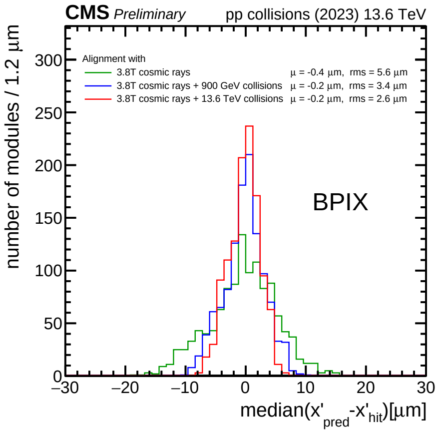


Figure 3: Distribution of median residuals in the local x coordinate on the barrel pixel detector according to alignment geometries derived iteratively in 2023.

3. Summary

The CMS tracker system plays a critical role in data-taking, enabling precise reconstruction of charged particle positions and momenta, even under the challenging conditions of Run 3, with a peak pileup of about 62 interactions per beam crossing. This article has discussed the performance of the Pixel and Silicon Strip detectors during Run 3, highlighting the continuous efforts to maintain exceptional performance and the role of the Tracker alignment in ensuring high-quality data.

References

[1] CMS collaboration, “The CMS experiment at the CERN LHC”, *Journal of Instrumentation* **3** (2008) S08004.

[2] CMS collaboration, “Development of the CMS detector for the CERN LHC Run 3”, `2309.05466`.

[3] L.O.S. Noehte and on behalf of the Tracker Group of the CMS collaboration, “CMS Phase-1 pixel detector refurbishment during LS2 and readiness towards the LHC Run 3”, *Journal of Instrumentation* (2022) .

[4] CMS Collaboration, “Pixel Detector Performance in early 2023”, CMS-DP-2023-041, `https://cds.cern.ch/record/2865842`, (2023).

[5] T.A. Vami for the CMS Collaboration, "Calibration and performance of the CMS pixel detector in LHC Run 2", Proceedings of 7th Annual Conference on Large Hadron Collider Physics — PoS(LHCP2019) `https://doi.org/10.22323/1.350.0010`, (2019).

[6] CMS Collaboration, "Pixel Detector Performance in Run 3", CMS-DP-2022-067, `https://cds.cern.ch/record/2844889`, (2022).

[7] CMS Collaboration, "CMS Silicon Strip Tracker Performance in 2023", CMS-DP-2023-040, `https://cds.cern.ch/record/2865841`, (2023).

[8] CMS Collaboration, "Strategies and performance of the CMS silicon tracker alignment during LHC Run 2", *Nucl. Inst and Meth. A* **1037** (2022) 166795.

[9] A.V. Barroso for the CMS Collaboration, "Tracker Alignment in CMS: Interplay with Pixel Local Reconstruction", Proceedings of Science — PoS(Pixel2022), arXiv:2303.16642, (2023). 10.22323/1.350.0010.

[10] CMS Collaboration, "Tracker alignment performance in early 2023", CMS-DP-2023-039, `https://cds.cern.ch/record/2865840`, (2023).

appendix C

Triggers for the Run 3 Beyond Standard Model H→bb analysis

The analysis triggers used in the search for additional Higgs bosons in final states with b-quarks were deployed in preparation for and during the beginning of the Run 3 data-taking period. Recent upgrades, particularly in the online b-tagging component, were targeted. These triggers are carefully designed to suppress the large multi-jet background rate while maintaining good efficiency in selecting potential signal events. Moreover, as mentioned in section 3.2.4, the HLT has a limited amount of resources, restricting the total output rate and the processing time per event. This chapter summarizes the studies performed to ensure the upgraded triggers meet all the requirements for Run 3 data-taking.

C.1 Level-1 seeds

Aiming to employ the same semi-leptonic (SL) and fully-hadronic (FH) Level-1 (L1) seeds utilized during Run 2 (see table 5.7), a feasibility study was conducted. For this, the Run 3 L1 rates were estimated and compared with the Run 2 values, taking them as a benchmark.

Table C.1 summarizes the results of this study. The second column shows the rates of the L1 seeds, determined according to Run 2 conditions at an average pile-up (PU) of 53. The expected performance during Run 3 at similar PU conditions was evaluated in two approaches, summarized in the third and fourth columns of table C.1. Run 2 data was re-emulated in the first approach to match the Run 3 upgrades in the L1 menu while using Run 2 trigger primitives [230]. Run 3 simulated proton-proton (p−p) collision data with average PU = 51 was employed in the second approach. The flat PU profile of these samples was re-weighted according to the CMS Luminosity physics objects group recommendations for Run 3 trigger development [230].

The rates estimated with Run 3 simulations slightly exceed the targeted rate. However, simulated data cannot fully emulate realistic beam conditions [230], and the rates yielded in simulations are generally larger. This is reflected in the total menu rate, which was expected to be around 63 kHz, while simulations yielded a total menu rate of 97 kHz. Therefore, the data-based estimations from the third column in table C.1 are taken as more accurate. As the expectations were matched within the uncertainty, the L1 seeds in table C.1 were included in the main HLT analysis paths in the corresponding categories since the beginning of Run 3.

	L1 rate (kHz)		
L1 seeds:	Run 2 at PU 53	Run 2 D at PU 53 (re-emulated)	Run 3 MC at PU 51 (reweighted)
SL: `L1_Mu12er2p3_Jet40er2p3_dR_Max0p4_DoubleJet40er2p3_dEta_Max1p6`	1.13 ± 0.42	1.28 ± 0.13	1.90 ± 0.27
FH: `L1_DoubleJet112er2p3_dEta_Max1p6`	3.38 ± 0.73	2.70 ± 0.19	4.39 ± 0.18

Table C.1: L1 rates estimated for the SL and FH seeds in preparation for Run 3.

C.2 HLT trigger rates and processing time

Several improvements concerning tracking, muon reconstruction, and b-tagging were implemented for the Run 3 HLT menu, the latter being the most relevant as the DeepJet algorithm [182] became available as part of the online selection. As discussed in section 4.2.5, this algorithm provided an enhanced efficiency compared to its predecessor, the DeepCSV algorithm [181].

The deployment of these improvements in the main HLT paths of the analysis was carried out according to the recommendations of the Trigger Studies Group (TSG) in Ref. [231]. The upgraded SL and FH paths are:

- **SL path**: `HLT_Mu12_DoublePFJets40MaxDeta1p6_DoublePFBTagDeepJet_pX`
- **FH path**: `HLT_DoublePFJets116MaxDeta1p6_DoublePFBTagDeepJet_pX`

where `pX` indicates that the b-tagging component features a misidentification rate of `0.X` %. This value was optimized to control the output rate while also guaranteeing good signal efficiency by tuning the online b-tagging thresholds. The approach followed is described next.

HLT rates

The budget assigned during Run 2 determined the target rates of the Run 3 SL and FH paths, i.e. 18 Hz and 10 Hz at an average PU = 53, respectively. Moreover, a target on the selection efficiency of signal events ranged from 0.4 to 11% for the SL path, and from 12 to 25% for the FH path. These numbers were estimated based on Run 2 simulated signal samples.

The expected rate of the upgraded HLT paths was calculated following the TSG recommendations in Ref. [232]. The dataset used consisted of 2018 p−p collision data with a maximum PU = 51. Several values of the online b-tag thresholds were used to determine the optimal configuration.

In parallel, the efficiency of selecting signal events was estimated with Run 3 simulated signal samples at hypothetical masses of 120, 350, 600, and 1200 GeV. The information from both studies is considered in the selection of the online b-tag threshold to guarantee a good compromise between the rate reduction and signal efficiency that enhances the significance of the search.

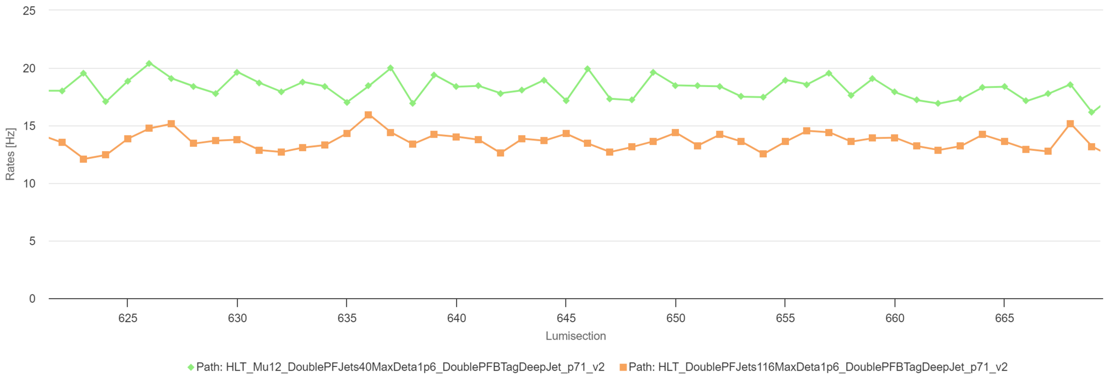


Figure C.1: Rate of the SL (green) and FH (orange) HLT paths during a brief period of time in 2022. In the x-axis, time is accounted for as lumisections. One lumisection is a short data-taking time where the instantaneous Luminosity does not change, defined as approximately 23.3 seconds.

An online b-tag threshold was set at 0.7, which yielded a rate of 18 Hz (13 Hz) on the SL (FH) HLT path, with a signal efficiency ranging approximately from 0.5−12% (15−28%) for the first HLT menu of Run 3 data-taking for physics in 2022. Shortly after, new jet energy corrections and simulated samples became available, and the online b-tag threshold was further tuned to 0.5, also considering the target rate and signal efficiency, from April 2022. Figure C.1 shows the HLT rates over time during a brief data-taking period in 2022, with similar PU conditions as those used in the rate estimation.

HLT timing

The deployed paths must perform efficiently to meet the processing time constraints at the HLT. The mean processing time per path should be in the order of 10 ms.

Three benchmark machines were provided by TSG to estimate the processing time per event, providing a meaningful assessment that approximates the performance at the HLT farm during data-taking. The recommendations in Ref. [233] were followed, and 10,000 Run 2 data events with an average PU of 50 were employed to derive the timing estimation. The results in Fig. C.2 for the SL (left) and FH (right) paths. These figures show interesting features as processing time increases. The first group of events peaking near 0 ms represent events that are rejected early in the sequence, for instance, events that did not pass the L1 selection. The second group of events originates from relatively fast processing parts of the HLT sequence, such as clustering algorithms. The final group of events towards higher processing time, featuring a long tail, corresponds to those that undergo further selection steps with increased complexity, such as track reconstruction and b-tagging. The mean processing time is also shown in the figures, resulting in 10 and 17 ms for the SL and FH paths. Both results lie within the acceptable markers, and were thus deployed for data-taking.

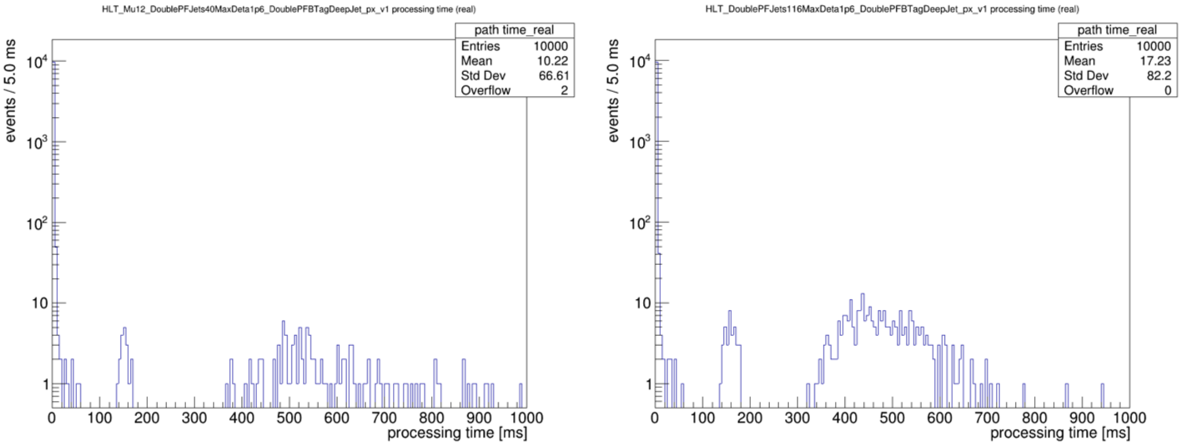


Figure C.2: Processing time of the SL (left) and FH (right) HLT paths. In both figures, the first group of events peaking near 0 ms corresponds to those that are rejected early in the HLT sequence, for instance, events that did not pass the L1 selection. The second group of events, between 100 and 200 ms, originates from relatively fast processing parts of the HLT sequence, such as clustering algorithms. The final group of events, starting at 300 ms, corresponds to events that undergo more complex selection steps, such as track reconstruction and b-tagging. The mean processing time is also shown in the figures.

C.3 Migration to ParticleNet

In preparation for 2024 data-taking, the b-tagging component of the BSM H→bb triggers was updated from the DeepJet algorithm to ParticleNet [234], which outperformed its predecessor. Fig. C.3 compares the performance of the ParticleNet and DeepJet algorithms. The optimal ParticleNet online b-tag thresholds were determined based on a similar study as the one outlined in previous sections for the migration from DeepCSV to the DeepJet algorithm. The new algorithm was deployed in the trigger menu from the beginning of the 2024 data-taking.

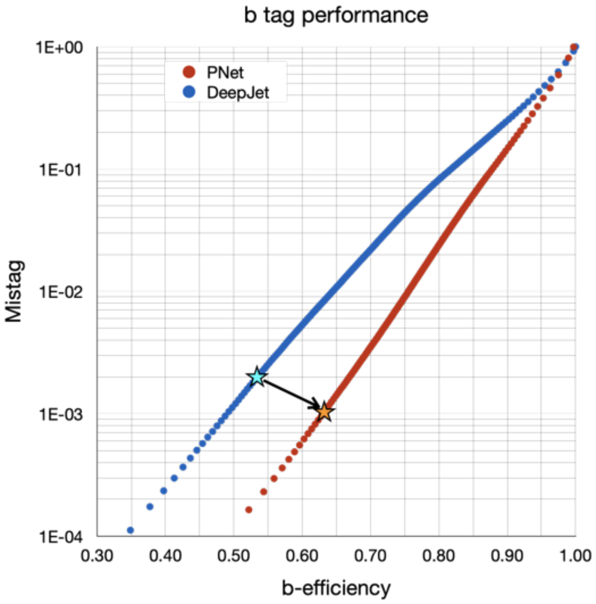


Figure C.3: Performance of the ParticleNet and DeepJet b-tagging algorithms represented as the misidentification rate with light flavours with respect to the b-tag efficiency. The star in each distribution represents the performance with the online b-tag thresholds used by the BSM H→bb triggers. The resulting improvement in the migration from DeepJet to ParticleNet is observed on the transition from the cyan star towards the orange star. Taken from [235].

appendix D

Systematic uncertainties affecting the 2017 signal shapes in the SL and FH categories

As described in Section 5.7.2, several systematic uncertainties can affect the shape and/or the normalization of the signal templates used in the analysis presented in this thesis. For visualization purposes, the effect of these systematic uncertainties in the 2017 SL and FH categories was not shown in the main body of this Thesis and will be displayed below.

Shape-altering systematic uncertainties
The uncertainties related to the jet energy corrections, i.e. JES and JER affect the shape and normalization of the signal distributions, as explained Section 5.7.2. Fig. D.1 show the effect of these uncertainties in the 2017 SL analysis, while Fig. D.2 and D.3 illustrates the effect 2017 FH category.

Shape-invariant systematic uncertainties
The uncertainties related to the kinematic trigger scale factors, online and offline b tag scale factors, and the L1 prefiring correction have a negligible effect on the shape of the M_{12}distribution. Similarly, the uncertainties on the muon ID scale factor affect the signal normalization in the 2017 SL category. To illustrate the impact of these uncertainties on the invariant di-jet mass, some exemplary results for selected mass points within each fit range from the 2017 SL category are shown in Figs. D.4 and D.5. The effects on the 2017 FH category in Figs. D.6, D.7, D.8, and D.9.

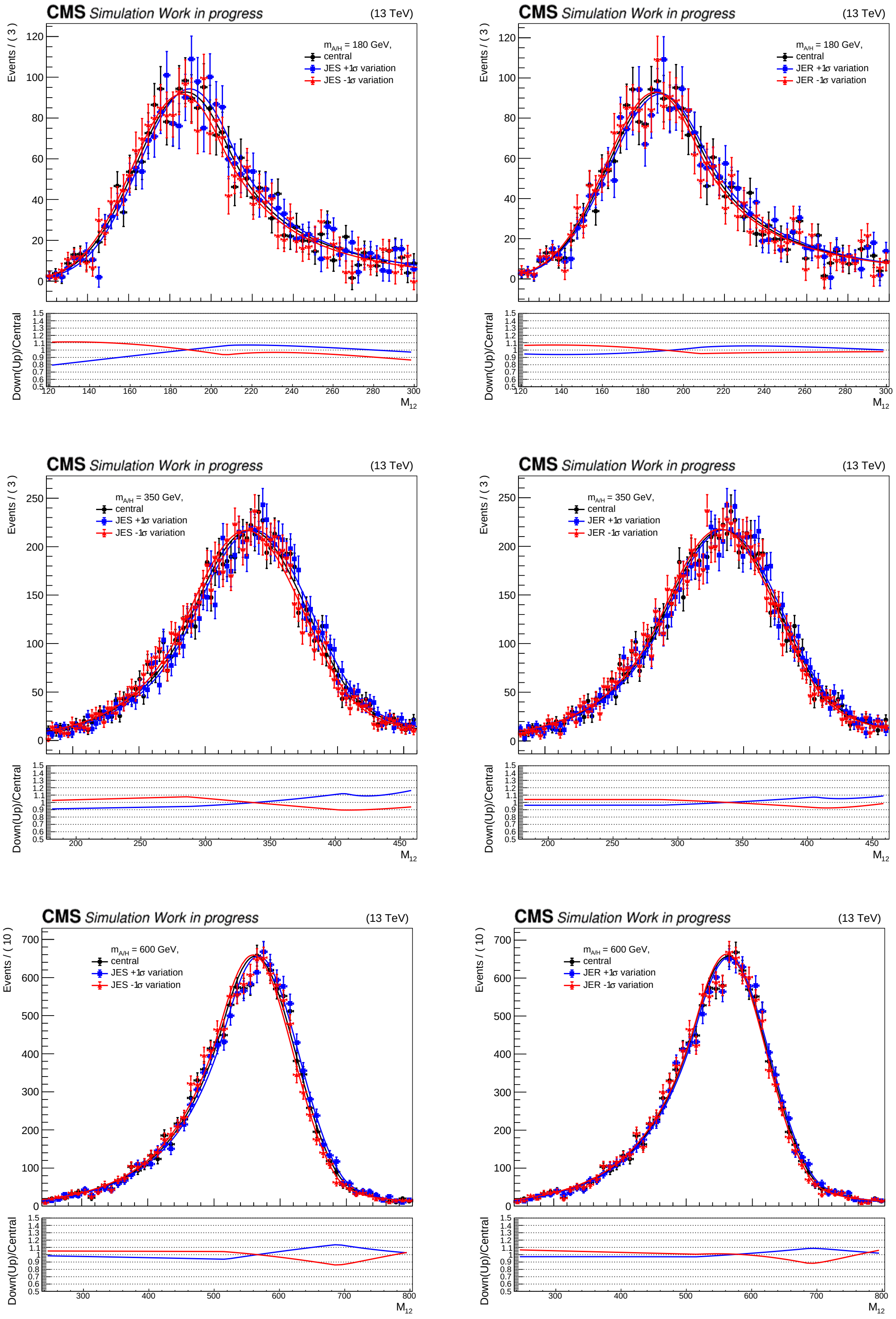


Figure D.1: Shape-altering systematic uncertainties from JES (left) and JER (right) uncertainties in the 2017 SL channel. The plots depict representative signal masses from each fit range: 180 GeV (top) in FR1, 350 GeV (middle) in FR2, and 600 GeV (bottom) in FR3. The parametrizations of up and down variations, as well as for the nominal, are shown. The bottom panel in each plot shows the ratio of the up/down parametrization with respect to the central distribution.

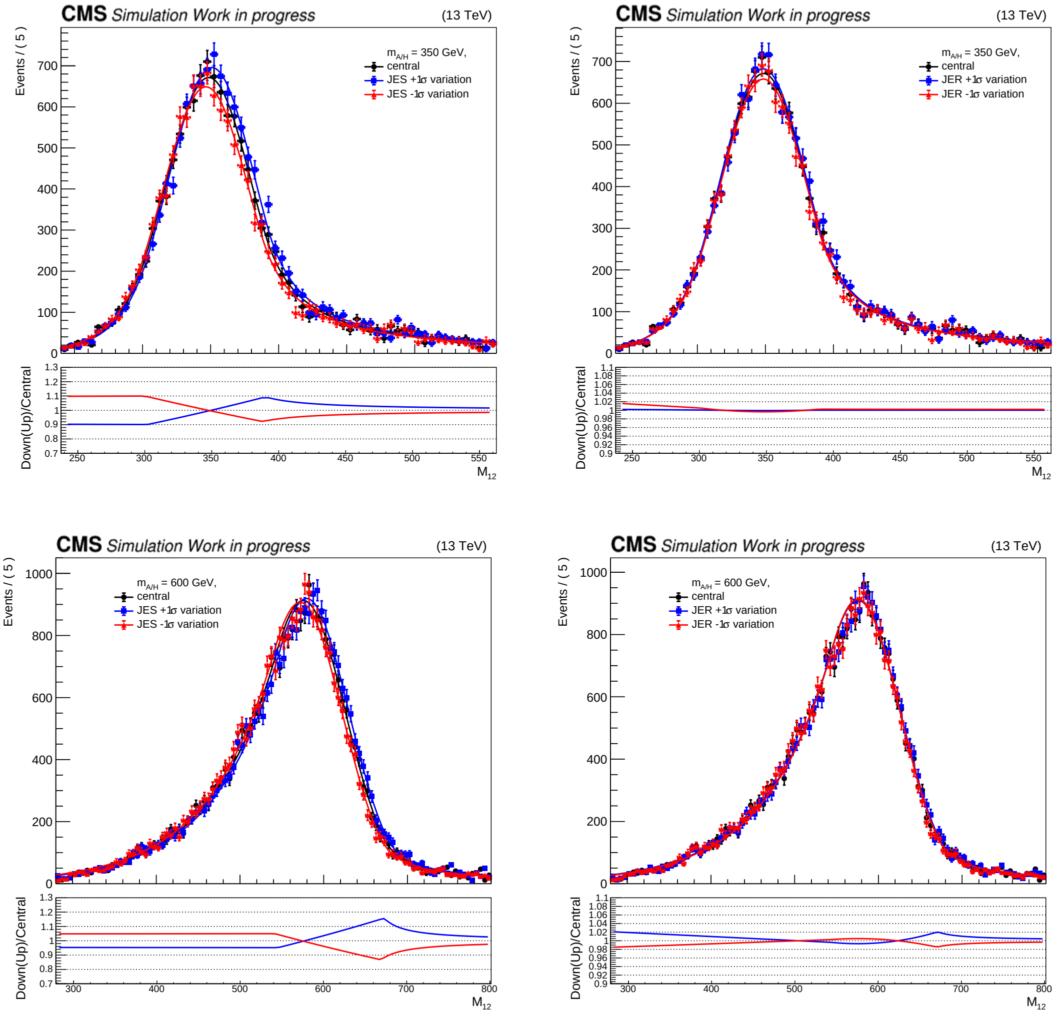


Figure D.2: Shape-altering systematic uncertainties from JES (left) and JER (right) uncertainties in the 2017 FH channel. The plots depict representative signal masses from each fit range: 350 GeV (top) in FR1 and 600 GeV (bottom) in FR2. The parametrizations of up and down variations, as well as for the nominal, are shown. The bottom panel in each plot shows the ratio of the up/down parametrization with respect to the central distribution.

