

Measurement of the $t\bar{t}H(b\bar{b})$ process in the dilepton final state with machine learning techniques at the CMS experiment

Dissertation

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

vorgelegt von

Angela Giraldi

Hamburg, 2023

Gutachter/innen der Dissertation:

Dr. Maria Aldaya Martin
Prof. Dr. Elisabetta Gallo

Zusammensetzung der Prüfungskommission:

Dr. Maria Aldaya Martin
Prof. Dr. Elisabetta Gallo
Prof. Dr. Dieter Horns
Dr. Alexei Raspereza
Prof. Dr. Sven-Olaf Moch

Vorsitzende/r der Prüfungskommission:

Prof. Dr. Dieter Horns

Datum der Disputation:

14.03.2023

Vorsitzender des Fach-Promotionsausschusses PHYSIK:

Prof. Dr. Günter H. W. Sigl

Leiter des Fachbereichs PHYSIK:

Prof. Dr. Wolfgang J. Parak

Dekan der Fakultät MIN:

Prof. Dr.-Ing. Norbert Ritter

ABSTRACT

This thesis presents the first measurement of the associated production of a Higgs boson and a top-quark-antiquark pair ($t\bar{t}H$) in final states with two leptons and where the Higgs boson decays into a bottom-quark-antiquark pair ($H \rightarrow b\bar{b}$). The measurement utilizes proton-proton collision data collected at the CERN LHC with the CMS experiment at a center-of-mass energy $\sqrt{s}=13$ TeV during 2016-2018 of Run 2, which corresponds to an integrated luminosity of 138 fb^{-1} . In the Standard Model (SM), the Higgs boson couples to elementary particles with a Yukawa-type interaction and a coupling strength proportional to the particle mass. Since the top quark is the heaviest elementary particle known to date, the $t\bar{t}H$ process is the best direct probe of the top-Higgs Yukawa coupling and, therefore, a vital element to verify the SM nature of the Higgs boson. Furthermore in the SM, the $H \rightarrow b\bar{b}$ decay presents the largest branching fraction and is thus experimentally attractive as a final state. The dominant background contributions arise from $t\bar{t}$ events containing additional jets that do not originate from the $t\bar{t}$ system ($t\bar{t} + \text{jets}$). In particular, $t\bar{t} + b\bar{b}$ processes constitute an almost irreducible background with respect to the $t\bar{t}H, H \rightarrow b\bar{b}$ signal. To improve the sensitivity of the measurement, dedicated multivariate analysis techniques are used to extract the signal.

This study includes detailed assessments of the performance and optimization of discriminants trained using, for the first time, Artificial Neural Networks, as well as improved modeling of the dominant background contributions. The results are interpreted in terms of the inclusive $t\bar{t}H$ signal-strength modifier ($\mu_{t\bar{t}H}$) and the first $t\bar{t}H, H \rightarrow b\bar{b}$ STXS measurement, as a powerful tool to test the SM predictions and search for new physics. The statistical model used for measurements effectively describes well the observed data. The observed best fit value is $\mu_{t\bar{t}H} = 0.34^{+0.17}_{-0.17}(\text{stat.})^{+0.20}_{-0.21}(\text{syst.})$ for the combination of all decay channels.

Additionally, this thesis presents an overview of the luminosity measurement for the CMS experiment calibrated using the Van der Meer method. The focus is on an improved method used to estimate the bias due to the assumption of factorizable proton densities of the beams using transverse densities determined from beam-imaging scan data, resulting in a significant reduction of the systematic uncertainty for the measurement of the integrated luminosity.

ZUSAMMENFASSUNG

Diese Arbeit beschreibt die erste experimentelle Messung der assoziierten Produktion eines Higgs Bosons und einem Top Quark-Antiquark Paar ($t\bar{t}H$) in Endzuständen mit zwei Leptonen, wobei das Higgs Boson in ein Bottom-Quark-Antiquark Paar ($H \rightarrow b\bar{b}$) zerfällt. Für die Messung werden Daten verwendet, die in Proton-Proton Kollisionen am CERN LHC mit dem CMS Experiment bei einer Schwerpunktsenergie von $\sqrt{s}=13$ TeV in den Jahren 2016-2018 in der zweiten Datennahmeperiode (Run 2) gewonnen wurden und einer integrierten Luminosität von 138 fb^{-1} entsprechen.

Im Standardmodell der Teilchenphysik (SM) koppelt das Higgs Boson durch Yukawa-Wechselwirkung an andere Elementarteilchen mit einer Stärke proportional zu ihrer Masse. Da das Top Quark das schwerste bisher bekannte Elementarteilchen ist, ist der $t\bar{t}H$ Prozess die beste bekannte Möglichkeit um die Top-Higgs Yukawakopplung zu studieren und bildet damit ein wesentliches Element die SM-Natur des Higgs Bosons zu überprüfen. Darüber hinaus weist der $H \rightarrow b\bar{b}$ Zerfallskanal das größte Verzweigungsverhältnis auf und ist daher experimentell sehr attraktiv. Dominierende Untergrundbeiträge sind $t\bar{t}$ Ereignisse, die zusätzliche Jets enthalten, welche nicht aus dem $t\bar{t}$ Zerfall stammen ($t\bar{t} + \text{jets}$). Vor allem stellt der $t\bar{t} + b\bar{b}$ Prozess einen nahezu irreduziblen Untergrund für die Messung des $t\bar{t}H, H \rightarrow b\bar{b}$ Signalprozesses dar. Um die Sensitivität der Messung zu verbessern, werden spezielle multivariate Analysetechniken verwendet um die Signalstärke zu extrahieren.

Diese Studie umfasst detaillierte Untersuchungen der Leistungsfähigkeit und Optimierung von Diskriminanten, die zum ersten Mal mit Hilfe von neuronalen Netzen trainiert wurden. Des Weiteren wird eine verbesserte Modellierung der dominanten Untergrundbeiträge diskutiert. Die Ergebnisse werden im Rahmen des Signalstärkemodifikators $\mu_{t\bar{t}H}$ und auch zum ersten Mal in Form einer $t\bar{t}H, H \rightarrow b\bar{b}$ STSX Messung interpretiert, was eine hervorragende Möglichkeit bietet, Standardmodell Vorhersagen zu überprüfen und nach neuer Physik zu suchen. Der resultierende gemessene Wert ist $\mu_{t\bar{t}H} = 0.34_{-0.17}^{+0.17}(\text{stat.})_{-0.21}^{+0.20}(\text{syst.})$ für die Kombination aus allen Zerfallskanälen.

Zusätzlich wird in dieser Arbeit die Messung der Luminosität des CMS-Experiments vorgestellt, welche mit der Van-der-Meer Methode kalibriert wird. Der Schwerpunkt liegt dabei auf einer verbesserten Methode um die systematische Verzerrung aufgrund der Annahme von faktorisierbaren Protonendichten der Strahlen durch transversale Dichten abzuschätzen. Hierzu werden Daten benutzt, welche in Strahlabbildungs-Scans gewonnen wurden. Dies führt zu einer signifikanten Verringerung der systematischen Unsicherheit bei der Messung der integrierten Luminosität.

CONTENTS

Introduction	1
1 The $t\bar{t}H$ process in the standard model of particle physics	5
1.1 The standard model of particle physics	5
1.2 Physics in proton-proton collisions	18
1.3 The $t\bar{t}H$ process at hadron colliders	21
2 The experimental setup	35
2.1 The Large Hadron Collider	35
2.2 The Compact Muon Solenoid detector	38
3 Event reconstruction	49
3.1 Basic objects	49
3.2 Physics objects	51
4 Luminosity measurement	59
4.1 Luminosity of collider experiments	59
4.2 Precision luminosity measurements at CMS in Run 2	60
5 Measurement of $t\bar{t}H$, $H \rightarrow b\bar{b}$ - Analysis strategy	75
5.1 Introduction	75
5.2 Event samples	76
5.3 Event selection	83
5.4 Corrections to the event simulation	87
5.5 $t\bar{t}$ + jets background model	92
5.6 Event categorization in the analysis regions	99
5.7 Observables for the parameter estimation	101
6 Measurement of $t\bar{t}H$, $H \rightarrow b\bar{b}$ - Analysis results	117
6.1 Uncertainty model	117
6.2 Measurement results	130
6.3 Contribution to the combined analysis	138
7 Toward differential measurements	145
7.1 Simplified template cross sections framework	145
7.2 STXS measurement in $t\bar{t}H$, $H \rightarrow b\bar{b}$ - Analysis strategy	147
7.3 STXS measurements results	154
Conclusion and outlook	167

Bibliography	171
A Machine learning and neural networks	191
A.1 Mathematical construction	191
A.2 Neural network training	192
B Statistical inference	199
B.1 The likelihood function	199
B.2 Inference methods	202
B.3 Saturated goodness-of-fit test	205
C MC settings for the $t\bar{t}$ + jets production	207
D Trigger paths	209
E Additional distributions for corrections to event simulation	211
E.1 Trigger efficiencies	211
E.2 b-tagging corrections	214
F Comparison 4FS vs 5FS	221
F.1 Bias tests	222
G Input variables to artificial neural networks	225
H Performances of the artificial neural networks	251
H.1 Taylor expansion	251
H.2 Input features validation	254
H.3 Loss function	260
I The ratio observable	261
J Yields tables	267
J.1 Inclusive measurement - full Run 2 dilepton fit	267
J.2 Inclusive measurement - full Run 2 all channels fit	270
J.3 STXS measurement - full Run 2 dilepton fit	273
J.4 STXS measurement - full Run 2 all channels fit	279
K Pre/Post-fit shape distributions	285
K.1 Inclusive measurement - full Run 2 dilepton fit	285
K.2 Inclusive measurement - full Run 2 all channels fit	288
K.3 STXS measurement - full Run 2 dilepton fit	291
K.4 STXS measurement - full Run 2 all channels fit	297
L Impacts	303
L.1 Inclusive measurement - full Run 2 dilepton fit	303
L.2 Inclusive measurement - full Run 2 all channels fit	307
L.3 STXS measurement - full Run 2 dilepton fit	317
L.4 STXS measurement - full Run 2 all channels fit	321

INTRODUCTION

High-energy physics seeks to explore the smallest constituents of nature, the elementary particles, and the underlying mechanisms that govern them with only four fundamental interactions. The standard model of particle physics (SM) describes our current knowledge and understanding regarding elementary particles. The most remarkable success of the SM is the proposal by Robert Brout, François Englert, and Peter W. Higgs in 1964 of the mechanism explaining the origin of the masses of the force-carrying particles via the existence of a new elementary particle, the so-called Higgs boson [1–7]. The search for this particle has driven the construction of the Large Hadron Collider (LHC) at CERN in Switzerland. Finally, in 2012, after 48 years, a particle compatible with the expected properties of the Higgs boson was observed with the A Toroidal LHC ApparatuS (ATLAS) and Compact Muon Solenoid (CMS) detectors [8, 9]. That discovery started a massive effort to probe the Higgs sector, to construct stringent tests of the SM and contributions from physics beyond (BSM). This year marks the tenth anniversary of the discovery, but there is still much potential for improvements in accurate measurements of its properties [10, 11].

The interaction of the Higgs boson with other elementary particles is expected to be proportional to their masses. Since the top quark is the heaviest elementary particle known, it couples to the Higgs boson with a Yukawa coupling constant close to unity. The research of this thesis aims to contribute to the precision measurement of the Higgs-top coupling via the most promising process: the production of a Higgs boson in association with a top quark-antiquark pair $t\bar{t}H$. This precise measurement is crucial not only for probing the validity of the SM [12, 13] but also because it is expected to be sensitive to BSM physics, which may predict different coupling strengths [14–17]. Indeed, although the predictions of the SM are compatible with observations a more fundamental theory is required: the SM does not include a quantum description of gravity and cannot account for observations in other fields, such as cosmology, where substantial evidence for non-radiative dark matter [18, 19] and dark energy was observed [20].

The decay of a Higgs boson into a top quark-antiquark pair ($t\bar{t}$) is kinematically forbidden, but the $t\bar{t}H$ production allows for direct probing of the top-Higgs coupling directly. Its cross section is $\sigma_{\text{SM}}^{t\bar{t}H} = 507_{-50}^{+35}$ fb at a center of mass of $\sqrt{s} = 13$ TeV, for a Higgs boson mass of $m_H = 125$ GeV and a top quark mass $m_t = 172.5$ GeV [21]. The $t\bar{t}H$ process has a complex and rich variety of final states resulting from the combinations of $t\bar{t}$ and Higgs boson decay channels. In 2018, the $t\bar{t}H$ was first observed based on a combined analysis of p-p collision data recorded with the CMS detector at $\sqrt{s} = 7, 8$ and 13 TeV, corresponding to an integrated luminosity of up to 5.1, 19.7, and 35.9 fb⁻¹, by taking into account the Higgs boson decays into pairs of W and Z bosons, photons, τ leptons, or bottom quark jets [22]. Shortly afterwards, both ATLAS and CMS Collaborations also reported on the observation of Higgs boson decays into bottom quark pairs, the one with the largest branching fraction of 0.58 ± 0.02 [21, 23, 24].

This thesis presents a search for $t\bar{t}H$ production into a pair of bottom quarks ($t\bar{t}H, H \rightarrow b\bar{b}$) and leptonic top quark pair decays based on p-p collisions recorded with the CMS detector

in the period 2016-2018 at a center of mass of $\sqrt{s}=13$ TeV, corresponding to an integrated luminosity of 138 fb^{-1} . The luminosity of a collider is a property that quantifies the rate of collisions occurring at the experiment, and its precise measurement is required for all cross section measurements. In the CMS experiment, it is calibrated with the Van der Meer method [25]. In the context of this thesis, an improved beam-imaging approach was used to investigate the factorization bias arising from the assumption of factorizable transverse profiles of the beams. This improved approach now takes into account additional contributions from beam-beam deflection, which were not ever previously considered. As a result, the factorization bias has been reduced from 1.5% (0.9%) in 2015 (2016) to a much lower 0.5% overall. These results are documented in Ref. [26], with main contribution from the author. In addition, the author has developed a new and faster curve fitting method to improve the beam-imaging approach with respect to Refs. [27, 28]. This method is currently being used to estimate the factorization bias for data from 2017, 2018 and ongoing data collection at the LHC.

In the $t\bar{t}H, H \rightarrow b\bar{b}$ analysis, handling the $t\bar{t}$ pair production process in association with a bottom-quark-antiquark pair ($t\bar{t} + b\bar{b}$) is highly critical. The $t\bar{t} + b\bar{b}$ is not only an irreducible background, but it also has a cross section $\mathcal{O}(10)$ times larger than the signal in the analyzed phase-spaces, and neither higher-order theoretical calculations nor experimental methods can constrain the expected rate of these contributions with good accuracy [29–46]. The sensitivity of the analysis is additionally limited by a combinatorial background caused by multiple b quark jets in the final state. To address these challenges, for the first time novel advanced multivariate analysis techniques such as neural networks have been developed by the author and deployed to enhance the separation between the signal and background processes and provide a high sensitivity to the $t\bar{t}H, H \rightarrow b\bar{b}$ process. The signal rate results of the inclusive $t\bar{t}H, H \rightarrow b\bar{b}$ analysis in the dilepton channel are obtained from a combined binned maximum likelihood fit of the classifier output distributions to the data, correlating processes and their uncertainties. They are presented in this work both considering only the dileptonic channel and also for the combination with the other channels, defined from the other top quark pair decays (single lepton and fully hadronic). Additionally, the available phase space is split into different regions of Higgs boson p_T , synchronized between experimental searches and the theoretical calculations, following the approach of the Simplified Template Cross Section (STXS) [47]. The results presented here are the first STXS measurement for the $t\bar{t}H, H \rightarrow b\bar{b}$ production, particularly in the dileptonic channel. The results obtained in the context of this thesis, are currently scrutinized by the CMS Collaboration and are going to be published in the first paper with a combined measurement of $t\bar{t}H$ and tH production in the $H \rightarrow b\bar{b}$ decay channel with 138 fb^{-1} of proton-proton collision data at 13 TeV.

This thesis is structured as follows. Chapter 1 provides a brief overview of the standard model, with a focus on the role of the top quark and Higgs boson. Chapter 2 introduces the CERN Large Hadron Collider LHC and the CMS experiment and Chapter 3 covers the main aspects of CMS event and physics reconstruction methods. Chapter 4 describes the luminosity measurement of the CMS experiment, with a emphasis on the author’s results for the factorization bias estimation. Chapter 5 outlines the strategy for the inclusive $t\bar{t}H, H \rightarrow b\bar{b}$ analysis in the dileptonic channel, including the definition of relevant physics processes, their simulation, and the use of neural-network based observables to improve analysis sensitivity. Chapter 6 presents the results of the inclusive $t\bar{t}H, H \rightarrow b\bar{b}$ analysis in the dilepton channel and its contribution to the combined $t\bar{t}H, H \rightarrow b\bar{b}$ analysis. Chapter 7 focuses on a new analysis method and the results for cross section measurement in the STXS framework. Finally, Chapter 8 summarizes the conclusions and provides an outlook for future data sensitivity projections.

THE $t\bar{t}H$ PROCESS IN THE STANDARD MODEL OF PARTICLE PHYSICS



1.1	The standard model of particle physics	5
1.1.1	The elementary particles	6
1.1.2	Gauge symmetries	8
1.1.3	Spontaneous symmetry breaking	13
1.1.4	Open questions of the standard model	16
1.2	Physics in proton-proton collisions	18
1.2.1	Factorization and parton distribution functions	18
1.2.2	Monte Carlo event simulation in proton-proton collisions	20
1.3	The $t\bar{t}H$ process at hadron colliders	21
1.3.1	The Higgs boson	22
1.3.2	The top quark	26
1.3.3	Signal and relevant background processes	30
1.3.4	From previous measurements to projected sensitivity	32

This chapter provides a comprehensive overview of the theoretical foundations that underlie the measurements presented in this thesis. The standard model of particle physics, which is the current theoretical framework that describes the fundamental building blocks of matter and their interactions, is discussed in detail. Additionally, the chapter highlights how high-energy data analysis relies largely on precision modeling, and the simulation of collision events is an essential step in the data analysis process. The essential characteristics of the event simulation are briefly explained, including the simulation tools, models used to describe the detector response and the methods used to generate the events. The Higgs mechanism and its phenomenology at the Large Hadron Collider (LHC) are also discussed, with a focus on the features that are crucial for describing proton-proton (p - p) collisions. The chapter then introduces the $t\bar{t}H$ event production and decay characteristics at hadron colliders, with a particular emphasis on the Higgs boson and top quark pair decay channels that are investigated in this thesis. Finally, the chapter concludes with a summary of results from prior measurements, highlighting the importance of the work presented in this thesis in advancing our understanding of the standard model and the search for new physics beyond it.

1.1 The standard model of particle physics

The standard model of particle physics (SM) is the most precise theoretical description of our current knowledge of elementary particles and their interactions. Experimental observations and direct tests at collider experiments have extensively corroborated the SM on an extensive range of energy scales, making it very solid and widely accepted. It is a renormalizable quantum field theory based on gauge invariance principles and provides a unified description of the strong, weak, and electromagnetic forces. However, theoretical considerations and observations from the subnuclear to the cosmic scales indicate that the

SM may be incomplete. It is necessary to extend the theory Beyond the SM (BSM) in search of higher-dimensional symmetries. For example, it lacks a description of the gravitational force, which is several orders of magnitude weaker compared to the other forces, and it is foreseen to become relevant at the so-called Planck scale, of the order of 10^{19} GeV.

The SM elegantly represents matter through 12 fundamental spin-1/2 particles known as fermions, which are divided into two classes based on whether they do interact strongly or not: quarks and leptons. The fermions interact with one another via spin-1 particles known as bosons, and three fundamental interactions govern them. A sketch of the SM constituents and their classification is shown in Figure 1.1. Mathematically, the SM is described by a renormalizable gauge-invariant quantum field theory (QFT) based on the local symmetry $SU(3) \times SU(2) \times U(1)$ gauge group, with a partial breaking of the symmetry induced by the Brout-Englert-Higgs mechanism in the $SU(2) \times U(1)$ electroweak (EWK) sector.

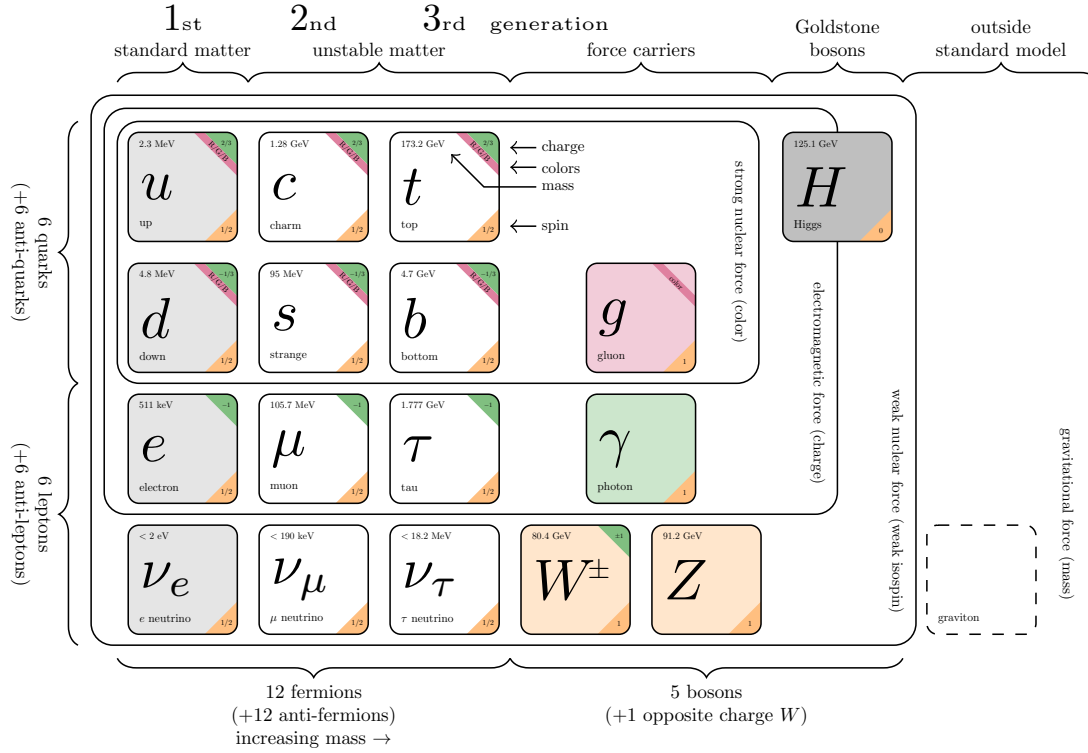


Figure 1.1: Particle content of the SM: the three generations of matter particles, the fermions, and the force carrier particles, the gauge bosons, in addition to the Higgs boson. Electric charge, spin, color, and the mass of each particle are also documented [48].

This chapter is meant to provide a brief overview of these topics; more detailed reviews can be found in Refs. [48–52].

1.1.1 The elementary particles

The fermionic fields, or fermions, are the elementary building components of matter and are classified into two categories: quarks and leptons. There are six massive quarks (q), which are color triplets (i.e., have a color charge that exists in three states: red (R), green (G), and blue (B)), carry EWK charges, and are subject to all three types of SM interactions mentioned

above. In addition, positively and negatively charged quarks are classified as up-type (u) or down-type (d). Then, there are six leptons, colorless but with EWK charges. Charged leptons (ℓ) interact via all forces except the strong interaction, whereas neutrinos (ν) are electromagnetically neutral, solely sensitive to the weak force, and assumed massless in the SM theory. While leptons can exist as free particles, quarks do not appear to be able to do so due to a strong force phenomenon known as color confinement (see Section 1.1.2). As a result, quarks combine in white color charge bound states (hadrons) comprising either a quark-antiquark pair (mesons) or three quarks (baryons).

All quarks and leptons have a corresponding antiparticle, according to Dirac's relativistic wave equation [53]. This prediction requires the presence of anti-matter, which has been experimentally validated by observing the positron (the antiparticle of the electron) [54]. Antiparticles have the same attributes as their equivalent partner particle (mass, spin, and lifetime) but have the opposite electric charge.

Fermions with the same quantum number are divided into three generations based on their electrical charge, weak isospin component, and mass hierarchy, beginning with the lightest and most stable particles (see Section 1.1.2), but there is currently no fundamental explanation for this subdivision. Doublets of leptons or quarks can be identified in each generation, and the same is true for antiparticles. For example, the doublet of leptons comprises a charged heavy lepton and a neutrino, assumed massless in the SM theory.

A fermionic field ψ can be decomposed into its left-handed ψ_L and its right-handed ψ_R components as $\psi = \psi_L + \psi_R$. The ψ_L and ψ_R components indicate the two irreducible representations of the restricted and orthochronous Lorentz group. The SM is a chiral theory, not invariant under parity changes, and therefore, ψ_L and ψ_R behave differently under the gauge symmetries transformations. For example, leptons and quarks can be represented in the following notation:

$$\begin{aligned}\psi_L^{lepton} &= \begin{pmatrix} \nu \\ l^- \end{pmatrix}_L & \psi_R^{lepton} &= l_R^- \\ \psi_L^{quark} &= \begin{pmatrix} u \\ d' \end{pmatrix}_L & \psi_R^{quark} &= u_R, d_R\end{aligned}$$

where all ψ_L are doublets and all ψ_R are singlets under $SU(2) \times U(1)$. The formers have a weak isospin component of $I_3 = +\frac{1}{2}$ and $I_3 = -\frac{1}{2}$ for the upper and lower rows, respectively, while for the latter $I_3 = 0$. Furthermore, the u -type quarks have an electric charge of $Q = +\frac{2}{3}e$ and are the up (u), charm (c) and top (t) quarks. The d -type quarks are the down (d), strange (s), and bottom (b) quarks, with $Q = -\frac{1}{3}e$. Finally, the charged leptons, with $Q = -e$ are the electron (e), muon (μ), and tau (τ) leptons. The elementary positive charge e corresponds to $1.602\,176\,634 \times 10^{-19}$ C [55]. Electron, muon, and tau neutrinos are electrically neutral and lack a right-handed counterpart.

The fermions interact via the exchange of five gauge bosons, which correspond to different types of interactions. Except for massless and electrically neutral neutrinos, the photon γ is the mediator of electromagnetic interactions for all fermions. When quarks interact strongly, eight gluons mediate the interaction; gluons are massless, electrically neutral, and carry a color charge. Three gauge bosons, the charged $W^\pm$ and the neutral Z , can convey the weak force for all fermions. The Higgs boson H , predicted by the SM and discovered in July 2012 at the LHC [8, 9] completes the picture of the SM and will be addressed more widely in Section 1.1.3.

1.1.2 Gauge symmetries

The SM is a gauge-invariant quantum field theory that explains strong, weak, and electromagnetic interactions by constructing a local invariant Lagrangian under the non-abelian group $SU(3) \times SU(2) \times U(1)$.

Local gauge symmetry is invariance under transformations that rotate the internal degrees of freedom at space-time point-dependent rotation angles. Different quantum numbers distinguish the various types of fermions, some of which correspond to a charge conserved under local transformations of the gauge invariance of the $SU(3) \times SU(2) \times U(1)$ group: rotations in hypercharge space for $U(1)$, rotations in weak isospin space for $SU(2)$, and rotations in color space for $SU(3)$.

Quantum electrodynamics

The relativistic quantum field theory of electrodynamics (QED), an abelian gauge theory with the symmetry group $U(1)$, describes the EM interactions of electrically charged particles. The QED Lagrangian density for a $\frac{1}{2}$ spin field interacting with the electromagnetic field is given in natural units by:

$$\mathcal{L}_{QED} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad (1.1)$$

where γ^μ are the Dirac matrices and $D_\mu \equiv \partial_\mu + iQA_\mu$ is the gauge covariant derivative, Q is the coupling constant that can be interpreted as the electric charge of the spinor field ψ and m its mass, A_μ is the covariant four-potential of the electromagnetic field, representing the photon, and the electromagnetic field tensor:

$$F_{\mu\nu} = \partial_\mu A_\nu(x) - \partial_\nu A_\mu(x). \quad (1.2)$$

The QED Lagrangian density is composed of three terms: a component $\psi\bar{\psi}$ to explain the kinematics of a free fermionic field, an A^2 component that describes the kinematics of a free photon, and a mixed $\bar{\psi}\gamma^\mu A_\mu\psi$ term that defines the interaction between the two fermion fields. The QED does not include the self-interaction of the photon field since it has an abelian $U(1)$ symmetry; also, a 4-point vertex (e.g., two fermions and two photons) does not exist due to renormalisability. As a result, only one elementary QED vertex exists, as illustrated in Figure 1.2.

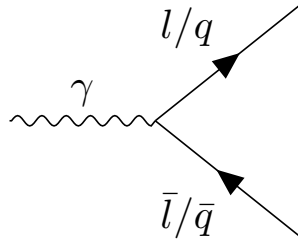


Figure 1.2: Feynman diagram of the elementary vertex of the EM interaction. γ represents the massless mediator of QED, the photon, while l refers to charged leptons only.

The Lagrangian density in Eq. 1.1 is invariant under the local transformation of the abelian unitary group $U(1)$:

$$\tilde{q}_0 = \begin{cases} \psi \rightarrow \psi' = e^{iQ\theta(x)}\psi \\ A_\mu \rightarrow A'_\mu = A_\mu - \partial_\mu\theta(x) \end{cases}$$

Q is the electric charge operator of the $U(1)$ group and $\theta(x)$ the phase depending on the space-time coordinates.

As a result of the gauge invariance, the mass term $m^2 A_\mu A^\mu$ is not allowed, leaving the photon massless. This property has been confirmed by several experimental observations, setting a limit on the photon mass $m_\gamma < 10^{-18}$ [55]. In addition, because of the abelian property of the group, the photon is not self-interacting. Photons can only couple to charged particles, with the coupling strength given by the constant $\alpha = \frac{e^2}{4\pi\epsilon} \approx \frac{1}{137} \ll 1$.

The formulation of QED following gauge invariance principles provides the necessary model for developing the SM as a gauge theory.

Quantum chromodynamics

The quantum chromodynamics (QCD) theory describes the strong interaction between quarks and gluons, which are responsible for holding ordinary matter together. The QCD theory is based on the more general theory introduced by Yang and Mills [56]. Based on the experimental evidence on the composition of hadrons, it was deduced that quarks carry an additional quantum number compared to leptons, the color charge, and that the strong interaction can be characterized by a non-abelian $SU(3)$ color symmetry group.

The strong interaction between quarks is carried by eight gluons, corresponding to physical gauge fields G_μ described by eight generators, the Gell-Mann matrices. The color charge is required by QCD to preserve the Pauli exclusion principle and is the fundamental representation of the group. It has three eigenstates (red, green, and blue) for the quarks and gluons. As quarks carry color, the antiquarks carry anti-color.

Quark fields are organized in color triplets of the $SU(3)$ group as $Q = (q_{\text{red}}, q_{\text{green}}, q_{\text{blue}})^T$. On the other hand, leptons are described by a singlet, carry no color charge, and do not participate in the strong interaction. The full gauge invariant QCD lagrangian is given by:

$$\mathcal{L}_{\text{QCD}} = \sum_{f=1}^{N_f} \bar{Q}(i\gamma^\mu D_\mu - m_q)Q - \frac{1}{4}G_{\mu\nu}^a G_a^{\mu\nu} \quad (1.3)$$

with the covariant derivative defined as:

$$D_\mu = \partial_\mu + ig_s T_a G_\mu^a$$

where $T_a = \lambda_a/2$ are the eight generators of the $SU(3)$ group, expressed in terms of the Gell-Mann matrices λ_a . N_f is the number of quark flavors, and the gauge invariant strength field tensor is:

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a - g_s f_{abc} G_\mu^b G_\nu^c, \quad a = 1, \dots, 8$$

where f_{abc} are the structure constants of the $SU(3)$ group defined by the commutation relation $[T_a, T_b] = if_{abc} T_c$ of its eight generators. The constant g_s is the dimensionless coupling constant strength of the strong interaction related to $\alpha_s = g_s^2/(4\pi)$.

The interaction between quarks and gluons resembles the one for QED. Conversely to the QED case where the photon is neutral, self-interaction is allowed by the symmetry group because also gluons carry a color charge. Both triple and quartic gauge couplings are present in QCD, representing the non-abelian property of the groups. The couplings of the ggg and $gggg$ vertices are of order g_s and g_s^2 , respectively, and these elementary vertices are shown in Figure 1.3.

The structure of QCD theory is simple, yet its content is vibrant. Among its remarkable properties are color confinement, asymptotic freedom, and the absence of CP violation.

A common aspect of QFT, and hence any SM interaction, is that the coupling constant depends on the energy of the considered process; this behavior is known as running of

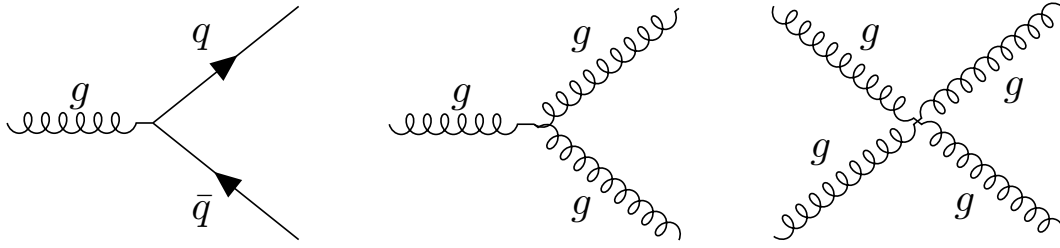


Figure 1.3: Feynman diagrams of the elementary vertices of the QCD interaction.

the coupling. This is a direct result of the theory's renormalizability: to eliminate divergent quantities a cut-off energy is typically introduced to allow the couplings to acquire a scale dependence by normalizing them to a known (measured) value at a given scale. The origin of these divergences is often interpreted as the SM being a low-energy effective field theory of a more fundamental, yet unknown, theory.

The running of the coupling can be viewed qualitatively as contributing to vacuum polarization via loop corrections from fermions and gauge bosons. These interfering virtual particles produce a change in the magnitude of the force. Due to their different symmetry groups, the electromagnetic, weak, and strong interactions behave differently. For example, the EM coupling is screened by a cloud of virtual electron-positron pairs, giving at a long distance (or low energy) the well-known value of the fine structure constant $\alpha_{EM} = \frac{e^2}{4\pi\epsilon} \approx \frac{1}{137}$, corresponding to the electron mass scale. At higher energies, such as the Z mass scale at which the unification of the EWK theory happens, the coupling value has been measured to be $\approx 1/129$ [57]. On the contrary, the gluon self-interactions in QCD produce a cloud of virtual gluons, which, together with the quark contribution, makes the coupling strength $\alpha_s = g_s^2/(4\pi)$ becoming stronger at large distances, or, equivalently, at low energies [58, 59]. When a quark-antiquark pair is produced, increasing amounts of energy are required to separate the two quarks, making forming a new quark-antiquark pair energetically favorable. This sequential quark pair production is called hadronization: the coupling becomes so strong to confine quarks that they eventually organize themselves into colorless bound states (hadrons) comprising either a quark-antiquark pair (mesons) or three quarks (baryons). This is a feature known as color confinement. As a result, quarks and gluons created in high-energy particle collisions cannot be observed freely as bare states but instead form collimated sprays of particles known as jets (see Section 3.2.5).

Conversely, another remarkable peculiarity of QCD is asymptotic freedom: the effective strength of the interaction decreases with increasing transferred momentum among the participants. This behavior of α_s shows that partons in the protons at small distances are almost free in this kinematic regime, i.e., asymptotically free. The description of their interactions can be approximated using a perturbative expansion in terms of α_s .

More precisely, low-energy processes, like hadronization, cannot be estimated analytically but must be modeled phenomenologically ($\alpha_s(m_\tau) \approx 0.3$). On the other hand, at high energies, α_s decreases, e.g. $\alpha_s(100 \text{ GeV}) \approx 0.118$, allowing quarks to be treated as quasi-free particles for perturbative computations.

The electroweak theory

The weak interaction has earned its name because its strength is several orders of magnitude less than the strong and electromagnetic forces. The first indications of the weak force

can be found in the nuclear β -decay research. However, the current formulation is credited to Glashow, Weinberg and Salam [60–62], who proposed a gauge theory unifying the weak and the electromagnetic interactions, called electroweak theory. Although both interactions differ at low energies, they can be merged into one gauge theory at high energy regimes beyond ≈ 100 GeV.

The three isospin weak operators $\vec{I} \equiv (I_1, I_2, I_3)$, and the weak hypercharge Y are the generators of the groups. It is possible to associate the $\vec{W}_\mu \equiv (W_\mu^1, W_\mu^2, W_\mu^3)$ gauge bosons to the isospin operators and B_μ gauge boson to the hypercharge operator thanks to the gauge invariance under $SU(2) \times U(1)$ transformations. This leads to the expression of the EWK Lagrangian density as:

$$\mathcal{L}_{EWK} = i\bar{\psi}_L \gamma^\mu D_\mu \psi_L + i\bar{\psi}_R \gamma^\mu D_\mu \psi_R - \frac{1}{4} \sum_{a=1}^3 \vec{W}_{\mu\nu}^a \cdot \vec{W}_a^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}, \quad (1.4)$$

where the left and right components of the spinors are split to emphasize the different transformations. The covariant derivative is here defined as:

$$D_\mu = \partial_\mu + i\frac{g'}{2} Y B_\mu + ig \sum_{a=1}^3 W_\mu^a I^a,$$

with g' and g the dimensionless coupling constants and $I^a = G^a/2$ are the generators of $SU(2)$. The field strength tensor $B_{\mu\nu}$ has a similar form as for the QED case (see Eq. 1.2); while for the $SU(2)$ group:

$$W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a - g\epsilon^{abc} W_\mu^b W_\nu^c,$$

where ϵ^{abc} is the totally antisymmetric Levi-Civita tensor.

The electroweak theory is a chiral theory; therefore fermion fields are split into left- and right-handed fermionic fields. Left- (right-) handed (anti-) fermions are organized in isospin doublets L of $SU(2)$ with chirality of -1 and eigenvalues (I, I_3) of $(1/2, \pm 1/2)$. On the contrary, right- (left-) handed (anti-) fermions form isospin singlets R with a chirality of $+1$ and eigenvalues of $I = I_3 = 0$. Unlike QED, the weak interaction is observed to be maximally parity violating [63], implying that only fermions with left chirality interact via the weak interaction. Moreover, the third component of the isospin (I_3) and the weak hypercharge Y , charges and quantum numbers of the $SU(2)$ and $U(1)$, are related to the electric charge by the Gell-Mann-Nishijima formula $Q = I_3 + Y/2$ [64, 65]. As a result, right-handed neutrinos have a zero hypercharge and hence do not participate in the electroweak interaction and thus are not characterized by the SM Lagrangian. In contrast, left-handed neutrinos only interact in the weak interaction.

All the interactions between gauge bosons and fermions can be derived from the equations above. Because of the composite nature of the symmetry built from the $SU(2)$ and $U(1)$ groups, the interpretation of the vector fields as propagators of the bosons mediating the electromagnetic and weak interactions is not as evident as in QED or QCD. However, the interpretation of the terms in the Lagrangian of Eq. 1.4 allows identifying the vector fields

corresponding to the interaction mediators. The boson fields are given by [66]:

$$\begin{aligned}
W_\mu^+ &= \frac{1}{\sqrt{2}} (W_\mu^1 - iW_\mu^2) \\
W_\mu^- &= \frac{1}{\sqrt{2}} (W_\mu^1 + iW_\mu^2) \\
Z_\mu &= \frac{1}{\sqrt{g^2 + g'^2}} (gW_\mu^3 - g'B_\mu^3) \\
A_\mu &= \frac{1}{\sqrt{g^2 + g'^2}} (g'W_\mu^3 + gB_\mu^3),
\end{aligned} \tag{1.5}$$

with the $W^\pm$ bosons responsible for the charged weak current, the neutral Z boson for the weak neutral current, and the photon γ for the electromagnetic interaction. Furthermore, the Z boson and the photon field are obtained by performing a rotation of the gauge fields of an angle θ_W , called Weinberg angle or weak-mixing angle, as:

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix}. \tag{1.6}$$

The electroweak mixing is further reflected in the ratio of the W and Z boson masses, $m_W = \cos(\theta_W) \times m_Z$; and both coupling constants g and g' relate to the electromagnetic charge e as $e = g \sin \theta_W = g' \cos \theta_W$. The Weinberg angle is a free parameter of the SM, and it is determined experimentally with very high precision to be 0.23122 ± 0.00002 [55].

The non-abelian structure of the SU(2) symmetry group also gives rise to self-interactions of the gauge bosons, as illustrated by the only conceivable boson combinations for the triple and quartic gauge couplings diagrams in Figure 1.4.

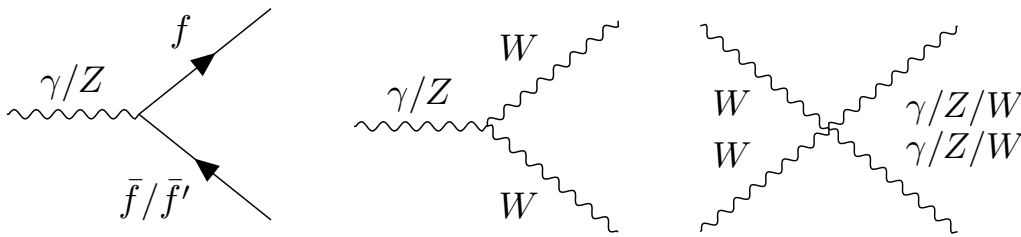


Figure 1.4: Feynman diagrams of the elementary vertices of the EWK interactions, including the self-interaction of gauge bosons. The symbol $f(\bar{f})$ represents any (anti) fermion.

Despite its elegance, as described here, the electroweak theory suffers from crucially significant limitations: it can not describe massive particles, but measurements have shown that the mediators of the weak force are instead massive. Furthermore, any attempt to add gauge boson or fermion mass terms breaks the theory's cornerstone, the Lagrangian's gauge invariance. Terms as $M^2 W^\mu W_\mu$ for vector bosons are not gauge invariant and would introduce unrenormalizable divergences.; similarly, terms of the form $-m^2 \bar{\psi} \psi$ for fermions would contain mixed chirality terms. The following section focuses on incorporating massive particles into the theory without breaking gauge invariance.

1.1.3 Spontaneous symmetry breaking

To corroborate the experimental results, P. W. Higgs, F. B. Englert, and R. Brout theorized and formalized the spontaneous symmetry breaking for local symmetries in 1964, which is today an essential component of the SM [1–3].

The so-called Higgs mechanism predicts the existence of a scalar boson, the Higgs boson, which gives mass to the fermions and the weak bosons, thus providing a way to generate mass terms while preserving gauge invariance in the SM Lagrangian. The concept of global symmetries is based on Goldstone’s theorem [49]. Then, the mechanism was further elaborated into the spontaneous breaking of the electroweak symmetry [67], which allows the symmetry of the gauge theory to be broken without disrupting the gauge invariance: the Lagrangian remains invariant under the symmetry, but the ground state of the theory no. As a result, the ground state degenerates, so choosing one eigenstate is necessary, breaking the symmetry. In particular, the Higgs field breaks three out of the four degrees of freedom of the $SU(2) \times U(1)$ group: the massless Goldstone bosons are absorbed by the three massive $W^\pm$ and Z bosons. The Higgs boson H is the massive scalar particle with the remaining degree of freedom.

Introducing a single colorless $SU(2)$ -doublet scalar field with a non-zero vacuum expectation value can break the symmetry and preserves the gauge invariance once added to the EWK Lagrangian. The field is chosen as a linear combination of four real scalar fields:

$$\phi \equiv \begin{pmatrix} \phi^\dagger \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix},$$

with hypercharge $Y = 1$, isospin $I = 1/2$ and its third component $I_3 = -1/2$ and therefore $Q = 0$.

The Lagrangian with a quartic potential (higher powers of ϕ would lead to a not renormalizable theory [68]) is considered to describe the main idea of symmetry breaking:

$$\mathcal{L}_{BEH} = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi^\dagger \phi).$$

With the covariant derivative being $D_\mu = \partial_\mu + i\frac{g'}{2}B_\mu + ig\tilde{W}_\mu^i \frac{\sigma_i}{2}$ as in Eq. 1.1.2, and the scalar potential describing the evolution of the field expressed as:

$$V(\phi, \phi^\dagger) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2. \quad (1.7)$$

The shape of potential $V(\phi)$ depends on the sign of the parameters:

- $\lambda > 0$ to have bound states inside the potential as $\phi \rightarrow \infty$;
- if $\mu^2 > 0$, $\phi = 0$ is the minimum of the potential with a U shape. The electroweak symmetry is unbroken in the vacuum because a gauge transformation acting on the ground state does not change it;
- if $\mu^2 < 0$, the minimum of the potential, a non-vanishing vacuum expectation value (VEV) v , is not uniquely identified but lies in the hyper-sphere $|\phi|^2 = -\frac{\mu^2}{2\lambda}$. The potential has a Mexican hat shape, as illustrated in Figure 1.5. The realization of one particular state spontaneously breaks the rotational symmetry of the problem.

The expansion of the ground state of the field ϕ around the minimum can be parameterized with the real scalar field h as:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}. \quad (1.8)$$

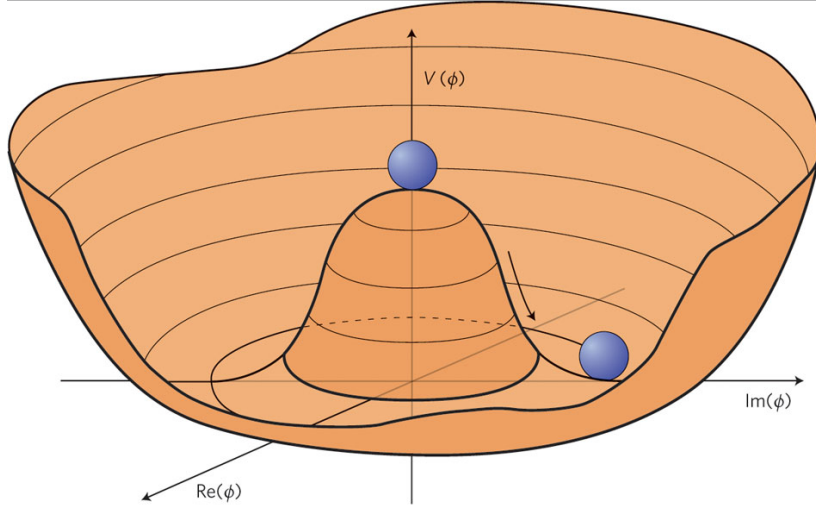


Figure 1.5: Shape of the Higgs potential for $\mu^2 < 0$ [69].

and finally, the potential $V(\phi)$ can be rewritten as:

$$V(\phi, \phi^\dagger) = -\lambda v^2 h^2(x) - \lambda v h^3(x) - \frac{1}{4} \lambda h^4(x). \quad (1.9)$$

The second and third terms represent the cubic and quartic self-coupling of the Higgs boson. Instead, from the first term, the value of the Higgs mass is obtained as:

$$m_h^2 = 2\lambda v^2 = -2\mu^2. \quad (1.10)$$

In contrast to the weak boson self-interactions, which have a gauge nature, the Higgs boson self-interactions are purely related to the parameters of the scalar sector of the theory and entirely determined from the Higgs boson mass and the VEV. Consequently, their experimental measurement represents a crucial test of the validity and coherence of the SM.

The massive gauge vector bosons can be obtained by replacing the vacuum expectation value of the ϕ field in the EWK Lagrangian with the addition invariant term due to the scalar field part. The H scalar field naturally acquires mass through self-interaction. When they interact with the vacuum, the gauge fields $W_\mu^1, W_\mu^2, W_\mu^3$ gain a mass term in the form of $\frac{1}{2} M^2 W_\mu^\mu W^\mu$, whereas the boson field associated with the photon remains massless. It means that, despite breaking all of the $SU(2) \times U(1)$ symmetries, the Higgs field maintains the $U(1)$ symmetry, leaving the vacuum electrically neutral. Remembering the relations in Eq. 1.5, the masses of the $W^\pm$ and Z bosons can be expressed with the weak coupling constants and the vacuum expectation value as:

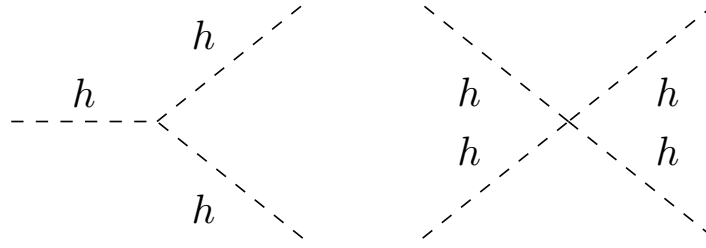
$$m_W = \frac{1}{2} v g \quad m_Z = \frac{1}{2} v \sqrt{g^2 + g'^2} = \frac{g v}{2 \cos \theta_W} \quad m_\gamma = 0. \quad (1.11)$$

The theory does not predict all the free parameters, nevertheless, the input is given from experimental values. For example, it is possible to determine the value of v , the scale of the electroweak theory, by measuring the Fermi constant G_F from the muon lifetime [70]. G_F is related to the W boson mass by $\frac{G_F}{\sqrt{2}} = \frac{g^2}{8m_W^2}$. The resulting value of v is [71]:

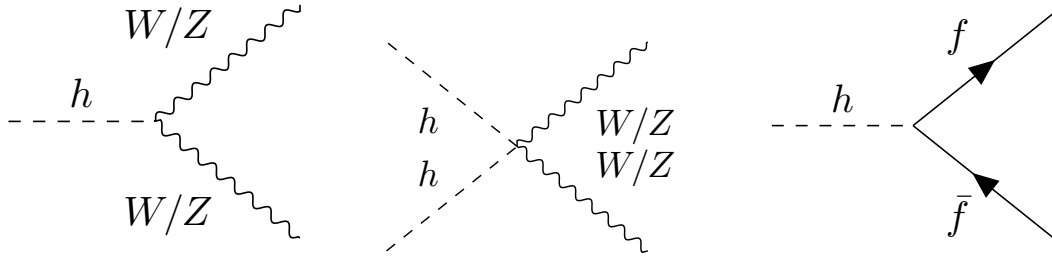
$$v^2 = \frac{1}{\sqrt{2} G_F} \approx (246 \text{ GeV})^2. \quad (1.12)$$

In addition, only with the relatively recent observation of the Higgs boson by the ATLAS and CMS Collaborations and the measurement of its mass, it was first possible to estimate the other free parameter: $\lambda \approx 0.129$.

In the lagrangian density equation, there are also terms corresponding to the couplings with a massive gauge boson (hVV and $hhVV$) and the trilinear and quartic self-couplings exhibit coupling strengths proportional to the Higgs boson mass of $\mu^2\lambda$ and λ , respectively. Once the boson masses and v are known, the SM predicts these couplings uniquely. Figure 1.6 depicts a sketch of the Feynmann diagrams of the SM Higgs interaction vertices.



SM Higgs - Self-interactions.



SM Higgs - Vector bosons interactions.

SM Higgs - Fermions interaction.

Figure 1.6: Feynman diagrams of the elementary vertices of the Higgs boson triple and quartic self-interactions (upper row) and the Higgs boson coupling with fermions and vector bosons (bottom row).

Although the mechanism described above does not incorporate interacting massive fermions, missing couplings can be introduced by extending the Lagrangian further with a term representing the Yukawa interaction that respects the gauge transformations. In particular, the coupling between the fermion field and the Higgs field can lead fermions to acquire their masses. The Yukawa coupling after the symmetry breaking results in the original mass terms proportional to the vacuum expectation value of the Higgs field: $m_f = \frac{\lambda_f v}{\sqrt{2}}$. Also in this case, the masses of the fermions are parameters in theory and are not predicted by the SM, and the input is given from experimental values. Additional terms representing the interaction between the Higgs and fermion fields are also formed, giving rise to the coupling with a strength proportional to the fermion mass.

Finally, theoretical indications and experimental proofs have led to the conclusion that quarks as free particles are mass eigenstates but appear as $SU(2)$ eigenstates in the electroweak interaction: the latter is a combination of the former, and vice versa [72]. In particular, the introduction of quark mixing between EWK q' and mass eigenstates q is needed to explain

the experimental observation of the transitions between the upper and lower components of a quark flavor doublet across different families, e.g. from a u to a s quark, as well as the violation of charge and parity (CP) conservation in the SM [73, 74]. This implies that the quark flavor eigenstates do not coincide with the weak isospin eigenstates, but the lower components of the doublets are rotated relative to one another. More explicitly, the corresponding mixing between the lower components of the weak isospin eigenstates d', s' and b' are related to the lower components of the flavor eigenstates d, s and b as:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \quad (1.13)$$

The rotation matrix V_{CKM} is the Cabibbo-Kobayashi-Maskawa (CKM) matrix describing the transition probabilities between flavors. In the SM, it is assumed to be unitary: the single matrix elements are free parameters of the theory, and all the values have been determined experimentally with very high precision [55, 75]. Large diagonal terms and small off-diagonal values characterize the CKM matrix. As a result, transitions between generations are significantly suppressed, and bottom quarks, for example, acquire a long lifetime. Additionally, as a consequence of quark mixing, flavor-changing neutral currents are not allowed at tree level [76], which explains why they have yet to be observed.

The existence of the mixing is related to many physical examples, e.g. charge-parity (CP) violation, Glashow-Iliopoulos-Maiani (GIM) mechanism, flavor changing neutral currents (FCNC). However, an analogue of this matrix for the leptons is not present in the SM, since no right-handed neutrinos are included in the theory. Only because neutrino oscillations have been detected experimentally [77–79], a similar mixing in the leptonic sector could be introduced to describe neutrinos as both flavor and mass eigenstates. This could be handled by the addition of a lepton mixing matrix, the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix [80, 81], which emerges naturally as a consequence of the seesaw mechanism [82] and has distinct phases depending on whether neutrinos are Dirac or Majorana particles [83]. The values of the masses of neutrinos are still unknown; however upper limits have been determined in experiments down to 0.8 eV [84].

1.1.4 Open questions of the standard model

The SM is a simple and powerful theory with remarkable predictive power. The discovery of the Higgs boson and its characterization by many LHC experiments with increasing precision strengthen the solidity of the SM. Nonetheless, there are some questions that our current knowledge of nature cannot answer, pointing towards the SM being neither complete nor final. Its most notable problems are divided into two categories, and some examples are shortly discussed in the following.

Observational puzzles

These are experimentally-motivated challenges the SM cannot currently explain.

- **Origin of neutrino masses.** Two out of every three neutrinos have a non-vanishing mass, according to neutrino flavor oscillation experiments on astrophysical and atmospheric sources. Even when right-handed neutrinos are included, the extremely small Yukawa coupling values pose a puzzling question. Another possibility could be the Majorana neutrinos, in which the neutrino is its own antiparticle, but this has been proven to result in lepton flavor and unitary violation at high energy [85].

- **Matter-antimatter asymmetry.** The excess of matter over anti-matter in the universe provides incontrovertible proof for the existence of physics beyond the SM. This imbalance suggests that particles and antiparticles behave differently, which neither the SM nor the inflationary cosmological model can explain. Even if there was an initial asymmetry at the Big Bang, the present prediction is extremely small compared to what is observed, suggesting the current theory is incomplete.
- **Presence of dark matter.** There is empirical evidence that the universe is full of an unknown type of matter, referred to as dark matter: rotation curves of galaxies [86], weak lensing measurements [87], microwave background experiments [88]. This state of matter is electromagnetically and strongly neutral, yet it is expected to be massive because it interacts with gravity.
- **flavor anomalies.** Flavor anomaly describes a possible violation of lepton flavor universality (LFU). It is based on the idea that lepton gauge couplings are flavor-independent, and all LEP measurements confirmed this hypothesis. However, recent observation from the BaBar, LHCb, and Belle experiments show tensions at the level of 2-3 sigma with the SM expectations [89–91].

Structural limitations

The structural problems are based on theoretical consideration and indicate that SM lacks robustness at higher energies. As mentioned above, the SM could be a low-energy manifestation of a more fundamental yet undiscovered theory.

- **High energy description of gravity** The general relativity framework already provides a successful gravity theory. However, a fully comprehensive QFT version is still lacking. Although quantum gravity calculations are performed as an effective field theory, the violation of unitarity at the Plank scale ($\Lambda_P \approx 10^{19}$ GeV) is the major obstacle to unifying the SM and the gravitational force.
- **Hierarchy of fundamental scales** The vast difference of magnitude (10^{16}) between the gravitational force and the electroweak scales has yet to be explained, and unifying such far-apart scales is a challenging task for new theories.
- **Origin of generations and mass hierarchy** The Yukawa coupling hides a pattern between the three generations. Many hypotheses attempt to extract an explanation from the first principles, but there is currently no satisfying answer.
- **Lack of strong CP-violation** To date, no CP-violating process in the strong interaction has been observed, causing an unmotivated asymmetry among the gauge interactions. To account for this lack, an additional U(1) symmetry can be introduced: this fine-tuning is known as the Peccei-Quinn theory and is linked to the potential existence of axions.

Numerous theories have been developed to overcome all or part of the above mentioned limitations of the SM. These models frequently include the introduction of additional particles, whose masses can vary widely. Aside from direct searches for new particles, evidence for BSM could be discovered indirectly by assessing deviations from SM expectations. These BSM theories are being extensively tested, primarily by the ATLAS and CMS experiments at the Large Hadron Collider. In this regard, the Higgs sector is vivid and remains relatively unexplored.

1.2 Physics in proton-proton collisions

Following the description of the particle composition of the SM and the theoretical construction of their interactions, this section introduces a phenomenological description of p-p collision physics and explains the concepts used to model collision events with characteristics comparable to real measurement conditions at the LHC. In the high-energy collisions further particles are created as a result of a hard interaction of the partons from both protons. In particular, the cross section of the particle production is a convolution of the structure of both protons and a partonic cross section. More thorough reviews of hadron collider physics is available in Refs. [55, 92]

1.2.1 Factorization and parton distribution functions

At the LHC, protons collide at the highest energies available in the laboratory today. Particle collisions are usually characterized by the center-of-mass energy $\sqrt{s}$ and the number of collision events $N = \sigma_X \cdot \mathcal{L}_{int}$, depending on the cross section and the total amount of recorded data. The center-of-mass energy in collider experiments, such as the LHC, is $\sqrt{s} = 2E$, with E representing the energy of each colliding proton. Being hadrons, protons are consisting of quarks, that for sufficient values of $\sqrt{s}$ are asymptotically free and take directly part in the collision. This has to be accounted for in the calculation of σ_X . The interaction of two incoming partons with high transverse momenta, known as hard scattering, is generally depicted in the simulations. The remaining part of the collision is the so-called underlying event, and it comprises the interaction with other incoming partons at lower transverse momenta as well as the emission of additional particles by the hard scattering partons, either in the initial or final state.

For two partons i and j , the partonic cross section $\hat{\sigma}_{ij \rightarrow X}$ of the given process $ij \rightarrow X$ is calculated perturbatively. Leading order (LO) and next-to-leading order (NLO) are the first and second orders in the perturbation series. In the context of p-p collision simulations, NLO generally refers to the perturbation order in QCD. The cross section of each parton interaction must be convoluted with the parton distribution function (PDF) of the proton components $f(x, q^2)$, which describes the probability of finding a parton carrying a specific fraction x of the total longitudinal momentum at a given scale q^2 . The scale-dependence of the PDFs can be calculated in perturbative QCD. On the contrary, the dependence on x are non-perturbative functions not predicted a priori by QCD and extracted from fits to experimental data, mainly electron-proton collisions [93]. The factorization scale is usually set to the hard scattering scale, which coincides with the normalization scale (the value at which α_s is evaluated). In any case this scale is not a physics parameter and the calculation should not depend on its value. Summing over all possible parton flavor i and j in a certain process, any hadronic cross section $\sigma_{pp \rightarrow X}$ can be factorized as [94]:

$$\sigma_{pp \rightarrow X} = \sum_{ij} \int \int f_i(x_i, q^2) f_j(x_j, q^2) \hat{\sigma}_{ij \rightarrow X}(x_i, x_j, q^2) dx_i dx_j. \quad (1.14)$$

Figure 1.7 illustrates the factorization in a hard scattering process.

The cross section $\hat{\sigma}_{ij \rightarrow X}$ is computed via the Fermi's golden rule with a calculation of the matrix element and the available phase space. The matrix element $\mathcal{M}$ describes the physics of the hard interaction of the initial partons and can be translated into Feynman diagrams. If the relevant couplings are small enough, the matrix element is calculated with the Lagrangian formalism and perturbation theory. As a result, cross section calculations are classified according to the order in which they are computed in perturbation theory. At the same time, using perturbation theory in the cross section calculations gives rise to two characteristic

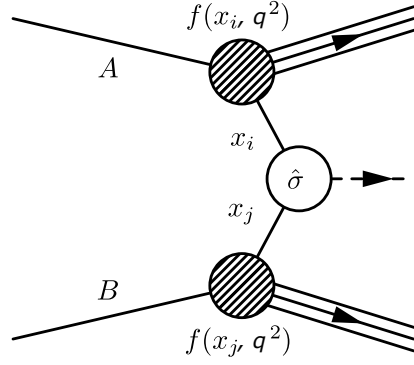


Figure 1.7: Diagram of the factorization in a hard scattering between two hadrons into the parton-level cross section $\hat{\sigma}$ and parton distribution functions $f(x_i, q^2)$.

energy scales [49]. The final cross section may contain nonphysical divergences, which can be removed to ensure meaningful interpretations by re-normalizing the calculation, which is defined by the re-normalization scale μ_R . The perturbative calculation is also limited at a certain threshold by the energy dependence of the coupling constants; the non-perturbative parts factorize at this characteristic energy scale μ_F and are incorporated into the PDFs extracted from observed data. Typically, a consistent scale $\mu_R = \mu_F$ is chosen close to a physics scale of the process. However, in experimental applications, the selection of a singular value is accounted for with systematic uncertainties using independent variations of μ_R and μ_F by factors of 1/2 and 2.

Despite the high predictive power, these calculations can present several mathematical challenges due to effect such as gluon splitting and that can lead to additional divergences. There are several techniques to avoid, one is what to exactly absorb into the PFDs. The scale μ_F is introduced to reduce infrared divergences caused by collinear starting state radiation, it generally represents the smallest scale at which a physical process may be resolved and absorbed into the non-perturbative PDF. In the scope of this thesis two of the schemes are used: for matrix elements derived using the five-flavor scheme (5FS) only the top quark is considered a heavy quark while the masses of the other quarks are negligible and their corresponding contribution is factorized into the PDF set; while for matrix elements calculated in the four-flavor scheme (4FS) the top and the bottom quarks are considered heavy. The latter calculations are generally more difficult, but they are expected to yield a more accurate description of the kinematic behavior of bottom quarks than 5FS derivations [95].

After the hard scattering is generated, the emission of initial state radiation (ISR) and final state radiation (FSR) from the hard scattering event is simulated. The colored final state particles can then emit further radiation, the parton shower (PS), which is modeled by a stepwise parton splitting criterion until the residual energy is too low to continue perturbative calculations. Conceptually, a phase space overlap between the calculation of the matrix element and the simulation of parton showers might occur: it is possible to create the same final state from LO matrix elements paired with parton shower emissions and NLO matrix elements that directly consider additional radiation. Therefore, the ME calculation and PS processes need to be appropriately matched to resolve the ambiguity.

Due to the principle of confinement there are no free color-charged particles, and as a result, these individual particles radiate through a process called hadronization until they are all bound in stable color-neutral hadrons. The partons involved in the initial hard scattering are eliminated from the bound proton state, which is no longer color-neutral. These particles hadronize or interact with one another, resulting in further hadronic activity in the event. This activity is known as the underlying event (UE), and it is described by multiple parton-parton

interactions inside a single collision. Finally, additional collisions in the same or close bunch crossings are simulated independently and then overlapped to the event of interest to obtain a realistic simulation of LHC p-p collision.

The chain of these interactions for collision events, starting from the hard interaction at high scales down to low scales for hadronization and particle decay, are tackled phenomenologically in dedicated tools based on the Monte Carlo technique.

1.2.2 Monte Carlo event simulation in proton-proton collisions

Usually, precise tests of the SM and statistical data analysis in general are carried out by performing comparison of observables measured at the collision experiment with the corresponding theory predictions. For this purpose a good description of the how the original partons (parton-level objects) arrive to the measured stable hadrons (detector level particles) is needed, starting from the hard interaction at high scales, via the fragmentation, down to hadronization at low scales and the decay of the unstable particles as well as the simulation of the detector response itself. In addition, most LHC analyses require the ability to realistically simulate the events for various physics processes with high accuracy; it is essential for calibrating the detector, tuning the trigger selection, optimizing the physics analysis and model possible BSM interactions. Processes in p-p collisions cannot be properly predicted on an event-by-event basis since they are expressions of quantum physics and in addition protons are composite objects. The theory based simulated data is commonly generated with Monte-Carlo (MC) event generators and the output can be directly compared to the observed data.

The steps of the simulation can be performed independently thanks to the factorization theorem. For each part of the chain, dedicated simulation tools are available to simulate the theory prediction with varying degrees of accuracy. Figure 1.8 depicts a $t\bar{t}H$ event sketch of a hadron collision event, including both the hard scattering mechanism and the underlying event, as could be simulated by a MC generator.

Nowadays, the vast majority of the calculations is obtained with numerical integration methods implemented in dedicated tools such as PYTHIA8 [96], POWHEG [97–99], HERWIG++ [100] and MADGRAPH5_AMC@NLO [101]. After the event simulation is performed, all particles generated in the preceding phases are interfaced to a simulation of the CMS detector based on the GEANT4 [102] toolkit to account for their interaction with detector components. To obtain a consistent comparison the same reconstruction technique as in the data is used (see Chapter 3). Simulation samples are generated individually for each year.

The distinction among the various stages of information processing in particle physics is established by using specific terminology. The simulated particles are referred to as at generator level for partons, or as at particle level (or at truth level in Chapter 7 in this thesis) for stable particles. Finally, measured and simulated data that have undergone detector simulation are referred to as at reconstruction level.

In general, for the analysis carried out in this thesis, the MEs for the signal and background processes are created at NLO of perturbation theory using POWHEG or MG5aMC(NLO), or at LO using PYTHIA8.2 (it will be referred to as PYTHIA8 in the following), depending on the process. The parton distribution functions supplied by the NNPDF [103] describe the probability density function for the momenta of the partons in the initial-state protons (the specifics are listed in Section 5.2.2). PYTHIA 8.2, based on the phenomenological Lund string model [104, 105], is used to simulate the parton showering and hadronization processes in all simulated samples. Because it is phenomenological, this model has tuneable parameters that must be adjusted using observed data. The energy scale utilized for matching the ME and parton-showering diagrams is one of these characteristics (ME-PS matching); the other

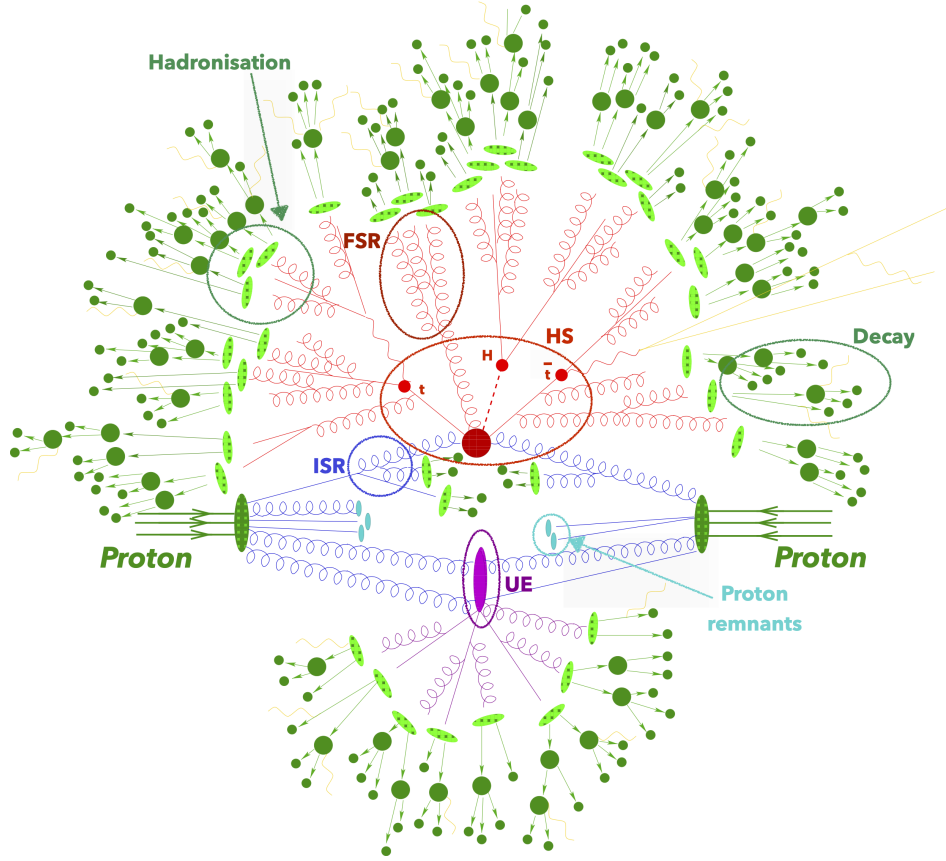


Figure 1.8: Sketch of a $t\bar{t}H$ event produced from a p - p collision as produced by an event generator. The hard scattering is shown (big red bullet), followed by the decay of both top quarks and the Higgs boson (small red bullets). Additional hard QCD radiation is produced as initial and final state radiation. Finally, additional multi-parton interactions take place: the underlying events (purple bullet). The hadronization products are shown in green, and photon radiation in yellow.

parameters, known as the UE tuning, are utilized to characterize the underlying events of the collisions.

Section 5.2 provides the details of the different simulated samples used in this thesis for the measurement of the $t\bar{t}H, H \rightarrow b\bar{b}$ process.

1.3 The $t\bar{t}H$ process at hadron colliders

With the final components from Section 1.1 in place, the SM is an incredibly accurate model of the known universe, but there are still questions to be answered.

Following the discovery of a new boson with a mass of 125 GeV consistent with the SM Higgs boson, the focus has shifted to measuring its unique properties. Motivated by this goal, the target here is the top-Higgs Yukawa coupling λ_t .

The direct measurement of the coupling between the top quark and the Higgs boson at tree level can occur when the Higgs is produced in association with a top-anti-top quark pair ($t\bar{t}H$). This coupling links a top quark and the Higgs boson, resulting in the production cross section being dependent only on the λ_t and other SM-known couplings. Therefore these measurements have been some of the highlights of Run 2 and are one of the main subjects of this thesis.

The main features of the Higgs boson and the top quark are discussed in the following: from their production mechanisms at hadron colliders to the signatures of their decay modes. Then the focus is on the $t\bar{t}H$ signal and the other physics processes that represent the major backgrounds of the analysis presented in this thesis. In particular, the production of a top quark pair with additional heavy-flavor jets is fundamental, and its theoretical description extremely challenging. The section concludes with a summary of experimental results from the previous analyses.

1.3.1 The Higgs boson

The Higgs boson represents the quantum of the scalar field introduced in Section in the discussion of the electroweak symmetry breaking. In the SM, the Higgs boson is expected to be electrically neutral, free of spin, and CP-even ($J^P = 0^+$) with a decay width of ≈ 4 MeV.

The LHC at the European Organization for Nuclear Research (CERN) (Section 2.1) was developed with the primary goal of verifying the SM hypothesis and discovering new physics beyond it via the exploration of the properties of the Higgs boson in particular.

Production mechanisms

The Higgs boson mass is a free parameter of the SM and needs to be determined experimentally. Therefore the production cross section and its decay branching ratios are usually computed as function of its mass, as shown in Figure 1.9 (a) and (b) at the center of mass $\sqrt{s} = 13$ TeV.

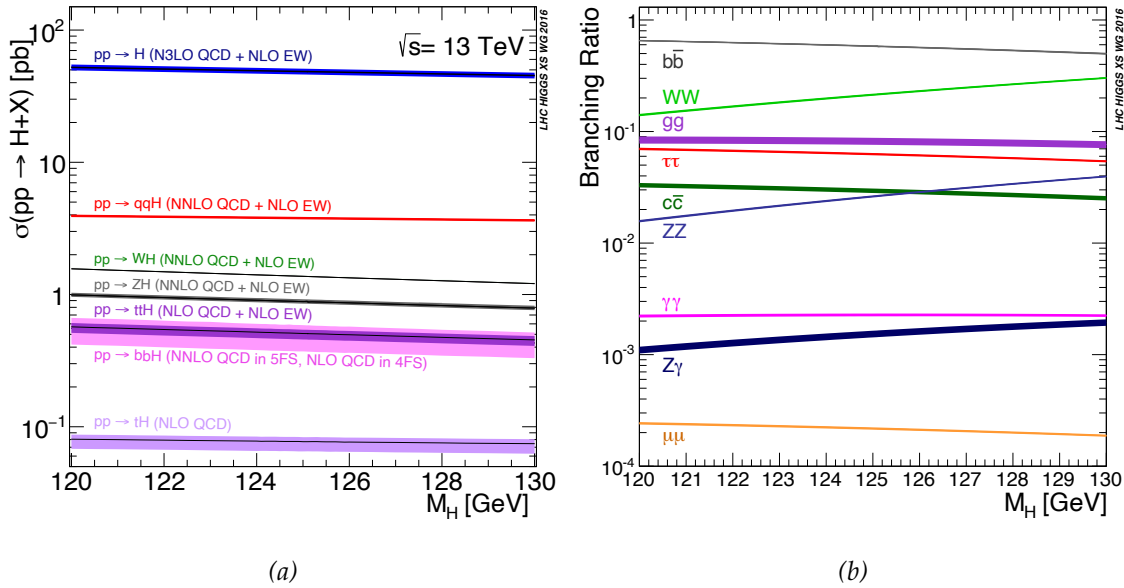


Figure 1.9: (a) The Higgs boson production cross sections and (b) decay branching ratios as a function of Higgs boson mass. Calculations are performed up to the order indicated in the plot, and the color bands represent the respective uncertainty [21].

At hadron colliders, there are several production mechanisms to produce a Higgs boson. The dominant mechanism originates from gluon fusion mediated by a virtual loop of heavy quarks (ggH , Figure 1.10 (a)). Because of their large mass, the top and bottom quarks contribute the most to the loop: this is the largest production cross section of almost 49 pb at $\sqrt{s} = 13$ TeV.

The second most common mechanism is the vector boson fusion (VBF or qqH) with a cross section of about 3.8 pb at $\sqrt{s} = 13$ TeV. In this case, the Higgs boson is produced by

the fusion of two vector bosons radiated by initial state quarks, as $qq \rightarrow V^*V^* \rightarrow qqH$, with $V = W/Z$ (see Figure 1.10 (b)). It can easily be discriminated against the background because the two additional quarks in the final state are detected as two hard jets with a tendency to be produced along the beam line, and they are highly energetic with high invariant mass.

In the low mass region there are other two mechanisms, the associate production with vector bosons (VH) (Figure 1.10 (c)) and with heavy quarks (tH or $t\bar{t}H$) (Figure 1.10 (d)).

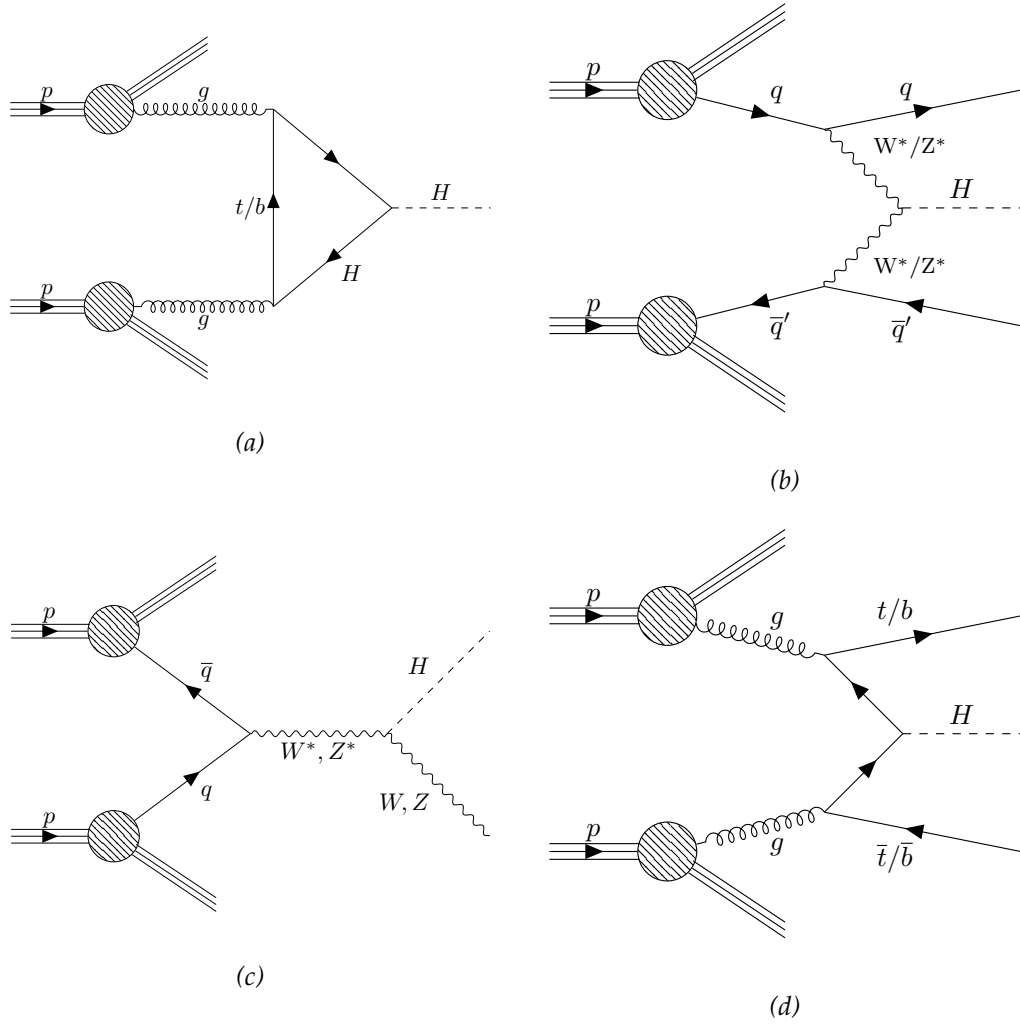


Figure 1.10: Leading Feynman diagrams for Higgs boson production via gluon-gluon fusion (a), vector boson fusion (b), and associate production with vector boson (c) and with heavy quarks (d).

The first process occurs when a virtual vector boson decays to its on-shell state, radiating a Higgs boson. Studying the VH (and VBF) production mechanisms allows for probing the Higgs boson coupling to vector bosons. Despite the lower cross section of 2.3 pb at $\sqrt{s} = 13$ TeV, of which ≈ 0.9 pb due to the ZH process, the VH production mechanism has a quite distinct signature: the presence in the final states of leptons coming from the electroweak decays of the vector bosons is helpful for triggering and background rejection.

The production in association with heavy flavor quarks, on the other hand, is particularly challenging due to the presence of a large number of hadronic jets in the event. Despite this and the smaller cross section of ≈ 0.5 pb at $\sqrt{s} = 13$ TeV with respect to other production modes, in $t\bar{t}H$ the coupling involves non-virtual top quarks and allows for direct measurement of the Higgs-top Yukawa coupling. In this respect, the production of a Higgs boson with

a single top quark is theoretically an option, but its cross section is dramatically reduced due to negative interference of leading-order diagrams. As a result, the analysis presented in this thesis concentrates on $t\bar{t}H$ production. In addition, for $t\bar{t}H$ measurements, prompt leptons from top quark decays provide a handle to select signal events. On the contrary, the Higgs production mode with a $b\bar{b}$ ($b\bar{b}H$) pair has a similar cross section, but it is experimentally very challenging to disentangle from the background.

Decay channels

In the SM, the Higgs boson has a mean lifetime of about 10^{-22} s, which is too short to detect unless through its decay products. The Higgs boson couples to fermions and vector bosons proportionally to their masses therefore it tends to decay in the heaviest particles allowed by the kinematics. As shown in Figure 1.9 (b) also, the theoretical branching fractions of the Higgs boson depend on the mass of the Higgs boson.

The dominant decay $H \rightarrow b\bar{b}$ into a pair of bottom quark-antiquark pairs ($b\bar{b}$), with a branching ratio of $\approx 58\%$, is fundamental to study the coupling of the Higgs boson with fermions but overall quite challenging to identify due to the presence of jets in a hadronic environment. The second fermionic decay mode is $H \rightarrow \tau^+\tau^-$, occurring $\approx 6\%$ of the time and easier to identify. Other decays into fermions have low branching ratios, however recently CMS has reported evidence for $H \rightarrow \mu^+\mu^-$ [106].

Decays into the vector bosons W and Z are suppressed since the Higgs boson is too light to decay into two vector bosons on their mass shell. If $m_H \geq 2m_V$ there is the possibility of the decay into real vector boson. Below this kinematic threshold, the Higgs boson decays into one or two virtual gauge bosons, decaying then into fermions. For a 125 GeV Higgs boson, the decay fractions to W^+W^- and ZZ^* are 21% and 2.6%, respectively.

The Higgs boson does not couple to massless particles. Therefore the decay modes in two gluons or two photons are induced through heavy particle loops with branching ratios at the percent and per-mille levels, respectively. Despite its sizable branching ratio, the decay into gluons is unlikely to be detected at hadron colliders. Conversely, the signature of $H \rightarrow \gamma\gamma$ is really clear with two highly energetic photons emitted back-to-back. It has a significant impact as one of the golden channels for Higgs boson searches at the LHC.

Higgs boson decays branching fractions are summarized in Table 1.1 for a Higgs boson of mass $m_H = 125.4$ GeV, currently the most precise mass measurement [107].

Decay mode	$\mathcal{B}$
$H \rightarrow b\bar{b}$	57.63 ± 0.70
$H \rightarrow W^+W^-$	22.00 ± 0.33
$H \rightarrow gg$	8.15 ± 0.42
$H \rightarrow l\bar{l}$	6.23 ± 0.09
$H \rightarrow c\bar{c}$	2.86 ± 0.09
$H \rightarrow ZZ^*$	2.71 ± 0.04
$H \rightarrow \gamma\gamma$	0.227 ± 0.005
$H \rightarrow Z\gamma$	0.157 ± 0.009

Table 1.1: Branching fractions of the main Higgs boson decay modes for a SM Higgs boson of mass $m_H = 125.38$ GeV [10].

Phenomenology at the LHC

The discovery of a new particle, the Higgs boson, with a mass of around 125 GeV, was announced on the 4th July 2012 by ATLAS [9] and CMS [108] Collaborations using the data collected at center of mass energies of 7 and 8 TeV. The combination of CMS data resulted in an observed (expected) statistical significance of 5.0 (5.8) standard deviation, driven mainly by the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ$ decay channels. It is worth mentioning here that these analyses, originally developed to identify a significant signal excess over a continuous background, have now evolved into a precision physics analysis in which many of the properties of the Higgs boson can be measured to probe in detail the SM predictions.

Following the discovery claim, the analyses focused on measuring the properties. First, using the angular distribution of leptons in the bosonic decay channels, the Higgs boson charge-parity (CP) and its spin are measured to be $J^P = 0^+$ [109, 110]. The natural width of the Higgs boson has been extracted and is found to be $\Gamma = 3.2 \pm 2.4$ MeV by using off-mass-shell and on-mass-shell Higgs boson production. The first measurement of its mass yields $m_H = 125.09 \pm 0.21(\text{stat}) \pm 0.11(\text{syst})$ [111, 112]; both mass and width measurements are found to be consistent with the predictions of the SM.

Currently, the most precise measurement of m_H results from a combination of the CMS $H \rightarrow ZZ \rightarrow 4\ell$ and $H \rightarrow \gamma\gamma$ decay channels using the full Run 1 statistics, and the 2016 Run 2 data [107]:

$$m_H = 125.38 \pm 0.11 (\text{stat}) \pm 0.08 (\text{syst}) \text{ GeV}. \quad (1.15)$$

A summary of the individual and combined measurements with the Run 1 and 2016 data sets is shown in Figure 1.11.

The observation of this new particle represented only the first step in exploring the scalar sector and the possibility of setting its properties. The precision measurement of the value of m_H is fundamental to setting the values of the production and branching ratios, thus determining which production mechanism and decay modes can be experimentally probed at the LHC. The observation of $t\bar{t}H$ production mode by ATLAS and CMS Collaborations in 2018 is worth mentioning [22, 113] (more in Section 1.3.4). As of today, all the production mechanisms but the $b\bar{b}H$ and tH have been observed at the LHC. Additionally, with the data collected at $\sqrt{s} = 13$ TeV, the CMS Collaboration observed the Higgs boson decaying into a pair of τ leptons with a significance of 5.9 standard deviations [114], into a pair of bottom quarks with a significance of 5.6 standard deviations [23]. In addition, a significance of 3 standard deviations has been reached for the Higgs boson decaying into muons [106]. The latter marks a major step forward in understanding the SM by confirming the expected proportionality between the fermionic couplings of the Higgs and the fermion masses. However, the couplings to first generation fermions still represent a daunting challenge [10, 11].

Figure 1.12 highlights the present status of the results by the CMS Collaboration for the measurement of the Higgs boson couplings to fermions and vector bosons, also showing in this case a remarkable agreement with SM predictions.

The agreement between the SM prediction and the observed measurement is usually quantified by the signal strength parameters μ . For given initial (i) and final (f) states, $i \rightarrow H \rightarrow f$ the signal strengths are defined as $\mu_i = \frac{\sigma_i}{\sigma_i^{\text{SM}}}$ and $\mu_f = \frac{\mathcal{B}_f}{\mathcal{B}_f^{\text{SM}}}$, where σ is the production cross section and $\mathcal{B}$ the branching fraction. Perfect agreement with SM expectations would correspond to all μ equal to one. By fitting all data from production modes and decay channels with a common parameter, the overall signal strength has been measured to be $\mu = 1.002 \pm 0.057$ [10]. Analogous measurements are performed by relaxing the assumption of a common signal-strength parameter and treating each production and each branching ratio signal strength independently. Figure 1.13 shows the best fit results.

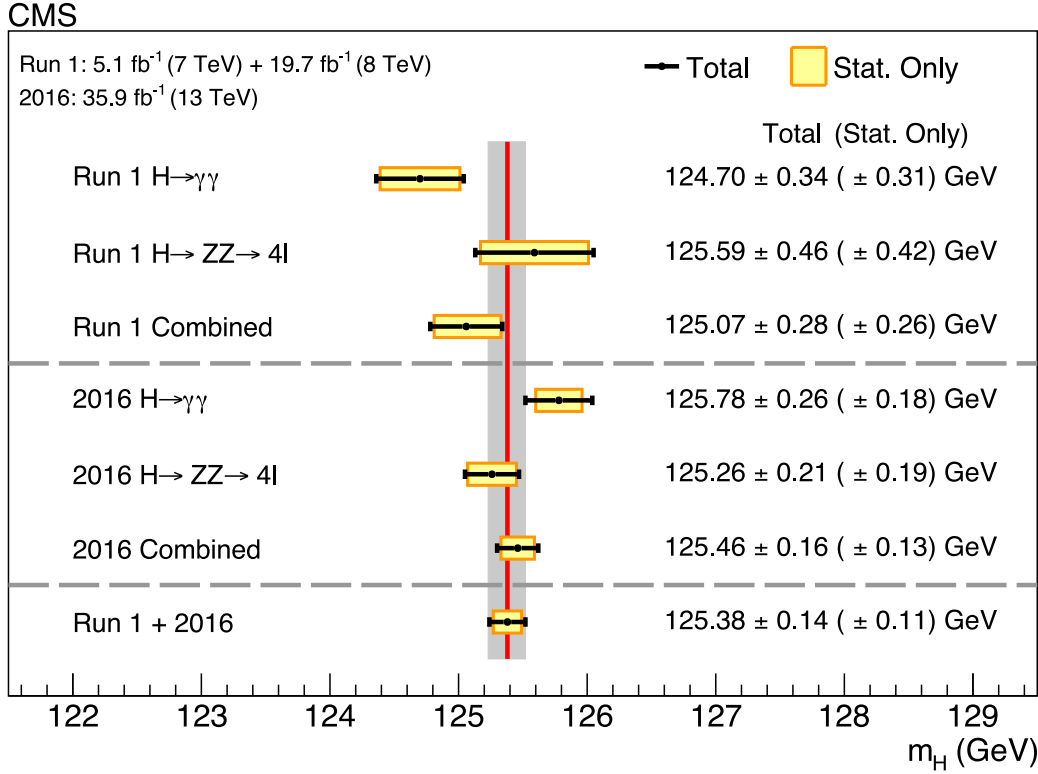


Figure 1.11: A summary of the measured Higgs boson mass in the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ \rightarrow 4\ell$ decay channels and for their combination. The statistical (wider, yellow-shaded bands) and total uncertainties (black error bars) are reported. The (red) vertical line and corresponding (gray) shaded column indicate the central value and the total uncertainty of the Run 1 + 2016 Run 2 combined measurement, respectively [107].

Currently, with access to a precise value of m_H , an even more stringent test of the SM is conceivable: a precision measurement of the self-coupling of the Higgs boson. The most natural way to measure this observable is to exploit the pair production of Higgs bosons. The search is performed by combining Higgs boson candidates reconstructed from different final states and expressed as an upper limit on the production cross section. With the current data set, the result is found to be less than 3.4 times the SM expectation at 95% confidence level [10]. Projections for the HL-LHC for $\mathcal{L} = 3000 \text{ fb}^{-1}$ [115] expect the limit to be below unity, showing enough sensitivity to establish its existence [10].

In conclusion, so far, all measurements are compatible with SM predictions, and searches for additional Higgs bosons that models beyond the SM can only explain were conducted without finding significant evidence. At the same time, the accuracy of analyses in the Higgs sector is still evolving, and signs of new physics may be observed in the future. The LHC experiments are entering the era of precision Higgs physics that will shed light on BSM physics [10, 11].

1.3.2 The top quark

The existence of the third-generation quarks was suggested to explain observations of CP violation. The observation of the b quark at Fermilab in 1977, confirmed the hypothesis and opened the quest for the top quark. The first experimental observation was reported independently by the CDF and DØ Collaborations at the Fermilab Tevatron using proton-

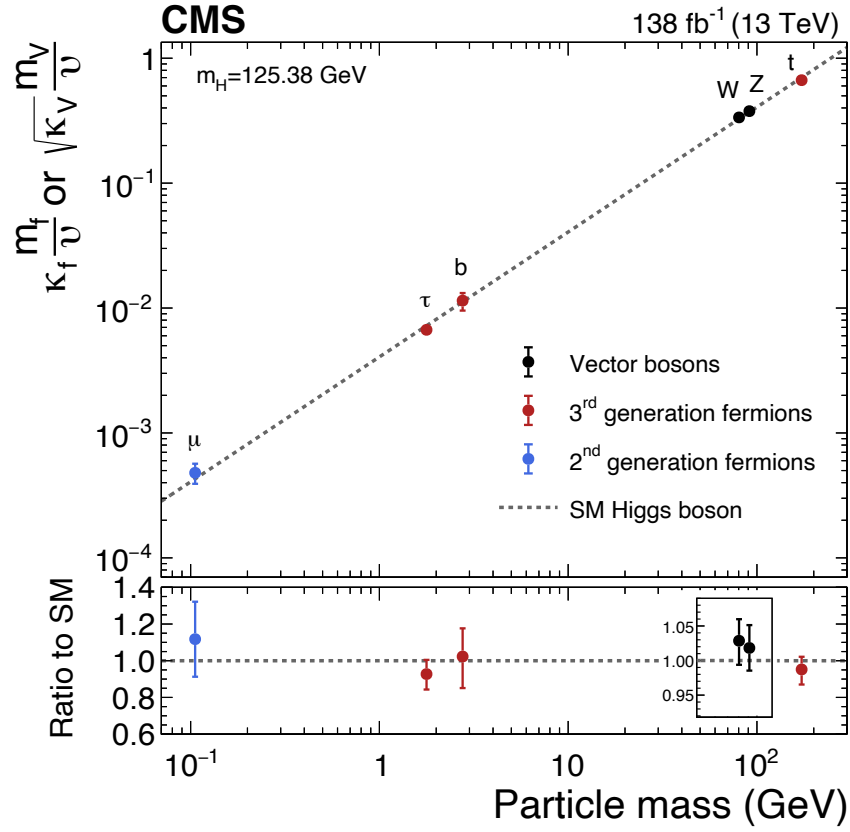


Figure 1.12: The measured coupling modifiers of the Higgs boson to fermions and heavy gauge bosons, as functions of fermion or gauge boson mass, where v is the vacuum expectation value. The square root of the coupling modifier is shown for gauge bosons to keep a linear proportionality to the mass. The p -value with respect to the SM prediction is 37.5% [10].

antiproton collisions at a center of mass energy of $\sqrt{s} = 1.8$ TeV. The following focuses on the production mechanism at p-p colliders, decay channels, and main properties.

The top quark has an electrical charge of $Q = +2/3 e$, a spin of $1/2$, and is subject to all three fundamental forces described by the SM. With $m_t = 172.5 \pm 0.7$ GeV, the top quark is the most massive particle known. It exhibits the largest Yukawa coupling to the Higgs boson; thus, the accurate measurement of Higgs- top coupling is essential for probing predictions of the SM and constraining BSM models. It has a total decay width of about 1.42 GeV and a lifetime $\tau_t \approx 5 \cdot 10^{-25}$ s, which is uniquely shorter than the time scale of hadronization processes. As a result, top quarks decay before forming a bound state, providing the unique opportunity to investigate the decay of a bare quark, i.e. the charge and polarization.

Production mechanisms

Top quarks are mostly produced as quark-antiquark pairs ($t\bar{t}$) in p-p collisions. At the LHC, the $t\bar{t}$ process is abundantly produced, and it allows for very precise SM measurements. At the same time, these mechanism represents a major background to the $t\bar{t}H, H \rightarrow b\bar{b}$ analysis reported in this thesis, as further described below.

The gluon-gluon fusion processes account for approximately 90% of the $t\bar{t}$ cross section for p-p collisions [55]. The leading-order Feynman diagrams are illustrated in Figure 1.14 (a) and (b). When considering higher orders, the $t\bar{t}$ system is often accompanied by emissions of additional jets.

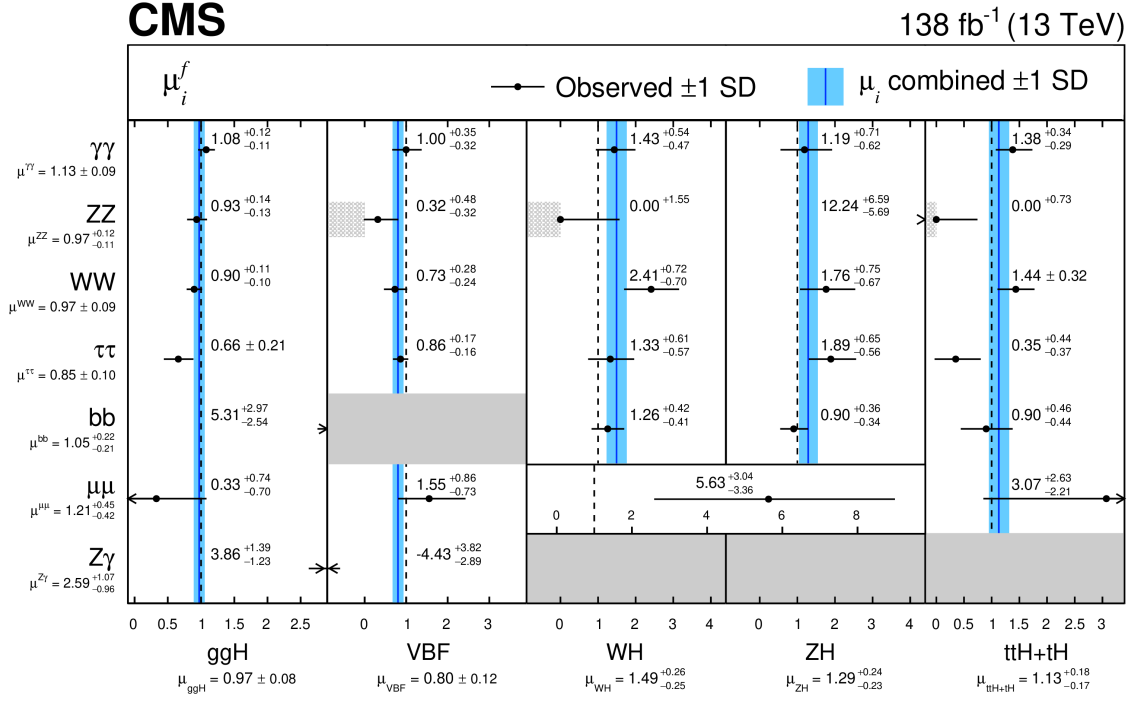


Figure 1.13: Signal strength parameters per individual production mode and decay channel, and combined. The dashed vertical lines at 1 represent the SM value. Light gray shading indicates that μ is forced to be positive; dark gray shading indicates the absence of measurement. The p -value with respect to the SM prediction is 5.8% [10].

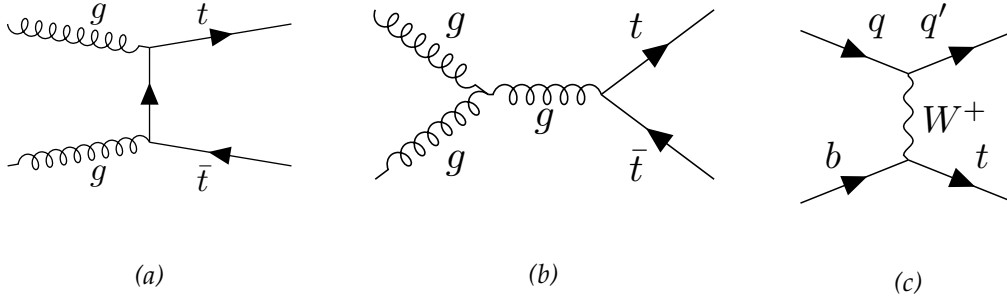


Figure 1.14: Feynman diagram representations for the most prominent top quark production: (a) and (b) top quark pair production processes at leading order and (c) single top quark production.

The $t\bar{t}$ production cross section has been studied at various center of mass energies for p-p collisions. The measurements are compared to the theoretical prediction, showing good agreement as in Figure 1.15. For $\sqrt{s} = 13$ TeV, the predicted $t\bar{t}$ cross section for $m_t = 172.5$ GeV is [116]:

$$\sigma_{t\bar{t}} = 833.9^{+20.5}_{-30.0}(\text{scale}) \pm 21.0(\text{PDF} + \alpha_s)^{+23.18}_{-22.45}(\text{mass})\text{pb}. \quad (1.16)$$

In the second group of processes, a single top quark is produced via weak interactions: three production channels can be distinguished based on the function of the involved W boson. An example is illustrated in Figure 1.14 (c). Assuming a top quark mass of $m_t = 172.5$ GeV the cross section of all production processes amounts to $299.01^{+9.83}_{-8.62}$ pb [116].

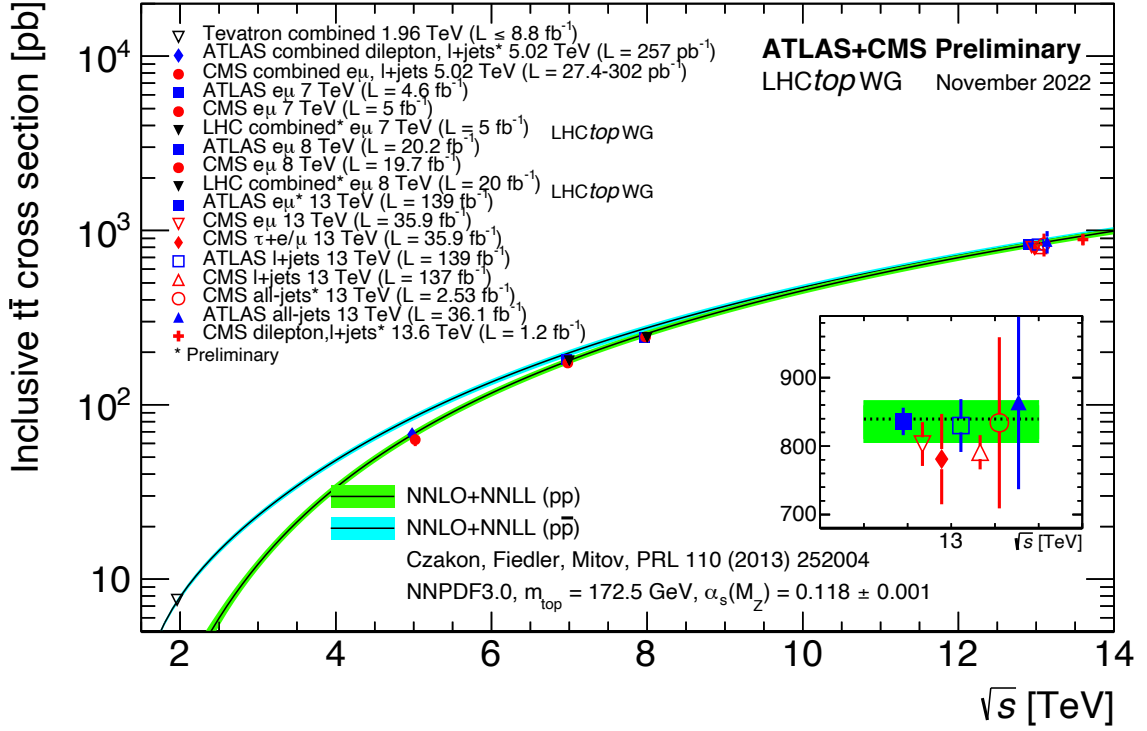


Figure 1.15: Measurements of the top-pair production cross section as a function of the center-of-mass energy [116]. The measurements and the theory calculation are quoted at $m_t = 172.5$ GeV. Measurements made at the same center-of-mass energy are slightly offset for clarity.

Decay channels

The top quark decays exclusively via electroweak interactions into a W boson and a b-type quark due to the CKM matrix element of $|V_{tb}| \approx 1$. Decays into strange or down quarks are theoretically allowed but highly suppressed due to the corresponding vanishing off-diagonal values $|V_{ts}|$ and $|V_{td}|$. The branching ratio is $\mathcal{B}_{bW} = (95.7 \pm 3.4)\%$ [55]; therefore the decays of the top quarks are typically categorized by the decays of the two involved W bosons as sketched in the pie chart in Figure 1.16.

The W boson decays into hadrons with a branching ratio of $\mathcal{B}_{\text{hadr}} = (67.41 \pm 0.27)\%$ [55], which then hadronize and are experimentally reconstructed as jets. In addition, leptonic decays are evenly distributed over the three generations due to the lepton flavor universality: $\mathcal{B}_e = (10.71 \pm 0.16)\%$ and $\mathcal{B}_\mu = (10.63 \pm 0.15)\%$ for the decays into electrons and muons, including their corresponding neutrinos, respectively. Decays involving τ leptons are not considered in the analysis presented in this thesis; however they decay weakly into additional leptons, e.g. $\tau^- \rightarrow \nu_\tau \ell^- \nu_\ell$, which could be potentially misidentified with those of the primary W boson decay.

The three distinct channels of the $t\bar{t}$ decays are:

- **fully-hadronic (FH)** Both W bosons decay into quarks, mostly as $W^+/W^- \rightarrow u\bar{d}/\bar{u}d$ and $W^+/W^- \rightarrow c\bar{s}/\bar{c}s$, they hadronize and lead to the experimental signature of six jets in the final state. Despite the high branching ratio of about 45%, it is challenging due to the background contributions from the multijet processes and the combinatorial ambiguities in the reconstruction of the events.

$q\bar{q}'$	Semileptonic (~14.4%)	Fully hadronic (~45.5%)
$\tau^+\tau^-$		
$\mu^+\mu^-$		
e^+e^-	Dileptonic (~4.6%)	Semileptonic (~14.4%)
$W \rightarrow e^+\nu_e \mu^+\nu_\mu \tau^+\nu_\tau$		
$q\bar{q}'$		

Figure 1.16: A schematic overview of the $t\bar{t}$ branching ratios and final states. The sections do not correspond to the actual branching fractions. Due to the difficulties in the reconstruction, leptonic decays involving τ leptons are not considered as signal.

- **semi-leptonic or single-lepton (SL)** Only one W decays into hadrons and the other decays into leptons: the final state comprises four jets, one charged lepton, and one neutrino in the final state. The branching ratio is $\approx 44\%$, but the multijet background is significantly experimentally suppressed by the isolation requirements on the charged lepton. Furthermore, to reconstruct the complete event kinematic, the momentum vector of a single neutrino can be retrieved with good efficiency from a missing transverse momentum measurement in combination with a constraint on the W boson mass.
- **dilepton (DL)** Both W bosons decay into leptons: the final state features two jets, two oppositely-charged leptons, and two neutrinos. The combined branching ratio $\approx 11\%$ is the smallest, but the signature is distinct. The full event reconstruction is challenging because the neutrinos are neither detectable nor unambiguously retrievable utilizing the missing transverse momentum balance.

This thesis explores the DL decay channel with electrons or muons in the final state, corresponding to a branching ratio of $\mathcal{B}_{t\bar{t},DL} = (\mathcal{B}_{e\nu} + \mathcal{B}_{\mu\nu})^2 = (4.6 \pm 0.1)\%$.

1.3.3 Signal and relevant background processes

Various experimental factors, ranging from object misidentification to detector resolution effects, could lead two processes to appear similar even though their nominal final states are different. Generally, a process is considered background if it cannot be distinguished from the process of interest by requiring specific thresholds on measurable observables.

In the following, the main features of the $t\bar{t}H$ signal process and its relevant background processes are presented.

Signal process

This analysis focuses on the $t\bar{t}H$ production, with the Higgs bosons decaying into a pair of bottom quarks.

Figure 1.17 (a) shows as example a leading-order Feynman diagram for $t\bar{t}H, H \rightarrow b\bar{b}$ in the DL decay channel. There are more diagrams describing $t\bar{t}H$ production, but similarly to $t\bar{t}$ production, the gluon-initiated processes have the largest contribution to the total cross section. At $\sqrt{s} = 13$ TeV, the net cross section is $\sigma_{t\bar{t}H, H \rightarrow b\bar{b}, \text{DL}} = \sigma_{t\bar{t}H} \mathcal{B}_{H \rightarrow b\bar{b}} \mathcal{B}_{t\bar{t}, \text{DL}} \approx 13.34$ fb, corresponding to ≈ 1800 produced events in the data set recorded by the CMS detector in the period 2016-2018 with an integrated luminosity of 138 fb^{-1} .

The Higgs bosons decaying into a pair of bottom quarks are considered here due to the largest branching ratio and because it explores the decay into 3rd generation quarks, which has not been investigated in much depth yet. The bottom quarks only decay via electroweak interactions, and the decay amplitudes are subject to small-off diagonal CKM matrix element V_{ub} and V_{cb} . Therefore when bottom quarks hadronize and form B hadrons, the lifetime is relatively long $\tau_0 \approx 1.6$ ps, and they travel a measurable flight distance before decaying at a secondary vertex. Typical distances are ≈ 1 cm (for $p_b = 100$ GeV, $d = c\tau_0 p_b / m_b$) and usually a secondary vertex is identified from the reconstruction of the decay products trajectories.

In the DL channel, the measurable final state consists of two isolated leptons with opposite charges and four jets, the result of B hadron decays, and a considerable amount of missing transverse momentum due to the undetectable neutrinos. The full reconstruction of the event is extremely challenging due to the high combinatorial complexity caused by the number of jets. More details in Section 5.2.2.

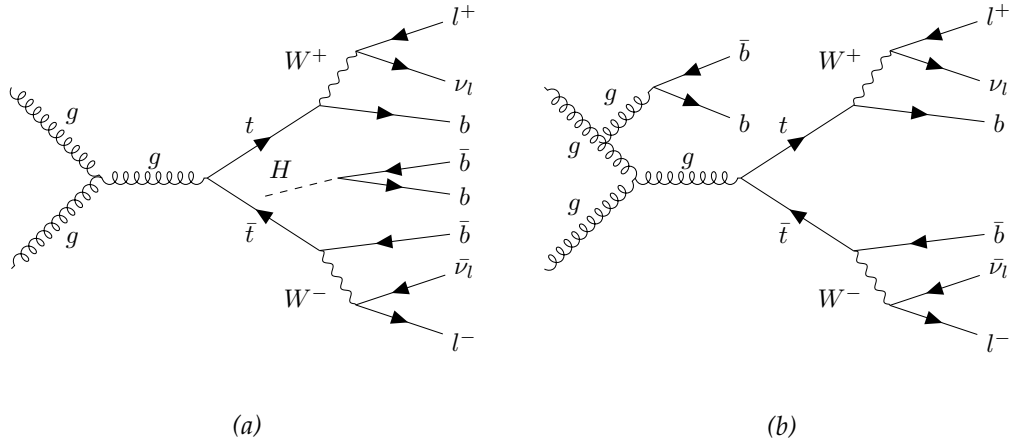


Figure 1.17: (a) Leading-order Feynman diagram for the signal process $t\bar{t}H, H \rightarrow b\bar{b}$ and (b) for the background $t\bar{t} + b\bar{b}$ process in the DL channel, with $\ell = e, \mu$ being the leptons considered for the scope of this analysis. It should be noted that more possible diagrams exist for both processes. To highlight, their final state is identical.

Background processes

The most dominant background for the $t\bar{t}H$ process arises from top quark production. Given the same final states of the $t\bar{t}$ system, the inclusive cross section is ≈ 2800 times larger. In the case of additionally emitted jets ($t\bar{t} + \text{jets}$), mostly heavy-flavor jets ($t\bar{t} + \text{HF}$), the process has very similar final-state content and kinematic topology as the $t\bar{t}H, H \rightarrow b\bar{b}$ one. Figure 1.17 (b) shows a Feynman diagram for top quark pair production in the DL channel, with two b jets in the final state.

From the theoretical point of view, the production mechanism of the $t\bar{t} + H$ contributions is ambiguous due to the multi-parton final state with two very different scales. Particularly important among these are $t\bar{t}$ events with additional b-jet production: processes with additional bottom quarks can be modeled via $gg \rightarrow t\bar{t}g$ matrix element calculations in (N)LO with collinear $g \rightarrow b\bar{b}$ splittings in the parton shower, $gb \rightarrow t\bar{t}b$ production using 4FS and 5FS calculations, or solely via collinear gluon splittings in the parton shower [44, 45, 117, 118]. In addition, dedicated measurements [119, 120] and observations in previous measurements of the $t\bar{t}H$, $H \rightarrow b\bar{b}$ process [121] show that the $t\bar{t} + b$ -jets contribution in data is larger than predicted by the simulation, often by a factor of 1.2-1.4, depending on the generator.

However, a robust and accurate process definition is essential to define background-enriched phase space regions to increase the measurement sensitivity and to properly assign the different theoretical uncertainties. In this analysis, the most relevant background contributions are also separated among each other (see Sections 5.2.2 and 5.7.2). The definition of the $t\bar{t}$ subprocesses is based on the number of additional jets and the content of b and c hadrons.

Additionally, other processes listed below are included in the overall expected background even if their final state signatures are relatively different because of comparably large cross sections and the possibility of additional emissions of jets and object misidentification. More detailed descriptions of the processes is in Section 5.2.2. A summary of their cross section is presented in Table 5.4; if not stated otherwise, cross sections are calculated at $\sqrt{s} = 13$ TeV, assuming $m_t = 172.5$ GeV, $m_W = 80.398$ GeV, and $m_Z = 91.1876$ GeV [55].

- **Single $t/\bar{t}$** Production of a single top (anti)quark as introduced in Section 1.3.2.
- **V+jets** Production of a single W or Z vector boson and additional jet emissions. These processes can resemble the signal, especially in decay into lepton.
- **Diboson** Production of a pair of vector bosons.
- **$t\bar{t} + V$** Production of a top quark pair in association with an additional vector boson (W,Z, γ).

1.3.4 From previous measurements to projected sensitivity

Several $t\bar{t}H$, $H \rightarrow b\bar{b}$ analyses in the leptonic channels have been performed at the LHC with p-p collisions at different center-of-mass energies, both by ATLAS and CMS experiments. The upper limits, the measured $\mu_{t\bar{t}H}$ signal strengths, and corresponding significances for each analysis are summarized in Table 1.2. It should be noted that the stated partial results of the CMS experiment at $\sqrt{13}$ TeV with 138 fb^{-1} include the work presented in this thesis.

All results are compatible with expectations derived from the SM.

At $\sqrt{s} = 13$ TeV, for both experiments, the measurement uncertainties are dominated by systematic effects contributions, mainly driven by the uncertainty prediction of $t\bar{t}$ production in association with additional heavy flavor jets, as will be further discussed in later chapters. To cope with this challenge arising from the background modeling, various analysis strategies involving sophisticated machine learning algorithms were employed. The strategy developed for the analysis presented here is discussed in detail in Section 5.7.

The most sensitive result to date is expected to be the most recent one performed by the CMS Collaboration at $\sqrt{s} = 13$ TeV, in which the work presented in this thesis is included. The measurement uses the recorded collision data corresponding to an integrated luminosity of 138 fb^{-1} and constitutes the first combined measurement of $t\bar{t}H$ and tH production in the $H \rightarrow b\bar{b}$ channel where all the three decay channels of the $t\bar{t}$ system. This is expected to

	$\sqrt{s}$ [TeV]	$\mathcal{L}$ [fb ⁻¹]	Strategy	$\sigma/\sigma_{\text{SM}}$	Significance (expected)	
ATLAS	8	20.3	MEM + NN	1.5 ± 1.1	1.4σ (1.1σ)	[122]
	13	20.3	MEM + NN	1.5 ± 1.1	1.4σ (1.1σ)	[122]
	13	36.1	MEM + BDT	$0.84^{+0.64}_{-0.61}$	1.4σ (1.6σ)	[123]
	13	139	MEM + BDT + NN	$0.35^{+0.36}_{-0.34}$	1.0σ (2.7σ)	[124]
CMS	7 (8)	5.0 (19.3)	BDT	2.8 ± 1.0	-	[125]
	8	19.5	MEM	$1.2^{+1.6}_{-1.5}$	-	[126]
	13	35.9	MEM + BDT + NN	0.72 ± 0.45	1.6σ (2.2σ)	[121]
	13	41.5	MEM + BDT + NN	$1.49^{+0.44}_{-0.40}$	3.7σ (2.6σ)	[127]
	13	77.4	MEM + BDT + NN	$1.15^{+0.32}_{-0.29}$	3.9σ (3.5σ)	[127]
	13	138	MEM + BDT + NN		this analysis	

Table 1.2: Summary of results for $t\bar{t}H, H \rightarrow b\bar{b}$ analyses in the leptonic channels: measured signal strengths, observed and expected significances are listed for each analyses at different center-of-mass energies $\sqrt{s}$ and luminosities. The different strategies to build observables for the final fits comprise the Matrix Element Method (MEM) (details in Appendix G), a machine learning algorithm in the form of Boosted Decision Trees (BDT) and one based on artificial neural network (NN).

increase the sensitivity by approximately 10%. In this thesis, not only is the measurement of the $t\bar{t}H$ and tH signal-strength modifiers performed, but additional interpretations, such as the $t\bar{t}H$ STXS measurement, are also conducted. This STXS framework (see Chapter 7) allows for the probing of individual parts of the total phase space due to its use of differential information. The experimental sensitivity to STXS contributions in the $t\bar{t}H, H \rightarrow b\bar{b}$ channel is gauged for the first time at the CMS experiment.

The analysis presented here improved the methods developed for the previous measurements at different levels. For example, the b-tagging method (see Section 3.2.5) has been improved, mainly affecting the baseline selection. This improvement has led to an enhanced event numbers while maintaining sensitivity to the signal process and stabilizing uncertainty estimation. Moreover, a major improvement has been made in the modeling of the critical $t\bar{t} + b\bar{b}$ background. In earlier analyses, the $t\bar{t} + b\bar{b}$ description was based on an inclusive NLO $t\bar{t}$ simulation in the 5FS, where only the first additional b jet is part of the matrix element calculation, and all the other jets originate from the parton shower. In the current analysis, instead, a dedicated NLO $t\bar{t} + b\bar{b}$ sample in the 4FS is used, with the additional b jets generated directly at the matrix element level (see Section 5.5 for further details). This results in a higher $t\bar{t} + b\bar{b}$ contribution in the signal-enriched categories, with a loss of sensitivity of approximately 10%. To fully exploit the improvements from the $t\bar{t} + b\bar{b}$ simulation, the a priori uncertainty on the additional $t\bar{t} + \text{heavy flavor}$ cross sections is removed and the normalizations are inferred from the data.

Additionally the classifier is refined and based on artificial neural networks, with an increased number of events for the training from the simulated samples for all three years simultaneously. The improved selection efficiency of the background processes translates into a better constrained background uncertainties. Furthermore, to increase the sensitivity, the current analysis exploits a dedicated likelihood ratio observable, as introduced in Section 5.7 and more detailed in Appendix I.

The results of the CMS $t\bar{t}H, H \rightarrow b\bar{b}$ analysis in the semi-leptonic channel with 35.9 fb^{-1} were projected to future scenarios as part of the formulation of the European Strategy for Particle Physics [128]. To note, given that the reference point for the projection is the analysis with the data set recorded in 2016, none of the above-mentioned improvements in the analysis technique since then are covered by the study. In any case, two scenarios were formulated by the experimental and theoretical communities to estimate the physics potential of the CMS detector:

- **Run 2 systematic uncertainties (S1)** In a conservative approach, both the CMS detector performance and the accuracy of the theoretical predictions are assumed constant.
- **YR systematic uncertainties (S2)** The more accurate techniques are reflected in considering half the values of the theoretical uncertainties; the experimental ones are scaled as a function of the square root of the luminosity till reaching a minimum value, which reflects the expected performances from the upgraded detector [129].

For both cases, the assumption is to have sufficiently large samples of simulated events.

The expected sensitivity is studied in terms of expected significance, and the corresponding uncertainty on the $t\bar{t}H, H \rightarrow b\bar{b}$ signal strength modifier $\mu_{t\bar{t}H}$, shown in Figure 1.18 as a function of the integrated luminosity [128]. The observation of the $t\bar{t}H, H \rightarrow b\bar{b}$ process is predicted in both scenarios with $\approx 200 \text{ fb}^{-1}$, opening the possibilities to more sophisticated measurements in future analyses like probing the top-Higgs Yukawa coupling.

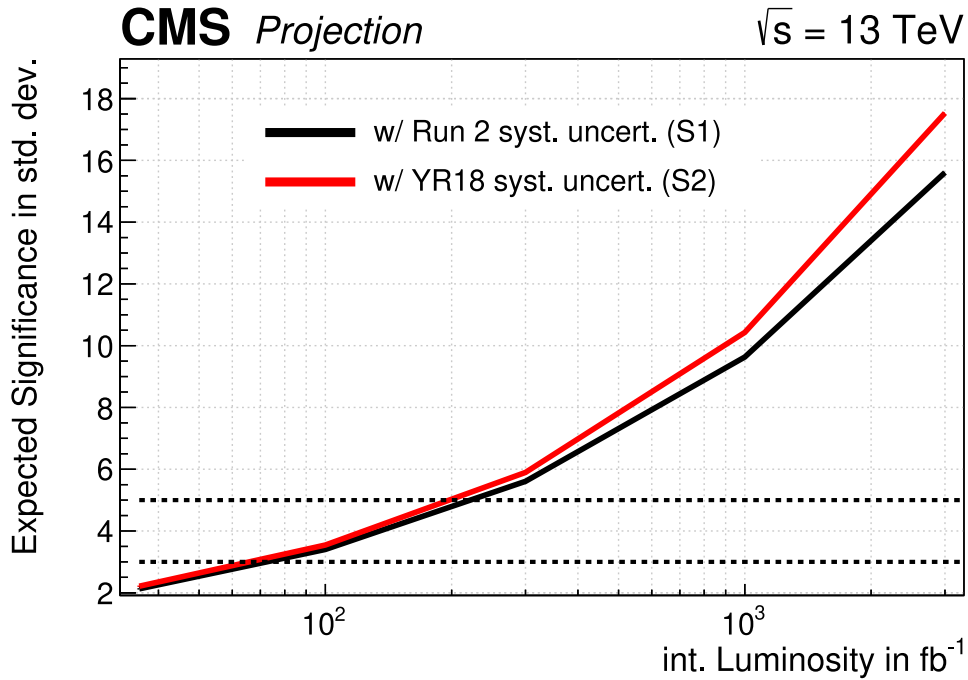


Figure 1.18: Comparison of the expected significance of the $t\bar{t}H, H \rightarrow b\bar{b}$ measurement as a function of the integrated luminosity in the two scenarios considered in the projection study. The dashed lines indicate the thresholds for evidence (3σ) and observation (5σ) [128].

THE EXPERIMENTAL SETUP

2.1	The Large Hadron Collider	35
2.1.1	The LHC complex	36
2.1.2	The LHC milestones	37
2.2	The Compact Muon Solenoid detector	38
2.2.1	The inner tracking system	41
2.2.2	The calorimeter system	41
2.2.3	The superconducting magnet	44
2.2.4	The muon system	44
2.2.5	The trigger system	45
2.2.6	The data processing and the computing system	46
2.2.7	The known detector issues	47

This chapter serves as an introduction to the experimental facility where the data analyzed in this thesis have been collected. The Large Hadron Collider [130] is the world's largest and most powerful particle accelerator, capable of producing extremely high-energy proton-proton and heavy-ion collisions. The Compact Muon Solenoid [131] detector is one of the two general-purpose detectors located on the LHC ring and is designed to detect a wide range of particles produced in these collisions. In this chapter, an overview of the key components of the CMS detector and the LHC accelerator system that are relevant to the analysis presented will be provided. This includes the subdetector components such as the tracker, calorimeter, and muon detector, as well as the trigger system which is responsible for selecting the most interesting collision events for further analysis. Additionally, data acquisition and processing techniques used to analyze the vast amount of data produced by the CMS detector and the LHC accelerator will be described.

2.1 The Large Hadron Collider

The Large Hadron Collider is a circular accelerator and collider for protons and heavy ions operated by the European Organization for Nuclear Research (CERN). It is situated in the 26.7 km circumference tunnel previously used for the Large Electron-Positron Collider (LEP) [132, 133], at a depth ranging from 45 m to 170 m underground. A schematic view of the LHC accelerator complex is shown in Figure 2.1. It was conceived to investigate the nature of the spontaneous symmetry breaking through the search of the Higgs boson, which was observed by the ATLAS and CMS Collaborations in 2012 [8, 9]. Ultimately, the goal is to scan the accessible phase space in search of new physics phenomena beyond the SM, aiming at favoring or ruling out the postulated scenarios, and thereby addressing some of the most outstanding challenges in the field. The LHC is designed to collide bunches of 10^{11} protons (or 10^8 lead ions) at center-of-mass energies $\sqrt{s}$ of up to 14 TeV (1.15 TeV), resulting in the world's most powerful accelerator. Two eras of physics operations have already been

conducted. During its first running period (Run 1) it operated at center-of-mass energies for p-p collisions of 7 TeV (2010-2011) and 8 TeV (2012). The second running period (Run 2) from 2015 to 2018 corresponded to p-p collisions at $\sqrt{s} = 13$ TeV. The third data-taking era (Run 3) started in July 2022 at $\sqrt{s} = 13.6$ TeV and is expected to last until 2025. After that, the LHC and the accelerator complex will undergo a profound upgrade towards the High Luminosity LHC (HL-LHC), scheduled to start in 2029.

The work described in this thesis makes use of data from p-p collisions collected by the CMS experiment during 2016-2018.

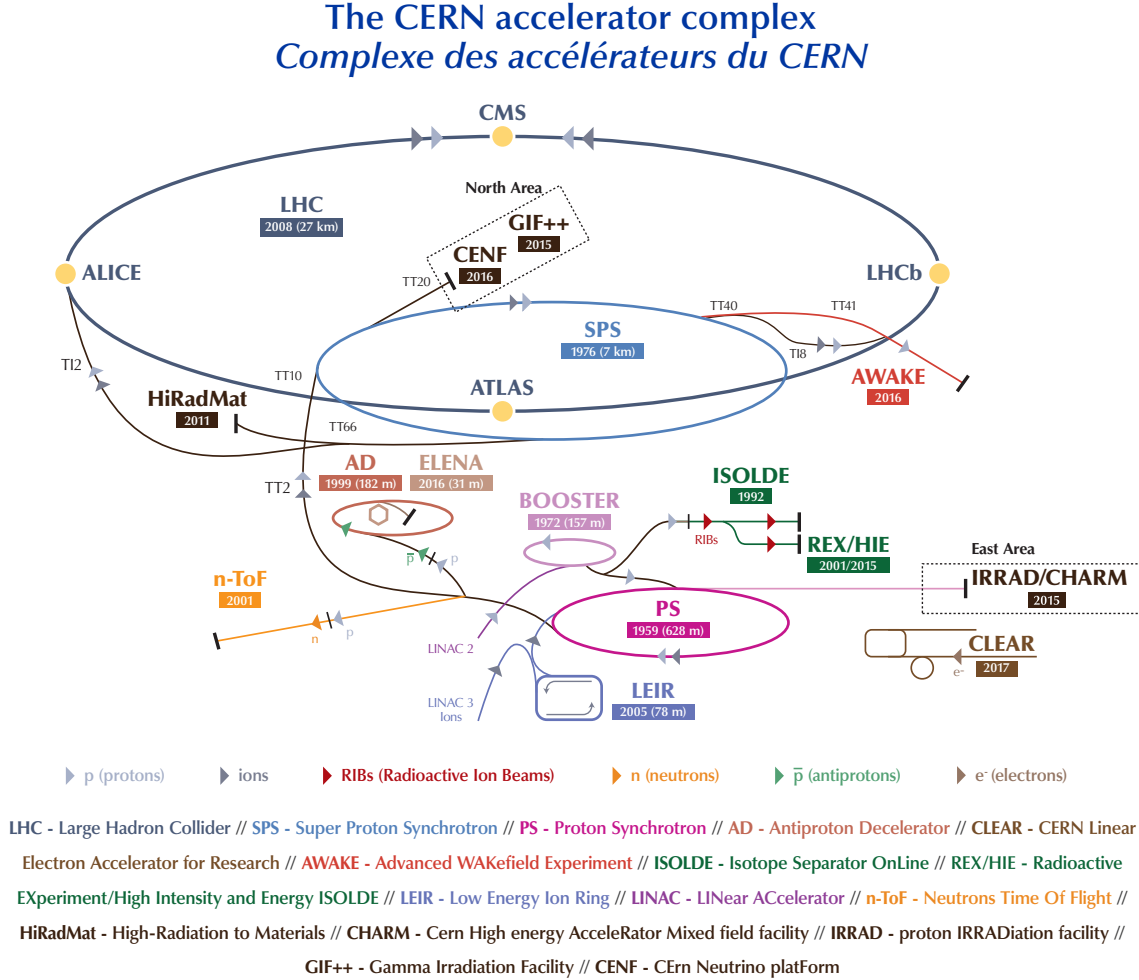


Figure 2.1: A schematic representation of the CERN accelerator facility, including the pre-accelerators and the four main experiments marked with yellow dots [134].

2.1.1 The LHC complex

The LHC is the final stage of the CERN accelerator complex system. Before entering the LHC machine, the protons are collected by ionization of the hydrogen gas source and pre-accelerated by a chain of machines used to increase beam energy gradually. The beam is initially accelerated to 50 MeV in the Linear Accelerator 2 (LINAC), then to 1.4 GeV in the Proton Synchrotron Booster (PBS). From there, the energy reaches 26 GeV in the Proton Synchrotron (PS), which splits the beams into 25 ns time-bunches. Next, the bunches are sent to the Super Proton Synchrotron SPS, where the protons are accelerated to 450 GeV

and injected into the two LHC rings with opposite rotation directions. The bending dipole and focusing quadrupole magnets are located in each of the eight arcs, while the straight sections host the interaction points (IPs) with detectors. Next, the beams are accelerated to collision energy by superconductive radio frequency (RF) cavities tuned at 400 MHz. A strong magnetic field up to 8.33 T guides the beams around the accelerator ring, which is maintained by 1232 superconducting niobium-titanium (NbTi) dipole magnets and cooled to 1.9 K.

There are four IPs in the LHC tunnel, where experiments with different characteristics and purposes are located. The ATLAS (A Toroidal LHC ApparatuS) [135] and CMS (Compact Muon Solenoid) [131] are general-purpose detectors meant to study the nature of electroweak symmetry breaking and possibly new physics signatures up to the TeV scale. They are located diametrically opposite and at the points of highest luminosity. ALICE (A Large Ion Collider Experiment) [136] is meant to examine the quark-gluon plasma produced by strongly interacting materials at extreme energy densities in heavy-ion collisions (Pb-Pb nuclei). Finally, LHCb (Large Hadron Collider beauty) [137] studies rare B hadron decays and measures CP violation to address the Universe's Matter-Antimatter imbalance.

The center-of-mass energy, the instantaneous and total luminosity, and the number of simultaneous collisions within the same bunch crossing, known as pileup, are all essential attributes of a collider. The center of mass energy of a collider with identical beam energies is defined by the beam energy E_{beam} as:

$$\sqrt{s} = 2 \cdot E_{\text{beam}}.$$

The instantaneous luminosity $\mathcal{L}$ of the machine substantially defines its performance. The production rate of events of a physical process generated in the collisions is given by $\nu^{\text{LHC}} = \sigma_{\text{process}}^{\text{LHC}} \cdot \mathcal{L}$, where $\sigma_{\text{process}}^{\text{LHC}}$ is the production run of the process under study. The general procedure to obtain the absolute luminosity is described in the dedicated Chapter 4.

2.1.2 The LHC milestones

The LHC began operating at the end of 2009, following a year-long technical delay caused by a faulty electrical connection between two magnets. The LHC established a new record in March 2010 by colliding proton beams at a $\sqrt{s} = 7$ TeV; the overall integrated luminosity was 44 pb^{-1} in 2010 and reached 6 fb^{-1} in 2011. A 4 TeV per beam ($\sqrt{s} = 8$ TeV) energy was supplied in 2012 [138]. After the end of the data-taking period, called Run 1, the total integrated luminosity delivered to LHC was 23.3 fb^{-1} , of which CMS recorded 21.79 fb^{-1} . During this period, the CMS total integrated luminosity was $7.67 \cdot 10^{33} \text{ cm}^{-2}\text{s}^{-1}$, whereas the LHC designed peak luminosity was $10^{34} \text{ cm}^{-2}\text{s}^{-1}$. In February 2013, the first long shutdown (LS) started to prepare the detector for increased energy and luminosity. Data collection began again in June 2015, and a month after protons collided with record-breaking energy at that time of 13 TeV. At the end of June 2016, the LHC reaches its design luminosity of $10^{34} \text{ cm}^{-2}\text{s}^{-1}$, the value that corresponds to the luminosity in the CMS and ATLAS IPs. Data taking continued till December 2018, marking the end of Run 2 and the start of the second LS.

The analysis described in this thesis is based on the LHC Run 2 p-p collision data collected by CMS between the years 2016 and 2018: the cumulative luminosity as a function of time delivered to the CMS collision point is reported in Figure 2.2 (a). During Run 2, LHC delivered a total integrated luminosity of 163 fb^{-1} , of which the CMS experiment recorded

150 fb^{-1} . The total integrated luminosity analyzed in this thesis is 138 fb^{-1} , corresponding to the collision with hard scattering processes.

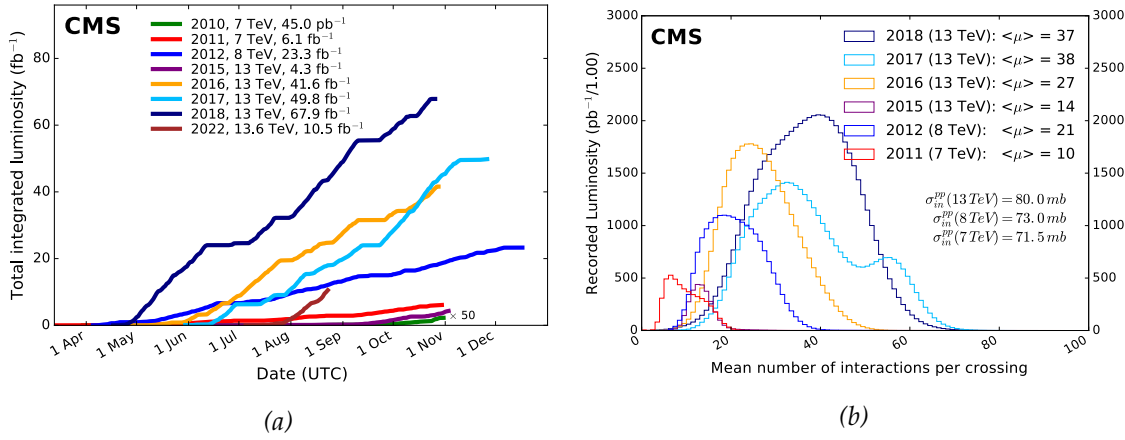


Figure 2.2: (a) Evolution of cumulative luminosity delivered to CMS versus day for p-p collisions during stable beams. The center-of-mass energy and the total luminosity are reported for each year of data-taking [139]. (b) Distribution of the mean number of inelastic interactions per bunch crossing, pileup, for each year of p-p collision. The average pileup values $\langle \mu \rangle$ are reported [139].

Physics studies usually benefit from the enhanced rates of uncommon events at high luminosity. The disadvantage arises from the overlap of several p-p collisions occurring during a single proton-bunch crossing, the pileup effect. The multiple interactions come from the same bunch crossing and different bunch crossing, out-of-time-pileup. However, the likelihood that several interactions result in an interesting process is usually considered negligible. Consequently, only the most energetic collision in each event is selected and studied, named the primary hard collision, while the remaining are called pileup (PU) interactions. Unavoidable PU events hinder data collection and reconstruction by creating additional overlapping particles across the detector, degrading measurement accuracy. Figure 2.2 (b) illustrates the pileup patterns of data obtained by the CMS detector over time.

The overall LHC operation schedule as a proton collider is summarized in Figure 2.3. From the end of 2018 until the end of 2021, the LHC was in its second long shutdown (LS2). At the time of writing, the Run 3 data-taking period has started with the 2022 Run. The beam energy increased from 6.5 to 6.8 TeV compared to the LHC parameters of 2018. Run 3 is expected to last four years and will collect an integrated luminosity of more than 150 fb^{-1} . A new LS should begin by the end of 2025, leading to a high-luminosity phase for the LHC known as the HL-LHC. Several upgrades to the accelerator complex are planned during this phase, allowing for an increased instantaneous luminosity of at least five times the nominal value. Several detector upgrades are also planned to account for the increased number of pileup interactions and collision rate. As a result, the HL-LHC is expected to collect between 3000 and 4000 fb^{-1} by 2040, increasing the total recorded luminosity by at least ten times.

2.2 The Compact Muon Solenoid detector

The Compact Muon Solenoid is one of the two multi-purpose particle detectors [141]. Its design was motivated by the desire to reconstruct each p-p collision event in its entirety and to achieve an excellent resolution of reconstructed leptons and photons, critical ingredients of the Higgs boson search in the golden channels $H \rightarrow \gamma\gamma$ and $H \rightarrow 4\ell$. The stated aims include verifying the SM consistency at the TeV scale, investigating the SM Higgs boson properties [8, 9], and searching for non-SM signals such as supersymmetry.

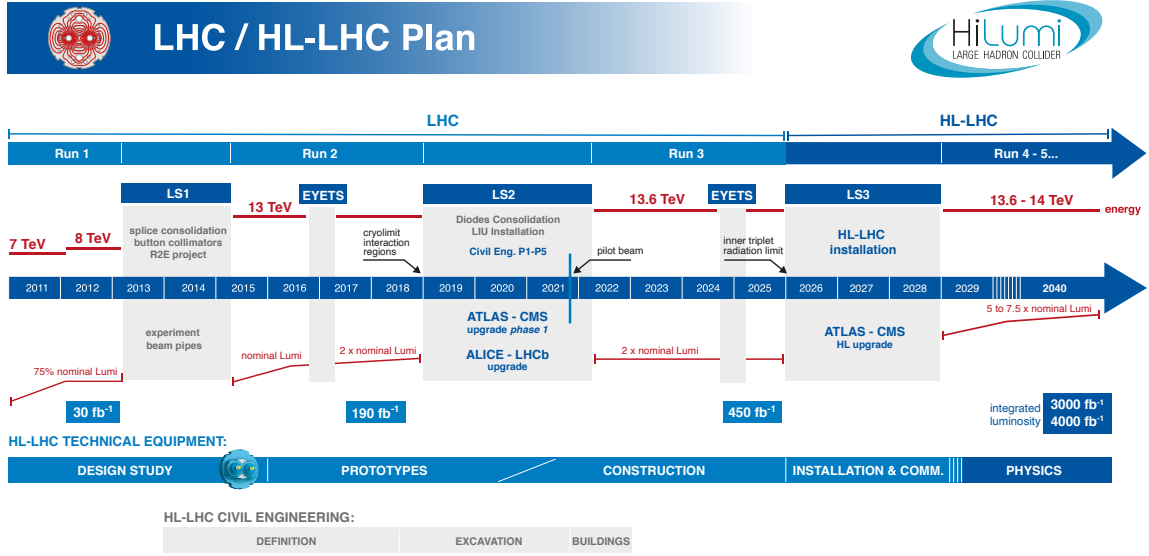


Figure 2.3: The present schedule for the LHC [140].