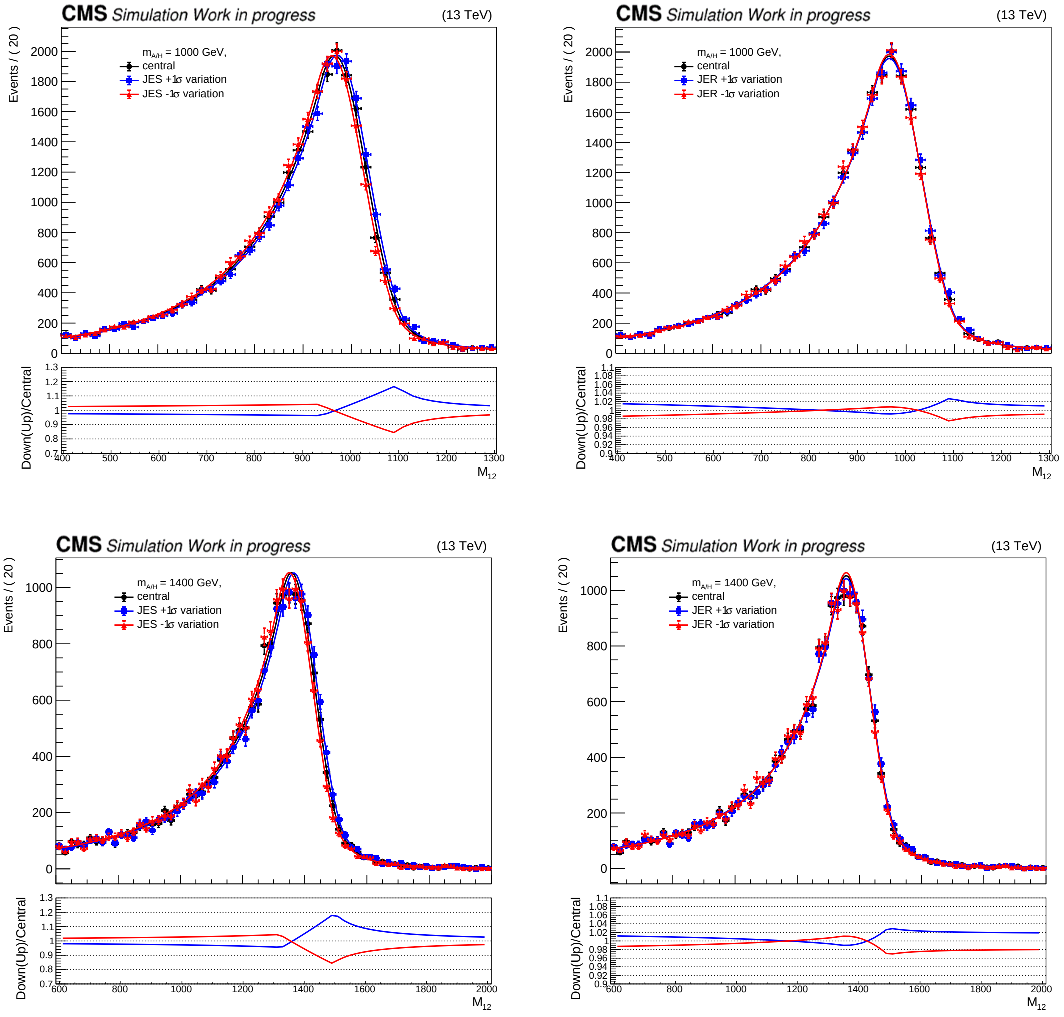


Figure D.3: Shape-altering systematic uncertainties from JES (left) and JER (right) uncertainties in the 2017 FH channel. The plots depict representative signal masses from each fit range: 1000 GeV (top) in FR3 and 1400 GeV (bottom) in FR4. The parametrizations of up and down variations, as well as for the nominal, are shown. The bottom panel in each plot shows the ratio of the up/down parametrization with respect to the central distribution.

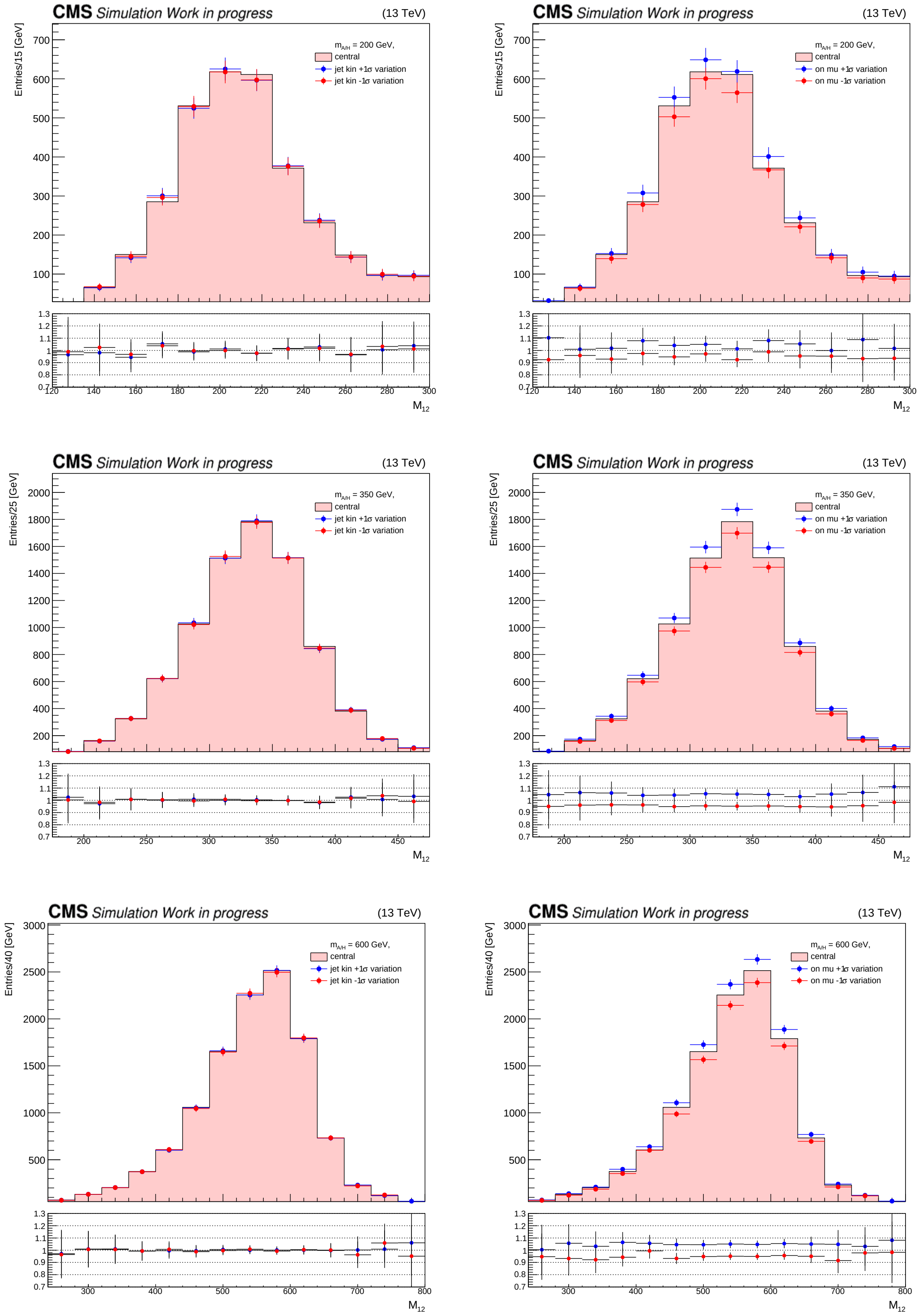


Figure D.4: Shape-invariant systematic uncertainties from trigger jet kinematic (left) and trigger muon kinematic (right) scale factor uncertainties in 2017 SL channel, for the representative signal masses from each fit range : 200 GeV (top) in FR1, 350 GeV (middle) in FR2, and FR3 600 GeV (bottom) in FR3, respectively.

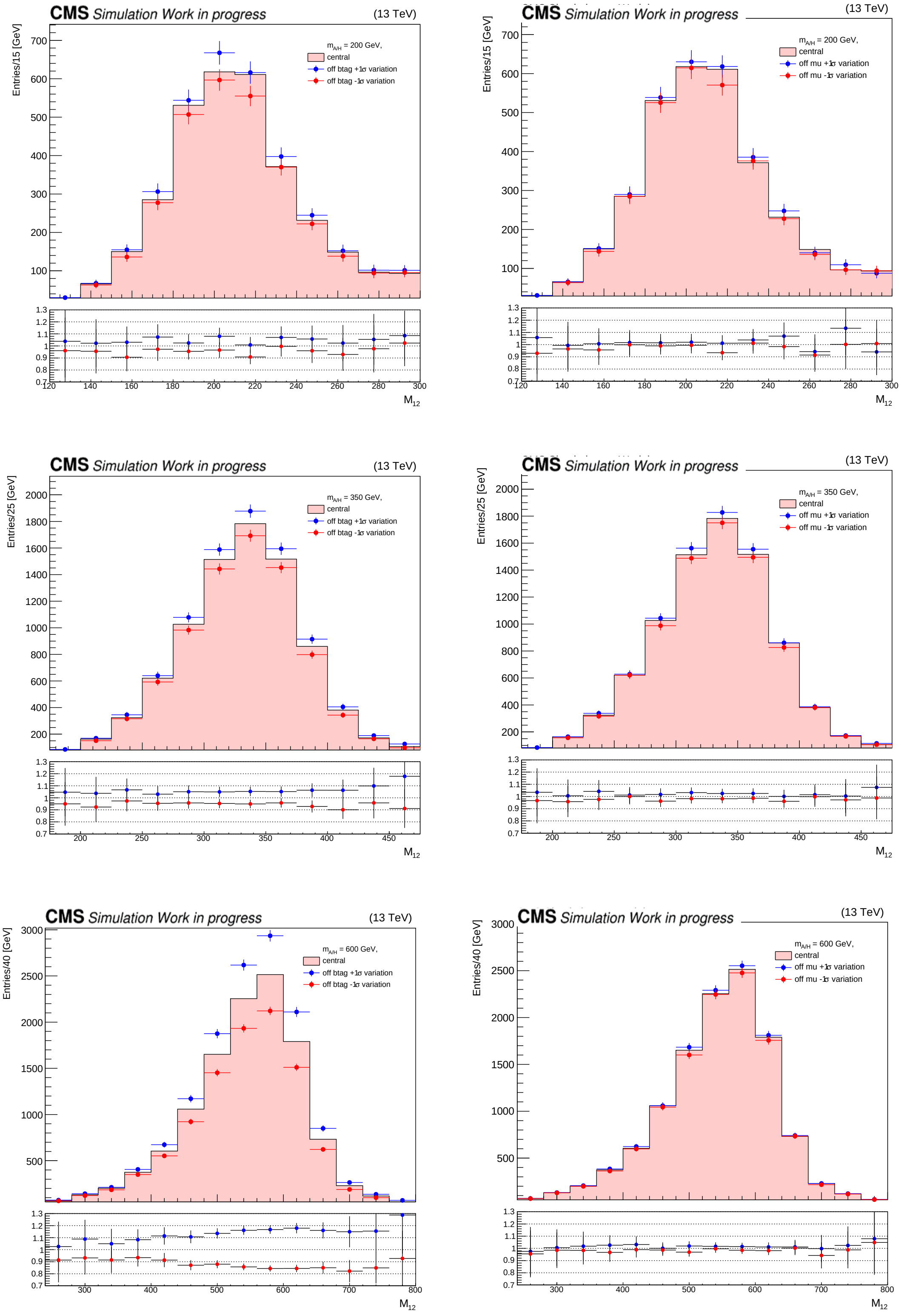


Figure D.5: Shape-invariant systematic uncertainties from offline b-tagging (left) and offline muon ID (right) scale factor uncertainties in 2017 SL channel, for the representative signal masses from each fit range : 200 GeV (top) in FR1, 350 GeV (middle) in FR2, and FR3 600 GeV (bottom) in FR3, respectively.

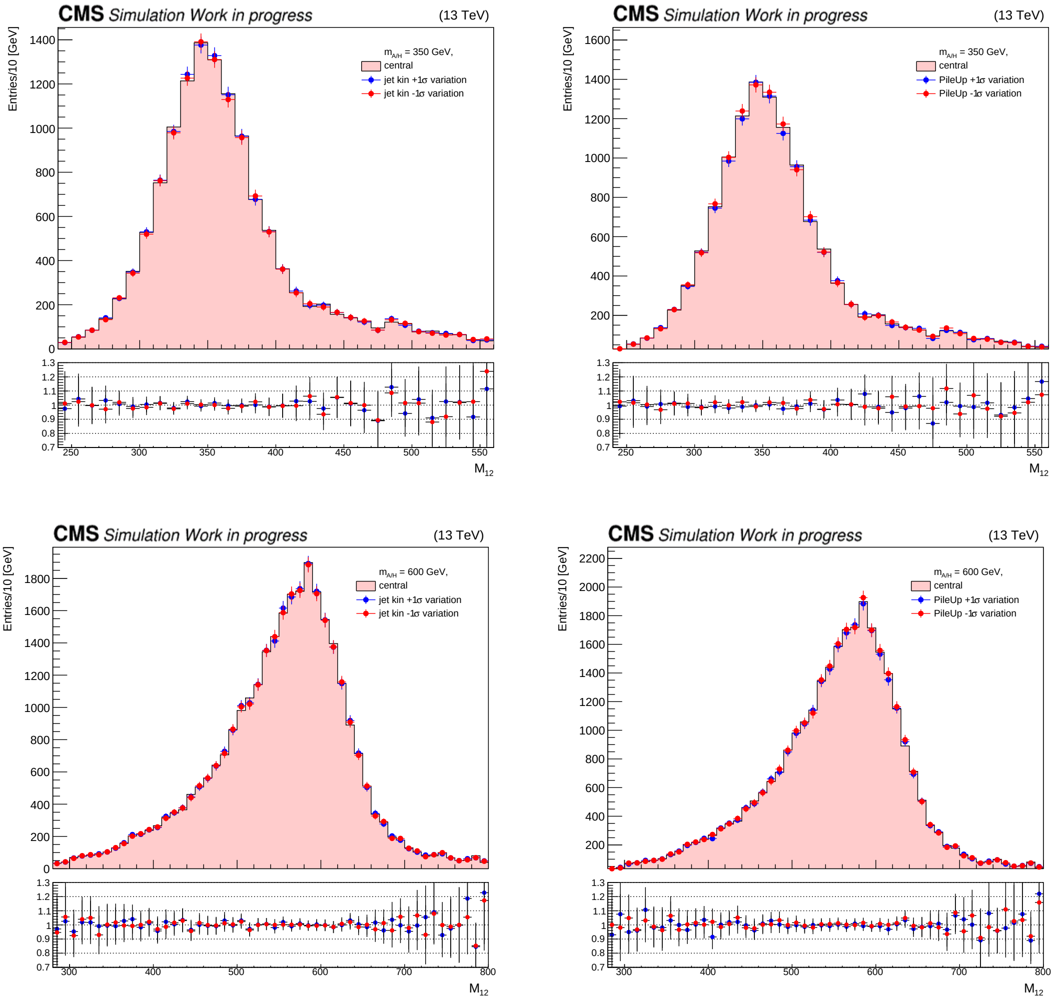


Figure D.6: Shape-invariant systematic uncertainties from trigger jet kinematic scale factor (left) and pile-up reweighting (right) uncertainties in 2017 FH channel, for the representative signal masses from each fit range : 350 GeV (top) in FR1, 600 GeV (bottom) in FR2, respectively.

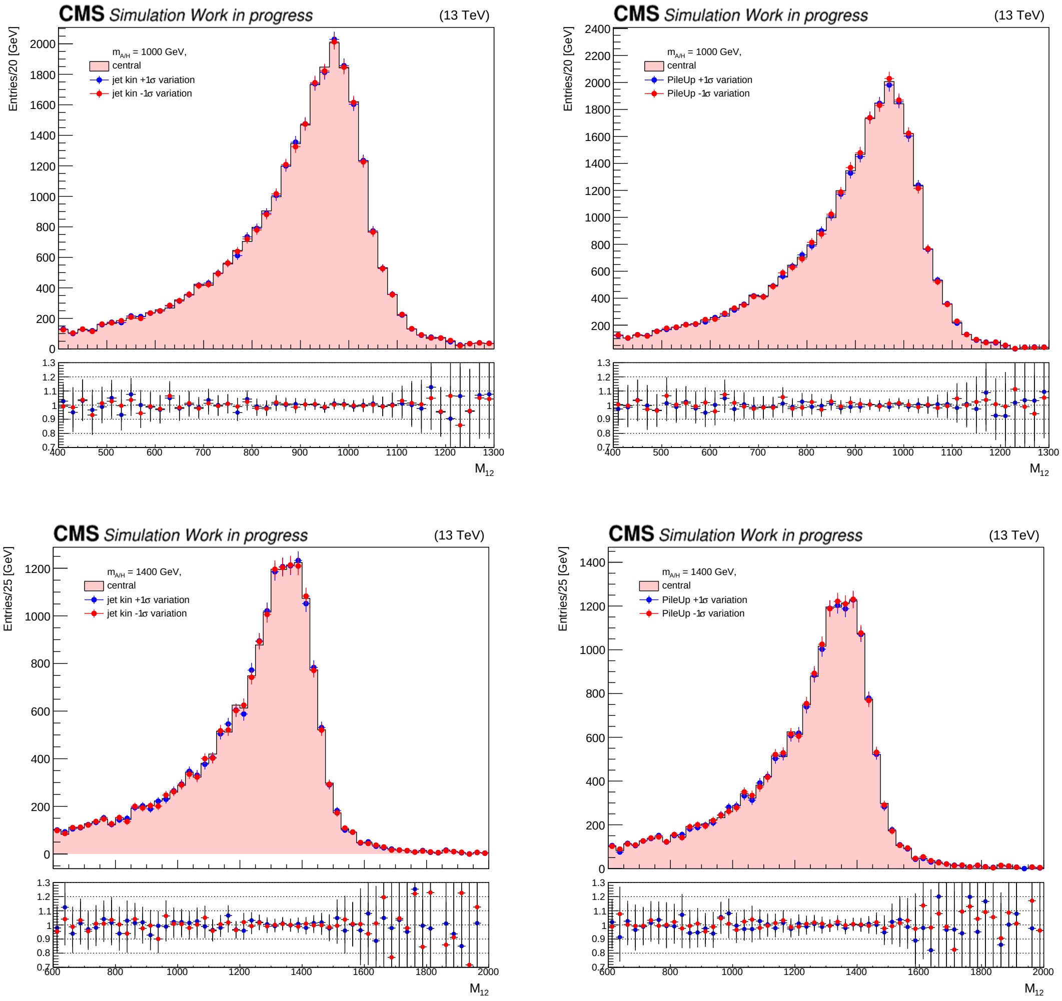


Figure D.7: Shape-invariant systematic uncertainties from trigger kinematic scale factor (left) and pile-up reweighting (right) uncertainties in 2017 FH channel, for the representative signal masses from each fit range : 1000 GeV (top) in FR3, 1400 GeV (bottom) in FR4, respectively.

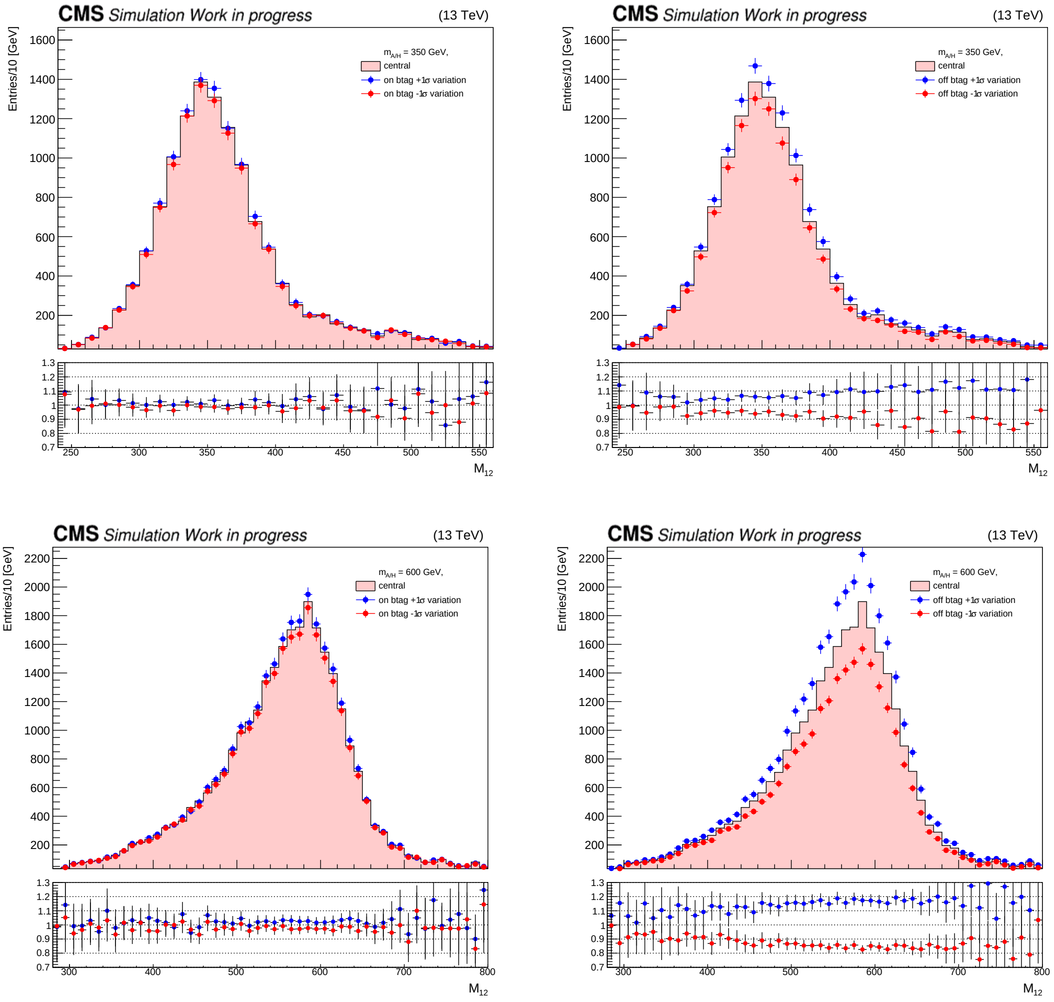


Figure D.8: Shape-invariant systematic uncertainties from trigger online b-tagging (left) and offline b-tagging (right) scale factor uncertainties in 2017 FH channel, for the representative signal masses from each fit range : 350 GeV (top) in FR1, 600 GeV (bottom) in FR2, respectively.

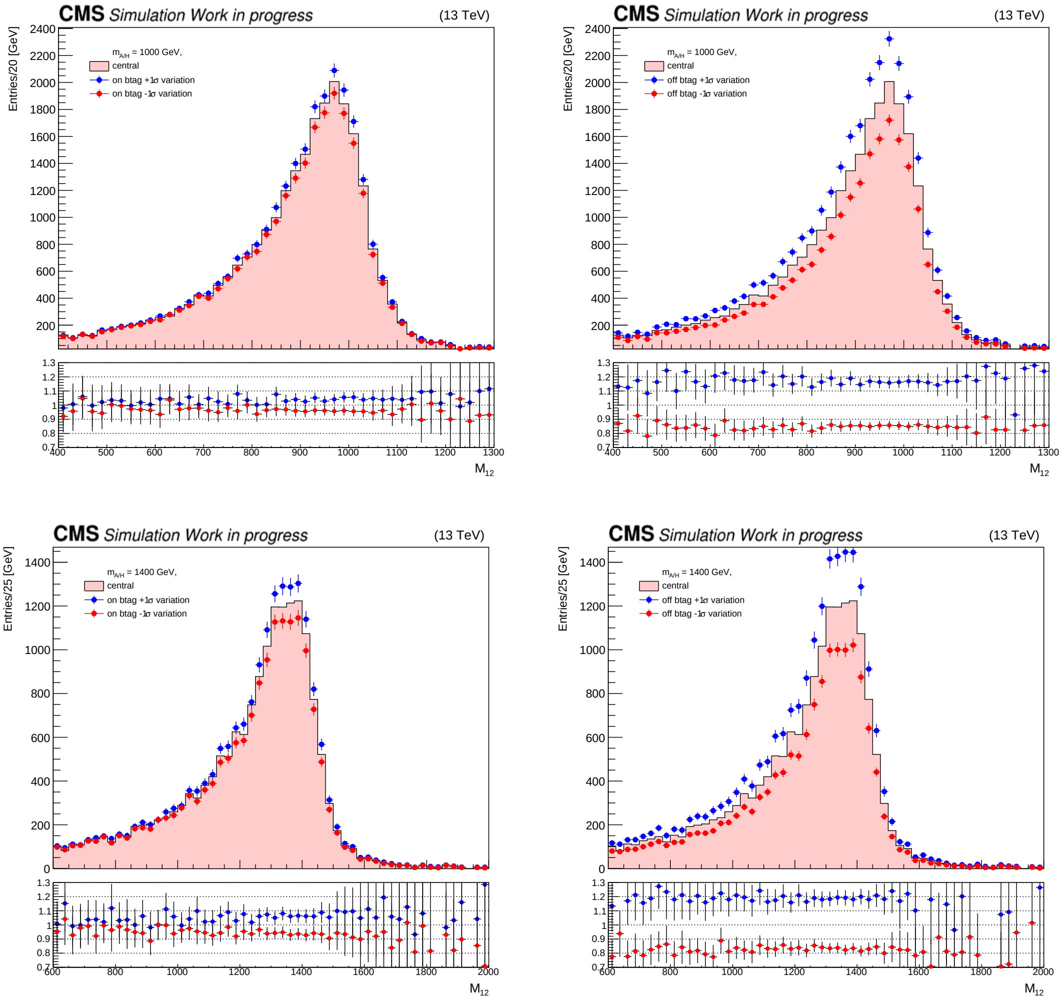


Figure D.9: Shape-invariant systematic uncertainties from trigger online b-tagging (left) and offline b-tagging (right) scale factor uncertainties in 2017 FH channel, for the representative signal masses from each fit range : 1000 GeV (top) in FR3, 1400 GeV (bottom) in FR4, respectively.

appendix E

Weighting method for b-tag selection in Transfer Factor studies

Section 5.8.2 presents the transfer factor utilized in the modelling of the background in the signal region. To obtain it, QCD Monte Carlo samples enriched with b-quarks are employed. However, these samples notoriously suffer from limited statistics, mainly caused by the triple b-tag selection used in the main analyses of this Thesis. A cut-based selection causes the loss of most events, affecting the statistical precision of the transfer factor studies relevant to this analysis.

To overcome this, a b-tag weighting technique is used for the MC-based estimation of the transfer factor shape. It relies on instead of applying hard cuts according to whether or not a jet satisfies a certain b-tag working point, a weight factor corresponding to the b-tag efficiency with respect to the true jet flavour is applied to the event. This way, the statistical uncertainty is significantly decreased, correspondingly improving the accuracy of the TF shape. The b-tag efficiencies by flavour employed in all analyses and closure tests comparing relevant distributions with the cut-based and the b-tag weights approach can be found in Appendix E. Although this technique has already been used earlier at an internal level in CMS analyses, the ATLAS collaboration recently used it in published results [236] and refers to it as “truth tagging”.

The b-tagging efficiencies for the medium working point depending on the true jet flavour are shown in Figs. E.1, E.2, and E.3 for 2017 SL, 2017 FH and 2018 FH respectively for the case of jets that triggered the online b-tag HLT component. Similarly, they can be found in Figs. E.4, E.5, and E.6 in the case where the jet did not fire the online b-tag, for 2017 SL, 2017 FH and 2018 FH respectively.

Several closure tests for the jet’s kinematic variables from the 2017 SL, 2017 FH, and 2018 FH analyses were performed. These tests compare the variable distributions obtained using the cut- and weight-based approaches. Figures E.7, E.8, and E.9 show the results corresponding to the leading jets, while figures E.10, E.11, and E.12 refer to the second leading jets. Similar figures for the third leading jet can be found in E.13, E.14, and E.15. The di-jet mass distribution reconstructed from the leading and second-leading jet resulting from the cut- and weight-based approaches are shown in Figs.E.16,E.17, and E.18 for 2017 SL, 2017 FH and 2018

FH respectively.

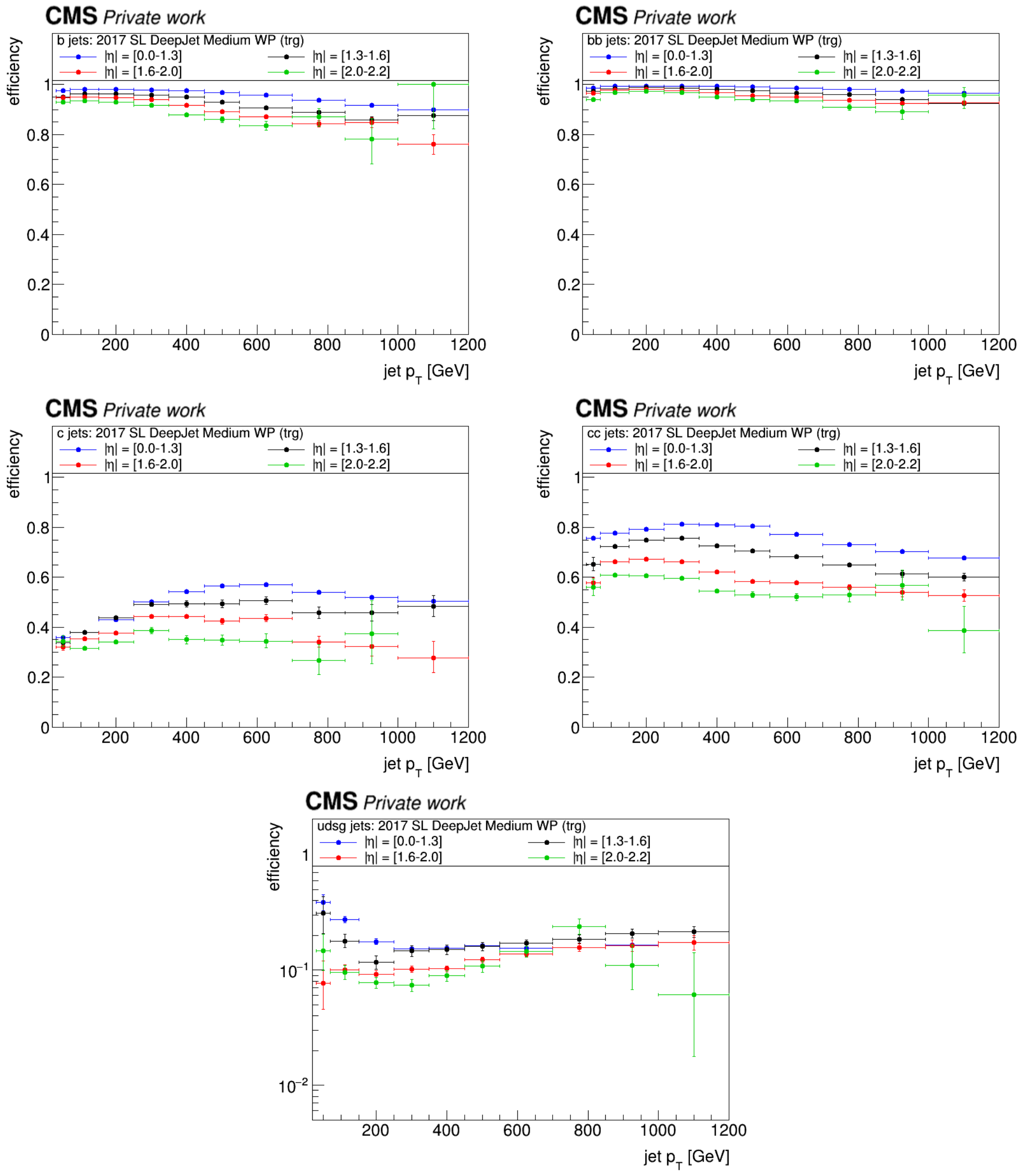


Figure E.1: B-tagging efficiencies for jets that triggered the online b-tagging, corresponding to the medium working point of the DeepJet algorithm and as a function of the jet's p_{T}, η and flavor in the 2017 SL channel. The jet flavors studied are b (top left), a merged bb jet (top right), c (center left), cc (center right), and light flavors udsg (bottom).

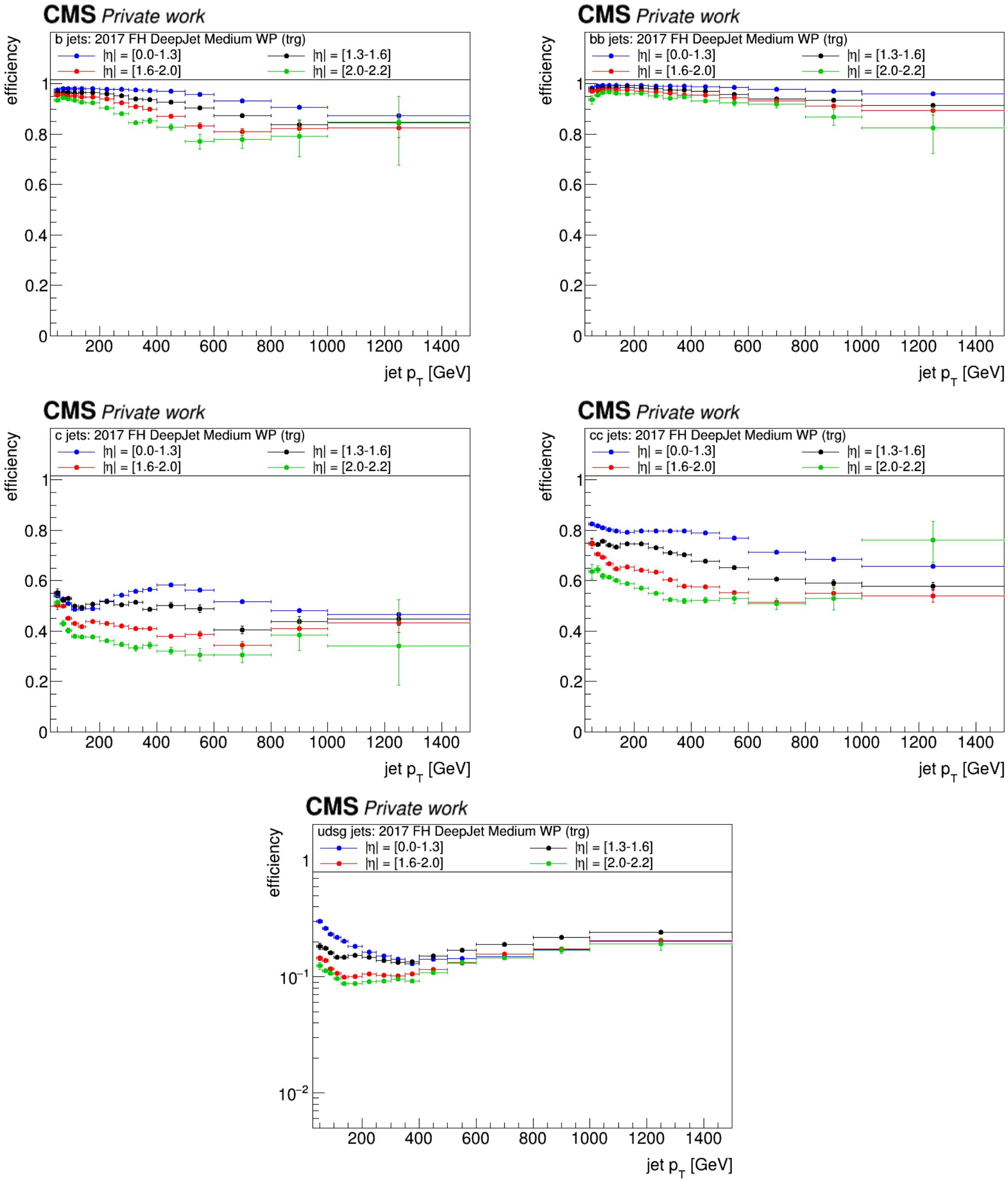


Figure E.2: B-tagging efficiencies for jets that triggered the online b-tagging, corresponding to the medium working point of the DeepJet algorithm and as a function of the jet's p_T, η and flavor in the 2017 FH channel. The jet flavors studied are b (top left), a merged bb jet (top right), c (center left), cc (center right), and light flavors udsg (bottom).

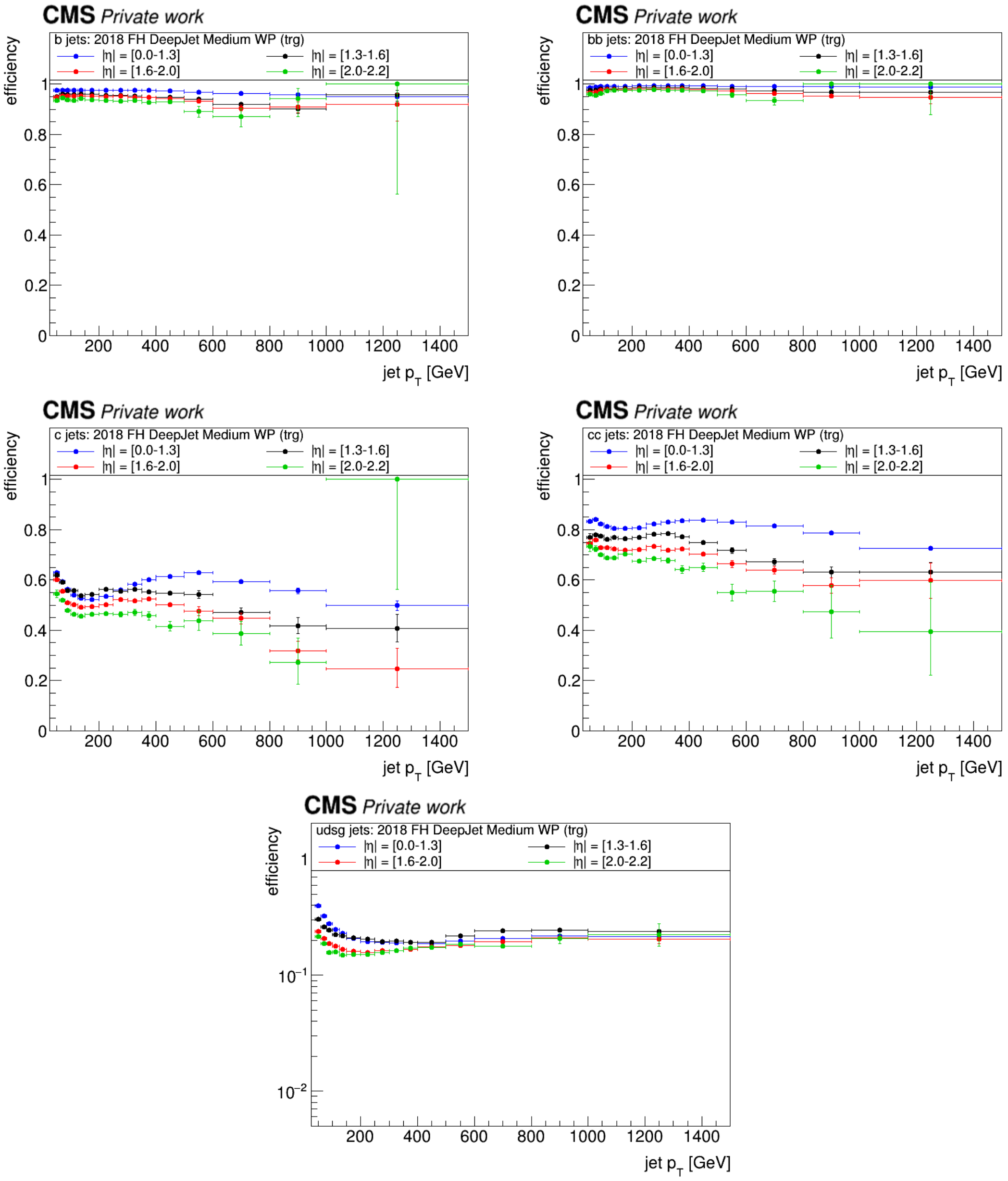


Figure E.3: B-tagging efficiencies for jets that triggered the online b-tagging, corresponding to the medium working point of the DeepJet algorithm and as a function of the jet's p_{T}, η and flavor in the 2018 FH channel. The jet flavors studied are b (top left), a merged bb jet (top right), c (center left), cc (center right), and light flavors udsg (bottom).

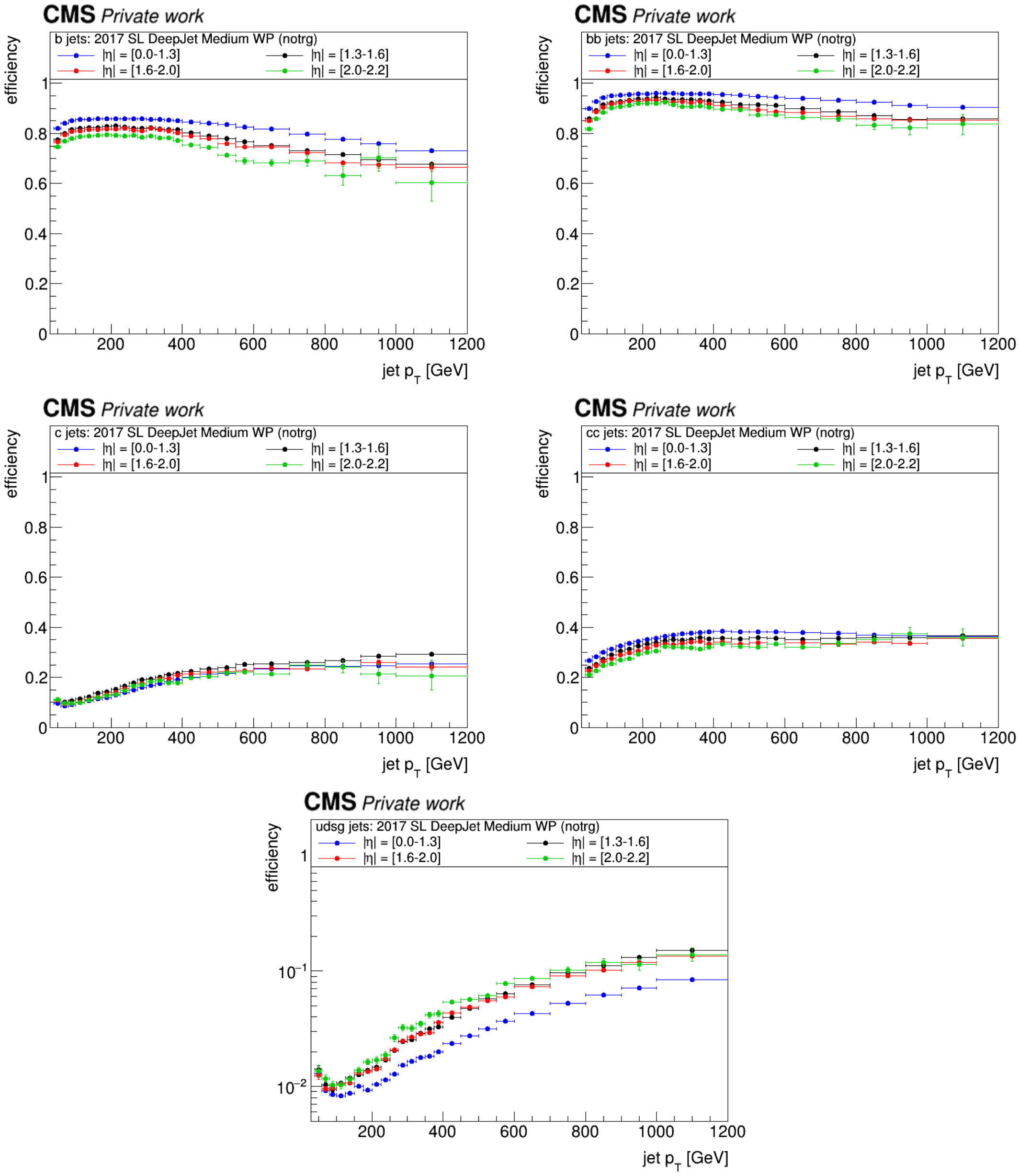


Figure E.4: B-tagging efficiencies for jets that did not fire the online b-tagging, corresponding to the medium working point of the DeepJet algorithm and as a function of the jet's p_{T}, η and flavor in the 2017 SL channel. The jet flavors studied are b (top left), a merged bb jet (top right), c (center left), cc (center right), and light flavors udsg (bottom).

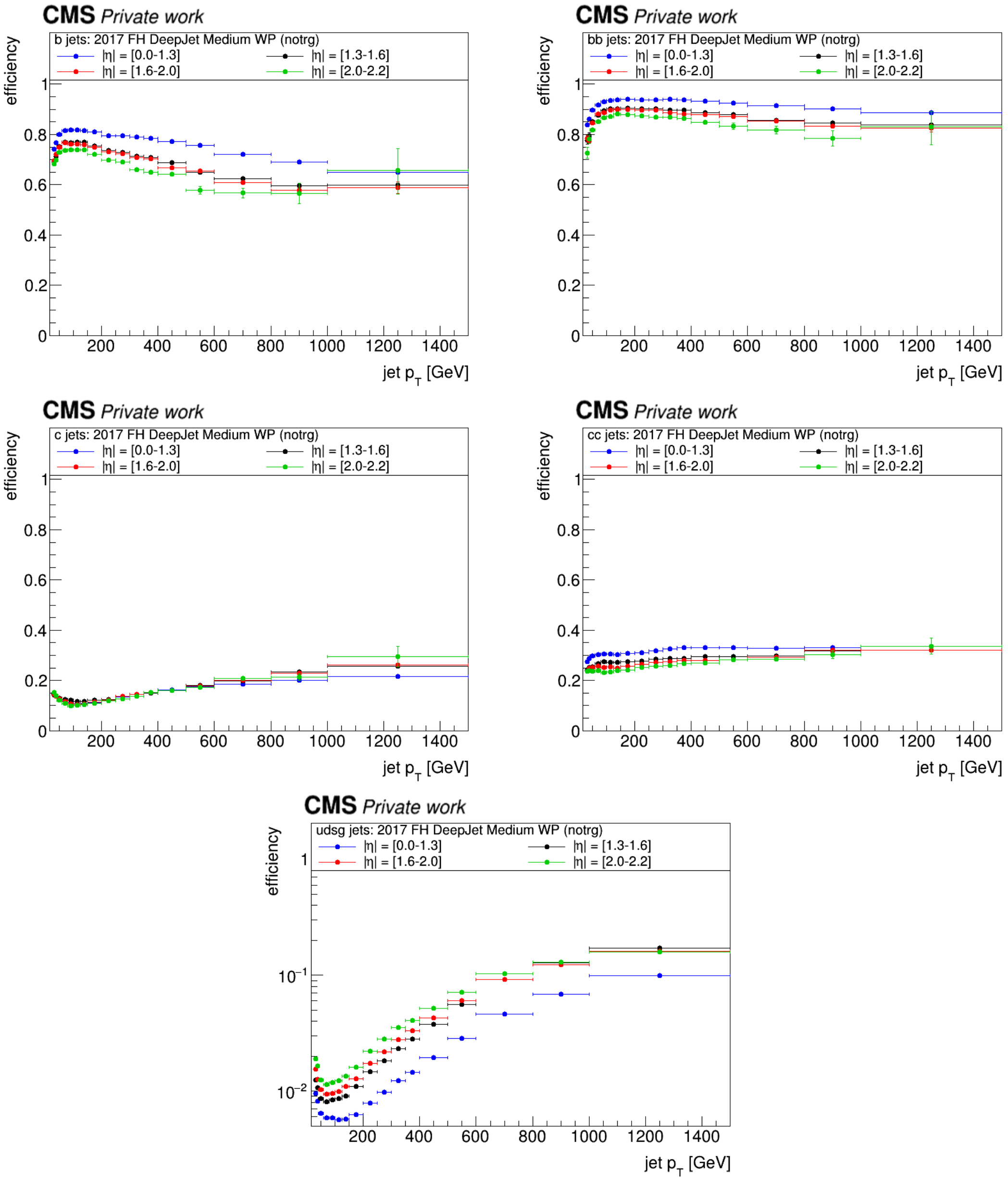


Figure E.5: B-tagging efficiencies for jets that did not fire the online b-tagging, corresponding to the medium working point of the DeepJet algorithm and as a function of the jet's p_{T}, η and flavor in the 2017 FH channel. The jet flavors studied are b (top left), a merged bb jet (top right), c (center left), cc (center right), and light flavors udsg (bottom).