Geometry of the detector

In CMS, two coordinate systems are used. They both originate at the nominal interaction point of the beams, located in the center of the detector. The first system is the Cartesian coordinates with left-handed orientation: the x -axis is horizontal and points toward the center of the LHC, the y -axis points vertically upward, and the z -axis points along the counter-clockwise beam direction. Thanks to the cylindrical shape, the second coordinate set (r, ϕ, η) can be naturally defined and expressed as:

$$r = \sqrt{x^2 + y^2} \quad \phi = \tan^{-1} \left(\frac{y}{x} \right) \quad \theta = \tan^{-1} \frac{r}{z} \quad \eta = -\ln \left(\tan \frac{\theta}{2} \right), \quad (2.1)$$

where r is the distance from the beam axis in the transverse plane, ϕ is the azimuthal angle measured in the $(x - y)$ plane with values in $[-\pi, \pi]$, θ is the polar angle measured from the z axis in the range $[0, \pi]$ and η is the pseudo-rapidity related to the angle θ at which a secondary particle emerges from a collision. The pseudo-rapidity is the relativistic limit of the rapidity of a particle: $y = \frac{1}{2} \ln \frac{E+p_z}{E-p_z}$, which depends on the particle energy E and longitudinal momentum p_z . η is also used to identify different detector regions according to their position with respect to the beamline and the nominal beam collision point.

Other useful variables used in high energy physics are the following:

-

$$\Delta R = \sqrt{\Delta \eta^2 + \Delta \phi^2}.$$

Since both $\Delta \phi$ and $\Delta \eta$ are invariant under a Lorentz boost along the z -axis, it is helpful to introduce a separation in the $(\eta - \phi)$ space that defines a z -boost invariant cone around a single particle or detector position, to study nearby detector activity.

-

$$p_T = |\vec{p} \cdot \sin \theta| = \sqrt{p_x^2 + p_y^2 + p_z^2} \sin \theta = \sqrt{p_x^2 + p_y^2}$$

is the transverse momentum, where p_x , p_y and p_z are the projections of the momentum of a particle along respectively the x , y and z axis.

Another important element is the beam spot, a luminous region near the coordinate origin where two proton beams collide. On the x -axis, this region is around $16\text{ }\mu\text{m}$ wide and $100\text{ }\mu\text{m}$ long, whereas on the z -axis, it is about 15 cm long. As discussed in Chapter 3, the beam point is the origin of all p-p interaction products.

Detector components

The detector is built around the most important feature of the experiment layout: the state-of-the-art superconducting solenoid, composed of a cylindrical coil of superconducting wire capable of generating a uniform magnetic field of 3.8 T . The solenoid is 13 m long and 6 m in diameter; it has the tracking system and the electromagnetic and hadron calorimeters within and outside the iron return yoke of the magnet hosts the muon chambers. The overall CMS detector is composed of a central part with $|\eta| \leq 1.2$, called barrel, made of sub-detectors positioned at increasing values of the cylinder radius and enclosed by two disks with $|\eta| > 1.2$, called endcaps, where sub-detectors are layered along the beam axis. It is 21.6 m long, 14.6 m in diameter, and its total weight is around $1.4 \cdot 10^7\text{ kg}$. Figure 2.4 shows a schematic overview of the CMS detector.

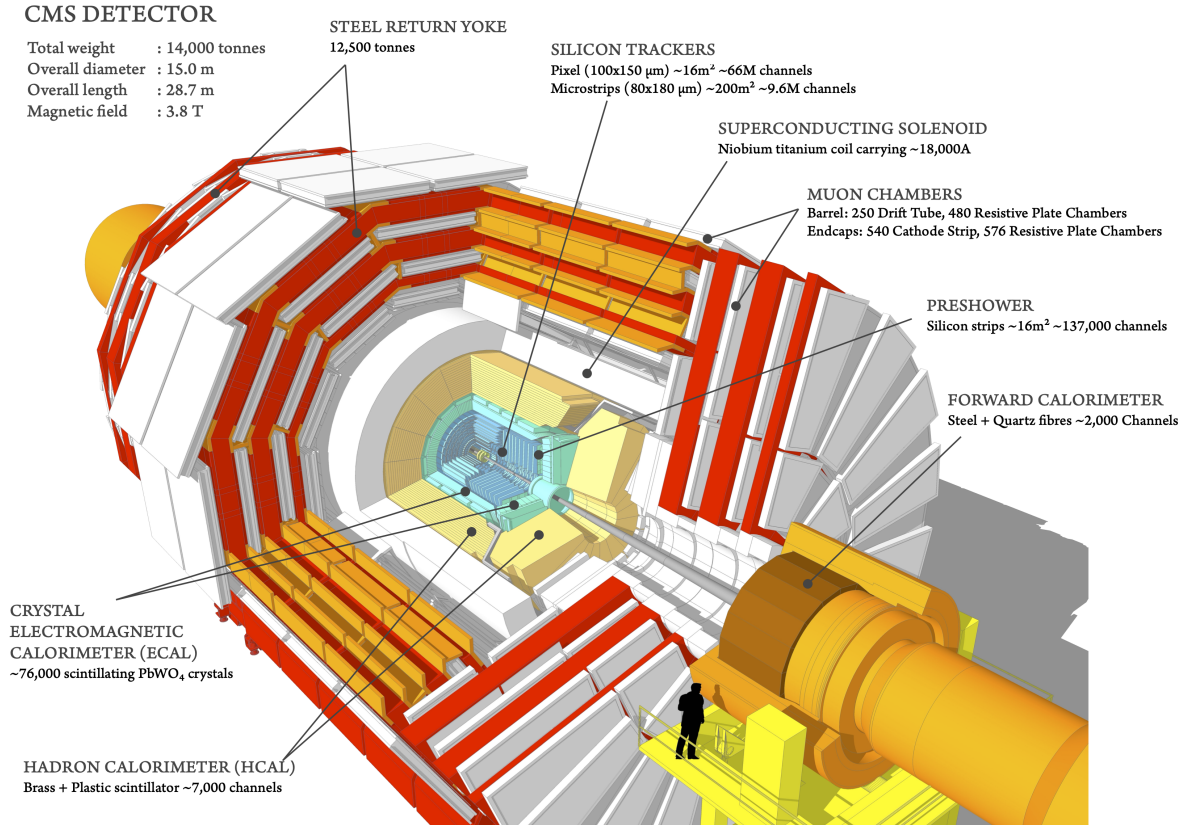


Figure 2.4: A schematic representation of the CMS detector sectional view and its sub-detectors [142].

There are two types of sub-detector: tracking devices that disclose the tracks of electrically charged particles by ionizing materials and calorimeters that measure energy by arresting particles and detecting the energy released. Two tracking devices are the pixel- and microstrip-based vertex detectors at the interaction point and the muon chambers in the outer layers. Calorimeters are categorized as electromagnetic, which uses scintillating crystals to analyze electrons and photons, or hadronic for jet energy measurement.

The following sub-detectors can be identified in CMS starting from the collision region:

- an **inner tracking system** providing excellent vertex reconstruction and particle momenta resolution; it is entirely composed of silicon sensors with pixels and strips.
- an **electromagnetic calorimeter (ECAL)** providing a precise measurement of electron and photon energies due to its granularity and extensive coverage.
- a **hadron calorimeter (HCAL)** providing a very good reconstruction of hadrons energies and the missing transverse momentum thanks to its sampling structure.
- a **solenoidal superconducting magnet**, large enough to house the inner tracking and calorimeter systems; it generates a nearly uniform magnetic field of 3.8 T along z , allowing the charged particle momenta in the inner tracking system to be measured.
- a **muonic system** between the iron plates of the return yoke; the comprehensive coverage and a redundant system of four muon chambers allow for accurate muon identification and momentum resolution even at very high energies.

To minimize performance degradation over time, detectors near the interaction point and front-end electronics are intended to be radiation-hard.

The operation conditions are characterized by high pileup, as illustrated in 2.2 (b), which implies a huge number of particles originating from the interaction region. Therefore, subsystems are made up of high granularity detectors to provide optimum physical performance and minimize pileup effects. These features and other crucial CMS subsystems are described thoroughly in the following subsections.

2.2.1 The inner tracking system

The inner tracking system is the innermost part of the CMS detector, located nearest to the beam spot in order to make a precise measurement of the momentum and the impact parameter of the tracks. It is further discussed in Refs. [143–145].

According to design characteristics, p-p interactions yield roughly 1000 particles for each bunch crossing. The tracking system consists of high granularity silicon detectors to achieve good accuracy in such a high flux environment. In particular, the ionization process is used to measure charged particle trajectories and momenta to discriminate among the vertices corresponding to the several interactions that may occur in the same bunch crossing.

The tracking system is composed by two parts, the pixel detector and the silicon strip detector. Both are based on semiconductor technology but satisfy different granularity and size criteria. The pixel detector is the innermost detector, situated a few centimeters away from the interaction point. Since a higher occupancy is expected in that region and since a high spacial resolution is needed closer to the interaction point for a precise measurement of the impact parameter, the pixel detector is very granular, with a typical resolution of 1.5% in p_T and 20–75 μm in the transverse impact parameter for non-isolated particles of $1 < p_T < 10 \text{ GeV}$ and up to $|\eta| < 3$. The pixel detector is divided in 3 (4) layers in the barrel and 2 (3) disks in the endcap before 2017 (from 2017). Figure 2.5 depicts the upgraded layout [146]. The silicon strip system is the second layer of subdetector located 20 to 116 cm away from the interaction point. Silicon strips are located along the z direction in the barrel, and along the radial component in the endcap.

2.2.2 The calorimeter system

The CMS experiment comprises four calorimeters to efficiently detect particle energies across a wide η range. ECAL (Electromagnetic CALorimeter) [147] and HCAL (Hadron CALorimeter) [148] are the two detectors utilized to achieve optimal measurement of

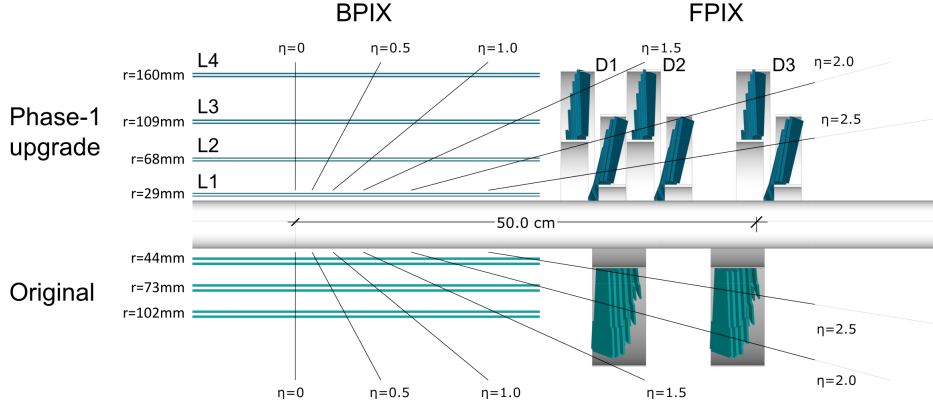


Figure 2.5: Comparison of layer and disk positions between the original and the Phase-1 pixel trackers [146].

electrons, photons, and hadrons across a wide pseudo-rapidity range ($|\eta| \leq 5.2$). CASTOR (Centrauro And Strange Object Research) and ZDC (Zero Degree Calorimeter) are two forward calorimeters designed to supplement CMS measurements in the extremely forward area ($|\eta| > 5.2$). This thesis relies solely on data collected from the first two detectors. More detailed information can be found in Refs. [149–157]

The two calorimeter layers surround the tracker and provide complementary information on the particle energy. A particle entering a calorimeter creates a shower of secondary particles used to compute the originating particle energy. In this destructive approach, the shower grows until the energies of the particles are low enough to be caught and absorbed by the detector material. Because of the longer characteristic length of hadronic showers and the presence of an electromagnetic component, electromagnetic calorimeters are typically placed before hadron calorimeters.

The ECAL was developed to attain high photon and electron energy and angular resolutions; the primary channel optimized was $H \rightarrow \gamma\gamma$. The HCAL was designed to detect and measure the energies of highly interacting particles and particle jets made primarily of hadrons. Additionally, accurate energy measurement is required to calculate the missing energy flow created by neutrinos.

Electromagnetic calorimeter and preshower

The CMS electromagnetic calorimeter is a hermetic, homogeneous scintillating crystal calorimeter with a photo-detector array positioned behind a pre-shower detector (ES). As illustrated in Figure 2.6, it is divided into barrel (EB) and endcap (EE) components. Because most of the energy from electrons or photons is deposited within the calorimeter chamber, it provides excellent energy resolution.

The barrel part, extending from $r = 1.3$ m to $r = 1.8$ m, covers the region $0 < |\eta| < 1.48$. A 20 cm thick pre-shower detector is placed in front of the endcap part covering the region $1.65 < |\eta| < 2.6$. The two ECAL endcaps cover the region $1.48 < |\eta| < 3.0$ and are located extending from $|z| = 3$ m to $|z| = 3.8$ m from the nominal interaction point. Each endcap has two D-shaped sections named Dees. With high density, short radiation length, fast response, and radiation hardness, the lead tungsten (PbWO_4) has been chosen as the active material for the scintillating crystals; thanks to these properties, the electromagnetic calorimeter has a compact shape and a fine granularity.

Combining the energy measurement in the ECAL with the momentum measurement in the tracker yields an estimate of electron momentum. The momentum resolution for electrons with $p_T \approx 45$ GeV decaying from $Z \rightarrow ee$ ranges from 1.6 to 5% [159]. It is generally better in

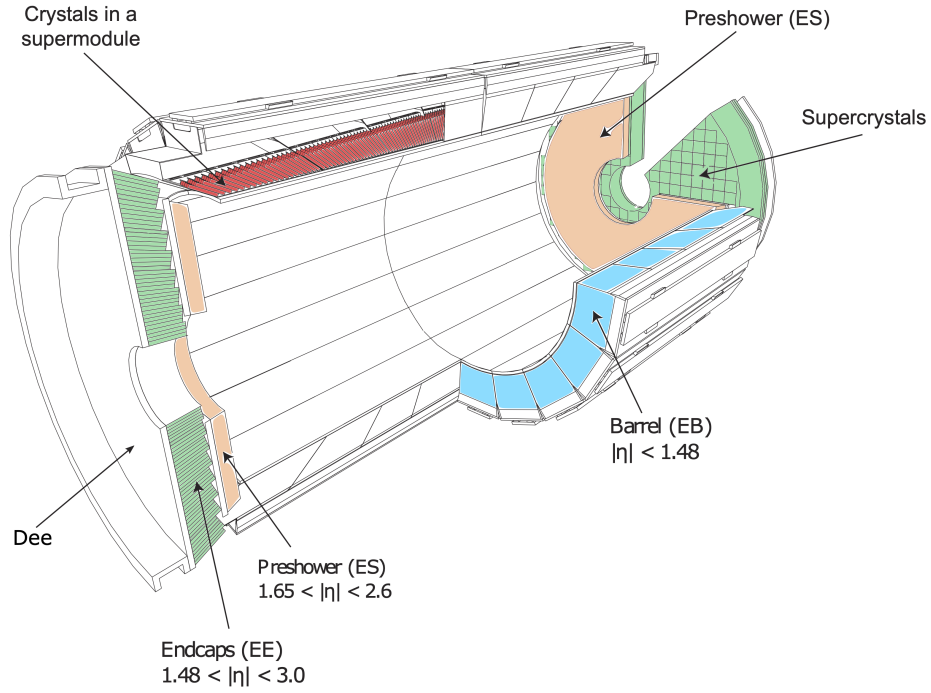


Figure 2.6: Longitudinal 3D view of the CMS ECAL. Adapted from [158].

the barrel region than at the endcaps, and it is also affected by the bremsstrahlung energy emitted by the electron as it passes through the material in front of the ECAL [152, 159].

Hadron calorimeter

The hadron calorimeter [148] is located surrounding the ECAL and is designed to maximize the amount of energy deposited by strongly interacting particles before reaching the magnet coil. The HCAL is a sampling calorimeter composed of four sub-detectors separated into alternating layers of absorbers and scintillators: the Hadron Barrel (HB) covers the central pseudo-rapidity region up to $\eta = 1.3$; the Hadron Outer (HO) in the same region as the HB; the Hadron Endcap (HE) covers the region between $1.3 < |\eta| < 3.0$ on each side and the Hadron Forward (HF) extends the coverage up to $|\eta| = 5.2$ in the forward region.

The longitudinal scheme of HCAL is shown in Figure 2.7. HB and HE are inside the solenoid magnet, demanding the employment of non-ferromagnetic materials: brass has been chosen as the absorbing material and plastic scintillating tiles are used as active material. HO is outside the magnet and completes the structure of the calorimeter in the barrel catching the tails of the hadron shower. Finally HF is placed outside the muon chambers and it covers the high-pseudo rapidity forward region. The absorbing plates are steel to withstand high particle fluxes, while the quartz-based fibers gather the Cherenkov radiation produced by showers in the absorbers. It is employed for the CMS luminosity measurements (see Chapter 4) since it gives good precision for energy measurements.

HCAL is meant to resolve energy well for hadronic jets (Section 3.2.5). This is due to the necessity to precisely quantify the missing transverse momentum (Section 3.2.6) per event and to minimize muon misidentification. When combining information from the entire detector, the jet energy resolution amounts typically to 15–20% at 30 GeV, 10% at 100 GeV, and 5% at 1 TeV [161].

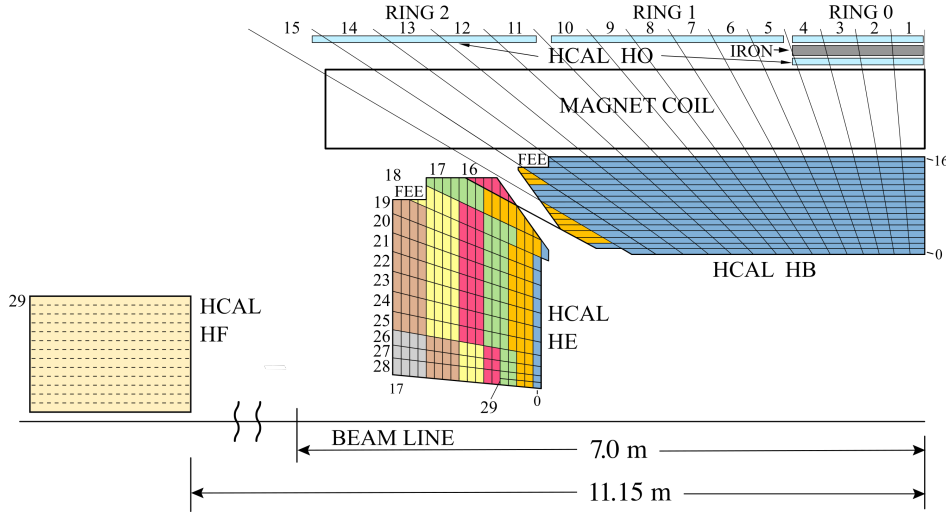


Figure 2.7: Longitudinal view of one quadrant of the CMS HCAL in the $(r-z)$ plane. The HB, HE, HO, and HF components of the HCAL detector are highlighted. The calorimeter towers are separated by black lines and indicated by numbers. Each color indicates the segmentation of scintillators layers available at the end of Run 2 [160].

2.2.3 The superconducting magnet

The CMS detector magnet [162] is one of its distinguishing features, the central device around which the entire experiment has been built. It bends the trajectory of charged particles from LHC collisions, allowing to measure their transverse momentum. The less a particle path is curved by the magnetic field, the more momentum it has. A stronger field bends paths more, and when combined with high-precision position measurements in the tracker and muon detectors, it is possible to determine the momentum of even high-energy particles accurately. The magnet of the CMS detector generates 3.8 T, allowing for a momentum resolution of better than 10% at 1 TeV. A cryogenic system maintains the magnet at an operating temperature of ≈ 4.5 K using liquid helium flows.

The magnet comprises a superconducting solenoid consisting of four layers of NbTi, surrounded by a massive iron yoke that closes the magnetic field loop and hosts the muon system. The returned magnetic field of 2 T in the iron yoke points in the opposite direction with respect to the field inside the coil. As a result, high momentum muons in the tracking and muon systems are bent in two opposed directions, boosting the resolution of their momentum measurement (see muon track example in Figure 3.1).

2.2.4 The muon system

The CMS design was optimized to reconstruct and identify muons and measure their momentum over a wide range. The ECAL and HCAL absorb the majority of particles produced in the interaction and their decays. The muon is therefore the most common particle (aside from neutrinos) that passes through the calorimetric system, with only minor radiative or ionization losses (less than 3%) in the energy range of 1 GeV to 1 TeV. As a result, the muon system is designed as the CMS outermost part [163–165]. As shown in Figure 2.8, the muon system has a cylindrical section along the barrel and two endcap planar regions positioned outside of the solenoid, and it covers the range $|\eta| \leq 2.4$. It detects muons by the use of three different gas-based detector technologies: drift tubes in the barrel region, cathode strip chambers in the endcap region, and resistive plate chambers in both sections.

The Drift tubes (DT) chambers in the barrel cover the region $|\eta| < 1.2$, organized into

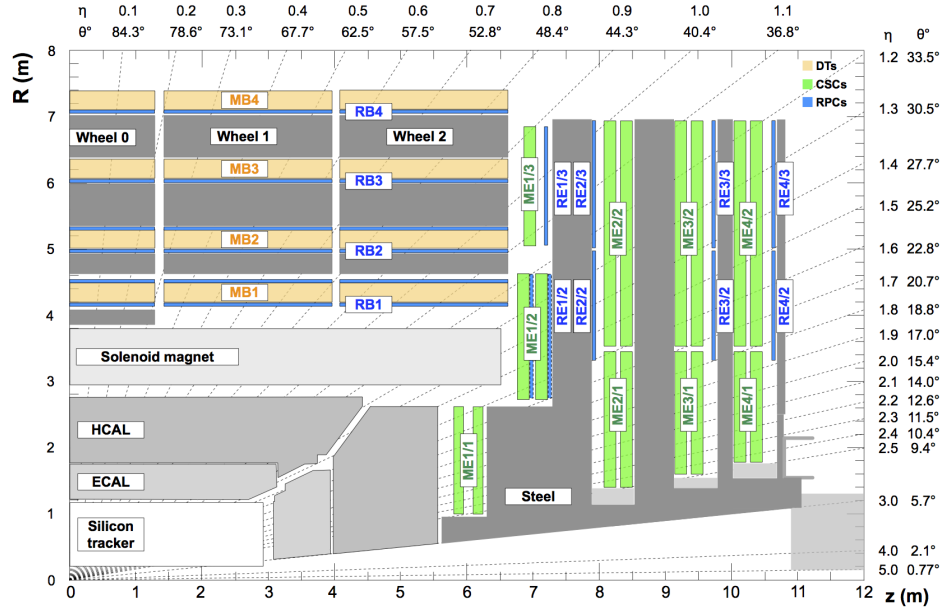


Figure 2.8: Quadrant of the CMS detector in the Run 2 configuration with the subdetectors of the muon system in color: DT in light orange, CSC in green, and RPC in blue [166].

four stations (MB1,..., MB4, where MB stand for Muon Barrel) with varying radii. Cathode Strip Chambers (CSC) have been installed in both endcaps allowing identification of muons in the region $0.9 < |\eta| < 2.4$. Each endcap section is composed of four stations (ME1, ..., ME4) arranged perpendicular to the beam and inter-layered between the magnet flux return yoke. The main features are fine segmentation, radiation resistance, and fast-time response. Each DT station (CSC chamber) has a resolution of $\approx 100 \mu\text{m}$ ($\approx 100 - 200 \mu\text{m}$, depending on the layer) in the position and $\approx 1 \text{ mrad}$ (10 mrad) in the direction. A crucial characteristic of the DT and CSC subsystems is their ability to trigger on the p_T of muons with good efficiency, high background rejection, and p_T resolution of about 10% and 25% in the barrel and endcap regions, respectively [165]. A highly effective standalone muon identification is also required for triggering purposes (see Section 2.2.5). In addition, Resistive Plate Chambers (RPC) are used to provide a complementary and independent trigger system with coverage in the region up to $|\eta| < 1.6$. The RPC are gaseous parallel-plate detectors made of double-gap modules operating in avalanche regime and with fast response and a high intrinsic temporal resolution of $\approx 2 \text{ ns}$.

The efficiency of muon reconstruction, identification, and isolation is determined to be better than 95%, with a misidentification rate of less than 1%. The muon resolution in a combined reconstruction that includes tracker information is on the order of 3% for muons with $p_T < 100 \text{ GeV}$ and around 10% for muons with p_T in the TeV range [166].

After the end of Run 2, gas electron multiplier (GEM) detectors [167] have been installed in the endcaps to accomplish the same physics goals of the RPCs, to extend the coverage in the very forward region ($1.6 < |\eta| < 2.4$). The first batch of chambers will be used during the ongoing Run 3 while others will follow for Phase-2 of the LHC.

2.2.5 The trigger system

The total p-p run at the nominal collision energy of $\sqrt{s} = 14 \text{ TeV}$ is expected to be roughly 100 mb [168]. This translates to roughly 20 simultaneous p-p collisions in the 25 ns bunch crossing interval at the design luminosity and a rate of approximately 10^9 events per second. Luminosity has already exceeded the designed level $\mathcal{L} = 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ and is expected to

increase further, as are other beam parameters [169]. As a result, the interaction rate is orders of magnitude higher than the maximum rate (~ 1000 Hz) that the data acquisition system can handle. Furthermore, the data size per event is 1 MB with zero suppression algorithms applied. Due to technical constraints, it is critical to store the information for all events; therefore, the experiments must select the interesting ones online. The production runs for the most interesting processes are many orders of magnitude smaller than the entire p-p collision cross section; consequently, the event selection criteria should be broad enough to account for unexpected new phenomena. The Trigger and Data Acquisition System (TriDAS) [170, 171] was developed to identify and record useful events in real time while limiting the amount of data collected for archiving and offline analysis.

To achieve the best performance, the CMS trigger reduces the rate in two steps: Level-1 (L1) triggers [172–174], designed to lower the rate from an input of 40 MHz up to 100 kHz and the High-Level Trigger (HLT) [171, 175], reducing the rate up to 10^3 events per second. The first stage is entirely hardware-based; it employs programmable custom-designed electronics to implement event selections using only calorimeters and muon system information and process the real data in less than 1 μ s. The HLT is a software-based system that decreases the event rate to about 1 kHz to match the capabilities of data storage and offline computing systems. Events accepted by the L1 trigger are read out using the complete detector information, reconstructed with algorithms similar to the offline reconstruction, and rejected as soon as there is enough information to make a choice. Consequentially, the HLT algorithms are run in order of increasing CPU complexity. The HLT data processing is organized around the concepts of Path and Menu. An HLT Path is a sequence of reconstruction algorithms and filters for one or more physics objects linked together by logical AND. An HLT Menu is the sum of logical OR of trigger paths that determine whether to reject or store an event.

In addition to the many physical requirements composing a trigger path, the CMS trigger system can discard events randomly. This feature is known as pre-scaling: when a trigger path has a lower threshold than expected, a pre-scale factor is given to the trigger, not to saturate the memory [172]. The HLT does not perform all pathways for each event to reduce execution time. Which HLT paths should be executed is determined by the L1 results; several L1 algorithms may seed one HLT path. The trigger system and its performances are further discussed in Refs. [176, 177].

2.2.6 The data processing and the computing system

After HLT, events accepted must be preserved and reprocessed to be used efficiently. Real data recorded in p-p collisions and simulated samples are stored in a structure known as the World LHC Computing Grid (WLCG) [178–180], a global computing infrastructure that stores, distributes, and analyzes data. The WLCG is divided into four Tiers, each offering a unique variety of services. The global data distribution has shown the importance of a standard infrastructure for data storage; the CMS experiment uses a number of data storage formats with variable volumes, reprocessing frequencies, and content.

For both data and simulated samples a format has been designed to provide event content for physics analysis. High-level physics objects are included in the Analysis Object Data (AOD) format, and by removing detector hit information event size is reduced to 400–500 kB/event. Many CMS studies can use high-level quantities like object kinematics and identification features, with little need for lower-level data. The benefit is more space savings and faster processing. The CMS Collaboration has devised two new formats that minimize event size by one (miniAOD [181]) and two (nanoAOD [182]) orders of magnitude when compared to the AOD format.

2.2.7 The known detector issues

In the course of the data-taking period, two issues with the recording of the data by the CMS detector occurred, that affect the analysis presented in this thesis.

During 2016 and 2017, a progressive timing shift of the ECAL was not correctly propagated to the L1 trigger inputs, resulting in a significant fraction of high- $|\eta|$ candidate objects being mistakenly associated with the previous bunch crossing. The issue is known as L1 prefiring: events with a high ECAL energy in the range of $2 < |\eta| < 3$ could veto themselves because of the L1 rules that forbid two consecutive bunch crossings to fire. Therefore, as discussed further in Section 6.1.1, the corresponding simulated samples are corrected, and a dedicated uncertainty is applied.

Furthermore, a portion of the detector was depleted entirely of HCAL energy deposits in 2018 due to a damaged HCAL sub-module: the so-called HEM problem. As a result, the number of reconstructed isolated electrons is significantly higher because HCAL information is used to discriminate between genuine electrons that deposit energy in the ECAL and jets that would also deposit energy in the HCAL. The effect on the analyses reported in this thesis is minor; nonetheless, this impact is considered with a dedicated uncertainty.

EVENT RECONSTRUCTION



3.1	Basic objects	49
3.1.1	Track and vertex reconstruction	50
3.1.2	Energy reconstruction in the calorimeter	51
3.2	Physics objects	51
3.2.1	Particle-flow algorithm	51
3.2.2	Muons	52
3.2.3	Electrons and photons	53
3.2.4	Hadrons and non-isolated photons	53
3.2.5	Jet reconstruction	54
3.2.6	Missing transverse momentum	55
3.2.7	Pileup mitigation	56
3.2.8	Heavy flavor jet identification	56

This chapter illustrates how detector information is employed to reconstruct high-level objects such as tracks, vertices and calorimeter clusters, and subsequently to build physics objects such as leptons and jets. The reconstruction of these objects is crucial for the identification and measurement of the various types of particles produced in LHC collisions. The CMS detector provides a wealth of information about the collision events, but this information needs to be processed and interpreted in order to extract the physics objects of interest. This chapter provides a detailed description of the various reconstruction algorithms and techniques used in CMS. The performance of these reconstruction algorithms is also discussed, with a focus on their accuracy and efficiency.

3.1 Basic objects

The raw data recorded by the detector consists of multiple silicon sensors and muon detector hits, and calorimeter clusters. In order to correctly measure and identify events of a given physics process like $t\bar{t}H$ production, it is essential to employ advanced algorithms to identify which combination of hits and clusters are caused by which particles. This is further challenged by the many simultaneous p-p collisions in an event and by electrical detector noise causing fake signals. Basic objects like tracks, particle trajectories, and calorimeter clusters are reconstructed in the first step. Then, the tracks allow for reconstructing the vertices of many p-p interactions and identifying the primary vertex. The algorithms utilized to reconstruct these basic objects are described briefly below.

An example of a transverse view of the CMS detector is shown in Figure 3.1, with the signatures of the detectable particles.

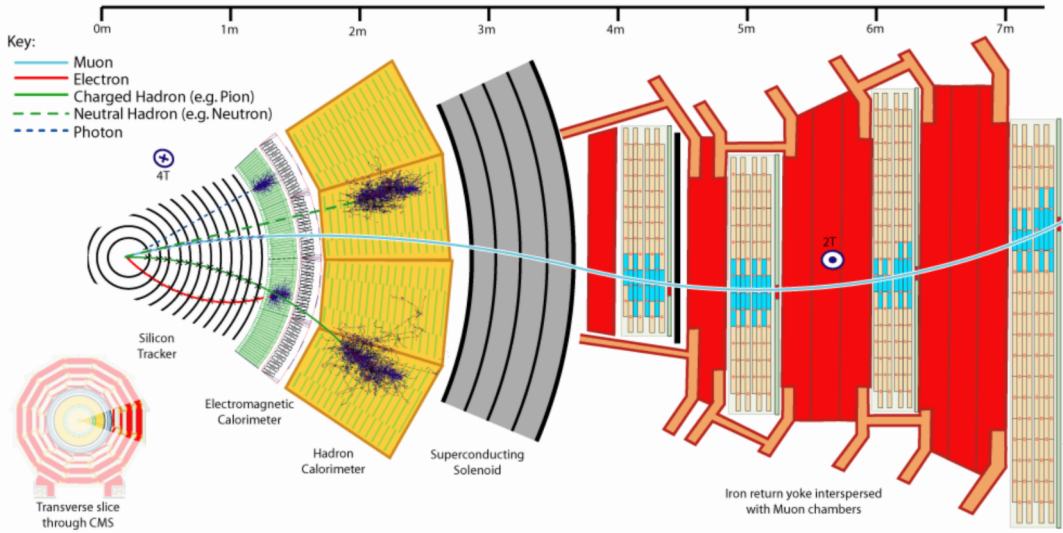


Figure 3.1: Transverse view of CMS detector in the barrel region, where typical signature of detectable particles are also highlighted [183].

3.1.1 Track and vertex reconstruction

The initial step in the procedure is the complex algorithm of track reconstruction that processes the inner tracking system readout (Section 2.2.1) to estimate the trajectory and momentum of charged particles. The magnetic field created by the superconducting magnet (see Section 2.2.3) is almost uniform in the tracker region, therefore a charged particle moves along a helical path described by five parameters: $d_0, z_0, \phi, \cot \theta$ and the transverse momentum p_T of the track defined at the impact point (x_0, y_0, z_0) . θ and ϕ are the polar and the azimuthal angles of the track momentum vector, respectively. $d_0 = -y_0 \cos \phi + x_0 \sin \phi$ is the impact parameter of the track in the transverse plane.

Pattern recognition algorithms [145] can recreate a particle's original trajectory based on charged deposits (hits), assigning hits to a track and evaluating the five helix parameters that best match the hits. The CMS track reconstruction algorithm is based on an iterative fit procedure based on the Kalman Filter (KF) [184, 185]; it starts with tracks that are easier to discover (prompt and high p_T) and progresses to more complicated ones (low p_T range or displaced vertices). The excellent knowledge of the exact position of the individual modules in the tracking system is required: this is achieved with the alignment procedure [186, 187]. Additional information are further described in Refs. [145, 188–190]

The tracking efficiency for the isolated muons (pions) is measured to be 99% (80-99%) [145], depending on the pseudorapidity range. The fake rate is between 5 and 15%. Excellent agreement is established between data and simulation, and the track p_T resolution is calculated to be between 1 and 10%, depending on the track p_T and η for muon tracks ranging from 1 to 100 GeV [145].

Numerous collisions occur when proton bunches collide at the LHC, yet only few of these induces hard scattering. A accurate reconstruction of all p-p interaction positions, known as primary vertices (PV), facilitates the rapid identification of the primary hard (high p_T) interaction products from all other pileup collisions. Furthermore, accurate identification of physical objects necessitates excellent primary and secondary vertex position measurements. Primary vertex reconstruction uses all reconstructed tracks with high-purity quality criteria [145]. First, tracks are selected to match the beam spot (within a distance of 24 cm

in z direction, and 2 cm in the $x - y$ plane) and then clustered based on a potential common vertex origin [191]. Afterwards, each cluster is fitted to retrieve the best estimate of the three-dimensional position of each vertex and their covariance matrix. The iterative fit process is repeated reducing the contribution of fake tracks until the convergence is reached.

The primary hard interaction vertex is defined as the one with the highest $\sum_{i=1}^{\#objects} p_T^2$, while the others are pileup vertices. The resolution of the primary vertex position spans between 10 and 100 μm , with an efficiency of more than 98% (99.8%) for vertices with 2 (or more) associated tracks [145].

3.1.2 Energy reconstruction in the calorimeter

The ECAL and HCAL systems collect the energy of electron and photon showers (see Section 2.2.2). However, electron bremsstrahlung and photon conversion prevent calorimeters from containing all energy. Therefore, jet clustering techniques are employed: to collect all the shower energy, nearby ECAL cells exceeding the noise level threshold are clustered around a seed (cell with higher energy). The clustering method improves the estimation of stable neutral particle energy and direction, reconstructs electron bremsstrahlung photons, and distinguishes neutral from charged particle energy deposits. Each subdetector (ECAL, HCAL, barrel, endcaps, and preshower) is independently clustered. Cluster energy determination is subject to a thorough calibration of the calorimeters' hadronic and electromagnetic energy responses to optimize identification efficiency and decrease misidentification [192].

3.2 Physics objects

The previous section introduced the basic objects; the next describes the algorithms used to connect tracks and calorimetric data to reconstruct electrons, muons, and jets.

3.2.1 Particle-flow algorithm

The Particle Flow (PF) [193–195], developed and used by the ALEPH experiment at LEP, has been applied for the first time at the hadron collider for CMS particle reconstruction and identification. The algorithm aims to optimally reconstruct and identify all stable particles linking the individual detector objects (charged-particle tracks, calorimeter clusters, and muon-system hits). Figure 3.2 shows the typical detector signatures for different particles. Fine-grained calorimeters and an advanced clustering algorithm are needed for accurate standalone reconstruction of each linked element.

PF blocks are formed by geometrically linking neighboring possible pairs of charged particle tracks, calorimeter clusters, and muon tracks in the $(\eta - \phi)$ plane. Each element can be assigned to multiple blocks, and the distance between two elements quantifies the quality of the link. Finally, all linked elements are combined to find the PF candidate. From the outermost hit in the tracker to the calorimeter system, a track is extrapolated and linked to all clusters, with an extrapolated track position falling within the cluster area. Tracks can be linked together also in case they originate from a common secondary vertex. Figure 3.3 depicts the concept behind the PF algorithm.

The particle-flow algorithm iteratively reconstructs and identifies PF particles from the blocks of linked PF elements. Blocks already assigned to PF particles are excluded from subsequent iterations to avoid double counting. The identification proceeds in the following order: muons are identified first, then electrons and photons, and finally charged or neutral hadrons [197].

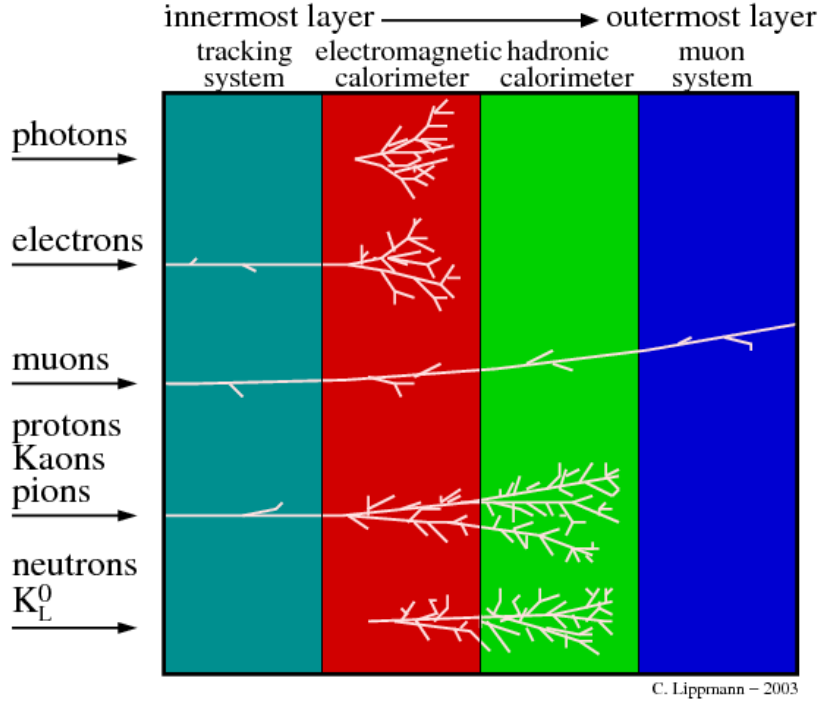


Figure 3.2: Examples of signatures in the detector for each particle type [196].

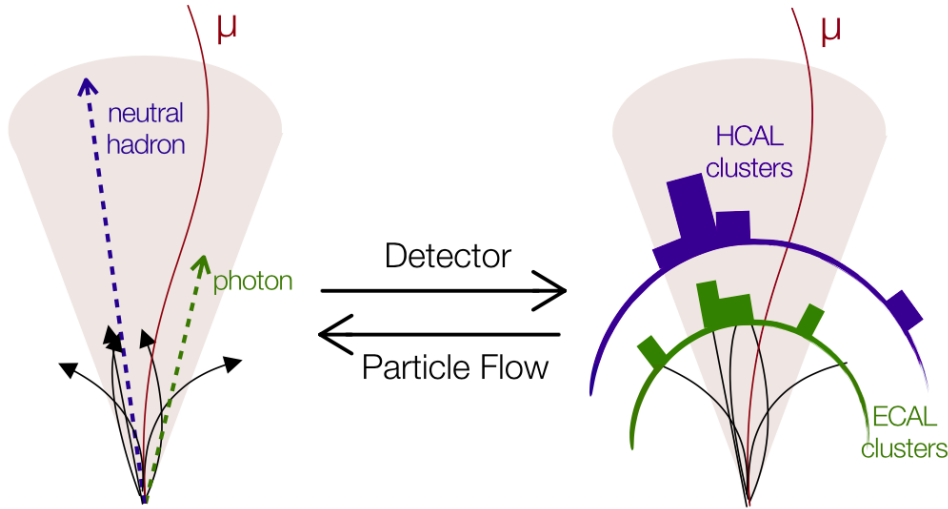


Figure 3.3: Schematic view of the PF algorithm principle [198].

3.2.2 Muons

Muon reconstruction is the first phase in the PF algorithm. Muons have a mean lifetime of $\approx 2.2 \mu\text{s}$, and only a small amount of energy is lost from interaction with tracker material over a wide range of p_T . This allows muons to pass through the entire detector, leaving tracks in both the inner tracker and the muon system [165]. Therefore, CMS can rely on excellent muon identification and reconstruction thanks to two independent systems.

The CMS muon collection consists of Standalone, Tracker, and Global muons. Standalone muons are reconstructed by the muon system tracks exclusively. CMS utilizes two off-line muon reconstruction approaches, the inside-out for the Tracker muons and the outside-in for

the Global muons.

Approximately 99% of muons produced in p-p collisions within the muon system geometrical acceptance and with enough momentum are reconstructed with one of the above methods. Global muon provides better momentum resolution for tracks with $p_T > 200$ GeV; the tracker-only fit performs better for tracks with $p_T < 5$ GeV.

The reconstructed muon candidates must also pass additional identification (ID) criteria. The CMS Collaboration employs several IDs, with varying efficiency and misidentification rates. The efficiency is defined as the number of reconstructed divided by the number of expected muons, whereas the fake rate is the rate of particles misidentified as muons. Among all, the Tight Muon criteria apply stringent requirements for the tracks and matching qualities, focusing on prompt muons and rejecting many secondary muons that originated in decays within the detector. Moreover, to improve the purity of the muon selection, the muon candidates must be isolated. The isolation criterion corresponds to having a momentum that is significantly larger than the energy flow within a cone of $\Delta R = 0.4$.

For the muons with $20 < p_T < 50$ GeV and $I_{rel} < 0.1$, the efficiency (measured by the tag-and-probe technique in $Z \rightarrow \mu\mu$) is above 92%, while the background efficiency (measured in the QCD-enhanced data set) is below 2.5% [165].

3.2.3 Electrons and photons

An electron can emit 33% of its energy interacting with the tracker material before reaching the ECAL in the central barrel region ($|\eta| \approx 0$) and up to 86% in the $|\eta| \approx 1.4$ area. On the contrary, photons deposit their energy only in the ECAL. Moreover, electrons and photons can interact with the material via Bremsstrahlung ($e^\pm \rightarrow \gamma e^\pm$) or photo conversion ($\gamma \rightarrow e^+ e^-$), spreading the energy deposited in the ECAL in the ϕ direction. The energy deposit from both the original electron and the Bremsstrahlung products must be combined to retrieve the unique properties of the resulting multi-particle showers. Dedicated algorithms for identifying photons and electrons combine electromagnetic calorimeter and tracker tracks [199]. The standalone algorithm uses ECAL super-clustering techniques, but combining standalone and PF approaches optimizes the electron reconstruction [192, 199].

The electron reconstruction efficiency is above 95% [159]. The final energy resolution is $\gtrsim 2\%$ in the barrel and $\gtrsim 5\%$ in the endcap for $p_T > 20$ GeV [159].

Similarly to muons, reconstructed electrons and photons must meet identification criteria. Jets and electrons from photon conversion or semi-leptonic b-quark and c-quark decays are the main sources of misidentification. To increase the identification sensitivity, the electrons are identified using a multivariate Boosted Decision Tree (BDT) discriminator [200]. Three working points (WPs) loose, medium and tight with efficiencies of 90%, 80%, and 70%, respectively, are defined.

3.2.4 Hadrons and non-isolated photons

After identifying muons, electrons, and photons, residual tracks and calorimeter clusters are used to reconstruct charged and neutral particles ($\pi^\pm$, $K^\pm$, K^0 , protons or neutrons). Unlinked calorimeter clusters are reconstructed as photons or hadrons. In particular, ECAL clusters without a matching HCAL are photons; otherwise, the PF candidate is a neutral hadron. The calorimeter clusters directly measure the energy of neutral particles. To infer extra neutral particles, the sum of the momentum of all linked tracks is compared to the calorimetric energy. If it exceeds the sum of the track momenta, the excess is attributed to extra PF candidates (photons or neutral hadrons for an ECAL or HCAL excess, respectively).

3.2.5 Jet reconstruction

Quantum chromodynamics processes dominate the underlying processes of the LHC p-p collisions. Except for the top quark, which has a lifetime ($\tau_t = 0.3 \times 10^{-24}$ s) long enough to decay into a W boson and a b quark before hadronization, all other color-charged particles cannot be observed free and undergo fragmentation and hadronization almost immediately. Quarks and gluons generate a collimated spray of hadrons due to the Lorentz boost. The final state products emerge in a narrow ($\eta - \phi$) cone as an energy cluster called jet (Figure 3.4). An analysis using jets must infer the kinematic properties of the initial color-charged parton by clustering the particles observed in the detector into a jet.

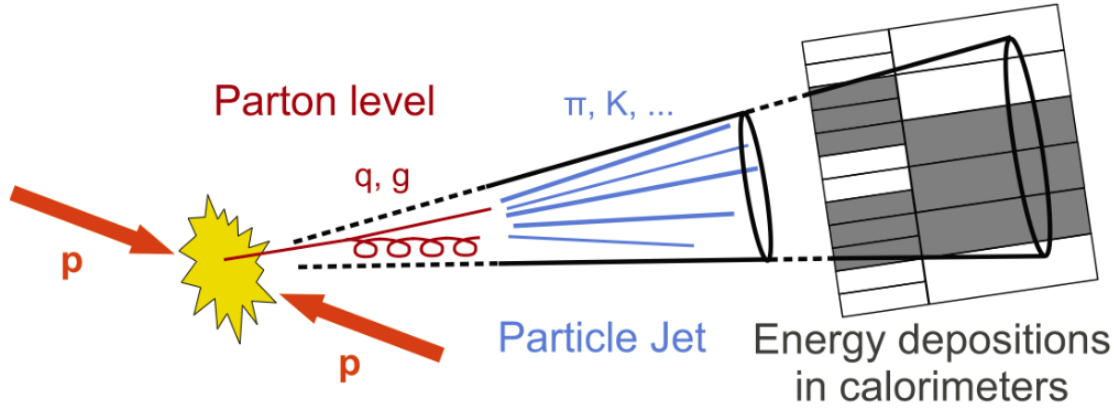


Figure 3.4: A simplified representation of a jet and its component.

The CMS algorithm used in this thesis starts from particle-flow objects, clustered using the anti- k_t algorithm as implemented in the FASTJET package [201, 202]. The anti- k_t algorithm is a sequential recombination algorithm that generalize the k_t [203] and Cambridge/Aachen [204, 205] jet finding algorithm. Jets reconstructed with this algorithm are resistant to soft radiation (infrared safety) and collinear decay products from particle splitting (collinear safety). In each iteration, the algorithm recombines the current set of entries (both PF particles and jet candidates) and calculate their mutual distances (d_{ij}) and update also the distance between an entry and the beamline (d_{iB}) [201]:

$$d_{ij} = \min(p_{T_i}^{2p}, p_{T_j}^{2p}) \frac{\Delta R_{ij}^2}{R^2}$$

$$d_{iB} = p_{T_i}^{2p}$$

where p_{T_i} is the transverse momentum of entry i ; $\Delta R_{ij} = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$ is the distance in the ($\eta - \phi$) plane between entries i and j ; R is the cone radius parameter, typically called jet radius; the parameter p characterizes each algorithm. $p = 1$ for the inclusive k_T algorithm, $p = 0$ corresponds to the inclusive Cambridge/Aachen algorithm and $p = -1$ refers to the anti- k_t algorithm. In this thesis jets with $R = 0.4$ are used.

The clustering proceeds by identifying the smallest between d_{ij} and d_{iB} : if d_{iB} is smaller than d_{ij} , the element i is promoted to a reconstructed jet, and it is excluded from further iterations. Otherwise, entries i and $j' = \min d_{ij}$, defined as the element that minimizes d_{ij} , are merged into a new entry i' and the algorithm proceeds to the next iteration. The algorithm stops as soon as all jet candidates are promoted to reconstructed jets. As a result, the anti- k_t algorithm initially clusters hard and close particles before moving on to soft and distant ones, providing a simple, intuitive, and circular cone structure in the ($\eta - \phi$) plane.

The reconstructed jet energy does not match the total energy of the particles in the jet because the calorimeter response is non-uniform and non-linear. The jet energy is corrected by applying a multiplicative factor to each of the four components of the raw momentum of the jet. A multi-step calibration technique is used to determine the jet energy scale (JES) and resolution (JER) in both real data and simulated events [161], as shown in Figure 3.5.



Figure 3.5: Simplified representation of the factorized approach used for jet calibration in CMS [161].

Several types of jet energy corrections (JEC) are used in a modular approach to account for contamination from neutral pileup particles and detector noise, non-linearity of the calorimeter response in the jet, and jet p_T ; an additional correction is used only in data events to account for residual differences between the jet response in data and simulation. These correction factors are calculated using the jet p_T and η to rescale the four-momentum of the original jet. The jet identification efficiency was measured with a tag-and-probe technique to be close to $\sim 100\%$ using 2.3 fb^{-1} of collision data recorded in 2015 at $\sqrt{s} = 13 \text{ TeV}$ [206].

The overall uncertainty associated to the jet energy calibration procedure for central jets with $p_T > 30 \text{ GeV}$ is substantially below 2% and is dominated by jet flavor dependency, with modest contributions from pileup offset, absolute, and relative scale corrections. In terms of dependence on η , the overall uncertainty is flat in the area $|\eta| < 2.5$ and calculated as 2.5% for jets with $p_T \approx 30 \text{ GeV}$. In the central detector, the jet flavor description predominates, whereas pileup modeling and methodology-related uncertainties gain importance in the more forward region. After, the jet transverse momentum resolution in simulation needs to be adjusted to match the jet resolution in data: they are measured in data and MC events and the resulting data/MC scale factors are used to correct the JER in simulated events. The typical jet energy resolution for jets in the barrel region is approximately 15-20% for $p_T \approx 30 \text{ GeV}$, 10% for $p_T \approx 100 \text{ GeV}$ and 5% for $p_T \approx 1 \text{ TeV}$ [161].

3.2.6 Missing transverse momentum

CMS can not directly detect neutrinos and other hypothetical neutral weakly interacting neutral particles due to their low interaction rate with the matter. An imbalance in the transverse momenta of all visible products in the event indicates the presence of neutrinos. Under the assumption that the colliding proton pair's transverse momentum is negligible, all p-p interaction products should have zero transverse moments. The missing transverse momentum $\vec{p}_T^{\text{miss}}$ is estimated as the negative vectorial sum of all PF particles' transverse momenta in an event and its magnitude is denoted as p_T^{miss} :

$$p_T^{\text{miss}} = |\vec{p}_T^{\text{miss}}| = - \sum_{\text{visible}} \vec{p}_T$$

The accuracy of p_T^{miss} is sensitive to tracker inefficiency, sub-detector calibration, reconstruction efficiency, mis-measurement of jet energy, and pileup effects. Because of the high

sensitivity to overall detector performance, some events may have very large values. Special noise-rejection algorithms are used to reject events with a bogus p_T^{miss} [207]. The bias on its measurement is minimized by using jet energy corrections and propagating them into the p_T^{miss} measurement as follows:

$$p_T^{\text{miss,corr}} = \vec{p}_T^{\text{miss}} - \sum_{\text{jets}} (\vec{p}_{T,\text{jet}}^{\text{corr}} - \vec{p}_{T,\text{jet}}),$$

where corr refers to the corrected values [207].

The analysis reported here employs a threshold to select events with appropriate neutrino activity based on the predicted $t\bar{t}H, H \rightarrow b\bar{b}$ signal process signature and to reduce background from QCD multijet processes.

3.2.7 Pileup mitigation

The PF algorithm effectively mitigates the effects of pileup events, but the CMS Collaboration also uses other techniques. For example, the Charged Hadron Subtraction (CHS) algorithm [197] is widely used in Run 2 analyses and this thesis. In the CHS algorithm, charged hadrons with reconstructed tracks associated with pileup vertices are removed from jet clustering. CHS cannot be applied to photons, neutral hadrons, and reconstructed particles outside the tracker acceptance. Due to limited coverage in η , no information is available on the origin of a particle for the latter case; therefore, the average p_T contribution of pileup events is estimated based on the assumption that the p_T density is uniform per unit area in the $(\eta - \phi)$ -plane. This approach has a limited impact because the additional corrections do not affect the jet shape or substructure. The pileup per particle identification (PUPPI) technique was introduced to overcome this limitation [208]. It calculates, event by event, and for each particle, the probability that it comes from the leading PV and scales down their energy based on that probability. The sketch in Figure 3.6 depicts the CHS and PUPPI algorithms.

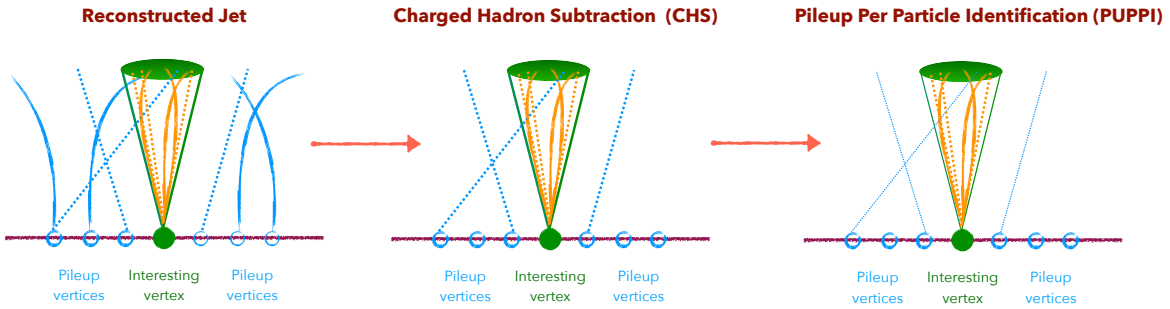


Figure 3.6: Sketch of a PU suppression techniques. Solid (dashed) lines represent charged (neutral) PF candidates. Thin lines represent down-weighted 4-momenta [209].

3.2.8 Heavy flavor jet identification

Many CMS analyses rely on accurately identifying the type of particle that generated a certain jet. A reliable classification (jet tagging) improves the selection efficiency of relevant events and reduces the large backgrounds coming mostly from light-flavor processes. Since $t \rightarrow Wb$, the identification of b jets (b tags), known as b tagging, has been used in SM Higgs boson and top-quark research. The $t\bar{t}H, H \rightarrow b\bar{b}$ signal process, investigated in this thesis, consists of four jets generated from decays of long-lived B hadrons, whose fragmentation

process cannot be fully described at the perturbative level. The b-tagging procedure relies on features of B hadrons, such as secondary vertices, kinematic properties (long lifetime and relatively large mass $m \sim 5$ GeV), and jet composition (the decay products are more energetic than those of light-quark hadrons). Figure 3.7 shows a simplified schematic of a b jet that travels a long distance in the detector before decaying at a secondary vertex and can be resolved thanks to the inner tracking system's high resolution. Secondary vertex (SV) reconstruction can yield powerful discriminating factors, but it's not always achievable. The SV mass, related to heavy-flavor hadron mass, and PV distance are efficient for flavor tagging. To recover B hadron decays from tracking data, specific techniques like the Inclusive Vertex Finding (IVF) algorithm have been deployed [210].

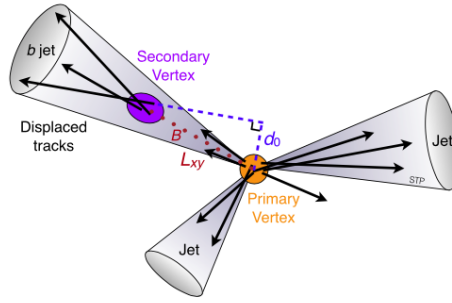


Figure 3.7: A simplified representation of a b jet [211].

The CMS Collaboration developed several b-tagging algorithms throughout Run 1 and Run 2. MVA b-tagging originally used high-level variables as inputs for b-jet recognition. Modern techniques use machine learning algorithms to incorporate these features into more advanced discriminators. Each method produces a single discriminator value for each jet; larger values indicate b-flavor, and lower values indicate light flavors. It is possible to compare algorithm performance based on tagging efficiency (fraction of jets created by b quarks that are accurately tagged) and mistag rate (fraction of light flavor or gluon tagged jets). A jet is attributed a b tag if the discriminator output exceeds a threshold value; these values are defined by a misidentification rate and summarized in Table 3.1.

Working Point	Min. b-tag discriminant value			Misidentification Rate (%)
	2016	2017	2018	
loose	0.0614	0.0521	0.0494	10
medium	0.3093	0.3033	0.2770	1
tight	0.7221	0.7489	0.7264	0.1

Table 3.1: Summary of the working points, the associated minimal values of the DEEPJET discriminant and the corresponding misidentification rates of light-flavor jets.

This analysis requires the best b tagger to ensure good signal identification efficiency. The DEEPJET jet-flavor identification algorithm [212] is the most efficient one; it uses high-level variables and low-level features to improve identification. The algorithm is a deep neural network that uses global event variables, charged and neutral PF jet-component properties, and SV information. A comparison between the performance of DEEPJET and DEEPCSV in the identification of b jets is in Figure 3.8. For the medium WP, with 1% misidentification probability, the DEEPJET tagger reaches b-jet identification efficiency of $\approx 82\%$.

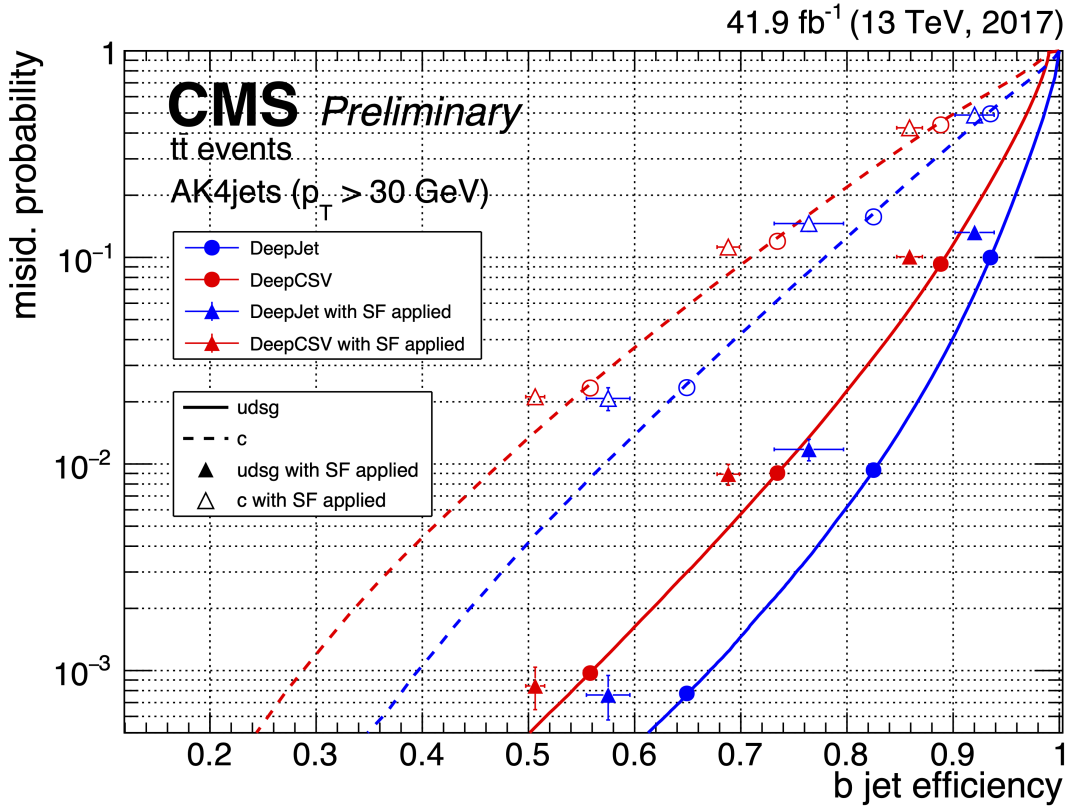


Figure 3.8: Performance of the DEEPIET and DEEPCSV algorithms. The plot shows the probability of misidentifying non b jets as b jets to the efficiency of correctly identifying b jets. The results shown were obtained using jets with $|\eta| < 2.5$ and $p_T > 30$ GeV from simulated top pair events. For the loose, medium, and tight working points, the data-to-simulation scale factors have been applied and are represented by the cross-hairs. Circular markers represent the performance of the respective working point in simulated data. [213].

LUMINOSITY MEASUREMENT

4

4.1	Luminosity of collider experiments	59
4.2	Precision luminosity measurements at CMS in Run 2	60
4.2.1	Factorization bias with beam-imaging method	65

This chapter presents an overview of the precision luminosity measurement performed with the CMS detector using p - p collisions at $\sqrt{s} = 13$ TeV during the Run 2 data-taking period. Luminosity is a crucial parameter in high-energy physics experiments as it determines the number of collisions that occur per unit of time and therefore the number of particles produced. The measurement of luminosity is therefore essential for the accurate calculation of cross sections and other physics observables. The focus of this chapter is on the general procedure used to obtain the absolute luminosity value and the investigation of specific systematic uncertainty sources. In particular, the chapter will discuss the beam imaging method employed to assess the systematic uncertainties associated with the so-called factorization bias.

4.1 Luminosity of collider experiments

The luminosity $\mathcal{L}$ is a measure of the collision rate, so a critical feature of any collider experiment. It corresponds to the proportionality factor between the event rate and the cross section of a process:

$$\frac{dN}{dt} = \mathcal{L} \cdot \sigma. \quad (4.1)$$

The cross section σ of a process can be estimated via the underlying interaction theory and depends on parameters like the center-of-mass energy $\sqrt{s}$ of the collision, the mass of the final state particle, and the coupling strength between the initial and final state particles. However, σ can also be calculated by counting the number of events N_{events} from a collision experiment as:

$$\sigma = \frac{N_{\text{events}}}{\mathcal{L}}. \quad (4.2)$$

As σ can be both estimated and calculated, comparing the values is fundamental to extracting the theory parameters. Therefore, the precision cross section measurements require an accurate understanding of the luminosity [214].

In general, the luminosity $\mathcal{L}$ can be derived from the knowledge of machine parameters:

$$\mathcal{L} = \frac{N_1 N_2 \nu}{A_{\text{transverse}}}, \quad (4.3)$$

with N_1 , N_2 the number of particles in each of the two colliding beams circulating at a revolutionary frequency ν , and $A_{\text{transverse}}$ the effective area where the collisions occur. Usually, ν is known, and the number of particles can be precisely assessed via the beam currents.

However, the poor knowledge of the overlap area limits the precision of the luminosity calculation [215].

Some collider experiments conduct precision cross section measurements of processes well-known with high accuracy, such that the inverse of Equation 4.2 provides an estimate of the luminosity. However, at hadron colliders, measuring the luminosity is more complicated due to the composite nature of the proton and the non-perturbativity of QCD at small energies. At the world's first p-p collider, the Intersecting Storage Rings at CERN, Simon van der Meer proposed to estimate the beam overlap area in Equation 4.3 using beam separation scans [25]. With this formulation, for the special case of factorizable particle densities $f_i(x, y) = f_i^x(x)f_i^y(y)$, the overlap area is expressed as $A_{\text{transverse}} = 2\pi\Sigma_x\Sigma_y$, via the effective widths of the beam overlap area along two orthogonal directions Σ_x, Σ_y . Through the so-called Van der Meer (VdM) scans, where the two beams are first significantly separated and then moved across each other in their respective directions, it is possible to derive the effective widths and estimate the luminosity for head-on collisions. Then, the absolute scale of the rate measurement of the luminosity detector, σ_{vis} , is calibrated with the measured luminosity [216].

The LHC experiments during the Run 1 p-p collisions at $\sqrt{s} = 8$ TeV reached a precision between 1.2% and 2.6% for the luminosity measurement [217–220], with the LHCb experiment achieving the best precision with a combination of the VdM method and results from a beam-gas imaging method. The methodology and the dominant sources of systematic for the measurements of the CMS experiment at Run 2 proton-proton collisions at $\sqrt{s} = 13$ TeV are presented in the following.

4.2 Precision luminosity measurements at CMS in Run 2

In the CMS experiment, the VdM method is first used to calibrate the absolute luminosity scale of the luminosity detectors (luminometers). Then, the VdM scans are used to compute the visible cross section of the luminometers from the measured rate measurement at head-on collisions $R(0, 0)$ as:

$$\sigma_{\text{vis}} = \frac{2\pi\Sigma_x\Sigma_y R(0, 0)}{\nu N_1 N_2}. \quad (4.4)$$

Afterward, the second part of this two-step procedure is integrating the rate of the luminosity detectors over the whole data-taking period and normalizing it with the obtained σ_{vis} to yield the integrated luminosity of the whole data set.

During the LHC Run 2 period, several luminometers were employed by the CMS experiment to monitor and measure the luminosity. The Pixel Luminosity Telescope (PLT), consisting of eight three-layer telescopes located 1.8 m from the interaction point in both directions [221], and the Fast Beam Conditions Monitor (BCM1F), made of silicon and diamond sensors with a time resolution of 6.25 ns, mounted on the same carriages as PLT, are two systems dedicated to luminosity measurement. Together with the luminosity system installed as part of the Hadron Forward Calorimeter (HF), they run on a separate data acquisition (DAQ) system, completely independent of the main CMS readout. HF uses two algorithms, one based on the transverse energy sum (HFET) and the other on the fraction of occupied towers (HFOC). The results are based on offline measurements made with the forward hadron calorimeter and the pixel cluster counting (PCC) method, which uses silicon pixels to determine the offline luminosity. In addition, the inner tracker information is employed in the vertex counting (VTX) method, the rate of the muon track in the muon barrel track finder is measured from the barrel drift tubes (DT), and also the rate of photons in ionization chambers from the CERN radiation and environmental monitoring system

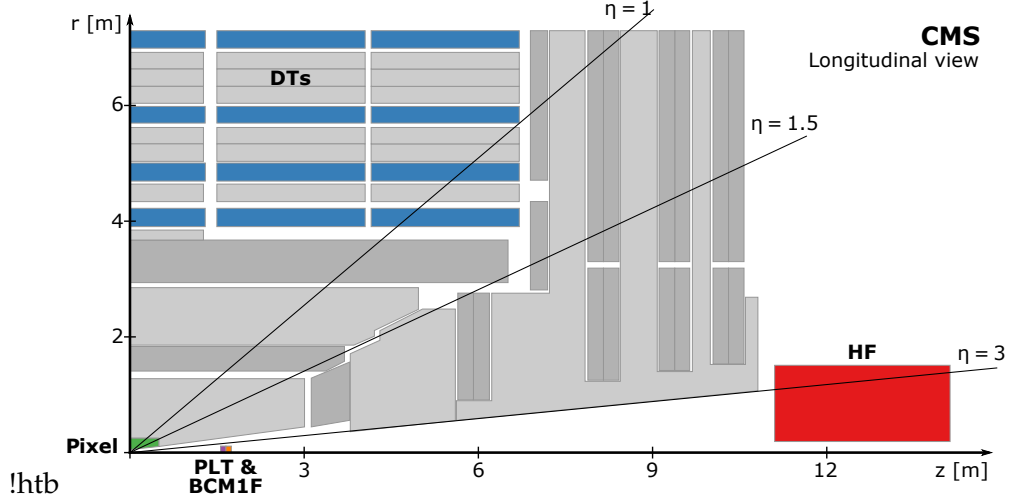


Figure 4.1: Schematic view of the CMS detector in the $(r-z)$ plane with the main luminometers [26]. The two RAMSES monitors are located directly behind HF. Only a quarter is shown here because the detector view is symmetric about the horizontal and vertical axes. The center of the detector is at the origin and corresponds to the collision point [26].

(RAMSES). Figure 4.1 shows a schematic cross section in the $(r-z)$ plane of the CMS detector, highlighting the main luminometers.

The VdM scans and other beam separation scans are performed by LHC using special beam optics to increase the transverse size of the bunches and allow precise measurements of the beam overlap region. In particular, small isolated and well-separated bunches are used to reduce beam-beam effects or other effects like long-range interactions away from the collisions.

Figure 4.2 represents the relative transverse beam positions as a function of time during the p-p data taking period in 2016 at $\sqrt{s} = 13$ TeV, with different kinds of beam separation scans. During the VdM scans, the beams are separated in the transverse plane and then moved in 25 steps; there are always two scans along the horizontal and vertical directions. Several scan pairs of this type are performed for reproducibility checks and monitoring of the changes. Then, there are beam-imaging (BI) scans, which are like the VdM scans but with only one beam moving in 19 steps across the other beam kept at the nominal head-on position; for both beams, both horizontally and vertically, this results in four beam-imaging scans employed in the beam-imaging analysis presented in Section 4.2.1. Finally, the length scale (LSC) scans, where the two beams are separated and move simultaneously to keep the separation constant, are employed to calibrate the beam separation scale.

Several systematic effects known as calibration uncertainties affect the value of σ_{vis} derived from the VdM scan analysis. In addition, there are also integration uncertainties arising from transferring the calibration of the luminometers from the special VdM fill conditions to the conditions in normal physics data-taking fills. The significant sources contributing to the uncertainty of the measured luminosity are the following:

- **Beam current measurement.** The number of protons per beam is provided by dedicated LHC devices and can be measured with an accuracy of 0.2 % [28, 222, 223].
- **Ghost and satellite charges.** Non-colliding protons can also contribute to the beam current measurement, but not the luminosity, via spurious charges in nominally empty bunch slots (ghost charges) and non-colliding buckets of the colliding bunches (satellite charges). Therefore they should be subtracted by measuring the longitudinal distribution of protons in empty bunch slots. The impact on σ_{vis} is about 0.1% [224].

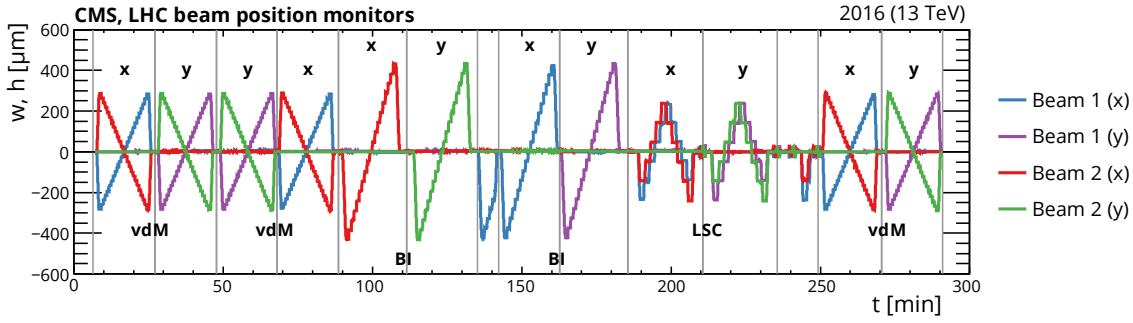


Figure 4.2: Relative change in the position of the two beams during a fill in 2016, in the horizontal (x) or vertical (y) directions, as a function of time [26]. The grey vertical lines delineate different kinds of beam separation scans.

- **VdM scan evaluation.** The effective beam overlap widths Σ_x (Σ_y) are obtained from a fit with a double-Gaussian function to the measured rates as a function of the beam separation in the horizontal (vertical) plane. An example fit result is shown in Figure 4.3 (a). An uncertainty is estimated from the consistency of Σ_x , Σ_y among the estimates from different pairs of VdM scans [26].
- **Beam separation scale.** The precise knowledge of the beam separation at each scan step is essential for any measurement evaluated from any beam separation scan. Several known features affect the nominal position inferred from the current setting of the steering magnets. The length scale scans calibrate the absolute beam separation scale by comparing nominal beam positions to the measured one of reconstructed interaction vertices with the CMS tracker [26, 28]. A time-dependent movement of the proton beams away from their regular orbit, known as orbit drift, has been observed and accounted for.
- **Beam-beam effects.** The electromagnetic interactions between the beams increase the transverse distance due to electric repulsion and result in a separation-dependent rise of the collision rate because of a variation of the transverse shape of the proton bunches due to (de)-focusing effect.
- **Residual differences.** Despite several VdM scan pairs being conducted during calibration and the bunch-to-bunch and scan-to-scan consistency, also using different luminometers, are examined to evaluate the systematic uncertainty of cross-detector calibration, a residual difference is consistently observed, see Figure 4.3 (b); corrections are taken into account.
- **Factorization bias.** The VdM method computes σ_{vis} using Equation 4.4, assuming that the beam proton densities are factorizable. If this assumption is not respected, then the estimated value of σ_{vis} is biased and needs to be corrected. The beam-imaging scans help reconstruct the two-dimensional transverse beam shape, and the uncertainty of the bias estimation is one of the dominant contributions [225–227]. The last Section of this chapter focuses on the method employed by the CMS experiment to account for this effect.
- **Out-of-time pileup.** Most detector measurements involve false signals from contributions caused by electronic spillover or activation of the detector material after collisions, leading to a sizable increase in the measured rate. Even though the calibration is not

affected, this yields a non-linear response of the detector. The observed rate from nominally empty bunches is used to assess the extent of these corrections.

- **Linearity.** Luminometers are required to have a linear response to the instantaneous luminosity when extrapolating the calibration from the special to the regular data-taking conditions. However, most detectors show some non-linearity, thus independently corrected and monitored with short VdM-like beam separation scans, so-called emittance scans, at the beginning and end of fills. The aging of photomultipliers and fibers in the forward hadron calorimeter is an example of changing detector conditions over time. The efficiency of the HFET response, as measured by emittance scans, is depicted in Figure 4.4 for 2017 and 2018, with a slope showing a significant reduction in efficiency with time. Additionally, the overall linearity is assessed with cross-detector comparisons. Finally, the ratio of two luminosity detectors is computed as a function of the instantaneous luminosity. A slope in the distribution is derived to quantify the residual non-linearity, up to 1%.
- **Stability.** Luminometers are also required to be stable against changes in the detector over time, which might be caused by radiation damage or operational issues. Corrections are calculated using the emittance scan data and cross-detector comparisons, similarly to linearity.

Table 4.1 summarizes the different sources contributing to the uncertainty of the measure of integrated luminosity for the p-p collision at $\sqrt{s} = 13$ TeV due to the VdM scan procedure (calibration) and the detector operations (integration).

In Run 2, the luminosity at the CMS experiment has been also measured for p-p collisions at $\sqrt{s} = 5.02$ TeV [228, 229], lead-lead collision at $\sqrt{s} = 5.02$ TeV [230], as well as for proton-lead collisions at $\sqrt{s_{NN}} = 8.16$ TeV [231].

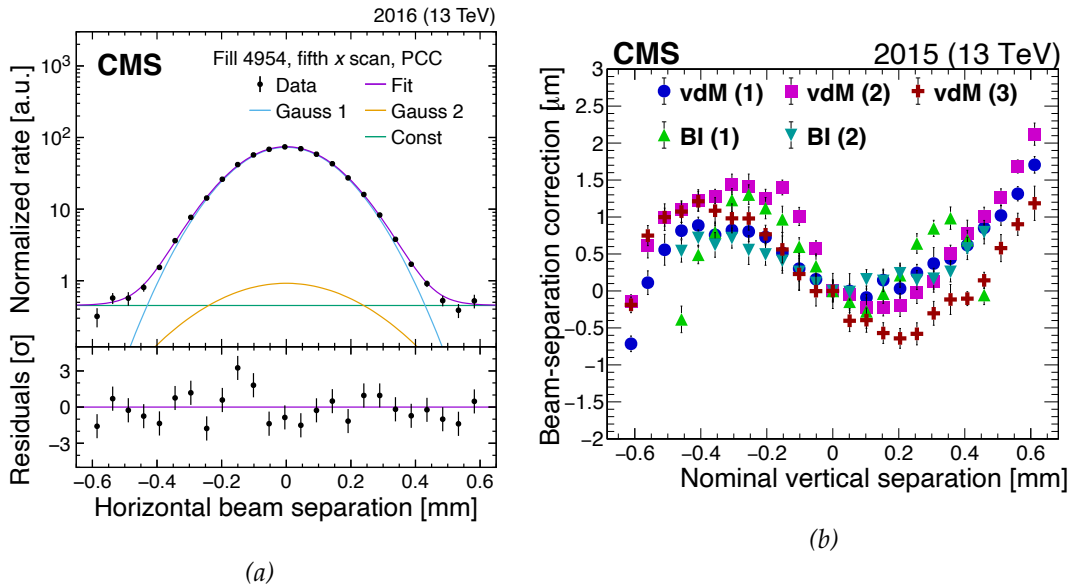


Figure 4.3: (a) Example of a VdM scan illustrating the normalized rate as a function of beam separation in the x direction and the fitted curves. The difference between the measured and fitted values is shown in the lower panel, divided by the statistical uncertainty [26]. (b) The beam-separation residuals in y [26].

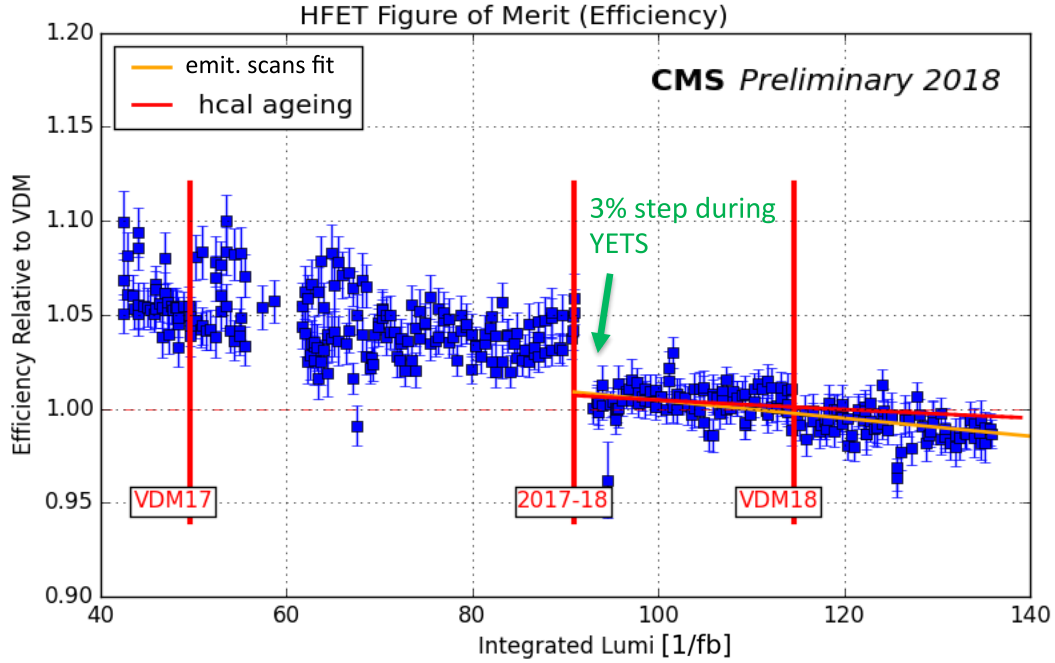


Figure 4.4: Relative efficiency of the HFET detector response as determined from emittance scans as a function of integrated luminosity for all LHC fills in 2017 and 2018 p - p data. The slope indicates instability [232].

Source		2015	2016	2017	2018
Calibration uncertainty					
Beam current normalization	[%]	0.2	0.2	0.3	0.2
Ghost and satellite charge	[%]	0.1	0.1	0.1	0.1
Length scale calibration	[%]	0.2	0.3	0.3	0.2
Orbit drift	[%]	0.2	0.2	0.2	0.2
Beam-beam effects	[%]	0.5	0.5	0.6	0.2
Residual differences	[%]	0.8	0.5	0.2	0.1
Transverse factorizability	[%]	0.5	0.5	0.8	2.0
Integration uncertainty					
Out-of-time pileup Type 1 corrections	[%]	0.3	0.3	0.2	0.1
Out-of-time pileup Type 2 corrections	[%]	0.1	0.3	0.3	0.4
Cross-detector stability	[%]	0.6	0.5	0.5	0.6
Linearity	[%]	0.5	0.3	1.5	1.5
Total uncertainty	[%]	1.6	1.2	2.3	2.5
Integrated luminosity	[fb ⁻¹]	2.27	36.3	41.5	59.8

Table 4.1: Summary of main contributions (in %) to the systematic uncertainty in σ_{vis} at $\sqrt{s}=13$ TeV [26–28].

4.2.1 Factorization bias with beam-imaging method

The VdM method uses two orthogonal transverse scans to estimate the width of the beam overlap region, which is assumed to be described by a function that factorizes into two one-dimensional function $f_i(x, y) = f_i^x(x) f_i^y(y)$, as in Equation 4.4. However, this assumption does not hold as soon as the width of the transverse area in one direction depends on the other, resulting in an overestimation of the beam overlap area.

At LHC, the first evidences of non-factorizable beam shapes were found in p-p collision at $\sqrt{s} = 8$ TeV. Starting from Run 2, the experiments began estimating the bias by computing the transverse area from a direct integration and comparing it to the beam overlap region computed from the widths along the two directions. In CMS, the beam-imaging method reconstructs the transverse proton densities from four beam-imaging scans. The distributions of measured vertices during the scans estimate the proton density, whose parameters are then extracted from a fit to the vertex data.

As described above, during the beam-imaging scans, one beam remains at the nominal head-on position while the other moves along one transverse direction in steps. At each step, the transverse distribution of vertices depends on the separation of the two beams and the two proton densities as:

$$N_{\text{vtx}}(x, y; \Delta x) \propto [f_1(x + \Delta x, y) \cdot f_2(x, y)] \otimes V. \quad (4.5)$$

for the case of the first beam moving horizontally and with a convolution to the vertex resolution function V . When all steps are combined (the so-called beam image), the transverse distributions of the vertices are described by the sum over the per-step distribution, which can be approximated for a small step size by the integral :

$$\sum_{\Delta x} N_{\text{vtx}}(x, y; \Delta x) \propto \left[\left(\int_{-\infty}^{+\infty} f_1(x + \Delta x, y) d\Delta x \right) \cdot f_2(x, y) \right] \otimes V. \quad (4.6)$$

After the integration over the horizontal direction, the dependence on x remains only in $f_2(x, y)$. Therefore, the horizontal dependence of the vertex distribution from all steps represents a measure for the horizontal profile of $f_2(x, y)$. Then, to completely determine the transverse proton distribution functions of the two beams, four beam images are needed from scans for each beam along the two orthogonal directions.

The presence of the vertex resolution V in Equation 4.5 represents a challenge in the beam-imaging method. V is usually modeled with a two-dimensional Gaussian function without a correlation parameter. Since the convolution of two Gaussian functions is another Gaussian function with parameters that can be estimated analytically, only Gaussian functions and sums of Gaussian functions are employed as proton density models.

A two-dimensional Gaussian distribution $\text{Gauss}^{2D}(x, y; \sigma_{i,x} \sigma_{i,y}, \rho)$ is the simplest model to describe the non-factorizable normalized transverse beam distributions:

$$f_i(x, y) = \frac{1}{2\pi\sigma_{i,x}\sigma_{i,y}\sqrt{1-\rho_i^2}} \exp \left[-\frac{1}{2(1-\rho_i^2)} \left(\frac{x^2}{\sigma_{i,x}^2} + \frac{y^2}{\sigma_{i,y}^2} - 2\rho_i \frac{x}{\sigma_{i,x}} \frac{y}{\sigma_{i,y}} \right) \right]. \quad (4.7)$$

The correlation parameter ρ represents the rotation of the factorization axes relative to the coordinate directions and is bounded to be $|\rho_i| < 1$. A value of $\rho_i = 0$ represents the case of the two-dimensional Gaussian factorizing into the two one-dimensional distributions.

Another possible way to describe the non-factorizable transverse beam distribution is via involved shape models built by the sums of Gaussian distributions. In this analysis, normalized sums of up to three Gaussian distributions are considered as fit models, as listed in Table 4.2. The additional Gaussian components with wider widths contribute to the description of

name	formula
single-Gaussian	$SG = \text{Gauss}_M$
double-Gaussian	$DG = w_M \text{Gauss}_M + w_W \text{Gauss}_W$
triple-Gaussian	$TG = w_M \text{Gauss}_M + w_W \text{Gauss}_W + w_E \text{Gauss}_E$
super-Gaussian	$\text{SupG} = -w_N \text{Gauss}_N + w_M \text{Gauss}_M$
super-double-Gaussian	$\text{SupDG} = -w_N \text{Gauss}_N + w_M \text{Gauss}_M + w_W \text{Gauss}_W$

Table 4.2: List of two-dimensional functions utilized in the beam-imaging method as fit models to represent the beam proton densities. The indices stand for the main (M), narrow (N), wide (W), and extra-wide (E) components. All models are constrained to have the sum of the weights set to one.

the tails. In contrast, narrow width and a small weight subtracted from the main Gaussian function in the super functions approximate the flattened central part.

For the beam-imaging analysis from p-p data-taking at $\sqrt{s} = 13$ TeV, the reconstructed vertices are selected if they meet standard quality criteria, have at least 14 associated tracks, and correspond to the time interval of beam-imaging scans.

Technically, the beam-imaging method does not explicitly rely on the beam positions but on the assumption that one beam stays at a fixed position. Nevertheless, as introduced in Section 4.2 there are beam-beam effects that could impact this assumption. First, the electric repulsion between the two proton beams causes a change in the position of the non-moving beam. The beam-beam deflection, analytically determined and dependent on the separation of the beams, is depicted in Figure 4.5 (a). Second, there is the so-called dynamic β^* effect, which describes the defocusing of one beam due to the quadrupole field of the other beam. This phenomenon affects the shape of the proton density differently at different distances, as shown in Figure 4.5 (b). Both these effects can be calculated analytically, and therefore they are fully taken into account in scaling the positions of the reconstructed vertices and as per-vertex weights, respectively.

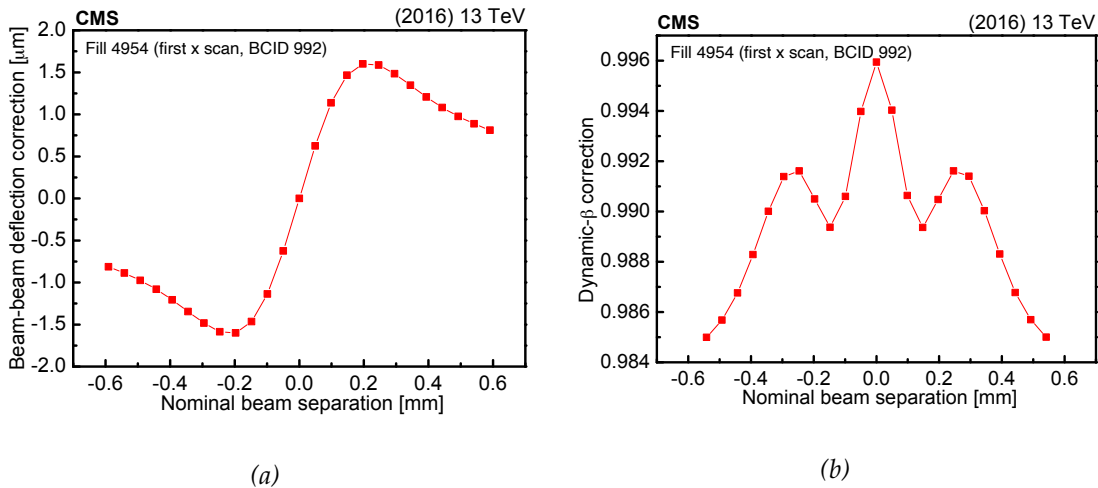


Figure 4.5: (a) Beam-beam deflection due to closed-orbit shift and (b) the rate correction due to the dynamic β^* effect, as a function of the nominal beam separation. Lines represent first-order polynomial interpolations between any two adjacent values [26].

A simultaneous fit to the four beam images is performed with the ROOFIT software [233]; for each bunch crossing number (BCID), the collision events are analyzed independently.

The agreement between the fitted model and the data is analyzed with a two-dimensional residual distribution for each beam image. The relative residual is obtained as the difference between the data and the fit model, divided by the statistical uncertainty of the data. From the sum of all relative residuals, a χ^2 is obtained to quantify the agreement. Two projections along the radial and angular coordinates are also provided for all fits. In particular, for the radial projection, the distance of each bin to the center of the vertex distribution is estimated, and the average relative residual is assessed. For the angular projection, the angle is determined between the positive x -axis and the ray from the center of the vertex distribution through the bin.

A fit of all available data sets is performed, starting from the most straightforward and still factorizable fit model of a single-Gaussian with the correlation parameter fixed to zero to the most complex super-double Gaussian. For example, three fit results for a specific fill and BCID are shown in Figure 4.6, 4.7 and 4.8 using single-Gaussian with $\rho \neq 0$, double-Gaussian, and super-double-Gaussian, respectively. Note the different scale on the y -axis of the radial and angular pulls in the single- and double-Gaussian results with respect to the super-double-Gaussian. The SG type still shows a strong correlation between pulls and angular directions, but as the complexity increases, the remaining structures get flattered and are well-described by the SupDG type fit model. Consistently, for both fills analyzed here and all BCIDS, the best data descriptions come from the SupDG fit, i.e., the sum of three Gaussian distributions, where the narrow component has a negative weight.

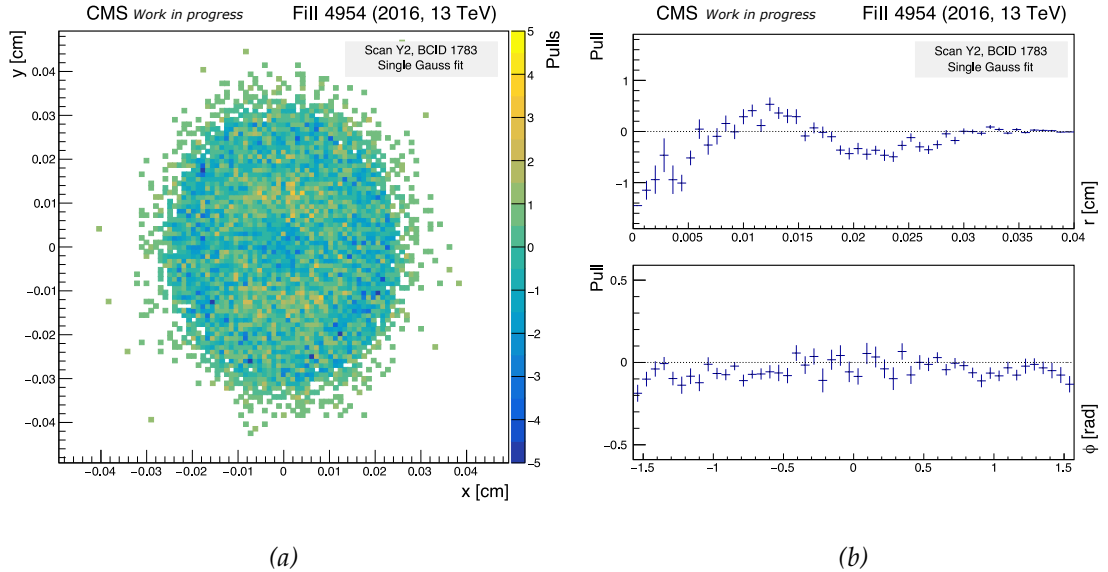


Figure 4.6: Example of the pull distributions of the single-Gaussian fit model with respect to the vertex distribution recorded in a fill in 2015. (a) The plot depicts the two-dimensional pull distributions. (b) The plots show the per-bin pulls averaged over the same radial distance (upper) or angle (lower). The error bars represent the standard error in the mean of the pulls in each bin.

The parameters describing the transverse proton densities of the two beams are the fit results, and they can be used as inputs to obtain the true beam overlap area $A_{\text{overlap}}^{\text{true}}$ from direct integration of the overlap of the bunch densities. The comparison to $A_{\text{overlap}}^{\text{VdM}}$, derived from a MC simulated VdM scan pair generated with the reconstructed bunch densities, yields the evaluation of the factorization bias, and a correction to the values of σ_{vis} is applied as:

$$\sigma_{\text{vis}}^{\text{corrected}} = \sigma_{\text{vis}}^{\text{biased}} \left(1 + \frac{A_{\text{overlap}}^{\text{VdM}} - A_{\text{overlap}}^{\text{true}}}{A_{\text{overlap}}^{\text{true}}} \right). \quad (4.8)$$

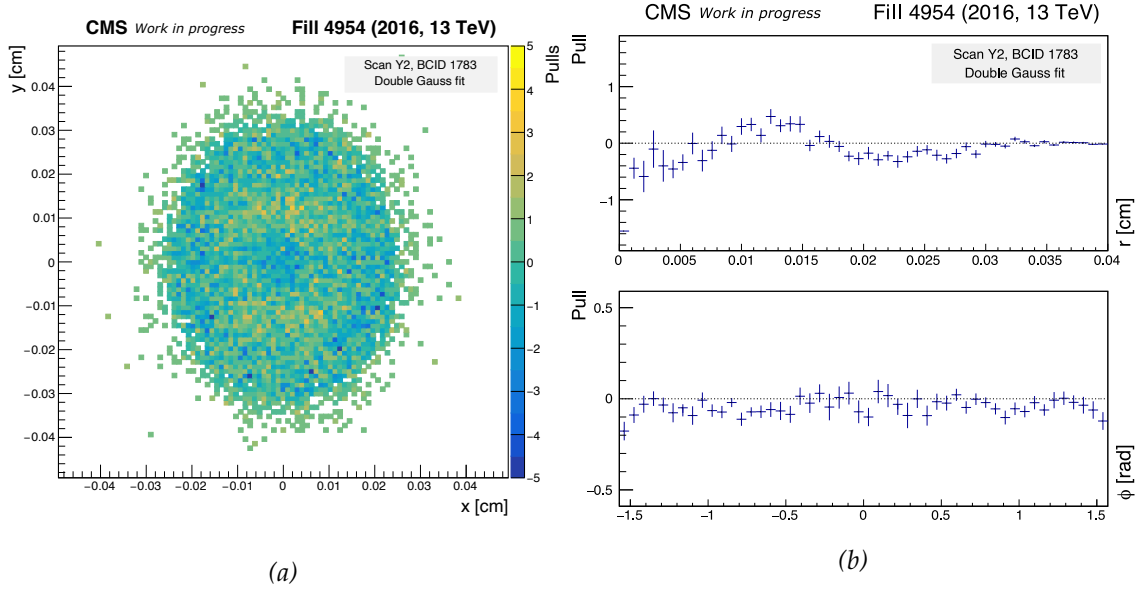


Figure 4.7: Example of the pull distributions of the double-Gaussian fit model with respect to the vertex distribution recorded in a fill in 2015. (a) The plot depicts the two-dimensional pull distributions. (b) The plots show the per-bin pulls averaged over the same radial distance (upper) or angle (lower). The error bars represent the standard error in the mean of the pulls in each bin.

The bias is calculated individually for each fill and each bunch crossing, and Figure 4.9 summarizes the results. The values for the estimated bias are averaged into a final correction value of 1.3 (0.9)% in σ_{vis} for 2015 (2016) because of the assumption of $x - y$ factorization.

The statistical uncertainty of 0.1% associated with the measured bias is estimated by 1000 repetitions of the MC simulation of the VdM scans with the same proton densities. Additionally, a systematic uncertainty of 0.5% is assessed via a closure test compared to the data fits of simulated beam-imaging scan data built by randomly drawing fit model parameters, then fitted with the same model and procedure [26, 234].

The beam-imaging method, and the simultaneous fit performed with the ROOFIT software, have been applied to provide preliminary results of 2017 and 2018 data [27, 28]. However, in contrast to the good description of the beam-imaging data from 2015 and 2016 using a SupDG model for proton densities, the fit results indicate a worse agreement between data and the fit model (Figure 4.10 and 4.11). The significant remaining residuals result in unreliable estimates of the corrections to σ_{vis} .

The reasons are not yet understood, but possible explanations could include alignment issues or correlations in the beam shapes not similar to those observed in previous years. Therefore, it is necessary to investigate the method further and address the issues.

The convolution with the two-dimensional vertex resolution function (see Equation 4.5) is one obvious limitation since it restricts the models describing the proton density to those that can be analytically convoluted with a Gaussian function. Alternative approaches have been examined by the author of this thesis to investigate the factorization bias estimation further. Among these are studies on the effect of vertex resolution value on final results and a feasibility study on using a deconvolution technique of beam-imaging data to fit arbitrary models directly. The aim is to quantify how the number of tracks associated with the reconstructed vertices affects the results. Figure 4.12 depicts the vertex resolution and number of events from a fill in 2017 data as a function of the minimum number of associated tracks: the loss in statistics is greater than the reduction in vertex resolution values. The decrease of vertices statistics results in a significant reduction of the steps sensitive to additional

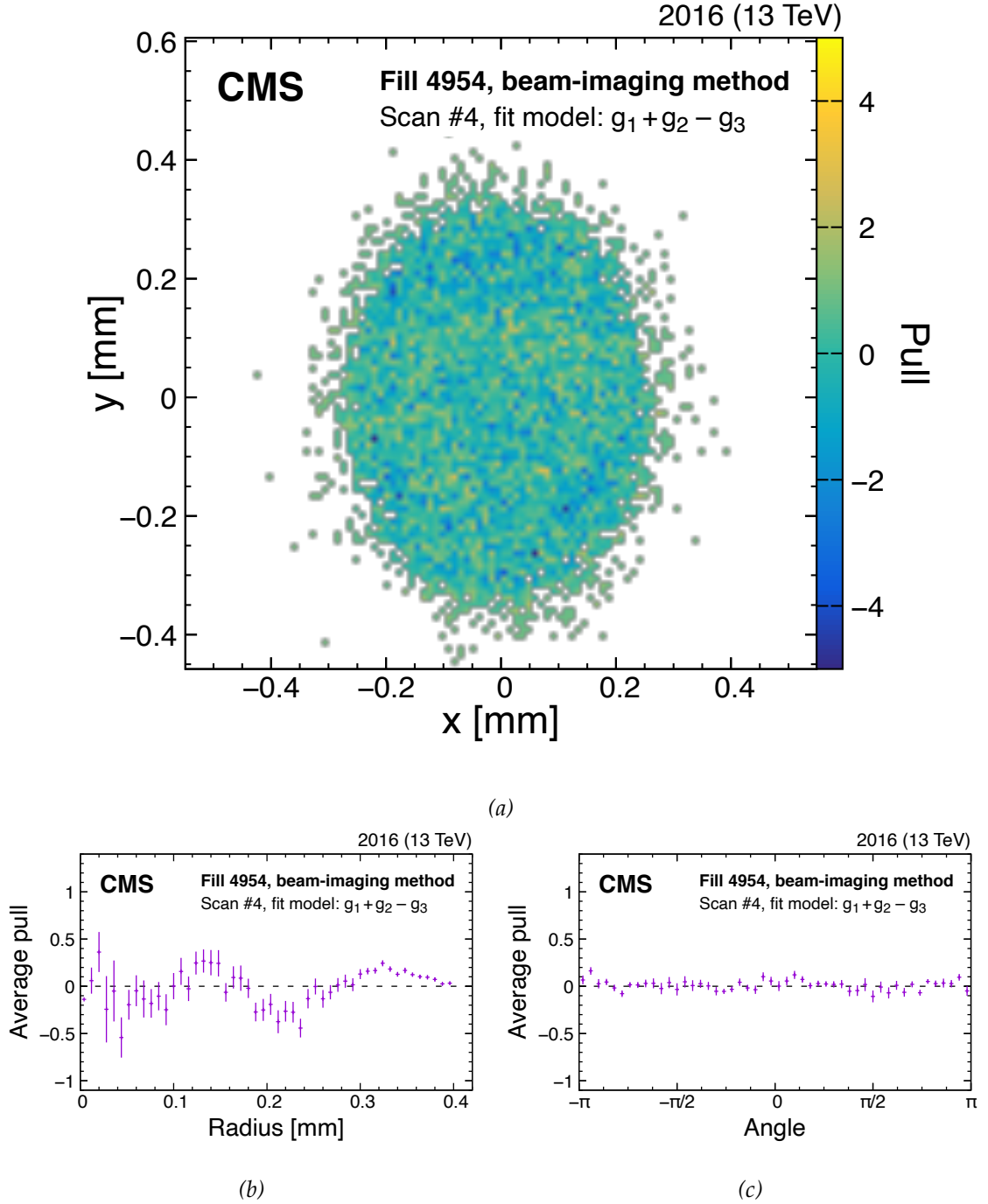


Figure 4.8: Example of the pull distributions of the super-double-Gaussian fit model with respect to the vertex distribution recorded in a fill in 2015. (a) The plot depicts the two-dimensional pull distributions. The lower plots show the per-bin pulls averaged over the same radial distance (b) or angle (c). The error bars represent the standard error in the mean of the pulls in each bin [26].

components, i.e., the outer steps. As a result, constraining the broader component of the double-gaussian models is impossible. Instead, the super components primarily affect the central part, unaffected when the minimum number of tracks is changed. Figure 4.13 depicts the factorization bias estimated from fits with the SG, DG, SupG, and SupDG fit models for a specific BCID. The corrections to σ_{vis} are almost unaffected by the number of tracks. In the limit of many tracks, the DG (SupDG) produces the same results as the SG (SupG), as

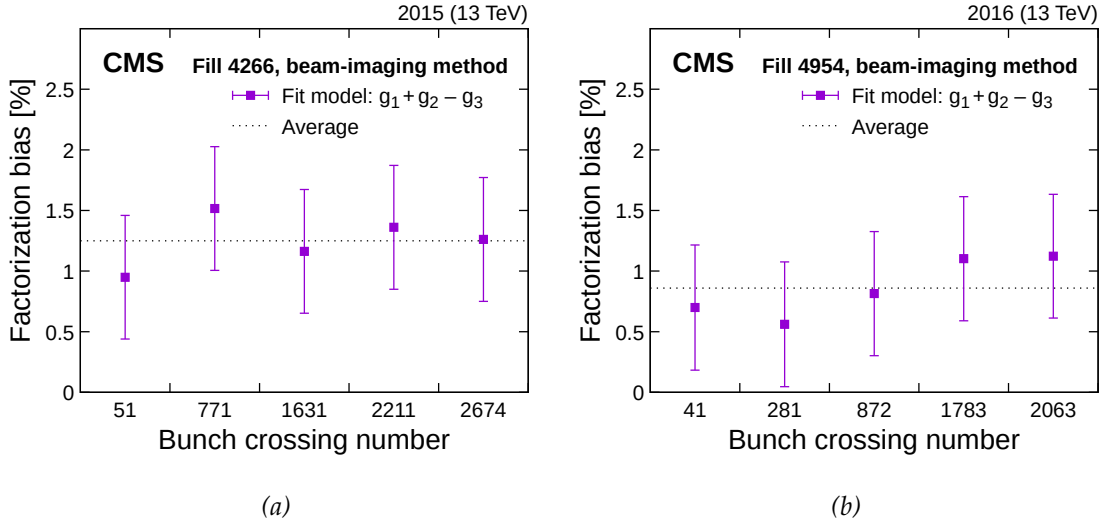


Figure 4.9: Factorization bias estimated from the fits to the beam imaging data in Fills 4266 (a) and 4945 (b). The error bars represent statistical and systematic sources of uncertainty that were added in quadrature [26].

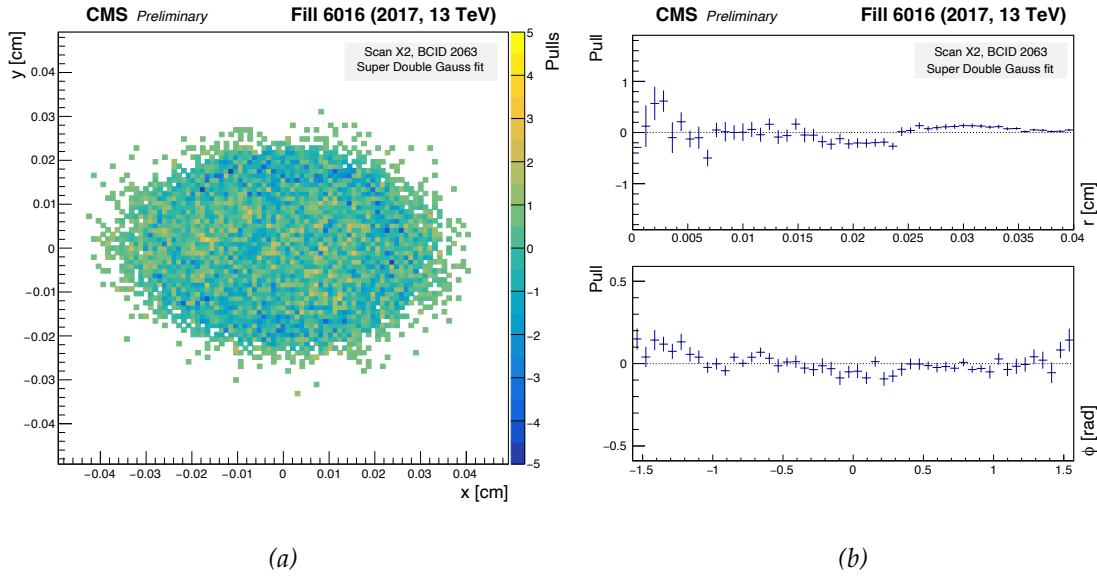


Figure 4.10: Example of the pull distributions of the super-double-Gaussian fit model with respect to the vertex distribution recorded in a fill in 2017. (a) The plot depicts the two-dimensional pull distributions. (b) The plots show the 1-D per-bin pulls projections averaged over the same radial distance (upper) or azimuthal angle (bottom) [27].

the sensitivity to the tails of the shapes is lost. The corrections associated with the SupDG are more scattered due to the little flexibility and complexity of the fit model. Considering the increase in the errors associated with the corrections, changing the requirements for the number of tracks has almost no effect on the magnitude of the factorization bias.

Another constraint is the processing time required to perform the currently implemented fits: more sophisticated models based on Gaussian functions cannot be executed in a reasonable CPU time. To overcome the bottleneck due to runtime, state-of-the-art techniques have been employed to build a new complementary framework using TENSORFLOW [235] and KERAS [236] and also running on graphics processing unit (GPU) machines. The four

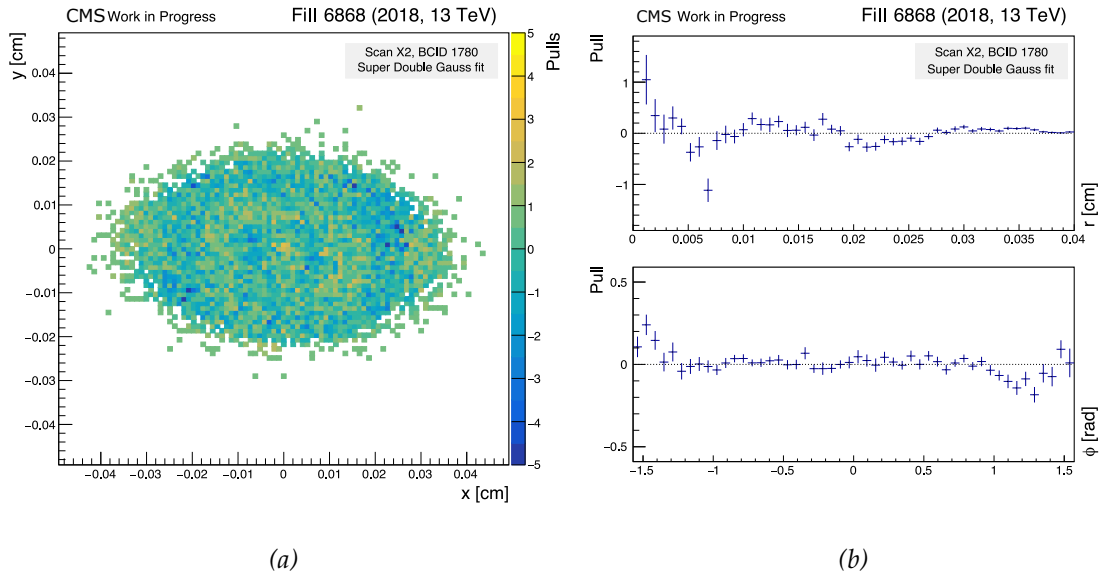


Figure 4.11: Example of the pull distributions of the super-double-Gaussian fit model with respect to the vertex distribution recorded in a fill in 2018. (a) The plot depicts the two-dimensional pull distributions. (b) The plots show the 1-D per-bin pulls projections averaged over the same radial distance (upper) or azimuthal angle (bottom).

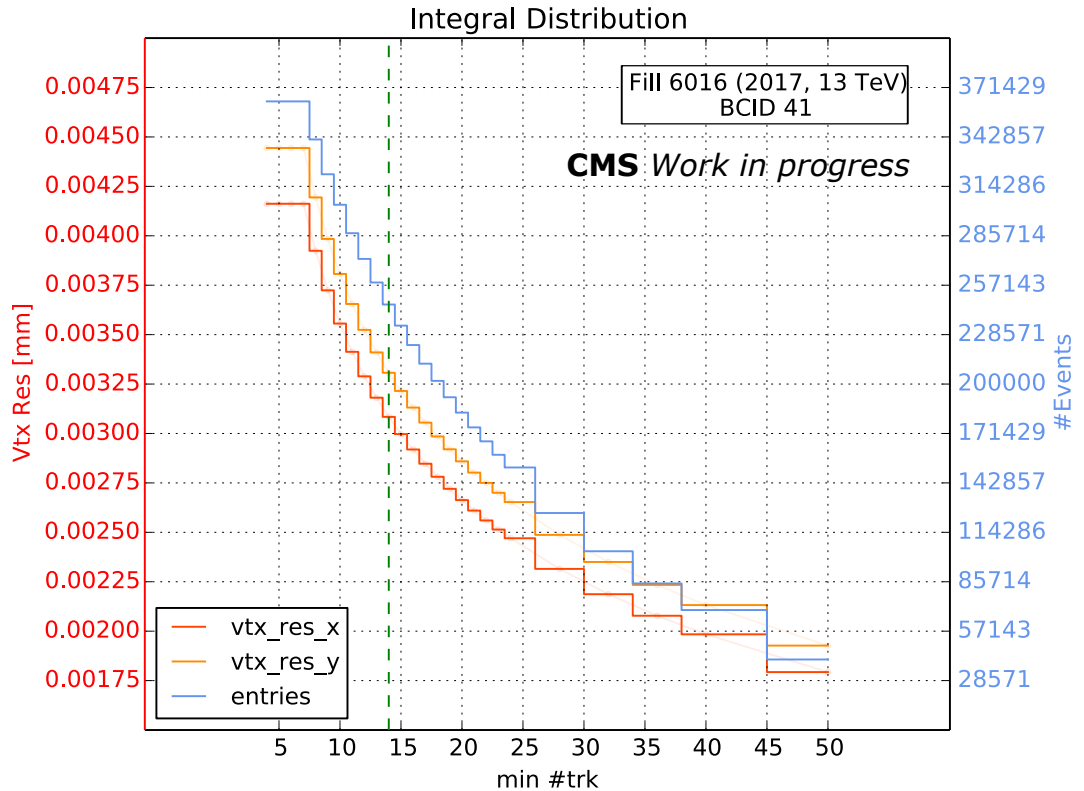


Figure 4.12: Distribution of the vertex resolution in the x - and y -axis and of the number of events as a function of the minimum number of tracks associated with the reconstructed vertices used in the beam image analysis for a fill in 2017.

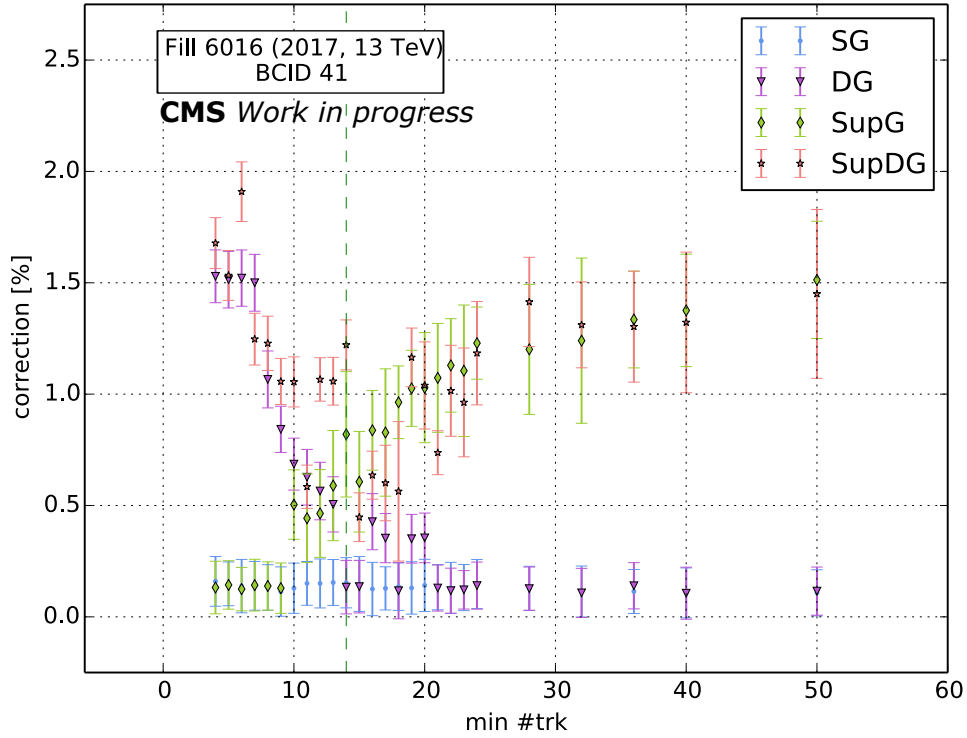


Figure 4.13: Distribution of the corrections to σ_{vis} for different fit models as a function of the minimum number of tracks associated with the reconstructed vertices used in the beam image analysis for a fill in 2017.

distributions with the collected vertex position from all steps of the four beam-imaging scans are the inputs to a custom layer. The analytical fit function is applied, returning the physics parameters of interest as trainable weights tensors. In this two-dimensional binned curve fitting approach with TENSORFLOW, the loss function mean squared error is responsible for providing the best fit to the data. Major improvements have also been achieved in the runtime for each fit model, especially for the more complex model. Figures 4.14 and 4.15 show the pull distributions for the 2017 and 2018 data, also angular and radial projections. The major features mimic what had been found with the previous method on Figures 4.10 and 4.11, but the projections here appear flatter overall.

At the time of writing this thesis, an effort is going to validate the method, compute the corrections to σ_{vis} and perform additional studies to address statistical and systematic uncertainties.

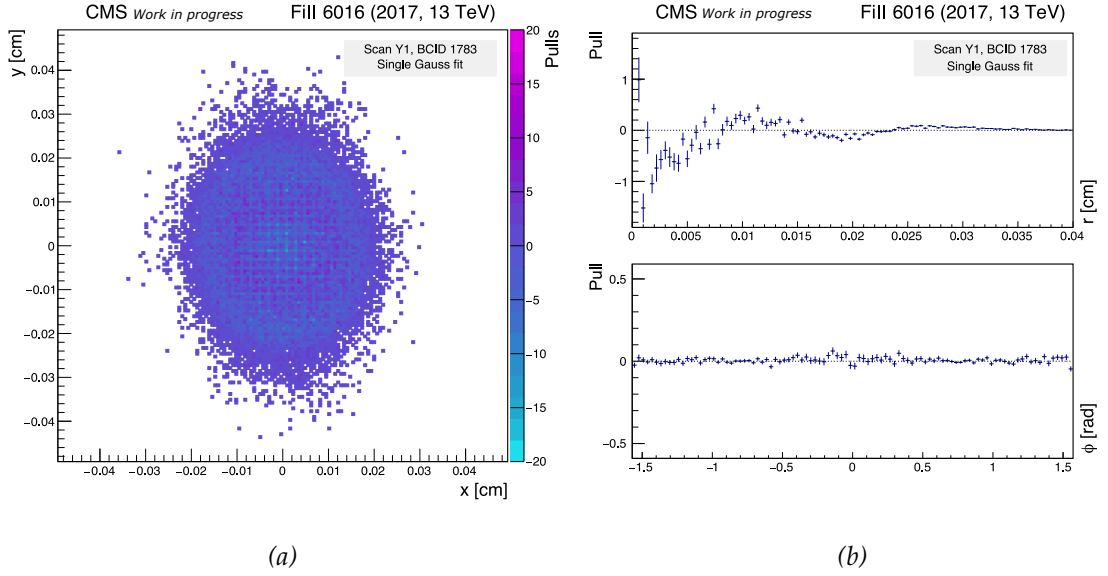


Figure 4.14: Example of the pull distributions of the super-double-Gaussian fit model with respect to the vertex distribution recorded in a fill in 2017 using the complementary framework with TENSORFLOW. (a) The plot depicts the two-dimensional pull distributions. (b) The plots show the 1-D per-bin pulls projections averaged over the same radial distance (upper) or azimuthal angle (bottom)

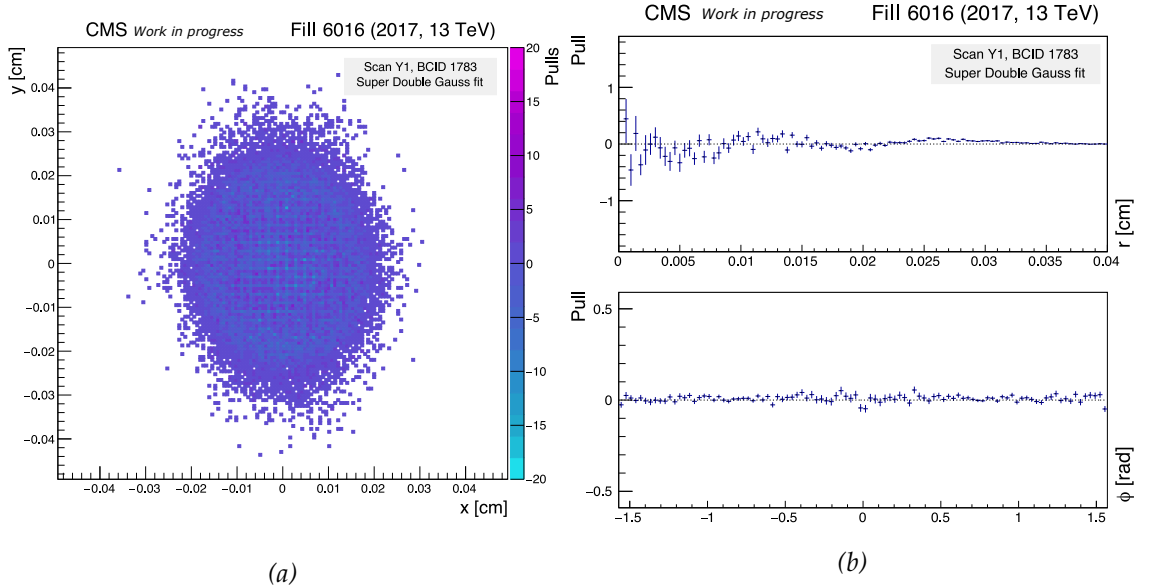


Figure 4.15: Example of the pull distributions of the super-double-Gaussian fit model with respect to the vertex distribution recorded in a fill in 2018 using the complementary framework with TENSORFLOW. (a) The plot depicts the two-dimensional pull distributions. (b) The plots show the 1-D per-bin pulls projections averaged over the same radial distance (upper) or azimuthal angle (bottom).

MEASUREMENT OF $t\bar{t}H$, $H \rightarrow b\bar{b}$ - ANALYSIS STRATEGY

5.1	Introduction	75
5.2	Event samples	76
5.2.1	Data samples	77
5.2.2	Simulation samples	77
5.3	Event selection	83
5.4	Corrections to the event simulation	87
5.5	$t\bar{t}$ + jets background model	92
5.6	Event categorization in the analysis regions	99
5.7	Observables for the parameter estimation	101
5.7.1	Input variables to the neural networks	103
5.7.2	Neural network architecture and training	105
5.7.3	Evaluation of training results	110
5.7.4	The other channels	112

This chapter presents the strategy applied to search for the $t\bar{t}H, H \rightarrow b\bar{b}$ process in the dilepton final states. The $t\bar{t}H, H \rightarrow b\bar{b}$ process, in which the Higgs boson is produced in association with a pair of top quarks, is a key process for studying the properties of the Higgs boson and its interactions with other particles. The chapter begins with an introduction to the physics process and corresponding final states, followed by a description of the recorded and simulated data sets used in the analysis. To ensure a good modeling of the data, the recorded data and simulated events are selected following certain quality requirements imposed on the physics objects. This is further discussed, including control distributions comparing data and simulation. The chapter then describes the event selection and categorization procedures used to enhance the purity of the signal. These procedures include the use of kinematic variables and machine learning techniques to discriminate the signal from the background. Finally, the strategy and the construction of the final observables to enhance the analysis sensitivity are presented, with a focus on validation of the input features to the neural network.

5.1 Introduction

The production of a Higgs boson in association with a top-quark-antiquark pair is particularly interesting for testing the SM and the Higgs mechanism. Experimental measurements of the $t\bar{t}H$ process are often performed separately based on the decay modes of the Higgs boson and the $t\bar{t}$ system, mainly due to experimental reasons, as they address different final states and require dedicated selection criteria. The Higgs boson decay mode relevant for this thesis is $H \rightarrow b\bar{b}$. As the top quark decays almost exclusively into a W boson and a bottom quark, the decay modes of the W boson into quarks or into a lepton and a neutrino identify the decay modes of the $t\bar{t}$ system, as discussed in Section 1.3.2 and illustrated in Figure 1.16.

This thesis focuses on DL $t\bar{t}$ decay channel, characterized by two highly energetic, isolated, oppositely-charged leptons (e^+e^- , $\mu^+\mu^-$, $e^\mp\mu^\pm$). Although this channel has a relatively small branching ratio, it benefits from significantly lower backgrounds than the other $t\bar{t}$ decay channels. Additionally, events in which the leptonic W boson decays into τ leptons are considered signal events if all τ leptons in the event further decay to electrons or muons; otherwise they are considered as background in this analysis.

The selection of $t\bar{t}H$ event candidates with $H \rightarrow b\bar{b}$ and $t\bar{t}$ decaying leptonically is designed to achieve high signal purity by imposing requirements on the physics objects that characterize the final state. This is challenging in practice, since the measured objects are subject to detector effects such as inefficiencies, limited acceptance, or incorrect identification of the actual objects. Moreover, the sensitivity of the analysis is limited by a combinatorial background due to multiple bottom quark jets in the final state, that does not allow to reconstruct the invariant mass peak of the Higgs boson in an unambiguous way. Therefore, the measurement strategy aims to increase the analysis sensitivity by categorizing events and constructing observables that exploit features in the underlying physics processes to further separate signal- and background-enriched phase space regions. Sophisticated analysis techniques, in particular, Artificial Neural Networks (ANN) are employed in this work for the first time to accomplish this separation.

To ensure good scientific practice and guarantee unbiased progress in the analysis of the data, a practice so called blinding procedure is applied. This is done by not initially examining the recorded data in the analysis region. Instead, an artificial data set is used for signal extraction. This data set is generated through simulation and all observables are estimated from it [237]. Once the agreement between the simulated data and observed data has been validated, the analysis is unblinded and a fit to the actual data is performed to estimate the parameter of the signal-strength modifier.

In this analysis, the signal is extracted in a binned maximum likelihood fit to the expected distribution of discriminating observables for both signal and background to the data across all categories, taking into account any correlations among corresponding processes. Systematic uncertainties are handled by incorporating nuisance parameters with appropriate constraints [238, 239], as discussed later in Section 6.1, and are correlated among processes, categories and data-taking periods as needed. The presence of different background compositions in the categories helps to constrain these uncertainties and improves the overall sensitivity. The focus of this analysis is to measure the $t\bar{t}H$ production rate, both inclusively and in different regions of Higgs boson p_T using the Simplified Template Cross Section approach [21] (presented in Chapter 7).

5.2 Event samples

As in almost every physics analysis, this work makes a wide usage of simulated events to develop the various procedures and methods used to extract the physics results. Simulated samples generated using Monte Carlo methods (see Section 1.2.2) are primarily used to parameterize detector acceptance and resolution effects and to derive the signal model, but are also relevant for studying the different signal and background processes and optimizing the analysis strategy accordingly.

The data samples used in this analysis, together with their simulated equivalents, correspond to p-p collision events collected by the CMS detector during the Run 2 phase of the LHC from 2016 to 2018 at $\sqrt{s} = 13$ TeV. Separate simulated samples are produced for each year of data taking with simulation parameters tuned to each year's particular conditions where possible. This section introduces the samples of measured and simulated events used in this analysis to search for $t\bar{t}H$, $H \rightarrow b\bar{b}$.

5.2.1 Data samples

The analysis is performed using the p – p collision data at $\sqrt{s} = 13\text{TeV}$ recorded from 2016 to 2018 with the CMS detector at the LHC at $\sqrt{s} = 13\text{TeV}$. The full data set corresponds to a total integrated luminosity of 137.60 fb^{-1} with per-year luminosity as listed in Table 5.1.

Year	Integrated luminosity
2016	36.33 fb^{-1}
2017	41.53 fb^{-1}
2018	59.74 fb^{-1}
Total	137.60 fb^{-1}

Table 5.1: Integrated luminosity per year and total [139].

Only data recorded in stable detector conditions and with optimal quality by CMS are taken into account in the analysis presented here.

5.2.2 Simulation samples

Dedicated tools are used to simulate events according to the theory prediction at different degrees of accuracy for each step of the simulation chain. Generally, the events for signal and the expected background contributions are generated at NLO of perturbation theory, as detailed in the following, depending on the process. The value of the Higgs boson mass m_H is set to 125 GeV and the top quark mass m_t is assumed to be 172.5 GeV. The simulation of the detector response is performed with GEANT4 [102]. Between 2016 and 2017, the tracker system has been upgraded, adding a layer; thus, the detector response in the simulation for 2016 differs from the other two years.

The probability density function of the partons in the initial-state protons are provided by the NNPDF3.1 set [103] as described in Section 1.2.2. In all simulated samples, the parton showering and the hadronization processes are simulated with PYTHIA8. They rely on a phenomenological model which entails tuneable parameters: the energy scale used for the matching of the ME and the parton-showering diagrams (ME-PS matching scale hdamp), and the ones used to describe the underlying events of the collisions (UE tune) are among these. The parameters for the underlying event description correspond to the CP5 tune [240] for all 2017 and 2018 samples, and for most of the relevant MC samples of 2016; for few background samples of 2016, the CUETP8M1 [241] tune is used (see Table 5.3).

To improve the accuracy of the simulation, a reweighing process is applied. This involves adjusting the weight of the events generated by the MC generators, which may not reproduce the state-of-art calculations of the theoretical cross sections. The reweighing considers any discrepancies in the theoretical cross sections used in the simulation and any differences between the simulated detector and actual experimental conditions. By doing so, the reweighing provides a more accurate representation of the data, leading to more meaningful comparisons and analysis.

The events in the simulated samples are normalized to the same integrated luminosity as the data sample, using event weights that include best known theoretical prediction for the SM cross section of the generated process and its branching ratio $(\sigma_p \times \mathcal{B})_{\text{Th., SM}}$, the integrated luminosity $\mathcal{L}$, and a per-event weight $w_{\text{gen},i}$ of the MC generator used. Thus, the

normalization for an event i is given by:

$$w_{p,i} = \frac{(\sigma_p \times \mathcal{B})_{\text{Th., SM}} \cdot \mathcal{L} \cdot w_{\text{gen},i} \cdot w_{\text{corr}} \cdot SF}{\sum_j^{N_{p,\text{gen}}} w_{\text{gen},i}}, \quad (5.1)$$

where the denominator sums over all $N_{p,\text{gen}}$ generator event weights so to get the expected yields for each process. Additional dedicated contributions as w_{corr} are introduced to match the MC to the observed data, like the L1 trigger primitives pre-firing issue described in Section 2.2.7, and as per-event scale factor SF , like for leptons' efficiency. More details about these additional contributions are described in Section 5.4. The specifics of the MC generators utilized to simulate the physical processes involved in the analysis, and the corresponding values for $(\sigma_p \times \mathcal{B})_{\text{Th., SM}}$ are detailed in the following sections, covering both signal and background processes.

Signal processes

The focus of this analysis is on Higgs bosons decaying into a pair of bottom quarks, $H \rightarrow b\bar{b}$, which is the dominant decay mode ($\mathcal{B} = 58\%$ for $m_H = 125$ GeV as introduced in Section 1.3.1). The hadronization of bottom quarks yields B hadrons, whose long lifetime makes them travel a measurable flight distance before decaying (see Section 1.3.1). A secondary vertex can be identified by reconstructing and extrapolating trajectories of the charged decay products (see Section 3.2.8). In the majority of the cases, the two b jets originating from the hadronization of bottom quarks should have a sufficiently large spatial separation allowing for their resolved measurement and identification. This allows to distinguish b jets from jets of other flavors.

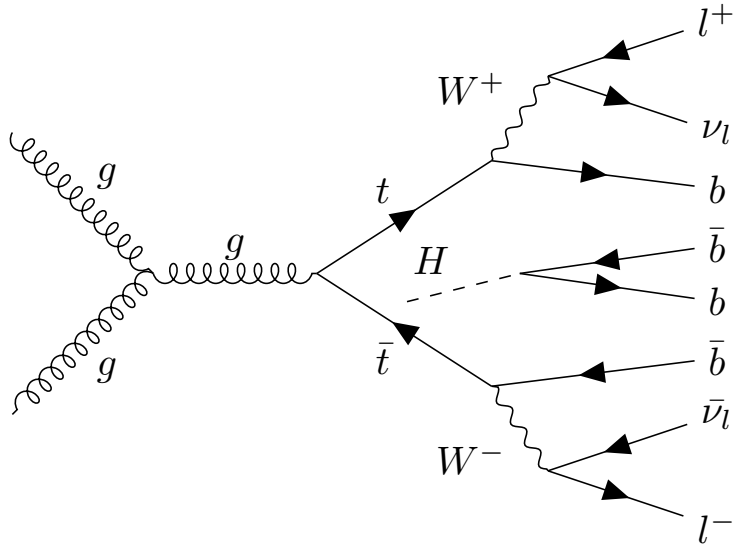


Figure 5.1: Leading-order Feynman diagram for the process $t\bar{t}H, H \rightarrow b\bar{b}$ in the DL channel, with $\ell = e, \mu$ being the leptons considered for the scope of this analysis.

Figure 5.1 shows as example a leading-order Feynman diagram for $t\bar{t}H, H \rightarrow b\bar{b}$ in the DL decay channel. More diagrams exist to describe $t\bar{t}H$ production, with gluon-initiated processes having the largest contribution to the total $t\bar{t}H$ cross section at $\sqrt{s} = 13$ TeV. The final state of interest features four b jets, two isolated leptons with opposite charge and a

significant amount of missing transverse momentum due to the presence of non-detectable neutrinos. Overall, the full reconstruction of the event is challenging due to the high jet multiplicity. The DEEJET algorithm (introduced in Section 3.2.8) used to identify the b jets has a non-negligible misidentification probability for the different quark flavors. Therefore, small contributions from other Higgs boson decay channels are expected. In order to correctly take into account these contributions, all decay products of the Higgs boson are considered as signal in the scope of this work.

Signal samples

The $t\bar{t}H$ signal process is simulated with POWHEG at NLO accuracy [98, 99, 242] and it is split according to the decay channel of the Higgs boson into $b\bar{b}$ pairs or any other decay. The inclusive $t\bar{t}H$ cross section is normalized to 507^{+35}_{-50} fb taken from calculations at NLO QCD and NLO EWK accuracy [21].

For all Higgs boson MC samples, the cross sections and branching ratios have been rescaled to the most recent and precisely measured Higgs boson mass in the diphoton channel of $m_H = 125.38 \pm 0.14$ GeV [107], using the correction factors given in Table 5.2.

	Value at $m_H = 125$ GeV	Scale Factor
$t\bar{t}H$ XS	0.507 pb	0.9928
tHW XS	0.015 pb	1.0000
tHq XS	0.077 pb	0.9987
$\mathcal{B}_{H \rightarrow b\bar{b}}$	0.5824	0.9896
$\mathcal{B}_{H \rightarrow W^+W^-}$	0.2137	1.0293
$\mathcal{B}_{H \rightarrow gg}$	0.0819	0.9953
$\mathcal{B}_{H \rightarrow l\bar{l}}$	0.0627	0.9903
$\mathcal{B}_{H \rightarrow c\bar{c}}$	0.0289	0.9898
$\mathcal{B}_{H \rightarrow ZZ}$	0.0262	1.0352
$\mathcal{B}_{H \rightarrow \gamma\gamma}$	0.0023	1.0000
$\mathcal{B}_{H \rightarrow Z\gamma}$	0.0015	1.0209

Table 5.2: Rescaling factors used to update the production cross section (XS, upper block) and branching ratios (lower block) to a Higgs boson mass of $m_H = 125.38 \pm 0.14$ GeV [107]. All values are taken from [21, 243, 244].

Main background processes

In the context of the $t\bar{t}H, H \rightarrow b\bar{b}$ measurement, $t\bar{t}$ production in association with additional jets ($t\bar{t} + \text{jets}$) is the dominant background contribution, as was mentioned in Section 1.3.3. This background is further divided in different sub-processes according to the flavor of the additional jets, determined at generator level via ghost-hadron matching [245–247]. Among these, $t\bar{t}$ events with additional b-jet production are particularly challenging processes, both from the experimental and the theoretical point of view. An example of a LO Feynman diagram for $t\bar{t}$ production in association with two bottom quarks ($t\bar{t} + b\bar{b}$) is shown for the DL channel in Figure 5.2. From the experimental side, the final state is identical to that of the $t\bar{t}H, H \rightarrow b\bar{b}$ signal process, and both processes have very similar kinematic properties. This makes it impossible to define a pure $t\bar{t} + b\bar{b}$ control region. For these reasons, the $t\bar{t} + b\bar{b}$

process is an irreducible background for the $t\bar{t}H, H \rightarrow b\bar{b}$ measurement. From the theoretical side, the calculation of the $t\bar{t} + b\bar{b}$ process is challenging because the production of $t\bar{t}$ pairs and the splitting of gluons into $b\bar{b}$ pairs usually does not happen at the same energy scale, ranging from about 4 GeV (bottom quark mass) to about 173 GeV (top quark mass). Due to the nature of such processes, large uncertainties arise from the choice of the renormalization and factorization scales even when calculated at NLO precision and are, therefore, difficult to model precisely. Dedicated measurements [43, 119, 120] and previous iterations of this analysis [121, 127] have shown that the $t\bar{t} + b$ jets contribution in data is larger than expected by the simulation by a factor of 1.2–1.4 depending on the generator. Overall, the control over this process is limited, and advanced techniques are required both in developing an analysis strategy (i.e., defining signal- and background-enriched phase space regions) and in statistical data analysis (i.e. assigning properly the different theoretical uncertainties).

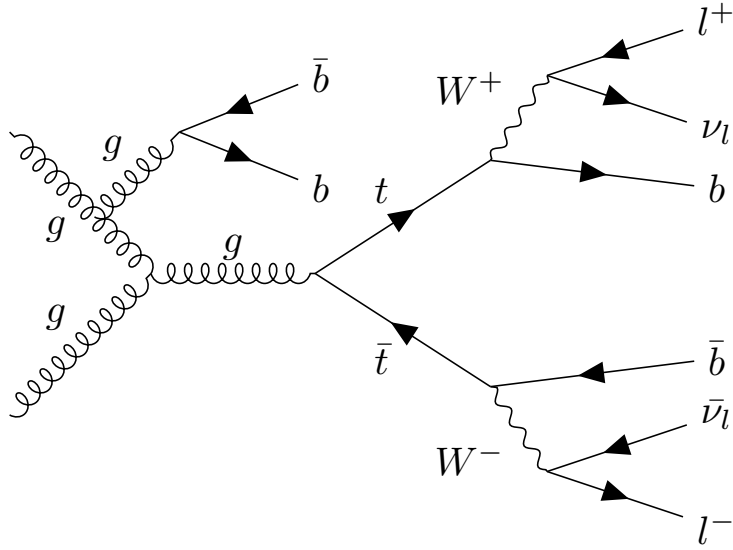


Figure 5.2: Leading-order Feynman diagram for the $t\bar{t} + b\bar{b}$ process in the DL channel, with $\ell = e, \mu$ being the leptons considered for the scope of this analysis.