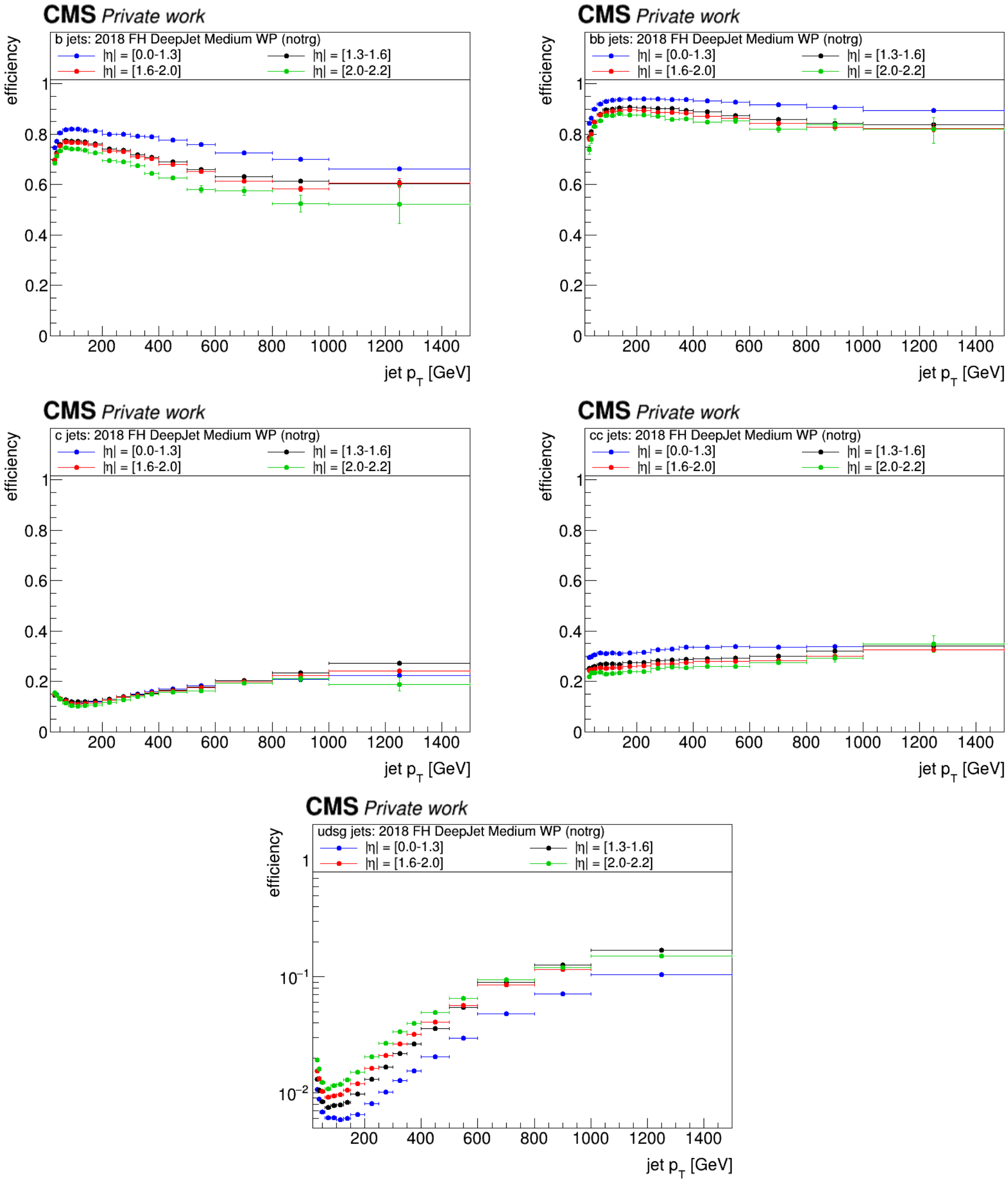


Figure E.6: B-tagging efficiencies for jets that did not fire the online b-tagging, corresponding to the medium working point of the DeepJet algorithm and as a function of the jet's p_T, η and flavor in the 2018 FH channel. The jet flavors studied are b (top left), a merged bb jet (top right), c (center left), cc (center right), and light flavors udsg (bottom).

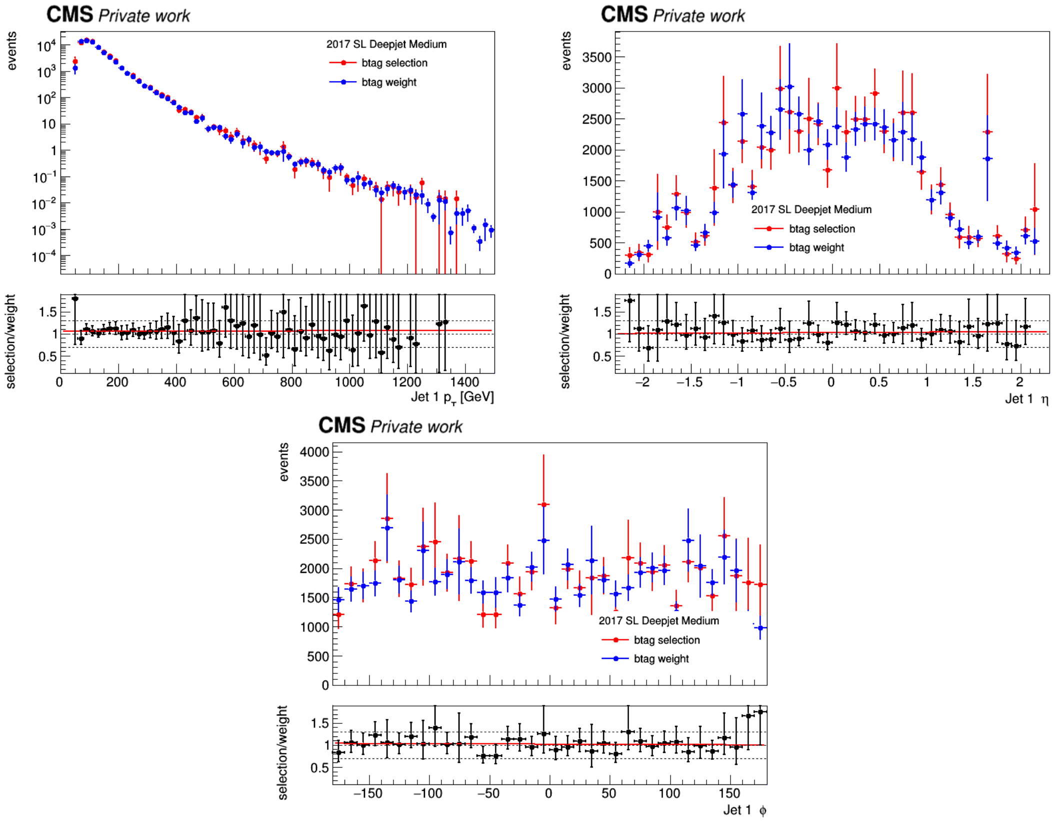


Figure E.7: Closure tests for the b-tag weighting method, showing comparisons of cut- and weight-based distributions of the leading jet's p_{T} (top left), η (top right) and ϕ (bottom) in 2017 SL analysis.

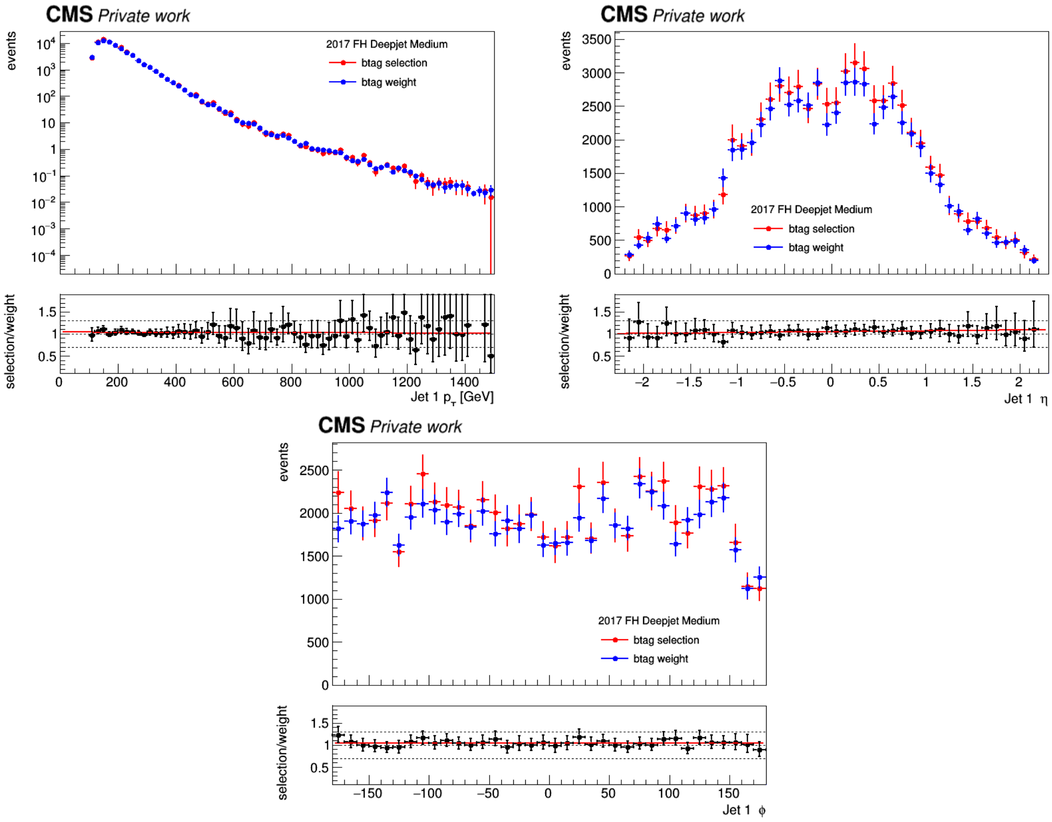


Figure E.8: Closure tests for the b-tag weighting method, showing comparisons of cut- and weight-based distributions of the leading jet's p_{T} (top left), η (top right) and ϕ (bottom) in 2017 FH analysis.

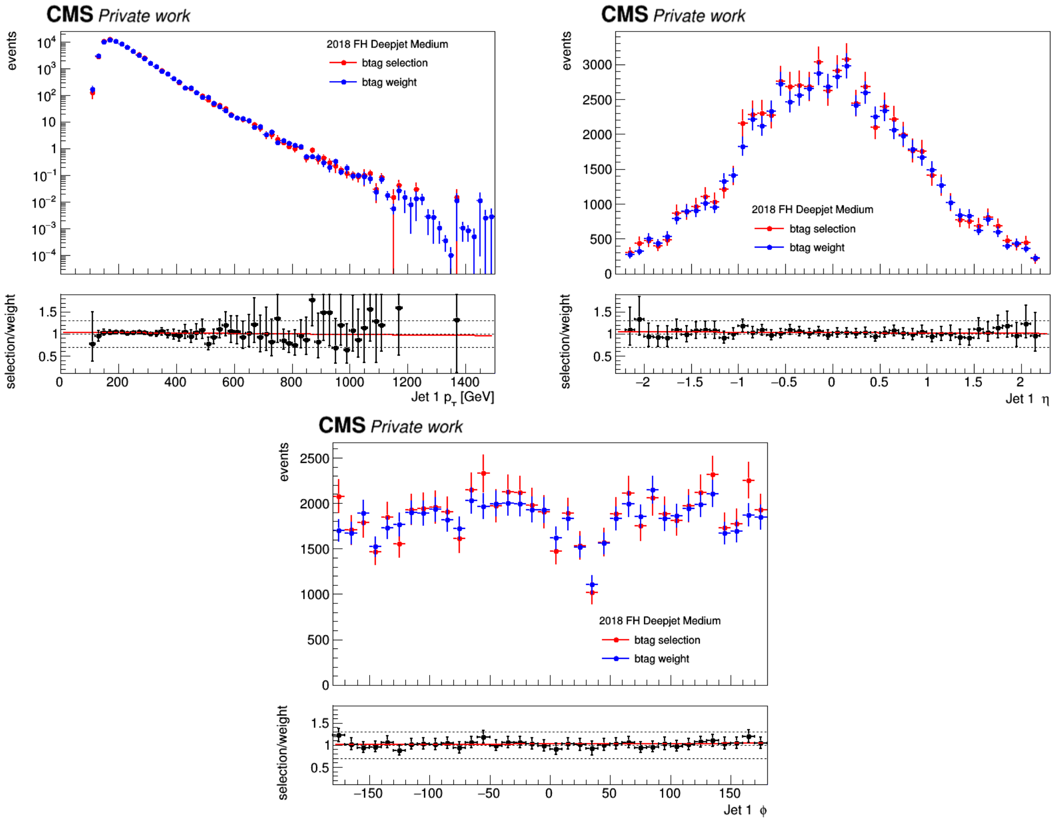


Figure E.9: Closure tests for the b-tag weighting method, showing comparisons of cut- and weight-based distributions of the leading jet's p_T (top left), η (top right) and ϕ (bottom) in 2018 FH analysis.

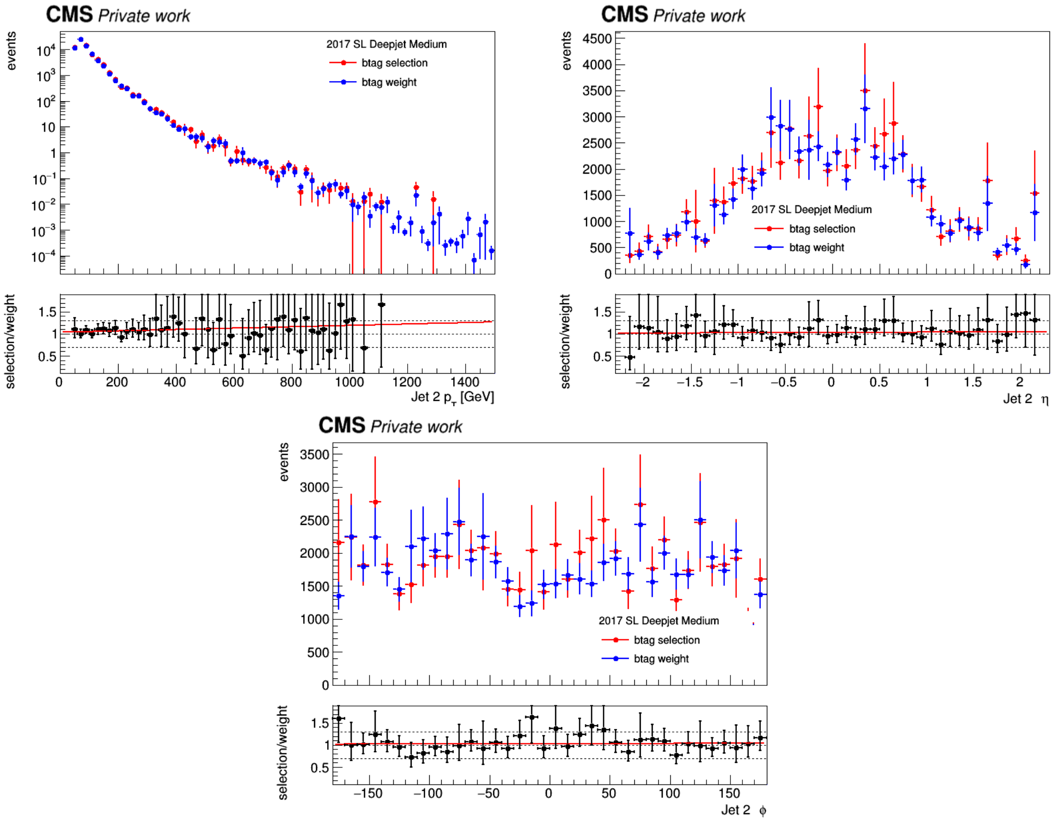


Figure E.10: Closure tests for the b-tag weighting method, showing comparisons of cut- and weight-based distributions of the second leading jet's p_T (top left), η (top right) and ϕ (bottom) in 2017 SL analysis.

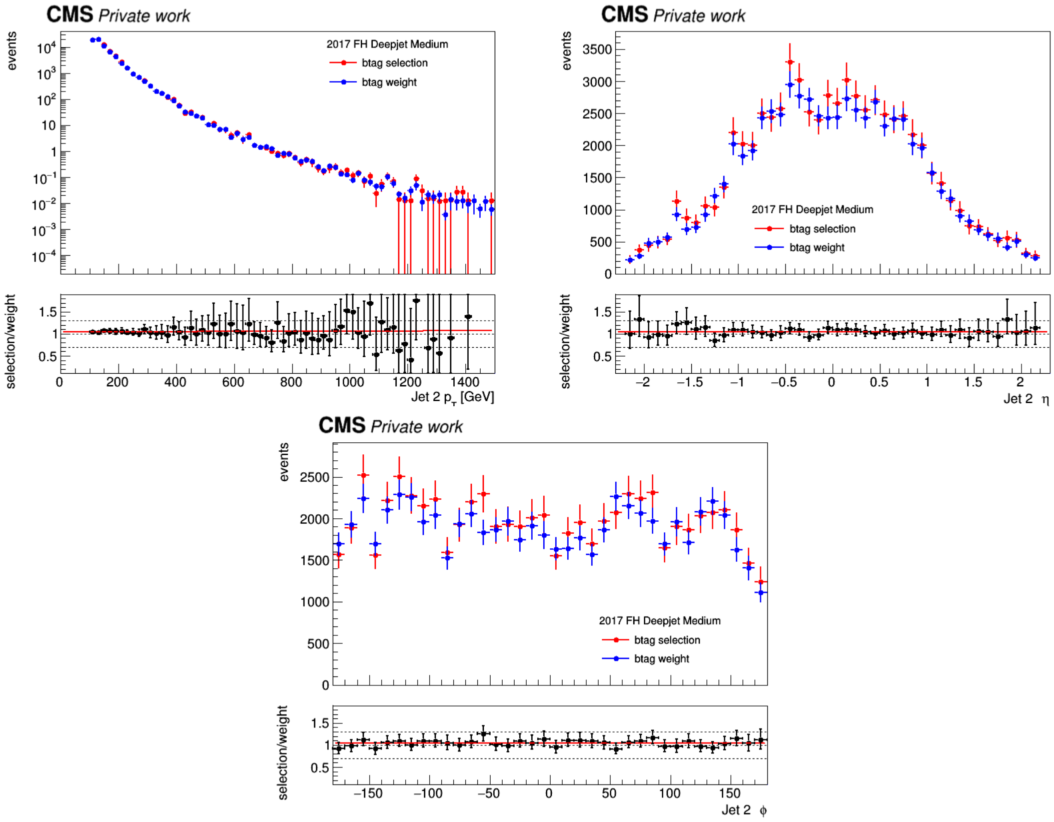


Figure E.11: Closure tests for the b-tag weighting method, showing comparisons of cut- and weight-based distributions of the second leading jet's p_{T} (top left), η (top right) and ϕ (bottom) in 2017 FH analysis.

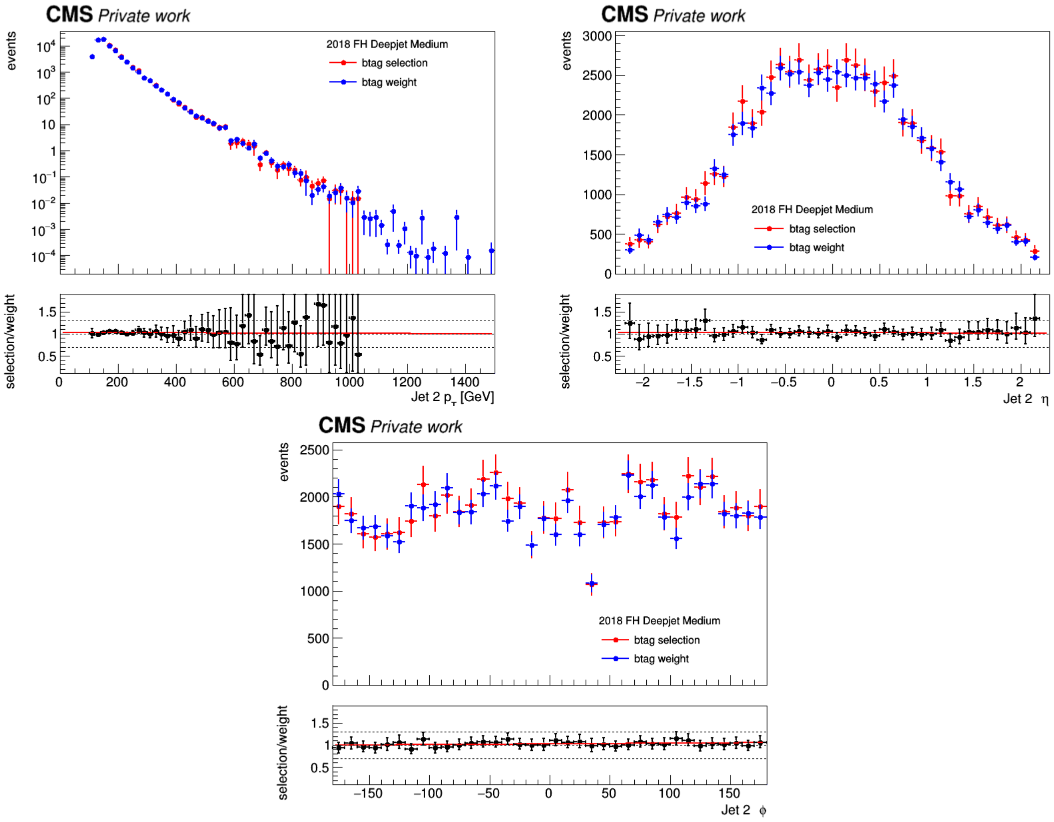


Figure E.12: Closure tests for the b-tag weighting method, showing comparisons of cut- and weight-based distributions of the second leading jet's p_{T} (top left), η (top right) and ϕ (bottom) in 2018 FH analysis.

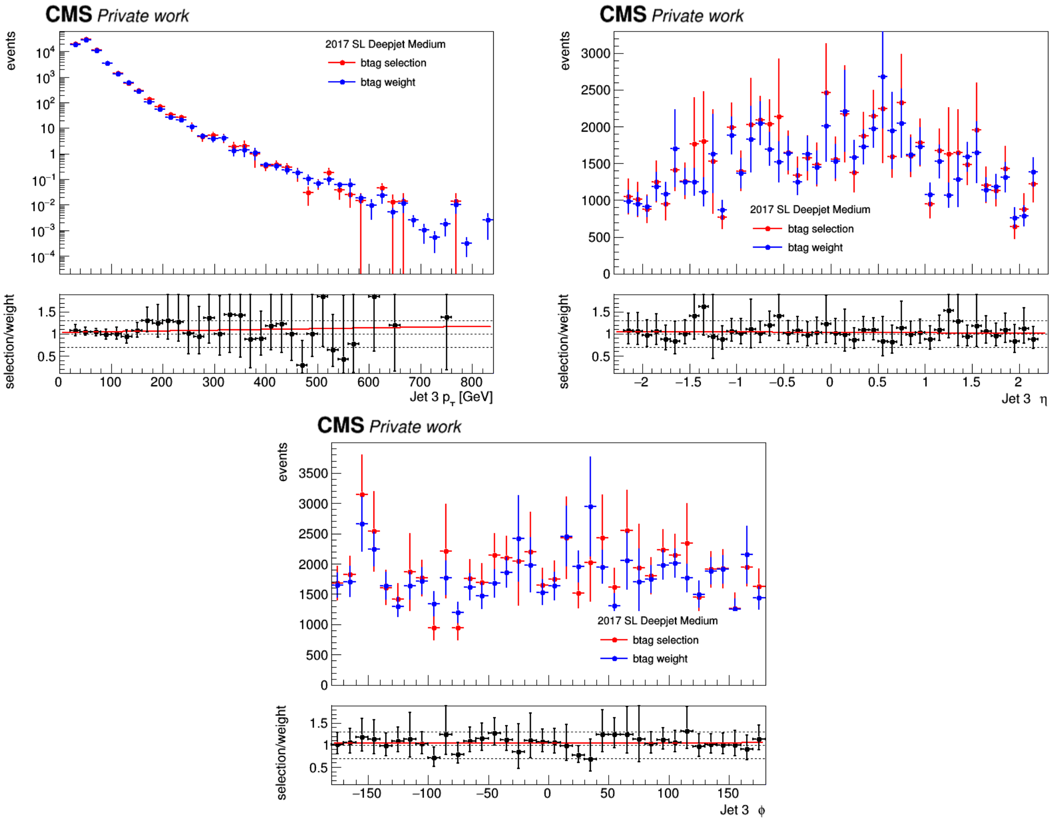


Figure E.13: Closure tests for the b-tag weighting method, showing comparisons of cut- and weight-based distributions of the third leading jet's p_{T} (top left), η (top right) and ϕ (bottom) in 2017 SL analysis.

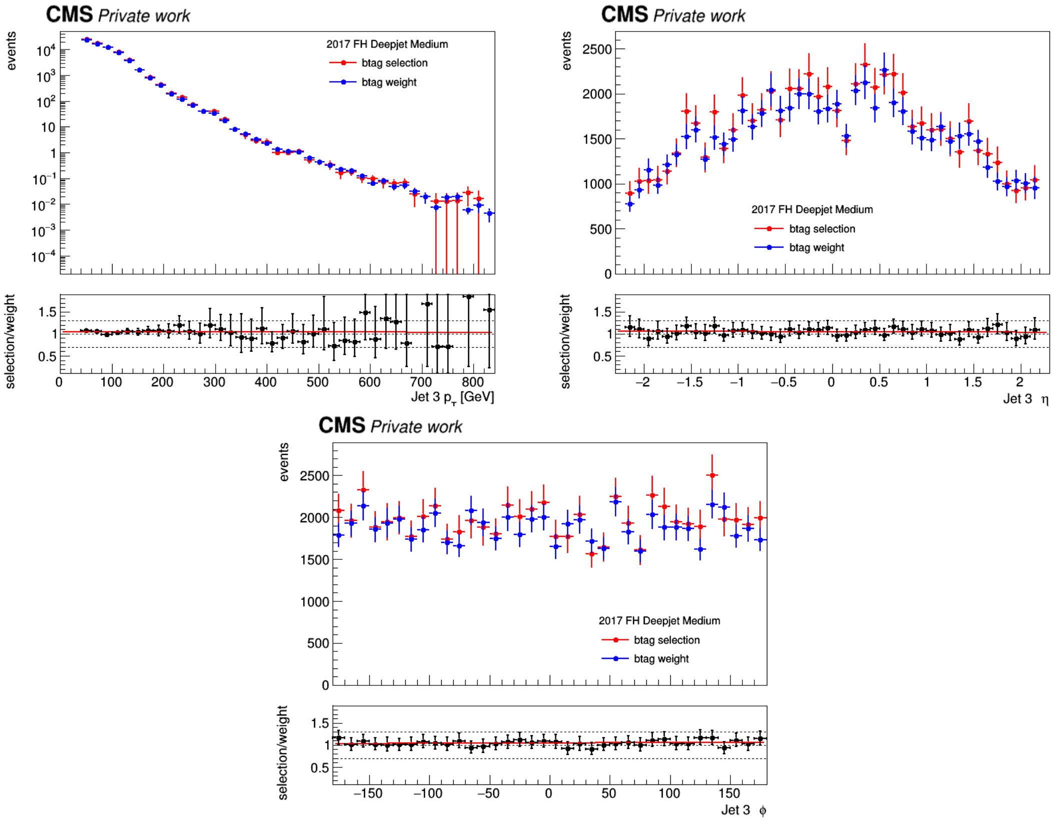


Figure E.14: Closure tests for the b-tag weighting method, showing comparisons of cut- and weight-based distributions of the third leading jet's p_{T} (top left), η (top right) and ϕ (bottom) in 2017 FH analysis.

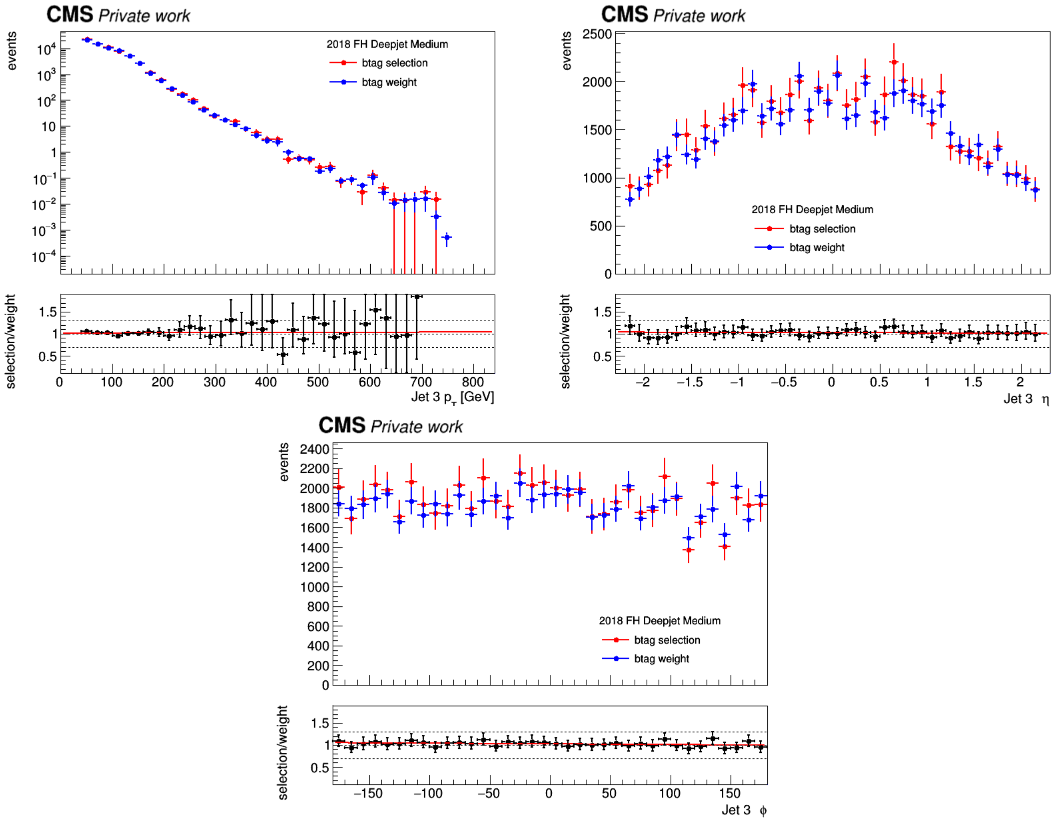


Figure E.15: Closure tests for the b-tag weighting method, showing comparisons of cut- and weight-based distributions of the third leading jet's p_{T} (top left), η (top right) and ϕ (bottom) in 2018 FH analysis.

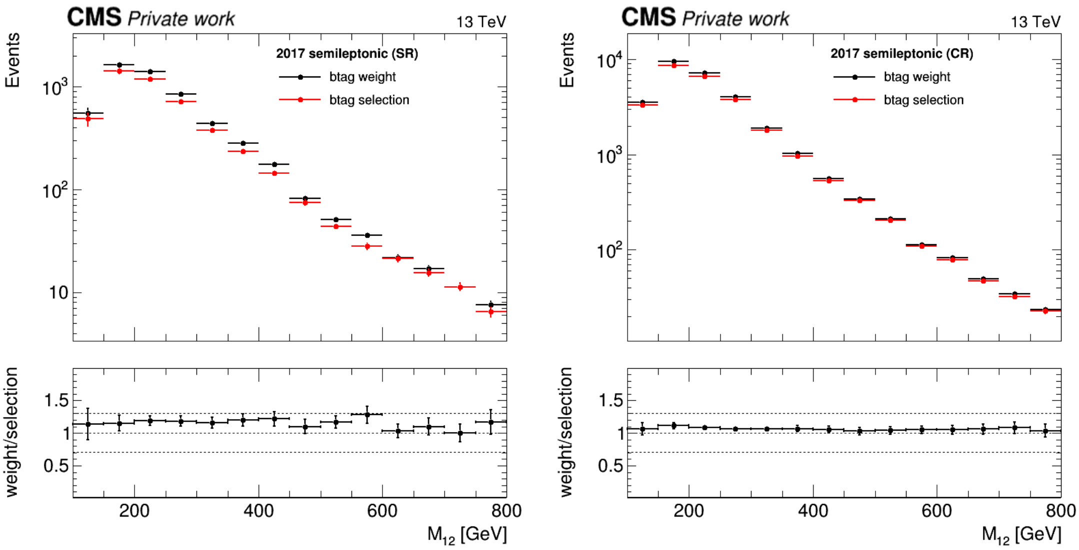


Figure E.16: Closure tests for the b-tag weighting method, showing comparisons of cut- and weight-based m_{12} distributions of 2017 SL analysis in the SR (left) and CR (right)

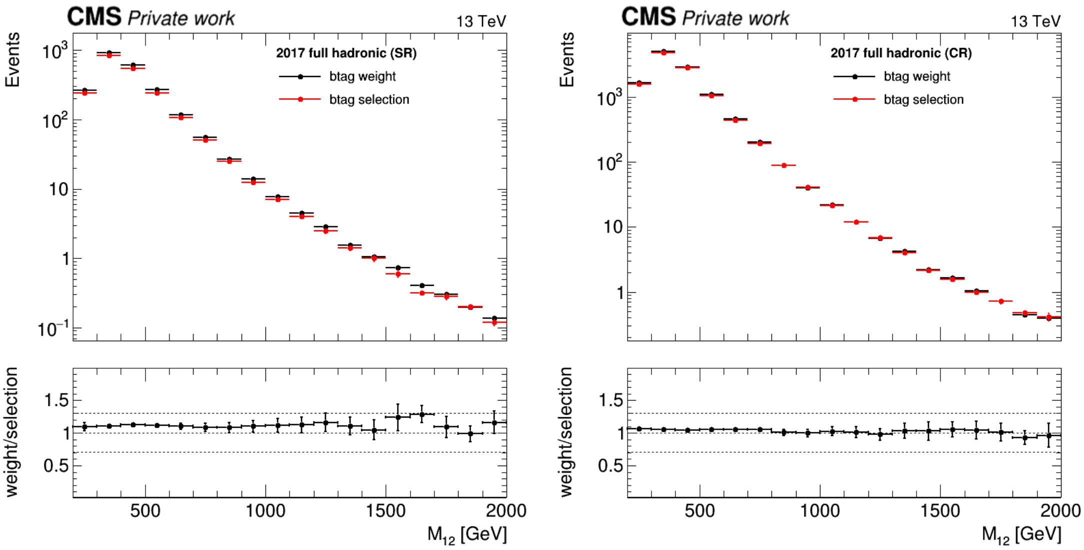


Figure E.17: Closure tests for the b-tag weighting method, showing comparisons of cut- and weight-based m_{12} distributions of 2017 FH analysis in the SR (left) and CR (right)

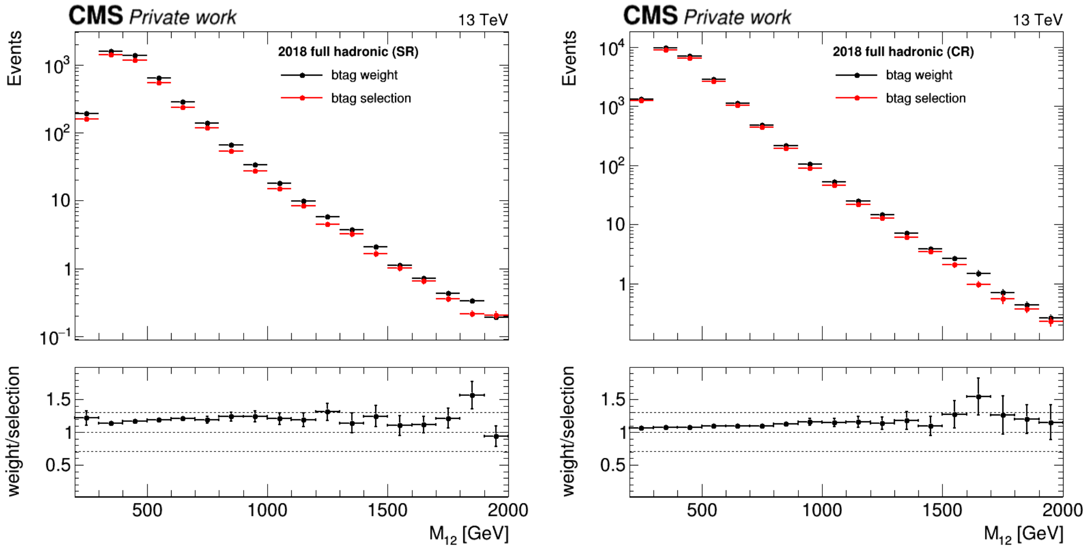


Figure E.18: Closure tests for the b-tag weighting method, showing comparisons of cut- and weight-based m_{12} distributions of 2018 FH analysis in the SR (left) and CR (right)

appendix F

Fisher tests for Transfer Factor studies

In the context of the analysis presented in this thesis, the background modelling in the signal region is described by the shape of a well-defined orthogonal control region and a transfer factor, whose shape is inspired by the ratio of the signal region and the control region in simulation. The transfer factor is parametrized with Chebychev polynomials (see section 5.8.2). As discussed in section 5.8.3, to evaluate possible systematic effects related to the selection of the transfer factor's functional form, different degree polynomials are investigated utilizing a discrete profiling method [210]. To determine which polynomial degrees must be considered, an F-test is performed [211]. A brief description of this method and the results of applying it to the transfer factor's parametrization in 2017 SL, 2017 FH and 2018 FH analyses are provided in this section.

The F-test

The F-test evaluates the null hypothesis that the simpler model f_1, with fewer parameters, is sufficient to explain the variability of the data against the alternative hypothesis that the more complex model f_2 significantly improves the fit. The F-test results are driven from the mathematical definition of the F-statistic:

$$F_{\text{stat}} = \frac{\frac{\chi_1^2 - \chi_2^2}{n_2^{par} - n_1^{par}}}{\frac{\chi_2^2}{ndf_2}}, \tag{F.1}$$

where:

F_{stat} : F-statistic,
χ_1^2 : Chi-squared statistic for the simpler model,
χ_2^2 : Chi-squared statistic for the more complex model,
n_1^{par} : Number of parameters in the simpler model,
n_2^{par} : Number of parameters in the more complex model,
ndf_2 : Number of degrees of freedom in the more complex model.

F_{stat} will follow an F-distribution of the order determined by $(n_2^{par} - n_1^{par}, ndf_2)$. The null hypothesis i rejected if F_{stat} is greater than a critical value F_{crit} for a desired false-rejection probability, which is commonly taken as 5%.

Results
The F-test results can be found in Table F.1, F.2, and F.3 for the transfer factor of the 2017 SL, 2017 FH, and 2018 FH analysis, respectively. Across all data-taking years and fit ranges, increasing beyond a second degree does not provide a significant improvement in the fit.

Fit-range	Polynomial orders tested	F_{stat}	F_{crit}	Conclusion
1	2 vs. 1	0.07	4.13	$F_{\text{stat}} < F_{crit}$
2	2 vs. 1	0.30	4.02	$F_{\text{stat}} < F_{crit}$
3	2 vs. 1	0.17	3.93	$F_{\text{stat}} < F_{crit}$

Table F.1: Summary of F-tests increasing Chebychev Polynomial order for the fit ranges of the 2017 SL analysis. The null hypothesis, or that increasing the polynomial order provides a significant improvement, is rejected only when $F_{\text{stat}} > F_{crit}$, where F_{crit} corresponds to a 5% false-rejection probability.

Fit-range	Polynomial orders tested	F_{stat}	F_{crit}	Conclusion
1	2 vs. 1	0.94	4.17	$F_{\text{stat}} < F_{crit}$
2	2 vs. 1	6.27	4.03	$F_{\text{stat}} > F_{crit}$
	3 vs. 2	0.18	4.04	$F_{\text{stat}} < F_{crit}$
3	2 vs. 1	0.88	3.95	$F_{\text{stat}} < F_{crit}$
4	2 vs. 1	18.50	4.23	$F_{\text{stat}} > F_{crit}$
	3 vs. 2	1.81	4.24	$F_{\text{stat}} < F_{crit}$

Table F.2: Summary of F-tests increasing Chebychev Polynomial order for the fit ranges of the 2017 FH analysis. The null hypothesis, or that increasing the polynomial order provides a significant improvement, is rejected only when $F_{\text{stat}} > F_{crit}$ F_{crit} corresponds to a 5% false-rejection probability.

Fit-range	Polynomial orders tested	F_{stat}	F_{crit}	Conclusion
1	2 vs. 1	1.12	4.21	$F_{\text{stat}} < F_{crit}$
2	2 vs. 1	0.70	4.05	$F_{\text{stat}} < F_{crit}$
3	2 vs. 1	1.35	4.06	$F_{\text{stat}} < F_{crit}$
4	2 vs. 1	5.91	4.20	$F_{\text{stat}} < F_{crit}$
	3 vs. 2	1.32	4.21	$F_{\text{stat}} < F_{crit}$

Table F.3: Summary of F-tests increasing Chebychev Polynomial order for the fit ranges of the 2018 FH analysis. The null hypothesis, or that increasing the polynomial order provides a significant improvement, is rejected only when $F_{\text{stat}} > F_{crit}$, where F_{crit} corresponds to a 5% false-rejection probability.

appendix G

Additional 2HDM plots

This chapter presents the branching fraction of the A and H Higgs bosons, predicted in 2HDM models, as described in section 2.2.2. They were calculated at next-to-leading-order (NLO) precision using the `SusHi v.1.6.1` [108], `2HDMC v.1.7.0` [109] and `LHAPDF v.6.1.6` [110] packages.

Figures G.1, G.2, G.3, and G.4, show the corresponding branching ratios of the H and A bosons with masses of 125 GeV, 400 GeV, 800 GeV and 1.2 TeV, respectively, in the $type-II$ and $flipped$ benchmark scenarios, at a fixed value of tanβ as indicated in the figures.

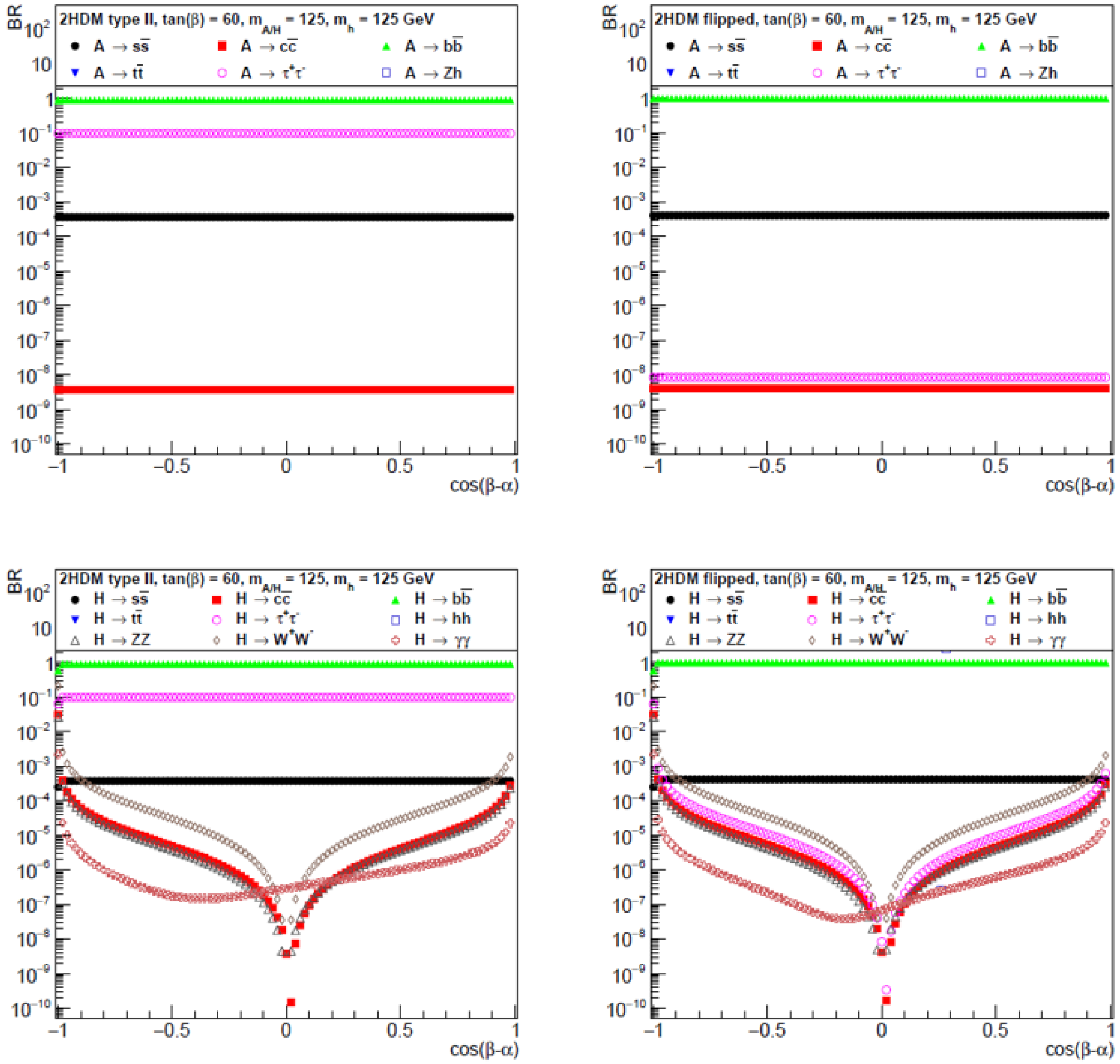


Figure G.1: Branching ratios of the A (top) and H (bottom) Higgs boson in 2HDM $type-II$ (left) and flipped (right) scenarios as a function of $\cos(\beta-\alpha)$, at a Higgs boson mass of 125 GeV and a tanβ value equal to 60. Taken from [39].