The different $t\bar{t} + \text{jets}$ subprocesses are defined using information from the event simulation. At simulation level, after ME generation, parton showering and hadronization processes (see Section 1.2.2), the final state particles are grouped into jets with the anti- k_T algorithm and a distance parameter $R=0.4$. All jets associated to the hard scattering process, i.e. originating from the top quark or hadronic W boson decays, or with a transverse momentum below 20 GeV are rejected. The remaining jets are split into five mutually exclusive $t\bar{t}$ processes at simulation level according to the number of additional jets and B and C hadrons, $t\bar{t} + X$, namely:

- **$t\bar{t} + b\bar{b}$ (resolved)** Events with at least two additional jets, each containing at least one B hadron within the acceptance of the applied selection;
- **$t\bar{t} + b$** Events with only one additional jet inside the acceptance of the applied selection and containing exactly one B hadron;
- **$t\bar{t} + 2b$** Events with one additional jet with two B hadrons clustered into the jet;

- $t\bar{t} + c\bar{c}$ (also labeled as $t\bar{t} + C/t\bar{t} + c\bar{c}$) None of above criteria stands, but the events have (at least) one additional jet containing (at least) one C hadron;
- $t\bar{t} + \text{LF}$ (also labeled as $t\bar{t} + \text{lf}$) None of above criteria applies.

The first four processes involve additional heavy-flavor jets, $t\bar{t} + \text{HF}$. On the contrary, the $t\bar{t} + \text{LF}$ are additional jets originating from other quark flavors (uds) as well as additional gluon emissions. Figure 5.3 schematically represents the $t\bar{t} + b\bar{b}$ sub-processes, based on resolution and acceptance effects in the detector. In this thesis, the first three processes are treated as a single process referred to as $t\bar{t} + b\bar{b}$ or $t\bar{t} + B$. At this stage no b-tagging algorithm is employed and this categorization holds only for simulated events. The calculation of the additional collinear gluon splitting effects in the $t\bar{t} + 2b$ subprocess is particularly challenging, and a dedicated source of systematic uncertainty is determined to account for it (see Section 6.1).

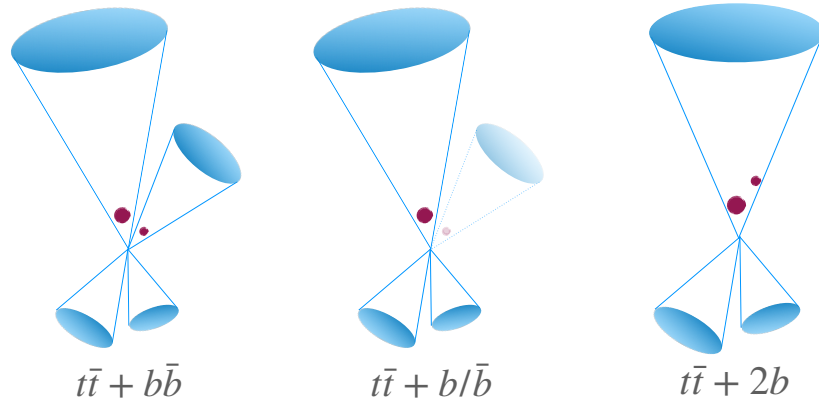


Figure 5.3: Classification of the $t\bar{t} + b\bar{b}$ sub-processes defined in the text. The cones represent reconstructed jets, while the purple dots are the bottom quarks in the system. Figure adapted from Ref. [248].

The $t\bar{t} + c\bar{c}$ process should be easily distinguished from the $t\bar{t}H, H \rightarrow b\bar{b}$ process, which requires the existence of b jets in the final state. However the b-tagging misidentification rate for charm quarks is not negligible (see Section 3.2.8). Similarly for the $t\bar{t} + \text{LF}$ process: even though the mistag probability in this case is considerably lower, the large branching ratio for these process makes their contribution significant in the analyzed phase space.

Main background samples

The $t\bar{t}$ background processes are simulated using POWHEG. The inclusive cross section is normalized to $831.76^{+19.77}_{-29.20}$ (scale) $^{+35.06}_{-35.06}$ (PDF + α_s) pb, which has been calculated at next-to-next-to leading order (NNLO) in QCD including resummation of next-to-next-to-leading logarithmic (NNLL) soft gluon terms with $\text{top}++2.0$ [29–35].

The $t\bar{t}$ events are categorized based on the flavor of additional jets, as detailed above. Specifically, irreducible background process events with additional b jets, are generated using the POWHEG MC generator with NLO accuracy in the 4FS. This is the preferred method as it better describes the hard interaction processes in the generation of the $t\bar{t} + b\bar{b}$ process. The LHC Higgs Cross Section Working Group has studied this simulation in detail in the scope of the full Run 2 $t\bar{t}H, H \rightarrow b\bar{b}$ analysis [42]. The other $t\bar{t} + \text{jets}$ events ($t\bar{t} + c\bar{c}$, $t\bar{t} + \text{LF}$) are generated using a POWHEG $t\bar{t}$ simulation at NLO accuracy in the 5FS, with only one of the additional jets described by the ME calculation. Details of the MC configurations can be found in Appendix C, and a more detailed description is in Section 5.5.

In the POWHEG samples, a damping parameter value of $h_{\text{damp}} = 237.9$ GeV has been used together with the CP5 tune, and a value of $h_{\text{damp}} = 172.5$ GeV together with the CUETP8M1 tune. The values are summarized in Table 5.3 for the different simulated samples. The uncertainties assigned as a result of the factorization and renormalization scales, as well as the PDF, are also listed in Section 6.1.

Year	Sample	UE Tune	ME-PS matching
2016	V + jets, $t\bar{t}$ + V, Diboson	CUETP8M1	172.5 GeV
	all others	CP5	237.9 GeV
2017	all	CP5	237.9 GeV
2018	all	CP5	237.9 GeV

Table 5.3: Parton-shower tune and damping-parameter values used in the simulation with PYTHIA 8.2.

The samples are merged to cover the entire $t\bar{t}$ + jets phase space: the $t\bar{t} + b\bar{b}$ 4FS prediction is normalized to the corresponding fraction that contributes to the inclusive $t\bar{t}$ cross section in the 5FS. The differential information for the $t\bar{t} + b\bar{b}$ process is obtained from the 4FS prediction, while the precise NNLO + NNLL inclusive $t\bar{t}$ cross section is required not to change. The $t\bar{t} + b\bar{b}$ and $t\bar{t} + c\bar{c}$ normalizations are left as a freely-floating parameter in the final fit to be determined in-situ to allow adjusting the predicted $t\bar{t}$ + jets yield to that observed in data.

Minor background processes and samples

In addition to the background processes from $t\bar{t}$ + jets production, there are several other physics processes that can mimic the $t\bar{t}H, H \rightarrow b\bar{b}$ final states and therefore contribute to the event yield in the analysis selection region. These are referred to as minor backgrounds and are discussed in the following.

The tH processes produces the same final state when additional radiation is reconstructed as part of the $t\bar{t}$ system. This process is split into the tHq and tHW processes to allow for further measurements like the CP properties of the Higgs boson and the sign of the top-Higgs-Yukawa coupling. The tHq and tHW processes are simulated with MADGRAPH5_AMC@NLO at LO accuracy. The tHq sample is generated in the 4FS, allowing for a more accurate description of the extra bottom quark from the initial-state gluon splitting. The tHW is instead modeled in the 5FS since it is so far the only way to accurately account for the interference effects between the 4FS prediction at LO and the 5FS prediction at NLO during the reweighting process. The inclusive cross sections are normalized to NLO accuracy in QCD (no NLO EWK corrections are available), computed in the 5FS [243, 249, 250]. The tH processes have little effect in the fiducial phase space of the $t\bar{t}H, H \rightarrow b\bar{b}$ analysis, thus they are actually considered as minor backgrounds. The contribution from s-channel tH production is negligible and therefore not considered. As for the $t\bar{t}H$ samples, all decays channels of the Higgs boson are taken into account.

The production of single top (anti)quark, single t , can also mimic the relevant final state. The overall production cross section is $O(400)$ times larger than the $t\bar{t}H$ one [251], so although the selection of this process is less likely, the contribution to the final phase-space is not negligible. The single-top background processes in the t - and tW -channels are simulated with POWHEG [252, 253], while MADGRAPH5_AMC@NLO is used to simulate the s-channel single top quark processes. The inclusive cross sections are normalized to approximate NNLO (tW channel [254]) and NLO (t - and s-channels [255, 256]).

Processes with a top quark pair produced in association with vector bosons (W, Z, γ) are also relevant for this analysis and referred to as $t\bar{t}V$. The cross sections for these processes are of the same order of magnitude as the $t\bar{t}H$ one [257], but the CKM matrix suppresses the relevant decays. Processes with a single W or Z/γ^* vector boson and extra jet emission can also resemble the signal, especially for decays into charged leptons; these are referred to as V +jets and their contributions are severely reduced with increasing of jet multiplicity. Another contribution comes from vector boson pair production (WW, WZ and ZZ), here referred to as diboson production. Specifically, these processes yield a similar final state as $t\bar{t}H, H \rightarrow b\bar{b}$ when the Z boson decays into bottom quark pairs and the light-flavor jets from the W boson are misreconstructed as b jets.

The V +jets and $t\bar{t}V$ processes are simulated with MADGRAPH5_AMC@NLO, where the matching of ME jets to PS is performed using the FxFx [258] prescription. Diboson production is simulated using the PYTHIA event generator. The inclusive cross sections are normalized to NNLO (WW and all V +jets, except Z +jets 2016 [259]), to NLO+NNLL QCD and NLO EWK ($t\bar{t}+V$ [257]), and to NLO (other diboson [260] and Z +jets 2016) accuracy.

A list of the cross section used for the normalization of the signal and background processes employed in this thesis is presented in Table 5.4, calculated at $\sqrt{s} = 13$ TeV and assuming $m_t = 172.5$ GeV, $m_W = 80.398$ GeV and $m_Z = 91.1876$ GeV [55], if not stated otherwise.

Tables 5.5, 5.6 and 6.5 provide a summary of the main simulated samples, as well as the generated number of events and the naming and color schemes used for the plots in this analysis.

5.3 Event selection

Generally, in a physics analysis, the selection of events has the goal of identifying the particular signal process under study, in this case, the $t\bar{t}H, H \rightarrow b\bar{b}$ signature. This can be achieved by exploiting differences in particle composition as well as kinematic features between signal and background events. Furthermore, the selection criteria should also aim to reduce the impact of systematic uncertainties.

When searching for the production of $t\bar{t}H$ events with $H \rightarrow b\bar{b}$ in the DL decays of the $t\bar{t}$ system, the final state consists of four b jets, two oppositely charged electrons or muons, and a considerable quantity of missing transverse momentum arising from the undetected neutrinos. The main background is due to the top quark pairs production with additional heavy flavor jets, whose events cannot be fully separated from the signal ones with only selection cuts in the phase space, as already discussed above in Section 5.2.2.

The most relevant objects for this thesis are electrons, muons, reconstructed jets and the missing transverse momentum originating from the neutrinos. These objects and their quality requirements are described in Section 3.2. Only events that pass the CMS trigger requirements are considered. In this analysis, events are selected by trigger paths requiring in the final state one or two isolated leptons (e, μ) with transverse momenta above predefined thresholds. The dielectron (dimuon) trigger checks for the presence of two electrons (muons) with one electron (muon) satisfying $p_T > 23$ (17) GeV and the other $p_T > 12$ (8) GeV. The $e^\mp \mu^\pm$ -trigger instead requires either an electron (muon) with $p_T > 23$ (23) GeV and muon (electron) with $p_T > 8$ (12) GeV. The dilepton trigger is not 100% efficient, therefore to increase the selection efficiency, events are selected involving also the complementary single lepton triggers. The requirements on the threshold values span from 24 to 27 GeV for muons, and from 27 to 32 GeV for the electrons, depending on the data-taking period and conditions. In order to compensate for the increased p_T threshold of the single-electron trigger in 2017 and 2018, a cross trigger requiring an electron with $p_T > 28$ GeV and the p_T sum of all jets $H_T > 150$ GeV

Process	Channel/phase space		Cross section	
$t\bar{t}H$	$H \rightarrow b\bar{b}$	$295.3^{+5.8\%}_{-9.2\%}(\text{scale}) \pm 3.6\% \text{ (PDF) fb}$	(NLO QCD and NLO EWK)	
	$H \rightarrow b\bar{b} \text{ DL}^1$	31.4 fb	(NLO QCD and NLO EWK)	
	$H \rightarrow b\bar{b} \text{ SL}^2$	129.8 fb	(NLO QCD and NLO EWK)	
	$t\bar{t}H, H \not\rightarrow b\bar{b}$	$211.8^{+5.8\%}_{-9.2\%}(\text{scale}) \pm 3.6\% \text{ (PDF) fb}$	(NLO QCD and NLO EWK)	
tH	tHq	$74.25^{+6.6\%}_{-14.9\%}(\text{scale}) \pm 3.6\% \text{ (PDF) fb}$	(NLO)	
	tHW	$15.17^{+4.9\%}_{-6.7\%}(\text{scale}) \pm 6.1\% \text{ (PDF) fb}$	(NLO)	
$t\bar{t}$	inclusive 5FS	$831.76^{+19.77}_{-29.20} \text{ (scale)} \pm 35.06 \text{ (PDF) pb}$	(NNLO+NNLL)	
	$t\bar{t} \text{ (SL)}$	365.46 pb	(NNLO+NNLL)	
	$t\bar{t} \text{ (DL)}$	88.34 pb	(NNLO+NNLL)	
	$t\bar{t} \text{ (FH)}$	377.96 pb	(NNLO+NNLL)	
$t\bar{t} + b\bar{b}$	inclusive 4FS	reweighted to $t\bar{t} + b\bar{b} \text{ (5FS)}$		
Single $t/\bar{t}$	s-channel	$W \rightarrow \text{had}$	$6.96^{+0.18}_{-0.15} \text{ (scale)} \pm 0.14 \text{ (PDF) pb}$	(NLO)
		$W \rightarrow \ell\nu$	$3.36^{+0.11}_{-0.09} \text{ (scale)} \pm 0.15 \text{ (PDF) pb}$	(NLO)
	t-channel	t	$136.02^{+4.09}_{-2.92} \text{ (scale)} \pm 3.52 \text{ (PDF) pb}$	(NLO)
		$\bar{t}$	$80.95^{+2.53}_{-1.71} \text{ (scale)} \pm 3.18 \text{ (PDF) pb}$	(NLO)
	tW-channel	$t/\bar{t}$	$35.85 \pm 0.90 \text{ (scale)} \pm 1.70 \text{ (PDF) pb}$	(NNLO)
W + jets	$W \rightarrow \ell\nu$	$62877^{+166}_{-88} \text{ (scale)} \pm 771 \text{ (PDF) pb}$	(NNLO)	
$Z/\gamma^* + \text{jets}$	$Z/\gamma^* \rightarrow \ell\ell$		22635 pb	
	$10 < m_{\ell\ell} < 50$			
	$70 < H_T \leq 100$		175.3 pb (LO)	
	$100 < H_T \leq 200$		147.4 pb (LO)	
	$200 < H_T \leq 400$		40.99 pb (LO)	
	$Z/\gamma^* \rightarrow \ell\ell$	$400 < H_T \leq 600$	5.678 pb (LO)	
		$50 < m_{\ell\ell} < \infty$		
	$600 < H_T \leq 800$	1.367 pb (LO)		
	$800 < H_T \leq 1200$	630.4 fb (LO)		
	$1200 < H_T \leq 2500$	151.4 fb (LO)		
$H_T > 2500$	3.565 fb (LO)			
Diboson	WW	118.7 pb	(NNLO)	
	WZ	65.54 pb	(NLO)	
	ZZ	15.83 pb	(NLO)	
$t\bar{t} + V$	$W \rightarrow \ell\nu$	179.2 fb	(NLO+NNLL QCD and NLO EWK)	
	$W \rightarrow q\bar{q}$	370.8 fb	(NLO+NNLL QCD and NLO EWK)	
	$Z \rightarrow \ell\ell/\nu\nu$	258.9 fb	(NLO+NNLL QCD and NLO EWK)	
	$Z \rightarrow q\bar{q}$	601.2 fb	(NLO+NNLL QCD and NLO EWK)	
	γ	3.697 pb	(NLO)	

Table 5.4: List of signal and background cross section at $\sqrt{s} = 13 \text{ TeV}$ for the $t\bar{t}H$ events with $H \rightarrow b\bar{b}$ analysis. The dilepton mass $m_{\ell\ell}$ and the sum of transverse jet momenta H_T are given in GeV. Uncertainties denoted by PDF include α_s variations.

Sample	Number of Events			Label	Color
	2016	2017	2018		
$t\bar{t}H, H \rightarrow b\bar{b}$	10 000 000	8 000 000	12 000 000	signal	
$t\bar{t}H, H \rightarrow b\bar{b}$ SL	10 000 000	9 558 675	10 000 000		
$t\bar{t}H, H \rightarrow b\bar{b}$ DL	10 000 000	9 794 620	10 000 000		
$t\bar{t}H, H \not\rightarrow b\bar{b}$	10 000 000	7 966 779	8 000 000		
tHq	10 000 000	10 000 000	30 000 000	tHq (SM)	
tHW	5 000 000	5 000 000	15 000 000	tHW (SM)	
inclusive $t\bar{t}$ (4FS)				$t\bar{t} + b\bar{b}$ (4FS)	
$t\bar{t}$ SL	110 000 000	159 889 425	114 520 000		
$t\bar{t}$ DL	70 000 000	69 155 808	70 960 000		
$t\bar{t}$ FH	70 000 000	130 797 650	134 808 000		
inclusive $t\bar{t}$ (5FS)				$t\bar{t} + c\bar{c}$	
$t\bar{t}$ SL	6 494 400	5 667 979	6 123 100		
$t\bar{t}$ DL	3 749 700	3 697 000	3 615 000	$t\bar{t} + \text{LF}$	
$t\bar{t}$ FH	6 077 198	6 727 500	5 914 500		

Table 5.5: Number of generated events produced for each year of data taking in the simulated signal and background samples used in this analysis. The label and color columns define the grouping of processes depicted in histograms throughout this thesis.

was also used. The inclusion of single-lepton triggers helps recovering about 10% of DL events. Appendix D lists the trigger paths used for the different years and the specific run ranges, as configured for use both in data and simulation. A logical OR among the different triggers is considered (e.g. the event is accepted if one or more trigger paths are fired). To avoid double counting, only triggers corresponding to the same lepton final state are considered for each dilepton final states by additionally vetoing the remaining trigger paths.

Selection criteria are then applied to the objects of interest to further purify the signal-to-background ratio and identify the phase space of events to be analyzed. The goal is also to ensure that the analysis is performed in a region where the simulation describes with good accuracy the recorded data. The same selection cuts on global event and physics quantities are applied both to data and simulation.

First of all, this analysis targets events containing two leptons in the final states, focusing on isolated electrons and muon and disregarding decays involving τ leptons. The DL channel is constituted by two oppositely charged leptons (e^+e^- , $e^\mp\mu^\pm$, $\mu^+\mu^-$). Events with two isolated leptons can potentially emerge from almost all processes considered in this analysis, so requiring opposite electrical charges reduces significantly some background contributions. In addition, the online and offline content of the event is required to be consistent, e.g. a dimuon event has to be recorded by a single or double-muon trigger. The purity of the isolated leptons should be high enough to reduce fake contributions, but also sufficiently large to retain a reasonable amount of events to analyze. This is usually achieved with cuts on kinematic properties. In case of two leptons a common approach is requiring the leading (i.e., the highest- p_T) lepton to satisfy rather tight selection criteria, and the sub-leading lepton to pass

Sample	Number of Events			Label	Color
	2016	2017	2018		
single t					
s-channel ($W \rightarrow \ell\nu$)	10 000 000	19 846 893	20 000 000	t	
s-channel ($W \rightarrow \text{had}$)	5 000 000	10 000 000	10 000 000		
t-channel	32 000 000	5 982 064	160 000 000		
$\bar{t}$ -channel	18 000 000	3 984 790	80 000 000		
tW-channel	5 000 000	7 945 242	10 000 000		
W-channel	5 000 000	7 977 430	7 745 276		
$t\bar{t}W$				$t\bar{t}V$	
$W \rightarrow \ell\nu$	2 160 168	9 994 543	5 000 000		
$W \rightarrow \text{had}$	833 298	811 306	830 000		
$t\bar{t}Z$					
$Z \rightarrow \ell\nu$	1 992 438	11 141 490	14 000 000		
$W \rightarrow \text{had}$	749 400	750 000	750 000		
$t\bar{t}\gamma^*$	1 474 885	1 607 839	4 691 915		
Diboson				VV	
WW	7 982 180	7 779 630	8 000 000		
WZ	3 997 571	3 928 630	4 000 000		
ZZ	1 988 098	1 925 931	2 000 000		
$W \rightarrow \ell\nu$	5 280 565	4 994 543	4 911 941	V+jets	
$Z/\gamma^* \rightarrow \ell\ell$	35 114 960	39 536 840	86 369 010		
$10 < m_{\ell\ell} < 50$					
$Z/\gamma^* \rightarrow \ell\ell$					
$50 < m_{\ell\ell} < \infty$					
$70 < H_T \leq 100$	9 691 660	9 344 037	10 019 684		
$100 < H_T \leq 200$	11 017 086	11 197 488	11 530 510		
$200 < H_T \leq 400$	9 754 170	10 728 447	11 225 887		
$400 < H_T \leq 600$	10 152 204	10 219 524	9 697 098		
$600 < H_T \leq 800$	8 292 957	8 743 640	8 862 104		
$800 < H_T \leq 1200$	2 673 066	3 114 980	3 138 129		
$1200 < H_T \leq 2500$	596 079	625 517	536 416		
$H_T > 2500$	399 492	419 308	427 951		

Table 5.6: (Continue) Number of generated events produced for each year of data taking in the simulated signal and background samples used in this analysis. The label column defines the groups of processes depicted in histograms throughout this thesis. The dilepton mass $m_{\ell\ell}$ and the sum of transverse jet momenta H_T are given in GeV.

partially loosened one. In particular, the transverse momenta and absolute pseudo-rapidities of leading leptons must have $p_T > 25$ GeV and $|\eta| < 2.4$. To suppress events from heavy-flavor resonances decays and low-mass Drell-Yan processes (i.e, photons decaying into pairs of leptons), the invariant mass of selected lepton pair $m_{\ell\ell}$ is required to be greater than 20 GeV. Moreover, same-flavor events (e^+e^- , $\mu^+\mu^-$) with $76 \text{ GeV} < m_{\ell\ell} < 106 \text{ GeV}$ are discarded to reject high-mass Drell-Yan processes. Effectively, these criteria suppress contributions from Z+jets and diboson processes. The $t\bar{t}H$ signal final state features four jets in the DL channel originating from decays of B hadrons. The majority of background processes, excluding the $t\bar{t} + b\bar{b}$ contribution, exhibit fewer jets or b tags and must be suppressed by selection cuts on their multiplicities. An event is accepted when it contains at least 2 jets (and of which one has to be b tagged) with $p_T > 30$ GeV and $|\eta| < 2.4$. Finally a cut on the amount of missing transverse momentum is applied to suppress background contributions that are not induced by processes involving W bosons originating from top quark decays. Dilepton events must have $p_T^{\text{miss}} > 40$ GeV to suppress sufficiently events from pure QCD multijet backgrounds, which are difficult to simulate with an adequate amount of statistics due to its high cross section and low selection efficiency. Table 5.7 provides a summary of all selection criteria, referred to as baseline selection, and exhibiting a good level of agreement between data and the total MC yields.

Observable	Selection
Number of leptons	2
Sign and flavor of leptons	e^+e^- , $e^\mp\mu^\pm$, $\mu^+\mu^-$
Min. p_T of p_T -leading electron [GeV]	25
Min. p_T of p_T -leading muon [GeV]	25
Min. p_T of p_T sub-leading lepton [GeV]	15
Max. $ \eta $ of leptons	2.4
Min. number of jets	2
Min. p_T of jets [GeV]	30
Max. $ \eta $ of jets	2.4
Min. number of b-tagged jets	1
Min. p_T^{miss} [GeV]	40

Table 5.7: Baseline selection criteria in this analysis.

After applying the baseline selection the yields observed in data and predicted by the simulation are listed in Table 5.8. As a reminder, the $t\bar{t}H$ signal includes $H \rightarrow b\bar{b}$ as well as all other Higgs boson decay modes. Background contribution from QCD-multijet production have been found to be negligible.

5.4 Corrections to the event simulation

Since the final measurement relies to a large extent on the simulated samples described above, it is very important to ensure a high quality description of the observed data by the MC simulation. This section describes the corrections applied to simulated events to smooth out the observed discrepancies, induced by various effects ranging from MC event generator tunings to parton showering, hadronization, and detector simulation and reconstruction algorithms.

Process	2016	2017	2018
Data	328586	400062	578521
$t\bar{t}H$	326.9	392.1	585.5
$\pm$ tot unc.	± 12.7	± 14.9	± 21.1
$t\bar{t} + \text{LF}$	273597.3	325525.9	487855.7
$\pm$ tot unc.	± 10409.1	± 12776.9	± 18711.8
$t\bar{t} + c\bar{c}$	23188.7	27692.7	41761.4
$\pm$ tot unc.	± 1753.3	± 1811.0	± 2432.3
$t\bar{t} + b\bar{b}$	8300.4	9850.2	14993.8
$\pm$ tot unc.	± 396.0	± 452.8	± 737.8
Single- t	14363.1	17436.9	26377.1
$\pm$ tot unc.	± 750.3	± 875.5	± 1107.6
$t\bar{t}V$	869.5	1048.3	1616.4
$\pm$ tot unc.	± 47.8	± 53.6	± 75.9
$V + \text{jets}$	1598.2	1951.3	3041.6
$\pm$ tot unc.	± 124.5	± 143.5	± 213.5
tH	18.0	5.3	7.6
$\pm$ tot unc.	± 0.4	± 0.1	± 0.2
Total bkg.	328259.2	394919.1	590386.0
$\pm$ tot unc.	± 12428.0	± 15134.9	± 21540.3

Table 5.8: Event yields observed in data and predicted by the simulation after the baseline selection for 2016, 2017 and 2018, prior to the fit to data. The quoted uncertainties correspond to the total statistical and all systematic uncertainties with a shape changing effect (see Section 6.1).

For these reasons, dedicated control regions are defined. In these control regions, the underlying physics is well known and understood, such that the comparison of the simulated events to the data can help validating the simulation. Correction factors, generally known as scale factors (SFs), are derived and used to reweight the simulation to reduce the differences between simulated and recorded data. These SFs can modify the individual event features or account for observed deviations by reweighting the events based on the distribution of a specific parameter. These corrections are centrally provided by the corresponding groups within the Collaboration.

At analysis level, the various corrections are applied as an event weight calculated as the product of all incorporated corrections. In the following, the different corrections applied to simulated events are described. The determination of the SF results in additional uncertainties that will be discussed in Section 6.1.

Pileup reweighting

The effect of additional p-p interactions in the same bunch crossing (pileup) is included in the event simulation by generating the final-state particles of minimum bias collision events and overlapping it to the process of interest. An event by event correction centrally

provided is then applied to simulated events, so that the pileup distributions in MC matches the one measured from the data. The approach necessitates a measure of the probability of p-p interaction during bunch crossing, which is given by the inelastic cross section σ_{inel} . The most recent recommended value is $\sigma_{\text{inel}} = 69.2 \text{ mb} \pm 4.6\%$. To obtain the pileup distribution in data, σ_{inel} is multiplied by the instantaneous luminosity for each bunch crossing.

Trigger efficiencies

The triggers considered are critical for selecting $t\bar{t}H, H \rightarrow b\bar{b}$ events; similarly to the lepton identification and isolation efficiencies, the trigger performance in simulation is corrected to account for any differences to data.

The DL $t\bar{t} + \text{jets}$ MC simulation sample is used to calculate dilepton trigger efficiency in simulation, and scale factors are calculated individually for each dilepton channel and year, taking into account both statistical and systematic uncertainties. These factors are also determined for different kinematic regions (high and low number of primary vertices, number of jets, and missing transverse momentum). The methods used are similar to previous $t\bar{t}H, H \rightarrow b\bar{b}$ analyses [121, 127]. The scale factors are determined to be close to 1 with uncertainties below 5%. In the dilepton channel of the $t\bar{t}H, H \rightarrow b\bar{b}$ analysis, 2D scale factors are derived as a function of the leading and sub-leading lepton's p_T . These factors are shown in Appendix E, Figures E.1-E.3 for the 2016, 2017, and 2018 eras.

Lepton corrections

Since the selection requires the presence of electrons and muons in the final state of this analysis, SFs need to be applied to account for differences of their reconstruction, identification and isolation efficiencies between data and MC. Furthermore, the muons transverse momenta are corrected for biases caused by detector misalignment. Corrections are derived in bins of p_T and η for the individual years of data taking and applied on an object-by-object basis. They are centrally provided in the CMS Collaboration by the groups dedicated to the reconstruction and measurement of muons, and the one dedicated to electron and photon.

All electron and muon selection criteria of the analysis are identical to those employed in the efficiency measurements, so that the derived weights are applicable to the phase space of this analysis without the need for extrapolation.

Jet Energy corrections

Measuring the energy of jets is extremely difficult because of the hadronization of color-charged particles; thus, the detector response is continuously monitored and calibrated. The JEC are classified into two categories: the jet energy scale and jet energy resolution.

The CMS Collaboration measures the correction to the JES [263] following several steps. First, the energy of pileup contributions during jet clustering is subtracted by comparing samples created with and without pileup overlay. Residual differences between data and the detector simulation as a function of η are determined in the observed data events and then the simulated jet response corrections are derived by comparing the reconstructed jet p_T to the truth-level transverse moment. These corrections are derived as a function of the jet p_T and η , making the detector response uniform in these variables. At this stage only minor discrepancies in the order of few percents are observed between data and simulation. On the one hand, the η dependence of the corrections is further refined by adjusting the response to a jet in a simulated dijet sample with respect to a jet of similar p_T in the barrel region of the detector. In addition, to account for the implications of the HEM problem outlined in Section 2.2.7, a separate JES source is used for both observed data and simulated samples for 2018.

For the second category, dedicated studies indicate that the Jet Energy Resolution is worse in data than in the event simulation at the CMS experiment. The simulated jet p_T distribution is smeared to match the JER in data [161].

b-tagging corrections

The expected final state in the $t\bar{t}H$, $H \rightarrow b\bar{b}$ analysis features multiple b jets in the final state. Jets originating from bottom quarks are identified using the DEEJET tagging algorithm (see Section 3.2.5), which is applied both to data and simulated events. However, since the underlying machine learning technique contains numerous input variables vulnerable to mismodeling, variations in the efficiency of this tagger in data and simulation have been reported [213, 264].

Generally, the output of the DEEJET tagging algorithm can be exploited in two ways: a certain threshold value can define a b tag, resulting in specific identification and misidentification efficiency, or the shape information of the tagger discriminant can be used. This analysis utilizes the discriminant values as inputs to neural network algorithms for event classification; for this reason a calibration of the discriminant shape is required.

B-tagging SFs are measured a function of the jet b-tagging score, p_T and η using a tag-and-probe technique, separately for heavy flavor (HF) and light flavor (LF) jets. The SFs are derived in dedicated control region of $t\bar{t}$ events using a selection requiring exactly two leptons and two jets. The distribution of the b-tag score of the probe jet is compared in p_T and η intervals between observed data and simulated samples, with the latter normalized to the observed event yield, since the SFs are only designed to do shape calibration rather than rate correction. In order to account for a possible contamination of the control samples, the expected LF contribution is subtracted from the data before the actual calculation of the SF. For example, the correction for the HF jets is calculated as:

$$\text{SF}(\text{btag}, p_T, \eta) = \frac{\text{Data} - s_{\text{LF}}}{s_{\text{HF}}},$$

where s_{LF} and s_{HF} denote the number of events in the simulated samples arising from light-flavor and heavy-flavor contributions, respectively. The calculation for the LF jets is performed analogously. The final SF for the event is the product of the corrections for the individual jets in the event.

The SFs are then interpolated over the corresponding p_T and η bins using an iterative fit approach. For all jet b-tag score, p_T and η values, the method gives continuous scale factors. It should be noted that the method is not supposed to change the expected rate of events as determined prior to the application of any b-tag selection criteria, which potentially requires an a-posteriori reweighting of events in bins of jet multiplicity.

Figure 5.4 shows exemplary scale factors for b jets (a) and for udsg jets (b) for the three years. An unexpected behavior is observed in the simulation after the application of the SFs: changes in the distribution of observables such as the jet multiplicity and the transverse momenta of the jets. The effect is almost negligible in 2016, while it becomes significant in 2017 and 2018: the largest deviation is observed in 2017, as visible in Figure 5.5 comparing blue to the black distributions.

This unwanted effect is compensated by including an additional event-by-event correction. This additional b-tag normalization scale factor is applied as multiplicative correction to the simulation and represents an additional step of the b-tag SF derivation, optimizing the SF for the specific analysis phase space. The weights are generated using simulated events that match all of the selection criteria listed in Section 5.3, except for the b-tagging condition, to ensure that the phase space is as close as possible to the one of this analysis. The b-tag normalization SF is derived separately for the major signal and background processes $t\bar{t}H$,

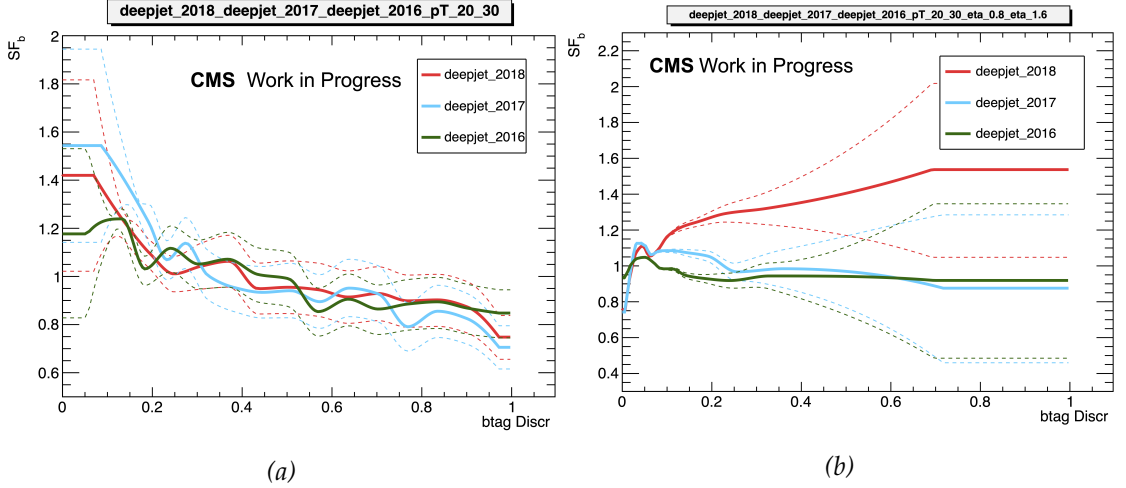


Figure 5.4: (a) Jet scale factors for b jets with $20 < p_T/GeV \leq 30$ and $|\eta| < 2.4$, and (b) for $udsg$ jets with $20 < p_T/GeV \leq 30$ and $0.8 < |\eta| < 1.6$ for the three different years of data-taking. Dotted lines represent total uncertainties of the scale factor calculation.

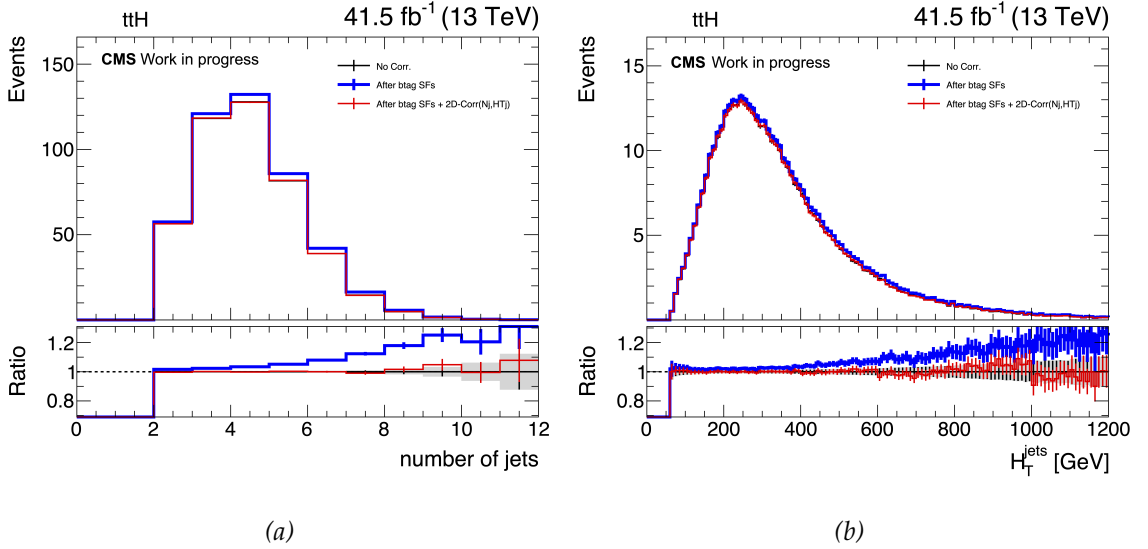


Figure 5.5: Jet multiplicity (a) and H_T (b) distributions for ttH events for 2017 after the baseline analysis selection. The distributions are shown before any b -tag correction (black), after applying the b -tag SF (blue), and after applying the b -tag SF and additional patch (red).

$t\bar{t} + b\bar{b}$, $t\bar{t} + c\bar{c}$ and $t\bar{t} + LF$; minor background processes are corrected with the weight for $t\bar{t} + LF$ events due to the poor statistics of simulated events. This is a purely MC-internal procedure added to the SF derivation to retain its internal consistency; no data is involved. First, two sets of 2D histograms of the jet multiplicity vs. sum of transverse jet momenta H_T are built for each background and signal processes, once with the b -tag scale factors applied and once without. The binning in jet multiplicity and H_T has been optimized for each class of processes such that the statistical uncertainty on the scale factor does not exceed 30% in most of the bins. The weights are then derived for each processes by dividing the 2D histogram where no b -tag SF is applied by the one with the b -tag SF applied. The event weights are shown in Appendix E in Figures E.4–E.7 for all the years.

As a validation, in Figure E.8 E.9 and E.10 of Appendix E, the DEEJET b -tagging discriminant distributions as well as jet kinematic distributions are shown for each process before

applying any b-tag correction (black), after applying the b-tag scale factors (blue), and after applying the b-tag reweighting patch. After application of both the b-tagging SF and the patch (red), the shape-altering effect of the b-tagging SF (blue) is corrected and the same shape as before applying any correction (black) is retained. Likewise, jet η distributions are not affected by the patch, while the jet p_T distribution, which is biased due to the application of the b-tagging SF, is properly corrected. This verifies that the reweighting patch corrects the bias introduced by the b-tag SF but does not change the shape of the relevant distributions, such as the DeepJet discriminator.

The data-to-MC agreement after the baseline selection is shown in Figure 5.6 for the jet and b-tagged jet multiplicity distributions for the three data-taking periods. Figures 5.7-5.9 illustrate the p_T and η distributions for the leading lepton and the two leading jets. Overall good agreement is observed within the uncertainties. The trend visible in the phase-space with higher jet and b-jet multiplicity is expected. This is because the $t\bar{t} + b\bar{b}$ (4FS) sample features harder jet and b-jet spectra (as described in the next Section), due to the two additional b jets from the ME which have higher p_T and thus more often pass the jet- p_T threshold.

5.5 $t\bar{t} + \text{jets}$ background model

Upon the baseline selection, the $t\bar{t} + \text{jets}$ production represents the major background contributions, accounting for approximately 95% of the total predicted background events in the DL channel (see Table 5.8). As already highlighted above in Section 5.2.2, the $t\bar{t}$ events with additional b jet production are highly crucial. Generally, they are theoretically difficult to model due to the multi-parton final state with two very different scales, and in the context of this analysis their topology is identical to the $t\bar{t}H$ signal events but with a cross section 10-20 times larger in the analyzed phase space. Dedicated measurements [43, 119, 120] and previous iteration of this analysis [121, 127] have demonstrated that the $t\bar{t} + b$ jets contribution in data is larger than expected by the simulation by a factor of 1.2–1.4.

The $t\bar{t} + \text{jets}$ production is modeled using MC simulation and the simulated events are split into mutually exclusive processes according to the flavor of the additional jets at truth-level not originating from the top quark decay (see Section 5.2.2 for detailed description). This distinction is critical since the processes are subject to distinct systematic uncertainties.

Previously, the entire $t\bar{t} + \text{jets}$ background, including $t\bar{t} + b$ -jets events, has been modeled using an inclusive $t\bar{t}$ simulation at NLO in the 5FS. There, the critical additional b jets not from the top decay are modeled by the PS and subject to PS tune and large uncertainties in general. Additional, empirically driven cross section uncertainties were also assigned to the different $t\bar{t} + b$ jets topologies. An improved background model was instead developed for this analysis. A dedicated exclusive $t\bar{t} + b\bar{b}$ simulation [44] at NLO in the 4 FS is used to model the $t\bar{t} + b$ jets events, and their overall normalization is left as a freely-floating parameter in the final fit. The three processes, $t\bar{t} + b\bar{b}$, $t\bar{t} + b$, and $t\bar{t} + 2b$, are now treated as one single process, $t\bar{t} + B$, subject to the same common normalization correction and the same scale uncertainties, which reflects the fact that they stem from the same underlying diagrams.

The process simulation has been investigated within the LHC Higgs Cross Section Working Group [42]: several variables relevant to this analysis have substantially different predicted distributions in the $t\bar{t}$ and $t\bar{t} + b\bar{b}$ simulations. The measurement precision is not yet good enough to distinguish between the models, but there is an a priori preference for the more fundamental ME-level prediction of the $t\bar{t} + b\bar{b}$ sample. In addition, the scale uncertainties of the ME computation, which are conceptually better defined than the uncertainties on the PS-tune parameters, are the major uncertainties. Therefore, the $t\bar{t} + b\bar{b}$ simulation is here employed to model the $t\bar{t} + b$ jets background.

The prediction of the $t\bar{t}$ (5FS) and $t\bar{t} + b\bar{b}$ (4FS) samples have been compared at truth-level:

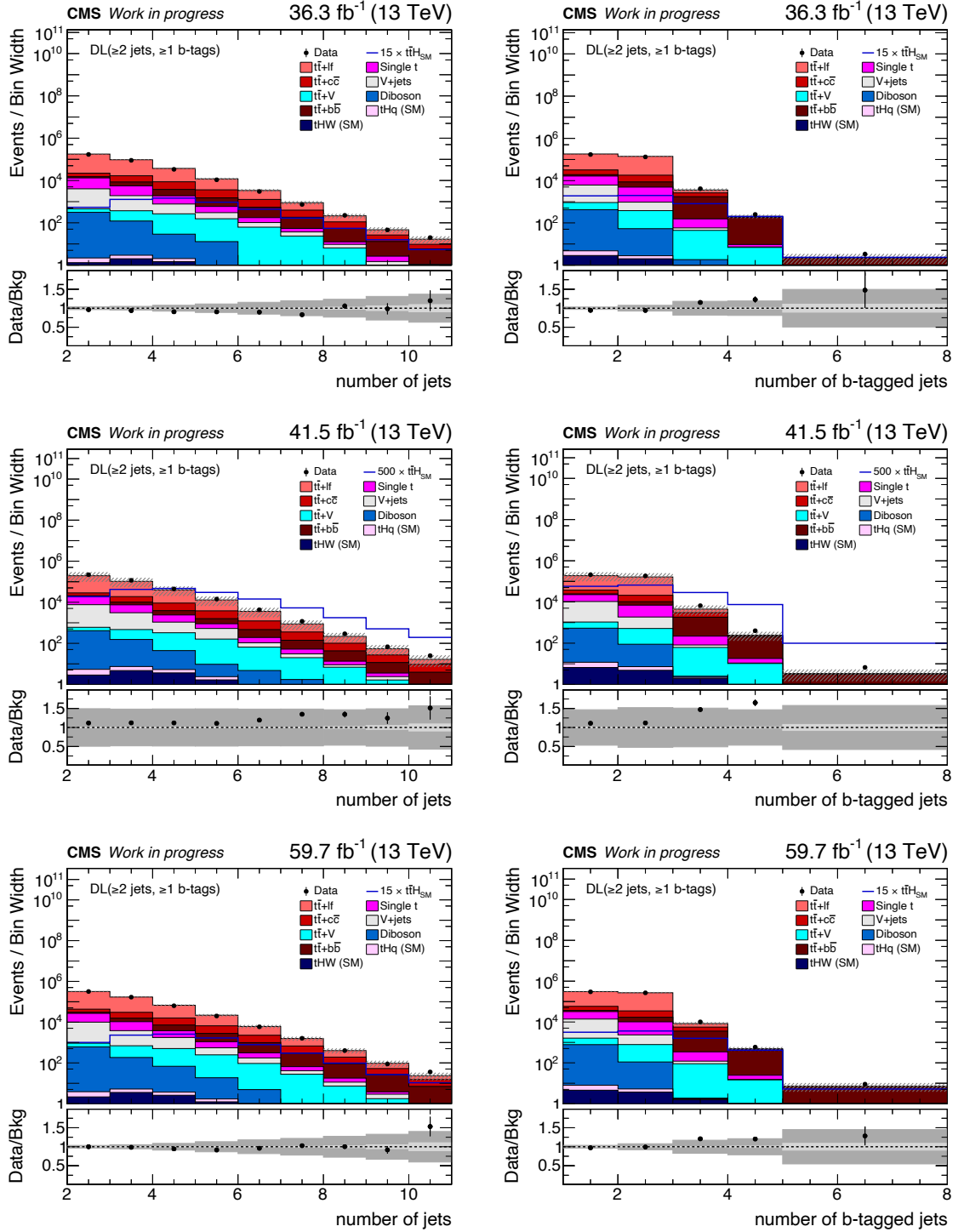


Figure 5.6: Jet (left) and b-tagged jet (right) multiplicity for 2016 (top), 2017 (middle), and 2018 (bottom) after the baseline selection. The expected background contributions (filled histograms) are stacked, and the expected signal distribution (line) is superimposed. The last bin includes overflow events. The hatched uncertainty bands correspond to the total statistical and systematic uncertainties, added in quadrature. The light gray band indicates the systematic uncertainties, while the dark gray band illustrates the statistical uncertainties in each bin of the background prediction. The lower plots show the ratio of the data to the background prediction.

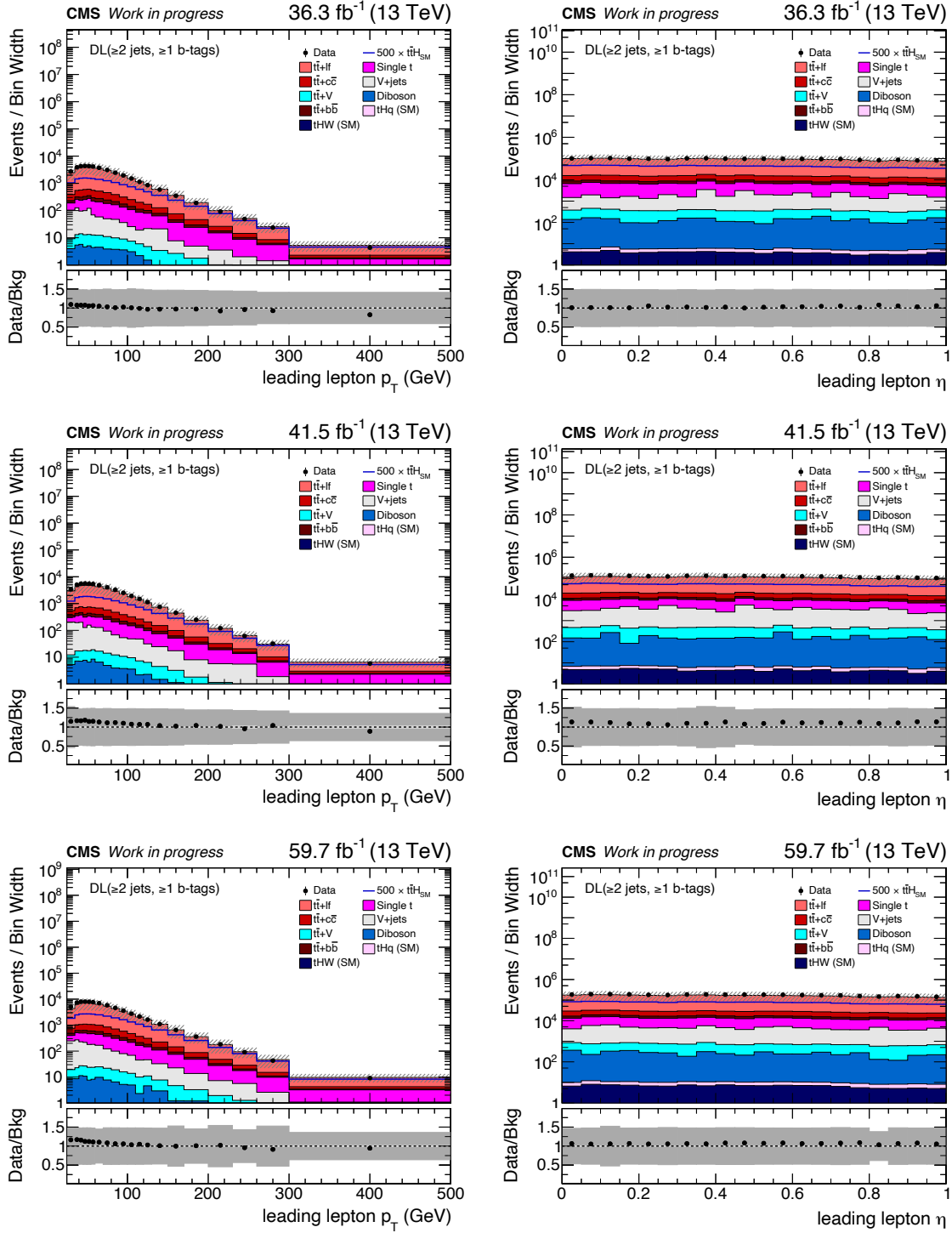


Figure 5.7: Leading lepton p_T (left) and η (right) for 2016 (top), 2017 (middle), and 2018 (bottom) after the baseline selection. The expected background contributions (filled histograms) are stacked, and the expected signal distribution (line) is superimposed. The last bin includes overflow events. The hatched uncertainty bands correspond to the total statistical and systematic uncertainties with a shape changing effect, added in quadrature. The light gray band indicates the systematic uncertainties, while the dark gray band illustrates the statistical uncertainties in each bin of the background prediction. The lower plots show the ratio of the data to the background prediction.

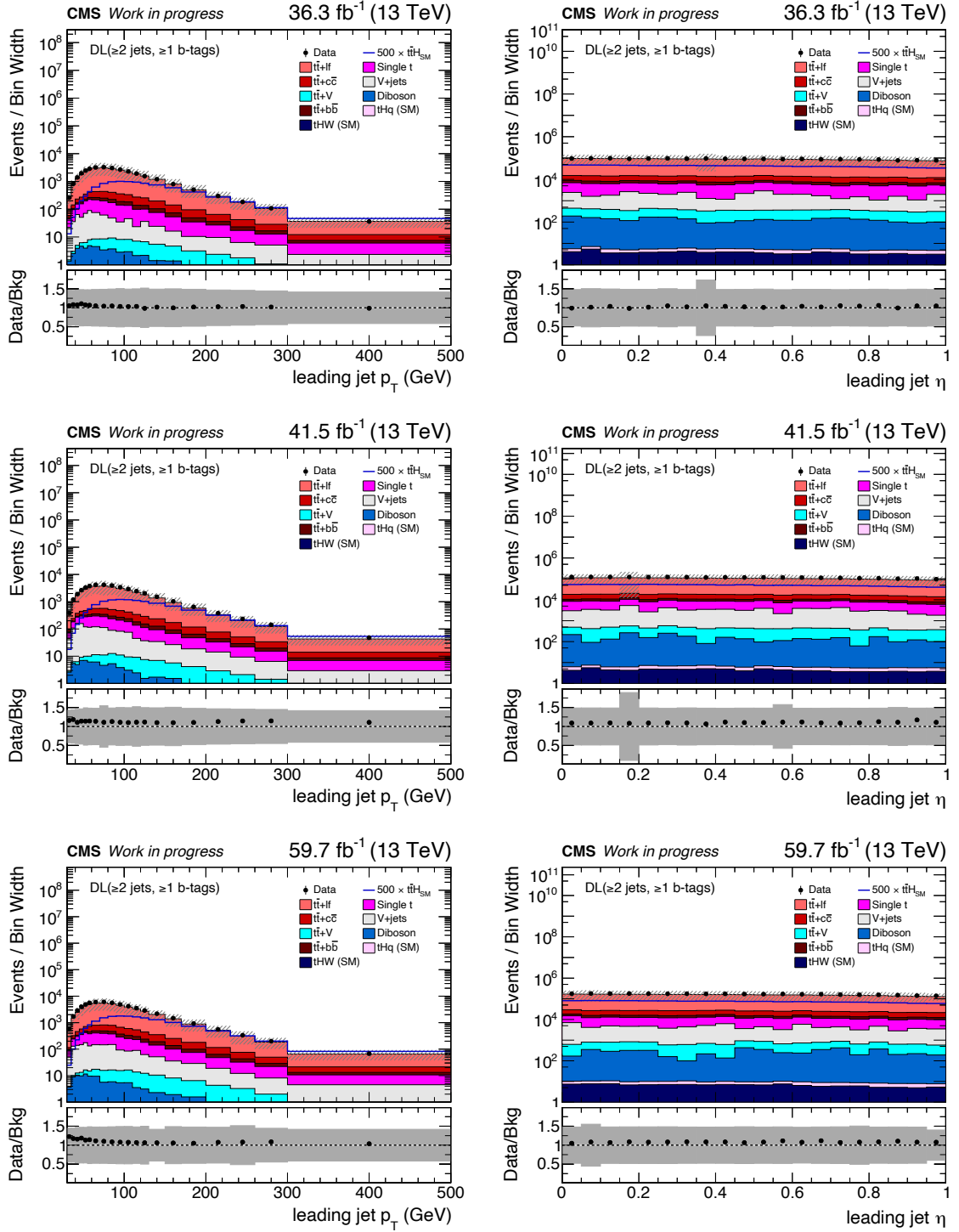


Figure 5.8: Leading jet p_T (left) and η (right) for 2016 (top), 2017 (middle), and 2018 (bottom) after the baseline selection. The expected background contributions (filled histograms) are stacked, and the expected signal distribution (line) is superimposed. The last bin includes overflow events. The hatched uncertainty bands correspond to the total statistical and systematic uncertainties with a shape changing effect, added in quadrature. The light gray band indicates the systematic uncertainties, while the dark gray band illustrates the statistical uncertainties in each bin of the background prediction. The lower plots show the ratio of the data to the background prediction.

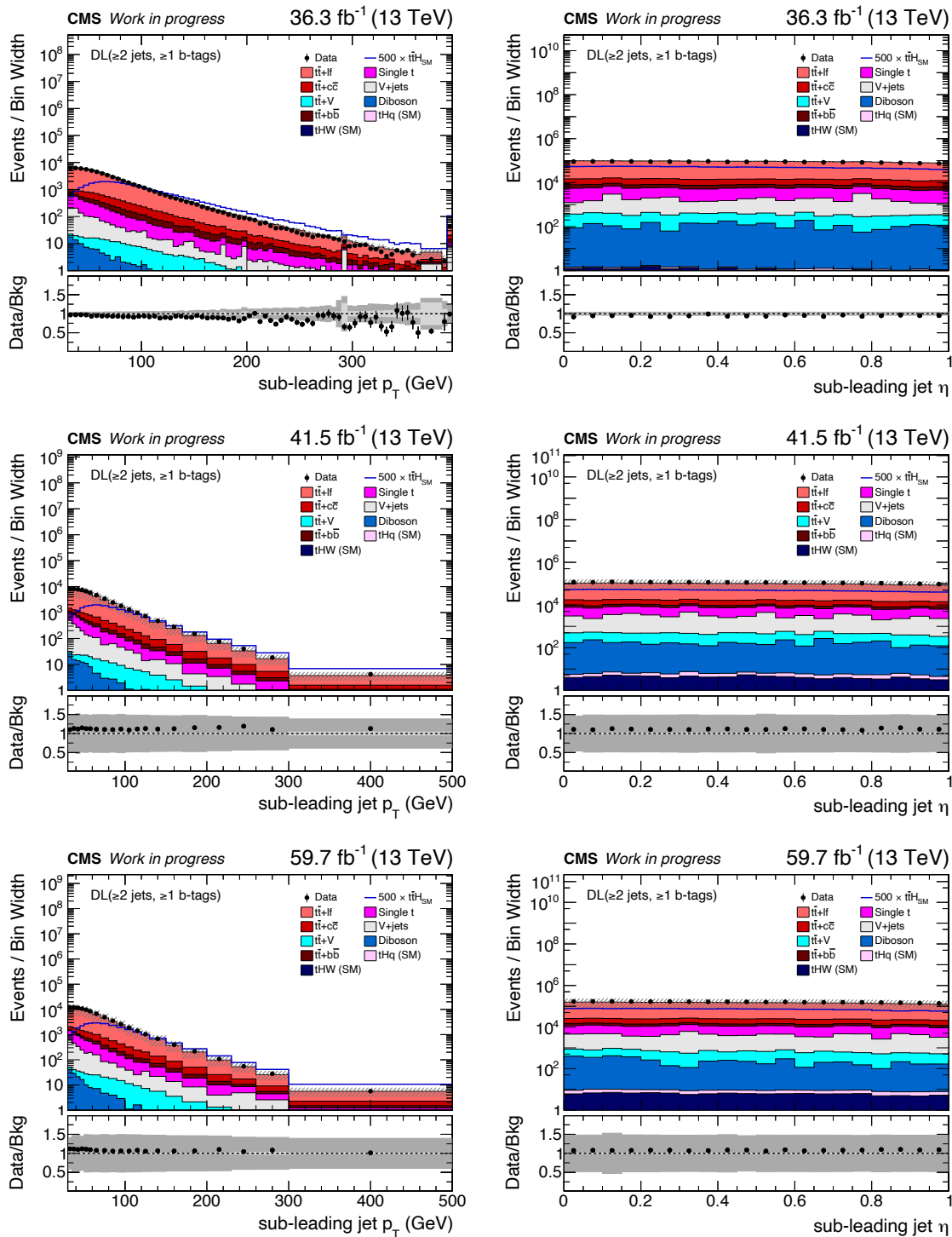


Figure 5.9: Sub-leading jet p_T (left) and η (right) for 2016 (top), 2017 (middle), and 2018 (bottom) after the baseline selection. The expected background contributions (filled histograms) are stacked, and the expected signal distribution (line) is superimposed. The last bin includes overflow events. The hatched uncertainty bands correspond to the total statistical and systematic uncertainties with a shape changing effect, added in quadrature. The light gray band indicates the systematic uncertainties, while the dark gray band illustrates the statistical uncertainties in each bin of the background prediction. The lower plots show the ratio of the data to the background prediction.

in general, the $t\bar{t} + b\bar{b}$ (4FS) sample features harder jet and b-jet spectra in the phase-space with high jet and b-jet multiplicity, especially close to the selection threshold. This behavior is expected because of the two additional b jets coming from the ME, generally having higher p_T . Observables related to the distance and invariant mass of the two additional b jets are different because the prediction of the kinematics can differ. The uncertainties are larger in the $t\bar{t} + b\bar{b}$ (4FS) sample than in the $t\bar{t}$ (5FS) sample, due to the additional b jets originating from the ME and subject to scale uncertainties.

The $t\bar{t} + c\bar{c}$ and $t\bar{t} + \text{LF}$ events instead are taken from the $t\bar{t}$ sample, since they do not include additional b jets within acceptance but can still pass the event selection due to mistagging. A merging procedure has been developed to combine both samples to cover the full $t\bar{t}$ + jets phase space without overlaps, as illustrated in Fig. 5.10. The $t\bar{t} + b\bar{b}$ simulation is expected to better model the kinematic but not the cross section prediction. For this reason, the fractional contributions of the $t\bar{t} + B$, $t\bar{t} + c\bar{c}$, and $t\bar{t} + \text{LF}$ events are chosen based on the $t\bar{t}$ sample and the total yield of the $t\bar{t}$ + jets events is scaled to the NNLO+NNLL $t\bar{t}$ inclusive cross section of 831.76 pb calculated in the 5FS scheme [261]. So the dedicated $t\bar{t} + b\bar{b}$ simulation is used to describe the shapes of the kinematic distributions of $t\bar{t} + b$ jets events, while their normalization is scaled to the yield predicted by the inclusive $t\bar{t}$ + jets 5FS sample. As a result, the overall yield of all $t\bar{t} + X$ events remains unchanged. The final yields of the $t\bar{t} + B$ and $t\bar{t} + c\bar{c}$ contributions are free parameters and determined in-situ in the fit to data, allowing flexibility to match the predicted $t\bar{t} + B$ yield to the one observed in data.

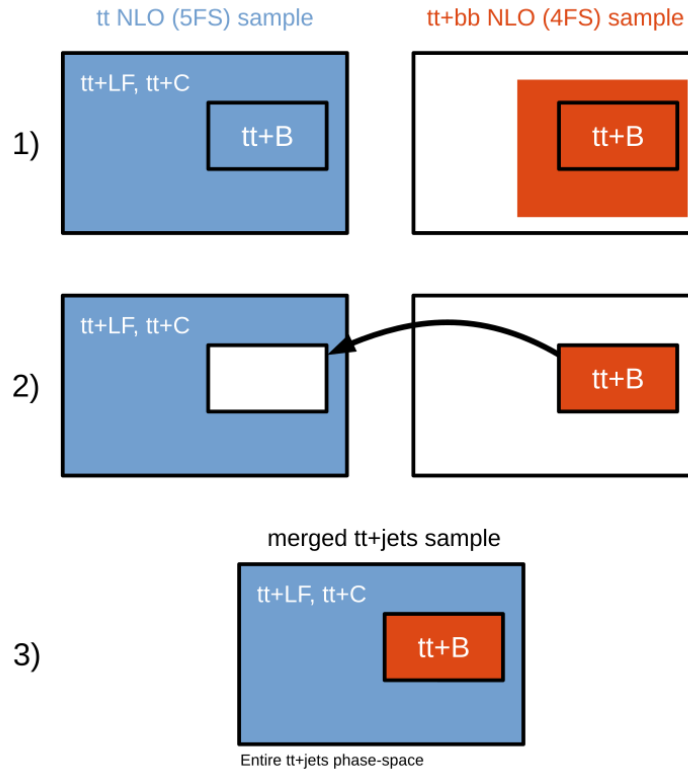


Figure 5.10: Illustration of the procedure to merge the $t\bar{t} + b\bar{b}$ and $t\bar{t}$ samples.

For the $t\bar{t}$ and $t\bar{t} + b\bar{b}$ samples, dedicated modeling uncertainties have been assigned when they have different definition for the 5FS vs. 4FS samples (more details in Section 6.1). In general, the relative yields of the $t\bar{t}$ + jets and $t\bar{t} + b\bar{b}$ samples can change when varying the ISR, FSR, μ_R and μ_F scale and PDF uncertainties. Therefore, for each variation, the $t\bar{t} + b\bar{b}$ sample is normalized such that its inclusive yield after the variation (and before any

event selection) corresponds to the one predicted by the inclusive $t\bar{t} + \text{jets}$ simulation after the simulation (and before any event selection). However, the $t\bar{t} + 2b$ process is affected by additional effects occurring in the collinear regime: additional gluon emission from the splitting gluon is probably underestimated [265]. Therefore, an additional 100% rate uncertainty is still assigned to the $t\bar{t} + 2b$ part.

Figure 5.11 shows a comparison of discriminant distributions used in the final fit (as defined in Section 5.7) for $t\bar{t} + B$ events in the categories most sensitive to the $t\bar{t}H$ signal, for the 2016 data. The distribution obtained with the dedicated $t\bar{t} + b\bar{b}$ sample ($t\bar{t} + b\bar{b}$ 4FS) used in this analysis is compared to the distribution obtained when using the $t\bar{t}$ sample ($t\bar{t} + b\bar{b}$ 5FS). The error bands show the total systematic uncertainties and the shape differences are well below or at the 1σ level. Again, the normalization differences are covered by the freely-floating normalization parameter in the fit. The same conclusions holds for all years, as visible in Figure F.1 in Appendix F.

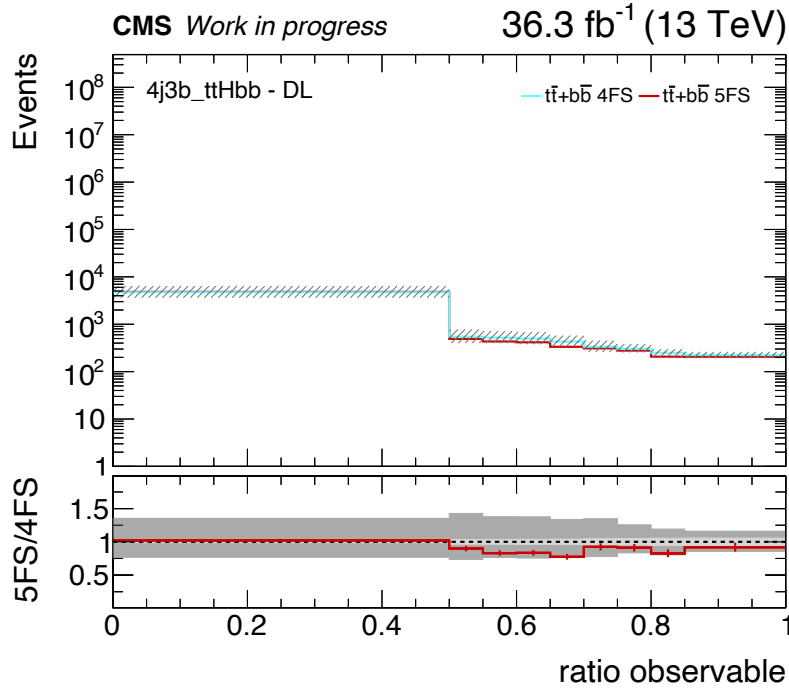


Figure 5.11: Distribution of $t\bar{t} + B$ events in the ≥ 4 jets, ≥ 3 b tags category for 2016 for the observable used in the final fit. The distribution obtained with the dedicated $t\bar{t} + b\bar{b}$ sample ($t\bar{t} + b\bar{b}$ 4FS, cyan) is compared to the one from the $t\bar{t}$ ($t\bar{t} + b\bar{b}$ 5FS, red) sample and the error bands contain the total systematic uncertainty.

The modeling of the $t\bar{t} + B$ background has been validated also in a comparison to the data in the different phase-space regions, also the ones relevant for the STXS measurement. As expected, in a phase-space region with more b jets the data is harder than in the simulation with the $t\bar{t}$ (5FS) sample and therefore much better described with the $t\bar{t} + b\bar{b}$ (4FS) sample. For example, the predicted distributions of the jet and b -tagged jet multiplicities (see Figure 5.6) after the baseline selection agree well with the ones observed in data within the uncertainties. As another example, the Higgs boson candidate identified using the two jets with the invariant mass closest to 125 GeV has also been validated in different regimes of Higgs-boson p_T . Figure 5.12 shows the distribution in a region of high jet and b -tagged jet multiplicity, considering all the combined data sets used in the analysis. The actual sensitivity to $t\bar{t}H$ in this observable is small so that the analysis is unblinded, but the background modeling can be tested in this way. The shape of the data distribution is well described by

the simulation. Differences in normalization are due to the expected under-prediction of the $t\bar{t} + B$ cross section by the simulation.

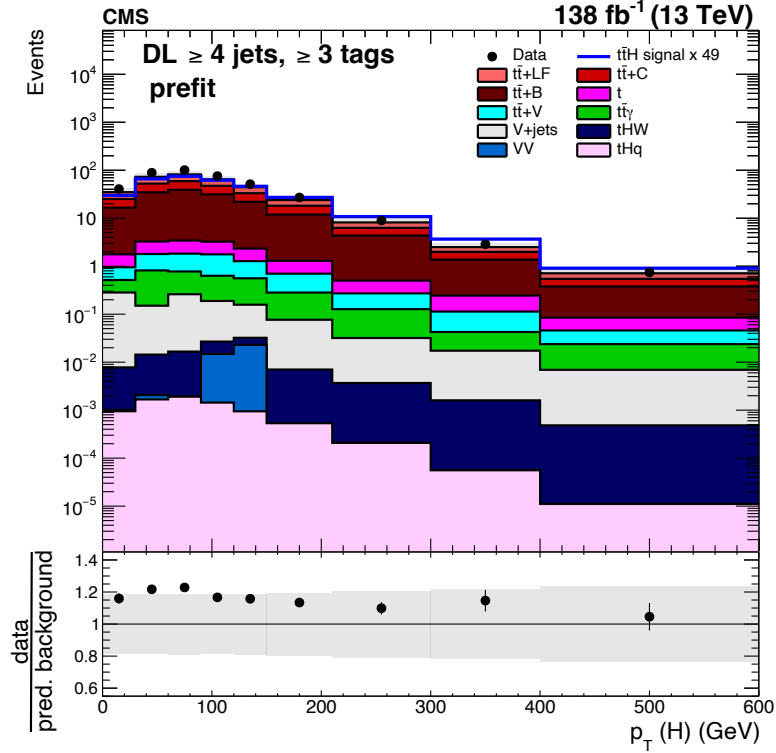


Figure 5.12: p_T of the Higgs boson candidate for events passing the baseline selection requirements in the ≥ 4 jets, ≥ 3 b tags category, considering all the combined data sets used in the analysis.

Furthermore, the stability and robustness of the statistical model against statistical fluctuations in the data and against potentially mismodeled $t\bar{t} + b\bar{b}$ background have been successfully tested using toy data as reported in Appendix F.

5.6 Event categorization in the analysis regions

After the MC simulated samples are confirmed to describe the recorded data well, the selected events are split into further analysis regions. To improve the sensitivity of the binned maximum likelihood fit, event categories are constructed such that each of them is enriched in either different background contributions or the $t\bar{t}H, H \rightarrow b\bar{b}$ signal. This allows constraining the nuisance parameters in the fit that are related to the background contributions. The description of the maximum-likelihood fit is given in Chapter 6.

Two mutually exclusive event categories are defined based on the number of jets and b-tagged jets: exactly 3 b-tagged jets $3 \text{ jets}, \geq 3 \text{ b tags}$ and at least 4 jets of which at least 3 are b-tagged $\geq 4 \text{ jets}, \geq 3 \text{ b tags}$. The first category is enriched in background events originating from the $t\bar{t} + \text{LF}$ process, while the relative contribution of signal events and $t\bar{t} + \text{HF}$ events increases dramatically in the $\geq 4 \text{ jets}, \geq 3 \text{ b tags}$ category. The splitting in these two analysis regions to increase the sensitivity to the signal comes at the expense of a decrease in the statistical precision. However, it has been proven to be the choice that provides an optimal balance between reaching higher signal purity and maintaining good control over the background processes and systematic uncertainties.

Table 5.9 reports the event yields in data and MC for the two event categories.

2016		
Process	3 jets, ≥ 3 b tags	≥ 4 jets, ≥ 3 b tags
Data	1697	2701
$t\bar{t}H$	13	55
$\pm$ tot unc.	± 2	± 5
$t\bar{t} + LF$	433	551
$t\bar{t} + c\bar{c}$	322	596
$t\bar{t} + b\bar{b}$	626	1155
Single-t	47	53
$t\bar{t}V$	8	42
V+jets	7	26
Total bkg.	1448	2434
$\pm$ tot unc.	± 360	± 332

2017		
Process	3 jets, ≥ 3 b tags	≥ 4 jets, ≥ 3 b tags
Data	2740	4297
$t\bar{t}H$	118	70
$\pm$ tot unc.	± 2	± 6
$t\bar{t} + LF$	722	925
$t\bar{t} + c\bar{c}$	505	905
$t\bar{t} + b\bar{b}$	859	1551
Single-t	60	85
$t\bar{t}V$	15	58
V+jets	12	29
Total bkg.	2183	3564
$\pm$ tot unc.	± 530	± 485

2018		
Process	3 jets, ≥ 3 b tags	≥ 4 jets, ≥ 3 b tags
Data	4167	6569
$t\bar{t}H$	27	105
$\pm$ tot unc.	± 4	± 10
$t\bar{t} + LF$	1234	1534
$t\bar{t} + c\bar{c}$	742	1370
$t\bar{t} + b\bar{b}$	1287	2340
Single-t	103	131
$t\bar{t}V$	23	80
V+jets	19	41
Total bkg.	3421	5526
$\pm$ tot unc.	± 758	± 737

Table 5.9: Event yields observed in data and predicted by the simulation after the categorization into 3 jets, ≥ 3 b tags and ≥ 4 jets, ≥ 3 b tags for 2016 (top), 2017 (middle) and 2018 (bottom), prior to the fit to data. The quoted uncertainties correspond to the total statistical and all systematic uncertainties with a shape changing effect (see Section 6.1).

5.7 Observables for the parameter estimation

As the search for $t\bar{t}H$ production is background-dominated, the choice of the observables for the parameter estimation is crucial to optimize the sensitivity of the $t\bar{t}H, H \rightarrow b\bar{b}$ analysis. Ideally, the observable should provide an adequate separation between signal and background events and also constrain background processes individually. In the final state under investigation, one could expect the reconstructed invariant mass of the two b -tagged jet emerging from the Higgs boson decay to be very sensitive, having a narrow resonance peak around the mass of the Higgs boson and a continuous distribution for the background processes. But this is not the case, as shown in Figure 5.13, due mostly to large background rate and experimental inefficiency effects in the jet reconstruction and the b tagging, which result in a non-unambiguous way of reconstructing the distinct invariant mass peak of the Higgs boson. This is based solely on the 2016 data set and should be considered indicative for the general pattern, as the distribution is consistent for the other years.

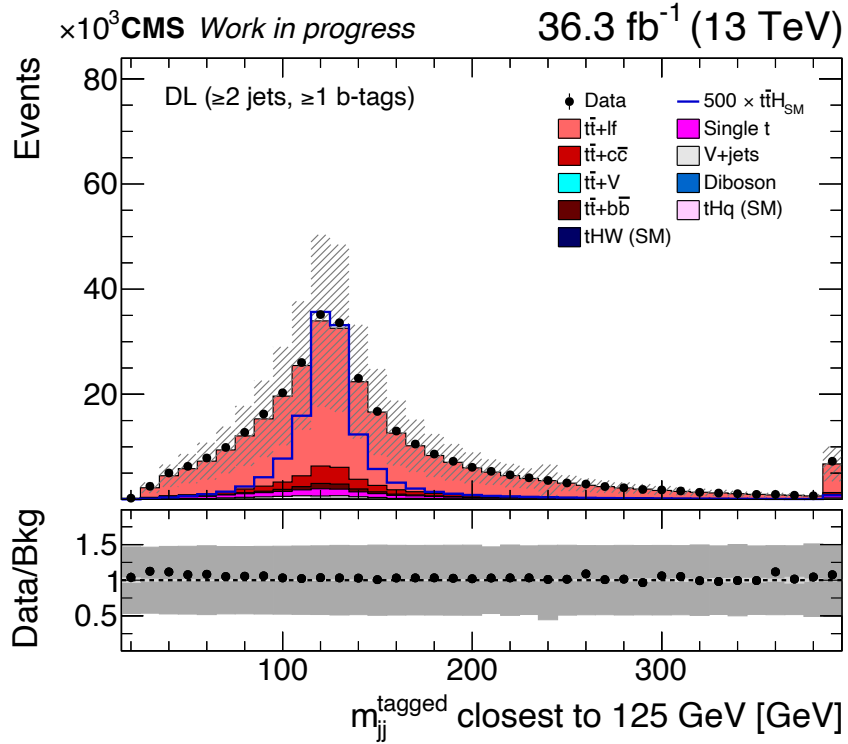


Figure 5.13: Reconstructed invariant mass of the two b -tagged jets closest to 125 GeV, for the 2016 data set. The stacked histograms show the distributions of the background processes. For a better comparison, the expected signal process is scaled with a normalization factor of 500. The black dots indicate the observed data. The bottom half of the figure shows the ratio of observed data with respect to the background prediction. The light gray band indicates the systematic uncertainties, while the dark gray band illustrates the statistical uncertainties in each bin of the background prediction.

Even after constructing the signal-enriched analysis region described above, the reconstructed invariant mass of the two b -tagged jets closest to the Higgs boson mass does not provide the desired separation between signal and background events due to the still large background contribution. Therefore, more suited discriminating observables need to be found that allow for a proper signal extraction. In principle, any distribution of any variable can be used, as long as it provides the necessary signal and background separation based on proper understanding of the features of all involved processes. This is usually accomplished by harnessing the power of multivariate analysis methods (MVA), which can more easily

identify differences and exploit the correlation among multiple variables of the reconstructed events.

Previous searches for $t\bar{t}H, H \rightarrow b\bar{b}$ employed the Matrix Element Method (MEM) (details in Appendix G), and a machine learning algorithm in the form of Boosted Decision Trees (BDT), as highlighted in Section 1.3.4. Both techniques thus differentiate just between signal and background, in which all distinct contributing processes are equally treated. The work presented in this thesis employs a newly designed multi-classification method as well as a collection of discriminating variables customized for the search of $t\bar{t}H, H \rightarrow b\bar{b}$ production in the DL final state. This algorithm is based on artificial neural networks (ANNs) with a novel approach for selecting the most discriminating variables and validating them to address higher order-correlations. An introduction to neural network is available in Appendix A. The strategy is designed not only to distinguish between signal and background events, but also to split the latter into distinct contributions, resulting in bins enriched with one of the relevant physics processes and improving the overall analysis sensitivity. As shown in Figure 5.14, a selected set of input variables measured for each event is fed into the ANN. It returns as output a vector of values between 0 and 1, reflecting the likelihood of a given event to originate from the signal or one of the three background $t\bar{t} + \text{jets}$ processes ($t\bar{t} + b\bar{b}, t\bar{t} + c\bar{c}, t\bar{t} + \text{LF}$). As a result, the ANN has four distinct classes in the most sensitive category. The vector elements sums up to one, and the classification into one of the classes is based on the most probable process, i.e. the class with the largest output value. The corresponding event mapping is unambiguous, preventing any double-counting. At the same time, the distributions of the output serve further to construct the final discriminant distributions that enter the fit.

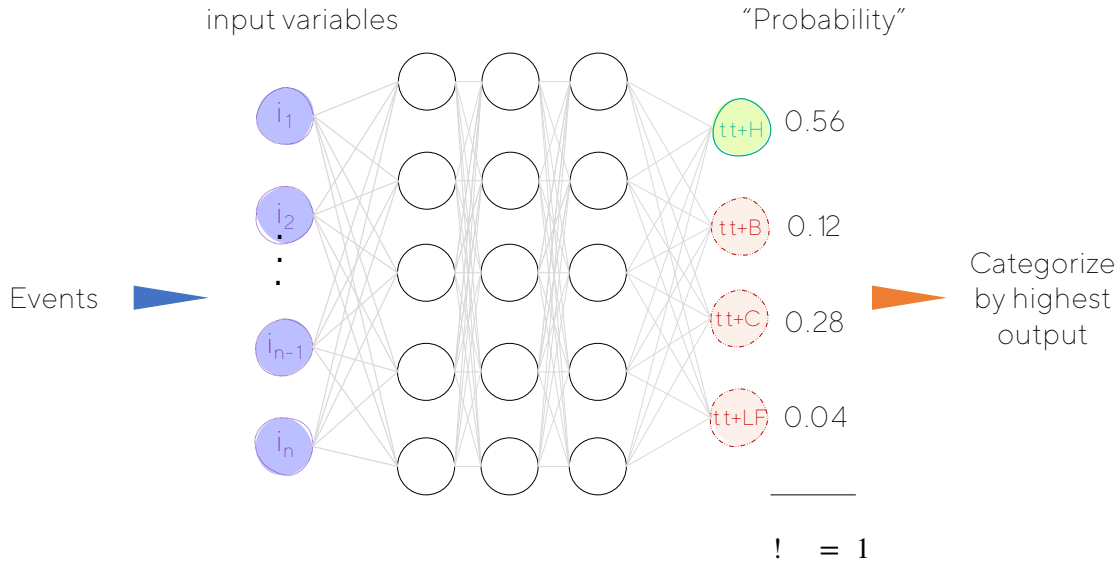


Figure 5.14: Example of strategy based on multi-class classification ANNs. Four values are assigned to events based on their likelihood of originating from the $t\bar{t}H$ signal or from one of the backgrounds. An event is assigned to the class with the greatest probability ($t\bar{t}H$ in this example).

Figure 5.15 displays the analysis workflow, from selection, categorization by the number of reconstructed jets/b jets, to the ANN trained and optimized in each jet/b-tag multiplicity category. In total this results in 5 categories with corresponding discriminants, which are employed simultaneously in the signal extraction procedure.

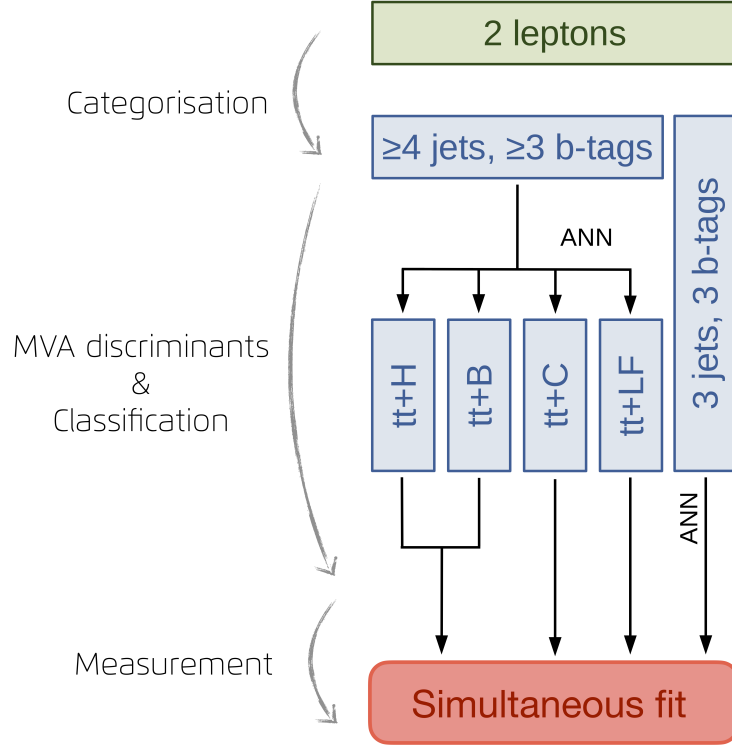


Figure 5.15: Representation of the analysis workflow with a focus on the classification and categorization. ANNs are trained in each jet/b-tag category to further classify the events based on their most likely process. The measurement is performed by a simultaneous fit of the expected ANN output distributions to data in all 5 categories with binned maximum likelihood fit.

5.7.1 Input variables to the neural networks

For each year and each jet/b-tag category, the most discriminating ANN input variables are selected based on their discrimination power between $t\bar{t}H$ signal and $t\bar{t}$ background events. The large set of inputs (~ 130) includes variables related to kinematic properties of single object quantities (e.g. transverse momenta of leptons and jets), combined variables of multiple physics objects (e.g. angular distributions between momentum vectors), jet b-tagging discriminant, event-shape variables but also more sophisticated observables like the b tagging likelihood ratio (BLR) and the MEM discriminants. The complete list with detailed definitions can be found in Appendix G, as well as a selection of their distributions.

In the first step, variables are ranked according to the output score of an unbinned one-dimensional Kolmogorov-Smirnov (KS) test [266, 267]: the most discriminating ones are selected and used in the subsequent steps. As expected, the MEM is among the input variables with the highest signal-vs-background discrimination power. The correlations among these first pre-selected variables are shown in Figures 5.16 for the ≥ 4 jets, ≥ 3 b tags category.

The second step is the construction of a preliminary fully-connected feed-forward network to gain further insight on the relative influence of each input to the ANN output. A Taylor expansion of the ANNs outputs [268] helps performing a second ranking based on the first-order derivatives of the output with respect to the individual input features. The first 20 (40 for the ≥ 4 jets, ≥ 3 b tags category) inputs are selected. Additional studies to improve the understanding of the relative importance among the input features are performed looking also at the Shapley metric [269]. This method originated in game theory and computes the importance of an input feature in the ensemble by calculating the Shapley metric with

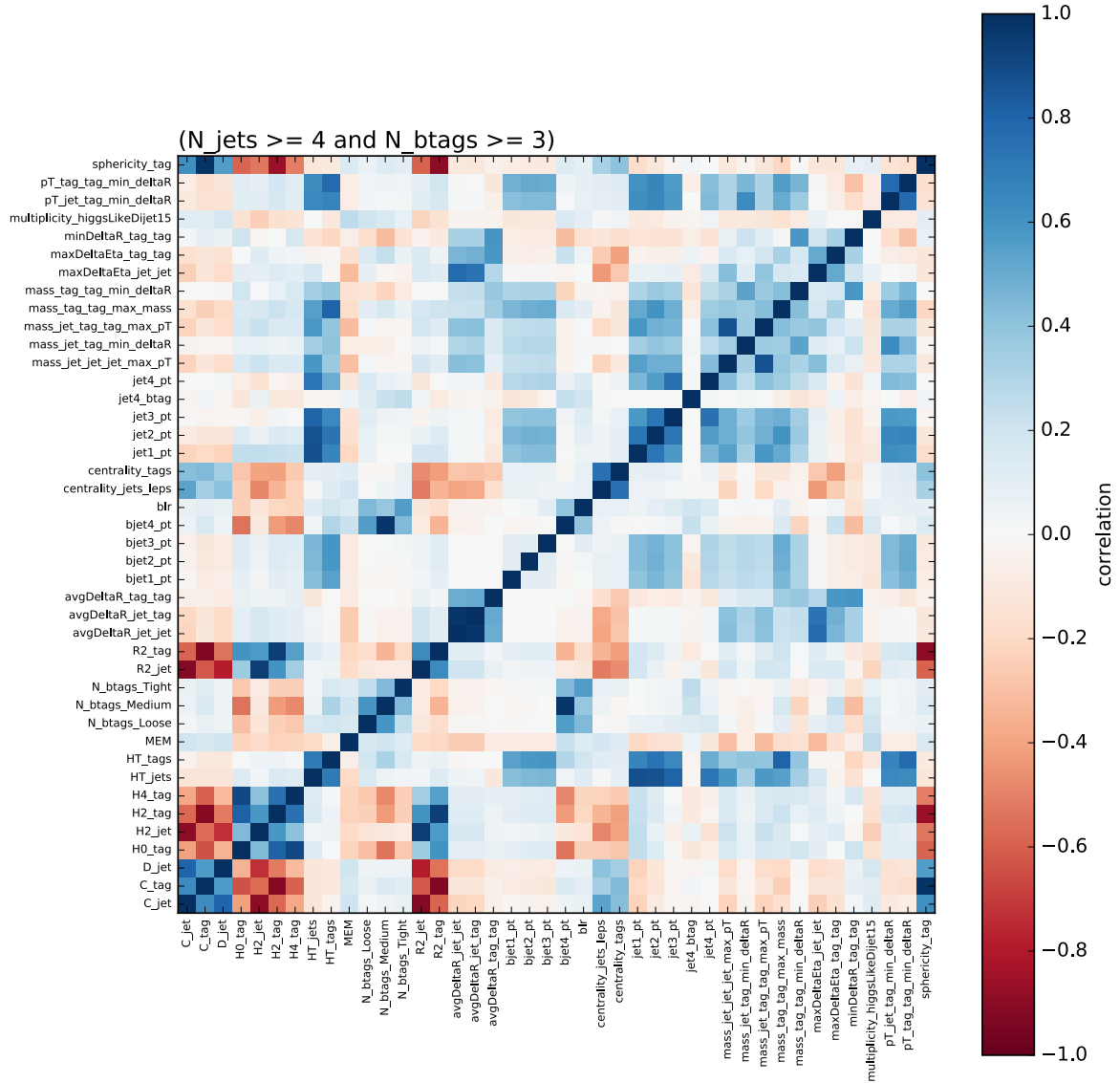


Figure 5.16: Correlations among the preselected input variables in the ≥ 4 jets, ≥ 3 b tags category for the 2018 data set.

and without a given observable, thus gauging the impact on the ANN. In Appendix H, Figures H.1–H.3 show the Taylor coefficients determined for the input variables used in the $3\text{jets}, \geq 3\text{ b tags}$ and $\geq 4\text{ jets}, \geq 3\text{ b tags}$ categories for 2016, 2017 and 2018, respectively. Figure 5.17 represents the Shapley values for the relative importance of the input features for the $\geq 4\text{ jets}, \geq 3\text{ b tags}$ category.

In general, the three distinct methods of ranking consistently identify the inputs with the highest discrimination power (first 5-10 inputs); in particular, the second and third rankings largely affect the remaining (sub-leading) inputs. The final rankings obtained with this procedure are consistent with the expectation. In the $\geq 4\text{ jets}, \geq 3\text{ b tags}$ category, the MEM and the BLR provide the best separation between the $t\bar{t}H$ signal and the $t\bar{t} + b\bar{b}$ background, with the BLR also helping the discrimination between $t\bar{t} + b\bar{b}$ and $t\bar{t} + \text{LF}$; other variables, such as $\Delta\eta_{b\bar{b}}$ or $m_{b\bar{b}}$ further contribute to the overall discrimination.

At this point a subset of approximately 30 variables, from the initial large set of observables considered as potential input features, is selected based on their importance. The distributions of the selected inputs are required to be properly described by the MC simulation with respect

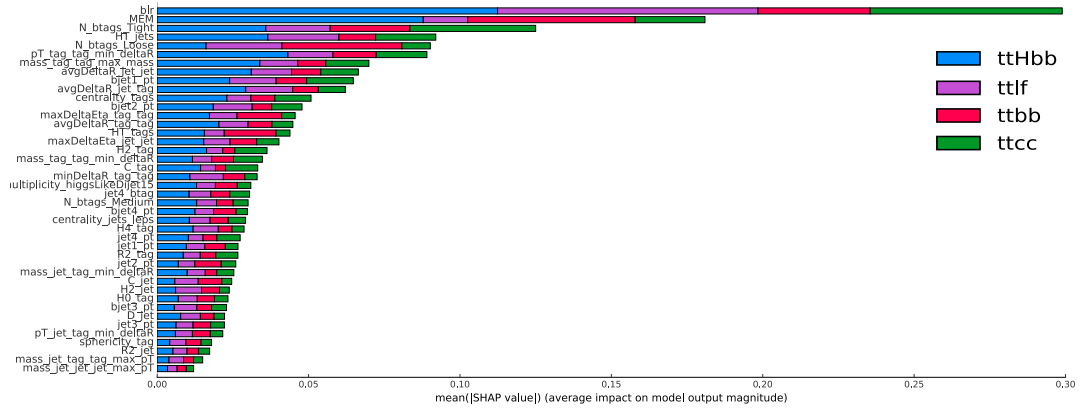


Figure 5.17: The x -axis shows the mean of the absolute Shapley values used for the ranking of the input variables in the final ANN training of the ≥ 4 jets, ≥ 3 b tags category, considering all three eras of data taking. The different colors of the bars indicate the dependence of the given category class on the respective input feature. Large values indicate a high importance.

to data. The central prediction and the correlation of the input features, together with their uncertainties, are the critical aspects to validate.

The validation is here performed with binned Goodness-Of-Fit (GoF) tests based on the full uncertainty model, i.e. including statistical and systematic uncertainties (see Section 6.1) using the saturated model method (see Appendix B.3). The aim is to guarantee that evident mismodelings between data and simulation (for example, detector effects not entirely accounted for by the uncertainty model) do not propagate to the output of the ANN classifiers.

In a first iteration, the 1D distribution of each potential input variable is considered. Subsequently, the 2D distribution for each pair of variables passing the 1D GoF test is tested, in order to ensure (to a certain extent) the proper modeling of the correlations between different inputs. From post-fit distributions of a test statistic constructed using the saturated model, a p value on the marginal distribution is determined, corresponding to the likelihood of observing the value of the test statistic calculated in observed data or a larger value given the probability density of the statistical model. Consequently, large p -values implies good agreement between the observed data and the simulated samples, whereas small values suggest a poor description. Variables with a p value below the threshold of 5% are further closely inspected and potentially not used in further analysis.

For the p value calculation, first, the observed value t_0 of the test statistic is computed by fitting the model to data. Next, the expected distribution $f(t)$ of the test statistic is constructed using pseudo data (300 toys) sampled from the post-fit model obtained from the fit to data. Finally, the p value is calculated with $p = \int_{t_0}^{\infty} f(t) dt$. An example of the obtained test statistic is shown in Fig. 5.18.

The p values for the 1D and 2D GoF tests of the selected variables are summarized in Figures H.4–H.9. The data/MC comparison of the variables shows good agreement as well (see Appendix E), providing further evidence that the variables are well modeled.

The input validation procedure generally rejects a small number of variables depending on the year and the category. The final selected and validated variables for each category are presented in Tables 5.10- 5.11.

5.7.2 Neural network architecture and training

The neural networks are now trained separately in each category. A binary (signal vs background) classifier is trained in the 3 jets, ≥ 3 b tags category, while in the most sensitive

Variable	Definition	3 jets, ≥ 3 b tags	≥ 4 jets, ≥ 3 b tags
MEM	matrix element method discriminant	–	+
BLR	likelihood ratio discriminating between events with 4 b-quark jets and 2 b-quark jets	–	+
$p_T(\text{jet } 2)$	p_T of the 2. jet, ranked in jet p_T	–	+
$p_T(\text{jet } 3)$	p_T of the 3. jet, ranked in jet p_T	+	–
$p_T(\text{b-jet } 1)$	p_T of the 1. b-tagged jet, ranked in jet p_T	–	+
$p_T(\text{b-jet } 2)$	p_T of the 2. b-tagged jet, ranked in jet p_T	–	+
$p_T(\text{b-jet } 3)$	p_T of the 3. b-tagged jet, ranked in jet p_T	–	+
$p_T(\text{b-jet } 4)$	p_T of the 4. b-tagged jet, ranked in jet p_T	–	+
N_b^{loose}	number of b-tagged jets at the loose working point	–	+
H_T	scalar sum of p_T of jets	–	+
H_T^b	scalar sum of p_T of b-tagged jets	–	+
$m_{b,b}^{\min \Delta R}$	mass of pair of b-tagged jets closest in ΔR	–	+
$m_{j,b}^{\min \Delta R}$	mass of pair of jet and b-tagged jet closest in ΔR	–	+
$m_{b,b}^{\max m}$	mass of pair of b-tagged jets with largest mass	+	+
$m_{j,b,b}^{\max p_T}$	mass of jet and two b-tagged jets system with highest p_T	–	+
$p_{T;b,b}^{\min \Delta R}$	sum p_T of pair of closest b-tagged jets	+	+
$p_{T;jj}^{\min \Delta R}$	sum p_T of pair of closest jets	+	–

Table 5.10: Input variables used in the DNNs in each category. Variables used in a given category are denoted by a + and unused variables by a –.

Variable	Definition	3 jets, ≥ 3 b tags	≥ 4 jets, ≥ 3 b tags
$p_{T;j,b}^{\min \Delta R}$	sum p_T of pair of closest jet and b-tagged jet	+	–
$\Delta R_{b,b}^{\text{avg}}$	average ΔR between b-tagged jets	–	+
$\Delta R_{jj}^{\min}$	minimal ΔR between any two jets	+	–
$\Delta R_{b,b}^{\min}$	minimal ΔR between any two b-tagged jets	+	+
$\Delta \eta_{jj}^{\max}$	largest $\Delta \eta$ between any two jets	+	+
$\Delta \eta_{b,b}^{\max}$	largest $\Delta \eta$ between any two b-tagged jets	+	–
R_2^b	ratio H_2^b/H_0^b of 0th and second Fox–Wolfram moment [215] computed with all b-tagged jets	–	+
R_2	ratio H_2/H_0 of 0th and second Fox–Wolfram moment [215] computed with all jets	–	+
C^b	scalar sum of the b-tagged jet p_T divided by the sum of the energies of all jets	–	+
H_2	2nd Fox–Wolfram moment [215] computed with all jets	–	+
H_0^b	0th Fox–Wolfram moment [215] computed with all b-tagged jets	–	+
H_4^b	4th Fox–Wolfram moment [215] computed with all b-tagged jets	–	+
$d(\text{jets})$	average b-tagging discriminant value of all jets	–	+
$d(\text{jet4})$	b-tagging discriminant value of 4. jet, ranked in jet p_T	–	+
N_{Higgs}^{jj}	multiplicity of dijet pairs with invariant mass between 100 GeV and 140 GeV	+	+
$m_{j,b,b}^{\text{median}}$	median mass of pair of jets	+	–
centrality(jets+leps)	centrality of leptons and jets	+	–
centrality(tags)	centrality of b-tagged jets	–	+
sphericity(tags)	$\frac{3}{2}(\lambda_2 + \lambda_3)$, with λ_i the eigenvalues of the momentum tensor computed with jets	–	+
$m_{b,b}^{\text{closest to 125}}$	mass of pair of b-tagged jets closest to 125 GeV	+	–

Table 5.11: Input variables used in the DNNs in each category. Variables used in a given category are denoted by a + and unused variables by a –.

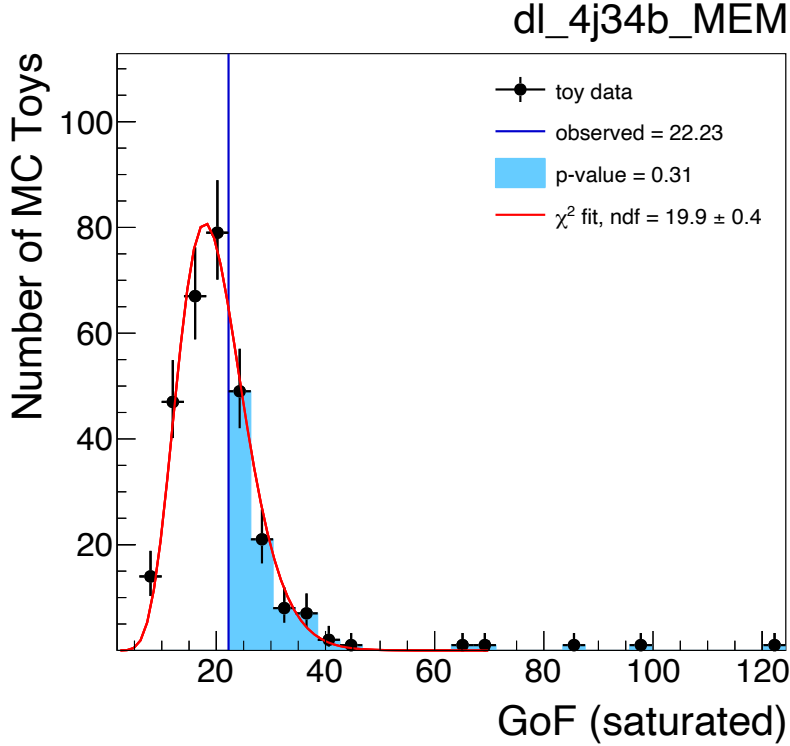


Figure 5.18: Goodness-of-Fit test for the MEM output in the ≥ 4 jets, ≥ 3 b tags category for 2017 data. Shown is the distribution $f(t)$ of the test statistic for the generated pseudo data (black dots), the values of test statistic when the observed data is evaluated (vertical blue line) and a comparison to a χ^2 distribution (red line). The p value is indicated by the shaded blue area.

≥ 4 jets, ≥ 3 b tags category, events are further separated based on a multi-class neural network with four output nodes, corresponding to the $t\bar{t}H$, $H \rightarrow b\bar{b}$, $t\bar{t} + b\bar{b}$, $t\bar{t} + c\bar{c}$ and $t\bar{t} + \text{LF}$ processes.

The signal events for the ANNs training are provided by a dedicated MC sample of $t\bar{t}H$, $H \rightarrow b\bar{b}$ events used exclusively for training, where both top quarks are forced to decay leptonically. The background samples used for training correspond to approximately half of the simulated DL $t\bar{t}$ MC sample; the other half is used to analyze the data. Each of the training samples is then randomly split into a subset for the real training (60%), a test subsample (20%) for cross-validation, and a validation set (20%) used to monitor the performances of the ANNs. The number of events used for training, evaluation and testing is summarized in Table 5.12.

Events are weighted in order to balance the impact of individual processes during the training. Additionally, before the actual training, a normalizing algorithm is applied so that all input variables have a mean value of zero and a standard deviation of one, ensuring that the value ranges of the inputs stay consistent with the domain of the chosen activation functions, as explained below.

For each of the two categories, only one single ANN is trained for all the three years of data-taking. The simulated events from all three years are weighted to reflect the different integrated luminosities. It has been checked that training one ANN valid for all three years rather than each year independently does not reduce the sensitivity of the resulting discriminants. In Table 5.13, fit results (considering all sources of systematic uncertainties) are summarized for an exemplary setup of ANNs in the most sensitive category. 95% CL

	3 jets, ≥ 3 b tags	≥ 4 jets, ≥ 3 b tags
events for training	321091	952373
events for validation	107030	317457
events for testing	107030	317457

	events per process	
$t\bar{t}H$	457731	1486906
$t\bar{t} + b\bar{b}$	40901	53618
$t\bar{t} + c\bar{c}$	16623	25809
$t\bar{t} + LF$	19897	20955

Table 5.12: Number of events used to train the ANNs per category. Also listed are the number of events per process and the number of events in the training, test, and validation subsamples.

expected upper limits are shown when using one single ANN per year or one combined ANN for all three years. The results seems to benefit from the increase in the training statistics. In any case, distinct final discriminant distributions are constructed per year during the evaluation step to constrain uncorrelated uncertainties between the years.

	2016	2017	2018	combined
3 ANNs	$1.58^{+0.72}_{-0.48}$	$1.57^{+0.69}_{-0.47}$	$1.28^{+0.55}_{-0.37}$	$1.84^{+0.37}_{-0.26}$
1 ANN	$1.54^{+0.69}_{-0.46}$	$1.55^{+0.68}_{-0.46}$	$1.30^{+0.56}_{-0.38}$	$1.82^{+0.36}_{-0.24}$

Table 5.13: 95% CL expected upper limits in an Asimov fit to pseudo data for the three data years and the combination. The fit has been evaluated when using a unique ANN per data era (3 ANNs) or one combined ANN for all three years (1 ANN). All systematic uncertainties have been included in the fit.

The architecture is a fully connected feed-forward network implemented in KERAS. The hyperparameters of such an architecture are the number of layers, the number of nodes per layer, the activation function after each layer, the optimizer and its learning rate. The number of nodes in the last layer and the last activation function must be set to the number of desired output classes and the function to be used to interpret each output as a probability, respectively. In the multi-classification training, the four output nodes are normalized to unity with a ‘softmax function, and the minimized cost function is the categorical cross entropy. A tanh output activation function and the squared hinge cost function are instead employed for the binary trainings. The ANN hyperparameters, for both the binary and multi-class classifiers, are optimized using a Bayesian approach [270, 271]. This technique allows to use information from previous trials to guide the next set of hyperparameters to explore. It generates a probability distribution function that evaluates how good a model is for a given set of hyperparameters. The hyperparameters have been optimized based on the performance on the test set. To rapidly assess the accuracy of the classifier, the maximized cost function is the approximate median significance [272] computed with the output signal and background histogram distributions. Additional requirements to minimize overtraining are applied already during the optimization procedure: the Kolmogorov-Smirnov test results between training and testing output distributions (both for signal and background) are required to be ≥ 0.2 . This optimization is carried out independently for each of the two

categories. The final ANNs are trained with the architectures summarized in Table 5.14.

	3 jets, ≥ 3 b tags	≥ 4 jets, ≥ 3 b tags
Optimizer	Adagrad	
Learning rate	$2 \cdot 10^{-4}$	$9 \cdot 10^{-4}$
Loss function	squared hinge	categorical cross-entropy
Number of input variables	13	29
Hidden Layers	3	
Nodes per Hidden Layer	[682,258,253]	[512,256,128]
Node Activation	leakyrelu	
Output Node activation	tanh	Softmax
L2 regularization	$2 \cdot 10^{-5}$	$7 \cdot 10^{-4}$
L1 regularization	10^{-5}	10^{-5}
Dropout Percentage	0.03	0.05
Early stopping min epochs	50	50
Batchsize	400	2048

Table 5.14: Hyperparameters of the neural networks per category.

Every training is performed over a maximum of 1000 epochs but some of the training don't reach such a number of epochs because an early-stopping criterion is applied to limit overtraining. Effectively, every training terminates when the performance on the validation set does not increase for 100 consecutive epochs. To reduce overtraining even further, both L1-L2-regularization and dropout are used, with their numerical values determined by the Bayesian optimization technique. Furthermore, the cost function value is required to agree within 2% or less between training and validation samples.

Figure H.10 shows two examples of the evolution of the ANN cost function as a function of the number of epochs, or the training and validation data sets.

5.7.3 Evaluation of training results

One commonly used method to demonstrate the classification ability of ANNs is through the Receiver Operating Characteristic (ROC) curve. This curve represents the accuracy of correctly identifying events as a proportion of events assigned to the incorrect class. In the context of high-energy physics, it reflects the accuracy of correctly identifying signal events versus incorrectly classifying background processes. The separation power of the ANN is further quantified by the Area Under the ROC curve (ROC-AUC), which is reported in Table 5.15. In addition, the thesis employs the use of confusion matrices, which serve as a gauge for classification efficiency and the probability of misidentification for each class. A confusion matrix shows ideal discrimination between classes when the diagonal entries are one and all other entries are zero. The confusion matrix for the 4-class ANN used in the ≥ 4 jets, ≥ 3 b tags category for three years can be found in Figure 5.19.

The outputs of the ANN discriminant (normalized to arbitrary units) for the different categories can be found in Figure 5.20. Additionally, the result values of the Kolmogorov-Smirnov test between training and testing distributions are also reported on the plots.

The highest value of the four output nodes for each event is used to allocate the event to a sub-category. The amount of signal and major background events for each category is represented in the pie charts in Figure 5.21 for the different years separately.

	3 jets, ≥ 3 b tags	≥ 4 jets, ≥ 3 b tags
ROC AUC	0.74	0.81

Table 5.15: Area under the curve of the receiver-operator-characteristic (ROC AUC) for the binary ANNs of the 3 jets, ≥ 3 b tags category and for the multiclass ANNs in the ≥ 4 jets, ≥ 3 b tags category.

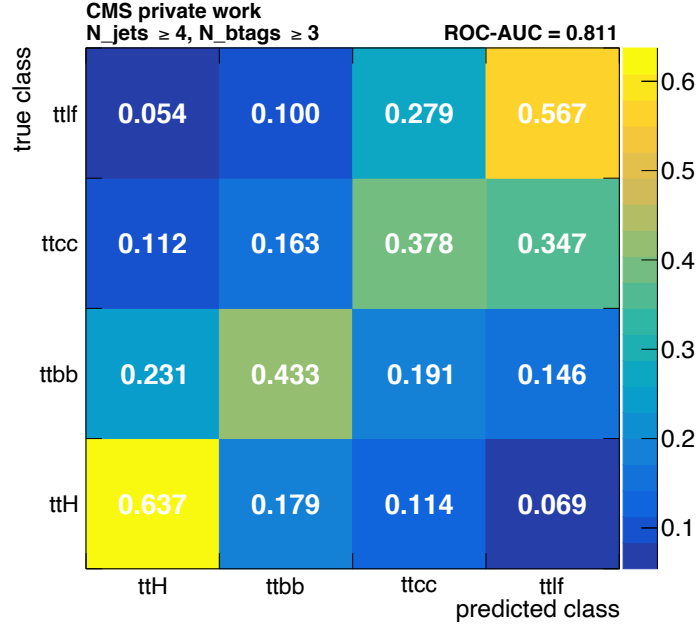


Figure 5.19: Confusion matrix after the ANN training of the 4-node ANN trainings in the ≥ 4 jets, ≥ 3 b tags category evaluated with the test data set. Quoted values correspond to the fraction of events of process X assigned to the output node (event sub-category) targeting process Y. The values are normalized to one per row, hence representing the classification accuracies (diagonal elements) and misclassification rates (off-diagonal elements) per process.

Then, the ANN output value $o(i)$ of the node i , corresponding to the process category, is employed directly as final discriminant for all events that fall in the $t\bar{t} + c\bar{c}$ and $t\bar{t} + \text{LF}$ nodes. To exploit more information of the ANN classification, only for the events in the $t\bar{t}H$ and $t\bar{t} + b\bar{b}$ process categories of the ≥ 4 jets, ≥ 3 b tags category, the final discriminant is a likelihood ratio computed from the ANN output values defined as

$$p_o = \frac{o(t\bar{t}H)}{o(t\bar{t}H) + o(ttbb)}, \quad (5.2)$$

where $o(i)$ is the ANN output value in the node i . This leads to a better separation of the $t\bar{t}H$ signal from the critical $t\bar{t} + b\bar{b}$ background, which is discussed further in Appendix I.

This approach leads to a total of four exclusive discriminant distributions:

- 1 likelihood ratio distribution p_o for the events in the $t\bar{t}H$ and $t\bar{t} + b\bar{b}$ process categories in ≥ 4 jets, ≥ 3 b tags category,
- 2 ANN output value distributions $o(i)$ for the events in the $t\bar{t} + c\bar{c}$ and $t\bar{t} + \text{LF}$ of the ≥ 4 jets, ≥ 3 b tags category,

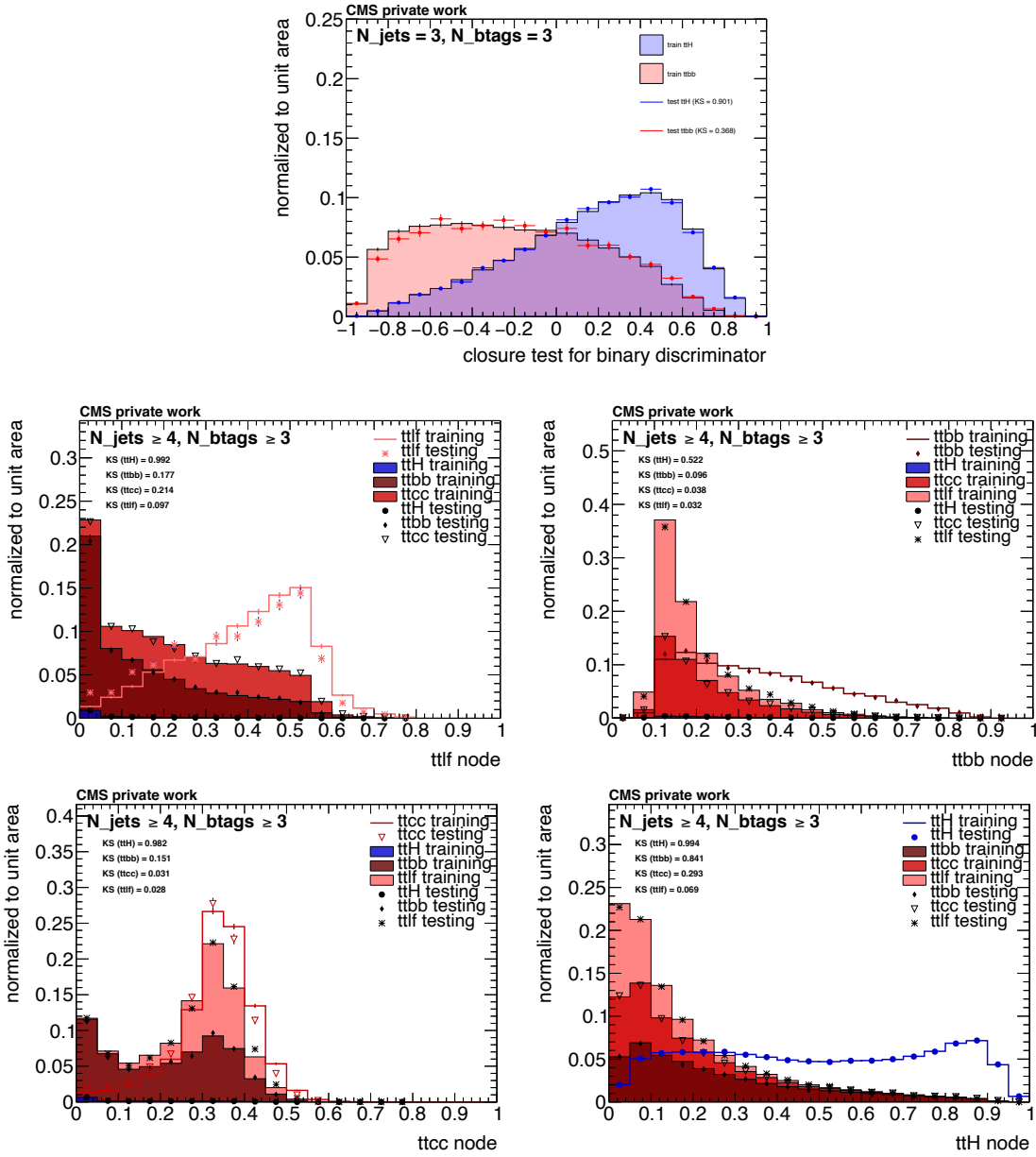


Figure 5.20: Distributions of the ANN binary classifiers in 3 jets, ≥ 3 b tags category (upper row) and of the multiclass ANNs in the ≥ 4 jets, ≥ 3 b tags category.

- and 1 binary ANN output distribution for the 3 jets, ≥ 3 b tags category,

which are ultimately combined in the statistical analysis for the signal extraction.

Figures 5.22-5.24 show the distributions of the discriminants for each separate year of data taking. Good data/MC agreement is observed overall, and the normalization differences are attributed to the underprediction of the inclusive $t\bar{t} + b\bar{b}$ cross section.

5.7.4 The other channels

The analysis of the DL channel presented in details in this thesis is only one of the input channels of the full $t\bar{t}H$, $H \rightarrow b\bar{b}$ analysis. Also the other decay channels of the $t\bar{t}$ system are considered: the fully-hadronic (FH) and the semileptonic (SL) channels. The whole analysis strategy is illustrated in Figure 5.25. The optimization has been performed separately in each

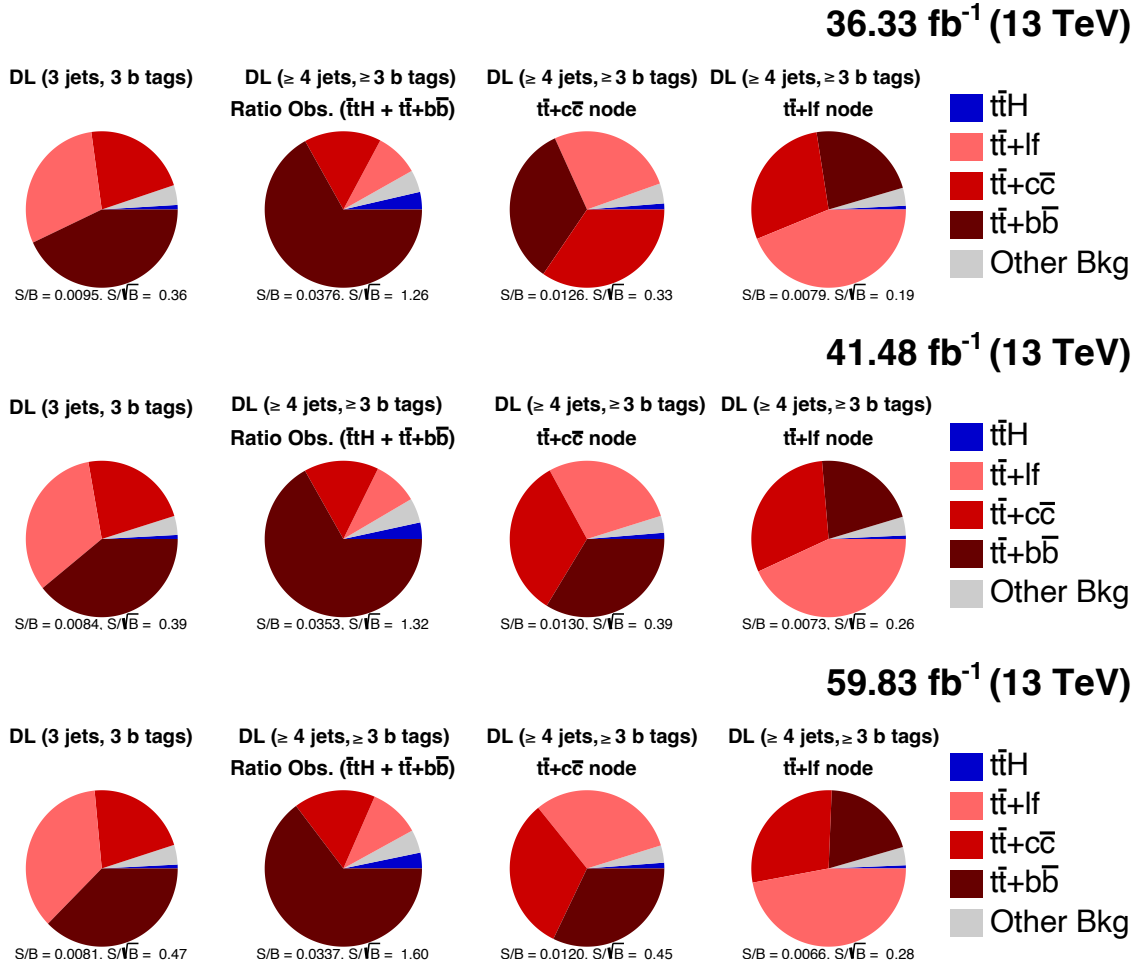


Figure 5.21: Pie charts with content of signal and background events for each category in 2016 (top row), 2017 (middle row) and 2018 (bottom row). The value of signal to background ratios are reported for each category.

channel based on the expected sensitivity to a SM signal evaluated with simulated data. The ultimate results for the full combination will be presented in Section 6.3.

In both channels the selected events are also split into categories based on the jet and b-tagged jet multiplicity. Dedicated ANNs are optimized to enhance the separation between the signal and the background processes.

In the FH channel, three categories with ($7\text{ jets}, \geq 4\text{ b tags}$), ($8\text{ jets}, \geq 4\text{ b tags}$), and ($\geq 9\text{ jets}, \geq 4\text{ b tags}$) are used for the signal extraction. In each signal region, a binary-classification valid for all three data-taking periods separates the $\bar{t}\bar{t}H$ signal from the background events; this choice was found to provide sufficient sensitivity for the measurement. The QCD multi-jet production, referred to as QCD, represent the dominant background; to avoid the implications of poor modeling and too few events in the relevant phase-space areas of the available QCD simulation, it is modeled from the data. Independently for each signal region, the yield is left as free parameter in the final fit. For each signal region, two mutually exclusive control regions with loser b-tagging requirements are defined and used to estimate the QCD-multijet background and to train the ANNs. As for DL, the ANN training is performed with simulated $\bar{t}\bar{t}H$ events, where only Higgs boson decays to $b\bar{b}$ are considered. The ANNs deploy input variables related to the kinematic properties of individual objects and event shape information as well as the MEM discriminant output. The input variable

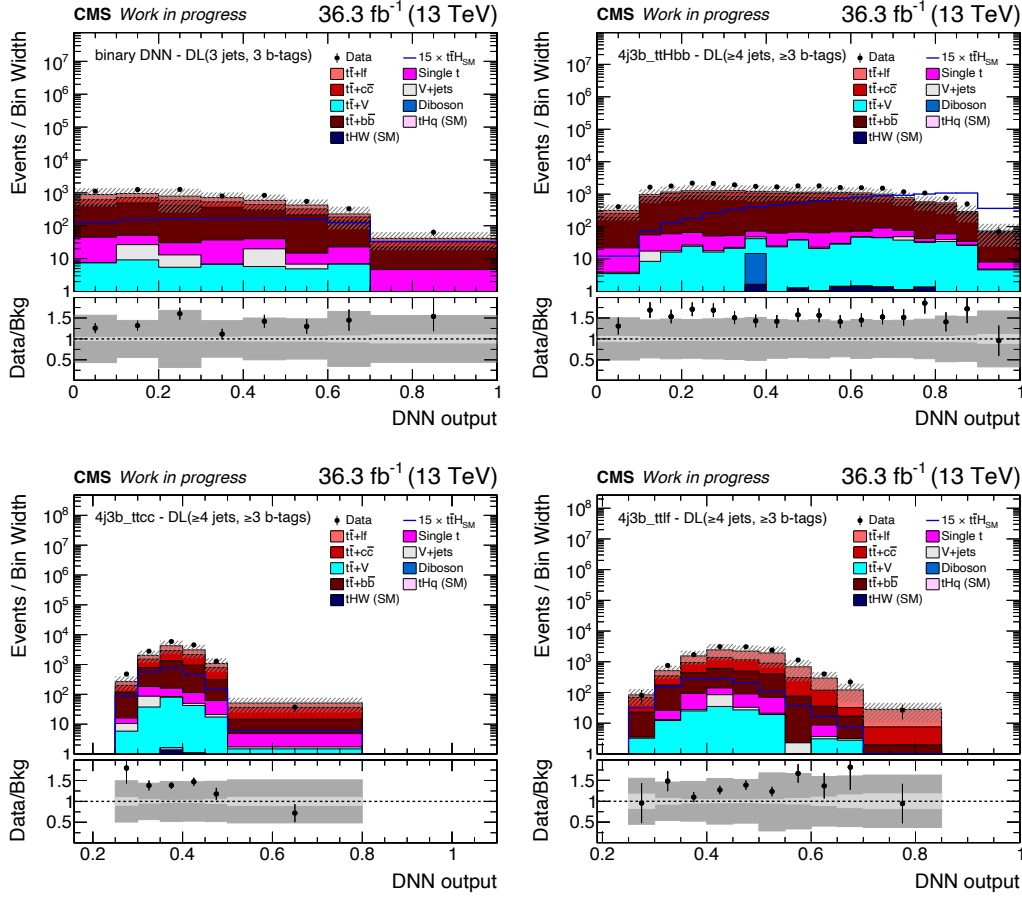


Figure 5.22: Discriminants distributions in the 3 jets, ≥ 3 b tags (top left), ≥ 4 jets, ≥ 3 b tags $t\bar{t}H + t\bar{t} + b\bar{b}$ ratio observable (top right), ≥ 4 jets, ≥ 3 b tags $t\bar{t} + c\bar{c}$ (bottom left) and ≥ 4 jets, ≥ 3 b tags $t\bar{t} + LF$ (bottom right) categories for 2016. The expected background contributions (filled histograms) are stacked, and the expected signal distribution (line) is superimposed. The black points represent the observed data. The hatched uncertainty bands correspond to the total statistical and systematic uncertainties with a shape changing effect, added in quadrature. The lower plots show the ratio of the data to the background prediction.

distributions are verified to model well the data based on a χ^2 criterion. The classifier output distribution is then directly used as the final discriminating observable.

In the SL channel, two categories with (≥ 6 jets, ≥ 4 b tags) and (5 jets, ≥ 4 b tags) are considered. The output of multi-classification ANNs is used to further categorize events; as in DL, the output value $o(i)$ are normalized to unity to be interpreted as probabilities of the event being of a certain process i . There are three signal processes $t\bar{t}H$, tHq , and tHW , and the four $t\bar{t} + \text{jets}$ background processes $t\bar{t} + LF$, $t\bar{t} + C$, $t\bar{t} + 2b$, and $t\bar{t} + B$. For events in the $t\bar{t}H$ and $t\bar{t} + b(\bar{b})$ process categories, a likelihood ratio discriminant is computed as final discriminant observable from the ANN output values as:

$$p_o^{\text{SL}} = \frac{o(t\bar{t}H)}{o(t\bar{t}H) + o(t\bar{t} + b(\bar{b})) + o(t\bar{t} + 2b)}.$$

In all of the remaining process categories, the ANN classifier output distribution of the node that matches the process category is used as the final discriminating observable, where in several categories effectively only the event yield is considered. In total, this leads to 12 categories in the SL channel per data-taking period, see Fig. 5.25.

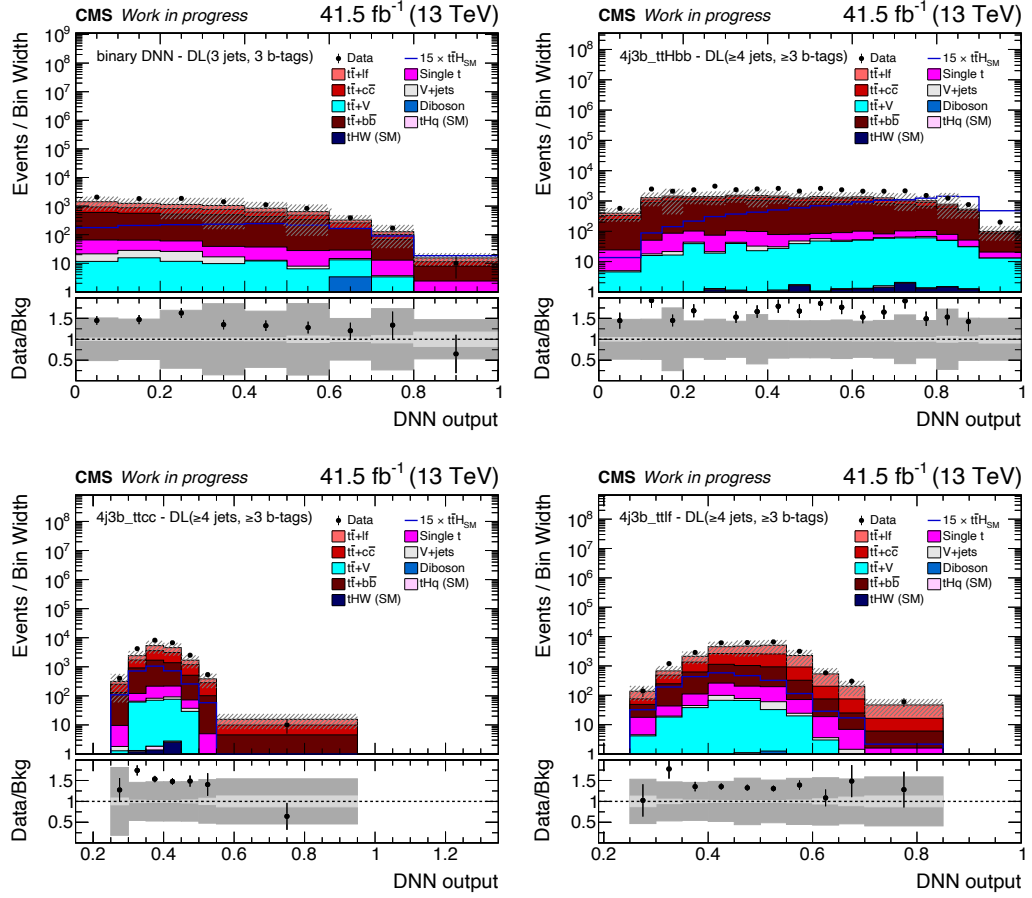


Figure 5.23: Discriminants distributions in the 3 jets, ≥ 3 b tags (top left), ≥ 4 jets, ≥ 3 b tags $t\bar{t}H + t\bar{t} + b\bar{b}$ ratio observable (top right), ≥ 4 jets, ≥ 3 b tags $t\bar{t} + c\bar{c}$ (bottom left) and ≥ 4 jets, ≥ 3 b tags $t\bar{t} + LF$ (bottom right) categories for 2017. The expected background contributions (filled histograms) are stacked, and the expected signal distribution (line) is superimposed. The black points represent the observed data. The hatched uncertainty bands correspond to the total statistical and systematic uncertainties with a shape changing effect, added in quadrature. The lower plots show the ratio of the data to the background prediction.

Additional variables constructed using information from the BDT-based event reconstruction are used as inputs. As in the DL channels, the input variables and their correlations have been validated to model well the data, based on a goodness-of-fit test with the full uncertainty model. In each category a single ANN is trained for all three data-taking periods weighted to reflect the different integrated luminosities to reduce statistical fluctuations. Only events with Higgs boson decays to $b\bar{b}$ are considered, and the $t\bar{t}$ system is required to decay to one lepton; for the tH events, only those with leptonic decays of the top quark are used. However, while evaluating the ANNs, distinct final discriminant distributions are constructed per data-taking period to allow constraining uncertainties that are uncorrelated between the periods. It has been thoroughly proven that the sensitivity of the analysis does not degrade when compared to training ANNs independently for each data-taking session.

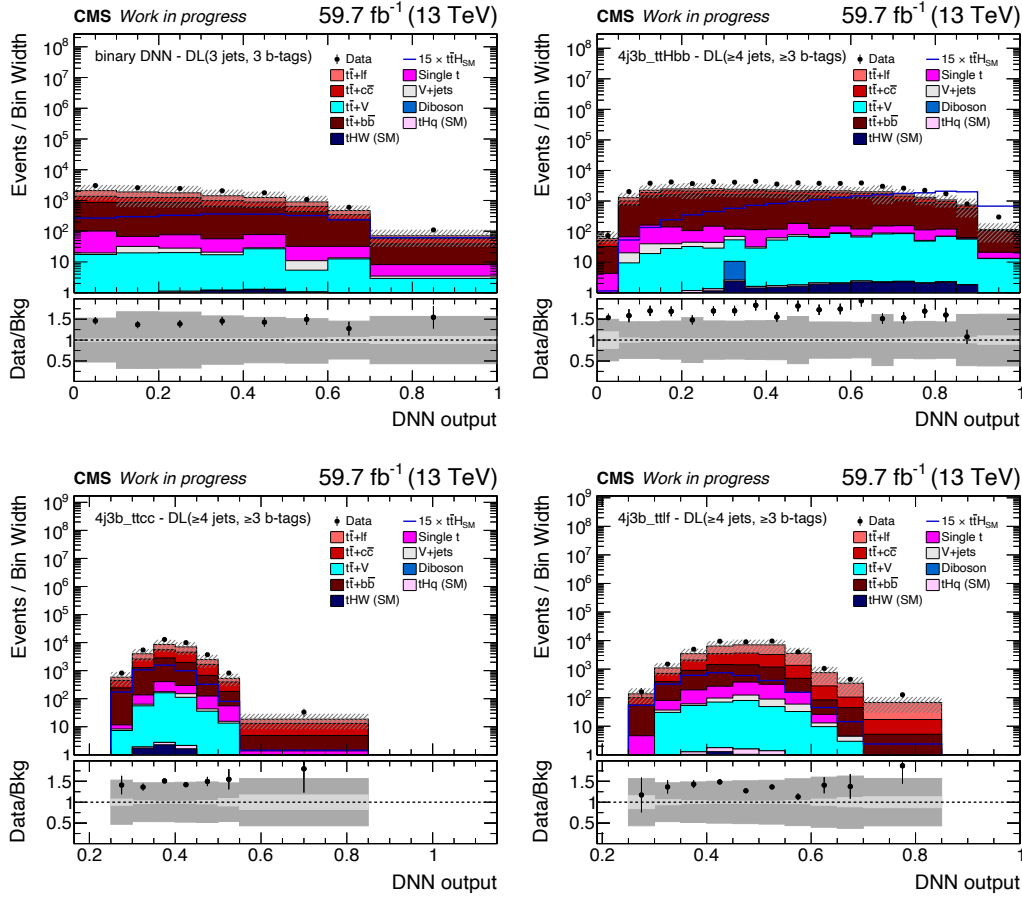


Figure 5.24: Discriminants distributions in the 3 jets, ≥ 3 b tags (top left), ≥ 4 jets, ≥ 3 b tags $t\bar{t}H + t\bar{t} + b\bar{b}$ ratio observable (top right), ≥ 4 jets, ≥ 3 b tags $t\bar{t} + c\bar{c}$ (bottom left) and ≥ 4 jets, ≥ 3 b tags $t\bar{t} + LF$ (bottom right) categories for 2018. The expected background contributions (filled histograms) are stacked, and the expected signal distribution (line) is superimposed. The black points represent the observed data. The hatched uncertainty bands correspond to the total statistical and systematic uncertainties with a shape changing effect, added in quadrature. The lower plots show the ratio of the data to the background prediction.

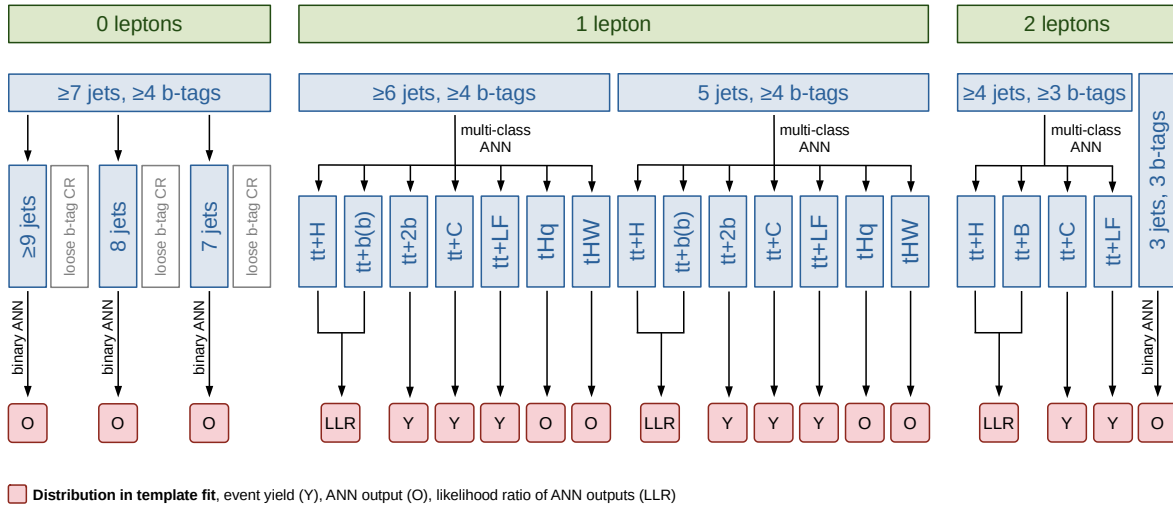


Figure 5.25: Illustration of the full $t\bar{t}H, H \rightarrow b\bar{b}$ analysis strategy for the inclusive measurement per data-taking period.

MEASUREMENT OF $t\bar{t}H$, $H \rightarrow b\bar{b}$ - ANALYSIS RESULTS

6.1	Uncertainty model	117
6.1.1	Experimental uncertainties	118
6.1.2	Theory and background modeling uncertainties	125
6.2	Measurement results	130
6.2.1	Distribution and yields	132
6.2.2	Goodness-of-Fit tests	132
6.2.3	Nuisance pulls and impacts	132
6.2.4	Parameter of interest estimation	135
6.3	Contribution to the combined analysis	138

This chapter presents the results of the inclusive measurement for $t\bar{t}H, H \rightarrow b\bar{b}$ production. The chapter begins with a discussion of the statistical model and the sources of uncertainties. This model provides the foundation for extracting the signal strength modifier in a fit to the data. The results are then presented, with a focus on the impact of individual sources of systematic uncertainties, such as the uncertainty on the modeling of the background processes. In the last section of the chapter, the contribution of the dilepton channel is discussed in the context of the combination with other channels (single lepton and fully hadronic). This includes a description of the methods used to combine the results from the different channels, and the impact of the dilepton channel on the overall sensitivity of the search.

6.1 Uncertainty model

This analysis focuses on the signal-strength modifier $\mu_{t\bar{t}H}$, defined as $\mu_{t\bar{t}H} = \sigma_{t\bar{t}H}^{\text{obs}} / \sigma_{t\bar{t}H}^{\text{SM}}$, the ratio of the measured cross section $\sigma_{t\bar{t}H}^{\text{obs}}$ to the value predicted by the SM $\sigma_{t\bar{t}H}^{\text{SM}}$. The signal extraction is based on a fit to data of a simulated parameterized variable distribution, given that an excess in the invariant mass spectrum of two b-tagged jets is not expected to be observable. In principle, any distribution of any arbitrary variable can be fitted to data, as long as it has the necessary discrimination power to separate signal from background. The model parameters derived in the fit include both the parameter of interest, namely the $t\bar{t}H$ signal strength modifier $\mu_{t\bar{t}H}$ as well as any other systematic uncertainties, known as the nuisance parameters. $\mu_{t\bar{t}H}$ is largely sensitive to the latter, especially in the context of such a background dominated analysis, since any discrepancy in the modeling of the data by the simulation is initially adjusted by shifting the central values of the nuisances. The parameter of interest (POI) and the collection of nuisance parameters (NPs) are the parameters with respect to which the test statistic is minimized and taken into account in the profile maximum-likelihood fit (see Appendix B).

The systematic uncertainties that are relevant for the presented analysis are discussed in the following. These can have an effect on both signal and background processes and are

correlated when necessary, i.e. the effects from the same source are treated as fully correlated. The probability of a specific nuisance is described by an a priori distribution, which can depend on a particular bin and/or a physics process. When there is no dependence on the bin, the nuisances only affect the overall yield of a process and are called rate uncertainties. The change in yield caused by the variations of these parameters follow a log-normal prior probability density function. Alternatively, uncertainties that impact both the yield and the shape of distributions are shape uncertainties. Internally, yield variations are treated as rate uncertainties. The shape variation, instead, is implemented as an interpolation between distributions representing a ± 1 standard deviation (std. dev.) change of the appropriate uncertainty source, as discussed in Appendix B. It should be mentioned that, rate changing nuisances can modify the composition of a process within a given category, resulting in effective changes of shape distributions.

The relevance of uncertainty sources varies with analysis, phase-space, and physical objects. Due to the high b-jet multiplicity required in this in this study, the uncertainties related to b-jet identification are expected to have a considerable effect on the measurement.

The correlation between uncertainty sources has to be addressed to ensure the correct treatment in profiled maximum-likelihood fits; the whole Run 2 data set analyzed here includes three data sets recorded from 2016 to 2018, so each data collection has distinct features that must be considered when assessing the correlation among the sources of uncertainty. All sources related to the statistical uncertainties are uncorrelated among eras, but systematic uncertainties that originate in the methodology of the derivation of a specific quantity are generally treated as fully correlated.

The considered sources of uncertainties relevant to the $t\bar{t}H, H \rightarrow b\bar{b}$ measurement are summarized in Table 6.1. Partial correlation means that some of the sub-sources of the systematic uncertainty group are fully correlated while others are not, as it will be explained in detail later in this Section. Their impact on the final result is discussed in Section 6.2.

6.1.1 Experimental uncertainties

Experimental sources of systematic uncertainties are those that originate from detector effects such as misidentification of particles, inefficiencies, and limited resolutions of detector sub-systems. This analysis takes into account uncertainties in the measurement of the integrated luminosity, the calibration of the jet energy scale and resolution, corrections applied to event weights and the limited number of simulated events that affect the precision of predicted distributions. They are detailed in the following.

Integrated luminosity

The uncertainties on the instantaneous and integrated luminosity estimates per year is provided in Table 6.2 and is required to accurately normalize the simulated samples for the final comparison to the observed data set. The uncertainties are assessed following the recommendation of the dedicated group within the CMS Collaboration. This has an impact on the rates of all processes; as advised, the uncertainties are partially correlated throughout the years.

Lepton scale-factors

p_T - and η - dependent lepton scale factors used as event weights are employed to increase the compatibility for the tracking (only electrons), reconstruction, ID, and isolation efficiency of the simulated samples with the observed data. The uncertainties in the calculation of these scale factors are taken into account in accordance with the recommendations provided by the CMS groups in charge of muon and electron reconstruction.

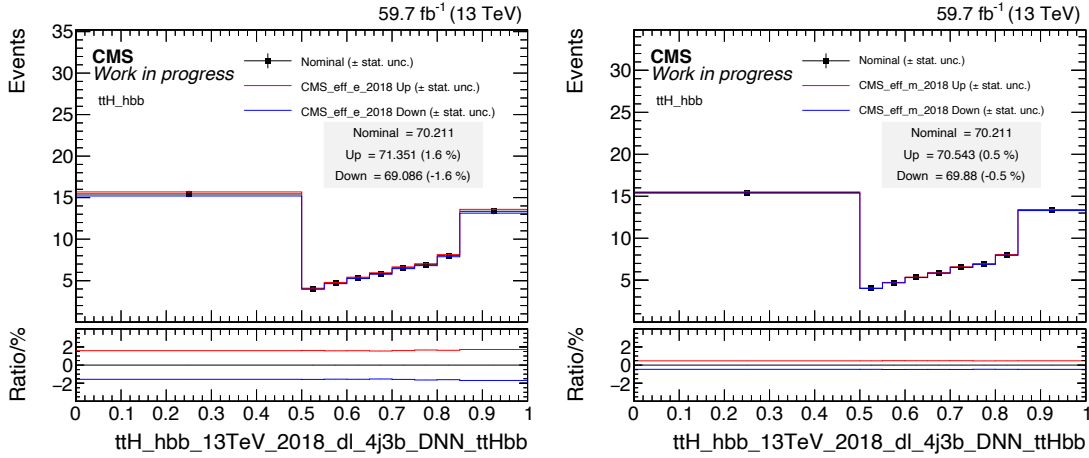
Source	Type	Correlation	Remarks
Integrated luminosity	R	partially	Signal and all backgrounds
Lepton ID/Iso	S	uncorrelated	Signal and all backgrounds
Trigger efficiency	S	uncorrelated	Signal and all backgrounds
L1 prefireing correction	S	uncorrelated	Signal and all backgrounds
Pileup	S	correlated	Signal and all backgrounds
Jet energy scale (11 scheme)	S	partially	Signal, $t\bar{t}$ + jets and Single-t
Jet energy resolution	S	uncorrelated	Signal, $t\bar{t}$ + jets and Single-t
b-tag HF/LF fraction	S	partially	Signal and all backgrounds
b-tag HF/LF stat (linear)	S	uncorrelated	Signal and all backgrounds
b-tag HF/LF stat (quadratic)	S	uncorrelated	Signal and all backgrounds
b-tag charm (linear)	S	partially	Signal and all backgrounds
b-tag charm (quadratic)	S	partially	Signal and all backgrounds
Size of the MC samples	S	uncorrelated	Statistical uncertainty of signal and background prediction due to limited sample size
Renorm./fact. scales	R	correlated	Scale uncertainty of (N)NLO prediction, independent for $t\bar{t}H$, tHq , tHW , $t\bar{t}$, t , V , VV
PDF+ α_s (gg)	R	correlated	PDF uncertainty for gg initiated processes, independent for $t\bar{t}H$, tHq , tHW , and others
PDF+ α_s (qq)	R	correlated	PDF uncertainty of qq initiated processes ($t\bar{t}W, W, Z$) except tHq
PDF+ α_s (gq)	R	correlated	PDF uncertainty of gq initiated processes (single t) except tHW
Collinear gluon splitting	R	correlated	Additional 100% rate uncertainty on $t\bar{t}$ + 2b
μ_R scale	S	correlated	Renormalization scale uncertainty of the ME generator, independent for $t\bar{t}H$, tHq , tHW , $t\bar{t}$ + B (4FS sample), other $t\bar{t}$ (5FS sample)
μ_F scale	S	correlated	Factorization scale uncertainty of the ME generator, independent for $t\bar{t}H$, tHq , tHW , $t\bar{t}$ + B (4FS sample), other $t\bar{t}$ (5FS sample)
PDF shape	S	correlated	From NNPDF variations, independent for $t\bar{t}$ + B (4FS sample), tHq , tHW and $t\bar{t}H$ /other $t\bar{t}$ (5FS sample)
PS scale ISR	S	correlated	Initial state radiation uncertainty of the PS (PYTHIA), independent for $t\bar{t}H$, $t\bar{t}$ + B (4FS sample), other $t\bar{t}$ (5FS sample)
PS scale FSR	S	correlated	Final state radiation uncertainty of the PS (PYTHIA), independent for $t\bar{t}H$, $t\bar{t}$ + B (4FS sample), other $t\bar{t}$ (5FS sample)
ME-PS matching ($t\bar{t}$)	R	correlated	NLO ME to PS matching (for $t\bar{t}$ + jets events), independent for additional jet flavors
Underlying event ($t\bar{t}$)	R	correlated	Underlying event (for all $t\bar{t}$ + jets events)
STXS migration	R	correlated	Signal, only in STXS measurement
STXS acceptance	S	correlated	Signal, only in STXS measurement

Table 6.1: Systematic uncertainties considered in the analysis. Type refers to rate (R) or shape (S) uncertainties. Correlation indicates whether the uncertainty is treated as correlated, partially correlated (as detailed in the main body), or uncorrelated across the years 2016–18.

Year	2016 in %	2017 in %	2018 in %
Uncorrelated Fits	1.2 [26]	2.3 [273]	2.5 [274]
Uncorrelated 2016	1.0	—	—
Uncorrelated 2017	—	2.0	—
Uncorrelated 2018	—	—	1.5
Correlated 2016, 2017, 2018	0.6	0.9	2.0
Correlated 2017, 2018	—	0.6	0.2

Table 6.2: Luminosity uncertainty in percent per year. In the fits for the individual years, the luminosity uncertainty is implemented as shown in the first row. For the combined fit corresponding to 137.60 fb^{-1} , the uncertainties are partially correlated.

The uncertainty in the lepton scale factor has no effect on the final outcome. As a result, the uncertainties are propagated to the final discriminant distributions as a total shape uncertainty, which corresponds to the quadratic sum of all statistical and systematic components, independently for electrons and muons. The scale factors are calculated for each year separately and are hence decorrelated between years. As an example, the effect on ratio observable discriminant for the signal process in 2018 is shown in Figure 6.1 (a) and (b) for electron and muons, respectively. The trend of similar effects holds true for the other two years as well.



(a) Combined electron efficiency uncertainty.

(b) Combined muon efficiency uncertainty.

Figure 6.1: Effect of uncertainty variations on the ratio observable discriminant in the ≥ 4 jets, ≥ 3 b tags category for the signal $t\bar{t}H$, $H \rightarrow b\bar{b}$ process in 2018. The bottom ratio plot shows relative shape variations. Changes of the total expected yield are quoted for the respective variation.

Trigger efficiency

The uncertainties in the derivation of the trigger scale factors introduced in Section 5.4 are also considered in the statistical model. Their impact on the final discriminants is estimated by varying the trigger scale factors within their uncertainties. The effect on the ratio observable discriminant in 2018 for the ≥ 4 jets, ≥ 3 b tags category for the $t\bar{t} + b\bar{b}$ background process is shown in Figure 6.2 (a). The trend of consistent results extends to the data from the other two years as well. Similar to the lepton scale factors, the trigger scale factors are derived for

each year individually. Consequently, these uncertainties are treated as uncorrelated across the years in the final fit.

L1 prefiring correction

The impacts of the L1 prefiring problem presented in Section 2.2.7 are corrected by adding special scale factors to the simulated events. The uncertainties on this corrections are accounted for by propagating the specified uncertainties on the event weights to the scale factors. The L1 prefiring adjustment is decorrelated between the data sets from 2016 and 2017 and does not exist in 2018.

Pileup

As mentioned in Section 5.4, the simulated sample distribution of pileup events is reweighted to enhance agreement with the observed distribution. To assess the effects of uncertainty in the distribution of the number of pileup interactions, the cross section used to predict the number of pileup interactions in the MC is varied by 4.6% from its nominal value. The changes in the weight factors are propagated to the final discriminant distributions and considered as rate and shape variation in the final fit. Because the underlying technique and its inputs are the same, the uncertainty is treated as fully correlated among all processes and years. The effect on the ratio observable discriminant in 2018 for the ≥ 4 jets, ≥ 3 b tags category for the $t\bar{t} + b\bar{b}$ background process is shown in Figure 6.2 (b). Similar effects are observed in the other two years.

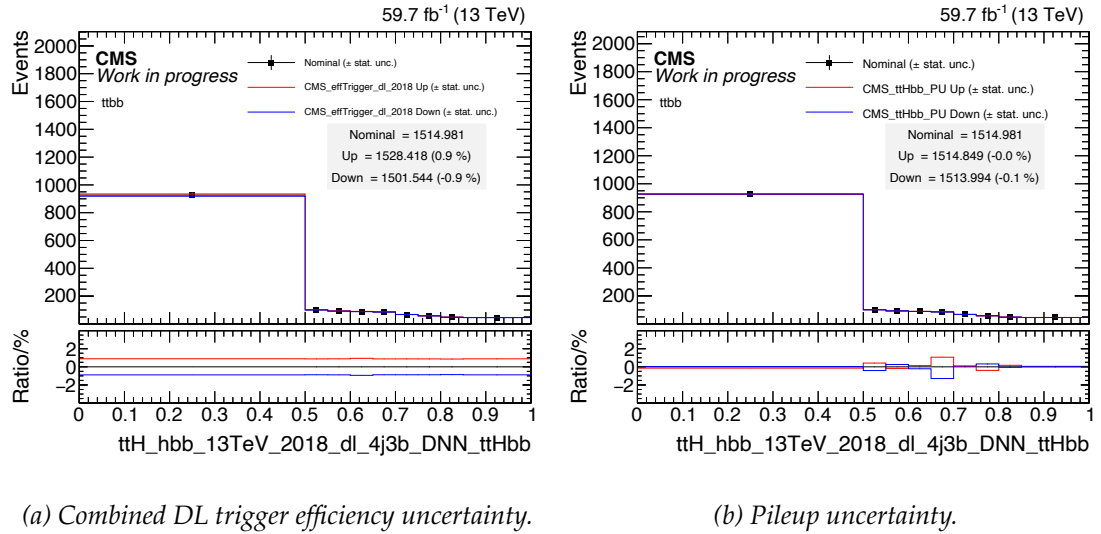


Figure 6.2: Effect of uncertainty variations on the ratio observable discriminant in the ≥ 4 jets, ≥ 3 b tags category for the background $t\bar{t} + b\bar{b}$ process in 2018. The bottom ratio plot shows relative shape variations. Changes of the total expected yield are quoted for the respective variation.

Jet energy scale

The JES is corrected to make usage of the latest calibrations of the detector, as presented in Section 5.4, but each part of this procedure is subject to various sources of uncertainties. These JES uncertainties are evaluated by adjusting the jet energy scale applied to the reconstructed jets up and down by one standard deviation. The events are then re-analyzed, including re-applying the jet-based selection and recalculating all kinematic parameters, to provide the corresponding final discriminant distributions.

They are considered as shape uncertainties in the final fit. The correlation of these sources across the different years follows the recommendation by the dedicated group within the CMS Collaboration.

Additionally, a dedicated uncertainty is applied in 2018 to account for the minor effects due to the HEM issue described in Section 2.2.7.

The effects on the ratio observable discriminant in 2018 for the ≥ 4 jets, ≥ 3 b tags category for the $t\bar{t} + b\bar{b}$ background process are depicted in Figures 6.2 (a), (b) and (c). Similar effects are observed in the other two years.

The large extent of the statistical uncertainties in the final varied distribution can cover possible shape variations and it is known that they can also artificially constrain the corresponding parameters in the estimation, therefore, these uncertainties are not applied to processes with negligible contributions in the analysis phase space. This does not alter the sensitivity of the signal strength modifier and the corresponding nuisance parameters.

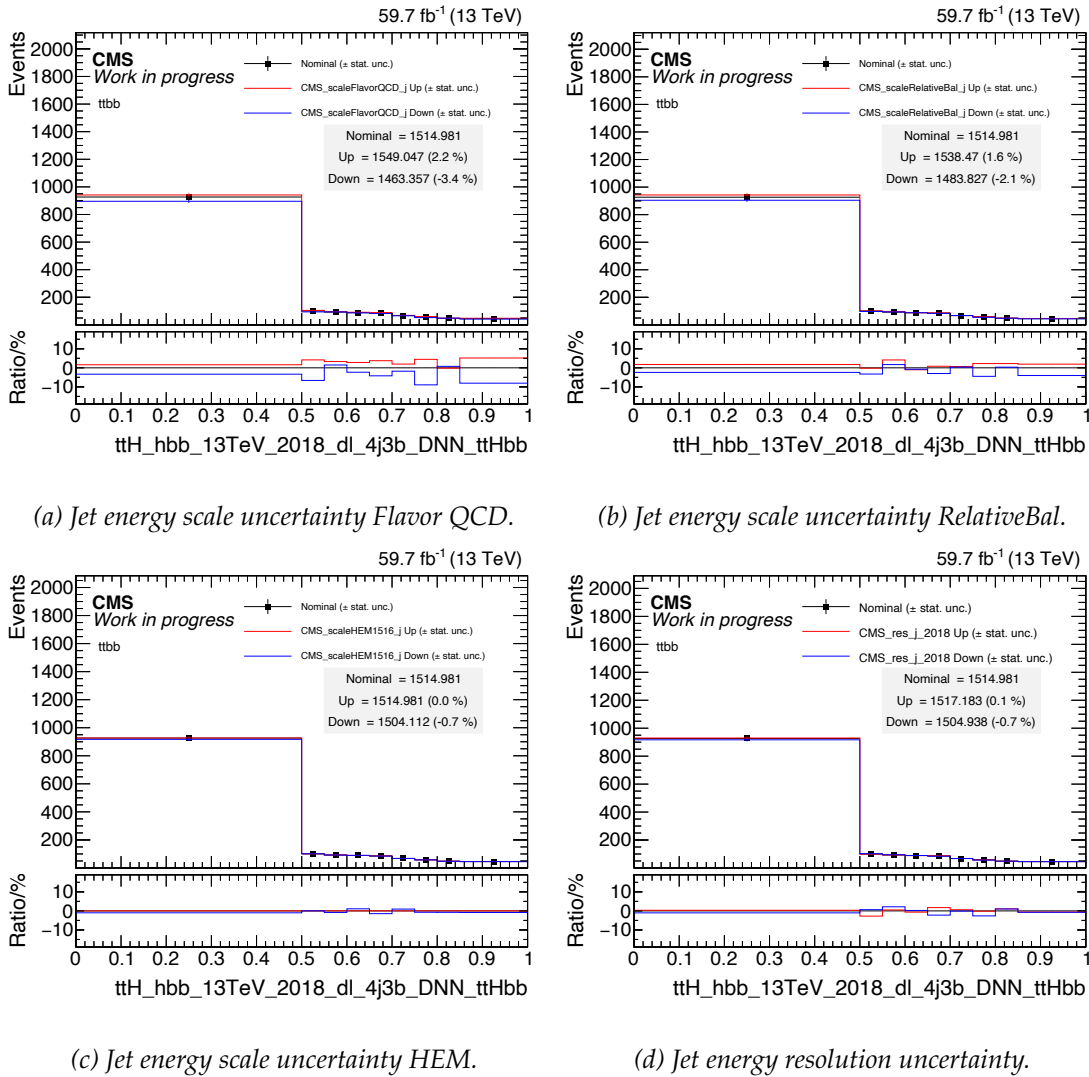


Figure 6.3: Effect of uncertainty variations on the ratio observable discriminant in the ≥ 4 jets, ≥ 3 b tags category for the background $t\bar{t} + b\bar{b}$ process in 2018. The bottom ratio plot shows relative shape variations. Changes of the total expected yield are quoted for the respective variation.

Jet energy resolution

The systematic uncertainty due to the JER is assessed by increasing/decreasing the difference between the reconstructed and truth-level jet energy in accordance with the CMS Collaboration’s prescription. To provide varied final discriminant distributions, the events are re-analyzed, including re-applying the jet-based selection and recalculating all kinematic parameters. This source is treated as a shape uncertainty in the final fit, which is completely decorrelated over the years. Its effect on the ratio observable discriminant in 2018 for the ≥ 4 jets, ≥ 3 b tags category for the $t\bar{t} + b\bar{b}$ background process is depicted in Figure 6.2 (d). Similar effects are observed in the other two years.

Similarly to JES uncertainties, the magnitude of statistical uncertainty in JER sources might be strongly dependent on the considered process; therefore, the treatment for this collection of uncertainty is the same as that described before for the JES uncertainties. It has been shown that this modification does not impact the sensitivity and constraints of the signal strength modifier and the corresponding nuisance parameters.

b-tag scale factors

The approach outlined in Section 5.4 for determining the SFs for the b-tag score is subject to three sources of systematic uncertainty on both the heavy flavor and light flavor scale factors:

Jet Energy Corrections: The b-tag uncertainty associated with the different JEC sources is evaluated at the same time of the corresponding source of JEC uncertainty (JES or JER). When the jet kinematics are shifted up or down by 1σ according to the JEC, the b-tag scale factor values are also shifted accordingly.

purity (HF/LF fraction): The uncertainties on the purity of the control samples used to derive the scale factors are controlled by a separate nuisance parameter for light- and for heavy-flavor jets. Shifting this parameter to $\pm 1\sigma$ corresponds to a change of the heavy-flavor (light-flavor) SF as caused by a higher or lower contamination by light-flavor (heavy-flavor) jets in the control sample. Due to the iterative nature of the scale factor calculation, this also has a small effect on the light-flavor (heavy-flavor) SF. The purity is estimated using the same MC simulation, and the uncertainty is correlated between 2017 and 2018. It is left uncorrelated with 2016 to allow for differences caused by the upgrade of the pixel detector.

HF/LF statistics: Two nuisance parameters for heavy-flavor jets and two nuisance parameters for light-flavor jets are utilized to control the impact of statistical uncertainty associated with the sample size used to derive the SF. These parameters correspond to the degrees of freedom in the interpolation of the SFs described in Section 5.4. The statistical uncertainties are treated as decorrelated between the years since they originate in the size of independent control samples.

Each of these systematic variations is considered separately. The effects on the ratio observable discriminant in 2018 for the ≥ 4 jets, ≥ 3 b tags category for the $t\bar{t} + b\bar{b}$ background process are depicted in Figure 6.4. Similar effects are observed in the other two years.

Due to the challenges involved in building an adequate control region, the same approach cannot be adopted with charm jets. Consequently, the linear and quadratic components of the uncertainty from the heavy flavor SFs are added to the purity component in quadrature, respectively. These numbers are multiplied by two and then utilized to develop two distinct nuisance parameters for controlling these uncertainties. The uncertainties associated with charm jets are treated independently from the uncertainties related to the HF and LF SFs.

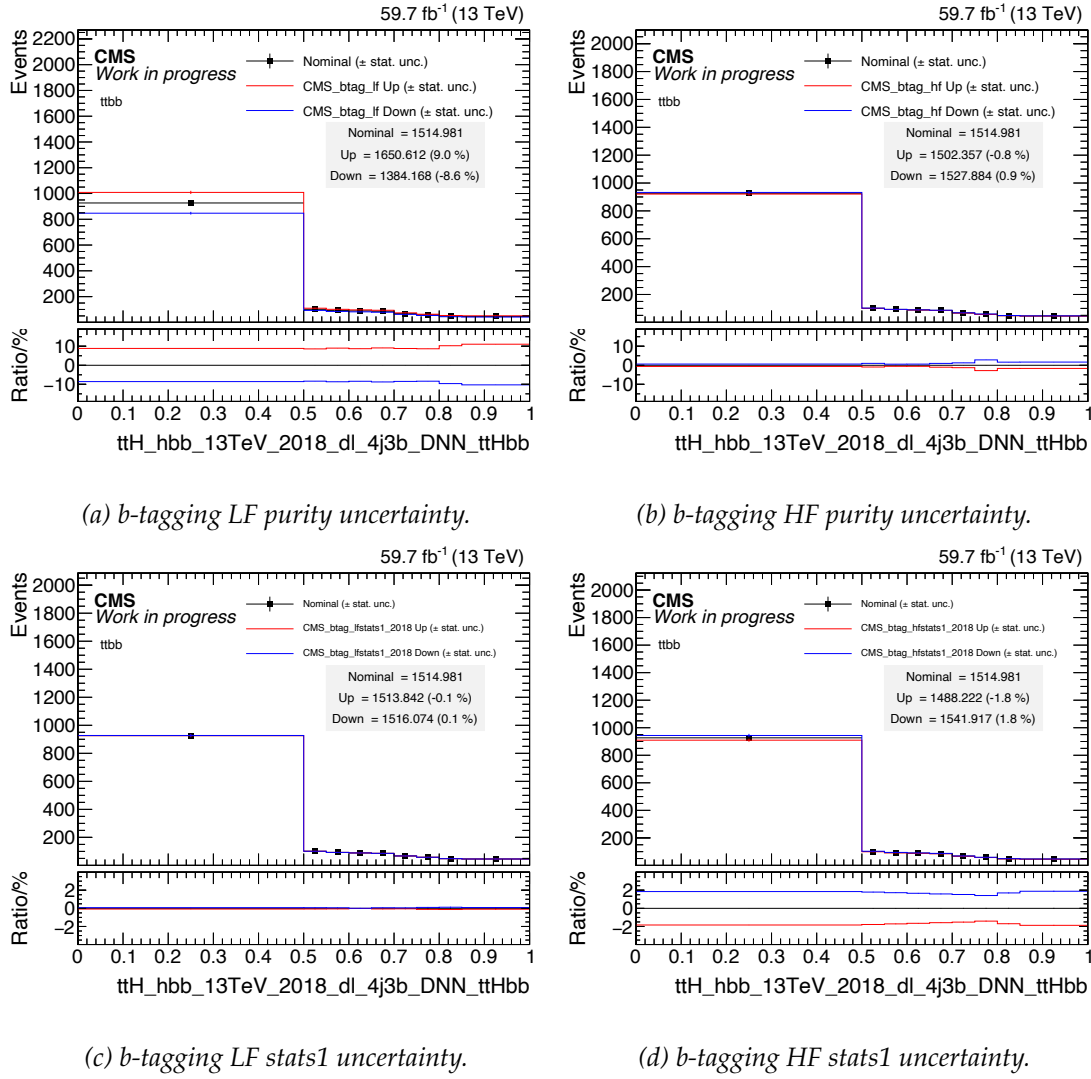


Figure 6.4: Effect of uncertainty variations on the ratio observable discriminant in the ≥ 4 jets, ≥ 3 b tags category for the background $t\bar{t} + b\bar{b}$ process in 2018. The bottom ratio plot shows relative shape variations. Changes of the total expected yield are quoted for the respective variation.

Taking into consideration the update to the CMS detector’s pixel subsystem in 2017, the purity uncertainties and uncertainties for the charm quark jets are decorrelated between 2016 and the other data years.

Size of the MC samples

The limited size of the simulated MC samples results in statistical fluctuations of the nominal prediction. This is taken into account via the Barlow-Beeston-lite approach [275]; technically, the implementation is supported by the CMS Collaboration’s centrally provided fitting tool and used in this analysis. A rate-changing nuisance parameter is incorporated into the statistical model for each discriminant bin, and it is connected with all predicted physics process contributions in that particular bin. For bins with more than ten effective entries, a single Gaussian-constrained nuisance parameter that scales the overall yield is introduced in each bin; for bins with fewer entries, a separate Poisson probability density function is employed for each process. Furthermore, the contributions of the signal distributions are not taken into account when deciding whether to employ a single Gaussian nuisance parameter

or a Poisson constrained parameter for each process, as recommended by the tool’s creators. In the morphing method presented in Appendix B.1, the rate and shape effects are considered independently.

6.1.2 Theory and background modeling uncertainties

The systematic uncertainty related to theoretical predictions and simulated event modeling are presented in the next paragraphs; they arise from assumptions and inaccuracies in the modeling of the MC simulations employed in the analysis for data comparisons and result derivation. In particular, uncertainties due to variations in parton distribution functions, renormalization and factorization scales, and generator-specific modeling uncertainties that impact the form of the expected $t\bar{t}$ background contributions. Unlike the treatment of experimental uncertainties, the majority of nuisance parameters introduced here influence either a specific physics process or a group of processes based on theoretical reasoning.

Inclusive cross sections

The inclusive signal and background cross section expectations are obtained from theoretical predictions with at least NLO accuracy. Uncertainties affecting these normalizations are divided into two components: those caused by QCD scale uncertainties in the matrix-element computation (renormalization/factorization scales) and those influenced by the PDF set and choice of α_s (PDF+ α_s) employed in the cross section computation. In the final fit, these uncertainties are accounted for as rate uncertainties; when the origins are shared among multiple processes they are correlated. The same cross section calculations are utilized for the simulated samples of the various processes during the three years of data collection, therefore, the uncertainties are correlated across the years. Tables 6.3 and 6.4 summarize the uncertainty.

Collinear gluon splitting

Calculating extra gluon splitting effects is difficult and poorly modeled. As a result, a rate uncertainty on the $t\bar{t} + 2b$ process is created to account for uncertainties coming from this effect. It is implemented as a shape uncertainty that provides a 100% prior uncertainty to the cross section of $t\bar{t} + b\bar{b}$ events emerging from the $t\bar{t} + 2b$ process. The effects on the ratio observable discriminant in 2018 for the ≥ 4 jets, ≥ 3 b tags category for the $t\bar{t} + b\bar{b}$ background process is visible in Figure 6.5 (a). Similar effects are observed in the other two years. The difficulties in calculating these extra splitting effects have an influence on the simulations for the corresponding years of data collecting. As a result, this uncertainty is treated as fully correlated across years.

PDF shape

The shape variation of the final discriminant distributions due to the uncertainty on the PDF set is calculated by reweighting all distributions according to the variations supplied by the NNPDF3.0 sets.

The total uncertainty is estimated as the envelope of all individual variations at the level of the final discriminant template, which means that for each PDF variation, first the final template distribution are constructed, then the residual difference between the varied and nominal templates in each bin computed, and finally the envelope of these residuals constructed. In particular, the final uncertainty is calculated as either the RMS or the quadratic sum of the residuals in each bin of the discriminant distribution, depending on the precise PDF set used for the individual samples. The effects on the ratio observable discriminant in

Process	Scales($t\bar{t}H$)	Scales(tHq)	Scales(tHW)
$t\bar{t}H$	+5.8% / -9.2% [243]		
tHq		+6.5% / -14.9% ⁽¹⁾ [243]	
tHW			+4.9% / -6.7% [243]

(a)

Process	Scales($t\bar{t}$)	Scales(t)	Scales(V)	Scales(VV)
$t\bar{t}$	+2.4% / -3.5% [261]			
single t ⁽²⁾		+3.1% / -2.1% [276]		
W + jets			$\pm 3.8\%$ [262]	
Z + jets			$\pm 2\%$ [262]	
$t\bar{t}W$	+25.5% / -16.4% [257]			
$t\bar{t}Z$	+8.1% / -9.3% [257]			
Diboson				$\pm 3\%$ [262]

⁽¹⁾: includes flavor-scheme dependence⁽²⁾: only dominant t channel considered

(b)

Table 6.3: Inclusive cross section uncertainties due to the renormalization and factorization scales on the signal (a) and the background (b) processes. Uncertainties in the same column for two different processes (different rows) are treated as correlated.

	gg $t\bar{t}H$	qb tHq	gb tHW
$t\bar{t}H$	$\pm 3.6\%$ [243]		
tHq		$\pm 3.7\%$ [243]	
tHW			$\pm 6.3\%$ [243]

(a)

	gg	qq	gq
$t\bar{t}$	$\pm 4.2\%$ [261]		
single t ⁽¹⁾			$\pm 2.8\%$ [276]
W + jets		-0.4% / + 0.8% [262]	
Z + jets		$\pm 0.2\%$ [262]	
$t\bar{t}W$		$\pm 3.6\%$ [257]	
$t\bar{t}Z$	$\pm 3.5\%$ [257]		
Diboson		$\pm 5\%$ [262]	

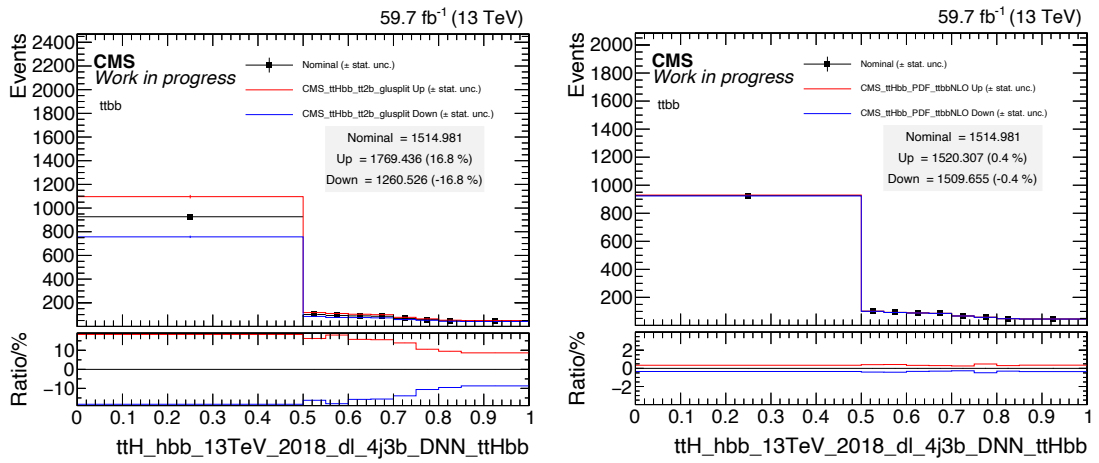
⁽¹⁾: only dominant t channel considered

(a)

Table 6.4: Inclusive cross section uncertainties due to the PDF+ α_s on the signal (a) and the background (b) processes for different initial states. Uncertainties for the same initial state (same column) for two different processes (different rows) are treated as correlated.

2018 for the ≥ 4 jets, ≥ 3 b tags category for the $t\bar{t} + b\bar{b}$ background process are depicted in Figure 6.5 (b). Similar effects are observed in the other two years.

Because the (N)NLO cross section uncertainties (Table 6.4) cover the total normalization uncertainties of the PDF, only the shape or acceptance changes are evaluated here. As a result, each variation is applied so that the total event yield is kept before any selection is retained. The PDF uncertainties are treated as uncorrelated between the $t\bar{t} + B$ (4FS sample) and the $t\bar{t}H$ and other $t\bar{t}$ (5FS samples) processes. The variations are taken from the respective PDF sets NNPDF31_nnlo_as_0118_nf_4 (4FS sample) and NNPDF31_nnlo_hessian_pdfas (5FS samples). There are also PDF uncertainties for tHq and tHW , which are calculated from the quadratic sum of all residuals. Individual PDF uncertainties are associated across all years since the same set is utilized for modeling of the corresponding processes in all data taking years evaluated.



(a) Collinear gluon splitting uncertainty on $t\bar{t} + 2b$.

(b) PDF shape variations on $t\bar{t} + b\bar{b}$.

Figure 6.5: Effect of uncertainty variations on the ratio observable discriminant in the ≥ 4 jets, ≥ 3 b tags category for the background $t\bar{t} + b\bar{b}$ process in 2018. The bottom ratio plot shows relative shape variations. Changes of the total expected yield are quoted for the respective variation.

μ_R, μ_F scales

The uncertainties caused by the choice of the renormalization μ_R and the factorization scale μ_F in the $t\bar{t} + b\bar{b}$ and $t\bar{t}$ matrix-element generators are modeled by varying independently the scales by factors of 0.5 or 2, respectively. This is accommodated via weights obtained directly from the generator information. Following the recommendations of the CMS Collaboration experts, the effects are propagated to the final discriminant distribution in the fit. Since the (N)NLO cross section uncertainties cover the normalization uncertainties of the matrix-element generator (Table 6.3), these parameters only cover effects on the shapes or acceptance variation of the discriminant distributions, i.e. the variations are scaled to retain the overall normalization as for the PDF shape uncertainty in the inclusive phase-space. Additionally, changes in the relative percentage of $t\bar{t} + B$ events in the inclusive $t\bar{t}$ phase space are also based on the composition of events in the $t\bar{t}$ (5FS) samples for these variations, using the same merging approach outlined in Section B.1.

The scale-variation uncertainties are implemented as shape uncertainties and are considered uncorrelated among the $t\bar{t}H$, tHq , tHW , $t\bar{t} + B$ (4FS sample) and other $t\bar{t}$ (5FS sample) processes. The scales used for the three data sets are the same for the simulated samples for the relevant processes. As a result, these uncertainties are correlated across all years. The

effects on the ratio observable discriminant in 2108 for the ≥ 4 jets, ≥ 3 b tags category, for the $t\bar{t} + b\bar{b}$ background process, are depicted in Figure 6.6 (a) and (b), respectively for the renormalization μ_R and the factorization scale μ_F . Similar effects are observed in the other two years.

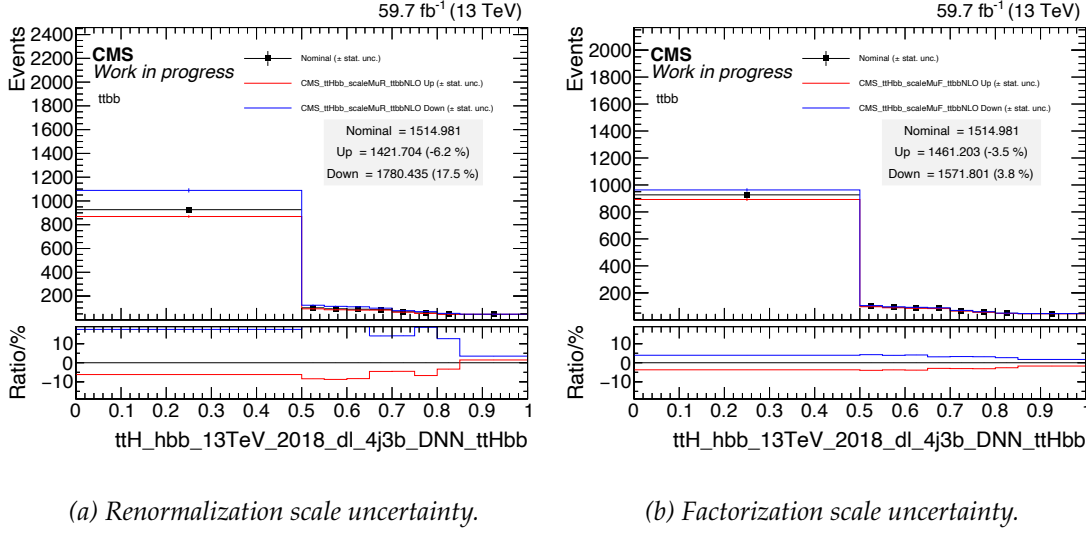


Figure 6.6: Effect of uncertainty variations on the ratio observable discriminant in the ≥ 4 jets, ≥ 3 b tags category for the background $t\bar{t} + b\bar{b}$ process in 2018. The bottom ratio plot shows relative shape variations. Changes of the total expected yield are quoted for the respective variation.

PS scales ISR/FSR

To assess the influence of α_s in the parton shower simulation, the relevant scale parameters in the shower simulation are modified by factors of 0.5 and 2. This is done separately for the initial-state radiation and final-state radiation showers. Because the (N)NLO cross section uncertainties also cover the normalization errors of the matrix-element generator, these uncertainties only assess the influence on the overall shapes of the final observables. (Table 6.3). Similarly to the μ_R , μ_F scales shape uncertainties, changes in the relative proportion of $t\bar{t} + B$ events to the inclusive $t\bar{t}$ phase-space are examined based on the composition of events in the $t\bar{t}$ (5FS) samples. The effects on the ratio observable discriminant in 2018 for the ≥ 4 jets, ≥ 3 b tags category for the $t\bar{t} + b\bar{b}$ background process are depicted in Figure 6.7 (a) and (b), respectively for the ISR and the FSR. Similar effects are observed in the other two years.

The ISR and FSR uncertainties are treated as uncorrelated among the $t\bar{t}H$, $t\bar{t} + B$ (4FS sample) and other $t\bar{t}$ (5FS sample) processes; but correlated across all years because the parton shower for the relevant simulations is the same.

ME-PS matching

The uncertainty in the matching between the matrix element generator POWHEG and the parton shower PYTHIA simulation is estimated using dedicated samples listed in Table 6.5 with varied $hdamp$ settings. The core simulation has $hdamp = 1.379 m_t$, while the variants have $hdamp = 2.305 m_t$ and $hdamp = 0.874 m_t$.

Ideally, the ME-PS matching uncertainties would be treated as shape uncertainties by building the distributions of the final discriminant from the appropriate dedicated simulations. However, the limited number of events available in the $t\bar{t} + B$ (4FS) and $t\bar{t}$ (5FS)

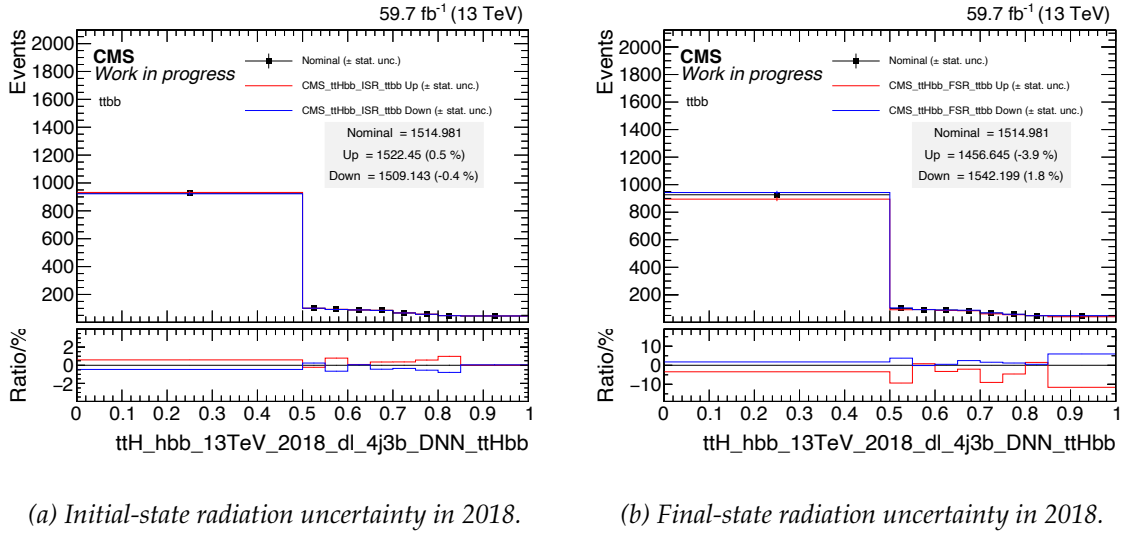


Figure 6.7: Effect of uncertainty variations on the ratio observable discriminant in the ≥ 4 jets, ≥ 3 b tags category for the background $t\bar{t} + b\bar{b}$ process in 2018. The bottom ratio plot shows relative shape variations. Changes of the total expected yield are quoted for the respective variation.

samples yields in substantial statistical fluctuations and consequently unrealistic shapes of the final distribution templates. The fluctuations between these distributions and the nominal expectations are statistically consistent with a pure rate variation (no shape variation). As a result, a rate uncertainty is computed per $t\bar{t} + X$ process and per category: the average of the absolute up/down rate changes is used as uncertainty, symmetrically around the nominal. In order to increase the statistical accuracy of the estimate, the average relative variation across all three years is computed. This leads in a rate change of 3-10%, with uncommon outliers of up to 20%, which is equivalent to the scale of the statistical variations.

An independent approach (TRF) is used to verify the premise that no substantial shape variation exist. For this, the varied distributions are produced from events in a control region with fewer b-tags than the analysis selection but identical event selection otherwise. The effect of the missing b-tag selection is then accounted for using event weights that operate as a transfer factor and are dependent on the p_T , η and flavor of the jets in the event. As a result, the statistical fluctuations of the various templates are reduced, allowing us to confirm that the effect of the matching uncertainty is compatible with a rate variation. A comparison of the variations derived from dedicated samples and the TRF approach is given in Figure 6.8 for 2018. Similar effects are observed in the other two years.

The uncertainties are treated uncorrelated across the many $t\bar{t} + \text{jets}$ background processes ($t\bar{t} + B$, $t\bar{t} + c\bar{c}$, $t\bar{t} + \text{LF}$) and correlated per process between the analysis categories and channels; because the fluctuations of these parameters are calculated as an average throughout the years, each of these uncertainties is fully correlated across the years.

Underlying event

The influence of uncertainty in the underlying event tune on the $t\bar{t} + \text{jets}$ processes production cross section is assessed using dedicated samples with varied setting, listed in Table 6.5; the settings for these samples correspond to the errors in the derivation of the tune [240, 241]. The variation of the underlying event tune affect the soft-particle regime and should not depend on the details of the simulation of the hard process. As a result, no dedicated samples are produced for the $t\bar{t} + B$ (4FS) samples, but the relative effect of the

Sample	Number of Events		
	2016	2017	2018
$t\bar{t}$ (SL)			
ME-PS matching Up	30 000 000	27 500 000	27 496 000
ME-PS matching Down	30 000 000	27 130 991	26 468 000
Underlying Event tune Up	30 000 000	20 122 010	27 482 000
Underlying Event tune Down	30 000 000	22 911 672	27 189 000
$t\bar{t}$ (DL)			
ME-PS matching Up	15 000 000	3 405 716	5 284 000
ME-PS matching Down	15 000 000	5 476 459	5 488 000
Underlying Event tune Up	15 000 000	5 500 010	5 494 000
Underlying Event tune Down	15 000 000	5 500 000	5 452 000
$t\bar{t}$ (FH)			
ME-PS matching Up	30 000 000	27 260 880	25 305 000
ME-PS matching Down	30 000 000	27 164 554	27 315 000
Underlying Event tune Up	30 000 000	27 301 272	24 064 000
Underlying Event tune Down	30 000 000	26 200 472	27 500 000
$t\bar{t} + b\bar{b}$ (SL)			
ME-PS matching Up	2 061 000	2 061 000	2 061 000
ME-PS matching Down	1 991 500	2 098 500	2 019 000
$t\bar{t} + b\bar{b}$ (DL)			
ME-PS matching Up	2 394 000	2 394 000	2 394 000
ME-PS matching Down	2 135 700	2 137 500	2 137 500
$t\bar{t} + b\bar{b}$ (FH)			
ME-PS matching Up	2 152 500	2 152 500	2 152 500
ME-PS matching Down	2 389 290	2 391 790	2 389 493

Table 6.5: Number of generated events produced for each year of data taking in the simulated SM samples used in this analysis for the estimation of the ME-PS matching and the underlying event uncertainties.

underlying event tune is extracted from the $t\bar{t} + B$ (5FS) prediction and applied to the nominal 4FS prediction.

The effect of the underlying event tuning accuracy, like the ME-PS matching uncertainties, would ideally be treated as shape uncertainties in the statistical model; however, the same challenges exist in the processing of these samples. As a result, the same technique as described above for the ME-PS matching scale uncertainties is employed to mitigate the impacts of statistical fluctuations induced by the restricted number of events in the dedicated samples. Because the underlying event tuning should be independent of the arising final state, the rate uncertainties are handled as fully correlated across all $t\bar{t} + \text{jets}$ background processes, categories and channels. The variations of these parameters are averaged across all years of data taking and are therefore treated as fully correlated across the years.

This technique stabilizes the expected results without significantly impacting the expected sensitivity on the signal strength and the constraints on the nuisance parameters under study.

6.2 Measurement results

As described in detail in Section 5.7, there are 5 different event categories per year entering the final fit: 1 likelihood ratio distribution $p_o = \frac{o(t\bar{t}H)}{o(t\bar{t}H) + o(t\bar{t}bb)}$ for the events in the $t\bar{t}H$ and

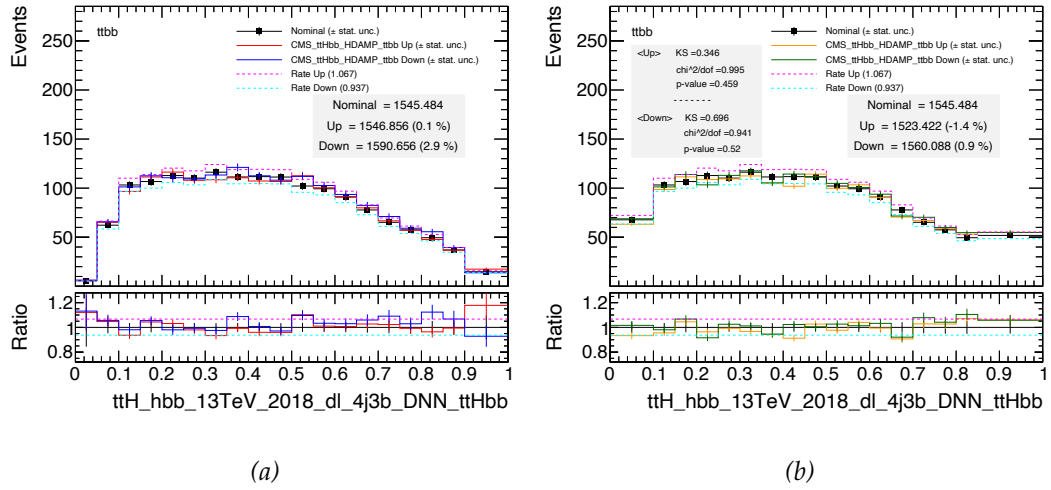


Figure 6.8: Effect of the $hdamp$ uncertainty on the ANN output distribution of $t\bar{t} + b\bar{b}$ events in the $t\bar{t}H$ category in 2018. (a) The nominal $t\bar{t} + b\bar{b}$ template (black solid line) is compared to the templates obtained when varying the $hdamp$ parameter by its uncertainties, i.e. from the events in the dedicated samples (blue and red solid lines). The dashed lines represent the rate variation obtained from this, which is used as uncertainty in the fit. (b) the nominal $t\bar{t} + b\bar{b}$ template (black solid line, same as in the left plot) is compared to the varied templates obtained using the TRF method (yellow and green solid lines). The KS, χ^2 , and p value probabilities indicate the compatibility of the varied templates with a rate-only (i.e. no shape) variation. For comparison, the rate variation used in the fit (dashed lines, same as in the left plot) is overlaid.

$t\bar{t} + b\bar{b}$ process categories in ≥ 4 jets, ≥ 3 b tags category, 2 ANN output value distributions $o(i)$ for the events in the $t\bar{t} + c\bar{c}$ and $t\bar{t} + LF$ of the ≥ 4 jets, ≥ 3 b tags category, and one binary ANN output distribution for the 3 jets, ≥ 3 b tags category. The binning of these distribution has been chosen to include at least three background events and the maximum relative statistical uncertainty of each bin is 30%, allowing for a meaningful statistical analysis.

The statistical combination of these regions use available information in the background dominated regions to constrain the background normalization uncertainties in the signal enriched regions.

As introduced in Section 5.1 and described in Appendix B, to ensure good scientific practice and guarantee unbiased progress, the recorded data is not initially assessed in the analysis region, with an approach referred to as blinding. The blinded analysis begins with the signal extraction using an artificial data set in which all observables are estimated from simulation [237]. Then, following comprehensive validation of the agreement of the simulated samples to the observed data, the analysis is unblinded and a fit to real data is performed. The pulls and constraints of the most important nuisance parameter are investigated and finally the parameter estimation of the signal-strength modifier are addressed. The structure of the following sections reflects this approach.

In the following, results are extracted through a binned maximum likelihood fit to data based on the combined statistical model explained above, i.e. the simultaneous parameter estimation in the 3 jets, ≥ 3 b tags and ≥ 4 jets, ≥ 3 b tags categories. The following are the usual quantities used to describe the expected and observed results of the analysis:

- **Significance** The excess of data over the expected background contribution is expressed in Gaussian standard deviation via conversion of p -values.
- **Best-fit value** The central value of the best estimate of μ and its uncertainty can be transformed to the measured cross section of the signal process. In the blinded anal-

ysis, the signal hypothesis helps to have an estimate for the expected accuracy of the measurement.

Appendix B describes the fit procedure and the signal extraction procedure by means of the statistical inference methods in depth.

6.2.1 Distribution and yields

A total of four categories per year contributes to the final fit. The corresponding fit and post-fit yields are listed in Appendix J.1.

In addition the corresponding distributions before and after the fit are shown in Figure 6.9 for the $t\bar{t}H$ and $t\bar{t} + b\bar{b}$ category, the one with the highest sensitivity to the $t\bar{t}H$, $H \rightarrow b\bar{b}$ signal. Here the shape discrimination between background and signal is clearly visible, enhanced by the effectiveness of the neural network based classification. In Appendix K.1 there are the distributions for the other category and nodes. As explained already the binning of the distribution has been carefully optimized to provide sufficient sensitivity but be flexible enough to mitigate possible differences between simulation and observed data. In particular, the categories with only one bin ($t\bar{t} + c\bar{c}$ and $t\bar{t} + \text{LF}$) help contributing to the parameter estimation only with the information of the process yields. All the other categories instead provide more differential information helping in the parameter estimation.

The post-fit distributions are the results of the unblinded combined fit of all categories and years together. Overall, the simulated predictions for the different processes are in good agreement with the observed data points. As a side note, because it is not defined, the prior uncertainty on the freely-floating parameters in the fit is not included in the prefit uncertainty bands. But all systematic uncertainties are included in the postfit uncertainty bands. The post-fit distribution has a smaller uncertainty band, reflecting that the fit is sensitive enough to constrain some of the available degrees of freedom.

6.2.2 Goodness-of-Fit tests

To further validate the agreement between the simulation and the observed data, the description of the latter is tested in the different analysis regions and in their combination with saturated Goodness-of-Fit (GoF) test, as described in Appendix B.3. The results are summarized in Table 6.6: all p-values are reported to be well above the 5% threshold, proving the exceptional quality of the observed data representation within the limits of the statistical model.

	2016	2017	2018	full Run 2
3 jets, ≥ 3 b tags	0.980	0.863	0.997	/
≥ 4 jets, ≥ 3 b tags	0.980	0.903	0.997	/
Combined	0.863	0.880	0.997	0.998

Table 6.6: Goodness-of-Fit test results for unblinded fit to the data. Each result correspond to the statistical model for either the 3 jets, ≥ 3 b tags category, the ≥ 4 jets, ≥ 3 b tags category or the combination of both, for each year individually and considering the full Run 2 combination.

6.2.3 Nuisance pulls and impacts

A detailed study of the effects of the individual nuisance parameters on the value of the signal strength allows to have a deeper understanding of the statistical model. In par-

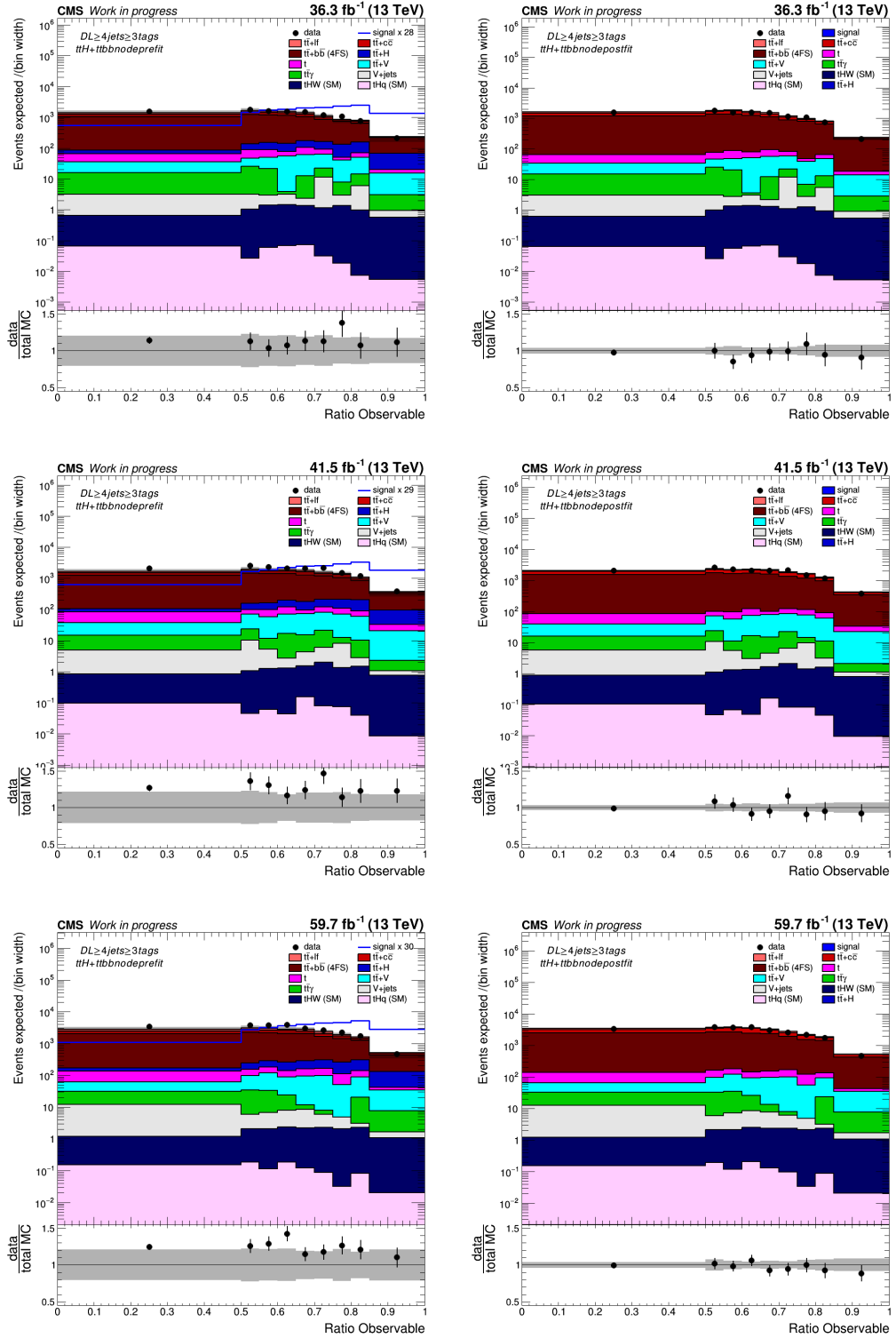


Figure 6.9: Prefit (left) and postfit (right) discriminant distributions in the ≥ 4 jets, ≥ 3 b tags $ttH + t\bar{t} + b\bar{b}$ ratio observable category for 2016 (top), 2017 (middle) and 2018 (bottom). The black points represent the observed data; the signal template is scaled to the total background contribution in the pre-fit plots; and the stacked distributions represent the simulations for the respective processes. The uncertainty band corresponds to the uncertainties in the full Run 2 combination fit.

ticular, the impact of a given nuisance parameter is calculated by fixing the parameter to its $\pm 1\sigma$ boundaries, re-estimating the signal-strength modifier μ and then computing the difference with respect to the nominal best estimate of μ . More details of the definition of pulls, constraints and impacts is in Appendix B.

Initially, the expected impacts are computed on a pseudo data set built on the sum of the signal and background predictions to define the predicted behavior of the statistical model. Afterwards, the ranking of the impact in the fit to observed data is calculated.

Figure 6.10 reports the 30 most relevant nuisance parameters having the largest expected impact on the signal inference. They are mostly related to the uncertainties of the $t\bar{t} + \text{jets}$ background and $t\bar{t}H$ signal or having a direct influence on the signal process itself. The fitted parameter values, expected and observed constraints are also shown. Overall the ranking of the expected impacts and the estimation of the nuisance parameters is reasonable and the observed ranking is compatible to the expected one, confirming the compatibility of the simulated predictions for the relevant processes and the observed data.

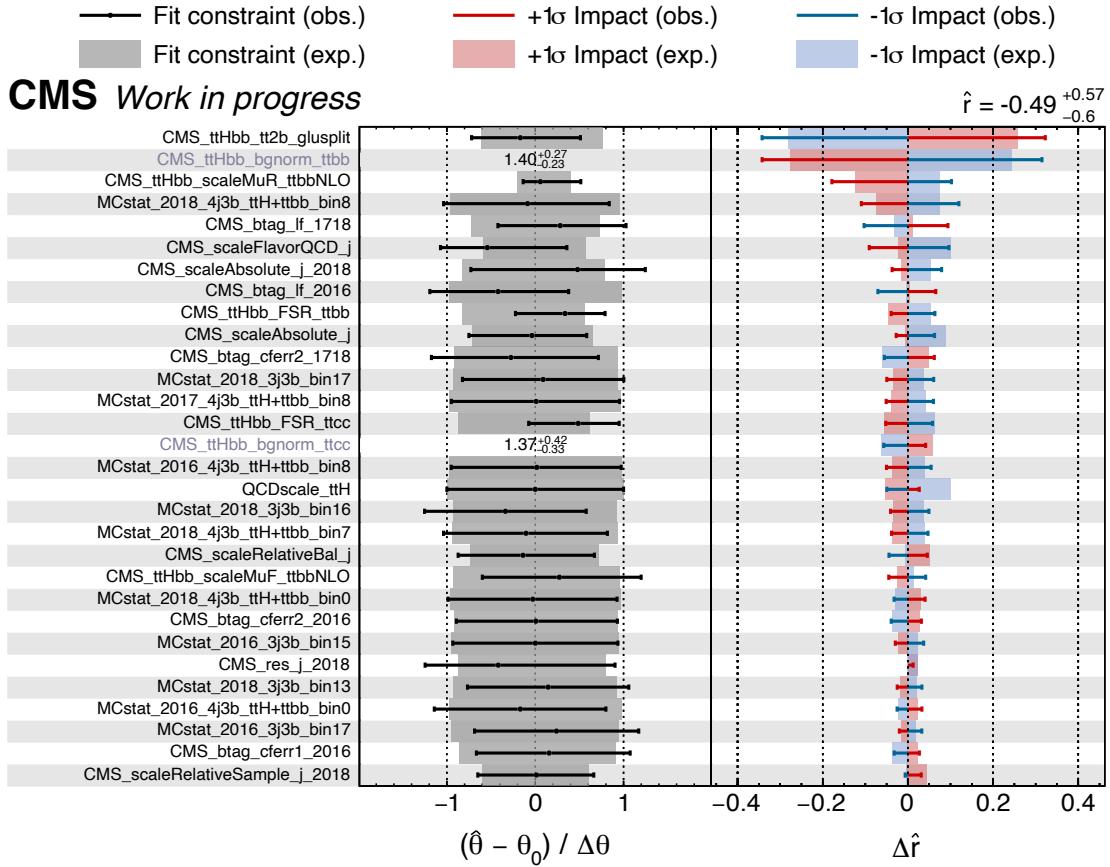


Figure 6.10: The 30 parameters with the largest impacts and their pulls and constraints on the signal strength obtained in the full Run 2 fit, both in the fit to pseudo data with the signal process expected from the SM and in the observed data. The order is relative to their impact in the fit to the data. The resulting fitted value for the signal strength is also given with its total uncertainty. The black lines represent the nuisance parameter values in comparison to their input values before the fit and relative to their uncertainty. The red and blue lines are the variation in $\Delta\mu$ when the nuisance parameter is varied up and down by its fitted uncertainty, respectively. The corresponding gray, red and blue bands are the expected values from fits to pseudo-data.

The variations of the irreducible $t\bar{t} + b\bar{b}$ background process are the dominant theory uncertainty, as expected, given that any variation of this process affects other processes with similar kinematic properties, such as the signal. Some of these parameters have a large a

priori uncertainties to account for the challenges in the computation of these predictions.: the 100% uncertainty on the collinear gluon splitting allows for large variations of the $t\bar{t} + b\bar{b}$ processes, yielding to corresponding changes in the signal process estimation, as well as the freely-floating background normalization of the $t\bar{t} + b\bar{b}$ process, modifying directly the yield of the irreducible background. Also the large impact of the signal modeling parameters, like the uncertainty on the initial and final state radiation, are reasonable as they directly influence the signal process itself.

The b-tagging method parameters, the JES and the JER are the experimental uncertainties with highest impact on the signal strength estimate. The final states have at least three b-tagged jets, therefore the parameters affecting the b-tagged objects reconstruction directly affect the signal inference. The $t\bar{t} + \text{LF}$ control region purity and charm quark tagging uncertainties have the greatest impact on b-tagging and also the highest a priori uncertainty, reaching up to 30%. Similarly, the JES and JER can affect the reconstruction of the various processes entering the analysis and model migration effects of events passing analysis phase space selections. Both in the expected and observed, this migration effect is shown as asymmetric or one-sided impacts.

Additionally, the uncertainties related to the limited size of the simulated samples are also present among the 30 with the highest impacts. In particular, these parameters modify either the first or last bins of the distributions considered. The bins at the beginning of the spectrum usually have the highest yields of simulated events, and any variation yield to impact the overall normalization of the various processes. Conversely, the last bins are enriched of signal process by construction and directly affect the signal strength estimation.

As expected from the fluctuations common in statistical inference, overall the nuisance parameters are found in an interval $\pm 1\sigma$ around the a priori expected values. All the constraints shown in Figure 6.10 are reasonable and compatible with the expectation of this analysis being sensitive to $t\bar{t} + \text{jets}$ background processes due to the large contribution to the analysis phase space. Finally, the post-fit value of the $t\bar{t} + b\bar{b}$ cross section normalization of $1.40^{+0.27}_{-0.23}$ is compatible with recent dedicated measurements [119, 277].

6.2.4 Parameter of interest estimation

The following section focuses on the sensitivity results of the analysis to the $t\bar{t}H, H \rightarrow b\bar{b}$ signal. Both expected and observed sensitivity have been studied. Pseudo data constructed from the sum of the SM prediction for both signal and background processes are used to obtain the expected sensitivity.

Figure 6.11 and Table 6.7 summarize the results of the signal-strength estimation for the individual years and the full Run 2 combination. It shows also the results of previous $t\bar{t}H, H \rightarrow b\bar{b}$ analyses using the 2016 [121] and 2017 [127] data sets. To notice, the statistical uncertainties in past analyses were significantly lower than the current ones. This is due to the additional cross section uncertainties for the $t\bar{t} + b\bar{b}$ and $t\bar{t} + c\bar{c}$ processes that were modeled using an a priori uncertainty of 50%, while now they are directly estimated from the data. Now they contribute to the statistical uncertainty whilst previously to the systematic uncertainties.

The original measurements of the 2016 and 2017 data yielded overall to a sensitivity compatible to the current one within uncertainties. The major differences in the observed signal strength are due to the treatment of the uncertainties of the $t\bar{t} + b\bar{b}$ and $t\bar{t} + c\bar{c}$ background processes and their post-fit values. Even with the full Run 2 the expected sensitivity for this channel is of 1.9σ with respect to the background hypothesis, not sensitive enough to claim any evidence of the $t\bar{t}H, H \rightarrow b\bar{b}$ process.

The total uncertainty of the measured best fit value μ is also studied through a breakdown into groups of specific uncertainties. In particular, the contribution of a source of systematic

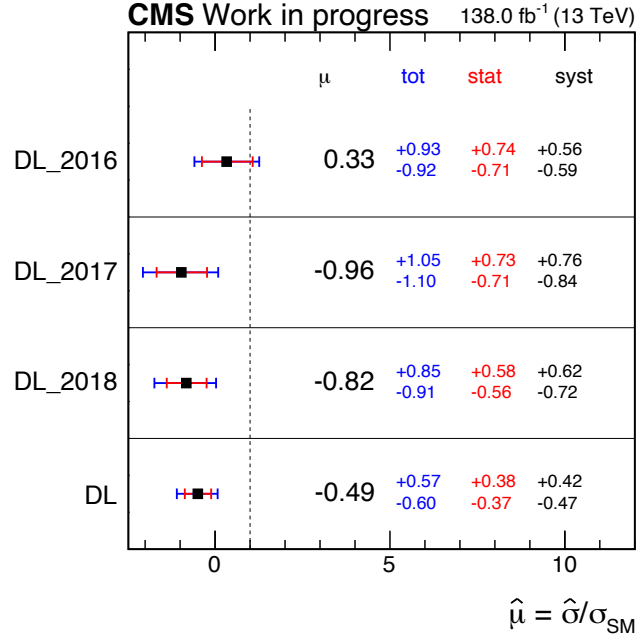


Figure 6.11: Best fit values of the signal strength modifiers μ obtained in the fit of the individual data set and in the full combination. Also shown are the 68% expected confidence intervals (outer error bar), also split into their statistical (inner error bar) and systematic components.

	Stat+Syst	Stat-Only	Significance	
			Observed	(Expected)
HIG-17-026 (2016 - DL channel)	-0.24 +1.21 -1.12	+0.63 -0.60	0.0	-
HIG-18-030 (2017 - DL channel)	1.62 +0.90 -0.85	+0.55 -0.55	1.9 σ	(1.2 σ)
this analysis (2016)	0.33 +0.93 -0.92	+0.74 -0.71	0.4 σ	(1.1 σ)
this analysis (2017)	-0.96 +1.05 -1.10	+0.73 -0.71	0.0	(1.2 σ)
this analysis (2018)	-0.82 +0.85 -0.91	+0.58 -0.56	0.0	(1.3 σ)
this analysis (full Run 2)	-0.49 +0.57 -0.60	+0.38 -0.37	0.0	(1.9 σ)

Table 6.7: Results in the measurement of the signal strength modifier for the individual years and their combination. For comparison, the first two rows show the results from previous measurements [121, 127].

uncertainty n modeled by θ_n is estimated as:

- the total uncertainty $\Delta\mu$ is obtained via the profile likelihood fit as above;
- the fit is repeated for the parameter under investigation fixing it to its post-fit value as given by the initial fit; this yields to a changed total uncertainty $\Delta\mu_{\text{fixed } n}$;
- the contribution to the overall uncertainty is given by the subtraction in quadrature as $\pm\Delta\mu_n = \sqrt{(\Delta\mu)^2 - (\Delta\mu_{\text{fixed } n})^2}$.

This procedure is also applied considering a group of nuisance parameters. Instead, for the statistical uncertainty, all nuisance parameters are fixed to their post-fit values.

Uncertainty source	$\Delta\mu_{\text{obs}}$
Theory	+0.23/-0.29
Signal	+0.08/-0.08
Background	+0.23/-0.27
tH	+0.08/-0.05
t $\bar{t}$	+0.24/-0.27
Freely floating t $\bar{t}$ + X normalization	+0.25/-0.35
Experimental	+0.20/-0.20
b tagging	+0.13/-0.12
Jet energy corrections	+0.16/-0.11
Luminosity	+0.08/-0.08
Statistics of simulated events	+0.21/-0.20
Total systematic	+0.42/-0.47
Statistical	+0.38/-0.37
Total	+0.57/-0.60

Table 6.8: Contributions of various groups of systematic uncertainties to the observed uncertainty of the best fit value μ_{DL} . Due to correlations among nuisance parameters within the fit, quadratic sums of individual uncertainty contributions differ from the quoted total values.

Following this method, a breakdown of the uncertainty for the inclusive measurement in the fit to the full Run 2 data is in Table 6.8. It should be noted that due to correlations between nuisance parameters, the quadratic sum of uncertainties may differ from the quoted total values. The overall uncertainty is split into its statistical and total systematic components. Other systematic uncertainties that are interdependent by construction are fixed simultaneously and treated as a group. With a contribution of +0.42/-0.47, at this point the systematic uncertainties are the limiting factor in the analysis; but the statistical uncertainties at +0.38/-0.37 is still dominant. The leading experimental sources are the uncertainties in b-tagging efficiency and jet energy corrections, understandable given the phase-space considered in the signal-enriched category.

In addition, to better understand the current results, Table 6.9 lists the parameters with the largest correlations to the signal strength modifier μ_{tH} . The order of magnitude of the correlations is similar both for the fit to the expected and observed data. As expected, the first parameters are related to the t $\bar{t}$ + b $\bar{b}$ process. In particular, the anti-correlation of the t $\bar{t}$ + b $\bar{b}$ cross section normalization reflects that any change in the background process yields result in an anti-correlated response in the estimation of the signal strength modifier. Moreover, there is also an uncertainty related to the limited size of the simulated sample affecting the content of the bin with the highest signal purity. Also in this case, the anti-correlation is well explainable: if the bin content is increased due to the uncertainty, the value of the signal-strength modifier decrease to mitigate the agreement between the simulation and the observed data. Figure 6.12 depicts the likelihood scans of the t $\bar{t}$ + b $\bar{b}$ cross section normalization parameter, where the tilt in the contours represent the correlations and overall showing no unexpected behavior.

Nuisance parameter	Correlation to $\mu_{t\bar{t}H}$ (%)	
	Blinded	Unblinded
Collinear gluon splitting	52.8	57.8
$t\bar{t} + b\bar{b}$ cross section normalization	-51.9	-57.0
Renormalization scale μ_R ($t\bar{t} + b\bar{b}$)	-17.0	-19.4
MC stat: ≥ 4 jets, ≥ 3 b tags, ratio observable bin 8 (2018)	-16.7	-19.7

Table 6.9: List of the four parameters with the largest correlation with the signal strength modifier in the analysis of the full Run 2 data set. Both the values of the correlations in the fit to the pseudo data and to the observed data are reported.

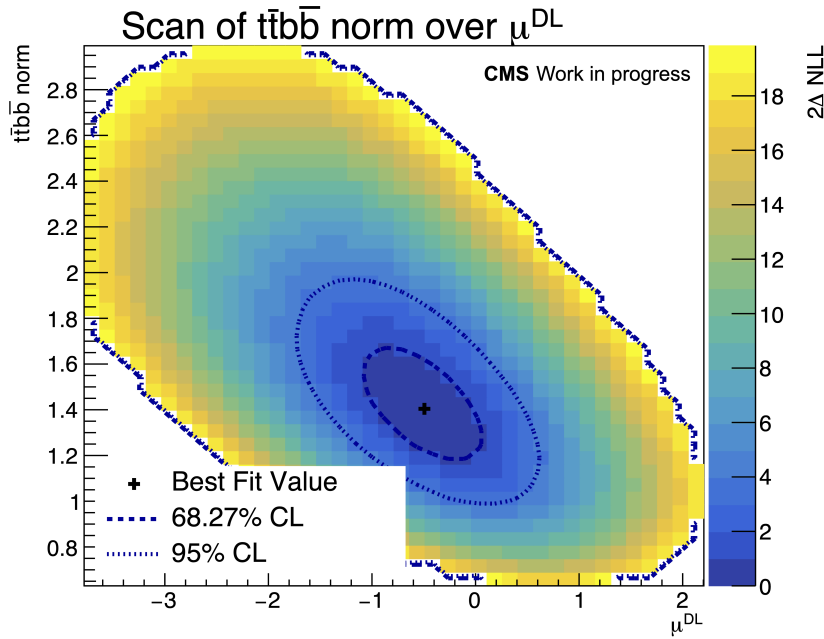


Figure 6.12: Likelihood scan of the freely-floating $t\bar{t} + b\bar{b}$ normalization as a function of the signal strength modifier μ^{DL} . The best fit value to the observed data (cross) and the 1σ and 2σ confidence intervals (dotted lines) are shown.

6.3 Contribution to the combined analysis

After presenting the results in the DL only channel, the focus of the following section is on the contribution of this specific channel to the full analysis phase space measurement. Indeed, as introduced in Chapter 5, the analysis of the DL channel is one of the input channels for the full $t\bar{t}H, H \rightarrow b\bar{b}$ analysis in combination with the other decay $t\bar{t}$ decay channels, the fully-hadronic (FH) and the semileptonic (SL) channels. The production rates of the $t\bar{t}H$ and $t\bar{t}H$ signal processes are determined in a binned maximum likelihood fit to the final discriminant distributions in all channels, categories, and data-taking. The rates of the $t\bar{t} + B$, $t\bar{t} + C$ as well as the QCD multijet background are separately left unconstrained in the fit. The different background composition in the categories helps to constrain the uncertainties and thus to increase the overall sensitivity.

At the time of writing this thesis, the $t\bar{t}H, H \rightarrow b\bar{b}$ analysis is still ongoing and under review. Therefore, here the preliminary current results are presented and discussed.

The results presented here follow the same skeleton as in the previous section for the dilepton only fit. First, the pre-fit and post-fit yields for each DL category and each year are listed in Appendix J.2. Then, the corresponding distributions are all shown in Appendix K.2, split per category, node and year. Obviously the focus is on the dilepton channel and how this is affected by the fit to the full phase space. In particular, Figure 6.13 shows a comparison of post-fit distributions for the most sensitive category of DL channel, i.e. the ratio observable in ≥ 4 jets, ≥ 3 b tags, for the fit using the DL-only statistical model and the full $t\bar{t}H, H \rightarrow b\bar{b}$ combination. Overall the agreement is good in both models, highlighting the absence of any tensions among the different channels entering the full combination fit. However, the main difference to appreciate is the different size of the uncertainty bands. Despite being compatible within few percent, the full combination benefits from the sensitivity arising from the other channels, mostly from the SL channel.

Additionally, the Goodness-of-fit test results are in Table 6.10 to quantify the agreement between the observed data and the simulation for each channel and year and their combinations. The p-values for the different inputs to the fit depict a good compatibility and well-modeled observed data in all channels. The fluctuations of the values are expected within a statistical data analysis. It has been properly investigated that the lower value for the full Run 2 combination among all channels with respect to the individual leptonic combinations is due to a tension with the FH channel. To control the QCD background, the FH analyses introduce a data-driven background estimation with additional nuisance parameters that have an interplay in the estimation of the $t\bar{t} + \text{jets}$ background modeling parameters.

	2016	2017	2018	full Run 2
FH Combined	0.818	0.349	0.492	0.462
SL Combined	0.810	0.999	0.712	0.951
DL Combined	0.832	0.873	0.998	0.998
All channels Combined	0.907	0.823	0.757	0.883

Table 6.10: Goodness-of-Fit test results for unblinded fits to observed data. Each result correspond to the statistical model for each channel or the combination of all, for each year individually and considering the full Run 2 combination.

Overall the different statistical models yield a good description of the observed data. In the following, the estimation of the nuisance parameters is presented with a thorough study of their pulls and impacts on the estimation of the best fit value of the signal strength. Figure 6.14 shows the 30 nuisance parameters with the largest impacts on the signal strength estimation, both in the fit to pseudo data and to observed data for the fully-combined fit across all $t\bar{t}H, H \rightarrow b\bar{b}$ channels. As expected, also in this case, the most important nuisance parameters are related to the normalization of the $t\bar{t} + b\bar{b}$ and QCD-multijet background modeling, and the uncertainty in the signal cross section as well as in the jet energy scale and b tagging. The high impact of the background model parameters reflects the challenge in the separation between this processes and the signal events. Since their shapes are really similar, parameters modifying the prediction of the background yield also impact the estimation of the signal yield.

For approximately all of the total number of nuisance parameters, the best fit values are within 1 standard deviation of the prior uncertainty. The fit constrains the nuisance parameters associated with the $t\bar{t} + B$ and $t\bar{t} + C$ background normalizations, as well as the conservatively assigned additional 100% prior uncertainty on the collinear gluon splitting model, as expected. A few more nuisance metrics related to jet energy scale and b-tagging

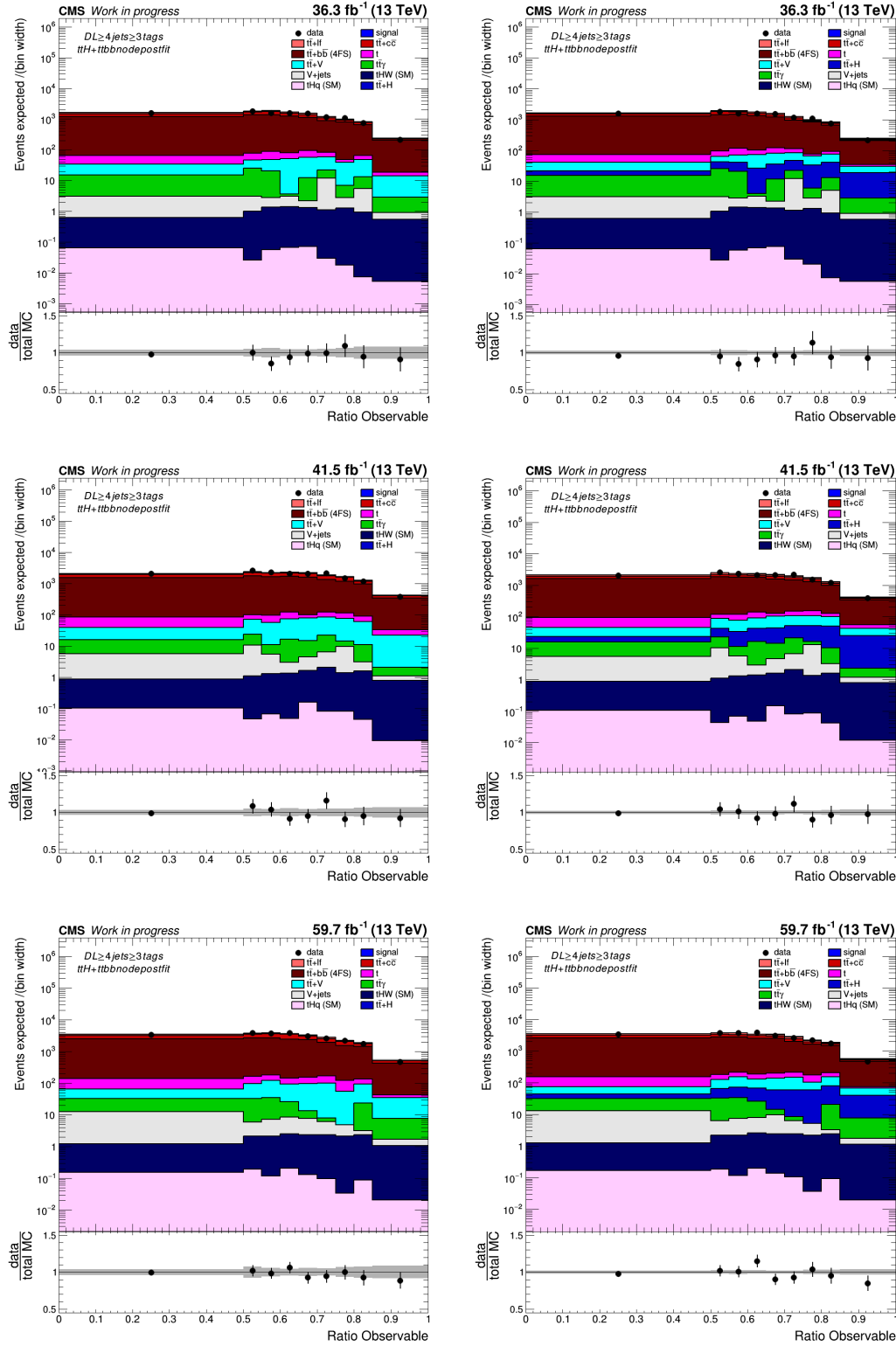


Figure 6.13: Post-fit discriminant distributions in the DL channel in the ≥ 4 jets, ≥ 3 b tags $t\bar{t}H + t\bar{t} + b\bar{b}$ ratio observable category for 2016 (top), 2017 (middle), and 2018 (bottom). The uncertainty band correspond to the uncertainty in the full Run 2 DL combination (left) and full Run 2 all channels combination fits. The black points represent the observed data; the stacked distributions correspond to the simulations for the respective processes.

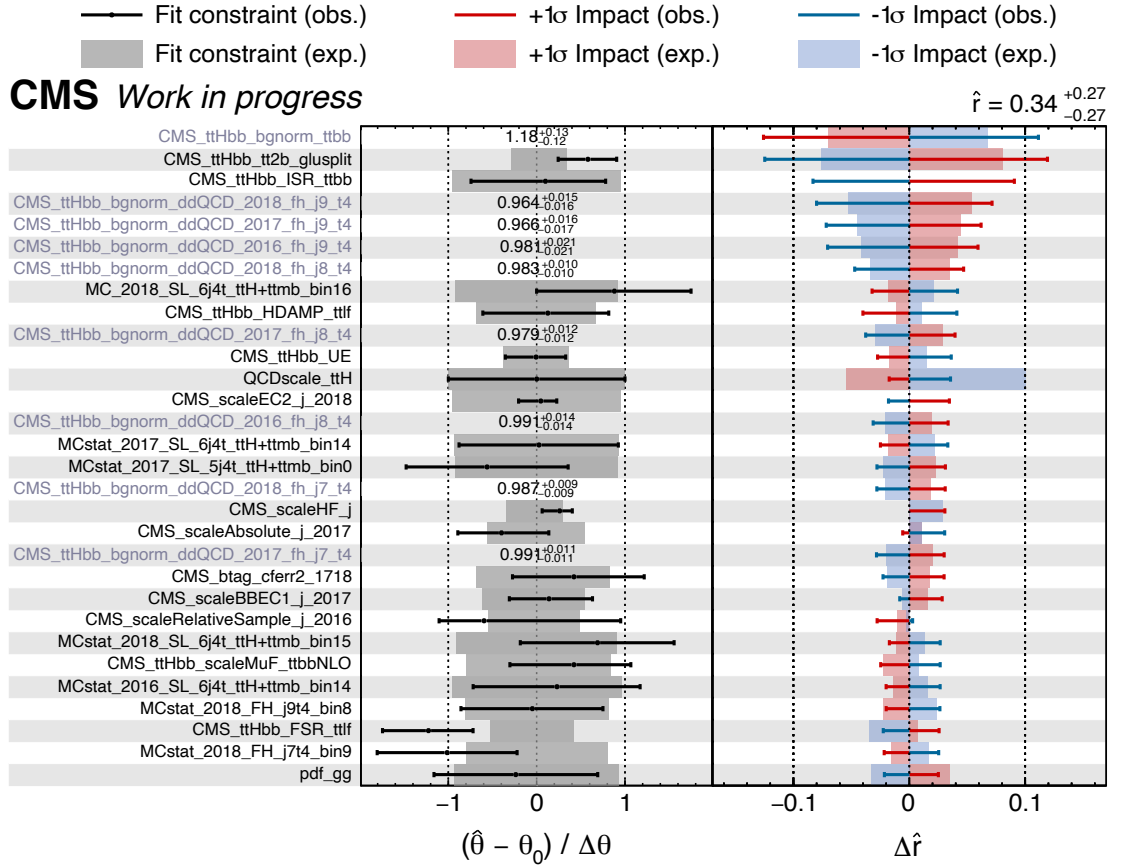


Figure 6.14: The 30 parameters with the largest impacts and their pulls and constraints on the signal strength obtained in the full Run 2 fit to the $t\bar{t}H, H \rightarrow b\bar{b}$, both in the fit to pseudo data with the signal process expected from the SM and in the observed data. The order is relative to their impact in the fit to the data. The resulting fitted value for the signal strength is also given with its total uncertainty. The black lines represent the nuisance parameter values in comparison to their input values before the fit and relative to their uncertainty. The red and blue lines are the variation in $\Delta\mu$ when the nuisance parameter is varied up and down by its fitted uncertainty, respectively. The corresponding gray, red and blue bands are the expected values from fits to pseudo-data.

uncertainties are constrained by a factor of about 40%. These constraints are explained by the fact that events are chosen based on different, large multiplicities of jets and b-tagged jets, increasing the sensitivity of the analysis to changes in the jet energy scale and b tagging efficiency, as measured, for example, by their effect on event yield per analysis category.

Overall the pulls and constraints here are compatible with what shown in Figure 6.10 for the fit with the DL only statistical model. However, it is remarkable how the uncertainties on nuisance parameters are further restricted and improved in the combined $t\bar{t}H, H \rightarrow b\bar{b}$ analysis. For example, the relative improvement in the uncertainty on the $t\bar{t} + b\bar{b}$ normalization estimation is 50%. This indicates again that the sensitivity to the most important systematic uncertainties is not driven by the DL channel and how much the estimation benefits from the full available phase space. Moreover, the ranking is slightly different. The presence of parameters related to the other $t\bar{t}H, H \rightarrow b\bar{b}$ channels change the ensemble of parameters and the correlations among them, induced by the changes in the statistical model itself. This results in the new parameters having a large impact, like the freely-floating parameters for the data-driven QCD background estimation in the FH channel; or in different estimates of the parameters, like FSR uncertainty for the $t\bar{t} + \text{LF}$ process. In the DL estimate, the post-fit value is compatible with the initial value, on the other hand in the full $t\bar{t}H, H \rightarrow b\bar{b}$ fit it

is pulled. Additionally, there is a clear decrease of the impact of experimental uncertainties when considering the full available phase space, that is now instead mostly dominated by parameters affecting the normalization of background processes and uncertainties due to the limited size of the simulated samples.

Also the post-fit values of the freely-floating $t\bar{t} + B$ and $t\bar{t} + C$ background normalization parameters and their uncertainties for the combined fit of all channels are presented in Table 6.11, for the single years and their combination. The results obtained from dedicated inclusive $t\bar{t} + b\bar{b}$ cross section measurements in a similar phases-space with 2016 data are at the level of 10-20% relative precision [119, 120].

Background	Blinded	Unblinded
$t\bar{t} + b\bar{b}$ (CMS_ $t\bar{t}Hbb_bgnorm_ttbb$) (36.33 fb^{-1})	$1^{+0.17}_{-0.14}$	$1.05^{+0.18}_{-0.15}$
$t\bar{t} + C$ (CMS_ $t\bar{t}Hbb_bgnorm_ttcc$) (36.33 fb^{-1})	$1^{+0.30}_{-0.26}$	$0.82^{+0.29}_{-0.26}$
$t\bar{t} + b\bar{b}$ (CMS_ $t\bar{t}Hbb_bgnorm_ttbb$) (41.53 fb^{-1})	$1^{+0.16}_{-0.14}$	$1.16^{+0.17}_{-0.15}$
$t\bar{t} + C$ (CMS_ $t\bar{t}Hbb_bgnorm_ttcc$) (41.53 fb^{-1})	$1^{+0.29}_{-0.26}$	$0.96^{+0.30}_{-0.27}$
$t\bar{t} + b\bar{b}$ (CMS_ $t\bar{t}Hbb_bgnorm_ttbb$) (59.74 fb^{-1})	$1^{+0.17}_{-0.14}$	$1.40^{+0.21}_{-0.19}$
$t\bar{t} + C$ (CMS_ $t\bar{t}Hbb_bgnorm_ttcc$) (59.74 fb^{-1})	$1^{+0.26}_{-0.24}$	$1.35^{+0.28}_{-0.26}$
$t\bar{t} + b\bar{b}$ (CMS_ $t\bar{t}Hbb_bgnorm_ttbb$) (137.60 fb^{-1})	$1^{+0.11}_{-0.10}$	$1.18^{+0.13}_{-0.12}$
$t\bar{t} + C$ (CMS_ $t\bar{t}Hbb_bgnorm_ttcc$) (137.60 fb^{-1})	$1^{+0.19}_{-0.17}$	$1.05^{+0.20}_{-0.19}$

Table 6.11: Post-fit values and uncertainties of the freely-floating background normalization parameters from the combined fit of all channels.

The following focuses on the sensitivity results of the full Run 2 analysis, looking at the best-fit $t\bar{t}H$ signal strength modifier $\mu_{t\bar{t}H}$ and the significance obtained from a fit to data in each channel alone and in the combination of all channels. To notice, in the fit to the FH channel alone, the sensitivity to the $t\bar{t} + B$ and $t\bar{t} + C$ normalization is minimal and therefore these uncertainties are parameterized with a 50% $\ln N$ uncertainties.

Figure 6.15 (a) illustrates the estimation of $\mu_{t\bar{t}H}$ in the different channels of the full Run 2 $t\bar{t}H, H \rightarrow b\bar{b}$ analysis and (b) how it evolves across the years. In the combination of all channels and years, an observed value of $\mu_{t\bar{t}H} = 0.34^{+0.26}_{-0.27}$ is obtained with a significance of the signal compared to the background-only hypothesis corresponding to 1.3 standard deviations, with an expectation of 4.1. The p -value compatibility of the combined result with the SM expectation of $\mu_{t\bar{t}H} = 1$ is 0.02. The statistical model considering the full phase space available yields the most sensitive results, with each channel individually contributing to the reduction of both the statistical and systematic uncertainties. Additionally Figure 6.15 (a) well depicts the contribution to the sensitivity coming from the single channels. The single-channel results are compatible with the combined result at a level corresponding to a p -value of 0.28. As well emphasized above, the reconstruction efficiency of the DL channel is quite high thanks to the two expected leptons in the final state, helping reducing the systematic uncertainty overall. But since the production probability is the lowest, the statistical uncertainty is the largest among the channels. With a higher production probability and a sufficient reconstruction efficiency, the SL channel is crucial and contributes the most to the sensitivity, yielding the smallest uncertainties. On the other hand, the FH channel has the smallest statistical uncertainties thanks to the large amount of data available (due to the branching ratio of the W boson to quarks). However, due to the large hadronic activity in any LHC event, making it difficult to reconstruct the $t\bar{t}H, H \rightarrow b\bar{b}$ system in this channel, it

results in the largest systematic uncertainties.

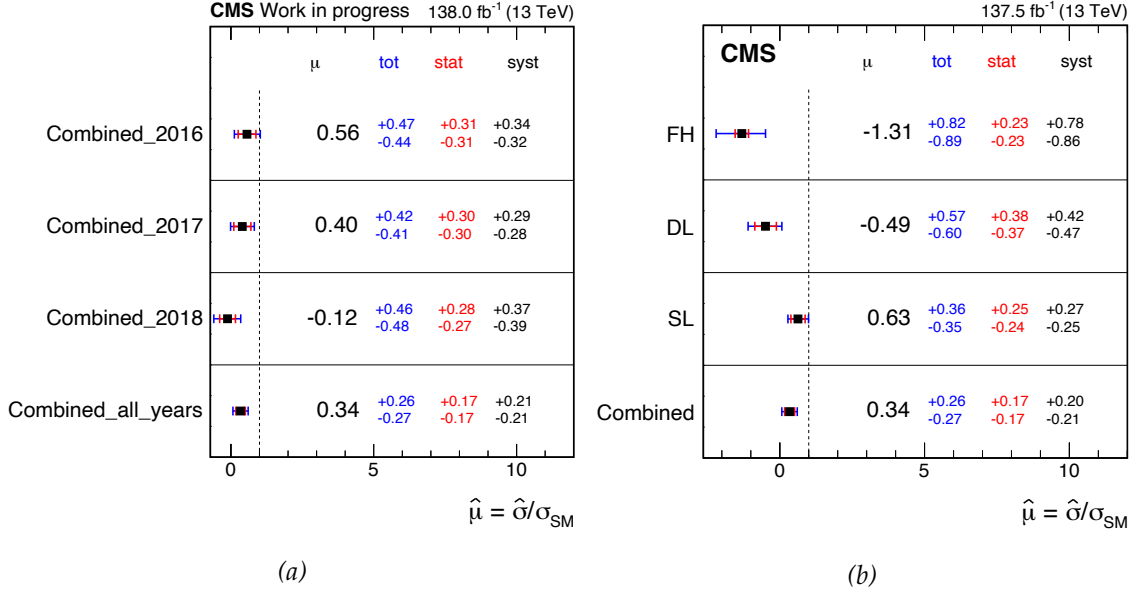


Figure 6.15: Best-fit results of the $t\bar{t}H$ signal-strength modifier obtained from a fit to observed data of 138 fb^{-1} in (a) each channel and their combination and (b) for the individual years.

The observed best fit value $\mu_{t\bar{t}H}$ is measured as:

$$\mu_{t\bar{t}H} = 0.34^{+0.26}_{-0.27} = 0.34^{+0.17}_{-0.17}(\text{stat.})^{+0.20}_{-0.21}(\text{syst.}).$$

Based on the SM prediction of $\sigma_{\text{SM}} = 507^{+35}_{-50} \text{ fb}$ with NLO QCD accuracy and NLO electroweak corrections, the cross section for the $t\bar{t}H$ production is:

$$\sigma_{t\bar{t}H} = 172^{+132}_{-137} = 172^{+86}_{-86}(\text{stat.})^{+101}_{-106}(\text{syst.}),$$

corresponding to a relative uncertainty of +77/-81%. Also for the full Run 2 all channels combination, the total uncertainty of the measured best fit value $\mu_{t\bar{t}H}$ is further studied through a breakdown into groups of specific uncertainties as reported in Table 6.12. The systematic and the statistical uncertainties, which include components from the background normalization, contribute at a similar level to the total uncertainty. Experimental uncertainties amount to $+0.10/-0.11$, dominated by the jet energy scale and resolution as well as b-tagging related uncertainties. Systematic uncertainties due to the size of the various simulated samples used to model the background and signal templates are at the same order and amount to ± 0.13 .

In order to check the compatibility of the fit result for the signal strength split between different channels and years, the p-value against the combined signal strength for full Run 2 data has been calculated to be 0.283. The results are well compatible with the full combined result. The p-value was also calculated to check the compatibility of the results with the SM expectation as in Table 6.13. The compatibility is around 2-3% which is consistent with the error bars obtained on the observed signal strength.

The compatibility with previous $t\bar{t}H, H \rightarrow b\bar{b}$ measurements performed with the 2016 and 2017 data sets by the CMS Collaboration was also quantified and it is listed in Table 6.14. Also here, the statistical uncertainties in the previous iterations have smaller values due to the parametrization of the $t\bar{t} + b\bar{b}$ and $t\bar{t} + c\bar{c}$ processes. Now their contributions is in the statistical uncertainty because they are estimated directly from the data. Additionally,

Uncertainty source	$\Delta\mu_{\text{obs}}$
Theory	+0.23/-0.23
Signal	+0.26/-0.27
Background	+0.24/-0.23
$t\bar{t}H$	+0.26/-0.27
$t\bar{t}$	+0.24/-0.23
Freely floating $t\bar{t} + X$ normalization	+0.12/-0.24
Experimental	+0.10/-0.11
b tagging	+0.06/-0.07
Jet energy corrections	+0.07/-0.08
Luminosity	+0.01/-0.02
Statistics of simulated events	+0.13/-0.13
Total systematic	0.20/-0.21
Statistical	+0.17/-0.17
Total	+0.26/-0.27

Table 6.12: Contributions of various groups of systematic uncertainties to the observed uncertainty of the best fit value $\mu_{t\bar{t}H}$. Due to correlations among nuisance parameters within the fit, quadratic sums of individual uncertainty contributions differ from the quoted total values.

Combination	p-value
Combined signal strength for full Run 2 data (1 POI)	0.015
Signal strength split between channels for full Run 2 data (3 POIs)	0.022

Table 6.13: Results of the p-value compatibility against the SM expectation.

an artificially decrease in the sensitivity is expected due to the new background model (4FS $t\bar{t} + b\bar{b}$ sample) predicting a larger background yield but improved description of the kinematic properties. However, these effects are mitigated by the refinements to the analysis strategy and the overall sensitivity of the current analysis is comparable with previous iterations of the $t\bar{t}H, H \rightarrow b\bar{b}$ measurement.

Previous result		Current result		p-value
2016 (SL +DL only)	$0.72^{+0.45}_{-0.45}$	2016	$0.56^{+0.47}_{-0.44}$	0.73
2016 (SL +DL only)	$0.72^{+0.45}_{-0.45}$	2016 (SL +DL only)	$0.32^{+0.50}_{-0.49}$	0.41
2016	$0.85^{+0.43}_{-0.41}$	2016	$0.56^{+0.47}_{-0.44}$	0.54
2017	$1.49^{+0.44}_{-0.40}$	2017	$0.40^{+0.42}_{-0.41}$	0.02
2016+2017	$1.15^{+0.32}_{-0.29}$	2016+2017	$0.55^{+0.31}_{-0.30}$	0.06

Table 6.14: Comparison results with previous measurements and their compatibility [121, 127].

TOWARD DIFFERENTIAL MEASUREMENTS



7.1	Simplified template cross sections framework	145
7.2	STXS measurement in $t\bar{t}H, H \rightarrow b\bar{b}$ - Analysis strategy	147
7.3	STXS measurements results	154
7.3.1	Contribution to the combined analysis	158

This chapter presents the first interpretation of $t\bar{t}H, H \rightarrow b\bar{b}$ towards a differential measurement within the Simplified Template Cross Section framework. The STXS framework provides a powerful tool for studying the properties of the Higgs boson and its interactions with other particles, by measuring the differential cross sections of the process as a function of various kinematic variables. After a short introduction about the general concept of the STXS measurement in the context of $t\bar{t}H$ analyses, the adopted analysis strategy is described. This includes a discussion of the event selection and categorization procedures used to enhance the purity of the signal, as well as the methods used to extract the STXS templates from the data. The results are presented, with a focus on the impact of individual sources of systematic uncertainties. The last section of the chapter focuses on the contribution of the dilepton channel in the context of the combination with other channels (single lepton and fully hadronic). This includes a description of the methods used to combine the results from the different channels, and the impact of the dilepton channel on the overall sensitivity of the STXS measurement.

7.1 Simplified template cross sections framework

Using the significant amount of data collected during the LHC Run 2, several couplings of the Higgs boson to vector bosons and second and third generation fermions have been directly observed and opened the possibility to precision measurements of the Higgs boson properties via differential measurements.

The STXS framework has been developed by the LHC Higgs Cross Section Working Group (LHCHWG) to evolve and move forward from signal strength measurements towards differential ones in as many Higgs production processes and final states as possible given the current experimental sensitivity. Furthermore, the STXS framework aims to differentiate between measurement and interpretation steps to systematically reduce the theory dependencies that are folded into the measurements, i.e., the dependence on theoretical uncertainties and the underlying physics model.

The premise of the STXS interpretation is to perform cross section measurements in well-defined and not overlapping regions of phase space of selected observables, which are referred to as STXS bins [21, 47, 278], defined in a common framework with all Higgs boson processes across experiments and the theoretical community to facilitate the combination of measurements and their interpretations. The most important feature is the synchronization across experiments and theory community on the definition of the STXS bins to help the global

combination of the measurements in all decay channels and the theoretical interpretations with more finely-grained experimental measurements.

Technically, the STXS interpretation is performed in an analogous way to a likelihood-based unfolding approach. The STXS bins are defined via observables at particle level and in stages of sequentially finer splitting, with the goals of minimizing the number of bins and the dependence on theoretical uncertainties but maximizing the experimental sensitivity and isolating BSM effects. First, at the so-called stage 0, all events with Higgs boson in the central region (i.e. absolute value of the Higgs boson rapidity $|Y| < 2.5$, to avoid the extrapolation to phase space regions out of the detectors acceptances) are split based on their production mechanism. Then, on stage 1, there is a splitting based on kinematic features of the event at the particle level, e.g. the transverse momentum of the Higgs boson (p_T^H) or the number of additional jets. An overview of the binning for the different Higgs production and decay channels is shown in Figure 7.1.

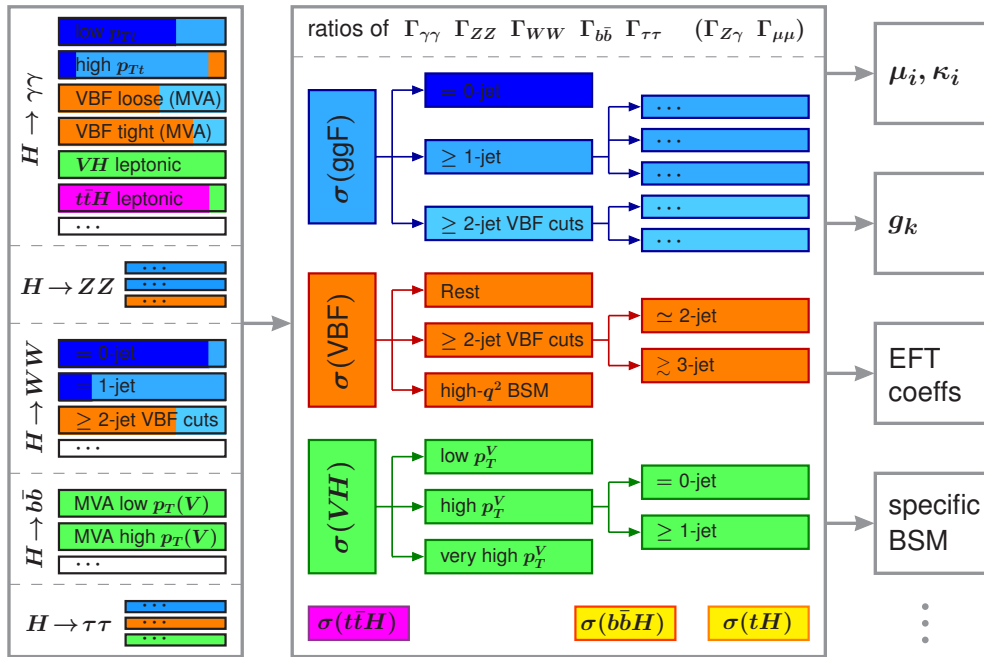


Figure 7.1: Schematic representation of the STXS framework [21]. The different colors represent different Higgs boson production modes. In the left, signal strength measurements are shown, split according to the decay channel of the Higgs boson and their own individual analysis categories. In the middle, STXS bins in different stages. On the right, the physical parameters that could be measured.

In the $t\bar{t}H$ production channel, the transverse momentum run is chosen as the observable of interest because it is sensitive to contributions from BSM operators (see Section 1.3).

The templates for the STXS bin are constructed by splitting the simulated signal events according to the edges of STXS bins defined at truth-level. Figure 7.2 (a) represents an illustrative example of the predicted spectrum of run split into two STXS bins. The well-defined separation of the two regions disappear after taking into account detector effects, as indicated in Figure 7.2 (b), since the finite resolution introduced by the detector response, parton shower, and hadronization leads to migration of events across bin boundaries. The STXS approach aims at correcting for these smearing effects by propagating the truth-level definition to the reconstruction after the detector simulation and performing a simultaneous measurement of the contribution of each STXS bin. This enables the inference of the truth-level properties while accounting for detector effects. Additionally, this whole approach also allows benefiting from MVAs to enrich STXS bins at truth-level in categories at reconstruction level.

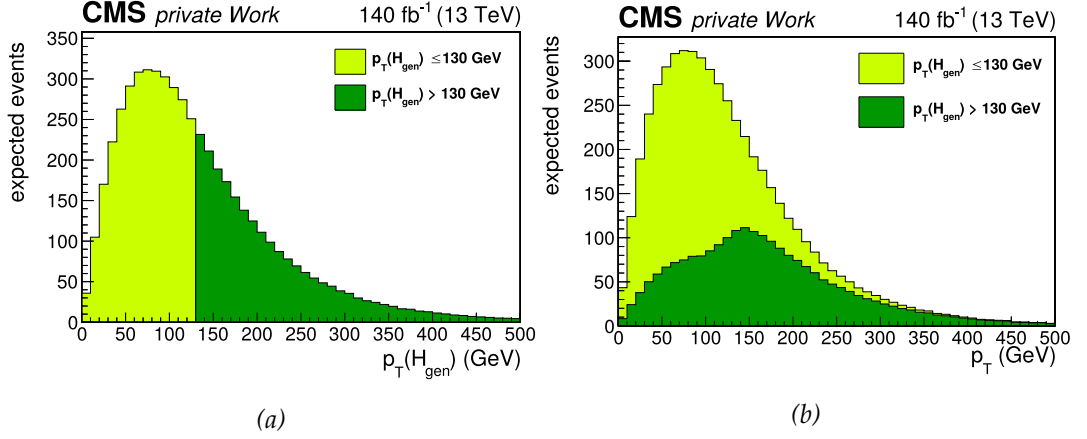


Figure 7.2: Illustration of the definition of two bins at truth-level (a) and after the reconstruction (b) [278].

The STXS bins in the $t\bar{t}H$ production channel have been designed to optimize the sensitivity in each STXS bin for different numbers of STXS bins but also for different bin boundaries [278]. The final design corresponds to the so-called stage 1.2, with the run spectrum divided into five STXS bins as in Figure 7.3. Then the $t\bar{t}H$ signal strength is measured as a function of the generated Higgs boson p_T .

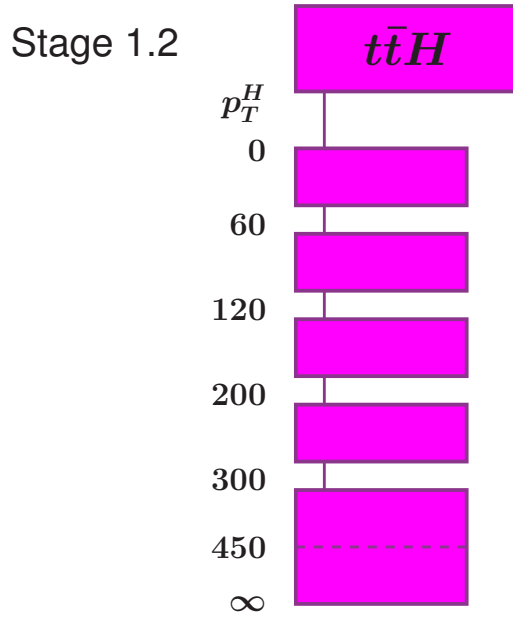


Figure 7.3: Sketch of the five STXS bins used in $t\bar{t}H$ -related measurements [21]. Additionally, the rapidity of the Higgs boson is restricted to be $|\Upsilon| < 2.5$.

7.2 STXS measurement in $t\bar{t}H, H \rightarrow b\bar{b}$ - Analysis strategy

The correct reconstruction of highly unstable, heavy particles like the top quark or the Higgs boson from their decay products is one of the main challenges in experimental particle physics. Therefore, the correct reconstruction of p_T^H is an important ingredient in the definition of the STXS analysis strategy.

In this analysis, the event categorization is built starting from the one optimized for the inclusive $t\bar{t}H$ measurement. In the most sensitive category, namely the ≥ 4 jets, ≥ 3 b tags,

after applying the baseline ANN multiclass leading to the $t\bar{t} + X$ categories, the events in the $t\bar{t}H$ or $t\bar{t} + B$ classes are further split further into sub-categories targeting the STXS bins, i.e. five bins in Higgs boson p_T^H with boundaries at $[0, 60, 120, 200, 300, 450]$ GeV. The events are assigned to the STXS bins according to the output of a further multi-class dedicated ANNs (STXS ANN), with output nodes targeting the 5 STXS bins. For the other classes of ≥ 4 jets, ≥ 3 b tags and 3 jets, ≥ 3 b tags categories, the same templates as in the $t\bar{t}H$ inclusive measurement are considered. The strategy with the independent final observables that enter the 5D simultaneous fit are schematically depicted in Figure 7.4.

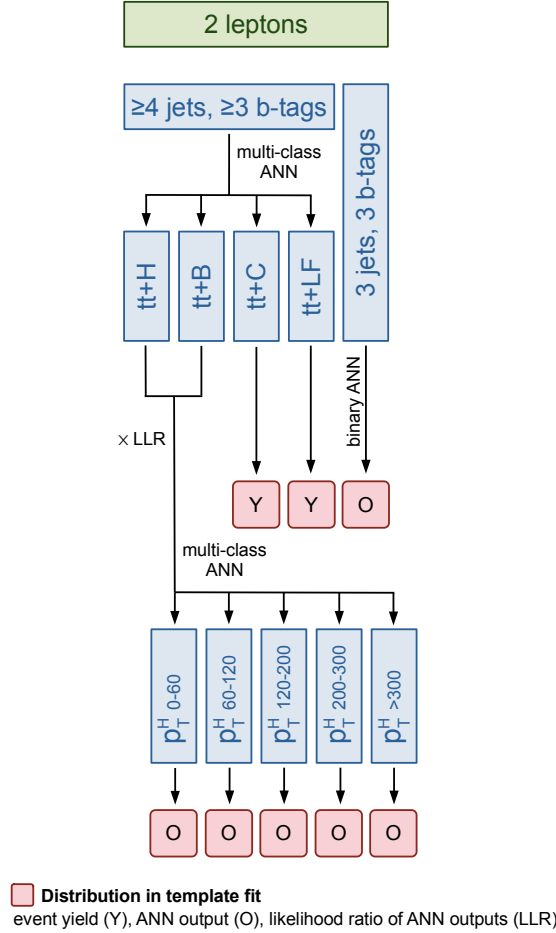


Figure 7.4: Illustration of the analysis strategy for the STXS measurement in the dilepton channel. The illustrated procedure is applied separately for all three years of data taking.

The dedicated STXS ANN is trained only for $t\bar{t}H$ simulated events in the ≥ 4 jets, ≥ 3 b tags category that are within acceptance, i.e. rapidity $|Y| < 2.5$. One single classifier is trained for all three data-taking years. The same $t\bar{t}H$ samples that were used to train the classifier for the inclusive measurement are also used here. The number of events used for training, evaluation, and testing is summarized in Table 7.1, and also the number of events per STXS category are reported.

The STXS ANN classifies events according to the STXS bins and assigns a probability-like number for each of the classes for each event. The STXS ANN features as input variables both the properties of the individual objects and the whole event. In addition to the set of input variables used previously in the inclusive analysis, the di-jet transverse momentum of the 2 jets with the mass closest to the nominal Higgs boson mass (125 GeV) has been

$\geq 4 \text{ jets}, \geq 3 \text{ b tags}$	
events for training	852660
events for validation	284219
events for testing	284219

events per process	
$p_T^H \leq 60$	285369
$60 \leq p_T^H \leq 120$	456848
$120 \leq p_T^H \leq 200$	403542
$200 \leq p_T^H \leq 300$	198754
$300 \leq p_T^H$	76585

Table 7.1: Number of events used to train the STXS ANNs in the $\geq 4 \text{ jets}, \geq 3 \text{ b tags}$ category. Also listed are the number of events per STXS category and the number of events in the training, test, and validation subsamples.

introduced to expand the set (see Figure 7.5). According to the unbinned one-dimensional Kolmogorov-Smirnov (KS) test and a Taylor expansion of the ANNs outputs, this variable has the largest discrimination power to further classify events as a function of the generated Higgs boson p_T^H . A simpler approach has been explored using just this variable alone as template for the fit, rather than as one of the inputs to the ANN classifier, but it was found to give worse performance.

The quality of each observable in the MC simulation with respect to data is evaluated with the binned Goodness-Of-Fit (GoF) tests, based on the full uncertainty model, using the saturated model method. In each analysis region, the test is performed for each individual observable and for the correlation between all pairs of input variables. Only input features with a p-value $\geq 5\%$ are further considered for the training.

The architecture is a fully connected feed-forward network implemented in KERAS. The optimization of the hyperparameters follows a similar strategy as for the inclusive measurement (see Section 5.7.2). The final architecture of the STXS ANN is summarized in Table 7.2.

Figure 7.6 (a) shows the training metric for the final training of the STXS-ANN classifier; no overtraining is observed. The confusion matrix for the STXS ANN classifier is shown in Figure 7.6 (b). The reconstruction efficiency implied by the matrix shows a larger amount of confusion in the bins below $p_T^H \leq 200 \text{ GeV}$ and in the neighboring classes. This classification power is reasonable because the behavior reflects the features of some of the input variables. The values of these observables are expected to be more distinct between the very low- and very high- p_T^H , whereas the difference between the neighboring STXS classes are less pronounced. For example, the distance between two b-tagged jets is expected to decrease with increasing p_T^H since the Higgs boson decay products become more collimated; or for the average transverse momentum of the b-tagged jets, since the Higgs boson decays into bottom quarks, the emerging jets are naturally expected to have a larger p_T on average with increasing p_T^H .

The normalized outputs of the ANN discriminants for the different STXS categories are shown in Figure 7.7.

It has been explored which metric is better to employ to separate the STXS bins at the

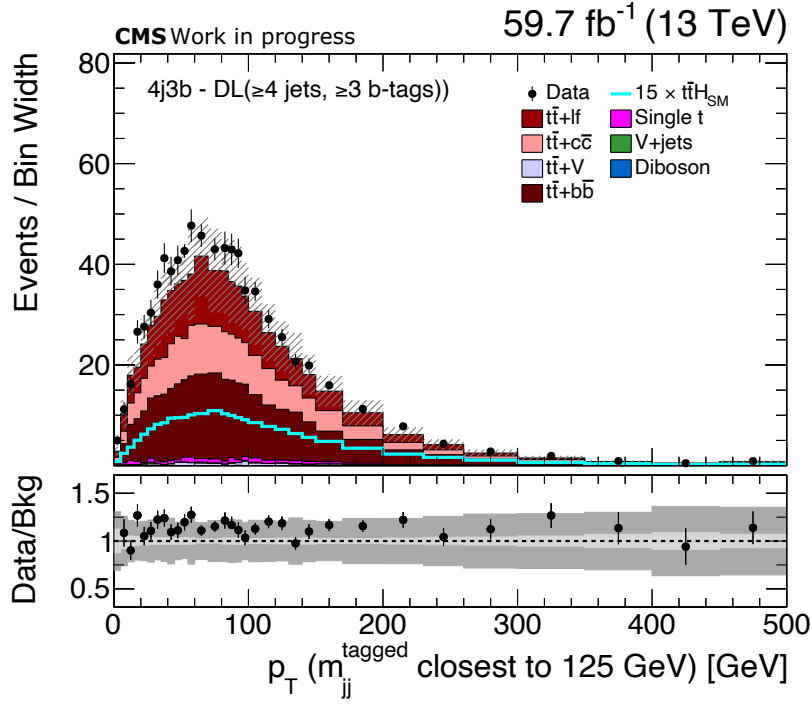


Figure 7.5: Di-jet transverse momentum of the two b -tagged jets with the mass closest to 125 GeV in the ≥ 4 jets, ≥ 3 b tags category for the 2018 data set. The stacked histograms show the distributions of the background processes. The black dots indicate the observed data. The bottom half of the figure shows the ratio of observed data with respect to the background prediction. The light gray band indicates the systematic uncertainties, while the dark gray band illustrates the statistical uncertainties in each bin of the background prediction.

≥ 4 jets, ≥ 3 b tags	
Optimizer	Adagrad
Loss function	categorical cross-entropy
Number of input variables	31
Hidden Layers	3
Nodes per Hidden Layer	[512,256,128]
Node Activation	leakyrelu
Output Node activation	Softmax
L2 regularization	10^{-4}
L1 regularization	10^{-5}
Dropout Percentage	0.0001
Early stopping min epochs	100
Batchsize	2048
Trainable parameters	184 837

Table 7.2: Hyperparameters of the STXS neural networks in the ≥ 4 jets, ≥ 3 b tags category.

reconstruction level. The two primary options are the most discriminating variable (i.e. the p_T of the di-jet system with invariant mass closest to 125 GeV as in Figure 7.5) or the output of the STXS multiclass ANN. Figure 7.8 shows the confusion matrices obtained in both scenarios

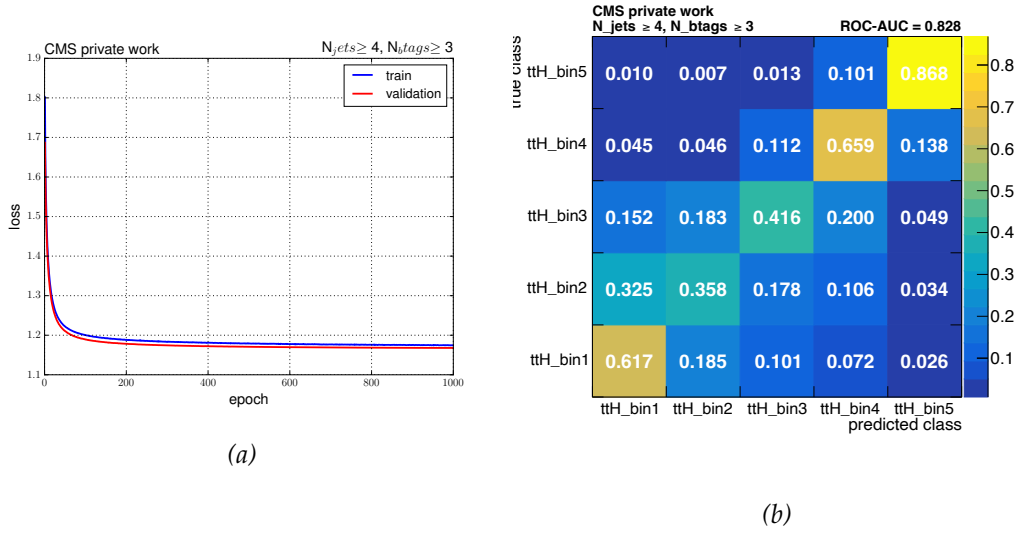


Figure 7.6: (a) Development of loss function values in the STXS ANN training of the ≥ 4 jets, ≥ 3 b tags category. (b) Confusion matrix for the STXS ANN classifier in the ≥ 4 jets, ≥ 3 b tags category evaluated with the test data set. The horizontal axis corresponds to the predicted class of an event and the vertical axis to its true class. The values are normalized to one per row, hence representing the classification accuracies (diagonal elements) and misclassification rates (off-diagonal elements) per process. The STXS bins are defined by the Higgs boson p_T with boundaries at 60, 120, 200 and 300 GeV.

after removing all background contributions: the STXS ANN is beneficial for enriching the STXS bins and having a more diagonal matrix. As depicted in Figure 7.9, these two scenarios have been tested also in terms of stability, purity, efficiency. The stability denotes the number of events that are generated and correctly reconstructed in a given bin i relative to the number of events that are generated in bin i but reconstructed anywhere. The purity denotes the number of events that are generated and correctly reconstructed in a given bin i relative to the number of events that are reconstructed in bin i but in this case generated anywhere. The efficiency is the number of events reconstructed in a given bin which are matched at generator level divided by the total number of events which have been matched at generator level. Following these studies, the STXS multiclass ANN has been chosen to define the STXS bins at reconstruction level.

In addition, it has been thoroughly investigated which distributions provide the best discriminating power to distinguish between the various STXS signal templates at generator levels as well as the ability to identify the signals from the background. Having established that the STXS bin at reconstruction level is defined by the STXS ANN, various templates have been developed, ranging from using the p_T of the di-jet system with the closest invariant mass to 125 GeV to the STXS ANN output weighted with the output of the baseline ANN (i.e. each event in the STXS categories is weighted additionally by the output of the previous inclusive categorization). This weighting procedure not only categorizes events into different STXS bins, but it also helps retain information about the distinction between signal and background processes from the first categorization. Table 7.3 presents the post-fit uncertainties of fits to SM-like Asimov data ($\mu = 1$) for STXS binning templates. Each STXS bin corresponds to a parameter of interest, and for this simultaneous multidimensional fit on SM like Asimov data all systematics were included. It was opted to utilize as the final template the scenario with the best performance: the STXS multiclass ANN weighted by the baseline ANN.

In conclusion, the final observable that enters the parameter estimation is the distribution

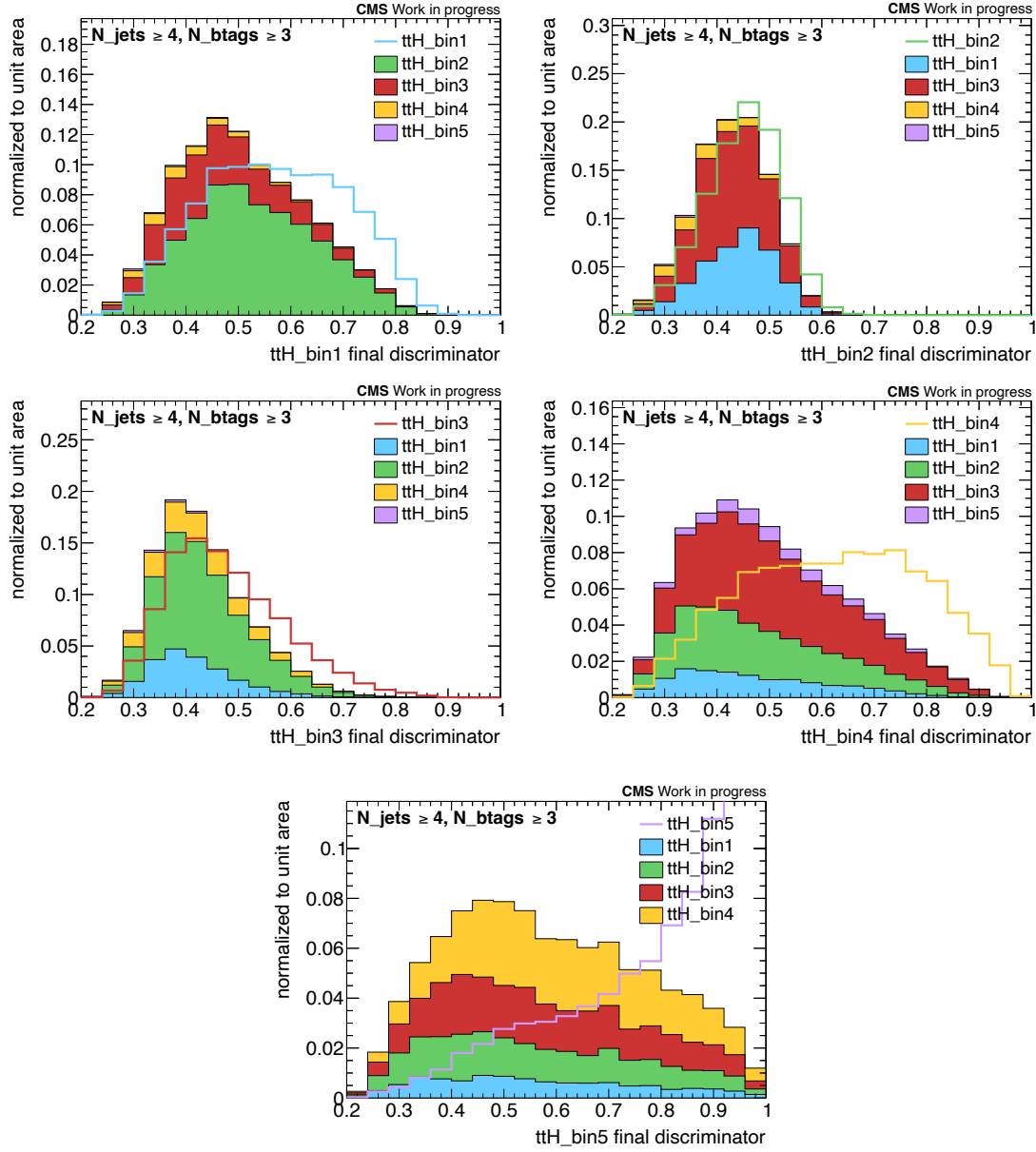


Figure 7.7: Distribution of STXS templates in each bin for each STXS multiclass ANN node in the ≥ 4 jets, ≥ 3 b tags category. The STXS bin template with the highest separation from the others is visible for each plot, and corresponds exactly to that ANN node.

Fitted templates	r_1	r_2	r_3	r_4	r_5
baseline ANN	-6.94/ + 7.08	-7.21/ + 7.71	-5.13/ + 5.21	-3.60/ + 3.78	-2.92/ + 3.31
STXS ANN	-7.94/ + 8.38	-8.56/ + 8.42	-5.35/ + 5.38	-3.69/ + 3.85	-3.45/ + 3.68
STXS ANN * baseline ANN	-5.85/ + 6.32	-6.78/ + 7.14	-3.88/ + 4.14	-2.68/ + 2.96	-2.71/ + 3.06

Table 7.3: Post fit uncertainties for each STXS bin for a multidimensional fit on SM like Asimov data for different fitting templates, with all systematics considered. For all scenarios the STXS bins at reconstruction level are defined by the STXS multiclass ANN and they correspond to 5 STXS Higgs boson p_T bins with boundaries at [0, 60, 120, 200, 300] GeV.

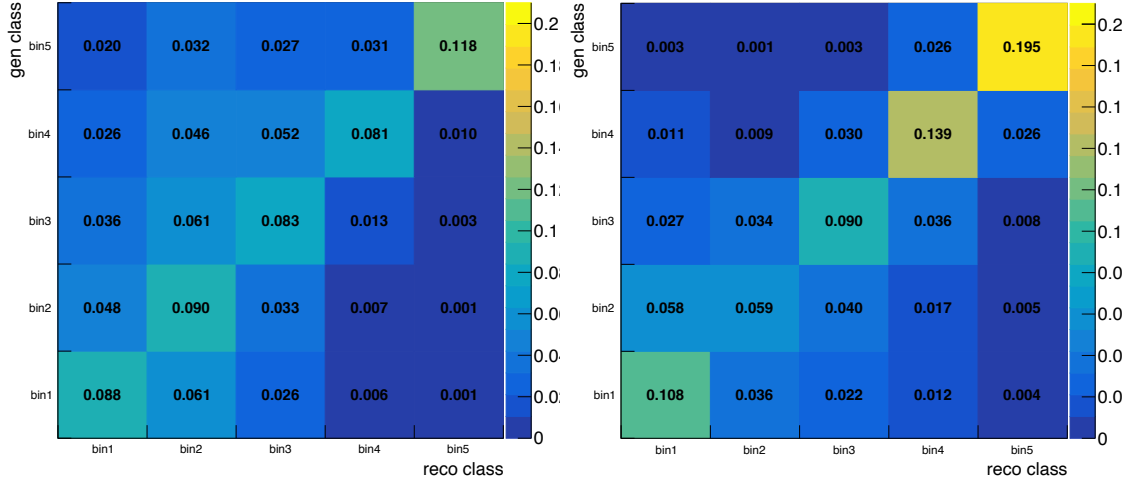


Figure 7.8: Confusion matrix (left) when the p_T of the di-jet system with invariant mass closest to 125 GeV is used to split the STXS bins at reconstruction level or confusion matrix (right) when the STXS multiclass ANN is considered. In both cases all background contributions were removed. The entire matrices are normalized to unity.

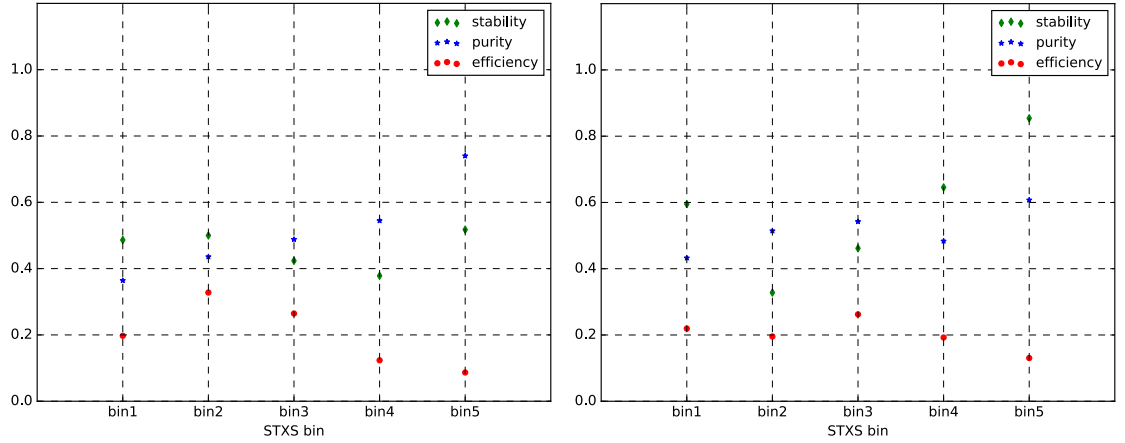


Figure 7.9: Values of stability, purity and efficiency when the p_T of the di-jet system with invariant mass closest to 125 GeV is used to split the STXS bins at Reconstruction level (left) or when the STXS multiclass ANN is considered (right).

of the STXS ANN output values weighted by the corresponding network employed for the process classification in the inclusive measurement. Therefore, this metric provides information regarding both the original process classification and the final assignment to a certain STXS bin. The corresponding prefit distributions for each year are in Figures 7.10 and 7.11 for two of the signal bins, and all the others in Appendix K.3. The output values in the background-related nodes are handled identically to those in the inclusive analysis. This results in powerful discriminators suited for the simultaneous cross section measurement of the p_T^H regions with high control over background processes contributing to the analysis phase space. Therefore, the statistical inference is based on eight categories per period of data-taking.

7.3 STXS measurements results

The results of the inclusive measurements reported in Chapter 6 are interpreted in terms of the signal strength modifier μ_i . Even though this method allows to compare the parameters directly, it is also sensitive to theoretical computations. Any updates to the calculation of the $t\bar{t}H, H \rightarrow b\bar{b}$ process and its uncertainties require re-evaluation of the measurement of the signal strength modifier. Furthermore, two measurements are not comparable they are not based on the same underlying theory calculations.

In the following the results of the final parameter estimation of $t\bar{t}H, H \rightarrow b\bar{b}$ STXS measurement are summarized. Overall the analysis strategy follows the line of the inclusive measurement, as described in Chapter 6. However, in this case, the expected yield y_i for each bin i is modified to allow the simultaneous measurement of the different STXS bins j :

$$\lambda = \sum_j \sigma_j s_{i,j}^{\text{t}\bar{t}\text{H}}(\vec{\theta}) + b_i(\vec{\theta}),$$

with each signal contribution $s_{i,j}$ being multiplied by a parameter σ_j .

The STXS measurement is performed as a direct measurement of the cross section in the various Higgs- p_T bins, and therefore all former theory uncertainties on the inclusive $t\bar{t}H$ signal cross section are omitted (rate parameters modeling effects arising from the renormalization and factorization scales, from additional initial and final state radiation as well as the PDF set). Instead, two types of additional uncertainty are provided to parameterize migration effects between STXS bins (migration uncertainties) and acceptance inside single STXS bins (acceptance uncertainties). They are constructed following central recommendation and they are implemented as rate and shape uncertainty, respectively, and treated as correlated across the years.

Migration uncertainties: The migration uncertainties are built using an adaptation of the Steward-Tackmann method [279] by studying the effect of all combinations of factorization and normalization scale variations across each Higgs- p_T boundary on the total cross section above this boundary. The boundary at 450 GeV from Figure 7.3 is only considered for the migration uncertainty and not in the actual measurement. Since the migration uncertainties only cause normalization effects, in the statistical model they are implemented as $\ln N$ distributed nuisance parameters and amount to 5-10 %.

Acceptance uncertainties: Acceptance uncertainties are constructed following the recommendation from the Higgs group in CMS. The QCD scale variations are utilized for this, where both the renormalization and factorization scales are varied at the same time. Then, for each STXS process, a normalizing factor is applied so that the cross section remains unaltered in an inclusive phase-space, i.e. without any Higgs boson p_T requirements. This uncertainty is decorrelated between low p_T STXS bins ($p_T^H < 300$ GeV) and the high p_T bins ($p_T^H \geq 300$ GeV). The resulting shape variation is typically 1-3% and as large as up to 7% depending on the classifier bin.

In addition to the afore-mentioned sources of uncertainty, the normalizations of the $t\bar{t} + b\bar{b}$ and $t\bar{t} + c\bar{c}$ processes are estimated in situ while the $t\bar{t}H, H \rightarrow b\bar{b}$ process is measured. This takes into consideration the challenging computation of these background processes and allows the statistical model to account for variations between simulation and data. Such discrepancies are to be expected as a result of earlier dedicated measurements [43, 46, 277].

To improve the stability of the final statistical model, some changes are made to the treatment of some systematic uncertainty sources. The JEC are not considered for the small

background processes $t\bar{t}H$, $t\bar{t}V$, $V + \text{jets}$, and diboson because the statistical fluctuations are large in the analysis phase space. Furthermore, each shape uncertainty is examined for a real shape effect on the distributions: if there is no shape effect, it is turned into a rate uncertainty. Finally, only rate uncertainties with an influence of at least 0.1% on a specific process are considered. In the final fit, these checks are done independently for each of the categories. This technique stabilizes the expected results without any impact the sensitivity on the signal strength modifiers and the constraints on the nuisance parameters in question.

The effort of scrutinizing the observed data within the CMS Collaboration is still ongoing at the time of writing, hence the conclusions presented in this chapter reflect the current status.

The discussion follows the same structure as in Chapter 6. Figure 7.10 and 7.11 illustrates the final discriminants prefit and postfit for the second and last STXS bins in the $\geq 4 \text{ jets}, \geq 3 \text{ b tags}$ category, respectively. The complete overview of the other distributions is in Appendix K.3. The separation between the STXS signals and the background process is on a comparable level to the inclusive analysis, due to the deployed strategy itself (i.e. the STXS classification occurs only after the process assignment by the initial neural network). However, the pre-fit plots highlight the separation among the various STXS classes. As expected, for each plot, the distribution of the correspondent STXS is enriched at high discriminant values. This reflects the express design achieved with the additional STXS neural network to be sensitive to that specific bin. At the same time, the discrimination to the other STXS classes increase with p_T^H : the distribution for the last STXS bin peaks at lower discriminant values. Additionally, the number of events categorized into this ANN output reflects this behavior. On the truth level, events from the first STXS bin contribute the most to this observable, while events from the last STXS category contribute the least, reflecting the above STXS ANN confusion matrix. The separation power between neighboring STXS bins is low, but it increases with high p_T^H differences. Also in all the other STXS categories, the observables show a good compatibility between the simulated samples with the observed data within the uncertainties.

The good compatibility is further assessed with the GoF method and the resulting p-values are listed in Tables 7.4.

	2016	2017	2018	full Run 2
Combined	0.71	0.70	0.99	0.99

Table 7.4: Goodness-of-Fit test results for unblinded fit to data with the statistical model used for the STXS measurement.

The nuisance parameters with the largest impact of the inference of the cross section for each STXS bins are reported in Figure 7.12, both for the expected with a pseudo data set corresponding to the SM expectation and for the real data.

Generally, the ranking and the order of magnitude of the impacts match the expectations and the post-fit values are compatible with the expected values within approximately two standard deviations, indicating good compatibility between simulated samples and real data. Despite the fact that the magnitude of the list of uncertainty sources is not exactly the same, especially for the low- p_T^H bins the list of uncertainty is comparable to the one from the inclusive $t\bar{t}H, H \rightarrow b\bar{b}$ analysis. This is indeed expected because the inclusive $t\bar{t}H, H \rightarrow b\bar{b}$ analysis is dominated by low- p_T^H effects, with most of the $t\bar{t}H, H \rightarrow b\bar{b}$ events being in the first STXS bins.

Overall, the modeling uncertainties on the $t\bar{t} + \text{jets}$ background and the experimental uncertainties due to the limited number of simulated samples are expected to be the dominant ones. In particular, the modeling parameters related to the irreducible $t\bar{t} + b\bar{b}$ process are

expected to have a large impact, such as the factorization scales, the ME-PS matching, the underlying event simulation and additional gluon splitting effects. The FSR uncertainties have a larger impact in the last STXS bins because the probability for emissions in the final state is proportional to the energy of the particles, and the last STXS bins are sensitive to effects at high energies by construction. Additionally, the freely-floating normalization parameters for the $t\bar{t} + b\bar{b}$ and the $t\bar{t} + c\bar{c}$ processes are highly ranked, as expected and already discussed for the inclusive measurement. For the parameters related to the limited size of simulated samples, the ones with the highest impact are the ones that vary the content to large discriminant values and expected to be mostly enriched in STXS events attributed to the correspondent STXS bin. The impact on the measurement in the final STXS bin is larger, considering the smaller number of events expected for the high- p_T^H region.

On the experimental side, there are some sources of JES and JER, mostly for the first STXS bins. These uncertainties model the fluctuations in the jet energy measurements, hence they can produce event migrations within the analysis phase space. These migration effects are more evident for jets with lower p_T for the jet selection, resulting in the low- p_T^H region bins being more affected by the JES and JER. Additionally, the b-tagging uncertainty simulating the impacts of $t\bar{t} + \text{LF}$ contamination in the heavy-flavor control region is expected to be noticeable in the low-energy regime because additional heavy-flavor jets are less likely in this region. Also the presence of the migration uncertainties among the top ranked nuisance parameters is expected since these uncertainties directly modify the expected signal yield for a given STXS signal process. The few cases of uncertainties strongly constrained have been studied, identifying the limited number of simulated events in the various distributions as the main cause.