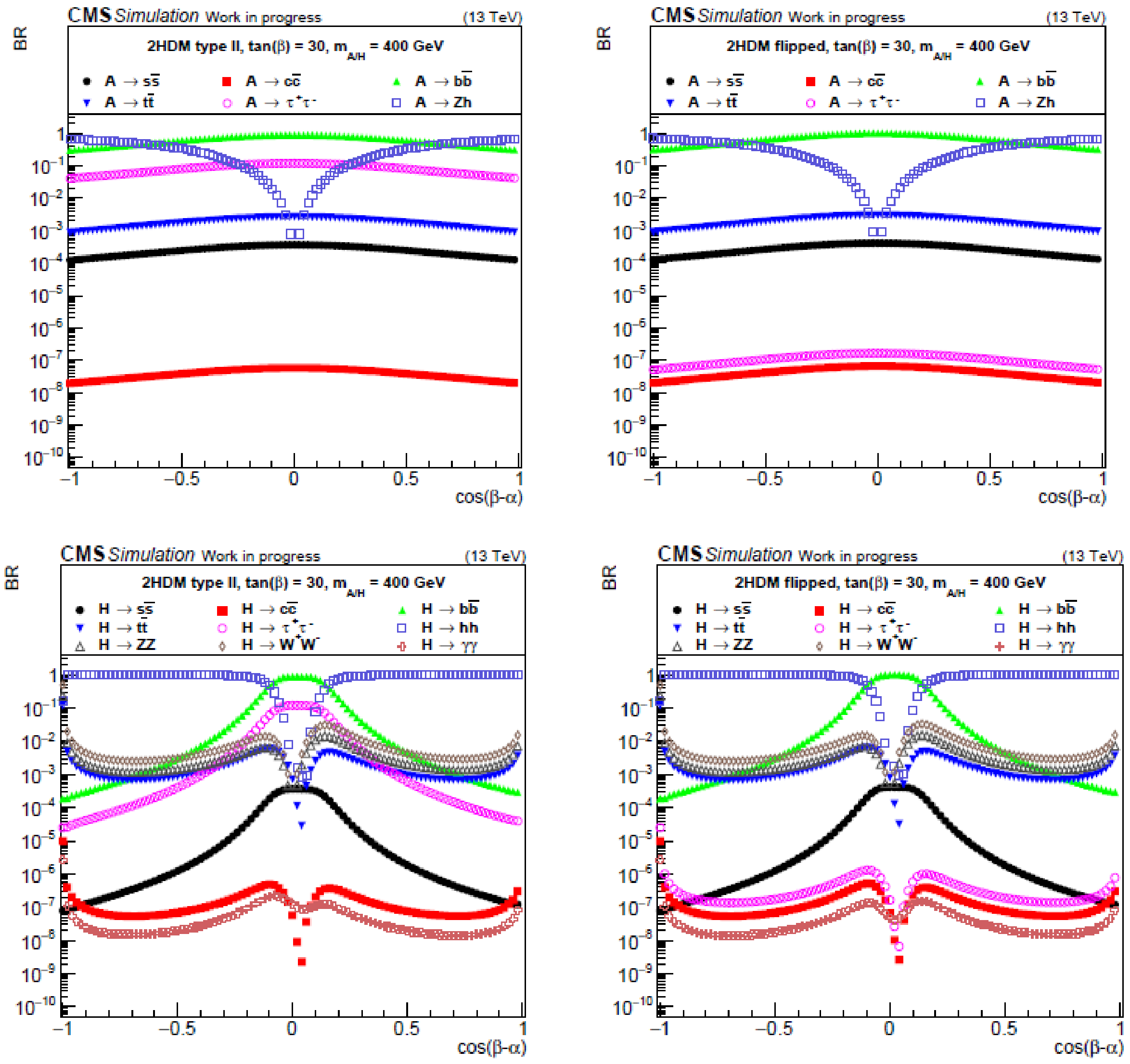


Figure G.2: Branching ratios of the A (top) and H (bottom) Higgs boson in 2HDM $type-II$ (left) and flipped (right) scenarios as a function of $\cos(\beta-\alpha)$, at a Higgs boson mass of 400 GeV and a tanβ value equal to 30. Taken from [40].

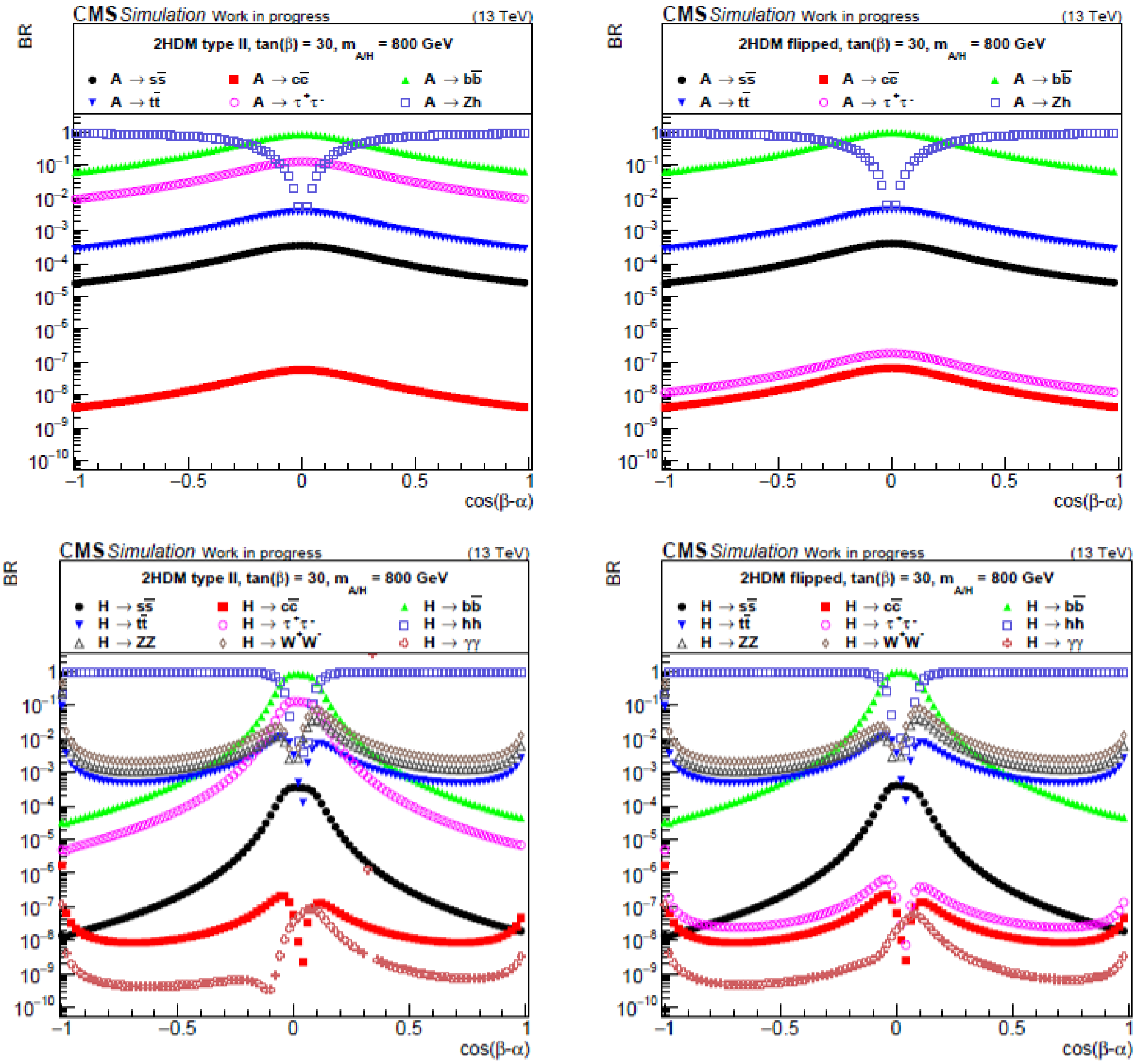


Figure G.3: Branching ratios of the A (top) and H (bottom) Higgs boson in 2HDM *type* $-II$ (left) and flipped (right) scenarios as a function of $\cos(\beta-\alpha)$, at a Higgs boson mass of 800 GeV and a tanβ value equal to 30. Taken from [40].

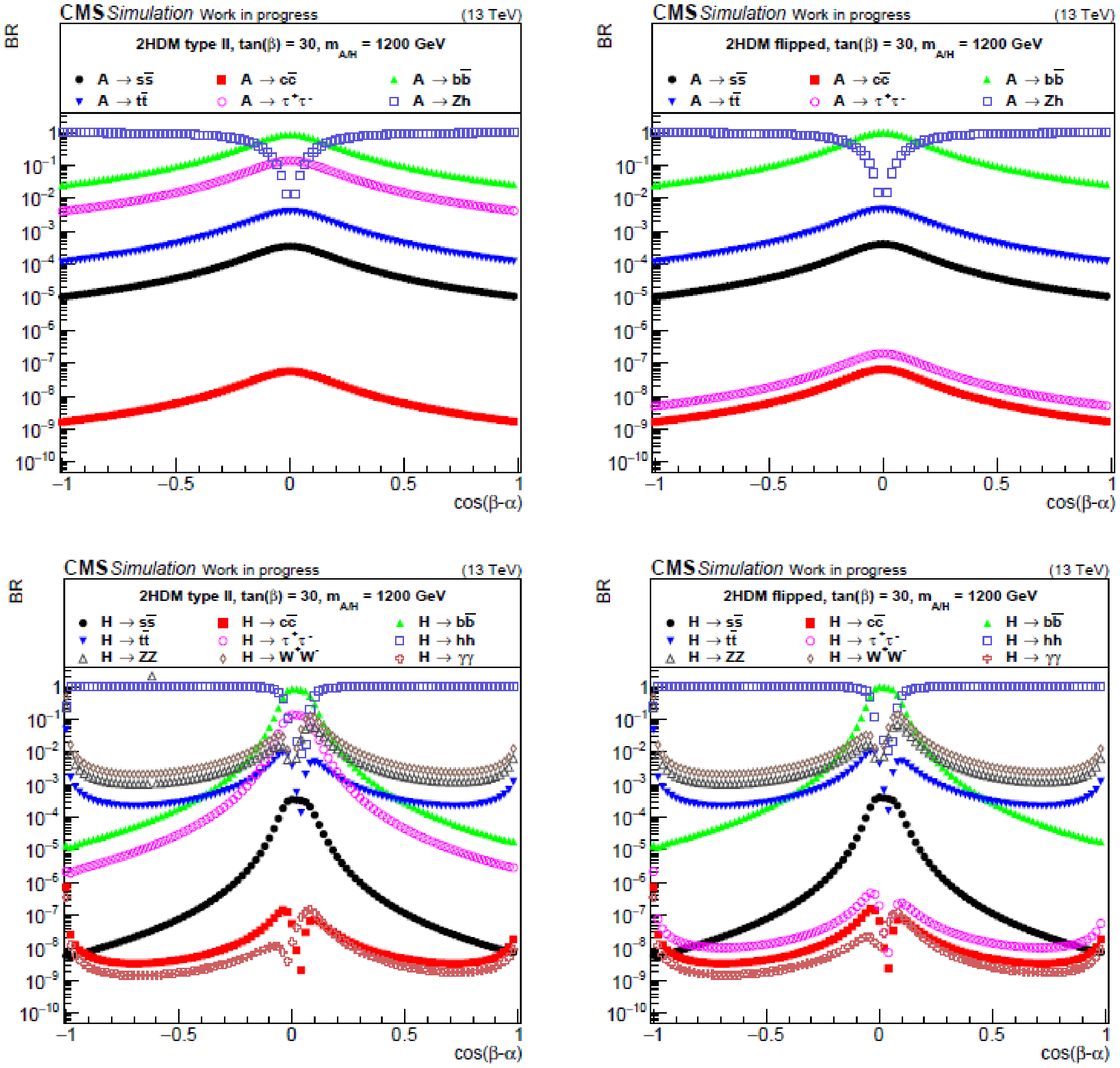


Figure G.4: Branching ratios of the A (top) and H (bottom) Higgs boson in 2HDM *type − II* (left) and flipped (right) scenarios as a function of $\cos(\beta - \alpha)$, at a Higgs boson mass of 1200 GeV and a tanβ value equal to 30. Taken from [40].

appendix H

Upper limits in the production cross-section times branching fraction

Mass [GeV]	-2σ	-1σ	Expected	+1σ	+2σ	Observed
125	37.096	71.487	121.750	189.207	273.352	113.860
130	24.814	46.897	82.500	134.458	206.306	66.302
140	16.523	31.007	52.875	81.749	116.636	43.419
160	11.388	20.846	35.125	53.466	74.929	30.379
180	8.921	19.291	32.625	49.661	69.596	23.994
200	6.987	14.914	26.500	41.816	59.382	18.938
250	10.941	14.225	18.188	23.335	30.953	38.993
300	8.417	10.820	13.812	17.722	23.507	26.030
350	3.767	5.005	6.938	9.730	13.149	4.631
400	2.726	3.651	5.094	7.185	9.813	4.978
450	1.848	2.462	3.453	4.898	6.758	2.333
500	1.282	1.716	2.414	3.463	4.811	1.344
600	0.818	1.097	1.562	2.266	3.218	1.585
700	0.860	1.159	1.656	2.415	3.450	3.168

Table H.1: Expected and observed upper limits at 95% confidence level in the 2017 SL analysis.

Mass [GeV]	-2σ	-1σ	Expected	+1σ	+2σ	Observed
300	11.859	15.821	22.000	30.682	41.292	19.242
350	5.424	7.236	10.062	14.114	19.072	10.802
400	2.855	3.809	5.297	7.450	10.122	5.350
450	1.221	1.630	2.500	3.845	5.460	6.086
500	0.979	1.310	1.844	2.770	4.155	2.920
600	0.624	0.841	1.191	2.094	3.205	1.040
700	0.509	0.687	0.973	1.419	2.014	0.926
800	0.419	0.562	0.801	1.165	1.656	0.708
900	0.339	0.459	0.652	0.951	1.351	0.691
1000	0.208	0.282	0.400	0.616	0.910	0.435
1200	0.155	0.210	0.301	0.449	0.698	0.298
1400	0.118	0.160	0.230	0.347	0.547	0.225
1600	0.091	0.125	0.187	0.305	0.464	0.334
1800	0.078	0.107	0.173	0.286	0.458	0.103

Table H.2: Expected and observed upper limits at 95% confidence level in the 2017 FH analysis.

Mass [GeV]	-2σ	-1σ	Expected	+1σ	+2σ	Observed
300	6.806	9.033	12.625	17.708	24.090	7.284
350	2.622	3.483	4.828	6.714	8.986	3.751
400	1.086	1.456	2.242	3.386	4.687	2.555
450	0.817	1.085	1.504	2.097	2.803	2.227
500	0.640	0.854	1.188	1.661	2.232	1.645
600	0.345	0.458	0.641	0.899	1.227	0.418
700	0.212	0.284	0.396	0.572	0.797	0.237
800	0.147	0.195	0.315	0.480	0.673	0.234
900	0.116	0.156	0.228	0.348	0.489	0.248
1000	0.097	0.129	0.180	0.274	0.385	0.329
1200	0.075	0.102	0.142	0.203	0.279	0.098
1400	0.057	0.077	0.109	0.157	0.220	0.077
1600	0.050	0.068	0.097	0.141	0.199	0.171
1800	0.046	0.065	0.098	0.153	0.227	0.144

Table H.3: Expected and observed upper limits at 95% confidence level in the 2018 FH analysis.

Mass [GeV]	-2σ	-1σ	Expected	+1σ	+2σ	Observed
125	37.236	71.408	123.000	194.091	285.464	114.764
130	24.814	46.897	82.500	133.143	203.963	66.167
140	16.562	31.081	53.000	81.943	117.531	43.488
160	11.457	20.702	35.125	53.606	75.844	30.410
180	8.921	19.291	32.625	49.661	69.983	24.015
200	8.483	11.307	18.250	27.780	39.148	18.228
250	10.979	14.273	18.250	23.560	31.170	39.228
300	4.110	5.448	8.094	11.159	14.469	8.720
350	1.936	2.541	3.516	4.875	6.488	2.540
400	0.953	1.271	1.820	2.756	3.809	2.013
450	0.641	0.851	1.180	1.645	2.229	1.915
500	0.486	0.649	0.902	1.258	1.694	1.088
600	0.278	0.369	0.508	0.714	0.964	0.350
700	0.183	0.243	0.340	0.478	0.665	0.252
800	0.138	0.184	0.285	0.434	0.608	0.207
900	0.106	0.142	0.198	0.303	0.423	0.230
1000	0.087	0.116	0.163	0.233	0.335	0.278
1200	0.068	0.091	0.127	0.181	0.249	0.090
1400	0.051	0.069	0.097	0.139	0.194	0.070
1600	0.044	0.059	0.083	0.120	0.169	0.169
1800	0.039	0.054	0.077	0.119	0.181	0.073

Table H.4: Expected and observed upper limits at 95% confidence level in the Run 2 combination.

appendix I

Transfer Factor fits with third-degree Chebychev polynomial in the FH category

As discussed in section 6.2, after the first unblinding stage of the analysis presented, it was found that a background model offering more flexibility was required in the FH category. A third-degree Chebyshev polynomial was thus included in the discrete profiling method, providing a significant improvement in the goodness-of-fit of several mass-points. Figs. I.1 and I.2 show the parametrization of the transfer factor derived from simulated QCD samples with a third-degree Chebyshev polynomial in the 2017 FH and 2018 FH categories. Exemplary scans of the negative logarithm of the likelihood ratio (NLL) after the inclusion, derived with background-only Asimov data, are shown in I.3.

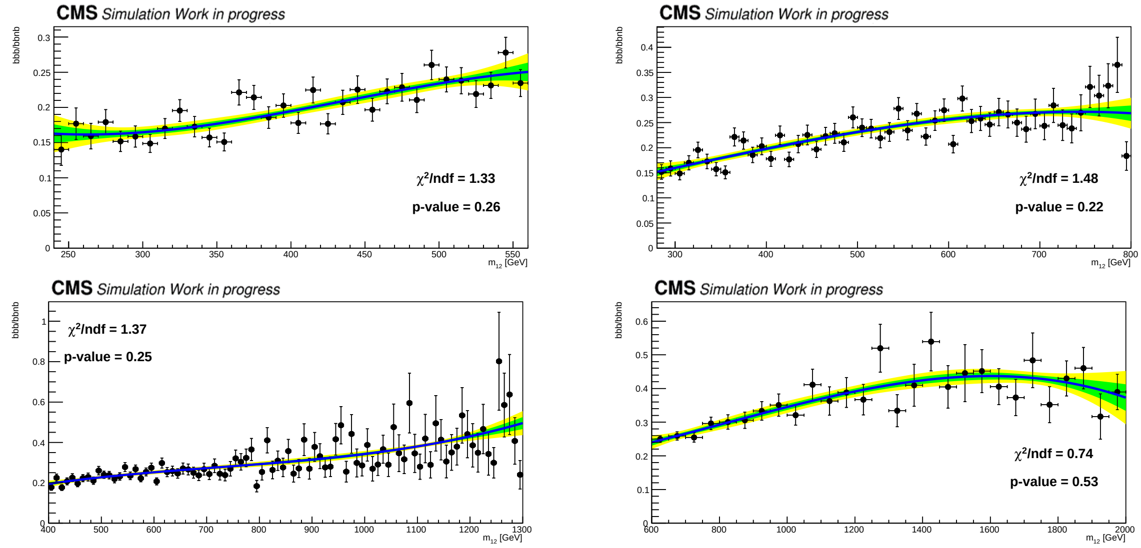


Figure I.1: Ratios of the invariant mass distributions of the SR ("bbb") and CR("bbnb") as determined in QCD multijet simulation, for the fit ranges FR1, FR2 and FR3 for 2017 FH channel. The function selected for the parametrization is a Chebychev polynomial of the third degree. Ultimately, the functional form will be chosen with a discrete profiling method.

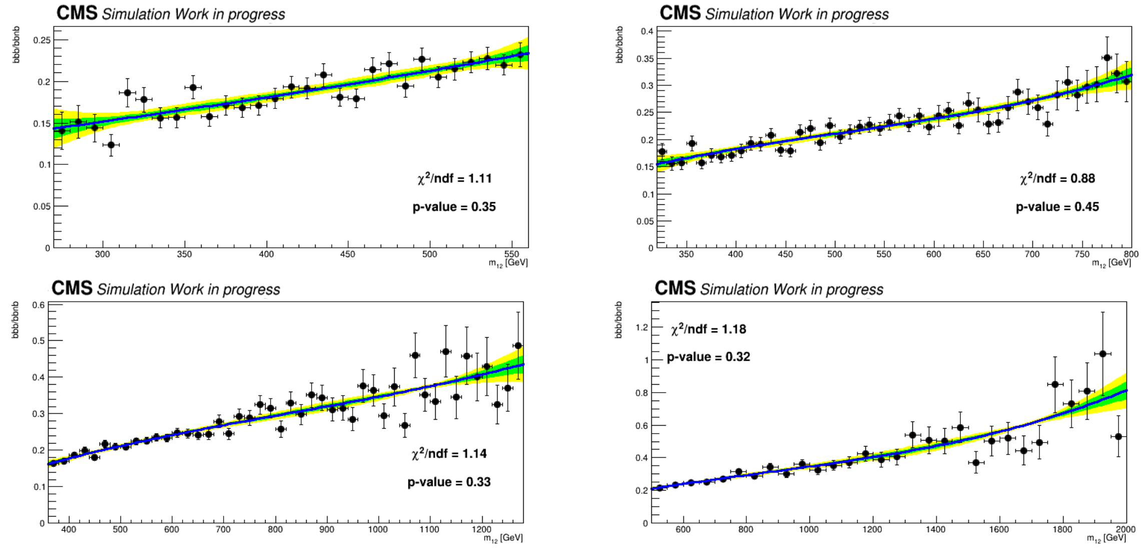


Figure I.2: Ratios of the invariant mass distributions of the SR ("bbb") and CR("bbnb") as determined in QCD multijet simulation, for the fit ranges FR1, FR2, FR3 and 4 for 2018 FH channel. The function selected for the parametrization is a Chebychev polynomial of the third degree. Ultimately, the functional form will be chosen with a discrete profiling method.

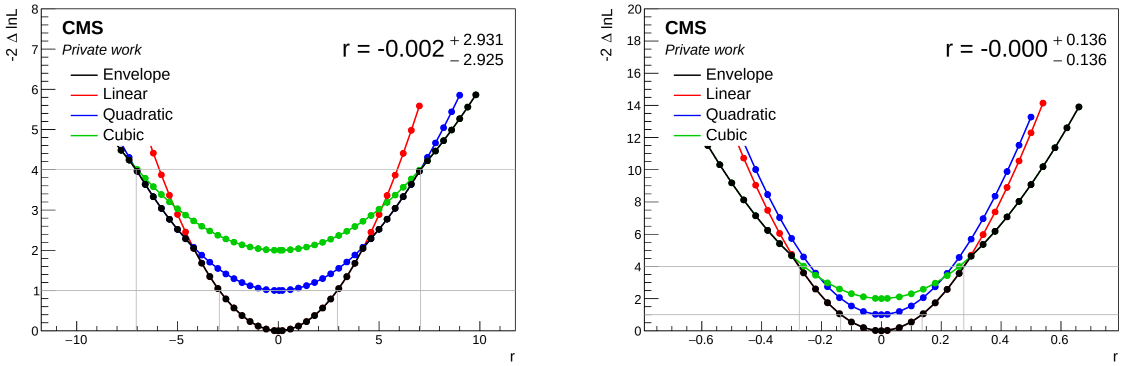


Figure I.3: Profiled scans of twice the negative of the logarithm of the likelihood ratio function at 350 GeV in the 2017 FH analysis (left) and at 800 GeV in the 2018 FH analysis (right).

Bibliography

[1] C. Grojean, “Physics of the Brout–Englert–Higgs boson: Theory”, *Nucl. Part. Phys. Proc.* **273-275** (2016) 11–20, `doi:10.1016/j.nuclphysbps.2015.09.003`.

[2] UA1 Collaboration, “Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s}$=540 GeV”, *Physics Letters B* **122** (1983), no. 1, 103–116, `doi:10.1016/0370-2693(83)91177-2`.

[3] UA2 Collaboration, “Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider”, *Physics Letters B* **122** (1983), no. 5, 476–485, `doi:10.1016/0370-2693(83)91605-2`.

[4] CDS Collaboration, “Observation of top quark production in $\bar{p}p$ collisions with the Collider Detector at Fermilab”, *Physical Review Letters* **74** (1995), no. 14, 2626–2631, `doi:10.1103/physrevlett.74.2626`.

[5] ATLAS Collaboration, “Improved W boson Mass Measurement using 7 TeV Proton-Proton Collisions with the ATLAS Detector”, technical report, CERN, Geneva, 2023.

[6] The ALEPH, DELPHI, L3, and OPAL Collaboration, “Electroweak Measurements in Electron-Positron Collisions at W-Boson-Pair Energies at LEP”, *Phys. Rep.* **532** (2013) 119–244, `doi:10.1016/j.physrep.2013.07.004`, `arXiv:1302.3415`.

[7] CMS Collaboration, “Precision measurement of the Z boson invisible width in pp collisions at $\sqrt{s} = 13$ TeV”, *Phys. Lett. B* **842** (2023) 137563, `doi:10.1016/j.physletb.2022.137563`, `arXiv:2206.07110`.

[8] CMS Collaboration, “Review of top quark mass measurements in CMS”, technical report, CERN, Geneva, 2024. `arXiv:2403.01313`.

[9] CMS Collaboration, “Precision measurements of the top quark mass and decay width using CMS experiment”, technical report, CERN, Geneva, 2024.

[10] M. Baak et al., “The electroweak fit of the standard model after the discovery of a new boson at the LHC”, *The European Physical Journal C* **72** (2012), no. 11, `doi:10.1140/epjc/s10052-012-2205-9`.