Figure 7.13 presents the signal strength and the corresponding uncertainties for each STXS bins, obtained on the preliminary observed results for each period of data-taking and for the full Run 2 combination. Overall the results are consistent with the SM expectation and same behaviors are visible across the years. As expected, because of the small $t\bar{t}H$ production individual cross section per STXS bin, all measurements are heavily dominated by statistical uncertainty. This is in line with the behavior previously observed in the impacts ranking of the nuisance parameters.

7.3.1 Contribution to the combined analysis

An overview of the STXS measurement in the full analysis phase space considering also the fully-hadronic and the semileptonic channels, briefly discussed in Chapter 6 in the context of the inclusive $t\bar{t}H$ measurement, is provided in the following.

In the FH channel, a Higgs boson candidate is reconstructed from the four jets with the highest b-tagging discriminant value and whose invariant dijet mass is most compatible with that of a 125 GeV Higgs boson, based on a χ^2 criterion that takes into account the experimental resolution of the jet p_T . The STXS signal region events are further classified and allocated to one of the five Higgs boson p_T ranges based on the p_T of the reconstructed Higgs boson candidate, resulting in around 45-59% correct assignments for $t\bar{t}H$ events. Thus, for each data-taking periods there are a total of 15 STXS categories in the FH channel and the ANN output distribution is used as the final discriminant in each category without any further optimization.

In the SL, the most signal-like events are categorized further by targeting the five regions in Higgs boson p_T , similarly to the DL approach. In particular these are all the events entering the likelihood ratio discriminant for the (5 jets, ≥ 4 b tags) and (≥ 6 jets, ≥ 4 b tags) categories. The STXS categorization is performed deploying an additional multi-classification ANN, and the STXS ANN output value in that node weighted by the observable value of the

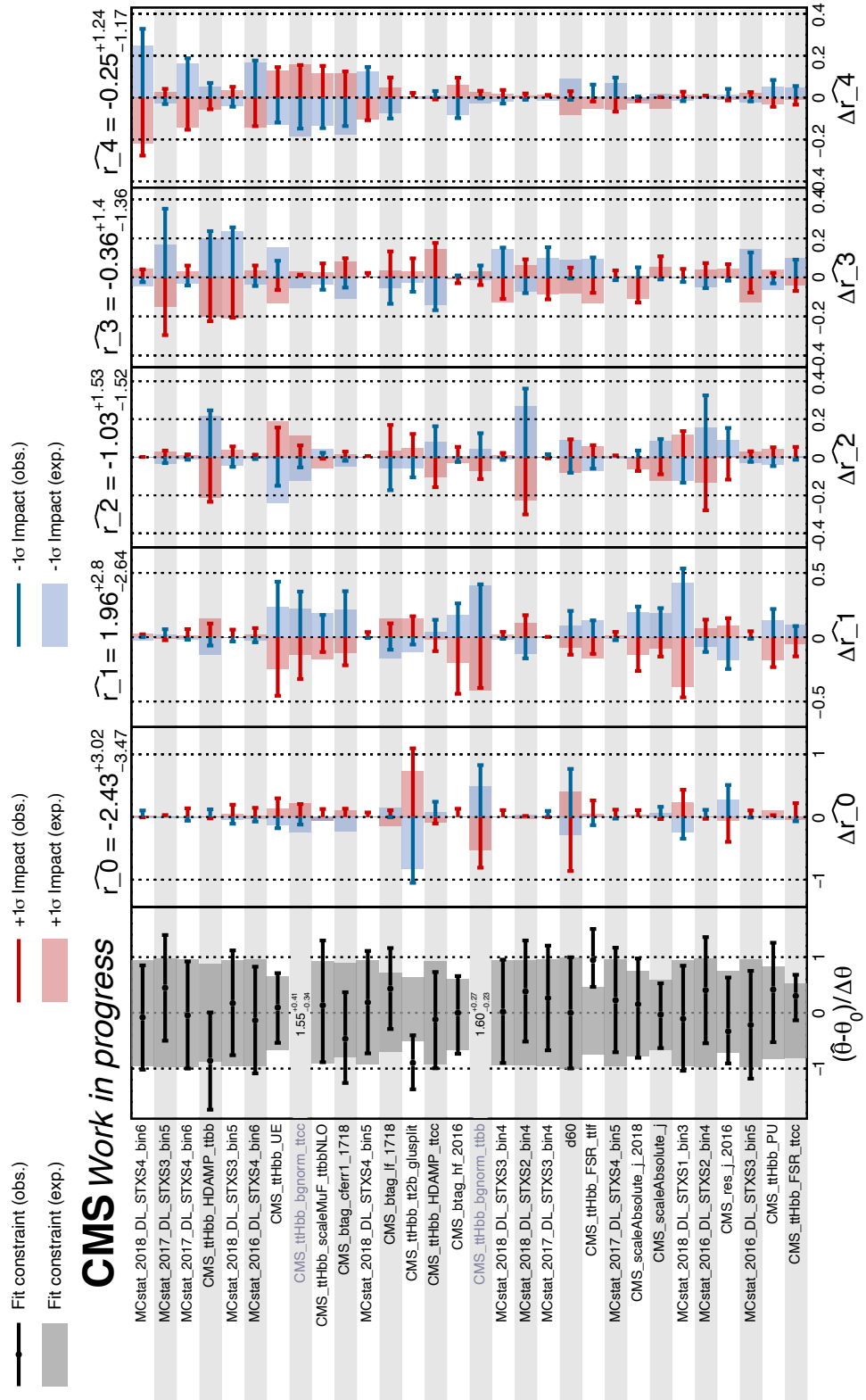


Figure 7.12: Impacts in the STXS bins for the full Run 2 combination. The middle panel displays the post-fit pulls and constraints, while the right-hand side panels highlight the magnitude of the impact on the signal inference for both a pseudo data set corresponding to the SM (expected) and the data (observed).

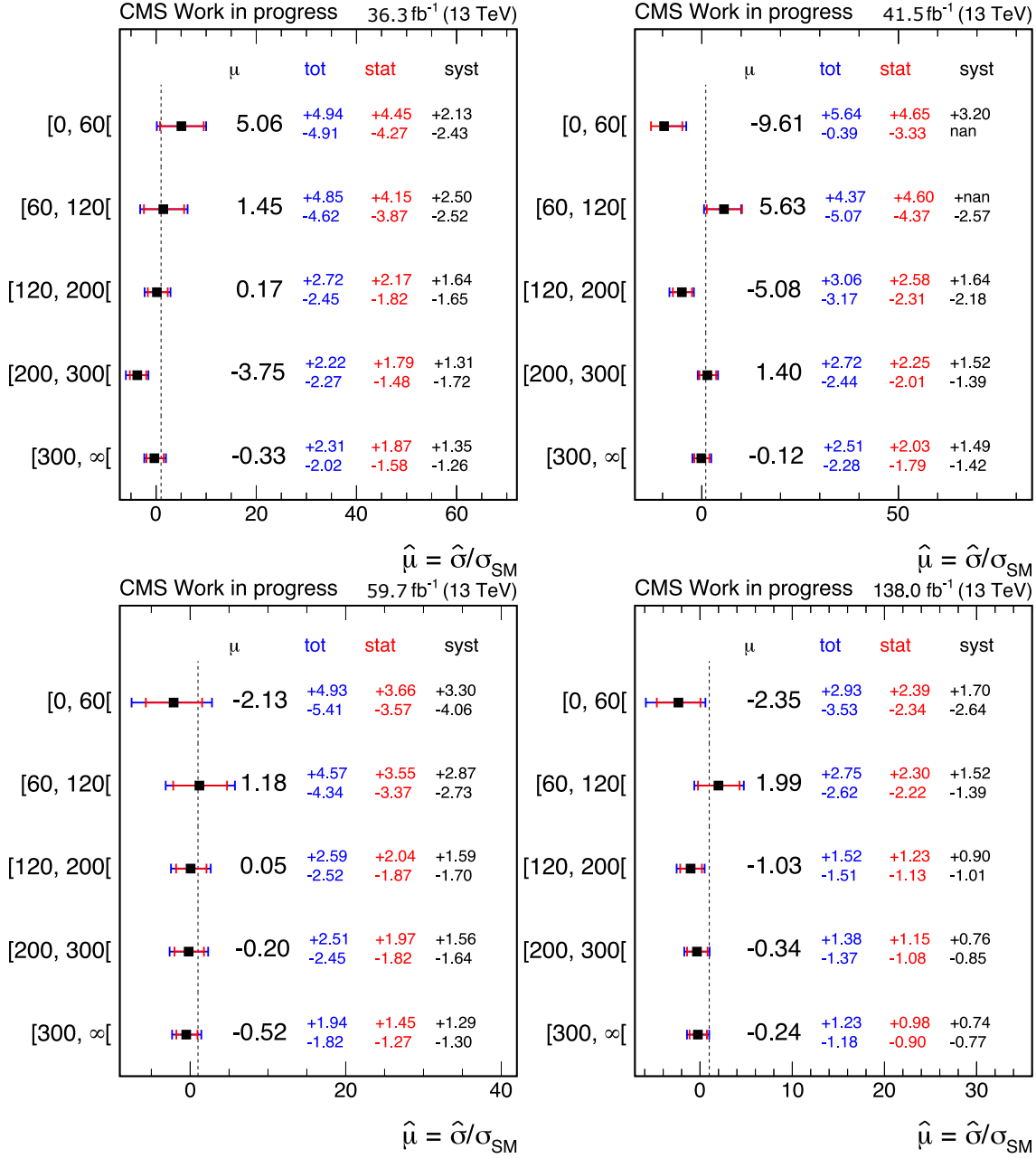


Figure 7.13: Best fit values of the signal strength modifiers μ obtained in the fit of the individual year data sets and in the full combination. Also shown are the 68% expected confidence intervals (outer error bar), also split into their statistical (inner error bar) and systematic components.

inclusive $t\bar{t}H$ measurement is used as final discriminating observable. Thus, in total, there are ten STXS categories in the SL channel, corresponding to the five Higgs boson p_T ranges in each of the jet multiplicity categories. All other events are classified in the same way as in the inclusive $t\bar{t}H$ measurement.

Figure 7.14 illustrate the analysis strategy for the STXS measurement in the whole phase space of the $t\bar{t}H$, $H \rightarrow b\bar{b}$ analysis, applied separately for all three years of data taking.

Also here the presentation of the results follows the same structure as above. Figures 7.15 and 7.16 illustrate the final discriminants postfit for the first and fourth STXS bins in the ≥ 4 jets, ≥ 3 b tags category, respectively, comparing the fit to only DL (left) to the one considering all channels (right). The complete overview of the other distributions is in Appendix K.4.

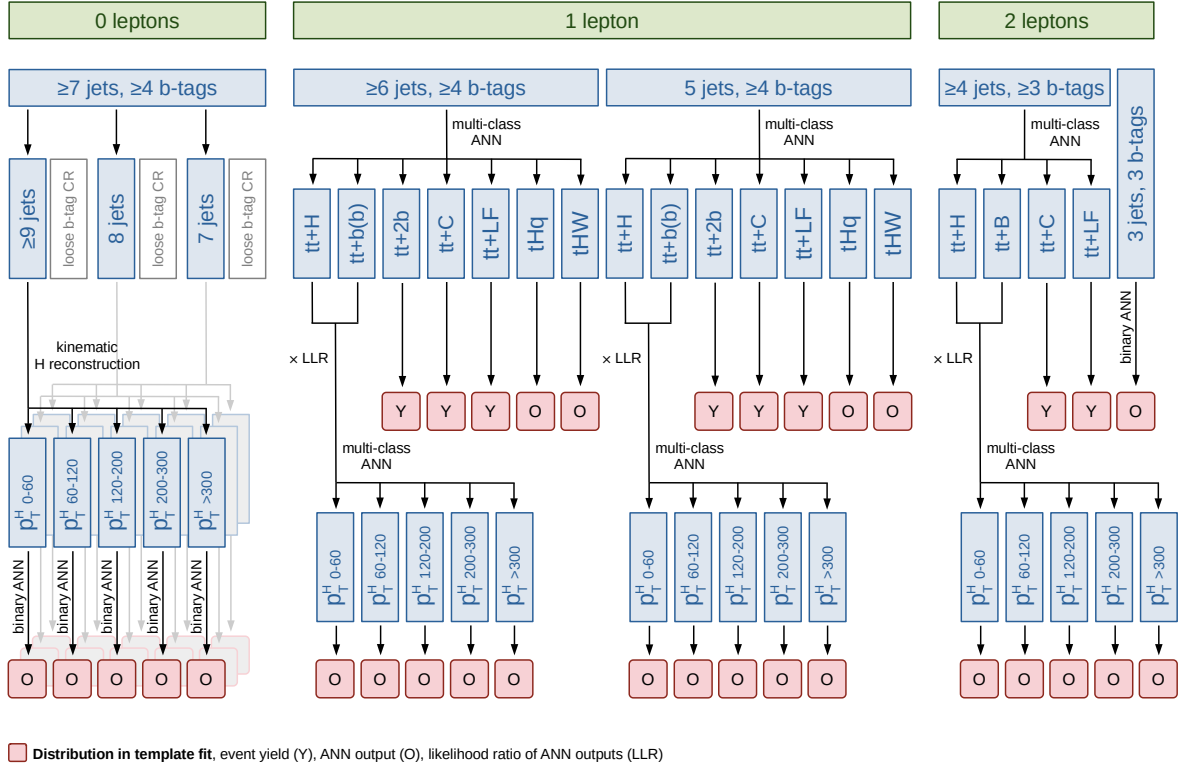


Figure 7.14: Illustration of the analysis strategy for the STXS measurement as it is applied separately for each year of data taking.

As for the inclusive measurement presented in Section 6.3, the overall agreement is good in both cases. However the full combination benefits from the sensitivity arising from the other channels, mostly from the SL channel and this results in the smaller uncertainties bands. The good compatibility is further evaluated using the Goodness-of-Fit (GoF) approach, with the resulting p-values shown in Table 7.5 for each channel and year.

	2016	2017	2018	full Run 2
FH Combined	0.87	0.60	0.79	0.85
SL Combined	0.97	0.99	0.94	0.98
DL Combined	0.71	0.70	0.99	0.99
All channels Combined	0.89	0.43	0.94	0.75

Table 7.5: Goodness-of-Fit test results for unblinded fit to data.

The inclusion of the fully-hadronic and the semileptonic $t\bar{t}$ decay modes enhances the sensitivity of the cross section measurement in the STXS framework. In Figure 7.18 the expected and observed high-ranked impacts are presented. In general, the post-fit estimations and constraints for the different parameter values in the full $t\bar{t}H, H \rightarrow b\bar{b}$ measurement are reasonable and in agreement with the estimation in the DL channel within one standard deviation, showing that the statistical models are capable of describing the observed data. However the overall ranking of the parameters agrees with the results obtained from a SL channel fit, reflecting the higher sensitivity of this channel. The order of magnitude of the impacts in the fully-combined $t\bar{t}H, H \rightarrow b\bar{b}$ measurement is smaller than in the

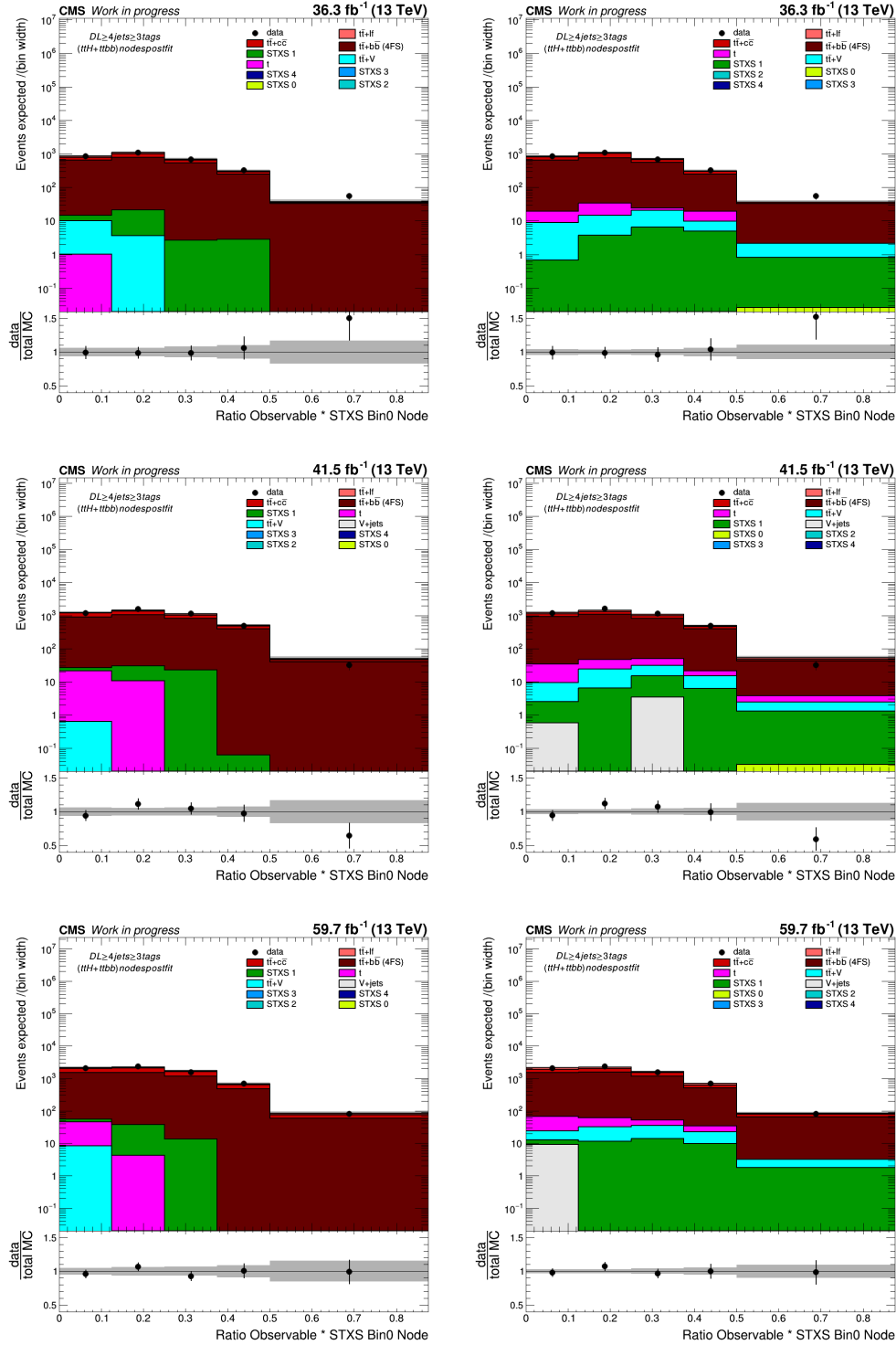


Figure 7.15: Discriminants in the DL channel in the second STXS bin ≥ 4 jets, ≥ 3 b tags $ttH + tt + b\bar{b}$ ratio observable category, for a fit considering only the DL statistical model (left) and all channels (right). Shown are the pre-fit (left) and post-fit (right) distributions for 2016 (top), 2017 (middle), and 2018 (bottom). The black points represent the observed data. The stacked distributions correspond to the simulations for the respective processes. The uncertainty band corresponds to the uncertainties in the full Run 2 combination. For better visibility, the signal template is scaled to the total background contribution in the pre-fit figures (cyan line).

single channels fits. The uncertainties due to the limited size of the simulated samples are mostly related to the SL channel and here in a larger fraction than in a single channel fit. The additional information from the other decay channels increases the control over the theoretical and experimental uncertainties in the fully-combined $t\bar{t}H, H \rightarrow b\bar{b}$ measurement, increasing the importance of the parameters related to the limited number of events available.

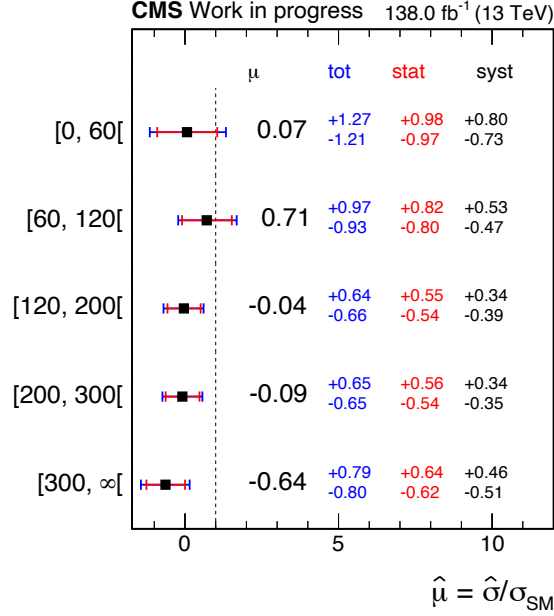


Figure 7.17: Best fit values of the signal strength modifiers μ obtained in the fit to the full combination for all years and channels. Also shown are the 68% expected confidence intervals (outer error bar), also split into their statistical (inner error bar) and systematic components.

Finally, Figure 7.17 shows the measurement results using the full $t\bar{t}H, H \rightarrow b\bar{b}$ combination. As expected, the additional information from the other channels greatly improves the sensitivity in all STXS bins by a better control over the systematic uncertainties. The contributions of the systematic uncertainties get reduced by around half in all bins. In addition, thanks to the more data available for the full $t\bar{t}H, H \rightarrow b\bar{b}$ phase space, also the statistical precision improve in all STXS bins by a factor 3. The sensitivity of this measurement is comparable with the latest results published by the ATLAS Collaboration presented in Ref. [124].

In the CMS Collaboration, this STXS measurement is the first statistical analysis providing details about p_T^H in the $t\bar{t}H, H \rightarrow b\bar{b}$ production channel, and the work presented in this thesis constitutes the first STXS result in the $t\bar{t}H, H \rightarrow b\bar{b}$ DL channel. Currently, the other available STXS measurements for the $t\bar{t}H$ production channel are for events with the Higgs boson decaying into two photons [280], in the leptonic WW decay [281] and in the four-lepton final state [282]. The plan is to perform a combination of these analyses, along with the one presented in this thesis and including $VH + H(b\bar{b})$, which could potentially increase significantly the sensitivity of future measurements.

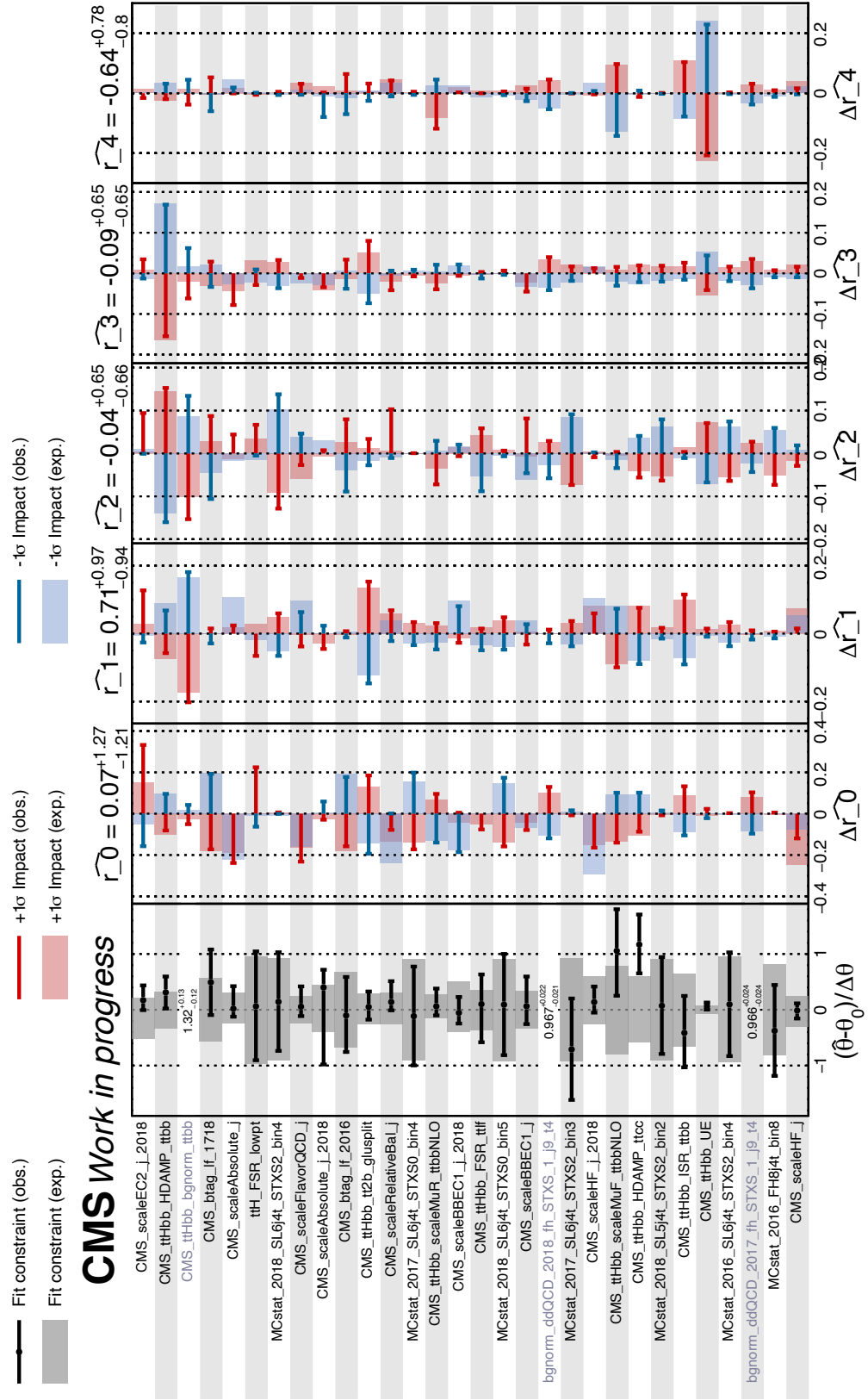


Figure 7.18: Impacts in the STXS bins for the full Run 2 combination and all channels. The middle panel displays the post-fit pulls and constraints, while the right-hand side panels highlight the magnitude of the impact on the signal inference for both a pseudo data set corresponding to the SM (expected) and the data (observed).

CONCLUSION AND OUTLOOK

In this thesis, two primary studies are presented. The first aims to estimate the major sources of uncertainty in the measurement of integrated luminosity with the Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider (LHC), and the second to present inclusive and differential cross section measurements in the context of the Simplified template cross section framework for the Higgs boson production in association with a top quark-antiquark pair in the dilepton decay channel, using data from proton-proton collisions at 13 TeV collected by the CMS experiment.

The CMS experiment utilizes the Van der Meer method for calibrating its luminosity measurement. This involves conducting beam separation scans to achieve a precise luminosity measurement during a specific LHC fill. This information is then employed to calculate the visible cross section of the assigned luminosity detector. However, the calibration process for the luminosity measurement is subject to certain systematic uncertainties, including uncertainties in the beam separation length scale and factorizable beam proton densities. Efforts have been made in the context of this thesis to improve the precision of the integrated luminosity measurement by reducing the bias caused by the assumption of factorizable proton densities. The beam-imaging method was used to reconstruct the transverse proton densities of the two beams, leading to a significant improvement in the experimental results from 1.5% (0.9%) in 2015 (2016) to 0.5% overall. These results have been published in the publication of the integrated luminosity measurement for the 2015 and 2016 data sets of the CMS experiment [26]. Additionally, a new and faster method using for the first time a machine learning approach, developed as part of this thesis, will be included in the upcoming publication for the 2017 and 2018 data sets and is being used for analyzing the data currently collected at the LHC.

Following the discovery of the Higgs boson, the ATLAS and CMS Collaborations accurately measured its properties and couplings to other particles. Among all fermions, the standard model (SM) predicts the top quark to couple with the Higgs boson with the strongest Yukawa coupling. This thesis presents the analysis of the associated production of a Higgs boson with a top quark pair ($t\bar{t}H$) using the full Run 2 data set recorded at the CMS experiment at a center-of-mass energy of $\sqrt{s} = 13$ TeV, with events where the Higgs boson decays into a pair of bottom quarks and in the dilepton final state of the $t\bar{t}$ system. In particular, the $t\bar{t}H, H \rightarrow b\bar{b}$ analysis is of special interest not only due to its direct access to the top-Higgs Yukawa coupling but also due to the possible sensitivity to contributions from physics beyond the standard model (BSM).

The primary challenge of the analysis in this thesis is the significant contribution of $t\bar{t} + \text{jets}$ processes, particularly the irreducible $t\bar{t} + b\bar{b}$ background. The critical $t\bar{t} + b\bar{b}$ background modeling was improved through simulation at next-to-leading order and a freely-floating normalization to address this challenge. To further reduce its impact, optimize the analysis strategy, and enhance the sensitivity to the $t\bar{t}H, H \rightarrow b\bar{b}$ signal, a sophisticated multiclass artificial neural network was developed and deployed for the first time in the dilepton

channel. Additionally, a more advanced validation technique was established for the input variables. The traditional one-dimensional comparison between simulated and recorded event variables was deemed insufficient for deep neural networks due to their ability to exploit deep correlations. This new method covers the validation of higher-order correlations by assessing fit quality using saturated goodness-of-fit tests on two-dimensional distributions of all variables. Only variables that produced high p-values in combination with most other variables are considered input variables to the neural networks.

In this thesis, a statistical model was developed and applied to measure the inclusive $t\bar{t}H, H \rightarrow b\bar{b}$ production. The model has been shown to describe the observed data well. Results were obtained by simultaneously analyzing data in multiple event categories in the $t\bar{t}$ dilepton channel, defined based on the number of jets and enriched with events of one of the considered physics processes determined by a neural network. For the less sensitive categories, a fit was performed on the distribution of the network output corresponding to the process category. For the most sensitive category, a better separation between the $t\bar{t}H$ signal and the crucial $t\bar{t} + b\bar{b}$ background was achieved by using a discriminant in the form of a likelihood ratio based on the network output values.

The observed results of this analysis are consistent with earlier $t\bar{t}H, H \rightarrow b\bar{b}$ measurements, both in the dilepton channel and in combination with other $t\bar{t}$ final states. As anticipated, the largest impact on the signal strength estimation arises from parameters related to the modeling of the $t\bar{t} + \text{jets}$ background, as well as uncertainties related to the b-tagging method and the limited size of the used simulated samples. The observed best-fit value for the signal strength modifier is:

$$\mu_{t\bar{t}H} = 0.34^{+0.26}_{-0.27} = 0.34^{+0.17}_{-0.17}(\text{stat.})^{+0.20}_{-0.21}(\text{syst.}).$$

This result is compatible with the Standard Model prediction and corresponds to an observed significance of 1.3σ compared to the background-only hypothesis. A value of $\mu_{t\bar{t}H}^{\text{exp.}} = 1.00^{+0.28}_{-0.25}$ and 4.1σ was expected from simulation.

The presented results improve compared to the previous CMS measurements in this channel due to increased integrated luminosity and refined analysis methods. The measured value agrees with the latest measurement performed by the ATLAS experiment using a similar amount of proton-proton collision data at a center-of-mass energy of 13TeV, but the analysis presented here is considerably more accurate.

Moreover, this thesis presents the first CMS measurement of the $t\bar{t}H, H \rightarrow b\bar{b}$ signal strength differentially as a function of the transverse momentum of the Higgs boson within the Simplified Template Cross Section (STXS) framework. Despite the large data set, the measurement is still dominated by statistical uncertainties, not only in the dilepton channel but also when combined with the other $t\bar{t}H, H \rightarrow b\bar{b}$ channels. As in the inclusive $t\bar{t}H, H \rightarrow b\bar{b}$ measurement, the cross section estimation in the low- p_T region is impacted by systematic uncertainties associated with $t\bar{t} + \text{jets}$ modeling and experimental uncertainties. Meanwhile, uncertainties in the high- p_T region primarily stem from the limited size of simulated samples and the modeling of the $t\bar{t} + b\bar{b}$ process. The $t\bar{t}H, H \rightarrow b\bar{b}$ measurements presented in this thesis are currently under review (“pre-approved”) within the CMS Collaboration, and the corresponding journal publication is ongoing.

The analysis of $t\bar{t}H, H \rightarrow b\bar{b}$ would benefit significantly from dedicated $t\bar{t} + b\bar{b}$ measurements and improvements in theoretical models to control better and describe these leading background processes. Currently, the cross-section measurements for $t\bar{t} + b$ jets take into account all processes that result in the production of the $t\bar{t}$ -system with additional b jets. A deeper understanding of multi-scale production can be achieved by separating the produc-

tion into its components from $t\bar{t} + b\bar{b}$, $t\bar{t} + b$, and $t\bar{t} + 2b$ processes and by determining their relative contributions in the total cross-sections. In addition, larger data sets will enable more accurate measurements of the $t\bar{t} + b\bar{b}$, $t\bar{t} + b$, $t\bar{t} + 2b$ cross-sections. Ultimately, a simultaneous measurement of similar processes like $t\bar{t}H, H \rightarrow b\bar{b}$ and $t\bar{t} + b\bar{b}$ production would enhance the precision of their cross section measurements and facilitate a more straightforward interpretation of new physics models that simultaneously affect all the processes. The precise measurement of $t\bar{t}H$ production is a key goal of the LHC physics program to understand the electroweak symmetry breaking. The $t\bar{t}H, H \rightarrow b\bar{b}$ channel will become even more important as more data becomes available. Future $t\bar{t}H, H \rightarrow b\bar{b}$ analyses may also be able to probe more of the phase space with differential measurements, particularly in the high- p_T regime where sensitivity to new physics is high. No significant deviations from the SM prediction have been observed at the time of writing this thesis. However, to account for possible new physics contributions, effective field theories (EFT) have been developed that approximate the low-energy effects of high-energy BSM physics. The EFT Lagrangian includes higher-dimensional operators increasingly suppressed at higher energy scales of BSM physics. The $t\bar{t}H$ process is expected to increase the sensitivity to the contributions of three operators in particular: the top-Higgs Yukawa coupling, which also affects pairwise Higgs production with a $t\bar{t}$ pair, the monochromatic dipole moment, and the effective gluon-Higgs coupling, both of which are expected to modify the shape of the Higgs boson p_T spectrum for high p_T values.

Experimental particle physics aims to understand the limitations of the Standard Model and discover new physics by collecting more data from high-energy proton-proton collisions at the LHC. The ongoing Run 3 of the LHC will collect more data at higher center-of-mass energy. By the 2030s, a data set of 3000 fb^{-1} is expected to be delivered to the CMS experiment at the High-Luminosity LHC with five to seven times larger instantaneous luminosity. The projected sensitivity used for the formulation of the European Strategy for Particle Physics for this analysis is around 10%, with the largest systematic uncertainties being related to the modeling of the signal process and the $t\bar{t} + \text{jets}$ background. However, the currently expected sensitivity has already exceeded the projected value for the then-expected Run 2 data set, indicating that even more sophisticated measurements may be possible to reach lower sensitivities in the future.

BIBLIOGRAPHY

- [1] F. Englert and R. Brout. *Broken Symmetry and the Mass of Gauge Vector Mesons*. *Phys. Rev. Lett.*, 13:321–323, 1964. doi:10.1103/PhysRevLett.13.321.
- [2] P. W. Higgs. *Broken Symmetries and the Masses of Gauge Bosons*. *Phys. Rev. Lett.*, 13:508–509, 1964. doi:10.1103/PhysRevLett.13.508.
- [3] P. W. Higgs. *Broken symmetries, massless particles and gauge fields*. *Phys. Lett.*, 12:132–133, 1964. doi:10.1016/0031-9163(64)91136-9.
- [4] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. *Global Conservation Laws and Massless Particles*. *Phys. Rev. Lett.*, 13:585–587, 1964. doi:10.1103/PhysRevLett.13.585.
- [5] Philip W. Anderson. *Plasmons, Gauge Invariance, and Mass*. *Phys. Rev.*, 130:439–442, 1963. doi:10.1103/PhysRev.130.439.
- [6] Fabio Maltoni, Eleni Vryonidou, and Cen Zhang. *Higgs production in association with a top-antitop pair in the Standard Model Effective Field Theory at NLO in QCD*. *JHEP*, 10:123, 2016. arXiv:1607.05330, doi:10.1007/JHEP10(2016)123.
- [7] Marco Battaglia, Massimiliano Grazzini, Michael Spira, and Marius Wiesemann. *Sensitivity to BSM effects in the Higgs p_T spectrum within SMEFT*. *JHEP*, 11:173, 2021. arXiv:2109.02987, doi:10.1007/JHEP11(2021)173.
- [8] CMS Collaboration. *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*. *Phys. Lett.*, B716:30–61, 2012. arXiv:1207.7235, doi:10.1016/j.physletb.2012.08.021.
- [9] ATLAS Collaboration. *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*. *Phys. Lett.*, B716:1–29, 2012. arXiv:1207.7214, doi:10.1016/j.physletb.2012.08.020.
- [10] *A portrait of the Higgs boson by the CMS experiment ten years after the discovery*. *Nature*, 607(7917):60–68, 2022. arXiv:2207.00043, doi:10.1038/s41586-022-04892-x.
- [11] *A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery*. *Nature*, 607(7917):52–59, 2022. arXiv:2207.00092, doi:10.1038/s41586-022-04893-w.
- [12] Giuseppe Degrandi, Stefano Di Vita, Joan Elias-Miro, Jose R. Espinosa, Gian F. Giudice, Gino Isidori, and Alessandro Strumia. *Higgs mass and vacuum stability in the Standard Model at NNLO*. *JHEP*, 08:098, 2012. arXiv:1205.6497, doi:10.1007/JHEP08(2012)098.
- [13] Fedor Bezrukov and Mikhail Shaposhnikov. *Why should we care about the top quark Yukawa coupling?* *J. Exp. Theor. Phys.*, 120:335–343, 2015. arXiv:1411.1923, doi:10.1134/S1063776115030152.
- [14] Stephen P. Martin. *A Supersymmetry primer*. *Adv. Ser. Direct. High Energy Phys.*, 18:1–98, 1998. arXiv:hep-ph/9709356, doi:10.1142/9789812839657_0001.
- [15] Giuliano Panico and Andrea Wulzer. *The Composite Nambu-Goldstone Higgs*, volume 913. Springer, 2016. arXiv:1506.01961, doi:10.1007/978-3-319-22617-0.

- [16] Gang Li, Ling-Xiao Xu, Bin Yan, and C. P. Yuan. *Resolving the degeneracy in top quark Yukawa coupling with Higgs pair production*. *Phys. Lett. B*, 800:135070, 2020. [arXiv:1904.12006](#), doi:10.1016/j.physletb.2019.135070.
- [17] Csaba Csáki and Philip Tanedo. *Beyond the Standard Model*. In *2013 European School of High-Energy Physics*, pages 169–268, 2015. [arXiv:1602.04228](#), doi:10.5170/CERN-2015-004.169.
- [18] Bruce Hoeneisen. *A Study of Dark Matter with Spiral Galaxy Rotation Curves*. *International Journal of Astronomy and Astrophysics*, 09:71–96, 01 2019. doi:10.4236/ijaa.2019.92007.
- [19] D. Paraficz, J.-P. Kneib, J. Richard, A. Morandi, M. Limousin, E. Jullo, and J. Martinez. *The Bullet cluster at its best: weighing stars, gas, and dark matter*. *Astronomy and Astrophysics*, 594:A121, oct 2016. URL: <https://doi.org/10.1051%2F0004-6361%2F201527959>, doi:10.1051/0004-6361/201527959.
- [20] Arindam Mazumdar, Subhendra Mohanty, and Priyank Parashari. *Evidence of dark energy in different cosmological observations*. *The European Physical Journal Special Topics*, 230, 07 2021. doi:10.1140/epjs/s11734-021-00212-y.
- [21] D. de Florian et al. *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*. 2/2017, 10 2016. [arXiv:1610.07922](#), doi:10.23731/CYRM-2017-002.
- [22] CMS Collaboration. *Observation of $t\bar{t}H$ production*. *Phys. Rev. Lett.*, 120(23):231801, 2018. [arXiv:1804.02610](#), doi:10.1103/PhysRevLett.120.231801.
- [23] A. M. Sirunyan et al. *Observation of Higgs boson decay to bottom quarks*. *Phys. Rev. Lett.*, 121(12):121801, 2018. [arXiv:1808.08242](#), doi:10.1103/PhysRevLett.121.121801.
- [24] Morad Aaboud et al. *Observation of $H \rightarrow b\bar{b}$ decays and VH production with the ATLAS detector*. *Phys. Lett. B*, 786:59–86, 2018. [arXiv:1808.08238](#), doi:10.1016/j.physletb.2018.09.013.
- [25] S. van der Meer. *Calibration of the effective beam height in the ISR*. Technical Report CERN-ISR-PO-68-31. ISR-PO-68-31, CERN, Geneva, 1968. URL: <https://cds.cern.ch/record/296752>.
- [26] CMS Collaboration. *Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS*. *Eur. Phys. J. C*, 81(9):800, 2021. [arXiv:2104.01927](#), doi:10.1140/epjc/s10052-021-09538-2.
- [27] CMS Collaboration. *CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV*. CMS Physics Analysis Summary CMS-PAS-LUM-17-004, CERN, Geneva, 2018. CMS-PAS-LUM-17-004<https://cds.cern.ch/record/2621960>. URL: <https://cds.cern.ch/record/2621960>.
- [28] CMS Collaboration. *CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV*. CMS Physics Analysis Summary CMS-PAS-LUM-18-002, CERN, Geneva, 2019. CMS-PAS-LUM-18-002<https://cds.cern.ch/record/2676164>. URL: <https://cds.cern.ch/record/2676164>.
- [29] M. Beneke, P. Falgari, S. Klein, and C. Schwinn. *Hadronic top-quark pair production with NNLL threshold resummation*. *Nucl. Phys. B*, 855:695, 2012. [arXiv:1109.1536](#), doi:10.1016/j.nuclphysb.2011.10.021.
- [30] Matteo Cacciari, Michal Czakon, Michelangelo Mangano, Alexander Mitov, and Paolo Nason. *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*. *Phys. Lett. B*, 710:612, 2012. [arXiv:1111.5869](#), doi:10.1016/j.physletb.2012.03.013.
- [31] Peter Bärnreuther, Michal Czakon, and Alexander Mitov. *Percent-level-precision physics at the Tevatron: next-to-next-to-leading order QCD corrections to $q\bar{q} \rightarrow t\bar{t} + X$* . *Phys. Rev. Lett.*, 109:132001, 2012. [arXiv:1204.5201](#), doi:10.1103/PhysRevLett.109.132001.

- [32] Michal Czakon and Alexander Mitov. *NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels*. *JHEP*, 12:054, 2012. [arXiv:1207.0236](#), [doi:10.1007/JHEP12\(2012\)054](#).
- [33] Michal Czakon and Alexander Mitov. *NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction*. *JHEP*, 01:080, 2013. [arXiv:1210.6832](#), [doi:10.1007/JHEP01\(2013\)080](#).
- [34] Michal Czakon, Paul Fiedler, and Alexander Mitov. *Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_s^4)$* . *Phys. Rev. Lett.*, 110:252004, 2013. [arXiv:1303.6254](#), [doi:10.1103/PhysRevLett.110.252004](#).
- [35] Michal Czakon and Alexander Mitov. *Top++: a program for the calculation of the top-pair cross-section at hadron colliders*. *Comput. Phys. Commun.*, 185:2930, 2014. [arXiv:1112.5675](#), [doi:10.1016/j.cpc.2014.06.021](#).
- [36] Michiel Botje et al. *The PDF4LHC Working Group Interim Recommendations*. 2011. [arXiv:1101.0538](#).
- [37] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt. *Parton distributions for the LHC*. *Eur. Phys. J.*, C63:189, 2009. [arXiv:0901.0002](#), [doi:10.1140/epjc/s10052-009-1072-5](#).
- [38] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt. *Uncertainties on α_s in global PDF analyses and implications for predicted hadronic cross sections*. *Eur. Phys. J.*, C64:653, 2009. [arXiv:0905.3531](#), [doi:10.1140/epjc/s10052-009-1164-2](#).
- [39] Hung-Liang Lai, Marco Guzzi, Joey Huston, Zhao Li, Pavel M. Nadolsky, Jon Pumplin, and C. P. Yuan. *New parton distributions for collider physics*. *Phys. Rev.*, D82:074024, 2010. [arXiv:1007.2241](#), [doi:10.1103/PhysRevD.82.074024](#).
- [40] Jun Gao, Marco Guzzi, Joey Huston, Hung-Liang Lai, Zhao Li, Pavel Nadolsky, Jon Pumplin, Daniel Stump, and C. P. Yuan. *CT10 next-to-next-to-leading order global analysis of QCD*. *Phys. Rev.*, D89:033009, 2014. [arXiv:1302.6246](#), [doi:10.1103/PhysRevD.89.033009](#).
- [41] Richard D. Ball et al. *Parton distributions with LHC data*. *Nucl. Phys.*, B867:244, 2013. [arXiv:1207.1303](#), [doi:10.1016/j.nuclphysb.2012.10.003](#).
- [42] *LHCHSWG working meeting: discussion on $t\bar{t}+b\bar{b}$ (MC comparisons)*. <https://indico.cern.ch/event/850548/>, October 2019.
- [43] Albert M Sirunyan et al. *Measurements of $t\bar{t}$ cross sections in association with b jets and inclusive jets and their ratio using dilepton final states in pp collisions at $\sqrt{s} = 13$ TeV*. *Phys. Lett. B*, 776:355, 2018. [arXiv:1705.10141](#), [doi:10.1016/j.physletb.2017.11.043](#).
- [44] Tomáš Ježo, Jonas M. Lindert, Niccolo Moretti, and Stefano Pozzorini. *New NLOPS predictions for $t\bar{t} + b$ -jet production at the LHC*. *Eur. Phys. J. C*, 78(6):502, 2018. [arXiv:1802.00426](#), [doi:10.1140/epjc/s10052-018-5956-0](#).
- [45] M. V. Garzelli, A. Kardos, and Z. Trócsányi. *Hadroproduction of $t\bar{t}b\bar{b}$ final states at LHC: predictions at NLO accuracy matched with Parton Shower*. *JHEP*, 03:083, 2015. [arXiv:1408.0266](#), [doi:10.1007/JHEP03\(2015\)083](#).
- [46] Andrej Saibel. *Phenomenology of $t\bar{t}H$ Production with Top Quark Running Mass and the Differential Cross-Section Measurement of $t\bar{t}+b$ -jets Production in the Dilepton Channel at $\sqrt{s} = 13$ TeV with the CMS Experiment*, 2021. Presented 04 Aug 2021. URL: <https://cds.cern.ch/record/2790968>.
- [47] Nicolas Berger et al. *Simplified Template Cross Sections - Stage 1.1*. 6 2019. [arXiv:1906.02754](#).
- [48] C. Burgard D. Galbraith. *UX: Standard Model of the Standard Model*, 2012. URL: <http://davidgalbraith.org/portfolio/ux-standard-model-of-the-standard-model/>.

- [49] M. D. Schwartz. *Quantum Field Theory and the Standard Model*. 2013. ISBN: 978-1107034730.
- [50] M. E. Peskin and D. V. Schroeder. *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995. ISBN: 978-0201503975.
- [51] David J Griffiths. *Introduction to elementary particles; 2nd rev. version*. Physics textbook. Wiley, New York, NY, 2008. URL: <https://cds.cern.ch/record/111880>.
- [52] W. A. Horowitz. *An Extremely Brief Introduction to Quantum Field Theory*. *Journal of Physics: Conference Series*, 802(1):012003, 2017. doi:10.1088/1742-6596/802/1/012003.
- [53] Paul A. M. Dirac. *The quantum theory of the electron*. *Proc. Roy. Soc. Lond.*, A117:610–624, 1928. doi:10.1098/rspa.1928.0023.
- [54] C. D. Anderson. *The Positive Electron*. *Phys. Rev.*, 43:491–494, Mar 1933. doi:10.1103/PhysRev.43.491.
- [55] P.A. Zyla et al. *Review of Particle Physics*. *PTEP*, 2020(8):083C01, 2020. and 2021 update. doi:10.1093/ptep/ptaa104.
- [56] Chen-Ning Yang and Robert L. Mills. *Conservation of Isotopic Spin and Isotopic Gauge Invariance*. *Phys. Rev.*, 96:191–195, 1954. doi:10.1103/PhysRev.96.191.
- [57] H. Burkhardt and B. Pietrzyk. *Recent BES measurements and the hadronic contribution to the QED vacuum polarization*. *Phys. Rev. D*, 84:037502, 2011. arXiv:1106.2991, doi:10.1103/PhysRevD.84.037502.
- [58] Alexandre Deur, Stanley J. Brodsky, and Guy F. de Teramond. *The QCD Running Coupling*. *Nucl. Phys.*, 90:1, 2016. arXiv:1604.08082, doi:10.1016/j.ppnp.2016.04.003.
- [59] K. Ackerstaff et al. *Measurement of the strong coupling constant $\alpha(s)$ and the vector and axial vector spectral functions in hadronic tau decays*. *Eur. Phys. J. C*, 7:571–593, 1999. arXiv:hep-ex/9808019, doi:10.1007/s100529901061.
- [60] S. L. Glashow. *Partial Symmetries of Weak Interactions*. *Nucl. Phys.*, 22:579–588, 1961. doi:10.1016/0029-5582(61)90469-2.
- [61] S. Weinberg. *A Model of Leptons*. *Phys. Rev. Lett.*, 19:1264–1266, 1967. doi:10.1103/PhysRevLett.19.1264.
- [62] A. Salam. *Weak and Electromagnetic Interactions*. *Conf. Proc.*, C680519:367–377, 1968.
- [63] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson. *Experimental Test of Parity Conservation in Beta Decay*. *Phys. Rev.*, 105:1413–1415, Feb 1957. URL: <https://link.aps.org/doi/10.1103/PhysRev.105.1413>, doi:10.1103/PhysRev.105.1413.
- [64] T. Nakano and K. Nishijima. *Charge Independence for V-particles*. *Prog. Theor. Phys.*, 10:581–582, 1953. doi:10.1143/PTP.10.581.
- [65] M. Gell-Mann. *The interpretation of the new particles as displaced charge multiplets*. *Nuovo Cim.*, 4(S2):848–866, 1956. doi:10.1007/BF02748000.
- [66] Francis Halzen and Alan Martin. *Quarks & Leptons: An introductory course in modern particle physics*. John Wiley & Sons, New York, USA, 1984.
- [67] T. W. B. Kibble. *Symmetry breaking in non-Abelian gauge theories*. *Phys. Rev.*, 155:1554–1561, 1967. doi:10.1103/PhysRev.155.1554.
- [68] Cedric Delaunay, Christophe Grojean, and James D. Wells. *Dynamics of Non-renormalizable Electroweak Symmetry Breaking*. *JHEP*, 04:029, 2008. arXiv:0711.2511, doi:10.1088/1126-6708/2008/04/029.

- [69] Wim de Boer. *The Discovery of the Higgs Boson with the CMS Detector and its Implications for Supersymmetry and Cosmology*. In *Time and Matter 2013 (TAM2013) Venice, Italy*, 2013. arXiv:1309.0721.
- [70] MuLan Collaboration. *Improved Measurement of the Positive-Muon Lifetime and Determination of the Fermi Constant*. *Phys. Rev. Lett.*, 99:032001, Jul 2007. doi:10.1103/PhysRevLett.99.032001.
- [71] Tilman Plehn and Michael Rauch. *The quartic higgs coupling at hadron colliders*. *Phys. Rev.*, D72:053008, 2005. arXiv:hep-ph/0507321, doi:10.1103/PhysRevD.72.053008.
- [72] Takeo Matsuoka. *The CKM matrix and its origin*. *Prog. Theor. Phys.*, 100:107–122, 1998. arXiv:hep-ph/9804329, doi:10.1143/PTP.100.107.
- [73] M. Kobayashi and T. Maskawa. *CP violation in the renormalizable theory of weak interaction*. *Prog. Theor. Phys.*, 49:652–657, 1973.
- [74] Nicola Cabibbo. *Unitary Symmetry and Leptonic Decays*. *Phys. Rev. Lett.*, 10:531–533, 1963. doi:10.1103/PhysRevLett.10.531.
- [75] Cristiano Alpigiani et al. *Unitarity Triangle Analysis in the Standard Model and Beyond*. In *5th Large Hadron Collider Physics Conference*, 10 2017. arXiv:1710.09644.
- [76] S. L. Glashow, J. Iliopoulos, and L. Maiani. *Weak Interactions with Lepton-Hadron Symmetry*. *Phys. Rev.*, D2:1285–1292, 1970. doi:10.1103/PhysRevD.2.1285.
- [77] Y. Fukuda et al. *Evidence for oscillation of atmospheric neutrinos*. *Phys. Rev. Lett.*, 81:1562–1567, 1998. arXiv:hep-ex/9807003, doi:10.1103/PhysRevLett.81.1562.
- [78] Q. R. Ahmad et al. *Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory*. *Phys. Rev. Lett.*, 89:011301, 2002. arXiv:nucl-ex/0204008, doi:10.1103/PhysRevLett.89.011301.
- [79] N. Agafonova et al. *Final Results of the OPERA Experiment on ν_τ Appearance in the CNGS Neutrino Beam*. *Phys. Rev. Lett.*, 120(21):211801, 2018. [Erratum: *Phys.Rev.Lett.* 121, 139901 (2018)]. arXiv:1804.04912, doi:10.1103/PhysRevLett.120.211801.
- [80] Ziro Maki, Masami Nakagawa, and Shoichi Sakata. *Remarks on the unified model of elementary particles*. *Prog. Theor. Phys.*, 28:870–880, 1962. doi:10.1143/PTP.28.870.
- [81] B. Pontecorvo. *Inverse beta processes and nonconservation of lepton charge*. *Zh. Eksp. Teor. Fiz.*, 34:247, 1957.
- [82] Rabindra N. Mohapatra and Goran Senjanovic. *Neutrino Mass and Spontaneous Parity Nonconservation*. *Phys. Rev. Lett.*, 44:912, 1980. doi:10.1103/PhysRevLett.44.912.
- [83] Samoil M. Bilenky, J. Hosek, and S. T. Petcov. *On Oscillations of Neutrinos with Dirac and Majorana Masses*. *Phys. Lett. B*, 94:495–498, 1980. doi:10.1016/0370-2693(80)90927-2.
- [84] M. Aker et al. *Direct neutrino-mass measurement with sub-electronvolt sensitivity*. *Nature Phys.*, 18(2):160–166, 2022. arXiv:2105.08533, doi:10.1038/s41567-021-01463-1.
- [85] G. Hernández-Tomé, J. I. Illana, M. Masip, G. López Castro, and P. Roig. *Effects of heavy Majorana neutrinos on lepton flavor violating processes*. *Phys. Rev. D*, 101(7):075020, 2020. arXiv:1912.13327, doi:10.1103/PhysRevD.101.075020.
- [86] Gianfranco Bertone and Dan Hooper. *History of dark matter*. *Rev. Mod. Phys.*, 90(4):045002, 2018. arXiv:1605.04909, doi:10.1103/RevModPhys.90.045002.
- [87] Alexandre Refregier. *Weak gravitational lensing by large scale structure*. *Ann. Rev. Astron. Astrophys.*, 41:645–668, 2003. arXiv:astro-ph/0307212, doi:10.1146/annurev.astro.41.111302.102207.

- [88] N. Aghanim et al. *Planck 2018 results. VI. Cosmological parameters. Astron. Astrophys.*, 641:A6, 2020. [Erratum: *Astron. Astrophys.* 652, C4 (2021)]. [arXiv:1807.06209](#), [doi:10.1051/0004-6361/201833910](#).
- [89] J. P. Lees et al. *Measurement of an Excess of $\bar{B} \rightarrow D^{(*)} \tau^- \bar{\nu}_\tau$ Decays and Implications for Charged Higgs Bosons. Phys. Rev. D*, 88(7):072012, 2013. [arXiv:1303.0571](#), [doi:10.1103/PhysRevD.88.072012](#).
- [90] G. Caria et al. *Measurement of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ with a semileptonic tagging method. Phys. Rev. Lett.*, 124(16):161803, 2020. [arXiv:1910.05864](#), [doi:10.1103/PhysRevLett.124.161803](#).
- [91] R. Aaij et al. *Test of Lepton Flavor Universality by the measurement of the $B^0 \rightarrow D^{*-} \tau^+ \nu_\tau$ branching fraction using three-prong τ decays. Phys. Rev. D*, 97(7):072013, 2018. [arXiv:1711.02505](#), [doi:10.1103/PhysRevD.97.072013](#).
- [92] Matthew D. Schwartz. *TASI Lectures on Collider Physics*, pages 65–100. 2018. [arXiv:1709.04533](#), [doi:10.1142/9789813233348_0002](#).
- [93] Jon Butterworth et al. *PDF4LHC recommendations for LHC Run II. J. Phys. G*, 43:023001, 2016. [arXiv:1510.03865](#), [doi:10.1088/0954-3899/43/2/023001](#).
- [94] John C. Collins, Davison E. Soper, and George F. Sterman. *Factorization for Short Distance Hadron - Hadron Scattering. Nucl. Phys. B*, 261:104–142, 1985. [doi:10.1016/0550-3213\(85\)90565-6](#).
- [95] Fabio Maltoni, Giovanni Ridolfi, and Maria Ubiali. *b-initiated processes at the LHC: a reappraisal. JHEP*, 07:022, 2012. [Erratum: *JHEP* 04, 095 (2013)]. [arXiv:1203.6393](#), [doi:10.1007/JHEP04\(2013\)095](#).
- [96] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. *An introduction to PYTHIA 8.2. Computer Physics Communications*, 191:159–177, Jun 2015. URL: <http://dx.doi.org/10.1016/j.cpc.2015.01.024>, [doi:10.1016/j.cpc.2015.01.024](#).
- [97] Paolo Nason. *A New Method for Combining NLO QCD with Shower Monte Carlo Algorithms. Journal of High Energy Physics*, 2004(11):040–040, Nov 2004. URL: <http://dx.doi.org/10.1088/1126-6708/2004/11/040>, [doi:10.1088/1126-6708/2004/11/040](#).
- [98] Stefano Frixione, Paolo Nason, and Carlo Oleari. *Matching NLO QCD computations with parton shower simulations: the POWHEG method. Journal of High Energy Physics*, 2007(11):070–070, Nov 2007. URL: <http://dx.doi.org/10.1088/1126-6708/2007/11/070>, [doi:10.1088/1126-6708/2007/11/070](#).
- [99] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. Journal of High Energy Physics*, 2010(6), Jun 2010. URL: [http://dx.doi.org/10.1007/JHEP06\(2010\)043](http://dx.doi.org/10.1007/JHEP06(2010)043), [doi:10.1007/jhep06\(2010\)043](#).
- [100] M. Bahr et al. *Herwig++ Physics and Manual. Eur. Phys. J. C*, 58:639–707, 2008. [arXiv:0803.0883](#), [doi:10.1140/epjc/s10052-008-0798-9](#).
- [101] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. JHEP*, 07:079, 2014. [arXiv:1405.0301](#), [doi:10.1007/JHEP07\(2014\)079](#).
- [102] S. Agostinelli et al. *GEANT4—a simulation toolkit. Nucl. Instrum. Meth. A*, 506:250–303, 2003. [doi:10.1016/S0168-9002\(03\)01368-8](#).
- [103] Richard D. Ball et al. *Parton distributions from high-precision collider data. Eur. Phys. J. C*, 77:663, 2017. [arXiv:1706.00428](#), [doi:10.1140/epjc/s10052-017-5199-5](#).

- [104] Torbjörn Sjöstrand. *Jet fragmentation of multiparton configurations in a string framework*. *Nuclear Physics B*, 248(2):469–502, December 1984. doi:10.1016/0550-3213(84)90607-2.
- [105] Bo Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand. *Parton Fragmentation and String Dynamics*. *Phys. Rept.*, 97:31–145, 1983. doi:10.1016/0370-1573(83)90080-7.
- [106] Albert M Sirunyan et al. *Evidence for Higgs boson decay to a pair of muons*. *JHEP*, 01:148, 2021. arXiv:2009.04363, doi:10.1007/JHEP01(2021)148.
- [107] Albert M Sirunyan et al. *A measurement of the Higgs boson mass in the diphoton decay channel*. *Physics Letters B*, 805:135425, 2020. URL: <http://www.sciencedirect.com/science/article/pii/S037026932030229X>, doi:<https://doi.org/10.1016/j.physletb.2020.135425>.
- [108] Serguei Chatrchyan et al. *Observation of a New Boson with Mass Near 125 GeV in pp Collisions at $\sqrt{s} = 7$ and 8 TeV*. *JHEP*, 06:081, 2013. arXiv:1303.4571, doi:10.1007/JHEP06(2013)081.
- [109] Georges Aad et al. *Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector*. *Eur. Phys. J. C*, 75(10):476, 2015. [Erratum: *Eur.Phys.J.C* 76, 152 (2016)]. arXiv:1506.05669, doi:10.1140/epjc/s10052-015-3685-1.
- [110] Vardan Khachatryan et al. *Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV*. *Phys. Rev. D*, 92(1):012004, 2015. arXiv:1411.3441, doi:10.1103/PhysRevD.92.012004.
- [111] Georges Aad et al. *Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments*. *Phys. Rev. Lett.*, 114:191803, 2015. arXiv:1503.07589, doi:10.1103/PhysRevLett.114.191803.
- [112] Vardan Khachatryan et al. *Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV*. *Eur. Phys. J. C*, 75(5):212, 2015. arXiv:1412.8662, doi:10.1140/epjc/s10052-015-3351-7.
- [113] Johannes Mellenthin. *Observation of the $t\bar{t}H$ production*. *J. Phys. Conf. Ser.*, 1390(1):012042, 2019. doi:10.1088/1742-6596/1390/1/012042.
- [114] Albert M Sirunyan et al. *Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector*. *Phys. Lett. B*, 779:283–316, 2018. arXiv:1708.00373, doi:10.1016/j.physletb.2018.02.004.
- [115] M. Cepeda et al. *Report from Working Group 2: Higgs Physics at the HL-LHC and HE-LHC*. *CERN Yellow Rep. Monogr.*, 7:221–584, 2019. arXiv:1902.00134, doi:10.23731/CYRM-2019-007.221.
- [116] LHCTopWG. *LHCTopWG Summary Plots*. URL: <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCTopWGSummaryPlots>.
- [117] Marcel Rieger. *Search for Higgs boson production in association with top quarks and decaying into bottom quarks using deep learning techniques with the CMS experiment*. Dissertation, RWTH Aachen University, Aachen, 2019. Veröffentlicht auf dem Publikationsserver der RWTH Aachen University; Dissertation, RWTH Aachen University, 2019. URL: <https://publications.rwth-aachen.de/record/763526>, doi:10.18154/RWTH-2019-06415.
- [118] Fabio Cascioli, Philipp Maierhöfer, Niccolò Moretti, Stefano Pozzorini, and Frank Siegert. *NLO matching for $t\bar{t}b\bar{b}$ production with massive b -quarks*. *Phys. Lett. B*, 734:210–214, 2014. arXiv:1309.5912, doi:10.1016/j.physletb.2014.05.040.
- [119] Albert M Sirunyan et al. *Measurement of the cross section for $t\bar{t}$ production with additional jets and b jets in pp collisions at $\sqrt{s} = 13$ TeV*. *JHEP*, 07:125, 2020. arXiv:2003.06467, doi:10.1007/JHEP07(2020)125.

- [120] Albert M Sirunyan et al. *Measurement of the $t\bar{t}b\bar{b}$ production cross section in the all-jet final state in pp collisions at $\sqrt{s} = 13$ TeV.* *Phys. Lett. B*, 803:135285, 2020. [arXiv:1909.05306](#), doi: 10.1016/j.physletb.2020.135285.
- [121] Albert M Sirunyan et al. *Search for $t\bar{t}H$ production in the $H \rightarrow b\bar{b}$ decay channel with leptonic $t\bar{t}$ decays in proton-proton collisions at $\sqrt{s} = 13$ TeV.* *JHEP*, 03:026, 2019. [arXiv:1804.03682](#), doi:10.1007/JHEP03(2019)026.
- [122] Georges Aad et al. *Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector.* *Eur. Phys. J. C*, 75(7):349, 2015. [arXiv:1503.05066](#), doi:10.1140/epjc/s10052-015-3543-1.
- [123] Morad Aaboud et al. *Search for the standard model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector.* *Phys. Rev. D*, 97(7):072016, 2018. [arXiv:1712.08895](#), doi:10.1103/PhysRevD.97.072016.
- [124] Georges Aad et al. *Measurement of Higgs boson decay into b -quarks in associated production with a top-quark pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector.* *JHEP*, 06:097, 2022. [arXiv:2111.06712](#), doi:10.1007/JHEP06(2022)097.
- [125] Vardan Khachatryan et al. *Search for the associated production of the Higgs boson with a top-quark pair.* *JHEP*, 09:087, 2014. [Erratum: *JHEP* 10, 106 (2014)]. [arXiv:1408.1682](#), doi: 10.1007/JHEP09(2014)087.
- [126] Vardan Khachatryan et al. *Search for a Standard Model Higgs Boson Produced in Association with a Top-Quark Pair and Decaying to Bottom Quarks Using a Matrix Element Method.* *Eur. Phys. J. C*, 75(6):251, 2015. [arXiv:1502.02485](#), doi:10.1140/epjc/s10052-015-3454-1.
- [127] *Measurement of $t\bar{t}H$ production in the $H \rightarrow b\bar{b}$ decay channel in 41.5 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV.* Technical Report CMS-PAS-HIG-18-030, CERN, Geneva, 2019. URL: <https://cds.cern.ch/record/2675023>.
- [128] *Sensitivity projections for Higgs boson properties measurements at the HL-LHC.* 2018.
- [129] The CMS Collaboration. *Expected performance of the physics objects with the upgraded CMS detector at the HL-LHC.* Technical report, CERN, Geneva, 2018. URL: <https://cds.cern.ch/record/2650976>.
- [130] *LHC Machine.* volume 3, page S08001. 2008. doi:10.1088/1748-0221/3/08/S08001.
- [131] CMS Collaboration. *The CMS Experiment at the CERN LHC.* *JINST*, 3:S08004, 2008. doi: 10.1088/1748-0221/3/08/S08004.
- [132] *LEP design report.* CERN, Geneva, 1984. URL: <https://cds.cern.ch/record/102083>.
- [133] Thomas Taylor and Daniel Treille. *The Large Electron Positron Collider (LEP): Probing the Standard Model.* *Adv. Ser. Direct. High Energy Phys.*, 27:217–261, 2017. doi:10.1142/9789814749145_0007.
- [134] Esma Mobs. *The CERN accelerator complex - August 2018.* *Complexe des accélérateurs du CERN - Aout 2018.* 2018. General Photo. URL: <https://cds.cern.ch/record/2636343>.
- [135] ATLAS Collaboration. *The ATLAS Experiment at the CERN Large Hadron Collider.* *JINST*, 3:S08003, 2008. doi:{10.1088/1748-0221/3/08/S08003}.
- [136] ALICE Collaboration. *The ALICE experiment at the CERN LHC.* *JINST*, 3:S08002, 2008. doi: 10.1088/1748-0221/3/08/S08002.
- [137] LHCb Collaboration. *The LHCb Detector at the LHC.* *JINST*, 3:S08005, 2008. doi:10.1088/1748-0221/3/08/S08005.

-
- [138] CERN Visual Media Office, M. Viguier, and P. Catapano. *CERN News - Physics restarts in the LHC at new record energy. CERN News - La physique reprend au LHC avec un nouveau record en energie.* Mar 2012. URL: <https://cds.cern.ch/record/1435792>.
- [139] CMS Collaboarton. *Public CMS Luminosity Information*, 2022. URL: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- [140] *The HiLumi LHC Design Study*, 2022. URL: <https://hilumilhcds.web.cern.ch/>.
- [141] G. L. Bayatian et al. *CMS physics: Technical design report*. 2006.
- [142] CMS Collaboration. *CMS detector design*. 2011. URL: <http://cms.web.cern.ch/news/cms-detector-design>.
- [143] CMS Collaboration. *The CMS tracker system project: Technical Design Report*. CERN, Geneva, 1997. URL: <https://cds.cern.ch/record/368412>.
- [144] *The CMS tracker: addendum to the Technical Design Report*. 2000.
- [145] Serguei Chatrchyan et al. *Description and performance of track and primary-vertex reconstruction with the CMS tracker*. *JINST*, 9(10):P10009, 2014. arXiv:1405.6569, doi:10.1088/1748-0221/9/10/P10009.
- [146] W. Adam et al. *The CMS Phase-1 Pixel Detector Upgrade*. *JINST*, 16(02):P02027, 2021. arXiv:2012.14304, doi:10.1088/1748-0221/16/02/P02027.
- [147] CMS Collaboration. *CMS: The electromagnetic calorimeter. Technical design report*. 1997. ISBN: 92908311221. URL: <https://cds.cern.ch/record/349375>.
- [148] CMS Collaboration. *The CMS hadron calorimeter project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997. ISBN: 9789290831099. URL: <https://cds.cern.ch/record/357153>.
- [149] *The CMS electromagnetic calorimeter project: Technical Design Report*. 1997.
- [150] *Addendum to the CMS ECAL technical design report: Changes to the CMS ECAL electronics*. 9 2002.
- [151] A. Benaglia. *The CMS ECAL performance with examples*. *JINST*, 9:C02008, 2014. doi:10.1088/1748-0221/9/02/C02008.
- [152] *ECAL 2016 refined calibration and Run2 summary plots*. 2020. URL: <https://cds.cern.ch/record/2717925>.
- [153] *The CMS hadron calorimeter project: Technical Design Report*. 1997.
- [154] S Chatrchyan et al. *Performance of the CMS Hadron Calorimeter with Cosmic Ray Muons and LHC Beam Data*. *JINST*, 5:T03012, 2010. arXiv:0911.4991, doi:10.1088/1748-0221/5/03/T03012.
- [155] *HCAL Energy Reconstruction Performance*. 2016. URL: <https://cds.cern.ch/record/2235509>.
- [156] *HCAL Calibration Status in Summer 2017*. 2017. URL: <https://cds.cern.ch/record/2281146>.
- [157] *Results related to the Phase1 HE upgrade*. 2018. URL: <https://cds.cern.ch/record/2320857>.
- [158] Dario Mapelli. *The CMS ECAL data acquisition system and its performance at LHC Run 2. PoS, LHCP2018:069*, 2018. doi:10.22323/1.321.0069.

- [159] Albert M Sirunyan et al. *Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC*. *JINST*, 16(05):P05014, 2021. arXiv:2012.06888, doi:10.1088/1748-0221/16/05/P05014.
- [160] CMS HCAL depth segmentation. URL: <https://home.fnal.gov/~chlebana/CMS/Phase1/depthSegmentation.html>.
- [161] Vardan Khachatryan et al. *Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV*. *JINST*, 12(02):P02014, 2017. arXiv:1607.03663, doi:10.1088/1748-0221/12/02/P02014.
- [162] CMS Collaboration. *The CMS magnet project: Technical design report*. 1997. ISBN: 9789290831013. URL: <http://cds.cern.ch/record/331056>.
- [163] *The CMS muon project: Technical Design Report*. 1997.
- [164] CMS Collaboration. *The CMS muon project : Technical Design Report*. 1997. ISBN: 9789290831211. URL: <http://cds.cern.ch/record/343814>.
- [165] Serguei Chatrchyan et al. *Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV*. *JINST*, 7:P10002, 2012. arXiv:1206.4071, doi:10.1088/1748-0221/7/10/P10002.
- [166] A. M. Sirunyan et al. *Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV*. *JINST*, 13(06):P06015, 2018. arXiv:1804.04528, doi:10.1088/1748-0221/13/06/P06015.
- [167] D. Abbaneo et al. *A GEM Detector System for an Upgrade of the High-eta Muon Endcap Stations GE1/1 + ME1/1 in CMS*. 2012. arXiv:1211.1494.
- [168] J. R. Cudell, V. V. Ezhela, P. Gauron, K. Kang K., Kuyanov V. Yu, S. B. Lugovsky, E. Martynov, B. Nicolescu, E. A. Razuvaev, and N. P. Tkachenko. *Benchmarks for the forward observables at RHIC, the Tevatron Run II and the LHC*. *Phys. Rev. Lett.*, 89:201801, 2002. arXiv:hep-ph/0206172, doi:10.1103/PhysRevLett.89.201801.
- [169] A. Tapper. *The CMS Level-1 Trigger for LHC Run II*. Technical Report CMS-CR-2016-303, CERN, Geneva, Oct 2016. ISBN: 9789290833901. URL: <https://cds.cern.ch/record/1556311>.
- [170] S. Dasu et al. *CMS TriDAS project: Technical design report, vol.1: The trigger systems*. 2000. ISBN: 9290831102. URL: <http://cds.cern.ch/record/706847>.
- [171] CMS: *The TriDAS project. Technical design report, Vol. 2: Data acquisition and high-level trigger*. 12 2002.
- [172] V. Khachatryan et al. *The CMS trigger system*. *JINST*, 12(01):P01020, 2017. arXiv:1609.02366, doi:10.1088/1748-0221/12/01/P01020.
- [173] L. Cadamuro. *The CMS Level-1 trigger system for LHC Run II*. *Journal of Instrumentation*, 12(03):C03021, 2017. doi:10.1088/1748-0221/12/03/C03021.
- [174] A Tapper and D. Acosta. *CMS Technical Design Report for the Level-1 Trigger Upgrade*. 2013. ISBN: 9789290833901. URL: <https://cds.cern.ch/record/1556311>.
- [175] W. Adam et al. *The CMS high level trigger*. *Eur. Phys. J.*, C46:605–667, 2006. arXiv:hep-ex/0512077, doi:10.1140/epjc/s2006-02495-8.
- [176] Albert M Sirunyan et al. *Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV*. *JINST*, 15(10):P10017, 2020. arXiv:2006.10165, doi:10.1088/1748-0221/15/10/P10017.
- [177] Silvio Donato. *CMS Trigger Performance*. *EPJ Web Conf.*, 182:02037, 2018. doi:10.1051/epjconf/201818202037.

- [178] *Worldwide LHC Computing Grid*, 2018. URL: <http://wlcg.web.cern.ch/>.
- [179] C. Eck et al. *LHC computing Grid: Technical Design Report. Version 1.06 (20 Jun 2005)*. Technical Design Report LCG. CERN, Geneva, 2005. ISBN: 9789290832539. URL: <https://cds.cern.ch/record/840543>.
- [180] CMS Collaboration. *CMS computing: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 2005. URL: <https://cds.cern.ch/record/838359>.
- [181] Giovanni Petrucciani, Andrea Rizzi, and Carl Vuosalo. *Mini-AOD: A New Analysis Data Format for CMS*. *J. Phys. Conf. Ser.*, 664(7):7, 2015. arXiv:1702.04685, doi:10.1088/1742-6596/664/7/072052.
- [182] Andrea Rizzi, Giovanni Petrucciani, and Marco Peruzzi. *A further reduction in CMS event data for analysis: the NANO AOD format*. *EPJ Web Conf.*, 214:06021. 6 p, 2019. URL: <https://cds.cern.ch/record/2699585>, doi:10.1051/epjconf/201921406021.
- [183] Siona Ruth Davis. *Interactive Slice of the CMS detector*. Aug 2016. URL: <https://cds.cern.ch/record/2205172>.
- [184] J. R. Cash and Alan H. Karp. *A Variable Order Runge-Kutta Method for Initial Value Problems with Rapidly Varying Right-hand Sides*. *ACM Trans. Math. Softw.*, 16(3):201–222, September 1990. doi:10.1145/79505.79507.
- [185] *Application of Kalman filtering to track and vertex fitting*.
- [186] Serguei Chatrchyan et al. *Alignment of the CMS tracker with LHC and cosmic ray data*. *JINST*, 9:P06009, 2014. arXiv:1403.2286, doi:10.1088/1748-0221/9/06/P06009.
- [187] Gregor Mittag and on behalf of the CMS Collaboration. *Alignment of the CMS Tracker: Latest Results from LHC Run-II*. *Journal of Physics: Conference Series*, 898(4):042014, oct 2017. URL: <https://dx.doi.org/10.1088/1742-6596/898/4/042014>, doi:10.1088/1742-6596/898/4/042014.
- [188] Cerati, Giuseppe, Elmer, Peter, Gravelle, Brian, Kortelainen, Matti, Krutelyov, Vyacheslav, Lantz, Steven, Lefebvre, Matthieu, Masciovecchio, Mario, McDermott, Kevin, Norris, Boyana, Reinsvold Hall, Allison, Riley, Daniel, Tadel, Matevz, Wittich, Peter, Würthwein, Frank, and Yagil, Avi. *Parallelized and Vectorized Tracking Using Kalman Filters with CMS Detector Geometry and Events*. *EPJ Web Conf.*, 214:02002, 2019. URL: <https://doi.org/10.1051/epjconf/201921402002>, doi:10.1051/epjconf/201921402002.
- [189] W. Adam, B. Mangano, T. Speer, and T. Todorov. *Track reconstruction in the CMS tracker*. 2005. URL: <http://cds.cern.ch/record/934067>.
- [190] Felice Pantaleo. *New Track Seeding Techniques for the CMS Experiment*, 2017. URL: <https://cds.cern.ch/record/2293435>.
- [191] Kenneth Rose. *Deterministic Annealing for Clustering, Compression, Classification, Regression, and Related Optimization Problems*. 1998. doi:10.1109/5.726788.
- [192] CMS Collaboration. *Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s}=8$ TeV*. *JINST*, 10(06):P06005, 2015. arXiv:1502.02701, doi:10.1088/1748-0221/10/06/P06005.
- [193] CMS Collaboration. *Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector*. Technical Report CMS-PAS-PFT-10-001, 2010. URL: <https://cds.cern.ch/record/1247373>.
- [194] CMS Collaboration. *Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET*. Technical Report CMS-PAS-PFT-09-001, CERN, Geneva, Apr 2009. URL: <https://cds.cern.ch/record/1194487>.

- [195] Milos Dordevic. *The CMS Particle Flow Algorithm*. *EPJ Web Conf.*, 191:02016, 2018. doi: 10.1051/epjconf/201819102016.
- [196] Christian Lippmann. *Particle identification*. *Nucl. Instrum. Meth. A*, 666:148–172, 2012. arXiv: 1101.3276, doi:10.1016/j.nima.2011.03.009.
- [197] A. M. Sirunyan et al. *Particle-flow reconstruction and global event description with the CMS detector*. *JINST*, 12(10):P10003, 2017. arXiv:1706.04965, doi:10.1088/1748-0221/12/10/P10003.
- [198] *PF concept in CMS*. URL: <http://bartosik.pp.ua/>.
- [199] S. Baffioni, C. Charlot, F. Ferri, D. Futyan, P. Meridiani, I. Puljak, C. Rovelli, R. Salerno, and Y. Sirois. *Electron reconstruction in CMS*. *Eur. Phys. J.*, C49:1099–1116, 2007. doi:10.1140/epjc/s10052-006-0175-5.
- [200] Andreas Hocker et al. *TMVA - Toolkit for Multivariate Data Analysis*. *PoS*, ACAT:040, 2007. arXiv:physics/0703039.
- [201] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. *The Anti- k_t jet clustering algorithm*. *JHEP*, 04:063, 2008. arXiv:0802.1189, doi:10.1088/1126-6708/2008/04/063.
- [202] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. *FastJet User Manual*. *Eur. Phys. J.*, C72:1896, 2012. arXiv:1111.6097, doi:10.1140/epjc/s10052-012-1896-2.
- [203] Stephen D. Ellis and Davison E. Soper. *Successive combination jet algorithm for hadron collisions*. *Phys. Rev. D*, 48:3160–3166, Oct 1993. doi:10.1103/PhysRevD.48.3160.
- [204] Yuri L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber. *Better jet clustering algorithms*. *JHEP*, 08:001, 1997. arXiv:hep-ph/9707323, doi:10.1088/1126-6708/1997/08/001.
- [205] M. Wobisch and T. Wengler. *Hadronization corrections to jet cross-sections in deep inelastic scattering*. In *Monte Carlo generators for HERA physics. Proceedings, Workshop, Hamburg, Germany, 1998-1999*, pages 270–279, 1998. arXiv:hep-ph/9907280.
- [206] *Jet algorithms performance in 13 TeV data*. Technical report, CERN, Geneva, 2017. URL: <https://cds.cern.ch/record/2256875>.
- [207] CMS Collaboration. *Performance of missing energy reconstruction in 13 TeV pp collision data using the CMS detector*. Technical Report CMS-PAS-JME-16-004, CERN, Geneva, 2016. URL: <http://cds.cern.ch/record/2205284>.
- [208] *Pileup-per-particle identification: optimisation for Run 2 Legacy and beyond*. Jan 2021. URL: <https://cds.cern.ch/record/2751563>.
- [209] *How CMS weeds out particles that pile up*. URL: <https://cms.cern/news/how-cms-weeds-out-particles-pile>.
- [210] Albert M Sirunyan et al. *Identification of heavy-flavor jets with the CMS detector in pp collisions at 13 TeV*. Submitted to: *JINST*, 2017. arXiv:1712.07158.
- [211] Stephen Poprocki. *Search for WZ+ZZ Production with Missing Transverse Energy and b Jets at CDF*. PhD thesis, Cornell U., 2013. URL: <http://lss.fnal.gov/archive/thesis/2000/fermilab-thesis-2013-01.pdf>.
- [212] Emil Bols, Jan Kieseler, Mauro Verzetti, Markus Stoye, and Anna Stakia. *Jet flavor Classification Using DeepJet*. *JINST*, 15(12):P12012, 2020. arXiv:2008.10519, doi:10.1088/1748-0221/15/12/P12012.
- [213] *Performance of the DeepJet b tagging algorithm using 41.9/fb of data from proton-proton collisions at 13 TeV with Phase 1 CMS detector*. CMS Detector Performance Summary CMS-DP-2018-058, CERN, 2018. URL: <http://cds.cern.ch/record/2646773>.

- [214] Michelangelo Mangano. *Motivations and precision targets for an accurate luminosity determination at the LHC. Motivations and precision targets for luminosity*. 2011. URL: <https://cds.cern.ch/record/1357612>.
- [215] G.C. Fox and S. Wolfram. *Event shapes in e^+e^- annihilation*. *Nuclear Physics B*, 157(3):543–544, 1979. URL: <http://www.sciencedirect.com/science/article/pii/0550321379901202>, doi:10.1016/0550-3213(79)90120-2.
- [216] P Grafstrom and W Kozanecki. *Luminosity determination at proton colliders*. *Prog. Part. Nucl. Phys.*, 81:97–148. 52 p, 2015. URL: <https://cds.cern.ch/record/2153734>, doi:10.1016/j.ppnp.2014.11.002.
- [217] CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update. Technical report, CERN, Geneva, 2013. URL: <https://cds.cern.ch/record/1598864>.
- [218] Roel Aaij et al. *Precision luminosity measurements at LHCb*. *JINST*, 9(12):P12005, 2014. arXiv:1410.0149, doi:10.1088/1748-0221/9/12/P12005.
- [219] Georges Aad et al. *Improved luminosity determination in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector at the LHC*. *Eur. Phys. J.*, C73(8):2518, 2013. arXiv:1302.4393, doi:10.1140/epjc/s10052-013-2518-3.
- [220] ALICE luminosity determination for pp collisions at $\sqrt{s} = 8$ TeV. Mar 2017. URL: <https://cds.cern.ch/record/2255216>.
- [221] Paul Lujan. *The Pixel Luminosity Telescope: A detector for luminosity measurement at CMS using silicon pixel sensors*. 6 2022. arXiv:2206.08870.
- [222] C Barschel, M Ferro-Luzzi, J-J Gras, M Ludwig, P Odier, and S Thoulet. *Results of the LHC DCCT Calibration Studies*. Feb 2012. URL: <https://cds.cern.ch/record/1425904>.
- [223] G Anders, N Bacchetta, V Balagura, C Barschel, D Belohrad, H Burkhardt, M Ferro-Luzzi, C Gabaldon, M Gagliardi, J J Gras, P Hopchev, A Jeff, W Kozanecki, M Ludwig, D Marlow, K Oyama, J Panman, S White, and A Zuranski. *STUDY OF THE RELATIVE LHC BUNCH POPULATIONS FOR LUMINOSITY CALIBRATION*. Feb 2012. URL: <https://cds.cern.ch/record/1427726>.
- [224] A Alici, G Anders, N Bacchetta, V Balagura, C Barschel, E Bravin, H Burkhardt, M Ferro-Luzzi, M Gagliardi, J J Gras, B Heinemann, P Hopchev, M Huhtinen, A Jeff, W Kozanecki, S Pagan Griso, J Panman, S White, and A Zuranski. *Study of the LHC ghost charge and satellite bunches for luminosity calibration*. Feb 2012. URL: <https://cds.cern.ch/record/1427728>.
- [225] Markus Klute, Catherine Medlock, and Jakob Salfeld-Nebgen. *Beam Imaging and Luminosity Calibration*. *JINST*, 12(03):P03018, 2017. arXiv:1603.03566, doi:10.1088/1748-0221/12/03/P03018.
- [226] J. Knolle (CMS Collaboration). *The factorization bias in the Van der Meer method: Run 2 experiences at the CMS experiment*. in proceeding of lhc lumi days 2019, 1910.14066, 2019. CMS-PAS-LUM-18-002 <https://cds.cern.ch/record/2676164>. URL: <https://cds.cern.ch/record/2676164>, arXiv:1910.14066.
- [227] Vladislav Balagura. *Van der Meer scan luminosity measurement and beam-beam correction*. *Eur. Phys. J. C*, 81(1):26, 2021. arXiv:2012.07752, doi:10.1140/epjc/s10052-021-08837-y.
- [228] CMS Luminosity Calibration for the pp Reference Run at $\sqrt{s} = 5.02$ TeV. Technical report, CERN, Geneva, 2016. URL: <https://cds.cern.ch/record/2235781>.
- [229] *Luminosity measurement in proton-proton collisions at 5.02 TeV in 2017 at CMS*. Technical report, CERN, Geneva, 2021. URL: <https://cds.cern.ch/record/2765655>.
- [230] *CMS luminosity measurement using nucleus-nucleus collisions at $\sqrt{s_{NN}} = 5.02$ TeV in 2018*. Technical report, CERN, Geneva, 2022. URL: <https://cds.cern.ch/record/2809613>.

- [231] CMS luminosity measurement using 2016 proton-nucleus collisions at nucleon-nucleon center-of-mass energy of 8.16 TeV. Technical report, CERN, Geneva, 2018. URL: <https://cds.cern.ch/record/2628652>.
- [232] CMS Collaboration. *Long-term monitoring of delivered luminosity and calibration stability in CMS*. CMS Detector Performance Note CMS-DP-2019-016, CERN, Geneva, Jun 2019. URL: <https://cds.cern.ch/record/2681803>.
- [233] Wouter Verkerke and David P. Kirkby. *The RooFit toolkit for data modeling*. eConf, C0303241:MOLT007, 2003. arXiv:physics/0306116.
- [234] Joscha Knolle. *Measuring luminosity and the $t\bar{t}Z$ production cross section with the CMS experiment*, Sep 2020. Presented 25 Jun 2020. URL: <https://cds.cern.ch/record/2744549>.
- [235] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. *Tensorflow: A system for large-scale machine learning*. In *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, pages 265–283, 2016.
- [236] Francois Chollet et al. *Keras*, 2015. URL: <https://github.com/fchollet/keras>.
- [237] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. *Asymptotic formulae for likelihood-based tests of new physics*. *Eur. Phys. J. C*, 71:1554, 2011. [Erratum: *Eur.Phys.J.C* 73, 2501 (2013)]. arXiv:1007.1727, doi:10.1140/epjc/s10052-011-1554-0.
- [238] The ATLAS and CMS Collaborations and The LHC Higgs Combination Group. *Procedure for the LHC Higgs boson search combination in Summer 2011*. Technical Report CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, CERN, Geneva, Aug 2011. URL: <https://cds.cern.ch/record/1379837>.
- [239] J. S. Conway. *Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra*, 2011. arXiv:1103.0354.
- [240] Albert M Sirunyan et al. *Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements*. *Eur. Phys. J. C*, 80(1):4, 2020. arXiv:1903.12179, doi:10.1140/epjc/s10052-019-7499-4.
- [241] Vardan Khachatryan et al. *Event generator tunes obtained from underlying event and multiparton scattering measurements*. *Eur. Phys. J. C*, 76:155, 2016. arXiv:1512.00815, doi:10.1140/epjc/s10052-016-3988-x.
- [242] H. B. Hartanto, B. Jäger, L. Reina, and D. Wackeroth. *Higgs boson production in association with top quarks in the POWHEG BOX*. *Phys. Rev. D*, 91:094003, May 2015. URL: <https://link.aps.org/doi/10.1103/PhysRevD.91.094003>, doi:10.1103/PhysRevD.91.094003.
- [243] LHC Higgs Cross Section Working Group. *SM Higgs production cross sections at $\sqrt{s} = 13$ TeV*. TWiki: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt13TeV_r23.
- [244] *SM Higgs production cross sections at $\sqrt{s} = 13$ TeV*. URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt13TeV#tH_Process.
- [245] Jet flavor identification (mc truth). Twiki: https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideBTagMCTools_r41.
- [246] N. Bartosik, J. Hauk, and A. Geiser. *Identification of the origin of heavy flavor jets at generator level in monte carlo simulations at cms*. CMS Analysis Note AN-14-093, CERN, 2014. URL: http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2014/093.

-
- [247] Nazar 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 $t\bar{t}$ +Higgs Production, 2015. Presented 03 Jul 2015. URL: <https://cds.cern.ch/record/2047049>.
 - [248] Maximilian Maria Horzela. *Merging of $t\bar{t}$ and $t\bar{t} + b\bar{t}b$ simulations for an improved background modeling for $t\bar{t}H(b\bar{t}b)$ measurements*. Master's thesis, Karlsruhe Institute of Technology (KIT), 2019.
 - [249] Federico Demartin, Fabio Maltoni, Kentarou Mawatari, and Marco Zaro. *Higgs production in association with a single top quark at the LHC*. *Eur. Phys. J.*, C75:267, 2015. [arXiv:1504.00611](#), doi:10.1140/epjc/s10052-015-3475-9.
 - [250] Federico Demartin, Benedikt Maier, Fabio Maltoni, Kentarou Mawatari, and Marco Zaro. *tWH associated production at the LHC*. *Eur. Phys. J.*, C77(1):34, 2017. [arXiv:1607.05862](#), doi:10.1140/epjc/s10052-017-4601-7.
 - [251] *NLO single-top channel cross sections ATLAS-CMS recommended predictions for single-top cross sections using the Hathor v2.1 program*. URL: TWiki: <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SingleTopRefXsec>.
 - [252] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. "NLO single-top production matched with shower in POWHEG: s- and t-channel contributions". *JHEP*, 09:111, 2009. [arXiv:0907.4076](#), doi:10.1088/1126-6708/2009/09/111.
 - [253] Emanuele Re. "Single-top Wt-channel production matched with parton showers using the POWHEG method". *Eur. Phys. J. C*, 71:1547, 2011. [arXiv:1009.2450](#), doi:10.1140/epjc/s10052-011-1547-z.
 - [254] N. Kidonakis. "two-loop soft anomalous dimensions for single top quark associated production with w^- or h^- ". *Phys. Rev. D*, 82:054018, 2010. [arXiv:1005.4451](#), doi:10.1103/PhysRevD.82.054018.
 - [255] M. Aliev, H. Lacker, U. Langenfeld, S. Moch, P. Uwer, and M. Wiedermann. "hathor: hadronic top and heavy quarks cross section calculator". *Comput. Phys. Commun.*, 182:1034, 2011. [arXiv:1007.1327](#), doi:10.1016/j.cpc.2010.12.040.
 - [256] P. Kant, O. M. Kind, T. Kintscher, T. Lohse, T. Martini, S. Mölbitz, P. Rieck, and P. Uwer. "hathor for single top-quark production: updated predictions and uncertainty estimates for single top-quark production in hadronic collisions".
 - [257] Kulesza, Anna and Motyka, Leszek and Schwartländer, Daniel and Stebel, Tomasz and Theeuwes, Vincent. *Associated production of a top quark pair with a heavy electroweak gauge boson at NLO+NNLL accuracy*. *Eur. Phys. J.*, C79:249, 2019. [arXiv:1812.08622](#), doi:10.1140/epjc/s10052-019-6746-z.
 - [258] Rikkert Frederix and Stefano Frixione. *Merging meets matching in MC@NLO*. *JHEP*, 12:061, 2012. [arXiv:1209.6215](#), doi:10.1007/JHEP12(2012)061.
 - [259] Gehrmann, T. and Grazzini, M. and Kallweit, S. and Maierhöfer, P. and von Manteuffel, A. and Pozzorini, S. and Rathlev, D. and Tancredi, L. *W^+W^- Production at Hadron Colliders in Next to Next to Leading Order QCD*. *Phys. Rev. Lett.*, 113:212001, 2014. [arXiv:1408.5243](#), doi:10.1103/PhysRevLett.113.212001.
 - [260] John M. Campbell, R. Keith Ellis, and Ciaran Williams. *Vector boson pair production at the LHC*. *JHEP*, 07:018, 2011. [arXiv:1105.0020](#), doi:10.1007/JHEP07(2011)018.
 - [261] *NNLO+NNLL top-quark-pair cross sections: ATLAS-CMS recommended predictions for top-quark-pair cross sections using the Top++v2.0 program*. TWiki: <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLOr16>.

- [262] CMS Collaboration. *Summaries of CMS cross section measurements*. URL: TWiki:<https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsCombined>.
- [263] *Jet energy scale and resolution performance with 13 TeV data collected by CMS in 2016-2018*. Apr 2020. URL: <https://cds.cern.ch/record/2715872>.
- [264] *B-tagging performance of the CMS Legacy dataset 2018*. Mar 2021. URL: <https://cds.cern.ch/record/2759970>.
- [265] Marco Zaro. *A fragmentation-based study of heavy quark production*. <https://indico.cern.ch/event/850548/#5-a-fragmentation-based-study>, October 2019.
- [266] A. Kolmogorov. *Sulla determinazione empirica di una legge di distribuzione*. URL: Inst.Ital. Attuari, Giorn. 4 (1933) 83–91.
- [267] N. V. Smirnov. *On the estimation of the discrepancy between empirical curves of distribution for two independent samples*. URL: Bull.Math.Univ.Moscou 2 (1939), no. 2, 3–14.
- [268] Stefan Wunsch, Raphael Friese, Roger Wolf, and Günter Quast. *Identifying the relevant dependencies of the neural network response on characteristics of the input space*. *Comput. Softw. Big Sci.*, 2(1):5, 2018. arXiv:1803.08782, doi:10.1007/s41781-018-0012-1.
- [269] Amirata Ghorbani and James Zou. *Data Shapley: Equitable Valuation of Data for Machine Learning*. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 2242–2251. PMLR, 09–15 Jun 2019. URL: <http://proceedings.mlr.press/v97/ghorbani19c.html>.
- [270] Snoek Jasper, Larochelle Hugo, and P. Adams Ryan. *Practical Bayesian Optimization of Machine Learning Algorithms*, 2012. arXiv:1206.2944.
- [271] Franceschi Luca, Donini Michele, Frasconi Paolo, and Pontil Massimiliano. *Forward and Reverse Gradient-Based Hyperparameter Optimization*, 06–11 Aug 2017. URL: <http://proceedings.mlr.press/v70/franceschi17a.html>.
- [272] Claire Adam-Bourdarios, Glen Cowan, Cécile Germain, Isabelle Guyon, Balázs Kégl, and David Rousseau. *The Higgs Boson Machine Learning Challenge*. In *Proceedings of the 2014 International Conference on High-Energy Physics and Machine Learning - Volume 42, HEPML’14*, pages 19–55. JMLR.org, 2014.
- [273] CMS Collaboration. *CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV*. CMS Physics Analysis Summary CMS-PAS-LUM-17-004, 2018. URL: <http://cms-results.web.cern.ch/cms-results/public-results/preliminary-results/LUM-17-004/index.html>.
- [274] CMS Collaboration. *CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV*. CMS Physics Analysis Summary CMS-PAS-LUM-18-002, 2019. URL: <http://cms-results.web.cern.ch/cms-results/public-results/preliminary-results/LUM-18-002/index.html>.
- [275] Roger J. Barlow and Christine Beeston. *Fitting using finite Monte Carlo samples*. *Comput. Phys. Commun.*, 77:219, 1993. doi:10.1016/0010-4655(93)90005-W.
- [276] *NLO single-top channel cross sections: ATLAS-CMS recommended predictions for single-top cross sections using the Hathor v2.1 program*. TWiki: <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SingleTopRefXsec> r34.
- [277] Albert M Sirunyan et al. *First measurement of the cross section for top quark pair production with additional charm jets using dileptonic final states in pp collisions at $s=13$ TeV*. *Phys. Lett. B*, 820:136565, 2021. arXiv:2012.09225, doi:10.1016/j.physletb.2021.136565.

- [278] Peter Kraemer. *Measurement of $t\bar{t}H$ Simplified Template Cross Sections in the $t\bar{t}H \rightarrow b\bar{b}$ channel at the CMS experiment*. PhD thesis, Feb 2020.
- [279] J. R. Andersen et al. *Les Houches 2017: Physics at TeV Colliders Standard Model Working Group Report*, 3 2018. [arXiv:1803.07977](https://arxiv.org/abs/1803.07977).
- [280] Albert M Sirunyan et al. *Measurements of Higgs boson production cross sections and couplings in the diphoton decay channel at $\sqrt{s} = 13$ TeV*. *JHEP*, 07:027, 2021. [arXiv:2103.06956](https://arxiv.org/abs/2103.06956), doi: 10.1007/JHEP07(2021)027.
- [281] Albert M Sirunyan et al. *Measurement of the inclusive and differential Higgs boson production cross sections in the leptonic WW decay mode at $\sqrt{s} = 13$ TeV*. *JHEP*, 03:003, 2021. [arXiv:2007.01984](https://arxiv.org/abs/2007.01984), doi: 10.1007/JHEP03(2021)003.
- [282] Albert M Sirunyan et al. *Measurements of production cross sections of the Higgs boson in the four-lepton final state in proton–proton collisions at $\sqrt{s} = 13$ TeV*. *Eur. Phys. J. C*, 81(6):488, 2021. [arXiv:2103.04956](https://arxiv.org/abs/2103.04956), doi: 10.1140/epjc/s10052-021-09200-x.
- [283] Ian J. Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, Cambridge, MA, USA, 2016. <http://www.deeplearningbook.org>.
- [284] Günter Klambauer, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter. *Self-Normalizing Neural Networks*. *CoRR*, abs/1706.02515, 2017. URL: <http://arxiv.org/abs/1706.02515>, [arXiv:1706.02515](https://arxiv.org/abs/1706.02515).
- [285] Eric B. Baum and Frank Wilczek. *Supervised Learning of Probability Distributions by Neural Networks*. In *NIPS*, pages 52–61, 1987. URL: <http://papers.nips.cc/paper/3-supervised-learning-of-probability-distributions-by-neural-networks>.
- [286] Elliott Gordon-Rodriguez, Gabriel Loaiza-Ganem, Geoff Pleiss, and John P. Cunningham. *Uses and Abuses of the Cross-Entropy Loss: Case Studies in Modern Deep Learning*, 2020. URL: <https://arxiv.org/abs/2011.05231>, doi: 10.48550/ARXIV.2011.05231.
- [287] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. *Learning Representations by Back-propagating Errors*. *Nature*, 323(6088):533–536, 1986. URL: <http://www.nature.com/articles/323533a0>, doi: 10.1038/323533a0.
- [288] Stuart J. Russell and Peter Norvig. *Artificial Intelligence: a modern approach*. Pearson, 3 edition, 2009.
- [289] John Duchi, Elad Hazan, and Yoram Singer. *Adaptive Subgradient Methods for Online Learning and Stochastic Optimization*. *J. Mach. Learn. Res.*, 12(null):2121–2159, jul 2011.
- [290] John C. Duchi, Elad Hazan, and Yoram Singer. *Adaptive Subgradient Methods for Online Learning and Stochastic Optimization*. *Journal of Machine Learning Research*, 12:2121–2159, 2011. URL: <http://dblp.uni-trier.de/db/journals/jmlr/jmlr12.html#DuchiHS11>.
- [291] Xavier Glorot and Yoshua Bengio. *Understanding the difficulty of training deep feedforward neural networks*. In Yee Whye Teh and Mike Titterton, editors, *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, volume 9 of *Proceedings of Machine Learning Research*, pages 249–256, Chia Laguna Resort, Sardinia, Italy, 13–15 May 2010. PMLR. URL: <https://proceedings.mlr.press/v9/glorot10a.html>.
- [292] Sergey Ioffe and Christian Szegedy. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. *CoRR*, abs/1502.03167, 2015. URL: <http://arxiv.org/abs/1502.03167>, [arXiv:1502.03167](https://arxiv.org/abs/1502.03167).
- [293] Alexander L. Read. *Presentation of search results: The CL(s) technique*. *J. Phys. G*, 28:2693–2704, 2002. doi: 10.1088/0954-3899/28/10/313.
- [294] *The CMS Higgs combine tool developers*. URL: <https://cms-hcomb.gitbooks.io/combine>.

- [295] J. Neyman and E. S. Pearson. *On the Problem of the Most Efficient Tests of Statistical Hypotheses*. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231:289–337, 1933. URL: <http://www.jstor.org/stable/91247>.
- [296] Vardan Khachatryan et al. *Event generator tunes obtained from underlying event and multiparton scattering measurements*. *Eur. Phys. J. C*, 76(3):155, 2016. arXiv:1512.00815, doi:10.1140/epjc/s10052-016-3988-x.
- [297] Federico Buccioni, Stefan Kallweit, Stefano Pozzorini, and Max F. Zoller. *NLO QCD predictions for $t\bar{t}b\bar{b}$ production in association with a light jet at the LHC*. *JHEP*, 12:015, 2019. arXiv:1907.13624, doi:10.1007/JHEP12(2019)015.
- [298] Catherine Bernaciak, Malte Sean Andreas Buschmann, Anja Butter, and Tilman Plehn. *Fox-Wolfram Moments in Higgs Physics*. *Phys. Rev. D*, 87:073014, 2013. arXiv:1212.4436, doi:10.1103/PhysRevD.87.073014.
- [299] Vernon D. Barger, J. Ohnemus, and R. J. N. Phillips. *Event shape criteria for single lepton top signals*. *Phys. Rev. D*, 48:R3953–R3956, 1993. arXiv:hep-ph/9308216, doi:10.1103/PhysRevD.48.R3953.
- [300] CMS Collaboration. *Search for $t\bar{t}H$ production in the $H \rightarrow b\bar{b}$ decay channel with with 2016 pp collision data at $\sqrt{s} = 13$ TeV*. CMS Physics Analysis Summary CMS-PAS-HIG-16-038, CERN, 2016. URL: <http://cms-results.web.cern.ch/cms-results/public-results/preliminary-results/HIG-16-038/index.html>.
- [301] CMS Collaboration. *Search for $t\bar{t}H$ production in the $H \rightarrow b\bar{b}$ decay channel with $\sqrt{s} = 13$ TeV pp collisions at the CMS experiment*. CMS Physics Analysis Summary CMS-PAS-HIG-16-004, CERN, 2016. URL: <http://cms-results.web.cern.ch/cms-results/public-results/preliminary-results/HIG-16-004/index.html>.
- [302] Claire Adam-Bourdarios, Glen Cowan, Cécile Germain, Isabelle Guyon, Balázs Kégl, and David Rousseau. *The Higgs Boson Machine Learning Challenge*. In *Proceedings of the 2014 International Conference on High-Energy Physics and Machine Learning -Volume 42*, HEPML’14, pages 19–55. JMLR.org, 2014.

MACHINE LEARNING AND NEURAL NETWORKS

As high-energy physics research advances, the measurements of processes with small production probability and where background processes dominate the physics process of interest are more and more frequent. Therefore, separating signal and background in these scenarios is crucial, exploiting powerful and cutting-edge techniques such as multivariate analysis (MVA) methods with artificial neural networks (ANN).

As deep learning and machine learning (ML) are constantly evolving research fields, the following appendix focuses on an overview of the fundamental concepts applied in this analysis. First, it introduces mathematical notation and foundations, then moves on to design strategies, training techniques, and the underlying minimization processes before concluding with training optimization and model generalization.

A.1 Mathematical construction

A neural network (NN) is a mapping assigning, based on a mathematical model, a collection of input features $\vec{x}$ to a set of output $\vec{y}$, the so-called prediction [283]. Typically, both $\vec{x}$ and $\vec{y}$ can have arbitrary dimensions, and the model is a composition of basic and complex mathematical operations with many free parameters to characterize the complexity of the input space. In the most fundamental type of neural network, the feed-forward neural network, the aim is to approximate an arbitrary, known or unknown, function by the mapping function $\vec{y} = f(\vec{x}; \theta)$, with θ describing the free parameters. These parameters are determined in the training process, an optimization technique that identifies the set producing the best approximation of the function of interest with an iterative approach. It is carried out based on the minimization of an objective function that compares predicted and true outputs, commonly referred to as loss function. The shape and behavior of this loss function are highly dependent on the specific task and application.

In the ANNs, the neurons represent the basic building blocks for architectures with more complexity. A schematic representation is in Figure A.1 with neurons stacked in so-called layers: an input layer ($l = 0$), $L - 1$ hidden layers, and the final output layer $l = L$. Each layer comprises a configurable number of units or nodes that receive inputs from all units in the previous layer, transform them into a single output value through mathematical operations, and then propagate it to all units in the layer after. The number of inputs and outputs units is defined by the specifics of the task, while the network dimension, i.e., the number and size of hidden layers, is optimized independently of the training process, as well as all the other so-called hyperparameters. The arrows in Figure A.1 depict the weights of the model, the connections between the nodes, which are multiplicative weight tensor W_{ij}^l between the node i in the layer $l - 1$, and the node j in the layer l . The grey circles represent the additional additive bias units. So the input to each node a_j^l is given by the weighted sum of the nodes in the previous layer, as:

$$z_i^l = \sum_j W_{ij}^{l-1} \cdot a_j^{l-1} + b_i^{l-1} \Rightarrow z^l = W^{l-1} \cdot a^{l-1} + b^{l-1}. \quad (\text{A.1})$$

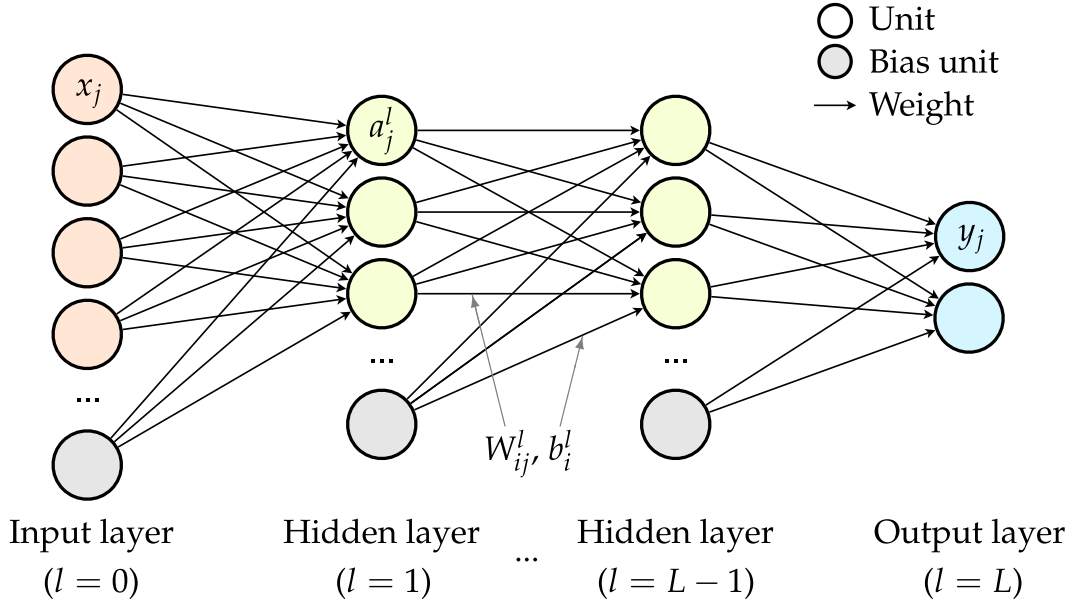


Figure A.1: Overview of a generic fully-connected neural network.

The right side represents the computation by matrix multiplication and vector addition. This linear operation, and any combination of this kind of operation, would prevent the neural network from learning non-linear behavior. Therefore, so-called non-linear activation functions σ^l are used to transform the output of the unit in order for the model to be able to approximate arbitrary functional relations. Hence, the output of each unit is:

$$a^l = \sigma^l(z^l) = \sigma^l(W^{l-1} \cdot a^{l-1} + b^{l-1}). \quad (\text{A.2})$$

The same activation function transforms all layer units, but different functions can be chosen per hidden layer. In particular, the choice of the output layer activation is highly dependent on the specifics of the individual application. Instead, the activation functions of the hidden layer are usually also defined during the hyperparameter optimization. Figure A.2 illustrates a small selection of commonly used activation functions. For multi-class classification tasks, the most common activation function used is the softmax, defined as:

$$\text{softmax}(z) = \frac{e^z}{\sum_i e^{z_i}}. \quad (\text{A.3})$$

It scales all outputs so that the total sum of all outputs equals unity and each output has the properties of a probability [284].

In general, the neural network model is a composition of nodes and mathematical expressions, given the set of free parameters θ [285, 286]:

$$\vec{y} = f(\vec{x}; \theta) = f(\vec{x}; \underbrace{W, b}_{\theta}) = \underbrace{(a^L \circ a^{L-1} \circ \dots \circ a^1)}_{\text{model}}(\vec{x}). \quad (\text{A.4})$$

The best set of parameters $\theta := (W, b)$ is determined during the neural network training process.

A.2 Neural network training

The neural network is trained in an iterative fitting approach to obtain the free parameters θ with a dataset representative of the underlying problem. Generally, several strategies can be

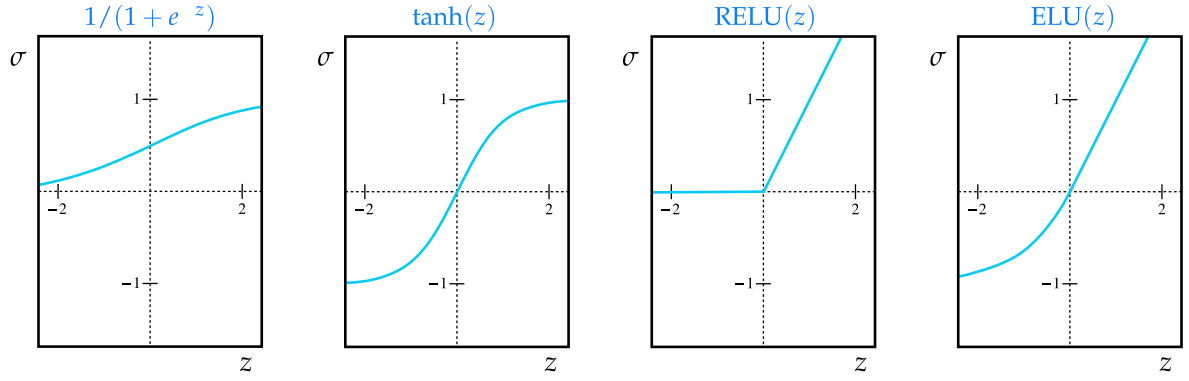


Figure A.2: Examples of commonly used activation functions for neural networks. From left to right: logistic, hyperbolic tangent, rectified linear unit (RELU), and exponential linear unit (ELU) function.

used for training. In this thesis, a supervised-learning approach is deployed, which features a task-specific loss function that maps the predicted output $\hat{y}$ to its true expected outcome $\vec{y}$ [285]. In particular, this mapping represents either true or false, i.e., one or zero, and the loss function L measures the difference between the final output $\hat{y}$ of the ANN and the predetermined label. The mean squared error (MSE) or the mean absolute percentage error (MAPE) loss is effective for the regression of a single quantity:

$$L(\vec{y}, \vec{\hat{y}})_{MSE} = \frac{1}{N} \sum_{i=0}^N (y_i - \hat{y}_i)^2 \quad (\text{A.5})$$

$$L(\vec{y}, \vec{\hat{y}})_{MAPE} = \frac{1}{N} \sum_{i=0}^N \frac{|y_i - \hat{y}_i|}{y_i}; \quad (\text{A.6})$$

while (multi-class) classification applications, as in the context of this thesis, often rely on the cross-entropy (CE) loss which considers the difference among all training classes:

$$L(\vec{y}, \vec{\hat{y}})_{MSE} = \frac{1}{N} \sum_{i=0}^N -\hat{y}_i \log y_i \quad (\text{A.7})$$

The dimension N represents the number of output classes, and $\hat{y}$ denotes a vector with length N whose elements have a value of one for a specific class and zero otherwise. The logarithm is evaluated element by element. The training aims to tweak the weight matrices θ so that the ANN output is as close to the true output classes $\vec{y}$ as possible. To achieve a statistically reliable estimate, the loss is frequently computed as an average over subsets of the total dataset, referred to as batch samples. This results in the training being a minimization task of the function L for a statistically significant number of examples. Hyperparameter optimization is also applied to find the best-performing batch size.

The minimization procedure is accomplished by an iterative, stochastic process known as batch gradient descent (BGD), which relates any modification in the loss function in each iteration t to a procedure for updating the model weights θ_i along the gradient towards the minimum. It can be expressed as:

$$\theta^{t+1} \leftarrow \theta^t + \Delta\theta = \sum_i \theta_i^t - \delta \frac{\partial L(\theta)}{\partial \theta_i} \quad (\text{A.8})$$

with δ commonly known as learning rate to regulate the derivative contribution of the respective weights in each training iteration. As a result, the loss change is negative, and the overall

loss is minimized. However, the repeated computation of the numerous and complex derivatives of L is inefficient and computationally very intensive. The so-called backpropagation algorithm provides a more straightforward solution [287]; using the layerwise structure of neural networks, it can be proved that, using δ^l as local error for each layer l , the derivatives can be expressed as [283, 288]:

$$\frac{\delta L}{\delta \theta_{ij}^l} = \begin{cases} a_j^{l-1} \cdot \delta_i^l, & \text{for the } \theta_{ij}^l \text{ weight vector} \\ \delta_j^l, & \text{for the } b_j^l \text{ is a bias vector,} \end{cases} \quad (\text{A.9})$$

with

$$\delta_i^l = \begin{cases} (\sum_j W_{ij}^{l+1} \delta_j^{l+1}) \cdot \sigma'(z^l), & \text{for } l < L \\ \frac{\partial L}{\partial y} \cdot \sigma'(z^l), & \text{for } l = L. \end{cases} \quad (\text{A.10})$$

As a result, the weights are more efficiently updated, and the complete neural network training with backpropagation proceeds until a minimum of the loss function is found. An epoch is an iteration where all the weights and biases are updated multiple times for all batches of data. Many other rules offer faster convergence in fewer epochs, based on different optimizers. The one considered in the scope of this thesis is the AdaGrad algorithm [289, 290], derived from the GradientDescent method.

The updates of the weights are crucially dependent on the layerwise computed local errors δ^l and so on the gradient $\sigma'(z^l)$ of the activation functions. Gradients close to zero cause negligible errors, vanishing weight updates, and prevent the network from learning additional information over time. This is a common issue when optimizing the weights of the NN in an iterative method, as mentioned above. Nevertheless, several ways can be used to address the so-called vanishing gradient problem [283, 288].

The input to the first hidden layer $z^1 = W^0 \cdot x + b^0$ has a generally unrestricted numeric domain because each input feature x_j has a different numeric domain. However, the activation functions tend to have preferable regions, usually around zero, in which their values, and derivatives, are most meaningful in the context of backpropagation. As a result, a straightforward solution is to bind the parameter range of weights and biases to prevent extreme output values. Another related method is to transform each input so that its distribution has a mean of zero and a standard deviation of one:

$$x_j \rightarrow \xi_j' = \frac{x_j - \mu_j}{\sigma_j}, \quad (\text{A.11})$$

where μ_j (σ_j) is the original distribution's mean (standard deviation). Further optimization can be achieved by ensuring an equally appropriate numeric domain of the layer output $a^1 = \sigma^1(z^1)$, by drawing the initial values for all node weights and bias from specific probability density distributions, such as the uniform or normal distribution. An example is the Glorot initialization process [291], in which biases are initialized to zero but updated during training to compensate for numeric offsets, and weights are taken from a uniform distribution from an interval respecting the dimension of the input layer. However, other methods exist to prevent vanishing gradients more efficiently, especially in deep networks. For example, when employing Batch Normalization, hidden layer activations are scaled with adaptive scaling coefficients per unit [292]. Similar results can be obtained when the SELU activation function as it exhibits intrinsic self-normalizing behaviour [284].

The optimization of the hyperparameters can be carried out with numerous methods. The simplest is the so-called grid-search strategy, in which all permutations of a distinct set of parameters from the available parameter space are tested. Faster convergence is achieved

with a randomized technique, in which only some permutations are used. Another strategy is based on Bayesian probability and, unlike the previous approaches, tracks previous iterations. An a priori distribution is updated towards a posteriori distribution based on previous trials to construct an approximate of the objective function. This posteriori function is then deployed to find the next set of hyperparameters for the subsequent iteration.

The Receiver Operating Characteristic (ROC), shown in Figure A.3 (a), is a regularly used metric to illustrate the separation power of ANNs. The ROC represents the proportion of events correctly identified as a function of the proportion of events assigned to the incorrect class (i.e., the rate of correctly detected signal events as a function of the wrongly classified background processes). Ideally, the ANN correctly classifies all occurrences, and the area under the ROC curve (ROC-AUC), assessed to quantify the separation power, equals one. However, when the ANN can not perform the classification task, the probability of randomly selecting real signal events or background events is the same, giving a ROC-AUC of 0.5. Another metric highly employed in this thesis is the confusion matrix, an example in Figure A.3 (b). All entries equal to one in the diagonal and all other elements equal to zero represent perfect discrimination among classes. Overall, the confusion matrix contains the classification performances and the misidentification probability for each class.

A.2.1 Solutions to prevent overtraining

The availability of data for training a neural network is generally limited. Where the set of training examples is an inadequate representation of the underlying probability distributions, the network is more sensitive to specific features of the dataset, which are more prone to statistical fluctuations. Therefore, the networks ability to generalize the learnt behaviour has to be monitored. The data is commonly divided into three statistically independent sets known as training, validation and testing. The training set is used to calculate all gradients and update free parameters during the training phase, while the loss is also calculated in each iteration on the statistically independent validation set with the goal of having it identical to

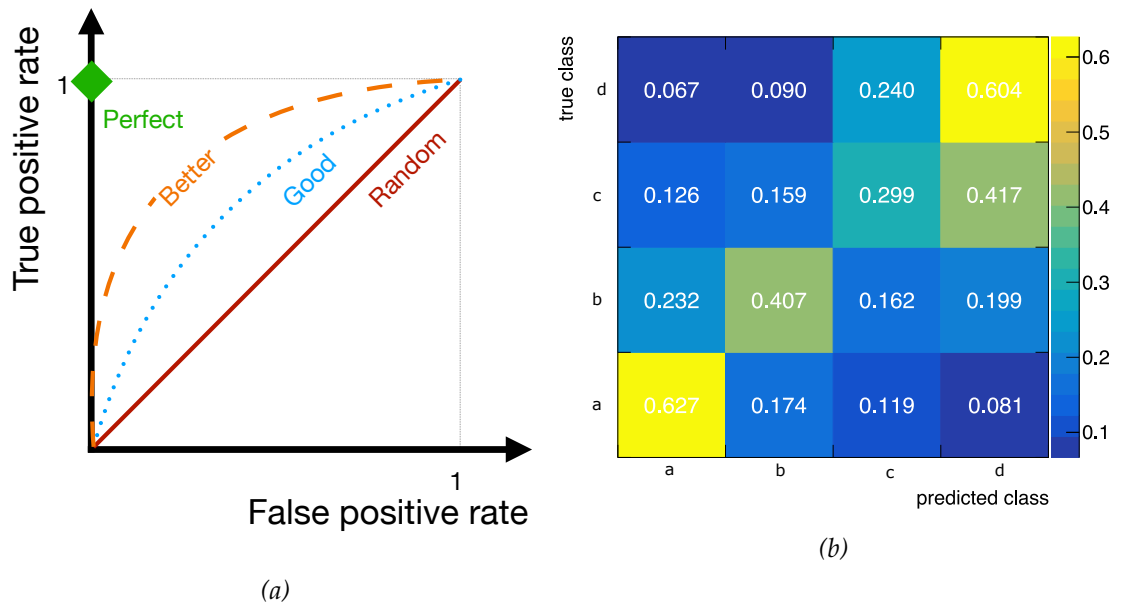


Figure A.3: (a) Exemplary curves for the ROC metric. For comparison, the metric is also shown for an ANN that cannot separate between the classes (red line). (b) Example of a confusion matrix for ANN classification with four classes. The values are normalized to one in each row of the matrix.

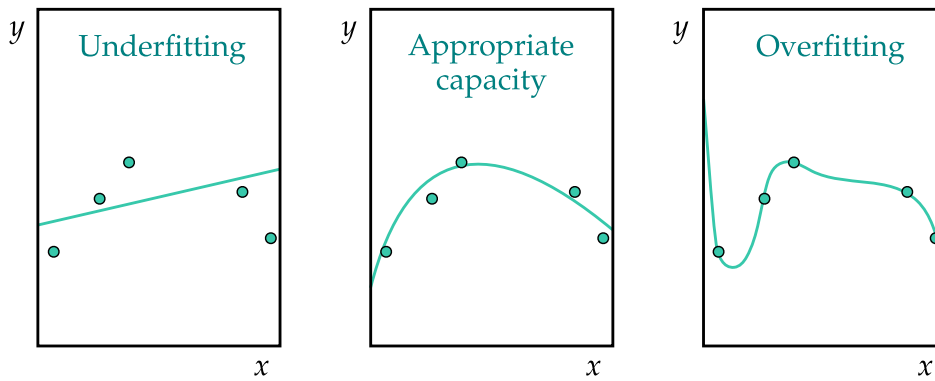


Figure A.4: Examples of typical states of capacity for a neural network model, which can result in underfitting (left), overfitting (right), or an optimal fit (center).

the training loss used for backpropagation. As soon as the model begins to fail to generalize on the validation set, the training is halted to prevent what is called overtraining. Given that the validation dataset is effectively used to optimize parameters, the third independent testing (or evaluation) dataset is necessary for final evaluation of the ANN. This technique assures that no artificially-augmented separation affects final discrimination, which can occur when training data is evaluated using an over-trained network.

Nevertheless, to identify hidden representation and higher-order features in data, advanced applications rely on networks with complex models involving many free parameters. This results then in a common trade-off situation, as depicted in Figure A.4. If the ability to learn complex models is too low or the training data set too small, the consequence often is underfitting the data. On the contrary, even though it is reasonable to have a high enough ability to describe the training data, this may cause problems with generalization when evaluating on new, independent data because no appropriate complex representation of the data has been learned. When training a neural network, it is important to account for two types of mistakes associated with varying capacity (i.e., the ability to approximate or learn complex models): bias and variance. The performance on the training set is affected by bias if the average model response deviates from the truth. In contrast, variance is linked to the flexibility to generalize to different related data sets. A large (low) capacity typically corresponds to a high variance (bias). As illustrated in Figure A.5, to identify the optimal capacity for a given task and model, it is necessary to find the right trade-off between bias and variance.

Several methods are deployed during the training to help limiting the variance and in general suppressing overtraining. Usually, when evaluated on the training data set, the loss function falls continuously, indicating a successful minimization. However, the loss function of statistically independent validation data set could increase, implying over-training of the ANN. This thesis uses early-stopping to avoid this effect. If the two loss function diverge beyond a threshold or if the difference does not reduce after a certain period of time during the training, the training is aborted and the current ANN is saved and used. The architecture hyperparameters include also the threshold and the time before which training is halted prematurely.

A simple strategy is to implement dropout units, which randomly set output nodes to zero based on a flat prior function and making the network less dependent on the performance of any single node. Another popular regularization strategy is the so-called 'weight decay or L2 regularization. In this case, the loss function is modified by including an extra penalty term with a given hyperparameter λ , which controls the intensity of the regularization ($\lambda = 0$ is equivalent to not employing any regularization at all). The penalty term applies to the

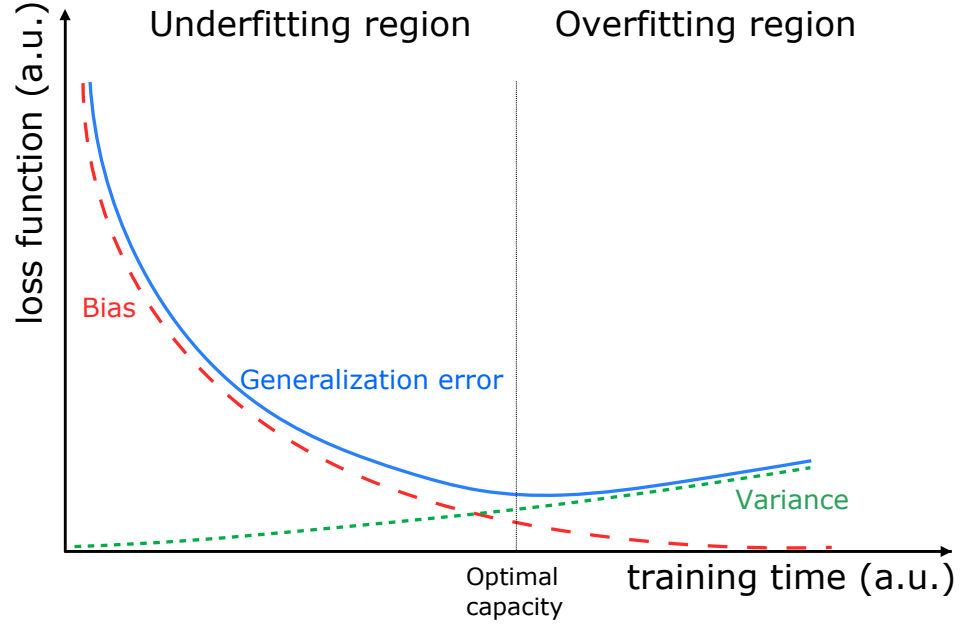


Figure A.5: A diagram showing the correlation between a neural network model's capability and its susceptibility to overfitting or underfitting, as well as its variance and bias. The optimal capacity is indicated by a dotted line and results in minimal bias and variance.

weights and not the bias as:

$$\lambda \sum_{i,j} W_{i,j}^2, \quad (\text{A.12})$$

so its effect is to prioritize smaller weights, which mitigates training bias towards specific inputs or nodes.

Finally, the predicted value distributions for the different classes are compared between the training and test data sets with a Kolmogorov-Smirnov and a χ^2 tests. That specific network is not considered in this analysis if the shapes of the distributions are not compatible within the statistical precision of the data sets.

STATISTICAL INFERENCE



A statistical inference is performed based on the confidence level CLs criteria [293] to make a quantitative statement on the possible presence of a signal.

The statistical foundation for estimating the t̄tH signal strength modifier $\mu = \sigma/\sigma_{\text{SM}}$ is described in this section. Systematic uncertainties, such as detector effects or theoretical cross section information, are considered nuisance parameters, whereas μ is considered an unconstrained model parameter. The signal extraction technique is carried out in all analysis categories at the same time, using a binned maximum likelihood fit of a simulated variable distribution to recorded data to test the hypotheses of background-only (b) and signal + background (s + b).

The first part focuses on developing the underlying statistical model as likelihood. Then there is an introduction to the various approaches used for inference of the results, ranging from the calculation of an upper limit on μ at a 95% confidence level to the exclusion significance of the background-only hypothesis and the profile likelihood fit for the extraction of μ . The final segment describes the χ^2 -based goodness-of-fit test, primarily used with a saturated model to quantify the agreement between simulated and measured events.

The technical implementation of the statistical model and signal extraction approach used in this research was developed at the LHC by the ATLAS and CMS Collaborations in the context of the Higgs boson search results [238].

The inference procedures are carried out using the CMS experiment's so-called `combine` tool in version 8.2.0 [294].

B.1 The likelihood function

The probability density function of the observed data, as a function of a set of parameters, is the likelihood function. This analysis is based on binned histograms, each with independent content, and can be viewed as the result of a counting experiment with an expected yield of events λ . A Poisson distribution governs the likelihood of observing a number of n events:

$$\mathcal{L}(n|\lambda) = \frac{\lambda^n}{n!} \exp^{-\lambda}.$$

This can be extended to the case of multiple N independent bins, each observing n_i and expecting λ_i occurrences; the likelihood function is the product of the likelihood functions for each counting experiment:

$$\mathcal{L}(\vec{n}|\vec{\lambda}) = \prod_{i=1}^N \frac{\lambda_i^{n_i}}{n_i!} \exp^{-\lambda_i}, \quad (\text{B.1})$$

with $\vec{n} = \{n_1, \dots, n_N\}$ and $\vec{\lambda} = \{\lambda_1, \dots, \lambda_N\}$ representing the observed and predicted number of events in each bin. In practice, the set of N bins consists of all histogram bins of specified variable distributions in various measurement categories.

The model for $\vec{\lambda}$ incorporates the knowledge of the production mechanism and rates of specific processes, the recorded luminosity of the data, detector acceptance, and reconstruction efficiencies. Systematic and statistical uncertainties in the model are integrated into the likelihood function as nuisance parameters $\vec{\theta} = \{\theta_1, \dots, \theta_L\}$ that affect the expectation values, i.e., $\vec{\lambda} = \vec{\lambda}(\vec{\theta})$. They are subject to prior probability distributions $\pi(\tilde{\theta}_j|\theta_j)$, where the nominal value $\tilde{\theta}_j$ is estimated through theoretical reasoning or specialized measurements where possible. The following changes are made to Eqn. B.1:

$$\mathcal{L}(\vec{n}|\vec{\lambda}) = \prod_{i=1}^N \frac{\lambda_i^{n_i}}{n_i!} \exp^{-\lambda_i} \prod_{j=1}^L \pi(\tilde{\theta}_j|\theta_j). \quad (\text{B.2})$$

For the sake of this analysis, it is more practical to express the number of predicted occurrences $\vec{\lambda}(\vec{\theta})$ as the sum of background and signal contributions:

$$\vec{\lambda} = \mu \cdot \vec{s}(\vec{\theta}) + \vec{b}(\vec{\theta})$$

with μ factorizing the signal yield and not having any a-priori assumptions. When this notation is applied to the likelihood function $\mathcal{L}(\vec{n}|\vec{\lambda})$ of Eqn. B.2, the signal strength μ is explicitly introduced as one of the parameters, in addition to the nuisances, $\mathcal{L}(\vec{n}|\mu, \vec{\theta})$. The values of μ and $\vec{\theta}$ are estimated by a binned maximum likelihood fit. Technically, the logarithm of the likelihood is typically fitted to find the maximum for a sum of terms rather than a product. The best-fit signal strength is the value of μ that minimizes the negative logarithm of the likelihood. The quantity $q = -2 \ln \mathcal{L}$ is plotted as a function of μ , and the boundaries of the $\pm 1\sigma$ interval are estimated by reading the μ values corresponding to the $\Delta q = 1$ line, with $\Delta q = q - q_{\min}$.

The total background b_i in a bin i is the sum of several processes p . Its dependence on the nuisance parameters can be implemented as follows:

$$b_i(\vec{\theta}) = \sum_p^{\text{processes}} \underbrace{b_{p,i}}_{\text{nominal yield}} \cdot \underbrace{\prod_{j=1}^L \pi_{p,i}(\theta_j|\tilde{\theta}_j)}_{\text{uncertainty model}},$$

and similarly for $s_i(\vec{\theta})$. The nominal expected number of events of a process p in bin i is represented by $b_{p,i}$. This expectation is modified by the nuisance parameters θ_j , by the function $\pi_{p,i}(\theta_j|\tilde{\theta}_j)$.

Not all nuisances have the same impact on all processes and bins. The approach presented in this Thesis distinguishes three types of nuisances based on the shape of their prior distributions and the corresponding impact on processes and bins.

1. **RATE-CHANGING** Uncertainties that solely affect the overall event yield of one or more processes, such as theoretical cross sections. Up and down variations, which describe central 68% confidence intervals, vary a template's integral equally in all of its bins without causing any shape change. They are represented by a log-normal distribution $\ln N(\theta)$ with parameter k :

$$\ln N(\theta) = \frac{1}{\theta} \cdot \frac{1}{\sqrt{2\pi} \ln(k)} \exp \left(-\frac{(\ln(\theta/\tilde{\theta}))^2}{2(\ln(k))^2} \right)$$

with k being the width of the distribution and the measured uncertainty [239]. The $\ln N$ distribution is preferred because it approaches zero at $\theta = 0$; negative nonphysical contributions to the observables are avoided without truncating relatively broad normal distributions, which usually lead to systematically overestimated event yields.

2. SHAPE-CHANGING

Uncertainties that shape the distribution structure by introducing changes in predicted event yields correlated across bins. The shape-changing variations are provided by two alternative shapes duplicating an analysis template but obtained by adjusting the nuisance parameter θ according to the bounds of its central 68% confidence level, so by $\pm 1\sigma$ around its original value. A continuous relationship between the value of θ and its impact on the template is required for the parameter θ to be applied within the likelihood: the interpolation between the up and down shapes is provided by a vertical morphing technique, conducted independently per bin. The morphing algorithm uses a 6th-degree polynomial to interpolate between the central and $\pm 1\sigma$ templates and a linear extrapolation outside. Fig B.1 shows an example of such a morphing.

In general, the normalization of the templates is not retained during the morphing method, resulting in rate and shape-changing effects.

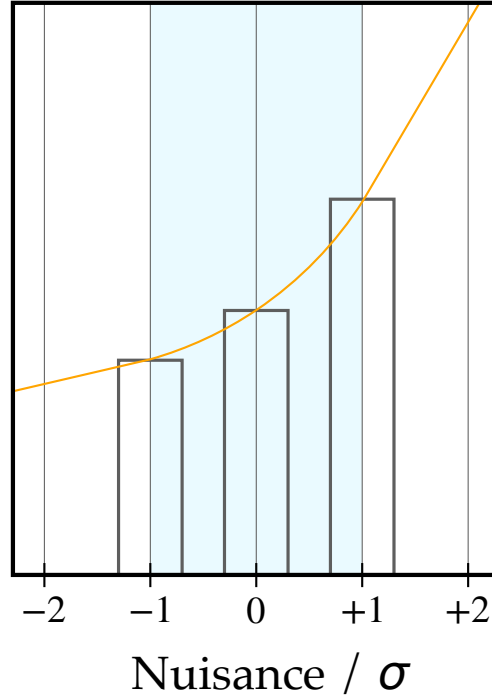


Figure B.1: Example of template morphing. The values at zero represent the nominal event yield, while the variations at $\pm 1\sigma$ represent one standard deviation shift due to a nuisance. The morphing function is given by a polynomial of 6th degree in the central region (light blue shaded area), while linear extrapolation is used outside.

3. **STATISTICAL** Uncertainties resulting from a limited number of available statistics of simulated events used to estimate the expected yield for each background process in each distribution bin. They are also known as bin-by-bin uncertainties and are a type of shape uncertainty in which each bin's behavior is independent of the others inside a single process. To model the statistical uncertainty, many nuisance parameters are introduced, one for each bin of each process, with a prior probability following a Poisson distribution. This results in a rapid increase in the number of nuisances in the likelihood. A single Gaussian distribution can approximate the sum of Poisson distributions for a sufficiently large enough number of events in a given bin; hence, one nuisance per bin is usually integrated into the statistical model representing this

distribution. A specific bin-merging algorithm addresses the requirements for minimum bin content.

The following section discusses likelihood ratios and their uses in statistical inference methods.

B.2 Inference methods

In high-energy physics, statistical inference methods are frequently used to interpret the results of a search for new physic processes by distinguishing between two hypotheses, such as the presence or absence of a hypothesized signal in a well-defined phase space in addition to a known background. A discovery is commonly claimed when the observed data is incompatible with the background-only hypothesis.

According to the Neyman-Pearson lemma, the likelihood ratio is the most powerful test statistic for evaluating two contradictory hypotheses. Even though it does not apply to models with several free parameters, a generalized technique based on likelihood profiling can be applied. For the likelihood introduced in Eqn. B.1, the profile likelihood ratio $\lambda(\mu)$ defined as [237]:

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \vec{\hat{\theta}})}{\mathcal{L}(\mu, \vec{\hat{\theta}})}$$

is commonly used to build the test statistic q_μ :

$$q_\mu = \begin{cases} -2 \ln \lambda(\mu) & \hat{\mu} \leq \mu, \\ 0, & \hat{\mu} \geq 0. \end{cases} \quad (\text{B.3})$$

It constitutes the basis for the following inference methods. First, the signal strength is fixed to the value under test in the numerator of the test statistic. The nuisance parameters $\vec{\theta}_\mu$ are fitted to maximize the likelihood for the given value of μ , leading to values $\vec{\hat{\theta}}_\mu$. Whereas in the denominator, both the signal strength and the nuisances are the maximum likelihood estimators (MLE), i.e., the parameters that maximize $\mathcal{L}$ and minimize $-\ln \mathcal{L}$, respectively. As a result, the likelihood ratio λ is exclusively reliant on μ ; this type of building is known as profiling.

The physical constraint $\hat{\mu} \geq 0$ specifies that the signal contribution cannot be negative. In contrast, the other constraint $\hat{\mu} \leq \mu$ ensures that upward fluctuations in the data, if larger than expected for a signal of strength μ , are not regarded as evidence against the signal hypothesis itself, defining a one-sided confidence interval. According to the formulation of q_μ , small values suggest compatibility between the data and the predicted value of μ . Based on the observed data, distinguishing between the signal-plus-background and background-only hypotheses is comparable to distinguishing between a given value of $\mu > 0$ and $\mu = 0$. The q_μ probability density functions can be calculated by generating pseudo-data with toy MC experiments, but this is very CPU consuming. The asymptotic features of Eq. B.3 may be used to construct the q_μ distribution under the signal-plus-background and background-only hypotheses using analytical equations [237]. The results reported in this Thesis are derived using this asymptotic approximation. It can be shown that q_μ can be used to construct confidence intervals corresponding to s Gaussian standard deviations for one POI in the limit of large samples as $q_\mu(s) = q_\mu^{min} + s^2$. An example of the test statistic and the construction of the confidence intervals is shown in Fig. B.2.

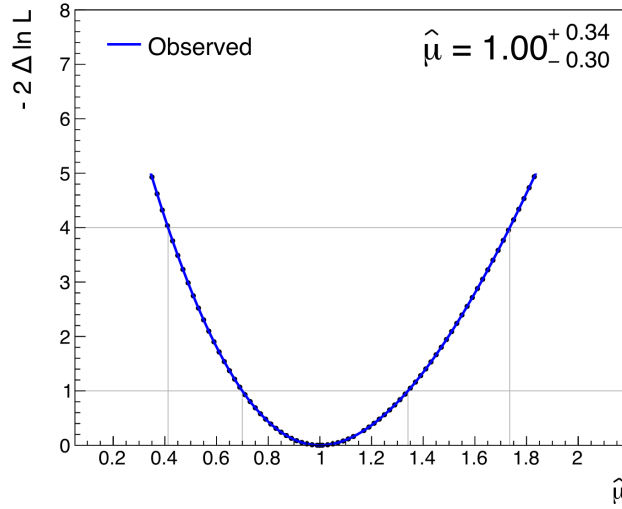


Figure B.2: A visual example of the test statistic as a function of the parameter $\hat{\mu}$. The best estimate for the parameter is located at the minimum; the confidence intervals are visualized by the gray lines, corresponding to one (two) Gaussian standard deviations constructed by crossing the test statistic at one (four).

B.2.1 Exclusion significance

An observation's claim is based on excluding the background-only hypothesis $\mu = 0$. The p-value for the latter hypothesis quantifies the significance of an excess of data above the background-only expectation. Starting with the test statistics described in Eq. B.3, where the value under test is 0, this may be calculated:

$$\tilde{q}_0 = \begin{cases} -2 \ln \lambda(0) & \hat{\mu} \geq 0, \\ 0 & \hat{\mu} < 0, \end{cases}$$

and the p-value expressing the measurement's probability assuming $\mu = 0$ reads as:

$$p_0 = P(q_0 \geq q_0^{\text{obs}}) = \int_{q_0^{\text{obs}}}^{\infty} f(q_0 | \vec{\theta}_0^{\text{obs}}) dq_0,$$

where $f(q_0 | \vec{\theta}_0^{\text{obs}})$ denotes the probability density function of the test statistic given q_0 and the nuisance parameters $\vec{\theta}_0^{\text{obs}}$, and it can be derived by producing pseudo-experiments (toys) using a Monte Carlo approach.

The exclusion significance is commonly given in terms of Gaussian standard deviations, regardless of whether the test statistic is normally distributed:

$$p_0 = \int_s^{\infty} \frac{\exp(-x^2/2)}{\sqrt{2\pi}} dx.$$

A significance larger than $s = 3$ ($p_0 = 1.3 \times 10^{-3}$) is considered evidence for the presence of a new physics process, whereas $s = 5$ ($p_0 = 2.9 \times 10^{-7}$) is considered observation.

B.2.2 Upper Limit

If the signal's significance is low, its cross section is certainly lower than some value, known as the upper limit. The test statistic in Eq. B.3, can be used to distinguish between

the hypothesis that a signal process is created at a specific rate and the alternative that the production rate is lower:

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

A single value of the test statistic $\hat{q}_{\text{obs}}$ is produced for a given measurement. The underlying probability distribution $f(\hat{q}_\mu | \mu, \vec{\theta}_\mu^{\text{obs}})$ is necessary to derive conclusions about the likelihood of measuring a value of $\hat{q}_{\text{obs}}$ or more. This can be accomplished by either sampling from the whole statistical model to generate a toy data set or employing a single Asimov data set. The latter is an approximate data set in which all free parameters are assigned to their predicted values from the simulation.

The p-value describing the signal strength compatibility is defined as:

$$CL_{S+B}(\mu) = \int_{\hat{q}_\mu^{\text{obs}}}^{\infty} f(q_\mu) dq_\mu,$$

At 95 % confidence, the frequentist upper limit on μ can be established by solving $CL_{S+B} = 0.05$. However, in the event of statistically significant downward fluctuations, this upper limit would become arbitrarily small without being able to exclude $\mu = 0$. Consequently, one defines:

$$CL_S = \frac{CL_{S+B}}{CL_B} = \frac{p_\mu}{1 - p_b}$$

with the denominator being the p-value of the consistency with the background-only hypothesis, $p_b = P(q_0 \geq q_0^{\text{obs}})$.

If $CL_S \leq \alpha$ the signal-plus-background hypothesis is rejected with $(1-\alpha)$ confidence level for $\mu = 1$. Similarly, the upper limit on μ at $(1-\alpha)$ confidence level is the value of such that $CL_S = \alpha$. The convention for setting exclusion limits is to use $\alpha = 0.05$ such that the exclusion is at 95% confidence level. In other words, the value produced by setting ' $CL_S = 0.05$ ' reflects the observed, frequentist upper limit above which the measurement with 95% confidence excludes the signal strength.

B.2.3 Profile Likelihood fit

Under the signal hypothesis, the parameter of interest μ is determined by minimizing the negative log-likelihood (NLL) $-2 \ln \mathcal{L}(\vec{n} | \mu, \vec{\theta}_\mu)$, which is commonly normalized as in Eq B.3 . This minimization is typically referred to as a fitting technique in physics, with the maximum likelihood estimators (MLE) comprising the best fit value $\hat{\mu}$. Parameter profiling can simplify the procedure in the scenario of many significant nuisances. First, the nuisance parameter values that would minimize the NLL are computed for a given value $\mu = \mu'$, and the associated minimum of the NLL is retained. Then, this operation is repeated for various values of μ' and represents a scan over all possible values of μ , with the overall minimum NLL value yielding the best fit value $\hat{\mu}$:

$$\hat{\mu} = \arg \min_{\mu} -2 \ln \mathcal{L}(\vec{n} | \mu, \vec{\theta}_\mu)$$

Similarly, the best-fit values and uncertainties of any nuisance parameters can be calculated by treating them as the scan's parameter of interest and profiling out all other

parameters, with the real parameter of interest optionally fixed at its best fit value. The a-posteriori, or post-fit nuisances θ_{post} , and especially the nuisances pulls:

$$\text{pull} := \frac{\theta_{\text{post}} - \theta_{\text{pre}}}{\Delta\theta_{\text{pre}}}$$

give useful information for determining the robustness of the underlying statistical model. The pull measures how far the nuisance parameter is pulled from its predicted value by the fit in number of standard deviations $\Delta\theta_{\text{pre}}$. When the pull is ≤ 1 , the situation is considered healthy; if this is not the case, more investigation is usually necessary. The post-fit nuisance parameter values can have a smaller uncertainty than the pre-fit values if the data can constrain them in the fit.

The first fit is done using the Asimov data set after all systematics have been incorporated inside the fit to detect any suspicious behavior. The Asimov data set contains the expected yields from the simulation as data. It is used to assess the fit performance and quantify the expected signal sensitivity and constraints on the nuisance parameters. The value of the pull from the fit to the Asimov data set is zero, by definition.

A crucial information for understanding the fit is the amount by which the value of the parameter of interest differs when the value of a nuisance parameter is changed. The influence of a nuisance is defined as:

$$\text{impact} := \Delta\mu^{\pm} = \hat{\mu}_{\hat{\theta} \pm \sigma_{\Theta}^{\text{post-fit}}} - \hat{\mu}$$

where $\hat{\mu}$ is the value that maximizes the likelihood and $\hat{\mu}_{\hat{\theta} \pm \sigma_{\Theta}}$ is the value of the parameter of interest extracted from a fit in which all nuisances are allowed to vary except θ which is fixed to the post-fit $\pm 1\sigma$. Each nuisance parameter has an influence, and not all are equally significant. A ranking graphic is often used to sort the nuisance parameters based on their impact.

B.2.4 Validation of the Statistical Model

The statistical approach described above offers a method for using measured data to quantify the presence of a signal or to set a limit on its magnitude. However, there is a risk of bias being introduced when the event selections and analysis procedures are set, which might skew the inference made from the data. A process known as blind analysis ensures that such a circumstance does not arise. The analytic technique is defined, developed, and validated using MC simulation or data in signal-free control zones. Only when the technique has been established is it allowed to unblind, examine the data in the signal areas, and determine whether or not a signal is present. To blind analyses is a common technique in CMS Collaboration searches: whenever a new data set becomes available, the search is built and optimized blindly, and tests of the correctness of the physics modeling precede the unblinding.

B.3 Saturated goodness-of-fit test

A goodness-of-fit (GOF) test is used to evaluate a hypothesis without a specified alternative. As a result, the Neyman-Pearson lemma for comparing two contradicting hypotheses is the appropriate prescription for doing GOF tests [295]. The analysis presented here uses the binned GOF tests to validate the modeling of variables used as inputs to deep neural networks and the correlations between them.

The event yields per bin are normally distributed for sufficiently large statistics, and the likelihood for a mean μ_i and variance σ_i^2 in the bin i , given a measurement x_i , is:

$$\mathcal{L}(\vec{x}|\vec{\mu}, \vec{\sigma}^2) = \prod_i^{\text{bins}} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(\frac{-(x_i - \mu_i)^2}{2\sigma_i^2}\right).$$

The likelihood of the so-called saturated model is calculated by assigning all parameters μ_i to the observed values x_i and formulating an artificial alternative hypothesis as

$$\mathcal{L}_{\text{saturated}}(\vec{x}|\vec{\mu} = \vec{x}, \vec{\sigma}^2) = \prod_i^{\text{bins}} \frac{1}{\sqrt{2\pi\sigma_i^2}}.$$

A ratio is then employed:

$$\lambda = \frac{\mathcal{L}(\vec{x}|\vec{\mu}, \vec{\sigma}^2)}{\mathcal{L}_{\text{saturated}}(\vec{x}|\vec{\mu} = \vec{x}, \vec{\sigma}^2)}$$

and its transformation $-2 \ln \lambda$ follows asymptotically a χ^2 distribution. This eventually gives origin to the p-value of the test, which states the likelihood of obtaining a value of the test statistic that is at least as large as the observed value for the corresponding degrees of freedom.

MC SETTINGS FOR THE $t\bar{t} + \text{JETS}$ PRODUCTION

in this analysis the major background contributions arise from $t\bar{t} + \text{jets}$ production, modelled with simulated events obtained from two different generators implemented in the POWHEG (v.2) framework [97–99, 242] interfaced with PYTHIA (v.8.230) [96] for the parton showering and hadronization simulation. The version and configuration of the two generators are summarised in Table C.1. The parameters for the underlying event description correspond to the CP5 tune [240] as for all signal and the other main background processes (except for the minor background processes in the samples matching the 2016 conditions where the CUETP8M1 tune [296] has been used). The choice of the renormalisation and factorisation scales $\mu_{R/F}$ for the $t\bar{t} + b\bar{b}$ samples is motivated by the presence of two different relevant scales related to the $t\bar{t}$ and $b\bar{b}$ systems [21, 44, 118, 297]. The PDF sets have been chosen consistently to the flavour scheme.

	$t\bar{t}$ sample	$t\bar{t}b\bar{b}$ sample
POWHEG version	Powheg v2	Powheg-Box-Res
PYTHIA version	8.230	8.230
Flavour scheme	5	4
PDF set	NNPDF3.1	NNPDF3.1
m_t	172.5 GeV	172.5 GeV
m_b	0	4.75 GeV
μ_R	$\sqrt{\frac{1}{2}(m_t^2 + m_t^2)}$	$\frac{1}{2}\sqrt[4]{m_t \cdot m_t \cdot m_b \cdot m_b}$
μ_F	μ_R	$\frac{1}{4}[m_t + m_t + m_b + m_b + m_g]$
h_{damp}	1.5	1.379
Tune	CP5	CP5

Table C.1: Generator version and configuration of the $t\bar{t}$ and $t\bar{t}b\bar{b}$ samples.

The combination of the two samples and further details of the $t\bar{t} + \text{jets}$ background model are presented in Section 5.5.

D

TRIGGER PATHS

Final state	Trigger paths	Run Era
e^+e^-	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ	B-H
	HLT_Ele27_WPTight_Gsf	B-H
$e^\pm\mu^\pm$	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL	B-H
	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ	B-H
	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL	B-H
	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ	B-H
	HLT_Ele27_WPTight_Gsf	B-H
	HLT_IsoMu24	B-H
	HLT_IsoTkMu24	B-H
$\mu^+\mu^-$	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL	B-G
	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ	H
	HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_	B-G
	HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ	H

Table D.1: List of triggers used for the 2016 data. Multiple triggers in the same run period are combined with a logical OR.

Final state	Trigger paths	Run Era
e^+e^-	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL	B-F
	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ	B-F
	HLT_Ele32_WPTight_Gsf	B-F
	(Emulated via HLT_Ele32_WPTight_Gsf_L1DoubleEG)	
$e^\pm\mu^\pm$	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL	B-F
	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ	B-F
	HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ	B-F
	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ	B-F
	HLT_Ele32_WPTight_Gsf	B-F
	(Emulated via HLT_Ele32_WPTight_Gsf_L1DoubleEG)	
	HLT_IsoMu24_eta2p1	B-D
$\mu^+\mu^-$	HLT_IsoMu27	B-F
	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ	B
	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8	C-F
	HLT_IsoMu24_eta2p1	B-D
	HLT_IsoMu27	B-F

Table D.2: List of triggers used for the 2017 data. Multiple triggers in the same run period are combined with a logical OR.

Final state	Trigger paths	Run Era
e^+e^-	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL	A-D
	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ	A-D
	HLT_Ele32_WPTight_Gsf	A-D
$e^\pm\mu^\pm$	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL	A-D
	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ	A-D
	HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ	A-D
	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ	A-D
	HLT_Ele32_WPTight_Gsf	A-D
	HLT_IsoMu24	A-D
$\mu^+\mu^-$	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8	A-D
	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8	A-D
	HLT_IsoMu24	A-D

Table D.3: List of triggers used for the 2018 data. Multiple triggers in the same run period are combined with a logical OR.

ADDITIONAL DISTRIBUTIONS FOR CORRECTIONS TO EVENT SIMULATION

E.1 Trigger efficiencies

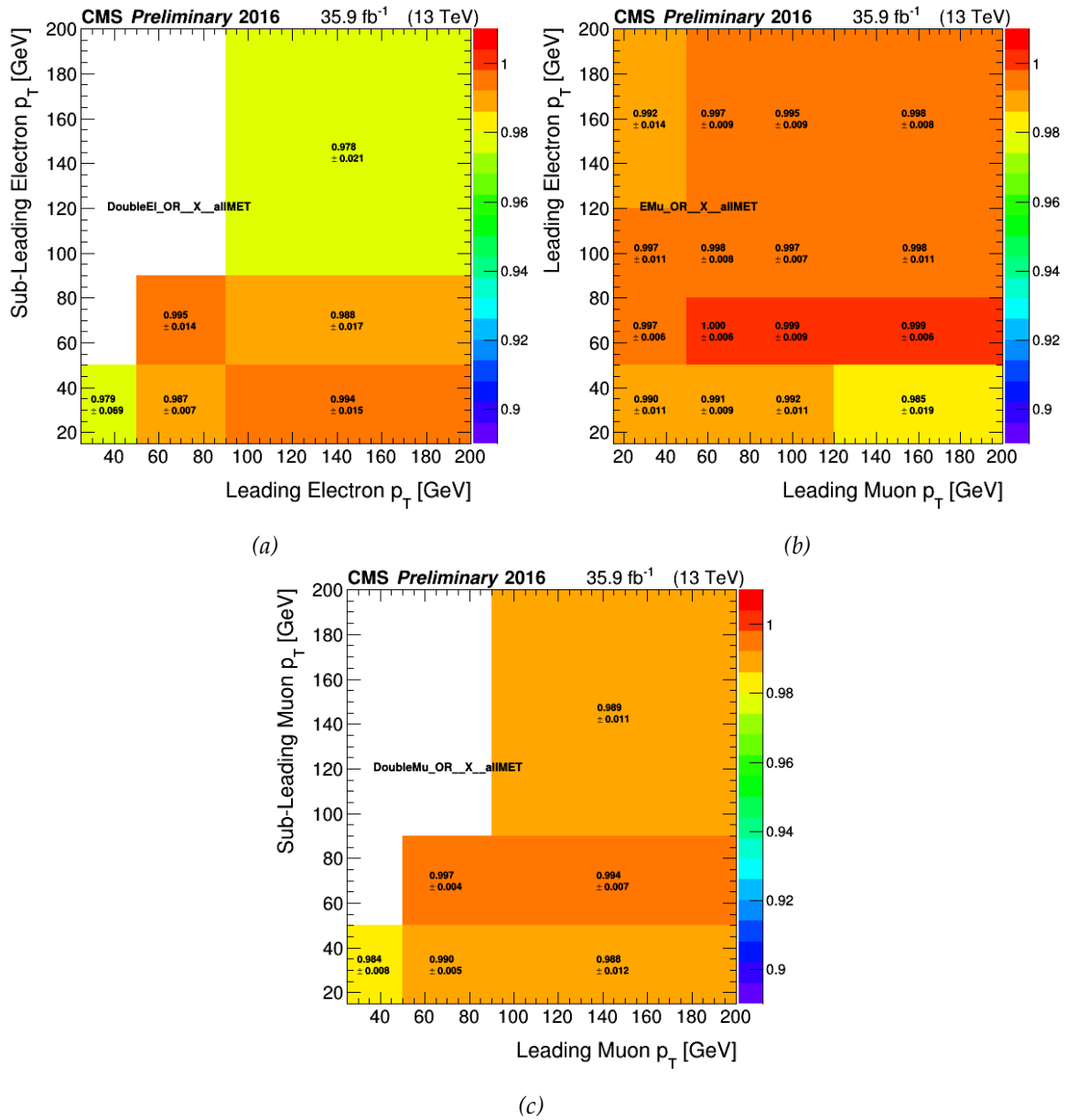


Figure E.1: Final 2D scale factors derived for the 2016 era, for the e^+e^- channel (a), the $e^\mp\mu^\pm$ channel (b) and the $\mu^+\mu^-$ channel (c), as a function of the leading and sub-leading lepton p_T showing the quadrature sum of statistical and systematic uncertainties.

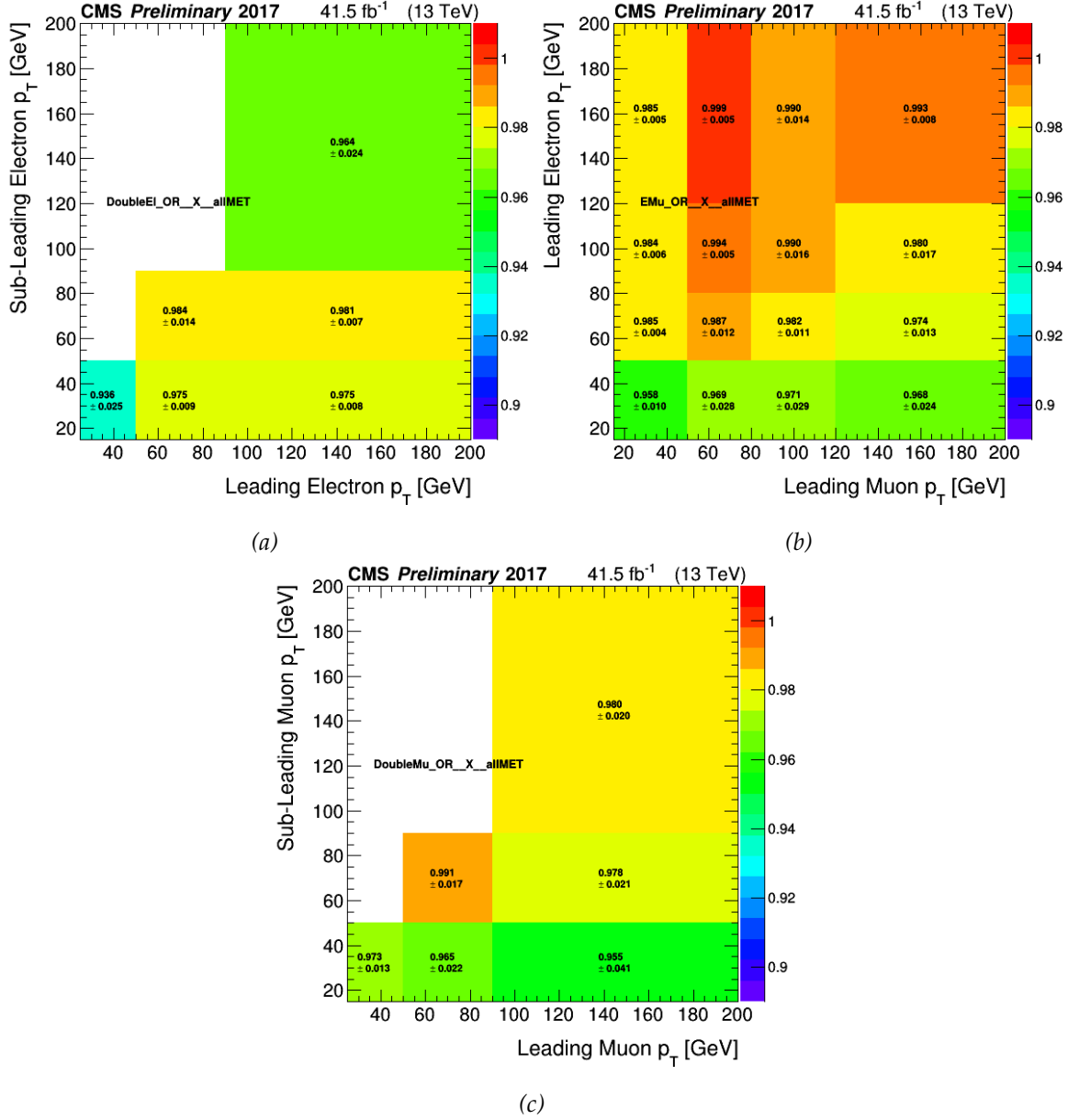


Figure E.2: Final 2D scale factors derived for the 2017 era, for the e^+e^- channel (a), the $e^\mp\mu^\pm$ channel (b) and the $\mu^+\mu^-$ channel (c), as a function of the leading and subleading lepton p_T showing the quadrature sum of statistical and systematic uncertainties.

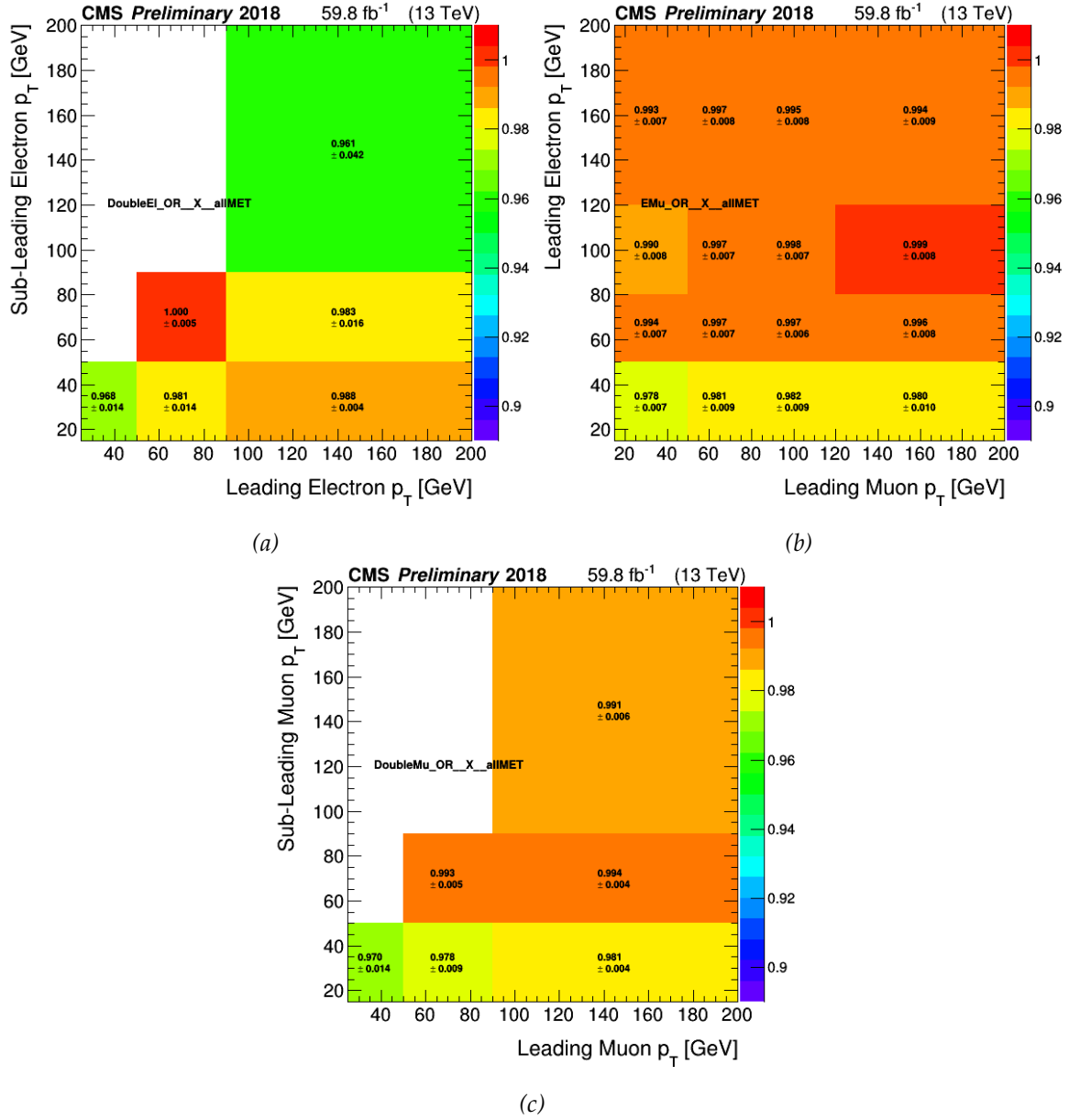


Figure E.3: Final 2D scale factors derived for the 2018 era, for the e^+e^- channel (a), the $e^\mp\mu^\pm$ channel (b) and the $\mu^+\mu^-$ channel (c), as a function of the leading and subleading lepton p_T showing the quadrature sum of statistical and systematic uncertainties.

E.2 b-tagging corrections

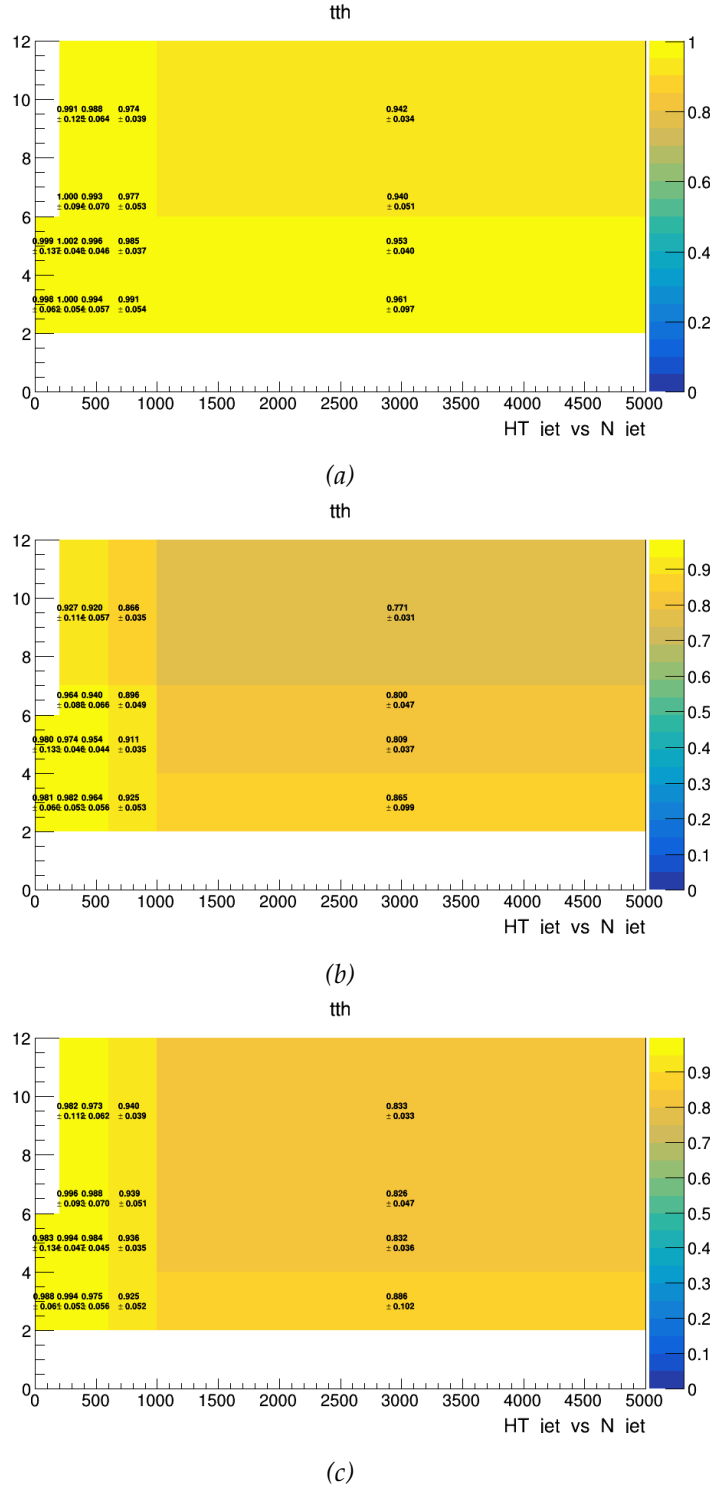
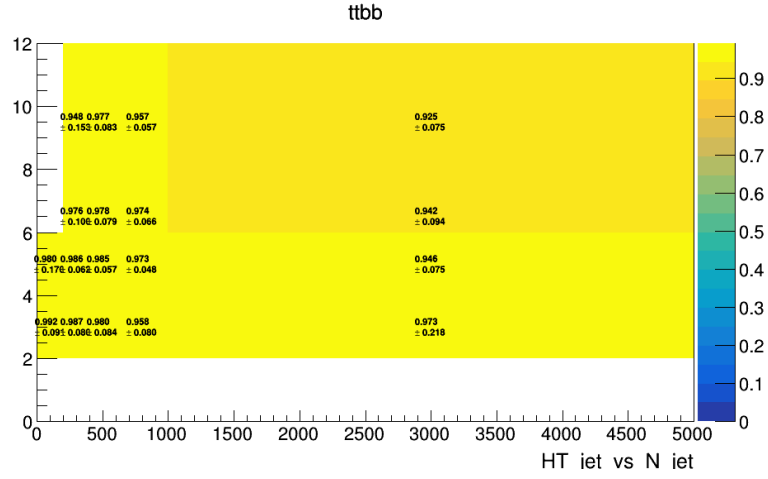
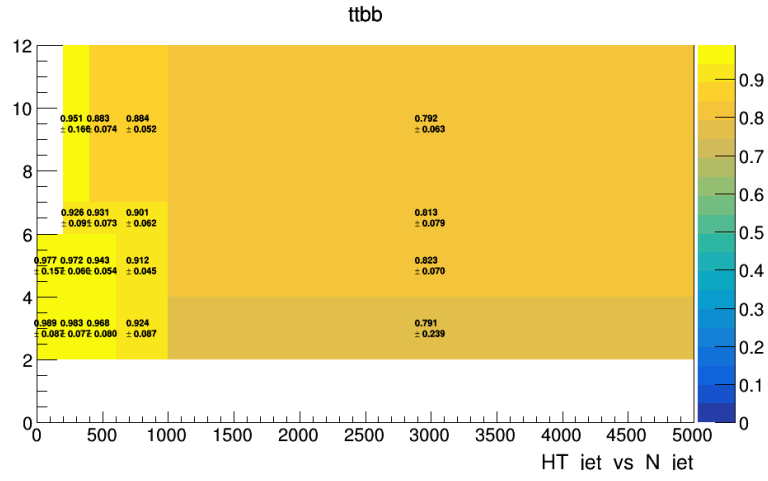


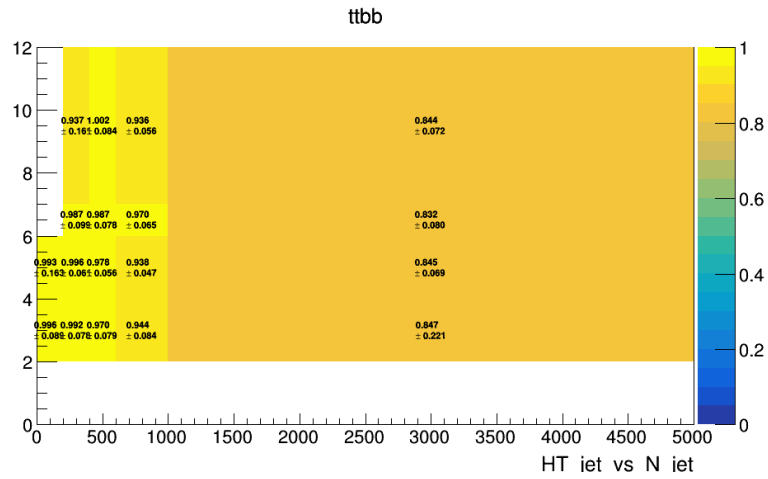
Figure E.4: Correction factors to account for the change of MC distributions induced by the application of the b -tagging scale factors, for the DEEJET tagger, for the $t\bar{t}H$ process, as function of the number of jets and the H_T of the events. The factors are shown for 2016 (a), 2017 (b) and 2018 (c) eras.



(a)



(b)



(c)

Figure E.5: Correction factors to account for the change of MC distributions induced by the application of the b -tagging scale factors, for the DEEPJET tagger, for the $t\bar{t} + b\bar{b}$ process, as function of the number of jets and the H_T of the events. The factors are shown for 2016 (a), 2017 (b) and 2018 (c) eras.

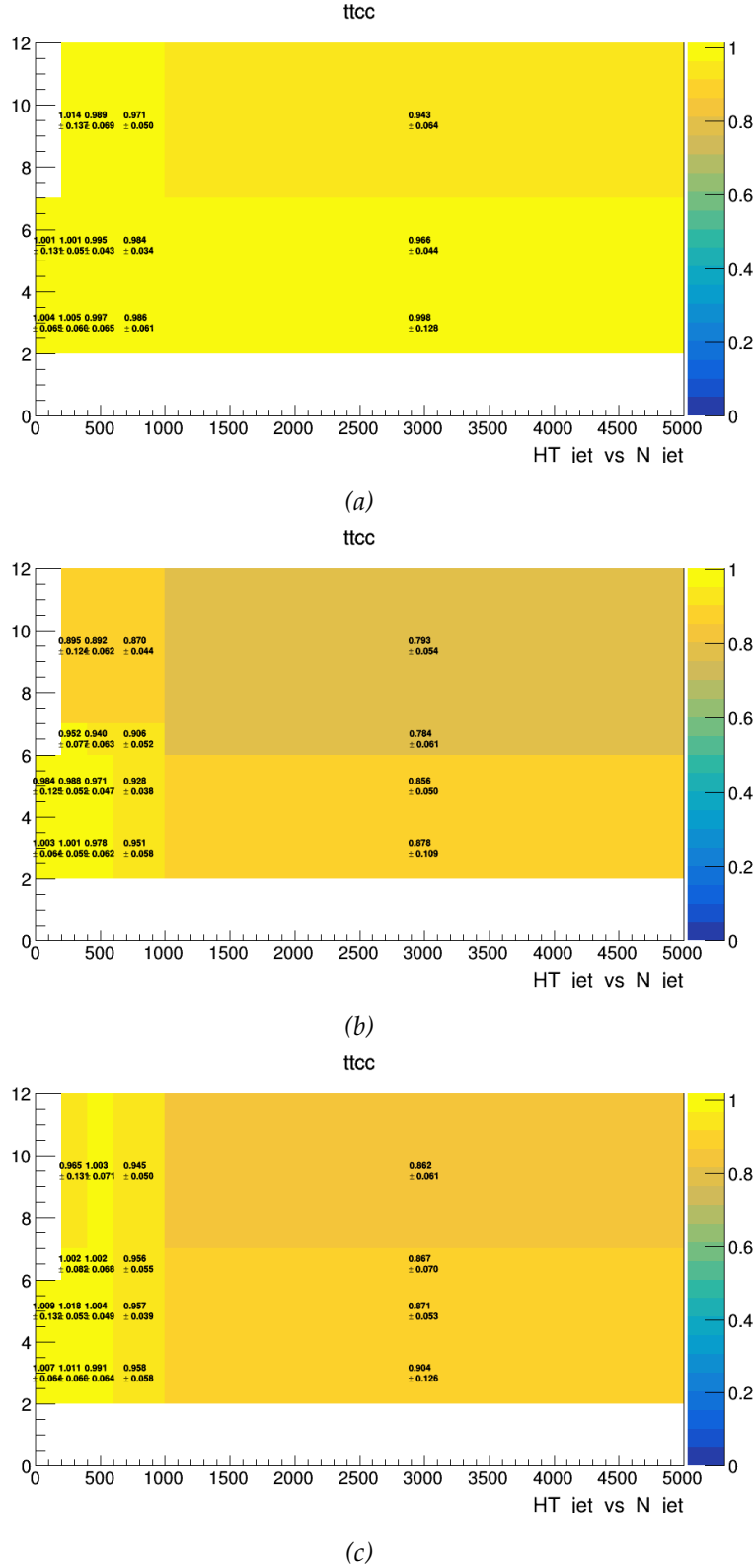
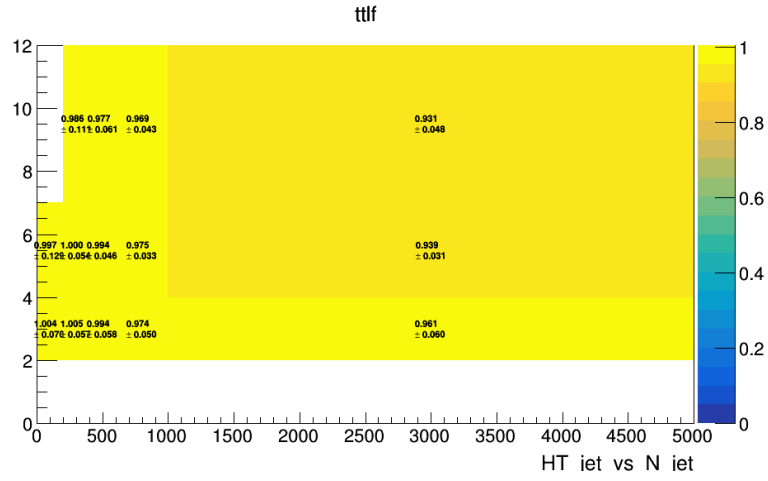
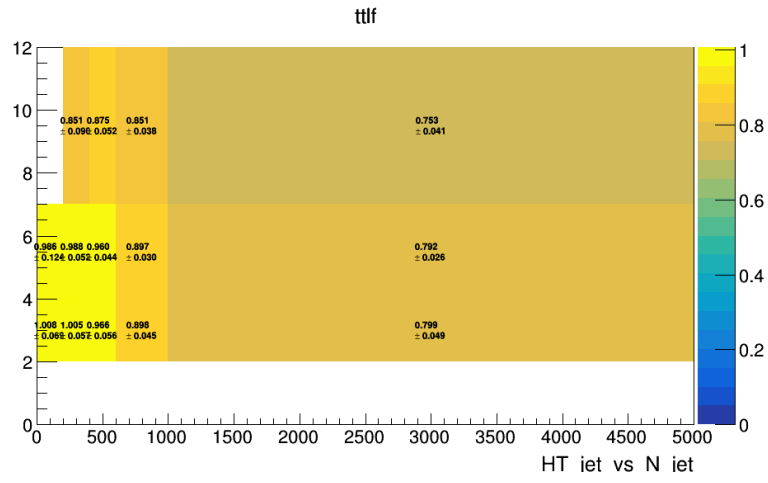


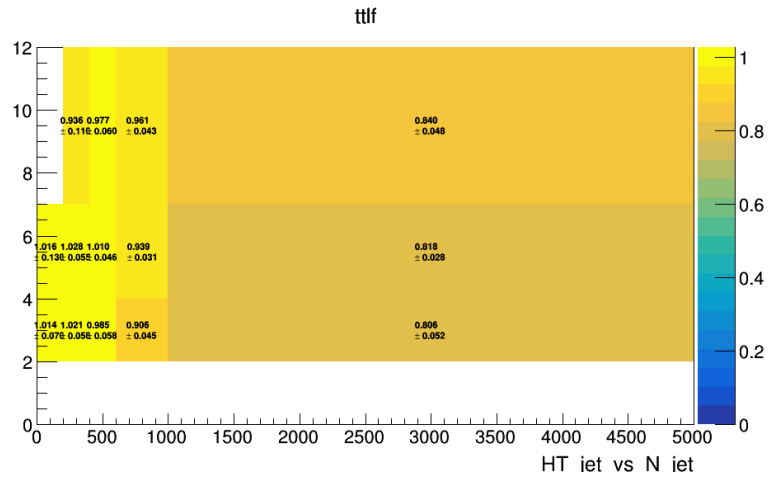
Figure E.6: Correction factors to account for the change of MC distributions induced by the application of the b -tagging scale factors, for the DEEPJET tagger, for the $t\bar{t} + c\bar{c}$ process, as function of the number of jets and the H_T of the events. The factors are shown for 2016 (a), 2017 (b) and 2018 (c) eras.



(a)



(b)



(c)

Figure E.7: Correction factors to account for the change of MC distributions induced by the application of the b -tagging scale factors, for the DEEPJET tagger, for the $t\bar{t}$ + LF process, as function of the number of jets and the H_T of the events. The factors are shown for 2016 (a), 2017 (b) and 2018 (c) eras.

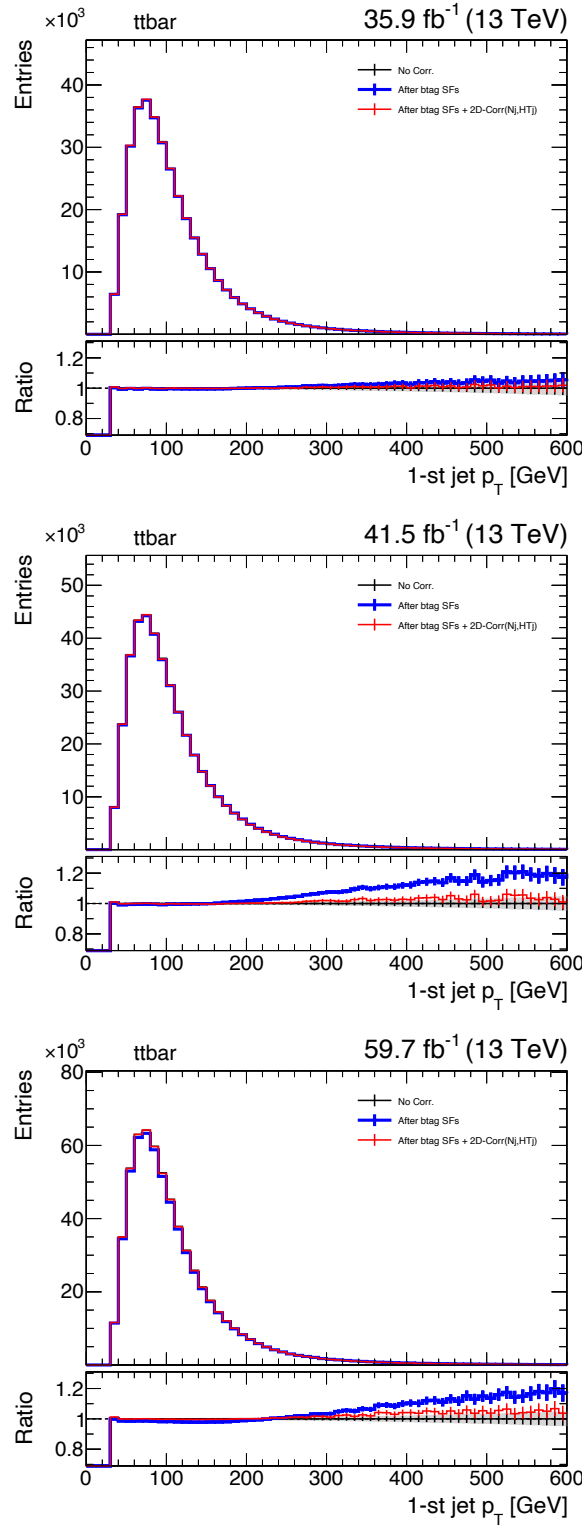


Figure E.8: Closure check for the leading jet p_T for 2016, 2017 and 2018 for all $t\bar{t}$ events in the DL channel. The black line shows the distribution before applying the b -tagging SF. The blue line shows the distribution after applying the b -tagging SF and the red line shows the distribution after applying the b -tagging SF as well as the normalization scale factor. These plots show that the application of the b -tagging normalization scale factor does not change the shape of the corrected b -tagging distribution.

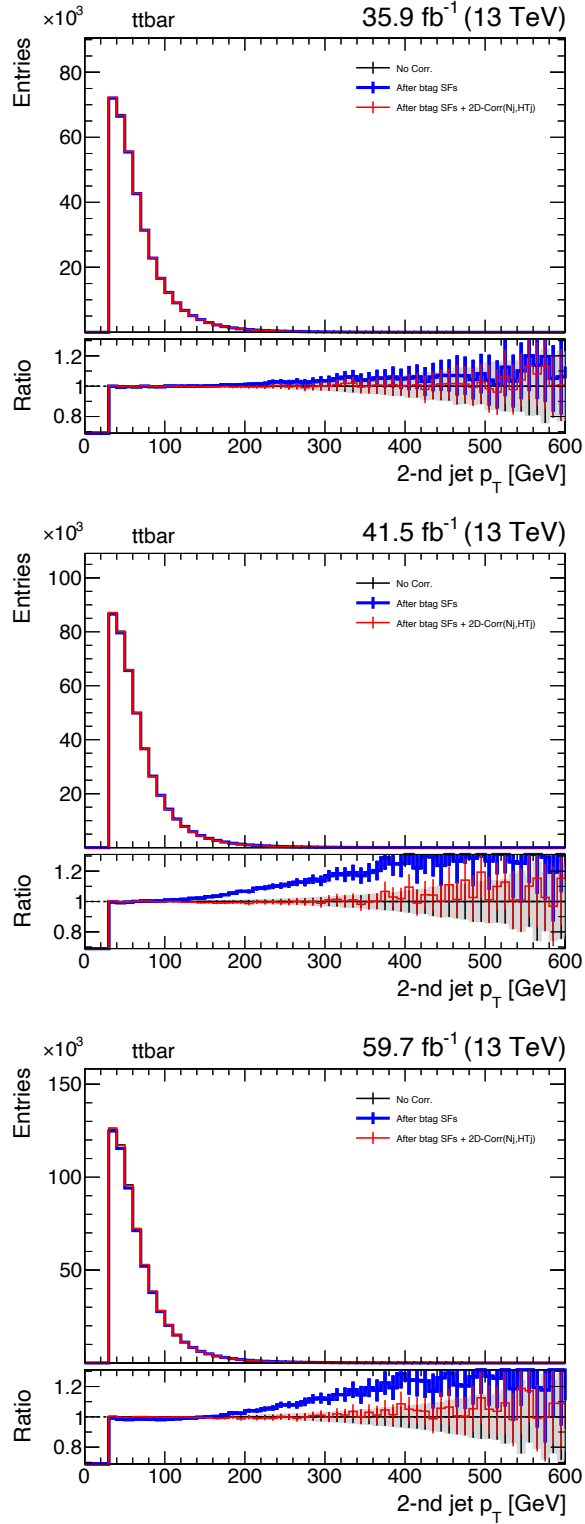


Figure E.9: Closure check for the sub-leading jet p_T for 2016, 2017 and 2018 for all $t\bar{t}$ events in the DL channel. The black line shows the distribution before applying the b -tagging SF. The blue line shows the distribution after applying the b -tagging SF and the red line shows the distribution after applying the b -tagging SF as well as the normalization scale factor. These plots show that the application of the b -tagging normalization scale factor does not change the shape of the corrected b -tagging distribution.

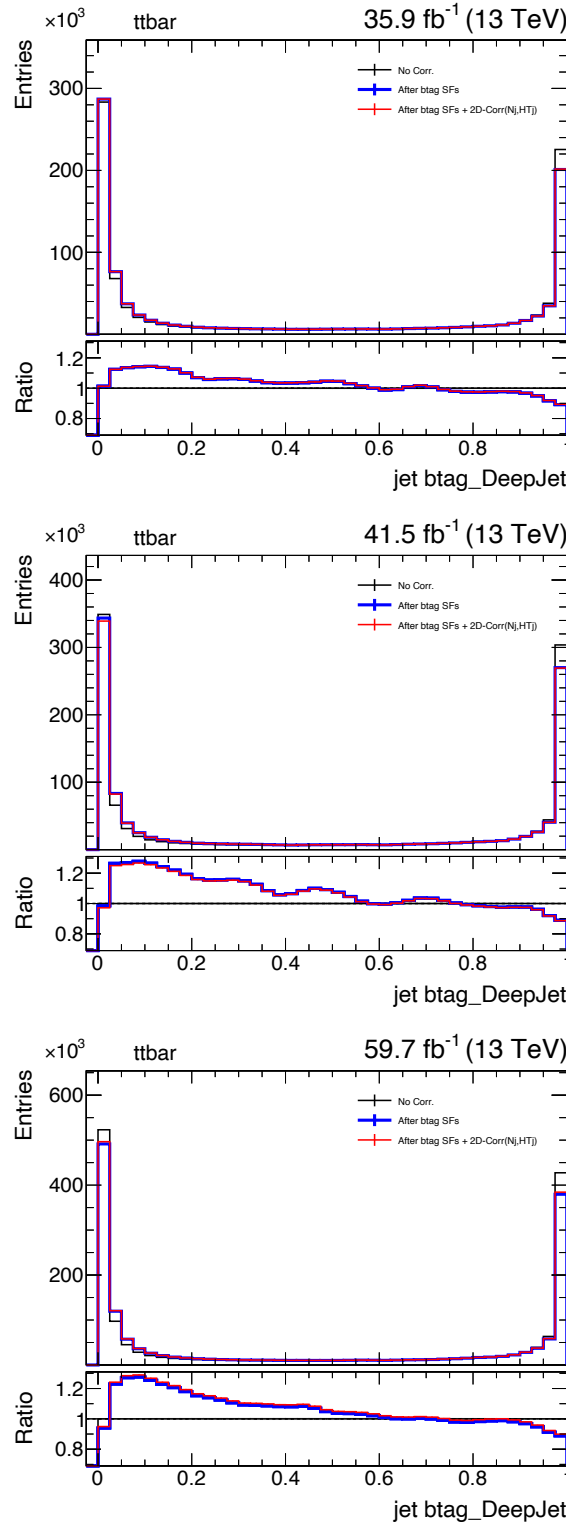


Figure E.10: Closure check for the b -tagging value of all jets for 2016, 2017 and 2018 for all $t\bar{t}$ events in the DL channel. The black line shows the distribution before applying the b -tagging SF. The blue line shows the distribution after applying the b -tagging SF and the red line shows the distribution after applying the b -tagging SF as well as the normalization scale factor. These plots show that the application of the b -tagging normalization scale factor does not change the shape of the corrected b -tagging distribution.

COMPARISON 4FS vs 5FS

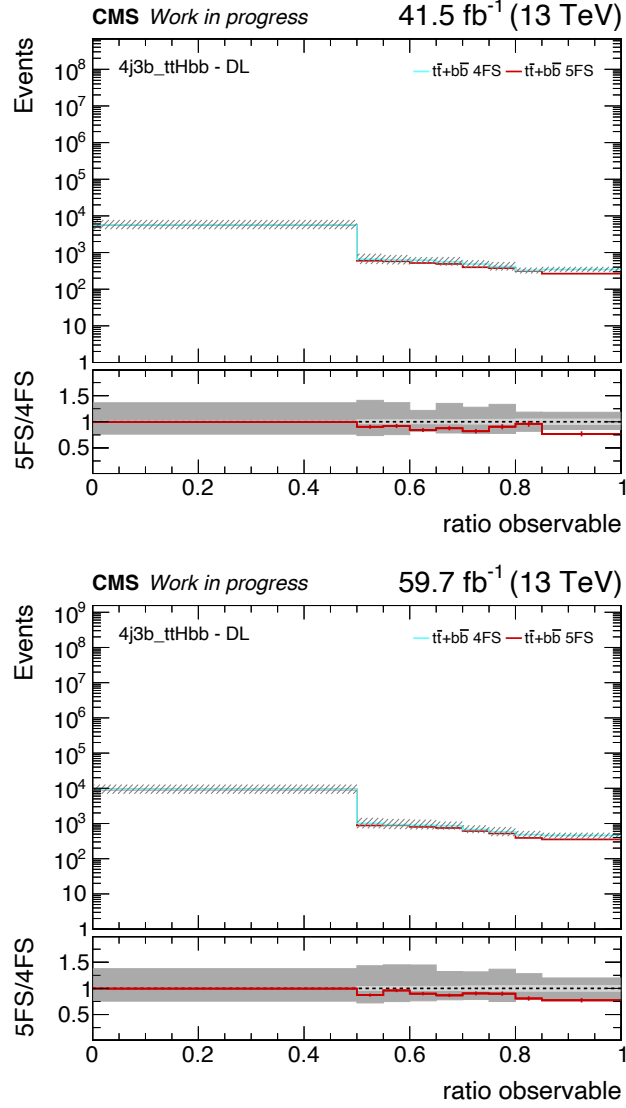


Figure F.1: Distribution of $t\bar{t} + B$ events in the ≥ 4 jets, ≥ 3 b tags category of the DL channel for 2017 (top) and 2018 (bottom) for the observable used in the final fit. The distribution obtained with the dedicated $t\bar{t} + b\bar{b}$ sample ($t\bar{t} + b\bar{b}$ 4FS, cyan) is compared to the one from the $t\bar{t}$ ($t\bar{t} + b\bar{b}$ 5FS, red) sample and the error bands contain the total systematic uncertainty.

F.1 Bias tests

The stability and robustness of the statistical model against statistical fluctuations and potentially mismodelled $t\bar{t} + b\bar{b}$ background have been assessed through tests using simulated data. The evaluations were done for the leptonic channels by generating 500 toy data sets under the signal+background hypothesis, i.e. $\mu_{t\bar{t}H}^{\text{injected}} = 1$. For each toy data set, the model was fit and evaluated based on the mean post-fit signal strength $\mu_{t\bar{t}H}$, its uncertainty, and the mean bias. The bias was calculated as the difference between the mean post-fit signal strength and the injected signal strength, divided by the uncertainty of the mean post-fit signal strength:

$$b = \frac{\mu_{t\bar{t}H} - \mu_{t\bar{t}H}^{\text{injected}}}{\sigma(\mu_{t\bar{t}H})}.$$

The toy data sets were generated using a frequentist method:

nominal A consistency check where the toy data is generated from the expected background prediction.

nominal, $t\bar{t} + B \times 1.2$ The toy data is generated from the expected background prediction but with the $t\bar{t} + B$ contribution increased by a factor of 1.2 to test the model’s flexibility to adjust for the expected underprediction of $t\bar{t} + B$.

5FS $t\bar{t} + b\bar{b}$ sample The toy data is generated from a different $t\bar{t} + B$ background prediction, the $t\bar{t}$ sample.

5FS $t\bar{t} + b\bar{b}$ sample, $t\bar{t} + B \times 1.2$ Similar to the previous test, but with an additional scaling of the $t\bar{t} + B$ contribution by 1.2.

unblinded bkg. model The toy data is generated based on the parameters from the fit to the observed data, with the SM signal injected.

The nominal statistical model is fit to these toy data sets. The results: the mean post-fit signal strength, bias values, and the background normalisations, are summarised in Tab. F.1. Examples distributions of the post-fit signal strength modifier and the $t\bar{t} + B$ background normalisation are shown in Fig. F.2.

Toy data	Fit results		
	$\langle \hat{\mu}_{t\bar{t}H} \rangle$	$\langle b \rangle$	$\langle t\bar{t} + B \text{ norm} \rangle$
nominal	1.03 ± 0.30	0.04 ± 0.94	1.01 ± 0.09
nominal, $t\bar{t} + B \times 1.2$	1.03 ± 0.32	0.07 ± 0.90	1.21 ± 0.15
5FS $t\bar{t} + b\bar{b}$ sample	1.06 ± 0.30	0.16 ± 0.91	1.03 ± 0.11
5FS $t\bar{t} + b\bar{b}$ sample, $t\bar{t} + B \times 1.2$	1.06 ± 0.32	0.14 ± 0.96	1.18 ± 0.12
unblinded bkg. model, SM signal	1.02 ± 0.32	-0.02 ± 0.92	1.19 ± 0.13

Table F.1: The results of the bias tests for the nominal fit model applied to various sets of toy data in the SL+DL channels are listed. The table shows the mean values and uncertainties (RMS of the distributions) of the best-fit signal strength modifier ($\hat{\mu}_{t\bar{t}H}$), the bias (b), and the $t\bar{t} + B$ background normalization parameter.

The results show that the fit model is stable and robust in all tested scenarios. The mean signal strength is found to be consistent with the injected value of 1, with biases lower than 1 standard deviation. When the background is sampled from the 5FS $t\bar{t} + B$ prediction, the

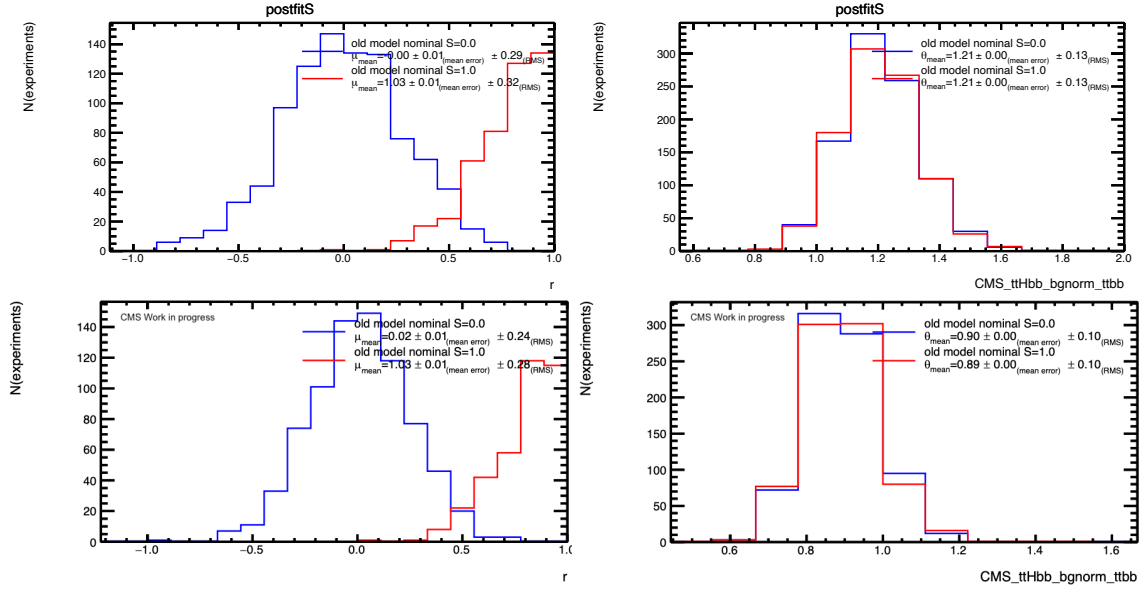


Figure F.2: Distributions of the post-fit signal strength modifier (left) and the $t\bar{t} + B$ background normalisation (right) for the signal+background (red) and background-only (blue) fits to the nominal, $t\bar{t} + B \times 1.2$ (top) and the 5FS $t\bar{t} + b\bar{b}$ sample (bottom) toy data sets described in the text.

model adjusts accordingly and compensates for the differences, leading to a post-fit signal strength close to 1 within uncertainties. The $t\bar{t} + B$ and $t\bar{t} + C$ background normalization parameters adjust as expected, with the model fitting to the injected background increase or lower background yield from the 5FS simulation. The analysis suggests that the fit model can effectively handle potential deviations in the $t\bar{t} + B$ data from the nominal prediction and compensate for the expected underprediction of the $t\bar{t} + B$ cross section.

INPUT VARIABLES TO ARTIFICIAL NEURAL NETWORKS

The set of variables that are potential neural networks inputs are presented here, subdivided into groups:

- **Kinematic properties of single object quantities.** Usually referred to as low-level variables, they are the transverse momenta and pseudorapidities of the two leptons and the first four jets (sorted by transverse momentum in descending order).
- **Combined variables of multiple physics objects.** The scalar sum of all transverse jet momenta H_T , the minimum and maximum distances $\Delta R = \Delta\eta^2 + \Delta\phi^2$ of all jet pairs, the minimum distance ΔR between a jet and the leading lepton, the average distance $\Delta\eta$ between jet pairs, are some example of variables entering this category. They are computed for different collections of jets: all selected jets, b-tagged and non b-tagged jets.
- **Mass variables.** Computed for pairs of particles using the four-vector momenta with $m = \sqrt{(E_1 + E_2)^2 - (p_1 + p_2)^2}$: from the average mass of all pairs of selected and b-tagged jets, the mass of the leading and sub-leading leptons, the mass $m_{\ell\ell}$ of the two leptons, the closest b-tagged jets by ΔR , the mass of the b-tagged and not jet pair closest to 125 GeV.
- **b-tagging related variables.** Information from jet b-tagging discriminants of the first four jets.
- **Event shape variables.** Usually single quantity that represent geometrical correlations among four-vectors of final state particles. The second, third, and fourth Fox-Wolfram moments, i.e., compositions of spherical harmonics, are calculated for all jets [298]. Further quantities are determined separately for all selected jets, b-tagged and non b-tagged jets. The centrality describes the ratio between the sums of transverse momenta and energies of involved jets, $\sum_j p_{T,j} / \sum_j E_j$. Sphericity S , transverse sphericity, S_T , and aplanarity A are defined through the eigenvalues λ_1 , λ_2 , and λ_3 of the momentum tensor M as [299]:

$$M = \frac{1}{\sum_j p_j^2} \sum_j \vec{p}_j \cdot \vec{p}_j^T \quad (\text{G.1})$$

$$\Rightarrow S = \frac{3}{2}(\lambda_2 + \lambda_3), \quad S_T = \frac{2\lambda_2}{\lambda_1 + \lambda_2}, \quad A = \frac{3}{2}\lambda_1. \quad (\text{G.2})$$

- **b tagging likelihood ratio (BLR) discriminant.** To identify events compatible with the presence of four b jets, a likelihood ratio criterion is used to test the hypothesis that four jets or two jets in the event originate from B hadron decays [126, 126, 300]. The likelihoods are based on the expected DEEJET b tagging discriminant probability densities from simulation, depending on the jet p_T and η . The four jets that maximize the ratio $\mathcal{L}_4 / (\mathcal{L}_2 + \mathcal{L}_4)$ are selected for associations to bottom quarks within permutations,

with $\mathcal{L}_4$ ($\mathcal{L}_2$) being the likelihood that express the compatibility of the four jets with the hypothesis of four (two) b jets. The maximum value yields a well separating variable, but its range of values is rather broad so also the transformed ratio $\log(b/1-b)$ is investigated.

- **Matrix element method (MEM) discriminant.** It has been intensively used in previous iterations of this analysis to compare and separate the contributions of $t\bar{t}H$ signals from the $t\bar{t} + b\bar{b}$ background events [121, 127]. It is constructed as a ratio of two probability density functions $w_{S/B}(x)$, describing the likelihood that the reconstructed event originate from a $t\bar{t}H$ or at $t\bar{t} + b\bar{b}$ event:

$$P_{S/B} = \frac{w_S(x)}{w_S(x) + k \cdot w_B(x)} \quad (\text{G.3})$$

The constant k is set to 0.1 as a result of an optimization procedure to maximise the signal to background discrimination. The four-vector x of all selected finale-state objects enter the computation of the densities $w_{S/B}(x)$; detector resolution effects are taken into account via transfer functions. This procedure is repeated for all possible permutations coming from the association at generator level of reconstructed jets and partons. Only the permutation with the highest probability is selected and the correspondent probability density values w_S and w_B used to compute the MEM discriminant $P_{S/B}$ [121, 126, 301]. The compatibility with the signal hypothesis corresponds with high discriminant values. To save computing time, the MEM discriminant is computed only for events in relevant phase space. In this search of $t\bar{t}H, H \rightarrow b\bar{b}$ in the dilepton channel, this corresponds to ≥ 4 jets, ≥ 3 b tags events passing the baseline selections.

The complete list of variables considered as initial set is in Table G.1.

A selection of the selected and validated input variable distributions is shown in the following, for each of the two analysis categories.

Variables		
N_{jets} $N_{\text{b}}^{\text{loose}}$ p_{T} (leading lepton) η (second lepton) p_{T} (jet 1) p_{T} (jet 4) η (jet 1) η (jet 4) p_{T} (b-jet 1) p_{T} (b-jet 4) η (b-jet 1) η (b-jet 4) b tag (jet 1) b tag (jet 4) b tag (jet 1) b tag (jet 4) $p_{\text{T}}^{\text{miss}}$ <b tag discriminator> ^{untagged} sphericity (tags) circularity (jets) isotropy (tags) D (jets) trans-sphericity (tags) isotropy (jets) centrality (tags) H_1^j H_2^b H_4^j R_1^b R_3^j R_4^b $\Delta R_{\text{j,j}}^{\text{avg}}$ $\Delta R_{\text{b,b}}^{\text{min}}$ $m_{\text{lep,b}}^{\text{min}\Delta R}$ $m_{\text{j,b}}^{\text{min}\Delta R}$ $m_{\text{b,b}}^{\text{closest to 125}}$ $p_{\text{T};\text{j,j}}^{\text{min}\Delta R}$ $\Delta\eta_{\text{j,j}}^{\text{max}}$	MEM $N_{\text{b}}^{\text{medium}}$ η (leading lepton) H_{T} p_{T} (jet 2) p_{T} (jet 5) η (jet 2) η (jet 5) p_{T} (b-jet 2) p_{T} (b-jet 5) η (b-jet 2) η (b-jet 5) b tag (jet 2) b tag (jet 5) b tag (jet 2) b tag (jet 5) ϕ^{miss} <b tag discriminator> ^{tagged} aplanarity (jets) circularity (tags) C (jets) D (tags) circularity (jets) isotropy (tags) H_0^j H_1^b H_3^j H_4^b R_2^j R_3^b $\Delta R_{\text{j,j}}^{\text{max}}$ $\Delta R_{\text{j,b}}^{\text{avg}}$ $\Delta R_{\text{lep,b}}^{\text{min}}$ $m_{\text{j,j}}^{\text{min}\Delta R}$ $m_{\text{b,b}}^{\text{max } m}$ $m_{\text{j,b}}^{\text{closest to 125}}$ $p_{\text{T};\text{b,b}}^{\text{min}\Delta R}$ $\Delta\eta_{\text{b,b}}^{\text{max}}$	BLR $N_{\text{b}}^{\text{tight}}$ p_{T} (second lepton) H_{T}^b p_{T} (jet 3) p_{T} (jet 6) η (jet 3) η (jet 6) p_{T} (b-jet 3) p_{T} (b-jet 6) η (b-jet 3) η (b-jet 6) b tag (jet 3) b tag (jet 6) b tag (jet 3) b tag (jet 6) $N_{\text{Higgs}}^{\text{jj}}$ sphericity (jets) aplanarity (tags) isotropy (jets) C (tags) trans-sphericity (jets) circularity (tags) centrality (jets) H_0^b H_2^j H_3^b R_1^j R_2^b R_4^j $\Delta R_{\text{b,b}}^{\text{avg}}$ $\Delta R_{\text{j,j}}^{\text{min}}$ $m_{\text{j}}^{\text{avg}}$ $m_{\text{b,b}}^{\text{min}\Delta R}$ $m_{\text{b,b}}^{\text{avg}}$ $m_{\text{j,b,b}}^{\text{max } p_{\text{T}}}$ $p_{\text{T};\text{j,b}}^{\text{min}\Delta R}$ $m_{\text{j,b,b}}^{\text{median}}$

Table G.1: Total set of input variables considered for the artificial neural network training.

3 jets, ≥ 3 b tags category

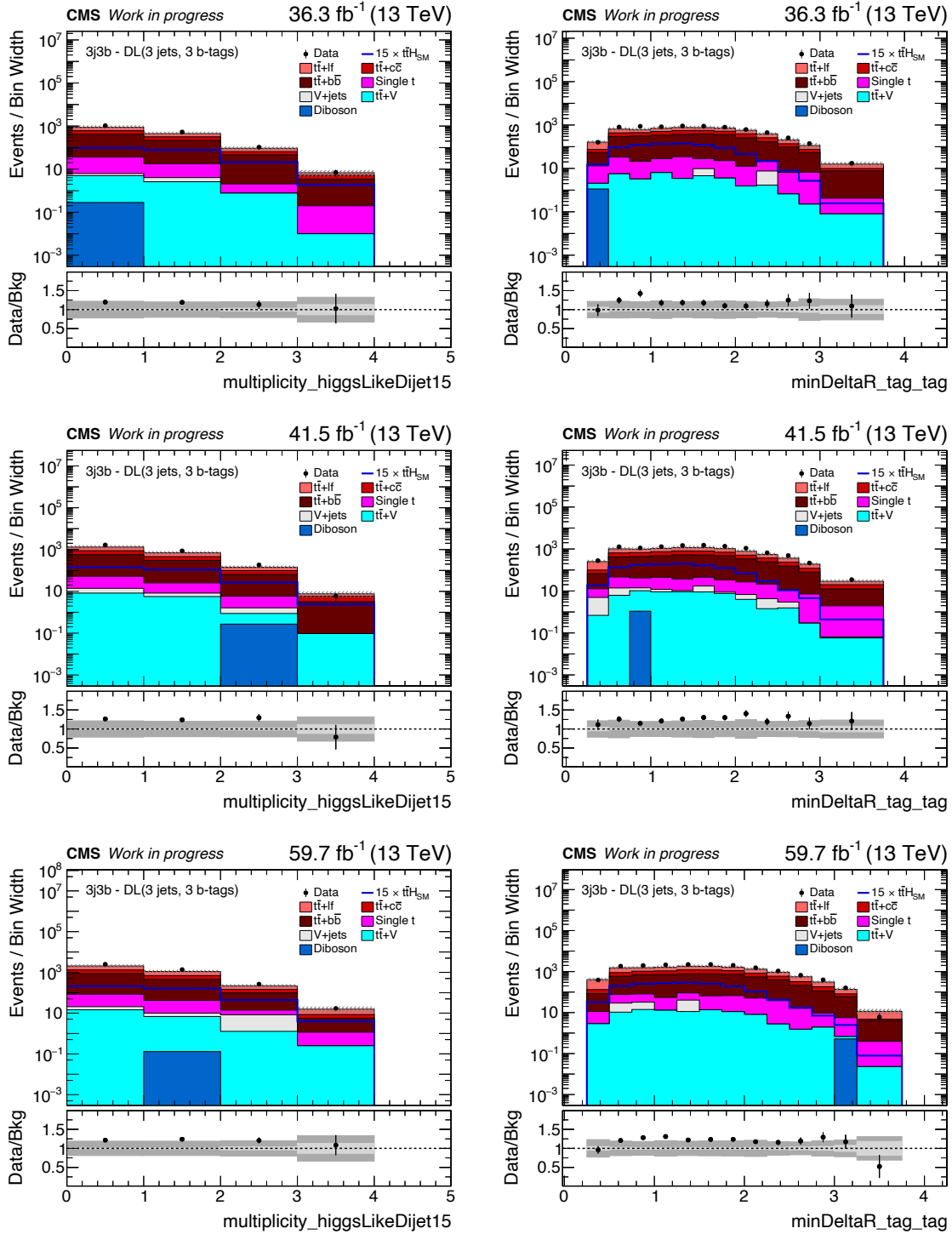


Figure G.1: Input Feature for DL ANN in all three years in the region with 3 jets, ≥ 3 b tags.

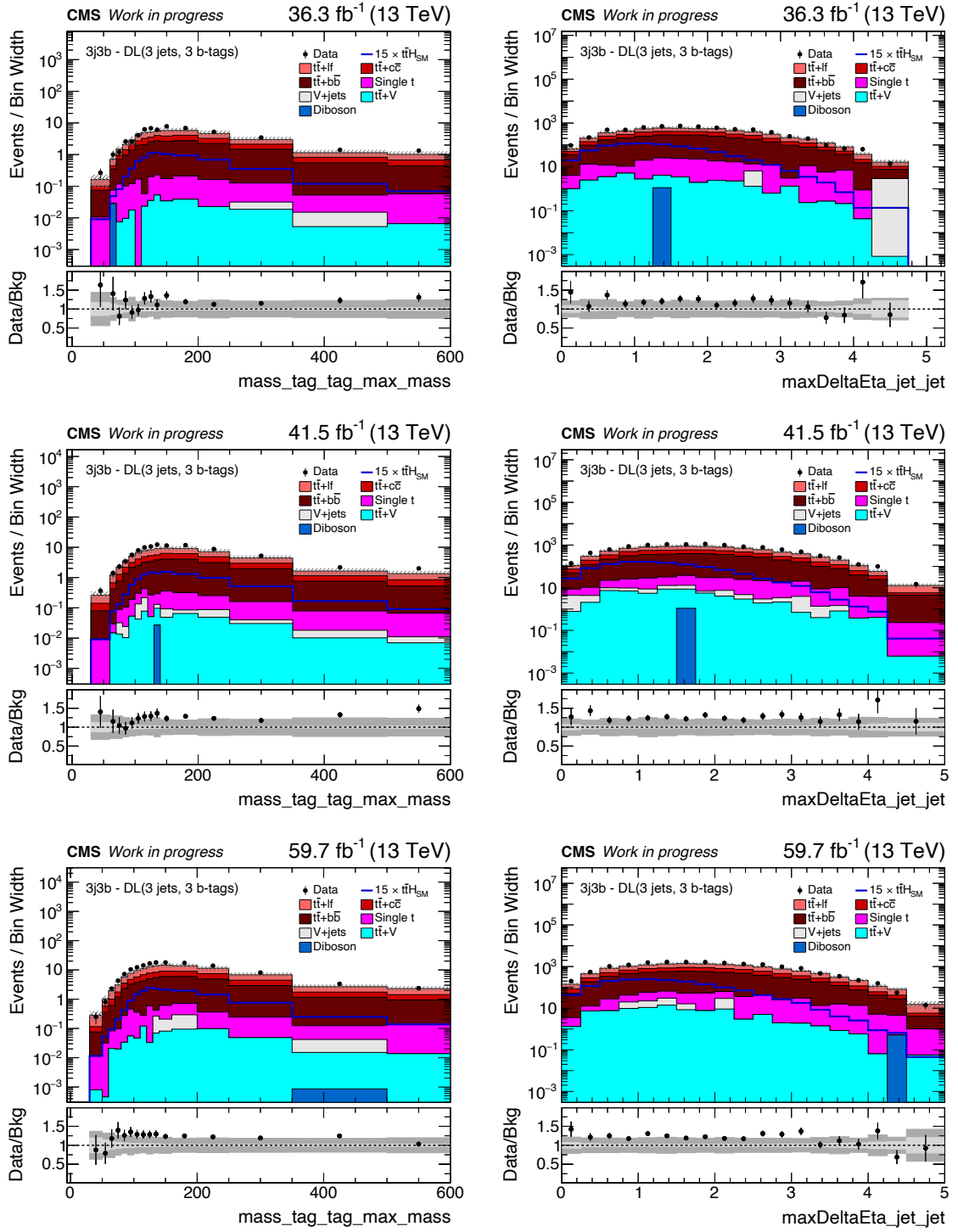


Figure G.2: Input Feature for DL ANN in all three years in the region with 3 jets, ≥ 3 b tags.

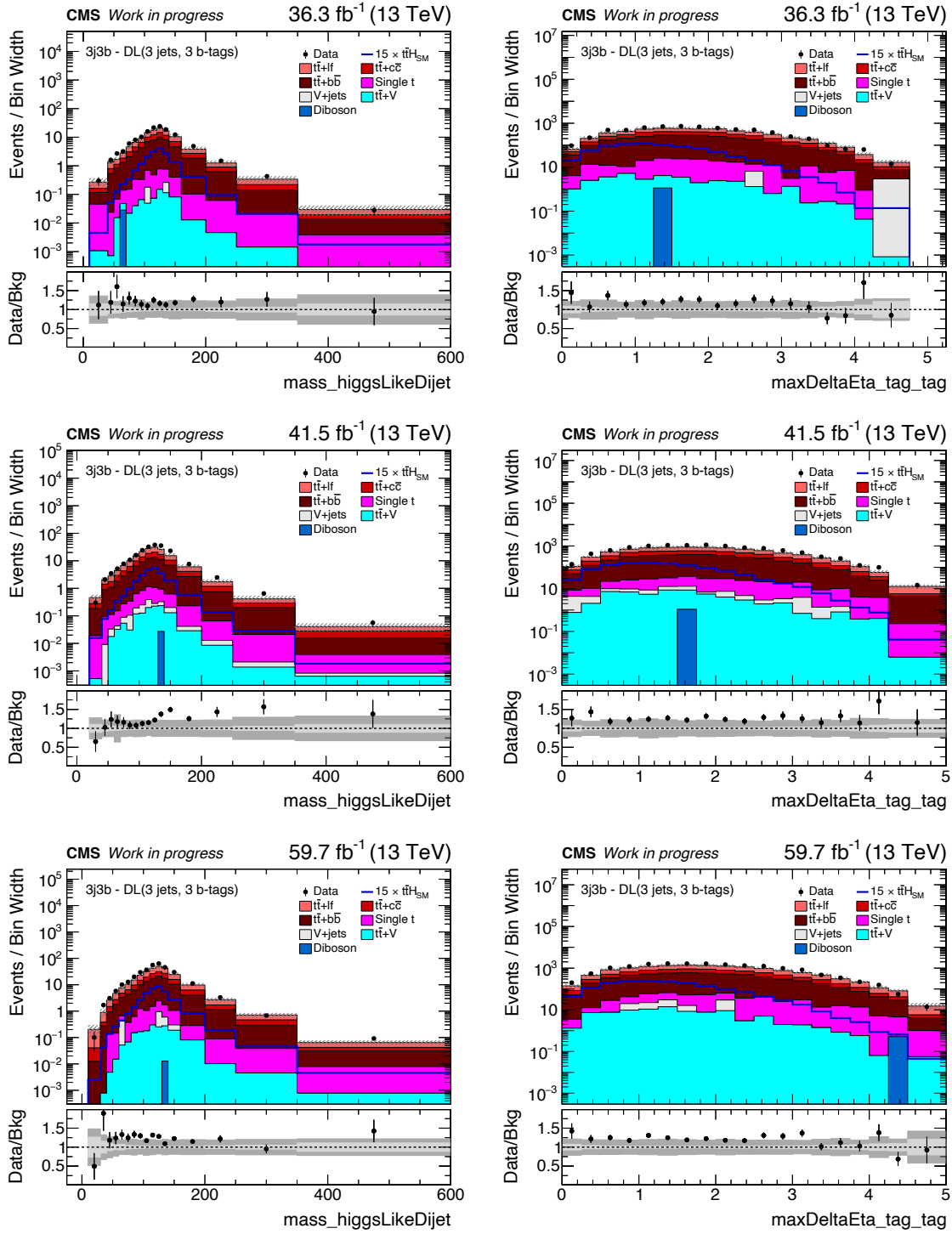


Figure G.3: Input Feature for DL ANN in all three years in the region with 3 jets and 3 b-tags.

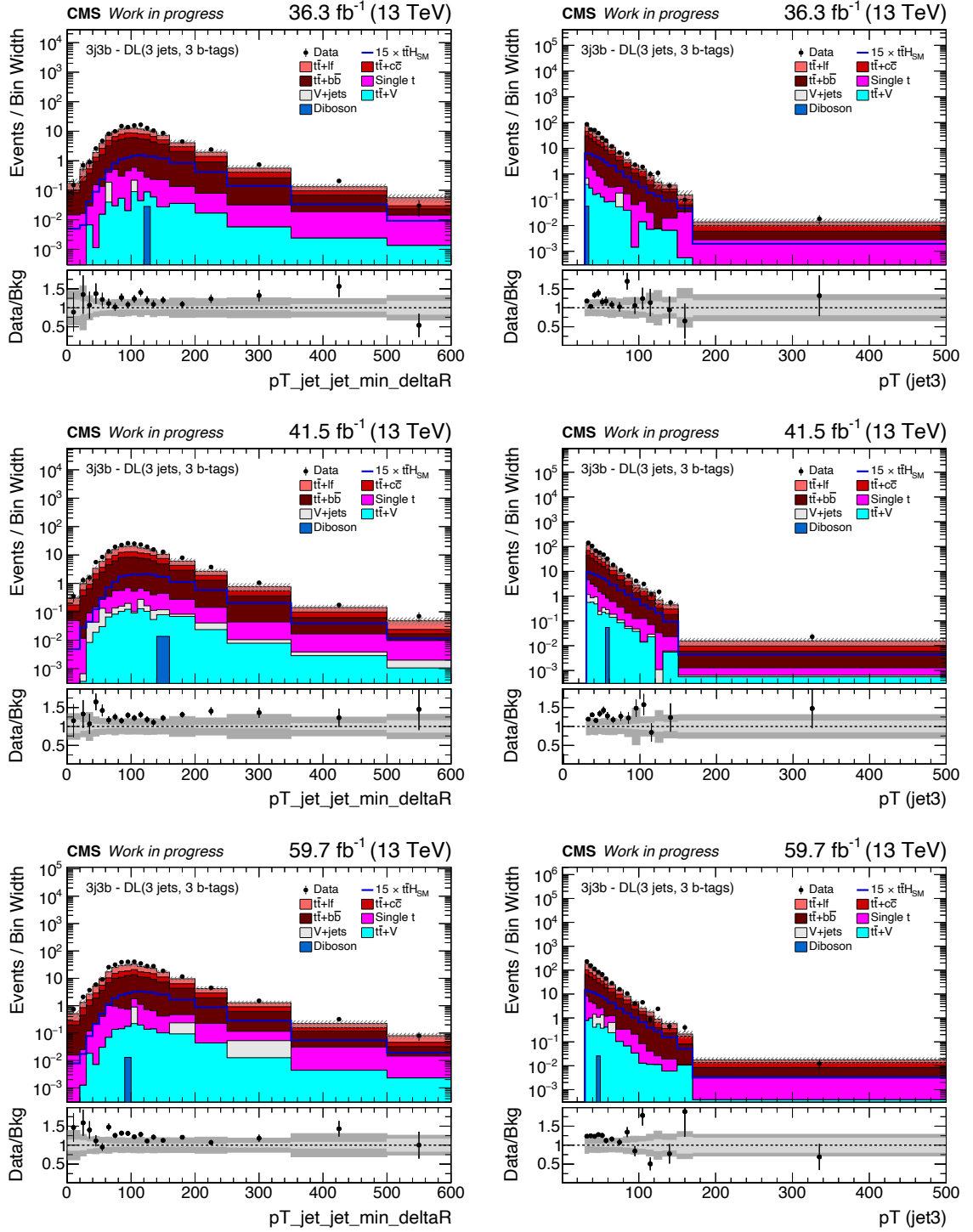


Figure G.4: Input Feature for DL ANN in all three years in the region with 3 jets and 3 b-tags.

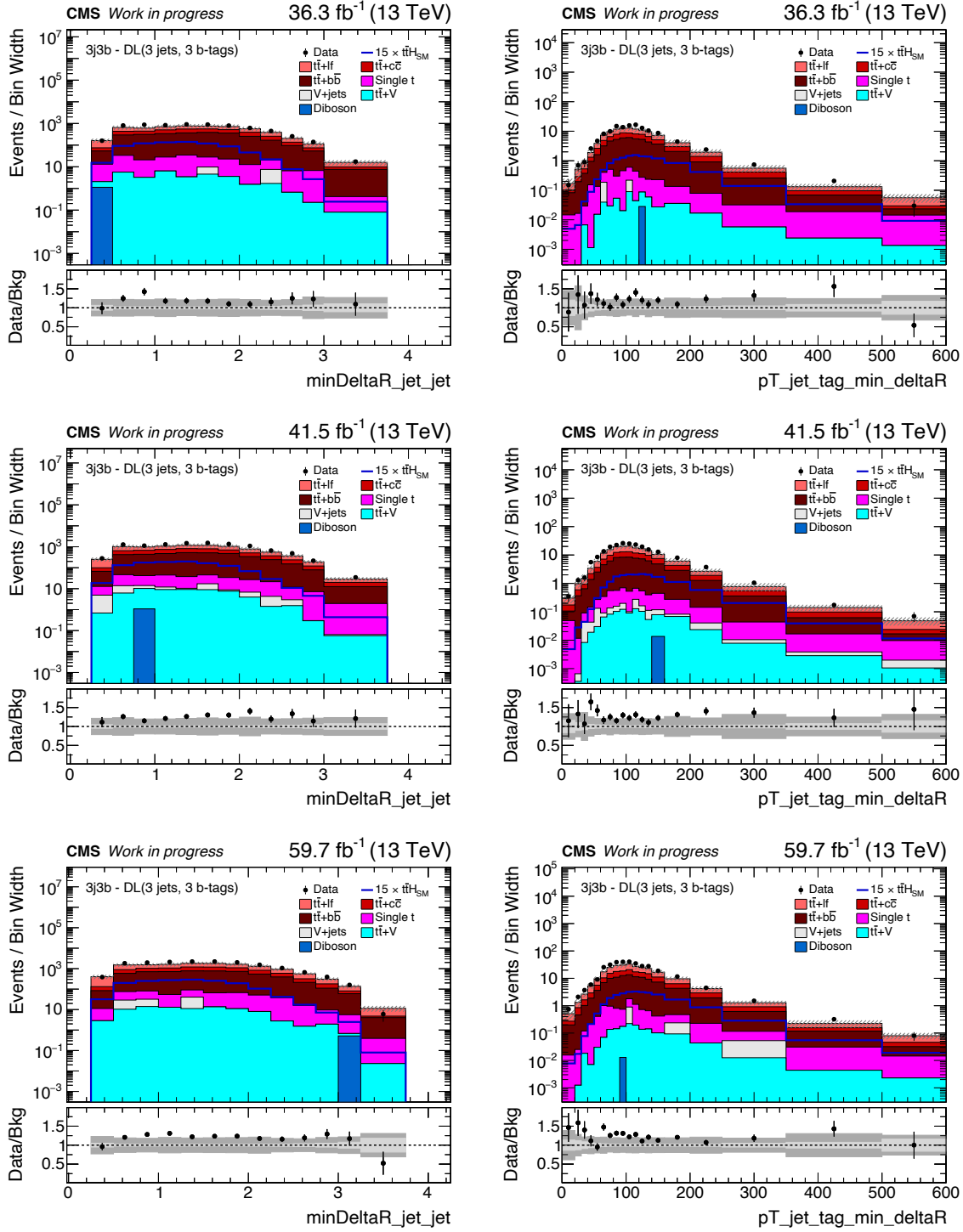


Figure G.5: Input Feature for DL ANN in all three years in the region with 3 jets and 3 b-tags.

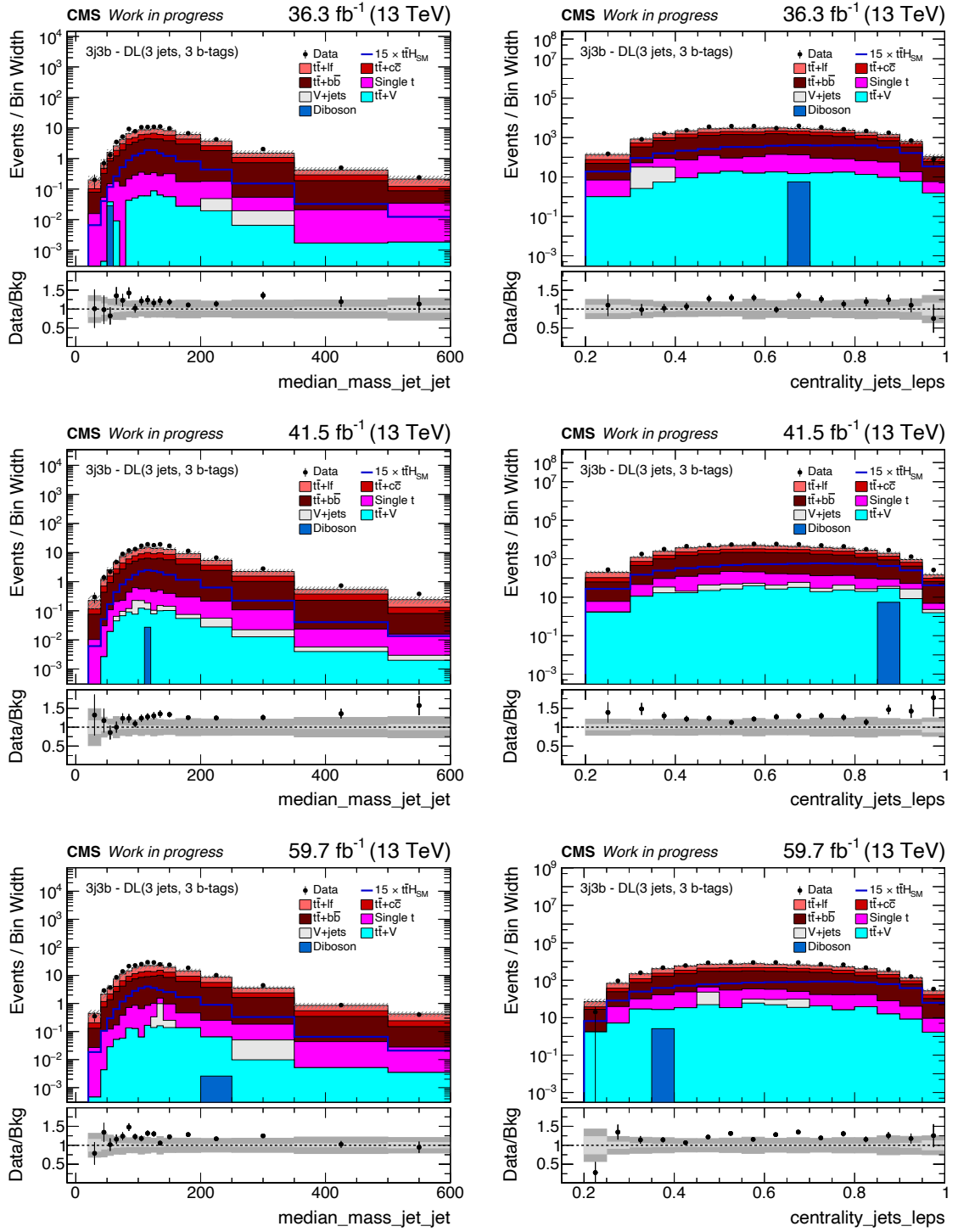


Figure G.6: Input Feature for DL ANN in all three years in the region with 3 jets and 3 b-tags.

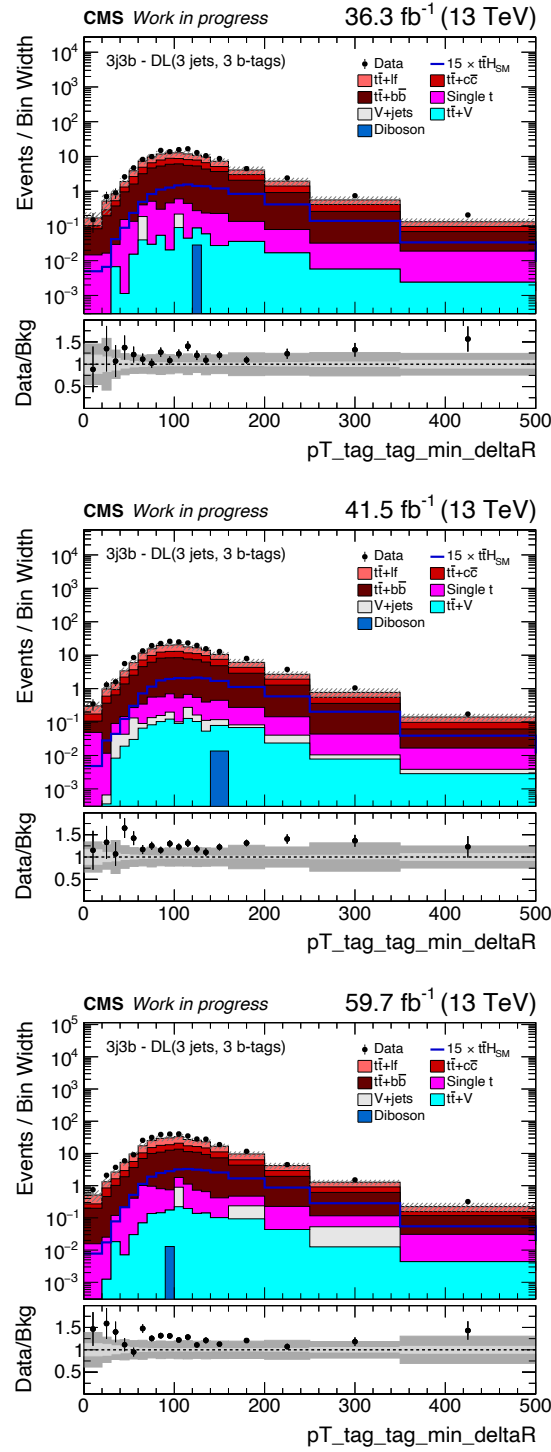


Figure G.7: Input Feature for DL ANN in all three years in the region with 3 jets and 3 b-tags.

≥ 4 jets, ≥ 3 b tags category

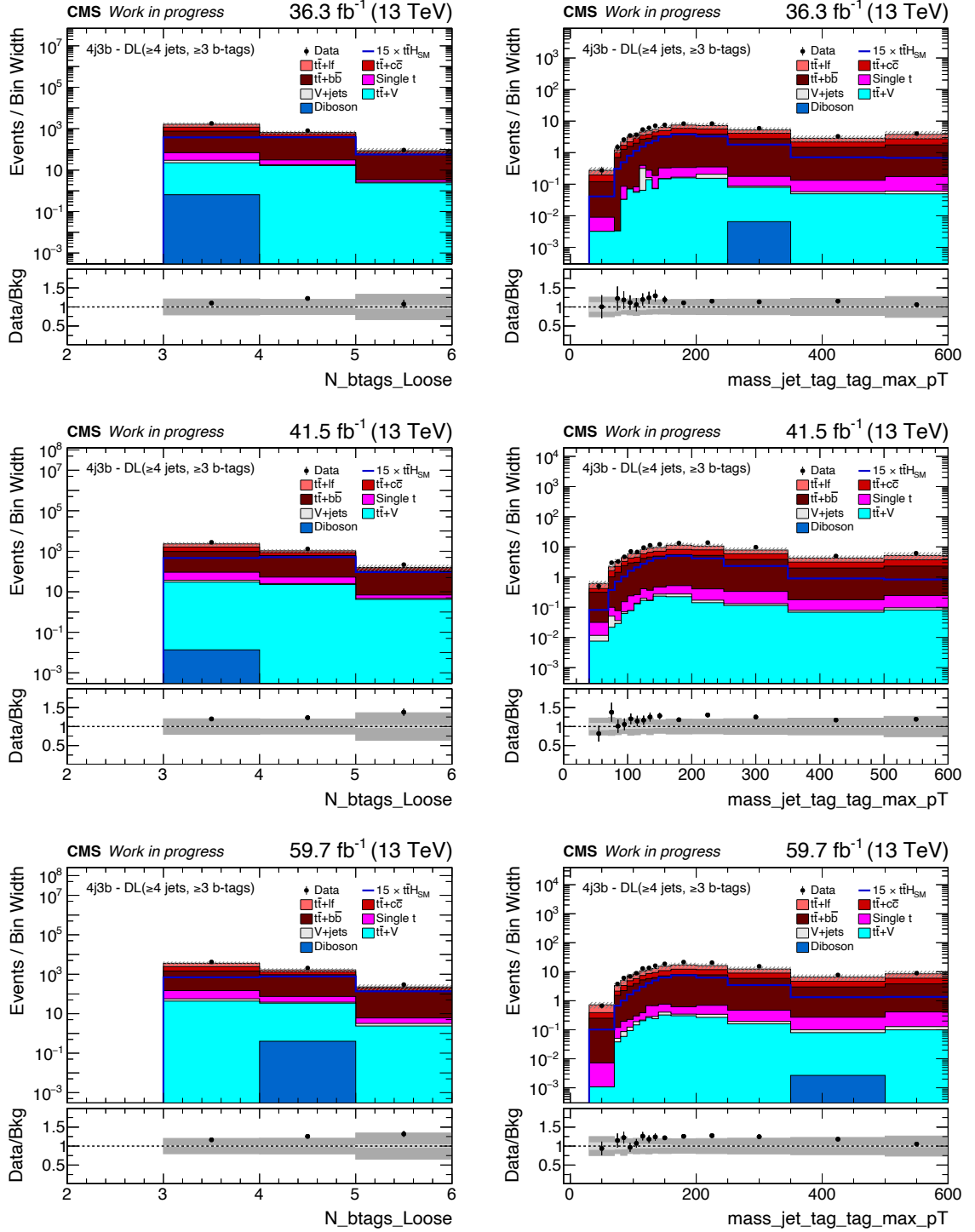


Figure G.8: Input Feature for DL ANN in all three years in the region ≥ 4 jets, ≥ 3 b tags.

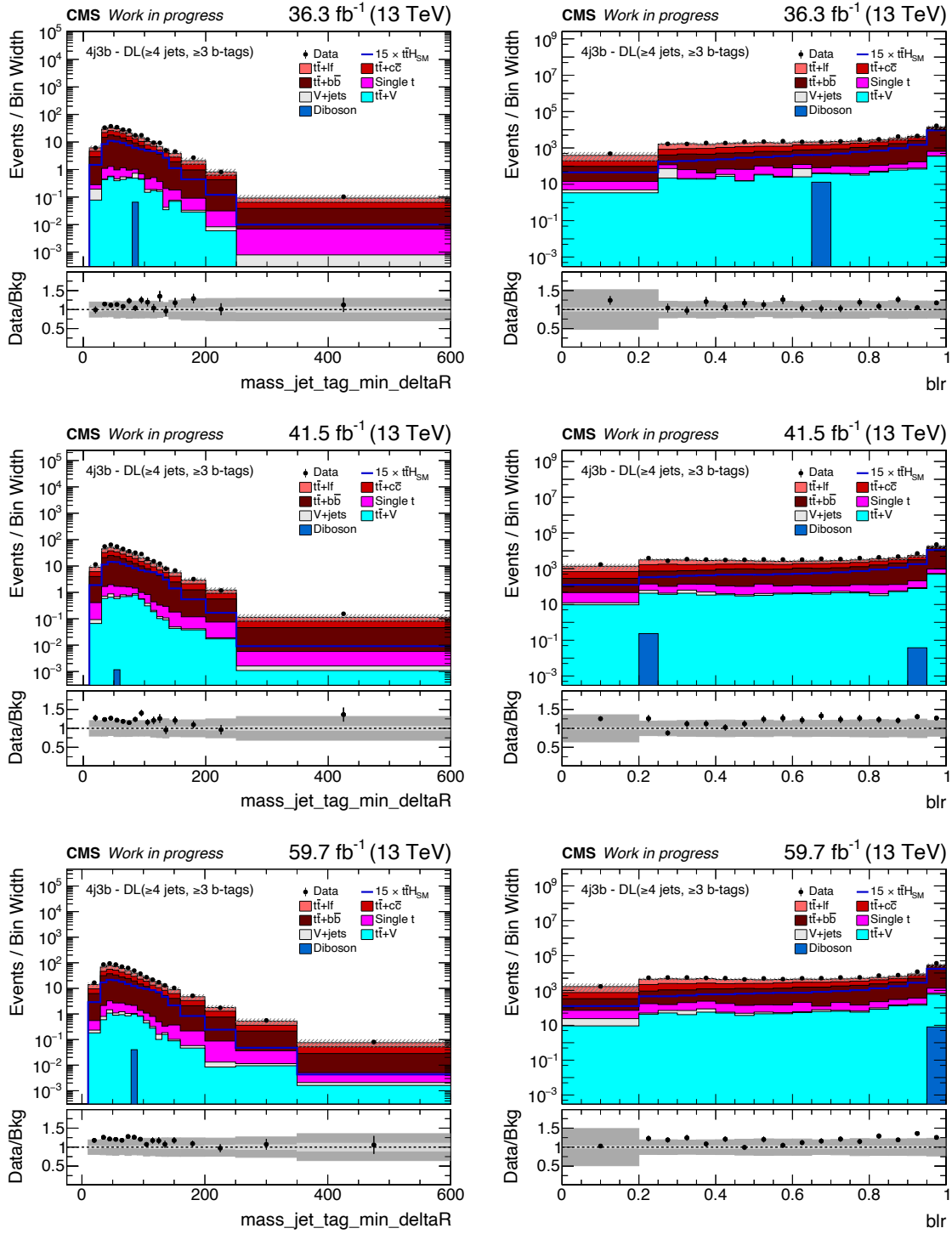


Figure G.9: Input Feature for DL ANN in all three years in the region ≥ 4 jets, ≥ 3 b tags.

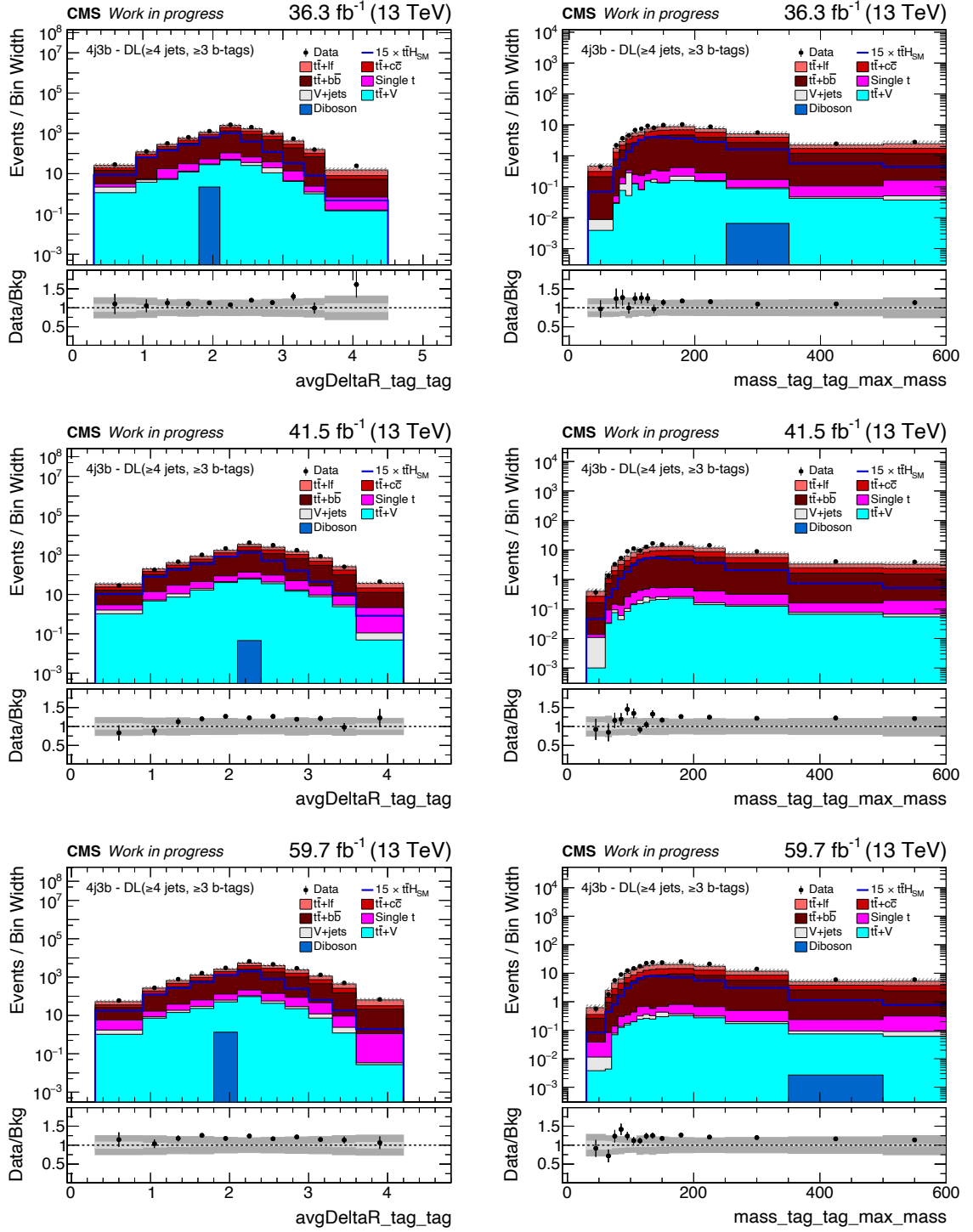


Figure G.10: Input Feature for DL ANN in all three years in the region ≥ 4 jets, ≥ 3 b tags.

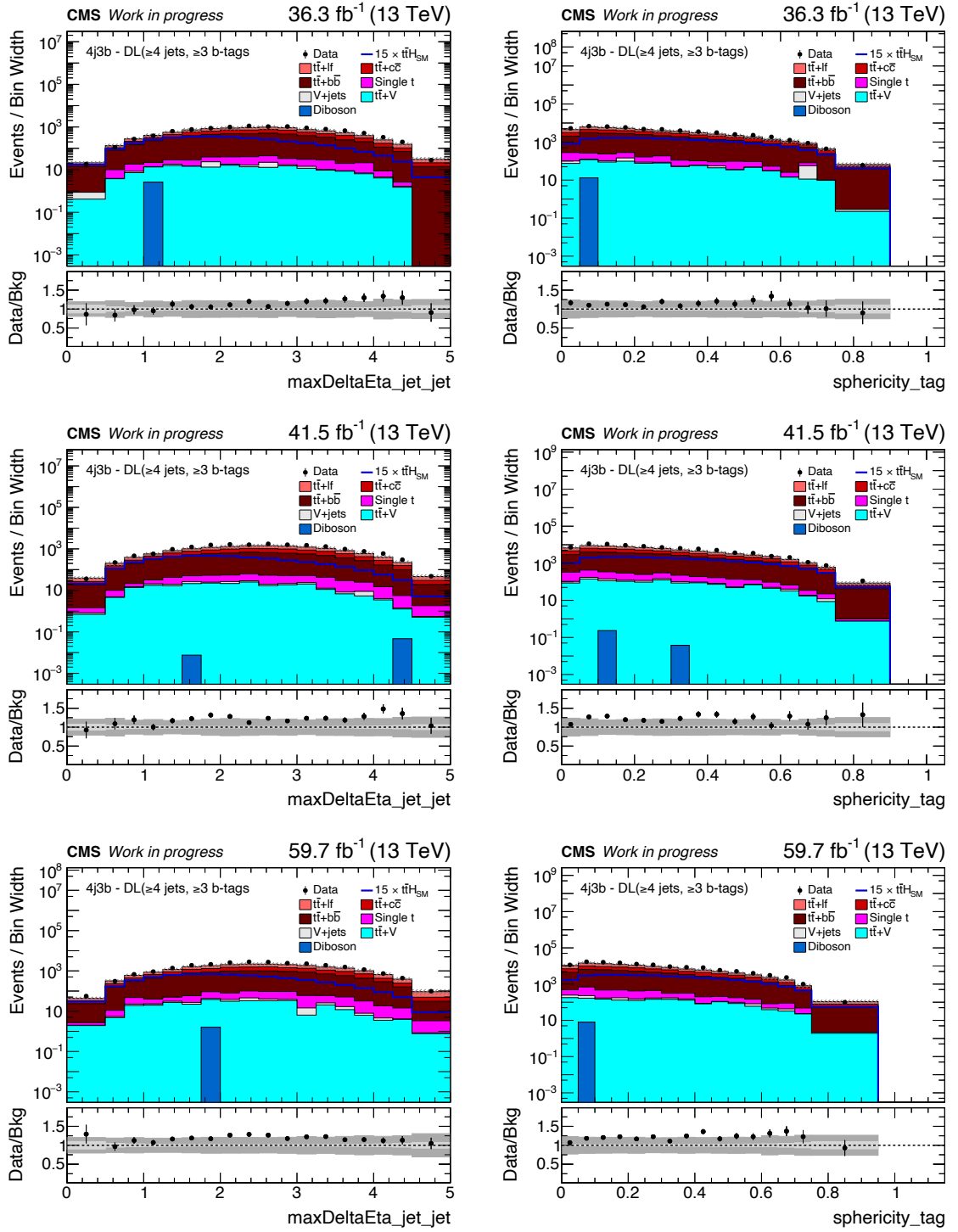


Figure G.12: Input Feature for DL ANN in all three years in the region $\geq 4 \text{ jets}, \geq 3 \text{ b tags}$.

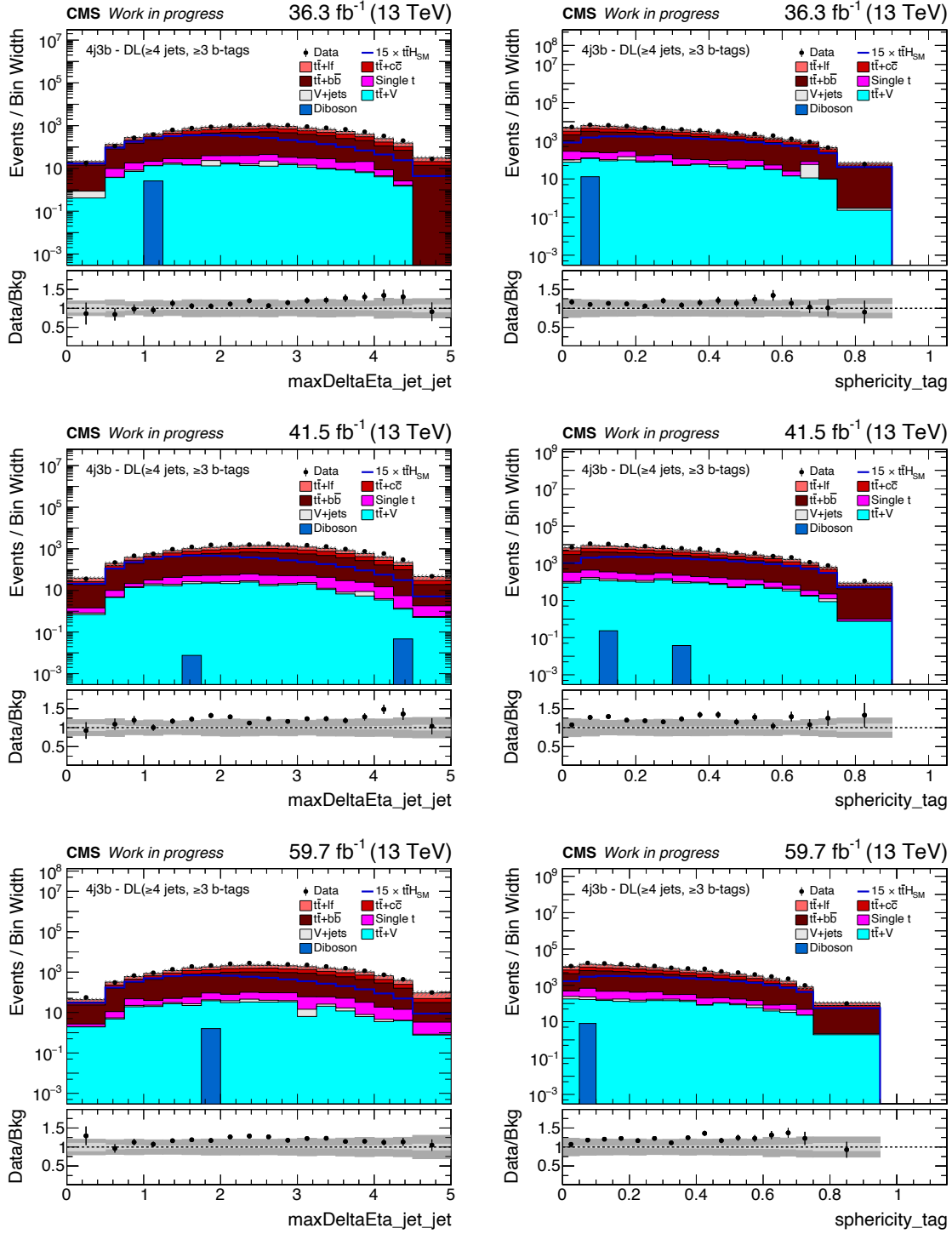


Figure G.13: Input Feature for DL ANN in all three years in the region ≥ 4 jets, ≥ 3 b tags.

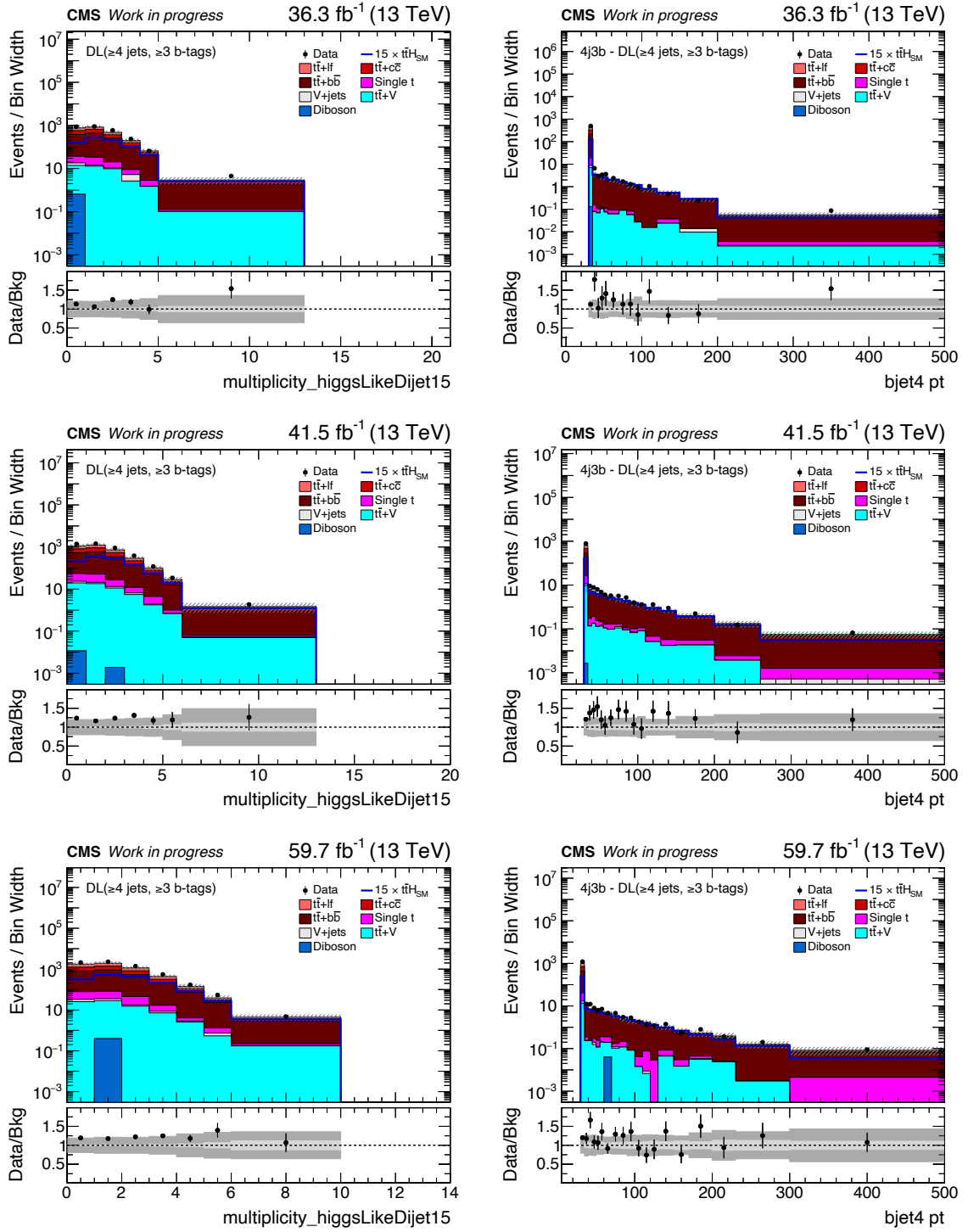


Figure G.14: Input Feature for DL ANN in all three years in the region ≥ 4 jets, ≥ 3 b tags.

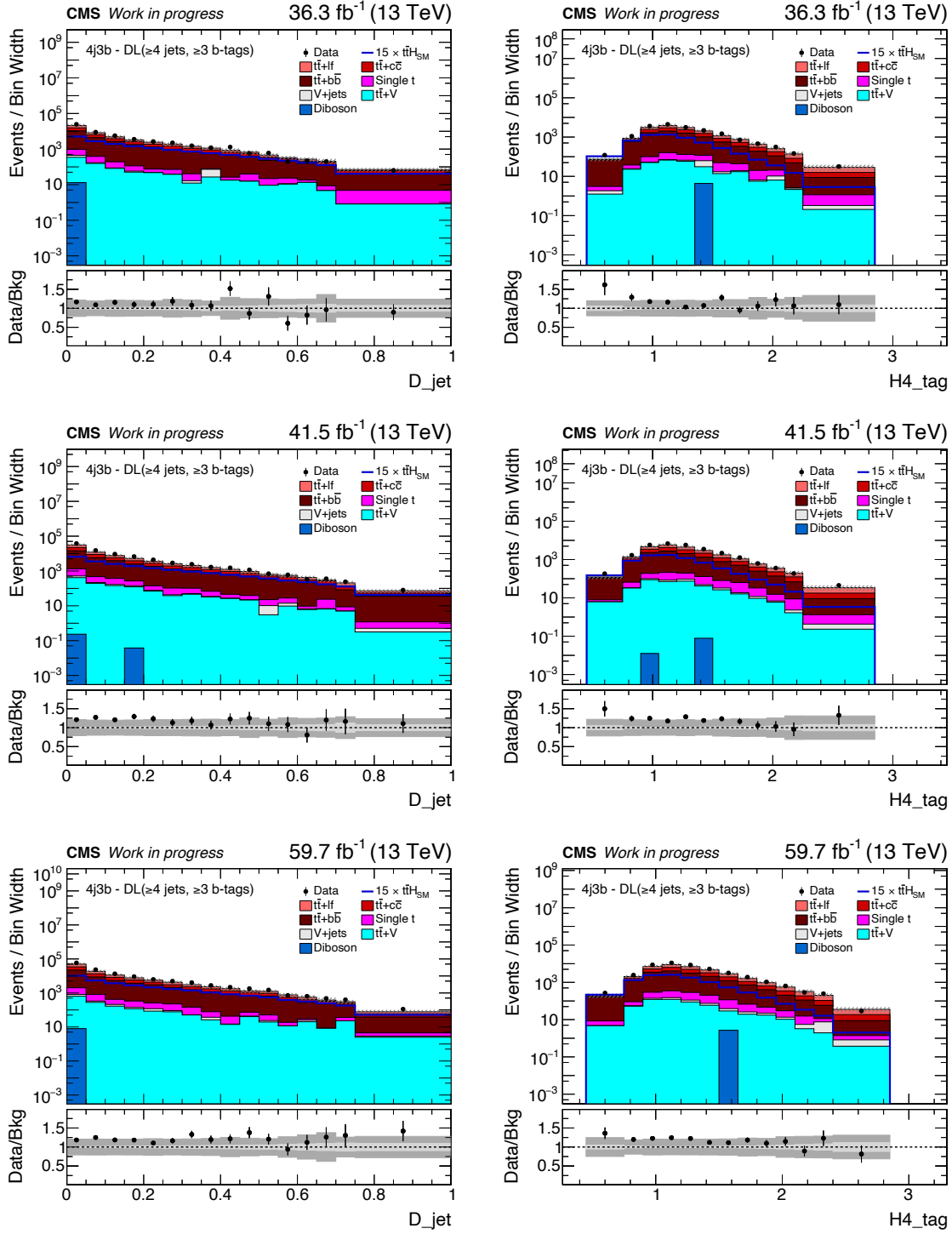


Figure G.15: Input Feature for DL ANN in all three years in the region ≥ 4 jets, ≥ 3 b tags.

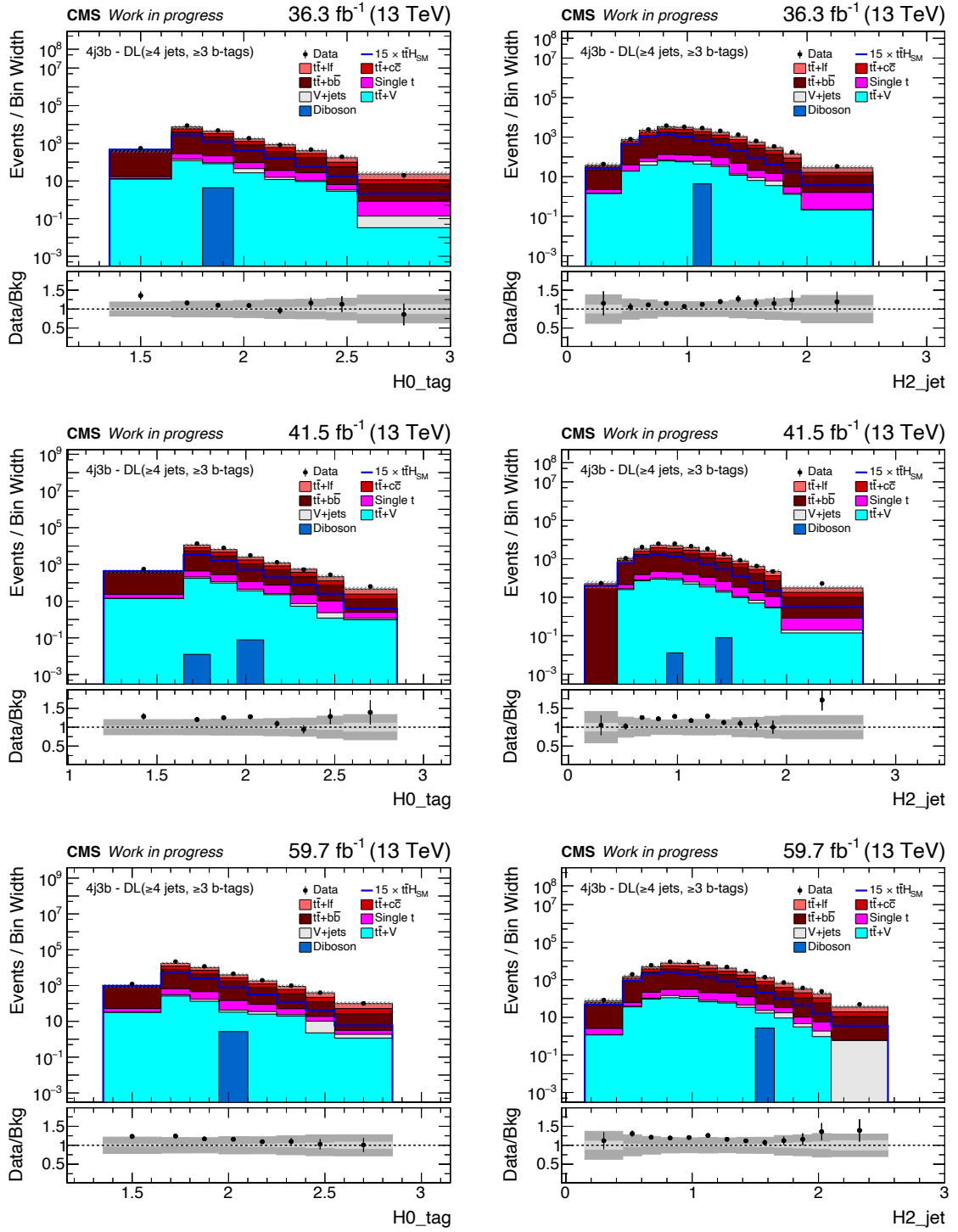


Figure G.16: Input Feature for DL ANN in all three years in the region $\geq 4 jets, \geq 3 b$ tags.

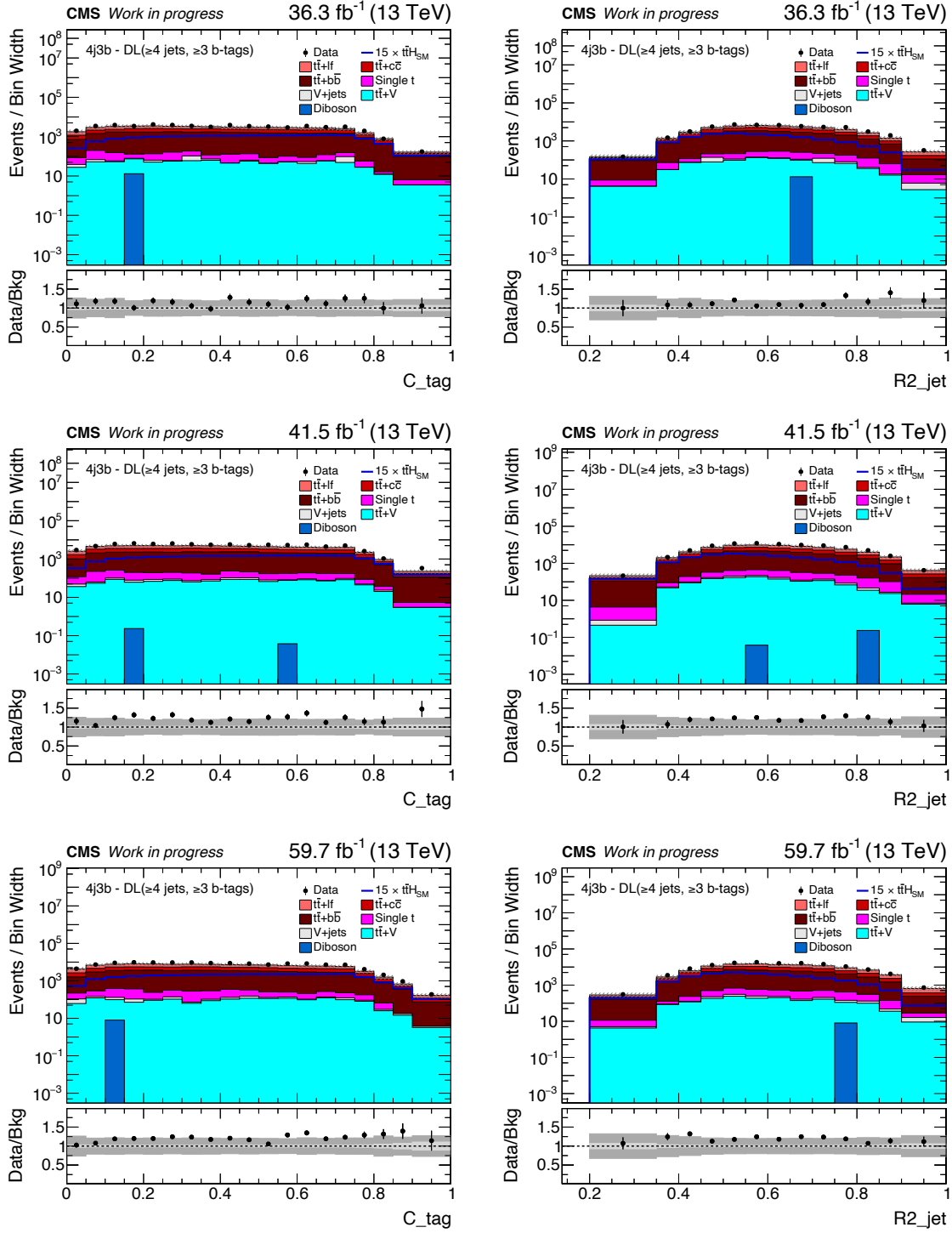


Figure G.17: Input Feature for DL ANN in all three years in the region ≥ 4 jets, ≥ 3 b tags.

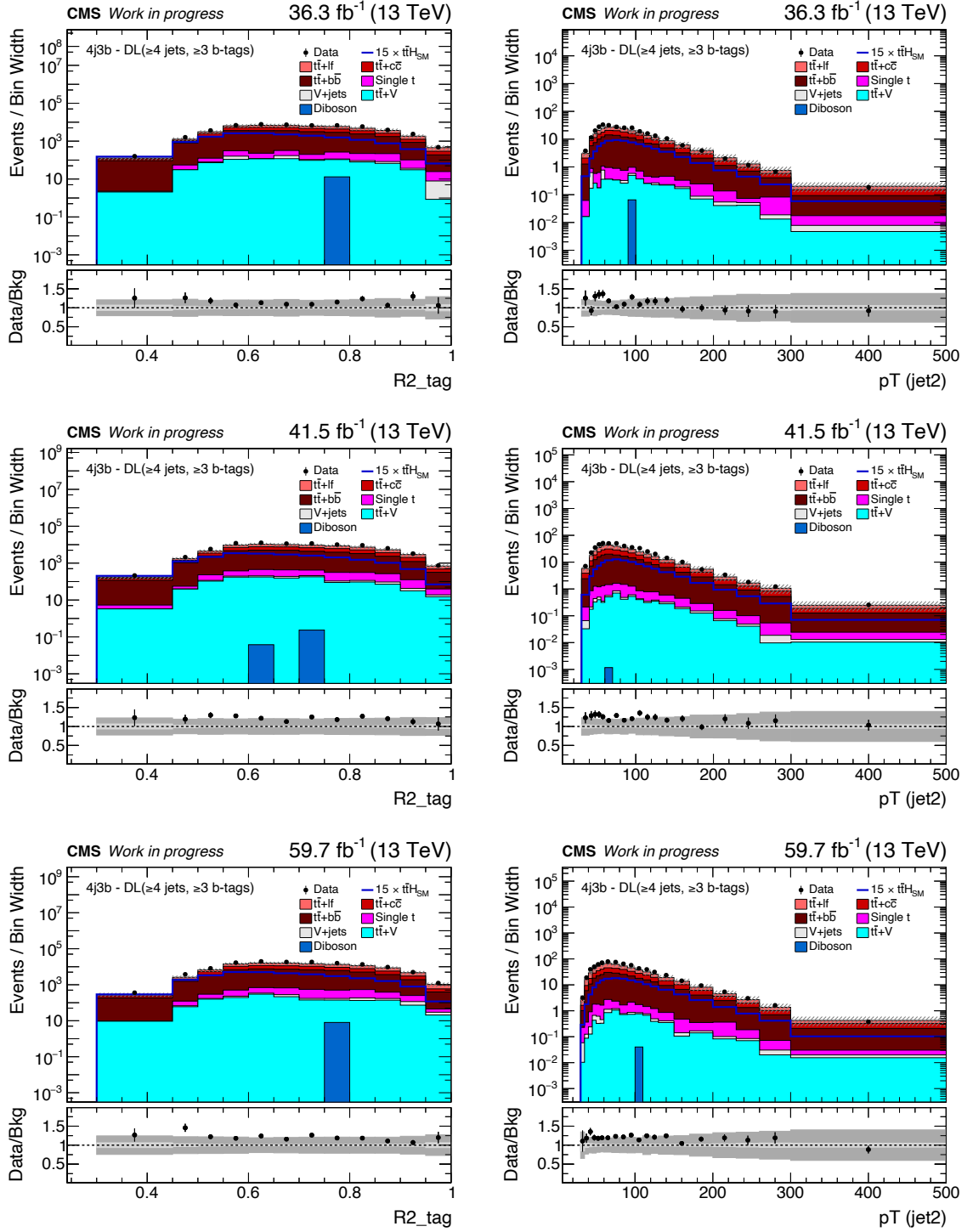


Figure G.18: Input Feature for DL ANN in all three years in the region ≥ 4 jets, ≥ 3 b tags.

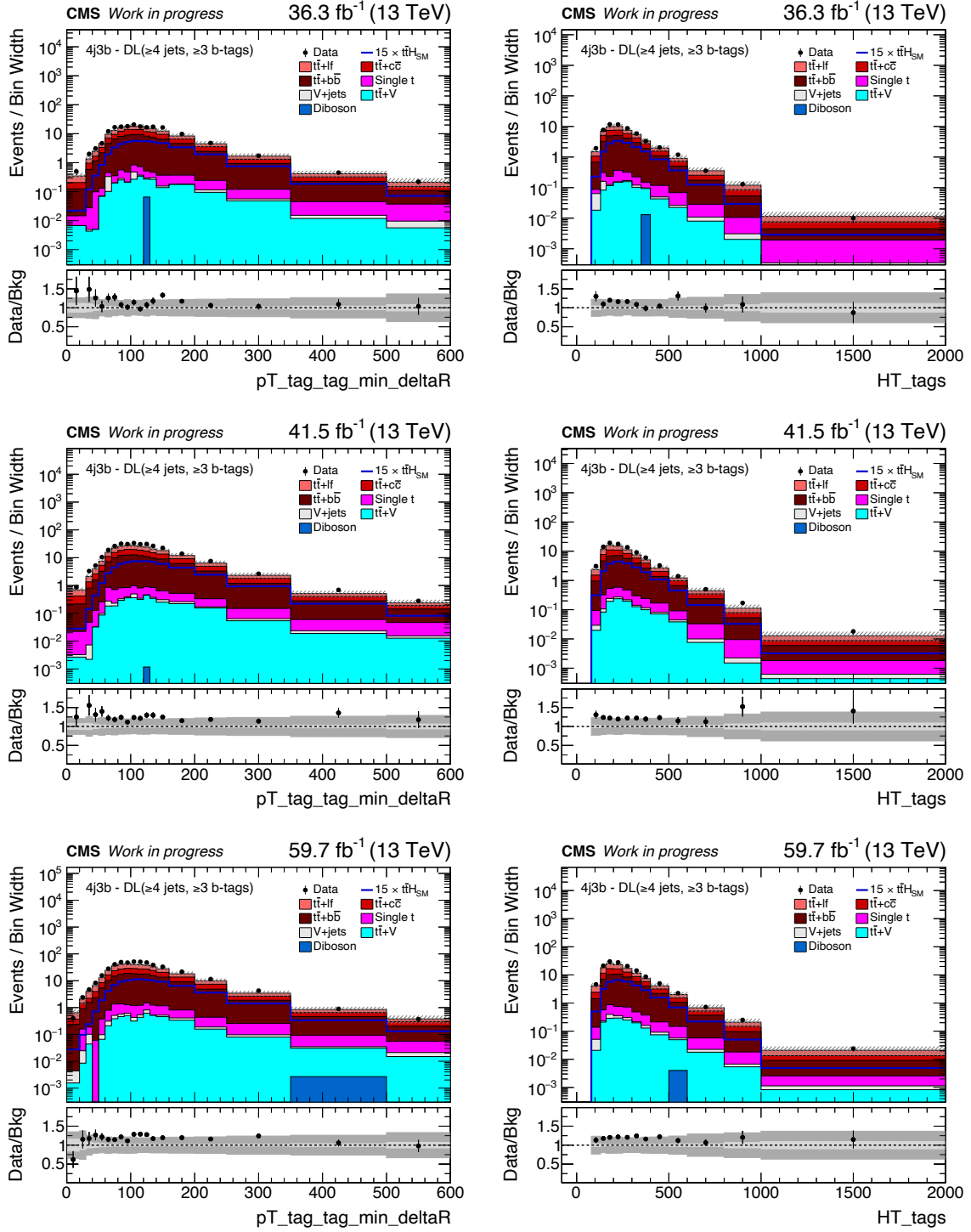


Figure G.19: Input Feature for DL ANN in all three years in the region ≥ 4 jets, ≥ 3 b tags.