[11] ATLAS Collaboration, E. Rossi, “SM (QCD+EW) in ATLAS and CMS”. `https://cds.cern.ch/record/2869578`, 2023. Accessed: 2024-01.

[12] CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”, *Physics Letters B* **716** (2012), no. 1, 30–61, `doi:10.1016/j.physletb.2012.08.021`.

[13] ATLAS Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”, *Physics Letters B* **716** (2012), no. 1, 1–29, `doi:10.1016/j.physletb.2012.08.020`.

[14] CMS Collaboration, “A portrait of the Higgs boson by the CMS experiment ten years after the discovery”, *Nature* (2022) `doi:10.1038/s41586-022-04892-x`, `arXiv:2207.00043`.

[15] G. P. Salam, L.-T. Wang, and G. Zanderighi, “The Higgs boson turns ten”, *Nature* **607** (2022) 41–47, `doi:10.1038/s41586-022-04899-4`.

[16] ATLAS Collaboration, “A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery”, *Nature* **607** (2022) 52–59, `doi:10.1038/s41586-022-04893-w`.

[17] The Super-Kamiokande Collaboration, “Evidence for Oscillation of Atmospheric Neutrinos”, *Physical Review Letters* **81** (1998), no. 8, 1562–1567, `doi:10.1103/physrevlett.81.1562`.

[18] A. Bellerive et al., “The Sudbury Neutrino Observatory”, *Nuclear Physics B* **908** (2016) 30–51, `doi:10.1016/j.nuclphysb.2016.04.035`.

[19] G. ’t Hooft, “Naturalness, chiral symmetry, and spontaneous chiral symmetry breaking”, *NATO Sci. Ser. B* **59** (1980) 135–157, `doi:10.1007/978-1-4684-7571-5_9`.

[20] European Space Agency, “Planck reveals an almost perfect Universe”. `https://www.esa.int/Science_Exploration/Space_Science/Planck/Planck_reveals_an_almost_perfect_Universe`, 2013. Accessed: 2024-02-09.

[21] Planck Collaboration, "Planck 2018 results - VI. Cosmological parameters", *AA* **641** (2020) A6, `doi:10.1051/0004-6361/201833910`.

[22] G. Branco et al., "Theory and phenomenology of two-Higgs-doublet models", *Physics Reports* **516** (2012), no. 1, 1–102, `doi:10.1016/j.physrep.2012.02.002`. Theory and phenomenology of two-Higgs-doublet models.

[23] H. Nilles, "Supersymmetry, supergravity and particle physics", *Physics Reports* **110** (1984), no. 1, 1–162, `doi:https://doi.org/10.1016/0370-1573(84)90008-5`.

[24] S. P. Martin, "A Supersymmetry primer", p. 1–98. World scientific, 1998. `doi:10.1142/9789812839657_0001`.

[25] S. Raby, "Grand Unified Theories", `arXiv:hep-ph/0608183`.

[26] G. Arcadi, A. Djouadi, and M. Raidal, "Dark Matter through the Higgs portal", *Physics Reports* **842** (2020) 1–180, `doi:10.1016/j.physrep.2019.11.003`.

[27] A. B. Lahanas, "LSP as a Candidate for Dark Matter", *Lect. Notes Phys.* **720** (2007) 35–68, `doi:10.1007/978-3-540-71013-4_2`, `arXiv:hep-ph/0607301`.

[28] A. Delgado and M. Quirós, "Higgsino dark matter in the MSSM", *Physical Review D* **103** (2021), no. 1, `doi:10.1103/physrevd.103.015024`.

[29] G. Bhattacharyya et al., "Can measurements of 2HDM parameters provide hints for high scale supersymmetry?", *Phys. Rev. D* **97** (2018) 095018, `doi:10.1103/PhysRevD.97.095018`.

[30] L. Wang et al., "A Concise Review on Some Higgs-Related New Physics Models in Light of Current Experiments", *Universe* **9** (2023), no. 4, `doi:10.3390/universe9040178`.

[31] J. Cao, Z. Heng, J. M. Yang et al., "A SM-like Higgs near 125 GeV in low energy SUSY: a comparative study for MSSM and NMSSM", *Journal of High Energy Physics* **2012** (2012), no. 86, `doi:10.1007/JHEP03(2012)086`.

[32] M. Krawczyk, D. Sokołowska, and B. Świeżewska, "2HDM with Z2 symmetry in light of new LHC data", *Journal of Physics: Conference Series* **447** (2013) 012050, `doi:10.1088/1742-6596/447/1/012050`.

[33] O. S. Brüning et al., "LHC Design Report", `doi:10.5170/CERN-2004-003-V-1`.

[34] CMS Collaboration, "The CMS Experiment at the CERN LHC", *Journal of Instrumentation* **3** (2008), no. 08, S08004, `doi:10.1088/1748-0221/3/08/S08004`.

[35] CMS Collaboration, "Search for beyond the standard model Higgs bosons decaying into a $\mathrm{b}\overline{\mathrm{b}}$ pair in pp collisions at $\sqrt{s} = 13$ TeV", *Journal of High Energy Physics* **2018** (2018), no. 8, `doi:10.1007/jhep08(2018)113`.

[36] E. Bagnaschi et al., "MSSM Higgs boson searches at the LHC: benchmark scenarios for Run 2 and beyond", *The European Physical Journal C* **79** (2019), no. 7, `doi:10.1140/epjc/s10052-019-7114-8`.

[37] CMS Collaboration, "Performance of heavy-flavour jet identification in boosted topologies in proton-proton collisions at $\sqrt{s} = 13$ TeV", technical report, CERN, Geneva, 2023.

[38] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", *Journal of Instrumentation* **13** (2018), no. 05, P05011, `doi:10.1088/1748-0221/13/05/p05011`.

[39] A. Vagnerini, "Search for Higgs bosons in the final state with b-quarks in the semi-leptonic channel with the CMS 2017 data". PhD thesis, Hamburg U., Hamburg, 2020. `doi:10.3204/PUBDB-2020-02277`.

[40] P. Asmuss, "Search for high-mass bosons of an extended Higgs sector in b quark final states using the 2017 data set of the CMS experiment". PhD thesis, Hamburg U., Hamburg, 2021. `doi:10.3204/PUBDB-2021-03020`.

[41] R. Shevchenko, "Search for high mass higgs bosons in the final state with b-quarks with the CMS 2016 data". PhD thesis, Hamburg U., Hamburg, 2019. `doi:10.3204/PUBDB-2019-01182`.

[42] L. O. S. Noehte and on behalf of the Tracker Group of the CMS collaboration, "CMS Phase-1 pixel detector refurbishment during LS2 and readiness towards the LHC Run 3", *Journal of Instrumentation* (2022) `doi:10.1088/1748-0221/17/09/C09017`.

[43] S. L. Glashow, "Partial-symmetries of weak interactions", *Nuclear Physics* **22** (1961), no. 4, 579–588, `doi:https://doi.org/10.1016/0029-5582(61)90469-2`.

[44] A. Salam, "Weak and Electromagnetic Interactions", *Conf. Proc. C* **680519** (1968) 367–377, `doi:10.1142/9789812795915_0034`.

[45] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (1967) 1264–1266, `doi:10.1103/PhysRevLett.19.1264`.

[46] H. D. Politzer, “Reliable Perturbative Results for Strong Interactions?”, *Phys. Rev. Lett.* **30** (1973) 1346–1349, `doi:10.1103/PhysRevLett.30.1346`.

[47] H. Fritzsch, M. Gell-Mann, and H. Leutwyler, “Advantages of the Color Octet Gluon Picture”, *Phys. Lett. B* **47** (1973) 365–368, `doi:10.1016/0370-2693(73)90625-4`.

[48] D. J. Gross and F. Wilczek, “Ultraviolet Behavior of Nonabelian Gauge Theories”, *Phys. Rev. Lett.* **30** (1973) 1343–1346, `doi:10.1103/PhysRevLett.30.1343`.

[49] G. Ecker, “Particles, Fields, Quanta: From Quantum Mechanics to the Standard Model of Particle Physics”, Springer, Cham, 2019, ISBN 978-3-030-14478-4, 978-3-030-14479-1.

[50] M. E. Peskin and D. V. Schroeder, “An Introduction to quantum field theory”, Addison-Wesley, Reading, USA, 1995, ISBN 978-0-201-50397-5.

[51] D. J. Griffiths, “Introduction to elementary particles”, Wiley, New York, 1987, ISBN 0-471-60386-4.

[52] M. Thomson, “Modern particle physics”, Cambridge University Press, New York, 2013, ISBN 978-1-107-03426-6.

[53] C. Burgard and D. Galbraith, “Standard Model of Physics”. `https://texample.net/tikz/examples/model-physics/`, 2016. Accessed: 2024-01.

[54] Particle Data Group Collaboration, “Review of Particle Physics”, *PTEP* **2022** (2022) 083C01, `doi:10.1093/ptep/ptac097`.

[55] J. Schwichtenberg, “Physics from Symmetry”, Undergraduate Lecture Notes in Physics, Springer International Publishing, Cham, 2018, ISBN 978-3-319-66630-3, 978-3-319-66631-0.

[56] E. Noether, “Invariant variation problems”, *Transport Theory and Statistical Physics* **1** (1971), no. 3, 186–207, `doi:10.1080/00411457108231446`.

[57] National Institute of Standards and Technology, “Value of Fine Structure Constant”. `https://physics.nist.gov/cgi-bin/cuu/Value?eqalph|search_for=fine+structure+constant`, 2024. Accessed: 2024-05-09.

[58] H. Fritzsch, “Fundamental Constants at High Energy”, *Fortschritte der Physik* **50** (2002), no. 5–7, 518–524, `doi:10.1002/1521-3978(200205)50:5/7<518::aid-prop518>3.0.co;2-f`.

[59] F. Halzen and A. Martin, "Quarks & Leptons: An introductory course in modern particle physics", John Wiley & Sons, New York, USA, 1984.

[60] C. S. Wu et al., "Experimental Test of Parity Conservation in Beta Decay", *Phys. Rev.* **105** (1957) 1413–1415, `doi:10.1103/PhysRev.105.1413`.

[61] LHCb Collaboration, "Erratum to "Observation of CP violation in B±→DK± decays"", *Physics Letters B* **713** (2012), no. 3, 351, `doi:10.1016/j.physletb.2012.05.060`.

[62] M. Kobayashi and T. Maskawa, "CP-Violation in the Renormalizable Theory of Weak Interaction", *Progress of Theoretical Physics* **49** (1973), no. 2, 652–657, `doi:10.1143/PTP.49.652`.

[63] T. Nakano and K. Nishijima, "Charge Independence for V-particles", *Progress of Theoretical Physics* **10** (1953), no. 5, 581–582, `doi:10.1143/PTP.10.581`.

[64] M. Gell-Mann, "The interpretation of the new particles as displaced charge multiplets", *Nuovo Cim.* **4** (1956), no. S2, 848–866, `doi:10.1007/BF02748000`.

[65] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.* **13** (1964) 321–323, `doi:10.1103/PhysRevLett.13.321`.

[66] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (1964) 508–509, `doi:10.1103/PhysRevLett.13.508`.

[67] P. W. Higgs, "Broken symmetries, massless particles and gauge fields", *Phys. Lett.* **12** (1964) 132–133, `doi:10.1016/0031-9163(64)91136-9`.

[68] P. W. Higgs, "Spontaneous Symmetry Breakdown without Massless Bosons", *Phys. Rev.* **145** (1966) 1156–1163, `doi:10.1103/PhysRev.145.1156`.

[69] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global Conservation Laws and Massless Particles", *Phys. Rev. Lett.* **13** (1964) 585–587, `doi:10.1103/PhysRevLett.13.585`.

[70] J. Ellis, M. K. Gaillard, and D. V. Nanopoulos, "A Historical Profile of the Higgs Boson. An Updated Historical Profile of the Higgs Boson", `doi:10.1142/9789814733519_0014`, `arXiv:1504.07217`.

[71] F. Mandl and G. Shaw, "Quantum Field Theory", A Wiley-Interscience publication, Wiley, 2010, ISBN 978-0-4714-9683-0.

[72] CMS Collaboration, "Observation of $t\bar{t}H$ Production", *Phys. Rev. Lett.* **120** (2018) 231801, `doi:10.1103/PhysRevLett.120.231801`.

[73] ATLAS Collaboration, "Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector", *Physics Letters B* **784** (2018) 173–191, `doi:10.1016/j.physletb.2018.07.035`.

[74] CMS Collaboration, "Search for the SM Higgs boson produced in association with bottom quarks in final states with leptons", technical report, CERN, Geneva, 2024.

[75] CERN, "CERN Yellow Reports: Monographs, Vol 2 (2017): Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector", 2017. `doi:10.23731/CYRM-2017-002`.

[76] CMS Collaboration, "Search for Higgs boson decays to invisible particles produced in association with a top-quark pair or a vector boson in proton-proton collisions at $\sqrt{s} = 13$ TeV and combination across Higgs production modes", technical report, CERN, Geneva, 2022.

[77] J. B. Hartle, "Quantum Mechanics at the Planck Scale", `arXiv:gr-qc/9508023`.

[78] C. Rovelli, "Quantum Gravity", Cambridge Monographs on Mathematical Physics, Cambridge University Press, 2004, ISBN 978-0-5218-3733-0.

[79] CMS Collaboration, "Combined measurements of Higgs boson couplings in proton–proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$", *The European Physical Journal C* **79** (2019), no. 5, `doi:10.1140/epjc/s10052-019-6909-y`.

[80] L. Girardello and M. Grisaru, "Soft breaking of supersymmetry", *Nuclear Physics B* **194** (1982), no. 1, 65–76, `doi:10.1016/0550-3213(82)90512-0`.

[81] I. Jack and D. Jones, "Non-standard soft supersymmetry breaking", *Physics Letters B* **457** (1999), no. 1–3, 101–108, `doi:10.1016/s0370-2693(99)00530-4`.

[82] M. Carena and H. Haber, "Higgs Boson theory and phenomenology", *Progress in Particle and Nuclear Physics* **50** (2003), no. 1, 63–152, `doi:10.1016/s0146-6410(02)00177-1`.

[83] M. Carena, S. Heinemeyer, C. E. M. Wagner, and G. Weiglein, "MSSM Higgs boson searches at the Tevatron and the LHC: Impact of different benchmark scenarios", *The European Physical Journal C* **45** (2006), no. 3, 797–814, `doi:10.1140/epjc/s2005-02470-y`.

[84] CERN LHC Physics Working Group, "LHC Physics Working Group: MSSM Neutral Higgs Bosons". `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHWGMSSMNeutral`, 2024. Accessed: 1 April 2024.

[85] M. Carena et al., "MSSM Higgs boson searches at the LHC: benchmark scenarios after the discovery of a Higgs-like particle", *The European Physical Journal C* **73** (2013), no. 9, `doi:10.1140/epjc/s10052-013-2552-1`.

[86] H. Bahl et al., "HL-LHC and ILC sensitivities in the hunt for heavy Higgs bosons", *The European Physical Journal C* **80** (2020), no. 10, `doi:10.1140/epjc/s10052-020-08472-z`.

[87] LEP Collaboration, "Search for neutral MSSM Higgs bosons at LEP", *The European Physical Journal C* **47** (2006), no. 3, `doi:10.1140/epjc/s2006-02569-7`.

[88] A. Djouadi and J. Quevillon, "The MSSM Higgs sector at a high M_{SUSY}: reopening the low tanβ regime and heavy Higgs searches", *JHEP* **10** (2013) 028, `doi:10.1007/JHEP10(2013)028`, `arXiv:1304.1787`.

[89] L. Maiani, A. D. Polosa, and V. Riquer, "Bounds to the Higgs Sector Masses in Minimal Supersymmetry from LHC Data", *Phys. Lett. B* **724** (2013) 274–277, `doi:10.1016/j.physletb.2013.06.026`, `arXiv:1305.2172`.

[90] A. Djouadi et al., "The post-Higgs MSSM scenario: Habemus MSSM?", *Eur. Phys. J. C* **73** (2013) 2650, `doi:10.1140/epjc/s10052-013-2650-0`, `arXiv:1307.5205`.

[91] A. Djouadi et al., "Fully covering the MSSM Higgs sector at the LHC", *JHEP* **06** (2015) 168, `doi:10.1007/JHEP06(2015)168`, `arXiv:1502.05653`.

[92] R. Harlander, M. Krämer, and M. Schumacher, "Bottom-quark associated Higgs-boson production: reconciling the four- and five-flavour scheme approach", `arXiv:1112.3478`.

[93] S. Forte, D. Napoletano, and M. Ubiali, "Higgs production in bottom-quark fusion in a matched scheme", *Physics Letters B* **751** (2015) 331–337, `doi:10.1016/j.physletb.2015.10.051`.

[94] S. Forte, D. Napoletano, and M. Ubiali, "Higgs production in bottom-quark fusion: Matching beyond leading order", *Physics Letters B* **763** (2016) 190–196, `doi:10.1016/j.physletb.2016.10.040`.

[95] S. Dittmaier, M. Krämer, and M. Spira, "Higgs radiation off bottom quarks at the Tevatron and the CERN LHC", *Phys. Rev. D* **70** (2004) 074010, `doi:10.1103/PhysRevD.70.074010`, `arXiv:hep-ph/0309204`.

[96] S. Dawson, C. B. Jackson, L. Reina, and D. Wackeroth, "Exclusive Higgs boson production with bottom quarks at hadron colliders", *Physical Review D* **69** (2004), no. 7, `doi:10.1103/physrevd.69.074027`.

[97] R. V. Harlander and W. B. Kilgore, "Higgs boson production in bottom quark fusion at next-to-next-to-leading order", *Physical Review D* **68** (2003), no. 1, `doi:10.1103/physrevd.68.013001`.

[98] A. Djouadi, J. Kalinowski, and M. Spira, "HDECAY: a Program for Higgs Boson Decays in the Standard Model and its Supersymmetric Extension", *Comput. Phys. Commun.* **108** (1998), no. 1, 56–74, `doi:10.1016/S0010-4655(97)00123-9`.

[99] A. Djouadi, M. M. Muhlleitner, and M. Spira, "Decays of Supersymmetric Particles: the program SUSY-HIT (SUspect-SdecaY-Hdecay-InTerface)", `arXiv:hep-ph/0609292`.

[100] G. Degrassi et al., "Towards high-precision predictions for the MSSM Higgs sector", *The European Physical Journal C* **28** (2003), no. 1, 133–143, `doi:10.1140/epjc/s2003-01152-2`.

[101] M. Frank et al., "The Higgs boson masses and mixings of the complex MSSM in the Feynman-diagrammatic approach", *Journal of High Energy Physics* **2007** (2007), no. 02, 047, `doi:10.1088/1126-6708/2007/02/047`.

[102] S. Heinemeyer, W. Hollik, and G. Weiglein, "FeynHiggs: a program for the calculation of the masses of the neutral CP-even Higgs bosons in the MSSM", *Computer Physics Communications* **124** (2000), no. 1, 76–89, `doi:10.1016/s0010-4655(99)00364-1`.

[103] S. Heinemeyer, W. Hollik, and G. Weiglein, "The masses of the neutral CP-even Higgs bosons in the MSSM: Accurate analysis at the two-loop level", *The European Physical Journal C* **9** (1999), no. 2, 343–366, `doi:10.1007/s100529900006`.

[104] S. L. Glashow and S. Weinberg, "Natural Conservation Laws for Neutral Currents", *Phys. Rev. D* **15** (1977) 1958, `doi:10.1103/PhysRevD.15.1958`.

[105] E. A. Paschos, "Diagonal Neutral Currents", *Phys. Rev. D* **15** (1977) 1966, `doi:10.1103/PhysRevD.15.1966`.

[106] H. E. Haber and O. Stål, "New LHC benchmarks for the CP-conserving two-Higgs-doublet model", *The European Physical Journal C* **75** (2015), no. 10, `doi:10.1140/epjc/s10052-015-3697-x`.

[107] X.-J. Xu, "Tree-level vacuum stability of two-Higgs-doublet models and new constraints on the scalar potential", *Physical Review D* **95** (2017), no. 11, `doi:10.1103/physrevd.95.115019`.

[108] R. V. Harlander, S. Liebler, and H. Mantler, "SusHi: A program for the calculation of Higgs production in gluon fusion and bottom-quark annihilation in the Standard Model and the MSSM", *Computer Physics Communications* **184** (2013), no. 6, 1605–1617, `doi:10.1016/j.cpc.2013.02.006`.

[109] D. Eriksson, J. Rathsman, and O. Stål, "2HDMC - two-Higgs-doublet model calculator", *Computer Physics Communications* **181** (2010), no. 1, 189–205, `doi:10.1016/j.cpc.2009.09.011`.

[110] A. Buckley et al., "LHAPDF6: parton density access in the LHC precision era", *The European Physical Journal C* **75** (2015), no. 3, `doi:10.1140/epjc/s10052-015-3318-8`.

[111] R. W. Assmann, M. Lamont, and S. Myers, "A Brief History of the LEP Collider", *Nucl. Phys. B, Proc. Suppl.* **109** (2002) 17–31, `doi:10.1016/S0920-5632(02)90005-8`.

[112] M. Benedikt et al., "LHC Design Report", `doi:10.5170/CERN-2004-003-V-3`.

[113] L. Arnaudon et al., "Linac4 Technical Design Report", technical report, CERN, Geneva, 2006.

[114] E. Mobs, "The CERN accelerator complex - August 2018. Complexe des accélérateurs du CERN - Août 2018". `https://cds.cern.ch/record/2636343`, 2018. General Photo.

[115] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *Journal of Instrumentation* **3** (2008), no. 08, S08002, `doi:10.1088/1748-0221/3/08/S08002`.

[116] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *Journal of Instrumentation* **3** (2008), no. 08, S08003, `doi:10.1088/1748-0221/3/08/S08003`.

[117] LHCb Collaboration, "The LHCb Detector at the LHC", *Journal of Instrumentation* **3** (2008), no. 08, S08005, `doi:10.1088/1748-0221/3/08/S08005`.

[118] S. van der Meer, “Calibration of the effective beam height in the ISR”, technical report, CERN, Geneva, 1968.

[119] CMS Collaboration, “CMS Public Luminosity Results”. `https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults`, 2024. Accessed: 2024-04-08.

[120] O. Aberle et al., “High-Luminosity Large Hadron Collider (HL-LHC): Technical design report”, `doi:10.23731/CYRM-2020-0010`.

[121] CMS Collaboration, “The Phase-2 Upgrade of the CMS Tracker”, technical report, CERN, Geneva, 2017. `doi:10.17181/CERN.QZ28.FLHW`.

[122] CMS Collaboration, “A MIP Timing Detector for the CMS Phase-2 Upgrade”, technical report, CERN, Geneva, 2019.

[123] CMS Collaboration, “The Phase-2 Upgrade of the CMS Muon Detectors”, technical report, CERN, Geneva, 2017.

[124] CMS Collaboration, “The Phase-2 Upgrade of the CMS Endcap Calorimeter”, technical report, CERN, Geneva, 2017. `doi:10.17181/CERN.IV8M.1JY2`.

[125] CMS Collaboration, “The CMS Experiment at the CERN LHC”, *JINST* **3** (2008) S08004, `doi:10.1088/1748-0221/3/08/S08004`.

[126] CMS Collaboration, “CMS Experiment: Detector”. `https://cms.cern/index.php/detector`. Accessed: 2024-04-11.

[127] CMS Collaboration, “The CMS trigger in Run 2”, technical report, CERN, Geneva, 2018. `doi:10.22323/1.314.0523`.

[128] The Tracker Group of the CMS Collaboration, “The CMS Phase-1 Pixel Detector Upgrade”, `arXiv:2012.14304`.

[129] Leyva Pernia, D. and Bayat Makou, M. on behalf of the CMS Collaboration, “Performance of the CMS Tracker during Run 3”, 2024. `doi:10.22323/1.449.0591`.

[130] “DPG Results TRK - CMS Public Wiki”. `https://twiki.cern.ch/twiki/bin/view/CMSPublic/DPGResultsTRK`. Accessed: 2024-04-12.

[131] CMS Collaboration, “The CMS Phase-1 Pixel Detector Upgrade”, *JINST* **16** (2021), no. 02, P02027, `doi:10.1088/1748-0221/16/02/P02027`, `arXiv:2012.14304`.