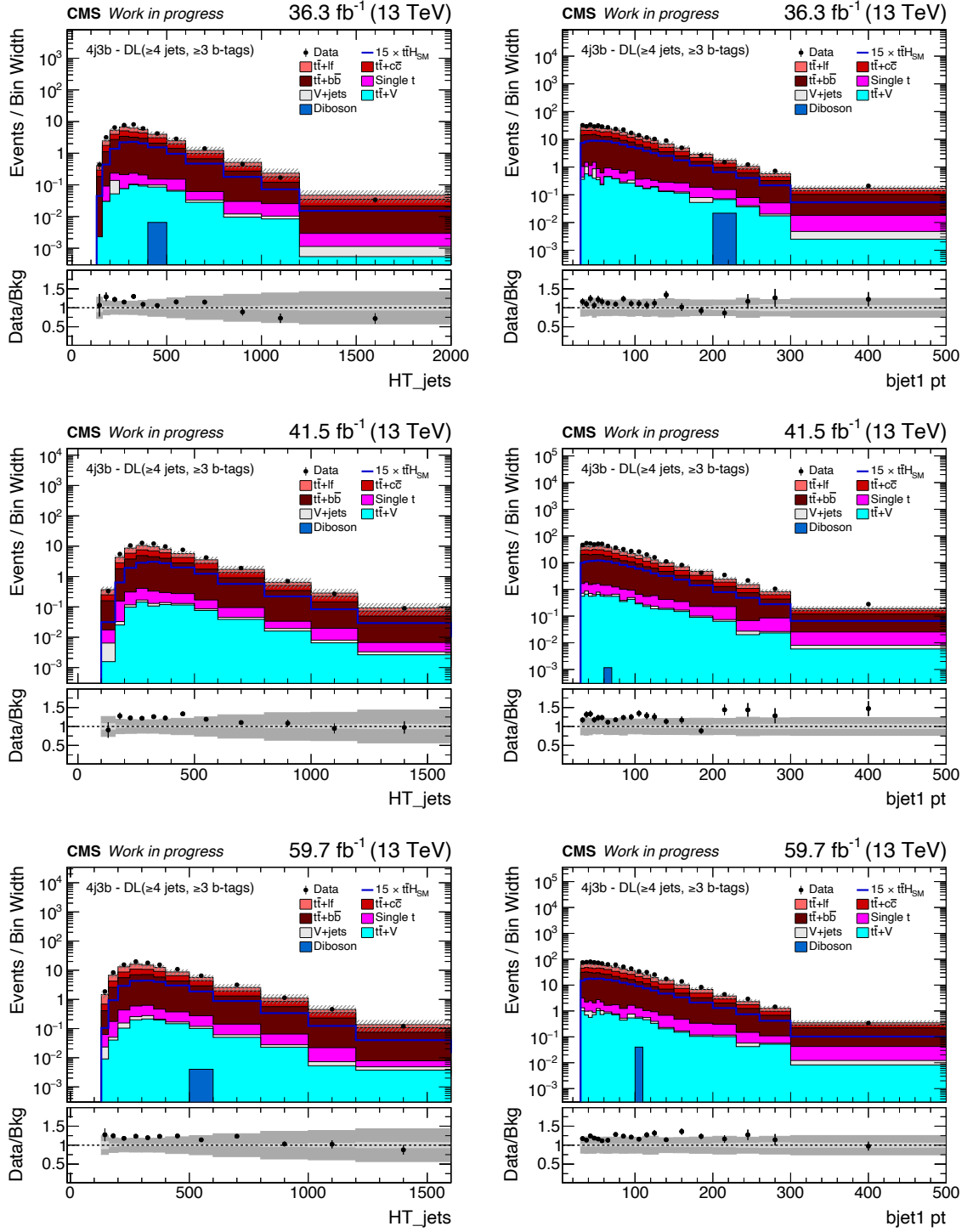


Figure G.20: Input Feature for DL ANN in all three years in the region ≥ 4 jets, ≥ 3 b tags.

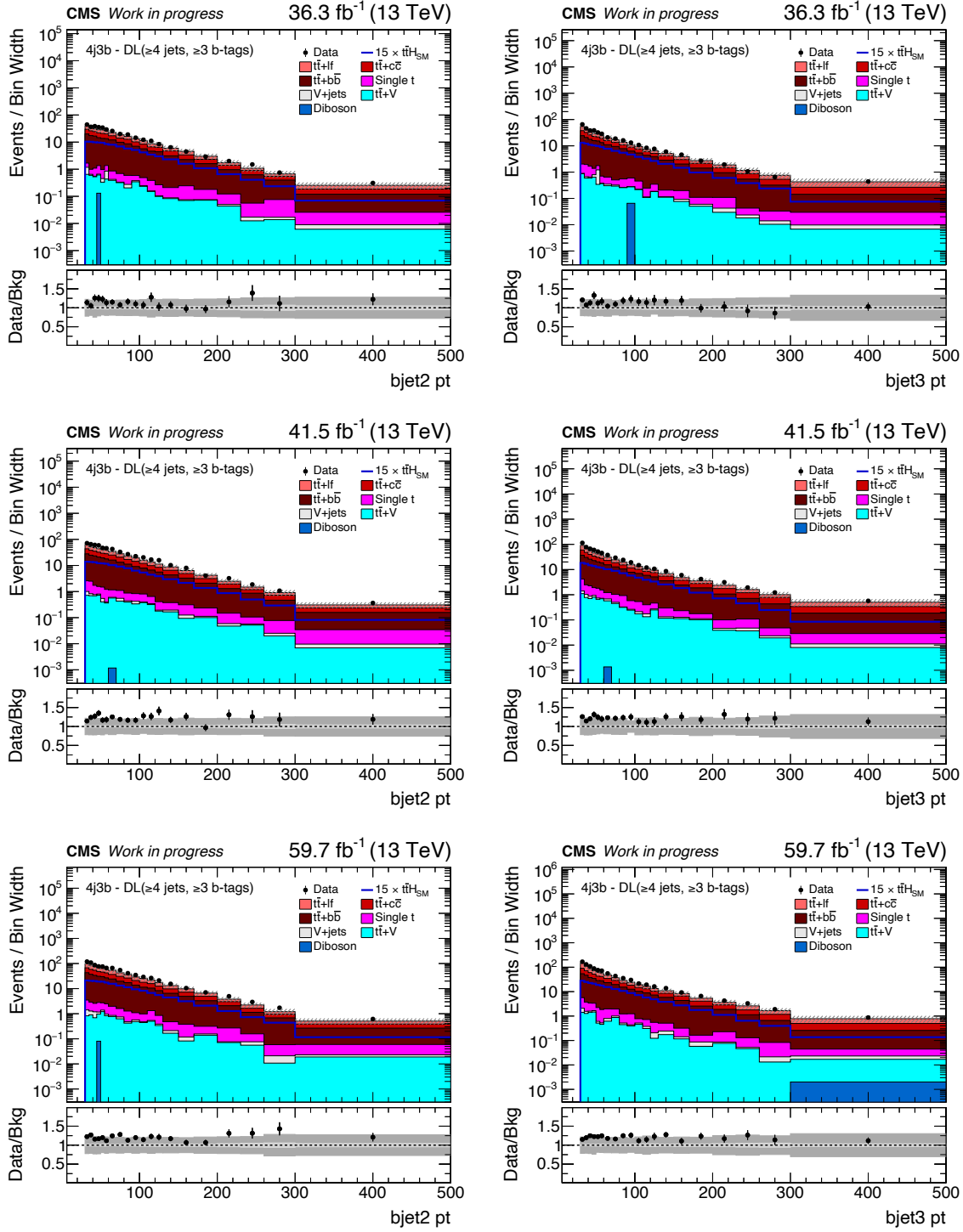


Figure G.21: Input Feature for DL ANN in all three years in the region ≥ 4 jets, ≥ 3 b tags.

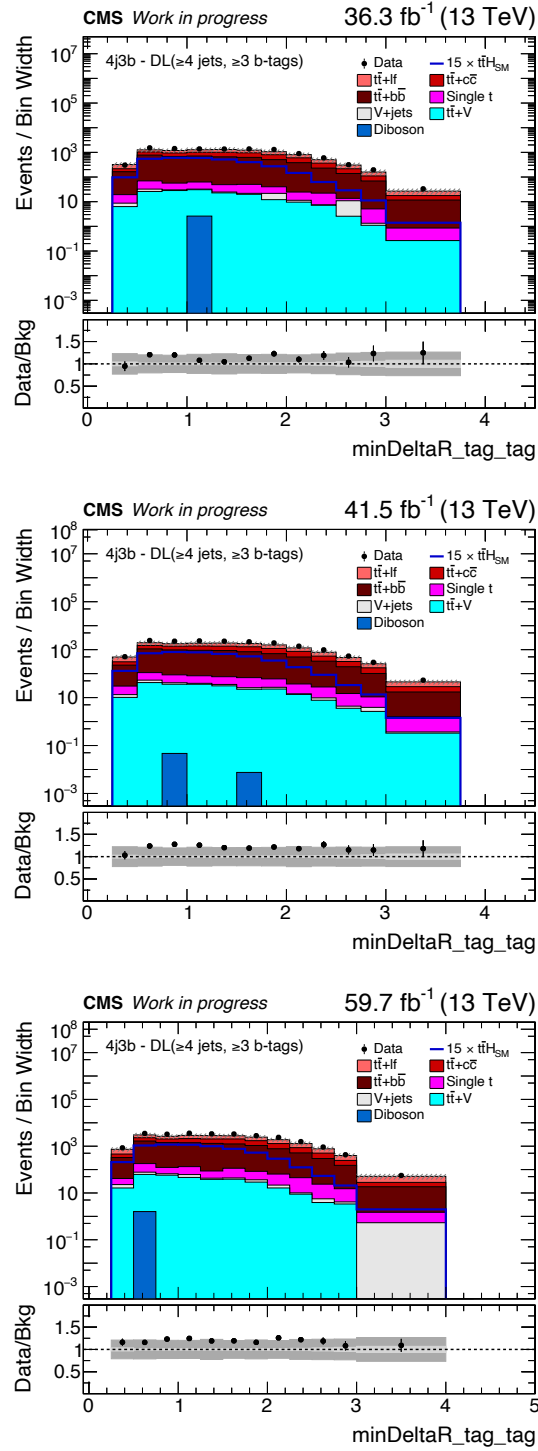


Figure G.22: Input Feature for DL ANN in all three years in the region ≥ 4 jets, ≥ 3 b tags.

PERFORMANCES OF THE ARTIFICIAL NEURAL NETWORKS

H.1 Taylor expansion

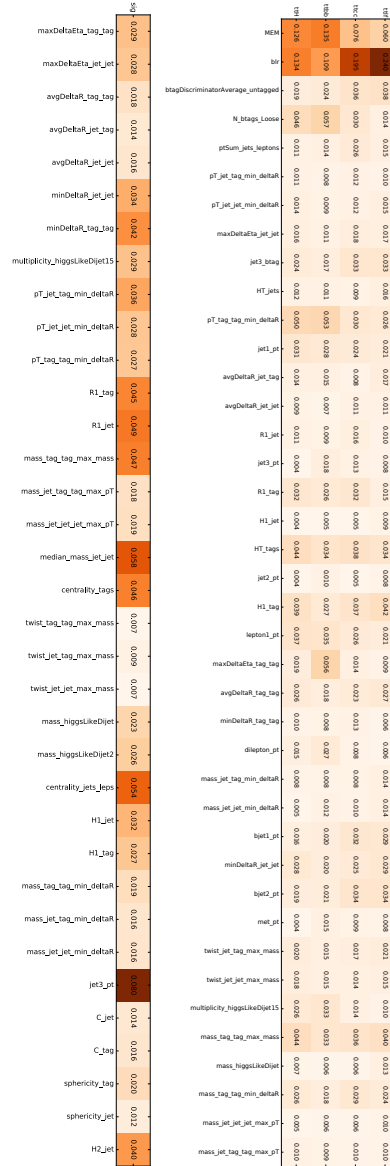


Figure H.1: First-order Taylor coefficients used for the ranking of the input variables used in the ANN training of the 3 jets, ≥ 3 b tags (top) and ≥ 4 jets, ≥ 3 b tags (bottom) categories in 2016.



Figure H.2: First-order Taylor coefficients used for the ranking of the input variables used in the ANN training of the ≥ 3 jets, ≥ 3 b tags (top) and ≥ 4 jets, ≥ 3 b tags (bottom) categories in 2017.

	ttf	ttc	ttb	ttf	MEM	ttf	ttc	ttb	MEM
maxDeltaEta_tag_tag	0.000	0.096	0.139	0.046	0.046	0.000	0.096	0.139	0.046
maxDeltaEta_jet_jet	0.048	0.134	0.112	0.015	0.015	0.048	0.134	0.112	0.015
avgDeltaR_tag_tag	0.018	0.030	0.027	0.018	0.018	0.018	0.030	0.027	0.018
avgDeltaR_jet_jet	0.017	0.016	0.004	0.017	0.017	0.017	0.016	0.004	0.017
minDeltaR_jet_jet	0.028	0.010	0.003	0.028	0.028	0.028	0.010	0.003	0.028
minDeltaR_tag_tag	0.027	0.021	0.013	0.027	0.027	0.027	0.021	0.013	0.027
R1_tag	0.016	0.021	0.013	0.016	0.016	0.016	0.021	0.013	0.016
R1_jet	0.046	0.039	0.018	0.046	0.046	0.046	0.039	0.018	0.046
mass_jet_tag_max_pt	0.027	0.014	0.026	0.027	0.027	0.027	0.014	0.026	0.027
mass_jet_jet_max_pt	0.045	0.036	0.018	0.045	0.045	0.045	0.036	0.018	0.045
mass_tag_max_mass	0.017	0.002	0.008	0.017	0.017	0.017	0.002	0.008	0.017
median_mass_jet_jet	0.031	0.013	0.017	0.031	0.031	0.031	0.013	0.017	0.031
multiplicity_higgsLikeDijet15	0.005	0.007	0.015	0.005	0.005	0.005	0.007	0.015	0.005
centrality_tags	0.014	0.027	0.017	0.014	0.014	0.014	0.027	0.017	0.014
pT_jet_min_deltaR	0.019	0.013	0.021	0.019	0.019	0.019	0.013	0.021	0.019
pT_jet_jet_min_deltaR	0.021	0.021	0.013	0.021	0.021	0.021	0.021	0.013	0.021
pT_tag_max_min_deltaR	0.021	0.018	0.024	0.021	0.021	0.021	0.018	0.024	0.021
twist_tag_max_mass	0.008	0.006	0.007	0.008	0.008	0.008	0.006	0.007	0.008
twist_jet_max_mass	0.009	0.005	0.007	0.009	0.009	0.009	0.005	0.007	0.009
twist_jet_jet_max_mass	0.011	0.007	0.018	0.011	0.011	0.011	0.007	0.018	0.011
H1_jet	0.011	0.013	0.010	0.011	0.011	0.011	0.013	0.010	0.011
H1_tag	0.012	0.012	0.007	0.012	0.012	0.012	0.012	0.007	0.012
mass_higgsLikeDijet	0.012	0.012	0.007	0.012	0.012	0.012	0.012	0.007	0.012
mass_higgsLikeDijet2	0.011	0.008	0.010	0.011	0.011	0.011	0.008	0.010	0.011
centrality_jetsleps	0.011	0.008	0.010	0.011	0.011	0.011	0.008	0.010	0.011
mass_tag_max_mass	0.011	0.008	0.010	0.011	0.011	0.011	0.008	0.010	0.011
D_jet	0.011	0.008	0.010	0.011	0.011	0.011	0.008	0.010	0.011
D_tag	0.011	0.008	0.010	0.011	0.011	0.011	0.008	0.010	0.011
mass_tag_min_deltaR	0.011	0.008	0.010	0.011	0.011	0.011	0.008	0.010	0.011
mass_jet_min_deltaR	0.011	0.008	0.010	0.011	0.011	0.011	0.008	0.010	0.011
mass_jet_jet_min_deltaR	0.011	0.008	0.010	0.011	0.011	0.011	0.008	0.010	0.011
C_jet	0.011	0.008	0.010	0.011	0.011	0.011	0.008	0.010	0.011
C_tag	0.011	0.008	0.010	0.011	0.011	0.011	0.008	0.010	0.011
sphericity_tag	0.011	0.008	0.010	0.011	0.011	0.011	0.008	0.010	0.011
sphericity_jet	0.011	0.008	0.010	0.011	0.011	0.011	0.008	0.010	0.011

Figure H.3: First-order Taylor coefficients used for the ranking of the input variables used in the ANN training of the 3 jets, ≥ 3 b tags (top) and ≥ 4 jets, ≥ 3 b tags (bottom) categories in 2018.

H.2 Input features validation

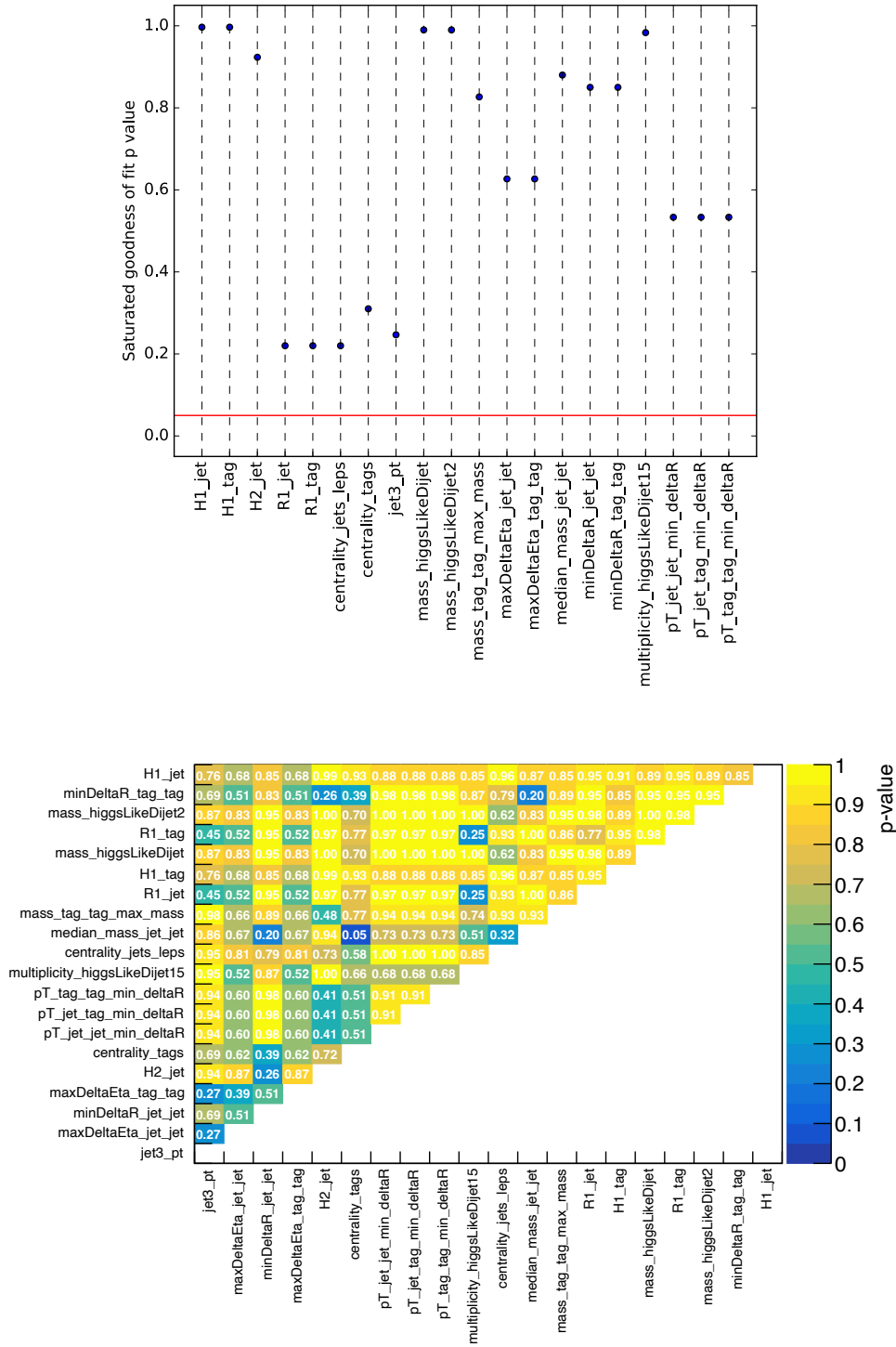


Figure H.4: p values of each variable obtained from 1D (top) and 2D (bottom) GoF test in the DL 3 jets, ≥ 3 b tags category for 2016.

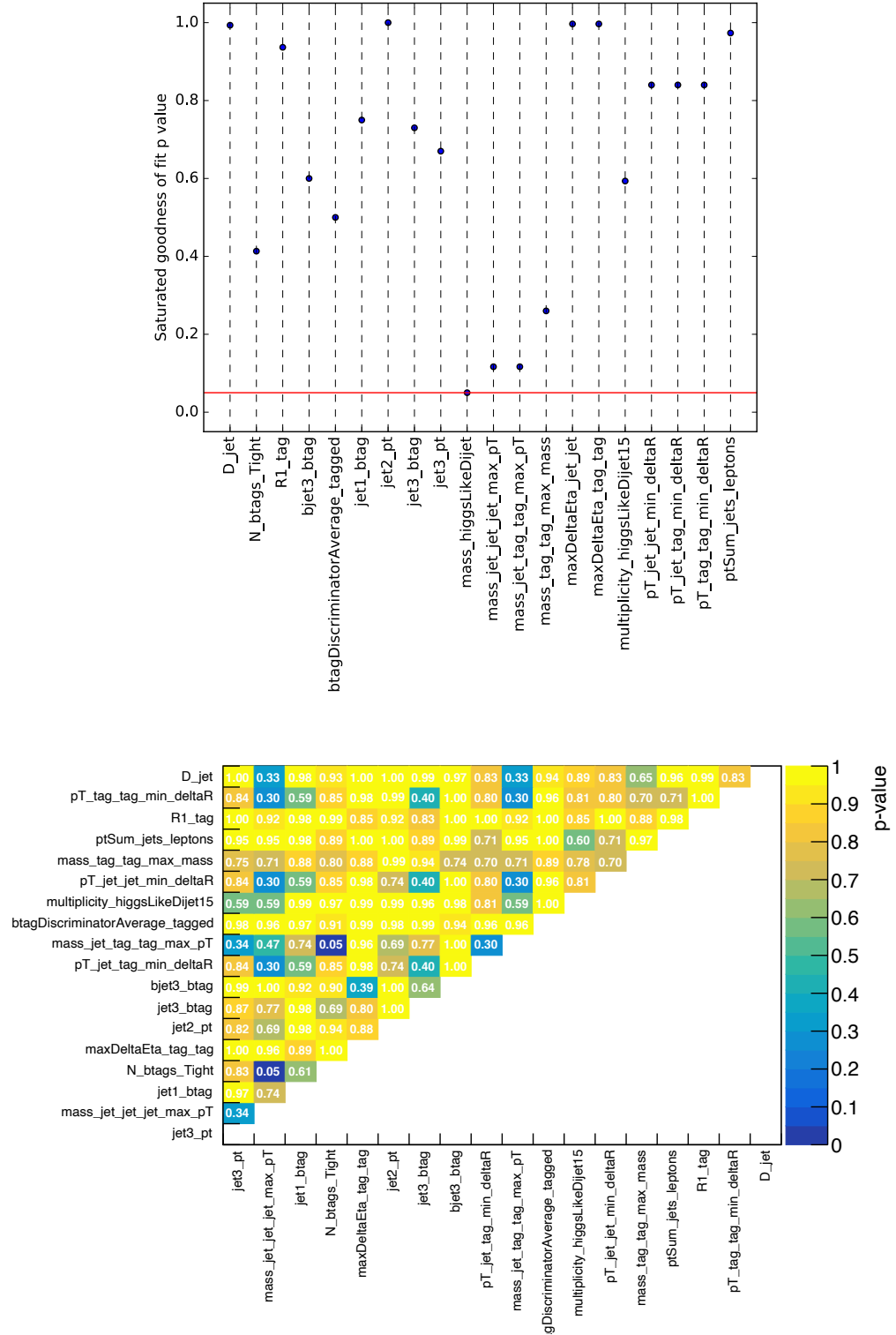


Figure H.5: p values of each variable obtained from 1D (top) and 2D (bottom) GoF test in the DL $3\text{ jets}, \geq 3\text{ b tags}$ category for 2017.

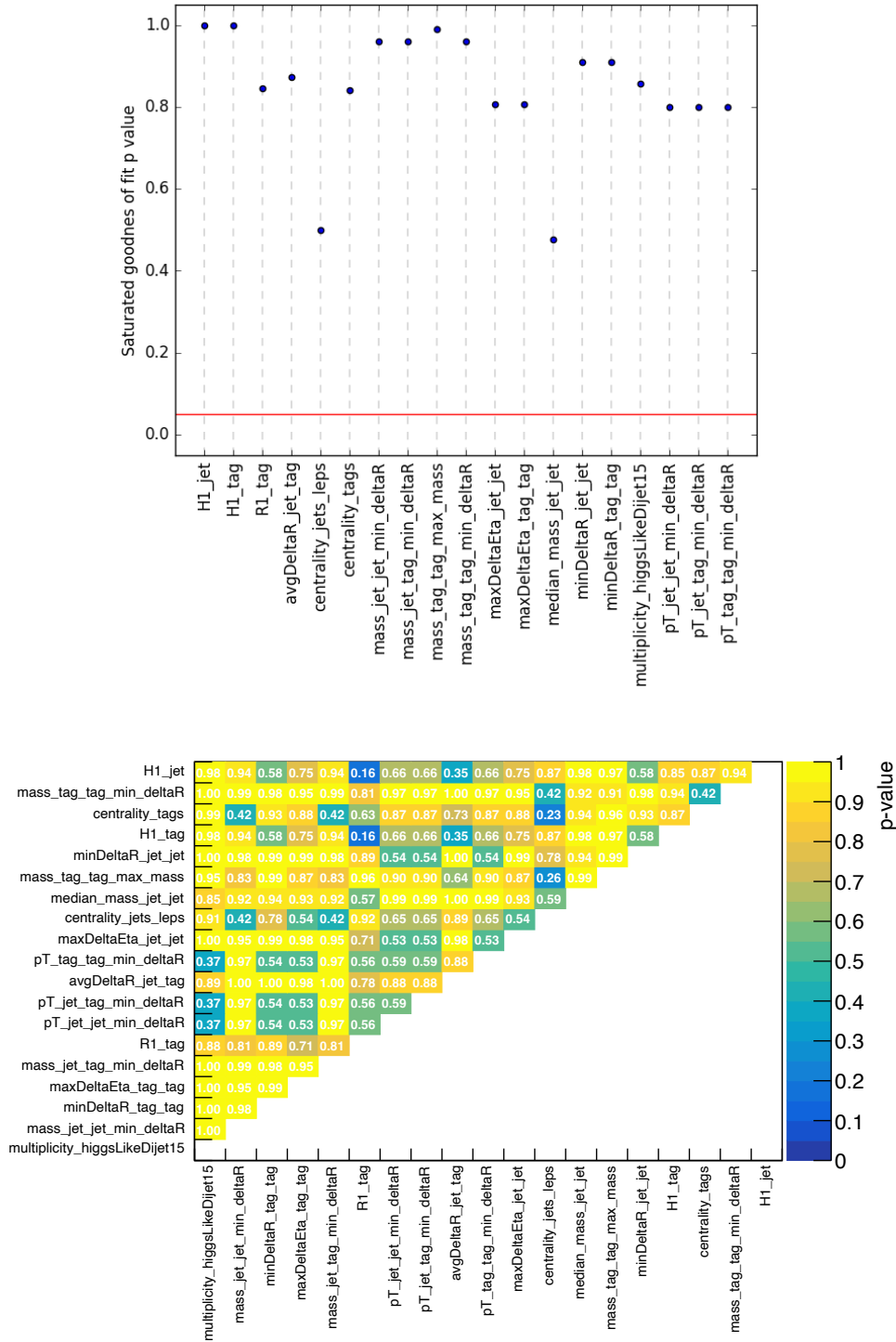


Figure H.6: p values of each variable obtained from 1D (top) and 2D (bottom) GoF test in the DL 3 jets, ≥ 3 b tags category for 2018.

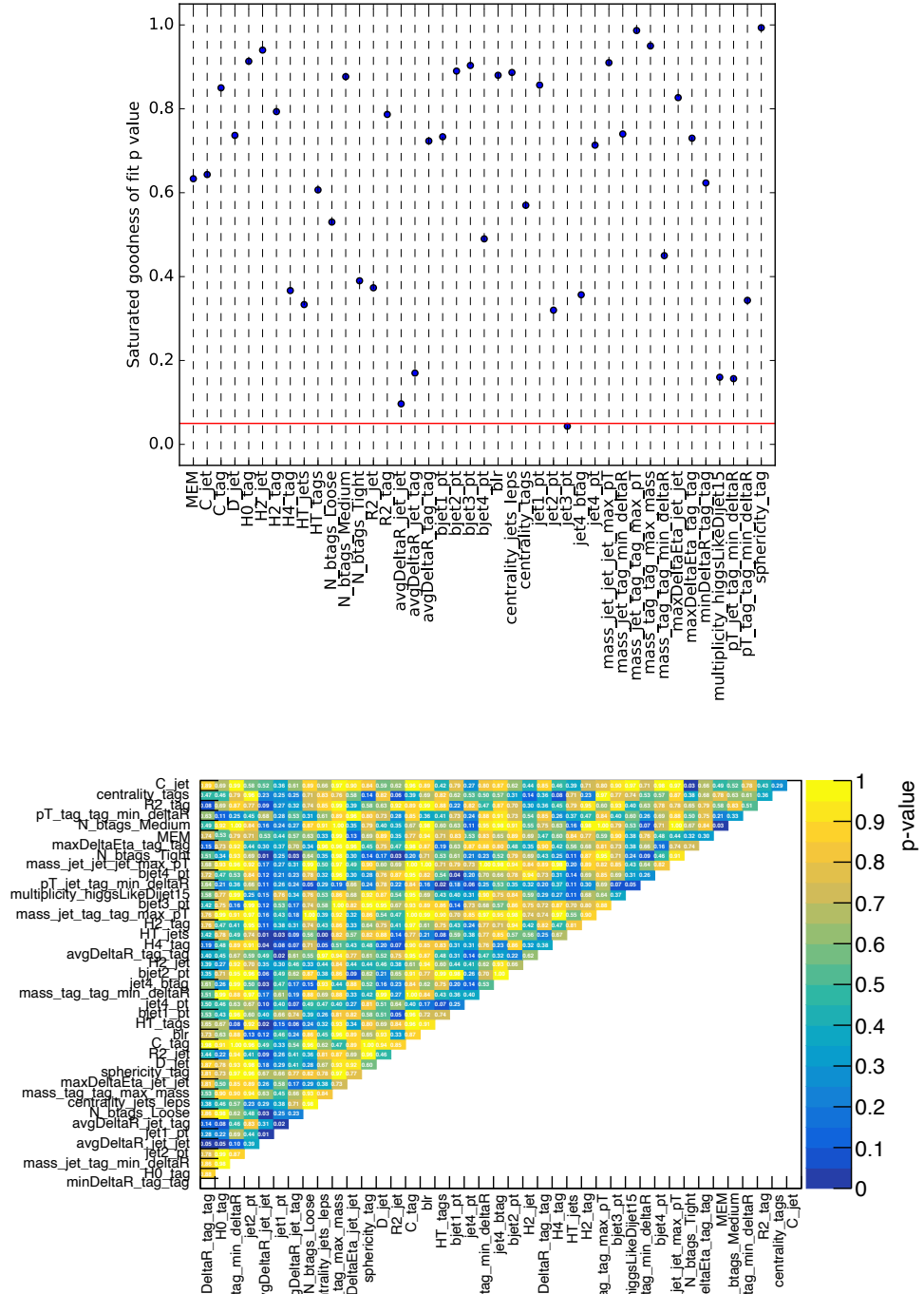


Figure H.7: p values of each variable obtained from 1D (top) and 2D (bottom) GoF test in the DL ≥ 4 jets, ≥ 3 b tags category for 2016.

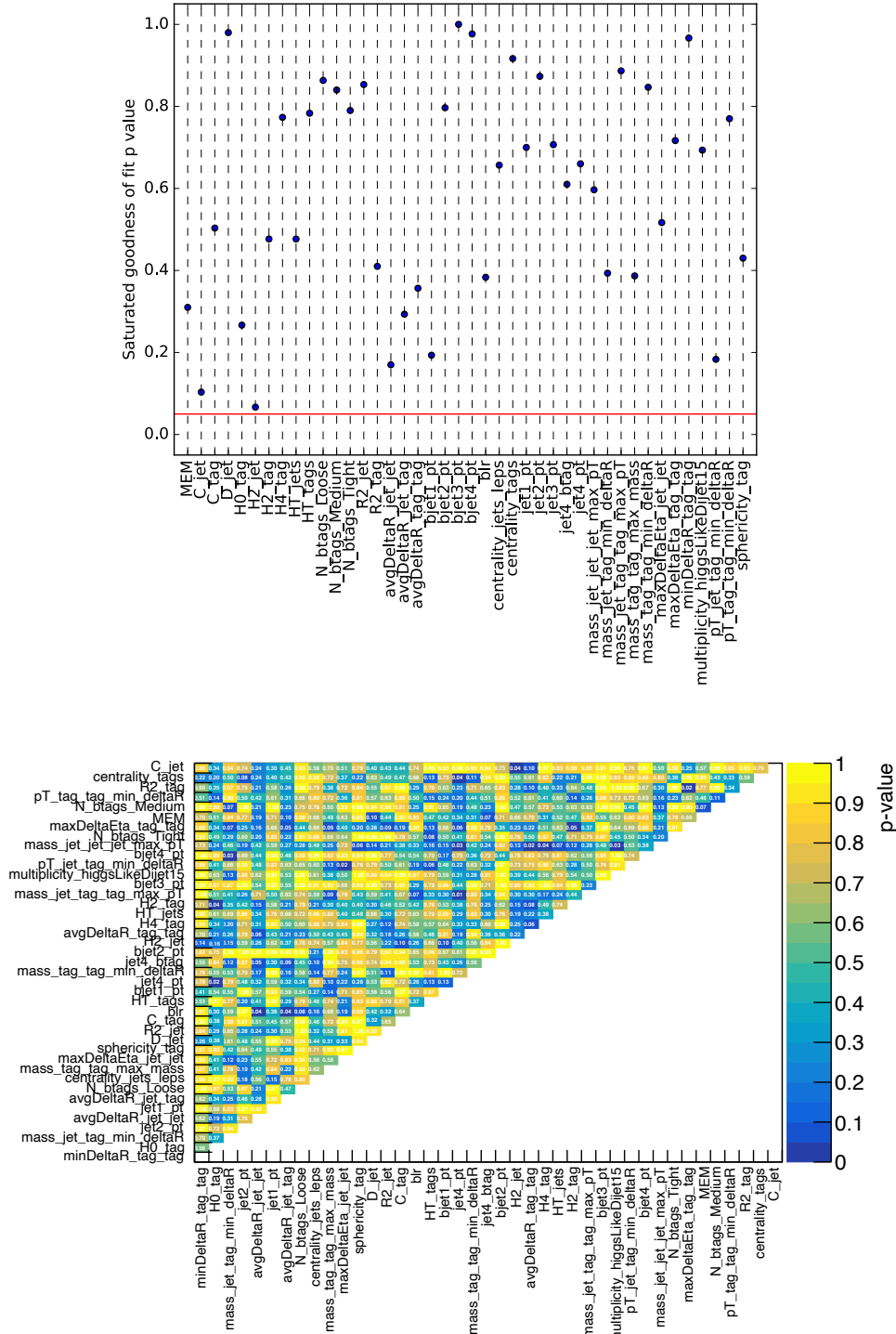


Figure H.8: p values of each variable obtained from 1D (top) and 2D (bottom) GoF test in the DL ≥ 4 jets, ≥ 3 b tags category for 2017.

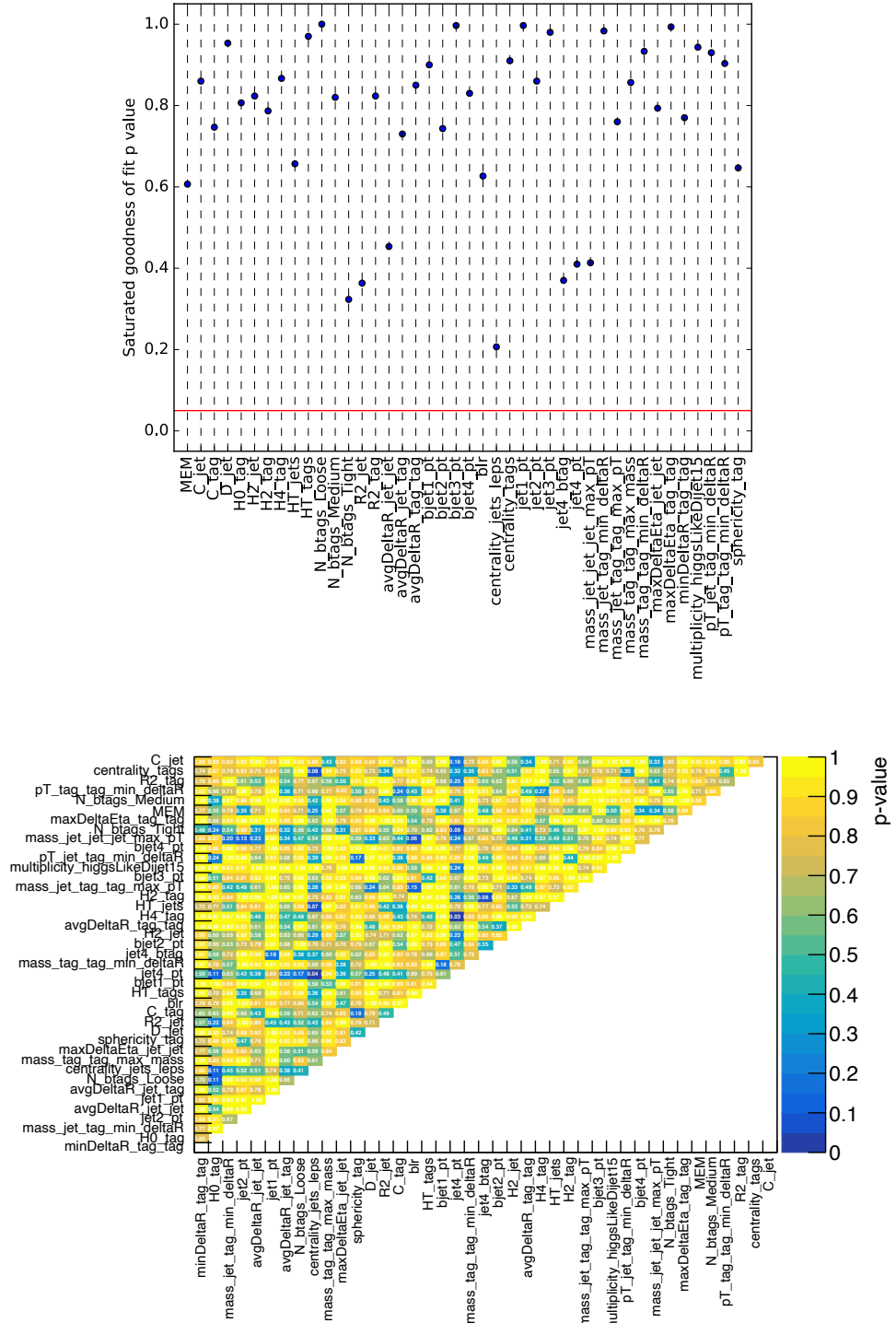


Figure H.9: p values of each variable obtained from 1D (top) and 2D (bottom) GoF test in the DL ≥ 4 jets, ≥ 3 b tags category for 2018.

H.3 Loss function

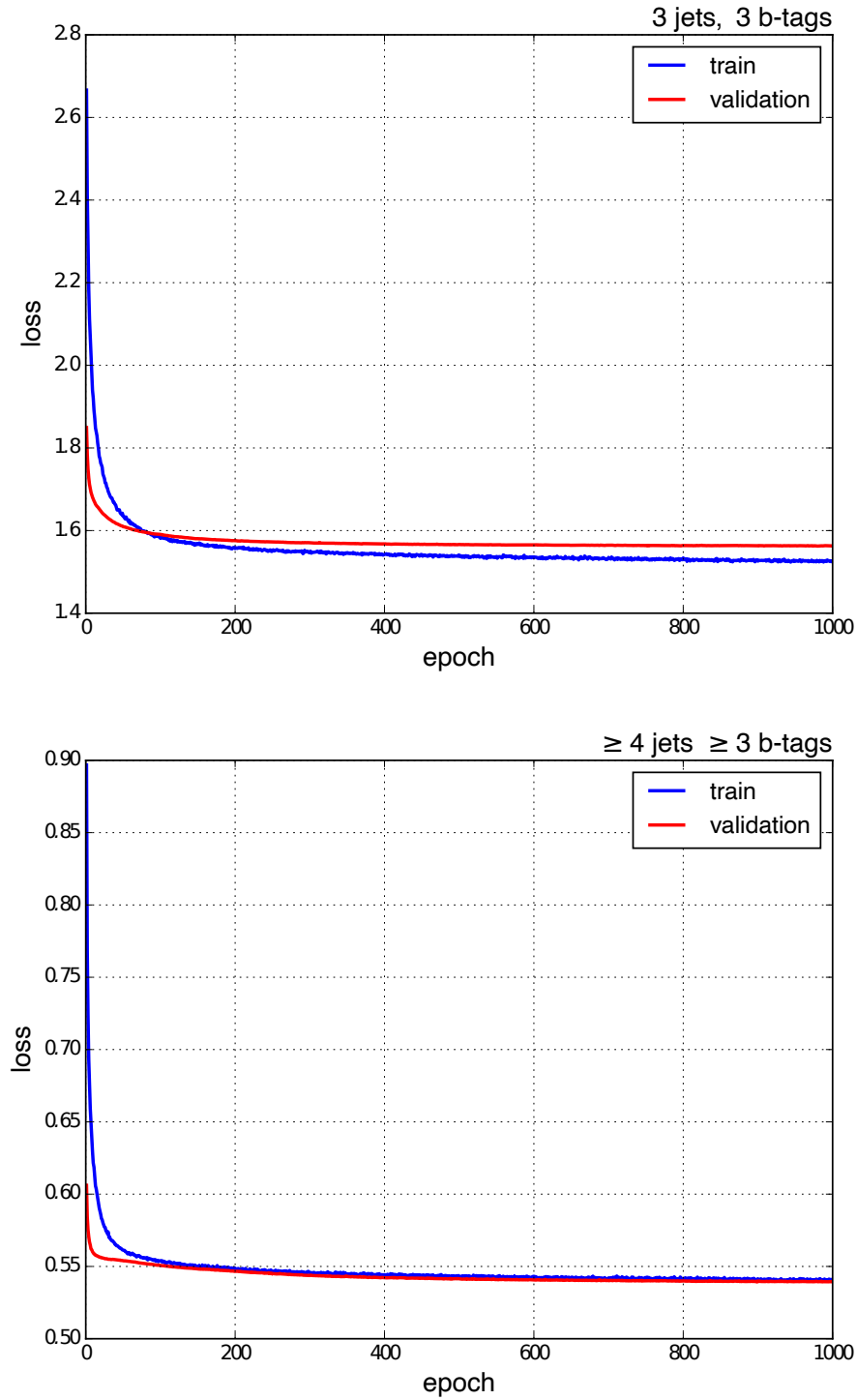


Figure H.10: Evolution of the loss as a function of the training epoch, for the networks in the 3 jets, ≥ 3 b tags (left) and ≥ 4 jets, ≥ 3 b tags (right) categories. The cost function is evaluated for both the training (blue) and test (red) data.

I THE RATIO OBSERVABLE

This section provides a summary of the studies that were carried out in order to investigate and justify the application of the likelihood ratio observable.

There is a relative large confusion in the Artificial Neural Network (ANN) categorization, in particular between the signal $t\bar{t}H$ and the main irreducible background $t\bar{t} + b\bar{b}$ processes, as illustrated in Figure 5.19. In order to strengthen the discrimination even further, a probability ratio observable based on the outputs of the ANN for the events in the respective process categories is implemented as:

$$p_o = \frac{o(t\bar{t}H)}{o(t\bar{t}H) + o(ttbb)} . \quad (\text{I.1})$$

Different configurations were tested and yielded to the choice described in Section 5.7 according to the best performances in terms of:

- discriminating power of each bin in the resulting templates, e.g. the approximate median significance (AMS) [302];
- the post-fit uncertainties and the expected significance in the final fit to pseudo data corresponding to the SM prediction, refereed to as Asimov data set in the following;
- the constraints on the individual nuisance parameters;
- the stability and robustness of the fit with respect to the statistical fluctuations expected in the templates as well as different scenarios for the $t\bar{t} + B$ prediction.

The most relevant tests done to validate the ratio observable are discussed here, and they were utilized to finalize the strategy employed for the dilepton $t\bar{t}H, H \rightarrow b\bar{b}$ analysis.

The comparison of the AMS metric, which provides a way to evaluate the quality of the classifier rapidly,

Table I.1 shows a comparison of the purity of the discriminator bins. The comparison considers the discriminants constructed from the ANN outputs of the $t\bar{t}H$ node with respect to the final ratio observable of Eq. 5.2. Overall, the signal purity in the ratio observable is higher, leading to an higher averaged AMS purity in each bin compared to the ANN discriminant distributions.

The expected sensitivity in a fit to Asimov data is shown in Table I.2 using distributions of the separate ANN nodes or the ratio observable. When employing the ratio observable, the total uncertainty in the parameter estimation slightly decreases. The statistical uncertainty is the same in all statistical models. However, because the models that use ratio observables include fewer bins for parameter estimation, the discrimination power of the ratio observable distributions is larger. The improvement in the expected significance is attributed to the improvement in the control over the systematic uncertainties, as further compared in the following. This is consistent with the behavior observed with the AMS. The improvement of the total uncertainty in the parameter estimation therefore originates from improved control over systematic uncertainties.

	$t\bar{t}H$ ANN output node	Ratio observable
$\sum_{\text{years}}$ Bins	72	36
highest AMS	0.88	0.94
second AMS	0.74	0.93
third AMS	0.63	0.80
forth AMS	0.61	0.67
fifth AMS	0.60	0.66
$\langle \text{AMS} \rangle$	2.46	2.54

Table I.1: Comparison of AMS of most discriminator bins in the distributions of the ANN output for the $t\bar{t}H$ (left) and the final ratio observable constructed for events in the $t\bar{t}H$ and $t\bar{t} + b(\bar{b})$ node (right).

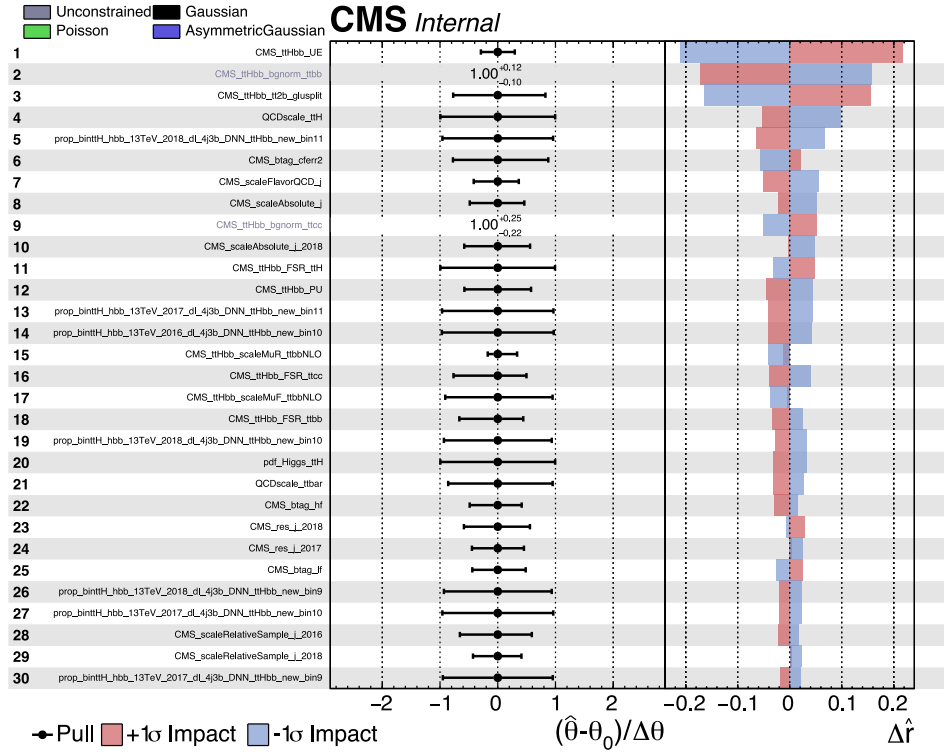
The comparison based on the ranking of the impact of the nuisance parameters on the signal inference is in Figure I.1. Overall, the model based on the ratio observable is less susceptible to the nuisance parameters, as seen by the smaller order of magnitude of the impacts when compared to the statistical model based directly on the ANN discrimination. Uncertainties modeling the effects of the simulated samples limited size, as well as modeling uncertainties on the signal processes and the $t\bar{t} + \text{jets}$ background, dominate the ranking. Given the discrimination power of the ratio observable, the large impact of the parameters directly affecting the corresponding bin contents is understandable. Furthermore, the high position of the modeling uncertainties in the impact ranking is expected. As a result, the impact ranking suggests better control over systematic uncertainties minimizing their impact on signal inference.

A comparison of the constraints extracted from the estimation of the nuisance parameters is listed in Table I.3 for all the highest ranked nuisances. Overall a reduction in the constraints is visible, which originate from the coarser binning of the distributions of the ratio observable and indicating the flexibility of the statistical model. Furthermore, having fewer bins helps in avoiding artificial constraints arising from the statistical fluctuation of the distributions used to construct them. Nonetheless, coarser binning of the ANN discriminant distributions in the $t\bar{t}H$ and $t\bar{t} + b\bar{b}$ nodes results in a loss of sensitivity to the $t\bar{t}H, H \rightarrow b\bar{b}$ process caused by a higher contribution of background processes in the individual bins. The statistical model based on the ratio observable can benefit from the advantages of using fewer bins while retaining high sensitivity to the signal strength modifier.

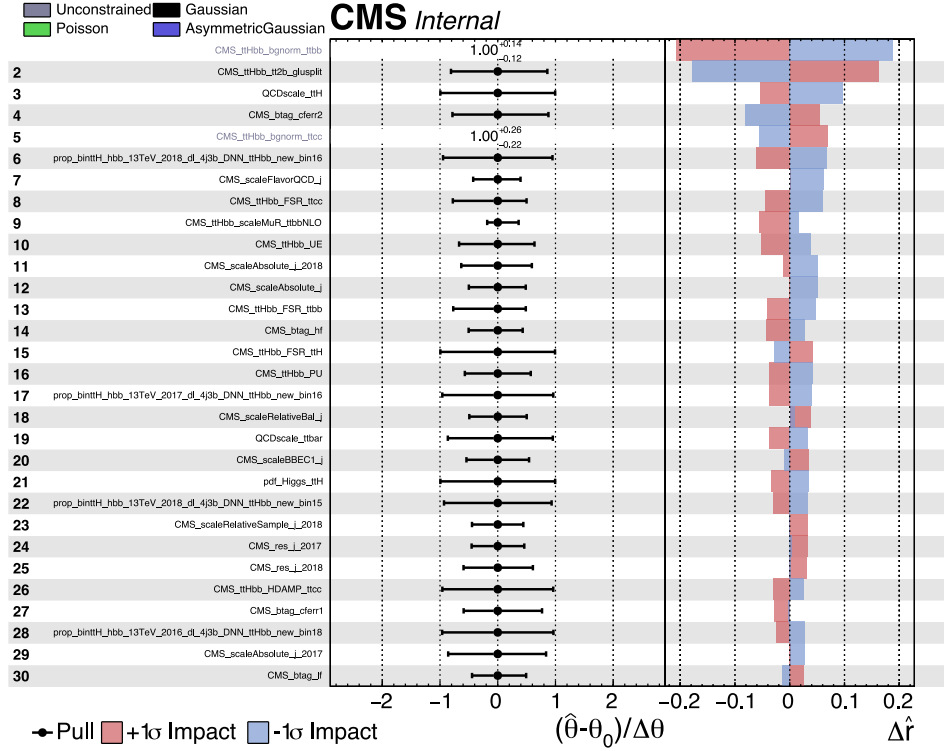
The stability of the signal strength estimation is assessed by fitting pseudo data generated from the Asimov data set and accounting for the statistical fluctuations expected from Poisson statistics. Two different sets are tested. The first set employs the $t\bar{t} + b\bar{b}$ processes as predicted by the simulation. The second set considers a 30% increased $t\bar{t} + b\bar{b}$ prediction, which is supposed to provide a more realistic test scenario given the most recent measurements of the $t\bar{t} + b\bar{b}$ process [119, 277]. The injected $t\bar{t}H, H \rightarrow b\bar{b}$ signal strength is one. Table I.4 displays the results, showing that the injected signal strength value of one is covered within the 68% confidence interval in the mean estimated signal strength in both sets of pseudo experiments and both using the ANN outputs or the ratio observable. However, the central values of the estimated signal strength is much closer to the true injected value in the scenario with

Fitted Observables	Stat+Syst	Stat-Only	Significance
ANN outputs	1.0 +0.49/-0.47	1.0 +0.33/-0.34	2.1
Ratio observable	1.0 +0.46/-0.44	1.0 +0.33/-0.34	2.2

Table I.2: Results in a Fit to Asimov Data with injected signal strength $\mu_{t\bar{t}H} = 1$.



(a)



(b)

Figure I.1: Comparison of ranking based on the impacts of the nuisance parameters on the signal inference based directly on the ANN discriminant distributions (a) and on the ratio observable (b).

Parameter	ANN outputs	Ratio observable	Relative difference (%)
CMS_ttHbb_bgnorm_ttbb	+0.115 -0.101	+0.141 -0.122	0.217
CMS_ttHbb_tt2b_glusplit	+0.826 -0.772	+0.859 -0.810	0.045
QCDscale_ttH	+0.994 -0.994	+0.994 -0.994	0.0
CMS_btag_cferr2	+0.875 -0.776	+0.878 -0.785	0.007
CMS_ttHbb_bgnorm_ttcc	+0.249 -0.218	+0.257 -0.225	0.03
CMS_scaleFlavorQCD_j	+0.365 -0.416	+0.394 -0.426	0.052
CMS_ttHbb_FSR_ttcc	+0.498 -0.766	+0.499 -0.778	0.009
CMS_ttHbb_scaleMuR_ttbbNLO	+0.335 -0.173	+0.362 -0.183	0.067
CMS_ttHbb_UE	+0.295 -0.292	+0.636 -0.670	1.226
CMS_scaleAbsolute_j_2018	+0.560 -0.580	+0.591 -0.632	0.072

Table I.3: Comparison of the parameter constraints with the largest impact originating from the parameter estimation of $\mu_{\text{t}\bar{\text{t}}\text{H}}$, using a statistical model based directly on the ANN discriminants (second column) and the ratio observable (third column). The relative difference in the last column is calculated with respect to the values in the second column.

statistical model based on the ratio observable, implying that the signal strength estimation is more stable in the presence of statistical fluctuation.

To evaluate the stability even further, the bias in signal strength estimation is calculated for each pseudo experiment as:

$$b = \frac{\mu_{\text{post-fit}} - \mu^{\text{injected}}}{\sigma_{\mu}^{\text{post-fit}}}. \quad (\text{I.2})$$

As above, the injected signal strength μ^{injected} is equal to one, and the $\sigma_{\mu}^{\text{post-fit}}$ represents the uncertainty on the estimated signal strength modifier. The distributions of this quantity in the fits to the different sets of pseudo experiments are shown in Figure I.2; this metric is zero if the parameter estimation closes perfectly. In the parameter estimations using the statistical model based directly on the ANN discriminants, the ideal value of zero is not compatible within the 68% confidence interval. Instead, the ideal value is compatible within the 68% interval when using the ratio observable. Furthermore, in the fit to pseudo experiments based on the nominal $\text{t}\bar{\text{t}} + \text{b}\bar{\text{b}}$ scenario, the mean value of the predicted bias in signal strength estimation drops by 23% and by 28% in the fits to the scaled $\text{t}\bar{\text{t}} + \text{b}\bar{\text{b}}$ scenario. Within the uncertainty of gaussian fit, the mean expected bias values of the two sets of pseudo experiments are compatible, which is not the case in the statistical model based solely on ANN discriminants, implying again that parameter estimation is more resistant to statistical fluctuations using the ratio observable.

Statistical model	$\text{t}\bar{\text{t}} + \text{b}\bar{\text{b}}$ hypothesis	Fit to Asimov	Fit to pseudo data
ANN outputs	$\text{t}\bar{\text{t}} + \text{b}\bar{\text{b}}$ (4FS)	1.0 ± 0.48	0.92 ± 0.50
	$\text{t}\bar{\text{t}} + \text{b}\bar{\text{b}}$ (4FS) $\times 1.3$	1.0 ± 0.54	0.88 ± 0.53
	$\text{t}\bar{\text{t}} + \text{b}\bar{\text{b}}$ (4FS)	1.0 ± 0.45	0.97 ± 0.44
Ratio observable	$\text{t}\bar{\text{t}} + \text{b}\bar{\text{b}}$ (4FS) $\times 1.3$	1.0 ± 0.50	0.95 ± 0.50

Table I.4: Comparison of the signal strength estimation in fits to pseudo data. The mean values of the estimated signal strength in fits to 1000 pseudo experiments generated from the statistical model specified in the second column are shown. The estimation in a fit to the Asimov data set is in the last column.

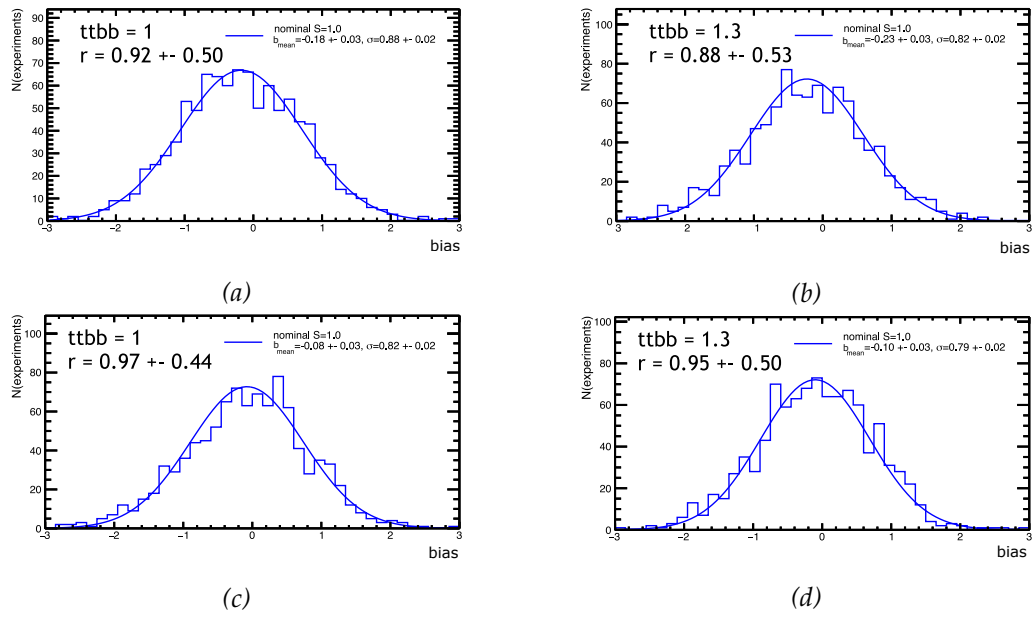


Figure I.2: Distributions of the bias in fits to pseudo data generated from the statistical model based on the ANN discriminants (upper row) and the ratio observable (bottom row). Two sets of $t\bar{t} + b\bar{b}$ hypotheses are tested: the one corresponding to the statistical model used in the main analysis (left column) or scaled by 30% (right column). The x-axis represents the bias. The mean bias and standard deviation estimations are derived from a gaussian fit (blue line) to the distributions.

YIELDS TABLES

J.1 Inclusive measurement - full Run 2 dilepton fit

Process	pre-fit (post-fit) yields							
	3j 3t		4j ≥ 3t Ratio (t $\bar{t}$ H + t $\bar{t}$ + b $\bar{b}$)		4j ≥ 3t t $\bar{t}$ + c $\bar{c}$		4j ≥ 3t t $\bar{t}$ + LF	
t $\bar{t}$ + LF	433	(410)	105	(101)	183	(166)	263	(242)
t $\bar{t}$ + C	322	(531)	186	(300)	239	(372)	171	(272)
t $\bar{t}$ + B	626	(797)	784	(935)	233	(276)	138	(162)
Single-t	47	(44)	26	(25)	15	(13)	12	(11)
V + jets	5	(5)	2	(2)	4	(3)	3	(3)
t $\bar{t}$ V	8	(8)	25	(23)	10	(10)	7	(7)
t $\bar{t}$ γ*	7	(7)	11	(10)	7	(7)	8	(8)
Diboson	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
tH	0.6	(0.6)	0.8	(0.8)	0.2	(0.2)	0.1	(0.1)
Total bkg.	1448	(1801)	1140	(1398)	691	(847)	603	(704)
± tot unc.	±360	(±90)	±240	(±60)	±174	(±47)	±149	(±42)
t $\bar{t}$ H	13	(0.0)	42	(0.0)	8	(0.0)	5	(0.0)
± tot unc.	±2	(±2.6)	±5	(±7.3)	±1	(±1.6)	±1	(±0.9)
Data	1697		1294		760		647	

Table J.1: Prefit (post-fit) yields in the 3 jets, ≥ 3 b tags category and ≥ 4 jets, ≥ 3 b tags t $\bar{t}$ H and t $\bar{t}$ + jets nodes for the 2016 dataset for the inclusive fit to dilepton channel. The prefit uncertainties are not including the freely-floating parameters; while the postfit uncertainties consider all nuisances in the fit to observed data.

Process	pre-fit (post-fit) yields							
	3j 3t		4j ≥ 3t	Ratio (t $\bar{t}$ H + t $\bar{t}$ + b $\bar{b}$)	4j ≥ 3t t $\bar{t}$ + c $\bar{c}$	4j ≥ 3t t $\bar{t}$ + LF		
t $\bar{t}$ + LF	722	(590)	134	(101)	263	(200)	528	(427)
t $\bar{t}$ + C	505	(1081)	223	(466)	310	(628)	372	(734)
t $\bar{t}$ + B	859	(1180)	971	(1253)	314	(399)	266	(331)
Single-t	60	(60)	37	(37)	18	(19)	30	(27)
V + jets	10	(11)	4	(4)	2	(2)	4	(4)
t $\bar{t}$ V	15	(15)	33	(34)	12	(12)	13	(13)
t $\bar{t}$ γ*	11	(11)	9	(9)	9	(9)	11	(11)
Diboson	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
tH	0.8	(0.8)	1	(1)	0.3	(0.4)	0.3	(0.3)
Total bkg.	2183	(2950)	1411	(1905)	929	(1269)	1224	(1548)
± tot unc.	±530	(±138)	±302	(±90)	±237	(±84)	±296	(±93)
t $\bar{t}$ H	18	(0.0)	49	(0.0)	12	(0.0)	9	(0.0)
± tot unc.	±2	(±4.0)	±6	(±11.1)	±2	(±1.5)	±1	(±1.8)
Data	2730		1795		1132		1370	

Table J.2: Prefit (post-fit) yields in the 3 jets, ≥ 3 b tags category and ≥ 4 jets, ≥ 3 b tags t $\bar{t}$ H and t $\bar{t}$ + jets nodes for the 2017 dataset for the inclusive fit to dilepton channel. The prefit uncertainties are not including the freely-floating parameters; while the postfit uncertainties consider all nuisances in the fit to observed data.

pre-fit (post-fit) yields								
Process	3j 3t		4j $\geq$ 3t	Ratio ($t\bar{t}H + t\bar{t} + b\bar{b}$)	4j $\geq$ 3t $t\bar{t} + c\bar{c}$	4j $\geq$ 3t $t\bar{t} + LF$		
$t\bar{t} + LF$	1234	(1032)	244	(193)	451	(354)	848	(724)
$t\bar{t} + C$	742	(1567)	393	(787)	463	(904)	514	(992)
$t\bar{t} + B$	1287	(1762)	1515	(1940)	466	(587)	359	(450)
Single-t	103	(101)	59	(59)	29	(28)	43	(42)
V + jets	13	(14)	7	(7)	4	(4)	8	(9)
$t\bar{t}V$	23	(23)	45	(46)	19	(20)	16	(16)
$t\bar{t}\gamma^*$	18	(18)	15	(16)	10	(9)	16	(15)
Diboson	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
tH	1	(1)	2	(2)	0.4	(0.4)	0.4	(0.4)
Total bkg.	3421	(4519)	2279	(3049)	1441	(1907)	1806	(2248)
$\pm$ tot unc.	± 758	(± 216)	± 489	(± 138)	± 342	(± 104)	± 433	(± 121)
$t\bar{t}H$	27	(0.0)	76	(0.0)	17	(0.0)	12	(0.0)
$\pm$ tot unc.	± 4	(± 4.6)	± 10	(± 16.8)	± 2	(± 3.5)	± 2	(± 2.2)
Data	4167		2842		1694		2033	

Table J.3: Prefit (post-fit) yields in the 3 jets, ≥ 3 b tags category and ≥ 4 jets, ≥ 3 b tags $t\bar{t}H$ and $t\bar{t} + jets$ nodes for the 2018 dataset for the inclusive fit to dilepton channel. The prefit uncertainties are not including the freely-floating parameters; while the postfit uncertainties consider all nuisances in the fit to observed data.

J.2 Inclusive measurement - full Run 2 all channels fit

Process	pre-fit (post-fit) yields							
	3j 3t		$4j \geq 3t$ ratio		$4j \geq 3t$ $t\bar{t} + c\bar{c}$		$4j \geq 3t$ $t\bar{t} + \text{LF}$	
$t\bar{t} + \text{LF}$	433	(397)	105	(99)	183	(170)	263	(254)
$t\bar{t} + C$	322	(338)	186	(195)	239	(253)	171	(171)
$t\bar{t} + b\bar{b}$	626	(860)	784	(979)	233	(296)	138	(181)
Single-t	47	(45)	26	(27)	15	(15)	12	(12)
V + jets	5	(5)	2	(2)	4	(4)	3	(4)
$t\bar{t}V$	8	(8)	25	(23)	10	(10)	7	(6)
$t\bar{t}\gamma^*$	7	(7)	11	(10)	7	(6)	8	(8)
Diboson	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
tH	0.6	(0.6)	0.8	(0.8)	0.2	(0.2)	0.1	(0.1)
TotalBkg	1449	(1659)	1140	(1337)	691	(753)	603	(635)
$\pm$ tot unc.	± 88	(± 18)	± 146	(± 24)	± 143	(± 24)	± 131	(± 20)
TotalSig	13	(4)	42	(14)	8	(3)	5	(2)
$\pm$ tot unc.	± 0	(± 1)	± 2	(± 3)	± 1	(± 2)	± 1	(± 1)
Data	1697		1294		760		647	

Table J.4: Prefit (post-fit) yields in the $3\text{ jets}, \geq 3\text{ }b\text{ tags}$ category and $\geq 4\text{ jets}, \geq 3\text{ }b\text{ tags}$ $t\bar{t}H$ and $t\bar{t} + \text{jets}$ nodes for the 2016 dataset for the inclusive fit to all channels. The prefit uncertainties are not including the freely-floating parameters; while the postfit uncertainties consider all nuisances in the fit to observed data.

Process	pre-fit (post-fit) yields							
	3j 3t		$4j \geq 3t$ ratio		$4j \geq 3t$ $t\bar{t} + c\bar{c}$		$4j \geq 3t$ $t\bar{t} + \text{LF}$	
$t\bar{t} + \text{LF}$	722	(812)	134	(171)	263	(317)	528	(606)
$t\bar{t} + C$	505	(514)	223	(240)	310	(323)	372	(372)
$t\bar{t} + b\bar{b}$	859	(1247)	971	(1295)	314	(423)	266	(361)
Single-t	60	(65)	37	(39)	18	(20)	30	(32)
V + jets	10	(12)	4	(4)	2	(2)	4	(5)
$t\bar{t}V$	15	(14)	33	(32)	12	(11)	13	(12)
$t\bar{t}\gamma^*$	11	(11)	9	(9)	9	(9)	11	(11)
Diboson	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
tH	0.8	(0.8)	1	(1)	0.3	(0.4)	0.3	(0.3)
TotalBkg	2183	(2677)	1411	(1791)	929	(1106)	1224	(1399)
$\pm$ tot unc.	± 127	(± 24)	± 181	(± 29)	± 200	(± 26)	± 254	(± 38)
TotalSig	18	(6)	49	(17)	12	(4)	9	(3)
$\pm$ tot unc.	± 1	(± 1)	± 2	(± 5)	± 2	(± 3)	± 1	(± 2)
Data	2730		1795		1132		1370	

Table J.5: Prefit (post-fit) yields in the 3 jets, $\geq 3b$ tags category and ≥ 4 jets, $\geq 3b$ tags $t\bar{t}H$ and $t\bar{t} + \text{jets}$ nodes for the 2017 dataset for the inclusive fit to all channels. The prefit uncertainties are not including the freely-floating parameters; while the postfit uncertainties consider all nuisances in the fit to observed data.

Process	pre-fit (post-fit) yields							
	3j 3t		$4j \geq 3t$ ratio		$4j \geq 3t$ $t\bar{t} + c\bar{c}$		$4j \geq 3t$ $t\bar{t} + \text{LF}$	
$t\bar{t} + \text{LF}$	1234	(1335)	244	(284)	451	(499)	848	(932)
$t\bar{t} + C$	742	(780)	393	(418)	463	(487)	514	(507)
$t\bar{t} + b\bar{b}$	1287	(1890)	1515	(2022)	466	(622)	359	(488)
Single-t	103	(111)	59	(62)	29	(30)	43	(45)
V + jets	13	(14)	7	(7)	4	(4)	8	(9)
$t\bar{t}V$	23	(23)	45	(44)	19	(20)	16	(16)
$t\bar{t}\gamma^*$	18	(18)	15	(15)	10	(9)	16	(16)
Diboson	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
$t\bar{t}H$	1	(1)	2	(2)	0.4	(0.4)	0.4	(0.4)
TotalBkg	3421	(4171)	2279	(2854)	1441	(1672)	1806	(2014)
$\pm$ tot unc.	± 183	(± 37)	± 297	(± 40)	± 285	(± 39)	± 379	(± 46)
TotalSig	27	(10)	76	(27)	17	(6)	12	(4)
$\pm$ tot unc.	± 1	(± 2)	± 4	(± 7)	± 2	(± 4)	± 2	(± 3)
Data	4167		2842		1694		2033	

Table J.6: Prefit (post-fit) yields in the 3 jets, $\geq 3b$ tags category and ≥ 4 jets, $\geq 3b$ tags $t\bar{t}H$ and $t\bar{t} + \text{jets}$ nodes for the 2018 dataset for the inclusive fit to all channels. The prefit uncertainties are not including the freely-floating parameters; while the postfit uncertainties consider all nuisances in the fit to observed data.

J.3 STXS measurement - full Run 2 dilepton fit

Process	pre-fit (post-fit) yields					
	3j 3t		4j $\geq$ 3t $\bar{t}\bar{t}$ + $c\bar{c}$		4j $\geq$ 3t $\bar{t}\bar{t}$ + LF	
$\bar{t}\bar{t}$ + LF	433	(410)	183	(168)	263	(236)
$\bar{t}\bar{t}$ + C	322	(423)	239	(305)	171	(225)
$\bar{t}\bar{t}$ + B	626	(791)	233	(262)	138	(157)
Single-t	47	(45)	15	(14)	12	(11)
V + jets	5	(5)	4	(4)	3	(3)
$\bar{t}\bar{t}V$	8	(8)	10	(9)	7	(6)
$\bar{t}\bar{t}\gamma^*$	7	(7)	7	(6)	8	(8)
Diboson	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
tH	0.6	(0.6)	0.0	(0.0)	0.0	(0.0)
$\bar{t}\bar{t}H$ H p_T [0, 60[	4	(0.0)	4	(0.0)	2	(0.0)
$\bar{t}\bar{t}H$ H p_T [60, 120[	6	(0.0)	5	(0.0)	3	(0.0)
$\bar{t}\bar{t}H$ H p_T [120, 200[	4	(0.0)	4	(0.0)	2	(0.0)
$\bar{t}\bar{t}H$ H p_T [200, 300[	2	(0.0)	3	(0.0)	2	(0.0)
$\bar{t}\bar{t}H$ H p_T [300, 450[	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
$\bar{t}\bar{t}H$ H p_T [450, ∞ [	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
Total bkg.	1448	(1688)	690	(767)	603	(646)
$\pm$ tot unc.	± 335	(± 142)	± 143	(± 37)	± 131	(± 32)
TotalSig	17	(0.0)	21	(0.0)	13	(0.0)
$\pm$ tot unc.	± 3	(± 0.0)	± 4	(± 0.0)	± 3	(± 0.0)
Data	0.0		0.0		0.0	

Table J.7: Prefit (post-fit) yields in the 3 jets, ≥ 3 b tags category and ≥ 4 jets, ≥ 3 b tags $\bar{t}\bar{t}H$ and $\bar{t}\bar{t}$ + jets nodes for the 2016 dataset for the STXS fit to dilepton channel. The prefit uncertainties are not including the freely-floating parameters; while the postfit uncertainties consider all nuisances in the fit to observed data.

Process	pre-fit (post-fit) yields									
	STXS Bin 0		STXS Bin 1		STXS Bin 2		STXS Bin 3		STXS Bin 4	
$t\bar{t} + \text{LF}$	30	(26)	24	(23)	22	(19)	17	(13)	17	(13)
$t\bar{t} + \text{C}$	53	(68)	45	(58)	38	(47)	34	(45)	34	(45)
$t\bar{t} + \text{B}$	228	(285)	193	(237)	144	(169)	138	(154)	138	(154)
Single- t	5	(5)	6	(6)	4	(4)	5	(5)	5	(5)
$V + \text{jets}$	0.0	(0.0)	0.4	(0.4)	0.3	(0.3)	0.5	(0.4)	0.5	(0.4)
$t\bar{t}V$	6	(6)	4	(4)	3	(3)	7	(6)	7	(6)
$t\bar{t}\gamma^*$	3	(3)	2	(2)	3	(2)	1	(1)	1	(1)
Diboson	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.7	(0.5)	0.7	(0.5)
$t\bar{t}H$	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
$t\bar{t}H H p_T [0, 60[$	5	(0.0)	2	(0.0)	2	(0.0)	1	(0.0)	1	(0.0)
$t\bar{t}H H p_T [60, 120[$	5	(0.0)	5	(0.0)	3	(0.0)	2	(0.0)	2	(0.0)
$t\bar{t}H H p_T [120, 200[$	2	(0.0)	2	(0.0)	5	(0.0)	3	(0.0)	3	(0.0)
$t\bar{t}H H p_T [200, 300[$	1.0	(0.0)	0.9	(0.0)	1	(0.0)	5	(0.0)	5	(0.0)
$t\bar{t}H H p_T [300, 450[$	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
$t\bar{t}H H p_T [450, \infty[$	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
Total bkg.	324	(392)	276	(330)	214	(244)	202	(226)	202	(226)
$\pm \text{tot unc.}$	± 62	(± 20)	± 55	(± 20)	± 43	(± 12)	± 45	(± 13)	± 45	(± 13)
TotalSig	15	(0.0)	12	(0.0)	12	(0.0)	13	(0.0)	13	(0.0)
$\pm \text{tot unc.}$	± 2	(± 0.0)	± 2	(± 0.0)	± 2	(± 0.0)	± 2	(± 0.0)	± 2	(± 0.0)
Data	0.0		0.0		0.0		0.0		0.0	

Table J.8: (Continue) Prefit (post-fit) yields in the $3 \text{ jets}, \geq 3 b \text{ tags}$ category and $\geq 4 \text{ jets}, \geq 3 b \text{ tags } t\bar{t}H$ and $t\bar{t} + \text{jets}$ nodes for the 2016 dataset for the STXS fit to dilepton channel. The prefit uncertainties are not including the freely-floating parameters; while the postfit uncertainties consider all nuisances in the fit to observed data.

Process	pre-fit (post-fit) yields					
	3j 3t		$4j \geq 3t \text{ } \bar{t}\bar{t} + c\bar{c}$		$4j \geq 3t \text{ } \bar{t}\bar{t} + \text{LF}$	
$\bar{t}\bar{t} + \text{LF}$	722	(554)	263	(171)	528	(397)
$\bar{t}\bar{t} + \text{C}$	505	(938)	310	(539)	372	(637)
$\bar{t}\bar{t} + \text{B}$	859	(1152)	314	(370)	266	(309)
Single-t	60	(60)	18	(18)	30	(27)
V + jets	10	(10)	2	(2)	4	(4)
$\bar{t}\bar{t}\text{V}$	15	(15)	12	(12)	13	(13)
$\bar{t}\bar{t}\gamma^*$	11	(11)	9	(9)	11	(10)
Diboson	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
tH	0.8	(0.8)	0.0	(0.0)	0.2	(0.2)
$\bar{t}\bar{t}\text{H H } p_{\text{T}} [0, 60[$	5	(0.0)	6	(0.0)	5	(0.0)
$\bar{t}\bar{t}\text{H H } p_{\text{T}} [60, 120[$	8	(0.0)	7	(0.0)	6	(0.0)
$\bar{t}\bar{t}\text{H H } p_{\text{T}} [120, 200[$	6	(0.0)	6	(0.0)	5	(0.0)
$\bar{t}\bar{t}\text{H H } p_{\text{T}} [200, 300[$	3	(0.0)	5	(0.0)	4	(0.0)
$\bar{t}\bar{t}\text{H H } p_{\text{T}} [300, 450[$	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
$\bar{t}\bar{t}\text{H H } p_{\text{T}} [450, \infty[$	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
Total bkg.	2183	(2742)	928	(1121)	1224	(1399)
$\pm \text{ tot unc.}$	± 488	(± 174)	± 200	(± 45)	± 254	(± 61)
TotalSig	26	(0.0)	31	(0.0)	26	(0.0)
$\pm \text{ tot unc.}$	± 4	(± 0.0)	± 6	(± 0.0)	± 6	(± 0.0)
Data	0.0		0.0		0.0	

Table J.9: Prefit (post-fit) yields in the 3 jets, $\geq 3b$ tags category and ≥ 4 jets, $\geq 3b$ tags $\bar{t}\bar{t}\text{H}$ and $\bar{t}\bar{t} + \text{jets}$ nodes for the 2017 dataset for the STXS fit to dilepton channel. The prefit uncertainties are not including the freely-floating parameters; while the postfit uncertainties consider all nuisances in the fit to observed data.

Process	pre-fit (post-fit) yields									
	STXS Bin 0		STXS Bin 1		STXS Bin 2		STXS Bin 3		STXS Bin 4	
$t\bar{t} + \text{LF}$	48	(33)	31	(20)	24	(15)	20	(12)	20	(12)
$t\bar{t} + \text{C}$	62	(111)	52	(96)	44	(77)	39	(71)	39	(71)
$t\bar{t} + \text{B}$	298	(402)	233	(290)	176	(219)	165	(199)	165	(199)
Single- t	9	(9)	7	(7)	6	(6)	7	(7)	7	(7)
$V + \text{jets}$	1	(1)	0.4	(0.4)	0.7	(0.7)	0.6	(0.7)	0.6	(0.7)
$t\bar{t}V$	7	(7)	6	(6)	5	(6)	8	(8)	8	(8)
$t\bar{t}\gamma^*$	2	(3)	2	(2)	2	(2)	1	(1)	1	(1)
Diboson	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
$t\bar{t}H$	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.3	(0.3)	0.3	(0.3)
$t\bar{t}H H p_T [0, 60[$	6	(0.0)	2	(0.0)	2	(0.0)	2	(0.0)	2	(0.0)
$t\bar{t}H H p_T [60, 120[$	5	(0.0)	6	(0.0)	3	(0.0)	3	(0.0)	3	(0.0)
$t\bar{t}H H p_T [120, 200[$	3	(0.0)	3	(0.0)	6	(0.0)	4	(0.0)	4	(0.0)
$t\bar{t}H H p_T [200, 300[$	1	(0.0)	1	(0.0)	2	(0.0)	5	(0.0)	5	(0.0)
$t\bar{t}H H p_T [300, 450[$	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
$t\bar{t}H H p_T [450, \infty[$	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
Total bkg.	427	(566)	331	(422)	257	(325)	241	(300)	241	(300)
$\pm \text{tot unc.}$	± 83	(± 28)	± 67	(± 21)	± 53	(± 17)	± 55	(± 18)	± 55	(± 18)
TotalSig	18	(0.0)	14	(0.0)	15	(0.0)	16	(0.0)	16	(0.0)
$\pm \text{tot unc.}$	± 3	(± 0.0)	± 2	(± 0.0)	± 2	(± 0.0)	± 3	(± 0.0)	± 3	(± 0.0)
Data	0.0		0.0		0.0		0.0		0.0	

Table J.10: (Continue) Prefit (post-fit) yields in the 3 jets, $\geq 3 b$ tags category and ≥ 4 jets, $\geq 3 b$ tags $t\bar{t}H$ and $t\bar{t} + \text{jets}$ nodes for the 2017 dataset for the STXS fit to dilepton channel. The prefit uncertainties are not including the freely-floating parameters; while the postfit uncertainties consider all nuisances in the fit to observed data.

Process	pre-fit (post-fit) yields					
	3j 3t		$4j \geq 3t \text{ } \bar{t}\bar{t} + c\bar{c}$		$4j \geq 3t \text{ } \bar{t}\bar{t} + \text{LF}$	
$\bar{t}\bar{t} + \text{LF}$	1234	(969)	451	(327)	848	(674)
$\bar{t}\bar{t} + \text{C}$	742	(1349)	463	(784)	514	(867)
$\bar{t}\bar{t} + \text{B}$	1287	(1700)	466	(542)	359	(421)
Single-t	103	(100)	29	(28)	43	(42)
V + jets	13	(14)	4	(4)	8	(8)
$\bar{t}\bar{t}\text{V}$	23	(23)	19	(20)	16	(16)
$\bar{t}\bar{t}\gamma^*$	18	(18)	10	(9)	16	(14)
Diboson	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
tH	1	(1)	0.0	(0.0)	0.3	(0.3)
$\bar{t}\bar{t}\text{H H } p_{\text{T}} [0, 60[$	8	(0.0)	8	(0.0)	6	(0.0)
$\bar{t}\bar{t}\text{H H } p_{\text{T}} [60, 120[$	12	(0.0)	10	(0.0)	7	(0.0)
$\bar{t}\bar{t}\text{H H } p_{\text{T}} [120, 200[$	9	(0.0)	9	(0.0)	6	(0.0)
$\bar{t}\bar{t}\text{H H } p_{\text{T}} [200, 300[$	5	(0.0)	7	(0.0)	5	(0.0)
$\bar{t}\bar{t}\text{H H } p_{\text{T}} [300, 450[$	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
$\bar{t}\bar{t}\text{H H } p_{\text{T}} [450, \infty[$	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
Total bkg.	3421	(4175)	1441	(1713)	1806	(2043)
$\pm \text{ tot unc.}$	± 698	(± 280)	± 285	(± 62)	± 379	(± 75)
TotalSig	38	(0.0)	47	(0.0)	35	(0.0)
$\pm \text{ tot unc.}$	± 6	(± 0.0)	± 9	(± 0.0)	± 7	(± 0.0)
Data	0.0		0.0		0.0	

Table J.11: Prefit (post-fit) yields in the 3 jets, $\geq 3b$ tags category and ≥ 4 jets, $\geq 3b$ tags $\bar{t}\bar{t}\text{H}$ and $\bar{t}\bar{t} + \text{jets}$ nodes for the 2018 dataset for the STXS fit to dilepton channel. The prefit uncertainties are not including the freely-floating parameters; while the postfit uncertainties consider all nuisances in the fit to observed data.

Process	pre-fit (post-fit) yields									
	STXS Bin 0		STXS Bin 1		STXS Bin 2		STXS Bin 3		STXS Bin 4	
$t\bar{t} + \text{LF}$	83	(62)	57	(36)	46	(29)	33	(25)	33	(25)
$t\bar{t} + \text{C}$	119	(213)	93	(162)	73	(126)	66	(116)	66	(116)
$t\bar{t} + \text{B}$	448	(594)	363	(454)	266	(316)	264	(324)	264	(324)
Single- t	13	(12)	8	(10)	10	(10)	11	(10)	11	(10)
$V + \text{jets}$	2	(2)	1	(1)	0.6	(0.6)	2	(1)	2	(1)
$t\bar{t}V$	9	(9)	8	(9)	9	(10)	9	(8)	9	(8)
$t\bar{t}\gamma^*$	4	(5)	3	(3)	1	(1)	2	(2)	2	(2)
Diboson	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
$t\bar{t}H$	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
$t\bar{t}H H p_T [0, 60[$	10	(0.0)	4	(0.0)	3	(0.0)	3	(0.0)	3	(0.0)
$t\bar{t}H H p_T [60, 120[$	9	(0.0)	9	(0.0)	5	(0.0)	4	(0.0)	4	(0.0)
$t\bar{t}H H p_T [120, 200[$	5	(0.0)	5	(0.0)	10	(0.0)	6	(0.0)	6	(0.0)
$t\bar{t}H H p_T [200, 300[$	2	(0.0)	2	(0.0)	3	(0.0)	8	(0.0)	8	(0.0)
$t\bar{t}H H p_T [300, 450[$	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
$t\bar{t}H H p_T [450, \infty[$	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
Total bkg.	678	(897)	536	(677)	407	(492)	386	(488)	386	(488)
$\pm \text{tot unc.}$	± 130	(± 41)	± 109	(± 32)	± 88	(± 26)	± 82	(± 23)	± 82	(± 23)
TotalSig	28	(0.0)	22	(0.0)	23	(0.0)	24	(0.0)	24	(0.0)
$\pm \text{tot unc.}$	± 5	(± 0.0)	± 3	(± 0.0)	± 3	(± 0.0)	± 4	(± 0.0)	± 4	(± 0.0)
Data	0.0		0.0		0.0		0.0		0.0	

Table J.12: (Continue) Prefit (post-fit) yields in the 3 jets, $\geq 3 b$ tags category and ≥ 4 jets, $\geq 3 b$ tags $t\bar{t}H$ and $t\bar{t} + \text{jets}$ nodes for the 2018 dataset for the STXS fit to dilepton channel. The prefit uncertainties are not including the freely-floating parameters; while the postfit uncertainties consider all nuisances in the fit to observed data.

J.4 STXS measurement - full Run 2 all channels fit

Process	pre-fit (post-fit) yields					
	3j 3t		$4j \geq 3t \text{ } t\bar{t} + c\bar{c}$		$4j \geq 3t \text{ } t\bar{t} + \text{LF}$	
$t\bar{t} + \text{LF}$	433	(362)	183	(145)	263	(222)
$t\bar{t} + \text{C}$	322	(427)	239	(313)	171	(220)
$t\bar{t} + \text{B}$	626	(833)	233	(278)	138	(168)
Single-t	47	(45)	15	(14)	12	(11)
V + jets	5	(5)	4	(4)	3	(3)
$t\bar{t}\text{V}$	8	(7)	10	(9)	7	(6)
$t\bar{t}\gamma^*$	7	(6)	7	(6)	8	(7)
Diboson	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
tH	0.6	(0.5)	0.0	(0.0)	0.0	(0.0)
$t\bar{t}\text{H H } p_{\text{T}} [0, 60[$	4	(0.0)	4	(0.0)	2	(0.0)
$t\bar{t}\text{H H } p_{\text{T}} [60, 120[$	6	(0.0)	5	(0.0)	3	(0.0)
$t\bar{t}\text{H H } p_{\text{T}} [120, 200[$	4	(0.0)	4	(0.0)	2	(0.0)
$t\bar{t}\text{H H } p_{\text{T}} [200, 300[$	2	(0.0)	3	(0.0)	2	(0.0)
$t\bar{t}\text{H H } p_{\text{T}} [300, 450[$	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
$t\bar{t}\text{H H } p_{\text{T}} [450, \infty[$	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
Total bkg.	1448	(1687)	690	(769)	603	(638)
$\pm \text{ tot unc.}$	± 335	(± 45)	± 143	(± 23)	± 131	(± 19)
TotalSig	17	(0.0)	21	(0.0)	13	(0.0)
$\pm \text{ tot unc.}$	± 3	(± 0.0)	± 4	(± 0.0)	± 3	(± 0.0)
Data	0.0		0.0		0.0	

Table J.13: Prefit (post-fit) yields in the 3 jets, ≥ 3 b tags category and ≥ 4 jets, ≥ 3 b tags $t\bar{t}\text{H}$ and $t\bar{t} + \text{jets}$ nodes for the 2016 dataset for the STXS fit to all channels. The prefit uncertainties are not including the freely-floating parameters; while the postfit uncertainties consider all nuisances in the fit to observed data.

Process	pre-fit (post-fit) yields									
	STXS Bin 0		STXS Bin 1		STXS Bin 2		STXS Bin 3		STXS Bin 4	
$t\bar{t} + \text{LF}$	30	(24)	24	(19)	22	(16)	17	(13)	17	(13)
$t\bar{t} + \text{C}$	53	(73)	45	(60)	38	(53)	34	(41)	34	(41)
$t\bar{t} + \text{B}$	228	(277)	193	(233)	144	(172)	138	(164)	138	(164)
Single- t	5	(5)	6	(6)	4	(4)	5	(5)	5	(5)
$V + \text{jets}$	0.0	(0.0)	0.4	(0.4)	0.3	(0.3)	0.5	(0.4)	0.5	(0.4)
$t\bar{t}V$	6	(5)	4	(4)	3	(3)	7	(6)	7	(6)
$t\bar{t}\gamma^*$	3	(2)	2	(2)	3	(2)	1	(1.0)	1	(1.0)
Diboson	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.7	(0.4)	0.7	(0.4)
$t\bar{t}H$	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
$t\bar{t}H H p_T [0, 60[$	5	(0.0)	2	(0.0)	2	(0.0)	1	(0.0)	1	(0.0)
$t\bar{t}H H p_T [60, 120[$	5	(0.0)	5	(0.0)	3	(0.0)	2	(0.0)	2	(0.0)
$t\bar{t}H H p_T [120, 200[$	2	(0.0)	2	(0.0)	5	(0.0)	3	(0.0)	3	(0.0)
$t\bar{t}H H p_T [200, 300[$	1.0	(0.0)	0.9	(0.0)	1	(0.0)	5	(0.0)	5	(0.0)
$t\bar{t}H H p_T [300, 450[$	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
$t\bar{t}H H p_T [450, \infty[$	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
Total bkg.	324	(387)	276	(325)	214	(250)	202	(231)	202	(231)
$\pm \text{tot unc.}$	± 62	(± 13)	± 55	(± 10)	± 43	(± 8)	± 45	(± 9)	± 45	(± 9)
TotalSig	15	(0.0)	12	(0.0)	12	(0.0)	13	(0.0)	13	(0.0)
$\pm \text{tot unc.}$	± 2	(± 0.0)	± 2	(± 0.0)	± 2	(± 0.0)	± 2	(± 0.0)	± 2	(± 0.0)
Data	0.0		0.0		0.0		0.0		0.0	

Table J.14: (Continue) Prefit (post-fit) yields in the 3 jets, $\geq 3 b$ tags category and ≥ 4 jets, $\geq 3 b$ tags $t\bar{t}H$ and $t\bar{t} + \text{jets}$ nodes for the 2016 dataset for the STXS fit to all channels. The prefit uncertainties are not including the freely-floating parameters; while the postfit uncertainties consider all nuisances in the fit to observed data.

Process	pre-fit (post-fit) yields					
	3j 3t		$4j \geq 3t \bar{t}\bar{t} + c\bar{c}$		$4j \geq 3t \bar{t}\bar{t} + \text{LF}$	
$\bar{t}\bar{t} + \text{LF}$	722	(676)	263	(235)	528	(501)
$\bar{t}\bar{t} + \text{C}$	505	(672)	310	(418)	372	(482)
$\bar{t}\bar{t} + \text{B}$	859	(1247)	314	(408)	266	(352)
Single-t	60	(63)	18	(20)	30	(31)
V + jets	10	(11)	2	(2)	4	(4)
$\bar{t}\bar{t}\text{V}$	15	(14)	12	(11)	13	(12)
$\bar{t}\bar{t}\gamma^*$	11	(10)	9	(8)	11	(10)
Diboson	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
tH	0.8	(0.8)	0.0	(0.0)	0.2	(0.2)
$\bar{t}\bar{t}\text{H H } p_{\text{T}} [0, 60[$	5	(0.0)	6	(0.0)	5	(0.0)
$\bar{t}\bar{t}\text{H H } p_{\text{T}} [60, 120[$	8	(0.0)	7	(0.0)	6	(0.0)
$\bar{t}\bar{t}\text{H H } p_{\text{T}} [120, 200[$	6	(0.0)	6	(0.0)	5	(0.0)
$\bar{t}\bar{t}\text{H H } p_{\text{T}} [200, 300[$	3	(0.0)	5	(0.0)	4	(0.0)
$\bar{t}\bar{t}\text{H H } p_{\text{T}} [300, 450[$	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
$\bar{t}\bar{t}\text{H H } p_{\text{T}} [450, \infty[$	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
Total bkg.	2183	(2695)	928	(1102)	1224	(1393)
$\pm$ tot unc.	± 488	(± 61)	± 200	(± 27)	± 254	(± 34)
TotalSig	26	(0.0)	31	(0.0)	26	(0.0)
$\pm$ tot unc.	± 4	(± 0.0)	± 6	(± 0.0)	± 6	(± 0.0)
Data	0.0		0.0		0.0	

Table J.15: Prefit (post-fit) yields in the 3 jets, ≥ 3 b tags category and ≥ 4 jets, ≥ 3 b tags $\bar{t}\bar{t}\text{H}$ and $\bar{t}\bar{t} + \text{jets}$ nodes for the 2017 dataset for the STXS fit to all channels. The prefit uncertainties are not including the freely-floating parameters; while the postfit uncertainties consider all nuisances in the fit to observed data.

Process	pre-fit (post-fit) yields									
	STXS Bin 0		STXS Bin 1		STXS Bin 2		STXS Bin 3		STXS Bin 4	
$t\bar{t} + \text{LF}$	48	(44)	31	(29)	24	(19)	20	(17)	20	(17)
$t\bar{t} + \text{C}$	62	(91)	52	(73)	44	(63)	39	(49)	39	(49)
$t\bar{t} + \text{B}$	298	(395)	233	(305)	176	(232)	165	(216)	165	(216)
Single- t	9	(9)	7	(7)	6	(6)	7	(7)	7	(7)
$\text{V} + \text{jets}$	1	(1)	0.4	(0.4)	0.7	(0.8)	0.6	(0.7)	0.6	(0.7)
$t\bar{t}\text{V}$	7	(7)	6	(6)	5	(5)	8	(7)	8	(7)
$t\bar{t}\gamma^*$	2	(2)	2	(2)	2	(2)	1	(1)	1	(1)
Diboson	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
$t\bar{t}\text{H}$	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.3	(0.3)	0.3	(0.3)
$t\bar{t}\text{H H } p_{\text{T}} [0, 60[$	6	(0.0)	2	(0.0)	2	(0.0)	2	(0.0)	2	(0.0)
$t\bar{t}\text{H H } p_{\text{T}} [60, 120[$	5	(0.0)	6	(0.0)	3	(0.0)	3	(0.0)	3	(0.0)
$t\bar{t}\text{H H } p_{\text{T}} [120, 200[$	3	(0.0)	3	(0.0)	6	(0.0)	4	(0.0)	4	(0.0)
$t\bar{t}\text{H H } p_{\text{T}} [200, 300[$	1	(0.0)	1	(0.0)	2	(0.0)	5	(0.0)	5	(0.0)
$t\bar{t}\text{H H } p_{\text{T}} [300, 450[$	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
$t\bar{t}\text{H H } p_{\text{T}} [450, \infty[$	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
Total bkg.	427	(551)	331	(423)	257	(327)	241	(299)	241	(299)
$\pm \text{tot unc.}$	± 83	(± 15)	± 67	(± 13)	± 53	(± 10)	± 55	(± 10)	± 55	(± 10)
TotalSig	18	(0.0)	14	(0.0)	15	(0.0)	16	(0.0)	16	(0.0)
$\pm \text{tot unc.}$	± 3	(± 0.0)	± 2	(± 0.0)	± 2	(± 0.0)	± 3	(± 0.0)	± 3	(± 0.0)
Data	0.0		0.0		0.0		0.0		0.0	

Table J.16: (Continue) Prefit (post-fit) yields in the 3 jets, $\geq 3 b$ tags category and ≥ 4 jets, $\geq 3 b$ tags $t\bar{t}\text{H}$ and $t\bar{t} + \text{jets}$ nodes for the 2017 dataset for the STXS fit to all channels. The prefit uncertainties are not including the freely-floating parameters; while the postfit uncertainties consider all nuisances in the fit to observed data.

Process	pre-fit (post-fit) yields					
	3j 3t		$4j \geq 3t \bar{t}\bar{t} + c\bar{c}$		$4j \geq 3t \bar{t}\bar{t} + \text{LF}$	
$\bar{t}\bar{t} + \text{LF}$	1234	(1137)	451	(396)	848	(797)
$\bar{t}\bar{t} + \text{C}$	742	(999)	463	(615)	514	(663)
$\bar{t}\bar{t} + \text{B}$	1287	(1863)	466	(602)	359	(474)
Single-t	103	(108)	29	(29)	43	(45)
V + jets	13	(14)	4	(4)	8	(8)
$\bar{t}\bar{t}\text{V}$	23	(22)	19	(19)	16	(15)
$\bar{t}\bar{t}\gamma^*$	18	(17)	10	(9)	16	(14)
Diboson	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
tH	1	(1)	0.0	(0.0)	0.3	(0.3)
$\bar{t}\bar{t}\text{H H } p_{\text{T}} [0, 60[$	8	(0.0)	8	(0.0)	6	(0.0)
$\bar{t}\bar{t}\text{H H } p_{\text{T}} [60, 120[$	12	(0.0)	10	(0.0)	7	(0.0)
$\bar{t}\bar{t}\text{H H } p_{\text{T}} [120, 200[$	9	(0.0)	9	(0.0)	6	(0.0)
$\bar{t}\bar{t}\text{H H } p_{\text{T}} [200, 300[$	5	(0.0)	7	(0.0)	5	(0.0)
$\bar{t}\bar{t}\text{H H } p_{\text{T}} [300, 450[$	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
$\bar{t}\bar{t}\text{H H } p_{\text{T}} [450, \infty[$	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
Total bkg.	3421	(4160)	1441	(1672)	1806	(2017)
$\pm$ tot unc.	± 698	(± 87)	± 285	(± 39)	± 379	(± 43)
TotalSig	38	(0.0)	47	(0.0)	35	(0.0)
$\pm$ tot unc.	± 6	(± 0.0)	± 9	(± 0.0)	± 7	(± 0.0)
Data	0.0		0.0		0.0	

Table J.17: Prefit (post-fit) yields in the 3 jets, ≥ 3 b tags category and ≥ 4 jets, ≥ 3 b tags $\bar{t}\bar{t}\text{H}$ and $\bar{t}\bar{t} + \text{jets}$ nodes for the 2018 dataset for the STXS fit to all channels. The prefit uncertainties are not including the freely-floating parameters; while the postfit uncertainties consider all nuisances in the fit to observed data.

Process	pre-fit (post-fit) yields									
	STXS Bin 0		STXS Bin 1		STXS Bin 2		STXS Bin 3		STXS Bin 4	
$t\bar{t} + \text{LF}$	83	(76)	57	(51)	46	(37)	33	(29)	33	(29)
$t\bar{t} + \text{C}$	119	(169)	93	(130)	73	(106)	66	(79)	66	(79)
$t\bar{t} + \text{B}$	448	(593)	363	(478)	266	(345)	264	(351)	264	(351)
Single- t	13	(13)	8	(9)	10	(10)	11	(12)	11	(12)
$V + \text{jets}$	2	(2)	1	(1)	0.6	(0.6)	2	(2)	2	(2)
$t\bar{t}V$	9	(9)	8	(8)	9	(9)	9	(8)	9	(8)
$t\bar{t}\gamma^*$	4	(4)	3	(3)	1	(1)	2	(2)	2	(2)
Diboson	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
$t\bar{t}H$	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
$t\bar{t}H H p_T [0, 60[$	10	(0.0)	4	(0.0)	3	(0.0)	3	(0.0)	3	(0.0)
$t\bar{t}H H p_T [60, 120[$	9	(0.0)	9	(0.0)	5	(0.0)	4	(0.0)	4	(0.0)
$t\bar{t}H H p_T [120, 200[$	5	(0.0)	5	(0.0)	10	(0.0)	6	(0.0)	6	(0.0)
$t\bar{t}H H p_T [200, 300[$	2	(0.0)	2	(0.0)	3	(0.0)	8	(0.0)	8	(0.0)
$t\bar{t}H H p_T [300, 450[$	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
$t\bar{t}H H p_T [450, \infty[$	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
Total bkg.	678	(866)	536	(681)	407	(508)	386	(482)	386	(482)
$\pm \text{tot unc.}$	± 130	(± 23)	± 109	(± 20)	± 88	(± 15)	± 82	(± 15)	± 82	(± 15)
TotalSig	28	(0.0)	22	(0.0)	23	(0.0)	24	(0.0)	24	(0.0)
$\pm \text{tot unc.}$	± 5	(± 0.0)	± 3	(± 0.0)	± 3	(± 0.0)	± 4	(± 0.0)	± 4	(± 0.0)
Data	0.0		0.0		0.0		0.0		0.0	

Table J.18: (Continue) Prefit (post-fit) yields in the 3 jets, $\geq 3 b$ tags category and ≥ 4 jets, $\geq 3 b$ tags $t\bar{t}H$ and $t\bar{t} + \text{jets}$ nodes for the 2018 dataset for the STXS fit to all channels. The prefit uncertainties are not including the freely-floating parameters; while the postfit uncertainties consider all nuisances in the fit to observed data.

PRE/POST-FIT SHAPE DISTRIBUTIONS



K.1 Inclusive measurement - full Run 2 dilepton fit

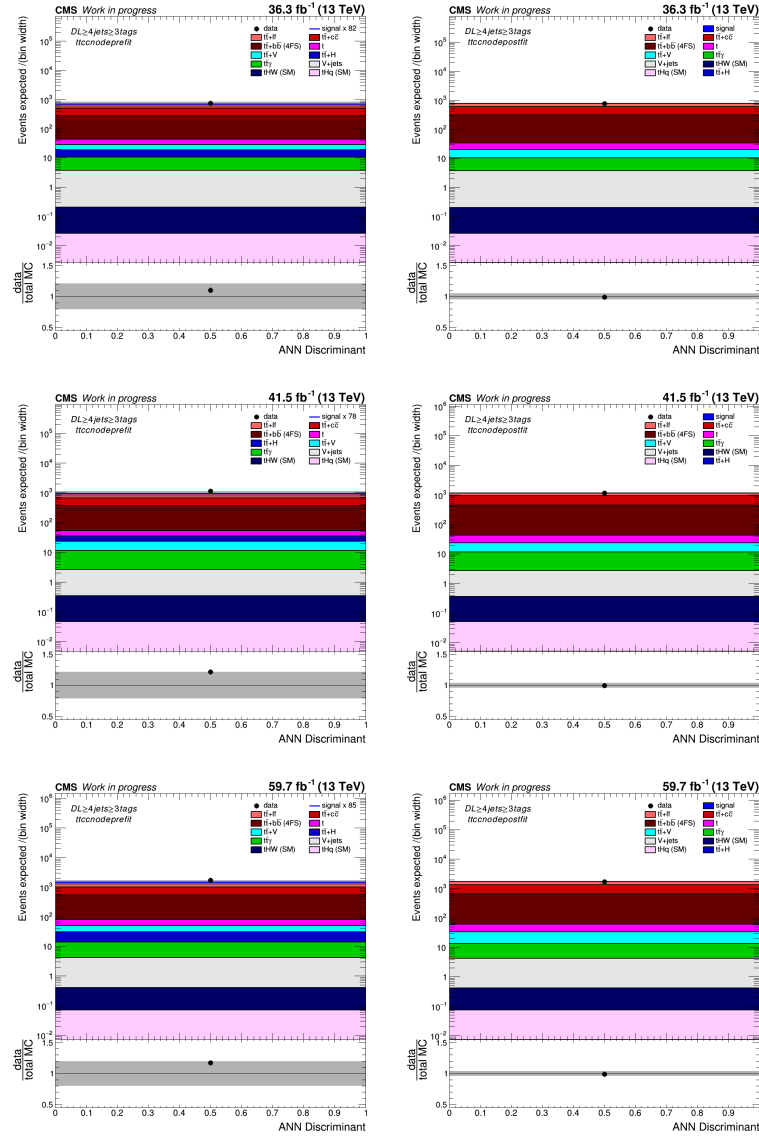


Figure K.1: Prefit (left) and postfit (right) discriminants distributions in the DL channel in the ≥ 4 jets, ≥ 3 b tags $t\bar{t} + c\bar{c}$ node category for 2016 (top), 2017 (middle) and 2018 (bottom). The black points are observed data; the signal template is scaled to the total background contribution in the pre-fit plots; and the stacked distributions represent the simulations for the respective processes. The uncertainty band corresponds to the uncertainties in the full Run 2 DL combination fit.

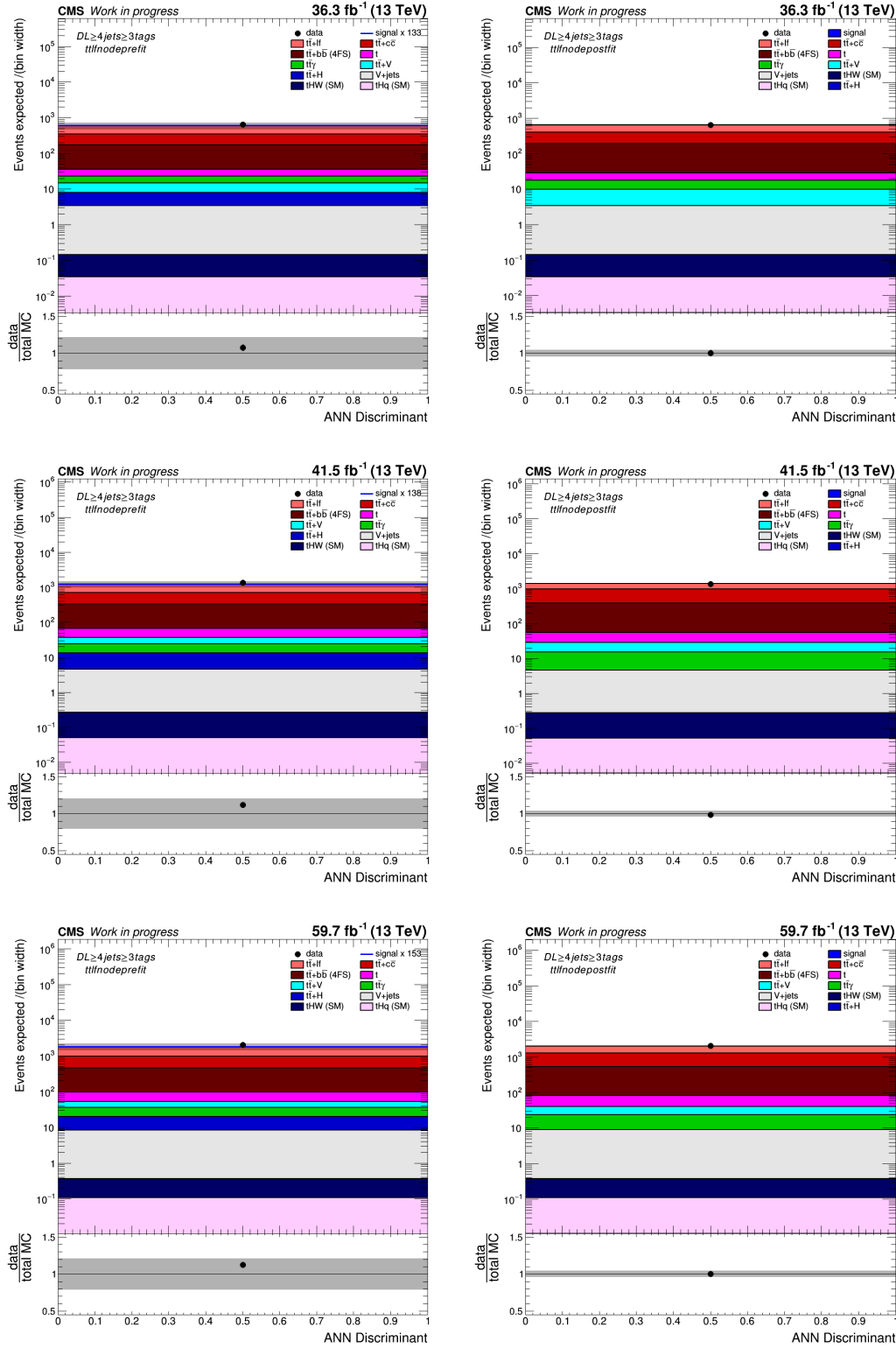


Figure K.2: Prefit (left) and postfit (right) discriminants distributions in the DL channel in the ≥ 4 jets, ≥ 3 b tags $t\bar{t}$ + LF node category for 2016 (top), 2017 (middle) and 2018 (bottom). The black points are observed data; the signal template is scaled to the total background contribution in the pre-fit plots; and the stacked distributions represent the simulations for the respective processes. The uncertainty band corresponds to the uncertainties in the full Run 2 DL combination fit.

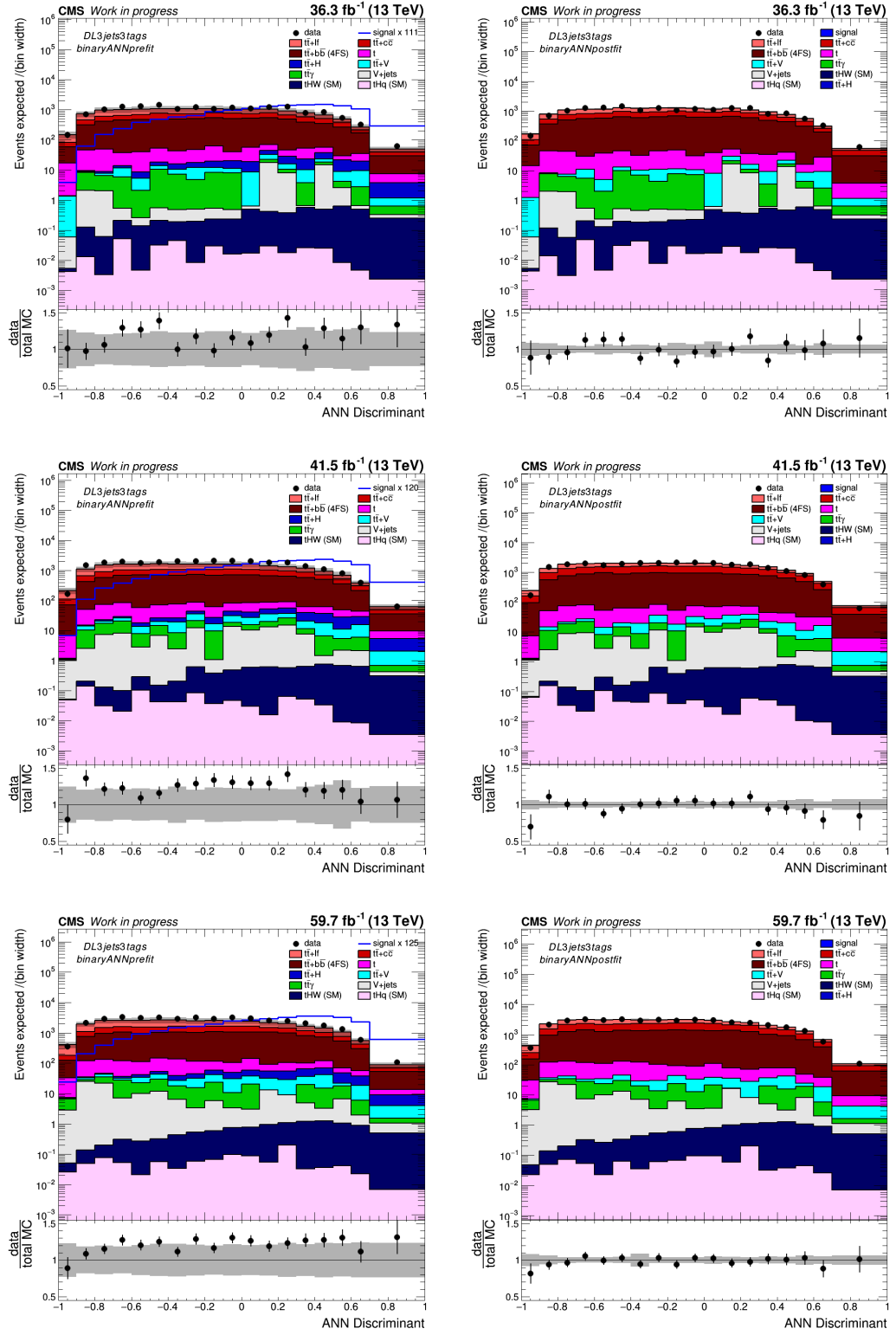


Figure K.3: Prefit (left) and postfit (right) discriminants distributions in the DL channel in the 3 jets, $\geq 3b$ tags node category for 2016 (top), 2017 (middle) and 2018 (bottom). The black points are observed data; the signal template is scaled to the total background contribution in the pre-fit plots; and the stacked distributions represent the simulations for the respective processes. The uncertainty band corresponds to the uncertainties in the full Run 2 DL combination fit.

K.2 Inclusive measurement - full Run 2 all channels fit

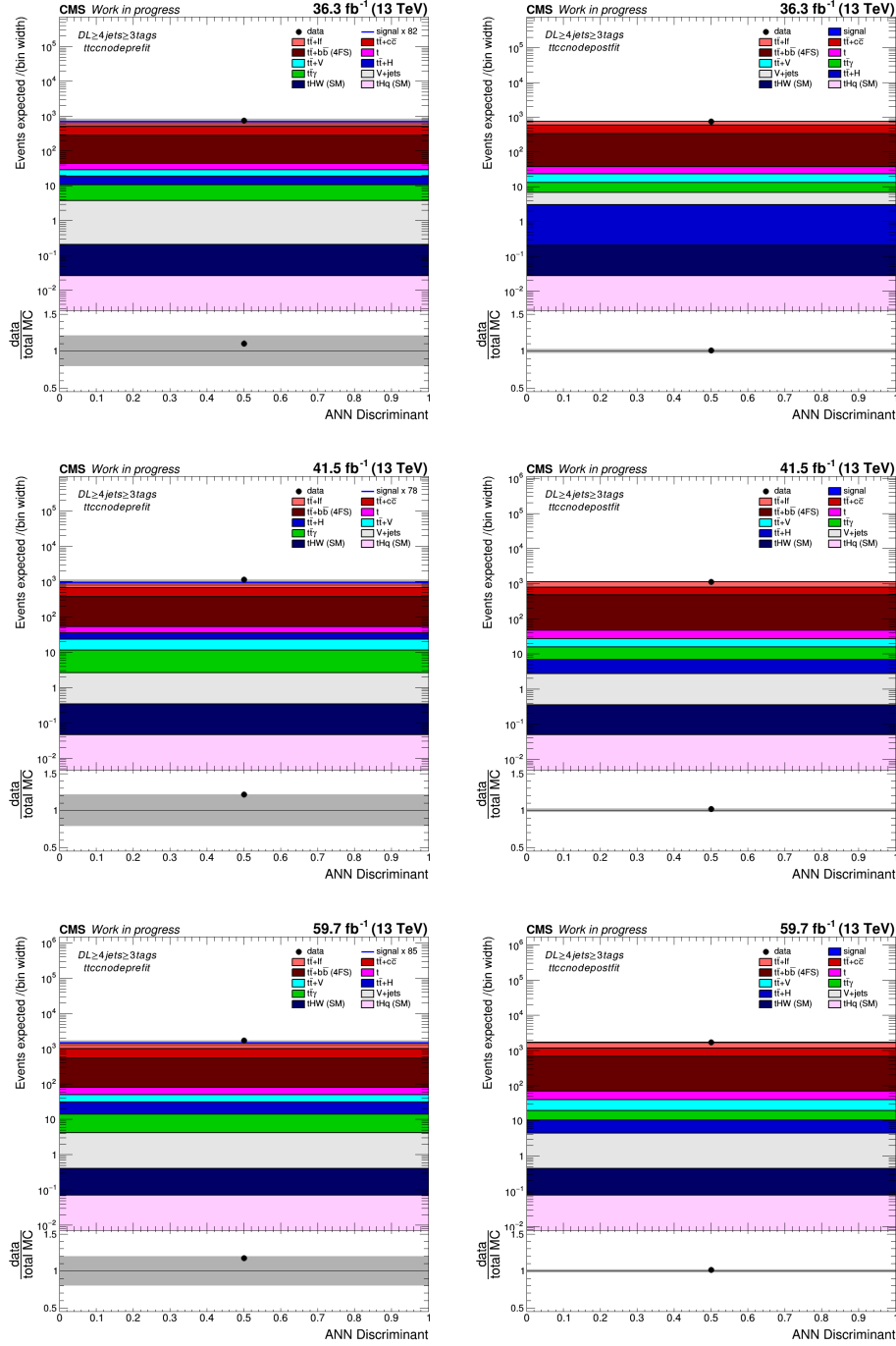


Figure K.4: Prefit (left) and postfit (right) discriminants distributions in the DL channel in the ≥ 4 jets, ≥ 3 b tags $t\bar{t} + c\bar{c}$ node category for 2016 (top), 2017 (middle) and 2018 (bottom) for the inclusive fit to all channels. The black points are observed data; the signal template is scaled to the total background contribution in the pre-fit plots; and the stacked distributions represent the simulations for the respective processes. The uncertainty band corresponds to the uncertainties in the full Run 2 combination fit to all channels.

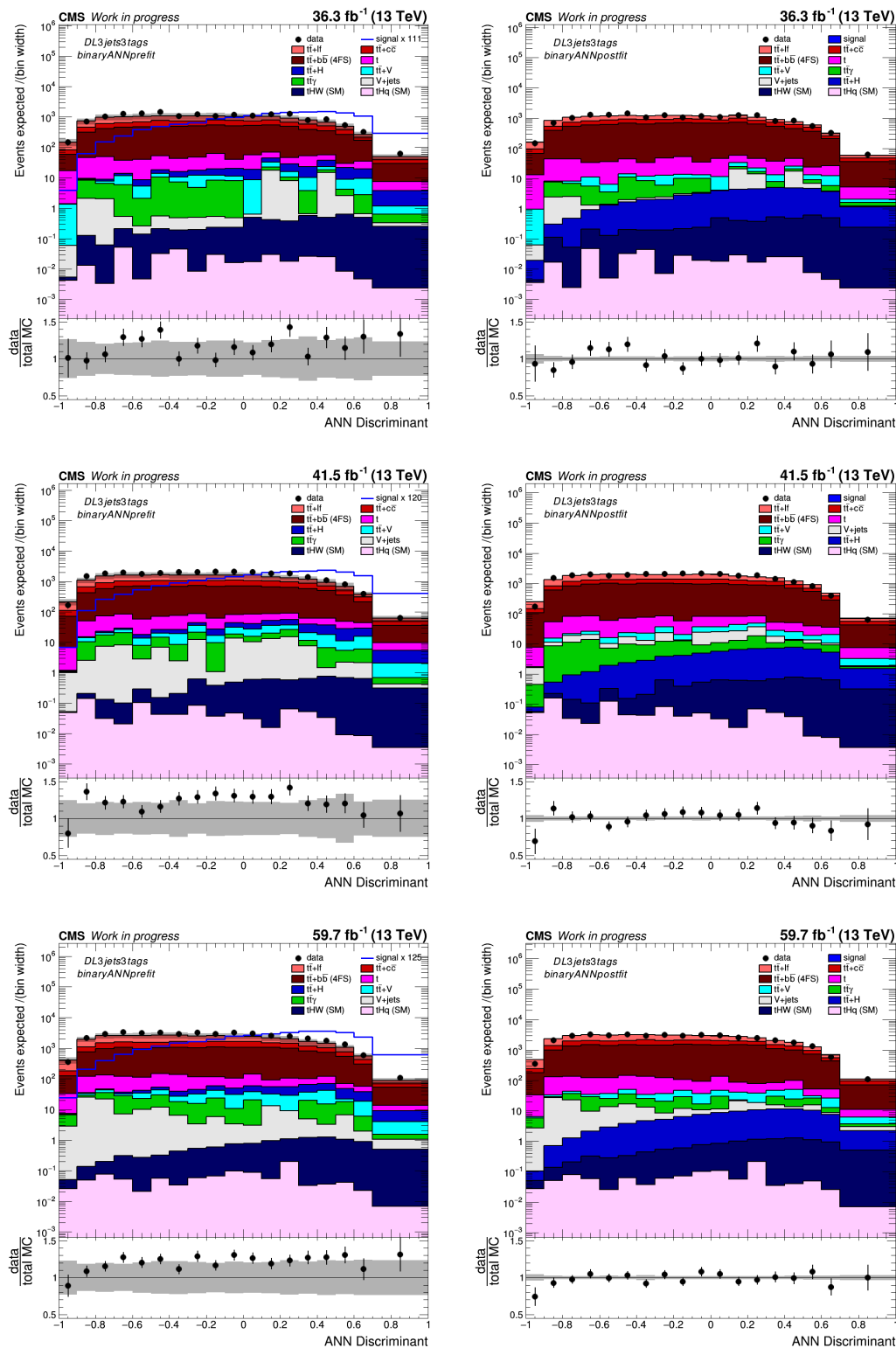


Figure K.6: Prefit (left) and postfit (right) discriminants distributions in the DL channel in the ≥ 4 jets, ≥ 3 b tags $t\bar{t} + c\bar{c}$ node category for 2016 (top), 2017 (middle) and 2018 (bottom) for the inclusive fit to all channels. The black points are observed data; the signal template is scaled to the total background contribution in the pre-fit plots; and the stacked distributions represent the simulations for the respective processes. The uncertainty band corresponds to the uncertainties in the full Run 2 combination fit to all channels.

K.3 STXS measurement - full Run 2 dilepton fit

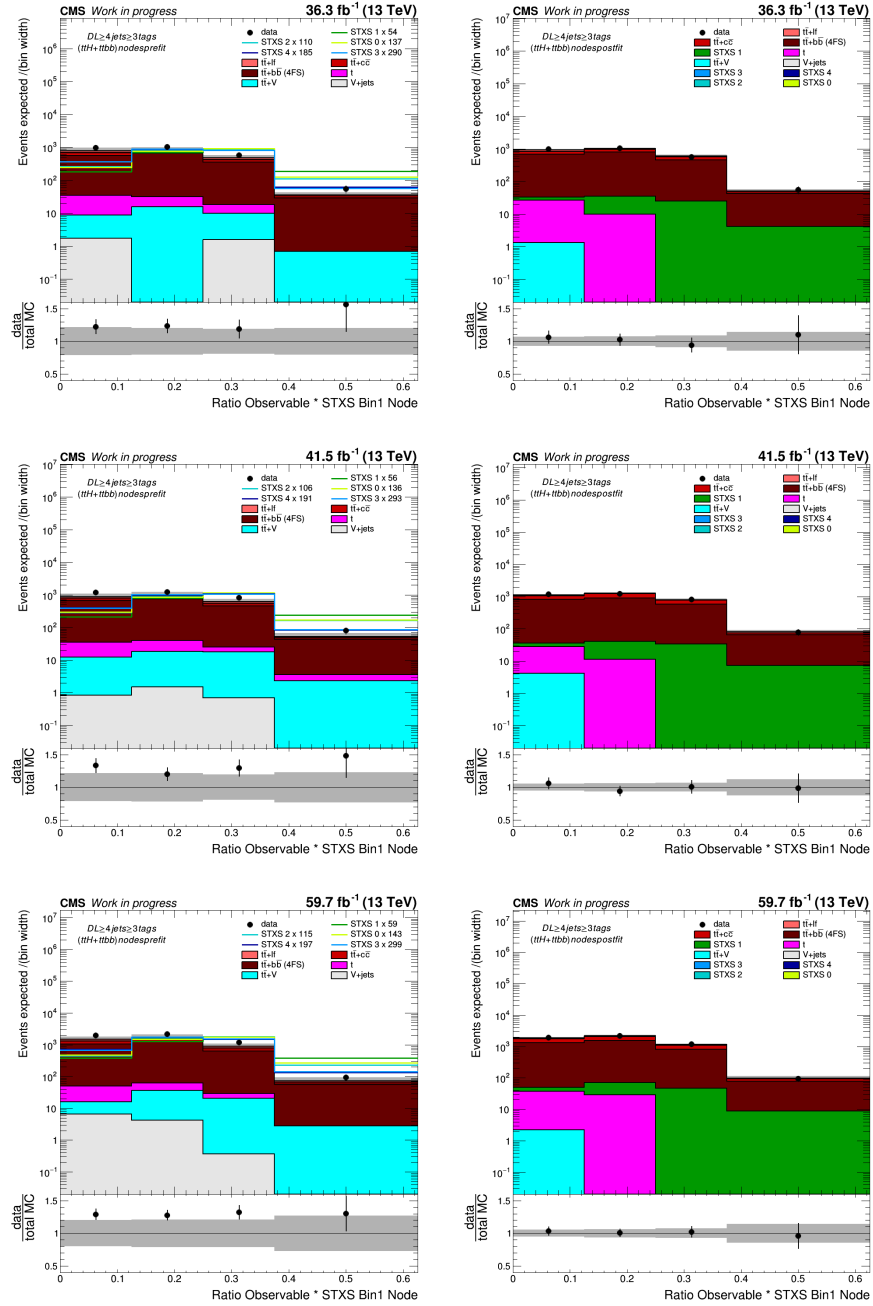


Figure K.7: Discriminants in the DL channel in the ≥ 4 jets, ≥ 3 b tags STXS bin 1 category. Shown are the pre-fit (left) and post-fit (right) distributions for 2016 (top), 2017 (middle), and 2018 (bottom). The black points represent the observed data. The stacked distributions correspond to the simulations for the respective processes. The uncertainty band corresponds to the uncertainties in the STXS full Run 2 DL combination fit. For better visibility, the signal template is scaled to the total background contribution in the pre-fit figures (cyan line).

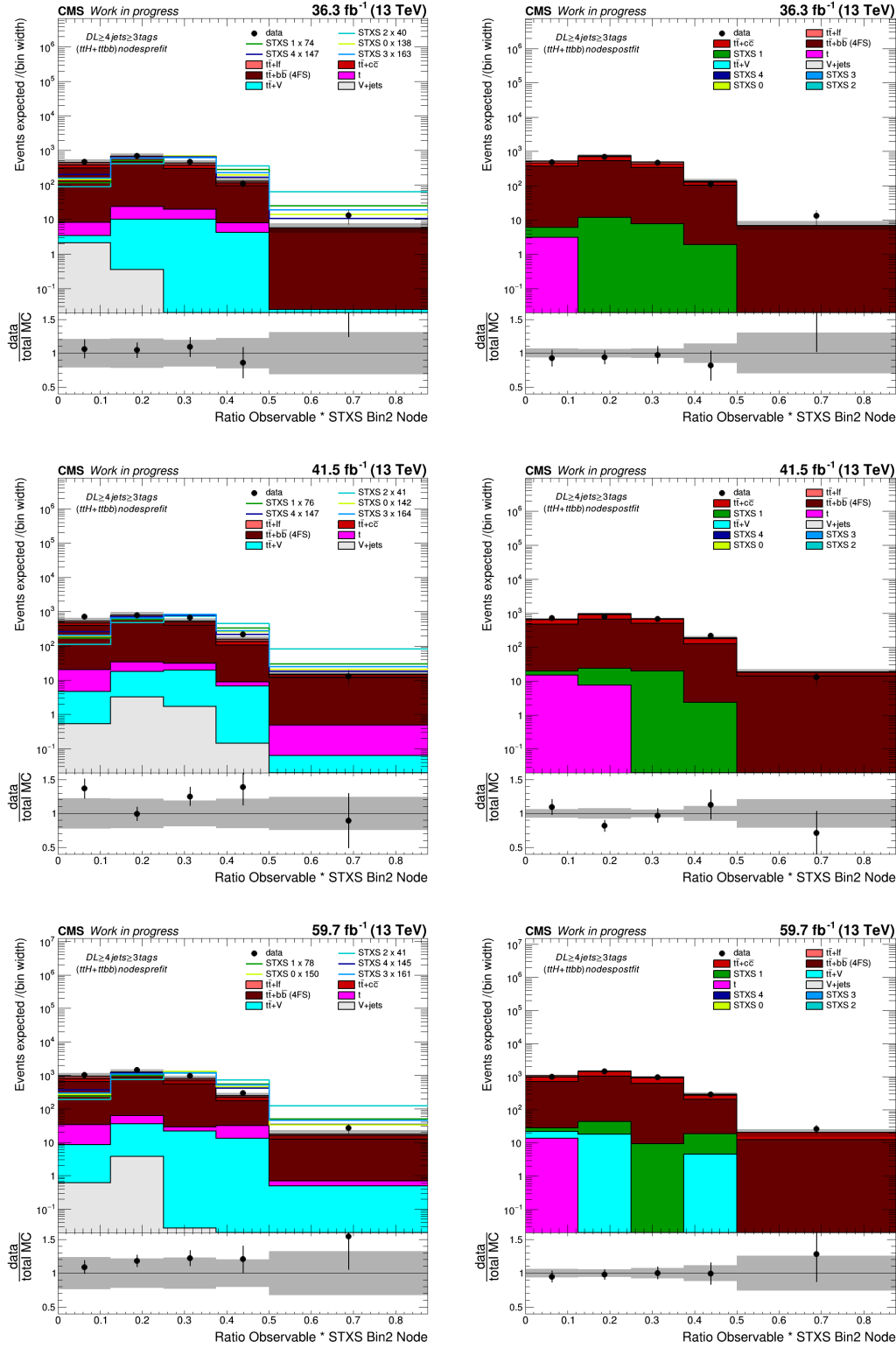


Figure K.8: Discriminants in the DL channel in the ≥ 4 jets, ≥ 3 b tags STXS bin 2 category. Shown are the pre-fit (left) and post-fit (right) distributions for 2016 (top), 2017 (middle), and 2018 (bottom). The black points represent the observed data. The stacked distributions correspond to the simulations for the respective processes. The uncertainty band corresponds to the uncertainties in the STXS full Run 2 DL combination fit. For better visibility, the signal template is scaled to the total background contribution in the pre-fit figures (cyan line).

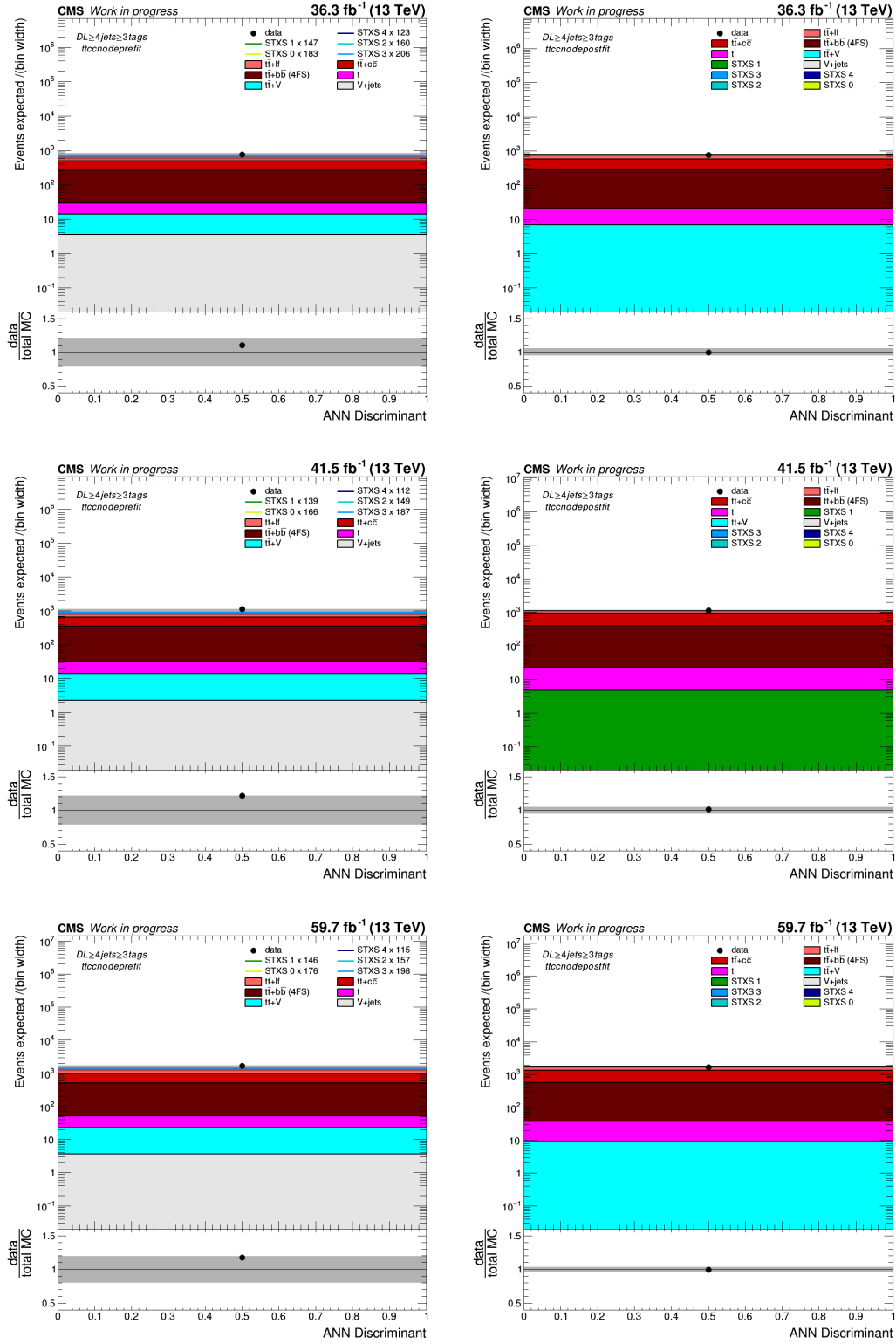


Figure K.10: Discriminants in the DL channel in the ≥ 4 jets, ≥ 3 b tags $t\bar{t} + c\bar{c}$ node category. Shown are the pre-fit (left) and post-fit (right) distributions for 2016 (top), 2017 (middle), and 2018 (bottom). The black points represent the observed data. The stacked distributions correspond to the simulations for the respective processes. The uncertainty band corresponds to the uncertainties in the STXS full Run 2 DL combination fit. For better visibility, the signal template is scaled to the total background contribution in the pre-fit figures (cyan line).

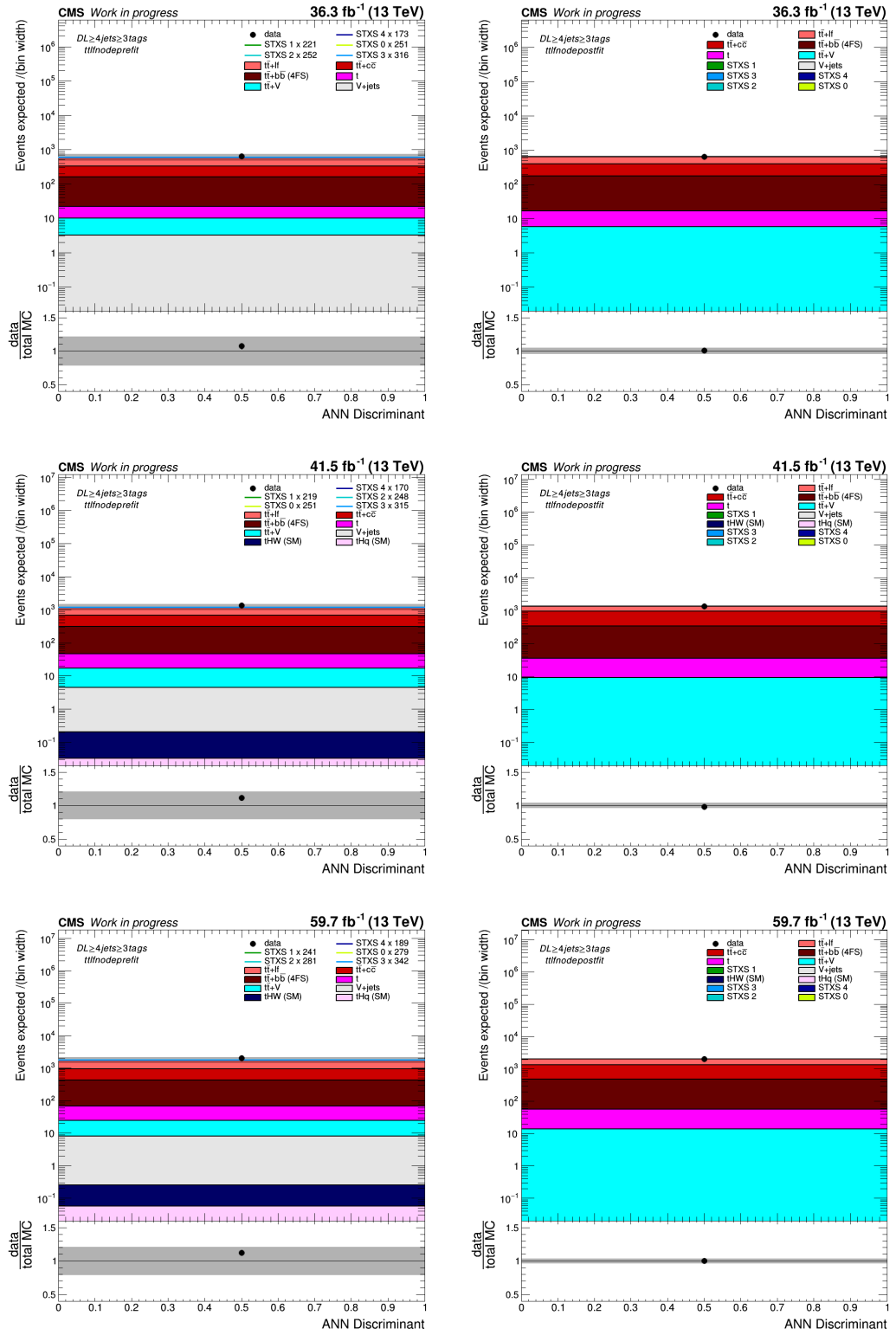


Figure K.11: Discriminants in the DL channel in the ≥ 4 jets, ≥ 3 b tags $t\bar{t}$ + LF node category. Shown are the pre-fit (left) and post-fit (right) distributions for 2016 (top), 2017 (middle), and 2018 (bottom). The black points represent the observed data. The stacked distributions correspond to the simulations for the respective processes. The uncertainty band corresponds to the uncertainties in the STXS full Run 2 DL combination fit. For better visibility, the signal template is scaled to the total background contribution in the pre-fit figures (cyan line).