[132] CMS Collaboration, "Development of the CMS detector for the CERN LHC Run 3", `arXiv:2309.05466`.

[133] CMS Collaboration, "Commissioning and performance of the CMS pixel tracker with cosmic ray muons", *Journal of Instrumentation* **5** (2010), no. 03, T03007, `doi:10.1088/1748-0221/5/03/t03007`.

[134] L. Borrello, A. Messineo, E. Focardi, and A. Macchiolo, "Sensor Design for the CMS Silicon Strip Tracker", technical report, 8, 2003.

[135] P. Azzurri, "The CMS Silicon Strip Tracker", *Journal of Physics: Conference Series* **41** (May, 2006) 127–134, `doi:10.1088/1742-6596/41/1/011`.

[136] CMS Collaboration, "Commissioning and Performance of the CMS Silicon Strip Tracker with Cosmic Ray Muons", *JINST* **5** (2010) T03008, `doi:10.1088/1748-0221/5/03/T03008`, `arXiv:0911.4996`.

[137] R. J. Bainbridge et al., "Local reconstruction software for the CMS silicon strip tracker", *J. Phys. Conf. Ser.* **119** (2008) 022009, `doi:10.1088/1742-6596/119/2/022009`.

[138] "The CMS tracker system project: Technical Design Report", Technical design report. CMS, CERN, Geneva, 1997.

[139] CMS Collaboration, "Strategies and performance of the CMS silicon tracker alignment during LHC Run 2", *Nucl. Inst and Meth. A* **1037** (2022) 166795, `doi:10.1016/j.nima.2022.166795`.

[140] CMS Collaboration, "First Tracker Alignment results with 2021 cosmic ray data",. Accessed: 2024-05.

[141] CMS Collaboration, "Tracker alignment performance in 2021",. Accessed: 2024-05.

[142] CMS Collaboration, "Tracker alignment performance in early 2023",. Accessed: 2024-05.

[143] C. W. Fabjan and F. Gianotti, "Calorimetry for Particle Physics", *Rev. Mod. Phys.* **75** (2003) 1243–1286, `doi:10.1103/RevModPhys.75.1243`.

[144] G. F. Knoll, "Radiation Detection and Measurement", John Wiley & Sons, 2000.

[145] The CMS Collaboration, "Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC", *Journal of Instrumentation* **16** (2021), no. 05, P05014, `doi:10.1088/1748-0221/16/05/p05014`.

[146] CMS Collaboration, “Performance of the CMS Electromagnetic Calorimeter in LHC Run2”, technical report, CERN, Geneva, 2020.

[147] CMS Collaboration, “The CMS ECAL performance with examples”, technical report, CERN, Geneva, 2014. `doi:10.1088/1748-0221/9/02/C02008`.

[148] CMS Collaboration, “Performance and operation of the CMS electromagnetic calorimeter”, *Journal of Instrumentation* **5** (2010), no. 03, T03010, `doi:10.1088/1748-0221/5/03/t03010`.

[149] M. Chadeeva and N. Lychkovskaya, “Calibration of the CMS hadron calorimeter in Run 2”, *Journal of Instrumentation* **13** (2018), no. 03, C03025, `doi:10.1088/1748-0221/13/03/C03025`.

[150] CMS Collaboration, “Performance of the CMS hadron calorimeter with cosmic ray muons and LHC beam data”, *Journal of Instrumentation* **5** (2010), no. 03, T03012, `doi:10.1088/1748-0221/5/03/T03012`.

[151] CMS Collaboration, “Performance of the local reconstruction algorithms for the CMS hadron calorimeter with Run 2 data”, *Journal of Instrumentation* **18** (2023), no. 11, P11017, `doi:10.1088/1748-0221/18/11/p11017`.

[152] CMS Collaboration, “Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV”, *Journal of Instrumentation* **7** (2012), no. 10, P10002, `doi:10.1088/1748-0221/7/10/P10002`.

[153] CMS Collaboration, “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at s=13 TeV”, *Journal of Instrumentation* **13** (2018), no. 06, P06015, `doi:10.1088/1748-0221/13/06/p06015`.

[154] CMS Collaboration, “The CMS trigger system”, *Journal of Instrumentation* **12** (2017), no. 01, P01020, `doi:10.1088/1748-0221/12/01/p01020`.

[155] CMS Collaboration, “CMS Technical Design Report for the Level-1 Trigger Upgrade”, technical report, 2013.

[156] CMS Collaboration, “The CMS trigger system”, *JINST* **12** (2017), no. 01, P01020, `doi:10.1088/1748-0221/12/01/P01020`, `arXiv:1609.02366`.

[157] CMS Collaboration, “New CMS trigger strategies for the Run 3 of the LHC”, technical report, CERN, Geneva, 2023.

[158] CMS Collaboration, “Commissioning CMS online reconstruction with GPUs”. `https://cds.cern.ch/record/2851656`, 2022. Accessed: 2024-03.

[159] CMS Collaboration, "Data Scouting and Data Parking with the CMS High level Trigger", *PoS* **EPS-HEP2019** (2020) 139, `doi:10.22323/1.364.0139`.

[160] Worldwide LHC Computing Grid, "WLCG - Worldwide LHC Computing Grid". `https://wlcg.web.cern.ch/`, 2024. Accessed: 2024-04-15.

[161] V. Azzolini et al., "The Data Quality Monitoring software for the CMS experiment at the LHC: past, present and future", *EPJ Web Conf.* **214** (2019) 02003, `doi:10.1051/epjconf/201921402003`.

[162] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *Journal of Instrumentation* **12** (2017), no. 10, P10003, `doi:10.1088/1748-0221/12/10/P10003`.

[163] GEANT4 Collaboration, "GEANT4–a simulation toolkikt. GEANT4. A Simulation toolkit", *Nucl. Instrum. Methods Phys. Res., A* **506** (2003), no. 3, 250–303, `doi:10.1016/S0168-9002(03)01368-8`.

[164] GEANT4 Collaboration, "Recent developments in Geant4", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **835** (2016) 186–225, `doi:https://doi.org/10.1016/j.nima.2016.06.125`.

[165] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *Journal of Instrumentation* **9** (2014), no. 10, P10009, `doi:10.1088/1748-0221/9/10/p10009`.

[166] P. Billiori, "Progressive track recognition with a kalman-like fitting procedure", *Comput. Phys. Commun.* **57** (1989) 390–394, `doi:10.1016/0010-4655(89)90249-X`.

[167] P. Billiori and S. Qian, "Simultaneous pattern recognition and track fitting by the kalman filtering method", *Nucl. Instrum. Methods Phys. Res. A* **294** (1990) 219–228, `doi:10.1016/0168-9002(90)91835-Y`.

[168] R. Mankel, "A concurrent track evolution algorithm for pattern recognition in the hera-b main tracking system", *Nucl. Instrum. Methods Phys. Res. A* **395** (1997) 169–184, `doi:10.1016/S0168-9002(97)00705-5`.

[169] R. Fruhwirth, "Application of kalman filtering to track and vertex fitting", *Nucl. Instrum. Methods Phys. Res. A* **262** (1987) 444–450, `doi:10.1016/0168-9002(87)90887-4`.

[170] CMS Collaboration, “Muon tracking performance in the CMS Run-2 Legacy data using the tag-and-probe technique”. `https://cds.cern.ch/record/2724492`, 2020. Accessed: 2024-04.

[171] K. Rose, “Deterministic annealing for clustering, compression, classification, regression, and related optimization problems”, *Proceedings of the IEEE* **86** (1998), no. 11, 2210–2239, `doi:10.1109/5.726788`.

[172] R. Fruhwirth, W. Waltenberger, and P. Vanlaer, “Adaptive vertex fitting”, CMS Public Note CMS-NOTE-2007-008, CERN, 2007.

[173] CMS Collaboration, “Track impact parameter resolution in the 2017 dataset with the CMS Phase-1 Pixel detector”. `https://cds.cern.ch/record/2723305`, 2020. Accessed: 2024-04.

[174] CMS Collaboration, “The CMS Particle Flow Algorithm”, *EPJ Web Conf.* **191** (2018) 02016, `doi:10.1051/epjconf/201819102016`.

[175] CMS Collaboration, “Performance of CMS muon reconstruction in pp collision events at$\sqrt{s} = 7$TeV”, *Journal of Instrumentation* **7** (October, 2012) P10002–P10002, `doi:10.1088/1748-0221/7/10/p10002`.

[176] T. C. Collaboration, “Performance of the reconstruction and identification of high-momentum muons in proton-proton collisions at s = 13 TeV”, *Journal of Instrumentation* **15** (2020), no. 02, P02027, `doi:10.1088/1748-0221/15/02/P02027`.

[177] CMS Muon POG, “Tag and Probe results for Muon with general tracks in Z Resonance for Run2017 UL”. `https://cms-muonpog.web.cern.ch/cms-muonpog/TagAndProbe/?object=muon&probe=generalTracks&resonance=Z&era=Run2017_UL&numden=NUM_LooseID_DEN_TrackerMuons`, 2020. Accessed: 2024-04-24.

[178] M. Cacciari, G. P. Salam, and G. Soyez, “The anti-kt jet clustering algorithm”, *Journal of High Energy Physics* **2008** (2008), no. 04, 063, `doi:10.1088/1126-6708/2008/04/063`.

[179] CMS Collaboration, “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV”, *Journal of Instrumentation* **12** (2017), no. 02, P02014, `doi:10.1088/1748-0221/12/02/p02014`.

[180] R. Frühwirth and A. Strandlie, “Secondary Vertex Reconstruction”, pp. 159–165. Springer International Publishing, Cham, 2021. `doi:10.1007/978-3-030-65771-0_9`.

[181] CMS Collaboration, “Identification of b-quark jets with the CMS experiment”, *Journal of Instrumentation* **8** (2013), no. 04, P04013, `doi:10.1088/1748-0221/8/04/P04013`.

[182] E. Bols et al., “Jet flavour classification using DeepJet”, *Journal of Instrumentation* **15** (2020), no. 12, P12012, `doi:10.1088/1748-0221/15/12/p12012`.

[183] CMS Collaboration, “Performance summary of AK4 jet b tagging with data from proton-proton collisions at 13 TeV with the CMS detector”. https://cds.cern.ch/record/2854609, 2023. Accessed: 2024-03.

[184] LHC Higgs Cross Section Working Group Collaboration, “Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group”, `doi:10.5170/CERN-2013-004`.

[185] S. Holmes, R. S. Moore, and V. Shiltsev, “Overview of the Tevatron collider complex: goals, operations and performance”, *Journal of Instrumentation* **6** (2011), no. 08, T08001, `doi:10.1088/1748-0221/6/08/t08001`.

[186] TeVatron Collaboration, “Search for neutral Higgs bosons in events with multiple bottom quarks at the Tevatron”, *Physical Review D* **86** (2012), no. 9, `doi:10.1103/physrevd.86.091101`.

[187] TeVatron Collaboration, “Search for Higgs-like particles produced in association with bottom quarks in proton-antiproton collisions”, *Physical Review D* **99** (2019), no. 5, `doi:10.1103/physrevd.99.052001`.

[188] R. Alemany-Fernandez et al., “Operation and Configuration of the LHC in Run 1”. `https://cds.cern.ch/record/1631030`, 2013. Accessed: 2024-02.

[189] CMS Collaboration, “Search for a Higgs boson decaying into a b-quark pair and produced in association with b quarks in proton–proton collisions at 7 TeV”, *Physics Letters B* **722** (2013) 207. CMS-HIG-12-033, CERN-PH-EP-2012-375, arXiv:1302.2892.

[190] CMS Collaboration, “Search for neutral MSSM Higgs bosons decaying into a pair of bottom quarks”, *JHEP* **11** (2015) 071, `doi:10.1007/JHEP11(2015)071`, `arXiv:1506.08329`. Submitted to J. High Energy Phys; published version.

[191] ATLAS Collaboration, “Search for heavy neutral Higgs bosons produced in association with b-quarks and decaying into b-quarks at $\sqrt{s} = 13$ TeV with the ATLAS detector”, *Phys. Rev. D* **102** (2020) 032004, `doi:10.1103/PhysRevD.102.032004`.

[192] C. Oleari, “The POWHEG BOX”, *Nuclear Physics B - Proceedings Supplements* **205–206** (2010) 36–41, `doi:10.1016/j.nuclphysbps.2010.08.016`.

[193] B. Jager, L. Reina, and D. Wackeroth, “Higgs boson production in association with b jets in the POWHEG BOX”, *Phys. Rev. D* **93** (2016), no. 1, 014030, `doi:10.1103/PhysRevD.93.014030`, `arXiv:1509.05843`.

[194] T. Sjöstrand et al., “An introduction to PYTHIA 8.2”, *Computer Physics Communications* **191** (2015) 159–177, `doi:10.1016/j.cpc.2015.01.024`.

[195] CMS Collaboration, “The CMS Phase-1 Pixel Detector Upgrade”, technical report, CERN, Geneva, 2020.

[196] CMS Collaboration, “Performance of jets and missing transverse momentum reconstruction at the High Level Trigger using Run 3 data from the CMS Experiment at CERN”. `https://cds.cern.ch/record/2856238`, 2023. Accessed: 2024-03.

[197] J. Kirk, “Introduction to Triggering”. `https://indico.stfc.ac.uk/event/119/contributions/1341/attachments/292/432/Trigger_RALPhDLecture.pdf`, 2020. Talk given at the lectures program: Advanced Graduate Lectures on practical Tools, Applications and Techniques in HEP.

[198] R. Van Kooten, “Experimental Techniques”. `https://indico.fnal.gov/event/3532/contributions/82033/attachments/52086/62343/vankooten_lecture1a.pdf`, 2010. Talk given at the 2010 Hadron Collider Physics Summer School.

[199] D.Acosta, “Introduction to Trigger and DAQ systems”. `https://indico.fnal.gov/event/43762/contributions/192710/attachments/133065/163909/TriggerLecture_day1.pdf`, 2020. Talk given at the 2020 Fermilab-CERN Hadron Collider Physics Summer School (HCPSS).

[200] “CERN Power Distribution Website”. `https://power-distribution.web.cern.ch/publications/#reports`, 2018. Accessed: 2024-04-01.

[201] Particle Data Group Collaboration, “Review of Particle Physics”, *Phys. Rev. D* **98** (2018), no. 3, 030001, `doi:10.1103/PhysRevD.98.030001`.

[202] CMS Collaboration, “Jet ID 13 TeV UL”. `https://twiki.cern.ch/twiki/bin/view/CMS/JetID13TeVUL`. Accessed: 2024-03-10.

[203] CMS Collaboration, “Pileup Jet ID UL”. `https://twiki.cern.ch/twiki/bin/view/CMS/PileupJetIDUL`. Accessed: 2024-03-10.

[204] CMS Collaboration, “A Deep Neural Network for Simultaneous Estimation of b Jet Energy and Resolution”, *Comput. Softw. Big Sci.* **4** (2020), no. 1, 10, `doi:10.1007/s41781-020-00041-z`, `arXiv:1912.06046`.

[205] CMS Higgs Working Group, “Recommendation for 13 TeV data analyses using b-jet regression”. `https://twiki.cern.ch/twiki/bin/view/CMS/HiggsWG/BJetRegression#Reccomendation_for_13_Main_TeV_d`. Accessed: 2024-03-10.

[206] CMS Collaboration, “Missing ET Optional Filters Run2”. `https://twiki.cern.ch/twiki/bin/viewauth/CMS/MissingETOptionalFiltersRun2`, 2024. Accessed: 2024-05-10.

[207] The CMS Collaboration, “Reweighting recipe to emulate Level 1 ECAL and Muon prefiring”. `https://twiki.cern.ch/twiki/bin/view/CMS/L1PrefiringWeightRecipe?rev=3`. Accessed: 2023-11.

[208] The CMS Collaboration, “Treatment of the HEM15/16 region in 2018 data”. `https://hypernews.cern.ch/HyperNews/CMS/get/JetMET/2000.html`, 2019. Accessed: 2022-01.

[209] Belle Collaboration, “A detailed test of the CsI(Tl) calorimeter for BELLE with photon beams of energy between 20-MeV and 5.4-GeV”, *Nucl. Instrum. Meth.* **A441** (2000) 401, `doi:10.1016/S0168-9002(99)00992-4`.

[210] P. Dauncey, M. Kenzie, N. Wardle, and G. Davies, “Handling uncertainties in background shapes: the discrete profiling method”, *Journal of Instrumentation* **10** (2015), no. 04, P04015–P04015, `doi:10.1088/1748-0221/10/04/p04015`.

[211] R. A. Fisher, “On the Interpretation of χ^2 from Contingency Tables, and the Calculation of P”, *Journal of the Royal Statistical Society* **85** (1922), no. 1, 87–94, `doi:10.2307/2340521`.

[212] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, “Asymptotic formulae for likelihood-based tests of new physics”, *The European Physical Journal C* **71** (2011), no. 2, `doi:10.1140/epjc/s10052-011-1554-0`.

[213] A. L. Read, “Presentation of search results: the CLs technique”, *Journal of Physics G: Nuclear and Particle Physics* **28** (2002), no. 10, 2693, `doi:10.1088/0954-3899/28/10/313`.

[214] ATLAS and CMS Collaborations, "Procedure for the LHC Higgs boson search combination in Summer 2011", *CMS Note-2011/005, ATL-PHYS-PUB-2011-11* (2011). `https://cds.cern.ch/record/1379837`.

[215] F. James, "MINUIT Function Minimization and Error Analysis: Reference Manual Version 94.1", 1994.

[216] G. Cowan, "Statistical data analysis", Oxford University Press, USA, 1998.

[217] S. S. Wilks, "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses", *Annals Math. Statist.* **9** (1938), no. 1, 60, `doi:10.1214/aoms/1177732360`.

[218] CMS Collaboration, "The CMS statistical analysis and combination tool: COMBINE", `arXiv:2404.06614`.

[219] R. D. Cousins, "Lectures on Statistics in Theory: Prelude to Statistics in Practice", `arXiv:1807.05996`.

[220] S. Dittmaier, M. Krämer, and M. Spira, "Higgs radiation off bottom quarks at the Fermilab Tevatron and the CERN LHC", *Phys. Rev. D* **70** (2004) 074010, `doi:10.1103/PhysRevD.70.074010`.

[221] CMS Collaboration, "Alignment of the CMS silicon tracker during commissioning with cosmic rays", *Journal of Instrumentation* **5** (2010), no. 03, T03009–T03009, `doi:10.1088/1748-0221/5/03/t03009`.

[222] N. Bartosik, "Associated top-quark-pair and b-jet production in the dilepton channel at $\sqrt{s} = 8$ TeV as test of QCD and background to tt+Higgs production". Dr., Universität Hamburg, Hamburg, 2015. Universität Hamburg, Diss., 2015. `doi:10.3204/DESY-THESIS-2015-035`.

[223] V. Blobel and C. Kleinwort, "A New Method for the High-Precision Alignment of Track Detectors", `arXiv:hep-ex/0208021`.

[224] V. Blobel, "Software alignment for tracking detectors", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **566** (2006), no. 1, 5–13, `doi:10.1016/j.nima.2006.05.157`.

[225] CMS Collaboration, "Alignment of the CMS tracker with LHC and cosmic ray data", *Journal of Instrumentation* **9** (2014), no. 06, P06009, `doi:10.1088/1748-0221/9/06/P06009`.

[226] G. Mittag, "Search for a Light NMSSM Higgs Boson in the b-Quark Final State at the LHC". Dissertation, Universität Hamburg, Hamburg, 2015. Dissertation, Universität Hamburg, 2015. `doi:10.3204/DESY-THESIS-2015-029`.

[227] CMS Collaboration, "Tracker alignment performance in 2022 (addendum)", 2022.

[228] R. Field, "Min-Bias and the Underlying Event at the LHC", `arXiv:1110.5530`.

[229] CMS Collaboration, "Development of the CMS detector for the CERN LHC Run 3", `arXiv:2309.05466`.

[230] CMS Collaboration, "How to L1 Trigger Menu". `https://twiki.cern.ch/twiki/bin/viewauth/CMS/HowToL1TriggerMenu`, Access Year. Accessed: 2024-04-24.

[231] CMS Collaboration, "CMS Public: Software Guide Global HLT". `https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideGlobalHLT?rev=3135`, 2021. Accessed: 2021-12-02.

[232] CMS Collaboration, "HLT Rates Calculation". `https://twiki.cern.ch/twiki/bin/viewauth/CMS/SteamHLTRatesCalculation`, 2024. Accessed: 2021-12-02.

[233] CMS Collaboration, "HLT Timing Estimation". `https://twiki.cern.ch/twiki/bin/view/CMS/TriggerStudiesTiming?rev=91`, 2024. Accessed: 2022-01-15.

[234] H. Qu and L. Gouskos, "Jet tagging via particle clouds", *Phys. Rev. D* **101** (Mar, 2020) 056019, `doi:10.1103/PhysRevD.101.056019`.

[235] Roberval Walsh and Daina Leyva, "MSSM Hbb trigger studies 2024". `https://indico.cern.ch/event/1360492/contributions/5727420/attachments/2782094/4849416/20240117_MssmHbbTriggers_TSG.pdf`, 2024. CMS internal discussion.

[236] ATLAS Collaboration, "Search for Higgs boson pair production in association with a vector boson in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector", *Eur. Phys. J. C* **83** (2023), no. 6, 519, `doi:10.1140/epjc/s10052-023-11559-y`, `arXiv:2210.05415`.

Acknowledgments

This thesis is the product of huge efforts carried out over three and a half years. During this time, I had the privilege of having the support of many people who made it possible, and today, I am glad to express my appreciation and gratitude towards them.

First, I would like to thank my supervisor, Dr. Roberval Walsh, for his immense support, guidance, and encouragement since the beginning of this PhD project. Even during the busiest times, he always made time to discuss the analysis, answer my numerous questions, and provide invaluable feedback that made this work possible. I am deeply grateful for all the expertise he shared in different areas of research, for encouraging my professional development and for guiding me through this journey. I am also thankful for his detailed proofreading, which significantly contributed to this thesis. It has been a pleasure to work under his supervision.

I would also like to thank my second supervisor, Prof. Dr. Elisabetta Gallo, for her constant support and guidance. Her advice and encouragement throughout this time were crucial to my professional growth in many areas, and her insights on the work presented have been extremely valuable. Working under her supervision has been a wonderful experience.

I am also pleased to thank my colleague, Dr. Chayanit Asawatangtrakuldee, for her support and contribution to this project. Her work on the analysis of the 2017 data in the SL and FH categories was crucial for completing this thesis. Furthermore, her expertise and hard work were instrumental in advancing this work towards the oncoming publication. Working with her has been delightful and inspiring.

I would like to thank Dr. Rainer Mankel and Dr. Alexei Raspereza for their valuable input. I appreciate their guidance, questions, and comments during the numerous Higgs working meetings and private discussions that have helped me move towards the completion of this thesis. Thank you for always making the time to answer my questions.

Special thanks to Dr. Sandra Consuegra for her guidance within the Tracker Alignment group and her help with the initial boost in the statistical inference of this analysis. Our discussions and her advice within and beyond academic life have been extremely helpful.

I am also very grateful to Dr. Aliya Nigamova for answering my numerous questions on statistical physics analysis and sharing her expertise with the CMS Combine Tool. I also want to dedicate special thanks Dr. Andrea Cardini for his support and advice in several aspects of my professional life.

I also want to thank my friends Chavely and Rafael, my friends and fellow PhD students Maryam, and Di, and once again, Aliya and Andrea, for the long conversations, dinners, and coffee we have shared. I would also like to thank Maria Luisa, who joined us during a brief but very fruitful period as an intern at DESY, for her nice work, and Punnawich, who contributed to the work presented in this thesis as part of the Bangkok team.

Lastly, I thank my family for their unconditional love and support. I am grateful

to my parents and grandmother for always encouraging and helping me. I also want to thank my sister Diana for inspiring me every day and for being such a great support. Thanks to her and Gustavo for always sharing their positive perspectives. Finally, I want to thank my husband for all his love and understanding. We have gone through many things together, including our PhD journey, always finding support in each other. His determination, kindness, and patience will always inspire me. I am very fortunate and grateful to have them in my life.