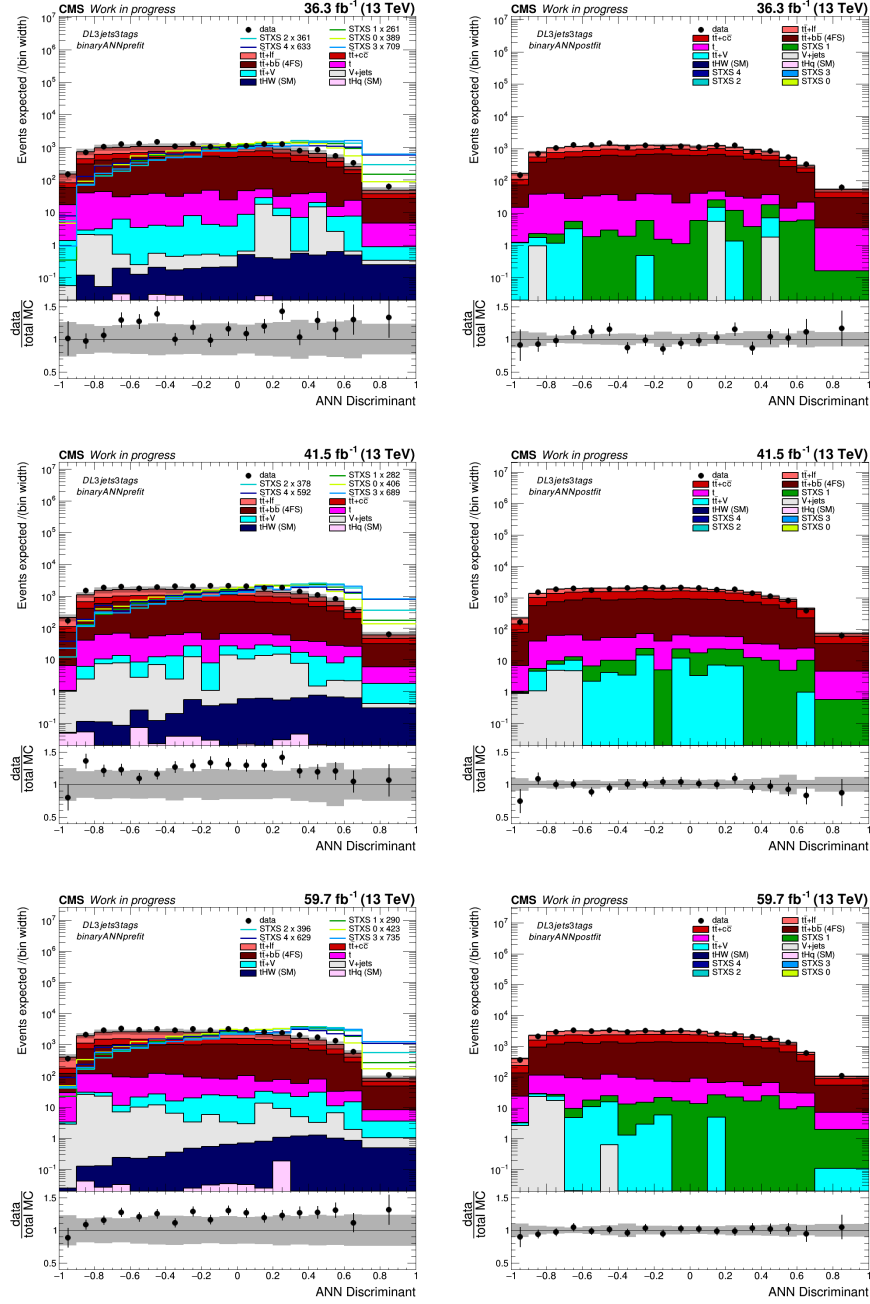


Figure K.12: Discriminants in the DL channel in the 3jets, $\geq 3b$ tags category. Shown are the pre-fit (left) and post-fit (right) distributions for 2016 (top), 2017 (middle), and 2018 (bottom). The black points represent the observed data. The stacked distributions correspond to the simulations for the respective processes. The uncertainty band corresponds to the uncertainties in the STXS full Run 2 DL combination fit. For better visibility, the signal template is scaled to the total background contribution in the pre-fit figures (cyan line).

K.4 STXS measurement - full Run 2 all channels fit

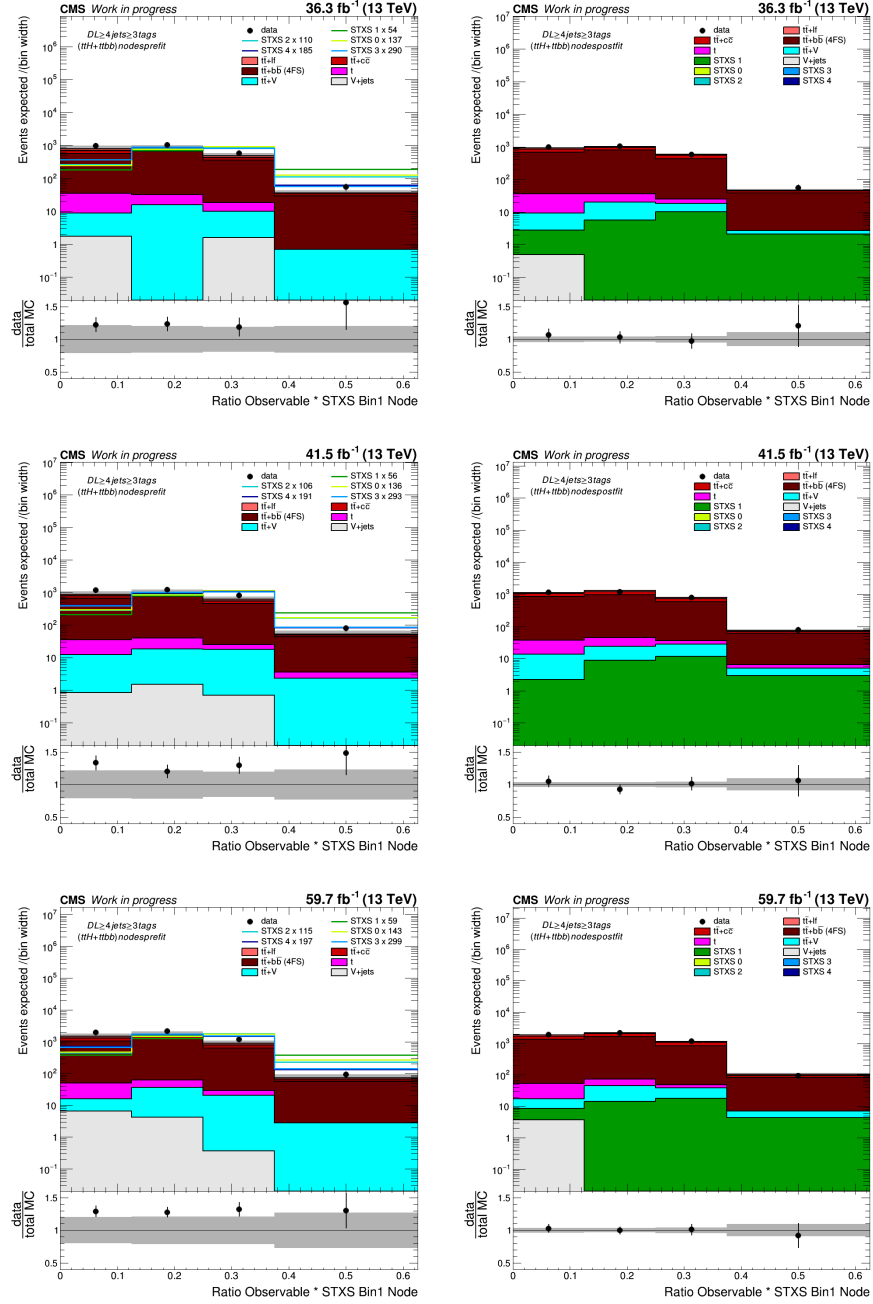


Figure K.13: Discriminants in the DL channel in the ≥ 4 jets, ≥ 3 b tags STXS bin 1 category. Shown are the pre-fit (left) and post-fit (right) distributions for 2016 (top), 2017 (middle), and 2018 (bottom). The black points represent the observed data. The stacked distributions correspond to the simulations for the respective processes. The uncertainty band corresponds to the uncertainties in the STXS full Run 2 fit to all channels. For better visibility, the signal template is scaled to the total background contribution in the pre-fit figures (cyan line).

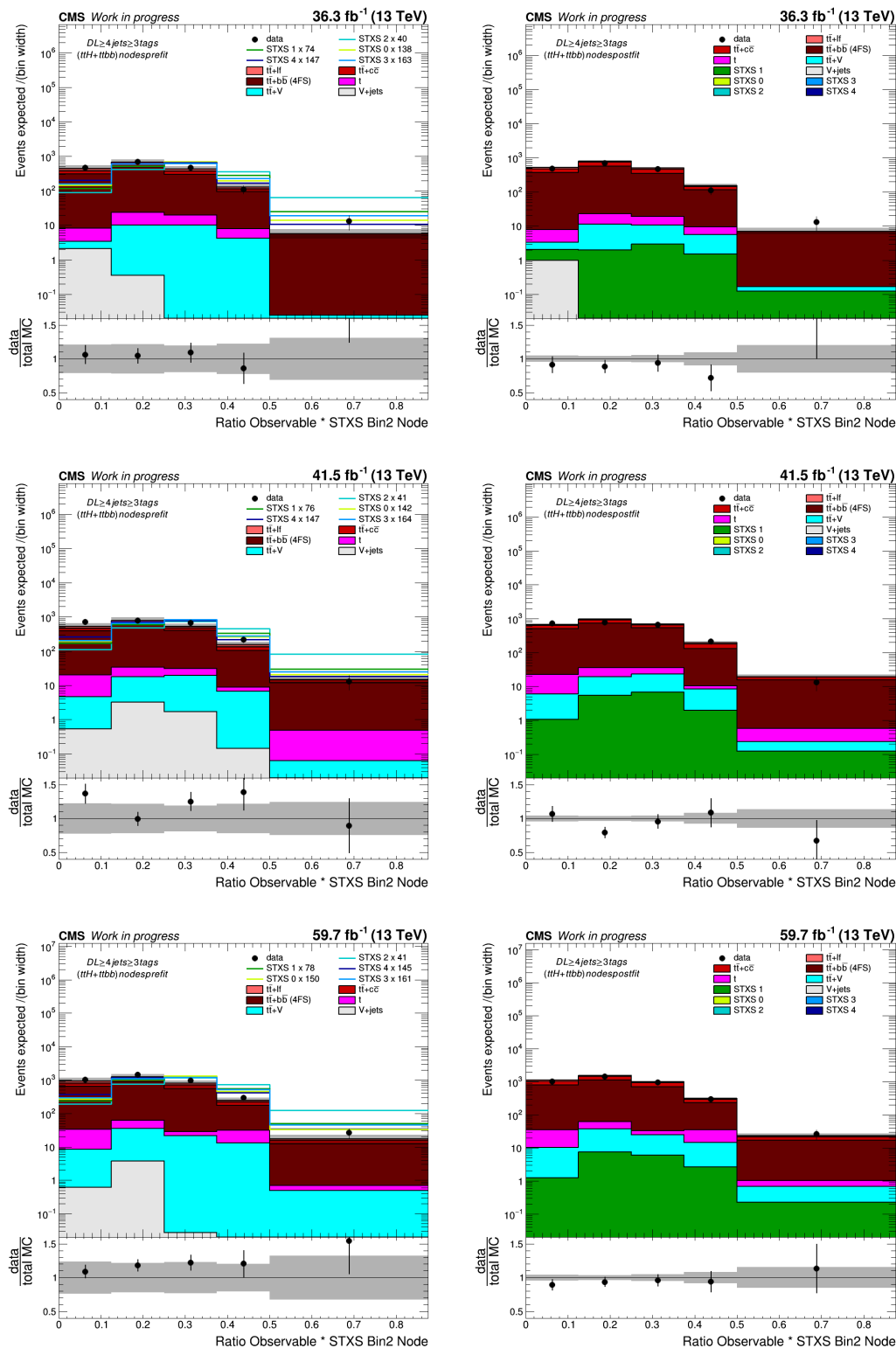


Figure K.14: *Discriminants in the DL channel in the ≥ 4 jets, ≥ 3 b tags STXS bin 2 category. Shown are the pre-fit (left) and post-fit (right) distributions for 2016 (top), 2017 (middle), and 2018 (bottom). The black points represent the observed data. The stacked distributions correspond to the simulations for the respective processes. The uncertainty band corresponds to the uncertainties in the STXS full Run 2 fit to all channels. For better visibility, the signal template is scaled to the total background contribution in the pre-fit figures (cyan line).*

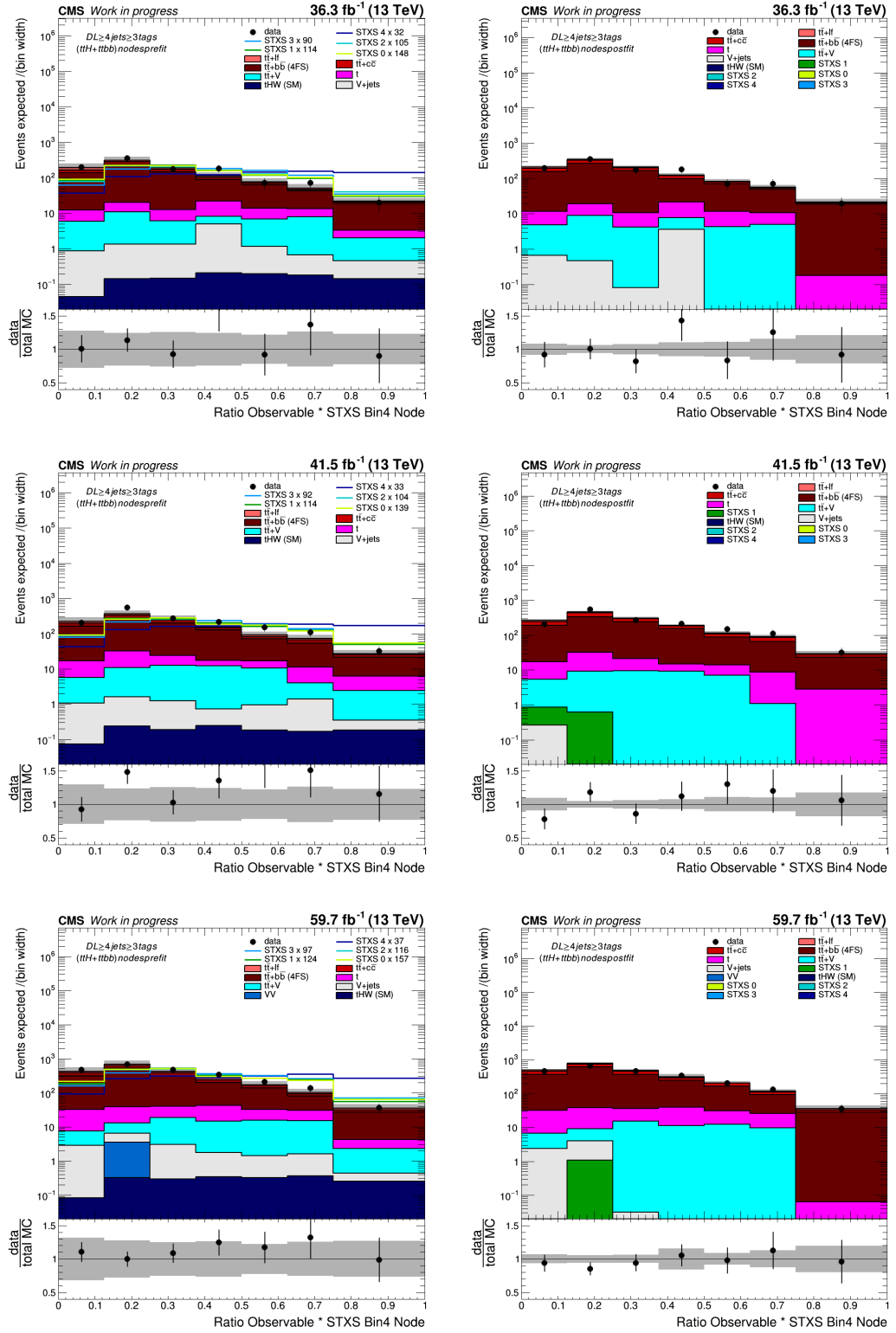


Figure K.15: Discriminants in the DL channel in the ≥ 4 jets, ≥ 3 b tags STXS bin 4 category. Shown are the pre-fit (left) and post-fit (right) distributions for 2016 (top), 2017 (middle), and 2018 (bottom). The black points represent the observed data. The stacked distributions correspond to the simulations for the respective processes. The uncertainty band corresponds to the uncertainties in the STXS full Run 2 fit to all channels. For better visibility, the signal template is scaled to the total background contribution in the pre-fit figures (cyan line).

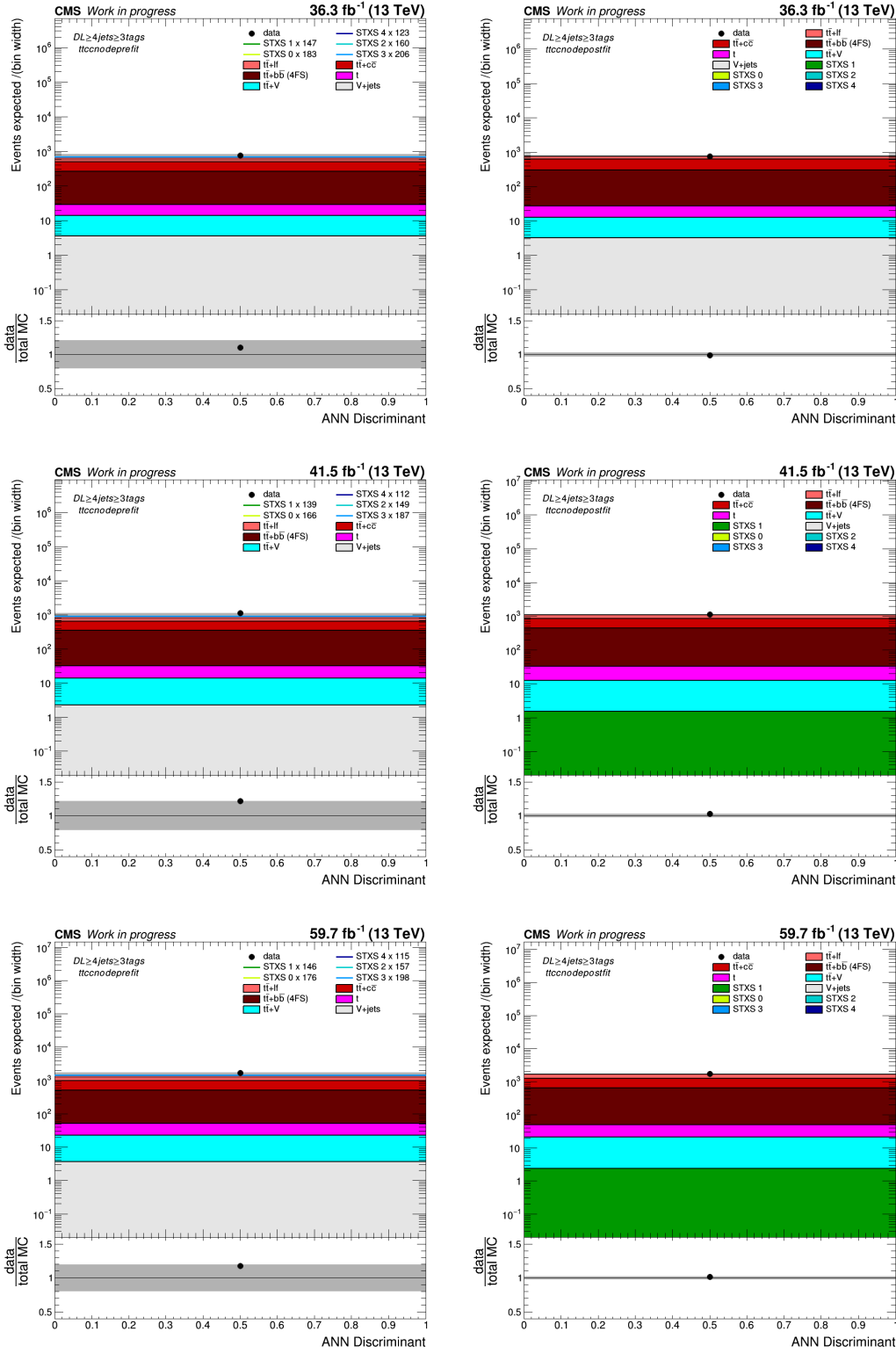


Figure K.16: Discriminants in the DL channel in the ≥ 4 jets, ≥ 3 b tags $t\bar{t} + c\bar{c}$ node category. Shown are the pre-fit (left) and post-fit (right) distributions for 2016 (top), 2017 (middle), and 2018 (bottom). The black points represent the observed data. The stacked distributions correspond to the simulations for the respective processes. The uncertainty band corresponds to the uncertainties in the STXS full Run 2 fit to all channels. For better visibility, the signal template is scaled to the total background contribution in the pre-fit figures (cyan line).

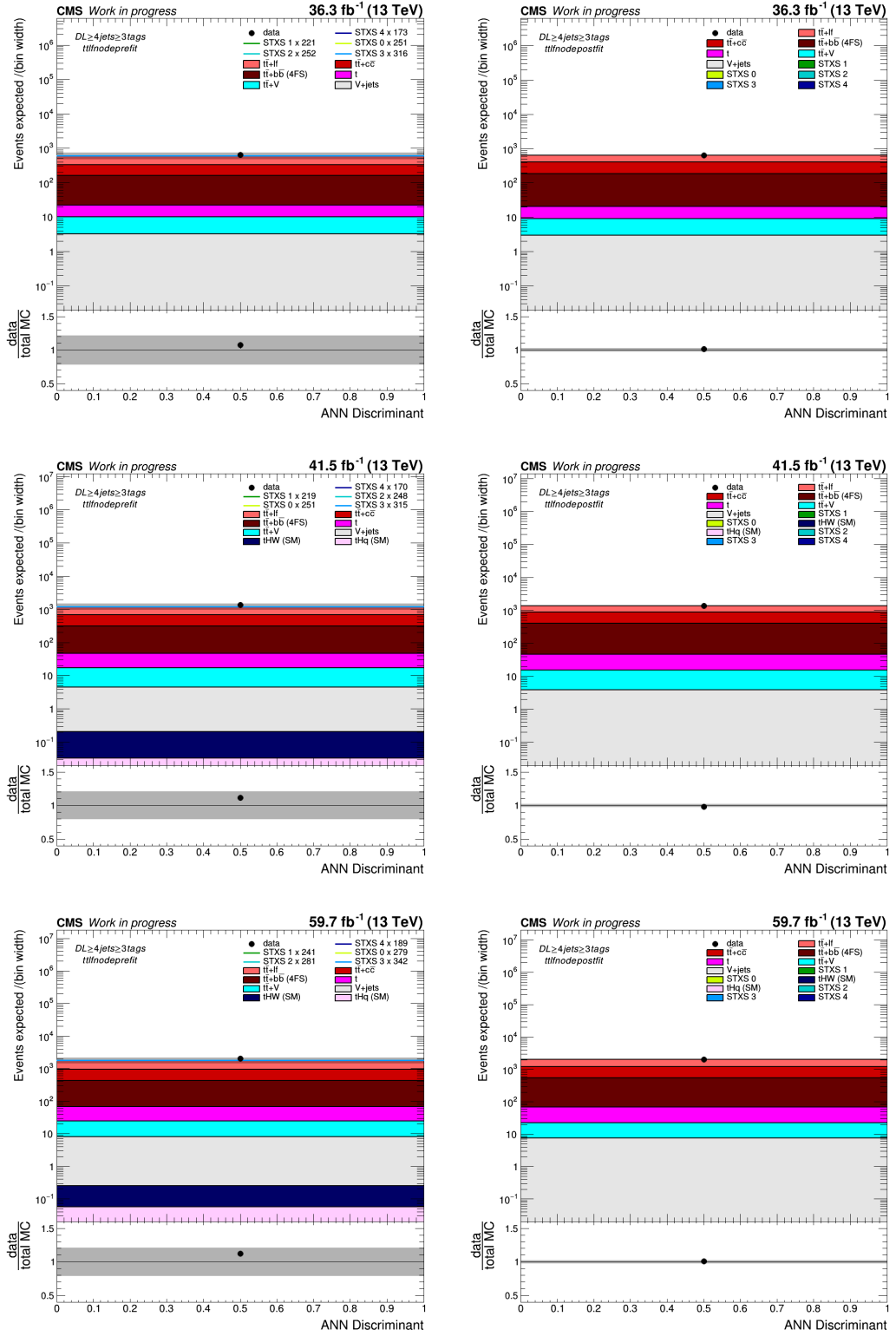


Figure K.17: Discriminants in the DL channel in the ≥ 4 jets, ≥ 3 b tags $t\bar{t}$ + LF node category. Shown are the pre-fit (left) and post-fit (right) distributions for 2016 (top), 2017 (middle), and 2018 (bottom). The black points represent the observed data. The stacked distributions correspond to the simulations for the respective processes. The uncertainty band corresponds to the uncertainties in the STXS full Run 2 fit to all channels. For better visibility, the signal template is scaled to the total background contribution in the pre-fit figures (cyan line).

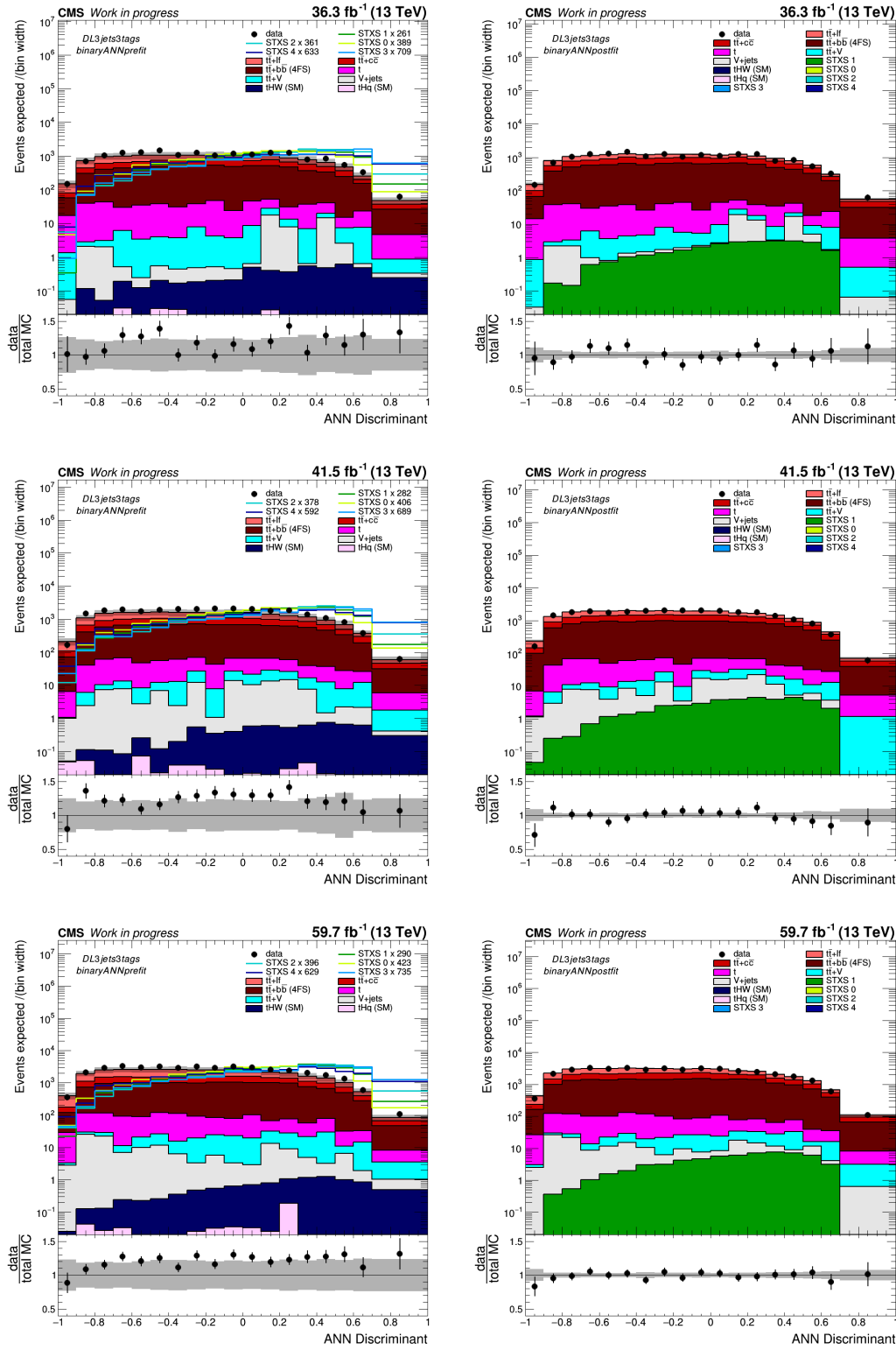


Figure K.18: Discriminants in the DL channel in the 3 jets, $\geq 3b$ tags category. Shown are the pre-fit (left) and post-fit (right) distributions for 2016 (top), 2017 (middle), and 2018 (bottom). The black points represent the observed data. The stacked distributions correspond to the simulations for the respective processes. The uncertainty band corresponds to the uncertainties in the STXS full Run 2 fit to all channels. For better visibility, the signal template is scaled to the total background contribution in the pre-fit figures (cyan line).

L IMPACTS

L.1 Inclusive measurement - full Run 2 dilepton fit

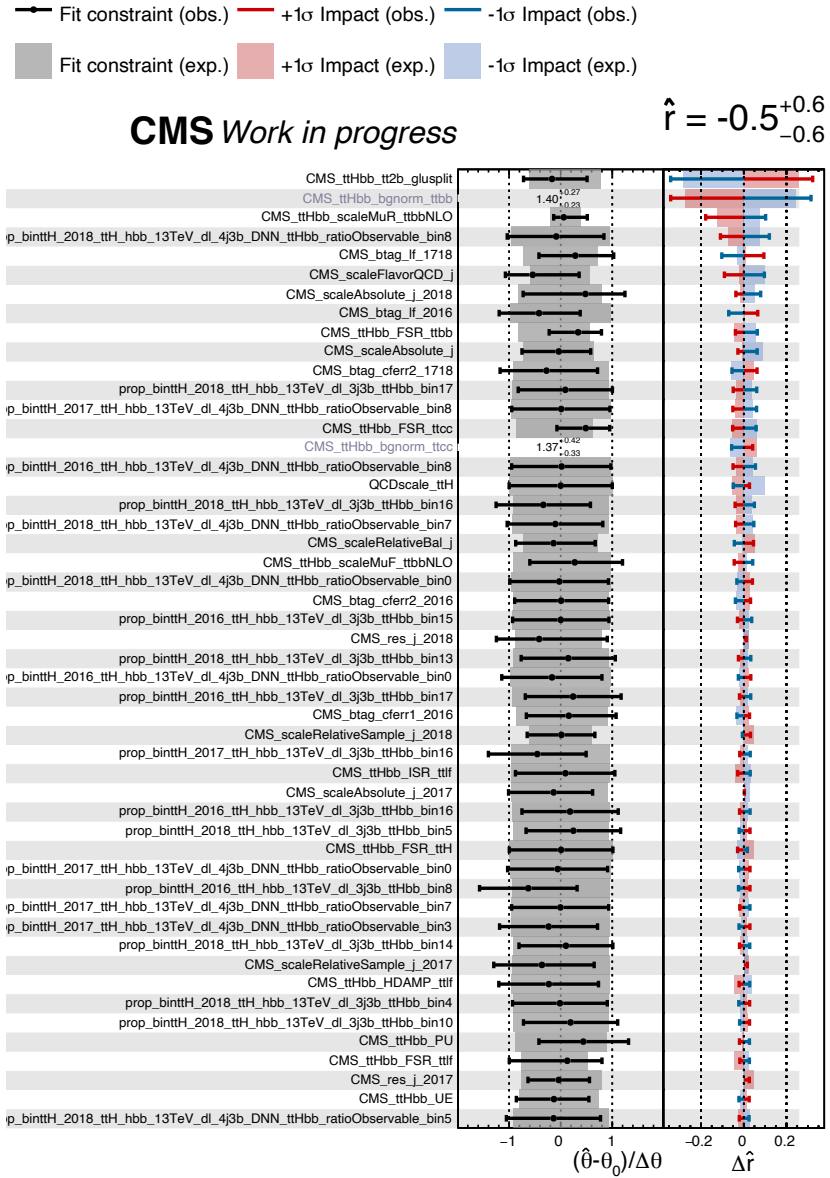


Figure L.1: The impacts and their pulls and constraints on the signal strength obtained in the full Run 2 dilepton fit, both in the fit to pseudo data with the signal process expected from the SM and in the observed data.

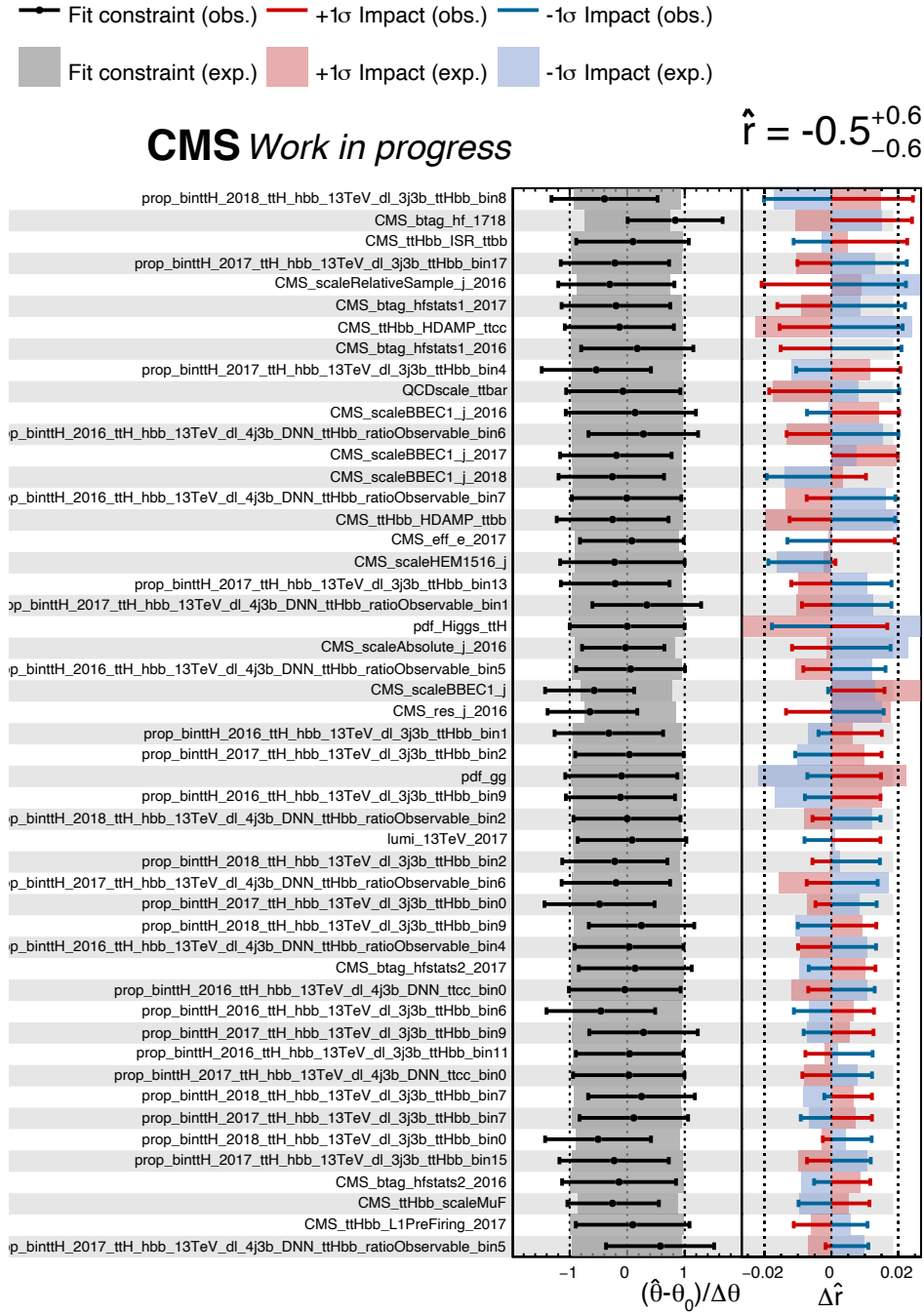


Figure L.2: (Continue) The impacts and their pulls and constraints on the signal strength obtained in the full Run 2 dilepton fit, both in the fit to pseudo data with the signal process expected from the SM and in the observed data. The order is relative to their impact in the fit to the data. The resulting fitted value for the signal strength is also given with its total uncertainty. The black lines represent the nuisance parameter values in comparison to their input values before the fit and relative to their uncertainty. The red and blue lines are the variation in $\Delta\mu$ when the nuisance parameter is varied up and down by its fitted uncertainty, respectively. The corresponding gray, red and blue bands are the expected values from fits to pseudo-data.

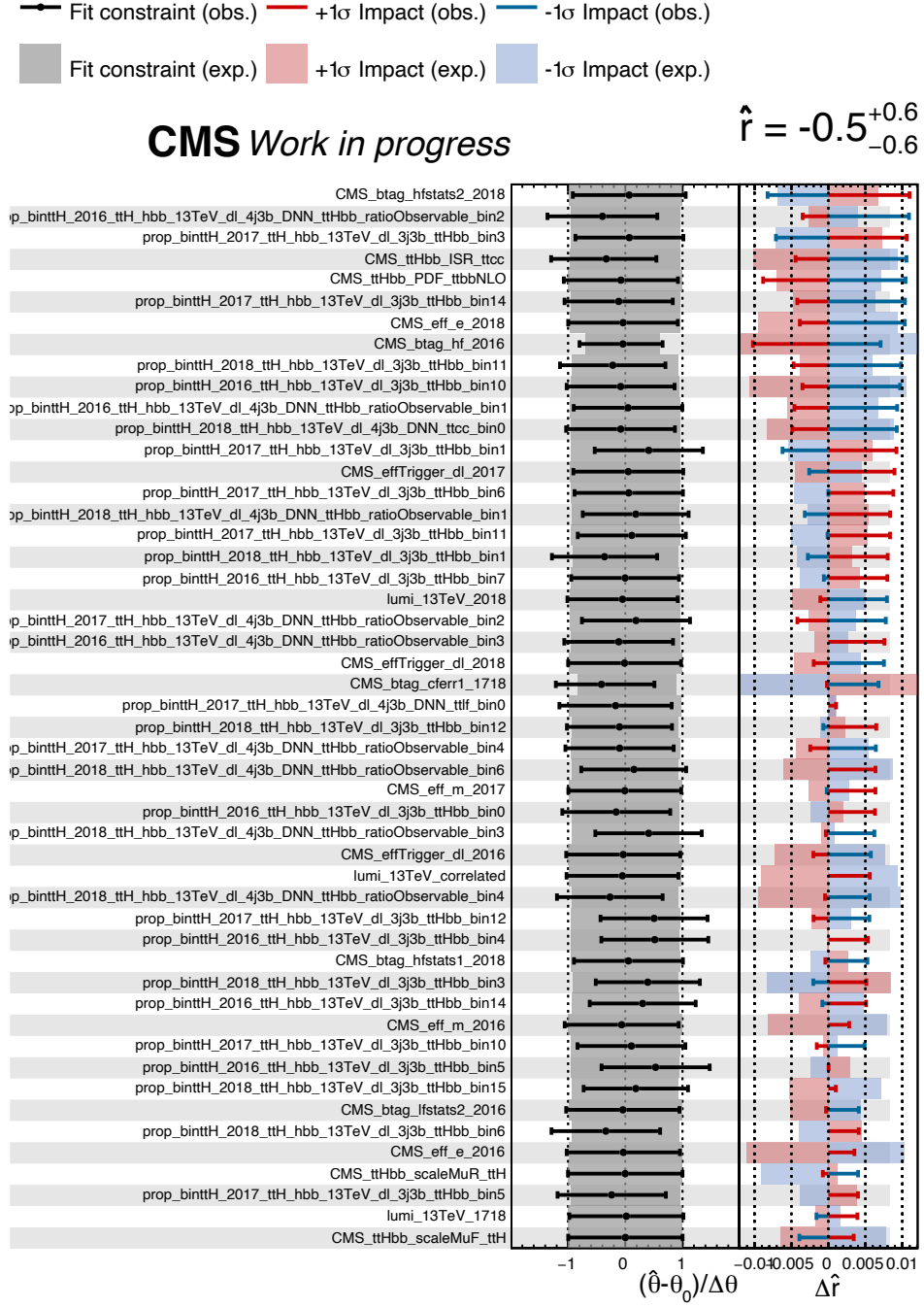


Figure L.3: (Continue) The impacts and their pulls and constraints on the signal strength obtained in the full Run 2 dilepton fit, both in the fit to pseudo data with the signal process expected from the SM and in the observed data. The order is relative to their impact in the fit to the data. The resulting fitted value for the signal strength is also given with its total uncertainty. The black lines represent the nuisance parameter values in comparison to their input values before the fit and relative to their uncertainty. The red and blue lines are the variation in $\Delta\mu$ when the nuisance parameter is varied up and down by its fitted uncertainty, respectively. The corresponding gray, red and blue bands are the expected values from fits to pseudo-data.

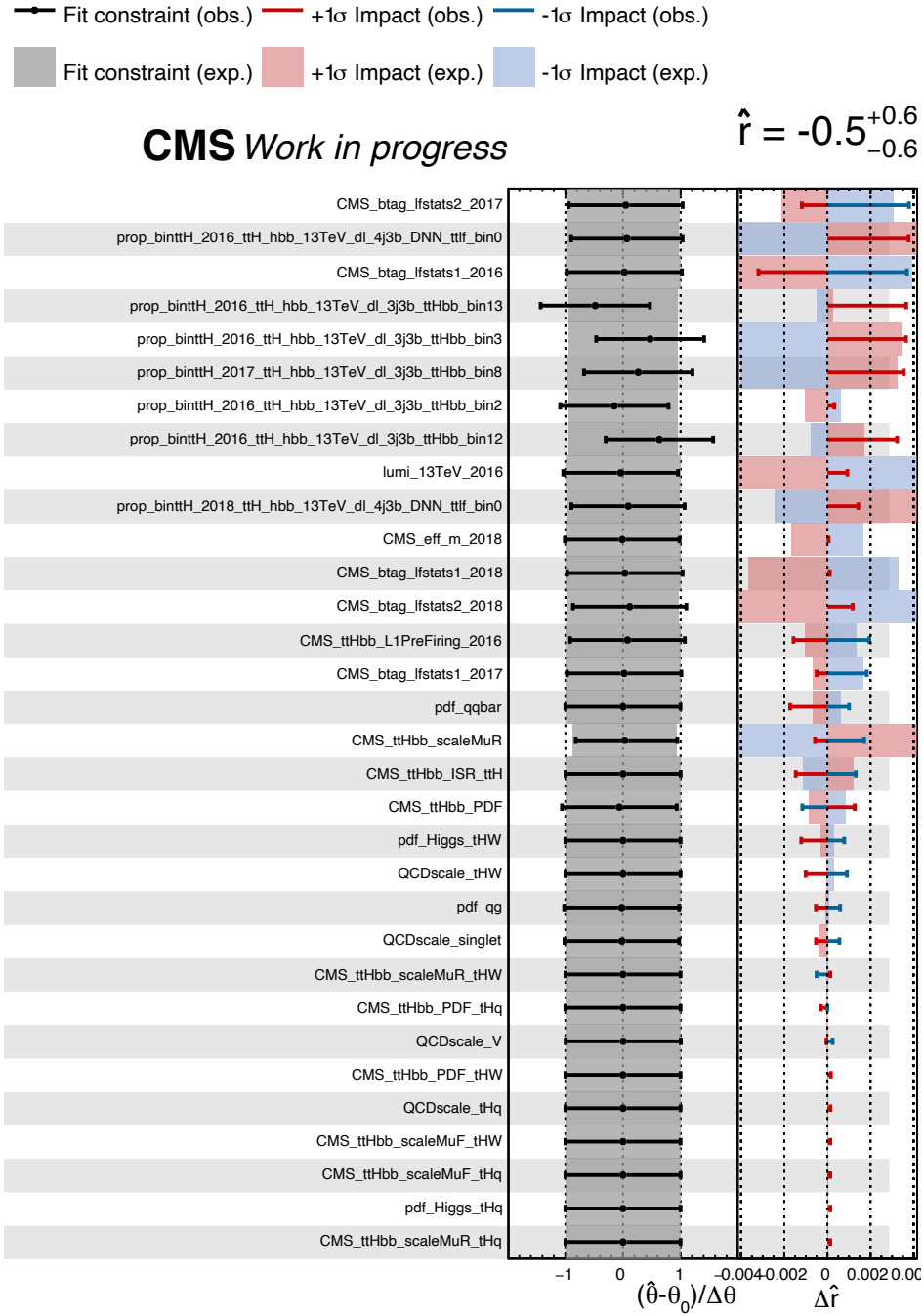


Figure L.4: (Continue) The impacts and their pulls and constraints on the signal strength obtained in the full Run 2 dilepton fit, both in the fit to pseudo data with the signal process expected from the SM and in the observed data. The order is relative to their impact in the fit to the data. The resulting fitted value for the signal strength is also given with its total uncertainty. The black lines represent the nuisance parameter values in comparison to their input values before the fit and relative to their uncertainty. The red and blue lines are the variation in $\Delta\mu$ when the nuisance parameter is varied up and down by its fitted uncertainty, respectively. The corresponding gray, red and blue bands are the expected values from fits to pseudo-data.

L.2 Inclusive measurement - full Run 2 all channels fit

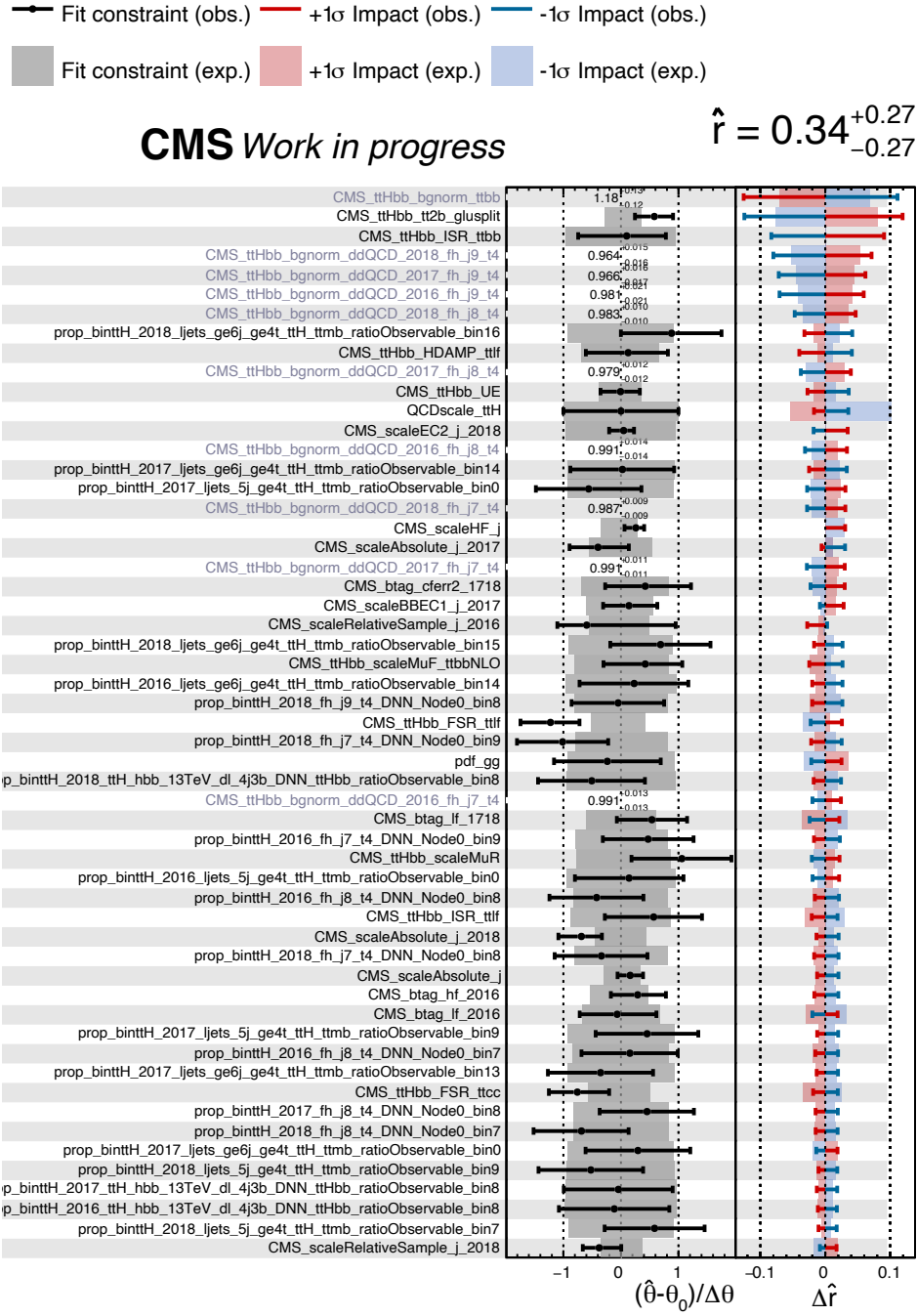


Figure L.5: The impacts and their pulls and constraints on the signal strength obtained in the full Run 2 combination fit to all channels, both in the fit to pseudo data with the signal process expected from the SM and in the observed data. The order is relative to their impact in the fit to the data. The resulting fitted value for the signal strength is also given with its total uncertainty. The black lines represent the nuisance parameter values in comparison to their input values before the fit and relative to their uncertainty. The red and blue lines are the variation in $\Delta\mu$ when the nuisance parameter is varied up and down by its fitted uncertainty, respectively. The corresponding gray, red and blue bands are the expected values from fits to pseudo-data.

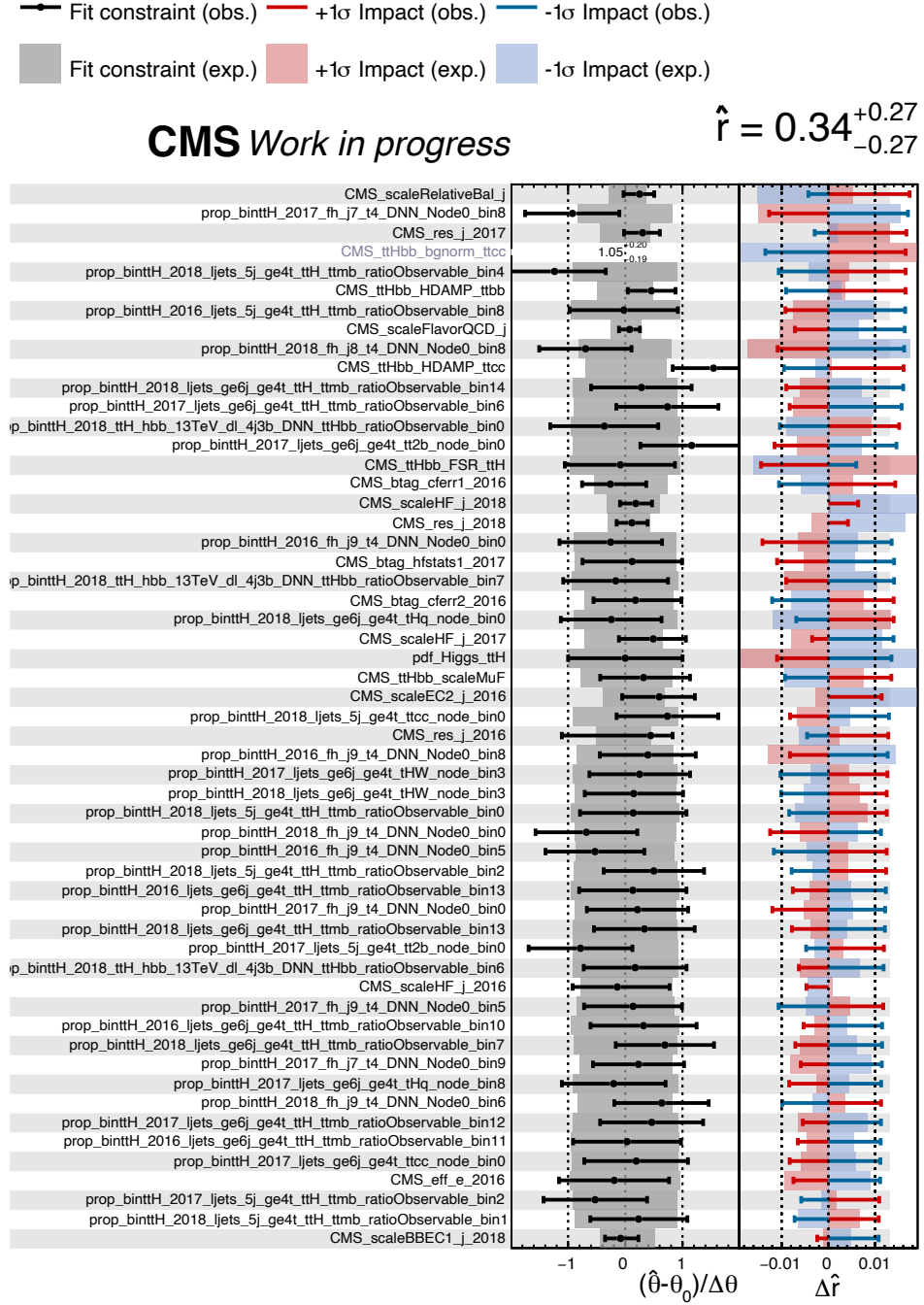


Figure L.6: (Continue) The impacts and their pulls and constraints on the signal strength obtained in the full Run 2 combination fit to all channels, both in the fit to pseudo data with the signal process expected from the SM and in the observed data. The order is relative to their impact in the fit to the data. The resulting fitted value for the signal strength is also given with its total uncertainty. The black lines represent the nuisance parameter values in comparison to their input values before the fit and relative to their uncertainty. The red and blue lines are the variation in $\Delta\mu$ when the nuisance parameter is varied up and down by its fitted uncertainty, respectively. The corresponding gray, red and blue bands are the expected values from fits to pseudo-data.

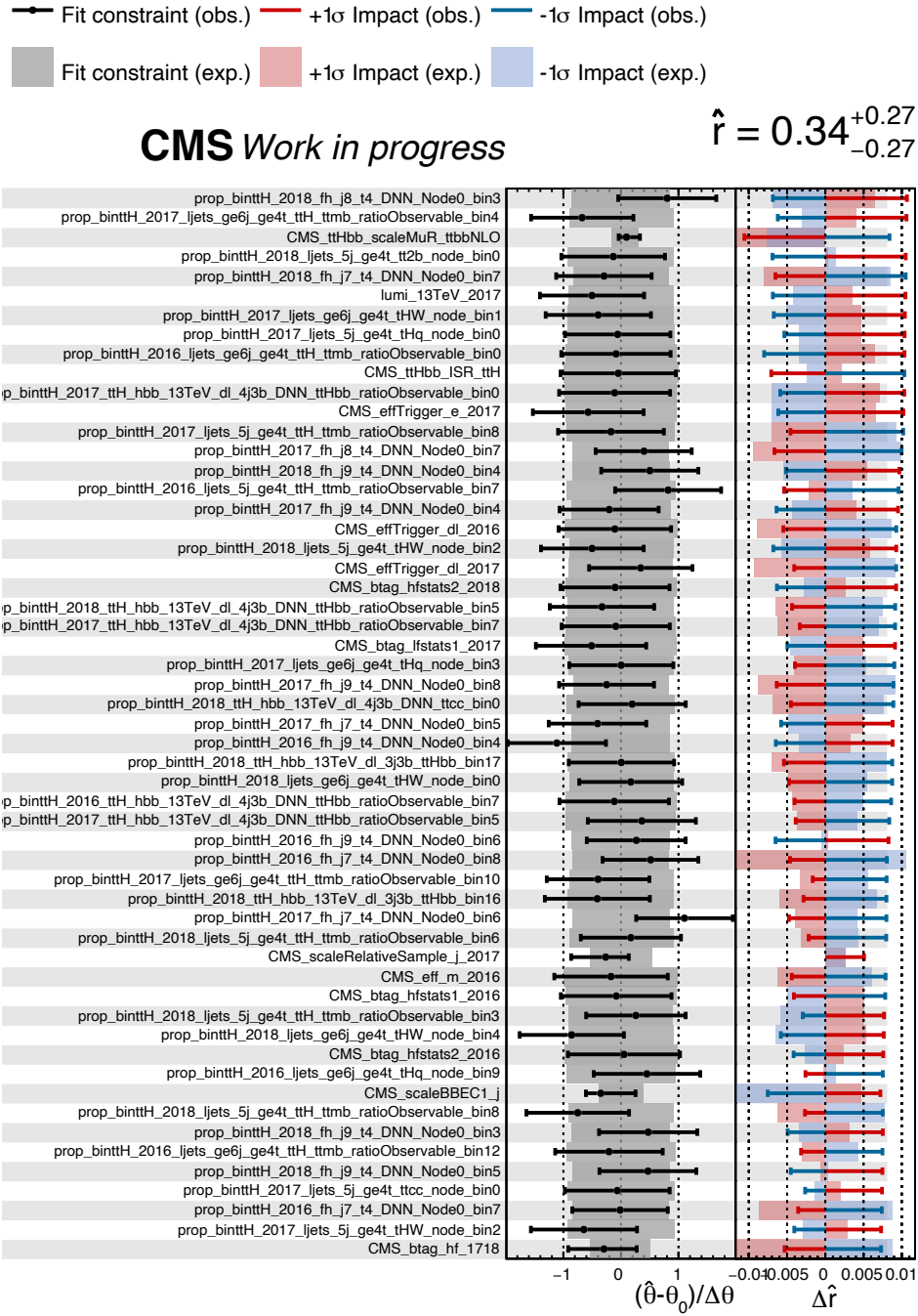


Figure L.7: (Continue) The impacts and their pulls and constraints on the signal strength obtained in the full Run 2 combination fit to all channels, both in the fit to pseudo data with the signal process expected from the SM and in the observed data. The order is relative to their impact in the fit to the data. The resulting fitted value for the signal strength is also given with its total uncertainty. The black lines represent the nuisance parameter values in comparison to their input values before the fit and relative to their uncertainty. The red and blue lines are the variation in $\Delta\mu$ when the nuisance parameter is varied up and down by its fitted uncertainty, respectively. The corresponding gray, red and blue bands are the expected values from fits to pseudo-data.

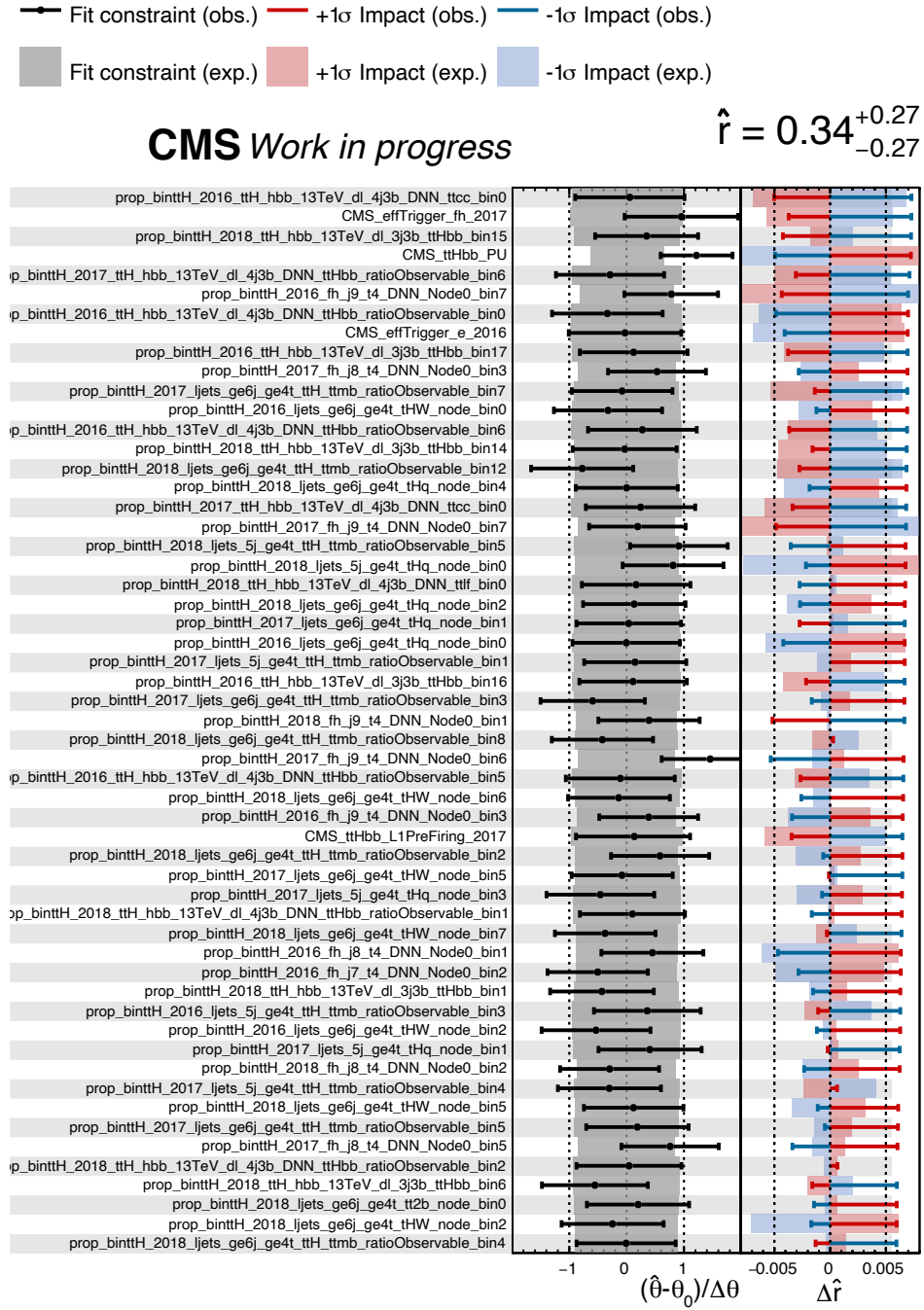


Figure L.8: (Continue) The impacts and their pulls and constraints on the signal strength obtained in the full Run 2 combination fit to all channels, both in the fit to pseudo data with the signal process expected from the SM and in the observed data. The order is relative to their impact in the fit to the data. The resulting fitted value for the signal strength is also given with its total uncertainty. The black lines represent the nuisance parameter values in comparison to their input values before the fit and relative to their uncertainty. The red and blue lines are the variation in $\Delta\mu$ when the nuisance parameter is varied up and down by its fitted uncertainty, respectively. The corresponding gray, red and blue bands are the expected values from fits to pseudo-data.

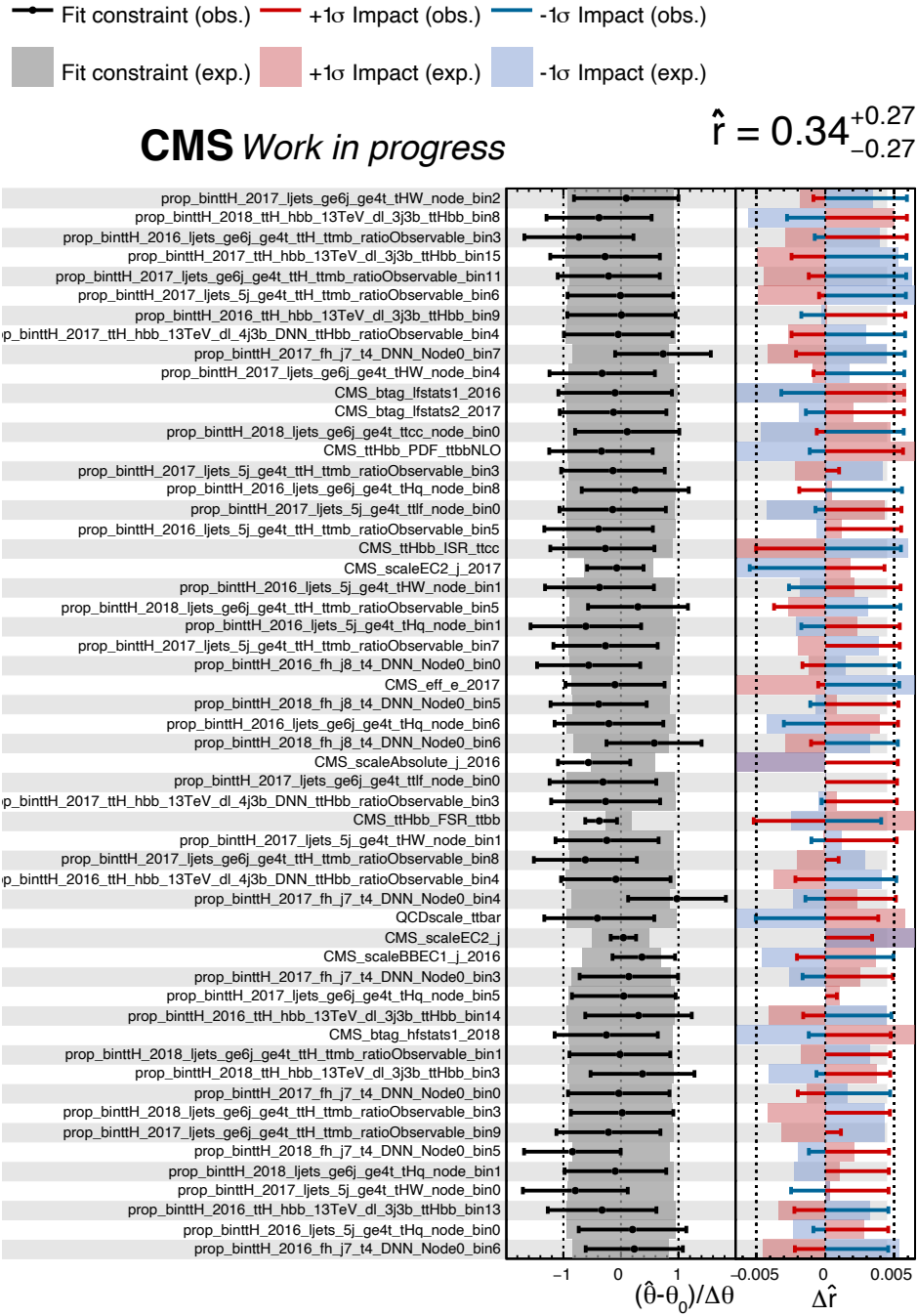


Figure L.9: (Continue) The impacts and their pulls and constraints on the signal strength obtained in the full Run 2 combination fit to all channels, both in the fit to pseudo data with the signal process expected from the SM and in the observed data. The order is relative to their impact in the fit to the data. The resulting fitted value for the signal strength is also given with its total uncertainty. The black lines represent the nuisance parameter values in comparison to their input values before the fit and relative to their uncertainty. The red and blue lines are the variation in $\Delta\mu$ when the nuisance parameter is varied up and down by its fitted uncertainty, respectively. The corresponding gray, red and blue bands are the expected values from fits to pseudo-data.

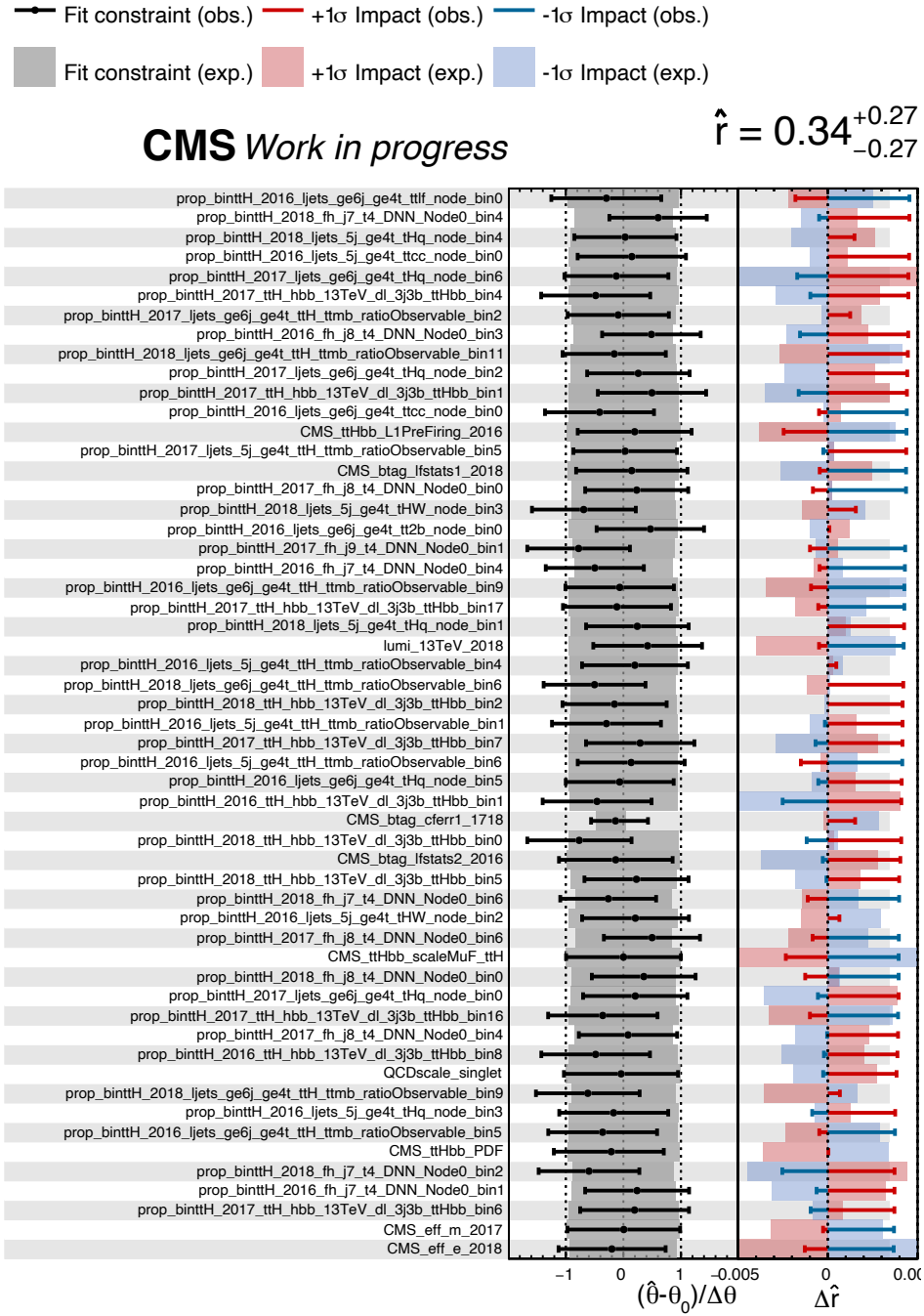


Figure L.10: (Continue) The impacts and their pulls and constraints on the signal strength obtained in the full Run 2 combination fit to all channels, both in the fit to pseudo data with the signal process expected from the SM and in the observed data. The order is relative to their impact in the fit to the data. The resulting fitted value for the signal strength is also given with its total uncertainty. The black lines represent the nuisance parameter values in comparison to their input values before the fit and relative to their uncertainty. The red and blue lines are the variation in $\Delta\mu$ when the nuisance parameter is varied up and down by its fitted uncertainty, respectively. The corresponding gray, red and blue bands are the expected values from fits to pseudo-data.

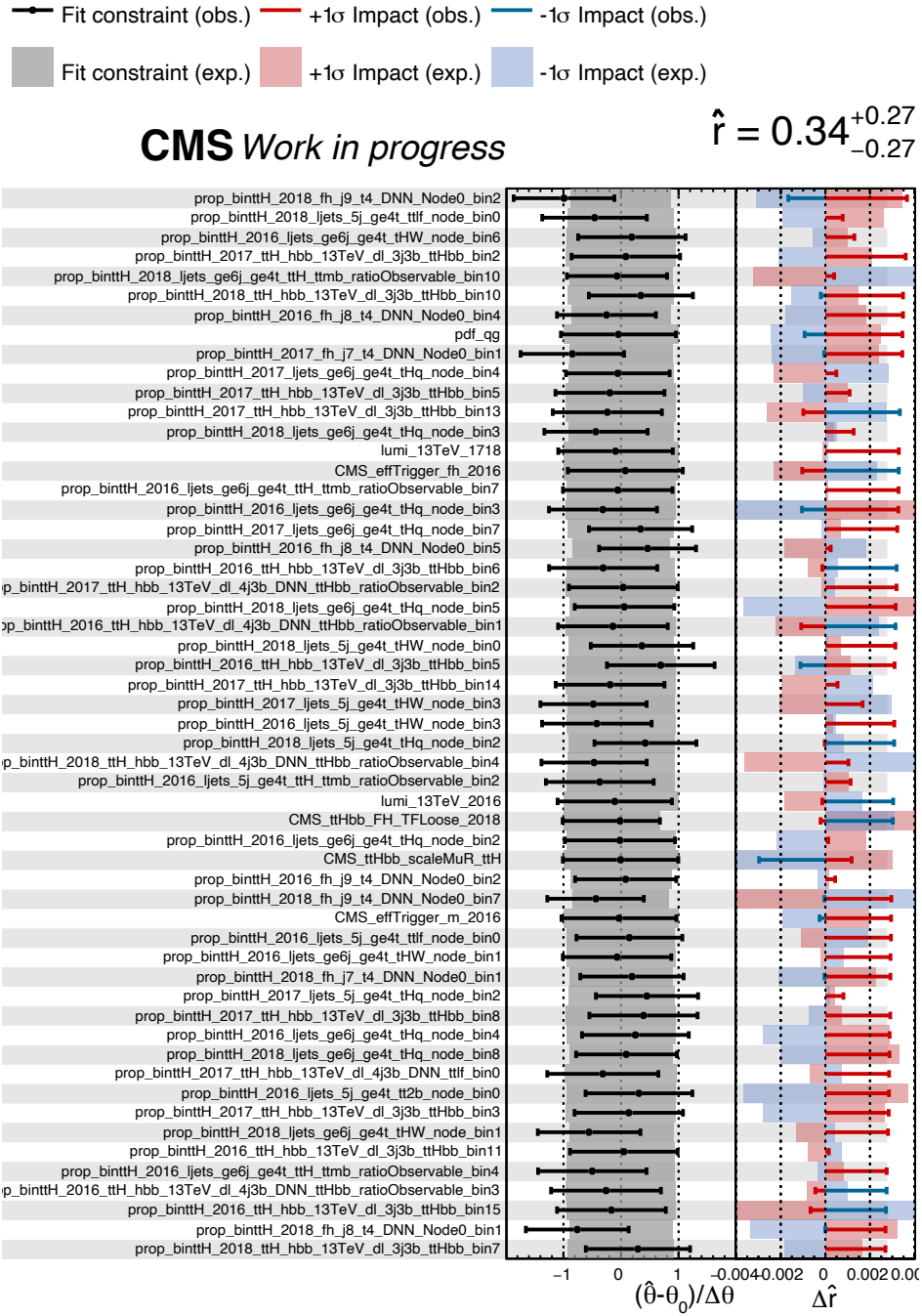


Figure L.11: (Continue) The impacts and their pulls and constraints on the signal strength obtained in the full Run 2 combination fit to all channels, both in the fit to pseudo data with the signal process expected from the SM and in the observed data. The order is relative to their impact in the fit to the data. The resulting fitted value for the signal strength is also given with its total uncertainty. The black lines represent the nuisance parameter values in comparison to their input values before the fit and relative to their uncertainty. The red and blue lines are the variation in $\Delta\mu$ when the nuisance parameter is varied up and down by its fitted uncertainty, respectively. The corresponding gray, red and blue bands are the expected values from fits to pseudo-data.

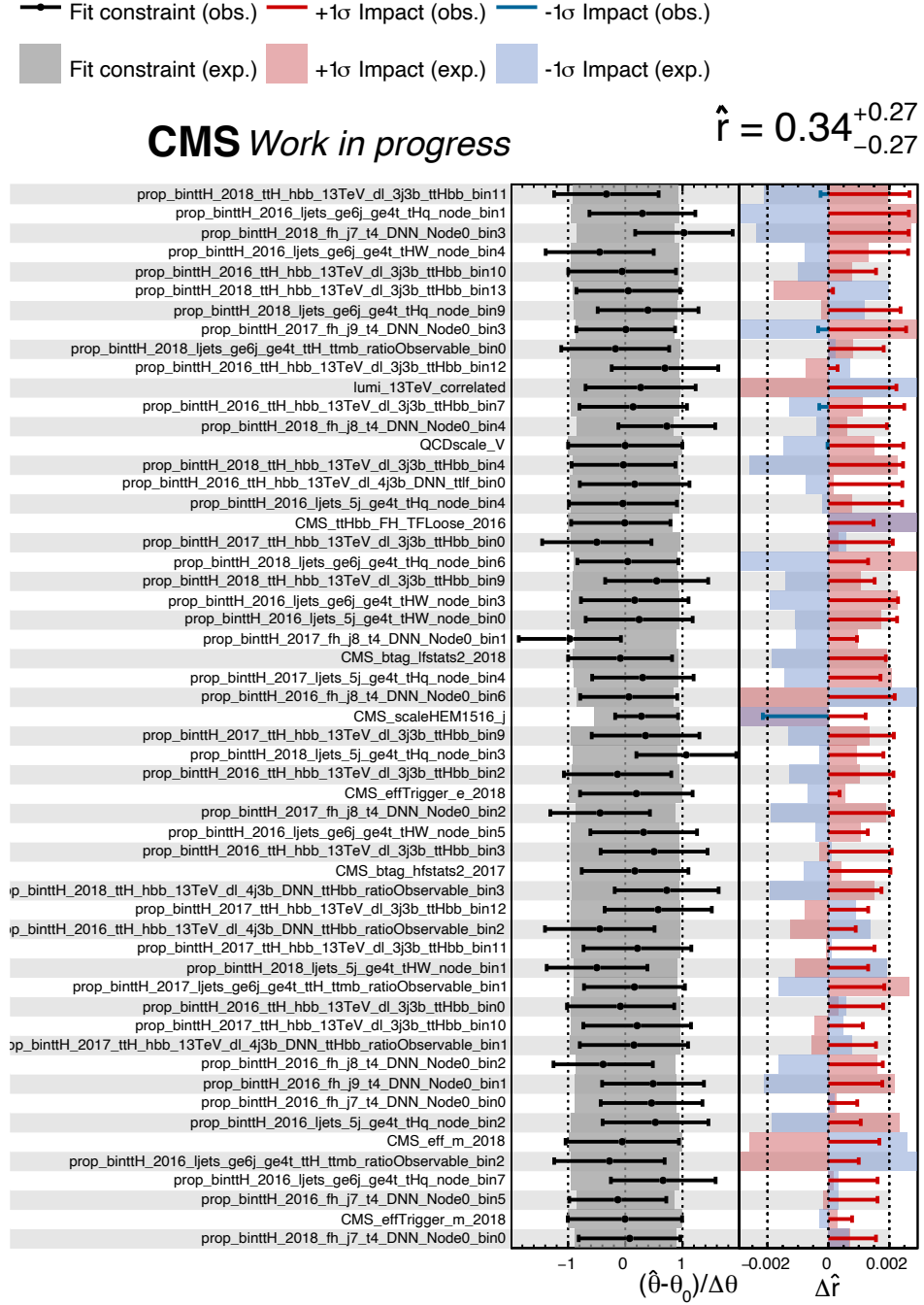


Figure L.12: (Continue) The impacts and their pulls and constraints on the signal strength obtained in the full Run 2 combination fit to all channels, both in the fit to pseudo data with the signal process expected from the SM and in the observed data. The order is relative to their impact in the fit to the data. The resulting fitted value for the signal strength is also given with its total uncertainty. The black lines represent the nuisance parameter values in comparison to their input values before the fit and relative to their uncertainty. The red and blue lines are the variation in $\Delta\mu$ when the nuisance parameter is varied up and down by its fitted uncertainty, respectively. The corresponding gray, red and blue bands are the expected values from fits to pseudo-data.

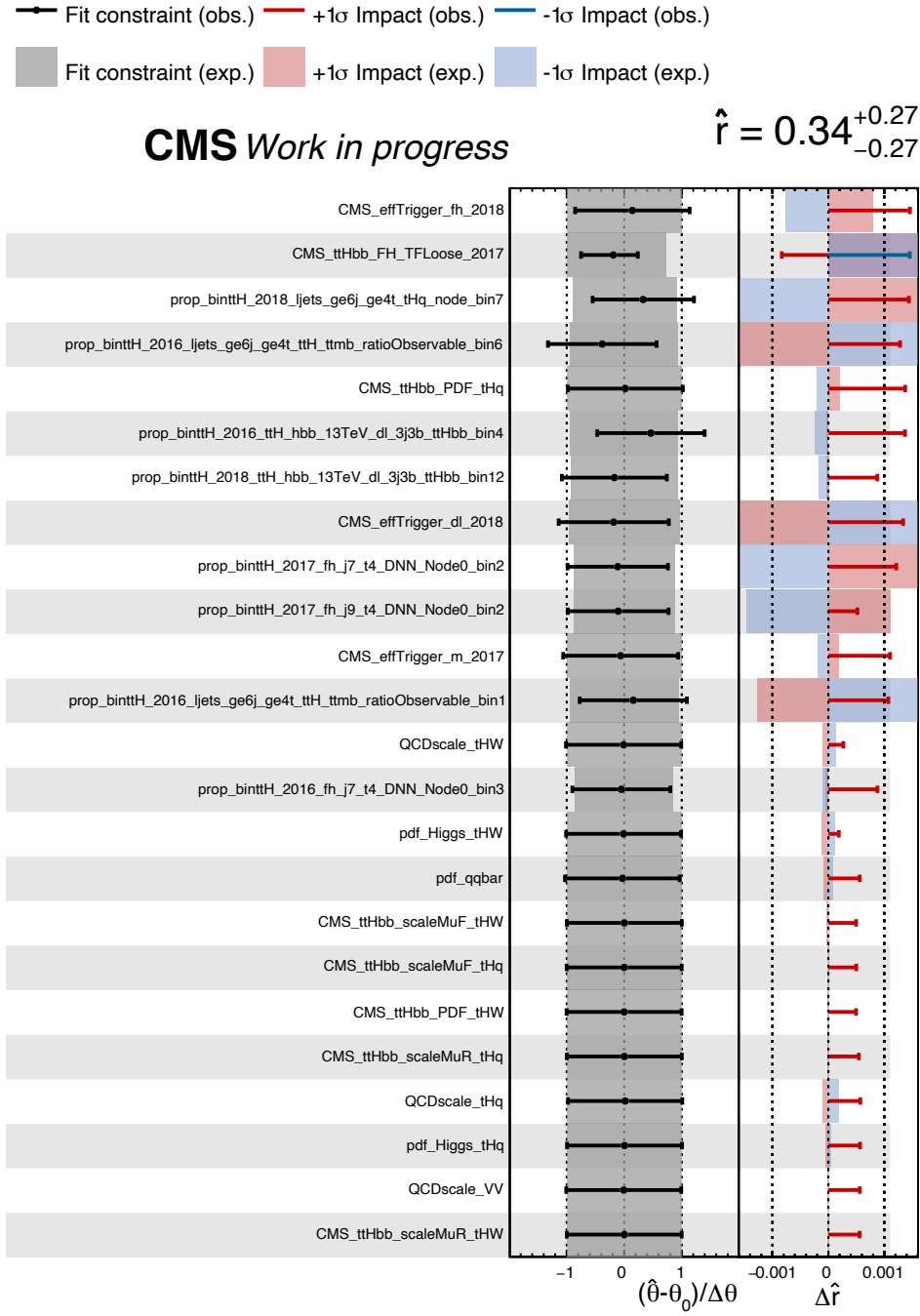
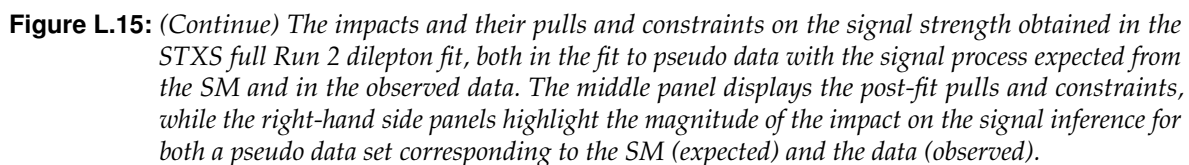


Figure L.13: (Continue) The impacts and their pulls and constraints on the signal strength obtained in the full Run 2 combination fit to all channels, both in the fit to pseudo data with the signal process expected from the SM and in the observed data. The order is relative to their impact in the fit to the data. The resulting fitted value for the signal strength is also given with its total uncertainty. The black lines represent the nuisance parameter values in comparison to their input values before the fit and relative to their uncertainty. The red and blue lines are the variation in $\Delta\mu$ when the nuisance parameter is varied up and down by its fitted uncertainty, respectively. The corresponding gray, red and blue bands are the expected values from fits to pseudo-data.



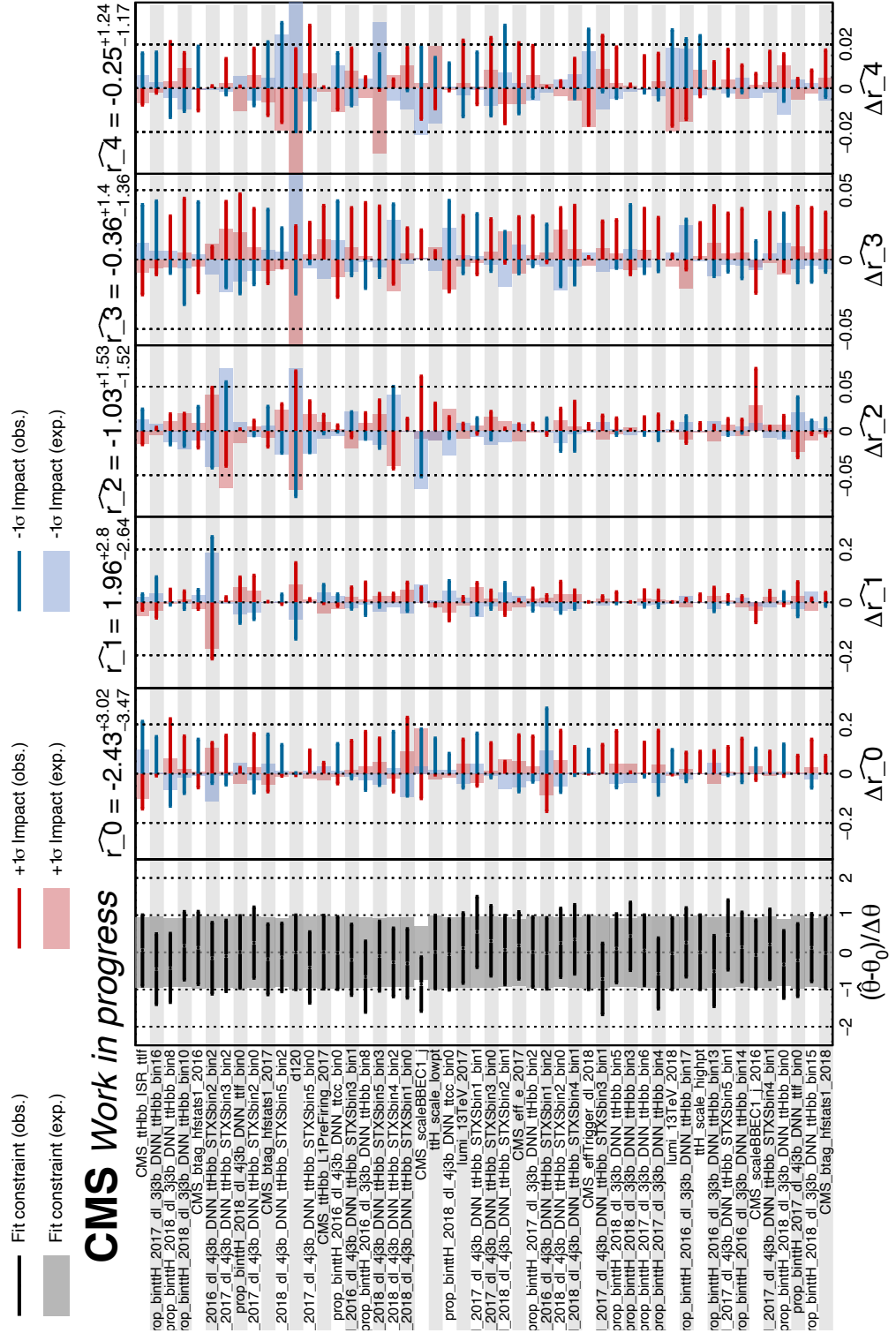


Figure L.16: (Continue) The impacts and their pulls and constraints on the signal strength obtained in the STXS full Run 2 dilepton fit, both in the fit to pseudo data with the signal process expected from the SM and in the observed data. The middle panel displays the post-fit pulls and constraints, while the right-hand side panels highlight the magnitude of the impact on the signal inference for both a pseudo data set corresponding to the SM (expected) and the data (observed).

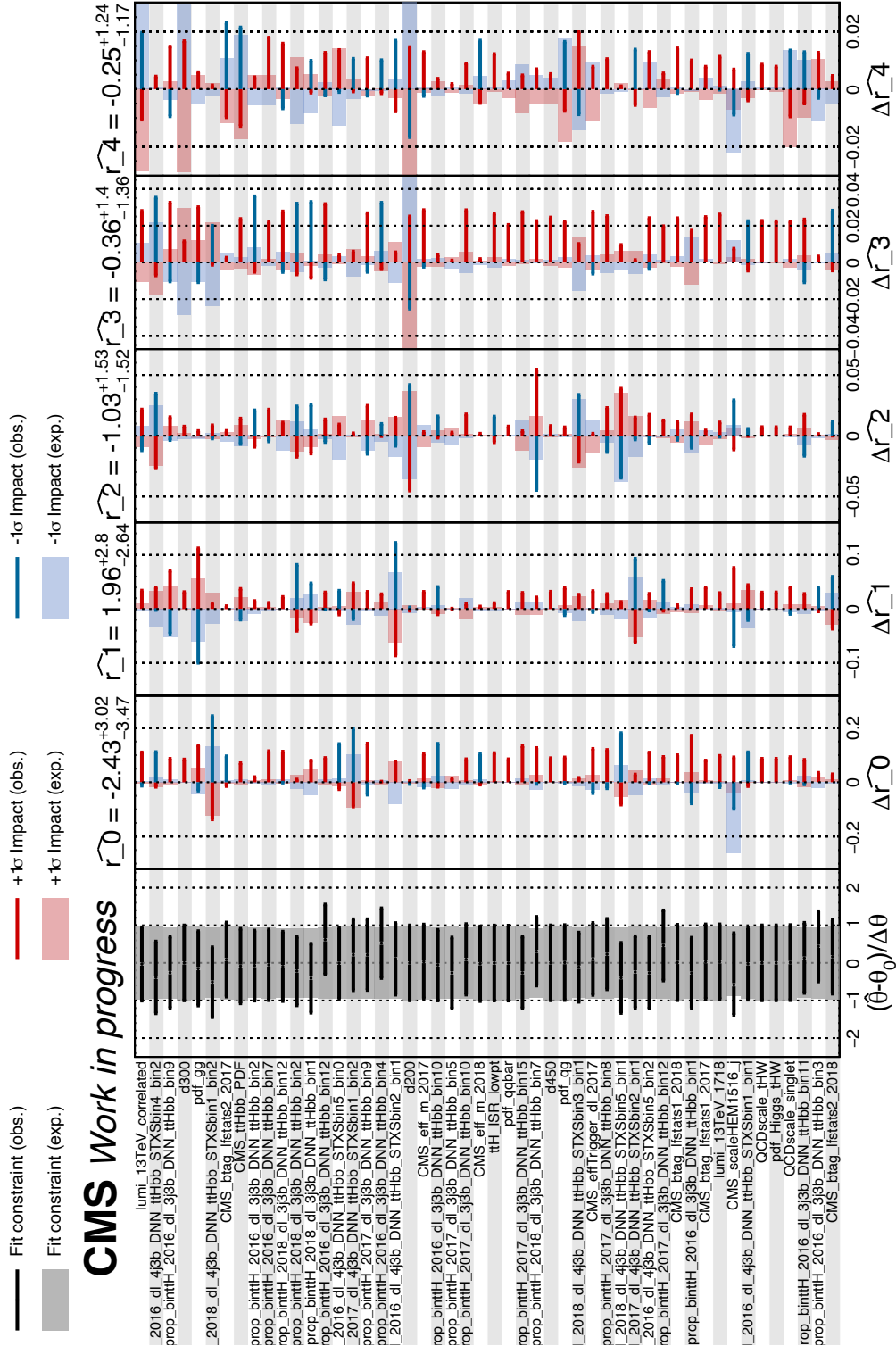
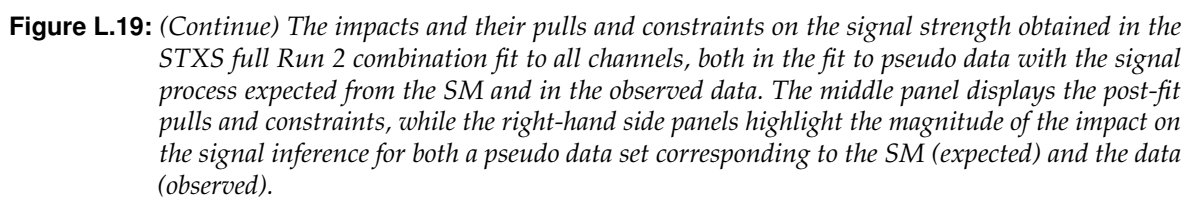


Figure L.17: (Continue) The impacts and their pulls and constraints on the signal strength obtained in the STXS full Run 2 dilepton fit, both in the fit to pseudo data with the signal process expected from the SM and in the observed data. The middle panel displays the post-fit pulls and constraints, while the right-hand side panels highlight the magnitude of the impact on the signal inference for both a pseudo data set corresponding to the SM (expected) and the data (observed).



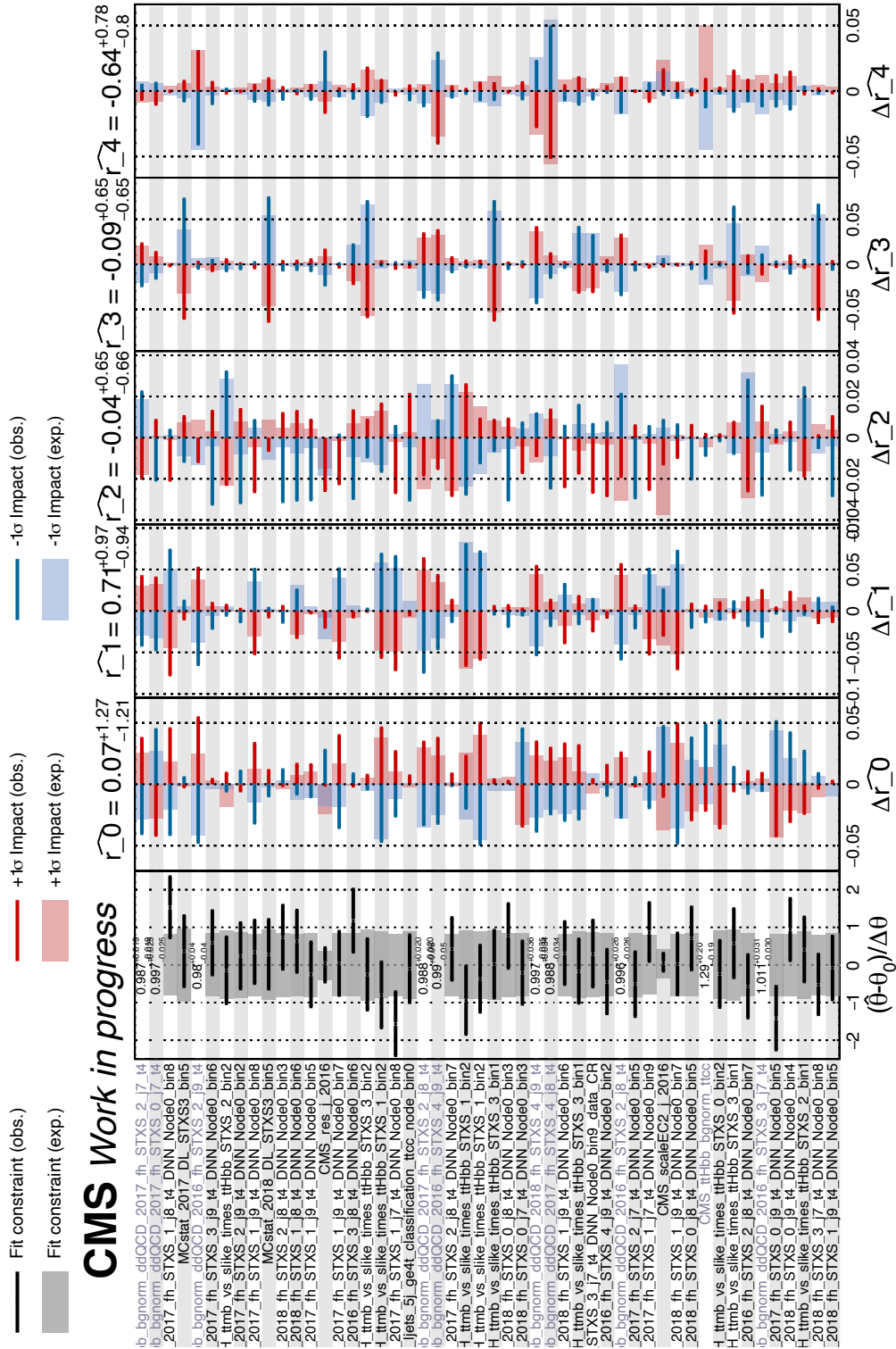


Figure L.21: (Continue) The impacts and their pulls and constraints on the signal strength obtained in the STXS full Run 2 combination fit to all channels, both in the fit to pseudo data with the signal process expected from the SM and in the observed data. The middle panel displays the post-fit pulls and constraints, while the right-hand side panels highlight the magnitude of the impact on the signal inference for both a pseudo data set corresponding to the SM (expected) and the data (observed).

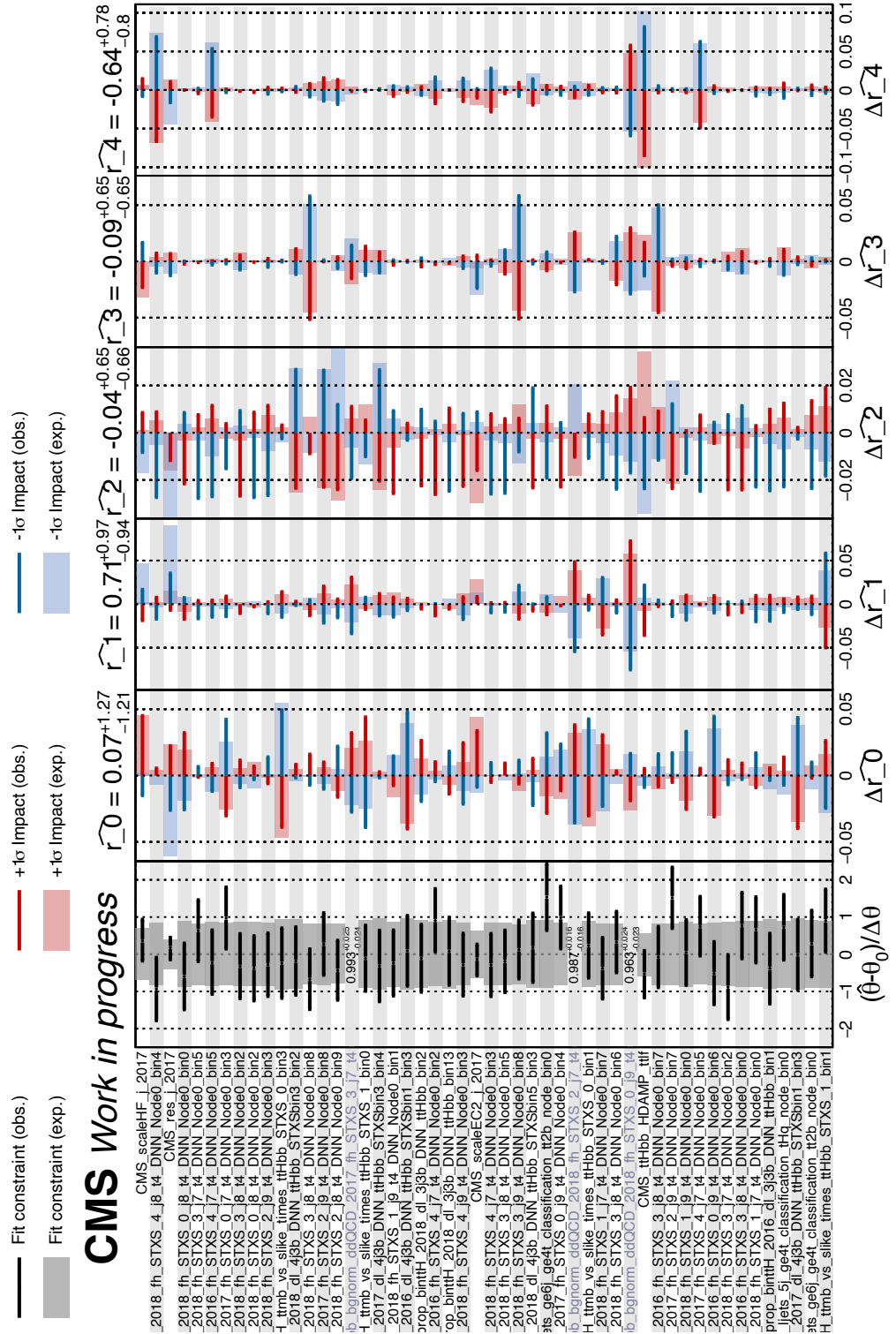


Figure L.22: (Continue) The impacts and their pulls and constraints on the signal strength obtained in the STXS full Run 2 combination fit to all channels, both in the fit to pseudo data with the signal process expected from the SM and in the observed data. The middle panel displays the post-fit pulls and constraints, while the right-hand side panels highlight the magnitude of the impact on the signal inference for both a pseudo data set corresponding to the SM (expected) and the data (observed).

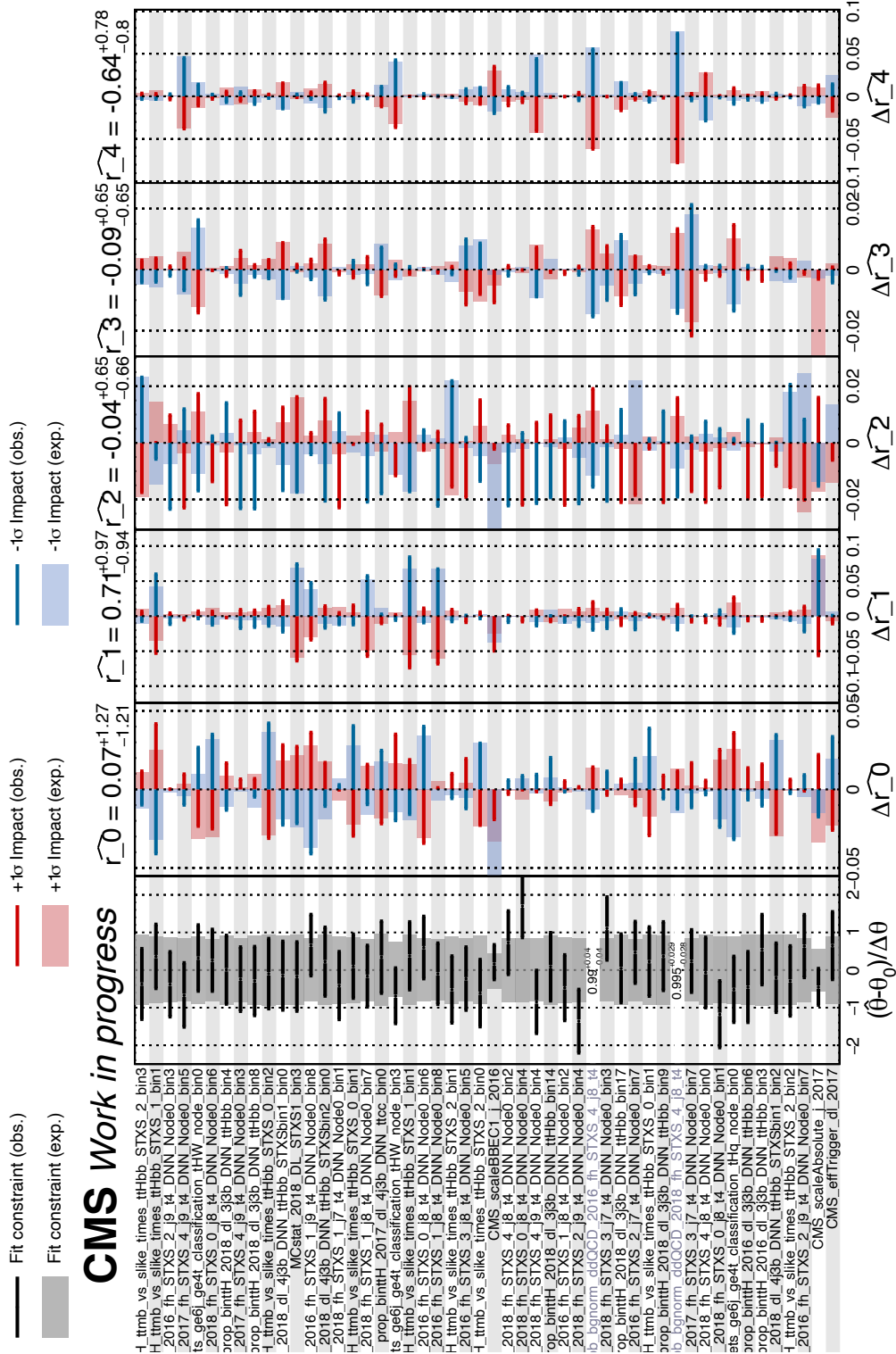


Figure L.23: (Continue) The impacts and their pulls and constraints on the signal strength obtained in the STXS full Run 2 combination fit to all channels, both in the fit to pseudo data with the signal process expected from the SM and in the observed data. The middle panel displays the post-fit pulls and constraints, while the right-hand side panels highlight the magnitude of the impact on the signal inference for both a pseudo data set corresponding to the SM (expected) and the data (observed).

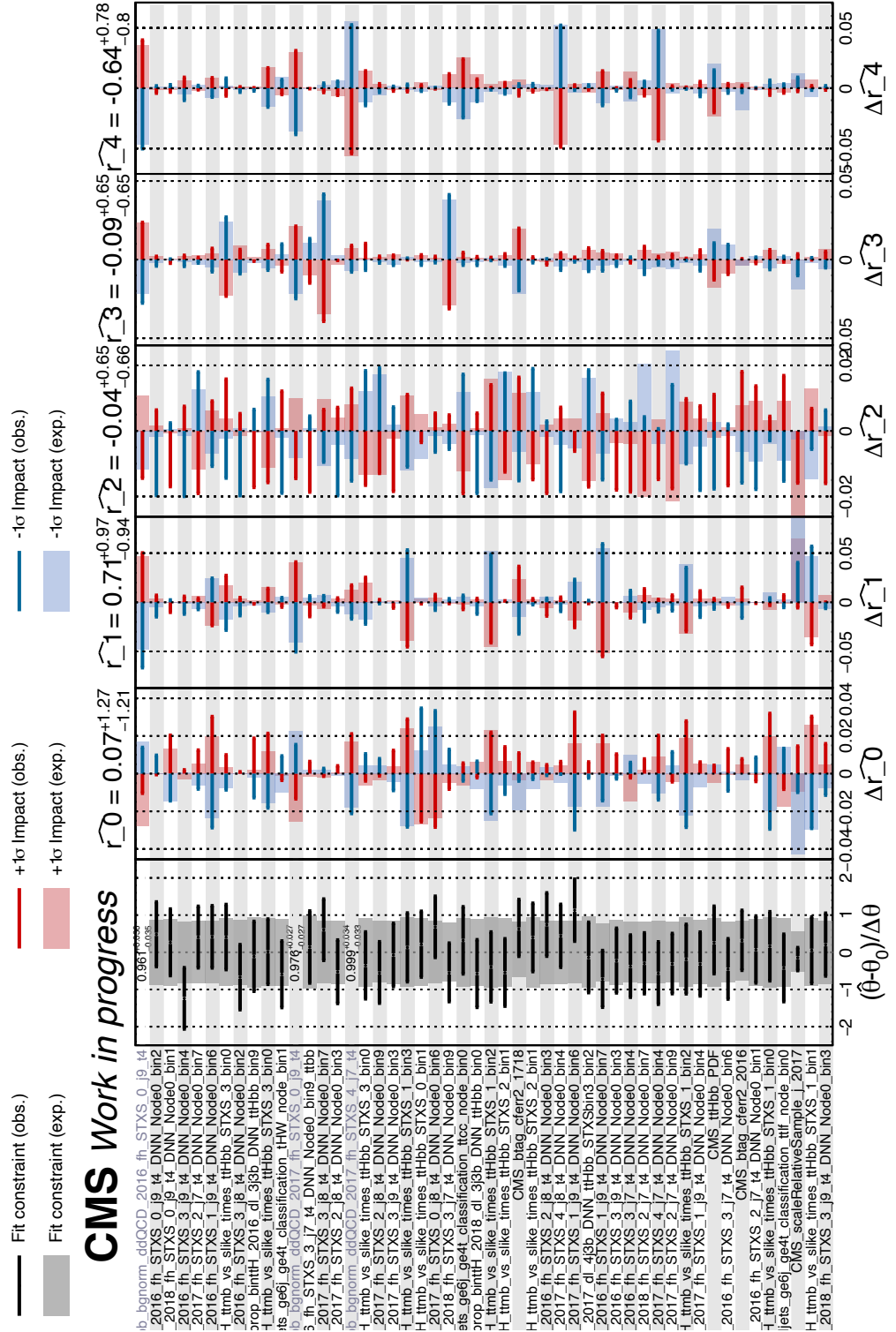


Figure L.24: (Continue) The impacts and their pulls and constraints on the signal strength obtained in the STXS full Run 2 combination fit to all channels, both in the fit to pseudo data with the signal process expected from the SM and in the observed data. The middle panel displays the post-fit pulls and constraints, while the right-hand side panels highlight the magnitude of the impact on the signal inference for both a pseudo data set corresponding to the SM (expected) and the data (observed).

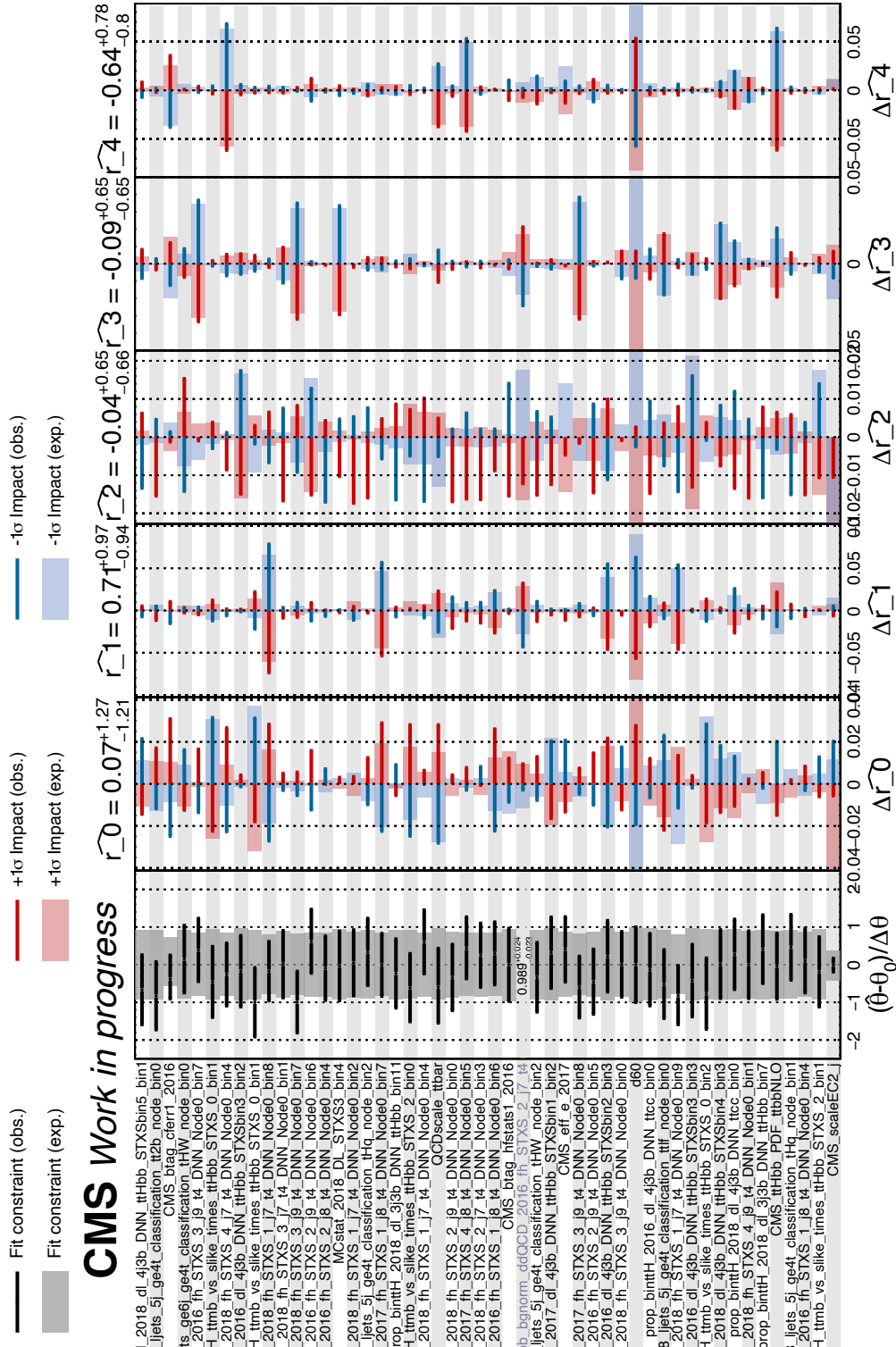


Figure L.25: (Continue) The impacts and their pulls and constraints on the signal strength obtained in the STXS full Run 2 combination fit to all channels, both in the fit to pseudo data with the signal process expected from the SM and in the observed data. The middle panel displays the post-fit pulls and constraints, while the right-hand side panels highlight the magnitude of the impact on the signal inference for both a pseudo data set corresponding to the SM (expected) and the data (observed).

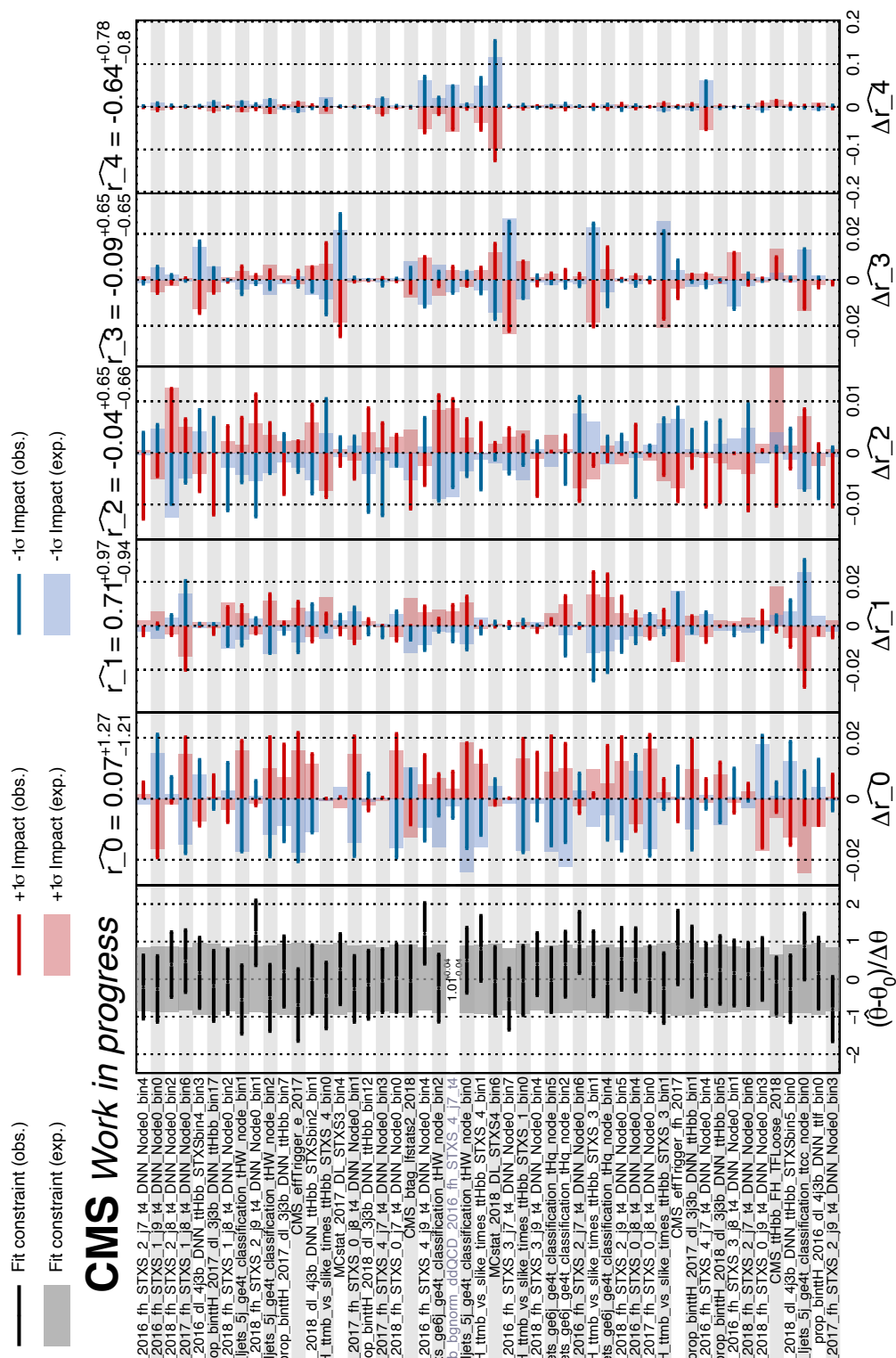


Figure L.27: (Continue) The impacts and their pulls and constraints on the signal strength obtained in the STXS full Run 2 combination fit to all channels, both in the fit to pseudo data with the signal process expected from the SM and in the observed data. The middle panel displays the post-fit pulls and constraints, while the right-hand side panels highlight the magnitude of the impact on the signal inference for both a pseudo data set corresponding to the SM (expected) and the data (observed).

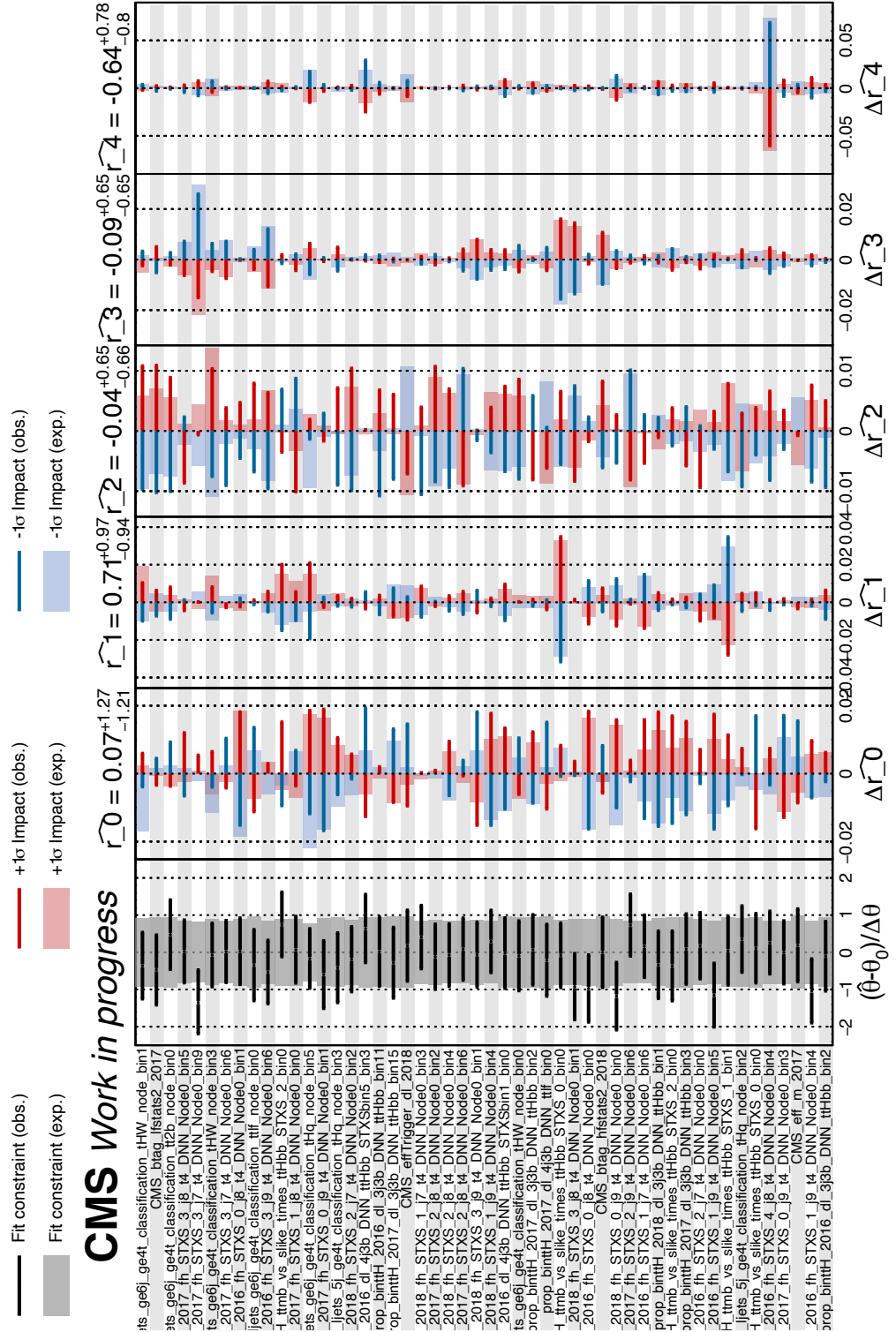


Figure L.28: (Continue) The impacts and their pulls and constraints on the signal strength obtained in the STXS full Run 2 combination fit to all channels, both in the fit to pseudo data with the signal process expected from the SM and in the observed data. The middle panel displays the post-fit pulls and constraints, while the right-hand side panels highlight the magnitude of the impact on the signal inference for both a pseudo data set corresponding to the SM (expected) and the data (observed).

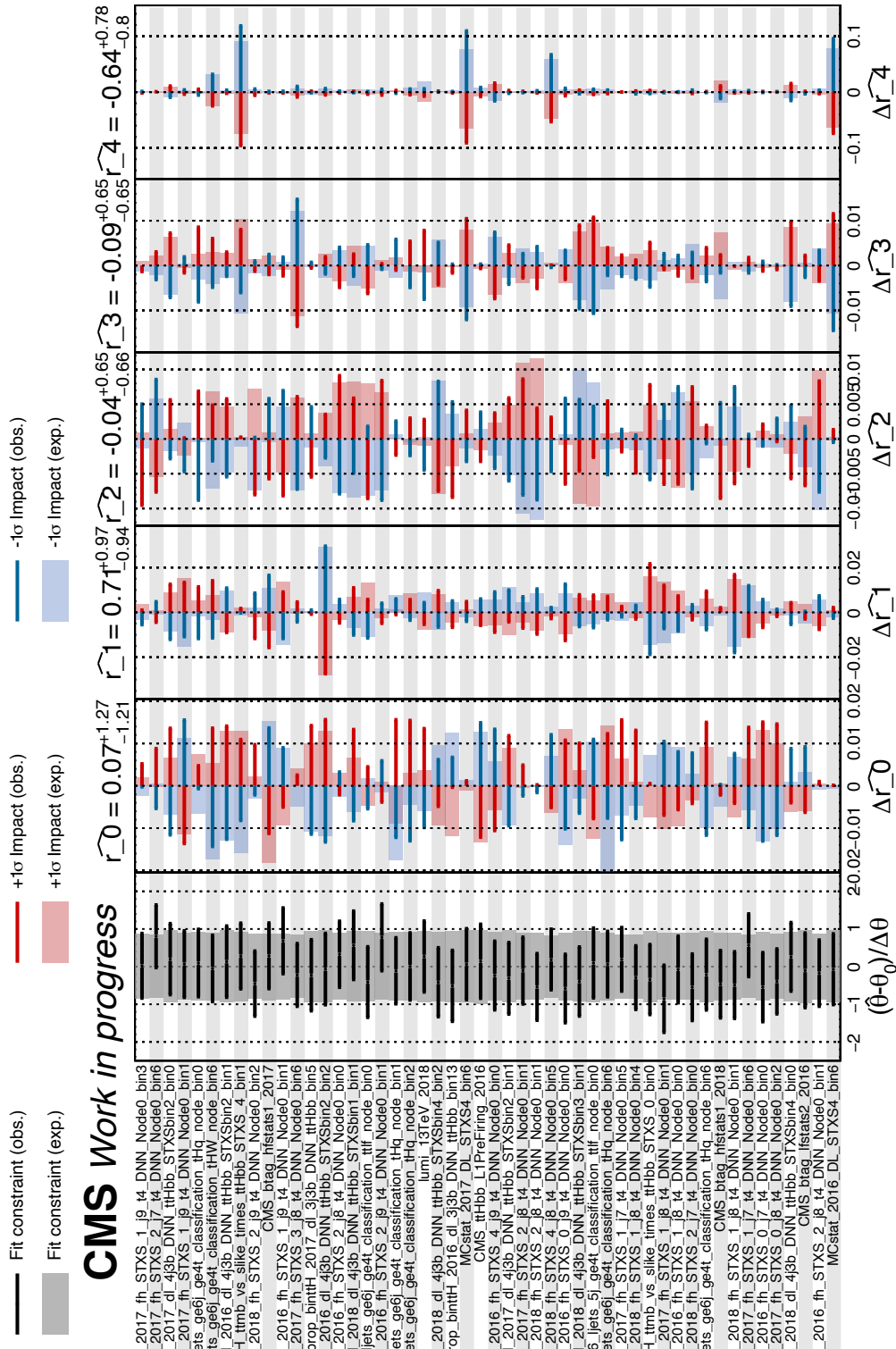


Figure L.29: (Continue) The impacts and their pulls and constraints on the signal strength obtained in the STXS full Run 2 combination fit to all channels, both in the fit to pseudo data with the signal process expected from the SM and in the observed data. The middle panel displays the post-fit pulls and constraints, while the right-hand side panels highlight the magnitude of the impact on the signal inference for both a pseudo data set corresponding to the SM (expected) and the data (observed).

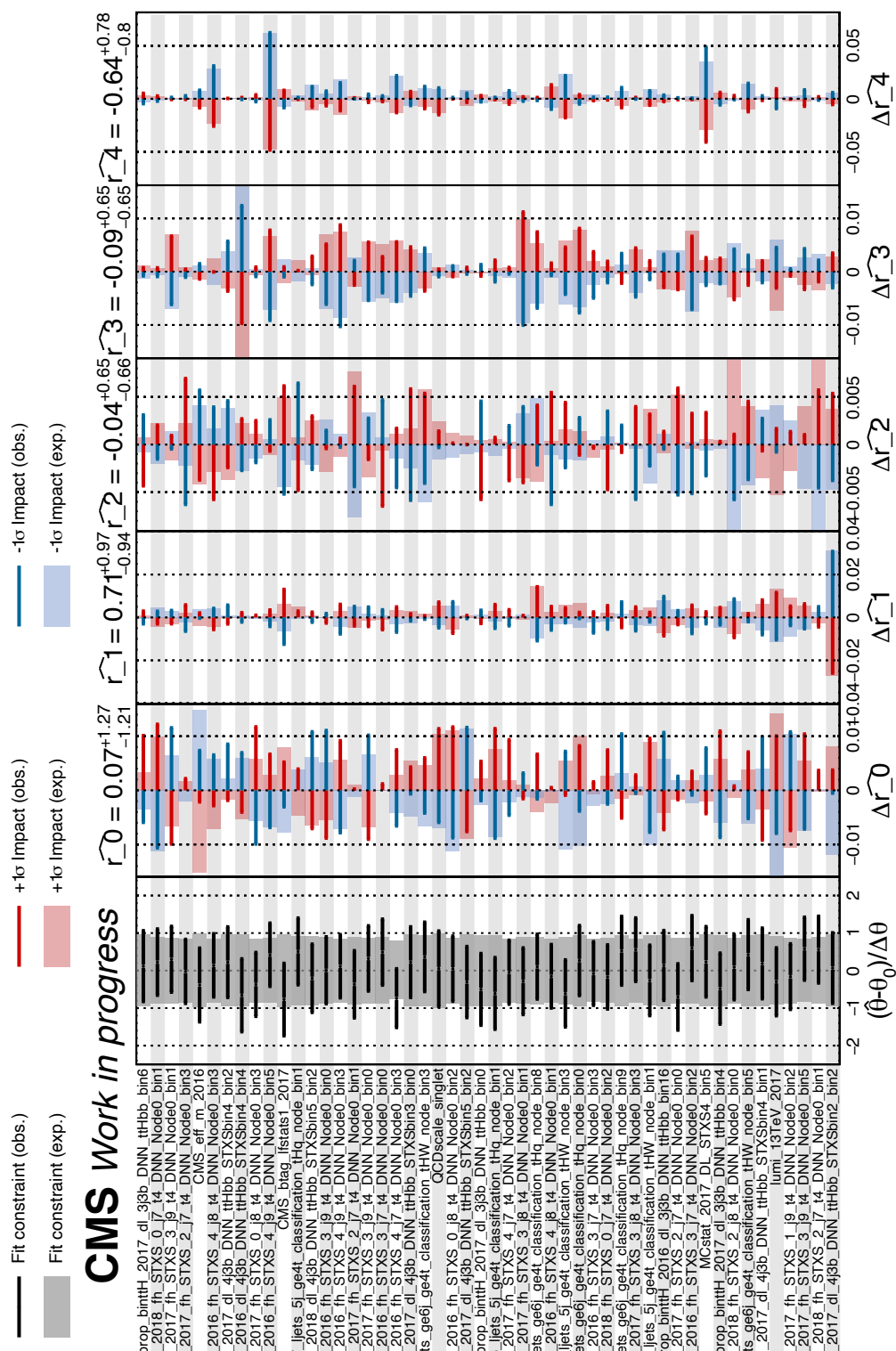


Figure L.31: (Continue) The impacts and their pulls and constraints on the signal strength obtained in the STXS full Run 2 combination fit to all channels, both in the fit to pseudo data with the signal process expected from the SM and in the observed data. The middle panel displays the post-fit pulls and constraints, while the right-hand side panels highlight the magnitude of the impact on the signal inference for both a pseudo data set corresponding to the SM (expected) and the data (observed).

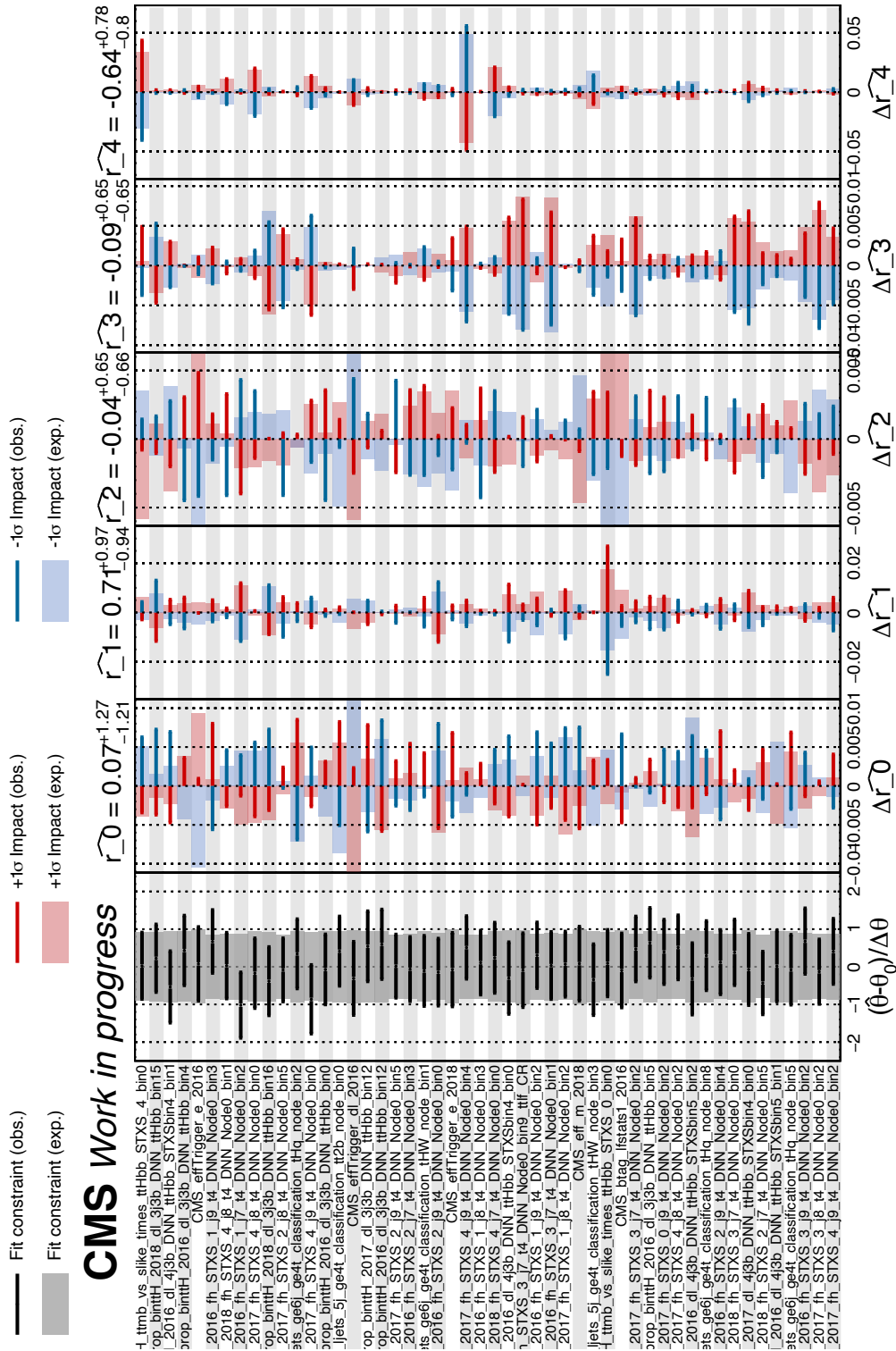


Figure L.33: (Continue) The impacts and their pulls and constraints on the signal strength obtained in the STXS full Run 2 combination fit to all channels, both in the fit to pseudo data with the signal process expected from the SM and in the observed data. The middle panel displays the post-fit pulls and constraints, while the right-hand side panels highlight the magnitude of the impact on the signal inference for both a pseudo data set corresponding to the SM (expected) and the data (observed).

Eidesstattliche Versicherung

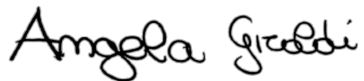
Hiermit versichere ich an Eides statt, die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt zu haben.

Die eingerichtete schriftliche Fassung entspricht der auf dem elektronischen Speichermedium.

Die Dissertation wurde in der vorgelegten oder einer ähnlichen Form nicht schon einmal in einem früheren Promotionsverfahren angenommen oder als ungenügend beurteilt.

Hamburg, February 15, 2023

Angela Giraldi

A handwritten signature in black ink, reading "Angela Giraldi". The signature is written in a cursive, flowing style with a large initial 'A'.